

# **Establishment and engineering of heterologous production platforms for cystobactamids and corramycins**

Dissertation

zur Erlangung des Grades

des Doktors der Naturwissenschaften

der Naturwissenschaftlich-Technischen Fakultät

der Universität des Saarlandes

von

**Sebastian Groß**

Saarbrücken

2020

Tag des Kolloquiums: 22.01.2021

Dekan: Prof. Dr. Jörn Walter

Berichterstatter: Prof. Dr. Rolf Müller

Prof. Dr. Andriy Luzhetskyy

Vorsitz: Prof. Dr. Alexander Titz

Akad. Mitarbeiter: Dr. Michael Kohlstedt

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

“Messieurs, c'est les microbes qui auront le dernier mot.”

**“Gentlemen, the microbes will have the last word.”**

Louis Pasteur

## Danksagung

Zuerst möchte ich mich bei meinem Doktorvater Prof. Dr. Rolf Müller dafür bedanken, dass er es mir ermöglicht hat in dieser Arbeitsgruppe meine Dissertation durchzuführen. Außerdem bedanke ich mich für das entgegengebrachte Vertrauen und die Möglichkeiten in den vielen interessanten und anspruchsvollen Projekten involviert sein zu dürfen. Die regelmäßigen wissenschaftlichen Diskussionen waren insbesondere gegen Ende meiner Promotion sehr hilfreich und motivierend.

Zudem möchte ich Prof. Dr. Andriy Luzhetskyy für die wissenschaftliche Begleitung und die Übernahme der Zweitkorrektur der Dissertation danken.

Ein ganz besonderer Dank geht an Dr. Silke Wenzel, die mich im ersten Jahr meiner Promotion betreute und mich mit ihrer motivierenden und geduldigen Art in die Projekte einarbeitete und somit maßgeblich am Gelingen einer Vielzahl an Experimenten beteiligt war. Ein weiterer Dank geht an Dr. Katja Gemperlein, die die Betreuung anschließend für kurze Zeit übernahm.

Desweiteren möchte ich mich bei allen ehemaligen und aktuellen Mitgliedern der Arbeitsgruppe Mikrobielle Naturstoffe bedanken, die mir in den letzten vier Jahren wissenschaftlich und organisatorisch zur Seite standen. Insbesondere danken möchte ich meinem Labornachbarn Dr. Domen Pogorevc dafür, dass er mich immer mit wissenschaftlichem Rat unterstützte und dass er meine Dissertation probegesehen hat. Selina Deckarm danke ich für die tatkräftige Unterstützung im Labor bei zahlreichen Klonierungsexperimenten. Ellen Merckel und Christina Decker sowie den technischen Assistenten möchte ich für die organisatorische Unterstützung danken.

Ein großer Dank geht außerdem an Jang Schlemmer, Patrick Alexander Haack, Dr. Fabian Panter, Chantal Bader, Dr. Ronald Garcia, Dr. Joachim Josef Hug, Nicotine, Dr. Sebastian Adam, Dr. Alexander Popoff, Dr. Jake Haeckl, Sari Rasheed, Joy Birkelbach und Sebastian Walesch für die humorvolle Zeit und die kreativen künstlerischen Verwirklichungen auf meiner Laborbank.

Zuletzt danke ich meiner Familie und Eva dafür, dass sie immer hinter mir stehen und mich unterstützen.

## Zusammenfassung

Die Entdeckung und Entwicklung neuartiger Antibiotika ist bedingt durch steigende bakterielle Resistenzen unabdinglich für die moderne Medizin. In den letzten Jahrzehnten wurde für Myxobakterien ein großes Potenzial zur Herstellung antibakterieller Naturstoffe gezeigt. Diese Dissertation konzentriert sich auf die Entwicklung, Assemblierung und heterologe Expression modifizierter biosynthetischer Gencluster zweier kürzlich entdeckter myxobakterieller Naturstoffklassen mit antibiotischer Wirkung. Die Etablierung einer heterologen Produktionsplattform im myxobakteriellen Modellstamm *Myxococcus xanthus* für Cystobactamide, lineare Peptide, die von einer nichtribosomalen Peptidsynthetase (NRPS) gebildet werden, führte zur Identifizierung bisher unbekannter natürlicher Analoga. Gezielte Deletionen von Genen und einer NRPS-Domäne führten zur Produktion unnatürlicher Cystobactamide und lenkten das Produktionsprofil hin zu wertvollen Derivaten. Zudem konnten zuvor unerklärte Biosyntheseschritte entschlüsselt werden, die die Produktion der völlig neuartigen Cystobactamid-Linkereinheit unter Beteiligung einer bifunktionalen Aminomutase- und Amid-Dehydratase-NRPS-Domäne erklären. Eine weitere heterologe Produktionsplattform in *M. xanthus* wurde für die peptidische Stoffklasse der Corramycine entwickelt. Die Manipulation der NRPS-Maschinerie führte neben der Produktionssteigerung zur Herstellung von verkürzten Corramycinen, die zur Semisynthese verbesserter Analoga verwendet werden können.

## Summary

The increase in bacterial resistance against antibiotics makes the discovery and development of novel compound classes a goal of utmost importance for the human health care system. In the last decades, myxobacteria were shown to have a great potential to produce antibacterial natural products. This thesis focuses on the design, assembly and heterologous expression of modified biosynthetic gene clusters of two recently discovered myxobacterial compound classes exhibiting antimicrobial activity. The establishment of a heterologous production platform in the myxobacterial model strain *Myxococcus xanthus* for cystobactamids, linear peptides formed by a nonribosomal peptide synthetase (NRPS), resulted in the identification of previously unknown natural analogs. Targeted gene and NRPS domain deletions led to the production of unnatural cystobactamids and directed the production profile towards advantageous derivatives. Additionally, the obscure biosynthesis steps of the unprecedented cystobactamid linker moiety were deciphered, involving a bifunctional aminomutation amide dehydratase NRPS domain. Another heterologous production platform in *M. xanthus* was developed for the peptidic compound class of corramycins. The manipulation of the NRPS assembly line resulted not only in an increase of production but also directed formation of truncated analogs usable for the semi-synthesis of improved congeners.

### **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:

### **Publikationen**

**Groß, S.\***, Schnell, B.\* , Haack A. P., Auerbach, D., Deckarm S. & Müller, R. *In vivo* and *in vitro* reconstitution of unique key steps in cystobactamid antibiotic biosynthesis. *Manuscript under revision in Nature comm.* (2020).

(\*: Authors contributed equally to the manuscript)

### **Tagungsbeiträge**

**Groß, S.**, Wenzel, S. C., Gemperlein, K. & Müller R. Design, Assembly and Heterologous Expression of a Synthetic Gene Cluster for the Production of Cystobactamids. **Poster Presentation**, 8<sup>th</sup> international HIPS symposium, Saarbrücken, Germany (2018).

**Groß, S.**, Wenzel, S. C., Gemperlein, K. & Müller R. Design, Assembly and Heterologous Expression of a Synthetic Gene Cluster for the Production of Cystobactamids. **Poster Presentation**, Summer Symposium of Interdisciplinary Graduate School, Saarbrücken, Germany (2018).

**Groß, S.**, Wenzel, S. C., Gemperlein, K. & Müller R. Design, Assembly and Heterologous Expression of a Synthetic Gene Cluster for the Production of Cystobactamids. **Poster Presentation**, International VAAM workshop, Jena, Germany (2019).

**Publikationen, die nicht Teil dieser Arbeit sind**

Lukežič, T., Fayad, A. A., Bader, C., Harmrolfs, K., Bartuli, J., **Groß, S.**, Lešnik, U., Hennesen, F., Herrmann, J., Pikel, S., Petković, H. & Müller, R. Engineering Atypical Tetracycline Formation in *Amycolatopsis sulphurea* for the Production of Modified Chelocardin Antibiotics. *ACS Chem. Biol.* **14**, 3, 468-477; 10.1021/acscchembio.8b01125 (2019).

## Table of Contents

1	Introduction.....	1
1.1	Natural products.....	1
1.2	Antibiotics and antimicrobial resistance.....	4
1.3	Myxobacteria as producers of NPs.....	9
1.4	Biosynthesis of natural products by multifunctional megasynthetases.....	13
1.5	Polyketide Synthases (PKSs).....	15
1.6	Nonribosomal peptide synthetases (NPRSs).....	17
1.6.1	Starter modules in NRPS biosynthesis.....	20
1.6.2	NRPS chain termination.....	20
1.6.3	Tailoring in NRPS biosynthesis.....	22
1.6.4	Combinatorial biosynthesis and NRPS engineering.....	24
1.7	Assembly and heterologous expression of complex natural product biosynthetic pathways .....	28
1.8	Outline of this work.....	32
1.9	References.....	34
2	<i>In vivo</i> and <i>in vitro</i> reconstitution of unique key steps in cystobactamid antibiotic biosynthesis .....	61
2.1	Abstract.....	61
2.2	Introduction.....	61
2.3	Results.....	64
2.3.1	Design, assembly, heterologous expression and manipulation of the cystobactamid gene cluster in <i>Myxococcus xanthus</i> DK1622.....	64
2.3.2	Biosynthesis of the linker moiety.....	66
2.3.3	Shuttling of the linker moiety to the assembly line by CysB.....	73
2.3.4	Revision of the complete cystobactamid biosynthesis model.....	75
2.4	Discussion.....	77
2.5	References.....	79

2.6	Methods.....	84
2.6.1	Cultivation of strains .....	84
2.6.2	<i>In silico</i> experiments and revision of the native BGC sequence .....	84
2.6.3	DNA synthesis and BGC and pMYC assembly .....	85
2.6.4	Genetic manipulation of expression constructs .....	85
2.6.5	Transformation of <i>M. xanthus</i> DK1622 and verification by colony PCR.....	85
2.6.6	Sample preparation and UPLC-ESI-HRMS analysis .....	86
2.6.7	Quantification of Cys919-1 production.....	86
2.6.8	Protein overexpression and purification.....	87
2.6.9	Substrate loading experiments.....	88
2.6.10	Direct intact protein UPLC-ESI-MS analysis .....	88
2.6.11	Cysteamine unloading assay and HPLC-MS analysis.....	89
2.7	Methods References .....	89
2.8	Supplementary Information .....	90
2.8.1	<i>In silico</i> analysis and revision of the native cystobactamid BGC sequence.....	90
2.8.2	<i>In silico</i> design of the modified BGC.....	91
2.8.3	<i>In silico</i> design of the cloning and expression vector system pMYC .....	91
2.8.4	Assembly of the modified BGC .....	92
2.8.5	Identification of the new cystobactamid derivatives .....	93
2.8.6	Analysis of CysJ activity on free L-asparagine using TLC .....	95
2.8.7	Synthesis of the di(ethylcarbonyl)asparaginyl-dicysteamine references.....	96
2.8.8	Supplementary Figures.....	97
2.8.9	Supplementary Tables .....	115
2.9	Supplementary Information References.....	129
3	Corramycin – a novel antibiotic class from myxobacteria hijacking inner membrane transporters of <i>Enterobacteriaceae</i> : Isolation, structure elucidation, biosynthesis, total synthesis and <i>in vivo</i> activity in infected mice.....	133
3.1	Abstract .....	133
3.2	Introduction.....	133

---

3.3	Results & discussion .....	135
3.3.1	Identification and isolation of a novel antibiotic class from myxobacteria.....	135
3.3.2	Elucidation of the flat structure and identification of corramycin derivatives .....	136
3.3.3	Identification of the corramycin biosynthetic gene cluster .....	138
3.3.4	Elucidation of the corramycin biosynthesis .....	140
3.3.5	<i>In silico</i> analysis-based prediction of eight stereocenters and elucidation of the absolute configuration of corramycin.....	143
3.3.6	Total synthesis of Cor1183.....	145
3.3.7	Frequency of resistance and corramycin uptake .....	149
3.3.8	<i>In vivo</i> activity of Cor1183 in infected mice .....	150
3.4	Conclusion and Outlook.....	152
3.5	Methods.....	153
3.5.1	Cultivation of strains .....	153
3.5.2	Sample preparation and UPLC-ESI-HRMS analysis .....	153
3.5.3	Purification and structural determination of Cor1183, Cor1199 and Cor1347 .....	154
3.5.4	Isolation of genomic DNA and illumina sequencing .....	155
3.5.5	<i>In silico</i> experiments .....	155
3.5.6	Protein purification and malachite green assay of the FAAL.....	156
3.5.7	Heterologous expression of <i>comD</i> and <i>comG</i> .....	156
3.6	References.....	157
3.7	Supplementary Information .....	163
3.7.1	<i>In silico</i> analysis of the corramycin BGC.....	163
3.7.2	Analysis of methyltransferase domains.....	163
3.7.3	Supplementary Figures.....	164
3.7.4	Supplementary Table.....	171
3.8	Supplementary Information References.....	172
4	Heterologous production of corramycins and their derivatives to allow for efficient semi-synthesis of improved congeners .....	175
4.1	Abstract.....	175

---

4.2	Introduction .....	175
4.3	Results & discussion .....	178
4.3.1	Design, assembly and heterologous expression of the corramycin biosynthetic gene cluster in <i>Myxococcus xanthus</i> DK1622 .....	178
4.3.2	Truncation of the assembly line to produce the N-terminal pharmacophore of corramycin .....	181
4.4	Methods.....	191
4.4.1	Cultivation of strains .....	191
4.4.2	<i>In silico</i> experiments.....	191
4.4.3	DNA synthesis and BGC assembly .....	192
4.4.4	Genetic manipulation of expression constructs .....	192
4.4.5	Transformation of <i>M. xanthus</i> DK1622 and verification by colony PCR.....	193
4.4.6	Sample preparation and UPLC-ESI-HRMS analysis .....	193
4.4.7	Relative quantification of truncated corramycins.....	194
4.4.8	Absolute quantification of the Cor666 production.....	194
4.4.9	MIC determination of Cor1183 against <i>M. xanthus</i> DK1622 .....	194
4.5	References .....	195
4.6	Supplementary Information .....	198
4.6.1	<i>In silico</i> analysis of the native corramycin BGC sequence .....	198
4.6.2	<i>In silico</i> design of the modified BGC.....	199
4.6.3	Assembly of the modified BGC .....	199
4.6.4	Relative quantification of truncated corramycin production.....	200
4.6.5	Supplementary Figures.....	201
4.6.6	Supplementary Tables .....	208
4.7	Supplementary Information References.....	213
5	Discussion.....	215
5.1	Major drawbacks in antibiotic development.....	215
5.2	Limitations and solution approaches in antibiotic drug discovery from microbial sources. .....	216

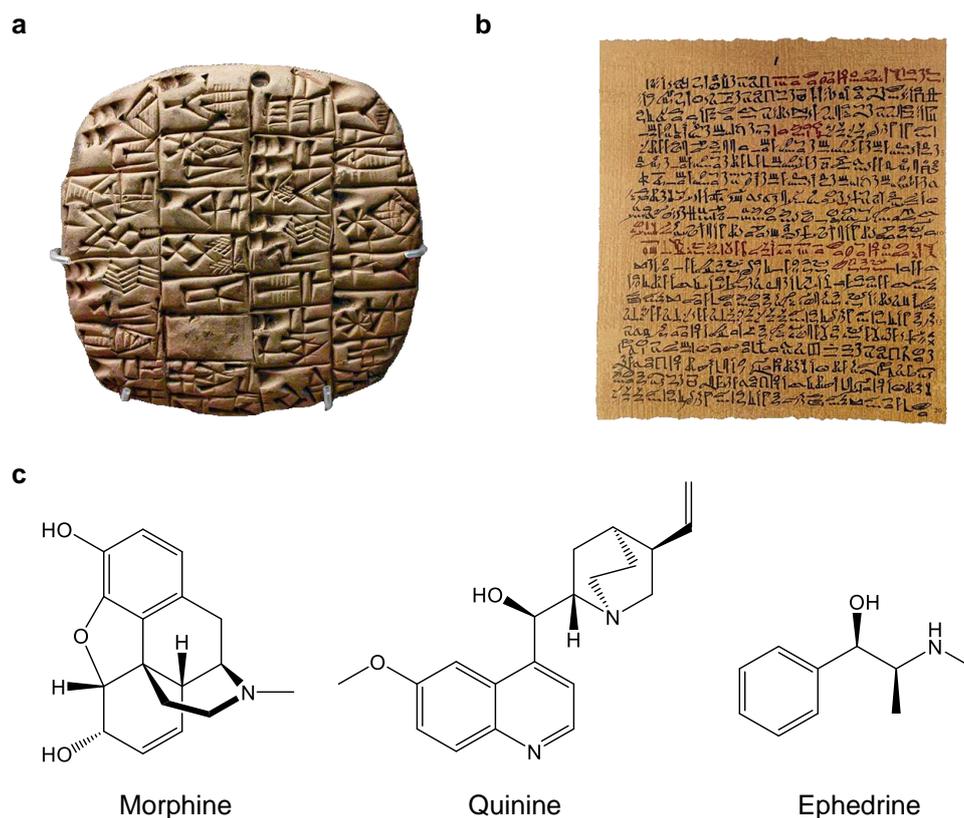
5.3	Heterologous expression of myxobacterial secondary metabolite pathways.....	220
5.4	Heterologous expression in the field of synthetic biology.....	227
5.5	Structure engineering of bacterial NPs .....	228
5.5.1	Semi-synthesis for structure diversification of NPs .....	229
5.5.2	Engineering of tailoring enzymes.....	232
5.5.3	Engineering of NRPS/PKS megasynthetases .....	237
5.6	Conclusion and Outlook.....	243
5.7	References.....	244

# 1 Introduction

## 1.1 Natural products

In the broadest sense, a natural product (NP) is a chemical substance that is produced by a biological system.<sup>1</sup> NPs are a large group of small molecules exhibiting an enormous structural variety and a broad spectrum of biological activities. However, since only a very small proportion of the entirety of NPs has been discovered today, the definition of a natural small molecule was designated as a “moving target”.<sup>2</sup> Narrower definitions relate NPs directly to secondary metabolites that are produced to gain evolutionary advantage over competitors but are not known to be involved in the internal economy of the organism.<sup>3</sup> In the public eye, a product of natural origin is often restricted to its production by an unaltered natural source. However in science, a substance found in nature is considered a NP, even if it was produced via chemical synthesis in a laboratory or by a metabolically engineered organism, as long as the resultant compound is chemically identical.

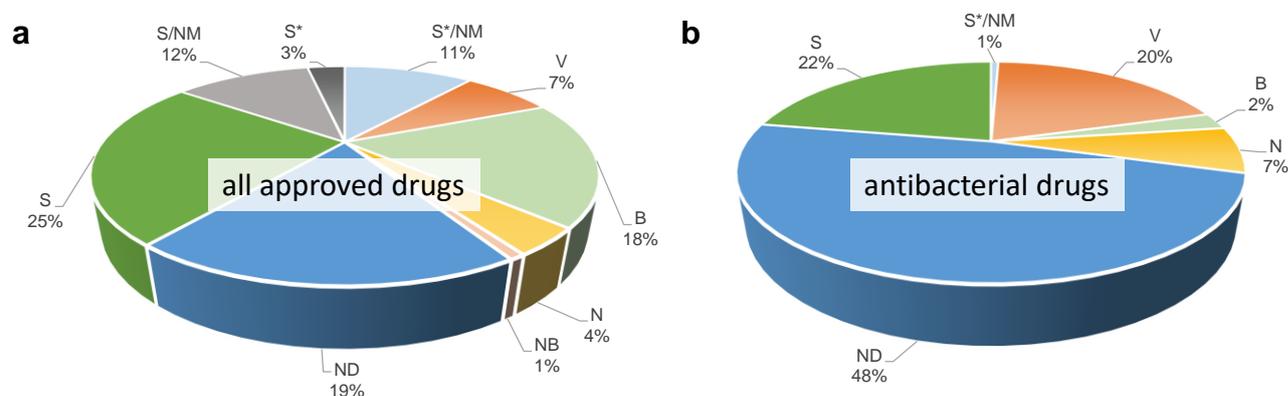
The medicinal use of NPs by humans has a long history. Mankind already used Nature as a source for medicinal products for millennia as there is evidence of plant material use in the Middle Paleolithic age.<sup>4</sup> The first historical records on Sumerian clay slab (Figure 1a) from Nagpur are approximately 5,000 years old documenting a dozen recipes for drug preparation based on 250 different plants.<sup>4-6</sup> Those substances were used in their crude form like oils from cedar and cypress or poppy juice and are still in use today to treat cough, colds, parasitic infections and inflammation. The Ebers Papyrus (16<sup>th</sup> century B.C.) was the first extensive collection of ancient Egyptian medical knowledge including around 800, mostly plant-derived, complex prescriptions for the treatment of a variety of illnesses like parasitic infections, injuries, tooth discomfort or migraine (Figure 1b).<sup>7</sup> The use of plants in traditional medicine have been extensively documented in China, India and many other cultures for thousands of years.<sup>8-12</sup> The medicinal knowledge of the ancient Western world in the Dark and Middle ages (5<sup>th</sup> to 12<sup>th</sup> centuries A.D.) was preeminently conserved by monasteries in England, Ireland, France and Germany, whilst the Arabs preserved the Greco-Roman expertise and even expanded their repertoire by Chinese and Indian plants.<sup>8</sup> In the modern era of medicine, the development in the field of chemistry allowed the isolation of the active ingredients from complex plant-derived mixtures that were used for centuries. In the 19<sup>th</sup> century various NPs were isolated that are still of vital importance in today’s medicine. For example the analgesic alkaloid morphine (1816), the antimalarial alkaloid drug quinine (1820), or ephedrine (1887), which served as starting point for anti-asthma agents (Figure 1c).<sup>8</sup> Although, advances in chemistry paved the way for modern medicine, the World Health Organization (WHO) estimated in 1985 that approximately 80 % of the world population still relied on traditional medicine to meet their primary health care demand.<sup>13</sup>



**Figure 1 | Sumerian clay slab, Ebers papyrus and examples of the first important medicinal NPs isolated in the 19<sup>th</sup> century.** **a:** Sumerian clay slab as it was used to document plant-derived recipes approximately 2,600 years B.C. Picture modified from: [https://s3.amazonaws.com/s3.timetoast.com/public/uploads/photos/5357336/cuneiform\\_660.jpg](https://s3.amazonaws.com/s3.timetoast.com/public/uploads/photos/5357336/cuneiform_660.jpg) (accessed 15.06.2020) **b:** Ebers papyrus documenting Egyptian medical knowledge 1,600 years B.C. Picture modified from: [https://commons.wikimedia.org/wiki/File:G.\\_Ebers\\_\(ed.\)\\_Papyrus\\_Ebers,\\_1875\\_Wellcome\\_L0016592.jpg](https://commons.wikimedia.org/wiki/File:G._Ebers_(ed.)_Papyrus_Ebers,_1875_Wellcome_L0016592.jpg) (accessed 15.06.2020) **c:** structures of morphine, quinine and ephedrine.

Technological advances in screening approaches lead to an enormous increase in NP discovery in the 20<sup>th</sup> century, reaching a maximum of 1,600 newly published compounds per year between the 1970s and 1980s.<sup>14</sup> The discovery rate slightly declined in the 1990s due to a change in screening methodology towards a high-throughput screening manner using large, ‘screen-friendly’ collections (libraries) of synthetic chemicals generated by combinatorial chemistry.<sup>15</sup> While screening of synthetic libraries could be performed with pure compounds, NP screening libraries were mainly based on crude extracts resulting in time- and resource-consuming downstream processes like bioactivity-guided isolation, production scale-up, extraction and purification steps.<sup>15</sup> Further obstacles in NP screening are the rapid identification of known compounds (dereplication), *de novo* structure elucidation and the fact that the potential target compounds often represent only a minor fraction in the crude extract (1 % or less of the total weight).<sup>15</sup> However, despite those disadvantages regarding the screening of NPs, roughly 58 % of antibacterial new chemical entities and 67 % of all approved drugs were NPs or their derivatives, synthetic drugs with NP pharmacophore, NP mimics or biological macromolecules in the period from 1981 to 2019 (Figure 2).<sup>16</sup> The combinatorial chemistry libraries, on the other hand, exhibited only very low hit rates, even though they were more

practical for high-throughput screening approaches. One reason for this could be that they were designed based on chemical accessibility and maximum achievable size, thereby excluding biologically relevant properties that are required in drug development.<sup>17</sup> The high proportion of NPs or NP-based compounds that are approved as drugs also justifies the question why NPs are that successful as template in drug development.



**Figure 2 | All drugs and antibacterial drugs approved between 1981 and 2019 by source.** B: biological macromolecule; NB: botanical drug; N: unaltered natural product; ND: natural product derivative; S: synthetic drug; S\*: synthetic drug with NP pharmacophore; V: vaccine; /NM: mimic of natural product. Definitions and values adapted from Ref. <sup>16</sup>.

NPs were shown to exhibit higher chemical diversity, biochemical specificity and other molecular properties making them favourable over synthetic chemicals as lead structures in drug development.<sup>15</sup> Analyses of NP libraries revealed a higher number of chiral centers, increased steric complexity, more oxygen but less nitrogen, sulphur and halogen groups compared to combinatorial synthesis libraries.<sup>18–20</sup> Nature also seems to favour aliphatic over aromatic rings since the ratio of aromatic rings to total heavy atoms is significantly lower in NPs, but they contain a larger fraction of  $sp^3$ -hybridized bridgehead atoms leading to less planar geometry and greater molecular rigidity.<sup>18,20–22</sup> Interestingly, around 20 % of ring systems in NPs are also found in trade drugs<sup>21</sup> and the rate of NP violations with the “Lipinsky rule of five”<sup>23</sup> equals approved drugs.<sup>20</sup> NPs exhibit well-defined 3D structures and functional groups, which enable precise interactions with specific conserved targets even though they are of completely different genetic origin.<sup>24</sup> However, apart from binding only to the active site of a certain enzyme, NPs are able to modify or inhibit protein-protein interactions by addressing multiple targets or binding-sites.<sup>15</sup> Therefore, NPs are considered privileged structures, which went through nature’s evolutionary screening process to optimize biological activity and are thus validated starting points for the design of screening-libraries.<sup>15,25</sup>

In the last two decades, a phase of NP-based drug discovery re-emerged in which combinatorial chemistry and diversity-oriented synthesis<sup>26</sup> offered the synthesis of structurally diverse screening libraries around a common NP-derived core structure.<sup>27</sup> The identification of the NP pharmacophore and its cellular target is thereby a critical step as it opens the door for fragment-based drug discovery<sup>28</sup> and fragment-based *de novo* design<sup>29</sup> with stepwise molecular linking of functional moieties.

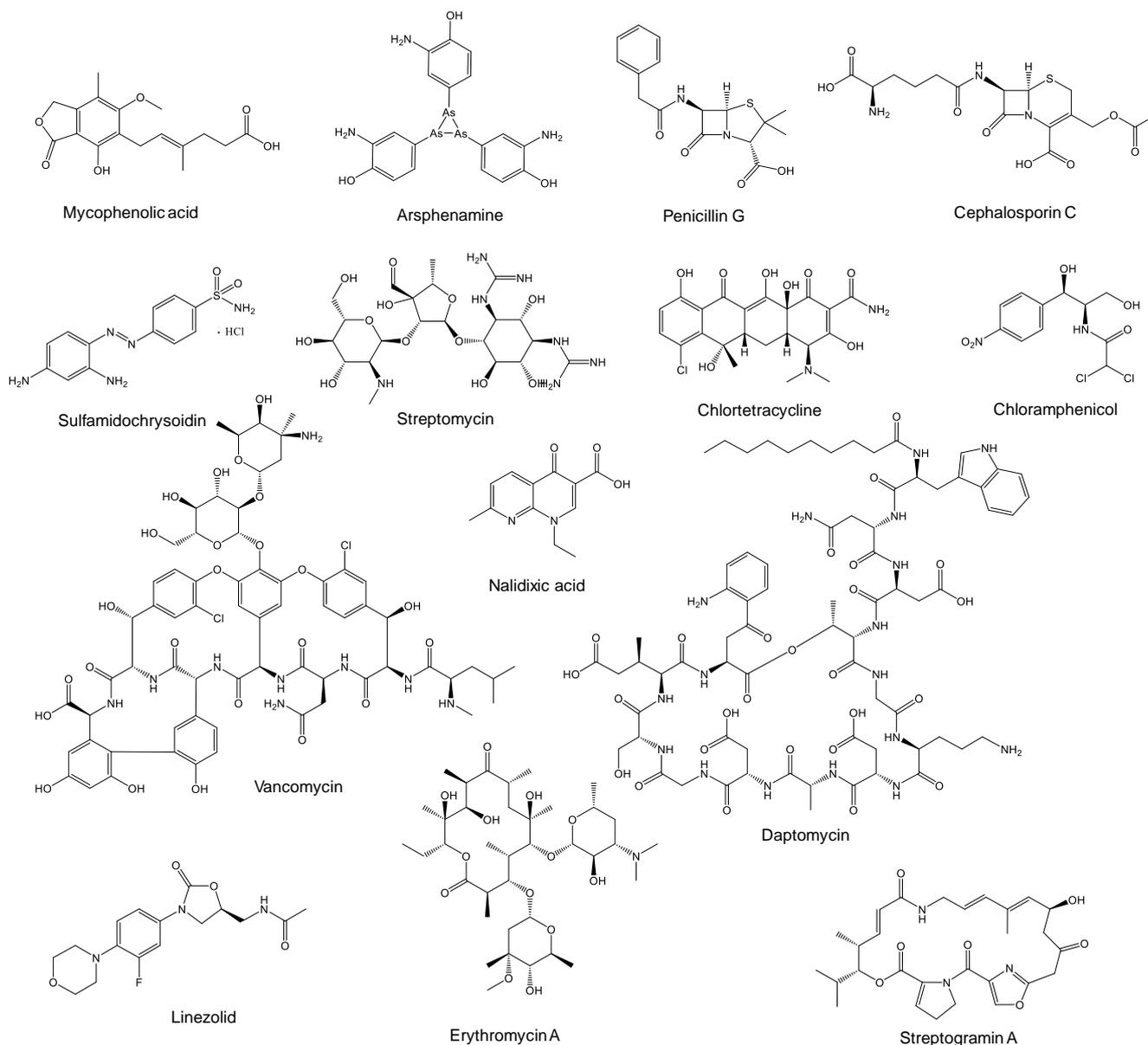
## 1.2 Antibiotics and antimicrobial resistance

For thousands of years infections represented the leading cause of death in humans, not infrequently with periodic outbreaks in epidemic proportions killing millions of people. It lasted until 1890, when Paul Vuillemin was the first to use the word ‘antibiose’ to describe the antagonistic action between different microorganisms.<sup>30,31</sup> Later, the term ‘antibiotic’ was derived from antibiose and was used for NPs that are produced by microorganisms and exhibiting either growth inhibitory (bacteriostatic) or killing (bactericidal) properties against bacteria or fungi. Today, the term is used in a broader sense describing designed molecules with specific activity against bacteria and fungi, but not viruses.<sup>30,32</sup> The first causative relationship between microbes and diseases was postulated by the German physician Robert Koch (together with Friedrich Loeffler based on the work of Jakob Henle) who later also isolated the causative bacteria *Mycobacterium tuberculosis* and *Vibrio cholerae* (together with Bernhardt Fischer and Georg Gaffky).<sup>30,33–36</sup>

In 1893, mycophenolic acid was the first antibiotic to be discovered and isolated from *Penicillium gaucum*.<sup>30,37,38</sup> Despite its antibacterial, antifungal, antitumor and anti-psoriasis activities, it remained unnoticed until re-discovery in 1913 in the U.S and it took another 40 years to elucidate the structure (shown in Figure 3).<sup>39–41</sup> In the meantime, the first synthetic antibiotic agent arsphenamine (Salvarsan) was developed by Alfred Bertheim in the laboratory of Paul Ehrlich and approved as a drug in 1910.<sup>42</sup> Salvarsan and improved derivatives were used for nearly 40 years as standard therapy for the treatment of syphilis.<sup>30,43–46</sup> In 1928, Alexander Fleming recognized the growth inhibition of a *Staphylococcus aureus* colony on a petri dish which was contaminated with a fungus (*Penicillium notatum*, now *P. chrysogenum*). Fleming grew the fungus as a pure culture, confirmed the antibacterial activity against several pathogenic bacterial strains and concluded that *Penicillium* must have excreted an antibacterial substance, which he finally named penicillin (structure shown in Figure 3).<sup>47</sup> However, it took twelve more years until penicillin was isolated by Howard Walter Florey, Ernst Boris Chain and Norman Heatley, tested in a mouse model and used for the treatment in humans.<sup>30</sup> Penicillin and other  $\beta$ -lactams act bactericidally by inhibiting the enzyme serine-type D-alanyl-D-alanine carboxypeptidase, which is involved in bacterial cell wall synthesis. Another sub-class of  $\beta$ -lactams, the cephalosporins, were discovered in 1945 by

Giuseppe Brotzu and re-discovered in 1955 with the isolation of cephalosporin C, which was not used in the clinic but inspired synthetic efforts leading to clinically relevant analogs.<sup>48,49</sup> The first sulfonamide sulfamidochrysoidin was synthesized as a dye by Fritz Mietzsch and Josef Klarer at Bayer in 1932 and marketed as a drug in 1935.<sup>30</sup> In contrast to  $\beta$ -lactams, sulfonamides exhibit bacteriostatic activity by inhibiting the folic acid biosynthesis. This initial phase of antibiotic drug discovery and development ushered into the “golden era of antibiotics” from the 1940s to the 1960s during which all the major antibiotic classes of today were discovered, most of them originated from microorganisms.

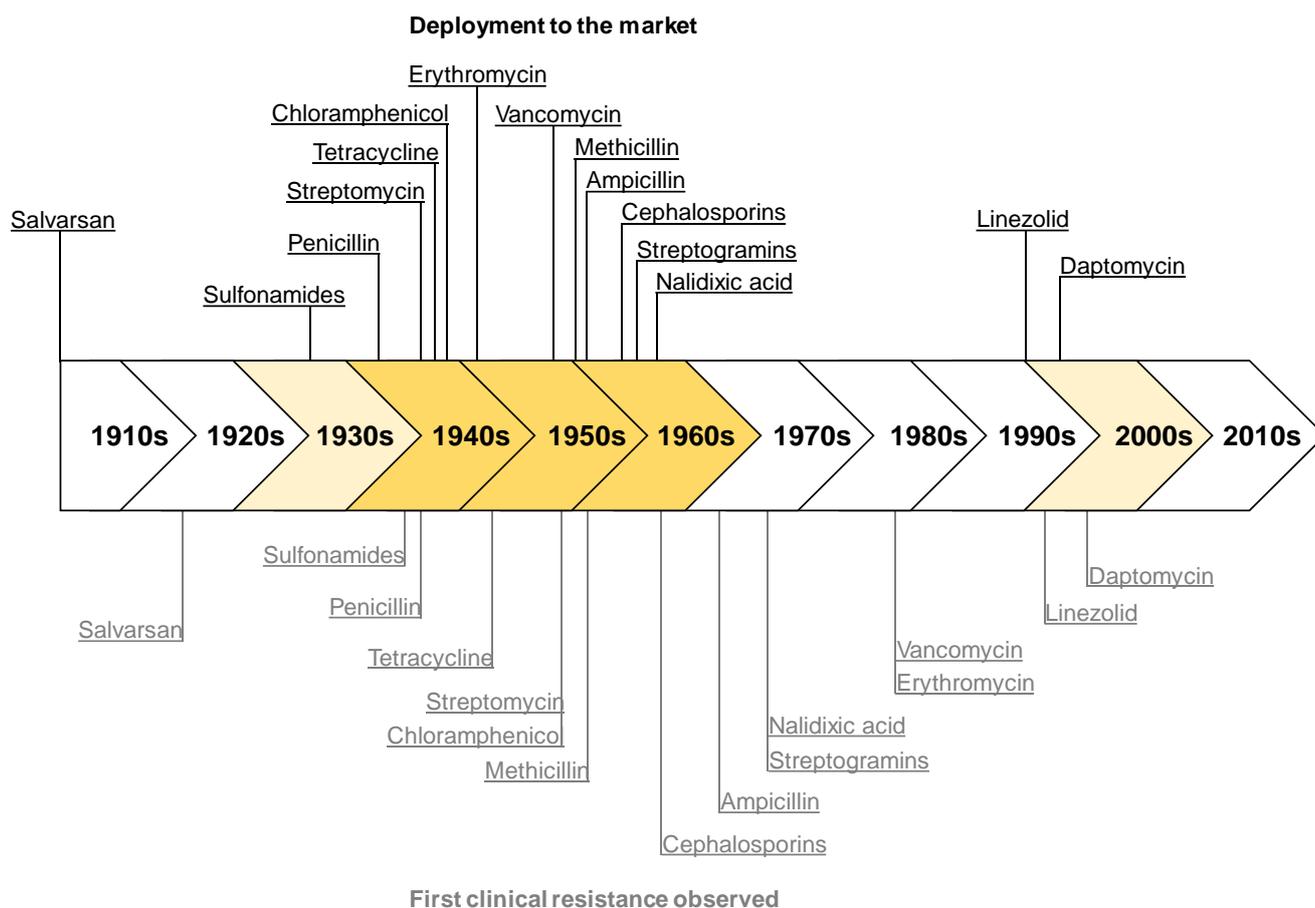
The first aminoglycoside antibiotic streptomycin was isolated from *Actinomyces griseus* (now *Streptomyces griseus*) in 1943 and approved for clinical use in 1946.<sup>50,51</sup> In 1945, the first tetracycline antibiotic chlortetracycline (formerly named aureomycin) was isolated from *Streptomyces aureofaciens* and approved in 1948 as broad-spectrum antibiotic with Gram-positive and Gram-negative activity. Both streptomycin and chlortetracycline exhibited bactericidal activity which was attributed to an inhibition of bacterial protein biosynthesis through binding to the 30S ribosome.<sup>52,53</sup> Shortly after, in 1947, the phenylpropanoid chloramphenicol was discovered, isolated from *Streptomyces venezuelae*, and already used in the clinic in the same year during a typhus epidemic in bolivia.<sup>30,54</sup> Chloramphenicol showed bacteriostatic activity by inhibition of the bacterial protein biosynthesis through binding to the 50S ribosomal subunit.<sup>55,56</sup> A similar mode of action was observed for the macrolide antibiotics erythromycins A and B which were isolated for the first time from *Saccharopolyspora erythraea* in 1952 and marketed as a mixture named Ilotycin in the same year.<sup>57,58</sup> The first glycopeptide antibiotic, later named vancomycin, was discovered and isolated from *Streptomyces orientalis* (now *Amycolatopsis orientalis*) in 1953, approved in 1958, and was shown to interfere with bacterial cell wall construction by binding to a D-Ala-D-Ala unit of a pentapeptide moiety in the peptidoglycan network.<sup>59,60</sup> In the same year streptogramin A and B were isolated from *Streptomyces graminofaciens*.<sup>61</sup> Interestingly, both compounds showed bacteriostatic activity by binding to the 50S ribosomal subunit, but combined administration resulted in synergistic effects and the development of bactericidal activity.<sup>62,63</sup> The first of the synthetic quinolone antibiotics, nalidixic acid, was synthesized in 1962 and was shown to target a subunit of the DNA gyrase and topoisomerase IV, thereby disrupting DNA biosynthesis.<sup>64–66</sup>



**Figure 3 | Structures of representative compounds of the main antibiotic classes.**

Since more and more antibiotics flooded the market in the “golden era of antibiotics”, the pharmaceutical industry began to lose interest in this field, followed by a steady decline in antibiotic research and development. In the past 20 years only two new antibiotic classes, the lipopeptides and the oxazolidinones, have been developed and eventually introduced to the international market.<sup>67</sup> The cyclic lipopeptide daptomycin (formerly A21978) was already isolated in the 1980s from *Streptomyces roseosporus* as a mixture of several compounds which were shown to inhibit the bacterial cell wall peptidoglycan biosynthesis in certain Gram-positive pathogens.<sup>68,69</sup> The first oxazolidinone was the synthetic drug linezolid which was shown to inhibit the protein biosynthesis in Gram-positive bacteria by binding to the 70S initiation complex.<sup>70</sup>

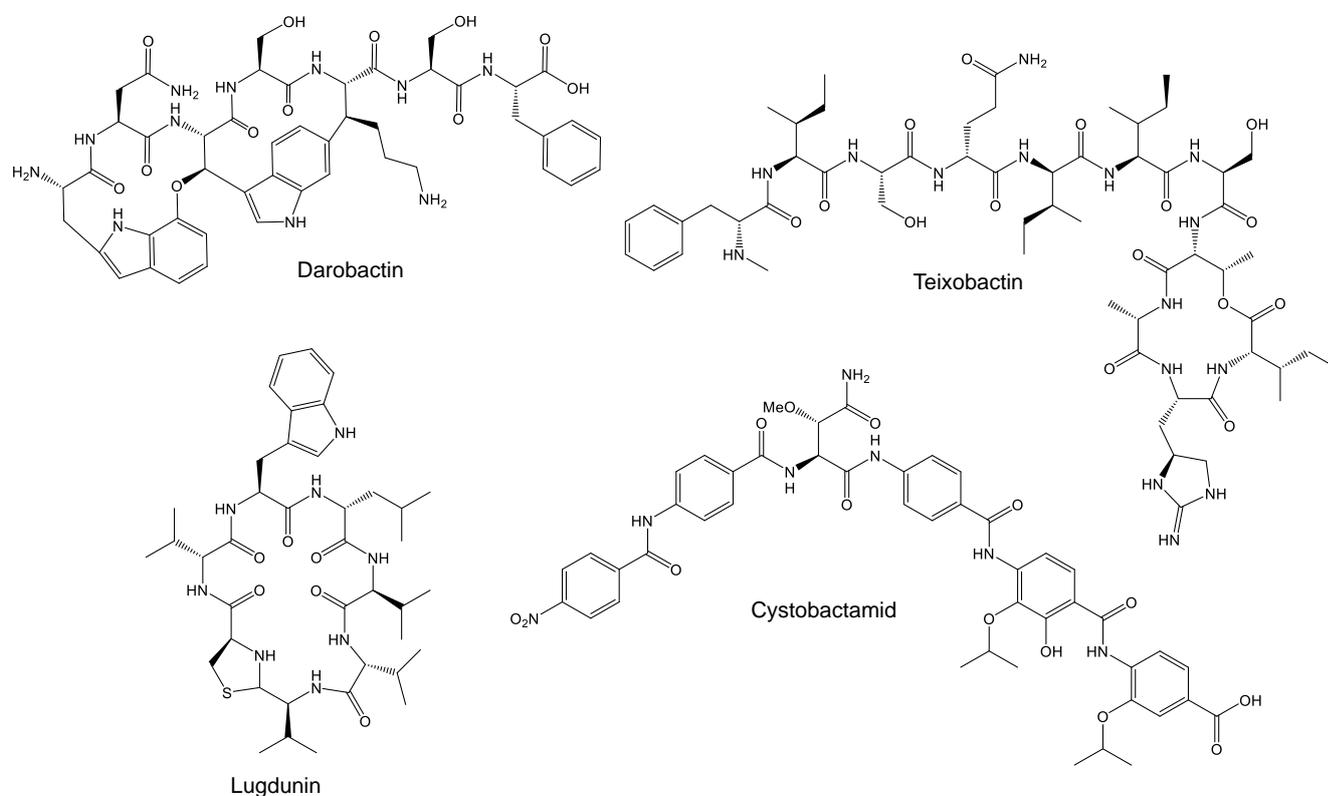
However, only a few years after market deployment antimicrobial resistance (AMR) has been observed for every antibiotic class against pathogenic bacteria that were initially susceptible to the respective antibiotics (Figure 4).<sup>71</sup> Even though it was shown that the evolution of antibiotic-resistant bacterial strains is a natural phenomenon occurring even without the anthropogenic use of antibiotics (primary resistance),<sup>72</sup> the acquired (secondary) resistance is one of the greatest challenges to the human health care system. Such secondary resistance development can be attributed to mutations leading to modifications of the biosynthetic pathways or alteration of target sites affected by the antibiotic as well as the inactivation of the antibiotic by specific enzymes or the increased expression of efflux transporters.<sup>73–75</sup> Lately, approximately 700,000 people died from infections caused by multidrug-resistant bacteria every year and it was predicted that around 10 million people might die in 2050 in case this trend is not stopped.<sup>76</sup> Especially pathogens belonging to the ESKAPE panel (vancomycin-resistant *Enterococci*, *Staphylococcus aureus*, *Klebsiella pneumoniae*, *Acinetobacter baumannii*, *Pseudomonas aeruginosa* and *Enterobacter* species), and in particular multidrug-resistant Gram-negative pathogens, cause difficult-to-treat nosocomial infections and were thus ranked with critical priority by the WHO.<sup>77–79</sup> Certain pathogenic strains developed resistance towards almost all approved antibiotics in the clinic.<sup>80</sup> However, the quinolones were the last novel antibiotic class introduced to the market with anti-Gram-negative activity.<sup>77</sup> In 2017, only around one third of the antibiotics in the development pipeline for intravenous use exhibited activity against Gram-negative bacteria<sup>77</sup> and all of them are modified agents from known antibiotics with known targets. Also most of the recently FDA-approved drugs derive from known antibiotic classes such as the  $\beta$ -lactam/ $\beta$ -lactamase inhibitor combinations ceftolozane/tazobactam, ceftazidime/avibactam,<sup>81</sup> meropenem/vaborbactam,<sup>82</sup> the aminoglycoside antibiotic plazomicin<sup>83</sup> and the tetracycline derivative eravacycline.<sup>84</sup> Reasons for that could be laborious discovery and development processes towards highly active and at the same time safe antibiotics with appropriate pharmacokinetic properties as well as difficult and expensive clinical trials. Drug design based on known antibiotic classes makes a rapid drug development but also resistance development more likely. Since AMR is only a question of time and all antibiotics therefore have a time-limited span of use, constant development of novel therapeutics with unprecedented modes of action is required to successfully fight infectious diseases.<sup>85</sup>



**Figure 4 | Timeline of market deployment and first clinical resistance of important antibiotics and antibiotic classes.** Deployment to the market is shown above the timeline (in black) and first clinical resistance observed is shown below the timeline (in grey). The “golden era of antibiotics” is colored in gold in the timeline and decades in which at least one new antibiotic class was introduced to the clinic is colored in light gold. References of the dates: Salvarsan,<sup>86</sup> streptogramins,<sup>87,88</sup> nalidixic acid,<sup>89,90</sup> linezolid,<sup>87,91,92</sup> daptomycin<sup>87,93–95</sup> and <sup>71</sup> if not further specified.

However, a number of new structures exhibiting new modes of action were recently described, for example the lipid II- (precursor of peptidoglycan) and lipid III-targeting (precursor of cell wall teichoic acid) antibiotic teixobactin (Figure 5), which was isolated from a previously uncultured bacterium of the genus *Aquabacteria* in 2015.<sup>96</sup> Likewise but only showing activity against Gram-positive pathogens, Lugdunin was isolated from the human commensal *Staphylococcus lugdunensis* in 2016.<sup>97</sup> Dissipation of the membrane potential in *S. aureus* demonstrated that the compound affects proton translocation.<sup>98</sup> Another recent example are the cystobactamids, which were first isolated from the myxobacterium *Cystobacter velatus* and show potent activity against a number of Gram-negative and multiresistant pathogens.<sup>99,100</sup> Cystobactamids are type II topoisomerase poisons similar to quinolones; however, they were suggested to have an overlapping but not identical binding site based on experiments testing cross-resistance with ciprofloxacin.<sup>99</sup> In 2019, the new cyclic peptide antibiotic darobactin was described, which was isolated from *Photobacterium kharii*.<sup>101</sup> Surprisingly, darobactin selectively kills Gram-negative pathogens by

binding to BamA, a chaperone and translocator of outer-membrane proteins. Thus, darobactin does not rely on permeating the outer-membrane, which is one of the major hurdle for antibiotics to overcome in Gram-negative bacteria.<sup>102</sup>



**Figure 5 | Structures of recently discovered antibiotics with new modes of action.** Shown are the prototypical structures of teixobactin, lugdunin, cystobactamid and darobactin as example of promising drug candidates with potential for development.

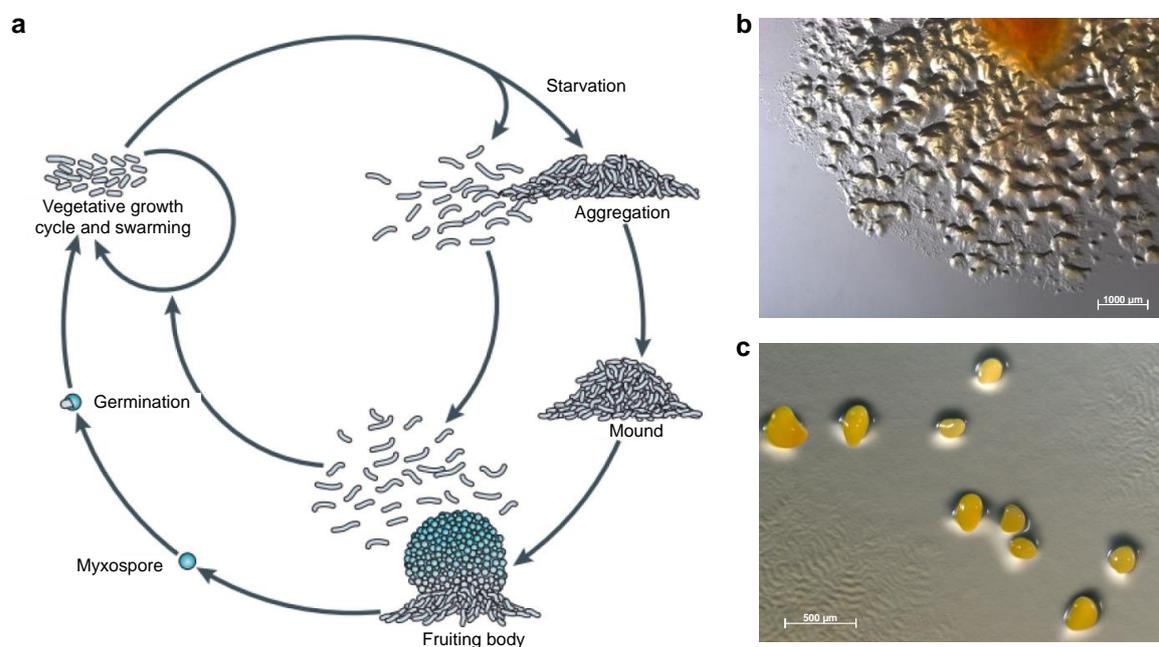
### 1.3 Myxobacteria as producers of NPs

Plants, fungi and eubacteria have always been the main sources of secondary metabolites. The metabolisms of plants and fungi are commonly associated with the production of alkaloids, polyketides, terpenes and non-ribosomal peptides (the latter not found in plants).<sup>103–105</sup> Furthermore, it is proposed that fungi, especially marine ones, are still an underrepresented resource of NPs.<sup>106</sup> Microorganisms in general, and bacteria in particular, have always been the leading force in production of antibiotics. Since bacteria compete for limited resources and space in all habitats around the world, evolution is the driving force for development of faster growth rates, complex social behaviour and production of toxic secondary metabolites to outcompete rival strains. Interestingly, only a small proportion of bacterial genera (2 % out of estimated  $1 \times 10^5$ ) have been classified based on 16S rRNA analyses and the vast majority of the bacteria have thus far not been cultured.<sup>107</sup> The most prominent bacterial source for anti-infectives are actinobacteria as they

produce two-thirds of all NP-derived antibiotics in current clinical use, but notably also anticancer, anthelmintic and antifungal compounds.<sup>108</sup> Starting with the isolation of streptothricin and streptomycin in the early 1940s, more than 5,000 antibiotics have been identified from actinobacteria until today.<sup>50,109,110</sup> The great potential of actinobacteria to produce valuable secondary metabolites with diverse bioactivities lies hidden in their genomes, which are extraordinarily large. Usually the number of genes encoding the protein machinery, which is necessary to produce a secondary metabolite, are located in close proximity in the genomes of their producers. These organized groups of genes are called biosynthetic gene clusters (BGCs). It was shown that the number of BGCs located in the genome positively correlates with the genome size, leading to an average of 20 to 50 BGCs for certain strains in the order of actinomycetales.<sup>111</sup> However, about 90 % of the BGCs predicted from genome sequencing are not or only weakly expressed when the respective strains are cultivated under laboratory growth conditions. Those BGCs are then referred to as cryptic or silent. The low-hanging fruits in the order of actinomycetales might already have been harvested because of excessive screening efforts in the past decades. For example, already a decade ago, more than 10<sup>7</sup> separate actinomycetes were screened per year at Cubist Pharmaceuticals for the production of antimicrobials.<sup>112</sup> One might think that the more screening efforts in the order of actinomycetales were made, the higher is the chance of their potential being depleted. However, it was estimated that solely the genus of *Streptomyces* is capable of producing antimicrobial compounds in the order of 100,000; which is only a tiny fraction of what has been unearthed so far.<sup>113</sup> Nevertheless, a change in the screening methodology, e.g. to get access to the NPs deriving from the production machinery encoded by silent BGCs, or the switch to a lesser explored bacterial order might increase the success rate of finding new antibiotics.

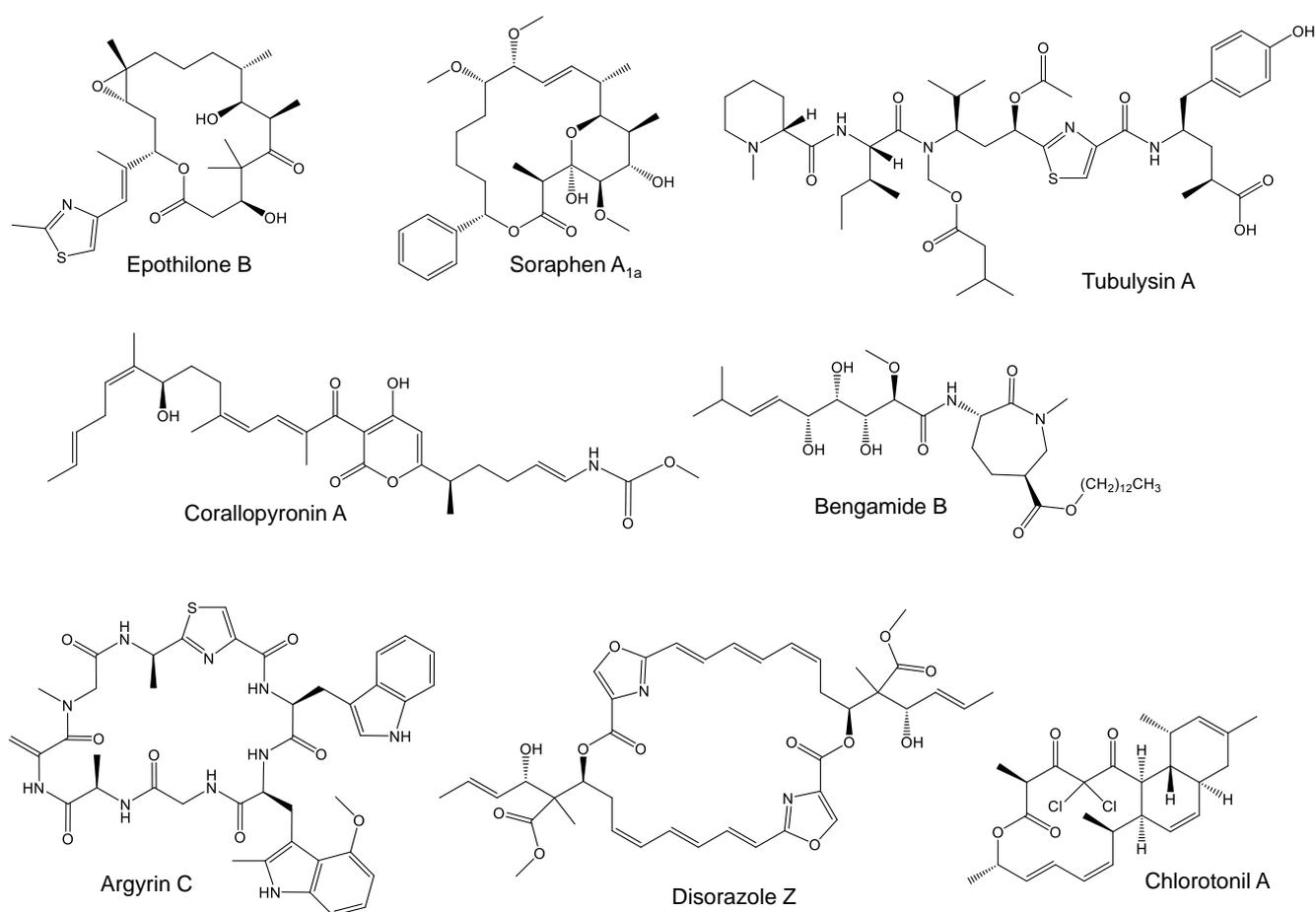
In the last decades, the bacterial order of myxococcales, the myxobacteria, came more and more into the focus of NP screening efforts. Starting with the description of the antifungal compounds ambruticin<sup>114</sup> and myxothiazol,<sup>115</sup> today already several hundreds of NPs with remarkable bioactivities and rare modes of action have been described.<sup>116-119</sup> Similar to actinomycetes, myxobacteria harbor the largest known genomes among the kingdom of bacteria, ranging up to 16 Mbp,<sup>120-122</sup> giving them great genetic capacity to encode secondary metabolite BGCs and develop complex social life behaviour. It was shown that each myxobacterial strain harbors several different secondary metabolite pathways, giving them the ability to produce numerous compound families.<sup>116</sup> One example is *Sorangium cellulosum* So ce12 producing chivosazols (cytotoxic, inhibit actin polymerization), disorazols (cytotoxic, antifungal, inhibit tubulin polymerization), sorangicins (antibacterial, inhibit RNA polymerase), sorangolids (antibacterial, disrupt membrane integrity) and sulfangolids (weak antibacterial).<sup>116,118,123-126</sup>

Myxobacteria are rod-shaped, highly GC-rich, Gram-negative bacteria belonging to the class of  $\delta$ -proteobacteria.<sup>127</sup> They are primarily soil organisms found in diverse biotopes as for example in steppes and deserts, arctic tundra or rain forests. They are often isolated from decaying plants or from herbivorous animal dung.<sup>116</sup> Notably, metagenome sequencing pointed to a widely distributed clade of marine myxobacteria; however, of those not a single strain has so far been isolated.<sup>128</sup> Apart from the outstanding capability of myxobacteria to produce NPs, they exhibit a complex life cycle (Figure 6a) including gliding and spreading over surfaces as multicellular swarms (Figure 6b) to hunt prey microorganisms cooperatively by secretion of NPs and enzymes.<sup>116</sup> Furthermore, myxobacteria are able to form so-called fruiting bodies by aggregation of hundreds of thousands of cells when they face nutritional limitations (Figure 6c). During the formation of fruiting bodies, myxobacteria undergo a differentiation process resulting in two possible types of cells: a monolayer of cells called peripheral rods<sup>129,130</sup> and the actual aggregate-forming cells building biofilm-like structures.<sup>120,131</sup> In the head of a fruiting body, resistant and reproductive myxospores are built within sporangioles. Spores are able to outlast the starvation conditions until sufficient nutrient conditions allow germination and formation of a swarm colony again.<sup>116</sup> Aggregated cells of the fruiting body often lose the ability to reproduce thus falling into a metabolic state of hibernation. Since monolayer cells surrounding the fruiting body often lyse, it was hypothesized that the released nutrients provide energy for the remaining cells to defend against competitive microorganisms.<sup>132</sup>



**Figure 6 | Myxobacterial life cycle. a:** Schematic life cycle of *Myxococcus xanthus*. Under starvation conditions vegetative cells aggregate into macroscopic mounds finally forming fruiting bodies. Persistent myxospores develop from rod-shaped cells inside the fruiting bodies. Germination of the myxospores occurs when nutrients are available again. Scheme adapted from Ref. <sup>133</sup>. **b:** Spreading of a multicellular *M. xanthus* DK1622 swarm on solid medium. **c:** Fruiting bodies of *M. xanthus* DK1622. Microscopic images were kindly provided by Dr. Ronald Garcia.

Currently, the FDA-approved drug ixabepilone,<sup>134</sup> a semisynthetic derivative of the myxobacterial NP epothilone (shown in Figure 7),<sup>135</sup> is already used in the clinic to treat breast cancer while several other compounds are in preclinical development owing to their promising modes of action. For example the antifungal macrolide sorafen A inhibits the eukaryotic acetyl-CoA carboxylase 1,<sup>136</sup> whereas the antimalarial macrolide chlorotoniil significantly reduced parasitemia *in vivo* in a *Plasmodium berghei* mouse model.<sup>137</sup> A bengamid derivative was studied in clinical trial phase I for its antiproliferative properties in cancer therapy<sup>138</sup> and also the antimicrotubulin tubulysin A<sup>139</sup> and disorazol Z,<sup>140</sup> which inhibit tubulin polymerization, showed promising results in preclinical evaluation. The antifilarial corallopyronins<sup>141</sup> is one more example of an auspicious compound class being in preclinical development. As previously mentioned, another promising candidate structure is provided by the highly potent, topoisomerase inhibitors cystobactamids (shown in Figure 5) that showed anti-Gram-negative activity against several clinically relevant pathogens like *Pseudomonas aeruginosa* and *Escherichia coli*.<sup>99,100</sup> Those examples underline the current potential of myxobacteria in drug discovery.



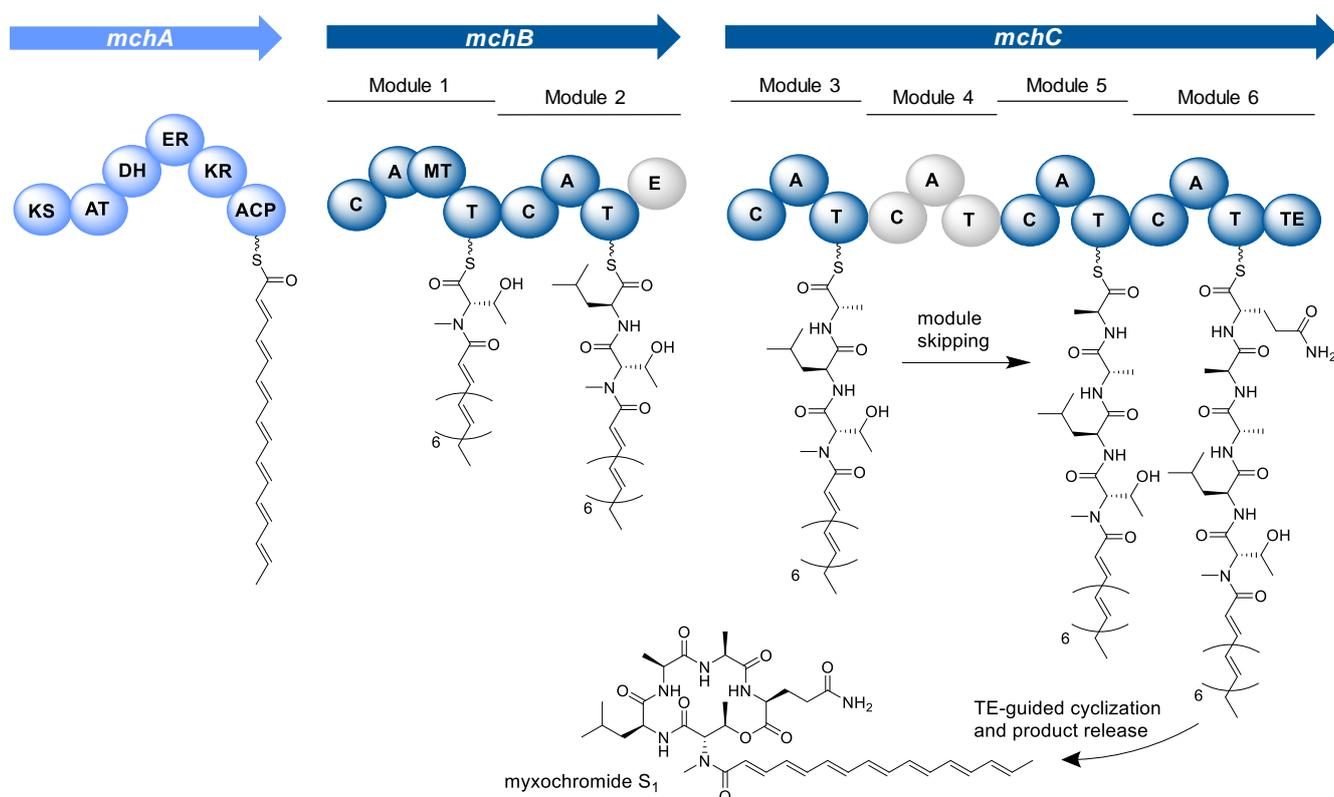
**Figure 7 | Structures of myxobacterial compounds with medically relevant properties.**

## 1.4 Biosynthesis of natural products by multifunctional megasynthetases

Complex secondary metabolites originate from biosynthetic pathways which use primary metabolism-derived simple monomeric building blocks. Linkage of a variety of those building blocks and numerous possible modifications allow for the creation of an immense chemical diversity in NP biosynthesis. Biosynthetic pathways are classified according to the type of building blocks used, as isoprenoids are used for terpene and isoprenoid biosynthesis, acetate-derived building blocks serve for polyketide (PK) biosynthesis and amino acids are utilized to form ribosomally produced and posttranslationally modified peptides (RiPPs), non-ribosomal peptides (NRPs) or alkaloids. The type and frequency of the pathways occurring in producers of NPs heavily depends on their taxonomy. In microorganisms, especially in bacteria, NP biosynthesis is predominantly executed via giant megasynthetase enzymes, the polyketide synthases (PKSs) and non-ribosomal peptide synthetases (NRPSs).<sup>142</sup>

Both PKSs and NRPSs exhibit a similar organization as assembly lines catalyzing step-wise coupling and modification of monomeric building blocks to a mature NP that is finally liberated. The megasynthetases are composed of specialized catalytic domains, each performing a specific function. The growing precursor chain is thereby handed over from one domain to the next. Those domains are connected by so-called ‘linker’ regions and are organized in functional units called modules.<sup>142</sup> The biosynthesis is initiated by the first module, the starter module, activating and covalently binding the first building block which is then sequentially extended by further building blocks incorporated by downstream modules. Covalent binding of the building block is mediated by a so-called ‘carrier protein’ domain, which has to be posttranslationally phosphopantetheinylated by a phosphopantetheinyl transferase (PPtase) to acquire its active *holo* form.<sup>143</sup> After reaching the terminal assembly line module, liberation of the product is achieved by cleaving the thioester bond, occasionally with further reduction, decarboxylation or cyclization reactions performed by a thioesterase (TE) domain.<sup>144–146</sup> Moreover, the compound intermediates are often modified during extension or after release from the assembly, either *in cis* by tailoring domains or *in trans* by independent tailoring enzymes.<sup>147</sup> A variety of tailoring modifications were described for NPs, such as methylation, hydroxylation, reduction, halogenation, glycosylation, cyclization and lipidation. PKSs and NRPSs use different building blocks which are acyl-coenzyme A (acyl-CoA) derivatives for the former and proteinogenic and non-proteinogenic amino acids as well as aryl acids for the latter.<sup>148</sup> However, the structural similarity of both systems allows their interplay in form of hybrid NRPS-PKS systems.<sup>148,149</sup> Notably, it was shown that hybrid NRPS-PKS systems represent a considerable number of NP biosynthetic pathways in myxobacteria.<sup>150</sup>

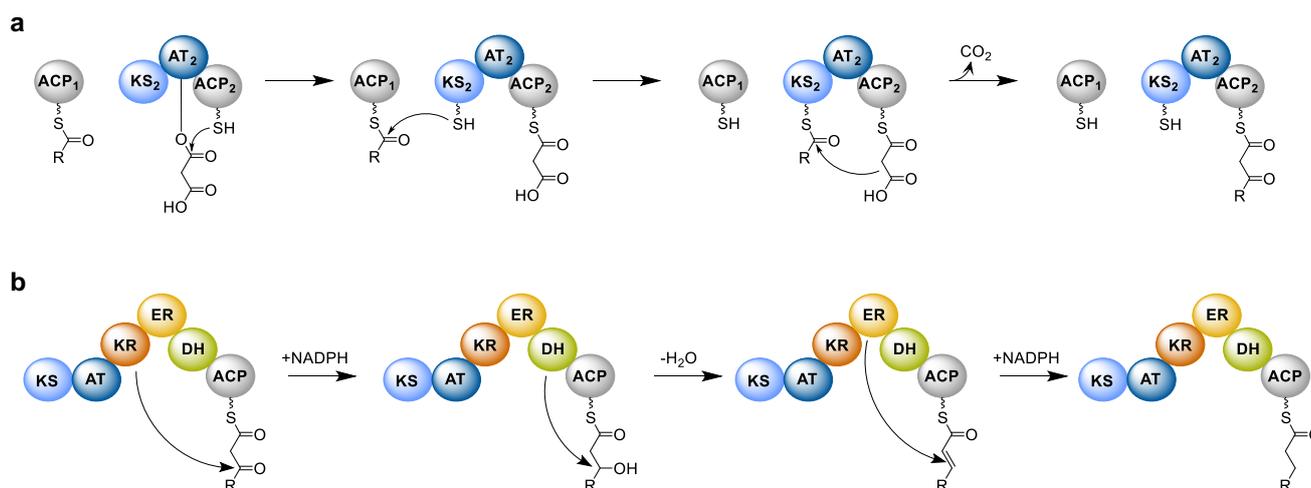
One example of a myxobacterial compound class produced by a megasynthetase are the myxochromides, which are cyclic lipopeptides produced by several *Myxococcus* and *Stigmatella* species.<sup>151–154</sup> As no cytotoxic or antimicrobial properties of myxochromides were detected,<sup>152,154</sup> an involvement in the developmental life cycle is discussed due to the prevalent production in numerous myxobacterial strains. Myxochromides are produced by a PKS-NRPS assembly line (Figure 8), which is encoded by three co-transcriptional genes: *mchA* encodes a monomodular, iterative type I PKS (described in the subsection PKS) and *mchB* and *mchC* encode NRPS subunits. The biosyntheses, for examples of myxochromide A and myxochromide S, slightly vary depending on the genetic origin of the PKS-NRPS system leading to penta-, hexa-, or heptapeptides linked to a polyunsaturated acyl chain with varying length. The biosynthesis (shown in Figure 8 on the example of myxochromide S)<sup>154</sup> includes some unusual features starting with MchA, which was proposed to synthesize the complete linear polyunsaturated PK in an iterative process. The NRPS subunits of MchB (modules 1 and 2) and MchC (modules 3 to 6) lead to the incorporation of the respective amino acids. Interestingly, module 4 is skipped, which was the first described example of module skipping in NRPSs.<sup>154</sup> Finally, liberation of the product is catalyzed by the TE of the terminal module after a cyclization reaction.



