

# **Exploring the biosynthetic logic of myxobacterial natural products**

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

von  
**Apotheker Joachim Josef Hug**  
Saarbrücken

2020

**Tag des Kolloquiums:**

**23.09.2020**

**Dekan:**

**Prof. Dr. Guido Kickelbick**

**Berichterstatter:**

**Prof. Dr. Rolf Müller**

**Prof. Dr. Rolf W. Hartmann**

**Prof. Dr. Andreas Bechthold**

**Vorsitz:**

**Prof. Dr. Andriy Luzhetskyy**

**Akad. Mitarbeiter:**

**PD Dr. Martin Frotscher**

Die vorliegende Arbeit wurde von April 2016 bis März 2020 unter der Anleitung von Herrn Prof. Dr. Rolf Müller am Helmholtz-Institut für Pharmazeutische Forschung Saarland (HIPS) angefertigt.

## **„Energie geht nicht verloren“**

Hermann Ludwig Ferdinand von Helmholtz (1821 – 1894)



## Danksagung

Als erstes möchte ich mich bei meinem Doktorvater Prof. Dr. Rolf Müller für die herzliche Aufnahme in die Arbeitsgruppe und dem entgegengebrachten Vertrauen zur eigenständigen Bearbeitung der anspruchsvollen Projekte im Bereich der Naturstoffbiosynthese bedanken. Seine immerwährende Unterstützung und Diskussionsbereitschaft in wissenschaftlichen Belangen lieferten die Impulse um mich als Wissenschaftler weiterzuentwickeln.

Zudem möchte ich mich bei Prof. Dr. Rolf Hartmann für die fruchtbaren Diskussionen während des Promotionskomitees und für die Unterstützung als wissenschaftlicher Begleiter bedanken.

Ein besonderer Dank geht an meinen wissenschaftlichen Betreuer Dr. Daniel Krug für sein Engagement, die vielen hilfreichen wissenschaftlichen Diskussionen, die kritische Überarbeitung meiner Manuskripte und für die zahlreichen innovativen Lösungsansätze für die diversen Experimente während meiner Doktorarbeit.

Mein Dank gilt auch der gesamten Arbeitsgruppe die mir in den 4 Jahren mit wissenschaftlichen Rat und Tat zur Seite standen. Durch euch bin ich nicht nur fachlich, sondern auch persönlich gewachsen.

Dabei will ich mich zunächst bei meinen Büopartnern Dr. Domen Pogorevc und Dr. Bastien Schnell, Dr. Fabian Panter aus dem Analytik Team, und bei Dr. *in spe* Alexander Popoff bedanken; ihr vier hattet immer ein offenes Ohr gehabt, für alle Fragen theoretischer und praktischer Natur rundum die myxobakterielle Naturstoffforschung. Ihr vier habt mir einen guten Start in meine Promotion ermöglicht.

Ein besonderer Dank geht an Jan Dastbaz. Durch die Betreuung deiner Diplomarbeit und der späteren Zusammenarbeit haben wir uns beide wissenschaftlich weiterentwickelt. Mich hat es sehr gefreut, dir die Wissenschaft näher gebracht zu haben und den Weg zur derzeitigen Promotion.

Weiterhin möchte ich mich bei den sogenannten „Hektikern“ Dr. Sebastian Adam, Kris-Raimund Hippler, Manuel Neubert, Schoko-Nico und Vanille-Nico bedanken für die wissenschaftliche aber auch freizeitliche Unterstützung.

Natürlich möchte ich mich bei den vielen anderen Kollegen und Kolleginnen bedanken wie Patrick Haack, Chantal Bader, Sebastian Groß, Jan Schlemmer, Maja Remškar, Asfandiyar Sikandar, Sari Rasheed, Anna-Lena Huber, Anika Diederich, Dr. Ronald Garcia, Sebastian Walesch, Joy Birkelbach, Hu Zeng, Dr. Chengzhang Fu, Claudia Helbig, Dr. Carsten Volz, Dr. Cisco Torres aber auch allen die ich hier leider vergessen habe zu erwähnen, für die großartige Zusammenarbeit und Unterstützung in allen Belangen aber auch für das freundschaftliche Miteinander.

Abschließend möchte ich mich bei der gesamten Arbeitsgruppe für diese großartige Zeit bedanken. In dieser Zeit wart ihr immer an meiner Seite und ein Hoffnungsschimmer gegen jeglichen Dissertationsstress. Ich habe in euch Freunde gefunden.

Ein ganz besonderer Dank geht an meine Freunde aus Nah und Fern für die Ermutigungen und Zuversicht.

Zu guter Letzt gilt mein ganz ausdrücklicher Dank an meine Eltern Salma und Randolph sowie an meinen Bruder Alexander, für die fortwährende Unterstützung in allen Lebenslagen und auch für euer Verständnis. Euer Rückhalt haben all das erst ermöglicht.

## Summary

Myxobacteria produce a wide range of natural products exhibiting intriguing structures and potent biological activities. Natural products are formed by a variety of biosynthetic machineries. The genetic information for those biosynthetic machineries is encoded in the genome of each bacterium. Genome sequencing of myxobacteria has revealed numerous biosynthetic pathways comprised of biosynthetic proteins which to date have not yet been associated with their corresponding natural products. This thesis covers the identification and exploration of the biosynthetic origin of six myxobacterial natural product classes, thereby employing diverse concepts and methods of natural product research. The biosyntheses of natural products were investigated through mass spectrometry and genomics-guided techniques, stable-isotope-labeled feeding experiments combined with *in silico* analysis, genetic engineering of myxobacteria, biochemical characterization of biosynthetic proteins and structure elucidation of natural products using spectroscopic methods. These efforts expanded the knowledge of myxobacterial natural product biosynthesis and revealed unprecedented biosynthetic pathways.

## **Zusammenfassung**

Myxobakterien produzieren eine Vielzahl von Naturstoffen, die beeindruckende Strukturen und potente biologische Aktivitäten aufweisen. Naturstoffe werden von einer Vielzahl von biosynthetischen Maschinerien produziert. Die genetische Information um diese biosynthetischen Maschinerien zu generieren ist in dem Genom des jeweiligen Bakteriums enkodiert. Genomsequenzierungen von Myxobakterien offenbarten zahlreiche biosynthetische Stoffwechselwege, welche aus biosynthetischen Proteinen bestehen und bis heute zu keinen korrespondierenden Naturstoffen zugeordnet werden konnten. Die vorliegende Dissertation umfasst die Identifizierung und Erforschung der biosynthetischen Ursprünge von sechs myxobakteriellen Naturstoffklassen und dementsprechend wurden diverse Konzepte und Methoden der Naturstoffforschung angewandt. Die Biosynthesen von Naturstoffen wurde durch massenspektrometrischen und genomgeleiteten Techniken, stabile isotoopenmarkierte Fütterungsexperimenten in Kombination mit *in silico* Analysen, gentechnischen Manipulationen von Myxobakterien, durch die Charakterisierung von biosynthetischen Proteinen und durch die Strukturaufklärung von Naturstoffen mit Hilfe spektroskopischer Methoden untersucht. Diese Arbeiten haben zur Erweiterung des Wissensstands der myxobakteriellen Naturstoffbiosynthese beigetragen und dabei beispiellose biosynthetische Stoffwechselwege offenbart.

## Vorveröffentlichungen der Dissertation

Teile dieser Arbeit wurden vorab mit Genehmigung der Naturwissenschaftlich-Technischen Fakultät III, vertreten durch den Mentor der Arbeit, in folgenden Beiträgen veröffentlicht oder sind derzeit in Vorbereitung zur Veröffentlichung:

### Publications

#### Research articles

Witte, S. N. R.\*, **Hug, J. J.\***, Géraldy, M., Müller, R. & Kalesse, M. Biosynthesis and Total Synthesis of Pyrronazol B: a Secondary Metabolite from *Nannocystis pusilla*. *Chem. Eur. J.*, **2017**, 23, 15917–15921.

**Hug, J. J.**, Panter, F., Krug, D. & Müller, R. Genome mining reveals uncommon alkylpyrones as type III PKS products from myxobacteria. *J. Ind. Microbiol. Biotechnol.* **2019**, 46, 319–334.

**Hug, J. J.\***, Dastbaz, J.\*, Adam, S., Revermann, O. Koehnke, J. Krug, D. & Müller, R. Biosynthesis of cittelins, unusual ribosomally synthesized and post-translationally modified peptides from *Myxococcus xanthus*. *bioRxiv* preprint. **2020**. *ACS Chem. Biol.* **2020**, submitted manuscript.

Okoth D. A.\*, **Hug, J. J.\***, Garcia, R., Sproer, C., Overmann, J. & Müller, R. 2-Hydroxysorangiadenosine: Structure and Biosynthesis of a Myxobacterial Sesquiterpene–Nucleoside. *Molecules* **2020**, accepted manuscript.

Popoff, A.\*, **Hug, J. J.\***, Walesch, S., Garcia, R. & Müller, R. Structure and Biosynthesis of Myxolipoxazoles and Myxopyrimidinols: Unique Myxobacterial Fatty Acids Featuring Isoxazole or 4-Pyrimidinol Heterocycles. *Manuscript ready for submission.* **2020**.

(\*: Authors contributed equally to the work)

#### Review articles

**Hug, J. J.**, Bader, C. D., Remškar, M., Cirnski, K. & Müller, R. Concepts and Methods to Access Novel Antibiotics from Actinomycetes. *Antibiotics* **2018**, 7 (2), 44.

Huo, L., **Hug, J. J.**, Fu, C.; Bian, X.; Zhang, Y.; Müller, R. Heterologous expression of bacterial natural product biosynthetic pathways. *Nat. Prod. Rep.* **2019**, 36, 1412–1436.

**Hug, J. J.**, Krug, D., Müller, R. Bacteria as genetically programmable producers of bioactive natural products. *Nat. Rev. Chem.* **2020**, 4, 172–193.

**Hug, J. J.**, Müller, R. Host Development for Heterologous Expression and Biosynthetic Studies of Myxobacterial Natural Products: *Comprehensive Natural Products III: Chemistry and Biology, Chapter 14818.* **2020**, in press.

## Conference Contributions (Posters and Oral Presentations)

**Hug, J. J.**, Panter, F., Krug, D. & Müller, R. Genomic investigation and activation of myxobacterial type III PKS gene clusters. **Poster Presentation**, 8<sup>th</sup> International HIPS Symposium, **2018**: Saarbrücken, **Germany**.

**Hug, J. J.**, Panter, F., Krug, D. & Müller, R. Genomic investigation and activation of myxobacterial type III PKS gene clusters. **Poster Presentation**, International VAAM Workshop, **2018**: Frankfurt am Main, **Germany**.

**Hug, J. J.**, Panter, F., Krug, D. & Müller, R. Self-resistance guided genome mining reveals novel topoisomerase inhibitors from myxobacteria. **Poster Presentation**, Copenhagen Bioscience Conference 16 - Natural Products: Discovery, Biosynthesis and Application, **2019**: Copenhagen, **Denmark**.

**Hug, J. J.**, Panter, F., Krug, D. & Müller, R. Self-resistance guided genome mining reveals novel topoisomerase inhibitors from myxobacteria. **Poster Presentation**, 10<sup>th</sup> International HIPS Symposium, **2019**: Saarbrücken, **Germany**.

**Hug, J. J.**, Panter, F., Krug, D. & Müller, R. Genomic investigation and activation of myxobacterial type III PKS gene clusters. **Poster Presentation**, International VAAM Workshop, **2019**: Jena, **Germany**.

Dastbaz, J., **Hug, J. J.**, Adam, S., Revermann, O. Koehnke, J. Krug, D. & Müller, R. Studies on the biosynthesis of cittelil, an unusual secondary metabolite from *Myxococcus xanthus*. **Poster Presentation**, International VAAM Workshop, **2019**: Jena, **Germany**.

## Table of Contents

<b>Chapter 1</b>	<b>Introduction</b>	<b>1</b>
<b>Chapter 2</b>	<b>Biosynthesis and Total Synthesis of Pyrronazol B</b> Biosynthesis and Total Synthesis of Pyrronazol B: a Secondary Metabolite from <i>Nannocystis pusilla</i>	<b>71</b>
<b>Chapter 3</b>	<b>Biosynthesis of Cittelins</b> Biosynthesis of cittelins, unusual ribosomally synthesized and post-translationally modified peptides from <i>Myxococcus xanthus</i>	<b>131</b>
<b>Chapter 4</b>	<b>2-Hydroxysorangadenosine: Structure and Biosynthesis</b> 2-Hydroxysorangadenosine: Structure and Biosynthesis of a Myxobacterial Sesquiterpene–Nucleoside	<b>213</b>
<b>Chapter 5</b>	<b>Biosynthesis of Myxolipoxazoles and Myxopyrimidinols</b> Structure and Biosynthesis of Myxolipoxazoles and Myxopyrimidinols: Unique Myxobacterial Fatty Acids Featuring Isoxazole or 4-Pyrimidinol Heterocycles	<b>267</b>
<b>Chapter 6</b>	<b>Type III PKS Products from Myxobacteria</b> Genome mining reveals uncommon alkylpyrones as type III PKS products from myxobacteria	<b>339</b>
<b>Chapter 7</b>	<b>Discussion and Outlook</b>	<b>407</b>

# Chapter 1

## 1 Introduction

### 1.1 Natural Products

Natural products are defined in the broadest sense as the entirety of biological molecules synthesized by living organisms<sup>1</sup>, whereas a more specific definition would refer only to molecules of low molecular weight (in general smaller than 3000 Da) which are not essential for the growth and development of the producing organism but offer evolutionary advantages under particular environmental conditions<sup>2</sup>. Beyond the basic biochemistry of primary metabolism including biomolecules such as nucleic acids, proteins, carbohydrates, lipids, and small molecules, which provide evolutionary conserved metabolites required for the growth and maintenance of cellular functions, many bacteria produce non-essential secondary metabolites, also termed natural products. These natural products feature inherent biological functions and thus have been an important source of therapeutics<sup>3</sup>.

Historically, the exploration of natural products mostly focused on plant-derived compounds for the treatment of wounds and diseases like the usage of oils from *Cupressus sempervirens* (cypress) and *Commiphora species* (myrrh), which are still in use nowadays to treat colds, coughs and inflammations<sup>4</sup>. Throughout human history, the use of natural products was closely connected to trial-and-error studies of various plants and animal parts, resulting in an extensive documentation of complex prescriptions to treat medical conditions<sup>5</sup>. The development of modern medicine parallels the transition from empirical preparation of well-known medicinal plants or their extracts to the use of the isolated natural product. Friedrich Sertürner initiated this transition in 1803, when he isolated the active ingredient morphine (along with several other alkaloids) from the opium poppy of *Papaver somniferum* L.<sup>6</sup>. In 1820 Pierre Joseph Pelletier and Joseph Bienaimé Caventou isolated quinine from the bark of *Cinchona officinalis* as an active natural product<sup>7,8</sup>, which was later intensively used for malaria treatment<sup>9</sup>. Another example is the isolation of the natural product salicin in 1828 from the bark of the willow tree *Salix alba* L. by Johann Andreas Buchner<sup>10</sup>. This was followed by the derivatization of the active metabolite salicylic acid to acetylsalicylic acid in 1853 by Charles Frédéric Gerhardt, setting the stage for production of the world-wide used painkiller Aspirin<sup>11</sup>.

One hallmark of natural product based medicine was the discovery of the  $\beta$ -lactam antibiotic penicillin in the broth filtrate of *Penicillium chrysogenum* in 1928 by Alexander Fleming<sup>12</sup>, which was one of the first known antibiotics (next to the synthetic antibiotics arsphenamine (1910)<sup>13</sup> and prontosil (1935)<sup>14</sup>). This discovery resulted in a paradigm shift because it emphasized that invaluable natural products do not exclusively originate from plants but also from microorganisms<sup>15</sup>. In addition, the discovery of streptomycin by Albert I. Schatz and Selman A. Waksman in 1943 from the soil dwelling bacterium, *Streptomyces griseus*, along with the clinical usage of penicillin in 1945 heralded the start of the “golden era” of anti-infective drug development lasting until the last quarter of the 20th century. During this time, several new classes of antibiotics originating from microorganisms were discovered, highlighting the indispensable value of natural products from microorganisms for clinical usage. The impact of

natural products on drug development and therapeutics is reflected by the origin of newly approved drugs: Almost 50% of all new approved drugs between 1981–2019 are either natural products, mimics thereof or feature a natural product pharmacophore (excluding biologicals and vaccines)<sup>3</sup>. Many natural products exhibit potent biological activity, including antibacterial activity (*e.g.* penicillin, vancomycin, erythromycin, rifamycin)<sup>16</sup>, cytotoxic activity (*e.g.* bleomycin, doxorubicin)<sup>17,18</sup>, antifungal activity (*e.g.* amphotericin, griseofulvin<sup>19,20</sup>), immunosuppressive effects (*e.g.* rapamycin, cyclosporine)<sup>21–24</sup>, among several other biological activities<sup>30</sup> (**Figure 1**). However, the end of the golden era of antibiotic drug development resulted in the gradual withdrawal of pharmaceutical companies from natural product research; this was primarily due to low hit rates and the dereplication issues – the strongly increasing rediscovery of known bioactive compounds<sup>25</sup>.

This withdrawal of pharmaceutical industry from natural product research arose in parallel with increasing emergence of antimicrobial resistance (AMR), which may evolve to be the greatest threat to prevention and treatment of an increasing number of infections<sup>26</sup>. The world health organization (WHO) released an action plan on global antibiotic resistance, the global antimicrobial surveillance system (GLASS), in order to track the resistance-related issues of antibiotics in use for treatment of hospital- and community-acquired infections across the participating countries<sup>27</sup>. Of particular interest are the so-called ESKAPE pathogens, which include *Enterococcus faecium*, *Staphylococcus aureus*, *Klebsiella pneumonia*, *Acinetobacter baumannii*, *Pseudomonas aeruginosa* and *Enterobacter* spp., since these pathogens are becoming increasingly resistant to most commonly prescribed antibiotics in hospitals<sup>28</sup>. In addition, several recent reports have also shown that these pathogens are becoming resistant even to antibiotics of the last resort, possibly leading to the worst-case scenario of a post-antibiotic era, pictured similarly to the time of the pre-antibiotic era marked by high mortality rate from infectious diseases. A review commissioned by the United Kingdom government entitled, “Antimicrobial Resistance: Tackling a crisis for the health and wealth of nations” (the AMR Review)<sup>29</sup>, estimated that AMR could cause 10 million deaths a year by 2050, which would exceed the number of deaths attributable to cancer disease. Although precise prediction of potential causalities have been criticized in the literature<sup>30</sup>, AMR is unquestionably jeopardizing clinical and public health of a vast majority in the developed and developing world.

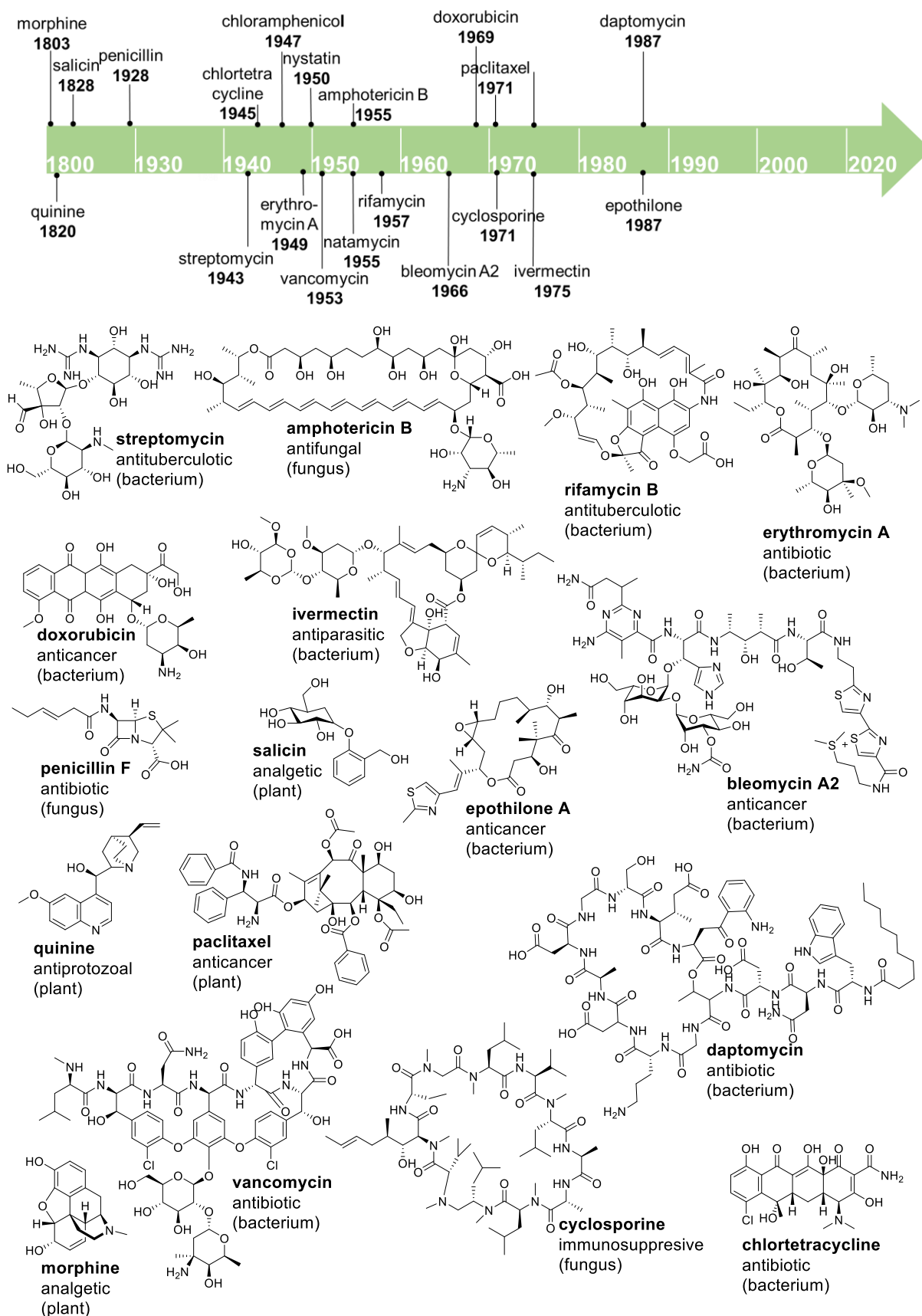
Despite the fact that AMR is undoubtedly accelerated by the extensive misuse of antibiotics in clinics and animal husbandry, the origin of AMR is much older than the implementation of antibiotic treatment in human and animal health<sup>31</sup>. In fact, antibiotic resistance among bacterial species is widespread in environmental bacteria, and those genes conferring resistance may be transferred from non-disease-causing bacteria to potentially pathogenic species, which are contributing to the present significant AMR<sup>32</sup>. As an example, soil-dwelling bacteria are living in competition with other bacterial species, and the production of chemical warfare such as antibiotics and the countermeasures in form of resistance mechanisms likely reflect common phenomena under specific environmental conditions<sup>33,34</sup>. Other studies suggest that bacteria are producing antibiotic natural products to defend their symbiotic eukaryotic host<sup>35</sup>. An impressive case highlighting this ancient environmental resistome was shown by the bacterial isolate *Paenibacillus* sp. LC231 that featured antibiotic-resistance to 26 of 40 antibiotics tested through five new resistance mechanisms and 12 orthologues of known resistance gene families<sup>36</sup>.



In accordance with the development of the environmental resistome, AMR in hospitals arises from anthropogenic overuse of synthetic and natural product-derived antimicrobials<sup>37</sup>. This misuse of antimicrobials induces natural selection – resembling the key process of evolution – leading to the prevalence of drug-resistance pathogens<sup>38</sup>. Bacterial pathogens are capable of expeditious evolutionary adaption due to their comparatively short doubling times. Beneficial genetic mutation of specific drug target genes and the rapidly exchange of genetic information among different bacterial species (known as horizontal gene transfer) through small extrachromosomal DNA molecules called plasmids, are important molecular mechanisms accounting for clinical relevant AMR<sup>39</sup>. As a result, pathogenic microorganisms, environmental bacteria as well as bacteria producing antibacterial natural products can curiously contribute to the immense AMR crisis due to extensive overuse and misuse of (natural product-derived) antibiotics<sup>40</sup>. In addition, soil-dwelling bacteria producing potent antibacterial natural products developed specific self-resistance mechanisms to prevent self-toxicity<sup>41</sup>; these self-resistance mechanisms might be in theory genetically transferred to other bacterial pathogens. This intense cross-linking between AMR, the biosynthesis of natural products and their mode of actions implies that under-investigated microbial sources of antimicrobials might help to overcome AMR or at least provide in-depth understanding of the underlying molecular mechanisms.

The fundamental information on how bacteria form biosynthetic enzymes to produce natural products is inscribed in the genome of each bacterium, and the set of genes responsible for the biosynthesis are typically clustered, which are referred to as biosynthetic gene clusters (BGCs). Together with the increasing availability of genomic data, this finding initiated a renewal in the field of natural product research as it enabled genomics-based investigations of bacteria. Among different genera of bacteria, the order actinomycetales has been proven to bear valuable producers of antibiotic, antifungal, antiprotazoal and antiviral natural products<sup>42</sup>. More than 5000 antimicrobials have been identified from the order actinomycetales, of which over 90% are produced solely by the genus *Streptomyces*. Prominent examples are the polyene antimycotics nystatin<sup>43</sup>, amphotericin B<sup>20</sup> and natamycin<sup>44</sup>, while chloramphenicol<sup>45</sup>, daptomycin<sup>46</sup> and chlortetracycline are the most prominent antibacterial examples<sup>47</sup>. Further biological activities are featured by natural products isolated from *Streptomyces* including the antiparasitic drug ivermectin<sup>48</sup> and the antineoplastic compound bleomycin<sup>18</sup>.

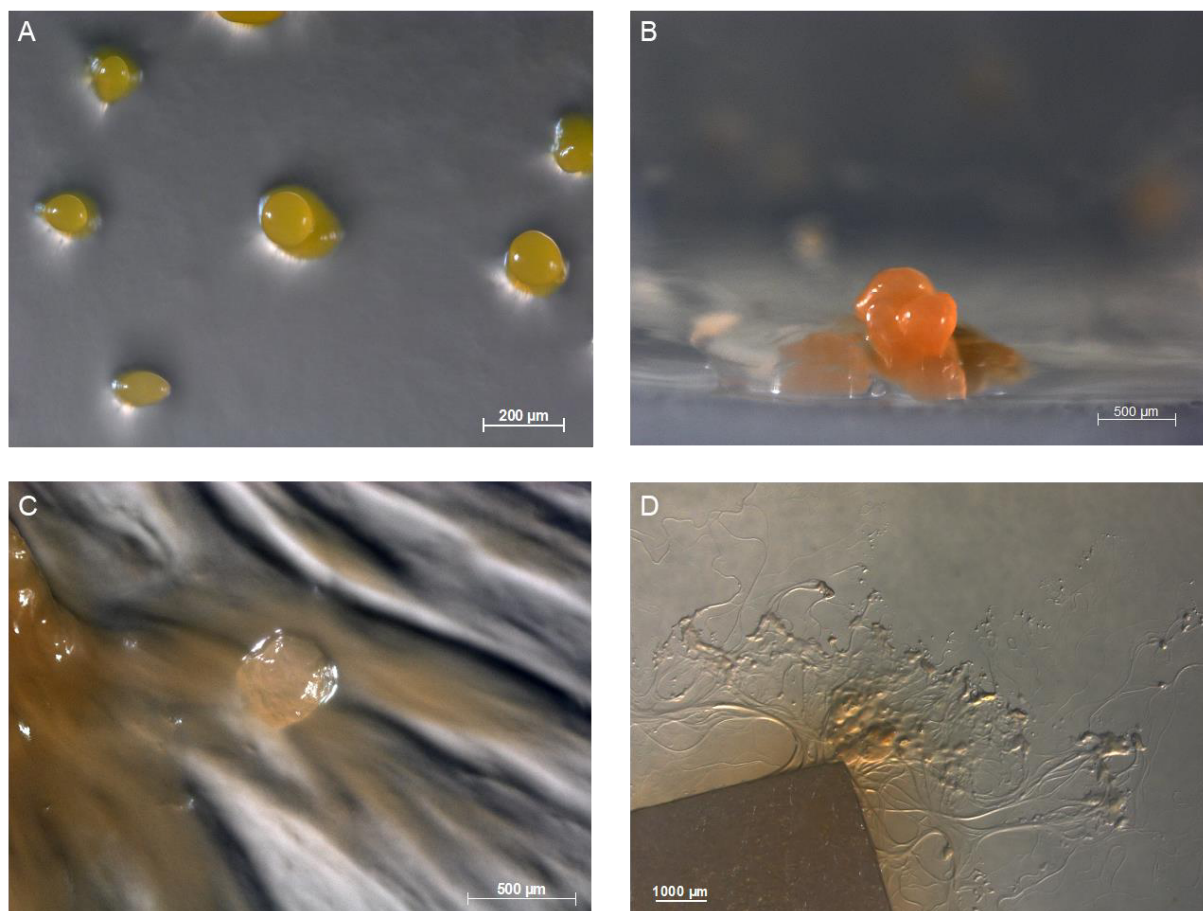
Nevertheless, the discovery of natural product-based drugs derived from streptomycetes has continued to decrease, shown by a 30% drop of natural product-based drugs in clinical studies between 2001 and 2008<sup>26,49</sup>. Encouragingly, over the past few decades under-investigated microbes including rare actinomycetes, cyanobacteria, plant endosymbionts, insect pathogenic bacteria and myxobacteria have entered the stage as prolific producers of bioactive natural products.



**Figure 1.** Discovery timeline of natural products and the chemical structures of some of the important natural products. The biological activity associated with each compound and the origin (in brackets) are mentioned.

## 1.2 Myxobacteria – Underexploited source of natural products

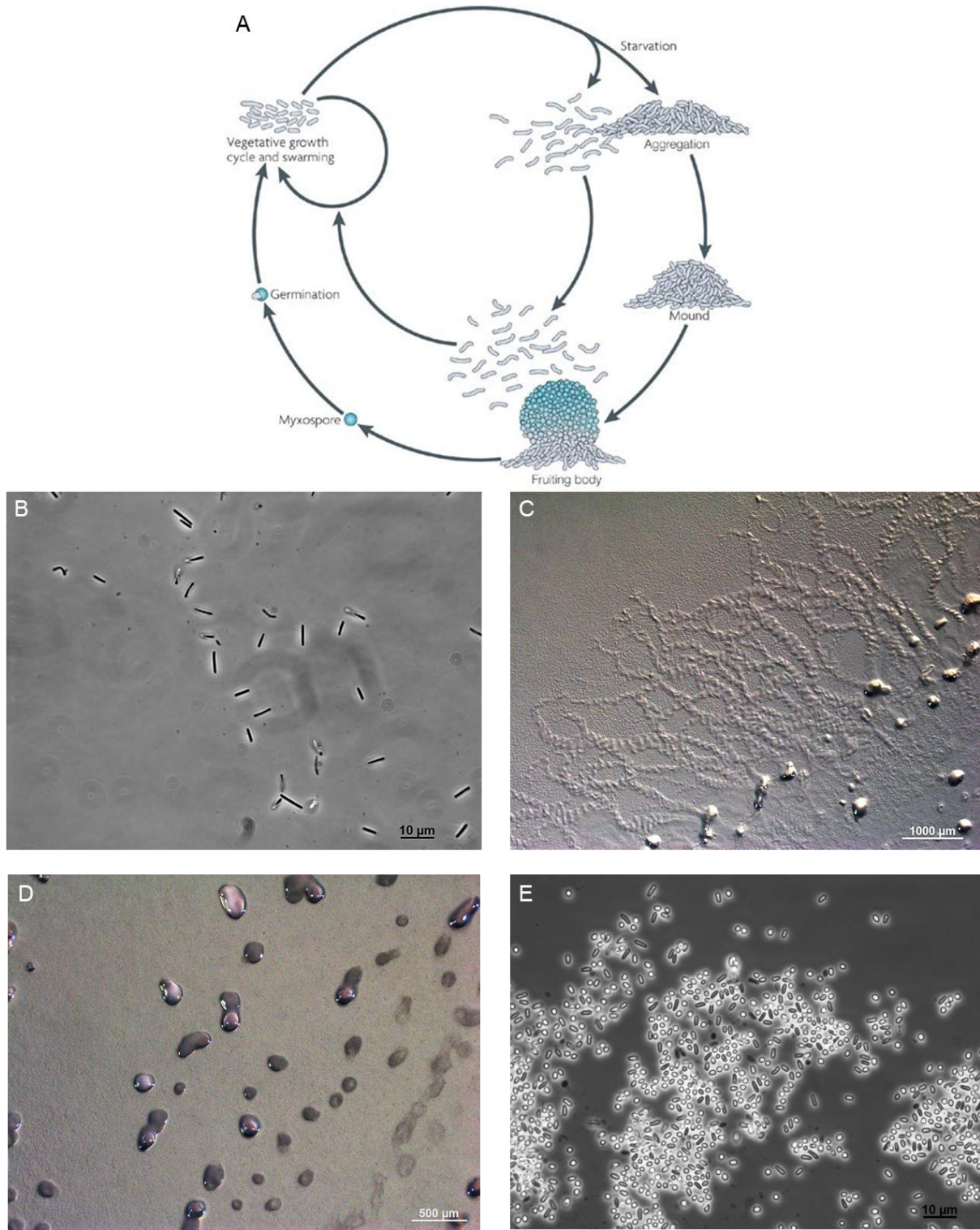
Myxobacteria are rod-shaped Gram-negative microorganisms belonging to the class of  $\delta$ -proteobacteria, which are phylogenetically united in the order *Myxococcales*. These bacteria are ubiquitously distributed in soil, where these microorganisms have developed several unique characteristics, such as a cooperative “social behavior” based on complex chemical communication systems and complex multicellular development stages, resembling the complexity observed in macroscopic organisms<sup>50</sup>. In addition to their particular life cycle, myxobacteria also feature unique characteristics such as gliding motility on surfaces and synchronized directional swarming based on chemical communication. Unlike other bacteria, the observed movement is not based on flagella but explained by two different motility systems<sup>51</sup>. The social (S)-motility system (also called twitching<sup>52</sup>) describes the movement of cell groups. The adventurous gliding and gliding transducer (Agl–Glt) system<sup>53</sup> on the other hand describes the individual single cell movement (also simply termed gliding)<sup>52</sup>. Myxobacteria are able to consume different biological macromolecules like cellulose as well as prey on microorganisms such as bacteria and yeast<sup>50</sup>. This nutritional characteristic concerning the degradation of biomacromolecules allows categorization of the order of *Myxococcales* into predators of other prokaryotes and eukaryotes, and cellulose-decomposers. Most myxobacteria known to date are classified as cooperative predators exhibiting wolf-pack predatory behaviour<sup>54</sup> comprising hunting, killing and the degradation of other microorganisms as prey via secretion of exoenzymes to lyse bacteria and yeasts to make use of the lysed material as nutrients<sup>55</sup>. These exoenzymes form holes in the cell wall of fungi and green algae, which allows the entry of myxobacteria into these “cracked” cells<sup>56</sup>. In contrast, only a minority of myxobacteria are cellulose-decomposers, represented by the genera *Sorangium* and *Byssovorax*<sup>57</sup>. Similar to some fungi, myxobacteria form complex multicellular fruiting bodies under starvation conditions<sup>58</sup> (**Figure 2**), via conversion of the vegetative cells into myxospores<sup>59</sup>, allowing the bacteria to withstand unfavorable environmental influences. As soon as myxobacteria face limited access to nutrition, hundred thousands of vegetative rod cells initiate a developmental program and exchange extracellular and physical contact signals to form cell aggregates. Afterwards these cell aggregates build up to 2 mm small diverse colorful and elaborate structures called fruiting bodies (**Figure 2**). The myxobacterial life cycle culminates in the development of dormant heat- and desiccation-resistant myxospores, which are stored within fruiting bodies to germinate and form a swarm colony upon improved nutrient conditions (**Figure 3**)<sup>60</sup>. The colors of the cells and fruiting bodies vary from yellow, orange, red, brown, milky to even black<sup>61</sup>. These colors are a result of carotenoid and melanoid pigments, providing protection against photooxidation<sup>62</sup>. In addition to these different developmental stages of myxobacteria, some strains such as *M. xanthus* sp. undergo phase variation to produce two colony phenotypes. The majority of colonies are classified as wild type yellow variant, which feature the characteristic color of *M. xanthus* sp. cells from the accumulation of the natural product DKxanthene<sup>63</sup> and produce the antibiotic myxovirescin that is required for sporulation and predation<sup>64</sup>; whereas, a minority features a “white” wild type tan variant. This tan variant usually lacks the ability to form mature fruiting bodies leading to the generation of fewer spores and display upregulation of iron acquisition systems culminating in higher production of myxochelins<sup>65</sup>. The discovery of homospermidine lipids produced by *M. xanthus* DK1622, exclusively during fruiting body formation, show the tight correlation between the developmental stage of myxobacteria and their secondary metabolism. The challenging task to isolate these homospermidine lipids from solid agar plates of *M. xanthus* DK1622 could be avoided, since metabolome analysis resulted in the identification of *Myxococcus fulvus* MCy9290 as an alternative producer of these homospermidine lipids in liquid culture as vegetative cells<sup>66</sup>.



**Figure 2.** Fruiting bodies of myxobacteria. (A) Fruiting bodies of *M. xanthus* sp. on agar (B) *Sorangiineae* SBSr124 fruiting body on agar (C) *Vitiosangium* sp. fruiting body on agar (D) *Sorangium cellulosum* So ce56 on filter paper. Figures courtesy of Dr. Ronald Garcia.

The tight connection between the production of secondary metabolites and cellular development in myxobacteria, shows the indispensable value of these natural products alongside their unique biological characteristics. Although myxobacterial species are significantly less investigated compared to other microorganisms, many bioactive compounds (several hundreds) have already been isolated from this family, some of which are shown in figure 4. Moreover, these compounds tend to display mode of actions which are uncommon for natural products isolated from other microorganisms<sup>67–69</sup>. Taken together these findings place myxobacteria as one of the most prolific producers of natural products next to actinomycetes<sup>70</sup>, *Bacillus* sp.<sup>71</sup> and fungi<sup>72</sup>. To date a wide variety of mode of actions have been documented for myxobacterial secondary metabolites with diverse bioactivities *e.g.* antifungal (soraphen A<sup>73</sup>, ambruticin<sup>74</sup>), immunomodulatory (argyrins<sup>75</sup>), antiparasitic (chlorotoni<sup>76</sup>), antibacterial (cystobactamid<sup>77</sup>, myxopyronin<sup>78</sup>), cytotoxic (epothilone<sup>79</sup>), antiviral (aetheramide<sup>80</sup>) and antiparasitic (*e.g.* corallopyronin<sup>81</sup>). Many myxobacterial natural products display antifungal bioactivity such as the myxobacterial cyclopropyl-polyene ambruticin, which was one of the first myxobacterial compounds isolated<sup>74</sup>. Ambruticin exerts its antifungal activity via a common mode of action, *i.e.* interfering with osmoregulation via the high-osmolarity glycerol (HOG) signaling pathway, which is shared among fungicides including phenylpyrroles (*e.g.* pyrrolnitrin and fludioxonil), dicarboximides (*e.g.* vinclozolin, iprodione, and procymidone), and aromatic hydrocarbons (*e.g.* chloroneb and quintozone)<sup>82</sup>.





**Figure 3.** A) *Myxococcus xanthus* life cycle as described by Zusman *et al.*<sup>83</sup>. Upon starvation, cells aggregate into macroscopic mounds, forming fruiting bodies in which the rod-shaped cells will undergo morphogenesis to form spherical myxospores. These myxospores can germinate when nutrient conditions improve. B–E) Developmental stages of *Myxococcus xanthus* Mx x48. (B) Vegetative cells as seen on a mounted slide. (C) Swarming cells of *M. xanthus* Mx x48. (D) Fruiting bodies of *M. xanthus* Mx x48 on agar. (E) Slide mount of *M. xanthus* Mx x48 spores on solid agar. Figures provided by Dr. Ronald Garcia.

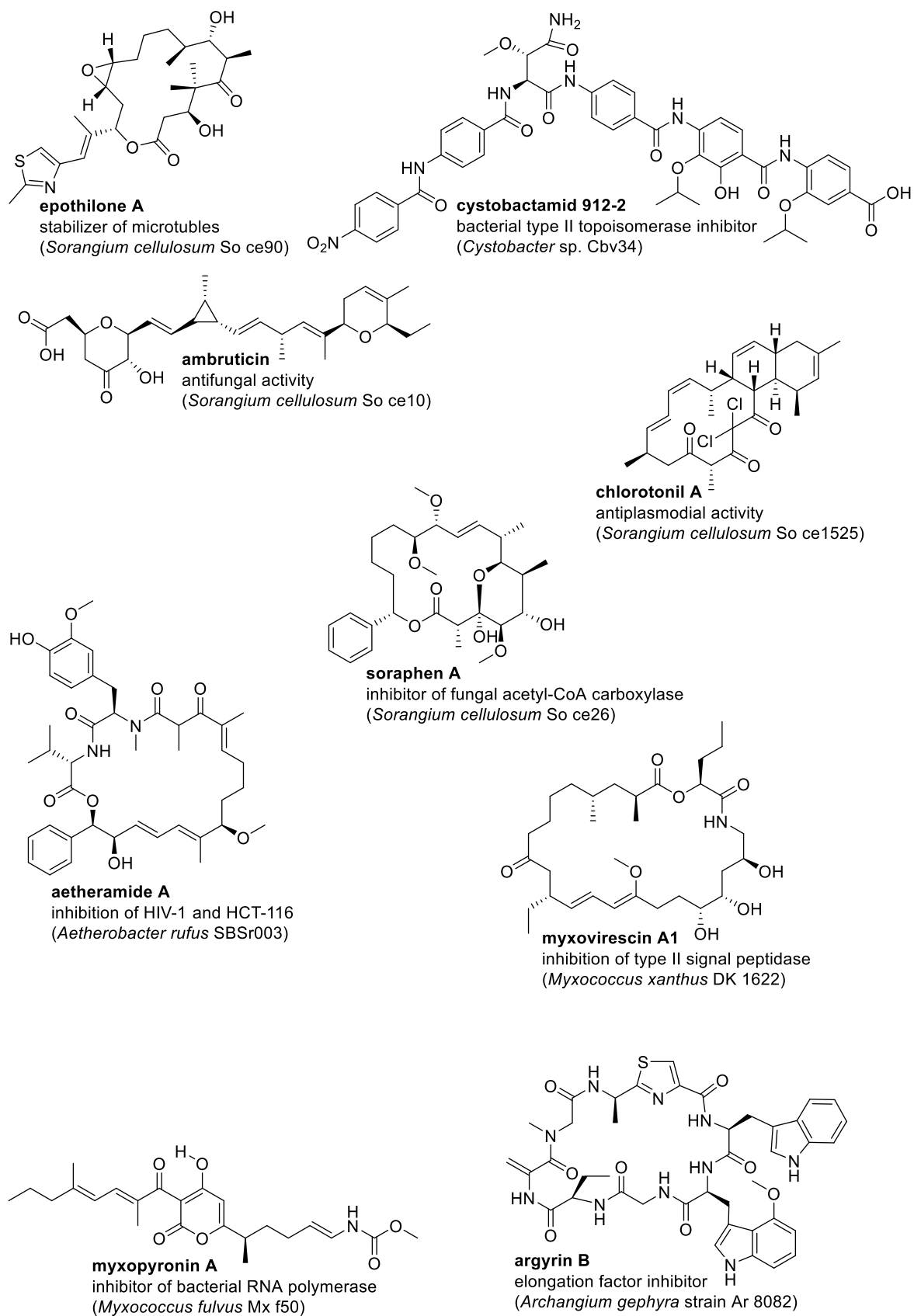
In contrast, soraphen A features a previously unknown antifungal target, the acetyl coenzyme A carboxylase (ACC) that has been identified from resistant *Saccharomyces cerevisiae*, revealing a mutation in a single genetic locus, the *acc1* gene<sup>84</sup>.

The  $\alpha$ -pyrone antibiotics, myxopyronin and coralpyronin, represent a promising compound class for the development of antibacterial agents including therapeutic applications against river blindness caused by *Wolbachia* bacteria<sup>81</sup>. Unlike commonly used inhibitors of bacterial RNA polymerase (RNAP), myxobacterial  $\alpha$ -pyrone antibiotics bind a novel target site on bacterial RNAP, the so-called “switch region”<sup>85</sup>. For this reason, these potential antibiotics are most likely active against rifampicin resistant bacteria. Similar to the  $\alpha$ -pyrone antibiotics, the peptide natural product cystobactamid 919-2 inhibits a well-known antibiotic target, the essential bacterial topoisomerase II (gyrase)<sup>86</sup>. Since no cross-resistance with other clinically used bacterial topoisomerase II inhibitors, such as the fluoroquinolones, has been observed, it is likely that the cystobactamids are exerting their potent antibacterial activity by binding to a unique site on the gyrase<sup>77</sup>.

One notable group of myxobacterial natural products are the epothilones, since the Food and Drug Administration (FDA) has already approved its semisynthetic derivative ixabepilone (Ixempra®) for breast cancer treatment in 2007<sup>87</sup>. Another group of myxobacterial natural product with promising potential for further drug development are the argyryns, a family of cyclic octapeptides with potent antimicrobial, antitumorigenic and immunosuppressant activities via inhibition of the elongation factor G (EF-G) in bacteria, and the homologue EF-G1 in yeast and mammalian cells<sup>88,89</sup>. Since the argyryns inhibit EF-G by a different binding site than the well-known EF-G inhibitor fusidic acid, some derivatives of argyryns are further developed as potential antibiotics with a new mode of protein synthesis inhibition<sup>88</sup>. Argryrin A might be used as a stabilizer of cyclin kinase inhibitor p27<sup>kip1</sup><sup>90,91</sup>, whereas, argyryrin F could be used as a therapeutic agent for pancreatic adenocarcinoma<sup>92</sup>.

Taken together, it can be stated that myxobacterial natural products are gifted secondary metabolites with diverse range of bioactivities and novel scaffolds, making them potential lead structures for future therapeutic agents. Encouragingly, the potential of myxobacteria as natural product source is not exhausted, since novel strains from ecological niches are still isolated frequently<sup>89,90</sup>. In particular, the investigation of myxobacterial strains from new genera rather than additional representatives within the same genus seems to be a good starting point to find new natural products as shown in a study of Hoffman *et al.* leading to the discovery of rowithocin A<sup>93</sup>. The emphasized correlation in the referenced study between taxonomic distance and the production of distinct natural product families supports focusing on under-investigated natural product producers, which are phylogenetically as distant as possible to other studied myxobacteria to isolate novel compounds.

Despite the various chemical structures of myxobacterial natural products, these complex molecules are produced from relatively “simple” monomeric building blocks deriving from primary metabolism. Previous molecular biological and biotechnological approaches have improved our knowledge of how Nature forms many of the intriguing structural moieties found in myxobacterial natural products, which will be further illustrated in the following subsection.



**Figure 4.** Chemical structure of unique myxobacterial natural products with potent biological activity.

## 1.3 Biosynthesis of natural products

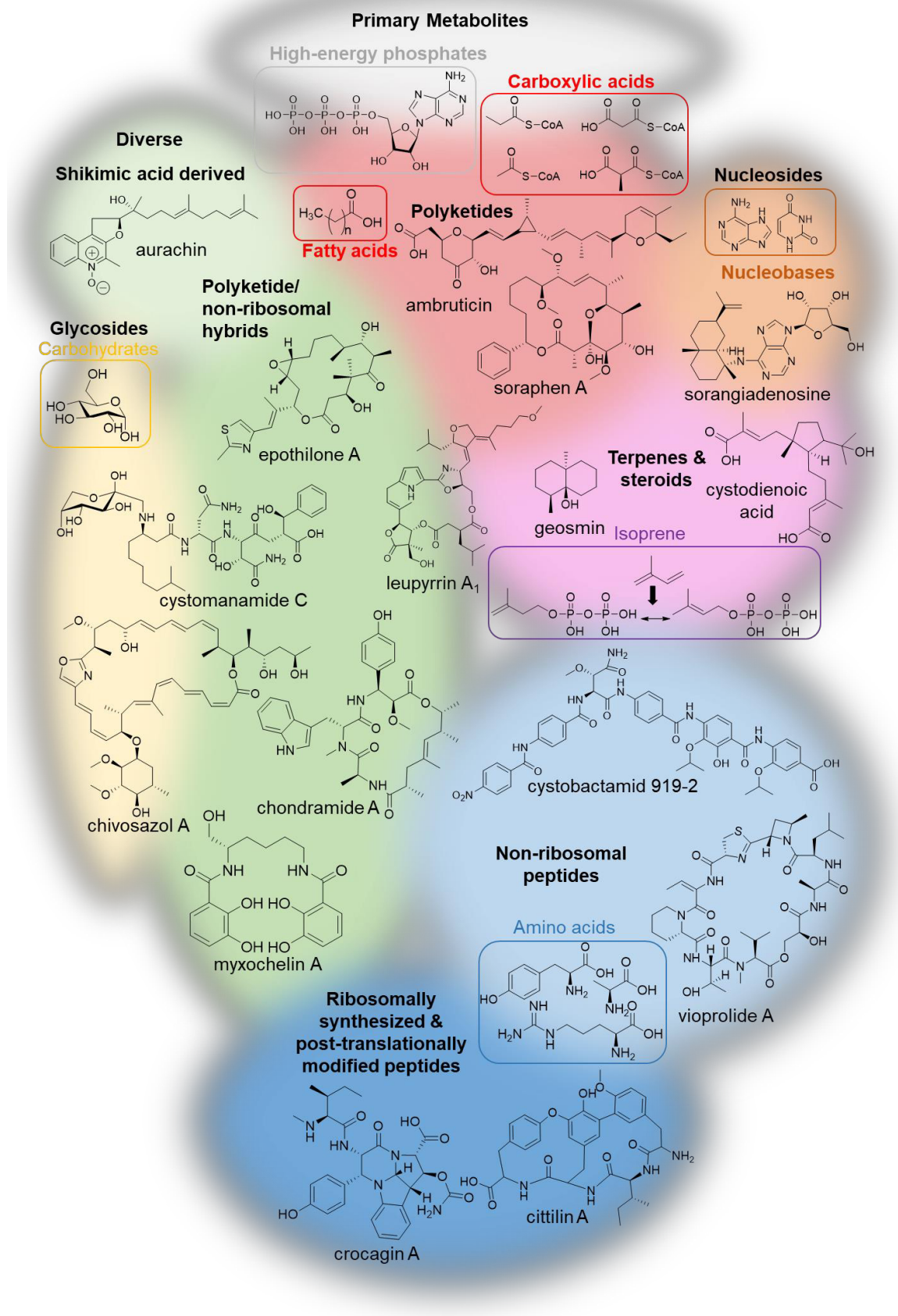
### 1.3.1 Chemical Space of Natural Products

The chemical scaffolds of natural products are considered as “privileged structures”<sup>94,95</sup>, since evolutionary optimization over millions of years for interactions with biological entities led to a diversity of chemical structures surpassing the chemical space of synthetic compounds<sup>96</sup> and natural products are more likely providing useful ligands for more than one receptor<sup>97</sup>. This optimization generated chemical structures with sophisticated stereochemistry, polycyclic structures and diverse biological functions with clinical relevance<sup>98</sup>. These evolutionarily optimized physiochemical properties enable natural products within a mass range of 200–3000 Da<sup>2</sup> to permeate cell membranes and bind their intracellular targets<sup>99</sup>, characteristics often missed in compound libraries.

Despite the highly diverse structures and biological functions, natural products can be classified into distinct groups based on their biosynthesis<sup>100</sup>. Generally, the biosynthesis of natural products incorporates a limited set of monomeric building blocks, originating from the primary metabolism, to assemble a variety of different secondary metabolites. Hence, the underlying biosynthetic pathway is classified and named after the type of building blocks incorporated into the chemical scaffold such as acetate, amino acids, isoprenoids, nucleosides *etc.* The three main producers of natural products namely plants, fungi and eubacteria, feature significant differences in their synthesized secondary metabolites concerning their structure and biosynthesis. Plants are producing mainly alkaloids (small nitrogen-rich natural products formed by using amino acids<sup>101</sup>), phenylpropanoids, terpenoids and polyketides<sup>102</sup>, whereas fungi produce additionally ribosomal peptides<sup>103</sup> and non-ribosomal peptides<sup>104</sup>. Bacteria in contrast provide a great variety of natural products deriving from numerous biosynthetic pathways, such as shown for myxobacteria (**Figure 5**). The occupied chemical space of myxobacterial natural products include linear and cyclic polyketides and polypeptides, alkaloids and terpenoids – although a few glycosylated natural products have also been isolated (**Figure 5**). In particular, the frequently observed hybrid structures of myxobacterial natural products originating from the linkage of carboxylic acids and amino acids mark a significant characteristic in the chemical space of natural products<sup>105,106</sup>. Within the order of *Myxococcales*, several secondary metabolites seem to feature a conserved function such as the iron transport metabolites namely the hydroxamate-type nannochelins<sup>107</sup> and the catecholate-type myxochelins<sup>108–110</sup>, cerebroside, ceramides, and carotenoid glycosides esterified with fatty acids<sup>111,112</sup>. Another common compound among myxobacteria is the sesquiterpenoid geosmin accounting for the earth smell of myxobacterial broth cultures<sup>113</sup>. Furthermore the analysis of *Stigmatella aurantiaca*<sup>114</sup> and *M. xanthus*<sup>115</sup> plate cultures revealed a large number of volatile substances from different compound classes, including ketones, esters, lactones, sulfur and nitrogen-containing molecules, and additional terpenes. The capability of bacteria to produce steroids such as cholesterol<sup>116</sup> and lanosterol<sup>111</sup> underlines the biochemical similarity of myxobacterial pathways to those of higher organisms such as fungi<sup>117</sup>.

In order to understand the biosynthetic logic of myxobacterial natural products investigated in this thesis, the following subsections elaborate on the underlying biosynthetic pathways responsible for the structural diversity of myxobacterial secondary metabolites.





**Figure 5.** Chemical space of myxobacterial natural products. Scheme adapted from Hug *et al.*<sup>118</sup>.

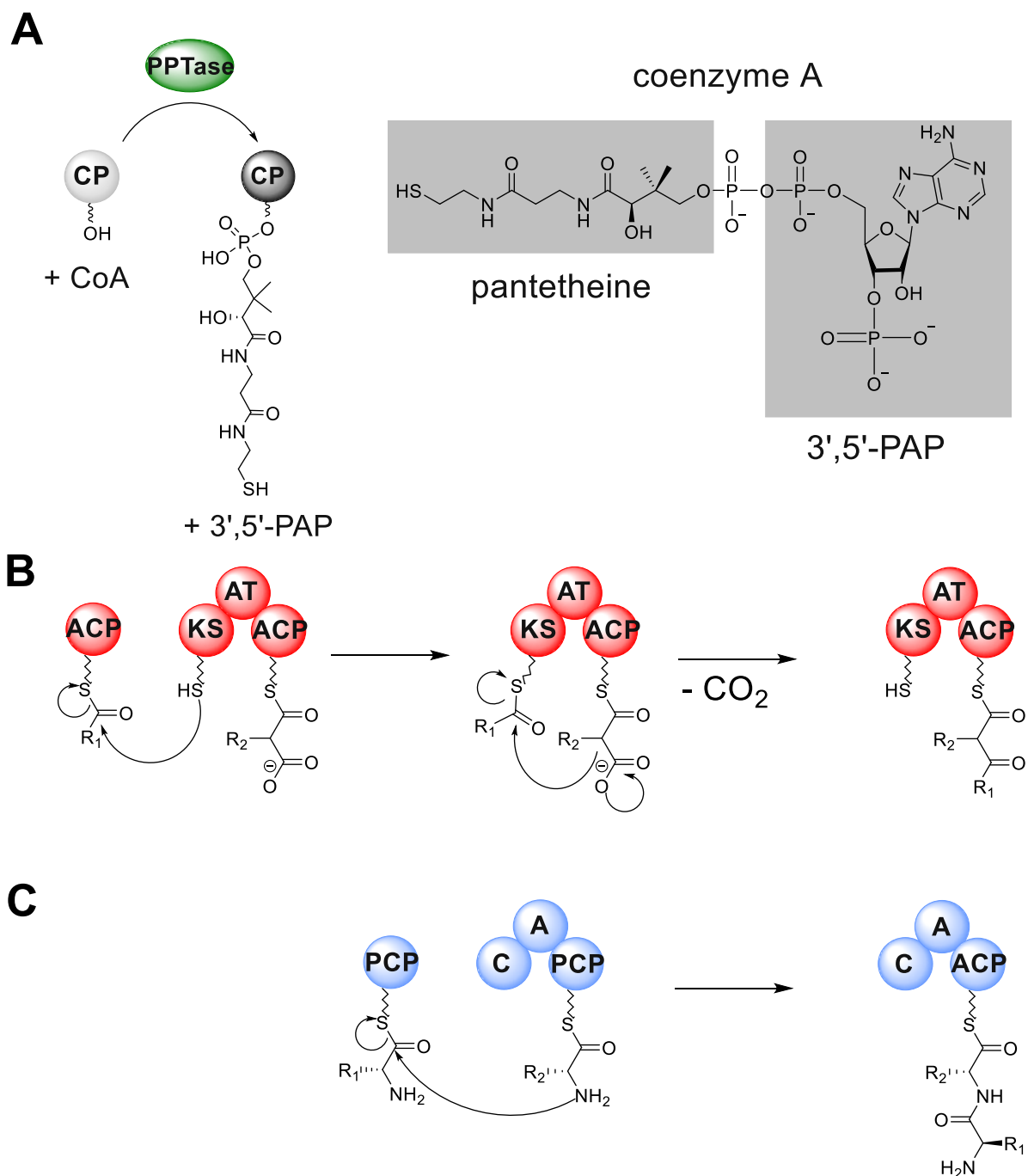
### 1.3.2 Multimodular Biosynthetic Pathways

Many myxobacterial natural products are produced by large, modular, multifunctional enzymes known as polyketide synthases (PKS) and non-ribosomal peptide synthetases (NRPS) forming polyketides, non-ribosomal peptides and their hybrids. Despite the fact, that both biosynthetic machineries show differences, such as the incorporated building blocks, activation of substrates and condensation mechanisms, both classes of enzymes share fundamental similarities resulting in hybrid PKS-NRPS biosynthetic pathways, which are frequently found in myxobacteria<sup>119</sup>. These large biosynthetic machineries are composed of catalytic domains fused in a linear manner, wherein each domain is responsible for catalyzing a single reaction step during the assembly of the polyketide or peptide backbone. During polyketide and non-ribosomal peptide biosynthesis, repetitive catalytic units called modules facilitate the incorporation of specific monomeric building blocks into the extending polyketide, polypeptide or hybrid chain. Each module consists of catalytic subunits called domains, which are in charge of individual biosynthetic steps such as loading, condensation, optional modification and release of the processed backbone. These chain elongation reactions are essential for the assembly line to further process the covalently bound intermediates. The nascent intermediates are covalently bound by carrier protein (CP) domains. These CPs have to undergo a post-translational modification by phosphopantetheinyl transferases (PPTase) catalyzing the *in trans* transfer of a 4'-phosphopantetheine (PPant) moiety from coenzyme A (CoA) to a highly conserved serine residue in the CP (**Figure 6A**)<sup>120</sup>. This tethering activates the catalytically inactive *apo*-CP into the active *holo*-CP. Afterwards the *holo*-CP can serve as a “swinging arm”, to transport the thioester bound biosynthetic intermediates to the catalytic sites of the PKS or NRPS domains<sup>121,122</sup>. The thioester-based biochemical activation of acyl monomers and aminoacyl monomers provides the required catalytic reactivity for this nucleophilic substitution and also presents kinetically accessible electrophiles (thioester) for the Claisen condensation (polyketide) or amide bond formation (non-ribosomal peptide) (**Figure 6B, C**)<sup>60,120</sup>. The number and order of modules often match the number and sequence of incorporated building blocks in the biosynthetic backbone – the so-called collinearity rule<sup>123,124</sup>, which dictate that each module catalyzes one extension.

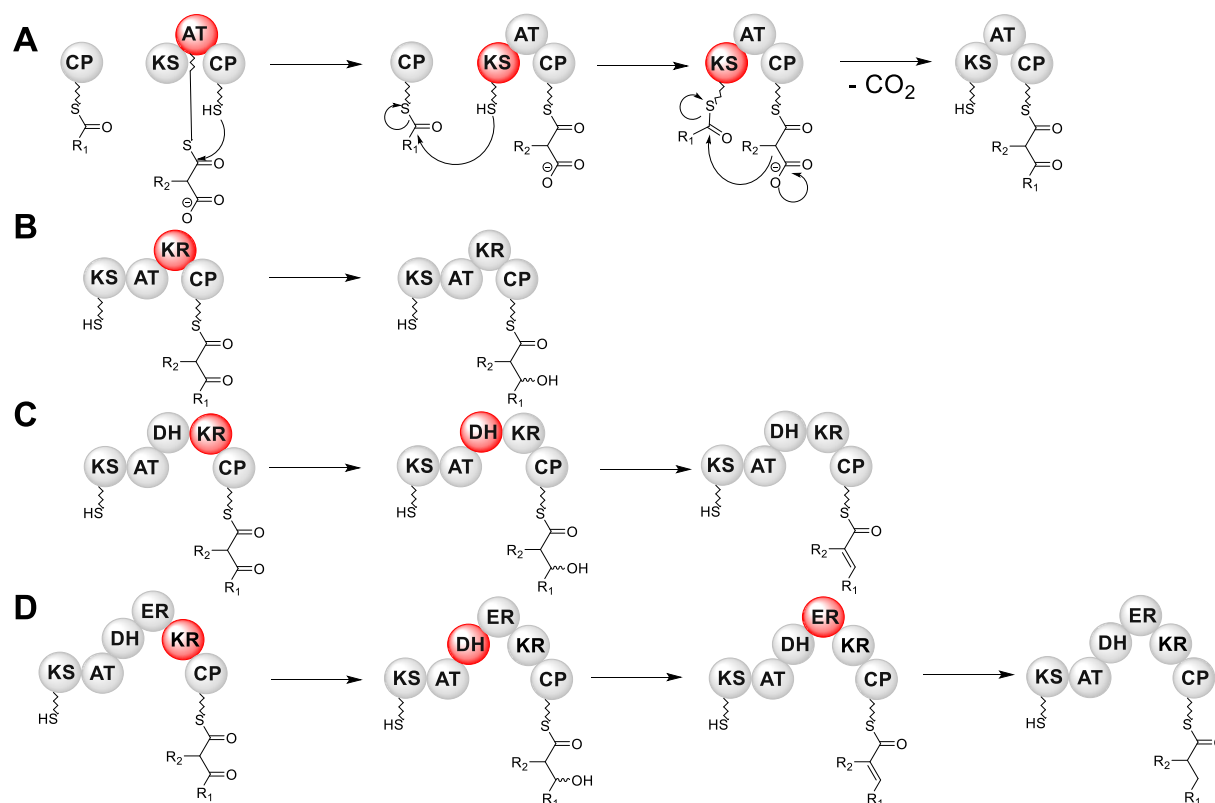
#### Modular type I PKS

Modular type I PKS systems are similar to fatty acid synthases from primary metabolism. Both catalyze the assembly of various acetate-derived building blocks via repetitive decarboxylative Claisen thioester condensations to form a long polyketide chain. The minimal PKS module consists of an acyltransferase (AT) domain, an acyl carrier protein (ACP) and a ketosynthase (KS) domain; wherein each domain is responsible for a particular step, resulting in the elongation of the chain by one extender unit. The first step of the chain elongation involves substrate recognition of an activated CoA extender unit, typically malonate or methylmalonate, by the AT domain. The building block is transferred to the ACP; afterwards the KS domain catalyzes a thio-Claisen condensation between the growing chain tethered to its active site thiol and the extender unit attached to the ACP (**Figure 7A**)<sup>112</sup>. Additional domains are also found between the AT domain and ACP, which catalyze different tailoring steps such as a ketoreductase (KR) domain to produce an alcohol (**Figure 7B**), dehydratase (DH) domain to produce a double bond (**Figure 7C**) and enoyl reductase (ER) domain that produces a saturated C2 unit (**Figure 7D**)<sup>125</sup>. The last module of the assembly line usually contains a thioesterase (TE) domain to catalyze the release of the matured backbone in linear, cyclic or branched cyclic forms<sup>126</sup>. Acetyl-CoA and

propionyl-CoA are often used as starter units, whereas malonyl-CoA serves as a typical extender unit. Alternative extender units such as methylmalonyl-CoA or methoxymalonyl-CoA can also be used in polyketide biosynthesis<sup>127</sup>. Substrate specificity for the different CoA esters as extender units is conferred by the AT domain<sup>128</sup>.



**Figure 6.** **A**) Transformation scheme of post-translational phosphopantetheinylation via a phosphopantetheinyl transferase (PPTase). The PPTase transfers the PPant moiety (pantetheine) from coenzyme A (CoA) to a conserved serine residue on the *apo*-CP (grey) to produce *holo*-CP (black). **B**) Simplified view of the decarboxylative Claisen condensation displaying polyketide chain elongation. **C**) Schematic view of non-ribosomal peptide chain elongation. Red spheres = PKS domains; Blue spheres = NRPS domains.; 3',5'-phosphoadenosine phosphate = 3',5'-PAP.



**Figure 7.** Assembly line of modular type I polyketide synthases (PKS) (A) Set of reactions in a typical polyketide synthase (PKS) module. A cycle of chain elongation in PKS module starts with the substrate recognition of a CoA activated extender unit through the acyltransferase (AT) domain. Afterwards the substrate is bound covalently as thioester onto a carrier protein (CP). Accessory domains such as a ketoreductase (KR) (B), dehydratase (DH) (C) and enoyl reductase (ER) (D)), catalyze redox adjustment onto the  $\beta$ -keto thioester during condensation with an acyl residue from the upstream module resulting. (B–D) Simplified schemes of different module organization in modular type I PKSs and the corresponding redox state of the C2 unit used as an extender. B highlights the minimal PKS module organization, whereas D constitutes the entire reductive loop of the PKS assembly line. The catalytically active domain in each step is highlighted in red. Scheme adapted from Hug *et al.*<sup>60</sup>.

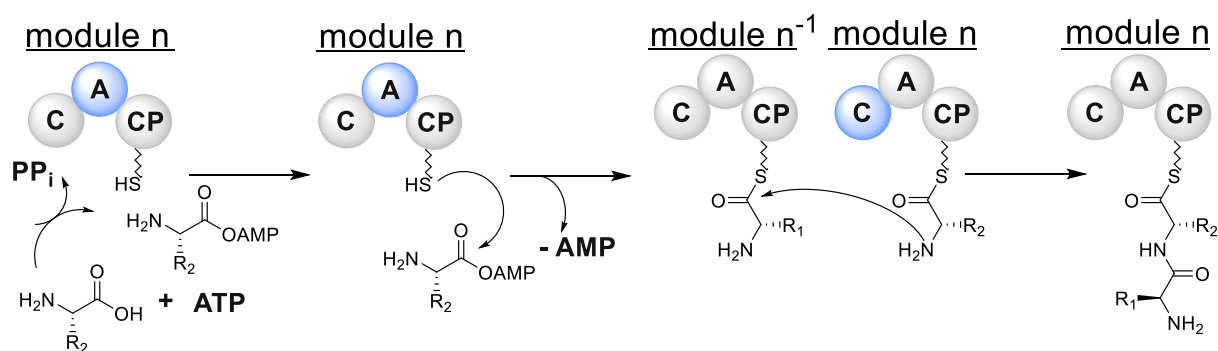
## Non-ribosomal peptide synthetases

The NRPS megaenzyme complex resembles the modular type I PKS; however, in case of the NRPSs the incorporated monomers comprise proteinogenic and non-proteinogenic amino acids<sup>129,130</sup>. The basic mechanism of PKS and NRPS chain elongation differs in the employed chemistry for activation and condensation of the starter and extender units. NRPSs produces structurally diverse peptides surpassing the anticipated structural scaffolds of non-modified proteinogenic amino acids<sup>131</sup>, by introducing unusual building blocks<sup>132</sup> and utilizing additional optional modifying domains<sup>133</sup> or separate proteins catalyzing reactions such as epimerization, hydroxylation, methylation, isomerization, and halogenation<sup>134,135</sup>. A loading or initiation module selects the first amino acid of the peptide chain, which comprises an adenylation (A) domain and a peptidyl carrier protein (PCP) domain but lacks a condensation (C) domain.

The set of catalyzed reactions in typical NRPS modules start with the recognition of an amino acid substrate by the upstream module, followed by the adenosine triphosphate (ATP)-dependent activation as an aminoacyl adenylate. The A domain catalyzes the covalent binding of the adenosine monophosphate (AMP)-activated amino acid as a thioester to a PCP. During the last step, the C domain catalyzes the condensation of the peptide bond via a nucleophilic attack of the loaded extender unit

(amino group) on the acyl thioester of the growing peptide chain to elongate the non-ribosomal peptide by one amino acid. The unloaded PCP domain can be now loaded with a successive substrate molecule, after the elongation of the peptide backbone<sup>136</sup> (**Figure 8**). This catalytic cycle leads to the conclusion, that the minimal module to perform substrate selection, binding and elongation of the growing peptide chain by peptide bond formation requires an A, a CP and a C domain<sup>137</sup>. The terminal module located downstream of the last elongation module usually contains a TE domain to release the full-length chain from the enzyme<sup>138,139</sup>. The TE domain can either simply hydrolyze the tethered peptide backbone to yield a linear peptide or perform a cyclization to release a macrolactam/macrolactone, depending on the reaction catalyzed by the TE domain<sup>140,141</sup>. Since the specific amino acid is selected by the A domain, the substrate specificity of these A domains can be predicted by the amino acid sequence of their substrate binding pocket, referred to as “Stachelhaus code”<sup>142,143</sup>. *In silico* prediction tools based on the “Stachelhaus code” proved their utility to assign myxobacterial A domain specificity in numerous studies<sup>144–146</sup>.

The structural diversity of the peptide backbone arises from various additional domains that may be present in each module such as *S*-adenosyl-L-methionine (SAM)-dependent *N*-, *O*- or *C*-methyltransferase (*N*-MT, *O*-MT, *C*-MT) domains. MT domains are prototypical accessory domains not only present in NRPSs, but also in PKSs or PKS-NRPS hybrid machineries. These domains catalyze the transfer of methyl groups from SAM to the carbon, nitrogen or oxygen atoms at various reactive positions on the backbones of polyketides and non-ribosomal peptides before the condensation of peptide bond formation<sup>133,147–149</sup>. These MT domains feature a bi-domain structure, where the first subdomain contains the binding site for methyl group donor and the second subdomain harbors the binding site for acceptor substrate<sup>147–150</sup>. Epimerase (E) domains convert L-amino acid into the respective D-configuration<sup>151</sup>, and a cyclization domain (Cy) replaces the conventional C domain to incorporate cysteine, serine or threonine residues to form a heterocycle<sup>152</sup>. These heterocycles can be further modified through oxidation domains to convert thiazoline and oxazoline residues into the corresponding thiazoles or oxazoles. In contrast, reduction domains can perform nicotinamide adenine dinucleotide reduced (NADH)-dependent formation of thiazolidine and oxazolidine. Soluble or free standing modification enzymes can further mature the non-ribosomal peptide intermediate via cytochrome P450 (CYP450) enzymes or glycosyltransferases<sup>153,154</sup>.



**Figure 8.** General reaction scheme and architecture of a NRPS module. The adenylation domain (A) activate the respective amino acid, using ATP. The AMP-activated aminoacyl is accepted by the peptidyl carrier protein (CP). Finally, the condensation (C) domain catalyzes peptide bond formation through a nucleophilic attack by the  $\alpha$ -amino acid of the loaded extender unit ( $\text{NH}_2$  group) onto the growing peptide chain (activated thioester) to elongate the non-ribosomal peptide by one amino acid. Optional *N*-methylation is catalyzed by *N*-methyltransferases (*N*-MT) and epimerization by epimerase (E) domains (not shown). The catalytically active domain in each step is highlighted in blue. Scheme adapted from Hug *et al.*<sup>60</sup>.

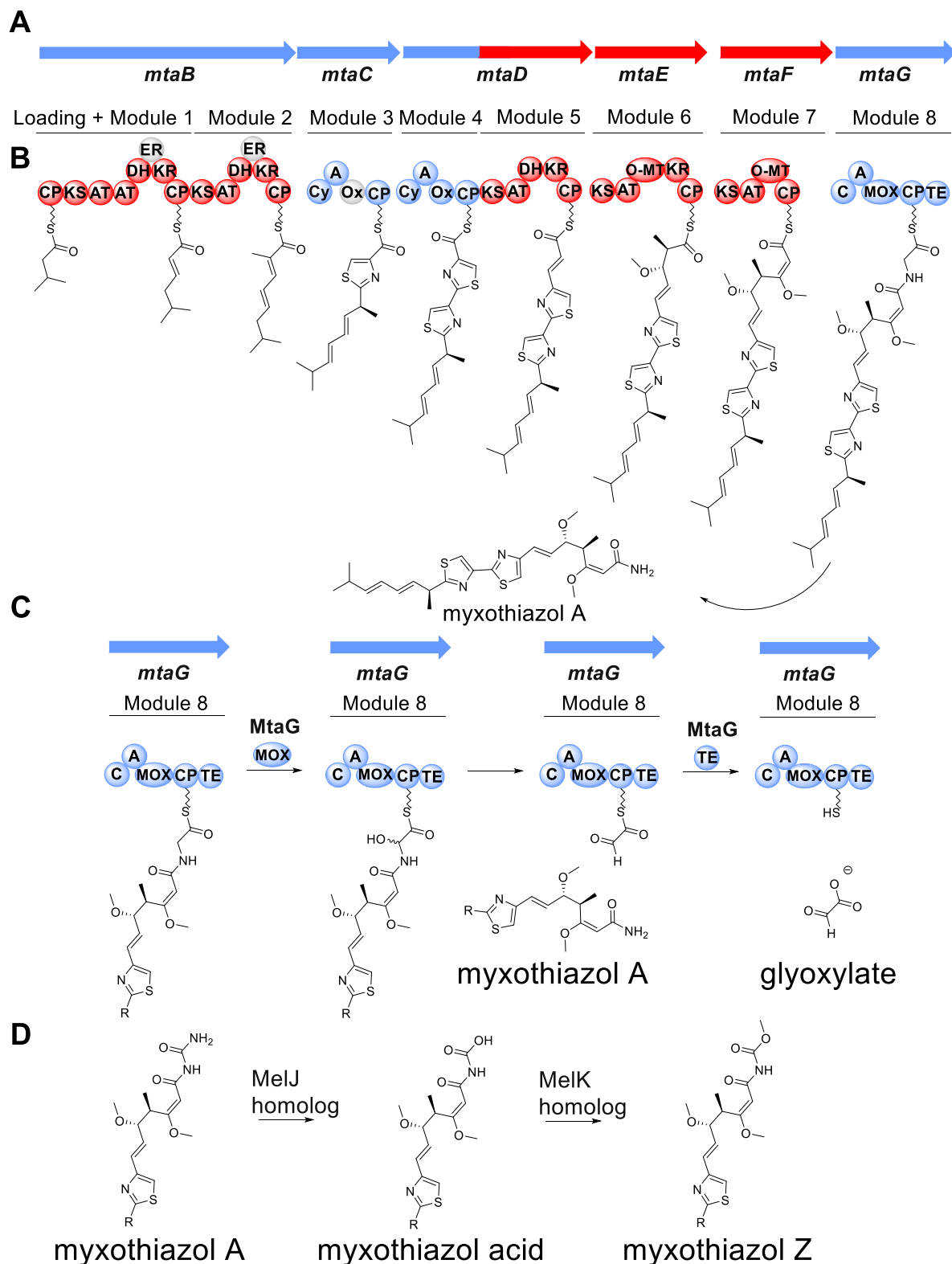
### PKS-NRPS hybrid systems

Since the modular NRPS and PKS machineries share similar biosynthetic mechanisms of building block selection, priming of a carrier protein and condensation with a downstream extender unit, it was in principle not unexpected to find hybrid PKS-NRPS assembly lines<sup>155</sup>. These hybrid assembly lines further diversify microbial natural products and the occupied chemical space by combining enzyme functions and features of PKS and NRPS assembly lines. Early biochemical characterization of PKS-NRPS hybrid systems<sup>156</sup> producing natural products such as bleomycin<sup>157</sup>, pristinamycin<sup>158</sup> but also the myxobacterial natural products epothilones and myxothiazols triggered the investigation of many more myxobacterial PKS-NRPS hybrid machineries. Similar to the well-known myxobacterial epothilones, the myxothiazols are biosynthesized via a PKS-NRPS hybrid machinery and features beside the potent biological activity several remarkable structural features (**Figure 9**).

The unusual starter unit isovaleryl-CoA, a bis-thiazole moiety, the  $\beta$ -methoxyacrylate pharmacophore and different terminal functional groups such as an amide in the structure of myxothiazol A or a methylester in myxothiazol Z are only some of the remarkable characteristics during the biosynthesis of the PKS-NRPS hybrid machinery. The biosynthesis is initiated by loading an isovaleryl-CoA starter unit catalyzed by an exceptional domain organization in module 1 harboring two AT domains, which is further condensed with malonate on module 1, methylmalonate on module 2 and cysteine unit on module 3. The downstream domain organization of the myxothiazol biosynthetic machinery is consistent with the enzymatic reactions required to build the PKS-NRPS hybrid molecule. Nevertheless several unusual deviations from the predicted biosynthetic logic were identified such as nonfunctional ER domains between each DH and KR domain in MtaB (referenced in the literature as spacer<sup>159</sup>), or the unusual terminal amide in myxothiazol A. The release mechanism for the biosynthetic backbone is significantly different from other NRPS, PKS or PKS-NRPS hybrid machineries. An additional chain extension step based on the incorporation of glycine on the terminal end catalyzes the release of the biosynthetic backbone and not a conventional TE-catalyzed release (**Figure 9C**). The last module MtaG in the myxothiazol assembly line harbors a monooxygenase (MOX) domain, which oxidizes the glycine extension on the intermediate followed by spontaneous or enzymatically catalyzed cleavage of the resulting unstable  $\alpha$ -hydroxylated intermediate. In conclusion, myxothiazol A is released from the PCP domain and the TE domain catalyzes the last reaction in this cascade, the release of the PCP-bound glyoxylate<sup>160,161</sup> (**Figure 9C**). The biosynthesis of myxothiazol exemplifies, that exceptions from the predicted “textbook” logic found in myxobacteria is a rather frequent finding.

Among the numerous myxobacterial PKS-NRPS hybrid assembly lines which have been characterized over the past years, structural and genetic investigations of the epothilone<sup>162</sup>, tubulysin<sup>163</sup> and myxothiazol<sup>164</sup> biosynthetic pathways provided important mechanistic insights concerning the dynamics of protein–protein interactions between PKS and NRPS subunits or modules. In principle, intersubunit recognition processes (the junction between two modules) in pure PKS systems are in general mediated by folded regions the so-called docking domains<sup>165</sup>, while the equivalent interaction between the NRPS interface are referred to as communication-mediating (COM) domains<sup>166</sup>. Both, the docking and the COM domains are located at the *N*- and *C*-termini of the respective protein subunits<sup>165,166</sup>, and generate non-covalent associations with the respective CP<sup>167</sup>. Despite the fact, that PKS-NRPS machineries feature the same domain organization and reaction mechanism of pure PKS and NRPS systems, the docking domains of PKS-NRPS hybrids are distinct<sup>167</sup>.





**Figure 9.** The myxothiazol gene cluster (**A**) and the encoded biosynthetic machinery (**B**) from *Stigmatella aurantiaca* DW4/3-1 represents a myxobacterial multimodular PKS-NRPS hybrid system. (**C**) Model for terminal amide formation in myxothiazol A biosynthesis. (**D**) Proposed methylester formation: The MeIJ homolog (hydrolase) hydrolyzes myxothiazol A to the intermediate myxothiazol acid, which is then methylated by a MeIK homolog (SAM-dependent methyltransferase) yielding myxothiazol Z; MOX: monooxygenase. Genes are shown as arrows, and modules of the biosynthetic machinery are indicated as shown, each containing a set of catalytic domains illustrated as circles (blue color: PKS genes/modules, red color: NRPS genes/modules) MT: methyltransferase domain; Cy: cyclization domain. Scheme adapted from Hug *et al.*<sup>60</sup>.

Hence, PKS-NRPS hybrid machineries feature not only different intersubunit modular interfaces between NRPS-PKS and PKS-NRPS, but also alternative docking domains between NRPS-NRPS and PKS-PKS interfaces. The biosynthetic domain organization encoded by the epothilone genes, exemplifies the nature of PKS-NRPS intersubunit interfaces of hybrid machineries<sup>168</sup>. The epothilone gene cluster encodes six genes (*epoA-F*), of which five proteins (EpoA, EpoC, EpoD, EpoE and EpoF) contain nine PKS modules and one protein (EpoB) serves as NRPS module<sup>169,170</sup> (**Figure 10A, B**). The PKS module EpoA and the NRPS module EpoB features a PKS-NRPS interface (**Figure 10C, i**), whereas the intersubunit NRPS-PKS interface is located between EpoB and the PKS module EpoC (**Figure 10C, ii**). At the PKS-NRPS interface, a Cy domain of the NRPS module EpoB recognizes the acetyl group bound to the CP domain of the PKS module EpoA for condensation with L-cysteine bound to the CP domain of the NRPS module EpoB. The CP domain of the PKS module EpoA contains a 38 amino acid long C-terminal docking domain (EpoA-Cdd) which interacts with the 55 amino acid long N-terminal docking domain (EpoB-Ndd) on the Cy domain of the NRPS module EpoB.

Both docking domains are not only essential for the formation of the 2-methylthiazole product, but also enables the replacement of EpoA and EpoB with a non-cognate protein, as long as the interface is not disrupted. Consequently, the replacement of EpoA through a CP domain of RapC from the rapamycin biosynthetic machinery was initially not recognized by the downstream module EpoB. However, a C-terminally added EpoA-Cdd onto the CP domain of RapC facilitated the production of 2-methylthiazole, even though at a lower rate than the native system<sup>171,172</sup>.

The crystal structure of the EpoB\_Cy domain revealed a monomeric crystal, in which EpoB-Ndd is exclusively connected to the downstream Cy domain through a flexible 20 amino acid residue linker. This finding suggests that any docking domain could possibly substituted for EpoB-Ndd without further modification of the protein surface of the Cy domain<sup>162</sup>. Homologs of the EpoB-Ndd docking domains are also found in other PKS-NRPS intersubunit interfaces such as in the bleomycin biosynthesis machinery (42% sequence identity)<sup>157</sup> or in the N-terminal docking domain of the NRPS module MtaG (45% sequence identity)<sup>160</sup>.

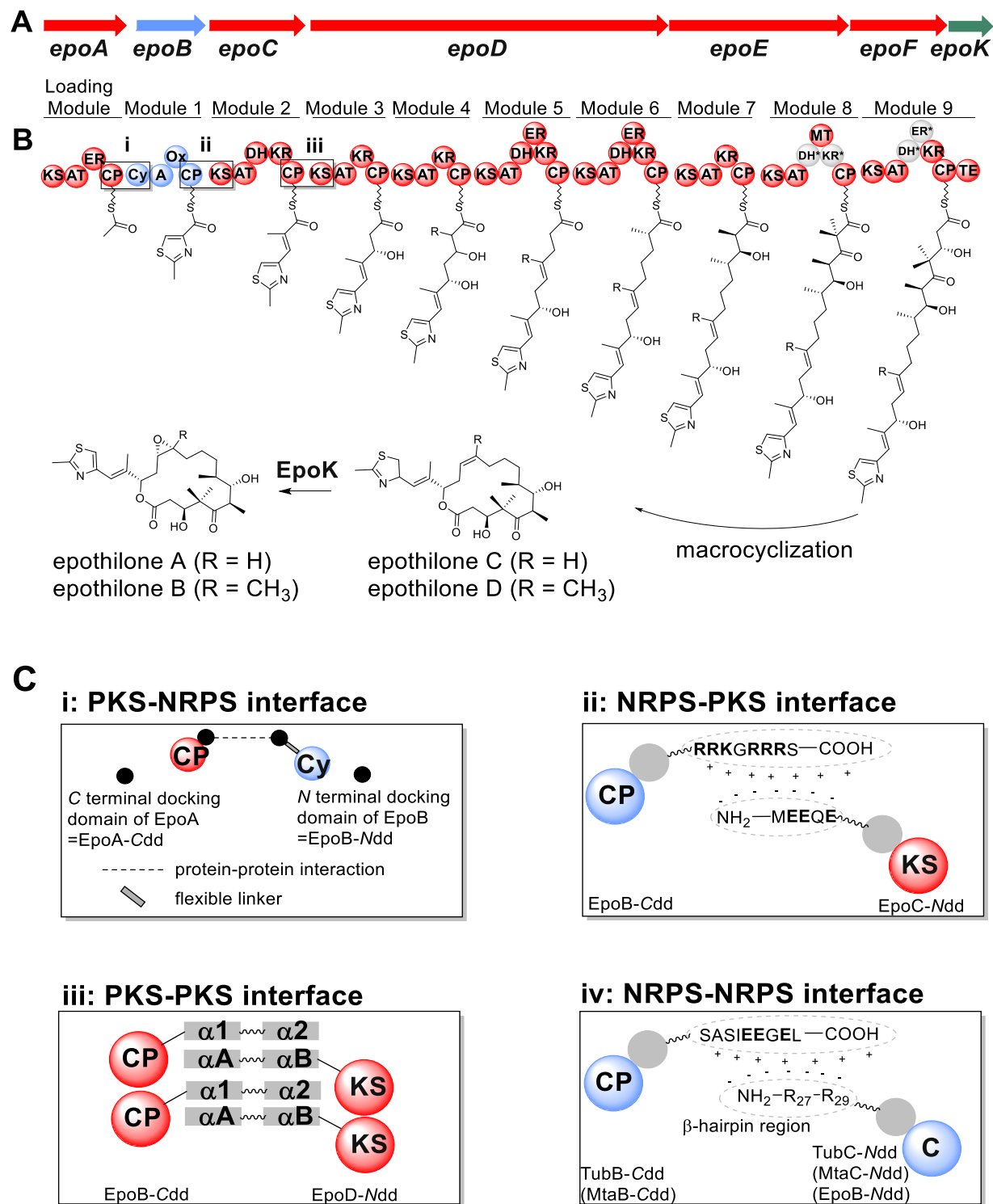
At the NRPS-PKS interface (**Figure 10C, ii**), the first KS domain of the PKS module EpoC accepts the 2-methylthiazole group from the CP domain of the upstream NRPS module EpoB for condensation with the methylmalonyl extender unit bound to the first CP domain of the downstream PKS module EpoC<sup>173</sup>. The characteristics of the C-terminal EpoB and N-terminal EpoC docking domains have been characterized through biochemical analysis using the standalone CP CouN5 from coumermycin biosynthesis<sup>172</sup>. The standalone CP CouN5 without modifications was not recognized by EpoC, whereas a CouN5 CP hybrid that contained the last eight residues of EpoB was recognized by EpoC. In conclusion, six positively charged residues in the C-terminus of EpoB works as a recognition element for interaction with three negatively charged Glu residues in the five N-terminal residue of EpoC (**Figure 10C, ii**).

Hence, electrostatic interactions play an important role in the assembly of the NRPS-PKS interface of EpoB-EpoC, which was also shown for other biosynthetic pathway such as bleomycin<sup>157</sup> and yersinicabactin<sup>174</sup>. Since structural information concerning the protein-protein interaction of the NRPS-PKS interface are still missing, the detailed mechanistic insights of docking domain-based recognition remains elusive<sup>167</sup>.



The docking domains in the PKS–PKS intersubunit interfaces such as in the PKS–NRPS hybrid biosynthetic machinery of epothilone, can be subdivided into class 1 and 2 docking domains. Class 1 docking domains are generally found in pure modular PKS systems<sup>165,175</sup> and in the biosynthetic machinery of rapamycin<sup>167</sup>, whereas class 2 docking domains are found in PKS–PKS intersubunit interfaces of cyanobacterial and myxobacterial PKS–NRPS hybrids<sup>176</sup>, such as in the interfaces of EpoC–EpoD, EpoD–EpoE and EpoE–EpoF in the epothilone machinery (**Figure 10C, iii**). The class 1 Cdd has two additional helices to form a four-helix bundle homodimer, whereas the class 2 Cdd is 40 amino acids shorter than the class 1 Cdd. In addition the class 2 Cdd lack the region responsible for homodimerization and is not contributing to PKS module dimerization. Nevertheless the class 2 Cdd features two docking helices, which interact and enable similar binding affinity ( $K_d$  is 2–20  $\mu$ M) to the respective class 2 Ndd<sup>167</sup>.

Some NRPS–NRPS interfaces of PKS–NRPS hybrids feature similarities to the EpoB–Ndd docking domain. For example, the tubulysin biosynthetic machinery features a C-terminal 25 amino acid docking domain TubB–Cdd on the NRPS module TubB which is recognized by the downstream docking domain TubC–Ndd (57 amino acids, 35% identity with EpoB–Ndd). In the myxobacterial interface between the NRPS module MtaC (MtaC–Cdd) and the downstream NRPS module MtaD (MtaD–Ndd) in the myxothiazol biosynthesis, MtaD–Ndd also features significant sequence identity to EpoB–Ndd (27%) and TubC–Ndd (40%). Despite the sequence similarity between EpoB–Ndd and TubC–Ndd (35%) and similar  $\alpha\beta\beta\alpha$  topology in their solution structure, the EpoB–Ndd\_C domain exist in a monomeric state in the crystal structure<sup>162</sup>, whereas the homologous N-terminal TubC NRPS module (TubC–Ndd) is a homodimer<sup>163</sup>. The  $\beta$ -hairpin region – which contains numerous charged amino acid residues in particular Arg27 and Arg29 – contributes to the dimer interface of the TubC NRPS module and is a conserved feature among TubC–Ndd and EpoB–Ndd type docking domains. Genetic mutation experiments and binding affinity studies revealed electrostatic interaction between Arg27 and Arg29 of the  $\beta$ -hairpin region TubC–Ndd and three glutamic acid residues in the terminal five residues of TubC–Cdd<sup>163</sup> (**Figure 10C, iv**). Genetic mutation studies altering the Glu residues in MtaC–Cdd<sup>164</sup> and the fact that this residue is also conserved in EpoB–Cdd-type docking domains supports its the significance in the assembly of functional docking domain complexes<sup>167</sup>. In general, the oligomerization states of PKS, NRPS and PKS–NRPS hybrid machineries are an important aspect concerning their quaternary structure. While type I modular PKSs are working as homodimeric enzyme complex, most of the NRPSs seem to work as monomers. The only so far characterized dimeric NRPS system is the vibriobactin NRPS VibF, where the catalytically inactive C1 domain enables dimerization<sup>177,178</sup>. In modular type I PKS systems, the homodimer structure is stabilized by KS and DH domains<sup>175</sup>, a dimerization element and through the dimerization  $\alpha$ -helices of the docking domain<sup>179,180</sup>. The apparent discrepancy of the oligomerization states of modular PKS and NRPS systems – which would severely impede functional compatibility of PKS–NRPS hybrid assembly lines – might be solved through an equilibrium state between monomeric and dimeric quaternary structure. Monomeric as well as dimeric structures have been found and it is assumed that NRPS subunits are able to switch between monomeric as well as dimeric structures depending on whether they interact with other NRPS subunits (monomeric) or PKS subunits (dimeric) in hybrid megasynthetases<sup>177</sup>. Further structural biology investigations might significantly contribute to rational engineering strategies of PKS–NRPS hybrid pathways to produce novel bioactive natural products.



**Figure 10.** The epothilone gene cluster (A) and the encoded biosynthetic machinery (B) from *S. cellulorum* So ce90 representing a multimodular polyketide synthase PKS-NRPS hybrid system. C) Examples of interfaces in PKS-NRPS hybrid machineries. Scheme adapted from Obwald *et al.*<sup>181</sup>.

### 1.3.3 Iterative Polyketide Biosynthesis

#### Type II and III PKS

The aforementioned type I PKS system is a large multimodular protein consisting of domains with defined catalytic functions, whereas the type II PKS machinery can be classified as tripartite aggregate of mono-functional proteins. The type II PKS system consists of two ketosynthase units KS $\alpha$  and KS $\beta$  and an ACP. KS $\alpha$  is the catalytically active site, while KS $\beta$  is controlling the chain length referred to as chain length factor (CLF), and the ACP carries the growing polyketide backbone (**Figure 11A**).

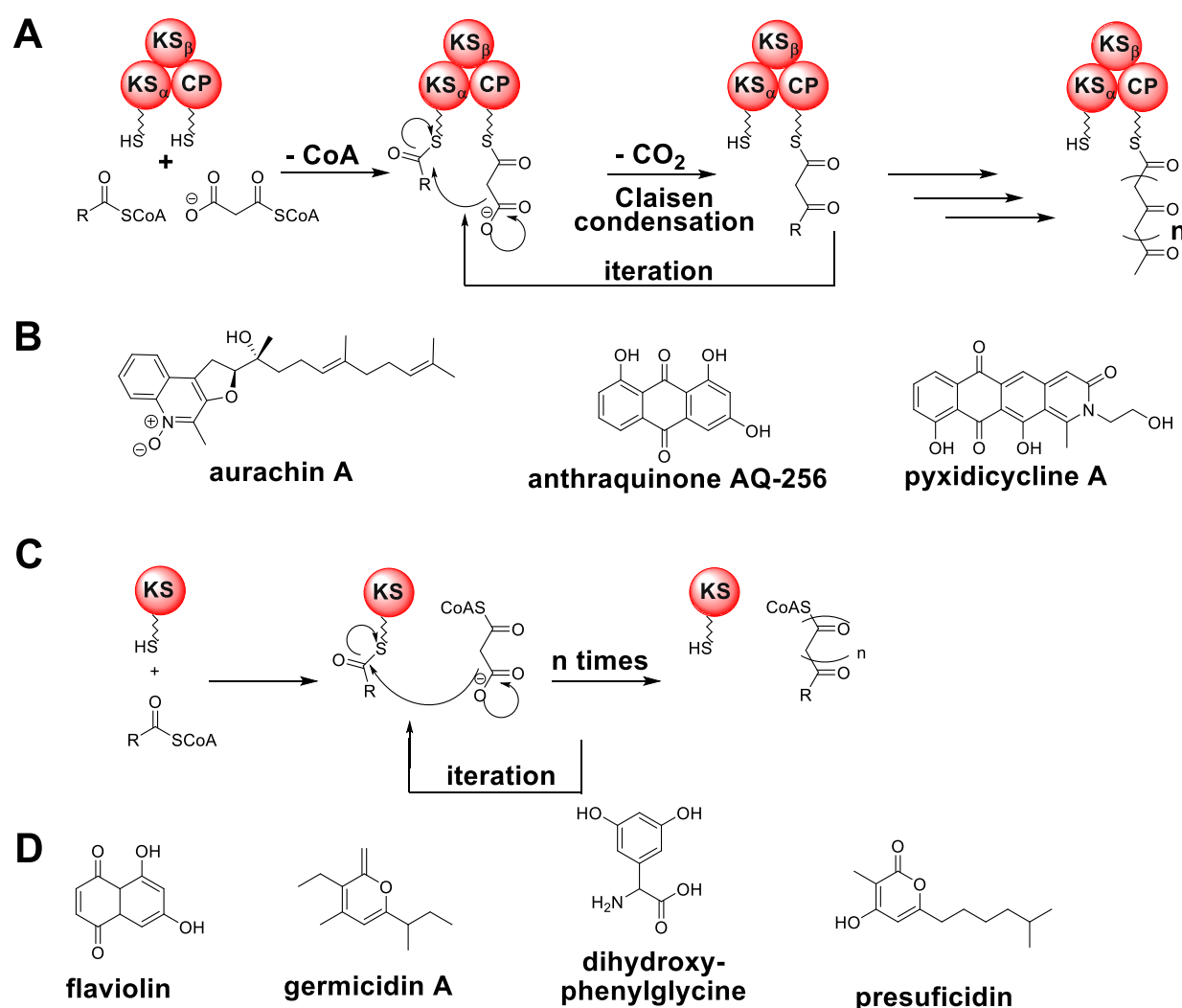
In conclusion, the catalytically active type II PKS machinery consist of three subunits forming a tightly interacting protein complex. Carboxylic acid, usually an acetate, is loaded onto an ACP that is subsequently transferred to the active site of the KS and undergoes iterative elongation using malonyl-CoA as extender units to form a nascent poly- $\beta$ -keto chain<sup>182–184</sup>. The way in which the nascent poly- $\beta$ -keto intermediate is folded is determined by the PKS subunit via a ketoreductase and other tailoring enzymes such as cyclases, aromatases along with oxygenases and in some cases transferases<sup>185–187</sup>. Additional modifications of the polyketide core skeleton are created by supplementary redox, group transfer, hydrolysis and rearrangement reactions<sup>188</sup>.

Type II PKS systems are exclusively present in bacterial microorganisms, in particular in Gram-positive bacteria belonging to the order actinomycetes such as the well-known streptomycetes producing tetracycline, and other actinobacteria such as *Amycolatopsis sulphurea*, the producer of chelocardin<sup>189,190</sup>. Only a few type II PKS products have been isolated from Gram-negative bacteria, such as the anthraquinones from *Photorehabdus luminescens*<sup>191</sup>, aurachin from *S. aurantiaca*<sup>192</sup> and recently the pyxidicyclines from *Pyxidicoccus fallax* An d48<sup>193</sup> (**Figure 11B**).

Type III PKS systems consist of a small homodimeric 40–47 kDa enzyme that utilizes a single active site to iteratively condense dicarboxylic extender units onto activated acyl-CoA<sup>194</sup>. Initially, type III PKSs were considered to be limited to plants, but in the last two decades more homologs from bacteria and fungi have been reported<sup>195</sup>. Unlike modular type I PKSs, type III PKSs generally accomplish an entire series of decarboxylative condensations and cyclization reactions in a single active site independent of ACPs<sup>196</sup>; however, ACP bound starter units as exemplified in NRPS biosynthesis are used by a few bacterial type III PKSs<sup>197</sup>.

The iterative mechanism of type III PKS systems works by transferring the growing polyketide chain in each catalytic cycle to its dimer to catalyze a decarboxylative thioester Claisen condensation where the growing polyketide chain remains bound to CoA instead of an ACP (**Figure 11C**). As soon as the polyketide backbone is extended to the preferred length (determined by the shape, volume, and chemical properties of the type III PKS active site cavity) the tautomeric keto-enol intermediate goes through one or more of three cyclization reactions, namely aldol condensation, Claisen condensation or lactonization<sup>198</sup>.

Type III PKSs synthesize natural products with relatively complex structures (**Figure 11D**) due to their flexible iteration, substrate tolerance and multiple cyclization reactions by solely utilizing a restricted substrate pool<sup>194,199</sup>. Type III PKS systems use acetyl, cinnamoyl, or fatty acyl as starter units and malonyl or methylmalonyl extender units. The type III polyketide scaffold is further decorated by modification enzymes, which can be classified as “upstream” reactions responsible for the production of type III PKS starter units or as tailoring enzymes performing “downstream” reactions of type III PKS products like acetylation, *O*-methylation, hydroxylation and oxidation<sup>198</sup>. These accessory genes are often clustered with type III PKS genes in microbial genomes<sup>198</sup>. Flaviolin<sup>200</sup> and its further oxidized product 1,8-dihydroxynaphthol<sup>201</sup>, the alkylquinones<sup>197</sup> and alkylpyrones (chapter 6)<sup>202</sup> are type III polyketides deriving from myxobacterial strains among other microorganisms.

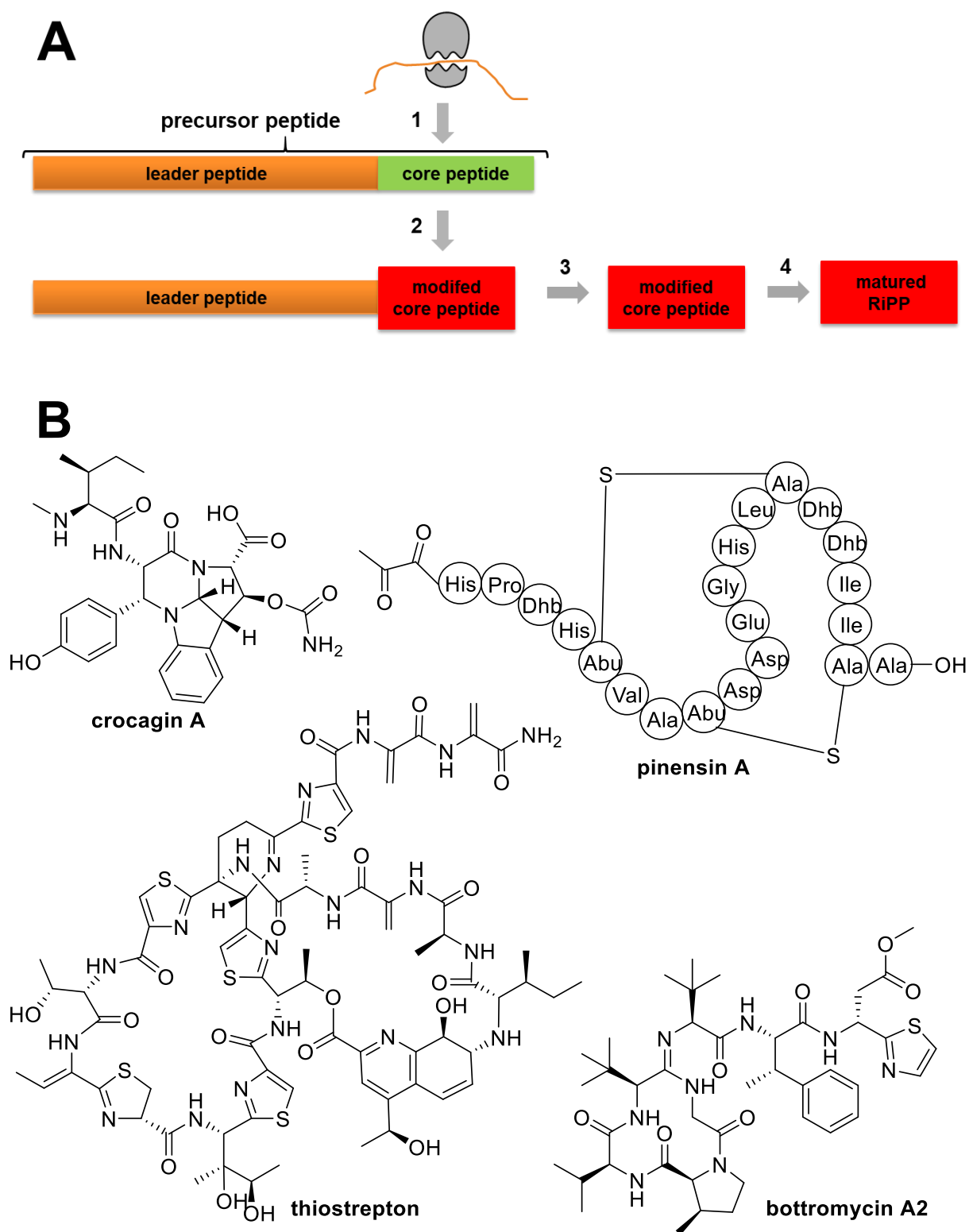


**Figure 11.** (A) Iterative polyketide extension in type II PKS systems and chemical structure of type II polyketides (B). The polyketide skeleton can be further modified by supplementary redox, group transfer, hydrolysis and rearrangement reactions (not shown). C) Iterative polyketide extension in type III PKS systems displaying the characteristic ACP-independent mechanism of the homodimeric single active enzymatic core that catalyzes loading, elongation and cleavage of the growing polyketide. (D) Chemical structure of natural products originating from type III PKSs. Scheme adapted from Hug *et al.*<sup>60</sup>.

### 1.3.4 Ribosomally synthesized and post-translationally modified peptides (RiPPs)

Biosynthesis of natural products does not exclusively require the presence of complex and specialized biosynthetic enzymes but can also rely on the very fundamental complex macromolecular machinery found in all living cells referred to as the ribosome. The production of ribosomally synthesized and post translationally modified peptides (RiPPs) begins with the generation of a precursor peptide (usually 20–110 amino acids), which undergoes post-translational enzymatic modifications<sup>203</sup>. This ribosomally produced precursor peptide consists of a leader peptide, usually located at the *N*-terminus, followed by the core peptide and an optional recognition sequence. The second step in the biosynthesis involves post-translational modification (PTM) enzymes, which recognize specific sequences within the leader peptide to modify the precursor peptide<sup>204</sup>. These PTM enzymes lead to the modification of the core peptide sequence on the precursor peptide. Afterwards the modified precursor peptide undergoes proteolysis to release the modified precursor peptide, which is finally further modified via secondary modifications to yield the mature peptide natural product, which can then be exported by transporters (**Figure 12A**). In some cases, removal of the leader peptide is the last step in the biosynthesis of RiPPs, which can be seen as a control mechanism to prevent natural product toxicity to the host itself<sup>205</sup>.

The PTM enzymes have a central role in the biosynthesis of RiPPs. These enzymes recognize in general secondary structures rather than specific sequences, providing the logical explanation for the given broad tolerance of leader peptide sequence<sup>204</sup>. In addition to these findings, recent studies supported the importance of core peptide recognition by these PTM enzymes, since the core peptide of their substrates may partly determine the ring topology and stereochemistry of the final products<sup>204</sup>. The implied general logic of PTM enzymes lies in the fact, that these enzymes perform their activities on the core peptide decoupled of the substrate recognition from the site of catalysis, what allows to some extent the replacement of amino acids within the core peptide. In contrast to linear peptides, extensive modification of the core peptide offers probably not only pharmacodynamics advantages to fulfill their biological function, but also resistance to proteolysis and lower entropic cost of target binding<sup>204</sup>. Despite the fact that the leader peptide plays a significant role in several steps of RiPP biosynthesis, many natural product intermediates have shown that specific posttranslational modifications are independent from the leader peptide. Examples for leader peptide independent modifications are C-terminal decarboxylation<sup>206</sup>, hydroxylation<sup>207</sup> and oxygenation<sup>208</sup> in lanthipeptides, as well as prenylation of cyanobactins<sup>209</sup>. Cyclization of the linear peptide is one of the common class defining PTM in RiPP pathways, which can be fundamental for bioactivity, since a cyclic shape of the matured natural product might be critical for receptor binding or protection from proteolytic enzymes<sup>210</sup>. These cyclizations include the formation of amide bonds, heterocyclization to form thiazolines or oxazolines<sup>211</sup>, oxidative carbon-carbon or aryl-aryl formation (chapter 3) and thioether cross-linking<sup>212</sup>. Some of these modifications are unique to RiPP biosynthesis, for example the formation of lanthionine<sup>210</sup>, YcaO protein-catalyzed heterocyclization<sup>213</sup> and radical SAM catalyzed thioether cross-links<sup>214</sup>. The tailoring reactions of PTM enzymes can form structurally complex RiPPs with non-proteinogenic amino acids in a peptide backbone resembling non-ribosomal peptides which in contrast require large multimodular enzyme complexes<sup>215</sup>. Because of their complex structures the thiopeptide antibiotic thiostrepton and the marine cytotoxic polytheonamide were initially thought to originate from



**Figure 12.** (A) Scheme of ribosomally synthesized and post-translationally modified peptide (RiPP) biosynthesis. The biosynthesis starts with ribosomal generation of the precursor peptide (1), which undergoes further posttranslational modification leading to a modified core peptide region (2). Subsequently, the modified precursor peptide undergoes proteolysis to release the modified precursor peptide (3). Finally, the modified precursor peptide can be further modified via secondary modifications to yield the mature peptide natural product (4). (B) Chemical/simplified structures of natural products originating from RiPP biosynthesis. Scheme adapted from Hug *et al.*<sup>60</sup>.

a non-ribosomal peptide biosynthesis<sup>216,217</sup>, however studies have confirmed that both are deriving from RiPP pathways<sup>218,219</sup>. Typical tailoring steps increasing the structural complexity of RiPP natural products involves hydroxylation reactions of thiopeptides like in thiostrepton<sup>220</sup>, epimerization, *i.e.* the catalytic conversion from L-configured amino acids to D-configured amino acid by epimerization of the  $\alpha$ -carbon<sup>221</sup>, methylation on heteroatoms in particular *O*- and *N*-methylation occurring after primary modifications. Remarkably, two peptides even contain an *N*-terminal di-methylated amino acid; arginine in plantazolicin<sup>222</sup> and alanine in cypemycin<sup>223</sup>. Moreover, *C*-methylation, halogenation, disulfide bond formation and prenylation for enhancing metabolic stability have been reported. Although primary modifications of those peptides are already sufficient for forming bioactive molecules, those tailoring steps offer enhanced stability or higher binding affinity to the biological target<sup>224</sup>. To date, there are over 30 distinct structural classes of RiPP natural products with the perspective that in the near future the number of classes likely increases further, as will be shown also in this thesis (chapter 3)<sup>225,226</sup>. In contrast to other microorganisms, myxobacterial RiPPs are under-investigated and the only published myxobacterial-exclusive RiPPs with characterized biosynthesis until recently were the polycyclic tripeptides named crocagins found in the myxobacterium *Chondromyces crocatus* Cm c5. The polycyclic tripeptides which contain an unique tetrahydropyrrolo [2,3-*b*] indoline core<sup>227</sup>, exemplify how myxobacterial natural products can enrich the diversity of RiPP encoded biosynthetic pathways. As shown in previous investigations of the myxobacterial natural product cittilin A<sup>228</sup> alongside with the chapters 3 and 7 in this thesis, myxobacteria might be a promising source for RiPPs.

## Scaffold peptide

A recent study highlighted an unprecedented strategy for the generation of amino acid-derived natural products such as reported for the biosynthesis of 3-thiaglutamate and cytotoxic pyrroloquinoline alkaloid ammosamide A<sup>229</sup>. The “scaffold peptide” biosynthesis starts with the ribosomal generation of a small peptide. In contrast to other characterized RiPP pathways, this small peptide serves exclusively as a scaffold for an additional non-ribosomal peptide extension and chemical modification and is not incorporated into the matured natural product. One specific amino acid, reflecting the chemical scaffold of the mature natural product, is transferred to the *C*-terminus of the peptide in an ATP- and tRNA-dependent reaction catalyzed by an aminoacyl-tRNA ligase (PEARL)<sup>230</sup>. Up to date, PEARL enzymes are known to use Cys-tRNA<sup>Cys</sup> leading to the biosynthesis of 3-thiaglutamate and to attach Trp-tRNA<sup>Trp</sup> in the ammosamide biosynthetic pathway. After addition of the amino acid cysteine as described in the biosynthesis of 3-thiaglutamate, modifications take place to furnish the 3-thiaglutamate by a series of enzymatic transformations that involve  $\beta$ -carbon excision, carboxymethylation, and proteolysis to yield the matured natural product derived from a single amino acid<sup>230</sup>. The scaffold peptide (considered as the biosynthetic machinery rather than as substrate) is regenerated after proteolytic release of the matured natural product and enters another round of biosynthesis providing a more efficient use of the ribosomally synthesized peptide than the stoichiometric use of leader peptides in typical RiPP pathways<sup>229</sup>. This finding provides an exciting variant of canonical RiPP biosynthesis, since the final product is not incorporated on a core peptide but added after ribosomal precursor peptide synthesis. The biosynthetic logic of scaffold peptide pathways resembles the NRPS biosynthesis, which uses the peptide carrier protein as a scaffold, and RiPP biosynthesis by using the same mechanism of binding of the scaffold peptide to an RRE domain<sup>229–231</sup>.

### 1.3.5 Terpenes

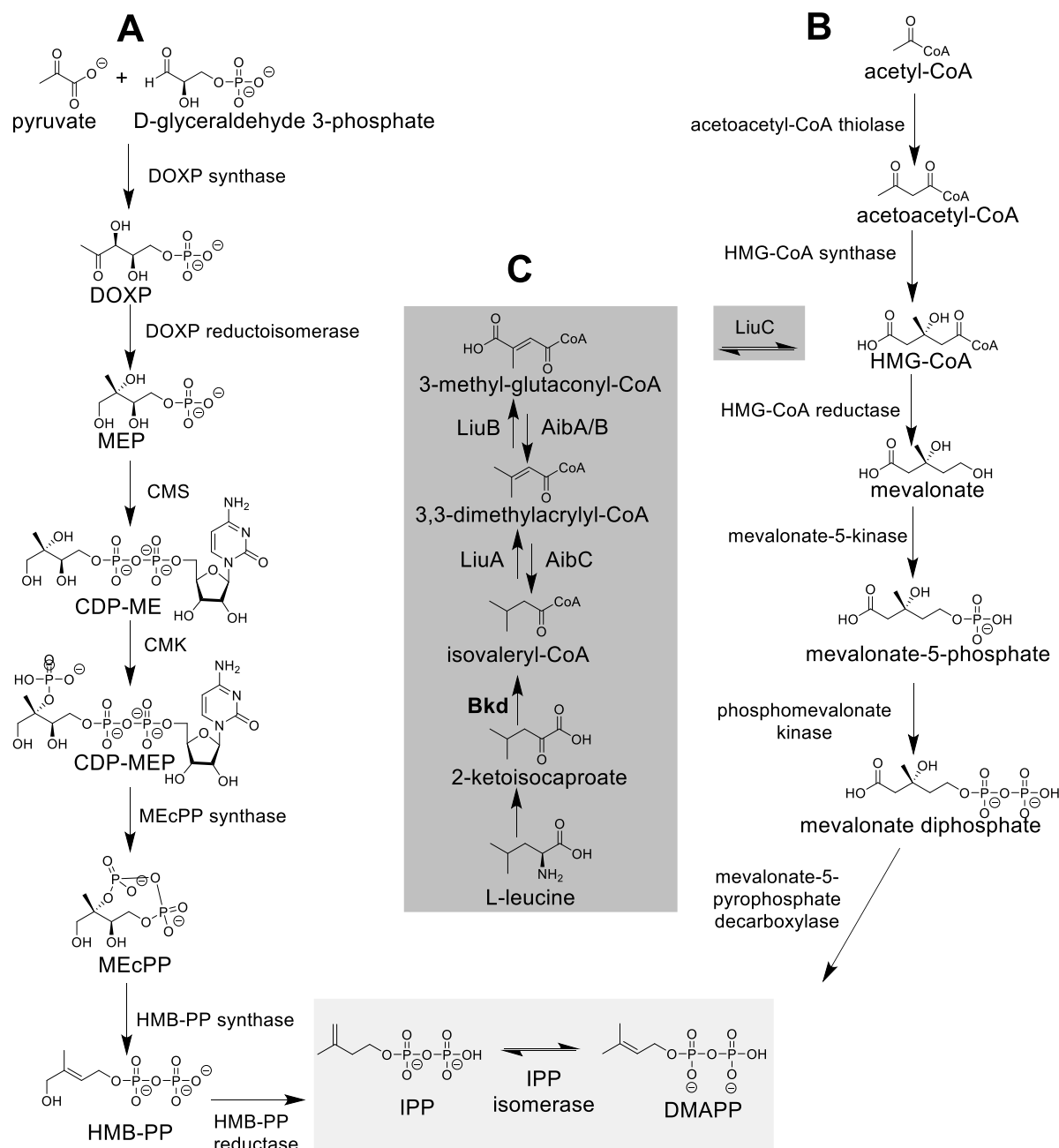
When all kingdoms of Life are taken into account as producers of natural products, terpenes and terpenoids (isoprenoids) are the largest class of natural products by spanning more than 70000 described compounds, which have been characterized and grouped into more than 400 structural families<sup>232</sup>. Terpenoids (also termed isoprenoids) contain oxygen in various functional groups and can be considered as modified terpenes, which in contrast are defined strictly as pure hydrocarbons<sup>233</sup>. To simplify further reading, both groups will be described in this thesis simply as terpenes, since both groups of natural products share isoprene as the common biosynthetic building block. Terpenes are synthesized through activated monomers of isoprene, namely isopentenyl (pyro)/diphosphate and dimethylallyl diphosphate, which are generated either by the mevalonate (or 3-hydroxy-3-methylglutaryl-CoA (HMG-CoA) reductase pathway<sup>234</sup> or via the 2-C-methyl-D-erythritol 4-phosphate/1-deoxy-D-xylulose-5-phosphate (MEP/DOXP)/non-mevalonate pathway<sup>235</sup>. The mevalonate pathway starts from acetyl-CoA whereas the non-mevalonate pathway starts from pyruvate and glyceraldehyde 3-phosphate to finally produce dimethylallyl pyrophosphate (DMAPP) and isopentenyl pyrophosphate (IPP) (**Figure 13A, B**). In myxobacteria DMAPP and IPP is provided by the mevalonate pathway for further biosynthesis of terpenes, but additional degradation of leucine and subsequent channeling of isovaleryl-CoA into HMG-CoA generation plays a central role in the biosynthesis of aurachin<sup>236</sup>.

In *Stigmatella aurantiaca* Sg a15, the mevalonate pathway is tightly connected to leucine catabolism, which was not exclusively caused to a complete degradation of leucine to acetyl-CoA. Genetic inactivation of the branched-chain  $\alpha$ -keto acid dehydrogenase (*bkd*) gene revealed an alternative pathway providing HMG-CoA starting from L-leucine (**Figure 13C**). This assumption was supported by feeding experiments with stable isotope-labeled precursors (isovaleryl ( $D_9$ ), and L-[ $D_{10}$ ]-leucine) in *S. aurantiaca* Sg a15 *bkd*-mutants – fermentation cultures showed predominantly a mass shift of + 5 and + 6 Da of the myxobacterial terpene aurachin<sup>236</sup>. Further studies (including chapter 4) showed also the presence of this biosynthetic route in other myxobacteria<sup>237</sup> and that isoprenoids are essential for fruiting body formation in *M. xanthus*<sup>238</sup>. The mevalonate pathway produces terpenes such as cholesterol, vitamin K, coenzyme  $Q_{10}$  and all steroid hormones<sup>239</sup>, while the non-mevalonate pathway, which is absent in eukaryotes and myxobacteria, presents a promising drug target against human pathogens like *Plasmodium falciparum*. The natural product fosmidomycin isolated from *Streptomyces lavendulae*<sup>240</sup>, specifically inhibits DOXP reductoisomerase, a key enzyme in the non-mevalonate pathway of isoprenoid biosynthesis<sup>241</sup>.

Whatever origin the monomers have, the biosynthesis of terpenes is initiated by oligoprenyl synthetases to assemble the hydrocarbon backbone using the monomers DMAPP and IPP to form geranyl pyrophosphate (GPP), farnesyl pyrophosphate (FPP), geranylgeranyl pyrophosphate (GGPP) or geranyl farnesyl pyrophosphate (GFPP) (**Figure 14A**). This linear polyene with branching methyl groups forms the core hydrocarbon structure in a single enzyme-catalyzed step. For that reason terpenes (including terpenoids) are classified according to the number of five-carbon units that form the hydrocarbon skeletons into hemiterpenes ( $C_5$ ), monoterpenes ( $C_{10}$ ), sesquiterpenes ( $C_{15}$ ), diterpenes ( $C_{20}$ ), sesterterpenes ( $C_{25}$ ), triterpenes ( $C_{30}$ ), tetraterpenes ( $C_{40}$ ) and polyterpenes ( $C_{5n}$ ). Among these groups the sesquiterpene and diterpenes are of particular interest since this group undergoes extensive oxidative modification and features molecular characteristics resembling those of known drugs<sup>232</sup>. After hydrocarbon backbone assembly the diversifying cyclization of the backbone is performed by catalytic



action of different terpene cyclases (TCs, also termed terpene synthases) to generate the basic terpene scaffold. Additional tailoring enzymes modify the terpene scaffold to yield decorated terpenes. The typical biosynthesis of terpenes can be subdivided into the first phase catalyzing the hydrocarbon backbone assembly and cyclization, and the second phase involving the terpene scaffold functionalization.



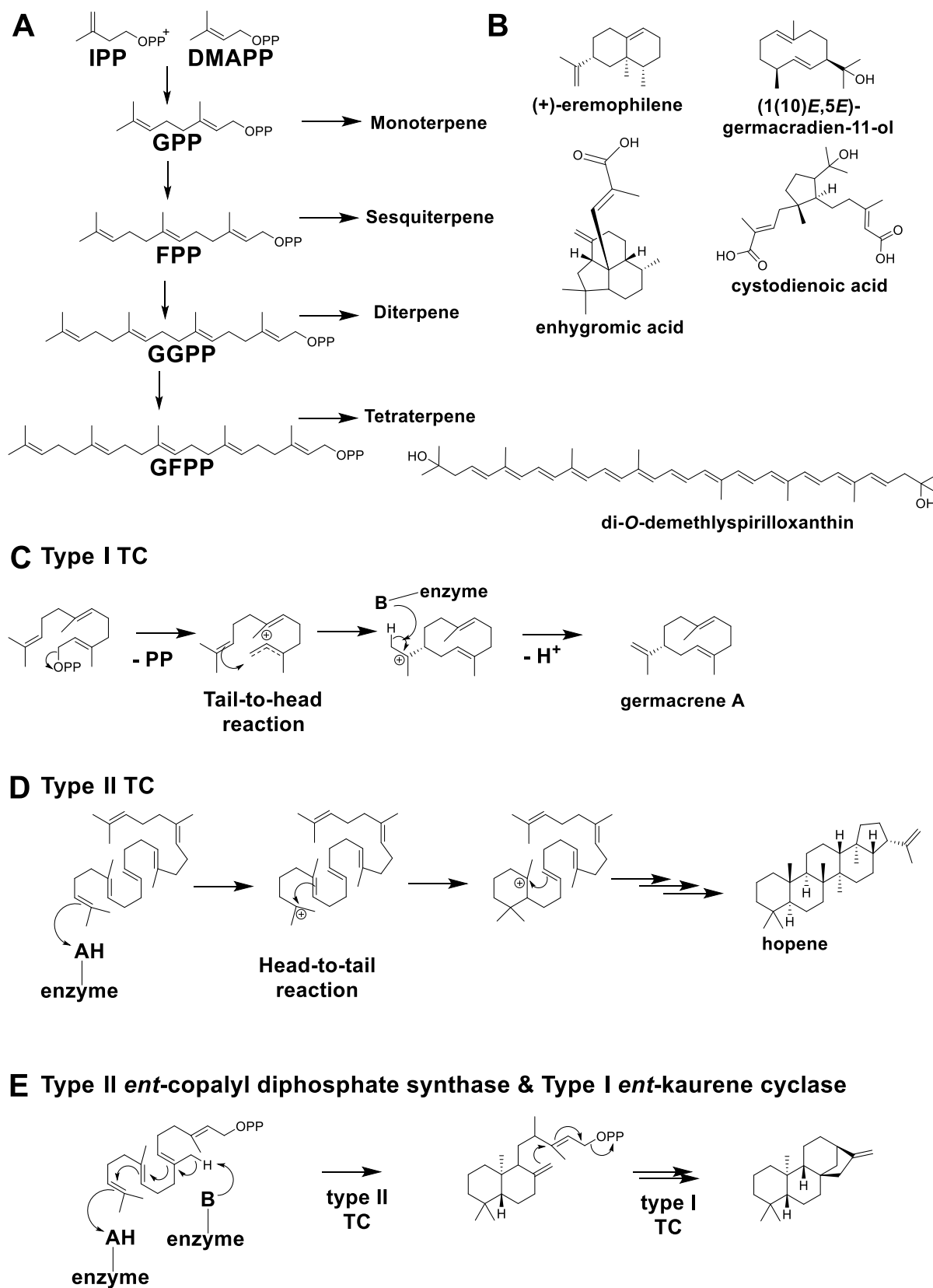
**Figure 13.** Biosynthesis of the isoprene monomers isopentenyl diphosphate (IPP) and dimethylallyl diphosphate (DMAPP) via the non-mevalonate pathway (**A**) or the mevalonate pathway (**B**). (**C**) In myxobacteria, leucine degradation actively channels generated isovaleryl-CoA into the biosynthesis of HMG-CoA to fuel the formation of IPP and DMAPP via the mevalonate pathway. Further information can be found in chapter 4. 1-deoxy-D-xylulose 5-phosphate (DOXP); 2-C-methyl-D-erythritol 4-phosphate (MEP); 4-diphosphocytidyl-2-C-methylerythritol (CDP-ME); 4-diphosphocytidyl-2-C-methyl-D-erythritol 2-phosphate (CDP-MEP); 2-C-methyl-D-erythritol-2,4-cyclopyrophosphate (MEcPP); (E)-4-Hydroxy-3-methylbut-2-enyl pyrophosphate (HMB-PP); 4-(cytidine 5'-diphospho)-2-C-methyl-D-erythritol kinase (CMK); 2-C-methyl-D-erythritol 4-phosphate cytidyltransferase (CMS).

### Terpene scaffold generation

The biosynthesis of the hydrocarbon backbone catalyzed by the oligoprenyl synthetases is a generic process, whereas the diversifying terpene cyclase is mostly responsible for the elaborate chemical scaffolds. In contrast to the aforementioned modular assembly lines with individual domains responsible for one specific biosynthetic step, TCs catalyze a cascade of cyclization reactions to guide the oligoprenyl diphosphate precursors in a chaperone-like mechanism<sup>242</sup>. TCs are categorized into two types, type I and II, featuring different mechanism for substrate recognition and protein folds. Type I TCs are metal-dependent enzymes catalyzing the formation of highly reactive allylic cations via heterolytic cleavage of the terminal diphosphate of FPP or GGPP<sup>243–245</sup>. The acyclic precursor is locked into a defined, pre-organized conformation that positions the leaving diphosphate group and the nucleophilic alkenes in proximity to initiate a C–C-bond formation and a carbocation-mediated cascade reaction<sup>242</sup>. In addition, intramolecular atom transfer, rearrangements including hydride or proton transfer and carbon shifts are assisted by type I TCs<sup>242</sup>. The highly reactive carbocation undergoes deprotonation via an E1-like mechanism or nucleophilic attack of water via a S<sub>N</sub>1-like reaction<sup>246</sup>.

Sesquiterpenes are exclusively formed by type I TCs, whereas diterpenes can either be generated by type I TCs or type II TCs<sup>243,247</sup>. Type II TCs use a conserved aspartate residue to protonate via Brønsted acid catalysis a terminal isoprene double bond or an epoxide ring to generate a carbocation<sup>248</sup>. In contrast to type I TCs this leads to an inverted direction of charge propagation along the polyisoprene chain in the subsequent cyclization cascade<sup>249</sup>. The type I TCs ring formation is defined as tail-to-head cyclization, since the allylic diphosphate ester bond (tail) is ionized followed by a cyclization of the terminal prenyl unit (head) (**Figure 14C**). In contrast, class II TCs results in a head-to-tail cyclization after electrophilic activation of a terminal prenyl unit (head) which is followed by a cyclization of a diene (**Figure 14D**). In conclusion, the cyclized product keeps the diphosphate functionality and can be used as substrate for a second cyclization catalyzed by type I TCs to generate more complex diterpene scaffolds<sup>247</sup>. Mixed type II/I cyclases and an increasing number of atypical terpene cyclization reactions have been recently described, many of which share the classical type I and II cyclization mechanisms initiated through formation of reactive temporary carbocation intermediates<sup>247</sup> (**Figure 14E**). Further non-canonical terpene cyclization can be catalyzed by flavin-dependent monooxygenases (FMO)<sup>250</sup> and CYP450 enzymes<sup>251</sup>. Oxidative<sup>252</sup> and reductive cyclizations<sup>253</sup> have been described, while the reductive cyclization unlike oxidative cyclization is catalyzed by two distinct enzymes.

These examples show that enzymes considered as terpene scaffold functionalization actors can also perform cyclization of the terpene scaffold. Among the tailoring enzymes important to further increase the structural complexity of terpenes, oxidoreductases (in particular CYP450 enzymes) are the most abundant tailoring enzymes. In the broadest sense, hybrid natural products featuring a terpene chemical scaffold hybridized with a chemical scaffold originating from another biosynthetic pathway can also be considered as additional decoration<sup>254</sup>. The terpene hybrid 2-hydroxysorangadenosine described in chapter 4 is an example of such a functionalized terpene.



**Figure 14.** Terpene biosynthesis. **(A)** Oligoprenyl synthetases use DMAPP and IPP to form linear polyenes with different number of five-carbon units to form the hydrocarbon skeletons. **(B)** Chemical structure of myxobacterial sesquiterpenes, diterpenes and a tetraterpene. Mechanism for type I **(C)**, type II **(D)** and type II/I tandem terpene cyclases **(E)**. Scheme adapted from Helfrich *et al.*<sup>232</sup>.

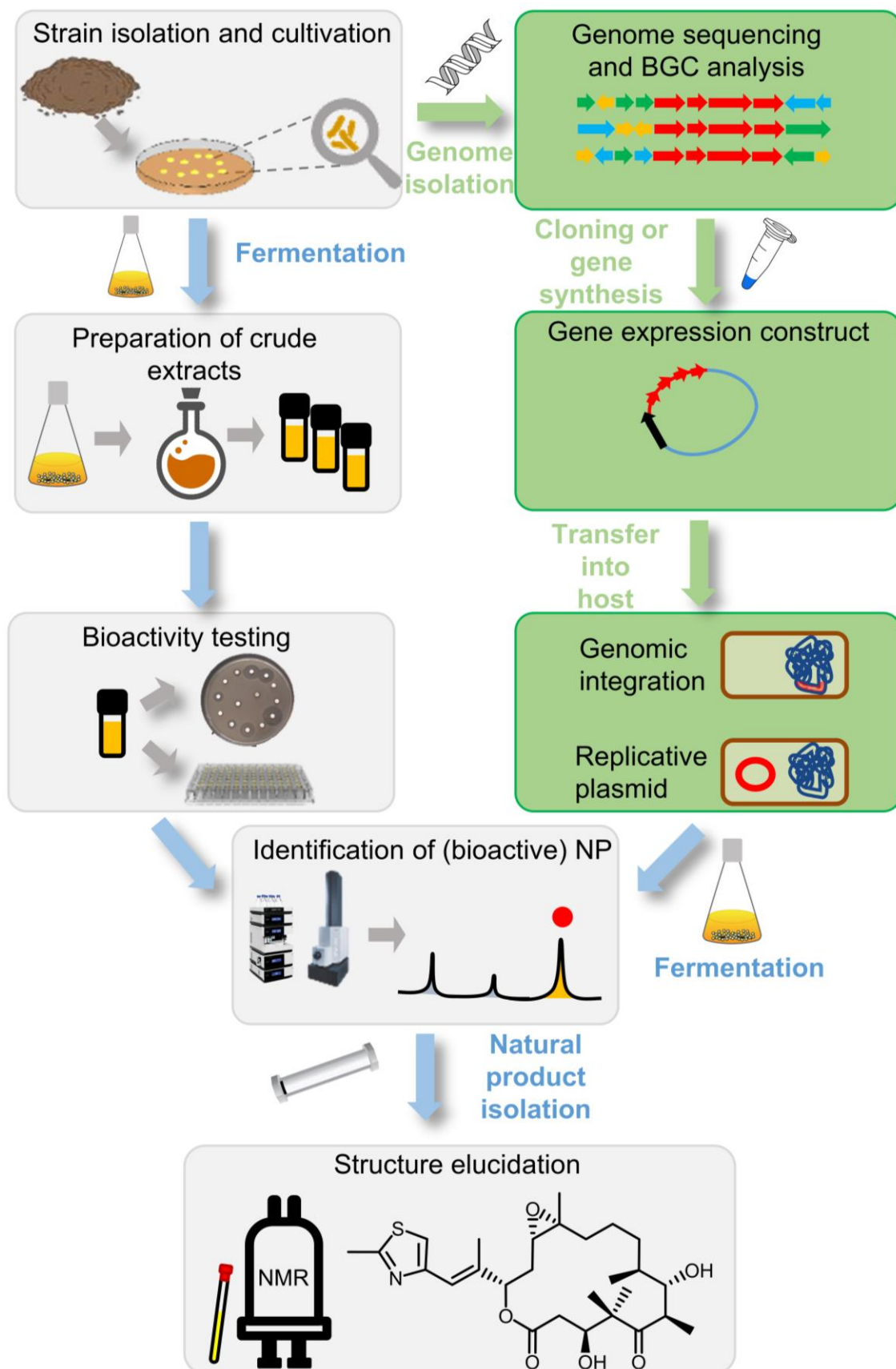
### 1.4 Natural product research – An interdisciplinary playground

Microbial natural product research significantly depends on the interplay of different disciplines of natural sciences. The traditional approach starts with microbiology of microorganisms through the isolation of new microbial secondary metabolite producers from the environment (**Figure 15**). The advent of affordable whole-genome sequencing transformed the process of discovering novel bacterial secondary metabolites immensely and (bio)informatics are nowadays inseparably connected to natural product research.

For example, genome-guided natural product discovery allows to access the biosynthetic potential of a microorganism directly from its genome sequence (by culture-dependent fermentation or metagenomics approaches). This in turns means that the BGC responsible for the production of a specific natural product can be identified *a priori* which paves the way for genetic manipulation procedures. The biosynthetic genes can then be transferred into a foreign host for the production of non-native molecules or the native strain can be manipulated to initiate or enhance the production of the corresponding natural product (**Figure 15**).

Molecular biology and biochemistry aims to characterize the biosynthetic machinery generating natural products. The overarching aim is not only to understand the underlying genetics, enzymology and regulation of the biosynthesis but also to enable pathway-modification within bacterial producers and the generation of engineered natural products of pharmaceutical or industrial interest. Accurate instrumentation and methods from the field of analytical chemistry are vitally important to identify and isolate natural products from a complex biological matrix. These analytical methods allow prioritization of minute amounts of natural products that escaped detection previously and opened the field for modern metabolomics studies. The complex structures of natural products involve organic chemistry to develop synthetic routes and reactions for total chemical synthesis or utilization of the complex chemical scaffolds of natural products to develop semisynthetic derivatives. The bioactive properties of natural products reveal often novel mode of actions and initiate the development of new biological assays. Large-scale production of a natural product ultimately requires further optimization through chemical engineering and biotechnological process optimization.

In light of the countless applied methods and concepts to access natural products – including those which are not described in this thesis – the following section aims to present a summary of selected technical and conceptual proceedings in the field of natural product research<sup>26</sup> relevant for the outline of this thesis. The focus is on the methods important to understand the biosynthetic logic of myxobacterial natural products. The first subsection 1.4.1 explains the analytical instrumentation to access natural products, while the second subsection 1.4.2 introduces the background of bacterial genetics, covering *in silico*-based biosynthetic pathway identification and investigation, and the tools available for the manipulation and engineering of biosynthetic pathways of myxobacteria.



**Figure 15.** Traditional natural product discovery on the left side and genome-guided natural product discovery workflow on the right (green boxes and arrows). Scheme adapted from Hug *et al.*<sup>118</sup>.

### 1.4.1 Analytical Instrumentation for Natural Product Discovery

As outlined above, one central part of natural product discovery is represented by the analytical setup for the identification, isolation and structure elucidation of the natural product. In that sense, the meaning of analytical chemistry-based methods remains to some extent unaffected from the major proceedings in the field of genomic-driven natural product discovery. This subsection does not intend to give a full overview of the analytical instrumentation used in the field of natural product research but provide a short overview of the fundamental instrumentation and methods applied in the context of this thesis. Several comprehensive reviews have summarized the instrumentation and methods applied to access bioactive natural products<sup>26,255</sup>.

In summary, the basic method to analyze and purify natural products relies on chromatographic separation of the culture extract, often coupled with liquid chromatography-mass spectrometry (LC-MS) monitoring. Methods for differential profiling using high-resolution mass spectrometric (hrMS) data coupled with statistical evaluation of profiles have been developed previously for myxobacteria and have been shown to be successful in detecting even subtle differences between numerous samples<sup>93</sup>. Following the detection of an unknown natural product, the structure elucidation is mainly conducted by advanced nuclear magnetic resonance (NMR) spectroscopy and the absolute configuration is determined with the help of different derivatization techniques<sup>256</sup>. Other conceptual innovations in analytical instrumentation for natural product discovery such as imaging mass spectrometry and supercritical fluid chromatography<sup>26</sup> are not addressed in this subsection for the matter of conciseness.

#### LC-MS

The classical “top-down” approach employed in the natural product discovery pipeline (**Figure 15**), mostly depended on sufficient production of the natural product of interest for reliable detection. Since the primary and secondary metabolome of bacteria are exceedingly complex, it is necessary to separate the metabolites in a bacterial extract from each other. For that reason, the advent of (ultra-)high performance liquid chromatography ((U)HPLC) and related separation techniques were a milestone for the reliable identification and isolation of natural products from the crude extracts. For a long time, natural product discovery was guided by LC-based separation and UV/Vis spectrometric detection of the eluting natural products. The combination of LC with hrMS detectors, capable of providing accurate mass-to-charge ratio ( $m/z$ ) information that supports the calculation of a proposed sum formula, revolutionized natural product discovery. From that time, significantly improved chromatographic separation (based on stationary phases with particles size smaller than 2  $\mu\text{m}$ ) enabled highly efficient separation of complex samples in very short run times. In addition, the coupled MS device enables not only the fast identification of known metabolites but can also provide via tandem mass spectrometry ( $\text{MS}^2$ ) fragmentation pattern valuable information concerning building blocks for uncharacterized metabolites. The high separation performance of LC-MS chromatography can also be used for fractionated bioactivity assays, by splitting the LC flow after elution; one part is used as usual for MS detection, whereas the residual flow is transferred into a 96 well plate, cultivated with the respective test organism. The bioactive LC-MS fractions can then be identified in 96 well plates and correlated to a specific time frame in the LC-MS run<sup>257</sup>. Liquid chromatography coupled to nuclear magnetic resonance spectroscopy (LC-NMR) might ideally accelerate the screening process by providing

additional structural information simultaneously to chromatographic separation that is not accessible by MS technique on its own. In addition, further steps that are necessary for the classical approaches *e.g.* utilizing various steps of fractionation, purification and subsequent structure elucidation using different spectroscopic methods<sup>26,258</sup>, can be circumvented through integration of a solid phase extraction (SPE) to perform online postcolumn enrichment of chromatographic peaks. Several studies showed an impressive value of LC-NMR for dereplication of myxobacterial crude extracts for various compound classes<sup>145,259–261</sup>. However, one of the major drawbacks of LC-NMR guided natural product research is the lower sensitivity in comparison to LC-MS<sup>262</sup>. Hence, most of the utilized LC-NMR setups run via a discontinuous workflow such as LC-SPE-NMR-MS systems<sup>263</sup>.

Gas chromatography-mass spectrometry (GC-MS) is predominantly used for the analysis of volatile and semi-volatile molecules and is a widespread analytical technique to identify and quantify molecules. GC-MS-based analysis is not only used in the field of natural product research but also in environmental science, forensics, health care, flavor and fragrances industry and food safety. The analysis of volatiles released by *M. xanthus*<sup>115</sup> and *S. aurantiaca*<sup>114</sup> leveraged the high sensitivity of GC-MS analyses to direct sample volatiles from agar plate cultures. Interpretation of mass spectra with respect to the molecular ion, the isotope pattern, known fragmentation pathways and using retention index increment systems can provide valuable information for a structural proposal of a new volatile natural product<sup>264</sup>. The structural proposal is afterwards supported by subsequent chemical synthesis, such as done for the volatiles released by *Chondromyces crocatus*<sup>265</sup>. GC-MS analysis is a powerful analytical platform in particular for terpene-like secondary metabolites, but the development of reliable and robust LC-MS platforms offers in comparison the advantages of quicker and less extensive extraction procedures, shorter run times and in particular the ability to identify and measure a broader range of natural products<sup>266</sup>.

### Metabolome mining based on LC-MS features

Since natural product analysis via LC-MS is very robust and highly reproducible, standardized workflows for the production of bacterial crude extracts and subsequent standardized LC-MS based measurement is an invaluable method to create a database for natural products. This database can then be useful not only to avoid rediscovery of known natural products, but also to help identify specific natural products in crude extracts prepared from uncharacterized strains. Annotation of natural products are typically based on retention time, accurate *m/z*, isotope pattern fit, and if applicable MS<sup>2</sup> data<sup>267</sup>. Reliable dereplication processes are essential to screen reliably for alternative bacterial producers of distinct natural products, but at the same time differentiate from derivatives thereof and potentially new chemical entities, as reported by Krug *et al.* for a set of 98 different strains of *M. xanthus*<sup>268</sup>. This study also underlines that manual investigation of numerous very complex LC-MS chromatograms of myxobacteria is neither feasible nor reliable. Decreasing relative intensity of a signal of interest and identification of the respective signal complicates the manual investigation. For that reason, the multivariate data set of LC-MS chromatograms can be dissected into molecular features, characterized by their exact mass (including all isotopic peaks and observed *m/z* of the monoisotopic peak) and standardized retention time. These molecular features are further processed to create matrices for comparative statistical analysis, a method termed “bucketing”. This type of processing is basically a two-dimensional binning (rectangular subspaces) of molecular features across all samples with respect to retention time and *m/z* value<sup>267</sup>. These subspaces are called buckets, which do not have intensity values (since these were not binned) but transferred the values as distinct entries to the buckets. The

created matrix termed as bucket table can then be used to apply comparative statistics based on multivariate analysis to highlight differences between distinct sample groups. One statistical procedure which utilizes bucket tables for comparative statistics in natural product research is principal component analysis (PCA)<sup>269</sup>. PCA is based on reducing the complexity of the bucket table, by transforming the complete multivariate LC-MS dataset into 'eigenvector' type data subsets defined as principle components (PCs)<sup>270</sup>. According to this transformation, the first PC (PC1) has the largest possible variance and each succeeding component in turn has the highest variance possible under the prerequisite that it is orthogonal to the preceding components, while the total number of principal components equals the number of data sets *e.g.* LC-MS chromatograms. A contribution analysis determines which bucket (based on molecular features) largely contribute to PC1, indicating these molecular features can be used to distinguish analyses from each other. PCA was mainly used in this work to investigate the differences in the secondary metabolism of generated genetic mutants in comparison to their wild types. Genetic disruption or induction of a biosynthetic gene that was beforehand not associated to a secondary metabolite, can be facilitated conveniently and better than by manual inspection via statistical evaluation of LC-MS datasets (**Figure 16**).

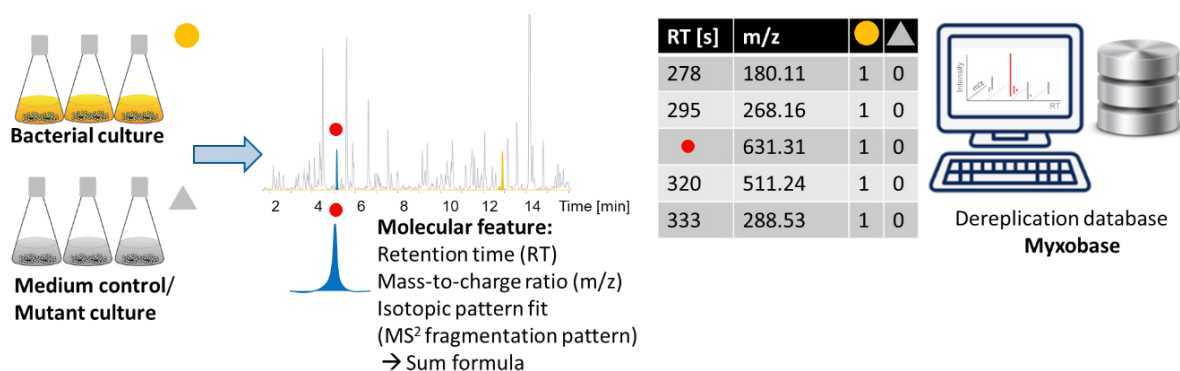
### Database-assisted dereplication

The statistical procedure is efficiently employed for myxobacterial secondary metabolomics such as the analysis of bacterial natural products and identification of medium-derived metabolites or the differential profiling of myxobacterial genetic mutants against and their corresponding wild types<sup>93,267,271</sup>. These myxobacterial secondary metabolomics studies relied on an assisted natural product dereplication database, storing accurate masses and retention times recorded on a standardized UHPLC-MS system. This in-house dereplication database called Myxobase, is used as a secondary metabolite annotation tool to annotate all the characterized metabolites in bacterial extracts thus limiting the possibilities of rediscovery (**Figure 16**).

In-house databases, which should include ideally all chromatographic data from known samples, are perfectly matched to the characteristics of the field of interest. The stored dataset in the Myxobase is adjusted to the secondary metabolites typically produced by myxobacteria. In contrast, several commercially available libraries such as the Dictionary of Natural Products<sup>272</sup>, AntiBase<sup>273</sup>, MarinLit<sup>274</sup> and MS databases such as MassBank<sup>275</sup>, Metlin<sup>276</sup>, mzCloud<sup>277</sup> and ReSpec<sup>278</sup> are also useful sources; however, these commercial libraries lack fundamental connectivity between each other. In addition, these databases are neither freely available, contain no hrMS data or do not allow customized use of its reference library<sup>26</sup>. For these reasons dereplication using commercial databases is rather complicated and a time-consuming process, which is a discouraging finding, since an improved connection between existing database libraries would lower the access barrier and a more intense data exchange between groups working in the field of natural products would be beneficial for the natural product research community<sup>26</sup>. An initial step towards a comprehensive data exchange was provided by Global Natural Products Social molecular networking (GNPS), an open-access knowledge platform for sharing MS<sup>2</sup> data<sup>279</sup>. The platform relies on spectral networking, in order to compare numerous spectra to each other, while spectra featuring similar fragmentation patterns are connected to each other. The ideal output from this spectral networking would show clustering of molecular features from secondary metabolites with similar fragmentation. For that reason, molecular features not clustered with known secondary metabolites or even featuring an independent MS<sup>2</sup> fragmentation pattern in this spectral network, might hint to a group of natural products with novel chemical scaffold. A recent study using GNPS for the



analysis of 146 marine *Salinispora* and *Streptomyces* strains yielded 15 molecular families of diverse natural products, showing that GNPS can be used for increased throughput screening<sup>280</sup>. Nevertheless, simple comparison of MS<sup>2</sup> spectra with spectra deposited in the GNPS is not sufficient for a comprehensive analysis of secondary metabolome data or provide an in-depth information on a structural level<sup>281</sup>.



**Figure 16.** Schematic depiction of LC-MS<sup>2</sup> based metabolome mining with database-assisted dereplication.

## Structure elucidation of natural products

During the so-called “golden age of antibiotics”<sup>282</sup>, structure elucidation of unknown natural products was a tedious procedure, requiring sometimes “indirect” methods to deduce the chemical structure of the unknown metabolite via decomposition into known fragments or total chemical synthesis. This time-consuming process also required substantial amount of pure compound (gram quantities) and often led to incorrectly assigned chemical structures, which resulted in structure revisions<sup>283</sup>. Since then, major technical progress has taken place to significantly reduce the time to characterize an unknown natural product and the amount of sample required for structure elucidation – these days 1 mg or less of compound may be enough for structure elucidation<sup>284</sup>. *De novo* structure elucidation relies majorly on NMR spectroscopy, even though methods like infrared (IR) spectroscopy and X-ray crystallography are useful to complement data obtained from NMR. In addition, mass spectrometry, a well-established analytical technique in natural product research (see above), can be efficiently used to support *de novo* structure elucidation of unknown secondary metabolites. The determination of the elemental composition of a molecule as well as gaining insights into sub-structures using MS<sup>n</sup> experiments are useful for the interpretation of NMR spectroscopic data<sup>285</sup>. In principle, the sample of interest is placed in a magnetic field and the NMR signal is generated via excitation of the nuclei sample with radio waves into nuclear magnetic resonance, which is detected with sensitive radio receivers. The NMR measurement is performed by excitation of all nuclei at the same time using a short radio frequency pulse. The excited nuclei emit the absorbed radiation generating a signal – a free-induction decay (FID) – that is then recorded as a function of ‘time after the pulse’<sup>286</sup>. The FID is the sum of all NMR responses originating from all excited nuclei with a superposition of all excited frequencies, resulting in a time-

domain signal<sup>286</sup>. This time-domain signal cannot be interpreted directly; however, the individual resonance frequencies can be extracted and the signal converted into the frequency domain spectrum applying a mathematical operation, the Fourier transformation. As the signal contains information on all resonance frequencies in only one measurement this signal can be repetitively measured in a shorter time, which results in an increased signal to noise ratio compared to the continuous wave mode, as it was applied in the beginning of the technique<sup>286,287</sup>. The non-destructive and non-invasive procedure of this technique underlines the general utility for natural product research despite its inherent lower sensitivity in comparison to optical spectroscopic methods<sup>288</sup>. Different NMR experiments are often required to elucidate the chemical structure of a natural product, such as one-dimensional (1D) <sup>1</sup>H, <sup>13</sup>C, <sup>15</sup>N NMR experiments as well as 2D homo- and heteronuclear correlation experiments<sup>286</sup>. The combination of different experiments allows it to connect the structural information to unambiguously assign the chemical structure of a newly discovered natural product. For elucidation of the absolute configuration of incorporated  $\alpha$  amino acids, two enantiomeric Marfey derivatization agents can be used such as reported for the absolute configuration assignment of the macyranones<sup>259</sup>.

As explained above, the structure elucidation of natural products is in general accomplished via NMR spectroscopy. X-ray crystallography techniques are an alternative procedure to elucidate the structure of several natural products such as described for the pyxidicyclines<sup>193</sup> or citreohybridonol<sup>289</sup>. Recently, a form of electron crystallography, known as microcrystal electron diffraction (MicroED), was also shown to be suitable for the structural determination of natural products but has yet to prove its broad applicability<sup>229,290</sup>.

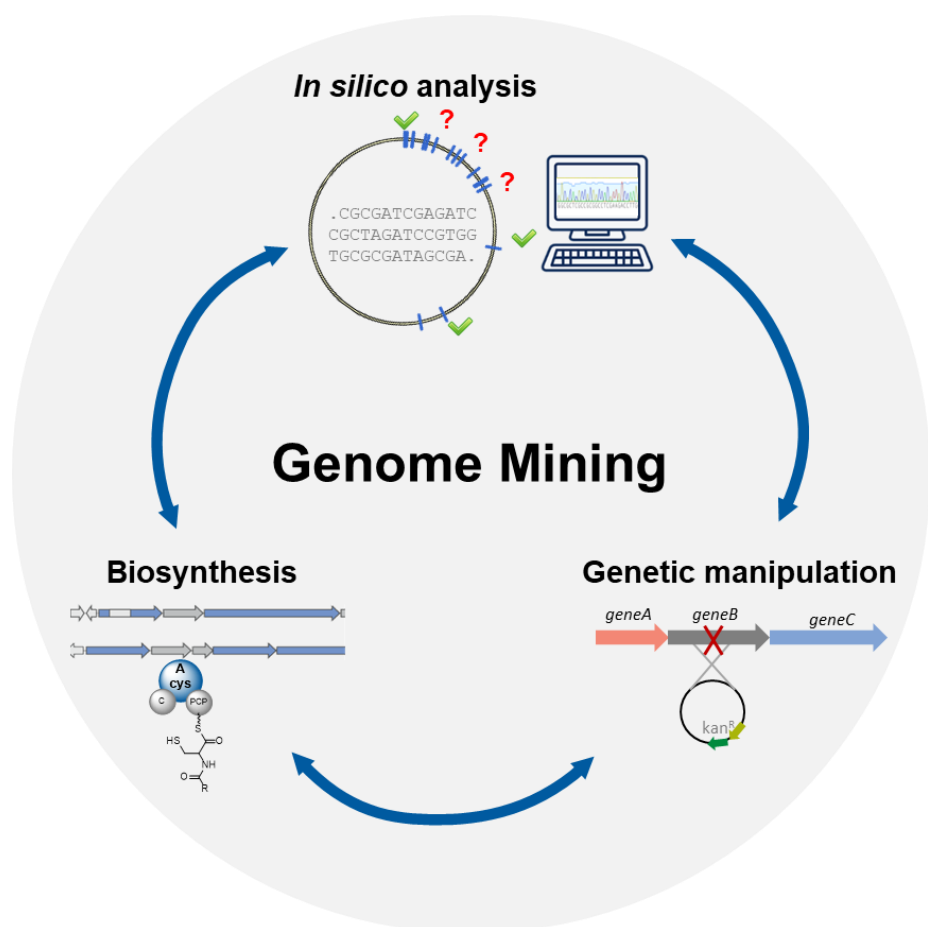
### 1.4.2 Bacterial Genetics – Key to the Biosynthetic Logic

Microbial and in particular bacterial biosynthetic pathways are known to consist of physically clustered groups of two or more genes within the genome. These clustered regions are referred to as biosynthetic gene clusters (BGCs, see above), which are necessary to encode all required biosynthetic enzymes for the production of specific natural products<sup>291</sup>. It is hypothesized, that evolutionary optimization of bacterial cellular processes, led to the clustering of genes encoding biosynthetic proteins within the bacterial genome as one or few polycistronic operons<sup>292</sup>. These clustered regions facilitates finding the connection between the biosynthetic genes and the produced secondary metabolite. In contrast, plant natural product genes are typically not found clustered in their chromosomes making identification of the full set of biosynthesis genes difficult<sup>293</sup>.

Remarkable progress in genetic engineering of microorganisms and the advent of affordable genome sequencing is transforming the field of natural product research. Because of this development, the paradigm in natural product discovery is shifting from traditional bioactivity-guided fractionation of crude extracts to genomic investigation and engineering of bacterial strains to produce complex secondary metabolites, which are available from the natural producer only under specific conditions (see also **Figure 15**). Accordingly, recent methods and strategies have seen development based on the availability of numerous whole-genome sequences. The general term for this bioinformatics-based approach to guide natural product discovery and investigation is genome mining<sup>294</sup>.

Genome mining comprises a broad spectrum of methods and technologies, starting from correlation of natural product production and the associated biosynthetic genes, up to genetically induced production of natural products, which are not produced under laboratory conditions (**Figure 17**). The major reason for the uprising of genome mining is that amenable whole-genome sequencing technologies accelerated the availability of genome sequences of important natural product producers and provided hereby the prerequisite for sophisticated genetic engineering of biosynthetic pathways. Furthermore numerous genome sequences revealed, to what extent bacteria are underexploited as resources of novel secondary metabolites – judged on the basis of genome-encoded pathways for natural product biosynthesis<sup>295</sup>. Computational analysis of genomes also points out that many secondary metabolite genes are downregulated under conventional laboratory conditions, which translates into a “silent” potential of the microbial world for the production of yet unknown natural products<sup>26</sup>. Regulatory genes play an essential role in bacterial metabolism and control the expression of one or more genes and can be divided into positive (promoting expression) and negative regulators (inhibiting expression).

In order to understand how the knowledge of bacterial genetics can access the biosynthesis of natural products, the next subsection briefly introduces the bioinformatics foundations and *in silico* analysis of bacterial genome sequences, while the following subsection presents the genetic tools available for myxobacteria and how the correlation of identified biosynthetic gene clusters can be accomplished.



**Figure 17.** Scheme of genome mining. The interplay of *in silico* analysis, genetic manipulation and the underlying biosynthesis can be harnessed for the discovery of new natural products and their biosynthetic logic next to genetic engineering of biosynthetic pathways.

### Bioinformatics – Identification of biosynthetic pathways

Bioinformatics prediction of undiscovered natural product structures based on biosynthetic genes is a powerful vision, since the experimental-based identification of natural products from microbes is a time-consuming and laborious procedure. The analysis of the whole-genome sequence of a given organism is nowadays a fast and relatively inexpensive process, which provides a comprehensive overview of the potential for the biosynthesis of natural products for a given strain. In addition, the full access to genetic information of a specific microorganism enables the correlation of a specific natural product with its BGC and the characterization of the underlying biosynthesis. Genome sequencing was initiated with the invention of the Sanger sequencing method in the seventies of the last century<sup>296–298</sup>, but in particular, the rapidly decreasing costs and increasing quality and quantity of acquired whole-genome sequencing data enabled the widespread usage of bioinformatics investigation in natural product research. Many new high-throughput sequencing methods have emerged such as 454/Roche, Illumina, PacBio, sequencing by synthesis and Ion semiconductor sequencing<sup>299–302</sup> that resulted in the deposition of large quantities of sequencing data from various organisms in public databases.

In contrast to other applications in natural sciences (such as phylogenetic studies, clinical bioanalysis, environmental studies), high quality genome sequences with very few contigs and no sequence ambiguity are fundamental for genetic investigation of natural product pathways<sup>303</sup>. BGCs are often composed of very long genes encoding multimodular machineries with repetitive domain architectures. Another challenge faced by genome sequencing of prolific producers of natural products are the guanine-cytosine rich genomes of actinobacteria<sup>26</sup> and myxobacteria<sup>67</sup>. An example underlining the importance of high-quality genome sequencing data is shown by the discrepancy of the previous draft genome of *Streptomyces clavuligerus* containing ambiguous and inaccurate nucleotides leading to the result that 30% (in total: 7163 protein coding genes) of all protein coding genes were incorrectly annotated<sup>304</sup>.

Nevertheless, genomic data alone does not provide any information without comprehension of biosynthetic pathways; this means that bioinformatics investigations dependent on the experimental data provided via molecular biology-based studies of genetic regulation of the encoded biosynthesis or biochemical characterization of biosynthetic enzymes (**Figure 17**). In reverse, the experimentally acquired knowledge of biosynthetic logics can be harnessed to provide a fundamental understanding of biosynthetic genetic regions and hence assist in the identification of potentially new natural products and its biosynthesis. This methodology has resulted in the identification of new classes of natural products such as the pyxidicyclines<sup>193</sup>.

Predictive bioinformatics tools have recently seen major advancements, which enabled more reliable identification of BGCs and the classification according to biosynthetic origin of the produced natural product, but also to predict genes responsible for secondary metabolite production according to their proposed function. The prime examples are natural products formed by multimodular biosynthetic machineries, where models have emerged that can predict substrate specificity of A and AT domains responsible for building block selection in each module of NRPS and PKS assembly lines. *In silico* prediction of substrate specificity relies in great measures on the collinear enzymatic logic in modular PKS and NRPS pathways and conserved catalytic recognition sequences (nevertheless deviations are possible and markedly frequent in myxobacteria)<sup>60</sup>. In particular, the discovery made by Stachelhaus

and coworkers regarding the basic rules of substrate specificity prediction of A domains, based on their primary amino acid sequence alignment, enabled the development of numerous A domain specificity prediction tools (*e.g.* NRPSpredictor2<sup>305</sup> or NRPSp<sup>306</sup>). *In silico* studies and experimentally conducted structure-function mutagenesis identified ten amino acid residues to account for the specificity-conferring code of substrate prediction of NRPSs within the binding pocket of the A domains commonly referred to as the non-ribosomal code<sup>143,307,308</sup>. This so-called non-ribosomal code was further refined to include residues in range of 8 Å from the binding pocket of the A domain<sup>309</sup>, which are anticipated to influence the A domain specificity upon interaction with the substrate.

This acquired knowledge and the developed tools are not only important to predict a structure of a natural product originating from a NRPS containing BGC, but also to engineer in a rational way the specificity of A domains in biosynthetic pathways to produce derivatives or entirely new natural products<sup>118</sup>. AntiSMASH 5.0<sup>310</sup> is currently the prediction tool that offers the broadest range of *in silico* analysis, whereas the recently developed bio- and chemoinformatics web application TransATor aims to enable *de novo* structural predictions for *trans*-AT polyketides, a special case of non-canonical type I PKS systems<sup>311</sup>. Nevertheless, *in silico* predictions are not universally reliable since many non-linear biosynthetic assembly lines exist in microorganisms. Iterative domain- or module usage, module skipping, split-modules or multiple copies of domains within a module present in PKS, NRPS and many hybrid PKS-NRPS pathways, makes structural prediction of the generated natural product very challenging<sup>312</sup>.

There are continuous processes to facilitate *in silico* prediction of specific RiPP-derived natural product classes such as the lanthipeptides – a group of RiPPs with polycyclic groups and intramolecular cross-linked thioether groups of two alanine residues called lanthionine – and the lassopeptides, featuring a “lasso” forming peptide chain, cross-linking the molecule with additional disulfide bonds<sup>313</sup>. Despite reported cases of recently discovered new classes of RiPPs, several obstacles hinder genome mining for novel RiPP BGCs. The minimal BGC encoding a functional RiPP pathway may comprise only two genes: a gene encoding the precursor peptide and another gene encoding a tailoring enzyme. Furthermore, the prevalence of numerous putative small peptides encoded throughout the genomes of potential producers makes it difficult to predict correctly either manually or by automated gene annotation software the presence of BGCs belonging to unidentified RiPPs. Furthermore, the lack of homology or signature protein domains encompassing different groups of RiPPs, impedes the identification of those gene clusters in contrast to terpenes, polyketides or non-ribosomal peptides<sup>314</sup>. RiPP biosynthetic genes are less unified in terms of conserved catalytic domains but instead by a basic logic of post-translationally modification of an unrestrained ribosomal peptide. It was described recently that undetected peptide-binding domains in tailoring enzymes, spanning the many of the known prokaryotic RiPP classes, contain a conserved domain, the so-called RiPP precursor peptide recognition element (RRE)<sup>231</sup>.

As set out above, the power of *in silico* prediction of biosynthetic pathways is already impressive and contributed in large measures to the acquired understanding of the myxobacterial biosynthetic pathways in this thesis. However, *in silico* methods feature inherent limitations connected to the knowledge of biosynthetic pathways previously characterized through genetic studies and subsequent in-depth biochemical investigations. The next subsection summarizes the available methods and genetic tools in myxobacteria to correlate the identified BGCs with actual natural products.

### Genetic tools

When a genetic locus can be provisionally connected with the biosynthesis of natural products via *in silico*-based methods, the confidence of this correlation significantly depends on the nature of the biosynthetic machinery. Other factors such as the sequence quality of the whole-genome sequence, the presence of other genetic loci putatively accounting for the formation of the natural product and the proportion of unprecedented biosynthetic building blocks or chemical moieties within the structure of the natural product are critical for the reliability of metabolome-genome correlation<sup>315</sup>. To this end, modular biosynthetic pathways such as PKS, NRPS and PKS-NRPS hybrids features several advantages concerning their biosynthetic logic (see above), that simplify the connection of metabolites to the biosynthetic genes.

There are three main approaches to conduct genome mining with myxobacteria (and other microorganisms) *i.e.* to correlate pathways from myxobacterial genomes to the observed secondary metabolome (**Figure 18A, B**). The first approach includes gene knockout or disruption experiments to directly establish the connection between a gene and the generated natural product. The subsequent “biosynthetic modification” leading to changes in MS signals serves as analytical readout to deduce the putative involvement of the targeted gene in the biosynthesis. This profiling method is less applicable to activate “silent” gene clusters, unless a negative regulator is disrupted.

Instead the second strategy, the insertion of a promoter in front of the putative biosynthetic gene cluster can lead to promoter-induced gene expression of a biosynthetic operon, which may sometimes allow the straightforward detection of a produced metabolite. This “activation” method faces limitations if the compound cannot be produced for reasons other than lacking gene expression, *e.g.* due to missing precursors or defective key enzymes. In addition, many biosynthetic gene clusters consists of multiple gene operons<sup>316</sup>, which would require the insertion of a promoter in front of each biosynthetic gene operon.

The third option is heterologous expression *i.e.* the transfer of biosynthetic genes into a well-studied foreign host. Although this is the most labor-intense and sophisticated route, with its own set of challenges, it offers great opportunities for the subsequent modification of the BGC in the heterologous producer. A great advantage of heterologous expression is to circumvent often unfavorable growth characteristics, the overall susceptibility to various stress factors and the lack of basic genetic tools and manipulation observed with many myxobacteria<sup>60</sup>. Heterologous expression of all biosynthetic genes independent from its origin can be conducted with reasonable effort and gives reliable correlation to the produced metabolite. In addition, heterologous expression of complete pathways is desirable as it offers tremendously expanded possibilities for rational genetic alteration of biosynthetic routes, since a well-studied host provides in general a greater availability of genetic tools than the native host does.

For example, microbial heterologous production platforms have become an attractive alternative to traditional isolation of phytochemicals via chemical extraction from natural host plants<sup>317</sup> or using plant cell culture<sup>318–320</sup>. The microbial production of plant-derived natural products is independent from agricultural land, crop yields, water, pests or extreme weather and can result in a less time-consuming supply of novel derivatives of the natural product by genetic engineering of the microbial producer<sup>293</sup>.

Reconstitution of the hydrocodone and thebaine biosynthesis<sup>321</sup> via a stepwise *in vivo* fermentation process of four engineered *E. coli* strains and the heterologous production of the paclitaxel precursors taxadiene<sup>322</sup> and oxygenated taxanes<sup>323</sup> in *E. coli*, underlines the applicability of bacteria to produce complex phytochemicals.

A number of myxobacterial BGCs have been successfully expressed in *M. xanthus* sp., which has been proven to work as a broad host to produce chemically diverse natural products<sup>60</sup>. Other strains within the suborder of *Cystobacterineae* can also be used as myxobacterial heterologous production platform such as *Stigmatella aurantiaca* DW4/3-1 (closely related to *M. xanthus* sp.) and the moderately thermophilic *Corallococcus macrosporus* GT-2<sup>60</sup>. Some strains from the suborder *Sorangineae*, in particular *Sorangium cellulosum* sp., are accessible to genetic manipulation that encourages the prospect to establish another myxobacterial expression platform. In contrast, strains from the suborder of *Nannocystineae* feature unfavorable characteristics such as slow growth and seem to be “inert” towards genetic manipulation<sup>60</sup>.

Moreover, organism-independent methods – *in vitro* reconstitution of the biosynthesis using recombinantly produced biosynthetic enzymes – provide not only experimental proof for the associated biosynthetic genes, but also allow in-depth biosynthetic characterization to elucidate single catalytic steps of a biosynthetic pathway. This approach can reveal profound insights concerning the biosynthetic mechanisms, but for many cases, these approaches are very time-consuming, challenging or not possible.

Genetic tools which are available for myxobacteria are still limited. The important considerations for the successful genetic manipulation of a BGC comprises the construction of a suitable genetic construct, the transfer of the genetic material into the myxobacterial strain and the stable integration and maintenance of the genetic construct in the host bacterium. There are numerous cloning, construction and mobilization methods of DNA, ranging from PCR-based strategies for relatively small genetic constructs to the usage of synthesized DNA fragments<sup>324</sup>. Since these methods are independent from the biological properties of myxobacteria, many different approaches have been used for cloning and mobilization of myxobacterial BGCs to generate genetic constructs.

A versatile classical method uses large insert libraries based on cosmids or fosmids that clone 30–35 kb regions or bacterial artificial chromosomes (BACs) that can contain over 100 kb. A major advantage of this method is that the genomic DNA might originate from a mixed population (*i.e.* environmental DNA) or from isolated bacterial cultures. Probing of specific cosmid libraries with PCR-primer for the presence of PKS-and/or NRPS-encoding genes enabled in former times the identification of the BGC of epothilone in *Sorangium cellulosum* So ce90<sup>325</sup>, the stigmatellin and the myxalamid BGC in *S. aurantiaca* Sg a15<sup>326</sup>. On the other hand, the successful heterologous expression of the oxytetracycline gene cluster<sup>327</sup>, tubulysin gene cluster<sup>328</sup> and haliangicin<sup>329</sup> in *M. xanthus* exemplifies the usefulness of these cosmid libraries for genetic engineering approaches.

Genetic modification of mobilized myxobacterial BGC pathways have been mainly conducted by using linear-circular homologous Red/ET recombination instead using conventional restriction ligation such as the reconstituted tubulysin gene cluster originating from two cosmids<sup>328</sup>. The invention of modern methods such as direct cloning of the BGC from genomic DNA, as described for the type III PKS BGC

in MCy9487 (chapter 6) or utilizing synthetic DNA fragments are continuing to replace many traditional methods, due to decreasing cost of synthetic DNA and less laborious cloning strategies. For example, synthetic modular versions of myxobacterial BGC were successfully expressed in *M. xanthus* to generate new derivatives of myxochromides<sup>330</sup>, vioprolides<sup>331</sup> and agryrins<sup>332</sup>.

In general, there are numerous concepts and methods available to access relatively huge biosynthetic gene clusters for subsequent heterologous expression, which rely on PCR amplified biosynthetic fragments and *in vitro* or *in vivo* homologous recombination<sup>324</sup>. For example, a PCR-based transformation-associated recombination (TAR) assembly utilizes the *in vivo* homologous recombination system of *Saccharomyces cerevisiae* to assemble numerous PCR fragments (with homology flanked DNA regions) with the vector backbone in yeast. This PCR-based strategy was used to reorganize the genetic architecture of a downregulated type II PKS BGC originating from *Pyxidococcus fallax* An d48<sup>193</sup>, which was afterwards transferred into the host *M. xanthus* and *S. aurantiaca* DW4/3-1 for heterologous production of the pyxidicyclines<sup>193</sup>.

The transfer of genetic material among bacteria is termed horizontal gene transfer that is significantly different from the vertical transmission of DNA from parent to offspring (reproduction). Horizontal gene transfer can be classified according to the specific mechanism into transformation (transfer of genetic material across the cell membrane), conjugation (transfer of genetic material between two bacterial cells in direct contact) and transduction (injection of foreign DNA by a bacteriophage virus into the host bacterium). All three gene transfer mechanisms are used as genetic tools for myxobacteria<sup>333,334</sup>. The primarily conducted process to genetically manipulate myxobacterial strains is transformation, which has the advantage that the genetic construct passes through the fermentation medium and the uptake is only dependent on the recipient bacterium<sup>335</sup>. The electroporation-based transformation procedure established for *M. xanthus* DK1622 proved to be very robust and forms the conceptual foundation to establish genetic manipulation strategies for newly discovered and yet uncharacterized myxobacterial strains<sup>336</sup>. The basic requirement for electroporation-based transformation is homogenous suspension-like growth in liquid culture in a fermentation medium.

While most strains within the genus *Myxococcus* and at least some strains within the suborder *Cystobacterineae* fulfill this requirement, other myxobacterial strains within the suborder *Sorangineae* and *Nannocystineae* tend to form cell clumps. A few myxobacterial strains from the suborder *Sorangineae* are genetically accessible by triparental conjugation procedures, such as *Streptomyces* sp. and *Pseudomonas* sp., which rely on bacterial conjugation procedures. For example the chivosazol gene cluster originating from *Sorangium cellulosum* So ce10<sup>337</sup> and *Sorangium cellulosum* So ce56<sup>338</sup>, and the crocagin RiPP pathway from *Chondromyces crocatus* Cm c5<sup>227</sup> have been identified by insert-specific homologous recombination through a triparental conjugation procedure in the respective native host.

After the successful transfer of the genetic construct (for heterologous expression or genetic manipulation), the last critical step for genetic engineering requires stable integration and maintenance of the genetic construct in the manipulated bacterium. There are some reports for self-replicating plasmids and their biosynthetic potential<sup>339–341</sup>. Nevertheless, the chromosomal integration of non-native DNA into myxobacterial hosts has been proven for a long time to work quite efficiently and enables direct genetic manipulation of the myxobacterial organism. Chromosomal integration of genetic



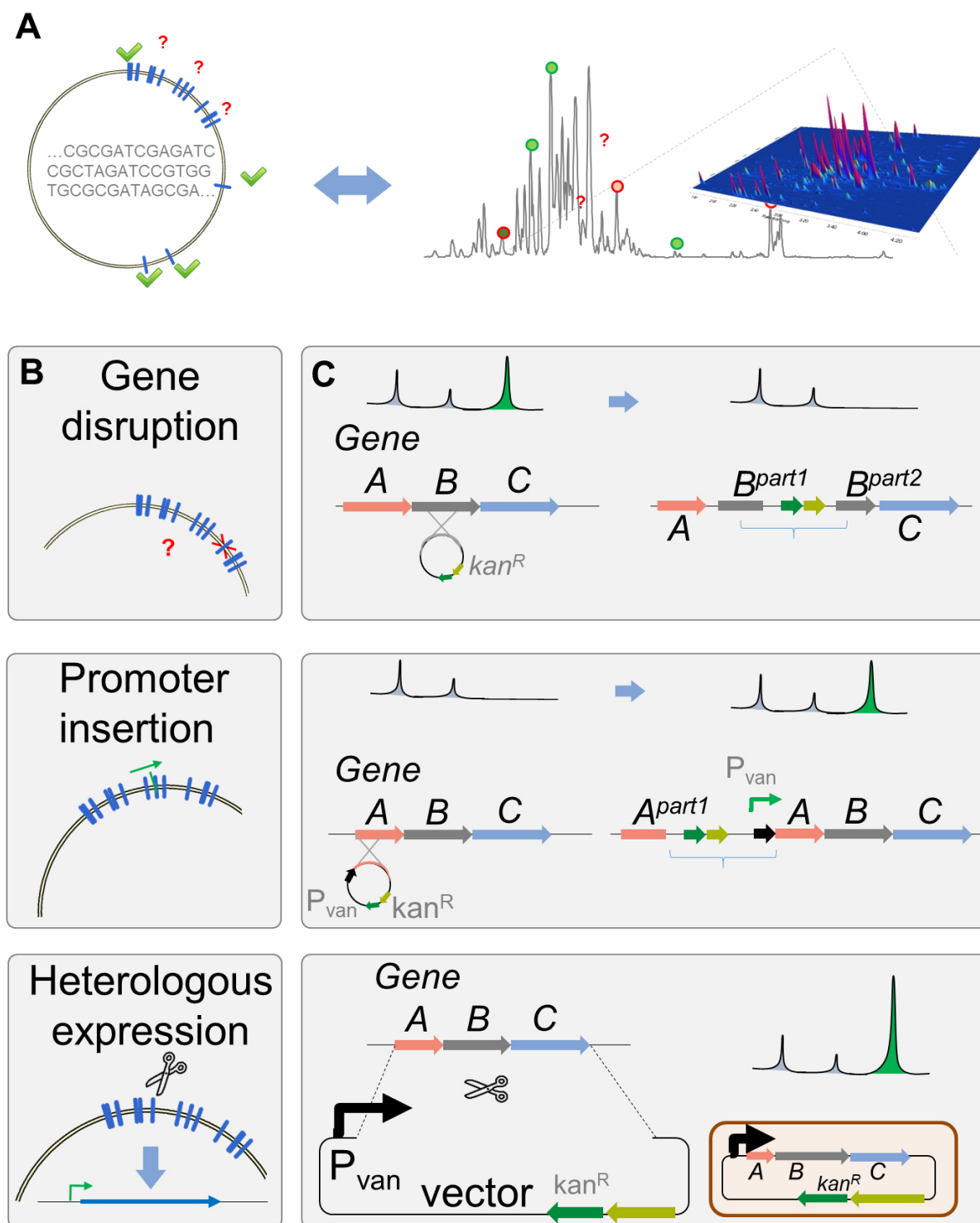
constructs can be accomplished in myxobacteria by using a transposon located on the plasmid pMycoMar<sup>342,343</sup>, which allows very efficient random integration into the myxobacterial genome. The Mx8 phage integrase performs in contrast to a transposon targeted genomic integration into the phage attachment sites attB1 or attB2 (if available in the myxobacterial chromosome). Both methods can efficiently be used to integrate large genetic constructs encoding biosynthetic pathways but lack definite specificity regarding the genomic insertion site.

For that reason site-specific chromosomal integration catalyzed by endogenous recombinases from the recipient myxobacterium according to the DNA homology region (around 1 kb)<sup>344</sup>, is the method of choice for biosynthetic investigation even though the genomic integration efficiency is lower than the aforementioned procedures.

This so-called single-crossover recombination strategy enables the integration of genetic constructs into the bacterial genome. With this approach, genes can be disrupted or partially engineered such as the replacement of a promoter sequence. Genetic manipulation procedures based on site-specific single-crossover recombination could be established in numerous strains<sup>60</sup>. Single-crossover recombination does not provide seamless genetic alterations, instead the entire genetic construct/plasmid is site-specifically integrated (**Figure 18C**).

An established but time-consuming and laborious protocol for seamless gene deletion/alteration in *M. xanthus* is based on double-crossover integration plasmids, including typically a kanamycin resistance gene and a counter-selectable genetic marker<sup>345</sup>. Recent protocols for site-specific and seamless genetic manipulation in *M. xanthus* are based on Cre/loxP recombination systems<sup>346</sup> or CRISPR/Cas9 systems<sup>347,348</sup>. All of these seamless gene deletion/alteration procedures have in common: they are very time consuming, laborious, less efficient in comparison to single-crossover-based recombination and have to be newly implemented for each host. Due to these limitations, the above-mentioned approaches are in principle less attractive to correlate the biosynthetic genes from new strains to their underlying secondary metabolites.

In summary, the available genetic tools for myxobacterial strains are still limited compared to *Streptomyces* or especially *E. coli*<sup>118</sup>. Nevertheless, these limitations in available genetic tools have not diminished the outcome of biosynthetic research in myxobacteria at all but have focused these investigations in certain directions as mentioned earlier. Moreover, most importantly these obstacles did not hinder genome mining approaches to enrich the biosynthetic insights of myxobacterial research.



**Figure 18.** (A) Metabolome-genome correlation in myxobacteria. (B) The main procedures to correlate biosynthetic genes to their produced natural products are gene disruption (top), the insertion of a promoter in front of downregulated biosynthetic genes (middle) and the transfer of biosynthetic genes into a heterologous host (bottom). (C) Scheme of single crossover-based recombination gene disruption (top) and promoter insertion (middle). The vector backbone is fully integrated into the chromosome. The chromosomal locus of foreign biosynthetic genes is often less critical for heterologous expression (bottom).

## 1.5 Outline of this thesis

The work of this thesis is focused on the biosynthetic logic of myxobacteria, with a particular interest to gain insights into the diversity of produced natural products including those that originate from biosynthetic pathways less frequently reported in myxobacteria (**Figure 19**).

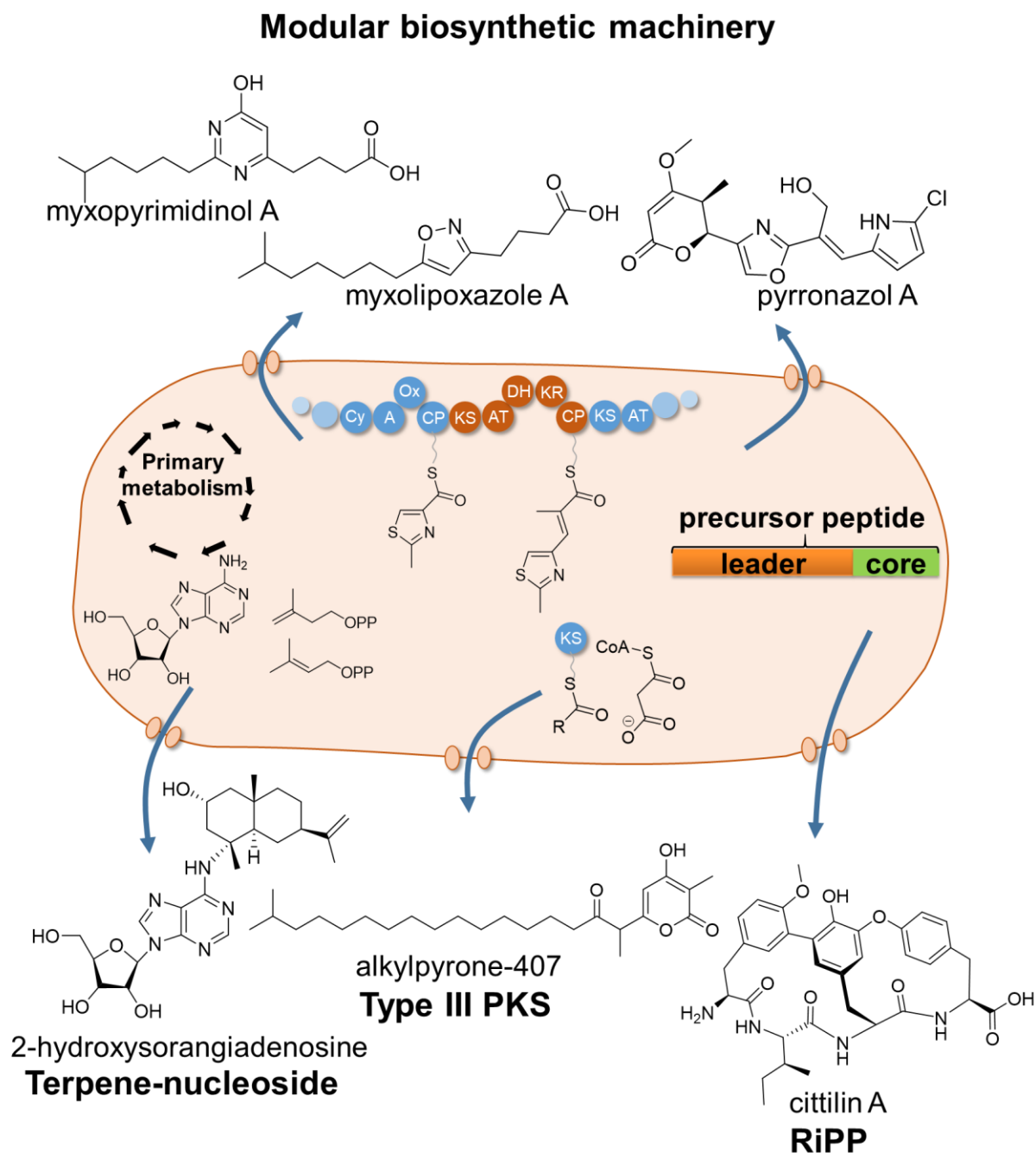
The first part (chapter 2) of thesis reflects the prototypical approach to elucidate the biosynthetic logic of a myxobacterial natural product. The pyrroazols have been isolated previously from *Nannocystis pusilla* Ari7 and *N. pusilla* strain Na a174; however at that time, the biosynthesis of these chlorinated pyrrole-oxazole-pyrones remained elusive<sup>349</sup>. In chapter 2, the genome of *N. pusilla* Ari7, the producer of pyrroazol A1 and B, was sequenced and investigated with bioinformatics methods. The identified BGC and the proposed biosynthetic PKS-NRPS pathway was supported by feeding experiments with stable isotopes of three major biosynthetic building blocks<sup>144</sup>. The primarily *in silico*-based elucidation of the biosynthetic pathway was possible, since the pyrroazols are generated by a multimodular biosynthetic assembly line, which follows the collinearity rule. This study highlights how specific architectures of myxobacterial natural products are biosynthesized and provides an instructive example of how powerful genome sequencing combined with feeding experiments has become in establishing biosynthetic routes.

Chapter 3 resembles the common natural product discovery workflow of the pyrroazols, where the isolation of the cittilins preceded their biosynthetic assignment. In contrast to the pyrroazols, genome sequence analysis and *in silico* investigation of confirmed cittilin producers were not sufficient to provide a proposal for their biosynthetic origin. To confirm the origin of cittilins as RiPPs, it was required to conduct gene disruption experiments, heterologous expression of the genes encoding all required biosynthetic enzymes producing cittilins, and *in vitro* biochemical characterization. The heterologous expression of the cittilin gene cluster in *Streptomyces albus* del14 and the *in vitro* reconstitution of important biosynthetic steps had a synergistic effect to clarify the underlying biosynthesis. The formation of the bicyclic ring – which requires the catalytic activity of only one CYP450 enzyme – and the fact that the cittilins are alongside with the crocagins<sup>227</sup> the only characterized RiPP natural products from myxobacteria, underlines the significance of the unprecedented cittilin gene cluster<sup>350</sup>. In addition, the outcome of this chapter provides the foundation for potential production of cittilin derivatives and to reveal the unique mechanism of this CYP450 enzyme reaction.

In chapter 4, a bioactivity-guided natural product screening of the myxobacterium *Vitiosangium cumulatum* MCy10943<sup>T</sup> enabled the discovery of the rare sesquiterpene-nucleoside 2-hydroxysorangiadenosine and the rediscovery of sorangiadenosine<sup>351</sup>. Sorangiadenosine has been isolated previously from the culture broth of *Sorangium cellulosum* KM1003<sup>352</sup>. Despite its biological activity against a broad range of bacterial strains and its hybridized chemical scaffold consisting of a sesquiterpene and a nucleoside, nothing was known about its biosynthetic origin. In this chapter, fundamental *in silico* analysis of the genome sequence of MCy10943 and stable isotope-labeled precursor feeding experiments enabled the identification of the biosynthetic origin and pathway of this unique hybrid natural product. This work also describes the first nucleoside-hybrid biosynthetic pathway in myxobacteria, which might be a stepping-stone for the identification of additional nucleosides or nucleoside-hybrid natural product pathways in the order of *Myxococcales*.

Chapter 5 describes how a metabolome-guided natural product isolation approach initiated the discovery of two different natural products, each displaying rare biosynthetic heterocycles. Both natural products, myxolipoxazole A and myxopyrimidinol A originate from an atypical PKS-NRPS hybrid biosynthetic pathway. The biosynthetic origin has been identified through retrobiosynthetic *in silico* investigation combined with feeding experiments using stable isotopes of biosynthetic building blocks and genetic manipulation of the producer strain *Myxococcus xanthus* Mx x48. The identified assembly line shows a high degree of biosynthetic variety in the production of chemically diversified natural products, which gave further hints towards the evolutionary development of this biosynthetic machinery. Genetic engineering of *M. xanthus* Mx x48 tremendously increased and stabilized the production of all genetically associated secondary metabolites, which paved the way to isolate the major natural products of this biosynthetic pathway. To the best of our knowledge, this is the first work that correlates an isoxazole- or 4-pyrimidinol-containing natural product to its multimodular biosynthetic machinery and has been confirmed by genetic experiments.

The discovery of the alkylpyrones in *M. xanthus* DK1622 (chapter 6) was initiated differently from the aforementioned studies. Myxobacteria have contributed majorly to the greater diversity and understanding the biosynthetic logic of microbial natural products. However, the focus was oftentimes on multimodular biosynthetic pathways, while natural products with other structural scaffolds like terpenes, phenyl-propanoids, alkaloids and type III polyketides are under-represented. This observation is exemplified by the fact, that no pure type III PKS-derived small molecule has previously been isolated from myxobacteria<sup>202</sup>. Additionally, bioinformatics insights indicated numerous type III PKSs to be present in myxobacterial genomes and none of the identified genes have been connected to a matching metabolite in the myxobacterial secondary metabolome. This finding stimulated speculation that the myxobacterial type III PKS pathways could be downregulated under laboratory fermentation conditions so the produced amount of natural product remains below the limit of detection. In this chapter, genomic investigation of myxobacteria focusing on type III polyketide synthase genes and a self-resistance-guided screening centered on co-localized pentapeptide repeat proteins resulted in the identification of the first uncommon alkylpyrones from the myxobacterial model strain *M. xanthus* DK1622. The heterologous expression of the biosynthetic type III PKS genes and promoter-induced gene expression in the native host enabled identification, structure elucidation and bioactivity testing of alkylpyrone-407, revealing its potent topoisomerase activity *in vitro*.



**Figure 19.** An illustrated summary of biosynthetic logic of myxobacterial natural products covered in this thesis.

## 1.6 References

1. Hanson, J. R. *Natural products. The secondary metabolites / James R. Hanson* (Royal Society of Chemistry, Cambridge, 2003).
2. Williams, D. H., Stone, M. J., Hauck, P. R. & Rahman, S. K. Why are secondary metabolites (natural products) biosynthesized? *J. Nat. Prod.* **52**, 1189–1208 (1989).
3. Newman, D. J. & Cragg, G. M. Natural Products as Sources of New Drugs over the Nearly Four Decades from 01/1981 to 09/2019. *J. Nat. Prod.* **83**, 770–803 (2020).
4. Cragg, G. M. & Newman, D. J. Natural products: a continuing source of novel drug leads. *Biochim. Biophys. Acta* **1830**, 3670–3695 (2013).
5. Bernardini, S., Tiezzi, A., Laghezza Masci, V. & Ovidi, E. Natural products for human health: an historical overview of the drug discovery approaches. *Nat. Prod. Res.* **32**, 1926–1950 (2018).
6. Krishnamurti, C. & Rao, S. C.C. The isolation of morphine by Serturmer. *Indian J. Anaesth.* **60**, 861 (2016).
7. Morveau, L.B.G. de *et al. Annales de chimie et de physique* (Masson, 1820).
8. Okombo, J., Ohuma, E., Picot, S. & Nzila, A. Update on genetic markers of quinine resistance in *Plasmodium falciparum*. *Mol. Biochem. Parasitol.* **177**, 77–82 (2011).
9. Briquet Pierre. *Traité thérapeutique du quinquina et de ses preparations* (1853).
10. Fischer, J. & Ganellin, C. R. *Analogue-based Drug Discovery II* (Wiley, 2010).
11. Jeffreys, D. *Aspirin: The Remarkable Story of a Wonder Drug* (Bloomsbury Publishing, 2008).
12. Fleming, A. On the Antibacterial Action of Cultures of a Penicillium, with Special Reference to their Use in the Isolation of *B. influenzae*. *Br. J. Exp. Pathol.* **10**, 226–236 (1929).
13. Williams, K. J. The introduction of ‘chemotherapy’ using arsphenamine – the first magic bullet. *J. R. Soc. Med.* **102**, 343–348 (2009).
14. Wainwright, M. The First Miracle Drugs: How the Sulfa Drugs Transformed Medicine. *Perspect. Biol. Med.* **50**, 639–642 (2007).
15. Gunatilaka, A. A. Natural products from plant-associated microorganisms: distribution, structural diversity, bioactivity, and implications of their occurrence. *J. Nat. Prod.* **69**, 509–526 (2006).
16. Dewick, P. M. *Medicinal Natural Products* (Wiley, 2009).
17. Arcamone, F. *et al.* Adriamycin, 14-hydroxydaimomycin, a new antitumor antibiotic from *S. peucetius* var. *caesius*. *Biotechnol. Bioeng.* **11**, 1101–1110 (1969).

18. Umezawa, H., Maeda, K., Takeuchi, T. & Okami, Y. New antibiotics, bleomycin A and B. *J. Antibiot.* **19**, 200–209 (1966).
19. Davenport-Hines, R. P. T. & Slinn, J. *Glaxo. A history to 1962 / R.P.T. Davenport-Hines and Judy Slinn* (Cambridge University Press, Cambridge, 1992).
20. Dutcher, J. D. The Discovery and Development of Amphotericin B. *Dis. Chest* **54**, 296–298 (1968).
21. Borel, J. F., Feurer, C., Gubler, H. U. & Stähelin, H. Biological effects of cyclosporin A: A new antilymphocytic agent. *Agents Actions* **6**, 468–475 (1976).
22. Merluzzi, V. J. & Adams, J. *The Search for Anti-Inflammatory Drugs: Case Histories from Concept to Clinic. The history of the discovery and development of Cyclosporin (Sandimmune®)* (Birkhäuser Boston, 2012).
23. Benjamin, D., Colombi, M., Moroni, C. & Hall, M. N. Rapamycin passes the torch: a new generation of mTOR inhibitors. *Nat. Rev. Drug Discov.* **10**, 868–880 (2011).
24. Schwecke, T. *et al.* The biosynthetic gene cluster for the polyketide immunosuppressant rapamycin. *Proc. Natl. Acad. Sci. USA* **92**, 7839–7843 (1995).
25. Pye, C. R., Bertin, M. J., Lokey, R. S., Gerwick, W. H. & Linington, R. G. Retrospective analysis of natural products provides insights for future discovery trends. *Proc. Natl. Acad. Sci. USA* **114**, 5601–5606 (2017).
26. Hug, J. J., Bader, C. D., Remškar, M., Cirnski, K. & Müller, R. Concepts and Methods to Access Novel Antibiotics from Actinomycetes. *Antibiotics* **7**, 44 (2018).
27. World Health Organization. Global Antimicrobial Resistance Surveillance System (GLASS) Report. Early implementation (2018).
28. Rice, L. B. Federal Funding for the Study of Antimicrobial Resistance in Nosocomial Pathogens: No ESKAPE. *J. Infect. Dis.* **197**, 1079–1081 (2008).
29. Jim O'Neill. Antimicrobial Resistance: Tackling a crisis for the health and wealth of nations., 2014.
30. de Kraker, Marlieke E. A., Stewardson, A. J. & Harbarth, S. Will 10 Million People Die a Year due to Antimicrobial Resistance by 2050? *PLoS Med.* **13** (2016).
31. D'Costa, V. M. *et al.* Antibiotic resistance is ancient. *Nature* **477**, 457–461 (2011).
32. Wright, G. D. Antibiotic resistance in the environment: a link to the clinic? *Curr. Opin. Microbiol.* **13**, 589–594 (2010).
33. Wright, G. D. The antibiotic resistome: the nexus of chemical and genetic diversity. *Nat. Rev. Microbiol.* **5**, 175–186 (2007).
34. van der Meij, A., Worsley, S. F., Hutchings, M. I. & van Wezel, G. P. Chemical ecology of antibiotic production by actinomycetes. *FEMS Microbiol. Rev.* **41**, 392–416 (2017).

35. Heine, D. *et al.* Chemical warfare between leafcutter ant symbionts and a co-evolved pathogen. *Nat. Commun.* **9**, 2208 (2018).
36. Pawlowski, A. C. *et al.* A diverse intrinsic antibiotic resistome from a cave bacterium. *Nat. Commun.* **7**, 1–10 (2016).
37. Shallcross, L. J. & Davies, D. S. C. Antibiotic overuse: a key driver of antimicrobial resistance. *The British journal of general practice : the journal of the Royal College of General Practitioners* **64**, 604–605 (2014).
38. Wise, R. *et al.* Antimicrobial resistance. Is a major threat to public health. *BMJ (Clinical research ed.)* **317**, 609–610 (1998).
39. Blair, J. M. A., Webber, M. A., Baylay, A. J., Ogbolu, D. O. & Piddock, L. J. V. Molecular mechanisms of antibiotic resistance. *Nat Rev Micro* **13**, 42–51 (2014).
40. Peterson, E. & Kaur, P. Antibiotic Resistance Mechanisms in Bacteria: Relationships Between Resistance Determinants of Antibiotic Producers, Environmental Bacteria, and Clinical Pathogens. *Front. Microbiol.* **9**, 2928 (2018).
41. Almabruk, K. H., Dinh, L. K. & Philmus, B. Self-Resistance of Natural Product Producers. Past, Present, and Future Focusing on Self-Resistant Protein Variants. *ACS Chem. Biol.* **13**, 1426–1437 (2018).
42. Procópio, R. E. d. L., Silva, I. R. d., Martins, M. K., Azevedo, J. L. d. & Araújo, J. M. d. Antibiotics produced by *Streptomyces*. *Braz. J. Infect. Dis.* **16**, 466–471 (2012).
43. Espinell-Ingroff, A. V. *Medical Mycology in the United States: A Historical Analysis (1894–1996)* (Springer Netherlands, 2013).
44. Struyk, A. P. *et al.* Pimaricin, a new antifungal antibiotic. *Antibiot. Annu.* **5**, 878–885 (1957).
45. Ehrlich, J., Bartz, Q. R., Smith, R. M., Joslyn, D. A. & Burkholder, P. R. Chloromycetin, a New Antibiotic From a Soil Actinomycete. *Science* **106**, 417 (1947).
46. Debono, M. *et al.* A21978C, a complex of new acidic peptide antibiotics: isolation, chemistry, and mass spectral structure elucidation. *J. Antibiot.* **40**, 761–777 (1987).
47. Pham, J. V. *et al.* A Review of the Microbial Production of Bioactive Natural Products and Biologics. *Front. Microbiol.* **10**, 1404 (2019).
48. Burg, R. W. *et al.* Avermectins, new family of potent anthelmintic agents: Producing organism and fermentation. *Antimicrob. Agents Chemother.* **15**, 361–367 (1979).
49. Li, J. W. H. & Vederas, J. C. Drug Discovery and Natural Products: End of an Era or an Endless Frontier? *Science* **325**, 161–165 (2009).
50. Munoz-Dorado, J., Marcos-Torres, F. J., Garcia-Bravo, E., Moraleda-Munoz, A. & Perez, J. Myxobacteria: Moving, Killing, Feeding, and Surviving Together. *Front. Microbiol.* **7**, 781 (2016).



51. Mauriello, E. M. F., Mignot, T., Yang, Z. & Zusman, D. R. Gliding motility revisited: how do the myxobacteria move without flagella? *Microbiol. Mol. Biol. Rev.* **74**, 229–249 (2010).
52. Sabass, B., Koch, M. D., Liu, G., Stone, H. A. & Shaevitz, J. W. Force generation by groups of migrating bacteria. *Proc. Natl. Acad. Sci. USA* **114**, 7266–7271 (2017).
53. Faure, L. M. *et al.* The mechanism of force transmission at bacterial focal adhesion complexes. *Nature* **539**, 530 (2016).
54. Velicer, G. J. & Vos, M. Sociobiology of the myxobacteria. *Annu. Rev. Microbiol.* **63**, 599–623 (2009).
55. Wenzel, S. C. & Müller, R. Myxobacterial natural product assembly lines: fascinating examples of curious biochemistry. *Nat. Prod. Rep.* **24**, 1211–1224 (2007).
56. Homma, Y. Perforation and lysis of hyphae of *rhizoctonia-solani* and conidia of *cochliobolus-miyabeanus* by soil myxobacteria. *Phytopathology* **74**, 1234–1239 (1984).
57. Reichenbach, H. The ecology of the myxobacteria. *Environ. Microbiol.* **1**, 15–21 (1999).
58. Shimkets, L. J. Intercellular signaling during fruiting-body development of *Myxococcus xanthus*. *Annu. Rev. Microbiol.* **53**, 525–549 (1999).
59. O'Connor, K. A. & Zusman, D. R. Development in *Myxococcus xanthus* involves differentiation into two cell types, peripheral rods and spores. *J. Bacteriol.* **173**, 3318–3333 (1991).
60. Hug, J. J. & Müller, R. Host Development for Heterologous Expression and Biosynthetic Studies of Myxobacterial Natural Products. *Comprehensive Natural Products III: Chemistry and Biology, Chapter 14818* **In press** (2020).
61. Dawid, W. Biology and global distribution of myxobacteria in soils. *FEMS Microbiol. Rev.* **24**, 403–427 (2000).
62. Burchard, R. P. & Dworkin, M. Light-Induced Lysis and Carotenogenesis in *Myxococcus xanthus*. *J. Bacteriol.* **91**, 535–545 (1966).
63. Meiser, P., Bode, H. B. & Müller, R. The unique DKxanthene secondary metabolite family from the myxobacterium *Myxococcus xanthus* is required for developmental sporulation. *Proc. Natl. Acad. Sci. U.S.A.* **103**, 19128–19133 (2006).
64. Xiao, Y., Wei, X. M., Ebright, R. & Wall, D. Antibiotic Production by Myxobacteria Plays a Role in Predation. *J. Bacteriol.* **193**, 4626–4633 (2011).
65. Furusawa, G., Dziewanowska, K., Stone, H., Settles, M. & Hartzell, P. Global analysis of phase variation in *Myxococcus xanthus*. *Mol. Microbiol.* **81**, 784–804 (2011).
66. Hoffmann, M. *et al.* Homospermidine Lipids: A compound class specifically formed during fruiting body formation of *Myxococcus xanthus* DK1622. *ACS Chem. Biol.* **13**, 273–280 (2018).

67. Herrmann, J., Fayad, A. A. & Müller, R. Natural products from myxobacteria: novel metabolites and bioactivities. *Nat. Prod. Rep.* **34**, 135–160 (2017).
68. Wenzel, S. C. & Müller, R. The biosynthetic potential of myxobacteria and their impact on drug discovery. *Curr. Opin. Drug Discov. Devel.* **12**, 220–230 (2009).
69. Weissman, K. J. & Müller, R. Myxobacterial secondary metabolites: bioactivities and modes-of-action. *Nat. Prod. Rep.* **27**, 1276–1295 (2010).
70. Landwehr, W., Wolf, C. & Wink, J. Actinobacteria and Myxobacteria—Two of the Most Important Bacterial Resources for Novel Antibiotics. *Curr. Top. Microbiol. Immunol.* (2016).
71. Sansinenea, E. & Ortiz, A. Secondary metabolites of soil *Bacillus* spp. *Biotechnol. Lett.* **33**, 1523–1538 (2011).
72. Schueffler, A. & Anke, T. Fungal natural products in research and development. *Nat. Prod. Rep.* **31**, 1425–1448 (2014).
73. Reichenbach, H. & Höfle, G. in *Scientific Annual Report* (Gesellschaft für Biotechnologische Forschung GmbH, Braunschweig, 1994), pp. 5–22.
74. Ringel, S. M. *et al.* Ambruticin (W7783), a new antifungal antibiotic. *J. Antibiot.* **30**, 371–375 (1977).
75. Sasse, F. *et al.* Argyrins, immunosuppressive cyclic peptides from myxobacteria. I. Production, isolation, physico-chemical and biological properties. *J. Antibiot.* **55**, 543–551 (2002).
76. Gerth, K., Steinmetz, H., Höfle, G. & Jansen, R. Chlorotonil A, a macrolide with a unique gem-dichloro-1,3-dione functionality from *Sorangium cellulosum*, So ce1525. *Angew. Chem. Int. Ed. Engl.* **47**, 600–602 (2008).
77. Baumann, S. *et al.* Cystobactamids: myxobacterial topoisomerase inhibitors exhibiting potent antibacterial activity. *Angew. Chem. Int. Ed.* **53**, 14605–14609 (2014).
78. Irschik, H., Gerth, K., Höfle, G., Kohl, W. & Reichenbach, H. The myxopyronins, new inhibitors of bacterial RNA synthesis from *Myxococcus fulvus* (Myxobacterales). *J. Antibiot.* **36**, 1651–1658 (1983).
79. Höfle, G. *et al.* Epothilone A and B—Novel 16-Membered macrolides with cytotoxic activity: Isolation, crystal structure, and conformation in solution. *Angew. Chem. Int. Ed. Engl.* **35**, 1567–1569 (1996).
80. Plaza, A. *et al.* Aetheramides A and B, potent HIV-inhibitory depsipeptides from a myxobacterium of the new genus "Aetherobacter". *Org. Lett.* **14**, 2854–2857 (2012).
81. Schäberle, T. F. *et al.* Coralopyronin A - A promising antibiotic for treatment of filariasis. *Int. J. Med. Microbiol.* **304**, 72–78 (2014).
82. Vetcher, L., Menzella, H. G., Kudo, T., Motoyama, T. & Katz, L. The antifungal polyketide ambruticin targets the HOG pathway. *Antimicrob. Agents Chemother.* **51**, 3734–3736 (2007).

83. Zusman, D. R., Scott, A. E., Yang, Z. & Kirby, J. R. Chemosensory pathways, motility and development in *Myxococcus xanthus*. *Nat. Rev. Microbiol.* **5**, 862–872 (2007).
84. Vahlensieck, H. F., Pridzun, L., Reichenbach, H. & Hinnen, A. Identification of the yeast *ACC1* gene product (acetyl-CoA carboxylase) as the target of the polyketide fungicide soraphen A. *Curr. Genet.* **25**, 95–100 (1994).
85. Mukhopadhyay, J. *et al.* The RNA polymerase "switch region" is a target of inhibitors. *Cell* **135**, 295–307 (2008).
86. Bax, B. D., Murshudov, G., Maxwell, A. & Germe, T. DNA Topoisomerase Inhibitors: Trapping a DNA-Cleaving Machine in Motion. *J. Mol. Biol.* **431**, 3427–3449 (2019).
87. Mulzer, J. (ed.). *The Epothilones, an Outstanding Family of Anti-Tumor Agents* (Springer, New York, 2009).
88. Nyfeler, B. *et al.* Identification of elongation factor G as the conserved cellular target of argyirin B. *PLoS ONE* **7**, e42657 (2012).
89. Bielecki, P. *et al.* Mutation in elongation factor G confers resistance to the antibiotic argyirin in the opportunistic pathogen *Pseudomonas aeruginosa*. *ChemBioChem* **13**, 2339–2345 (2012).
90. Nicleleit, I. *et al.* Argyrin A reveals a critical role for the tumor suppressor protein p27kip1 in mediating antitumor activities in response to proteasome inhibition. *Cancer Res.* **14**, 23–35 (2008).
91. Stauch, B. *et al.* Elucidation of the structure and intermolecular interactions of a reversible cyclic-peptide inhibitor of the proteasome by NMR spectroscopy and molecular modeling. *Angew. Chem. Int. Ed. Engl.* **49**, 3934–3938 (2010).
92. Chen, X. *et al.* Therapeutic effects of Argyrin F in pancreatic adenocarcinoma. *Cancer Lett.* **399**, 20–28 (2017).
93. Hoffmann, T. *et al.* Correlating chemical diversity with taxonomic distance for discovery of natural products in myxobacteria. *Nat. Commun.* **9**, 803 (2018).
94. Rodrigues, T., Reker, D., Schneider, P. & Schneider, G. Counting on natural products for drug design. *Nat. Chem.* **8**, 531–541 (2016).
95. Schneider, P. & Schneider, G. Privileged Structures Revisited. *Angew. Chem. Int. Ed. Engl.* **56**, 7971–7974 (2017).
96. Reymond, J.-L. & Awale, M. Exploring Chemical Space for Drug Discovery Using the Chemical Universe Database. *ACS Chem. Neurosci.* **3**, 649–657 (2012).
97. Evans, B. E. *et al.* Methods for drug discovery. Development of potent, selective, orally effective cholecystikinin antagonists. *J. Med. Chem.* **31**, 2235–2246 (1988).
98. Henkel, T., Brunne, R. M., Müller, H. & Reichel, F. Statistical investigation into the structural complementarity of natural products and synthetic compounds. *Angew. Chem. Int. Ed. Engl.* **38**, 643–647 (1999).

99. Veber, D. F. *et al.* Molecular properties that influence the oral bioavailability of drug candidates. *J. Med. Chem.* **45**, 2615–2623 (2002).
100. Koch, M. A. *et al.* Charting biologically relevant chemical space: A structural classification of natural products (SCONP). *Proc. Natl. Acad. Sci. USA* **102**, 17272–17277 (2005).
101. Facchini, P. J. Alkaloid biosynthesis in plants: Biochemistry, cell biology, molecular regulation, and metabolic engineering applications. *Annu. Rev. Plant Physiol. Plant Molec. Biol.* **52**, 29–66 (2001).
102. Osbourn, A. E. & Lanzotti, V. *Plant-derived natural products. Synthesis, function, and application* (Springer-Verlag New York, New York, NY, 2009).
103. Vogt, E. & Künzler, M. Discovery of novel fungal RiPP biosynthetic pathways and their application for the development of peptide therapeutics. *Appl. Microbiol. Biotechnol.* **103**, 5567–5581 (2019).
104. Keller, N. P., Turner, G. & Bennett, J. W. Fungal secondary metabolism - from biochemistry to genomics. *Nat. Rev. Microbiol.* **3**, 937–947 (2005).
105. Wang, H., Fewer, D. P., Holm, L., Rouhiainen, L. & Sivonen, K. Atlas of nonribosomal peptide and polyketide biosynthetic pathways reveals common occurrence of nonmodular enzymes. *Proc. Natl. Acad. Sci. USA* **111**, 9259–9264 (2014).
106. Dejong, C. A. *et al.* Polyketide and nonribosomal peptide retro-biosynthesis and global gene cluster matching. *Nat. Chem. Biol.* **12**, 1007–1014 (2016).
107. Kunze, B., Trowitzsch-Kienast, W., Höfle, G. & Reichenbach, H. Nannochelins A, B and C, new iron-chelating compounds from *Nannocystis exedens* (myxobacteria). Production, isolation, physico-chemical and biological properties. *J. Antibiot.* **45**, 147–150 (1992).
108. Ambrosi, H. D., Hartmann, V., Pistorius, D., Reissbrodt, R. & Trowitzsch-Kienast, W. Myxochelins B, C, D, E and F: A new structural principle for powerful siderophores imitating nature. *Eur. J. Org. Chem.*, 541–551 (1998).
109. Kunze, B., Bedorf, N., Kohl, W., Höfle, G. & Reichenbach, H. Myxochelin A, a new iron-chelating compound from *Angiococcus disciformis* (Myxobacterales). Production, isolation, physico-chemical and biological properties. *J. Antibiot.* **42**, 14–17 (1989).
110. Silakowski, B. *et al.* The myxochelin iron transport regulon of the myxobacterium *Stigmatella aurantiaca* Sg a15. *Eur. J. Biochem.* **267**, 6476–6485 (2000).
111. Reichenbach, H. Myxobacteria, producers of novel bioactive substances. *J. Ind. Microbiol. Biotechnol.* **27**, 149–156 (2001).
112. Weissman, K. J. & Müller, R. A brief tour of myxobacterial secondary metabolism. *Bioorg. Med. Chem.* **17**, 2121–2136 (2009).
113. Trowitsch, W., Witte, L. & Reichenbach, H. Geosmin from earthly smelling cultures of *Nannocystis exedens* (Myxobacterales). *FEMS Microbiol. Lett.* **12**, 257–260 (1981).

114. Dickschat, J. S., Bode, H. B., Wenzel, S. C., Müller, R. & Schulz, S. Biosynthesis and identification of volatiles released by the myxobacterium *Stigmatella aurantiaca*. *ChemBioChem* **6**, 2023–2033 (2005).
115. Dickschat, J. S., Wenzel, S. C., Bode, H. B., Müller, R. & Schulz, S. Biosynthesis of volatiles by the myxobacterium *Myxococcus xanthus*. *ChemBioChem* **5**, 778–787 (2004).
116. Kohl, W., Gloe, A. & Reichenbach, H. Steroids from the myxobacterium *Nannocystis exedens*. *J. Gen. Microbiol.* **129**, 1629–1635 (1983).
117. Bode, H. B. *et al.* Steroid biosynthesis in prokaryotes: identification of myxobacterial steroids and cloning of the first bacterial 2,3(*S*)-oxidosqualene cyclase from the myxobacterium *Stigmatella aurantiaca*. *Mol. Microbiol.* **47**, 471–481 (2003).
118. Hug, J. J., Krug, D. & Müller, R. Bacteria as genetically programmable producers of bioactive natural products. *Nat. Rev. Chem.* **4**, 172–193 (2020).
119. Gregory, K., Salvador, L. A., Akbar, S., Adaikpoh, B. I. & Stevens, D. C. Survey of Biosynthetic Gene Clusters from Sequenced Myxobacteria Reveals Unexplored Biosynthetic Potential. *Microorganisms* **7** (2019).
120. Huo, L. Synthetic biotechnology to study and engineer natural product biosynthesis in actinomycetes. Dissertation. Saarland University, 2014.
121. Stachelhaus, T., Schneider, A. & Marahiel, M. A. Engineered biosynthesis of peptide antibiotics. *Biochem. Pharmacol.* **52**, 177–186 (1996).
122. Stachelhaus, T., Huser, A. & Marahiel, M. A. Biochemical characterization of peptidyl carrier protein (PCP), the thiolation domain of multifunctional peptide synthetases. *Chem. Biol.* **3**, 913–921 (1996).
123. Donadio, S., Staver, M. J., McAlpine, J. B., Swanson, S. J. & Katz, L. Modular organization of genes required for complex polyketide biosynthesis. *Science* **252**, 675–679 (1991).
124. Walsh, C. T. Polyketide and nonribosomal peptide antibiotics: modularity and versatility. *Science* **303**, 1805–1810 (2004).
125. Dutta, S. *et al.* Structure of a modular polyketide synthase. *Nature* **510**, 512–517 (2014).
126. Horsman, M. E., Hari, T. P. A. & Boddy, C. N. Polyketide synthase and non-ribosomal peptide synthetase thioesterase selectivity: logic gate or a victim of fate? *Nat. Prod. Rep.* **33**, 183–202 (2016).
127. Wenzel, S. C. *et al.* On the biosynthetic origin of methoxymalonyl-acyl carrier protein, the substrate for incorporation of "glycolate" units into ansamitocin and soraphen A. *J. Am. Chem. Soc.* **128**, 14325–14336 (2006).
128. Dunn, B. J., Cane, E. D. & Khosla, C. Mechanism and specificity of an acyltransferase domain from a modular polyketide synthase. *Biochemistry* **52**, 1839–1841 (2013).

129. Marahiel, M. A. Protein templates for the biosynthesis of peptide antibiotics. *Chem. Biol.* **4**, 561–567 (1997).
130. Marahiel, M. A., Stachelhaus, T. & Mootz, H. D. Modular Peptide Synthetases Involved in Nonribosomal Peptide Synthesis. *Chem. Rev.* **97**, 2651–2674 (1997).
131. Caboche, S. *et al.* Norine: a database of nonribosomal peptides. *Nucleic Acids Res.* **36**, D326–D331 (2008).
132. Pupin, M. *et al.* Norine: A powerful resource for novel nonribosomal peptide discovery. *Synth. Syst. Biotechnol.* **1**, 89–94 (2016).
133. Walsh, C. T. *et al.* Tailoring enzymes that modify nonribosomal peptides during and after chain elongation on NRPS assembly lines. *Curr. Opin. Chem. Biol.* **5**, 525–534 (2001).
134. Krug, D. & Müller, R. Discovery of additional members of the tyrosine aminomutase enzyme family and the mutational analysis of CmdF. *ChemBioChem* **10**, 741–750 (2009).
135. Walsh, C. T., O'Brien, R. V. & Khosla, C. Nonproteinogenic amino acid building blocks for nonribosomal peptide and hybrid polyketide scaffolds. *Angew. Chem. Int. Ed. Engl.* **52**, 7098–7124 (2013).
136. Stachelhaus, T., Mootz, H. D., Bergendahl, V. & Marahiel, M. A. Peptide bond formation in nonribosomal peptide biosynthesis. Catalytic role of the condensation domain. *J. Biol. Chem.* **273**, 22773–22781 (1998).
137. Sieber, S. A. & Marahiel, M. A. Molecular mechanisms underlying nonribosomal peptide synthesis: approaches to new antibiotics. *Chem Rev* **105**, 715–738 (2005).
138. Schwarzer, D., Finking, R. & Marahiel, M. A. Nonribosomal peptides: from genes to products. *Nat. Prod. Rep.* **20**, 275–287 (2003).
139. Mootz, H. D., Schwarzer, D. & Marahiel, M. A. Ways of assembling complex natural products on modular nonribosomal peptide synthetases. *ChemBioChem* **3**, 490–504 (2002).
140. Trauger, J. W., Kohli, R. M., Mootz, H. D., Marahiel, M. A. & Walsh, C. T. Peptide cyclization catalysed by the thioesterase domain of tyrocidine synthetase. *Nature* **407**, 215–218 (2000).
141. Kohli, R. M., Trauger, J. W., Schwarzer, D., Marahiel, M. A. & Walsh, C. T. Generality of peptide cyclization catalyzed by isolated thioesterase domains of nonribosomal peptide synthetases. *Biochemistry* **40**, 7099–7108 (2001).
142. Mootz, H. D. & Marahiel, M. A. Biosynthetic systems for nonribosomal peptide antibiotic assembly. *Curr. Opin. Chem. Biol.* **1**, 543–551 (1997).
143. Stachelhaus, T., Mootz, H. D. & Marahiel, M. A. The specificity-conferring code of adenylation domains in nonribosomal peptide synthetases. *Chem. Biol.* **6**, 493–505 (1999).

144. Witte, S. N. R., Hug, J. J., Géraldy, M., Müller, R. & Kalesse, M. Biosynthesis and Total Synthesis of Pyrronazol B: a Secondary Metabolite from *Nannocystis pusilla*. *Chem. Eur. J.* **23**, 15917–15921 (2017).
145. Plaza, A., Viehrig, K., Garcia, R. & Müller, R. Jahnellamides,  $\alpha$ -keto- $\beta$ -methionine-containing peptides from the terrestrial myxobacterium *Jahnella* sp.: structure and biosynthesis. *Org. Lett.* **15**, 5882–5885 (2013).
146. Gorges, J. *et al.* Structure, Total Synthesis, and Biosynthesis of Chloromyxamides: Myxobacterial Tetrapeptides Featuring an Uncommon 6-Chloromethyl-5-methoxy-pipecolic Acid Building Block. *Angew. Chem. Int. Ed. Engl.* **57**, 14270–14275 (2018).
147. Miller, D. J. *et al.* Crystal complexes of a predicted *S*-adenosylmethionine-dependent methyltransferase reveal a typical AdoMet binding domain and a substrate recognition domain. *Protein Sci.* **12**, 1432–1442 (2003).
148. Ansari, M. Z., Sharma, J., Gokhale, R. S. & Mohanty, D. *In silico* analysis of methyltransferase domains involved in biosynthesis of secondary metabolites. *Bmc Bioinform.* **9** (2008).
149. Martin, J. SAM (dependent) I AM: the *S*-adenosylmethionine-dependent methyltransferase fold. *Curr. Opin. Struct. Biol.* **12**, 783–793 (2002).
150. Martin, J. L. & McMillan, F. M. SAM (dependent) I AM: the *S*-adenosylmethionine-dependent methyltransferase fold. *Curr. Opin. Struct. Biol.* **12**, 783–793 (2002).
151. Stachelhaus, T. & Walsh, C. T. Mutational Analysis of the Epimerization Domain in the Initiation Module PheATE of Gramicidin S Synthetase. *Biochemistry* **39**, 5775–5787 (2000).
152. Linne, U. & Marahiel, M. A. in *Methods in Enzymology : Protein Engineering* (Academic Press 2004), pp. 293–315.
153. Woithe, K. *et al.* Oxidative phenol coupling reactions catalyzed by OxyB: a cytochrome P450 from the vancomycin producing organism. Implications for vancomycin biosynthesis. *J. Am. Chem. Soc.* **129**, 6887–6895 (2007).
154. Losey, H. C. *et al.* Incorporation of glucose analogs by GtfE and GtfD from the vancomycin biosynthetic pathway to generate variant glycopeptides. *Chem. Biol.* **9**, 1305–1314 (2002).
155. Du, L. & Shen, B. Biosynthesis of hybrid peptide-polyketide natural products. *Curr. Opin. Drug Discov. Devel.* **4**, 215–228 (2001).
156. Du, L. H., Sanchez, C. & Shen, B. Hybrid peptide-polyketide natural products: biosynthesis and prospects toward engineering novel molecules. *Metab. Eng.* **3**, 78–95 (2001).
157. Du, L., Sanchez, C., Chen, M. T., Edwards, D. J. & Shen, B. The biosynthetic gene cluster for the antitumor drug bleomycin from *Streptomyces verticillus* ATCC15003 supporting functional interactions between nonribosomal peptide synthetases and a polyketide synthase. *Chem. Biol.* **7**, 623–642 (2000).

158. Crecy-Lagard, V. de *et al.* Pristinamycin I biosynthesis in *Streptomyces pristinaespiralis*: molecular characterization of the first two structural peptide synthetase genes. *J. Bacteriol.* **179**, 705–713 (1997).
159. Perlova, O. *et al.* Reconstitution of myxothiazol biosynthetic gene cluster by Red/ET recombination and heterologous expression in *Myxococcus xanthus*. *Appl. Environ. Microbiol.* **72**, 7485–7494 (2006).
160. Silakowski, B. *et al.* New lessons for combinatorial biosynthesis from myxobacteria. The myxothiazol biosynthetic gene cluster of *Stigmatella aurantiaca* DW4/3-1. *J. Biol. Chem.* **274**, 37391–37399 (1999).
161. Müller, I. *et al.* A unique mechanism for methyl ester formation via an amide intermediate found in myxobacteria. *ChemBioChem* **7**, 1197–1205 (2006).
162. Dowling, D. P. *et al.* Structural elements of an NRPS cyclization domain and its intermodule docking domain. *Proc. Natl. Acad. Sci. U.S.A.* **113**, 12432–12437 (2016).
163. Richter, C. D., Nietlispach, D., Broadhurst, R. W. & Weissman, K. J. Multienzyme docking in hybrid megasynthetases. *Nat. Chem. Biol.* **4**, 75–81 (2007).
164. Li, Y., Weissman, K. J. & Müller, R. Insights into multienzyme docking in hybrid PKS-NRPS megasynthetases revealed by heterologous expression and genetic engineering. *ChemBioChem* **11**, 1069–1075 (2010).
165. Broadhurst, R. W., Nietlispach, D., Wheatcroft, M. P., Leadlay, P. F. & Weissman, K. J. The structure of docking domains in modular polyketide synthases. *Chem. Biol.* **10**, 723–731 (2003).
166. Hahn, M. & Stachelhaus, T. Selective interaction between nonribosomal peptide synthetases is facilitated by short communication-mediating domains. *Proc. Natl. Acad. Sci. U.S.A.* **101**, 15585–15590 (2004).
167. Miyanaga, A., Kudo, F. & Eguchi, T. Protein–protein interactions in polyketide synthase–nonribosomal peptide synthetase hybrid assembly lines. *Nat. Prod. Rep.* (2018).
168. Chen, H., O'Connor, S., Cane, D. E. & Walsh, C. T. Epothilone biosynthesis: assembly of the methylthiazolylcarboxy starter unit on the EpoB subunit. *Chem. Biol.* **8**, 899–912 (2001).
169. Molnár, I. *et al.* The biosynthetic gene cluster for the microtubule-stabilizing agents epothilones A and B from *Sorangium cellulosum* So ce90. *Chem. Biol.* **7**, 97–109 (2000).
170. Tang, L. *et al.* Cloning and heterologous expression of the epothilone gene cluster. *Science* **287**, 640–642 (2000).
171. O'Connor, S. E., Walsh, C. T. & Liu, F. Biosynthesis of epothilone intermediates with alternate starter units: engineering polyketide-nonribosomal interfaces. *Angew. Chem. Int. Ed. Engl.* **42**, 3917–3921 (2003).



172. Liu, F., Garneau, S. & Walsh, C. T. Hybrid nonribosomal peptide-polyketide interfaces in epothilone biosynthesis: Minimal requirements at N and C termini of EpoB for elongation. *Chem. Biol.* **11**, 1533–1542 (2004).
173. O'Connor, S. E., Chen, H. W. & Walsh, C. T. Enzymatic assembly of epothilones: The EpoC subunit and reconstitution of the EpoA-ACP/B/C polyketide and nonribosomal peptide interfaces. *Biochemistry* **41**, 5685–5694 (2002).
174. Miller, D. A., Luo, L. S., Hillson, N., Keating, T. A. & Walsh, C. T. Yersiniabactin synthetase: A four-protein assembly line producing the nonribosomal peptide/polyketide hybrid siderophore of *Yersinia pestis*. *Chem. Biol.* **9**, 333–344 (2002).
175. Weissman, K. J. & Müller, R. Protein-protein interactions in multienzyme megasynthetases. *ChemBioChem* **9**, 826–848 (2008).
176. Whicher, J. R. *et al.* Cyanobacterial polyketide synthase docking domains: A tool for engineering natural product biosynthesis. *Chem. Biol.* **20**, 1340–1351 (2013).
177. Hillson, N. J. & Walsh, C. T. Dimeric structure of the six-domain VibF subunit of vibriobactin synthetase: Mutant domain activity regain and ultracentrifugation studies. *Biochemistry* **42**, 766–775 (2003).
178. Hillson, N. J., Balibar, C. J. & Walsh, C. T. Catalytically Inactive Condensation Domain C1 is Responsible for the Dimerization of the VibF Subunit of Vibriobactin Synthetase. *Biochemistry* (2004).
179. Tang, Y., Kim, C. Y., Mathews, I. I., Cane, D. E. & Khosla, C. The 2.7-Angstrom crystal structure of a 194-kDa homodimeric fragment of the 6-deoxyerythronolide B synthase. *Proc. Natl. Acad. Sci. USA* **103**, 11124–11129 (2006).
180. Zheng, J., Fage, C. D., Demeler, B., Hoffman, D. W. & Keatinge-Clay, A. T. The missing linker: a dimerization motif located within polyketide synthase modules. *ACS Chem. Biol.* **8**, 1263–1270 (2013).
181. Oßwald, C. *et al.* A highly unusual polyketide synthase directs dawenol polyene biosynthesis in *Stigmatella aurantiaca*. *J. Biotechnol.* **191**, 54–63 (2014).
182. Das, A. & Khosla, C. Biosynthesis of aromatic polyketides in bacteria. *Acc. Chem. Res.* **42**, 631–639 (2009).
183. Zhou, H., Li, Y. & Tang, Y. Cyclization of aromatic polyketides from bacteria and fungi. *Nat. Prod. Rep.* **27**, 839–868 (2010).
184. Hertweck, C. The Biosynthetic Logic of Polyketide Diversity. *Angew. Chem. Int. Ed. Engl.* **48**, 4688–4716 (2009).
185. Hertweck, C., Luzhetskyy, A., Rebets, Y. & Bechthold, A. Type II polyketide synthases: gaining a deeper insight into enzymatic teamwork. *Nat. Prod. Rep.* **24**, 162–190 (2007).

186. Leeper, F. J. & Vederas, J. C. (eds.). *Biosynthesis. Aromatic Polyketides, Isoprenoids, Alkaloids* (Springer-Verlag Berlin Heidelberg; Springer e-books, Berlin, Heidelberg, 2000).
187. Rawlings, B. J. Biosynthesis of polyketides (other than actinomycete macrolides). *Nat. Prod. Rep.* **16**, 425–484 (1999).
188. Zhang, Z., Pan, H.-X. & Tang, G.-L. New insights into bacterial type II polyketide biosynthesis. *F1000Research* **6** (2017).
189. Lukežič, T. *et al.* Identification of the chelocardin biosynthetic gene cluster from *Amycolatopsis sulphurea*: a platform for producing novel tetracycline antibiotics. *Microbiology* **159**, 2524–2532 (2013).
190. Stepanek, J. J., Lukežič, T., Teichert, I., Petković, H. & Bandow, J. E. Dual mechanism of action of the atypical tetracycline chelocardin. *Biochim. Biophys. Acta* **1864**, 645–654 (2016).
191. Brachmann, A. O. *et al.* A type II polyketide synthase is responsible for anthraquinone biosynthesis in *Photorhabdus luminescens*. *ChemBioChem* **8**, 1721–1728 (2007).
192. Sandmann, A. *et al.* A type II polyketide synthase from the gram-negative bacterium *Stigmatella aurantiaca* is involved in aurachin alkaloid biosynthesis. *Angew. Chem. Int. Ed. Engl.* **46**, 2712–2716 (2007).
193. Panter, F., Krug, D., Baumann, S. & Müller, R. Self-resistance guided genome mining uncovers new topoisomerase inhibitors from myxobacteria. *Chem. Sci.* **9**, 4898–4908 (2018).
194. Parvez, A. *et al.* Novel Type III Polyketide Synthases Biosynthesize Methylated Polyketides in *Mycobacterium marinum*. *Sci. Rep.* **8**, 6529 (2018).
195. Reiter, S., Cahn, J. K. B., Wiebach, V., Ueoka, R. & Piel, J. Characterization of an Orphan Type III Polyketide Synthase Conserved in Uncultivated "Entotheonella" Sponge Symbionts. *ChemBioChem* (2019).
196. Lim, Y. P., Go, M. K. & Yew, W. S. Exploiting the Biosynthetic Potential of Type III Polyketide Synthases. *Molecules (Basel, Switzerland)* **21** (2016).
197. Hayashi, T., Kitamura, Y., Funa, N., Ohnishi, Y. & Horinouchi, S. Fatty acyl-AMP ligase involvement in the production of alkylresorcylic acid by a *Myxococcus xanthus* type III polyketide synthase. *ChemBioChem* **12**, 2166–2176 (2011).
198. Shimizu, Y., Ogata, H. & Goto, S. Type III Polyketide Synthases. Functional Classification and Phylogenomics. *ChemBioChem* **18**, 50–65 (2017).
199. Yu, D., Xu, F., Zeng, J. & Zhan, J. Type III polyketide synthases in natural product biosynthesis. *IUBMB Life* **64**, 285–295 (2012).
200. Gross, F. *et al.* Bacterial type III polyketide synthases: Phylogenetic analysis and potential for the production of novel secondary metabolites by heterologous expression in pseudomonads. *Arch. Microbiol.* **185**, 28–38 (2006).

201. Sone, Y. *et al.* Identification and characterization of bacterial enzymes catalyzing the synthesis of 1,8-dihydroxynaphthalene, a key precursor of dihydroxynaphthalene melanin, from *Sorangium cellulosum*. *Appl. Environ. Microbiol.*, AEM.00258-18 (2018).
202. Hug, J. J., Panter, F., Krug, D. & Müller, R. Genome mining reveals uncommon alkylpyrones as type III PKS products from myxobacteria. *J. Ind. Microbiol. Biotechnol.* **46**, 319–334 (2019).
203. Arnison, P. G. *et al.* Ribosomally synthesized and post-translationally modified peptide natural products: overview and recommendations for a universal nomenclature. *Nat. Prod. Rep.* **30**, 108–160 (2013).
204. Yang, X. & van der Donk, W. A. Ribosomally synthesized and post-translationally modified peptide natural products: new insights into the role of leader and core peptides during biosynthesis. *Chem. Eur. J.* **19**, 7662–7677 (2013).
205. Hudson, G. A. & Mitchell, D. A. RiPP antibiotics: biosynthesis and engineering potential. *Curr. Opin. Microbiol.* **45**, 61–69 (2018).
206. Kupke, T., Kempter, C., Jung, G. & Gotz, F. Oxidative decarboxylation of peptides catalyzed by flavoprotein EpiD. Determination of substrate specificity using peptide libraries and neutral loss mass spectrometry. *J. Biol. Chem.* **270**, 11282–11289 (1995).
207. Shi, Y. X., Yang, X. A., Garg, N. & van der Donk, W. A. Production of lantipeptides in *Escherichia coli*. *J. Am. Chem. Soc.* **133**, 2338–2341 (2011).
208. Boakes, S., Cortés, J., Appleyard, A. N., Rudd, B. A. M. & Dawson, M. J. Organization of the genes encoding the biosynthesis of actagardine and engineering of a variant generation system. *Mol. Microbiol.* **72**, 1126–1136 (2009).
209. McIntosh, J. A., Donia, M. S., Nair, S. K. & Schmidt, E. W. Enzymatic basis of ribosomal peptide prenylation in cyanobacteria. *J. Am. Chem. Soc.* **133**, 13698–13705 (2011).
210. Truman, A. W. Cyclisation mechanisms in the biosynthesis of ribosomally synthesised and post-translationally modified peptides. *Beilstein J. Org. Chem.* **12**, 1250–1268 (2016).
211. Melby, J. O., Nard, N. J. & Mitchell, D. A. Thiazole/oxazole-modified microcins: complex natural products from ribosomal templates. *Curr. Opin. Chem. Biol.* **15**, 369–378 (2011).
212. Flühe, L. *et al.* The radical SAM enzyme AlbA catalyzes thioether bond formation in subtilisin A. *Nat. Chem. Biol.* **8**, 350–357 (2012).
213. Franz, L., Adam, S., Santos-Aberturas, J., Truman, A. W. & Koehnke, J. Macroamidine Formation in Bottromycins Is Catalyzed by a Divergent YcaO Enzyme. *J. Am. Chem. Soc.* **139**, 18158–18161 (2017).
214. Grell, T. A. J., Goldman, P. J. & Drennan, C. L. SPASM and Twitch Domains in S-Adenosylmethionine (SAM) Radical Enzymes. *J. Biol. Chem.* **290**, 3964–3971 (2015).

215. Ortega, M. A. & van der Donk, Wilfred A. New Insights into the Biosynthetic Logic of Ribosomally Synthesized and Post-translationally Modified Peptide Natural Products. *Cell Chem. Biol.* **23**, 31–44 (2016).
216. Inoue, M. *et al.* Total synthesis of the large non-ribosomal peptide polytheonamide B. *Nat. Chem.* **2**, 280–285 (2010).
217. Shipley, P. R. The biosynthesis of the thiopeptide antibiotic thiostrepton. Dissertation. University of Washington, 1999.
218. Kelly, W. L., Pan, L. & Li, C. Thiostrepton Biosynthesis: Prototype for a New Family of Bacteriocins. *J. Am. Chem. Soc.* (2009).
219. Freeman, M. F. *et al.* Metagenome mining reveals polytheonamides as posttranslationally modified ribosomal peptides. *Science* **338**, 387–390 (2012).
220. Zheng, Q., Fang, H. & Liu, W. Post-translational modifications involved in the biosynthesis of thiopeptide antibiotics. *Org. Biomol. Chem.* **15**, 3376–3390 (2017).
221. Morinaka, B. I. *et al.* Radical *S*-adenosyl methionine epimerases: regioselective introduction of diverse D-amino acid patterns into peptide natural products. *Angew. Chem. Int. Ed. Engl.* **53** (2014).
222. Kalyon, B. *et al.* Plantazolicin A and B. Structure elucidation of ribosomally synthesized thiazole/oxazole peptides from *Bacillus amyloliquefaciens* FZB42. *Org. Lett.* **13**, 2996–2999 (2011).
223. Minami, Y. *et al.* Structure of cypemycin, a new peptide antibiotic. *Tetrahedron Lett.* **35**, 8001–8004 (1994).
224. Repka, L. M., Chekan, J. R., Nair, S. K. & van der Donk, Wilfred A. Mechanistic Understanding of Lanthipeptide Biosynthetic Enzymes. *Chem. Rev.* **117**, 5457–5520 (2017).
225. Tietz, J. I. *et al.* A new genome-mining tool redefines the lasso peptide biosynthetic landscape. *Nat. Chem. Biol.* **13**, 470–478 (2017).
226. Merwin, N. J. *et al.* DeepRiPP integrates multiomics data to automate discovery of novel ribosomally synthesized natural products. *Proc. Natl. Acad. Sci. USA* **117**, 371–380 (2020).
227. Viehrig, K. *et al.* Structure and biosynthesis of crocagins: polycyclic postranslationally modified ribosomal peptides from *Chondromyces crocatus*. *Angew. Chem.*, 1–5 (2017).
228. Revermann, O. Novel secondary metabolites from myxobacteria and their biosynthetic machinery. Dissertation. Saarland University, 2012.
229. Ting, C. P. *et al.* Use of a scaffold peptide in the biosynthesis of amino acid–derived natural products. *Science* **365**, 280–284 (2019).
230. Zhang, Z. & van der Donk, Wilfred A. Nonribosomal Peptide Extension by a Peptide Amino-Acyl tRNA Ligase. *J. Am. Chem. Soc.* **141**, 19625–19633 (2019).

231. Burkhart, B. J., Hudson, G. A., Dunbar, K. L. & Mitchell, D. A. A prevalent peptide-binding domain guides ribosomal natural product biosynthesis. *Nat. Chem. Biol.* **11**, 564–570 (2015).
232. Helfrich, E. J. N., Lin, G.-M., Voigt, C. A. & Clardy, J. Bacterial terpene biosynthesis: challenges and opportunities for pathway engineering. *Beilstein J. Org. Chem.* **15**, 2889–2906 (2019).
233. Nič, M., Jirát, J., Košata, B., Jenkins, A. & McNaught, A. *IUPAC Compendium of Chemical Terminology* (IUPAC, Research Triangle Park, NC, 2009).
234. Buhaescu, I. & Izzedine, H. Mevalonate pathway: A review of clinical and therapeutical implications. *Clin. Biochem.* **40**, 575–584 (2007).
235. Rohmer, M. The discovery of a mevalonate-independent pathway for isoprenoid biosynthesis in bacteria, algae and higher plants†. *Nat. Prod. Rep.* **16**, 565–574 (1999).
236. Mahmud, T. *et al.* A novel biosynthetic pathway to isovaleryl-CoA in myxobacteria: The involvement of the mevalonate pathway. *ChemBioChem* **6**, 322–330 (2005).
237. Bode, H. B. *et al.* Identification of additional players in the alternative biosynthesis pathway to isovaleryl-CoA in the myxobacterium *Myxococcus xanthus*. *ChemBioChem* **10**, 128–140 (2009).
238. Lorenzen, W., Ring, M. W., Schwär, G. & Bode, H. B. Isoprenoids are essential for fruiting body formation in *Myxococcus xanthus*. *J. Bacteriol.* **191**, 5849–5853 (2009).
239. Holstein, S. A. & Hohl, R. J. Isoprenoids: remarkable diversity of form and function. *Lipids* **39**, 293–309 (2004).
240. Iguchi, E., Okuhara, M., Kohsaka, M., Aoki, H. & Imanaka, H. Studies on new phosphonic acid antibiotics. *J. Antibiot.* **33**, 18–23 (1980).
241. Jomaa, H. *et al.* Inhibitors of the nonmevalonate pathway of isoprenoid biosynthesis as antimalarial drugs. *Science* **285**, 1573–1576 (1999).
242. Driller, R. *et al.* Towards a comprehensive understanding of the structural dynamics of a bacterial diterpene synthase during catalysis. *Nat. Commun.* **9**, 1–8.
243. Dickschat, J. S. Bacterial terpene cyclases. *Nat. Prod. Rep.* **33**, 87–110 (2015).
244. Baunach, M., Franke, J. & Hertweck, C. Terpenoid Biosynthesis Off the Beaten Track: Unconventional Cyclases and Their Impact on Biomimetic Synthesis. *Angew. Chem. Int. Ed.* **54**, 2604–2626 (2015).
245. Baer, P. *et al.* Induced-Fit Mechanism in Class I Terpene Cyclases. *Angew. Chem. Int. Ed. Engl.* **53**, 7652–7656 (2014).
246. Davis, E. M. & Croteau, R. in *Biosynthesis*, edited by F. J. Leeper & J. C. Vederas (Springer-Verlag Berlin Heidelberg; Springer e-books, Berlin, Heidelberg, 2000), pp. 53–95.

247. Dickschat, J. S. Bacterial Diterpene Biosynthesis. *Angew. Chem. Int. Ed.* **58**, 15964–15976 (2019).
248. Smanski, M. J., Peterson, R. M., Huang, S.-X. & Shen, B. Bacterial diterpene synthases: new opportunities for mechanistic enzymology and engineered biosynthesis. *Curr. Opin. Chem. Biol.* **16**, 132–141 (2012).
249. Pronin, S. V. & Shenvi, R. A. Synthesis of highly strained terpenes by non-stop tail-to-head polycyclization. *Nat. Chem.* **4**, 915–920 (2012).
250. Xu, Z., Baunach, M., Ding, L. & Hertweck, C. Bacterial synthesis of diverse indole terpene alkaloids by an unparalleled cyclization sequence. *Angew. Chem. Int. Ed.* **51**, 10293–10297 (2012).
251. Chooi, Y. H., Hong, Y. J., Cacho, R. A., Tantillo, D. J. & Tang, Y. A cytochrome P450 serves as an unexpected terpene cyclase during fungal meroterpenoid biosynthesis. *J. Am. Chem. Soc.* **135**, 16805–16808 (2013).
252. Taura, F., Morimoto, S., Shoyama, Y. & Mechoulam, R. First direct evidence for the mechanism of  $\Delta^1$ -tetrahydrocannabinolic acid biosynthesis. *J. Am. Chem. Soc.* **117**, 9766–9767 (1995).
253. Lichman, B. R. *et al.* Uncoupled activation and cyclization in catmint reductive terpenoid biosynthesis. *Nat. Chem. Biol.* **15**, 71–79 (2019).
254. Yee, D. A. *et al.* Genome Mining of Alkaloidal Terpenoids from a Hybrid Terpene and Nonribosomal Peptide Biosynthetic Pathway. *J. Am. Chem. Soc.* **142**, 710–714 (2020).
255. Majik, M. S., Gawas, U. B. & Mandrekar, V. K. in *Advances in biological science research*, edited by S. N. Meena & M. M. Naik (Academic Press, an imprint of Elsevier, London, 2019), pp. 395–409.
256. Seco, J. M., Quinoa, E. & Riguera, R. The assignment of absolute configuration by NMR. *Chem. Rev.* **104**, 17–117 (2004).
257. Garcia, R., Krug, D. & Müller, R. in *Methods in Enzymology*. Part A: Overview Articles and Peptides, edited by David A. Hopwood (2009). Part A: Overview Articles and Peptides, pp. 59–91.
258. Schroeder, F. C. & Gronquist, M. Extending the scope of NMR spectroscopy with microcoil probes. *Angew. Chem. Int. Ed. Engl.* **45**, 7122–7131 (2006).
259. Keller, L. *et al.* Macyranonones: Structure, Biosynthesis, and Binding Mode of an Unprecedented Epoxyketone that Targets the 20S Proteasome. *J. Am. Chem. Soc.* **137**, 8121–8130 (2015).
260. Nadmid, S. *et al.* Hyalachelins A-C, unusual siderophores isolated from the terrestrial myxobacterium *Hyalangium minutum*. *Org. Lett.* **16**, 4130–4133 (2014).
261. Etzbach, L., Plaza, A., Garcia, R., Baumann, S. & Müller, R. Cystomanamides: structure and biosynthetic pathway of a family of glycosylated lipopeptides from myxobacteria. *Org. Lett.* **16**, 2414–2417 (2014).

262. Tokunaga, T., Akagi, K.-I. & Okamoto, M. Sensitivity enhancement by chromatographic peak concentration with ultra-high performance liquid chromatography-nuclear magnetic resonance spectroscopy for minor impurity analysis. *J. Chromatogr. A* **1508**, 163–168 (2017).
263. Schlotterbeck, G. & Ceccarelli, S. M. LC–SPE–NMR–MS: a total analysis system for bioanalysis. *Bioanalysis* **1**, 549–559 (2009).
264. Dickschat, J. S. Capturing volatile natural products by mass spectrometry. *Nat. Prod. Rep.* **31**, 838–861 (2014).
265. Schulz, S., Fuhlendorff, J. & Reichenbach, H. Identification and synthesis of volatiles released by the myxobacterium *Chondromyces crocatus*. *Tetrahedron* **60**, 3863–3872 (2004).
266. Perez, E. R. *et al.* Comparison of LC–MS–MS and GC–MS Analysis of Benzodiazepine Compounds Included in the Drug Demand Reduction Urinalysis Program. *J. Anal. Toxicol.* **40**, 201–207 (2016).
267. Hoffmann, T., Krug, D., Hüttel, S. & Müller, R. Improving natural products identification through targeted LC–MS/MS in an untargeted secondary metabolomics workflow. *Anal. Chem.* **86**, 10780–10788 (2014).
268. Krug, D. *et al.* Discovering the Hidden Secondary Metabolome of *Myxococcus xanthus*: a Study of Intraspecific Diversity. *Appl. Environ. Microbiol.* **74**, 3058–3068 (2008).
269. Kuhnert, N. *et al.* Scope and limitations of principal component analysis of high resolution LC–TOF–MS data: the analysis of the chlorogenic acid fraction in green coffee beans as a case study. *Anal. Methods* **3**, 144–155 (2011).
270. Abdi, H. & Williams, L. J. Principal component analysis. *WIREs Computational Statistics* **2**, 433–459 (2010).
271. Cortina, N. S., Krug, D., Plaza, A., Revermann, O. & Müller, R. Myxoprincomide: a natural product from *Myxococcus xanthus* discovered by comprehensive analysis of the secondary metabolome. *Angew. Chem. Int. Ed. Engl.* **51**, 811–816 (2012).
272. Dictionary of Natural Products.
273. Laatsch, H. *AntiBase 2014: The natural compound identifier* (Wiley-VCH; Weinheim, 2014).
274. Blunt, J. W., Copp, B. R., Keyzers, R. A., Munro, Murray H. G. & Prinsep, M. R. Marine natural products. *Nat. Prod. Rep.* **34**, 235–294 (2017).
275. Horai, H. *et al.* MassBank: a public repository for sharing mass spectral data for life sciences. *J. Mass Spectrom.* **45**, 703–714 (2010).
276. Smith, C. A. *et al.* METLIN: a metabolite mass spectral database. *Ther. Drug Monit.* **27**, 747–751 (2005).
277. mzCloud. Advanced Mass Spectral Database.

278. Sawada, Y. *et al.* RIKEN tandem mass spectral database (ReSpect) for phytochemicals. A plant-specific MS/MS-based data resource and database. *Phytochemistry* **82**, 38–45 (2012).
279. Wang, M. *et al.* Sharing and community curation of mass spectrometry data with Global Natural Products Social Molecular Networking. *Nat. Biotechnol.* **34**, 828–837 (2016).
280. Crüsemann, M. *et al.* Prioritizing Natural Product Diversity in a Collection of 146 Bacterial Strains Based on Growth and Extraction Protocols. *J. Nat. Prod.* (2016).
281. Johnston, C. W. *et al.* An automated Genomes-to-Natural Products platform (GNP) for the discovery of modular natural products. *Nat. Commun.* **6**, 8421 (2015).
282. Gould, K. Antibiotics: from prehistory to the present day. *J. Antimicrob. Chemother.* **71**, 572–575 (2016).
283. Nicolaou, K. C. & Snyder, S. A. Chasing molecules that were never there: misassigned natural products and the role of chemical synthesis in modern structure elucidation. *Angew. Chem. Int. Ed.* **44**, 1012–1044 (2005).
284. Reynolds, W. F. & Mazzola, E. P. in *Progress in the chemistry of organic natural products. 100*, edited by U. Wagner (Springer, Cham, 2015), pp. 223–309.
285. Kind, T. & Fiehn, O. Advances in structure elucidation of small molecules using mass spectrometry. *Bioanal Rev.* **2**, 23–60 (2010).
286. Etzbach, L. Exploring the biosynthetic potential of *Cystobacter fuscus* - characterization of new structures and studies on their biosynthesis. Dissertation. Saarland University, 2015.
287. Webb, A. Increasing the sensitivity of magnetic resonance spectroscopy and imaging. *Anal. Chem.* **84**, 9–16 (2012).
288. Emwas, A.-H. M. The strengths and weaknesses of NMR spectroscopy and mass spectrometry with particular focus on metabolomics research. *Methods in molecular biology (Clifton, N.J.)* **1277**, 161–193 (2015).
289. Özkaya, F. C. *et al.* Isolation and X-ray structure analysis of citreohydrinol from marine-derived *Penicillium atrovenerum*. *Nat. Prod. Res.* **32**, 840–843 (2018).
290. Jones, C. G. *et al.* The CryoEM Method MicroED as a Powerful Tool for Small Molecule Structure Determination. *ACS Cent. Sci.* **4**, 1587–1592 (2018).
291. Medema, M. H. *et al.* Minimum information about a biosynthetic gene cluster. *Nat. Chem. Biol.* **11**, 625–631 (2015).
292. Rocha, E. P. C. The organization of the bacterial genome. *Annu. Rev. Genet.* **42**, 211–233 (2008).
293. Cravens, A., Payne, J. & Smolke, C. D. Synthetic biology strategies for microbial biosynthesis of plant natural products. *Nat. Commun.* **10**, 2142 (2019).



294. Ziemert, N., Alanjary, M. & Weber, T. The evolution of genome mining in microbes - a review. *Nat. Prod. Rep.* **33**, 988–1005 (2016).
295. Baltz, R. H. Natural product drug discovery in the genomic era: realities, conjectures, misconceptions, and opportunities. *J. Ind. Microbiol.* **46**, 281–299 (2018).
296. Staden, R. A strategy of DNA sequencing employing computer programs. *Nucleic Acids Res.* **6**, 2601–2610 (1979).
297. Sanger, F., Nicklen, S. & Coulson, A. R. DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. USA* **74**, 5463–5467 (1977).
298. Sanger, F. *et al.* Nucleotide sequence of bacteriophage  $\phi$  X174 DNA. *Nature* **265**, 687–695 (1977).
299. Shendure, J. & Ji, H. Next-generation DNA sequencing. *Nat. Biotechnol.* **26**, 1135–1145 (2008).
300. Bentley, D. R. *et al.* Accurate whole human genome sequencing using reversible terminator chemistry. *Nature* **456**, 53–59 (2008).
301. Rusk, N. Torrents of sequence. *Nat. Methods* **8**, 44 (2011).
302. Levene, M. J. *et al.* Zero-mode waveguides for single-molecule analysis at high concentrations. *Science (New York, N.Y.)* **299**, 682–686 (2003).
303. Lee, N. *et al.* Thirty complete *Streptomyces* genome sequences for mining novel secondary metabolite biosynthetic gene clusters. *Sci. Data* **7**, 55 (2020).
304. Hwang, S. *et al.* Primary transcriptome and translome analysis determines transcriptional and translational regulatory elements encoded in the *Streptomyces clavuligerus* genome. *Nucleic Acids Res.* **47**, 6114–6129 (2019).
305. Röttig, M. *et al.* NRPSpredictor2—a web server for predicting NRPS adenylation domain specificity. *Nucleic Acids Res.* **39**, W362–W367 (2011).
306. Prieto, C., García-Estrada, C., Lorenzana, D. & Martín, J. F. NRPSsp: non-ribosomal peptide synthase substrate predictor. *Bioinformatics (Oxford, England)* **28**, 426–427 (2012).
307. Conti, E., Stachelhaus, T., Marahiel, M. A. & Brick, P. Structural basis for the activation of phenylalanine in the non-ribosomal biosynthesis of gramicidin S. *EMBO J.* **16**, 4174–4183 (1997).
308. Döhren, H. von, Dieckmann, R. & Pavela-Vrancic, M. The nonribosomal code. *Chem. Biol.* **6**, R273–R279 (1999).
309. Rausch, C., Weber, T., Kohlbacher, O., Wohlleben, W. & Huson, D. H. Specificity prediction of adenylation domains in nonribosomal peptide synthetases (NRPS) using transductive support vector machines (TSVMs). *Nucleic Acids Res.* **33**, 5799–5808 (2005).

310. Blin, K. *et al.* antiSMASH 5.0: updates to the secondary metabolite genome mining pipeline. *Nucleic Acids Res.*, W81-W87 (2019).
311. Helfrich, E. J. N. *et al.* Automated structure prediction of *trans*-acyltransferase polyketide synthase products. *Nat. Chem. Biol.* **15**, 813–821 (2019).
312. Khater, S., Anand, S. & Mohanty, D. *In silico* methods for linking genes and secondary metabolites. The way forward. *Synth. Syst. Biotechnol.* **1**, 80–88 (2016).
313. Knappe, T. A., Linne, U., Robbel, L. & Marahiel, M. A. Insights into the biosynthesis and stability of the lasso peptide capistruin. *Chem. Biol.* **16**, 1290–1298 (2009).
314. Agrawal, P., Khater, S., Gupta, M., Sain, N. & Mohanty, D. RiPPMiner: a bioinformatics resource for deciphering chemical structures of RiPPs based on prediction of cleavage and cross-links. *Nucleic Acids Res.* **45**, W80-W88 (2017).
315. Soldatou, S., Eldjarn, G. H., Huerta-Urbe, A., Rogers, S. & Duncan, K. R. Linking biosynthetic and chemical space to accelerate microbial secondary metabolite discovery. *FEMS Microbiol. Lett.* **366** (2019).
316. Cimermancic, P. *et al.* Insights into secondary metabolism from a global analysis of prokaryotic biosynthetic gene clusters. *Cell* **158**, 412–421 (2014).
317. Hussain, M. S. *et al.* Current approaches toward production of secondary plant metabolites. *J. Pharm. Bioallied Sci.* **4**, 10–20 (2012).
318. Ochoa-Villarreal, M. *et al.* Plant cell culture strategies for the production of natural products. *BMB Rep.* **49**, 149–158.
319. Smetanska, I. in *Food biotechnology*, edited by U. Stahl, U. E. B. Donalies, E. Nevoigt & D. B. Archer (Springer, Berlin, 2008), pp. 187–228.
320. Mustafa, N. R., Winter, W. d., van Iren, F. & Verpoorte, R. Initiation, growth and cryopreservation of plant cell suspension cultures. *Nat. Protoc.* **6**, 715 (2011).
321. Nakagawa, A. *et al.* Total biosynthesis of opiates by stepwise fermentation using engineered *Escherichia coli*. *Nat. Commun.* **7**, 10390 (2016).
322. Ajikumar, P. K. *et al.* Isoprenoid pathway optimization for Taxol precursor overproduction in *Escherichia coli*. *Science* **330**, 70–74 (2010).
323. Biggs, B. W. *et al.* Overcoming heterologous protein interdependency to optimize P450-mediated Taxol precursor synthesis in *Escherichia coli*. *Proc. Natl. Acad. Sci. USA* **113**, 3209–3214 (2016).
324. Huo, L. *et al.* Heterologous expression of bacterial natural product biosynthetic pathways. *Nat. Prod. Rep.* **36**, 1412-1436 (2019).
325. Molnar, I. *et al.* The biosynthetic gene cluster for the microtubule-stabilizing agents epothilones A and B from *Sorangium cellulosum* So ce90. *Chem. Biol.* **7**, 97–109 (2000).

326. Beyer, S., Kunze, B., Silakowski, B. & Müller, R. Metabolic diversity in myxobacteria: identification of the myxalamid and the stigmatellin biosynthetic gene cluster of *Stigmatella aurantiaca* Sg a15 and a combined polyketide-(poly)peptide gene cluster from the epothilone producing strain *Sorangium cellulosum* So ce90. *Biochim. Biophys. Acta* **1445**, 185–195 (1999).
327. Stevens, D. C., Henry, M. R., Murphy, K. A. & Boddy, C. N. Heterologous expression of the oxytetracycline biosynthetic pathway in *Myxococcus xanthus*. *Appl. Environ. Microbiol.* **76**, 2681–2683 (2010).
328. Chai, Y. *et al.* Heterologous expression and genetic engineering of the tubulysin biosynthetic gene cluster using Red/ET recombineering and inactivation mutagenesis. *Chem. Biol.* **19**, 361–371 (2012).
329. Sun, Y. *et al.* Heterologous Production of the Marine Myxobacterial Antibiotic Haliangicin and Its Unnatural Analogues Generated by Engineering of the Biochemical Pathway. *Sci. Rep.* **6**, 22091 (2016).
330. Yan, F. *et al.* Synthetic biology approaches and combinatorial biosynthesis towards heterologous lipopeptide production. *Chem. Sci.* **9**, 7510–7519 (2018).
331. Yan, F. *et al.* Biosynthesis and Heterologous Production of Vioprolides: Rational Biosynthetic Engineering and Unprecedented 4-Methylazetidinecarboxylic Acid Formation. *Angew. Chem. Int. Ed. Engl.* **57**, 8754–8759 (2018).
332. Pogorevc, D. *et al.* Biosynthesis and Heterologous Production of Argyrins. *ACS Synth. Biol.* **8**, 1121–1133 (2019).
333. Whitworth, D. (ed.). *Myxobacteria: Multicellularity and differentiation* (ASM Press, Chicago, 2007).
334. Streips, U. N. & Yasbin, R. E. *Modern microbial genetics* (John Wiley & Sons, 2004).
335. Johnston, C., Martin, B., Fichant, G., Polard, P. & Claverys, J.-P. Bacterial transformation: distribution, shared mechanisms and divergent control. *Nat. Rev. Microbiol.* **12**, 181–196 (2014).
336. Wenzel, S. C. & Müller, R. The impact of genomics on the exploitation of the myxobacterial secondary metabolome. *Nat. Prod. Rep.* **26**, 1385–1407 (2009).
337. Kopp, M. *et al.* Critical variations of conjugational DNA transfer into secondary metabolite multiproducing *Sorangium cellulosum* strains So ce12 and So ce56: development of a *mariner*-based transposon mutagenesis system. *J. Biotechnol.* **107**, 29–40 (2004).
338. Perlova, O., Gerth, K., Hans, A., Kaiser, O. & Müller, R. Identification and analysis of the chivosazol biosynthetic gene cluster from the myxobacterial model strain *Sorangium cellulosum* So ce56. *J. Biotechnol.* **121**, 174–191 (2006).
339. Zhao, J.-Y. *et al.* Discovery of the autonomously replicating plasmid pMF1 from *Myxococcus fulvus* and development of a gene cloning system in *Myxococcus xanthus*. *Appl. Environ. Microbiol.* **74**, 1980–1987 (2008).

340. Bhat, S., Zhu, X., Patel, R. P., Orlando, R. & Shimkets, L. J. Identification and Localization of *Myxococcus xanthus* Porins and Lipoproteins. *PLoS ONE* **6**, e27475 (2011).
341. Korp, J., Winand, L., Sester, A. & Nett, M. Engineering Pseudochelin Production in *Myxococcus xanthus*. *Appl. Environ. Microbiol.* **84**, e01789-18 (2018).
342. Rubin, E. J. *et al.* *In vivo* transposition of mariner-based elements in enteric bacteria and mycobacteria. *Proc. Natl. Acad. Sci. USA* **96**, 1645–1650 (1999).
343. Fu, J. *et al.* Efficient transfer of two large secondary metabolite pathway gene clusters into heterologous hosts by transposition. *Nucleic Acids Res.* **36**, e113 (2008).
344. Fujitani, Y., Yamamoto, K. & Kobayashi, I. Dependence of Frequency of Homologous Recombination on the Homology Length. *Genetics* **140**, 797–809 (1995).
345. Ueki, T., Inouye, S. & Inouye, M. Positive-negative KG cassettes for construction of multi-gene deletions using a single drug marker. *Gene* **183**, 153–157 (1996).
346. Yang, Y.-J. *et al.* Genome editing in model strain *Myxococcus xanthus* DK1622 by a Site-Specific Cre/loxP Recombination System. *Biomolecules* **8** (2018).
347. Yang, Y.-J. *et al.* Increasing on-target cleavage efficiency for CRISPR/Cas9-induced large fragment deletion in *Myxococcus xanthus*. *Microb. Cell Fact.* **16**, 142 (2017).
348. Peng, R. *et al.* CRISPR/dCas9-mediated transcriptional improvement of the biosynthetic gene cluster for the epothilone production in *Myxococcus xanthus*. *Microb. Cell Fact.* **17**, 15 (2018).
349. Jansen, R. *et al.* Pyrroazols, metabolites from the myxobacteria *Nannocystis pusilla* and *N. exedens*, are unusual chlorinated pyrone-oxazole-pyrroles. *J. Nat. Prod.* **77**, 320–326 (2014).
350. Hug, J. J. *et al.* Biosynthesis of citilins, unusual ribosomally synthesized and post-translationally modified peptides from *Myxococcus xanthus*. *ACS Chem. Biol.* **submitted** (2020).
351. Okoth Dorothy A. *et al.* 2-Hydroxysorangiadenosine: Structure and Biosynthesis of a Myxobacterial Sesquiterpene–Nucleoside. *Molecules (Basel, Switzerland)* **accepted** (2020).
352. Ahn, J. W., Jang, K. H., Chung, S. C., Oh, K. B. & Shin, J. Sorangiadenosine, a new sesquiterpene adenoside from the myxobacterium *Sorangium cellulosum*. *Org. Lett.* **10**, 1167–1169 (2008).

## Chapter 2

# Biosynthesis and Total Synthesis of Pyrronazol B: a Secondary Metabolite from *Nannocystis pusilla*

Swjatoslaw N. R. Witte<sup>†</sup>, **Joachim J. Hug<sup>†</sup>**, Magalie N. E. Geraldys, Rolf Müller  
and Markus Kalesse

<sup>†</sup> These authors contributed equally to this work

Dedicated to Professor Stefan Schulz on the occasion of his 60th birthday

*Chemistry – A European Journal*, **2017**, 23, 15917–15921

DOI: 10.1002/chem.201703782

Published online: 24.09.2017

### **Contributions to the presented work**

#### **Author's contribution**

The author significantly contributed to the conception of this study, designed and performed experiments, evaluated and interpreted resulting data. The author developed and performed feeding experiments and analyzed the corresponding LC-MS data. Thereby the author identified the biosynthetic gene cluster and conceived a biosynthetic model for the pyrronazols. Furthermore, the author contributed significantly to conceiving and writing this manuscript.

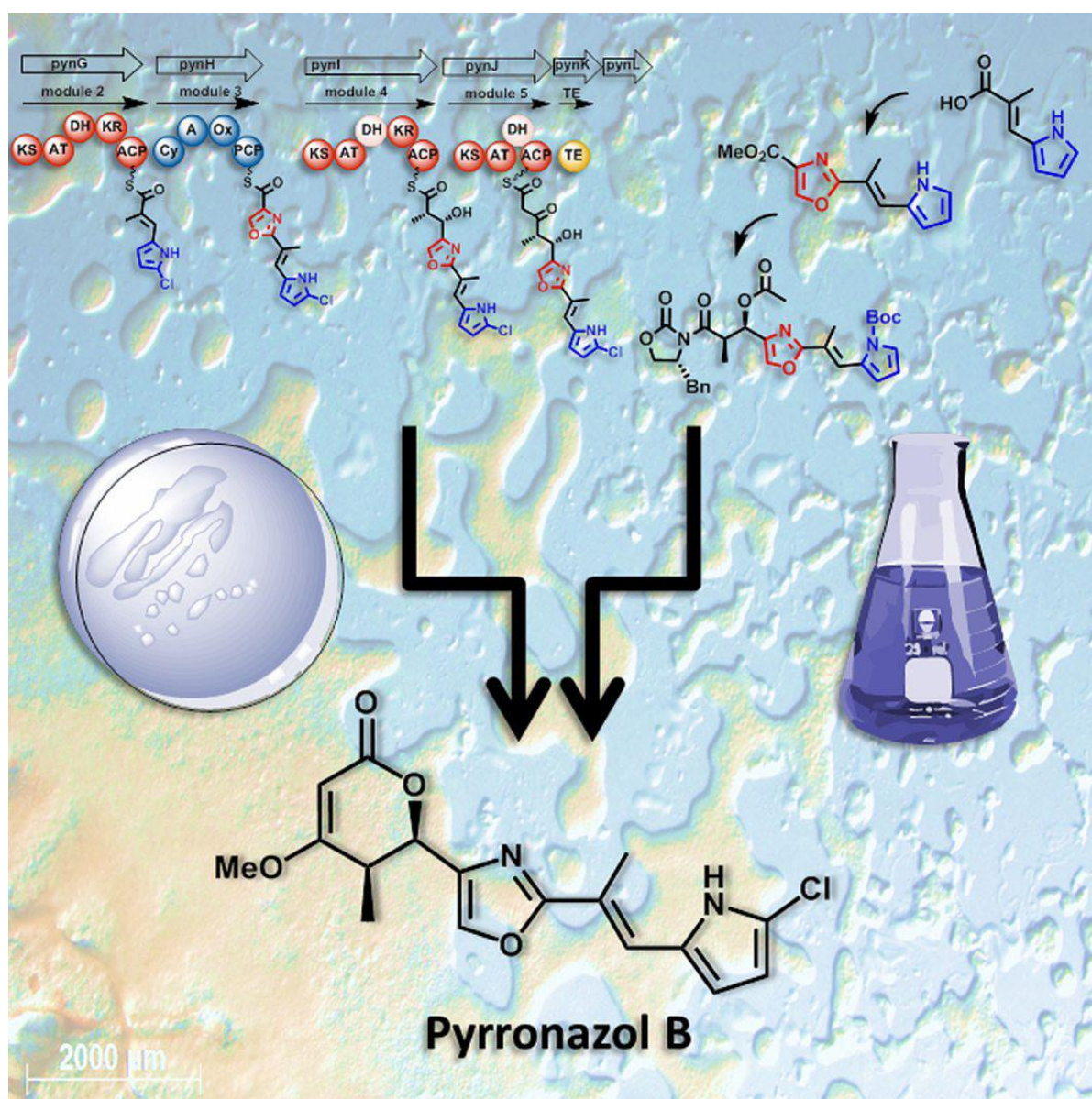
#### **Contribution by others**

Swjatoslaw N. R. Witte significantly contributed to the conception of this study, designed and performed experiments, evaluated and interpreted resulting data. He developed a retrosynthetic route to pyrronazol B and performed all experiments related to its total synthesis. In addition he contributed significantly to conceiving and writing this manuscript. Magalie N. E. Geraldly performed experiments, evaluated and interpreted resulting data and contributed to the conception of this study. Markus Kalesse and Rolf Müller contributed by supervision of this project and conceiving, editing and proofreading of this manuscript.

## 2 Biosynthesis and Total Synthesis of Pyrronazol B: a Secondary Metabolite from *Nannocystis pusilla*

### 2.1 Abstract

The first stereoselective total synthesis of the natural product pyrronazol B, which contains a chlorinated pyrrole–oxazole–pyrone framework, has been achieved. Genome sequencing of the myxobacterial producer strain *Nannocystis pusilla* Ari7 led to the identification of the putative biosynthetic gene cluster. The proposed biosynthetic pathway was supported by feeding experiments with stable isotopes of three biosynthetic building blocks, namely L-proline, L-serine, and L-methionine.



Frontispiece

## 2.2 Introduction

Myxobacterial microorganisms constitute a prolific source for novel natural compounds, exhibiting great potential for the discovery of new drugs against human diseases with great chemical diversity and unusual mode of actions<sup>1</sup>. Pyrronazol A1 and B are produced in the myxobacterium *Nannocystis pusilla* Ari7 and have been isolated from the fermentation broth by Müller and colleagues in 2014<sup>2</sup>. Their tricyclic carbon skeleton including a halogenated pyrrole moiety represents a novel structural class of secondary metabolites. The structure of the pyrronazol derivatives suggests that their biosynthetic route is based on a nonribosomal peptide synthetase-polyketide synthase (NRPS–PKS) hybrid biochemistry.

We report herein the total synthesis of pyrronazol B and the *in silico* characterization of pyrronazols biosynthesis based on feeding experiments.

## 2.3 Results and Discussion

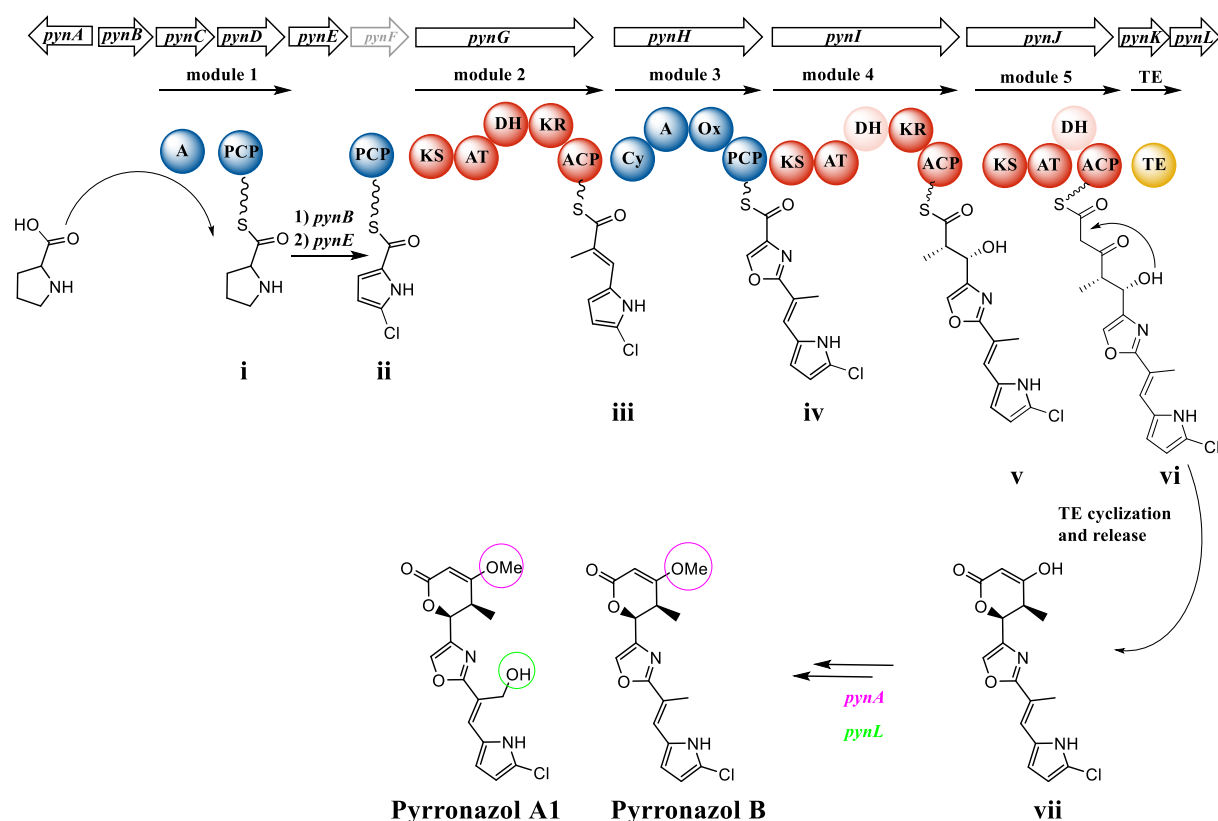
*In silico* investigation of putative biosynthetic gene clusters in the natural producer *Nannocystis pusilla* strain Ari7 lead to one candidate gene cluster matching the biosynthetic structure of pyrronazols. The cluster spans approximately 29.9 kb and consists of twelve genes encoding a NRPS–PKS hybrid product, regulators, transporters and tailoring enzymes. Because it was not possible to genetically manipulate *Nannocystis pusilla* Ari7, the identification of the gene cluster and the following biosynthesis proposal is based on *in silico* analysis supported by feeding experiments with stable isotopes of three biosynthetic building blocks L-methionine-(methyl-<sup>13</sup>C), L-serine-1-<sup>13</sup>C and L-proline-<sup>15</sup>N to observe through incorporation a mass shift of +1 Da in the isotopic pattern of pyrronazol A1 and B. The putative biosynthetic gene cluster contains twelve open reading frames (ORFs). The NRPS–PKS hybrid cluster contains five modules; three of these are PKS modules (*pynG*, *pynI*, *pynJ*), whereas two modules are NRPS derived (*pynC*, *pynD*, *pynH*) (**Scheme 1**).

The assembly line starts with the activation of L-proline to L-prolyl-adenosine monophosphate (AMP) by the adenylation (A) domain PynC. The Stachelhaus specificity-conferring code is identical to the A-domains of leupyrrin, DKxanthen and pyoluteorin<sup>3–5</sup>. Feeding studies with [<sup>15</sup>N] labeled L-proline clearly show incorporation with a mass shift of +1 Da (see Figures S2 and S5) for pyrronazol A1 and B. It appears likely that L-prolyl-AMP is loaded onto the PCP domain PynD resulting L-proline intermediate **i**. The amino acid sequence of PynD is similar to the crystallized PCP domain PltL (35%), which is responsible for loading of AMP activated L-proline during the biosynthesis of pyoluteorin in the natural producer *Pseudomonas fluorescens* Pf-5. The amino acid sequence of PynD is in accordance with the proposed conserved residues and polarity necessary for pyrrole PCPs<sup>6</sup>. The tethered L-prolyl-S-PCP protein is probably first modified through a four-electron, two-step tandem desaturation by the action of the putative dehydrogenase PynB to the pyrrolyl-2-carboxyl-S–PCP as described in the pyoluteorin, prodigiosin, coumermycin A1 and chlorobiocin biosynthesis<sup>7</sup>.

Monochlorination at C<sub>5</sub> is assumed to take place while the heterocycle is tethered to the PCP domain PynD by the action of the halogenase PynE to yield 5-chloropyrrolyl-2-thioester **ii**. The amino acid sequence of PynE is similar to the crystallized dichlorinating enzyme required for pyoluteorin



biosynthesis in *Pseudomonas fluorescens* Pf-5 PltA (57%) and to Mpy16 (60%) which is involved in dichlorination of marinopyrrole. The PKS module PynF located between the halogenase *pynE* and the PKS module encoding *pynG* is probably not functional. The module PynF encodes a ketosynthase (KS), a dehydratase (DH) and an acyl carrier protein (ACP) domain. The KS domain neither contains the catalytic cysteine embedded in the motif TACSS nor the neighboring histidine residues. Furthermore, the DH domain does not contain the catalytic aspartic acid in the HPALLD motif or the necessary histidine in the HxxxGxxxxP motif. Missing these important motifs in both domains concludes that this module might not be functional and not contribute to the overall biosynthesis of pyrronazols. Another interesting detail within the biosynthetic gene cluster architecture is the presence of a polyketide cyclase *orf7*. The first chain extension occurs with condensation of 5-chloropyrrolyl-2-thioester **ii** with methylmalonate, catalyzed by the PKS module PynG to generate intermediate **iii**. The fingerprint motif of the acyl transferase (AT) domain in *pynG* concludes the specificity for (2*S*)-methylmalonyl-CoA<sup>8</sup>.



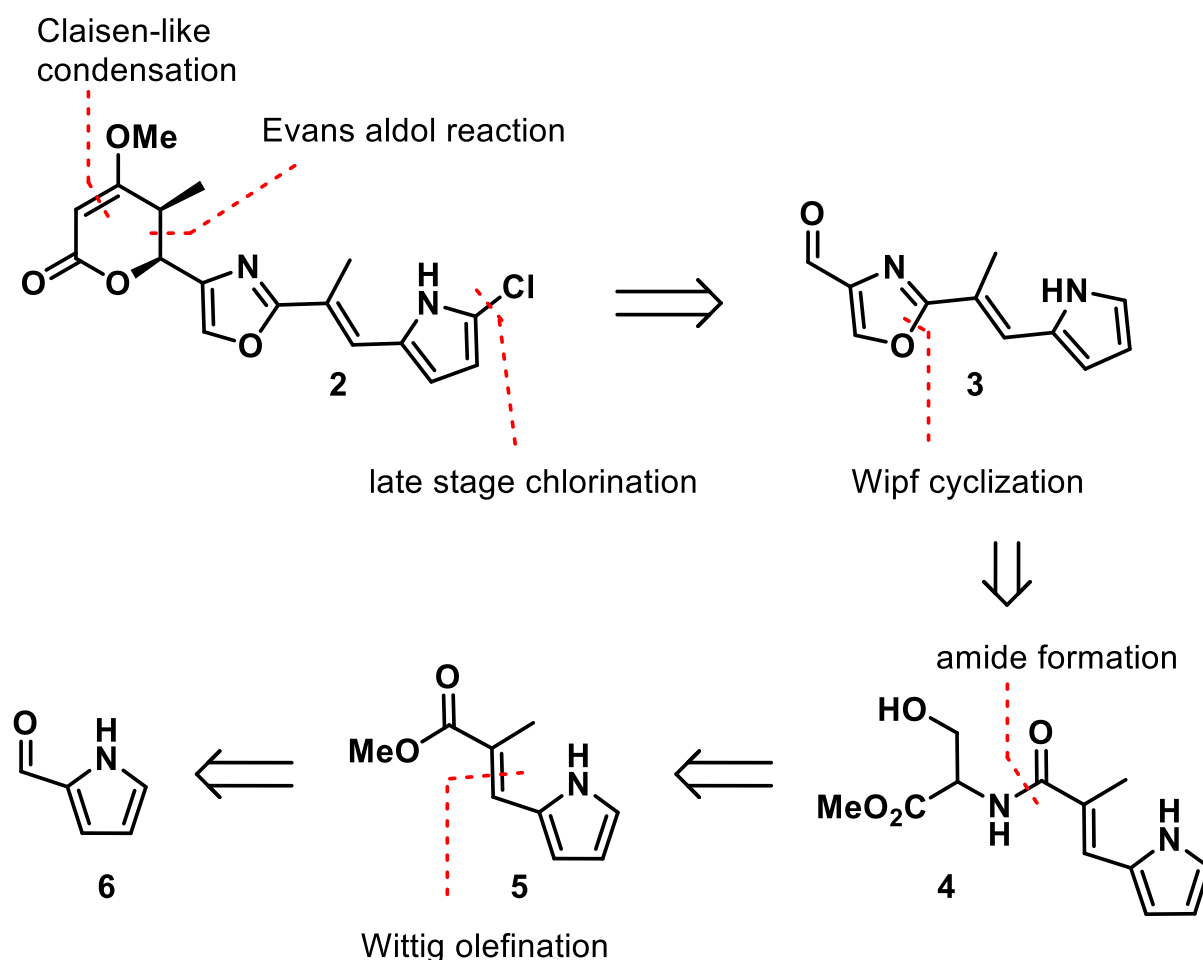
The next building block incorporated into the pyrronazol structure is L-serine, generating oxazole **iv** as supported by feeding studies using [ $^{13}\text{C}$ ] labeled L-serine (see Figures S3 and S6). The similar NRPS module EpoB (46.1%) is described in *S. cellulorum* for the biosynthesis of epothilone B<sup>9</sup>. The acyl chain on the ACP domain of PynG is transferred to the amino group of the serine-S-PynH, generating a transient amide intermediate. Cyclization and dehydration yields the five-membered ring oxazolinyll-S-PynH. Finally the flavin-mononucleotide-containing oxidation domain oxidizes the five-membered intermediate to the heteroaromatic oxazole<sup>10</sup>. In this NRPS module the conventional condensation (C) domain is replaced by a cyclization domain (Cy) containing the catalytic motif DxxxxDxxS instead of the common motif HHxxxDG<sup>11</sup>. The A domain of PynH seems to be specific for L-serine; however *in silico* prediction based on the Stachelhaus code suggests aspecificity for L-cysteine. The A-domain involved in the ajudazol biosynthesis and in the leupyrrin biosynthesis exhibits like the A-domain of pyrronazol an unusual Stachelhaus code motif specific for L-serine (see Supporting Information Table S2)<sup>12</sup>. Nevertheless the result of the feeding study, the overall structure of pyrronazol and the lack of any thiazole containing pyrronazol derivatives, leads to the conclusion that the A domain of PynH is in fact activating L-serine instead of L-cysteine.

The next PKS module PynI loads a second methylmalonyl- CoA extender unit onto the growing pyrronazol structure **v**. In accordance to the reduction state observed in pyrronazols the DH-domain within the PKS module is inactive since it does not contain the required catalytic aspartic acid in the HPALLD motif or the necessary histidine in the HxxxGxxxxP motif<sup>13</sup>. Hence only the ketoreductase (KR) domain is active and generates the hydroxyl group by reduction of the extender unit. In addition, our statistical analysis of the ketoreductase gave a ScoreDiff value of 42.65 for the secondary alcohol and 10.80 for the methyl group. The oxy-functionality and the methyl branch were therefore predicted to be d-configured as they are generated by a B-type KR domain<sup>14</sup>. The last step on the assembly line is catalyzed by the PKS module PynJ which also contains a nonfunctional DH domain. The linear NRPS-PKS chain **vi** is probably released and cyclized by the action of the thioesterase PynK, containing the GxSxG closely located to the N-terminus and the motif GxH located nearby the C-terminus both harboring the catalytic serine and the histidine, respectively.

Finally, the released and cyclized precursor **vii** might be modified by the action of the O-methyltransferase PynA and the oxidoreductase PynL. The presence of the non-hydroxylated pyrronazol B and the hydroxylated pyrronazol A1 supports the hypothesis of a post-PKS oxidation, whereas the O-methyltransferase could also be active before the cyclization and release. In order to support the proposed mechanism of the S-adenosyl-L-methionine dependent methyltransferase PynA, feeding with [ $^{13}\text{C}$ ]-L-methionine was performed, corroborating the origin of the methyl group (see **Figure S1** and **S4**).

In the context of our program to establish synthetic access to hybride natural products derived from polyketidal and nonribosomal fragments<sup>15–21</sup> we aimed at synthesizing pyrronazol B. This synthetic endeavor was initiated to confirm the proposed structure and to gain access to derivatives for medicinal chemistry. Our retrosynthetic analysis parallels the proposed biosynthesis of pyrronazol regarding the direction of the assembly. As in the biosynthesis the synthesis starts from a pyrrole derivative and extends the carbon framework towards the carboxylate terminus. In synthetic direction this is achieved through a Wittig olefination, a Wipf cyclization of amide **4** and finally by an aldol reaction and an ester condensation (**Scheme 2**).

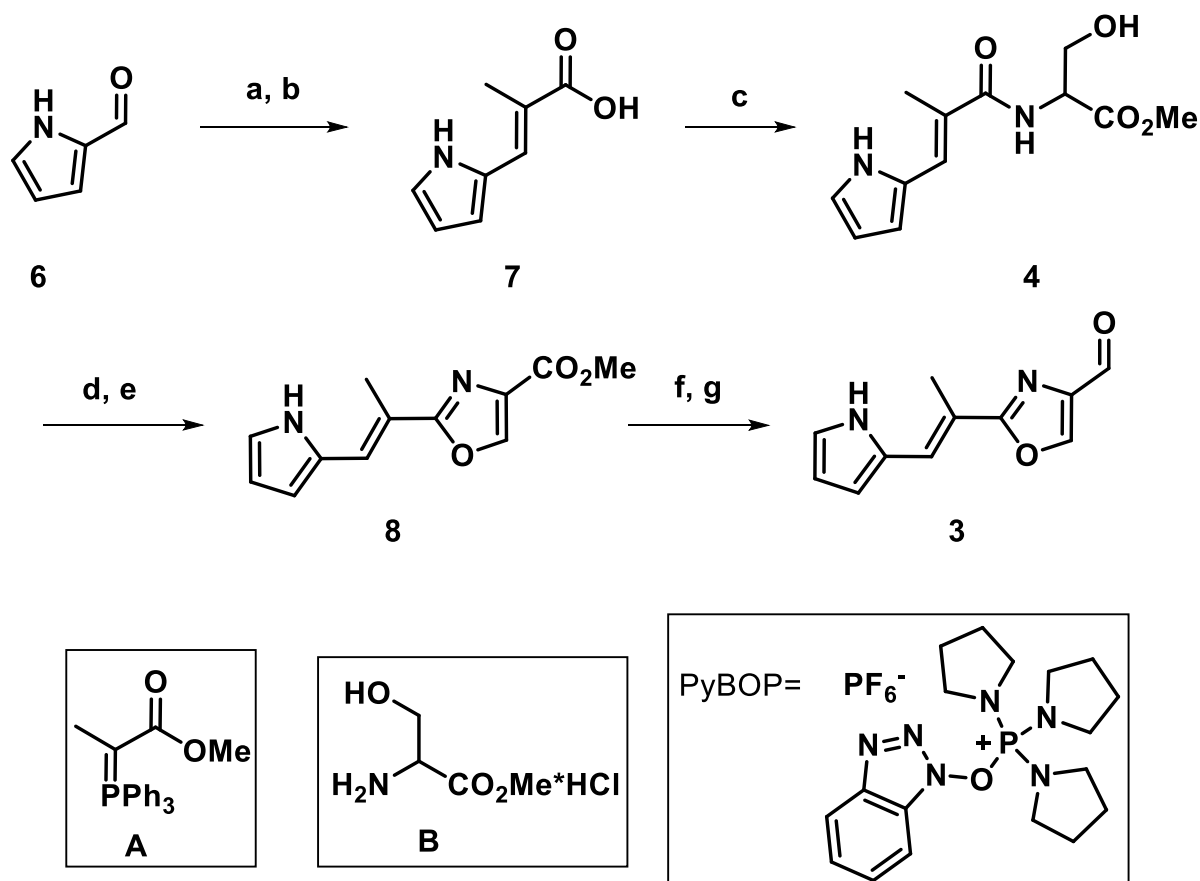
Starting from pyrrole-2-carboxyaldehyde (**6**) a Wittig olefination with A in benzene generated the corresponding unsaturated ester in 90%<sup>22</sup>. The ester was then hydrolyzed with KOH to quantitatively yield acid **7** which was used in the subsequent amide formation using DL-serine methylester hydrochloride (B) and PyBOP as the coupling reagent. The so-generated amide **4** was used in a two-step Wipf cyclization<sup>23</sup>. This sequence includes a diethylaminosulfur trifluoride (DAST)-mediated oxazoline formation followed by an oxidation with BrCCl<sub>3</sub> and 1,8-diazabicyclo[5.4.0]undec-7-ene (DBU) to generate oxazole **8** in 46% yield over two steps. The ester group in **8** was reduced with LiAlH<sub>4</sub> to the corresponding alcohol in 68% yield, followed by Hoover–Stahl<sup>24</sup> oxidation to generate aldehyde **3** in 90% yield (**Scheme 3**). It is worth mentioning at this point that only the Hoover–Stahl protocol was able to oxidize the alcohol to its corresponding aldehyde. Oxidations using MnO<sub>2</sub>, pyridinium dichromate (PDC), Dess–Martin periodinane (DMP) or the Swern protocol did not generate the aldehyde or led to the decomposition of the starting material.



**Scheme 2.** Retrosynthetic analysis of pyrronazol B.

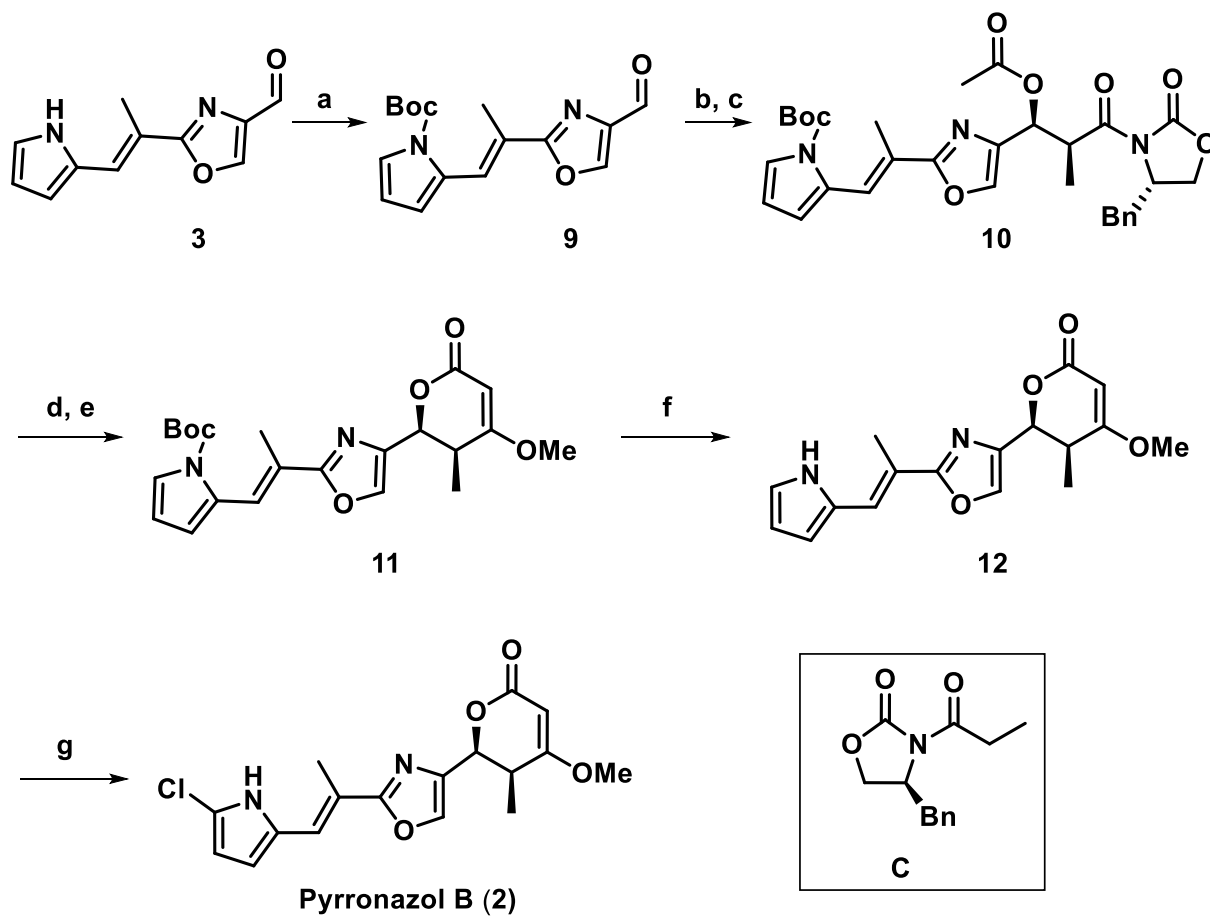
With aldehyde **3** in hands, we started the construction of the unsaturated lactone. Therefore, aldehyde **3** was Boc-protected to provide **9**, and to enable the subsequent proton-sensitive transformations. The successive Evans aldol reaction was performed with TiCl<sub>4</sub><sup>25</sup> and provided a 4:1 diastereomeric mixture of the Evans and non-Evans product in favor of the desired isomer. In this transformation TiCl<sub>4</sub> proved to be superior compared to Bu<sub>2</sub>BOTf for which only decomposition of the starting material was observed.

Next, the product mixture was subjected to an acetylation with Ac<sub>2</sub>O and 4-dimethylaminopyridine (4-DMAP). At this point it was possible to separate the diastereoisomers by flash column chromatography and to obtain **10** in 58% yield over two steps. This acetylation set the stage for an intramolecular Claisen condensation<sup>26</sup> which was initiated by treatment with LiHMDS and provided the desired lactone. However, separation of the cyclization product and the Evans auxiliary was only possible after treatment with Me<sub>2</sub>SO<sub>4</sub> and K<sub>2</sub>CO<sub>3</sub> (**Scheme 4**).



**Scheme 3.** Synthesis of aldehyde **3**: a) **A**, C<sub>6</sub>H<sub>6</sub>, 80 °C, 3 h, 90%; b) KOH, MeOH:H<sub>2</sub>O 2:1, rt, 24 h, quant.; c) **B**, DMF, 0 °C, PyBOP, DIPEA, 0 °C to rt, 24 h, 95%; d) DAST, K<sub>2</sub>CO<sub>3</sub>, CH<sub>2</sub>Cl<sub>2</sub>, -78 °C, 24 h; e) BrCCl<sub>3</sub>, DBU, CH<sub>2</sub>Cl<sub>2</sub>, 0 °C, 1.5 h, 46% over two steps; f) LiAlH<sub>4</sub>, THF, 0 °C, 1 h, 68%; g) 5% Cu(MeCN)<sub>4</sub> (OTf), 5% bpy, 5% TEMPO, 5% 4-DMAP, air, MeCN, rt, 1.5 h, 90%. rt = room temperature, DIPEA = *N,N*-diisopropylethylamine, TEMPO = 2,2,6,6-tetramethylpiperidin-1-yl)oxyl.

The last steps of the synthesis included the removal of the Boc group. Experimentation with various conditions such as treatment with 0.5 M HCl or K<sub>2</sub>CO<sub>3</sub> in methanol led only to decomposition of the starting material. Finally, treatment of **11** with water at 100 °C followed by purification via HPLC gave clean deprotected **12** which was chlorinated with *N*-chlorosuccinimide (NCS) to form pyrroazol B in 23% yield (**Scheme 4**). Other chlorination methods using solvents such as acetone or THF did not provide improved yields. The optical rotation (authentic material ([α]<sub>D</sub><sup>25</sup> = -132.3, synthetic material ([α]<sub>D</sub><sup>27</sup> = -125.2) as well as the other spectroscopic data of the synthetic and authentic material were in full agreement.



**Scheme 4.** Synthesis of the lactone **11** a)  $\text{Boc}_2\text{O}$ , 4-DMAP,  $\text{CH}_2\text{Cl}_2$ , rt, 1 h, 76%; b) **C**,  $\text{TiCl}_4$ , DIPEA,  $\text{CH}_2\text{Cl}_2$ , 0 °C, 30 min, then, **9**, -78 °C, 3 h; c)  $\text{Ac}_2\text{O}$ , 4-DMAP,  $\text{CH}_2\text{Cl}_2$ , rt, 30 min, 58% % over two steps; d) LiHMDS, THF, -78 °C, 1.5 h;<sup>24</sup> e)  $\text{Me}_2\text{SO}_4$ ,  $\text{K}_2\text{CO}_3$ , acetone, 50 °C, 30 min, 43% % over two steps; f)  $\text{H}_2\text{O}$ , 100 °C, 1 h, 56%; g) NCS, THF, rt, 24 h, 23%. rt = room temperature, LiHMDS = lithium bis(trimethylsilyl)amide.

## 2.4 Conclusions

We have provided the putative biosynthetic gene cluster of pyrronazol, a secondary metabolite isolated from *Nannocystis pusilla* Ari7. Additionally, we were able to provide synthetic access to pyrronazol B and thus confirm the proposed structure of the natural product. Based on the biosynthetic analysis and synthetic protocol, it will be possible to provide new, optimized analogues of the pyrronazol family.

## 2.5 Experimental Section

Experimental procedures, spectral data, gene cluster analysis and copies of  $^1\text{H}$ ,  $^{13}\text{C}$  NMR spectra, LC-MS and HPLC chromatograms are available in the **Supporting Information**.

## 2.6 Acknowledgements

We thank Bastien Schnell and Fabian Panter for biosynthetic discussions. We are grateful to Dr. Nestor Zaburannyi for bioinformatic support. We thank the Boehringer Ingelheim Fonds for financial support.

## 2.7 Conflict of Interest

The authors declare no conflict of interest.

## 2.8 Key words

biosynthetic gene cluster, deprotection, pyrronazol, total synthesis.

## 2.9 References

1. Herrmann, J., Fayad, A. A. & Müller, R. Natural products from myxobacteria: novel metabolites and bioactivities. *Nat. Prod. Rep.* **34**, 135–160 (2017).
2. Jansen, R. *et al.* Pyrronazols, metabolites from the myxobacteria *Nannocystis pusilla* and *N. exedens*, are unusual chlorinated pyrone-oxazole-pyrroles. *J. Nat. Prod.* **77**, 320–326 (2014).
3. Kopp, M. *et al.* Insights into the complex biosynthesis of the leupyrrins in *Sorangium cellulosum* So ce690. *Mol. Biosyst.* **7**, 1549–1563 (2011).
4. Meiser, P. *et al.* DKxanthene biosynthesis - understanding the basis for diversity-oriented synthesis in myxobacterial secondary metabolism. *Chem. Biol.* **15**, 771–781 (2008).
5. Nowak-Thompson, B., Chaney, N., Wing, J. S., Gould, S. J. & Loper, J. E. Characterization of the Pyoluteorin Biosynthetic Gene Cluster of *Pseudomonas fluorescens* Pf-5. *J. Bacteriol.* **181**, 2166–2174 (1999).
6. Jaremko, M. J., Lee, D. J., Opella, S. J. & Burkart, M. D. Structure and substrate sequestration in the pyoluteorin type II peptidyl carrier protein PltL. *J. Am. Chem. Soc.* **137**, 11546–11549 (2015).
7. Walsh, C. T., Garneau-Tsodikova, S. & Howard-Jones, A. R. Biological formation of pyrroles: nature's logic and enzymatic machinery. *Nat. Prod. Rep.* **23**, 517–531 (2006).
8. Yadav, G., Gokhale, R. S. & Mohanty, D. Computational approach for prediction of domain organization and substrate specificity of modular polyketide synthases. *J. Mol. Biol.* **328**, 335–363 (2003).
9. Dowling, D. P. *et al.* Structural elements of an NRPS cyclization domain and its intermodule docking domain. *Proc. Natl. Acad. Sci. U.S.A.* **113**, 12432–12437 (2016).

10. Chen, H., O'Connor, S., Cane, D. E. & Walsh, C. T. Epothilone biosynthesis: assembly of the methylthiazolylcarboxy starter unit on the EpoB subunit. *Chem. Biol.* **8**, 899–912 (2001).
11. Duerfahrt, T., Eppelmann, K., Müller, R. & Marahiel, M. A. Rational design of a bimodular model system for the investigation of heterocyclization in nonribosomal peptide biosynthesis. *Chem. Biol.* **11**, 261–271 (2004).
12. Buntin, K., Weissman, K. J. & Müller, R. An unusual thioesterase promotes isochromanone ring formation in ajudazol biosynthesis. *ChemBioChem* **11**, 1137–1146 (2010).
13. Keating-Clay, A. T. The structures of type I polyketide synthases. *Nat. Prod. Rep.* **29**, 1050–1073 (2012).
14. Kitsche, A. & Kalesse, M. Configurational assignment of secondary hydroxyl groups and methyl branches in polyketide natural products through bioinformatic analysis of the ketoreductase domain. *ChemBioChem* **14**, 851–861 (2013).
15. Brodmann, T., Janssen, D. & Kalesse, M. Total synthesis of chivosazole F. *J. Am. Chem. Soc.* **132**, 13610–13611 (2010).
16. Jahns, C. *et al.* Pellasoren: structure elucidation, biosynthesis, and total synthesis of a cytotoxic secondary metabolite from *Sorangium cellulosum*. *Angew. Chem. Int. Ed. Engl.* **51**, 5239–5243 (2012).
17. Rentsch, A. & Kalesse, M. The total synthesis of corallopyronin A and myxopyronin B. *Angew. Chem. Int. Ed. Engl.* **51**, 11381–11384 (2012).
18. Rentsch, A. & Kalesse, M. Die Totalsynthesen von Corallopyronin A und Myxopyronin B. *Angewandte Chemie (Weinheim an der Bergstrasse, Germany)* **124**, 11543–11547 (2012).
19. Hartmann, O. & Kalesse, M. The structure elucidation and total synthesis of  $\beta$ -lipomycin. *Angew. Chem. Int. Ed. Engl.* **53**, 7335–7338 (2014).
20. Tautz, T. *et al.* Isolation, Structure Elucidation, Biosynthesis, and Synthesis of Antalid, a Secondary Metabolite from *Polyangium species*. *Org. Lett.* **18**, 2560–2563 (2016).
21. Steinmetz, H. *et al.* Isolation, structure elucidation, and (bio)synthesis of Haprolid, a cell-type-specific myxobacterial cytotoxin. *Angew. Chem. Int. Ed. Engl.* **55**, 10113–10117 (2016).
22. Campbell, S. E. *et al.* Synthesis of pyrrolizin-3-ones by flash vacuum pyrolysis of pyrrol-2-ylmethylidene Meldrum's acid derivatives and 3-(pyrrol-2-yl)propenoic esters. *J. Chem. Soc. [Perkin 1]*, 2195–2202 (1997).
23. Wipf, P. & Graham, T. H. Total Synthesis of (–)-Disorazole C<sub>1</sub>. *J. Am. Chem. Soc.* **126**, 15346–15347 (2004).
24. Hoover, J. M., Ryland, B. L. & Stahl, S. S. Mechanism of copper(I)/TEMPO-catalyzed aerobic alcohol oxidation. *J. Am. Chem. Soc.* **135**, 2357–2367 (2013).

25. Crimmins, M. T., King, B. W., Tabet, E. A. & Chaudhary, K. Asymmetric aldol additions: use of titanium tetrachloride and (–)-sparteine for the soft enolization of *N*-acyl oxazolidinones, oxazolidinethiones, and thiazolidinethiones. *J. Org. Chem.* **66**, 894–902 (2001).
26. Hinterding, K., Singhanat, S. & Oberer, L. Stereoselective synthesis of polyketide fragments using a novel intramolecular Claisen-like condensation/reduction sequence. *Tetrahedron Lett.* **42**, 8463–8465 (2001).



## Chapter 2

### Supporting Information

#### **Biosynthesis and Total Synthesis of Pyrronazol B: a Secondary Metabolite from *Nannocystis pusilla***

Swjatoslaw N. R. Witte<sup>[a]</sup>, **Joachim J. Hug**<sup>[b]</sup>, Magalie N. E. Geraldys<sup>[a]</sup>, Rolf Müller<sup>[b]</sup> and Markus Kalesse<sup>[a,c]</sup>

a) Institute for Organic Chemistry, Leibniz University Hannover, Schneiderberg 1B, 30167 Hannover; Centre of Biomolecular Drug Research (BMWZ), Schneiderberg 38, 30167 Hannover, Germany

b) Department of Microbial Natural Products, Helmholtz-Institute for Pharmaceutical Research Saarland (HIPS), Helmholtz Centre for Infection Research (HZI) Saarland University, building E8.1, D-66123 Saarbruecken, Germany

c) Helmholtz Centre for Infection Research (HZI), Inhoffenstrasse 7, D-38124 Braunschweig, Germany  
Helmholtz Centre for Infection Research (HZI)

## Table of Contents

General material and methods .....	85
Results: .....	87
1. [Partial LC_MS spectra].....	87
2. Genes encoded in the pyrronazol biosynthetic gene cluster.....	93
3. Stereochemistry prediction of the secondary alcohol moiety and methyl model during chain elongation in module 4 .....	94
4. <i>In silico</i> analysis of the A domain of pynH.....	95
5. Comparison of NMR-spectroscopic data of authentic and synthetic Pyrronazol B in CDCl <sub>3</sub> .....	96
6. Total synthesis of Pyrronazol B .....	99
7. Copies of <sup>1</sup> H and <sup>13</sup> C NMR spectra .....	108
8. Copies of preparative HPLC chromatograms.....	127
References .....	129

## General material and methods

### [Identification of the Pyrroazol Biosynthetic Gene Cluster in *Nannocystis pusilla* Ari7]

The gene sequence of the pyrroazol gene cluster has been deposited in GenBank and is accessible under the accession number MF817819. Genomic DNA of *Nannocystis pusilla* Ari7 was sequenced by Illumina sequencing technology.<sup>1</sup> Initial prediction and analysis of the pyrroazol biosynthesis gene cluster was done by using antiSMASH 3.0 (<https://antismash.secondarymetabolites.org/#!/start>).<sup>2</sup> Detailed prediction of NRPS A domain specificity was performed by using NRPSpredictor2 (<http://nrps.informatik.uni-tuebingen.de/Controller?cmd=SubmitJob>).<sup>3</sup> The functional prediction of open reading frames (ORFs) encoding proteins was performed using protein blast and/or blastx program (<https://blast.ncbi.nlm.nih.gov/Blast.cgi>) and Pfam (<http://pfam.xfam.org>).<sup>4</sup> Routine DNA analysis such as ORF identification, alignments and searching for specific motifs was done using Geneious software (Biomatters Ltd., Auckland, New Zealand)

### [Feeding experiment with stable isotope labeled building blocks]

*Nannocystis pusilla* Ari7 was maintained as a tenfold concentrated cryo sample stored at -80 °C. 250 µl of the cryo sample was used to inoculate 20 mL of CyH medium (Soy flour 0.2%, Bacto yeast extract 0.2%, Bacto casiton 0.3%, glucose 0.2%, soluble starch 0.8%, CaCl<sub>2</sub>·2H<sub>2</sub>O 0.1%, MgSO<sub>4</sub>·7H<sub>2</sub>O 0.1%, HEPES 50 mM, pH adjusted to 7.2 with KOH before autoclaving, after autoclaving Fe-EDTA (8 mg/L) was added in a 100 mL flask. 1 mL of the preculture was used to inoculate 20 mL of CyH medium and for four consecutive days 20 µl of a 0.1 M solution of either L-methionine-(methyl-<sup>13</sup>C), L-serine-1-<sup>13</sup>C or L-proline-<sup>15</sup>N in sterile water was added. Additionally one sample without feeding was prepared used as control. 12 h after the last supplementation with stable isotope labeled building blocks, the culture broths were complemented with 2% amberlite resin XAD-16 (Sigma). After 10 days of incubation at 30 °C and 200 rpm, the culture broths were harvested by centrifugation at 8000 rpm for 10 min, the supernatant was discarded, and the residual material extracted was extracted first with 25 mL methanol, stirred for 1 h, filtered through filter paper into a round bottom flasks and afterwards this procedure was repeated with 25 mL acetone. Filtered extracts were evaporated and subsequently redissolved in 2 mL methanol. For metabolic analysis the redissolved sample was centrifuged and analyzed by a Dionex HPLC coupled to the TOF mass spectrometer maXis 4G as described further below.

**[Analysis of secondary metabolism of broth extracts]**

For analyzing the secondary metabolism of broth extracts of *Nannocystis pusilla* Ari7 LC-HRESI-DAD-MS measurements were acquired on a Bruker maXis 4G mass spectrometer coupled with a Dionex Ultimate 3000 RSLC system using a BEH C18 column (100 × 2.1 mm, 1.7 µm, Waters, Germany) with a gradient of 5–95% MeCN + 0.1% FA in H<sub>2</sub>O + 0.1% FA at 0.6 mL/min and 45 °C over 18 min with UV detection by a diode array detector at 200–600 nm. Mass spectra were acquired from 150 to 2000 m/z at 2 Hz. The detection was performed in the positive MS mode.

**[Total synthesis of Pyrroazonol B]**

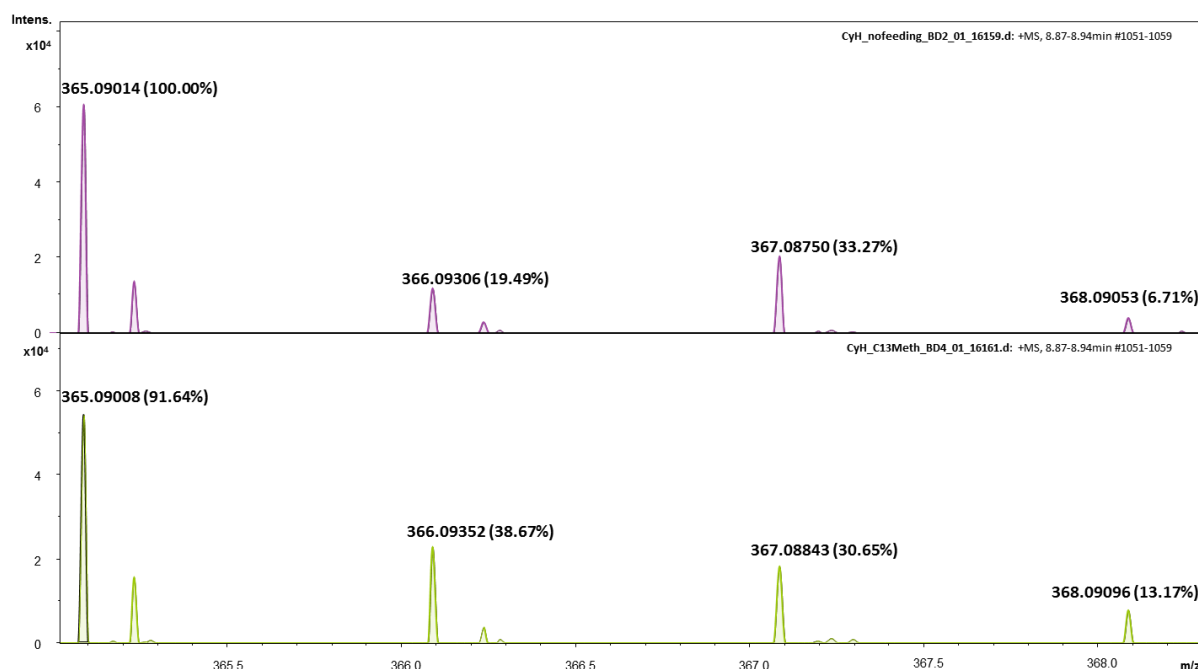
All reactions were carried out under dry argon unless otherwise stated. Commercially available reagents were used without further purification. Solvents were dried accordingly to standard procedures. Organic solvents were evaporated with reduced pressure using a rotary evaporator. Analytical thin layer chromatography (TLC) was performed using aluminium plates with silica gel (0.20 mm). TLC plates were visualized by exposure to UV light (UV), and then were visualized by either vanillin stain, bromocresol green, CAM or 2,4-NP followed by brief heating. Flash column chromatography was performed using silica gel 60 M (0.04–0.063 mm, Macherey-Nagel GmbH Co. KG) with the indicated solvents. <sup>1</sup>H and <sup>13</sup>C spectra were recorded on Bruker avance III HD (400 MHz). <sup>1</sup>H NMR spectra are represented as follows: chemical shift, multiplicity (s = singlet, d = doublet, t = triplet, q = quartet, m = multiplet), integration and coupling constant (*J*) in hertz (Hz). <sup>1</sup>H chemical shifts are reported relative to CDCl<sub>3</sub> (7.26 ppm) unless otherwise stated. <sup>13</sup>C NMR was recorded relative to the central line of CDCl<sub>3</sub> (77.16 ppm) unless otherwise stated. Purification *via* HPLC was performed on AlphaCrom Laboratory Equipment Agilent Technologies PrepStar. The column was packed with Resprospher 100 C18-DE, 10 µm, 250 mm x 20 mm, ID. The flow rate was 15 mL/min at ambient temperature, solvent A: water, solvent B: MeCN.

**Table 1:** Preparative HPLC conditions.

time [min]	% A	% B
0	80	20
5	80	20
15	20	80
20	20	80
25	5	95
30	5	95
34	0	100
39	80	20

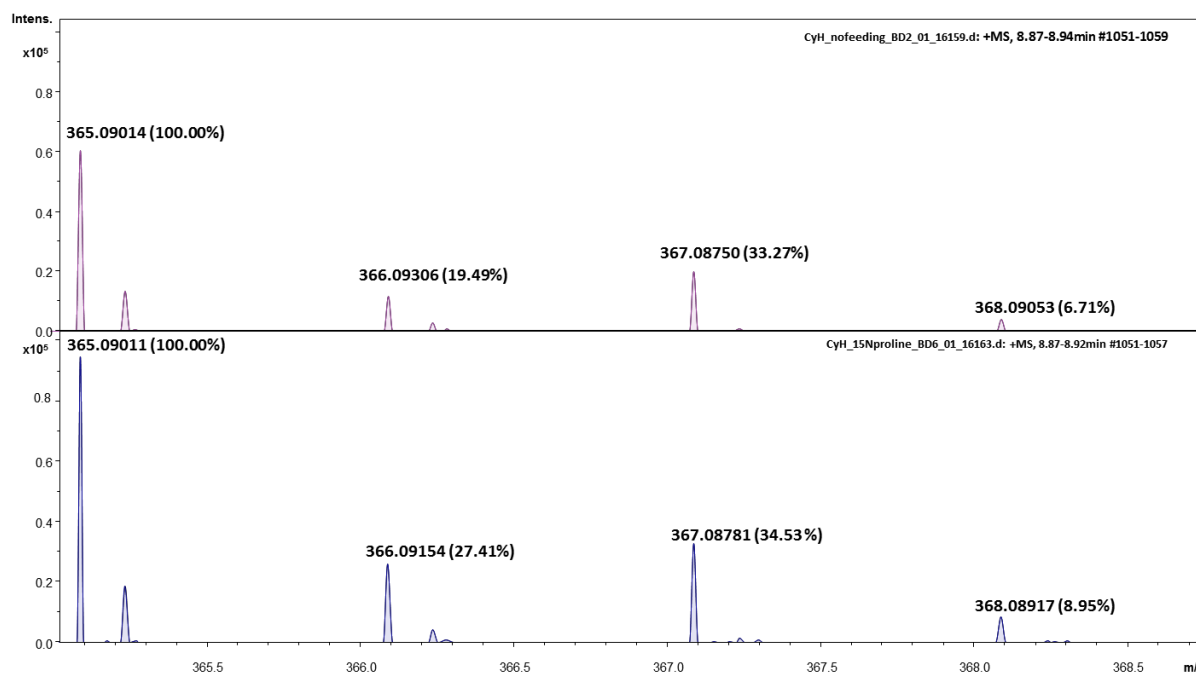
## Results:

### 1. [Partial LC\_MS spectra]



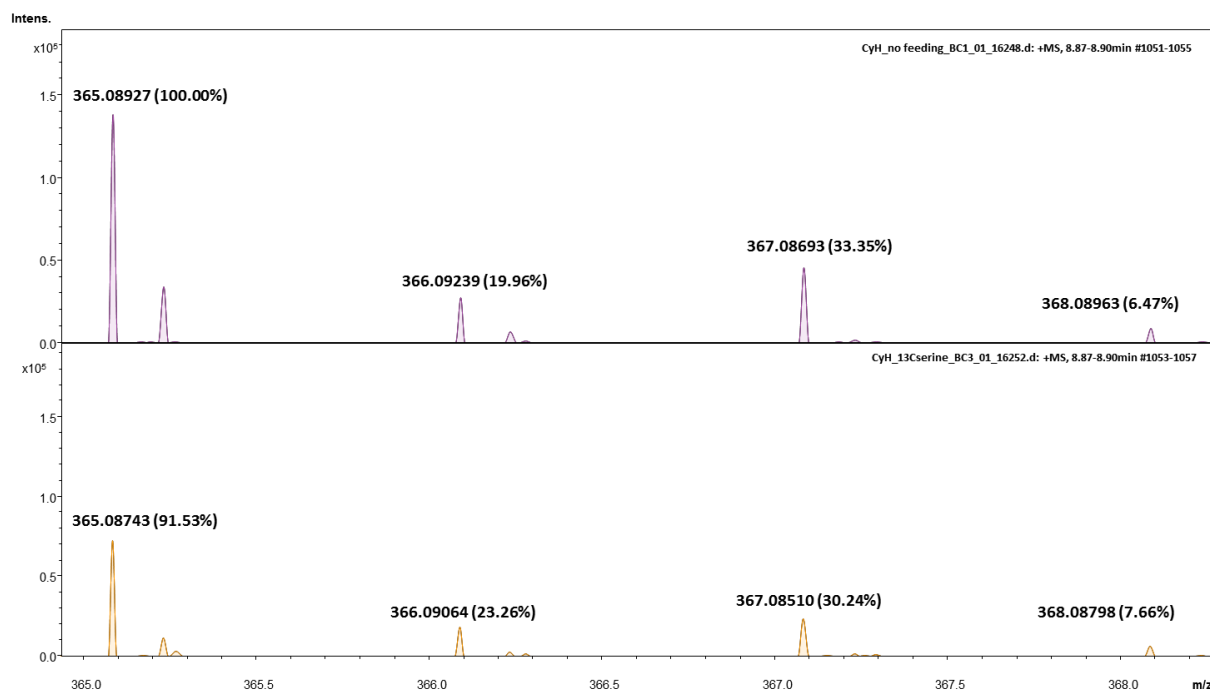
**Figure S1.** Partial ESI+MS spectra for pyrronazol A with L-methionine-(methyl- $^{13}\text{C}$ ) feeding (bottom) and culture broth without precursor feeding as control (top)

The isotopic patterns of the pyrronazol A show intensified peaks at  $m/z$  366 and 368 corresponding to a mass shift of +1 Da, caused by the incorporation of L-methionine-(methyl- $^{13}\text{C}$ ). Incorporation of the methyl- $^{13}\text{C}$  moiety was more efficient than L-proline- $^{15}\text{N}$  or L-serine-1- $^{13}\text{C}$  incorporation. The reason might be the comparatively specific consumption of the L-methionine methyl group in the metabolism of SAM-e cycle rather than in the primary metabolism.



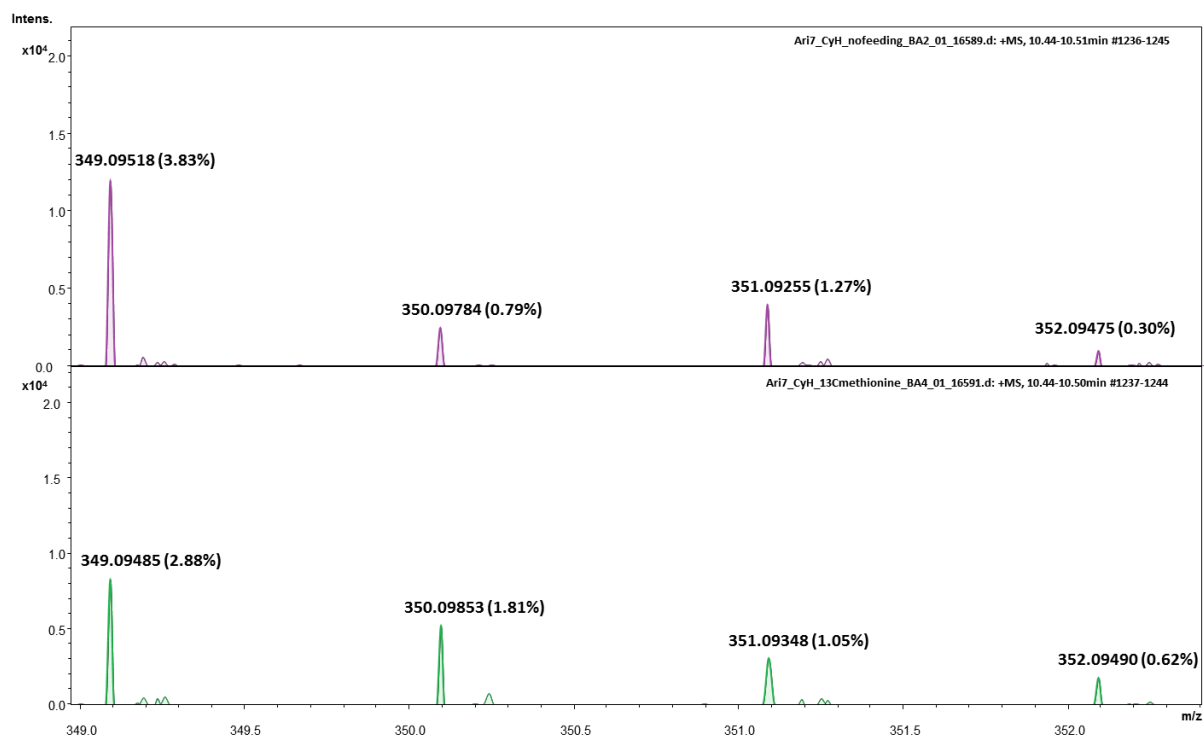
**Figure S2.** Partial ESI+MS spectra for pyrronazol A with L-proline-<sup>15</sup>N feeding (bottom) and culture broth without precursor feeding as control (top)

The isotopic patterns of the pyrronazol A show intensified peaks at m/z 366 and 368 corresponding to a mass shift of +1 Da, caused by the incorporation of L-proline-<sup>15</sup>N. Incorporation of L-proline-<sup>15</sup>N was less efficient than the incorporation of the L-methionine-(methyl-<sup>13</sup>C) methyl group due to the broader involvement in primary metabolism.



**Figure S3.** Partial ESI+MS spectra for pyrronazol A with L-serine-1-<sup>13</sup>C feeding (bottom) and culture broth without precursor feeding as control (top)

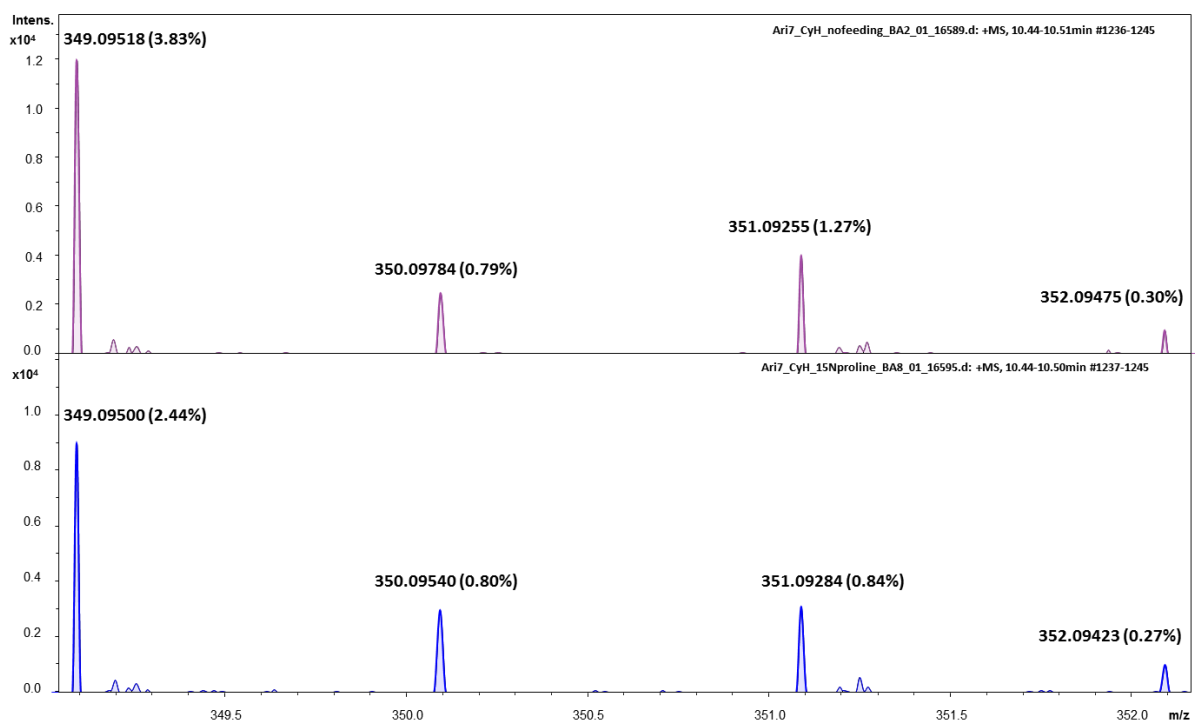
The isotopic patterns of the pyrronazol A show intensified peaks at m/z 366 and 368 corresponding to a mass shift of +1 Da, caused by the incorporation of L-serine-1-<sup>13</sup>C. The incorporation of L-serine-1-<sup>13</sup>C was very low but still detectable. The primary metabolism in *Nannocystis pusilla* Ari7 is probably consuming most of the supplemented L-serine-1-<sup>13</sup>C.



**Figure S4.** Partial ESI+MS spectra for pyrronazol B with L-methionine-(methyl-<sup>13</sup>C) feeding (bottom) and culture broth without precursor feeding as control (top)

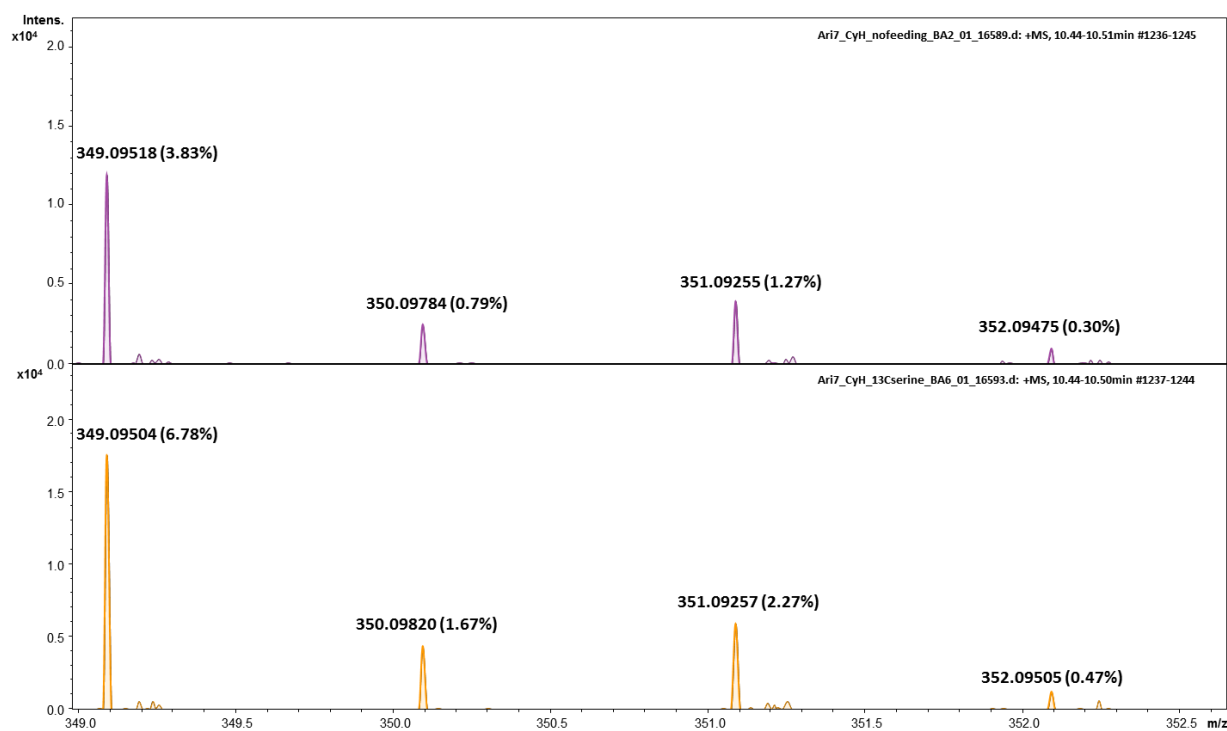
The isotopic patterns of the pyrronazol B show intensified peaks at m/z 350 and 352 corresponding to a mass shift of +1 Da, caused by the incorporation of L-methionine-(methyl-<sup>13</sup>C). Incorporation of the methyl-<sup>13</sup>C moiety was more efficient than L-proline-<sup>15</sup>N or L-serine-1-<sup>13</sup>C incorporation. The reason might be the comparatively specific consumption of the L-methionine methyl group in the metabolism of SAM-e cycle rather than in the primary metabolism.





**Figure S5.** Partial ESI+MS spectra for pyrronazol B with L-proline-<sup>15</sup>N feeding (bottom) and culture broth without precursor feeding as control (top)

The isotopic patterns of the pyrronazol B show intensified peaks at  $m/z$  350 and 352 corresponding to a mass shift of +1 Da, caused by the incorporation of L-proline-<sup>15</sup>N. Incorporation of L-proline-<sup>15</sup>N was less efficient than the incorporation of the L-methionine-(methyl-<sup>13</sup>C) methyl group due to the broader involvement in primary metabolism.



**Figure S6.** Partial ESI+MS spectra for pyrronazol B with L-serine-1-<sup>13</sup>C feeding (bottom) and culture broth without precursor feeding as control (top)

The isotopic patterns of the pyrronazol B show intensified peaks at m/z 350 corresponding to a mass shift of +1 Da, caused by the incorporation of L-serine-1-<sup>13</sup>C. The incorporation of L-serine-1-<sup>13</sup>C was very low but still detectable. The primary metabolism in *Nannocystis pusilla* Ari7 is probably consuming most of the supplemented L-serine-1-<sup>13</sup>C.

## 2. Genes encoded in the pyrronazol biosynthetic gene cluster

**Table S1:** Table of all open reading frames assigned to the putative pyrronazol biosynthetic gene cluster with proposed function and closest homologue according to blastp search in the nr (non-redundant protein sequences) protein database at NCBI.

name	length (bp)	proposed function	closest homologue	accession No.	identity (%)
<i>pynA</i>	726	Tailoring enzyme	Methyltransferase	SFE85624	93.0
<i>pynB</i>	1143	Proline oxidation	Proline dehydrogenase	SFE85639	96.1
<i>pynC</i>	1512	NRPS biosynthesis	Proline ligase	SFE85654	92.8
<i>pynD</i>	255	Starter unit tethering	ACP	SFE85666	96.4
<i>pynE</i>	1347	Halogenation	FADH <sub>2</sub> O <sub>2</sub> -dependent halogenase	SFE85682	97.5
<i>pynF</i>	2196	-	Nonfunctional PKS module	SFE85699	88.2
<i>pynG</i>	5541	PKS biosynthesis	MxaE and MxaD	SFE85712	86.2
<i>pynH</i>	4191	NRPS biosynthesis	EpoB	SFE85729	93.8
<i>pynI</i>	5850	PKS biosynthesis	PKS module	SFE85764	81.5
<i>pynJ</i>	4611	PKS biosynthesis	PKS module	SFE85775	77.7
<i>pynK</i>	768	Chain release	Thioesterase	SFE85792	92.6
<i>pynL</i>	1230	Hydroxylation	Cytochrome P450 enzyme	SFE85811	97.1
<i>orf1</i>	582	-	Hypothetical protein	SFF42573	71.7
<i>orf2</i>	585	-	Hypothetical protein	ORZ24869	36.4
<i>orf3</i>	837	-	Hypothetical protein	SFE85862	77.7
<i>orf4</i>	780	-	NAD(P)H dehydrogenase	SFE85879	90.3
<i>orf5</i>	903	-	Transcriptional regulator, LysR family	SFE85890	87.3
<i>orf6</i>	375	-	Enoyl-CoA hydratase	SFK55502	39.3
<i>orf7</i>	456	-	Polyketide cyclase	WP_056108953	72.6
<i>orf8</i>	642	-	Glutathione S-transferase family protein	WP_057205710	77.0
<i>orf9</i>	885	-	AraC family transcriptional regulator	WP_012241458	76.9
<i>orf10</i>	900	-	AraC-type DNA-binding protein	SFE85934	86.8

### 3. Stereochemistry prediction of the secondary alcohol moiety and methyl model during chain elongation in module 4

Profile hidden Markov models of 80 amino acid sequences of KR domains with known characteristics in PKS assembly line enables prediction of the stereochemistry of the secondary hydroxyl group product of the KR domain in module 4. Therefore the profile HMM online tool ([https://akitsche.shinyapps.io/profileHMM\\_App/](https://akitsche.shinyapps.io/profileHMM_App/) (accessed June 2017) was used.<sup>5</sup>

The query sequence used for stereochemistry prediction is the translation of the KR domain in module 4 on the *pynI* gene (FASTA format):

> **pynI KR**

```
ATYLITGGLGGLGLLAAGWLVDGRGARHLALLGRSAPGPAQAEAVAALTRRGARVELVRVDV  
ADLDALAAALHRLDAAMPPLRGILHAAGALDDGLIAGQSLERLRPVLAPKLHGACNLHRLTR  
ARPLDFFVLYSSVSAVLGAPGLANYVAANAALDALAHRRRLGLPATSINWGLF
```

*In silico* prediction score for the D-form of secondary alcohol is 42.65 while the score for the D-form for the methyl branch is 10.8. The D-configuration of the secondary alcohol and the methyl branch fits the structure experimentally determined by NMR as described previously.<sup>6</sup>

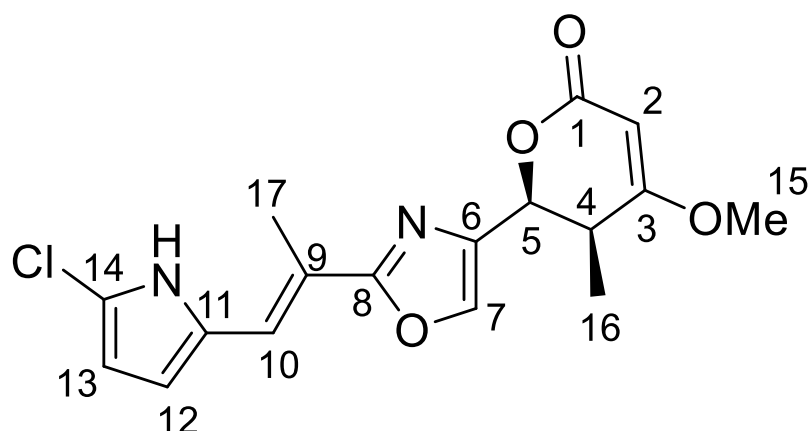
#### 4. *In silico* analysis of the A domain of PynH

The Stachelhaus code sequence of the A-domain of PynH is interestingly not assigned to L-serine but L-cysteine. Since the framework of the pyrronazols consists of an oxazole rather than thiazole the incorporated amino acid has to be L-serine rather than L-cysteine. However to confirm this, the Stachelhaus code sequence of several cysteine and serine incorporating myxobacterial A domains within oxazole/oxazoline or thiazole/thiazoline forming modules were compared to the A domain of PynH. The second and the fourth position in the Stachelhaus signature of the A domain of PynH is highly conserved in A domains incorporating L-cysteine whereas the third position is common for A domains incorporating L-serine. The Stachelhaus code sequence of the A domains involved in the ajudazol and the leupyrrin biosynthesis are also assigned incorrectly to L-cysteine by *in silico* analysis.

**Table S2:** Stachelhaus code sequence of different cysteine and serine incorporating myxobacterial A domains.

name/position	235	236	239	278	299	301	322	330	331	517
Epothilone	D	L	Y	N	L	S	L	I	W	K
Antalid	D	L	W	N	L	S	L	I	W	K
Vioprolid	D	L	Y	N	L	S	L	I	W	K
Tubulysin	D	L	Y	N	M	S	L	I	W	K
Thuggacin	D	L	W	N	L	S	L	I	W	K
Myxothiazol	D	L	Y	N	L or M	S	L	I	W	K
Melithiazol	D	L	Y	N	L or M	S	L	I	W	K
Lipothiazol	D	L	Y	N	L or I	S	L or S	I	W	K
Argyrimycin	D	L	Y	N	M	S	M	I	W	K
Althiomycin	D	L	Y	N	L	S	L	I	W	K
Disorazol	D	V	W	H	F	S	L	V	D	K
Rhizopodin	D	V	W	H	F	S	L	V	D	K
Ajudazol	D	L	Y	N	L	A	L	V	W	K
Leupyrrin	D	M	F	N	F	G	V	L	W	K
Pyrronazol	D	L	W	N	I	A	S	I	W	K

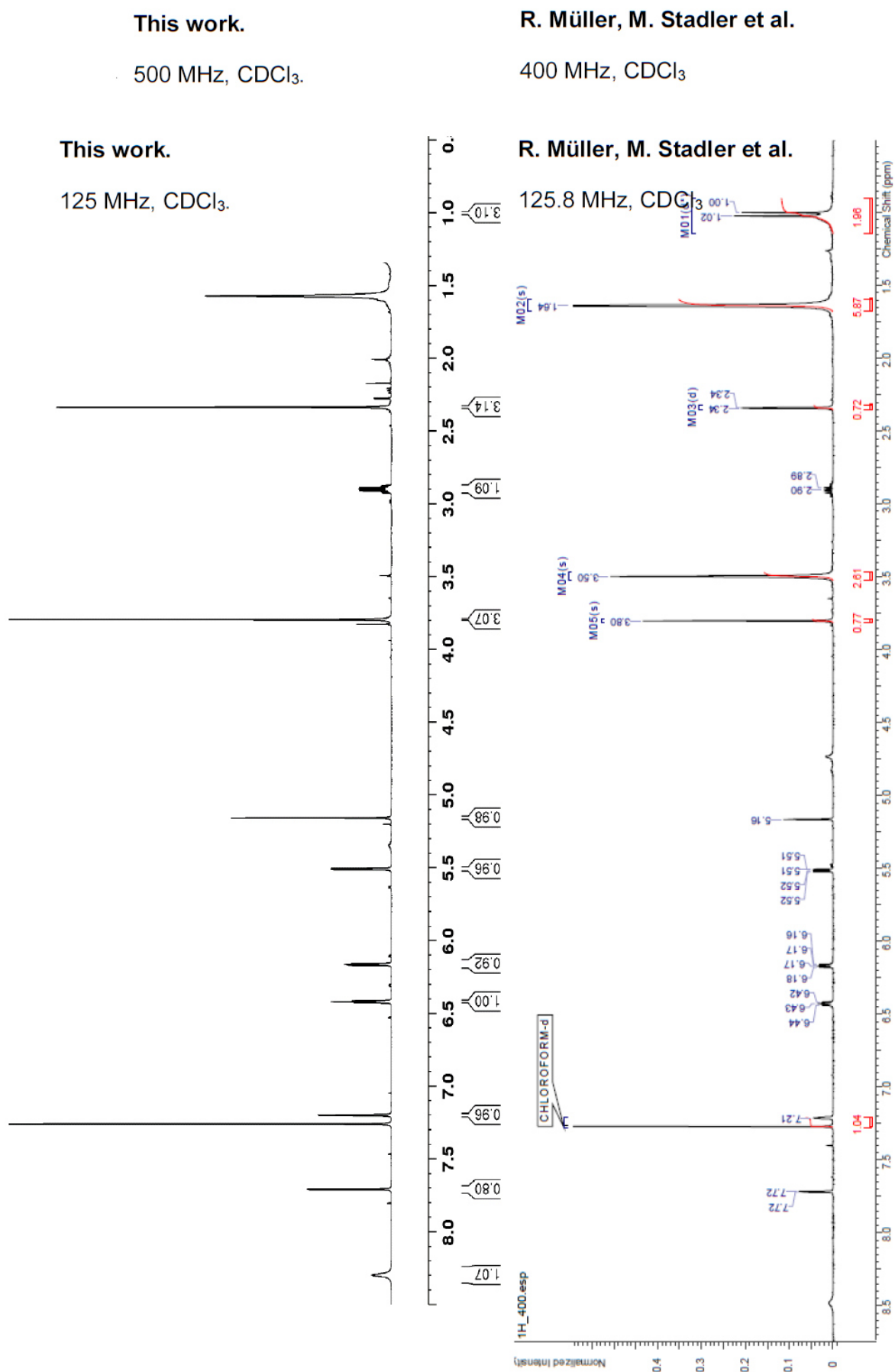
## 5. Comparison of NMR-spectroscopic data of authentic and synthetic Pyrronazol B in CDCl<sub>3</sub>

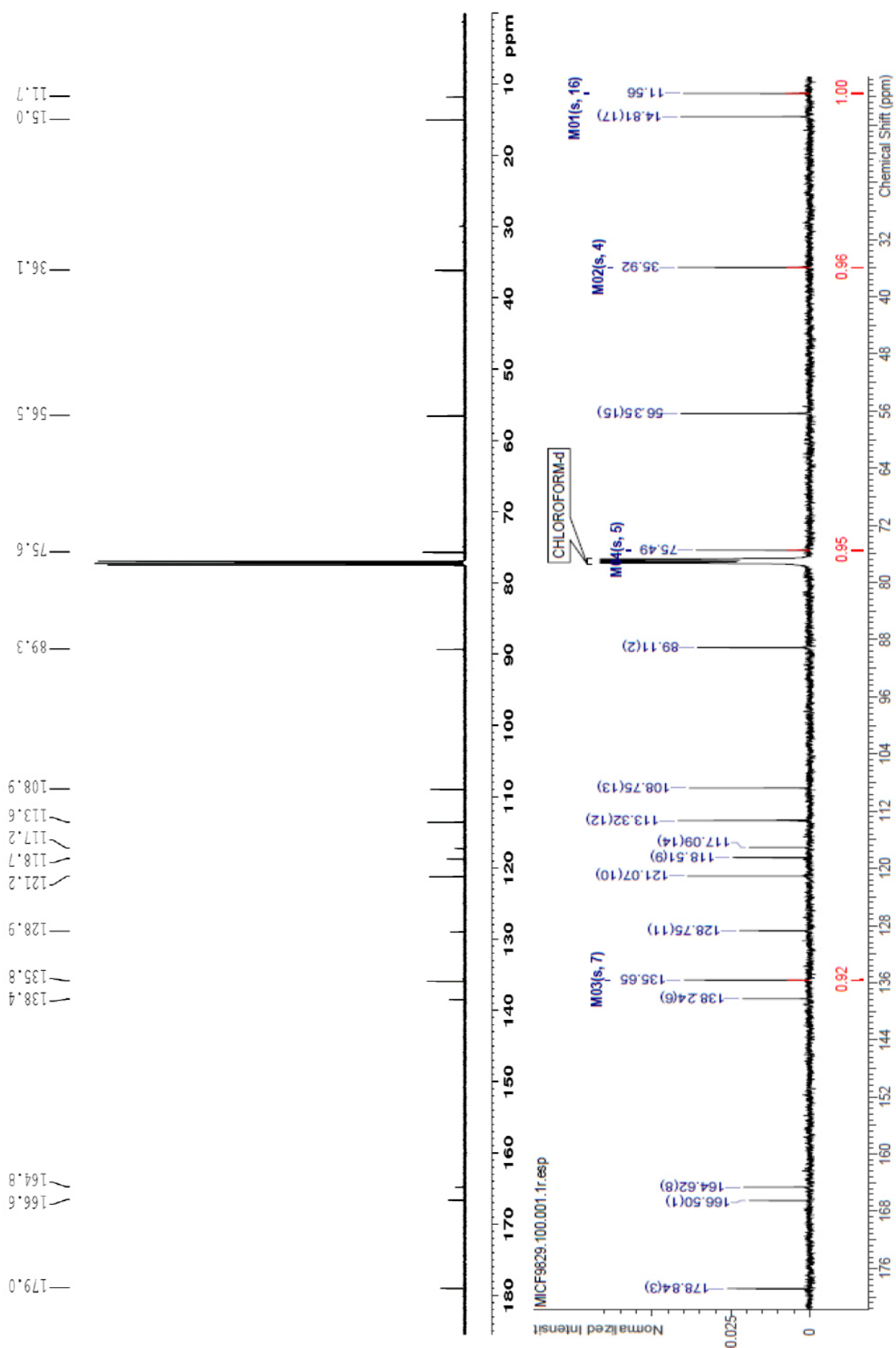


**Table S3:** NMR-spectroscopic data for authentic and synthetic Pyrronazol B.

no.	Authentic Pyrronazol B					Synthetic Pyrronazol B				
	<sup>13</sup> C	m <sup>b</sup>	<sup>1</sup> H <sup>a</sup>	m	J [HZ]	<sup>13</sup> C	m <sup>d</sup>	<sup>1</sup> H <sup>c</sup>	m	J [HZ]
1	166.4	C	-	-	-	166.6	C	-	-	-
2	89.1	CH	5.16	s	-	89.3	CH	5.16	s	-
3	178.8	C	-	-	-	179.0	C	-	-	-
4	35.9	CH	2.91	qd	7.0, 3.4	36.1	CH	2.90	m	-
5	75.4	CH	5.52	dd	3.4, 1.5	75.6	CH	5.51	q	1.6
6	138.2	C	-	-	-	138.4	C	-	-	-
7	135.7	CH	7.71	d	1.5	135.8	CH	7.71	d	1.4
8	164.7	C	-	-	-	164.8	C	-	-	-
9	118.5	C	-	-	-	118.7	C	-	-	-
10	121.1	CH	7.21	m	-	121.2	CH	7.20	s	-
11	128.7	C	-	-	-	128.9	C	-	-	-
12	113.3	CH	12	t	3.4	113.6	CH	6.42	t	3.4
13	108.8	CH	6.17	dd	3.8, 2.6	108.9	CH	6.16	q	2.1
14	117.1	C	-	-	-	117.2	C	-	-	-
15	56.4	CH <sub>3</sub>	3.80	s	-	56.5	CH <sub>3</sub>	3.79	s	-
16	11.6	CH <sub>3</sub>	1.01	d	7.2	11.7	CH <sub>3</sub>	1.00	d	7.2
17	14.8	CH <sub>3</sub>	2.34	d	1.1	15.0	CH <sub>3</sub>	2.34	d	1.1
NH	-	-	8.47	s	br	-	-	8.30	s	br

<sup>a</sup> <sup>1</sup>H 300 and 600 MHz. <sup>b</sup> <sup>13</sup>C (75 MHz) from DEPT and HMQC. <sup>c</sup> <sup>1</sup>H 500 MHz. <sup>d</sup> <sup>13</sup>C 125 MHz

Figure S7. Comparison of <sup>1</sup>H-NMR spectra.

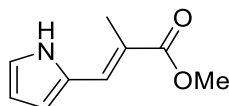
**Figure S8.** Comparison of  $^{13}\text{C}$ -NMR spectra.



## 6. Total synthesis of Pyrronazol B

### Methyl (*E*)-2-methyl-3-(1*H*-pyrrol-2-yl)acrylate (**5**)

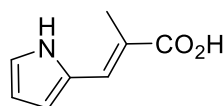
C/C(C(OC)=O)=C\C1=CC=CN1



Ester **5** was synthesized *via* a literature known procedure by H. McNab *et al.*<sup>7</sup> The spectroscopic data obtained are consistent with those reported in literature (McNab 1997)<sup>7</sup>.

### (*E*)-2-Methyl-3-(1*H*-pyrrol-2-yl)acrylic acid (**7**)

C/C(C(O)=O)=C\C1=CC=CN1



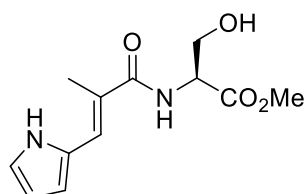
The following reaction was carried out under air. Ester **6** (4.63 g, 28.1 mmol, 1.0 eq.) was solved in a mixture of water/methanol (280 mL, 1:2). Potassium hydroxide (15.8 g, 281 mmol, 10 eq.) was added to this mixture at room temperature. The reaction mixture was stirred for 24 h. After completion of the reaction (monitored by TLC), it was quenched with 1M HCl until pH ~ 4 (monitored by pH indicator paper). The resulting solution was extracted with EtOAc and washed with brine. The organic layers were dried over anhydrous MgSO<sub>4</sub> and concentrated *in vacuo*. The resulting residue was used without further purification in the next reaction.

**TLC** R<sub>f</sub>=0.7 (CH<sub>2</sub>Cl<sub>2</sub>:MeOH = 9:1)

**HRMS** (ESI) Calcd for (C<sub>8</sub>H<sub>9</sub>NO<sub>2</sub> + Na<sup>+</sup>): 174.0531. Found: 174.0533.

### Methyl (*E*)-(2-methyl-3-(1*H*-pyrrol-2-yl)acryloyl)serinate (**4**)

C/C(C(N[C@H](C(OC)=O)CO)=O)=C\C1=CC=CN1



To a solution of crude **7** (4.00 g, 26.5 mmol, 1.0 eq.) in DMF (26 mL) L-serine methyl ester hydrochloride **B** (4.53 g, 29.1 mmol, 1.1 eq.) was added. The reaction mixture was cooled with an ice bath to 0 °C. PyBOP (15.2 g, 29.1 mmol, 1.1 eq.) and DIPEA (15.8 mL, 92.6 mmol, 3.5 eq.) were added and further cooled for 10 minutes. The ice bath was removed and the reaction was stirred for 24 h at room temperature. After completion of the reaction (monitored by TLC) it was quenched with aqueous NaHCO<sub>3</sub>, extracted with EtOAc, washed with 1M citric acid and brine. The organic layers were dried over anhydrous Na<sub>2</sub>SO<sub>4</sub> and concentrated *in vacuo*. The resulting residue was purified by flash column chromatography on silica gel (PE:EtOAc = 1:2) to afford Amide **4** (6.35 g, 25.1 mmol, 95%) as an orange colored solid.

**TLC** R<sub>f</sub> = 0.2 (PE: EtOAc = 1:2).

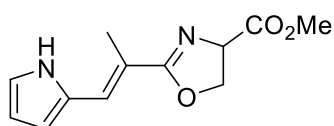
**<sup>1</sup>H NMR** (400 MHz, DMSO-d<sub>6</sub>) δ 11.18 (s, 1H), 7.82 (d, *J* = 7.4 Hz, 1H), 7.24 (s, 1H), 6.93 (m, *J* = 1.3 Hz, 1H), 6.41 (s, 1H), 6.20 (q, *J* = 2.7 Hz, 1H), 5.04 (t, *J* = 6.2 Hz, 1H), 4.42 (m, 1H), 3.75 (q, *J* = 5.4 Hz, 2H), 3.64 (s, 3H), 2.06 (d, *J* = 0.9 Hz, 3H).

**<sup>13</sup>C NMR** (100 MHz, DMSO-d<sub>6</sub>) δ 171.3, 169.0, 128.3, 124.9, 123.6, 120.4, 111.4, 109.9, 61.17, 55.6, 51.8, 14.5).

**HRMS** (ESI) Calcd for (C<sub>12</sub>H<sub>16</sub>N<sub>2</sub>O<sub>4</sub> + Na<sup>+</sup>): 275.1008. Found: 275.1010.

### Methyl (*E*)-2-(1-(1*H*-pyrrol-2-yl)prop-1-en-2-yl)-4,5-dihydrooxazole-4-carboxylate (**13**)

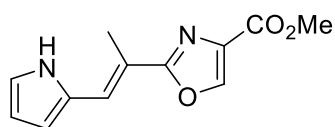
C/C(C1=NC(C(OC)=O)CO1)=C\C2=CC=CN2



Amide **4** (2.88 g, 11.4 mmol, 1.0 eq) was dissolved in CH<sub>2</sub>Cl<sub>2</sub> (265 mL) and cooled to -78 °C. DAST (1.50 mL, 11.4 mmol, 1.0 eq.) was added dropwise at -78 °C and stirred for 1 h. K<sub>2</sub>CO<sub>3</sub> (3.16 g, 22.9 mmol, 2.0 eq.) was added at -78 °C and stirred for another 20 minutes. The reaction mixture was warmed to 0 °C and stirred for 1 h. The ice bath was removed and the mixture was stirred at room temperature for 24 h. After completion of the reaction (monitored by TLC) it was quenched with aqueous NaHCO<sub>3</sub>, extracted with EtOAc and washed with brine. The organic layers were dried over anhydrous Na<sub>2</sub>SO<sub>4</sub> and concentrated *in vacuo*. The resulting residue was used in the next reaction without further purification.

**TLC** R<sub>f</sub> = 0.4 (PE:EtOAc = 1:2).

**HRMS** (ESI) Calcd for (C<sub>12</sub>H<sub>14</sub>N<sub>2</sub>O<sub>3</sub> + Na<sup>+</sup>): 257.0902. Found: 275.0902.

**Methyl (E)-2-(1-(1H-pyrrol-2-yl)prop-1-en-2-yl)oxazole-4-carboxylate (8)**
C/C(C1=NC(C(OC)=O)=CO1)=C\C2=CC=CN2


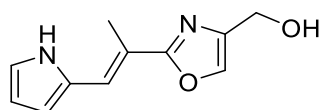
Crude **13** was dissolved in CH<sub>2</sub>Cl<sub>2</sub> (100 mL) and cooled to 0 °C. BrCCl<sub>3</sub> (2.2 mL, 22.8 mmol, 2.0 eq) and DBU (3.7 mL, 25.1 mmol, 2.2 eq) were added at 0 °C and the reaction was stirred for 1.5 h at 0 °C. After completion of the reaction (monitored by TLC) the reaction was quenched with NaHCO<sub>3</sub>, extracted with CH<sub>2</sub>Cl<sub>2</sub> and washed with brine. The organic layers were dried over anhydrous Na<sub>2</sub>SO<sub>4</sub> and concentrated *in vacuo*. The resulting residue was purified by flash column chromatography on silica gel (PE:EtOAc = 3:1 to 2:1) to afford **8** (1.23 g, 5.31 mmol, 46% over two steps) as a pale yellow solid.

**TLC** R<sub>f</sub> = 0.3 (PE:EtOAc = 2:1).

**<sup>1</sup>H NMR** (400 MHz, DMSO-d<sub>6</sub>) δ 11.29 (s, 1H), 8.80 (s, 1H), 7.44 (s, 1H), 7.00 (m, *J* = 1.3 Hz, 1H), 6.51 (s, 1H), 6.24 (q, *J* = 2.7 Hz, 1H), 3.82 (s, 3H), 2.26 (d, *J* = 1.0 Hz, 3H).

**<sup>13</sup>C NMR** (100 MHz, DMSO-d<sub>6</sub>) δ 164.4, 161.3, 144.9, 133.2, 128.4, 123.5, 120.9, 115.4, 111.7, 110.4, 51.7, 14.6.

**HRMS** (ESI) Calcd for (C<sub>12</sub>H<sub>12</sub>N<sub>2</sub>O<sub>3</sub> + Na<sup>+</sup>): 255.0746. Found: 255.0745.

**(E)-2-(1-(1H-Pyrrol-2-yl)prop-1-en-2-yl)oxazol-4-yl)methanol (14)**
C/C(C1=NC(CO)=CO1)=C\C2=CC=CN2


A solution of **8** (1.34 g, 5.77 mmol, 1.0 eq) in THF (50 mL) was cooled to 0 °C. LiAlH<sub>4</sub> (2.4 M in THF, 2.9 mL, 6.92 mmol, 1.2 eq) was added dropwise to the reaction mixture. The reaction mixture was stirred at 0 °C for 1 h. After completion of the reaction (monitored by TLC) Et<sub>2</sub>O (20 mL) was added at 0 °C. To this reaction 0.3 mL water and 0.3 mL 2 M NaOH were added. To this mixture another 1 mL Water was added and the reaction was allowed to warm to room temperature while stirring. Anhydrous MgSO<sub>4</sub> was added and the resulting residue was filtered and concentrated *in vacuo*. The crude product was purified by flash column chromatography (PE:EtOAc = 1:2) to afford **14** (800 mg, 3.94 mmol, 68%) as a bright yellow solid.

**TLC**  $R_f$  = 0.1 (PE:EtOAc = 1:2).

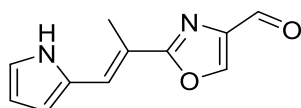
**$^1\text{H}$  NMR** (400 MHz, DMSO- $d_6$ )  $\delta$  11.19 (s, 1H), 7.83 (s, 1H), 7.34 (s, 1H), 6.95 (s, 1H), 6.44 (s, 1H), 6.21 (d,  $J$  = 2.8 Hz, 1H), 5.15 (t,  $J$  = 5.6 Hz, 1H), 4.37 (d,  $J$  = 5.6 Hz, 2H), 2.25 (s, 3H).

**$^{13}\text{C}$  NMR** (100 MHz, DMSO- $d_6$ )  $\delta$  163.5, 142.5, 134.9, 128.7, 121.7, 120.2, 116.7, 110.9, 110.1, 55.8, 14.7.

**HRMS** (ESI) Calcd for ( $\text{C}_{11}\text{H}_{12}\text{N}_2\text{O}_2 + \text{H}^+$ ): 205.0977. Found: 205.0976.

**(E)-2-(1-(1H-pyrrol-2-yl)prop-1-en-2-yl)oxazole-4-carbaldehyde (3)**

C/C(C1=NC(C=O)=CO1)=C\C2=CC=CN2



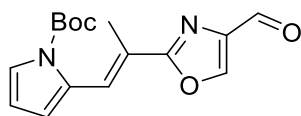
The following reaction was carried out under air in an open flask. Alcohol **14** (800 mg, 3.92 mmol, 1.0 eq) was dissolved in MeCN (39 mL). To this solution  $\text{Cu}(\text{MeCN})_4(\text{OTf})$  (73.8 mg, 0.20 mmol, 5 mol%), bpy (30.6 mg, 0.20 mmol, 5 mol%), TEMPO (30.6 mg, 0.20 mmol, 5 mol%) and 4-DMAP (47.9 mg, 0.39 mmol, 10 mol%) were added. The reaction flask was stirred at room temperature for 1.5 h. After completion of the reaction (monitored by TLC) the reaction was quenched with  $\text{NH}_4\text{Cl}$ , extracted with EtOAc and washed with brine. The organic layers were dried over anhydrous  $\text{Na}_2\text{SO}_4$  and concentrated *in vacuo*. The resulting residue was purified by flash column chromatography (PE:EtOAc = 4:1 to 3:1) to afford **3** (716 mg, 3.54 mmol, 90%) as a yellow solid.

**TLC**  $R_f$  = 0.5 ( $\text{CH}_2\text{Cl}_2$ :MeOH = 95:5).

**$^1\text{H}$  NMR** (400 MHz, DMSO- $d_6$ )  $\delta$  11.31 (s, 1H), 9.87 (s, 1H), 8.94 (s, 1H), 7.47 (s, 1H), 7.02 (m,  $J$  = 1.3 Hz, 1H), 6.53 (s, 1H), 6.25 (q,  $J$  = 2.7 Hz, 1H), 2.28 (d,  $J$  = 1.1 Hz, 3H).

**$^{13}\text{C}$  NMR** (100 MHz, DMSO- $d_6$ )  $\delta$  184.2, 164.9, 148.3, 141.3, 128.4, 124.0, 121.1, 115.2, 112.0, 110.5, 14.7.

**HRMS** (ESI) Calcd for ( $\text{C}_{11}\text{H}_{10}\text{N}_2\text{O}_2 + \text{Na}^+$ ): 225.0640. Found: 225.0641.

***tert*-Butyl (*E*)-2-(2-(4-formyloxazol-2-yl)prop-1-en-1-yl)-1*H*-pyrrole-1-carboxylate (**9**)**
C/C(C1=NC(C=O)=CO1)=C\C2=CC=CN2C(OC(C)(C)C)=O


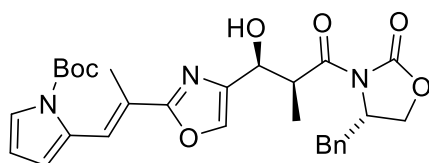
To a solution of **3** (716 mg, 3.54 mmol, 1.0 eq) in CH<sub>2</sub>Cl<sub>2</sub> (35 mL) were added 4-DMAP (433 mg, 3.54 mmol, 1.0 eq) and Boc<sub>2</sub>O (927 mg, 4.25 mmol, 1.2 eq) at room temperature. The reaction mixture was stirred at room temperature for 1 h. After completion of the reaction (monitored by TLC) it was quenched with water, extracted with CH<sub>2</sub>Cl<sub>2</sub> and washed with brine. The organic layers were dried over anhydrous Na<sub>2</sub>SO<sub>4</sub> and concentrated *in vacuo*. The resulting residue was purified by flash column chromatography (PE:EtOAc = 5:1) to afford **9** (830 mg, 2.74 mmol, 76%) as a yellow oil.

**TLC** R<sub>f</sub> = 0.4 (PE:EtOAc = 4:1).

**<sup>1</sup>H NMR** (400 MHz, DMSO) δ 9.89 (s, 1H), 8.99 (s, 1H), 8.06 (s, 1H), 7.44 (q, *J* = 1.6 Hz, 1H), 6.66 (d, *J* = 3.5 Hz, 1H), 6.39 (t, *J* = 3.4 Hz, 1H), 2.29 (d, *J* = 1.4 Hz, 3H), 1.57 (s, 9H).

**<sup>13</sup>C NMR** (101 MHz, DMSO) δ 184.3, 164.3, 148.7, 148.6, 141.3, 129.8, 123.8, 123.6, 120.3, 117.4, 111.7, 84.6, 27.5, 14.9.

**HRMS** (ESI) Calcd for (C<sub>16</sub>H<sub>18</sub>N<sub>2</sub>O<sub>4</sub> + Na<sup>+</sup>): 325.1164. Found: 325.1161.

***tert*-Butyl-2-((*E*)-2-(4-((1*S*,2*S*)-3-((*S*)-4-benzyl-2-oxooxazolidin-3-yl)-1-hydroxy-2-methyl-3-oxopropyl)oxazol-2-yl)prop-1-en-1-yl)-1*H*-pyrrole-1-carboxylate (**15**)**
C/C(C1=NC([C@H]([C@H](C)C(N2[C@@H](CC3=CC=CC=C3)COC2=O)=O)O)=CO1)=C\C4=C(C=CN4C(OC(C)(C)C)=O


Evans auxiliary **C** (280 mg, 0.93 mmol, 1.1 eq) was dissolved in CH<sub>2</sub>Cl<sub>2</sub> (5 mL) and cooled to 0 °C. To this solution was added a solution of TiCl<sub>4</sub> in CH<sub>2</sub>Cl<sub>2</sub> (1 M, 0.92 mL, 0.92 mmol, 1.0 eq) and DIPEA (0.4 mL, 2.3 mmol, 2.5 eq). The resulting mixture was stirred for 30 min at 0 °C and the cooled to -78 °C. Aldehyde **9** was solved in CH<sub>2</sub>Cl<sub>2</sub> (0.5 mL) and added dropwise to the reaction mixture. The reaction was stirred for 3 h at -78 °C. After completion of the reaction (monitored by TLC) it was quenched with pH 7 buffer at -78 °C, slowly warmed to room temperature, extracted with CH<sub>2</sub>Cl<sub>2</sub> and washed with

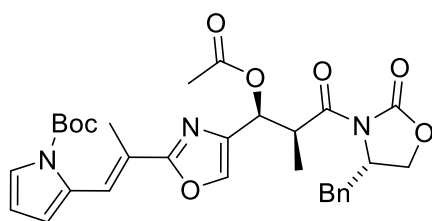
brine. The organic layers were dried over anhydrous  $\text{Na}_2\text{SO}_4$  and concentrated *in vacuo*. The resulting residue was purified by flash column chromatography (PE:EtOAc = 2:1) to afford a mixture of the Evans product **15** and non-Evans product in a 4:1 ratio, which could not be separated by chromatography, as a yellow oil.

**TLC**  $R_f$  = 0.2 (PE:EtOAc = 2:1).

**HRMS** (ESI) Calcd for  $(\text{C}_{29}\text{H}_{33}\text{N}_3\text{O}_7 + \text{Na}^+)$ : 558.2216. Found: 558.2211.

**tert-Butyl-2-((E)-2-(4-((1S,2S)-1-acetoxy-3-((S)-4-benzyl-2-oxooxazolidin-3-yl)-2-methyl-3-oxopropyl)oxazol-2-yl)prop-1-en-1-yl)-1H-pyrrole-1-carboxylate (10)**

C/C(C1=NC([C@H]([C@H](C)C(N2[C@@H](CC3=CC=CC=C3)COC2=O)=O)OC(C)=O)=CO1)=C\C4=CC=CN4C(OC(C)(C)C)=O



To a solution of **15** in  $\text{CH}_2\text{Cl}_2$  (10 mL) were added 4-DMAP (102 mg, 0.47 mmol, 0.5 eq.) and  $\text{Ac}_2\text{O}$  (0.18 mL, 1.87 mmol, 2.0 eq) at room temperature. The reaction mixture was stirred for 30 min at room temperature. After completion of the reaction (monitored by TLC) it was quenched with methanol and concentrated *in vacuo*. The residue was purified by flash column chromatography (PE:Acetone = 4:1 +2.5% MeOH) to afford **10** (314 mg, 0.54 mmol, 58% over two steps) as a yellow oil.

**TLC**  $R_f$  = 0.3 (PE:Acetone = 4:1 +2.5% MeOH).

**$^1\text{H}$  NMR** (400 MHz,  $\text{CDCl}_3$ )  $\delta$  7.92 (s, 1H), 7.59 (s, 1H), 7.32 (q,  $J$  = 5.3 Hz, 4H), 7.19 (d,  $J$  = 6.9 Hz, 2H), 6.41 (s, 1H), 6.25 (t,  $J$  = 3.3 Hz, 1H), 6.14 (d,  $J$  = 7.9 Hz, 1H), 4.58 (m,  $J$  = 4.7 Hz, 2H), 4.16 (m, 2H), 3.24 (q,  $J$  = 5.5 Hz, 1H), 2.78 (q,  $J$  = 7.6 Hz, 1H), 2.27 (s, 3H), 2.11 (s, 3H), 1.59 (s, 9H), 1.36 (d,  $J$  = 6.9 Hz, 3H).

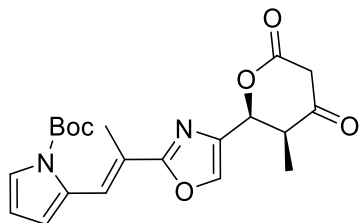
**$^{13}\text{C}$  NMR** (100 MHz,  $\text{CDCl}_3$ )  $\delta$  174.1, 170.6, 164.0, 153.2, 149.6, 139.3, 136.8, 135.4, 131.1, 129.6, 129.1, 127.5, 123.01, 122.1, 116.7, 111.3, 84.5, 69.2, 66.4, 55.6, 40.9, 38.1, 28.2, 21.2, 15.3, 13.6.

**HRMS** (ESI) Calcd for  $(\text{C}_{31}\text{H}_{35}\text{N}_3\text{O}_8 + \text{Na}^+)$ : 600.2322. Found: 600.2323.

$[\alpha]_D^{25}$  = 32.4 ( $c$  = 9.9 mg/ $\text{cm}^3$ , MeOH)

***tert*-Butyl-2-((*E*)-2-(4-((2*S*,3*S*)-3-methyl-4,6-dioxotetrahydro-2H-pyran-2-yl)oxazol-2-yl)prop-1-en-1-yl)-1H-pyrrole-1-carboxylate (**16**)**

C/C(C1=NC([C@@H]2[C@H](C)C(CC(O2)=O)=O)=CO1)=C\C3=CC=CN3C(OC(C)(C)C)=O



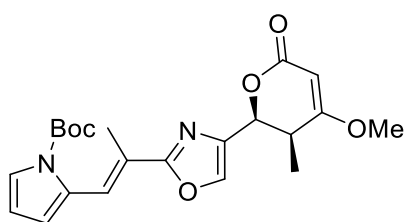
Acetate **10** (314 mg, 0.54 mmol, 1.0 eq) was solved in THF (3 mL) and cooled to  $-78^{\circ}\text{C}$ . LiHMDS in THF (1 M, 1.62 mL, 1.62 mmol, 3.0 eq) was added dropwise at  $-78^{\circ}\text{C}$ . The reaction mixture was stirred for 1.5 h at  $-78^{\circ}\text{C}$ . After completion of the reaction (monitored by TLC) it was quenched with  $\text{NH}_4\text{Cl}$ , slowly warmed to room temperature, extracted with EtOAc and washed with brine. The organic layers were dried over anhydrous  $\text{Na}_2\text{SO}_4$  and concentrated *in vacuo*. The resulting residue was used without further purification in the next reaction.

TLC  $R_f = 0.5$  (PE:Acetone = 1:1 + 1% MeOH)

HRMS (ESI) Calcd for  $(\text{C}_{21}\text{H}_{24}\text{N}_2\text{O}_6 + \text{Na}^+)$ : 423.1532. Found: 423.1532.

***tert*-Butyl-2-((*E*)-2-(4-((2*S*,3*S*)-4-methoxy-3-methyl-6-oxo-3,6-dihydro-2H-pyran-2-yl)oxazol-2-yl)prop-1-en-1-yl)-1H-pyrrole-1-carboxylate (**11**)**

C/C(C1=NC([C@@H]2[C@H](C)C(OC)=CC(O2)=O)=CO1)=C\C3=CC=CN3C(OC(C)(C)C)=O



Ketone **16** was solved in dry acetone (3 mL).  $\text{K}_2\text{CO}_3$  (82 mg, 0.59 mmol, 1.1 eq) and  $\text{Me}_2\text{SO}_4$  (68  $\mu\text{L}$ , 0.54 mmol, 1.0 eq) were added at room temperature. The reaction mixture was warmed to  $50^{\circ}\text{C}$  and stirred for 30 min. After completion of the reaction (monitored by TLC) it was allowed to cool down to room temperature, quenched with  $\text{NaHCO}_3$ , extracted with  $\text{CH}_2\text{Cl}_2$  and washed with brine. The organic layers were dried over anhydrous  $\text{Na}_2\text{SO}_4$  and concentrated *in vacuo*. The resulting residue was purified by flash column chromatography (PE:Acetone = 7:2 + 1% MeOH) to afford **11** (83 mg, 0.23 mmol, 43% over two steps) as a pale yellow oil.

**TLC**  $R_f$  = 0.3 (PE:Acetone = 7:2 + 1% MeOH).

**$^1\text{H}$  NMR** (400 MHz,  $\text{C}_6\text{D}_6$ )  $\delta$  8.39 (s, 1H), 7.44 (d,  $J$  = 1.5 Hz, 1H), 7.40 (q,  $J$  = 1.6 Hz, 1H), 6.31 (t,  $J$  = 1.7 Hz, 1H), 6.10 (t,  $J$  = 3.3 Hz, 1H), 5.26 (q,  $J$  = 1.7 Hz, 1H), 4.90 (s, 1H), 2.81 (s, 3H), 2.73 (m, 7.1 Hz, 1H), 2.29 (d,  $J$  = 1.4 Hz, 3H), 1.28 (s, 9H), 0.87 (d,  $J$  = 7.1 Hz, 3H).

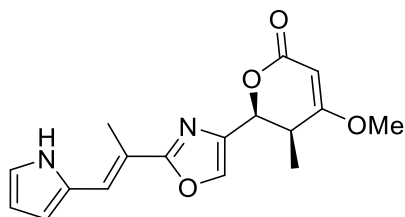
**$^{13}\text{C}$  NMR** (100 MHz,  $\text{C}_6\text{D}_6$ )  $\delta$  177.5, 164.8, 139.4, 136.1, 131.4.

**HRMS** (ESI) Calcd for ( $\text{C}_{22}\text{H}_{26}\text{N}_2\text{O}_6 + \text{Na}^+$ ): 437.1689. Found: 437.1689.

$[\alpha]_D^{27} = -103.0$  ( $c$  = 10.0  $\text{mg}/\text{cm}^3$ , MeOH)

**(5*S*,6*S*)-6-(2-((*E*)-1-(1*H*-pyrrol-2-yl)prop-1-en-2-yl)oxazol-4-yl)-4-methoxy-5-methyl-5,6-dihydro-2*H*-pyran-2-one (12)**

C/C(C1=NC([C@@H]2[C@H](C)C(OC)=CC(O2)=O)=CO1)=C\C3=CC=CN3



Compound **11** (40 mg, 97  $\mu\text{mol}$ , 1.0 eq) was suspended in  $\text{H}_2\text{O}$  (2 mL) and heated to 100  $^\circ\text{C}$ . The reaction was stirred for 1 h at 100  $^\circ\text{C}$ . After completion of the reaction (monitored by TLC) it was allowed to cool down to room temperature and extracted with EtOAc. The organic layers were dried over anhydrous  $\text{Na}_2\text{SO}_4$  and concentrated *in vacuo*. The resulting residue was purified by HPLC to afford **12** (17 mg, 54  $\mu\text{mol}$ , 56%) as a pale green oil.

**TLC**  $R_f$  = 0.4 (PE:Acetone 2:1 +1% MeOH)

**HPLC**  $R_t$  = 17 min.

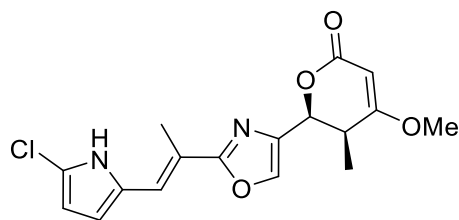
**$^1\text{H}$  NMR** (400 MHz,  $\text{C}_6\text{D}_6$ )  $\delta$  7.65 (s, 1H), 7.53 (d,  $J$  = 1.5 Hz, 1H), 7.32 (s, 1H), 6.49 (s, 1H), 6.46 (m,  $J$  = 1.3 Hz, 1H), 6.30 (q,  $J$  = 2.8 Hz, 1H), 5.31 (q,  $J$  = 1.7 Hz, 1H), 4.91 (s, 1H), 2.81 (s, 3H), 2.78 (m, 1H), 2.29 (d,  $J$  = 1.1 Hz, 3H), 0.91 (d,  $J$  = 7.1 Hz, 3H).

**$^{13}\text{C}$  NMR** (100 MHz,  $\text{C}_6\text{D}_6$ )  $\delta$  177.7, 165.4, 165.2, 139.3, 135.7, 129.5, 122.7, 120.2, 118.1, 112.6, 111.0, 89.9, 75.4, 55.3, 36.2, 15.0, 11.6.

**HRMS** (ESI) Calcd for ( $\text{C}_{17}\text{H}_{18}\text{N}_2\text{O}_4 + \text{Na}^+$ ): 337.1164. Found: 337.1164.

$[\alpha]_D^{27} = -136.9$  ( $c$  = 9.0  $\text{mg}/\text{cm}^3$ , MeOH)



**Pyrronazol B (2)**
C/C(C1=NC([C@@H]2[C@H](C)C(OC)=CC(O2)=O)=C\C3=CC=C(Cl)N3


A solution of **11** (12 mg, 38  $\mu$ mol, 1.0 eq) was solved in THF (0.3 mL). To this solution NCS (5.1 mg, 0.04 mmol, 1.0 eq) was added at room temperature. The reaction was stirred at room temperature for 24h. After completion of the reaction (monitored by TLC) it was quenched with water, extracted with EtOAc and washed with brine. The organic layers were dried over anhydrous  $\text{Na}_2\text{SO}_4$  and concentrated *in vacuo*. The resulting residue was purified by HPLC to afford **2** (3.1 mg, 8.8  $\mu$ mol, 23%) as a pale white solid.

**TLC**  $R_f$  = 0.3 (PE:Acetone 2:1 +1 % MeOH)

**HPLC**  $R_t$ : 19 min

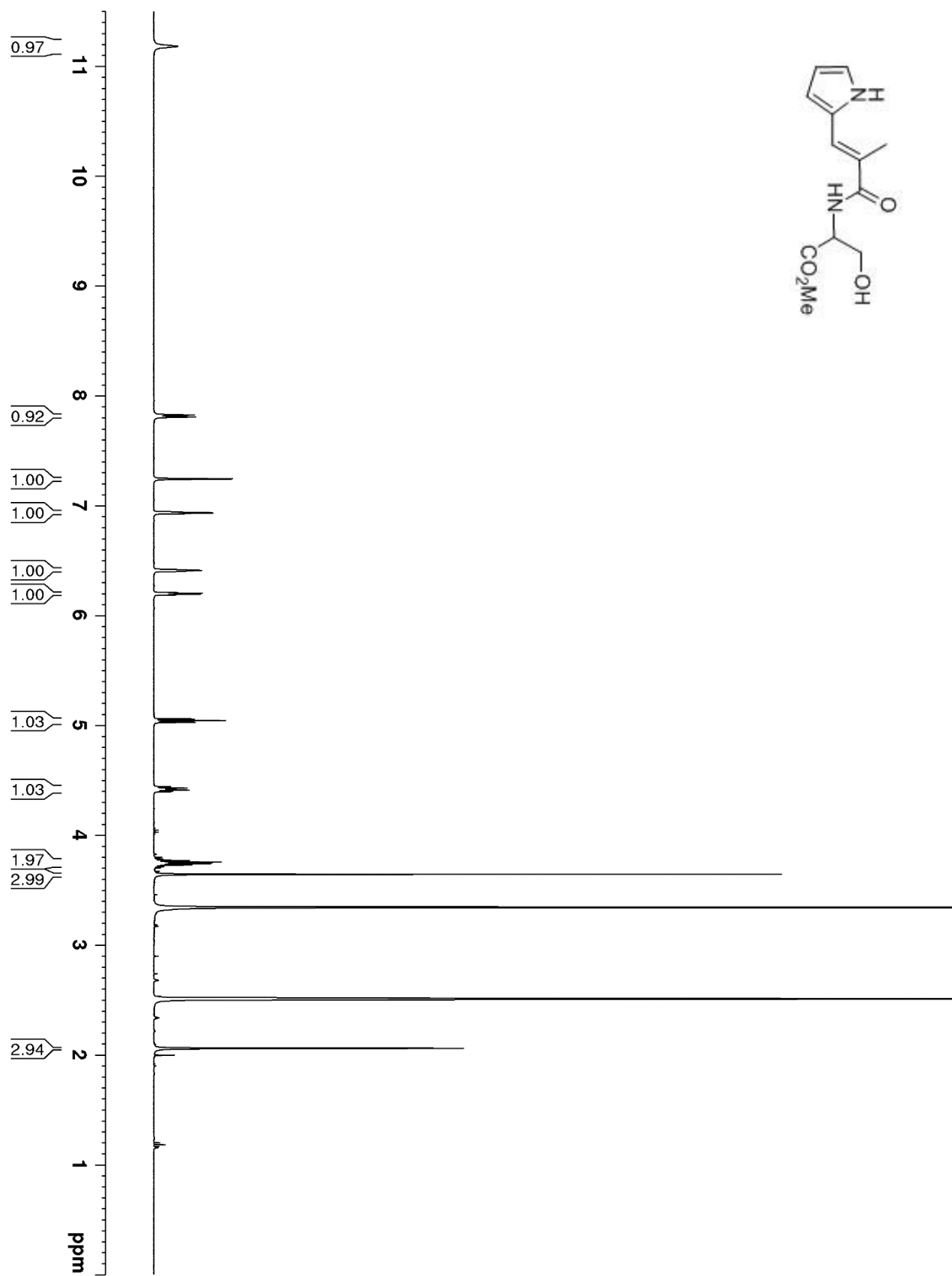
**$^1\text{H}$  NMR** (500 MHz,  $\text{CDCl}_3$ )  $\delta$  8.30 (s, 1H), 7.71 (d,  $J$  = 1.4 Hz, 1H), 7.20 (s, 1H), 6.42 (t,  $J$  = 3.4 Hz, 1H), 6.16 (q,  $J$  = 2.1 Hz, 1H), 5.51 (q,  $J$  = 1.6 Hz, 1H), 5.16 (s, 1H), 3.79 (s, 3H), 2.90 (m, 7.1 Hz, 1H), 2.34 (d,  $J$  = 1.1 Hz, 3H), 1.00 (d,  $J$  = 7.2 Hz, 3H).