**Figure 8 | Biosynthesis of myxochromide S<sub>1</sub> by a PKS-NRPS megasynthetase.** PKS gene and domains (KS: ketosynthase, AT: acyltransferase, DH: dehydrogenase, ER: enoylreductase, KR: ketoreductase, ACP: acyl carrier protein) are shown in light blue, NRPS genes and domains (C: condensation, A: adenylation, T: thiolation, E: epimerization, TE: thioesterase) in dark blue or grey (presumably inactive domains). Module 4 skipping leads to direct transfer from the T domain of module 3 to the T domain of module 5. Scheme adapted from Ref. <sup>154</sup>.

## 1.5 Polyketide Synthases (PKSs)

The basic biosynthetic steps performed by a PKS are similar to those performed by fatty acid synthases (FASs) as both systems share an evolutionary connection.<sup>155</sup> In the process, a basic PK chain is formed by repetitive Claisen thioester condensation reactions linking simple acyl-CoA precursors, which are most commonly malonyl-CoA, methylmalonyl-CoA, acetyl-CoA or propionyl-CoA.<sup>156</sup> Therefore, typically three catalytic domains are required: The acyltransferase (AT) domain selecting starter and extender units, the phosphopantetheinylated acyl carrier protein (ACP) for covalent binding and shuttling of the building blocks, and the ketosynthase (KS) domain catalyzing the actual C-C bond forming condensation reaction (Figure 9).<sup>156</sup> The resulting basic PK chain contains various  $\beta$ -keto groups which can be further processed and even fully reduced by the action of three additional domains: the ketoreductase (KR) domain reduces the  $\beta$ -keto group generating a  $\beta$ -hydroxyl group, whereas the dehydratase (DH) domain dehydrates the  $\beta$ -hydroxyl group forming a C-C double bond, which can be further reduced by the enoylreductase (ER) domain.<sup>157</sup>



**Figure 9 | Polyketide chain elongation and optional  $\beta$ -keto processing steps.** **a:** mechanism of polyketide chain elongation. ACP: acyl carrier protein (grey); KS: ketosynthase (light blue); AT: acyltransferase (dark blue). **b:** optional  $\beta$ -keto processing steps in PKS or fatty acid biosynthesis. KR: ketoreductase (orange); ER: enoyl reductase (yellow); DH: dehydratase (green).

In fatty acid biosynthesis each monomer is commonly fully reduced including the action of all mentioned domains, while in PK biosynthesis the reaction of the latter three are optional. In comparison to fatty acids, this necessarily leads to a higher structural diversity in PKs including the presence of C-C double bonds and  $\beta$ -keto,  $\beta$ -hydroxyl or enoyl groups. Furthermore, additional tailoring domains such as C-methyltransferase (MT) domains contribute to the diversification of PKS products. Liberation of the final products is obtained by hydrolysis or lactonization by a TE domain which is part of the terminal module.<sup>158</sup> Depending on the enzymatic structure and

differences in their functionality, PKSs can be classified into modular type I and iterative type II or type III systems.

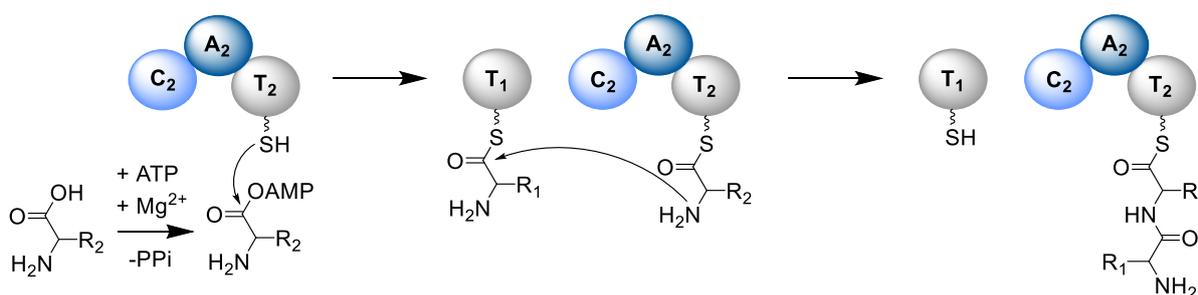
Type I PKSs are giant multimodular enzyme complexes in which the catalytic domains are covalently fused via linkers in a linear manner. Most frequently those systems are noniterative, meaning that the number of extension cycles and the degree of  $\beta$ -keto processing correlates with the number of modules carrying a KR domain.<sup>156</sup> Noniterative type I PKSs are mainly found in prokaryotes as for example the 6-deoxyerythronolide B synthase, which is involved in the biosynthesis of the macrocyclic erythromycin aglycon 6-deoxyerythronolide B.<sup>159</sup> Two examples of myxobacterial NRPS-PKS systems with involvement of noniterative type I PKSs are the melithiazol biosynthetic pathway, notably in which a dehydro-isobutyrate was described as starter molecule, and the epothilone pathway.<sup>160,161</sup> However, it is noteworthy that noniterative type I PKS systems have recently been described in protozoans as well.<sup>162,163</sup> On the other hand, iterative type I PKS systems include distinct families of bacterial origin, the mycocerosic acid synthase-like PKSs (MAS-like PKSs) and the fungal iterative type I PKSs.<sup>164</sup> As previously mentioned, there is evidence for an iterative type I PKS involved in the biosynthesis of the myxobacterial compound myxochromide S<sub>1-3</sub>.<sup>154</sup>

Type II PKS systems consist of a minimal set of dissociable, monofunctional and iteratively used enzymes ('minimal PKS'), including two KS units (KS $\alpha$  and KS $\beta$ ) and an ACP.<sup>165</sup> In this constellation KS $\alpha$  catalyzes the Claisen thioester condensation and KS $\beta$ , also referred to as 'chain length factor', determines the PK chain length. Further processing or folding of the nascent poly- $\beta$ -keto chain into an aromatic ring system is performed by PKS-associated enzymes like KR, DHs, aromatases and cyclases. Type II PKSs have exclusively been identified in prokaryotes and most of them in actinomycetes. Some examples of medicinally relevant compound classes are the anthracyclines, such as the anticancer drug doxorubicin, and the tetracycline antibiotics.<sup>166,167</sup> An example for a type II PKS-derived structure class from myxobacteria are the aurachins, which are rare quinoline alkaloids exhibiting various bioactivities such as antiparasitic, antibacterial, antifungal, cytotoxic and depletion of the mitochondrial membrane potential.<sup>168,169</sup>

Type III PKSs were found in plants, bacteria and fungi<sup>170-172</sup> and are similar to type II PKSs consisting of dissociable, monofunctional and iteratively used enzymes. However, type III PKS proteins form homodimers and use free CoA-activated thioester building blocks without using an ACP.<sup>173</sup> Characteristic compound classes produced by type III PKSs are flavonoids and stilbenoids. Examples for myxobacterial compounds produced by type III PKSs are flaviolin and a class of uncommon alkylpyrones, the latter showing potent inhibition of the bacterial topoisomerase.<sup>174,175</sup>

## 1.6 Nonribosomal peptide synthetases (NPRSs)

NRPSs share a similar biosynthetic logic with PKSs as they step-wise incorporate monomeric building blocks into a growing polypeptide chain in an assembly line manner. In contrast to ribosomally produced peptides, NRPSs are not limited to the 20 proteinogenic amino acids, selenocysteine and their possible products. They are estimated to incorporate over 500 different amino acids including proteinogenic and non-proteinogenic amino acids,  $\beta$ -amino acids, D-amino acids, heterocyclized amino acids and aryl acids.<sup>176</sup> The minimal NRPS module consists of three core domains: The adenylation (A) domain, the thiolation (T) domain, which is analogous to the ACP in PKS systems and also known as peptidyl carrier protein (PCP), and the condensation (C) domain. The classical architecture of such a module is C-A-T.<sup>177</sup> The A domain selectively activates amino acids by adenylation in an ATP- and  $Mg^{2+}$ -dependent reaction. The generated aminoacyl-adenylate substrate is subsequently transferred to the T domain after a conformational rearrangement of the A domain (Figure 10).<sup>178</sup> Linking of the substrate to the T domain occurs after posttranslational phosphopantetheinylation of the T domain by a PPTase as it is also the case for ACPs in PKS biosynthesis.<sup>143</sup> The T domain-bound substrate is then shuttled to the adjacent C domain which catalyzes the peptide bond formation between the carboxyl group of the donor and the amine group of the T domain-bound acceptor substrate of the downstream NRPS module, resulting in elongation of the peptide chain. Notably, C domains are pseudo dimers with high substrate selectivity on the acceptor side and lower substrate selectivity on the donor side, thus taking a role as second gate keepers in NRPS biosynthesis beside the A domains.<sup>179,180</sup> Furthermore, C domains are further divided into multiple functional subtypes. E.g. a  $^L C_L$  domain catalyzes the condensation between two L-amino acids, whereas  $^D C_L$  domains link a D- with an L-amino acid.<sup>181</sup>



**Figure 10 | Peptide chain elongation in NRPS biosynthesis.** C: condensation domain (light blue), A: adenylation domain (dark blue), T: thiolation domain (grey). The A domain activates an amino acid substrate by adenylation. The substrate is then tethered to the phosphopantetheinylated T domain. The C domain catalyzes peptide bond formation between the T domain-bound substrates of two adjacent NRPS modules.

Apart from the core domains, optional auxiliary tailoring domains such as methylation,<sup>182</sup> epimerization<sup>183</sup> and oxidation<sup>184</sup> domains modify the T domain-bound substrates on-line (*in cis*) prior to condensation, thereby generating the huge structural variety in NRP synthesis.<sup>147</sup>

Furthermore, those modifications of the growing polypeptide chain can also occur during or after product release from the assembly line by independent tailoring enzymes *in trans*.<sup>147</sup> Similar to PKS chain termination, liberation of the full-length product in NRP biosynthesis is usually performed by a TE domain of the terminal NRPS module.

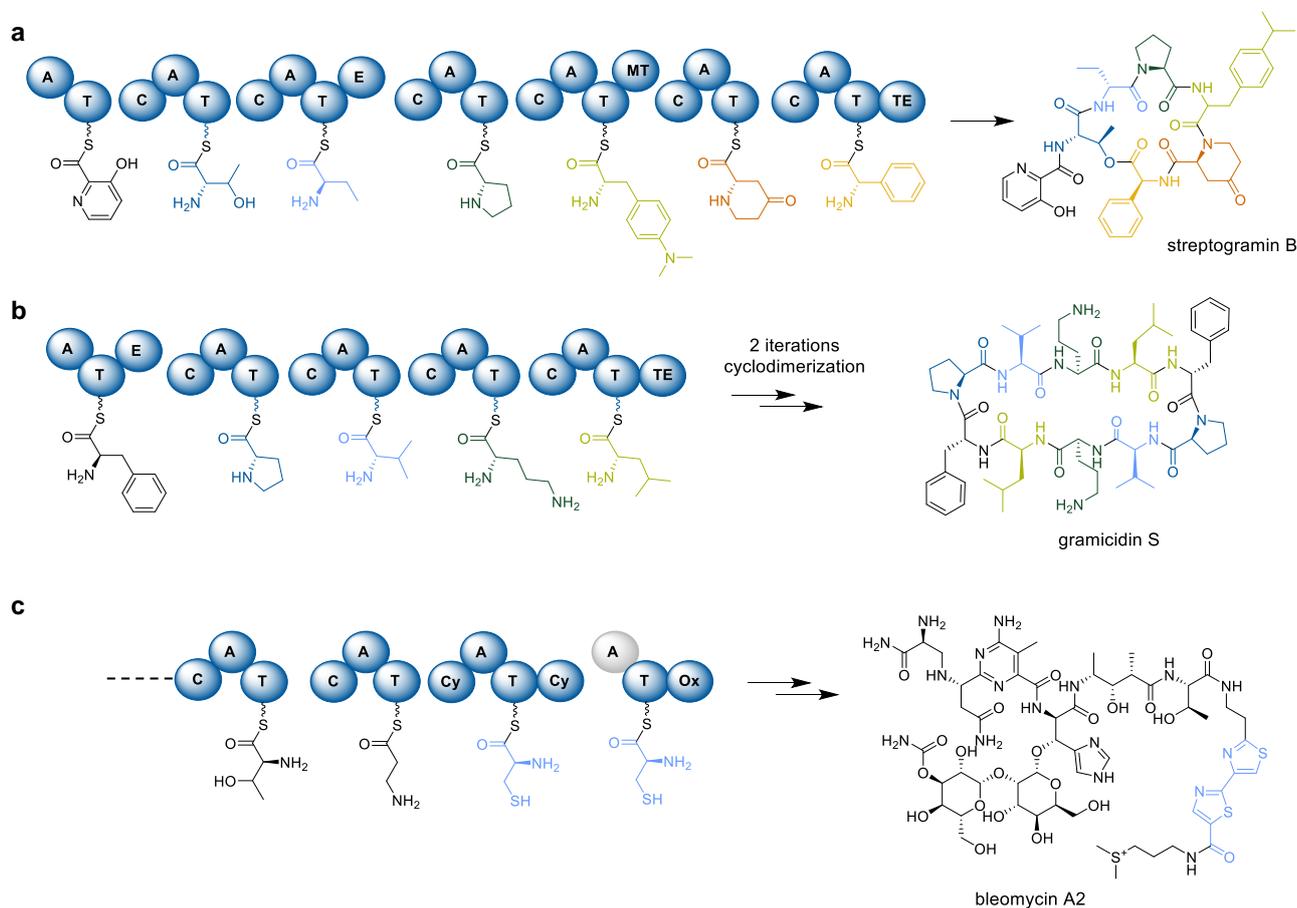
Commonly, NRPS biosynthesis follows two principles, the principle of collinearity, which states that each NRPS module catalyzes the incorporation of one building block,<sup>185</sup> and the principle of processivity, stating that the biosynthesis starts with the first module and proceeds sequentially to the downstream modules. However, exceptions to those principles were identified. Those exceptions in their biosynthetic logic lead to the classification of NRPSs into different subtypes (Figure 11): Subtype A classifies NRPSs which follow the principle of collinearity (linear NRPSs), subtype B NRPSs follow an iterative logic and subtype C include nonlinear NRPSs.<sup>149,177</sup>

In type A NRPSs, the resulting polypeptide sequence is completely determined by the number and order of modules, which have the prototypical core domain arrangement C-A-T.<sup>177</sup> Some examples of prominent compound classes produced by linear NRPSs are the antibiotics streptogramin (Figure 11a),<sup>186</sup> tyrocidine<sup>187</sup> and vancomycin-type antibiotics.<sup>188</sup> However, also for linear NRPSs exceptions were described such as the previously mentioned module skipping in the biosynthesis of the myxobacterial compound class of myxochromides.<sup>154</sup>

Type B NRPSs use their modules iteratively during the biosynthesis of one single product as it was observed for the iron-chelating siderophore enterobactin, which is a cyclic trimer of dihydroxybenzoyl-serines.<sup>177,189</sup> Another example is gramicidin S (Figure 11b), which consists of two identical head-to-tail condensed pentapeptide halves.<sup>190</sup> Those NRPS systems work with a reduced number of modules, which produce the repetitive polypeptide sequences that are finally fused to a multimeric product by the terminal T or TE domain.<sup>177</sup> For gramicidin and enterobactin cyclization and product release, one or two of the repetitive polypeptide chains are attached to the active site serine residue of the TE, respectively, thereby deacylating the upstream T domain and opening the door for the synthesis of the next polypeptide chain.<sup>177,191</sup>

NRPSs are classified as type C if they act nonlinear, meaning that one domain (not an entire module) is used several times during one NRP biosynthesis round.<sup>149</sup> Type C NRPS systems show dissociated assembly lines into stand-alone modules or even domains thus deviating from the linear NRPS C-A-T domain architecture.<sup>149</sup> One important example is the biosynthesis of the cytostatic and antibiotic glykopeptide bleomycin (Figure 11c) in which one cysteine-activating A domain is reused to load the T domain of the same module and additionally a *trans*-acting T domain, finally leading to the formation of the bisthiazole moiety by the action of the flanking cyclization (Cy) domain.<sup>184,192</sup> In syringomycin biosynthesis, one NRPS subunit (SyrB) consists of an A-T didomain which

activates L-Thr and delivers the substrate *in trans* via SyrC to another NRPS subunit (SyrE) harboring a module with an unusual C-T-TE domain architecture.<sup>177,193</sup> Similar *in trans* mechanisms and activation of amino acids from stand-alone NRPS subunits were proposed for the myxobacterial compound class of cystobactamids and the related albicidins.<sup>99,194</sup>



**Figure 11 | Subtypes of NRPSs and their underlying principles. a:** In linear type A NRPSs, exemplified by the streptogramin assembly line, each module is used only once during one biosynthesis cycle. The substrates of the different modules are highlighted in different colors to show that each building block only occurs once in the final product. **b:** Iterative type B NRPSs use certain modules or the entire assembly line more than once as shown for gramicidin S biosynthesis. The terminal TE domain catalyzes cyclodimerization and release of the product. Each building block is present twice in the final product as indicated by the substrate coloring. **c:** Nonlinear type C NRPSs, such as the bleomycin NRPS, reuse single domains during one biocatalytic cycle. The cysteine building block (shown in blue) is incorporated by the two terminal modules, but only activated by one A domain, because the A domain of the terminal module (shown in grey) is inactive. For reasons of simplicity, only the four terminal modules are depicted (indicated by the dashed line). The entire assembly line includes five more modules - four NRPS modules and one PKS module.

Notably, other classification systems designated NRPSs as type I or type II, similar to the classification of FASs and PKSs.<sup>195</sup> Type I NRPSs, likewise type A NRPSs, are classified as linear, modular NRPSs in which the transfer of the activated aminoacyl adenylate can only occur between an A domain and its cognate T domain. In type II NRPSs, however, the T domain has no cognate A domain of its own and can thereby be aminoacylated for example by an A domain from another NRPS module. Even though this classification system differentiates between entire type I and type II

NRPS systems, the type I and type II classification is also often used only for single domains, describing their existence as part of multienzyme complexes as well as stand-alone domains or modules. Type II NRPS systems can therefore be treated as equivalent to type C NRPSs. However, the growing number of exceptions in the structure and functionality of NRPSs and PKSs, as well as transition states between differently classified biosynthetic systems, makes the use of such classification systems increasingly difficult. This was discussed for the classification of PKSs nearly two decades ago.<sup>196</sup>

### 1.6.1 Starter modules in NRPS biosynthesis

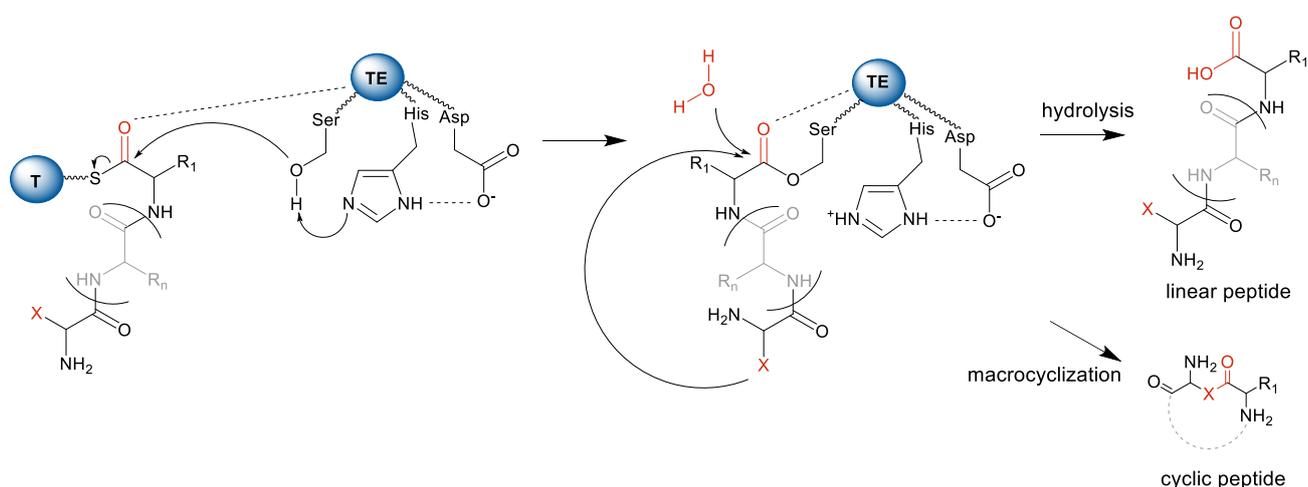
The C domains in NRPS initiation modules are often absent (A-T didomain) or inactive in pure NRPS systems since no upstream module is present that provides an activated aminoacyl adenylate substrate for condensation. However, in hybrid NRPSs, e.g. PKS-NRPS or NRPSs involved in lipopeptide biosynthesis, the starter C domains are active and catalyze the connection of fatty acids or PKs to the first amino acid of the peptide chain.<sup>181</sup> For example, in the biosynthesis of the surfactant and antibiotic lipopeptide lichenysin a  $\beta$ -hydroxyl fatty acid is transferred to the starter C domain from an ACP.<sup>197</sup> Other examples are the biosyntheses of streptogramin (pristinamycin), enterobactin and actinomycin in which the respective starter C domains catalyze a bond formation between  $\beta$ -hydroxyl carboxylic acids and an amino acid.<sup>181</sup> Therefore, the starter C domains require specificity for both fatty acids or carboxylic acids and amino acids explaining why they do not cluster in the same subtrees with other C domain subtypes in phylogenetic analyses.<sup>181</sup>

### 1.6.2 NRPS chain termination

In contrast to fungal NRPSs which usually have a C domain as terminal product-releasing and macrocyclization domain,<sup>198</sup> many bacterial NRPSs have type I TE domains as terminal domains instead acting as release catalyst for the full-length polypeptide chain.<sup>199</sup> However, reductase domains or spontaneous intramolecular cyclization reactions were also described as possible releasing mechanisms in NRPSs.<sup>200</sup> Furthermore, an unusual hydrolytic release domain with similarity to C domains was described in the biosynthesis of the myxobacterial antifungal and cytotoxic compound class of crocacins.<sup>201</sup>

Biochemically type I TE domains catalyze the nucleophilic attack of an active site serine residue (part of the catalytic Ser-His-Asp triad) on the carbonyl carbon of the T domain-bound peptidylacyl thioester substrate to yield a TE domain-bound oxoester substrate (Figure 12). The TE domain-bound substrate is then attacked either by an external nucleophile like water leading to the release of a linear product, or by an internal nucleophile leading to macrocyclization.<sup>200</sup> Especially cyclization reactions are assumed to require strategic interaction of the peptidylacyl intermediate

with residues in the active site of the TE domain. The interpretation of the TE domains structural data suggested that electrophilic and nucleophilic residues of the substrate determine the substrate specificity of the TE domain in NRPSs.<sup>200</sup> Moreover, the solved structure of the T-TE didomain underlined the importance of domain interactions but also intradomain motions,<sup>202</sup> thus showing that TE domains are not only specific for their substrate but also that interaction with the upstream T domain is a specific prerequisite.



**Figure 12 | Peptide release mechanism of type I TE domains.** The nucleophilic attack of the active site serine yields a TE domain-bound oxoester substrate. A second nucleophilic attack by water leads to hydrolysis and release of a linear peptide, whereas an internal nucleophilic attack (here from X moiety as nucleophile) results in macrocyclization and release of a cyclic peptide.

In contrast to type I TEs, type II TEs are stand-alone proteins which are often associated with NRPS, PKS or FAS pathways. In NRP biosynthesis sometimes aberrant substrates compete with required substrates leading to activation of amino acids by the A domain that are not accepted by the respective C domain. This may lead to stalling of the biosynthesis and thereby blocking of the assembly line. Type II TEs perform proofreading functions by catalyzing hydrolysis and thus removal of amino acids and peptides that block NRPSs.<sup>203</sup> Therefore, they exhibit inverse specificity for the correct substrate to only cleave the aberrant substrates. The importance of type II TEs for PKS-NRPS systems was underlined for instance when a complete loss of production was observed after deletion of the gene encoding the type II TE in the streptolydigin BGC.<sup>204</sup> Another function of type II TEs is the release of pathway intermediates from a T domain allowing further processing steps by additional enzymes as it was shown in the biosyntheses of kutznerides and nikkomycin.<sup>205,206</sup> Further functions were assigned to type II TEs in PKS systems, as for example starter unit selection, chain length control and product release.<sup>203</sup>

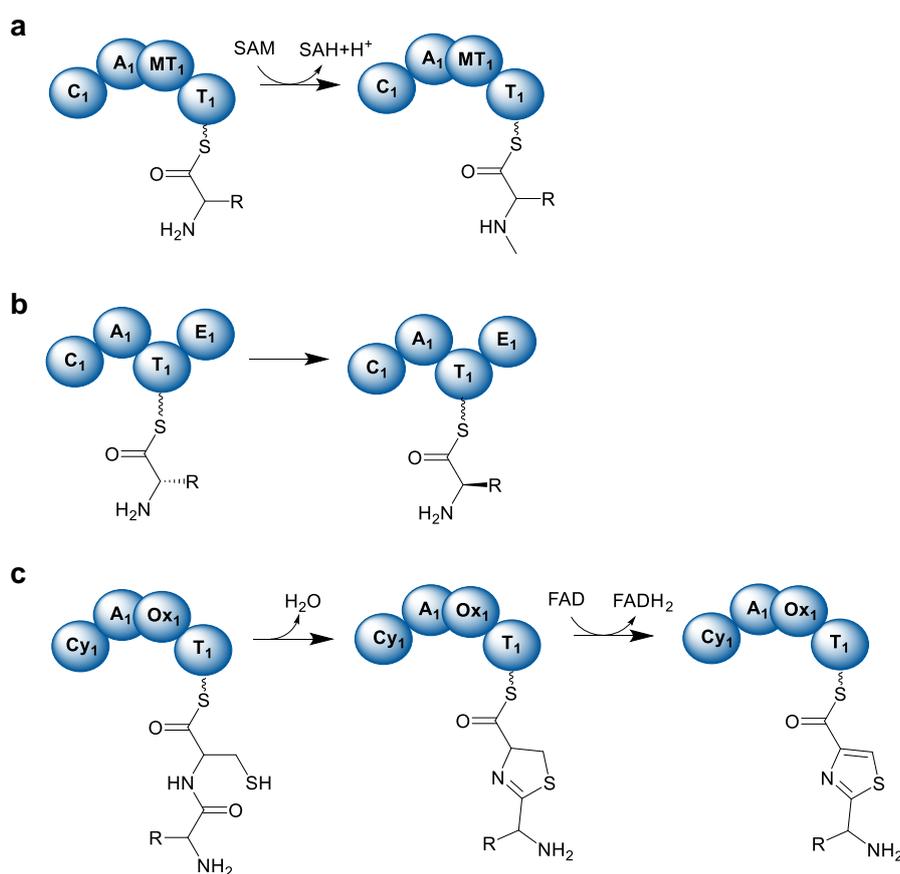
### 1.6.3 Tailoring in NRPS biosynthesis

As already mentioned, in NRPS biosynthesis tailoring enzymes optionally modify the NRP either during chain elongation, while the growing peptidylacyl thioester is attached to a T domain, or after product release from the assembly line. *In trans* reactions describe modifications by independent enzymes, whereas *in cis* reactions are performed by tailoring domains which are embedded within the assembly line. Numerous modifications are known in NRPS biosynthesis, such as methylations, oxidations, formylations, epimerizations, glycosylations and cyclizations.<sup>149</sup>

A variety of NRPs contain *N*-, *C*-, *S*- or *O*-methylated amino acids.<sup>147</sup> *N*-methylations in NRPs are the most abundant among the methylations and were discussed having a significant impact on polarity, conformational freedom and resistance towards proteases.<sup>207</sup> Apart from *trans*-acting *N*-MTs, the majority of methylations in NRPs are catalyzed *in cis* by *N*-MTs (shown in Figure 13a) embedded in the A domain leading to a C-A<sub>core</sub>-*N*-MT-A<sub>sub</sub>-T module architecture.<sup>149</sup> One prominent example is the cyclosporine synthetase, which contains seven modules with C-A<sub>core</sub>-*N*-MT-A<sub>sub</sub>-T architecture, leading to *N*-methylation of seven out of eleven amino acids in cyclosporine A.<sup>208</sup> However, there are also deviations such as the diaminopropionic acid side-chain *N*-methylating *N*-MT involved in paenilamicin biosynthesis, which is localized upstream of the respective A domain.<sup>209</sup> Compared to *N*-methylations, *O* and *S*-methylations are only rarely found in NRPs as for example a serine residue *O*-methylation in the biosynthesis of kutznerides<sup>210</sup> and a *S*-methylation of a cysteine residue in the biosynthesis of thiocoraline.<sup>211</sup> One example for NRPS-mediated *C*-methylation is found in the yersiniabactin biosynthesis: a cysteine residue is first cyclized to a thiazoline group and subsequently *C*-methylated *in cis* by a *C*-MT embedded between the respective Cy and T domains.<sup>212</sup>

Epimerization (E) domains usually epimerize A domain-activated L-amino acids to D-amino acids (shown in Figure 13b), which not only tremendously increases the stereochemical diversity in NRPs compared to RiPPs, but also increases the resistance of NRPs towards peptidases.<sup>149</sup> Even though A domains were described that activate D-amino acids after conversion of an L-amino acid through an independent racemase,<sup>213</sup> the common mechanism includes the A domain-driven activation of an L-amino acid and the subsequent epimerization of the T domain-bound aminoacyl thioester by an E domain. Notably, E domains produce an equilibrated mixture of L- and D-amino acids as it was shown for the E domain of module 1 in gramicidin S synthetase.<sup>183</sup> However, the downstream C domain normally belongs to the <sup>D</sup>C<sub>L</sub> subtype and thus acts as a gate keeper specifically catalyzing a condensation reaction with the D-amino acid as substrate.

Another modification in NRP biosynthesis are cyclizations, which are performed by Cy domains. Cy domains descend from C domains and gained the function not only to catalyze peptide bond formation but also to perform heterocyclization, usually using cysteine, serine or threonine as substrates (shown in Figure 13c).<sup>149</sup> Those domains are found in the megasynthetases of compound classes harboring thiazoline- or oxazoline-based heterocycles as for example in EpoB, which is involved in epothilone biosynthesis.<sup>214</sup> Furthermore, apart from common *in trans*-acting monooxygenases, oxidation (Ox) domains, which are integrated into A domains, have also been described. Ox domains are regularly found in modules with Cy domains, catalyzing the transformation of thiazoline and oxazoline heterocycles into thiazole or oxazole rings, respectively (Figure 13c).<sup>149</sup> Another function of Ox domains was for example described in myxothiazol and melithiazol biosyntheses, where Ox domain-catalyzed C $\alpha$  hydroxylation of an A domain-activated glycine residue lead to spontaneous decomposition and release of the product harboring a C-terminal amide group.<sup>160,215</sup>



**Figure 13 | Common tailoring reactions by optional NRPS domains. a:** *N*-methylation by a *N*-MT domain. **b:** epimerization by an E domain. **c:** cyclization of cysteine and formation of a thiazoline ring by a Cy domain. Subsequent transformation to a thiazole heterocycle is catalyzed by an Ox domain.

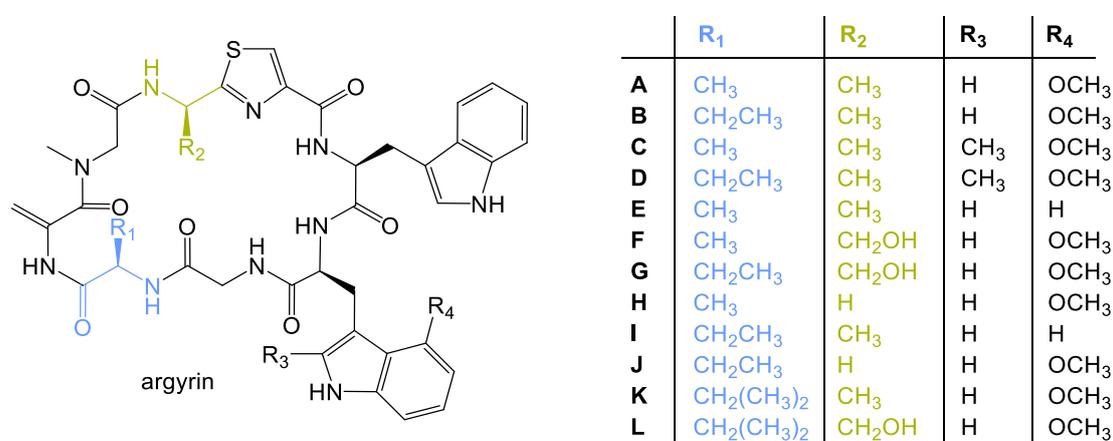
Some NRPs harbor aldehyde or primary alcohol groups that originate from a reduction of the T domain-bound peptidyl thioester by NAD(P)H-dependent reductase (R) domains. R domains are members of the short-chain dehydrogenase/reductase protein family and are organized in a C-terminal subdomain, which is involved in substrate recognition, and a catalytic N-terminal subdomain.<sup>149</sup> One example of a R domain is described in the biosynthesis of myxalamid antibiotics, potent inhibitors of the respiratory electron transport chain.<sup>216,217</sup> In this pathway, the R domain is the terminal domain replacing a TE domain and catalyzing the release and primary alcohol formation in myxalamids. Another, notably very rare, modification in NRP biosynthesis is the formylation, which was observed in linear gramicidin biosynthesis catalyzed by a formyltetrahydrofolate-dependent formylation domain on T domain-tethered substrate.<sup>218</sup>

*In trans* acting auxiliary enzymes such as monooxygenases, methyltransferases, *O*-carbamoyl transferases and glycosyltransferases were shown to further increase the chemical complexity of NRPs. However, they do not only modify the released product or T domain-tethered peptidyl thioester, but their actions are sometimes also required in a strict order to make the biosynthesis of the final product possible.<sup>149</sup>

#### **1.6.4 Combinatorial biosynthesis and NRPS engineering**

As mentioned before, one reason for the high structural diversity within the group of NRPs is the large pool of different substrates that can be accessed by NRPSs. However, those substrates are not randomly chosen. Instead selection occurs at several stages by numerous instances. The first specificity-mediating gatekeeper in substrate selection is the A domain as it was shown in several studies.<sup>219,220</sup> The elucidation of the first crystal structures of adenylating enzymes, namely the firefly luciferase<sup>221</sup> and the GrsA domain from the gramicidin S NRPS,<sup>222</sup> was an important step to understand the functionality of the substrate selection in A domains. The crystal structure of the latter was determined in complex with AMP and the substrate L-phenylalanine which enabled the identification of ten critical residues that are involved in substrate recognition and are located in the A domain binding pocket.<sup>222</sup> Based on this work, Stachelhaus and coworkers compared the essential residues of the L-phenylalanine-binding A domain with moieties in other A domains and developed general rules for determining substrate specificities of A domains, which was designated the “specificity-conferring code of A domains” or the “nonribosomal code”.<sup>223,224</sup> Initially, the nonribosomal code was based only on those ten residues identified in the GrsA crystal structure; however, later work included further important residues and core motifs thereby expanding the nonribosomal code.<sup>225</sup> This understanding set the basis for the development of bioinformatic tools<sup>225–228</sup> that are able to identify the specificity of A domains only based on the *in silico* analysis of the primary protein or DNA sequence. However, it is worth mentioning that A domains are not

necessarily specific for only one substrate as the incorporation of different amino acids at the same position in one NRP class was observed. For example in the biosynthesis of the myxobacterial compound class of argyriins, the A domain in module 1 activates either L-alanine, L- $\alpha$ -aminobutyric acid or L-valine while the A domain in module 4 activates either L-alanine, L-serine or glycine.<sup>229</sup> Notably, all substrates (except for glycine) are subsequently epimerized and accepted by the downstream C domains, leading to argyriin congeners with various combinations of D-amino acids in the two positions (Figure 14). Another example are the anabaenopeptins, a compound class exhibiting various bioactivities isolated from cyanobacteria, in which two distinct amino acids, arginine or tyrosine, are incorporated into different anabaenopeptin derivatives from the same A domain.<sup>230</sup>



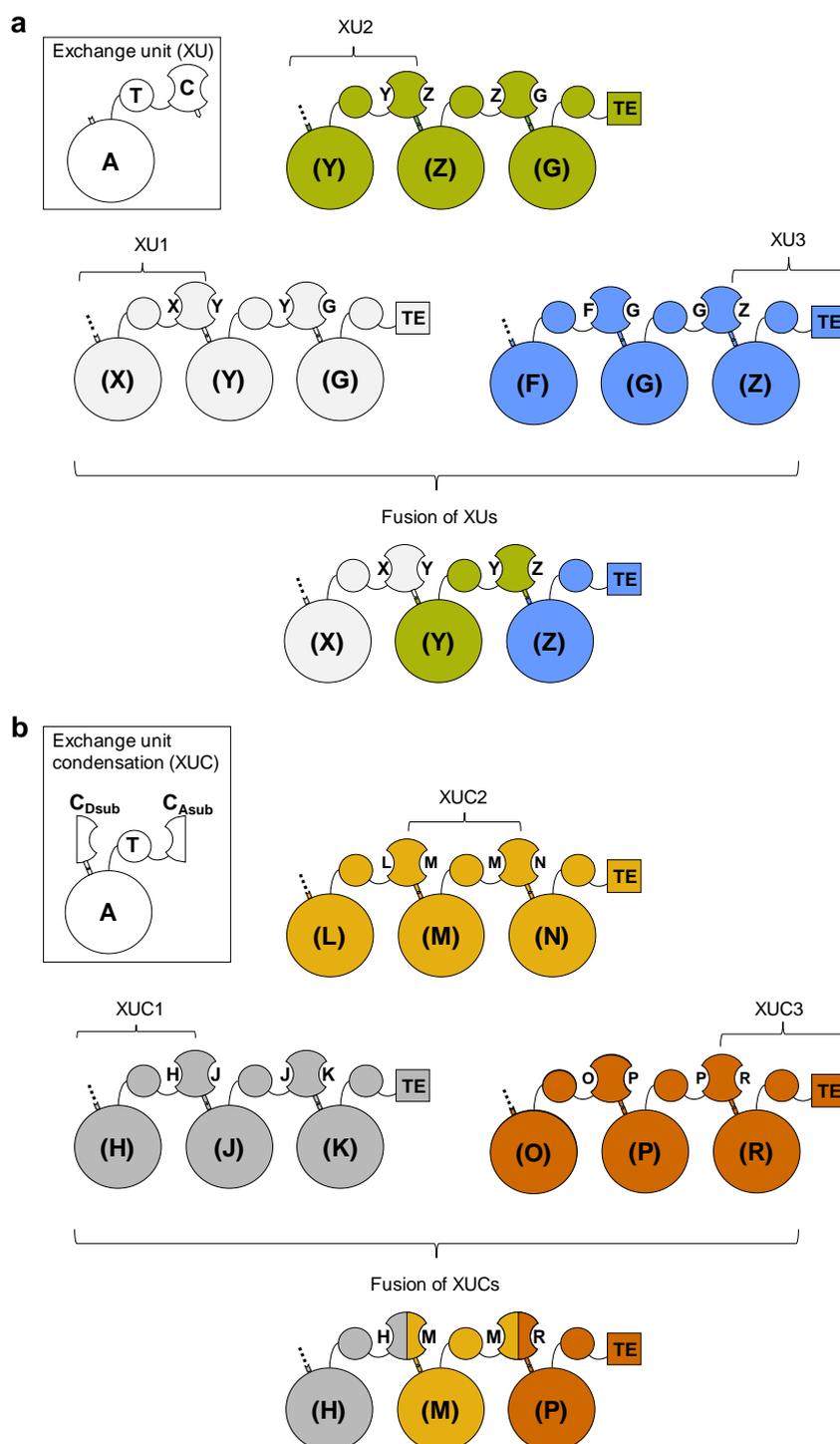
**Figure 14 | Structural diversity of argyriins introduced through promiscuous activity of A domains.** Shown is the prototypic structure of argyriin on the left. Two amino acids are colored in light-blue and green as different substrates were observed in those positions. Possible amino acid side chains are listed in the table on the right. This figure was modified from Ref. <sup>231</sup>.

The gained knowledge about A domains opened the door for manipulation attempts, in which the substrate specificities were altered only by exchange of one or a few residues in the A domain binding pocket.<sup>232–235</sup> Systematic mutagenesis can also be used to narrow down the substrate tolerance of a promiscuous A domain.<sup>236</sup> However, despite all methodological options to manipulate A domains, downstream domains and modules in NRPS systems may hinder further processing of the substrate as especially C domains act as second gatekeepers in NRP biosynthesis.<sup>179</sup> In early attempts to manipulate NRP biosynthesis, precursor-directed biosynthesis was the method of choice. Supplementation of modified or synthetic amino acids was performed during cultivation of wild-type NRP producer strains with the hope to achieve incorporation of those building blocks by NRPS modules, which exhibit relaxed substrate specificity.<sup>237</sup> As the fed precursors compete with the natural endogenous precursors, the yields of novel isolated compounds were often low and production optimization attempts were limited. This issue was addressed by mutasynthesis, in which engineered organisms with partly deficient enzyme machinery for the formation of natural

precursors were cultivated under supplementation of precursor analogs.<sup>237</sup> Apart from exogenous supplementation of unnatural amino acids, heterologous expression (described in the section below) of additional tailoring enzymes or deletion of tailoring enzymes was used to produce new NRP derivatives.<sup>237</sup> However, in contrast to systematic engineering of A domain specificities, those attempts did not aim directly at engineering the NRPS machinery.

Since the core peptide sequence of the NRP product is determined by the substrate specificity of NRPS modules, exchanges of entire NRPS subunits were performed to produce new peptidic core scaffolds. A prominent example is the engineering of the daptomycin assembly line, in which the terminal two modules (C-A-T-C-A-T-TE architecture) were deleted and complemented with respective modules from the highly similar A54145 and CDA (calcium-dependent antibiotic) pathways, resulting in the production of two new daptomycin analogs.<sup>238</sup> Also the exchange of entire modules (C-A-T or C-A-T-E), didomains (C-A) or single A domains resulted in the production of new peptide analogs; however, often accompanied by reduced production titers or no production at all.<sup>237,239,240</sup> Another route to engineer NRP biosynthesis was to change the length of peptide chains by insertion or deletion of modules. For example in balhimycin, a glycopeptide antibiotic, an additional amino acid was inserted by adding a chimeric NRPS module, consisting of the C-A from module 4 and T-E from module 5 balhimycin NRPS, to the assembly line.<sup>241</sup> Apart from the extended new balhimycin analog, truncated products were also obtained. This underlines how sensitive NRPS systems are towards engineering approaches and how important domain-domain interactions and the linker regions connecting the NRPS domains are.

Recently, Bozhüyük and coworkers developed two new strategies to generate NRPS systems *de novo* from NRPS subunits deriving from different biosynthetic pathways.<sup>242,243</sup> Based on a helical secondary structure in the linkers between C and A domains they established specific engineering sites and thus developed the exchange unit (XU) strategy in which not C-A-T but A-T-C units were fused (Figure 15a).<sup>242</sup> With this strategy they were able to combine up to four XUs from different assembly lines resulting in the production of numerous artificial peptides. In contrast to the previously mentioned methods, only a moderate drop in production titer was observed, however only when one or two XUs were exchanged. The major drawback of the XU strategy was the limitation of possible combinations of XUs due to the specificities of downstream C domains. In the second study Bozhüyük and coworkers overcame this problem by identifying specific assembly points within C domains (Figure 15b).<sup>243</sup> Since C domains have a pseudodimeric structure, they chose the linker region connecting the pseudodimers as engineering site. With the exchange unit condensation domain (XUC) concept, which also works in combination with the XU strategy, they were able to combine XUCs from different assembly lines and produce novel peptides and peptide libraries with natural and unnatural amino acids.



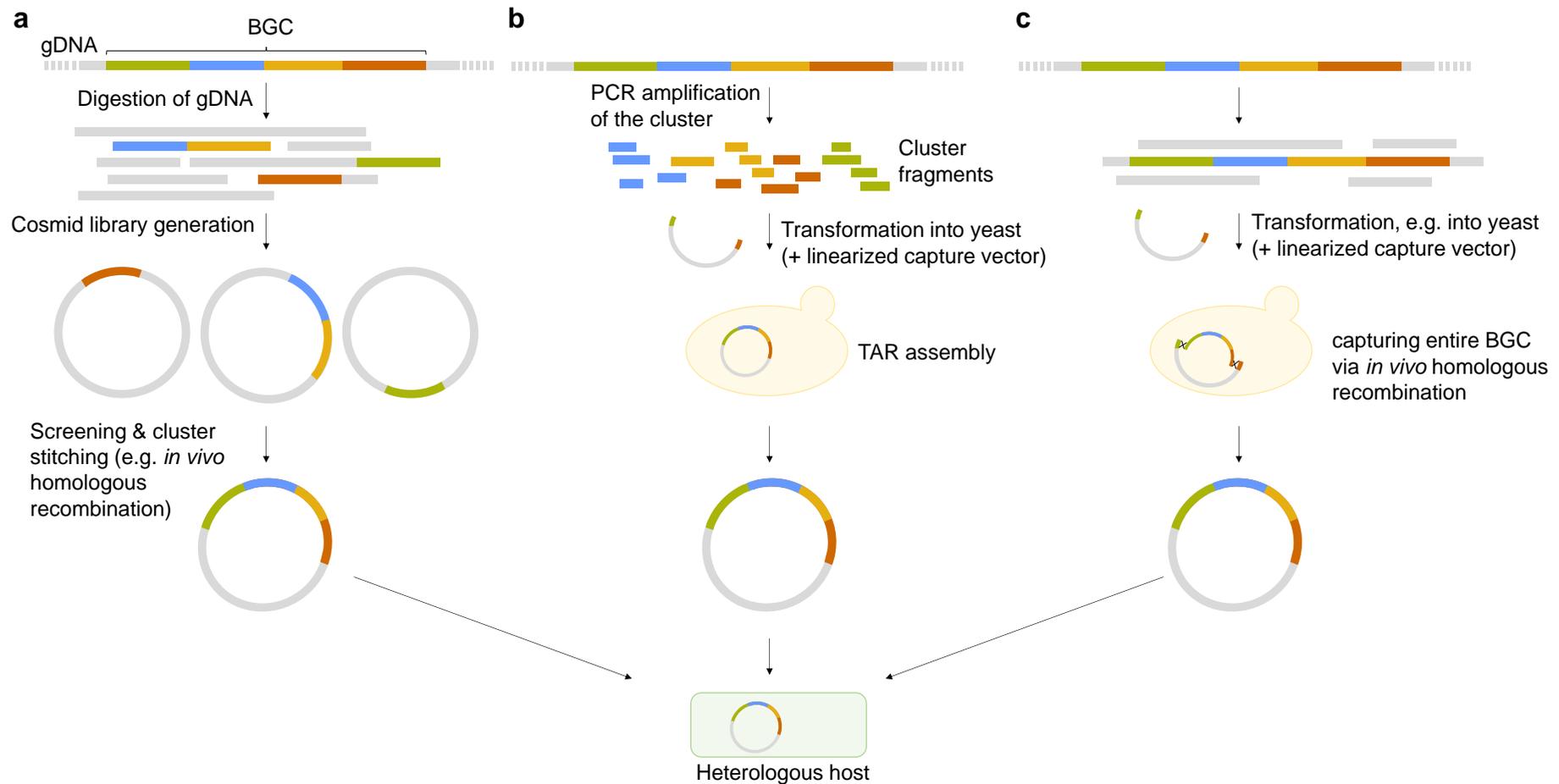
**Figure 15 | Exchange unit (XU) and exchange unit condensation (XUC) strategies. a:** One XU consists of an A-T-C NRPS building block (shown in the frame on the top left). XUs from three different NRPSs (light-grey, green and blue) were fused using fusion points within the C-A linker region. The letters in brackets on the A domains and next to the C domains indicate the specificities for hypothetical substrates. **b:** One XUC consists of an  $C_{Dsub}$ -A-T- $C_{Dsub}$  NRPS building block (shown in the frame on the top left). XUCs from three different NRPSs (dark-grey, orange and red) were fused using fusion points between the C domain pseudodimers. The letters in brackets on the A domains and next to the C domains indicate the specificities for hypothetical substrates. This scheme was adapted from Refs. <sup>242,243</sup>.

Notably, the combination of XUs or XUCs of Gram-negative and Gram-positive origin failed, emphasizing that phylogenetic relatedness of the host organisms of the NRPS machinery remains a limiting factor in NRPS engineering. Since the XU and XUC strategy was so far only successfully tested using NRPSs from phylogenetically closely related *Xenorhabdus* and *Photorhabdus* strains, the limitations of both methods has to be addressed in future experiments.

### **1.7 Assembly and heterologous expression of complex natural product biosynthetic pathways**

Since there is often only little knowledge about native bacterial producer strains of NPs regarding optimal cultivation conditions and genetic manipulability, heterologous production of the respective NPs using well-characterized host strains has been proven indispensable.<sup>244,245</sup> Keeping in mind that 99 % of the potential microbial species on earth are undiscovered and only a small fraction of those can be cultured using current techniques,<sup>246</sup> the few culturable producers with promising production profiles often show slow growth rates and poor production titers. As most of the native producers are genetically not accessible, increases in gene expression levels by introducing strong promoters or positive regulators remain difficult to realize. Using model host strains is therefore sometimes more efficient than investing time and resources in the development of cultivation and manipulation methods for native producers. Typical model strains are *E. coli*, *Bacilli* (e.g. *B. megaterium*, *B. subtilis* and *B. brevis*),<sup>247</sup> *Streptomyces* strains (e.g. *S. albus* and *S. coelicolor*),<sup>248</sup> cyanobacterial *Anabaena* sp. and *Synechocystis* sp. strains, *Pseudomonas putida* and *Burkholderia* strains.<sup>244</sup> For myxobacterial NPs *Myxococcus xanthus* DK1622 has proven a suitable heterologous host strain with relatively fast growth, well-described life cycle (see previous section) and a remarkable number of successfully heterologously expressed BGCs.<sup>116,249</sup>

The complex machinery of megasynthetases, additional tailoring enzymes, resistance-mediating enzymes and regulatory proteins involved in the biosyntheses of the NPs are encoded by a set of genes, usually located in close proximity (clustered) in the genomes of the producer strains. Bacterial NP BGCs range in their size from a few kb up to several hundred kb, whereas type I PKS clusters are the largest like for example the quinolidomycin BGC with over 200 kb.<sup>250</sup> Since it is extremely difficult to clone and transform BGCs with sizes over 100 kb, only a few of them were heterologously expressed so far.<sup>250–252</sup> Before transformation of the potential heterologous host strain, the BGC has to - in most cases - be isolated from the native producer genome. Traditionally, cosmid libraries were generated and screened for the presence of the BGC; however, especially in the case of large BGCs, the desired genes were often located on several cosmids, requiring downstream cloning steps such as *in vivo* recombination in *E. coli* to assemble the complete BGC (Figure 16a).<sup>215,253</sup>

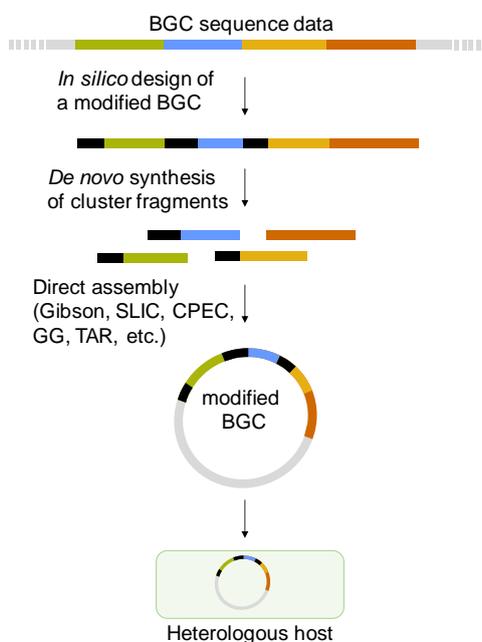


**Figure 16 | Strategies for heterologous expression of a BGC from genomic DNA.** **a:** digestion of genomic DNA (gDNA), cosmid library generation, subsequent screening of clones harboring cosmids with cluster fragments and cluster stitching (e.g. by *in vivo* homologous recombination). **b:** amplification of BGC DNA fragments by PCR and BGC assembly via transformation-associated recombination (TAR). **c:** direct cloning approach: digestion of gDNA without disruption of the BGC, transformation into yeast with capture vector and BGC “capturing”, e.g. via TAR. This scheme was adapted from Ref. <sup>261</sup>.

Another method includes the isolation of genomic DNA of the native producer strain, amplification of the complete BGC DNA by numerous polymerase chain reactions (PCRs), assembly, for example via transformation-associated recombination, and subsequent heterologous expression (Figure 16c) as it was shown for the BGCs of the *Streptomyces*-derived antibiotic grecoacycline and the myxobacterial topoisomerase inhibitor pyxidicycline.<sup>254,255</sup> Potential drawbacks of this method are the variety of PCRs that have to be performed, especially when amplifying large BGCs, and the risk of introducing random mutations as a result of the DNA polymerase error rate.<sup>255</sup> Further methods to assemble BGCs from DNA fragments are for example Gibson assembly,<sup>256</sup> sequence- and ligation-independent cloning,<sup>257</sup> circular polymerase extension cloning<sup>258</sup> and Golden Gate assembly.<sup>259</sup> A rapid method to isolate BGCs from genomic DNA is direct cloning,<sup>260,261</sup> in which a genomic DNA stretch harboring the BGC of interest is transformed together with a so-called capture vector into yeast or *E. coli* (Figure 16b). Subsequent homologous recombination based on homologous sequences on the vector and BGC-flanking regions leads to integration of the complete BGC into the vector. There are numerous examples of bacterial secondary metabolite pathways that were captured by direct cloning and heterologously expressed yielding the production of various compounds with antibiotic or anti-cancer properties.<sup>262–264</sup>

To achieve successful heterologous expression, the BGC is usually modified prior to transformation, including the introduction of host-specific promoters, regulators, terminators, rearrangement of the gene order and cloning onto a suitable expression vector system, the latter being self-replicating or integrative and with appropriate resistance-mediating genes for positive clone selection.<sup>245</sup> Furthermore, for the cloning of extremely large BGCs, the use of low-copy origin of replication (*ori*) vectors, such as bacterial artificial chromosome (BAC) vectors, were shown to be advantageous since they have a higher capacity for stable maintenance of large DNA pieces.<sup>250</sup> As the genes of the BGC have to be transcribed and translated in the heterologous host, including correct folding and posttranslational modifications of the proteins, close phylogenetic relatedness of the native producer strain with the heterologous host is often beneficial. For example, myxobacteria exhibit an average GC content of around 70 %<sup>116</sup> and therefore also have a different codon usage bias than non-GC-rich organisms, which could make heterologous expression of a myxobacterial BGC in an AT-rich host strain inefficient. Codons which are common in the native producer strain may be rare in a potential heterologous host and could thereby lead to ribosomal stalling and misfolded, nonfunctional proteins.

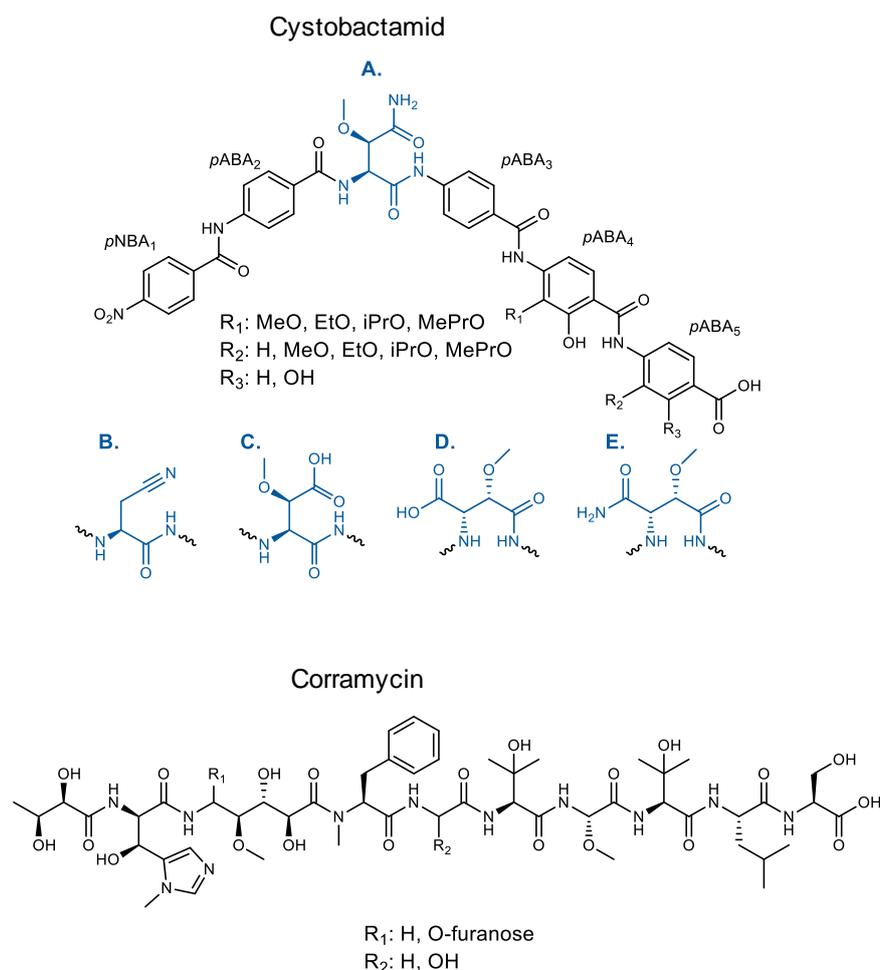
In the last decade, advances in *de novo* DNA synthesis<sup>265</sup> opened up the possibility to design biosynthetic pathways completely *in silico* and order modified BGCs directly from the gene synthesis industry (Figure 17). Based on the sequence data of native producer strains, modifications such as promoter insertions, restriction site engineering and gene engineering like rearrangement, deletion and addition of whole genes or site-specific mutagenesis of specific core motifs were possible only with mouse clicks without additional cloning effort. Furthermore, adaptation of the codon usage bias of a BGC enables the heterologous expression even in host strains that are phylogenetically only distantly related to the native producer strain. For example, the heterologous production of carbapenem antibiotics in *E. coli* using a codon-optimized BGC from *Pectobacterium carotovorum* resulted in production titers surpassing the titers observed in native producers like *Streptomyces cattleya* by a factor of ten and more.<sup>266</sup> However, there are numerous examples of codon-optimized BGCs that were expressed in less well-characterized strains (compared to *E. coli*) resulting in loss of production or no production titer improvement. For example the heterologous expression of a codon-optimized argyirin BGC in *M. xanthus* DK1622 yielded only 100  $\mu\text{g L}^{-1}$  argyirin A/B, whereas a second (not codon-optimized) BGC version resulted in 35  $\text{mg L}^{-1}$  production of argyirin A/B.<sup>229</sup> Notably, *de novo* DNA synthesis also has its limits as the synthesis of DNA fragments exceeding several kb is expensive and currently impractical, especially for GC-rich DNA. Therefore, the assembly of entire BGCs from DNA synthesis fragments still relies on previously mentioned classical assembly techniques which has to be kept in mind during *in silico* design phases.



**Figure 17 | Strategy for the heterologous expression of a modified BGC from DNA synthesis fragments.** The native BGC sequence is modified *in silico* (modifications shown in black), DNA cluster fragments are synthesized *de novo* and assembled using one or several techniques such as Gibson assembly, sequence- and ligation-independent cloning (SLIC), circular polymerase extension cloning (CPEC) and golden gate (GG) assembly.

## 1.8 Outline of this work

This work focuses on two myxobacterial NP classes, the cystobactamids and the corramycins (Figure 18), which are both in preclinical development as promising antibiotics. Both compound classes include highly potent congeners with anti-Gram-negative activity against important human pathogens. Furthermore, total syntheses of native compounds and synthetic derivatives with acceptable (cystobactamid)<sup>100,267–269</sup> and poor (corramycin; unpublished data; part of this work in chapter 3) yields were established. For both compound families the biosynthetic pathways have already been identified and initial studies on the biosynthesis were carried out.<sup>99,270</sup> Since the native producer strains were genetically not or only poorly accessible for manipulation, the primary objective of this work was the establishment of versatile heterologous expression platforms for both compound families to gain insights into their biosynthesis, produce novel derivatives and to pave the way for semi-synthetic production approaches for corramycins.



**Figure 18 | Structure of cystobactamid and corramycin.** *p*NBA: *para*-nitrobenzoic acid, *p*ABA: *para*-aminobenzoic acid, A.-E.: different linker moieties.

The first part of this thesis describes the work on the bacterial topoisomerase IIA inhibitor cystobactamid, which is a NRP produced by different *Cystobacter*, *Myxococcus* and *Coralloccoccus* species.<sup>99,100,271</sup> This compound class contains several in NP rarely observed or unprecedented structural features, such as the tailored *para*-aminobenzoic acid, *para*-nitrobenzoic acid moieties and the L-isoasparagine or L-asparagine linker groups. Naturally, a whole cocktail of cystobactamids is produced and structure activity relationship (SAR) studies revealed that certain derivatives with tailored L-asparagine linker exhibit superior antibacterial activity compared to other congeners.<sup>100</sup> The biosynthesis of cystobactamids was initially hypothesized based on *in silico* analysis of the BGC and feeding experiments.<sup>99</sup> Experimental proof of most of the biosynthetic steps was achieved afterwards by heterologously expressing the enzymes in *E. coli*, followed by their purification and *in vitro* characterization of their activity.<sup>272</sup> However, some activities could not be proven in those experiments, such as the function of the proposed bifunctional aminomutation/dehydration (AMDH) domain. Therefore, it was planned to design a modified cystobactamid BGC and an appropriate cloning and expression vector system *in silico* and assemble those after *de novo* DNA synthesis of the cluster and vector fragments. The first goal was the heterologous expression of the modified BGC in the myxobacterial model strain *M. xanthus* DK1622 and the production of native cystobactamids. The second goal was the genetic manipulation of the BGC, for example the targeted deletion of tailoring genes or the AMDH domain, and subsequent heterologous expression of the manipulated construct. The objectives of this second part were as follows: (i) The production of novel cystobactamid derivatives; (ii) the direction of the production profile towards congeners with superior antibacterial activity; (iii) the elucidation of the missing biosynthesis steps that could not be elucidated by the *in vitro* experiments, particularly the ones involved in linker formation.

The second part of this work describes the linear NRPS-PKS product corramycin, which was isolated from two *Coralloccoccus coralloides* species and showed promising antibacterial activity against *E. coli*.<sup>270</sup> The structure of corramycin was elucidated via NMR revealing some unprecedented features like the (2*R*,3*S*)- $\gamma$ -*N*-methyl- $\beta$ -hydroxy-histidine moiety. The corramycin BGC was identified after genome sequencing of the native producer strains and a biosynthesis model was provided based on *in silico* analysis of the 12-modular assembly line and feeding experiments.<sup>270</sup> Furthermore, the development of a total synthesis route for corramycin was achieved and promising *in vivo* activity of one corramycin derivative in an *E. coli* septicemia mouse model was shown (unpublished data; part of this work in chapter 3). The first goal in this project was the establishment of a heterologous production platform for native corramycins using *M. xanthus* DK1622 and based on a modified BGC that is obtained by the assembly of *de novo*-synthesized DNA fragments. Since the total synthesis of corramycin is inefficient due to the complex (2*R*,3*S*)- $\gamma$ -*N*-methyl- $\beta$ -hydroxy-histidine building block (unpublished data; personal communication with Stephane Renard, Evotec),

the second goal is the manipulation of the BGC and subsequent heterologous expression in *M. xanthus* DK1622 for the production of C-terminally truncated derivatives. Those derivatives could be used in a semi-synthetic approach to obtain higher yields of corramycin and to produce novel derivatives.

## 1.9 References

1. Hanson, J. R. *Natural products: The secondary metabolites* (Royal Society of Chemistry, Cambridge, 2003).
2. All natural. *Nat. Chem. Biol.* **3**, 351; 10.1038/nchembio0707-351 (2007).
3. 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).
4. Lietava, J. Medicinal plants in a Middle Paleolithic grave Shanidar IV? *J. Ethnopharmacol.* **35**, 263–266; 10.1016/0378-8741(92)90023-k (1992).
5. Petrovska, B. B. Historical review of medicinal plants' usage. *Pharmacogn. Rev.* **6**, 1–5; 10.4103/0973-7847.95849 (2012).
6. Scurlock, J. *Sourcebook for Ancient Mesopotamian Medicine* (Society of Biblical Literature, 2014).
7. Ebers, G. ed. *Papyrus Ebers: Das Hermetische Buch über die Arzneimittel der alten Ägypter in hieratischer Schrift* (Leipzig, 1875).
8. Cragg, G. M. & Newman, D. J. Biodiversity: A continuing source of novel drug leads. *Pure Appl. Chem.* **77**; 10.1351/pac200577010007 (2005).
9. Chang, H.-M., But, P. P.-H., Yao, S.-C., Wang, L.-L. & Yeung, S. C.-S. *Pharmacology and Applications of Chinese Materia Medica* (World Scientific, 1986).
10. Kapoor, L. D. *Handbook of Ayurvedic Medicinal Plants. Herbal Reference Library* (CRC Press, 2000).
11. Iwu, M. M. *Handbook of African Medicinal Plants, Second Edition* (CRC Press, 2014).
12. Cowan, R. S. & Ayensu, E. S. Medicinal Plants of the West Indies. *Taxon* **32**, 325; 10.2307/1222001 (1983).
13. Farnsworth, N. R., Akerele, O., Bingel, A. S., Soejarto, D. D. & Guo, Z. Medicinal plants in therapy. *Bulletin of the World Health Organization* **63**, 965–981 (1985).

14. 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; 10.1073/pnas.1614680114 (2017).
15. Koehn, F. E. & Carter, G. T. The evolving role of natural products in drug discovery. *Nat. Rev. Drug Discovery* **4**, 206–220; 10.1038/nrd1657 (2005).
16. 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; 10.1021/acs.jnatprod.9b01285 (2020).
17. Martin, E. J. & Critchlow, R. E. Beyond Mere Diversity: Tailoring Combinatorial Libraries for Drug Discovery. *J. Comb. Chem.* **1**, 32–45; 10.1021/cc9800024 (1999).
18. Feher, M. & Schmidt, J. M. Property Distributions: Differences between Drugs, Natural Products, and Molecules from Combinatorial Chemistry. *J. Chem. Inf. Comp. Sci.* **43**, 218–227; 10.1021/ci0200467 (2003).
19. 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).
20. Stahura, F. L., Godden, J. W., Xue, L. & Bajorath, J. Distinguishing between Natural Products and Synthetic Molecules by Descriptor Shannon Entropy Analysis and Binary QSAR Calculations. *J. Chem. Inf. Comp. Sci.* **40**, 1245–1252; 10.1021/ci0003303 (2000).
21. Lee, M. L. & Schneider, G. Scaffold architecture and pharmacophoric properties of natural products and trade drugs: application in the design of natural product-based combinatorial libraries. *J. Comb. Chem.* **3**, 284–289; 10.1021/cc0000971 (2001).
22. 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; 10.1002/(SICI)1521-3773(19990301)38:5<643::AID-ANIE643>3.0.CO;2-G (1999).
23. Lipinski, C. A., Lombardo, F., Dominy, B. W. & Feeney, P. J. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv. Drug Delivery Rev.* **23**, 3–25; 10.1016/s0169-409x(96)00423-1 (1997).

24. Anantharaman, V., Aravind, L. & Koonin, E. V. Emergence of diverse biochemical activities in evolutionarily conserved structural scaffolds of proteins. *Curr. Opin. Chem. Biol.* **7**, 12–20; 10.1016/s1367-5931(02)00018-2 (2003).
25. Breinbauer R., Vetter I. R. & Waldmann H. From Protein Domains to Drug Candidates - Natural Products as Guiding Principles in the Design and Synthesis of Compound Libraries. *Angew. Chem. Int. Ed. Engl.* **41**, 2878–2890; 10.1002/1521-3773(20020816)41:16<2878::AID-ANIE2878>3.0.CO;2-B (2002).
26. Burke M. D. & Schreiber S. L. A Planning Strategy for Diversity-Oriented Synthesis. *Angew. Chem. Int. Ed. Engl.* **43**, 46–58; 10.1002/anie.200300626 (2004).
27. Basu, S. & Waldmann, H. Polymer supported synthesis of a natural product-inspired oxepane library. *Bioorg. Med. Chem.* **22**, 4430–4444; 10.1016/j.bmc.2014.05.039 (2014).
28. Murray, C. W. & Rees, D. C. The rise of fragment-based drug discovery. *Nat. Chem.* **1**, 187–192; 10.1038/nchem.217 (2009).
29. Reutlinger, M., Rodrigues, T., Schneider, P. & Schneider, G. Multi-Objective Molecular De Novo Design by Adaptive Fragment Prioritization. *Angew. Chem. Int. Ed. Engl.* **53**, 4244–4248; 10.1002/anie.201310864 (2014).
30. Nicolaou K. C. & Rigol S. A brief history of antibiotics and select advances in their synthesis. *J. Antibiot.* **71**, 153–184; 10.1038/ja.2017.62 (2018).
31. Wallenfels, K. Symbiose und Antibiose. *Die Chemie* **58**, 1–16; 10.1002/ange.19450580102 (1945).
32. Bentley, R. & Bennett, J.W. What Is an Antibiotic? Revisited. *Adv. Appl. Microbiol.* **52**, 303–331; 10.1016/S0065-2164(03)01012-8 (2003).
33. Henle, F. G. J. Pathologische Untersuchungen. *Verlag von August Hirschfeld, Berlin, Prussia*, 1–82 (1840).
34. Koch, R. Die Aetiologie der Tuberculose. *Berl. Klin. Wochenschr.* **19**, 287–296 (1882).
35. Koch, R. Die Aetiologie der Tuberculose. *Mitth. Kais. Gesundheits* **2**, 1–88 (1884).
36. Loeffler, F. Untersuchung über die Bedeutung der Mikroorganismen für die Entstehung der Diphtherie beim Menschen, bei der Taube und beim Kalbe. *Mitth. Kais. Gesundheits* **2**, 421–499 (1884).

37. Gosio, B. Contributo all'etiologia della pellagra. Ricerche chimiche e batteriologiche sulle alterazioni del mais. *G. Accad. Med. Torino* 61, 464–487 (1893). *G. Accad. Med. Torino* **61**, 464 (1893).
38. Gosio, B. Ricerche batteriologiche e chimiche sulle alterazioni del mais. Contributo all'etiologia della pellagra. *Riv. d'Ig. San. Pubbl.* 7, 825–849 (1896). *Riv. d'Ig. San. Pubbl.* **7**, 825 (1896).
39. Silverman Kitchin, J. E., Pomeranz, M. K., Pak, G., Washenik, K. & Shupack, J. L. Rediscovering mycophenolic acid: A review of its mechanism, side effects, and potential uses. *J. Am. Acad. Derm.* **37**, 445–449; 10.1016/s0190-9622(97)70147-6 (1997).
40. Alsberg, C. L. & Black, O. F. Contributions to the study of maize deterioration. Biochemical and toxicological investigations of *Penicillium puberulum* and *Penicillium stoloniferum*. *Bull. US Bur. Pl. Ind.* 270, 7–48 (1913) (1913).
41. Birkinshaw, J. H., Raistrick, H. & Ross, D. J. The molecular constitution of mycophenolic acid, a metabolic product of *Penicillium brevi-compactum* Dierckx. Part 3. Further observations on the structural formula for mycophenolic acid. *Biochem. J.* **50**, 630–634; 10.1042/bj0500630 (1952).
42. Ehrlich, P. & Bertheim, A. Über das salzsaure 3.3'-Diamino-4.4'-dioxy-arsenobenzol und seine nächsten Verwandten. *Ber. dtsh. Chem. Ges.* **45**, 756–766; 10.1002/cber.191204501110 (1912).
43. Farbwerke vorm. Meister Lucius & Brüning in Höchst Verfahren zur Darstellung von Oxyarylarsenoxyden. *Deutsches Reichspatent No. 213594 (1908)* (2135).
44. Farbwerke vorm. Meister Lucius & Brüning in Höchst Verfahren zur Darstellung von Aminoderivaten der Oxyarylarsinsäuren und deren Reduktionsprodukten. *Deutsches Reichspatent No. 224953 (1909)* (2249).
45. Farbwerke vorm. Meister Lucius & Brüning in Höchst Verfahren zur Darstellung von Aminoxyarylarsenoxyden. *Deutsches Reichspatent No. 235391 (1909)* (2353).
46. Williams, K. J. The introduction of 'chemotherapy' using arsphenamine – the first magic bullet. *J. Royal Soci. Med.* **102**, 343–348; 10.1258/jrsm.2009.09k036 (2009).
47. 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).

48. Bo, G. Giuseppe Brotzu and the discovery of cephalosporins. *Clin. Microbiol. Infect.* **6**, 6–8; 10.1111/j.1469-0691.2000.tb02032.x (2000).
49. Newton, G. G. F. & Abraham, E. P. Cephalosporin C, a New Antibiotic containing Sulphur and D- $\alpha$ -Aminoadipic Acid. *Nature* **175**, 548; 10.1038/175548a0 (1955).
50. Schatz, A., Bugle, E. & Waksman, S. A. Streptomycin, a Substance Exhibiting Antibiotic Activity Against Gram-Positive and Gram-Negative Bacteria. *Exp. Biol. and Med.* **55**, 66–69; 10.3181/00379727-55-14461 (1944).
51. Jones, D., Metzger, H. J., Schatz, A. & Waksman, S. A. Control of Gram-negative Bacteria in Experimental Animals by Streptomycin. *Science* **100**, 103–105; 10.1126/science.100.2588.103 (1944).
52. DeMirci, H. *et al.* A structural basis for streptomycin-induced misreading of the genetic code. *Nat. Commun.* **4**; 10.1038/ncomms2346 (2013).
53. Chopra, I. & Roberts, M. Tetracycline Antibiotics: Mode of Action, Applications, Molecular Biology, and Epidemiology of Bacterial Resistance. *Microbiol. Mol. Bio. Rev.* **65**, 232–260; 10.1128/mubr.65.2.232-260.2001 (2001).
54. 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; 10.1126/science.106.2757.417 (1947).
55. Drainas, D., Kalpaxis, D. L. & Coutsogeorgopoulos, C. Inhibition of ribosomal peptidyltransferase by chloramphenicol. Kinetic studies. *Eur.J. Biochem.* **164**, 53–58; 10.1111/j.1432-1033.1987.tb10991.x (1987).
56. Schlunzen, F. *et al.* Structural basis for the interaction of antibiotics with the peptidyl transferase centre in eubacteria. *Nature* **413**, 814–821 (2001).
57. McGuire, J. M. *et al.* Ilotycin, ein neues Antibiotikum. *Schweiz. Med. Wochenschr.* **82**, 1064-1065 (1952).
58. Wolfe, A. D. & Hahn, F. E. Erythromycin: Mode of Action. *Science* **143**, 1445–1446; 10.1126/science.143.3613.1445 (1964).
59. Griffith, R. S. Introduction to Vancomycin. *Clin. Infect. Dis.* **3**, S200-S204; 10.1093/clinids/3.supplement\_2.s200 (1981).

60. Courvalin, P. Vancomycin Resistance in Gram-Positive Cocci. *Clin. Infect. Dis.* **42**, S25-S34; 10.1086/491711 (2006).
61. Charney, J., Fisher, W. P., Curran, C., Machlowitz, R. A. & Tytell, A. A. Streptogramin, a new antibiotic. *Antibiot. Chemother. (Northfield, Ill.)* **3**, 1283 (1953).
62. Ennis, H. L. Inhibition of protein synthesis by polypeptide antibiotics I. Inhibition in intact bacteria. *J. Bacteriol.* **90**, 1102-1108 (1965).
63. Cocito, C. Metabolism of macromolecules in bacteria treated with virginiamycin. *Microbiology* **57**, 179-194 (1969).
64. Leshner, G. Y., Froelich, E. J., Gruett, M. D., Bailey, J. H. & Brundage, R. P. 1,8-Naphthyridine Derivatives. A New Class of Chemotherapeutic Agents. *J. Med. Pharm. Chem.* **5**, 1063–1065; 10.1021/jm01240a021 (1962).
65. Sugino, A., Peebles, C. L., Kreuzer, K. N. & Cozzarelli, N. R. Mechanism of action of nalidixic acid: Purification of *Escherichia coli* nalA gene product and its relationship to DNA gyrase and a novel nicking-closing enzyme. *Proc. Natl. Acad. Sci.* **74**, 4767–4771; 10.1073/pnas.74.11.4767 (1977).
66. Fournier, B., Zhao, X., Lu, T., Drlica, K. & Hooper, D. C. Selective Targeting of Topoisomerase IV and DNA Gyrase in *Staphylococcus aureus*: Different Patterns of Quinolone- Induced Inhibition of DNA Synthesis. *Antimicrob. Agents Chemother.* **44**, 2160–2165; 10.1128/aac.44.8.2160-2165.2000 (2000).
67. Katherine H. Luepke *et al.* Past, Present, and Future of Antibacterial Economics: Increasing Bacterial Resistance, Limited Antibiotic Pipeline, and Societal Implications. *Pharmacotherapy: The Journal of Hum. Pharmacol. Drug Ther.* **37**, 71–84; 10.1002/phar.1868 (2017).
68. Debono M. *et al.* A21978C, A Complex of New Acidic Peptide Antibiotics: Isolation, Chemistry, and Mass Spectral Structure Elucidation. *J. Antibiot.* **40**, 761–777; 10.7164/antibiotics.40.761 (1987).
69. Eliopoulos, G. M., Thauvin, C., Gerson B. & Moellering R. C. *In vitro* activity and mechanism of action of A21978C1, a novel cyclic lipopeptide antibiotic. *Antimicrob. Agents Chemother.* **27**, 357–362; 10.1128/AAC.27.3.357 (1985).
70. Ford, C. W., Zurenko G. E. & Barbachyn M. R. The Discovery of Linezolid, the First Oxazolidinone Antibacterial Agent. *Curr. Drug Targets - Infect. Disord.* **1**, 181–199 (2001).

71. Palumbi, S. R. Humans as the world's greatest evolutionary force. *Science (New York, N.Y.)* **293**, 1786–1790; 10.1126/science.293.5536.1786 (2001).
72. Bhullar, K. *et al.* Antibiotic resistance is prevalent in an isolated cave microbiome. *PLoS ONE* **7**, e34953; 10.1371/journal.pone.0034953 (2012).
73. Li, X.-Z. & Nikaido, H. Efflux-Mediated Drug Resistance in Bacteria. *Drugs* **69**, 1555–1623; 10.2165/11317030-000000000-00000 (2009).
74. Blair, J. M. A., Webber, M. A., Baylay, A. J., Ogbolu, D. O. & Piddock, L. J. V. Molecular mechanisms of antibiotic resistance. *Nat. Rev. Microbiol.* **13**, 42–51; 10.1038/nrmicro3380 (2014).
75. Aminov, R. I. & Mackie, R. I. Evolution and ecology of antibiotic resistance genes. *FEMS microbiol. Lett.* **271**, 147–161; 10.1111/j.1574-6968.2007.00757.x (2007).
76. O'Neil, J. Antimicrobial Resistance: Tackling a crisis for the health and wealth of nations (2014). Available at [https://amr-review.org/sites/default/files/AMR%20Review%20Paper%20-%20Tackling%20a%20crisis%20for%20the%20health%20and%20wealth%20of%20nations\\_1.pdf](https://amr-review.org/sites/default/files/AMR%20Review%20Paper%20-%20Tackling%20a%20crisis%20for%20the%20health%20and%20wealth%20of%20nations_1.pdf) (accessed 14.09.2020).
77. Tacconelli, E. *et al.* Discovery, research, and development of new antibiotics: the WHO priority list of antibiotic-resistant bacteria and tuberculosis. *Lancet Infect. Dis.* **18**, 318–327; 10.1016/S1473-3099(17)30753-3 (2018).
78. World Health Organization. Global Antimicrobial Resistance Surveillance System (GLASS) Report. Early implementation (2018). Available at <http://apps.who.int/iris/bitstream/handle/10665/259744/9789241513449-eng.pdf;jsessionid=97E045EBFE3B29F286CBA3EE8360F0D1?sequence=1> (accessed 14.09.2020).
79. Boucher, H. W. *et al.* Bad Bugs, No Drugs: No ESCAPE! An Update from the Infectious Diseases Society of America. *Clin. Infect. Dis.* **48**, 1–12; 10.1086/595011 (2009).
80. Kupferschmidt, K. Resistance fighters. *Science* **352**, 758–761; 10.1126/science.352.6287.758 (2016).
81. van Duin, D. & Bonomo, R. A. Ceftazidime/Avibactam and Ceftolozane/Tazobactam: Second-generation  $\beta$ -Lactam/ $\beta$ -Lactamase Inhibitor Combinations. *Clin. Infect. Dis.* **63**, 234–241; 10.1093/cid/ciw243 (2016).

- 
82. Wunderink, R. G. *et al.* Effect and Safety of Meropenem-Vaborbactam versus Best-Available Therapy in Patients with Carbapenem-Resistant Enterobacteriaceae Infections: The TANGO II Randomized Clinical Trial. *Infect. Dis. Ther.* **7**, 439–455; 10.1007/s40121-018-0214-1 (2018).
83. Saravolatz, L. D. & Stein, G. E. Plazomicin: A New Aminoglycoside. *Clin. Infect. Dis.* **70**, 704–709; 10.1093/cid/ciz640 (2020).
84. Sutcliffe, J. A., O'Brien, W., Fyfe, C. & Grossman, T. H. Antibacterial activity of eravacycline (TP-434), a novel fluorocycline, against hospital and community pathogens. *Antimicrob. Agents Chemother.* **57**, 5548–5558; 10.1128/AAC.01288-13 (2013).
85. Kostyanev, T. *et al.* The Innovative Medicines Initiative's New Drugs for Bad Bugs programme: European public-private partnerships for the development of new strategies to tackle antibiotic resistance. *J. Antimicrob. Chemother.* **71**, 290–295; 10.1093/jac/dkv339 (2016).
86. Stekel, D. First report of antimicrobial resistance pre-dates penicillin. *Nature* **562**; 10.1038/d41586-018-06983-0 (2018).
87. Hutchings, M., Truman, A. & Wilkinson, B. Antibiotics: past, present and future. *Curr. Opin. Microbiol.* **51**, 72–80; 10.1016/j.mib.2019.10.008 (2019).
88. Dublanchet, A., Soussy, C. J., Squinazi, F. & Duval, J. Résistance de *Staphylococcus aureus* aux streptogramines. *Ann. Inst. Pasteur Microbiol.* **128A**, 277–287 (1977).
89. Emmerson, A. M. & Jones, A. M. The quinolones: decades of development and use. *J. Antimicrob. Chemother.* **51**, 13–20; 10.1093/jac/dkg208 (2003).
90. Gellert, M., Mizuuchi, K., O'Dea, M. H., Itoh, T. & Tomizawa, J.-I. Nalidixic acid resistance: A second genetic character involved in DNA gyrase activity. *PNAS* **74**, 4772–4776; 10.1073/pnas.74.11.4772 (1977).
91. Tsiodras, S. *et al.* Linezolid resistance in a clinical isolate of *Staphylococcus aureus*. *The Lancet* **358**, 207–208; 10.1016/s0140-6736(01)05410-1 (2001).
92. Gonzales, R. D. *et al.* Infections due to vancomycin-resistant *Enterococcus faecium* resistant to linezolid. *The Lancet* **357**, 1179; 10.1016/s0140-6736(00)04376-2 (2001).
93. Mangili, A., Bica, I., Snyderman, D. R. & Hamer, D. H. Daptomycin-Resistant, Methicillin-Resistant *Staphylococcus aureus* Bacteremia. *Clin. Infect. Dis.* **40**, 1058–1060; 10.1086/428616 (2005).

94. Hayden, M. K. *et al.* Development of Daptomycin Resistance *In Vivo* in Methicillin-Resistant *Staphylococcus aureus*. *J. Clin. Microbiol.* **43**, 5285–5287; 10.1128/JCM.43.10.5285-5287.2005 (2005).
95. Sabol, K. *et al.* Emergence of Daptomycin Resistance in *Enterococcus faecium* during Daptomycin Therapy. *Antimicrob. Agents Chemother.* **49**, 1664–1665; 10.1128/AAC.49.4.1664-1665.2005 (2005).
96. Ling, L. L. *et al.* A new antibiotic kills pathogens without detectable resistance. *Nature* **517**, 455–459; 10.1038/nature14098 (2015).
97. Zipperer, A. *et al.* Human commensals producing a novel antibiotic impair pathogen colonization. *Nature* **535**, 511–516; 10.1038/nature18634 (2016).
98. Schilling, N. A. *et al.* Synthetic Lugdunin Analogues Reveal Essential Structural Motifs for Antimicrobial Action and Proton Translocation Capability. *Angew. Chem. Int. Ed. Engl.* **58**, 9234–9238; 10.1002/anie.201901589 (2019).
99. Baumann, S. *et al.* Cystobactamids: myxobacterial topoisomerase inhibitors exhibiting potent antibacterial activity. *Angew. Chem. Int. Ed.* **53**, 14605–14609; 10.1002/anie.201409964 (2014).
100. Hüttel, S. *et al.* Discovery and Total Synthesis of Natural Cystobactamid Derivatives with Superior Activity against Gram-Negative Pathogens. *Angew. Chem. Int. Ed. Engl.* **56**, 12760–12764; 10.1002/anie.201705913 (2017).
101. Imai, Y. *et al.* A new antibiotic selectively kills Gram-negative pathogens. *Nature*; 10.1038/s41586-019-1791-1 (2019).
102. Nikaido, H. Outer membrane barrier as a mechanism of antimicrobial resistance. *Antimicrob. Agents Chemother.* **33**, 1831–1836; 10.1128/AAC.33.11.1831 (1989).
103. Facchini, P. J. Alkaloid Biosynthesis in Plants: Biochemistry, Cell Biology, Molecular Regulation, and Metabolic Engineering Applications. *Ann. Rev. Plant Biol.* **52**, 29-66; 10.1146/annurev.arplant.52.1.29 (2001).
104. Lange, B. M. The evolution of plant secretory structures and emergence of terpenoid chemical diversity. *Ann. Rev. Plant Biol.* **66**, 139–159; 10.1146/annurev-arplant-043014-114639 (2015).

105. Keller, N. P., Turner, G. & Bennett, J. W. Fungal secondary metabolism - from biochemistry to genomics. *Nat. Rev. Microbiol.* **3**, 937–947; 10.1038/nrmicro1286 (2005).
106. Imhoff, J. F. Natural Products from Marine Fungi — Still an Underrepresented Resource. *Marine drugs* **14**, 19; 10.3390/md14010019 (2016).
107. Yarza, P. *et al.* Uniting the classification of cultured and uncultured bacteria and archaea using 16S rRNA gene sequences. *Nat. Rev. Microbiol.* **12**, 635–645; 10.1038/nrmicro3330 (2014).
108. Barka, E. A. *et al.* Taxonomy, Physiology, and Natural Products of Actinobacteria. *Microbiol. Mol. Biol. Rev.* **80**, 1–43; 10.1128/MMBR.00019-15 (2016).
109. Waksman, S. A. Production and Activity of Streptothricin<sup>12</sup>. *J. Bacteriol.* **46**, 299–310 (1943).
110. Challinor, V. L. & Bode, H. B. Bioactive natural products from novel microbial sources. *Ann. N. Y. Acad. Sci.* **1354**, 82–97; 10.1111/nyas.12954 (2015).
111. Baltz, R. H. Gifted microbes for genome mining and natural product discovery. *J. Ind. Microbiol. Biotechnol.*; 10.1007/s10295-016-1815-x (2016).
112. Baltz, R. Antimicrobials from Actinomycetes: Back to the Future. *Microbe* **2**, 125–131 (2007).
113. Watve, M. G., Tickoo, R., Jog, M. M. & Bhole, B. D. How many antibiotics are produced by the genus *Streptomyces*? *Arch. Microbiol.* **176**, 386–390; 10.1007/s002030100345 (2001).
114. Ringel, S. M. *et al.* Ambruticin (W7783), a new antifungal antibiotic. *J. Antibiot.* **30**, 371–375 (1977).
115. Gerth, K., Irschik, H., Reichenbach, H. & Trowitzsch, W. Myxothiazol, an antibiotic from *Myxococcus fulvus* (myxobacterales). I. Cultivation, isolation, physico-chemical and biological properties. *J. Antibiot.* **33**, 1474–1479 (1980).
116. Wenzel, S. C. & Müller, R. in *Industrial biotechnology, Microorganisms Volume 3a and 3b*, edited by C. Wittmann & J. Liao (Wiley-VCH, Weinheim, Germany, 2017), pp. 453–485.
117. 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).

118. Weissman, K. J. & Müller, R. Myxobacterial secondary metabolites: bioactivities and modes-of-action. *Nat. Prod. Rep.* **27**, 1276–1295; 10.1039/c001260m (2010).
119. Reichenbach, H. Myxobacteria, producers of novel bioactive substances. *J. Ind. Microbiol. Biotechnol.* **27**, 149–156 (2001).
120. 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; 10.3389/fmicb.2016.00781 (2016).
121. Zaburanyi, N., Bunk, B., Maier, J., Overmann, J. & Müller, R. Genome analysis of the fruiting body forming myxobacterium *Chondromyces crocatus* reveals high potential for natural product Biosynthesis. *Appl. Environ. Microbiol.* **82**, 1945–1957; 10.1128/AEM.03011-15 (2016).
122. Garcia, R., Gemperlein, K. & Müller, R. *Minicystis rosea* gen. nov., sp. nov., a polyunsaturated fatty acid-rich and steroid-producing soil myxobacterium. *Int. J. Syst. Evol. Microbiol.* **64**, 3733–3742; 10.1099/ijs.0.068270-0 (2014).
123. Jansen, R., Irschik, H., Reichenbach, H. & Höfle, G. Antibiotics from gliding bacteria, LXXX. Chivosazoles A-F: Novel antifungal and cytotoxic macrolides from *Sorangium cellulosum* (Myxobacteria). *Liebigs Ann. Chem.* **1997**, 1725–1732 (1997).
124. Irschik, H., Jansen, R., Gerth, K., Höfle, G. & Reichenbach, H. Sorangiolid A, a new antibiotic isolated from the myxobacterium *Sorangium cellulosum* So ce 12. *J. Antibiot.* **48**, 886–887 (1995).
125. Zander, W. *et al.* Sulfangolids, macrolide sulfate esters from *Sorangium cellulosum*. *Chem. Eur. J.* **18**, 6264–6271; 10.1002/chem.201100851 (2012).
126. Irschik, H., Jansen, R., Gerth, K., Hofle, G. & Reichenbach, H. The sorangicins, novel and powerful inhibitors of eubacterial RNA polymerase isolated from myxobacteria. *J. Antibiot.* **40**, 7–13 (1987).
127. Reichenbach, H. The ecology of the myxobacteria. *Environ. Microbiol.* **1**, 15–21 (1999).
128. Brinkhoff, T. *et al.* Biogeography and phylogenetic diversity of a cluster of exclusively marine myxobacteria. *ISME J.* **6**, 1260–1272; 10.1038/ismej.2011.190 (2012).
129. O'Connor, K. A. & Zusman, D. R. Behavior of peripheral rods and their role in the life cycle of *Myxococcus xanthus*. *J. Bacteriol.* **173**, 3342–3355 (1991).

130. 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).
131. Shimkets, L. J. Intercellular signaling during fruiting-body development of *Myxococcus xanthus*. *Annu. Rev. Microbiol.* **53**, 525–549 (1999).
132. 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; 10.1128/MMBR.00043-09 (2010).
133. 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; 10.1038/nrmicro1770 (2007).
134. Conlin, A., Fournier, M., Hudis, C., Kar, S. & Kirkpatrick, P. Ixabepilone. *Nat. Rev. Drug Discov.* **6**, 953–954; 10.1038/nrd2469 (2007).
135. Mulzer, J. *The epothilones: An outstanding family of anti-tumour Agents: from soil to the clinic* (Springer, New York, 2009).
136. Berod, L. *et al.* De novo fatty acid synthesis controls the fate between regulatory T and T helper 17 cells. *Nat. Med.* **20**, 1327–1333; 10.1038/nm.3704 (2014).
137. Held, J. *et al.* Antimalarial activity of the myxobacterial macrolide chlorotonil A. *Antimicrob. Agents Chemother.* **58**, 6378–6384; 10.1128/AAC.03326-14 (2014).
138. Dumez, H. *et al.* A phase I and pharmacokinetic study of LAF389 administered to patients with advanced cancer. *Anticancer. Drugs* **18**, 219–225 (2007).
139. Kaur, G. *et al.* Biological evaluation of tubulysin A: a potential anticancer and antiangiogenic natural product. *Biochem. J.* **396**, 235–242; 10.1042/BJ20051735 (2006).
140. Guenther, E. *et al.* Targeting Disorazol Z to LHRH-receptor positive tumors by the cytotoxic conjugate AEZS-125. *Cancer Res* **68**, 782 (2008).
141. Schäberle, T. F. *et al.* Corallopyronin A - A promising antibiotic for treatment of filariasis. *Int. J. Med. Microbiol.* **304**, 72–78; 10.1016/j.ijmm.2013.08.010 (2014).
142. Weissman, K. J. The structural biology of biosynthetic megaenzymes. *Nat. Chem. Biol.* **11**, 660–670; 10.1038/nchembio.1883 (2015).

143. Walsh, C. T., Gehring, A. M., Weinreb, P. H., Quadri, L. E. & Flugel, R. S. Post-translational modification of polyketide and nonribosomal peptide synthases. *Curr. Opin. Chem. Biol.* **1**, 309–315 (1997).
144. Gehret, J. J. *et al.* Terminal alkene formation by the thioesterase of curacin a biosynthesis structure of a decarboxylating thioesterase. *J. Biol. Chem.* **286**, 14445–14454; 10.1074/jbc.M110.214635 (2011).
145. Gokhale, R. S., Hunziker, D., Cane, D. E. & Khosla, C. Mechanism and specificity of the terminal thioesterase domain from the erythromycin polyketide synthase. *Chem. Biol.* **6**, 117–125 (1999).
146. Strieker, M., Tanović, A. & Marahiel, M. A. Nonribosomal peptide synthetases. Structures and dynamics. *Curr. Opin. Struct. Biol.* **20**, 234–240; 10.1016/j.sbi.2010.01.009 (2010).
147. 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).
148. Fischbach, M. A. & Walsh, C. T. Assembly-line enzymology for polyketide and nonribosomal Peptide antibiotics: logic, machinery, and mechanisms. *Chem. Rev.* **106**, 3468–3496; 10.1021/cr0503097 (2006).
149. Süßmuth, R. D. & Mainz, A. Nonribosomal peptide synthesis - Principles and prospects. *Angew. Chem. Int. Ed.* **56**, 3770–3821; 10.1002/anie.201609079 (2017).
150. Silakowski, B., Kunze, B. & Müller, R. Multiple hybrid polyketide synthase/non-ribosomal peptide synthetase gene clusters in the myxobacterium *Stigmatella aurantiaca*. *Gene* **275**, 233–240 (2001).
151. Wenzel, S. C., Meiser, P., Binz, T. M., Mahmud, T. & Müller, R. Nonribosomal Peptide Biosynthesis: Point Mutations and Module Skipping Lead to Chemical Diversity. *Angew. Chem. Int. Ed. Engl.* **45**, 2296–2301; 10.1002/anie.200503737 (2006).
152. Ohlendorf, B., Kehraus, S. & König, G. M. Myxochromide B<sub>3</sub>, a New Member of the Myxochromide Family of Secondary Metabolites. *J. Nat. Prod.* **71**, 1708–1713; 10.1021/np800319v (2008).
153. Burgard, C. *et al.* Genomics-Guided Exploitation of Lipopeptide Diversity in Myxobacteria. *ACS Chem. Biol.* **12**, 779–786; 10.1021/acscchembio.6b00953 (2017).

154. Wenzel, S. C. *et al.* Structure and Biosynthesis of Myxochromides S<sub>1-3</sub> in *Stigmatella aurantiaca*: Evidence for An Iterative Bacterial Type I Polyketide Synthase and for Module Skipping in Nonribosomal Peptide Biosynthesis. *ChemBioChem* **6**, 375–385; 10.1002/cbic.200400282 (2005).
155. Jenke-Kodama, H., Sandmann, A., Müller, R. & Dittmann, E. Evolutionary implications of bacterial polyketide synthases. *Mol. Biol. Evol.* **22**, 2027–2039; 10.1093/molbev/msi193 (2005).
156. Hertweck, C. The Biosynthetic Logic of Polyketide Diversity. *Angew. Chem. Int. Ed. Engl.* **48**, 4688–4716; 10.1002/anie.200806121 (2009).
157. Staunton, J. & Weissman, K. J. Polyketide biosynthesis: a millennium review. *Nat. Prod. Rep.* **18**, 380–416 (2001).
158. Kohli, R. M. & Walsh, C. T. Enzymology of acyl chain macrocyclization in natural product biosynthesis. *Chem. Commun.*, 297–307; 10.1039/b208333g (2003).
159. Rawlings, B. J. Type I polyketide biosynthesis in bacteria (Part A-erythromycin biosynthesis). *Nat. Prod. Rep.* **18**, 190–227 (2001).
160. Weinig, S., Hecht, H.-J., Mahmud, T. & Müller, R. Melithiazol biosynthesis: further insights into myxobacterial PKS/NRPS systems and evidence for a new subclass of methyl transferases. *Chem. Biol.* **10**, 939–952; 10.1016/j.chembiol.2003.09.012 (2003).
161. Müller, R. in *The Epothilones, an Outstanding Family of Anti-Tumor Agents*, edited by J. Mulzer (Springer, New York, 2009), pp. 29–53.
162. Zhu, G. *et al.* *Cryptosporidium parvum*: the first protist known to encode a putative polyketide synthase. *Gene* **298**, 79–89 (2002).
163. Snyder, R. V. *et al.* Polyketide synthase genes from marine dinoflagellates. *Mar. Biotechnol.* **5**, 1–12; 10.1007/s10126-002-0077-y (2003).
164. Herbst, D. A., Townsend C. A. & Maier T. The architectures of iterative type I PKS and FAS. *Nat. Prod. Rep.* **35**, 1046–1069; 10.1039/C8NP00039E (2018).
165. 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; 10.1039/b507395m (2007).

166. Hutchinson, C. R. Biosynthetic Studies of Daunorubicin and Tetracenomyacin C. *Chem. Rev.* **97**, 2525–2536 (1997).
167. Pickens, L. B. & Tang, Y. Oxytetracycline biosynthesis. *J. Biol. Chem.* **285**, 27509–27515; 10.1074/jbc.R110.130419 (2010).
168. 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; 10.1002/anie.200603513 (2007).
169. Li, X. W. *et al.* Synthesis and biological activities of the respiratory chain inhibitor aurachin D and new ring versus chain analogues. *Beilstein J. Org. Chem.* **9**, 1551–1558; 10.3762/bjoc.9.176 (2013).
170. Moore, B. S. & Höpke, J. N. Discovery of a new bacterial polyketide biosynthetic pathway. *ChemBioChem* **2**, 35–38 (2001).
171. 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; 10.1074/jbc.M106580200 (2001).
172. Seshime, Y., Juvvadi, P. R., Fujii, I. & Kitamoto, K. Discovery of a novel superfamily of type III polyketide synthases in *Aspergillus oryzae*. *Biochem. Biophys. Res. Commun.* **331**, 253–260 (2005).
173. Austin, M. B. & Noel, J. P. The chalcone synthase superfamily of type III polyketide synthases. *Nat. Prod. Rep.* **20**, 79–110; 10.1039/B100917F (2003).
174. 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; 10.1007/s10295-018-2105-6 (2019).
175. 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; 10.1007/s00203-005-0059-3 (2006).
176. Caboche, S. *et al.* Norine: a database of nonribosomal peptides. *Nucleic Acids Res.* **36**, D326–D331; 10.1093/nar/gkm792 (2008).
177. Mootz, H. D., Schwarzer, D. & Marahiel, M. A. Ways of assembling complex natural products on modular nonribosomal peptide synthetases. *ChemBioChem* **3**, 490–504 (2002).

- 
178. Drake, E. J. *et al.* Structures of two distinct conformations of holo-non-ribosomal peptide synthetases. *Nature* **529**, 235–238; 10.1038/nature16163 (2016).
179. Belshaw, P. J., Walsh, C. T. & Stachelhaus, T. Aminoacyl-CoAs as probes of condensation domain selectivity in nonribosomal peptide synthesis. *Science* **284**, 486–489 (1999).
180. Keating, T. A., Marshall, C. G., Walsh, C. T. & Keating, A. E. The structure of VibH represents nonribosomal peptide synthetase condensation, cyclization and epimerization domains. *Nat. Struct. Biol.* **9**, 522–526; 10.1038/nsb810 (2002).
181. 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; 10.1186/1471-2148-7-78 (2007).
182. Weckwerth, W. *et al.* Biosynthesis of PF1022A and related cyclooctadepsipeptides. *J. Biol. Chem.* **275**, 17909–17915; 10.1074/jbc.M001084200 (2000).
183. Stachelhaus, T. & Walsh, C. T. Mutational Analysis of the Epimerization Domain in the Initiation Module PheATE of Gramicidin S Synthetase. *Biochemistry* **39**, 5775–5787; 10.1021/bi9929002 (2000).
184. Schneider, T. L., Shen, B. & Walsh, C. T. Oxidase Domains in Epothilone and Bleomycin Biosynthesis: Thiazoline to Thiazole Oxidation during Chain Elongation. *Biochemistry* **42**, 9722–9730; 10.1021/bi034792w (2003).
185. Stachelhaus, T. & Marahiel, M. A. Modular structure of genes encoding multifunctional peptide synthetases required for non-ribosomal peptide synthesis. *FEMS Microbiol. Lett.* **125**, 3–14 (1995).
186. 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).
187. Mootz, H. D. & Marahiel, M. A. The tyrocidine biosynthesis operon of *Bacillus brevis*: complete nucleotide sequence and biochemical characterization of functional internal adenylation domains. *J. Bacteriol.* **179**, 6843–6850 (1997).
188. Grace Yim, Maulik N Thaker, Kalinka Koteva & Gerard Wright. Glycopeptide antibiotic biosynthesis. *J. Antibiot.* **67**, 31–41; 10.1038/ja.2013.117 (2014).

189. Gehring, A. M., Mori, I. & Walsh, C. T. Reconstitution and characterization of the *Escherichia coli* enterobactin synthetase from EntB, EntE, and EntF. *Biochemistry* **37**, 2648–2659 (1998).
190. 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; 10.1021/bi010036j (2001).
191. Shaw-Reid, C. A. *et al.* Assembly line enzymology by multimodular nonribosomal peptide synthetases: the thioesterase domain of *E. coli* EntF catalyzes both elongation and cyclolactonization. *Chem. Biol.* **6**, 385–400 (1999).
192. Du, L., Sánchez, C., Chen, M., 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; 10.1016/S1074-5521(00)00011-9 (2000).
193. Singh, G. M., Vaillancourt, F. H., Yin, J. & Walsh, C. T. Characterization of SyrC, an aminoacyltransferase shuttling threonyl and chlorothreonyl residues in the syringomycin biosynthetic assembly line. *Chem. Biol.* **14**, 31–40; 10.1016/j.chembiol.2006.11.005 (2007).
194. Cociancich, S. *et al.* The gyrase inhibitor albicidin consists of *p*-aminobenzoic acids and cyanoalanine. *Nat. Chem. Biol.* **11**, 195–197; 10.1038/nchembio.1734 (2015).
195. Du, L. & Shen, B. Identification and characterization of a type II peptidyl carrier protein from the bleomycin producer *Streptomyces verticillus* ATCC 15003. *Chemistry & Biology* **6**, 507–517; 10.1016/s1074-5521(99)80083-0 (1999).
196. Müller, R. Don't classify polyketide synthases. *Chem. Biol.* **11**, 4–6; 10.1016/j.chembiol.2004.01.005 (2004).
197. Konz, D., Doekel, S. & Marahiel, M. A. Molecular and biochemical characterization of the protein template controlling biosynthesis of the lipopeptide lichenysin. *J. Bacteriol.* **181**, 133–140 (1999).
198. Gao, X. *et al.* Cyclization of fungal nonribosomal peptides by a terminal condensation-like domain. *Nat. Chem. Biol.* **8**, 823–830; 10.1038/nchembio.1047. (2012).
199. Walsh, C. T. Insights into the chemical logic and enzymatic machinery of NRPS assembly lines. *Nat. Prod. Rep.* **33**, 127–135; 10.1039/c5np00035a (2016).

- 
200. Du, L. & Lou, L. PKS and NRPS release mechanisms. *Nat. Prod. Rep.* **27**, 255–278; 10.1039/b912037h (2010).
201. Müller, S. *et al.* Biosynthesis of crocacin involves an unusual hydrolytic release domain showing similarity to condensation domains. *Chem. Biol.* **21**, 855–865; 10.1016/j.chembiol.2014.05.012 (2014).
202. Frueh, D. P. *et al.* Dynamic thiolation-thioesterase structure of a non-ribosomal peptide synthetase. *Nature* **454**, 903–906; 10.1038/nature07162 (2008).
203. Kotowska, M. & Pawlik, K. Roles of type II thioesterases and their application for secondary metabolite yield improvement. *Appl. Microbiol. Biotechnol.* **98**, 7735–7746; 10.1007/s00253-014-5952-8 (2014).
204. Yu, F.-M., Qiao, B., Zhu, F., Wu, J.-C. & Yuan, Y.-J. Functional Analysis of Type II Thioesterase of *Streptomyces lydicus* AS 4.2501. *ABAB* **135**, 145–158; 10.1385/abab:135:2:145 (2006).
205. Fujimori, D. G. *et al.* Cloning and characterization of the biosynthetic gene cluster for kutznerides. *Proc. Natl. Acad. Sci. USA* **104**, 16498–16503 (2007).
206. Chen, H., Hubbard, B. K., O'Connor, S. E. & Walsh, C. T. Formation of beta-hydroxy histidine in the biosynthesis of nikkomycin antibiotics. *Chem. Biol.* **9**, 103–112 (2002).
207. Chatterjee, J., Rechenmacher, F. & Kessler, H. *N*-Methylation of Peptides and Proteins: An Important Element for Modulating Biological Functions. *Angew. Chem. Int. Ed. Engl.* **52**, 254–269; 10.1002/anie.201205674 (2013).
208. Velkov, T. *et al.* Characterization of the *N*-methyltransferase activities of the multifunctional polypeptide cyclosporin synthetase. *Chemistry & biology* **18**, 464–475; 10.1016/j.chembiol.2011.01.017 (2011).
209. Müller, S. *et al.* Paenilamicin: structure and biosynthesis of a hybrid nonribosomal peptide/polyketide antibiotic from the bee pathogen *Paenibacillus larvae*. *Angew. Chem. Int. Ed. Engl.*; 10.1002/anie.201404572 (2014).
210. Zolova, O. E. & Garneau-Tsodikova, S. KtzJ-dependent serine activation and *O*-methylation by KtzH for kutznerides biosynthesis. *J. Antibiot.* **67**, 59–64; 10.1038/ja.2013.98 (2014).

211. Al-Mestarihi, A. H. *et al.* Adenylation and S-Methylation of Cysteine by the Bifunctional Enzyme TioN in Thiocoraline Biosynthesis. *J. Am. Chem. Soc.* **136**, 17350–17354; 10.1021/ja510489j (2014).
212. Miller, D. A., Walsh, C. T. & Luo, L. S. C-methyltransferase and cyclization domain activity at the intraprotein PK/NRP switch point of yersiniabactin synthetase. *J. Am. Chem. Soc.* **123**, 8434–8435; 10.1021/ja016398w (2001).
213. Hoffmann, K., Schneider-Scherzer, E., Kleinkauf, H. & Zocher, R. Purification and characterization of eucaryotic alanine racemase acting as key enzyme in cyclosporin biosynthesis. *J. Biol. Chem.* **269**, 12710–12714 (1994).
214. 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; 10.1073/pnas.1608615113 (2016).
215. 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; 10.1128/AEM.01503-06 (2006).
216. Gerth, K. *et al.* The myxalamids, new antibiotics from *Myxococcus xanthus* (Myxobacteriales). I. Production, physico-chemical and biological properties, and mechanism of action. *J. Antibiot.* **36**, 1150–1156 (1983).
217. Barajas, J. F. *et al.* Comprehensive Structural and Biochemical Analysis of the Terminal Myxalamid Reductase Domain for the Engineered Production of Primary Alcohols. *Chemistry & biology* **22**, 1018–1029; 10.1016/j.chembiol.2015.06.022 (2015).
218. Schoenafinger, G., Schracke, N., Linne, U. & Marahiel, M. A. Formylation domain: An essential modifying enzyme for the nonribosomal biosynthesis of linear gramicidin. *J. Am. Chem. Soc.* **128**, 7406–7407; 10.1021/ja0611240 (2006).
219. Mootz, H. D. & Marahiel, M. A. The tyrocidine biosynthesis operon of *Bacillus brevis*: complete nucleotide sequence and biochemical characterization of functional internal adenylation domains. *J. Bacteriol.* **179**, 6843–6850 (1997).
220. Stachelhaus, T. & Marahiel, M. A. Modular structure of peptide synthetases revealed by dissection of the multifunctional enzyme GrsA. *J. Biol. Chem.* **270**, 6163–6169 (1995).
221. Conti, E., Franks, N. P. & Brick, P. Crystal structure of firefly luciferase throws light on a superfamily of adenylate-forming enzymes. *Structure* **4**, 287–298 (1996).

- 
222. 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).
223. Stachelhaus, T., Mootz, H. D. & Marahiel, M. A. The specificity-conferring code of adenylation domains in nonribosomal peptide synthetases. *Chem. Biol.* **6**, 493–505; 10.1016/S1074-5521(99)80082-9 (1999).
224. Döhren, H. von, Dieckmann, R. & Pavela-Vrancic, M. The nonribosomal code. *Chem. Biol.* **6**, R273-R279 (1999).
225. 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; 10.1093/nar/gki885 (2005).
226. Prieto, C., García-Estrada, C., Lorenzana, D. & Martín, J. F. NRPSsp: non-ribosomal peptide synthase substrate predictor. *Bioinformatics (Oxford, England)* **28**, 426–427; 10.1093/bioinformatics/btr659 (2012).
227. Röttig, M. *et al.* NRPSpredictor2—a web server for predicting NRPS adenylation domain specificity. *Nucleic Acids Res.* **39**, W362-7; 10.1093/nar/gkr323 (2011).
228. Damir Baranašić *et al.* Predicting substrate specificity of adenylation domains of nonribosomal peptide synthetases and other protein properties by latent semantic indexing. *J. Ind. Microbiol. Biotechnol.* **41**, 461–467; 10.1007/s10295-013-1322-2 (2014).
229. Pogorevc, D. *et al.* Biosynthesis and Heterologous Production of Argyrins. *ACS Synth. Biol.* **8**, 1121–1133; 10.1021/acssynbio.9b00023 (2019).
230. Christiansen, G., Philmus, B., Hemscheidt, T. & Kurmayer, R. Genetic Variation of Adenylation Domains of the Anabaenopeptin Synthesis Operon and Evolution of Substrate Promiscuity. *J. Bacteriol.* **193**, 3822–3831; 10.1128/JB.00360-11 (2011).
231. Pogorevc, D. Thesis. Saarland University (2019).
232. 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; 10.1002/anie.201802479 (2018).

233. Chen, C.-Y., Georgiev, I., Anderson, A. C. & Donald, B. R. Computational structure-based redesign of enzyme activity. *Proc. Natl. Acad. Sci. USA* **106**, 3764–3769; 10.1073/pnas.0900266106 (2009).
234. Eppelmann, K., Stachelhaus, T. & Marahiel, M. A. Exploitation of the selectivity-conferring code of nonribosomal peptide synthetases for the rational design of novel peptide antibiotics. *Biochemistry* **41**, 9718–9726; 10.1021/bi0259406 (2002).
235. Stevens, B. W., Lilien, R. H., Georgiev, I., Donald, B. R. & Anderson, A. C. Redesigning the PheA domain of gramicidin synthetase leads to a new understanding of the enzyme's mechanism and selectivity. *Biochemistry* **45**, 15495–15504; 10.1021/bi061788m (2006).
236. Kaljunen, H. *et al.* Structural Elucidation of the Bispecificity of A Domains as a Basis for Activating Non-natural Amino Acids. *Angew. Chem. Int. Ed. Engl.* **54**, 8833–8836; 10.1002/anie.201503275 (2015).
237. Winn, M., Fyans, J. K., Zhuo, Y. & Micklefield, J. Recent advances in engineering nonribosomal peptide assembly lines. *Nat. Prod. Rep.* **33**, 317–347; 10.1039/C5NP00099H (2016).
238. Miao, V. *et al.* Genetic Engineering in *Streptomyces roseosporus* to Produce Hybrid Lipopeptide Antibiotics. *Chemistry & biology* **13**, 269–276; 10.1016/j.chembiol.2005.12.012 (2006).
239. Nguyen, K. T. *et al.* Combinatorial biosynthesis of novel antibiotics related to daptomycin. *Proc. Natl. Acad. Sci. USA* **103**, 17462–17467; 10.1073/pnas.0608589103 (2006).
240. Calcott, M. J. & Ackerley, D. F. Genetic manipulation of non-ribosomal peptide synthetases to generate novel bioactive peptide products. *Biotechnol. Lett.* **36**, 2407–2416; 10.1007/s10529-014-1642-y (2014).
241. Butz, D. *et al.* Module extension of a non-ribosomal peptide synthetase of the glycopeptide antibiotic balhimycin produced by *Amycolatopsis balhimycina*. *ChemBioChem* **9**, 1195–1200; 10.1002/cbic.200800068 (2008).
242. Bozhüyük, K. A. J. *et al.* De novo design and engineering of non-ribosomal peptide synthetases. *Nat. Chem.* **10**, 275–281; 10.1038/nchem.2890 (2018).
243. Bozhüyük, K. A. J. *et al.* Modification and de novo design of non-ribosomal peptide synthetases using specific assembly points within condensation domains. *Nat. Chem.* **11**, 653–661; 10.1038/s41557-019-0276-z (2019).

- 
244. Huo, L. *et al.* Heterologous expression of bacterial natural product biosynthetic pathways. *Nat. Prod. Rep.* **36**, 1412–1436; 10.1039/C8NP00091C (2019).
245. Ongley, S., Bian, X., Neilan, B. A. & Müller, R. Recent advances in the heterologous expression of microbial natural product biosynthetic pathways. *Nat. Prod. Rep.* **30**, 1121–1138; 10.1039/c3np70034h (2013).
246. Attila Bodor *et al.* Challenges of unculturable bacteria: environmental perspectives. *Rev. Environ. Sci. Biotechnol.* **19**, 1–22; 10.1007/s11157-020-09522-4 (2020).
247. Terpe, K. Overview of bacterial expression systems for heterologous protein production: from molecular and biochemical fundamentals to commercial systems. *Appl. Microbiol. Biotechnol.* **72**, 211–222; 10.1007/s00253-006-0465-8 (2006).
248. Myronovskiy, M. & Luzhetskyy, A. Heterologous production of small molecules in the optimized *Streptomyces* hosts. *Nat. Prod. Rep.*; 10.1039/c9np00023b (2019).
249. 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*; 10.1016/B978-0-12-409547-2.14818-8 (2020).
250. Hashimoto, T. *et al.* Biosynthesis of quinolidomicin, the largest known macrolide of terrestrial origin. Identification and Heterologous Expression of a Biosynthetic Gene Cluster over 200 kb. *Org. Lett.*; 10.1021/acs.orglett.8b03570 (2018).
251. Hashimoto, T. *et al.* Biosynthesis of versipelostatin: identification of an enzyme-catalyzed [4+2]-cycloaddition required for macrocyclization of spirotetronate-containing polyketides. *J. Am. Chem. Soc.* **137**, 572–575; 10.1021/ja510711x (2015).
252. Nah, H.-J., Pyeon, H.-R., Kang, S.-H., Choi, S.-S. & Kim, E.-S. Cloning and Heterologous Expression of a Large-sized Natural Product Biosynthetic Gene Cluster in *Streptomyces* Species. *Front. Microbiol.* **8**, 394; 10.3389/fmicb.2017.00394 (2017).
253. Wenzel, S. C. *et al.* Heterologous expression of a myxobacterial natural products assembly line in pseudomonads via red/ET recombineering. *Chem. Biol.* **12**, 349–356; 10.1016/j.chembiol.2004.12.012 (2005).
254. Panter, F., Krug, D., Baumann, S. & Müller, R. Self-resistance guided genome mining uncovers new topoisomerase inhibitors from myxobacteria. *Chem. Sci.* **9**, 4898–4908; 10.1039/C8SC01325J (2018).

255. Bilyk, O., Sekurova, O. N., Zotchev, S. B. & Luzhetskyy, A. Cloning and Heterologous Expression of the Grecoacycline Biosynthetic Gene Cluster. *PLoS ONE* **11**, e0158682; 10.1371/journal.pone.0158682 (2016).
256. Gibson, D. G. *et al.* Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat. Methods* **6**, 343–345 (2009).
257. Li, M. Z. & Elledge, S. J. Harnessing homologous recombination in vitro to generate recombinant DNA via SLIC. *Nat. Methods* **4**, 251–256; 10.1038/nmeth1010 (2007).
258. Quan, J. & Tian, J. Circular Polymerase Extension Cloning of Complex Gene Libraries and Pathways. *PLoS ONE* **4**, e6441; 10.1371/journal.pone.0006441 (2009).
259. Engler, C., Kandzia, R. & Marillonnet, S. A One Pot, One Step, Precision Cloning Method with High-Throughput Capability. *PLoS ONE* **3**, e3647; 10.1371/journal.pone.0003647 (2008).
260. Fu, J. *et al.* Full-length RecE enhances linear-linear homologous recombination and facilitates direct cloning for bioprospecting. *Nat. Biotechnol.* **30**, 440–446; 10.1038/nbt.2183 (2012).
261. Cobb, R. E. & Zhao, H. Direct cloning of large genomic sequences. *Nat. Biotechnol.* **30**, 405–406; 10.1038/nbt.2207 (2012).
262. Bian, X. *et al.* Direct cloning, genetic engineering, and heterologous expression of the syringolin biosynthetic gene cluster in *E. coli* through Red/ET recombineering. *ChemBioChem* **13**, 1946–1952; 10.1002/cbic.201200310 (2012).
263. Yin, J. *et al.* Direct cloning and heterologous expression of the salinomycin biosynthetic gene cluster from *Streptomyces albus* DSM41398 in *Streptomyces coelicolor* A3(2). *Sci. Rep.* **5**, 15081; 10.1038/srep15081 (2015).
264. Yamanaka, K. *et al.* Direct cloning and refactoring of a silent lipopeptide biosynthetic gene cluster yields the antibiotic taromycin A. *Proc. Natl. Acad. Sci. USA* **111**, 1957–1962; 10.1073/pnas.1319584111 (2014).
265. Kosuri, S. & Church, G. M. Large-scale de novo DNA synthesis: technologies and applications. *Nat. Methods* **11**, 499–507; 10.1038/nmeth.2918 (2014).
266. Shomar, H. *et al.* Metabolic engineering of a carbapenem antibiotic synthesis pathway in *Escherichia coli*. *Nat. Chem. Biol.* **14**, 794–800; 10.1038/s41589-018-0084-6 (2018).

267. Testolin, G. *et al.* Synthetic studies of cystobactamids as antibiotics and bacterial imaging carriers lead to compounds with high *in vivo* efficacy. *Chemical Science* **70**, 3; 10.1039/C9SC04769G (2020).
268. Trauner, D., Cheng, B. & Müller, R. Total syntheses of cystobactamids and structural confirmation of cystobactamid 919-2. *Angew. Chem. Int. Ed.* **56**, 7407–7410; 10.1002/anie.201705387 (2017).
269. Kirschning, A., Planke, T., Cirnski, K., Herrmann, J. & Müller, R. Synthetic and biologic studies on new urea and triazole containing cystobactamid derivatives. *Chemistry (Weinheim an der Bergstrasse, Germany)*; 10.1002/chem.201904073 (2019).
270. von Tesmar, A. Dissertation. Saarland University (2017).
271. Kim, Y. J. *et al.* Isolation of Coralmycins A and B, Potent Anti-Gram Negative Compounds from the Myxobacteria *Coralloccoccus coralloides* M23. *J. Nat. Prod.*; 10.1021/acs.jnatprod.6b00294 (2016).
272. Schnell, B. Thesis. Saarland University (2019).



## Chapter 2

### ***In vivo* and *in vitro* reconstitution of unique key steps in cystobactamid antibiotic biosynthesis**

#### **List of authors**

Sebastian Groß<sup>1,2,3</sup>, Bastien Schnell<sup>1,2,3</sup>, Patrick A. Haack<sup>1,2</sup>, David Auerbach<sup>1,2</sup>, Selina Deckarm<sup>1,2</sup>,  
Rolf Müller<sup>1-2\*</sup>

<sup>1</sup> Helmholtz Institute for Pharmaceutical Research Saarland (HIPS), Helmholtz Centre for Infection Research and Department of Pharmacy at Saarland University, Saarland University Campus, Building E8.1, 66123 Saarbrücken, Germany; <sup>2</sup> German Centre for Infection Research (DZIF), Partner Site Hannover-Braunschweig, Germany; <sup>3</sup> These authors contributed equally: S. Groß, B. Schnell.

(Manuscript under revision in *Nature communications*)

## **Contributions to the presented work**

Sebastian Groß designed, cloned and manipulated the modified BGC and performed heterologous expression experiments in *M. xanthus* DK1622. Bastien Schnell performed overexpression and purification of enzymes from *E. coli* BL21 and *in vitro* investigation of their activity. Patrick A. Haack and David Auerbach developed the direct intact protein UPLC-ESI-MS methods, performed measurements and analyzed data. Selina Deckarm was significantly involved in cloning experiments to obtain the targeted gene deletions in the modified BGC. Rolf Müller supervised the research. Sebastian Groß, Bastien Schnell and Rolf Müller wrote the manuscript.

## **Acknowledgement**

We thank Fred P. J. Haeckl, Joachim J. Hug and Domen Pogorevc for helpful proofreading of the manuscript and Sascha Baumann for fruitful discussions about *in vitro* experiments. We are especially grateful to Silke C. Wenzel for important contributions to the design of the heterologous expression constructs. This project was funded by German Center for Infection Research (DZIF).

## 2 *In vivo* and *in vitro* reconstitution of unique key steps in cystobactamid antibiotic biosynthesis

### 2.1 Abstract

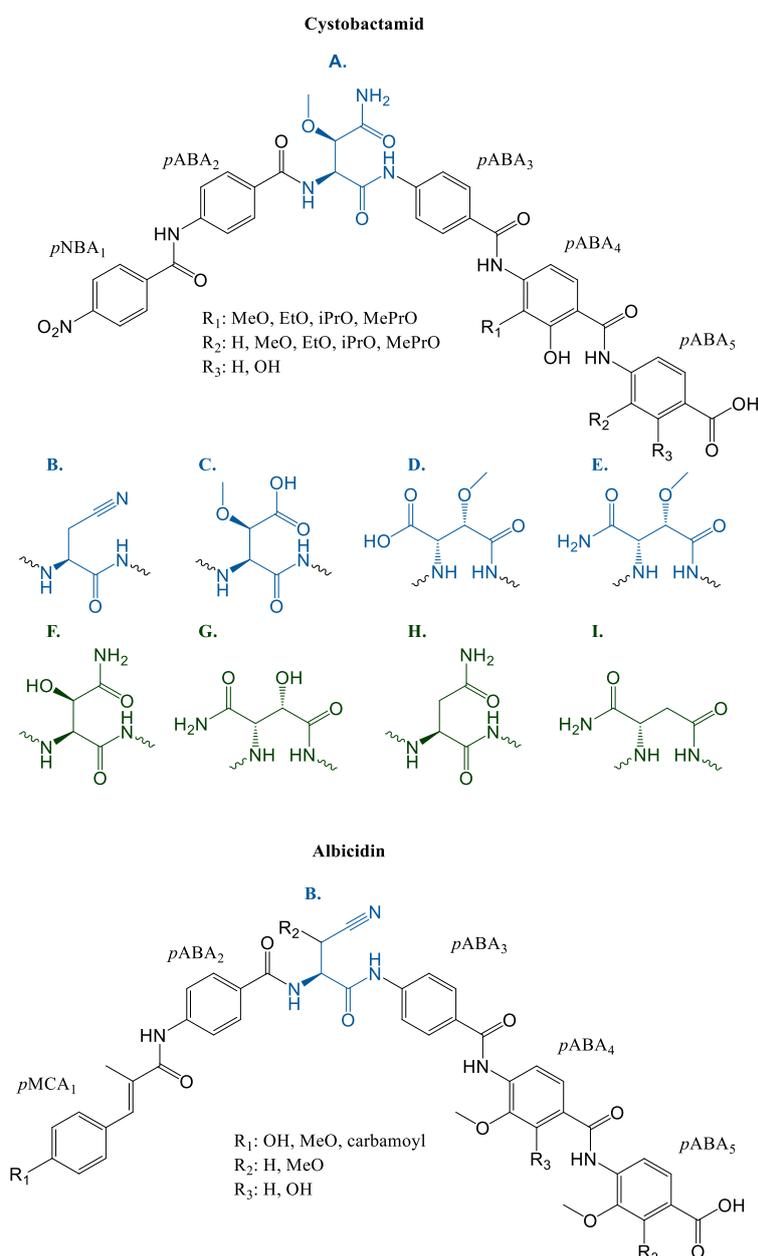
Cystobactamids are myxobacteria-derived topoisomerase inhibitors with potent anti-Gram-negative activity. They are formed by a non-ribosomal peptide synthetase (NRPS) and consist of tailored *para*-aminobenzoic acids, connected by an unprecedented  $\alpha$ -methoxy-L-isoasparagine or a  $\beta$ -methoxy-L-asparagine linker moiety. We describe the heterologous expression of the cystobactamid biosynthetic gene cluster (BGC) in *Myxococcus xanthus*. Targeted gene deletions produced several novel unnatural cystobactamids. Using *in vitro* experiments, we reconstituted the key biosynthetic steps of linker formation and shuttling via CysB to the NRPS. The biosynthetic logic involved a novel bifunctional domain found in the stand-alone NRPS module CysH, albicidin biosynthesis and numerous BGCs of unknown natural products. This domain performs either an aminomutase (AM) or an amide dehydratase (DH) type of reaction, depending on the activity of CysJ which hydroxylates CysH-bound L-asparagine. Furthermore, CysQ *O*-methylates hydroxyl-L-(iso)asparagine only in presence of the AMDH domain. Taken together, these findings provide direct evidence for unique steps in cystobactamid biosynthesis.

### 2.2 Introduction

The cystobactamids are a family of non-ribosomally synthesized peptide antibiotics produced by different myxobacteria such as *Cystobacter velatus* Cbv34 and *Myxococcus fulvus* SBMx122.<sup>1,2</sup> They target the bacterial topoisomerase IIA, but no cross-resistance was observed to clinically used gyrase inhibitors of the fluoroquinolone family, which share the same target.<sup>1</sup> The major derivatives in Cbv34 extracts are cystobactamid (Cys) 919-1, Cys919-2 and Cys507.<sup>1</sup> The prototypical structures (shown in Figure 1) feature one *para*-nitrobenzoic acid (*p*NBA), four *para*-aminobenzoic acids (*p*ABA) and an unusual L-isoasparagine or L-asparagine linker moiety (*p*NBA<sub>1</sub>-*p*ABA<sub>2</sub>-(iso)Asn-*p*ABA<sub>3</sub>-*p*ABA<sub>4</sub>-*p*ABA<sub>5</sub>). L-isoasparagine or L-asparagine are usually  $\alpha$ - or  $\beta$ -methoxylated, respectively. *p*ABA<sub>4</sub> and *p*ABA<sub>5</sub> are commonly isopropoxylated in position 2 and *p*ABA<sub>5</sub> is additionally hydroxylated in position 3.<sup>1</sup> Cys507, however, only consists of the three (tailored) C-terminal *p*ABA<sub>4-6</sub> moieties. In total, thirteen native cystobactamid derivatives were described, which differ in the structure of their linker moiety and the tailoring pattern of *p*ABA<sub>4</sub> and *p*ABA<sub>5</sub>.<sup>2</sup>

Interestingly, antibacterial activity is nearly limited to derivatives which belong to cystobactamid series 2 harboring an L-asparagine linker. Particularly interesting is the native derivative Cys861-2

showing low micromolar activity against *Acinetobacter baumannii*, *Pseudomonas aeruginosa*, *Escherichia coli* and other pathogens<sup>2</sup> that are classified with high- to critical-priority by the WHO.<sup>3</sup> Most of the cystobactamids with other linker moieties are inactive or show only weak antibacterial activity. Furthermore, total syntheses for several native cystobactamids and novel synthetic derivatives with improved antibacterial activity or metabolic stability were described.<sup>2,4-7</sup>



**Figure 1 | Structural variations among native and unnatural cystobactamids and structure comparison with albicidin.** *para*-Nitrobenzoic acid (*pNBA*) and *para*-aminobenzoic acid (*pABA*) with possible substitutions ( $R_1$ ,  $R_2$ ,  $R_3$ ) are shown in black. Different linker moieties of natural cystobactamids (shown in blue; linker A-E) and unnatural cystobactamids (shown in green; linker F-I; Table 1): A:  $\beta$ -methoxy-L-asparagine, B:  $\beta$ -cyano-L-alanine, C:  $\beta$ -methoxy-L-aspartate, D:  $\alpha$ -methoxy-L-isoaspartate, E:  $\alpha$ -methoxy-L-isoasparagine, F:  $\beta$ -hydroxy-L-asparagine, G:  $\alpha$ -hydroxy-L-isoasparagine, H: L-asparagine, I: L-isoasparagine. Scheme was adapted from Hüttel *et al.* Albicidin carries an N-terminal *para*-methylcoumaric acid (*pMCA*<sub>1</sub>), two *pABAs*, two substituted *pABAs* and a (possibly modified)  $\beta$ -cyano-L-alanine (B) or  $\beta$ -methoxy-L-asparagine (A) linker. Different possible substitutions ( $R_1$ ,  $R_2$ ,  $R_3$ ) are given.

Notably, cystobactamids show structural similarity with albicidins (shown in Figure 1), aPKS/NRPS product class isolated from *Xanthomonas albilineans*, which also shows antibacterial activity.<sup>8-10</sup> However, significant structural differences between both compound classes are found in the N-terminal parts and the linker moieties. A methylated *para*-coumaric acid moiety (*pMCA*<sub>1</sub>) typically forms the N-terminal part in albicidins, which can be further tailored e.g. with a carbamoyl group,<sup>10</sup> whereas native cystobactamids are restricted to *pNBA*<sub>1</sub>. The linker moieties arising in albicidins are  $\beta$ -cyano-L-alanine, which was also observed in Cys871,<sup>2</sup> and L-asparagine, both optionally methoxylated,<sup>10</sup> but no L-isoasparagine linker was described so far.

Non-ribosomal peptide synthetases (NRPSs) are large enzyme complexes with a multimodular architecture, in which each module is subdivided into independent domains, each usually catalyzing a single reaction. Adenylation (A) domains activate amino acids using ATP, thiolation (T) domains tether the activated amino acid or the growing peptide and condensation (C) domains catalyse peptide bond formation.<sup>13-15</sup> Tailoring of the product happens either after product release or on the assembly line. In the latter case, both *in-trans* tailoring by independent enzymes and *in-cis* modifications by tailoring domains, such as epimerization (E), heterocyclization (Cy) or methyltransferase (MT) domains, can occur.<sup>16</sup> NRPS biosynthesis is not limited to proteinogenic amino acids, thus allowing a great variability of chemical scaffolds,<sup>17</sup> such as shown for the daptomycins, a natural product featuring the unusual amino acid L-kynurenine.<sup>18</sup> Notably, NRPSs typically follow two important rules: First, the collinearity rule states that each module catalyzes the incorporation of a single building block into a growing peptide chain.<sup>19</sup> Second, the processivity rule states that the biosynthesis starts from the first module and proceeds sequentially to the next modules.

A model for the biosynthesis of cystobactamids was proposed by Baumann and coworkers based on *in silico* analyses of the BGC and feeding experiments with isotope-labelled amino acids.<sup>1</sup> In this model the modules 1, 2 and 4-6 on the NRPS enzymes CysK and CysG incorporate the two N-terminal and three C-terminal *pABA* units, *pABA*<sub>1-2</sub> and *pABA*<sub>4-6</sub>, respectively. However, the A<sub>3</sub> domain of module 3 in CysK was assumed inactive, because the core motif A10<sup>20</sup> lacks the catalytically essential lysine residue.<sup>21,22</sup> Therefore, the authors proposed that the T<sub>3</sub> domain in CysK is primed *in trans* by the stand-alone NRPS CysH with the help of the putative shuttling protein CysB. The stand-alone NRPS module CysH was proposed to activate L-asparagine, which is either used directly to prime T<sub>3</sub> or isomerized by an unusual ammonia/amine-ligase-like domain in CysH. Finally, Baumann and coworkers assumed that the linear hexapeptide is released from the assembly line by the TE of CysG being further modified by various tailoring enzymes afterwards. By comparison, in the  $\beta$ -cyano-L-alanine linker biosynthesis in albicidin, Cociancich and colleagues proposed activation of L-asparagine by the stand-alone NRPS module 2\* (AlbIV) and

phosphorylation of the side chain amide oxygen *in cis* by a domain harboring an ATP-binding motif. They postulated that subsequent dephosphorylation would lead to formal elimination of water and formation of  $\beta$ -cyano-L-alanine (shown in Supplementary Figure 9b). However, none of the reaction steps were experimentally proven in either of the previous publications.

Native myxobacterial producing strains are often difficult to cultivate and genetically manipulate, making the study of their natural products biosynthesis challenging.<sup>11</sup> Heterologous expression of biosynthetic gene clusters (BGCs) can circumvent these issues, but identification of proper host strains and subsequent cloning of the complex clusters remain significant bottlenecks. Nevertheless, a number of myxobacterial BGCs have been heterologously expressed in *Myxococcus xanthus* DK1622.<sup>11,12</sup> In addition to heterologous expression systems, overexpression of individual proteins and their biochemical analysis *in vitro* can be used to gain further insights into the biosynthesis of natural products.

Herein, we describe the design, assembly and heterologous expression of a modified cystobactamid BGC in *M. xanthus* DK1622. We identified 13 new natural cystobactamids upon expression of all biosynthetic genes and 5 new unnatural major derivatives after targeted gene deletions. Targeted gene deletions in combination with *in vitro* investigation of the enzyme activities allowed us to explain the unique biosynthesis steps of the  $\alpha$ -methoxy-L-isoasparagine linker and its shuttling to the assembly line. This building block is synthesized by the independent NRPS module CysH and the bifunctional *in cis* tailoring aminomutase/amide dehydratase (AMDH) domain, working in tandem with the oxygenase CysJ and the *O*-methyltransferase CysQ. Finally, this moiety is transferred onto module 3 of CysK by the shuttling protein CysB. Furthermore, we confirmed that the biosynthesis of the *N*-terminally truncated derivative Cys507 starts from the middle of the assembly line, bending the processivity rule. With these results we were able to decipher most of the obscure and unique steps of cystobactamid linker biosynthesis. We provide a heterologous production platform for topoisomerase inhibitors and discovered an unprecedented plasticity of NRPS biosynthesis.

## 2.3 Results

### 2.3.1 Design, assembly, heterologous expression and manipulation of the cystobactamid gene cluster in *Myxococcus xanthus* DK1622

A modified BGC together with a cloning and expression vector system were designed *in silico* for the heterologous production of cystobactamids in *M. xanthus* DK1622 (described in Supplementary Information; Supplementary Figures 1-3 and Supplementary Tables 4-8). The revised template sequence of the BGC originated from *C. velatus* Cbv34, including the 25 biosynthetic genes *cysA-T*

and *Orf1-5* as previously described.<sup>1</sup> The modified BGC was chemically synthesized in fragments, because of the size, GC content and repetitive sequence segments in the cluster. Assembly of the cluster fragments was done by combination of *in vivo* transformation-associated recombination (TAR) cloning in yeast<sup>23</sup> and a previously described *in vitro* three-step restriction/ligation cloning strategy<sup>24</sup> using *BsaI* (cloning steps are summarized in Supplementary Figures 4-5 and Supplementary Table 9). The final expression construct pMYC20Cys\_v2 was integrated into the *M. xanthus* DK1622 genome via the *Mx8* phage integrase. UPLC-HRMS analysis and MS<sup>2</sup> experiments confirmed the heterologous production of 13 new and 9 known cystobactamids (Table 1; see Supplementary Information; Supplementary Figures 6-8 and 14; Supplementary Table 10).

**Table 1 | Natural and unnatural cystobactamids heterologously produced by *M. xanthus* DK1622.** Linker and R<sub>1</sub>, R<sub>2</sub>, R<sub>3</sub> labelling was adapted from Figure 1. pMYC20Cys\_v2 includes all genes from the cystobactamid BGC described by Baumann *et al.* Heterologously produced derivatives that were described previously are marked in dark grey. Known derivatives that were not identified in the heterologous producer are marked in light grey. Derivatives in bold are major products in native producer strains and the heterologous producer. For deletion constructs only major products, which are relevant for the elucidation of the linker biosynthesis, are shown. <sup>[a]</sup> Cys889-1b and Cys889-2b (reported as Coralmycin D)<sup>25</sup> carry an N-terminal amine rather than nitro-group.

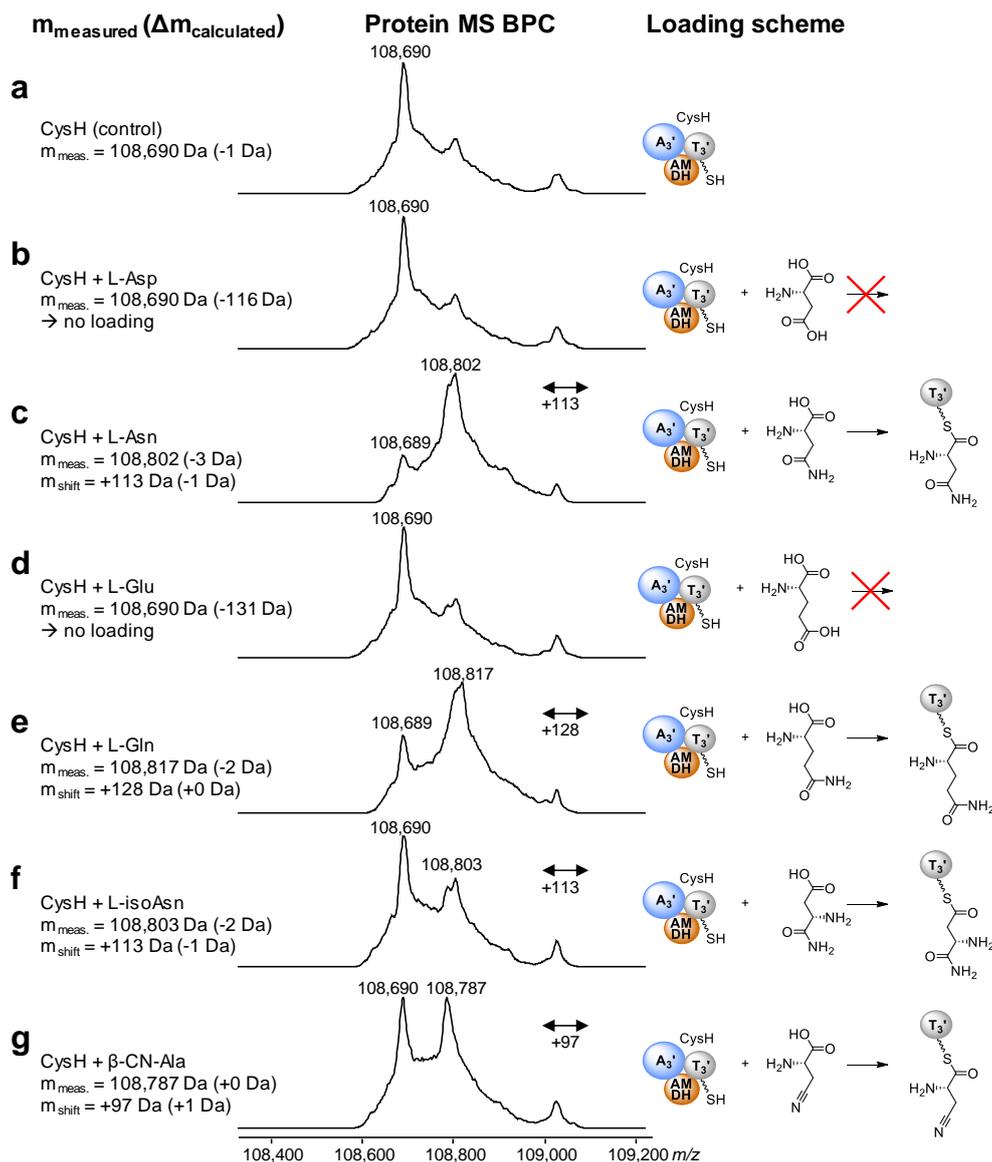
Natural cystobactamids						Unnatural cystobactamids						
Construct	Product	Linker	R <sub>1</sub>	R <sub>2</sub>	R <sub>3</sub>	Construct	Product	Linker	R <sub>1</sub>	R <sub>2</sub>	R <sub>3</sub>	
pMYC20 Cys_v2	Cys449	-	<i>iPrO</i>	H	H	pMYC20Cys_v2	ΔAMDH	Cys905-2c	F	<i>iPrO</i>	<i>iPrO</i>	H
	Cys507	-	<i>iPrO</i>	<i>iPrO</i>	H	pMYC20Cys_v4	Cys905-1c	G	<i>iPrO</i>	<i>iPrO</i>	H	
	Cys861-1	E	<i>iPrO</i>	H	H	Δ <i>cysQ</i>	Cys905-2c	F	<i>iPrO</i>	<i>iPrO</i>	H	
	Cys861-2	A	<i>iPrO</i>	H	H	pMYC20Cys_v4	Cys889-1a	I	<i>iPrO</i>	<i>iPrO</i>	H	
	Cys871	B	<i>iPrO</i>	<i>iPrO</i>	H	Δ <i>cysJ</i>	Cys889-2a	H	<i>iPrO</i>	<i>iPrO</i>	H	
	Cys877-1	E	EtO	MeO	H	Cys871	B	<i>iPrO</i>	<i>iPrO</i>	H		
	Cys877-2	A	EtO	MeO	H	pMYC20Cys_v4	Cys889-2a	H	<i>iPrO</i>	<i>iPrO</i>	H	
	Cys891-1a	E	EtO	EtO	H	Δ <i>cysJ</i> ΔAMDH	Cys889-2a	H	<i>iPrO</i>	<i>iPrO</i>	H	
	Cys891-1b	E	<i>iPrO</i>	MeO	H	pMYC20Cys_v4	Cys507	-	<i>iPrO</i>	<i>iPrO</i>	H	
	Cys891-2a	A	EtO	EtO	H	Δ <i>cysB</i>	Cys889-1b <sup>[a]</sup>	E	<i>iPrO</i>	<i>iPrO</i>	H	
	Cys891-2	A	<i>iPrO</i>	MeO	H	pMYC20Cys_v4	Cys889-2b <sup>[a]</sup>	A	<i>iPrO</i>	<i>iPrO</i>	H	
	Cys905-1a	E	<i>iPrO</i>	EtO	H							
	Cys905-1b	E	EtO	<i>iPrO</i>	H							
	Cys905-2a	A	EtO	<i>iPrO</i>	H							
	Cys905-2	A	<i>iPrO</i>	EtO	H							
	<b>Cys919-1</b>	<b>E</b>	<b><i>iPrO</i></b>	<b><i>iPrO</i></b>	<b>H</b>							
	<b>Cys919-2</b>	<b>A</b>	<b><i>iPrO</i></b>	<b><i>iPrO</i></b>	<b>H</b>							
	Cys920-1	C	<i>iPrO</i>	<i>iPrO</i>	H							
	Cys920-2	D	<i>iPrO</i>	<i>iPrO</i>	H							
	Cys933-1a	E	<i>iPrO</i>	MePrO	H							
	Cys933-1b	E	MePrO	<i>iPrO</i>	H							
	Cys933-2a	A	<i>iPrO</i>	MePrO	H							
	Cys933-2b	A	MePrO	<i>iPrO</i>	H							
	Cys934-2	D	<i>iPrO</i>	MePrO	H							
	Cys935-1	E	<i>iPrO</i>	<i>iPrO</i>	OH							
	Cys935-2	A	<i>iPrO</i>	<i>iPrO</i>	OH							

Notably, these 13 new derivatives were also produced in native producer strains under the same cultivation conditions. The production titer of the major product Cys919-1 was 8.1 mg·L<sup>-1</sup> in the heterologous producer as compared to 3.6 mg·L<sup>-1</sup> in *Myxococcus fulvus* SBMx122. Formerly mentioned production yields in native producer strains were much lower with 60–100 µg compound isolated per 1·L culture.<sup>1,2</sup> However, difficulties in upscaling and compound loss during purification resulted in yields of isolated products similar to those originally reported. Subsequently, Red/ET recombineering in combination with *Bsa*I restriction and ligation was used to generate scarless gene deletions of *cysQ*, *cysJ*, the AMDH domain of *cysH*, *cysB* and *cysR* independently. Using this strategy, we identified five new major unnatural cystobactamid derivatives (plus one recently reported as Coralmycin D<sup>25</sup> and four minor derivatives) by heterologous expression of the manipulated constructs (Table 1; see Supplementary Information; Supplementary Figures 6-7 and 15; Supplementary Table 10).