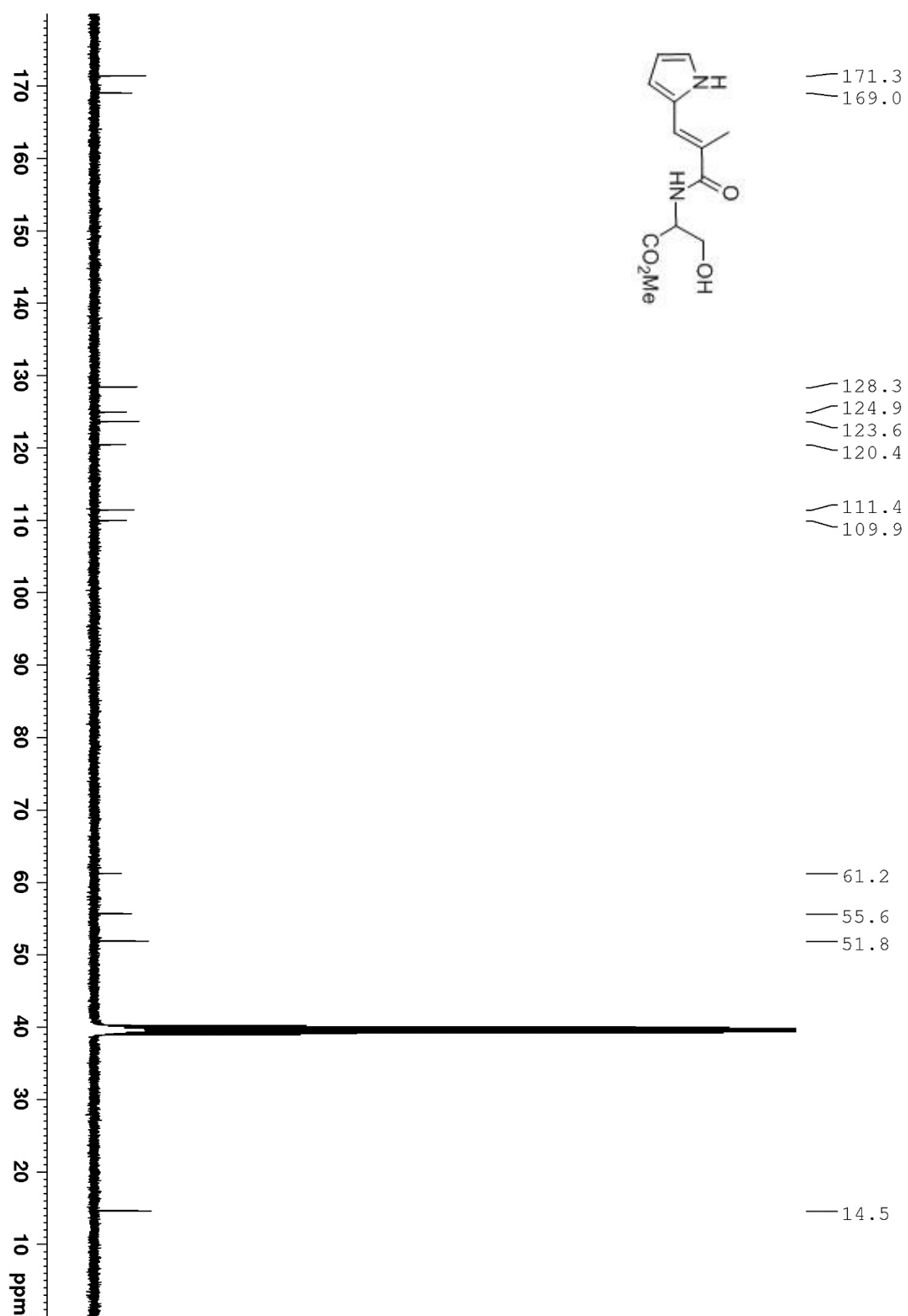
**$^{13}\text{C}$  NMR** (126 MHz, MeOD)  $\delta$  179.0, 166.6, 164.8, 138.4, 135.8, 128.9, 121.2, 118.7, 117.2, 113.6, 108.9, 89.3, 75.6, 56.5, 36.1, 15.0, 11.7.

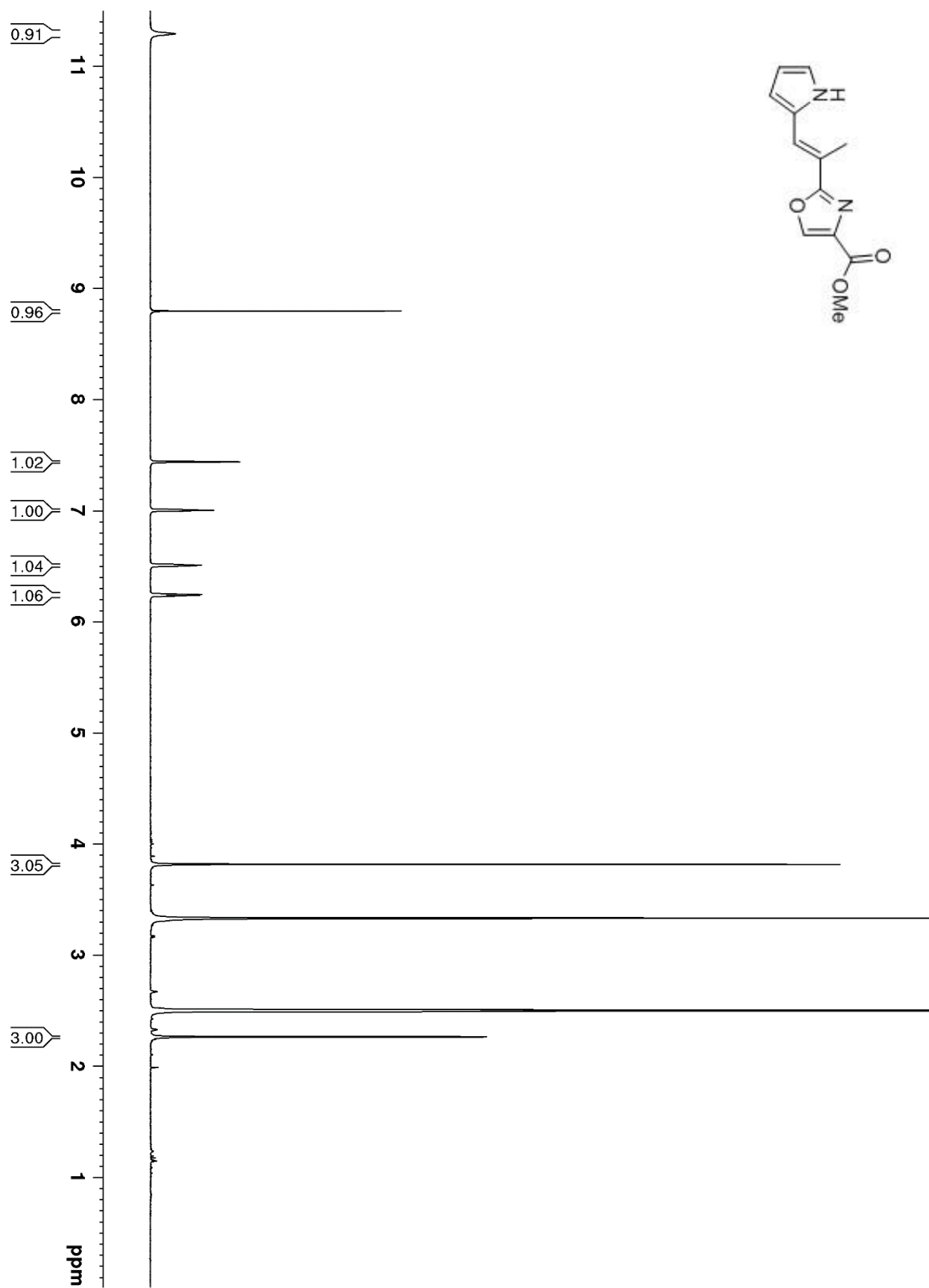
**HRMS** (ESI) Calcd for ( $\text{C}_{17}\text{H}_{18}\text{N}_2\text{O}_4\text{Cl} + \text{Na}^+$ ): 349.0955. Found: 349.0957

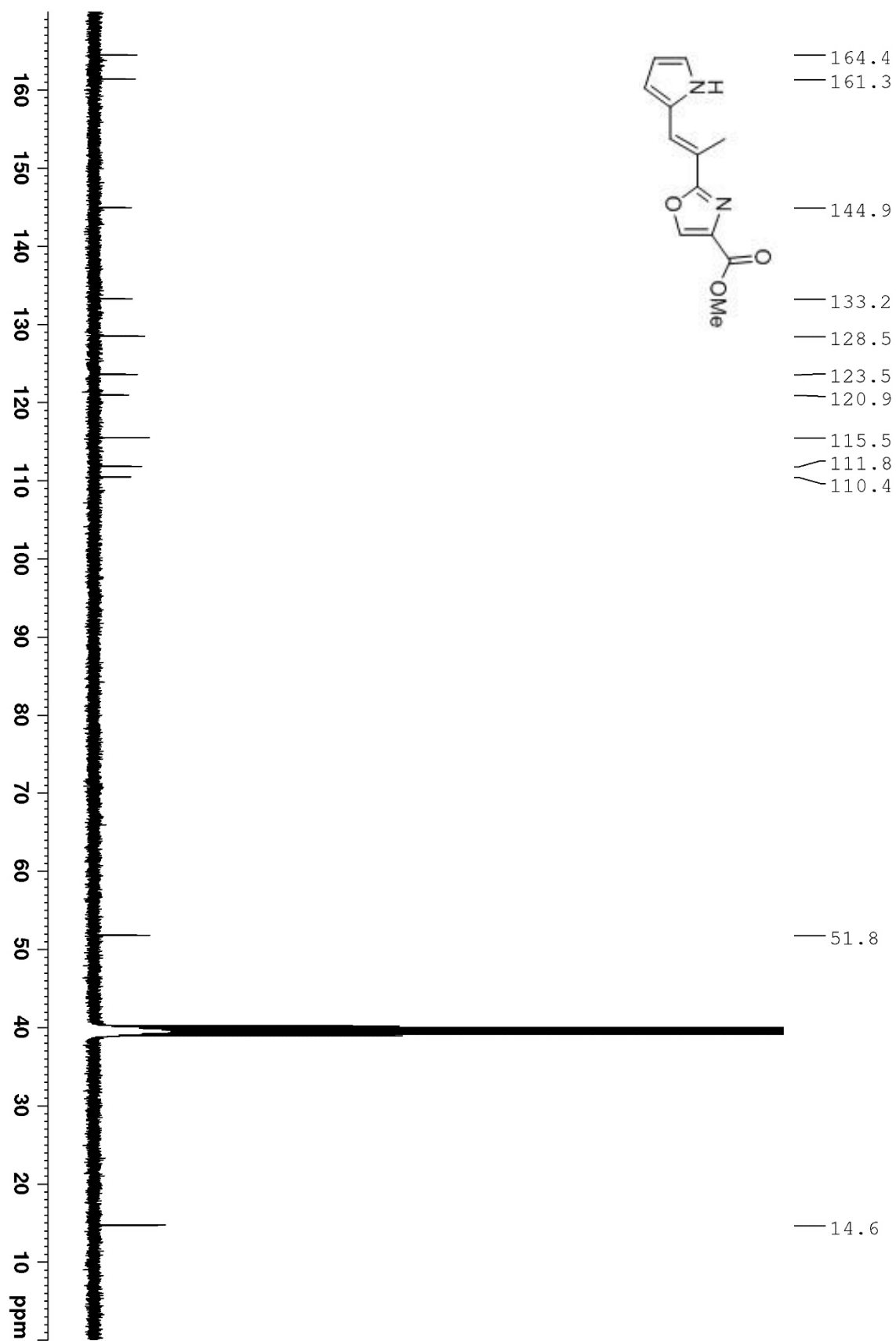
$[\alpha]_D^{27}$  = -125.2 ( $c$  = 1.6  $\text{mg}/\text{cm}^3$ , MeOH)

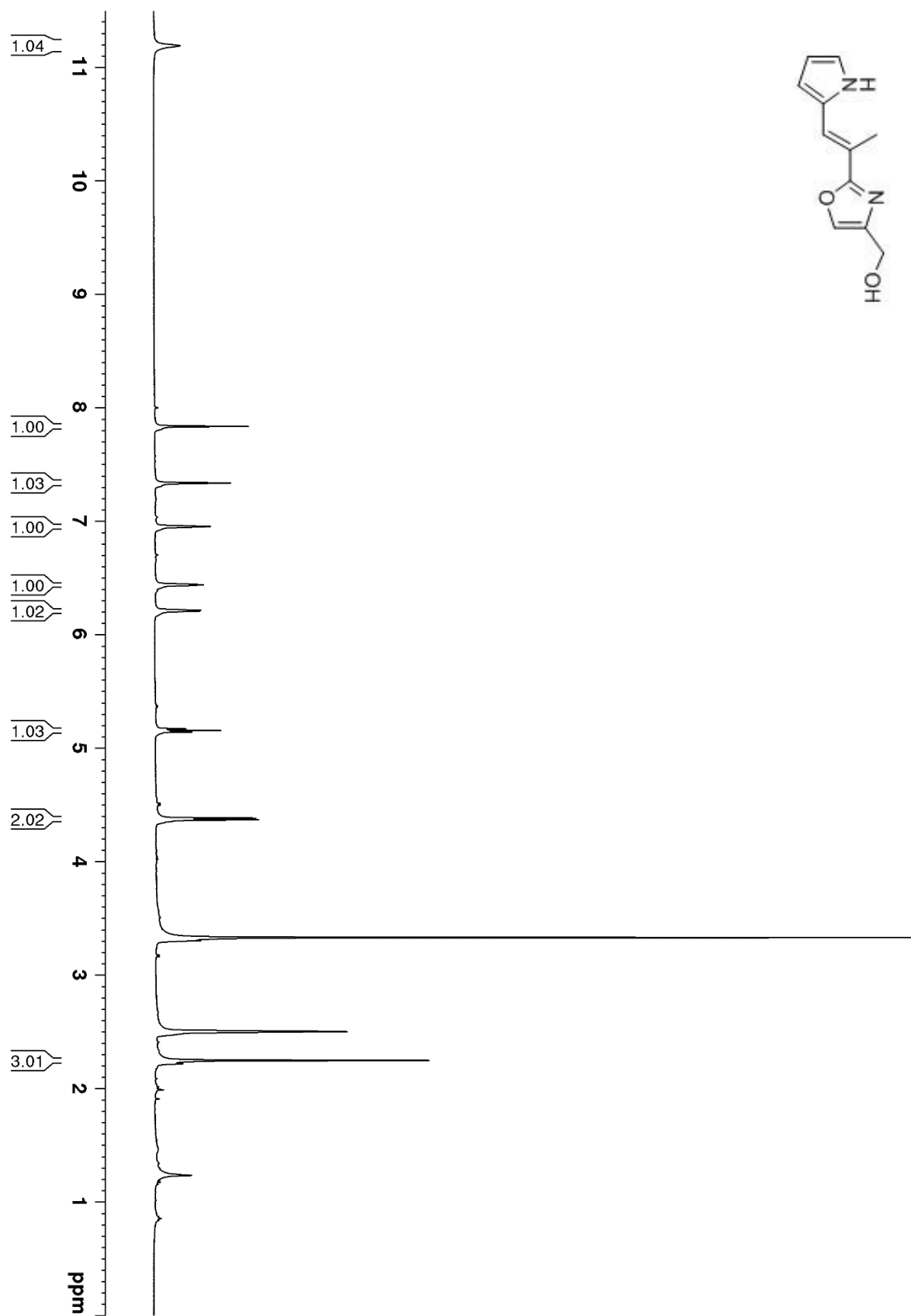
## **7. Copies of $^1\text{H}$ and $^{13}\text{C}$ NMR spectra**

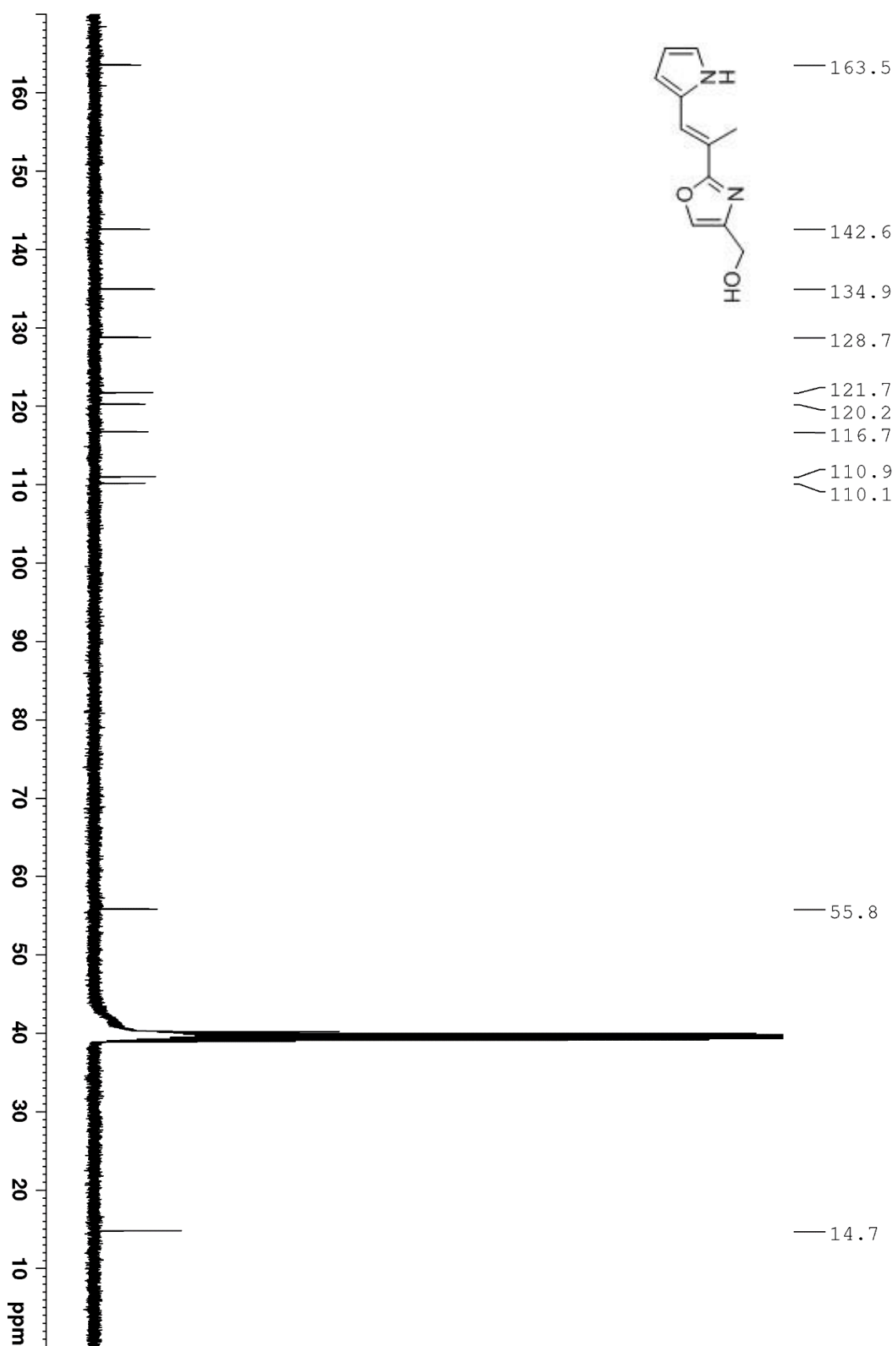




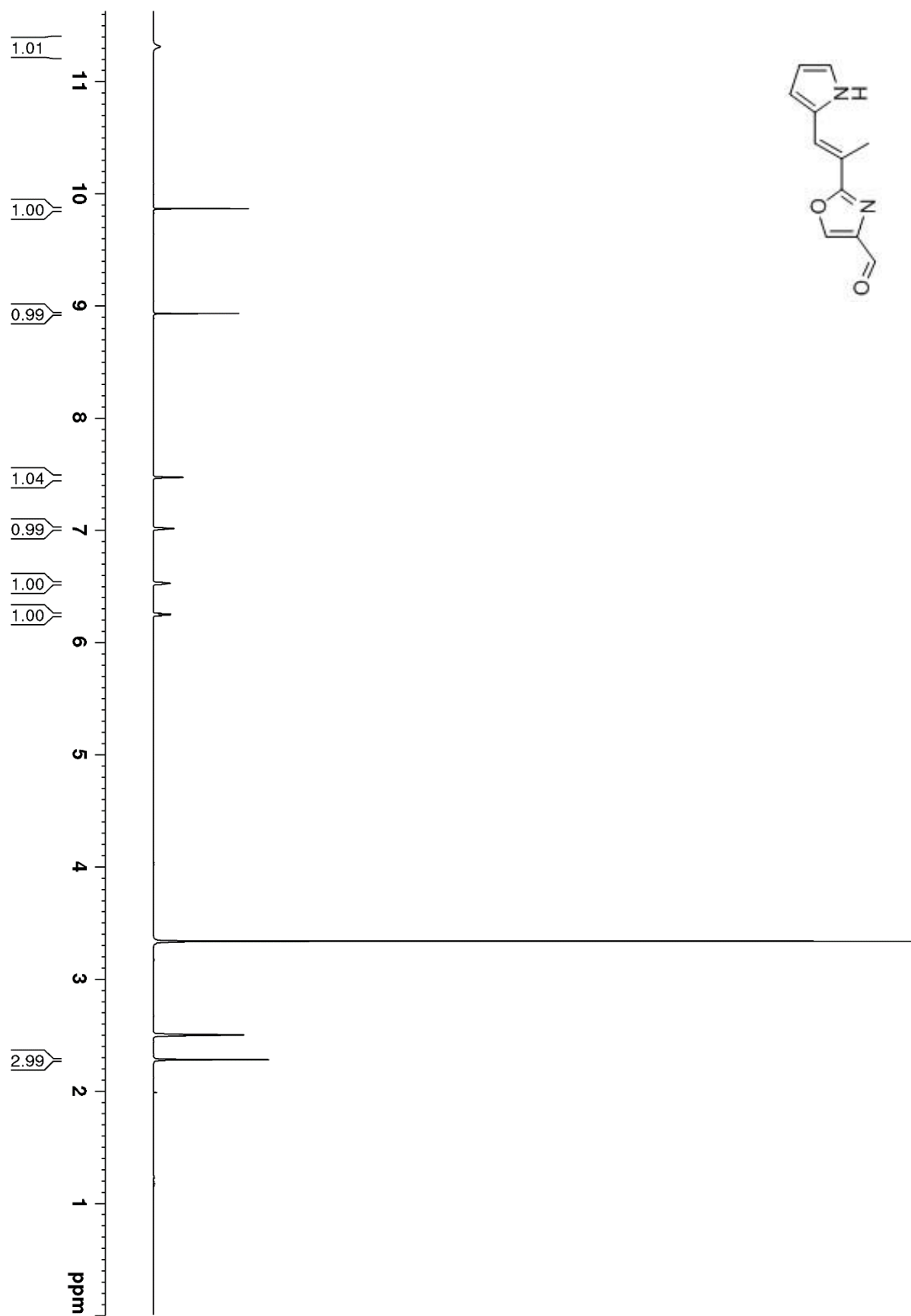


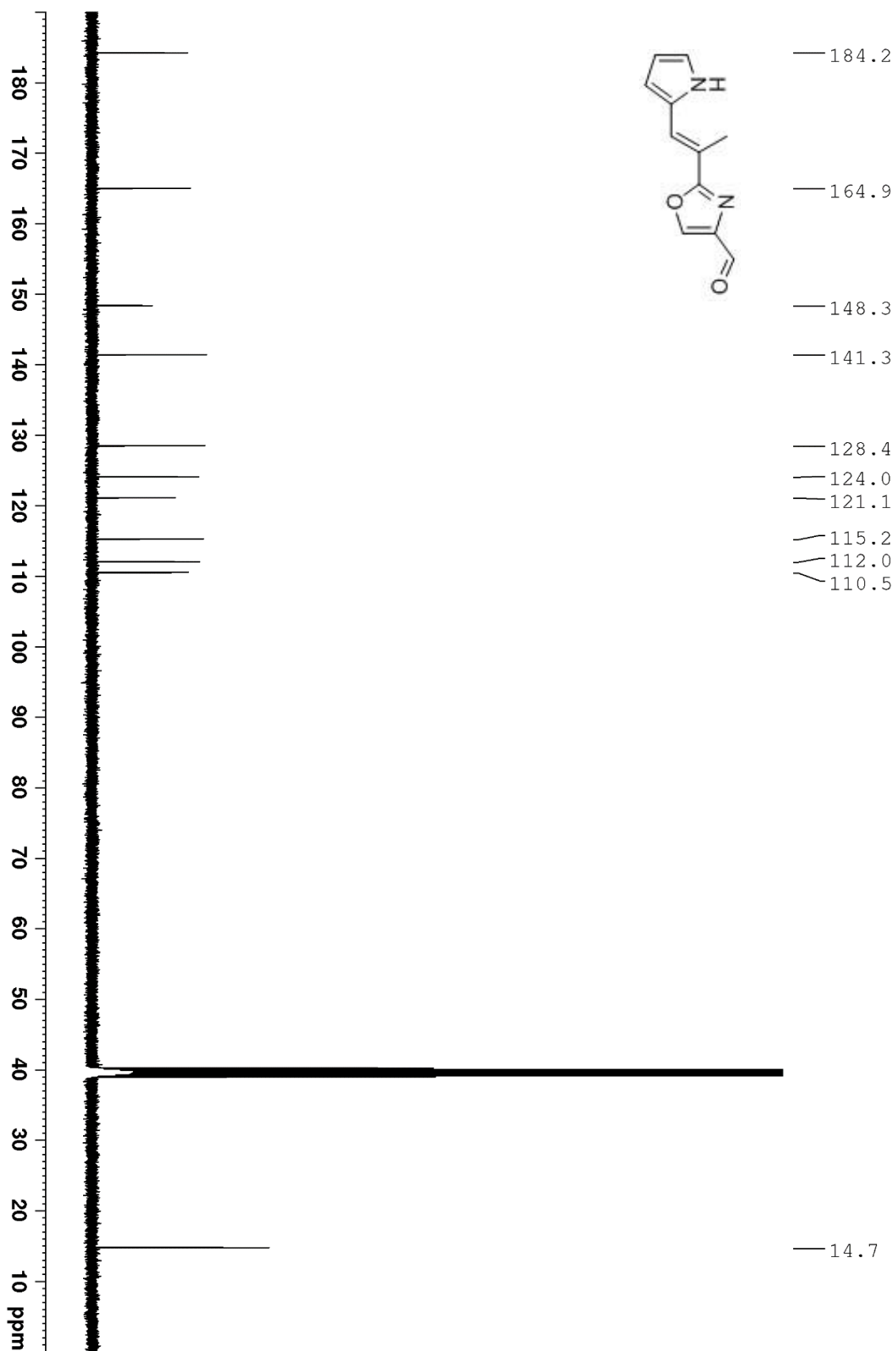


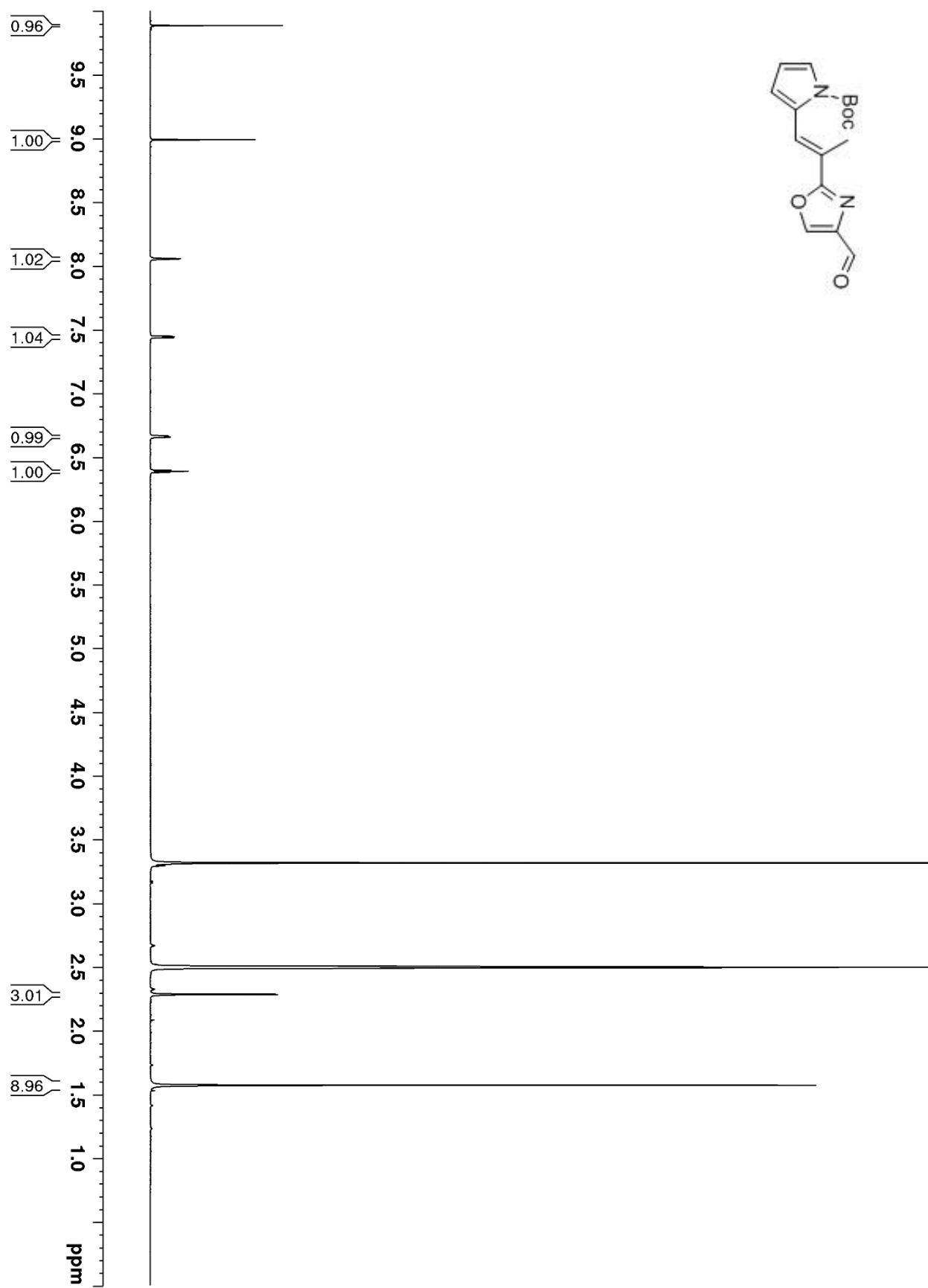


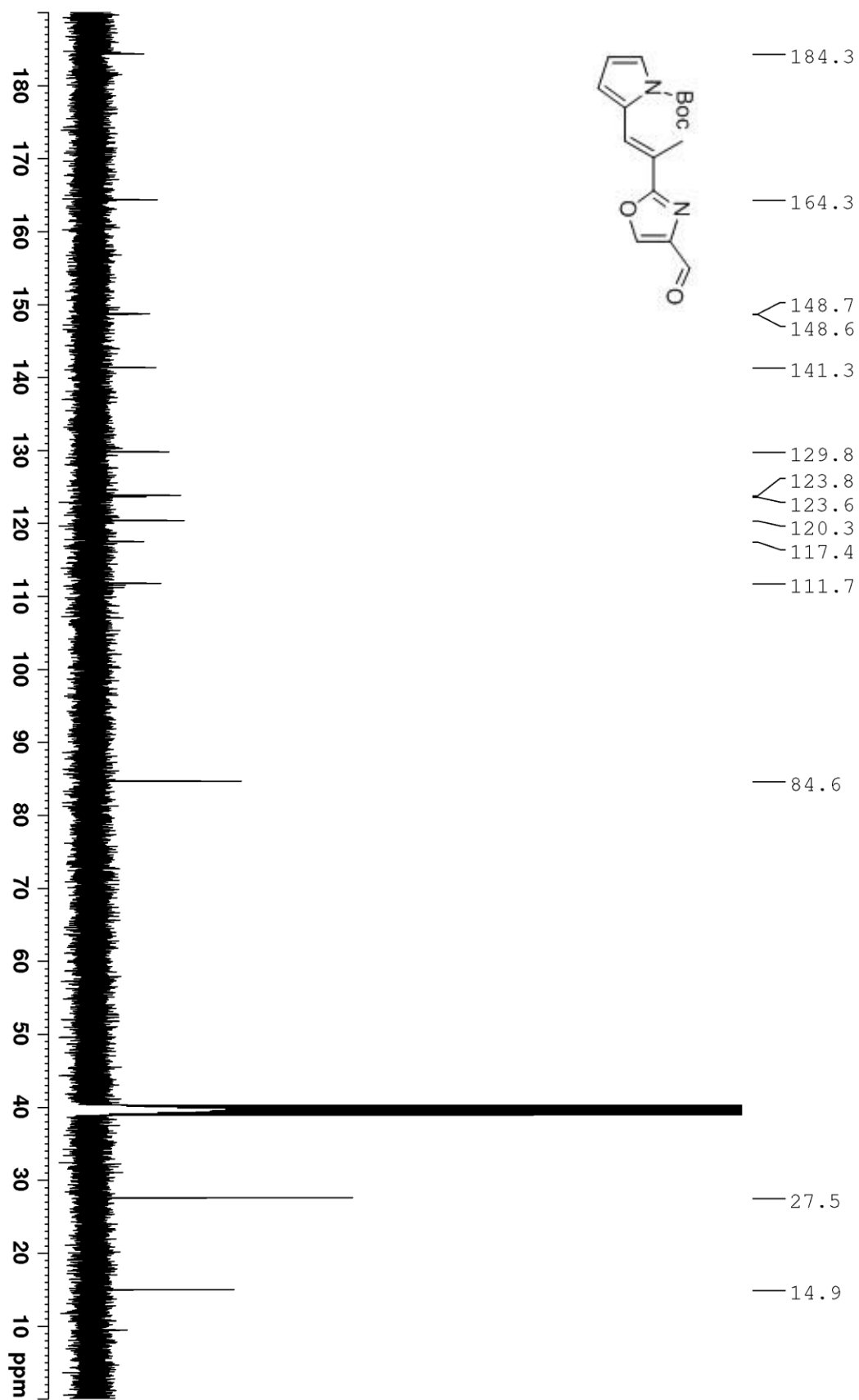


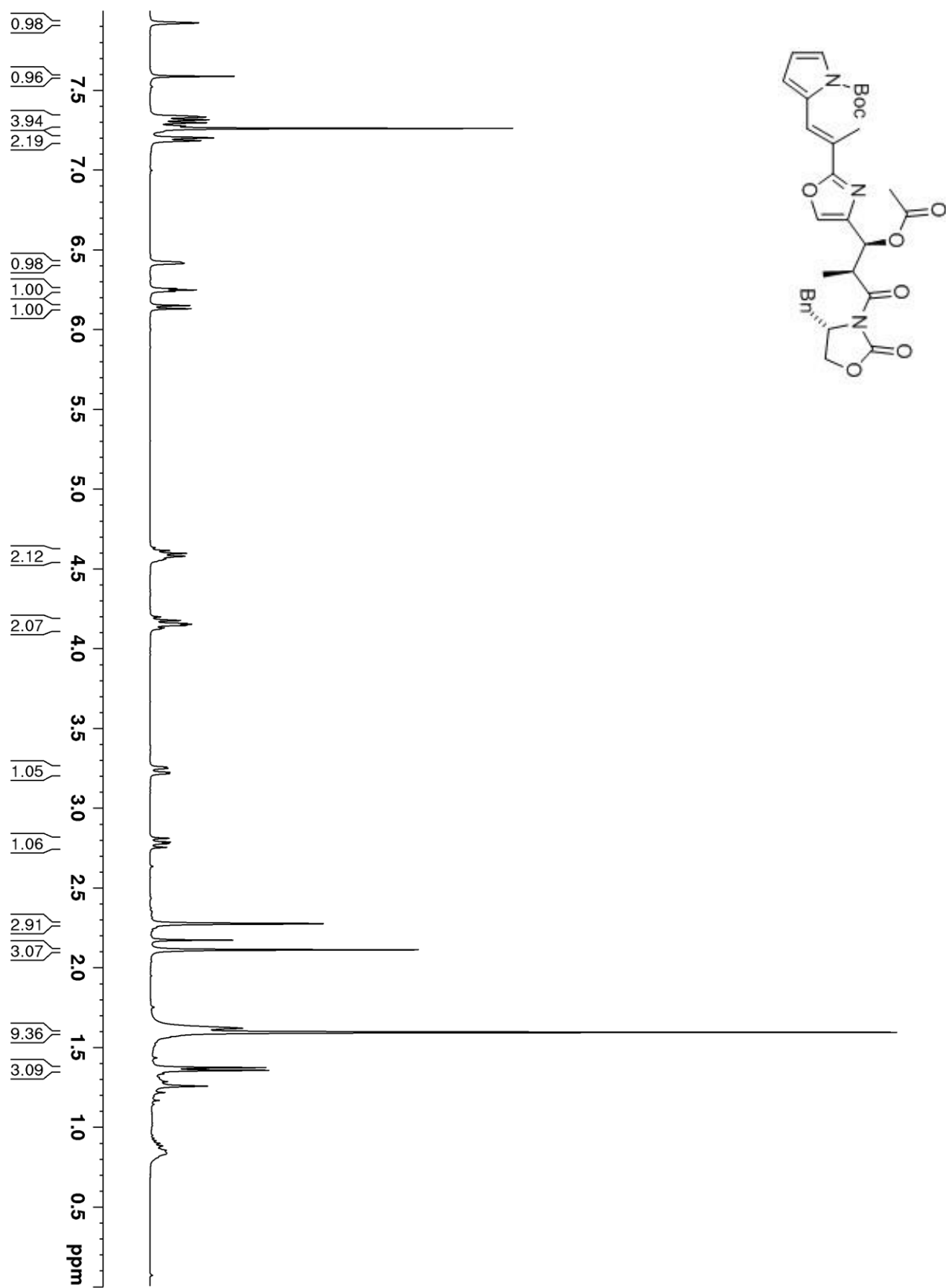


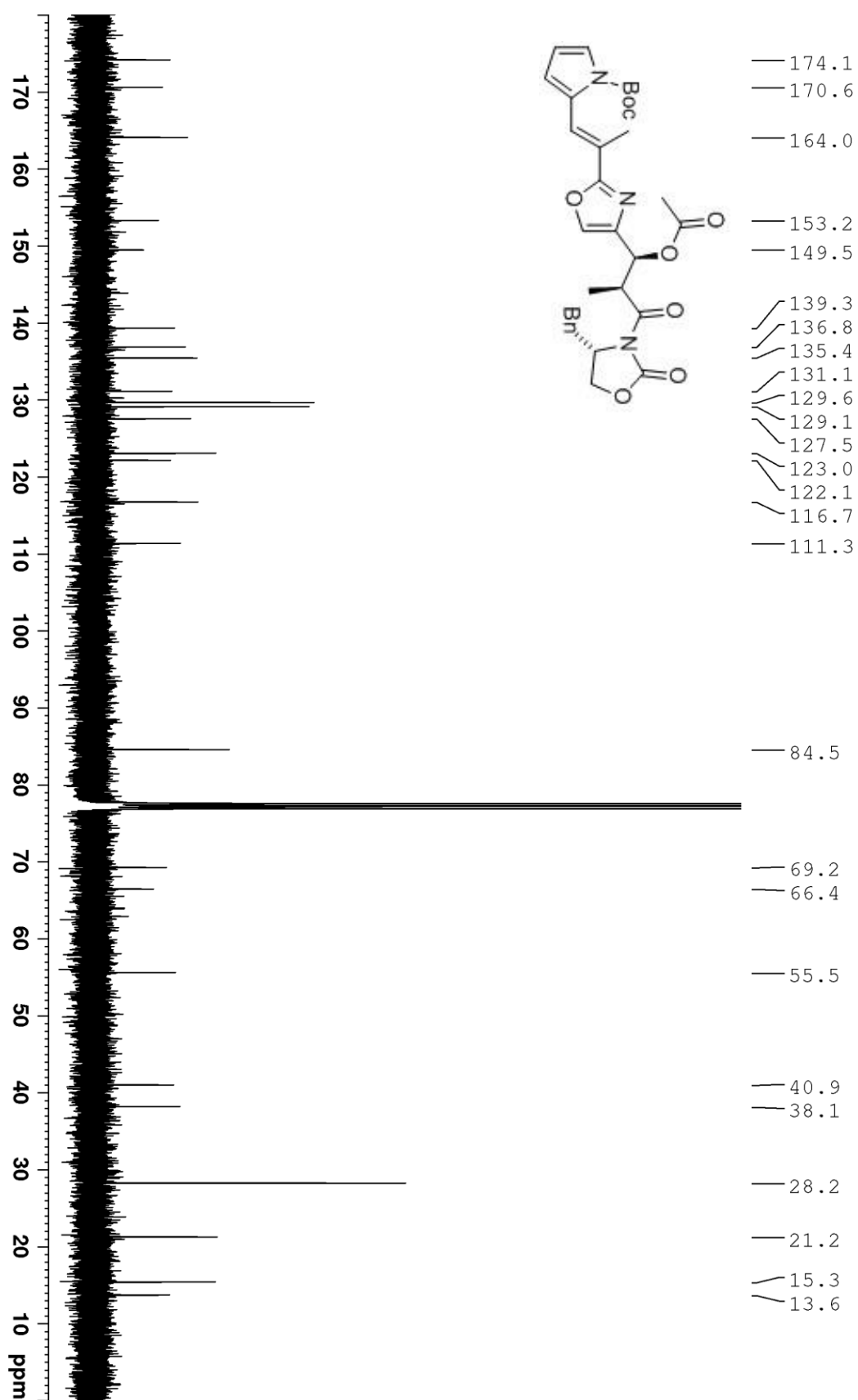


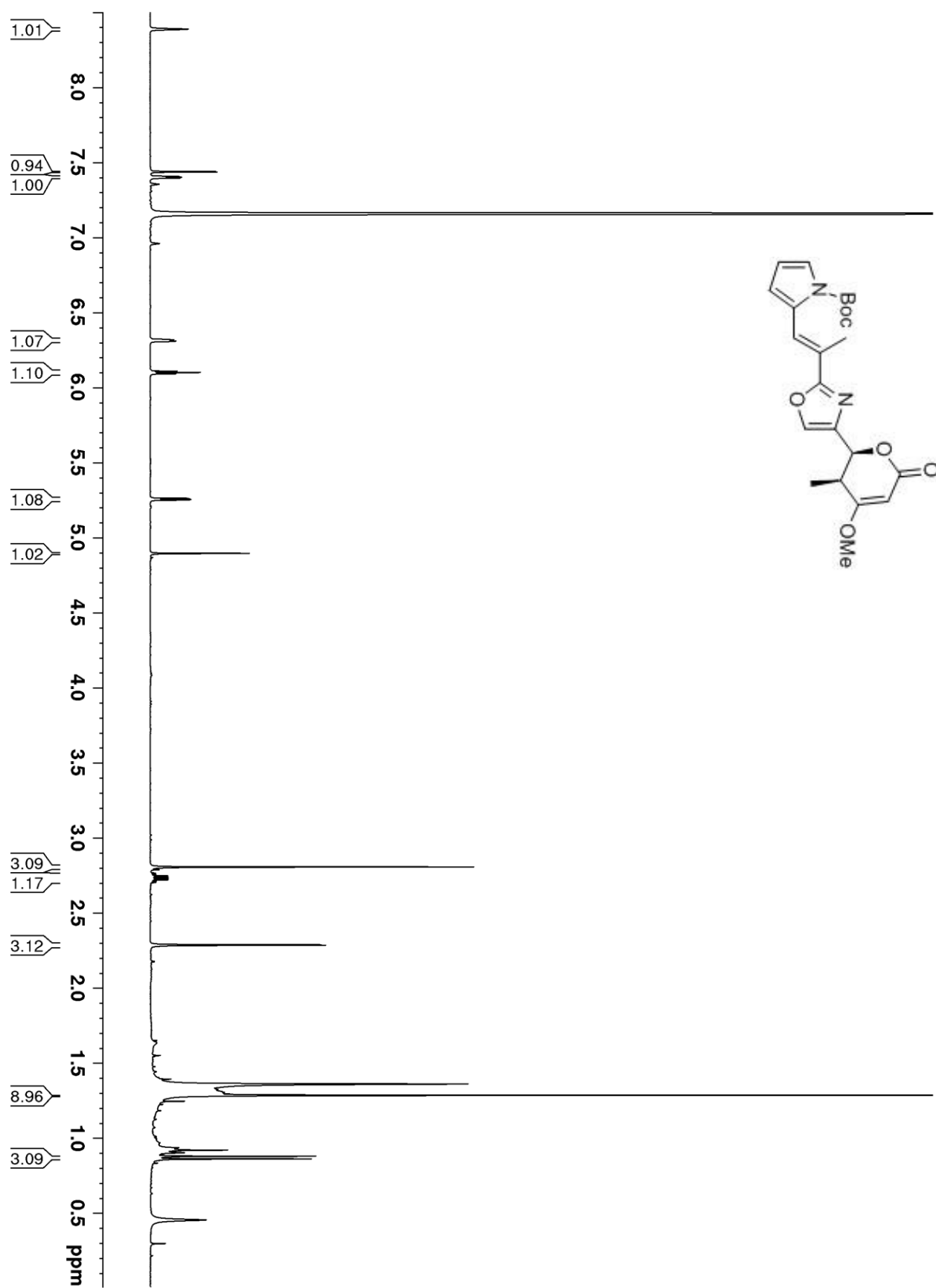


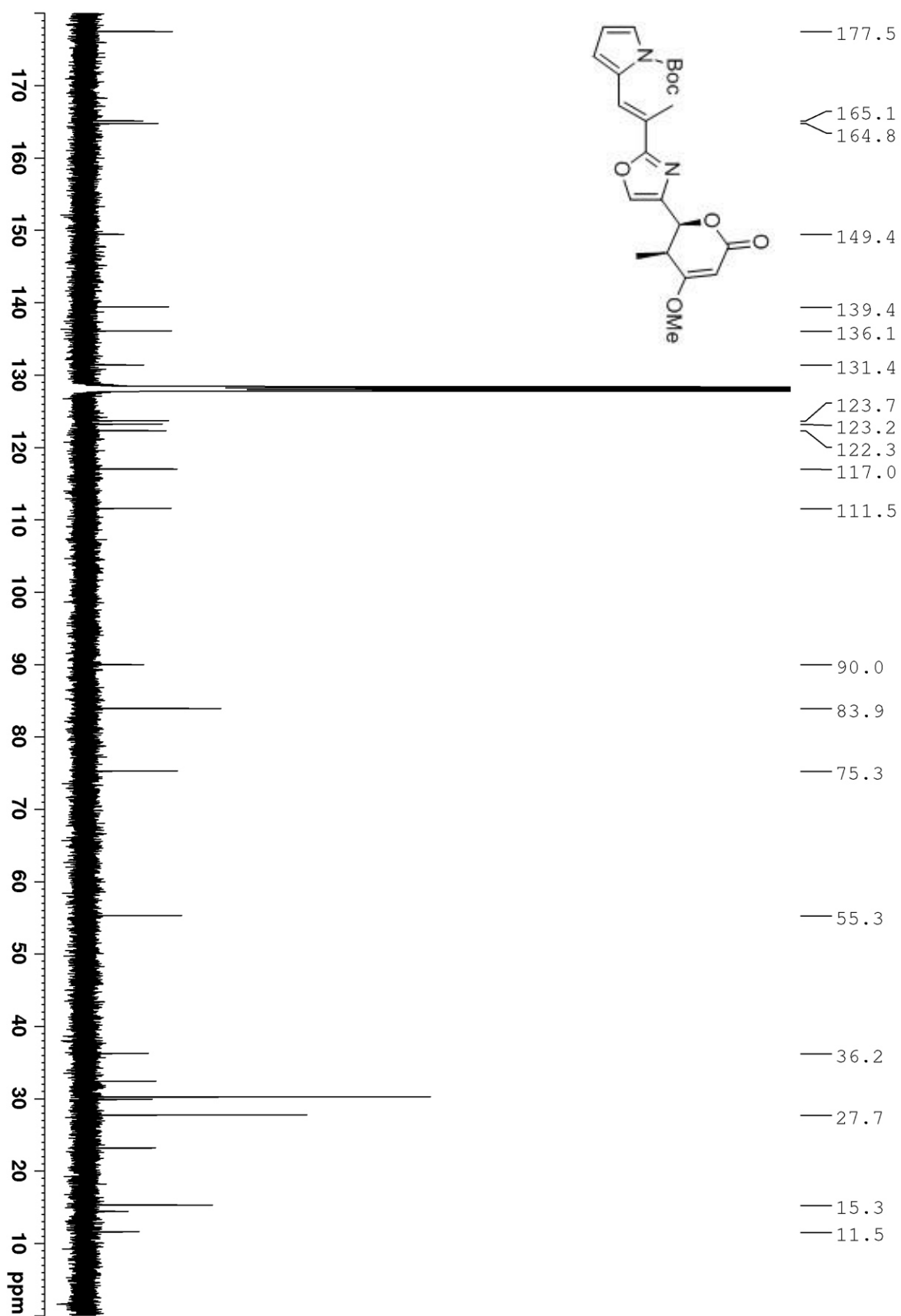




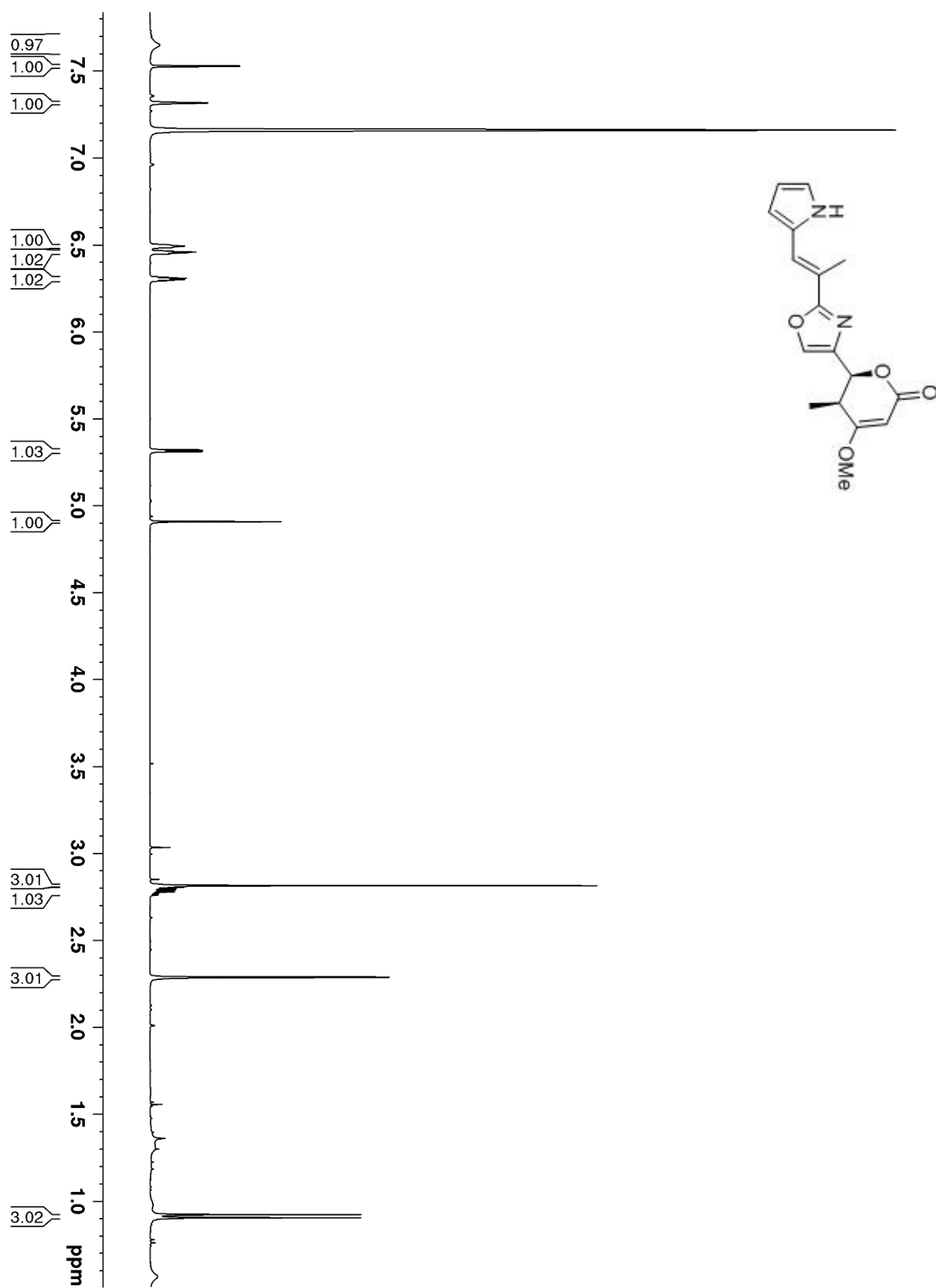


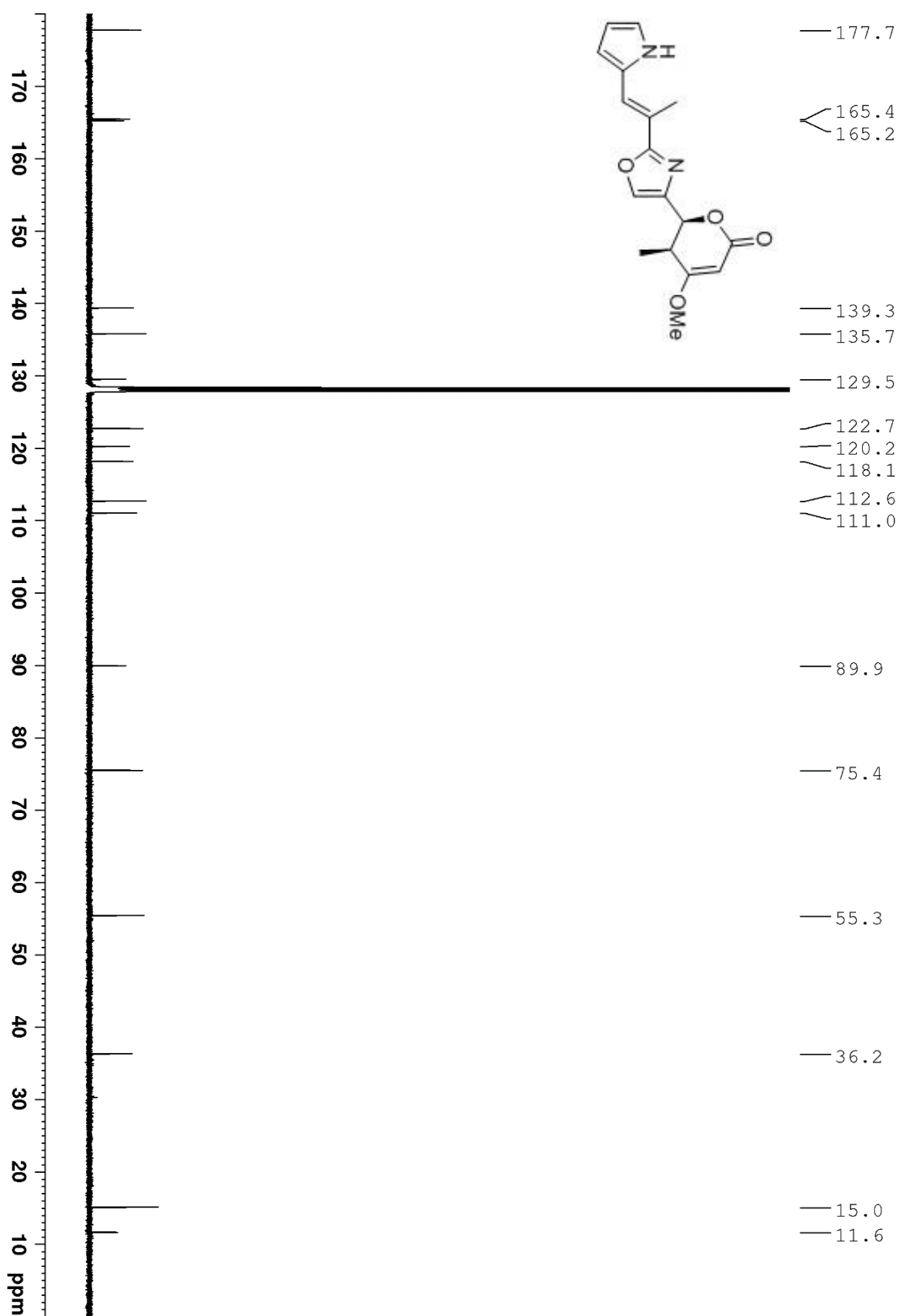


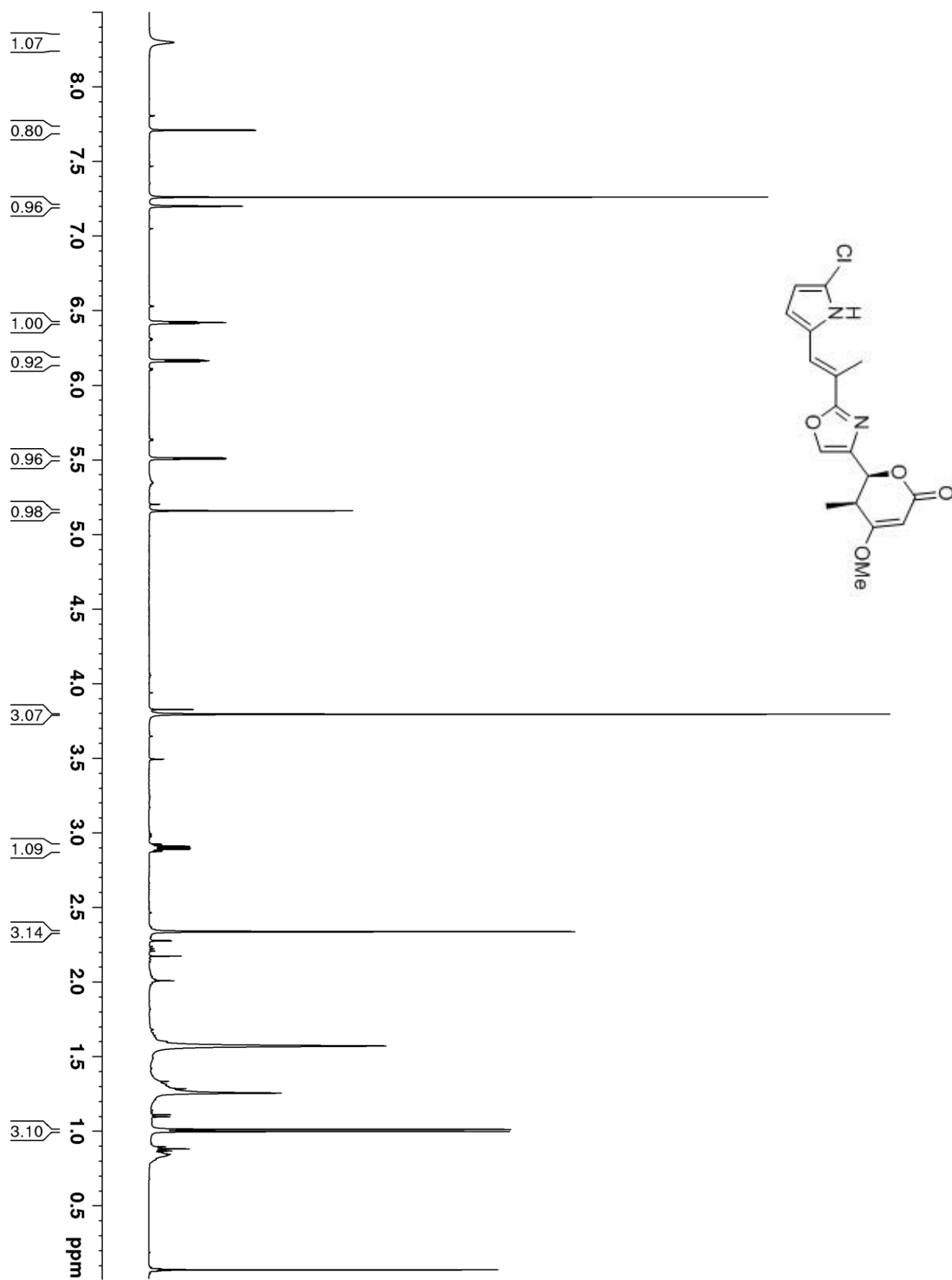


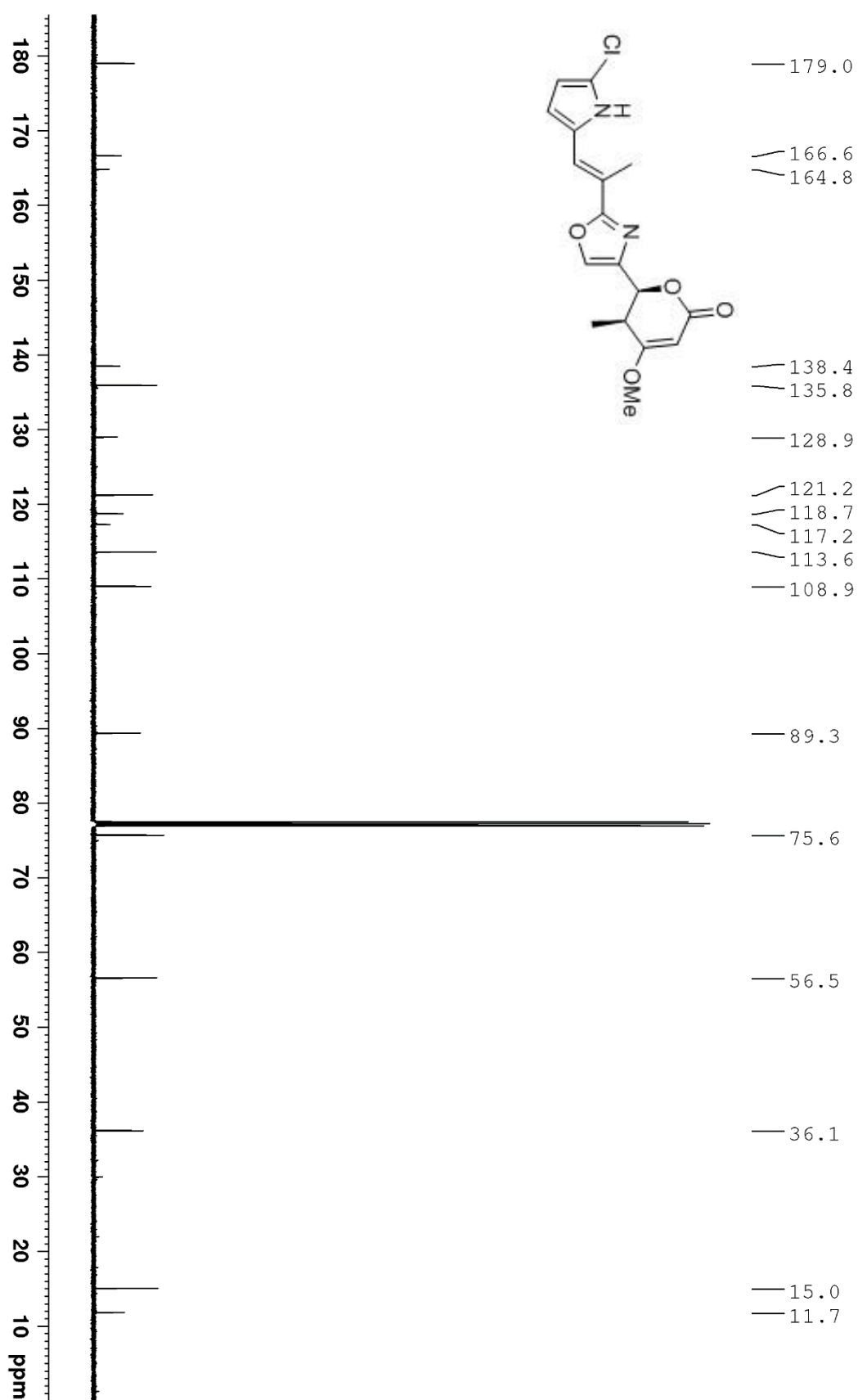






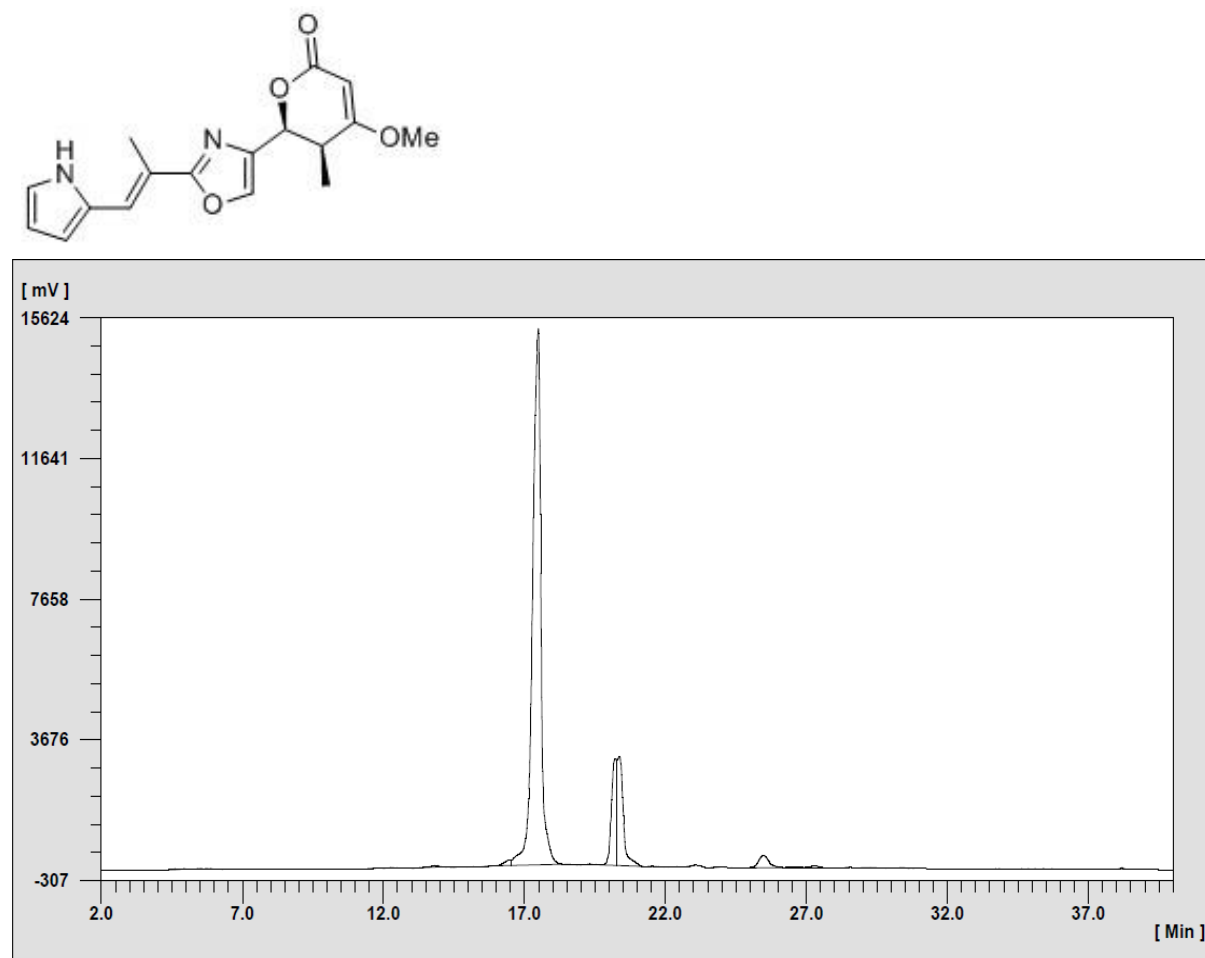






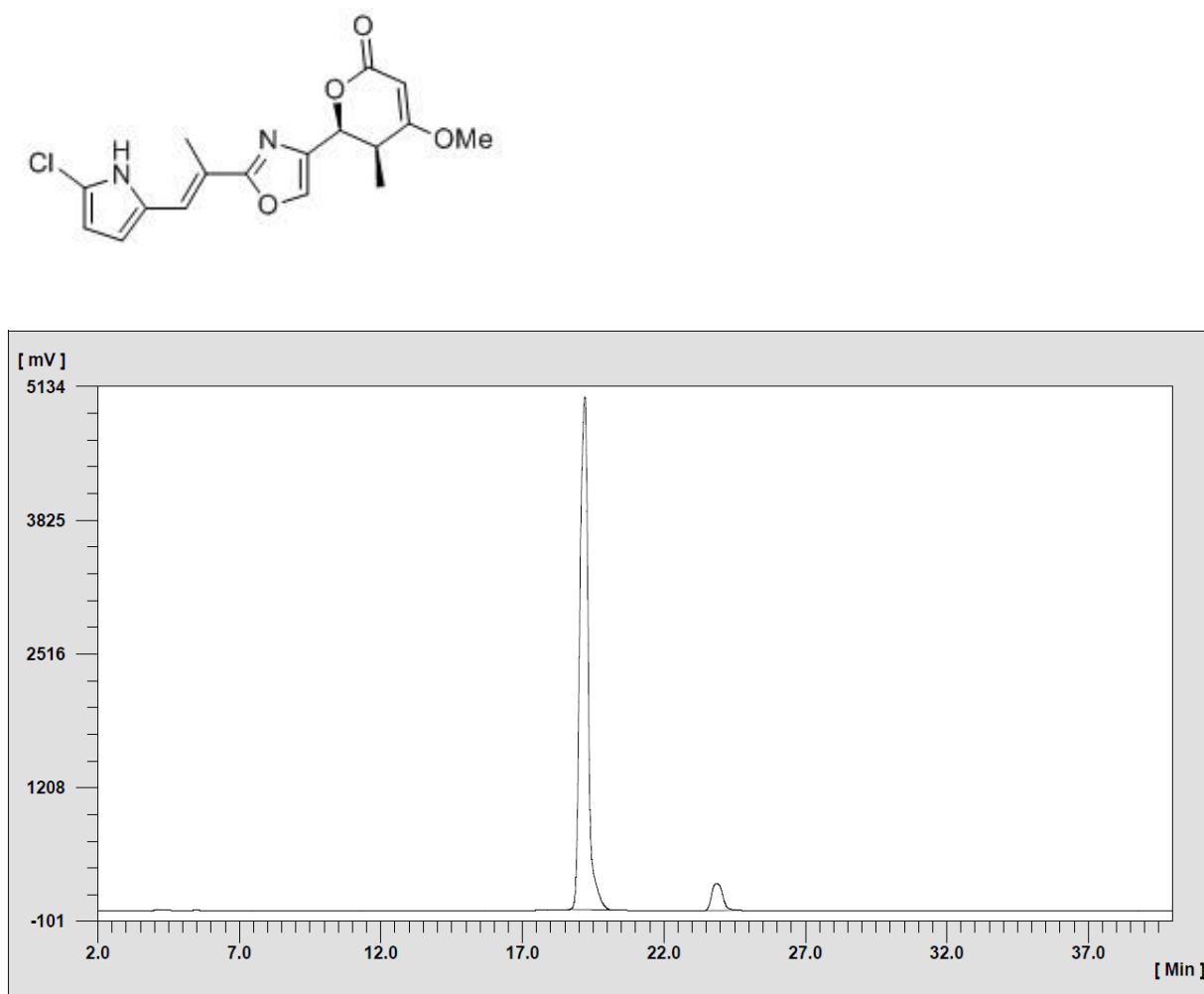
## 8. Copies of preparative HPLC chromatograms

The HPLC chromatograms of the purifications of **12** and pyrronazol B (**2**) are shown.



**Figure S9.** HPLC trace for the purification of **12**.

Channel	$t_R$	Height [mV]	Integration
1 (220 nm)	17.48	15176.78	313697.30
1 (220 nm)	20.23	3013.28	39675.26



**Figure S10.** HPLC trace for the purification of Pyrronazol B (2).

Channel	$t_R$	Height [mV]	Integration
1 (220 nm)	19.20	5030.24	102577.10
1 (220 nm)	23.83	260.96	7045.27

## References

1. Bennett, S. Solexa Ltd. *Pharmacogenomics* **5**, 433–438 (2004).
2. Weber, T. *et al.* antiSMASH 3.0—a comprehensive resource for the genome mining of biosynthetic gene clusters. *Nucleic Acids Res.* **43**, W237–W243 (2015).
3. Röttig, M. *et al.* NRPSpredictor2—a web server for predicting NRPS adenylation domain specificity. *Nucleic Acids Res.* **39**, W362–W367 (2011).
4. Finn, R. D. *et al.* The Pfam protein families database: towards a more sustainable future. *Nucleic Acids Res.* **44**, D279–285 (2016).
5. Kitsche, A. & Kalesse, M. Configurational assignment of secondary hydroxyl groups and methyl branches in polyketide natural products through bioinformatic analysis of the ketoreductase domain. *ChemBioChem* **14**, 851–861 (2013).
6. Jansen, R. *et al.* Pyrroazols, metabolites from the myxobacteria *Nannocystis pusilla* and *N. exedens*, are unusual chlorinated pyrone-oxazole-pyrroles. *J. Nat. Prod.* **77**, 320–326 (2014).
7. Campbell, S. E. *et al.* Synthesis of pyrrolizin-3-ones by flash vacuum pyrolysis of pyrrol-2-ylmethylidene Meldrum's acid derivatives and 3-(pyrrol-2-yl)propenoic esters. *J. Chem. Soc. [Perkin 1]*, 2195–2202 (1997).





## Chapter 3

# **Biosynthesis of cittilins, unusual ribosomally synthesized and post-translationally modified peptides from *Myxococcus xanthus***

**Joachim J. Hug<sup>†</sup>**, Jan Dastbaz<sup>†</sup>, Sebastian Adam, Ole Revermann, Jesko Koehnke, Daniel Krug and Rolf Müller

<sup>†</sup> These authors contributed equally to this work

*bioRxiv* preprint

DOI: 10.1101/2020.05.25.114512

*ACS Chemical Biology*

Submitted manuscript: 28.05.2020

### Contributions to the presented work

#### Author's contribution

The author contributed to the conception of this study, designed and performed experiments, evaluated and interpreted resulting data. The laboratory and *in silico* work regarding gene cluster analysis, genetic manipulation, heterologous expression, and biochemical analysis was performed by the author. Furthermore, he contributed significantly to conceiving and writing this manuscript.

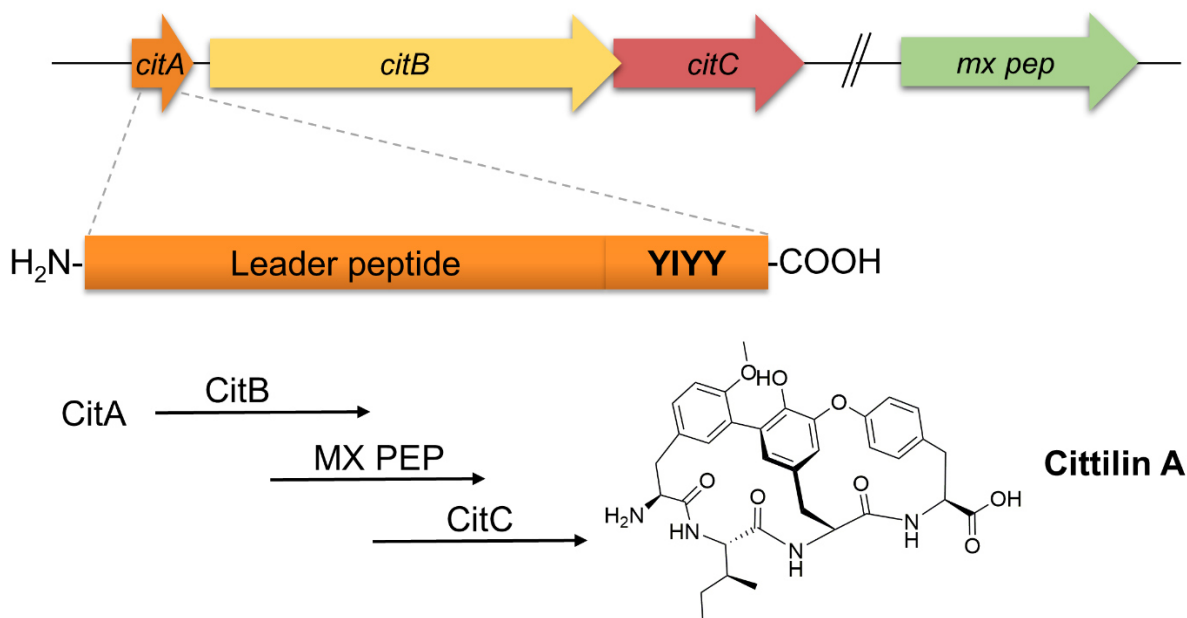
#### Contribution by others

Jan Dastbaz contributed to the conception of this study, designed and performed experiments, evaluated and interpreted resulting data. He performed cloning experiments, heterologous expression of the cittilin gene cluster and biochemical characterization of biosynthetic key enzymes. In addition, he contributed significantly to conceiving and writing this manuscript. Sebastian Adam was planning and conducting biochemical experiments and designed *in vitro* experiments. Furthermore, he contributed to conceiving and writing this manuscript. Ole Revermann initiated biosynthetic investigation of the cittilins and confirmed via isolation and NMR-based structure elucidation the correct assignment of cittilin A and B. Jesko Koehnke contributed by editing and proofreading this manuscript. Daniel Krug identified the cittilin biosynthetic gene cluster in *M. xanthus* DK1622 and contributed significantly to the conception and supervision of this study. Furthermore he contributed by editing and proofreading this manuscript. Rolf Müller contributed by supervision of the project and conceiving, editing and proofreading of this manuscript.

### 3 Biosynthesis of cittilins, unusual ribosomally synthesized and post-translationally modified peptides from *Myxococcus xanthus*

#### 3.1 Abstract

Cittilins are secondary metabolites from myxobacteria comprised of three L-tyrosines and one L-isoleucine forming a bicyclic tetrapeptide scaffold with biaryl and aryl-oxygen-aryl ether bonds. Here we reveal that cittilins belong to the ribosomally synthesized and post-translationally modified peptide (RiPP) family of natural products, for which only the crocagins have been reported from myxobacteria. A 27 amino acid precursor peptide harbors a C-terminal four amino acid core peptide, which is enzymatically modified and finally exported to yield cittilins. The small biosynthetic gene cluster responsible for cittilin biosynthesis also encodes a cytochrome P450 enzyme and a methyltransferase, whereas a gene encoding a prolyl endopeptidase for the cleavage of the precursor peptide is located outside of the cittilin biosynthetic gene cluster. We confirm the roles of the biosynthetic genes responsible for the formation of cittilins using targeted gene inactivation and heterologous expression in *Streptomyces*. We also report first steps towards the biochemical characterization of the proposed biosynthetic pathway *in vitro*. An investigation of the cellular uptake properties of cittilin A connected it to a potential biological function as an inhibitor of the prokaryotic carbon storage regulator A (CsrA).



**Graphical Abstract**

## 3.2 Introduction

Myxobacteria provide a multitude of natural products that exhibit diverse biological activities, which frequently feature novel modes-of-action<sup>1</sup>. Many natural products produced by myxobacteria are derived from large biosynthetic enzyme complexes of the modular non-ribosomal peptide synthetase (NRPS) and polyketide synthase (PKS) types<sup>2</sup>. The combinatorial nature of these megasynthetases contributes to the structural diversity of myxobacterial secondary metabolites, and the high mutual similarity of catalytic domains involved in their biosynthesis greatly facilitates the *in silico* identification and characterization of genes encoding these microbial biosynthetic pathways<sup>3</sup>.

Tools like the “antibiotics and secondary metabolite analysis shell” (antiSMASH) allow the rapid genome-wide identification, annotation and analysis of biosynthetic gene clusters (BGCs)<sup>4</sup> and are therefore instrumental for the discovery of new microbial natural products by genome-mining. However, other types of biosynthetic pathways are smaller, or the involved genes are less obviously clustered than the large multi-domain NRPS and PKS genes, making them more difficult to identify in bacterial genomes.

An example for such BGCs is the relatively young but steadily expanding family of ribosomally produced and post-translationally modified peptides (RiPPs), which includes the lantibiotics pinensins<sup>5</sup>, bicereucin<sup>6</sup> and the labyrinthopeptins<sup>7</sup>, lasso peptides such as fusilassin<sup>8</sup>, the antibiotic bottromycins, which feature a unique macoramidine linkage<sup>9,10</sup>, and the marine-derived divamides<sup>11</sup>, just to name a few. Recently developed bioinformatics tools allow the automated detection of certain classes of RiPPs in bacterial genomes based on signature genes and canonical recognition elements<sup>12–17</sup>. However, the pace of novel RiPPs discovery in recent years suggests that an abundance of yet undiscovered RiPP families likely exist. Most of these might until now have evaded identification because their structures and the underlying biosynthetic pathways deviate from previously described patterns. The only myxobacterial RiPP for example that has been characterized biosynthetically are crocagins, which contain a tetrahydropyrrolo [2,3-*b*] indoline core<sup>18</sup>.

In this study we reveal citalins (**Figure 1**) as a new class of RiPPs and connect them to a small BGC in the myxobacterium *Myxococcus xanthus* DK1622. This strain is a model organism for the study of bacterial motility and multicellular differentiation<sup>19</sup> and exemplifies the concept of a myxobacterial multiproducer of secondary metabolites<sup>20</sup>. Its genome contains 18 BGCs encoding NRPS, PKS, and NRPS/PKS hybrid systems<sup>21</sup>.

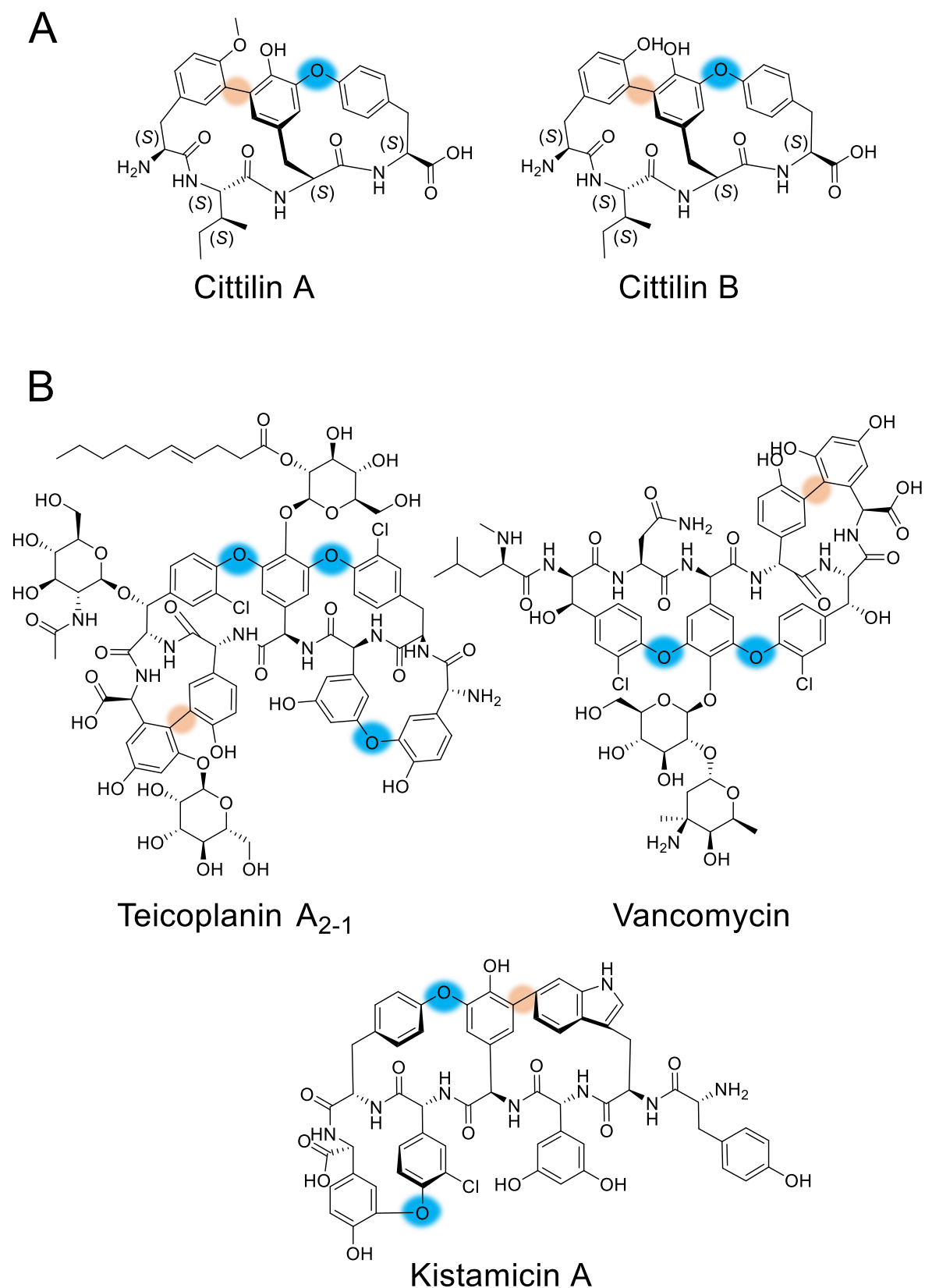
Citalins are also produced by *M. xanthus* DK1622, and in fact by more than half of all *M. xanthus* strains analyzed to date<sup>22</sup>. Citalin A and B (**Figure 1**) were first isolated in a study on the myxobacterial metabolite saframycin in *Myxococcus xanthus* Mx x48<sup>23–25</sup>. In a later screening, citalin B was isolated independently from *Streptomyces* strain 9738 and therefore named RP-66453<sup>26</sup>. Trowitzsch-Kienast *et al.* elucidated the planar structure of citalins, whereas the stereochemistry of all five carbonic stereocenters and the configuration of the biaryl-atropisomer was determined via total synthesis<sup>27–30</sup>. Stereochemical determination revealed the absolute configuration of all carbonic stereocenters as *S* and the molecule axis has been assigned as *R*. Hence, all four amino acid building blocks are L-configured.

The bicyclic structure of cittilins resembles the cross-linked structures of the glycopeptide antibiotics vancomycin, teicoplanin and kistamicin (**Figure 1**) which are generated via known modular NRPS machineries and a subsequent oxidative cyclization cascade of cytochrome P450 enzymes. These cytochrome P450 enzymes perform stepwise cyclization of the NRPS-bound heptapeptide to generate rigid, active glycopeptide antibiotics<sup>31</sup>.

In addition to antibiotics, there are also natural products that target bacterial virulence by abolishing pathogenic features without killing the bacterial pathogens<sup>32</sup>. The identification of these natural products is more challenging and requires sophisticated screening platforms, but at the same time these screening efforts pave the way for new antimicrobial treatments with less compound-induced selection pressure, leading to reduced rate of resistance development<sup>33</sup>.

The carbon storage regulator protein A (CsrA) was identified in many Gram-negative bacteria (such as *Helicobacter pylori*<sup>33</sup> and *Salmonella thyphimurium*<sup>34</sup>) to be involved in the regulation and translation of numerous virulence factors by binding the 5' untranslated region of mRNA. This CsrA-RNA interaction represses the translation of target transcripts through competition with the ribosome for RNA binding and modulates expression of multiple virulence-relevant processes, necessary for successful host infection<sup>35</sup>. A test system based on surface plasmon resonance and fluorescence polarization technologies was established and validated to test approximately 1000 small-molecules as potential inhibitors of the CsrA-RNA interaction<sup>36</sup>. This study revealed cittilin A (named in the publication as the undisclosed structure "MM14") as the most potent inhibitor of the CsrA-RNA interaction of all compounds tested (IC<sub>50</sub>: 4  $\mu$ M)<sup>36</sup>.

In our present study, we report the identification and heterologous expression of the genes encoding the biosynthetic machinery required to produce cittilin A and B. We delineate the biosynthetic origin of cittilins as RiPPs through targeted gene inactivation experiments and *in vitro* biochemical characterization of involved enzymes. In addition, we provide further biological characterization of cittilin A by investigating its cell entry properties and link it to a potential biological function as an inhibitor of the CsrA<sup>36</sup>.



**Figure 1.** Chemical structures of the biaryl tetrapeptides cittilin A and B (**A**), and glycopeptide antibiotics vancomycin, teicoplanin A<sub>2-1</sub> and kistamicin A (**B**) showing related substructures. Blue circles indicate aryl-oxygen-aryl ether bonds and beige-colored circles indicate biaryl bonds.

### 3.3 Results and discussion

#### Identification of cittilin biosynthesis genes

Cittilin A and B are commonly found in extracts derived from the species *M. xanthus* including the model strain DK1622, as judged on the basis of high-resolution LC-MS data from myxobacterial secondary metabolome studies<sup>22,37</sup>. We first verified the structures of cittilin A and B using NMR spectroscopy, since no analytical data for these compounds isolated from myxobacteria was available in the literature<sup>38</sup> (**Supporting Information 1.3**). Our analysis confirmed that the structure of myxobacterial cittilin B is identical to RP-66453 from *Streptomyces* strain 9738<sup>38</sup> (**Supporting Information, Table S8 and Figure S1**).

We then set out to investigate the genome of *M. xanthus* DK1622 for a potential cittilin BGC. However, retrobiosynthetic scrutiny of the *M. xanthus* DK1622 genome sequence based on NRPS biosynthesis concepts did not highlight biosynthetic genes that could plausibly be responsible for the production of cittilins. This *in silico* analysis result agrees with a previous study, where the construction of a targeted mutant library including genetic disruption of nine NRPS-related BGCs in the genome of *M. xanthus* DK1622 did not yield any mutant with abolished production of cittilins<sup>21</sup>.

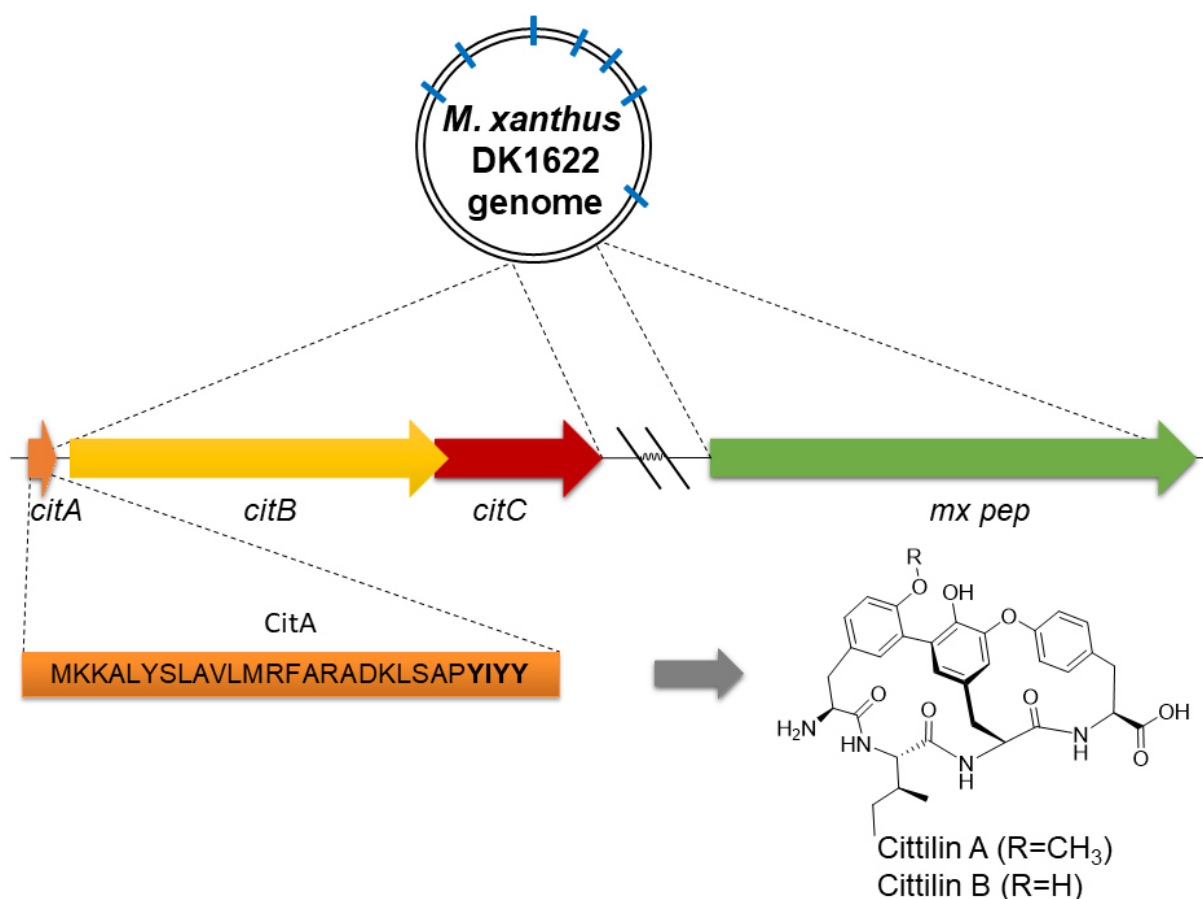
Given the peptide-based chemical structure of cittilins, we decided to explore the hypothesis that cittilin biogenesis could be a hitherto undescribed type of RiPP pathway. Formation of these peptide-derived compounds starts with ribosomal translation to give a precursor peptide, which typically undergoes diverse enzymatic modifications while it matures into the fully decorated natural product. Most often RiPP precursor peptides consist of a leader peptide important for substrate recognition by the modifying enzymes and a core peptide, which undergoes enzymatic modifications.

We thus searched the genome of *M. xanthus* DK1622 for a locus encoding a core peptide with the amino acid sequence YIYY. One such genetic locus was found encoding the tetrapeptide YIYY, located at the end of a small 84-base pairs open reading frame (ORF) we named *citA*. Upstream of *citA*, which might encode the precursor peptide CitA, we identified four ABC transporter genes (*citT<sub>a</sub>-T<sub>d</sub>*), while downstream one gene encoding a cytochrome P450-type enzyme (*citB*) and a gene encoding a methyltransferase (*citC*) are located (**Figure 2**).

Since several *M. xanthus* strains can produce cittilins, we investigated several confirmed myxobacterial producers of cittilins with available genome sequence data for the presence of this genetic locus. In fact, the genetic region putatively responsible for the biosynthesis of cittilins was identified in eight strains with available genome sequences (**Supporting Information, Figure S2**). Comparison of these additional BGCs assisted the annotation of the region thought to be responsible for the production of cittilin A and B. Gene cluster comparison revealed that the core region of cittilin biosynthesis is very small and the surrounding genes are highly variable. In fact, only the genes encoding CitA, CitB and CitC are clustered and thus likely constitute the essential cittilin biosynthesis operon spanning around 2.2 kbp.

Next, publicly available genome sequences were investigated to find genes resembling the putative cittilin gene cluster. From the genomes accessible in the GenBank database, six *M. xanthus* strains and *Myxococcus fulvus* HW-1 harbor gene clusters with high similarity to the cittilin gene cluster from *M. xanthus* DK1622; *Myxococcus hansupus mixupus* features the aforementioned biosynthetic organization to afford cittilins but has a nucleotide sequence encoding the deviant core peptide YHYY<sup>39</sup> (**Supporting Information, Figure S6**). Furthermore, the genome sequence of *Streptomyces* sp. Ncost-T10-10d contains a gene encoding a putative precursor peptide consisting of 27 amino acids (albeit the core peptide ends with the amino acid sequence YSYY). Similar to the myxobacterial biosynthetic gene cluster organization, this hypothetical cittilin-related operon also contains a gene encoding a CitB homolog. However, different from the myxobacterial cittilin gene cluster architecture, no *citC* homolog was found downstream of the *citB* homolog in *Streptomyces* sp.

This finding is consistent with an apparent difference in decoration of the core scaffold between *Streptomyces* and *Myxococcales*: it appears that *Streptomyces* lacks a methyltransferase, whereas all genome sequence data available for myxobacteria strictly shows the presence of a gene encoding a CitC homolog.



**Figure 2.** Retrobiosynthetic investigation of cittilin biosynthesis. The chemical structures of cittilin A and B consist of the proteinogenic amino acids L-tyrosine, L-isoleucine, L-tyrosine and L-tyrosine (YIYY). One genetic locus in the genome of *Myxococcus xanthus* DK1622 encodes the tetrapeptide YIYY, located at the end of a small open reading frame. Gene cluster comparison with other cittilin producers confirmed that genes encoding the precursor peptide (*citA*), the cytochrome P450 enzyme (*citB*) and the methyltransferase (*citC*) are strictly conserved and are thus likely to constitute the essential biosynthetic core region. Blue bars on the genome denote loci of seven biosynthetic gene clusters, which have been characterized and correlated to the produced natural products<sup>20</sup>.



## Cittilin biosynthesis in *M. xanthus* DK1622

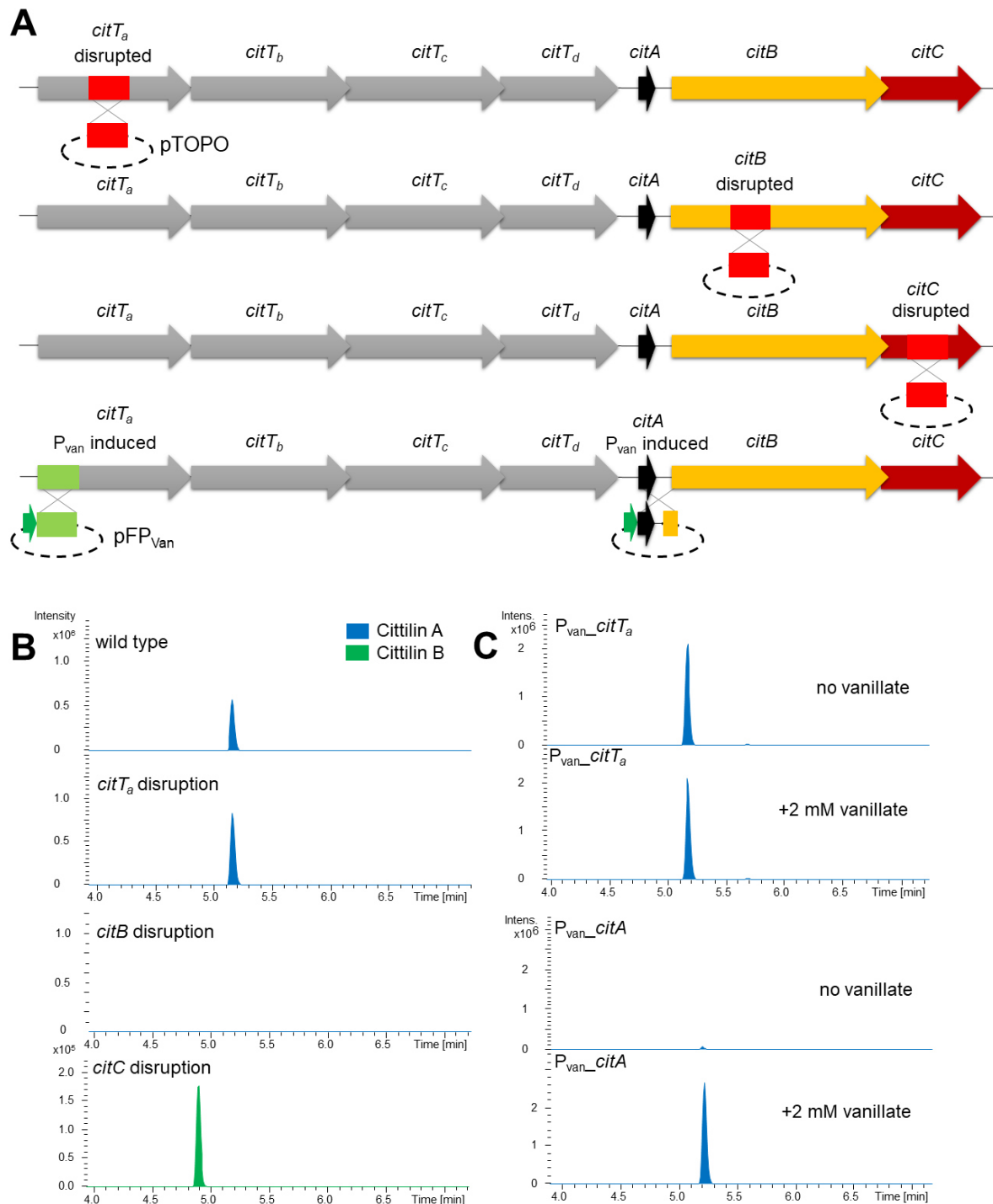
To confirm the assignment of cittilins to the identified candidate BGC in the native producer *M. xanthus* DK1622, we conducted gene disruption experiments by single crossover recombination of a plasmid into the putative biosynthetic genes (**Figure 3A**). In addition, a vanillate-inducible promoter system was inserted via single crossover recombination in front of the ABC transporter genes *citT<sub>a</sub>*–*citT<sub>d</sub>* or the identified ORF encoding CitA, respectively, to control the expression of the cittilin operon in *M. xanthus* DK1622 through supplementation of vanillate (**Figure 3A**).

Genetic disruption of *citB* abolished the production of cittilin A and B. Disruption of *citC* located within the cittilin operon also abolished the production of cittilin A as predicted, whereas the production rate of cittilin B increased significantly (**Figure 3B**). In contrast, disruption of *citT<sub>a</sub>* intended to abolish tetrameric ABC transporter formation, did not significantly influence the production rate of cittilin A or B in *M. xanthus* DK1622 (**Figure 3B**).

Since the nucleotide sequence length of *citA* was deemed too short for genetic disruption via single crossover recombination, a vanillate-inducible promoter system was inserted upstream of the identified precursor peptide gene *citA*. Subsequent HPLC-MS analysis clearly showed the influence of vanillate-induced expression of *citA* but not *citT<sub>a</sub>* on the production rate of cittilin A and B in the generated mutant (**Figure 3C**).

These results confirmed the proposed cittilin biosynthetic locus and the operon structure of the essential biosynthetic core region, consisting of the precursor peptide, the cytochrome P450 enzyme and the methyltransferase encoded by *citA*, *citB* and *citC*, respectively. Nevertheless, at least one additional enzyme was likely to be involved in the biosynthesis of the cittilins. Rather surprisingly, the cittilin operon does not contain a gene encoding a peptidase that is commonly found in RiPP biosynthetic gene clusters and is usually required to catalyze the cleavage of the leader peptide and thus releases the modified core peptide<sup>40</sup>.

Therefore, we attempted to define which enzymes are required for the biosynthesis of cittilins via heterologous expression of the biosynthetic genes encoded in the cittilin BGC.



**Figure 3.** Schematic overview of the cittilin biosynthetic gene cluster identified in *M. xanthus* DK1622, including the ABC transporter genes *citT<sub>a</sub>*–*citT<sub>d</sub>*, the precursor peptide gene *citA*, the cytochrome P450 enzyme gene *citB* and methyltransferase gene *citC* (A). Genetic modifications comprise *citT<sub>a</sub>* (first example on the top), *citB* (second example), and *citC* (third example) disruption and vanillate-inducible promoter insertion in front of *citT<sub>a</sub>* and *citA* (fourth example on the bottom). B) HPLC-MS EIC of crude extracts from *M. xanthus* DK1622 wild type and the respective gene disruption mutant. C) HPLC-MS EIC of crude extracts from *M. xanthus* DK1622\_ *P<sub>van</sub>\_citT<sub>a</sub>* and *M. xanthus* DK1622\_ *P<sub>van</sub>\_citA* with and without supplementation of 2 mM vanillate. EIC: Extracted ion chromatogram, blue: 631.2768 m/z, with a width of 7.9 ppm, cittilin A [M+H]<sup>+</sup>; green: 617.2611 m/z, with a width of 7.9 ppm, cittilin B [M+H]<sup>+</sup>.

## Heterologous expression of cittilin BGC

To confirm the functions of the identified gene products for the production of cittilins, we aimed to perform heterologous expression of candidate genes in the distantly related order of *Streptomyces*. Using a streptomycete as the heterologous host to produce cittilin appeared promising, since cittilin B is known from *Streptomyces* strain 9738<sup>26</sup> and we thus reasoned that the streptomycete host should be compatible with cittilin production.

The operon consisting of *citA*, *citB* and *citC* was amplified by PCR and subcloned into the expression vector based on plasmid pSET152 (**Supporting Information 2.4**). Following transformation into *Streptomyces albus* del14<sup>41</sup> by conjugation, this initial heterologous expression experiment did yield cittilin A at very low concentrations according to LC-MS analysis of the exconjugants (**Figure 4A**).

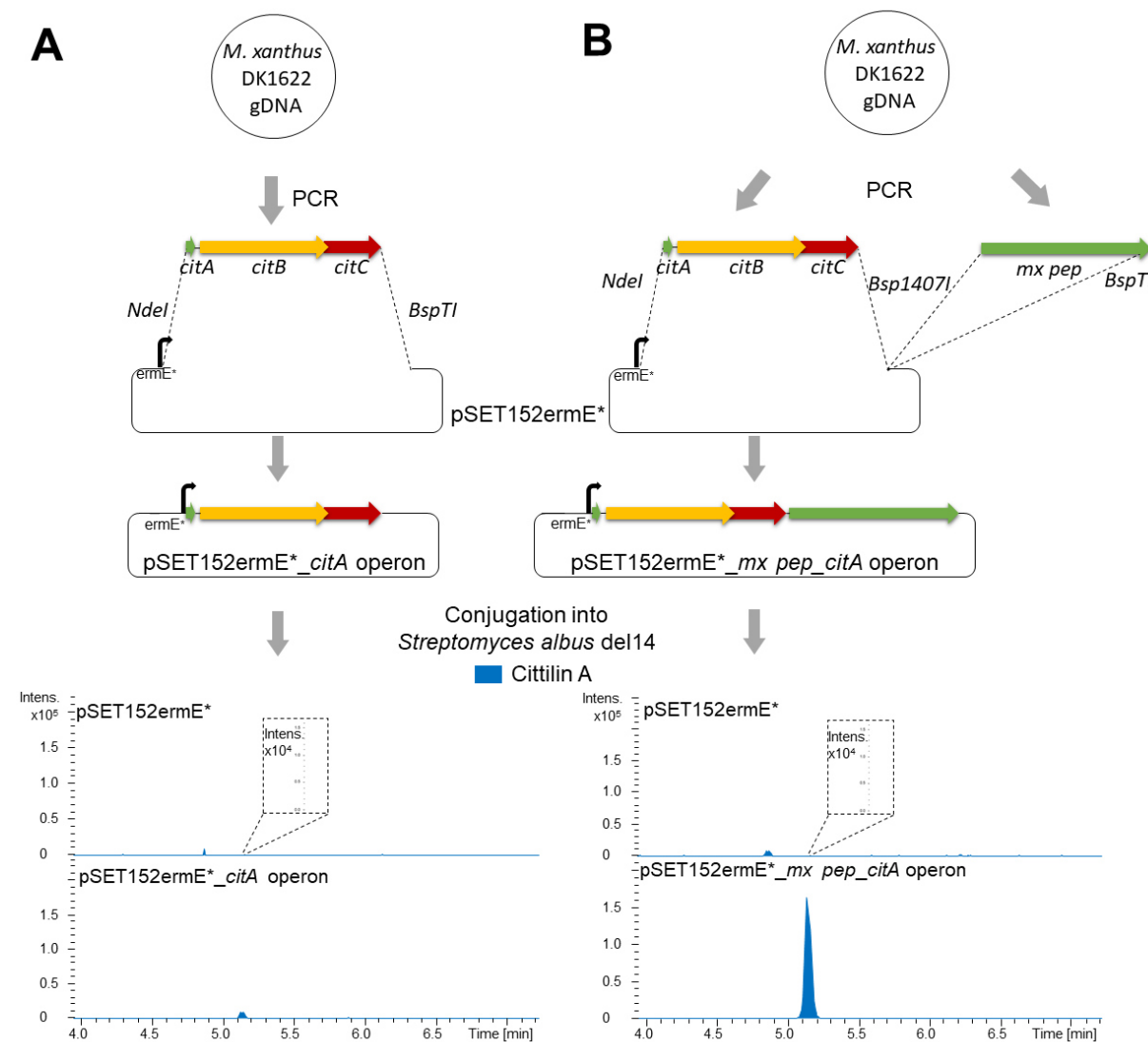
As mentioned above, the identified genetic locus for production of cittilin A and B lacks an ORF for a peptidase that removes the leader peptide during a typical RiPP biosynthesis<sup>42</sup>. Since L-proline is located in front of the core peptide amino acid sequence YIYY, one would expect a prolyl endopeptidase to catalyze the cleavage of the precursor peptide.

Among the numerous peptidase genes present in the genome sequence of *M. xanthus* DK 1622, one specific prolyl endopeptidase gene (*mx pep*) had been investigated previously<sup>43</sup>. Since MX PEP showed a higher preference for cleavage of Pro-Tyr/Phe bonds over Pro-Gln sites<sup>44,45</sup> and no homolog of *mx pep* was present in the genome of the heterologous host, we postulated that MX PEP could be involved in the formation of cittilins.

We thus constructed a fused cittilin operon that included the prolyl endopeptidase MX PEP. This operon yielded higher production of cittilin A (**Figure 4B**) in the heterologous host *S. albus* del14<sup>41</sup>. The corresponding LC-MS results readily showed the production of cittilin A in the heterologous host *S. albus* del14, and the identity of heterologously produced cittilin A was unambiguously confirmed through tandem MS (MS/MS) measurements, and by evaluation of fragment masses in the obtained MS/MS spectra compared to the purified cittilin A as reference (**Supporting Information, Figure S9 and S10**).

Following these promising results from the heterologous expression platform, we sought to verify the influence of MX PEP on the production of cittilins in the native producer *M. xanthus* DK1622. However, to our surprise genetic disruption of the *mx pep* gene in *M. xanthus* DK1622 did not significantly influence the production rates of cittilin A or B.

The fact that addition of MX PEP helps the heterologous host to cleave off the leader peptide argues for its involvement in cittilin formation, whereas the finding that its inactivation in *M. xanthus* DK1622 does not diminish cittilin production underpins the notion that MX PEP is not the only protease involved in the cleavage of CitA in the myxobacterial host.



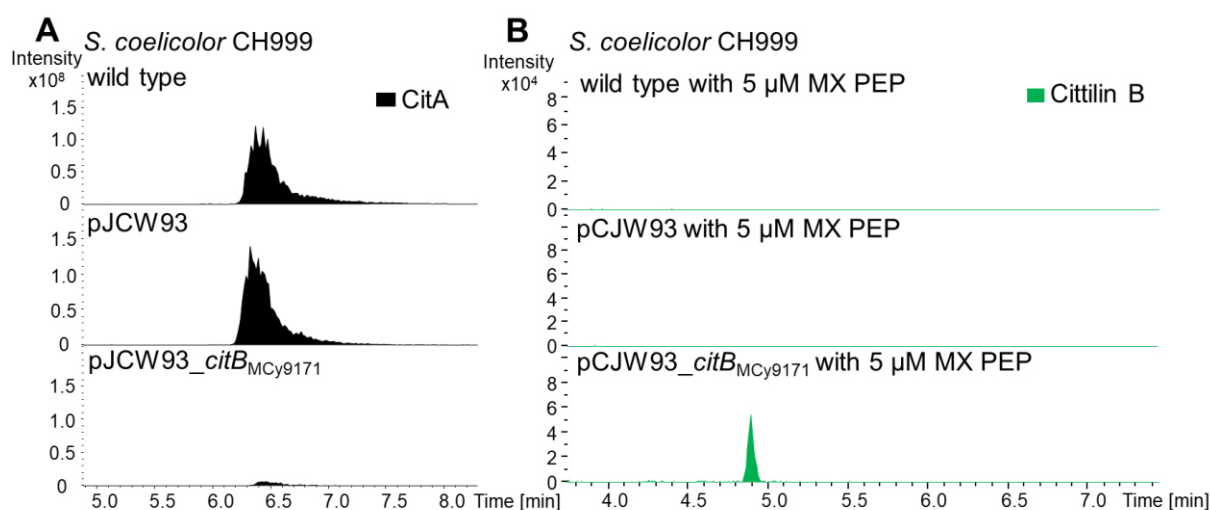
**Figure 4.** Heterologous production of cittilin A. **(A)** Heterologous expression of the cittilin operon *citA*–*citC* (*citA* operon) based on the constructed expression vector pSET152 resulted in low production of cittilin A. **(B)** Heterologous expression of *citA* operon and the gene encoding a previously characterized prolyl endopeptidase MX PEP, increased the amount of produced cittilin A. EIC: Extracted ion chromatogram, blue: 631.2768 m/z, with a width of 7.9 ppm, cittilin A [M+H]<sup>+</sup>.

### *In vitro* investigation of cittilin biosynthesis

The successful heterologous expression of cittilin genes provided the conceptual proof for the enzymatic steps needed for cittilin biosynthesis. Next, we aimed to shed light on the sequence of enzymatic steps leading to cittilin formation. Following the conventional biosynthetic logic of RiPP pathways<sup>40</sup>, the overall process can be subdivided into four steps: 1) production of the ribosomal precursor peptide, 2) posttranslational modification of the core peptide, usually dependent on leader peptide recognition (primary modifications), 3) proteolysis of the leader peptide to yield the modified core peptide and 4) further modifications (also termed secondary modifications) and export of the mature peptide-derived natural product.

In the case of cittilin, following the biosynthesis of the 27 amino acid precursor peptide CitA, the next step would involve the catalysis of the biaryl and aryl-oxygen-aryl linkage reactions by the cittilin cytochrome P450 enzyme CitB. From the heterologous expression experiments, it was concluded that CitB alone is likely to catalyze both of these intriguing biosynthetic transformations.

Heterologous production of recombinant CitB in *E. coli* could only be achieved through co-expression of the heat shock chaperone system *groEL/groES* in low yields, despite serious efforts (using different strains, genetic constructs and expression conditions, see **Supporting Information 2.5**). Production of the CitB homolog (CitB<sub>MCy9171</sub>) from MCy9171 (a confirmed alternative producer of cittilin A and B, see **Supporting Information, Figure S18**) was ultimately achieved in higher yields and without chaperone gene co-expression in the heterologous host *Streptomyces coelicolor* CH999. However, the carbon monoxide (CO)-spectral analysis of this homolog after purification shows an absorption maximum at 420 nm. The spectral peak at 420 nm for the ferrous (Fe<sup>2+</sup>) CitB<sub>MCy9171</sub>-CO complex indicates denaturation of cytochrome P450 enzyme in the process of isolation<sup>46</sup>. Accordingly, during catalytic activity testing of recombinantly produced CitB<sub>MCy9171</sub> in various experiments (including different co-factors and the Fdx/FdR reductase pair system from *Spinacia oleracea*, see **Supporting Information 2.5**) no obvious conversion of the precursor peptide was observed. These results, combined with the observation that CitB<sub>DK1622</sub> could only be obtained in *E. coli* in the presence of co-produced chaperone system GroEL/GroES, hint towards improperly folded protein.



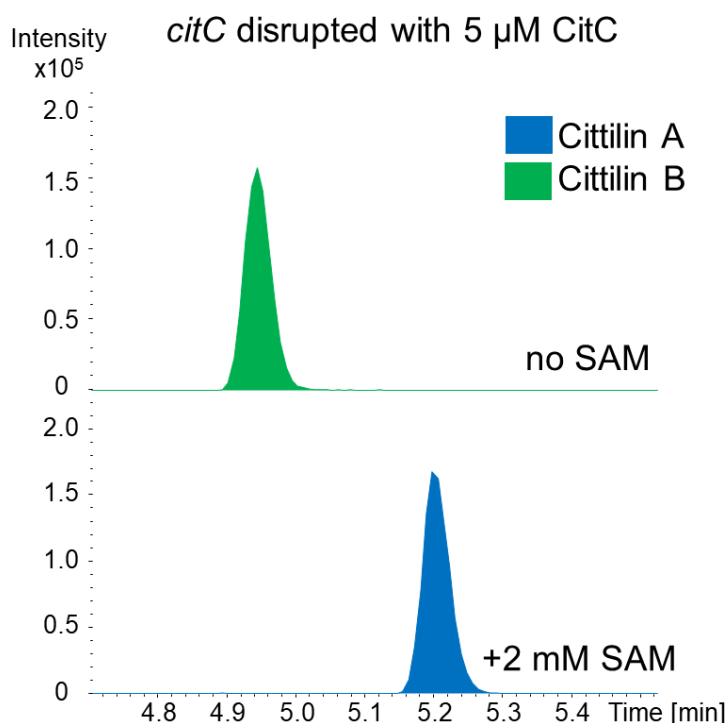
**Figure 5.** Investigation of cittilin biosynthesis through cell free lysate reactions in the heterologous host *Streptomyces coelicolor* CH999 (A) HPLC-MS EIC (3051.75 m/z, chemically synthesized precursor peptide [M+H]<sup>+</sup>) of cell-free lysates of *S. coelicolor* CH999, *S. coelicolor* CH999 with pCJW93 and *S. coelicolor* CH999 with pCJW93\_citB<sub>MCy9171</sub> are shown to observe conversion of supplemented chemically synthesized precursor peptide (62.5 μM). Only the cell-free lysate of *S. coelicolor* CH999 with pCJW93\_citB<sub>MCy9171</sub> shows significant consumption of the supplemented precursor peptide. (B) HPLC-MS EIC (617.2611 m/z cittilin B [M+H]<sup>+</sup>) of cell-free lysates of *S. coelicolor* CH999, *S. coelicolor* CH999 with pCJW93 and *S. coelicolor* CH999 with pCJW93\_citB<sub>MCy9171</sub> are shown to observe conversion of supplemented chemically synthesized precursor peptide (62.5 μM) in the presence of recombinantly produced prolyl endopeptidase MX PEP (5 μM). Only the cell-free lysate of *S. coelicolor* CH999 with pCJW93\_citB<sub>MCy9171</sub> with added recombinant MX PEP shows production of cittilin B.

As a workaround and to link CitB to the conversion of the precursor peptide *in vitro*, cell-free lysates of induced *S. coelicolor* CH999, containing different expression plasmids were incubated with chemically synthesized CitA (**Supporting Information 2.5**). These cell-free lysate reactions displayed consumption of the supplemented precursor peptide only for cell-free lysate originating from *S. coelicolor* CH999 expressing CitB<sub>MCy9171</sub>, but not the negative controls (**Figure 5A**). Supplying the

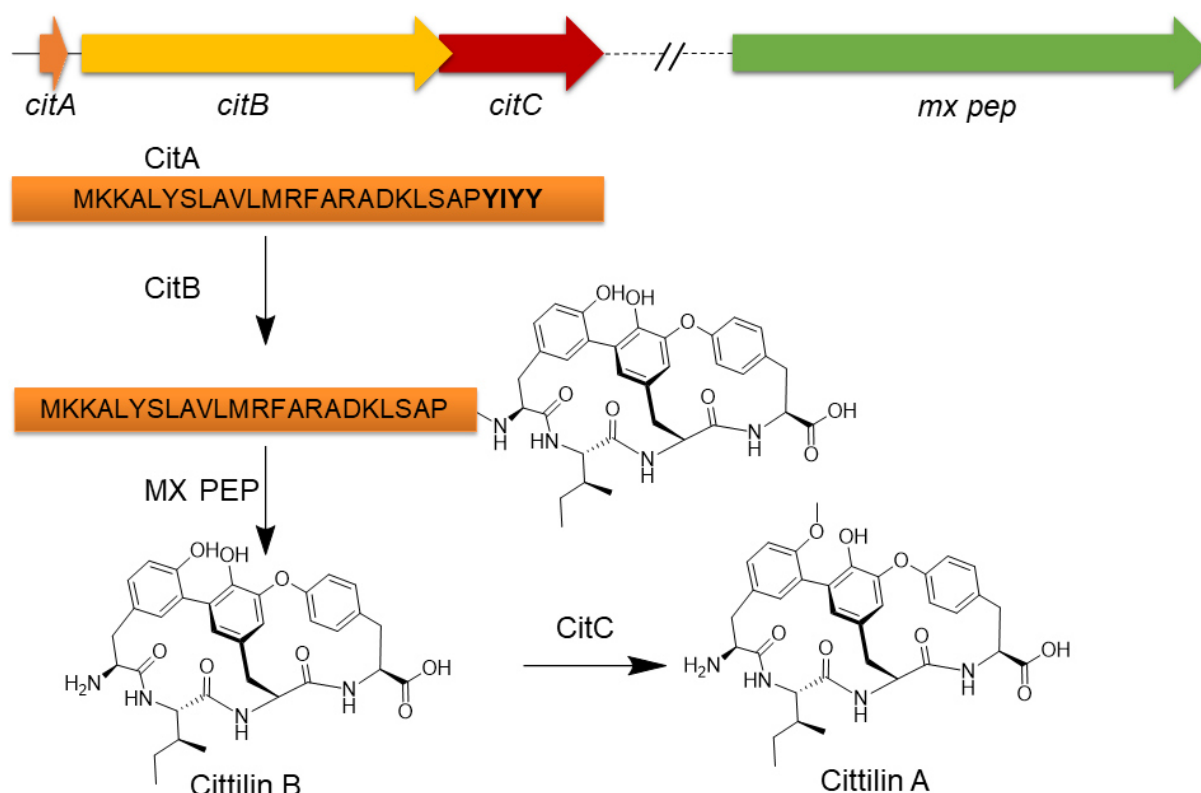
core peptide YIYY instead of full-length CitA did also not result in peptide conversion, which suggests that the leader peptide is essential for processing. Next, we simultaneously supplemented recombinant MX PEP and precursor peptide to cell free lysate from CitB<sub>MCy9171</sub>-producing *S. coelicolor* CH999 and observed the formation of cittilin B (**Figure 5B**).

The methyltransferase CitC was suspected to catalyze the last modification step in the biosynthesis, since the native host is naturally producing both, cittilin A and B. In addition, genetic disruption of *citC* resulted in the exclusive production of cittilin B (see above). We could also observe the CitC-dependent methylation of cittilin B with the expected mass shift of +14 Da yielding cittilin A, when a concentrated crude extract (containing cittilin B, but not cittilin A) of a *M. xanthus* DK1622 mutant with *citC* disruption was supplemented with recombinantly produced CitC (**Figure 6**). In contrast, recombinantly produced CitC showed no methylation activity when incubated with CitA or unmodified core peptide (YIYY) in *in vitro* assays (**Supporting Information 2.5**).

We conclude that the first enzyme to act on CitA is CitB, which catalyzes both, the biaryl and aryl-oxygen-aryl bond formation. The modified CitA is proteolytically processed by MX PEP, which removes the leader peptide from the modified core peptide (cittilin B). Finally, the cittilin methyltransferase CitC specifically methylates cittilin B, to yield the major derivative cittilin A (**Figure 7**). In conclusion, we confirmed the proposed biosynthetic pathway as shown in **Figure 7** through cell-free lysate experiments for the catalytic investigation of CitB and *in vitro* experiments of CitC and MX PEP. These findings set the stage for further in-depth biochemical analysis; in particular, the formation of the bicyclic ring system in cittilin catalyzed by CitB should become the subject of future studies.



**Figure 6.** *In vitro* conversion of cittilin B to cittilin A through recombinantly produced methyltransferase CitC. HPLC-MS EIC shows the catalytic conversion of cittilin B to cittilin A via CitC in dependence of *S*-adenosyl-L-methionine (SAM) as a co-factor. EIC: Extracted ion chromatogram, blue: 631.2768 m/z, with a width of 7.9 ppm, cittilin A [M+H]<sup>+</sup>; green: 617.2611 m/z, with a width of 7.9 ppm, cittilin B [M+H]<sup>+</sup>.



**Figure 7.** Schematic overview of cittilin biosynthesis in *M. xanthus* DK1622. The proposed biosynthesis of cittilin A starts with the ribosomally synthesized precursor peptide CitA, which is then further processed by the cytochrome P450 enzyme CitB. Subsequently, the modified precursor peptide (hypothetical chemical structure of the cyclized tetrapeptide is shown) is cleaved by the action of a prolyl endopeptidase (MX PEP being a candidate according to heterologous expression results), which yields cittilin B. Finally, cittilin B is undergoing methylation via the methyltransferase CitC yielding the major myxobacterial compound cittilin A.

## Biological function of cittilin A

Since the production of cittilin occurs widespread among strains of the genus *Myxococcus*<sup>22</sup>, the question of the biological function of this natural product arises. Cittilin A appears to serve none of the biological functions often assigned to other myxobacterial natural products since no significant antibiotic, antifungal or cytotoxic activity was observed (**Supporting Information 2.6**)<sup>47</sup>. Biological activities of cittilin A were previously listed in the literature as a moderate inhibitor of pancreatic elastase, whereas cittilin B was considered as a neurotensin inhibitor ( $IC_{50}$ :  $30 \mu\text{g mL}^{-1}$ )<sup>26</sup>, which even led to synthesis efforts towards the drug development of peptides resembling the tetrapeptidic core structure of cittilin<sup>48</sup>.

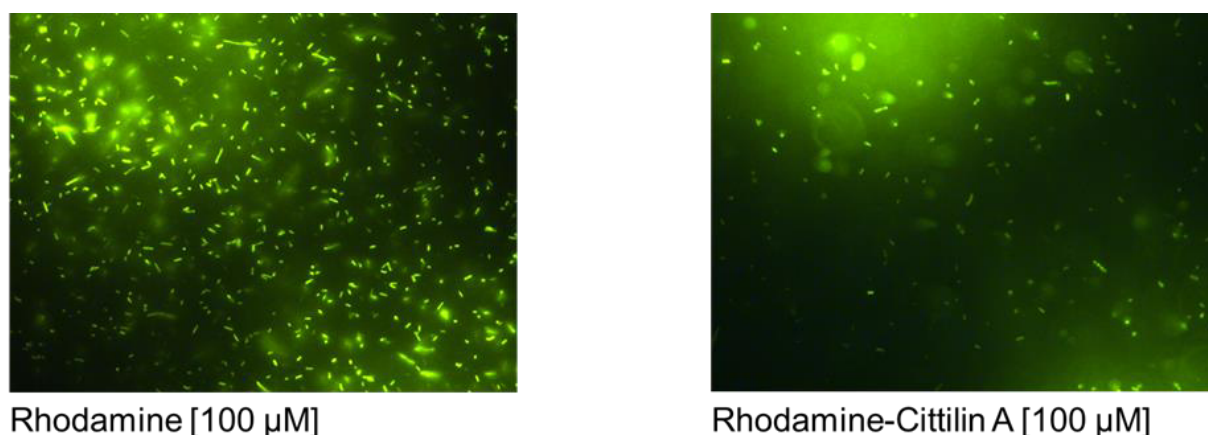
In a recently conducted screen for inhibitors of CsrA, cittilin A was found to be the most potent inhibitor<sup>36</sup>. Cittilin A (named in the publication as the undisclosed structure “MM14”) featured the highest *in vitro* affinity for inhibition of the interaction of CsrA with RNA ( $IC_{50}$ :  $4 \mu\text{M}$ )<sup>36</sup>. Cittilin A was consequently considered as a potential small-molecule inhibitor of this essential virulence factor in opportunistic pathogens.



The obvious discrepancy of cittilin A between its potent *in vitro* inhibition of CsrA and the lack of any biological effect against Gram-negative pathogens, implies that pharmacokinetic obstacles account for the absence of its *in cellulo* activity. Therefore, we chose to investigate the cellular uptake of cittilin A with the help of LC-MS analysis as well as fluorescence microscopy.

Cell entry of synthesized rhodamine-tagged cittilin A (**Supporting Information, Figure S21**) was tested in *E. coli* by fluorescence microscopy of treated bacterial cells in comparison with free rhodamine. Microscopic analysis reveals that *E. coli* cells incubated with free rhodamine display strong fluorescence, whereas cultures treated with rhodamine-tagged cittilin A show significantly less fluorescence (**Figure 8**). Furthermore, the supernatant of *E. coli* TolC mutants treated with either cittilin A or free rhodamine were extracted separately and subjected to LC-MS analysis to compare cell entry properties of both compounds (or strong adhesion to the cell membrane or cell wall respectively). Only free rhodamine was taken up by the cells, according to LC-MS analysis (**Supporting Information, Figure S22**).

Given these results, we conclude that cittilin A has poor cell entry capability for *E. coli*, which impedes the inhibition of intracellular CsrA and may explain why cittilin A showed no inhibition of pyocyanin production when administered to *Pseudomonas aeruginosa*, an assay commonly used to evaluate anti-virulence efficiency of small-molecules (**Supporting Information 2.6**)<sup>49</sup>. To support the assumption of poor pharmacokinetics, studies using cittilin A with improved cell entry properties through coupling with iron chelators might give further insights into its biological function. At present we cannot exclude that the biological function of cittilins is distinct from its potential activity as an anti-infective pathoblocker<sup>50</sup>.



**Figure 8.** Test for bacterial cell entry of rhodamine (left) and rhodamine-tagged cittilin A (right) in the Gram-negative bacterium *Escherichia coli* (TolC efflux deficient *E. coli* mutant<sup>51</sup>, **Supporting Information 2.6**).



### 3.4 Conclusions

We established the molecular basis for the biosynthesis of cittilin A and B, two common natural products of myxobacteria from the genus *Myxococcus*. Even though the rare bicyclic structure of cittilins resembles those of the cross-linked substructures known from the glycopeptide antibiotics vancomycin, teicoplanin and kistamicin, their biosynthesis and biological function deviates significantly. The small BGC involved in the biosynthesis of the cittilins consists of a gene encoding the 27 amino acid precursor peptide CitA, the cytochrome P450 enzyme CitB and the methyltransferase CitC, which is unprecedented for RiPPs. Although the recently described darobactin genetic locus remotely resembles the cittilin BGC, the precursor peptide of darobactin is twice the size of the cittilin precursor peptide and the proposed radical *S*-adenosyl-L-methionine (SAM)-catalyzed cyclization differs significantly from the biaryl and aryl-oxygen-aryl formation in cittilin biosynthesis<sup>52</sup>. Strikingly, the formation of the bicyclic ring system in cittilin apparently requires the enzymatic activity of only one cytochrome P450 enzyme. These features are not shared by any of the known RiPP compound classes<sup>40</sup> and lead us to the conclusion that cittilins constitute a new class of RiPPs. Our study provides the foundation for further investigation of this small, but intriguing biosynthetic pathway. *In silico* analysis based on publicly available genome sequences already hints towards proposed cittilin derivatives, which feature the alternative core peptide sequences YHYY and YSYY in *M. hansupus mixupus* and *Streptomyces* sp. Ncost-T10-10d, respectively.

The biotechnological generation of bicyclic tetrapeptides resembling the cittilin scaffold are now conceivable both in the native and the heterologous host, but also further investigation of the underlying mechanism of the interesting cytochrome P450 enzyme reaction yielding the bicyclic linkage will be feasible once reaction conditions to establish the respective steps purely *in vitro* are established. In addition, identification of the biosynthetic genes producing putative cittilin derivatives across phylogenetic borders might further help to determine the biological function of these intriguing natural products.

### 3.5 Methods

Details of experimental procedures are provided in the **Supporting Information**.

### 3.6 Associated Content

The sequence of the cittilin biosynthetic gene cluster originating from *M. xanthus* DK1622 has been deposited in the Minimum Information about a Biosynthetic Gene cluster (MIBiG) database under the accession number BGC0002043.

### 3.7 Acknowledgements

The authors thank W. Hofer for the synthesis of the rhodamine-coupled cittilin A and recording the NMR spectra of the cittilins, J. Herrmann and A. Boese for microscopic fluorescence imaging, S. Schmidt for performing bioactivity assays, S. Amann for conducting the pyocyanin assay, A. Abdulmughni for CO-spectral analysis and N. Zaburannyi for bioinformatic support. J. Hug acknowledges funding by a PhD fellowship of the Boehringer Ingelheim Fonds. Research in R. Müller's laboratory is funded by the Deutsche Forschungsgemeinschaft (DFG) and the Bundesministerium für Bildung und Forschung (BMBF) and the Deutsches Zentrum für Infektionsforschung Standort Hannover-Braunschweig.

### 3.8 Notes

The authors declare no conflict of interest.

### 3.9 Key words

myxobacteria, biosynthesis, *Myxococcus xanthus* DK1622, natural products, RiPPs, cittilin, cytochrome P450, biaryls.

### 3.10 References

1. Herrmann, J., Fayad, A. A. & Müller, R. Natural products from myxobacteria: novel metabolites and bioactivities. *Nat. Prod. Rep.* **34**, 135–160 (2017).
2. Wenzel, S. C. & Müller, R. Myxobacteria—'microbial factories' for the production of bioactive secondary metabolites. *Mol. Biosyst.* **5**, 567–574 (2009).
3. Witte, S. N. R., Hug, J. J., Géraldy, M., Müller, R. & Kalesse, M. Biosynthesis and Total Synthesis of Pyrrolic B: a Secondary Metabolite from *Nannocystis pusilla*. *Chem. Eur. J.* **23**, 15917–15921 (2017).
4. Blin, K. *et al.* antiSMASH 5.0: updates to the secondary metabolite genome mining pipeline. *Nucleic Acids Res.*, W81–W87 (2019).
5. Mohr, K. I. *et al.* Pinensins: the first antifungal lantibiotics. *Angew. Chem. Int. Ed. Engl.* **127**, 11254–11258 (2015).
6. Huo, L. & van der Donk, Wilfred A. Discovery and Characterization of Bicereucin, an Unusual D-Amino Acid-Containing Mixed Two-Component Lantibiotic. *J. Am. Chem. Soc.* **138**, 5254–5257 (2016).

7. Meindl, K. *et al.* Labyrinthopeptins: a new class of carbacyclic lantibiotics. *Angew. Chem. Int. Ed.* **49**, 1151–1154 (2010).
8. DiCaprio, A. J., Firouzbakht, A., Hudson, G. A. & Mitchell, D. A. Enzymatic Reconstitution and Biosynthetic Investigation of the Lasso Peptide Fusilassin. *J. Am. Chem. Soc.* **141**, 290–297 (2018).
9. Huo, L., Rachid, S., Stadler, M., Wenzel, S. C. & Müller, R. Synthetic biotechnology to study and engineer ribosomal bottromycin biosynthesis. *Chem. Biol.* **19**, 1278–1287 (2012).
10. Franz, L., Adam, S., Santos-Aberturas, J., Truman, A. W. & Koehnke, J. Macroamidine Formation in Bottromycins Is Catalyzed by a Divergent YcaO Enzyme. *J. Am. Chem. Soc.* **139**, 18158–18161 (2017).
11. Smith, T. E. *et al.* Accessing chemical diversity from the uncultivated symbionts of small marine animals. *Nat. Chem. Biol.* **14**, 179 (2018).
12. Agrawal, P., Khater, S., Gupta, M., Sain, N. & Mohanty, D. RiPPMiner: a bioinformatics resource for deciphering chemical structures of RiPPs based on prediction of cleavage and cross-links. *Nucleic Acids Res.* **45**, W80–W88 (2017).
13. Tietz, J. I. *et al.* A new genome-mining tool redefines the lasso peptide biosynthetic landscape. *Nat. Chem. Biol.* **13**, 470–478 (2017).
14. Kloosterman, A. M., Shelton, K. E., van Wezel, G. P., Medema, M. H. & Mitchell, D. A. RRE-Finder: A Genome-Mining Tool for Class-Independent RiPP Discovery. *bioRxiv* (2020).
15. Los Santos, E. L. C. de. NeuRiPP: Neural network identification of RiPP precursor peptides. *Sci. Rep.* **9**, 13406 (2019).
16. Merwin, N. J. *et al.* DeepRiPP integrates multiomics data to automate discovery of novel ribosomally synthesized natural products. *Proc. Natl. Acad. Sci. USA* **117**, 371–380 (2020).
17. Santos-Aberturas, J. *et al.* Uncovering the unexplored diversity of thioamidated ribosomal peptides in Actinobacteria using the RiPPER genome mining tool. *Nucleic Acids Res.* **47**, 4624–4637 (2019).
18. Viehrig, K. *et al.* Structure and biosynthesis of crocagins: polycyclic postranslationally modified ribosomal peptides from *Chondromyces crocatus*. *Angew. Chem.*, 1–5 (2017).
19. Gong, Y. *et al.* Competitive Interactions Between Incompatible Mutants of the Social Bacterium *Myxococcus xanthus* DK1622. *Front. Microbiol.* **9**, 1200 (2018).
20. Bader, C. D., Panter, F. & Müller, R. In depth natural product discovery - Myxobacterial strains that provided multiple secondary metabolites. *Biotechnol Adv.* **39**, 107480 (2020).
21. Cortina, N. S., Krug, D., Plaza, A., Revermann, O. & Müller, R. Myxoprincomide: a natural product from *Myxococcus xanthus* discovered by comprehensive analysis of the secondary metabolome. *Angew. Chem. Int. Ed. Engl.* **51**, 811–816 (2012).
22. Krug, D. *et al.* Discovering the Hidden Secondary Metabolome of *Myxococcus xanthus*: a Study of Intraspecific Diversity. *Appl. Environ. Microbiol.* **74**, 3058–3068 (2008).

23. Irschik, H., Trowitzsch-Kienast, W., Gerth, K., Höfle, G. & Reichenbach, H. Saframycin Mx1, a new natural saframycin isolated from a myxobacterium. *J. Antibiot.* **41**, 993–998 (1988).
24. Trowitzsch-Kienast, W. (ed.). *Citalins: Bicyclic Isotriptyrosines from Myxococcus xanthus*. 24th General meeting German Chemists' Society (1993).
25. Reichenbach, H. & Höfle, G. in *Drug Discovery from Nature*, edited by S. Grabley & R. Thiericke (Springer, Berlin, 1999), pp. 149–179.
26. Helynck, G., Dubertret, C., Frechet, D. & Leboul, J. Isolation of RP 66453, a new secondary peptide metabolite from *Streptomyces* sp. useful as a lead for neurotensin antagonists. *J. Antibiot.* **51**, 512–514 (1998).
27. Boissard, S., Carbonnelle, A. C. & Zhu, J. P. Studies on the total synthesis of RP 66453: Synthesis of fully functionalized 15-membered biaryl-containing macrocycle. *Org. Lett.* **3**, 2061–2064 (2001).
28. Boissard, S. & Zhu, J. P. Studies toward the total synthesis of RP-66453. *Tetrahedron Lett.* **43**, 2577–2580 (2002).
29. Krenitsky, P. J. & Boger, D. L. Preparation of the 14-membered L,L-cycloisodityrosine subunit of RP 66453. *Tetrahedron Lett.* **43**, 407–410 (2002).
30. Krenitsky, P. J. & Boger, D. L. Synthesis of the (S,S,S)-diastereomer of the 15-membered biaryl ring system of RP 66453. *Tetrahedron Lett.* **44**, 4019–4022 (2003).
31. Greule, A. *et al.* Kistamicin biosynthesis reveals the biosynthetic requirements for production of highly crosslinked glycopeptide antibiotics. *Nat. Commun.* **10**, 1–15 (2019).
32. Silva, L. N., Zimmer, K. R., Macedo, A. J. & Trentin, D. S. Plant Natural Products Targeting Bacterial Virulence Factors. *Chem. Rev.* **116**, 9162–9236 (2016).
33. Barnard, F. M. *et al.* Global regulation of virulence and the stress response by CsrA in the highly adapted human gastric pathogen *Helicobacter pylori*. *Mol. Microbiol.* **51**, 15–32 (2004).
34. Altier, C., Suyemoto, M. & Lawhon, S. D. Regulation of *Salmonella enterica* serovar typhimurium invasion genes by *csrA*. *Infect. Immun.* **68**, 6790–6797 (2000).
35. Baker, C. S., Morozov, I., Suzuki, K., Romeo, T. & Babitzke, P. CsrA regulates glycogen biosynthesis by preventing translation of *glgC* in *Escherichia coli*. *Mol. Microbiol.* **44**, 1599–1610 (2002).
36. Maurer, C. K. *et al.* Discovery of the first small-molecule CsrA-RNA interaction inhibitors using biophysical screening technologies. *Future Med. Chem.* **8**, 931–947 (2016).
37. Krug, D., Zurek, G., Schneider, B., Garcia, R. & Müller, R. Efficient mining of myxobacterial metabolite profiles enabled by liquid chromatography-electrospray ionization-time-of-flight mass spectrometry and compound-based principal component analysis. *Anal. Chim. Acta* **624**, 97–106 (2008).

38. Revermann, O. Novel secondary metabolites from myxobacteria and their biosynthetic machinery. Dissertation. Saarland University, 2012.
39. Sharma, G., Narwani, T., Subramanian, S. & Gao, F. Complete genome sequence and comparative genomics of a novel myxobacterium *Myxococcus hansupus*. *PLoS ONE* **11**, e0148593 (2016).
40. Arnison, P. G. *et al.* Ribosomally synthesized and post-translationally modified peptide natural products: overview and recommendations for a universal nomenclature. *Nat. Prod. Rep.* **30**, 108–160 (2013).
41. Myronovskyi, M. *et al.* Generation of a cluster-free *Streptomyces albus* chassis strains for improved heterologous expression of secondary metabolite clusters. *Metab. Eng.* **49**, 316–324 (2018).
42. Ortega, M. A. & van der Donk, Wilfred A. New Insights into the Biosynthetic Logic of Ribosomally Synthesized and Post-translationally Modified Peptide Natural Products. *Cell Chem. Biol.* **23**, 31–44 (2016).
43. Shan, L., Martin, T., Sollid, L. M., Gray, G. M. & Khosla, C. Comparative biochemical analysis of three bacterial prolyl endopeptidases: implications for coeliac sprue. *Biochem. J.* **383**, 311–318 (2004).
44. Shan, L., Mathews, I. I. & Khosla, C. Structural and mechanistic analysis of two prolyl endopeptidases: Role of interdomain dynamics in catalysis and specificity. *Proc. Natl. Acad. Sci. USA* **102**, 3599–3604 (2005).
45. Kocadag Kocazorbaz, E. & Zihnioğlu, F. Purification, characterization and the use of recombinant prolyl oligopeptidase from *Myxococcus xanthus* for gluten hydrolysis. *Protein Expr. Purif.* **129**, 101–107 (2017).
46. Danielson, P. B. The Cytochrome P450 Superfamily: Biochemistry, Evolution and Drug Metabolism in Humans. *Curr. Drug Metab.* **3**, 561–597 (2002).
47. Newman, D. J. & Cragg, G. M. Natural Products as Sources of New Drugs over the Nearly Four Decades from 01/1981 to 09/2019. *J. Nat. Prod.* **83**, 770–803 (2020).
48. Clerc, F.-F., Dubroeuq, M.-C., Helynck, G., Leboul, J. & Martin, J.-P. (EN) *Neurotensin Antagonist Peptides (FR) Peptides Antagonistes de la Neurotensine*. Patent WO/1995/032218 (1995).
49. Morkunas, B. *et al.* Discovery of an inhibitor of the production of the *Pseudomonas aeruginosa* virulence factor pyocyanin in wild-type cells. *Beilstein J. Org. Chem.* **12**, 1428–1433 (2016).
50. Schütz, C. & Empting, M. Targeting the *Pseudomonas* quinolone signal quorum sensing system for the discovery of novel anti-infective pathoblockers. *Beilstein J. Org. Chem.* **14**, 2627–2645 (2018).
51. Donner, J. *et al.* The Biofilm Inhibitor Carolacton Enters Gram-Negative Cells: Studies Using a TolC-Deficient Strain of *Escherichia coli*. *mSphere* **2** (2017).
52. Imai, Y. *et al.* A new antibiotic selectively kills Gram-negative pathogens. *Nature* **576**, 459–464 (2019).



# Chapter 3

## Supporting Information

### **Biosynthesis of cittilins, unusual ribosomally synthesized and post-translationally modified peptides from *Myxococcus xanthus***

**Joachim J. Hug<sup>1,3</sup>**, Jan Dastbaz<sup>1,3</sup>, Sebastian Adam<sup>2</sup>, Ole Revermann<sup>1,3</sup>, Jesko Koehnke<sup>2</sup>, Daniel Krug<sup>1,3</sup> and Rolf Müller<sup>1,3\*</sup>

1 Department Microbial Natural Products, Helmholtz-Institute for Pharmaceutical Research Saarland (HIPS), Helmholtz Centre for Infection Research (HZI) and Department of Pharmacy Saarland University, Campus E8.1, 66123 Saarbrücken, Germany

2 Department Structural Biology of Biosynthetic Enzymes, Helmholtz-Institute for Pharmaceutical Research Saarland (HIPS), Helmholtz Centre for Infection Research (HZI) and Department of Pharmaceutical Biotechnology Saarland University, Campus E8.1, 66123 Saarbrücken, Germany

3 German Center for Infection Research (DZIF), Partner Site Hannover-Braunschweig, Germany

\* Correspondence: rolf.mueller@helmholtz-hips.de, Tel.: +4968198806-3000

### Table of Contents

1 General materials and methods .....	155
1.1 Applied software, sequence analysis and bioinformatics methods .....	155
1.2 Maintenance of bacterial cultures, molecular cloning and construction of plasmids <sup>14</sup> .....	156
Bacterial cultures and preparation of cryogenic long-term stocks .....	163
Crude extracts preparation for analysis of secondary metabolism.....	167
1.3 Compound isolation .....	168
Analysis during purification .....	168
Size exclusion chromatography via gel filtration .....	168
Semi-preparative HPLC chromatography .....	168
2 Results.....	169
2.1 NMR spectroscopic data.....	169
2.2 Bioinformatics investigation of different cittilin A and B producing myxobacterial strains .....	171
2.3 Biosynthetic investigation of the cittilin biosynthesis in the native producer.....	177
2.4 Heterologous production of cittilin .....	179
2.5 <i>In vitro</i> and cell free lysate investigation of cittilin biosynthesis .....	182
2.6 Biological function of cittilin A.....	198
Cell based bioactivity profiling.....	198
Synthesis and purification of rhodamine-coupled cittilin A .....	200
Bacterial cell entry test .....	200
2.7 <sup>1</sup> H and <sup>13</sup> C NMR spectra of cittilin A and rhodamine-coupled cittilin A.....	203
3. References .....	208



## 1 General materials and methods

### 1.1 Applied software, sequence analysis and bioinformatics methods

Geneious prime (Biomatters Ltd., Auckland, New Zealand) was used to design primers for PCR and sequencing, to create plasmid maps, to find open reading frames (ORF) and to predict molecular weights of proteins. Furthermore all pairwise and multiple alignments of nucleotide or amino acid sequences were performed with the plugin software from Geneious by using the MUSCLE (Multiple Sequence Comparison by Log- Expectation) alignment (3.8.425 by Robert C. Edgar) since it claims to achieve higher average accuracy and better speed than ClustalW2 or T-Coffee algorithm <sup>1</sup>. In order to find homologous genes or proteins, either the nucleotide or amino acid sequence of interest was aligned with the basic local alignment search tool (BLAST) against our in-house genome database or the publically available nucleotide database. Raw data from alignments for *in silico* evaluation of citilin biosynthetic gene clusters (BGCs) are deposited on our in-house server. The functional prediction of ORFs was performed by either using protein blast and/or blastx program <sup>2</sup>. The in-house standard extract database embedded in the software bundle Mxbase Explorer 3.2.27 <sup>3</sup> was used for the search of alternative producers of citilins. The molecular formula and experimentally determined retention times of ions typically observed from citilins were used as data input. The BGC sequences of confirmed producers of citilin deriving from our in-house database have been deposited in GenBank and are accessible under the accession number as displayed in **Table S1**. In addition, the sequence of the citilin BGC originating from *Myxococcus xanthus* DK1622 has been deposited in the Minimum Information about a Biosynthetic Gene cluster (MIBiG) database under the accession number BGC0002043.

**Table S1.** In-house genome sequence analysis of myxobacterial strains

Citilin A producing strain	Reference	Accession number BGC
<i>M. xanthus</i> Mxx48 (MCy8278)	<sup>4</sup>	MN731370
<i>Archangium violaceus</i> Cb vi35 (MCy8337)	<sup>5</sup>	MN731363
<i>Archangium gephyra</i> (MCy8375)	<sup>6</sup>	MN731362
<i>Myxococcus fulvus</i> (MCy10608)	<sup>7</sup>	MN731365
<i>Myxococcus xanthus</i> DK897 (MCy8986)	<sup>8</sup>	MN731369
<i>Myxococcus xanthus</i> DK1622 (MCy9151)	<sup>9</sup>	MN731368
<i>Myxococcus fulvus</i> strain ATCC BAA-855/HW-1 (MCy11108)	<sup>10</sup>	NC_015711.1
<i>Myxococcus virescens</i> ST200611 (MCy11474)	<sup>11</sup>	MN731367
<i>Myxococcus</i> sp. MCy9171	<sup>12</sup>	MN731366
<i>Myxococcus fulvus</i> MCy8286 (Mxf65)	<sup>13</sup>	MN731364

## 1.2 Maintenance of bacterial cultures, molecular cloning and construction of plasmids<sup>14</sup>

Routine handling of nucleic acids, such as isolation of plasmid DNA, restriction endonuclease digestions, DNA ligations, and other DNA manipulations, was performed according to the standard protocols<sup>15</sup>. *Escherichia coli* HS996 (Invitrogen) was used as host for standard cloning experiments and *E. coli* SCS110 (Stratagene) for preparation of plasmid DNA free of Dam or Dcm methylation. *E. coli* strains were cultured in LB liquid medium or on LB agar (1% tryptone, 0.5% yeast extract, 0.5% NaCl, (1.5% agar) at 30–37 °C and 200 rpm) overnight (o/n). Antibiotics were used at the following final concentrations: 100 µg/mL ampicillin, 50 µg/mL kanamycin, 12 µg/mL oxytetracycline, 50 µg/mL apramycin, 25 µg/mL chloramphenicol and 25 µg/mL nalidixic acid. Transformation of *E. coli* strains was achieved via electroporation in 0.1 cm wide cuvettes at 1250 V, a resistance of 200 Ω, and a capacity of 25 µF. Plasmids were purified either by standard alkaline lysis<sup>15</sup> or by using the GeneJet Plasmid Miniprep Kit (Thermo Fisher Scientific™) or the NucleoBond PC100 kit (Macherey-Nagel). Restriction endonucleases, alkaline phosphatase (FastAP) and T4 DNA ligase were purchased from Thermo Fisher Scientific™. Oligonucleotides used for PCR and sequencing were acquired from Sigma-Aldrich and are listed in **Table S2 and Table S3**. PCRs were carried out in a Mastercycler® pro (Eppendorf) using Phusion™ High-Fidelity according to the manufacturer's protocol. Temperature and duration setting for each thermocycling step in PCR with Phusion™ High-Fidelity polymerase were performed as follows: Initial denaturation (30 s, 98 °C); 33 cycles of denaturation (15 s, 98 °C), annealing (15 s, 53–72 °C, depending on the melting temperature of primers) and elongation (based on PCR product length 30 s/1 kb, 72 °C); and final extension (10 min, 72 °C). When amplification of desired PCR product could not be achieved according to the manufacturer's instructions, additives like DMSO (3–8%), glycerol (4–8%) and betaine (0.5 mM) were supplemented. PCR products or DNA fragments from restriction digestions were purified by agarose gel electrophoresis and isolated using the PCR clean-up gel extraction kit using Nucleo Spin® (Macherey-Nagel). To recover DNA fragments larger than 8 kb, the Agarose Gel DNA Extraction Kit from Roche was used. After selection with suitable antibiotics, clones harboring correct ligation products were identified by plasmid isolation and restriction analysis with a set of different restriction endonucleases. In addition to restriction analysis, integrity of the constructs for induced gene expression was verified by sequencing.

### *Streptomyces* conjugation procedure

Conjugation procedure between *E. coli* and *Streptomyces* was performed following a procedure adapted from Mazodier et al.<sup>16</sup>. For conjugation SM-agar plates were used. The plates were prepared one week before in order to dry the plates. *E. coli* ET12567 harboring pUZ8002 was transformed with the respective plasmid/genetic construct for heterologous production of cittilin in *Streptomyces albus* del14 and *Streptomyces coelicolor* CH999 for recombinant protein production (cittilin cytochrome P450 enzyme). O/n cultures of *E. coli* ET12567 + pUZ8002 harboring the genetic construct of interest were prepared: 5 mL of 2TY medium were inoculated with the respective *E. coli* strains from a cryogenic long term stock with 25 µg/mL chloramphenicol, 50 µg/mL kanamycin and 50 µg/mL apramycin. The next day, the o/n cultures were used with 0.75 mL, 1.00 mL and 1.25 mL as inoculation volume to inoculate 20 mL of 2TY medium with same antibiotics and concentrations for (o/n) cultures. When OD<sub>600</sub> reached 0.4, the cultures were centrifuged for 10 min at 4000 rpm. The supernatant (SN) was discarded and the cell pellet (CP) was re-suspended in 10 mL 2TY medium. Applying same centrifugation parameters, the cells were centrifuged, the SN was discarded and the CP was washed with 2TY medium again. After centrifugation according the same parameters as before, the SN was discarded and the CP was re-suspended in 1 mL 2TY medium and stored on ice.

Spores of *S. coelicolor* CH999 or *S. albus* del14, which have been aliquoted and frozen at -80 °C were thawed on ice. The spore aliquots of about 100 µL were washed with 1 mL 2TY medium by pipetting up and down and centrifuged for 10 min at 6000 rpm. A second washing procedure with 1 mL of 2TY medium followed by centrifugation for 2 min at 6000 rpm was performed. After discarding the SN, the spore pellet was re-suspended in 0.5 mL 2TY medium and heat shock procedure was conducted for 5 min at 50 °C in a HLC thermo-block heater and then stored on ice. *Streptomyces* spores and *E. coli* were combined in a volume ratio of 2:1; 0.5 mL spores and 0.25 mL of *E. coli* cells were suspended in an Eppendorf tube and 350 µL of the mixture was plated on SM agar plates. On the next day, each plate was overlaid with 25 µg/mL nalidixic acid and 50 µg/mL apramycin (final concentration in the plate) in 1 mL MQ-H<sub>2</sub>O. The plates were incubated at 30 °C. As soon as exconjugants of *Streptomyces* were detected, these single colonies were transferred to a new SM-agar plate with 25 µg/mL nalidixic acid and 50 µg/mL apramycin. After these isolated exconjugants showed vital growth, these cultures were transferred again to a SM-agar plate containing this time only 50 µg/mL apramycin. This agar culture was incubated at 30°C until sporulation occurred.

**Table S2.** List of oligonucleotides used in this study.

No.	Primer name	Primer sequence 5'–3'
1	JHuFw_ <i>citT<sub>a</sub></i> _DK1622_KO	ATATAAGCTTGACATGTTCCCCAGCACCT
2	JHuRv_ <i>citT<sub>a</sub></i> _DK1622_KO	ATATACTAGTTTCATTGTCCAGCACCTGCC
3	JHuFw_ <i>citB</i> _DK1622_KO	ATATAAGCTTGACCTGATTGAACGCGTCTCT
4	JHuRv_ <i>citB</i> _DK1622_KO	ATATACTAGTCCGAACGGGAAGTAGACGTA
5	JHuFw_ <i>citC</i> _DK1622_KO	ATATAAGCTTAGTTCTTCCGCTCGGCATACC
6	JHuRv_ <i>citC</i> _DK1622_KO	ATATACTAGTGGGTCTGGTCTGCTCGTAGTG
7	JHuFw_ <i>mx<sub>pep</sub></i> _KO	ATATAAGCTTCCTGGGACGGCAAGAAGGTG
8	JHuRv_ <i>mx<sub>pep</sub></i> _KO	ATATACTAGTCGGTGCTGACGGACGTCTTGT AG
9	JHuFw_ <i>citA</i> _activate	ATATCATATGAAGAAGGCCCTGTACTCTTTG
10	JHuRv_ <i>citA</i> _activate	ATATGAATTCCGTGCCCCGGACATGATCATC
11	JHuFw_ <i>citT<sub>a</sub></i> _activate	ATATCATATGAGTCCATCCAACAGGCGTT
12	JHuRv_ <i>citT<sub>a</sub></i> _activate	ATGAATTCATCTGCATGATGAGGCCCG
13	JHuFw_PermE*	ATATTTAATTAAAAGCGAGCGAAGCCACTG AG
14	JHuRv_PermE*	ATATTCTAGAATATCTTAAGGGCCATATGTG GGGTCCTCC
15	JHuFw_ <i>citlin</i> _operon_NdeI	ATATCATATGAAGAAGGCCCTGTACTCTTTG G
16	JHuRv_ <i>citlin</i> _operon_BspTI	ATATCTTAAGTCGAACTGCTGGCGGAGTGA
17	JHuRv_ <i>citlin</i> _operon_BspI407I	ATATTGTACATCGAACTGCTGGCGGAGTG
18	JHuFw_DK1622_ <i>mx<sub>pep</sub></i>	ATATTGTACACCCGGTGTTGTCTGGTTGAC
19	JHuRv_DK1622_ <i>mx<sub>pep</sub></i>	ATATCTTAAGCATCAAGGGAACACCCGAGG

20	JHuFw_DK1622_ <i>citC</i>	ATATCATATGGAAAACCTGTATTTTCAGGGC GGCATGCGGCGGGAGCATGAAGG
21	JHuRv_DK1622_ <i>citC</i>	ATATGAATTCCGCCTGAGTCAGTCCTCTGTC G
22	JHuFw_DK1622_ <i>mx_pep</i>	ATATCATATGTCCTACCCGGCGACC
23	JHuRv_DK1622_ <i>mx_pep</i>	ATATAAGCTTTCAGCGGCCCTGCGCCGCCAC
24	JHuFw_DK1622_ <i>citB</i>	ATATCCATGGGGTTGGGTCTCAAGAGCTGG TCGA
25	JHuRv_DK1622_ <i>citB</i>	ATATAAGCTTCATGCTCCCGCCGCATGC
26	JHuFw_DK1622_ <i>citB</i> _Strepto	CTTCATATGGAAAACCTGTATTTTCAGGGCG GCATGGAGCGCATGGTTCGCT
27	JHuRv_DK1622_ <i>citB</i> _Strepto	ATGAATTCCCTCATGCTCCCGCCGC
28	JHuFw_Cbvi35_ <i>citB</i> _Strepto	CTTCATATGGAAAACCTGTATTTTCAGGGCG GCATGGCGGGGGAGGACTCCT
29	JHuRv_Cbvi35_ <i>citB</i> _Strepto	ATAAGCTTTCATGCGCTCACCTCTTG
30	JHuFw_MCy9171_ <i>citB</i> _Strepto	CTTCATATGGAAAACCTGTATTTTCAGGGCG GCATGTGACGCGGGAGGAAT
31	JHuRv_MCy9171_ <i>citB</i> _Strepto	ATAAGCTTTCAGGCTCGTGCTGCATG
32	JHuFw_pCJW93_exchange_noHis	CGTCAGACCCCGTAGAAAAG
33	JHuRv_pCJW93_exchange_noHis	ATATCATATGTGTCCGCTCCCTTCTCTG
34	JDa14Rv_pCJW93_test1	CGCTGCTGTGATGATGATG
35	JDa14Rv_pCJW93_test2	GATGATGATGATGGCTGCTG
36	JHuFw_DK1622_ <i>citB</i> _Strepto_no_stop_ C-His6	CTTCATATGGAGCGCATGGTTCGCT
37	JHuRv_DK1622_ <i>citB</i> _Strepto_no_stop_ C-His6	ATGAATTCTCAATGGTGATGGTGATGGTGG CCGGCGCCTGCTCCCGCCGCATG
38	JHuFw_Cbvi35_ <i>citB</i> _Strepto_no_stop_ C-His6	CTTCATATGGCGGGGGAGGACTCCT
39	JHuRv_Cbvi35_ <i>citB</i> _Strepto_no_stop_ C-His6	ATGAATTCTCAATGGTGATGGTGATGGTGG CCGGCGCCTGCGCTCACCTTTGAGGAAG
40	JHuFw_MCy9171_ <i>citB</i> _Strepto_no_stop_ C-His6	CTTCATATGTGACGCGGGAGGAAT
41	JHuRv_MCy9171_ <i>citB</i> _Strepto_no_stop_ C-His6	ATGAATTCTCAATGGTGATGGTGATGGTGG CCGGCGCCGGCTCGTGCTGCATGCC

**Table S3.** List of oligonucleotides used as sequencing primers.

No.	Primer name	Primer sequence 5'–3'
1	Test_Fw1_DK1622_citT <sub>a</sub> _KO	CGACTACCGTGAGTCCATCCAACA
2	Test_Rv2_DK1622_citT <sub>a</sub> _KO	TCCACGCGTTGACCGTCTGAA
3	Test_Rv3_DK1622_citT <sub>a</sub> _KO	TCTACGTGTTCCGCTTCCTTTAGCAG
4	Test_Fw4_DK1622_citT <sub>a</sub> _KO	CCTTTGAGTGAGCTGATACCGCTCG
5	Test_Fw1_DK1622_citB_KO	TGGGTCTCAAGAGCTGGTCGAC
6	Test_Rv2_DK1622_citB_KO	CAGGTATGCCGAGCGGAAGAAC
7	Test_Rv3_DK1622_citB_KO	ATTCAGGCTGCGCAACTGTTGG
8	Test_Fw4_DK1622_citB_KO	GCCACCTCTGACTTGAGCGTC
9	Test_Fw1_DK1622_citC_KO	CTCTCACACAACACAGGAAGGCG
10	Test_Rv2_DK1622_citC_KO	CTTCACCTGCTGTCACCCCG
11	Test_Rv3_DK1622_citC_KO	CGCCCAGTCTAGCTATCGCCA
12	Test_Fw4_DK1622_citC_KO	CTTCCGGCTCGTATGTTGTGTGG
13	Test_Fw1_DK1622_mx_pep_KO	CCACAAGGACAAGGAGAAGGCC
14	Test_Rv2_DK1622_mx_pep_KO	CGTTGCTGCCGCCGTAGAT
15	Test_Rv3_DK1622_mx_pep_KO	TCTTCGCTATTACGCCAGCTGG
16	Test_Fw4_DK1622_mx_pep_KO	CGTATTACCGCCTTTGAGTGAGCTG
17	Test_Fw1_DK1622_citA_activate	CTGTGCCACATTTACCGCG
18	Test_Rv2_DK1622_citA_activate	CTCACCGCCAGCATCGTTCC
19	Test_Rv3_DK1622_citA_activate	CTGCGTTATCCCCTGATTCTGTGG
20	Test_Fw4_DK1622_citA_activate	TGTCAAGCTGCTGTTTTTCGCCG
21	Test_Fw1_DK1622_citT <sub>a</sub> _activate	AGAAGAAACACCCAGGCTTTGAC
22	Test_Rv2_DK1622_citT <sub>a</sub> _activate	CACCATCATCTTCCCGAGCA
23	Test_Rv3_DK1622_citT <sub>a</sub> _activate	TTAGCTCACTCATTAGGCACCC
24	Test_Fw4_DK1622_citT <sub>a</sub> _activate	GGATCCAATAGGTCGCCGAA
25	Test_Fw1_Seq_pCJW93_citB	GACAAAACCTTTAGATCTGGG
26	Test_Rv2_Seq_pCJW93_citB	GAACGTCCGGGCTTGAC

**Table S4.** List of plasmids used in this study.

No.	Plasmid name/ characteristic	Size [kb]	Function	Reference
1	pCR2.1-TOPO (EcoRI, religated)	3.931	Used as template to PCR-amplify <i>tn5_kanR</i> gene	TOPO®TA Cloning® Kit Thermo Fisher Scientific™
2	pFP <sub>van</sub> <i>pcyA</i>	6.181	pCR2.1-TOPO derivative used as vector to ligate PCR products for subsequent vanillate-induced gene expression in <i>M. xanthus</i> DK1622	<sup>17</sup>
3	pSET152	5.549	Shuttle vector for expression of secondary metabolites in <i>Streptomyces</i> , harboring <i>apraR</i>	<sup>18</sup>
4	pAB03	5.429	Used as template to PCR-amplify P <sub>ERME</sub> *	<sup>19</sup>
5	pET-28b	5.368	Used as expression vector for recombinant protein production	(Novagen)
6	pHisTEV	5.365	Used as expression vector for recombinant protein production	<sup>20,21</sup>
7	pCJW93	8.672	Used as expression vector for recombinant protein production in <i>Streptomyces coelicolor</i> CH999	<sup>22</sup> , Kindly provided by Peter F. Leadlay

**Table S5.** List of PCR-amplified constructs.

No.	PCR product name/ characteristic	Size [bp]	Template	Primer used
1	<i>citT<sub>a</sub></i> _homology	593	gDNA from <i>M. xanthus</i> DK1622	primer No.1 primer No. 2
2	<i>citB</i> _homology	994	gDNA from <i>M. xanthus</i> DK1622	primer No. 3 primer No. 4
3	<i>citC</i> _homology	531	gDNA from <i>M. xanthus</i> DK1622	primer No. 5 primer No. 6

4	<i>mx_pep_homology</i>	1021	gDNA from <i>M. xanthus</i> DK1622	primer No. 7 primer No. 8
5	<i>citA_activate</i>	970	gDNA from <i>M. xanthus</i> DK1622	primer No. 9 primer No. 10
6	<i>citT<sub>a</sub>_activate</i>	1226	gDNA from <i>M. xanthus</i> DK1622	primer No. 11 primer No. 12
7	P <sub>ermE</sub> *	470	pAB03erm*	primer No.13 primer No.14
8	DK1622_ <i>citA</i> – <i>citC</i> _operon_NdeI_BspTI	2233	gDNA from <i>M. xanthus</i> DK1622	primer No.15 primer No.16
9	DK1622_ <i>citA</i> – <i>citC</i> _operon_NdeI_Bsp1407I	2233	gDNA from <i>M. xanthus</i> DK1622	primer No.15 primer No.17
10	DK1622_ <i>mx_pep</i> _Bsp1407I_ BspTI	2407	gDNA from <i>M. xanthus</i> DK1622	primer No.18 primer No.19
11	Cittilin_DK1622_CitC_recombinant_protein	667	gDNA from <i>M. xanthus</i> DK1622	primer No. 20 primer No. 21
12	Cittilin_DK1622_MX_PEP_recombinant_protein	2087	gDNA from <i>M. xanthus</i> DK1622	primer No. 22 primer No. 23
13	Cittilin_DK1622_CitB_recombinant_protein	1461	gDNA from <i>M. xanthus</i> DK1622	primer No. 24 primer No. 25
14	Cittilin_DK1622_CitB_recombinant_protein_Strepto	1456	gDNA from <i>M. xanthus</i> DK1622	primer No. 26 primer No. 27
15	Cittilin_Cbvi35_CitB_recombinant_protein_Strepto	1475	gDNA from <i>Cystobacter</i> Cb vi35	primer No. 28 primer No. 29
16	Cittilin_Mcy9171_CitB_recombinant_protein_Strepto	1475	gDNA from MCy9171	primer No. 30 primer No. 31
17	pCJW93_noHistag_exchange_construct	1388	pCJW93	primer No. 32 primer No. 33
18	Cittilin_DK1622_CitB_recombinant_protein_Strepto_no_stop_C-His	1472	gDNA from <i>M. xanthus</i> DK1622	primer No. 36 primer No. 37
19	Cittilin_Cbvi35_CitB_recombinant_protein_Strepto_no_stop_C-His	1393	gDNA from <i>Cystobacter</i> Cb vi35	primer No. 38 primer No. 39
20	Cittilin_Mcy9171_CitB_recombinant_protein_Strepto_no_stop_C-His	1471	gDNA from MCy9171	primer No. 40 primer No. 41

**Table S6.** List of genetic constructs generated in this study.

No.	Plasmid name	Construction details/ characteristic
1	pCR2.1-TOPO_ <i>citTa</i> _KO	Construct obtained by conventional restriction ligation of plasmid pCR2.1-TOPO and PCR product No. 1
2	pCR2.1-TOPO_ <i>citB</i> _KO	Construct obtained by conventional restriction ligation of plasmid pCR2.1-TOPO and PCR product No. 2
3	pCR2.1-TOPO_ <i>citC</i> _KO	Construct obtained by conventional restriction ligation of plasmid pCR2.1-TOPO and PCR product No. 3
4	pCR2.1-TOPO_ <i>mx_peg</i> _KO	Construct obtained by conventional restriction ligation of plasmid pCR2.1-TOPO and PCR product No. 4
5	pFPVan_ <i>citA</i> _activate	Construct obtained by conventional restriction ligation of plasmid pFPVan_ncyA and PCR product No. 5
6	pFPVan_ <i>citTa</i> _activate	Construct obtained by conventional restriction ligation of plasmid pFPVan_ncyA and PCR product No. 6
7	pSET152_P <sub>ermE</sub> *	Construct obtained by conventional restriction ligation of plasmid pSET152 and PCR product No. 7
8	pSET152_P <sub>ermE</sub> *_ <i>citA</i> - <i>citC</i> _operon	Construct obtained by conventional restriction ligation of construct No. 7 and PCR product No. 8
9	pSET152_P <sub>ermE</sub> *_ <i>citA</i> - <i>citC</i> _operon_ <i>mx_peg</i>	Construct obtained by conventional restriction ligation of construct No. 7 and PCR products No. 8 and No. 9
10	pET28b_DK1622_CitC	Construct obtained by conventional restriction ligation of plasmid pET28b and PCR product No. 11
11	pET28b_DK1622_MX_PEP	Construct obtained by conventional restriction ligation of plasmid pET28b and PCR product No. 12
12	pHisTEV_DK1622_CitB	Construct obtained by conventional restriction ligation of plasmid pHisTEV and PCR product No. 13
13	pCJW93_DK1622_CitB	Construct obtained by conventional restriction ligation of plasmid pCJW93 and PCR product No. 14
14	pCJW93_Cbvi35_CitB	Construct obtained by conventional restriction ligation of plasmid pCJW93 and PCR product No. 15
15	pCJW93_MCy9171_CitB	Construct obtained by conventional restriction ligation of plasmid pCJW93 and PCR product No. 16
16	pCJW93noHis	Construct obtained by conventional restriction ligation of plasmid pCJW93 and PCR product No. 17
17	pCJW93noHis_DK1622_CitB	Construct obtained by conventional restriction ligation of genetic construct No. 16 and PCR product No. 18
18	pCJW93noHis_Cbvi35_CitB	Construct obtained by conventional restriction ligation of genetic construct No. 16 and PCR product No. 19



19	pCJW93noHis_MCy9171_CitB	Construct obtained by conventional restriction ligation of genetic construct No. 16 and PCR product No. 20
----	--------------------------	--

### Bacterial cultures and preparation of cryogenic long-term stocks

Myxobacteria were preserved at -80 °C. Ten mL of liquid culture was transferred to a 2 mL Eppendorf tube and centrifuged for 2 min, 7000 rpm. 1.5 mL of SN was discarded, leaving 0.5 mL SN in the Eppendorf tube. The CP was re-suspended by pipetting up and down and then mixed with 0.5 mL glycerol (50%). The mixture was transferred into a cryogenic vial.

For preservation of spores of *Streptomyces* spp. at -80 °C, 2 mL of glycerol (20%) was added on a sporulating agar culture. The glycerol was spread with a cell spreader on the plate to collect the spores. The glycerol-spore mixture was afterwards transferred from plate via pipetting into a cryogenic vial.

Bacterial strains for the experiments of this work, their relevant characteristics and sources are listed in the **Table S7**

**Table S7.** Bacterial strains used in this study.

Strain	Function	Origin
<i>E. coli</i> HS996	Standard cloning host	Invitrogen
<i>E. coli</i> ET12567	Donor for <i>Streptomyces</i> conjugation	ATCC® BAA-525™
<i>E. coli</i> Lemo21	Recombinant protein production	Novagen
<i>E. coli</i> BL21	Recombinant protein production	Novagen
<i>Myxococcus xanthus</i> DK1622	Investigated strain in this study	Internal strain collection
<i>Streptomyces albus</i> del14	Host for heterologous production of citilin A	Kindly provided by Andriy Luzhetskyy <sup>23</sup>
<i>Streptomyces coelicolor</i> CH999	Host for recombinant protein production	Kindly provided by Peter F. Leadlay <sup>24</sup>
<i>E. coli</i> (TolC-deficient mutant)	Gram-negative bacterium to test bacterial cell entry	Internal strain collection <sup>25</sup>

### Cultivation media and buffers

The pH of all cultivation media was adjusted before autoclaving. Autoclaving was conducted at 121°C for 20 min.

#### 2TY-medium

Ingredient	Concentration
Tryptone	16 g/L
Yeast extract	20 g/L
NaCl	5 g/L

pH 6.8 +/- 0.2 with NaOH<sub>aq</sub>

#### COM-medium

Ingredient	Concentration
Glucose	25 g/L
Soybean flour	25 g/L
Baker's yeast (fresh)	3 g/L
NaCl	2 g/L
(NH <sub>4</sub> ) <sub>2</sub> SO <sub>4</sub>	2 g/L
CaCO <sub>3</sub>	2 g/L
K <sub>2</sub> HPO <sub>4</sub>	0.15 g/L

pH 8.4 +/- 0.2 with NaOH<sub>aq</sub>

#### CTT-medium

Ingredient	Concentration
Casitone	10 g/L
Tris	10 mM
KH <sub>2</sub> PO <sub>4</sub>	1 mM
MgSO <sub>4</sub>	8 mM

pH 7.6 with KOH<sub>aq</sub>

#### LB-medium

Ingredient	Concentration
Yeast extract	5 g/L
Tryptone	10 g/L
NaCl	5 g/L

pH 7.2 with KOH<sub>aq</sub>

## SM-agar

Ingredient	Concentration
D-mannitol	20 g/L
Soybean flour	20 g/L
MgCl <sub>2</sub> *6H <sub>2</sub> O	10 mM
Agar	20 g/L

## Super YEME-medium

Ingredient	Concentration
Yeast extract	3 g/L
Peptone	5 g/L
Glucose	10 g/L
Malt extract	3 g/L
Sucrose	340 g/L
Glycine	5 g/L
MgCl <sub>2</sub> x 6H <sub>2</sub> O	2.35 g/L
L-proline	75 mg/L
L-arginine	75 mg/L
L-cysteine	75 mg/L
L-histidine	100 mg/L
Uracil	15 mg/L

## TSB-medium

Ingredient	Concentration
Tryptic soy broth	30 g/L

## Lysis buffer

Ingredient	Concentration
NaCl	500 mM
BIS-TRIS	20 mM
Imidazole	20 mM
Glycerol	10% (m/m)

pH 6.8 with NaOH<sub>aq</sub>

### Elution buffer

<b>Ingredient</b>	<b>Concentration</b>
NaCl	500 mM
BIS-TRIS	20 mM
Imidazole	250 mM
Glycerol	10% (m/m)

pH 6.8 with NaOH<sub>aq</sub>

### Protein buffer

<b>Ingredient</b>	<b>Concentration</b>
NaCl	500 mM
BIS-TRIS	20 mM
Glycerol	5% (m/m)

pH 6.8 with NaOH<sub>aq</sub>

### Lysis buffer No. 3

<b>Ingredient</b>	<b>Concentration</b>
NaCl	500 mM
BIS-TRIS	20 mM
Imidazole	20 mM

pH 6.8 with NaOH<sub>aq</sub>

### Elution buffer No. 3

<b>Ingredient</b>	<b>Concentration</b>
NaCl	500 mM
BIS-TRIS	20 mM
Imidazole	250 mM

pH 6.8 with NaOH<sub>aq</sub>

### Protein buffer No. 3

<b>Ingredient</b>	<b>Concentration</b>
NaCl	500 mM
BIS-TRIS	20 mM

pH 6.8 with NaOH<sub>aq</sub>

## Crude extracts preparation for analysis of secondary metabolism

*Myxococcus xanthus* DK 1622 and mutant strains were grown on CTT agar plates for 3–5 days at 30 °C. For the preparation of liquid pre-cultures, a suitable portion of overgrown agar was transferred to inoculate 20 mL of CTT medium in an Erlenmeyer flask (300 mL) and incubated at 30 °C and 200 rpm for 2–4 days. 1 mL of the pre-culture was used to inoculate 50 mL of CTT medium containing appropriate antibiotic selection and 2% amberlite resin XAD-16 (Sigma Aldrich).

Spores of *Streptomyces albus* del14 and mutant strains were used to inoculate 20 mL of TSB medium in an Erlenmeyer flask (300 mL) and incubated at 30 °C and 200 rpm for 2–4 days. 1–2 mL of the pre-culture was used to inoculate 50 mL of COM medium containing appropriate antibiotic selection and 2% amberlite resin XAD-16 (Sigma Aldrich).

In order to obtain statistically significant results, three independent transformants of each recombinant strain were selected and fermentations were performed at least in triplicates. The resulting fermentation broth were harvested by centrifugation at 8000 rpm for 10 min (Eppendorf centrifuge 5810 R). The SN was discarded, whereas the residual material was extracted first with 25 mL MeOH, stirred for 1 h, filtered through filter paper (folded filters grade: 3hw from Sartorius) into a round bottom flask and afterwards this procedure was repeated with 25 mL acetone. The solvent of the filtered extracts was removed under vacuum (BÜCHI Rotavapor R-210) and the extracts re-dissolved in 1.5 mL MeOH and stored at -20 °C. The re-dissolved extract were diluted with MeOH (1:3 (extract/MeOH (v:v))) centrifuged at 13000 g for 10 min (VWR centrifuge ECN521-3601, Hitachi Koki Co., Ltd) and 1 µL of the SN was subjected to HPLC-MS analysis as described further below.

## Analysis of secondary metabolites in broth extracts, and enzymatic reactions

The secondary metabolism of broth extracts was analyzed by HPLC-HRESI-DAD-MS on a Bruker maXis 4G mass spectrometer coupled with a Dionex Ultimate 3000 RSLC system using a BEH C18 column (100 × 2.1 mm, 1.7 µm, Waters, Germany) with a gradient of 5–95% acetonitrile (ACN) + 0.1% formic acid (FA) in H<sub>2</sub>O + 0.1% FA at 0.6 mL/min and 45 °C over 9 or 18 min with UV detection by a diode array detector at 200–600 nm. Mass spectra were acquired from 150 to 2000 m/z at 2 Hz. The detection was performed in the positive MS mode. The plugin for Chromeleon Xpress (Dionex) was used for operation of UltiMate 3000 LC System. HyStar (Bruker Daltonic) was used to operate on maXis 4G speed MS system. HPLC-MS mass spectra were analyzed with DataAnalysis 4.2 (Bruker Daltonic).

## Analysis of recombinant proteins via LC-MS

Recombinant proteins were analyzed on a Dionex Ultimate 3000 UPLC system (Thermo Scientific™) coupled with an maXis4G Q-TOF MS (Bruker) using an ESI in positive mode. The samples were run on a Aeris Widepore XB-C8, 150 x 2.1 mm, 3.6 µm dp column (Phenomenex, USA). LC conditions: A: H<sub>2</sub>O dd + 0.1% FA; B: Acetonitrile + 0.1% FA at a flow rate of 300 µL/min and 45 °C. 0 min: 98% A / 2% B, 0.5 min: 98% A / 2% B, 10.5 min: 25% A / 75% B, 13.5 min: 25% A / 75% B, 14 min: 98% A / 2% B. The LC flow was split to 75 µL/min before entering the maXis4G hr-ToF mass spectrometer (Bruker Daltonics, Bremen, Germany) using the standard Bruker ESI source. In the source region, the temperature was set to 180 °C, the capillary voltage was 4000 V, the dry-gas flow was 6.0 L/min and the nebulizer was set to 1.1 bar. Mass spectra were acquired in positive ionization mode ranging from 150–2500 m/z at 2.5 Hz scan rate. Protein masses were deconvoluted by using the Maximum Entropy algorithm (Copyright 1991-2004 Spectrum Square Associates, Inc.).

### 1.3 Compound isolation

#### Analysis during purification

All measurements to analyze the mass of cittilin A and B during purification were performed on a Dionex Ultimate 3000 RSLC system coupled to the amaZon iontrap MS using a BEH C18, 100 x 2.1mm, 1.7  $\mu$ 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 (FA) to (**B**) ACN + 0.1% FA at a flow rate of 0.6 mL/min and 45 °C. The gradient was initiated by a 0.5 min isocratic step at 5% **B**, followed by linear increase to 95% **B** in 18 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 200–600 nm. The detection was performed in the positive ESI MS/MS mode.

#### Size exclusion chromatography via gel filtration

The CP of a 10 L fermentation culture with 2% XAD-16 was used to prepare a crude extract by mixing the CP with 400 mL (MeOH), stirring for 1 h and filtering through glass wool into a round bottom flask. This extraction was repeated once; afterwards the extraction fraction was dried using the rotary evaporator (BÜCHI Rotavapor R–210). The dried residue was re-dissolved MeOH, in order to run the obtained extract on a methanolic Sephadex® LH20 column. The cittilin-containing fraction was placed on a silica gel column, which was run with a CHCl<sub>3</sub>:MeOH gradient (starting 9:1 (v/v), ending with pure MeOH). For further purification another Sephadex® LH20 column was performed with H<sub>2</sub>O/ MeOH (3:2).

#### Semi-preparative HPLC chromatography

Semi-preparative HPLC purification was done using a Dionex Ultimate 3000 SDLC low pressure gradient system on a XBridge® peptide BEH C18 OBD™ prep column (138 Å, 5  $\mu$ m, 10 mm x 250 mm). Column temperature was stabilized at 45 °C with the eluents H<sub>2</sub>O + 0.1% FA as **A** and ACN + 0.1% FA as **B**, at a flow rate of 1.5 mL/min. Detection of cittilin A was facilitated via mass spectrometry on the Agilent 1100 series coupled to the HCT 3D ion trap or with a UV detector on the Dionex 3000 SL systems by UV absorption at 256 nm and 320 nm. The gradient starts with a plateau at 95% **A** for 2 min followed by a ramp to 73% **A** during 5.3 min. Then, **A** content was kept to 73% during 11 min and finally ramped to 5% **A** during 2 min. **A** content is kept at 5% for 2 min and then ramped back to 95% during 30 s. The column was re-equilibrated at 95% **A** for 3 min. The white compound was subsequently dried by lyophilization yielding cittilin A and B.

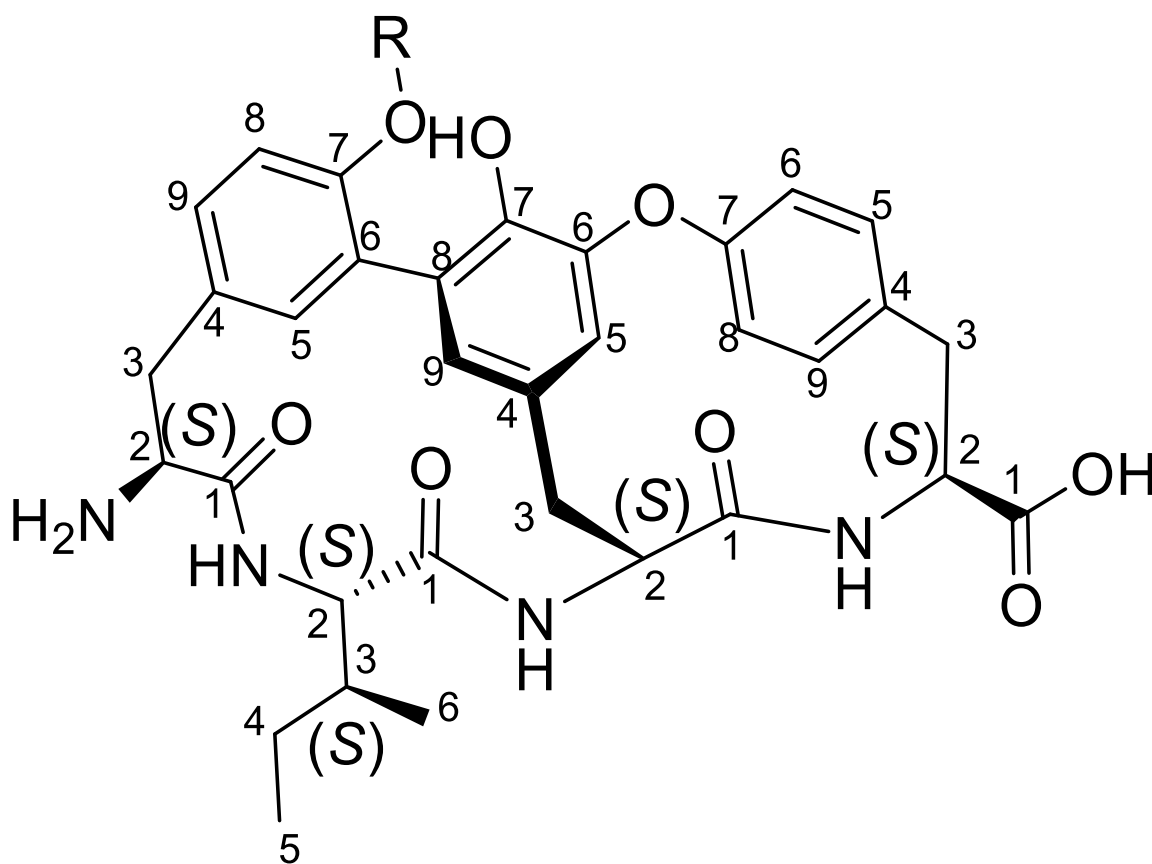
## 2 Results

### 2.1 NMR spectroscopic data

**Table S8.**  $^1\text{H}$  and  $^{13}\text{C}$  NMR spectral data of cittilin A and B; recorded at 500 and 125 MHz in  $\text{CD}_3\text{OD}$ ; d: doublet; dd: doublet of doublets; s: singlet; m: multiplet.

Amino acid	Position	Cittilin A $^1\text{H}$ (J) ppm (Hz)	$^{13}\text{C}$ ppm	Cittilin B $^1\text{H}$ (J) ppm (Hz)	$^{13}\text{C}$ ppm	Multiplicity
<b>L-tyrosine 1</b>	1-CO	---	169.9	---	170.1	---
	2-CH	4.21	53.8	4.08	54.1	M
	3a-CHH	3.46	36.7	3.41	37.6	M
	3b-CHH	3.05 (6/15)	36.7	2.99 (6/15)	37.6	dd
	4-C	---	129.7	---	128.0	---
	5-CH	6.92	137.0	6.90	137.0	m
	6-C	---	125.9	---	125.6	---
	7-C	---	157.9	---	155.9	---
	8-CH	7.21 (2/8)	131.7	7.19 (2/8)	131.7	dd
	9-CH	7.01 (9)	112.0	6.98 (8)	111.8	d
	O-CH <sub>3</sub>	3.88	56.3	---	---	s
<b>L-isoleucine</b>	1-CO	---	169.9	---	171.7	---
	2-CH	4.20	60.1	4.16	60.0	m
	3-CH	1.77	37.1	1.77	37.1	m
	4a-CHH	1.61	26.3	1.62	26.4	m
	4b-CHH	1.27	26.3	1.26	26.4	m
	5-CH <sub>3</sub>	0.95 (7)	15.7	0.96 (6)	15.9	m
	6-CH <sub>3</sub>	0.92 (7)	11.0	0.92 (7)	10.9	t
<b>L-tyrosine 2</b>	1-CO	---	173.0	---	171.5	---
	2-CH	3.68	59.8	3.68	59.8	m
	3a-CHH	3.10 (2/14)	40.7	3.16 (2/13)	40.7	dd
	3b-CHH	2.47 (4/13)	41.0	2.47 (4/14)	40.9	dd
	4-C	---	136.6	---	135.4	---
	5-CH	5.78 (2)	120.7	5.80 (2)	120.8	d
	6-C	---	154.2	---	152.6	---
	7-C	---	145.2	---	143.3	---
	8-CH	---	129.7	---	129.6	---
<b>L-tyrosine 3</b>	9-CH	6.62 (2)	126.3	6.63 (2)	126.4	d
	1-CO	---	176.8	---	175.2	---
	2-CH	4.30	54.2	4.26	54.7	m
	3a-CHH	3.61	37.3	3.61	37.4	m
	3b-CHH	2.98 (7/14)	37.3	2.94 (m)	37.4	dd

	4-C	---	135.5	---	133.9	---
	5-CH	7.22	135.4	7.19	135.4	m
	6-C	6.83 (3/8)	124.8	6.79 (3/9)	124.6	dd
	7-C	---	164.5	---	162.5	---
	8-CH	7.47 (2/8)	127.0	7.45 (3/9)	127.0	dd
	9-CH	7.35 (2/8)	130.7	7.36 (2/8)	131.0	dd



cittilin A (R=CH<sub>3</sub>)

cittilin B/RP-66453 (R=H)

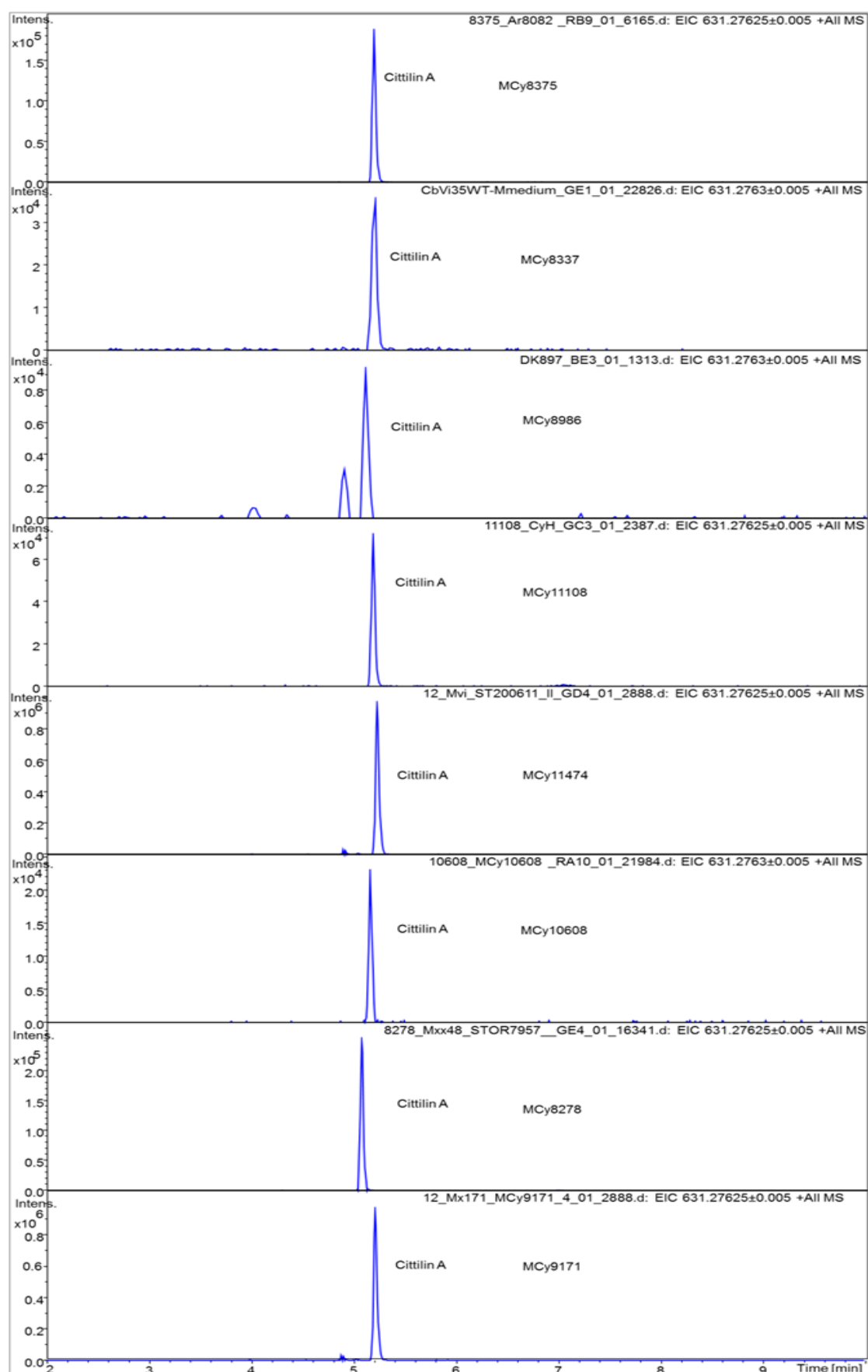
**Figure S1.** Structure and carbon numbering of cittilin A and B.



## 2.2 Bioinformatics investigation of different citilin A and B producing myxobacterial strains

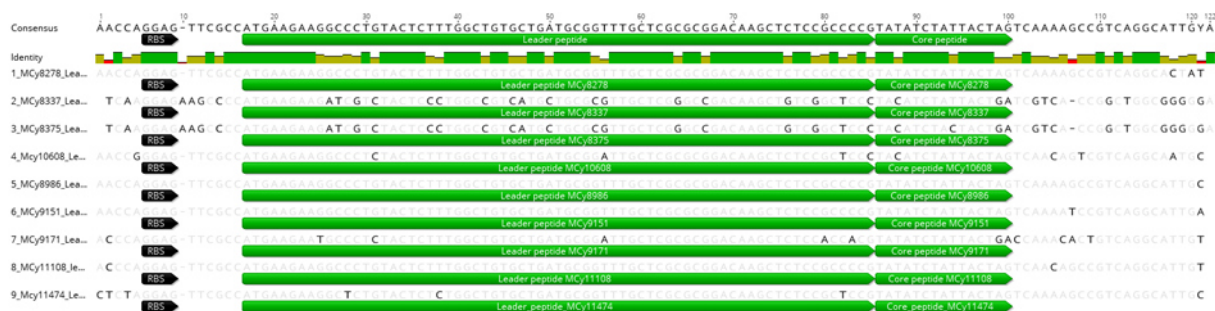
The biosynthetic pathway of citilin and their genetic organization was investigated by *in silico* analysis. As previously described by Krug et al.<sup>26</sup> several producers of citilins are known, in particular the well described *M. xanthus* DK1622 (MCy9151), but also the closely related strains DK897 (MCy8986) and ST200611 (MCy11474). All those strains are LC-MS-confirmed producers with available whole genome sequences. In order to find more producer strains, the translated nucleotide sequence of the biosynthetically involved cytochrome P450 enzyme (CitB) from the characterized producer MCy9151 was used as query input to find via discontinuous MEGABLAST in the private MINS genome database and the public available database further candidate BGCs for citilin A.

The list of similar citB homologs in different strains showed several potential results in the private MINS genome database and three hits in the public database, however only hits with identical sites up to 78% and with query coverage higher than 90% were considered as relevant hits. The following strains from the private MINS genome database were further investigated: MCy8278, MCy8337, MCy8375, MCy10608, MCy9171. From the publically available genome sequences eight hits were found. Six of these hits are strains, which are phylogenetically closely related to *M. xanthus* DK1622 (*Myxococcus xanthus* strain KF3.2.8c11, *Myxococcus xanthus* strain GH5.1.9c20, *Myxococcus xanthus* strain MC3.5.9c15, *Myxococcus xanthus* strain MC3.3.5c16, *Myxococcus xanthus* strain GH3.5.6c2, *Myxococcus xanthus* strain KF4.3.9c1)<sup>27</sup>, whereas *Myxococcus fulvus* HW-1 (MCy11108) and *Myxococcus hansupus mixupus* are phylogenetically less related. Roughly around 100 bp upstream of *citB*, a sequence encoding the core peptide YIYY followed by a stop codon could be found associated with all those hits, except for the strain *Myxococcus hansupus mixupus*<sup>28</sup>. Remarkably, it has a nucleotide sequence encoding the tetrapeptide YHYIY. The precursor peptide (CitA), the cytochrome P450 enzyme (CitB) and the methyltransferase (CitC) amino acid sequence are highlighting high similarity to the characterized sequence from the strain MCy9151 that implies the BGC might be actively expressing a natural citilin derivative. Since no metabolomic data or the strain itself is available, it was unfortunately not possible to search for this putative derivative. Interestingly the hit showing the highest percentage of identical sites to the query sequence derived from the strain A47 with 97% by showing at the same time very low query coverage of 12.96%. This strain has already been described as a non-producer of citilins; however the genome harbors the same precursor peptide architecture like the confirmed citilin producer strain MCy9151<sup>29</sup>. The cytochrome P450 enzyme gene is truncated which is plausibly responsible for the lack of any citilin in the metabolome of A47 and responsible for the low query coverage in the BLAST result. For this reason for further investigations only strains characterized by genomic data and being confirmed producers of citilins by previously performed LC-MS measurements (**Fig. S2**), were considered for further *in silico* investigations. The following strains were used for in-depth *in silico* characterization: 1) MCy8278, 2) MCy8337, 3) MCy8375, 4) MCy10608, 5) MCy8986, 6) MCy9171, 8) MCy11108, 9) MCy11474. One benefit of having several BGCs is to use the information to determine the cluster borders. The final confirmation of the assumed borders is heterologous expression of the specific BGC in a heterologous host (assuming that this host is not capable of complementing any involved biosynthetic enzyme).



**Fig. S2** HPLC-MS extracted ion chromatograms of cittilin producers. The retention time and mass of cittilin A of 631.2768  $m/z$  with a width of 7.9 ppm  $m/z$  in different strains with available genome data is shown. Cittilin B was detectable as well in lower concentration (not displayed).

To validate *a priori* that the precursor peptide gene (*citA*) consists of the 23 amino acid (aa) leader peptide gene and the four aa core peptide gene, the precursor peptide sequences of those nine different confirmed citilin producer strains were aligned. The alignment is highlighting the identity among all producers within the leader and core peptide genes. Interestingly the putative ribosome binding site (RBS) sequence A(G)/GGAG is conserved in all citilin producers. Directly upstream of the RBS and downstream of the core peptide the identity is significantly decreasing. Another hint for the start of peptide translation is the increasing 3<sup>rd</sup>-position GC content of all analyzed sequences<sup>30</sup>. Therefore, it can be assumed that for heterologous expression experiments the precursor peptide and the core peptide genes are sufficiently well defined for this work. Interestingly, the comparison of different leader peptide sequences showed two different residues upstream the core peptide. Unlike the other leader peptide sequences, the sequence in the strain MCy9171 is encoding the amino acids TT instead of AP. This position is likely critical for recognition of the putatively involved prolyl endopeptidase (MX PEP) to process the precursor peptide CitA. Therefore, for future *in vitro* studies with recombinant enzymes involved in the formation of citilins, one has to consider the possibility that substrate specificity of the peptidase might deviate among different strains.

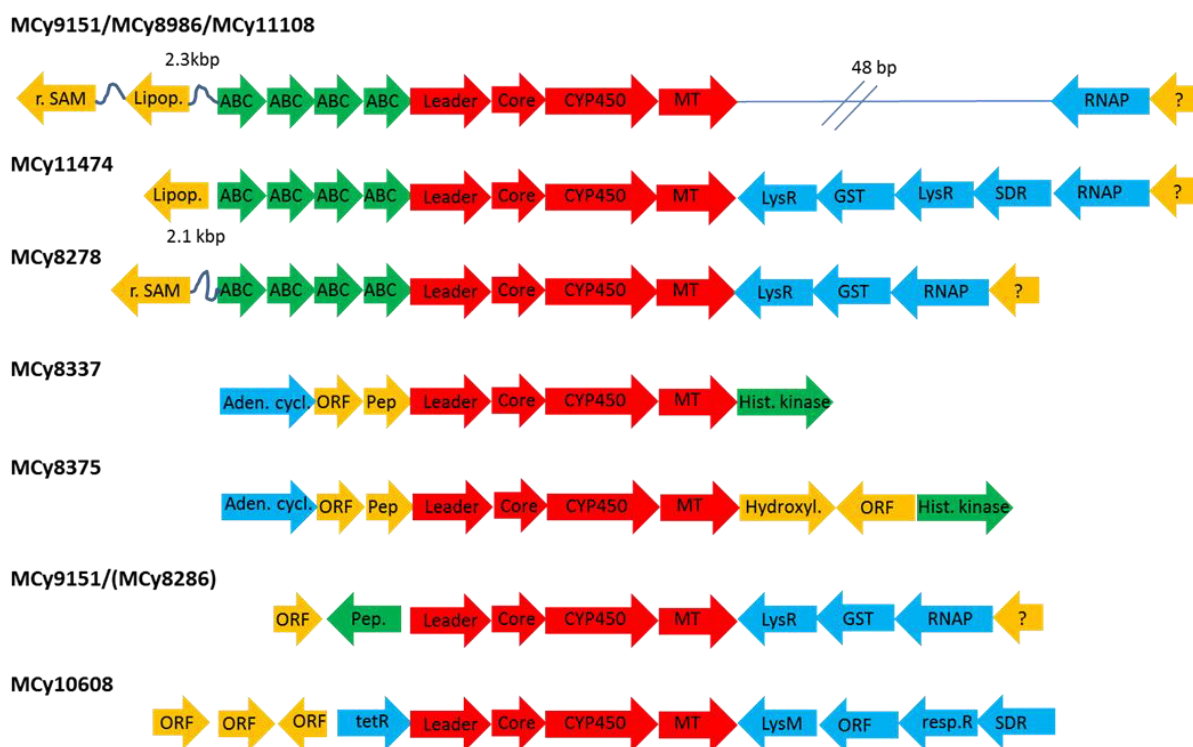


**Figure S3.** Nucleotide alignment of precursor peptide sequence of nine confirmed citilin producers. The nucleotide sequence is highly conserved in particular the putative ribosome binding site (RBS) sequence A(G)/GGAG is conserved in all citilin producers. Directly upstream of the RBS and downstream of the core peptide the similarity is significantly decreasing.



**Figure S4.** Protein alignment of precursor peptide sequences of nine confirmed citilin producers. The amino acid sequence is highly conserved in particular the core peptide residues in the middle. However, the two different residues downstream the core peptide can vary. Unlike the other leader peptide sequences, the sequence in the strain MCy9171 is encoding the amino acids TT instead of AP (region indicated by red ring). Directly upstream of the core peptide the amino acid identity is significantly decreasing.

To narrow down essential genes for the production of citilin, nine citilin gene loci from confirmed producers were analyzed in detail. The strains MCy8278, MCy9151, MCy8986, MCy11474 and the non-producing strain A47 harboring the truncated cytochrome P450 enzyme gene *citB*, possess downstream of the precursor peptide gene *citA* four genes encoding an ABC transporter (*citT<sub>a</sub>*–*citT<sub>d</sub>*). The ABC transporter sequences are sharing high similarity in those strains and might play a plausible role for export of produced citilin. Interestingly there exists another type of genetic organization, which has no ABC transporter upstream of the precursor peptide gene but an operon with an adenylate cyclase, a hypothetical protein and a peptidase. The BGCs which do not have an ABC transporter upstream of the precursor gene might have the gene somewhere else located in the genome, like it has already been demonstrated for the prolyl endopeptidase *mx pep*. MCy8837 and MCy8375 are sharing this gene cluster organization. The strain MCy10608 features upstream of the precursor peptide a tetracycline regulator gene and downstream of the methyltransferase gene *citC* a peptidoglycan-binding *lysM* gene. The surrounding genomic area in the strain of MCy10608 is unique in comparison to other citilin producers. The producer strain MCy9171 and non-producing strain MCy8286 are not only very similar in the remarkable variation of their leader peptide but also the surrounding area is nearly identical.

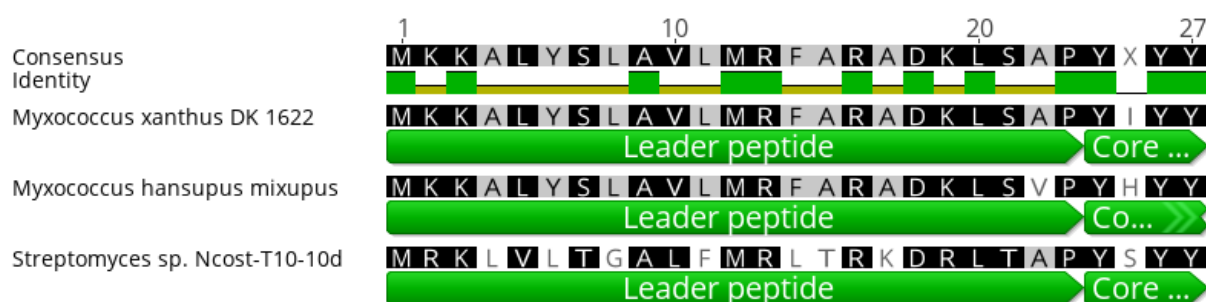


**Figure S5.** Schematic overview of citilin BGCs in different strains. Fundamentally the structural genes consisting of the precursor peptide gene (displayed separately), the cytochrome P450 enzyme and the methyltransferase, are present in the BGC. It seems to be that the ABC transporter upstream of the precursor peptide gene is conserved for the strains MCy9151, MCy8986, MCy11108, MCy11474 and MCy8278. The strains MCy9151, MCy8986 and MCy11108 are sharing a highly similar architecture not only the structural genes but also the surrounding genes 10 kb up- and downstream of the structural genes are almost identical according to the nucleotide alignments. Red: structural crucial genes, Green: Putative resistance conferring genes, Blue: Putative regulatory genes, Orange: Unknown function. ABC: ABC transporter, MT: Methyltransferase, rSAM: radical *S*-adenosyl-L-methionine (SAM) containing protein, Lipop: Lipoprotein, RNAP: RNA polymerase, GST: Glutathione *S* transferase, SDR: Short-chain dehydrogenase/reductase, Hist. kinase: Histidine kinase, ORF: Open reading frame or hypothetical protein, *tetR*: Tetracycline regulator, Aden. cycl: Adenylate cyclase.

Amino acid comparison of the cytochrome P450 enzyme (CitB) homologs, emphasized highly conserved residues, especially the heme-binding domain motif FxxGxRxCxG(S)<sup>31</sup> and the ExxR motif in the k-helix, which is important for the stabilization of the meander loop and the tertiary structure of the cytochrome<sup>32,33</sup>. Additionally it became clear that the translation most probably starts with the amino acid sequence motif MP/S(QVR/TAL)LP, whereas the C-terminal end of the sequence shows decreasing similarity with higher variability. Comparison of MEGABLAST search results showed that the sequence is conserved. The primary amino acid sequence of the cytochrome P450 enzymes of any of those nine strains was used to search in the RCSB PDB database for structurally elucidated homologous proteins. The hit with the highest similarity (27 % identity for MCy9151-CitB) was found for the albaflavenone monooxygenase deriving from *Streptomyces coelicolor* A3(2)<sup>34</sup>.

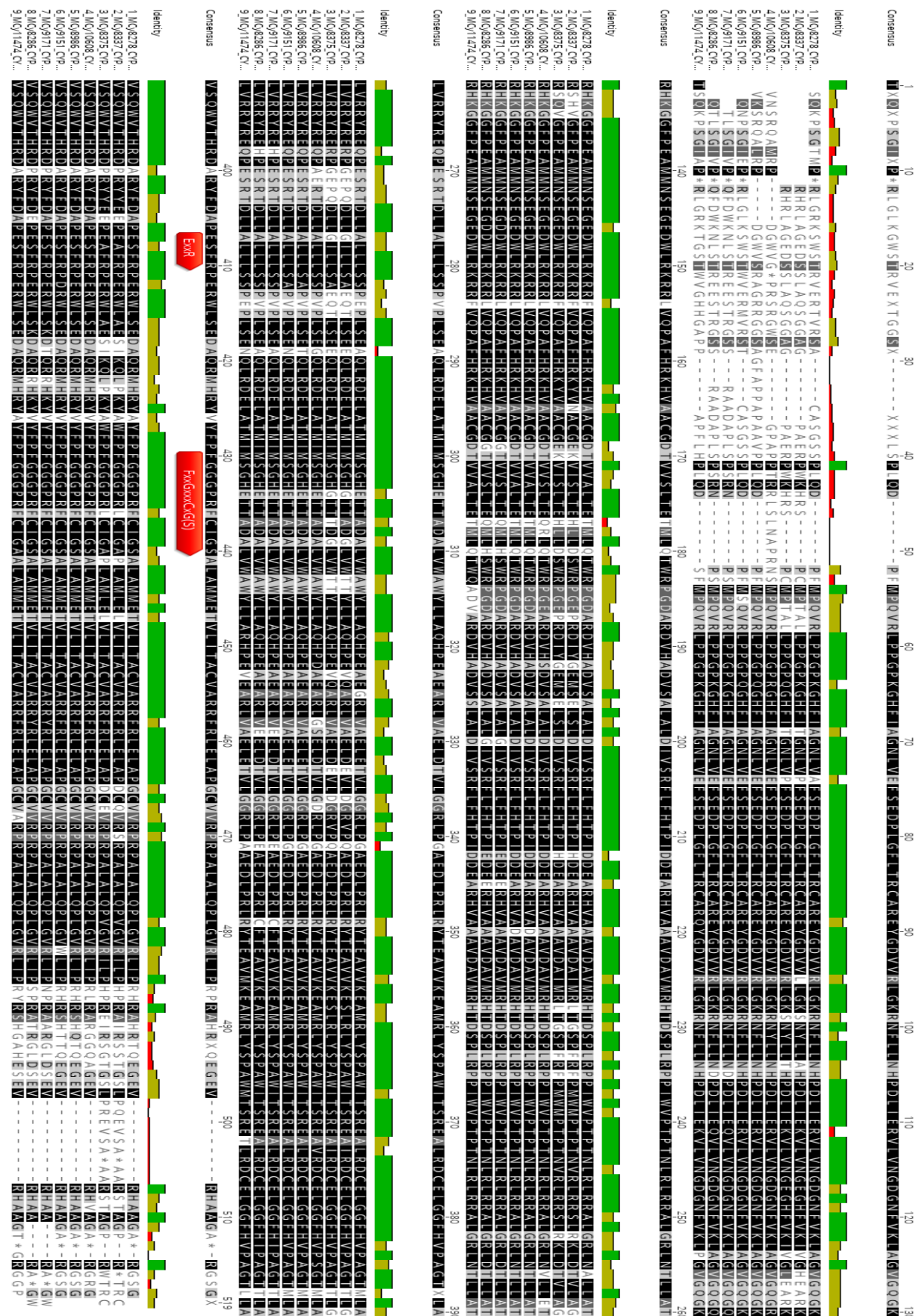
The albaflavenone monooxygenase is catalyzing two steps in the biosynthesis, first the hydroxylation of epi-isozizaene to the individual epimers of albaflavenol and second the oxidation of both epimers to the same final antibiotic albaflavenone. Interestingly within the typical cytochrome P450 enzyme structure the active site of the functional farnesene synthase is located, catalyzing the conversion from farnesyl diphosphate to *E*- $\beta$ -farnesene (61%), (3*E*,6*E*)- $\alpha$ -farnesene (26%), (3*Z*,6*E*)- $\alpha$ -farnesene (6.8%), as well as nerolidol (4.9%) and farnesol (1.8%). However the two putative Mg<sup>2+</sup> binding motifs the aspartate-rich sequence DDXX(D/E) and the residues (N/D)DXX(S/T)XXXE, (NSE or DTE triad), which are conserved in all terpene synthases of microbial and plant origin could not be found in any of the cytochrome P450 homologs<sup>35</sup>. For the recombinant overexpression of the cytochrome P450 enzyme gene from the strains MCy9151, MCy9171 and MCy8337 the translation start was set earlier than the sequence alignment suggested. One reason for this approach is the low increase of the 3<sup>rd</sup>-letter GC content across the conserved residues.

### Putative cittilin derivatives



**Figure S6.** Protein alignment of precursor peptide sequences of putative cittilin derivative producers. The leader peptide amino acid sequence of *M. xanthus* DK1622 and *M. hansupus mixupus* are almost identical except for the motif DKLSAPY (*M. hansupus mixupus* DKLSVPY), whereas the leader peptide amino acid sequence of *Streptomyces* sp. Ncost-T10 shows more dissimilarities.





**Figure S7.** Protein alignment of CitB homologs from nine confirmed citilin producer strains. The comparison shows clearly that the translation most probably starts with the amino acid sequence motif MP/S(QVR/TAL)LP. At the C-terminal end of the sequence, the similarity is decreasing significantly and the variability is relatively high.

## 2.3 Biosynthetic investigation of the cittilin biosynthesis in the native producer

### Generation of disruption and induced-gene constructs to connect the identified genes to cittilin production

A 631–1021 bp homology sequence of the genes encoding the ABC transporter (only *citT<sub>a</sub>*, to abolish tetrameric ABC transporter expression), the cytochrome P450 enzyme (*citB*), the methyltransferase (*citC*) and the prolyl endopeptidase (*mx\_ppe*) has been PCR amplified by the primers as shown in **Table S2**. The specific homology sequence was subcloned via conventional restriction ligation into the pCR2.1 vector from the TOPO-TA cloning kit (Thermo Scientific™ TOPO-TA cloning Kit). Similarly to the creation of disruption constructs, a 1200 bp homology sequence starting from the translational start (not the RBS, but the coding sequence) of the genes encoding CitT<sub>a</sub> or the identified precursor peptide CitA has been PCR amplified by the primers as shown in **Table S2**. The homology sequence was subcloned via conventional restriction ligation into the pFP<sub>van</sub> vector, which has been constructed and utilized previously to express several genes in myxobacteria<sup>17,36,37</sup>. The pFP<sub>van</sub> vector is a derivative of the pCR2.1 vector featuring a vanillate-inducible promoter which is fused to the vanillate-responsive repressor as described in literature<sup>38</sup>.

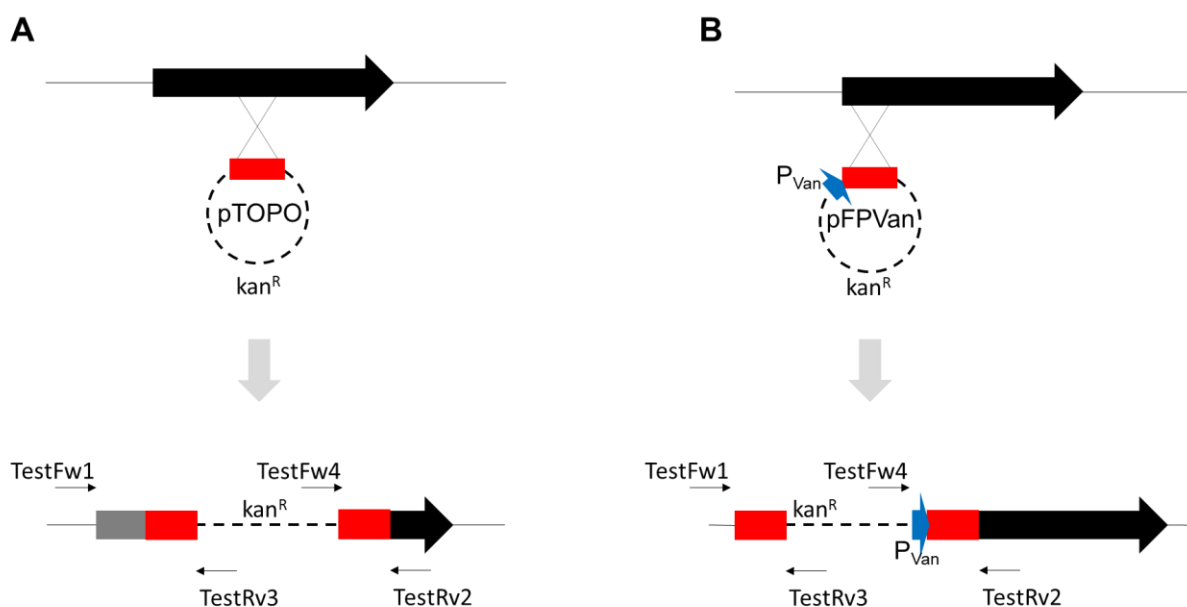
### Transfer and chromosomal integration of the constructs into the host *Myxococcus xanthus* DK1622

According to a previously established electroporation procedure for *Myxococcus xanthus* DK1622<sup>39</sup> the strain *M. xanthus* DK1622 was transformed with the generated disruption and induced-gene constructs (**Table S6**, genetic constructs 1–6). *M. xanthus* DK1622 transformants were routinely cultivated at 30 °C in CTT medium or CTT agar. Liquid cultures were grown in Erlenmeyer flasks on an orbital shaker at 180 rpm for 3–6 days. *M. xanthus* DK1622 transformants were selected by adding 50 µg/µL kanamycin to the fermentation culture. Correct chromosomal integration of the expression constructs via homologous recombination into the site-specific locus was confirmed by PCR (**Figure S8**). PCRs were performed according to the settings described above. Genomic DNA of the transformants were isolated using the Gentra® Puregene® Yeast/Bact. Genomic DNA Purification Kit (Qiagen) according to manufacturer's instructions. For each expression construct, correct chromosomal integration was confirmed using two different primer combinations revealing PCR products of the expected sizes:

- Construct No. 1, Seq. primer No.1/3 (1622 bp), and primer No.4/2 (1548 bp)
- Construct No. 2, Seq. primer No.5/7 (1499 bp), and primer No.8/5 (1719 bp)
- Construct No. 3, Seq. primer No.9/11 (1482 bp), and primer No.12/10 (1612 bp)
- Construct No. 4, Seq. primer No.13/15 (1547 bp), and primer No.16/14 (1542 bp)
- Construct No. 5, Seq. primer No.17/19 (1431 bp), and primer No.20/18 (1499 bp)
- Construct No. 6, Seq. primer No.21/23 (1477 bp), and primer No.24/22 (1373 bp)

Genomic DNA of *M. xanthus* DK1622 was used as negative control. A complementary experiment using the following primer combinations revealed a specific PCR product for *M. xanthus* DK1622, but not for any of the transformants of *M. xanthus* DK1622 harboring one of the generated constructs.

- Construct No.1, Seq. primer No. 1/2 (1515 bp PCR product for *M. xanthus* DK1622 wild type)
- Construct No.2, Seq. primer No. 5/6 (1482 bp PCR product for *M. xanthus* DK1622 wild type)
- Construct No.3, Seq. primer No. 9/10 (1597 bp PCR product for *M. xanthus* DK1622 wild type)
- Construct No.4, Seq. primer No. 13/14 (1343 bp PCR product for *M. xanthus* DK1622 wild type)
- Construct No.5, Seq. primer No. 17/18 (1294 bp PCR product for *M. xanthus* DK1622 wild type)
- Construct No.6, Seq. primer No. 21/22 (1425 bp PCR product for *M. xanthus* DK1622 wild type)



**Figure S8.** Genotypic verification procedure of correct chromosomal integration via PCR. Genetic disruption (A) and induced-gene expression constructs (B) were verified by multiplex PCR using four primers: two integrations site-specific primers (TestFw1 and TestRv2) and two vector specific primers (Test Rv3 and Test Fw4).



## 2.4 Heterologous production of cittilin

### Generation of the expression vector pSET152\_P<sub>ermE</sub>\*

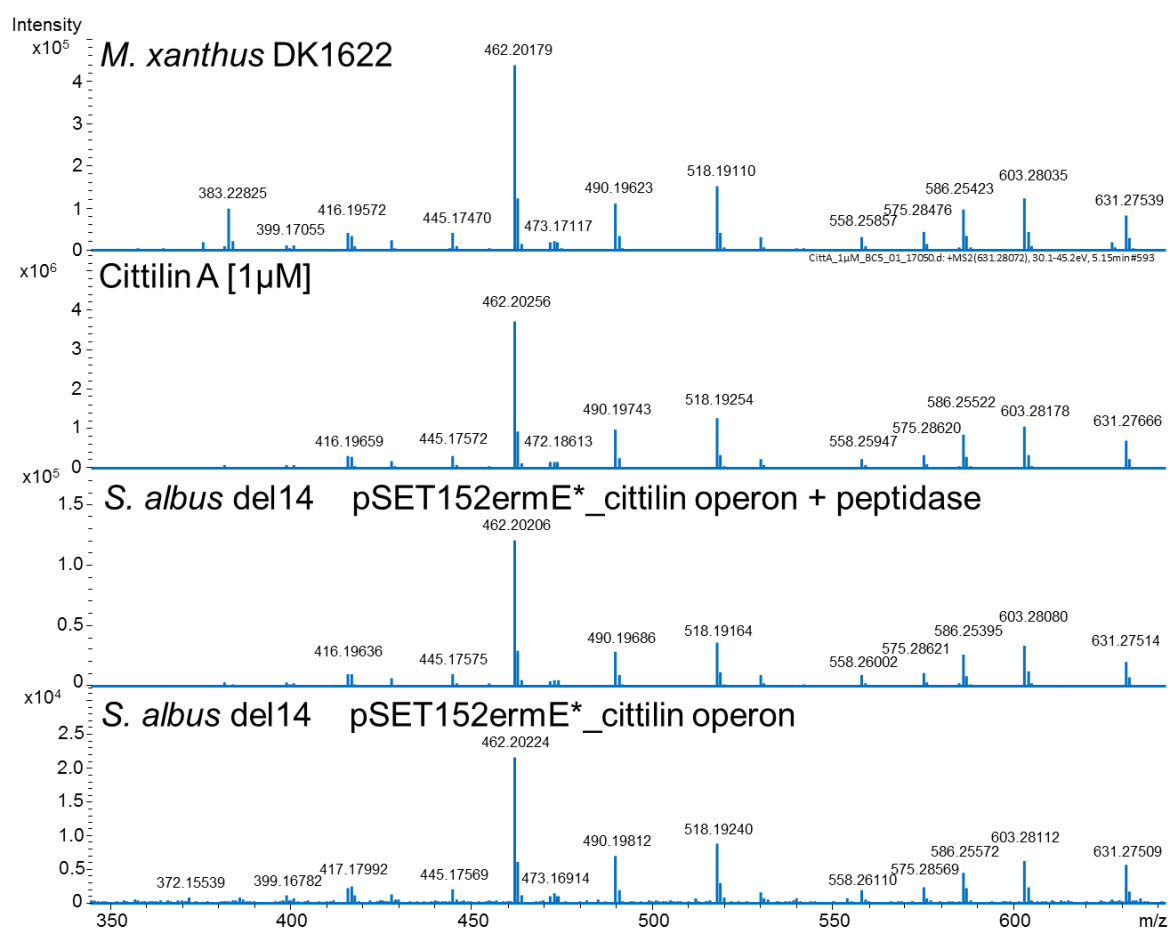
To express the cittilin BGC in *Streptomyces spp.*, an expression vector was constructed based on the plasmid pSET152. The native promoter sequence of pSET152 was exchanged by the ermE\* promoter derived from *Saccharopolyspora erythraea* with nested deletions, finally to achieve higher promoter activity in *Streptomyces lividans* TK24<sup>40</sup>. Therefore the backbone of the generated expression construct derives from the vector pSET152, whereas the ermE\* promoter sequence was amplified via PCR from the plasmid pAB03 (Table S5, PCR construct 7), with additional restriction enzyme specific sequences for PvuI and XbaI. Afterwards the PCR-amplified ermE\* promoter sequence was subcloned into the PacI (the compatible cohesive end of PvuI generated at this position the new restriction site for MseI) and XbaI site of pSET152 to generate the expression vector pSET152\_P<sub>ermE</sub>\* (Table S6, genetic construct 7).

### PCR-based cloning of cittilin operon

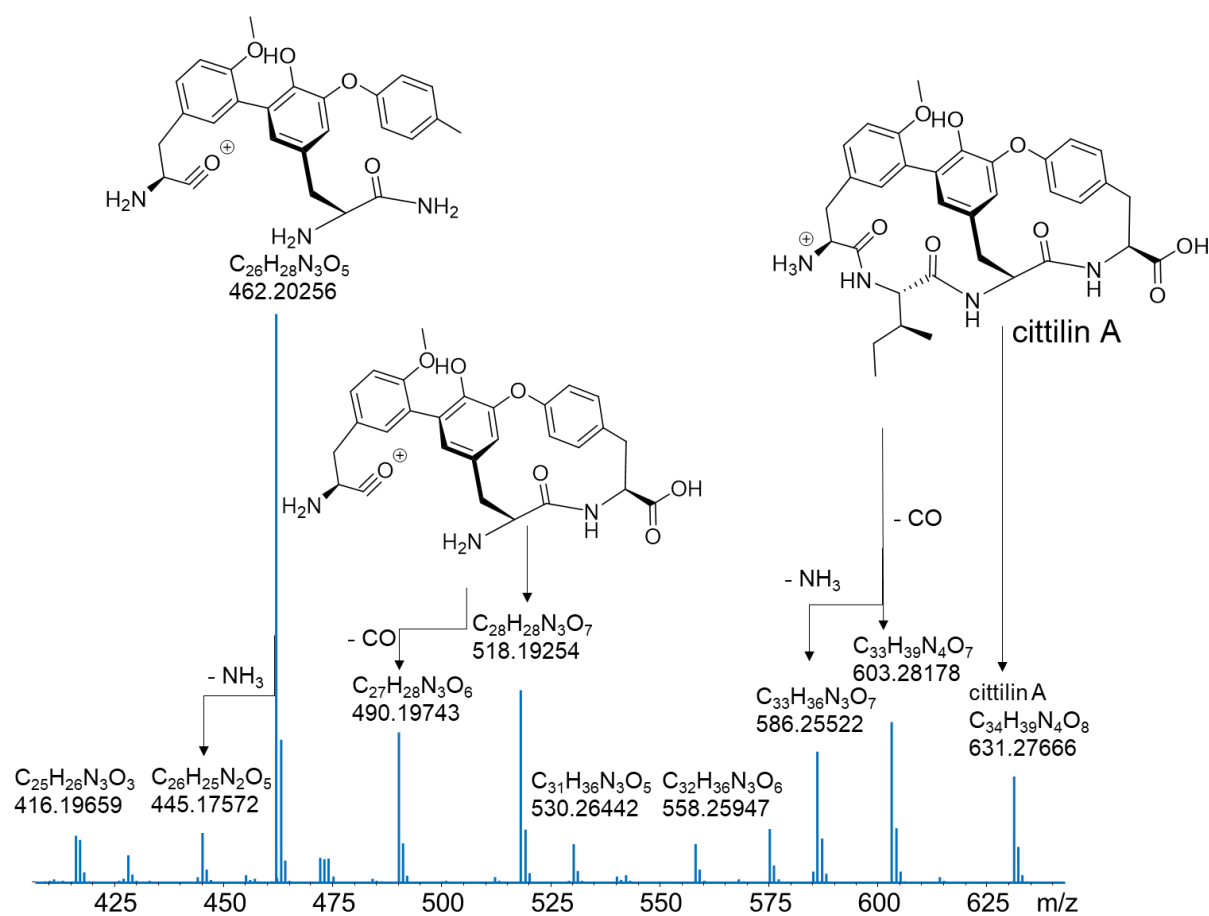
The cittilin operon consisting of the genes *citA–citC* has been PCR-amplified (Table S5, PCR construct 8) and subcloned into the constructed expression vector pSET152\_P<sub>ermE</sub>\* via the restriction site NdeI and BspTI yielding the expression vector pSET152\_P<sub>ermE</sub>\*\_cit<sub>itlin</sub>\_operon (Table S6, genetic construct 8). For the construct including additionally *mx pep*, the cloning strategy was slightly changed. The cittilin operon *citA–citC* was PCR-amplified with a different reverse (Rv) primer harboring the restriction site Bsp1407I (Table S5, PCR construct 9), while *mx pep* was PCR-amplified with the restriction sites Bsp1407I and BspTI (Table S5, PCR construct 10). Subsequently both PCR amplified inserts were subcloned simultaneously into the constructed expression vector pSET152\_P<sub>ermE</sub>\*, via the restriction site NdeI and BspTI yielding the expression vector pSET152\_P<sub>ermE</sub>\*\_cit<sub>itlin</sub>\_operon\_*mx pep* (Table S6, genetic construct 9).

### Intergeneric conjugation of generated genetic constructs

The generated genetic constructs were transformed for intergeneric conjugation with *Streptomyces albus* del14 into the methylation deficient *E. coli* ET12567 strain harboring the RK2 derivative pUZ8002, since many streptomycetes possess a potent methylation-specific restriction system, which effectively prevents the introduction of heterologous DNA<sup>41</sup> (see above). The plasmid pUZ8002 supplies transfer functions to the oriT carrying plasmid backbone of pSET152 without transferring itself due to a mutation in its own oriT<sup>42</sup>. After conjugation, the generated mutants *Streptomyces albus* del14\_pSET152\_P<sub>ermE</sub>\*, *Streptomyces albus* del14\_pSET152\_P<sub>ermE</sub>\*\_cit<sub>itlin</sub>\_operon and *Streptomyces albus* del14\_pSET152\_P<sub>ermE</sub>\*\_cit<sub>itlin</sub>\_operon\_*mx pep* were used from a sporulating agar plate to inoculated 20 mL of TSB medium to obtain seed cultures. After three days, the seed cultures were well grown and used to inoculate with 1–2 mL COM medium containing XAD-16. Four days later, the cultures were centrifuged, CPs and XAD-16 extracted (50% methanol/ 50% acetone (v:v)) and measured via HPLC-MS. As reference, a crude extract from *M. xanthus* DK1622 and a pure reference of cittilin A was used. Cittilin A production was observed in the mutants *Streptomyces albus* del14\_pSET152\_P<sub>ermE</sub>\*\_cit<sub>itlin</sub>\_operon and *Streptomyces albus* del14\_pSET152\_P<sub>ermE</sub>\*\_cit<sub>itlin</sub>\_operon\_*mx pep*. Furthermore, MS<sup>2</sup>-spectra and fragmentation pattern confirmed the identity of heterologously produced cittilin A (Figure S9, Figure S10).



**Figure S9.** MS<sup>2</sup>-spectra of citilin A. Citilin A produced by *M. xanthus* DK1622, citilin A pure compound as reference and the heterologous produced citilin A in *Streptomyces albus* del14 pSET152ermE\*\_citilin operon with and without coproduction of MX PEP. All four MS<sup>2</sup>-spectra exhibit the same fragmentation pattern, which confirms the successful heterologous production of citilin A. Due to the selective MS/MS fragmentation it was not possible to show a time interval of the MS/MS fragmentation pattern.



**Figure S10.** MS<sup>2</sup>-spectrum and fragmentation pattern of cirtilin A.

## 2.5 *In vitro* and cell free lysate investigation of cittilin biosynthesis

### *In vitro* investigation of CitC

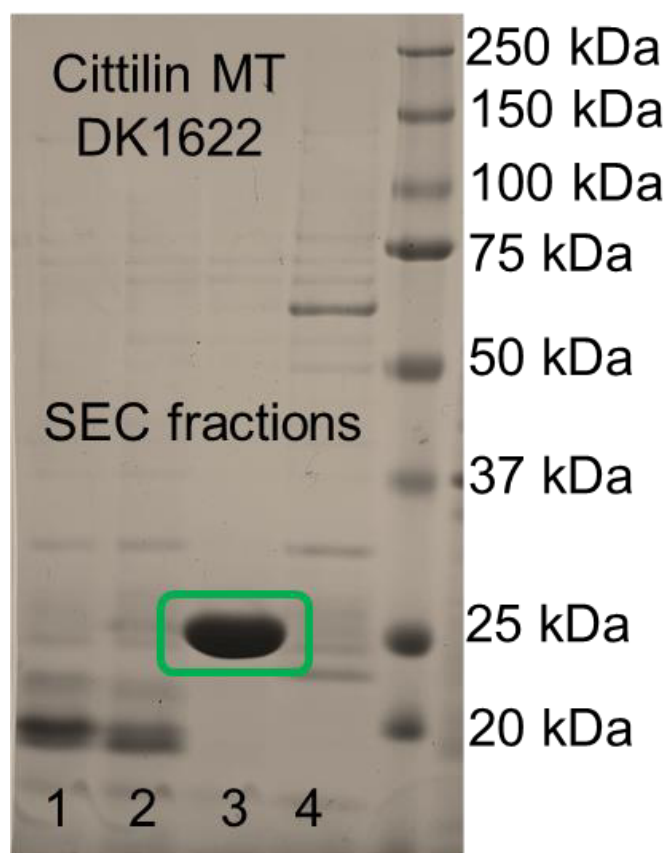
#### Cloning of *citC* into overexpression vector

The gene encoding the cittilin-specific methyltransferase CitC in *M. xanthus* DK1622 has been PCR-amplified by the primers as shown in **Tab. S2**. The amplified DNA fragment encoding CitC was subcloned into the expression vector pET28b via the restriction site NdeI and EcoRI yielding the expression vector pET28b\_DK1622\_CitC (**Table S6**, genetic construct 10)

#### *E. coli*-based recombinant CitC production

The generated expression vector pET28b\_DK1622\_CitC was transformed into *E. coli* Lemo21 for the recombinant production of CitC containing the *N*-terminal His<sub>6</sub>-tag and Tobacco etch virus (TEV) cleavage site and an IPTG-inducible *lac*-operon. The generated *E. coli* strain was grown o/n in LB medium containing 50 µg/mL kanamycin at 37 °C. Antibiotic supplemented LB medium was inoculated with the o/n culture and incubated at 37 °C until an OD<sub>600</sub> of 0.6 was reached. The culture was equilibrated at 16 °C before expression of His<sub>6</sub>-tagged *citC* was induced with 1 mM IPTG. The culture was incubated at 16 °C for 20 h. Subsequently the cells were harvested at 3400 g for 10 min at 4 °C. The CP of *E. coli* Lemo21\_pET28b\_DK1622\_CitC was re-suspended in lysis buffer and cells were lysed using a CD-017a constant cell disruption system. Cell debris was removed by centrifugation (15 min at 50000 g) and the SN was loaded to a 5 mL HisTrap FF column (GE Healthcare) with 5 mL/min on an ÄKTA™ pure system (GE Healthcare) after the column was equilibrated with five column volumes (CVs) lysis buffer. The column loaded with recombinant protein was washed with 10 CV lysis buffer. Elution was performed isocratically with 5 CV elution buffer at a flow rate of 5 mL/min. Protein-containing fractions, protein identity and purity were assessed by SDS-PAGE analysis (12% acrylamide) and LC-MS. Combined fractions containing CitC were concentrated via centrifugal filtration using Amicon Ultra-30 columns (MW 30000 Da, Merck). Size exclusion chromatography was performed using a Superdex 200 Increase preppacked column. After equilibration with 1.2 CV protein buffer, the recombinant protein solution was loaded via a 2 mL loading loop. The size excluded protein fractions were identified via SDS-PAGE (Figure S11). The combined protein fraction was digested with TEV protease (1.5 mg/10 mg recombinant protein) (o/n) at 4 °C. The next day, a second Ni-affinity purification was performed using a 5 mL HisTrap FF column (GE Healthcare) with 5 mL/min on an ÄKTA™ pure system (GE Healthcare). The column was equilibrated with 5 CV lysis buffer and the TEV digested protein solution was loaded on the HisTrap FF column with an flow rate of 5 mL/min and 10 CV column wash followed. The column with bound His<sub>6</sub>-TEV protease and His<sub>6</sub>-TEV site was isocratically eluted with 5 CV elution buffer. The column wash fractions with CitC without His<sub>6</sub>-tag were combined and protein identity was assessed by SDS-PAGE. The protein solution was concentrated via centrifugal filtration using Amicon Ultra-30 columns (MW 30000 Da, Merck). Size exclusion chromatography followed using a Superdex 200 Increase preppacked columns. After equilibration with 1.2 CV protein buffer with a flow rate of 1 mL/min the recombinant protein solution was loaded via a 2 mL loading loop. Protein-containing fractions, protein identity and purity were assessed by SDS-PAGE and LC-MS. Combined CitC fractions were concentrated via centrifugal filtration using Amicon Ultra-30 columns (MW 30000 Da, Merck). The concentrated CitC solution was adjusted at 10% glycerol and 100 µM and aliquoted into PCR tubes. Protein concentrations were determined by UV spectroscopy

(with  $\epsilon_{280}$  nm values) using Thermo Scientific™ NanoDrop™ 2000/2000c. The aliquots were frozen immediately in liquid nitrogen and stored at -80 °C.



**Figure S11.** SDS-PAGE gel of CitC purification. Green circle indicate CitC after size exclusion chromatography.

### Catalytic activity testing of recombinant CitC

Catalytic activity of the purified recombinant CitC was tested in a reaction mixture (50  $\mu$ L end volume) containing 2  $\mu$ M CitC, 2 mM *S*-adenosyl-L-methionine (SAM), 5  $\mu$ L of concentrated crude extract (see below) in MQ water. Negative control testing were performed by omitting SAM or CitC (data not shown). Replacing concentrated crude extract with 100  $\mu$ M precursor ((synthesized and commercially purchased peptide with the sequence KKALYSLAVLMRFARADKLSAPYIYY (DK1622 motif)) peptide or core peptide (synthesized and commercially purchased peptide with the sequence YIYY) showed no methylation of these substrates. The reaction was carried out o/n at 30°C. The reaction was terminated by adding MeOH (50  $\mu$ L, final concentration 50% v/v). The mixture was transferred to -80 °C for at least 1 h, centrifuged at 13000 g for 15 min at 4°C (VWR centrifuge ECN521-3601, Hitachi Koki Co., Ltd) and 1  $\mu$ L of the SN was subjected to HPLC-MS analysis as described previously.

### Preparation of cittilin B as substrate from crude extract

One mL of crude extract from the mutant *M. xanthus* DK1622\_ citC\_disruption (obtained previously see above) was dried under N<sub>2</sub>. Residue was re-dissolved with 100  $\mu$ L MeOH. Centrifugation for 15 min at 15000 rpm, 4 °C. The SN was used subsequently as substrate (cittilin B).

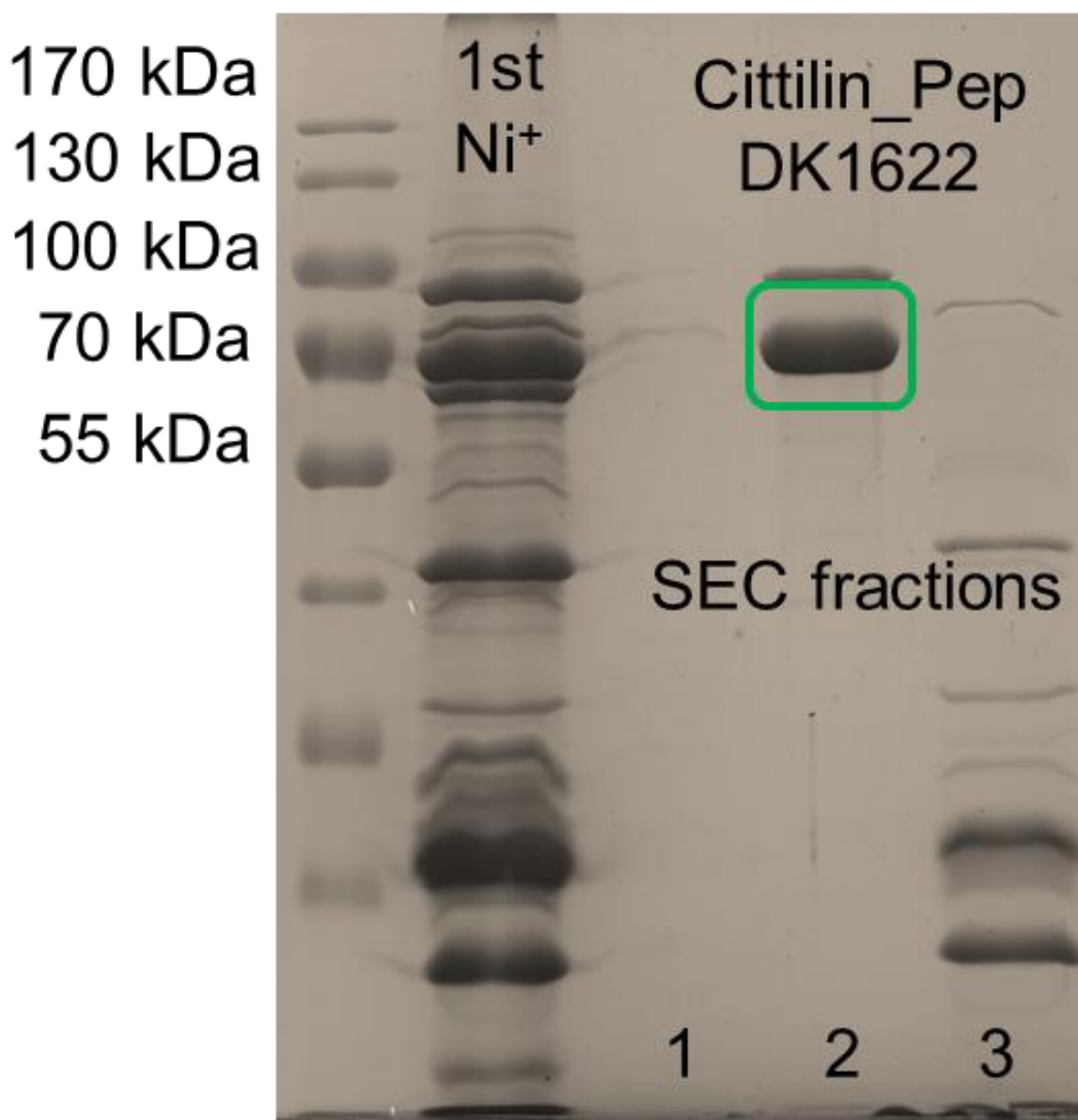
### ***In vitro* investigation of the prolyl endopeptidase MX PEP**

#### **Cloning of the *mx pep* into overexpression vector**

The gene encoding the prolyl endopeptidase MX PEP in *M. xanthus* DK 1622 has been PCR-amplified by the primers as shown in **Tab. S2**. The amplified DNA fragment encoding MX PEP was subcloned into the expression vector pET28b via the restriction site NdeI and HindIII, yielding the expression vector pET28b\_DK1622\_MX\_PEP (**Table S6**, genetic construct 11).

#### ***E. coli*-based recombinant MX PEP production**

The generated expression vector pET28b\_DK1622\_MX\_PEP was transformed into *E. coli* BL21 ( $\lambda$ DE3) for the recombinant production of MX PEP containing the *N*-terminal His<sub>6</sub>-tag and TEV cleavage site and an IPTG-inducible *lac*-operon. The *E. coli* host harboring the expression vector was grown o/n in LB medium containing 50  $\mu$ g/mL kanamycin at 37 °C. Antibiotic supplemented LB medium was inoculated with the o/n culture and incubated at 37 °C until an OD<sub>600</sub> of 0.6 was reached. The culture was equilibrated at 16 °C before the expression of His<sub>6</sub>-tagged *mx pep* was induced with 1 mM IPTG. The culture was incubated at 16 °C for 20 h. Subsequently the cells were harvested at 3400 g for 10 min at 4 °C. The CP of *E. coli* BL21 ( $\lambda$ DE3) with pET28b\_DK1622\_MX\_PEP was re-suspended in lysis buffer and cells were lysed using a CD-017a constant cell disruption system. Cell debris was removed by centrifugation (15 min at 50000 g) and the SN was loaded to a 5 mL HisTrap FF column (GE Healthcare) with 5 mL/min on an ÄKTA™ pure system (GE Healthcare) after the column was equilibrated with 5 CV lysis buffer. The column loaded with recombinant protein was washed with 10 CV lysis buffer. Elution was performed via a linear gradient up to 100%, with 5 CV elution buffer at a flow rate of 5 mL/min. Protein-containing fractions, protein identity and purity were assessed by SDS-PAGE analysis (12% acrylamide) and LC-MS. Combined fractions containing MX PEP were concentrated via centrifugal filtration using Amicon Ultra-30 columns (MW 50000 Da, Merck). Size exclusion chromatography was performed using a Superdex 200 Increase prepacked column. After equilibration with 1.2 CV protein buffer, the recombinant protein solution was loaded via a 2 mL loading loop. The size excluded protein fractions were identified via SDS-PAGE (Figure S12). Combined MX PEP fractions were concentrated via centrifugal filtration using Amicon Ultra-30 columns (MW 30000 Da, Merck). The concentrated protein solution was adjusted to 10% glycerol and 100  $\mu$ M and aliquoted into PCR tubes. Protein concentrations were determined by UV spectroscopy (with  $\epsilon$ <sub>280 nm</sub> values) using Thermo Scientific™ NanoDrop™ 2000/2000c. The aliquots were frozen immediately in liquid nitrogen and stored at -80 °C.

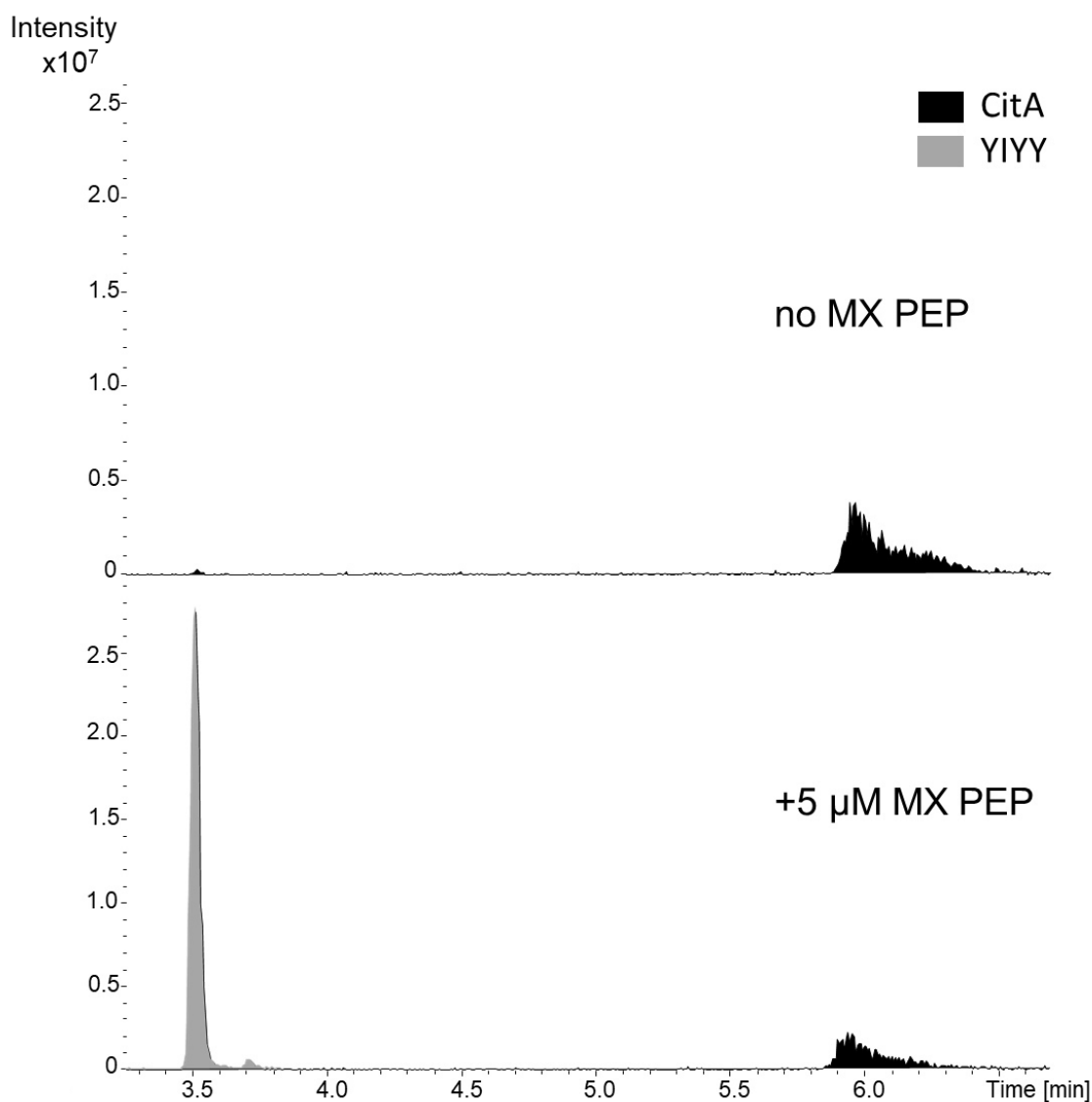


**Figure S12.** SDS-PAGE gel of MX PEP purification. Green box indicates recombinant MX PEP after size exclusion chromatography (SEC) fractionation.

### Catalytic activity testing recombinant MX PEP

Since the modified precursor peptide was not available as a substrate for MX PEP, we decided to first investigate conversion of the unmodified precursor peptide. Catalytic activity of the purified recombinant MX PEP was tested in a reaction mixture containing 5  $\mu$ M MX PEP and 100  $\mu$ M precursor peptide (synthesized and commercially purchased peptide with the sequence KKALYSLAVLMRFARADKLSAPYIYY (DK1622 motif) in 50 mM NaCl, 20 mM Bis-TRIS buffer (pH 6.8). Negative control testing were performed by omitting recombinant MX PEP. The reaction was carried out (o/n) at room temperature. The reaction was terminated by adding MeOH (final concentration 50% v/v). The mixture was transferred to -80 °C for at least 1 h, centrifuged at 13000 g for 15 min at 4 °C (VWR centrifuge ECN521-3601, Hitachi Koki Co., Ltd) and 1  $\mu$ L of the SN was subjected to HPLC-MS analysis as described previously.

The recombinantly produced prolyl endopeptidase originating from *M. xanthus* DK1622 did catalyze the cleavage of the leader peptide yielding the linear core peptide, which was confirmed by mass spectrometric detection of a 621.28 Da fragment (**Figure S13**). In conclusion, the identified prolyl endopeptidase MX PEP catalyzes the cleavage of different peptides as reported previously<sup>43,44</sup>, and is not exclusively optimized for the cleavage of the modified or native cittilin precursor peptide.



**Figure S13.** HPLC-MS EIC chromatograms (3051.75 m/z (grey) and 621.28 m/z (black), precursor peptide  $[M+H]^+$  and core peptide  $[M+H]^+$ ) shows the catalytic cleavage of the precursor peptide into the core peptide via recombinantly produced prolyl endopeptidase.



## ***In vitro* investigation of the cittilin cytochrome P450 enzyme CitB**

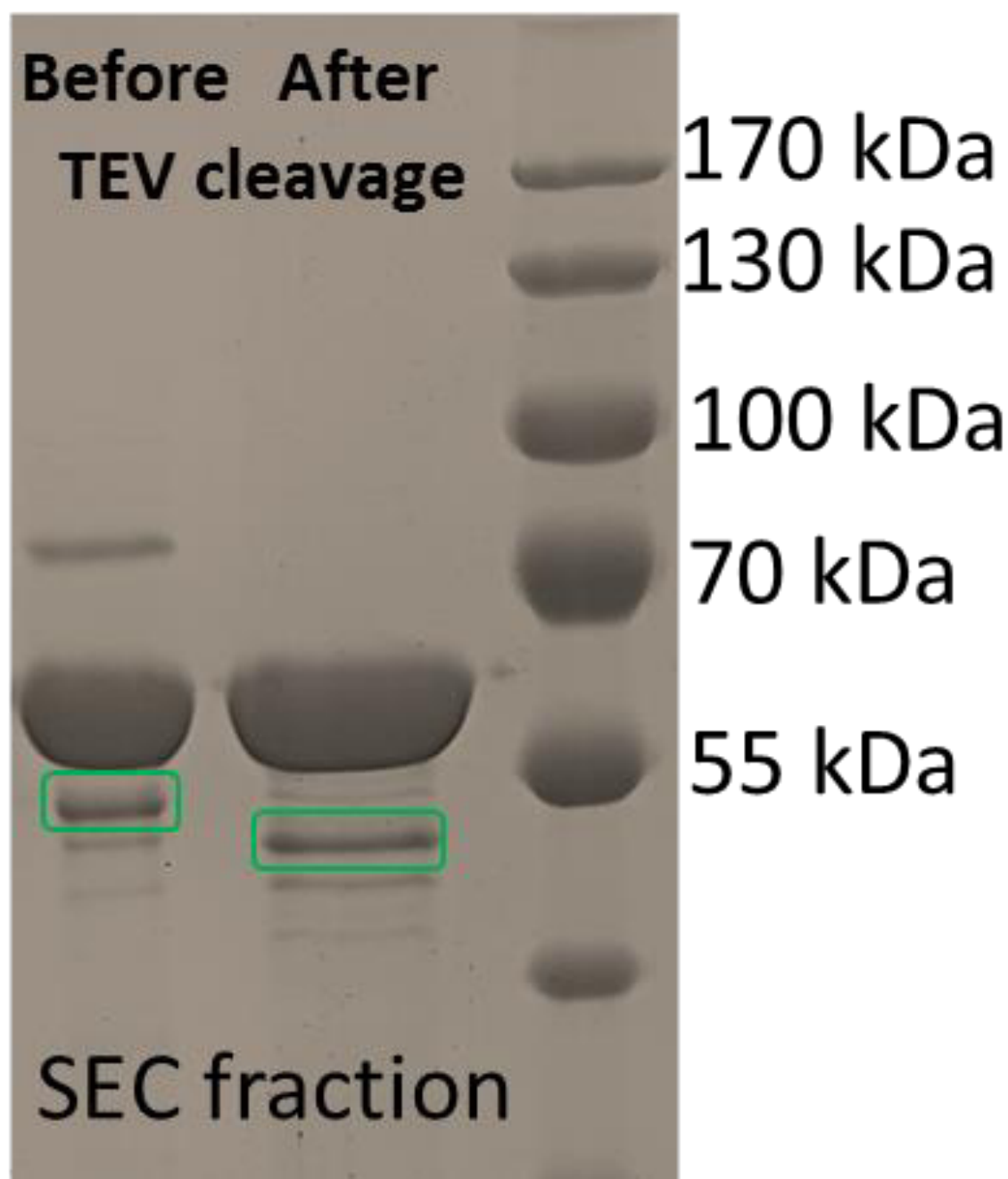
### **Cloning of *citB* into overexpression vector for *E. coli*-based expression**

The gene encoding the cittilin cytochrome P450 enzyme CitB in *M. xanthus* DK 1622 has been PCR-amplified by the primers as shown in Tab. S2. The amplified DNA fragment encoding CitB<sub>DK1622</sub> was subcloned into the expression vector pHisTEV via the restriction site NcoI and HindIII, yielding the expression vector pHisTEV\_DK1622\_CitB (Table S6, genetic construct 12).

### ***E. coli*-based recombinant CitB<sub>DK1622</sub> production**

Recombinant production of CitB was coupled to *in vivo* co-production of chaperons, using recombinant vectors from a commercially available chaperone plasmid set (Takara Bio Inc.). The generated recombinant expression vector pHisTEV\_DK1622\_CitB and pGro7 (Takara Chaperone plasmid set) for co-production of GroEL/GroES were co-transformed into *E. coli* C43. *E. coli* C43 with pHisTEV\_DK1622\_CitB and pGro7 was grown o/n in LB medium containing 50 µg/mL kanamycin and 25 µg/mL chloramphenicol at 37 °C. Antibiotic supplemented TB medium was inoculated with the (o/n) culture and incubated at 37 °C until an OD<sub>600</sub> of 0.6 was reached. The culture was equilibrated at 16 °C before the expression of His<sub>6</sub>-tagged *citB*<sub>DK1622</sub> was induced with 0.5 mM IPTG and co-expression of *groEL/groES* with 3 mg/mL L-arabinose. The culture was incubated at 16 °C for 18 h. Subsequently, cells were harvested at 3400 g for 10 min at 4 °C. The CP of *E. coli* C43 with pHisTEV\_DK1622\_CitB and pGro7 was re-suspended in lysis buffer and cells were lysed using a CD-017a constant cell disruption system. Cell debris was removed by centrifugation (15 min at 50000 g) and the SN was loaded to a 5 mL HisTrap HP column (GE Healthcare) with 5 mL/min on an ÄKTA™ pure system (GE Healthcare) after the column was equilibrated with 5 CV lysis buffer. Subsequently the column with loaded recombinant protein was washed with 10 CV lysis buffer. Elution was performed isocratically with 5 CV elution buffer and a flow rate with 5 mL/min. Protein-containing fractions, protein identity and purity were assessed by SDS-PAGE. Combined CitB<sub>DK1622</sub> were concentrated via centrifugal filtration using Amicon Ultra-30 columns (MW 30000 Da, Merck). Size exclusion chromatography was performed using a Superdex 200 Increase prepacked columns. After equilibration with 1.2 CV protein buffer, the recombinant protein solution was loaded via a 5 mL loading loop. The size excluded protein fractions were identified via SDS-PAGE. The combined protein fractions were digested with TEV protease (1.5 mg/ 10 mg recombinant protein) o/n at 4 °C. The next day, a second Ni-affinity purification was performed using a 5 mL HisTrap FF column (GE Healthcare) with 5 mL/min on an ÄKTA™ pure system (GE Healthcare). The column was equilibrated with 5 CV lysis and the TEV digested protein solution was loaded on the HisTrap FF column with a flow rate of 5 mL/min and 10 CV column wash followed. The column with bound His<sub>6</sub>-TEV protease and His<sub>6</sub>-TEV-site was isocratically eluted with 5 CV elution buffer. The column wash fractions with CitB<sub>DK1622</sub> without His<sub>6</sub>-tag were combined and protein identity was assessed by SDS-PAGE. On the gel, the chaperone GroEL is present, referred with the binding of the chaperone to the substrate protein. LC-MS measurements failed due to overlapping peaks in the chromatogram, further it was not possible to deconvolute an exact mass off the protein. The protein solution was concentrated via centrifugal filtration using Amicon Ultra- 30 columns (MW 30,000 Da, Merck). Size exclusion chromatography followed using a Superdex 200 Increase prepacked columns. After equilibration with 1.2 CV protein buffer, the recombinant protein solution was loaded via a 2 mL loading loop. Protein-containing fractions, protein identity and purity were assessed by SDS-PAGE (Figure S14). Combined CitB<sub>DK1622</sub> fractions were concentrated via centrifugal filtration using Amicon Ultra-30 columns (MW 30000 Da, Merck).

The concentrated protein solution was adjusted at 10% glycerol and 100  $\mu$ M and aliquoted into PCR tubes. The aliquots were frozen immediately in liquid nitrogen and stored at -80 °C. Protein concentrations were determined by UV spectroscopy (with  $\epsilon_{280}$  nm values) using Thermo Scientific™ NanoDrop™ 2000/2000c. The amount of protein was high, related to the high amount of recombinant chaperon bound to the substrate protein.



**Figure S14.** SDS-PAGE gel of recombinant CitBDK<sub>1622</sub> purification. Green rectangles indicate pure CitBDK<sub>1622</sub> (left; with His<sub>6</sub>-TEV-site, right; without His<sub>6</sub>-TEV-site), whereas the upper bands indicate co-produced GroEL.

### Catalytic activity testing of recombinant CitB<sub>DK1622</sub> from *E. coli* based production

Catalytic activity of the recombinant CitB<sub>DK1622</sub> was tested in a reaction mixture containing approx. 5  $\mu$ M CitB<sub>DK1622</sub> and 100  $\mu$ M precursor peptide (synthesized and commercially purchased peptide with the sequence KKALYSLAVLMRFARADKLSAPYIYY (DK1622 motif)) or core peptide (synthesized and commercially purchased peptide with the sequence YIYY). In addition, 500  $\mu$ M MgCl<sub>2</sub>, 1 mM ATP and 2 mM of co-factor (one of the following; FMN, FAD<sup>+</sup>, NAD<sup>+</sup>, NADP<sup>+</sup>, NADPH, NADH) in 500 mM NaCl, 20 mM Bis-TRIS buffer (pH 6.8) was added. As alternative approach to achieve co-factor regeneration, the commercially available Fdx/FdR reductase pair system from *Spinacia oleracea* was performed with 2.5  $\mu$ M Fdx, 2.5  $\mu$ M FdR and 2 mM NADPH. Control testing were performed by omitting CitB<sub>DK1622</sub>. The reaction was carried out for 10 min, 1 h and (o/n) at room temperature and 30 °C. The reaction was terminated by adding MeOH (final concentration 50% v/v). The mixture was transferred to -80 °C for at least 1 h, centrifuged at 13000 g for 15 min at 4 °C (VWR centrifuge ECN521-3601, Hitachi Koki Co., Ltd) and 1  $\mu$ L of the SN was subjected to HPLC-MS analysis as described previously. No activity of recombinant CitB<sub>DK1622</sub> was observed.

### Cloning of the *citB* into overexpression vector for *Streptomyces*-based expression

Since the CitB<sub>DK1622</sub> seemed to be recalcitrant for efficient recombinant production in *E. coli*, an alternative strategy was conducted utilizing *Streptomyces* as heterologous host for recombinant production of CitB. The successful heterologous expression of the citilin BGC in *S. albus* del14 and previous successful recombinant cytochrome P450 enzyme production in *Streptomyces*<sup>45</sup> underline the potential of this approach.

In order to increase the prospects for successful recombinant protein production, not only *citB* from *M. xanthus* DK1622 was cloned and tested for heterologous expression but also homologs from *Cystobacter* spp. (MCy9171) and Cb vi35 (MCy8337) were cloned. The expression vector pCJW93, which has an *N*-terminal His<sub>6</sub>-tag and thrombin cleavage site (LVPRGS) was used for recombinant CitB production. Since the available thrombin cleavage site has several disadvantages for later performed purification steps<sup>46</sup>, the TEV cleavage site was additionally amplified for *N*-terminal His<sub>6</sub>-tag constructs to facilitate the cleavage of the *N*-terminal His<sub>6</sub>-tag during protein purification. In order to test the influence of either an *N*-terminal or *C*-terminal His<sub>6</sub>-tag for the functional production of cytochrome P450 enzyme, the His<sub>6</sub>-tag was modified by cloning procedures to obtain the shuttle vector pCJW93\_noHis without any His<sub>6</sub>-tag and thrombin cleavage site, in which *citB* with amplified *C*-terminal His<sub>6</sub>-tag can be cloned.

This modification was conducted through restriction digestion with Alw44I and NdeI of the shuttle vector pCJW93, to yield two different DNA fragments. The smaller DNA fragment was exchanged through a PCR amplified fragment (pCJW93\_noHistag\_exchange\_construct, **Table S5**, No. 17), which was digested with the restriction enzymes Alw44I and NdeI. The native larger DNA fragment pCJW93\_backbone and the smaller PCR amplified pCJW93no His<sub>6</sub>-tag fragment were ligated via Alw44I and NdeI to obtain the pCJW93noHis plasmid after transformation of the ligation product into *E. coli* HS996. The shuttle vectors pCJW93 and pCJW93noHis were used for further cloning of different *citB* homologs.

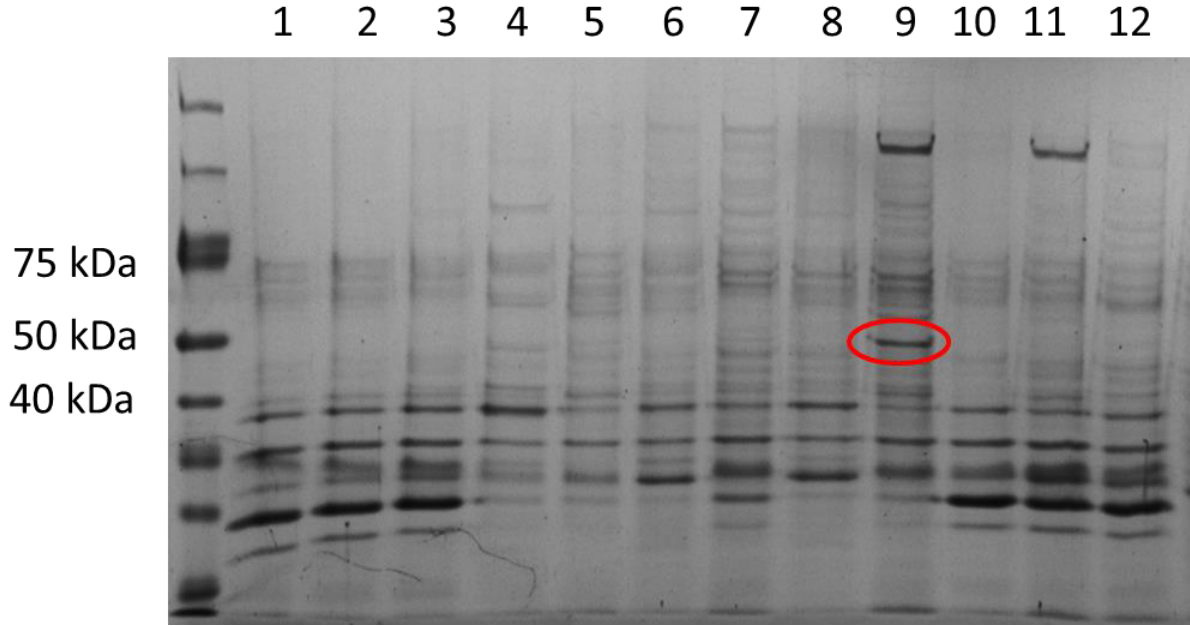
In general, all CitB constructs designed for production of recombinant protein with *N*-terminal His<sub>6</sub>-tag, the PCR products contained the additional TEV cleavage site for further purification steps, whereas constructs designed for production of recombinant protein with *C*-terminal His<sub>6</sub>-tag, the PCR products had to be amplified with a *C*-terminal His<sub>6</sub>-tag.

- *citB*<sub>MCy8337</sub> from strain Cb vi35 (MCy8337) with *C*-terminal His<sub>6</sub>-tag and *citB*<sub>MCy9171</sub> from MCy9171 with *N*-terminal His<sub>6</sub>-tag were digested with EcoRI and HindIII and cloned into the EcoRI and HindIII site of pCJW93 and respectively of pCJW93noHis.
- *citB*<sub>DK1622</sub> from DK1622 gene with *N*-terminal His<sub>6</sub>-tag, *citB*<sub>DK1622</sub> from DK1622 gene with *C*-terminal His<sub>6</sub>-tag, *citB*<sub>MCy9171</sub> from MCy9171 with *C*-terminal His<sub>6</sub>-tag and *citB*<sub>vi35</sub> from Cb vi35 with *N*-terminal His<sub>6</sub>-tag was cloned into the *NdeI* and *EcoRI* site of pCJW93 and respectively of pCJW93noHis.

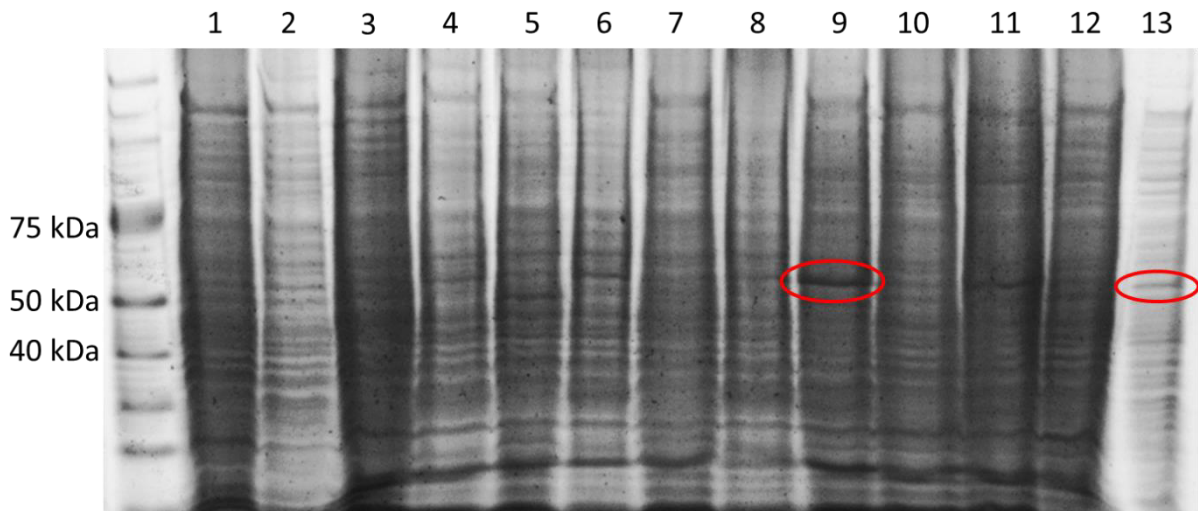
The gene encoding CitB in *M. xanthus* DK 1622 and MCy9171 has been PCR-amplified by the primers as shown in Tab. S2. The amplified DNA fragment encoding the CitB was subcloned into the expression vector pCJW93 via the restriction site *NdeI* and *HindIII*, yielding the expression vector pCJW93\_DK1622\_CitB (Table S6, genetic construct 13).

### ***Streptomyces*-based recombinant production of CitB; test productions**

The generated plasmids were conjugated into *Streptomyces coelicolor* CH999 with *E. coli* ET12567 harboring the plasmid pUZ8002 as donor strain as described previously (see above). After conjugation, the generated *S. coelicolor* CH999 mutants were used to inoculate TSB media to obtain seed cultures. After three days, the seed cultures were used to inoculate duplicates of 100 mL of Super YEME media with 5 mL of the seed culture and were incubated at 30 °C and 180 rpm for at least two days. The cultures were prepared in 250 mL Erlenmeyer flasks with metal spirals to ensure a suspension cell culture. After two days, the cultures were well grown and gene expression was induced with thiostrepton (working concentration 10 µg/mL). The duplicates of cultures were separated into two parts; one was incubated at 25 °C, 180 rpm, the other at 30 °C, 180 rpm. 24 h later, the incubation was stopped and 10 mL of each culture was transferred into a Falcon tube and centrifuged for at 4000 rpm for 10 min at 4 °C. After discarding the SN, the CP was re-suspended in 500 µL lysis buffer No. 3. The cell suspension was sonicated, centrifuged and nickel pulldown was performed (see below). SDS-PAGE revealed that the *citB*<sub>MCy9171</sub> from *Cystobacter* MCy9171 could be recombinantly overexpressed according to the described parameters. **Figure S15** and **Figure S16** display a specific band with the size of 50 kDa, which can be connected to soluble CitB<sub>MCy9171</sub> (Figure S15) and CitB<sub>MCy9171</sub> in the inclusion bodies (**Figure S16**).

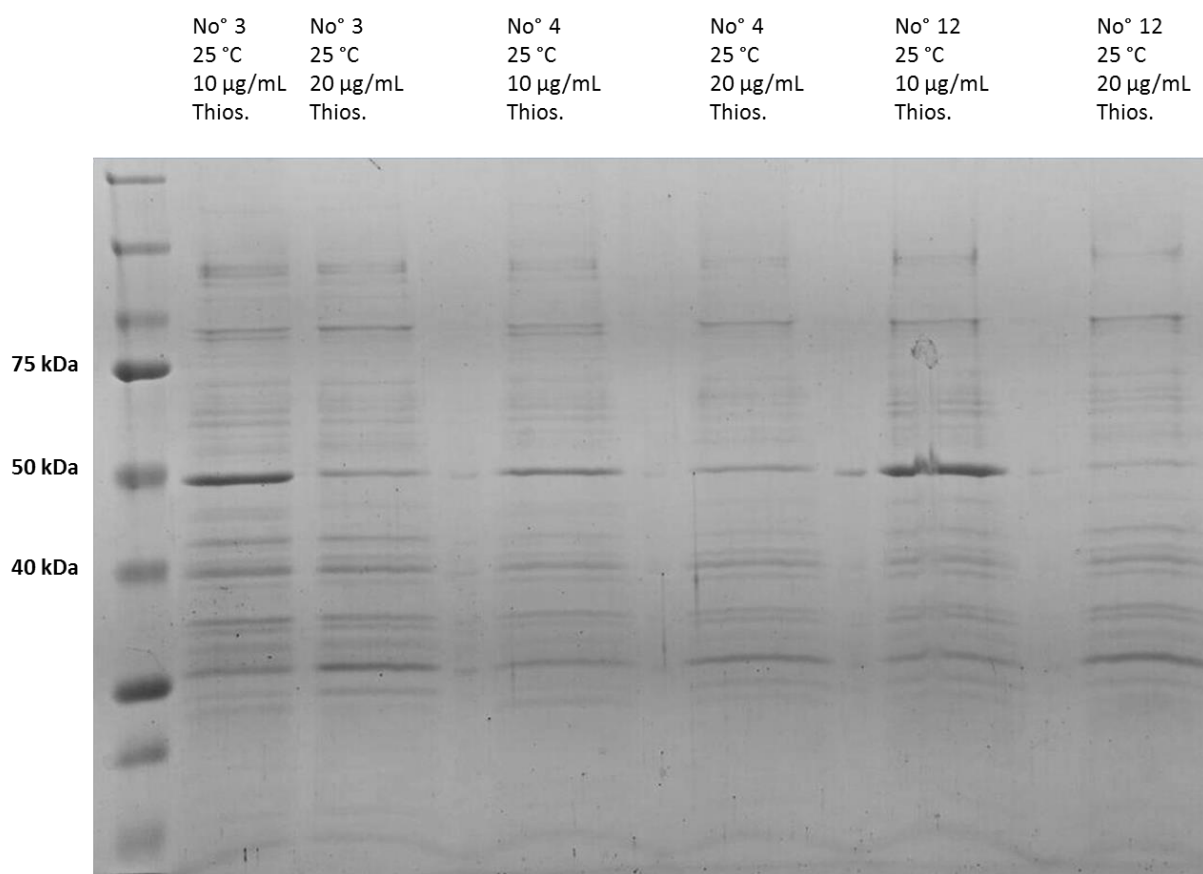


**Figure S15.** SDS-PAGE of *citB* expression tests in *S. coelicolor* CH999 after nickel pulldown. The temperature of incubation after induction was 25 °C and 30 °C respectively. 1) *S. coelicolor* CH999\_pCJW93\_DK1622\_CitB at 25 °C; 2) *S. coelicolor* CH999\_pCJW93noHis\_DK1622\_CitB at 25 °C; 3) *S. coelicolor* CH999\_pCJW93\_DK1622\_CitB at 30 °C; 4) *S. coelicolor* CH999\_pCJW93noHis\_DK1622\_CitB at 30 °C; 5) *S. coelicolor* CH999\_pCJW93\_Cb\_vi35\_CitB at 25 °C; 6) *S. coelicolor* CH999\_pCJW93noHis\_Cb\_vi35\_CitB at 25 °C; 7) *S. coelicolor* CH999\_pCJW93\_Cb\_vi35\_CitB at 30 °C; 8) *S. coelicolor* CH999\_pCJW93noHis\_Cb\_vi35\_CitB at 30 °C; 9) *S. coelicolor* CH999\_pCJW93\_MCy9171\_CitB at 25 °C; 10) *S. coelicolor* CH999\_pCJW93noHis\_MCy9171\_CitB at 25 °C; 11) *S. coelicolor* CH999\_pCJW93\_MCy9171\_CitB at 30 °C; 12) *S. coelicolor* CH999\_pCJW93noHis\_MCy9171\_CitB at 30 °C.



**Figure S16.** SDS-PAGE (cell pellets (CPs) *citB* expression tests in *S. coelicolor*. The temperature of incubation after induction was 25 °C and 30 °C respectively. 20 µL of CP after sonication and centrifugation of 1) *S. coelicolor* CH999\_pCJW93\_DK1622\_CitB at 25 °C; 2) *S. coelicolor* CH999\_pCJW93noHis\_DK1622\_CitB at 25 °C; 3) *S. coelicolor* CH999\_pCJW93\_DK1622\_CitB at 30 °C; 4) *S. coelicolor* CH999\_pCJW93noHis\_DK1622\_CitB at 30 °C; 5) *S. coelicolor* CH999\_pCJW93\_Cb\_vi35\_CitB at 25 °C; 6) *S. coelicolor* CH999\_pCJW93noHis\_Cb\_vi35\_CitB at 25 °C; 7) *S. coelicolor* CH999\_pCJW93\_Cb\_vi35\_CitB at 30 °C; 8) *S. coelicolor* CH999\_pCJW93noHis\_Cb\_vi35\_CitB at 30 °C; 9) *S. coelicolor* CH999\_pCJW93\_MCy9171\_CitB at 25 °C; 10) *S. coelicolor* CH999\_pCJW93noHis\_MCy9171\_CitB at 25 °C; 11) *S. coelicolor* CH999\_pCJW93\_MCy9171\_CitB at 30 °C; 12) *S. coelicolor* CH999\_pCJW93noHis\_MCy9171 at 30 °C 13) Supernatant (SN) of *S. coelicolor* CH999\_pCJW93\_MCy9171\_CitB at 25 °C.

In **Figure S15** and **Figure S16** the band of the expected protein is visible in the CP and in the SN, revealing an overexpression of *citB<sub>MCy9171</sub>*. The amount of protein depends on the composition of lysis buffer, used to extract the protein for further purification steps. For that reason, different lysis buffers were used to solute the protein after mechanic cell lysis. To test the correlation of protein purification and different lysis buffers, 20 mL of TSB medium was inoculated with spores of *S. coelicolor* CH999 harboring the constructed shuttle vector pCJW93\_MCy9171\_CitB from a glycerol (20%) cryogenic long-term stock to cultivate a seed culture. After three days, the culture was well-grown and suitable to inoculate two times 100 mL of Super YEME media containing 50 µg/mL apramycin in a 250 mL shake flask with metal spiral. After four days, the cultures were induced with either 10 or 20 µg/mL thioestrepton. The cultures were incubated for 24 h at 25 °C with 180 rpm and after incubation 10 mL was transferred in a Falcon tube, centrifuged at 4000 rpm, for 10 min at 4 °C, whereas the SN was discarded. The CPs were re-suspended in lysis buffer No. 3, No. 4 (NaCl 500 mM, Tris pH 8.0 20 mM, imidazole pH 8.0 20 mM) and 12 (NaCl 500 mM, Bis-Tris pH 6.8 20 mM, imidazole pH 8.0 20 mM, glycerol 10% m/m). After sonication and centrifugation at 15000 rpm for 10 min at 4 °C, nickel pulldown was performed. The SDS-PAGE revealed that lysis buffer No. 3 and 12 with a thioestrepton concentration of 10 µg/mL to induce gene expression in the fermentation culture is optimal in order to obtain the highest amount of recombinantly expressed protein (**Figure S17**).



**Figure S17.** SDS-PAGE of *citB<sub>MCy9171</sub>* homolog expression tests from *Cystobacter* MCy9171. Each band illustrates the use of different lysis buffers and thioestrepton concentration for induced gene expression. Lysis buffers No. 3 and 12, with a thioestrepton concentration of 10 µg/mL showed the highest yield of recombinant protein.

***Streptomyces*-based recombinant production of CitB<sub>MCy9171</sub> from *Cystobacter* MCy9171**

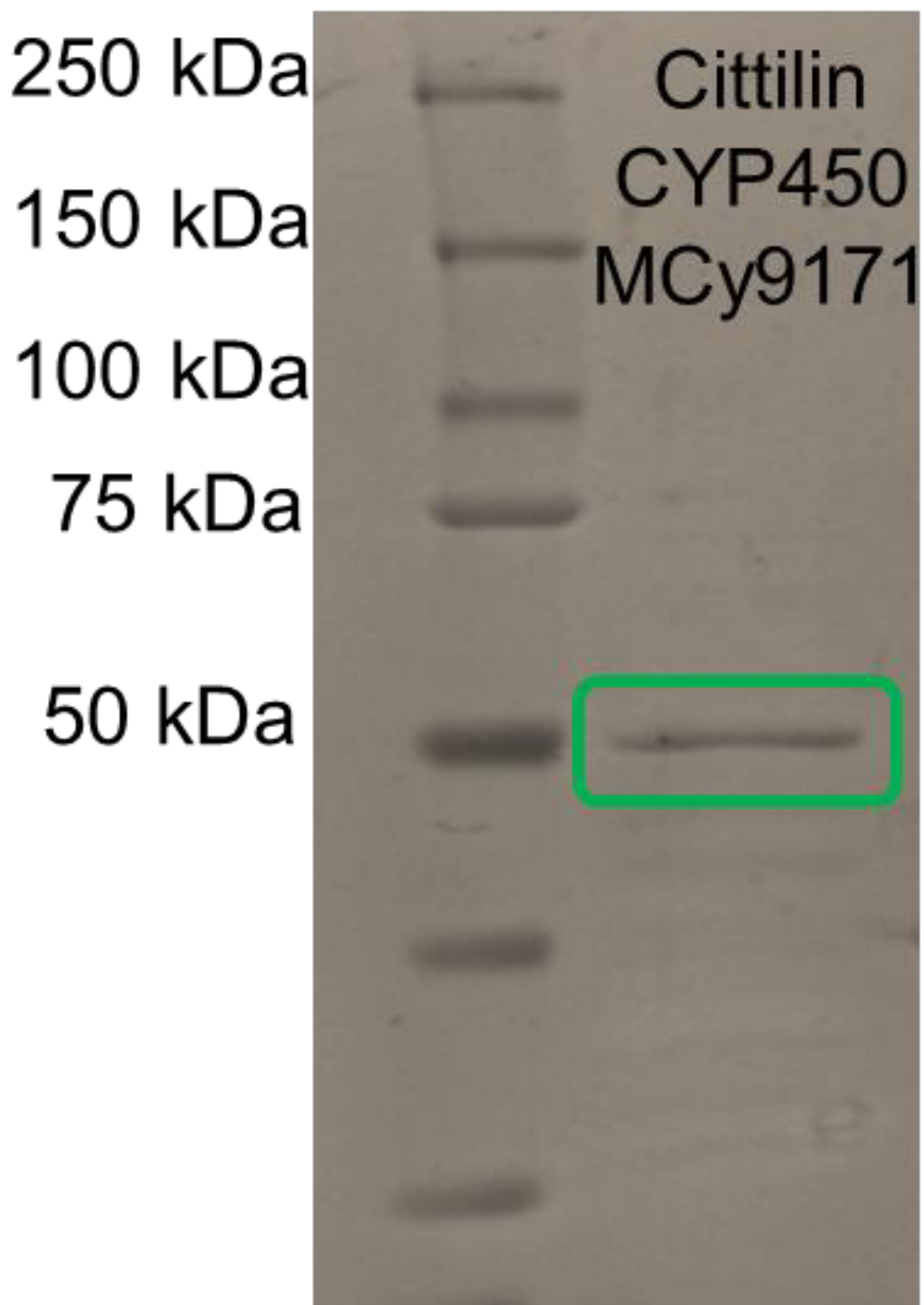
The conclusions of the conducted test-expression experiments of the *citB*<sub>MCy9171</sub> homolog from *Cystobacter* MCy9171 were directly implemented for targeted protein purification in larger scale. 20 mL of TSB medium was inoculated with spores of *S. coelicolor* CH999 harboring the constructed shuttle vector pCJW93\_MCy9171\_CitB from a glycerol (20%) cryogenic long-term stock to cultivate a seed culture.

After three days, the seed culture was used to inoculate with 20 mL, 100 mL TSB medium as pre-culture. After two days, this pre-culture was densely grown and used to inoculate with 5 mL, 18 x 100 mL of Super YEME media incubated at 30 °C and 180 rpm for at least two days. The cultures were prepared in 250 mL Erlenmeyer flasks with metal spirals to ensure cell growth in suspension. After two days, the cultures were well grown and gene expression was induced with thiostrepton (working concentration 10 µg/mL).

The cultures were incubated at 25 °C, 180 rpm after thiostrepton induction. Twenty-four hours later, the incubation was stopped and the cell broth was centrifuged at 4000 rpm for 10 min at 4 °C. After discarding the SN, the CP (27 g) was re-suspended in 100 mL ice-cold lysis buffer No. 3. Two protease inhibitor cocktail tablets (Roche diagnostics) and 10.8 mg of bovine pancreas were added to the cell suspension. Subsequently the cells were lysed using a CD-017a constant cell disruption system. Cell debris was removed by centrifugation (15 min at 50000 g) and the SN was loaded to a 5 mL HisTrap FF column (GE Healthcare) with 5 mL/min on an ÄKTA™ pure system (GE Healthcare) after the column was equilibrated with 5 CV lysis buffer No. 3. The column loaded with recombinant protein was washed with 30 CV lysis buffer.

Elution was performed a linear gradient up to 100%, with 5 CV elution buffer No. 3 at a flow rate of 5 mL/min. Protein-containing fractions, protein identity and purity were assessed by SDS-PAGE analysis (12% acrylamide) and LC-MS. Combined fractions containing CitB<sub>MCy9171</sub> were concentrated via centrifugal filtration using Amicon Ultra-30 columns (MW 30000 Da, Merck). Size exclusion chromatography was performed using a Superdex 200 Increase prepacked column. After equilibration with 1.2 CV protein buffer 3, the recombinant protein solution was loaded via a 2 mL loading loop.

The size excluded protein fractions were identified via SDS-PAGE (Figure S18). Combined CitB<sub>MCy917</sub> fractions were concentrated via centrifugal filtration using Amicon Ultra-30 columns (MW 30000 Da, Merck). The concentrated protein solution was adjusted at 100 µM and aliquoted into PCR tubes. Protein concentrations were determined by UV spectroscopy (with ε<sub>280 nm</sub> values) using Thermo Scientific™ NanoDrop™ 2000/2000c. The aliquots were frozen immediately in liquid nitrogen and stored at -80 °C.



**Figure S18.** SDS-PAGE gel of CitB<sub>MCy9171</sub> purification from *Streptomyces coelicolor* CH999. Green box indicate purified CitB<sub>MCy9171</sub>.



### Catalytic activity testing recombinant CitB<sub>MCy9171</sub> from *Streptomyces* based production

Catalytic activity of the recombinant CitB<sub>MCy9171</sub> was tested in a reaction mixture containing 4.2  $\mu$ M CitB<sub>MCy9171</sub>, 125  $\mu$ M precursor peptide (synthesized and commercially purchased peptide with the sequence KKALYSLAVLMRFARADKLSAPYIYY (DK1622 motif)) or core peptide (synthesized and commercially purchased peptide with the sequence YIYY), 10  $\mu$ M iron(II) sulfate and iron(III) citrate, 500  $\mu$ M of co-factor (one of the following: FAD<sup>+</sup>, NAD<sup>+</sup>, NADP<sup>+</sup>, NADPH, NADH) in 500 mM NaCl, 20 mM Bis-TRIS buffer (pH 6.8). As alternative approach to achieve co-factor regeneration, the commercially available Fdx/FdR reductase pair system from *Spinacia oleracea* was performed with 2.5  $\mu$ M Fdx, 2.5  $\mu$ M FdR and 2 mM NADPH. Controls were performed by omitting recombinant citililn cytochrome P450 enzyme. The reaction was carried out for 10 min, 1 h and (o/n) at room temperature and 30°C. The reaction was terminated by adding MeOH (final concentration 50% v/v). The mixture was transferred to -80 °C for at least 1 h, centrifuged at 13000 g for 15 min at 4 °C (VWR centrifuge ECN521-3601, Hitachi Koki Co., Ltd) and 1  $\mu$ L of the SN was subjected to HPLC-MS analysis as described previously. As described above catalytic activity of recombinant CitB<sub>MCy9171</sub> could not be observed.

### Analysis of recombinant produced CitB by nickel pulldown and SDS-PAGE

10 mL of the liquid cultures of *S. coelicolor* CH999 harboring pCJW93 constructs, which are potential recombinant producers of recombinant CitB were centrifuged for 10 min at 4000 rpm and 4 °C. The CP was re-suspended with 500  $\mu$ L lysis buffer (500 mM NaCl, 20 mM Bis-Tris pH 6.8, 20 mM imidazole pH 8.0, 10% glycerol, 3 mM  $\beta$ -mercaptoethanol). The mixture was transferred to an 1.5 mL Eppendorf tube and sonicated via “sonics Vibra-cell” (ZinsserAnalytic) with an amplitude of 80% for 15 s on-time and 15 s off-time for a total of 2 min of on-time.

After sonication, the tubes were centrifuged for 10 min at 15000 rpm and 4 °C. The SN was transferred to a new 1.5 mL Eppendorf tube and stored on ice. The “KINGFISHER mL” (Thermo Scientific™) was used to perform a nickel pulldown. The principle is based on the use of magnetic nickel beads, which have an affinity to the His<sub>6</sub>-tag of the recombinant protein. Five hundred  $\mu$ L of cell lysate was mixed with 50  $\mu$ L of magnetic nickel beads to bind the His<sub>6</sub>-tagged protein.

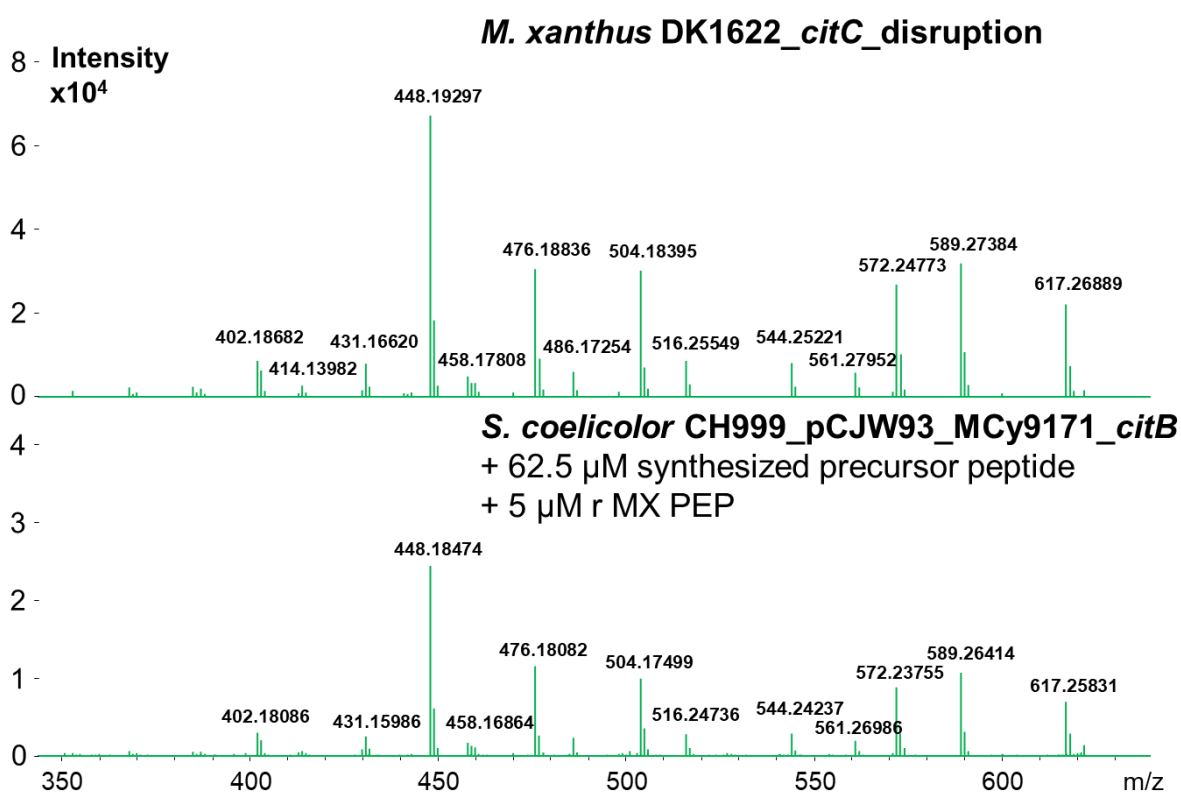
Afterwards, the mixture was washed twice with 500  $\mu$ L of lysis buffer, before the protein was eluted with 50  $\mu$ L of elution buffer (250 mM imidazole pH 8.0). After nickel pulldown, 20  $\mu$ L of the elution fraction was transferred to a 1.5 mL Eppendorf tube and heated up to 95 °C for 2 min with 4  $\mu$ L of SDS loading dye (6x). The SDS-PAGE was loaded with 20  $\mu$ L of the prepared samples; 2  $\mu$ L of “PageRuler Unstained BroadRange Protein Ladder” (Thermo Fisher Scientific™) was loaded for monitoring the progress of SDS-PAGE and for estimating the approximate size of separated proteins after staining of the gel.

The electrophoresis was performed in a “Mini-PROTEAN® Tetra System” (BIO RAD) with SDS (1x) Laemmli buffer with a voltage of 120 V for 90 min. After electrophoresis was finished, the SDS-PAGE was heated in a microwave with staining solution (Coomassie-blue 0.5%, MeOH 50%, acetic acid 7%, H<sub>2</sub>O 43 %) and stored in water for 24 h.

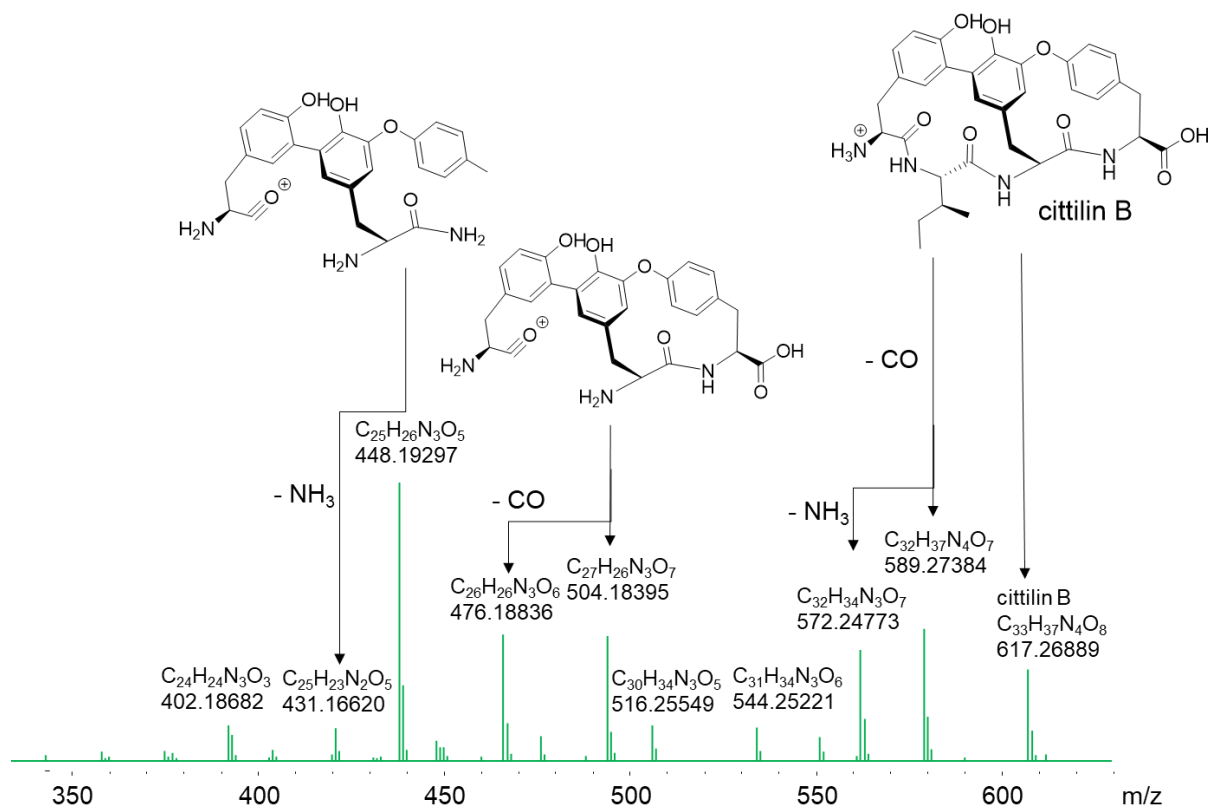
***In vitro* enzymatic conversion of CitB<sub>MCy9171</sub> in cell-free lysate**

10 mL of liquid cultures of thiostrepton-induced cultures of *S. coelicolor* CH999 + pCJW93, *S. coelicolor* CH999 + pCJW93\_MCy9171\_CitB and non-induced *S. coelicolor* CH999 wild type were centrifuged for 10 min at 4000 rpm and 4 °C. The CP was re-suspended with 500 µL lysis buffer (500 mM NaCl, 20 mM Bis-Tris pH 6.8, 10% glycerol (v/v)). The mixture was transferred to an 1.5 mL Eppendorf tube and sonicated via “sonics Vibra- cell” (ZinsserAnalytic) with an amplitude of 80% for 15 s on-time and 15 s off-time for a total of 2 min of on-time. After sonication, the tubes were centrifuged for 10 min at 15000 rpm and 4 °C. The SN was transferred to a new 1.5 mL Eppendorf tube and stored on ice as cell-free lysate.

Catalytic activity of the respective cell-free lysate was tested in a reaction mixture containing 97.5 µL of the freshly prepared cell-free lysate and 62.5 µM precursor peptide (KKALYSLAVLMRFARADKLSAPYIYY (DK1622 motif)) (total reaction volume: 100 µL). Controls were performed by replacing 62.5 µM precursor peptide through 100 µM core peptide (synthesized and commercially purchased peptide with the sequence YIYY). In order to detect the formation of citilin B, similar reaction set-up as mentioned above was prepared, with additional supplementation of 5 µM recombinantly produced prolyl endopeptidase. The reaction was carried out for 5 h at 30°C. The reaction was terminated by adding MeOH (final concentration 50% v/v). The mixture was transferred to -80 °C for at least 1 h, centrifuged at 13000 g for 15 min at 4 °C (VWR centrifuge ECN521-3601, Hitachi Koki Co., Ltd) and 1 µL of the SN was subjected to HPLC-MS analysis as described previously.



**Figure S19.** MS<sup>2</sup>-spectra of citilin B. Citilin B produced by *M. xanthus* DK1622\_ citC\_disruption mutant (**top**) and cell-free lysate of *S. coelicolor* CH999\_pCJW93\_MCy9171\_citB supplemented with chemically synthesized precursor peptide (62.5 µM) and recombinantly produced prolyl endopeptidase MX PEP (5 µM) (**bottom**). Due to the selective MS/MS fragmentation, it was not possible to show a time interval of the MS/MS fragmentation pattern.



**Figure S20.** MS<sup>2</sup>-spectrum and fragmentation pattern of cittilin B.

### CO difference spectra of CitB<sub>MCy9171</sub>

Spectroscopic properties of purified recombinant CitB<sub>MCy9171</sub> were analyzed using a double-beam spectrophotometer (UV-2101PC, Shimadzu, Japan). Recombinant CitB<sub>MCy9171</sub> was diluted to 2.5  $\mu$ M (water or buffer, see above), reduced by using 10  $\mu$ L of a saturated solution of sodium dithionate and CO was generated by the addition of a small spatula of sodium boranocarbonate. All UV-visible absorbance spectra were recorded from 200 to 700 nm<sup>47,48</sup> (data not shown).

### 2.6 Biological function of cittilin A

#### Cell based bioactivity profiling

##### Antimicrobial assay

Standard sterile microbiological techniques were maintained throughout. All microorganisms were handled according to standard procedures and were obtained from the German Collection of Microorganisms and Cell Cultures (Deutsche Sammlung für Mikroorganismen und Zellkulturen, DSMZ) or were part of our internal strain collection. Cittilin A was tested in microbroth dilution assays on the following panel of microorganisms: *E. coli* DSM-1116, *E. coli* JW0451-2 (*acrB*-efflux pump deletion mutant of *E. coli* BW25113), *Pseudomonas aeruginosa* PA14, *Bacillus subtilis* DSM-10, *Mycobacterium smegmatis* mc2-155, *Staphylococcus aureus* Newman, *Candida albicans* DSM-1665, *Citrobacter freundii* DSM 30039, *Pichia anomala* DSM-6766 and *Acinetobacter baumannii* DSM30007. Microbroth dilution assays were conducted with prepared o/n cultures from cryogenically preserved long-term cultures and were diluted to achieve a final inoculum of  $10^4$ – $10^5$  cfu/mL. Serial dilutions of compounds were prepared in sterile 96-well plates in the respective test medium. The cell suspension was added and microorganisms were grown for 18–48 h at 37, 30 °C, respectively. Growth inhibition was evaluated by visual inspection and given as minimum inhibitory concentration (MIC) values (the lowest concentration of antibiotic at which no visible growth was observed). No inhibition of one of the tested microorganisms was observed at concentration up to 64 µg/mL of cittilin A.

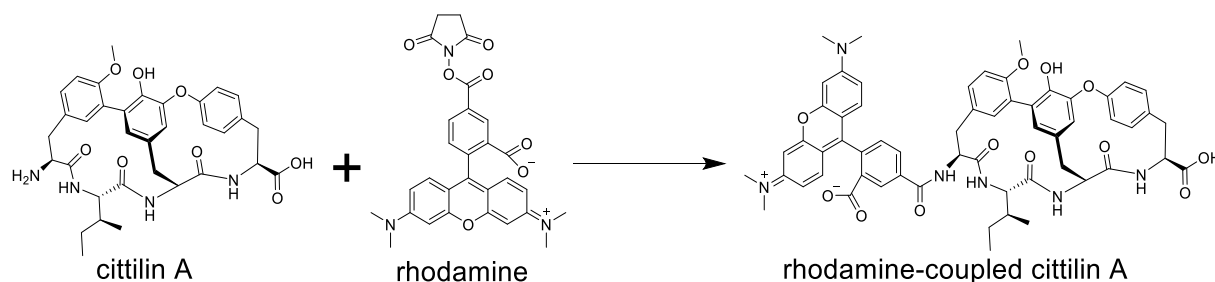
##### Cytotoxic activity

Cell lines were obtained from the German Collection of Microorganisms and Cell Cultures (Deutsche Sammlung für Mikroorganismen und Zellkulturen, DSMZ) or were part of our internal collection and were cultured under conditions recommended by the depositor. HCT-116 (human colon carcinoma cell line, DSMZ No. ACC 581) and KB-3-1 (cervix carcinoma cell line, DSMZ No. ACC 158) cells were seeded at  $6 \times 10^3$  cells per well of 96-well plates in 180 µL complete medium and treated with cittilin A in serial dilution after 2 h equilibration. After 5 days incubation, 20 µL of 5 mg/mL MTT (thiazolyl blue tetrazolium bromide) in phosphate buffered saline (PBS) was added per well and it was further incubated for 2 h at 37°C. The medium was discarded and cells were washed with 100 µL PBS before adding 100 µL isopropanol/10 N HCl (250:1) in order to dissolve formazan granules. The absorbance at 570 nm was measured using a microplate reader (Tecan Infinite M200Pro), and cell viability was expressed as percentage relative to the respective MeOH control. IC<sub>50</sub> values were determined by sigmoidal curve fitting. The IC<sub>50</sub> of cittilin A against HCT-116 cells was determined to 110.4 µg/mL and against KB-3-1 cells to 74.8 µg/mL.

**Pyocyanin assay<sup>49</sup>**

For determination of extracellular levels of pyocyanin produced by *Pseudomonas aeruginosa* strain PA14, cultivation was performed in the following way: cultures (initial OD<sub>600</sub> = 0.02) were incubated with or without inhibitor (final DMSO concentration 1%, v/v) at 37 °C, 200 rpm and a humidity of 75% for 16 h in 24-well Greiner BioOne. Cellstar plates containing 1.5 mL of PPGAS medium per well. Pyocyanin produced by PA14 was quantified using the method of Essar et al.<sup>50</sup> with some modifications, as described in detail by Klein et al.<sup>51</sup>. Briefly, 900 µL of each culture were extracted with 900 µL of CHCl<sub>3</sub> and 800 µL of the organic phase re-extracted with 250 µL of 0.2 M HCl. OD<sub>520</sub> was measured in the aqueous phase using FLUOstar Omega. For each sample, cultivation and sample work-up were performed in triplicates. Inhibition values of pyocyanin formation were normalized to OD<sub>600</sub>. Ten µM of cittilin A, inhibited 3.34% of pyocyanin production.

## Synthesis and purification of rhodamine-coupled cittilin A



**Figure S21.** Synthesis and purification of rhodamine-coupled cittilin A.

The following reaction was carried out under nitrogen atmosphere; To a solution of cittilin A (2.18 mg, 0.0035 mmol) in anhydrous DCM (3.0 mL) with molecular sieves 4 Å, a solution of 5-carboxy-tetramethylrhodamine *N*-succinimidyl ester (2.489 mg, 0.0047 mmol) was added in anhydrous DMF (0.5 mL) followed by *N,N*-Diisopropylethylamine (12 µL, 0.691 mmol). The mixture was stirred for six days at 28 °C and monitored by LC-MS. The mixture was dried (o/n) under vacuum to remove DCM and DMF. The residue was re-dissolved in 550 µL bidistilled MeOH and directly subjected to preparative RP-HPLC without further workup. Semi-preparative HPLC purification was done using a Dionex Ultimate 3000 SDLC low pressure gradient system on a XBridge peptide BEH C18 column 138 Å, 4.6 mm × 2500 mm column. Column temperature was stabilized at 45 °C with the eluents H<sub>2</sub>O + 0.1% FA as **A** and ACN + 0.1% FA as **B**, at a flow rate of 1.5 mL/min. Detection of rhodamine-coupled cittilin A was facilitated via mass spectrometry on the Agilent 1100 series coupled to the HCT 3D ion trap or with a UV detector on the Dionex 3000 SL systems by UV absorption at 256 nm and 320 nm. The gradient starts with a plateau at 95% **A** for 2 min followed by a ramp to 62% **A** during 6.3 min. Then, **A** content was kept to 62% during 6 min and finally ramped to 5% **A** during 1 min. **A** content is kept at 5% for 1 min and then ramped back to 95% during 30 s. The column was re-equilibrated at 95% **A** for 3 min. The pure bright pink compound was subsequently dried by lyophilization yielding 0.4 mg rhodamine-coupled cittilin A.

## Bacterial cell entry test

Cell entry of rhodamine-tagged cittilin A, cittilin A and free rhodamine was tested in Gram-negative bacterial cells (TolC efflux deficient *E. coli* mutant, from internal strain collection) by fluorescence microscopy of treated cultures. The bacterial cells were prepared as follows before testing cell entry:

- 5 mL pre-cultures of *E. coli* TolC were used (o/n, incubation at 37 °C, LB medium, OD<sub>600</sub>: ~1)
- O/n cultures were centrifuged (5 min, 8000 rpm, 4 °C in 15 mL Falcon tubes)
- SN was discarded, re-suspended CP with 5 mL PBS buffer, repeated centrifugation
- Discarded SN, re-suspended CP with 5 mL PBS buffer, split up re-suspended *E. coli* cells

**PFA fixation**

Transferred 4 x 500  $\mu$ L of re-suspended *E. coli* cells (in PBS buffer) to 4 x 2 mL Eppendorf tubes

500  $\mu$ L of re-suspended *E. coli* cells + 10  $\mu$ L of rhodamine-coupled cistilin A [1 mg/mL]

500  $\mu$ L of re-suspended *E. coli* cells + 21.7  $\mu$ L of free rhodamine [1 mg/mL]

500  $\mu$ L of re-suspended *E. coli* cells + 16.5  $\mu$ L of cistilin A [1 mg/mL]

500  $\mu$ L of re-suspended *E. coli* cells + 10  $\mu$ L of MeOH [100%]

**LC-MS analytic**

Transferred 4 x 500  $\mu$ L of re-suspended *E. coli* cells (in PBS buffer) to 4 x 2 mL Eppendorf tubes

500  $\mu$ L of re-suspended *E. coli* cells + 20  $\mu$ L of rhodamine-coupled cistilin A [1 mg/mL]

500  $\mu$ L of re-suspended *E. coli* cells + 43.4  $\mu$ L of free rhodamine [1 mg/mL]

500  $\mu$ L of re-suspended *E. coli* cells + 33  $\mu$ L of cistilin A [1 mg/mL]

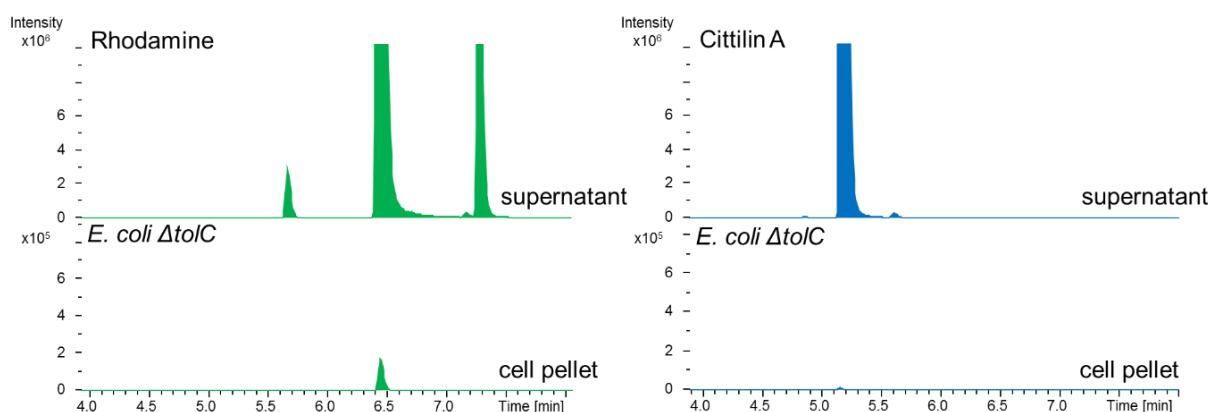
500  $\mu$ L of re-suspended *E. coli* cells + 10  $\mu$ L of MeOH [100%]

After supplementation of compounds/MeOH, samples were incubated for 30 min, 37 °C, 400 rpm (Eppendorf incubator), light protected. (A: confocal microscopy/ B: LC-MS analysis)

- Cultures were centrifuged (5 min, 8000 rpm, 4 °C)
- A) discarded SN, re-suspended CP with 500  $\mu$ L PBS buffer
- B) collected SN, re-suspended CP with 500  $\mu$ L PBS buffer
- Repeated centrifugation (5 min, 8000 rpm, 4 °C)
- A) discarded SN, added 1 mL PBS buffer + 4% PFA, mixed gently, incubation for 20 min, RT
- B) collected SN, re-suspended CP with 500  $\mu$ L PBS buffer, repeated centrifugation
- B) collected SN and CP. CP and SN were stored at -20 °C.
- A) After 20 min of PFA fixation, mixtures were centrifuged (5 min, 8000 rpm, 4°C)
- A) discarded SN, added 1 mL PBS buffer, re-suspended cells, repeated centrifugation
- A) discarded SN, CP contains mounted cells, which are now accessible for confocal microscopy

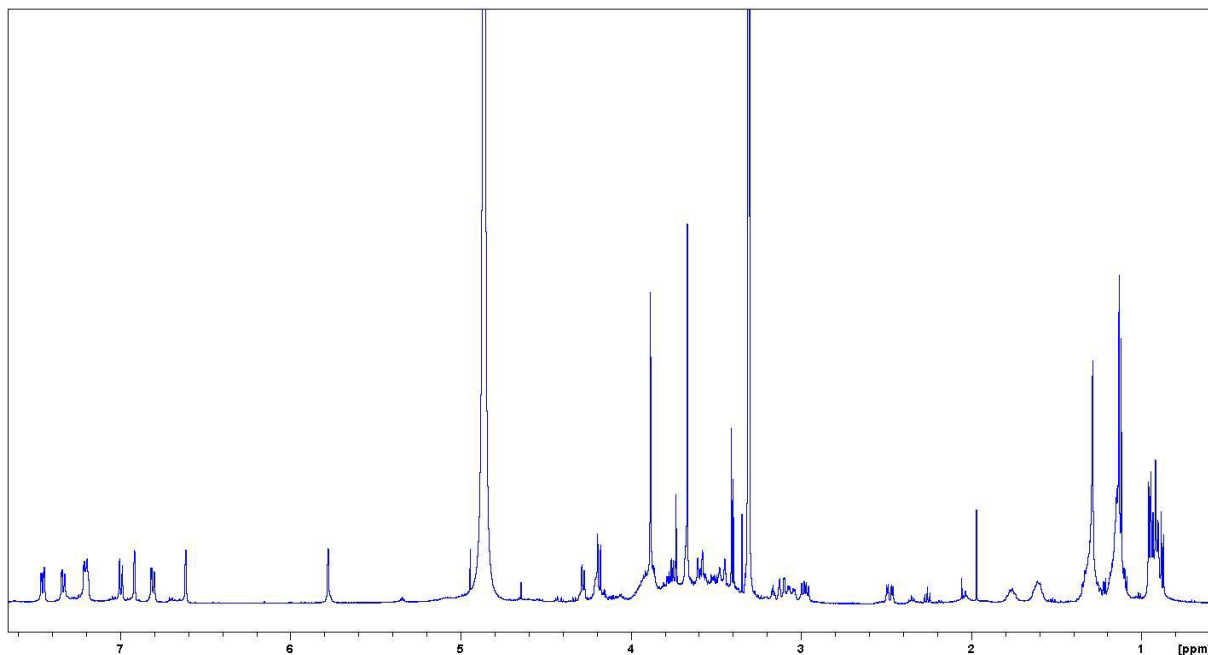
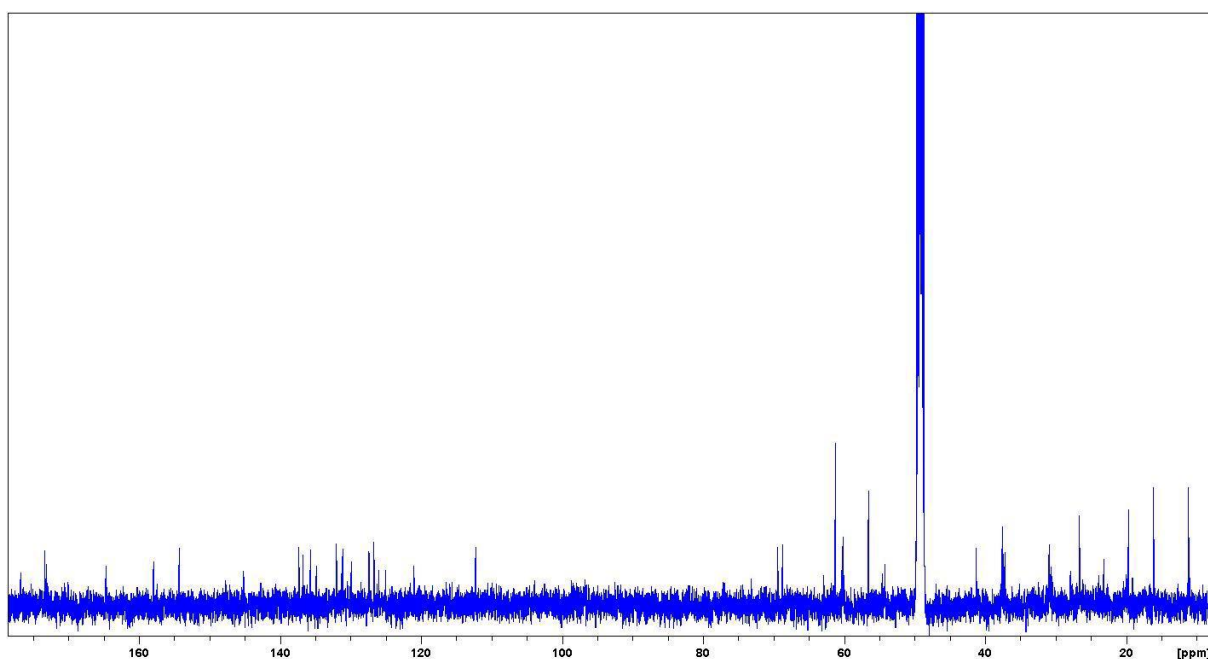
**LC-MS analysis**

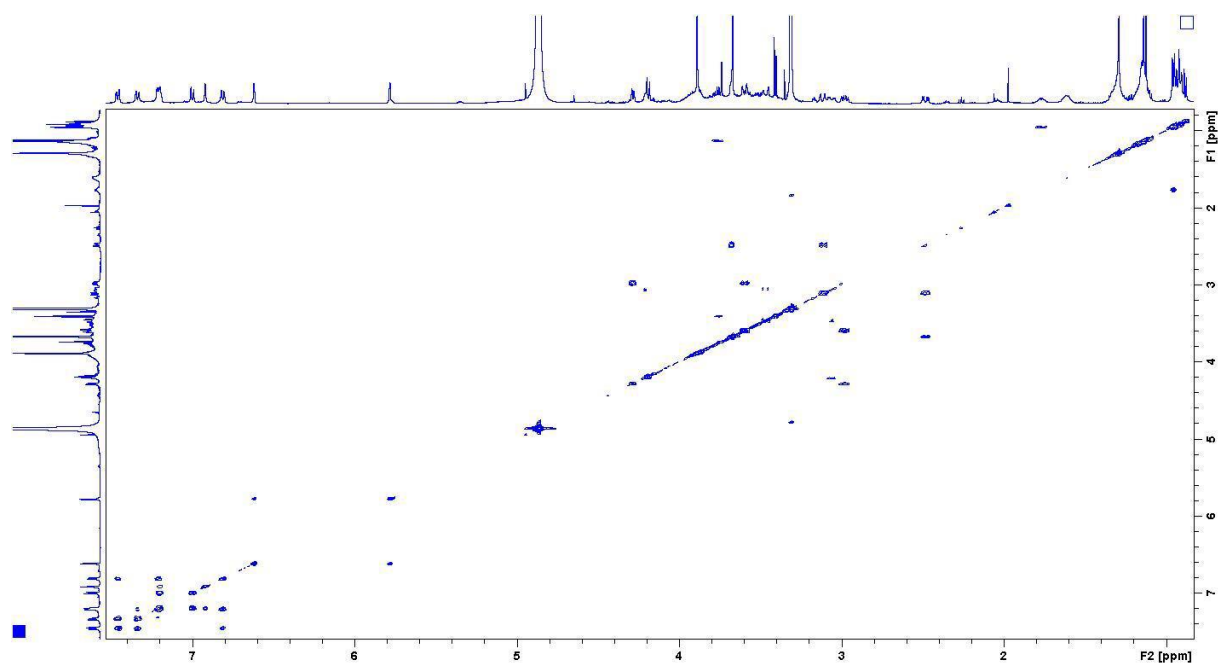
- Added to SN samples (four samples) 1.7 mL of MeOH and re-dissolved samples via sonication bathing (30 °C) for 5 min
- Added to CP samples (four samples) 0.5 mL MeOH and 0.5 mL acetone, extracted samples via sonication bathing (30 °C) for 5 min
- Centrifuged all samples (eight samples) for at least 10 min, max. speed, 4 °C (all samples in 2 mL Eppendorf tubes)
- Afterwards transferred SN of samples into brown glass vials (1.5 mL), dried samples under N<sub>2</sub> flow. Re-dissolved samples in 200 µL MeOH.
- Transferred 60 µL of each sample to 1.5 mL Eppendorf tubes. Centrifuged all samples for at least 10 min, max. speed, 4 °C (all samples in 2 mL Eppendorf tubes). Submitted undiluted samples for maXis 4G measurement as described above.



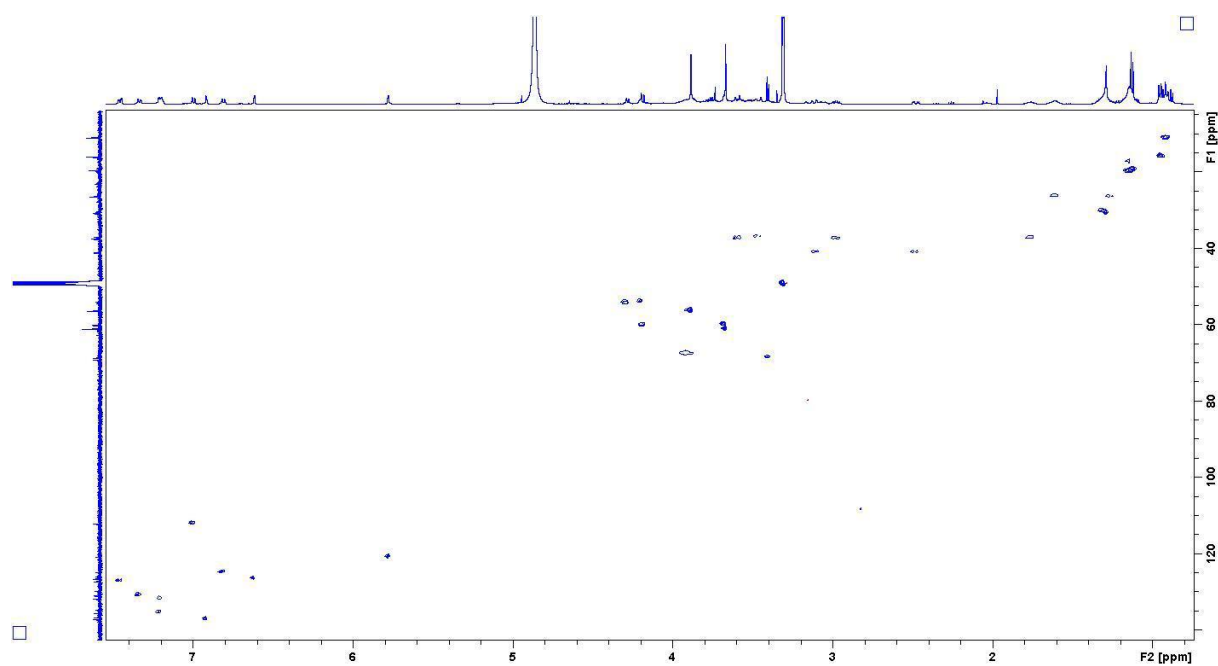
**Figure S22.** Test for bacterial uptake of cirtilin A by TolC-deficient *E. coli*. Under the conditions tested, cirtilin A is not detected in the bacterial extract. EIC: Extracted ion chromatogram, green: 431.1600 m/z, with a width of 7.9 ppm, free rhodamine [M+H]<sup>+</sup>; blue: 631.2768 m/z, with a width of 7.9 ppm, cirtilin A [M+H]<sup>+</sup>.



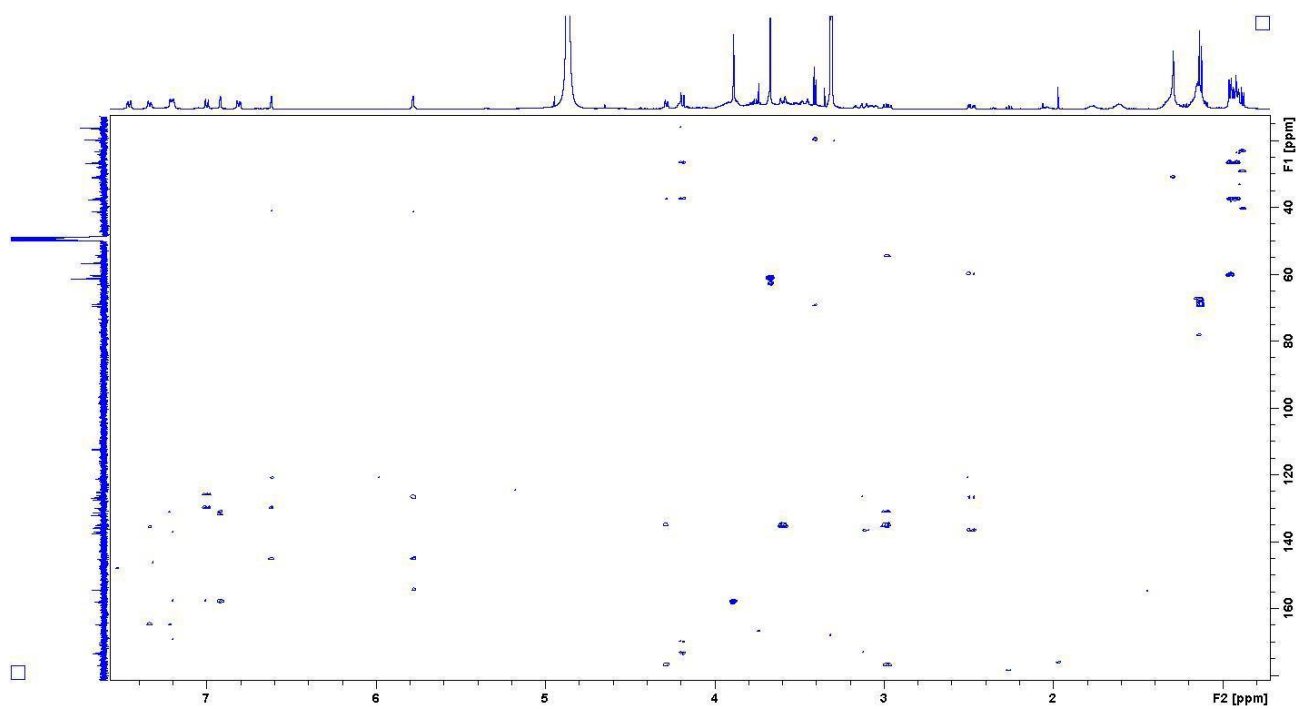
**2.7  $^1\text{H}$  and  $^{13}\text{C}$  NMR spectra of cittilin A and rhodamine-coupled cittilin A****Figure S23.**  $^1\text{H}$  NMR spectrum of cittilin A in MeOD.**Figure S24.**  $^{13}\text{C}$  NMR spectrum of cittilin A in MeOD.



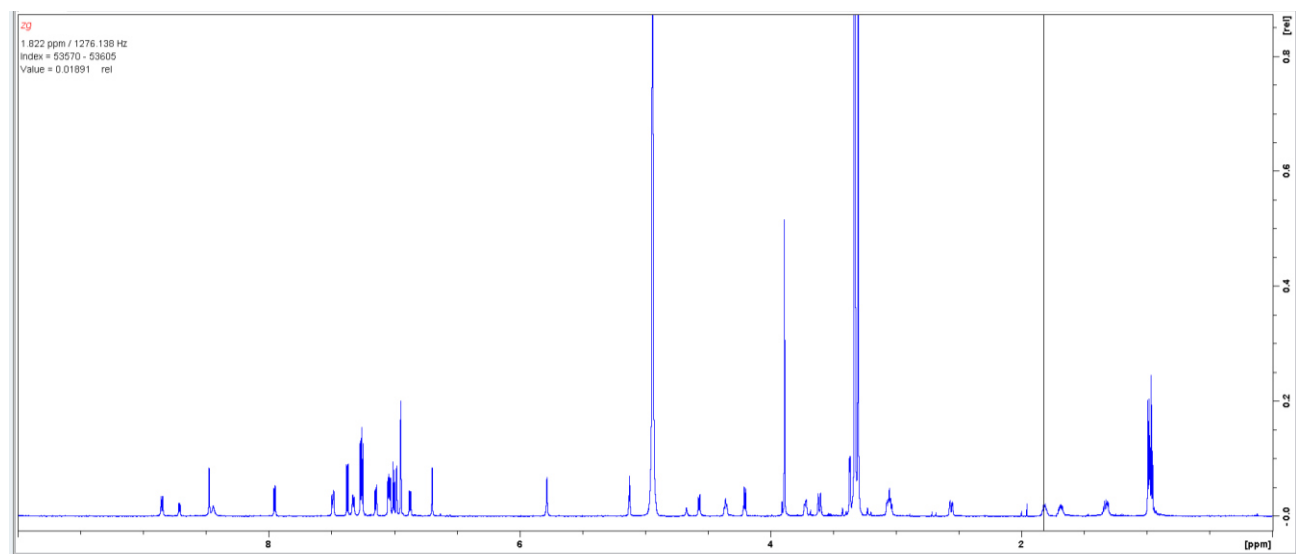
**Figure S25.** COSY spectrum of citilin A in MeOD.



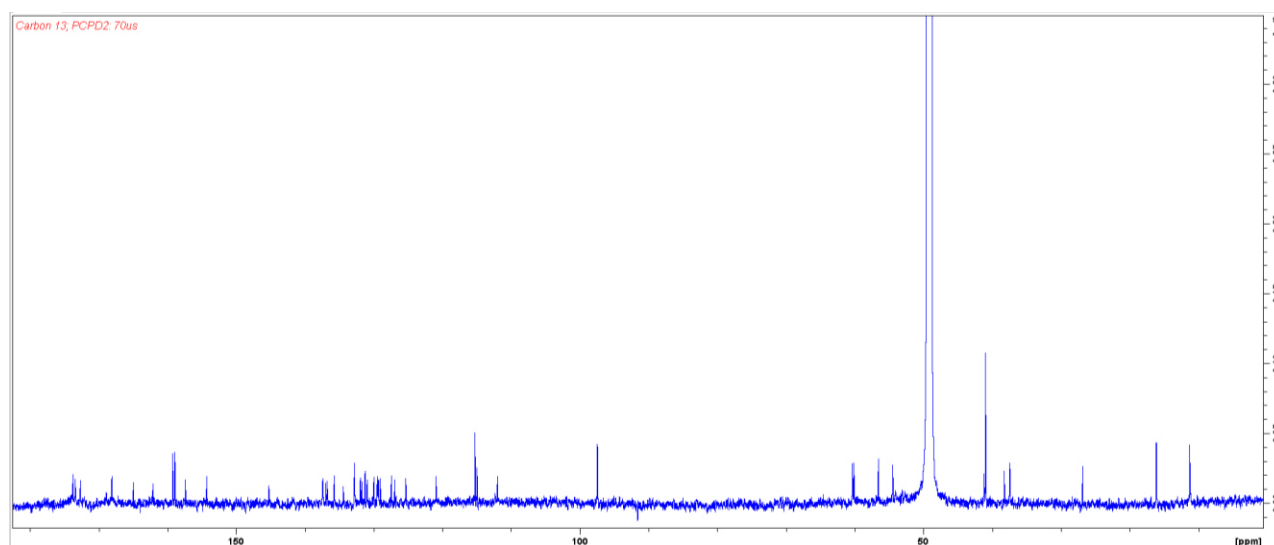
**Figure S26.** HSQC spectrum of citilin A in MeOD.



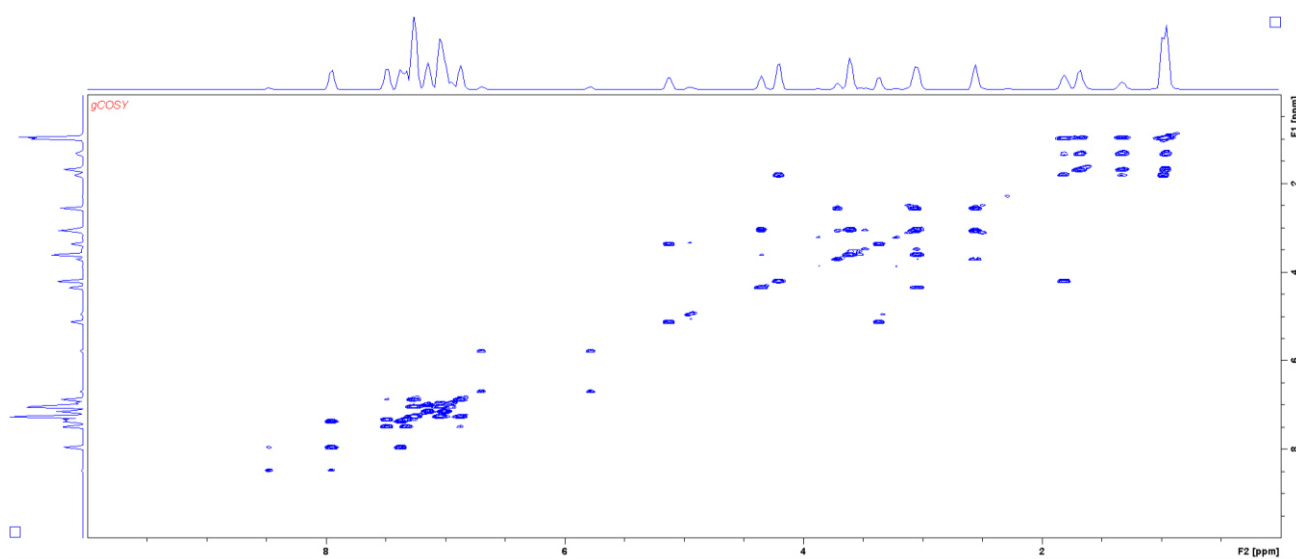
**Figure S27.** *HMBC* spectrum of citilin A in MeOD.



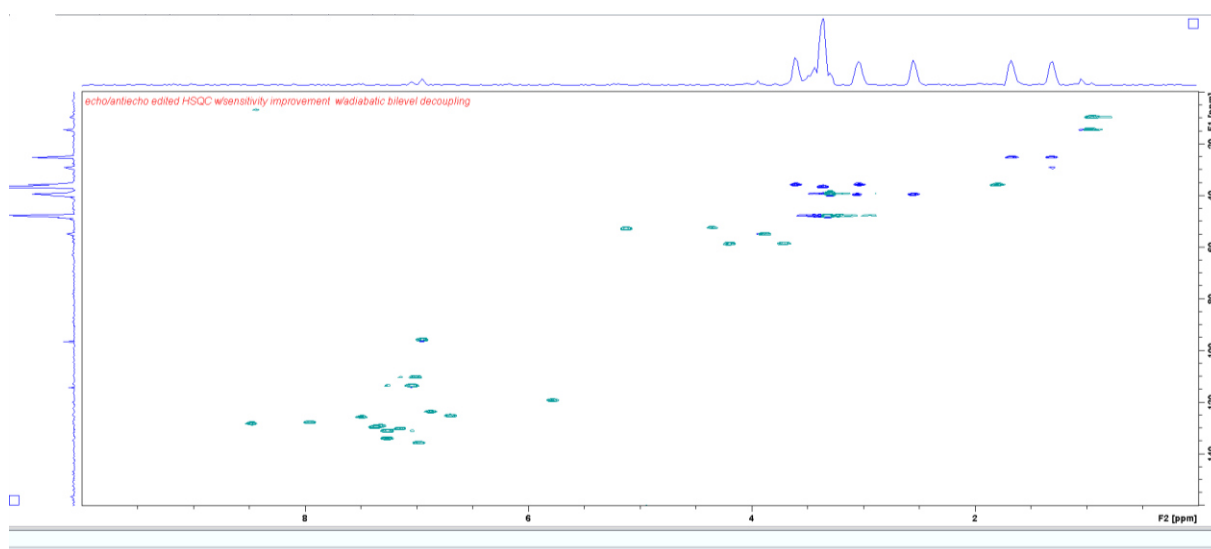
**Figure S28.**  $^1\text{H}$  NMR spectrum of rhodamine-coupled citilin A in MeOD.



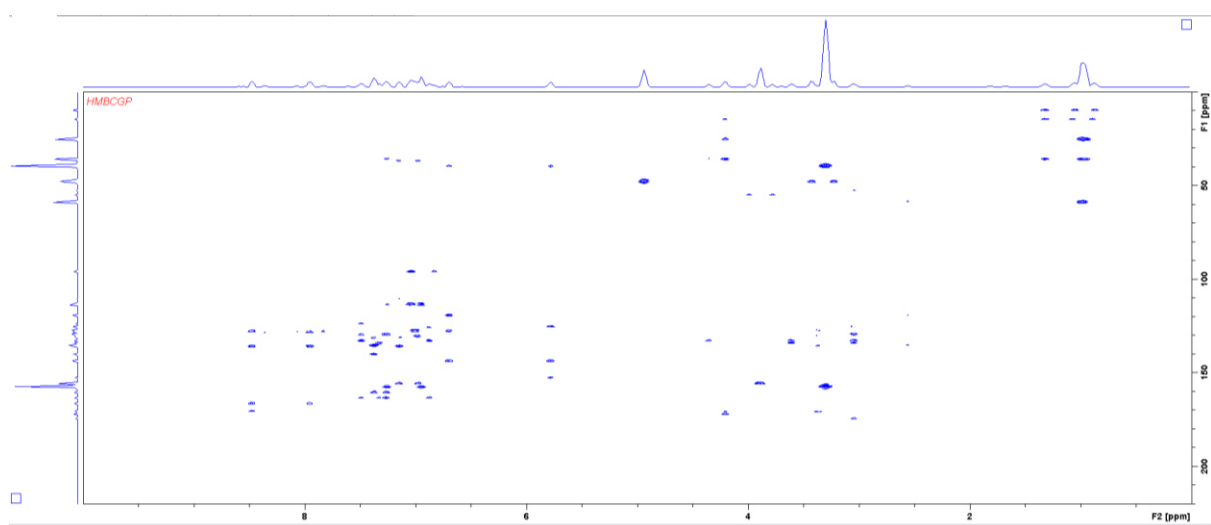
**Figure S29.**  $^{13}\text{C}$  NMR spectrum of rhodamine-coupled cittilin A in MeOD.



**Figure S30.** COSY NMR spectrum of rhodamine-coupled cittilin A in MeOD.



**Figure S31.** *Echo Antiecho* spectrum of rhodamine-coupled citilin A in MeOD.



**Figure S32.** *HMBC* of rhodamine-coupled citilin A in MeOD.

### 3. References

1. Edgar, R. C. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **32**, 1792–1797 (2004).
2. Finn, R. D. *et al.* The Pfam protein families database: towards a more sustainable future. *Nucleic Acids Res.* **44**, D279–285 (2016).
3. Hoffmann, T. *et al.* Correlating chemical diversity with taxonomic distance for discovery of natural products in myxobacteria. *Nat. Commun.* **9**, 803 (2018).
4. Irschik, H., Trowitzsch-Kienast, W., Gerth, K., Höfle, G. & Reichenbach, H. Saframycin Mx1, a new natural saframycin isolated from a myxobacterium. *J. Antibiot.* **41**, 993–998 (1988).
5. Yan, F. *et al.* Biosynthesis and Heterologous Production of Vioprolides: Rational Biosynthetic Engineering and Unprecedented 4-Methylazetidinecarboxylic Acid Formation. *Angew. Chem. Int. Ed. Engl.* **57**, 8754–8759 (2018).
6. Sasse, F. *et al.* Argyrins, immunosuppressive cyclic peptides from myxobacteria. I. Production, isolation, physico-chemical and biological properties. *J. Antibiot.* **55**, 543–551 (2002).
7. Gorges, J. *et al.* Structure, Total Synthesis, and Biosynthesis of Chloromyxamides: Myxobacterial Tetrapeptides Featuring an Uncommon 6-Chloromethyl-5-methoxypipericolic Acid Building Block. *Angew. Chem. Int. Ed. Engl.* **57**, 14270–14275 (2018).
8. Cortina, N. S., Revermann, O., Krug, D. & Müller, R. Identification and characterization of the althiomycin biosynthetic gene cluster in *Myxococcus xanthus* DK897. *ChemBioChem* **12**, 1411–1416 (2011).
9. Goldman, B. S. *et al.* Evolution of sensory complexity recorded in a myxobacterial genome. *Proc. Natl. Acad. Sci. USA* **103**, 15200–15205 (2006).
10. Li, Z. F. *et al.* Genome sequence of the halotolerant marine bacterium *Myxococcus fulvus* HW-1. *J. Bacteriol.* **193**, 5015–5016 (2011).
11. Wenzel, S. C. *et al.* Production of the bengamide class of marine natural products in myxobacteria: biosynthesis and structure-activity relationships. *Angew. Chem. Int. Ed. Engl.* **54**, 15560–15564 (2015).
12. Duddela, S. A bioinformatics approach for conceptual genome mining. Dissertation. Saarland University, 2015.
13. Irschik, H., Gerth, K., Kemmer, T., Steinmetz, H. & Reichenbach, H. The Myxovalargins, new peptide antibiotics from *Myxococcus fulvus* (Myxobacterales) I. Cultivation, isolation, and some chemical and biological properties. *J. Antibiot.* **36**, 6–12 (1983).

14. Hug, J. J., Panter, F., Krug, D. & Müller, R. Genome mining reveals uncommon alkylpyrones as type III PKS products from myxobacteria. *J. Ind. Microbiol. Biotechnol.* **46**, 319–334 (2019).
15. Green, M. R. & Sambrook, J. *Molecular cloning. A laboratory manual*. 4th ed. (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., 2012).
16. Mazodier, P., Petter, R. & Thompson, C. Intergeneric conjugation between *Escherichia coli* and *Streptomyces* species. *J. Bacteriol.* **171**, 3583–3585 (1989).
17. Panter, F., Krug, D., Baumann, S. & Müller, R. Self-resistance guided genome mining uncovers new topoisomerase inhibitors from myxobacteria. *Chem. Sci.* **9**, 4898–4908 (2018).
18. Luzhetskii, A. N., Ostash, B. E. & Fedorenko, V. A. Intergeneric Conjugation *Escherichia coli*–*Streptomyces globisporus* 1912 Using Integrative Plasmid pSET152 and Its Derivatives. *Russ. J. Genet.* **37**, 1123–1129 (2001).
19. Lukežič, T. *et al.* Identification of the chelocardin biosynthetic gene cluster from *Amycolatopsis sulphurea*: a platform for producing novel tetracycline antibiotics. *Microbiology* **159**, 2524–2532 (2013).
20. Liu, H. & Naismith, J. H. A simple and efficient expression and purification system using two newly constructed vectors. *Protein Expr. Purif.* **63**, 102–111 (2009).
21. Franz, L., Adam, S., Santos-Aberturas, J., Truman, A. W. & Koehnke, J. Macroamidine Formation in Botromycins Is Catalyzed by a Divergent YcaO Enzyme. *J. Am. Chem. Soc.* **139**, 18158–18161 (2017).
22. Wilkinson, C. J. *et al.* Increasing the efficiency of heterologous promoters in actinomycetes. *J. Mol. Microbiol. Biotechnol.* **4**, 417–426 (2002).
23. Myronovskyi, M. *et al.* Generation of a cluster-free *Streptomyces albus* chassis strains for improved heterologous expression of secondary metabolite clusters. *Metab. Eng.* **49**, 316–324 (2018).
24. Ziermann, R. & Betlach, M. C. Recombinant polyketide synthesis in *Streptomyces*: engineering of improved host strains. *Biotechniques* **26**, 106–110 (1999).
25. Donner, J. *et al.* The Biofilm Inhibitor Carolacton Enters Gram-Negative Cells: Studies Using a TolC-Deficient Strain of *Escherichia coli*. *mSphere* **2** (2017).
26. Krug, D. *et al.* Discovering the Hidden Secondary Metabolome of *Myxococcus xanthus*: a Study of Intraspecific Diversity. *Appl. Environ. Microbiol.* **74**, 3058–3068 (2008).
27. Wielgoss, S., Wolfensberger, R., Sun, L., Fiegna, F. & Velicer, G. J. Social genes are selection hotspots in kin groups of a soil microbe. *Science (New York, N.Y.)* **363**, 1342–1345 (2019).
28. Sharma, G., Narwani, T., Subramanian, S. & Gao, F. Complete genome sequence and comparative genomics of a novel myxobacterium *Myxococcus hansupus*. *PLoS ONE* **11**, e0148593 (2016).

29. Revermann, O. Novel secondary metabolites from myxobacteria and their biosynthetic machinery. Dissertation. Saarland University, 2012.
30. Mizuno, M. & Kanehisa, M. Distribution profiles of GC content around the translation initiation site in different species. *FEBS letters* **352**, 7–10 (1994).
31. Hasemann, C. A., Kurumbail, R. G., Boddupalli, S. S., Peterson, J. A. & Deisenhofer, J. Structure and function of cytochromes P450. A comparative analysis of three crystal structures. *Structure* **3**, 41–62 (1995).
32. Seifert, A. & Pleiss, J. Identification of selectivity-determining residues in cytochrome P450 monooxygenases. A systematic analysis of the substrate recognition site 5. *Proteins* **74**, 1028–1035 (2009).
33. Syed, K. & Mashele, S. S. Comparative analysis of P450 signature motifs EXXR and CXG in the large and diverse kingdom of fungi: identification of evolutionarily conserved amino acid patterns characteristic of P450 family. *PLoS ONE* **9**, e95616 (2014).
34. Zhao, B. *et al.* Crystal structure of albaflavenone monooxygenase containing a moonlighting terpene synthase active site. *J. Biol. Chem.* **284**, 36711–36719 (2009).
35. Christianson, D. W. Structural biology and chemistry of the terpenoid cyclases. *Chem. Rev.* **106**, 3412–3442 (2006).
36. Panter, F., Krug, D. & Müller, R. Novel Methoxymethacrylate Natural Products Uncovered by Statistics-Based Mining of the *Myxococcus fulvus* Secondary Metabolome. *ACS Chem. Biol.* **14**, 88–98 (2019).
37. Hug, J. J., Panter, F., Krug, D. & Müller, R. Genome mining reveals uncommon alkylpyrones as type III PKS products from myxobacteria. *J. Ind. Microbiol. Biotechnol.* **46**, 319–334 (2019).
38. Iniesta, A. A., García-Heras, F., Abellón-Ruiz, J., Gallego-García, A. & Elías-Arnanz, M. Two systems for conditional gene expression in *Myxococcus xanthus* inducible by isopropyl-β-D-thiogalactopyranoside or vanillate. *J. Bacteriol.* **194**, 5875–5885 (2012).
39. Kashefi, K. & Hartzell, P. L. Genetic suppression and phenotypic masking of a *Myxococcus xanthus* *frzF*- defect. *Mol. Microbiol.* **15**, 483–494 (1995).
40. Bibb, M. J., White, J., Ward, J. M. & Janssen, G. R. The mRNA for the 23S rRNA methylase encoded by the *ermE* gene of *Saccharopolyspora erythraea* is translated in the absence of a conventional ribosome-binding site. *Mol. Microbiol.* **14**, 533–545 (1994).
41. Flett, F., Mersinias, V. & Smith, C. P. High efficiency intergeneric conjugal transfer of plasmid DNA from *Escherichia coli* to methyl DNA-restricting streptomycetes. *FEMS Microbiol. Lett.* **155**, 223–229 (1997).



42. Paget, M. S. B., Chamberlin, L., Atrih, A., Foster, S. J. & Buttner, M. J. Evidence that the Extracytoplasmic Function Sigma Factor  $\sigma^E$  Is Required for Normal Cell Wall Structure in *Streptomyces coelicolor* A3(2). *J. Bacteriol.* **181**, 204–211 (1999).
43. Shan, L., Mathews, I. I. & Khosla, C. Structural and mechanistic analysis of two prolyl endopeptidases: Role of interdomain dynamics in catalysis and specificity. *Proc. Natl. Acad. Sci. USA* **102**, 3599–3604 (2005).
44. Kocadag Kocazorbaz, E. & Zihnioglu, F. Purification, characterization and the use of recombinant prolyl oligopeptidase from *Myxococcus xanthus* for gluten hydrolysis. *Protein Expr. Purif.* **129**, 101–107 (2017).
45. Nam-Sil, P., Hyun-Joo, P., Kyu-Boem, H. & Eung-Soo, K. Heterologous Expression of Novel Cytochrome P450 Hydroxylase Genes from *Sebekia benihana*. *J. Microbiol. Biotechnol.* **16**, 295–298 (2006).
46. Waugh, D. S. An overview of enzymatic reagents for the removal of affinity tags. *Protein Expr. Purif.* **80**, 283–293 (2011).
47. Omura, T. & Sato, R. The carbon monoxide-binding pigment of liver microsomes. I. Evidence for its hemoprotein nature. *J. Biol. Chem.* **239**, 2370–2378 (1964).
48. Abdulmughni, A. *et al.* Biochemical and structural characterization of CYP109A2, a vitamin D<sub>3</sub> 25-hydroxylase from *Bacillus megaterium*. *FEBS J.* **284**, 3881–3894 (2017).
49. Storz, M. P., Allegretta, G., Kirsch, B., Empting, M. & Hartmann, R. W. From *in vitro* to *in cellulo*: structure–activity relationship of (2-nitrophenyl)methanol derivatives as inhibitors of PqsD in *Pseudomonas aeruginosa*. *Org. Biomol. Chem.* **12**, 6094–6104 (2014).
50. Essar, D. W., Eberly, L., Hadero, A. & Crawford, I. P. Identification and characterization of genes for a second anthranilate synthase in *Pseudomonas aeruginosa*: interchangeability of the two anthranilate synthases and evolutionary implications. *J. Bacteriol.* **172**, 884–900 (1990).
51. Klein, T. *et al.* Identification of small-molecule antagonists of the *Pseudomonas aeruginosa* transcriptional regulator PqsR: biophysically guided hit discovery and optimization. *ACS Chem. Biol.* **7**, 1496–1501 (2012).



## Chapter 4

# 2-Hydroxysorangiadenosine: Structure and Biosynthesis of a Myxobacterial Sesquiterpene–Nucleoside

Dorothy A. Okoth<sup>†</sup>, **Joachim J. Hug<sup>†</sup>**, Ronald Garcia, Cathrin Sproer, Jörg Overmann and Rolf Müller

<sup>†</sup> These authors contributed equally to this work

*Molecules*

DOI: 10.3390/molecules25112676

Accepted manuscript: 05.06.2020

### Contributions to the presented work

#### Author's contribution

The author significantly contributed to the conception of this study, designed and performed experiments, evaluated and interpreted resulting data. He performed secondary metabolome-genome investigations of numerous myxobacterial strains including the producer of 2-hydroxysorangiadenosine, *Vitiosangium cumulatum* MCy10943<sup>T</sup>. In addition, the author conducted biosynthetic *in silico* analysis and investigated corresponding LC-MS data. Thereby the author identified the biosynthetic gene cluster and conceived a biosynthetic model for these unique myxobacterial sesquiterpene-nucleosides. Furthermore, the author contributed significantly to conceiving and writing this manuscript.

#### Contribution by others

Dorothy A. Okoth significantly contributed to the conception of this study, designed and performed experiments, evaluated and interpreted resulting data. She discovered and isolated 2-hydroxysorangiadenosine and sorangiadenosine through bioactivity-guided metabolome screening, performed NMR-based structure elucidation, and contributed to the proposed biosynthetic model. Furthermore, she contributed significantly to conceiving and writing this manuscript. Ronald Garcia performed large-scale fermentation of *Vitiosangium cumulatum* MCy10943<sup>T</sup> and prepared the strain for genome sequencing. Furthermore, he performed feeding experiments with stable isotope-labeled building blocks. Cathrin Sproer and Jörg Overmann performed genome sequencing and assembly of *Vitiosangium cumulatum* MCy10943<sup>T</sup>. Rolf Müller contributed by supervision of the project and conceiving, editing and proofreading of this manuscript.

## 4 2-Hydroxysorangiadenosine: Structure and biosynthesis of a myxobacterial sesquiterpene-nucleoside

### 4.1 Abstract

Myxobacteria represent an underinvestigated source for biologically active natural products featuring intriguing structural moieties with potential applications, e.g. in pharmaceutical industry. Sorangiadenosine and the here discovered 2-hydroxysorangiadenosine are myxobacterial sesquiterpene-nucleosides with an unusual structural moiety, a bicyclic eudesmane-type sesquiterpene. As the biosynthesis of these rare terpene-nucleoside hybrid natural products remained elusive, we investigated secondary metabolomes and genomes of several 2-hydroxysorangiadenosine producing myxobacteria. We report the isolation and full structure elucidation of 2-hydroxysorangiadenosine, its cytotoxic and antibiotic activities and propose a biosynthetic pathway in the myxobacterium *Vitiosangium cumulatum* MCy10943<sup>T</sup>.

### 4.2 Introduction

Myxobacteria are remarkable Gram-negative bacteria exhibiting several unusual characteristics such as coordinated movement by gliding or creeping on surfaces in a swarm-like pattern<sup>1</sup>. Unlike other prokaryotes, they show an unique cooperative “social behavior”, based on a complex chemical communication systems<sup>2</sup>. A wide variety of mode of actions have been documented for myxobacterial secondary metabolites with diverse bioactivities<sup>3</sup> such as the antifungal soraphen A<sup>4</sup>, the antibacterial cystobactamid 919-2<sup>5</sup>, the antiplasmodial chlorotonil<sup>6</sup>, the antiviral aetheramides<sup>7</sup> and the cytotoxic pretubulysin<sup>8,9</sup>. Most of these compounds derive biosynthetically from giant biosynthetic enzyme complexes such as modular polyketide synthases (PKSs), non-ribosomal peptide synthetases (NRPSs) and hybrids thereof, while natural products from other biosynthetic machineries have been less frequently isolated.

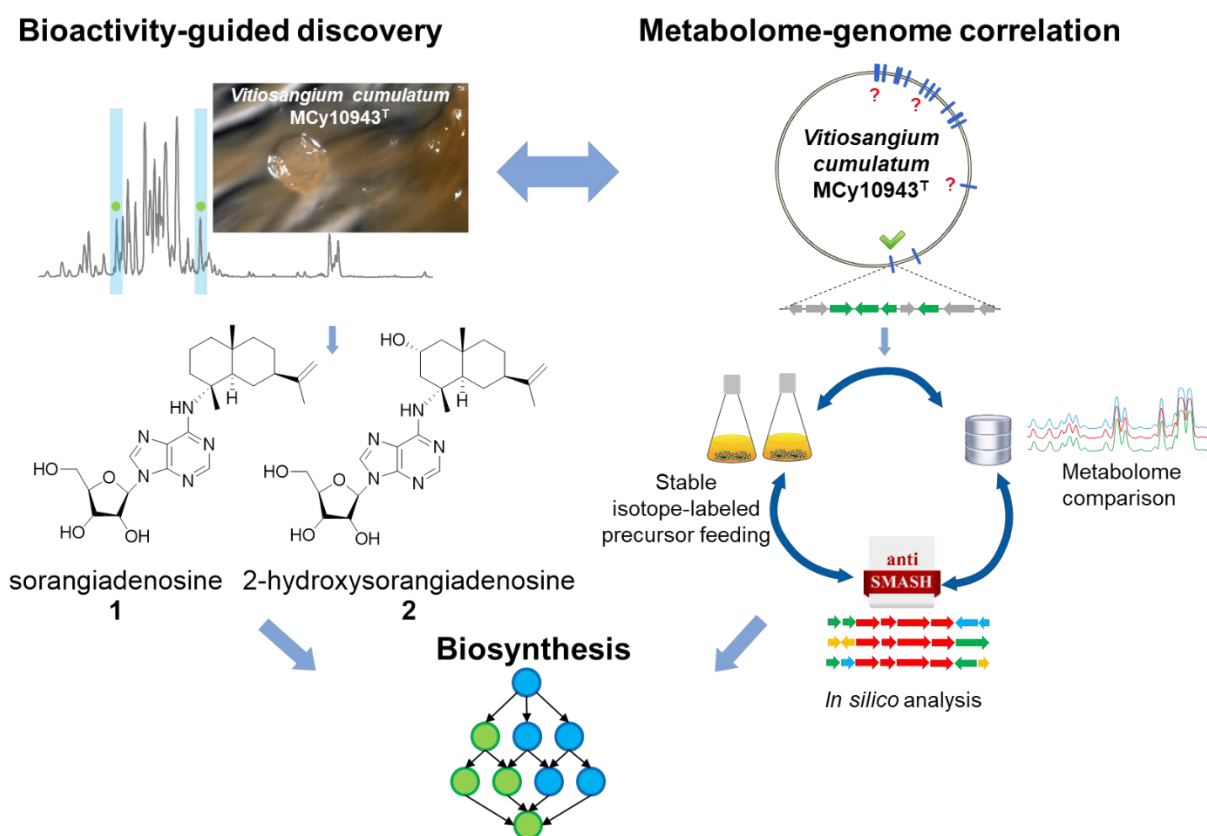
Terpene natural products in myxobacteria have been primarily investigated by analysis of the volatiles emitted from cell cultures of *Myxococcus xanthus* sp.<sup>10</sup> and *Stigmatella aurantiaca* sp.<sup>11</sup> through a closed-loop stripping apparatus and subsequent gas chromatography-mass spectrometry (GC-MS) analysis. These investigations revealed not only the biosynthesis of several myxobacterial volatiles through stable isotope-labeled precursor feeding, but permitted also further insights into the formation of the well-known terpenes geosmin<sup>12</sup> and 2-methylisoborneol<sup>13,14</sup>. The biosynthetic characterization of (+)-eremophilene biosynthesis<sup>15</sup> and the function of terpene synthase Soce6369 (10-*epi*-cubenol synthase)<sup>16</sup> in *Sorangium cellulosum* So ce56 provided additional in-depth knowledge of terpene biosynthesis in myxobacteria.

The sesquiterpene adenoside sorangiadenosine (**1**) has been isolated from the crude organic extract of *Sorangium cellulosum* KM1003<sup>17</sup>. Evaluation of its biological function indicated moderate antibacterial activity against the test strains *Staphylococcus aureus* ATCC 6538p, *Bacillus subtilis* ATCC 6633,

*Micrococcus luteus* IFC 12708, *Proteus vulgaris* ATCC 3851, *Salmonella typhimurium* ATCC 14028, and *Escherichia coli* ATCC 25922<sup>17</sup>.

Our current interest was focused on deciphering the biosynthetic logic of this terpene-nucleoside hybrid. Terpene-nucleoside hybrids are a rarely described class of natural products, reflected by the fact that **1** is the only known terpene-nucleoside hybrid from a myxobacterium.

We report here the isolation, and the cytotoxic and antibacterial activities together with the full structure elucidation of the new sorangiadenosine derivative 2-hydroxysorangiadenosine (**2**) from the myxobacterial producer *Vitosangium cumulatum* MCy10943<sup>T</sup><sup>18</sup>, which belongs to a different myxobacterial family than the previously described producer of **1**. In addition, genome sequencing of the producer strain *V. cumulatum* MCy10943<sup>T</sup> and comparative metabolome analysis of several additional myxobacterial strains led to the identification of a putative biosynthetic gene cluster (BGC) involved in the formation of **1** and **2** (Figure 1).



**Figure 1.** Bioactivity-guided metabolome screening of *Vitosangium cumulatum* MCy10943<sup>T</sup> led to the re-discovery of sorangiadenosine (**1**) and the discovery of 2-hydroxysorangiadenosine (**2**). The molecular weight, sum formula and the MS<sup>2</sup> fragmentation pattern of the unknown molecules enabled identification and isolation of **1** and **2**. The corresponding biosynthetic gene cluster was identified via *in silico* investigation of the MCy10943<sup>T</sup> genome sequence. Stable isotope-labeled precursor feeding experiments and comparative metabolome investigation of genome sequenced myxobacterial strains corroborated this analysis. Finally, the outcome of this study led to a proposed biosynthetic route responsible for the biosynthesis of these unique sesquiterpene-adenosine hybrids.

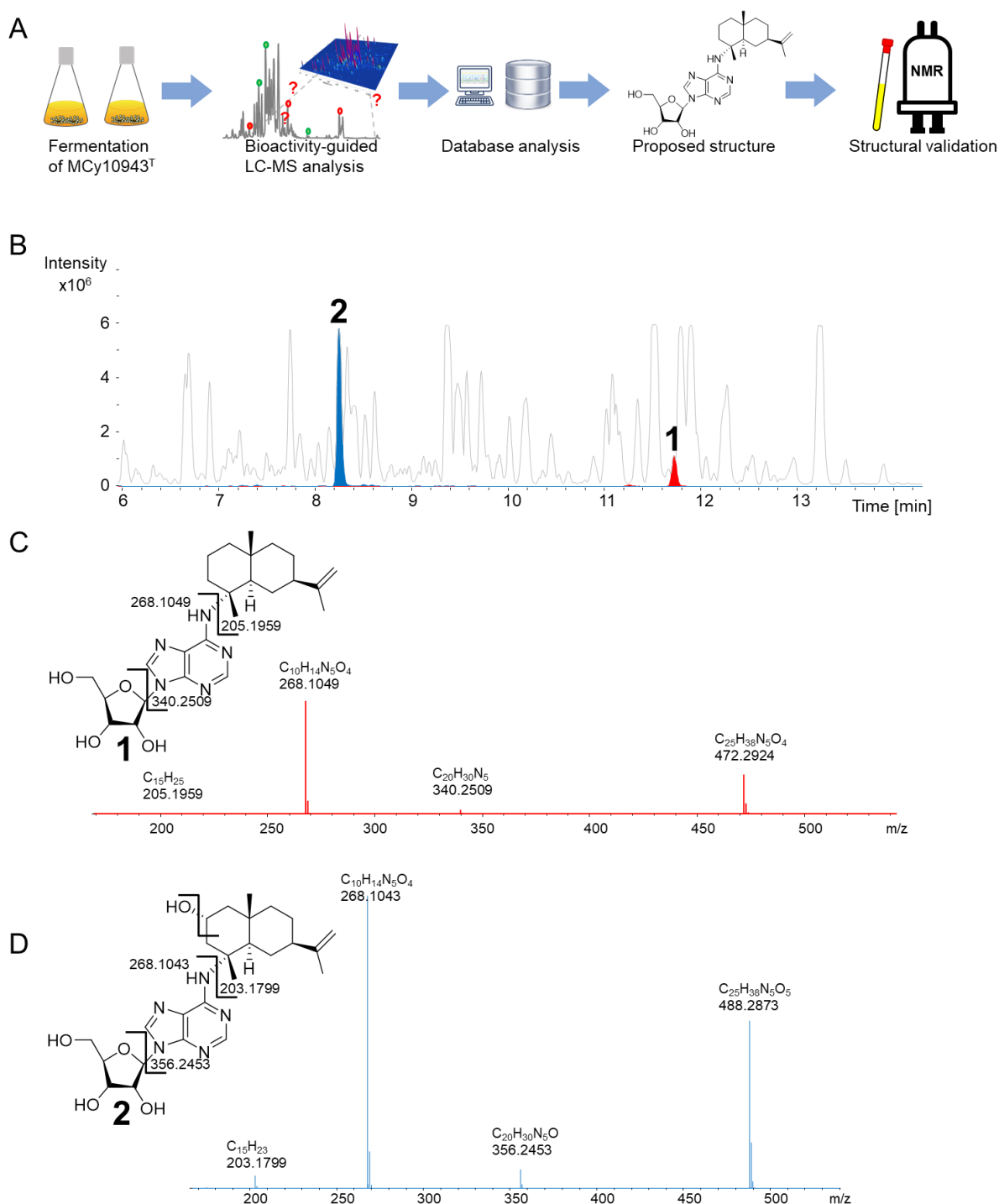
### 4.3 Results and Discussion

#### Discovery of 2-hydroxysorangiadenosine and sorangiadenosine

In the course of our bioactivity-guided secondary metabolite screening of the recently isolated strain *V. cumulatum* MCy10943<sup>T 18</sup> (**Figure 2A**), HPLC-fractions of the crude extract active against *Bacillus subtilis* revealed two target masses. One secondary metabolite features an ion peak in the LC-MS chromatogram at  $m/z$  472.2924  $[M+H]^+$  supporting the deduced molecular formula  $C_{25}H_{38}N_5O_4$  at the retention time of 11.72 min, while the second target mass features an ion peak at  $m/z$  488.2873  $[M+H]^+$  supporting the deduced molecular formula  $C_{25}H_{38}N_5O_5$  at the retention time of 8.24 min (**Figure 2B**). A search in several chemical structure databases revealed that the target mass of 472.2924  $[M+H]^+$  might correspond to the previously isolated natural product **1**, whereas the second natural product **2** could be a derivative thereof, according to the observed MS<sup>2</sup> fragmentation pattern of both compounds. The MS<sup>2</sup> fragment of **1** was characterized by  $m/z$  340.2509 ( $C_{20}H_{30}N_5^+$ ), 268.1047 ( $C_{10}H_{14}N_5O_4^+$ ) and 205.1959 ( $C_{15}H_{25}^+$ ) consistent with  $[M+H+sugar]^+$ , adenosine and sesquiterpene  $[M+H]^+$  fragment ions (**Figure 2C**). Similarly the MS<sup>2</sup> fragment of compound **2** was characterized by  $m/z$  356.2453  $[M+H]^+$  ( $C_{20}H_{30}N_5O^+$ ),  $m/z$  268.1043 ( $C_{10}H_{14}N_5O_4^+$ ) and  $m/z$  203.1799 ( $C_{15}H_{23}^+$ ) consistent with  $[M+H+sugar]^+$ , [adenosine+H]<sup>+</sup>, and  $[M+H+adenosine+OH]^+$  ion peaks (**Figure 2D**). This suggested that both compounds were made of a sesquiterpene, adenine and a sugar moiety with compound **2** having an additional oxygen atom (**Figure 2C, D**). Since *S. cellulosum* KM1003, the original myxobacterial producer of **1** was not deposited in a publicly accessible collection, and reference substance was not commercially available, the identity of the detected compound was validated via isolation and structure elucidation of the identified compounds. The recorded proton (<sup>1</sup>H) nuclear magnetic resonance (NMR) and carbon-13 (<sup>13</sup>C) NMR of both purified compounds proved the identity of these compounds as **1** and the new derivative **2** (**Figure 3**).

The <sup>1</sup>H NMR and <sup>13</sup>C NMR data of **1** was consistent with the observed correlations of the previously isolated sorangiadenosine from a *Sorangium cellulosum* (KM1003)<sup>17</sup>. The <sup>1</sup>H and <sup>13</sup>C NMR data of **2** resembled that of **1**. The heteronuclear single quantum coherence/correlation (HSQC) and <sup>13</sup>C NMR of **1** showed 25 carbons; six quaternary, eight methines, eight methylenes and three methyl groups; the HSQC and <sup>13</sup>C NMR for **2** showed 25 carbons; six quaternary, nine methines, seven methylenes and three methyl groups. Like **1**, a combination of one-dimensional (1D) and 2D NMR indicated that **2** also consists of a heteroaromatic moiety (adenine), a pentose sugar and a sesquiterpene. The structure of the sesquiterpene was fully elucidated based on its <sup>1</sup>H and <sup>13</sup>C NMR data. Fifteen carbons (3-quaternary carbons, three methyl groups, six methylenes, two methines, one oxymethine) were assigned to the sesquiterpene moiety. The NMR spectra of **2** were similar to **1**<sup>17</sup>, with the exception of a secondary alcohol ( $\delta_C$  65.6,  $\delta_H$  3.94, 1H, m). <sup>1</sup>H-<sup>1</sup>H correlation spectroscopy (COSY) correlations were observed between the oxymethine proton ( $\delta_H$  3.94, m) and two pairs of methylene protons H-1 ( $\delta_H$  1.27, 1H, *t* and 1.80, 1H, *dd*) and H-3 ( $\delta_H$  2.52, 1H, *t* and 2.41, 1H, m) indicated a partial structure of -CH<sub>2</sub>-CH(OH)-CH<sub>2</sub>-. Further COSY correlations of H-6 ( $\delta_H$  1.70, 1H, m and  $\delta_H$  1.38, 1H, *t*) to H-5 ( $\delta_H$  2.60, 1H, *dd*) and H-7 ( $\delta_H$  1.94, 1H, m), H-8 ( $\delta_H$  1.56, 1H, m and  $\delta_H$  1.38, 1H, *t*) to H-7 ( $\delta_H$  1.94, 1H, m) and H-9 ( $\delta_H$  1.52, 1H, m and  $\delta_H$  1.38, 1H, *t*) lead to formation of another partial structure of -CH-CH<sub>2</sub>-CH-CH<sub>2</sub>-CH<sub>2</sub>-. <sup>1</sup>H-<sup>13</sup>C heteronuclear multiple bond correlation (spectroscopy) (HMBC) correlations between H-3 protons ( $\delta_H$  2.52 1H, *t* and 2.41, 1H, m) and C-15 ( $\delta_C$  22.4), C-1 ( $\delta_C$  50.8), C-2 ( $\delta_C$  65.6) and C-4 ( $\delta_C$  60.4). Further <sup>1</sup>H-<sup>13</sup>C HMBC correlations were observed between (H-1 protons)  $\delta_H$  1.27 and 1.80 and

C-14 ( $\delta_c$  20.7), C-3 ( $\delta_c$  46.6), C-10 ( $\delta_c$  35.6), C-2 ( $\delta_c$  65.6) and C-9 ( $\delta_c$  46.4). These enabled the assignment of the alcohol at C-2.



**Figure 2.** A) Workflow of secondary metabolite screening in MCy10943<sup>T</sup>, resulting in the independent re-isolation of sorangiadenosine (**1**) and the discovery of 2-hydroxysorangiadenosine (**2**). B) HPLC-MS BPC (grey) and EIC of **1** ( $[M+H]^+$  472.2923  $m/z$ , red) and **2** ( $[M+H]^+$  488.2873  $m/z$ , blue) from MCy10943<sup>T</sup> crude extracts. Fragmentation of sorangiadenosine (C) and 2-hydroxysorangiadenosine (D) in MS<sup>2</sup> experiments.

Additional HMBC cross peaks between H-5 proton ( $\delta_H$  2.60) and C-1 ( $\delta_c$  50.8), C-4 ( $\delta_c$  60.4), C-10 ( $\delta_c$  35.6), C-14 ( $\delta_c$  20.7) and that between H-14 methyl protons to C-10 ( $\delta_c$  35.6), C-9 ( $\delta_c$  46.4) and C-5

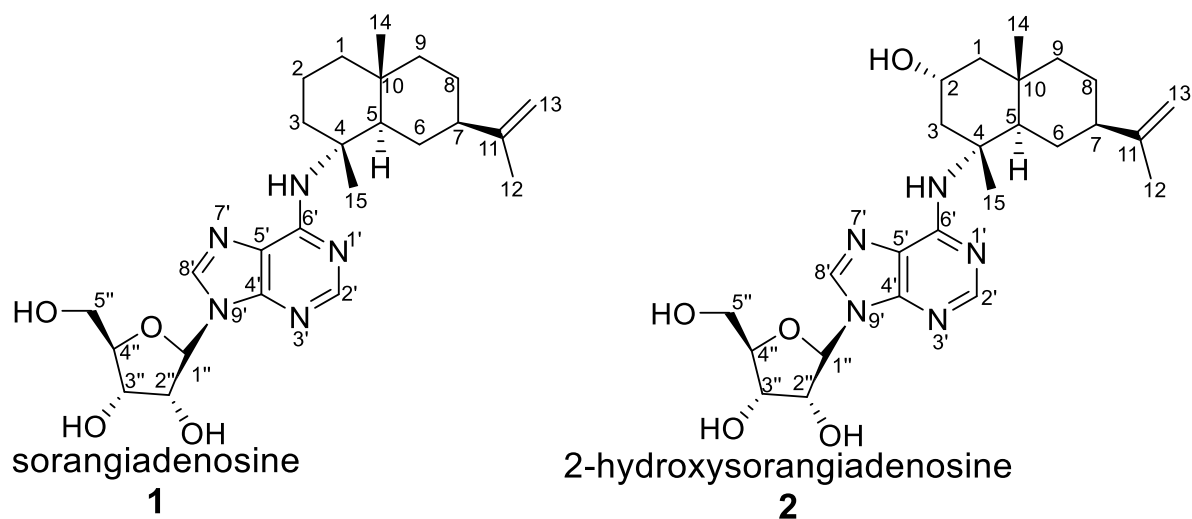


( $\delta_C$  48.5) allowed the completion of the bicyclic ring. The vinyl signals ( $\delta_C$  109.0 (C-13),  $\delta_H$  4.65 (H-13) and  $\delta_C$  151.5 (C-11)) and a methyl group ( $\delta_C$  21.1,  $\delta_H$  1.68) corresponded to the sesquiterpene head in eudesmane-type sesquiterpenes. HMBC correlations between the H-7 proton ( $\delta_H$  1.94, 1H, m) and C-13 ( $\delta_C$  109.0), C-11 ( $\delta_C$  151.5) and C-12 ( $\delta_C$  21.1) confirmed the attachment of the isopropylene unit at C-7 of the sesquiterpene ring.

The relative stereochemistry at the asymmetric carbon centers at C-4, C-5, C-7, and C-10 was assigned by rotating frame nuclear Overhauser enhancement spectroscopy (ROESY) experiments. A number of cross-peaks at H-2 $\beta$  ( $\delta$  3.94)/H-1 $\beta$  ( $\delta$  1.80), H-2  $\beta$ /H-3 $\beta$  ( $\delta$  2.41), H-2  $\beta$  ( $\delta$  3.94)/H-14 ( $\delta$  1.06), H-2  $\beta$ /H-15 ( $\delta$  1.46), H-6  $\beta$  ( $\delta$  1.38)/H-8 ( $\delta$  1.56), H-6  $\beta$ /H-9  $\beta$  ( $\delta$  1.52), H-6  $\beta$ /H-14, H-6  $\beta$ /H-15, H-8  $\beta$  ( $\delta$  1.56)/H-14, H-14/H-15 were observed. These 1,3-diaxial correlations of H-14 and H-15 methyl protons and neighboring protons indicated that their orientation was axial. In contrast other series of ROESY correlations at H-1 $\alpha$  ( $\delta$  1.30)/H-3  $\alpha$  ( $\delta$  2.52), H-1/H-5 ( $\delta$  2.60), H-1/H-9  $\alpha$  ( $\delta$  1.38), H-3/H-5, H-5/H-6  $\alpha$  ( $\delta$  1.70), H-5/H-7 ( $\delta$  1.94), H-5/H-12 ( $\delta$  1.68), H-5/H-9  $\alpha$  ( $\delta$  1.38), showed these protons had an opposite orientation at the decalin plane.

Thus, the relative configurations were assigned to be *trans* (C-5 and C-10) and 2*R*, 4*R*, 5*R*, 7*R*, 10*S* for the ring junction asymmetric carbons, respectively. Assignment of the relative stereochemical configuration of the C-2 hydroxyl group was based on ROESY correlations observed between the H-2 $\beta$  proton ( $\delta_H$  3.94) and C-14 ( $\delta_H$  1.06) and C-15 ( $\delta_H$  1.46) methyl groups. The presence of two downfield aromatic protons at  $\delta_H$  8.24 (1H, s, H-2') and  $\delta_H$  8.26 (1H, s, H-8') suggested they were connected to nitrogen atoms. These two protons were HMBC coupled to five downfield carbons  $\delta_C$  152.9 (C-2'), 148.8 (C-4'), 121.6 (C-5'), 155.7 (C-6') and 141.2 (C-8') confirmed the presence of the adenine base<sup>19</sup>. Long-range HMBC correlations between the C-15 methyl proton signals ( $\delta_H$  1.46) attached to C-4 ( $\delta_C$  60.4) of sesquiterpene and C-6' ( $\delta_C$  155.7) of the adenine confirmed the position of linkage between the C-4 of eudesme-11-ne and the adenine base. The absence of any observable HMBC correlation between the C-15 methyl protons and the C-2' carbon of the adenine, confirms that the point of sesquiterpene attachment cannot be at the N-1' position of adenine.