### 2.3.2 Biosynthesis of the linker moiety

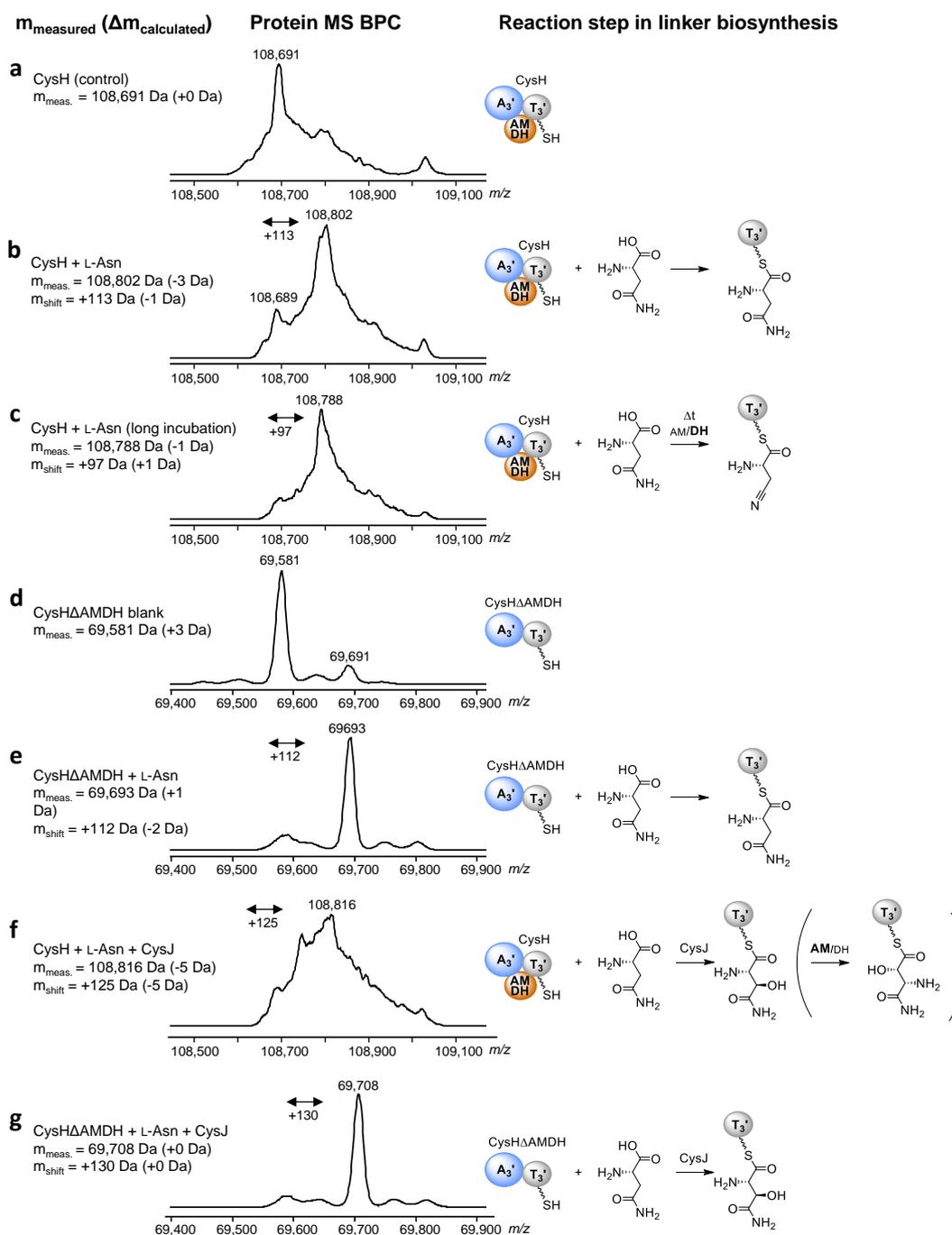
β-cyano-L-alanine linkers were found both in albicidin and in minor cystobactamid derivatives (Figure 1).<sup>2,9</sup> Interestingly, we found high structural similarity between the unknown domain of the single-standing NRPS AlbIV, which was hypothesized to catalyze dehydration of L-asparagine,<sup>9</sup> with the 38 kDa domain found inserted in the stand-alone NRPS CysH (69% identity / 83% similarity), which is involved in cystobactamid linker biosynthesis (Supplementary Figure 9a).<sup>1</sup> Despite high structural similarity between these two unusual domains, completely different reaction mechanisms were proposed (Supplementary Figure 9b and c). The major cystobactamid derivative harbors a modified L-isoasparagine linker and Baumann *et al.* proposed that the unusual domain catalyzes the isomerization of CysH-bound L-asparagine to L-isoasparagine.<sup>1</sup> Feeding of <sup>15</sup>N<sub>2</sub>-<sup>13</sup>C<sub>4</sub>-labelled L-asparagine during fermentation of native producer *C. velatus* Cbv34 confirmed full conservation of all carbons and nitrogens from L-asparagine in L-isoasparagine<sup>1</sup> indicating an aminomutase-type reaction. Since cystobactamids contain β-cyano-L-alanine or L-isoasparagine linkers, we hypothesize that the unusual domain in CysH catalyzes either aminomutation (AM) or dehydration (DH) of L-asparagine. Hence, we named the domain AMDH.

We overexpressed and purified the enzymes CysH, CysH without AMDH domain (CysHΔAMDH) and CysJ from *E. coli* BL21. The enzymes were incubated *in vitro* individually or in combination using different substrates. Loading of the substrate onto T domain and subsequent biochemical conversion resulted in mass shifts observed after deconvolution of direct intact protein ESI-MS spectra.<sup>26</sup> First, CysH was incubated with different amino acids to test substrate specificity. Although L-asparagine was favored as substrate by CysH, we also observed loading of L-glutamine, β-cyano-L-alanine and L-isoasparagine (Figure 2).



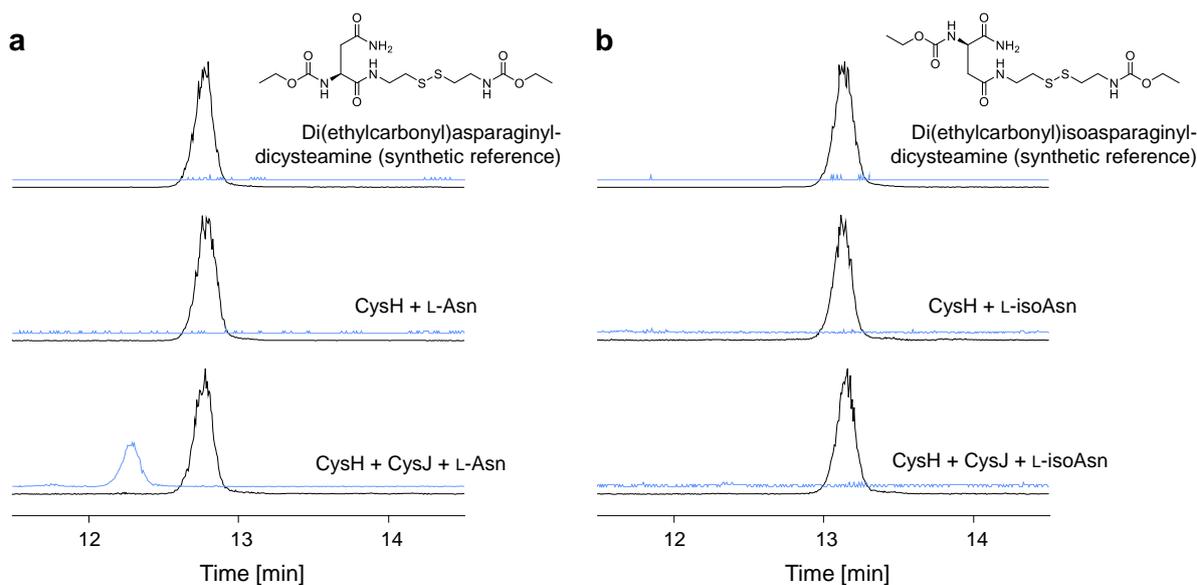
**Figure 2 | Substrate specificity of CysH.** Observed mass shifts in deconvoluted protein MS BPCs reveal loading of L-asparagine (c), L-glutamine (e), L-isoasparagine (f) and β-cyano-L-alanine (g) onto CysH. L-asparagic acid (b) and L-glutamic acid (d) were not accepted by CysH (control: a).

All naturally occurring cystobactamids (except of the linker-free derivatives Cys449 and Cys507), which were identified thus far, harbor linker moieties deriving from L-asparagine. Since we also observed acceptance of other substrates by CysH, we assume that substrate specificities of downstream modules in the assembly line hinder incorporation of different amino acids than L-asparagine. Interestingly, incubation of CysH with L-asparagine for longer than five minutes at room temperature led to a mass increase of +96 *m/z* instead of +114 *m/z* indicating substrate dehydration (-18 *m/z*) and formation of β-cyano-L-alanine (Figure 3a-c). Incubation of CysHΔAMDH with L-asparagine only resulted in substrate loading but not in dehydration since only the mass shift expected for L-asparagine was observed even after prolonged incubation (Figure 3d-e). This experiment was a first indication for the dehydratase activity of the AMDH domain.



**Figure 3 | Loading and hydroxylation of L-asparagine on CysH or CysH $\Delta$ AMDH.** Deconvoluted protein MS BPCs (left) reveal different reaction mechanisms (models shown on the right) after *in vitro* incubation of CysH or CysH $\Delta$ AMDH with L-asparagine with and/or without CysJ. **a:** CysH control. **b:** Loading of L-asparagine onto CysH. **c:** The dehydratase activity of the AMDH domain leads to dehydration of L-asparagine and formation of  $\beta$ -cyano-L-alanine after prolonged incubation times. **d:** CysH $\Delta$ AMDH control. **e:** CysH $\Delta$ AMDH incubated with L-asparagine. No dehydration of L-asparagine was observed since CysH has no AMDH domain. **f:** CysH incubated with L-asparagine and CysJ leads to loading of the substrate onto CysH with subsequent hydroxylation by CysJ (see Figure 4). The isomerization of ( $\beta$ -hydroxy)-L-asparagine to ( $\alpha$  hydroxy)-L-isoasparagine is shown in parenthesis, because this step cannot be observed by MS. The isomerization was confirmed by deletion of the AMDH domain from the BGC and analysis of the production profile after heterologous expression of the respective construct in *M. xanthus* DK1622 (see Figure 5). **g:** CysH $\Delta$ AMDH incubated with L-asparagine and CysJ leads only to the formation of  $\beta$ -hydroxy-L-asparagine.

Next, we investigated the  $\beta$ -hydroxylation of L-asparagine, which was speculated to be catalyzed by CysJ on T domain-bound substrate in an  $\alpha$ -ketoglutarate ( $\alpha$ -KG)-dependent reaction.<sup>1</sup> Incubation of CysJ with free L-asparagine and  $\alpha$ -KG with subsequent analysis using thin layer chromatography (TLC) indicated no hydroxylation of free L-asparagine (Supplementary Figure 16). Since CysJ also revealed high structural similarity to SyrP,<sup>27</sup> which has been shown to catalyze  $\beta$ -hydroxylation of T domain-bound aspartyl residues in syringomycin biosynthesis, we expected an *in trans* tailoring step of CysJ in coordination with CysH. However, we observed significant peak broadening during protein MS upon incubation of CysH with L-asparagine and CysJ (Figure 3f), preventing clear indication of the  $\beta$ -hydroxylation. Interestingly, incubation of CysH $\Delta$ AMDH with L-asparagine and CysJ could be analyzed without peak broadening and  $\beta$ -hydroxylation of L-asparagine could be confirmed (Figure 3g). To prove  $\beta$ -hydroxylation of L-asparagine in presence of the AMDH domain, we incubated CysH with CysJ and L-asparagine and subsequently unloaded the carrier protein bound intermediate via trans-thioesterification using cysteamine,<sup>28</sup> which was analyzed using HPLC-MS after further derivatization (Figure 4 and Supplementary Figure 17).

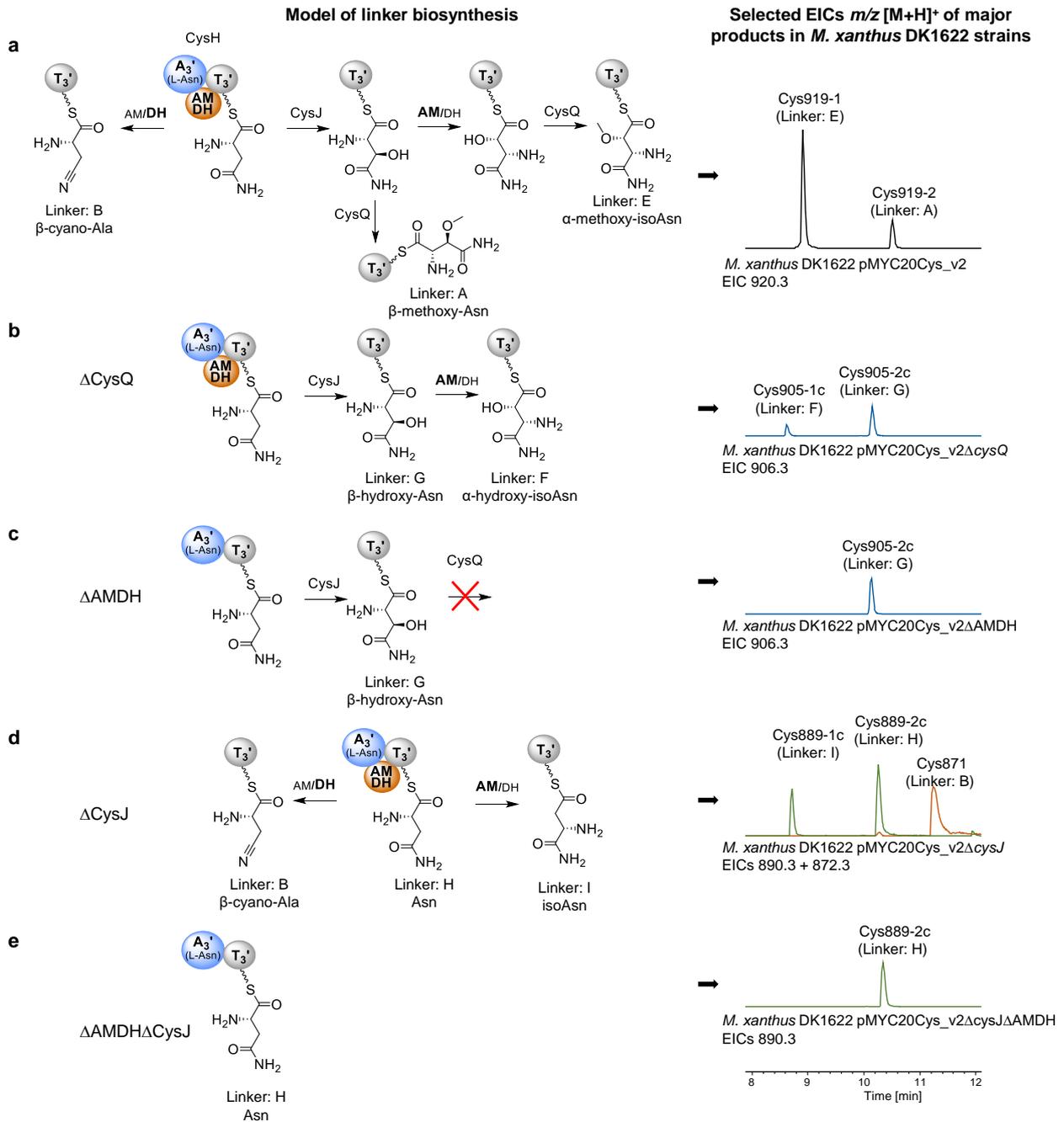


**Figure 4 | Hydroxylation of CysH-bound L-asparagine by CysJ.** HPLC-MS analysis of cysteamine-unloaded and derivatized substrate from the CysH protein. EICs 411.1  $m/z$   $[M+H]^+$  are shown in black and EICs 427.1  $m/z$   $[M+H]^+$  (Di(ethylcarbonyl)-hydroxy-L-(iso)asparaginyl-dicysteamine) are shown in blue. **a:** Di(ethylcarbonyl)-L-asparaginyl-dicysteamine synthetic reference. CysH-bound L-asparagine unloaded and derivatized showed same retention time as the synthetic reference. Incubation of CysH with CysJ and L-asparagine lead to hydroxylation of L-asparagine and different retention time of the unloaded substrate compared to the reference. **b:** Di(ethylcarbonyl)-L-isoasparaginyl-dicysteamine synthetic reference. CysH-bound L-isoasparagine unloaded and derivatized showed same retention time as the synthetic reference. No hydroxylation occurred upon addition of CysJ.

We observed a different mass and retention time of the unloaded substrate in the presence of CysJ which confirms the *in trans*  $\beta$ -hydroxylation of CysH-bound L-asparagine. Surprisingly, we could

not observe  $\alpha$ -hydroxylation of CysH-bound L-isoasparagine by CysJ. Consequently, the  $\beta$ -hydroxylation of L-asparagine occurs prior to the expected aminomutase reaction. However, in this set of *in vitro* experiments we were unable to detect the  $\alpha,\beta$ -aminomutase activity of the AMDH domain, because the expected isomerization of L-asparagine cannot be observed by MS.

We thus performed numerous targeted gene and domain deletion experiments with subsequent heterologous expression of the modified BGC in *M. xanthus* DK1622 (Figure 5 and Supplementary Figure 15). Most importantly, the deletion of the AMDH domain in *cysH* resulted in the abolishment of the production of all cystobactamids with L-isoasparagine or  $\beta$ -cyano-L-alanine linkers in the heterologous producer. This result can be taken as the first experimental proof confirming the AMDH domain asparaginyl  $\alpha,\beta$ -aminomutase activity. Surprisingly, the major product of this construct was the unnatural derivative Cys905-2c carrying a  $\beta$ -hydroxy-L-asparagine linker instead of the expected  $\beta$ -methoxy-L-asparagine (Figure 5c). The production of the unnatural Cys905-1c and Cys905-2c derivatives, both lacking *O*-methylation in the linkers, was also achieved through deletion of the gene *cysQ* encoding an *O*-methyltransferase (Figure 5b). We speculate that the abolishment of *O*-methylation in absence of the AMDH domain in CysH is linked to protein-protein interaction between AMDH and CysQ. These results allowed us devise a biosynthesis model for the production of the native  $\beta$ -methoxy-L-asparagine (linker A), the  $\alpha$ -methoxy-L-isoasparagine (linker E) and the  $\beta$ -cyano-L-alanine (linker B) moieties, as shown in Figure 5a. In the presence of CysH (including the AMDH domain), CysJ and CysQ, the major products were Cys919-1 and Cys919-2 harboring linkers A and E, respectively. However, Cys871 with linker B was not detected in the cultivation broth of the heterologous producer, but previously described in native producer strains.<sup>2</sup> As shown in Figure 3b-c, dehydration of L-asparagine by the AMDH domain mainly occurs in absence of CysJ. We assume that the activity of CysJ in the heterologous producer prevented the AMDH domain from dehydrating L-asparagine. To test this hypothesis, we deleted *cysJ* in the modified BGC, which indeed lead to heterologous production of substantial amounts of Cys871 (Figure 5d) confirming our previous assumption. Consequently, we hypothesize that, in presence of CysJ,  $\beta$ -hydroxylation of L-asparagine occurs much faster than dehydration of L-asparagine by the AMDH domain. Furthermore, two novel unnatural major derivatives, Cys889-1a and Cys889-2a were produced (Figure 5d). Cys889-1a and Cys889-2a lack the methoxy group in the linker, thus only having either L-isoasparagine (linker I) or L-asparagine (linker H), because neither hydroxylation by CysJ nor *O*-methylation by CysQ can occur. Interestingly, isomerization of L-asparagine still occurs, but now leads to a much less abundant product. We speculate that the isomerization of  $\beta$ -hydroxyl-L-asparagine by the AMDH domain is more efficient than the isomerization of L-asparagine, or that CysH, CysJ and CysQ form a protein complex influencing the reactivity of the AMDH domain.



**Figure 5 | Summary of biosynthesis pathways for cystobactamid linkers.** **a:** Hydroxylation of CysH-bound L-asparagine by CysJ with subsequent isomerization by AMDH plus *O*-methylation by CysQ leading to  $\alpha$ -methoxy-L-isoasparagine (E linker); *O*-methylation without isomerization leads to  $\beta$ -methoxy-L-asparagine (A linker); direct dehydration by AMDH domain leads to the formation of  $\beta$ -cyano-L-alanine (B linker). EIC 920.3  $[M+H]^+$  (black) shows production of Cys919-1 (E linker) and Cys919-2 (A linker) in *M. xanthus* DK1622 pMYC20Cys\_v2. **b:** Absence/deletion of CysQ leads to formation of  $\beta$ -hydroxy-L-asparagine (G linker) or  $\alpha$ -hydroxy-L-isoasparagine (F linker) after isomerization by AMDH. EIC 906.3  $[M+H]^+$  (blue) shows production of Cys905-1c (F linker) and Cys905-2c (G linker), which lack one methyl group in the linker compared to Cys919 (-14 Da shift). **c:** Deletion of the AMDH domain leads to hydroxylation of CysH-bound L-asparagine by CysJ, but *O*-methylation by CysQ does not occur (only formation of  $\beta$ -hydroxy-L-asparagine/production of Cys905-2c). **d:** Absence/deletion of CysJ leads to formation of L-asparagine (H linker), L-isoasparagine (I linker) or  $\beta$ -cyano-L-alanine (B linker). Overlay of EIC 890.3  $[M+H]^+$  (green) and EIC 872.3  $[M+H]^+$  (orange) shows production of Cys889-1a (I linker), Cys889-2a (H linker) and Cys871 (B linker) **e:** Deletion of CysJ and the AMDH domain prevents any modification of asparagine leading only to the production of Cys889-2a.

If CysJ is deleted, the isomerization reaction may occur much slower, thus leading to a decreased L-isoasparagine/L-asparagine linker ratio and an increased probability of L-asparagine dehydration. Finally, we generated an expression construct with a double deletion of both *cysJ* and the AMDH domain. As expected, only Cys889-2a featuring a simple L-asparagine linker was produced, whereas neither Cys889-1a nor Cys871 could be detected (Figure 5e). This experiment again proves that the AMDH domain catalyzes either dehydration or aminomutation of L-asparagine. The type of reaction catalyzed by the AMDH domain depends on preceding hydroxylation of the substrate by CysJ. Notably, we also identified four minor cystobactamid derivatives, two after deletion of the AMDH domain and two after deletion of *cysJ*. An expanded explanation about those derivatives, for which we were not able to provide reliable structures based on MS<sup>2</sup> experiments, is given in the Supplementary Information (Supplementary Figure 15).

Including all results of the *in vitro* and *in vivo* experiments, we are able to provide biosynthesis schemes for the unprecedented  $\alpha$ -methoxy-L-isoasparagine linker and also other linker derivatives, which are summarized in Figure 5. Although CysH loads a variety of amino acids, L-asparagine is the favored substrate, which is either directly dehydrated by AMDH to form  $\beta$ -cyano-L-alanine or  $\beta$ -hydroxylated by CysJ with subsequent isomerization by AMDH. CysQ performs the *O*-methylation of  $\beta$ -hydroxy-L-asparagine or  $\alpha$ -hydroxy-L-isoasparagine leading to formation of  $\beta$ -methoxy-L-asparagine or  $\alpha$ -methoxy-L-isoasparagine (Figure 5a). Furthermore, we speculated that the product ratio of cystobactamid derivatives with different linkers is highly dependent on the reaction kinetics of the enzymes involved in linker biosynthesis. In absence of CysJ substantially higher production of cystobactamid with dehydrated linker was observed (Figure 5d). Furthermore, heterologous expression of constructs with deleted AMDH domain resulted in elimination of production of cystobactamids with L-isoasparagine linkers, which only showed weak or no antibacterial activity.<sup>2</sup>

Our findings suggest that AMDH is an unprecedented domain resulting in both amide dehydration and aminomutation via an unknown mechanism. Structure prediction was precluded by unsuccessful crystallization attempts. Additionally, no known templates with a crystal structure were available, preventing *in silico* 3D modelling. A protein BLAST query of the AMDH domain identified numerous homologs, all inserted into A domains with L-asparagine-specificity based on Stachelhaus prediction.<sup>20</sup> Since none of these homologs could be linked to a known secondary metabolite, we speculate that the dehydration or isomerization of L-asparagine is a common mechanism in the biosynthesis of hitherto unknown natural products. We analyzed the 25 closest BLAST homologues of the AMDH domain *in silico* and identified seven conserved core motif regions (Supplementary Figure 10) that might be required for catalysis or folding. Core region 1 contains a highly conserved ATP-binding motif (SGGKD), which was also found in the homologous domain in AlbIV

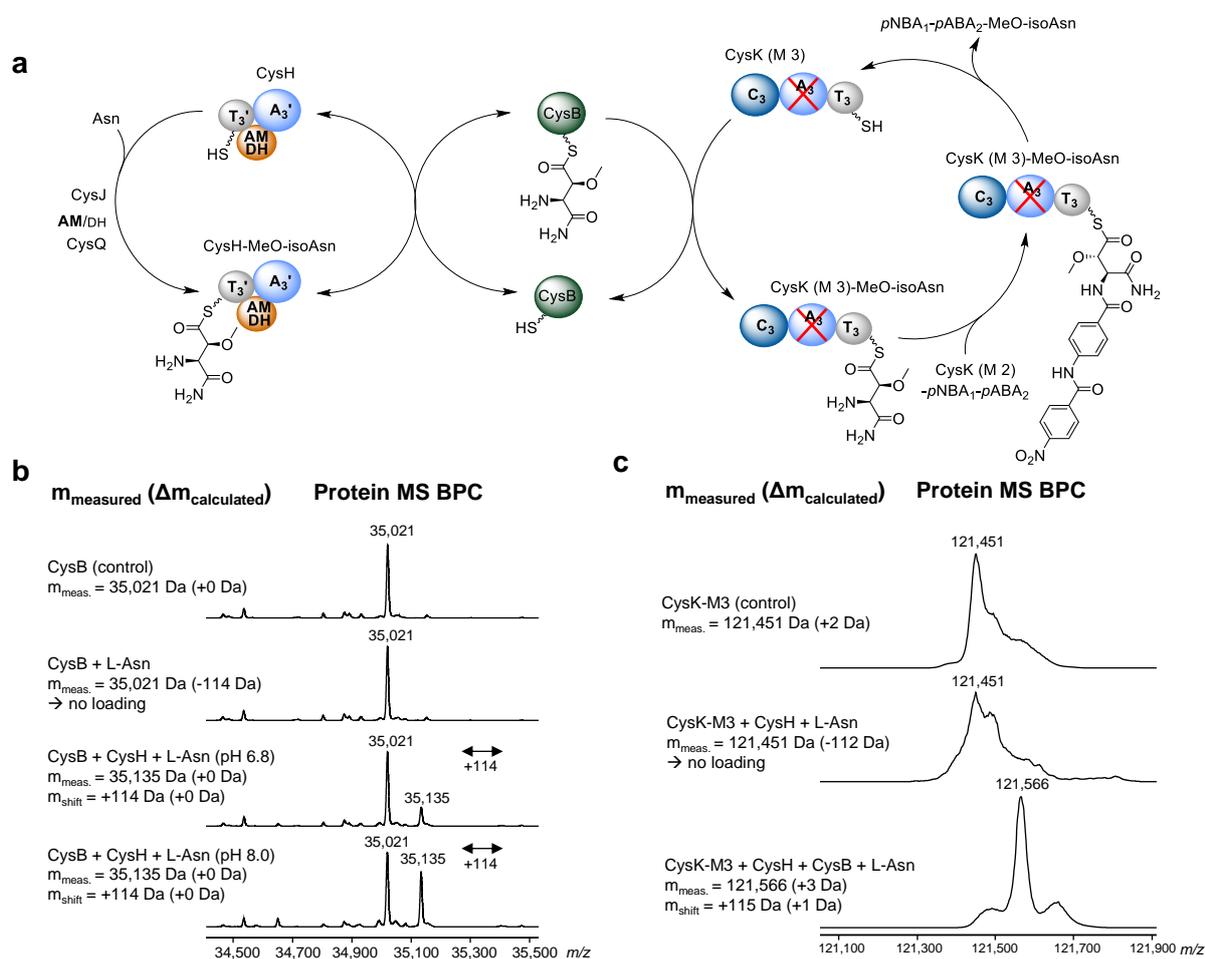
(Supplementary Figure 9a) and hypothesized to be involved in L-asparagine dehydration.<sup>9</sup> To investigate if the AMDH domain shares any homology with known aminomutases, we searched for conserved sequence motifs such as an ASG motif described in tyrosine and phenylalanine aminomutases that contain the cofactor 4-methylideneimidazole-5-one.<sup>29,30</sup> No such motif was identified, preventing comparison to this class of aminomutases. Interestingly, we observed that CysH shows a dark brown color after overexpression and purification from *E. coli*, indicating the presence of a metal as cofactor. Since we did not identify a CxxCxxxC motif serving as Fe-S cluster binding site,<sup>31</sup> we assume that a radical mode of action is unlikely for the AMDH domain, even if a few radical-SAM proteins were described not harboring this motif.<sup>32</sup> Although the enzymatic mechanism of amide dehydration to nitriles is not yet known, the reverse reaction catalyzed by a heterodimeric nitrile hydratase has been described. Interestingly, this reaction relies on a single, cysteine-bound iron atom.<sup>33</sup> Summarized, we consider a radical mode of action or a similar one to known aminomutases unlikely. We speculate that the AMDH domain involves a new mode of catalysis, which may require a metal cofactor or might be similar to the ATP-dependent reaction proposed for albicidin formation and certainly deserves future investigation.

### 2.3.3 Shuttling of the linker moiety to the assembly line by CysB

Another intriguing feature of cystobactamid biosynthesis is the incorporation of the linker moiety into the NRPS assembly line. Interestingly, the A domain of module 3 in CysK was proposed to be inactive, because the catalytic lysine residue in core motif A10 is missing.<sup>1</sup> We performed direct intact protein MS analysis to confirm experimentally that L-asparagine is not accepted by module 3 (Figure 6c), which was separately overexpressed and purified, because of the large size of CysK (507 kDa). Initially the overexpression yields of CysK-M3 were very low, but we overcame this problem by coexpressing CysA, which is an MbtH type A domain activator protein supposed to be required for expression and activity of NRPS modules.<sup>34</sup> Incubation of CysK-M3 with L-asparagine and subsequent full protein MS analysis confirmed that no substrate loading occurred. Thus, we hypothesize that loading of the respective T<sub>3</sub> domain requires an *in trans* shuttling process between CysH and CysK via a third enzyme.

BLAST analysis of the genes in the cystobactamid BGC showed similarity of CysB to SyrC (35 % similarity / 20 % identity), which is a (chloro)threonyl aminoacyl transferase in syringomycin biosynthesis.<sup>35</sup> Another CysB homolog, CmaE, was shown to shuttle aminoacyl groups between carrier protein domains in coronamic acid biosynthesis.<sup>36</sup> With CysB being our candidate enzyme for the hypothesized shuttling process, we overexpressed and purified CysB, CysH and CysK-M3 from *E. coli* BL21. First, we analyzed the aminoacyl transfer reaction of L-asparagine from CysH to CysB using protein MS. We used commercially available L-asparagine instead of methoxylated

L-(iso)asparagine derivatives since the deletion of *cysJ* from the BGC with subsequent heterologous expression in *M. xanthus* DK1622 showed that L-asparagine is also accepted by the assembly line (Figure 5d).



**Figure 6 | CysB-mediated transfer of  $\beta$ -methoxy-L-isoasparagine from CysH to CysK.** **a:** Model for the CysB-mediated shuttling process. L-asparagine is activated by CysH and modified by CysJ, the AMDH domain and CysQ as shown in Figure 5.  $\beta$ -methoxy-L-isoasparagine is transferred from CysH to module 3 of CysK (M3) by CysB. Condensation of the linker moiety with the pNBA<sub>1</sub>-pABA<sub>2</sub> dipeptide leads to the formation of the shown tripeptide. **b:** Deconvoluted protein MS analysis of CysB control; CysB incubated with free L-asparagine does not result in CysB loading; CysB incubated with L-asparagine and CysH leads to pH-sensitive loading of CysB (+114 m/z shift). **c:** Protein MS analysis verifies the transfer of L-asparagine from CysB to CysK (M3) in the presence of CysH.

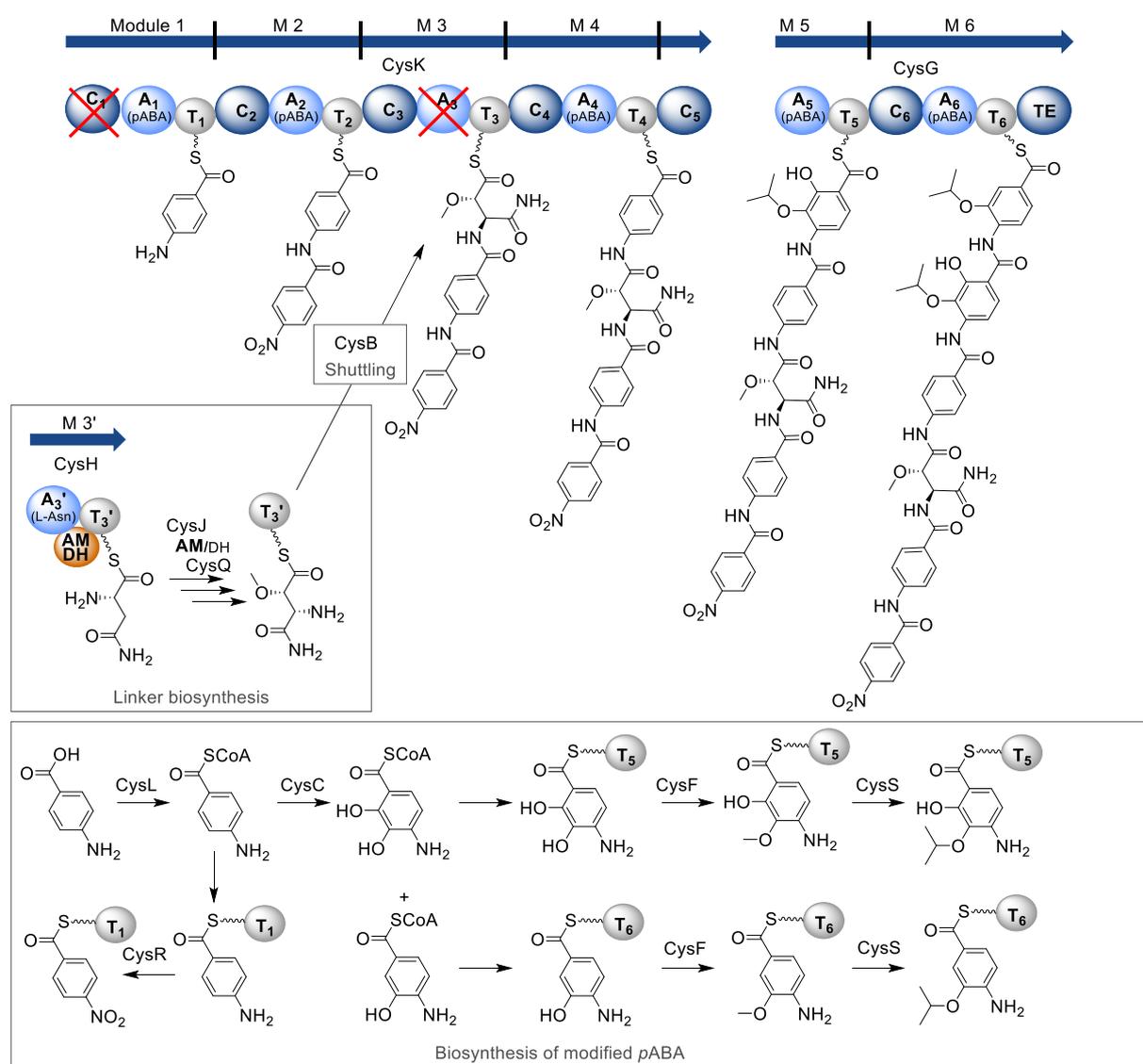
Protein MS analysis showed that CysB was only partially loaded with L-asparagine in the presence of CysH, because the major peak still derived from unloaded CysB, whereas no free L-asparagine was loaded (Figure 6b). To exclude that the partial loading is caused by inappropriate reaction conditions, we tested different pH values. Even though the equilibrium between L-asparagine-loaded CysH and CysB shifted pH-dependently, we still observed partial loading of CysB. Thus, we conclude that the reaction is reversible. Next, we analyzed the transfer of CysB-loaded L-asparagine to CysK-M3. Interestingly, loading of L-asparagine from CysB to module 3 of CysK was almost

stoichiometric (Figure 6c) implying that this second part of the shuttling process probably drives the cycle towards the transfer from CysH to CysK-M3. With those experiments we confirmed the hypothesis that CysB mediates the shuttling process of the linker moiety between CysH and CysK-M3 and provide a respective model in Figure 6a.

### 2.3.4 Revision of the complete cystobactamid biosynthesis model

Based on our new findings regarding the linker biosynthesis and transfer and incorporation into the assembly line, we provide a revised biosynthesis model on the example of Cys919-1 (Figure 7). The biosynthesis of the cystobactamid peptide scaffold starts with CysK. Separate overexpression and purification of modules 1, 2 and 4 (CysK) with subsequent *in vitro* loading experiments revealed in the protein MS analyses that various *pABA* derivatives are accepted (Supplementary Figure 11).

Interestingly, module 1 does not accept *pNBA* as substrate, which means that the oxygenation of *pABA*<sub>1</sub> is performed *in trans* or after final product release. Deletion of *cysR* and subsequent heterologous expression of the respective construct in *M. xanthus* DK1622 lead to the production of derivatives with an N-terminal amine, thus proving that CysR is the *pABA-N*-oxygenase (Supplementary Figure 12). CysL is assumed to be a *pABA*-CoA ligase which activates free *pABA*.<sup>1</sup> We assume that the oxidation of CoA-bound *pABA* is performed by CysC forming 3-hydroxy-*pABA* and 2,3-dihydroxy-*pABA* prior to incorporation into the assembly line by modules 5 and 6 (CysG), respectively. CysC is homologous to the benzoate oxidase BoxB, which is described as dioxygenase requiring a CoA-activated substrate.<sup>37</sup> Additionally, deletion of *cysC* and heterologous expression of the construct in *M. xanthus* DK1622 lead to complete abolishment of cystobactamid production. This underlines the importance of *pABA*<sub>5</sub> and *pABA*<sub>6</sub> hydroxylation prior to their activation by A<sub>5</sub> and A<sub>6</sub> from CysG. We separately overexpressed and purified modules 5 and 6 from *E. coli* BL21 and analyzed their substrate specificity *in vitro*. Likewise for modules 1,2 and 4, protein MS analysis revealed loading of various *pABA* derivatives by modules 5 and 6, respectively (Supplementary Figure 13). CysF is a SAM-dependent methyl transferase assumed to be involved in the formation of 2-hydroxy-3-methoxy-*pABA* on module 5 and 3-methoxy-*pABA* on module 6. The final tailoring steps of *pABA*<sub>5</sub> and *pABA*<sub>6</sub> are iterative methyl group alkylations leading to various branched alkoxy groups. Those reactions are performed by CysS, a cobalamin-dependent radical-SAM enzyme.<sup>38</sup>



**Figure 7 | Revised biosynthesis of cystobactamids on the example of Cys919-1.** C: condensation domain (dark blue); A: adenylation domain (light blue); T: thiolation domain (grey); TE: thioesterase domain (dark blue); AMDH: aminomutase dehydratase domain (orange); red cross: inactive domains. Biosynthesis of the  $\alpha$ -methoxy-L-isoasparagine linker moiety is described in more detail in Figure 5. CysR converts *pABA* to *pNBA in-trans* or after product release from the assembly line. Biosynthesis of 2-isopropoxyl-*pABA* and 2-isopropoxyl-3-hydroxy-*pABA* is presumably catalyzed by CysC, CysF and CysS. *pABA* is incorporated by M 1 and M 2, respectively. The linker moiety is transferred from T<sub>3</sub>' (CysH) to T<sub>3</sub> (CysK) by CysB (see Figure 6). Another *pABA* is incorporated by M 4. Two tailored *pABAs* are incorporated by M 5 and M 6.

We observed another special feature of the cystobactamid biosynthesis when we deleted *cysB* from the BGC and subsequently expressed the modified construct in *M. xanthus* DK1622. Surprisingly, Cys507, a tripeptide consisting only of the three C-terminal (tailored) *pABAs*, was still produced as the only derivative (Supplementary Figure 18). This not only proves CysB being indispensable for the biosynthesis of full-length cystobactamids, but also that Cys507 is not a degradation product as initially thought.<sup>1</sup> Instead, the biosynthesis can also start from module 4 on, which is contradictory to the processivity rule in NRPSs.

## 2.4 Discussion

Our results present novel features in NRPS synthesis demonstrating that the AMDH domain performs both dehydration and aminomutation of L-asparagine in a single domain. The bifunctionality and ability to perform completely different biochemical reactions dependent on preceding tailoring steps enables the production of a variety of different compounds with a yet unknown mechanism. We excluded a radical mode of action and refuted similar mechanisms to known tyrosine- and phenylalanine aminomutases.<sup>29,30</sup> Further biochemical characterization and crystallization experiments are necessary to elucidate the underlying mechanisms of the AMDH domain in the future. Interestingly, we found numerous unannotated homologous domains in several BGCs of unknown function, indicating that this AMDH domain is involved in the biosyntheses of a significant number of yet to be identified natural products. Consequently, the AMDH domain can be queried to identify additional rarely observed modified L-asparagine-containing natural products and their derivatives.

Interestingly, no albicidin derivative harboring an L-isoasparagine linker has been identified so far, indicating that the AMDH domain homolog in the albicidin biosynthesis is not able to perform an aminomutation-type reaction. Furthermore, von Eckardstein and coworkers described an albicidin derivative with a methoxylated  $\beta$ -cyano-L-alanine linker, a linker-type not found in cystobactamids. Based on our experiments, we hypothesized that  $\beta$ -hydroxylation of L-asparagine occurs much faster than the dehydration and that the  $\beta$ -hydroxylation prevents the AMDH domain from dehydrating the substrate. This explains why we only identified a considerable amount of Cys871 harboring a  $\beta$ -cyano-L-alanine linker after deletion of the hydroxylase CysJ. However, the existence of the methoxylated  $\beta$ -cyano-L-alanine linker in albicidin either means that AlbVIII (the homolog of CysJ) is able to hydroxylate  $\beta$ -cyano-L-alanine or that the AMDH domain homolog in AlbIV is able to dehydrate  $\beta$ -hydroxy-L-asparagine (both of which was not observed in the cystobactamid biosynthesis) or that another enzyme than AlbVIII catalyzes the hydroxylation of  $\beta$ -cyano-L-alanine. Since Cys871 is only a very minor derivative in the presence of CysJ, one could also speculate that CysJ, likewise to AlbVIII, is able to hydroxylate  $\beta$ -cyano-L-alanine but the generated cystobactamid derivatives with methoxylated  $\beta$ -cyano-L-alanine linkers are only produced in such minor amounts that the ion intensity in MS does not exceed the detection limit. In this case the inactivation of the aminomutation function of the AMDH domain in CysH might potentially lead to an increased production of Cys871 and the detection of methoxylated  $\beta$ -cyano-L-alanine linkers in cystobactamids. An alternative experiment would be the exchange of the AMDH domain in CysH by its homolog from AlbIV, which could potentially lead to a more pronounced dehydration reaction and subsequent hydroxylation by CysJ. In any case it needs further experimental investigation to

understand the order and under which circumstances the respective linker modifications take place in the cystobactamid and albicidin biosyntheses.

The modification of amino acids by independent NRPS modules and subsequent incorporation into nascent polypeptides in the assembly line is not an entirely new phenomenon in natural product biosynthesis as similar findings have been reported for novobiocin, nikkomycin and vancomycin.<sup>39–41</sup> In those examples, a TE releases the modified amino acid from the T domain of the independent module on which the modification reaction occurs. The free modified amino acid is then tethered to the core peptide by A domain reactivation or by a specific ligase. However, in cystobactamid biosynthesis, transport of the modified L-asparagine between the independent module CysH and the assembly line is mediated by CysB in a hitherto unprecedented shuttling mechanism. The differences in the kinetics of the first (CysH to CysB) and second (CysB to CysK-M3) part of the reaction are highly pH-dependent, thus indicating that the reaction might be driven by pI differences between the T domains of CysH (pI = 5.6) and CysK-M3 (pI = 6.3). A similar correlation was also proposed for the CysB homologue CmaE.<sup>36</sup> Notably, we demonstrated that the assembly line is able to start the biosynthesis from module 4 on, skipping the first three modules, when we deleted *cysB* and heterologously expressed the modified construct in *M. xanthus* DK1622. This leads to an interruption of the shuttling process and the production of the N-terminally truncated, linker-free cystobactamid derivative Cys507. Consequently, the cystobactamid biosynthesis shows exceptions for two common NRPS rules, the collinearity and the processivity rule, which underlines the diversity in the functionality of NRPS systems. The production of Cys507 even in the presence of the shuttling protein CysB shows that the transfer of the linker moiety to the assembly line is a bottleneck for the production of full-length cystobactamids.

Finally, we demonstrated that the heterologous expression platform can be used to produce a variety of new cystobactamids by genetic engineering of the BGC. Despite serious efforts, we were not able to isolate the novel natural derivatives, because they are present in much lower concentrations compared to the major product Cys919-1. Furthermore, the cystobactamid production decreased substantially when we scaled up the cultivation from 50 mL to 1.5 L, which may also explain the low formerly reported yields. This drop in production was even worse for new unnatural cystobactamids, for which even the production of the major products was a fraction compared to Cys919-1. Therefore, we relied on exact HRMS data and MS<sup>2</sup> fragmentation in this study. We assigned the stereocenters of the new derivatives based on the stereochemistry of previously described cystobactamids. Furthermore, we assigned the linker moieties with same masses (e.g. linker A and E) based on different retention times that were also observed for previously reported cystobactamids, in which derivatives harboring L-isoasparagine linkers always eluted first. However, the establishment of a robust fermentation process combined with media optimization has

to be addressed in future experiments to enable purification and NMR verification of the novel derivatives from this study. Moreover, the deletion of the AMDH domain may be used in the future to drive the heterologous production profile towards cystobactamids with L-asparagine rather than L-isoasparagine linkers. Notably, cystobactamids with L-asparagine linker showed superior antibacterial activity against numerous human pathogens like *A. baumannii*, *Citrobacter freundii*, *E. coli*, *Enterobacter cloacae*, *P. aeruginosa*, *Proteus vulgaris*, *Bacillus subtilis*, *Staphylococcus aureus* and *Streptococcus pneumoniae*.<sup>1,2</sup> The question arises why Nature established such a complex biosynthesis route including a *trans*-acting independent NRPS module with a shuttling process to produce cystobactamids which are biologically less active? It was previously shown<sup>1,2</sup> and confirmed in this study that naturally a whole cocktail of cystobactamids is produced. Even though cystobactamids with L-asparagine linkers exhibited superior antibacterial activity against a small panel of tested human pathogens, the natural producer strains have to outcompete a myriad of rival strains in their natural environment. It thus appears likely that the diversity of cystobactamids produced helps the natural producers to gain advantage over a variety of their competitors. Furthermore, it cannot be excluded that cystobactamids possess another function apart from their antibacterial activity, e.g. the involvement in developmental processes of the cell. However, from a human point of view, the simplest solution to produce more active cystobactamids with medicinal relevance harboring L-asparagine linkers would be the existence of an active L-asparagine-specific CysK-A<sub>3</sub> domain. Restoring the activity of the natively inactive A<sub>3</sub> domain by genetic engineering of the assembly line and thus bypassing the production-limiting shuttling process will be addressed in future experiments.

## 2.5 References

1. Baumann, S. *et al.* Cystobactamids: myxobacterial topoisomerase inhibitors exhibiting potent antibacterial activity. *Angew. Chem. Int. Ed.* **53**, 14605–14609; 10.1002/anie.201409964 (2014).
2. Hüttel, S. *et al.* Discovery and Total Synthesis of Natural Cystobactamid Derivatives with Superior Activity against Gram-Negative Pathogens. *Angew. Chem. Int. Ed. Engl.* **56**, 12760–12764; 10.1002/anie.201705913 (2017).
3. Tacconelli, E. *et al.* Discovery, research, and development of new antibiotics: the WHO priority list of antibiotic-resistant bacteria and tuberculosis. *Lancet Infect. Dis.* **18**, 318–327; 10.1016/S1473-3099(17)30753-3 (2018).

- Moreno, M. *et al.* Synthesis and biological evaluation of cystobactamid 507: a bacterial topoisomerase inhibitor from *Cystobacter* sp. *Synlett* **26**, 1175–1178; 10.1055/s-0034-1380509 (2015).
- Moeller, M. *et al.* Scalable Syntheses of Methoxyaspartate and Preparation of the Antibiotic Cystobactamid 861-2 and Highly Potent Derivatives. *Org. Lett.* **21**, 8369–8372; 10.1021/acs.orglett.9b03143 (2019).
- Kirschning, A., Planke, T., Cirnski, K., Herrmann, J. & Müller, R. Synthetic and biologic studies on new urea and triazole containing cystobactamid derivatives. *Chem. Eur. J.* **26**, 4289–4296; 10.1002/chem.201904073 (2020).
- Elgaher, W. A. M. *et al.* Cystobactamid 507: Concise Synthesis, Mode of Action and Optimization toward More Potent Antibiotics. *Chem. Eur. J.* **26**, 7219–7225; 10.1002/chem.202000117 (2020).
- Birch, R. G. & Patil, S. S. Preliminary characterization of an antibiotic produced by *Xanthomonas albilineans* which inhibits DNA synthesis in *Escherichia coli*. *J. Gen. Microbiol.* **131**, 1069–1075; 10.1099/00221287-131-5-1069 (1985).
- Cociancich, S. *et al.* The gyrase inhibitor albicidin consists of *p*-aminobenzoic acids and cyanoalanine. *Nat. Chem. Biol.* **11**, 195–197; 10.1038/nchembio.1734 (2015).
- Eckardstein, L. von *et al.* Total Synthesis and Biological Assessment of Novel Albicidins Discovered by Mass Spectrometric Networking. *Chemistry* **23**, 15316–15321; 10.1002/chem.201704074 (2017).
- 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**; 10.1016/B978-0-12-409547-2.14818-8 (2020).
- Herrmann, J., Fayad, A. A. & Müller, R. Natural products from myxobacteria: novel metabolites and bioactivities. *Nat. Prod. Rep.* **34**, 135–160; 10.1039/C6NP00106H (2017).
- Marahiel, M. A. A structural model for multimodular NRPS assembly lines. *Nat. Prod. Rep.* **33**, 136–140; 10.1039/c5np00082c (2016).
- Finking, R. & Marahiel, M. A. Biosynthesis of nonribosomal peptides. *Annu. Rev. Microbiol.* **58**, 453–488; 10.1146/annurev.micro.58.030603.123615 (2004).

15. Süßmuth, R. D. & Mainz, A. Nonribosomal peptide synthesis - Principles and prospects. *Angew. Chem. Int. Ed.* **56**, 3770–3821; 10.1002/anie.201609079 (2017).
16. 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).
17. 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; 10.1002/anie.201208344 (2013).
18. Miao, V. *et al.* Daptomycin biosynthesis in *Streptomyces roseosporus*: Cloning and analysis of the gene cluster and revision of peptide stereochemistry. *Microbiol.* **151**, 1507–1523; 10.1099/mic.0.27757-0 (2005).
19. Stachelhaus, T. & Marahiel, M. A. Modular structure of genes encoding multifunctional peptide synthetases required for non-ribosomal peptide synthesis. *FEMS Microbiol. Lett.* **125**, 3–14 (1995).
20. Stachelhaus, T., Mootz, H. D. & Marahiel, M. A. The specificity-conferring code of adenylation domains in nonribosomal peptide synthetases. *Chem. Biol.* **6**, 493–505; 10.1016/S1074-5521(99)80082-9 (1999).
21. 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).
22. Branchini, B. R., Murtiashaw, M. H., Magyar, R. A. & Anderson, S. M. The role of lysine 529, a conserved residue of the acyl-adenylate-forming enzyme superfamily, in firefly luciferase. *Biochemistry* **39**, 5433–5440; 10.1021/bi9928804 (2000).
23. Kouprina, N. & Larionov, V. Transformation-associated recombination (TAR) cloning for genomics studies and synthetic biology. *Chromosoma* **125**, 621–632; 10.1007/s00412-016-0588-3 (2016).
24. Yan, F. *et al.* Synthetic biology approaches and combinatorial biosynthesis towards heterologous lipopeptide production. *Chem. Sci.* **9**, 7510–7519; 10.1039/c8sc02046a (2018).
25. Kim, B.-M., van Minh, N., Choi, H.-Y. & Kim, W.-G. Coralmycin Derivatives with Potent Anti-Gram Negative Activity Produced by the Myxobacteria *Coralloccoccus coralloides* M23. *Molecules* **24**, 1390; 10.3390/molecules24071390 (2019).

26. Tesmar, A. von *et al.* Total Biosynthesis of the Pyrrolo4,2benzodiazepine Scaffold Tomaymycin on an In Vitro Reconstituted NRPS System. *Cell Chem. Biol.* **24**, 1216-1227.e8; 10.1016/j.chembiol.2017.08.001 (2017).
27. Singh, G. M., Fortin, P. D., Koglin, A. & Walsh, C. T. beta-Hydroxylation of the aspartyl residue in the phytotoxin syringomycin E: characterization of two candidate hydroxylases AspH and SyrP in *Pseudomonas syringae*. *Biochemistry* **47**, 11310–11320; 10.1021/bi801322z (2008).
28. Belecki, K. & Townsend, C. A. Biochemical determination of enzyme-bound metabolites: preferential accumulation of a programmed octaketide on the enediyne polyketide synthase CalE8. *J. Am. Chem. Soc.* **135**, 14339–14348; 10.1021/ja406697t (2013).
29. Wu, B., Szymanski, W., Heberling, M. M., Feringa, B. L. & Janssen, D. B. Aminomutases: mechanistic diversity, biotechnological applications and future perspectives. *Trends Biotechnol.* **29**, 352–362; 10.1016/j.tibtech.2011.02.005 (2011).
30. Krug, D. & Müller, R. Discovery of additional members of the tyrosine aminomutase enzyme family and the mutational analysis of CmdF. *ChemBioChem* **10**, 741–750; 10.1002/cbic.200800748 (2009).
31. Frey, P. A., Hegeman, A. D. & Ruzicka, F. J. The Radical SAM Superfamily. *Crit. Rev. Biochem. Mol. Biol.* **43**, 63–88 (2008).
32. Parent, A. *et al.* The B12-Radical SAM Enzyme PoyC Catalyzes Valine C $\beta$ -Methylation during Polytheonamide Biosynthesis. *J. Am. Chem. Soc.* **138**, 15515–15518; 10.1021/jacs.6b06697 (2016).
33. Huang, W. *et al.* Crystal structure of nitrile hydratase reveals a novel iron centre in a novel fold. *Structure* **5**, 691–699; 10.1016/S0969-2126(97)00223-2 (1997).
34. Herbst, D. A., Boll, B., Zocher, G., Stehle, T. & Heide, L. Structural basis of the interaction of mbth-like proteins, putative regulators of nonribosomal peptide biosynthesis, with adenylating enzymes. *J. Biol. Chem.* **288**, 1991–2003; 10.1007/s10295-011-1022-8 (2013).
35. Singh, G. M., Vaillancourt, F. H., Yin, J. & Walsh, C. T. Characterization of SyrC, an aminoacyltransferase shuttling threonyl and chlorothreonyl residues in the syringomycin biosynthetic assembly line. *Chem. Biol.* **14**, 31–40; 10.1016/j.chembiol.2006.11.005 (2007).

36. Strieter, E. R., Vaillancourt, F. H. & Walsh, C. T. CmaE: A transferase shuttling aminoacyl groups between carrier protein domains in the coronamic acid biosynthetic pathway. *Biochemistry* **46**, 7549–7557; 10.1021/bi700243h (2007).
37. Gescher, J., Eisenreich, W., Wörth, J., Bacher, A. & Fuchs, G. Aerobic benzoyl-CoA catabolic pathway in *Azoarcus evansii*: studies on the non-oxygenolytic ring cleavage enzyme. *Mol. Microbiol.* **56**, 1586–1600; 10.1111/j.1365-2958.2005.04637.x (2005).
38. Wang, Y., Schnell, B., Baumann, S., Müller, R. & Begley, T. P. Biosynthesis of branched alkoxy groups: iterative methyl group alkylation by a Cobalamin-Dependent Radical SAM Enzyme. *J. Am. Chem. Soc.* **139**, 1742–1745; 10.1021/jacs.6b10901 (2016).
39. Hubbard, B. K. & Walsh, C. T. Vancomycin assembly: nature's way. *Angew. Chem. Int. Ed. Engl.* **42**, 730–765 (2003).
40. Chen, H. W. & Walsh, C. T. Coumarin formation in novobiocin biosynthesis: beta-hydroxylation of the aminoacyl enzyme tyrosyl-S-NovH by a cytochrome P450 NovI. *Chem. Biol.* **8**, 301–312 (2001).
41. Chen, H., Hubbard, B. K., O'Connor, S. E. & Walsh, C. T. Formation of beta-hydroxy histidine in the biosynthesis of nikkomycin antibiotics. *Chem. Biol.* **9**, 103–112 (2002).

## 2.6 Methods

### 2.6.1 Cultivation of strains

*E. coli* DH10 $\beta$ , HS996 and NEB10 $\beta$  strains were used for cloning of the modified BGC. *E. coli* BL21 (DE3) was used for recombinant protein expression. Cultivation was performed in LB medium (10 g L<sup>-1</sup> tryptone, 5 g L<sup>-1</sup> NaCl, 5 g L<sup>-1</sup> yeast extract, pH 7.6) at 37 °C or 30 °C (handling plasmids larger than 15 kb). Protein expression experiments were carried out at 37 °C and 16 °C after induction (see protein expression section). Ampicillin (100  $\mu$ g mL<sup>-1</sup>), chloramphenicol (34  $\mu$ g mL<sup>-1</sup>), kanamycin (50  $\mu$ g mL<sup>-1</sup>) and oxytetracyclin (10  $\mu$ g mL<sup>-1</sup>) were used as selection markers. *Myxococcus xanthus* DK1622 was used as heterologous expression host. *Cystobacter velatus* Cbv34 and *Myxococcus fulvus* SBMx122 are native cystobactamid producer strains and were used for the isolation of genomic DNA or controls in production screening experiments, respectively. Cultivation was done in CTT medium (10 g L<sup>-1</sup> casitone, 1.21 g L<sup>-1</sup> TRIS, 8 mM MgSO<sub>4</sub>, 1 mM KH<sub>2</sub>PO<sub>4</sub>, pH 7.6) to grow cells for genomic DNA isolation, transformations or for starting cultures prior to production screening cultivations. M7/s4 medium (5 g L<sup>-1</sup> soy flour, 5 g L<sup>-1</sup> corn starch, 2 g L<sup>-1</sup> glucose, 1 g L<sup>-1</sup> yeast extract, 1 g L<sup>-1</sup> MgSO<sub>4</sub> × 7H<sub>2</sub>O, 1 g L<sup>-1</sup> CaCl<sub>2</sub> × 2H<sub>2</sub>O, 10 g L<sup>-1</sup> HEPES, pH 7.4; supplemented with 0.1 mg L<sup>-1</sup> of vitamin B<sub>12</sub> and 5 mg L<sup>-1</sup> of FeCl<sub>3</sub> after autoclaving) was used for 50 mL screening cultures. M7/s4 pre cultures (without supplements) were inoculated from CTT agar starting cultures. Screening cultures were inoculated from 1-3 d old M7/s4 pre cultures (10 % (v/v) inoculation volume) and cultivated for 5 d. Heterologous gene expression in *M. xanthus* was induced after 1 d by adding vanillate (1 mM final concentration). XAD16 absorber resin was added after 2 d (2 % (v/v)). All liquid cultivations were performed in baffled Erlenmeyer flasks on an orbital shaker at 160 rpm at 30 °C. Kanamycin (50  $\mu$ g mL<sup>-1</sup>) and oxytetracycline (10  $\mu$ g mL<sup>-1</sup>) were used as selection markers when cultivating heterologous *M. xanthus* strains. *S. cerevisiae* ATCC4004247 was used for TAR cloning. Cultivations were performed at 30 °C in YPAD medium (20 g L<sup>-1</sup> glucose, 10 g L<sup>-1</sup> peptone, 10 g L<sup>-1</sup> yeast extract, 100 mg L<sup>-1</sup> adenine-hemisulfate, pH 7.0). YNB medium (20 g L<sup>-1</sup> glucose, 8 g L<sup>-1</sup> YNB base w/o leucine, 2 g L<sup>-1</sup> amino acid mix w/o leucine, 100 mg L<sup>-1</sup> adenine-hemisulfate, pH 7.0) was used for selection of transformants.

### 2.6.2 *In silico* experiments and revision of the native BGC sequence

*Geneious* v10.1.3 (Biomatters Ltd.) was used to analyze the native cystobactamid BGC sequence of Cbv34 (GenBank: KP836244) and to design the modified BGC and the cloning and expression vector system pMYC *in silico*. Repetitive sequence segments in *cysK* were analyzed using dotplot (*EMBOSS* 6.5.7 tool *dottup*: <http://emboss.sourceforge.net/>). Resequencing of *cysK* was performed using Sanger sequencing. The revised BGC sequence is accessible under GenBank accession

number: will be supplied upon acceptance of the manuscript. All *in silico* experiments and the revision of the Cbv34 BGC sequence are described in detail in Supplementary Information.

### 2.6.3 DNA synthesis and BGC and pMYC assembly

The modified BGC and the cloning and expression vector system pMYC were synthesized in fragments (fragment description listed in Supplementary Table 6). DNA synthesis was carried out by ATG: biosynthetics GmbH. Sequence-verified DNA synthesis fragments were delivered in pGH standard vector harboring an *ampR (bla)* gene for selection on ampicillin. Restriction endonuclease hydrolysis, DNA ligation, *E. coli* transformation and plasmid DNA isolation were done according to standard protocols.<sup>42</sup> pMYC20 and pMYC21 were generated from the DNA fragments pMYC and Mx8-tetR or Mx9-kanR by ligation after hydrolysis with *PacI* and *XmaII*, respectively. Both operons of the modified BGC (CysOp1 and CysOp2) were assembled in two separate TAR cloning reactions using pMYC20 or pMYC21, respectively. TAR cloning was performed according to standard protocols.<sup>43</sup> Some DNA fragments were assembled *in vitro* to obtain larger fragments before TAR assembly. *cysK* was replaced by a dummy sequence in order to minimize the risk of unspecific recombination during TAR assembly. The fragments of *cysK* harboring repetitive sequence segments were assembled separately *in vitro* by a three-step restriction/ligation cloning strategy using *BsaI* (Supplementary Figure 4) prior to its assembly with the rest of the cluster. Supplementary Figure 5 schematically depicts all cloning steps performed to obtain the final expression construct pMYC20Cys\_v2. Supplementary Table 9 summarizes all *in vitro* cloning steps performed in this work.

### 2.6.4 Genetic manipulation of expression constructs

Red/ET recombineering<sup>44</sup> in combination with restriction hydrolysis and re-ligation was used to delete (part of) genes from the plasmids. Amplification of *ampR (bla)* gene from pUC18 was done via PCR. Apart from pUC18 binding site, primers contained *BsaI* R-sites and 50 bp sequences that are homologous to the gene, which was deleted. Supplementary Table 3 lists all primers used for this experiment. After Red/ET recombineering, we selected clones harboring the correct recombination products on ampicillin and oxytetracycline. After the plasmid isolation, we verified the clones by restriction analysis. Next, the recombination product was hydrolysed with *BsaI* and re-ligated to remove *ampR (bla)* from the construct. Clones, which lost their resistance towards ampicillin, were selected for plasmid isolation and restriction analysis. Supplementary Table 11 lists all manipulated plasmids that were generated in this study. Supplementary Tables 1 and 2 list all strains and plasmids generated during the cloning process, respectively.

### 2.6.5 Transformation of *M. xanthus* DK1622 and verification by colony PCR

Expression constructs were transformed into *M. xanthus* DK1622 via electroporation using established standard protocols.<sup>45</sup> Integration of the constructs into the chromosome occurred by site-specific phage recombination in Mx8 attachment site. Integration was verified by colony PCR using different combinations of primers Mx8-attB-up2, Mx8-attB-down, Mx8-attP-up2 and Mx8-attP-down (SupplementaryTable 3). DNA preparation, suitable primer combinations, reaction conditions and PCR product sizes are described by Pogorevc *et al.*<sup>46</sup> Supplementary Table 1 lists all expression strains generated in this work.

### 2.6.6 Sample preparation and UPLC-ESI-HRMS analysis

Cells and XAD16 absorber resin of 50 mL screening cultures were harvested by centrifugation at  $3,200 \times g$  for 15 min at 4 °C. Extraction was done 2x60 min with 30 mL methanol under stirring at room temperature (RT). Extracts were filtered using folded filter paper (8 – 12  $\mu\text{m}$  pore size) and dried using a rotary evaporator. Dried extract was dissolved in 3 mL methanol and analyzed using UPLC-HRMS. An UltiMate 3000 LC System (Dionex) with a Acquity UPLC BEH C-18 column (1.7  $\mu\text{m}$ , 100 x 2 mm; Waters), equipped with a VanGuard BEH C-18 (1.7  $\mu\text{m}$ ; Waters) guard column, was coupled to an Apollo II ESI source (Bruker) and hyphenated to maXis 4G ToF mass spectrometer (Bruker). Separation was performed at a flow rate of 0.6 mL min<sup>-1</sup> (eluent A: deionized water + 0.1 % formic acid (FA), eluent B: acetonitrile + 0.1 % FA) at 45 °C using the following gradient: 5 % B for 30 s, followed by a linear gradient up to 95 % B in 18 min and a constant percentage of 95 % B for further 2 min. Original conditions were adjusted with 5 % B within 30 s and kept constant for 1.5 min. The LC flow was split to 75  $\mu\text{L}$  min<sup>-1</sup> before entering the mass spectrometer. Mass spectra were acquired in centroid mode ranging from 150-2,500  $m/z$  at a 2 Hz full scan rate. Mass spectrometry source parameters were set to 500 V as end plate offset, 4,000 V as capillary voltage, 1 bar nebulizer gas pressure, 5 L min<sup>-1</sup> dry gas flow and 200 °C dry temperature. For MS<sup>2</sup> experiments, CID (collision-induced dissociation) energy was ramped from 35 eV for 500  $m/z$  to 45 eV for 1,000  $m/z$ . MS full scan acquisition rate was set to 2 Hz and MS/MS spectra acquisition rates were ramped from 1 to 4 Hz for precursor ion intensities of 10 kcts to 1,000 kcts. We used *Compass DataAnalysis* version 4.4 (Bruker) to interpret MS data.

### 2.6.7 Quantification of Cys919-1 production

Quantification of Cys919-1 in the heterologous producer *M. xanthus* DK1622 pMYC20Cys\_v2 and native producer strain *M. fulvus* SBMx122 was done using an amaZon speed 3D ion trap MS system (Bruker) with an Apollo II ESI source. ESI source settings were identical as described above. We measured Cys919-1 standard solutions with concentrations of 0.001 mg mL<sup>-1</sup>, 0.005 mg mL<sup>-1</sup>, 0.01 mg mL<sup>-1</sup>, 0.05 mg mL<sup>-1</sup>, and 0.1 mg mL<sup>-1</sup>. Solutions for each concentration were prepared three times and measured two times. EIC  $m/z$  920.3 [M+H]<sup>+</sup> peak surface of Cys919-1 were integrated

manually using *Compass DataAnalysis* version 4.4. The mean values of each measurement were used to construct a regression line that was used to calculate the quantity of Cys919-1 in the cultivation extracts.

### 2.6.8 Protein overexpression and purification

Standard protocols were used for DNA amplification by PCR, cloning procedures, transformation of *E. coli* and plasmid DNA purification after standard alkaline lysis.<sup>42</sup> Genomic DNA was extracted from *C. velatus* Cbv34 with Gentra Puregene DNA Purification Kit (Qiagen). DNA fragments encoding CysJ, CysH, CysH $\Delta$ AMDH, CysB, CysA, the separate modules 1-4 of CysK and modules 5-6 of CysG were amplified by PCR using the gDNA of Cbv34 as template and the following primers: CysJ for and CysJ rev, CysH for and CysH rev, CysB for and CysB rev, CysA for and CysA rev, CysK1 for and CysK1 rev 1, CysK2 for 0 and CysK2 rev 1/2, CysK3 for 1/2 and CysK3 rev 1, CysK4 for 1/2 and CysK4 rev, CysG5 for and CysG5 rev 1, CysG6 for 0 and CysG6 rev, respectively. CysH $\Delta$ AMDH was amplified from pMYC20Cys\_v2 $\Delta$ AMDH using CysH for and CysH rev primers.

The amplified DNA fragment encoding CysJ was hydrolyzed with *Nde*I and *Bam*HI and cloned into pET-28b (Novagen) with an N-terminal 6xHis tag. CysH and CysB encoding fragments were hydrolyzed with *Nco*I and *Bam*HI and cloned into pHISUMOTEV<sup>47</sup> with an N-terminal 6xHis tag, respectively. The CysA encoding fragment was hydrolyzed with *Nde*I and *Bgl*II and cloned into the second multiple cloning site (MCS) of pETduet-1 (Novagen) generating pETdeut-1-cysA. CysK-M1-4 and CysG-M5-6 encoding fragments were digested with *Bam*HI and *Hind*III and cloned into the first MCS of pETduet-1-cysA to yield N-terminal 6xHis tag and TEV protease site fusion constructs. All generated constructs were Sanger sequenced (LGC Genomics GmbH) to verify that no mutation had been introduced during PCR.

Recombinant protein expression was carried out in *E. coli* BL21 (DE3) grown in LB medium and supplemented with 50  $\mu\text{g mL}^{-1}$  kanamycin or 100  $\mu\text{g mL}^{-1}$  ampicillin. The culture was inoculated with 1/10 volume from a fully-grown overnight culture and cultivated at 37°C until OD<sub>600</sub> 0.6-0.9 was reached. The temperature was decreased to 16 °C and after 30 min at 16°C 0.1 mM isopropyl- $\beta$ ,D-thiogalactopyranoside (IPTG) was supplemented to induce gene expression. The cells were harvested 16 h after induction by centrifugation at 8,000  $\times$  g for 10 min at 4°C and resuspended in lysis buffer (25 mM TRIS (pH 7.5), 150 mM NaCl, 20 mM imidazole). The cells were lysed by passage at 1,500 bar at 4 °C using an M-110P Microfluidizer (Microfluidics). The crude extract was centrifuged at 45,000  $\times$  g for 15 min at 4 °C. The supernatant was treated as soluble protein fraction.

Affinity chromatography was performed on an Äkta Avant system and size exclusion chromatography (SEC) was performed on an Äkta Pure system. The soluble protein fraction was applied to 5 mL Ni-NTA cartridge (GE) equilibrated with lysis buffer. The Ni-NTA column was washed with 10 CV (column volumes) lysis buffer prior to elution with elution buffer containing 250 mM imidazole. For CysJ the pooled fractions were concentrated to less than 5 mL and loaded on a HiLoad 16/600 Superdex 200 pg SEC column equilibrated in running buffer (25 mM TRIS (pH 7.5), 150 mM NaCl, 2 mM DTT). For CysH, CysB, CysK-M1-4 and CysG-M5-6 the pooled fractions were applied to a HiPrep 26/10 desalting column equilibrated in running buffer. The resulting fractions were pooled and incubated overnight at 4 °C with TEV protease (1 mg/20 mg protein). After 16 h incubation 20 mM imidazole was added to the solution prior to loading on a 5 mL Ni-NTA cartridge (GE) equilibrated with lysis buffer. The Ni-NTA column was washed with 10 CV lysis buffer prior to elution with elution buffer containing 250 mM imidazole. The pooled fractions were concentrated to less than 5 ml and loaded on a HiLoad 16/600 Superdex 200 pg SEC column equilibrated in running buffer. Protein purity was determined by SDS-PAGE. Proteins were stored at -80 °C in 25 % glycerol.

### 2.6.9 Substrate loading experiments

All substrate loading experiments were performed in 50 µL scale using 25 mM TRIS (pH 7.5), 150 mM NaCl, 10 mM MgCl<sub>2</sub> and 1 mM ATP. For CysH, CysHΔAMDH, CysK-M1-4 and CysG-M5-6, 5 µM protein and 1 mM amino acid were used to test loading of different substrates. Hydroxylation of CysH-bound L-asparagine by CysJ was tested by additionally supplementing 1 mM α-KG, 50 µM FeSO<sub>4</sub> and 500 nM CysJ. To analyze the transfer of L-asparagine from CysH to CysB, 1 µM CysH was incubated with 1 mM L-asparagine and 5 µM CysB. The transfer of L-asparagine from CysB to CysK-M3 was tested by incubating 5 µM CysK-M3, 1 µM CysH, 500 nM CysB with 1 mM L-asparagine. All reactions were incubated for either 5 min or 2 h at RT, respectively. Direct intact protein UPLC-ESI-MS analysis was done as described below.

### 2.6.10 Direct intact protein UPLC-ESI-MS analysis

Direct intact protein UPLC-ESI-MS analysis was performed using an UltiMate 3000 UPLC system coupled with a maXis4G Q-ToF mass spectrometer using an Apollo II ESI source in positive mode. The samples were separated using an Aeris Widepore XB-C8 column (3.6 µm, 150 x 2.1 mm; Phenomenex). Separation was performed at a flow rate of 0.3 mL min<sup>-1</sup> (eluent A: deionized water + 0.1 % FA, eluent B: acetonitrile + 0.1 % FA) at 45 °C using the following gradient: 2 % B for 30 s, followed by a linear gradient up to 75 % B in 10 min and a constant percentage of 75 % B for further 3 min. Original conditions were adjusted with 2 % B within 30 s and kept constant for 3 min. The LC flow was split to 75 µL min<sup>-1</sup> before entering the mass spectrometer. Mass spectra were

acquired in centroid mode ranging from 150-2,500  $m/z$  at a 2 Hz full scan rate. Mass spectrometry source parameters were set to 500 V as end plate offset, 4,000 V as capillary voltage, 1.1 bar nebulizer gas pressure, 6 L  $\text{min}^{-1}$  dry gas flow and 180 °C dry temperature. Protein masses were deconvoluted by using the Maximum Entropy deconvolution algorithm in *Compass DataAnalysis* version 4.4.

### 2.6.11 Cysteamine unloading assay and HPLC-MS analysis

To unload CysH-bound L-asparagine or  $\beta$ -hydroxy-L-asparagine, cysteamine was added to a final concentration of 100 mM and incubated at 30 °C for 1 h with slow shaking. The free amines of the unloaded substrate were derivatized with ethyl carbamates prior to HPLC-MS analysis by adding 45  $\mu\text{L}$  ethanol:pyridine (4:1) solution and 5  $\mu\text{L}$  ethyl chloroformate (ECF). After addition of 200  $\mu\text{L}$  deionized water the derivatized NN-diethoxycarbonyl  $\beta$ -hydroxyasparaginyl dicysteamine was extracted twice with 300  $\mu\text{L}$  ethyl acetate and 1 % ECF. The collected organic layers were dried, dissolved in methanol and analyzed through HPLC-MS. Chemical synthesis of the di(ethylcarbonyl)asparaginyl-dicysteamine references is described in Supplementary Information.

All measurements were performed using an UltiMate 3000 RSLC system coupled with a with an amaZon speed 3D ion trap mass spectrometer using an Apollo II ESI source in positive mode. The samples were separated using a BEH C18 column (1.7  $\mu\text{m}$ , 100 x 2.1 mm; Waters). Separation was performed at a flow rate of 0.6  $\text{mL min}^{-1}$  (eluent A: deionized water + 0.1 % FA, eluent B: methanol + 0.1 % FA) at 45 °C using the following gradient: starting conditions 5 % B for 30 s, linear gradient up to 20 % B in 1 min, linear gradient up to 30 % B in 13 min, linear gradient up to 95 % B in 3 min, constant percentage of 95 % B for 3 min, adjusting of original conditions (5 % B) in 30 s. The LC flow was split to 75  $\mu\text{L min}^{-1}$  before entering the mass spectrometer. Mass spectra were acquired in centroid mode ranging from 150-1,500  $m/z$ . Mass spectrometry source parameters were set to 500 V as end plate offset, 4,000 V as capillary voltage, 1 bar nebulizer gas pressure, 5 L  $\text{min}^{-1}$  dry gas flow and 200 °C dry temperature. MS spectra were interpreted using *Compass DataAnalysis* version 4.4.

## 2.7 Methods References

42. Sambrook, J. & Russell, D. W. *Molecular cloning: A laboratory manual* (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 2001).
43. Agatep, R., Kirkpatrick, R. D., Parchaliuk, D. L., Woods, R. A. & Gietz, R. D. Transformation of *Saccharomyces cerevisiae* by the lithium acetate/single-stranded carrier DNA/polyethylene glycol protocol. *Technical Tips Online* **3**, 133–137; 10.1016/S1366-2120(08)70121-1 (1998).

44. Zhang, Y., Buchholz, F., Muyrers, J. P.P. & Stewart, A. F. A new logic for DNA engineering using recombination in *Escherichia coli*. *Nature Genetics* **20**, 123; 10.1038/2417 (1998).
45. Kashefi, K. & Hartzell, P. L. Genetic suppression and phenotypic masking of a *Myxococcus xanthus* *frzF*- defect. *Mol. Microbiol.* **15**, 483–494 (1995).
46. Pogorevc, D. *et al.* Production optimization and biosynthesis revision of coralopyronin A, a potent anti-filarial antibiotic. *Metab. Eng.* **55**, 201–211; 10.1016/j.ymben.2019.07.010 (2019).
47. Liu, H. & Naismith, J. H. A simple and efficient expression and purification system using two newly constructed vectors. *Protein Expr. Purif.* **63**, 102–111; 10.1016/j.pep.2008.09.008 (2009).

## 2.8 Supplementary Information

### 2.8.1 *In silico* analysis and revision of the native cystobactamid BGC sequence

The previously published cystobactamid BGC sequence (GenBank accession number: KP836244) of *C. velatus* Cbv34 includes 52,081 bp with a GC content of 65 %. During the dotplot analysis (*EMBOSS* 6.5.7 tool *dottup*: <http://emboss.sourceforge.net/>) of the BGC we found three repetitive sequence segments in *cysK* (27,422 bp to 41,059 bp) (Supplementary Figure 1). The sequence repeats were identified between modules 1 (C1, A1, T1) and 2 (C2, A2, T2), between modules 1 and 4 (C4, A4, T4), and between modules 2 and 4. Therefore, NRPS modules 1-4 of *cysK* were separately amplified by PCR and re-sequenced using Sanger sequencing (LGC Genomics GmbH). Primer binding sites were located in unique linker regions between the modules to avoid multiple binding in repetitive sequence segments. We used one to two forward and reverse primers for the amplification of the same sequence stretch, respectively (two separate PCR reactions per module). We used the primers CysK1 for, CysK1 rev 1, CysK 1 rev 1/2 to amplify module 1, CysK2 for 0, CysK2 rev 1, CysK2 rev 1/2 to amplify module 2, CysK3 for 0, CysK3 rev 1, CysK3 rev 1/2 to amplify module 3 and CysK4 for 0, CysK4 rev, CysK2 rev 1/2 to amplify module 4 (Supplementary Table 3). PCR products were ligated into pETDeut-1 after vector/insert hydrolysis using *Bam*HI and *Hind*III. We found 71 differences compared to the previously published sequence (GenBank accession number: KP836244) leading to the revision of the sequence (GenBank accession number will be supplied upon acceptance of the manuscript).

### 2.8.2 In silico design of the modified BGC

The unrevised Cbv34 BGC sequence (GenBank: KP836244) was initially used as template sequence to design the modified BGC. The modified BGC (GenBank accession number will be supplied upon acceptance of the manuscript) was organized in two transcriptional units CysOp1 and CysOp2 (Supplementary Figure 2). CysOp1 contains native operon *cysA – N* (without native promoter, RBS of *cysA* and terminator). CysOp2 combines two native operons *cysO – T* and *Orf1 – Orf5*. *Orf1 – Orf5* was engineered downstream of *cysO – T* (native promoters, terminators and RBS of *cysO* excluded; RBS of *Orf5* included). Vanillate-inducible promoter system<sup>1</sup> ( $P_{van}$ ) was used for inducible gene expression in *M. xanthus* DK1622.  $P_{van}$ , including *vanR* encoding repressor gene, was engineered upstream of CysOp1 and only  $P_{van}$  (without *vanR*) was added upstream of CysOp2. *tD1* terminator sequence from *M. xanthus* bacteriophage Mx8<sup>2</sup> was engineered downstream of both modified operons, respectively. For TAR assembly, we engineered *LEU2* gene encoding  $\beta$ -isopropylmalate dehydrogenase auxotrophy marker downstream and 100 bp sequences originating from *URA3* gene downstream and upstream of both operons, respectively. We added unique R-sites for cloning purposes. Supplementary Table 4 lists all genetic elements including sequence origin and unique R-sites used for the design of CysOp1 and CysOp2. We removed R-sites (e.g. *BsaI*) by synonymous codon substitutions while keeping the codon usage bias for a single amino acid as similar as possible to the native codon usage bias in the BGC (Supplementary Table 5). To reduce costs and turnaround time for DNA synthesis, we divided the modified BGC into twelve fragments (Supplementary Table 6). We added 100 bp homologous sequences to all adjacent cluster fragments used for TAR cloning. R-sites were engineered at the 5' and 3' ends of all fragments for DNA synthesis vector release or for step-wise assembly in cloning vectors. We flanked some fragments by splitter elements (SEs) if they were used for three-step assembly in cloning vectors.<sup>3</sup> SEs consist of unique 'conventional' type II R-sites flanked by two *BsaI* recognition sequences extended with 5 bp sequences. *BsaI* is a type IIS restriction endonuclease cutting outside of the recognition sequence. This allowed generation of variable and unique 5 bp sticky ends for ligation. Unique 'conventional' R-sites allowed stepwise cloning of several fragments into a cloning vector.

### 2.8.3 In silico design of the cloning and expression vector system pMYC

Three pMYC vector system fragments were designed for TAR assembly of the cystobactamid BGC in yeast, standard cloning procedures in *E. coli* and transformation and integration of the cluster into *M. xanthus* DK1622 genome. The basic pMYC building block contains *p15A* origin of replication (*ori*) for replication in *E. coli*, *cat* gene encoding chloramphenicol acetyl transferase (mediating resistance towards chloramphenicol), origin of transfer (*oriT*) and transfer gene (*traJ*) for

conjugation (not used in this study), *CEN6/ARS4* ori for replication in yeast, and *URA3* encoding orotidine 5'-phosphate decarboxylase for counter-selection in yeast. A unique *EcoRV* R-site was located in *URA3* for vector linearization prior to TAR cloning. The flanking 100 bp sequences upstream and downstream of the *EcoRV* R-site are homologous to short sequences added to CysOp1 and CysOp2 to capture the marginal DNA synthesis fragments during TAR cloning. A second pMYC building block (Mx8-tetR) contained the genes *mx8 int* and *tetR*. *mx8 int* encodes *mx8* integrase for integration of plasmids via phage attachment site (*attP*) in the genome of *M. xanthus* DK1622 or other strains harboring the appropriate *attB* site. *tetR* encodes an efflux transporter mediating oxytetracycline resistance. A third pMYC building block (Mx9-kanR) contained genes *mx9 int* and *aph(3')-Ia*. *mx9 int* encodes *mx9* integrase for *M. xanthus* DK1622 genome integration and *aph(3')-Ia* encodes aminoglycoside-3'-phosphotransferase-Ia mediating kanamycin. All genetic elements including unique R-sites and their sequence origin are listed in Supplementary Table 7. Synonymous codon substitutions were performed to remove R-sites from pMYC building blocks (Supplementary Table 8). Supplementary Figure 3 shows vector maps of basic pMYC, pMYC20 (basic pMYC with Mx8-tetR) and pMYC21 (basic pMYC with Mx9-kanR). GenBank accession numbers will be supplied upon acceptance of the manuscript.

#### 2.8.4 Assembly of the modified BGC

At the time of sequence revision, we already ordered cluster fragments with sequences based on the previously published *cysK* sequence. Thus, fragments K1, K2 and K3 were completely (and 3-GHIJ5-K partly) synthesized again and named K1\_v2, K2\_v2, K3\_v2 and 3-GHIJ5-K\_flong\_v2, respectively. 3-GHIJ5-K\_flong\_v2 contains 1,222 bp of the 3' end of 3-GHIJ5-K (without mutations) and 1,796 bp of pGH DNA synthesis vector, including 559 bp of *ampR* (*bla*). We used unique R-sites to exchange the mutated part in 3-GHIJ5-K for the generation of 3-GHIJ5-K\_v2. The DNA fragments hPvanABCDEF5-G, K123\_v2, 3-KLNtD1LEU2h (all part of CysOp1) and ABC3-2345tD1LEU2h (CysOp2) were assembled using pSynbio1<sup>3</sup> as cloning vector. Fragments K123\_v2 and 3-KLNtD1LEU2h contained SEs, which were removed by restriction hydrolysis with *BsaI* and re-ligation (rejoining) using T4 ligase. After rejoining of the fragments, the generated construct did not contain additional R-sites (Supplementary Figure 4). Release of the final fragments from pSynbio1 was achieved by *BsaI* hydrolysis. Generation of 3-GHIJ5-K\_v2 was achieved by ligating 3-GHIJ5-K\_flong\_v2 into pGH-3-GHIJ5-K. Release of 3-GHIJ5-K\_v2, G, hPvanOPQRS and T-ABC15-2 from pGH was achieved by *BsaI* hydrolysis.

KDummy was generated via PCR using K-rpsLF and K-rpsLR primers pairs (Supplementary Table 3) and pSW1b plasmid as template (unpublished data). pSW1b contains the *rpsL* counter selection marker (mediating streptomycin sensitivity), flanked by *BsaI* R-sites and 50 bp sequences

of *cysK*, which are homologous to the 3' end of 3-GHIJ5-K\_v2 and the 5' end of 3-KLNtD1LEU2h. To remove a single point mutation from KDummy (after sequence revision), KDummy was cloned into pJET1.2blunt vector. We used two overlapping, complementary, mutagenic primers KrpsLF\_v2\_SDM and KrpsLR\_v2\_SDM to generate pJET1.2KDummy\_v2 from pJET1.2KDummy via PCR. Template DNA was hydrolysed with *DpnI*. KDummy\_v2 was released from pJET1.2 using *XbaI/XhoI* hydrolysis.

TAR cloning was used to assemble pMYC20preCysOp1\_v2 and pMYC21CorOp2. For pMYC20preCysOp1\_v2 assembly, DNA fragments hPvanABCDEFG, G, 3-GHIJ5-K\_v2, KDummy\_v2, 3-KLNtD1LEU2h and *EcoRV*-linearized pMYC20 were used. KDummy\_v2 was used instead of *cysK* fragments (K123\_v2) to avoid unspecific recombination caused by repetitive sequence segments. pMYC21CysOp2 was cloned from fragments hPvanOPQRS, T-ABC15-2, ABC3-2345tD1LEU2h and *EcoRV*-linearized pMYC21. During TAR assembly, the counter selection marker *URA3* in the vectors was disrupted whereas *LEU2* was introduced into the plasmid together with the cluster fragments. Transformation of *S. cerevisiae* ATCC4004247 was done according to the *standard high-efficiency transformation protocol I* described by Agatep et al.<sup>4</sup> However, centrifuging was done at  $3,200 \times g$  and heat shock lasted for 45 min. Constructed plasmids were isolated according to the protocol from Kouprina & Larionov<sup>5</sup> (*Identification of gene-positive pools*) and transformed into *E. coli* DH10 $\beta$ . Clones harboring the correct construct were verified by restriction analysis.

KDummy\_v2 was replaced by K123\_v2 in pMYC20preCysOp1\_v2 via restriction/ligation to generate pMYC20CysOp1\_v2. To generate final expression construct pMYC20Cys\_v2, we ligated CysOp2 into pMYC20CysOp1\_v2. Supplementary Figure 5 schematically depicts all cloning steps performed to obtain the final expression construct pMYC20Cys\_v2. Supplementary Table 9 lists all conventional cloning steps for the generation of pMYC20Cys\_v2. We verified all generated constructs by restriction analysis. Additionally, the Illumina paired-end technology on a MiSeq PE300 platform (in-house) was used to verify the sequences of pMYC20CysOp1\_v2 (2,244-fold mean sequencing coverage), pMYC21CysOp2 (2,856-fold) and pMYC20Cys\_v2 (1,108-fold). Supplementary Table 1 and Supplementary Table 2 list all strains and plasmids generated during cloning process.

### **2.8.5 Identification of the new cystobactamid derivatives**

We identified a total of 22 new cystobactamid derivatives (13 natural and 9 unnatural) based on high-resolution masses and the MS<sup>2</sup> fragmentation patterns. The fragment ions that we used to identify the structural differences of the new derivatives, including the respective calculated fragment masses, are shown in Supplementary Figure 6 and Supplementary Table 10. We only

focused on the three fragments a, b and c, because fragment a only contains the N-terminal cystobactamid part ( $pNBA_1$ - $pABA_2$ ), whereas fragment b contains the N-terminal part plus the linker and one additional  $pABA$  ( $pNBA_1$ - $pABA_2$ -linker- $pABA_3$ ). Fragment c contains one additional (tailored)  $pABA$  ( $pNBA_1$ - $pABA_2$ -linker- $pABA_3$ - $pABA_4$ ) compared to fragment b. Those three fragments were also described previously by Baumann and coworkers (in the SI).<sup>6</sup> We decided to look at those three fragments, because it enabled us to find structural differences in the N-terminal part, e.g. after deletion of *cysR* ( $m/z$  269.0562  $\rightarrow$  239.0821 [M+H]<sup>+</sup>; NO<sub>2</sub>  $\rightarrow$  NH<sub>2</sub>), in the linker, e.g. after deletion of *cysQ* ( $m/z$  532.1468  $\rightarrow$  518.1312 [M+H]<sup>+</sup>; methoxy-(iso)Asn  $\rightarrow$  hydroxy-(iso)Asn), and in the C-terminal part, e.g.  $m/z$  725.2207  $\rightarrow$  711.2051 [M+H]<sup>+</sup>; isopropoxyl- $pABA_4$   $\rightarrow$  ethoxy- $pABA_4$ ). The differences between experimental masses and the calculated masses of the new derivatives (and their corresponding fragments) are listed in Supplementary Table 10. Based on the high-resolution mass of the new derivatives and their fragmentation pattern we were able to propose the structures (Supplementary Figure 7) using known cystobactamids as template (example depicted in Supplementary Figure 14).

E.g. after the deletion of *cysQ*, the gene encoding the *O*-methyltransferase that acts on the linker, and heterologous expression of the deletion construct in *M. xanthus* DK1622, we did not find the two major cystobactamid peaks (Cys919-1 and Cys919-2) in the EIC with  $m/z$  920.3097 [M+H]<sup>+</sup>. Instead we found two new masses with  $m/z$  906.3004 (Cys905-1c) and 906.3007 (Cys905-2c) [M+H]<sup>+</sup> (red box in Supplementary Figure 14). The fragmentation pattern shows that fragment a (blue box), which is the N-terminal cystobactamid part ( $pNBA_1$ - $pABA_2$ ) of Cys905-1c/2c is the same as for Cys919-1/2. However, fragment b (green box) differs when comparing Cys919 with Cys905, as there is a mass shift of 14 Da ( $m/z$  532.1468  $\rightarrow$  518.1312 [M+H]<sup>+</sup>), indicating that a methyl group is missing in Cys905 (structural difference shown in red). In fragment c (yellow box) and in the mass of the molecular ion (red box) the mass shift of 14 Da is still present, which means that no further structural difference is found between Cys919 and Cys905. Assignment of the linker moieties with the same masses (e.g. linker G and F for Cys905-1c and Cys905-2c, respectively) was done based on the different retention times that were also observed for previously reported cystobactamids, in which derivatives harboring L-isoasparagine linkers eluted between 1.4 and 1.6 min earlier than derivatives harboring L-asparagine linkers.

However, for four unnatural cystobactamids, we were not able to propose a putative structure: Cys905-2d, Cys905-2e, Cys919-2b<sub>1</sub> and Cys919-2b<sub>2</sub>. In *M. xanthus* DK1622 pMYC20Cys\_v4Δ*cysJ* and *M. xanthus* DK1622 pMYC20Cys\_v4Δ*cysJ*Δ*AMDH* we found the two new minor Cys905 derivatives, Cys905-2d and Cys905-2e (Supplementary Figure 15). The RTs are different to the one observed for Cys905-1c/2c (major products in the extract of *M. xanthus* DK1622 pMYC20Cys\_v4Δ*cysQ*), but surprisingly, the fragmentation pattern is the same (Supplementary

Table 10). Since we found Cys905-2d and Cys905-2e in the extracts of the *cysJ* and the *cysJ*-AMDH deletion strains, we speculate that the linker is an L-asparagine rather than an L-isoasparagine derivative, because L-isoasparagine linkers only occur if the AMDH domain is present. In the *cysJ* deletion strains we expected the production of exclusively desmethoxylated linker derivatives (fragment b:  $m/z$  502.1363 [M+H]<sup>+</sup>). However, the fragment ion b has an  $m/z$  of 518.1312 [M+H]<sup>+</sup>, which means that fragment b potentially contains an additional hydroxyl group even though *cysJ* was deleted in the producer strains. Since the RT are different compared to Cys905-1c/2c, we can only hypothesize that the hydroxylation occurs in a different position and that it might be catalyzed by a different enzyme encoded in the genome of *M. xanthus* DK1622. However, this enzyme must be active also in the other strains harboring different construct. Thus, we searched for cystobactamid derivatives with double-hydroxylated fragment b ions, but we did not find any of them. Finally, purification and subsequent NMR analysis would be required to understand where the hydroxylation in Cys905-2d/e occurs and why those two derivatives have a different retention time despite the same potential linker moiety. However, due to the very low production titer, this is not possible at this time.

Furthermore, we found two new Cys919 derivatives, Cys919-2b<sub>1</sub> and Cys919-2b<sub>2</sub> (Supplementary Figure 15), in the extract of the AMDH deletion strain. Both derivatives show the same fragmentation pattern compared to each other, but they have a different fragmentation pattern compared to Cys919-2 (Supplementary Table 10). The fragmentation pattern shows that the linker is desmethylated (fragment b:  $m/z$  518.1312 [M+H]<sup>+</sup> compared to  $m/z$  532.1468 [M+H]<sup>+</sup> for Cys919-2), because CysQ is not able to operate in case the AMDH domain is deleted. Since fragment c has a difference of 14 Da compared to fragment c of Cys919-2, the additional methyl group has to be located in the terminal pABA<sub>5</sub> moiety. We suggest that R<sub>2</sub> (see Supplementary Figure 6 and Supplementary Table 10) is a 1-methylpropoxy moiety as this moiety has already been described in native cystobactamids. However, since we identified two derivatives with only slightly different retention times (0.22 min), the methyl group might also be located elsewhere or at different positions in Cys919-2b<sub>1</sub> and Cys919-2b<sub>2</sub>. We also cannot exclude the possibility that there is another structural difference elsewhere in the molecule, which we cannot explain at this time and only based on MS<sup>2</sup> data. As for Cys905-2d/e, we would also have to isolate those derivatives, which is not possible due to the very low production titers.

### 2.8.6 Analysis of CysJ activity on free L-asparagine using TLC

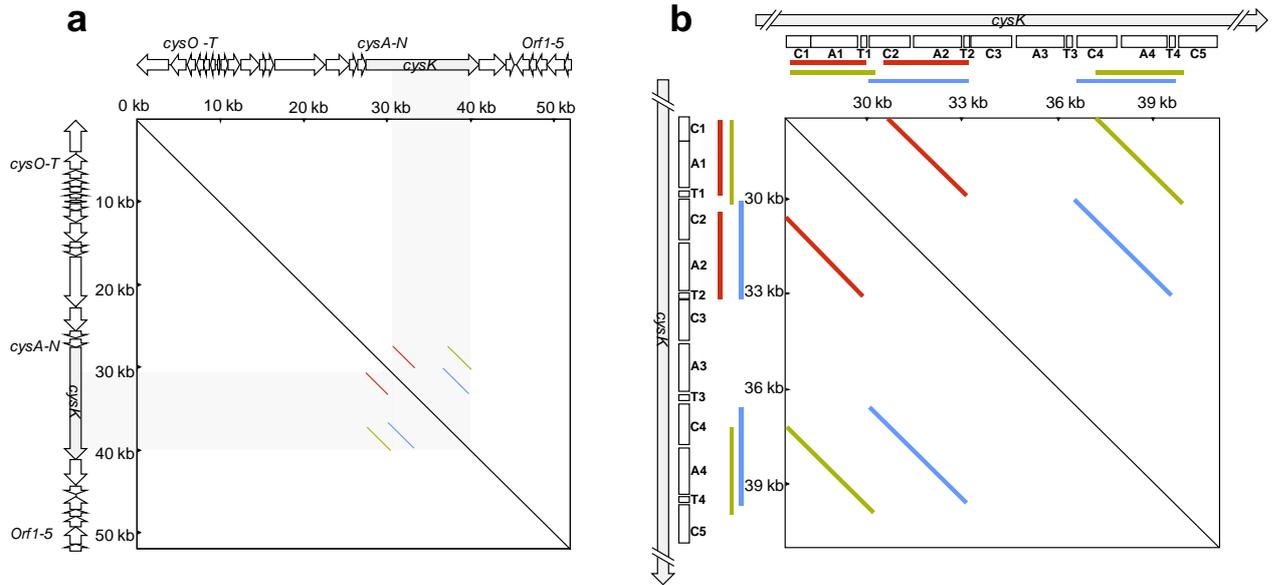
To test if free L-asparagine is hydroxylated by CysJ, 500 nM CysJ was incubated with 1 mM  $\alpha$ -KG and 1 mM L-asparagine and analyzed using TLC with ninhydrine (Supplementary Figure 16 lane 5). We observed different colors for L-aspartate (lane 2; control) and L-isoasparagine (lane 3; control),

respectively, and a different retention factor of L-isoasparagine compared to L-asparagine (lane 1; control). The incubation of L-asparagine with CysJ and  $\alpha$ -KG (lane 5) prior to TLC analysis did not result in visible turnover or changing retention factors compared to the L-asparagine control, L-asparagine incubated with CysJ without  $\alpha$ -KG (lane 4), or L-asparagine incubated with denatured CysJ and  $\alpha$ -KG (lane 6).

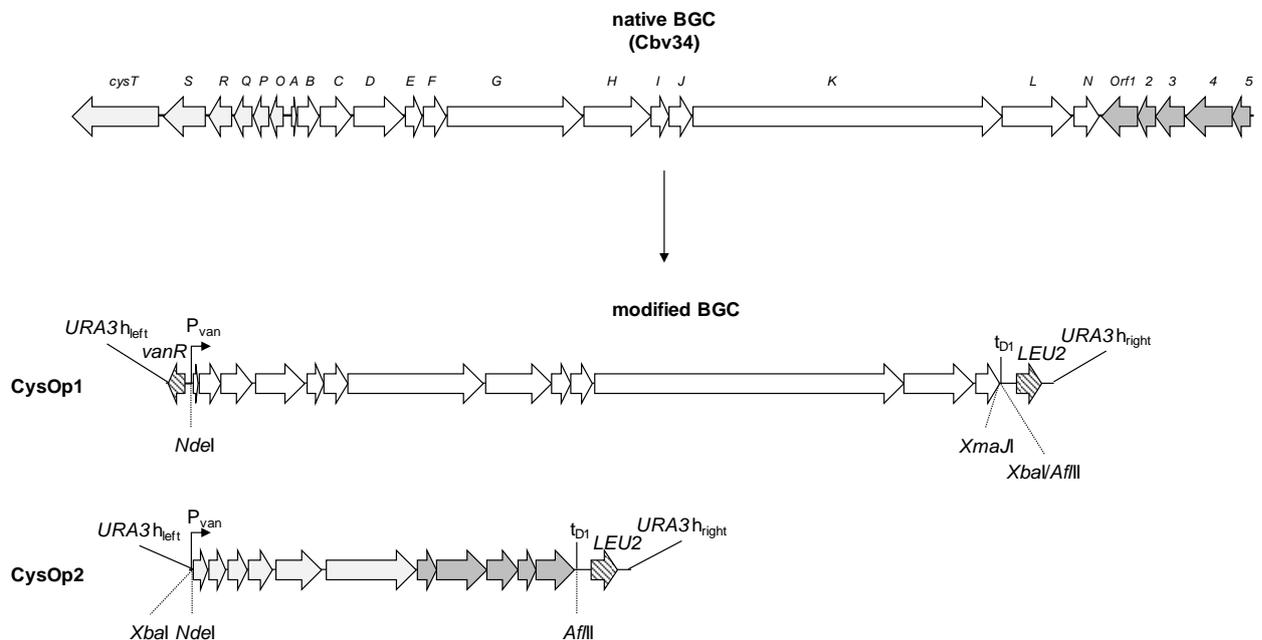
### 2.8.7 Synthesis of the di(ethylcarbonyl)asparaginyldicysteamine references

Supplementary Figure 17 depicts all below mentioned reaction steps for the synthesis of the di(ethylcarbonyl)asparaginyldicysteamine references. 55.6 mg Trityl chloride (0.2 mmol, 2eq.), 19  $\mu$ L acetic anhydride (0.2 mmol, 2eq.) and 6.1  $\mu$ L concentrated H<sub>2</sub>SO<sub>4</sub> (0.115 mmol, 1.15eq.) are added to 500  $\mu$ L glacial acetic acid until dissolution. 35.4 mg L-isoasparagine (0.1 mmol, 1eq.) and 50  $\mu$ L DMF (dimethylformamide) are added and left to react overnight at 60°C in an oil bath. The solution was slowly poured on 2 mL ice cold H<sub>2</sub>O, the pH was adjusted to 6.0 by addition of 10 M NaOH and left on ice for 1 h. The precipitate was filtrated using silica gel, washed with water and subsequently dissolved in DCM (dichloromethane). 37.4 mg of Trityl-L-isoasparagine (100  $\mu$ mol, 1eq.) and 27.9  $\mu$ L TEA (trimethylamine; 200  $\mu$ mol, 2eq.) were dissolved in 500  $\mu$ L THF (tetrahydrofuran) and 34.5  $\mu$ L di-tert-butyl dicarbonate were added and left to react at room temperature for 2 h. Upon completion the reaction was quenched in water, the pH was acidified by addition of 1 M HCl and the boc-L-isoasparagine(trt)-OH was extracted with DCM. 50 mg boc-L-asparagine (trt)-OH or boc-L-isoasparagine (trt)-OH (105  $\mu$ mol, 1eq.), 35.2  $\mu$ L DIPEA (N,N-diisopropylethylamine; 367  $\mu$ mol, 3.5 eq.) and 60.1 mg PyBop (115  $\mu$ mol, 1.1 eq.) were dissolved in 2 mL DCM and stirred at room temperature for 30 min. 29.8 mg cysteamine hydrochloride (262  $\mu$ mol, 2.5eq.) and 1 mL DMSO were added and incubated overnight at room temperature. The reaction was quenched in H<sub>2</sub>O and the product was extracted twice with DCM. The organic fractions were dried, and the white precipitate was washed with pentane. The obtained powder was left to react at room temperature overnight in 2 mL of an 88 % TFA, 5 % phenol, 5 % H<sub>2</sub>O, 2 % TIPS (triisopropylsilane) solution with addition of 2 eq. of cysteamine to prevent thiol disulfide exchange. The reaction was quenched in water, washed three times with DCM and the aqueous layer was lyophilized to yield pure asparaginyldicysteamine. The compound was dissolved in ethanol and left to react with 5% ECF before HPLC-MS analysis.

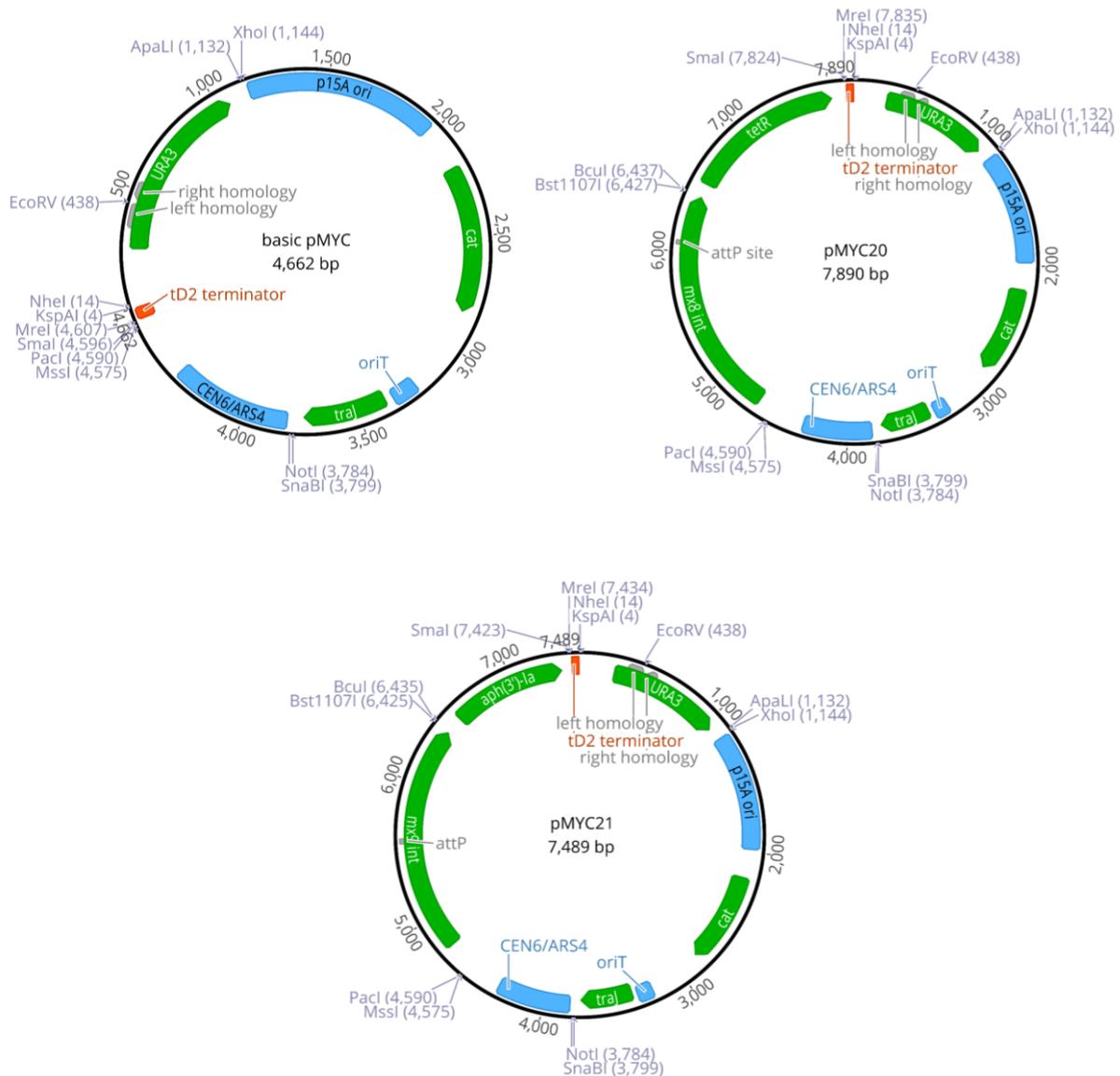
### 2.8.8 Supplementary Figures



**Supplementary Figure 1 | Dotplots of the entire cystobactamid gene cluster and *cysK*.** **a:** Dotplot of entire cystobactamid BGC. Three repetitive sequence segments (red, blue and green lines) span from about 30 kb to 40 kb (highlighted in light grey) in *cysK*. **b:** Dotplot of *cysK*. Magnification of the grey area in a. C: condensation domain, A: adenylation domain, T: thiolation domain. The Dotplot was created using the *EMBOSS 6.5.7 tool dottup* in Geneious.

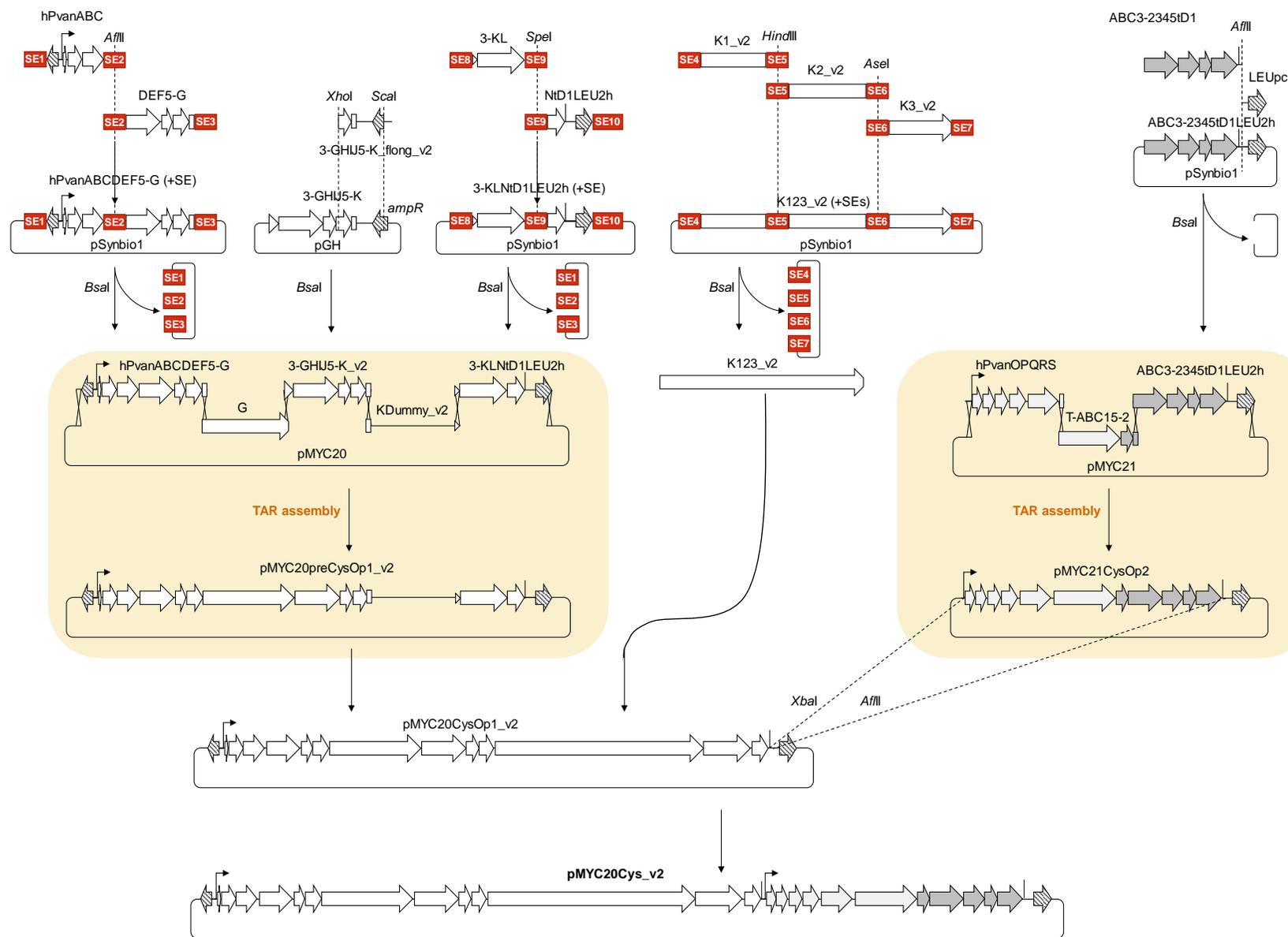


**Supplementary Figure 2 | Design of the modified BGC based on the native BGC.** The modified BGC consists of two operons: CysOp1: *URA3* homology left for TAR assembly, *vanR* repressor, vanillate inducible promoter *P<sub>van</sub>*, biosynthetic genes *cysA-N*, tD1 terminator, *LEU2* auxotrophy marker and *URA3* homology right for TAR assembly. CysOp2: *URA3* homology left, *P<sub>van</sub>*, biosynthetic genes *cysO-T* and *Orf1-5*, tD1, *LEU2* and *URA3* homology right. Location of important R-sites are shown by dashed lines.

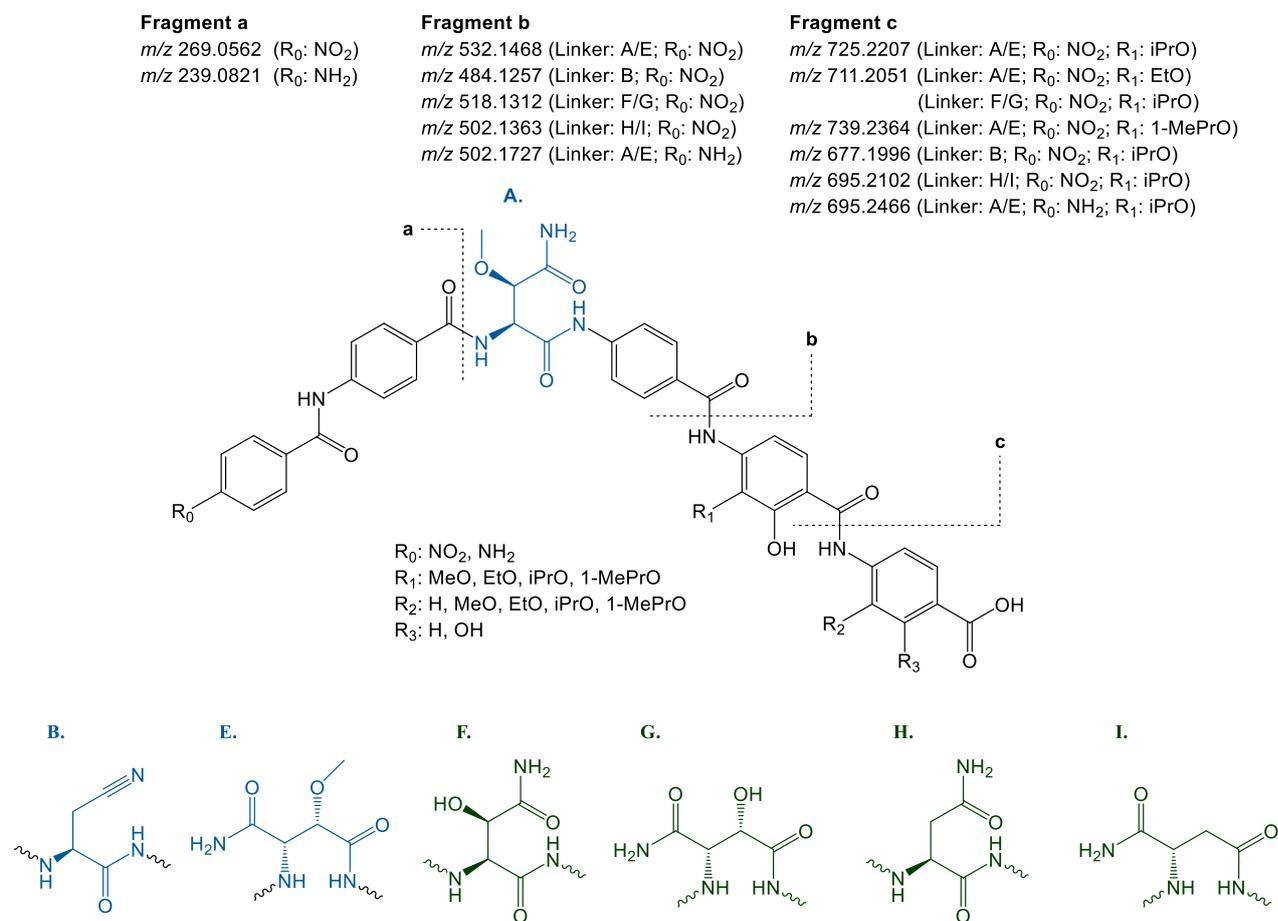


**Supplementary Figure 3 | Plasmid maps of basic pMYC, pMYC20 and pMYC21.** Mx8-tetR building block engineered into basic pMYC results in pMYC20. Likewise, Mx9-kanR building block and basic pMYC generate pMYC21. Genes are annotated using green arrows. Origins of replication or transfer are annotated using blue boxes. Terminator sequence is labeled using red box and homology sequences for TAR cloning are shown in grey. Important unique R-sites are shown in dark blue. Vector maps generated using *Geneious* v10.1.3.

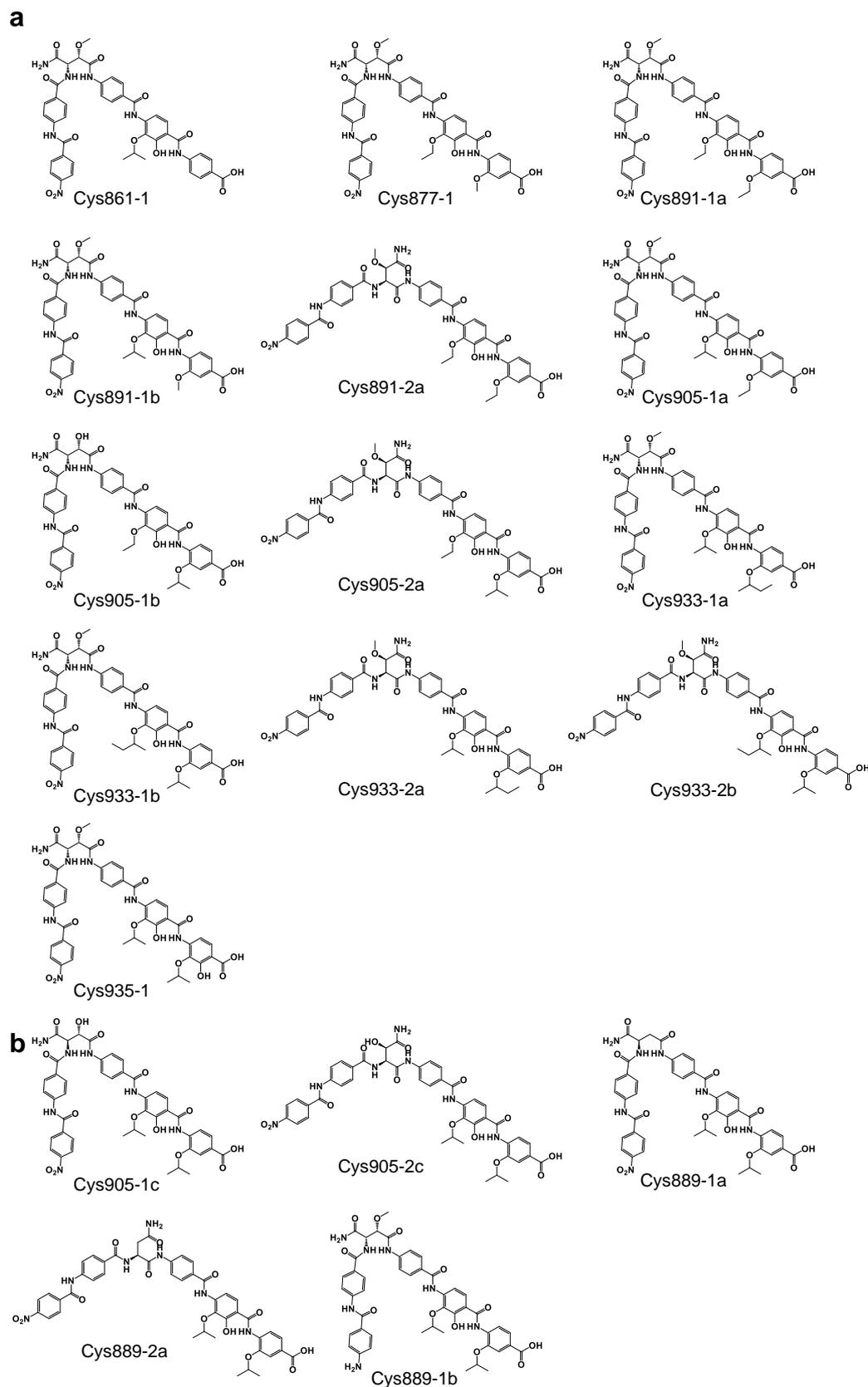




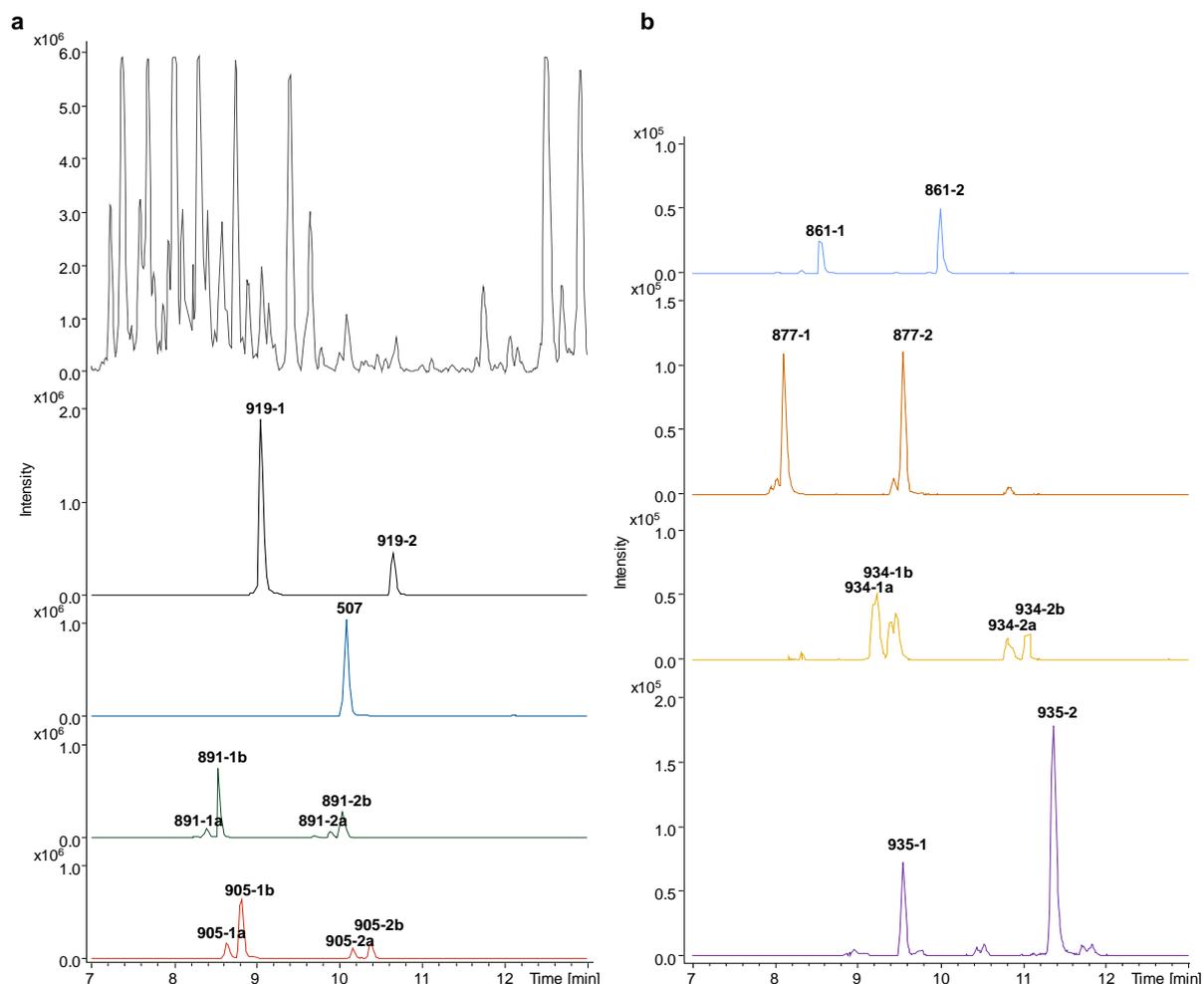
**Supplementary Figure 5 | Assembly strategy of the modified gene cluster.** Both operons of the modified gene cluster are shown on top (see also Supplementary Figure 2). Conventional restriction/ligation-based cloning steps were used to assemble hPvanABCDE5-G, 3-GHIJ5-K\_v2, 3-KLNtD1LEU2h, K123 (CysOp1) and ABC3-2345tD1LEU2h (CysOp2) from synthesized DNA fragments. Splitter elements (SEs; shown in red) were used for stepwise cloning into pSynbio1 (depicted in Supplementary Figure 4). TAR cloning (highlighted in beige) was used to assemble fragments hPvanABCDE5-G, G, 3-GHIJ5-K\_v2, KDummy, 3-KLNtD1LEU2h and pMYC20 to generate pMYC20preCysOp1\_v2. Likewise, pMYC21CysOp2 was assembled via TAR from hPvanOPQRS, T-ABC15-2, ABC3-2345tD1LEU2h and pMYC21. KDummy was replaced by K123 in a restriction/ligation-based cloning step to generate pMYC20CysOp1\_v2. Final expression construct pMYC20Cys\_v2 was generated by cloning CysOp2 downstream of CysOp1 in pMYC20CysOp1\_v2. Supplementary Table 9 lists all restriction/ligation-based cloning steps.



**Supplementary Figure 6 | Fragmentation scheme of natural and unnatural cystobactamids.** Fragments a, b and c were used in MS<sup>2</sup> fragmentation experiments to identify novel natural and unnatural cystobactamids (Table 1, Supplementary Table 10). The calculated  $m/z$   $[M+H]^+$  values of the fragments depend on tailoring of  $R_{0-3}$  and linker moiety.



**Supplementary Figure 7 | Proposed structures of novel natural and unnatural cystobactamids.** Natural (a) and unnatural (b) cystobactamid structures were proposed based on exact mass from HRMS analysis and MS<sub>2</sub> fragmentation.

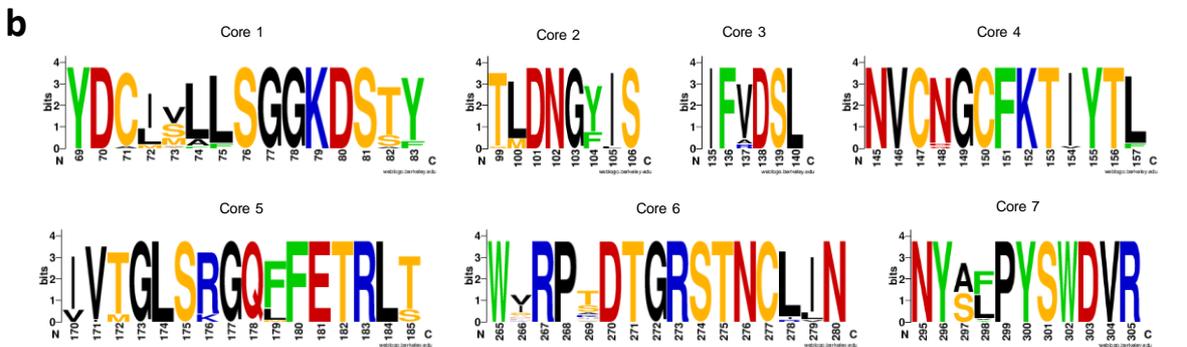
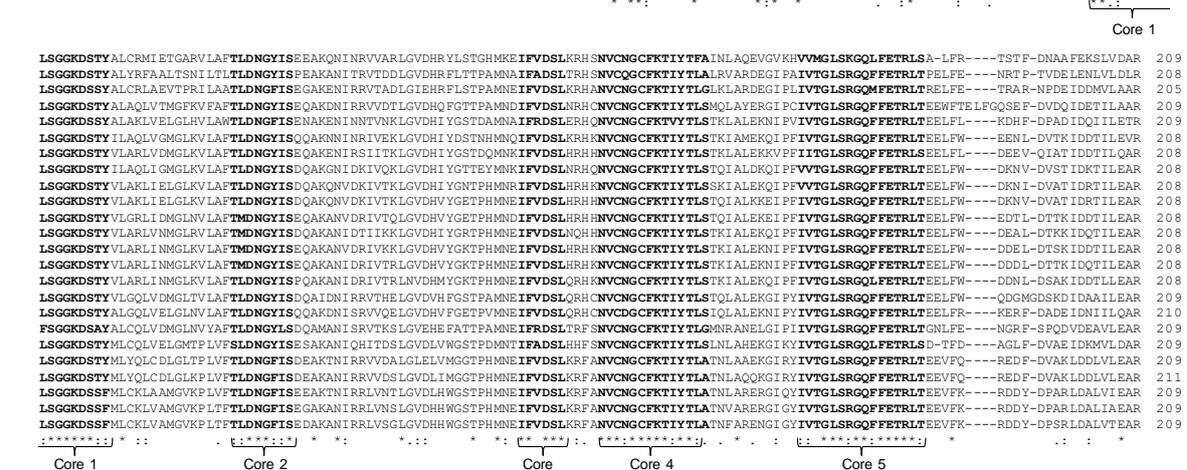


**Supplementary Figure 8 | HPLC-MS analysis of major and minor cystobactamid derivatives produced by *M. xanthus* DK1622 pMYC20Cys\_v2. a:** BPC (dark grey), EIC 920.30 (black), EIC 508.20 (dark blue), EIC 892.27 (dark green) and EIC 906.29 (red) (all EICs  $[M+H]^+ \pm 0.02$  Da) show existence of Cys919-1/2, Cys507, Cys891-1a/1b/2a/2b and Cys9051a/1b/2a/2b. **b:** BPC and EICs of minor native cystobactamid derivatives in methanolic *M. xanthus* DK1622 pMYC20Cys\_v2 extracts. EIC 862.26 (blue), EIC 878.26 (dark orange), EIC 935.30 (red) and EIC 936.30 (purple) (all EICs  $[M+H]^+ \pm 0.02$  Da) show existence of Cys861-1/2, Cys877-1/2, Cys934-1a/1b/2a/2b and Cys935-1/2. Corresponding MS<sup>2</sup> data are listed in Supplementary Table 10.