The ribofuranosyl was characterized by six proton signals in the region 3.74–5.94 in the <sup>1</sup>H NMR spectrum. Based on COSY, HSQC and HMBC correlations five carbons  $\delta_C$  91.3 (C-1'), 88.3 (C-4'), 75.4 (C-2'), 72.7 (C-3'), and 63.5 (C-5') were assigned to the sugar moiety. A coupling constant of 6.5 Hz of the anomeric proton H-1'' ( $\delta_H$  5.94), a larger magnitude (>2.15 ppm) in the difference between H-2' ( $\delta_H$  8.24) and H-1'' ( $\delta_H$  5.94) protons indicated a  $\beta$ -anomer sugar<sup>20</sup>. The linkage between the adenine and ribose moiety was based on HMBC correlation between anomeric proton H-1'' ( $\delta_H$  5.94) and  $\delta_C$  148.8 (C-4') and 141.2 (C-8'). ROESY correlations observed between the anomeric proton H-1''/H-4', H-1''/H-2'', H-1''/H-3'' and that between H-2''/H-3'' and H-2'/H-4' suggests a *cis* conformation of the H-2'' and H-3'' hydroxyl groups. Comparison of the NMR data with literature values suggested a D-ribose sugar<sup>21</sup>. In addition, the observed optical rotation of both compounds (**1** [ $\alpha$ ]<sub>D</sub><sup>20</sup> -76.92 (MeOH) and **2** [ $\alpha$ ]<sub>D</sub><sup>20</sup> -70.92 (MeOH) were negative as in the case of previously isolated **1** [ $\alpha$ ]<sub>D</sub><sup>20</sup> -78.7<sup>17</sup>. Moreover, the circular dichroism (CD) spectrum of **2** shows two negative cotton effect bands at 260 and 215 nm (**Supporting Information, Figure S13**), consistent with the CD spectrum from adenosine<sup>22–24</sup>, which features like most of the naturally-occurring monosaccharides a D-configured ribose<sup>22–24</sup>. Finally, the subsequently identified biosynthetic origin of **1** and **2**, strikingly suggest a D-configured ribose (see below).



**Figure 3.** Chemical structure and numbering of sorangiadenosine (**1**) and 2-hydroxysorangiadenosine (**2**).

### Biological activity of 2-hydroxysorangiadenosine and sorangiadenosine

The sesquiterpene-nucleoside compounds **1** and **2** showed biological activity against the Gram-positive bacteria *B. subtilis* DSM 10 and *S. aureus* Newman (**Table 1**). **1** was cytotoxic towards HCT and KB3.1 cells at IC<sub>50</sub> of 30.0 µg/mL and 39.46 µg/mL while **2** featured activity against HCT cell lines at IC<sub>50</sub> of 52.0 µg/mL (**Table 2**). In conclusion, **1** shows better bioactivity compared to **2**. This may be due to the additional oxygen atom in **2**, which increases significantly the polarity of the sesquiterpene-nucleoside scaffold and limits the cell membrane penetration capability of **2** (a finding also reflected in the difference of the retention times).

The results of our bioactivity profiling of **1** (and **2**) are resembling the findings of the previously observed antimicrobial range of **1**<sup>17</sup>. Albeit the fact that former bioactivity screening of **1** displayed moderate bioactivity against the Gram-negative bacterium *Proteus vulgaris* ATCC 3851 (Minimum inhibitory concentration (MIC) at 6.25 µg/mL) and *Salmonella typhimurium* ATCC 14028 (MIC at 12.5 µg/mL)<sup>17</sup>, the biological activity of these sesquiterpene-nucleoside seems to be focused on Gram-positive bacteria according to our findings. The lack of biological activity of **1** and **2** against *E. coli* test strains (*E. coli* wild type (WT) (DSM 1116) and *E. coli* *acrB* JW0451-2 see **Table 1**, *Escherichia coli* ATCC 25922 see Ahn *et al.*<sup>17</sup>) underlines this conclusion and highlights the difficulties to find new compounds capable to presumably penetrate the outer membrane of Gram-negative bacteria<sup>25,26</sup>.

**Table 1.** Minimum inhibitory concentration (MIC) values of sorangiadenosine and 2-hydroxysorangiadenosine (**1** and **2**) against common microbial pathogens.

Microorganism	MIC (µg/mL) of sorangiadenosine ( <b>1</b> )	MIC (µg/mL) of 2-hydroxysorangiadenosine ( <b>2</b> )
<i>B. subtilis</i> DSM 10	16	64
<i>E. coli</i> WT (DSM 1116)	>128	>128
<i>E. coli acrB</i> JW0451-2	>128	>128
<i>P. aeruginosa</i> PA14	>128	>128
<i>S. aureus</i> Newman	32	128
<i>C. freundii</i> DSM 30039	>128	>128
<i>A. baumannii</i> DSM 30007	>128	>128
<i>M. hiemalis</i> DSM 2656	128	>128
<i>P. anomala</i> DSM 6766	>128	>128
<i>M. smegmatis</i> MC <sup>2</sup> 155	>128	>128
<i>C. albicans</i> DSM 1665	128	>128

**Table 2.** Half-maximal inhibitory concentrations (IC<sub>50</sub> values in µg/mL) of sorangiadenosine, 2-hydroxysorangiadenosine (**1** and **2**) and doxorubicin (used as cytotoxic positive control) against HCT-116 (human colon carcinoma cell line, DSMZ No. ACC 581) and KB-3-1 (cervix carcinoma cell line, DSMZ No. ACC 158).

Cancer cell line	IC <sub>50</sub> (µg/mL) of sorangiadenosine <b>1</b>	IC <sub>50</sub> (µg/mL) of 2-hydroxysorangiadenosine <b>2</b>	IC <sub>50</sub> (µg/mL) of doxorubicin
HCT-116	30.00	52.00	0.6
KB-3-1	39.46	>111.1	0.09

## Identification of the 2-hydroxysorangiadenosine biosynthetic gene cluster

According to retrobiosynthetic considerations, **1** and **2** are made of the building blocks adenosine (available from primary metabolism) and the eudesmane-type sesquiterpene, which can be generated from farnesyl pyrophosphate (FFP). Supplementation with sodium acetate <sup>13</sup>C<sub>2</sub> – representing the starting precursor for the mevalonate pathway – did not result in a significant mass shift in the isotopic pattern of **1** and **2** (**Supporting Information Figure S1, S4**).

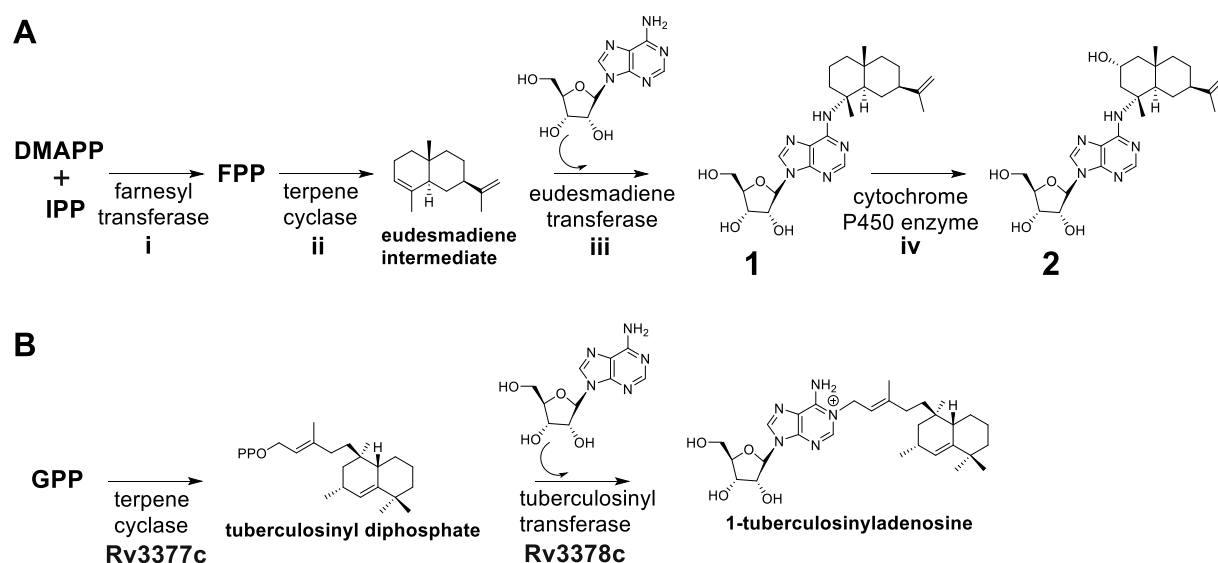
Therefore the previously characterized myxobacterial L-leucine degradation pathway (including the formation of isovaleryl-CoA, dimethylacrylyl-CoA and 3-methylglutaconyl-CoA)<sup>27</sup> that branches from the mevalonate-dependent isoprenoid biosynthesis pathway might be involved in the formation of **1** and **2**. The observed mass shifts of +5 Da, +10 Da and +14 Da in the isotopic pattern of **1** and **2** upon dimethyl acrylic acid d<sub>6</sub> supplementation suggest that the sesquiterpene scaffold originates from this myxobacterial isoprenoid pathway (**Supporting Information Figure S2, S5**).

In addition, genomic investigation of *V. cumulatum* MCy10943<sup>T</sup> revealed the presence of all genes encoding the proteins which are required for the leucine degradation pathway<sup>27</sup> and the alternative biosynthesis of isovaleryl-CoA<sup>28</sup> (**Supporting Information, Figure S14, Table S1**). Corresponding mass shifts of +4 Da and +5 Da upon incorporation of adenosine monophosphate (<sup>15</sup>N<sub>5</sub>) (**Supporting Information Figure S3, S6**), underlines that adenosine is transferred as a purine nucleoside to the eudesmadiene scaffold rather than being generated *de novo*. In conclusion, these feeding experiments, proved that **1** and **2** are most likely generated from the building blocks isopentenyl diphosphate (IPP) or dimethylallyl diphosphate (DMAPP) and adenosine.

Based on the elucidated chemical structure of **1** and **2** along with the findings from feeding experiments, we propose both to be biosynthesized by a farnesyl transferase (**i**) which produces FFP, a terpene cyclase (**ii**) which converts the generated FFP to an eudesmane-type sesquiterpene (eudesmadiene intermediate, see **SI**) and an eudesmadiene transferase (**iii**) to transfer the adenosine building block onto the eudesmane-type sesquiterpene (**Figure 4A**). The mechanism to transfer adenosine, which is present in most microorganisms, might work via hydroamination (direct addition of ammonia or primary and secondary amines to non-activated alkenes and alkynes) or alkylation<sup>29,30</sup>.

A last maturation step involves the hydroxylation of **1** leading to the herein described structure of **2**. Since most bacterial terpene biosynthetic pathways involve tailoring steps through cytochrome P450 enzyme hydroxylations<sup>31</sup>, we assume that the last tailoring step in the biosynthesis of **2** might be catalyzed as well by a cytochrome P450 enzyme (**iv**) (**Figure 4A**).

This assumption is also supported by the fact, that the secondary metabolome of MCy10943<sup>T</sup> shows no signal corresponding to the sum formula of a hydroxylated eudesmadiene intermediate or intermedeol. In addition, three secondary metabolites with the mass of 205 m/z and the sum formula C<sub>15</sub>H<sub>25</sub> (which can be linked to non-hydroxylated eudesmadiene intermediates) are present in the crude extract of MCy10943<sup>T</sup> (**Supporting Information, Figure S16**).



**Figure 4.** A) Proposed biosynthetic route to **1** and **2** based on retrobiosynthetic considerations. B) Biosynthesis of 1-tuberculosinyladenosine<sup>32</sup>.

In contrast to the first, second and the last biosynthetic step, the catalytic reaction of a specific eudesmadiene transferase is unprecedented. This biosynthetic enzyme would act different from classical cis-prenyl transferases by catalyzing the condensation of eudesmadiene pyrophosphate and adenosine to generate the previously discovered structure of **1**. A similar reaction has been described for the biosynthesis of 1-tuberculosinyladenosine, where the second biosynthetic enzyme Rv3378c works as a tuberculosinyl transferase, catalyzing the condensation of tuberculosinyl pyrophosphate and adenosine to generate 1-tuberculosinyladenosine<sup>32</sup> (**Figure 4B**).

For that reason, genome sequence analysis was focused on gene clusters harboring genes encoding terpene or diterpene cyclases to identify the biosynthetic locus accounting for the production of **1** and **2**. MCy10943<sup>T</sup> harbors according to the antibiotics and secondary metabolite analysis shell (antiSMASH)<sup>33</sup> nine putative terpene gene clusters (**Table 3**). *In silico* genome sequence analysis of several myxobacterial strains (including alternative producers of **1** and **2**, and strains with BGCs homologous to three terpene gene clusters found in MCy10943<sup>T</sup>) in combination with metabolomic profiling of corresponding crude extracts facilitated the identification of a candidate biosynthetic gene cluster (**Supporting Information 2.3**).

The candidate terpene gene cluster No. 1 harbors all genes necessary for the formation of **1** and **2** (**Table 3**). Terpene gene cluster No. 1 contains 22 open reading frames (*sora1*–*22*) and comprises 27,920 bp (**Figure 5A**, **Table 4**). The myxobacterial strains *Archangium violaceum* Cb vi76<sup>34</sup> and *Archangium* sp. Cb g35<sup>35</sup> feature similar biosynthetic gene clusters organization according to antiSMASH evaluation (however these gene clusters are missing *sora12*, see below). The gene *sora9* encodes a terpene cyclase which could produce the eudesmane-type sesquiterpene building block, while *sora8* and *sora12* are presumably catalyzing the transfer of the adenosine onto the terpene scaffold (**Figure 5B**). Both transferases share high mutual sequence similarity and probably exert the same catalytic function. Only the 2-hydroxysorangiadenosine gene cluster in *V. cumulatum* MCy10943<sup>T</sup> contains *sora12*, which might be the reason for the tremendously improved production in comparison to the identified alternative myxobacterial producers (**Supporting Information 1.3**).

**Table 3.** Biosynthetic gene clusters identified through antiSMASH analysis harboring a terpene cyclase.

No.	Gene cluster	Size (bp)	Location	Terpene cyclase	Associated Biosynthesis <sup>1</sup>
1	Terpene	27920	653776–674051	1 x Type I, 2 x Type II	This study
2	Terpene	42274	1269631–1311904	1 x Type I	Geosmin <sup>36,37</sup>
3	Terpene	41089	3415277–3456365	1 x Type I	Genome-Metab
4	Terpene	41071	3447909–3488979	1 x Type I	-----
5	Terpene	40978	3603511–3644488	1 x Type I	Genome-Metab
6	Terpene	41041	4852259–4893299	1 x Type I	Genome-Metab
7	Terpene/ THII_PKS	69038	7156603–7225640	1 x Type II	Carotenoid <sup>38–41</sup>
8	Terpene/ TfuA-rel.	50108	8404785–8454892	1 x Type I	-----
9	Terpene/ RiPP	78053	12603875–12681927	1 x Type I	Geosmin <sup>36,37</sup>

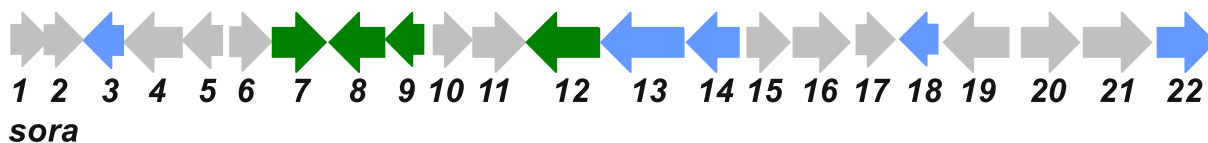
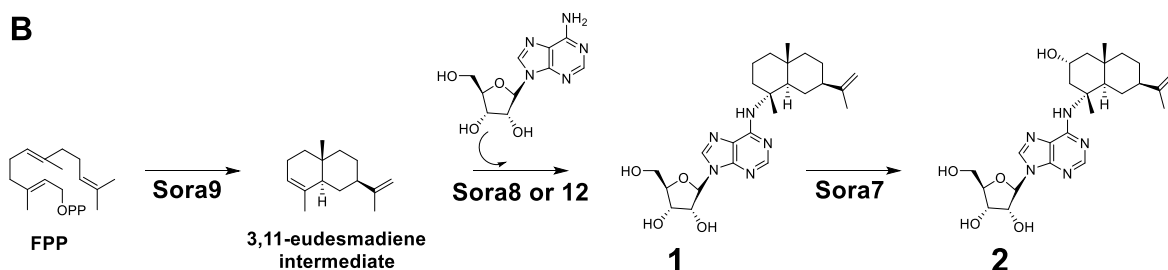
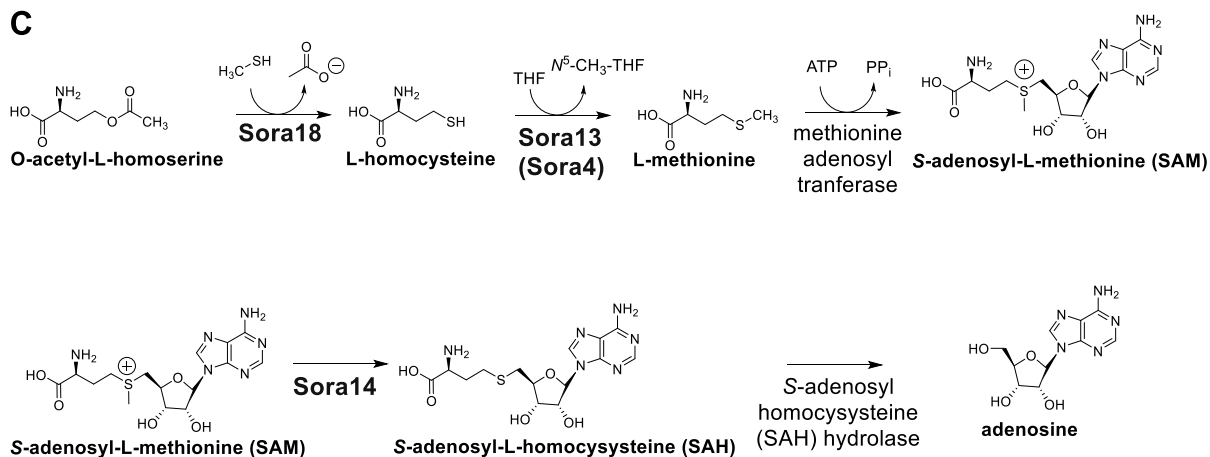
<sup>1</sup>Genome-Metab.; these gene clusters have been excluded to be responsible for the formation of **1** and **2**, due to genome-metabolome correlation (myxobacterial strains with homologues BGCs featured no detectable production of **1** or **2**, see **SI**).

The identified type I terpene cyclase Sora9 contains all typical sequence motifs such as the aspartate-rich motif DDxxxD, the pyrophosphate sensor (R), the NSE triad and the RY dimer<sup>42</sup>. In addition, terpene gene cluster No. 1 harbors several genes responsible for adenosine supply. In prokaryotes, adenosine can be formed either by salvage pathways or via *de novo* synthesis, which starts from simple primary metabolites and forms inositol monophosphate (IMP) as branch-point intermediate to form further guanosine monophosphate (GMP) and adenosine monophosphate (AMP), of which the later can be converted to adenosine via adenosine/deoxycytidine kinase or by the catalytic action of 5'-nucleotidase.

In addition, adenosine can also be produced in context of L-homocysteine metabolism, where *S*-adenosyl-homocysteine (SAH) is converted to adenosine via the SAH hydrolase. The gene for the conversion of *O*-acetyl-homoserine to L-homocysteine via an *O*-acetylhomoserine sulfhydrylase (*sora18*) is located within the terpene gene cluster No. 1. The intermediate L-homocysteine can then subsequently be further converted by a vitamin B<sub>12</sub> independent synthase (encoded by *sora4*) and a methionine synthase (encoded by *sora13*) to L-methionine.

An ATP-dependent methionine adenosyltransferase catalyze the production of *S*-adenosyl-L-methionine (SAM), which can be methylated by a SAM methyltransferase encoded by *sora14*. The catalytic product of Sora14, *S*-adenosyl-homocysteine (SAH) finally undergoes hydrolytic cleavage to yield adenosine which in turn can be used as a building block for **1** and **2** (**Figure 5C**). These proteins further contribute to the availability of adenosine. In addition, these findings further supports the assigned D-configuration of the ribose in **1** and **2** (see **2.1**), since these genes are inseparably connected with the D-configured ribose, in particular the SAM-dependent methyltransferase Sora14. The truncated 2-hydroxysorangiadenosine BGC in the alternative producer *Cystobacter* sp. strain MCy9101 (SBCb004) also harbors *sora13*, *sora14* and *sora18* (**Supporting Information Figure S18**). Therefore, it seems likely, that these encoded proteins are directly involved in adenosine supply for **1** and **2**. As shown for the biosynthesis of 1-tuberculosinyladenosine, the gene responsible for the transfer of the adenosine scaffold is closely located to the terpene cyclase gene<sup>32</sup>.

However, it cannot be excluded that the genes responsible for the generation of adenosine are located elsewhere in the genome. In conclusion, these findings strongly support our conclusion that the identified BGC is involved in the proposed biosynthetic route leading to the formation of **1** and **2**.

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

**Figure 5.** A) Genetic organization of the candidate biosynthetic gene cluster responsible for the biosynthesis of sorangiadenosine (**1**) and 2-hydroxysorangiadenosine (**2**). B) Proposed biosynthetic route leading to **1** and **2**. C) Catalyzed reactions to increase the supply of adenosine.

**Table 4.** Predicted functions of the proteins encoded by the (2-hydroxy)sorangiadenosine biosynthetic gene cluster.

Gene	Size (aa)	Proposed function	Closest Homologue	Coverage/ Similarity (%)
<i>sora1</i>	378	Oxidoreductase	WP_108075222	100/95.99
<i>sora2</i>	72	Hypothetical protein	WP_108075223	100/93.90
<i>sora3</i>	173	Hypothetical protein	WP_108075224	100/82.92
<i>sora4</i>	364	Methionine synthase (MetE)	WP_108075225	100/95.87
<i>sora5</i>	304	Hydrolase	WP_108075226	100/90.43
<i>sora6</i>	284	Hypothetical protein	WP_108075227	95/92.25
<i>sora7</i>	457	Cytochrome P450 enzyme	WP_108069092	98/95.67
<i>sora8</i>	518	Eudesmadiene transferase	WP_047856205	99/80.31
<i>sora9</i>	333	Terpene cyclase	WP_073564250	99.78/79.09
<i>sora10</i>	387	Hypothetical protein	WP_095976240	91/56.91
<i>sora11</i>	609	sensory transducer	WP_146210122	92/74.69
<i>sora12</i>	509	Eudesmadiene transferase	WP_108075230	100/73.80
<i>sora13</i>	1171	Methionine synthase (MetH)	WP_108075232	100/97.61
<i>sora14</i>	299	SAM-dependent methyltransferase	WP_108075233	100/97.32
<i>sora15</i>	292	Patatin lipid acyl hydrolases	WP_108075234	100/98.97
<i>sora16</i>	375	Dehydrogenase	WP_052519033	91/82.11
<i>sora17</i>	205	Tet <sup>R</sup> transcriptional regulator	WP_073564266	100/79.62
<i>sora18</i>	147	<i>O</i> -acetylhomoserine sulfhydrylase	WP_108075236	100/97.26
<i>sora19</i>	302	Hypothetical protein	WP_108075237	100/88.70
<i>sora20</i>	449	Thioredoxin	WP_140874099	99/54.81
<i>sora21</i>	279	Hypothetical protein	WP_158079939	100/77.34
<i>sora22</i>	630	Phosphotransferase	WP_073564278	100/79.83

## 4.4 Discussion

The weak antibacterial myxobacterial sesquiterpene-nucleoside **1** was isolated from *S. cellulosum* KM1003 in 2008<sup>17</sup>. Distinct structural features of **1** such as the heteroaromatic adenosine, the pentose sugar D-ribofuranose and the bicyclic sesquiterpene of eudesmane-type are of particular biosynthetic interest, since they have not been reported for other myxobacterial natural products and are rare among other producers. The biosynthetic pathway and gene cluster leading to the formation of **1** remained elusive.

Unlike naturally-occurring nucleotides and nucleoside natural products and terpenes, there are only few terpene-nucleoside/nucleotide hybrids known up-to date, in particular from microorganisms, such as 1-tuberculosinyladenosine produced by *Mycobacterium tuberculosis*<sup>32</sup> or 2-methylthio-*N*<sup>6</sup>-( $\Delta^2$ -isopentenyl)adenosine from *E. coli*<sup>43</sup>. The cytokines are described as plant-derived terpene-purine derivatives involved in cell growth and differentiation and act complementary with auxin as plant growth hormones<sup>44,45</sup>. The zeatins (adenine-type cytokinins like lupinic acid<sup>46</sup>) feature a hydroxylated isopentenyl building block on the N-6 position of adenosine. Furthermore, the agelasines<sup>47,48</sup>, agelasimines<sup>49</sup> and asmarines<sup>50</sup> are terpene-purines originating from different genera of sponges (*Agela* sp., *Raspailia* sp.) which feature a diterpene attachment on the N-6 (and N-7 in case of asmarines) position of adenine<sup>51</sup>.



In this study we report the isolation and structure elucidation of the new natural product **2** alongside with the known derivative **1** from the myxobacterium *V. cumulatum* MCy10943<sup>T</sup>. For the first time we provide biosynthetic insights leading to the production of these unique sesquiterpene-adenosine hybrids in myxobacteria. The presented discovery and *in silico* biosynthetic study paves the way for further genetic investigation of this intriguing pathway and the underlying production of these unique bioactive chemical scaffolds.

The development of genetic tools for the myxobacterium *V. cumulatum* MCy10943<sup>T</sup> – which can be a difficult endeavor for myxobacterial strains<sup>52</sup> – would allow experimental correlation between the identified BGC and the production of **1** and **2**. After genetic confirmation, the key biosynthetic genes could be heterologously expressed in the myxobacterial model host *M. xanthus* DK1622 in order to genetically engineer the biosynthetic pathway for an improved production rate or streamline the generation of new derivatives. In-depth biochemical investigation of the biosynthesis of **1** and **2** would be attained through recombinantly generated biosynthetic proteins to uncover single biosynthetic reactions or reconstitute the biosynthesis *in vitro*.

In conclusion, the presented discovery and the hypothesized biosynthetic pathway of **1** and **2** are essential prerequisites for further biosynthetic studies, production optimization and the generation of new sorangiadenosine derivatives.

## 4.5 Material and Methods

### Applied software, sequence analysis and bioinformatics methods

DNA was isolated using Qiagen Genomic-tip 100/G (Qiagen, Hilden Germany) according to the instructions of the manufacturer. A SMRTbell™ template library was prepared according to the instructions from Pacific Biosciences (Menlo Park, CA, USA), following the Procedure & Checklist – Greater Than 10 kb Template Preparation. Briefly, for preparation of 15 kb libraries, 8 µg genomic DNA was sheared using g-tubes™ from Covaris (Woburn, MA, USA), according manufacturer's instructions. DNA was end-repaired and ligated overnight to hairpin adapters applying components from the DNA/Polymerase Binding Kit P6 from Pacific Biosciences (Menlo Park, CA USA). Reactions were carried out according manufacturer's instructions. BluePippin™ Size-Selection to greater than 7 kb was performed according to the manufacturer's instructions (Sage Science, Beverly, MA, USA). Conditions for annealing of sequencing primers and binding of polymerase to purified SMRTbell™ template were assessed with the Calculator in RS Remote, Pacific Biosciences (Menlo Park, CA, USA). 2 SMRT cells were sequenced on the PacBio RSII (Pacific Biosciences, Menlo Park, CA, USA) taking 240-minutes movies.

The neutral sum formula of the identified two target masses (C<sub>25</sub>H<sub>37</sub>N<sub>5</sub>O<sub>4</sub> corresponding to **1** and C<sub>25</sub>H<sub>37</sub>N<sub>5</sub>O<sub>5</sub> corresponding to **2**) were used to search in several chemical structure databases (SciFinder, Dictionary of Natural Products, ChemSpider, PubChem, ChEMBL, ZINC and Super Natural II). The *Vitiosangium cumulatum* MCy10943<sup>T</sup> genome was screened for secondary metabolite BGCs using the antiSMASH 5.0<sup>33</sup> online tool and the software Geneious prime® (Biomatters Ltd., Auckland, New

Zealand, 2020.0.5)<sup>53</sup>. The nucleotide or amino acid sequence of interest was aligned with the basic local alignment search tool (BLAST) against our in-house genome database or the publicly available nucleotide database, in order to find homologous genes or proteins. The functional prediction of ORFs was performed by either using protein blast and/or blastx program and Pfam<sup>54</sup>. The nucleotide sequence of the 2-hydroxysorangiadenosine BGC originating from MCy10943<sup>T</sup> has been deposited in GenBank and is accessible under the accession number MT520811. Furthermore, the same nucleotide sequence will be implemented in the Minimum Information about a Biosynthetic Gene cluster (MIBiG) database. In addition, the nucleotide sequence of the other eight putative terpene gene clusters from MCy10943<sup>T</sup> have been deposited in GenBank under the following accession numbers:

- No2\_MCy10943\_terpene\_gene\_cluster (MT520812)
- No3\_MCy10943\_terpene\_gene\_cluster (MT520813)
- No4\_MCy10943\_terpene\_gene\_cluster (MT520814)
- No5\_MCy10943\_terpene\_gene\_cluster (MT520815)
- No6\_MCy10943\_terpene\_gene\_cluster (MT520816)
- No7\_MCy10943\_terpene\_typeIIIPKS\_gene\_cluster (MT520817)
- No8\_MCy10943\_terpene\_TfuA-related (MT520818)
- No9\_MCy10943\_terpene\_RiPP\_gene\_cluster (MT520819)

Further information concerning gene sequences can be found in the **Supporting Information**.

### **Maintenance of bacterial cultures and feeding experiments with stable isotope-labeled building blocks**

*Vitiosangium cumulatum* MCy10943<sup>T</sup> was routinely cultivated at 30 °C in CyHv3 medium or agar plates (0.2% soytone (BD), 0.3% casitone (BD), 0.2% glucose, 0.8% soluble starch (Roth), 0.15% yeast extract (BD), 0.1% CaCl<sub>2</sub> x 2H<sub>2</sub>O, 0.1% MgSO<sub>4</sub> x 7H<sub>2</sub>O, 50 mM HEPES, (for agar plate cultures 15 g/L agar,) adjusted pH to 7.2 with 10N KOH before autoclaving, added after autoclaving 8 mg/L Fe-EDTA). Liquid cultures were grown in Erlenmeyer flasks on an orbital shaker at 180 rpm for 3–6 days. Feeding experiments were performed by cultivating the strain in 20 mL CyHv3 broth using 100 µL inoculum. The cultures were supplemented with 1 mL (v/v) sterile amberlite resin XAD-16 (Sigma-Aldrich Chemie GmbH, Taufkirchen, Germany) and fed for five consecutive days with 20 µL of a 0.1 M solution of either sodium acetate (<sup>13</sup>C<sub>2</sub>), sodium dimethyl acrylic acid (d<sub>6</sub>) or adenosine monophosphate (<sup>15</sup>N<sub>5</sub>) at 30 °C, at 180 rpm. The combined cells and resin were harvested by centrifugation after seven days of incubation before extraction. The supernatant was discarded, whereas the combined cells and resin were extracted with a mixture of 25 mL methanol (MeOH) and 25 mL acetone, stirred for 2 h, filtered through filter paper and the solvent of the extracts were removed under vacuum. The re-dissolved extracts (2 mL) were diluted with MeOH (1:3 (extract/MeOH, v:v) centrifuged, and the supernatant was subjected to HPLC-MS analysis as described further below.

## Analysis of secondary metabolism of broth extracts

The secondary metabolism of broth extracts was analyzed by HPLC-HRESI-DAD-MS on a maXis 4G mass spectrometer (Bruker Daltonics, Billerica, MA, USA) coupled with a Dionex Ultimate 3000 RSLC system (Thermo Fisher Scientific, Waltham, MA, USA) using a BEH C18 column (100 × 2.1 mm, 1.7 μm) (Waters, Eschborn, Germany) with a gradient of 5–95% acetonitrile (ACN) + 0.1% formic acid (FA) in H<sub>2</sub>O + 0.1% FA at 0.6 mL/min and 45 °C over 18 min with UV detection by a diode array detector at 200–600 nm. Mass spectra were acquired from 150 to 2000 m/z at 2 Hz. The detection was performed in the positive MS mode. The plugin for Chromeleon Xpress (Thermo Fisher Scientific, Waltham, MA, USA, version 6.8) was used for operation of Dionex UltiMate 3000 RSLC system. HyStar (Bruker Daltonics, Billerica, MA, USA, version 3.2) was used to operate on maXis 4G mass spectrometer system. HPLC-MS mass spectra were analyzed with DataAnalysis (Bruker Daltonics, Billerica, MA, USA, version 4.2).

## Compound isolation

At the end of fermentation (60 L), wet cell mass and adsorbent resin XAD-16 were harvested by centrifugation and extracted three times with 2 L of acetone. The extract was dried under vacuum resulting in 7.68 g of crude extract. The crude extract was dissolved in 2 L MeOH and partitioned using 2 L hexane solvent. The MeOH layer was dried (5.34 g) and dissolved in 2 L deionized water followed by further partitioning using chloroform (CHCl<sub>3</sub>). The CHCl<sub>3</sub> layer was dried under vacuum to yield 3.95 g of extract. 3.95 g of the CHCl<sub>3</sub> extract was subjected to flash chromatography on a Isolera™ One (Biotage) with a SNAP 100 g column packed with silica gel (60 Å, 70–230 mesh, 63–200 μm), using 100% *n*-hexane (5 column volume (CV)), 95% *n*-hexane/ethyl acetate (EA) (5 CV), 95% *n*-hexane/EA to 100% EA (25 CV), 100% EA (5 CV), 95% EA/MeOH gradient as eluents. The flow rate was 50 mL/min while the UV/Vis absorption was set at 260 nm and 320 nm. Fractions of 45 mL were collected and measured on a Dionex Ultimate 3000 RSLC system coupled to the amaZon iontrap MS using a BEH C18, 100 × 2.1 mm, 1.7 μm dp column equipped with a C18 precolumn (Waters) to detect the fractions containing **1** or **2**. Fractions 83–95 were pooled and dried under vacuum to yield 1.23 g extract. Purification of sample was done using an UltiMate™ 3000 semi preparative system coupled to a Thermo Scientific Dionex UltiMate 3000 Series Automated fraction collector. Separation was performed on a C18 Phenomenex Luna (100 Å, 5 μm, 10 × 250 mm) LC column and eluted with water (0.1% FA) as Solvent **A** and ACN (0.1% FA) as solvent **B** at a flow rate of 5 mL/min. The initial gradient was held at 5% ACN for 2 min and then elevated to 60% ACN within 5 min. This was followed by an increase from 61 to 70% ACN during a period of 25 min and once more raised to 95% ACN within 5 min. The gradient was held at 95% ACN for 2 min and then ramped back to 5% ACN during 1 min. The column was re-equilibrated at 5% ACN for 5 min. Detection of the sesquiterpene-nucleosides was facilitated via mass spectrometry on the Agilent 1100 series coupled to the HCT 3D ion trap or with a UV detector on the Dionex 3000 SL systems by UV absorption at 220 nm, 260 nm, 320 nm and 400 nm.

The pure compounds were subsequently dried by lyophilization and resulted in 4.91 mg of compound **2**, 3.88 mg of compound **1** at a retention time of 18.23 min and 22.14 min respectively

### NMR-based structure elucidation, chiroptical and CD measurement.

The chemical structures of **1** and **2** were determined via multidimensional NMR analysis. <sup>1</sup>H NMR, <sup>13</sup>C NMR and 2D spectra were recorded at 500 MHz (<sup>1</sup>H)/175 MHz (<sup>13</sup>C) conducting an Ascend 500 spectrometer using a cryogenically cooled triple resonance probe (Bruker Biospin, Rheinstetten, Germany). Samples were dissolved in methanol-d<sub>4</sub>. Chemical shifts are reported in ppm relative to tetramethylsilane; the solvent was used as the internal standard (**Supporting Information**).

Chiroptical measurements of **1** and **2** in MeOH ([α]<sub>D</sub>) were obtained on a Model 341 polarimeter (PerkinElmer Inc., Waltham, MA, USA) in a 100 x 2 mm cell at 20 °C. Circular dichroism measurements were performed for **2** at 1.0 mM in MeOH (400–190 nm) with the J-1500 CD spectrophotometer (JASCO, Easton, MD, USA).

### Bioactivity profiling

Standard sterile microbiological techniques were maintained throughout. All microorganisms and cell lines were handled according to standard procedures and were obtained from the German Collection of Microorganisms and Cell Cultures (Deutsche Sammlung für Mikroorganismen und Zellkulturen, DSMZ) or were part of our internal strain collection and were cultured under conditions recommended by the depositor. Both **1** and **2** were tested in microbroth dilution assays on the following panel of microorganisms: *Escherichia coli* WT (DSM 1116), *E. coli* JW0451-2 (*acrB*-efflux pump deletion mutant of *E. coli* BW25113), *Pseudomonas aeruginosa* PA14, *Bacillus subtilis* DSM 10, *Mycobacterium smegmatis* MC<sup>2</sup>-155 (DSM 43756), *Staphylococcus aureus* Newman, *Candida albicans* DSM 1665, *Citrobacter freundii* DSM 30039, *Pichia anomala* DSM 6766 and *Acinetobacter baumannii* DSM 30007.

For microbroth dilution assays, overnight cultures were prepared from cryogenic-preserved cultures and were diluted to achieve a final inoculum of 10<sup>4</sup>–10<sup>5</sup> cfu/mL. Serial dilutions of compounds were prepared in sterile 96-well plates in the respective test medium. The cell suspension was added and microorganism were grown for 18–48 h at either 30 °C or 37 °C. Growth inhibition was evaluated by visual inspection and given MIC values are the lowest concentration of antibiotic at which no visible growth was observed. To evaluate cytotoxic capabilities of **1** and **2**, HCT-116 (human colon carcinoma cell line, DSMZ No. ACC 581) and KB-3-1 (cervix carcinoma cell line, DSMZ No. ACC 158) cells were seeded at 6 x 10<sup>3</sup> cells per well of 96-well plates in 180 µL complete medium and treated with **1** or **2** in serial dilution after 2 h equilibration. After five days of incubation, 20 µL of 5 mg/mL MTT (thiazolyl blue tetrazolium bromide) in phosphate buffered saline (PBS) was added per well and it was further incubated for 2 h at 37 °C. The medium was discarded and cells were washed with 100 µL PBS before adding 100 µL isopropanol/10 N HCl (250:1) in order to dissolve formazan granules. The absorbance at 570 nm was measured using the microplate reader Infinite® M200Pro (Tecan Group Ltd., Männedorf, Switzerland), and cell viability was expressed as percentage relative to the respective MeOH control. IC<sub>50</sub> values were determined by sigmoidal curve fitting.

## 4.6 Acknowledgements

The authors thank Stefanie Schmidt, Irene Kochems, and Victoria Schmitt for performing bioactivity assays, Bettina Ehlert and Carola Berg for excellent technical assistance with sequencing, and Nestor Zaburannyi and Boyke Bunk for bioinformatic support. Research in Rolf Müller's laboratory is funded by the Deutsche Forschungsgemeinschaft (DFG), the Bundesministerium für Bildung und Forschung (BMBF) and the Deutsches Zentrum für Infektionsforschung Standort Hannover-Braunschweig.

## 4.7 Conflict of Interest

The authors declare no conflict of interest.

## 4.8 Key words

Myxobacteria, terpene-nucleoside, biosynthesis, antibiotics, sorangiadenosine, 2-hydroxysorangiadenosine, secondary metabolites, genome-mining, antibiotics, natural products discovery

## 4.9 References

1. Munoz-Dorado, J., Marcos-Torres, F. J., Garcia-Bravo, E., Moraleda-Munoz, A. & Perez, J. Myxobacteria: Moving, Killing, Feeding, and Surviving Together. *Front. Microbiol.* **7**, 781 (2016).
2. Cao, P., Dey, A., Vassallo, C. N. & Wall, D. How Myxobacteria Cooperate. *J. Mol. Biol.* (2015).
3. Herrmann, J., Fayad, A. A. & Müller, R. Natural products from myxobacteria: novel metabolites and bioactivities. *Nat. Prod. Rep.* **34**, 135–160 (2017).
4. Gerth, K., Bedorf, N., Irschik, H., Höfle, G. & Reichenbach, H. The soraphens: a family of novel antifungal compounds from *Sorangium cellulosum* (Myxobacteria). I. Soraphen A<sub>1α</sub>: fermentation, isolation, biological properties. *J. Antibiot.* **47**, 23–31 (1994).
5. Baumann, S. *et al.* Cystobactamids: myxobacterial topoisomerase inhibitors exhibiting potent antibacterial activity. *Angew. Chem. Int. Ed.* **53**, 14605–14609 (2014).
6. Gerth, K., Steinmetz, H., Höfle, G. & Jansen, R. Chlorotonil A, a macrolide with a unique gem-dichloro-1,3-dione functionality from *Sorangium cellulosum*, So ce1525. *Angew. Chem. Int. Ed. Engl.* **47**, 600–602 (2008).
7. Plaza, A. *et al.* Aetheramides A and B, potent HIV-inhibitory depsipeptides from a myxobacterium of the new genus "Aetherobacter". *Org. Lett.* **14**, 2854–2857 (2012).

8. Sasse, F., Steinmetz, H., Heil, J., Höfle, G. & Reichenbach, H. Tubulysins, new cytostatic peptides from myxobacteria acting on microtubuli. Production, isolation, physico-chemical and biological properties. *J. Antibiot.* **53**, 879–885 (2000).
9. Ullrich, A. *et al.* Pretubulysin, a potent and chemically accessible tubulysin precursor from *Angiococcus disciformis*. *Angew. Chem. Int. Ed. Engl.* **48**, 4422–4425 (2009).
10. Dickschat, J. S., Wenzel, S. C., Bode, H. B., Müller, R. & Schulz, S. Biosynthesis of volatiles by the myxobacterium *Myxococcus xanthus*. *ChemBioChem* **5**, 778–787 (2004).
11. Dickschat, J. S., Bode, H. B., Wenzel, S. C., Müller, R. & Schulz, S. Biosynthesis and identification of volatiles released by the myxobacterium *Stigmatella aurantiaca*. *ChemBioChem* **6**, 2023–2033 (2005).
12. Dickschat, J. S., Bode, H. B., Mahmud, T., Müller, R. & Schulz, S. A novel type of geosmin biosynthesis in myxobacteria. *J. Org. Chem.* **70**, 5174–5182 (2005).
13. Brock, N. L., Ravella, S. R., Schulz, S. & Dickschat, J. S. A Detailed View of 2-Methylisoborneol Biosynthesis. *Angew. Chem. Int. Ed. Engl.* **52**, 2100–2104 (2013).
14. Dickschat, J. S. *et al.* Biosynthesis of the off-flavor 2-methylisoborneol by the myxobacterium *Nannocystis exedens*. *Angew. Chem. Int. Ed. Engl.* **46**, 8287–8290 (2007).
15. Schiffrin, A. *et al.* Characterization of the Gene Cluster CYP264B1-*geoA* from *Sorangium cellulosum* So ce56: Biosynthesis of (+)-Eremophilene and Its Hydroxylation. *ChemBioChem* **16**, 337–344 (2015).
16. Schiffrin, A. *et al.* A single terpene synthase is responsible for a wide variety of sesquiterpenes in *Sorangium cellulosum* So ce56. *Org. Biomol. Chem.* **14**, 3385–3393 (2016).
17. Ahn, J. W., Jang, K. H., Chung, S. C., Oh, K. B. & Shin, J. Sorangiadenosine, a new sesquiterpene adenoside from the myxobacterium *Sorangium cellulosum*. *Org. Lett.* **10**, 1167–1169 (2008).
18. Awal, R. P. *et al.* *Vitiosangium cumulatum* gen. nov., sp. nov. and *Vitiosangium subalbum* sp. nov., soil myxobacteria, and emended descriptions of the genera *Archangium* and *Angiococcus*, and of the family *Cystobacteraceae*. *Int. J. Syst. Evol. Microbiol.* **67**, 1422–1430 (2017).
19. Nguyen Thi Mai, Nguyen Thi Cuc, Tran Hong Quang. Chemical constituents of *Datura metel* L. *Viet. J. Chem.* **55**, 188–195 (2017).
20. Ciuffreda, P., Casati, S. & Manzocchi, A. Complete <sup>1</sup>H and <sup>13</sup>C NMR spectral assignment of α- and β-adenosine, 2'-deoxyadenosine and their acetate derivatives. *Magn. Reson. Chem.* **45**, 781–784 (2007).
21. Nakamukai, S., Ise, Y., Ohtsuka, S., Okada, S. & Matsunaga, S. Isolation and identification of N<sup>6</sup>-isopentenyladenosine as the cytotoxic constituent of a marine sponge *Oceanapia* sp. *Biosci. Biotech. Biochem.* **83**, 1985–1988 (2019).
22. Brunner, W. C. & Maestre, M. F. Circular dichroism of some mononucleosides. *Biopolymers* **14**, 555–565 (1975).

23. Moore, D. S. Circular dichroism of nucleic acid monomers. I. Calculated adenosine and 2'-deoxyadenosine CD spectra. *Biopolymers* **19**, 1017–1038 (1980).
24. Miles, D. W., Farmer, M. & Eyring, H. Calculations of the circular dichroism of adenosine derivatives constrained in the *syn* form. *Proc. Natl. Acad. Sci. U.S.A.* **77**, 3398–3402 (1980).
25. Lewis, K. Platforms for antibiotic discovery. *Nat. Rev. Drug Discov.* **12**, 371–387 (2013).
26. Rojas, E. R. *et al.* The outer membrane is an essential load-bearing element in Gram-negative bacteria. *Nature* **559**, 617–621 (2018).
27. Bode, H. B. *et al.* Identification of additional players in the alternative biosynthesis pathway to isovaleryl-CoA in the myxobacterium *Myxococcus xanthus*. *ChemBioChem* **10**, 128–140 (2009).
28. Li, Y., Luxenburger, E. & Müller, R. An alternative isovaleryl CoA biosynthetic pathway involving a previously unknown 3-methylglutaconyl CoA decarboxylase. *Angew. Chem. Int. Ed. Engl.* **52**, 1304–1308 (2012).
29. Li, R. *et al.* Computational redesign of enzymes for regio- and enantioselective hydroamination. *Nat. Chem. Biol.* **14**, 664–670 (2018).
30. Severin, R. & Doye, S. The catalytic hydroamination of alkynes. *Chem. Soc. Rev.* **36**, 1407–1420 (2007).
31. Helfrich, E. J. N., Lin, G.-M., Voigt, C. A. & Clardy, J. Bacterial terpene biosynthesis: challenges and opportunities for pathway engineering. *Beilstein J. Org. Chem.* **15**, 2889–2906 (2019).
32. Layre, E. *et al.* Molecular profiling of *Mycobacterium tuberculosis* identifies tuberculosinyl nucleoside products of the virulence-associated enzyme Rv3378c. *Proc. Natl. Acad. Sci. U.S.A.* **111**, 2978–2983 (2014).
33. Blin, K. *et al.* antiSMASH 5.0: updates to the secondary metabolite genome mining pipeline. *Nucleic Acids Res.*, W81–W87 (2019).
34. Stevens, D. C., Young, J., Carmichael, R., Tan, J. & Taylor, R. E. Draft Genome Sequence of Gephyronic Acid Producer *Cystobacter violaceus* Strain Cb vi76. *Genome announc.* **2** (2014).
35. Adaikpoh, B. I., Dowd, S. E. & Stevens, D. C. Draft Genome Sequence of *Archangium* sp. Strain Cb G35. *Genome announc.* **5** (2017).
36. Giglio, S., Jiang, J., Saint, C. P., Cane, D. E. & Monis, P. T. Isolation and characterization of the gene associated with geosmin production in cyanobacteria. *Environ. Sci. Technol.* **42**, 8027–8032 (2008).
37. Jiang, J. Y., He, X. F. & Cane, D. E. Biosynthesis of the earthy odorant geosmin by a bifunctional *Streptomyces coelicolor* enzyme. *Nat. Chem. Biol.* **3**, 711–715 (2007).
38. Botella, J. A., Murillo, F. J. & Ruiz-Vazquez, R. A cluster of structural and regulatory genes for light-induced carotenogenesis in *Myxococcus xanthus*. *Eur. J. Biochem.* **233**, 238–248 (1995).

39. Lopez-Rubio, J. J., Elias-Arnanz, M., Padmanabhan, S. & Murillo, F. J. A repressor-antirepressor pair links two loci controlling light-induced carotenogenesis in *Myxococcus xanthus*. *J. Biol. Chem.* **277**, 7262–7270 (2002).
40. Perez-Marin, M. C., Padmanabhan, S., Polanco, M. C., Murillo, F. J. & Elias-Arnanz, M. Vitamin B<sub>12</sub> partners the CarH repressor to downregulate a photoinducible promoter in *Myxococcus xanthus*. *Mol. Microbiol.* **67**, 804–819 (2008).
41. Cervantes, M. & Murillo, F. J. Role for Vitamin B<sub>12</sub> in Light Induction of Gene Expression in the Bacterium *Myxococcus xanthus*. *J. Bacteriol.* **184**, 2215–2224 (2002).
42. Dickschat, J. S. Bacterial terpene cyclases. *Nat. Prod. Rep.* **33**, 87–110 (2015).
43. Harada, F. *et al.* 2-Methylthio N<sup>6</sup>-( $\Delta^2$ -isopentenyl) adenosine: A component of *E. coli* tyrosine transfer RNA. *Biochem. Biophys. Res. Commun.* **33**, 299–306 (1968).
44. Großkinsky, D. K. & Petrášek, J. Auxins and cytokinins – the dynamic duo of growth-regulating phytohormones heading for new shores. *New Phytol.* **221**, 1187–1190 (2019).
45. Schaller, G. E., Bishopp, A. & Kieber, J. J. The yin-yang of hormones: cytokinin and auxin interactions in plant development. *Plant Cell* **27**, 44–63 (2015).
46. MacLeod, J. K., Summons, R. E., Parker, C. W. & Letham, D. S. Lupinic acid, a purinyl amino acid and a novel metabolite of zeatin. *J. Chem. Soc., Chem. Commun.*, 809 (1975).
47. Abdjul, D. B. *et al.* Structures and Biological Evaluations of Agelasines Isolated from the Okinawan Marine Sponge *Agelas nakamurai*. *J. Nat. Prod.* **78**, 1428–1433 (2015).
48. Cullen, E. & Devlin, J. P. Agelasine: A Novel Quaternary 9-Methyladenine from the Sponge *Agelas dispar*. *Can. J. Chem.* **53**, 1690–1691 (1975).
49. Fathi-Afshar, R. & Allen, T. M. Biologically active metabolites from *Agelas mauritiana*. *Can. J. Chem.* **66**, 45–50 (1988).
50. Yosief, T., Rudi, A. & Kashman, Y. Asmarines A-F, novel cytotoxic compounds from the marine sponge *Raspailia species*. *J. Nat. Prod.* **63**, 299–304 (2000).
51. Gordaliza, M. Terpenyl-purines from the sea. *Mar. Drugs* **7**, 833–849 (2009).
52. Hug, J. J. & Müller, R. Host Development for Heterologous Expression and Biosynthetic Studies of Myxobacterial Natural Products. *Comprehensive Natural Products III: Chemistry and Biology, Chapter 14818* **In press** (2020).
53. Kearse, M. *et al.* Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* **28**, 1647–1649 (2012).
54. Finn, R. D. *et al.* The Pfam protein families database: towards a more sustainable future. *Nucleic Acids Res.* **44**, D279–285 (2016).



# Chapter 4

## Supporting Information

### **2-Hydroxysorangiadenosine: Structure and Biosynthesis of a Myxobacterial Sesquiterpene–Nucleoside**

Dorothy A. Okoth<sup>1,2</sup>, **Joachim J. Hug**<sup>1,2</sup>, Ronald Garcia<sup>1,2</sup>, Cathrin Spröer<sup>3</sup>,  
Jörg Overmann<sup>2,3</sup> and Rolf Müller<sup>1,2\*</sup>

1 Department Microbial Natural Products, Helmholtz-Institute for Pharmaceutical Research Saarland (HIPS), Helmholtz Centre for Infection Research (HZI) and Department of Pharmacy, Saarland University, Campus E8 1, 66123 Saarbrücken, Germany

2 German Center for Infection Research (DZIF), Partner Site Hannover-Braunschweig, Germany

3 Leibniz Institute DSMZ-German Collection of Microorganisms and Cell Cultures, Braunschweig, Germany

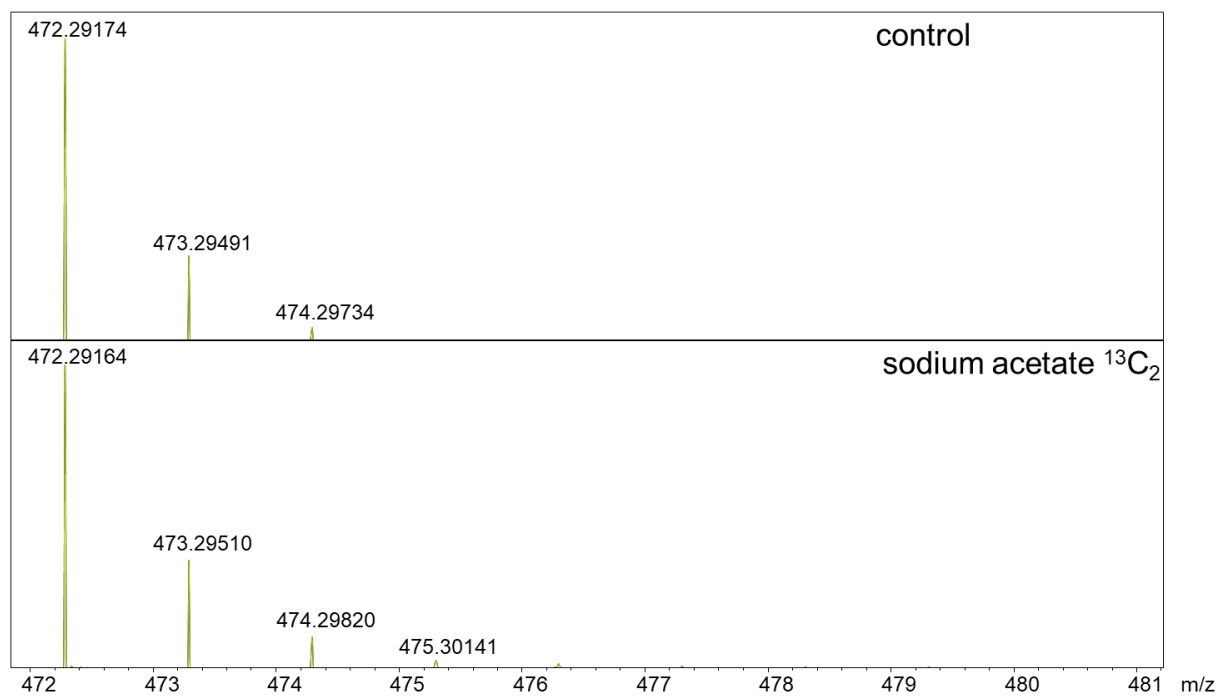
\* Correspondence: rolf.mueller@helmholtz-hips.de; Tel.: +49-681-98806-3000

### Table of Contents

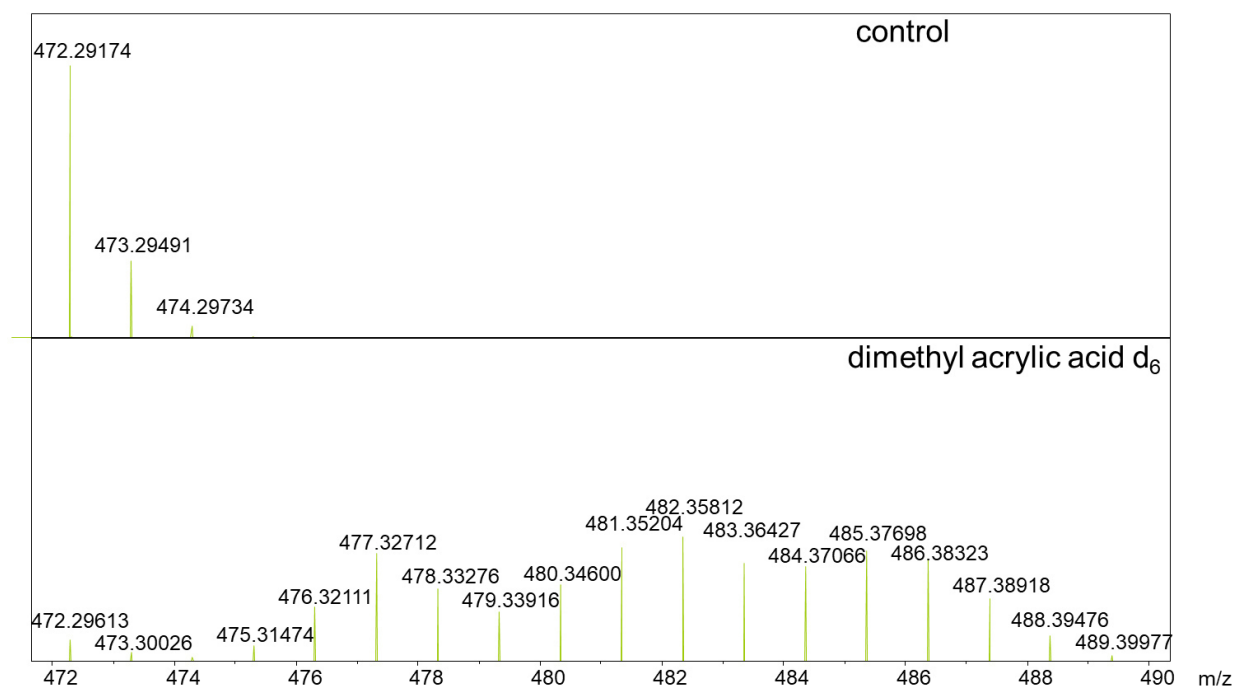
1. Partial LC-MS spectra.....	237
1.1 Feeding experiments .....	237
1.2 Preparative LC-MS spectra.....	241
1.3 Alternative producer .....	242
1.4 CD spectrum .....	246
2. Biosynthetic <i>in silico</i> investigation.....	247
2.1 Leucine degradation (and alternative isovaleryl CoA biosynthesis).....	247
2.2 Formation of sesquiterpene scaffolds and eudesmane-type sesquiterpenes.....	251
2.3 Metabolome-genome correlation .....	254
3. Structure elucidation .....	256
3.1 NMR spectroscopic data .....	256
3.2 <sup>1</sup> H and <sup>13</sup> C NMR spectra of sorangiaadenosine and 2-hydroxysorangiaadenosine.....	259
4. References.....	266

## 1. Partial LC-MS spectra

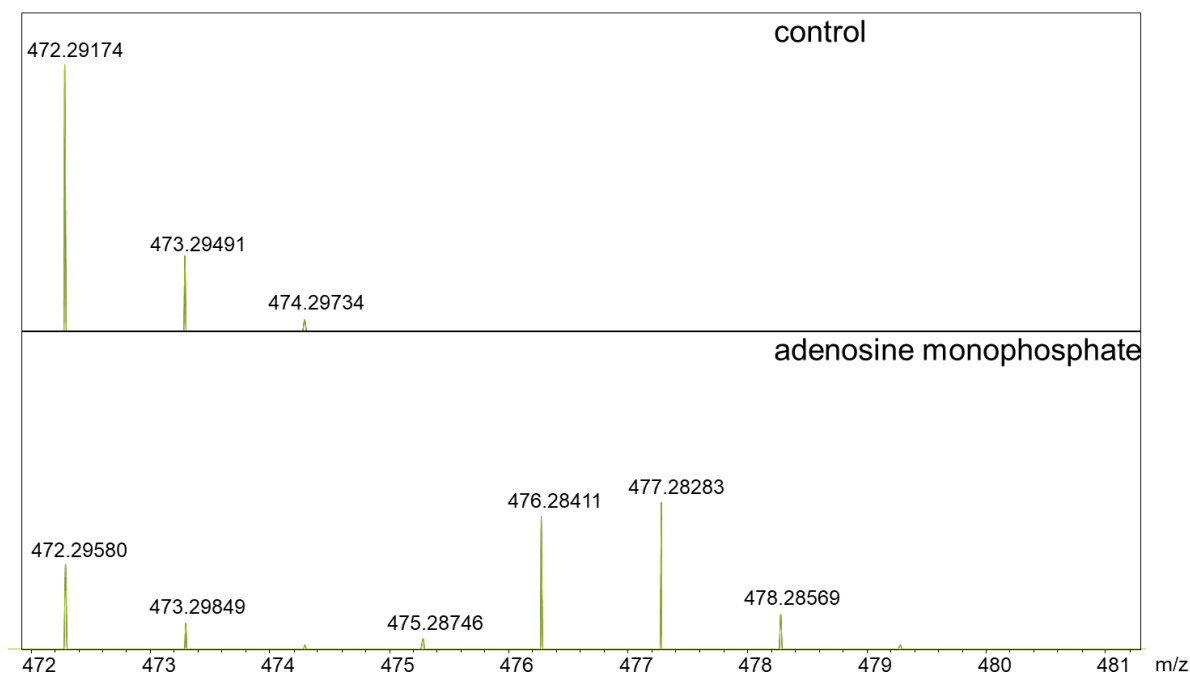
### 1.1 Feeding experiments



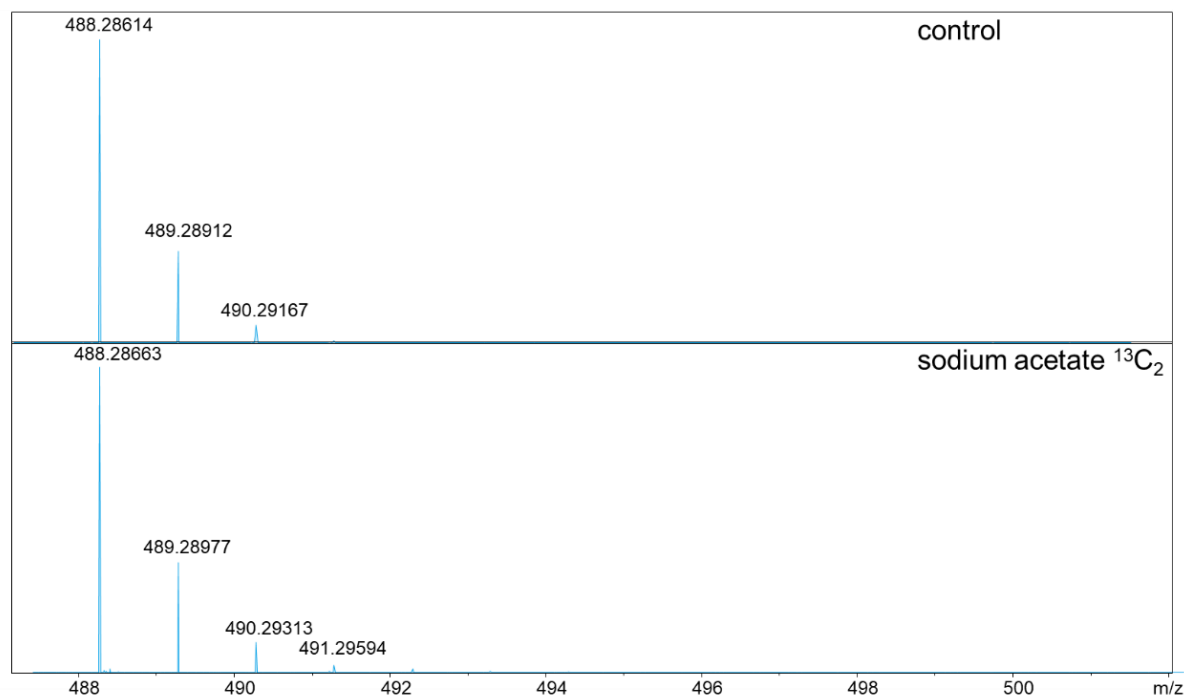
**Figure S1.** Partial ESI+MS spectra for sorangiaenosine (**1**) supplemented with sodium acetate (<sup>13</sup>C<sub>2</sub>) (bottom) and culture broth without precursor feeding as control (top).



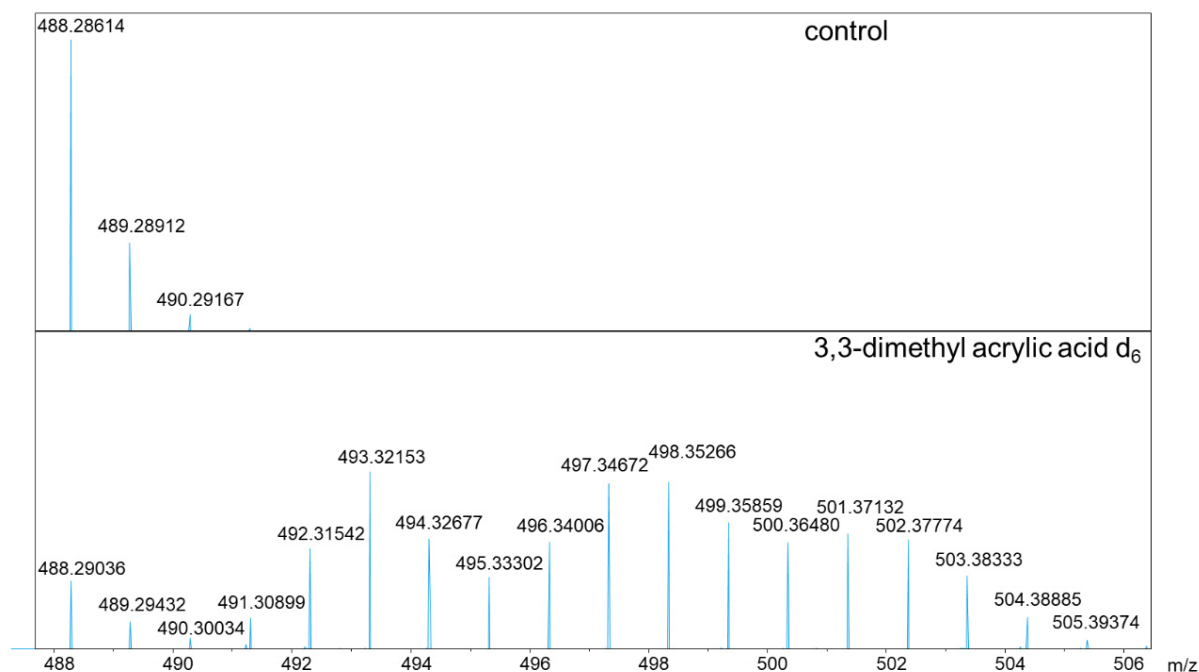
**Figure S2.** Partial ESI+MS spectra for sorangiaenosine (**1**) supplemented with sodium dimethyl acrylic acid (d<sub>6</sub>) (bottom) and culture broth without precursor feeding as control (top).



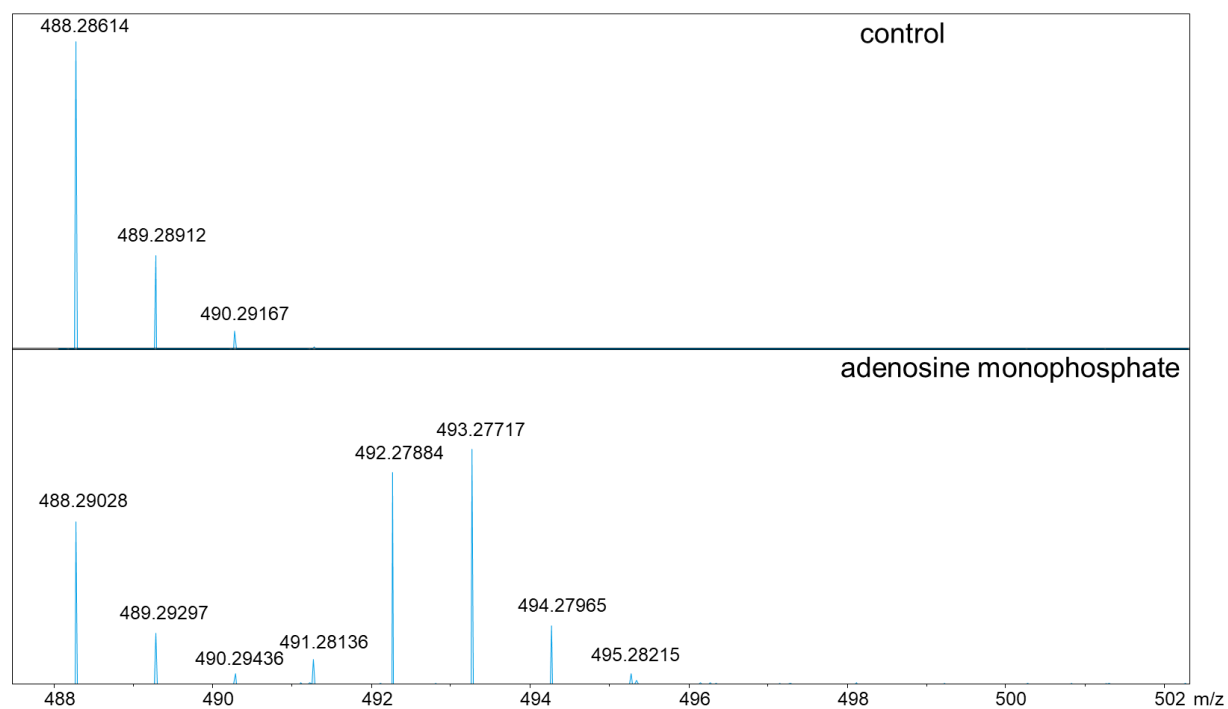
**Figure S3.** Partial ESI+MS spectra for sorangiaenosine (**1**) supplemented with adenosine monophosphate (<sup>15</sup>N<sub>5</sub>) (bottom) and culture broth without precursor feeding as control (top).



**Figure S4.** Partial ESI+MS spectra for 2-hydroxysorangiadenosine (**2**) supplemented with sodium acetate ( $^{13}\text{C}_2$ ) (bottom) and culture broth without precursor feeding as control (top).

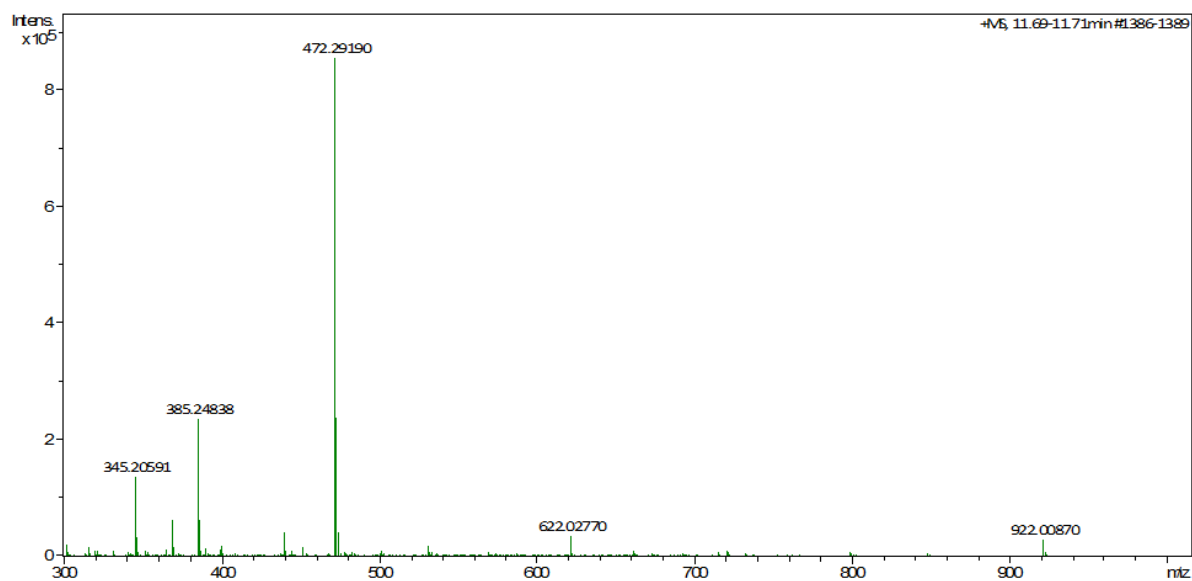


**Figure S5.** Partial ESI+MS spectra for 2-hydroxysorangiadenosine (**2**) supplemented with sodium dimethyl acrylic acid ( $\text{d}_6$ ) (bottom) and culture broth without precursor feeding as control (top).

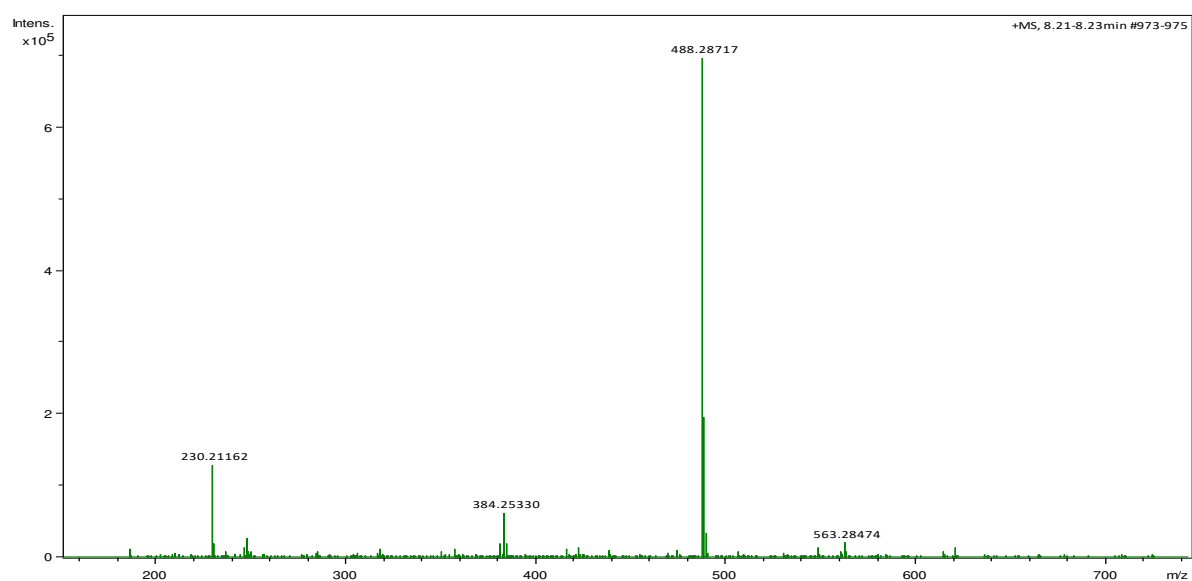


**Figure S6.** Partial ESI+MS spectra for 2-hydroxysorangiadenosine (**2**) supplemented with adenosine monophosphate ( $^{15}\text{N}_5$ ) (bottom) and culture broth without precursor feeding as control (top).

## 1.2 Preparative LC-MS spectra

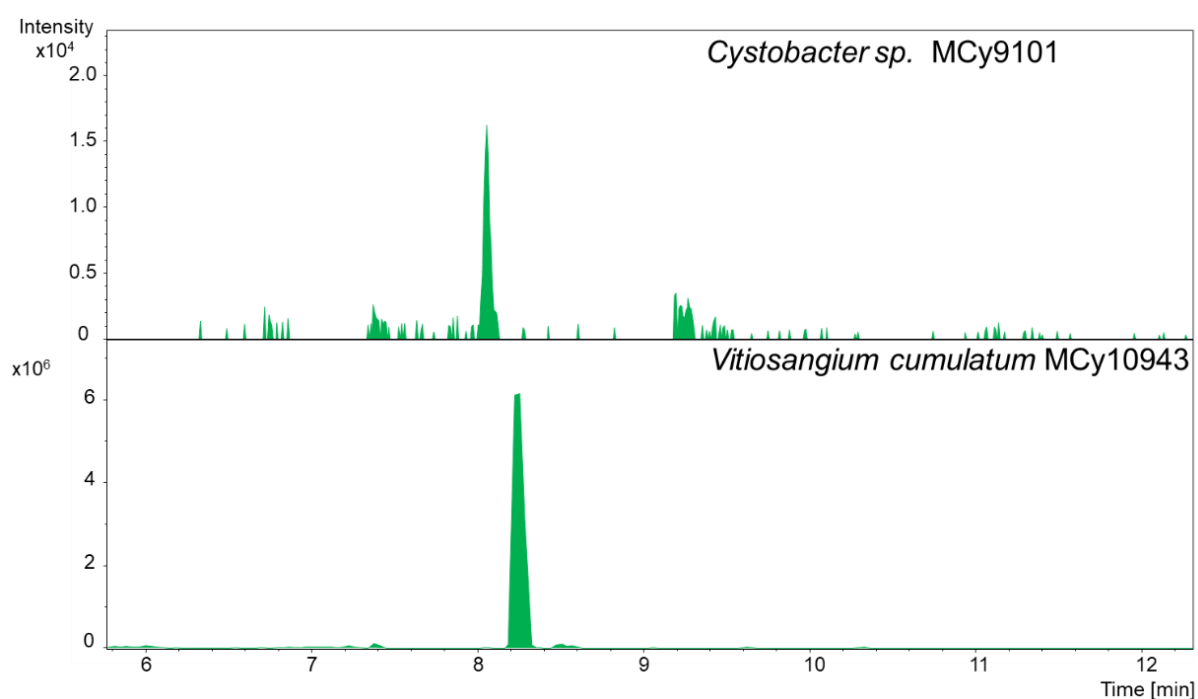


**Figure S7.** Partial ESI+MS spectra for sorangiadenosine (1).



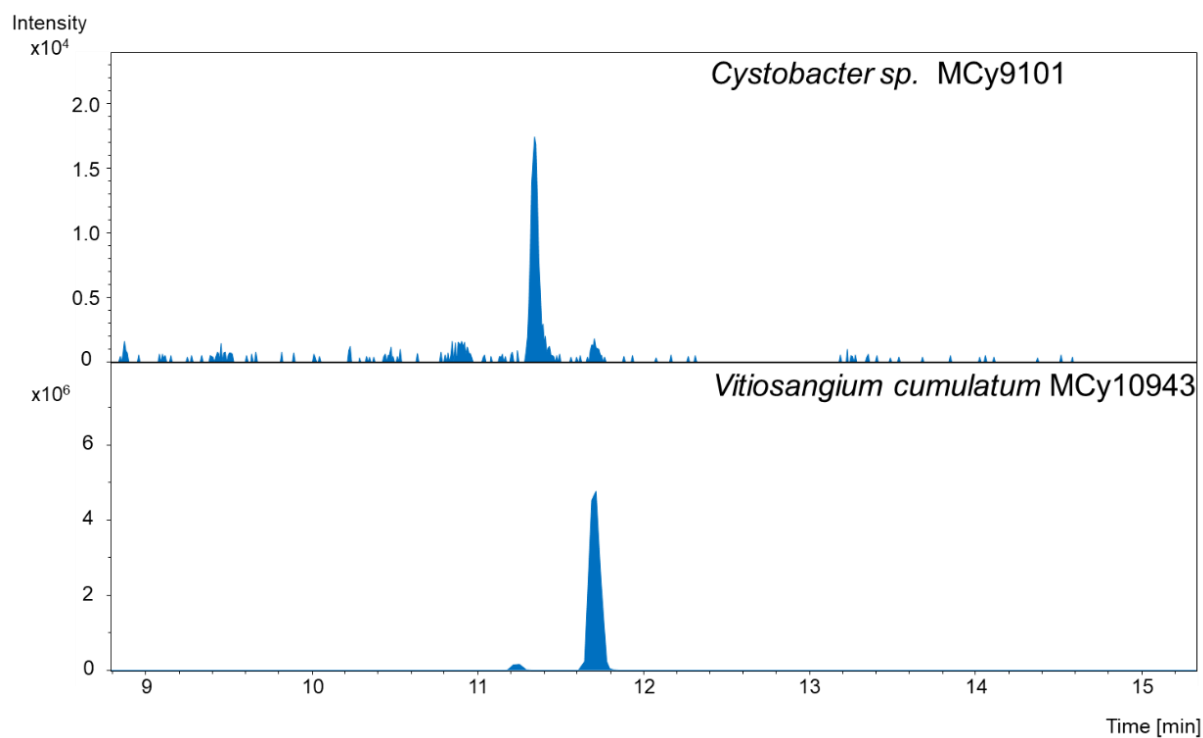
**Figure S8.** Partial ESI+MS spectra for 2-hydroxysorangiadenosine (2).

### 1.3 Alternative producer investigation

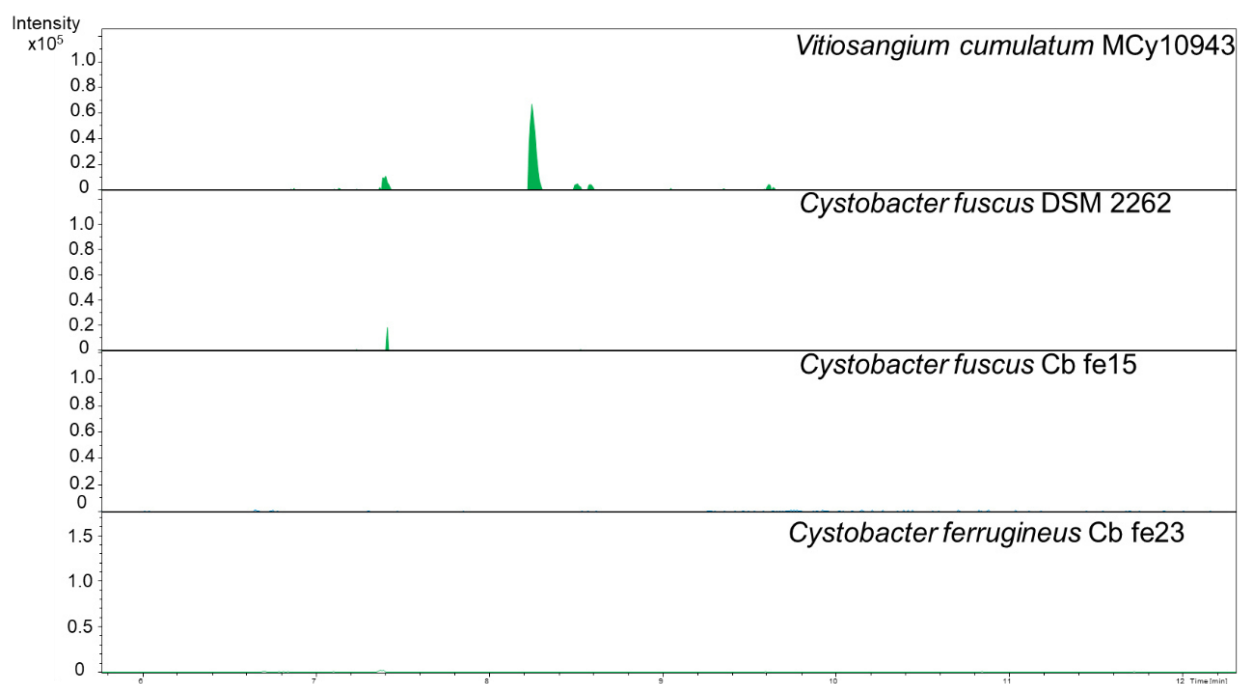


**Figure S9.** HPLC-MS EICs of crude extracts of *Cystobacter* sp. MCy9101 (top) and *Vitosangium cumulatum* MCy10943<sup>T</sup> (bottom). EIC: Extracted ion chromatogram, green: 488.2873 m/z, with a width of 7.9 ppm, 2-hydroxysorangiadenosine [M+H]<sup>+</sup>.

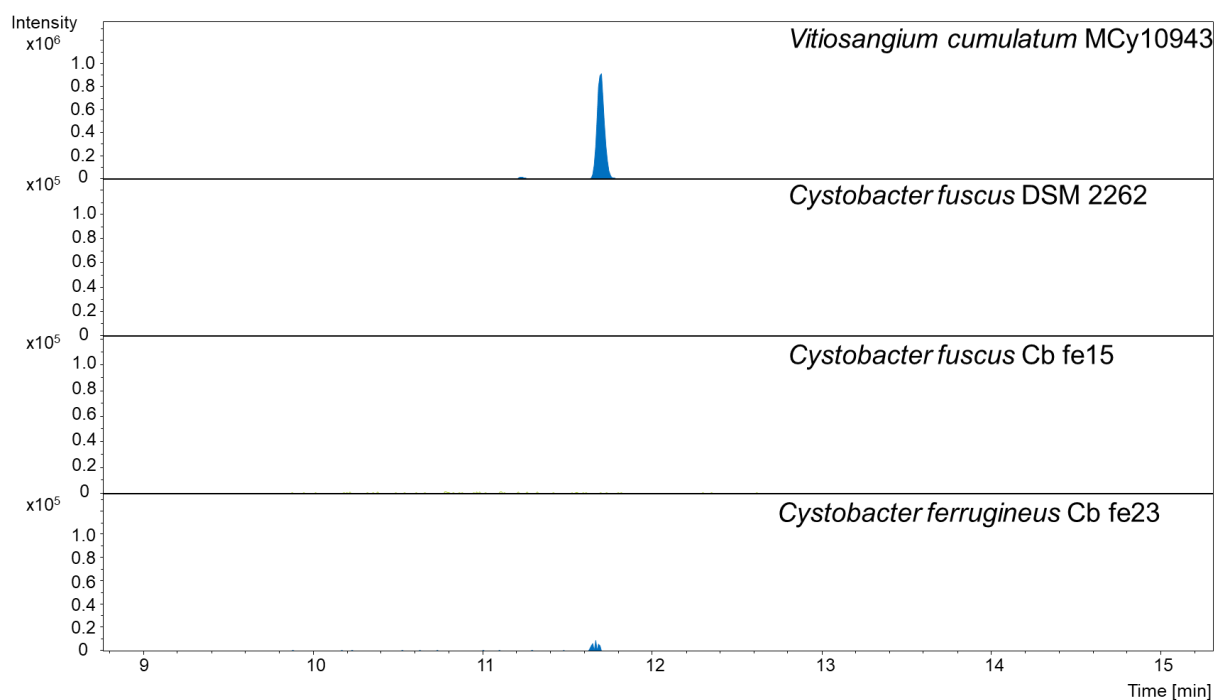




**Figure S10.** HPLC-MS EICs of crude extracts of *Cystobacter sp.* MCy9101 (top) and *Vitosangium cumulatum* MCy10943<sup>T</sup> (bottom). EIC: Extracted ion chromatogram, blue: 472.2924 m/z, with a width of 7.9 ppm, sorangiadenosine [M+H]<sup>+</sup>.

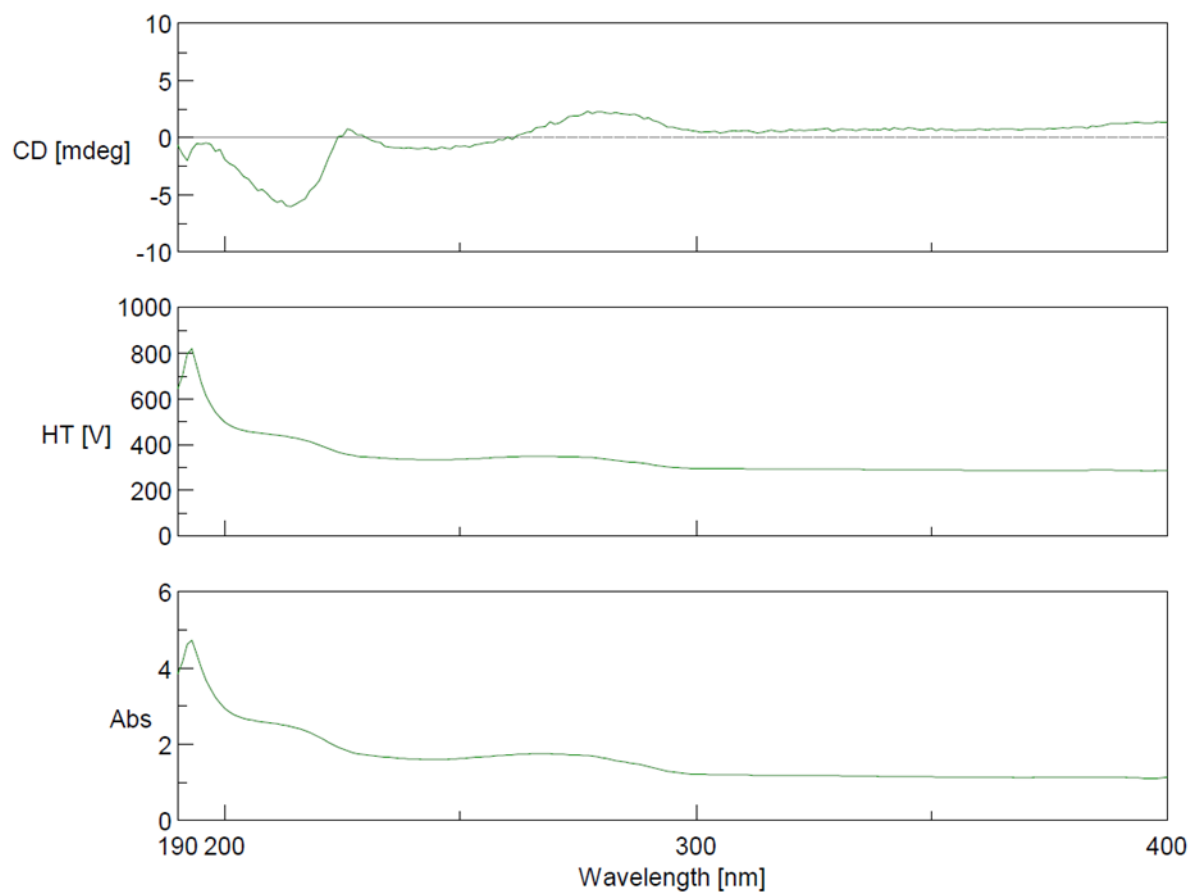


**Figure S11.** HPLC-MS EICs of crude extracts of *Vitiosangium cumulatum* MCy10943<sup>T</sup>, *Cystobacter fuscus* DSM 2262, *Cystobacter fuscus* Cb fe15 and *Cystobacter ferrugineus* Cb fe23. EIC: Extracted ion chromatogram, green: 488.2873 m/z, with a width of 7.9 ppm, 2-hydroxysorangiadenosine [M+H]<sup>+</sup>.



**Figure S12.** HPLC-MS EICs of crude extracts of *Vitiosangium cumulatum* MCy10943<sup>T</sup>, *Cystobacter fuscus* DSM 2262, *Cystobacter fuscus* Cb fe15 and *Cystobacter ferrugineus* Cb fe23. EIC: Extracted ion chromatogram, green: 472.2924 m/z, with a width of 7.9 ppm, sorangiaenosine [M+H]<sup>+</sup>.

## 1.4 CD spectrum



**Figure S13.** CD spectrum of 2-hydroxysorangiadenosine at 1.0 mM in MeOH in the area 190–400 nm, shows two negative cotton effect bands at 215 nm and 260 nm (top).

## 2. Biosynthetic *in silico* investigation

### 2.1 Leucine degradation (and alternative isovaleryl CoA biosynthesis)

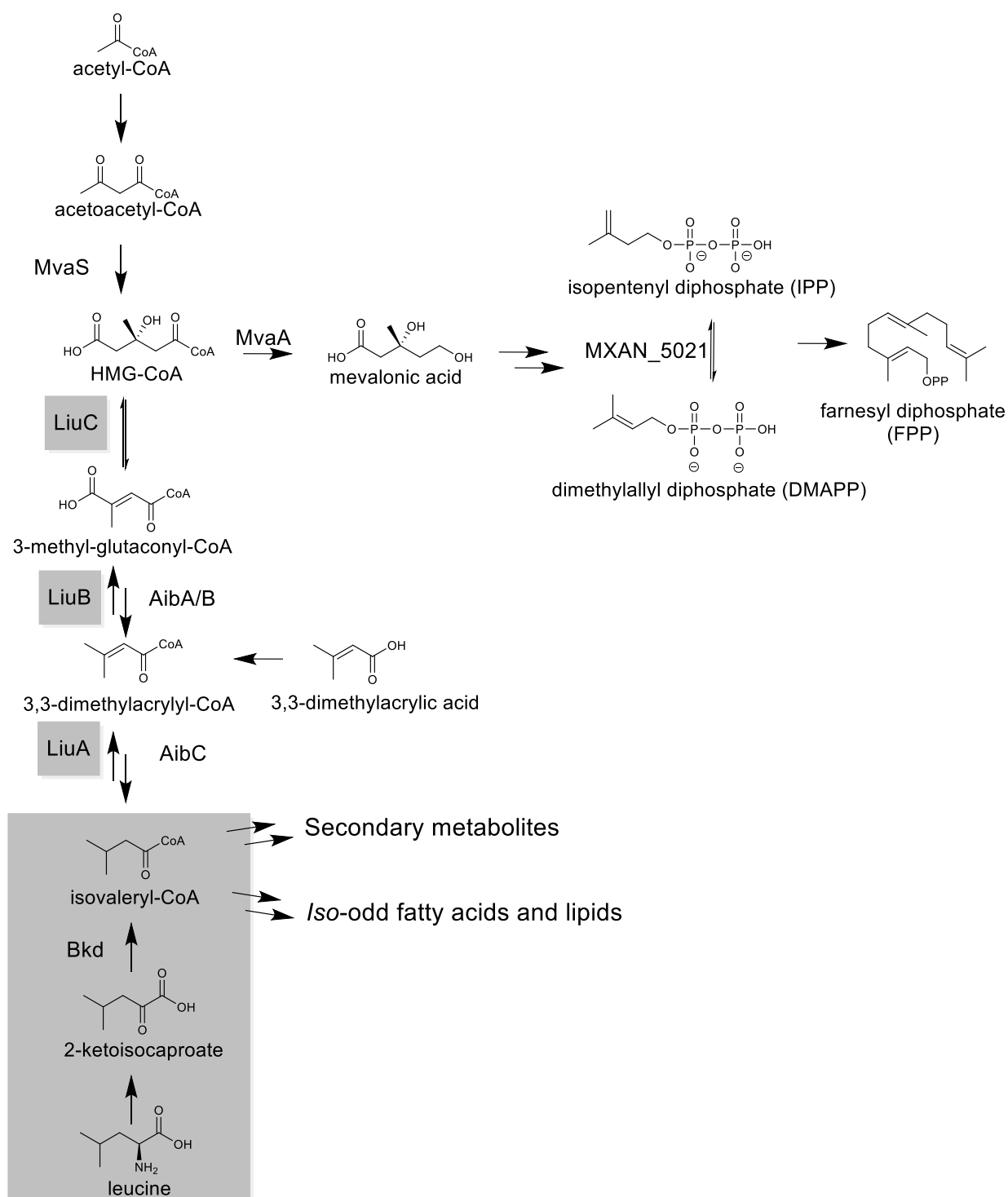
In myxobacteria, the essential terpene and terpenoid precursors isopentenyl (pyro)/diphosphate (IPP) and dimethylallyl diphosphate (DMAPP) can be biosynthesized through the mevalonate pathway, but also through degradation of leucine. This leucine degradation pathway has been identified and characterized previously in the myxobacteria *Myxococcus xanthus*<sup>1,2</sup> and *Stigmatella aurantiaca* Sg a15 and DW 4/3–1<sup>3</sup> (**Figure S14**).

The genes in the myxobacterium *M. xanthus* DK1622, which encode the proteins involved in leucine degradation and the alternative biosynthesis of isovaleryl-CoA<sup>2</sup> have been described previously. For that reason, the genome sequence of *Vitosangium cumulatum* MCy10943<sup>T</sup> was investigated accordingly to find homologues genes. All genes encoding proteins which are required for the leucine degradation pathway and the alternative biosynthesis of isovaleryl CoA from *M. xanthus* DK1622 and their corresponding homologues in *Vitosangium cumulatum* MCy10943<sup>T</sup> are listed in **Table S1**.

**Table S1.** Proteins involved in leucine degradation and in the alternative isovaleryl-CoA biosynthesis in *Myxococcus xanthus* DK1622 and their identified homologues in *Vitosangium cumulatum* MCy10943<sup>T</sup>.

Protein	Gene locus	Size (aa)	Deduced function	Coverage/ Similarity(%)
LiuA	MXAN_3760	381	short chain acyl-CoA DH	100/88.2
LiuB	MXAN_3759	513	carboxyl transferase ( $\alpha$ & $\beta$ SU)	99.81/90.1
LiuC	MXAN_3757	258	3-methylglutaconyl-CoA hydratase	100/85.7
AibR	MXAN_4263	228	TetR-like transcriptional regulator	82.89/76.2
AibA <sup>a</sup>	MXAN_4264	265	MG-CoA decarboxylase	100/80.8
AibB <sup>a</sup>	MXAN_4265	246	MG-CoA decarboxylase	100/84.6
AibC	MXAN_4266	345	dehydrogenase, Zn binding	99.71/84.0
MvaS	MXAN_4267	418	HMG-CoA synthase	99.76/82.3
BkdE1a	MXAN_4564	336	2-oxoisovalerate DH complex E1 $\alpha$ SU	100/80.4
BkdE1b	MXAN_4565	352	2-oxoisovalerate DH complex E1 $\beta$ SU	100/84.9
MvaA	MXAN_5020	442	HMG-CoA synthase	99.55/81.1
–	MXAN_5021	352	isopentenyl diphosphate isomerase	99.72/82.9

<sup>1</sup>Liu<sup>1</sup>: leucine and isovalerate utilization// Aib <sup>2</sup>: alternative isovaleryl-CoA biosynthesis/ AibA/AibB: Catalytically active as complex<sup>4</sup>. Bkd: (branched chain keto acid dehydrogenase).



**Figure S14.** Leucine degradation pathway (grey boxes) and the alternative biosynthesis of isovaleryl CoA observed in *Myxococcus xanthus* DK1622. Proteins are listed in **Table S1**.

## Amino acid sequence of enzymes involved in leucine degradation and alternative isovaleryl-CoA biosynthesis homologues in MCy10943<sup>T</sup>

### >LiuA homologue:

MDFELPESHRLQSSLRDFCERKVKPYAREWDKEEFPMEEVVRELGELGVLGMLVSEYGGGAAMDSL  
AVAVAVEEIARYDGLSLALTVASHNGLGTSHLRVFGNDAQRRKYLPLKATGEYLGAWGLTEPGSGSDA  
AGMKTTAVRKGDKWVLNGTKMFIQTGTGVSFVVLAVTSPEKKQKGITAFILEKGLPGFSQRSIHGKL  
GMRSSDTAELVMENVEVPDSQRLGEVDHGFIDTMKILDKGRITIGALAVGLARGALEESVRYAQERTA  
FGQPIADFQGLRWMFADMKTETDAARLLVHRAAFLADAGQPYSQEASMAKLFASEVATRACGKAVQ  
IHGGYGYTREFPVERYLRDAKLCEIGEGTSEIQRTHIAREVFKKA

### >LiuB homologue:

MPPGAMSFDQKLEKIAQVEKGGAEKYHAKNRETGKLFARERIRLLVDEGSFVEDAKLANNLDAELPS  
DGVIIGVGKVAGRAVAIMANDSTVKAGSWGARTVEKILRIQETAKQLRCPLMYLVDSAGARITDQVE  
MFPGRRGAGRIFYNEVHMSGEVPQVCLLFGPSAAGGAYIPAFCDLVIMVDGNASMYLGSPRMAEMVI  
GEKVTLEEMGGAKMHCSVSGVGDVLVKTEQEAIEAAKKYISFFPENFTQVPPRADPRAPKSSGKRVE  
IIPADQNKPFDMYALINELIDEGSWFEVKKLFAQELITGLARIDGRPVGIVANQPKYKGGVLFVDSADK  
AARFIWLCDAFNIPLLYLADVPGFMIGTKVERAGIIRAGAKMISAVSEASVPKICVVVRKAYGAGLYAM  
SGPGFAPDATLALPQAMIAVMGPEAAVNNAVYFNKIQEKPEAERAAYVQQLRDEYRQDVDIYKLASEL  
VVDEIVPGDRLRHELQQRYYELYSRRFQPRETKKHGVYPV

### >LiuC homologue:

MAEFKVDARGPIEIWTIDGEGRRNAISRAMLKEFEEMVARVSHGHDTRAVVVTGAGDKAFCAGADLK  
ERSTMSEPEVRAFLDGLRRTLRSIEKSDCVFIAAINGAAFGGGTELALSCDLRVAAPAAELGLTEVKLGI  
IPGGGGTQRLTRLVGPGRAKDLILTGRRLNAAEAFSIGLVNRLAPEGHLLDTAYSMAESIVENAPIAVAT  
AKHAIDEGLSLELDEALALELRHYEKLVLATEDRLEGLKAFAEKRKPVYKGR

### >AibR homologue:

MQSGRRPDDGERYRAILETAARLICERGFEGTSMQEIAAACRMTKAGLYHHVQNKEQLLFDIMSYGM  
DAFEQQVLDKVRSNPDPVERLRECMRLNIHLVTGGCIKEVIIIHEHATLRGEARAFIDSRRKAYVRFL  
DSFSEAVRMGRIRPVQPTVAAFSFLGMVLWYKWFQPDGRLSADQVANEMVELLFAGLVTPAAAAAP  
GAGAPMLALVPPKAVGGES

### >AibA homologue:

MRPARWGSVTELVASIPDGALLATGGFMLGRAPMALVLELIAQGRRNLQLISLPNPLPAEFLVAGGCL  
ARVELPFGALNLEGRVRPMPCLKRAIEQGRIDWREHDGYRVVQRLRAASMGLPFIPAPDADVSELADA  
EPLQTVVDPFTGQRPVPEPAVYPDVALIHARAADERGNLFIEDPTDLLVAGAARRVLATAEERVTRL  
RVTIPGFQVERVALARDGALPTGCTGLYPHDDEMLADYLALAEAGREAEFLDSLRLTRRAA

### >AibB homologue:

MSVESSVSPAEEVVVSLAREIDDGAVVATGVASPLAILAIAVARATHAPGLTYLACVGSGLDPLPCLLP  
SSEDLGYLEGRSAEISIPDLFDHARRGRVDTVFFGAAEVDVQGRTNMTASGSLEKPRTKFPGVAGAATL  
RQWVRRPVLVVRQSRRLVPEVQVVTTRDDRRPVKLISDLGVFELGAGGARLLSRHPWASLDTIGER  
TGFSFQVEESLPVTPLPDARTVSAIRAIDSHRLRDQLVGA

### >AibC homologue:

MKAVVLRFGESNLRMESVPDRPGRGEVLIRVHACGVCYHDVINRRGNLPRTHVPAILGHEAAGEV  
VEVGPDTPGWKVGDVRVATLQRLSCGECALCKSGRNSLCKKDNRRFFGEEISGGYAQYMTAPVAGLGRV  
PANLPWPVAATVCCTLGTAVHTVRTRARVRAGETVLITGASGGVGLAAVQLAKLDGARVIAVTSGEA  
KVQPLREAGADEVIVSRGLDFAAETRKRTGGEGVNVAIEIVGSATFGQTLKAMAPGGRVGVVGNLESG  
IVELNPGLVIVKELEILGAYATTREELDESLRLTADGKIRPFVSESVPLEAAARAHFRLNREIAGRLVLV  
PPELQ

### >MvaS homologue:

MKKQVGIEALAIAPRRYVDIEDLARARGVDPKFTVGLGAKEMAVADPGEDSVALAATAARLIQR  
NDVDPKVGMLVVGTTGTVDHSAVASHVQGLLKLPRSMRTYDAQHACYGGTAGLMAA VEWIASG  
AGAGRSAIVVCSDIARYGLNTAGEPTQGGGAVALLVSEQPDLLALDIGLNGVCSNDVYDFWRPLGRRE  
AVVDGHYSISCYLEALSGAYRGWRERALAHEVVRWGETLPGEQLARILYHVPFCKMARKAHTQLRLC  
DLEDAPGAGASTPAAREEIAKSSASYDAQVASSLGINARVGNVYTASLYLALAGLLQGESAALAGKRI  
GLLSYSGGCCAEFYSGVVGAAAAQRMADKADVESVLAKRERITL EYERIMRLSSDAPERVNPASGEFR  
LVEIRDHRRYYAAG

### >BkdE1a homologue:

MPKPRLLNREEDSLPLERELLVRMHDLMVKARVLEERLIQMYKQGHGYFWIGGPGEAFNVPLGLLM  
KKGQGPAYDYLHAHYRQSATLLALGEEPIGSLRQMKNTATDPYSGGRNFAGHYSKREWNIA PVSSPIE  
VQYVMAPGTALAQKRHGGDGITIVTGGDAGTAEGDFASCLIWSSRPGNELPLLIIVTNNKWGISTPAET  
QHGETHVADRGKAFNIRSKTIDGNDPINAYLELKEAMEYVRKERKPFLEAMVSRLYGHSSASGANF  
VGNELDCLARFEERLEKNGVLTRKEMDDLNRNYTEDMAAMARQVREEPLPAPETIWNHIFAERK

### > BkdE1b homologue:

MANMAQAIRMAHYAENLGVTDVFGEDVGAPLGGVFTATQGLKTAWNSPLDERGIIGMAIGLAMA  
GQKPVAEIQFADYIYNTIDLLKIAGNTCWSHGDWNVPMVVRTPVSGSIRGSIYHSHSFDATATHIPGW  
KIVMPSNPLDAYGLLISACQEINPVMYLEPKALLRIKGEERIPGEPDDDKLLSKMIDAPLGD RSQWKPQ  
WPQQLEAYAVPIGKGKVVRSGSQVTVVSYGRTLPLCKKAADDLASAGVDAEVIDMRTLWPYDWELI  
KGSVEKTGRVLYVNEDTEVTNFGEHLVRRTVEELFFKL VAPPRLLAGKFVPGIGLADTLEMASVPQLG  
DITAAIRSLAAEQP

### >MvaA homologue:

MSETLTSRLSGFHKLSMVERHEKLAQMLGLDEMDLAQLQGISA LQPGLANQMIENAVGTFSPLPLGLL  
NLQINGRDYLVPMAVEEPSVVA AVSFASRIVREAGGFSAEADDPIMIGQVQLTRYGDPTEATKKILA AK  
EALLALANSFHPMSMVKRGGGCKDIEVRVLPAPEGPRGEPLLVHLLI DTQEAMGANLINTMAEGVAPLL  
EQLTGKVFRLRILSNLADRRLARATCRIPVEALADFDMPGHVIAEGIYQASRFALADPYRAATHNKGIM  
NGIDSAAIAAGQDWRAIEAGAHAFACRDGQYRPLSTWHVDDGHLVGRIELPMALGLVGGPIKVHPGV  
QVAMKILRVESARELAMVFAAVGLAQNF AAVRALGSIGIQKGHMA LHARCVA VTAGARGDWVEKIA  
ELLVSAGHVKVEKAREIIASLSAEDFRAATGTDS

### >MXAN\_5021 homologue:

MVPKQGRRDVTGEEATAKRKDAHL DLCATGEVEPAENSTLLEYVRLVHCAMP EMAVEEVDLSTEF LG  
KKLRYPLITGMTGTERAGVVNRDLALLAERHGLAFGVGSQRAMAENPQA AETYVVRHVAPTVP LL  
GNIGLYQAIELGVDGVRRLADAIGADGMALHLNAGQELTQPEGDRDFRGGYTVVEGLVRVFGARLLV  
KETGCGIGPEVARRLVDLGVRLNDVSGLGGTSWVRVEQLRASGVQAQVGA EFSSWGIPTAAAIASVRR  
AVGPEPRLVASGGLRTGLDAAKVLALGADLAGMALPLFRAQQAGGLEGA EQALAVILSGLRQALVLT  
GSRNCGELRQKPRVIMGQLKDWLAAL

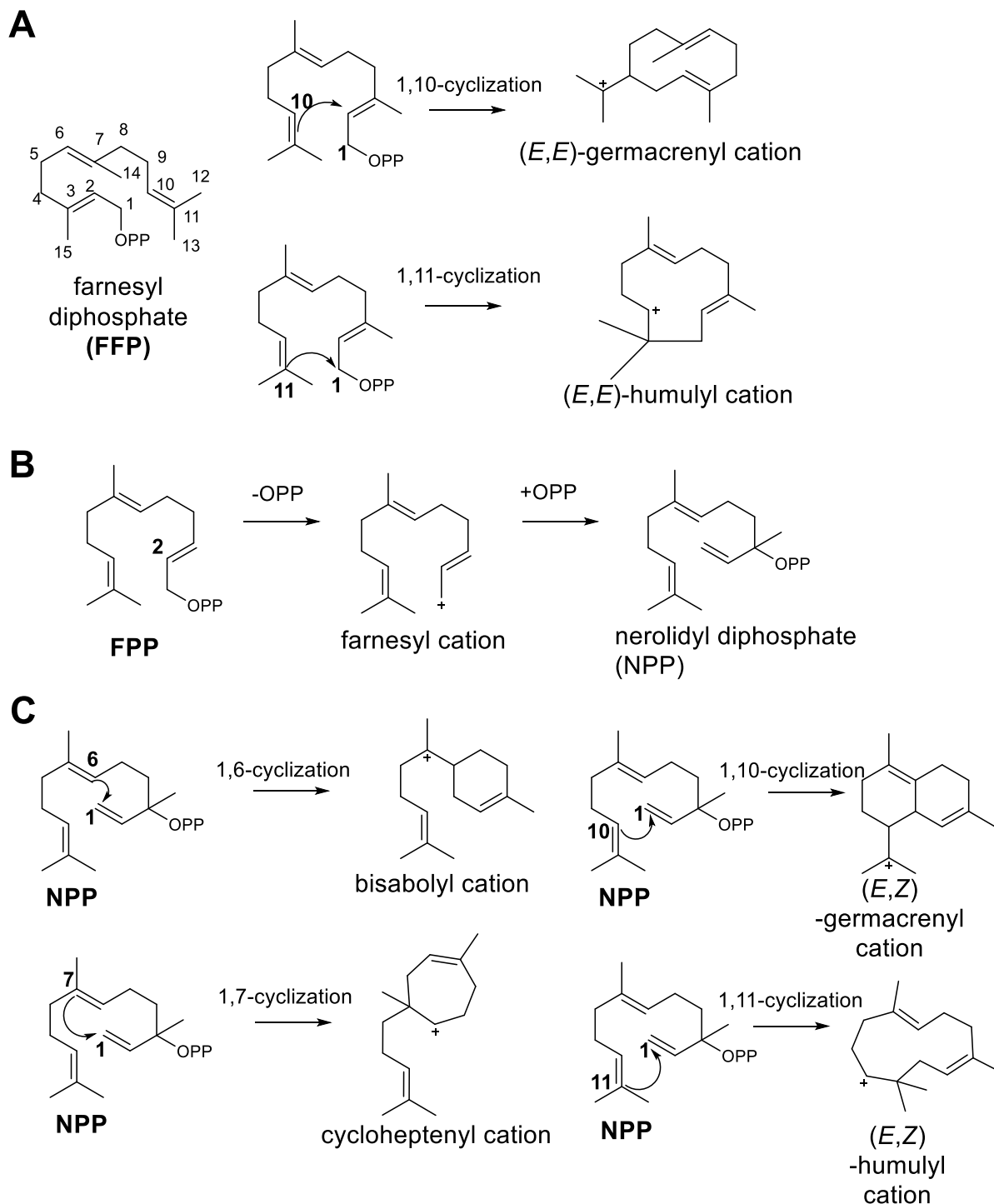


## 2.2 Formation of sesquiterpene scaffolds and eudesmane-type sesquiterpenes

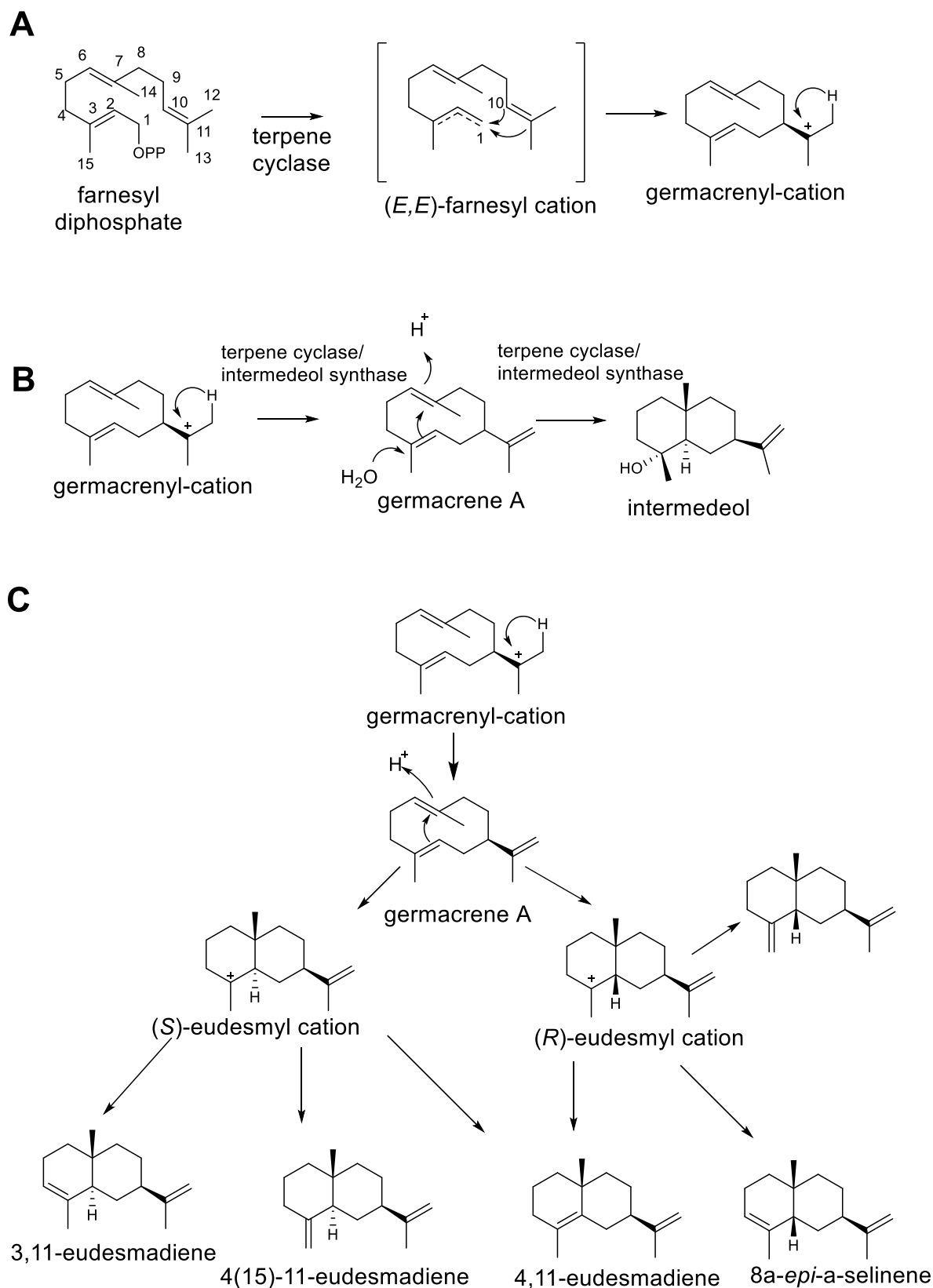
The formation of the sesquiterpene scaffold starts from farnesyl pyrophosphate (FPP) with subsequent cyclization via a sesquiterpene cyclase, which is known to catalyze the formation of diverse (poly)cyclic sesquiterpene skeletons<sup>5</sup>. Six different initial cyclization reactions are possible starting from farnesyl diphosphate<sup>5</sup> (**Figure S15A**). These include the direct conversion of farnesyl diphosphate to (*E,E*)-germacrenyl cation via 1-10-cyclization or 1-11 cyclization to the (*E,E*)-humulyl cation. The 2-*E* double bond of FPP can be isomerized via ionization leading to the farnesyl cation, followed by reattachment of diphosphate at C-3 to yield nerolidyl diphosphate (NPP) (**Figure S15B**). This allows NPP a 1,6-cyclization to the bisabolyl cation, a 1,7-cyclization to the cycloheptenyl cation, a 1,10-cyclization to the (*E,Z*)-germacrenyl cation or a 1,11-cyclization to the (*E,Z*)-humulyl cation (**Figure S15C**).

According to the featured eudesmane-type sesquiterpene structure, a logical proposal would start from FPP to yield via a 1,10-cyclization the (*E,E*) germacrenyl cation (**Figure S16A**). The germacrenyl cation could also be cyclized to intermedeol, such as described in the literature for intermedeol cyclases (**Figure S16B**)<sup>6</sup>. However, the secondary metabolome of MCy10943<sup>T</sup> shows no signal corresponding to the sum formula or mass of intermedeol or a hydroxylated eudesmadiene intermediate. On the contrary, three compounds with the mass of 205 m/z and the sum formula C<sub>15</sub>H<sub>25</sub> are present in the crude extract of MCy10943<sup>T</sup>, which correspond to the non-hydroxylated eudesmadiene intermediates (**Figure S16C**, **Figure S17**). Cyclization of eudesmadiene begins with a 1-10 bond formation to yield germacrenyl A which is then protonated at C<sub>6</sub> to facilitate C<sub>2</sub>-C<sub>7</sub> bond formation resulting in the eudesmane carbocations.

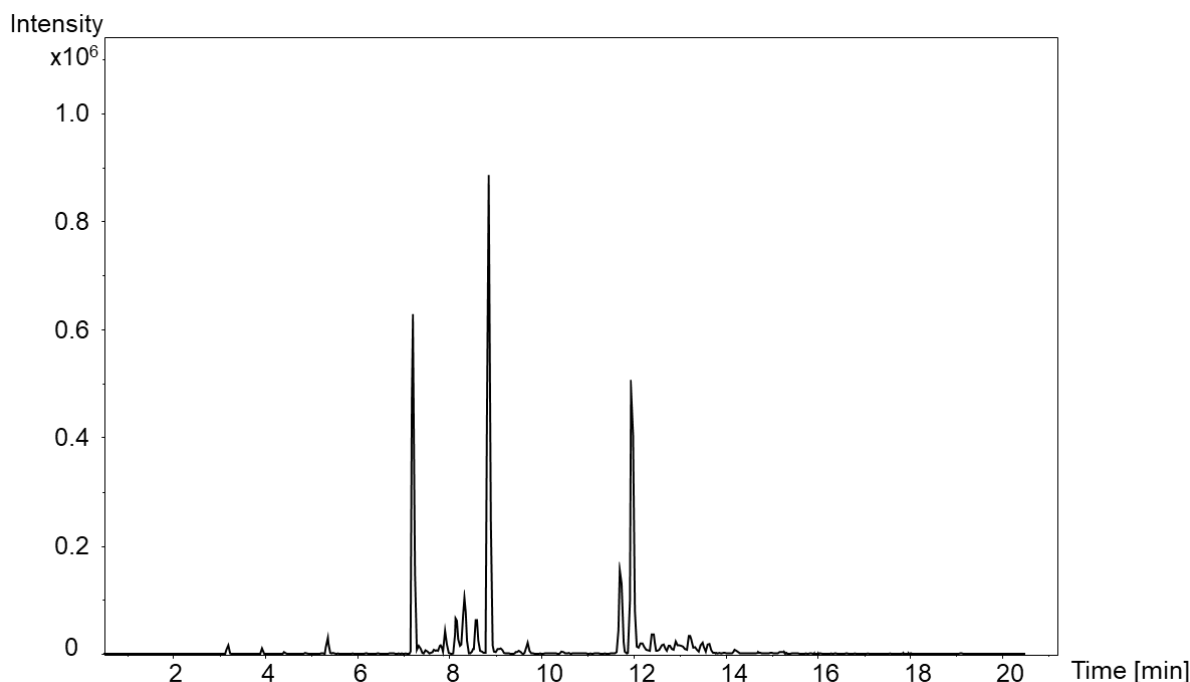
During the process, the C<sub>2</sub>-C<sub>3</sub> bond of the farnesyl diphosphate retains its *trans* configuration. Elimination of protons at  $\alpha$ ,  $\beta$  or  $\gamma$  to the (*S*)-eudesmyl carbocation center accounts for the formation of  $\alpha$ -selinene,  $\beta$ -selinene and selina-4(11)-diene. Elimination of protons at  $\alpha$ ,  $\beta$  or  $\gamma$  to the (*R*)-eudesmyl carbocation center accounts for the formation of alternative products like 8a-*epi*- $\alpha$ -selinene, selina-4(11)-diene. Considering the stereochemistry of the products it is assumed that **1** derives from the (*S*)-eudesmyl carbocation in the second stage.



**Figure S15.** 1,10-cyclization and 1,11-cyclization from farnesylpyrophosphate (FFP) yield the (*E,E*)-germacrenyl cation, which is the proposed precursor for the generation of eudesmane-type sesquiterpene required for the biosynthesis of **1** and **2** or (*E,E*)-humulyl cation, respectively. Both enantiomers of the (*E,E*)-germacrenyl cation can be formed. **B**) Isomerization mechanism of FFP to nerolidyl diphosphate (NPP). **C**) NPP can be cyclized to a bisabolyl cation via 1,6-cyclization, to the cycloheptenyl cation via 1,7-cyclization, to the (*E,Z*)-germacrenyl cation via a 1,10-cyclization or to the (*E,Z*)-humulyl cation via a 1,11-cyclization. Both enantiomeric products of the 1,6-cyclization, 1,10-cyclization and 1,11-cyclization can be generated.



**Figure S16.** A) Proposed start of the sesquiterpene part during the biosynthesis of sorangiadenosine yielding the intermediate germacrenyl-cation. B) Possible formation of intermedeol from the intermediate germacrenyl-cation. C) Proposed formation of the different eudesmadiene building blocks.



**Figure S17.** HPLC-MS EIC of *Vitiosangium cumulatum* MCy10943<sup>T</sup> crude extract. EIC: Extracted ion chromatogram, black: 205.1956 m/z, with a width of 7.9 ppm, proposed non-hydroxylated eudesmadiene intermediates [M+H]<sup>+</sup>.

## 2.3 Metabolome-genome correlation

Initial *in silico* analysis already excluded two of the antiSMASH-identified biosynthetic gene clusters (BGCs) to produce **1** and **2**, since both terpene cyclases share high sequence similarity to the geosmin terpene cyclase. In addition, a study investigating different terpene cyclases from *S. cellulorum* So ce56 suggested, that the geosmin cyclase gene (*sce1440*) is specific in terms of its production profile<sup>7</sup>. In order to narrow down the number of suitable BGCs for the production of **1** and **2** we performed metabolomic profiling of the myxobacterial strains *Cystobacter ferrugineus* Cb fe23, *Cystobacter fuscus* Cb fe15 (DSM 52655) and *Cystobacter fuscus* DSM 2262<sup>T</sup>. These three strains harbor according antiSMASH and manual evaluation similar terpene gene clusters to No. 3, 5 and 6 (GenBank Accession numbers upon acceptance) (**Table 1**). In the respective crude extracts, no metabolites corresponding to the retention time, exact molecular mass and sum formula connected to **1** and **2** was found in the LC-MS chromatogram of these three myxobacterial strains (**Figure S11** and **S12**). In addition, the gene clusters No. 3, 5, 6 were further excluded, since as proposed above, each gene cluster is missing at least either a gene encoding an oligoprenyl transferase or a cytochrome P450 gene. Taken together, genomic investigation combined with metabolomic profiling seemingly reduced the likelihood that the terpene gene clusters No. 3, 5 and 6 are responsible for the production of **1** and **2**. However, since genes responsible for the production of the terpenes might be poorly expressed in some organisms and genes pivotal for the formation of the terpene scaffold can be located outside of the biosynthetic core region (terpene or diterpene cyclase genes), it cannot be entirely excluded that the terpene gene clusters No. 3, 5 and 6 are in some way involved in the biosynthesis of **1** and **2**.

Consequently, this metabolome-genome correlation indicates that only the terpene gene clusters 1, 4 and 8 might be involved in the biosynthesis of **1** and **2**. The terpene cyclase in cluster No. 4 identified

as pentalene cyclase could account for the biosynthesis of an eudesmane-type sesquiterpene according to structural relationship to known eudesmane/selinadiene forming terpene cyclases. In addition, at least one aldo-keto oxidoreductase within terpene gene cluster No. 4 could be responsible for the conversion of **1** to **2**, albeit this enzyme would rather change the redox state of the alcohol/ketone than install a hydroxyl group. However, the identified terpene gene cluster No. 4 lacks presumably a gene which might account for transferring the adenosine to the sesquiterpene. Terpene gene cluster No. 8 harbors neither genes responsible for hydroxylation of **1** leading to **2**, nor it features genes accounting for the transfer of adenosine onto the eudesmane-type sesquiterpene.

The alternative producers of **1** and **2** with available in-house genome sequences *Archangium* sp. strain MCy8383 (Ar3548) (GenBank Accession number MT520809), *Archangium gephyra* strain MCy8375 (Ar8082) (GenBank Accession number MT520808) and *Cystobacter* sp. strain MCy9101 (SBCb004) (GenBank Accession number MT520810) harbor a terpene BGC resembling the candidate 2-hydroxysorangiadenosine BGC. The proposed hydroxysorangiadenosine BGCs in *Archangium* sp. strain MCy8383 (Ar3548) and *Archangium gephyra* strain MCy8375 (Ar8082) are highly similar to the candidate gene cluster in *Vitiosangium cumulatum* MCy10943<sup>T</sup>, whereas the homologue BGC in *Cystobacter* sp. strain MCy9101 (SBCb004) deviates significantly from the herein described gene cluster organization; nevertheless all three homologues BGCs are missing the gene *sora12*. The genetic organization of the 2-hydroxysorangiadenosine BGC in *Cystobacter* sp. strain MCy9101 (SBCb004) is reduced to *sora7–9*, *sora11–12*, *sora18–19* and *sora21–22*. **Figure S18** shows this truncated 2-hydroxysorangiadenosine BGC in *Cystobacter* sp. strain MCy9101 (SBCb004) and **Figure S9** and **S10** demonstrate that this truncated BGC is presumably sufficient to produce **1** and **2** in low concentration.



**Figure S18.** Genetic organization of the truncated 2-hydroxysorangiadenosine BGC from *Cystobacter* sp. strain MCy9101 (SBCb004).

### 3. Structure elucidation

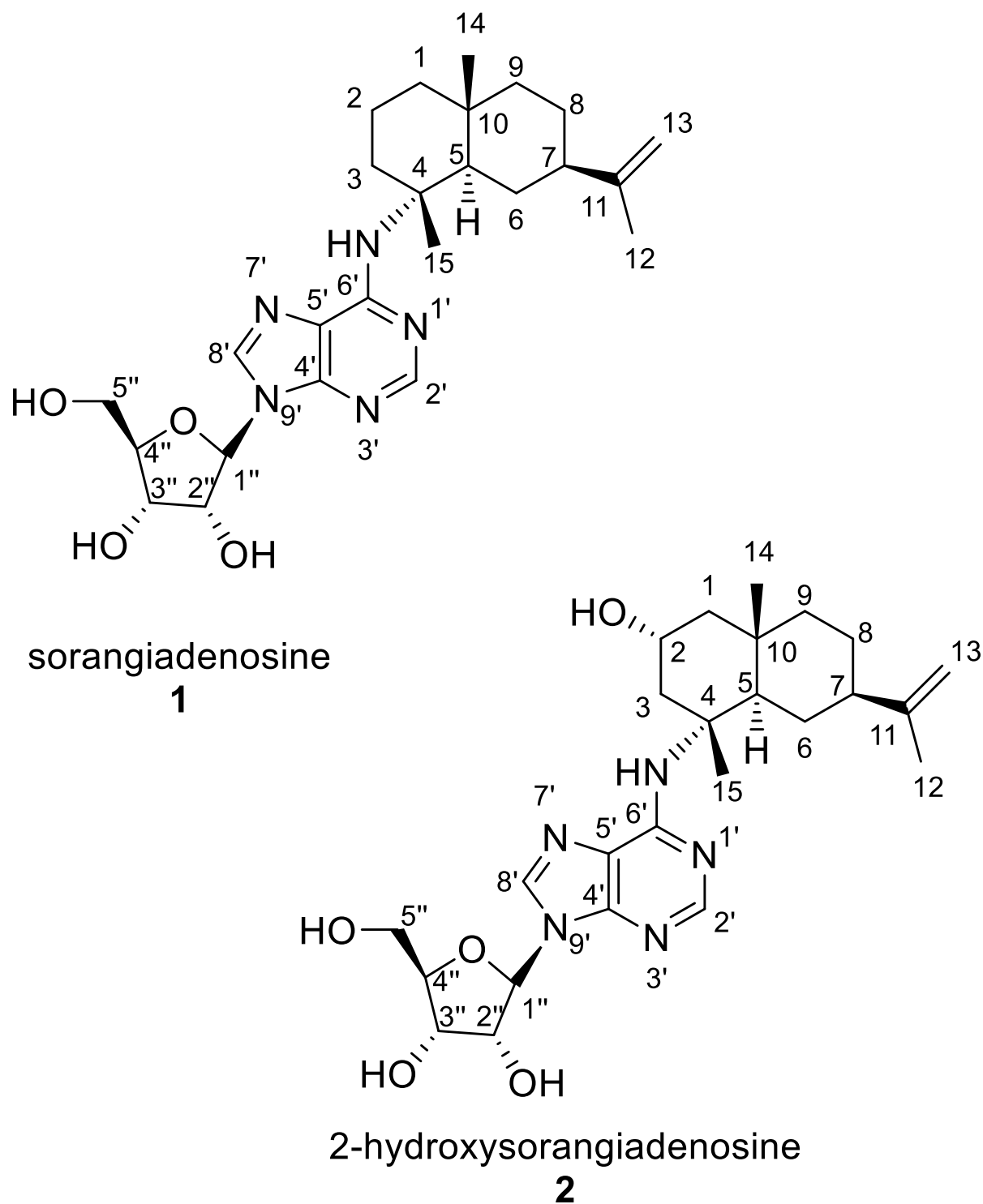
#### 3.1 NMR spectroscopic data

**Table S2.** NMR spectroscopic data for sorangiadenosine (**1**) in methanol-d<sub>4</sub>.

sorangiadenosine ( <b>1</b> )					
position	$\delta_{\text{C}}$ , type	$\delta_{\text{H}}$ , ( <i>J</i> in Hz)	COSY-DQF	gCOSY	HMBC
1	41.9	1.47 m 1.30 m	1.75	1.04, 1.59	2, 3,5,9,10,14
2	20.3	1.75 m 1.59 m	-	1.47, 1.30, 2.11	1, 3, 4
3	38.13	2.48 td (4.4/13.5) 2.11 br d (13.5)	1.75, 1.59	1.75, 1.30, 1.47	1, 2, 4, 5, 15
4	59.9	-	-	-	-
5	49.5	2.54 dd (2.0/12.5)	1.38	1.69, 1.38	3, 4, 6, 7, 10, 14, 15
6	27.9	1.69 m 1.38 d (12.5)	-	2.54, 1.94	5, 7, 8, 10, 11
7	47.7	1.94 m	1.69, 1.38	1.69, 1.38, 1.57	6, 8, 11, 12,13
8	28.3	1.57 m 1.49 m	-	1.94, 1.33, 1.47	6,7, 8, 9,10
9	46.5	1.47 m 1.33 m	1.49, 1.57	1.49, 1.57, 1.04	5, 8, 10, 14
10	35.8	-	-	-	-
11	151.7	-	-	-	-
12	21.1	1.68 s 3H	-	4.65	7, 11, 13
13	108.8	4.65 d 2H (16.1)	1.68	1.68	1, 7, 12
14	19.6	1.04 s 3H	-	1.30	1, 5, 9, 10
15	21.3	1.46 s 3H	-	1.75, 2.11, 2.48, 2.54	3, 4, 5, 6'
1'	-	-	-	-	-
2'	152.9	8.22 s	-	-	4', 5', 6'
3'	-	-	-	-	-
4'	148.7	-	-	-	-
5'	121.6	-	-	-	-
6'	155.9	-	-	-	-
7'	-	-	-	-	-
8'	141.1	8.24 s	-	5.94	1'', 4', 5'
9'	-	-	-	-	-
1''	91.3	5.94 d (6.5)	4.74	3.74, 4.17, 4.31, 4.74	2'',3'', 4', 8'
2''	75.4	4.74 dd (5.4, 6.5)	4.31	4.31, 5.94	1'', 4''
3''	72.7	4.31 dd (2.3/5.4)	4.17	4.17, 4.74	1'', 4'', 5''
4''	88.3	4.17 q (2.3)	3.88, 3.74	3.74, 3.88, 4.31	3''
5''	63.5	3.88 dd (2.3/12.6) 3.74 dd (2.3/12.6)	-	4.17	3'', 4''

**Table S3.** NMR spectroscopic data for 2-hydroxysorangiadenosine (**2**) methanol-d<sub>4</sub>.

2-hydroxysorangiadenosine ( <b>2</b> )					
position	$\delta_c$ , type	$\delta_H$ , (J in Hz)	COSY-DQF	gCOSY	HMBC
1	50.8	1.80 dd (3.8/12.2) 1.30 t (9.1/12.2)	2.41, 3.94	1.06, 2.41, 3.94	2, 3, 5, 10, 14
2	65.6	3.94 m	1.27, 1.80, 2.52, 2.41	1.27, 1.80, 2.41, 2.52	1, 3
3	46.6	2.52 t (12.1) 2.41 m	3.94	1.46, 1.80, 3.94	2, 4, 5, 15
4	60.4	-	-	-	-
5	48.5	2.60 dd (2.2/12.4)	1.70, 1.38	1.70, 1.38	3, 4, 6, 7, 10, 14, 15
6	27.6	1.70 m 1.38 t (12.4)	1.94, 2.60	2.60, 1.94	4, 7, 8, 10, 11, 14
7	47.6	1.94 m	1.56, 1.48, 1.70, 1.38	1.70, 1.48, 1.56, 4.65	11, 12, 13
8	28.0	1.56 m 1.48 m	1.94, 1.52, 1.38	1.94, 1.38, 1.52	6, 7, 10, 11
9	46.4	1.52 m 1.38 t (12.4)	-	-	1, 5, 8, 10, 14
10	35.6	-	-	-	-
11	151.5	-	-	-	-
12	21.1	1.68 s 3H	4.64	4.64	7, 11, 13
13	109.0	4.65 d 2H (15.4)	1.68	1.68	7, 11, 12
14	20.7	1.06 s 3H	-	1.27, 1.38	1, 5, 9, 10
15	22.4	1.46 s 3H	-	1.94, 2.60	3, 4, 5, 6'
1'	-	-	-	-	-
2'	152.9	8.24 s	-	-	4', 5', 6'
3'	-	-	-	-	-
4'	148.8	-	-	-	-
5'	121.6	-	-	-	-
6'	155.7	-	-	-	-
7'	-	-	-	-	-
8'	141.2	8.26 s	-	5.94	1'', 4', 5', 6'
9'	-	-	-	-	-
1''	91.3	5.94 d (6.5)	4.74	4.17, 4.74	2'', 3'', 4'', 4', 8'
2''	75.4	4.74 dd (5.1/6.4)	4.31	4.31, 5.94	1'', 3'', 4''
3''	72.7	4.31 dd (2.5/5.1)	4.17	4.17, 4.74	1'', 4'', 5''
4''	88.3	4.17 q (2.5)	3.88, 3.74, 4.31	3.74, 3.88, 4.31	3'', 5''
5''	63.5	3.88 dd (2.5/12.6) 3.74 dd (2.5/12.6)	4.17	4.17	3'', 4''



**Figure S19.** Chemical structure and numbering of sorangiadenosine (1) and 2-hydroxysorangiadenosine (2).



## 3.2 $^1\text{H}$ and $^{13}\text{C}$ NMR spectra of sorangadenosine and 2-hydroxysorangadenosine

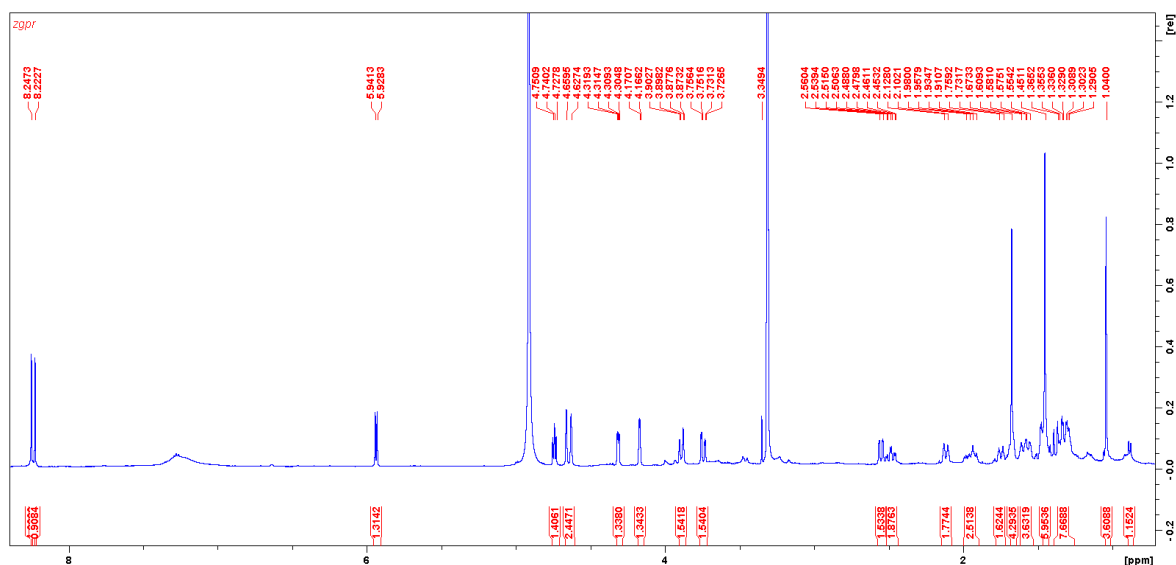


Figure S20.  $^1\text{H}$  NMR spectrum of sorangadenosine (1) in methanol- $\text{d}_4$ .

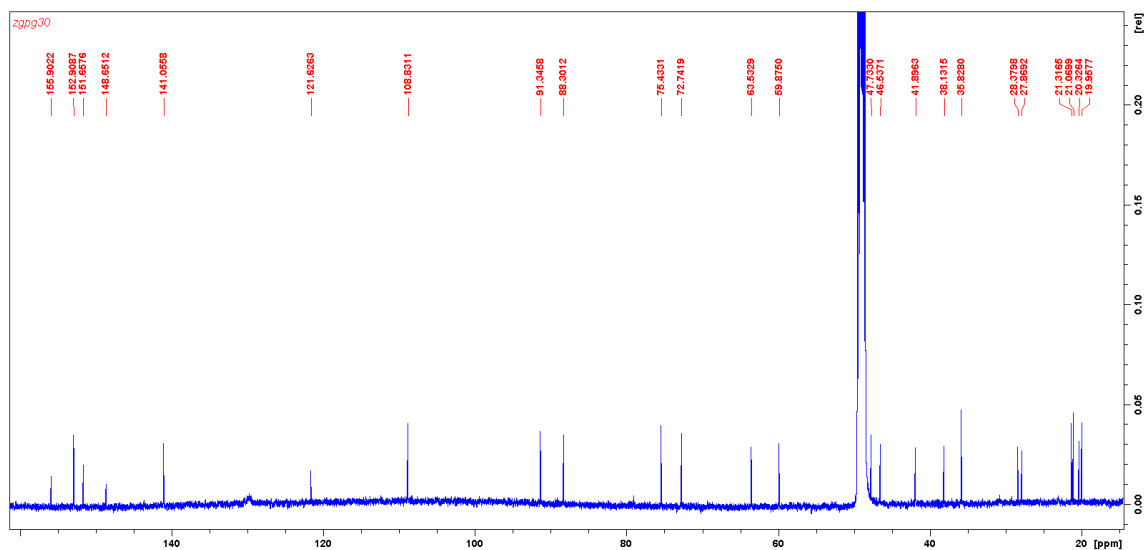
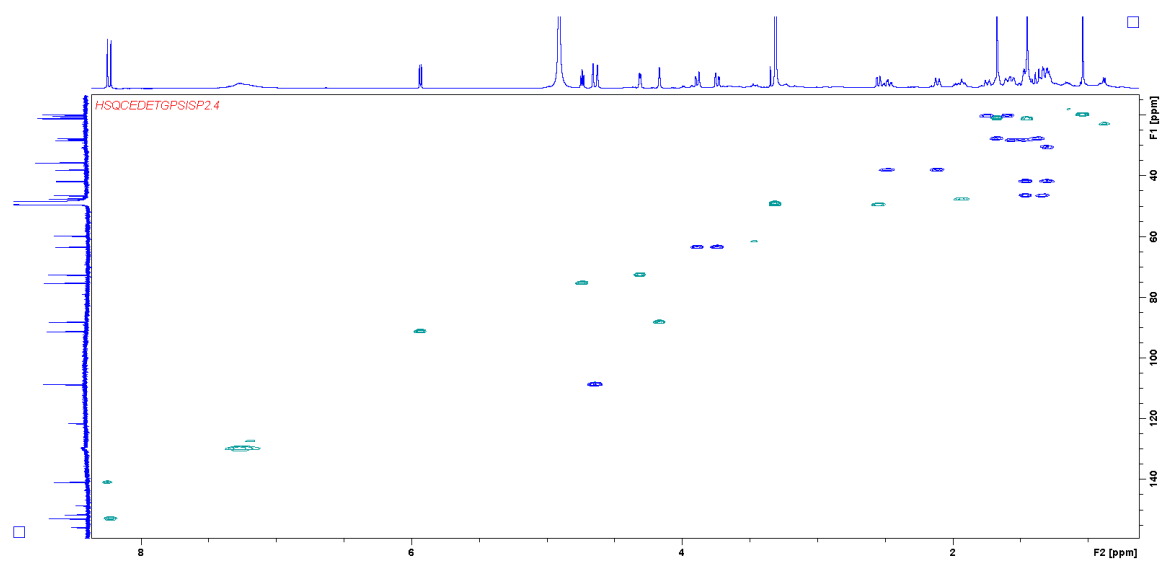
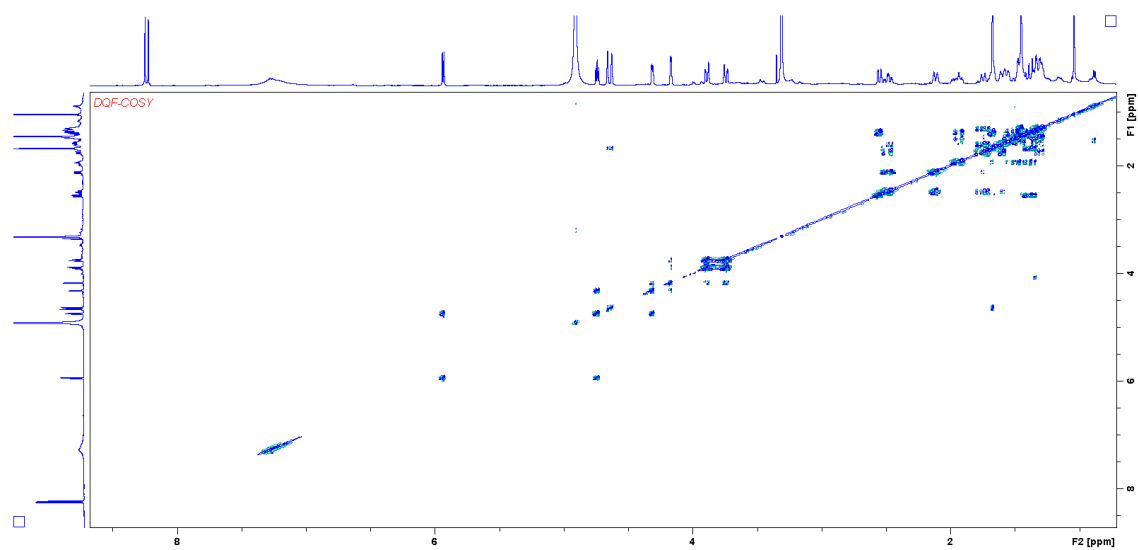


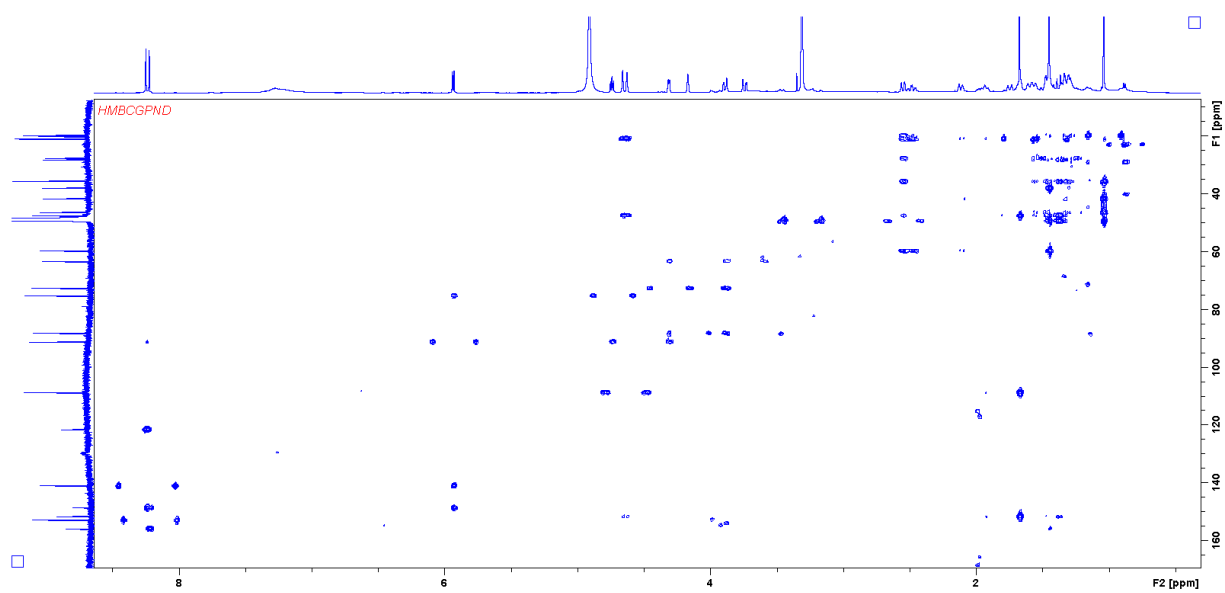
Figure S21.  $^{13}\text{C}$  NMR spectrum of sorangadenosine (1) in methanol- $\text{d}_4$ .



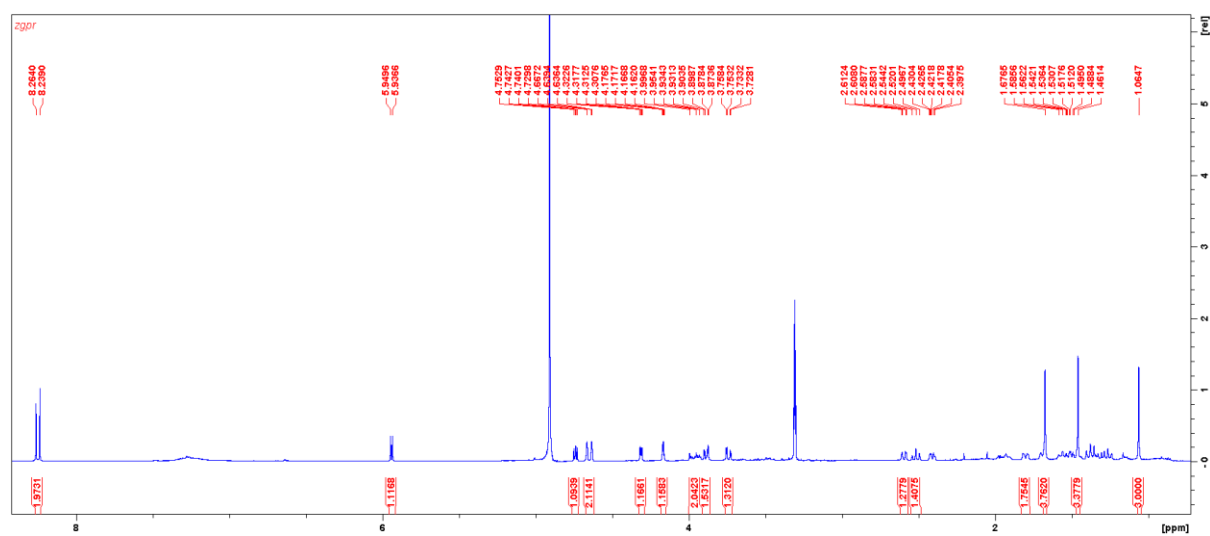
**Figure S22.** *HSQC* spectrum of sorangiadenosine (**1**) in methanol- $d_4$ .



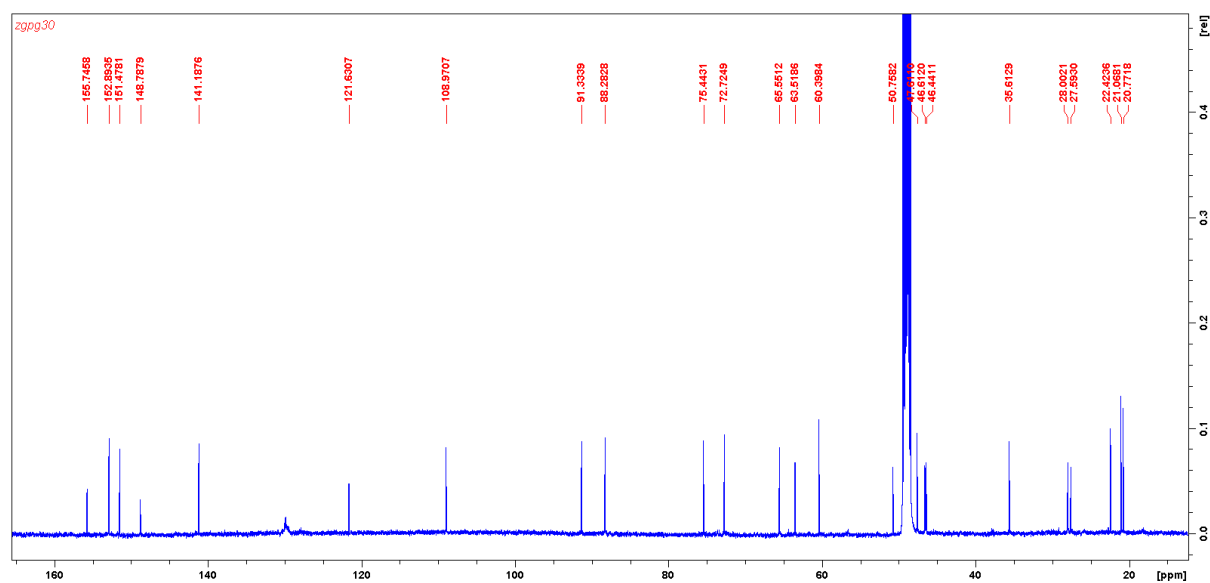
**Figure S23.** *DQF-COSY* spectrum of sorangiadenosine (**1**) in methanol- $d_4$ .



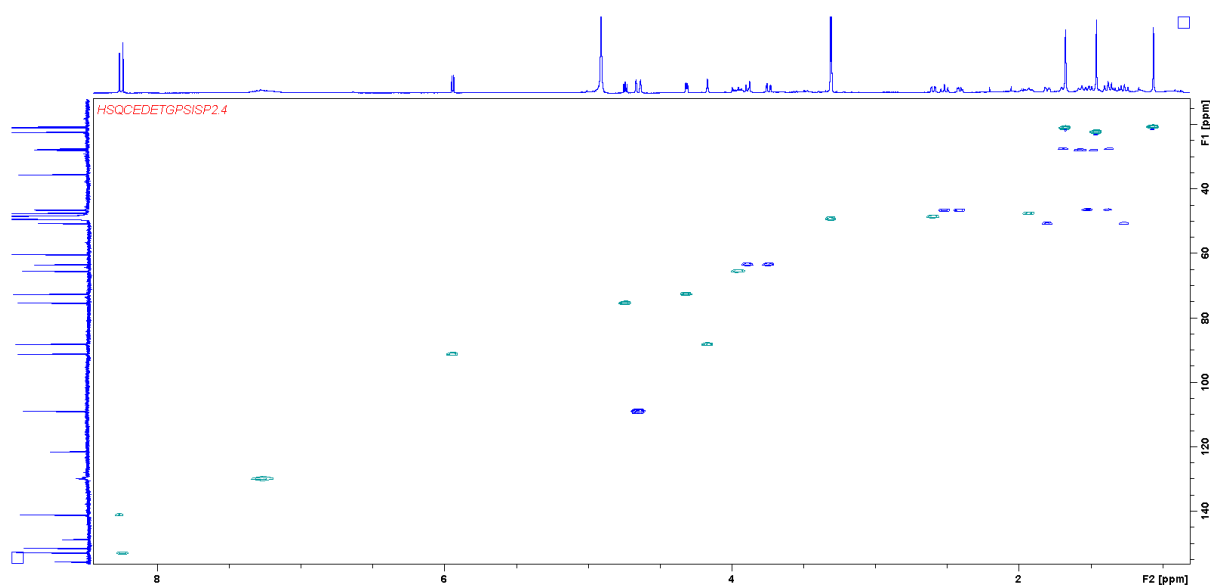
**Figure S24.** *HMBC* spectrum of sorangiaenosine (**1**) in methanol- $d_4$ .



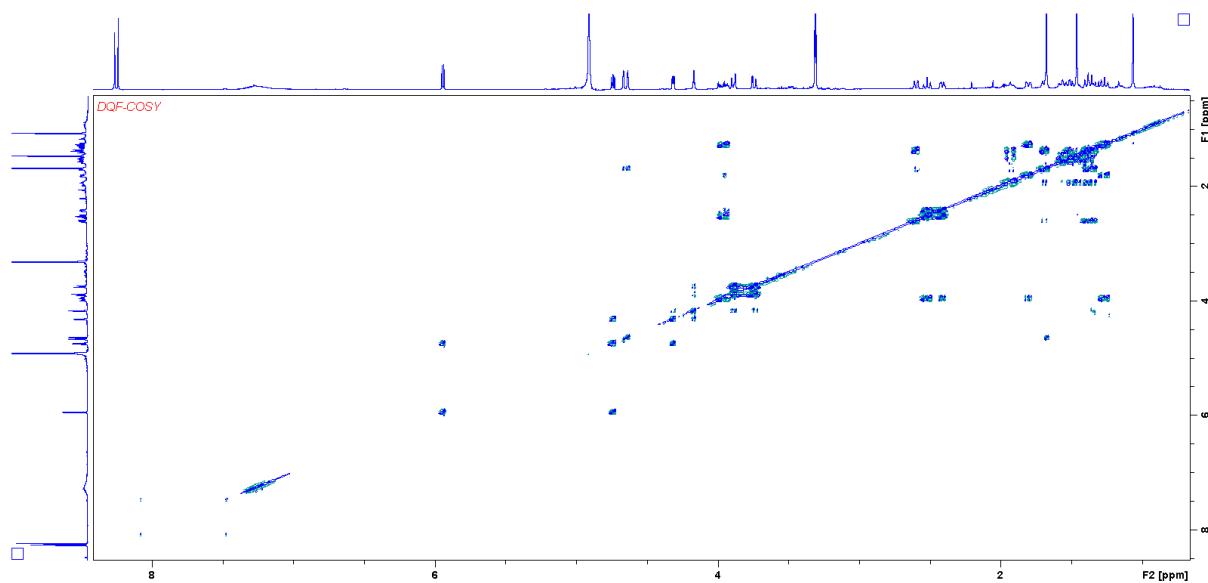
**Figure S25.**  $^1\text{H}$  NMR spectrum of 2-hydroxysorangiaenosine (**2**) in methanol- $d_4$ .



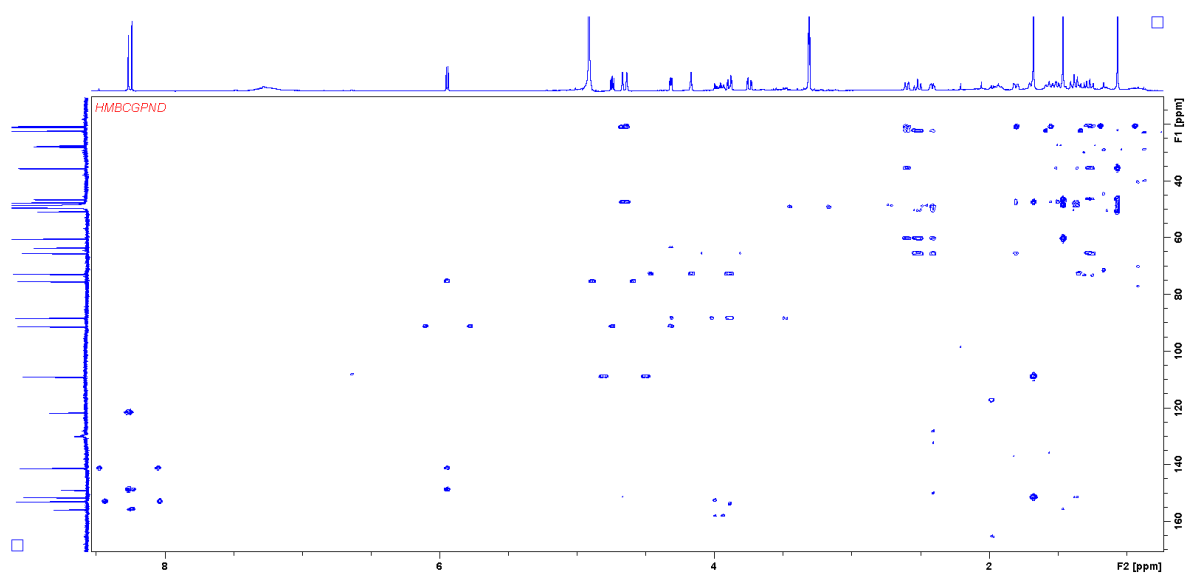
**Figure S26.** <sup>13</sup>C NMR spectrum of 2-hydroxysorangiadenosine (2) in methanol-d<sub>4</sub>.



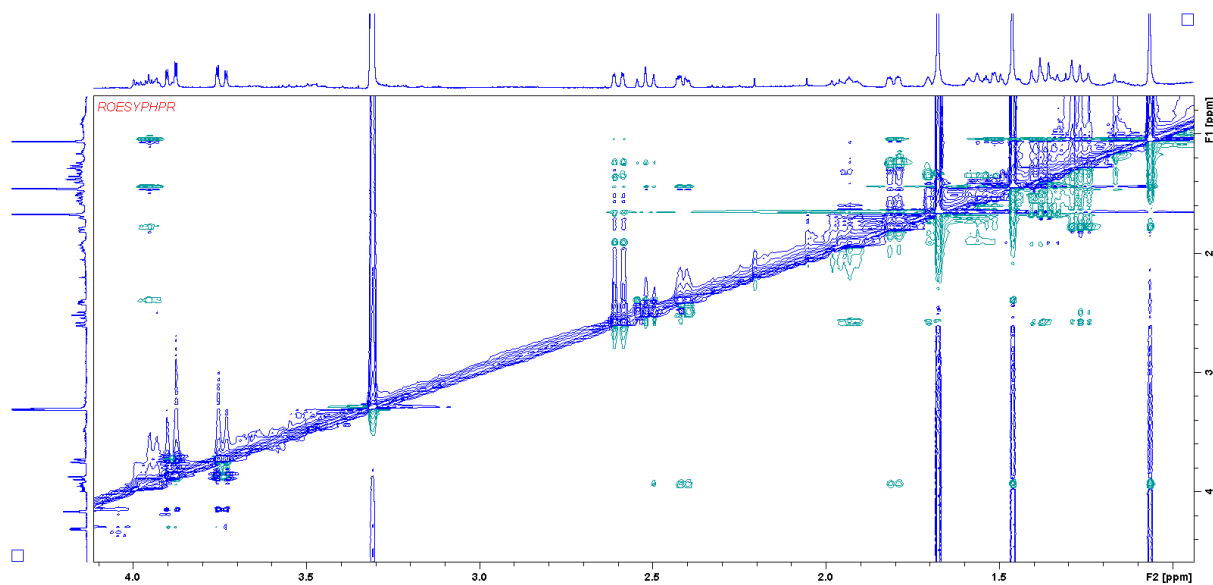
**Figure S27.** HSQC spectrum of 2-hydroxysorangiadenosine (2) in methanol-d<sub>4</sub>.



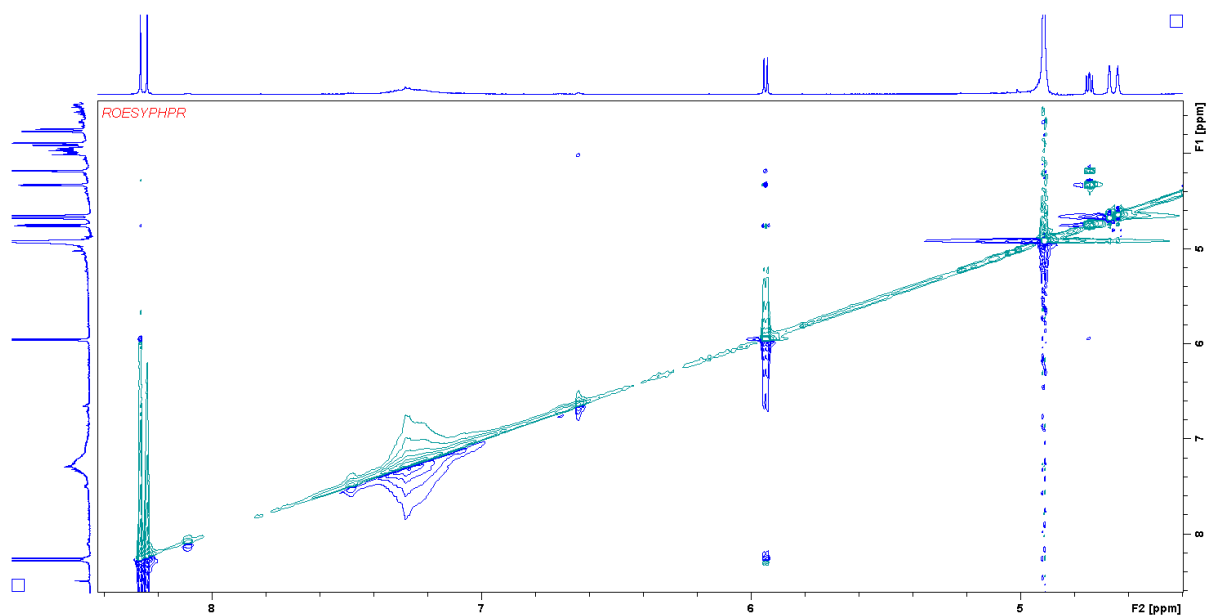
**Figure S28.** *DQF-COSY* spectrum of 2-hydroxysorangiadenosine (**2**) in methanol-*d*<sub>4</sub>.



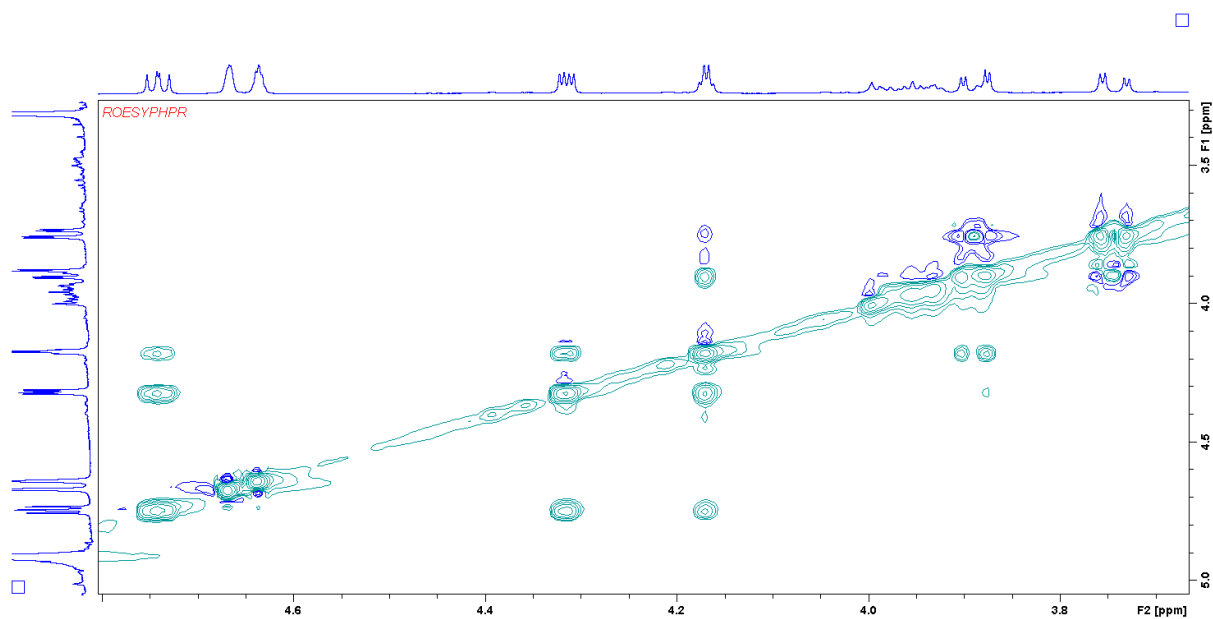
**Figure S29.** *HMBC* spectrum of 2-hydroxysorangiadenosine (**2**) in methanol-*d*<sub>4</sub>.



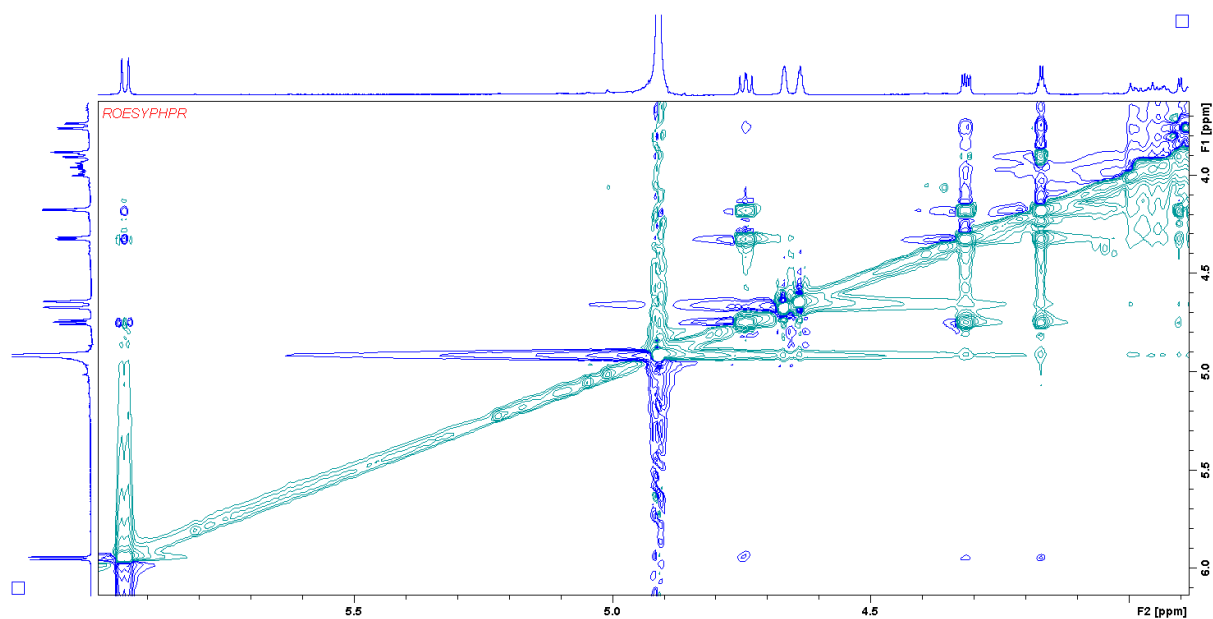
**Figure S30.** *ROESYphpr* spectrum No. 1 of 2-hydroxysorangiadenosine (**2**) in methanol- $d_4$ .



**Figure S31.** *ROESYphpr* spectrum No. 2 of 2-hydroxysorangiadenosine (**2**) in methanol- $d_4$ .



**Figure S32.** ROESYphpr spectrum 3 of 2-hydroxysorangiadenosine (**2**) in methanol-d<sub>4</sub>.



**Figure S33.** ROESYphpr spectrum 4 (expanded region) of 2-hydroxysorangiadenosine (**2**) in methanol-d<sub>4</sub>.

## 4. References

1. Bode, H. B. *et al.* Identification of additional players in the alternative biosynthesis pathway to isovaleryl-CoA in the myxobacterium *Myxococcus xanthus*. *ChemBioChem* **10**, 128–140 (2009).
2. Li, Y., Luxenburger, E. & Müller, R. An alternative isovaleryl CoA biosynthetic pathway involving a previously unknown 3-methylglutaconyl CoA decarboxylase. *Angew. Chem. Int. Ed. Engl.* **52**, 1304–1308 (2012).
3. Mahmud, T. *et al.* A novel biosynthetic pathway to isovaleryl-CoA in myxobacteria: The involvement of the mevalonate pathway. *ChemBioChem* **6**, 322–330 (2005).
4. Bock, T. *et al.* AibA/AibB induces an intramolecular decarboxylation in isovalerate biosynthesis by *Myxococcus xanthus*. *Angew. Chem. Int. Ed.* **56**, 9986–9989 (2017).
5. Dickschat, J. S. Bacterial terpene cyclases. *Nat. Prod. Rep.* **33**, 87–110 (2015).
6. Rabe, P., Rinkel, J., Klapschinski, T. A., Barra, L. & Dickschat, J. S. A method for investigating the stereochemical course of terpene cyclisations. *Org. Biomol. Chem.* **14**, 158–164 (2015).
7. Schiffrin, A. *et al.* A single terpene synthase is responsible for a wide variety of sesquiterpenes in *Sorangium cellulosum* So ce56. *Org. Biomol. Chem.* **14**, 3385–3393 (2016).



## Chapter 5

# **Structure and Biosynthesis of Myxolipoxazoles and Myxopyrimidinols: Unique Myxobacterial Fatty Acids Featuring Isoxazole or 4-Pyrimidinol Heterocycles**

Alexander Popoff<sup>†</sup>, **Joachim J. Hug<sup>†</sup>**, Sebastian Walesch, Ronald Garcia and Rolf Müller

<sup>†</sup> These authors contributed equally to this work

*Manuscript ready for submission*

### Contributions to the presented work

#### Author's contribution

The author contributed to the conception of this study, designed and performed experiments, evaluated and interpreted resulting data. The author developed and performed feeding and genetic experiments and analyzed the corresponding LC-MS data. Thereby the author identified the biosynthetic gene cluster and conceived a biosynthetic model for the myxolipoxazoles and myxopyrimidinols. Furthermore, the author contributed to conceiving and writing this manuscript.

#### Contribution by others

Alexander Popoff contributed to the conception of this study, designed and performed experiments, evaluated and interpreted resulting data. He discovered and isolated myxolipoxazole A, performed NMR-based structure elucidation and performed feeding experiments to conceive the biosynthesis. Furthermore, he contributed to conceiving and writing this manuscript. Sebastian Walesch contributed to the conception of this study, designed and performed experiments, evaluated and interpreted resulting data. He confirmed the alternative producers of myxolipoxazole A experimentally, developed a protocol to genetically manipulate Mx x48 and isolated myxopyrimidinol A. Ronald Garcia isolated, characterized and cultivated the myxobacterium *Corallococcus* sp. MCy9072. Rolf Müller contributed by supervision of the project and conceiving, editing and proofreading of this manuscript.

## 5 Structure and Biosynthesis of Myxolipoxazoles and Myxopyrimidinols: Unique Myxobacterial Fatty Acids featuring Isoxazole or 4-Pyrimidinol Heterocycles

### 5.1 Abstract

A classical metabolite-guided screening approach in *Corallococcus* sp. MCy9072 resulted in the isolation of myxolipoxazole A, a natural product containing the biosynthetic isoxazole substructure, which has been rarely described in other natural products. *In silico* analysis of the genome sequence of the alternative producer *Myxococcus xanthus* Mx x48 along with feeding experiments using stable isotope-labeled precursors led to the identification of a candidate biosynthetic gene cluster. The roles of the biosynthetic genes responsible for the formation of myxolipoxazole A were confirmed through targeted gene inactivation. Induced gene expression of the respective gene operon led to increased production and identification of putative myxolipoxazole derivatives provided by the underlying biosynthetic gene cluster, leading to the isolation of myxopyrimidinol A. We report here the full structure elucidation of myxolipoxazole A and myxopyrimidinol A, and to the best of our knowledge, the first biosynthetic investigation correlating isoxazole-containing and 4-pyrimidinol-containing natural products to their biosynthetic gene cluster.

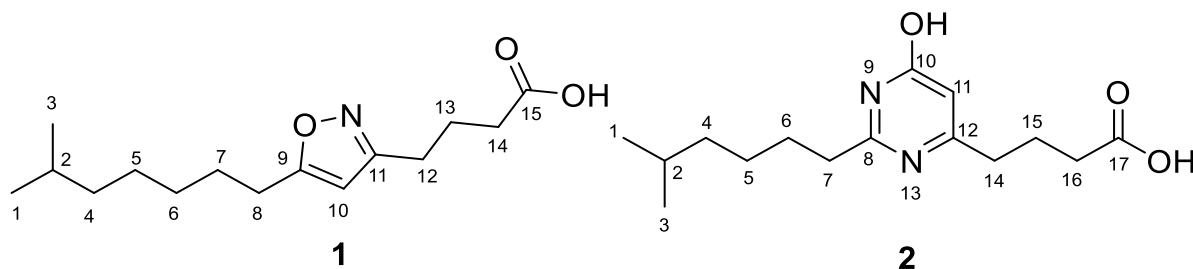
### 5.2 Introduction

Natural products from microorganisms are indispensable for drug discovery by showing unique and specific biological effects through their underlying biosynthesis incorporating a wide range of intriguing chemical scaffolds<sup>1</sup>. Many of these natural products or derivatives thereof are in clinical usage not only as antibacterial and antifungal agents, but also as immunosuppressant and cytotoxic anticancer drugs<sup>2</sup>. However, the discovery of new chemical scaffolds has become increasingly more challenging, since numerous secondary metabolites considered as the low-hanging fruits<sup>3</sup> are characterized in-depth and several prolific producers of natural products seems to be exploited<sup>4</sup>. Hence, rediscovery of known natural products is becoming a major bottleneck in natural product discovery pipelines<sup>5</sup>. Encouragingly underexploited producers of natural products are entering the stage such as myxobacteria, with the propitious prospect to provide new chemical entities<sup>6</sup>. A classical metabolite-guided screening program of the newly isolated *Corallococcus* sp. MCy9072 led to the identification and structure elucidation of myxolipoxazole A (**1**), a new isoxazole-containing natural product and its biosynthetic correlation to an unprecedented biosynthetic pathway. In addition, genetic modification of the *in silico* identified biosynthetic genes in the alternative producer *Myxococcus xanthus* Mx x48 led to the discovery of myxolipoxazole derivatives provided by the underlying biosynthetic pathway and culminated in the isolation of myxopyrimidinol A (**2**), featuring a different biosynthetic heterocyclic scaffold than **1**, namely a 4-pyrimidinol (**Figure 1**).

### 5.3 Results and Discussion

The *Corallococcus* sp. strain MCy9072 was isolated from a Chinese soil sample taken in 1980 from the HZI soil collection (formerly called German Center for Biotechnology (GBF)). Initial analysis of the secondary metabolome of the crude extract of *Corallococcus* sp. MCy9072 was based on liquid chromatography (LC) coupled with high-resolution mass spectrometry (*hrMS*). The in-house LC-MS metabolome database termed Myxobase<sup>7</sup>, which contains assigned chromatographic data of myxobacterial natural products generated by related myxobacterial strains, enabled the identification of an uncommon secondary metabolite in the crude extract of *Corallococcus* sp. MCy9072. The identified target metabolite displayed an ion peak at  $m/z = 268.1919$   $[M+H]^+$  with the retention time of 11.80 min. The molecular formula  $C_{15}H_{25}N_1O_3$  was calculated for the single charged ion ( $[M+H]^+_{calc}$ : 268.1901;  $[M+H]^+_{meads}$ : 268.1919  $\Delta=4.47$  ppm) with 3.5 degrees of unsaturation. Large-scale fermentation of *Corallococcus* sp. MCy9072 permitted the isolation of **1** in sufficient amounts for full structure elucidation (experimental condition for production, see **Supporting Information 1.2** and **1.3**).

The  $^1H$  spectrum of **1** in  $CD_3OD$  exhibited one signal at  $\delta_{H-10}$  6.08, characteristic for a proton in an aromatic system. Furthermore, four methylene signals accounting for eight protons were observed in the spectrum at  $\delta_{H2-14}$  2.23,  $\delta_{H2-13}$  1.92,  $\delta_{H2-7}$  1.69 and  $\delta_{H2-4}$  1.19 with resonances typically found in alkyl chains. In addition, one methine group at  $\delta_{H-2}$  1.53 was found and one intense signal at  $\delta_{H3-1/3}$  0.88 indicated the presence of two methyl groups. Examination of the heteronuclear single quantum coherence (HSQC) spectrum revealed the presence of four additional methylene groups ( $\delta_{C-12}$  26.7,  $\delta_{H2-13}$  2.64,  $\delta_{C-8}$  27.5,  $\delta_{H2-8}$  2.72,  $\delta_{C-6}$  30.4,  $\delta_{H2-6}$  1.36 and  $\delta_{C-5}$  28.1,  $\delta_{H2-5}$  1.35). Examination of correlation spectroscopy (COSY) cross peaks indicated that both methyl groups,  $H_3-1$  and  $H_3-3$ , are linked via the methine group at position two. Further COSY correlations from the methine proton H-2 showed that both, the methyl-protons  $H_3-1/3$  and H-2 form a spin system consisting of five additional methylene protons ( $H_2-4$ ,  $H_2-5$ ,  $H_2-6$ ,  $H_2-7$  and  $H_2-8$ ), indicated presence of an *iso*-branched alkyl chain. In addition, a second shorter alkyl chain which is comprised of three methylene protons ( $H_2-12$ ,  $H_2-13$  and  $H_2-14$ ) was identified by analysis of COSY cross peaks. As indicated by heteronuclear multiple bond correlation (spectroscopy) (HMBC) correlations from  $H_2-13$  and  $H_2-14$  to a characteristic carbon resonance at  $\delta_{C-15}$  181.4, this alkyl group is terminated by a carboxyl group located next to position 14. In order to establish connectivity between both alkyl chains, the HMBC spectrum was examined. The methylene group at  $\delta_{H2-8}$  2.72 showed HMBC correlations to a carbon signal at  $\delta_{C-9}$  174.1 and to the carbon signal of an aromatic methine group ( $\delta_{H-10}$  6.08;  $\delta_{C-10}$  101.9). From H-10, HMBC correlations to a carbon resonance at 163.1 ppm ( $\delta_{C-11}$ ) were observed. In addition,  $H_2-12$  ( $\delta_H$  2.64) showed HMBC correlations to C-11 at  $\delta_C$  163.1, suggesting both alkyl chains are connected by a heteroaromatic system. The shift values of C-9 ( $\delta_C$  174.1), C-10 ( $\delta_C$  101.9) and C-11 ( $\delta_C$  163.1) strongly implied that these carbons are part of an oxazole-like heteroaromatic system, but the comparison with known examples revealed major deviations of the carbon resonances from oxazole heterocycles<sup>8</sup>. Instead, the carbon resonances resemble those of isoxazole type heterocycles<sup>9</sup>. In addition, the determined  $^{15}N$  chemical shift of 361 ppm ( $\delta_{N-C11}$ ) is more characteristic for an isoxazole-type heterocycle with reported chemical shift values ranging from 377 to 388 ppm<sup>10</sup> than for an oxazole-type heterocycle where  $^{15}N$  chemical shifts values are more upfield shifted to values between 222 and 250 ppm<sup>11</sup>. The results obtained by Fourier-transform infrared spectroscopy (FT-IR) supported the presence of an isoxazole heterocycle manifested as strong N–O stretch at  $1567\text{ cm}^{-1}$  (**Supporting Information, Figure S9** and **S10**).

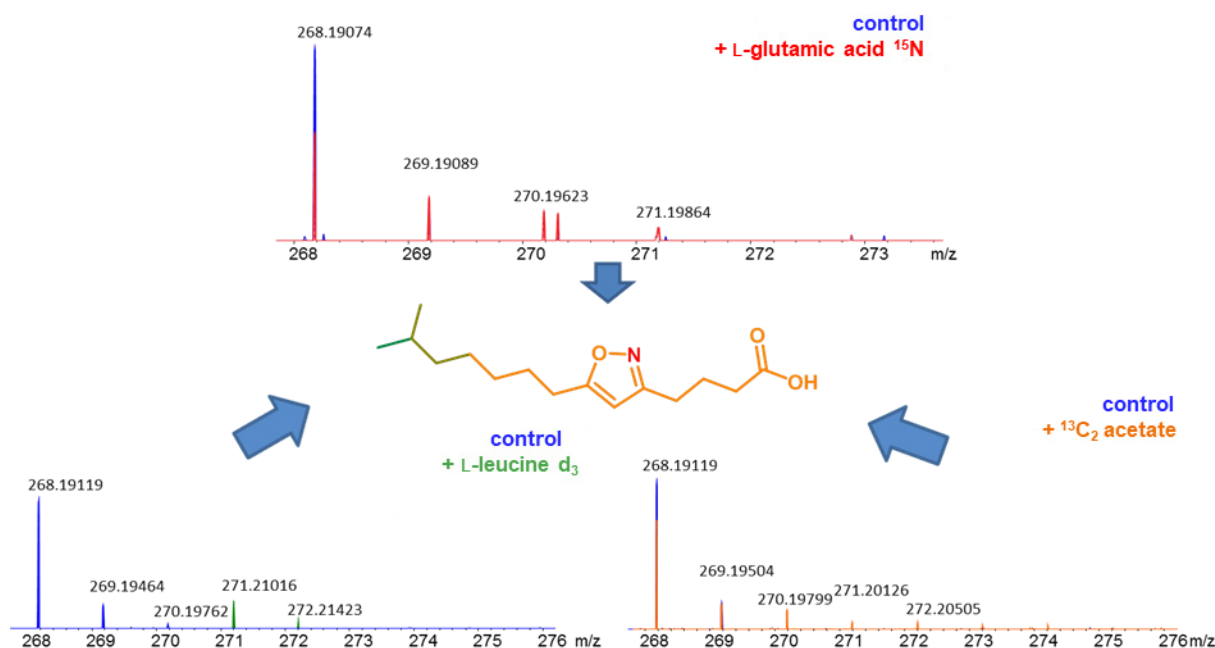


**Figure 1.** Chemical structure of myxolipoxazole A (**1**) and myxopyrimidinol A (**2**).

Taken together, these analyses resulted in the unambiguous assignment of the overall structure of **1** (**Figure 1**). The displayed isoxazole heterocycle is a chemical feature rarely found in natural products.

Among the few natural products featuring either an isoxazoline or isoxazolol scaffold (which are more common than the isoxazole scaffold)<sup>12</sup>, only the well-known cycloserins<sup>13,14</sup> from *Streptomyces* sp., acivicin and 4-hydroxyacivicin from *Streptomyces sviveus*<sup>15</sup>, muscimol and ibotenic acid from *Amanita muscaria*<sup>16</sup> feature an isoxazole moiety. The biosynthesis of ibotenic acid is proposed to start from 3-hydroxy-glutamate<sup>17</sup>. Identification and recombinant production of the glutamate hydroxylase IboH from *A. muscaria* provided preliminary support for this hypothesis. The gene *iboH* encoding the glutamate hydroxylase is surrounded by six other genes which are proposed to be involved in the biosynthesis of ibotenic acid<sup>18</sup>. In contrast to *A. muscaria*, the bacterial producer of acivicin and 4-hydroxyacivicin named *S. sviveus* and its genome sequence are publicly available, which encouraged us to experimentally correlate the biosynthetic genes responsible for acivicin production. However, it was not possible to detect any acivicin or 4-hydroxyacivicin in the secondary metabolome of *S. sviveus* (**Supporting Information, Figure S3**).

In order to obtain further insights concerning the biosynthesis of **1** and in particular the formation of its isoxazole heterocycle, feeding experiments with stable isotope-labeled biosynthetic building blocks were conducted. The observed mass shift of +3 Da in the isotopic pattern of **1** upon L-leucine *d*<sub>3</sub> supplementation suggest that the isovaleryl scaffold originates from L-leucine degradation, which has been described previously among the order of *Myxococcales*<sup>19</sup>. In contrast, supplementation with L-valine *d*<sub>8</sub> did not result in any observable mass shift in the isotopic pattern of **1**. The nitrogen of the isoxazole heterocycle might be incorporated through an aminotransferase-catalyzed reaction. A corresponding mass shift of +1 Da upon incorporation of L-glutamic acid-<sup>15</sup>N, supports this hypothesis. In addition, acetate-<sup>13</sup>C<sub>2</sub> supplementation of the bacterial culture broth displayed incorporation of acetate into the fatty acid backbone (**Figure 2**).



**Figure 2.** Feeding experiments indicate, that the isovaleryl scaffold might originate from L-leucine, the carbon-hydrogen skeleton from acetate and the nitrogen in the heterocycle from a glutamine transferase reaction, respectively.