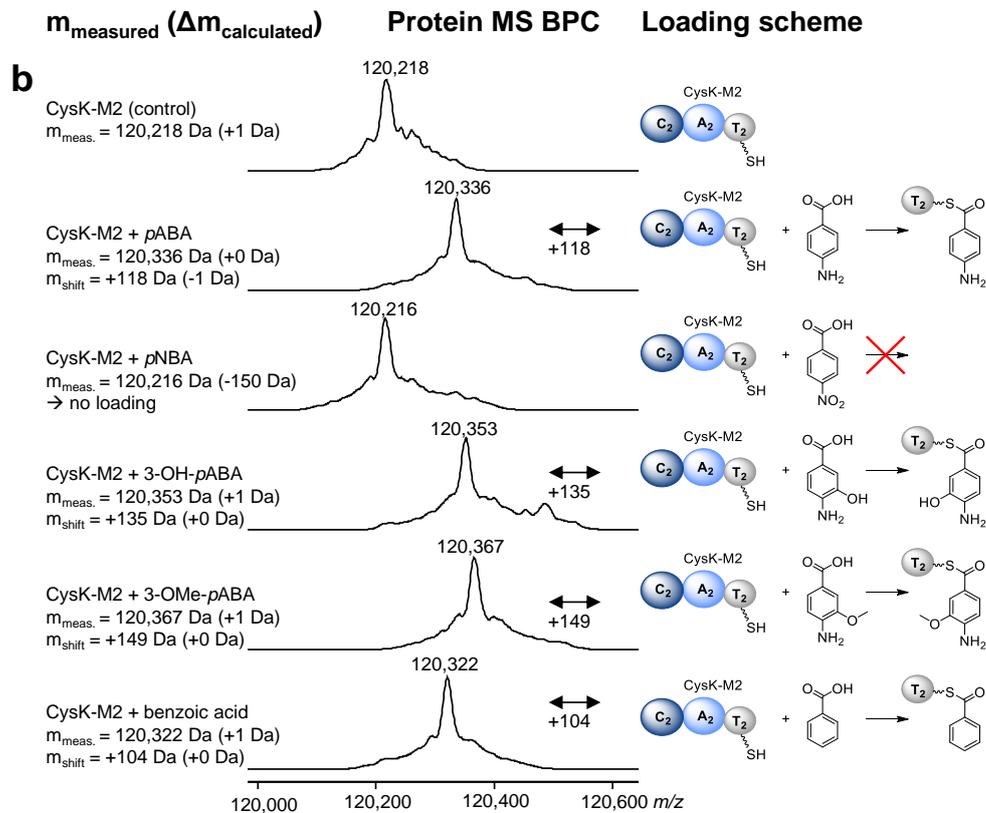
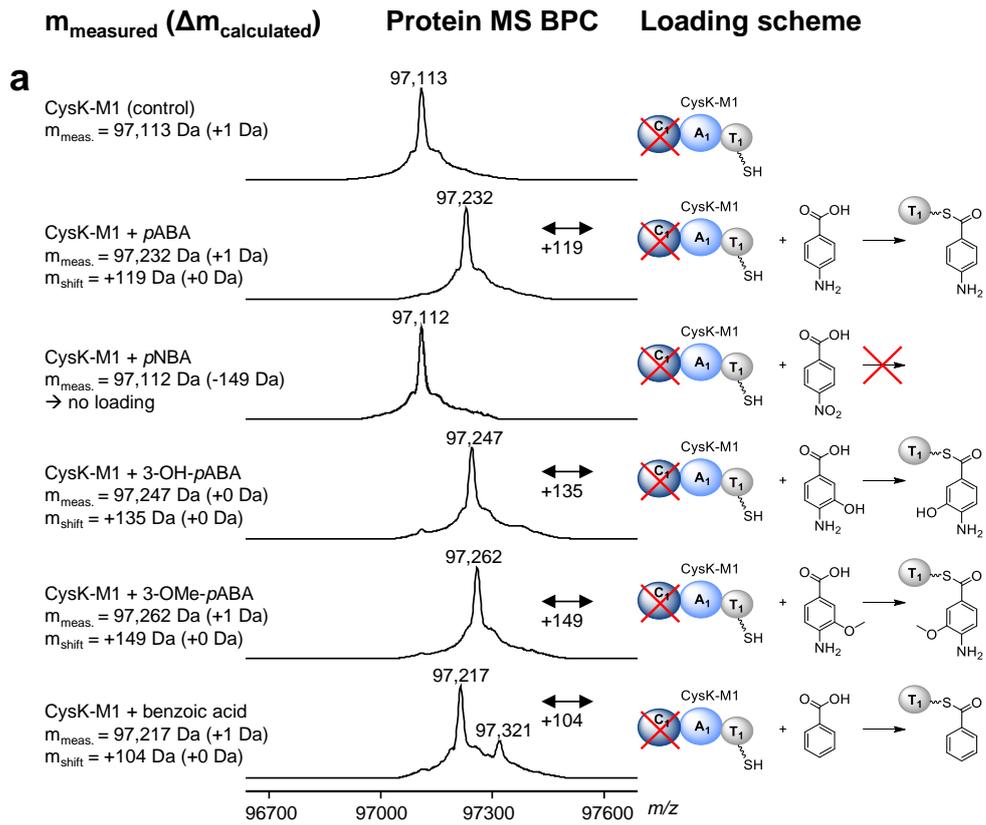


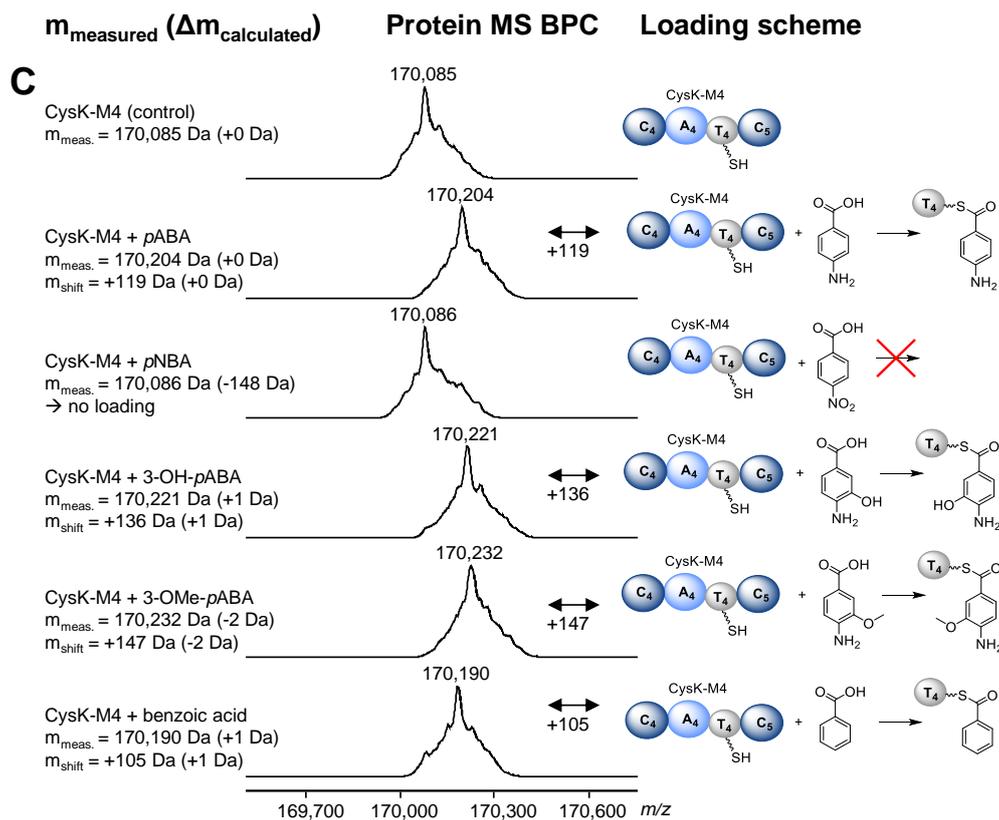
**a**

[Cystobacter velatus Cbv34] CysH_AMDH_domain	-AAQLYTCTKGLASSFPNTTYSAEGVNCHCAFQDKYR.SVVDDY FSTMDLQSIIVTEMKSIHN---SKYDCIVA	74
[Robiginitomaculum_sp.]_P_HR62283.1	-PKQANSTRCRGI.DSVPGT.KHT-NGICNCAFGEFYKHRAAAAYFKTPEALQNLVLTLPARKS---GKYDAIVL	73
[Tropoclonema_sedimicicola]_WP_08923480.8.1	---IAHCTRCGVTGDTGFI.EIDDTGVCTCS.GFDSFRERASEY.FDGLGAFAGIVLDRARFR---GDYDCIML	70
[Leptotyngbya_sp._Heron_Island_J]_WP_023070733.1	-QSSGVYTCRCGI.ASTY.PGI.TFDHGVCDICQ.TYVETVDRTEQY.FKS.IDE.FKSRLEAARQRQ---REYDCIML	74
[Algoriphagus_haloophilus]_WP_074224043.1	-YQADPFYCKNIGI.PSNYPDAS.FNNEGVNCLCE.SFEEYQKIVASY.FKSMN.LKSEIL.GLGNKE---REYDCIML	74
[Confluentibacter_sp._3B]_PK044143.1	-EVEKVINICNGLP.SNYP.EIDFDDQGVCHLCAFKNYK.DKVQY.FRTEAE.LKSLIL.KNKT-NN---PSYDCIL	74
[Arenibacter_sp._AK53]_WP_086478236.1	-PEKDVNCTQCGLP.SNYPKT.DFDES.GVCHLCAFNFYSK.DKALRY.FKTEDE.LAS.LLMSNRG-KN---PTDYCISL	74
[Kriegella_aquimaris]_WP_089884457.1	-SENEVICTECGLP.SNYPNT.DFDENG.VCHLCAFNGYK.DEAQRY.FKT.TDE.LRD.LI.LSKRG-KN---PNYDCISL	74
[Arenibacter_certesii]_WP_026813687.1	-PENEVNICACGLP.SNYPNV.DFDENG.VCHLCAFNFYK.QQAQRY.FKTENE.LRD.LL.LSKRG-NS---PNYDCISL	74
[Arenibacter_latericius]_WP_026810787.1	-PENEVNICACGLP.SNYPNV.DFDENG.VCHLCAFNFYK.QQAQRY.FKTENE.LRD.LL.LSKRG-NS---PNYDCISL	74
[Prisia_antarctica]_WP_091870771.1	-PENEVNICACGLP.SNYP.EIFENQGV.CQLCT.SFSNBYE.DKVKRY.FKNDDE.LVR.LL.LSKRG-ES---PKYDCISL	74
[Gelidiibacter_algens]_WP_066429701.1	-PEEDVNVIC.ECGLP.SNYPNT.DFNENG.VCHLCAF.SFYK.VKAERY.FKNDDE.LQV.LL.LSKRG-QN---PNYDCISL	74
[Zobellia_uliginosa]_WP_038233587.1	-PEEVANCTC.GLPSNYPNA.DFNEDGV.CHLCT.AFETRY.DKTERY.FKNDDE.LV.SLL.LSKRG-QN---RSYDCISL	74
[Flavobacteriaceae_bacterium]_WP_100868958.1	-PEEVANCTC.GLPSNYPST.DFDQGV.CHLCAFADTYK.EKAHRY.FKNDDE.LV.SLL.SKRD-KS---VNYDCISL	74
[Pseudozobellia_thermophila]_WP_072994154.1	-PEEETNC.SRGLP.SNYPGT.DFDQGV.CHLCAFADSYK.EKVDGH.FRNDG.LV.LL.LSKRG-QN---PDYDCISL	74
[Lewinella_marina]_WP_099106946.1	-DSDPDRFCTRCGLP.ANYP.SAFDEAGV.CQLCR.GFADYE.EKARAY.FRT.PED.FRAL.FA.DRPVDRG---APYDCIML	74
[Lewinella_nigricans]_WP_099148200.1	AAKVAHDNHCRCGLP.ANYP.TAE.FDEQGV.CHLCS.FQHYQEN.VQVY.FRT.MAD.LKAL.FE.KP.GSAD---RAYDCIML	75
[Verrucumicrobiae_bacterium_DG1235]_WP_008102120.1	-PTGAAAHCTRCGLS.SHYFNI.VFNAE.GVCS.VCD.SYHS.IK.GEASHV.FSD.PS.LKEL.FKASRKR---GNVDCMFV	74
[Candidatus_Thiomargarita_nelsonii]_OAD19576.1	-QPTI.HYCVKGLP.SNYPST.FDQGV.CNCTC.DFETHQAKVQY.FKTKAD.LQAI.LA.QAKATV---GKYDCIAL	74
[Gemmatimonadetes_bacterium]_OLC11591.1	-AAKLAHCTRCGLS.SVNPVT.SYD.AG.VN.CVCR.GFDAY.DKAQY.FK.PEA.LKALVA.EMKART---GNVDCMLV	74
[Gemmatimonadetes_bacterium]_OLC6088.1	-AAKRLTYCTRCGLS.SVNPVT.SYD.AG.VN.CVCR.GLDTY.AKAQY.FKS.PDE.LKALVA.QMKAT.TGGG.YDCILV	74
[Microvira_floculans]_WP_084021052.1	-HRTVEYR.CERCGLS.SNMPGT.FD.EAG.I.CN.CRAFADTY.DKAQY.FKT.PED.LEAL.IA.EMQANR.K---GPYDCIVL	74
[Bradyrhizobium_jicamae]_WP_083519209.1	-PATQVYR.CERCGLS.DLPGT.FD.AAG.VN.CN.CRAFADTY.DKAQY.FKT.SDD.LQV.LI.EA.MQANR.K---GPYDCIVL	74
[Microvira_lotononidis]_WP_009491688.1	-AATQVYR.CERCGLS.SDMPGT.TYD.AG.VN.CN.CQAFDTY.EKAQY.FKT.PDD.LQSL.IV.EMKANR.K---GPYDCIVL	74

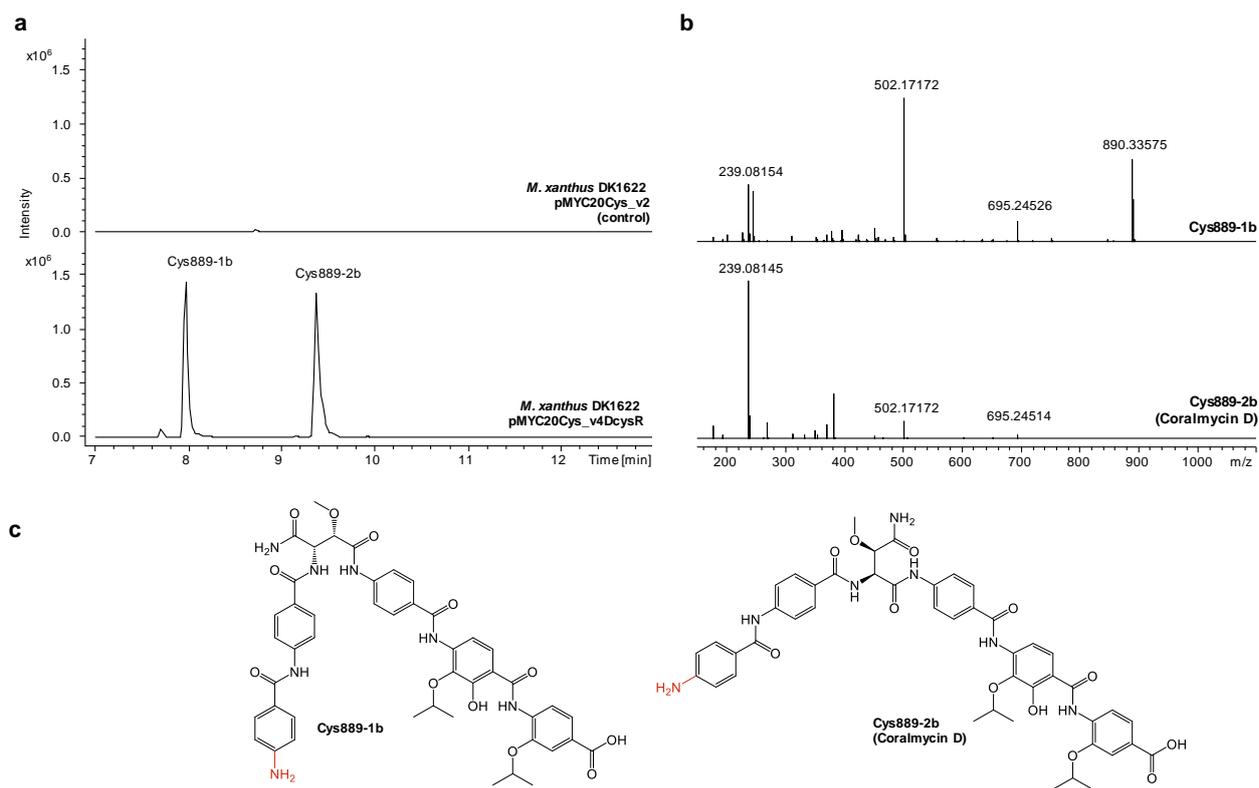


Supplementary Figure 10 | Comparison of the AMDH domain with 25 closest homologs. **a**: Sequence alignment of the AMDH domain with the 25 closest BLAST homologs. **b**: Sequence logo of the seven core motifs (generated using <https://weblogo.berkeley.edu/>). Core motif 1 harbors a ATP-binding motif (SGGK) that was also found in the unusual domain in AlbIV.<sup>8</sup>

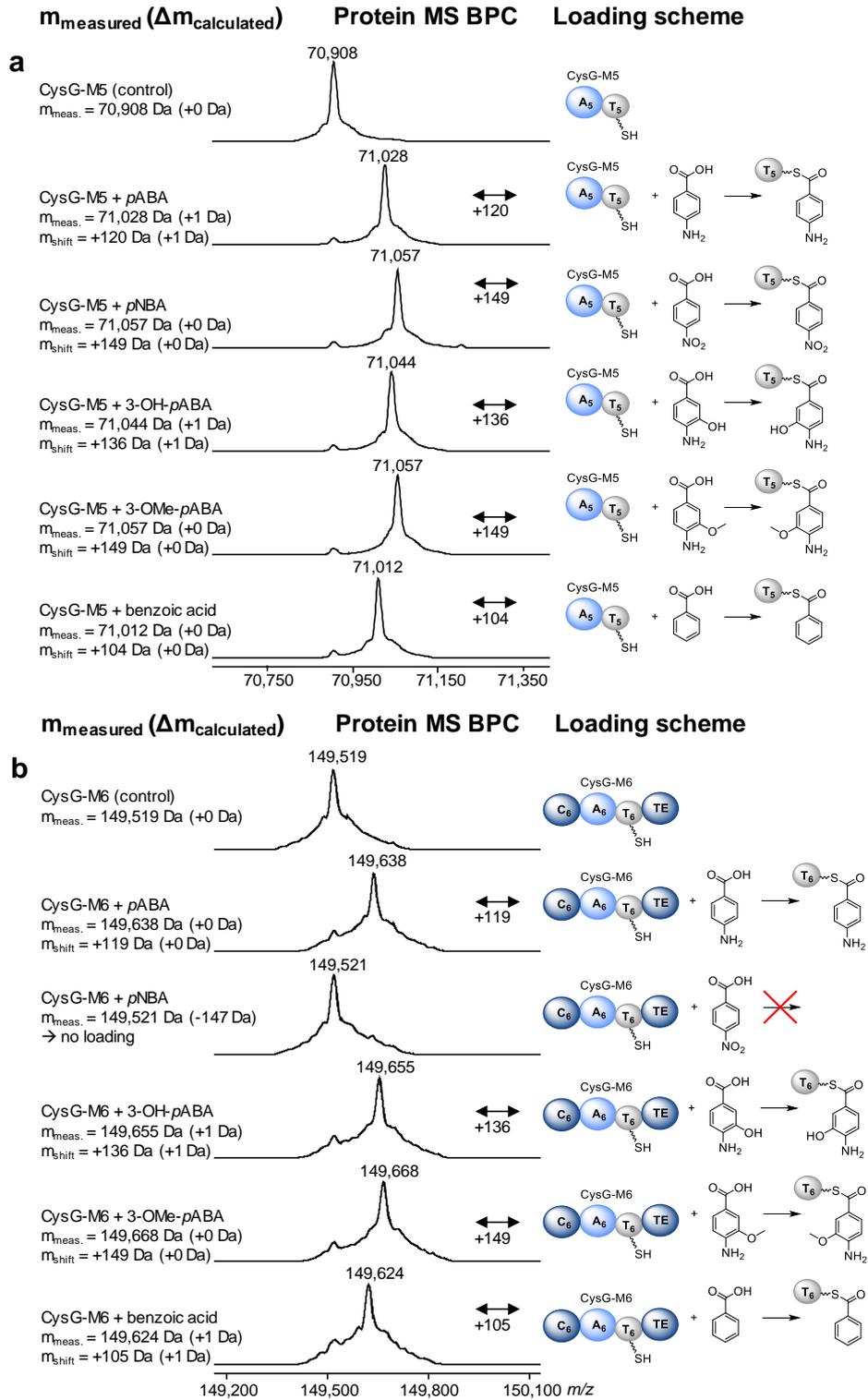




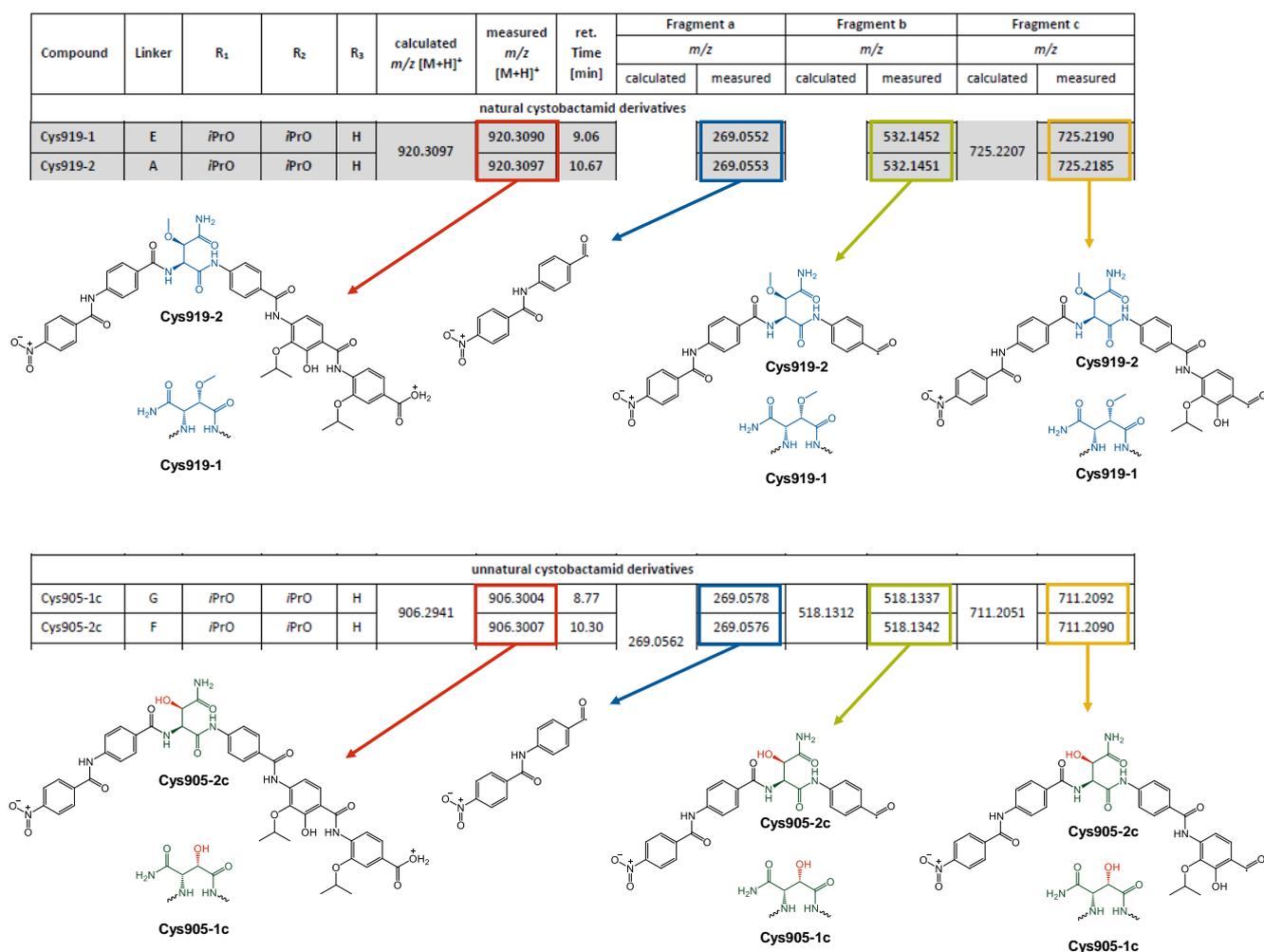
**Supplementary Figure 11 | Substrate specificity of CysK-M1, CysK-M2 and CysK-M4.** Observed mass shifts in deconvoluted protein MS BPCs reveal loading of *p*ABA, 3-OH-*p*ABA, 3-OMe-*p*ABA and benzoic acid but not of *p*NBA by CysK modules 1 (a), 2 (b) and 4 (c), respectively.



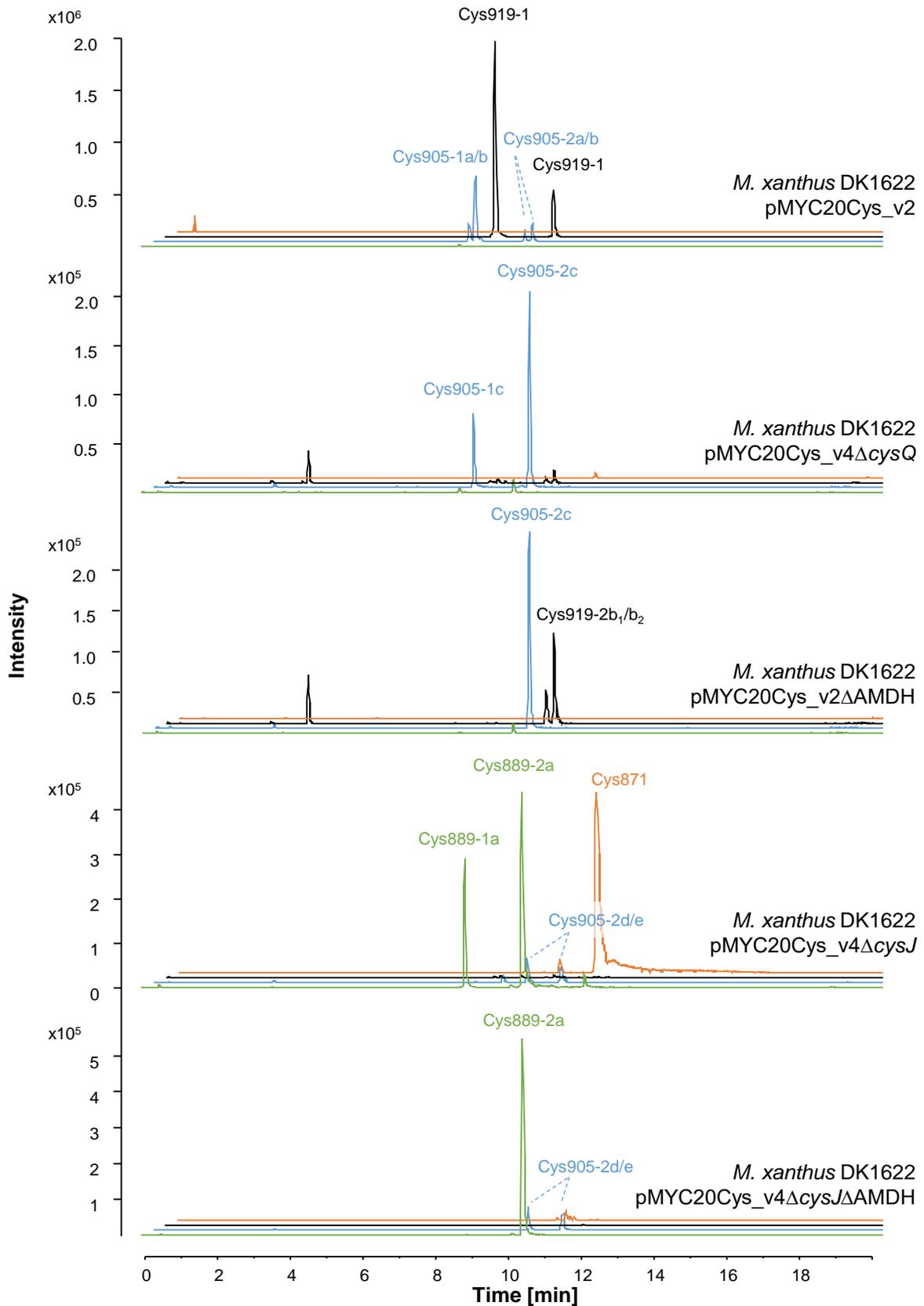
**Supplementary Figure 12 | UPLC-HRMS data and structures of Cys889-1b/2b produced in *M. xanthus* DK1622 pMYC20Cys\_v4ΔcysR. a:** EIC  $m/z$  890.29  $[M+H]^+$ . **b:** MS<sup>2</sup> spectra of Cys889-1b/2b. Fragment with  $m/z$  239 Da  $[M+H]^+$  (30 Da mass shift to natural cystobactamids) confirms that nitro group is replaced by amide group in N-terminal *pABA*. **c:** structure proposal of Cys889-1b and Cys889-2b (Coralmycin D).<sup>10</sup>



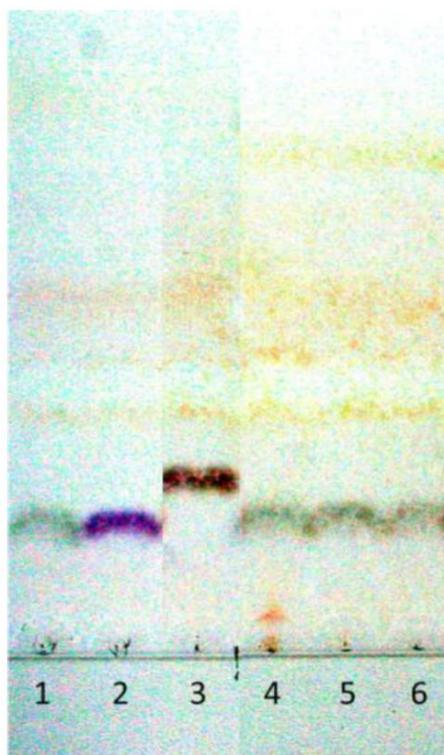
**Supplementary Figure 13 | Substrate specificity of CysG-M5 and CysG-M6.** Observed mass shifts in deconvoluted protein MS BPCs reveal loading of *p*ABA, *p*NBA, 3-OH-*p*ABA, 3-OMe-*p*ABA and benzoic acid by CysG module 5 (a). The same substrate loading, except for *p*NBA, was observed for CysG module 6 (b).



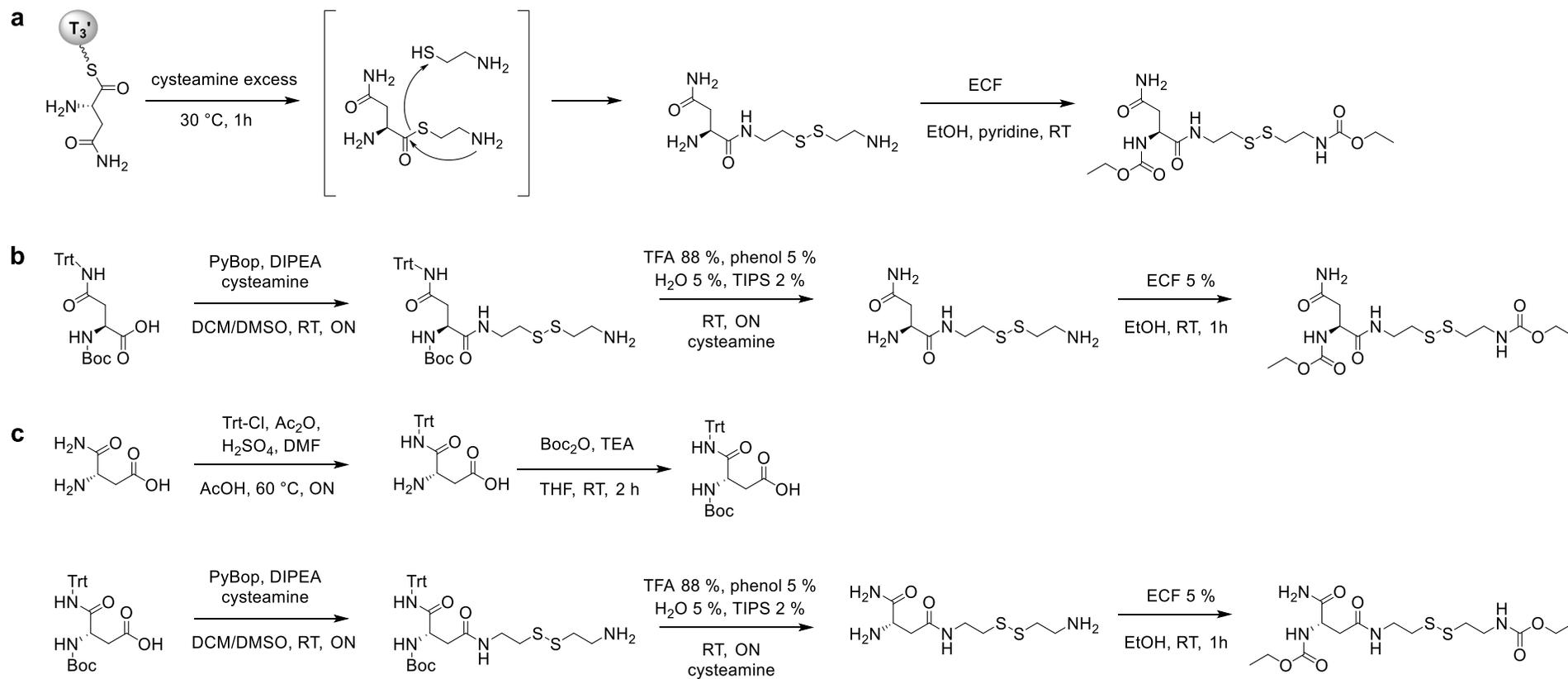
**Supplementary Figure 14 | Example of how structures were assigned to new cystobactamid derivatives.** On top a cropped part of Supplementary Table 10 with the example Cys919-1 and Cys919-2 is shown. The measured high-resolution mass of the molecular ions is shown in the red box. The corresponding structures are shown below. The fragment ion masses are shown in the blue, green and yellow boxes with the corresponding fragment ions below. After deletion of *cysQ*, Cys905-1c and Cys905-2c were produced instead of Cys919-1 and Cys919-2. The measured high-resolution masses of the molecular ion and the fragment ions are shown in the red, blue, green and yellow boxes. The difference of 14 Da in the molecular ion and fragments b and c is due to a loss of a methyl group in the linker, which led to a hydroxyl group (labelled in red) instead of a methoxy group. The linker type was assigned based on the retention time difference, whereas derivatives with L-isoasparagine linkers eluted 1.4–1.6 min earlier than derivatives with L-asparagine linkers.



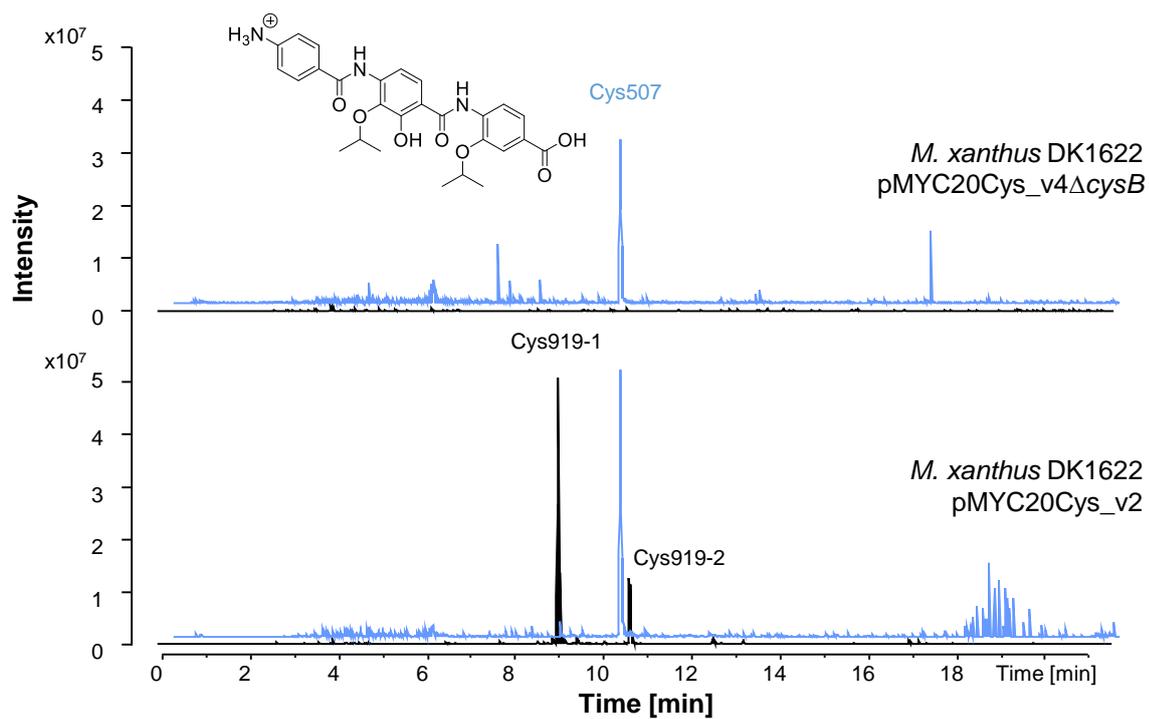
**Supplementary Figure 15 | Full HPLC-ESI-MS EIC traces of methanolic extracts from different heterologous strains.** Stacked/overlaid view of the EICs  $m/z$  920.3 [M+H]<sup>+</sup> (black), 906.3 (blue), 890.3 (green) and 872.3 (orange).



**Supplementary Figure 16 | TLC analysis of CysJ activity on free L-asparagine. 1:** L-asparagine without CysJ (control). **2:** L-aspartate without CysJ (control). **3:** L-isoasparagine without CysJ (control). **4:** L-asparagine with CysJ. **5:** L-asparagine with CysJ and  $\alpha$ -KG. **6:** L-asparagine with denaturated CysJ and  $\alpha$ -KG.



**Supplementary Figure 17 | Derivatization of unloaded L-asparagine and organic synthesis of di(ethylcarbonyl)asparaginyl-dicysteamine references. a:** Derivatization of cysteamine unloaded L-asparagine. **b:** Organic synthesis of di(ethylcarbonyl)asparaginyl-dicysteamine. **c:** Organic synthesis of di(ethylcarbonyl)isoasparaginyl-dicysteamine.



**Supplementary Figure 18 | HPLC-ESI-MS EIC traces of methanolic extracts from *M. xanthus* DK1622 pMYC20Cys\_v4ΔcysB.** Stacked view of the EICs  $m/z$  920.3 [M+H]<sup>+</sup> (black) and 507.3 (blue). The structure of Cys507 is depicted on top.

## 2.8.9 Supplementary Tables

**Supplementary Table 1 | Strains used and generated in this work.**

Bacterial strain	Genotype	Reference
<b>Native cystobactamid producer strains</b>		
<i>C. velatus</i> Cbv34	-	HIPS/MINS
<i>M. fulvus</i> SBMx122	-	HIPS/MINS
<b>Cloning strains</b>		
<i>E. coli</i> DH10β	F, <i>mcrA</i> , Δ( <i>mrr-hsdRMS-mcrBC</i> ), Φ80 <i>lacZ</i> ΔM15, Δ <i>lacX74</i> , <i>recA1</i> , <i>araD139</i> , Δ( <i>ara-leu</i> )7697, <i>galU</i> , <i>galk</i> , <i>rpsL</i> (Str <sup>R</sup> ), <i>endA1</i> , <i>nupG</i> , λ	Invitrogen
<i>E. coli</i> HS996	F, <i>mcrA</i> , Δ( <i>mrr-hsdRMS-mcrBC</i> ), Φ80 <i>lacZ</i> ΔM15, Δ <i>lacX74</i> , <i>recA1</i> , <i>araD139</i> , Δ( <i>ara-leu</i> )7697, <i>galU</i> , <i>galk</i> , <i>rpsL</i> (Str <sup>R</sup> ), <i>endA1</i> , <i>nupG</i> , <i>fhuA</i> ::IS2	Invitrogen
<i>E. coli</i> NEB10β	<i>mcrA</i> , spoT1Δ( <i>mrr-hsdRMS-mcrBC</i> ), Φ80d( <i>lacZ</i> ΔM15) <i>recA1</i> , <i>relA1</i> , Δ <i>lacX74</i> , <i>recA1</i> , <i>araD139</i> , Δ( <i>ara-leu</i> )7697, <i>galK16</i> , <i>galE15</i> , <i>rpsL</i> (Str <sup>R</sup> ), <i>endA1</i> , <i>nupG</i> , <i>fhuA</i>	New England Biolabs
<i>E. coli</i> GB05-red	F, <i>mcrA</i> , Δ( <i>mrr-hsdRMS-mcrBC</i> ), Φ80 <i>lacZ</i> ΔM15, Δ <i>lacX74</i> , <i>recA1</i> , <i>araD139</i> , Δ( <i>ara-leu</i> )7697, <i>galU</i> , <i>galk</i> , <i>rpsL</i> (Str <sup>R</sup> ), <i>endA1</i> , <i>nupG</i> , λ, Δ <i>fhuA</i> , P <sub>BAD</sub> - <i>gbaA</i> Δ <i>ybcC</i> , Δ <i>recE</i> T19	Gene Bridges
<i>S. cerevisiae</i> ATCC4004247	MATa, <i>his3</i> Δ1, <i>leu2</i> Δ0, <i>met15</i> Δ0, <i>ura3</i> Δ0, <i>ydr411c</i> ::KanMX4	ATCC
<i>E. coli</i> DH10β pGH-hPvanABC	<i>E. coli</i> DH10β pGH-hPvanABC, Amp <sup>R</sup>	This work
<i>E. coli</i> DH10β pGH-DEF5-G	<i>E. coli</i> DH10β pGH-DEF5-G, Amp <sup>R</sup>	This work
<i>E. coli</i> DH10β pGH-G	<i>E. coli</i> DH10β pGH-G, Amp <sup>R</sup>	This work
<i>E. coli</i> DH10β pGH-3-GHIJ5-K	<i>E. coli</i> DH10β pGH-3-GHIJ5-K, Amp <sup>R</sup>	This work
<i>E. coli</i> DH10β Dev-3-GHIJ5-K_flong_v2	<i>E. coli</i> DH10β Dev-3-GHIJ5-K_flong_v2, Amp <sup>R</sup>	This work
<i>E. coli</i> DH10β pGH-K1_v2	<i>E. coli</i> DH10β pGH-K1_v2, Amp <sup>R</sup>	This work
<i>E. coli</i> DH10β pGH-K2_v2	<i>E. coli</i> DH10β pGH-K2_v2, Amp <sup>R</sup>	This work
<i>E. coli</i> DH10β pGH-K3_v2	<i>E. coli</i> DH10β pGH-K3_v2, Amp <sup>R</sup>	This work
<i>E. coli</i> DH10β pGH-3-KL	<i>E. coli</i> DH10β pGH-3-KL, Amp <sup>R</sup>	This work
<i>E. coli</i> DH10β pGH-NtD1LEU2h	<i>E. coli</i> DH10β pGH-NtD1LEU2h, Amp <sup>R</sup>	This work
<i>E. coli</i> DH10β pGH-hPvanOPQRS	<i>E. coli</i> DH10β pGH-hPvanOPQRS, Amp <sup>R</sup>	This work
<i>E. coli</i> DH10β pGH-T-ABC15-2	<i>E. coli</i> DH10β pGH-T-ABC15-2, Amp <sup>R</sup>	This work
<i>E. coli</i> DH10β pGH-3-2345tD1	<i>E. coli</i> DH10β pGH-3-2345tD1, Amp <sup>R</sup>	This work
<i>E. coli</i> DH10β pGH-LEU2pc	<i>E. coli</i> DH10β pGH-LEU2pc, Amp <sup>R</sup>	This work
<i>E. coli</i> DH10β pGH-pMYC	<i>E. coli</i> DH10β pGH-pMYC, Amp <sup>R</sup> , Cm <sup>R</sup>	This work
<i>E. coli</i> DH10β pGH-KanR-Mx9	<i>E. coli</i> DH10β pGH-KanR-Mx9, Amp <sup>R</sup> , Kan <sup>R</sup>	This work
<i>E. coli</i> DH10β pGH-TetR-Mx8	<i>E. coli</i> DH10β pGH-TetR-Mx8, Amp <sup>R</sup> , Tet <sup>R</sup>	This work
<i>E. coli</i> DH10β pGH-AmpR-tnp	<i>E. coli</i> DH10β pGH-AmpR-tnp, Amp <sup>R</sup>	This work
<i>E. coli</i> DH10β pJET1.2KDummy_v2	<i>E. coli</i> HS996 pJET1.2KDummy_v2, Amp <sup>R</sup>	This work
<i>E. coli</i> DH10β pSynbio1hPvanABC	<i>E. coli</i> DH10β pSynbio1hPvanABC, Amp <sup>R</sup>	This work
<i>E. coli</i> DH10β pSynbio1hPvanABCDEF5-G (+SE)	<i>E. coli</i> DH10β pSynbio1hPvanABCDEF5-G (+SE), Amp <sup>R</sup>	This work
<i>E. coli</i> DH10β pSynbio1hPvanABCDEF5-G	<i>E. coli</i> DH10β pSynbio1hPvanABCDEF5-G, Amp <sup>R</sup>	This work
<i>E. coli</i> DH10β pGH-GHIJ5-K_v2	<i>E. coli</i> DH10β pGH-GHIJ5-K_v2, Amp <sup>R</sup>	This work
<i>E. coli</i> DH10β pSynbio1K1_v2	<i>E. coli</i> DH10β pSynbio1K1_v2, Amp <sup>R</sup>	This work
<i>E. coli</i> DH10β pSynbio1K12_v2	<i>E. coli</i> DH10β pSynbio1K12_v2, Amp <sup>R</sup>	This work
<i>E. coli</i> DH10β pSynbio1K123_v2 (+SE)	<i>E. coli</i> DH10β pSynbio1K123_v2 (+SE), Amp <sup>R</sup>	This work
<i>E. coli</i> DH10β pSynbio1K123_v2	<i>E. coli</i> DH10β pSynbio1K123_v2, Amp <sup>R</sup>	This work

Bacterial strain	Genotype	Reference
<b>Cloning strains</b>		
<i>E. coli</i> DH10β pSynbio13-KL	<i>E. coli</i> DH10β pSynbio13-KL, Amp <sup>R</sup>	This work
<i>E. coli</i> DH10β pSynbio13-KLNtD1LEU2h (+SE)	<i>E. coli</i> DH10β pSynbio13-KLNtD1LEU2h (+SE), Amp <sup>R</sup>	This work
<i>E. coli</i> DH10β pSynbio13-KLNtD1LEU2h (+SE)	<i>E. coli</i> DH10β pSynbio13-KLNtD1LEU2h (+SE), Amp <sup>R</sup>	This work
<i>E. coli</i> HS996 pSynbio1ABC3-2345tD1	<i>E. coli</i> HS996 pSynbio1ABC3-2345tD1, Amp <sup>R</sup>	This work
<i>E. coli</i> HS996 pSynbio1ABC3-2345tD1LEU2h	<i>E. coli</i> HS996 pSynbio1ABC3-2345tD1LEU2h, Amp <sup>R</sup>	This work
<i>S. cerevisiae</i> ATCC4004247 pMYC20preCysOp1_v2	<i>S. cerevisiae</i> ATCC4004247 pMYC20preCysOp1_v2, <i>LEU2</i>	This work
<i>S. cerevisiae</i> ATCC4004247 pMYC21CysOp2	<i>S. cerevisiae</i> ATCC4004247 pMYC21CysOp2, <i>LEU2</i>	This work
<i>E. coli</i> DH10β pMYC20preCysOp1_v2	<i>E. coli</i> DH10β pMYC20preCysOp1_v2, Cm <sup>R</sup> , Otc <sup>R</sup>	This work
<i>E. coli</i> DH10β pMYC21CysOp2	<i>E. coli</i> DH10β pMYC21CysOp2, Cm <sup>R</sup> , Kan <sup>R</sup>	This work
<i>E. coli</i> DH10β pMYC20CysOp1_v2	<i>E. coli</i> DH10β pMYC20CysOp1_v2, Cm <sup>R</sup> , Otc <sup>R</sup>	This work
<i>E. coli</i> DH10β pMYC20Cys_v2	<i>E. coli</i> DH10β pMYC20Cys_v2, Cm <sup>R</sup> , Otc <sup>R</sup>	This work
<i>E. coli</i> GB05-red pMYC20Cys_v2	<i>E. coli</i> GB2005 pMYC20Cys_v2, Cm <sup>R</sup> , Otc <sup>R</sup>	This work
<i>E. coli</i> NEB10β pMYC20Cys_v2ΔAMDHampR	<i>E. coli</i> NEB10β pMYC20Cys_v2ΔAMDHampR, Cm <sup>R</sup> , Otc <sup>R</sup> , Amp <sup>R</sup>	This work
<i>E. coli</i> NEB10β pMYC20Cys_v2ΔAMDH	<i>E. coli</i> NEB10β pMYC20Cys_v2ΔAMDH, Cm <sup>R</sup> , Otc <sup>R</sup>	This work
<i>E. coli</i> NEB10β pMYC20Cys_v2ΔcysQampR	<i>E. coli</i> NEB10β pMYC20Cys_v2ΔcysQampR, Cm <sup>R</sup> , Otc <sup>R</sup> , Amp <sup>R</sup>	This work
<i>E. coli</i> NEB10β pMYC20Cys_v4ΔcysQ	<i>E. coli</i> NEB10β pMYC20Cys_v4ΔcysQ, Cm <sup>R</sup> , Otc <sup>R</sup>	This work
<i>E. coli</i> NEB10β pMYC20Cys_v2ΔcysJampR	<i>E. coli</i> NEB10β pMYC20Cys_v2ΔcysJampR, Cm <sup>R</sup> , Otc <sup>R</sup> , Amp <sup>R</sup>	This work
<i>E. coli</i> NEB10β pMYC20Cys_v4ΔcysJ	<i>E. coli</i> NEB10β pMYC20Cys_v4ΔcysJ, Cm <sup>R</sup> , Otc <sup>R</sup>	This work
<i>E. coli</i> GB05-red pMYC20Cys_v4ΔcysJ	<i>E. coli</i> GB2005 pMYC20Cys_v4ΔcysJ, Cm <sup>R</sup> , Otc <sup>R</sup>	This work
<i>E. coli</i> NEB10β pMYC20Cys_v4ΔcysJΔAMDHampR	<i>E. coli</i> NEB10β pMYC20Cys_v4ΔcysJΔAMDHampR, Cm <sup>R</sup> , Otc <sup>R</sup> , Amp <sup>R</sup>	This work
<i>E. coli</i> NEB10β pMYC20Cys_v4ΔcysJΔAMDH	<i>E. coli</i> NEB10β pMYC20Cys_v4ΔcysJΔAMDH, Cm <sup>R</sup> , Otc <sup>R</sup>	This work
<i>E. coli</i> NEB10β pMYC20Cys_v2ΔcysBampR	<i>E. coli</i> NEB10β pMYC20Cys_v2ΔcysBampR, Cm <sup>R</sup> , Otc <sup>R</sup> , Amp <sup>R</sup>	This work
<i>E. coli</i> NEB10β pMYC20Cys_v4ΔcysB	<i>E. coli</i> NEB10β pMYC20Cys_v4ΔcysB, Cm <sup>R</sup> , Otc <sup>R</sup>	This work
<i>E. coli</i> NEB10β pMYC20Cys_v2ΔcysRampR	<i>E. coli</i> NEB10β pMYC20Cys_v2ΔcysRampR, Cm <sup>R</sup> , Otc <sup>R</sup> , Amp <sup>R</sup>	This work
<i>E. coli</i> NEB10β pMYC20Cys_v4ΔcysR	<i>E. coli</i> NEB10β pMYC20Cys_v4ΔcysR, Cm <sup>R</sup> , Otc <sup>R</sup>	This work
<i>E. coli</i> NEB10β pMYC20Cys_v2ΔcysCampR	<i>E. coli</i> NEB10β pMYC20Cys_v2ΔcysCampR, Cm <sup>R</sup> , Otc <sup>R</sup> , Amp <sup>R</sup>	This work
<i>E. coli</i> NEB10β pMYC20Cys_v4ΔcysC	<i>E. coli</i> NEB10β pMYC20Cys_v4ΔcysC, Cm <sup>R</sup> , Otc <sup>R</sup>	This work
<i>E. coli</i> DH10β pET-28bcysJ	<i>E. coli</i> DH10β pET-28bcysJ, Kan <sup>R</sup>	This work

Bacterial strain	Genotype	Reference
<b>Cloning strains</b>		
<i>E. coli</i> DH10β pHisSUMOTEV <sub>cysH</sub>	<i>E. coli</i> DH10β pHisSUMOTEV <sub>cysH</sub> , Kan <sup>R</sup>	This work
<i>E. coli</i> DH10β pHisSUMOTEV <sub>cysHΔAMDH</sub>	<i>E. coli</i> DH10β pHisSUMOTEV <sub>cysHΔAMDH</sub> , Kan <sup>R</sup>	This work
<i>E. coli</i> DH10β pHisSUMOTEV <sub>cysB</sub>	<i>E. coli</i> DH10β pHisSUMOTEV <sub>cysB</sub> , Kan <sup>R</sup>	This work
<i>E. coli</i> DH10β pETduet-1 <sub>cysA</sub>	<i>E. coli</i> DH10β pETduet-1 <sub>cysA</sub> , Kan <sup>R</sup>	This work
<i>E. coli</i> DH10β pETduet-1 <sub>cysA</sub> cysK-M1	<i>E. coli</i> DH10β pETduet-1 <sub>cysA-cysK-M1</sub> , Kan <sup>R</sup>	This work
<i>E. coli</i> DH10β pETduet-1 <sub>cysA</sub> cysK-M2	<i>E. coli</i> DH10β pETduet-1 <sub>cysA-cysK-M2</sub> , Kan <sup>R</sup>	This work
<i>E. coli</i> DH10β pETduet-1 <sub>cysA</sub> cysK-M3	<i>E. coli</i> DH10β pETduet-1 <sub>cysA-cysK-M3</sub> , Kan <sup>R</sup>	This work
<i>E. coli</i> DH10β pETduet-1 <sub>cysA</sub> cysK-M4	<i>E. coli</i> DH10β pETduet-1 <sub>cysA-cysK-M4</sub> , Kan <sup>R</sup>	This work
<i>E. coli</i> DH10β pETduet-1 <sub>cysA</sub> cysG-M5	<i>E. coli</i> DH10β pETduet-1 <sub>cysA-cysG-M5</sub> , Kan <sup>R</sup>	This work
<i>E. coli</i> DH10β pETduet-1 <sub>cysA</sub> cysG-M6	<i>E. coli</i> DH10β pETduet-1 <sub>cysA-cysG-M6</sub> , Kan <sup>R</sup>	This work
<b>Heterologous host and production strains</b>		
<i>M. xanthus</i> DK1622	-	HIPS/MINS
<i>M. xanthus</i> DK1622 pMYC20Cys_v2	<i>M. xanthus</i> DK1622 pMYC20Cys_v2, Otc <sup>R</sup>	This work
<i>M. xanthus</i> DK1622 pMYC20Cys_v2ΔAMDH	<i>M. xanthus</i> DK1622 pMYC20Cys_v2ΔAMDH, Otc <sup>R</sup>	This work
<i>M. xanthus</i> DK1622 pMYC20Cys_v4ΔcysQ	<i>M. xanthus</i> DK1622 pMYC20Cys_v4ΔcysQ, Otc <sup>R</sup>	This work
<i>M. xanthus</i> DK1622 pMYC20Cys_v4ΔcysJ	<i>M. xanthus</i> DK1622 pMYC20Cys_v4ΔcysJ, Otc <sup>R</sup>	This work
<i>M. xanthus</i> DK1622 pMYC20Cys_v4ΔcysJΔAMDH	<i>M. xanthus</i> DK1622 pMYC20Cys_v4ΔcysJΔAMDH, Otc <sup>R</sup>	This work
<i>M. xanthus</i> DK1622 pMYC20Cys_v4ΔcysB	<i>M. xanthus</i> DK1622 pMYC20Cys_v4ΔcysB, Otc <sup>R</sup>	This work
<i>M. xanthus</i> DK1622 pMYC20Cys_v4ΔcysC	<i>M. xanthus</i> DK1622 pMYC20Cys_v4ΔcysC, Otc <sup>R</sup>	This work
<i>M. xanthus</i> DK1622 pMYC20Cys_v4ΔcysR	<i>M. xanthus</i> DK1622 pMYC20Cys_v4ΔcysR, Otc <sup>R</sup>	This work
<b>Protein overexpression strains</b>		
<i>E. coli</i> BL21 (DE3) pET-28bcysJ	<i>E. coli</i> DH10β pET-28bcysJ, Kan <sup>R</sup>	This work
<i>E. coli</i> BL21 (DE3) pHisSUMOTEV <sub>cysH</sub>	<i>E. coli</i> DH10β pHisSUMOTEV <sub>cysH</sub> , Kan <sup>R</sup>	This work
<i>E. coli</i> BL21 (DE3) pHisSUMOTEV <sub>cysHΔAMDH</sub>	<i>E. coli</i> DH10β pHisSUMOTEV <sub>cysHΔAMDH</sub> , Kan <sup>R</sup>	This work
<i>E. coli</i> BL21 (DE3) pHisSUMOTEV <sub>cysB</sub>	<i>E. coli</i> DH10β pHisSUMOTEV <sub>cysB</sub> , Kan <sup>R</sup>	This work
<i>E. coli</i> BL21 (DE3) pETduet- 1 <sub>cysA-cysK-M1</sub>	<i>E. coli</i> DH10β pETduet-1 <sub>cysAcysK-M1</sub> , Kan <sup>R</sup>	This work
<i>E. coli</i> BL21 (DE3) pETduet-1 <sub>cysA-cysK-M2</sub>	<i>E. coli</i> DH10β pETduet-1 <sub>cysAcysK-M2</sub> , Kan <sup>R</sup>	This work
<i>E. coli</i> BL21 (DE3) pETduet-1 <sub>cysA</sub> cysK-M3	<i>E. coli</i> DH10β pETduet-1 <sub>cysAcysK-M3</sub> , Kan <sup>R</sup>	This work

Bacterial strain	Genotype	Reference
<b>Protein overexpression strains</b>		
<i>E. coli</i> BL21 (DE3) pETduet-1 <i>cysA-cysK</i> -M4	<i>E. coli</i> DH10 $\beta$ pETduet-1 <i>cysAcysK</i> -M4, Kan <sup>R</sup>	This work
<i>E. coli</i> BL21 (DE3) pETduet-1 <i>cysA-cysG</i> -M5	<i>E. coli</i> DH10 $\beta$ pETduet-1 <i>cysA cysG</i> -M5, Kan <sup>R</sup>	This work
<i>E. coli</i> BL21 (DE3) pETduet-1 <i>cysA-cysG</i> -M6	<i>E. coli</i> DH10 $\beta$ pETduet-1 <i>cysA cysG</i> -M6, Kan <sup>R</sup>	This work

**Supplementary Table 2 | Plasmids used and generated in this work.**

Plasmid	Genotype	Reference
pSynbio1	Non-integrative plasmid for cloning in <i>E. coli</i> ; <i>oriV</i> and <i>trfA</i> from RK2 plasmid, <i>bla</i> (Amp <sup>R</sup> ), MCS	3
pMYC	pMYC vector backbone; <i>p15A</i> ori, <i>cat</i> (Cm <sup>R</sup> ) from pACYC184, <i>traJ</i> , <i>oriT</i> , <i>CEN6/ARS4</i> , <i>URA3</i> , <i>t<sub>D1</sub></i> terminator from <i>Myxococcus</i> phage Mx8	This work
pMYC20	TetR-mx8 cloned into pMYC; <i>tetR</i> (Otc <sup>R</sup> ) from pALTER(R)-1, <i>mx8</i> integrase from <i>Myxococcus</i> phage Mx8	This work
pMYC21	KanR-mx9 cloned into pMYC; <i>kanR</i> (Kan <sup>R</sup> ) from pACYC177, <i>mx9</i> integrase from <i>Myxococcus</i> phage Mx9	This work
pJET1.2	pUC ori and <i>rep</i> from pMB1, P <sub>lacUV5-eco47I</sub> (endonuclease)/T7 promoter-MCS, <i>bla</i> (Amp <sup>R</sup> )	Thermo Fisher Scientific
pET-28b	ColE1 ori from pBR322, T7 promoter-MCS, <i>kanR</i> (Kan <sup>R</sup> ), <i>lacI</i>	Novagen
pHisSUMOTEV	-	11
pETduet-1	ColE1 ori from pBR322, T7 promoter-MCS, <i>bla</i> (Amp <sup>R</sup> ), <i>lacI</i>	Novagen
pSynbio1hPvanABC	Gene synthesis product hPvanABC cloned into pSynbio1; h <i>URA3</i> homology from pRS415, P <sub>van</sub> promoter and <i>vanR</i> repressor gene, <i>cysA</i> , <i>cysB</i> , <i>cysC</i> from native cystobactamid producer strain <i>C. velatus</i> Cbv34	This work
pSynbio1hPvanABCDEF5-G (+SEs)	Gene synthesis product DEF5-G cloned into pSynbio1hPvanABC; <i>cysD</i> , <i>cysE</i> , 5' end of <i>cysG</i> , including splitter elements (SEs) Cbv34	This work
pSynbio1hPvanABCDEF5-G	pSynbio1ABCDEF5-G (+SEs) w/o SEs	This work
pSynbio1K1_v2	Gene synthesis product K1_v2 cloned into pSynbio1; K1 is the first part of <i>cysK</i> from Cbv34	This work
pSynbio1K12_v2	Gene synthesis product K2_v2 cloned into pSynbio1K1_v2; K2 is the second part of <i>cysK</i>	This work
pSynbio1K123_v2 (+SEs)	Gene synthesis product K3_v2 cloned into pSynbio1K12_v2; K3 is the third part of <i>cysK</i> ; including SEs	This work
pSynbio1K123_v2	pSynbio1K123 (+SEs) w/o SEs	This work
pSynbio13-KL	Gene synthesis product 3-KL cloned into pSynbio1; 3' end of <i>cysK</i> , <i>cysL</i> from Cbv34	This work
pSynbio13-KLNtD1LEU2h (+SEs)	Gene synthesis product NtD1LEU2h cloned into pSynbio13-KL; <i>cysN</i> from Cbv34, <i>t<sub>D1</sub></i> terminator, <i>LEU2</i> from pRS415 plasmid, h <i>URA3</i> homology, including SEs	This work
pSynbio13-KLNtD1LEU2h	pSynbio13-KLNtD1LEU2h (+SEs) w/o SEs	This work
pSynbio1ABC3-2345tD1	Gene synthesis product ABC3-2345tD1 cloned into pSynbio1; 3' end of <i>ABC2</i> (ABC transporter permease), <i>ABC3</i> (efflux ABC transporter), <i>ABC4</i> (hypothetical protein), <i>ABC5</i> (nucleoporin) from Cbv34, <i>t<sub>D1</sub></i> terminator	This work
pSynbio1ABC3-2345tD1LEU2h	pSynbio1ABC3-2345tD1, <i>LEU2</i>	This work
pMYC20preCysOp1_v2	pMYC20, hPvanABCDEF5'-G, G ( <i>cysG</i> from Cbv34), 3-GHJ5-K_v2 (3' end of <i>cysG</i> , <i>cysH</i> , <i>cysI</i> , <i>cysJ</i> , 5' end of <i>cysK_v2</i> ), KDummy_v2 ( <i>rpsL</i> and <i>cysK_v2</i> homologies), 3-KLNtD1LEU2h	This work
pMYC20CysOp1_v2	K123_v2 from pSynbio1K123_v2 cloned into pMYC20preCysOp1_v2	This work
pMYC21CysOp2	pMYC21, hPvanOPQRS (h <i>URA3</i> homology, <i>cysP</i> , <i>cysQ</i> , <i>cysS</i> from Cbv34), T-ABC15-2 ( <i>cysT</i> , <i>ABC1</i> (ABC transporter ABC-binding protein) and 5' end of <i>ABC2</i> (permease) from Cbv34)	This work
pMYC20Cys_v2	CysOp2 from pMYC21CysOp2 cloned into pMYC20CysOp1_v2	This work
pMYC20Cys_v2 $\Delta$ AMDHam pR	pMYC20Cys_v2 $\Delta$ AMDH ( <i>cysH</i> ) <i>bla</i>	This work
pMYC20Cys_v2 $\Delta$ AMDH	pMYC20Cys_v2 $\Delta$ AMDH ( <i>cysH</i> )	This work

Plasmid	Genotype	Reference
pMYC20Cys_v2ΔcysQampR	pMYC20Cys_v2ΔcysQ <i>bla</i>	This work
pMYC20Cys_v4ΔcysQ	pMYC20Cys_v2ΔcysQ	This work
pMYC20Cys_v2ΔcysJampR	pMYC20Cys_v2ΔcysJ <i>bla</i>	This work
pMYC20Cys_v4ΔcysJ	pMYC20Cys_v2ΔcysJ	
pMYC20Cys_v4ΔcysJΔAMD DHampR	pMYC20Cys_v2ΔcysJΔAMDH ( <i>cysH</i> ) <i>bla</i>	This work
pMYC20Cys_v4ΔcysJΔAMD DH	pMYC20Cys_v2ΔcysJΔAMDH ( <i>cysH</i> )	This work
pMYC20Cys_v2ΔcysBampR	pMYC20Cys_v2ΔcysB <i>bla</i>	This work
pMYC20Cys_v4ΔcysB	pMYC20Cys_v2ΔcysB	This work
pMYC20Cys_v2ΔcysRampR	pMYC20Cys_v2ΔcysR <i>bla</i>	This work
pMYC20Cys_v4ΔcysR	pMYC20Cys_v2ΔcysR	This work
pMYC20Cys_v2ΔcysCampR	pMYC20Cys_v2ΔcysC <i>bla</i>	This work
pMYC20Cys_v4ΔcysC	pMYC20Cys_v2ΔcysC	This work
pET-28bcysJ	PCR-amplified <i>cysJ</i> from Cbv34 cloned into pET-28b	This work
pHisSUMOTEV <i>cysH</i>	PCR-amplified <i>cysH</i> from Cbv34 cloned into pHisSUMOTEV	This work
pHisSUMOTEV <i>cysH</i> ΔAMD H	PCR-amplified <i>cysH</i> ΔAMDH from pMYC20Cys_v2cysHΔAMDH cloned into pHisSUMOTEV	This work
pHisSUMOTEV <i>cysB</i>	PCR-amplified <i>cysB</i> from Cbv34 cloned into pHisSUMOTEV	This work
pETduet-1 <i>cysA</i>	PCR-amplified <i>cysA</i> from Cbv34 cloned into pETduet-1	This work
pETduet-1 <i>cysA-cysK-M1</i>	PCR-amplified <i>cysK-M1</i> from Cbv34 cloned into pETduet-1 <i>cysA</i>	This work
pETduet-1 <i>cysA-cysK-M2</i>	PCR-amplified <i>cysK-M2</i> from Cbv34 cloned into pETduet-1 <i>cysA</i>	This work
pETduet-1 <i>cysA-cysK-M3</i>	PCR-amplified <i>cysK-M3</i> from Cbv34 cloned into pETduet-1 <i>cysA</i>	This work
pETduet-1 <i>cysA-cysK-M4</i>	PCR-amplified <i>cysK-M4</i> from Cbv34 cloned into pETduet-1 <i>cysA</i>	This work
pETduet-1 <i>cysA-cysG-M5</i>	PCR-amplified <i>cysK-G5</i> from Cbv34 cloned into pETduet-1 <i>cysA</i>	This work
pETduet-1 <i>cysA-cysG-M6</i>	PCR-amplified <i>cysK-G6</i> from Cbv34 cloned into pETduet-1 <i>cysA</i>	This work

**Supplementary Table 3 | Oligonucleotides used in this work.**

Oligonucleotide	Sequence (5'-3')
K-rpsLF	TAGCTACCCGCCCGAGCTGGCGAGGAAGGTGGCGGAGCTCAGCCGGG AGCCGAGACCGGCCTGGTGATGATGGCGGGATCGTTGTAT
K-rpsLR	CGAGCCCCCAGCGTCGCGCACATGAACTCTATGACCTTCTGGTGACGC TGCGAGACCGGCCTTACTTAACGGAGAACCATTAA
K-rpsLF_v2_SDM	GATTAGCTACCCGCCCGCGCTGGCG
K-rpsLR_v2_SDM	CGCCAGCGCGGGCGGGTAGCTAATC
Mx8-attP-up2	CGACGGTGCCGACAAATAC
Mx8-attB-up2	GCGCACTGGACCATCACGTC
Mx8-attP-down	GGCTTGTGCCAGTCAACTGCC
Mx8-attB-down	CGGATAGCTCAGCGGTAGAG
5-PcysH-X-KO	AGCGCCCTTCTCTCCTATCCGGGGATCAAGGAATGCATCGTCGATGTG GCGAGACCGACGAAAGGCCTCGTGATAC
3-PcysH-X-KO	TCCGCGGAGACATAGTAGGCGACCAGCTGGGCTTCCGCCGAGCCACC ACCAGACCTTACCAATGCTTAATCAGTGAG
5-PcysQ-KO	ATCCGCAACGGGCGGCGTTCGCGCTTCGCGAATCCAGCCACGGACGCT GACGAGACCGACGAAAGGCCTCGTGATAC
3-PcysQ-KO	TTGCTGCATTGAAAGGGGAGCGAGCGCCTGCGGGCGCTGGTTCGCGCGC GCTCAGCGAGACCTTACCAATGCTTAATCAGTGAG
5-PcysB-KO	ACATGAGGCCGAAATCGCTACGGGAAGCCCTGACGCGCAGCAACTGCT GACGAGACCGACGAAAGGCCTCGTGATAC
3-PcysB-KO	TGCGTCCGCTCGTCGAGGCCGATGTTGTTGGGAAGTATCATGTTGTCTC CTCAGCGAGACCTTACCAATGCTTAATCAGTGAG
5-PcysC-KO	TCGACACCAGGGCGTCGAGGCTGCCGCTCCTGCGGTGGCGGGAGCGT AGCGAGACCGACGAAAGGCCTCGTGATAC

Oligonucleotide	Sequence (5'-3')
3-PeysC-KO	TTCCACGAACTGCTCGCTCTTCTCCCGTCAACGATCCGACGGCTTCGT TCTACCGAGACCTTACCAATGCTTAATCAGTGAG
5-PeysJ-KO	CGACGACGGCGCGTTTGTCTGAAAGACCTTTGCCAGGGTGTGTACTTTT GACGAGACCGACGAAAGGGCCTCGTGATAC
3-PeysJ-KO	CCGACTGTTTTGCGCAACGGCGAAAGCGACACCACCATCGAGCCCGGC CCTCAACGAGACCTTACCAATGCTTAATCAGTGAG
5-PeysR-KO	GCACCCCGGTGTGCCCCAGGGATACCTGGTGCACGGAGTCAAGCGCT GACGAGACCGACGAAAGGGCCTCGTGATAC
3-PeysR-KO	TAGCTGAAAAATTTACTCTCCGGCACTCTCATGTTCTGGGTCTGCGGG CTCAGCGAGACCTTACCAATGCTTAATCAGTGAG
CysJ for	TATCATATGACCGTAATTTGGATAGCGCGG
CysJ rev	TATGGA-TCCTTACGAGCGCCCTGAGTTCGTTGC
CysH for	TATCCATGGACAATCGAGAGATCGC
CysH rev	TATGGATCCTTATCCCCTGTATGCAGGCG
CysB for	TATCCATGGGTACGCCAGCAGCAGG
CysB rev	TATGGATCCCTACGCTCCCGCCACCGCAG
CysA for	TATCATATGAGCATGAACGGGGACG
CysA rev	TATAGATCTTCAGCAGTTGCTGCGCG
CysK1 for	TATGGATCCGAAAAACCTGTATTTTCAGGGCATGCTGCTGGAGGGAGA GCT
CysK1 rev 1	TATAAGCTTTCACTGCGTCCGCTCGACC
CysK1 rev 1/2	TATAAGCTTTACGCGACAACGTTGGAGAGC
CysK2 for 0	TATGGATCCGAAAAACCTGTATTTTCAGGGCGAGATTCCGCTCTCTAC CTGC
CysK2 for 1/2	TATGGATCCGAAAAACCTGTATTTTCAGGGCGTCGAGCGGACGCAGG
CysK2 rev 1	TATAAGCTTTCACTCGGTGTTCCCGACG
CysK2 rev 1/2	TATAAGCTTTACGACGGAGCGAGGGC
CysK3 for 0	TATGGATCCGAAAAACCTGTATTTTCAGGGCGCGGTGCTCTCGTTTCG
CysK3 for 1/2	TATGGATCCGAAAAACCTGTATTTTCAGGGCAACACCGAGGCGGTGCT
CysK3 rev 1	TATAAGCTTCCGATGGATCGACGACAC
CysK3 rev 1/2	TATAAGCTTTCCGGTCCGCTCCACC
CysK4 for 0	TATGGATCCGAAAAACCTGTATTTTCAGGGCCCGCTGCCTCTGGCGTA
CysK4 for 1/2	TATGGATCCGAAAAACCTGTATTTTCAGGGCGTGGAGCGGACCGGAC
CysK4 rev	TATAAGCTTTACCGAGCCCCCAGC
CysG5 for	TATGGATCTGAAAAACCTGTATTTTCAGGGCATGGCCACCAAATTGTCT GACTTC
CysG5 rev 1	TATAAGCTTTACATGCTGATCAGCCTCTGCG
CysG6 for 0	TATGGATCCGAAAAACCTGTATTTTCAGGGCGCGCTTCCGCTGTGCG
CysG6 rev	TATAAGCTTTACGAAGCTCGCGTCCTC

**Supplementary Table 4 | Genetic elements used in the design of the modified gene cluster.** Nucleotide position numbering in modified cluster refers to CysOp1 and CysOp2 separately. Numbering in brackets refers to full modified cluster (GenBank accession numbers will be supplied upon acceptance of the manuscript). Genetic elements used for TAR assembly are labelled in light blue.

Genetic element	Nucleotide position in modified cluster	Sequence origin (GenBank accession)	Nucleotide position in original sequence
<b>CysOp1</b>			
<i>URA3</i> homology left	0 – 100	pRS416 (U03450)	606 - 705
Pvan (+ <i>vanR</i> )	101 – 1,180	pMR3679 <sup>1</sup>	1862 – 2941
<i>cysA</i> - <i>cysN</i>	1,181 – 36,816	Cbv34 gene cluster (KP836244)	9,687 - 45,325
<i>Xma</i> II	36,817 – 36,822	-	-
tD1 terminator	36,823 – 36,871	<i>Myxococcus xanthus</i> phage Mx8 <sup>2</sup>	-
<i>Xba</i> I, spacer, <i>Afl</i> III	36,872 – 36,889	-	-
<i>LEU2</i>	36,890 – 39,124	pRS415 (U03449)	3498 - 5732
<i>URA3</i> homology right	39,125 – 39,224	pRS416 (U03450)	506 - 605
<b>CysOp2</b>			
<i>URA3</i> homology left	0 – 100	pRS416 (U03450)	606 - 705
<i>Xba</i> I	101 – 106 (36,872 – 36,889)	-	-
Pvan	107 – 236 (36,878 – 37,007)	pMR3679 <sup>1</sup>	2,812 – 2,941
<i>cysO</i> - <i>cysT</i>	237 – 9,583 (37,008 – 46,354)	Cbv34 gene cluster (KP836244)	9,347 - 0
<i>Orf5</i> – <i>Orf1</i>	9,584 – 16,171 (46,355 – 52,942)	Cbv34 gene cluster (KP836244)	52,049 - 45,462
tD1 terminator	16,172 – 16,220 (52,943 – 52,991)	<i>Myxococcus xanthus</i> phage Mx8 <sup>2</sup>	-
<i>Afl</i> III	16,221 – 16,226	-	-
<i>LEU2</i>	16,227 – 18,461	pRS415 (U03449)	3498 - 5732
<i>URA3</i> homology right	18,462 – 18,561	pRS416 (U03450)	506 - 605

**Supplementary Table 5 | Removal of restriction sites (R-sites) by point mutations from the modified gene cluster.** Recognition sequence of restriction endonucleases are underlined. Silent point mutations (labelled red) were made to remove R-sites inside coding sequences. Random point mutations were made to remove R-sites in intergenic regions. Nucleotide position numbering refers to original sequence. Numbering in brackets refers to revised gene cluster sequence. Sequences which do not contain blanks are non-coding sequences. Removal of R-site in genetic elements used for TAR cloning are labelled in light blue.