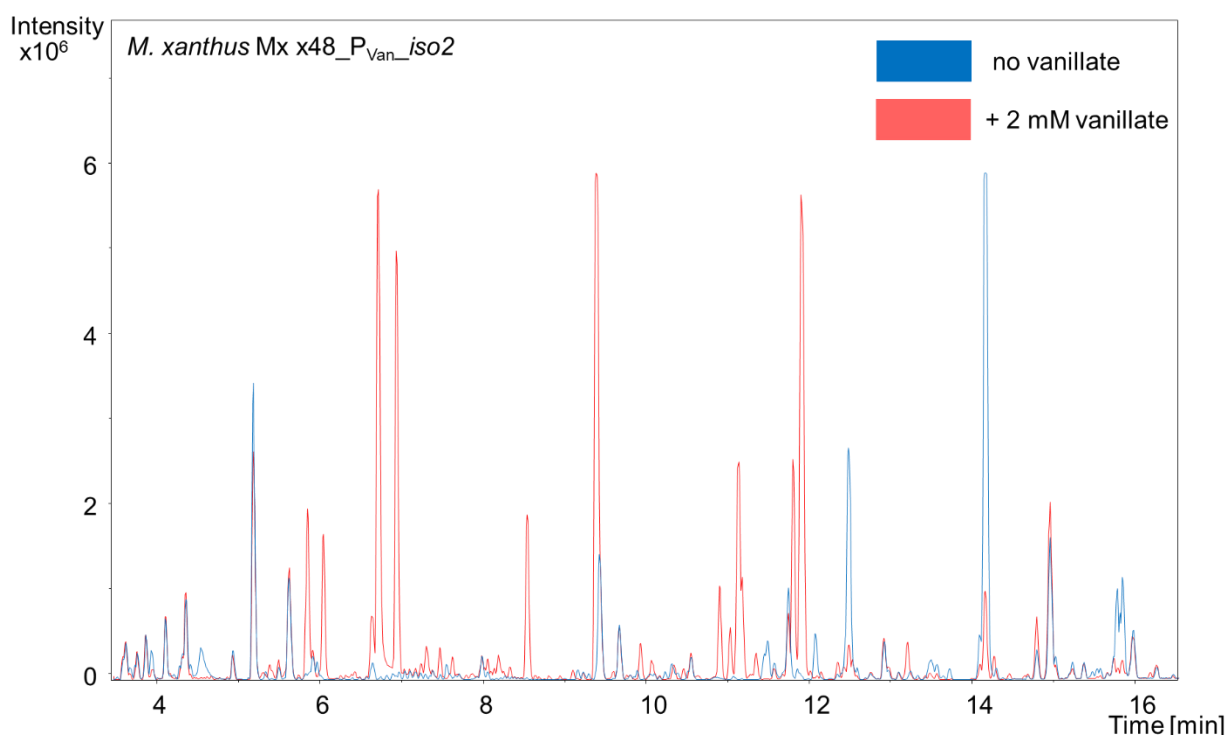
Based on the elucidated chemical structure of **1** along with the findings from the feeding experiments, we propose **1** to be biosynthesized by a modular polyketide pathway, which features at least one aminotransferase. Since no genome sequence of the strain *Corallococcus* sp. MCy9072 was available, the Myxobase was investigated for alternative producers with available genome sequence. This survey revealed that *Myxococcus* Mx x48, *Stigmatella aurantiaca* Sg a32 and *Stigmatella erecta* MCy6787 (Pd e42) are alternative producers of **1** (**Supporting Information 1.1**). AntiSMASH analysis of all three genome sequences led to the identification of a candidate PKS/NRPS hybrid gene cluster putatively responsible for the biosynthesis of **1**. Investigation of the publicly available genome sequences revealed four myxobacterial strains harboring gene clusters with high similarity to the identified candidate PKS/NRPS hybrid gene cluster (**Supporting Information 2.5**).

The putative biosynthetic gene cluster contains nine open reading frames (*iso1–iso9*) (**Figure 4A**). The NRPS-PKS hybrid cluster contains four modules; three of these are PKS modules (Iso2, Iso5, Iso9), whereas one module is NRPS derived (Iso6). In comparison to the *Stigmatella* strains, *M. xanthus* Mx x48<sup>20</sup> offers beneficial growth characteristics and is phylogenetically closely related to the myxobacterial model strain *Myxococcus xanthus* DK1622<sup>21,22</sup>. For that reason, *M. xanthus* Mx x48 was selected to establish a genetic manipulation procedure for gene disruption experiments on the proposed biosynthetic gene cluster.

The role of the biosynthetic genes *iso1–6* and *iso8–9* was confirmed by targeted single crossover based gene inactivation (**Supporting Information 2.6**) leading to loss of production. This result connected the *in silico* identified biosynthetic gene cluster to **1** and additional comparative metabolome analyses revealed further secondary metabolites produced by this biosynthetic pathway. Most of these revealed

secondary metabolites, which are probably derivatives of **1**, featuring the generic sum formula  $C_{(12+w)}H_{(20+x)}N_{(1+y)}O_{(1+z)}$  ( $w=0-9$ ,  $x=0-9$ ,  $y=0-1$ ,  $z=0-3$ ).

In addition, the vanillate-inducible promoter system, which has been used in numerous studies describing genetic engineering of myxobacterial pathways<sup>23</sup>, was inserted via single crossover recombination in front of *iso2* in *M. xanthus* Mx x48 to control the genetic operon *iso2-9*. In particular, the previously poor and fluctuating production of most of the associated derivatives of **1** was stabilized and increased upon this genetic engineering approach (**Figure 3** and **Figure S22**). This will pave the way for the isolation and structure elucidation of the remaining secondary metabolites afforded by the *iso* gene cluster.



**Figure 3.** High-performance liquid chromatography-mass spectrometry base peak chromatogram (HPLC-MS BPC) of *M. xanthus* Mx x48 P<sub>van</sub>\_iso2 with and without supplementation of 2 mM vanillate. The chromatogram reveals tremendously increased production of many secondary metabolites associated with the *iso* BGC.

Secondary metabolome analysis of *M. xanthus* Mx x48, characterizing all corresponding peaks of previously identified myxobacterial compounds (**Supporting Information 2.1**) combined with the identified secondary metabolites from the gene disruption experiments led to the identification of a new target metabolite, displaying an ion peak at  $m/z = 281.1851$   $[M+H]^+$  with the retention time of 6.93 min. The molecular formula  $C_{15}H_{25}N_2O_3$  was calculated for the single charged ion ( $[M+H]^+$ <sub>calc</sub>: 281.1859;  $[M+H]^+$ <sub>meas</sub>: 281.1851,  $\Delta = 2.94$  ppm) with 4.5 degrees of unsaturation. Large-scale fermentation of *M. xanthus* Mx x48 permitted the isolation of **2** in sufficient amounts for full structure elucidation.

The collected NMR data of **2** was acquired in DMSO  $d_6$  due to poor solubility in  $CDCl_3$ , but strongly resembled the spectrum of **1** in  $CD_3OD$ . The  $^1H$ -spectrum exhibited one signal at  $\delta_{H-11}$  5.92, characteristic for a  $sp^2$ -hybridized methine proton. In addition, the spectrum showed five methylene protons at  $\delta$  2.38 ( $H_2$ -14),  $\delta$  2.09 ( $H_2$ -16),  $\delta$  1.75 ( $H_2$ -15),  $\delta$  1.60 ( $H_2$ -6) and  $\delta$  1.15 ( $H_2$ -4). Furthermore, one signal with a characteristic shift value for methyl protons was identified at  $\delta_{H3-1/3}$  0.83. Integration of this signal implied the presence of two methyl groups. Inspection of the HSQC spectrum showed one additional methylene group ( $\delta_{C-7}$  34.2,  $\delta_{H2-7}$  2.48) and one methine group ( $\delta_{C-2}$  27.0,  $\delta_{H-2}$  1.49). Examination of COSY cross peaks revealed presence of an *iso*-branched alkyl chain like in **1**, consisting of two methyl protons, ( $H_3$ -1/3) one methine proton ( $H$ -2) and four methylene protons ( $H_2$ -4,  $H_2$ -5,  $H_2$ -6 and  $H_2$ -7). In addition, a short carboxylic acid chain, similar to the chain described in **1**, was identified by analysis of COSY and HMBC correlations. The chain consists of three methylene groups ( $H_2$ -14,  $H_2$ -15 and  $H_2$ -16) and one carboxyl group C-17 ( $\delta_C$  175.1).

In order to elucidate connectivity between the short fatty acid chain and the *iso*-branched alkyl chain, HMBC correlations were analyzed. From  $\delta_{H2-7}$  2.48 and  $\delta_{H2-6}$  1.60 in the *iso*-branched alkyl chain, HMBC correlations to a carbon resonance at  $\delta_{C-8}$  161.7 were observed. This shift value indicated a  $sp^2$ -hybridized carbon in a heteroaromatic system. The methylene group  $H_2$ -14 ( $\delta_H$  2.38) in the carboxylic acid chain shows HMBC correlations to a carbon resonance at  $\delta_{C-12}$  166.8 and an aromatic methine group ( $\delta_{H-11}$  5.92;  $\delta_{C-11}$  109.1). The methine proton  $H$ -11 exhibited further HMBC correlations to carbon resonances at  $\delta_{C-10}$  162.9,  $\delta_{C-12}$  167.2 and  $\delta_{C-14}$  35.8 together with a weak HMBC correlation to  $\delta_{C-7}$  34.2. These HMBC correlations and the obtained sum formula indicated that both, the *iso*-branched alkyl chain and the carboxylic acid, are linked by a pyrimidinol-type heterocycle with the hydroxyl group at ring position 2, 4 or 5 (2-pyrimidinol, 4-pyrimidinol and 5-pyrimidinol, respectively, **Supporting Information, Figure S8**). The absence of HMBC correlations between the *iso*-branched alkyl chain and carboxylic acid chain indicated that chains are attached in either *meta* or *para* position to the pyrimidinol ring.

A simulation of carbon resonances to generate  $^{13}C$  NMR spectra of hypothetical myxopyrimidinols featuring a 2-pyrimidinol, 4-pyrimidinol or 5-pyrimidinol heterocycle, showed that a pyrimidinol ring with the hydroxyl group at ring position 2 or 4 is in best agreement to experimentally obtained carbon resonances (**Supporting Information, Table S13**). Considering the observed HMBC correlations, a 4-pyrimidinol ring is the most plausible structure. The *iso*-branched alkyl chain at ring-position 2 showed only HMBC correlations (from  $\delta_{H2-7}$  2.48 and  $\delta_{H2-6}$  1.60) to one ring carbon ( $\delta_{C-8}$  161.7), while the methylene protons  $H_2$ -14 ( $\delta$  2.38) of the carboxylic acid chain at ring-position 6 showed correlations to two ring carbons ( $\delta_{C-11}$  109.1 and  $\delta_{C-12}$  167.2). Uncharacteristic  $^5J_{(C,H)}$  correlations like the observed weak correlation from  $H$ -11 ( $\delta$  5.92) to  $C$ -7 ( $\delta$  34.2) occasionally appear in standard HMBC spectra<sup>24</sup>. In addition, a proton and a carbon spectrum of the commercially available standard 2-isopropyl-6-methyl-4-pyrimidinol were acquired and the carbon signals of the pyrimidinol ring were compared to the signals of **2**. The signals were differing less than 2 ppm (**Supporting Information, Table S12, Figure S6 and S7**), a fact that additionally supports the proposed structure of **2** (**Figure 1**).

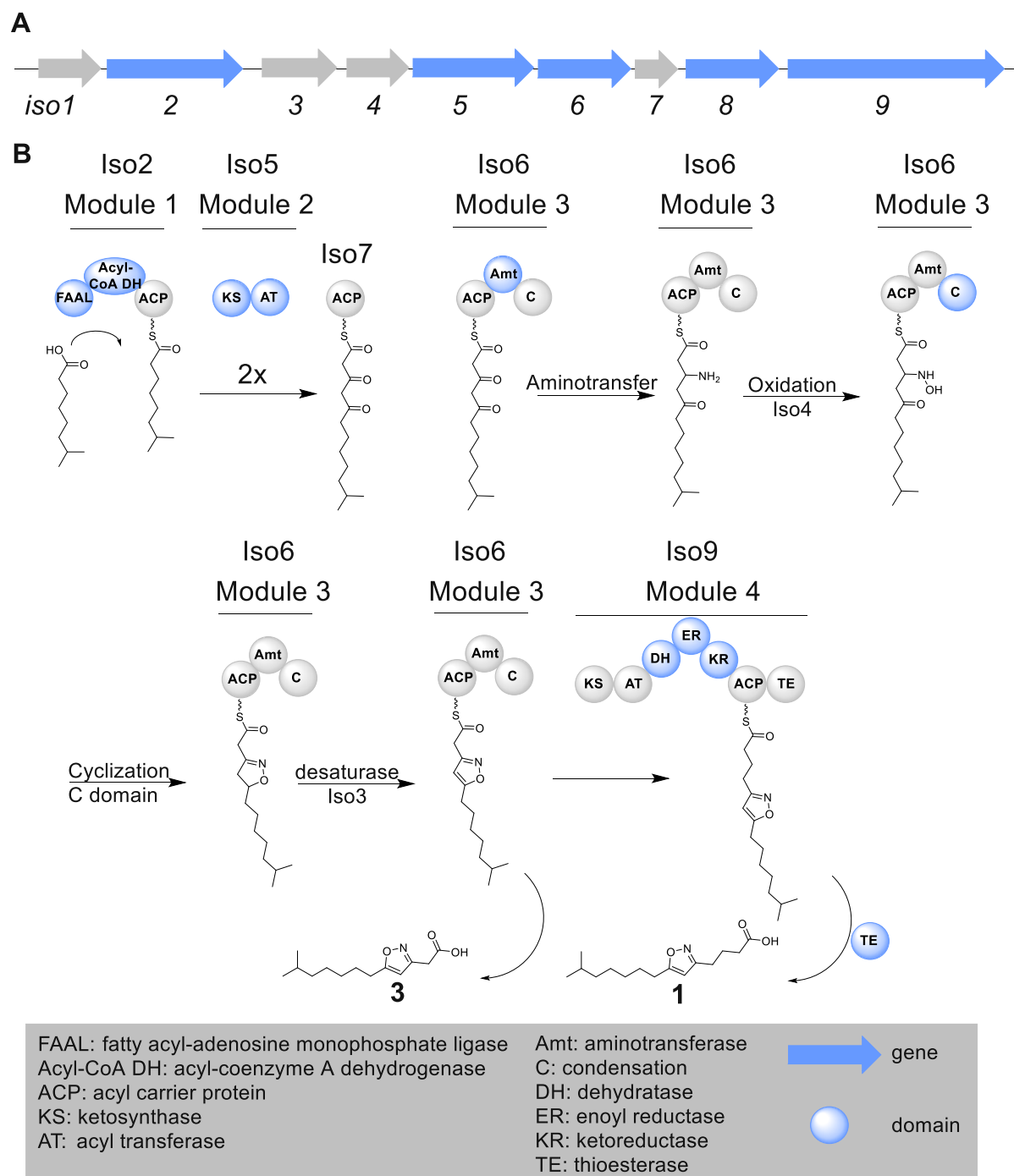
The different heterocycles featured in **1** and **2** underlines the inherent interest on how these chemical scaffolds are formed through the biosynthetic proteins encoded by the identified PKS-NRPS hybrid gene cluster *iso1-iso9*. Biosynthesis of **1** starts with the activation of 7-methyloctanic acid to 7-methyloctanic acid-adenosine monophosphate (AMP) by the fatty acyl-AMP ligase (FAAL) domain on Iso2 (**Figure 4B**). The conducted feeding experiment with L-leucine  $d_3$  provides experimental support for the proposed starter unit, since degradation of L-leucine and further conversion can lead to



the generation of *iso*-odd fatty acids and lipids<sup>25</sup>. The first chain extension occurs with the condensation of 7-methyloctanic acid with malonate, catalyzed by the PKS module Iso5. The standalone acyl carrier protein (ACP) Iso7 acts in this case as the swinging arm for the polyketide backbone. Iterative use of module 2 logically explains the incorporation of two ketide units onto the growing polyketide backbone, which provides the fundamental requirements for the subsequent cyclization on module 3. Iso2 harbors a putatively functional acyl-coenzyme A dehydrogenase (acyl-CoA DH) which seems to be required during the formation of **2** (see below). Therefore, an alternative route leading to **1** would start from 9-methyldecanoic acid, which undergoes AMP-based activation and subsequent  $\beta$ -oxidation of the fatty acid to yield (*E*)-9-methyldec-3-enoic acid. Single use of module 2 would explain the incorporation of another keto unit onto the growing polyketide backbone, which provides together with the *trans* double bond the requirements for the subsequent cyclization on module 3.

The next NRPS module accepts the acyl chain on Iso7 without backbone elongation and catalyzes an aminotransferase reaction on the second carbonyl of the tethered backbone. The generated amino group is afterwards hydroxylated by Iso4, and subsequent cyclization of the tethered backbone leads to a proposed isoxazoline intermediate. The condensation (C) domain from the NRPS module Iso6 is phylogenetically closely related to heterocyclization domains (see **Supporting Information, Figure S19**) for phylogenetic analysis of characterized C domains) such as described for the biosynthesis of epothilone and myxothiazol<sup>23</sup>, and might perform the cyclization of the generated oximes to yield the isoxazoline scaffold and the 4-pyrimidinol scaffold (see below). This isoxazoline intermediate might then be oxidized by the fatty acid desaturase Iso3 to yield the oxidized derivative myxolipoxazole B (**3**) as the growing backbone. The last step of the myxolipoxazole biosynthesis involves loading of a malonyl-CoA extender unit onto the scaffold of intermediate **3** onto the PKS module Iso9. In accordance to the reduction state observed in **1**, module 4 performs a full reduction loop of the loaded extender unit. The linear NRPS-PKS chain is released by the action of the thioesterase (TE) domain on module 4.

Gene disruption of the last PKS module Iso9 resulted in the abolishment of **1**, and **2** and the associated identified metabolites (**Supporting Information Table S18**). Interestingly the significant increase of several metabolites corresponding to derivatives reduced by C<sub>2</sub>H<sub>4</sub> logically supports the proposed function of Iso9. Accordingly, the proposed biosynthetic pathway for the myxolipoxazoles combined with the observed MS<sup>2</sup> fragmentation pattern of the identified metabolites enabled the clear assignment of the direct precursor **3** (**Supporting Information, Figure S23** and **S24**).



**Figure 4.** A) Genetic operon *iso1–9* and proposed biosynthetic route leading to myxolipoxazole A (**1**) and B (**3**) (B). The catalytic domains for each biosynthetic step are colored in blue.

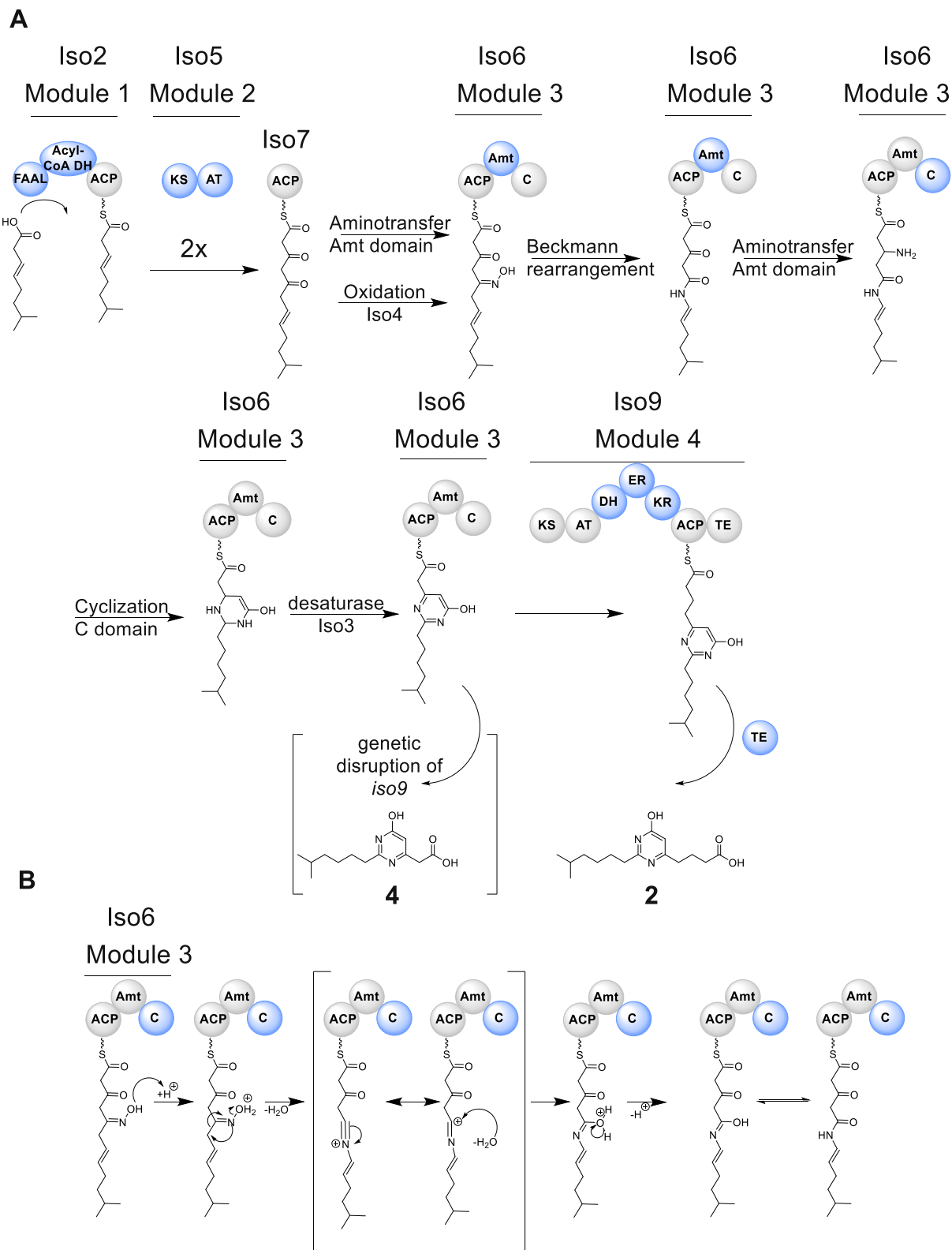
The biosynthesis of **2** starts similar to **1** with the activation and compulsory  $\beta$ -oxidation of a fatty acid through the FAAL and acyl-CoA DH domain leading to (*E*)-7-methyloct-3-enoic acid attached on the ACP of Iso2 (**Figure 5A**). Further processing resembles the biosynthesis of **1**, with different catalysis on the NRPS module 3. The NRPS module 3 catalyzes the aminotransferase reaction on the third carbonyl of the tethered backbone and hydroxylates this amino group afterwards.

The generated oxime is proposed to undergo a Beckmann rearrangement catalyzed by the C domain on Iso6 to yield an amide (**Figure 5B**); such rearrangement reactions are proposed for the biosynthesis of marine oxazole-containing natural products<sup>26</sup>. A second aminotransferase reaction on the second carbonyl of the tethered backbone combined with the nucleophilic attack of this generated amino group to the diene would lead to a cyclized tetrahydropyrimidinol scaffold on the growing backbone. Subsequent oxidation of the fatty acid desaturase Iso3 would yield the oxidized derivative myxopyrimidinol B (**4**).

The last step is identical to the myxolipoxazole biosynthesis, by loading a malonyl-CoA extender unit onto the scaffold of intermediate **4** onto the PKS module Iso9. The linear NRPS-PKS chain is released by the action of the TE domain on module 4 to yield **2**. Since gene disruption of *iso9* still allows the production of the derivative **3** and reveals the intermediate **4** (which could not be found in the crude extracts of *M. xanthus* Mx x48 wild type or the genetically engineered vanillate-induced producer, see **Supporting Information, Figure S22**) this finding underlines that the proposed biosynthesis works independently from module 4. The precursor **4** was structurally assigned through observed MS<sup>2</sup> fragmentation pattern compared with the one of **2** and the findings provided through genetic and biosynthetic investigations (**Supporting Information, Figure S25 and S26**).

The proposed biosynthetic pathway consequentially explains the order of biosynthetic steps to generate the isoxazole moiety in **1** and **3** and the 4-pyrimidinol scaffold in **2** and **4**; however it cannot be excluded that the steps differ from the proposed biosynthetic pathway, since the characterization of these biosynthetic formations are unprecedented and not proven stepwise.

Compound **1** showed weak antifungal and cytotoxic activity towards *Mucor hiemalis* DSM 2656 (MIC: 32  $\mu$ g/mL) and *Cellosaurus* KB-3-1 cell line (MIC: 24.06  $\mu$ g/mL). The intriguing structural feature of an isoxazole heterocycle in **1** is described for chemical crop protection agents<sup>27</sup>, whereas the 4-pyrimidinol scaffold in **2** has been described in several tankyrase inhibitors<sup>28,29</sup>. These implications foster further evaluation of the biological function of the myxolipoxazoles and myxopyrimidinols.



**Figure 5.** A) Proposed biosynthetic route leading to myxopyrimidinol A (**2**) and B (**4**). B) Proposed mechanism of Beckmann rearrangement during myxopyrimidinol biosynthesis on module 3. The catalytic domains for each biosynthetic step are colored in blue.

## 5.4 Conclusions

The findings summarized in this study highlight myxobacteria as producers of myxolipoxazoles and myxopyrimidinols, two new natural product classes featuring heterocycles rarely found in biosynthetic pathways. We identified the myxolipoxazoles as new secondary metabolites with unique chemical structure and report genetic investigation of the biosynthesis of an isoxazole-containing natural product. In contrast to oxazole heterocycles, the presence of an isoxazole chemical scaffold is exceptional for natural products and the underlying biosynthesis of isoxazole-containing natural products remained in large measures elusive. The biosynthetic logic of the unusual PKS-NRPS hybrid machinery in *M. xanthus* Mx x48 concerning the formation of the isoxazole scaffold in myxolipoxazole significantly differs from the proposed pathways of acivicin, ibotenic acid or D-cycloserine. Further in-depth investigation of the biosynthetic pathway leading to the formation of the myxolipoxazoles, myxopyrimidinols and their derivatives should be considered to fully characterize this unique biosynthetic pathway. Heterologous expression of the identified genetic operon *iso1–iso9* in the myxobacterial model host *M. xanthus* DK1622 would allow profound genetic engineering, such as seamless gene deletion of each gene (*iso1–iso9*) to investigate the effects on the production of corresponding secondary metabolites. In combination with recombinantly produced key enzymes involved in the biosynthesis of myxolipoxazoles and myxopyrimidinols, an *in vitro* system could be established to reveal the biosynthetic logic of this fascinating pathway. In closing, the identification of two different structural classes of myxobacterial natural products in *M. xanthus* Mx x48 exemplifies the notable biosynthetic capabilities of myxobacteria.

## 5.5 Experimental Section

Experimental procedures, spectral data, gene cluster analysis and copies of  $^1\text{H}$ ,  $^{13}\text{C}$  NMR spectra, LC-MS and HPLC chromatograms are available in the **Supporting Information**.

## 5.6 Acknowledgements

The authors thank Dr. Lena Keller for her assistance in structure elucidation of myxopyrimidinol A and myxolipoxazole A, Dr. Bastien Schnell, Dr. Fabian Panter and Dr. Chengzhang Fu for biosynthetic discussions, Stefanie Schmidt, Irene Kochems, and Victoria Schmitt for performing bioactivity assays. We are grateful to Dr. Nestor Zaburannyi for bioinformatics support.

Joachim J. Hug acknowledges funding by a PhD fellowship of the Boehringer Ingelheim Fonds. Research in Rolf Müller's laboratory is funded by the Deutsche Forschungsgemeinschaft (DFG), the Bundesministerium für Bildung und Forschung (BMBF) and the Deutsches Zentrum für Infektionsforschung Standort Hannover-Braunschweig.

## 5.7 Conflict of Interest

The authors declare no conflict of interest

## 5.8 Key words

Myxobacteria, PKS-NRPS hybrid, Biosynthesis, Natural Products, Secondary Metabolites, Isoxazole, Fatty Acids, 4-Pyrimidinol, Heterocycles.

## 5.9 References

1. Chen, Y., Garcia de Lomana, M., Friedrich, N.-O. & Kirchmair, J. Characterization of the Chemical Space of Known and Readily Obtainable Natural Products. *J. Chem. Inf. Model.* **58**, 1518–1532 (2018).
2. Newman, D. J. & Cragg, G. M. Natural Products as Sources of New Drugs over the Nearly Four Decades from 01/1981 to 09/2019. *J. Nat. Prod.* **83**, 770–803 (2020).
3. Flemming, A. Antibacterials: Resistance-guided discovery of new antibiotics. *Nat. Rev. Chem.* **12**, 826 (2013).
4. Pye, C. R., Bertin, M. J., Lokey, R. S., Gerwick, W. H. & Linington, R. G. Retrospective analysis of natural products provides insights for future discovery trends. *Proc. Natl. Acad. Sci. U.S.A.* **114**, 5601–5606 (2017).
5. Hug, J. J., Bader, C. D., Remškar, M., Cirnski, K. & Müller, R. Concepts and Methods to Access Novel Antibiotics from Actinomycetes. *Antibiotics* **7**, 44 (2018).
6. Skinnider, M. A. & Magarvey, N. A. Statistical reanalysis of natural products reveals increasing chemical diversity. *Proc. Natl. Acad. Sci. U.S.A.* **114**, E6271-E6272 (2017).
7. Hoffmann, T. *et al.* Correlating chemical diversity with taxonomic distance for discovery of natural products in myxobacteria. *Nat. Commun.* **9**, 803 (2018).
8. Hiemstra, H., Houwing, H. A., Possel, O. & van Leusen, A. M. Carbon-13 nuclear magnetic resonance spectra of oxazoles. *Can. J. Chem.* **57**, 3168–3170 (1979).
9. Pinto, D. C. G. A., Santos, C. M. M. & Silva, A.M.S. in *Recent research developments in heterocyclic chemistry 2007*, pp. 397–475.
10. Thorn, K. A. <sup>13</sup>C and <sup>15</sup>N NMR identification of product compound classes from aqueous and solid phase photodegradation of 2,4,6-trinitrotoluene. *PLoS ONE* **14**, e0224112 (2019).

11. Kalyon, B. *et al.* Plantazolicin A and B. Structure elucidation of ribosomally synthesized thiazole/oxazole peptides from *Bacillus amyloliquefaciens* FZB42. *Org. Lett.* **13**, 2996–2999 (2011).
12. Rahbæk, L. & Christophersen, C. in *The alkaloids*, edited by K. R. Alper, *et al.* (Elsevier\_1983–1989), pp. 185–233.
13. Kumagai, T. *et al.* Molecular cloning and heterologous expression of a biosynthetic gene cluster for the antitubercular agent D-cycloserine produced by *Streptomyces lavendulae*. *Antimicrob. Agents Chemother.* **54**, 1132–1139 (2010).
14. Svensson, M.-L. & Gatenbeck, S. The pathway of D-cycloserine biosynthesis in *Streptomyces garyphalus*. *Arch. Microbiol.* **131**, 129–131 (1982).
15. Gould, S. J. & Ju, S. Biosynthesis of Acivicin. 3. Incorporation of Ornithine and  $N^{\delta}$ -Hydroxyornithine. *J. Am. Chem. Soc.* **114**, 10166–10172 (1992).
16. Bowden, K., Drysdale, A. C. & Moge, G. A. Constituents of *Amanita muscaria*. *Nature* **206**, 1359–1360 (1965).
17. Matsumoto, T., Trueb, W., Gwinner, R. & Eugster, C. H. Isolierung von (–)-R-4-Hydroxypyrrolidon-(2) und einigen weiteren Verbindungen aus *Amanita muscaria*. 31. Mitteilung ber Inhaltsstoffe von Fliegenpilzen. *HCA* **52**, 716–720 (1969).
18. Müller, M. & Obermaier, S. Ibotenic Acid Biosynthesis in the Fly Agaric Is Initiated by Glutamate Hydroxylation. *Angew. Chem. Int. Ed.* (2020).
19. Bode, H. B. *et al.* Identification of additional players in the alternative biosynthesis pathway to isovaleryl-CoA in the myxobacterium *Myxococcus xanthus*. *ChemBioChem* **10**, 128–140 (2009).
20. Freese, A., Reichenbach, H. & Lünsdorf, H. Further characterization and in situ localization of chain-like aggregates of the gliding bacteria *Myxococcus fulvus* and *Myxococcus xanthus*. *J. Bacteriol.* **179**, 1246–1252 (1997).
21. Krug, D., Garcia, R. & Müller, R. Myxobakterielle Naturstofffabriken. *Biospektrum* **26**, 32–36 (2020).
22. Yang, Y.-J. *et al.* Genome editing in model strain *Myxococcus xanthus* DK1622 by a Site-Specific Cre/loxP Recombination System. *Biomolecules* **8** (2018).
23. Hug, J. J. & Müller, R. Host Development for Heterologous Expression and Biosynthetic Studies of Myxobacterial Natural Products. *Comprehensive Natural Products III: Chemistry and Biology, Chapter 14818* **In press** (2020).
24. Araya-Maturana, R., Gavín-Sazatornil, J. A., Heredia-Moya, J., Pessoa-Mahana, H. & Weiss-López, B. Long-range correlations ( $^nJ_{C,H} > 3$ ) in the HMBC spectra of 3-(4-oxo-4H-chromen-3-yl)-acrylic acid ethyl esters. *J. Braz. Chem. Soc.* **16**, 657–661 (2005).
25. Massey, L. K., Sokatch, J. R. & Conrad, R. S. Branched-Chain Amino Acid Catabolism in Bacteria. *Bacteriol. Rev.* **40**, 42–54 (1976).

26. Ichino, T., Arimoto, H. & Uemura, D. Possibility of a non-amino acid pathway in the biosynthesis of marine-derived oxazoles. *Chem. Commun.*, 1742–1744 (2006).
27. Lamberth, C. Oxazole and Isoxazole Chemistry in Crop Protection. *J. Heterocycl. Chem.* **55**, 2035–2045 (2018).
28. Kim, M. K. Novel insight into the function of tankyrase. *Oncol. Lett.* **16**, 6895–6902 (2018).
29. Buchstaller, H.-P. & Dorsch, D. *Heterocyclyl-Butanamide Derivatives*. *Patent* (2015).



# Chapter 5

## Supporting Information

### Structure and Biosynthesis of Myxolipoxazoles and Myxopyrimidinols: Unique Myxobacterial Lipids Featuring Isoxazole or 4-Pyrimidinol Heterocycles

Alexander Popoff<sup>1,2</sup>, **Joachim J. Hug**<sup>1,2</sup>, Sebastian Walesch<sup>1,2</sup>, Ronald Garcia<sup>1,2</sup>  
and Rolf Müller<sup>1,2\*</sup>

1 Department Microbial Natural Products, Helmholtz-Institute for Pharmaceutical Research Saarland (HIPS), Helmholtz Centre for Infection Research (HZI) and Department of Pharmacy Saarland University, Campus E8.1, 66123 Saarbrücken, Germany

2 German Center for Infection Research (DZIF), Partner Site Hannover-Braunschweig, Germany

\* Correspondence: rolf.mueller@helmholtz-hips.de, Tel.: +4968198806-3000

**Table of Contents**

1. General materials and methods .....	285
1.1 Applied software, sequence analysis and bioinformatics methods .....	285
1.2 Maintenance of bacterial cultures, molecular cloning and construction of plasmids <sup>5-7</sup> .....	286
Bacterial cultures and preparation of cryogenic long-term stocks .....	290
Crude extract preparation for analysis of secondary metabolism.....	293
1.3 Compound isolation .....	295
Myxolipoxazole A.....	295
Myxopyrimidinol A.....	296
2. Results .....	297
2.1 Metabolome analysis .....	297
2.2 NMR spectroscopic data .....	300
2.3 IR data .....	307
2.4 Feeding experiments.....	308
2.5 <i>In silico</i> investigation of the myxolipoxazole/myxopyrimidinol biosynthesis.....	311
2.6 Biosynthetic investigation and genetic manipulation of the myxolipoxazole/myxopyrimidinol biosynthesis in the producer <i>M. xanthus</i> Mx x48.....	320
Analysis of secondary metabolism of genetic mutants .....	322
2.7 Biological function of myxolipoxazoles and myxopyrimidinols .....	326
Cell based bioactivity profiling .....	326
2.8 <sup>1</sup> H and <sup>13</sup> C NMR spectra of myxolipoxazole A and myxopyrimidinol A .....	327
NMR data myxolipoxazole A.....	327
NMR data myxopyrimidinol A .....	333
3. References .....	335

## 1. General materials and methods

### 1.1 Applied software, sequence analysis and bioinformatics methods

Geneious prime (Biomatters Ltd., Auckland, New Zealand) was used to design primers for PCR and sequencing, to create plasmid maps and to find open reading frames (ORF). Furthermore all pairwise and multiple alignments of nucleotide or amino acid sequences were performed with the plugin software from Geneious by using the MUSCLE (Multiple Sequence Comparison by Log-Expectation) alignment (3.8.425 by Robert C. Edgar) since it claims to achieve higher average accuracy and to be faster than ClustalW2 or the T-Coffee algorithm<sup>1</sup>. In order to find homologous genes or proteins, either the nucleotide or amino acid sequence of interest was aligned with the basic local alignment search tool (BLAST) against the in-house genome database or the publicly available nucleotide database. Raw data from alignments for *in silico* evaluation of the *iso* biosynthetic gene clusters (BGCs) are stored on the in-house server. The functional prediction of ORFs was performed by either using protein blast and/or blastx program (<https://blast.ncbi.nlm.nih.gov/Blast.cgi>) and Pfam (<http://pfam.xfam.org>)<sup>2</sup>. To obtain further information concerning the catalytic function of identified biosynthetic proteins, the amino acid sequences were evaluated by the *in silico* protein homology analogy recognition engine 2 (Phyre2)<sup>3</sup>. The in-house standard extract database embedded in the software bundle Mxbase Explorer 3.2.27 was used for the search of alternative producers of myxolipoxazoles and the myxopyrimidinols. The molecular formula and experimentally determined retention times of ions typically observed from the myxolipoxazoles and the myxopyrimidinols were used as data input. The BGC sequences of alternative confirmed producers of myxolipoxazoles and the myxopyrimidinols deriving from our in-house database have been deposited in GenBank and are accessible under the accession number as displayed in **Table S1**. In addition, the sequence of the myxolipoxazole and the myxopyrimidinol biosynthetic gene cluster (BGC) originating from *Myxococcus xanthus* Mx x48 will be deposited in the Minimum Information about a Biosynthetic Gene cluster (MIBiG) database under the accession number (upon acceptance)

**Table S1.** In-house genome sequence analysis of myxobacterial strains

Myxolipoxazoles and myxopyrimidinols producing strains	Reference	Accession number BGC
<i>M. xanthus</i> Mx x48 (MCy8278)	<sup>4</sup>	MT513750
<i>Stigmatella aurantiaca</i> Sg a32 DSM No. 53785	This study	MT513751
<i>Stigmatella erecta</i> MCy6787 (Pd e42)	This study	MT513752

### 1.2 Maintenance of bacterial cultures, molecular cloning and construction of plasmids<sup>5-7</sup>

Routine handling of nucleic acids, such as isolation of plasmid DNA, restriction endonuclease digestions, DNA ligations, and other DNA manipulations, was performed according standard procedures<sup>5</sup>. *Escherichia coli* HS996 (Invitrogen) was used as host for standard cloning experiments and *E. coli* SCS110 (Stratagene) for preparation of plasmid DNA free of Dam or Dcm methylation. *E. coli* strains were cultured in LB liquid medium or on LB agar (1% tryptone, 0.5% yeast extract, 0.5% NaCl, (1.5% agar) at 30–37 °C and 200 rpm) overnight. Antibiotics were used at the following final concentrations: 100 µg/mL ampicillin, 50 µg/mL kanamycin and 25 µg/mL chloramphenicol. Transformation of *E. coli* strains was achieved via electroporation in 0.1 cm wide cuvettes at 1250 V, a resistance of 200 Ω, and a capacity of 25 µF. Plasmids (**Table S4**) were purified either by standard alkaline lysis<sup>5</sup> or by using the GeneJet Plasmid Miniprep Kit (Thermo Fisher Scientific) or the NucleoBond PC100 kit (Macherey-Nagel). Restriction endonucleases, alkaline phosphatase (FastAP) and T4 DNA ligase were purchased from Thermo Fisher Scientific. Oligonucleotides used for PCR and sequencing were acquired from Sigma-Aldrich and are listed in **Table S2** and **Table S3** PCRs were carried out in a Mastercycler® pro (Eppendorf) using Phusion™ High-Fidelity according to the manufacturer's protocol.

Temperature and duration setting for each thermocycling step for PCRs using Phusion™ High-Fidelity polymerase were performed as follows: Initial denaturation (30 s, 98 °C); 33 cycles of denaturation (15 s, 98 °C), annealing (15 s, 53–72 °C, depending on the melting temperature of primers) and elongation (based on PCR product length 30 s/1 kb, 72 °C); and final extension (10 min, 72 °C). PCR products or DNA fragments from restriction digestions were purified by agarose gel electrophoresis and isolated using the PCR clean-up gel extraction kit using Nucleo Spin® (Macherey-Nagel). After selection with suitable antibiotics, clones harboring correct ligation constructs were identified by plasmid isolation and restriction analysis with a set of different restriction endonucleases. In addition to restriction analysis, integrity of the constructs for genetic disruption experiments and induced gene expression was verified by sequencing.

**Table S2.** List of oligonucleotides used in this study.

No.	Primer name	Primer sequence 5'–3'
1	JHuFw_Mxx48_iso1_KO	ATATAAGCTTACTCCACGCCCTGCTGATTG
2	JHuRv_Mxx48_iso1_KO	ATATACTAGTTACGGATGACCTTCGCGGT
3	JHuFw_Mxx48_iso2_KO	ATATAAGCTTGCAGCATCGTCCACAACAAC
4	JHuRv_Mxx48_iso2_KO	ATATACTAGTTTCGATGACGAACGAGTTCAGC
5	JHuFw_Mxx48_iso3_KO	ATATAAGCTTGCAGCTCTCGCTCTACCTGTTT
6	JHuRv_Mxx48_iso3_KO	ATATACTAGTAGTTGTAGAAGCGCCCCGTG
7	JHuFw_Mxx48_iso4_KO	ATATAAGCTTAGATTTCCCGCCGCATCGTG
8	JHuRv_Mxx48_iso4_KO	ATATACTAGTTGGGCCGTAGTAGGGGATGA
9	JHuFw_Mxx48_iso5_KO	ATATAAGCTTAGATTGACATGGCCATCGTC
10	JHuRv_Mxx48_iso5_KO	ATATACTAGTTACTCTAACGCGAACAGGGA
11	JHuFw_Mxx48_iso6_KO	ATATAAGCTTCAACACGTACCTCGACTTCA
12	JHuRv_Mxx48_iso6_KO	ATATACTAGTAATACCAGAGGGCTTCCTGG
13	JHuFw_Mxx48_iso8_KO	ATATAAGCTTCCTGGTCGCGGAGGGACTTC
14	JHuRv_Mxx48_iso8_KO	ATATACTAGTGCCAGGATGCGCACCGAGAC
15	JHuFw_Mxx48_iso9_KO	ATATAAGCTTTCATCTCGAACGTGACAGGC
16	JHuRv_Mxx48_iso9_KO	ATATACTAGTAGAGCTTGTCTTCCAGGTG
17	JHuFw_Mxx48_activate_iso2	ATATCATATGAGTCTCGAAGGGGTCATGACGG
18	JHuRv_Mxx48_activate_iso2	ATATGAATTCATGCCCCACCAGCGTCTGCAC

**Table S3.** List of oligonucleotides used as sequencing primers.

No.	Primer name	Primer sequence 5'–3'
1	Test_Fw1_Mxx48_iso1_KO	CGCGACAATAAACATACGCAA
2	Test_Rv2_Mxx48_iso1_KO	TCCACCTTCTCGATGCAGAT
3	Test_Rv3_Mxx48_iso1_KO	CTATCGCCATGTAAGCCCACT
4	Test_Fw4_Mxx48_iso1_KO	CCGATTCATTAATGCAGCTGG
5	Test_Fw1_Mxx48_iso2_KO	CCTTCCTCCAGTACACGTCCGGTTC
6	Test_Rv2_Mxx48_iso2_KO	GAGCAGCTCCATCACCCGCAG
7	Test_Rv3_Mxx48_iso2_KO	GCGATGGCCCACTACGTGAACC
8	Test_Fw4_Mxx48_iso2_KO	GTGATGCTCGTCAGGGGGGC
9	Test_Fw1_Mxx48_iso3_KO	CTTCTGGGAAGGTTGAACAAG
10	Test_Rv2_Mxx48_iso3_KO	GTGGATGCTGTGATAGCCAT
11	Test_Rv3_Mxx48_iso3_KO	GTCTAGCTATCGCCATGTAAG
12	Test_Fw4_Mxx48_iso3_KO	CCGATTCATTAATGCAGCTGG
13	Test_Fw1_Mxx48_iso4_KO	CTGAAGCGACGAAGCATGAC

14	Test_Rv2_Mxx48_iso4_KO	AGTTGCCGATGCCGAAGTAG
15	Test_Rv3_Mxx48_iso4_KO	CTATCGCCATGTAAGCCCACT
16	Test_Fw4_Mxx48_iso4_KO	CCGATTCATTAATGCAGCTGGC
17	Test_Fw1_Mxx48_iso5_KO	ACTCACCGCGCTGCACTTCG
18	Test_Rv2_Mxx48_iso5_KO	GGAGTGAAAGGCATGGCGTGTCC
19	Test_Rv3_Mxx48_iso5_KO	GCGATGGCCCACTACGTGAACC
20	Test_Fw4_Mxx48_iso5_KO	GTGATGCTCGTCAGGGGGGC
21	Test_Fw1_Mxx48_iso6_KO	GTGTTTCCCATCGTCGCGGAGC
22	Test_Rv2_Mxx48_iso6_KO	CGTTGTAGAGGGCCGCGATGC
23	Test_Rv3_Mxx48_iso6_KO	GCGATGGCCCACTACGTGAACC
24	Test_Fw4_Mxx48_iso6_KO	GTGATGCTCGTCAGGGGGGC
25	Test_Fw1_Mxx48_iso8_KO	CTGACGAAGTACGACCTCTG
26	Test_Rv2_Mxx48_iso8_KO	ACCGTCCTGATCATCTTGAAG
27	Test_Rv3_Mxx48_iso8_KO	AGTCTAGCTATCGCCATGTAAG
28	Test_Fw4_Mxx48_iso8_KO	CCGATTCATTAATGCAGCTGG
29	Test_Fw1_Mxx48_iso9_KO	CACTCGGAGCTGATGCGGCC
30	Test_Rv2_Mxx48_iso9_KO	GCTGGCGTGTCGGGCATCTC
31	Test_Rv3_Mxx48_iso9_KO	GCGATGGCCCACTACGTGAACC
32	Test_Fw4_Mxx48_iso9_KO	GTGATGCTCGTCAGGGGGGC
33	Test_Fw1_Mxx48_iso2_KO_activate	CTTCCAGACCTTTCCTCGTC
34	Test_Rv2_Mxx48_iso2_KO_activate	TGCTCGTCTTGGAATCATC
35	Test_Rv3_Mxx48_iso2_KO_activate	CTGATTCTGTGGATAACCGTATTAC
36	Test_Fw4_Mxx48_iso2_KO_activate	GACTCTAGCCGACCGACTGA

Table S4. List of plasmids used in this study

No.	Plasmid name/ characteristic	Size [kb]	Function	Reference
1	pCR2.1-TOPO (EcoRI, religated)	3.931	Used as template to PCR-amplify <i>tn5_kanR</i> gene	TOPO®TA Cloning® Kit Thermo Fisher Scientific
2	pFP <sub>van</sub> <i>pcyA</i>	6.181	pCR2.1-TOPO derivative used as vector to ligate PCR products for subsequent vanillate-induced gene expression in <i>M. xanthus</i> DK1622	<sup>8</sup>

**Table S5.** List of PCR-amplified constructs

No.	PCR product name / characteristic	Size [bp]	Template	Primer used
1	Mxx48_iso1_ homology	317	gDNA from <i>M. xanthus</i> Mxx48	primer No. 1 primer No. 2
2	Mxx48_iso2_ homology	1269	gDNA from <i>M. xanthus</i> Mxx48	primer No. 3 primer No. 4
3	Mxx48_iso3_ homology	598	gDNA from <i>M. xanthus</i> Mxx48	primer No. 5 primer No. 6
4	Mxx48_iso4_ homology	583	gDNA from <i>M. xanthus</i> Mxx48	primer No. 7 primer No. 8
5	Mxx48_iso5_ homology	1230	gDNA from <i>M. xanthus</i> Mxx48	primer No. 9 primer No. 10
6	Mxx48_iso6_ homology	1357	gDNA from <i>M. xanthus</i> Mxx48	primer No. 11 primer No. 12
7	Mxx48_iso8_ homology	938	gDNA from <i>M. xanthus</i> Mxx48	primer No. 13 primer No. 14
8	Mxx48_iso9_ homology	1373	gDNA from <i>M. xanthus</i> Mxx48	primer No. 15 primer No. 16
9	Mxx48_iso2_ activate	1117	gDNA from <i>M. xanthus</i> Mxx48	primer No. 17 primer No. 18

**Table S6.** List of genetic constructs generated in this study

No.	Plasmid name	Construction details/ characteristic
1	pCR2.1-TOPO_ Mxx48_iso1_KO	Construct obtained by conventional restriction ligation of plasmid pCR2.1-TOPO and PCR product No. 1
2	pCR2.1-TOPO_ Mxx48_iso2_KO	Construct obtained by conventional restriction ligation of plasmid pCR2.1-TOPO and PCR product No. 2
3	pCR2.1-TOPO_ Mxx48_iso3_KO	Construct obtained by conventional restriction ligation of plasmid pCR2.1-TOPO and PCR product No. 3
4	pCR2.1-TOPO_ Mxx48_iso4_KO	Construct obtained by conventional restriction ligation of plasmid pCR2.1-TOPO and PCR product No. 4
5	pCR2.1-TOPO_ Mxx48_iso5_KO	Construct obtained by conventional restriction ligation of plasmid pCR2.1-TOPO and PCR product No. 5

6	pCR2.1-TOPO_ Mxx48_iso6_KO	Construct obtained by conventional restriction ligation of plasmid pCR2.1-TOPO and PCR product No. 6
7	pCR2.1-TOPO_ Mxx48_iso8_KO	Construct obtained by conventional restriction ligation of plasmid pCR2.1-TOPO and PCR product No. 7
8	pCR2.1-TOPO_ Mxx48_iso9_KO	Construct obtained by conventional restriction ligation of plasmid pCR2.1-TOPO and PCR product No. 8
9	pFP <sub>van</sub> _iso2_precursor_activate	Construct obtained by conventional restriction ligation of plasmid pFP <sub>van</sub> _pcyA and PCR product No. 9

### Bacterial cultures and preparation of cryogenic long-term stocks

The myxobacterial strain *Corallococcus* sp. MCy9072 was recognized for its swarming and fruiting body formation in the standard *Escherichia coli* bacterial baiting method<sup>9</sup>. Repeated transfer of the swarm edge onto fresh VY/2 agar led to the isolation of the strain (**Figure S1**). Cultivation and maintenance was kept at VY/2 agar for 3–4 weeks at 30 °C.

Long-term storage of *Corallococcus* sp. MCy9072, *Myxococcus xanthus* Mx x48 (wild type and mutants), *Stigmatella aurantiaca* Sg a32 and *Stigmatella erecta* MCy6787 (Pd e42) was achieved by cryogenic preservation at -80 °C as tenfold concentrated glycerol stocks. Ten mL of well-grown liquid culture was transferred to a 50-mL Falcon tube and centrifuged for 5 min at 8000 rpm (Eppendorf Centrifuge 5804R). Nine mL of the supernatant was discarded, leaving 1.0 mL supernatant in the Falcon tube. The cell pellet was re-suspended by pipetting up and down and then mixed with 1.0 mL glycerol (50%). The myxobacterial glycerol sample (25% glycerol) was transferred into a cryogenic vial and maintained at -80 °C.



**Figure S1.** Orange scattered fruiting bodies of *Corallococcus* sp. MCy9072 on VY/2 agar with typical ridge and hump-shaped morphologies.



Bacterial strains for the experiments of this work, their relevant characteristics and sources are listed in the **Table S7**.

**Table S7.** Bacterial strains used in this study.

Strain	Function	Origin
<i>E. coli</i> HS996	Standard cloning host	Invitrogen
<i>Corallococcus</i> sp. MCy9072	Producer of myxolipoxazoles and myxopyrimidinols: Identification and isolation of myxolipoxazole A	This study
<i>Myxococcus xanthus</i> Mx x48	Producer of myxolipoxazoles and myxopyrimidinols: Identification and isolation of myxopyrimidinol, genetic characterization	MINS lab
<i>Stigmatella aurantiaca</i> Sg a32	Producer of myxolipoxazoles and myxopyrimidinols: Alternative producer with available genome and metabolome data	MINS lab
<i>Stigmatella erecta</i> MCy6787 (Pd e42)	Producer of myxolipoxazoles and myxopyrimidinols: Alternative producer with available genome and metabolome data	MINS lab

### Cultivation media

The pH of all cultivation media was adjusted before autoclaving. Autoclaving was conducted at 121°C for 20 min.

#### CTT-medium

Ingredient	Concentration
Casitone	10 g/L
TRIS	10 mM
KH <sub>2</sub> PO <sub>4</sub>	1 mM
MgSO <sub>4</sub> x 6H <sub>2</sub> O	8 mM

pH 7.6 with KOH<sub>aq</sub>

#### VY/2-medium

Ingredient	Concentration
Baker's yeast	5 g/L
CaCl <sub>2</sub> x 2H <sub>2</sub> O	1 g/L

pH 7.2 with KOH<sub>aq</sub>, after autoclaving medium was supplemented with 0.5 mg/L vitamin B<sub>12</sub>.

### LB-medium

Ingredient	Concentration
Yeast extract	5 g/L
Tryptone	10 g/L
NaCl	5 g/L

pH 7.2 with KOH<sub>aq</sub>

### M7/s4-medium

Ingredient	Concentration
Soy flour	5 g/L
Corn starch	5 g/L
Glucose	2 g/L
Yeast extract	1 g/L
MgSO <sub>4</sub> x 6H <sub>2</sub> O	1 g/L
CaCl <sub>2</sub> x 2H <sub>2</sub> O	1 g/L
HEPES	10 g/L

pH 7.4 with NaOH<sub>aq</sub>, after autoclaving medium was supplemented with 0.1 mg/L vitamin B<sub>12</sub>.

### CyH-medium

Ingredient	Concentration
Casiton (Difco)	1.5 g/L
Yeast extract (Difco)	1.5 g/L
Soluble starch (Roth)	4.0 g/L
Soy flour, defatted (Hensel)	1.0 g/L
D-glucose	1 g/L
MgSO <sub>4</sub> x 6H <sub>2</sub> O	0.5 g/L
CaCl <sub>2</sub> x 2H <sub>2</sub> O	1 g/L
HEPES	11.9 g/L

pH 7.4 with KOH<sub>aq</sub>, after autoclaving medium was supplemented with 0.05 mg/L vitamin B<sub>12</sub> and 0.8 mg/L Fe-EDTA.

**DSM924-seed medium adapted from Steven J. Gould and Shyh-Chen Ju<sup>10,11</sup>**

Ingredient	Concentration
D-glucose	10 g/L
Peptone (Difco)	10.0 g/L
Yeast extract (Difco)	2.6 g/L

pH 7.2 with NaOH<sub>aq</sub>**DSM924-production medium adapted from Steven J. Gould and Shyh-Chen Ju<sup>10,11</sup>**

Ingredient	Concentration
D-glucose	2.0 g/L
Yeast extract (Difco)	2.5 g/L
Soy flour, defatted (Hensel)	20.0 g/L
Corn starch	10.0 g/L
NH <sub>4</sub> Cl	5 g/L
“Lard oil“ (Hausmacher Zwiebelschmalz (Ponnath))	5 g/L

pH 7.2 with NaOH<sub>aq</sub>**Crude extract preparation for analysis of secondary metabolism**

The initial screening for *Corallocccus* sp. MCy9072 secondary metabolites was performed in CyH medium supplemented with 2% (v/v) amberlite resin XAD-16 and cultivated for 10 d at 30 °C, 180 rpm.

*Myxococcus xanthus* Mx x48 and mutant strains were grown on CTT agar plates for 3–5 days at 30 °C. For the preparation of liquid pre-cultures, a suitable portion of overgrown agar was transferred to inoculate 20 mL of CTT medium in an Erlenmeyer flask (100 mL) and incubated at 30 °C and 180 rpm for 2–4 days. 3 mL of the pre-culture was used to inoculate 50 mL of M7/s4 or CTT medium containing appropriate antibiotic selection and 2% amberlite resin XAD-16 (Sigma Aldrich) (after one day of incubation at 30 °C and 200 rpm). In order to obtain statistically significant results, three independent transformants of each recombinant strain were selected and fermentations were performed at least in triplicates.

After eight days of fermentation at 30 °C and 200 rpm, the culture broths were harvested and lyophilized. Afterwards the residual material was extracted by stirring with 25 mL MeOH and 25 mL acetone for 2 h, filtered through filter paper (folded filters grade: 3hw from Sartorius) into a round bottom flask. The solvent of the filtered extracts was removed under vacuum (BÜCHI Rotavapor R-210) and the extracts re-dissolved in 1.5 mL MeOH and stored at -20 °C. The re-dissolved extracts were diluted with MeOH (1:3 (extract/MeOH (v:v))) centrifuged at 13000 g for 10 min (VWR centrifuge ECN521-3601, Hitachi Koki Co., Ltd) and 1 µL of the supernatant was subjected to HPLC-MS analysis as described further below.

### Feeding experiments with stable isotope-labeled precursors

0.5 mL of myxobacterial pre-culture was used to inoculate 10 mL of CTT or M7/s4 medium and for three consecutive days, two times a day, 10  $\mu$ L (in total 60  $\mu$ L) of a 0.1 M solution of either L-glutamine  $^{15}\text{N}_2^{13}\text{C}_6$ , L-asparagine  $^{15}\text{N}_2^{13}\text{C}_5$ , either L-glutamic acid  $^{15}\text{N}$  or acetate  $^{13}\text{C}_2$  in  $\text{H}_2\text{O}$  dd + MeOH [1:1 vol.] was added. Additionally, one sample without feeding was prepared used as control. 12 h after the last supplementation with the respective stable isotope-labeled precursor, the culture broths were complemented with 2% amberlite resin XAD-16 (Sigma). After 10 days of incubation at 30 °C and 200 rpm, the culture broths were harvested by centrifugation at 8000 rpm for 10 min, the supernatant was discarded, and the residual material was extracted first with 25 mL MeOH, stirred for 1 h, filtered through filter paper into a round bottom flask and afterwards this procedure was repeated with 25 mL acetone. Filtered extracts were evaporated and subsequently re-dissolved in 2 mL MeOH. For metabolic analysis the re-dissolved sample was centrifuged and analyzed by a Dionex HPLC coupled to the TOF mass spectrometer maXis 4G as described further below.

### Analysis of secondary metabolites in broth extracts

The secondary metabolism of broth extracts was analyzed by HPLC-HRESI-DAD-MS on a Bruker maXis 4G mass spectrometer coupled with a Dionex Ultimate 3000 RSLC system using a BEH C18 column (100  $\times$  2.1 mm, 1.7  $\mu$ m, Waters, Germany) with a gradient of 5–95% acetonitrile (ACN) + 0.1% formic acid (FA) in  $\text{H}_2\text{O}$  + 0.1% FA at 0.6 mL/min and 45 °C over 9 or 18 min with UV detection by a diode array detector at 200–600 nm. Mass spectra were acquired from 150 to 2000  $m/z$  at 2 Hz. The detection was performed in the positive MS mode. The plugin for Chromeleon Xpress (Dionex) was used for operation of UltiMate 3000 LC System. HyStar (Bruker Daltonic) was used to operate on maXis 4G speed MS system. HPLC-MS mass spectra were analyzed with DataAnalysis 4.2 (Bruker Daltonic).

### Large-scale fermentation of *Myxococcus xanthus* Mx x48

*M. xanthus* Mx x48 was grown on CTT agar plates for 3–5 days at 30 °C. For the preparation of liquid pre-cultures, a suitable portion of overgrown agar was transferred to inoculate 50 mL of CTT medium in an Erlenmeyer flask (300 mL) and incubated at 30 °C and 200 rpm for 2–4 days. 8  $\times$  5 mL of the pre-culture was used to inoculate 8  $\times$  250 mL of M7/s4 medium (1 L shake flask with baffles) and incubated at 30 °C and 200 rpm for 2–4 days. Afterwards 12  $\times$  150 mL of pre-culture was used to inoculate 12  $\times$  1.25 L M7/s4 medium (5 L shake flask with baffles) and incubated at 30 °C and 160 rpm for nine days. After two days of incubation, each shake flask was supplemented with 2% amberlite resin XAD-16 (Sigma Aldrich). The fermentation was terminated through centrifugation of the whole cell broth; the supernatant was discarded, whereas the cell pellet and amberlite resin XAD-16 was lyophilized. The lyophilized cell pellet with XAD-16 was used to prepare a crude extract (see below).

### 1.3 Compound isolation

#### Myxolipoxazole A

##### Analysis during purification

All measurements to analyze the mass of myxolipoxazole A during purification were performed on a Dionex Ultimate 3000 RSLC system coupled to the amaZon iontrap MS using a BEH C18, 100 x 2.1 mm, 1.7  $\mu$ 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 (FA) to (B) ACN + 0.1% FA at a flow rate of 0.6 mL/min and 45 °C. The gradient was initiated by a 0.5 min isocratic step at 5% B, followed by linear increase to 95% B in 18 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 200–600 nm. The detection was performed in the positive ESI MS positive mode

##### Semi-preparative HPLC chromatography – Isolation of Myxolipoxazole A

A 2.5 L culture of MCy9072 (25 x 100 ml in 300 mL flasks) in CyH medium was harvested by pouring the culture through a 100  $\mu$ m sieve to separate the medium from cells and the absorber resin XAD-16. The cells and the XAD-16 were lyophilized for two days and extracted 3 x with 150 mL acetone/MeOH 1:1. The solvents were poured through a funnel with glass wool and the solvent was evaporated. The dried extract was re-dissolved in 100 mL hexane and 100 ml MeOH and transferred into a separatory funnel. The layers were separated, and the liquid/liquid extraction was repeated with 100 mL hexane. The MeOH layer was dried and re-dissolved in 300 mL H<sub>2</sub>O and 150 mL CHCl<sub>3</sub>. The layers were separated in a separatory funnel and the aqueous layer was re-extracted with 150 mL CHCl<sub>3</sub>. After extraction with CHCl<sub>3</sub> the liquid/liquid extraction was repeated twice with 150 mL ethyl acetate. During extraction with ethyl acetate the aqueous layer was acidified with 1 N HCl. The layers were dried, dissolved in MeOH and united. For further purification a size exclusion chromatography was performed using Sephadex LH-20 resin and MeOH as eluent. A flow rate of 1 drop/sec was applied and 500 drops were collected for each fraction. An aliquot of each fraction was taken and measured by LC-MS to screen for the myxolipoxazole A. The fractions containing myxolipoxazole A were pooled and the solvent was evaporated. For further purification HPLC was used.

Following setup was used for HPLC: Dionex Ultimate 3000 coupled with Bruker High capacity trap mass spectrometer (HCT); Column: Luna 250 x 10 mm, 5  $\mu$ m, flowrate 6 mL/min and column temperature 40 °C with H<sub>2</sub>O + 0.1% FA as eluent A and ACN + 0.1% FA as eluent B. The following gradient was applied 0–35 min linear increase of eluent B to 95%; 30–34 min 95% eluent B; 34–34.5 min linear decrease of eluent B to 5%; 34.5–46 min re-equilibration with 5% eluent B. Myxolipoxazole A was collected from 23.75–24 min. The collected fraction was dried and dissolved in 4 mL MeOH.

### Myxopyrimidinol A

#### Analysis during purification

All measurements to analyze the mass of Myxopyrimidinol A during purification were performed on a Dionex Ultimate 3000 RSLC system coupled to the amaZon iontrap MS using a BEH C18, 50 x 2.1 mm, 1.7  $\mu$ m dp column equipped with a C18 precolumn (Waters). Samples of 2  $\mu$ L were separated by a gradient from (A) H<sub>2</sub>O + 0.1% formic acid (FA) to (B) ACN + 0.1% FA at a flow rate of 0.6 mL/min and a column temperature of 45 °C. The gradient was initiated by a 0.5 min isocratic step at 5% B, followed by linear increase to 95% B in 9 min. After a 1 min step at 95% B, B was decreased to 5% in 0.5 min and the system was re-equilibrated for 1.5 min. UV-spectra were recorded by a DAD in the range from 200–600 nm. The detection was performed in the positive ESI MS/MS mode.

#### Supercritical fluid extraction (SFE)

The lyophilized cell pellet of a 12.5 L fermentation culture with 2% XAD-16 was extracted with the MV-10 ASFE system (Waters). The sample was grinded to pulverize and homogenize it. It was then filled into a 10 mL extraction vessel (Waters) and tightly packed. The vessel was plugged to tubes in the column oven for extraction. Extraction was performed with following parameters for all fractions: oven temperature: 40 °C, extraction pressure: 170 bar, flow rate (CO<sub>2</sub> with or without organic solvent): 10 mL/min. To transport extracted compounds to the collector, MeOH to a final flow of 3.5 mL/min organic solvent was added to the elution mixture after the extraction vessel. Extraction parameters were composed as follows: **SFE-1**: 100% CO<sub>2</sub>; **SFE-2**: 80% CO<sub>2</sub>/20% ethyl acetate; **SFE-3**: 80% CO<sub>2</sub>/20% MeOH and **SFE-4**: 50% CO<sub>2</sub>/50% MeOH.

Every fraction was extracted for 10 min and collected in 100 mL Schott bottles. Fractions were afterwards dried using the rotary evaporator (BÜCHI Rotavapor R-210). The dried residue was re-dissolved in MeOH or, if insoluble in MeOH, in ethyl acetate and stored at -20 °C. Myxopyrimidinol A was present in **SFE-2** and **SFE-3**.

#### Semi-preparative HPLC chromatography – Isolation of Myxopyrimidinol A

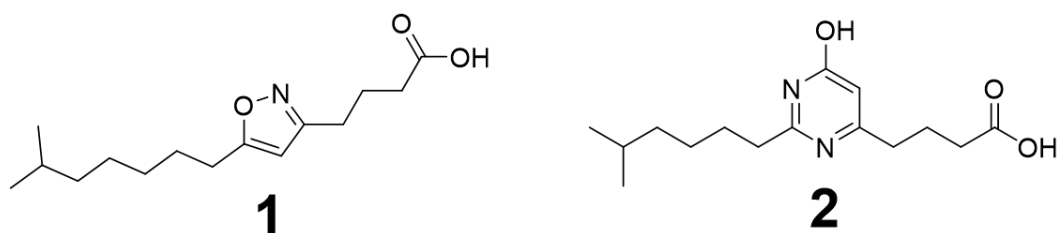
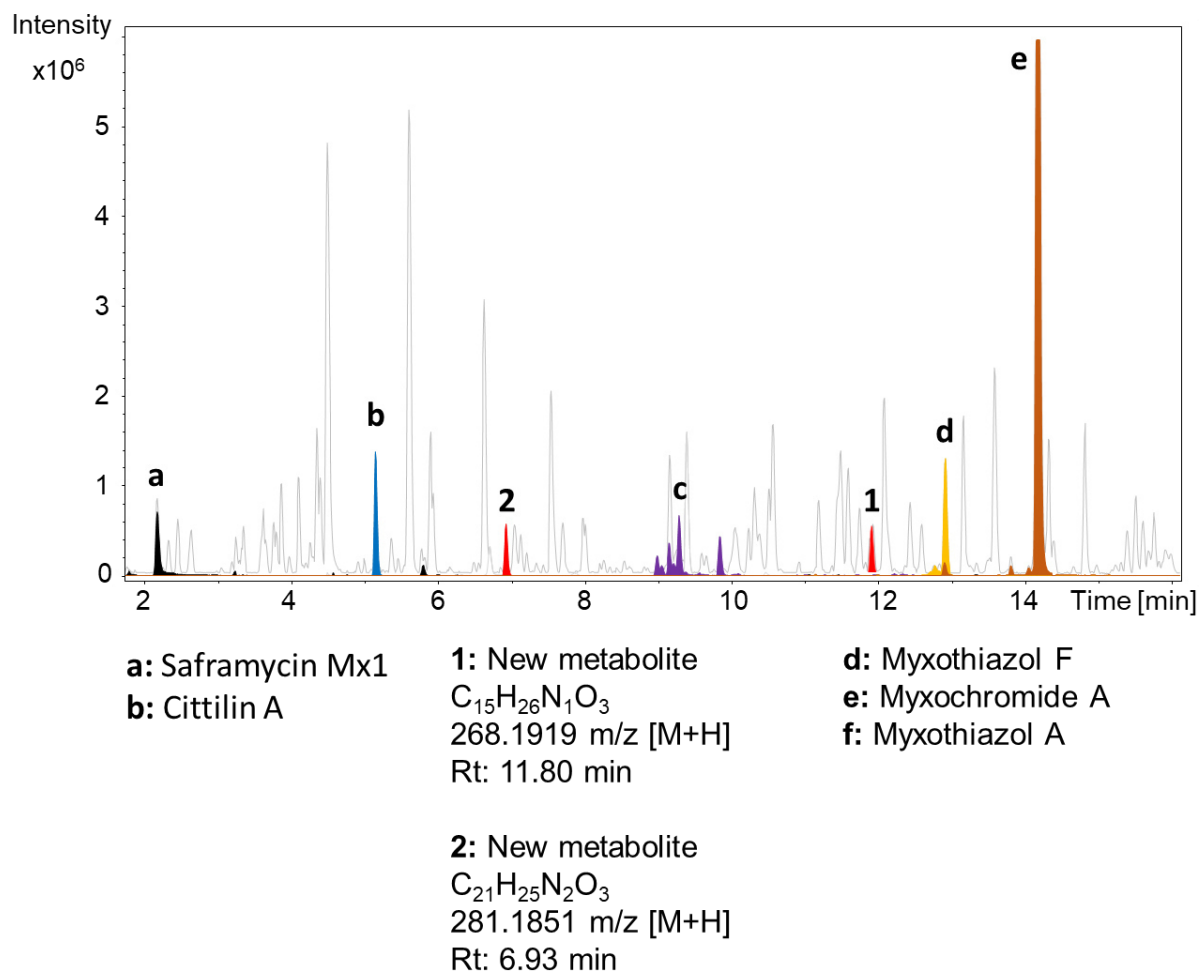
Semi-preparative HPLC purification was done using a Dionex Ultimate 3000 SL systems consisting of a SWPS 3000 SL Autosampler, P680 Pump, TCC100 column oven, PDA100 UV-Detector and AFL 3000 fraction collector linked to a Bruker Daltonics HCT plus ion trap system with standard ESI ion source.

An xSelect® Peptide CSH C18, 10 mm x 250 mm column (Waters) was used as stationary phase, H<sub>2</sub>O + 0.1% FA (Solvent A) and ACN + 0.1% FA (Solvent B) as mobile phase. Semi-preparative HPLC was carried out with a flow rate of 5 mL/min and following multi-step gradient: the gradient was initiated by a 2 min isocratic step at 5% B, followed by a steep ramp to 30% B in 0.5 min. Then B was increased to 42.5% within 15 min and afterwards to 95% in 0.5 min. After a 2 min isocratic step at 95% B, it was reduced to 5% B within 1 min and the system was re-equilibrated for 2 min.

Myxopyrimidinol A was collected 12.10–12.50 min, dried using the rotary evaporator (BÜCHI Rotavapor R-210), re-dissolved in MeOH, re-dried under N<sub>2</sub>-flow and stored at -20 °C.

## 2. Results

### 2.1 Metabolome analysis



**Figure S2.** LC-MS analysis of *Myxococcus xanthus* Mx x48. The corresponding peaks of previously identified myxobacterial compounds are colored. The chemical structure of **1** and **2** are displayed below the chromatogram.

**Table S8.** Initial metabolome analysis of *M. xanthus* Mx x48.

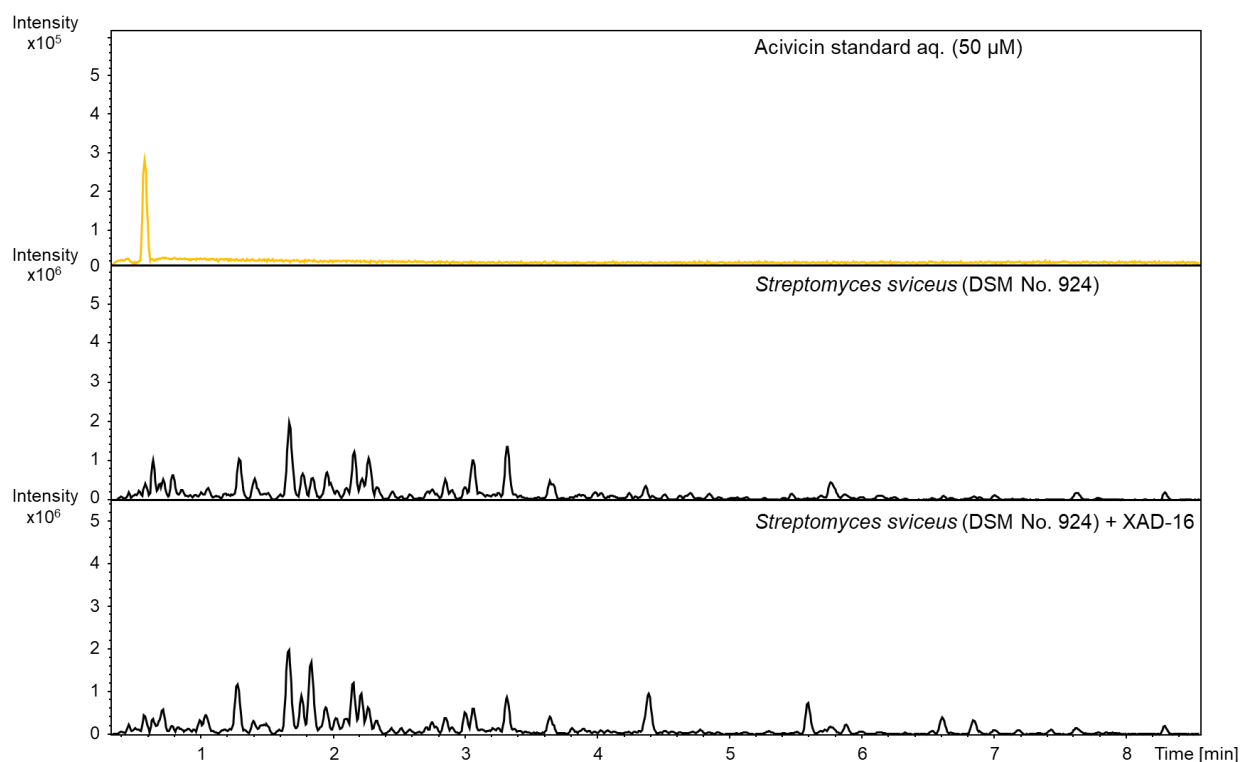
Identified metabolite	Mass [M+H]	Sum formula	Retention time [min]	Reference
Cittilin A	631.17590	C <sub>34</sub> H <sub>39</sub> N <sub>4</sub> O <sub>8</sub>	5.15	<sup>12</sup>
Cittilin B	617.26007	C <sub>33</sub> H <sub>37</sub> N <sub>4</sub> O <sub>8</sub>	4.92	<sup>12</sup>
cyclo(Phe-Pro)_R63.1	245.12912	C <sub>14</sub> H <sub>16</sub> N <sub>2</sub> O <sub>2</sub>	4.11	-
cyclic-dipeptide02	245.12922	C <sub>14</sub> H <sub>16</sub> N <sub>2</sub> O <sub>2</sub>	4.36	-
cyclo(Pro-isoLeu)	211.14445	C <sub>11</sub> H <sub>18</sub> N <sub>2</sub> O <sub>2</sub>	3.49	-
cyclo(Pro-Leu)	211.14446	C <sub>11</sub> H <sub>18</sub> N <sub>2</sub> O <sub>2</sub>	3.77	-
Cyclo(leucylprolyl)	211.14455	C <sub>11</sub> H <sub>18</sub> N <sub>2</sub> O <sub>2</sub>	3.88	-
cyclic-dipeptide01	197.12870	C <sub>10</sub> H <sub>16</sub> N <sub>2</sub> O <sub>2</sub>	2.64	-
cyclo(Tyr-Pro)	261.12381	C <sub>14</sub> H <sub>16</sub> N <sub>2</sub> O <sub>3</sub>	2.46	-
cyclo(Tyr-Pro)	261.12389	C <sub>14</sub> H <sub>16</sub> N <sub>2</sub> O <sub>3</sub>	2.35	-
Saframycin Mx1	568.25351 [M-H <sub>2</sub> O+H]	C <sub>29</sub> H <sub>36</sub> N <sub>4</sub> O <sub>8</sub>	2.42 [285.13379]	<sup>4</sup>
Myxopyrimidinol A	281.18557	C <sub>21</sub> H <sub>25</sub> N <sub>2</sub> O <sub>3</sub>	6.93	This study
Myxothiazol F	521.20132	C <sub>25</sub> H <sub>35</sub> N <sub>3</sub> O <sub>5</sub> S <sub>2</sub>	8.99/9.16/9.29	Unpublished
DKxanthen-534	535.25537	C <sub>29</sub> H <sub>34</sub> N <sub>4</sub> O <sub>6</sub>	7.19	<sup>13</sup>
Myxolipoxazole A	268.19192	C <sub>15</sub> H <sub>26</sub> NO <sub>3</sub>	11.80	This study
Myxochromide A	846.4770	C <sub>45</sub> H <sub>63</sub> N <sub>7</sub> O <sub>9</sub>	12.91	<sup>14</sup>
Myxothiazol A	456.17774	C <sub>24</sub> H <sub>29</sub> N <sub>3</sub> O <sub>2</sub> S <sub>2</sub>	14.04	<sup>15</sup>

### Metabolome analysis of *Streptomyces sviveus* (DSM No. 924)

*Streptomyces sviveus* (DSM No. 924), which is deposited at the German Collection of Microorganisms and Cell Cultures GmbH as producer of acivicin and 4-hydroxyacivicin, was directly used (agar culture from DSMZ) to inoculate 50 mL of DSM-924 seed culture (300 mL baffled shake flask). After three days of incubation at 30 °C and 200 rpm, 5 mL of well-grown seed culture was used to inoculate 200 mL of DSM-924 production culture (1 L baffled shake flask); control without inoculum and sample with additional supplementation of 2% XAD-16 was prepared as well. After five days of fermentation at 30 °C and 200 rpm, the culture broths (cells, supernatant (and XAD-16)) were harvested and lyophilized. Afterwards the residual material was extracted by stirring with 25 mL MeOH and 25 mL MQ water for 2 h, filtered through filter paper (folded filters grade: 3hw from Sartorius) into a round bottom flask. The solvent of the filtered extracts was removed under vacuum (BÜCHI Rotavapor R-210) and the extracts re-dissolved in 1.5 mL MeOH : H<sub>2</sub>O (1:1 (v:v)) and stored at -20 °C. The re-dissolved extracts were diluted 1:5 (v:v) with MeOH : H<sub>2</sub>O (1:1 (v:v)) and centrifuged at 13000 g for 10 min (VWR centrifuge ECN521-3601, Hitachi Koki Co., Ltd) and 1 µL of the supernatant was subjected to HPLC-MS analysis as described above.



In none of the crude extracts a secondary metabolite with an ion peak of  $m/z = 179.0223$  [M+H] at the retention time of 0.58 min was found, such as shown for commercially purchased acivicin (Sigma-Aldrich, CAS number 42228-92-2) (**Figure S3**).



**Figure S3.** HPLC-MS BPCs of acivicin standard (top) and *Streptomyces sviveus* (DSM No. 924) crude extracts without (middle) and with (bottom) supplementation of 2% XAD-16.

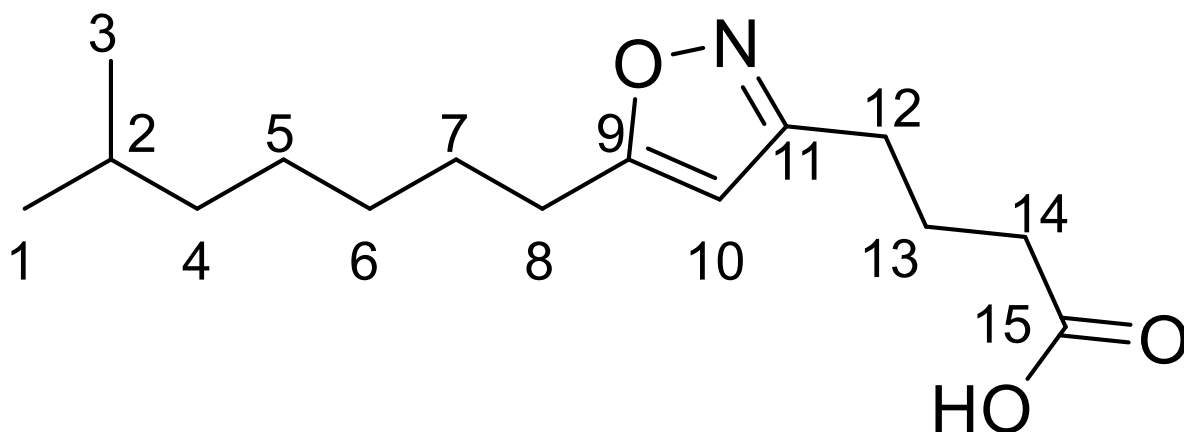
## 2.2 NMR spectroscopic data

**Table S9.**  $^1\text{H}$  and  $^{13}\text{C}$  NMR spectral data of myxolipoxazole A; recorded at 500 and 125 MHz in  $\text{CDCl}_3$ ; d: doublet; dd: doublet of doublets; s: singlet; m: multiplet.

Position	$\delta$ $^1\text{H}$ ppm ( $J=\text{Hz}$ )	Multiplicity	$\delta$ $^{13}\text{C}$ ppm	COSY	HMBC
1	0.86 (6.6)	d	23.1	2	2,3,4
2	1.52	m	28.4	1,3,4	1,3,4,5
3	0.86 (6.6)	d	23.1	1,2, 4	4,1,2
4	1.16	m	39.2	5,2	1,2,3,5,6
5	1.31	m	27.3	4,6	2,4,6
6	1.32	m	29.7	5,7	4,5,7,8
7	1.67 (7.49)	quin	28.0	6, 8	6,8,9
8	2.69	m	27.2	7,10	6,7,9,10
9			174.1	---	7,8,10
10	5.83	s	100.5	8	8,9,11,12
11			163.1	---	10,12,13
12	2.70	m	25.7	13	10,11,13,14
13	1.99	m	23.7	12,14	11,12,14,15
14	2.42 t (7.3)	t	33.6	13	12,13,15
15			178.2		13,14

**Table S10.**  $^1\text{H}$  and  $^{13}\text{C}$  NMR spectral data of myxolipoxazole A; recorded at 500 and 125 MHz in  $\text{CD}_3\text{OD}$ ; d: doublet; dd: doublet of doublets; s: singlet; m: multiplet.

Position	$\delta$ $^1\text{H}$ ppm ( $J=\text{Hz}$ )	Multiplicity	$\delta$ $^{13}\text{C}$ ppm	COSY	HMBC
<b>1</b>	0.88 d 6.6	<b>d</b>	23.0	2	2,3,4
<b>2</b>	1.53 m	<b>m</b>	29.1	1,3,4	2,3
<b>3</b>	0.88 d 6.6	<b>d</b>	23.0	2	1,2,4
<b>4</b>	1.19 m	<b>m</b>	40.0	2,5	1,3,6
<b>5</b>	1.35 m	<b>m</b>	28.1	<b>4,6,7</b>	<b>7</b>
<b>6</b>	1.36 m	<b>m</b>	30.4	<b>5</b>	<b>4,7,8</b>
<b>7</b>	1.69 m	<b>m</b>	28.7	6,8	5,6,8
<b>8</b>	2.72 t 7.6	<b>t</b>	27.5	7,10	6,7,9,10
<b>9</b>			175.1		<b>8,10</b>
<b>10</b>	6.08 s	<b>s</b>	101.9	<b>8</b>	<b>8,9,11,12</b>
<b>11</b>			165.4		<b>10,12,13</b>
<b>11-N</b>			361 <sup>e</sup>		<b>12</b>
<b>12</b>	2.64 t 7.6	<b>t</b>	26.7	13	10,11,11-N,13,14
<b>13</b>	1.92 m	<b>m</b>	26.4	12,14	11,12,14,15
<b>14</b>	2.23 t 7.5	<b>t</b>	38.2	13	12,13,15
<b>15</b>			181.4		<b>13,14</b>

<sup>a</sup> acquired at 125 MHz and assigned from 2D NMR spectra, referenced to solvent signal  $\text{CD}_3\text{OD}$  at  $\delta$  49.00 ppm.<sup>b</sup> acquired at 500 MHz, referenced to solvent signal  $\text{CD}_3\text{OD}$  at  $\delta$  3.31 ppm.<sup>c</sup> proton showing COSY correlation to indicated proton.<sup>d</sup> proton showing HMBC correlations to indicated carbons.<sup>e</sup>  $^{15}\text{N}$  shift value obtained from  $^{15}\text{N}$  HMBC acquired at 70 MHz**Figure S4.** Structure and carbon numbering of myxolipoxazole A (**1**).

**Table S11.**  $^1\text{H}$  and  $^{13}\text{C}$  NMR spectral data of myxopyrimidinol A; d: doublet; dd: doublet of doublets; s: singlet; m: multiplet.

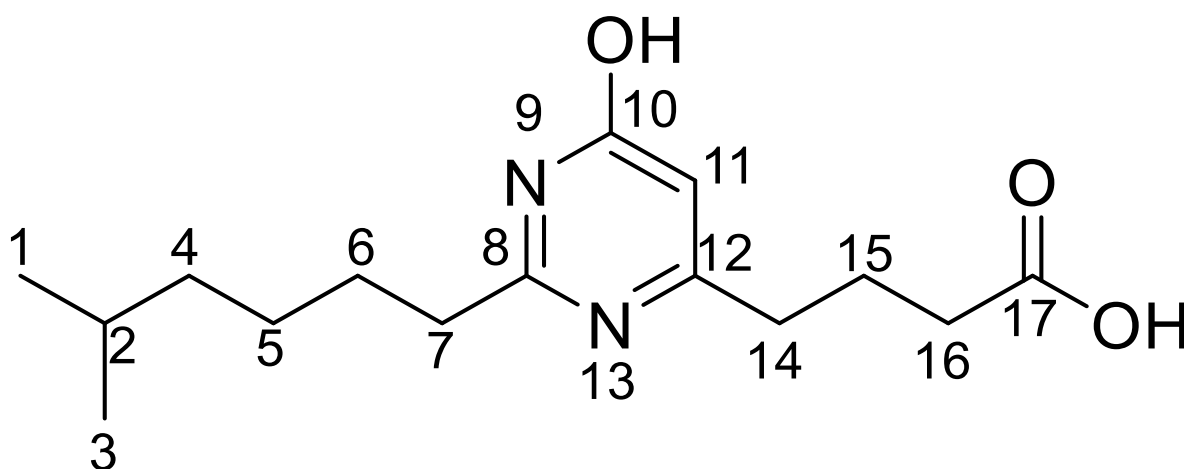
Position	$\delta$ $^1\text{H}$ ppm ( $J=\text{Hz}$ )	Multiplicity	$\delta$ $^{13}\text{C}$ ppm	COSY	HMBC
1	0.83 (6.4)	d	22.5	2	2,3,4
2	1.49	m	27.2	1,3,4	1,3,4,5
3	0.83 (6.4)	d	22.1	2	2,3,4
4	1.15 (6.9/7.6)	d, t	38.0	3,5	1,2,3,5,6
5	1.24	m	26.2	4,6	2,4,6,7
6	1.60 (7.5/7.5)	t, t	27.0	5,7	4,5,7,8
7	2.48	m	34.2	6	5,6,8
8	---	---	161.7	---	6,7
9	---	---	<i>N</i>	---	---
10	---	---	162.9	---	11
11	5.92	s	109.1	---	7,10,14
12	---	---	167.2	---	14
13	---	---	<i>N</i>	---	---
14	2.38 (7.3)	t	35.8	15	11,12,15,16
15	1.75 (7.2/7.4)	t, t	23.6	14,16	14,16,17
16	2.09	m	34.4	15	14,15,17
17	---	---	175.1	---	15,16

<sup>a</sup> acquired at 175 MHz and assigned from 2D NMR spectra, referenced to solvent signal DMSO  $d_6$  at  $\delta$  39.52 ppm.

<sup>b</sup> acquired at 700 MHz, referenced to solvent signal DMSO  $d_6$  at  $\delta$  2.50 ppm.

<sup>c</sup> proton showing COSY correlation to indicated proton.

<sup>d</sup> proton showing HMBC correlations to indicated carbons.

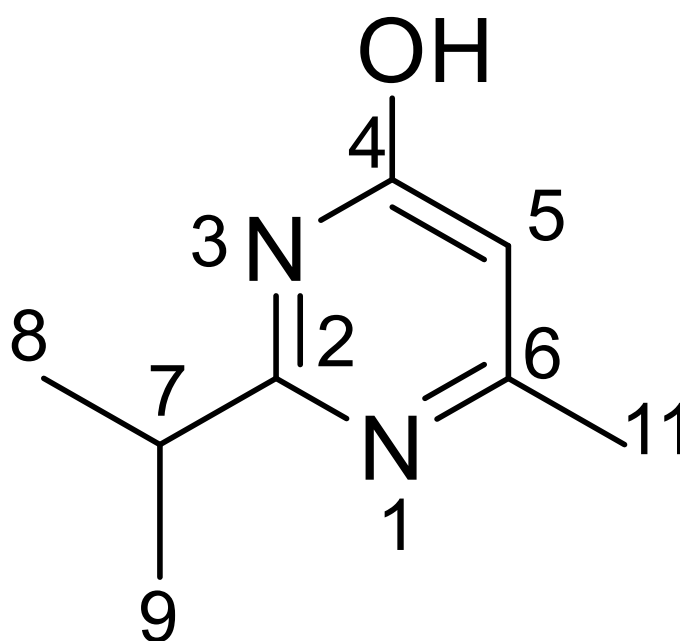
**Figure S5.** Structure and carbon numbering of myxopyrimidinol A.

**Table S12.**  $^1\text{H}$  and  $^{13}\text{C}$  NMR spectral data of commercially purchased 2-isopropyl-6-methyl-4-pyrimidinol (Sigma Aldrich); d: doublet; dd: doublet of doublets; s: singlet; m: multiplet.

Position	$\delta$ $^1\text{H}$ ppm	Multiplicity ( $J=\text{Hz}$ )	$\delta$ $^{13}\text{C}$ ppm
1	---	---	<i>N</i>
2	---	---	162.9
3	---	---	<i>N</i>
4	---	---	164.4
5	5.99	s	109.9
6	---	---	165.9
7	2.77	q (6.8), d (6.8)	33.2
8	1.33	d (7.9)	20.3
9	1.33	d (7.0)	20.3
10	---	---	---
11	2.13	s	23.6

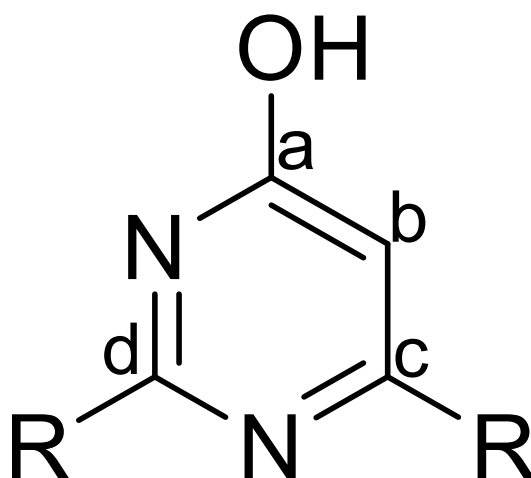
<sup>a</sup> acquired at 175 MHz and assigned from 2D NMR spectra, referenced to solvent signal DMSO  $d_6$  at  $\delta$  39.52 ppm.

<sup>b</sup> acquired at 700 MHz, referenced to solvent signal DMSO  $d_6$  at  $\delta$  2.50 ppm.

**Figure S6.** Structure and carbon numbering of commercially purchased 2-isopropyl-6-methyl-4-pyrimidinol.

Comparison of observed  $^{13}\text{C}$  shifts

Position	$\delta^{13}\text{C}$ ppm myxopyrimidinol A	$\delta^{13}\text{C}$ ppm 2-isopropyl-6-methyl-4- pyrimidinol	$\Delta\delta^{13}\text{C}$ ppm
<b>a</b>	162.9	164.4	1.5
<b>b</b>	109.9	109.9	0
<b>c</b>	167.2	165.9	1.3
<b>d</b>	161.2	162.9	1.7

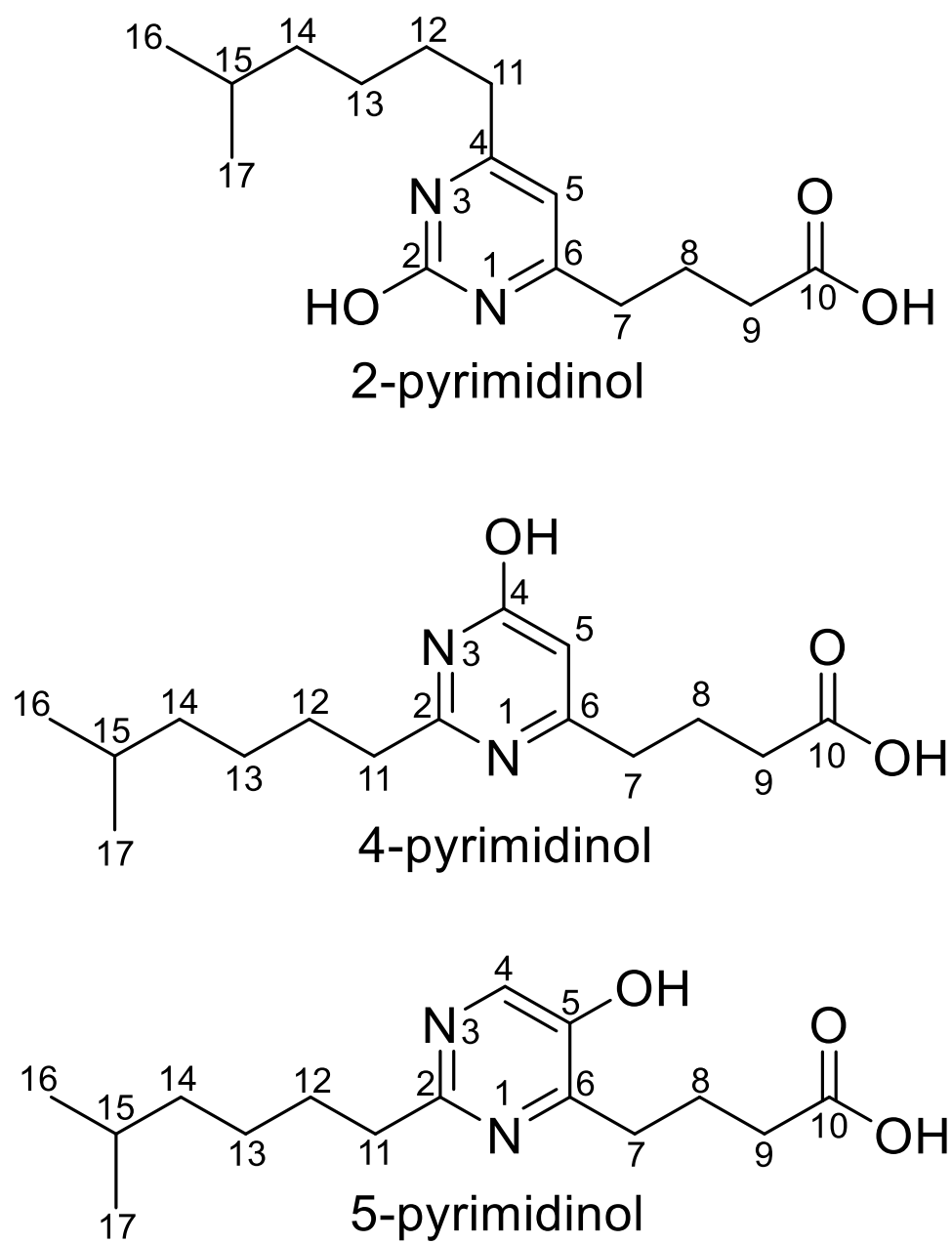
**Figure S7.** Structure and carbon positions of the 4-pyrimidinol scaffold.

## Carbon resonance simulation

In order to further elucidate and underline the structural proposal of myxopyrimidinol A, the ACD/Labs software ACD/ChemSketch 2018.2.5 was used to generate via the implemented function *C+H NMR Predictors and DB* simulated  $^{13}\text{C}$  NMR spectra of hypothetical myxopyrimidinols featuring a 2-pyrimidinol, 4-pyrimidinol or 5-pyrimidinol heterocycle (simply termed “2-pyrimidinol”, “4-pyrimidinol” and “5-pyrimidinol”).

**Table S13.** Simulated  $^{13}\text{C}$  NMR spectral data of 2-pyrimidinol, 4-pyrimidinol and 5-pyrimidinol, and experimentally obtained carbon resonances of myxopyrimidinol A.

Position	$\delta^{13}\text{C}$ ppm “2-pyrimidinol”	$\delta^{13}\text{C}$ ppm “4-pyrimidinol”	$\delta^{13}\text{C}$ ppm “5-pyrimidinol”	$\delta^{13}\text{C}$ ppm myxopyrimidinol A
<b>1</b>	<i>N</i>	<i>N</i>	<i>N</i>	<i>N</i>
<b>2</b>	163.6	163.0	158.2	161.7
<b>3</b>	<i>N</i>	<i>N</i>	<i>N</i>	<i>N</i>
<b>4</b>	168.2	169.8	161.7	162.9
<b>5</b>	108.8	102.9	145.9	109.1
<b>6</b>	168.6	166.5	144.7	167.2
<b>7</b>	37.5	37.5	32.6	35.8
<b>8</b>	24.5	24.3	23.4	23.6
<b>9</b>	34.0	34.0	33.6	34.4
<b>10</b>	176.2	176.2	176.6	175.1
<b>11</b>	37.9	38.4	38.2	34.2
<b>12</b>	29.4	29.6	29.7	27.0
<b>13</b>	26.9	26.9	26.8	26.2
<b>14</b>	38.9	38.7	38.8	38.0
<b>15</b>	28.0	28.0	28.0	27.2
<b>16</b>	22.7	22.7	22.7	22.1
<b>17</b>	22.7	22.7	22.7	22.5



**Figure S8.** Structure and carbon numbering of “2-pyrimidinol”, “4-pyrimidinol” and “5-pyrimidinol”.



### 2.3 IR data

Myxolipoxazole A was re-dissolved in MeOH (approx. 8 mg/mL) and 2  $\mu$ L were samples were measured on the Spectrum™ 100 FT-IR Spectrometer (PerkinElmer) equipped with an UATR accessory (Diamond/ZnSe). Pure methanol was used as a blank for background subtraction. Absorbance spectra were collected between 400  $\text{cm}^{-1}$  and 4000  $\text{cm}^{-1}$  at a spectral resolution of 4  $\text{cm}^{-1}$  and 64 scans were co-added and averaged.

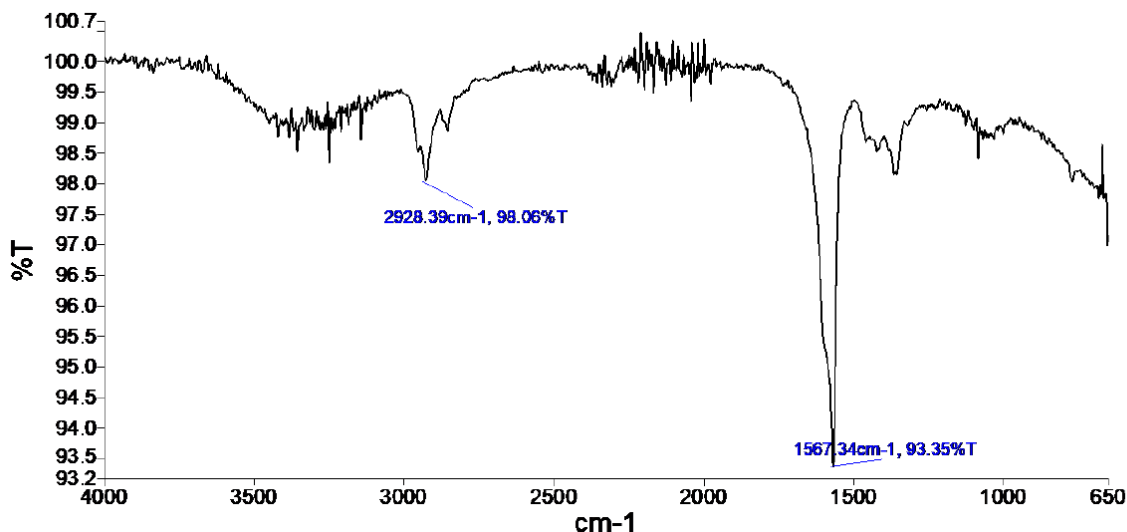


Figure S9. FT-IR spectrum of myxolipoxazole A in MeOH at the range from 4000–650  $\text{cm}^{-1}$ .

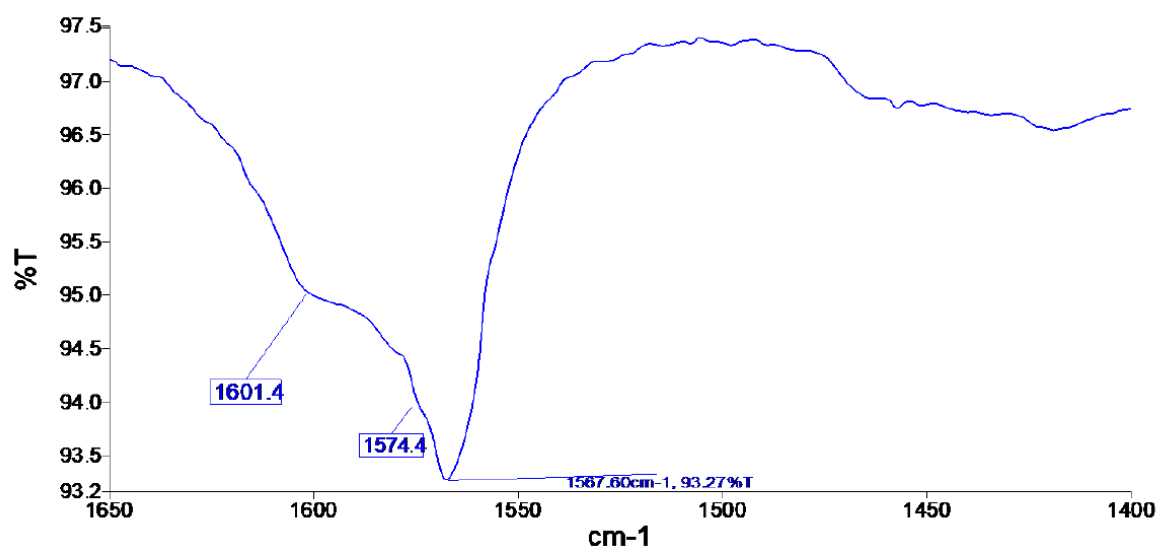
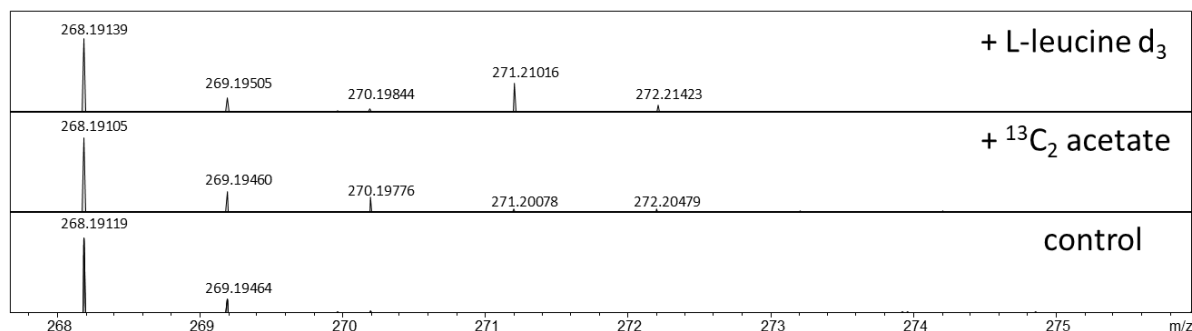


Figure S10. FT-IR spectrum of myxolipoxazole A in MeOH at the range from 1650–1400  $\text{cm}^{-1}$ .

## 2.4 Feeding experiments

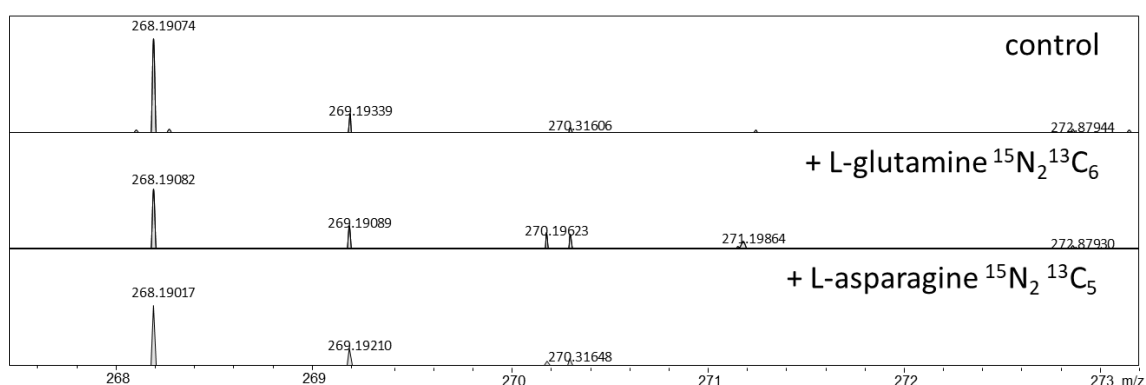
## Partial LC-MS chromatogram: myxolipoxazole A



**Figure S11.** Partial ESI+MS spectra for myxolipoxazole A with L-leucine  $d_3$  and acetate  $^{13}C_2$  supplementation and culture broth without precursor supplementation as control.

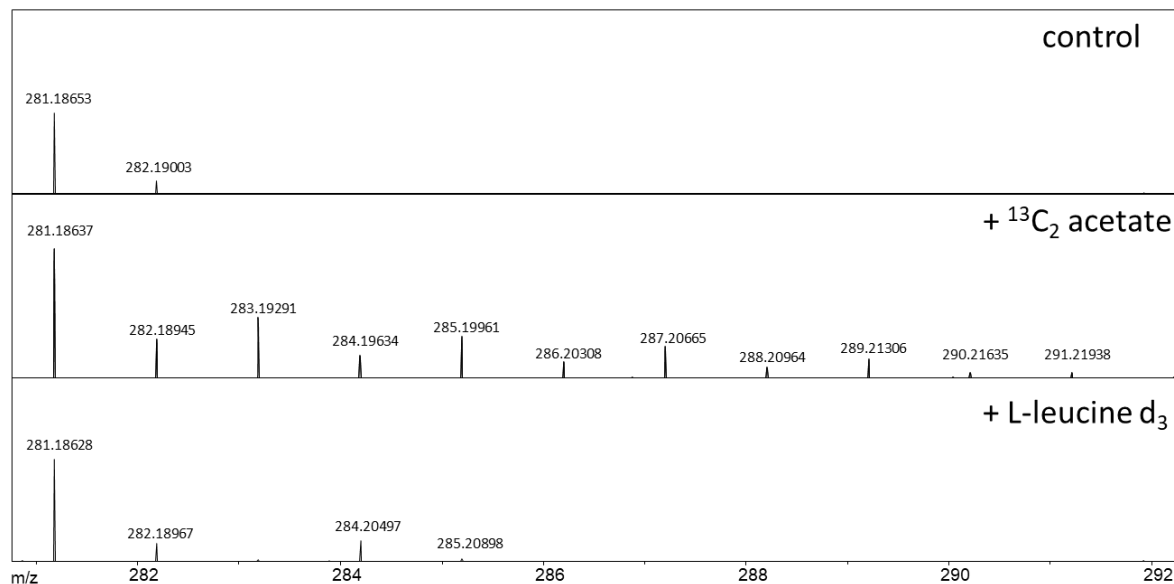
The isotopic patterns of myxolipoxazole A show intensified peaks at  $m/z$  271 and 272 corresponding to a mass shift of +3 Da, caused by the incorporation of L-leucine  $d_3$ . The incorporation of the branched-chain amino acid L-leucine  $d_3$  can be explained through its conversion to isovaleryl-CoA, which serves as a starter unit during the biosynthesis of myxolipoxazole A.

Acetate  $^{13}C_2$  supplementation shows intensified peaks at  $m/z$  270, 271 and 272 in the isotopic patterns of the myxolipoxazole A corresponding to a mass shift of +2 Da.



**Figure S12.** Partial ESI+MS spectra for myxolipoxazole A with L-glutamine  $^{15}N_2^{13}C_6$  and L-asparagine  $^{15}N_2^{13}C_5$  supplementation and culture broth without precursor supplementation as control.

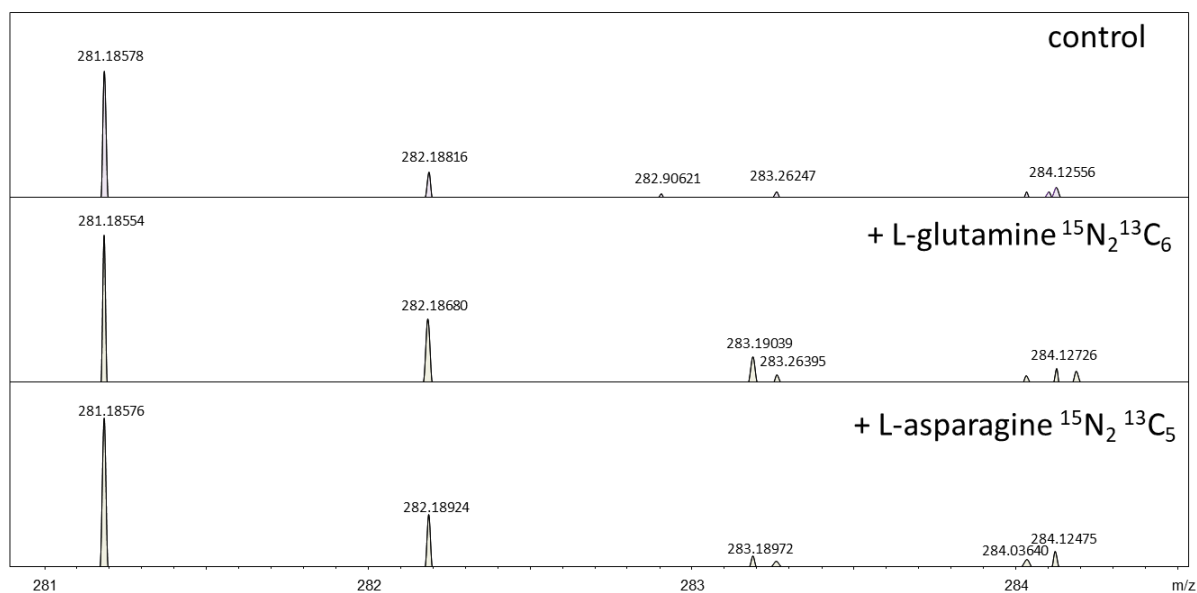
The isotopic patterns of myxolipoxazole A show intensified peaks at  $m/z$  269 and 270 (in particular the changed ratio between 268 and 269) corresponds to a mass shift of +1 Da, caused by the incorporation of L-glutamine  $^{15}N_2^{13}C_6$ . The incorporation of L-asparagine  $^{15}N_2^{13}C_5$  is not unambiguous. The incorporation of L-glutamine  $^{15}N_2^{13}C_6$  can be explained through an aminotransferase-catalyzed incorporation of the only nitrogen in the chemical scaffold of the myxolipoxazole A.

**Partial LC-MS spectrum: myxopyrimidinol A**

**Figure S13.** Partial ESI+MS spectra for myxopyrimidinol with L-leucine  $\text{d}_3$  and acetate  $^{13}\text{C}_2$  supplementation and culture broth without precursor supplementation as control.

The isotopic patterns of myxopyrimidinol A show intensified peaks at  $m/z$  284 and 285 corresponding to a mass shift of +3 Da, caused by the incorporation of L-leucine  $\text{d}_3$ . The incorporation of the branched-chain amino acid L-leucine  $\text{d}_3$  can be explained through its conversion to isovaleryl-CoA, which serves as a starter unit during the biosynthesis of myxopyrimidinol A. Therefore, both the myxolipoxazoles and the myxopyrimidinols are sharing the same building blocks as starter unit for their biosyntheses.

Acetate  $^{13}\text{C}_2$  supplementation shows intensified peaks at  $m/z$  283, 285, 287 and 289 in the isotopic patterns of the myxolipoxazole A corresponding to a mass shift of +8 Da.



**Figure S14.** Partial ESI+MS spectra for myxopyrimidinol A with L-glutamine  $^{15}\text{N}_2\ ^{13}\text{C}_6$  and L-asparagine  $^{15}\text{N}_2\ ^{13}\text{C}_5$  supplementation and culture broth without precursor supplementation as control.

The isotopic patterns of the myxopyrimidinol show intensified peaks at  $m/z$  282 and 283 (in particular the changed ratio between 282 and 283) corresponds to a mass shift of +1 Da, caused by the incorporation of L-glutamine  $^{15}\text{N}_2\ ^{13}\text{C}_6$ . The incorporation of L-asparagine  $^{15}\text{N}_2\ ^{13}\text{C}_5$  is not unambiguous. The incorporation of L-glutamine  $^{15}\text{N}_2\ ^{13}\text{C}_6$  can be explained through an aminotransferase-catalyzed incorporation of one nitrogen in the chemical scaffold of myxopyrimidinol A. The second nitrogen might originate from a different building block.

## 2.5 *In silico* investigation of the myxolipoxazole/myxopyrimidinol biosynthesis

### Genes encoded in the myxolipoxazole/myxopyrimidinol gene cluster (*iso1–9*)

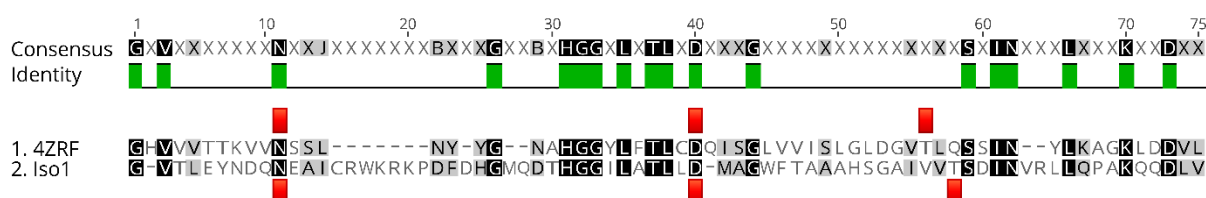
**Table S14.** Table of all open reading frames assigned to the putative myxolipoxazole/myxopyrimidinol gene cluster (*iso1–9*) in *M. xanthus* Mx x48 with proposed function and closest homologue according to blastp search in the nr (non-redundant protein sequences) protein database at NCBI.

name	length (bp)	proposed function	closest homologue	coverage/ identity (%)
<i>iso1</i>	414	thioesterase	WP_108068573.1	99/96
<i>iso2</i>	3651	starter unit tethering FAAL-ACP	WP_158619977.1	99/76
<i>iso3</i>	843	fatty acid desaturase	WP_120624300.1	96/86
<i>iso4</i>	843	iron containing redox enzyme	WP_143901697.1	99/85
<i>iso5</i>	2766	PKS biosynthesis KS-AT	WP_108069076.1	88/77
<i>iso6</i>	3288	PKS/NRPS biosynthesis ACP-Aminotransferase-C- domain	WP_143901701.1	99/75
<i>iso7</i>	375	ACP	WP_161665324.1	99/69
<i>iso8</i>	1581	Adenylation domain	WP_161665323.1	99/85
<i>iso9</i>	7383	PKS biosynthesis KS-AT-DH-ER-KR-ACP-TE	WP_143901706.1	98/79

The identified biosynthetic gene cluster was also found in the genome sequence of the myxobacterial strains *Coralloccoccus* sp. Z5C101001, *Coralloccoccus sicarius* sp. nov. (CA040B<sup>T</sup> = DSM 108850<sup>T</sup> = NBRC 113890<sup>T</sup>)<sup>16</sup>, *Vitiosangium* sp. GDMCC 1.1324<sup>17</sup> and *Coralloccoccus* sp. c25j21. The comparison of these strains assisted to determine precisely the exact biosynthetic gene cluster borders, since upstream of *iso1* and downstream of *iso9* the genetic region is significantly different between the different producers.

*iso1*

The gene *iso1* encodes a thioesterase (TE) which is a member of the thioesterase family 13 (TE13)<sup>18</sup> such as the phenylacetic acid degradation protein PaaI<sup>19,20</sup>. Typical TEs in PKS and NRPS pathways are classified as type I thioesterase (TE I; thioesterase family 16) integrated in the multimodular architecture of FASs, PKSs, or NRPs such as the TE in Iso9 or type II thioesterase (TE I; thioesterase family 18) which are independently working TEs<sup>21</sup>. These TEs feature a  $\alpha/\beta$ -hydrolase fold<sup>22</sup>, whereas the structure of Iso1 putatively exhibits a Hotdog fold<sup>23</sup> such as other members of TE13. Gene disruption of *iso1* abolished the production of all myxolipoxazole and myxopyrimidinol derivatives implying the fundamental role of Iso1 for their biosynthesis. The reason for that finding remains to some extent elusive, since gene disruption of *iso9* still produces the truncated derivative **3** and reveals **4**. One explanation would be, that the intermediates have to be released from module 3 before further modification of module 4 can be achieved, or the genetic disruption of *iso1* interferes with the genetic region of the promoter and disable successful expression of the downstream gene operon. The thioesterase from *Streptococcus pneumoniae* (PDB: 4ZRF) *SpPaaI* revealed a dual specificity for phenylacetyl-CoA and medium-chain fatty acyl-CoAs (C<sub>4</sub>–C<sub>14</sub>, C<sub>10</sub> highest turnover)<sup>24</sup>, which would fit with the expected substrate specificity of Iso1. In addition the putative active site residues of *SpPaaI* (Asn<sup>37</sup>, Asp<sup>52</sup>, and Thr<sup>68</sup>) which are important for catalysis, were also found in Iso1 with small deviations (insertion of seven amino acids between Asn<sup>37</sup> and Asp<sup>52</sup>, and the residue Thr<sup>68</sup> is presumably shifted by two amino acids) (**Figure S15**).



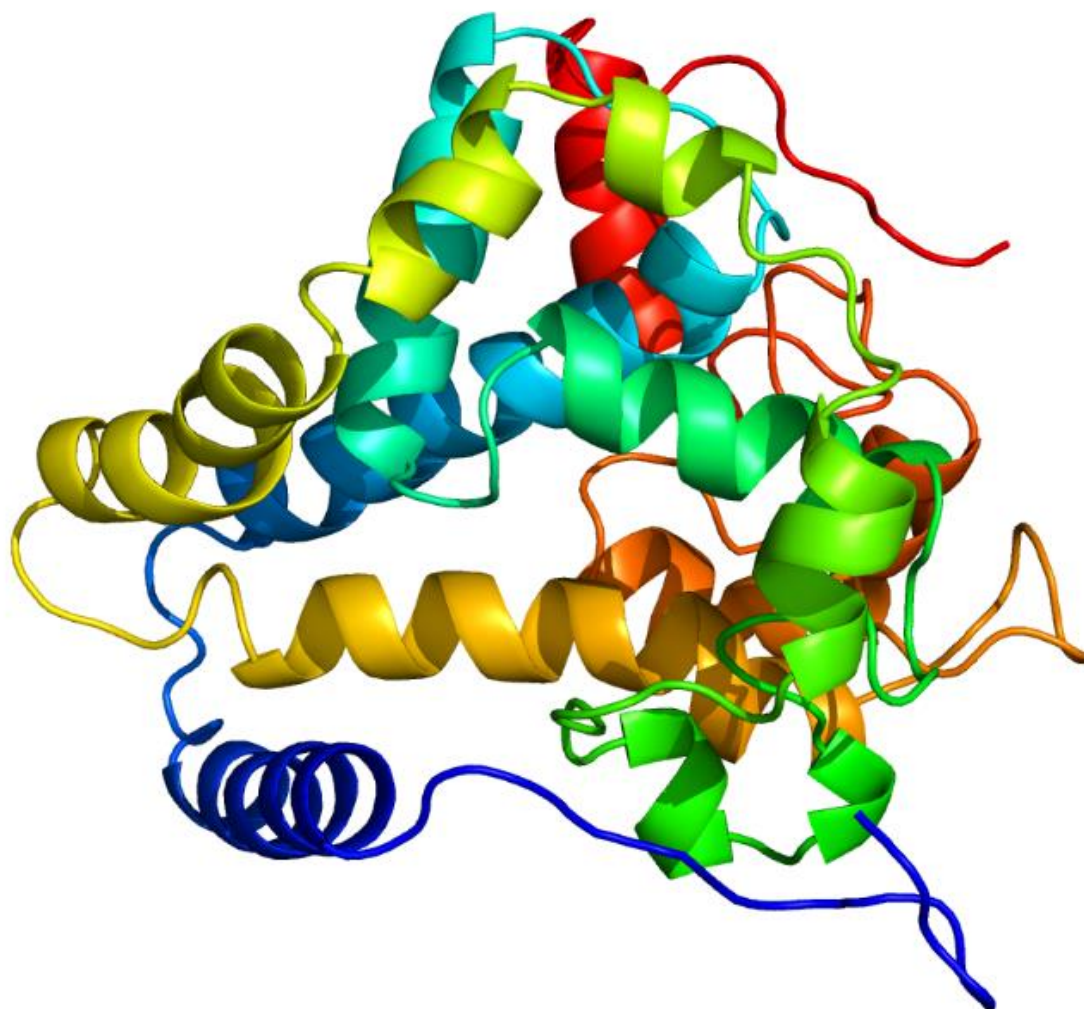
**Figure S15.** Amino acid sequence alignment of *SpPaaI* (labeled as 4ZRF) and Iso1. The red boxes indicate from left to right the active site residues of Asn<sup>37</sup>, Asp<sup>52</sup>, and Thr<sup>68</sup>.

*iso2*

The gene *iso2* encodes a fatty acyl AMP ligase (FAAL), an acyl CoA dehydrogenase (DH) domain and an ACP domain. The FAAL shows the highest structural similarity according Phyre 2 analysis to the FAAL from *L. pneumophila* (*LpFAAL*, PDB: 3KXW)<sup>25</sup>. The FAAL in Iso2 features the insertion motif (distinguishing it from FACLS), the hinge region, the gate motif and the highly conserved loop region in the C-domain, which are interacting with each other for substrate binding and product release<sup>26</sup>. The product is then transferred to the ACP on Iso2. The ACP is not featuring the motif with the catalytic serine of (D/E)xGxDSL or any related DSL/R motif<sup>27</sup>. A similar ACP is described for AmbG, responsible for the biosynthesis of ambruticin<sup>28, 29</sup>. The optional  $\beta$ -oxidation, of the fatty acid substrate, is catalyzed by the Acyl-CoA DH in Iso2. The Acyl-CoA DH has the highest sequence similarity with the Acyl-CoA DH NocC in *Nostoc* sp. CCAP 1453/38<sup>30</sup>.

*iso3*

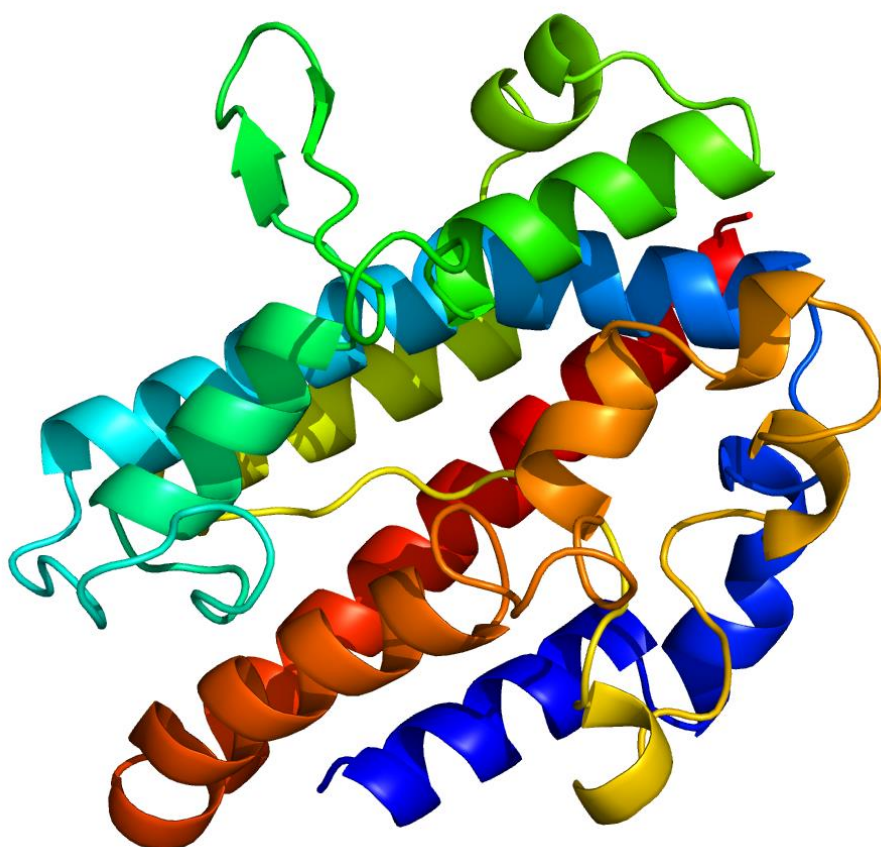
The amino acid sequence and proposed structure *in silico* prediction of Iso3 shows the closest similarity to stearyl-CoA desaturases (c4zyoA, c4ymkA) (**Figure S16**).



**Figure S16.** Phyre-2 structure homology model of the putative stearyl-CoA desaturase encoded by *iso3* (model based on template c4zyoA; figure colored by rainbow *N* → *C* terminus).

*iso4*

The gene *iso4* putatively encodes a cryptic protein, which might be responsible for the hydroxylation of the generated free secondary amine. Iso4 shows low sequence similarity with the imidazolonepropionase (BBA66509.1) from the biosynthesis of actinoallolides<sup>31</sup> and low structural similarity to crystal structure of a tena/thi-4 domain-containing protein from *Sulfolobus solfataricus* (PDB: 4LQX) and the oxygenase PqqC. The proposed mechanism could work by using molecular oxygen to generate the oxime<sup>32</sup>.



**Figure S17.** Phyre-2 structure homology model of a putative tena/thi-4 domain-containing protein (oxidoreductase, PDB: 4LQX) encoded by *iso4* (model based on template d1otva; figure colored by rainbow *N* → *C* terminus).



*iso5*

The gene *iso5* encodes a KS and AT domain responsible for conventional polyketide chain extension incorporating malonyl-CoA building blocks. The active site residue of the AT domain of Iso5 and Iso9 (**Tab. S13 and Tab. S14**) and clearly indicates that the extension building block has to be malonyl-CoA rather than being methylmalonate CoA or any other unusual substrate. The reason why Iso5 might act iteratively as hypothesized remains elusive.

**Table S15.** Acyl transferase (AT) active residues that correlate with domain specificity.

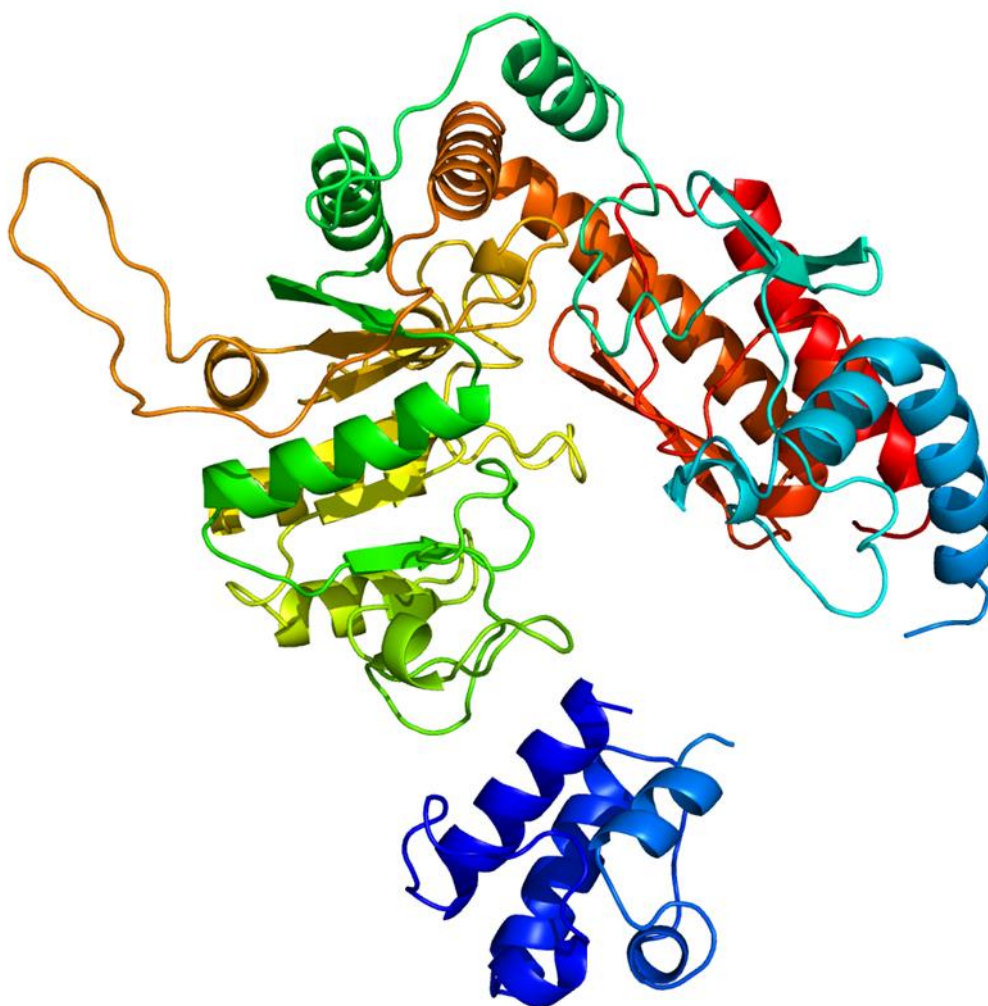
Domain /residue	11	63	90	91	92	93	94	117	200	201	231	250	255
Methylmalonate cons.	Q	Q	G	H	S	QMI	G	R	S	H	T	NS	V
malonate cons.	Q	Q	G	H	S	LVIFAM	G	R	FP	H	ANTGEDS*	NHQ	V
epoC_AT_1	Q	Q	G	H	S	I	G	R	F	H	N	H	V
epoA_AT_1	Q	Q	G	H	S	I	G	R	F	H	N	H	V
epoB_AT_1	Q	Q	G	H	S	M	G	R	S	H	N	H	V
Iso5 AT	Q	Q	G	H	S	L	G	R	F	H	T	N	V
Iso9 AT	Q	Q	G	H	S	V	G	R	F	H	N	H	V

**Table S16.** Acyl transferase (AT) additional conserved residues that correlate with domain specificity.

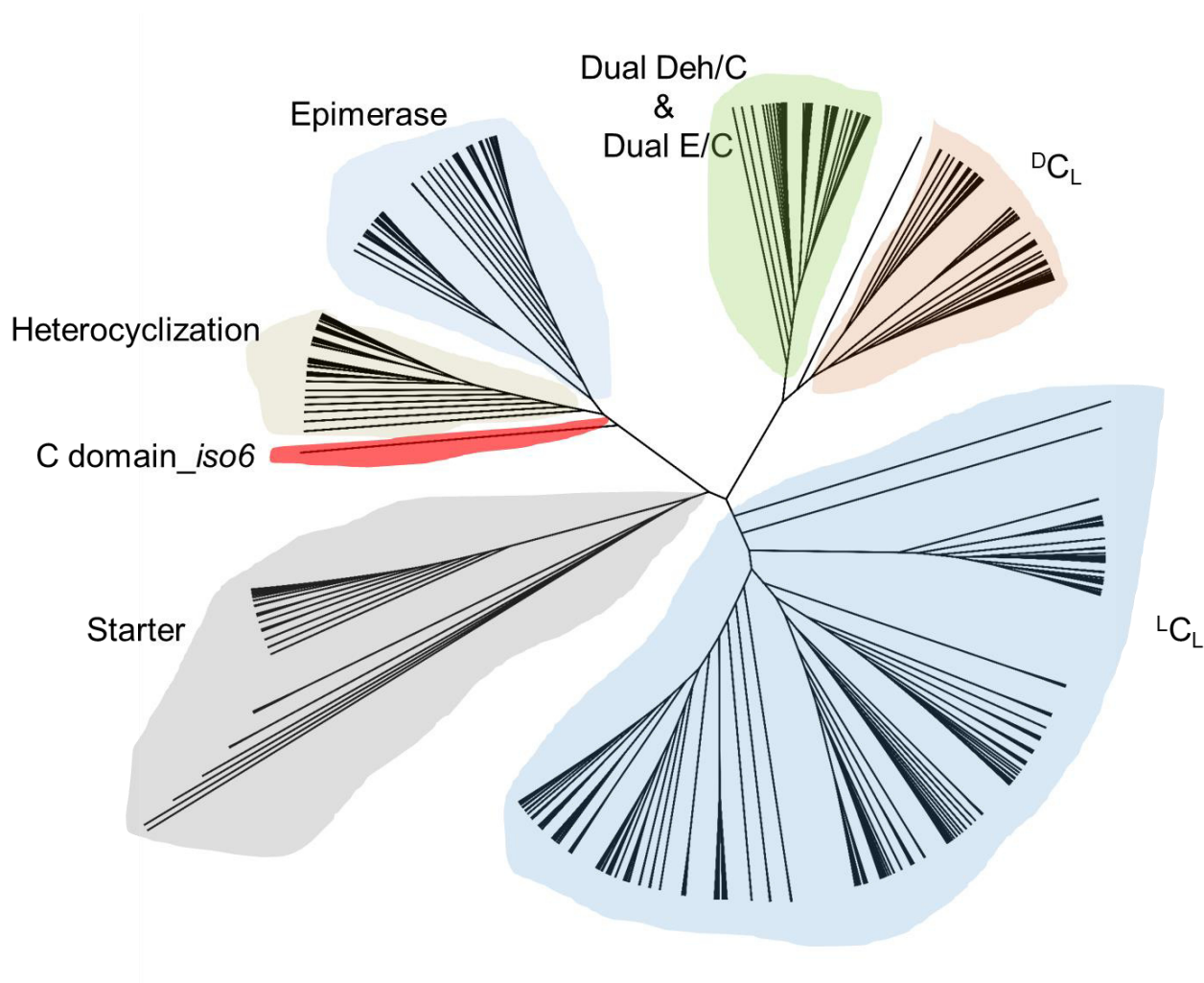
Domain/residue	15	58	59	60	61	62	70	72	197	198	199
Methylmalonate cons.	W	REDQ	VIDA	D	V	VLI	MAEQ	SAG	D	YV	A
Malonate cons.	R	RQED	T	GRLE	YFW	TAS	EQ	AG	SN	H	A
epoC_AT_1	T	E	T	A	F	T	E	A	S	H	A
epoA_AT_1	T	Q	T	A	F	T	E	A	S	H	A
epoB_AT_1	W	R	I	D	V	V	A	A	D	V	A
Iso5 AT	F	D	T	H	A	A	E	A	R	H	A
Iso9 AT	Y	Q	T	K	Y	T	E	A	S	H	A

*iso6*

*iso6* encodes module 3, harboring an ACP, an aminotransferase domain and a C domain. The ACP features the typical D/ExGxDSL motif with the catalytically active serine. The ACP-aminotransferase didomain architecture shows striking similarity to the ACP-Amt fusion domain found in the mycosubtilin pathway (PDB: c6kfuA) (**Figure S18**). The C domain found in Iso6 is hypothesized to perform the cyclization of the generated oximes to yield the isoxazole/4-pyrimidinol scaffold. Phylogenetic analysis of this C domain reveals that it might represent a distinct novel clade besides the six C domain subtypes distinguished by Rausch *et al.*<sup>33</sup> and the new C domain subtype described by Pogorevc *et al.*<sup>34</sup>. However, the C domain in Iso6 features the common catalytic motif HHxxxDG and is not replaced through the motif DxxxxDxxS which is described for several heterocyclization domains<sup>35</sup>.



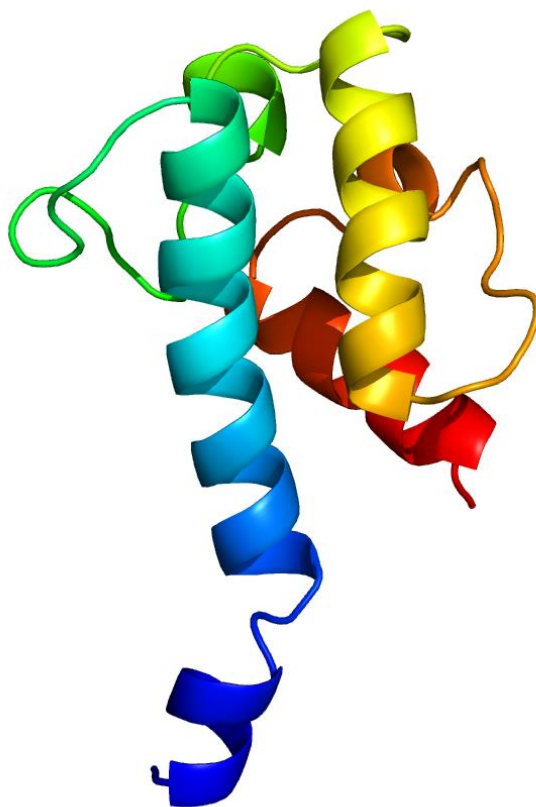
**Figure S18.** Phyre2 structure homology model of the ACP-Amt di-domain encoded by *iso6* (model based on template c6kfuA4; figure colored by rainbow *N* → *C* terminus).



**Figure S19.** Phylogenetic tree of different C domain subtypes. Phylogenetic tree of all known C domain subtypes ( $^L C_L$ ,  $^D C_L$ , starter, dual E/C, dual Deh/C, epimerization and heterocyclization domains). The phylogeny was reconstructed by using phym1, a gamma-distributed rate variation with four categories and employing the JTT model of amino acid substitution. Support values are based on 100-fold bootstrapping. The C domain list includes in total 538 domains from phylogenetic studies by Rausch *et al.*<sup>33</sup> and Pogorevc *et al.*<sup>34</sup>. The C-domain from the NRPS module Iso6 seems to be phylogenetically closely related to heterocyclization domains such as described for epothilone or myxothiazol biosynthesis<sup>36</sup>.

### *iso7*

*iso7* encodes a standalone ACP domain. The ACP domain shows similarity with the structure of the acyl carrier 2 protein domain from module 2 of the 6-deoxyerythronolide b<sub>3</sub> synthase<sup>37</sup> (**Figure S20**).



**Figure S20.** Phyre2 structure homology model of the standalone ACP-Amt encoded by *iso7* (model based on template c2ju2A; figure colored by rainbow *N*→*C* terminus).

**iso8**

*iso8* encodes a standalone adenylation (A) domain. The specificity-conferring code of adenylation domains in NRPSs (Stachelhaus code) implies that the substrate of Iso8 cannot be an  $\alpha$ -amino acid, since residue 235 features a histidine rather an aspartic acid (**Table S17**). The exact role of Iso8 remains elusive, but the function of this adenylation domain is essential, since genetic disruption of *iso8* abolishes the production of all myxolipoxazoles and myxopyrimidinols in the secondary metabolome of *M. xanthus* Mx x48. It might eventually be responsible for the incorporation of the second nitrogen in the 4-pyrimidinol heterocycles of myxopyrimidinols.

**Table S17.** Stachelhaus code sequence of different A domains including the standalone A domain Iso8.

Name /position	235	236	239	278	299	301	322	330	331	517	building block
Iso8	H	V	W	A	V	G	N	N	M	K	---
AIL50189.1 (BGC0000213)	F	G	C	G	F	G	A	I	C	K	aminolevulinic acid
AMU1	D	A	W	T	I	A	A	I	C	K	Phe
cdaps003_A_002_1amu	D	G	W	A	V	A	S	V	C	K	Trp
ATY72527.1 (BGC0001574)	V	G	W	C	V	G	I	A	C	K	fatty acyl
natF (BGC0001695)	H	G	V	W	L	Y	V	G	P	K	2-amino benzoic acid
AZF85933.1 (BGC0001963)	I	G	W	C	A	G	L	N	T	K	fatty acyl

Accession number according the MIBiG database are shown in brackets.

**iso9**

*iso9* encodes a PKS module consisting of an AT, ACP and KS domain equipped with the full set of accessory domains (KR, DH and ER) to catalyze a full reductive loop onto the  $\beta$ -keto thioester during condensation. *In silico* analysis and the conducted genetic disruption experiment of *iso9* confirmed the functionality of the last module, which provides optional backbone extension of the myxolipoxazoles and myxopyrimidinols.

## 2.6 Biosynthetic investigation and genetic manipulation of the myxolipoxazole/myxopyrimidinol biosynthesis in the producer *M. xanthus* Mx x48

### Generation of gene-disruption and induced-gene mutants of *M. xanthus* Mx x48 to connect the identified genes to the production of myxolipoxazole/myxopyrimidinol

A 317–1373 bp homology sequence of the genes *iso1*, *iso2*, *iso3*, *iso4*, *iso5*, *iso6*, *iso8* and *iso9* have been PCR amplified by the primers as shown in **Table S2**. The specific homology sequence was subcloned via conventional restriction ligation (HindIII, SpeI) into the pCR2.1 vector from the TOPO-TA cloning kit (Thermo scientific TOPO-TA cloning Kit). Similarly to the construction of the disruption constructs, a 1117 bp homology sequence starting from the translational start (not the RBS, but the coding sequence) of the gene *iso2* encoding the starter module has been PCR amplified by the primers as shown in **Table S2**. The homology sequence was subcloned via conventional restriction ligation into the pFP<sub>van</sub> vector (NdeI, EcoRI), which has been constructed and utilized previously to express several genes in myxobacteria<sup>8,38,39</sup>. The pFP<sub>van</sub> vector is a derivative of the pCR2.1 vector featuring a vanillate-inducible promoter which is fused to the vanillate-responsive repressor as described in literature<sup>40</sup>.

### Transfer and chromosomal integration of the constructs into the host *M. xanthus* Mx x48

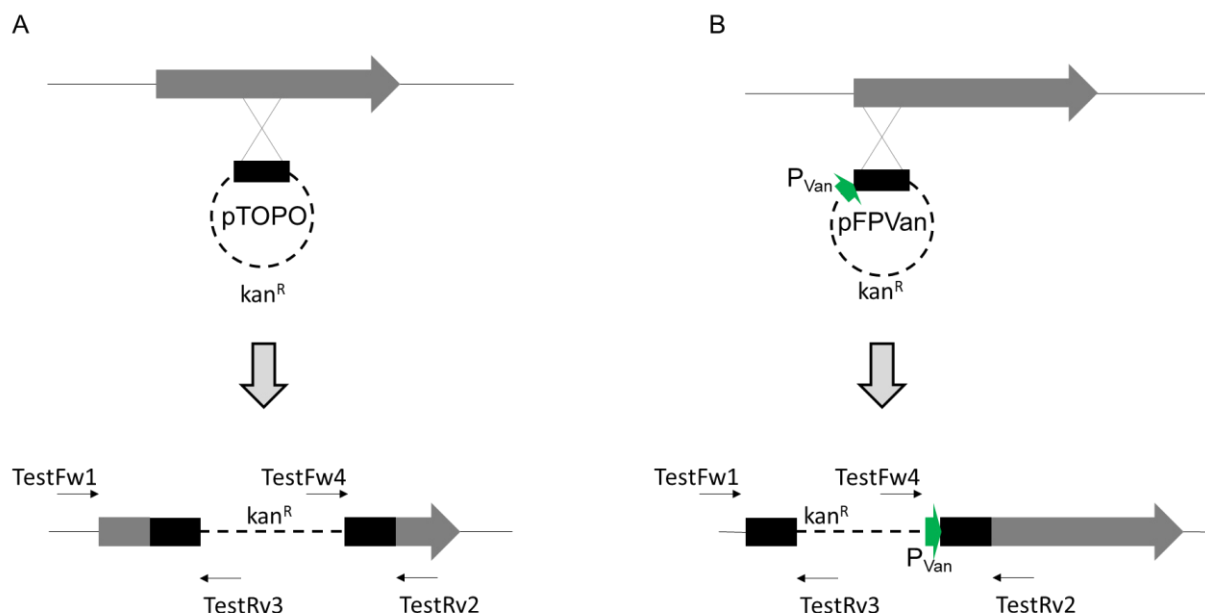
According to a previously established electroporation procedure for *Myxococcus xanthus* DK1622<sup>41</sup> the strain *M. xanthus* Mx x48 was transformed with the generated disruption and induced-gene constructs (**Table S6**, genetic constructs 1–6). *M. xanthus* Mx x48 transformants were routinely cultivated at 30 °C in CTT medium or CTT agar. Liquid cultures were grown in Erlenmeyer flasks on an orbital shaker at 180 rpm for 3–6 days. *M. xanthus* Mx x48 transformants were selected by adding 50 µg/mL kanamycin to the fermentation culture. Correct chromosomal integration of the expression constructs via homologous recombination into the site-specific locus was confirmed by PCR (Figure S21). PCRs were performed according to the settings described above. Genomic DNA of the transformants were isolated using the Gentra® Puregene® Yeast/Bact. Genomic DNA Purification Kit (Qiagen) according to manufacturer's instructions. For each expression construct, correct chromosomal integration was confirmed using two different primer combinations revealing PCR products of the expected sizes:

- Construct No. 1, Seq. primer No.1/3 (1495 bp), and primer No.4/2 (1609 bp)
- Construct No. 2, Seq. primer No.5/7 (1844 bp), and primer No.8/6 (1836 bp)
- Construct No. 3, Seq. primer No.9/11 (1659 bp), and primer No.12/10 (1449 bp)
- Construct No. 4, Seq. primer No.13/15 (1561 bp), and primer No.16/14 (1548 bp)
- Construct No. 5, Seq. primer No.17/19 (1789 bp), and primer No.20/18 (2037 bp)
- Construct No. 6, Seq. primer No.21/23 (1928 bp), and primer No.24/22 (2079 bp)

- Construct No. 7, Seq. primer No.25/27 (2032 bp), and primer No.28/26 (1533 bp)
- Construct No. 8, Seq. primer No.29/31 (1970 bp), and primer No.32/30 (1859 bp)
- Construct No. 9, Seq. primer No.33/35 (1620 bp), and primer No.36/34 (1745 bp)

Genomic DNA of *M. xanthus* Mx x48 was used as negative control. A complementary experiment using the following primer combinations revealed a specific PCR product for *M. xanthus* Mx x48, but not for any of the transformants of *M. xanthus* Mx x48 harboring one of the generated constructs.

- Construct No.1, Seq. primer No.1/2 (1716 bp PCR product for *M. xanthus* Mx x48 wild type)
- Construct No.2, Seq. primer No. 5/6 (1442 bp PCR product for *M. xanthus* Mx x48 wild type)
- Construct No.3, Seq. primer No. 9/10 (1634 bp PCR product for *M. xanthus* Mx x48 wild type)
- Construct No.4, Seq. primer No. 13/14 (1455 bp PCR product for *M. xanthus* Mx x48 wild type)
- Construct No.5, Seq. primer No.17/18 (1627 bp PCR product for *M. xanthus* Mx x48 wild type)
- Construct No.6, Seq. primer No. 21/22 (1681 bp PCR product for *M. xanthus* Mx x48 wild type)
- Construct No.7, Seq. primer No.25/26 (1549 bp PCR product for *M. xanthus* Mx x48 wild type)
- Construct No.8, Seq. primer No. 29/30 (1487 bp PCR product for *M. xanthus* Mx x48 wild type)
- Construct No.9, Seq. primer No. 33/34 (1754 bp PCR product for *M. xanthus* Mx x48 wild type)



**Figure S21.** Genotypic verification procedure using PCR. Correct chromosomal integration of the genetic disruption (A) and induced-gene expression constructs were verified by multiplex PCR using four primers: two integrations site-specific primers (TestFw1 and TestRv2) and two vector specific primers (Test Rv3 and Test Fw4).

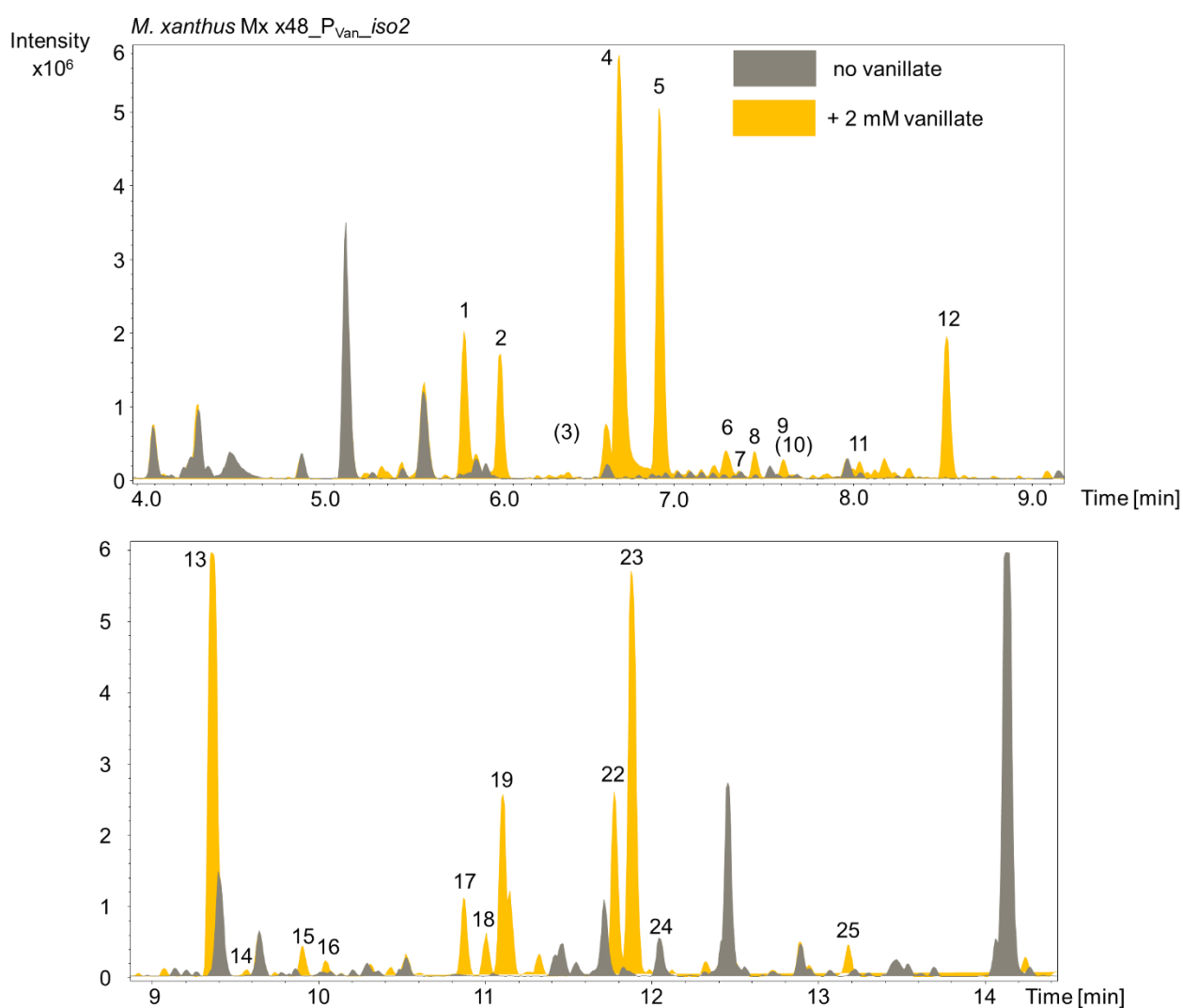
## Analysis of secondary metabolism of genetic mutants

Table S18. Secondary metabolites associated with BGC *iso1–9* in *M. xanthus* Mx x48.

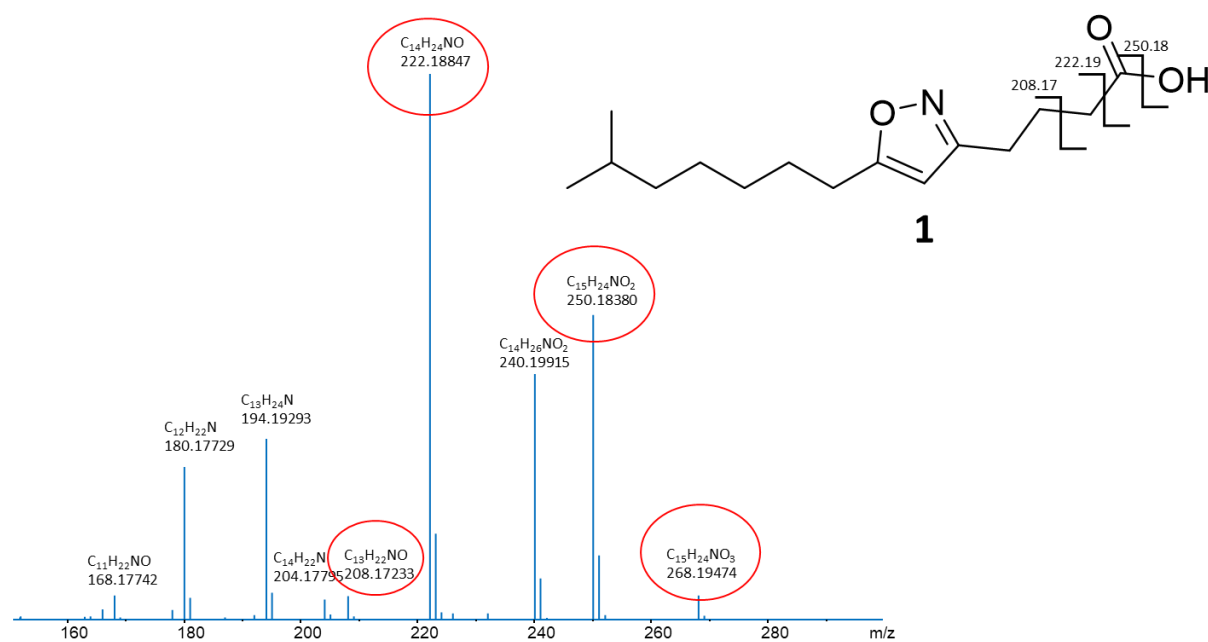
No.	Retention time [min]	Mass [M+H]	Sum formula	Identification	Metabolite
1	5.83	270.17022 (252.15864)	C <sub>14</sub> H <sub>24</sub> NO <sub>4</sub> (C <sub>14</sub> H <sub>22</sub> NO <sub>3</sub> )	<i>iso2</i> activation	---
2	6.03	267.17060	C <sub>14</sub> H <sub>23</sub> N <sub>2</sub> O <sub>3</sub>	<i>iso2</i> activation	---
3	6.43	253.1553	C <sub>13</sub> H <sub>20</sub> N <sub>2</sub> O <sub>3</sub>	<i>iso9</i> disruption; not produced through <i>iso2</i> activation	myxopyrimidinol B
4	6.70	284.18599 (266.17454)	C <sub>15</sub> H <sub>26</sub> NO <sub>4</sub> (C <sub>15</sub> H <sub>24</sub> NO <sub>3</sub> )	<i>iso2</i> activation	---
5	6.93	281.18557	C <sub>21</sub> H <sub>25</sub> N <sub>2</sub> O <sub>3</sub>	<i>iso9</i> disruption and <i>iso2</i> activation	myxopyrimidinol A
6	7.29	266.17542	C <sub>15</sub> H <sub>24</sub> NO <sub>3</sub>	<i>iso2</i> activation	--
7	7.38	194.1544	---	<i>iso9</i> disruption & <i>iso2</i> activation	---
8	7.41	228.19573	C <sub>13</sub> H <sub>26</sub> NO <sub>2</sub>	<i>iso2</i> activation	---
9	7.61	325.21214	C <sub>17</sub> H <sub>29</sub> N <sub>1</sub> O <sub>4</sub>	<i>iso2</i> activation	---
10	7.63	518.2861	---	<i>iso9</i> disruption	---
11	8.04	546.3182	---	slightly increased through <i>iso2</i> activation	---
12	8.52	270.17027 (252.15920)	C <sub>14</sub> H <sub>24</sub> NO <sub>4</sub> (C <sub>14</sub> H <sub>22</sub> NO <sub>3</sub> )	<i>iso2</i> activation	---
13	9.37	284.18612 (266.17462) (248.16495)	C <sub>15</sub> H <sub>26</sub> NO <sub>4</sub> C <sub>15</sub> H <sub>24</sub> NO <sub>3</sub> C <sub>15</sub> H <sub>22</sub> NO <sub>2</sub>	<i>iso2</i> activation	---
14	9.57	282.1706	C <sub>15</sub> H <sub>24</sub> NO <sub>4</sub>	<i>iso2</i> activation	---
15	9.91	268.19042	C <sub>15</sub> H <sub>26</sub> NO <sub>3</sub>	<i>iso2</i> activation	---
16	10.05	250.14307 268.15369	C <sub>14</sub> H <sub>20</sub> NO <sub>3</sub> C <sub>14</sub> H <sub>22</sub> NO <sub>4</sub>	<i>iso2</i> activation	---
17	10.88	282.17043 (264.15929)	C <sub>15</sub> H <sub>24</sub> NO <sub>4</sub> (C <sub>15</sub> H <sub>22</sub> NO <sub>3</sub> )	<i>iso2</i> activation	---
18	11.01	254.17528	C <sub>14</sub> H <sub>24</sub> NO <sub>3</sub>	<i>iso2</i> activation	---
19	11.11	254.17555	C <sub>14</sub> H <sub>24</sub> NO <sub>3</sub>	<i>iso2</i> activation	---
20	11.18	240.15910	C <sub>13</sub> H <sub>22</sub> NO <sub>3</sub>	<i>iso9</i> disruption and increase through <i>iso2</i> activation	myxolipoxazole B
21	11.33	266.12050	C <sub>17</sub> H <sub>16</sub> NO <sub>2</sub>	<i>iso2</i> activation	---



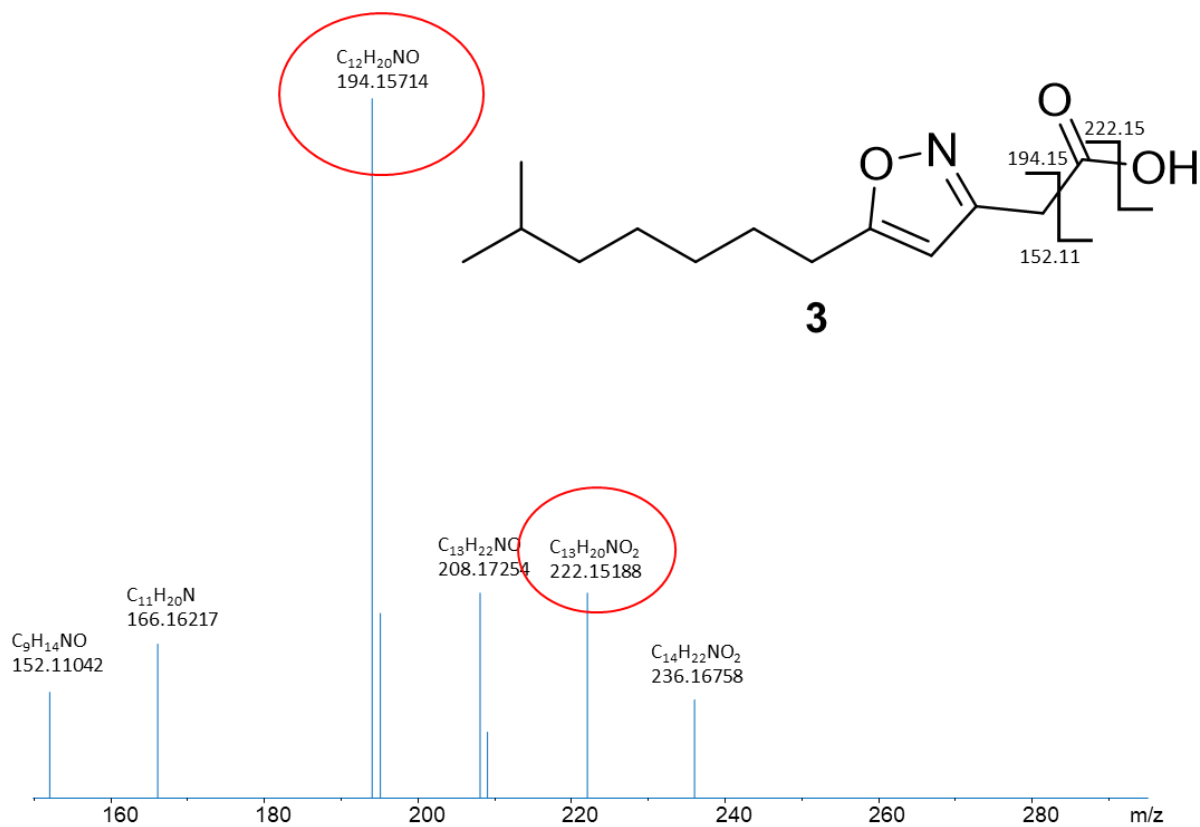
				insensitive knockouts	to
22	11.80	268.19192	C <sub>15</sub> H <sub>26</sub> NO <sub>3</sub>	myxolipoxazole A	myxolipoxazole A
23	11.91	268.19192	C <sub>15</sub> H <sub>26</sub> NO <sub>3</sub>	myxolipoxazole A	myxolipoxazole A
24	12.03	266.1757	C <sub>15</sub> H <sub>24</sub> NO <sub>3</sub>	myxolipoxazole double bond derivative	---
25	13.18	200.20087	C <sub>12</sub> H <sub>26</sub> NO	<i>iso9</i> disruption	---



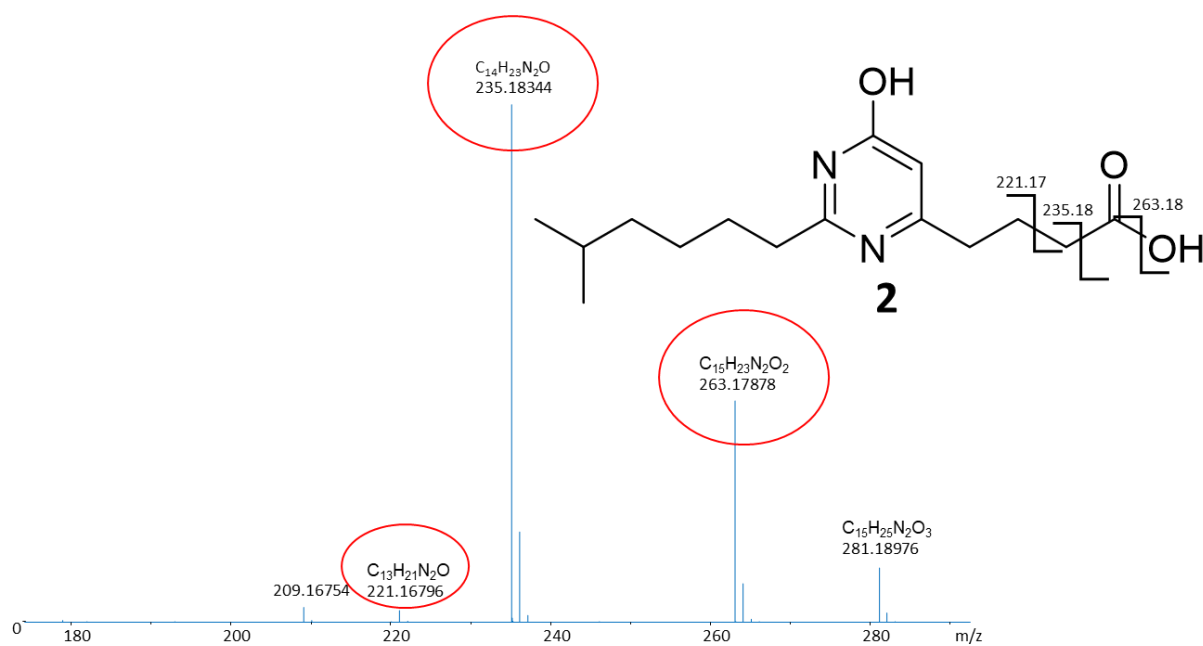
**Figure S22.** HPLC-MS BPC of *M. xanthus* Mx x48\_P<sub>van</sub>\_iso2 with and without supplementation of 2 mM vanillate. The chromatogram reveals tremendously increased production of many secondary metabolites associated with the *iso* BGC. All identified and correlated secondary metabolites are labeled with a number and listed in **Table S18**. Secondary metabolites labeled with a number in brackets are only present in the crude extract of *M. xanthus* Mx x48\_iso9\_disruption.



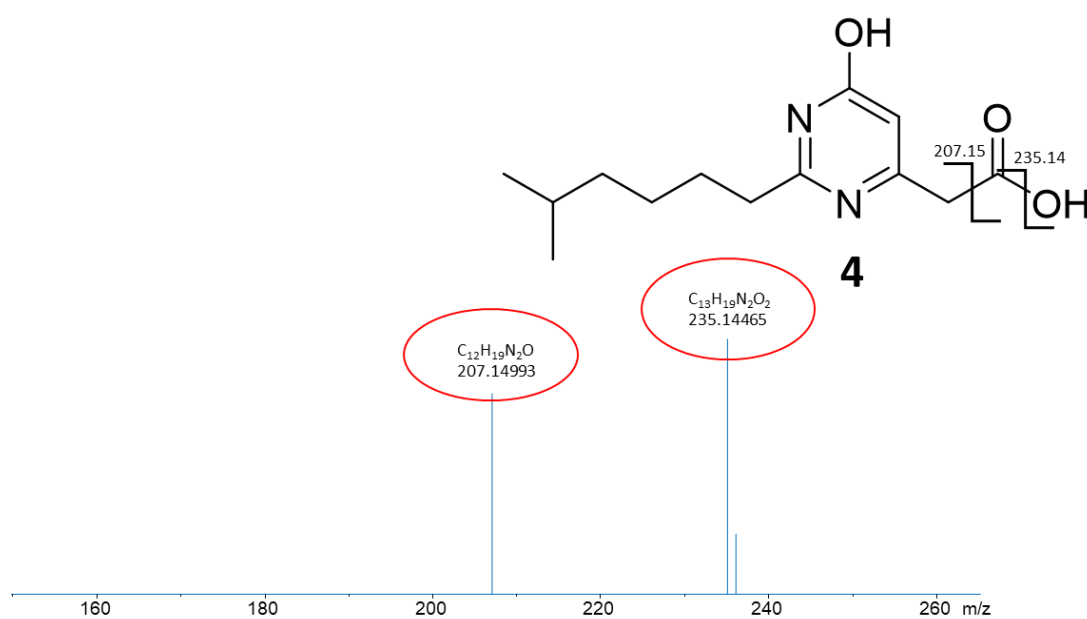
**Figure S23.** MS<sup>2</sup>-spectrum and fragmentation pattern of myxolipoxazole A (1). Red rings indicate specific fragments.



**Figure S24.** MS<sup>2</sup>-spectrum and fragmentation pattern of myxolipoxazole B (3). Red rings indicate specific fragments.



**Figure S25.** MS<sup>2</sup>-spectrum and fragmentation pattern of myxopyrimidinol A (**2**). Red rings indicate specific fragments.



**Figure S26.** MS<sup>2</sup>-spectrum and fragmentation pattern of myxopyrimidinol B (**4**). Red rings indicate specific fragments.

## 2.7 Biological function of myxolipoxazoles and myxopyrimidinols

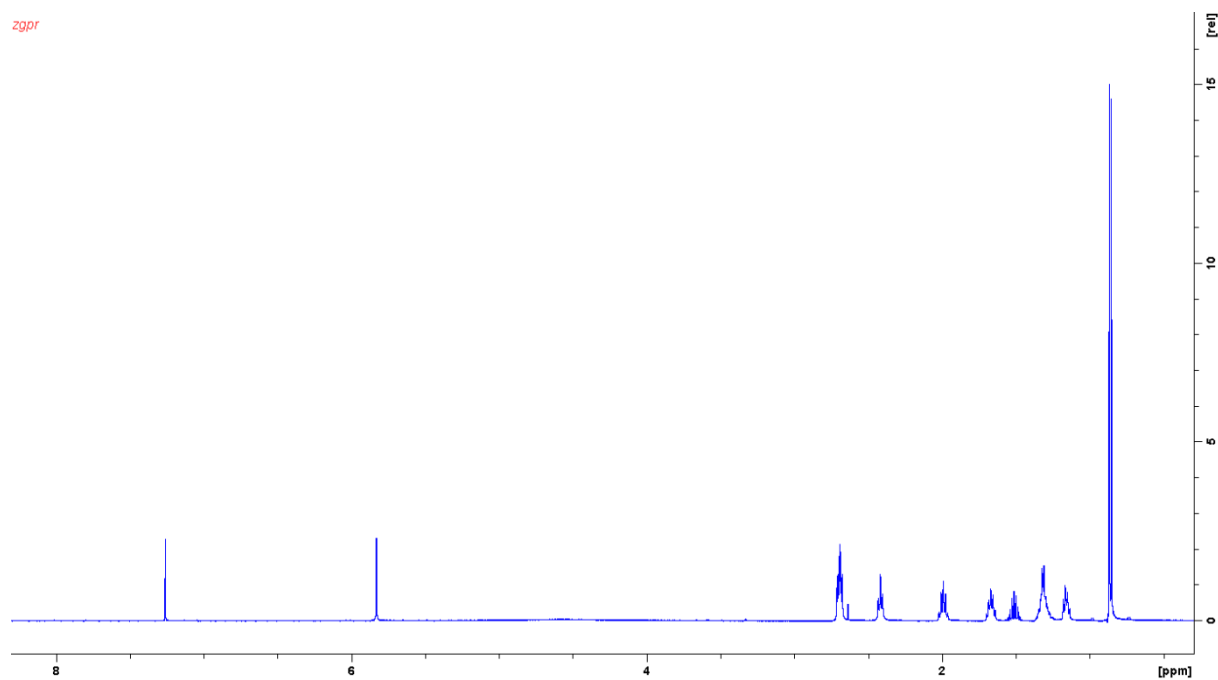
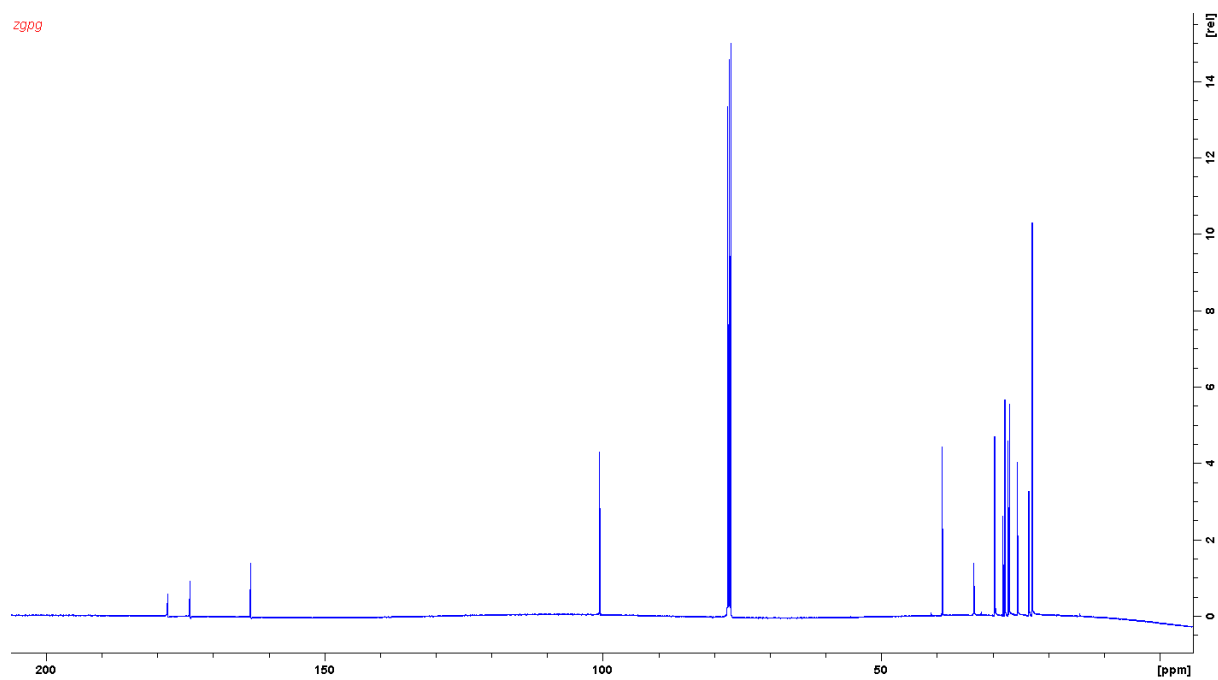
### Cell based bioactivity profiling

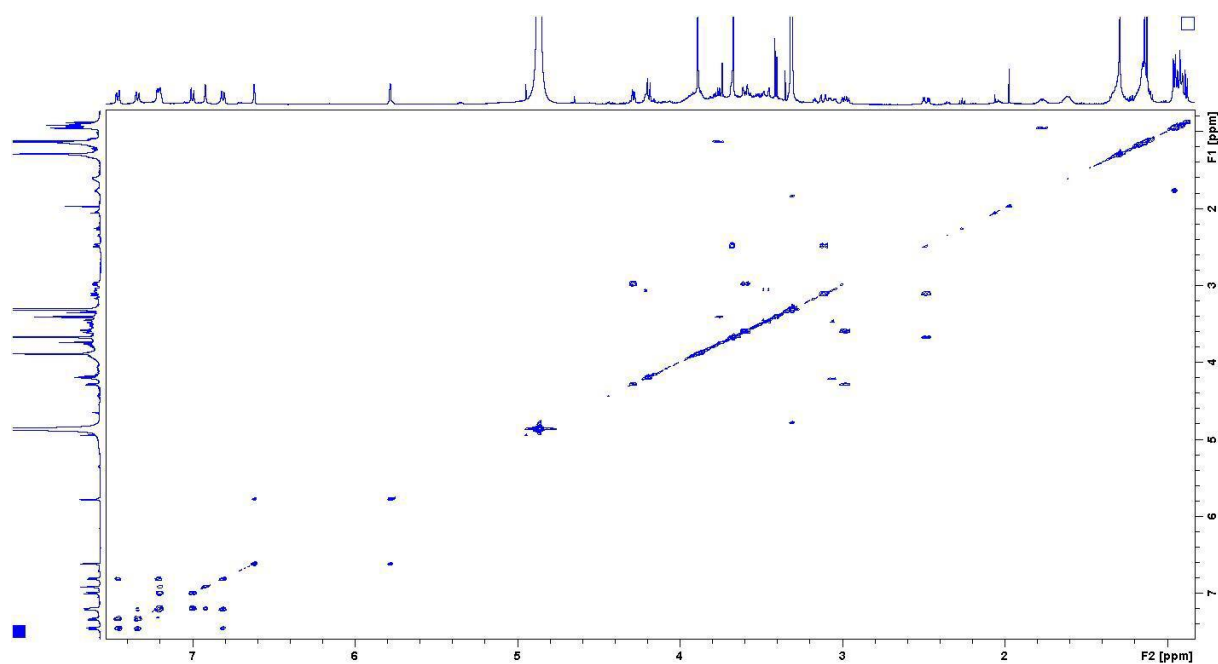
#### Antimicrobial assay

Standard sterile microbiological techniques were maintained throughout. All microorganisms were handled according to standard procedures and were obtained from the German Collection of Microorganisms and Cell Cultures (Deutsche Sammlung für Mikroorganismen und Zellkulturen, DSMZ) or were part of our internal strain collection. Myxolipoxazole and myxopyrimidinol was tested in microbroth dilution assays on the following panel of microorganisms: *Escherichia coli* DSM-1116, *E. coli* JW0451-2 (*acrB*-efflux pump deletion mutant of *E. coli* BW25113), *Pseudomonas aeruginosa* PA14, *Bacillus subtilis* DSM-10, *Mycobacterium smegmatis* mc2-155, *Staphylococcus aureus* Newman, *Candida albicans* DSM-1665, *Citrobacter freundii* DSM 30039, *Pichia anomala* DSM-6766 and *Acinetobacter baumannii* DSM30007. For microdilution assays, overnight cultures were prepared from cryo-preserved cultures and were diluted to achieve a final inoculum of  $10^4$ – $10^5$  cfu/mL. Serial dilutions of compounds were prepared in sterile 96-well plates in the respective test medium. The cell suspension was added and microorganism were grown for 18–48 h at either 30 °C or 37 °C. Growth inhibition was evaluated by visual inspection and given minimum inhibitory concentration (MIC) values are the lowest concentration of antibiotic at which no visible growth was observed. The MIC of myxolipoxazole A against *Mucor hiemalis* DSM2656 was observed at 32 µg/mL. No inhibition of one of the above-mentioned tested microorganisms was observed at concentration up to 64 µg/mL of myxolipoxazole A.

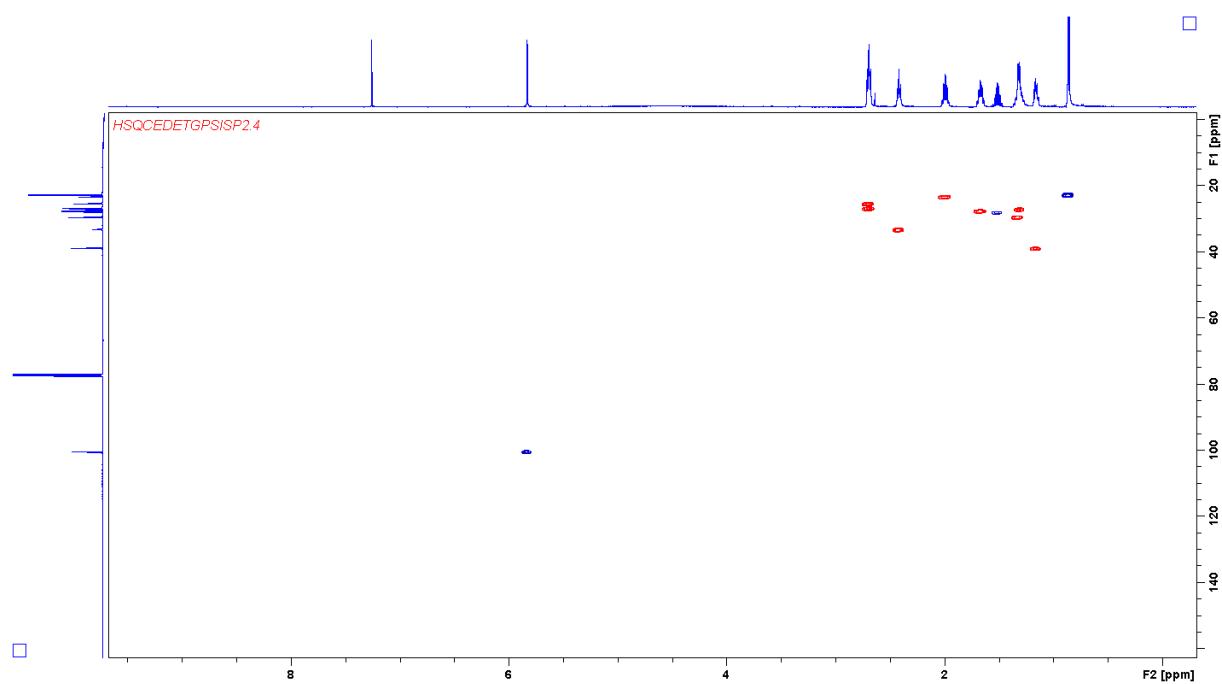
#### Cytotoxic activity

Cell lines were obtained from the German Collection of Microorganisms and Cell Cultures (Deutsche Sammlung für Mikroorganismen und Zellkulturen, DSMZ) or were part of our internal collection and were cultured under conditions recommended by the depositor. HCT-116 (human colon carcinoma cell line, DSMZ No. ACC 581) and KB-3-1 (cervix carcinoma cell line, DSMZ No. ACC 158) cells were seeded at  $6 \times 10^3$  cells per well of 96-well plates in 180 µL complete medium and treated with myxolipoxazole and myxopyrimidinol in serial dilution after 2 h equilibration. After 5 days incubation, 20 µL of 5 mg/mL MTT (thiazolyl blue tetrazolium bromide) in phosphate buffered saline (PBS) was added per well and it was further incubated for 2 h at 37°C. The medium was discarded and cells were washed with 100 µL PBS before adding 100 µL isopropanol/10 N HCl (250:1) in order to dissolve formazan granules. The absorbance at 570 nm was measured using a microplate reader (Tecan Infinite M200Pro), and cell viability was expressed as percentage relative to the respective MeOH control. IC<sub>50</sub> values were determined by sigmoidal curve fitting. The IC<sub>50</sub> of myxolipoxazole A against KB-3-1 cells was determined to 24.06 µg/mL.

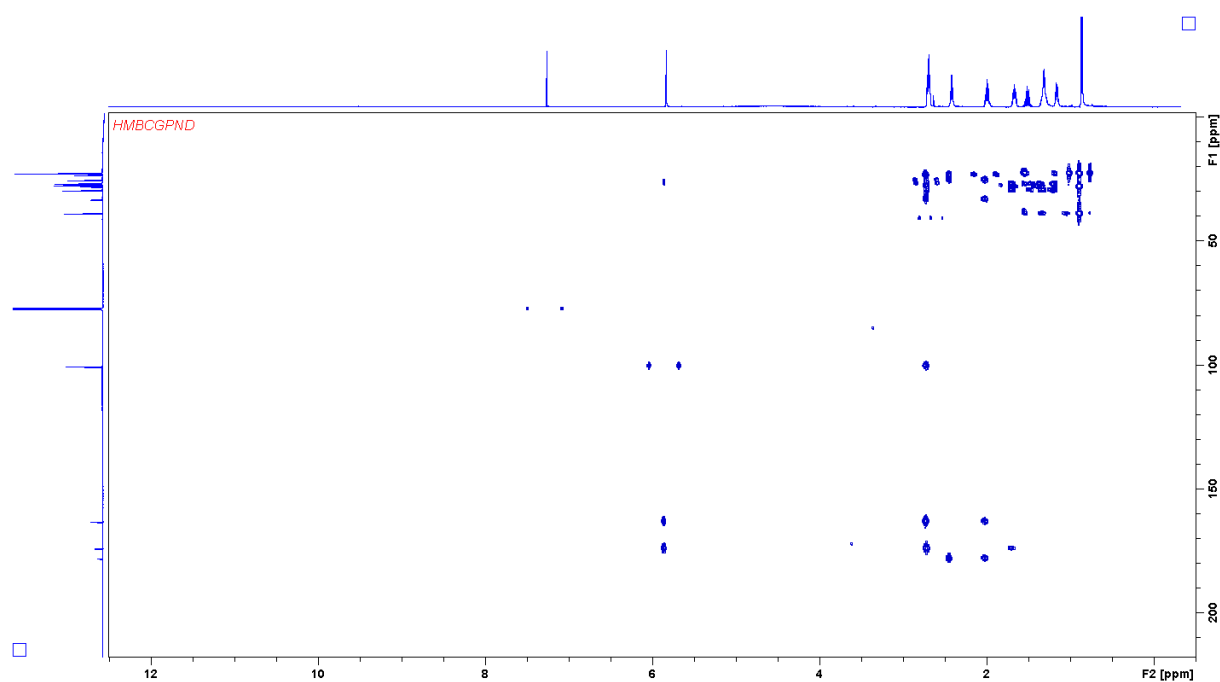
**2.8  $^1\text{H}$  and  $^{13}\text{C}$  NMR spectra of myxolipoxazole A and myxopyrimidinol A****NMR data myxolipoxazole A****Figure S27.**  $^1\text{H}$  NMR spectrum myxolipoxazole A in  $\text{CDCl}_3$ .**Figure S28.**  $^{13}\text{C}$  NMR spectrum myxolipoxazole A in  $\text{CDCl}_3$ .



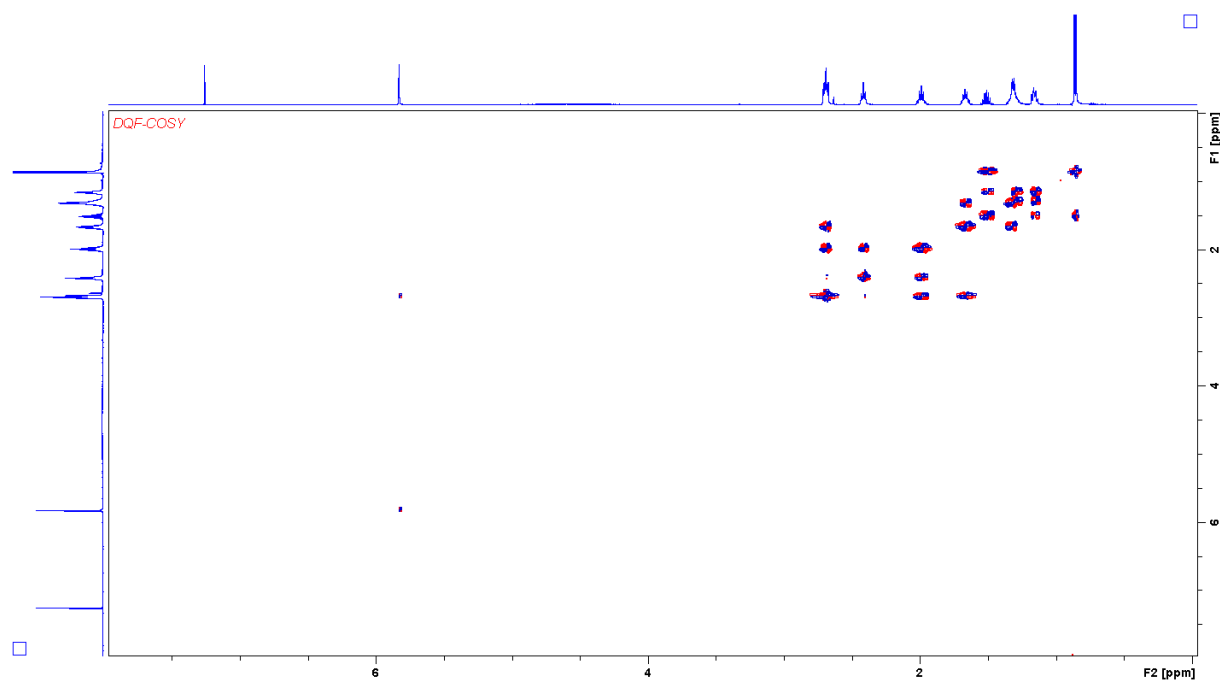
**Figure S29.** *COSY* spectrum of myxolipoxazole A in  $\text{CD}_3\text{OD}$ .



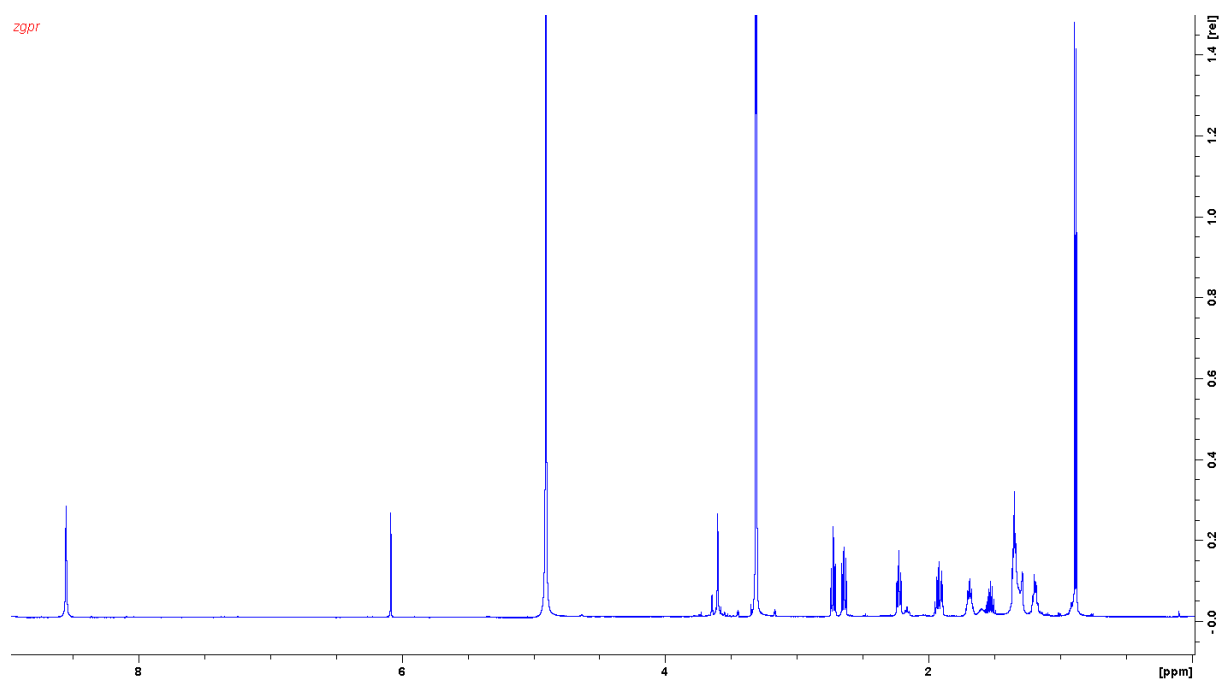
**Figure S30.** *HSQC* spectrum of myxolipoxazole A in  $\text{CDCl}_3$ .



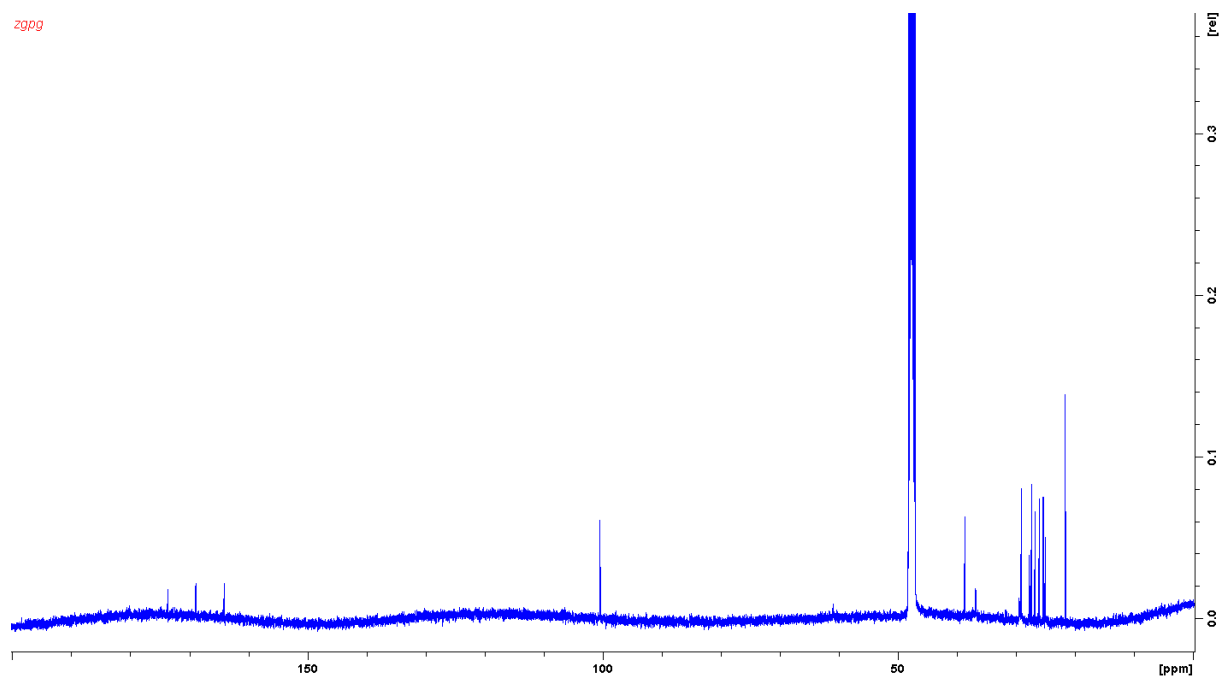
**Figure S31.** HMBC spectrum of myxolipoxazole A in CDCl<sub>3</sub>.



**Figure S32.** <sup>1</sup>H HMBC spectrum of myxolipoxazole A in CDCl<sub>3</sub>.

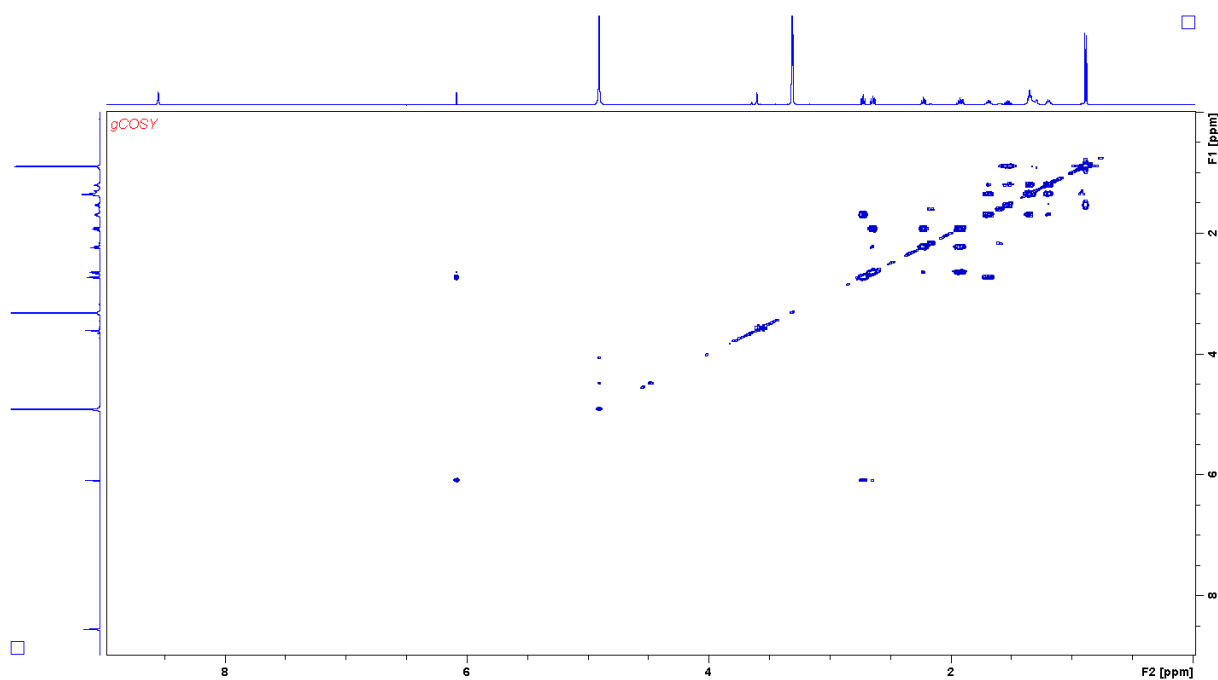


**Figure S33.**  $^1\text{H}$  spectrum of myxolipoxazole A acquired in  $\text{CD}_3\text{OD}$  at 500 MHz.

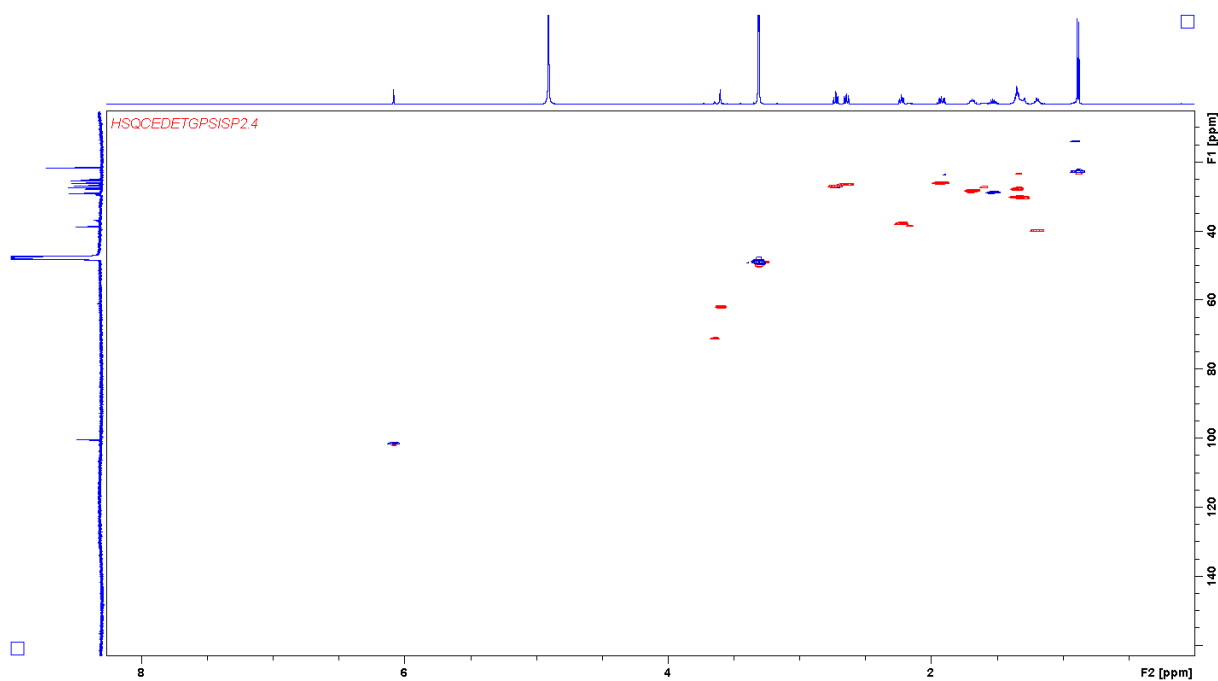


**Figure S34.**  $^{13}\text{C}$  spectrum of myxolipoxazole A acquired in  $\text{CD}_3\text{OD}$  at 500 MHz.

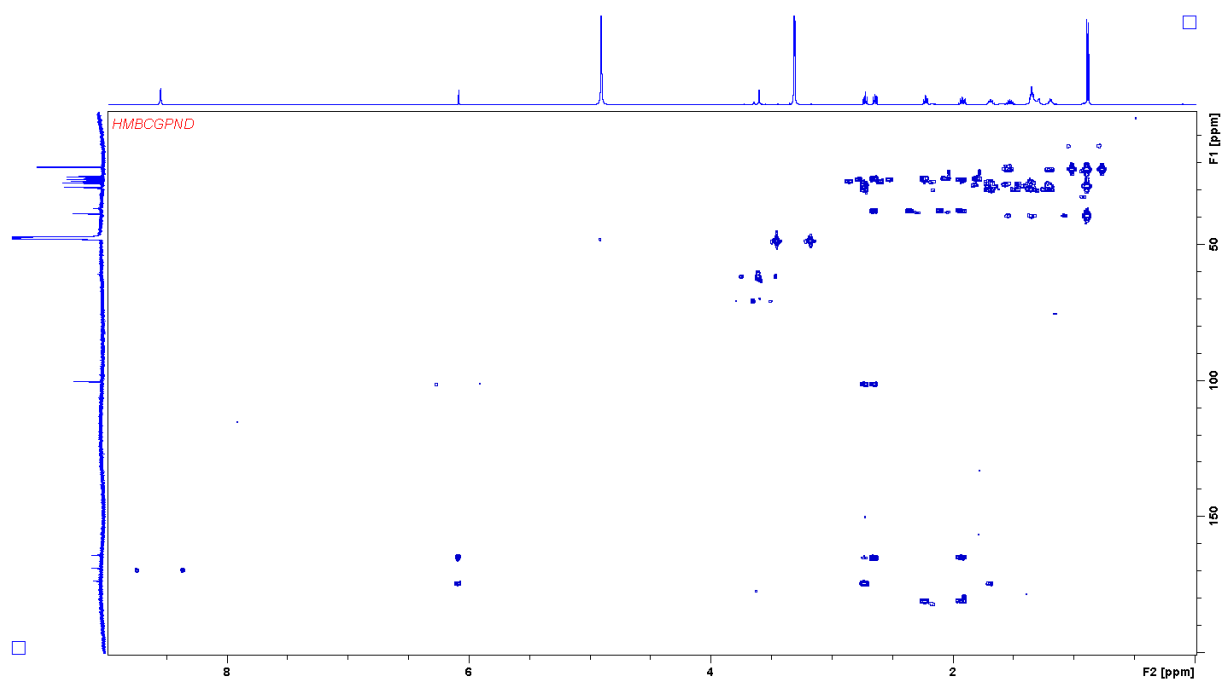




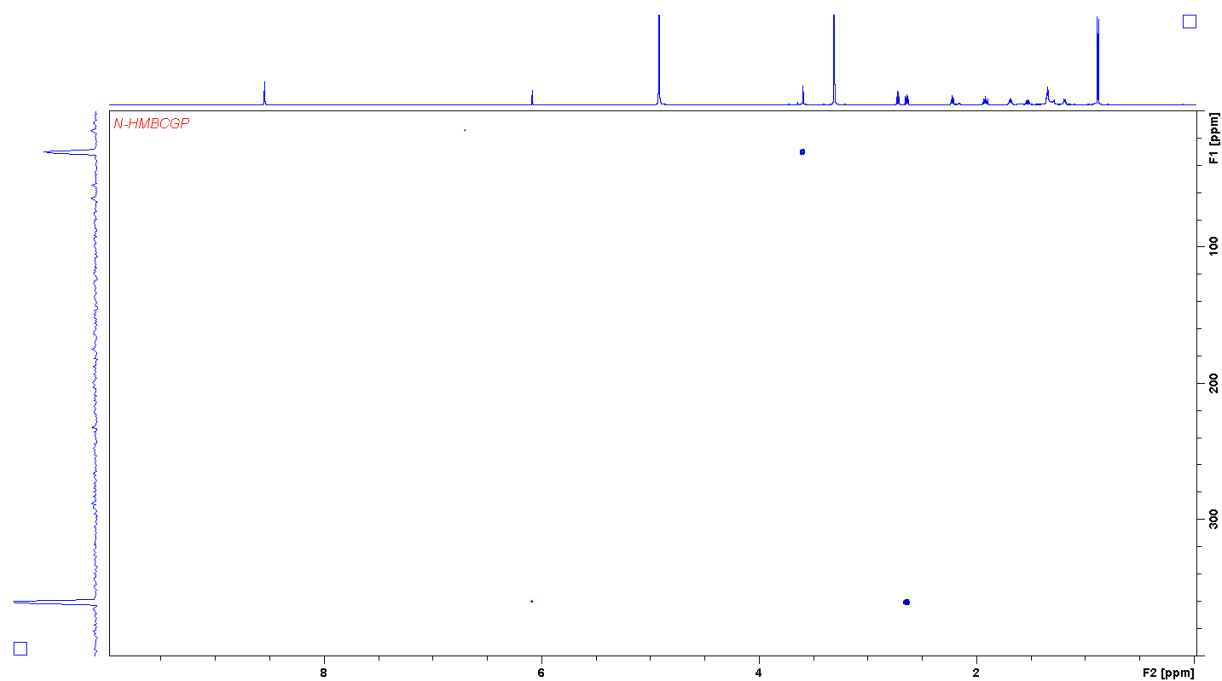
**Figure S35.** COSY spectrum of myxolipoxazole A acquired in CD<sub>3</sub>OD at 500 MHz.



**Figure S36.** HSQC spectrum of myxolipoxazole A acquired in CD<sub>3</sub>OD at 500 MHz.

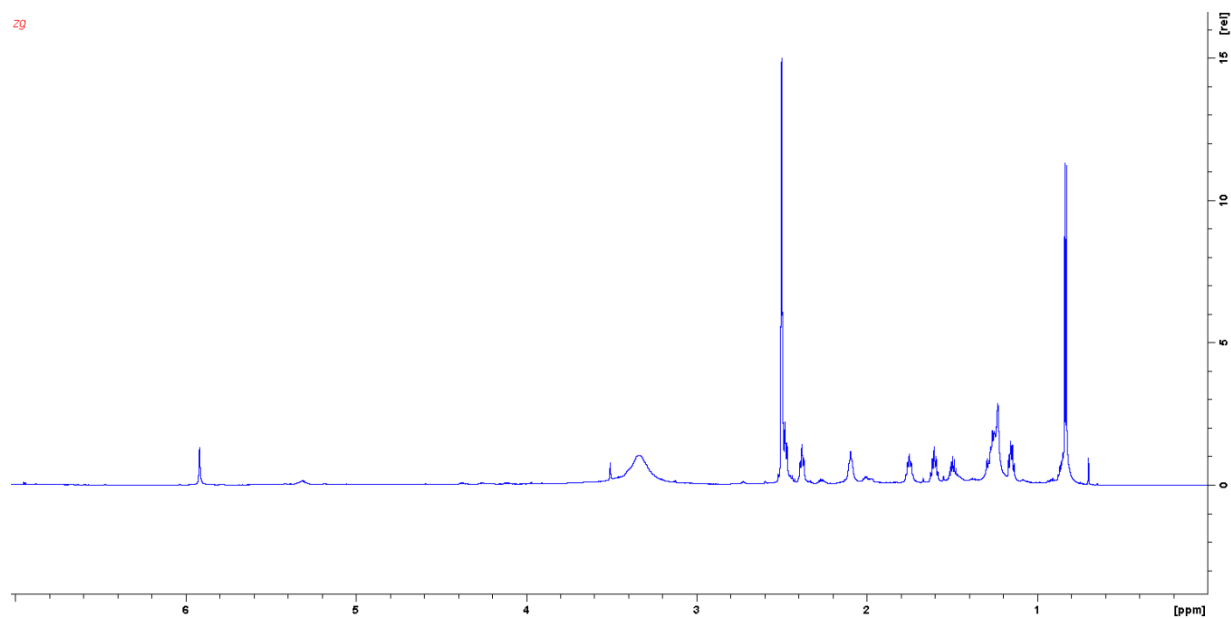


**Figure S37.** HMBC spectrum of myxolipoxazole A acquired in CD<sub>3</sub>OD at 500 MHz.

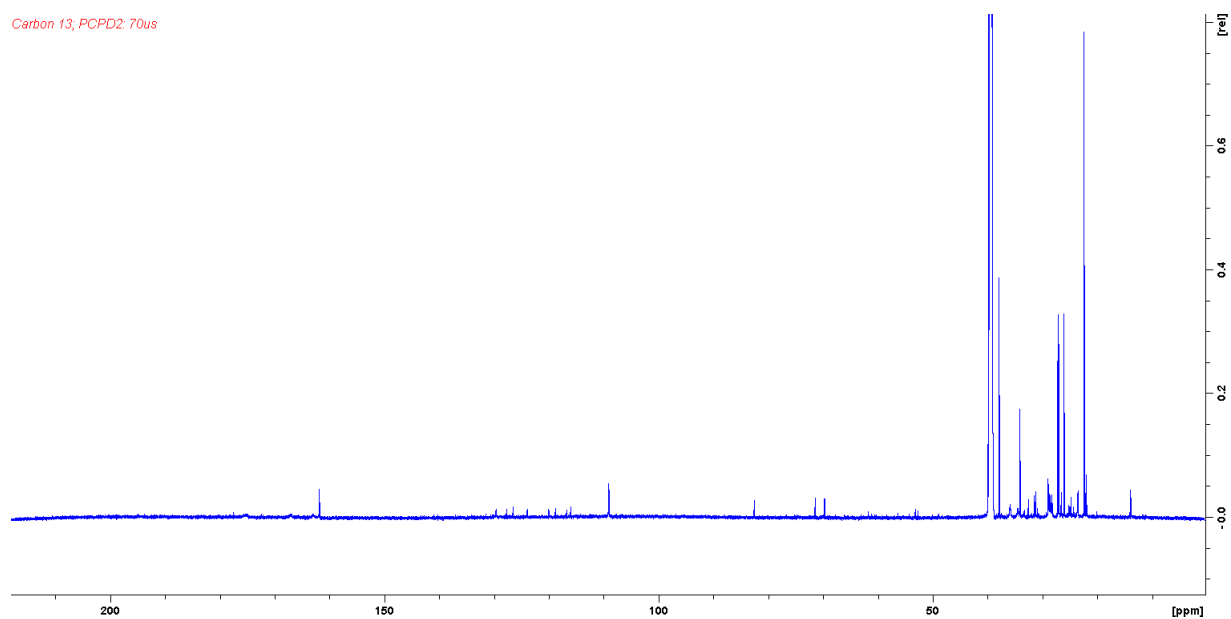


**Figure S38.** <sup>15</sup>N-HMBC spectrum of myxolipoxazole A acquired in CD<sub>3</sub>OD at 700 MHz.

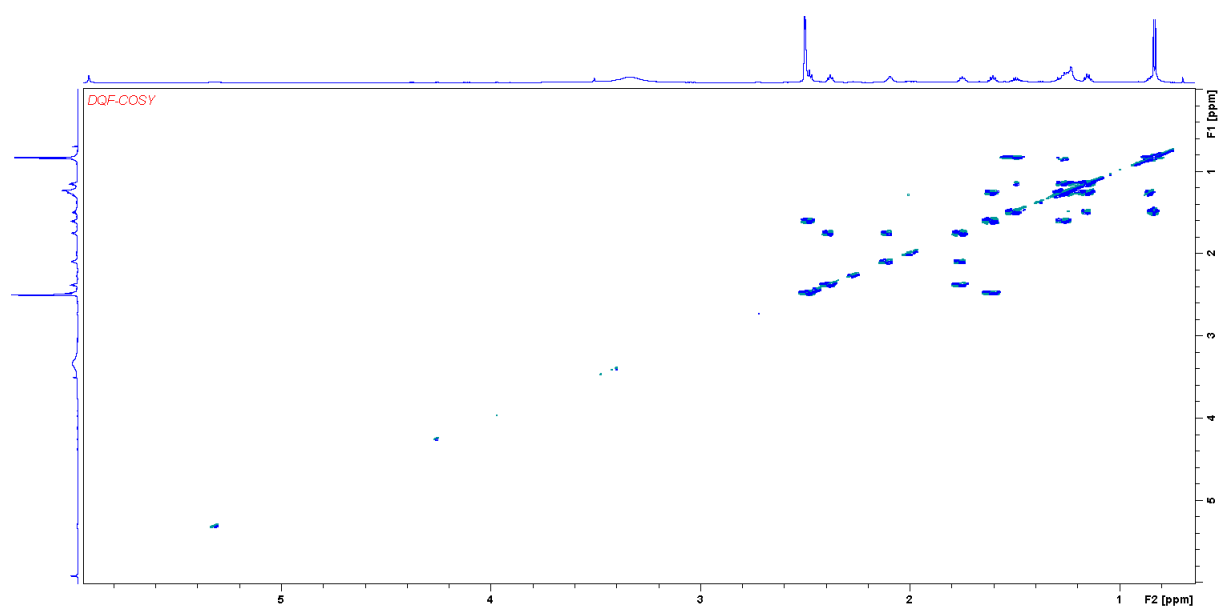
## NMR data myxopyrimidinol A



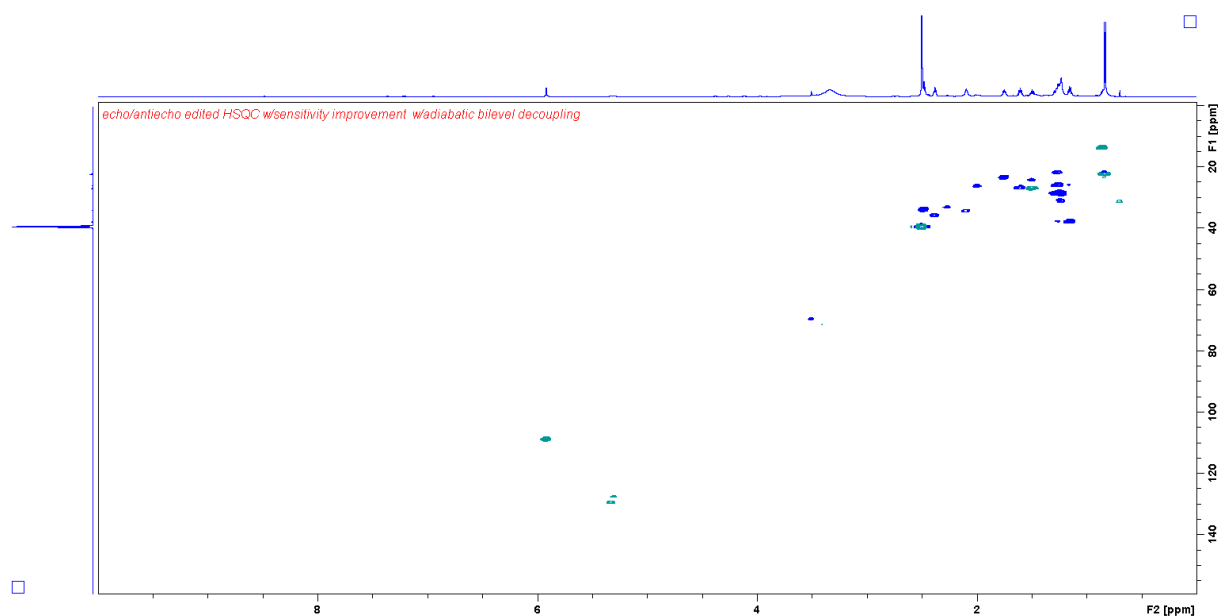
**Figure S39.**  $^1\text{H}$  spectrum of myxopyrimidinol A acquired in  $\text{CD}_3\text{OD}$  at 700 MHz.



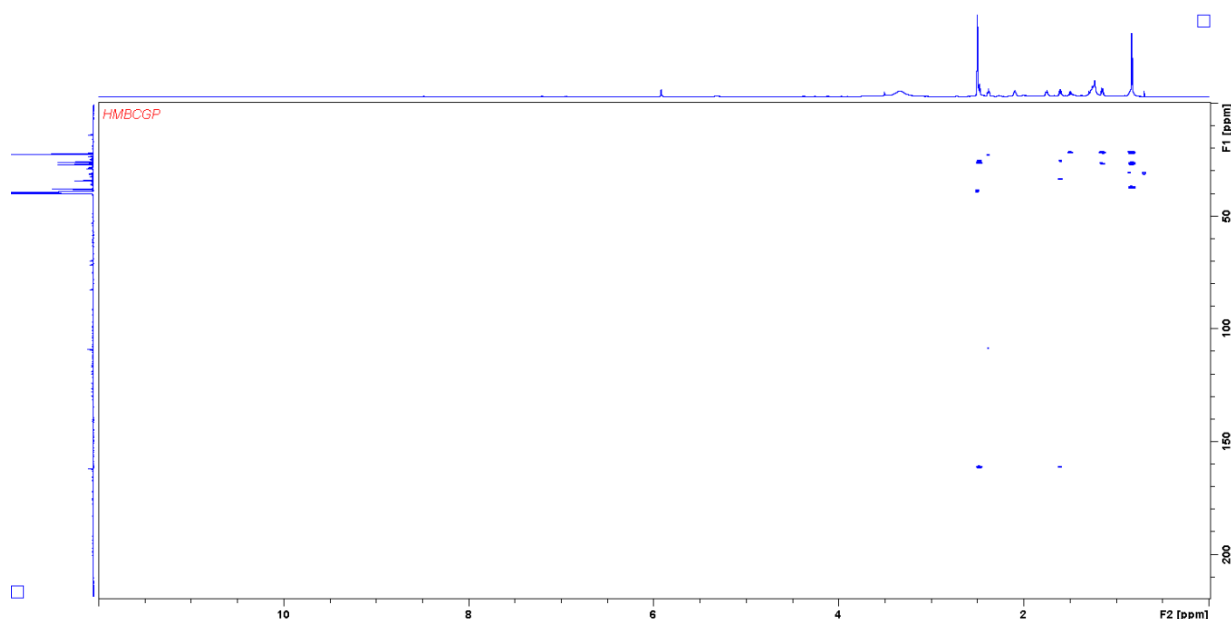
**Figure S40.**  $^{13}\text{C}$  spectrum of myxopyrimidinol A acquired in  $\text{CD}_3\text{OD}$  at 700 MHz.



**Figure S41.** DQF COSY spectrum of myxopyrimidinol A acquired in CD<sub>3</sub>OD at 700 MHz.



**Figure S42.** HSQC spectrum of myxopyrimidinol A acquired in CD<sub>3</sub>OD at 700 MHz.



**Figure S43.** HMBC spectrum of myxopyrimidinol A acquired in CD<sub>3</sub>OD at 700 MHz.

### 3. References

1. Edgar, R. C. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **32**, 1792–1797 (2004).
2. Finn, R. D. *et al.* The Pfam protein families database: towards a more sustainable future. *Nucleic Acids Res.* **44**, D279–285 (2016).
3. Kelley, L. A., Mezulis, S., Yates, C. M., Wass, M. N. & Sternberg, M. J. E. The Phyre2 web portal for protein modeling, prediction and analysis. *Nat. Protoc.* **10**, 845–858 (2015).
4. Irschik, H., Trowitzsch-Kienast, W., Gerth, K., Höfle, G. & Reichenbach, H. Saframycin Mx1, a new natural saframycin isolated from a myxobacterium. *J. Antibiot.* **41**, 993–998 (1988).
5. Green, M. R. & Sambrook, J. *Molecular cloning. A laboratory manual*. 4th ed. (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., 2012).
6. Sucipto, H., Pogorevc, D., Luxemburger, E., Wenzel, S. C. & Müller, R. Heterologous production of myxobacterial  $\alpha$ -pyrone antibiotics in *Myxococcus xanthus*. *Metab. Eng.* **44**, 160–170 (2017).
7. Pogorevc, D. *et al.* Production optimization and biosynthesis revision of corallopyronin A, a potent anti-filarial antibiotic. *Metab. Eng.* **55**, 201–211 (2019).
8. Panter, F., Krug, D., Baumann, S. & Müller, R. Self-resistance guided genome mining uncovers new topoisomerase inhibitors from myxobacteria. *Chem. Sci.* **9**, 4898–4908 (2018).
9. Garcia, R. & Müller, R. in *The Prokaryotes*, edited by E. Rosenberg, E. F. DeLong, S. Lory, E. Stackebrandt & F. Thompson (Springer Berlin Heidelberg, Berlin, Heidelberg, s.l., 2014), pp. 191–212.

10. Gould, S. J. & Ju, S. Biosynthesis of Acivicin. 3. Incorporation of Ornithine and  $N^{\delta}$ -Hydroxyornithine. *J. Am. Chem. Soc.* **114**, 10166–10172 (1992).
11. Shyh-Chen Ju. Biosynthesis of Acivicin and 4-Hydroxyacivicin. Thesis. Oregon State University, September 1988.
12. Hug, J. J. *et al.* Biosynthesis of citilins, unusual ribosomally synthesized and post-translationally modified peptides from *Myxococcus xanthus*. *ACS Chem. Biol.* **submitted** (2020).
13. Meiser, P. *et al.* DKxanthene biosynthesis—understanding the basis for diversity-oriented synthesis in myxobacterial secondary metabolism. *Chem. Biol.* **15**, 771–781 (2008).
14. Burgard, C. *et al.* Genomics-Guided Exploitation of Lipopeptide Diversity in Myxobacteria. *ACS Chem. Biol.* **12**, 779–786 (2017).
15. Trowitzsch, W., Reifensahl, G., Wray, V. & Höfle, G. Myxothiazol, an antibiotic from *Myxococcus fulvus* (Myxobacterales) II. Structure elucidation. *J. Antibiot.* **33**, 1480–1490 (1980).
16. Livingstone, P. G. *et al.* Predatory organisms with untapped biosynthetic potential. A description of eight novel *Coralloccoccus* species: *Coralloccoccus aberystwythiensis* sp. nov., *Coralloccoccus carmarthensis* sp. nov., *Coralloccoccus exercitus* sp. nov., *Coralloccoccus interemptor* sp. nov., *Coralloccoccus llansteffanensis* sp. nov., *Coralloccoccus praedator* sp. nov., *Coralloccoccus sicarius* sp. nov., and *Coralloccoccus terminator* sp. nov. *Appl. Environ. Microbiol.* (2019).
17. Albataineh, H., Duke, M., Misra, S. K., Sharp, J. S. & Stevens, D. C. Solo acylhomoserine lactone synthase from predatory myxobacterium suggests beneficial participation in interspecies cross talk. *bioRxiv* **65** (2019).
18. Cantu, D. C., Chen, Y. & Reilly, P. J. Thioesterases: A new perspective based on their primary and tertiary structures. *Protein Sci.* **19**, 1281–1295 (2010).
19. Chen, X. *et al.* Phenylacetic acid catabolism and its transcriptional regulation in *Corynebacterium glutamicum*. *Appl. Environ. Microbiol.* **78**, 5796–5804 (2012).
20. Ferrández, A. *et al.* Catabolism of phenylacetic acid in *Escherichia coli*. Characterization of a new aerobic hybrid pathway. *J. Biol. Chem.* **273**, 25974–25986 (1998).
21. Heathcote, M. L., Staunton, J. & Leadlay, P. F. Role of type II thioesterases: evidence for removal of short acyl chains produced by aberrant decarboxylation of chain extender units. *Chem. Biol.* **8**, 207–220 (2001).
22. Ollis, D. L. *et al.* The alpha/beta-hydrolase fold. *Protein Eng.* **5**, 197–211 (1992).
23. Dillon, S. C. & Bateman, A. The Hotdog fold: wrapping up a superfamily of thioesterases and dehydratases. *BMC bioinformatics* **5**, 109 (2004).
24. Khandokar, Y. B. *et al.* Structural and Functional Characterization of the PaaI Thioesterase from *Streptococcus pneumoniae* Reveals a Dual Specificity for Phenylacetyl-CoA and Medium-chain Fatty Acyl-CoAs and a Novel CoA-induced Fit Mechanism. *J. Biol. Chem.* **291**, 1866–1876 (2016).
25. Zhang, Z. *et al.* Structural and functional studies of fatty acyl adenylate ligases from *E. coli* and *L. pneumophila*. *J. Mol. Biol.* **406**, 313–324 (2011).

26. Clark, L., Leatherby, D., Krilich, E., Ropelewski, A. J. & Perozich, J. *In silico* analysis of class I adenylate-forming enzymes reveals family and group-specific conservations. *PLoS ONE* **13**, e0203218 (2018).
27. Keating-Clay, A. T. The structures of type I polyketide synthases. *Nat. Prod. Rep.* **29**, 1050–1073 (2012).
28. Hemmerling, F., Lebe, K. E., Wunderlich, J. & Hahn, F. An Unusual Fatty Acyl:Adenylate Ligase (FAAL)–Acyl Carrier Protein (ACP) Didomain in Ambruticin Biosynthesis. *ChemBioChem* (2018).
29. Marchetti, P. M., van Kelly, Simpson, J. P., Ward, M. & Campopiano, D. J. The carbon chain-selective adenylation enzyme TamA: the missing link between fatty acid and pyrrole natural product biosynthesis. *Org. Biomol. Chem.* **16**, 2735–2740 (2018).
30. Voráčová, K. *et al.* The cyanobacterial metabolite nocuolin a is a natural oxadiazine that triggers apoptosis in human cancer cells. *PLoS ONE* **12**, e0172850 (2017).
31. Inahashi, Y. *et al.* Identification and heterologous expression of the actinoallolide biosynthetic gene cluster. *J Antibiot* **71**, 749–752 (2018).
32. Puehringer, S., Metlitzky, M. & Schwarzenbacher, R. The pyrroloquinoline quinone biosynthesis pathway revisited: a structural approach. *BMC Biochem.* **9**, 8 (2008).
33. Rausch, C., Hoof, I., Weber, T., Wohlleben, W. & Huson, D. H. Phylogenetic analysis of condensation domains in NRPS sheds light on their functional evolution. *BMC Evol. Biol.* **7**, 78–92 (2007).
34. Pogorevc, D. *et al.* Biosynthesis and Heterologous Production of Argyrins. *ACS Synth. Biol.* **8**, 1121–1133 (2019).
35. Duerfahrt, T., Eppelmann, K., Müller, R. & Marahiel, M. A. Rational design of a bimodular model system for the investigation of heterocyclization in nonribosomal peptide biosynthesis. *Chem. Biol.* **11**, 261–271 (2004).
36. Hug, J. J. & Müller, R. Host Development for Heterologous Expression and Biosynthetic Studies of Myxobacterial Natural Products. *Comprehensive Natural Products III: Chemistry and Biology, Chapter 14818 In press* (2020).
37. Alekseyev, V. Y., Liu, C. W., Cane, D. E., Puglisi, J. D. & Khosla, C. Solution structure and proposed domain domain recognition interface of an acyl carrier protein domain from a modular polyketide synthase. *Protein Sci.* **16**, 2093–2107 (2007).
38. Panter, F., Krug, D. & Müller, R. Novel Methoxymethacrylate Natural Products Uncovered by Statistics-Based Mining of the *Myxococcus fulvus* Secondary Metabolome. *ACS Chem. Biol.* **14**, 88–98 (2019).
39. Hug, J. J., Panter, F., Krug, D. & Müller, R. Genome mining reveals uncommon alkylpyrones as type III PKS products from myxobacteria. *J. Ind. Microbiol. Biotechnol.* **46**, 319–334 (2019).

40. Iniesta, A. A., García-Heras, F., Abellón-Ruiz, J., Gallego-García, A. & Elías-Arnanz, M. Two systems for conditional gene expression in *Myxococcus xanthus* inducible by isopropyl- $\beta$ -D-thiogalactopyranoside or vanillate. *J. Bacteriol.* **194**, 5875–5885 (2012).
41. Kashefi, K. & Hartzell, P. L. Genetic suppression and phenotypic masking of a *Myxococcus xanthus* *frzF*- defect. *Mol. Microbiol.* **15**, 483–494 (1995).



## Chapter 6

### **Genome mining reveals uncommon alkylpyrones as type III PKS products from myxobacteria**

**Joachim J. Hug**, Fabian Panter, Daniel Krug and Rolf Müller

*Journal of Industrial Microbiology and Biotechnology*, **2019**, 46, 319–334

DOI: 10.1007/s10295-018-2105-6

Published online: 01.12.2018

## **Contributions to the presented work**

### **Author's contribution**

The author significantly contributed to the conception of this study, designed and performed experiments, evaluated and interpreted resulting data. The laboratory and *in silico* work regarding gene cluster analysis, genetic manipulation, heterologous expression efforts, isolation and structure elucidation of the alkylpyrones was performed by the author. Furthermore, the author contributed significantly to conceiving and writing this manuscript.

### **Contribution by others**

Fabian Panter performed topoisomerase inhibition assays and evaluated the resulting data. He significantly contributed to the conception of this study through *in silico* identification of the target biosynthetic gene cluster in the genome of MCy9487. Daniel contributed to the conception and supervision of this study. Furthermore he also contributed through editorial supervision and proofreading of this manuscript. Rolf Müller contributed by supervision of the project and conceiving, editing and proofreading of this manuscript.

## 6 Genome mining reveals uncommon alkylpyrones as type III PKS products from myxobacteria

### 6.1 Abstract

Type III polyketide synthases (PKSs) are comparatively small homodimeric enzymes affording natural products with diverse structures and functions. While type III PKS biosynthetic pathways have been studied thoroughly in plants, their counterparts from bacteria and fungi are to date scarcely characterized. This gap is exemplified by myxobacteria from which no type III PKS-derived small molecule has previously been isolated. In this study, we conducted a genomic survey of myxobacterial type III PKSs and report the identification of uncommon alkylpyrones as the products of type III PKS biosynthesis from the myxobacterial model strain *Myxococcus xanthus* DK1622 through a self-resistance-guided screening approach focusing on genes encoding pentapeptide repeat proteins, proficient to confer resistance to topoisomerase inhibitors. Using promoter-induced gene expression in the native host as well as heterologous expression of biosynthetic type III PKS genes, sufficient amounts of material could be obtained for structural elucidation and bioactivity testing, revealing potent topoisomerase activity *in vitro*.

### 6.2 Introduction

There is a constant need for the discovery of novel natural products to support the urgent development of new therapeutic options for the treatment of numerous diseases<sup>1</sup>. While screening efforts have been centered around certain groups of microorganisms and classes of secondary metabolites in the past, new genomics- and metabolomics-based methods are now increasingly applied to uncover unidentified natural products from less investigated microbes including rare actinomycetes, cyanobacteria, plant endosymbionts and the myxobacteria<sup>2,3</sup>. Myxobacteria are ubiquitous Gram-negative soil bacteria of the order *Myxococcales*, subclassified in the  $\delta$ -group of proteobacteria. They move by gliding on surfaces and show a unique cooperative “social behavior” including synchronized swarming directed through chemical communication. Similar to some fungi, myxobacteria build complex fruiting bodies consisting of vegetative cells when nutrients become limited, and differentiate into myxospores, which can withstand unfavorable environmental influences<sup>4</sup>. Some myxobacteria are “predators” of other prokaryotes by excretion of exoenzymes to lyse bacteria and yeast to use the lysed material as nutrients<sup>5</sup>. In addition to these remarkable biological characteristics, myxobacteria are a prolific source for natural products featuring great structural diversity and often new modes of action<sup>3</sup>. Many intriguing secondary metabolites are synthesized by multimodular enzyme complexes of the modular non-ribosomal peptide synthetases (NRPS), polyketide synthases (PKS) and in particular, PKS-NRPS hybrids systems, whereas natural products with other structural scaffolds like terpenoids, phenyl-propanoids and alkaloids have been less frequently isolated. One of these underrepresented families of secondary metabolites are natural products originating from type III PKSs. Type III PKSs are comparatively small homodimeric enzymes, which act iteratively for polyketide chain extension<sup>6</sup>. In general, these small proteins utilize a single active site to iteratively condense dicarboxylic malonyl-coenzyme-A (CoA) or methylmalonyl-CoA extender units onto activated CoA analogues acting as starter substrates<sup>7</sup>; however some bacterial type III PKSs use starter units bound to acyl carrier proteins as exemplified in NRPS

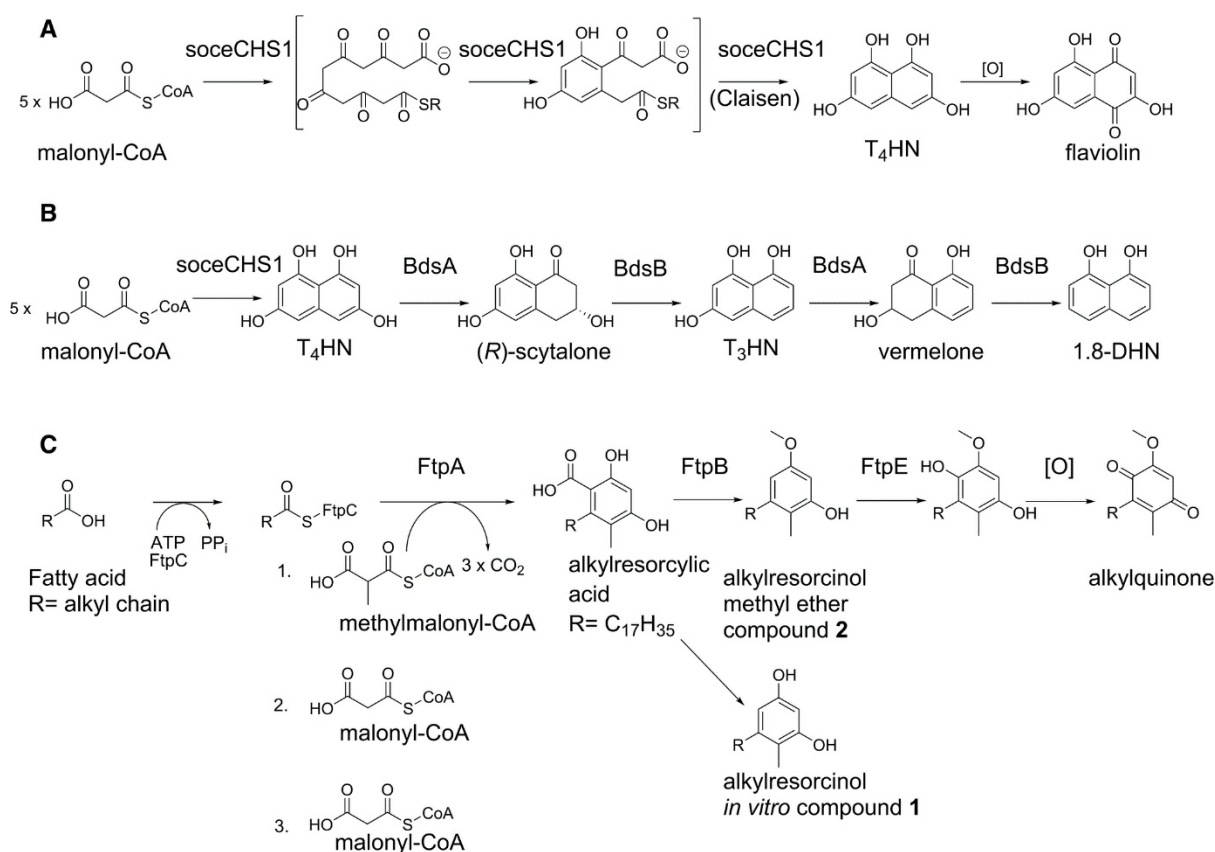
biosynthesis. Despite their structural simplicity, type III PKSs synthesize natural products with diverse structures due to their substrate variability, flexible iteration and multiple cyclization reactions by utilizing a limited substrate pool<sup>7,8</sup>. Notably, the broad substrate promiscuity exhibited by type III PKSs can be utilized to incorporate unnatural building blocks to generate novel “unnatural natural products” in order to expand the chemical diversity of polyketides by enzymatic engineering<sup>9,10</sup>. In addition, chemical modifications of type III polyketide scaffolds are performed by tailoring enzymes, which have been shown to cluster with type III PKS genes in microbial genomes<sup>11</sup>. These modification enzymes can be classified as “upstream reactions” – responsible for the production of type III PKS starters like the type I fatty acid synthases (FASs) ArsA and ArsD<sup>12</sup>, which generate fatty acid starter units, and the fatty acyl-AMP ligase FtpD producing fatty acyl-ACP as starter units<sup>13</sup> – or as tailoring enzymes performing “downstream reactions” of type III PKS products like acetylation, O-methylation, hydroxylation and oxidation<sup>11</sup>. Although type III PKSs have been thoroughly investigated since the 1980s<sup>14</sup>, in-depth biosynthetic characterization of the underlying decarboxylative condensation machinery and subsequent cyclization of the polyketide can only be pursued following structural elucidation of the specific type III polyketide. Accordingly, prediction of the synthesized small molecule scaffold from sequence information alone is generally not possible for type III PKS pathways. In contrast, high mutual similarity of canonical domain types, in particular ketosynthase (KS) and acyltransferase (AT) domains in non-iterative PKS and condensation (C) and adenylation (A) domains in NRPS, greatly facilitates the *in silico* prediction of approximate chemical scaffolds produced by these gigantic microbial biosynthetic pathway machineries, since they are (to varying degrees) following a so-called collinearity rule<sup>15,16</sup>. *In silico* characterization of bacterial type III PKSs are further complicated as these, unlike the plant type III PKSs, usually share less than 50% sequence identity with each other. Circumventing these bioinformatics obstacles, crystallographic analyses of the bacterial type III PKSs 1,3,6,8-tetrahydroxynaphthlene synthase (THNS) (*S. coelicolor*<sup>17</sup>), PKS11 and PKS18 (*M. tuberculosis*<sup>18,19</sup>) and ArsC (*A. vinelandii*<sup>20</sup>) enabled the understanding of starter unit selectivity, chain elongation and cyclization of the linear polyketide intermediate. Biochemical characterization is a crucial prerequisite to leverage the biotechnological potential of type III PKS systems through engineering of their activity and substrate specificity, since even subtle modifications in the active site cavity can alter the scaffold of the produced natural products<sup>21</sup>.

Among the few reports of type III PKS natural products from myxobacteria are flaviolin – the product of non-enzymatic oxidation of 1,3,6,8-tetrahydroxynaphthalene (T<sub>4</sub>HN) – obtained by heterologous expression of a type III PKS from the myxobacterium *Sorangium cellulosum* So ce56 in *Pseudomonas putida* KT2440<sup>22</sup> and 1,8-dihydroxynaphthalene (1,8-DHN) generated by heterologous expression of the associated gene operon consisting of *soceCHS1* (type III PKS), *bdsA* (naphthol reductase) and *bdsB* (scytalone/vermelone dehydratase) in *Streptomyces coelicolor*<sup>23</sup> (**Figure 1A, B**). In addition, the biosynthesis of 1,8-DHN was reconstituted *in vitro* through simultaneous incubation of recombinantly produced SoceCHS1, BdsA and BdsB with malonyl-CoA and NADPH. Furthermore, production of the alkylresorcinol 5-heptadecyl-4-methylbenzene-1,3-diol (*in vitro* compound **1**) has been achieved through *in vitro* catalysis of stearic acid with recombinant FtpA (type III PKS), FtpC (ACP) and FtpD (fatty acyl-AMP ligase) from *Myxococcus xanthus* Beebe 1941 (DSM: 52389). The alkylresorcinol methyl ether 5-methoxy-2-methyl-3-heptadecyl-phenol (compound **2**) was isolated from heterologous expression of the *ftpABCD* operon (FtpB: methyltransferase) by *Streptomyces lividans*, harboring pIJ6021-*ftpABCD*<sup>13</sup> (**Figure 1C**). The last gene in the operon *ftpABCDE*, *ftpE*, encodes a FAD-dependent monooxygenase classified as class A monooxygenase<sup>24</sup>, that shows significant structural similarity to the crystallized flavin-dependent monooxygenase PgaE<sup>25</sup> and sequence similarity to the hydroxylase SrsC<sup>26</sup>, deriving from *Streptomyces griseus* IFO 13350. Therefore, we assume that FtpE

like SrsC hydroxylates compound **2** to produce 6-alkyl-2-methoxy-5-methylhydroquinone, which is then auto-oxidized to alkylquinone (**Figure 1C**).

In contrast to bioinformatic insights that indicate numerous type III PKSs to be present in myxobacterial genomes, none of the examples above have connected genes to a matching metabolite in the myxobacterial secondary metabolome to date. These findings have been fueling speculation that the myxobacterial type III PKS pathways could be downregulated under laboratory fermentation conditions so that production titers remain below the limit of detection.

In this study, we describe combined genomic and metabolomic investigations of myxobacteria with regard to type III polyketide synthases. A self-resistance guided screening approach focusing on pentapeptide repeat proteins (PRPs) co-localized with myxobacterial type III polyketide synthases, served as an indicator for the biological function of the afforded natural product of a specific type III BGC. Induced gene activation and heterologous expression of this selected type III PKS pathway lead to the production of uncommon alkylpyrones, allowing for their full structural elucidation and ultimately their biological characterization as potent topoisomerase inhibitors.



**Figure 1.** Myxobacterial type III PKS described in literature by heterologous expression and through *in vitro* reconstitution of key enzymes involved in type III polyketide formation. A) SoceCHS1 originating from *Sorangium cellulosum* So ce56 has been heterologously produced in *Pseudomonas putida* KT2440 yielding flaviolin<sup>22</sup>, while flaviolin is the product of non-enzymatic oxidation of 1,3,6,8-tetrahydroxynaphthalene (T<sub>4</sub>HN). B) Heterologous expression of the gene operon *soceCHS1* (type III PKS), *bdsA* (naphthol reductase) and *bdsB* (scytalone/vermelone dehydratase) originating from *Sorangium cellulosum* So ce56 in *Streptomyces coelicolor* generated 1,8-dihydroxynaphthalene (1,8-DHN)<sup>23</sup>. C) Proposed biosynthetic pathway carried out by the Ftp proteins, originating from *Myxococcus xanthus* Beebe 1941<sup>13</sup>.

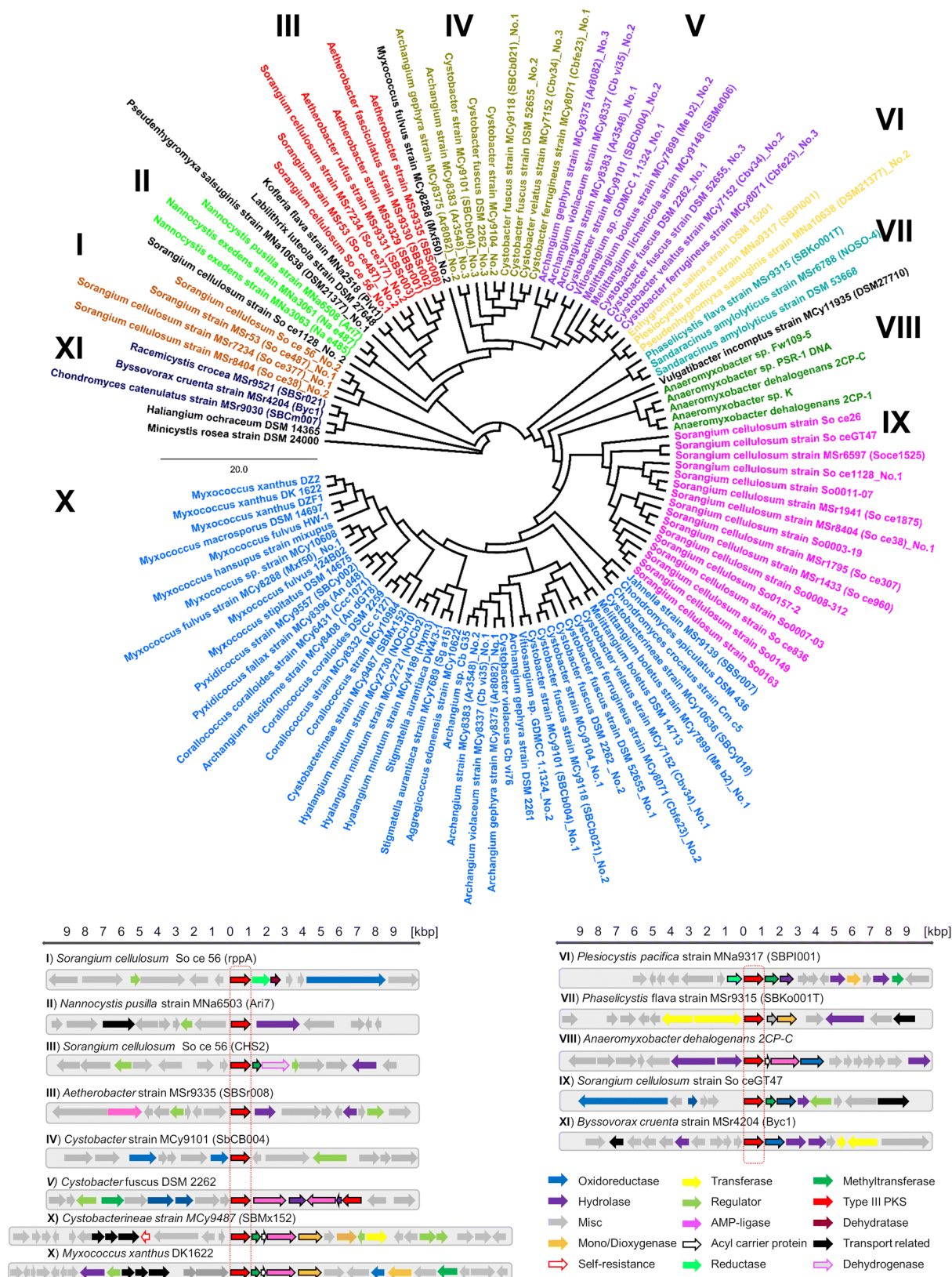
## 6.3 Results and discussion

### Survey of type III PKS biosynthesis pathways in myxobacterial genomes

The starting point of this study is defined by our aim to assess the hidden potential of myxobacteria for production of type III polyketides. An initial search for genes encoding type III polyketide synthases yielded 116 candidate loci spread on 90 myxobacterial genome sequences. Based on primary amino acid sequence similarity these genes were aligned in a phylogenetic tree to highlight distinct families of type III PKS pathways (**Figure 2**). Type III PKSs were considered as members of the same family, when they shared more than 60% identical sites in the amino acid alignment (**Suppl. 2.1**). However, since fungal and bacterial type III PKSs typically show less than 50% mutual similarity<sup>27</sup>, even members placed in different families by this method may produce the same secondary metabolites. Moreover, genes encoding accessory enzymes responsible for upstream or downstream tailoring reactions – which are usually but not exclusively organized within an operon with the type III PKS core genes – might alter the produced scaffold, leading to different type III PKS metabolites from the same sequence subtype. In light of these obstacles, the phylogenetic tree (**Figure 2**) should be regarded as a rough overview on the genetic level, illustrating approximate type III PKS diversity in myxobacteria through the apparent grouping into pathway families (**Suppl. 2.1**).

Notably, only two of the sequence families permit a reasonable hypothesis concerning the small molecule product possibly formed by these pathways. Due to sequence similarity to *rppA*-like type III polyketides family **I** is termed the *rppA* family (**Figure 2**, red-brown color), likely producing the intermediate flaviolin through non-enzymatic oxidation<sup>22</sup>. *RppA* homologues are known from *Streptomyces* species within various genetic contexts, accounting for different subsequent tailoring reactions. In *Streptomyces antibioticus* IFO13271 a monooxygenase MomA catalyzes the oxygenation of T<sub>4</sub>HN to flaviolin and subsequently to mompain<sup>28</sup>, whereas in *Streptomyces griseus* IFO13350 the biosynthetic pathway comprising *RppA* and a cytochrome P450 enzyme namely P450mel yields 1,4,6,7,9,12-hexahydroxyperylene-3,10-quinone (HPQ) which is directly autopolymerized to generate HPQ melanin<sup>29</sup>.

As recently described by Sone *et al.*<sup>23</sup> the *rppA*-like gene *soceCHS1* in *Sorangium cellulosum* So ce56 and its connected operon consisting of *soceCHS1*, *bdsA* and *bdsB* could be found in three other myxobacterial genome sequences (*Sorangium cellulosum* So ce487, *Sorangium cellulosum* So ce377 and *Sorangium cellulosum* So ce38), concluding that these strains are also capable to produce flaviolin or the further processed 1,8-DHN. The specific key amino acid residues in *rppA*-like type III PKSs that are responsible for reaction priming and chain elongation (C138, H270 and N303)<sup>30</sup> are present in this family of type III PKSs, unlike in the phylogenetically neighbored gene clusters from *Chondromyces catenulatus* MSr9030 or *Nannocystis pusilla* MNa6508 (**Suppl. 2.1**). The basic local alignment search tool (BLAST) for homologues of *momA* from *Streptomyces antibioticus* IFO13271 or the cytochrome P-450 enzyme namely P-450mel homologues from *Streptomyces griseus* IFO13350 did not reveal any myxobacterial homologue, making it less likely to find either mompain or HPQ in myxobacteria.



**Figure 2.** Circular phylogenetic tree based on gene sequence similarity for genes encoding type III polyketide synthases in myxobacterial genomes. Colored clades highlight distinct families of type III PKSs. Below the phylogenetic tree, the schematic genetic organization of representative members from each family is presented. See **Suppl. 2.1** for detailed comments on all families.

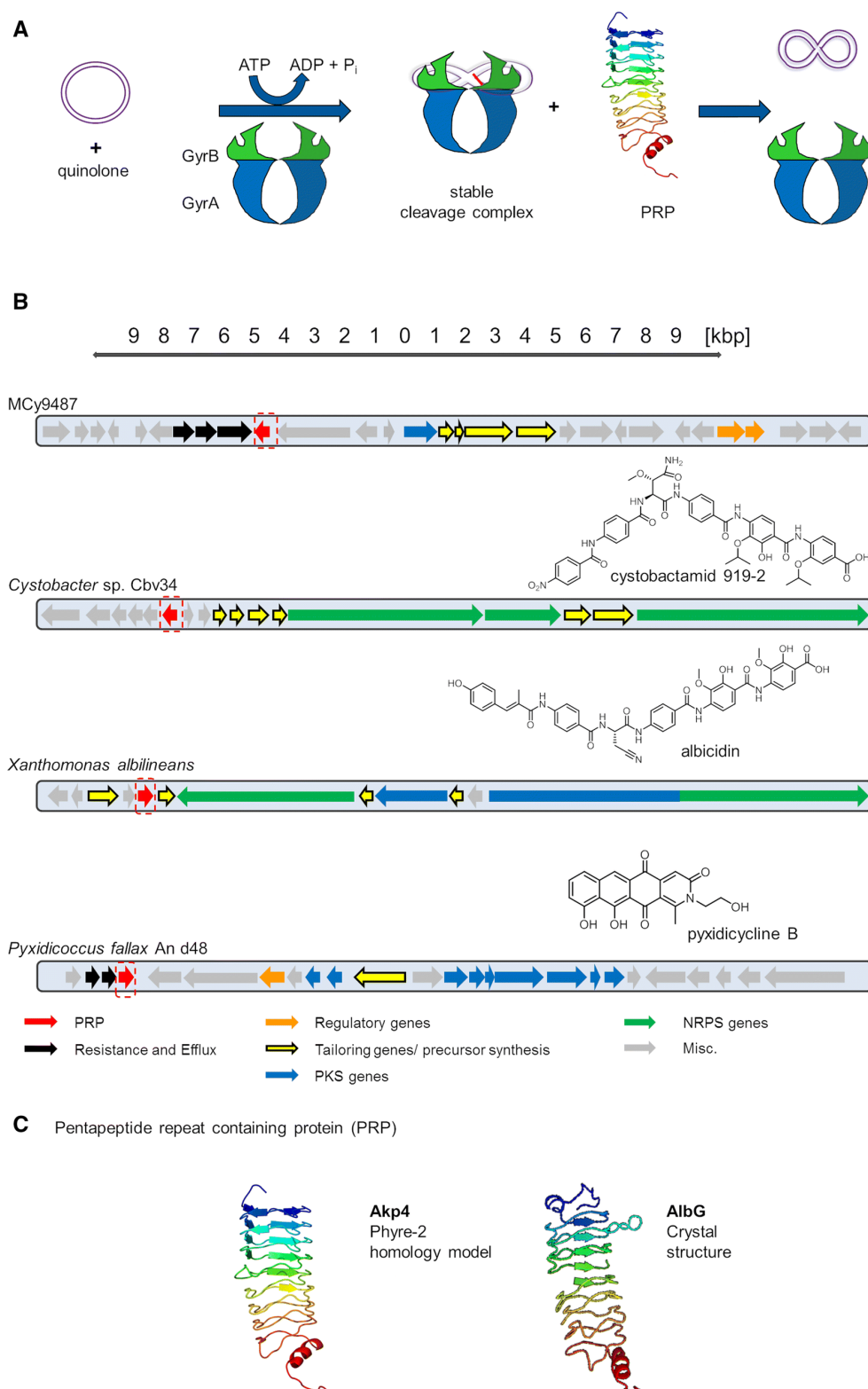
The largest family of type III PKS genes in myxobacteria (**Figure 2**, light blue, **X**) – termed the *ftp*-like gene cluster family – occurs in nearly half of all sequenced myxobacterial genomes, indicating a highly conserved (albeit to date unknown) function in the order of *Myxococcales*. All 44 members of the *ftp*-like gene cluster family harbor the aforementioned *ftpABCDE* operon structure. One interesting subgroup within this family **X** consists of eight members with the defining detail being the shortened *ftp* gene operon consisting solely of *ftpA* (type III PKS), *ftpC* (ACP) and *ftpD* (fatty acyl-AMP ligase) and missing *ftpB* (methyltransferase) and *ftpE* (monooxygenase) (**Suppl. 2.1**). This truncated operon might hint at a preference of myxobacterial type III PKSs to produce non-oxidized and non-methylated alkylresorcinols or alkylpyrones.

However, since the final polyketides produced by type III PKSs are not only dependent on the tailoring enzymes closely located to the structural genes but could also be altered by enzymes encoded in distant loci, it cannot be excluded that these pathways are producing alkylquinones instead of alkylpyrones or alkylresorcinols.

### ***Ftp*-like myxobacterial gene cluster family and an unique variant from MCy9487**

Based on *in silico* analysis of the myxobacterial type III PKS gene sequences it became obvious that the *ftp*-like PKSs – the largest family of myxobacterial type III PKS genes – are particularly widespread in myxobacteria. A detailed investigation of the *ftp* operon by *in vitro* catalysis of key enzymes and heterologous expression of the *ftpABCD* operon in *S. lividans* has been published previously<sup>13</sup>, but several open questions remain. Firstly, the apparent absence of the compounds detected *in vitro* from the secondary metabolome of *Myxococcus xanthus* according to HPLC-MS analysis raises doubts whether the compounds generated through *in vitro* reconstitution and heterologous production are in fact identical to those produced in the natural host. Secondly, given the wide distribution of *ftp*-like gene clusters among the order of myxobacteria, in particular in the suborder of *Cystobacterineae*, we were suspicious that failure to find the associated metabolites could be actually attributed to downregulation of this pathway and whether they might be “reactivated”. Furthermore, we reasoned that the function of these metabolites for myxobacteria could be possibly defined more precisely if conditions for their production were established and the corresponding products could be structurally elucidated. With regard to the alleged biological function of compounds originating from the *ftp*-like type III PKSs, one particular BGC from the myxobacterial strain MCy9487 stands out: MCy9487 harbors a type III PKS BGC starting with genes encoding putative self-resistance mechanisms, namely an ABC transporter and a gene encoding a PRP. PRPs are known to bind topoisomerases and mimic DNA in order to protect DNA gyrase and topoisomerase IV from inhibitory quinolone-like compounds<sup>31</sup> (**Figure 3A**). Gyrase is a target of quinolones facilitated by binding to a pocket consisting of the G-segment and residues of both GyrA and GyrB<sup>32</sup>. Religation is disfavored and the cleaved strands stabilize the complex of gyrase and DNA (stable cleavage complex); in this state further supercoiling is not possible. The original hypothesis (termed as G-segment-mimicry model<sup>33</sup>) proposed that putative DNA-mimicking PRPs bind to gyrase in competition with DNA<sup>34</sup> and protects gyrase from inhibitors – this inhibition model seems to be problematic since inhibition of the gyrase would also occur in the absence of potential inhibitors and PRP production would need a tight regulation. Furthermore this model appears incompatible with *in vitro* results where quinolone protection without supercoiling inhibition was observed for several PRPs<sup>33</sup>.





**Figure 3.** A) The proposed mechanism of PRPs (T-segment mimicry model) to confer host self-resistance<sup>33</sup>. During the catalytic cycle of gyrase, quinolone-binding leads to DNA stabilized in complex with gyrase termed as “stable cleavage complex”. PRP is able to recognize this conformation and destabilize the complex, that leads to the loss of quinolone binding, which allows the enzyme to continue its catalytic cycle including the supercoiling of DNA. B) Genetic organization of the cryptic type III PKS gene cluster MCy9487, *Cystobacter* sp. Cbv34, producer of cystobactamid 919-2<sup>35</sup>, *Xanthomonas albilineans*, producer of albicidin<sup>36</sup>, and *Pyxidicoccus fallax* An d48, producer of pyxidicycline B<sup>37</sup> are shown. Dashed boxes in red indicate genes encoding pentapeptide repeat proteins (PRPs) to confer resistance to the produced natural product. C) Comparison of the Phyre-2 structure homology model of the putative PRP encoded by *akp4* and the crystal structure of the PRP AlbG.

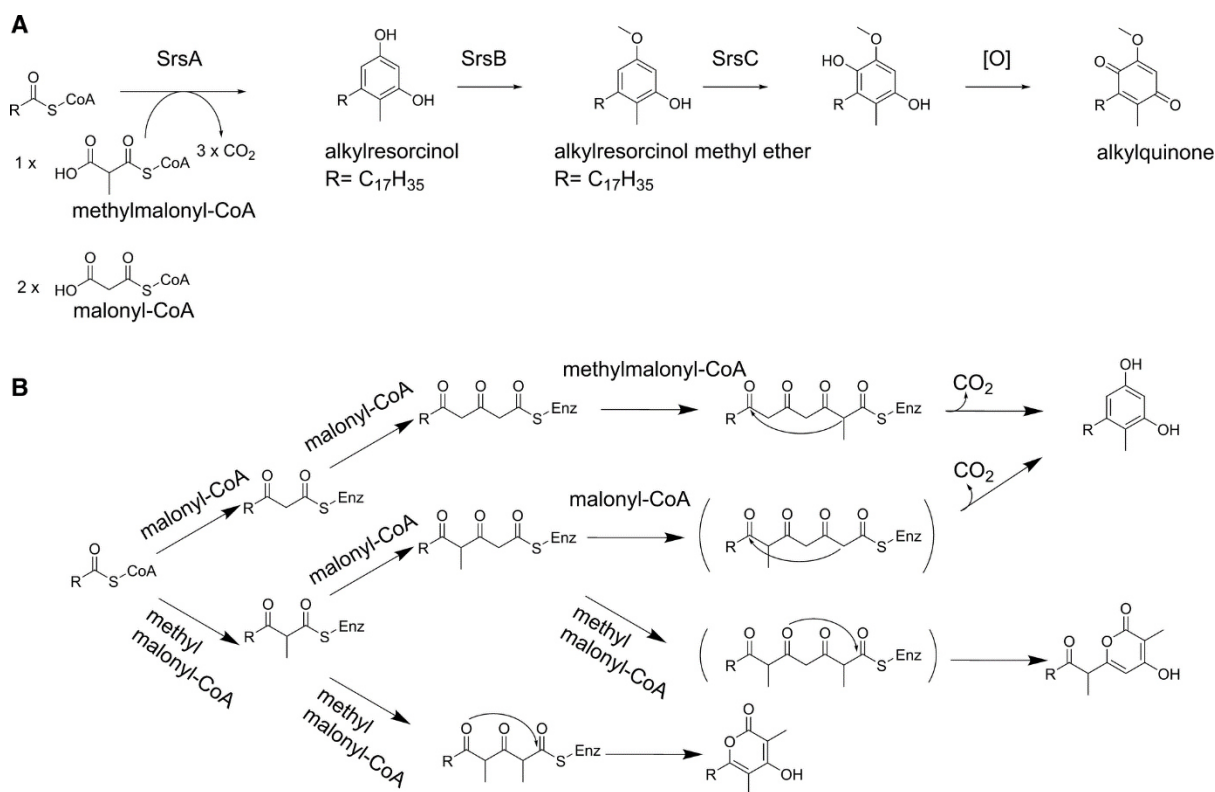
The alternative model (termed as T-segment-mimicry model<sup>33</sup>) describes that PRPs are able to recognize this stable cleavage complex and destabilize this conformation, leading to the loss of quinolone binding, which allows the enzyme to continue its catalytic cycle including the supercoiling of DNA. The inhibition of gyrase at higher concentration of PRP would be a side effect of a weaker affinity to unbound or DNA-bound topoisomerase. A few PRPs have been characterized to confer self-resistance against topoisomerase poisons like in the producers of albicidin<sup>38</sup> and cystobactamids<sup>35</sup>. Conducted supercoiling assays of purified AlbG (0.65  $\mu$ M) protected gyrase from the effects of albicidin yielding 2–4 fold increased half-maximal inhibitory concentration ( $IC_{50}$ ) values without no significant effect on the sensitivity to ciprofloxacin<sup>38</sup> and partially inhibited the supercoiling activity of DNA gyrase in the absence of albicidin with an  $IC_{50}$  of 6  $\mu$ M. Crystallographic analysis provided the crystal structure of AlbG at 2 Å<sup>39</sup>; however without the crystal structure of the AlbG and gyrase/topoisomerase IV interaction complex, the exact interaction of AlbG remains elusive.

Following the rationale that unassigned BGCs encoding PRPs could actually be responsible for production of topoisomerase inhibitors recently led to the discovery of pyxidicyclines<sup>37</sup> (**Figure 3B**). Thus, the finding in MCy9487 could be seen as a hint towards the putative mode of action for the encoded type III PKS product and could ultimately lead to assignment of a biological function of the yet unidentified metabolite. To strengthen this hypothesis, the putative PRP in this type III PKS cluster named *akp4* was further evaluated by *in silico* protein homology analogy recognition engine 2 (Phyre2), and results indeed underline the structural similarity of Akp4 to previously described PRPs such as AlbG<sup>40</sup> (**Figure 3C**). The *akp* operon structure in MCy9487 closely resembles the architecture previously described from *Myxococcus xanthus* Beebe 1941 (DSM No. 52389) used for *in vitro* reconstitution of the fatty acyl-AMP ligase-dependent variant of type III polyketide synthases<sup>13</sup>. Furthermore, a highly similar BGC also exists in the myxobacterial model strain *M. xanthus* DK1622, leading to the assumption that these type III PKS gene operons could be capable of generating the same small molecule products.

Antibiotics and secondary metabolite analysis shell (antiSMASH) analysis of the type III PKS genes/operons in MCy9487 and *M. xanthus* DK1622 reveals high sequence similarity (BlastP: MCy9487/DK1622; 52% identity, 96% coverage) to the alkylresorcinol biosynthetic gene cluster from *Streptomyces griseus* IFO 13350<sup>41</sup>. This alkylresorcinol biosynthetic gene cluster consists of the *srsABC* operon, in which *srsA* encodes a type III PKS, *srsB* encodes a methyltransferase, and *srsC* encodes a flavoprotein hydrolase. Heterologous expression of the *srs* genes in *Streptomyces lividans* showed that SrsA synthesizes phenolic lipids, alkylresorcinols and alkylpyrones, SrsB methylates the generated phenolic lipids to yield alkylresorcinols methyl ethers while SrsC is acting as hydroxylase (**Figure 4**)<sup>26</sup>. However, the operon lacks an acyl carrier protein and fatty acyl-AMP ligase to activate the incorporated fatty acid and subsequently activate its transfer onto the acyl carrier protein to form finally the type III PKS metabolite. While *in vitro* reconstitution of the type III PKS pathway from *Myxococcus* yielded alkylresorcinols of different chain lengths, the observed BGC aberration makes the *in vivo* metabolites produced by this gene operon appear rather elusive. We therefore set out to investigate the *ftp*-like BGCs from DK1622 and MCy9487 with the objective of connecting a small molecule product to these biosynthetic pathways using two different experimental approaches.

The first approach comprises the heterologous expression of the PRP-accompanied type III PKS operon including adjacent genes from MCy9487, since this strain was not amenable to direct genetic manipulation despite considerable efforts to establish a method using electroporation or conjugation-

mediated single crossover recombination, thus impeding the activation of the BGC through promoter insertion (**Suppl. 1.2**)<sup>42</sup>. The second approach involves targeted overexpression of the *ftp*-like type III PKS operon within the myxobacterial model host *M. xanthus* DK1622, since genetic manipulation by site-specific single crossover recombination is straight-forward for this strain.



**Figure 4.** Reactions catalyzed by Srs proteins adapted from Funabashi *et al.*<sup>26</sup> **A**) Alkylquinone is formed by the Srs proteins SrsA, SrsB and SrsC. The last conversion occurs non-enzymatically. **B**) Reaction routes yielding phenolic lipids, alkylresorcinols and alkylpyrones.

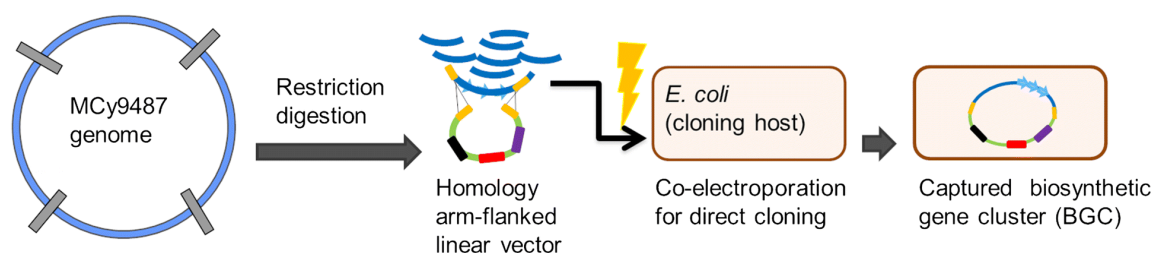
## Heterologous expression and modification of a cryptic type III PKS gene cluster from MCy9487

For heterologous expression of the candidate BGC from MCy9487 two cluster fragments of different size were used for the subsequent direct cloning approach. That way we determined the borders of the biosynthetic gene cluster. The larger cluster fragment consists of 28.126 kb and includes genes between *orf7* downstream of the biosynthetic gene cluster and *orf23* upstream of the biosynthetic gene cluster, while the smaller cluster fragment consist of 19.816 kb comprising only the essential type III PKS BGC (**Figure 5C**). The digested genomic DNA was co-electroporated with the homology arm-flanked linear vector into arabinose-induced *E. coli* GB05RedTrfA for RecET direct cloning (**Figure 5A**)<sup>43</sup>. Candidate clones were analyzed by restriction digestion to reveal correct homologous recombination of plasmid backbone and gene cluster fragment. Subsequent modifications involved the replacement of the original ampicillin resistance gene through a tetracycline resistance gene followed by the incorporation of a kanamycin resistance gene.

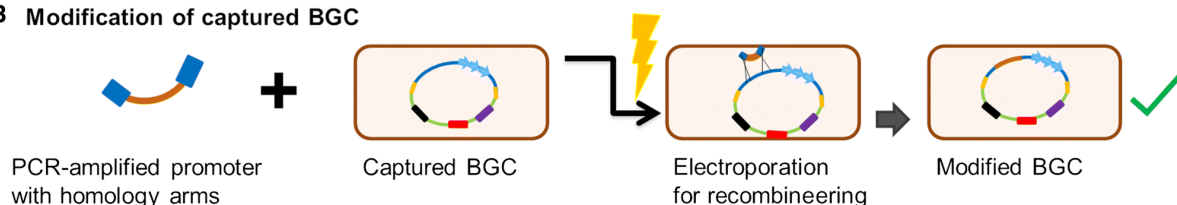
This second modification was conducted either by Red $\alpha\beta$  recombineering (**Figure 5B**) to generate two constructs harboring the artificial promoter system in front of the type III PKS operon or by conventional restriction ligation where the kanamycin-tn5 promoter system is located outside of the biosynthetic gene cluster.

The kanamycin resistance gene with the heterologous tn5 promoter was used on the one hand as selection marker in the heterologous host *M. xanthus* DK1622  $\Delta mxchrA-tetR$  and on the other hand for constitutive gene expression of the putative “silent” type III PKS gene cluster, for those constructs with the kanamycin-tn5 promoter system in front of the type III PKS. HPLC-MS analysis of extracts derived from mutant strain cultivations revealed that only constructs with the kanamycin-tn5 promoter system in front of the type III PKS afforded the apparent production of two new metabolites. The “late” retention time of 16.8 min and 17.2 min of the newly emerging signals may be seen as an early hint that the expected aromatic structure possesses a fatty acyl side chain (**Figure 6**).

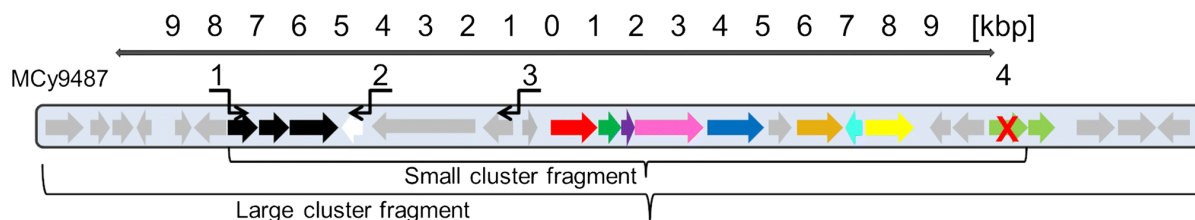
#### A Direct cloning of MCy9487 BGC



#### B Modification of captured BGC



#### C Performed modification on MCy9487 BGC



**Figure 5.** Schematic direct cloning strategy and modification for heterologous expression of the cryptic type III PKS BGC from MCy9487<sup>43</sup>. **A)** Digested genomic DNA is co-electroporated with the homology arm-flanked linear vector into arabinose-induced *E. coli* GB05RedTrfA for the RecET direct cloning to “capture” the BGC. **B)** Genetic manipulation within the BGC can be performed efficiently in *E. coli* with PCR-amplified cassettes flanked by homology arms to enable Red $\alpha\beta$  recombineering. **C)** Schematic overview of the genetic organization of the small and large cluster fragment. Modifications 1, 2 and 3 comprises insertion of a vanillate-inducible promoter system fused with an ampicillin resistance gene in front of an ABC transporter (1), pentapeptide repeat protein (2) or in front of a methyltransferase (3), whereas (4) involves the replacement of the putative repressor through an ampicillin cassette.

## Induced gene expression of type III PKS gene cluster from DK1622

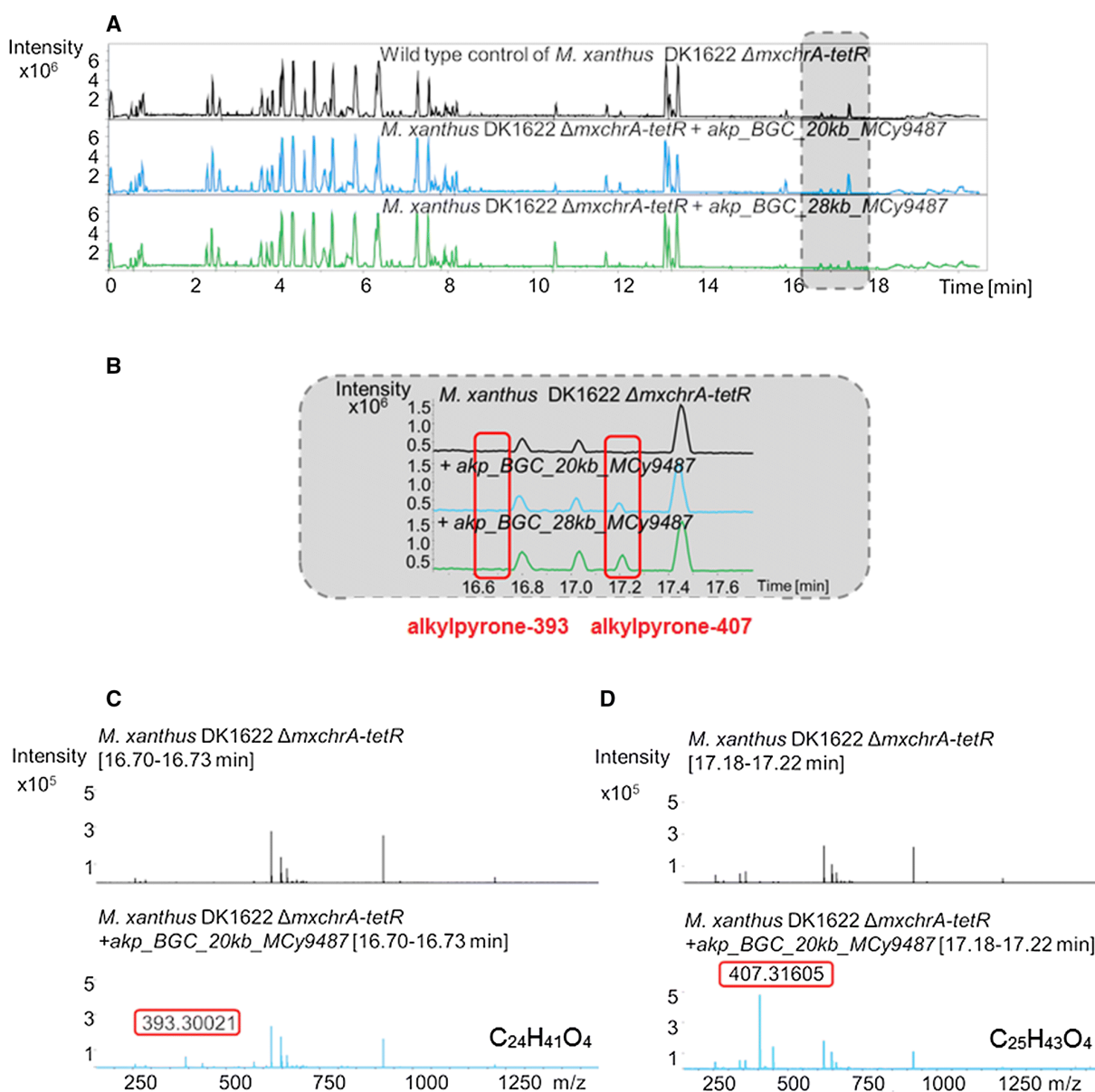
As described previously, homologues of the type III PKSs investigated in this work are very widespread among myxobacteria. Incidentally, a homologue of the minimal PKS system observed in MCy9487 is also present in the myxobacterial model strain *M. xanthus* DK1622. Since the myxobacterial model host *M. xanthus* DK1622 is accessible by genetic manipulation, a different strategy was conducted to investigate the representative of the large family of *ftp*-like type III PKS genes found in this strain. A vanillate-inducible promoter system was inserted in front of the *ftp* operon in *M. xanthus* DK1622 *ΔmxchrA-tetR* by single crossover recombination to control the *ftp* operon through supplementation of vanillate. Upon HPLC-MS analysis of mutant cultivations, the outcome of the vanillate-induced gene expression actually turned out to be identical to the heterologous expression of the *ftp*-like operon originating from MCy9487 (**Figure S1**).

## Results of heterologous expression and induced gene expression

Since the obtained high resolution masses from heterologous expression and induced gene-expression are precisely the same and the alleged new metabolites also display identical retention times, the secondary metabolites produced by the type III PKS from MCy9487 and DK1622 are considered to be the same pair of small molecules. As gene expression of the *ftp* operon in *M. xanthus* DK1622 and the *ftp*-like biosynthetic gene cluster in the native host MCy9487 appear to be downregulated under laboratory fermentation, the biosynthetic gene cluster of the heterologously expressed type III PKS BGC from MCy9487 was modified by removing a putative repressor. The BLAST result of *akp16* highlighted a gene encoding a putative transcriptional regulator of the GntR family that shows significant structural similarity to the crystallized autorepressor and transcriptional regulator GabR<sup>44</sup> which might cause downregulation of expression. Therefore, the gene *akp16* was replaced by an ampicillin cassette through direct cloning in the constructs, in which the *ftp*-like operon is under the control of the constitutive kanamycin-tn5 promoter and in those constructs where the system is located outside the biosynthetic gene cluster. However, the metabolic profiles did not display improved production compared to the previous constructs.

Further modification steps were inspired by the heterologous expression of the bottromycin BGC in *Streptomyces coelicolor*, which enabled a 20-fold higher production of bottromycin through insertion of the *PerME\** promoter in front of the efflux pump *botT*<sup>45</sup>. A vanillate-inducible promoter system fused with an ampicillin resistance gene in front of the putative self-resistance genes in the *ftp*-like operon was inserted. One insertion was made in front of the gene operon *akp1–3*, whereas other approaches involved the insertion in front of *akp4* or in front of *akp6* (**Figure 5C**) to induce the expression of the ABC transporter (*akp1–3*), the expression of the putative self-resistance mechanism of the pentapeptide repeat protein (*akp4*) or the expression of the putative operon system consisting of genes encoding a methyltransferase (*akp6*), an isoamylase (*akp5*) and a pentapeptide repeat protein (*akp4*). For this purpose the constructs pBR322-*tetR-tetReg-tetO*\_MCy9487\_akpBGC\_large\_fragment and pBR322-*tetR-tetReg-tetO*\_MCy9487\_akpBGC\_small\_fragment (genetic constructs No. 3 and No. 5, **Table S5**) were used to obtain finally six modified constructs (genetic constructs No. 13–18, **Table S5**).

However, in none of these constructs the production of new metabolites could be observed by HPLC-MS profiling. These results point towards the conclusion that limited genetic expression of the structural genes are not caused by polar effects from expression of the investigated neighboring genes. Additional evidence that these genes are not playing a critical role for the expression of the *ftp*-like operon comes from the observation that other *ftp*-like operons in other myxobacterial strains are not harboring any of these tailoring enzymes.



**Figure 6.** A) HPLC-MS BPC chromatograms of crude extracts from *M. xanthus* DK1622  $\Delta$ *mxchrA-tetR* + *akp\_BGC\_20kb\_MCy9487* and *M. xanthus* DK1622  $\Delta$ *mxchrA-tetR* + *akp\_BGC\_28kb\_MCy9487*. Red dashed box indicates magnified area of the HPLC-MS BPC chromatograms, shown in B. The crude extract of *M. xanthus* DK1622  $\Delta$ *mxchrA-tetR* was used as control. Red boxes in B indicate monitoring time of C and D. C, D) MS-spectra of heterologous expression extracts of the type III PKS from MCy9487. MS-spectra C shows the major occurring mass 393.300 Da ( $C_{24}H_{41}O_4$ ) and MS-spectra D is showing the major occurring mass 407.316 Da ( $C_{25}H_{43}O_4$ ) (D). The MS-spectrum of *M. xanthus* *mxchrA-tetR* DK1622 was used as control.

## Structure elucidation of myxobacterial type III polyketides and investigation of molecular target

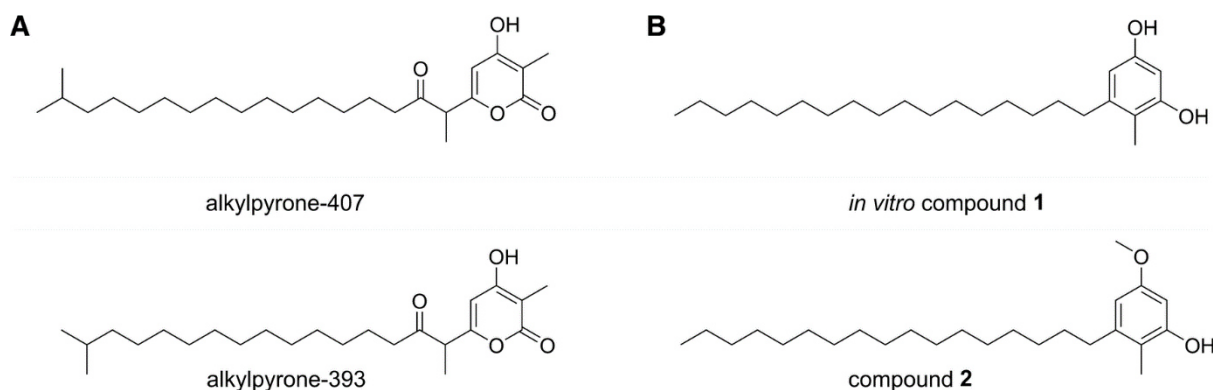
For the purpose of full structural elucidation, the two candidate compounds with masses 407.316 m/z ( $C_{25}H_{43}O_4$ ) and 393.29 m/z ( $C_{24}H_{41}O_4$ ) had to be produced and purified in large-scale. Since the highest production rate of both major target masses was observed in *M. xanthus* DK1622  $\Delta mxchrA-tetR$  harboring the larger MCy9487\_ftp gene cluster fragment, this mutant was selected for large-scale fermentation.

Both purified candidate compounds were dissolved in methanol- $d_4$  for subsequent NMR measurement. Structural elucidation confirmed the amphiphilic structure of a fatty acid connected to a polar cyclic moiety. Both structures are different from the *in vitro* produced and literature-reported *in vitro* compound **1** (through FtpA, FtpC, FtpD) or compound **2**, which was heterologously produced through *S. lividans* harboring pIJ6021-ftpABCD (**Figure 7**). This result is unexpected, since induced gene expression of the operon ftpABCDE in the native host did only yield very diminutive trace amounts of alkylresorcinols, but decent amount of tetraketide alkylpyrones (**Figure S4**). As reported by Hayashi *et al.*, the *in vitro* product produced by the Ftp proteins are quite expected to be different from the natural product produced by *M. xanthus*, hence the converted stearic acid is not detected in the *M. xanthus* lipid composition<sup>46</sup>. Instead, the fatty acids incorporated are rather unusual ones namely 14-methylpentadecanoic acid and 14-methylhexadecanoic acid, both detectable in the lipid composition of *M. xanthus*<sup>46</sup>. In light of the finding that no formation of alkylpyrones were reported either in conducted *in vitro* experiments with recombinant produced FtpA, FtpC and FtpD or through heterologous expression in *S. lividans* harboring pIJ6021-ftpABCD, the reason why induced gene expression of the ftp genes in the native host did only yield tetraketide alkylpyrones remains elusive.

In addition, heterologous overexpression in *Streptomyces lividans* of the related srsA gene previously afforded the formation of alkylpyrone-407 and alkylpyrone-393 in low amounts, along with their resorcinol derivatives in higher amount. Furthermore, targeted re-evaluation of the secondary metabolome of MCy9487 wild type with a method optimized for higher sensitivity in the target mass range reveals trace amounts alkylpyrone-407 and alkylpyrone-393. Still, no metabolites in the reported range of retention time indicate the presence of the related alkylquinones, alkylresorcinols or alkylresorcinol methyl ether, supporting the conclusion of alkylpyrones being the major type III polyketides afforded *in vivo* by the ftp-like operons (**Figure S3**).

The function of the produced polyketides of the Srs proteins, had been previously hypothesized to be associated with the cytoplasmic membrane, since srsA gene disruption in the native host *S. griseus* conferred resistance to  $\beta$ -lactam antibiotics on the host<sup>26</sup> (**Figure 4**). Since a PRP is located in close proximity to the alkylpyrone BGC from MCy9487, we propose the compounds afforded by this biosynthetic gene cluster to act as topoisomerase inhibitors providing an alternative explanation for the biological function of these produced polyketides. Structural similarity of alkylpyrone-407 and alkylpyrone-393 to hexylresorcinol (**Figure 7**) – indeed described as topoisomerase IV inhibitor with an  $IC_{50}$  of 40  $\mu M$  – strengthened the hypothesis that the alkylpyrones might indeed target topoisomerases<sup>47</sup>. Hence, the topoisomerase inhibition activity toward bacterial topoisomerases was tested, namely *E. coli* topoisomerase II (gyrase) and topoisomerase IV.





**Figure 7.** A) Isolated myxobacterial compounds alkylpyrone-407 and alkylpyrone-393. B) *In vitro* compound 1 could be produced through *in vitro* catalysis of stearic acid with recombinantly produced FtpA, FtpC and FtpD, compound 2 was isolated from heterologous expression of the operon *ftpABCD* by *Streptomyces lividans*, harboring pIJ6021-*ftpABCD*<sup>13</sup>.

The bacterial topoisomerase IV was inhibited with IC<sub>50</sub> values of 1.6–3.2 µg/mL (3.94–7.88 µM) by alkylpyrone-407 (**Figure S7**), unlike bacterial topoisomerase II (gyrase) whose activity was not inhibited at concentrations up to 50 µg/mL (123.06 µM). In comparison to the described topoisomerase inhibitors like albicidin, the cystobactamids and the pyxidicyclines, the topoisomerase IV relaxation assay of alkylpyrone-407 displayed potent inhibition with IC<sub>50</sub> values comparable to those of the pyxidicyclines (**Table 1**). These IC<sub>50</sub> values do not match with the observed results of bactericidal activity testing.

We reason that either transport-related effects might be accountable for the absence of any antimicrobial activity or the molecular target is not bacterial topoisomerase but has eukaryotic origin. As the alkylpyrone-407 exhibited cytotoxicity in our bioassay, inhibition of human topoisomerase I was evaluated. This enzyme is the human counterpart for unwinding topoisomerases such as topoisomerase IV. The IC<sub>50</sub> values 1.6–3.2 µg/mL of the alkylpyrone-407 against human topoisomerase I was identical to the result of topoisomerase IV (**Figure S8**). The MIC of alkylpyrone-407 against HCT-116 cells was determined to 12 µg/mL. Taken together, alkylpyrone-407 is a type III polyketide with significant *in vitro* bioactivity.

**Table 1.** Half-maximal inhibitory concentrations (IC<sub>50</sub> values in µM) of alkylpyrone-407, albicidin, cystobactamide 912-1 and 912-2, pyxidicycline A and B in gyrase supercoiling assays and topoisomerase (topo) IV relaxation assays.

	alkylpyrone 407	albicidin	cystobactamid 912-1	cystobactamid 912-2	pyxidicycline A	pyxidicycline B
<i>E. coli</i> gyrase	–	0.04	21.3 ± 6.0	0.26 ± 0.06	–	–
<i>E. coli</i> topo IV	3.94–7.88	0.3	89.5 ± 0.8	38.0 ± 0.7	8.76–17.12	4.38–8.76



## 6.4 Conclusions

This study sets the stage for a comprehensive approach to assess the hidden potential of myxobacteria to produce type III polyketides, considering that no such metabolites have been assigned to the corresponding BGCs until now. To the best of our knowledge, for the first time a type III PKS BGC from myxobacterial origin has been activated by heterologous/induced gene expression and isolated from a myxobacterium, namely the model host *M. xanthus* DK1622. As logical follow-up to the identification of alkylpyrones and their characterization as topoisomerase inhibitors, the same overexpression platform established in this project can now be applied to representatives of the diverse BGC architectures found in myxobacterial genomes. In that sense the *ftp*-like family – present in nearly half of all sequenced strains in our myxobacterial in-house genome database – has now been taken off the genome mining candidate list, with plenty of intriguing type III PKS operons still remaining to be analyzed. Our *in silico* survey highlights that myxobacterial strains are harboring indeed a promising repertoire for several new type III PKSs synthesizing likely unanticipated type III PKSs. In particular, the type III PKSs from the family **III**, **IV** and **V** are representing type III PKS BGC showing conserved genetic organization with diverse tailoring reactions. Future type III polyketide discovery in myxobacteria appears promising when conducted through careful heterologous expression of these promising while phylogenetically distant myxobacterial type III PKS BGCs.

## 6.5 Material and Methods

### Applied software, sequence analysis and bioinformatic methods

A comprehensive dataset containing myxobacterial genome sequences from our in-house and publically available database has been analyzed. Biosynthetic gene cluster identification of genes encoding type III PKSs and phylogenetic clustering of BGCs were automatically performed by our in-house software “Clust-o-matic”, which yielded 116 biosynthetic gene clusters of which 48 derives from our private genome bank (**Table S1**). Verification of automated phylogenetic clustering, hierarchical clustering of type III PKSs and separation into different subclades was performed manually. Geneious 10.2.2 (Biomatters Ltd., Auckland, New Zealand) was used to design primers for PCR and sequencing, to create plasmid maps and to identify open reading frames (ORF). All pairwise and multiple alignments of nucleotide or amino acid sequences were performed with the plugin software from Geneious by using the MUSCLE (Multiple Sequence Comparison by Log- Expectation) alignment (3.8.425 by Robert C. Edgar) since it claims to achieve higher average accuracy and higher speed than ClustalW2 or T-Coffee algorithm<sup>48,49</sup>. In order to find homologue genes or proteins, either the nucleotide or amino acid sequence of interest was aligned with the BLAST against our in-house genome database or the publically available nucleotide or amino acid database. The functional prediction of ORFs was performed by using protein blast and/or blastx program (<https://blast.ncbi.nlm.nih.gov/Blast.cgi>) and Pfam (<http://pfam.xfam.org>)<sup>50</sup>. The in-house standard extract database embedded in the software bundle Mxbase Explorer 3.2.27 was used for the search of metabolites for the subsequent correlation to their corresponding type III polyketide genes.

## Molecular cloning, construction of plasmids, maintenance of bacterial cultures

Routine handling of nucleic acids, such as isolation of plasmid DNA, restriction endonuclease digestions, DNA ligations, and other DNA manipulations, were performed according to standard protocols<sup>51</sup>. *E. coli* HS996 (Invitrogen) was used as host for standard cloning experiments and *E. coli* SCS110 (Stratagene) for preparation of plasmid DNA free of *Dam* or *Dcm* methylation. *E. coli* strains were cultured in LB liquid medium or on LB agar (1% tryptone, 0.5% yeast extract, 0.5% NaCl, (1.5% agar) at 30–37 °C and 200 rpm) overnight. Antibiotics were used at the following final concentrations: 100 µg/mL ampicillin, 50 µg/mL kanamycin, 12 µg/mL oxytetracycline. Transformation of *E. coli* strains were achieved via electroporation in 0.1 cm-wide cuvettes at 1250 V, a resistance of 200 Ω, and a capacity of 25 µF. Plasmids were purified either by standard alkaline lysis<sup>51</sup> or by using the GeneJet Plasmid Miniprep Kit (Thermo Fisher Scientific) or the NucleoBond PC100 kit (Macherey-Nagel). Restriction endonucleases, alkaline phosphatase (FastAP) and T4 DNA ligase were purchased from Thermo Fisher Scientific. Oligonucleotides used for PCR and sequencing were acquired from Sigma Aldrich and are listed in **Table S2**. PCRs were carried out in a Mastercycler® pro (Eppendorf) using Phusion™ High-Fidelity according to manufacturer's protocol. Temperature and duration setting for each thermocycling step in PCR with Phusion™ High-Fidelity polymerase were performed as follows: Initial denaturation (30 s, 98 °C); 33 cycles of denaturation (15 s, 98 °C), annealing (15 s, 53–72 °C, depending on melting temperature of primers) and elongation (based on PCR product length 30 s/ 1 kb, 72 °C); and final extension (10 min, 72 °C). When amplification of desired PCR product could not be achieved according manufacturer's instructions, additives like DMSO (5%), glycerol (8%) and betaine (0.5 M) were supplemented. PCR products or DNA fragments from restriction digestions were purified by agarose gel electrophoresis and isolated using the PCR clean-up gel extraction kit using Nucleo Spin® (Macherey-Nagel). To recover DNA fragments larger than 10 kb the Agarose Gel DNA Extraction Kit from Roche was used. Rec/ET direct cloning of MCy9487 type III PKS BGC and Redαβ recombineering experiments for plasmid modifications<sup>43,52</sup> using the strains *E. coli* GB05RedTrfA/pSC101-BAD-ETgA-tet or *E. coli* GB08-red were performed according to authors' s instructions<sup>43</sup> (**Suppl. 1.2**). After selection with suitable antibiotics, clones harboring correct recombination products were identified by plasmid isolation and restriction analysis with a set of different restriction endonucleases. In addition to restriction analysis, integrity of the constructs for induced gene expression was verified by sequencing.

According to a previously established electroporation procedure for *Myxococcus xanthus* DK1622<sup>53</sup> the host strain *M. xanthus* DK1622  $\Delta$ *mxchrA-tetR*<sup>54</sup> was transformed with the generated expression constructs (**Table S5**). *M. xanthus* DK1622 mutants were routinely cultivated at 30 °C in CTT medium or an CTT agar (1% casitone, 10 mM Tris buffer pH 7.6, 1 mM KH<sub>2</sub>PO<sub>4</sub> pH 7.6, 8 mM MgSO<sub>4</sub> (1.5% agar) pH adjusted to 7.6). Liquid cultures were grown in Erlenmeyer flasks on an orbital shaker at 180 rpm for 3–6 days. *M. xanthus* transformants were selected by adding 50 µg/mL kanamycin to the fermentation culture. Correct chromosomal integration of the expression constructs via homologous recombination into the *tetR* locus was confirmed by PCR. Genomic DNA of the transformants was isolated using the Gentra® Puregene® Yeast/Bact. Genomic DNA Purification Kit (Qiagen) according to manufacturer's instructions.

## Analysis of secondary metabolism of broth extracts

The secondary metabolism of broth extracts was analyzed by HPLC-HRESI-DAD-MS on a Bruker maXis 4G mass spectrometer coupled with a Dionex Ultimate 3000 RSLC system using a BEH C18 column (100 × 2.1 mm, 1.7 μm, Waters, Germany) with a gradient of 5–95% acetonitrile (ACN) + 0.1% formic acid (FA) in H<sub>2</sub>O + 0.1% FA at 0.6 mL/min and 45 °C over 18 min with UV detection by a diode array detector at 200–600 nm. Mass spectra were acquired from 150 to 2000 m/z at 2 Hz. The detection was performed in the positive MS mode. The plugin for Chromeleon Xpress (Dionex) was used for operation of UltiMate 3000 LC System. HyStar (Bruker Daltonic) was used to operate on maXis 4G speed MS system. HPLC-MS mass spectra were analyzed with DataAnalysis 4.2 (Bruker Daltonic).

## Compound isolation

The cell pellet of a 10 L fermentation culture with 2% XAD-16 was used to perform liquid-liquid extraction to concentrate the alkyldipyrone in the ethyl acetate phase (**Suppl. 1.3**). Subsequent fractionation was performed on the flash chromatography system Biotage Isolera One (**Suppl. 1.3**). Semi-preparative HPLC purification was done using a Dionex Ultimate 3000 SDLC low pressure gradient system on a Phenomenex Synergi 4μm Fusion RP 80A, 250 x 10 mm column. Column temperature was stabilized at 45 °C with the eluents H<sub>2</sub>O + 0.1% FA as **A** and ACN + 0.1% FA as **B**, at a flow rate of 4.72 mL/min. Detection of the alkyldipyrone was facilitated via mass spectrometry on the Agilent 1100 series coupled to the HCT 3D ion trap or with a UV detector on the Dionex 3000 SL systems by UV absorption at 256 nm and 320 nm. The gradient that starts with a plateau at 95% **A** for 2 min followed by a ramp to 59% **A** during 6 min. Then **A** content is ramped to 30% during 17 min and finally ramped to 5% **A** during 1 min. **A** content is kept at 5% for 1 min and then ramped back to 95% during 30 s. The column was re-equilibrated at 95% **A** for 5 min. The pure white compounds are subsequently dried by lyophilization. 0.3 mg of alkyldipyrone-393 and 0.6 mg of alkyldipyrone-407 have been isolated applying this protocol. The compounds are obtained as colorless solids.

## NMR-based structure elucidation

The chemical structures of the alkyldipyrone-393 and alkyldipyrone-407 were determined via multidimensional NMR analysis. <sup>1</sup>H NMR, <sup>13</sup>C NMR and 2D spectra were recorded at 700 MHz (1H)/175 MHz (<sup>13</sup>C) conducting an Ascend 700 spectrometer using a cryogenically cooled triple resonance probe (Bruker Biospin, Rheinstetten, Germany). Samples were dissolved in methanol-*d*<sub>4</sub>. Chemical shifts are reported in ppm relative to TMS, the solvent was used as the internal standard (**Suppl. 2.4**).

### Bioactivity profiling

Cell based biological assays to valuate MIC values on bacteria and cancer cell lines were performed as described in **Suppl. 1.4**. Topoisomerase inhibition assays were performed to determine the molecular target of alkylpyrone-407. All topoisomerase related inhibition assays were performed using topoisomerase inhibition evaluation kits for *E. coli* topoisomerase II, *E. coli* topoisomerase IV and human topoisomerase I supplied by Inspiralis (Inspiralis Ltd., Norwich, U.K.). The kits were used according to manufacturer's instructions. Compound was tested as freshly prepared DMSO solution. Total volume of the assay reaction was 30  $\mu$ L. All assays were carried out in a 96-well micro titer plate. Topoisomerase inhibition assays were evaluated by agarose gel electrophoresis using 0.8% agarose gels in 1 x TAE buffer (**Figure S7, S8**). No inhibition of *E. coli* DNA gyrase was observed at concentration up to 50  $\mu$ g/mL of alkylpyrone-407 (data not shown).

### 6.6 Acknowledgements

The authors thank Alexander Popoff for recording the NMR spectra of the alkylpyrones, Viktoria Schmitt and Jennifer Herrmann for performing bioactivity assays, Ronald Garcia for microbiological assistance and Nestor Zaburannyi for bioinformatic support. Joachim J. Hug was supported by a PhD fellowship of the Boehringer Ingelheim Fonds.

### 6.7 Conflict of Interest

The authors declare no conflict of interest.

### 6.8 Key words

Myxobacteria; Genome mining; Natural products; Type III polyketide synthase; Self-resistance

## 6.9 References

1. Huang, T. & Lin, S. Microbial Natural Products. A Promising Source for Drug Discovery. *Appl. Microbiol. Biotechnol.* **1** (2017).
2. Hoffmann, T., Krug, D., Hüttel, S. & Müller, R. Improving natural products identification through targeted LC-MS/MS in an untargeted secondary metabolomics workflow. *Anal. Chem.* **86**, 10780–10788 (2014).
3. Herrmann, J., Fayad, A. A. & Müller, R. Natural products from myxobacteria: novel metabolites and bioactivities. *Nat. Prod. Rep.* **34**, 135–160 (2017).
4. Wenzel, S. C. & Müller, R. The impact of genomics on the exploitation of the myxobacterial secondary metabolome. *Nat. Prod. Rep.* **26**, 1385–1407 (2009).
5. Wenzel, S. C. & Müller, R. Myxobacterial natural product assembly lines: fascinating examples of curious biochemistry. *Nat. Prod. Rep.* **24**, 1211–1224 (2007).
6. Austin, M. B. & Noel, J. P. The chalcone synthase superfamily of type III polyketide synthases. *Nat. Prod. Rep.* **20**, 79–110 (2003).
7. Parvez, A. *et al.* Novel Type III Polyketide Synthases Biosynthesize Methylated Polyketides in *Mycobacterium marinum*. *Sci. Rep.* **8**, 6529 (2018).
8. Yu, D., Xu, F., Zeng, J. & Zhan, J. Type III polyketide synthases in natural product biosynthesis. *IUBMB Life* **64**, 285–295 (2012).
9. Morita, H. *et al.* Synthesis of unnatural alkaloid scaffolds by exploiting plant polyketide synthase. *Proc. Natl. Acad. Sci. USA* **108**, 13504–13509 (2011).
10. Chemler, J. A. *et al.* Biochemical and structural characterization of germicidin synthase. Analysis of a type III polyketide synthase that employs acyl-ACP as a starter unit donor. *J. Am. Chem. Soc.* **134**, 7359–7366 (2012).
11. Shimizu, Y., Ogata, H. & Goto, S. Type III Polyketide Synthases. Functional Classification and Phylogenomics. *ChemBioChem* **18**, 50–65 (2017).
12. Miyanaga, A., Funa, N., Awakawa, T. & Horinouchi, S. Direct transfer of starter substrates from type I fatty acid synthase to type III polyketide synthases in phenolic lipid synthesis. *Proc. Natl. Acad. Sci. USA* **105**, 871–876 (2008).
13. Hayashi, T., Kitamura, Y., Funa, N., Ohnishi, Y. & Horinouchi, S. Fatty acyl-AMP ligase involvement in the production of alkylresorcylic acid by a *Myxococcus xanthus* type III polyketide synthase. *ChemBioChem* **12**, 2166–2176 (2011).
14. Schuz, R., Heller, W. & Hahlbrock, K. Substrate specificity of chalcone synthase from *Petroselinum hortense*. Formation of phloroglucinol derivatives from aliphatic substrates. *J. Biol. Chem.* **258**, 6730–6734 (1983).

15. Minowa, Y., Araki, M. & Kanehisa, M. Comprehensive Analysis of Distinctive Polyketide and Nonribosomal Peptide Structural Motifs Encoded in Microbial Genomes. *J. Mol. Biol.* **368**, 1500–1517 (2007).
16. Witte, S. N. R., Hug, J. J., Géraldy, M., Müller, R. & Kalesse, M. Biosynthesis and Total Synthesis of Pyrroazonol B: a Secondary Metabolite from *Nannocystis pusilla*. *Chem. Eur. J.* **23**, 15917–15921 (2017).
17. Austin, M. B. *et al.* Crystal Structure of a Bacterial Type III Polyketide Synthase and Enzymatic Control of Reactive Polyketide Intermediates. *J. Biol. Chem.* **279**, 45162–45174 (2004).
18. Gokulan, K. *et al.* Crystal structure of *Mycobacterium tuberculosis* polyketide synthase 11 (PKS11) reveals intermediates in the synthesis of methyl-branched alkylpyrones. *J. Biol. Chem.* **228**, 16484–16494 (2013).
19. Sankaranarayanan, R. *et al.* A novel tunnel in mycobacterial type III polyketide synthase reveals the structural basis for generating diverse metabolites. *Nat. Struct. Mol. Biol.* **11**, 894–900 (2004).
20. Satou, R. *et al.* Structural basis for cyclization specificity of two *Azotobacter* type III polyketide synthases. A single amino acid substitution reverses their cyclization specificity. *J. Biol. Chem.* **288**, 34146–34157 (2013).
21. Zha, W., Rubin-Pitel, S. B. & Zhao, H. Characterization of the substrate specificity of PhlD, a type III polyketide synthase from *Pseudomonas fluorescens*. *J. Biol. Chem.* **281**, 32036–32047 (2006).
22. Gross, F. *et al.* Bacterial type III polyketide synthases: Phylogenetic analysis and potential for the production of novel secondary metabolites by heterologous expression in pseudomonads. *Arch. Microbiol.* **185**, 28–38 (2006).
23. Sone, Y. *et al.* Identification and characterization of bacterial enzymes catalyzing the synthesis of 1,8-dihydroxynaphthalene, a key precursor of dihydroxynaphthalene melanin, from *Sorangium cellulosum*. *Appl. Environ. Microbiol.*, AEM.00258-18 (2018).
24. Mascotti, M. L., Juri Ayub, M., Furnham, N., Thornton, J. M. & Laskowski, R. A. Chopping and Changing: the Evolution of the Flavin-dependent Monooxygenases. *J. Mol. Biol.* **428**, 3131–3146 (2016).
25. Koskiniemi, H. *et al.* Crystal structures of two aromatic hydroxylases involved in the early tailoring steps of angucycline biosynthesis. *J. Mol. Biol.* **372**, 633–648 (2007).
26. Funabashi, M., Funa, N. & Horinouchi, S. Phenolic lipids synthesized by type III polyketide synthase confer penicillin resistance on *Streptomyces griseus*. *J. Biol. Chem.* **283**, 13983–13991 (2008).
27. Mallika, V., Sivakumar, K. C. & Soniya, E. V. Evolutionary Implications and Physicochemical Analyses of Selected Proteins of Type III Polyketide Synthase Family. *Evol. Bioinform.* **7**, 41–53 (2011).

28. Funa, N., Funabashi, M., Yoshimura, E. & Horinouchi, S. A novel quinone-forming monooxygenase family involved in modification of aromatic polyketides. *J. Biol. Chem.* **280**, 14514–14523 (2005).
29. Funa, N., Funabashi, M., Ohnishi, Y. & Horinouchi, S. Biosynthesis of hexahydroxyperylenequinone melanin via oxidative aryl coupling by cytochrome P-450 in *Streptomyces griseus*. *J. Bacteriol.* **187**, 8149–8155 (2005).
30. Funa, N., Ohnishi, Y., Ebizuka, Y. & Horinouchi, S. Properties and substrate specificity of RppA, a chalcone synthase-related polyketide synthase in *Streptomyces griseus*. *J. Biol. Chem.* **277**, 4628–4635 (2002).
31. Vetting, M. W. *et al.* Pentapeptide repeat proteins. *Biochemistry* **45**, 1–10 (2006).
32. Collin, F., Karkare, S. & Maxwell, A. Exploiting bacterial DNA gyrase as a drug target. Current state and perspectives. *Appl. Microbiol. Biotechnol.* **92**, 479–497 (2011).
33. Shah, S. & Heddle, J. G. Squaring up to DNA. Pentapeptide repeat proteins and DNA mimicry. *Appl. Biochem. Biotechnol.* **98**, 9545–9560 (2014).
34. Hegde, S. S. *et al.* A fluoroquinolone resistance protein from *Mycobacterium tuberculosis* that mimics DNA. *Science* **308**, 1480–1483 (2005).
35. Baumann, S. *et al.* Cystobactamids: myxobacterial topoisomerase inhibitors exhibiting potent antibacterial activity. *Angew. Chem. Int. Ed.* **53**, 14605–14609 (2014).
36. Cociancich, S. *et al.* The gyrase inhibitor albicidin consists of *p*-aminobenzoic acids and cyanoalanine. *Nat. Chem. Biol.* **11**, 195–197 (2015).
37. Panter, F., Krug, D., Baumann, S. & Müller, R. Self-resistance guided genome mining uncovers new topoisomerase inhibitors from myxobacteria. *Chem. Sci.* **9**, 4898–4908 (2018).
38. Hashimi, S. M., Wall, M. K., Smith, A. B., Maxwell, A. & Birch, R. G. The Phytotoxin Albicidin is a Novel Inhibitor of DNA Gyrase. *Antimicrob. Agents Chemother.* **51**, 181–187 (2007).
39. Vetting, M. W., Hegde, S. S., Zhang, Y. & Blanchard, J. S. Pentapeptide-repeat proteins that act as topoisomerase poison resistance factors have a common dimer interface. *Acta Crystallogr. Sect. F* **67**, 296–302 (2011).
40. Kelley, L. A., Mezulis, S., Yates, C. M., Wass, M. N. & Sternberg, M. J. E. The Phyre2 web portal for protein modeling, prediction and analysis. *Nat. Protoc.* **10**, 845–858 (2015).
41. Ohnishi, Y. *et al.* Genome sequence of the streptomycin-producing microorganism *Streptomyces griseus* IFO 13350. *J. Bacteriol.* **190**, 4050–4060 (2008).
42. Iniesta, A. A., García-Heras, F., Abellón-Ruiz, J., Gallego-García, A. & Elías-Arnanz, M. Two systems for conditional gene expression in *Myxococcus xanthus* inducible by isopropyl- $\beta$ -D-thiogalactopyranoside or vanillate. *J. Bacteriol.* **194**, 5875–5885 (2012).

43. Wang, H. *et al.* RecET direct cloning and Red $\alpha\beta$  recombineering of biosynthetic gene clusters, large operons or single genes for heterologous expression. *Nat. Protoc.* **11**, 1175–1190 (2016).
44. Edayathumangalam, R. *et al.* Crystal structure of *Bacillus subtilis* GabR, an autorepressor and transcriptional activator of *gabT*. *Proc. Natl. Acad. Sci. USA* **110**, 17820–17825 (2013).
45. Huo, L., Rachid, S., Stadler, M., Wenzel, S. C. & Müller, R. Synthetic biotechnology to study and engineer ribosomal bottromycin biosynthesis. *Chem. Biol.* **19**, 1278–1287 (2012).
46. Bode, H. B. *et al.* Identification of additional players in the alternative biosynthesis pathway to isovaleryl-CoA in the myxobacterium *Myxococcus xanthus*. *ChemBioChem* **10**, 128–140 (2009).
47. Taylor, J. A., Burton, N. P. & Maxwell, A. High-throughput microtitre plate-based assay for DNA topoisomerases. *Methods Mol. Biol.* **815**, 229–239 (2012).
48. Edgar, R. C. MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *Bmc Bioinform.* **5**, 113 (2004).
49. Edgar, R. C. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **32**, 1792–1797 (2004).
50. Finn, R. D. *et al.* The Pfam protein families database: towards a more sustainable future. *Nucleic Acids Res.* **44**, D279–285 (2016).
51. Sambrook, J. & Russell, D. W. *Molecular cloning: A laboratory manual* (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 2001).
52. Zhang, Y., Buchholz, F., Muyrers, J. P. & Stewart, F. A. A new logic for DNA engineering using recombination in *Escherichia coli*. *Nat. Genet.* **20**, 123–128 (1998).
53. Kashefi, K. & Hartzell, P. L. Genetic suppression and phenotypic masking of a *Myxococcus xanthus* *frzF*- defect. *Mol. Microbiol.* **15**, 483–494 (1995).
54. Sucipto, H., Pogorevc, D., Luxenburger, E., Wenzel, S. C. & Müller, R. Heterologous production of myxobacterial  $\alpha$ -pyrone antibiotics in *Myxococcus xanthus*. *Metab. Eng.* **44**, 160–170 (2017).



# Chapter 6

## Supporting Information

### Genome mining reveals uncommon alkylpyrones as type III PKS products from myxobacteria

**Joachim J. Hug<sup>1,2,3</sup>, Fabian Panter<sup>1,2,3</sup>, Daniel Krug<sup>1,2,3</sup> and Rolf Müller<sup>1,2,3\*</sup>**

<sup>1</sup>Department Microbial Natural Products, Helmholtz-Institute for Pharmaceutical Research Saarland (HIPS), Helmholtz Centre for Infection Research (HZI), Campus E8 1, 66123 Saarbrücken, Germany

<sup>2</sup>Department of Pharmacy, Saarland University, Campus E8 1, 66123 Saarbrücken, Germany

<sup>3</sup>German Center for Infection Research (DZIF), Partner Site Hannover-Braunschweig, Germany

Special Issue: Natural Product Discovery and Development in the Genomic Era. Dedicated to Professor Satoshi Ōmura for his numerous contributions to the field of natural products.

\* Correspondence: [rolf.mueller@helmholtz-hips.de](mailto:rolf.mueller@helmholtz-hips.de), Tel.: +4968198806-3000

## Table of Contents

1	General materials and methods .....	365
1.1	Applied software, sequence analysis and bioinformatics methods .....	365
1.2	Molecular cloning, construction of plasmids, maintenance of bacterial cultures .....	368
	Approaches to establish reliable mutagenesis protocol for the myxobacterium <i>Cystobacterineae</i> strain MCy9487 (SBMx152) .....	374
	Heterologous expression of type III polyketide synthase gene cluster from MCy9487 via RecET direct cloning .....	375
1.3	Compound isolation .....	378
	Analysis during purification.....	378
	Liquid/liquid extraction.....	379
	Flash chromatography.....	379
1.4	Bioactivity profiling .....	380
	Cell based biological assays to evaluate MIC values on bacteria and cancer cell lines.....	380
2	Results.....	381
2.1	Bioinformatic investigation .....	381
2.2	Genes encoded in the ftp-like biosynthetic gene cluster from MCy9487 .....	384
2.3	Investigation MCy9487 crude extract and secondary metabolism <i>M. xanthus</i> DK1622_vanillate kanamycin mutant for <i>srs</i> operon compounds .....	386
2.4	NMR spectroscopic data .....	389
2.5	Topoisomerase inhibition assays of alkylpyrone-407.....	391
2.6	Copies of <sup>1</sup> H and <sup>13</sup> C NMR spectra of alkylpyrone-407 and alkylpyrone-393 .....	392
3	References.....	402

## 1 General materials and methods

### 1.1 Applied software, sequence analysis and bioinformatics methods

Biosynthetic gene cluster (BGC) sequences putatively encoding type III polyketide synthases (PKS) deriving from our in-house database have been deposited in GenBank and are accessible under the accession number as displayed in **Table S1**. The *ftp* BGC from *M. xanthus* DK 1622 was deposited in MiBiG under the accession number BGC0001831.

**Tab. S1** In-house genome sequence analysis of myxobacterial strains type III PKS genes

Name of myxobacterial strain	Reference	Accession number
<i>Aetherobacter fasciculatus</i> strain MSr9330 (SBSr002)	<sup>1</sup>	MH908877
<i>Aetherobacter rufus</i> strain MSr9331 (SBSr003)	<sup>1</sup>	MH908878
<i>Aetherobacter</i> strain MSr9329 (SBSr001)	<sup>1</sup>	MH908879
<i>Aetherobacter rufus</i> strain MSr9335 (SBSr008)	<sup>1</sup>	MH908880
<i>Aggregicoccus edonensis</i> strain MCy10622	<sup>2</sup>	MH908881
<i>Archangium gephyra</i> Mcy8375 (Ar8082)_No.1	<sup>3</sup>	MH908909
<i>Archangium gephyra</i> Mcy8375 (Ar8082)_No.2	<sup>3</sup>	MH908875
<i>Archangium gephyra</i> Mcy8375 (Ar8082)_No.3	<sup>3</sup>	MH908909
<i>Archangium</i> sp. Mcy8383 (Ar3548)_No.1	<sup>4</sup>	MH908911
<i>Archangium</i> sp. Mcy8383 (Ar3548)_No.2	<sup>4</sup>	MH908911
<i>Archangium</i> sp. Mcy8383 (Ar3548)_No.3	<sup>4</sup>	MH908876
<i>Byssovorax cruenta</i> strain MSr4204 (Byc1)	<sup>5</sup>	MH908916
<i>Chondromyces</i> sp. MSr9030	<sup>6</sup>	MH908917
<i>Corallococcus coralloides</i> strain MCy6431 ( <i>Ccc1071</i> )	<sup>7</sup>	MH908913
<i>Corallococcus coralloides</i> strain MCy8332 ( <i>Ccc127</i> )	<sup>8</sup>	MH908874
<i>Cystobacter</i> MCy9101 (SBCb004)_No.1	<sup>9</sup>	MH908910
<i>Cystobacter</i> MCy9101 (SBCb004)_No.2	<sup>9</sup>	MH908868
<i>Cystobacter</i> MCy9101 (SBCb004)_No.3	<sup>9</sup>	MH908873
<i>Cystobacter</i> sp. MCy9104_No.1	<sup>10</sup>	MH908897
<i>Cystobacter</i> sp. MCy9104_No.2	<sup>10</sup>	MH908903
<i>Cystobacterineae</i> strain MCy10636 (SBCy018)	<sup>11</sup>	MH908884
<i>Cystobacterineae</i> strain MCy9487(SBMx152)	This work	MH908886
<i>Cystobacter ferrugineus</i> strain MCy8071 (Cbfe23)_No.1	<sup>12</sup>	MH908905
<i>Cystobacter ferrugineus</i> strain MCy8071 (Cbfe23)_No.2	<sup>12</sup>	MH908900

<i>Cystobacter ferrugineus</i> strain MCy8071 (Cbfe23)_No.3	12	MH908900
<i>Cystobacter fuscus</i> MCy9118_No.1	13	MH908908
<i>Cystobacter fuscus</i> MCy9118_No.2	13	MH908899
<i>Cystobacter fuscus</i> MCy9118_No.2		MH908899
<i>Cystobacter fuscus</i> strain DSM 52655_No.1	14	MH908898
<i>Cystobacter fuscus</i> strain DSM 52655_No.2	14	MH908904
<i>Cystobacter fuscus</i> strain DSM 52655_No.3	14	MH908898
<i>Cystobacter velatus</i> MCy7152 (Cbv34)_No.1 (DSM: 14718)	15	MH908901
<i>Cystobacter velatus</i> MCy7152 (Cbv34)_No.2 (DSM: 14718)	15	MH908901
<i>Cystobacter velatus</i> MCy7152 (Cbv34)_No.3 (DSM: 14718)	15	MH908906
<i>Archangium violaceus</i> MCy8337 Cb vi35_No.1	16	MH908907
<i>Archangium violaceus</i> MCy8337 Cb vi35_No.2	16	MH908907
<i>Hyalangium minutum</i> MCy4189 (Hym3)	17	MH908871
<i>Hyalangium minutum</i> MCy2730 (NOCb10)	18	MH908872
<i>Jahnella thaxteri</i> MSr9139 (SBSr007)	19	MH908889
<i>Kofleria flava</i> MNa2518 (DSM 14601)	20	MH908912
<i>Melittangium boletus</i> Me b2 MCy7899 (DSM: 52889)_No.1	21	MH908896
<i>Melittangium boletus</i> Me b2 MCy7899 (DSM: 52889)_No.2	21	MH908896
<i>Melittangium lichenicola</i> strain MCy9148 (SBMe006)	22	MH908902
<i>Myxococcus fulvus</i> MCy8288 (Mxf50)_No.1	23	MH908885
<i>Myxococcus fulvus</i> MCy8288 (Mxf50)_No.2	23	MH908890
<i>Myxococcus fulvus</i> MCy10608 (Mxf188)	24	MH908923
<i>Nannocystis exedens</i> strain MNa3061 (Nae487)	21	MH908893
<i>Nannocystis exedens</i> strain MNa3063 (Nae485)	25	MH969426
<i>Nannocystis pusilla</i> strain MNa6508 (Ari7)	26	MH908866
<i>Phaselicystis flava</i> MSr9315 (SBKo001) (DSM: 21295)	27	MH908922
<i>Pseudenygromyxa salsuginis</i> MNa10638 (DSM 21377)_No.1	28	MH908895
<i>Pseudenygromyxa salsuginis</i> MNa10638 (DSM 21377)_No.2	28	MH908882
<i>Pyxidicoccus fallax</i> MCy8396 (And48)	29	MH908915
<i>Pyxidicoccus fallax</i> Mcy8408 (AndGT8)	30	MH908883

<i>Pyxidicoccus</i> strain MCy9557 (SBCy002)	<sup>31</sup>	MH908914
<i>Racemicystis crocea</i> MSr9521	<sup>32</sup>	MH908924
<i>Sandaracinus amylolyticus</i> MSr6788 (NOSO-4)	<sup>33</sup>	MH908864
<i>Sorangium cellulosum</i> MSr8404 (So ce38)_No.1	<sup>1</sup>	MH908870
<i>Sorangium cellulosum</i> MSr8404 (So ce38)_No.2	<sup>1</sup>	MH908888
<i>Sorangium cellulosum</i> So ce377 (DSM 53452)_No.1	<sup>7</sup>	MH908887
<i>Sorangium cellulosum</i> So ce377 (DSM 53452)_No.2	<sup>7</sup>	MH908894
<i>Sorangium cellulosum</i> MSr53 (So ce487) (DSM 53452)_No.1	<sup>1</sup>	MH908865
<i>Sorangium cellulosum</i> MSr53 (So ce487) (DSM 53452)_No.2	<sup>1</sup>	MH908863
<i>Sorangium cellulosum</i> So ce836	<sup>1</sup>	MH908891
<i>Sorangium cellulosum</i> So ce880	<sup>34</sup>	MH908867
<i>Sorangium cellulosum</i> So ce1128_No.1	<sup>34</sup>	MH908921
<i>Sorangium cellulosum</i> So ce1128_No.2	<sup>34</sup>	MH908920
<i>Sorangium cellulosum</i> So ce1525	<sup>1</sup>	MH908919
<i>Sorangium cellulosum</i> So ce1875 MSr1941 (DSM: 53600)	<sup>21</sup>	MH908918
<i>Sorangium cellulosum</i> So ceGT47	<sup>1</sup>	MH908892
<i>Stigmatella aurantiaca</i> Sg a15	<sup>35</sup>	MH908869

## 1.2 Molecular cloning, construction of plasmids, maintenance of bacterial cultures

**Tab. S2** List of oligonucleotides used in this study.

No.	Primer name	Primer sequence 5'–3'
1	MCy9487_akpBGC_large_Fw_left	GGAGGTCTACCGCACGTTTCGCGGAAGACTA CATGGCGATGCCGGTGTGACCGGGCGCAA GACGGAGTCGGAGAAGTTCGGATATCAGAT CCGAAAACCCCAAGTTACG
2	MCy9487_akpBGC_large_Rv_right	TGCCCCGGCCCCGGGCGCGCATGCGTGTAC AACGTCCCATCGATGATCTCGCCCACCACGC CCTCGGGCAGGGCCTCGATTTAATTAAGATA TCAGATCCTTTCTCCTCTTTAGATC
3	MCy9487_akpBGC_small_Fw_left	CCAGGACAACACCTGGACCGCCCCAAATGC CCGCGCACGAGACTACAAGTTGACCCGGCC TCCCCCCCCGCTAATGAATGATATCAGATC CGAAAACCCCAAGTTACG
4	MCy9487_akpBGC_smallRv_right_	AGCGCGCGCGAAGGCGCTGATGCTGACTCG GCCGTCCGGCACCTCCTGCGTGCCTTGACAG GTGGTGGGCTCGCGCCAATTTAATTAAGATA TCAGATCCTTTCTCCTCTTTAGATC
5	MCy9487_akpBGC_tetR_Fw	ATACTAGTTTATAACCTGAAGTCAGCCCCAT ACG
6	MCy9487_akpBGC_tetR_Rv	ATTTAATTAATTCTGCCAAGGGTTGGTTTG
7	MCy9487_akpBGC_Fw_tn5kan_outside	ATACTAGTGGTTTTATGGACAGCAAGCGAAC
8	MCy9487_akpBGC_Rv_tn5kan_outside	ATTTATAATCAGAAGAACTCGTCAAGAAGG C
9	MCy9487_akpBGC_Fw_tn5kan_inside	CCGTTTCGCCGGCTAAGCGTGGGGATTTCGATT TCCAGGGTGGGACCCTTAAGTGGACAGCAA GCGAACCGG
10	MCy9487_akpBGC_Rv_tn5kan_inside	CAACGAAGGCGACGGCGAGGGGTGCTGACG GGTCGTGCTGAGCATAATCTGTACCTCCTTA AGTCAGAAGAACTCGTCAAGA

11	MCy9487_akpBGC_verificationprimer_Fw_1	CGAGCAATCCGCTATTGGC
12	MCy9487_akpBGC_verificationprimer_Fw_3_kan_inside	ATCTGCGCTCTGCTGAAGC
13	MCy9487_akpBGC_verificationprimer_Rv_4	CTGTGTCCTTCTGCGACGC
14	MCy9487_akpBGC_large_verificationprimer_Rv_2	CGCATGCGTGTACAACGTC
15	MCy9487_akpBGC_small_verificationprimer_Rv_2	ACCTCCTGCGTGCCTTGAC
16	MCy9487_akp16KO_AmpR_Fw	CGGAAGCCATCCTTGCCCGGTAAGTGGTAGG GTCATGAAGACCACGTATACGACGAAAGGG CCTCGTGATAC
17	MCy9487_akp16KO_AmpR_Rv	CCACCTGTCAAGGCACGCAGGAGGTGCCGG ACGGCCGAGTCAGCAGAGACCTTACCAATG CTTAATCAGTGAG
18	MCy9487_akp1_tn5Kan_induction_Fw	GCGGTCCAGGTGTTGTCTGGAAAAGTCCCC GGAGGTGCATGGCGTTTAAAGCTTACATGGC GATAGCTAG
19	MCy9487_akp1_tn5Kan_induction_Rv	CACCGCCACCAGACCCACCAGGAGCCACAG CCACGAAAATCGCATACTCTTCCTTTTTTCAA TTCAG
20	MCy9487_akp4_tn5Kan_induction_Fw	ACTACCGATGGGCTCGCACCTATCGCCAGGA AGGACCCTGTTCTTTTAAAGCTTACATGGC GATAGCTAG
21	MCy9487_akp4_tn5Kan_induction_Rv	GAGCCGGCGAAGCAGCGCGGTGGACGCGTT TTCTTCCGTGGTCACACTCTTCCTTTTTCAAT TCAG
22	MCy9487_akp6_tn5Kan_induction_Fw	CTCCACTCCGACGCGGGACCCACCCTACCGC CAGGAGGAAGCCCCTTTAAAGCTTACATGG

		CGATAGCTAG
23	MCy9487_akp6_tn5Kan_induction_Rv	GTTGTTGAACCACAGGAACGTGACGATCTTC TGGACGTCGGGCATACTCTTCCTTTTCAATT CAG
24	DK1622_ftpA_vanillate_induction_Fw	ATATCATATGCTCAGTGCCTCAGGCC
25	DK1622_ftpA_vanillate_induction_Rv	ATGAATTCTGAAGAGGGCATGCGTGGAG
26	DK1622_ftpA_vanillate_verificationprimer_Fw _1	GTCACCTCTCCACCGACGCG
27	DK1622_ftpA_vanillate_verificationprimer_Rv _2	GCAGCGAGTCAGTGAGCGAG
28	DK1622_ftpA_vanillate_verificationprimer_Fw _3	GGCTTCCGGAGTCTTCAGC
29	DK1622_ftpA_vanillate_verificationprimer_Rv _4	GCGACATAGTTGGGGTGGC
30	MCy9487_akpBGC_verificationprimer_Fw_3_ kan_outside	CGCTTCCTCGTGCTTTACG

**Tab. S3.** List of plasmids used in this study.

No.	Plasmid name/ characteristic	Size [kb]	Function	Reference
1	pBR322-amp-tetReg-tetO-hyg-ccdB	4.454	Used as template to PCR-amplify capture vectors with homology arms	<sup>36</sup>
2	pACYC184	4.245	Used as template to PCR amplify <i>tetR</i> gene	<sup>37</sup>
3	pCR2.1-TOPO (EcoRI, religated)	3.931	Used as template to PCR-amplify <i>tn5_kanR</i> gene	TOPO®TA Cloning® Kit Thermo Fisher Scientific
4	pUC18	2.686	Used as template to PCR-amplify <i>ampR</i> gene	<sup>38</sup>



5	pFPV <sub>an_pcyA</sub>	6.181	pCR 2.1 topo derivative used as vector to ligate PCR product No.10 for subsequent vanillate-induced gene overexpression in DK1622	<sup>29</sup>
6	pMycoMar	---	Used to test mutagenesis of MCy9487	<sup>39</sup>

**Tab. S4** List of PCR-amplified constructs.

No.	PCR product name/ characteristic	Size [kb]	Template	Primer used
1	pBR322_capture_large_fragment (pBR322-amp-tetReg-tetO)	3.027	pBR322-amp-tetReg-tetO-hyg-ccdB	primer No.1 primer No.2
2	pBR322_capture_small_fragment (pBR322-amp-tetReg-tetO)	3.027	pBR322-amp-tetReg-tetO-hyg-ccdB	primer No.3 primer No.4
3	<i>tetR</i> _gene	1.436	pACYC184	primer No.5 primer No.6
4	kanR_gene_restriction_ligation_outside	0.971	pCR2.1-TOPO	primer No.7 primer No.8
5	kanR_gene_Red $\alpha\beta$ _recombineering_inside	1.062	pCR2.1-TOPO	primer No.9 primer No.10
6	ampR_gene_replace_akp16_ Red $\alpha\beta$ _recombineering_inside	1.163	pUC18	primer No.16 primer No.17
7	tn5kanR_gene_induced_gene_expression_akp1	1.095	pCR2.1-TOPO	primer No.18 primer No.19
8	tn5kanR_gene_induced_gene_expression_akp4	1.095	pCR2.1-TOPO	primer No.20 primer No.21
9	tn5kanR_gene_induced_gene_expression_akp6	1.095	pCR2.1-TOPO	primer No.22 primer No.23
10	DK1622_ <i>ftpA</i> _for_vanillate_induced_gene_expression	1.118	gDNA_ <i>M. xanthus</i> DK1622	primer No.24 primer No.25

## Chapter 6: Supporting Information

**Tab. S5** List of genetic constructs generated in this study.

No.	Plasmid name	Construction details/ characteristic
1	pBR322-amp-tetReg-tetO_MCy9487_akpBGC_large_fragment	Captured MCy9487_akpBGC_large_fragment from predigested MCy9487 gDNA using PCR product No.1 via RecET direct cloning
2	pBR322-amp-tetReg-tetO_MCy9487_akpBGC_small_fragment	Captured MCy9487_akpBGC_small_fragment from predigested MCy9487 gDNA using PCR product No.2 via RecET direct cloning
3	pBR322-tetR-tetReg-tetO_MCy9487_akpBGC_large_fragment	Construct obtained by restriction digestion of No.1; introduced PCR product No. 3 by conventional restriction ligation
4	pBR322-tetR-tetReg-tetO_MCy9487_akpBGC_small_fragment	Construct obtained by restriction digestion of No.2; introduced PCR product No. 3 by conventional restriction ligation
5	pBR322-tetR-tetReg-tetO_MCy9487_akpBGC_large_fragment_tn5kan_outside	Construct obtained by restriction digestion of No.3; introduced PCR product No. 4 by conventional restriction ligation
6	pBR322-tetR-tetReg-tetO_MCy9487_akpBGC_small_fragment_tn5kan_outside	Construct obtained by restriction digestion of No.4; introduced PCR product No. 4 by conventional restriction ligation
7	pBR322-tetR-tetReg-tetO_MCy9487_akpBGC_large_fragment_tn5kan_inside	Construct obtained by Red $\alpha\beta$ recombineering of construct No.3 and PCR product No. 5
8	pBR322-tetR-tetReg-tetO_MCy9487_akpBGC_small_fragment_tn5kan_inside	Construct obtained by Red $\alpha\beta$ recombineering of construct No.4 and PCR product No. 5
9	pBR322-tetR-tetReg-tetO_MCy9487_akpBGC_large_fragment_tn5kan_outside_ampR_Δakp16	Construct obtained by Red $\alpha\beta$ recombineering of construct No.5 and PCR product No. 6
10	pBR322-tetR-tetReg-tetO_MCy9487_akpBGC_small_fragment_tn5kan_outside_ampR_Δakp16	Construct obtained by Red $\alpha\beta$ recombineering of construct No.6 and PCR product No. 6
11	pBR322-tetR-tetReg-tetO_MCy9487_akpBGC_large_fragment_tn5kan_inside_ampR_Δakp16	Construct obtained by Red $\alpha\beta$ recombineering of construct No.7 and PCR product No. 6
12	pBR322-tetR-tetReg-tetO_MCy9487_akpBGC_small_fragment_tn5kan_inside_ampR_Δakp16	Construct obtained by Red $\alpha\beta$ recombineering of construct No.8 and PCR product No. 6
13	pBR322-tetR-tetReg-	Construct obtained by Red $\alpha\beta$ recombineering of

	tetO_MCy9487_akpBGC_large_fragment_tn5_kanR_akp1	construct No.3 and PCR product No. 7
14	pBR322-tetR-tetReg-tetO_MCy9487_akpBGC_small_fragment_tn5_kanR_akp1	Construct obtained by Red $\alpha$ $\beta$ recombineering of construct No.4 and PCR product No. 7
15	pBR322-tetR-tetReg-tetO_MCy9487_akpBGC_large_fragment_tn5_kanR_akp4	Construct obtained by Red $\alpha$ $\beta$ recombineering of construct No.3 and PCR product No. 8
16	pBR322-tetR-tetReg-tetO_MCy9487_akpBGC_small_fragment_tn5_kanR_akp4	Construct obtained by Red $\alpha$ $\beta$ recombineering of construct No.4 and PCR product No. 8
17	pBR322-tetR-tetReg-tetO_MCy9487_akpBGC_large_fragment_tn5_kanR_akp6	Construct obtained by Red $\alpha$ $\beta$ recombineering of construct No.3 and PCR product No. 9
18	pBR322-tetR-tetReg-tetO_MCy9487_akpBGC_small_fragment_tn5_kanR_akp6	Construct obtained by Red $\alpha$ $\beta$ recombineering of construct No.4 and PCR product No. 9
19	pFPVan_ DK1622_ftpA	Construct obtained by restriction digestion of plasmid No.6; introduced PCR product No. 10 by conventional restriction ligation

### Approaches to establish reliable mutagenesis protocol for the myxobacterium *Cystobacterineae* strain MCy9487 (SBMx152)

Reliable mutagenesis via transformation can be achieved in myxobacteria by electroporation exemplified by *Myxococcus xanthus* DK1622<sup>40</sup>, via biparental conjugation procedures as shown for *Chondromyces crocatus* Cm c5<sup>41</sup> or via triparental mating e.g. established in *Sorangium cellulosum* So ce56<sup>42</sup>. Electrocompetence and triparental conjugation of the *Cystobacterineae* strain SBMx152 (MCy9487) was evaluated with the plasmid pMycoMar<sup>39</sup>, conferring kanamycin resistance by the defective transposon magellan-4<sup>43</sup>. MCy9487 was routinely cultivated at 30 °C in RGAE2 medium (0.35% Bacto casitone (BD), 0.1% Bacto soytone (BD), 0.05% Bacto yeast extract (BD), 0.2% glucose, 0.2% soluble starch, 0.1% sucrose, 0.1% maltose x H<sub>2</sub>O, 0.2% cellobiose, 0.05% CaCl<sub>2</sub> x 2H<sub>2</sub>O, 0.1% MgSO<sub>4</sub> x 7H<sub>2</sub>O, 0.238% HEPES( 2-[4-(2-hydroxyethyl)piperazin-1-yl]ethanesulfonic acid), adjusted with KOH to pH 7.0). MCy9487 features resistance during liquid fermentation and cultivation on agar plate against hygromycin (up to 150 µg/mL) and ampicillin (up to 100 µg/mL), while the strain is sensitive to kanamycin (50 µg/mL), tetracycline (5 µg/mL) and oxytetracycline (5 µg/mL).

### Procedure testing electrocompetence of MCy9487

For preparation of MCy9487 electrocompetent cells, 20 mL of RGAE2 medium in a 300 mL shake flask was inoculated with a suitable colony from a RGAE2 agar plate and incubated at 30 °C at 200 rpm for 3–5 days. As soon as the OD<sub>600</sub> reached 0.8–0.9, 1–3 mL of this culture was reinoculated in 50 mL RGAE2 medium in a 300 mL shake flask and incubated 3–5 days at 30 °C, 200 rpm. Cultures with OD<sub>600</sub> 0.6–0.9 were used and subsequently divided into aliquots à 2 mL (Eppendorf tubes) and centrifuged for 2 min, 7000 rpm at room temperature (VWR centrifuge ECN521-3601, Hitachi Koki Co., Ltd). After centrifugation, the supernatant was discarded and the cell pellet was resuspended in 1 mL autoclaved MQ water (obtained through Milli-Q® Reference Water Purification System, Merck), followed by centrifugation for 2 min at 7000 rpm. The supernatant was discarded; the second wash step was performed with 800 µL MQ water, followed by centrifugation for 2 min at 7000 rpm. The cell pellet was resuspended in 50 µL MQ water by pipetting up and down and was directly used for electroporation. 4–6 µL of pMycoMar vector solution (100–200 ng/µL) was added to the resuspended electrocompetent cells and transferred into an electroporation cuvette (volume 1000 µL; electrode distance 1 mm).

The electroporation was performed under different conditions 200–1200 V/cm, 25 µF, and 200–800 Ω (Epurator V1.01, Eppendorf). Afterwards the cells were resuspended in 1 mL RGAE2 medium and incubated at 30 °C, 950 rpm (HLC Heating-Thermomixer, DitaBis) for 4 h, 6 h, 8 h, 10 h or 12 h. Afterwards 1 mL of this culture was mixed with 3 mL of RGAE2 soft agar (7.5% agar) containing kanamycin for the selection of transformants. Subsequently the mixture was quickly poured on RGAE2 agar plates containing the same antibiotic concentration. The plates were incubated at 30 °C for 10–14 days. Applying this procedure, no transformants could be observed.

### Investigation of mutagenesis via triparental conjugation

Overnight cultures of *E. coli* ET12567 (pMycoMar) and *E. coli* HB101 (pRK600), grown in LB media with appropriate selection marker (pMycoMar; 50 µg/mL kanamycin, pRK600; 30 µg/mL chloramphenicol) were used to inoculate 25 mL of LB media. Cells were harvested at OD<sub>600</sub> 0.4–1.0, washed three times (centrifugation for 2 min, 7000 rpm at 4 °C, ) with LB media and adjusted to a cell density of approximately 10<sup>10</sup> cells/mL. Cells were resuspended in 1 mL of LB media and stored at 4 °C before conducting conjugation. MCy9487 was grown in 50 mL of RGAE2 for 3–5 days, by avoiding the appearance of cell aggregates. When the culture reached an OD<sub>600</sub> 0.4–1.3, 1 or 10 mL were harvested by centrifugation and subsequently washed three times with RGAE2. Cells were resuspended in 1 mL of RGAE2 media and stored at room temperature before conducting conjugation. Then, 100 µL of *E. coli* HB101 (RK600), *E. coli* ET12567 (pMycoMar) and MCy9487 were carefully mixed and spotted onto RGAE2-medium agar plates without antibiotic selection. The cells were conjugated for either 1 h, 24 h or 36 h at 37 °C. The cell mixture was scraped from the agar after conjugation and resuspended in 1 mL of RGAE2 liquid medium and plated as 500 µL aliquots onto RGAE2 agar plates containing 50 µg/mL kanamycin (clonal selection of MCy9487 conjugants with genomic integration of pMycoMar) and 120 µg/mL tobramycin (counter-selection to avoid growth of *E. coli*). The occurrence of transconjugants could not be detected within 14 days of incubation at 30 °C.

### Heterologous expression of type III polyketide synthase gene cluster from MCy9487 via RecET direct cloning

#### Genomic DNA isolation procedure

For the successfully performed RecET direct cloning strategy, high quality genomic DNA is required. Genomic DNA was prepared from lysed cells of MCy9487 via phenol-chloroform extraction and ethanol (EtOH) precipitation according to established procedures <sup>36</sup>. Therefore, MCy9487 wild type was cultivated at 30 °C in RGAE2 medium. For the cultivation of liquid cultures, the strain was grown in Erlenmeyer flasks on an orbital shaker at 180 rpm for 3–5 days. 100 mL liquid culture of MCy9487 was centrifuged for 10 min at 8000 rpm (Eppendorf Centrifuge 5804R). The cell pellet was resuspended in 15 mL SET buffer (75 mM NaCl, 25 mM EDTA pH 8.0, 20 mM Tris-HCl pH 7.5) and centrifuged for 10 min at 8000 rpm (Eppendorf Centrifuge 5804R). Afterwards the cell pellet was resuspended in 5 mL SET buffer and mixed with 50 µL of RNase A stock solution (10 mg/mL). After mixing 300 µL proteinase K solution ((10 mg/mL Proteinase K, 50 mM Tris-HCl pH 8.0, 1 mM CaCl<sub>2</sub> (final concentration: 0.5 mg/mL) and 600 µL 10% SDS solution (final concentration: 1% (w/v)) was added to the mixture. The mixture was incubated for 2 h at 55 °C in a water bath with occasional inversion. Afterwards 6 mL phenol-chloroform-isoamylalcohol mixture (25:24:1) was added and inverted for 1 h.

Subsequently the mixture was centrifuged for 5 min at 8000 rpm at room temperature. The upper phase was transferred by using an end-cut 1 mL tip to prevent shearing of genomic DNA. Afterwards 6 mL phenol-chloroform-isoamyl alcohol mixture (25:24:1) was added and inverted for 1 h. The procedure of adding, inversion and centrifugation of 6 mL phenol-chloroform-isoamylalcohol mixture (25:24:1) was repeated as often as necessary, usually three times. 5 mL of chloroform-isoamylalcohol (24:1) was added to the combined upper phase and inverted for 1 h. Afterwards the mixture was centrifuged for 10 min at 8000 rpm. The supernatant was transferred using an end-cut 1 mL tip to prevent shearing of genomic DNA. 440 µL of 3 M sodium acetate (pH 7.5) was added and mixed by

inversion. Afterwards 11 mL of 100% ice-cold EtOH was added to the separated aqueous phase and inverted gently. The occurring genomic DNA was collected by rolling the DNA on the end of the Pasteur pipette, washed with 70% EtOH, dried by attaching the inner site of a fresh 2 mL tube to remove the EtOH drops and transferred to another Eppendorf tube. The genomic DNA was resuspended in 2 mL of MQ water and stored at 4 °C.

### Restriction digestion of genomic DNA

In order to release the DNA fragment containing roughly the “silent” biosynthetic gene cluster the obtained intact high quality genomic DNA was digested by appropriate restriction enzymes. The larger cluster fragment consisting of 28.126 kb (*orf7–orf23*) was released by the restriction enzymes MreI and EcoRV, the smaller cluster fragment consisting of 19.816 kb (*akp1–akp16*) by SspI. The following protocol was adapted as described previously<sup>36</sup>. Digestion of genomic DNA was performed in a volume of 400 µL, containing around 50 µg genomic DNA and the restriction endonucleases in the appropriate buffer according to manufacturer’s instructions. The reactions were allowed to proceed for at least 12 h at appropriate temperature, depending on the restriction endonuclease used. Reactions were stopped by addition of gel loading buffer. 400 µL of phenol-chloroform-isoamyl alcohol mixture (25:24:1) was added to the remaining digestion mixture: after mixing by inversion to create an emulsion, mixture was centrifuged for 10 min at room temperature at 9400 g. An end-cut wide-bore 1 mL tip was used to gently pipette 300 µL of the aqueous phase (top layer) into a new 1.5 mL microcentrifuge tube. 20 µL of 3 M sodium acetate (pH 7.5) was added and gently inverted to mix. 800 µL EtOH was added and gently inverted. Afterwards the mixture was centrifuged at 9400 g for 2 min at room temperature to pellet the digested genomic DNA. The supernatant was discarded; the tube was inverted on a paper towel to drain the liquid off the pellet. The pellet was washed carefully by adding 1 mL of 70% EtOH and centrifuged at 9400 g for 1 min at room temperature. The supernatant was discarded again afterwards the tube was inverted on a paper towel to drain the pellet. The pellet was dried on a thermomixer at 45 °C without shaking for 10 min until no liquid was left. The DNA pellet was resuspended in 12 µL of autoclaved MQ water and the NanoDrop UV spectrophotometer (Thermo Scientific™) was used to measure the concentration: the concentration should be around 2 µg/µL. The digested genomic DNA was stored at 4°C.

### RecET direct cloning protocol

Digested genomic DNA was co-electroporated with the homology arm-flanked linear PCR product (PCR product No.1 or No.2) into arabinose-induced *E. coli* GB05RedTrfA for RecET direct cloning. Overnight culture of *E. coli* GB05RedTrfA harboring pSC101-BAD-ETgA-tet was used to inoculate 1.4 mL LB medium in lid-punctured 2 mL Eppendorf tubes containing 6 µg/mL tetracycline and grown at 30 °C until OD<sub>600</sub> reached 0.35 and 0.40. Afterwards 35 µL of 10% L-arabinose was added to the culture and incubated at 37 °C for 40 min, to facilitate homologous recombination by induced gene expression. Electrocompetent cells were then prepared from the culture by washing the cells twice with autoclaved MQ water. For transformation, 5 µL (10 µg) ice-cold digested genomic DNA and 5 µL (1 µg) of backbone PCR product No.1 or No.2 with specific homology arms was added to aliquots of electrocompetent cells to perform direct cloning of digested genomic DNA. Mixture was transferred into 0.1 cm-wide electroporation cuvettes and subjected to electrical pulse at voltage of 1350 V, capacity of 25 µF, resistance of 600 Ω and 1 mm cuvette (Eporator V1.01 from Eppendorf). The pulsed cells were diluted in 1 mL LB and incubated at 37 °C for 1 h followed by plating on LB agar plates containing 100 µg/mL ampicillin and incubating at 37 °C overnight to avoid replication of the temperature sensitive pSC101-BAD-ETgA.

### Transfer and chromosomal integration of the expression constructs into the host *Myxococcus xanthus* DK1622

According to a previously established electroporation procedure for *Myxococcus xanthus* DK1622<sup>40</sup> the host strain *M. xanthus* DK1622  $\Delta mxchrA-tetR$ <sup>44</sup> was transformed with the generated expression constructs (**Tab. S5**, constructs 3–19). *M. xanthus* DK1622  $\Delta mxchrA-tetR$  is a mutant in which the myxochromide A gene cluster (*mxchrA*) has been deleted, since myxochromide A is one of the major compounds produced by *M. xanthus* DK1622, and its gene cluster deletion was expected to provide an increased precursor pool for production of other secondary metabolites like type III polyketides.

Therefore *mxchrA* was deleted from *M. xanthus* DK1622 and replaced by a *tetR* gene, which served as the target site for the chromosomal integration of the generated expression constructs (**Tab. S5**, constructs 3–18) via single crossover recombination. *M. xanthus* DK1622 mutants were routinely cultivated at 30 °C in CTT medium or CTT agar (1% casitone, 10 mM Tris buffer pH 7.6, 1 mM KH<sub>2</sub>PO<sub>4</sub> pH 7.6, 8 mM MgSO<sub>4</sub> (1.5% agar) pH adjusted to 7.6). Liquid cultures were grown in Erlenmeyer flasks on an orbital shaker at 180 rpm for 3–6 days.

*M. xanthus* transformants were selected by adding 50 µg/µL kanamycin to the fermentation culture. Correct chromosomal integration of the expression constructs via homologous recombination into the *tetR* locus was confirmed by PCR. Genomic DNA of the transformants were isolated using the Gentra® Puregene® Yeast/Bact. Genomic DNA Purification Kit (Qiagen) according to manufacturer's instructions. For each expression construct, correct chromosomal integration was confirmed using two different primer combinations revealing PCR products of the expected sizes:

Constructs No.3, 7, 11, 13, 15 and 17 primer No.11/14 (1484 bp), and primer No.12/13 (1682 bp)

Constructs No.5 and 7 primer No.11/14 (1484 bp), and primer No.30/13 (1547 bp)

Constructs No.4, 6, 8, 10, 12, 14, 16 and 18 primer No.11/15 (1541 bp), and primer No.12/13 (1682 bp)

Constructs No.6 and 8 primer No.11/15 (1541 bp), and primer No.12/13 (1547 bp)

Genomic DNA of *M. xanthus* DK1622  $\Delta mxchrA-tetR$  was used as negative control. A complementary experiment using primer No.11/13 revealed a 1461 bp PCR product for *M. xanthus* DK1622  $\Delta mxchrA-tetR$ , but not for any of the transformants of *M. xanthus* DK1622  $\Delta mxchrA-tetR$  harboring one of the generated constructs. PCRs were performed according to the settings described above.

### Crude extracts preparation for analysis of secondary metabolism

*Myxococcus xanthus* DK 1622  $\Delta$ *mxchrA-tetR* and mutant strains were grown on CTT agar plates for 3–5 days at 30 °C. For the preparation of liquid pre-cultures, a suitable portion of overgrown agar was transferred to inoculate 20 mL of CTT medium in an Erlenmeyer flask (300 ml) and incubated at 30 °C and 200 rpm for 2–4 days. 1 mL of the preculture was used to inoculate 50 mL of CTT medium containing appropriate antibiotic selection and 2% amberlite resin XAD-16 (Sigma). In order to obtain statistically significant results, three independent transformants of each recombinant strain were selected and fermentations were performed at least in triplicates. The resulting fermentation broth were harvested by centrifugation at 8000 rpm for 10 min (Eppendorf centrifuge 5810 R). The supernatant was discarded, whereas the residual material was extracted first with 25 mL MeOH, stirred for 1 h, filtered through filter paper (folded filters grade: 3hw from Sartorius) into a round bottom flask and afterwards this procedure was repeated with 25 mL acetone. The solvent of the filtered extracts was removed under vacuum (BÜCHI Rotavapor R-210) and the extracts dissolved in 2 mL MeOH and stored at -20 °C. The redissolved extract were diluted with MeOH (1:3 (extract/MeOH (v:v))) centrifuged at 13000 g for 10 min (VWR centrifuge ECN521-3601, Hitachi Koki Co., Ltd) and 1 µL of the supernatant was subjected to HPLC-MS analysis as described further below.

### 1.3 Compound isolation

#### Analysis during purification

All measurements to analyze the mass of alkylpyrone-393 and alkylpyrone-407 during purification were performed on a Dionex Ultimate 3000 RSLC system coupled to the amaZon iontrap MS using a BEH C18, 100 x 2.1mm, 1.7 µm dp column equipped with a C18 precolumn (Waters). Samples of 1 µL were separated by a gradient from (A) H<sub>2</sub>O + 0.1% formic acid (FA) to (B) ACN + 0.1% FA at a flow rate of 0.6 mL/min and 45 °C. The gradient was initiated by a 0.5 min isocratic step at 5% B, followed by linear increase to 95% B in 18 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 200–600 nm. The detection was performed in the positive ESI MS/MS mode.



### Liquid/liquid extraction

The crude extract was prepared by mixing the cell pellet of 10 L fermentation culture with 2% XAD-16 with 200 mL (MeOH) and 200 mL acetone, stirring for 1 h and filtering through glass wool into a round bottom flask. This extraction was repeated once; afterwards the extraction fraction was dried using the rotary evaporator (BÜCHI Rotavapor R-210). The dried residue was redissolved in 400 mL MeOH and 300 mL hexane and transferred to a separation funnel. After mixing thoroughly, precipitates were collected (precip.1) and both phases were separated. 300 mL hexane was added to the MeOH phase to perform a second liquid-liquid extraction. Both phases were dried by rotary evaporation. The hexane extract residue was redissolved in 6 mL MeOH and stored at -20°C. The MeOH extract residue was redissolved in 400 mL MQ water and 300 mL chloroform (CHCl<sub>3</sub>) to repeat liquid-liquid extraction by adding 300 mL CHCl<sub>3</sub> to the water phase after the first separation. To clear the turbid phases, HCl was added until the phases could be distinguished. Precipitates were collected (precip. 2) and both phases were separated; the CHCl<sub>3</sub> phase was dried by rotary evaporation and redissolved in ethyl acetate (EE) to dissolve the same spectrum of metabolites as CHCl<sub>3</sub>. The water phase was mixed twice with 300 mL EE to perform liquid-liquid extraction as described before. The precipitates were collected again (Precip.3) and the EE phase was dried by rotary evaporation and redissolved in 8 mL MeOH. Collected precipitates were partly redissolved and subjected to HPLC-MS analysis as described further above.

### Flash chromatography

Further fractionation was performed on the flash chromatography system Biotage Isolera One. The Biotage system was used with a 50 g SNAP cartridge KP Sil column using a solvent system consisting of CHCl<sub>3</sub> (**A**), EE (**B**) and MeOH (**C**). A flow rate of 50 mL/min was used, and 45 aliquots à 30 mL were collected with testing tubes. The column was equilibrated with three column volumes (CV) of 100% solvent **A**. The following gradient was used: 2 x CV 100% solvent **A**; 8 x CV linear increase from 0–100% Solvent **B**; 2 x CV 100% Solvent **B**; 5 x CV linear increase from 10–50% solvent **C**; 5 x CV 50% solvent **B** and 50% solvent **C**. After flash chromatography, the aliquots were analyzed for the target masses using a shorter template gradient program. Samples of 1 µL were separated by a gradient from (**A**) H<sub>2</sub>O + 0.1% FA to (**B**) ACN + 0.1% FA at a flow rate of 0.6 mL/min and 45 °C. The gradient was initiated by a 0.5 min isocratic step at 5% **B**, followed by linear increase to 95% **B** in 6 min. After a 1.5 min step at 95% **B** the system was re-equilibrated to the initial conditions (5% **B**).

## 1.4 Bioactivity profiling

### Cell based biological assays to evaluate MIC values on bacteria and cancer cell lines

#### Antimicrobial assay

Standard sterile microbiological techniques were maintained throughout. All microorganisms were handled according to standard procedures and were obtained from the German Collection of Microorganisms and Cell Cultures (Deutsche Sammlung für Mikroorganismen und Zellkulturen, DSMZ) or were part of our internal strain collection. Alkylpyrone-407 was tested in microbroth dilution assays on the following panel of microorganisms: *Escherichia coli* DSM-1116, *E. coli* (TolC-deficient), *Pseudomonas aeruginosa* PA14, *Bacillus subtilis* DSM-10, *Micrococcus luteus* DSM-1790, *Mycobacterium smegmatis* mc2-155, *Staphylococcus aureus* Newman, *Mucor hiemalis* DSM-2656, *Candida albicans* DSM-1665 and *Pichia anomala* DSM-6766. For microdilution assays, overnight cultures were prepared from cryopreserved cultures and were diluted to achieve a final inoculum of  $10^4$ – $10^5$  cfu/mL. Serial dilutions of compounds were prepared in sterile 96-well plates in the respective test medium. The cell suspension was added and microorganism were grown for 18–48 h at either 30 °C or 37 °C. Growth inhibition was evaluated by visual inspection and given minimum inhibitory concentration (MIC) values are the lowest concentration of antibiotic at which no visible growth was observed. No inhibition of one of the tested microorganisms was observed at concentration up to 64 µg/mL of alkylpyrone-407.

#### Cytotoxic activity

Cell lines were obtained from the German Collection of Microorganisms and Cell Cultures (Deutsche Sammlung für Mikroorganismen und Zellkulturen, DSMZ) or were part of our internal collection and were cultured under conditions recommended by the depositor. Cells (HCT-116 (human colon carcinoma cell line)) were seeded at  $6 \times 10^3$  cells per well of 96-well plates in 180 µL complete medium and treated with alkylpyrone-407 in serial dilution after 2 h equilibration. After 5 days incubation, 20 µL of 5 mg/mL MTT (thiazolyl blue tetrazolium bromide) in phosphate buffered saline (PBS) was added per well and it was further incubated for 2 h at 37°C. The medium was discarded and cells were washed with 100 µL PBS before adding 100 µL 2-propanol/10 N HCl (250:1) in order to dissolve formazan granules. The absorbance at 570 nm was measured using a microplate reader (Tecan Infinite M200Pro), and cell viability was expressed as percentage relative to the respective MeOH control. IC<sub>50</sub> values were determined by sigmoidal curve fitting.

## 2 Results

### 2.1 Bioinformatic investigation

The majority of myxobacterial genomes harbor one gene encoding a type III PKS, whereas 19 myxobacterial genomes harbor more than one gene encoding a type III PKS. Twelve myxobacterial strains harbor two type III PKS genes and seven myxobacterial strains even three type III PKS in their genomes. Noteworthy 23 type III polyketide synthases are co-localized with different putative biosynthetic gene clusters. Some examples in the literature describe that type III polyketides are providing building blocks for different secondary metabolites, e.g. the type I polyketide kendomycin from the producer *Streptomyces violaceoruber* (strain 3844-33C) <sup>45</sup> or the non-ribosomal peptide balhimycin <sup>46</sup>. Furthermore napyradiomycins as well as other meroterpenoids, such as naphterpin <sup>47</sup>, furaquinocins <sup>48</sup>, marinone <sup>49</sup> and neomarinone <sup>50</sup>, are biosynthesized from the type III polyketide building block 1,3,6,8-tetrahydroxynaphthalene (T<sub>4</sub>HN) and isoprenoid units deriving in most cases from the mevalonic acid biosynthetic pathway <sup>51</sup>. However, it seems to be less likely from a genetic point of view that most of the type III PKS genes are involved in the biosynthesis of their neighbored biosynthetic genes, since the BGCs are clearly separated from each other. It is more likely, that the automatic annotation of antibiotics and secondary metabolite analysis shell (antiSMASH) fused these closely located BGCs. Type III PKSs were considered as members of the same family, when they shared more than 60% identical sites in the amino acid alignment. The rationale of this cut-off was provided through phylogenetic analysis of the closely related but separated type III PKS family **IX** and **X** which are not sharing more than 60% identical site in their amino acid sequence. The second reason to fix the cut-off to 60% was the observation that the type III PKS family **XI**, consisting of three type III PKS genes, are sharing 60% identical sites in the amino acid alignment, unlike the excluded type III PKS gene from *Haliangium ochraceum* DSM14365 located in the same clade.

Family **II** of type III PKS genes (bright green color) contains myxobacterial strains of the suborder *Nannocystineae*. The type III PKS gene is probably not organized in a gene operon with further tailoring enzymes. Only around 0.6 kb downstream of the type III PKS gene, a gene is located encoding a GTPase. Upstream of the type III PKS from Ari7 a hypothetical protein and a transcriptional regulator is located: the type III PKS from Nae487 harbors a hypothetical protein upstream of the type III PKS, whereas Nae485 has a similar transcriptional regulator. The pentapeptide repeat protein (PRP) gene which is located around 6 kb from the type III PKS gene in Nae487, seems to have no self-resistance function, since the *in silico* protein homology analysis did propose as structural model the H7 effector protein NleL from *E. coli* O157:H7 (pdb: 3NB2) <sup>52</sup> and the usual length of the self-resistant conferring pentapeptide repeat protein (PRP) would exceed the usual size of known members <sup>53</sup>.

The third family (red colored strains **III**) comprises seven type III PKS genes, with basically two different architectures of the genetic operon, exemplified by the strains *Sorangium cellulosum* So ce56 and *Aetherobacter* strain MSr9335 (SBSr008). The strains *Sorangium cellulosum* So ce56, *Sorangium cellulosum* strain MSr7234 (So ce377), *Sorangium* strain MSr53 (So ce487) are sharing an overlapping operon consisting of a type III PKS, and methyltransferase and a dehydrogenase. The strains *Aetherobacter* strain MSr9335 (SBSr008), *Aetherobacter rufus* strain MSr9331 (SBSr003) *Aetherobacter fasciculatus* strain MSr9330 (SBSr002) and *Aetherobacter* strain MSr9329 (SBSr001) have a putative dimethylformamidase downstream of the type III PKS in common and further genes surrounding this operon are highly conserved. However, the meaning of those surrounding genes remains elusive. The phylogenetically closest neighbors of family **III** namely *Kofleria flava* strain

MNa2518, *Labilithrix luteola* strain DSM 27648 and *Pseudenhygromyxa salsuginis* strain MNa10638 (DSM21377) are quite inhomogeneous in their sequence similarity and genetic organization. Nevertheless, these type III PKS genes from the underinvestigated suborder of *Nannocystineae* present further putative genes encoding type III PKSs that might afford new type III polyketides.

The two families **IV** and **V** of type III PKS genes originate from myxobacterial strains, which harbor often two or three type III PKS genes. The type III PKS gene representing family **IV** (grey-green color) derives from the strain MCy9101 and the type III PKS gene representing family **V** (purple color) derives from the strain *Cystobacter fuscus* DSM 2262 (MCy9127). Family **IV** of type III PKSs harbors nine different type III PKS genes, divided in two subgroups according to their arrangement of tailoring enzymes. The first subgroup contains type III PKS genes from MCy9127, *Cystobacter velatus* strain MCy7152 (Cbv34), *Cystobacter ferrugineus* strain MCy8071 (Cbfe23), *Cystobacter* strain MCy9104, *Cystobacter fuscus* strain MCy9118 (SBCb021) and *Cystobacter fuscus* strain DSM 52655. These type III PKS BGCs harbor several tailoring enzymes organized in a putative genetic operon. In all members four genes are located upstream of the type III PKS, namely a glucose-6-phosphate-1-dehydrogenase, a hypothetical protein, a short-chain dehydrogenases/reductase (SDR) and a second smaller hypothetical protein. Additionally upstream of the type III PKS gene, a flavin-binding monooxygenase gene, a transcriptional regulator gene and a peptidase gene are located in all members of this family. However, the orientation of these genes and the insertion of a serine/threonine protein kinase gene between the type III PKS gene and flavin-binding monooxygenase gene in the genome of MCy7152, MCy8071, DSM2262 and DSM52655 makes it less likely that these genes are associated with the type III PKS BGC. Furthermore the genetic organization of the second subgroup consisting of MCy9101, *Archangium* strain MCy8383 (Ar3548) and *Archangium gephyra* strain MCy8375 (Ar8082), which solely share the short-chain dehydrogenases/reductase (SDR) the hypothetical protein and the hypothetical protein might indicate the operon structure of this type III PKS family.

Family **V** of type III PKSs harbors eleven different type III PKS genes, which are divided in several subgroups according to the sequence similarity of the type III PKS gene; however according to their arrangement of tailoring enzymes all of these eleven members are highly similar. Upstream of the type III PKS gene, an overlapping gene encoding a small hypothetical protein is located, followed by a gene encoding a methyl-accepting chemotaxis protein and a SDR gene whereas downstream an acyl-CoA synthetase gene and a gene encoding a metalloprotease-like protein are located. Some variation is given upstream of this conserved operon by a second operon consisting either of two genes encoding a methyltransferase and a transmembrane protein followed by a facultative FAD-dependent oxidoreductase gene. Noteworthy with the exception of the type III PKS operon from MCy9101 all members in this family harbor downstream of their BGC another type III PKS operon, which are phylogenetically clustered in family **X**.

Family **VI** of type III PKS genes (yellow color) comprises three type III PKSs, two of these derives from the suborder *Nannocystineae* (*Pseudenhygromyxa salsuginis* strain DSM21377 (MNa10638) and *Plesiocystis pacifica* strain SBPI001 (MNa9317) whereas the type III PKS from *Enhygromyxa salina* strain DSM 15201 is classified into the suborder *Sorangineae*. All three type III PKS genes are organized in a putative genetic operon harboring upstream of the gene encoding a type III PKS, genes encoding a small hypothetical protein, a nitroreductase and downstream of the type III PKS gene one gene encoding a hypothetical protein. Only in the genetic operons of MNa10638 and MNa9317 a second gene encoding a hypothetical protein is located.

Family **VII** (turquoise color) comprises three type III PKS genes originating from the suborder *Nannocystineae*. The type III PKS operon from *Phaselicystis flava* strain SBKo001 (MSr9315) is co-localized with a terpene and PKS-NRPS hybrid biosynthetic gene cluster. AntiSMASH analysis yielded as relevant hits for homologous known gene clusters the microsclerodermins biosynthetic gene cluster (21% of genes show similarity)<sup>54</sup> characterizing the PKS-NRPS hybrid architecture, whereas the type III PKS gene and the associated methyltransferase showed similarity to the alkylresorcinol biosynthetic gene cluster consisting of the operon *srsABC*<sup>55</sup>. The type III PKS operon from MSr9315 is the most distant phylogenetic relative of type III PKSs that has been still recognized as homologue of the alkylresorcinol biosynthetic gene cluster. The putative operon structures of MSr6788 and DSM 53668 seems to be quite conserved. However it is still questionable if all the genes organized in this operon are indeed necessary for the formation of the underlying type III polyketide.

Family **VIII** (green color) harbors type III PKSs from *Anaeromyxobacter* sp. Fw109-5, *Anaeromyxobacter* sp. PSR-1 DNA, *Anaeromyxobacter dehalogenans* 2CP-C, *Anaeromyxobacter dehalogenans* 2CP-1, *Anaeromyxobacter* sp. K., whereas the type III PKS genes from *Anaeromyxobacter dehalogenans* 2CP-1 and *Anaeromyxobacter* sp. K are identical. All members of this family are sharing high amino acid sequence similarity to the type III PKS PKS11 from *Mycobacterium tuberculosis* and show a conserved biosynthetic gene cluster arrangement. Upstream of the gene encoding a type III PKS, a gene is located encoding a PDZ-domain containing peptidase. A similar operon structure as described for the *ftp-like* genes is located downstream of the structural type III PKS consisting of genes encoding an acyl carrier protein, an AMP-dependent synthetase and a FAD-binding protein.

Family **IX** (pink color) of type III PKSs is sharing comparatively high homology (more than 54% identical sites) to the *ftp-like* family **X** (blue colored strains **X**); however, the operon structure clearly varies in terms of the putative tailoring enzymes. The absence of the genes *ftpC* (acyl carrier protein) and *ftpD* (fatty acyl AMP-ligase) is observed, accounting for a similar operon structure like the *srsABC* operon<sup>55</sup>. However if the putative gene product is either an alkylresorcinol or simply a resorcinol cannot be answered by solely *in silico* analysis. Within this family of 16 members, one specific member is remarkable, because of the co-localized PKS-NRPS hybrid biosynthetic gene cluster. AntiSMASH analysis did not reveal any known homologues biosynthetic gene cluster.

The largest family of type III PKS genes in myxobacteria (**Figure 2**, light blue, **X**) – termed the *ftp-like* gene cluster family – occurs in nearly half of all sequenced myxobacterial genomes, indicating a highly conserved (albeit to date unknown) function in the order of *Myxococcales*. One interesting subgroup within this family **X** consists of eight members with the defining detail being the shortened *ftp* gene operon consisting solely of *ftpA* (type III PKS), *ftpC* (ACP) and *ftpD* (fatty acyl-AMP ligase) and missing *ftpB* (methyltransferase) and *ftpE* (monooxygenase). Seven strains harboring these type III PKS operons have been described in family **IV** and **V** (Meb2, Cbv34, Cbfe23, DSM52655, DSM2262, MCy9104, SbCb021) which are co-localized with type III PKS gene operons from family **V**, whereas *Melittangium boletus* DSM 14713 harbors only this type III PKS operon. This truncated operon might hint at a preference of myxobacterial type III PKSs to produce non-oxidized and non-methylated alkylresorcinols or alkylpyrones.

The circular phylogenetic tree reveals that family **XI** (dark blue color) is most distantly related to the *ftp*-like gene clusters. The family comprises three type III PKS genes which share solely an operon consisting of genes encoding a type III PKS and a FAD dependent oxidoreductase. All three strains were probably imprecisely annotated as mixed type III PKS biosynthetic gene clusters; *Byssosvorax cruenta* strain MSr4204 (Byc1) as type III PKS-lantipeptide BGC, *Chondromyces catenulatus* strain MSr9030 (SBCm007) as bacteriocin-type III PKS-type I PKS BGC and *Racemicystis crocea* MSr9521 (SBSr021) as bacteriocin-type III PKS-type I PKS-NRPS BGC.

## 2.2 Genes encoded in the *ftp*-like biosynthetic gene cluster from MCy9487

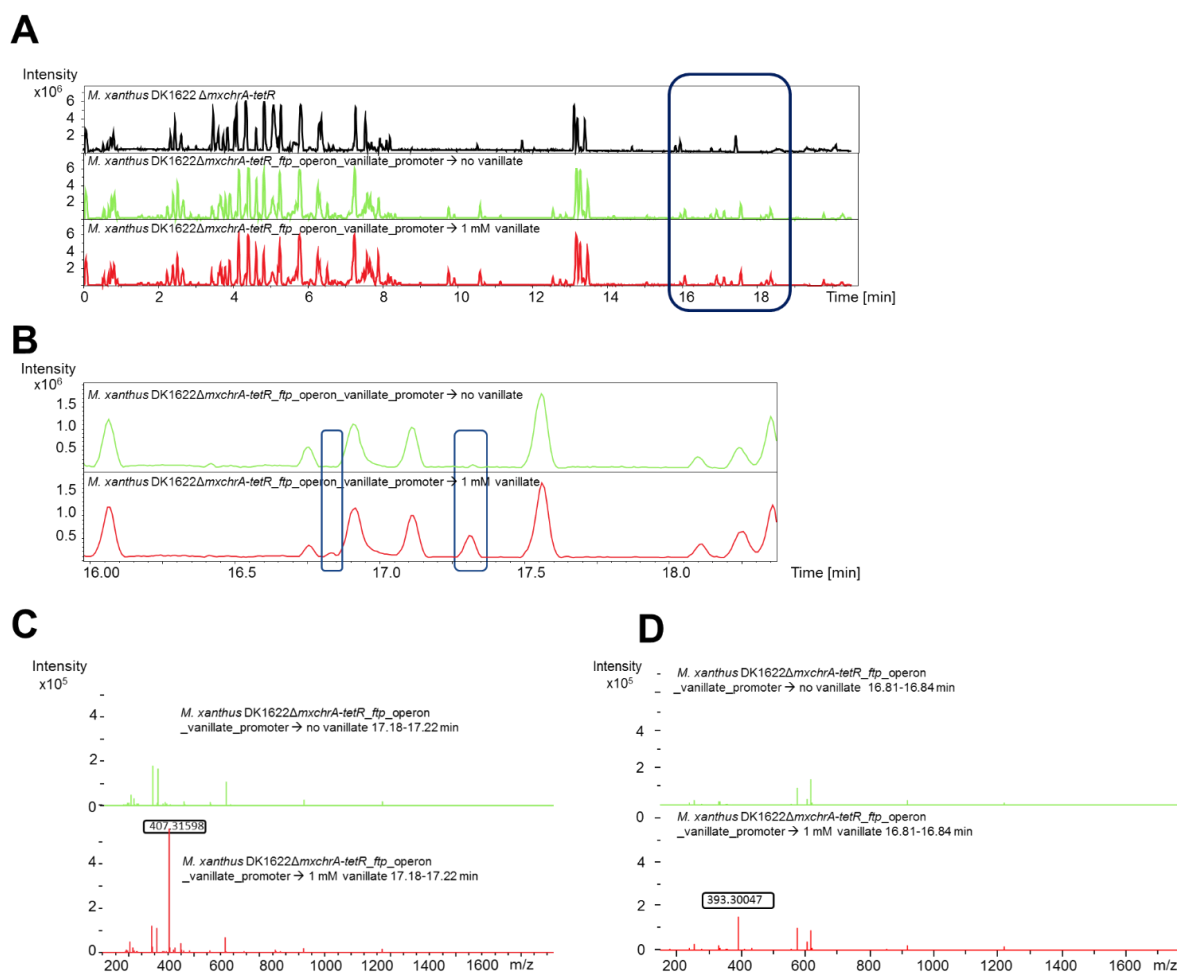
**Table S6:** Table of all open reading frames (orfs) assigned by antiSMASH to the putative *ftp*-like biosynthetic gene cluster with proposed function and closest homologue according to blastp search in the nr (non-redundant protein sequences) protein database at NCBI.

Name	Length (bp)	Proposed function	Closest homologue	Accession No.	Identity (%)	Query coverage (%)
<i>orf1</i>	2154	-	Hypothetical protein	WP_014400034	66.3	99.86
<i>orf2</i>	660	-	Peptidase	WP_014400033	96.1	90.45
<i>orf3</i>	2202	-	Sensor histidine kinase	WP_014400032	83.5	99.73
<i>orf4</i>	555	-	RNA 2',3'-cyclic phosphodiesterase	WP_014400031	78.0	98.38
<i>orf5</i>	543	-	Metal-dependent hydrolase	WP_014400030	82.1	95.03
<i>orf6</i>	1242	-	Adenylate/guanylate cyclase domain-containing protein	WP_014400029	82.0	99.52
<i>orf7</i>	1434	-	Prolyl-tRNA ligase	WP_014400028	92.5	99.79
<i>orf8</i>	696	-	DUF4336 domain-containing protein	WP_014400027	71.0	99.57
<i>orf9</i>	630	-	Hypothetical protein	WP_014400026	79.0	92.38
<i>orf10</i>	876	-	Hypothetical protein	WP_014400025	77.7	95.55
<i>orf11</i>	555	-	Hypothetical protein	AFE07418	83.2	98.92
<i>orf12</i>	573	-	Hypothetical protein COCOR_07240	AFE07417	86.1	97.91
<i>orf13</i>	594	-	MarR family transcriptional regulator	WP_043322285	83.8	80.81
<i>akp1</i>	816	Resistance	ABC transporter	WP_014400021	77.5	100.0
<i>akp2</i>	696	Resistance	ABC transporter permease subunit	WP_014400020	93.3	96.97
<i>akp3</i>	1068	Resistance	ABC transporter ATP-	WP_014400019	89.5	99.44

			binding protein			
<i>akp4</i>	663	Self-resistance	Pentapeptide repeat protein	AFE09361	70.0	95.02
<i>akp5</i>	2421	-	Glycogen-debranching protein	WP_014400015	77.0	99.88
<i>akp6</i>	489	-	VOC family protein	WP_014393563	69.0	96.93
<i>akp7</i>	309	-	Serine/threonine protein phosphatase	OUW87376	32.9	79.61
<i>akp8</i>	1089	Type III PKS	Stilbene synthase [Corallococcus coralloides]	WP_014399992	91.8	99.72
<i>akp9</i>	579	Methyl-transferase	Isoprenylcysteine carboxyl methyltransferase	WP_014399991	89.5	99.48
<i>akp10</i>	276	Fatty acyl-ACP	Acyl carrier protein	WP_014399990	84.6	98.91
<i>akp11</i>	1755	Fatty acyl-ACP catalysis	AMP-binding protein	AFE08735	88.5	99.83
<i>akp12</i>	1098	Hydroxylation	NAD(P)/FAD-dependent oxidoreductase	WP_014399988	89.3	99.73
<i>akp13</i>	630	-	Hypothetical protein	WP_014399986	87.1	88.57
<i>akp14</i>	1476	-	Endonuclease	WP_014399985	81.2	99.19
<i>akp15</i>	327	-	Transcriptional regulator	WP_054208909	69.9	95.37
<i>akp16</i>	996	-	Quinone oxidoreductase	WP_037238406	48.3	99.70
<i>orf14</i>	501	-	Hypothetical protein	WP_002636395	80.7	96.41
<i>orf15</i>	711	-	NIPSNAP family protein	WP_014399979	71.8	97.47
<i>orf16</i>	1515	-	PLP-dependent aminotransferase family protein [	WP_002636397	84.0	97.82
<i>orf17</i>	378	-	ArsR family transcriptional regulator	WP_014399977	88.2	87.30
<i>orf18</i>	465	-	ATPase	WP_014399976	62.8	92.90
<i>orf19</i>	486	-	Hypothetical protein	WP_014399975	90.1	99.38
<i>orf20</i>	1071	-	Hypothetical protein	WP_014399974	86.2	99.44
<i>orf21</i>	579	-	Uma2 family endonuclease	WP_014399972	79.1	98.96
<i>orf22</i>	1041	-	O-methyltransferase	WP_014399970	94.5	99.71
<i>orf23</i>	1320	-	Response regulator	WP_014399968	96.6	99.77
<i>orf24</i>	642	-	Hypothetical protein	WP_014399967	95.3	99.07
<i>orf25</i>	483	-	Hypothetical protein	WP_014399966	74.2	96.27
<i>orf26</i>	606	-	Thioredoxin-dependent thiol peroxidase	WP_014399965	92.8	75.74
<i>orf27</i>	516	-	Hypothetical protein	WP_085960277	75.1	99.42

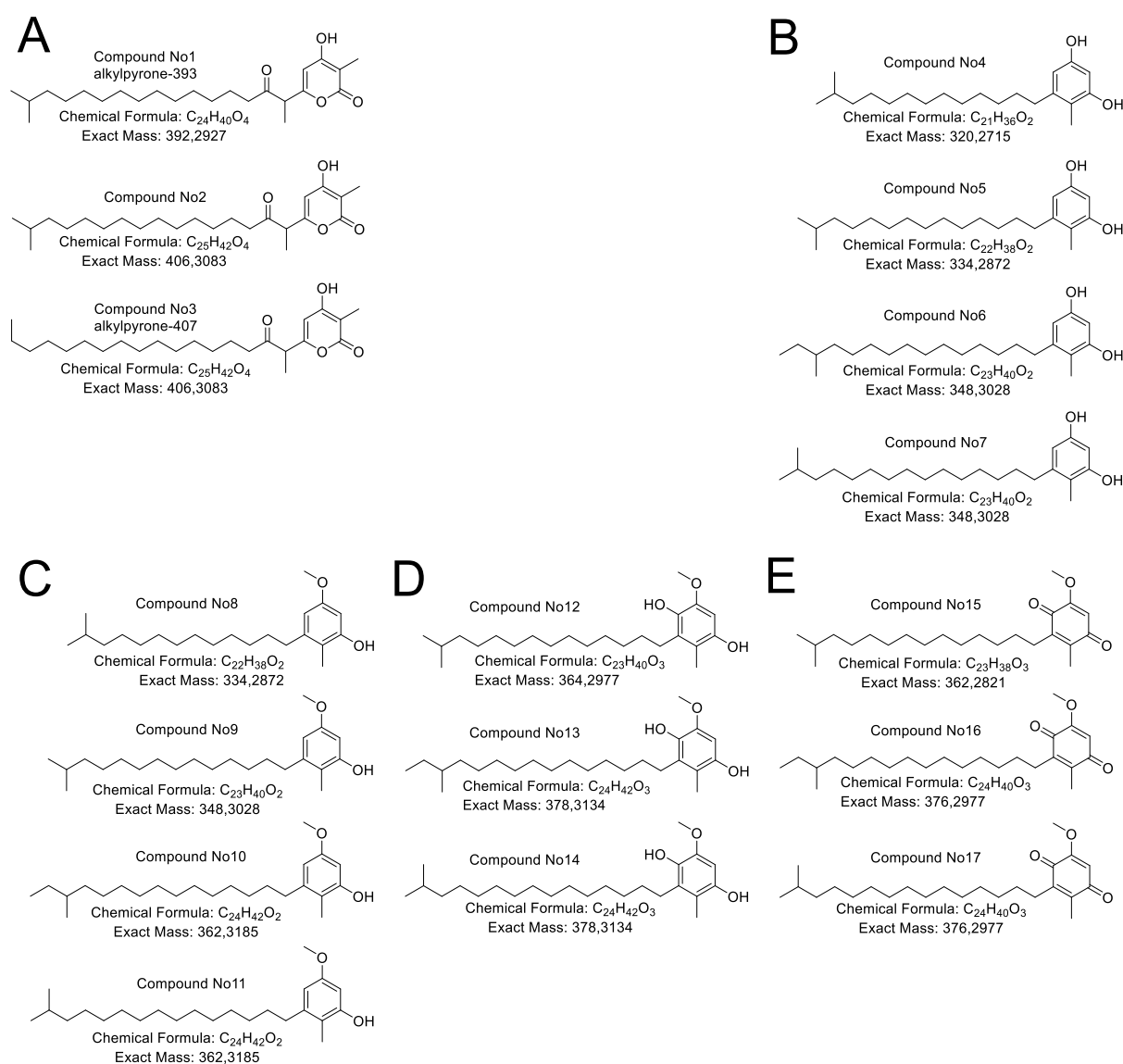
## 2.3 Investigation MCy9487 crude extract and secondary metabolism *M. xanthus* DK1622\_vanillate kanamycin mutant for *srs* operon compounds

Acquired HPLC-MS data of the extracts of MCy9487 (**Figure S1**) and *M. xanthus* DK1622\_vanillate\_promoter mutant (**Figure S2**) were investigated for the compounds ([M+H]<sup>+</sup>) isolated from recombinant *S. lividans* TK21 carrying *srs* genes (**Figure S3**), since both isolated type III polyketides afforded by myxobacterial *ftp*-like operons are identical to the alkylpyrones described from the *srs*-operon. In the metabolome of the crude extract of MCy9487 only the mass of 407.3161 m/z could be detected, which can be connected to either compound No2 or No3 (**Figure S2**). None of the other masses (B, D–J) corresponding to the characterized compounds generated by the *srs* operon could be detected. In contrast, the metabolome of *M. xanthus* DK1622\_vanillate\_promoter mutant highlighted not only the mass of alkylpyrone-393 (B) and alkylpyrone-407 (C) but also in lower intensity the mass of 365.3056 m/z was detected, which could be connected to compound No12, (see **Figure S2**). Trace amounts of mass 363.3263 m/z could be detected, which might be connected to either compound No10, No11 or No15 (see **Figure S2**). This investigation supports the assumption that myxobacteria harboring the *ftp*-operon (harboring the necessary tailoring genes to afford alkylquinones) preferably produce the alkylpyrone products.