R-site removed	R-site position in original sequence	Sequence before	Sequence after
<b>Pvan (+vanR) (from pMR3679<sup>1</sup>)</b>			
<i>MreI</i>	1,900 – 1,907	<u>CGC GCC GGC GCT</u>	CGC GCC <b>CGC</b> GCT
<i>BsaI</i>	1,998 – 2,003	<u>CGA GAC CGC</u>	CGA GAC <b>GCC</b>
<i>BsaI</i>	2,282 – 2,287	<u>GGT CTC</u>	<b>CGT</b> CTC
<b>Chv34 gene cluster (</b>			
<i>BsaI</i>	2,390 - 2,395	<u>GGA GAC CGT</u>	GGA GAC <b>GGT</b>
<i>BsaI</i>	4,903 - 4,908	<u>GGT CTC</u>	<b>CGT</b> CTC
<i>BsaI</i>	5,057 - 5,062	<u>CGA GAC CGG</u>	CGA GAC <b>GGG</b>
<i>BsaI</i>	6,047 - 6,052	<u>GGTCTC</u>	GGTCT <b>G</b>
<i>BsaI</i>	6,769 - 6,774	<u>GGT CTC</u>	<b>CGT</b> CTC
<i>BsaI</i>	7,872 - 7,877	<u>GGT CTC</u>	<b>CGT</b> CTC
<i>BsaI</i>	8,793 - 8,798	<u>GGT CTC</u>	<b>CGT</b> CTC
<i>BsaI</i>	13,177 - 13,182	<u>GAG ACC</u>	GAG <b>ACG</b>
<i>BsaI</i>	13,618 - 13,623	<u>GAG ACC</u>	GAG <b>ACG</b>
<i>BsaI</i>	14,650 - 14,655	<u>GAGACC</u>	GAG <b>ACG</b>
<i>BsaI</i>	16,384 - 16,389	<u>GAG ACC</u>	GAG <b>ACG</b>
<i>BsaI</i>	17,038 - 17,043	<u>GAG ACC</u>	GAG <b>ACG</b>
<i>BsaI</i>	18,364 - 18,369	<u>GAG ACC</u>	GAG <b>ACG</b>
<i>BsaI</i>	18,628 - 18,233	<u>GAG ACC</u>	GAG <b>ACG</b>
<i>BsaI</i>	20,418 - 20,423	<u>GCG GTC TCG</u>	GCG GTC <b>AGC</b>
<i>BsaI</i>	20,674 - 20,679	<u>GAG ACC</u>	GAG <b>ACG</b>
<i>BsaI</i>	21,829 - 21,834	<u>GAG ACC</u>	GAG <b>ACG</b>
<i>BsaI</i>	22,885 - 22,890	<u>CTG GTC TCA</u>	CTG GTC <b>AGT</b>
<i>BsaI</i>	23,090 - 23,095	<u>GAG ACC</u>	GAG <b>ACG</b>
<i>BsaI</i>	23,573 - 23,578	<u>GAG ACC</u>	GAG <b>ACG</b>
<i>BsaI</i>	26,441 - 26,446	<u>TGG GTC TCA</u>	TGG GTC <b>AGT</b>
<i>BsaI</i>	27,176 - 27,181	<u>AAG GTC TCG</u>	AAG <b>GTG</b> TCC
<i>BsaI</i>	33,118 - 33,123 (33,115 – 33,120)	<u>GTG GTC TCA</u>	<b>GTC</b> <b>GTG</b> TCA
<i>BsaI</i>	41,424 - 41,429 (41,421 – 41,426)	<u>CTG GTC TCT</u>	<b>CTC</b> <b>GTG</b> TCT
<i>BsaI</i>	42,270 - 42,275 (42,267 - 42,272)	<u>CTG AGA CCG</u>	<b>CTC</b> <b>CGA</b> CCG
<i>BsaI</i>	44,504 - 44,509 (44,501 - 44,506)	<u>GAG ACC</u>	GAG <b>ACG</b>
<i>BsaI</i>	46,013 - 46,018 (44,501 - 44,506)	<u>GGA GAC CTG</u>	GG <b>T</b> GAC <u>CTG</u>
<i>BsaI</i>	49,164 - 49,169 (49,164 - 49,169)	<u>CGG GTC TCG</u>	<b>CGC</b> GTC TCG
<i>NdeI</i>	9,345 – 9,350	<u>CATATG</u>	CAT <b>CGT</b>
<b>LEU2 (U03449)</b>			
<i>KspAI</i>	5,483 – 5,488	<u>GTTAAC</u>	GTT <b>AGC</b>
<i>AflIII</i>	4,749 – 4,754	<u>TCT TAA GTT</u>	TCT <b>CAA</b> GTT

**Supplementary Table 6 | Name, size and restriction sites or splitter elements (SEs) of the thirteen gene synthesis cluster fragments.** Spacer sequence (sp) was introduced between SE restriction sites.

Cluster fragment	Description	Size [bp]	Flanking restriction sites or SEs
hPvanABC	<i>URA3</i> homology, <i>vanR</i> , Pvan, <i>cysA</i> , <i>cysB</i> , <i>cysC</i>	3,878	SE1 (5'): <i>KpnI</i> -sp(CGTTAA)- <i>NheI</i> - <i>BsaI</i> SE2 (3'): <i>BsaI</i> - <i>AflIII</i> -sp(GATTGC)- <i>PmeI</i>
DEF5-G	<i>cysD</i> , <i>cysE</i> , <i>cysF</i> , (5') 150 bp of <i>cysG</i>	4,413	SE2 (5'): <i>AflIII</i> - <i>BsaI</i> SE3 (3'): <i>BsaI</i> - <i>AvrII</i> -sp(AGCCTA)- <i>PmeI</i> -3'
G	<i>cysG</i>	5,872	5'- <i>BsaI</i> 3'- <i>BsaI</i>
3-GHIJ5-K	(3') 173 bp of <i>cysG</i> , <i>cysH</i> , <i>cysI</i> , <i>cysJ</i> , (5') 250 bp of <i>cysK</i>	5,330	5'- <i>BsaI</i> 3'- <i>BsaI</i>
3-GHIJ5-K_flong_v2	(3') 1,222 bp of 3-GHIJ5-K fragment, 1,796 bp of pGH cloning vector backbone	3,018	-
K1_v2	<i>cysK</i>	2,542	SE4 (5'): <i>KpnI</i> -sp(GTTACG)- <i>PacI</i> - <i>BsaI</i> SE5 (3'): <i>BsaI</i> - <i>HindIII</i> -sp(GACCTA)- <i>PmeI</i>
K2_v2	<i>cysK</i>	5,499	SE5 (5'): <i>KpnI</i> -sp(CCAGCT)- <i>HindIII</i> - <i>BsaI</i> SE6 (3'): <i>BsaI</i> - <i>AseI</i> -sp(AGCCAT)- <i>PmeI</i>
K3_v2	<i>cysK</i>	5,563	SE6 (5'): <i>KpnI</i> -sp(TATCCG)- <i>AseI</i> - <i>BsaI</i> SE7 (3'): <i>BsaI</i> - <i>EcoRI</i> -sp(AGCCAT)- <i>PmeI</i>
3-KL	(3') 150 bp of <i>cysK</i> , <i>cysL</i>	3,322	SE8 (5'): <i>KpnI</i> -sp(AGGCGT)- <i>MreI</i> - <i>BsaI</i> SE9 (3'): <i>BsaI</i> - <i>SpeI</i> -sp(GACTCC)- <i>PmeI</i>
NtD1LEU2h	<i>cysN</i> , tD1, <i>LEU2</i> , <i>URA3</i> homology	3,602	SE9 (5'): <i>SpeI</i> - <i>BsaI</i> SE10 (3'): <i>BsaI</i> - <i>NotI</i> -sp(GCAGTC)- <i>PmeI</i>
hPvanOPQRS	<i>URA3</i> homology, Pvan, <i>cysO</i> , <i>cysP</i> , <i>cysQ</i> , <i>cysR</i> , <i>cysS</i>	5,687	5'- <i>BsaI</i> 3'- <i>BsaI</i>
T-ABC15-2	<i>cysT</i> , <i>ABC1</i> ( <i>Orf5</i> ), (5') 827 bp of <i>ABC2</i> ( <i>Orf4</i> )	5,666	5'- <i>BsaI</i> 3'- <i>BsaI</i>
ABC3-2345tD1	(3') 1379 bp of <i>ABC2</i> ( <i>Orf4</i> ), <i>ABC3</i> ( <i>Orf3</i> ), <i>ABC4</i> ( <i>Orf2</i> ), <i>ABC5</i> ( <i>Orf1</i> ), tD1	5,139	SE11 (5'): <i>SwaI</i> - <i>BsaI</i> SE12 (3'): <i>AflIII</i> - <sup>^</sup> GTTTT <sup>^</sup> C- <i>BsaI</i> - <i>PmeI</i>
LEU2pc	<i>URA3</i> homology, <i>LEU2</i> , <i>URA3</i> homology	2,457	5'- <i>H<sub>L</sub></i> - <i>AflIII</i> 3'- <i>H<sub>R</sub></i> - <sup>^</sup> GTTTT <sup>^</sup> C- <i>BsaI</i> -3'

Supplementary Table 7 | Genetic elements used in the design of pMYC vector system.

Genetic element	Nucleotide position in pMYC building block	Sequence origin (GenBank accession)	Nucleotide position in original sequence
<b>basic pMYC</b>			
<i>KspAI</i> , spacer, <i>NheI</i>	0 – 18	-	-
<i>URA3</i>	19 – 1,130	pRS416(U03450) <sup>1</sup>	187 – 1,298
<i>ApaLI</i> , spacer, <i>XhoI</i>	1,131 – 1.148	-	-
p15A ori	1,149 – 1,981	pACYC177 (X06402) <sup>12</sup>	766 - 1,598
<i>cmR</i> ( <i>cat</i> )	1,982 – 3,041	pACYC184 (X06403) <sup>13</sup>	3,605 - 419
IR2	3,042 – 3,090	pFNLTP16 H3 (DQ236098)	4,277 - 4,325
<i>oriT/traJ</i>	3,091 – 3,781	<i>Pseudomonas aeruginosa</i> plasmid Birmingham IncP-alpha (L27758)	50,687 - 51,377
<i>NotI</i> , spacer, <i>SnaBI</i>	3,782 – 3,801	-	-
<i>CEN6/ARS4</i>	3,802 - 4,570	pRS415 (U03449)	2,729 - 3,497
<i>MssI</i> , spacer, <i>PacI</i> , <i>SmaI</i> , spacer, <i>MreI</i>	4,571 – 4,612	-	-
tD2 terminator	4,613 - 4,662	<i>Myxococcus xanthus</i> bacteriophage Mx8 <sup>2</sup>	-
<b>Mx8-tetR</b>			
<i>MssI</i> , spacer, <i>PacI</i>	0 – 22	-	-
<i>mx8</i> integrase	23 - 1,853	Bacteriophage Mx8 imm (BMU64984)	4,979 - 6,809
<i>Bst1107I</i> , spacer, <i>SpeI</i>	1,854 - 1,871	-	-
<i>tetR</i>	1,872 – 3,250	pALTER1(R) (X65334)	451 - 1,829
<i>SmaI</i> , spacer, <i>MreI</i>	3,251 – 3,270	-	-
<b>Mx9-kanR</b>			
<i>MssI</i> , spacer, <i>PacI</i>	0 – 22	-	-
<i>mx9</i> integrase	23 - 1,851	<i>mx9</i> sequence (AY247757)	430 – 2,255
<i>Bst1107I</i> , spacer, <i>SpeI</i>	1,852 – 1,869	-	-
<i>kanR</i> ( <i>aph(3')-Ia</i> )	1,870 – 2,849	pACYC177 (X06402) <sup>12</sup>	1,816 - 2,797
<i>SmaI</i> , spacer, <i>MreI</i>	2,850 – 2,869	-	-

**Supplementary Table 8 | Removal of restriction sites (R-sites) by point mutations from pMYC building blocks.** Recognition sequence of restriction endonucleases are underlined. Silent point mutations (labelled red) were made to remove R-sites inside coding sequences. Random point mutations were made to remove R-sites in intergenic regions. Nucleotide position numbering refers to original sequence. Sequences which do not contain blanks are non-coding sequences.

R-site removed	R-site position in original sequence	Sequence before	Sequence after
<b>p15A (X06402)</b>			
<i>Bst</i> 11071	1,577 - 1582	<u>GTATAC</u>	GCATAC
<i>Nhe</i> I	1,588 - 1,593	<u>GCTAGC</u>	GCTACC
<b><i>oriT/traJ</i> (L27758)</b>			
<i>Mre</i> I	51,117 - 51,124	<u>CGCCGGCG</u>	CCCCGGCG
<b>CEN6/ARS4 (U03449)</b>			
<i>Apa</i> LI	3,488 - 3,493	<u>GTGCAC</u>	GTGGAC
<b>URA3 (U03450)</b>			
<i>Nde</i> I	329 - 334	<u>CATATG</u>	CATATC
<i>Bst</i> 11071	741 - 746	<u>GTA TAC</u>	GTC TAC
<i>Bsa</i> I	1,085 - 1,090	GTG <u>GTC TCT</u>	GTT GTC TCT
<i>Bst</i> 11071	1,246 - 1,251	<u>GTA TAC</u>	GTC TAC
<b><i>tnp</i> (DQ236098)</b>			
<i>Ksp</i> AI	981 - 986	AGT TAA <u>CAG</u>	AGT TAA CAG
<i>Eco</i> RV	1,119 - 1,124	TGA TAT <u>CTT</u>	ACT TAT CTT
<i>Nde</i> I	1,409 - 1,414	<u>CATATG</u>	CATATC
<b><i>ampR</i> (<i>bla</i>) (X06402)</b>			
<i>Apa</i> LI	8,342 - 8,347	GGT GCA <u>CGA</u>	GGT GCC CGA
<i>Bsa</i> I	8,942 - 8,947	GGG TCT <u>CGC</u>	GGG TCA CGC
<b><i>mx8</i> (BMU64984)</b>			
<i>Bsa</i> I	5,937 - 5,942	<u>GAG ACC</u>	GAG ACG
<i>Not</i> I	6,711 - 6,718	<u>GCG GCC GCC</u>	GCC GCG GCC
<b><i>tetR</i> (X65334)</b>			
<i>Eco</i> RV	640 - 645	<u>GAT ATC</u>	GAC ATT
<i>Nhe</i> I	684 - 689	CTG CTA <u>GCG</u>	CTG CTT GCG
<b><i>mx9</i> (AY247757)</b>			
<i>Nde</i> I	985 - 990	<u>CAT ATG</u>	CAC ATG
<i>Bst</i> 11071	2,057 - 2,062	GGT ATA <u>CCG</u>	GGG ATA CCG
<i>Mre</i> I	2,139 - 2,146	GCC GCC GGC <u>GTC</u>	GCC GCC GGT GTC
<b><i>kanR</i> (<i>aph</i>(3')-Ia) (X06402)</b>			
<i>Xho</i> I	1,952 - 1,957	TGC TCG <u>AGG</u>	TGC TCA AGG
<i>Sma</i> I	2,226 - 2,231	<u>CCC GGG</u>	CCA GGG

**Supplementary Table 9 | Source of vector and insert DNA and restriction endonucleases used for construction of plasmids in this work.** <sup>[a]</sup> SEs were removed afterwards during desplitting with *BsaI*, <sup>[b]</sup> blunt end ligation since pSynbio1 was linearized using only *PmeI*.

Product generated	Vector	Insert	Insert source	Restriction enzymes
pSynbio1hPvanABC	pSynbio1	hPvanABC	Gene synthesis fragment	<i>KpnI/PmeI</i>
pSynbio1hPvanABCDEF5-G (+SE)	pSynbio1hPvanABC	DEF5-G	Gene synthesis fragment	<i>AflIII/PmeI</i>
pSynbio1hPvanABCDEF5-G	pSynbio1hPvanABCDEF5-G (+SE)	-	-	<i>BsaI</i>
pGH-GHIJ5-K_v2	pGH-3-GHIJ5-K	3-GHIJ5-K_flong_v2	Gene synthesis fragment	<i>XhoI/ScaI</i>
pSynbio1K1_v2	pSynbio1	K1	Gene synthesis fragment	<i>KpnI/PmeI</i>
pSynbio1K12_v2	pSynbio1K1	K2	Gene synthesis fragment	<i>HindIII/PmeI</i>
pSynbio1K123_v2 (+SE) <sup>[a]</sup>	pSynbio1K12	K3	Gene synthesis fragment	<i>AseI/PmeI</i>
pSynbio1K123_v2	pSynbio1K123 (+SE)	-	-	<i>BsaI</i>
pSynbio13-KL	pSynbio1	3-KL	Gene synthesis fragment	<i>KpnI/PmeI</i>
pSynbio13-KLNtD1LEU2h (+SE) <sup>[a]</sup>	pSynbio13-KL	NtD1LEU2h	Gene synthesis fragment	<i>SpeI/PmeI</i>
pSynbio13-KLNtD1LEU2h	pSynbio13-KLNtD1LEU2h (+SE)	-	-	<i>BsaI</i>
pSynbio1ABC3-2345tD1	pSynbio1	ABC3-2345tD1	Gene synthesis fragment	<i>SwaI/PmeI</i> <sup>[b]</sup>
pSynbio1ABC3-2345tD1LEU2h	pSynbio1ABC3-2345tD1	LEU2pc	Gene synthesis fragment	<i>AflIII/BsaI</i>
pMYC20	pMYC	TetR-mx8	Gene synthesis fragment	<i>SmaI/PacI</i>
pMYC21	pMYC	KanR-mx9	Gene synthesis fragment	<i>SmaI/PacI</i>
pMYC20CysOp1_v2	pMYC20preCysOp1	K123_v2	pSynbio1K123_v2	<i>BsaI</i>
pMYC20Cys_v2	pMYC20CysOp1_v2	CysOp2	pMYC21CysOp2	<i>XbaI/AflIII</i>

**Supplementary Table 10 | UPLC-HRMS and MS<sup>2</sup> data of natural and unnatural cystobactamids produced in the heterologous *M. xanthus* DK1622 strains.** Linker and R<sub>1</sub>, R<sub>2</sub>, R<sub>3</sub> classification shown in Figure 1. Fragmentation pattern shown in Supplementary Figure 6. BPC and EIC of minor and major natural cystobactamid derivatives shown in Supplementary Figure 8. Previously described derivatives are marked in grey.

Compound	Linker	R <sub>1</sub>	R <sub>2</sub>	R <sub>3</sub>	calculated m/z [M+H] <sup>+</sup>	measured m/z [M+H] <sup>+</sup>	ret. Time [min]	Fragment a		Fragment b		Fragment c	
								m/z		m/z		m/z	
								calculated	measured	calculated	measured	calculated	measured
<b>natural cystobactamid derivatives</b>													
Cys449	-	<i>i</i> PrO	H	H	450.1660	450.1674	9.15	-	-	-	-	-	-
Cys507	-	<i>i</i> PrO	<i>i</i> PrO	H	508.2078	508.2067	10.09	-	-	-	-	-	-
Cys871	B	<i>i</i> PrO	<i>i</i> PrO	H	872.2886	872.2926	11.43	269.0562	269.0569	484.1257	484.1270	677.1996	677.2012
Cys861-1	E	<i>i</i> PrO	H	H	862.2679	862.2659	8.58		269.0551	532.1468	532.1452	725.2207	725.2178
Cys861-2	A	<i>i</i> PrO	H	H		862.2667	10.01		269.0552		532.1442		725.2185
Cys877-1	E	EtO	MeO	H	878.2628	878.2615	8.12		269.0553		532.1448	711.2051	711.2027
Cys877-2	A	EtO	MeO	H		878.2616	9.58		269.0554		532.1454		711.2017
Cys891-1a	E	EtO	EtO	H	892.2784	892.2773	8.41		269.0555		532.1451	711.2031	
Cys891-1b	E	<i>i</i> PrO	MeO	H		892.2771	8.54		269.0551		532.1450	725.2207	725.2183
Cys891-2a	A	EtO	EtO	H		892.2765	9.90		269.0553		532.1446	711.2051	711.2019
Cys891-2b	A	<i>i</i> PrO	MeO	H		892.2768	10.05		269.0551		532.1446	725.2207	725.2177
Cys905-1a	E	<i>i</i> PrO	EtO	H		906.2941	906.2935		8.65		269.0556	532.1453	725.2192
Cys905-1b	E	EtO	<i>i</i> PrO	H	906.2943		8.83		269.0555		532.1455	711.2051	711.2033
Cys905-2a	A	EtO	<i>i</i> PrO	H	906.2923		10.18		269.0551		532.1439	711.2028	
Cys905-2b	A	<i>i</i> PrO	EtO	H	906.2932		10.40		269.0553		532.1451	725.2182	
Cys919-1	E	<i>i</i> PrO	<i>i</i> PrO	H	920.3097		920.3090		9.06		269.0552	532.1452	725.2190
Cys919-2	A	<i>i</i> PrO	<i>i</i> PrO	H		920.3097	10.67		269.0553		532.1451	725.2207	725.2185
Cys933-1a	E	<i>i</i> PrO	1- MePrO	H		934.3254	934.3233		9.20		269.0549	532.1450	725.2180
Cys933-1b	E	1-MePrO	<i>i</i> PrO	H	934.3241		9.25		269.0554		532.1453	739.2364	739.2350
Cys933-2b <sub>1</sub>	A	<i>i</i> PrO	1- MePrO	H	934.3237		9.42	269.0555	532.1454		725.2207	725.2171	
Cys933-2b <sub>2</sub>	A	1-MePrO	<i>i</i> PrO	H	934.3236		9.47	269.0553	532.1445	739.2364	739.2328		

Compound	Linker	R <sub>1</sub>	R <sub>2</sub>	R <sub>3</sub>	calculated m/z [M+H] <sup>+</sup>	measured m/z [M+H] <sup>+</sup>	ret. Time [min]	Fragment a		Fragment b		Fragment c	
								m/z		m/z		m/z	
								calculated	measured	calculated	measured	calculated	measured
<b>natural cystobactamid derivatives</b>													
Cys935-1	E	<i>i</i> PrO	<i>i</i> PrO	OH	936.3046	936.3037	9.58	269.0562	269.0558	532.1468	532.1436	725.2207	725.2200
Cys935-2	A	<i>i</i> PrO	<i>i</i> PrO	OH		936.3038	11.38		269.0552		532.1446		725.2182
<b>unnatural cystobactamid derivatives</b>													
Cys905-1c	G	<i>i</i> PrO	<i>i</i> PrO	H	906.2941	906.3004	8.77	269.0562	269.0578	518.1312	518.1337	711.2051	711.2092
Cys905-2c	F	<i>i</i> PrO	<i>i</i> PrO	H		906.3007	10.30		269.0576		518.1342		711.2090
Cys905-2d	?	<i>i</i> PrO	<i>i</i> PrO	H		906.2928	10.24		269.0555		518.1298		711.2025
Cys905-2e	?	<i>i</i> PrO	<i>i</i> PrO	H		906.2917	11.15						
Cys919-2b <sub>1</sub>	A	<i>i</i> PrO	1- MePrO	H	920.3058	920.3083	10.45		269.0555	518.1310	711.2025		
Cys919-2b <sub>2</sub>	A	<i>i</i> PrO	1- MePrO	H	920.3058	920.3086	10.67		269.0557	518.1294	711.2036		
Cys889-1a	I	<i>i</i> PrO	<i>i</i> PrO	H	890.2992	890.3031	8.88		269.0565	502.1363	502.1371	695.2102	695.2116
Cys889-2a	H	<i>i</i> PrO	<i>i</i> PrO	H		890.3033	10.44		269.0565	502.1374	695.2113		
Cys889-1b	E	<i>i</i> PrO	<i>i</i> PrO	H	890.3355	890.3396	7.98	239.0823	502.1727	502.1739	695.2466	695.2483	
Cys889-2b	A	<i>i</i> PrO	<i>i</i> PrO	H		890.3386	9.38	239.0819	502.1734	695.2474			

**Supplementary Table 11 | Manipulated plasmids generated via Red/ET recombineering and restriction hydrolysis/re-ligation.**

Product generated	Description	Vector used for Red/ET	Red/ET product
pMYC20Cys_v2ΔAMDH	Deletion of X domain in <i>cysH</i>	pMYC20Cys_v2	pMYC20Cys_v2ΔAMDHampR
pMYC20Cys_v4ΔcysQ	Deletion of <i>cysQ</i>	pMYC20Cys_v2	pMYC20Cys_v2ΔcysQampR
pMYC20Cys_v4ΔcysJ	Deletion of <i>cysJ</i>	pMYC20Cys_v2	pMYC20Cys_v2ΔcysJampR
pMYC20Cys_v4ΔcysJΔAMDH	Deletion of <i>cysJ</i> and X domain in <i>cysH</i>	pMYC20Cys_v4ΔcysJ	pMYC20Cys_v4ΔcysJΔAMDHampR
pMYC20Cys_v4ΔcysB	Deletion of <i>cysB</i>	pMYC20Cys_v2	pMYC20Cys_v2ΔcysBampR
pMYC20Cys_v4ΔcysR	Deletion of <i>cysR</i>	pMYC20Cys_v2	pMYC20Cys_v2ΔcysRampR
pMYC20Cys_v4ΔcysC	Deletion of <i>cysC</i>	pMYC20Cys_v2	pMYC20Cys_v2ΔcysCampR

## 2.9 Supplementary Information References

1. 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; 10.1128/JB.01110-12 (2012).
2. Magrini, V., Creighton, C. & Youderian, P. Site-specific recombination of temperate *Myxococcus xanthus* phage Mx8: Genetic elements required for integration. *J. Bacteriol.* **181**, 4050–4061 (1999).
3. Yan, F. *et al.* Synthetic biology approaches and combinatorial biosynthesis towards heterologous lipopeptide production. *Chem. Sci.* **9**, 7510–7519; 10.1039/c8sc02046a (2018).
4. Agatep, R., Kirkpatrick, R. D., Parchaliuk, D. L., Woods, R. A. & Gietz, R. D. Transformation of *Saccharomyces cerevisiae* by the lithium acetate/single-stranded carrier DNA/polyethylene glycol protocol. *Technical Tips Online* **3**, 133–137; 10.1016/S1366-2120(08)70121-1 (1998).
5. Kouprina, N. & Larionov, V. Selective isolation of genomic loci from complex genomes by transformation-associated recombination cloning in the yeast *Saccharomyces cerevisiae*. *Nat. Protoc.* **3**, 371–377; 10.1038/nprot.2008.5 (2008).
6. Baumann, S. *et al.* Cystobactamids: myxobacterial topoisomerase inhibitors exhibiting potent antibacterial activity. *Angew. Chem. Int. Ed.* **53**, 14605–14609; 10.1002/anie.201409964 (2014).
7. Edgar, R. C. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **32**, 1792–1797; 10.1093/nar/gkh340 (2004).
8. Cociancich, S. *et al.* The gyrase inhibitor albicidin consists of *p*-aminobenzoic acids and cyanoalanine. *Nat. Chem. Biol.* **11**, 195–197; 10.1038/nchembio.1734 (2015).

9. Crooks, G. E., Hon, G., Chandonia, J. M. & Brenner, S. E. WebLogo: a sequence logo generator. *Genome Res.* **14**, 1188–1190 (2004).
10. Kim, B.-M., van Minh, N., Choi, H.-Y. & Kim, W.-G. Coralmycin Derivatives with Potent Anti-Gram Negative Activity Produced by the Myxobacteria *Coralloccoccus coralloides* M23. *Molecules* **24**; 10.3390/molecules24071390 (2019).
11. 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; 10.1021/jacs.7b09898 (2017).
12. Rose, R. E. The nucleotide sequence of pACYC177. *Nucleic Acids Res.* **16**, 356 (1988).
13. Rose, R. E. The nucleotide sequence of pACYC184. *Nucleic Acids Res.* **16**, 355 (1988).

## Chapter 3

### **Corramycin – a novel antibiotic class from myxobacteria hijacking inner membrane transporters of *Enterobacteriaceae*: Isolation, structure elucidation, biosynthesis, total synthesis and *in vivo* activity in infected mice**

#### **Preliminary author list**

Alexander von Tesmar<sup>1,2,3</sup>, Sebastian Groß<sup>1,2,3</sup>, Michael Hoffmann<sup>2,3</sup>, Ram Prasad Awal<sup>2,3</sup>, Nestor Zaburanyi<sup>2,3</sup>, Kirsten Harmrolfs<sup>2,3</sup>, Rolf Müller<sup>2,3,5</sup> and Stephane Renard<sup>4,5</sup>

<sup>1</sup> first author; <sup>2</sup> Helmholtz Center for Infection Research (HZI); <sup>3</sup> Helmholtz-Institute for Pharmaceutical Research Saarland (HIPS); <sup>4</sup> Evotec; <sup>5</sup> corresponding authors

This manuscript is in preparation. The work presented herein was performed by HZI/HIPS, Sanofi and Evotec. The author list and author contribution section is in a preliminary state and will be finalized soon after announcement of the authors from Sanofi and Evotec. The author contribution section is limited to the authors from HIPS. All experiments that are not mentioned in the author contribution section were performed by Sanofi and Evotec. Missing data, e.g. experimental protocols or NMR data, will be provided in a later stage of writing. The method section only contains protocols from the work performed at HIPS.

## **Contributions to the presented work**

### **HIPS author contributions**

Alexander von Tesmar identified and performed the *in silico* analysis of the corramycin biosynthetic gene cluster, designed and performed all experiments to elucidate the biosynthesis and the putative pre-drug mechanism. He also identified and described new corramycin derivatives and was involved in writing of the manuscript. Sebastian Groß performed *in silico* analysis of the corramycin biosynthetic gene cluster and wrote the manuscript.

### **Contributions by other HIPS authors**

Nestor Zaburannyi assembled and refined the raw sequence data from the *Coralloccoccus coralloides* producer strain genomes. Ram Prasad Awal isolated *C. coralloides* Mcy10984. Michael Hoffmann purified corramycin derivative Cor1347 and Kirsten Harmrolfs performed structure elucidation. The project was supervised by Rolf Müller.

### **Acknowledgement**

We thank Joachim Wink for providing the producer strain *C. coralloides* ST201330.

### 3 Corramycin – a novel antibiotic class from myxobacteria hijacking inner membrane transporters of *Enterobacteriaceae*: Isolation, structure elucidation, biosynthesis, total synthesis and *in vivo* activity in infected mice

#### 3.1 Abstract

Corramycins were isolated as novel secondary metabolites in a bioactivity-guided approach from the myxobacterium *Coralloccoccus coralloides*. They represent a novel peptide antibiotic class containing hitherto unique chemical features like a (2*R*,3*S*)- $\gamma$ -*N*-methyl- $\beta$ -hydroxy-histidine moiety. Genome sequencing of two producer strains enabled identification of the corresponding biosynthetic gene cluster (BGC). We propose a model for the biosynthesis of corramycins by a 12-module megasynthetase assembly line based on *in silico* analysis and feeding experiments with isotope-labeled precursors. The assembly line harbors a fatty-acid activating FAAL domain in the starter module that does not correspond to the molecules structure. *In vitro* reconstitution of fatty acid activation by FAAL indicates the presence of a pre-drug mechanism, involving acylated corramycin. Furthermore, eight stereocenters could be assigned by prediction from the BGC, eventually being verified by achieving a total synthesis of the molecule, which also allowed elucidation of the absolute configuration. One corramycin derivative, Cor1183, shows activity in the low  $\mu$ M range against *Escherichia coli* and no cross-resistance with any of the known antibiotics. Therefore, an unprecedented mode of action is postulated and shown to be dependent on uptake of Cor1183 via two redundant systems, SbmA and YejABEF. Administration of Cor1183 in a mouse model of *E. coli* septicemia lead to a 100 % survival rate of the mice after 4 days starting from 20 mg kg<sup>-1</sup> administration. Taken together, the novel compound class of corramycins is a promising starting point for the development of a potent antibacterial drug to tackle hospital-acquired infections.

#### 3.2 Introduction

The rise of antimicrobial resistance (AMR), especially in pathogens belonging to the ESKAPE panel (*Enterococcus faecium*, *Staphylococcus aureus*, *Klebsiella pneumonia*, *Actinetobacter baumannii*, *Pseudomonas aeruginosa* and *Enterobacter* species), is a global threat for the human health care system.<sup>1,2</sup> Particularly some multidrug-resistant Gram-negative pathogens were ranked with critical priority by the WHO (World Health Organization). The insight that resistance is only a question of time and all antibiotics therefore have only a limited life-span of use induces a constant need for novel therapeutics to successfully fight infectious diseases.<sup>3-5</sup> Most of the recently FDA-approved drugs derive from known antibiotic classes such as the  $\beta$ -lactam/ $\beta$ -lactamase inhibitor combinations

ceftolozane/tazobactam, ceftazidime/avibactam,<sup>6</sup> meropenem/vaborbactam,<sup>7</sup> the aminoglycoside antibiotic plazomicin<sup>8</sup> and the tetracycline derivative eravacycline,<sup>9</sup> which makes a rapid resistance development likely. However, a number of new structures with a new mode of action were recently described, for example the Gram-negative outer membrane-targeting compounds darobactin,<sup>10</sup> a chimeric peptidomimetic combining pharmacophores of murepavadin and polymyxin B<sub>1</sub><sup>11</sup> and the small molecule MRL-494.<sup>12</sup> Another example are the *Streptomyces*-derived griselimycins, which are highly active against *Mycobacterium tuberculosis* by inhibiting the new antimicrobial target DnaN, the DNA polymerase sliding clamp.<sup>13</sup> Nevertheless, those antimicrobials are still under investigation and the discovery of antibiotic classes with new modes of action against Gram-negative pathogens remains indispensable.

Bacterial natural products remain a main resource in the drug discovery process.<sup>14</sup> Myxobacteria are Gram-negative  $\delta$ -proteobacteria exhibiting a complex life-cycle and advanced multicellular social behavior, inhabiting almost every living space, including soil, deep-sea sediments, sweet water and hydrothermal vents.<sup>15</sup> Although poorly studied in comparison to actinomycetes, the order myxococcales produces a large quantity of natural products with remarkable chemical diversity.<sup>16–18</sup> Furthermore, myxobacteria have the largest known genomes in the bacterial domain with genome sizes ranging from 9 to 16 Mbp. The genomes contain 6–10 % of biosynthetic genes in 10 to 20 different clusters. Therefore, myxobacteria are thought to exhibit an equal potential for the biosynthesis of natural products as actinomycetes.<sup>19–21</sup>

Microbial natural products, including therapeutically valuable antibacterials, are structurally highly diverse and exhibit complex chemical scaffolds. They are most often produced by multifunctional enzyme complexes via two major pathways: the polyketide synthases (PKSs) and non-ribosomal peptide synthetases (NRPSs).<sup>22</sup> The multimodularity, functionality and plasticity of PKS and NRPS systems are described in detail elsewhere.<sup>23–25</sup> Briefly summarized, both systems follow a similar assembly line logic, in which a cascade of condensation reactions links simple monomeric building blocks. *In silico* identification and analysis of secondary metabolite gene clusters encoding those megaenzymes in bacterial genomes, e.g. with software like antiSMASH,<sup>26</sup> improved genome mining approaches in the past. Moreover, the automated annotation of potential biosynthetic gene clusters (BGCs), the prediction of domain organization in PKS and NRPS modules or their substrate specificity<sup>27–29</sup> enables assignment of known metabolites to previously unknown clusters or predictions of the stereochemistry based on sequence analysis.<sup>30</sup> The stereochemistry of amino acids in nonribosomal peptides can often be predicted by determination of the functional subtype of a condensation (C) domain.<sup>29</sup> <sup>L</sup>C<sub>L</sub> domains catalyze peptide bond formation between two L-amino acids, whereas <sup>D</sup>C<sub>L</sub> domains connect a D-amino acid with an L-amino acid.

Herein, we report the discovery of the novel natural product class of corramycins, which we isolated from the myxobacterium *Coralloccoccus coralloides*. Structure elucidation via NMR revealed some unprecedented structural features as for example the (2*R*,3*S*)- $\gamma$ -*N*-methyl- $\beta$ -hydroxy-histidine moiety. One derivative, corramycin 1183 (Cor1183), showed antibacterial activity against the ESKAPE pathogen *Escherichia coli* in the low  $\mu$ M range. Furthermore, we identified the corresponding BGC after genome sequencing of two producer strains. Based on *in silico* analysis of the BGC, we provide a biosynthesis model by a hybrid NRPS-PKS assembly line, including the probable presence of a pre-drug mechanism, and predicted the stereochemistry of eight stereocenters. The correctness of this prediction was confirmed by the total synthesis of Cor1183. Furthermore, the absolute configuration of the molecule was elucidated by the synthesis of every possible diastereomer and subsequent structural comparison with Cor1183. Moreover, we identified two Cor1183 uptake systems of *E. coli*, SbmA and YejABEF, and observed a low frequency of resistance in *E. coli* using minimal medium. Finally, we observed promising results in an infected mouse model with a 100 % survival rate after 4 days when administering 20 mg kg<sup>-1</sup> of Cor1183.

### 3.3 Results & discussion

#### 3.3.1 Identification and isolation of a novel antibiotic class from myxobacteria

We screened 3.896 myxobacterial extracts for their antimicrobial activity. Over 200 extracts were selectively active against *E. coli* ATCC35218. One of them contained an unknown active metabolite with an  $m/z$  [M+H]<sup>+</sup> value of 1184.56, which was subsequently isolated in an activity-guided isolation process from *C. coralloides* ST201330. We named the compound corramycin 1183 (Cor1183). Notably, we identified another strain *Coralloccoccus coralloides* MCy10984 producing the same compound.

To confirm the antibacterial activity of the purified compound, the MIC (minimal inhibitory concentration) of Cor1183 against *E. coli* ATCC35218, other *E. coli* strains and some important human pathogens was determined using a standard CLSI (Clinical and Laboratory Standards Institute) protocol.<sup>31</sup> Cor1183 is mainly active against *E. coli* ATCC35218 with a MIC of 1-4  $\mu$ g mL<sup>-1</sup>, but also against additional *E. coli* strains, including multidrug-resistant ones, and *S. typhimurium* (4  $\mu$ g mL<sup>-1</sup>) (Table 1). However, Cor1183 showed only a very limited or no activity against several Gram-negative and Gram-positive pathogens causing severe nosocomial infections such as *Staphylococcus aureus*, *Pseudomonas aeruginosa*, *Klebsiella pneumoniae* and others (Supplementary Table 1).

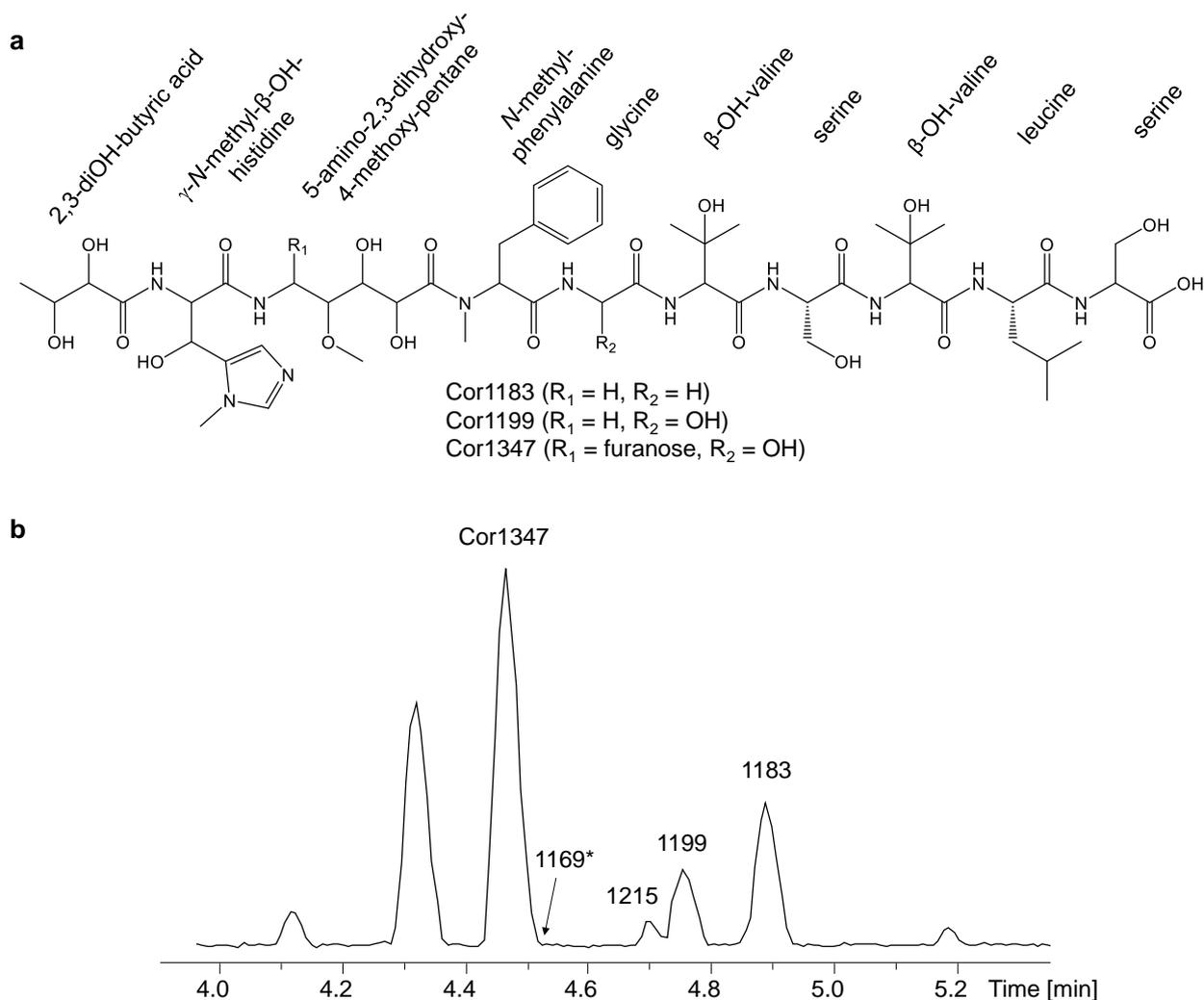
**Table 1 | MIC values of Cor1183 against selected Gram-negative and Gram-positive pathogens.**

<b>Gram-negative species</b>	<b>Strain</b>	<b>MIC (<math>\mu\text{g mL}^{-1}</math>)</b>
<i>E. coli</i>	NCTC 10418	8
	ATCC 25922	6
	ATCC 35218	1-4
	DSM 46345	32
	NCTC13441	4
	ATCC25922	2-8
	1705863	8
	1705878	16
	MG1655 WT	8-16
<i>S. typhimurium</i>	ATCC 13311	4
<i>E. cloacae</i>	17059482	>64
	DSM 46348	32
<i>K. pneumoniae</i>	1705966	>64
	1705949	64
	ATCC 13883	64
<b>Gram-positive species</b>	<b>Strain</b>	<b>MIC (<math>\mu\text{g mL}^{-1}</math>)</b>
<i>E. faecium</i>	A6349	>64
	DSM 17050	64
<i>E. faecalis</i>	1069 VanB	16

Albeit the narrow activity spectrum of Cor1183, we tested if Cor1183 exhibits cross-resistance with known antibiotic classes or whether there are any indications for a potentially new mode of action (unpublished data, personal communication with Stephane Renard). As we could not observe cross-resistance to any of the known antibiotic classes, we assessed Cor1183 as a good starting point to develop a novel anti-Gram-negative therapeutic.

### 3.3.2 Elucidation of the flat structure and identification of corramycin derivatives

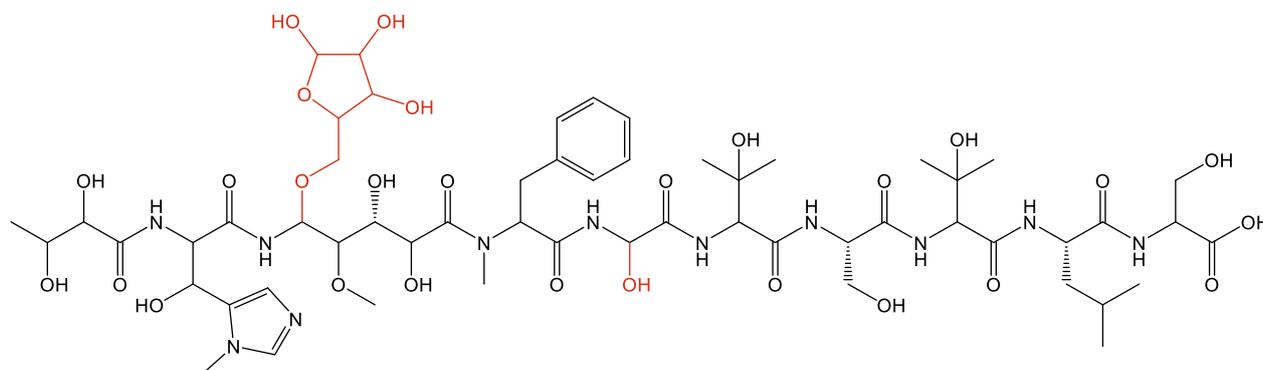
Next, we aimed for the elucidation of the flat structure of Cor1183 using intense NMR studies (data will be provided in the Supplementary Information in a later stage of the writing process). Cor1183 exhibits a linear peptidic structure with eight  $\beta$ -amino acids including several unique structural features such as the so far unknown  $\gamma$ -*N*-methyl- $\beta$ -hydroxy-histidine, an 5-amino-2,3-dihydroxy-4-methoxy-pentanoyl moiety and an N-terminal dihydroxy butyric acid moiety (Figure 1).



**Figure 1 | Flat structure of corramycin and BPC of a methanolic crude extract from *C. coralloides* Mcy10984.** **a:** The structures of Cor1183, Cor1199 and Cor1347 were verified by NMR. **b:** Base peak chromatogram (BPC) of methanolic crude extract from *C. coralloides* Mcy10984. The different corramycin derivatives are labelled. \*Peak intensity of Cor1169 was close to detection limit.

Interestingly, the supplementation of isotope-labelled precursors in the fermentation process coupled with high-resolution ESI-MS<sup>n</sup> analysis enabled the identification of several corramycin derivatives (Table 2, Supplementary Figure 1 and Supplementary Figure 2), including demethylated (Cor1169), hydroxylated (Cor1199) and double-hydroxylated (Cor1215) corramycins. We were able to purify Cor1183 and Cor1199 and elucidate their flat structures by NMR (data will be provided in the Supplementary Information in a later stage of the writing process); however, the concentrations of Cor1169 and Cor1215 in the fermentation broth were too low to allow compound isolation. Furthermore, a species with 1348  $m/z$   $[M+H]^+$  (Cor1347) showed identical mass shifts as Cor1183 upon supplementation of isotope labeled precursors and showed the highest peak intensities of all corramycin derivatives (Figure 1b). Cor1347 was isolated and the structure was elucidated via 2D NMR (HSQC and HMBC) experiments (data will be provided in the Supplementary Information in a later stage of the writing process) (Figure 2). Compared to Cor1183, Cor1347 is hydroxylated

at the CH<sub>2</sub> group of glycine and additionally exhibits a second hydroxy group, which is glycosylated with a furanose moiety and attached at position 1 of the 5-amino-2,3-dihydroxy-4-methoxy-pentanoyl moiety. Notably, we also examined the antibacterial activity of the isolated corramycin derivatives (data not shown), but surprisingly all of them lack antibiotic activity in the tested panel. Thus, our follow-up experiments focused on the only active congener, Cor1183.



**Figure 2 | Flat structure of Cor1347.** Structural differences compared to Cor1183 are labelled in red.

**Table 2 | Supplementation experiments of isotope-labelled corramycin precursors and detection of corramycin derivatives.** On the left  $\Delta m/z$  shows the observed mass shift in Cor1183 after supplementation of the respective isotope-labelled precursor. The listed corramycin derivatives were identified after isotope-labelled precursor supplementation.  $\Delta m/z$  refers to the mass difference compared to Cor1183.

Supplementation				Corramycin derivatives			
Isotope	$\Delta m/z$	Moiety	Module	Derivative	$m/z$	$\Delta m/z$	Modification
d <sub>3</sub> -Met	12	4 x Me	-	Cor1183 <sup>[a]</sup>	1184	-	
<sup>13</sup> C <sub>4</sub> - <sup>15</sup> N-Asp	3	$\beta$ -Ala	4	Cor1169	1170	- 14	- CH <sub>3</sub>
1- <sup>13</sup> C-Ac	1	-	-	Cor1199 <sup>[a]</sup>	1200	+ 16	+ OH
2- <sup>13</sup> C-Ac	1	Ac	5	Cor1215	1216	+ 32	+ 2 OH
<sup>13</sup> C <sub>2</sub> -Ac	2	-	-	Cor1347 <sup>[a]</sup>	1348	+ 164	+ 2 OH + pentose
d <sub>5</sub> -Phe	5	Phe	6				
d <sub>2</sub> -Gly	2	Gly	7				
d <sub>8</sub> -Val	14	2 x OH-Val	8, 10				
d <sub>3</sub> -Ser	6	2 x Ser	9, 12				
d <sub>3</sub> -Leu	3	Leu	11				
d <sub>6</sub> -OH-Val	-	-	-				
<sup>13</sup> C <sub>4</sub> - <sup>15</sup> N-Thr	-	-	-				
<sup>13</sup> C <sub>3</sub> -Glycerol	-	-	-				

<sup>[a]</sup> Structure determined by NMR.

### 3.3.3 Identification of the corramycin biosynthetic gene cluster

We next aimed for the identification of the corresponding BGC to gain insights into the biosynthesis and the stereochemistry of the unique structure of corramycins. Considering the peptidic structure

and the presence of the N-terminal dihydroxy butyric acid and the 5-amino-2,3-dihydroxy-4-methoxy-pentanoyl moiety, we assumed that they are produced by a hybrid NRPS-PKS. Thus, we sequenced the genomic DNA of both producer strains and screened for respective clusters in the genome of *C. coralloides* ST201330 after antiSMASH analysis.<sup>32</sup> We identified a locus harboring three NRPS modules including an *N*-methyltransferase domain (*N*-MT) with predicted adenylation (A) domain specificities for Phe-(*N*-MT)-Gly-Val based on the Stachelhaus code.<sup>28</sup> This prediction agreed with the chemical structure of Cor1183 and enabled the identification of the complete BGC in both producer strains (for details see Supplementary Information). The 12-modular hybrid NRPS-PKS assembly line of corramycin is encoded by five core genes (*comK-O*), which are organized in one putative operon. To determine the cluster borders, we compared the genomic regions surrounding the five genes in both producer strains with the phylogenetically closely related non-producer strain *Coralloccoccus coralloides* DSM2259 (Supplementary Figure 3). Fifteen putative genes were identified that presumably belong to the corramycin BGC (Table 3).

**Table 3 | Proposed function of the biosynthetic genes involved in corramycin biosynthesis.**

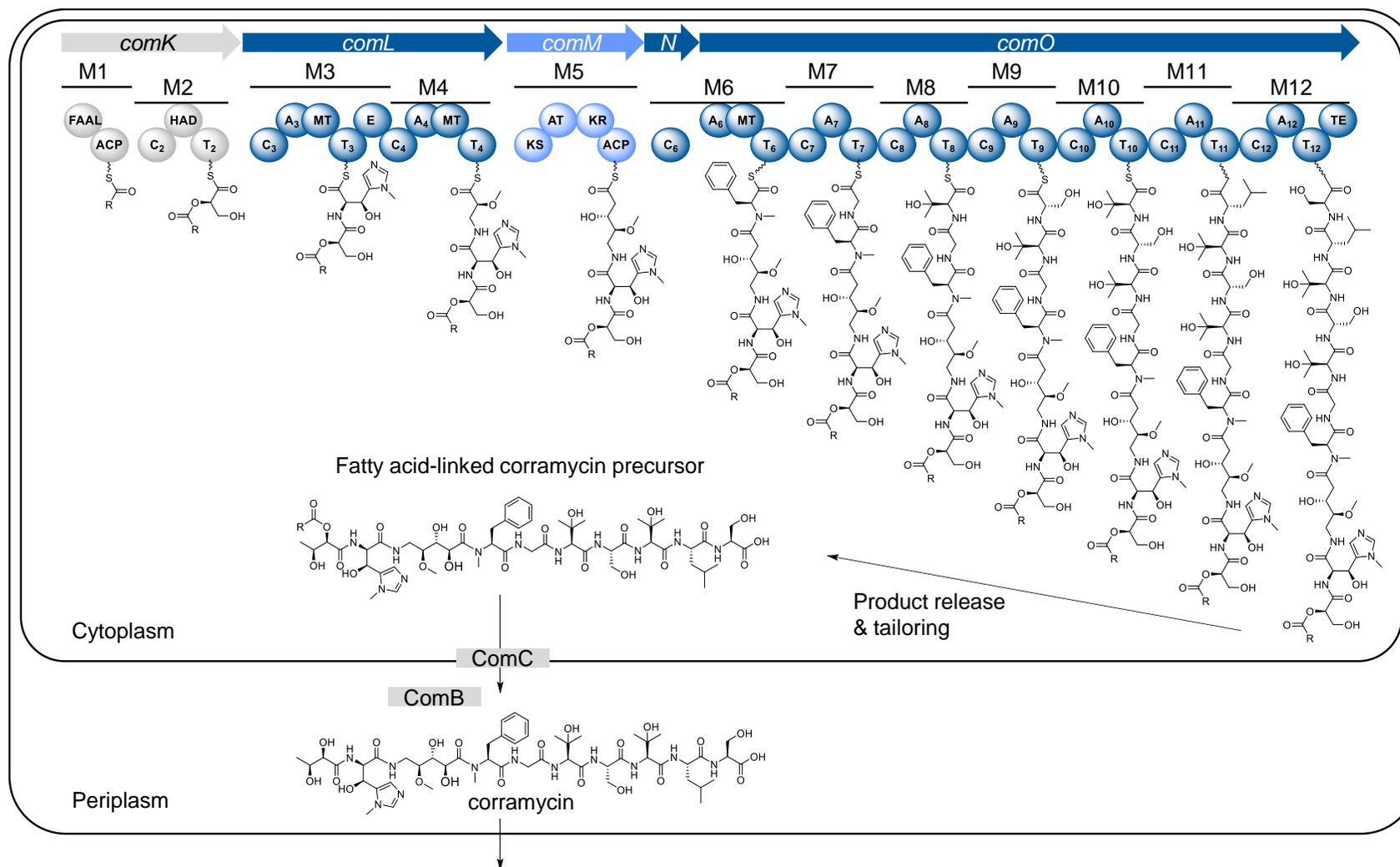
Gene	Size (aa)	Proposed function		Homolog	Identity (%)
<i>comA</i>	370	JmjC domain containing protein, hydroxylase		JMJ30	25
<i>comB</i>	378	o-Phtalyl amidase / lipase		POC2Y0	25
<i>comC</i>	424	Glycosyltransferase (membrane-bound)		arnF	31
<i>comD</i>	383	DNA replication and repair		RecF	38
<i>comE</i>	130	Aspartate decarboxylase		panD	78
<i>comF</i>	342	Luciferase-like monooxygenase		BtrO	27
<i>comG</i>	364	Kinase		murA	26
<i>comH</i>	388	JmjC domain containing protein, hydroxylase		KDM8	32
<i>comI</i>	372	Limonene 1,2-monooxygenase		limB	24
<i>comJ</i>	524	Vitamin B <sub>12</sub> -dependent radical-SAM methyltransferase	<b>Predicted domain specificity</b>	bchE	26
<i>comK</i>	5.379	NRPS	M1 (FAAL-ACP) M2 (C-HAD-T)	Acyl D-1,3-BPG	
<i>comL</i>	9.690	NRPS	M3 (C-A-MT-T-E) M4 (C-A-MT-T)	L-His β-Ala	
<i>comM</i>	4.554	PKS	M5 (KS-AT-KR-ACP)	Ac	
<i>comN</i>	1.419	NRPS	M6 (C)	L-Phe	
<i>comO</i>	23.328	NRPS	M6 (A-MT-T) M7 (C-A-T) M8 (C-A-T) M9 (C-A-T) M10 (C-A-T) M11 (C-A-T) M12 (C-A-T-TE)	Gly L-Val L-Ser L-Val L-Leu L-Ser	

### 3.3.4 Elucidation of the corramycin biosynthesis

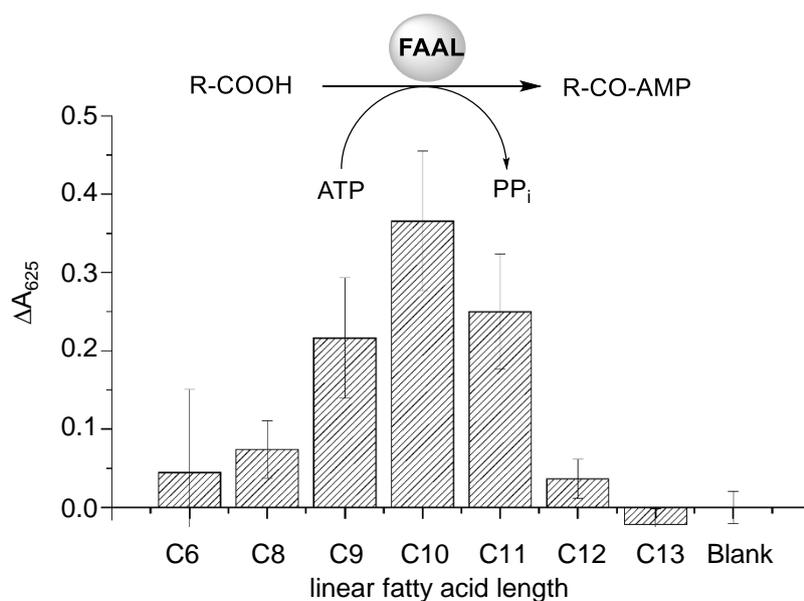
Based on the *in silico* analysis of the putative genes involved in corramycin biosynthesis and feeding experiments with isotope-labelled precursors (summarized in Table 2), we were able to propose a biosynthesis model shown in Figure 3.

The first module of the assembly line, encoded by *comK*, contains a fatty-acyl AMP ligase (FAAL), a member of the adenylate forming enzymes. FAALs are often associated with NRPS/PKS pathways as a starter unit catalyzing the acylation of the first building block.<sup>33–35</sup> *In silico* analysis revealed that all active sites in the corramycin FAAL are present (Supplementary Figure 4) and predicted a substrate specificity for long-chain fatty acids ranging from 5-12 C-atoms. To examine if the FAAL domain is indeed active and to investigate its substrate specificity, we heterologously overexpressed it in *E. coli* BL21 (DE3) and performed a malachite green assay<sup>36</sup> after protein purification. We observed activation of numerous linear fatty acids with different lengths by the FAAL domain (Figure 4). The highest activity was observed for decanoic acid and the spectrum of activated fatty acid ceased at tridecylic acid and hexanoic acid. The broad substrate specificity of the FAAL is not unexpected because lipopeptides often occur with various fatty acid chain lengths on the same peptide backbone.<sup>35,37,38</sup> However, upon re-inspection we did not find fatty acid-linked corramycin derivatives in the extracts of the producer strains.

There are numerous examples of lipopeptides harboring FAAL-incorporated fatty acid residues which are vital for their bioactivity.<sup>39–44</sup> On the contrary, in the biosynthesis of the myxobacterial lipopeptide vioprolide, a fatty acid-linked precursor is biosynthesized and maturation of the active product is achieved by hydrolysis of the fatty acid.<sup>45</sup> Since the corramycin FAAL domain is evidently active, we speculate that inactive, acylated corramycins are produced in terms of self-protection of the producer strain and, similarly to vioprolide biosynthesis, subsequently hydrolyzed to their active form. However, in contrast to vioprolide case, we could thus far not identify any acylated corramycin precursors, requiring further investigations in future experiments. A hypothetical pre-drug mechanism in corramycin maturation could include ComB and ComC, a putative hydrolase and flippase-like exporter, respectively, the latter showing homology with glycosyl transferases. Both harbor an *N*-terminal peptide signal for periplasmic localization. The membrane bound ComC might be involved in the export of putatively inactive, acylated corramycin precursors to the periplasm, where hydrolysis by ComB facilitates maturation to the final, active product.



**Figure 3 | Proposed biosynthesis scheme of corramycin.** The 12-modular NRPS-PKS assembly line produces a fatty acid-linked corramycin precursor which is hypothesized to be deacylated and transported into the periplasm by ComC and ComB prior to export. FAAL: fatty-acyl AMP ligase; ACP: acyl carrier protein; C: condensation domain; HAD: haloacid dehalogenase; T: thiolation domain; A: adenylation domain; MT: methyltransferase domain; E: epimerization domain; KS: ketosynthase; AT: acyl transferase; KR: ketoreductase; TE: thioesterase domain.



**Figure 4 | *In vitro* adenylation of linear fatty acids by the FAAL domain.** Pyrophosphate ( $PP_i$ ) production was detected in a malachite green assay. The shift in absorption at 625 nm ( $\Delta A_{625}$ ) was detected and used to determine the substrate specificity of the FAAL domain.

Another self-protection mechanism of *C. corralloides* includes ComG, a member of the protein-kinase-like superfamily. Heterologous expression of ComG in *E. coli* BL21 (DE3) increased the resistance against Cor1183 by a factor of  $10^3$  (Supplementary Figure 5). With ComG showing homology to kinases, we speculate that corramycin may be inactivated by phosphorylation similar to aminoglycoside resistance mechanisms. Interestingly, heterologous expression of ComD in *E. coli* BL21 (DE3) also lead to a 1,000-fold increase in resistance against Cor1183; however, at this time we did not find any well-described homologs of ComD that could enable us the proposal of a putative resistance mechanism. Future experiments may address those self-resistance mechanisms, which could possibly lead to identification of the cellular target of corramycins and pave the way to elucidate the mode of action.

The second module in corramycin biosynthesis, also encoded by *comK*, harbors an Fkbh-like domain belonging to the haloacid dehalogenase (HAD) superfamily. The same architecture of the first two modules (FAAL-ACP-C-HAD-T) was described for vioprolide, in which an unusual esterification process leading to the incorporation of a glycerate building block was described.<sup>46</sup> However, corramycins harbor an N-terminal C-4 butyric acid moiety and glycerate only contains three C-atoms. Thus, we assume that a glycerate moiety is incorporated and subsequently C-methylated in corramycin biosynthesis. Based on the feeding of  $d_3$ -methionine during fermentation of *C. corralloides* MCy10984 with subsequent analysis using LC-MS (Supplementary Figure 1), we propose this reaction to be performed by ComJ, a radical-SAM methyltransferase.

One intriguing chemical feature of corramycins is the unprecedented  $\gamma$ -*N*-methyl- $\beta$ -hydroxy-histidine moiety. Based on the architecture of module 3 including a putative *N*-methyltransferase and an epimerization (E) domain, we assume that L-histidine is activated by the A domain, *N*-methylated *in cis* and subsequently epimerized to  $\gamma$ -*N*-methyl-D-histidine. Interestingly, no NRPS-associated *N*-methyltransferase acting *in cis* on an aromatic substrate was described before. So far, also no biosynthetic pathway for  $\beta$ -hydroxylation of histidine has been described. NikQ, a heme-protein, has been described to facilitate the  $\beta$ -hydroxylation of histidine in the nikkomycin biosynthesis.<sup>47</sup> Interestingly, free histidine is not accepted and has to be loaded onto a T domain by a respective A domain, before it is recognized by NikQ. However, we could not identify any cytochrome P450 or other heme-containing proteins in the corramycin BGC. Thus, we assume a different mechanism taking place in corramycin biosynthesis that has to be addressed in future experiments.

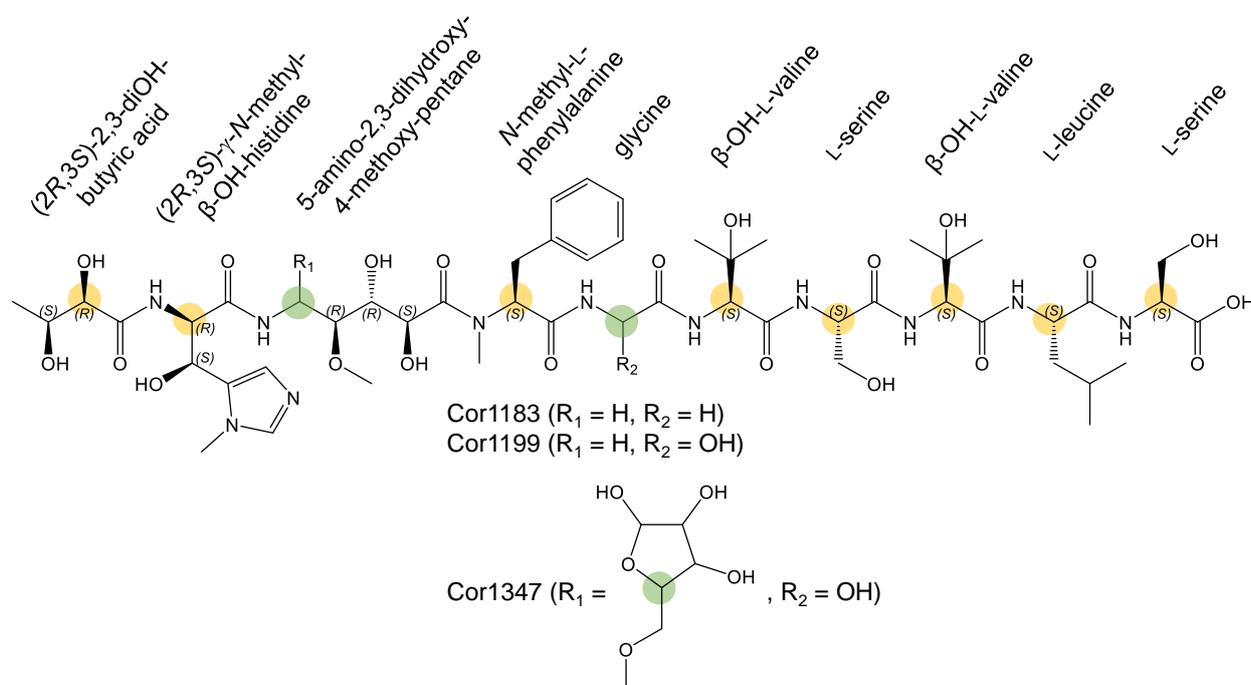
Another intriguing feature of corramycins is the unusual 5-amino-2,3-dihydroxy-4-methoxy-pentanoyl moiety. Based on the *in silico* analysis of modules 4 and 5, coupled with feeding experiments, we propose incorporation of  $\beta$ -alanine by module 4 with decarboxylation by CorE and an *O*-methylation performed *in cis*, which is followed by a *C*-2 extension catalyzed by the type I PKS module 5.

Interestingly, two  $\beta$ -hydroxy-valines are found in corramycins, which is a rarely observed building block in natural products. Independent supplementation of  $d_8$ -L-valine,  $d_6$ -L-hydroxy-valine and  $^{13}C_4$ - $^{15}N$ -L-threonine during fermentation showed that only valine was accepted by the NRPS (Table 2). Subsequently, hydroxylation occurs either on-line or after the product release. The  $\beta$ -hydroxy-valine moiety was also found in the myxobacterial nonribosomal peptide myxoprincomide,<sup>48</sup> however the biosynthetic origin has thus far not been identified. Notably, no enzyme was identified in the myxoprincomide BGC that catalyzes hydroxylation of valine indicating that the responsible genes are located elsewhere in the genome. For corramycin biosynthesis, further investigations are necessary to clarify the biosynthetic origin of  $\beta$ -hydroxy-valine, for example by the independent deletion of the genes encoding putative hydroxylases.

### **3.3.5 *In silico* analysis-based prediction of eight stereocenters and elucidation of the absolute configuration of corramycin**

Based on the *in silico* analysis of the module and domain architecture of the corramycin assembly line, we were able to predict the configuration of eight stereocenters (labelled in yellow in Figure 5). The HAD (FkbH) domain was proposed to load and dephosphorylate D-1,3-bisphosphoglycerate and was experimentally shown to activate D-3-phosphoglycerate by the vioprolide FkbH domain.<sup>46</sup>

Consequently, we proposed an *R* configuration for position 2 in corramycin. The analysis of all C domain subtypes (Supplementary Figure 6) revealed an <sup>L</sup>C<sub>L</sub> architecture for the C domain of module 3 (histidine) and a <sup>D</sup>C<sub>L</sub> architecture of the C domain of module 4. Since module 3 harbors an E domain, we postulate that L-histidine is activated by module 3 and subsequently epimerized to a D-histidine, which leads to an *R* configuration at the position 3 in corramycin. The C domain subtypes of modules 6 and 8-12 were determined as <sup>L</sup>C<sub>L</sub> subtype and no additional epimerization domains were identified within the respective modules. Consequently, we proposed *S* configuration for phenylalanine, both β-hydroxy-valines, both serines and leucine. Since the remaining five stereocenters biosynthetically either arose by hydroxylation through tailoring enzymes or by downstream PKS modifications, we were not able to predict configurations due to a lack of bioinformatics tools.

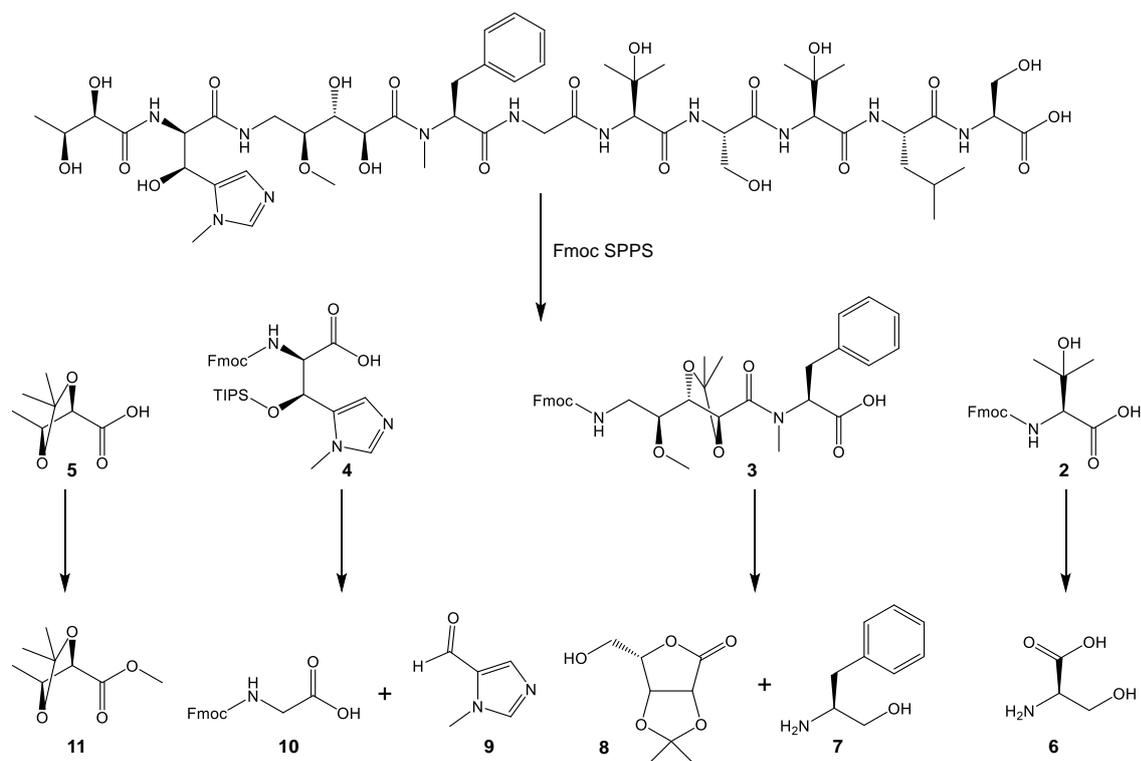


**Figure 5 | Absolute configuration of corramycin.** The configuration of the thirteen stereocenters of Cor1183 is given. Stereocenters which were correctly predicted after the *in silico* analysis are labelled in yellow. The remaining stereocenters were elucidated by comparison with numerous chemically synthesized diastereomers. Stereocenters which are only present in the corramycin derivatives Cor1199 and Cor1347 and have to be elucidated in future experiments, are labelled in green.