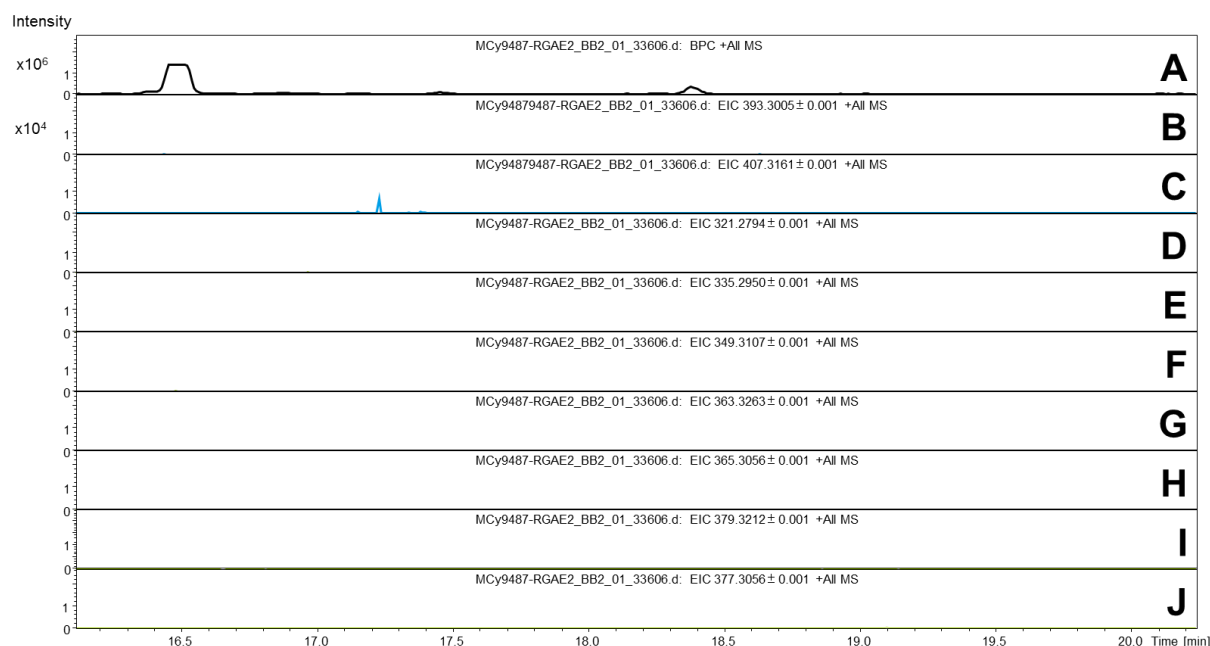




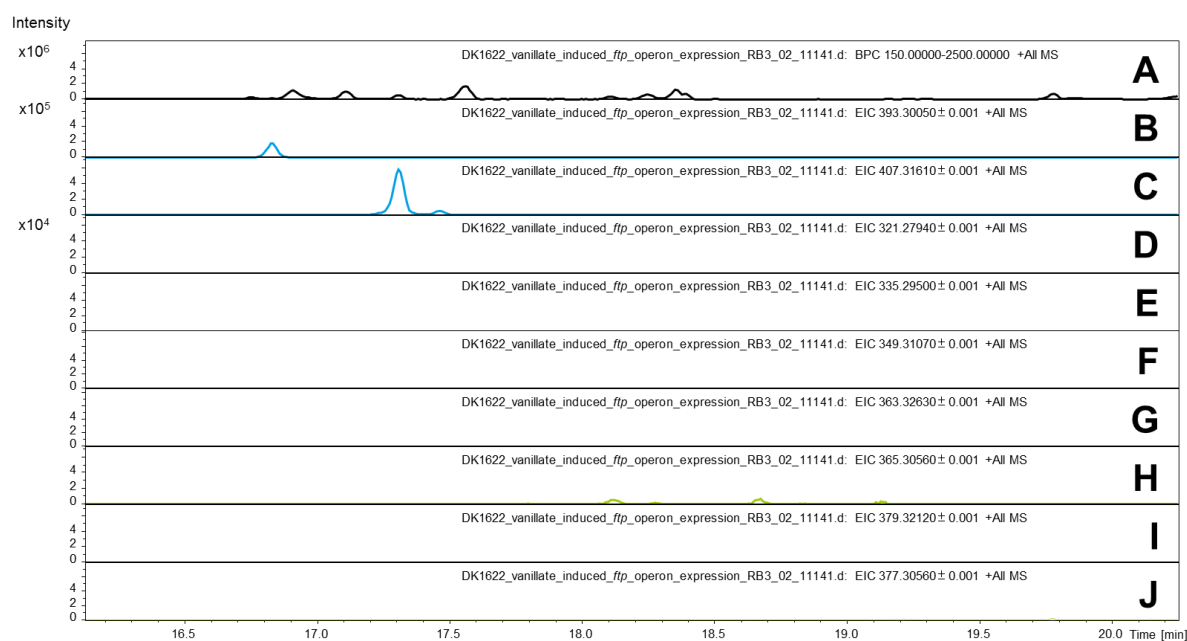
**Figure S1.** A) HPLC-MS BPCs of crude extracts from *M. xanthus* DK1622  $\Delta$ *mxchrA-tetR\_fip\_operon\_vanillate\_promoter* without vanillate supplementation and with 1 mM vanillate supplementation. Red dashed box indicates magnified area of the HPLC-MS BPCs, shown in **B**. The crude extract of *M. xanthus* DK1622 *mxchrA-tetR* was used as control. Red boxes in **B** indicate monitoring time of **C** and **D**. **C** and **D**) MS-spectra of heterologous expression of the type III PKS from MCy9487. MS-spectra **C** is showing the major occurring mass 407.316 Da and MS-spectra **D** is showing the major occurring mass 393.300 Da. The MS-spectrum of *M. xanthus* *mxchrA-tetR* DK1622 was used as control



**Figure S2.** Structures isolated from recombinant *S. lividans* TK21 carrying *srs* genes<sup>55</sup>. **A**) Alkylpyrones produced by *srsA*. **B**) Alkylresorcinols produced by *SrsA*. **C**) Methylresorcinols produced by *SrsA* and *SrsB* **D**) Methylated Alkylhydroxyquinols afforded by *SrsA*, *SrsB* and *SrsC*. **E**) Alkylquinones afforded by *SrsA*, *SrsB* and *SrsC*.



**Figure S3.** Partial BPC (A) and extracted ion chromatograms (B–J) of MCy9487 wild type extract. The mass of 407.3161 m/z could be detected in the crude extract of MCy9487, which can be connected to either compound No2 or No3 (**Figure S2**). Trace amounts of 393.3005 m/z (B, not visible on the extracted ion chromatogram) could be detected, which might be connected to compound No1 (**Figure S2**). None of the other masses (B, D–J) corresponding to the characterized compounds generated by the *srs* operon could be detected.

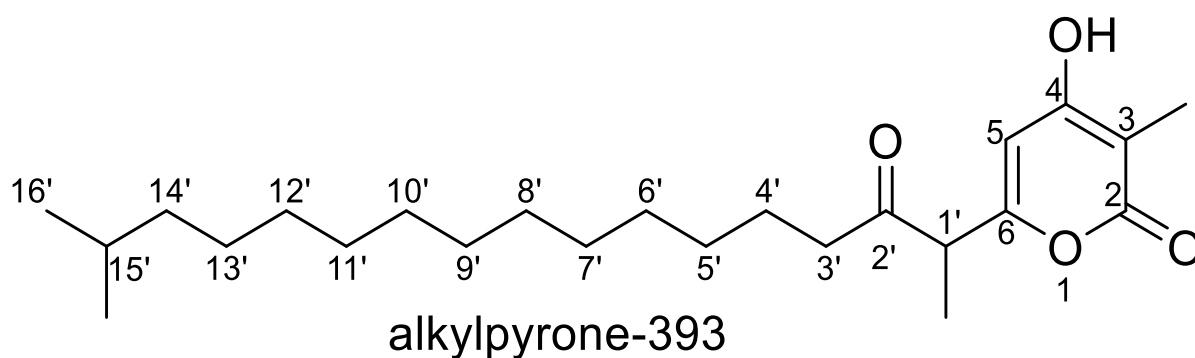


**Figure S4.** Partial BPC (A) and extracted ion chromatograms (B–J) of vanillate induced *M. xanthus* DK1622\_vanillate kanamycin mutant extracts. As reported the mass of alkylpyrone-393 (B) and alkylpyrone-407 (C) could be detected in the crude extract. In comparison, the mass of 365.3056 m/z was detected in significant lower amount (H) which could be connected to compound No12, (**Figure S2**). Trace amounts of mass 363.3263 m/z (G, not visible on the extracted ion chromatogram) could be detected, which might be connected to either compound No10, No11 or No15 (**Figure S2**). None of the other masses (D–F, I, J) corresponding to further characterized compounds generated by the *srs* operon could be detected.

## 2.4 NMR spectroscopic data

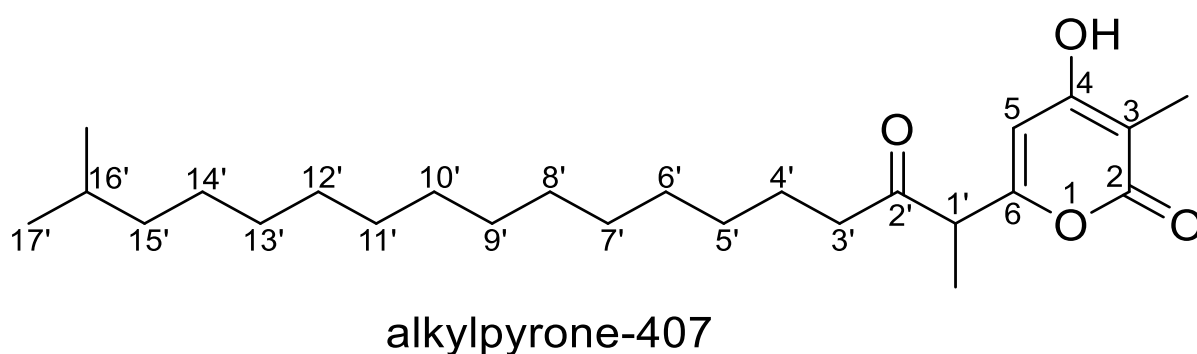
**Table S7.** NMR spectroscopic data of alkylpyrone-393

No.	$^{13}\text{C}$ ( $\delta$ )	$^1\text{H}$ ( $\delta$ )	$J$ [Hz]	Multiplicity	Chemistry
1	---	---	---	---	O
2	173.4	---	---	---	C
3	98.4	---	---	---	C
3-CH <sub>3</sub>	8.6	1.83	---	s	CH <sub>3</sub>
4	169.1	---	---	---	C
5	105.6	5.99	---	s	CH
6	161.4	---	---	---	C
1'	51.9	3.66	7.1	q	CH
1'-CH <sub>3</sub>	14.1	1.32	7.1	d	CH <sub>3</sub>
2'	209.5	---	---	---	C
3'	41.7	2.53	---	m	CH <sub>2</sub>
4'	24.7	1.53	---	m	CH <sub>2</sub>
5'	30.1	1.25-1.27	---	m	CH <sub>2</sub>
6'-12'	30.4-30.8	1.25-1.27	---	m	CH <sub>2</sub>
13'	28.5	1.25-1.27	---	m	CH <sub>2</sub>
14'	40.3	1.17	---	m	CH <sub>2</sub>
15'	29.1	1.52	---	m	CH
15'-CH <sub>3</sub>	23.0	0.88	6.6	d	CH <sub>3</sub>
16'	23.0	0.88	6.6	d	CH <sub>3</sub>

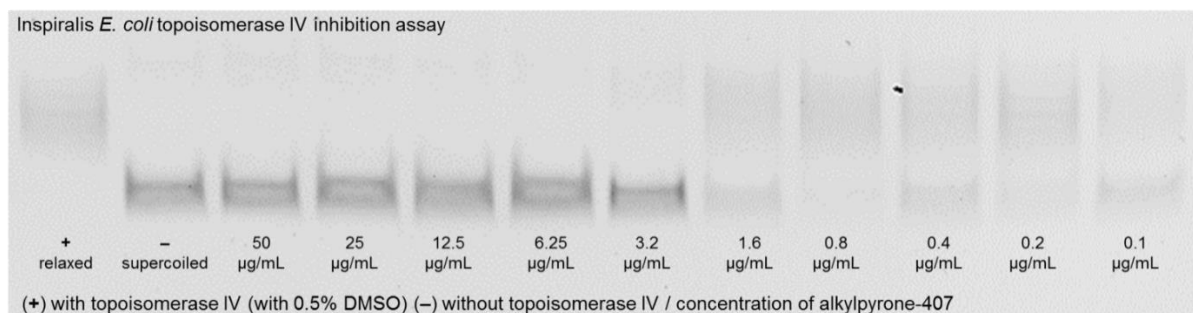
**Figure S5.** Structure and carbon numbering of alkylpyrone-393.

**Tab. S8** NMR spectroscopic data of alkylpyrone-393

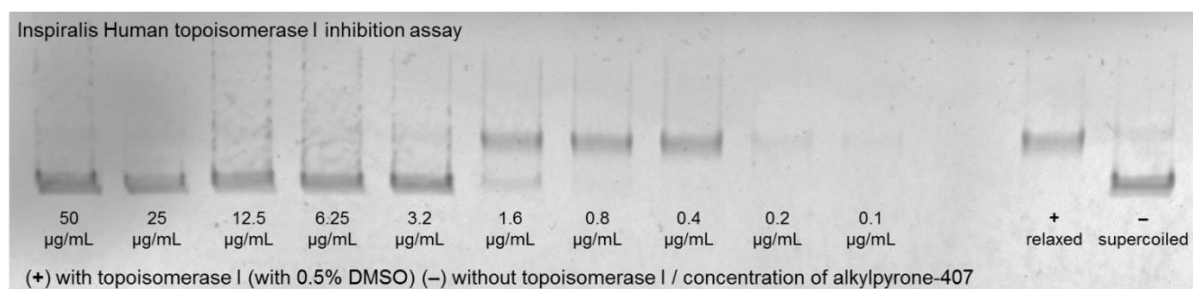
No.	$^{13}\text{C}$ ( $\delta$ )	$^1\text{H}$ ( $\delta$ )	$J$ [Hz]	Multiplicity	Chemistry
1	---	---	---	---	O
2	173,5	---	---	---	C
3	97,5	---	---	---	C
3-CH <sub>3</sub>	8,3	1,83	---	s	CH <sub>3</sub>
4	169,1	---	---	---	C
5	105,6	5,99	---	s	CH
6	160,7	---	---	---	C
1'	51,6	3,65	7.1	q	CH
1'-CH <sub>3</sub>	13,8	1,32	7.1	d	CH <sub>3</sub>
2'	208,9	---	---	---	C
3'	41,8	2,53	---	m	CH <sub>2</sub>
4'	24,4	1,53	---	m	CH <sub>2</sub>
5'	30,1	1,26	---	m	CH <sub>2</sub>
6'-13'	30,5-31,0	1.25-1.27	---	m	CH <sub>2</sub>
14'	28,9	1,30	---	m	CH <sub>2</sub>
15'	40,0	1,17	---	m	CH <sub>2</sub>
16'	29,9	1,53	---	m	CH
16'-CH <sub>3</sub>	22,8	0,87	6.6	d	CH <sub>3</sub>
17'	22,8	0,87	6.6	d	CH <sub>3</sub>

**Figure S6.** Structure and carbon numbering of alkylpyrone-407.

## 2.5 Topoisomerase inhibition assays of alkylpyrone-407



**Figure S7.** *E. coli* DNA topoisomerase IV inhibition assay. The MIC of alkylpyrone-407 to inhibit topoisomerase IV is 1.6–3.2 µg/mL. Inhibition assay was performed using topoisomerase inhibition evaluation kits for *E. coli* topoisomerase IV supplied by Inspiralis (Inspiralis Ltd., Norwich, U.K.).



**Figure S8.** Inspiralis Human topoisomerase I inhibition assay. The MIC of alkylpyrone-407 to inhibit topoisomerase I is 1.6–3.2 µg/mL. Inhibition assay was performed using topoisomerase inhibition evaluation kits for human topoisomerase I supplied by Inspiralis (Inspiralis Ltd., Norwich, U.K.).

## 2.6 Copies of $^1\text{H}$ and $^{13}\text{C}$ NMR spectra of alkylpyrone-407 and alkylpyrone-393

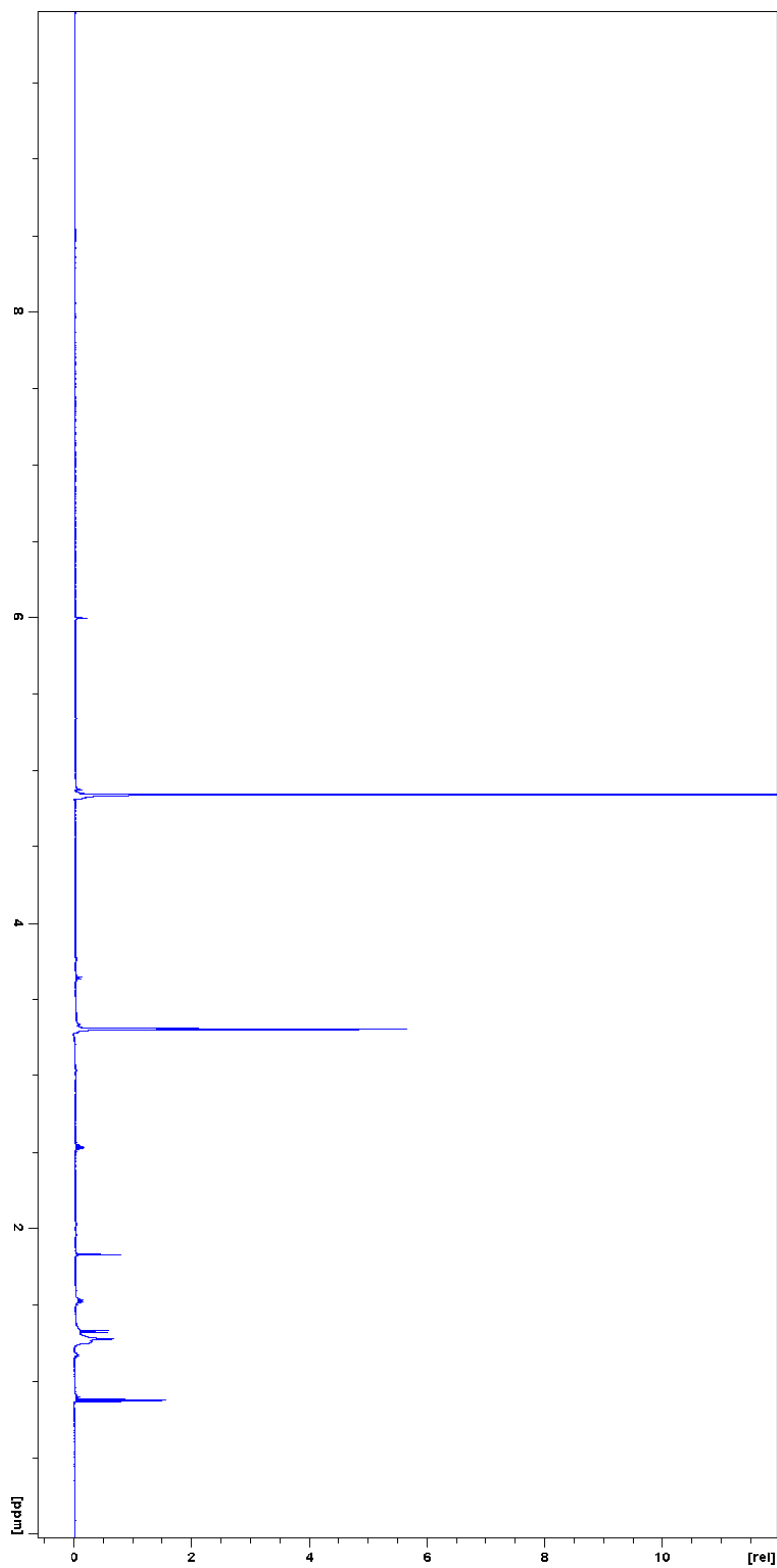


Figure S9.  $^1\text{H}$  NMR spectrum of alkylpyrone-407 in MeOD.

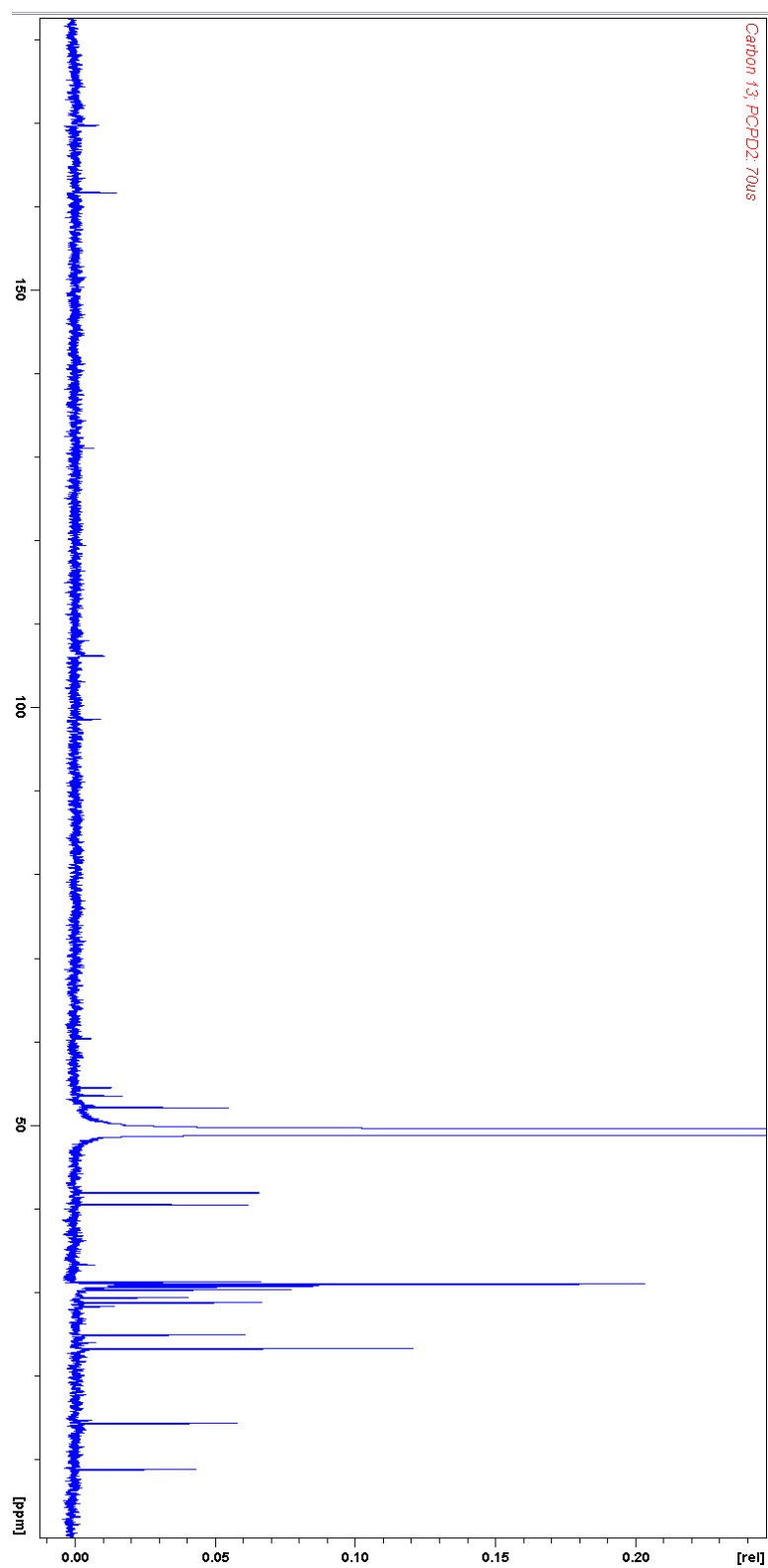
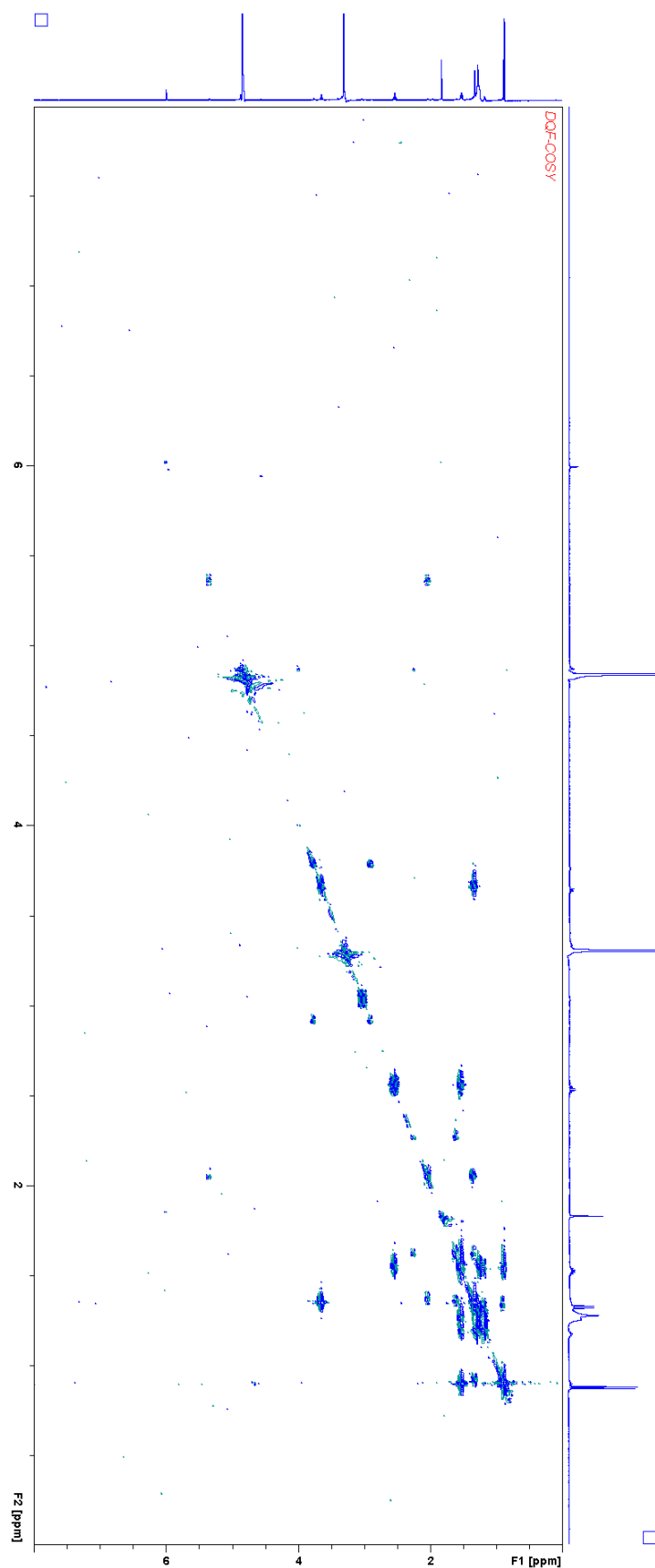
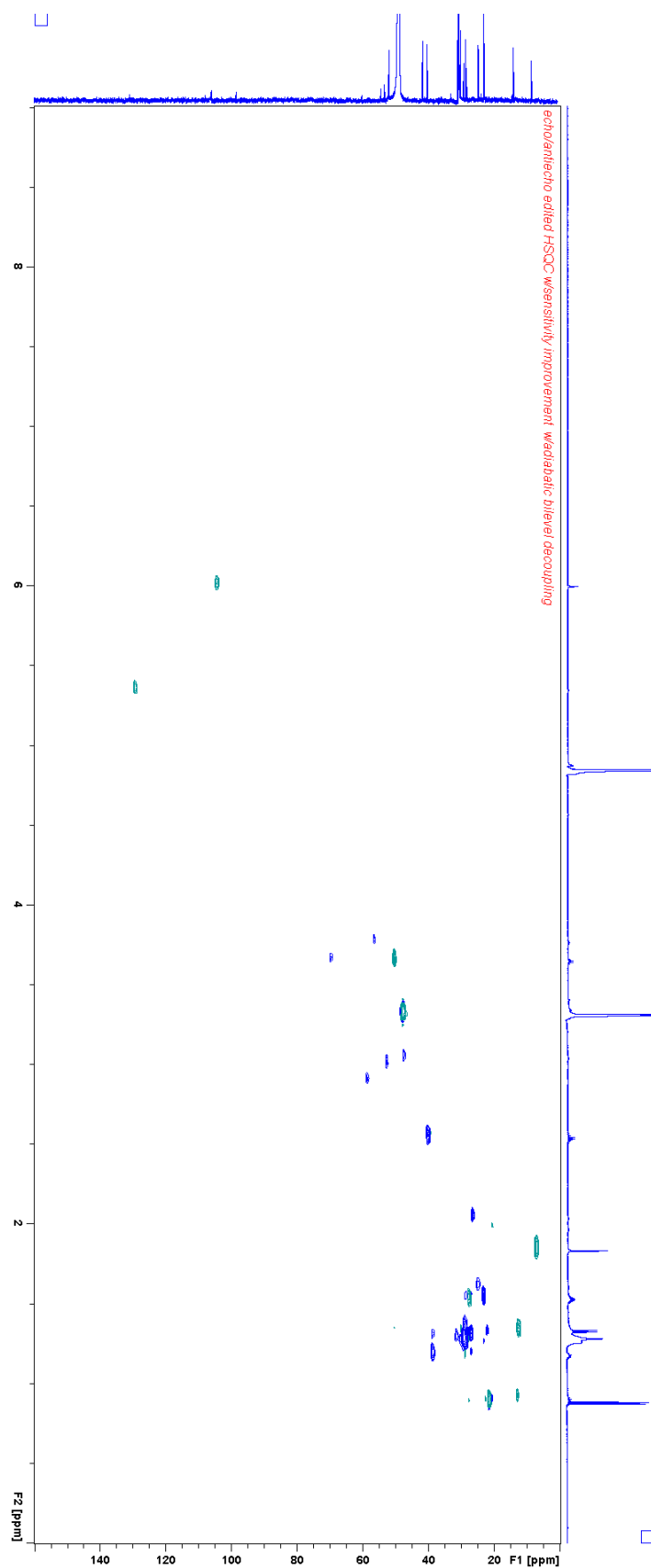


Figure S10.  $^{13}\text{C}$  NMR spectrum of alkylpyrone-407 in MeOD.



**Figure S11.** DQF COSY NMR spectrum of alkylpyrone-407 in MeOD.





**Figure S12.** echo/antiecho edited HSQC NMR spectrum of alkylpyrone-407 in MeOD.

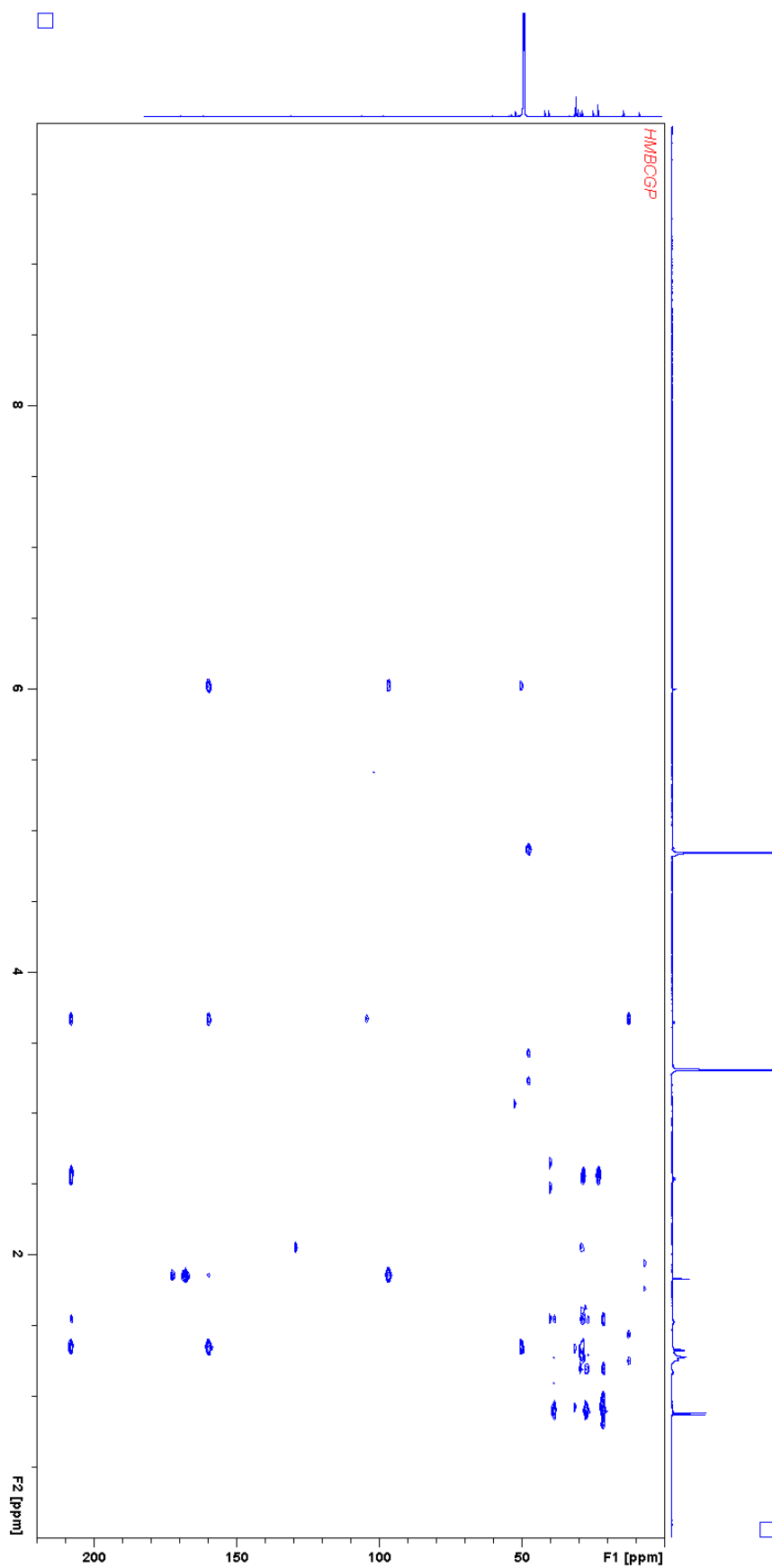
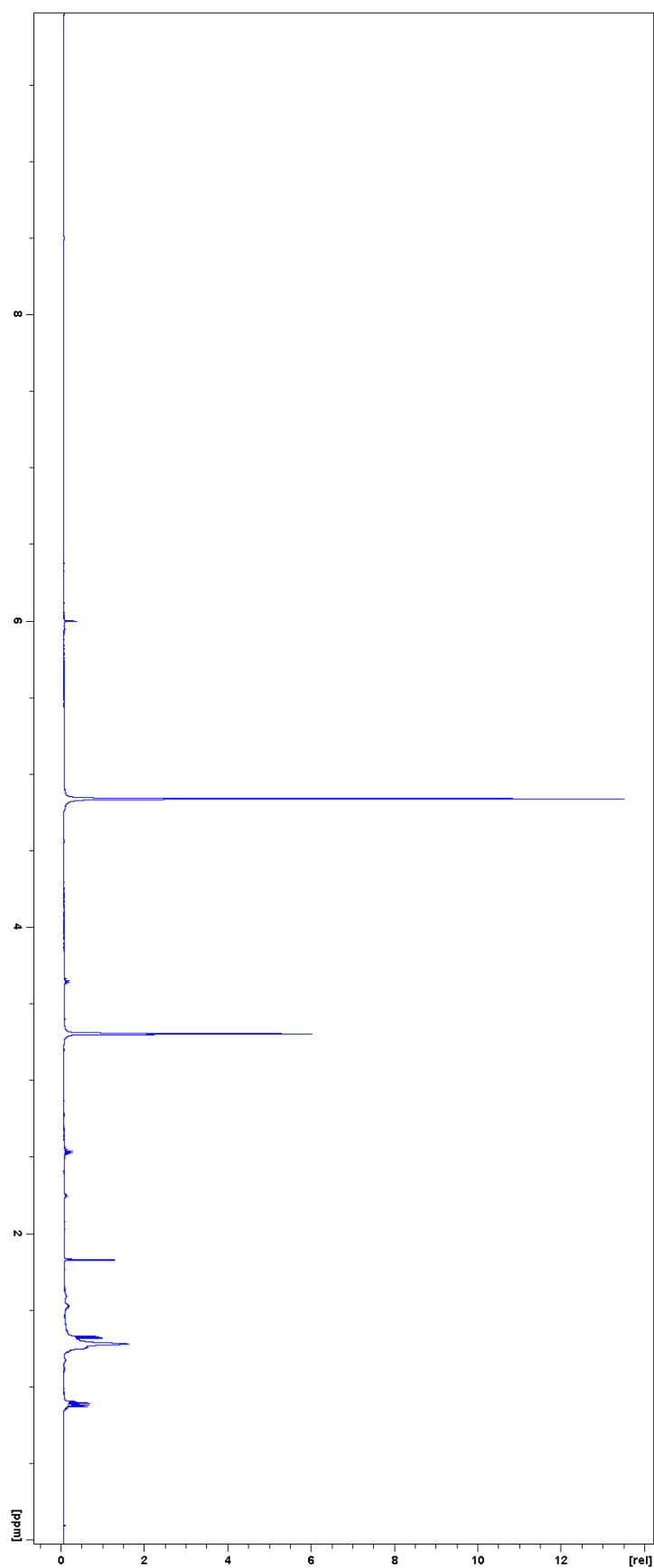


Figure S13. HMBC of alkylpyrone-407 in MeOD.



**Figure S14.**  $^1\text{H}$  NMR spectrum of alkylpyrone-393 in  $\text{MeOD}$ .

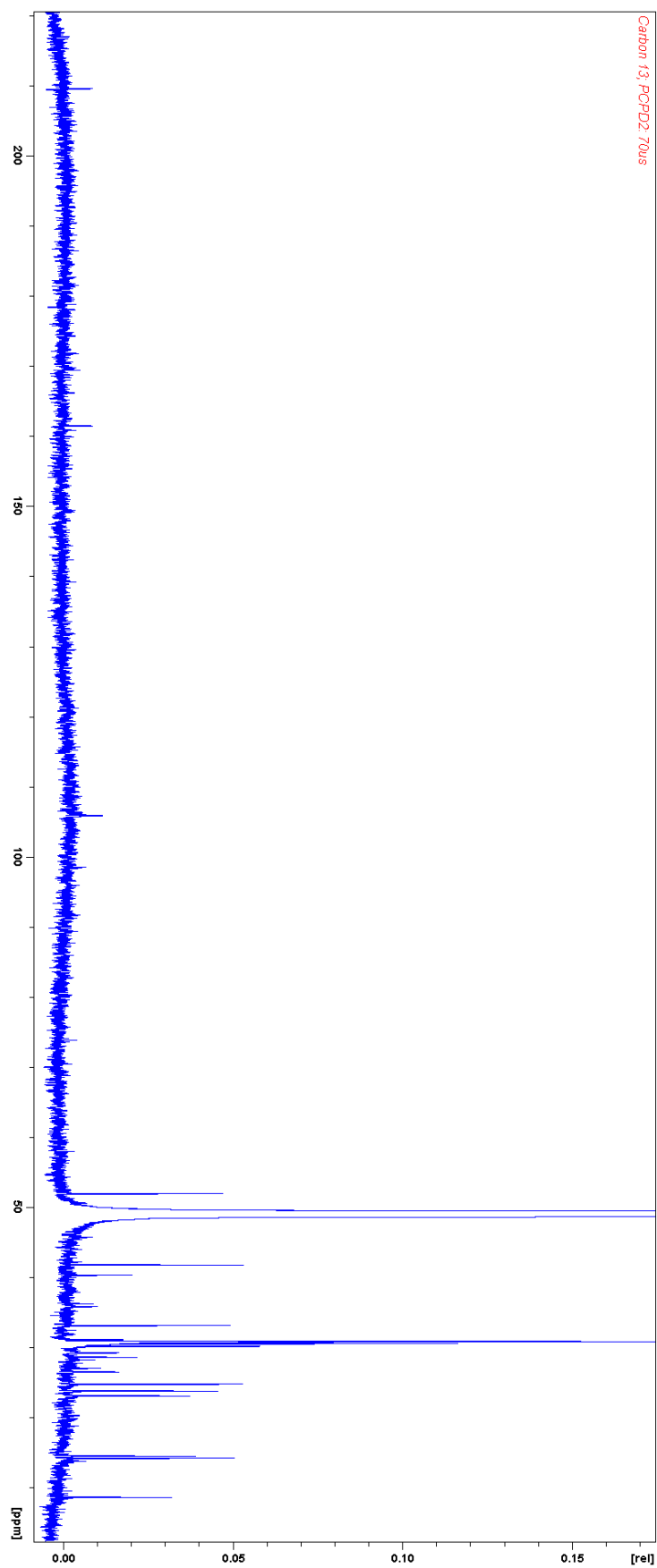
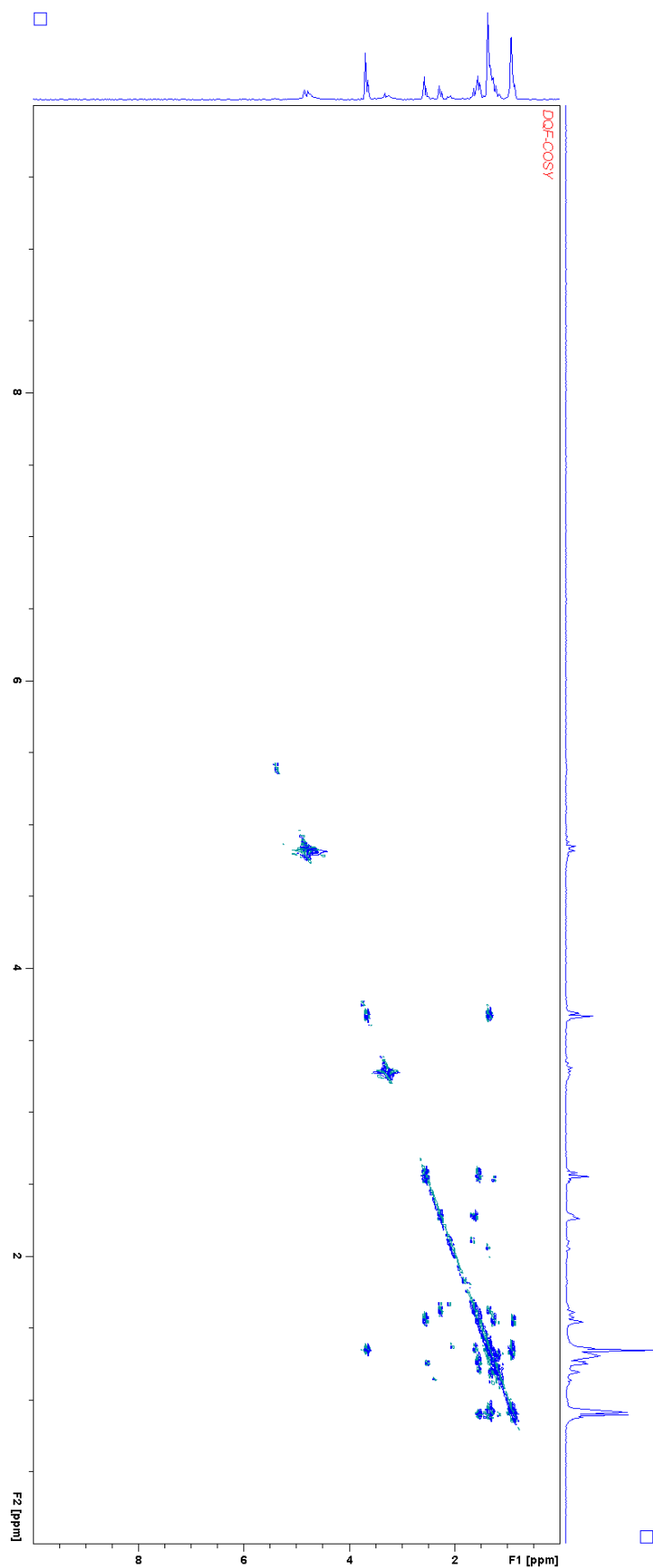
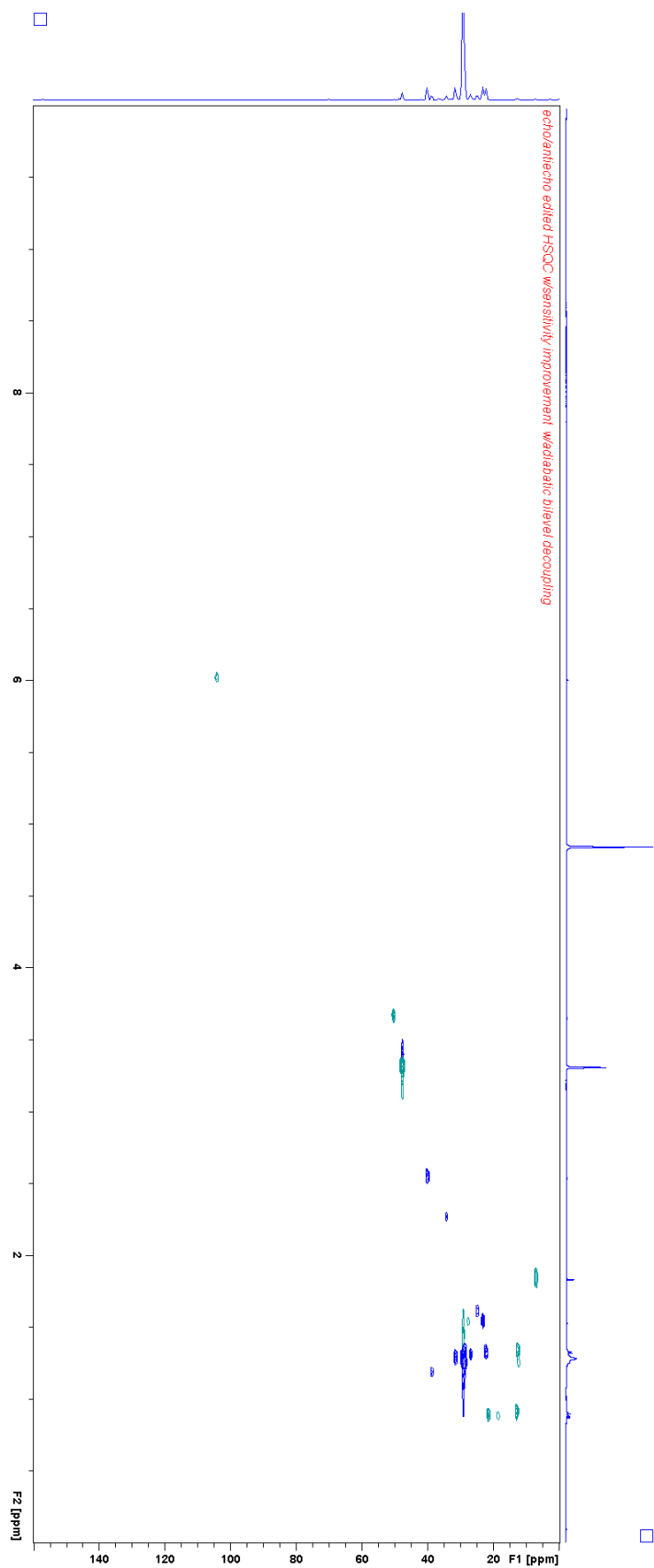


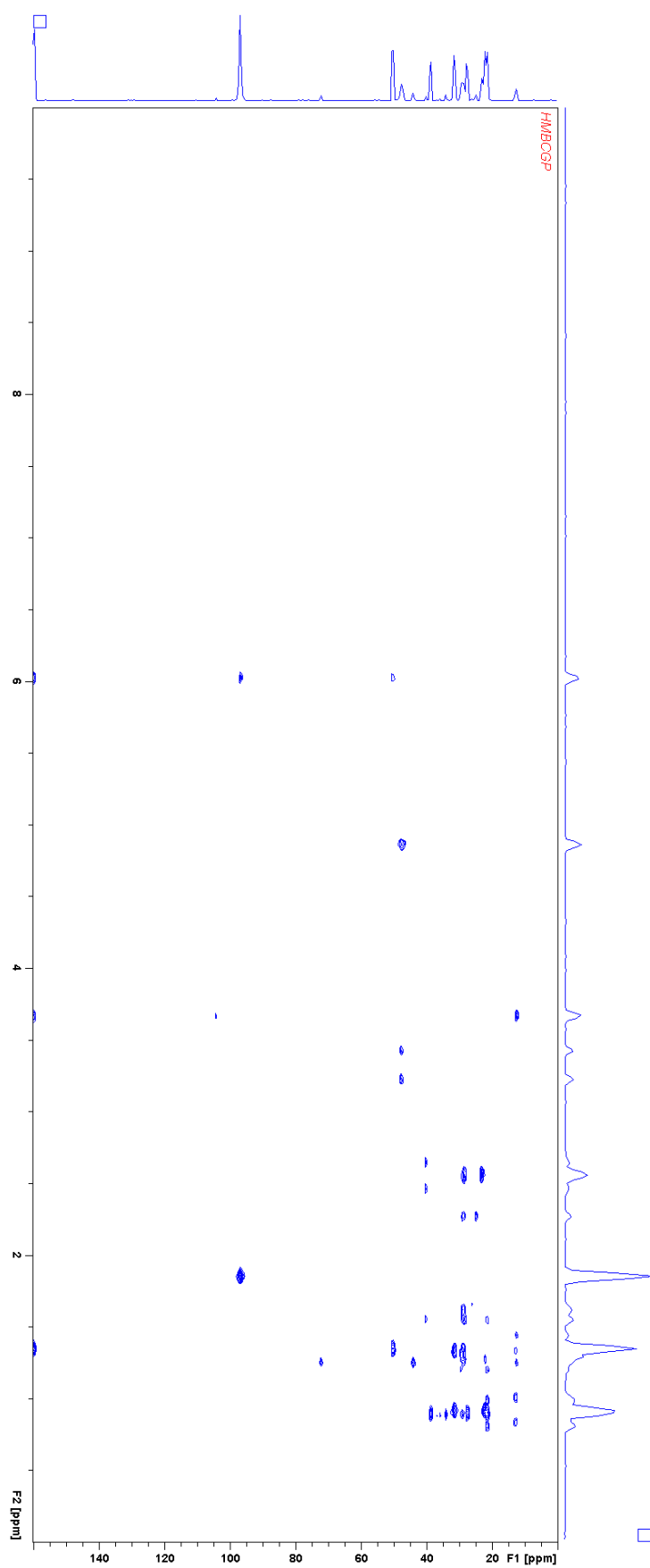
Figure S15.  $^{13}\text{C}$  NMR spectrum of alkylpyrone-393 in MeOD.



**Figure S16.** DQF COSY NMR spectrum of alkylpyrone-393 in MeOD.



**Figure S17.** Echo/antiecho edited HSQC NMR spectrum of alkylpyrone-393 in MeOD.



**Figure S18.** HMBC of alkylpyrone-393 in MeOD.

### 3 References

1. Gemperlein, K., Rachid, S., Garcia, R. O., Wenzel, S. C. & Müller, R. Polyunsaturated fatty acid biosynthesis in myxobacteria. Different PUFA synthases and their product diversity. *Chem. Sci.* **5**, 1733–1741 (2014).
2. Sood, S. *et al.* *Aggregicoccus edonensis* gen. nov., sp. nov., an unusually aggregating myxobacterium isolated from a soil sample. *Int. J. Syst. Evol. Microbiol.* **65**, 745–753 (2015).
3. Sasse, F. *et al.* Argyrins, immunosuppressive cyclic peptides from myxobacteria. I. Production, isolation, physico-chemical and biological properties. *J. Antibiot.* **55**, 543–551 (2002).
4. Sasse, F., Steinmetz, H., Hofle, G. & Reichenbach, H. Archazolid, new cytotoxic macrolactone from *Archangium gephyra* (Myxobacteria). Production, isolation, physico-chemical and biological properties. *J. Antibiot.* **56**, 520–525 (2003).
5. Jundt, L. *et al.* Isolation and Structure Elucidation of Cruentarens A and B - Novel Members of the Benzolactone Class of ATPase Inhibitors from the Myxobacterium *Byssovorax cruenta*. *Eur. J. Org. Chem.*, 5036–5044 (2006).
6. Herrmann, J., Hüttel, S. & Müller, R. Discovery and biological activity of new chondramides from *Chondromyces* sp. *ChemBioChem* **14**, 1573–1580 (2013).
7. Fu, C. *et al.* Solving the puzzle of the one-carbon loss in ripostatin biosynthesis. *Angew. Chem. Int. Ed. Engl.* **56**, 2192–2197 (2017).
8. Irschik, H. & Reichenbach, H. An unusual pattern of carbohydrate utilization in *Corallococcus* (*Myxococcus*) *coralloides* (Myxobacteriales). *Arch. Microbiol.* **142**, 40–44 (1985).
9. Chai, Y. *et al.* Discovery of 23 natural tubulysins from *Angiococcus disciformis* An d48 and *Cystobacter* SBCb004. *Chem. Biol.* **17**, 296–309 (2010).
10. Nadmid, S., Plaza, A., Garcia, R. & Müller, R. Cystochromones, unusual chromone-containing polyketides from the myxobacterium *Cystobacter* sp. MCy9104. *J. Nat. Prod.* **78**, 2023–2028 (2015).
11. Garcia, R. & Müller, R. *Simulacricoccus ruber* gen. nov., sp. nov., a microaerotolerant, non-fruiting, myxospore-forming soil myxobacterium and emended description of the family *Myxococcaceae*. *Int. J. Syst. Evol. Microbiol.* **68**, 3101–3110 (2018).
12. Akbar, S., Dowd, S. E. & Stevens, D. C. Draft Genome Sequence of *Cystobacter ferrugineus* Strain Cbfe23. *Genome Announc.* **5** (2017).



13. Keller, L. *et al.* Macyranonones: Structure, Biosynthesis, and Binding Mode of an Unprecedented Epoxyketone that Targets the 20S Proteasome. *J. Am. Chem. Soc.* **137**, 8121–8130 (2015).
14. Treuner-Lange, A., Bruckskotten, M., Rupp, O., Goesmann, A. & Søgaaard-Andersen, L. Whole-Genome Sequence of the Fruiting Myxobacterium *Cystobacter fuscus* DSM 52655. *Genome announc.* **5** (2017).
15. Baumann, S. *et al.* Cystobactamids: myxobacterial topoisomerase inhibitors exhibiting potent antibacterial activity. *Angew. Chem. Int. Ed.* **53**, 14605–14609 (2014).
16. Yan, F. *et al.* Biosynthesis and Heterologous Production of Vioprolides: Rational Biosynthetic Engineering and Unprecedented 4-Methylazetidinecarboxylic Acid Formation. *Angew. Chem. Int. Ed. Engl.* **57**, 8754–8759 (2018).
17. Okanya, P. W. *et al.* Hyafurones, hyapyrrolines, and hyapyrones: polyketides from *Hyalangium minutum*. *J. Nat. Prod.* **77**, 1420–1429 (2014).
18. Okanya, P. W. Isolation and structure elucidation of secondary metabolites from the gliding bacteria *Ohtaekwangia kribbensis* and *Hyalangium minutum*, 2012.
19. Klefisch, T. Untersuchungen methylierter Myxocheline und neuer Sekundärmetabolitfamilien aus Myxobakterien. Dissertation. Saarland University, 2014.
20. Reichenbach, H. in *Bergey's manual of systematic bacteriology*, edited by D.J. Brenner, N.R. Krieg & J.T. Staley (Springer2005), Vol. 2, pp. 1059–1144.
21. Burgard, C. *et al.* Genomics-Guided Exploitation of Lipopeptide Diversity in Myxobacteria. *ACS Chem. Biol.* **12**, 779–786 (2017).
22. Lang, E. & Sproer, C. Replacement of ATCC 25944T, the current type strain of *Melittangium lichenicola*, with ATCC 25946. Request for an opinion. *Int. J. Syst. Evol. Microbiol.* **58**, 2991–2992 (2008).
23. Sucipto, H., Wenzel, S. C. & Müller, R. Exploring chemical diversity of a-pyrone antibiotics: molecular basis of myxopyronin biosynthesis. *ChemBioChem* **14**, 1581–1589 (2013).
24. Gorges, J. *et al.* Structure, Total Synthesis, and Biosynthesis of Chloromyxamides: Myxobacterial Tetrapeptides Featuring an Uncommon 6-Chloromethyl-5-methoxypiperic Acid Building Block. *Angew. Chem. Int. Ed. Engl.* **57**, 14270–14275 (2018).
25. Kunze, B., Trowitzsch-Kienast, W., Höfle, G. & Reichenbach, H. Nannochelins A, B and C, new iron-chelating compounds from *Nannocystis exedens* (myxobacteria). Production, isolation, physico-chemical and biological properties. *J. Antibiot.* **45**, 147–150 (1992).

26. Witte, S. N. R., Hug, J. J., Géraldy, M., Müller, R. & Kalesse, M. Biosynthesis and Total Synthesis of Pyrroazol B: a Secondary Metabolite from *Nannocystis pusilla*. *Chem. Eur. J.* **23**, 15917–15921 (2017).
27. Garcia, R. O., Reichenbach, H., Ring, M. W. & Müller, R. *Phaselicystis flava* gen. nov., sp. nov., an arachidonic acid-containing soil myxobacterium, and the description of *Phaselicystidaceae* fam. nov. *Int. J. Syst. Evol. Microbiol.* **59**, 1524–1530 (2009).
28. Iizuka, T. *et al.* *Pseudenhymyxa salsuginis* gen. nov., sp. nov., a myxobacterium isolated from an estuarine marsh. *Int. J. Syst. Evol. Microbiol.* **63**, 1360–1369 (2013).
29. Panter, F., Krug, D., Baumann, S. & Müller, R. Self-resistance guided genome mining uncovers new topoisomerase inhibitors from myxobacteria. *Chem. Sci.* **9**, 4898–4908 (2018).
30. Surup, F. *et al.* Disciformycins A and B: 12-membered macrolide glycoside antibiotics from the myxobacterium *Pyxidicoccus fallax* active against multiresistant staphylococci. *Angew. Chem. Int. Ed. Engl.* **49**, 13588–13591 (2014).
31. Kjaerulff, L. *et al.* Pyxipyrrolones: Structure elucidation and biosynthesis of cytotoxic myxobacterial metabolites. *Angew. Chem. Int. Ed.* **56**, 9614–9618 (2017).
32. Awal, R. P., Garcia, R. & Müller, R. *Racemicystis crocea* gen. nov., sp. nov., a novel soil myxobacterium in *Polyangiaceae*. *Int. J. Syst. Evol. Microbiol.*, 2389–2395 (2016).
33. Steinmetz, H. *et al.* Indiacens A and B: prenyl indoles from the myxobacterium *Sandaracinus amyolyticus*. *J. Nat. Prod.* **75**, 1803–1805 (2012).
34. Herrmann, M., Bohlendorf, B., Irschik, H., Reichenbach, H. & Höfle, G. Maracin and maracen: New types of ethynyl vinyl ether and alpha-chloro divinyl ether antibiotics from *Sorangium cellulosum* with specific activity against mycobacteria. *Angew. Chem. Int. Ed. Engl.* **37**, 1253–1255 (1998).
35. Kunze, B., Kemmer, T., Höfle, G. & Reichenbach, H. Stigmatellin, a new antibiotic from *Stigmatella aurantiaca* (Myxobacterales). I. Production, physico-chemical and biological properties. *J. Antibiot.* **37**, 454–461 (1984).
36. Wang, H. *et al.* RecET direct cloning and Red $\alpha\beta$  recombineering of biosynthetic gene clusters, large operons or single genes for heterologous expression. *Nat. Protoc.* **11**, 1175–1190 (2016).
37. Chang, A. C. & Cohen, S. N. Construction and characterization of amplifiable multicopy DNA cloning vehicles derived from the P15A cryptic miniplasmid. *J. Bacteriol.* **134**, 1141–1156 (1978).

38. Yanisch-Perron, C., Vieira, J. & Messing, J. Improved M13 phage cloning vectors and host strains: nucleotide sequences of the M13mp18 and pUC19 vectors. *Gene* **33**, 103–119 (1985).
39. Rubin, E. J. *et al.* *In vivo* transposition of mariner-based elements in enteric bacteria and mycobacteria. *Proc. Natl. Acad. Sci. USA* **96**, 1645–1650 (1999).
40. Kashefi, K. & Hartzell, P. L. Genetic suppression and phenotypic masking of a *Myxococcus xanthus* *frzF*- defect. *Mol. Microbiol.* **15**, 483–494 (1995).
41. Surup, F. *et al.* Crocadsins-Depsipeptides from the Myxobacterium *Chondromyces crocatus* Found by a Genome Mining Approach. *ACS Chem. Biol.* **13**, 267–272 (2018).
42. Pradella, S. *et al.* Characterisation, genome size and genetic manipulation of the myxobacterium *Sorangium cellulosum* So ce56. *Arch. Microbiol.* **178**, 484–492 (2002).
43. Youderian, P., Burke, N., White, D. J. & Hartzell, P. L. Identification of genes required for adventurous gliding motility in *Myxococcus xanthus* with the transposable element mariner. *Mol. Microbiol.* **49**, 555–570 (2003).
44. Sucipto, H., Pogorevc, D., Luxenburger, E., Wenzel, S. C. & Müller, R. Heterologous production of myxobacterial  $\alpha$ -pyrone antibiotics in *Myxococcus xanthus*. *Metab. Eng.* **44**, 160–170 (2017).
45. Wenzel, S. C., Bode, H. B., Kochems, I. & Müller, R. A type I/type III polyketide synthase hybrid biosynthetic pathway for the structurally unique *ansa* compound kendomycin. *ChemBioChem* **9**, 2711–2721 (2008).
46. Pfeifer, V. *et al.* A polyketide synthase in glycopeptide biosynthesis - The biosynthesis of the non-proteinogenic amino acid (*S*)-3,5-dihydroxyphenylglycine. *J. Biol. Chem.* **276**, 38370–38377 (2001).
47. Kuzuyama, T., Noel, J. P. & Richard, S. B. Structural basis for the promiscuous biosynthetic prenylation of aromatic natural products. *Nature* **435**, 983–987 (2005).
48. Kawasaki, T. *et al.* Biosynthesis of a natural polyketide-isoprenoid hybrid compound, furaquinocin A: Identification and heterologous expression of the gene cluster. *J. Bacteriol.* **188**, 1236–1244 (2006).
49. Pathirana, C., Jensen, P. R. & Fenical, W. Marinone and debromomarinone. Antibiotic sesquiterpenoid naphthoquinones of a new structure class from a marine bacterium. *Tetrahedron Lett.* **33**, 7663–7666 (1992).
50. Kalaitzis, J. A., Hamano, Y., Nilsen, G. & Moore, B. S. Biosynthesis and structural revision of neomarinone. *Org. Lett.* **5**, 4449–4452 (2003).

51. Winter, J. M. *et al.* Molecular basis for chloronium-mediated meroterpene cyclization: cloning, sequencing, and heterologous expression of the napyradiomycin biosynthetic gene cluster. *J. Biol. Chem.* **282**, 16362–16368 (2007).
52. Kelley, L. A., Mezulis, S., Yates, C. M., Wass, M. N. & Sternberg, M. J. E. The Phyre2 web portal for protein modeling, prediction and analysis. *Nat. Protoc.* **10**, 845–858 (2015).
53. Shah, S. & Heddle, J. G. Squaring up to DNA. Pentapeptide repeat proteins and DNA mimicry. *Appl. Biochem. Biotechnol.* **98**, 9545–9560 (2014).
54. Hoffmann, T., Müller, S., Nadmid, S., Garcia, R. & Müller, R. Microsclerodermins from terrestrial myxobacteria: An intriguing biosynthesis likely connected to a sponge symbiont. *J. Am. Chem. Soc.* **45**, 16904–16911 (2013).
55. Funabashi, M., Funa, N. & Horinouchi, S. Phenolic lipids synthesized by type III polyketide synthase confer penicillin resistance on *Streptomyces griseus*. *J. Biol. Chem.* **283**, 13983–13991 (2008).

## 7 Discussion and Outlook

### 7.1 General summary and scope of this work

The overarching aim of this thesis was to study the diversity of myxobacterial biosynthesis, with a special focus on natural products originating from uncommon biosynthetic pathways. The biosynthetic characterization of the pyrronazols (chapter 2) and the cittilins (chapter 3) was initiated from a retrobiosynthetic workflow, where the myxobacterial natural products have been described beforehand without prior knowledge of the biosynthesis. The biosynthetic comparison of these myxobacterial natural products features striking differences in their biosynthetic logic. While the pyrronazols are prototypical PKS-NRPS hybrids enabling precise *in silico* prediction of their biosynthesis<sup>1</sup>, the cittilins display a more enigmatic ribosomal synthesis route due to an unprecedented cyclization<sup>2</sup>. In contrast, the *in silico* biosynthetic investigation of the novel 2-hydroxysorangiadenosine<sup>3</sup> and its derivative sorangiadenosine<sup>4</sup> (chapter 4) and particularly the discovery of the novel natural product classes, the myxolipoxazoles and myxopyrimidinols (chapter 5), provided a better understanding of the biosynthesis of “rare” heterocyclic scaffolds. The acquired knowledge concerning the biosynthetic pathways of the sesquiterpene adenosides, the myxolipoxazoles and myxopyrimidinols, lays the groundwork for future in-depth characterization of their underlying biochemical logic. Myxobacterial type III PKSs can be considered as a disregarded natural product class, since no type III PKS derived secondary metabolite have been isolated up to date<sup>5</sup>. In contrast, myxobacterial genome sequences showed a variety of type III PKS genes, a finding that initiated the genome mining-guided isolation of type III PKS-derived natural products. Although the outcome did not achieve the identification of an entirely novel natural product, this study revealed hidden type III polyketides from *M. xanthus* DK1622 and enriched the knowledge of their biosynthetic logic.

The significance of this thesis can be subdivided into the applied concepts and techniques in chapter 2–6 and the thereby explored biosynthetic space of myxobacteria.

The first part of this chapter deals with the concepts and methods to access the biosynthesis of natural products, exemplifying in detail how the biosynthesis of natural products can be explored, with an emphasis on retrobiosynthetic investigations based on bioinformatics, stable isotope-labeled precursor feeding and genetic manipulation. It is important to mention that these genome-metabolome correlation approaches in order to identify, investigate and characterize biosynthetic pathways (as mainly conducted in this thesis) are essential perquisites for in-depth biochemical characterization of singular catalytic steps. Another approach starts from *in silico*-annotated genes to access the biosynthesis of previously unknown or not associated natural products. This genome mining-based approach can be applied to produce natural products, which would have been not found by conventional isolation-based workflows.

The second part of this chapter summarizes the biosynthetic insights of myxobacterial natural products that were provided by the findings presented in this thesis. Thereby, the biosynthetic space of myxobacteria was explored and expanded our knowledge of unprecedented biosynthetic pathways.

## 7.2 Concepts and methods to access the biosynthesis of natural products

### 7.2.1 Retrobiosynthesis: Retrospective Genome Mining

The concepts and methods aiming to access the biosynthesis of natural products differ significantly from the classical natural product discovery workflow. During the golden age of anti-infective natural product discovery, many new natural products were isolated and investigated for their potential as clinical drugs. However, most of these natural products have not made their way into the clinics and usually little biosynthetic information of these natural products was obtained<sup>6</sup>. Considering the laborious and limited methods available to obtain biosynthetic insights, primarily through feeding of stable isotope-labeled precursors, it is not surprising that the biosynthetic investigation of most of these natural products was not pursued.

With the advent of affordable whole-genome sequencing, it becomes a routine that upon the isolation of a new natural product, its underlying biosynthesis is also investigated. Genetic experiments or *in silico* investigation can lead to the identification of natural product derivatives produced by the biosynthetic machinery, structure elucidation or even enable initial natural product isolation such as described for the myxopyrimidinols (chapter 5). Understanding the biosynthetic logic of the natural product of interest can also help to set-up an appropriate analytical and preparative instrumentation<sup>7</sup>.

The typical retrobiosynthetic workflow starts from the manual investigation of biosynthetic building blocks harbored in the isolated natural product (**Figure 1**). Amino acids, nucleic acids, carbohydrates, acetate, and isoprene units are relatively easy to identify, and the acquired knowledge of their biosynthesis from primary or secondary metabolism can already suggest, which enzymes provide the building blocks. In general, non-typical cyclization reactions such as described for the cyclization of the pyrronazol-lactone ring (chapter 2), or very rare heterocycles such as the case for the myxolipoxazoles (chapter 5), impedes the direct identification of a biosynthetic origin.

If the biosynthetic building block of a specific scaffold in the natural product is ambiguous, feeding with stable isotope-labeled precursors might reveal the metabolic origin of the building block (**Figure 1**). The major advantage of stable isotope-labeled precursor feeding lies in the relatively simple experimental procedure, which is less dependent from the biological characteristics of the organism. The incorporation of stable isotope-labeled precursors in the final natural product results in a specific *m/z* shift in the isotopic pattern of the natural product, which can be observed by HPLC-MS measurement. From the incorporated building blocks, the genes encoding the enzymes involved in the biosynthesis can be searched for in the genome sequence.

Hence, the next step involves the usage of automated *in silico* annotation tools, such as antiSMASH 5.0<sup>8</sup> to narrow down the BGC (**Figure 1**). Manual comparison of the structural scaffold displayed by the natural product and the *in silico* identified BGCs, can already lead to a reliable correlation of the biosynthetic pathway to the produced natural product. However, prolific producers of natural products such as the members of the order of *Myxococcales* harbor usually numerous BGCs that have not been

correlated to any natural product in the secondary metabolome yet<sup>9</sup>, and this holds true for even well investigated members such as *M. xanthus* DK1622<sup>10</sup>. For this reason, antiSMASH often helps to narrow down the putative BGCs responsible for the biosynthesis of a specific metabolite. Several studies investigating the biosynthesis of myxobacterial natural products have provided a reliable biosynthetic model mostly based on *in silico* identification of the genetic locus and subsequent investigation of the biosynthetic architecture<sup>1,11–14</sup>. In order to provide some experimental support for the proposed biosynthetic pathway, stable isotope-labeled precursor feeding turned out to be a widely used approach for *in silico* characterization of biosynthesis<sup>12,15,16</sup>. Interestingly, this approach is a combination of one of the oldest methods to investigate the biosynthesis of natural products with the recent development of affordable whole-genome sequencing. The biosynthesis of numerous myxobacterial natural products such as pyxipyrrolones<sup>11</sup>, chloromyxamides<sup>13</sup>, jahnellamides<sup>14</sup>, cystobactamids<sup>16</sup>, pyrroazols<sup>1</sup> (chapter 2) and sorangiadenosines<sup>3</sup> (chapter 4) have been identified and investigated by this combined approach. The finding that many myxobacteria are to date intractable to genetic manipulation tools – a major bottleneck for experimental biosynthetic investigations – underlines the indispensable value of stable isotope-labeled precursor feeding combined with *in silico* investigation.

The identification of several strains producing the same natural products has indispensable advantages to attain the comprehensive biosynthetic logic. If the alternative producer is phylogenetically related to the investigated producer, it is probable, that both producers share high similarity of the genome sequence. This finding impedes the reliable identification of their biosynthesis, since most BGCs present in each strain are very similar or even identical to each other. On the other hand, a phylogenetically distantly related alternative producer increases the likelihood that the BGC of interest features the only similarity between both producers. In addition, in phylogenetically distantly related producers of a specific natural product, the surrounding genetic region might be different to each other, which assist in predicting the BGC borders with high accuracy and precision. The prediction of the borders of BGCs is still a difficult challenge for automated *in silico* annotation tools<sup>17</sup>, but nevertheless essential for further genetic processes such as the successful heterologous production of the identified natural product. The advantages underline the benefit of the analysis of more than one producer strain to understand the biosynthesis of the natural product. Another advantage of an alternative producer appears when this producer is amenable for genetic manipulation.

The important connection between the biosynthetic genes and the produced natural products eventually has to be accomplished by genetic manipulation *in vivo* or through biochemical *in vitro* study with recombinantly produced proteins originating from the proposed biosynthetic genes (**Figure 1**). As discussed in chapter 1, genetic disruption or deletion of the biosynthetic genes, can establish the correlation of required genes and therefore enzymes required for the formation of the natural product. Since only one specific gene or the associated gene cluster is investigated concerning the role of the biosynthesis, several independent genetic knockouts/disruptions have to be made to piece together each individual result into a proposed biosynthetic pathway. In addition the studies on the myxobacterial type III PKS system in *M. xanthus* DK1622 proved, that the strategy of overexpression of putative silent BGCs might be more efficient in finding unassigned metabolites rather than performing genetic knockouts<sup>5,18</sup>. Nevertheless, the activation of silent BGCs requires reliable functional genetic tools such as the constitutive promoter P<sub>nptIII</sub><sup>19</sup> (often referred in the literature as P<sub>tn5</sub> promoter) or the vanillate-inducible promoter P<sub>van</sub><sup>20</sup>. Both promoters are employed successfully for the activation (in the native host) and heterologous expression of several BGCs in myxobacteria<sup>21</sup>.

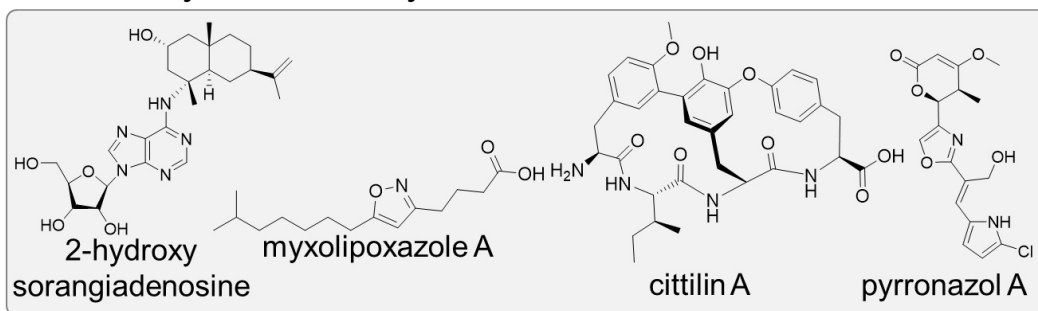
In a scenario where none of the producer strains is accessible by genetic manipulation, a culture-independent genetic correlation through heterologous expression of all biosynthetic genes can become the method of choice. There are several obstacles associated with this method such as different codon usage, rare tRNAs usage, tightly regulated promoters, toxicity and stability of the produced natural product or insufficient precursor supply from the primary metabolism<sup>21–23</sup>. Some of these obstacles can be addressed by extensive genetic and metabolic engineering of the heterologous host, which can lead to the heterologous production of the natural product of interest, help to ultimately achieve gene cluster correlation and to accomplish the retrobiosynthetic investigation.

As mentioned above, retrobiosynthetic investigation approaches rely fundamentally on the information acquired from the genome sequence of the bacterial producer. In addition, analytical data on alternative producers with available genome sequences can enable or improve the identification of the associated BGC. Retrobiosynthetic approaches can also be considered as retrospective genome mining. This approach starts from microbial fermentation, followed by the isolation of the natural product, and in retrospective manner, the genome sequence is investigated to find the associated BGC.

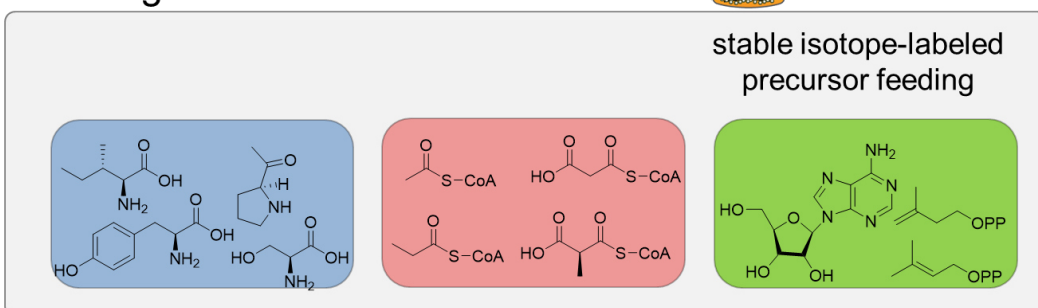
The prospective genome mining approach starts from the investigation of the genome sequence to identify BGCs of interest and afterwards cultivation of the wild type or genetically engineered strain to yield the natural product<sup>24</sup>.



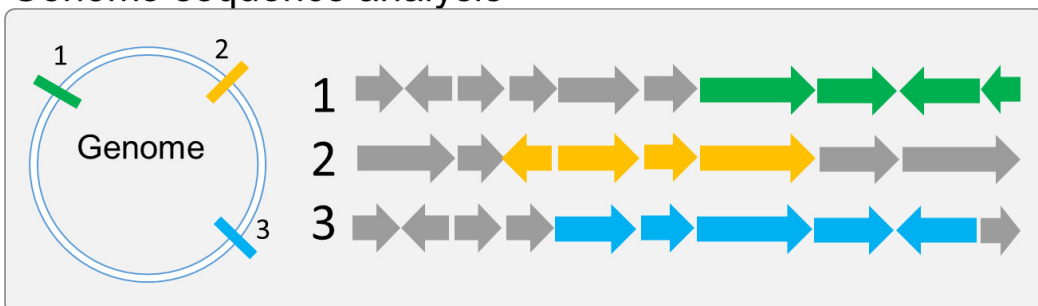
## Retrobiosynthetic analysis



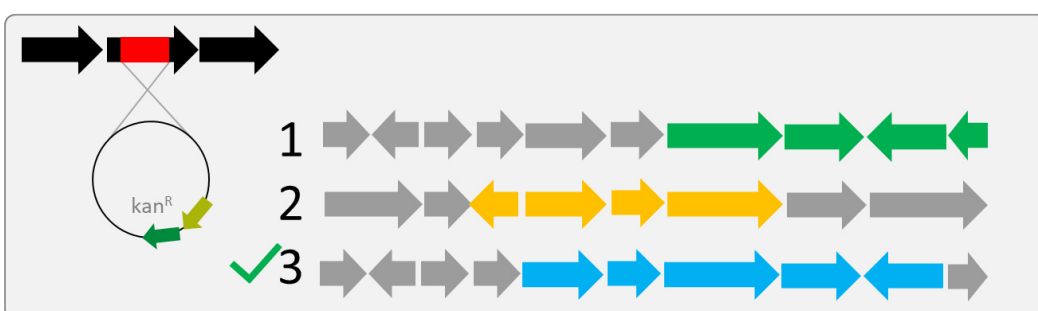
## Building block identification



## Genome sequence analysis



## Genetic confirmation

**Figure 1.** Retrobiosynthetic workflow to identify the biosynthetic genes responsible for the formation of the natural products.

### 7.2.2 The Pleasures and Pitfalls of prospective Genome Mining

The availability of affordable whole-genome sequencing combined with automated biosynthetic *in silico* prediction tools and the development of diverse genetic tools, paved the way for a paradigm shift in natural product research. The challenge is now to find a way to prioritize *in silico* the most promising BGCs in terms of the production of structurally novel or bioactive natural products<sup>25</sup>. A chemocentric screening approach relies on finding novel derivatives of known classes of bioactive natural products. Chemocentric screening is based on finding well-characterized BGCs of known bioactive natural products – but these newly identified BGCs should feature sufficient dissimilarities that novel derivatives can be expected. A chemocentric screening approach led to the identification of rifamycin-like BGCs with additional tailoring genes leading to the production of the rifamycin congeners kanglemycins, which display a new mode of action and evade existing resistance mechanisms<sup>24,26</sup>. In another study, *in silico* global chemocentric screening of 7395 bacterial genome sequences led to the identification of NRPS enzymes that were predicted to utilize cationic amino acid precursors. Secondary metabolome analysis of the fermentation broth of *Brevibacillus laterosporus* DSM 25, *Brevibacillus laterosporus* ATCC 9141 and *Paenibacillus alvei* DSM 29 then revealed two new NRPS natural products – brevicidine and laterocidine – displaying anti-Gram-negative bioactivity<sup>27</sup>.

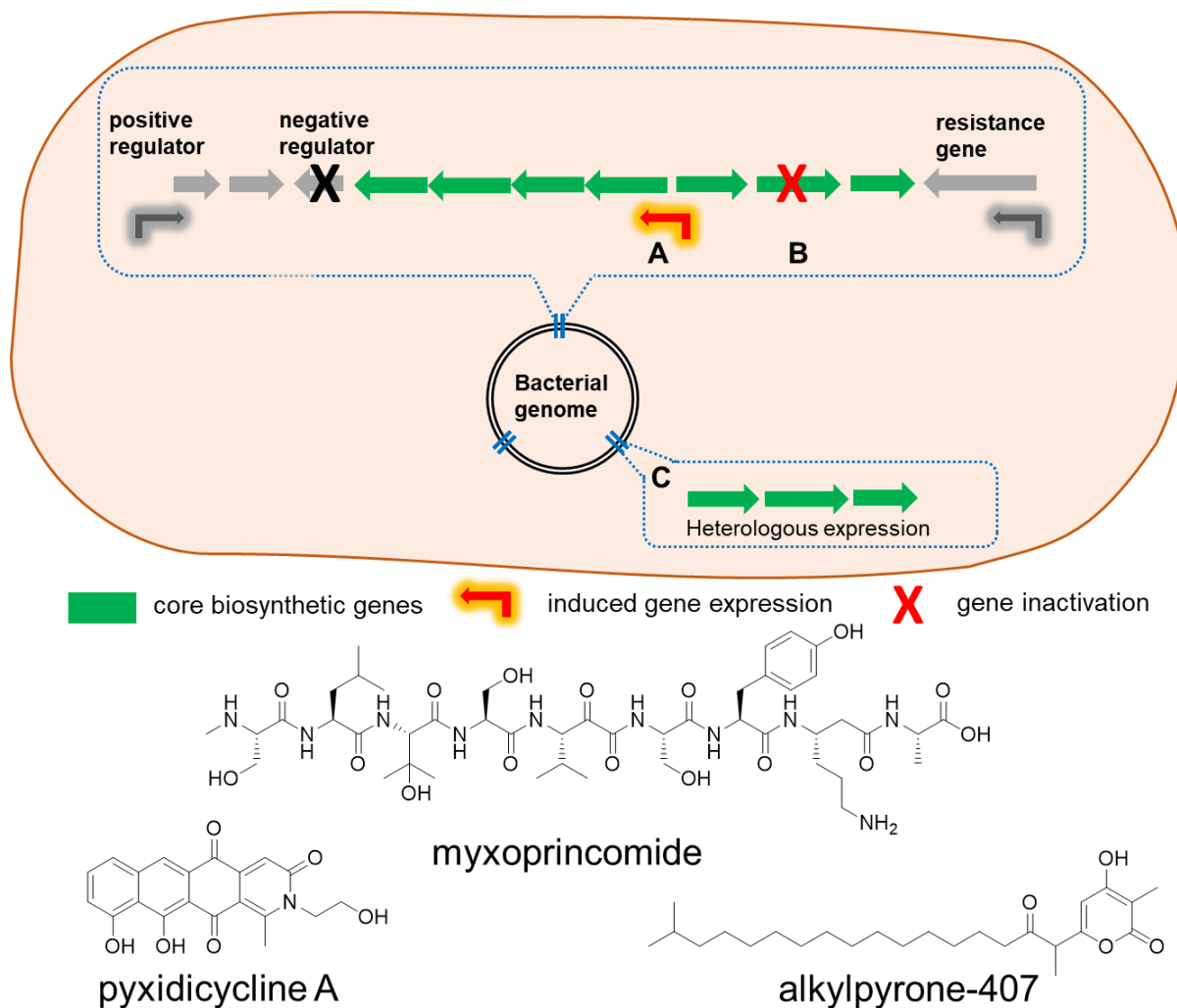
Nevertheless, an important aspect of prospective genome mining approaches relies on the accessibility of genetic tools in the native strains or the efficient transfer of the silent BGC to conduct heterologous expression in a well-characterized host. Genome mining enables to obtain insights into biosynthetic pathways and natural products from uncultured environmental bacteria via the expression of biosynthetic genes originating from *in silico* metagenomic data, as described for the discovery of the calcium-dependent antibiotics malacidin A and B<sup>28</sup>. Prospective genome mining can help to harness silent BGCs to find chemical and biosynthetic novelty but mostly neglects the biological function of the underlying natural product. This ‘chemocentric’ screening mode might provide – due to structural similarity to other BGCs – already some hints to the associated biological function, but reliable prediction of the specific mode of action by this procedure is not possible. For that reason, prioritizing BGCs also encoding factors for host self-resistance for characterization as putative antibiotics is a very promising strategy. As an example, Kling *et al.* reported within an NRPS gene cluster the occurrence of a *dnaN* gene analog in *Streptomyces griseus* – this *dnaN* analog encodes the DNA polymerase sliding clamp for self-resistance to the produced antibiotic griselimycin, which is a novel anti-tuberculosis natural product with an unprecedented mode of action<sup>29</sup>. Griselimycin binds the DNA polymerase sliding clamp, which inhibits the DNA replication; thus, expression of *dnaN* analog confers self-resistance to griselimycin. The observation of duplicated copies of essential housekeeping genes associated with BGCs to provide self-resistance to the host is one of the fundamental criteria for the automated bioinformatics tool “Antibiotic Resistant Target Seeker” (ARTS) to find possible resistant housekeeping genes within BGCs<sup>30</sup>.

Prime examples of successfully conducted self-resistance guided genome mining approaches to access bioactive (myxo)bacterial compounds from under-investigated biosynthetic pathways are the discovery of the alkylpyrones<sup>5</sup> (chapter 6) and the pyxidicyclines<sup>31</sup> (**Figure 2**). The genome mining approach was initiated through a self-resistance guided *in silico* search, focusing on genes encoding pentapeptide repeat proteins (see chapter 6) co-localized with BGCs in myxobacterial genome sequences. This

self-resistance guided genome mining approach was inspired by a previous study reporting the isolation of cystobactamids from *Cystobacter* sp. Cbv34, in which the cystobactamid gene cluster harbors a gene encoding a pentapeptide repeat protein responsible for self-resistance against the gyrase inhibition activity of cystobactamids<sup>16</sup>. Consequently, the presence of this specific pentapeptide repeat protein gene enabled to determine the mode of action of the cystobactamids<sup>16</sup>. The type III PKS biosynthetic machinery for the production of putative topoisomerase inhibitors found in the genome of the myxobacterium MCy9487, culminated in the discovery of the alkylpyrones. Similarly, the conducted *in silico* analysis for genes encoding a functional biosynthetic machinery and co-localized host self-resistance led to the discovery of a presumably unexpressed type II PKS BGC in *Pyxidicoccus fallax* An d48. In both cases the discovery and structure elucidation of pyxidicyclines (type II polyketides), and the alkylpyrones (type III polyketides) required induced gene expression of the associated BGC in the native host or heterologous host. In summary, these studies underline the huge potential of prospective genome mining-based approaches to discover entirely new natural products but also to associate previously uncharacterized BGCs to natural products. The discovery of myxoprincomide is another example of a natural product identified in myxobacteria through prospective genome mining<sup>10</sup> (**Figure 2**).

Since genome mining approaches relies in great measures on previously established knowledge of the biosynthetic logic of biosynthetic pathways, a certain degree of unpredictability concerning the functionality and novelty of the BGCs exists. It is worth mentioning that there is a huge discrepancy between the automated identification of certain classes of BGCs and the interpretation of the underlying biosynthesis. In that sense the identification, predictability and the comparison of genes encoding biosynthetic proteins with known biosynthetic machineries are more reliable for modular biosynthetic pathways, due to their collinear biosynthetic logic. For modular biosynthetic pathways, *in silico* analyses provide already profound information of the proposed biosynthetic logic and associated natural product. In contrast to these modular biosynthetic pathways, in iteratively working biosynthetic systems it is more challenging to interpret accurately the underlying biosynthesis. In other words, the probability to rediscover known natural products via prospective genome mining based approaches is higher for iteratively working biosynthetic enzymes such as iterative type I PKS, type II and III PKSs, and terpene cyclases, than for natural products formed by modular biosynthetic pathways such as NRPS, PKS and hybrids thereof. For that reason, revelation and rediscovery of the alkylpyrones from *M. xanthus* DK1622 was not entirely unexpected, since sequence similarity and *in vitro* experiments from previous studies of type III PKS BGCs point towards this outcome<sup>32,33</sup>. Nevertheless, this rediscovery is not only connecting the alkylpyrones as structural lead for potential topoisomerase inhibitors but also revealed for the first time a myxobacterial type III PKS to its actual *in vivo* generated natural product<sup>5</sup>. The presence of conserved residues within the nucleotide (amino acid) sequences of iteratively working biosynthetic pathways like in the sequences of type III PKSs and terpene cyclases enables fast and reliable identification. In contrast, an accurate prediction of the natural products formed by the identified iteratively working enzyme is often not possible. *In silico* characterization of bacterial type III PKSs and terpene cyclases are severely impeded by the fact that both classes share low overall sequence similarity to each other<sup>5,34</sup>. This means that newly identified type III PKS and terpene cyclase genes are difficult to compare *in silico* with previously genetically or biochemically characterized relatives across phylogenetic kingdoms. The *in silico* identification and proposed biosynthesis of 2-hydroxysorangiadenosine (chapter 4) was possible, due to the requirement of specific tailoring modifications to yield a hydroxylated and adenosinylated sesquiterpene. A simple *in silico* identification of terpene cyclase alone would have not led to the elucidation of the biosynthetic pathway of this unique class of natural product.

## Prospective genome mining in myxobacteria



**Figure 2.** Prospective genome mining. Generic scheme of a biosynthetic gene cluster (BGC) and concepts and methods to conduct prospective genome mining is shown. Examples of natural products identified in myxobacteria through prospective genome mining approaches are shown below. Genetic disruption of core biosynthetic gene led to the discovery of myxoprincomide, while activating silent biosynthetic genes, combined with heterologous expression enabled the discovery of the pyxidicyclines and alkylpyrones.

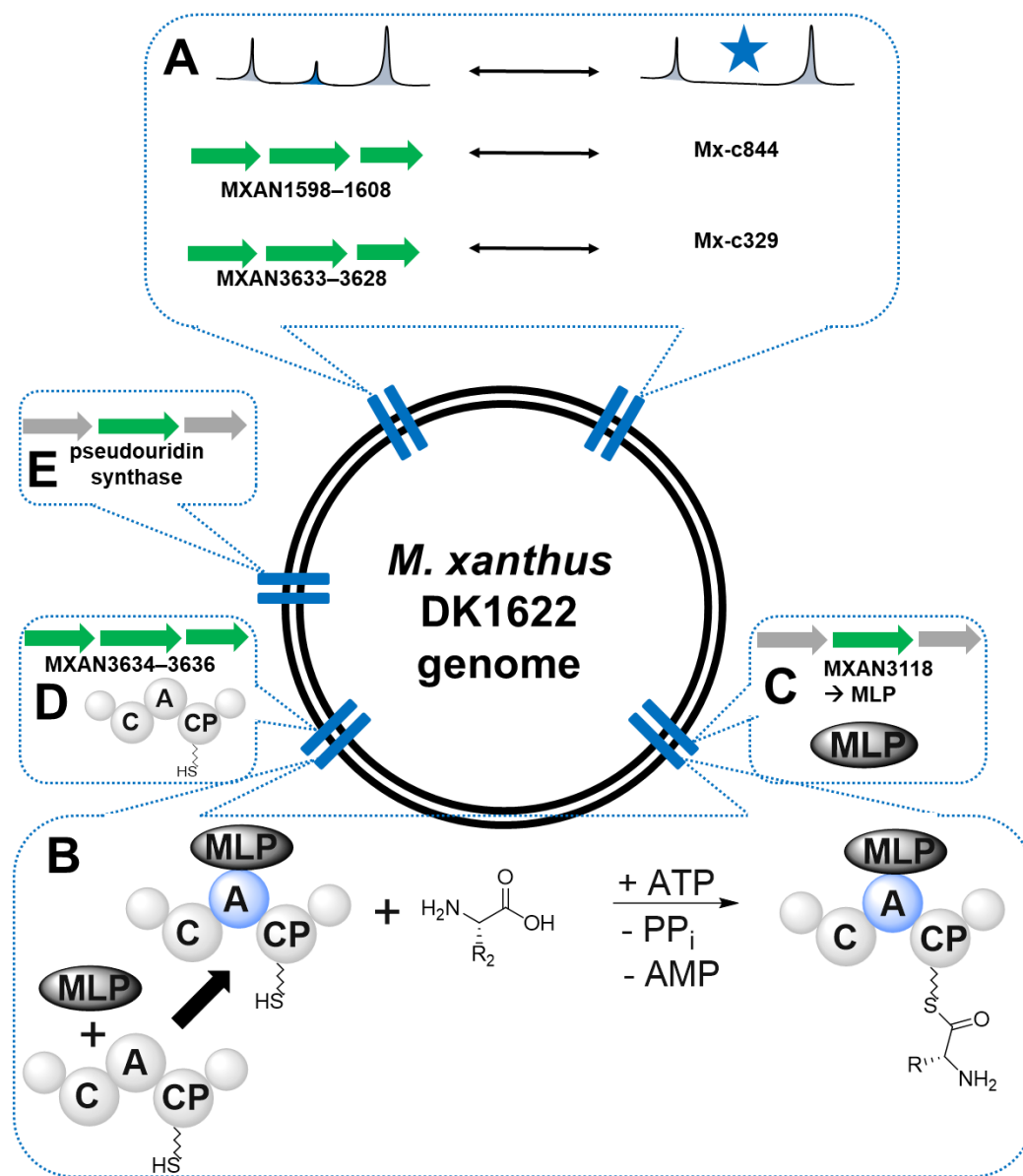
In conclusion, these bioinformatics obstacles can be addressed by structural studies of the specific bacterial type III PKS or the terpene cyclase, ideally in combination with *in vitro* experiments. However, a recent study of Dickschat *et al.* showed the discrepancy of an *in vitro* reconstituted terpene biosynthesis, producing a terpene skeleton different from the one found in the native or in the heterologous host<sup>35</sup>. In addition, even the choice of the heterologous host in which the bacterial terpene cyclase gene was expressed can lead to the formation of different terpene skeletons<sup>36</sup>. The reasons for this observation remain in large parts elusive – some subtle differences in folding the terpene backbone caused by the reaction environment could account for this variation. Like the commonly observed discrepancy of *in vitro* and natively or heterologously produced terpenes, the outcome of the myxobacterial type III polyketides displays a similar observation. The *in vivo* induced gene expression of the type III PKS in *M. xanthus* DK1622, led to the production of the authentic type III PKS compound, the alkylpyrones, that significantly differs from the *in vitro* and heterologously produced

natural products, mainly alkylresorcinols and alkylquinones, respectively. These differences in the produced natural products underline the importance of precursor supply, regulation mechanisms of type III PKSs and the resulting cyclization pattern. The different incorporation of the extender units malonyl-CoA and methylmalonyl-CoA leads to the structure of the alkylpyrones, alkylresorcinols or alkylquinones. Reiter *et al.* investigated the “*Entotheonella*” PKS (*cep*) cluster, encoding a type III PKS CepA and the putative methyltransferase CepB, via *in vitro* enzyme characterization of recombinant CepA<sup>37</sup>. The *in vitro* produced type III polyketides, mainly phenolic lipids, have not been detected in theonellid sponges or in *Theonella* crude extracts, either directly or by molecular network analysis<sup>38</sup>. This discrepancy parallels the findings of the alkylpyrones in this work; the produced authentic compound and the type III polyketides found in the *in vitro* setting do not share the same structure. Another reason could simply be the low abundance of phenolic lipids in the crude extracts of *Theonella* (in contrast to crude extracts of *Haliclona sponges*<sup>39</sup>) which underlines the importance of the analytical set-up<sup>37</sup>. Several studies have shown the differences in signal intensities of MS-based analytical set-ups of specific groups of natural products, implying that a BGC which is considered to be silent might actively generate an associated natural product (possibly in low abundance), which might not have been detected by a previously used analytical set-up.

The discrepancy between genome-inscribed capacity for the production of natural products and the set of known secondary metabolite classes detected from many microorganisms under laboratory cultivation conditions is astounding even for well characterized strains such as *M. xanthus* DK1622. This so-called myxobacterial model host is a prolific producer of several well-investigated natural products, and still features a huge untapped potential for the production of “hidden” secondary metabolites. In this thesis, the biosynthetic logic of a neglected, but commonly found natural product within the species of *M. xanthus* namely the cittelins, was presented in chapter 3. Furthermore, the alkylpyrones as secondary metabolites in the metabolome of *M. xanthus* DK1622 have been revealed by a prospective genome mining approach. In addition, bioinformatics analysis of the genome sequence of *M. xanthus* DK1622 reveals, that numerous predicted biosynthetic genetic loci have still not been characterized or connected to secondary metabolites. In particular, the presence of six antiSMASH-annotated RiPP biosynthetic pathways – two lanthipeptide BGCs (MXAN2852–2857 and MXAN6388 + MXAN6387), three bacteriocin BGCs (MXAN3552–3559, MXAN4156–4165 and MXAN4951–4959) and one thiopeptide BGC (MXAN4931–4945) – opens the way for further exploration of RiPPs originating from myxobacteria.

*In silico* investigation emphasizes the huge potential for further genome mining in *M. xanthus* DK1622 next to experimental studies originating from various disciplines in natural product research. The following three studies connected directly or indirectly to the myxobacterial model host *M. xanthus* DK1622 showcase potential starting points for further prospective genome mining approaches. The first study which aimed to explore the hidden biosynthetic potential of *M. xanthus* DK1622 revealed a new NRPS-PKS hybrid natural product class termed myxoprincomide. The discovery and isolation of myxoprincomide was enabled by site-specific single crossover disruption of every assigned BGC combined with applied PCA<sup>10</sup>. In addition, this approach also led to the identification of two new classes of natural products. These two still structurally unassigned natural products are associated with the BGCs MXAN1598–1608 (compound Mx-c844) and MXAN3633–3628 (compound Mx-329) (**Figure 3A**). In order to obtain the chemical structure and elucidate the underlying biosynthesis of these two natural products, site-specific promoter insertion to increase the production or classical production optimization of the host strain or alternative producers could be conducted.

## Genome mining potential in *M. xanthus* DK1622



**Figure 3.** Prospective genome mining outlook for the model strain *M. xanthus* DK1622 (explanation in main text).

In contrast to the first example another study investigated the genome of *M. xanthus* DK1622 to shed light on a particular detail in the biosynthetic logic of NRPS. It has been shown that some A domains of NRPS systems interact with small proteins named after the MbtH protein from the *Mycobacterium tuberculosis* mycobactin operon<sup>40</sup>. MbtH-like proteins (MLPs) can induce the activation of acyl-adenylate formation<sup>41,42</sup>, by increasing the solubility and activity of recombinant A domains *in vitro*<sup>43</sup> (**Figure 3B**). The important role of MLPs for some A domains and the dynamic characteristics of NRPS modules, including the module core formed by the A and C domains as well as the orientation of the mobile TE domain<sup>44</sup>, was elucidated via the obtained crystal structure of the full-length NRPS EntF (comprised of a C, A, PCP and TE domain) bound to the MLPs from *Escherichia coli* and *Pseudomonas aeruginosa*. The authors of this study revealed, that MLP is binding to the A

domain in EntF during its entire reaction cycle but is not altering the structure of the A domain, whereas the TE domain of EntF changes its conformations and locations to adopt a different orientation loosely based on the location of the PCP. In addition, MLP binding induces changes to the orientation of the C–A didomain interface of EntF and conformational changes at the PCP binding site within the C domain. These findings underline the dynamic architecture of NRPS modules and that the biosynthetic function of MLPs is probably not limited to bind A domains<sup>44</sup>. Esquilín-Lebrón *et al.*<sup>45</sup> made the observation that only one gene in *M. xanthus* DK1622 (MXAN3118, **Figure 3C**) encodes an orphan MbtH-like protein (MLP), in contrast to other bacterial strains, which encode separate MLPs for different biosynthetic gene clusters<sup>46</sup>. This MLP was investigated with respect to the interaction with several NRPS gene clusters. Nine NRPS gene clusters displayed interaction with the MLP according to bacterial two-hybrid screening and seven of those interactions were confirmed by solubility and co-purification assays<sup>45</sup>. The outcome of this study suggests that the BGC (MXAN3634–3636, **Figure 3D**) encodes a NRPS pathway that possibly produces a natural product that inhibits the clumping of *M. xanthus* DK1622 in liquid culture.

Deletion of the MLP gene or the BGC (MXAN3634–3636) led to aggregation of *M. xanthus* DK1622. The phenotypical difference in *M. xanthus* DK1622 wild type and its deletion mutant ( $\Delta$ MXAN3634–3636) implies that the production of a natural product was abolished. However, the authors were not capable to correlate any natural product in the secondary metabolome of *M. xanthus* DK1622 to the BGC (MXAN3634–3636), presumably due to the low production rate of the hypothetical natural product in the wild type. In order to identify the putative natural product associated with clumping of *M. xanthus* DK1622 in minimal medium and to obtain subsequently sufficient material for structure elucidation and investigate the underlying biosynthesis of this natural product, induced gene expression of the BGC (MXAN3634–3636) in the native host or heterologous expression of the BGC could be performed.

The identification of the pseudouridimycin BGC in the producer strain *Streptomyces* sp. ID38640, and the associated bioinformatics investigation of the essential pseudouridine synthase gene in the publicly available genome database of myxobacterial strains, point towards the presence of a putative nucleoside BGC in several myxobacterial strains that might produce similar pseudouridine-containing nucleoside antibiotics<sup>47</sup> (**Figure 3E**). This phylogenomic analysis encourages further prospective genome mining approaches to uncover the hidden metabolome of myxobacterial strains but also the secondary metabolome of *M. xanthus* DK1622.

In summary, prospective genome mining of myxobacterial strains will enrich the understanding of the biosynthetic logics of natural products despite the drawbacks of possible rediscovery of known natural products. In combination with classical retrospective genome mining procedures, genome mining could establish itself as an important cornerstone of myxobacterial natural product research.

Nevertheless, structural and biological novelty of natural products and unprecedented insights into the biosynthetic logic of myxobacteria equally arises from classical natural product isolation processes.

### 7.3 Biosynthetic space of myxobacteria – Beyond PKS and NRPS

The biosynthetic space of myxobacteria is occupied by numerous natural products produced by large multimodular enzyme complexes such as the PKS, NRPS and PKS/NRPS hybrids. Many of these biosynthetic machineries afforded myxobacterial natural products with impressive biological activities such as the epothilones, aetheramides and nannocystins<sup>48,49</sup>. A specific characteristic of myxobacterial biosynthetic space is the prolific generation of hybrid PKS/NRPS natural products fusing amino acids and carboxylic acids to enhance the structural diversity afforded by these multimodular biosynthetic machineries. For example, the heterocyclic oxazole moiety in pyrrolozoles (chapter 2) feature a common chemical scaffold in myxobacterial natural products<sup>50–52</sup>, which is generated via a cyclization (Cy) domain to fuse an amino acid and acetate-derived building block to a heteroaromatic five-ring.

A brief overview considering the unexplored biosynthetic space within the order of *Myxococcales* can be given by simply viewing at all automatically identified BGCs in our in-house and the publicly available databases. This manual overview reveals that many of the uncharacterized BGCs within the order of *Myxococcales* are classified as RiPPs, hinting at a huge unrecognized potential in myxobacteria (such as mentioned above for *M. xanthus* DK 1622). It is worth mentioning that these uncharacterized BGCs reflect only RiPPs pathways, which can be recognized by antiSMASH. Many RiPP biosynthetic pathways are not detectable via automated annotation tools<sup>53</sup>. Instead, the biosynthetic logic of the respective RiPP has to be provided manually such as for the citilins and crocagins. Since the only two characterized myxobacterial RiPP pathways are not classified into one of the large families such as lanthipeptides, head-to-tail cyclized peptides, lassopeptides and thiopeptides<sup>54</sup> it might be possible there are more novel myxobacterial RiPP families. Another notable characteristic of myxobacterial biosynthetic space is the absence of any BGC associated with production of saccharide natural products<sup>55–57</sup> (one of the seven structure-based classes according MIBiG<sup>58</sup>) including the aminoglycoside and aminocyclitol subtypes<sup>59–61</sup>. Consequently no saccharide natural product (probably due to prioritization and targeted isolation of less polar natural products during the common screening workflow<sup>62</sup>) and only few glycosylated derivatives of myxobacterial natural products have been isolated such as the chivosazols<sup>51</sup>, the cystomanamides<sup>63</sup>, disciformycin<sup>12</sup>, icumazols and noricumazols<sup>64</sup>. Recently, heterologous expression of the myxobacterial acylhomoserine lactone synthase *agpI* from *Archangium gephyra* in *E. coli* led to the production of three acylhomoserine lactones, of which one seems to be novel<sup>65</sup>. This study supports the finding that quorum signals can be sensed and addressed by *M. xanthus*<sup>66</sup>, which is probably connected to the presence of butyrolactone and homoserine lactone gene clusters within the genome sequences of myxobacteria. These BGCs are typically connected with the production of quorum sensing molecules produced by numerous *Proteobacteria* sp. and *Streptomyces* sp.<sup>67–70</sup>. In summary, the genomic potential of myxobacterial producers exceeds the number of currently known biosynthetic machineries<sup>9,71</sup>. Furthermore, the biosynthetic space of myxobacteria is most likely more diverse than suggested by the reported case studies of heterologously produced myxobacterial secondary metabolites in *M. xanthus* DK1622<sup>21</sup> or the recent summary of bioactive natural products from myxobacteria<sup>48</sup>. The myxobacterial natural products in this thesis are corroborating this assumption, since these are biosynthesized from different biosynthetic machineries such as modular PKS-NRPS hybrid machineries, the ribosome, an iteratively working type III PKS and a terpene cyclase. In the following subsections, the underexplored biosynthetic space of myxobacterial natural products is highlighted based on the case studies reported in the chapters 3–6.



### 7.3.1 The Cittilins – RiPPs in Myxobacteria

The production of the cittilins is conserved in numerous myxobacterial strains belonging to the genus *Myxococcus*. For this reason cittilin A was also previously investigated in an intraspecific diversity study of many *M. xanthus* strains and in a study combining LC-MS analysis with subsequently performed statistics to investigate the secondary metabolome and the degree of metabolic diversity within a single species<sup>72,73</sup>. Despite the knowledge of the chemical structure and analytical characteristics, little was known about the biosynthesis of cittilins.

Since retrobiosynthetic analysis did not identify suitable NRPS genes to assemble the cittilin tetrapeptide YIYY in the genome of *M. xanthus* DK1622, the sequence was subsequently screened for genes that directly encode this peptide sequence, resulting in the identification of the 27 amino acid long precursor peptide CitA and the cittilin operon (doctoral thesis: Ole Revermann<sup>74</sup>). The experiments conducted in chapter 3 proved that the identified biosynthetic locus is associated with the biosynthesis of the cittilins. Similar to the previously published myxobacterial RiPPs – the crocagins<sup>75</sup> – the cittilins were discovered by conventional metabolite-guided approaches and not through prospective genome mining. The work conducted in chapter 3, enables from now on, *in silico* detection of cittilin-like BGCs in bacterial genomes (due to the deposited cittilin BGC in the MIBiG database).

A CYP450-type enzyme termed CitB plays a central role in the biosynthesis of this novel RiPP pathway. It catalyzes the ring formation of the precursor peptide to yield after proteolytic cleavage cittilin B. The functionality of CitB within the heterologously expressed BGC was confirmed in the distantly related host *Streptomyces albus* del14<sup>76</sup>. In addition, the conducted cell-free lysate reactions in the heterologous host *Streptomyces coelicolor* CH999, supports the proposed biosynthetic pathway and set the stage for future in-depth biochemical analysis.

Catalytically active recombinant CitB would be required for further biochemical investigation of this catalyzed cyclization in a purely *in vitro* setting. Furthermore, a functional redox partner for the CYP450 enzyme CitB has yet to be found to regenerate the active site after the first catalytic turnover. After obtaining functional CitB with an appropriate redox system, further *in vitro* assays can then be performed in order to test the activity, specificity and its dependency on the amino acid sequence of different cittilin precursor peptides. *In vitro* reconstitution of the cittilin biosynthesis through all three enzymes, namely the prolyl endopeptidase MX PEP, the CYP450 enzyme CitB and the methyltransferase CitC would further confirm the cittilin formation and would allow finalizing the order of enzymatic transformations.

The observations from the cell-free lysate reactions already hint towards the catalyzed biochemical reaction of the CYP450 enzyme CitB. However, the inherent complex biological matrix in which these catalytic conversions take place might hinder to reveal the exact chemical mechanism of the performed cyclization. The occurrence of radical intermediates and electron transfer to the heme co-factor in CYP450 enzyme systems such as described for the CYP450 enzyme OxyC<sup>77</sup> during the biosynthesis of vancomycin, limits severely *in vivo* biochemical reaction investigations.

The bicyclic structure of the citilins shares some similarity to the cross-linked structures of the glycopeptide antibiotics vancomycin, teicoplanin and kistamicin, but the formation of the bicyclic ring in the citilin scaffold apparently requires only the enzymatic activity of one CYP450 enzyme (**Figure 4A**). In contrast, vancomycin biosynthesis studies have shown that the phenol-coupling is catalyzed by two CYP450 enzymes (OxyA and OxyB)<sup>78</sup>, while the C–C bond formation is catalyzed by a third CYP450 enzyme (OxyC, **Figure 4B**)<sup>77</sup>. The cyclization cascade catalyzed by CitB could either be initiated via a phenol-coupling reaction, resembling the described mechanism for OxyA and OxyB or via a biaryl ether linkage such as described for OxyC. The second cyclization could occur concurrently with the first cyclization (due to increased chemical reactivity) or by a separate second catalytic cyclization. The exact mechanisms for these cyclizations are very challenging and might involve yet unknown protein–protein interactions or additional binding partners, which would be an explanation for functional CitB during the heterologous production of citilin A or citilin B (in the cell-free lysate reactions).

For example, the sequential CYP450 enzyme-catalyzed *in vitro* oxidative cross-linking of the glycopeptide antibiotics vancomycin and teicoplanin relies strictly on the recruitment of the PCP-X di-domain on the terminal NRPS module (**Figure 4B**)<sup>79,80</sup>. Consequently, the oxidative cross-linking during vancomycin biosynthesis requires the X-domain (**X**) on the terminal NRPS module 7 to recruit PCP-bound substrate and the OxyA, OxyB and OxyC CYP450 enzymes<sup>79,80</sup>. The oxidative cross-linking during the vancomycin biosynthesis involves OxyB-catalyzed phenol-coupling, followed by halogenation via VhaA at the  $\beta$ -OH-Tyr residues, and subsequently the second phenol-coupling takes place via the CYP450 enzyme OxyA. The biaryl C–C bond formation catalyzed by OxyC is the last oxidative cross-linking reaction in the biosynthesis of vancomycin. The cyclized peptide is released by the TE domain, glycosylated (by GtfA and B) and undergoes MtfA-catalyzed *N*-methylation<sup>78</sup> to finally yield vancomycin<sup>78</sup>.

Other enzymes are also known to catalyze C–C bond formation during RiPP biosynthesis. Radical SAM (rSAM)-catalyzed cyclization<sup>81</sup> between L-lysine and L-tryptophan has been shown for streptide, a streptococcal macrocyclic peptide<sup>82</sup>, and proposed for the linkage in darobactin<sup>83</sup>. Paralleling the cyclization of citilin CYP450 enzyme, the DarE rSAM in the biosynthesis of darobactin thought to proceed via the formation of the Trp–Lys C–C bond and the C–O–C Trp–Trp ether bond (**Figure 4C**). However, the exact biochemical mechanism of this rSAM remains elusive. Furthermore, the respective protease to cleave the leader peptide has not been identified yet<sup>83</sup>. In contrast, the citilin CYP450 enzyme already features great potential for further investigation and advancing the catalytic tools for enzymatic conversions. The very small set of enzymes might represent the least complex biosynthetic system to generate bicyclic tetrapeptides.

The *in silico* identification of two potential citilin derivatives is an interesting finding towards prospective genome mining of RiPPs in myxobacteria. The genome sequence of *Myxococcus hansupus mixupus* features the same biosynthetic organization of all described citilins except for the nucleotide sequence encoding the core peptide YHYY. In contrast, the genome sequence of *Streptomyces* sp. Ncost-T10-10d features no citilin-specific methyltransferase, unlike the myxobacterial gene cluster organization. Therefore, the core peptide with the terminal amino acid sequence YSY, would lead to a non-methylated citilin derivative. Since the strains and the associated metabolomic data are not available, it was unfortunately not possible to search for these putative derivatives.

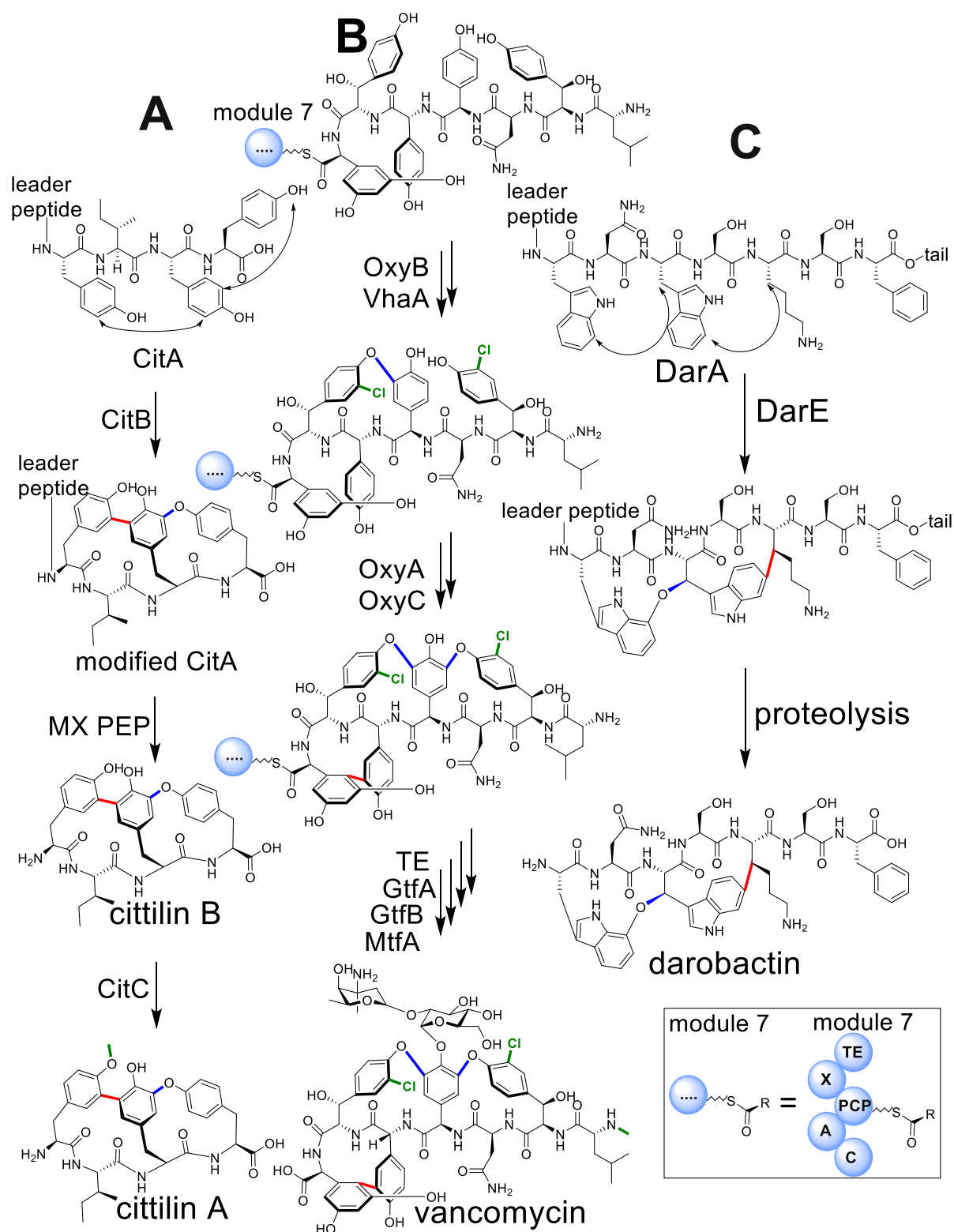
The heterologous production platform in *S. albus* del14 could be used to generate numerous different cittilin derivatives including the genomically identified histidine and serine derivatives. As this heterologous platform featured a significantly lower production rate than the native myxobacterial host, an alternative approach could also involve genetic engineering of *M. xanthus* DK1622. A simple insertion of a vanillate-inducible promoter system in front of the identified precursor peptide by single crossover recombination could be harnessed to produce cittilin derivatives.

This method would – as described in chapter 3 – use the genetic homology region, which comprises the gene encoding the precursor peptide including a part of the downstream CYP450 enzyme gene except for a site-specific mutation in the core peptide-encoding region. As single crossover recombination inserts the whole genetic construct into the genome, this method can be considered as a “quick and dirty” approach in contrast to the time-consuming double crossover-based manipulation procedures mentioned earlier for *M. xanthus* DK1622 (see chapter 1).

The *in silico* identification of only two different cittilin derivatives and the widespread occurrence of cittilins in the genus of *Myxococcus* raises the question of its biological function. The observed CsrA *in vitro* inhibition through cittilin A might not hint at the actual evolutionary developed function, since the poor cellular uptake of cittilin A seems to render it unsuitable as chemical defense against other bacteria.

The relative potent inhibition of CsrA by cittilin A (IC<sub>50</sub>: 4 µM) and its level of polarity (according the lipophilicity descriptor clogD<sub>pH 7.1</sub>: -0.9) indicate at the first glance towards advantageous physiochemical properties for anti-Gram-negative bioactivity<sup>84</sup>. However, the molecular weight (MW) of cittilin A exceeds the empirically defined cutoff at 600 Da for the average MW of Gram-negative antibacterials<sup>85</sup>, which is also reflected through its low ligand efficiency value of 0.18<sup>84</sup>. For that reason, further investigation of the main biological function of the cittilins is warranted and might reveal further insights into the complex biology of myxobacteria. Initial steps could involve further biological investigation such as fruiting body formation of single crossover or double crossover mutants, deficient in cittilin production.

In conclusion, we learn from the biosynthesis of cittilins, that myxobacterial RiPP gene clusters despite their small size, feature biosynthetic enzymes with unique characteristics and that we likely have only started to tap into the reservoir of myxobacterial RiPPs.



**Figure 4.** Cross-linking pathways in different natural products. **A)** Cyclization of citilin requires the precursor peptide CitA as substrate, cyclization via the CYP450 enzyme CitB and subsequent proteolysis of modified CitA is catalyzed by the prolyl endopeptidase MX PEP. The last maturation step involves specific methylation of citilin B via CitC to yield citilin A<sup>2</sup>. **B)** The oxidative cross-linking of vancomycin requires the X-domain (X) on the terminal NRPS module 7 to recruit PCP-bound substrate and the OxyA, OxyB and OxyC CYP450 enzymes<sup>79,80</sup>. **C)** The precursor peptide DarA is processed by the DarE rSAM in the biosynthesis of darobactin. DarE catalyzes presumably the formation of the C–C bond (red color) and the C–O–C ether bond (blue). The respective protease to cleave the leader peptide has not been identified yet<sup>83</sup>. Green color indicates different modifications.

### 7.3.2 2-Hydroxysorangiadenosine – Terpene Fused with a Nucleoside

The secondary metabolome of myxobacteria is known for its diversity of less complex terpenes, mainly volatiles<sup>86,87</sup>, but also for 2-methylisoborneol<sup>88</sup>, and geosmin<sup>89,90</sup> that accounts for the earthy smell of myxobacterial broth cultures. Additional terpenes from myxobacteria were either discovered through biochemical-guided investigation of myxobacterial terpene cyclases such as described for the (+) eremophilene synthase Sce8552<sup>91</sup> and 10-*epi*-cubebol synthase Sce6369 in *Sorangium cellulosum* So ce56<sup>92</sup>, or by classical bioactivity-guided isolation described for the diterpenes cystodienoic acid<sup>93</sup> and enhygromic acid<sup>94</sup>.

The surprisingly low number of myxobacterial terpene natural products reflects the general under-investigation (in comparison to modular biosynthetic pathways) of bacterial terpenes. The antiSMASH database comprises the DNA-sequence information of more than 4000 bacterial terpene BGCs, whereas only 127 have been investigated and deposited in the MIBiG database<sup>36</sup>. One reason attributing for this neglect is the poor *in silico* predictability and comparison, concerning rational structure prediction, of identified terpene biosynthetic pathways. Like terpene cyclases in plants, biosynthetic investigation of bacterial terpene biosynthesis is equally challenging. The identification of the bacterial terpene cyclase is comparatively simple, whereas the prediction of the associated natural product is in most cases simply not possible. The second reason is the low production titer of bacterial terpenes under standard laboratory conditions. Connected to this, the physicochemical properties of many terpenes – which are often non-polar, lack functional groups and feature poor ionization properties – are severely impeding targeted isolation of terpenes through LC-MS guided screening. These reasons explain why most of the terpenes described in myxobacteria had to feature an interesting chemical scaffold or possess pronounced bioactivity in order to be detected. The leupyrrins are an example of a myxobacterial hybrid natural product featuring PKS, NRPS and terpene-derived building blocks<sup>95</sup>.

The first discovery of sorangiadenosine was also initiated via bioactivity-guided natural product isolation<sup>96</sup>, such as the rediscovery and biosynthetic investigation of the sorangiadenosine and the associated discovery of 2-hydroxysorangiadenosine in chapter 4<sup>3</sup>. In that sense, sorangiadenosine is one of the few myxobacterial terpene natural products, which features not only significant bioactivity against human pathogens, but also a complex terpene-adenosine chemical scaffold (**Figure 5A**). Since the biosynthesis of the sorangiadenosine is more complex than that of regular terpenes, it was possible to correlate the biosynthetic locus of 2-hydroxysorangiadenosine with the afforded natural products. The biosynthesis starts from the basic terpene building blocks, IPP and DMAPP, to generate farnesyl diphosphate that undergoes a type I cyclization to yield the eudesmadiene core structure (**Figure 5B**). Subsequent transfer of adenosine and hydroxylation leads to the formation of 2-hydroxysorangiadenosine.

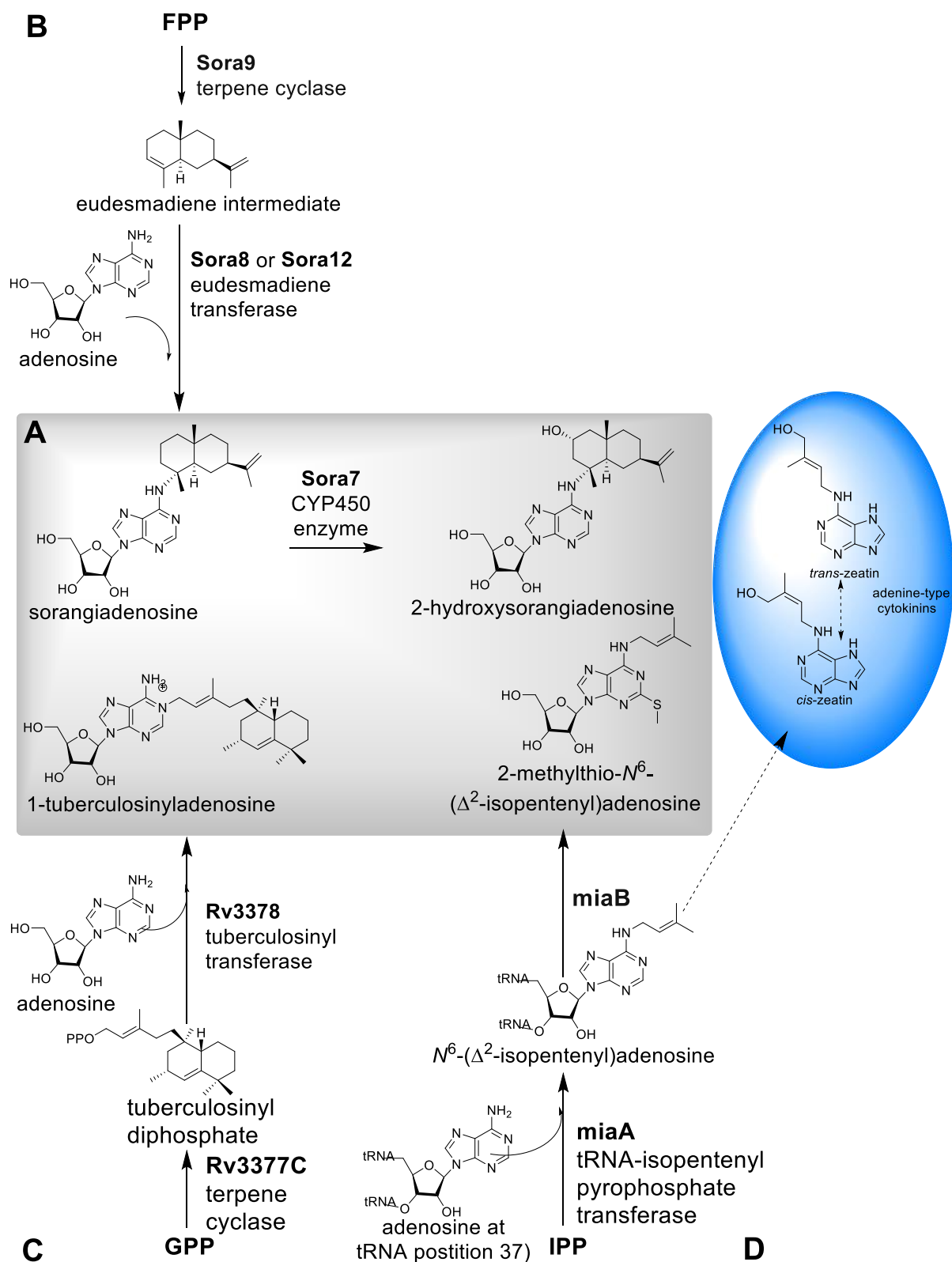
Unlike naturally-occurring nucleotides, nucleoside natural products and terpenes, there are only few biosynthetically investigated terpene-nucleoside/nucleotide hybrid natural products known up-to date, in particular from microorganisms, such as the lipid 1-tuberculosinyladenosine produced by *Mycobacterium tuberculosis*<sup>97</sup>. The biosynthesis of 1-tuberculosinyladenosine starts from geranylgeranyl pyrophosphate (GGPP) to generate tuberculosinyl pyrophosphate catalyzed by the terpene cyclase Rv3377c, whereas the second biosynthetic enzyme Rv3378c acts distinguished from classical *cis*-prenyl transferases via a dual-substrate pocket. Hence, Rv3378c works as tuberculosinyl

transferase, catalyzing the condensation of tuberculosinyl pyrophosphate and adenosine to generate 1-tuberculosinyladenosine<sup>97</sup> (**Figure 5C**). Another hybrid, which features a less complex terpene scaffold is 2-methylthio-*N*<sup>6</sup>-( $\Delta^2$ -isopentenyl)adenosine. The biosynthesis of 2-methylthio-*N*<sup>6</sup>-( $\Delta^2$ -isopentenyl) adenosine in *E. coli* starts from adenosine and DMAPP which are converted to *N*<sup>6</sup>-( $\Delta^2$ -isopentenyl) adenosine by the tRNA prenyltransferase MiaA<sup>98</sup>. Subsequently *N*<sup>6</sup>-( $\Delta^2$ -isopentenyl)adenosine is converted by the methylthiotransferase MiaB<sup>99</sup> (**Figure 5D**).

The connection of other terpene-derived natural products to a biosynthetic locus in bacterial genomes is often not possible by *in silico*-based methods. The strain *Vitiosangium cumulatum* MCy10943<sup>T</sup> appears promising for prospective genome mining to investigate the numerous myxobacterial terpene cyclases located within the genome. The development of a genetic manipulation protocol in MCy10943<sup>T</sup> would be instrumental to connect the remaining uncharacterized terpene cyclases to specific metabolites in the secondary metabolome and to confirm the *in silico* characterized 2-hydroxysorangiadenosine gene cluster through genetic manipulation. Besides the unknown ratio of successfully correlated natural products, this approach however harbors the great risk of rediscovery of known natural products, since it seems nearly impossible to predict the biosynthesized natural product.

The underlying biosynthetic genes could also be heterologously expressed in the well-characterized host *M. xanthus* DK1622 to perform further genetic engineering to significantly increase the production rate of this biologically valuable scaffold or derivatives thereof, such as described for the argyrins<sup>100</sup>. Additional biochemical investigation of recombinant terpene cyclase (Sora9), CYP450 enzyme (Sora7) and the phosphatase-eudesmadiene transferase (Sora8 or Sora12) might provide valuable insights into the biosynthetic formation of sorangiadenosine and 2-hydroxysorangiadenosine.

In conclusion, the strain *V. cumulatum* MCy10943<sup>T 101</sup> is an excellent example for a novel myxobacterium with great potential for classical metabolome and bioactivity guided natural product isolation and prospective genome mining approaches.



**Figure 5.** A) Different terpene-adenosine hybrids are displayed in the grey box in the middle. B) Biosynthetic pathway leading to sorangiadenosine and 2-hydroxysorangiadenosine in *Vitisangium cumulatatum* MCy10943<sup>T</sup>. C) Biosynthesis of 1-tuberculosinyladenosine in *Mycobacterium tuberculosis* and D) Formation of 2-methylthio- $N^6$ -( $\Delta^2$ -isopentenyl)adenosine in *E. coli*.  $N^6$ -( $\Delta^2$ -isopentenyl)adenosine provides the building block for the adenine-type cytokinins *trans*-zeatin and *cis*-zeatin (small blue circle).

### 7.3.3 The Peculiar World of Silent Type III Polyketide Genes in Myxobacteria

The plethora of type III PKS genes present in myxobacterial genomes stood in contrast to matching natural products in the myxobacterial secondary metabolome. The work in chapter 6 elucidated for the first time type III polyketides from *M. xanthus* DK1622 *in vivo* via the activation of a gene cluster. Similar to the field of terpene-based investigation (mentioned above), the only generated type III polyketides from myxobacteria so far originated from heterologous expression of myxobacterial type III PKS genes or *in vitro* production of recombinant myxobacterial type III PKSs<sup>32,102,103</sup>. In addition, the *in silico* survey highlights that myxobacterial strains harbor a promising repertoire for several new type III PKSs likely synthesizing unanticipated natural products (**Figure 6A, B**).

One follow-up strategy could involve the usage of the same overexpression platform established in chapter 6 to overexpress representatives of diverse type III PKS BGC-architectures found in myxobacterial genomes (**Figure 6B**). This heterologous expression approach could potentially suffer from the experimentally observed discrepancy (see above and chapter 6) between the authentic metabolites produced in the native strain and the metabolites found in heterologous expression constructs of the biosynthetic genes. *In vitro* conducted catalytic reactions of specific representative myxobacterial type III PKSs would have the advantage to provide detailed insights into the substrate specificity of the enzyme, but this approach could also fail to mimic the authentic catalytic environment and consequently to elucidate the actual natural products (**Figure 6 B**). An analytical-driven approach would aim to identify the secondary metabolites connected to the respective type III PKS gene clusters by comparative statistical analysis of several strains harboring the same groups of type III PKS gene clusters (according to the groups in chapter 6). Difficulties with this approach would arise from the less specific fragmentation patterns of type III polyketides *e.g.* in comparison to those of peptide-based natural products. Another challenge is the low abundance of type III polyketides in the secondary metabolome of myxobacteria, since the expression of type III PKSs seems to be downregulated among myxobacteria<sup>18,102</sup>. The fermentation of many myxobacterial strains in different media could help to overcome this obstacle, since one of these strains might be under specific conditions a decent producer of type III polyketides. From that point of view, the biological reason for the abundance of type III PKS BGCs in myxobacteria might be interesting.

One of the most promising approaches seems to be the induction of gene expression of these downregulated BGCs in the native hosts (**Figure 6A**); however, it might be challenging to develop appropriate genetic manipulation procedures for several myxobacteria. In addition, induced gene expression might not lead to the production of authentic natural products from the respective microorganism, but to increased production of biosynthetic shunt metabolites as shown during the gene activation strategy in *P. fallax* An d48<sup>31</sup>. Induced gene expression of the biosynthetic operon *pcyA–I* within the *in silico* identified and unexpressed type II PKS BGC in *P. fallax* An d48, led to the production of polyketide synthesis derailment products<sup>104</sup> such as anthraquinone 340<sup>31</sup>. It was necessary to use two promoters, in order to induce simultaneously the biosynthetic operons *pcyA–I* and *pcyJ–P* to yield the authentic natural products pyxidicycline A and B. Furthermore, it was essential that one of the applied promoters had to be inducible, since initial genetic manipulation, relying only on constitutive promoters was yielding no transformants. This finding hints at significant auto-toxicity challenges,

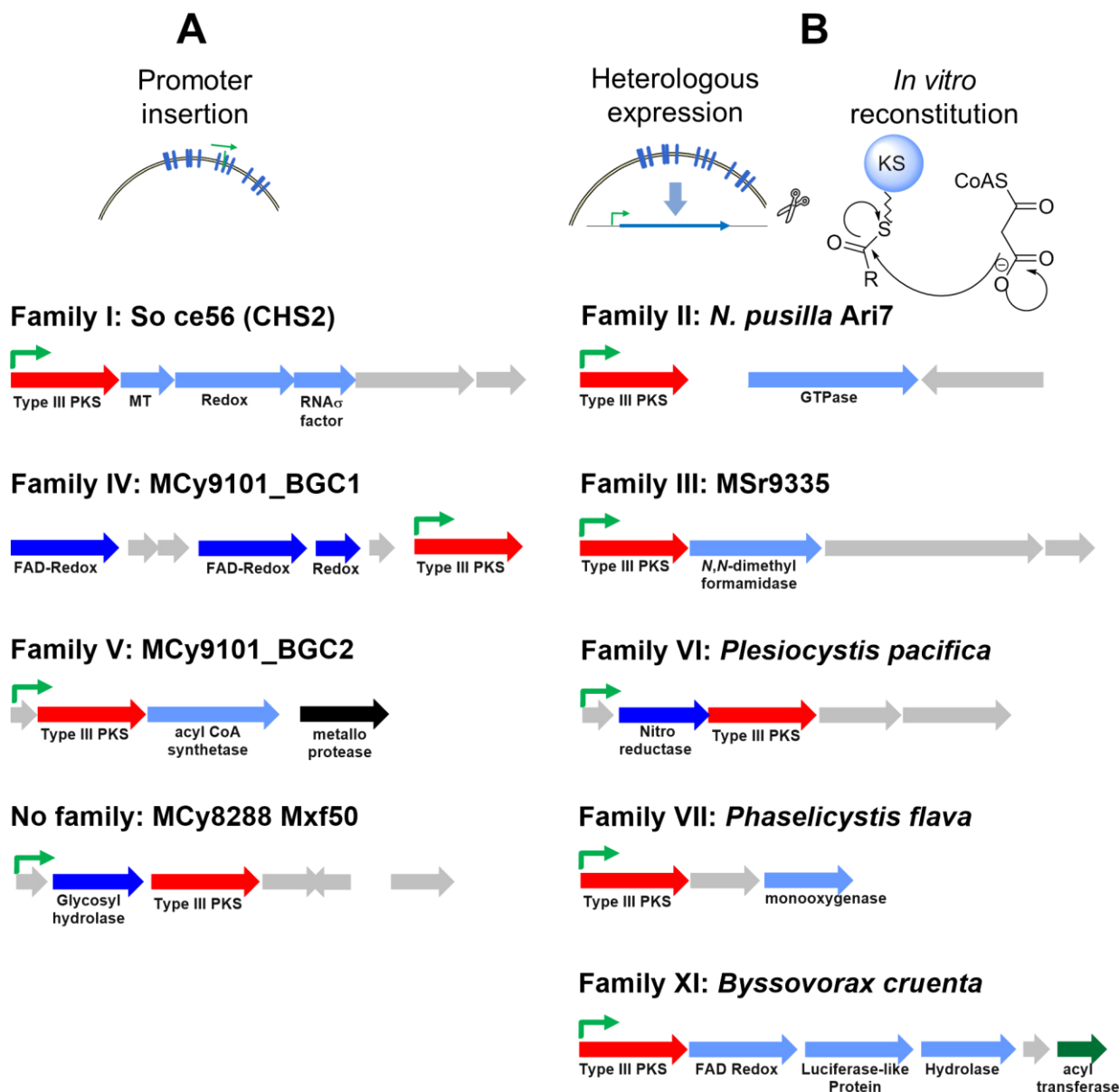


when artificial promoters are used; therefore, controlled gene regulation is essential for successful genetic activation strategies.

In conclusion, all aforementioned approaches are plausible to access the hidden secondary metabolome encoded by myxobacterial type III polyketides. The genomic survey in chapter 6 revealed also that 23 type III polyketide synthases are co-localized with different putative BGCs. Examples from literature exhibit that type III polyketides can provide building blocks for different natural products such as kendomycin (type I PKS)<sup>105</sup> and balhimycin (NRPS)<sup>106</sup> (further examples see chapter 6). However, it seems currently more plausible, that the automatic antiSMASH annotation fused these closely located BGCs to hypothetical hybrid gene clusters.

Beside natural hybrid type III PKS systems, latest developments brought genetically engineered hybrid biosynthetic machineries consisting of bacterial type II PKSs and plant type III PKS. Three of these artificial *de novo* hybrid biosynthetic machineries have been successfully expressed to direct the production of valuable natural products such as carminic acid<sup>107</sup> or octaketides for the biosynthesis of benzoisochromanequinone antibiotics<sup>108</sup>. The rationale behind these engineered hybrid biosynthetic pathways is that the type II bacterial PKSs cannot be expressed in eukaryotes<sup>109</sup>, therefore the biosynthetic machinery for the production of non-reduced octaketides relies in these hybrids on type III PKS enzymes. On the other hand, type III PKS pathways in plants lack characterized cyclases, aromatasases and tailoring enzymes that severely limit the use of type III PKSs for versatile polyketide biosynthesis. The synergy of these hybrid biosynthetic pathways is impressive since it shows the potential of engineered combination of different biosynthetic principles. In that sense, the provided insights of chapter 6 might initiate hybrid engineered myxobacterial pathways.

Although the investigation conducted in chapter 6 did not lead to the discovery of new myxobacterial type III polyketides, these studies connected for the first time a myxobacterial type III PKS to its actual *in vivo* generated natural product. Thus, this study only presents the beginning of prospective genome mining to access the hidden secondary metabolome of type III PKS machineries.



**Figure 6.** Different type III PKS gene cluster families identified in myxobacteria and future prospective genome mining approaches to access the underlying hidden natural products. **A)** Promoter insertion in the native host, might reveal the associated natural products originating from the type III PKS gene family I, IV, V and the type PKS gene from MCy8288 Mxf50, since all myxobacterial strains can be genetically manipulated. **B)** Type III PKS genes from other families would require a heterologous host and/or biosynthetic characterization via *in vitro* reconstituted systems. Further information concerning the identified type III PKS gene cluster families can be found in chapter 6.

### 7.3.4 Myxolipoxazoles and Myxopyrimidinols – Atypical PKS-NRPS Hybrid Machinery produces rare Chemical Scaffolds

In chapter 5, a classical metabolome-guided screening of myxobacterial natural products culminated in the discovery of the rare 1,2-oxazole and 4-pyrimidinol scaffolds, displayed by myxolipoxazole A and myxopyrimidinol A, respectively. Heterocyclic chemical scaffolds are common for numerous myxobacterial natural products such as the thiazole rings in epothilone and myxothiazol<sup>21</sup>, the oxazole ring formation during the biosynthesis of ajudazols<sup>110</sup> and disorazols<sup>52</sup>, and six-ring formations during the biosynthesis of ambruticin<sup>111</sup> and pyrroazols<sup>1</sup>. Many heterocycles formed during the biosynthesis of NRPS natural products are made by cyclization domains within a NRPS module that performs the amide bond forming-condensation of L-serine or L-cysteine and the cyclization to form an oxazoline/thiazoline intermediate, which can be subsequently oxidized by oxidase domains to yield the unsaturated heterocycles oxazole and thiazole, respectively. The thioesterase, usually present on the last elongation module, catalyzes in those cases not only the release from the assembly line but also the formation of macrocyclic structures. The pyrroazols follow a typical NRPS-catalyzed oxazole ring formation; the  $\delta$ -lactone arises probably from a thioesterase-catalyzed ring formation and the 5-chloro-pyrrole moiety from L-proline.

In contrast to 1,3-oxazoles, the 1,2-oxazole motif (simply termed isoxazole) is rarely presented in natural products and the biosynthesis for most of these compounds remained elusive<sup>112–114</sup>. The 1,2-oxazole scaffold in myxolipoxazole A is generated via a multimodular PKS-NRPS hybrid machinery, which is to the best of our knowledge unprecedented (**Figure 7A**). The firstly discovered and well-known isoxazole-containing (isoxazolidone) natural product is the cyclic analog of D-alanine termed D-cycloserine<sup>115–118</sup>, which is produced by *Streptomyces garyphalus*<sup>119</sup> and *Streptomyces lavendulae*<sup>120</sup>. This isoxazolidone features potent antibacterial bioactivity through its phosphorylated form, which enables inhibition of the D-alanine:D-alanine ligase<sup>121</sup>. Despite known neurological toxicity issues of D-cycloserine<sup>122</sup>, it is used as an oral bacteriostatic second line-medication against multidrug-resistant tuberculosis caused by a variant of *Mycobacterium tuberculosis*<sup>123</sup>. The biosynthesis of D-cycloserine starts from L-serine, which is *O*-acetylated by the homoserine *O*-acetyltransferase DcsE to generate *O*-acetyl-L-serine<sup>124,125</sup>, which reacts subsequently with hydroxyurea to yield *O*-ureido-L-serine catalyzed by *O*-acetylserine sulfhydrylase DcsD (pyridoxal phosphate (PLP)-dependent enzyme)<sup>126</sup>. *O*-ureido-L-serine undergoes racemization by the PLP-independent racemase DcsC<sup>126,127</sup>, followed by heterocyclization with DcsG (ATP-grasp fold family of protein) yielding D-cycloserine. The precursor L-arginine is involved in the biosynthesis of D-cycloserine, by DcsA-catalyzed conversion (heme protein) to *N*<sup>ω</sup>-hydroxy-L-arginine, which in turns provide hydroxyurea as building block through hydrolysis by DcsB<sup>128</sup> (**Figure 7B**).

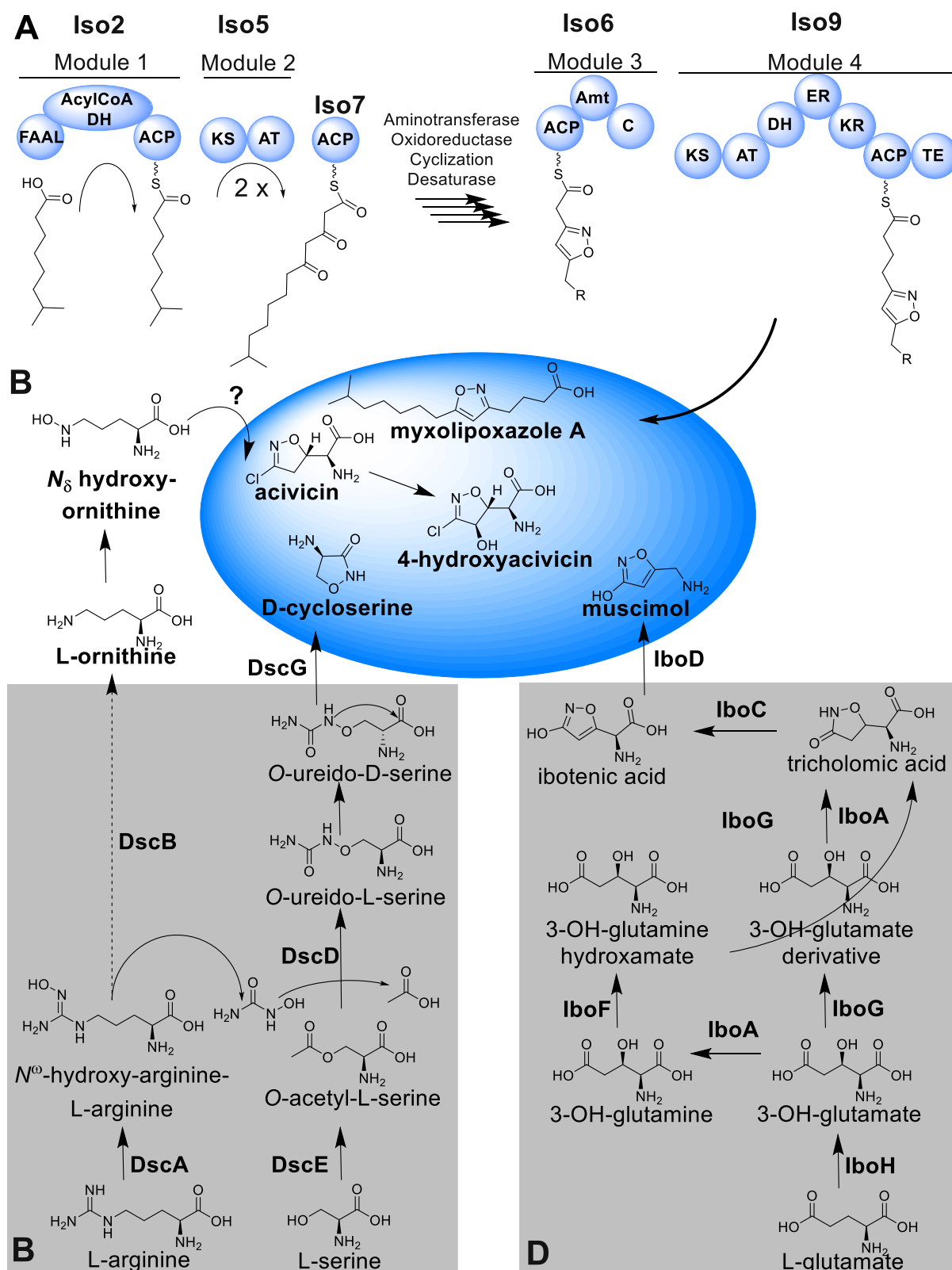
The isoxazoline-containing bacterial natural products acivicin and 4-hydroxyacivicin represent another class of 1,2-oxazole compounds from *Streptomyces*. *Streptomyces sviveus* the producer of acivicin and 4-hydroxyacivicin has been investigated by stable isotope-labeled precursor feeding which revealed that the pathway might start from L-ornithine through *N*<sup>δ</sup>-hydroxyornithine to acivicin to finally yield 4-hydroxyacivicin<sup>129</sup> (**Figure 7C**). Despite the availability of the producer strain and its genome sequence, no genetic correlation to a corresponding BGC has been made yet. This finding is highly surprising since acivicin and 4-hydroxyacivicin feature a rare chemical scaffold, show diverse potent biological activities<sup>130,131</sup> and the producing strain has amenable characteristics for further genetic

investigations. This finding might be connected to the observation made in chapter 5, that neither acivicin nor 4-hydroxyacivicin was detected in the secondary metabolome of *Streptomyces sviveus* grown in our laboratory. It would be highly interesting to find the biosynthetic genes associated with the production of acivicin and 4-hydroxyacivicin.

Paralleling the biosynthetic knowledge of acivicin, until recently, little was known concerning the biosynthesis of ibotenic acid and muscimol from *Amanita muscaria*<sup>132</sup>. Recent findings support the proposed biosynthesis of ibotenic acid to start from 3-hydroxy (3-OH)-glutamate<sup>133</sup>, supported through identification and recombinant production of the glutamate hydroxylase IboH from *A. muscaria*<sup>134</sup>. The gene *iboH* encoding the glutamate hydroxylase is surrounded by six other genes which are proposed to be involved in the biosynthesis of ibotenic acid<sup>134</sup> (**Figure 7D**). The proposed biosynthetic route to muscimol starts from L-glutamate, which undergoes hydroxylation by IboH to yield 3-OH-glutamate. Afterwards the biosynthetic pathway could be further proceeded by the adenylating enzyme IboA, the flavin-dependent monooxygenase IboF and the PLP-dependent enzyme IboG to yield the intermediates 3-OH-glutamine, 3-OH-glutamine hydroxamate and tricholomic acid, respectively. An alternative route would convert 3-OH-glutamate to a 3-OH-glutamate derivative catalyzed by IboG, and subsequently convert this derivative by IboA. The precursor tricholomic acid is proposed to be converted to ibotenic acid by the CYP450 enzyme IboC, and the last step in the biosynthesis involves the decarboxylase IboD to decarboxylate ibotenic acid to muscimol (**Figure 7D**)<sup>134</sup>.

The discovery of the myxolipoxazoles along with the myxopyrimidinols in chapter 5 is in that sense unexpected, since the heterocyclic moieties featured in both natural product classes are rare in natural products and the underlying biosynthesis remained up to date entirely elusive. In general, heterocyclic–fatty acid hybrids are interesting compounds in the field of medicinal chemistry<sup>135</sup>, since these hybrids might benefit from both chemical scaffolds to display potent biological activity as shown for the natural products jaspine B<sup>136</sup> or evocarpine<sup>137</sup>. The biosynthesis of the isoxazole heterocycle in myxolipoxazole A is unique and different from the proposed mechanisms in acivicin, ibotenic acid or D-cycloserine biosynthesis. An atypical modular PKS pathway guides the major part of the underlying biosynthesis. A single NRPS module guides the incorporation of the nitrogen via an aminotransferase mechanism and a stand-alone A domain continues the further processing of the biosynthetic pathway.

The exact order of the isoxazole or 4-pyrimidinol cyclization remains elusive, but the work in chapter 5 investigated, identified and provided genetic proof for the partial biosynthesis of two groups of myxobacterial natural products featuring two unique biosynthetic chemical scaffolds. In addition, single crossover-based genetic engineering of the producer strain *Myxococcus xanthus* Mx x48 significantly improved the production of natural products associated with this atypical PKS-NRPS pathway. This induced gene expression also revealed that most of the afforded natural products can be put into the group of myxolipoxazoles, while the myxopyrimidinols are less prominent. Further investigation should involve heterologous expression of the identified genetic operon (*iso1–iso9*) ideally in the phylogenetically related strain *M. xanthus* DK1622 to further investigate the biosynthesis of the myxolipoxazoles and myxopyrimidinols. This approach allows profound genetic engineering of the identified genetic operon in the intermediate host *E. coli*, to simplify seamless gene deletion of each gene needed to investigate the effects on the corresponding secondary metabolites. Nevertheless, an *in vitro* system with recombinantly produced biosynthetic proteins catalyzing the cyclization reactions in the biosynthesis of myxolipoxazoles and myxopyrimidinols is essential to entirely reveal the biosynthetic logic of this unique pathway.



**Figure 7.** The biosynthetic logic of isoxazole formation in natural products (blue circle). **A**) The myxolipoxazoles are biosynthesized by a PKS-NRPS hybrid pathway. **B**) The biosynthesis of D-cycloserine has been characterized previously and starts from L-serine and L-arginine<sup>119,128,138</sup>. **C**) The biosynthetic pathway of 4-hydroxyacivicin remains elusive; nevertheless L-ornithine and N<sup>δ</sup>-hydroxyornithine has been identified as building blocks<sup>129</sup>. **D**) A proposed biosynthetic gene cluster connected with the generation of muscimol has been discovered; the proposed biosynthetic pathway in *Amanita muscaria* starts from L-glutamate, via the well known intermediates tricholomic acid and ibotenic acid<sup>134</sup>.

## 7.4 Final Conclusion

The primary objective of this thesis was to explore the biosynthetic potential of myxobacteria and characterize the biosynthesis of several myxobacterial natural products. In particular, the numerous biosynthetic origins of the investigated natural products in this thesis emphasize the diversity in the biosynthetic logic that has evolved in myxobacteria. This includes biosynthetic pathways of modular prototypical and atypical PKS-NRPS hybrids, the generation of ribosomal peptides with unique posttranslational modifications, single iteratively working catalytic enzymes to afford unprecedented terpene-nucleoside hybrids and previously unidentified type III polyketides.

Along these lines, the most significant findings in this thesis were the isolation of the first type III polyketide from *Mxyococcus xanthus* DK1622 after activation of its PKS gene cluster, the genetic correlation of the myxolipoxazoles and myxopyrimidinols to their atypical PKS-NRPS hybrid biosynthetic pathway and the identification and investigation of the first nucleoside-hybrid biosynthetic pathway in myxobacteria to bring about the unique 2-hydroxysorangiadenosine.

The knowledge of these newly uncovered biosynthetic pathways is foundational for an improved understanding of the biosynthetic logic in myxobacteria.

## 7.5 References

1. Witte, S. N. R., Hug, J. J., Géraldy, M., Müller, R. & Kalesse, M. Biosynthesis and Total Synthesis of Pyrronazol B: a Secondary Metabolite from *Nannocystis pusilla*. *Chem. Eur. J.* **23**, 15917–15921 (2017).
2. Hug, J. J. *et al.* Biosynthesis of citilins, unusual ribosomally synthesized and post-translationally modified peptides from *Myxococcus xanthus*. *ACS Chem. Biol.* **submitted** (2020).
3. Okoth Dorothy A. *et al.* 2-Hydroxysorangiadenosine: Structure and Biosynthesis of a Myxobacterial Sesquiterpene–Nucleoside. *Molecules (Basel, Switzerland)* **accepted** (2020).
4. Ahn, J. W., Jang, K. H., Chung, S. C., Oh, K. B. & Shin, J. Sorangiadenosine, a new sesquiterpene adenoside from the myxobacterium *Sorangium cellulosum*. *Org. Lett.* **10**, 1167–1169 (2008).
5. Hug, J. J., Panter, F., Krug, D. & Müller, R. Genome mining reveals uncommon alkylpyrones as type III PKS products from myxobacteria. *J. Ind. Microbiol. Biotechnol.* **46**, 319–334 (2019).
6. Shen, B. A New Golden Age of Natural Products Drug Discovery. *Cell* **163**, 1297–1300 (2015).
7. Hubert, J., Nuzillard, J.-M. & Renault, J.-H. Dereplication strategies in natural product research. How many tools and methodologies behind the same concept? *Phytochem. Rev.* **16**, 55–95 (2017).
8. Blin, K. *et al.* antiSMASH 5.0: updates to the secondary metabolite genome mining pipeline. *Nucleic Acids Res.*, W81–W87 (2019).
9. Amiri Moghaddam, J. *et al.* Analysis of the Genome and Metabolome of Marine Myxobacteria Reveals High Potential for Biosynthesis of Novel Specialized Metabolites. *Sci. Rep.* **8**, 16600 (2018).
10. Cortina, N. S., Krug, D., Plaza, A., Revermann, O. & Müller, R. Myxoprincomide: a natural product from *Myxococcus xanthus* discovered by comprehensive analysis of the secondary metabolome. *Angew. Chem. Int. Ed. Engl.* **51**, 811–816 (2012).
11. Kjaerulff, L. *et al.* Pyxipyrrolones: Structure elucidation and biosynthesis of cytotoxic myxobacterial metabolites. *Angew. Chem. Int. Ed.* **56**, 9614–9618 (2017).
12. Surup, F. *et al.* Disciformycins A and B: 12-membered macrolide glycoside antibiotics from the myxobacterium *Pyxidicoccus fallax* active against multiresistant staphylococci. *Angew. Chem. Int. Ed. Engl.* **49**, 13588–13591 (2014).
13. Gorges, J. *et al.* Structure, Total Synthesis, and Biosynthesis of Chloromyxamides: Myxobacterial Tetrapeptides Featuring an Uncommon 6-Chloromethyl-5-methoxypiperic Acid Building Block. *Angew. Chem. Int. Ed. Engl.* **57**, 14270–14275 (2018).
14. Plaza, A., Viehrig, K., Garcia, R. & Müller, R. Jahnellamides,  $\alpha$ -keto- $\beta$ -methionine-containing peptides from the terrestrial myxobacterium *Jahnella* sp.: structure and biosynthesis. *Org. Lett.* **15**, 5882–5885 (2013).

15. Rinkel, J. & Dickschat, J. S. Recent highlights in biosynthesis research using stable isotopes. *Beilstein J. Org. Chem.* **11**, 2493–2508 (2015).
16. Baumann, S. *et al.* Cystobactamids: myxobacterial topoisomerase inhibitors exhibiting potent antibacterial activity. *Angew. Chem. Int. Ed.* **53**, 14605–14609 (2014).
17. Hannigan, G. D. *et al.* A deep learning genome-mining strategy for biosynthetic gene cluster prediction. *Nucleic Acids Res.* **47**, e110 (2019).
18. Li, Y. & Müller, R. Non-modular polyketide synthases in myxobacteria. *Phytochemistry* **70**, 1850–1857 (2009).
19. Beck, E., Ludwig, G., Auerswald, E. A., Reiss, B. & Schaller, H. Nucleotide sequence and exact localization of the neomycin phosphotransferase gene from transposon Tn5. *Gene* **19**, 327–336 (1982).
20. Iniesta, A. A., García-Heras, F., Abellón-Ruiz, J., Gallego-García, A. & Elías-Arnanz, M. Two systems for conditional gene expression in *Myxococcus xanthus* inducible by isopropyl-β-D-thiogalactopyranoside or vanillate. *J. Bacteriol.* **194**, 5875–5885 (2012).
21. Hug, J. J. & Müller, R. Host Development for Heterologous Expression and Biosynthetic Studies of Myxobacterial Natural Products. *Comprehensive Natural Products III: Chemistry and Biology, Chapter 14818* **In press** (2020).
22. Huo, L. *et al.* Heterologous expression of bacterial natural product biosynthetic pathways. *Nat. Prod. Rep.* **36**, 1412–1436 (2019).
23. Hug, J. J., Bader, C. D., Remškar, M., Cirnski, K. & Müller, R. Concepts and Methods to Access Novel Antibiotics from Actinomycetes. *Antibiotics* **7**, 44 (2018).
24. Foulston, L. Genome mining and prospects for antibiotic discovery. *Curr. Opin. Microbiol.* **51**, 1–8 (2019).
25. Rudolf, J. D., Yan, X. & Shen, B. Genome neighborhood network reveals insights into enediynes biosynthesis and facilitates prediction and prioritization for discovery. *J Ind Microbiol Biotechnol* **43**, 261–276 (2016).
26. Peek, J. *et al.* Rifamycin congeners kanglemycins are active against rifampicin-resistant bacteria via a distinct mechanism. *Nat. Commun.* **9**, 4147 (2018).
27. Li, Y.-X., Zhong, Z., Zhang, W.-P. & Qian, P.-Y. Discovery of cationic nonribosomal peptides as Gram-negative antibiotics through global genome mining. *Nat. Commun.* **9**, 3273 (2018).
28. Hover, B. M. *et al.* Culture-independent discovery of the malacidins as calcium-dependent antibiotics with activity against multidrug-resistant Gram-positive pathogens. *Nat. Microbiol.* **3**, 415–422 (2018).
29. Kling, A. *et al.* Targeting DnaN for tuberculosis therapy using novel griselimycins. *Science* **348**, 1106–1112 (2015).



30. Alanjary, M. *et al.* The Antibiotic Resistant Target Seeker (ARTS), an exploration engine for antibiotic cluster prioritization and novel drug target discovery. *Nucleic Acids Res.* **45**, W42-W48 (2017).
31. Panter, F., Krug, D., Baumann, S. & Müller, R. Self-resistance guided genome mining uncovers new topoisomerase inhibitors from myxobacteria. *Chem. Sci.* **9**, 4898–4908 (2018).
32. Hayashi, T., Kitamura, Y., Funa, N., Ohnishi, Y. & Horinouchi, S. Fatty acyl-AMP ligase involvement in the production of alkylresorcylic acid by a *Myxococcus xanthus* type III polyketide synthase. *ChemBioChem* **12**, 2166–2176 (2011).
33. Funabashi, M., Funa, N. & Horinouchi, S. Phenolic lipids synthesized by type III polyketide synthase confer penicillin resistance on *Streptomyces griseus*. *J. Biol. Chem.* **283**, 13983–13991 (2008).
34. Yamada, Y. *et al.* Terpene synthases are widely distributed in bacteria. *Proc. Natl. Acad. Sci. USA* **112**, 857–862 (2015).
35. Dickschat, J. S., Pahirulzaman, K. A. K., Rabe, P. & Klapschinski, T. A. An Improved Technique for the Rapid Chemical Characterisation of Bacterial Terpene Cyclases. *ChemBioChem* **15**, 810–814 (2014).
36. Helfrich, E. J. N., Lin, G.-M., Voigt, C. A. & Clardy, J. Bacterial terpene biosynthesis: challenges and opportunities for pathway engineering. *Beilstein J. Org. Chem.* **15**, 2889–2906 (2019).
37. Reiter, S., Cahn, J. K. B., Wiebach, V., Ueoka, R. & Piel, J. Characterization of an Orphan Type III Polyketide Synthase Conserved in Uncultivated "*Entotheonella*" Sponge Symbionts. *ChemBioChem* (2019).
38. Wang, M. *et al.* Sharing and community curation of mass spectrometry data with Global Natural Products Social Molecular Networking. *Nat. Biotechnol.* **34**, 828–837 (2016).
39. Barrow, R. A. & Capon, R. J. Alkyl and Alkenyl Resorcinols From an Australian Marine Sponge, *Haliclona* Sp (Haplosclerida : Haliclonidae). *Aust. J. Chem.* **44**, 1393–1405 (1991).
40. Quadri, L. E., Sello, J., Keating, T. A., Weinreb, P. H. & Walsh, C. T. Identification of a *Mycobacterium tuberculosis* gene cluster encoding the biosynthetic enzymes for assembly of the virulence-conferring siderophore mycobactin. *Chem. Biol.* **5**, 631–645 (1998).
41. Felnagle, E. A. *et al.* MbtH-like proteins as integral components of bacterial nonribosomal peptide synthetases. *Biochemistry* **49**, 8815–8817 (2010).
42. Zhang, W., Heemstra, J. R., Walsh, C. T. & Imker, H. J. Activation of the Pacidamycin PacL Adenylation Domain by MbtH-Like Proteins. *Biochemistry* (2010).
43. Boll, B., Taubitz, T. & Helde, L. Role of MbtH-like proteins in the adenylation of tyrosine during aminocoumarin and vancomycin biosynthesis. *J. Biol. Chem.* **286**, 36281–36290 (2011).

44. Miller, B. R., Drake, E. J., Shi, C., Aldrich, C. C. & Gulick, A. M. Structures of a Nonribosomal Peptide Synthetase Module Bound to MbtH-Like Proteins Support a Highly Dynamic Domain Architecture. *J. Biol. Chem.*, 22559–22571 (2016).
45. Esquilín-Lebrón, K. J., Boynton, T. O., Shinkets, L. J. & Thomas, M. G. An orphan MbtH-like protein interacts with multiple nonribosomal peptide synthetases in *Myxococcus xanthus* DK1622. *J. Bacteriol.* (2018).
46. Baltz, R. H. Function of MbtH homologs in nonribosomal peptide biosynthesis and applications in secondary metabolite discovery. *J. Ind. Microbiol. Biotechnol.* **38**, 1747–1760 (2011).
47. Sosio, M. *et al.* Analysis of the Pseudouridimycin Biosynthetic Pathway Provides Insights into the Formation of C-nucleoside Antibiotics. *Cell Chem. Biol.* (2018).
48. Herrmann, J., Fayad, A. A. & Müller, R. Natural products from myxobacteria: novel metabolites and bioactivities. *Nat. Prod. Rep.* **34**, 135–160 (2017).
49. Weissman, K. J. & Müller, R. Myxobacterial secondary metabolites: bioactivities and modes-of-action. *Nat. Prod. Rep.* **27**, 1276–1295 (2010).
50. Buntin, K., Weissman, K. J. & Müller, R. An unusual thioesterase promotes isochromanone ring formation in ajudazol biosynthesis. *ChemBioChem* **11**, 1137–1146 (2010).
51. Perlova, O., Gerth, K., Hans, A., Kaiser, O. & Müller, R. Identification and analysis of the chivosazol biosynthetic gene cluster from the myxobacterial model strain *Sorangium cellulosum* So ce56. *J. Biotechnol.* **121**, 174–191 (2006).
52. Kopp, M., Irschik, H., Pradella, S. & Müller, R. Production of the tubulin destabilizer disorazol in *Sorangium cellulosum*: biosynthetic machinery and regulatory genes. *ChemBioChem* **6**, 1277–1286 (2005).
53. Los Santos, E. L. C. de. NeuRiPP: Neural network identification of RiPP precursor peptides. *Sci. Rep.* **9**, 13406 (2019).
54. Agrawal, P., Khater, S., Gupta, M., Sain, N. & Mohanty, D. RiPPMiner: a bioinformatics resource for deciphering chemical structures of RiPPs based on prediction of cleavage and cross-links. *Nucleic Acids Res.* **45**, W80–W88 (2017).
55. Rodriguez, J., O'Neill, S. & Walczak, M. A. Constrained saccharides: a review of structure, biology, and synthesis. *Nat. Prod. Rep.* **35**, 220–229 (2018).
56. Cimermancic, P. *et al.* Insights into secondary metabolism from a global analysis of prokaryotic biosynthetic gene clusters. *Cell* **158**, 412–421 (2014).
57. McCranie, E. K. & Bachmann, B. O. Bioactive oligosaccharide natural products. *Nat. Prod. Rep.* **31**, 1026–1042 (2014).
58. Kautsar, S. A. *et al.* MIBiG 2.0: a repository for biosynthetic gene clusters of known function. *Nucleic Acids Res.* (2019).

59. Flatt, P. M. & Mahmud, T. Biosynthesis of aminocyclitol-aminoglycoside antibiotics and related compounds. *Nat. Prod. Rep.* **24**, 358–392 (2007).
60. Kudo, F. & Eguchi, T. Aminoglycoside antibiotics. New Insights into the Biosynthetic Machinery of Old Drugs. *Chem. Rec.* **16**, 4–18 (2016).
61. Yu, Y., Zhang, Q. & Deng, Z. Parallel pathways in the biosynthesis of aminoglycoside antibiotics. *F1000Research* **6** (2017).
62. Wolfender, J.-L., Litaudon, M., Touboul, D. & Queiroz, E. F. Innovative omics-based approaches for prioritisation and targeted isolation of natural products - new strategies for drug discovery. *Nat. Prod. Rep.* **36**, 855–868 (2019).
63. Etzbach, L., Plaza, A., Garcia, R., Baumann, S. & Müller, R. Cystomanamides: structure and biosynthetic pathway of a family of glycosylated lipopeptides from myxobacteria. *Org. Lett.* **16**, 2414–2417 (2014).
64. Barbier, J. *et al.* Isolation and total synthesis of icumazoles and noricumazoles—antifungal antibiotics and cation-channel blockers from *Sorangium cellulosum*. *Angew. Chem. Int. Ed. Engl.* **51**, 1256–1260 (2012).
65. Albataineh, H., Duke, M., Misra, S. K., Sharp, J. S. & Stevens, D. C. Solo acylhomoserine lactone synthase from predatory myxobacterium suggests beneficial participation in interspecies cross talk. *bioRxiv* **65** (2019).
66. Lloyd, D. G. & Whitworth, D. E. The myxobacterium *Myxococcus xanthus* can sense and respond to the quorum signals secreted by potential prey organisms. *Front. Microbiol.* **8**, 439 (2017).
67. Camilli, A. & Bassler, B. L. Bacterial small-molecule signaling pathways. *Science (New York, N.Y.)* **311**, 1113–1116 (2006).
68. Takano, E.  $\gamma$ -Butyrolactones: *Streptomyces* signalling molecules regulating antibiotic production and differentiation. *Curr. Opin. Microbiol.* **9**, 287–294 (2006).
69. Papenfort, K. & Bassler, B. L. Quorum sensing signal-response systems in Gram-negative bacteria. *Nat. Rev. Chem.* **14**, 576–588 (2016).
70. Polkade, A. V., Mantri, S. S., Patwekar, U. J. & Jangid, K. Quorum Sensing: An Under-Explored Phenomenon in the Phylum Actinobacteria. *Front. Microbiol.* **7**, 131 (2016).
71. Gregory, K., Salvador, L. A., Akbar, S., Adaikpoh, B. I. & Stevens, D. C. Survey of Biosynthetic Gene Clusters from Sequenced Myxobacteria Reveals Unexplored Biosynthetic Potential. *Microorganisms* **7** (2019).
72. Krug, D. *et al.* Discovering the Hidden Secondary Metabolome of *Myxococcus xanthus*: a Study of Intraspecific Diversity. *Appl. Environ. Microbiol.* **74**, 3058–3068 (2008).
73. Krug, D., Zurek, G., Schneider, B., Garcia, R. & Müller, R. Efficient mining of myxobacterial metabolite profiles enabled by liquid chromatography-electrospray ionization-time-of-flight mass

- spectrometry and compound-based principal component analysis. *Anal. Chim. Acta* **624**, 97–106 (2008).
74. Revermann, O. Novel secondary metabolites from myxobacteria and their biosynthetic machinery. Dissertation. Saarland University, 2012.
75. Viehrig, K. *et al.* Structure and biosynthesis of crocagins: polycyclic postranslationally modified ribosomal peptides from *Chondromyces crocatus*. *Angew. Chem.*, 1–5 (2017).
76. Myronovskyi, M. *et al.* Generation of a cluster-free *Streptomyces albus* chassis strains for improved heterologous expression of secondary metabolite clusters. *Metab. Eng.* **49**, 316–324 (2018).
77. Forneris, C. C. & Seyedsayamdost, M. R. In Vitro Reconstitution of OxyC Activity Enables Total Chemoenzymatic Syntheses of Vancomycin Aglycone Variants. *Angewandte. Chem. Int. Ed. Engl. (Angewandte Chemie International Edition in English)* **57**, 8048–8052 (2018).
78. Forneris, C. C., Ozturk, S., Gibson, M. I., Sorensen, E. J. & Seyedsayamdost, M. R. In Vitro Reconstitution of OxyA Enzymatic Activity Clarifies Late Steps in Vancomycin Biosynthesis. *ACS Chem. Biol.* **12**, 2248–2253 (2017).
79. Brieke, C., Peschke, M., Haslinger, K. & Cryle, M. J. Sequential In Vitro Cyclization by Cytochrome P450 Enzymes of Glycopeptide Antibiotic Precursors Bearing the X-Domain from Nonribosomal Peptide Biosynthesis. *Angew. Chem. Int. Ed.* **54**, 15715–15719 (2015).
80. Haslinger, K., Peschke, M., Brieke, C., Maximowitsch, E. & Cryle, M. J. X-domain of peptide synthetases recruits oxygenases crucial for glycopeptide biosynthesis. *Nature* **521**, 105–109 (2015).
81. Yokoyama, K. & Lilla, E. A. C–C bond forming radical SAM enzymes involved in the construction of carbon skeletons of cofactors and natural products. *Nat. Prod. Rep.* (2018).
82. Schramma, K. R., Bushin, L. B. & Seyedsayamdost, M. R. Structure and biosynthesis of a macrocyclic peptide containing an unprecedented lysine-to-tryptophan crosslink. *Nat. Chem.* **7**, 431–437 (2015).
83. Imai, Y. *et al.* A new antibiotic selectively kills Gram-negative pathogens. *Nature* **576**, 459–464 (2019).
84. Maurer, C. K. *et al.* Discovery of the first small-molecule CsrA-RNA interaction inhibitors using biophysical screening technologies. *Future Med. Chem.* **8**, 931–947 (2016).
85. O'Shea, R. & Moser, H. E. Physicochemical properties of antibacterial compounds: implications for drug discovery. *J. Med. Chem.* **51**, 2871–2878 (2008).
86. Dickschat, J. S., Bode, H. B., Wenzel, S. C., Müller, R. & Schulz, S. Biosynthesis and identification of volatiles released by the myxobacterium *Stigmatella aurantiaca*. *ChemBioChem* **6**, 2023–2033 (2005).
87. Dickschat, J. S., Wenzel, S. C., Bode, H. B., Müller, R. & Schulz, S. Biosynthesis of volatiles by the myxobacterium *Myxococcus xanthus*. *ChemBioChem* **5**, 778–787 (2004).

88. Dickschat, J. S. *et al.* Biosynthesis of the off-flavor 2-methylisoborneol by the myxobacterium *Nannocystis exedens*. *Angew. Chem. Int. Ed. Engl.* **46**, 8287–8290 (2007).
89. Nawrath, T. *et al.* Identification of (8*S*,9*S*,10*S*)-8,10-dimethyl-1-octalin, a key intermediate in the biosynthesis of geosmin in bacteria. *J. Am. Chem. Soc.* **130**, 430–431 (2008).
90. Dickschat, J. S., Bode, H. B., Mahmud, T., Müller, R. & Schulz, S. A novel type of geosmin biosynthesis in myxobacteria. *J. Org. Chem.* **70**, 5174–5182 (2005).
91. Schiffrin, A. *et al.* Characterization of the Gene Cluster CYP264B1-geoA from *Sorangium cellulosum* So ce56: Biosynthesis of (+)-Eremophilene and Its Hydroxylation. *ChemBioChem* **16**, 337–344 (2015).
92. Schiffrin, A. *et al.* A single terpene synthase is responsible for a wide variety of sesquiterpenes in *Sorangium cellulosum* So ce56. *Org. Biomol. Chem.* **14**, 3385–3393 (2016).
93. Raju, R., Mohr, K. I., Bernecker, S., Herrmann, J. & Müller, R. Cystodienoic acid: A new diterpene isolated from the myxobacterium *Cystobacter* sp. *J. Antibiot.* **68**, 473–475 (2015).
94. Tomura, T. *et al.* An Unusual Diterpene—Enhygromic Acid and Deoxyenhygrolides from a Marine Myxobacterium, *Enhygromyxa* sp. *Mar. Drugs* **15** (2017).
95. Kopp, M. *et al.* Insights into the complex biosynthesis of the leupyrrins in *Sorangium cellulosum* So ce690. *Mol. Biosyst.* **7**, 1549–1563 (2011).
96. Ahn, M., Y., Jung, J., H., Na, Y., J. & Kim, H., S. A natural histone deacetylase inhibitor, Psammaphin A, induces cell cycle arrest and apoptosis in human endometrial cancer cells. *Gynecologic Oncology* **108**, 27–33 (2008).
97. Layre, E. *et al.* Molecular profiling of *Mycobacterium tuberculosis* identifies tuberculosinyl nucleoside products of the virulence-associated enzyme Rv3378c. *Proc. Natl. Acad. Sci. U.S.A.* **111**, 2978–2983 (2014).
98. Leung, H. C., Chen, Y. & Winkler, M. E. Regulation of substrate recognition by the MiaA tRNA prenyltransferase modification enzyme of *Escherichia coli* K-12. *J. Biol. Chem.* **272**, 13073–13083 (1997).
99. Pierrel, F., Hernandez, H. L., Johnson, M. K., Fontecave, M. & Atta, M. MiaB protein from *Thermotoga maritima*. Characterization of an extremely thermophilic tRNA-methylthiotransferase. *Journal of Biological Chemistry* **278**, 29515–29524 (2003).
100. Pogorevc, D. *et al.* Biosynthesis and Heterologous Production of Argyrins. *ACS Synth. Biol.* **8**, 1121–1133 (2019).
101. Awal, R. P. *et al.* *Vitosangium cumulatum* gen. nov., sp. nov. and *Vitosangium subalbum* sp. nov., soil myxobacteria, and emended descriptions of the genera *Archangium* and *Angiococcus*, and of the family *Cystobacteraceae*. *Int. J. Syst. Evol. Microbiol.* **67**, 1422–1430 (2017).

102. Gross, F. *et al.* Bacterial type III polyketide synthases: Phylogenetic analysis and potential for the production of novel secondary metabolites by heterologous expression in pseudomonads. *Arch. Microbiol.* **185**, 28–38 (2006).
103. Sone, Y. *et al.* Identification and characterization of bacterial enzymes catalyzing the synthesis of 1,8-dihydroxynaphthalene, a key precursor of dihydroxynaphthalene melanin, from *Sorangium cellulosum*. *Appl. Environ. Microbiol.*, AEM.00258-18 (2018).
104. Zhou, Q. *et al.* Molecular mechanism of polyketide shortening in anthraquinone biosynthesis of *Photorhabdus luminescens*. *Chemical Science* **48**, 4688 (2019).
105. Wenzel, S. C., Bode, H. B., Kochems, I. & Müller, R. A type I/type III polyketide synthase hybrid biosynthetic pathway for the structurally unique *ansa* compound kendomycin. *ChemBioChem* **9**, 2711–2721 (2008).
106. Pelzer, S. *et al.* Identification and analysis of the balhimycin biosynthetic gene cluster and its use for manipulating glycopeptide biosynthesis in *Amycolatopsis mediterranei* DSM5908. *Antimicrob. Agents Chemother.* **43**, 1565–1573 (1999).
107. Frandsen, R. J. N. *et al.* Heterologous production of the widely used natural food colorant carminic acid in *Aspergillus nidulans*. *Sci. Rep.* **8**, 12853 (2018).
108. Jakočiūnas, T. *et al.* Programmable polyketide biosynthesis platform for production of aromatic compounds in yeast. *Synth. Syst. Biotechnol.* **5**, 11–18 (2020).
109. Hertweck, C., Luzhetskyy, A., Rebets, Y. & Bechthold, A. Type II polyketide synthases: gaining a deeper insight into enzymatic teamwork. *Nat. Prod. Rep.* **24**, 162–190 (2007).
110. Essig, S., Bretzke, S., Müller, R. & Menche, D. Full stereochemical determination of ajudazols a and B by bioinformatics gene cluster analysis and total synthesis of ajudazol B by an asymmetric ortholithiation strategy. *J. Am. Chem. Soc.* **134**, 19362–19365 (2012).
111. Julien, B., Tian, Z. Q., Reid, R. & Reeves, C. D. Analysis of the ambruticin and jerangolid gene clusters of *Sorangium cellulosum* reveals unusual mechanisms of polyketide biosynthesis. *Chem. Biol.* **13**, 1277–1286 (2006).
112. Rahbæk, L. & Christophersen, C. in *The alkaloids*, edited by K. R. Alper, *et al.* (Elsevier\_1983–1989), pp. 185–233.
113. Kochetkov, N. K. & Sokolov, S. D. in *Advances in heterocyclic chemistry*, edited by A. R. Katritzky (Academic Press, New York, London, 1963), pp. 365–422.
114. Kaur, K., Kumar, V., Sharma, A. K. & Gupta, G. K. Isoxazoline containing natural products as anticancer agents: a review. *Eur. J. Med. Chem.* **77**, 121–133 (2014).
115. Kuehl, F. A. *et al.* D-4-Amino-3-Isoxazolidone, A New Antibiotic\*. *J. Am. Chem. Soc.* **77**, 2344–2345 (1955).
116. Harned, R. L., Hidy, P. H. & La Baw, E. K. Cycloserine. I. A preliminary report. *Antibiotics & chemotherapy (Northfield, Ill.)* **5**, 204–205 (1955).

117. Harris, D. A. *et al.* Discovery, development, and antimicrobial properties of D-4-amino-3-isoxazolidone (oxamycin), a new antibiotic produced by *Streptomyces garyphalus* n. sp. *Antibiotics & chemotherapy (Northfield, Ill.)* **5**, 183–190 (1955).
118. Shull, G. M. & Sardinas, J. L. PA-94, an antibiotic identical with D-4-amino-3-isoxazolidinone (cycloserine, oxamycin). *Antibiotics & chemotherapy (Northfield, Ill.)* **5**, 398–399 (1955).
119. Svensson, M.-L. & Gatenbeck, S. The pathway of D-cycloserine biosynthesis in *Streptomyces garyphalus*. *Arch. Microbiol.* **131**, 129–131 (1982).
120. Noda, M. *et al.* Self-protection Mechanism in D-Cycloserine-producing *Streptomyces lavendulae*. Gene cloning, characterization, and kinetics of its alanine racemase and D-alanyl-D-alanine ligase, which are target enzymes of D-cycloserine. *J. Biol. Chem.* **279**, 46143–46152 (2004).
121. Batson, S. *et al.* Inhibition of D-Ala:D-Ala ligase through a phosphorylated form of the antibiotic D-cycloserine. *Nat. Commun.* **8**, 1939 (2017).
122. Cohen, A. C. Pyridoxine in the prevention and treatment of convulsions and neurotoxicity due to cycloserine. *Ann. N. Y. Acad. Sci.* **166**, 346–349 (1969).
123. Li, Y. *et al.* Cycloserine for treatment of multidrug-resistant tuberculosis: a retrospective cohort study in China. *Infection and drug resistance* **12**, 721–731 (2019).
124. Kumagai, T. *et al.* Molecular cloning and heterologous expression of a biosynthetic gene cluster for the antitubercular agent D-cycloserine produced by *Streptomyces lavendulae*. *Antimicrob. Agents Chemother.* **54**, 1132–1139 (2010).
125. Oda, K., Matoba, Y., Kumagai, T., Noda, M. & Sugiyama, M. Crystallographic study to determine the substrate specificity of an L-serine-acetylating enzyme found in the D-cycloserine biosynthetic pathway. *J. Bacteriol.* **195**, 1741–1749 (2013).
126. Uda, N. *et al.* Establishment of an *in vitro* D-cycloserine-synthesizing system by using O-ureido-L-serine synthase and D-cycloserine synthetase found in the biosynthetic pathway. *Antimicrob. Agents Chemother.* **57**, 2603–2612 (2013).
127. Dietrich, D., van Belkum, M. J. & Vederas, J. C. Characterization of DcsC, a PLP-independent racemase involved in the biosynthesis of D-cycloserine. *Org. Biomol. Chem.* **10**, 2248–2254 (2012).
128. Kumagai, T. *et al.* Heme protein and hydroxyarginase necessary for biosynthesis of D-cycloserine. *Antimicrob. Agents Chemother.* **56**, 3682–3689 (2012).
129. Gould, S. J. & Ju, S. Biosynthesis of Acivicin. 3. Incorporation of Ornithine and N<sup>δ</sup>-Hydroxyornithine. *J. Am. Chem. Soc.* **114**, 10166–10172 (1992).
130. Kreuzer, J., Bach, N. C., Forler, D. & Sieber, S. A. Target discovery of acivicin in cancer cells elucidates its mechanism of growth inhibition. *Chem. Sci.* **6**, 237–245 (2014).
131. Duke, S. O. & Dayan, F. E. Modes of action of microbially-produced phytotoxins. *Toxins* **3**, 1038–1064 (2011).

132. Bowden, K., Drysdale, A. C. & Moge, G. A. Constituents of *Amanita muscaria*. *Nature* **206**, 1359–1360 (1965).
133. Matsumoto, T., Trueb, W., Gwinner, R. & Eugster, C. H. Isolierung von (–)-R-4-Hydroxypyrrolidon-(2) und einigen weiteren Verbindungen aus *Amanita muscaria*. 31. Mitteilung über Inhaltsstoffe von Fliegenpilzen. *HCA* **52**, 716–720 (1969).
134. Müller, M. & Obermaier, S. Ibotenic Acid Biosynthesis in the Fly Agaric Is Initiated by Glutamate Hydroxylation. *Angew. Chem. Int. Ed.* (2020).
135. Venepally, V. & Reddy Jala, R. C. An insight into the biological activities of heterocyclic–fatty acid hybrid molecules. *Eur. J. Med. Chem.* **141**, 113–137 (2017).
136. Salma, Y. *et al.* The natural marine anhydrophytosphingosine, Jaspine B, induces apoptosis in melanoma cells by interfering with ceramide metabolism. *Biochem. Pharmacol.* **78**, 477–485 (2009).
137. Yamahara, J., Kobayashi, G., Matsuda, H. & Fujimura, H. The vasorelaxant effect of evocarpine in isolated aortic strips: mode of action. *Eur. J. Pharmacol.* **155**, 139–143 (1988).
138. Kumagai, T. *et al.* High-Level Heterologous Production of D-Cycloserine by *Escherichia coli*. *Appl. Environ. Microbiol.* **81**, 7881–7887 (2015).