To verify the correctness of the predicted configuration of the eight stereocenters and to elucidate the configuration of the remaining five stereocenters, we aimed for the comparison of the isolated Cor1183 from fermentation with synthetic standards. Therefore, we established a total synthesis route to synthesize all possible diastereomers, finally allowing the elucidation of the absolute configuration of Cor1183 (the synthesis of all diastereomers and comparison with Cor1183 by NMR will be added to the Supplementary Information in a later stage of writing) (Figure 5).

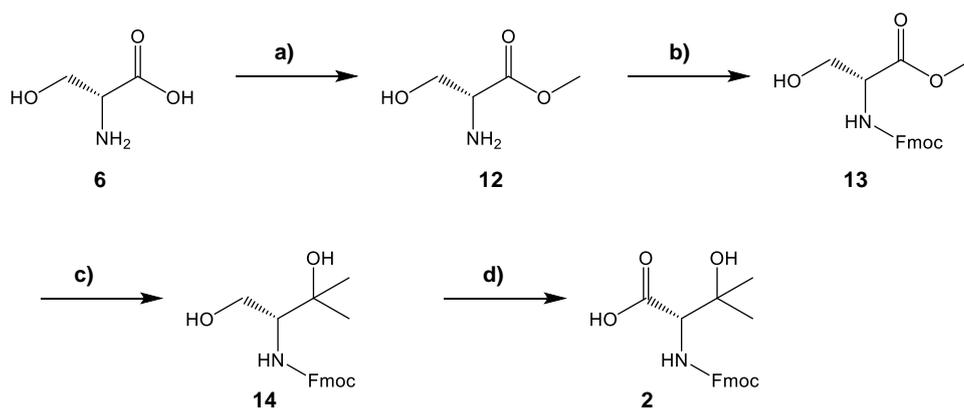
### 3.3.6 Total synthesis of Cor1183

The total synthesis of Cor1183 was not only established to elucidate the absolute configuration of the molecule, but also to overcome the low availability by fermentation and to set the basis for structure engineering towards analogs with improved antibacterial activity profile and favorable pharmaceutical properties. Therefore, we planned to start a solid-phase peptide synthesis (SPPS) approach beginning from the C-terminal L-serine using a traditional Fmoc-strategy with hydroxyl group-protected amino acids. However, since Cor1183 contains unprecedented and commercially unavailable building blocks, we had to perform a retrosynthetic analysis, in which we separated the peptidic C-terminal part of the molecule, including the  $\beta$ -hydroxy-L-valines (subsequently referred to as **2**), from the N-terminal part, including building blocks we had to synthesize separately, such as the 5-amino-2,3-dihydroxy-4-methoxy-pentanoyl-phenylalanine moiety (**3**), the (2*R*,3*S*)- $\gamma$ -*N*-methyl- $\beta$ -hydroxy-histidine (**4**) and the dihydroxy butyric acid (**5**) (Figure 6).



**Figure 6 | Retrosynthetic analysis of Cor1183.**

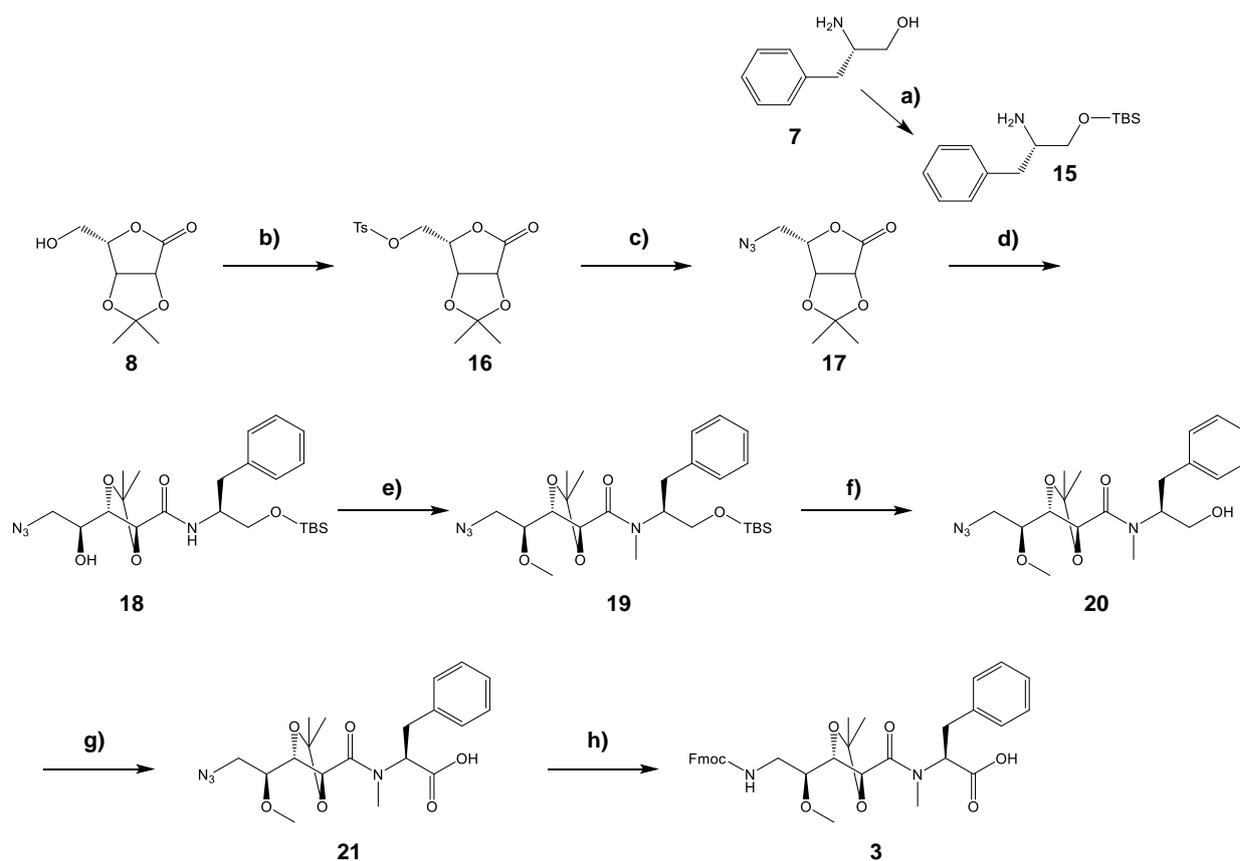
The synthesis of **2** was possible by using a previously described strategy starting from D-Serine (Figure 7). The esterification of the carboxy group in D-Serine by addition of a methyl group was followed by Fmoc protection of the amino group resulting in intermediate **13**. The subsequent addition of the Grignard reagent MeMgBr delivered a tertiary alcohol, followed by the oxidation of the alcohol, resulting in an acceptable overall yield of **2** (33 % over four steps).



**Figure 7 | Synthesis of building block 2.** Reagents and conditions: **a)**  $\text{SOCl}_2$ , MeOH, 0 °C to room temperature (RT) to reflux, 3 h (yield 94 %); **b)** FmocOSu, Dioxane-water (1:1),  $\text{NaHCO}_3$ , 0 °C to RT, 2 h (yield 72 %); **c)** MeMgBr, THF, -78 °C to 0 °C to RT, (yield 69 %); **d)** Tempo, NaOCl,  $\text{NaClO}_2$ , pH 4 buffer,  $\text{CH}_3\text{CN}$ , 0 °C to rt (yield 72 %).

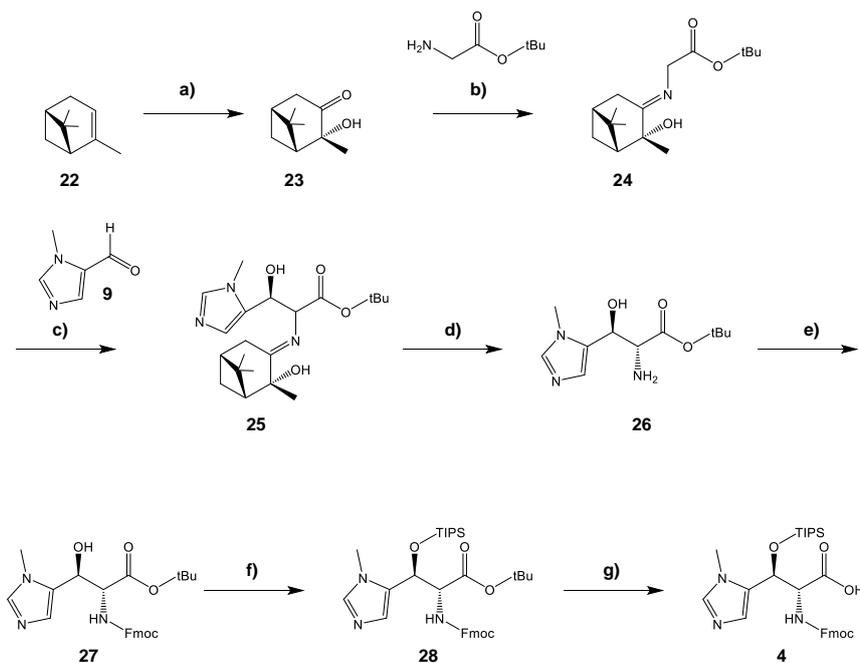
More steps were required for the synthesis of **3** (Figure 8). The commercial availability of a protected D-Ribose derivative (**8**) had the advantage having three stereocenters according to the corramycin structure. We used this building block as starting point for the synthesis of **3**. In the first two steps, the free alcohol was replaced by an azido group to yield intermediate **17**. Then the lactone ring was opened using a derivative of L-phenylalanine (**7**) to yield intermediate **18** (57 % yield). Methylation of the free alcohol and amide group, followed by deprotection of primary alcohol with TBAF and oxidation, furnished derivative **20**. Finally, the transformation of the azido group into an Fmoc-protected amine under hydrogenation conditions allowed to obtain building block **3** in an overall acceptable yield of 12 % over seven synthesis steps.

Due to the presence of two stereocenters in building block **4**, we envisaged to create those by enantioselective aldolisation using a titanium enolate derived from a chiral iminoglycinate (Figure 9). Using  $\alpha$ -pinene as chiral agent, an oxidation step was followed by imine formation with protected glycine, delivering **24**. The aldolisation step with imidazole (**9**) using  $\text{ClTi}(\text{OiPr})_3$  as Lewis acid followed by acidic deprotection delivered **26**, which was directly protected with a Fmoc group to furnish **27**. Alcohol protection and carboxylic acid deprotection yielded the final building block **4** in a good overall yield (5 % for seven steps). Building block **5** was synthesized by a saponification of the commercial starting material **11** with lithine in a THF/water mixture at room temperature for 15 h with a yield of 100 % (The synthesis scheme will be provided in a later stage of writing).

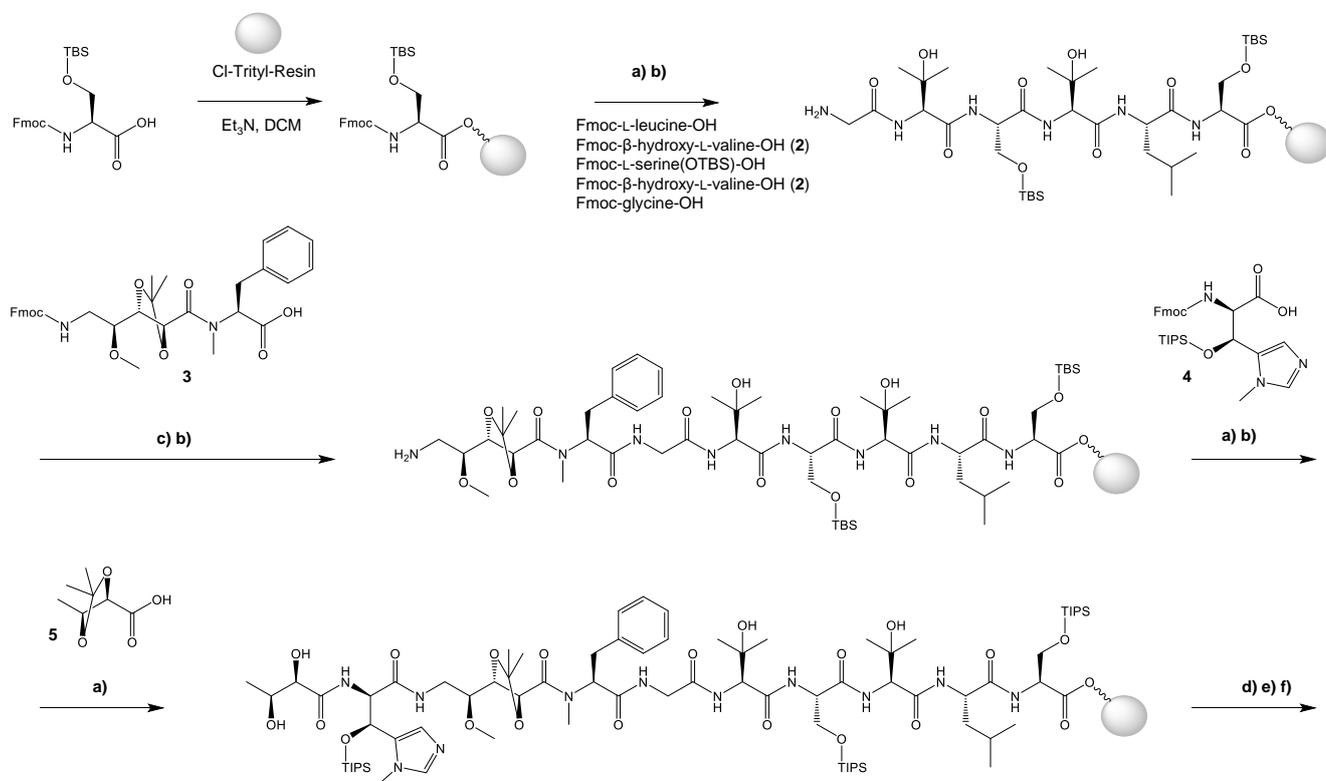


**Figure 8 | Synthesis of building block 3.** Reagents and conditions: **a)** Et<sub>3</sub>N, TBSCl, imidazole, DCM, 0-15 °C, 16 h (yield 47 %); **b)** TsCl, pyridine, DCM, 0-15 °C, 16 h (yield 82 %); **c)** NaN<sub>3</sub>, DMF, 15-60 °C, 3 h (yield 87 %); **d)** 15-80 °C, 16 h (yield 57 %); **e)** MeI, NaHMDS, THF, -70-15 °C, 16 h (yield 94 %); **f)** NH<sub>4</sub>F, MeOH, 80 °C, 16 h (yield 80 %); **g)** NaClO, TEMPO, NaClO<sub>2</sub>, MeCN, H<sub>2</sub>O, 10-15 °C, 16 h (yield 86 %); **h)** H<sub>2</sub> (15-30 psi), Pd/C, FmocOSu, THF, 25 °C, 16 h (yield 45 %).

Finally, the synthesis of Cor1183 was completed by coupling all the amino acids and respective building blocks (Figure 10). Therefore, we used traditional Fmoc-SPPS in standard conditions (HATU, DIEA, DMF, room temperature) starting from Fmoc-TBDMSO-L-serine using Fmoc-L-Leu, Fmoc-β-OH-L-valine, Fmoc-TBDMSO-L-serine, Fmoc-β-OH-L-valine and glycine. Incorporation of **3** was achieved at -20 °C in order to minimize epimerization of C<sub>α</sub> in the phenylalanine moiety. Around 10 % of the D diastereoisomer was generally observed at room temperature, whereas only around 2 % was present at -20 °C. After incorporation of **4**, the final coupling with building block **5** was achieved. The silyl protecting groups were removed with TBAF followed by cleavage of the peptide from the resin with HFIP. Final acidic deprotection was followed by a reverse-phase purification step, yielding pure Cor1183. In conclusion, the first total synthesis of Corramycin has been accomplished with the confirmation of the structure via NMR and biological activities in comparison with purified Cor1183.



**Figure 9 | Synthesis of building block 4.** Reagents and conditions: **a)**  $\text{KMnO}_4$ , acetone/water, 0-5 °C, 26 h (yield 21 %); **b)**  $\text{BF}_3 \cdot \text{Et}_2\text{O}$ , toluene, 25-110 °C, 7 h; **c)** **9**,  $\text{ClTi}(\text{OiPr})_3$ ,  $\text{Et}_3\text{N}$ ,  $\text{CH}_2\text{Cl}_2$ , 0 °C, 3 h; **d)**  $\text{HCl}$  1.2N, THF, 25 °C, 16 h; **e)**  $\text{FmocOSu}$ , acetone/water, 15 °C, 12 h; **f)**  $\text{TIPSOTf}$ , 2,6-lutidine, DCM, 15 °C, 12 h (yield 72 %, five steps); **g)**  $\text{TFA}$ , DCM, 25 °C, 12 h (yield 35 %).



**Figure 10 | Synthesis of Cor1183.** Reagents and conditions: **a)**  $\text{HATU}$ ,  $\text{DIPEA}$ , RT, 18 h, DMF; **b)** Piperidine/DMF; **c)**  $\text{HATU}$ ,  $\text{DIPEA}$ , -25 °C, 1 h then RT, 18 h, DMF; **d)**  $\text{TBAF}$  1N in THF, RT, 12 h; **e)** HFIP/DCM, RT, 20 min; **f)**  $\text{TFA}/\text{H}_2\text{O}/\text{Et}_3\text{SiH}$  (95/2.5/2.5), RT, 50 min.

### 3.3.7 Frequency of resistance and corramycin uptake

Next, we planned to investigate the potential of Cor1183 as a starting point for the development of a new antimicrobial drug. Therefore, we first performed a time-kill curve experiment using *E. coli* ATCC25922, which was cultivated in MHB full medium or M9 minimal medium, and the 4x and 8x MIC of Cor1183 (Supplementary Figure 7). This experiment revealed that Cor1183 is rapidly bactericidal, independent of the medium used. However, we observed regrowth in MHB after already 6 hours indicating that *E. coli* gained resistance towards Cor1183 with a high-frequency of resistance (FoR) under these conditions. Therefore, we performed a FoR experiment using *E. coli* ATCC25922 in MHB and M9 medium with a 4x MIC of Cor1183. The FoR was much higher in MHB medium ( $3.8 \times 10^{-6}$ ) compared to M9 medium ( $3 \times 10^{-10}$ ). Next, we sequenced the genome of resistant *E. coli* clones cultivated in MHB medium to identify potentially responsible mutations. A frameshift in the gene encoding the SbmA transporter gave a first hint about the bacterial uptake systems of Cor1183 as the potential inactivation of SbmA, caused by the frameshift, may confer resistance towards Cor1183. Interestingly, clones harboring the mutated SbmA were only resistant towards Cor1183 in MHB medium but not in M9 medium. Thus, we performed an additional resistance experiment with SbmA mutants in M9 medium. We sequenced the genome of clones which gained resistance towards Cor1183 in M9 medium and found additional mutations (frameshifts) in YejE or YejF. Independent or combined deletions of SbmA and YejEF in *E. coli* ATCC25922 using  $\lambda$  red-mediated homologous recombination (will be described in the Method section and Supplementary Information in a later stage of writing) resulted in strains that showed medium-dependent resistance towards Cor1183 (Table 4). The uptake appears to be dependent on the single SbmA transporter in MHB medium and both SbmA and YejEF transporters in M9 medium. Functional complementation of SbmA and YejEF in *E. coli* ATCC25922  $\Delta$ sbmA/ $\Delta$ yejEF (will be described in the Supplementary Information in a later stage of writing) revealed that solely SbmA is able to restore susceptibility to Corramycin in MHB medium.

**Table 4 | Susceptibility of *E. coli* towards Cor1183 after deletion and complementation of SbmA or YejEF. MIC values are given in  $\mu\text{g mL}^{-1}$ .**

Strain/Medium	<i>E. coli</i> ATCC25922			
	wild type	$\Delta$ sbmA	$\Delta$ yejEF	$\Delta$ sbmA/ $\Delta$ yejEF
MHB	4	64-128	4-8	>128
M9	2	2	2	>128
Strain/Plasmid/Medium	<i>E. coli</i> ATCC25922 $\Delta$ sbmA/ $\Delta$ yejEF			
	empty pBAD	pBAD-yejABEF	empty pTRC	pTRC-sbmA
MHB	>128	>128	>128	2

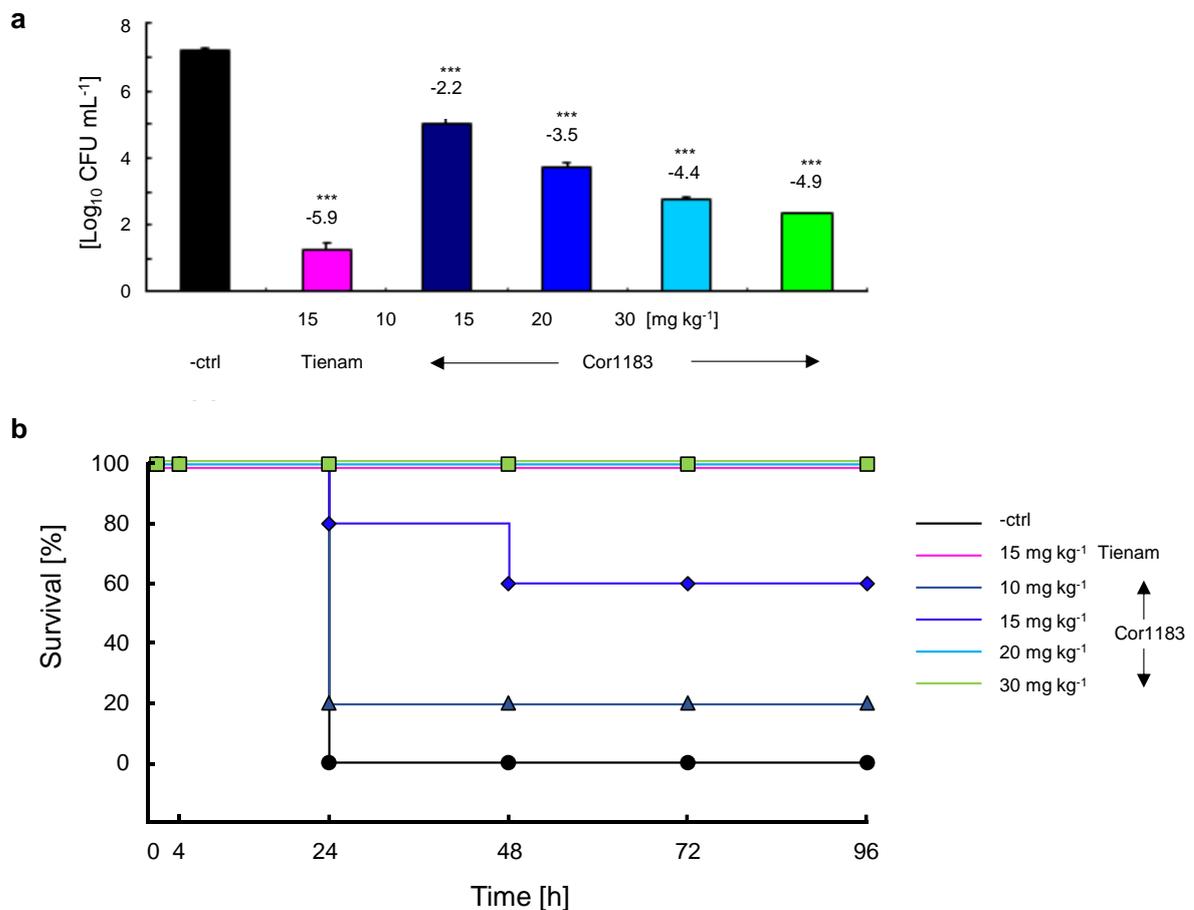
One recognized bottleneck for the identification of novel agents with specific antimicrobial activity against Gram-negative bacteria is the difficulty for compounds to penetrate the robust permeability barrier caused by the differing properties of outer and inner membranes.<sup>49</sup> Cor1183 seems to hijack several bacterial transporters in a seemingly redundant manner. SbmA is an ABC transporter that has been described being responsible for the uptake of several antibacterial molecules such as microcin J25<sup>50</sup> and bleomycin.<sup>51</sup> SbmA is present in *E. coli* and *K. pneumoniae* but absent in other enterobacteriaceae where an ortholog can be identified under the name of YddA.<sup>52</sup> YejABEF is an inner membrane ABC transporter, which is conserved in Gram-negative bacteria and was shown to be responsible for the uptake of microcin C.<sup>53</sup> Both transporters were shown to play a vital role in the uptake of Cor1183; however, SbmA seems to be the main uptake system when *E. coli* was cultivated in MHB medium. The dominant role of SbmA for Cor1183 susceptibility in MHB medium was also confirmed using other *E. coli* strains, such as K12, ATCC35218 and BAA2469 (data not shown).

Since compound penetration into bacteria is one of the major bottlenecks in antibiotic development, potential drugs hijacking only a single transporter make quick resistance development likely, which was e.g. shown for fosfomicin,<sup>54,55</sup> kazugamycin<sup>56</sup> and GE81112.<sup>57</sup> This was also shown here, where SbmA is the only transporter leading to Cor1183 uptake in MHB medium. The high FoR against Cor1183 observed in MHB medium would potentially be an exclusion criterion regarding the further development of corramycin towards a marketable drug. However, the FoR was much lower in M9 minimal medium, where two transporter systems had to be mutated to decrease Cor1183 susceptibility. As bacterial resistance development in minimal medium is more representative to an actual infection *in vivo*, we decided to test the *in vivo* efficacy of Cor1183 in infected mice, despite the poor FoR experiment results in full medium.

### **3.3.8 *In vivo* activity of Cor1183 in infected mice**

Therefore, we evaluated the *in vivo* antibacterial activity of Cor1183 in a peritonitis mouse model of infection (Figure 11). The mice were infected intraperitoneally with *E. coli* ATCC35218 and treated with different doses of Cor1183 using a double administration 1 h and 3 h after infection. We observed a dose-dependent reduction of the bacterial load 4 h after infection resulting in a 100 % survival rate of the mice after 4 days starting from 20 mg kg<sup>-1</sup> total dose. Notably, Cor1183 is rapidly eliminated after i.v. administration, keeping the plasma concentration above the MIC for less than 30 min. However, despite this rapid elimination, the antibacterial effect was strong enough in this animal model to achieve survival.

Since we previously identified two potential transporters (SbmA and YejEF) that confer resistance towards Cor1183, we investigated the susceptibility of *E. coli* ATCC25922 with independent or combined deletion of those transporters in the animal model. The ED50 (effective dose for 50 % of the population) of the compound could not be evaluated in blood, because the tested doses were very efficacious for the wild type *E. coli* and the single deletion mutants. However, we could evaluate the ED50 in the spleen of the animals (Table 5). The deletion of a single transporter in *E. coli* resulted in a mild decrease of susceptibility to Cor1183 compared to the wild type control strain, whereas the double deletion of SbmA and YejEF lead to a significant increase in resistance.



**Figure 11 | *In vivo* antibacterial activity of Cor1183 in infected mice. a:** -ctrl (black; negative control): no treatment; Tienam (pink; positive control) imipenem/cilastatin treatment; Cor1183 (dark blue, blue, light blue and green) treatment with different doses. CFU (colony forming units) load in the blood 4 h post infection. **b:** Survival of the mice 96 hours after infection using the respective doses per day.

**Table 5 | ED50 of Cor1183 in mice infected with different *E. coli* strains.**

Strain	<i>E. coli</i> ATCC25922			
	wild type	$\Delta$ sbmA	$\Delta$ yejEF	$\Delta$ sbmA/ $\Delta$ yejEF
<b>ED50</b> [mg kg <sup>-1</sup> ]	6.1	23.3	21.4	>40

Taken together, the promising *in vivo* activity of Cor1183 in infected mice underlines the potential of Cor1183 as starting point for an antibiotic drug development process. The decreased susceptibility towards Cor1183 in *E. coli* strains lacking the SbmA and YejEF transporter systems results in an increased ED50 in the infected mice, thus complementing the results that we obtained in our previous *in vitro* experiments.

### 3.4 Conclusion and Outlook

In summary, corramycin is a novel antibiotic scaffold with unique or in natural products rarely observed structural features and promising Gram-negative activity against numerous *E. coli* strains and *S. typhimurium*. Based on the unprecedented chemical structure of corramycin and no observed cross-resistance with any of the known antibiotic classes, we assume a novel, not yet elucidated, mode of action. The identification of the biosynthetic pathway in *C. coralloides* undermines that myxobacterial species are a promising source of new antibiotic natural products with completely different chemistry compared to previously identified antibiotics, e.g. from actinobacteria. Furthermore, we proposed a biosynthesis model revealing interesting features such as a possible detoxification mechanism including the production of fatty acid-linked corramycin precursors, which are exported and deacylated to secrete the active compound. Nevertheless, many proposed biosynthetic steps require further experimental investigations. Notably, the correct prediction of eight stereocenters, exclusively based on the *in silico* analysis of the assembly line, underlines the potential of bioinformatic tools in structure prediction of NRPS-derived natural products. Although there was a huge discrepancy between the FoR in full medium compared to minimal medium, the involvement of two transport systems in the corramycin uptake and low FoR in minimal medium plus the promising *in vivo* activity of Cor1183 in infected mice, makes corramycin a perfect candidate for further development. With the establishment of a total synthesis route we already set a milestone for this developmental process. Current research aims at the (semi-)synthesis of novel corramycin derivatives with improved antimicrobial activity and a broadened activity spectrum. Given the numerous structural and biosynthetic features of corramycins, we hope that this compound class will serve as starting point for the development of a marketable antibiotic.

## 3.5 Methods

### 3.5.1 Cultivation of strains

Cultivation of *C. coralloides* Mcy10984 for corramycin production screening was carried out in 30 mL AMB medium (0.5 % (w/v) soluble starch, 0.25 % (w/v) casitone (bacto), 0.05 % (w/v)  $\text{MgSO}_4 \times 7\text{H}_2\text{O}$ , 0.025 % (w/v)  $\text{K}_2\text{HPO}_4$ , 10 mM HEPES, pH 7.0) for 4 – 5 days at 30 °C and 200 rpm on an orbital shaker. Additionally 0.00005 % (w/v) vitamin B<sub>12</sub> and 2 % (v/v) Amperlite XAD16 absorber resin were supplemented after medium sterilization. For isotope-labelled precursor supplementation experiments, the cultivation volume and time were 10 mL and 3 days, respectively, and 1 mM of the following isotope-labelled precursors were fed: L-leucine-5,5,5-*d*<sub>3</sub>, L-serine-2,3,3-*d*<sub>3</sub>, L-valine-*d*<sub>8</sub>, L-threonine-<sup>13</sup>C<sub>4</sub>-<sup>15</sup>N, L-phenylalanine-*d*<sub>5</sub>, L-methionine (methyl-*d*<sub>3</sub>), sodium 1-<sup>13</sup>C-acetate, sodium 2-<sup>13</sup>C-acetate, 1,2-<sup>13</sup>C-acetate, L-aspartate-<sup>13</sup>C<sub>4</sub>-<sup>15</sup>N, glycine-2,2-*d*<sub>2</sub>, β-hydroxy-L-valine-*d*<sub>6</sub> and glycerol-<sup>13</sup>C<sub>3</sub>. For genomic DNA isolation, *C. coralloides* ST201330 and *C. coralloides* MCy10984 were cultivated in 50 mL AMB medium for 5 days at 30 °C and 180 rpm on an orbital shaker. For compound isolation, 1 L production cultures were cultivated in 5 L Erlenmeyer flasks using E medium (0.4 % skimmed milk, 0.4 % (w/v) soy meal, 0.2 % (w/v) yeast extract, 1 % (w/v) soluble starch, 0.1 % (w/v)  $\text{MgSO}_4 \times 7\text{H}_2\text{O}$ , 10 mM HEPES, pH 7.4 and supplementation with 8 mg L<sup>-1</sup> Fe-EDTA, 0.5 % glycerole and 0.00005 % vitamin B<sub>12</sub> after autoclaving). Cultivation was carried out for 5 days at 30 °C and 200 rpm. For Cor1183 isolation, 2 % Amperlite XAD16 absorber resin were added. Cultivation of *E. coli* for FAAL protein expression or heterologous expression of *comB* and *comG* for resistance studies is described in the respective section below.

### 3.5.2 Sample preparation and UPLC-ESI-HRMS analysis

Cells and XAD16 absorber resin of 10 mL and 30 mL screening cultures were harvested by centrifugation at 3,200 × *g* for 20 min at 20 °C. Extraction was done two times for 60 min with 15 mL and 30 mL methanol under stirring at room temperature (RT), respectively. Extracts were filtered using folded filter paper (8-12 μm pore size) and dried using a rotary evaporator. Dried extracts were dissolved in 100 μL (for isotope-labelled precursor supplementation experiments) and 1 mL methanol and analyzed using UPLC-HRMS.

Isotope-labelled precursor supplementation samples were analyzed using a Dionex Ultimate 3000 RSLC system and a BEH C18 (1.7 μm, 50 x 2.1 mm; Waters) column with an injection volume of 2 μL. Separation was achieved by a linear gradient with deionized water (A) + 0.1 % FA (formic acid) and acetonitrile (B) + 0.1 % FA at a flow rate of 600 μL min<sup>-1</sup> and 45 °C. The gradient was initiated by a 0.33 min isocratic step at 5 % B, followed by an increase to 95 % B in 9 min to end up

with a 1 min flush step at 95 % B before re-equilibration to initial conditions. The HPLC system was coupled to an Orbitrap mass spectrometer (Thermo Scientific) by a Triversa NanoMate nano-ESI system (Advion). Mass spectra were acquired in centroid mode ranging from 200 to 2,000  $m/z$  at a resolution of  $R = 30,000$ .

For the analysis of production screening cultures, an UltiMate 3000 LC System (Dionex) with a Acquity UPLC BEH C-18 column (1.7  $\mu\text{m}$ , 100 x 2 mm; Waters), equipped with a VanGuard BEH C-18 (1.7  $\mu\text{m}$ ; Waters) guard column, was coupled to an Apollo II ESI source (Bruker) and hyphenated to maXis 4G ToF mass spectrometer (Bruker). Separation was performed at a flow rate of 0.6 mL/min (eluent A: deionized water + 0.1 % FA, eluent B: acetonitrile + 0.1 % FA) at 45 °C using the following gradient: 5 % B for 30 s, followed by a linear gradient up to 95 % B in 18 min and a constant percentage of 95 % B for further 2 min. Original conditions were adjusted with 5 % B within 30 s and kept constant for 1.5 min. The LC flow was split to 75  $\mu\text{L min}^{-1}$  before entering the mass spectrometer. Mass spectra were acquired in centroid mode ranging from 150–2,500  $m/z$  at a 2 Hz full scan rate. Mass spectrometry source parameters were set to 500 V as end plate offset, 4,000 V as capillary voltage, 1 bar nebulizer gas pressure, 5 L  $\text{min}^{-1}$  dry gas flow and 200 °C dry temperature. For MS<sup>2</sup> experiments, CID (collision-induced dissociation) energy varied linearly from 30, 35, 45, to 55 eV with respect to the precursor  $m/z$  from 300, 600, 1,000, to 2,000  $m/z$ . MS full scan acquisition rate was set to 2 Hz and MS/MS spectra acquisition rates were ramped from 1 to 3 Hz for precursor ion intensities of 10 kcts to 1,000 kcts. We used *Compass DataAnalysis* version 4.4 (Bruker) to interpret MS data.

### 3.5.3 Purification and structural determination of Cor1183, Cor1199 and Cor1347

Cor1183 and Cor1199 purification from ST201330 and structure elucidation was performed previously (unpublished data, personal communication with Stephane Renard). Cor1347 was purified directly from the supernatant without supplementation of XAD16 absorber resin. The supernatant was dried by lyophilization or *in vacua* and extracted using methanol. The subsequent preparative HPLC purification step was performed with an Autopurifier System (APS; Waters) equipped with a Biphenyl (5  $\mu\text{m}$ , 250 x 22 mm) column at a flow rate of 25 mL  $\text{min}^{-1}$ . Deionized water (eluent A) + 0.1 % FA and acetonitrile (eluent B) + 0.1 % FA were used as solvents. The separation gradient started with a 1 min isocratic step at 25 % B, followed by a gradient to 29 % B in 22 min, a steep increase to 95 % B in 1 min to end up with a plateau for 4 min prior to return and re-equilibration with the initial conditions. The injection volume was 800  $\mu\text{L}$ . Fraction collection was controlled by a mass trigger which was set to 1348.5  $m/z$ , respectively. The solvent was removed from the collected fraction and the dry extract was resolved in a small amount of methanol for further analysis.

NMR spectra were recorded on a 700 MHz Avance III (Ascend) spectrometer (Bruker BioSpin GmbH), equipped with a 5 mm TXI cryoprobe, at 298 K. Chemical shift values of  $^1\text{H}$ - and  $^{13}\text{C}$ -NMR spectra are reported in ppm relative to the residual solvent signal given as an internal standard.  $^{13}\text{C}$ -signals were assigned via 2D- $CH$  and  $CCH$  correlations (HSQC and HMBC).

### 3.5.4 Isolation of genomic DNA and Illumina sequencing

ST201330 and MCy10984 cultures were harvested by centrifugation at  $3,200 \times g$  for 15 min at  $4^\circ\text{C}$  and the cell pellet was resuspended in 5 mL 10 mM Tris-HCl (pH 7.5) plus 300  $\mu\text{L}$  Proteinase K solution ( $10 \text{ mg mL}^{-1}$ ) and 600  $\mu\text{L}$  SDS-solution (10 %). The mixture was incubated at  $55^\circ\text{C}$  for 2 h. Subsequent extraction using phenol:chloroform:isoamylalcohol (P:C:I; 25:24:1) was conducted three times: One volume of P:C:I was added and the mixture was incubated at room temperature (RT) for 60 min under constant tube inversion (5 rpm). After centrifugation at  $3,200 \times g$  for 10 min at RT, the upper phase was extracted again. A fourth extraction of the upper phase with C:I (24:1) was performed prior to addition of 1/10 volume of 3 M Na-acetate (pH 5.5) and 2.5 volumes of ice-cold 100 % ethanol. The precipitated DNA was washed once in 1 mL of 70 % ethanol and subsequently air-dried overnight. The DNA-pellet was resuspended in 500  $\mu\text{L}$  of 10 mM Tris-HCl buffer (pH 7.5).

Illumina sequencing of strain ST201330 was performed using a MiSeq platform (SEQ-IT GmbH & Co. KG). The raw sequencing data comprised 26,054,366 paired-end reads and 14,483,850 mate-pair reads with a read length of 250 bp, respectively. Assembly into contigs with Abyss-pe assembler (version 1.3.6), resulting in 26 sequences with total length of 10,243,981 bp. Genome sequencing of strain MCy10984 was performed using Illumina sequencing technology (MiSeq platform; in-house service). The obtained raw sequencing data included 9,509,890 paired-end reads, each with a length of 250 bp. Assembly of the raw sequencing data in 59 sequences resulted in a genome sequence with a total length of 11,107,395 bp.

### 3.5.5 *In silico* experiments

*Geneious* v10.1.3 (Biomatters Ltd.) was used for routine DNA and protein sequence analysis tasks. BLAST (basic local alignment search tool; <https://blast.ncbi.nlm.nih.gov/Blast.cgi>) and MUSCLE alignment<sup>58</sup> were used to compare protein sequences of the cluster with sequences of characterized protein members of the same family. The prediction of NRPS A domain substrate specificities was performed using NRPSpredictor2.<sup>59</sup> Construction of phylogenetic trees was done with NaPDOS online server.<sup>60</sup>

### 3.5.6 Protein purification and malachite green assay of the FAAL

To purify the FAAL domain, the FAAL-encoding sequence was amplified from genomic DNA of MCy10984 via PCR according to standard protocols<sup>61</sup> using the primers FAAL-FP (AAAAAAACATGTTGGACCTCGCTCGGA) and FAAL-RP (AAAAAAAAGCTTCTACGCCGGCGAG). The PCR construct was cloned into a petM44 expression vector with an N-terminal 6xHis-MBP tag and heterologously expressed in *E. coli* BL21 (DE3) cells. Cultivation was performed in 500 mL LB at 16 °C for 16 h after gene expression by addition of 0.1 mM IPTG. Afterwards the culture was harvested at 3,200 × g for 10 min at 4 °C and the cell pellet was resuspended in lysis buffer (150 mM NaCl, 25 mM Tris, 40 mM imidazole, pH 7.5), sonicated and centrifuged under the same conditions as before. The supernatant was loaded onto a gravity flow column, which contained NiNTA-loaded sepharose, followed by a washing step with lysis buffer prior to elution using imidazole (250 mM). The MBP tag was cleaved using HRV3C protease during overnight dialysis against SEC buffer (150 mM NaCl, 25 mM Tris, pH 7.5) at 4 °C, and removed by a second Ni-NTA chromatography step. The protein was concentrated using a 30 kDa cutoff filter after running through a Superdex 200 16/60 pg column (GE Healthcare Life Sciences). Storage of the protein occurred at -80 °C in 10 % glycerol. Protein purity was determined by SDS-PAGE. The concentration of the protein was determined spectrophotometrically by measuring the extinction coefficient and comparison with the calculated value obtained from PROTPARAM webserver (<http://web.expasy.org/protparam/>).<sup>62</sup>

The substrate specificity of the FAAL domain was determined in a malachite green assay as described elsewhere.<sup>36</sup> 4 μM FAAL was incubated with 2 μM ATP and 2 μM linear fatty acid (C<sub>6</sub>-C<sub>13</sub>) in reaction buffer (150 mM NaCl, 25 mM Tris-HCl, pH 7.5) containing inorganic phosphatase at 37 °C for 1 h. The reaction was stopped by adding 20 μL malachite green solution and the absorption at 625 nm was measured after 10 min. The relative increase of absorption was calculated based on a negative control without fatty acid substrate. Therefore, all reactions were performed in triplicates.

### 3.5.7 Heterologous expression of *comD* and *comG*

First, *comD* and *comG* were amplified from genomic DNA of *C. coralloides* ST201330 via PCR using the primers ORF-2-nt-FP (AAAAAACCATGGTCATGTTTCGTTCAAGAGC) and ORF-2-nt-RP (AAAAAAAAGAATTCCTACTCCTCCAAAGCGCCGTC) for *comD* amplification and ORF-5-nt-FP (AAAAAAACATGTTAGTGGGTAACAACCTCGCGTG) and ORF-5-nt-RP (AAAAAAAAGAA TTCTCAAGGACCCGGCG) for *comG* amplification. The PCR products were subsequently cloned into pET-28b vector according to standard protocols.<sup>61</sup> Next, the vector backbone of pET-28b*comD* was changed by cloning two PCR fragments originating from

pET-28b*comD* and pACYC-177 to generate pAvT1-*comD*. The primers for vector backbone extension were pAvT1-177-FP (AAAAAACCCGGGCATCAGAAGGGCACTGGTGC), pAvT1-177-RP (AAAAAACACACGTGT CGCTCACTGACTCGCT) using pACYC177 as template and pAvT1-28-FP (AAAAAACACGTTGT GCGCCAATCCGGATA) and pAvT1-28-RP (AAAAAACCCGGGGAAACGTTTGGTGGCG) using pET28b-*comD* as template. After independent and combined transformation of *E. coli* BL21 (DE3) with pET-28b*comD*, pET-28b*comG* and pAvT1-*comD*, overnight cultures (LB medium, 37 °C, 200 rpm) were diluted 1:100 (in LB) and protein expression was induced by addition of 0.1 M IPTG. 150 µL aliquots of the diluted overnight cultures were exposed to increasing Cor1183 concentrations from 1 nM to 100 µM and incubated at 30 °C and 200 rpm for 16 h. The optical density at 600 nm was measured to assess resistance development.

### 3.6 References

1. Tacconelli, E. *et al.* Discovery, research, and development of new antibiotics: the WHO priority list of antibiotic-resistant bacteria and tuberculosis. *Lancet Infect. Dis.* **18**, 318–327; 10.1016/S1473-3099(17)30753-3 (2018).
2. World Health Organization. Global Antimicrobial Resistance Surveillance System (GLASS) Report. Early implementation (2018). Available at <http://apps.who.int/iris/bitstream/handle/10665/259744/9789241513449-eng.pdf;jsessionid=97E045EBFE3B29F286CBA3EE8360F0D1?sequence=1> (accessed 15.09.2020).
3. Fischbach, M. A. & Walsh, C. T. Antibiotics for emerging pathogens. *Science* **325**, 1089–1093; 10.1126/science.1176667 (2009).
4. Walsh, C. T. & Wencewicz, T. A. Prospects for new antibiotics: a molecule-centered perspective. *J. Antibiot.* **67**, 7–22; 10.1038/ja.2013.49 (2014).
5. Walsh, C. & Wencewicz, T. *Antibiotics: Challenges, mechanisms, opportunities* (American Society for Microbiology, 2016).
6. van Duin, D. & Bonomo, R. A. Ceftazidime/Avibactam and Ceftolozane/Tazobactam: Second-generation β-Lactam/β-Lactamase Inhibitor Combinations. *Clin. Infect. Dis.* **63**, 234–241; 10.1093/cid/ciw243 (2016).

7. Wunderink, R. G. *et al.* Effect and Safety of Meropenem-Vaborbactam versus Best-Available Therapy in Patients with Carbapenem-Resistant Enterobacteriaceae Infections: The TANGO II Randomized Clinical Trial. *Infect. Dis. Ther.* **7**, 439–455; 10.1007/s40121-018-0214-1 (2018).
8. Saravolatz, L. D. & Stein, G. E. Plazomicin: A New Aminoglycoside. *Clin. Infect. Dis.* **70**, 704–709; 10.1093/cid/ciz640 (2020).
9. Sutcliffe, J. A., O'Brien, W., Fyfe, C. & Grossman, T. H. Antibacterial activity of eravacycline (TP-434), a novel fluorocycline, against hospital and community pathogens. *Antimicrob. Agents Chemother.* **57**, 5548–5558; 10.1128/AAC.01288-13 (2013).
10. Imai, Y. *et al.* A new antibiotic selectively kills Gram-negative pathogens. *Nature*; 10.1038/s41586-019-1791-1 (2019).
11. Luther, A. *et al.* Chimeric peptidomimetic antibiotics against Gram-negative bacteria. *Nature* **576**, 452–458; 10.1038/s41586-019-1665-6 (2019).
12. Hart, E. M. *et al.* A small-molecule inhibitor of BamA impervious to efflux and the outer membrane permeability barrier. *Proc. Natl. Acad. Sci. U.S.A.* **116**, 21748–21757; 10.1073/pnas.1912345116 (2019).
13. Kling, A. *et al.* Targeting DnaN for tuberculosis therapy using novel griselimycins. *Science* **348**, 1106–1112; 10.1126/science.aaa4690 (2015).
14. Newman, D. J. & Cragg, G. M. Natural Products as Sources of New Drugs from 1981 to 2014. *J. Nat. Prod.* **79**, 629–661; 10.1021/acs.jnatprod.5b01055 (2016).
15. Munoz-Dorado, J., Marcos-Torres, F. J., Garcia-Bravo, E., Moraleta-Munoz, A. & Perez, J. Myxobacteria: Moving, Killing, Feeding, and Surviving Together. *Front. Microbiol.* **7**, 781; 10.3389/fmicb.2016.00781 (2016).
16. Reichenbach, H. Myxobacteria, producers of novel bioactive substances. *J. Ind. Microbiol. Biotechnol.* **27**, 149–156 (2001).
17. Wenzel, S. C. & Muller, R. Myxobacteria—‘microbial factories’ for the production of bioactive secondary metabolites. *Mol. Biosyst.* **5**, 567–574; 10.1039/b901287g (2009).
18. Herrmann, J., Fayad, A. A. & Muller, R. Natural products from myxobacteria: novel metabolites and bioactivities. *Nat. Prod. Rep.* **34**, 135–160; 10.1039/c6np00106h (2016).

19. Ikeda, H. *et al.* Complete genome sequence and comparative analysis of the industrial microorganism *Streptomyces avermitilis*. *Nat. Biotechnol.* **21**, 526–531; 10.1038/nbt820 (2003).
20. Goldman, B. S. *et al.* Evolution of sensory complexity recorded in a myxobacterial genome. *Proc. Natl. Acad. Sci.* **103**, 15200–15205; 10.1073/pnas.0607335103 (2006).
21. Oliynyk, M. *et al.* Complete genome sequence of the erythromycin-producing bacterium *Saccharopolyspora erythraea* NRRL23338. *Nat. Biotechnol.* **25**, 447–453; 10.1038/nbt1297 (2007).
22. 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. U.S.A.* **111**, 9259–9264; 10.1073/pnas.1401734111 (2014).
23. Marahiel, M. A. Working outside the protein-synthesis rules: insights into non-ribosomal peptide synthesis. *J. Pept. Sci.* **15**, 799–807; 10.1002/psc.1183 (2009).
24. Weissman, K. J. The structural biology of biosynthetic megaenzymes. *Nat. Chem. Biol.* **11**, 660–670; 10.1038/nchembio.1883 (2015).
25. Staunton, J. & Weissman, K. J. Polyketide biosynthesis: a millennium review. *Nat. Prod. Rep.* **18**, 380–416; 10.1039/a909079g (2001).
26. Weber, T. *et al.* antiSMASH 3.0—a comprehensive resource for the genome mining of biosynthetic gene clusters. *Nucleic Acids Res.* **43**, W237–43; 10.1093/nar/gkv437 (2015).
27. Röttig, M. *et al.* NRPSpredictor2—a web server for predicting NRPS adenylation domain specificity. *Nucleic Acids Res.* **39**, W362–W367; 10.1093/nar/gkr323 (2011).
28. 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).
29. 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; 10.1186/1471-2148-7-78 (2007).
30. Miao, V. *et al.* Daptomycin biosynthesis in *Streptomyces roseosporus*: Cloning and analysis of the gene cluster and revision of peptide stereochemistry. *Microbiology* **151**, 1507–1523; 10.1099/mic.0.27757-0 (2005).

31. Romney M. Humphries *et al.* CLSI Methods Development and Standardization Working Group Best Practices for Evaluation of Antimicrobial Susceptibility Tests. *J. Clin. Microbiol.* **56**; 10.1128/JCM.01934-17 (2018).
32. Weber, T. *et al.* antiSMASH 3.0—a comprehensive resource for the genome mining of biosynthetic gene clusters. *Nucleic Acids Res.* **43**, W237-W243; 10.1093/nar/gkv437 (2015).
33. Arora, P. *et al.* Mechanistic and functional insights into fatty acid activation in *Mycobacterium tuberculosis*. *Nat. Chem. Biol.* **5**, 166–173; 10.1038/nchembio.143 (2009).
34. Trivedi, O. A. *et al.* Enzymic activation and transfer of fatty acids as acyl-adenylates in mycobacteria. *Nature* **428**, 441–445; 10.1038/nature02406. (2004).
35. Mares, J., Hajek, J., Urajova, P., Kopecky, J. & Hrouzek, P. A hybrid non-ribosomal peptide/polyketide synthetase containing fatty-acyl ligase (FAAL) synthesizes the beta-amino fatty acid lipopeptides puwainaphycins in the Cyanobacterium *Cylindrospermum alatosporum*. *PLoS ONE* **9**, e111904; 10.1371/journal.pone.0111904 (2014).
36. McQuade, T. J. *et al.* A nonradioactive high-throughput assay for screening and characterization of adenylation domains for nonribosomal peptide combinatorial biosynthesis. *Anal. Biochem.* **386**, 244–250; 10.1016/j.ab.2008.12.014 (2009).
37. 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).
38. Morikawa, M., Hirata, Y. & Imanaka, T. A study on the structure–function relationship of lipopeptide biosurfactants. *Biochim. Biophys. Acta, Mol. Cell. Biol. Lipids* **1488**(3), 211–218; 10.1016/S1388-1981(00)00124-4 (2000).
39. Arima, K., Kakinuma, A. & Tamura, G. Surfactin, a crystalline peptidelipid surfactant produced by *Bacillus subtilis*: isolation, characterization and its inhibition of fibrin clot formation. *Biochem. Biophys. Res. Commun.* **31**, 488–494 (1968).
40. Peypoux, F. *et al.* Revised structure of mycosubtilin, a peptidolipid antibiotic from *Bacillus subtilis*. *J. Antibiot.* **39**, 636–641 (1986).
41. Maget-Dana, R. & Peypoux, F. Iturins, a special class of pore-forming lipopeptides: biological and physicochemical properties. *Toxicology* **87**, 151–174 (1994).

42. Duitman, E. H. *et al.* The mycosubtilin synthetase of *Bacillus subtilis* ATCC6633: a multifunctional hybrid between a peptide synthetase, an amino transferase, and a fatty acid synthase. *Proc. Natl. Acad. Sci.* **96**, 13294–13299 (1999).
43. Tsuge, K., Akiyama, T. & Shoda, M. Cloning, sequencing, and characterization of the iturin A operon. *J. Bacteriol.* **183**, 6265–6273; 10.1128/JB.183.21.6265-6273.2001 (2001).
44. Yao, S. *et al.* Cloning, sequencing, and characterization of the genetic region relevant to biosynthesis of the lipopeptides iturin A and surfactin in *Bacillus subtilis*. *Curr. Microbiol.* **47**, 272–277 (2003).
45. Yan, F. *et al.* Biosynthesis and Heterologous Production of Vioprolides: Rational Biosynthetic Engineering and Unprecedented 4-Methylazetidincarboxylic Acid Formation. *Angew. Chem. Int. Ed. Engl.* **57**, 8754–8759; 10.1002/anie.201802479 (2018).
46. Auerbach, D., Yan, F., Zhang, Y. & Müller, R. Characterization of an Unusual Glycerate Esterification Process in Vioprolide Biosynthesis. *ACS Chem. Biol.* **13**, 3123–3130; 10.1021/acscchembio.8b00826 (2018).
47. Chen, H., Hubbard, B. K., O'Connor, S. E. & Walsh, C. T. Formation of beta-hydroxy histidine in the biosynthesis of nikkomycin antibiotics. *Chem. Biol.* **9**, 103–112 (2002).
48. Cortina, N. S., Krug, D., Plaza, A., Revermann, O. & Muller, R. Myxoprincomide: a natural product from *Myxococcus xanthus* discovered by comprehensive analysis of the secondary metabolome. *Angew. Chem. Int. Ed. Engl.* **51**, 811–816; 10.1002/anie.201106305 (2012).
49. Silver, L. L. Appropriate Targets for Antibacterial Drugs. *Cold Spring Harb Perspect. Med.* **6**, a030239; 10.1101/cshperspect.a030239 (2016).
50. Salomón, R. A. & Farías, R. N. The peptide antibiotic microcin 25 is imported through the TonB pathway and the SbmA protein. *J. Bacteriol.* **177**, 3323–3325; 10.1128/jb.177.11.3323-3325.1995 (1995).
51. Corbalan, N. *et al.* Functional and Structural Study of the Dimeric Inner Membrane Protein SbmA. *J. Bacteriol.* **195**, 5352–5361; 10.1128/JB.00824-13 (2013).
52. Feng, Z. *et al.* Cloning and Functional Characterization of Putative *Escherichia coli* ABC Multidrug Efflux Transporter YddA. *J. Microbiol. Biotechnol.* **30**, 982–995; 10.4014/jmb.2003.03003 (2020).

53. Novikova, M. *et al.* The *Escherichia coli* Yej Transporter Is Required for the Uptake of Translation Inhibitor Microcin C. *J. Bacteriol.* **189**, 8361–8365; 10.1128/JB.01028-07 (2007).
54. Cattoir, V. & Guérin, F. How is fosfomycin resistance developed in *Escherichia coli*? *Future Microbiol.* **13**, 1693–1696; 10.2217/fmb-2018-0294 (2018).
55. Takahata, S. *et al.* Molecular mechanisms of fosfomycin resistance in clinical isolates of *Escherichia coli*. *Int. J. Antimicrob. Agents* **35**, 333–337; 10.1016/j.ijantimicag.2009.11.011 (2010).
56. Shiver, A. L. *et al.* A Chemical-Genomic Screen of Neglected Antibiotics Reveals Illicit Transport of Kasugamycin and Blasticidin S. *PLOS Genetics* **12**, e1006124; 10.1371/journal.pgen.1006124 (2016).
57. Maio, A., Brandi, L., Donadio, S. & Gualerzi, C. O. The Oligopeptide Permease Opp Mediates Illicit Transport of the Bacterial P-site Decoding Inhibitor GE81112. *Antibiot.* **5**, 17; 10.3390/antibiotics5020017 (2016).
58. Madeira, F. *et al.* The EMBL-EBI search and sequence analysis tools APIs in 2019. *Nucleic Acids Res.* **47**, W636-W641; 10.1093/nar/gkz268 (2019).
59. Rottig, M. *et al.* NRPSpredictor2-a web server for predicting NRPS adenylation domain specificity. *Nucleic Acids Res.* **39**, W362-7; 10.1093/nar/gkr323 (2011).
60. Ziemert, N. *et al.* The natural product domain seeker NaPDoS: a phylogeny based bioinformatic tool to classify secondary metabolite gene diversity. *PLoS ONE* **7**, e34064; 10.1371/journal.pone.0034064 (2012).
61. Sambrook, J. & Russell, D. W. *Molecular cloning: A laboratory manual* (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 2001).
62. Gasteiger, E. *et al.* in *The proteomics protocols handbook*, edited by J. M. Walker (Humana Press, Totowa, N.J., 2005), pp. 571–607.
63. Quevillon, E. *et al.* InterProScan: protein domains identifier. *Nucleic Acids Res.* **33**, W116-W120; 10.1093/nar/gki442 (2005).

## 3.7 Supplementary Information

### 3.7.1 *In silico* analysis of the corramycin BGC

#### 3.7.1.1 General organization of the corramycin BGC

The cluster borders of the corramycin BGC were determined by comparing genome sequences of the two producer strains, *C. coralloides* MCy10984 and *C. coralloides* ST201330, with the non-producer strain *C. coralloides* DSM2259 using MUSCLE alignment (Supplementary Figure 3).<sup>58</sup> The BGC represents a well-defined insertion into the genome of *C. coralloides* and has a size of 59 kb with a GC content of 71 %. Fifteen genes are presumably involved in corramycin biosynthesis. Based on the coding strand on which the genes are located and the small intergenic regions between adjacent genes on the same coding strand, we assume that the BGC is organized in five putative transcriptional units: *comA-C*, *comF-D* (reverse direction), *comG*, *comH* and *comI-O*.

#### 3.7.1.2 Analysis of *comA*

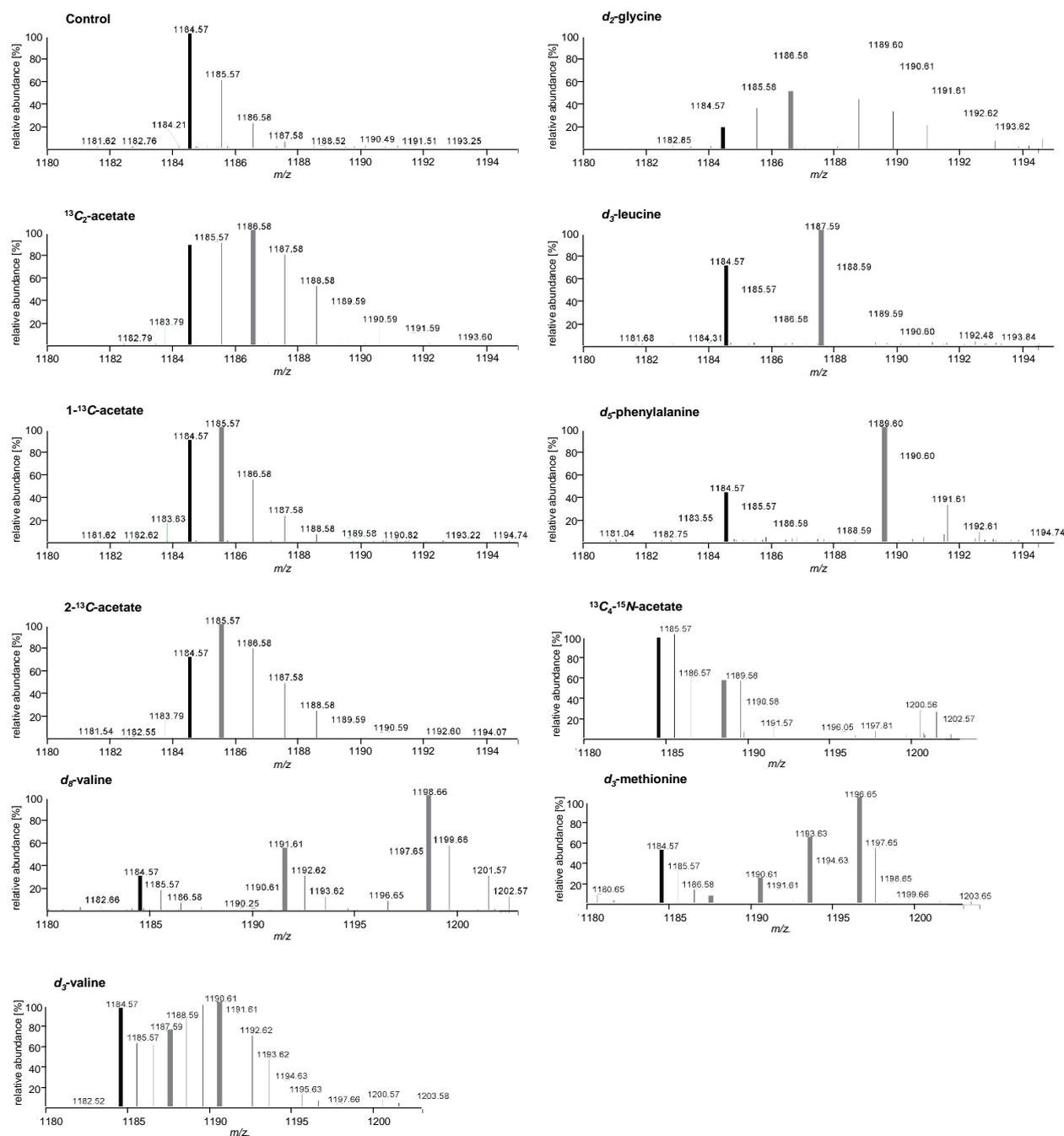
BLAST analysis of *comA* found cupin-like transcription factors as closest homologs. However, *InterPro* scan<sup>63</sup> showed that ComA presumably contains a JmjC domain, which was shown to be involved in the hydroxylation of aspartate residues.<sup>64</sup> The JmjC domain was also described in JmjC domain-containing factor inhibiting hypoxia, which shows homology to cupin-like metalloenzyme domains.<sup>65</sup> Furthermore, ComA showed homology to clavamate synthase-like 2-oxoglutarate-dependent dioxygenase with an active center containing Fe (II). Including this, ComA is most probably a hydroxylase. Likewise, *comF*, *comH* and *comI* encode for various hydroxylases. The mentioned hydroxylases are presumably catalyzing the  $\beta$ -hydroxylation of L- or D-histidine, L-valine and the two hydroxylations of  $\beta$ -alanine and the PKS incorporated acetate moiety.

### 3.7.2 Analysis of methyltransferase domains

The assembly line includes three methyltransferase (MT) domains. Two *N*-MT domains were identified in modules 3 and 6 embedded in the respective A domains. The third MT is located downstream of module 4 and was not annotated automatically by standard NRPS prediction tools. A BLAST search of module 4 revealed a stretch of approximately 300 amino acids as an FkbM-like methyltransferase. The three known members of this MT protein family are all stand-alone *O*-MTs.<sup>66</sup> Phylogenetic analysis of the three MTs in the corramycin assembly line with 54 MTs using the NaPDoS online server<sup>60</sup> showed that the *N*-MT of module 6 clusters with NRPS-associated *N*-MTs (Supplementary Figure 8) that have been shown to catalyze methylation of amide nitrogens.<sup>67,68</sup> This is in accordance with the corramycin structure, since the A domain of module 6 is specific for

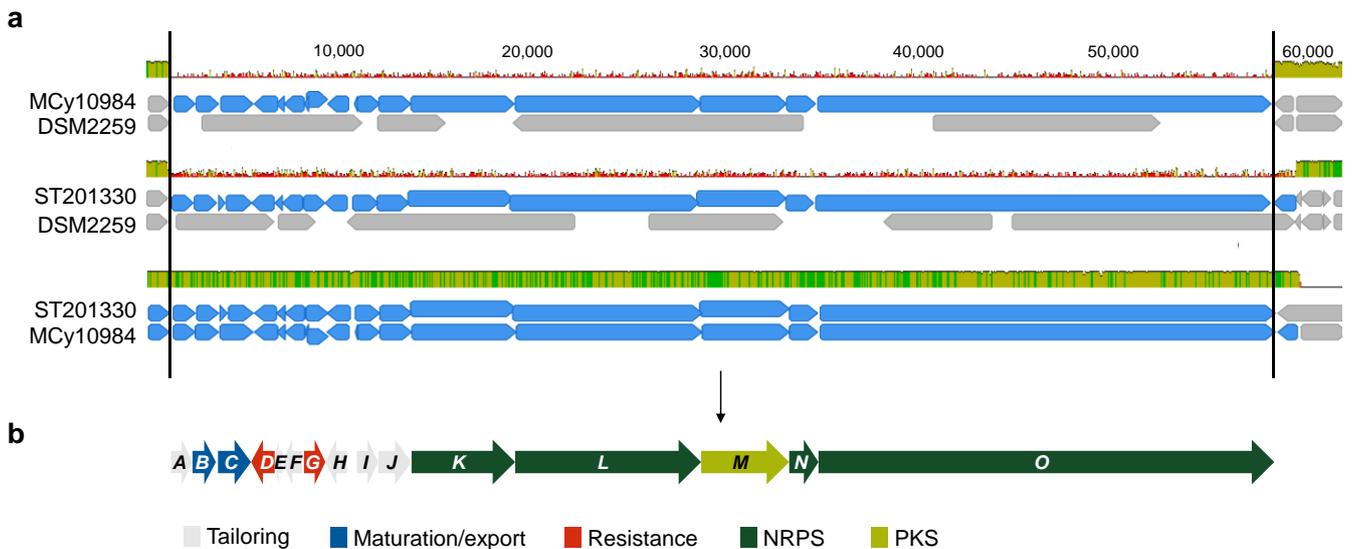
phenylalanine (Table 3) and the phenylalanine moiety is *N*-methylated. The designated *N*-MT of module 3 does not cluster with any known MT domain found in natural product biosynthetic pathways. Thus, we assume that this domain catalyzes the *N*-methylation of histidine. So far no NRPS-associated *in cis* *N*-MT acting on an aromatic substrate was described. Furthermore, we assume that the stand-alone *O*-MT of module 4 catalyzes the *O*-methylation of the  $\beta$ -alanine moiety.

### 3.7.3 Supplementary Figures

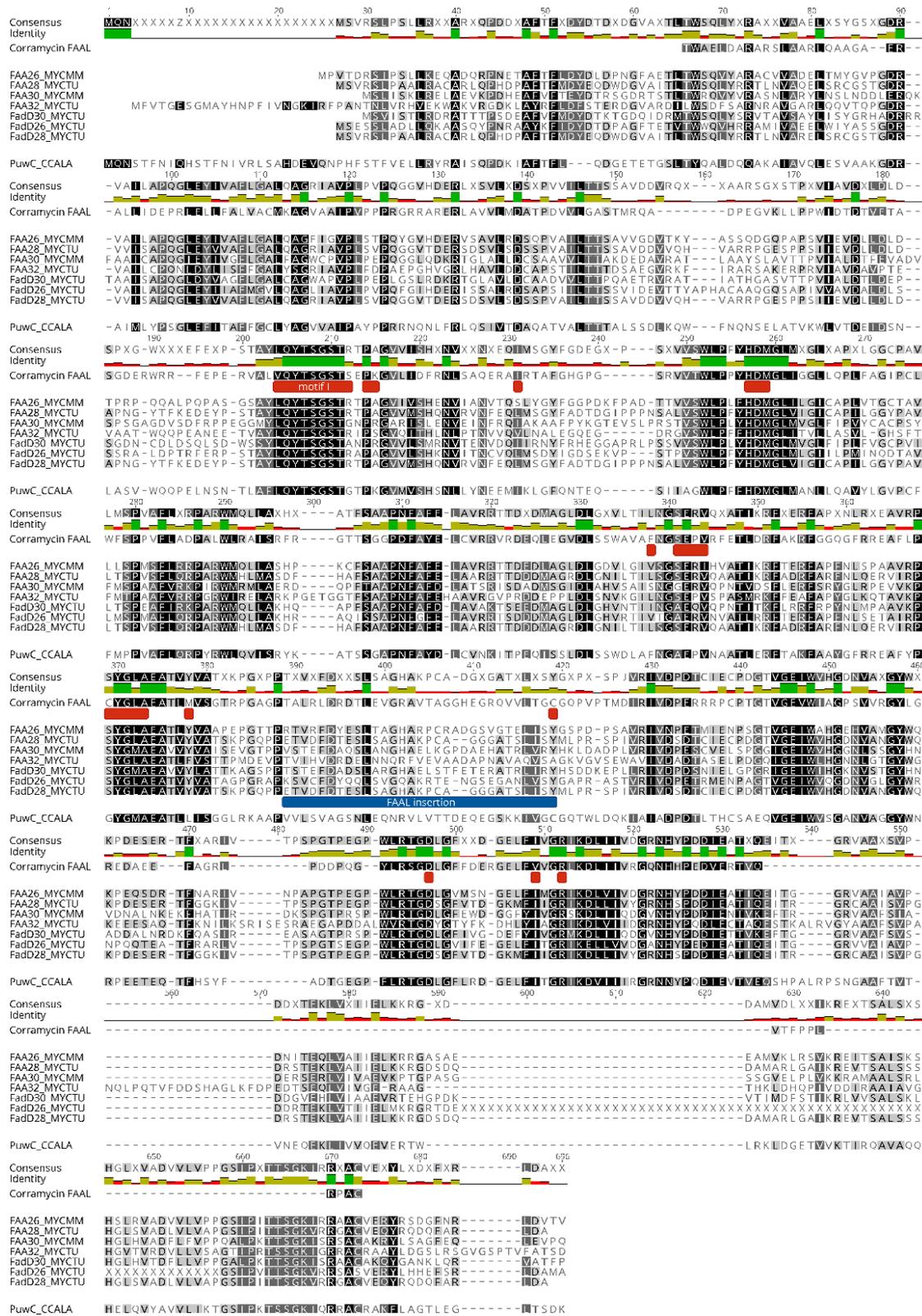


**Supplementary Figure 1 | Observed mass shifts after supplementation of isotope-labelled precursors during strain cultivation.** The Cor1183 mass peak (thick black line) and the new isotope peaks (thick grey lines) are highlighted.

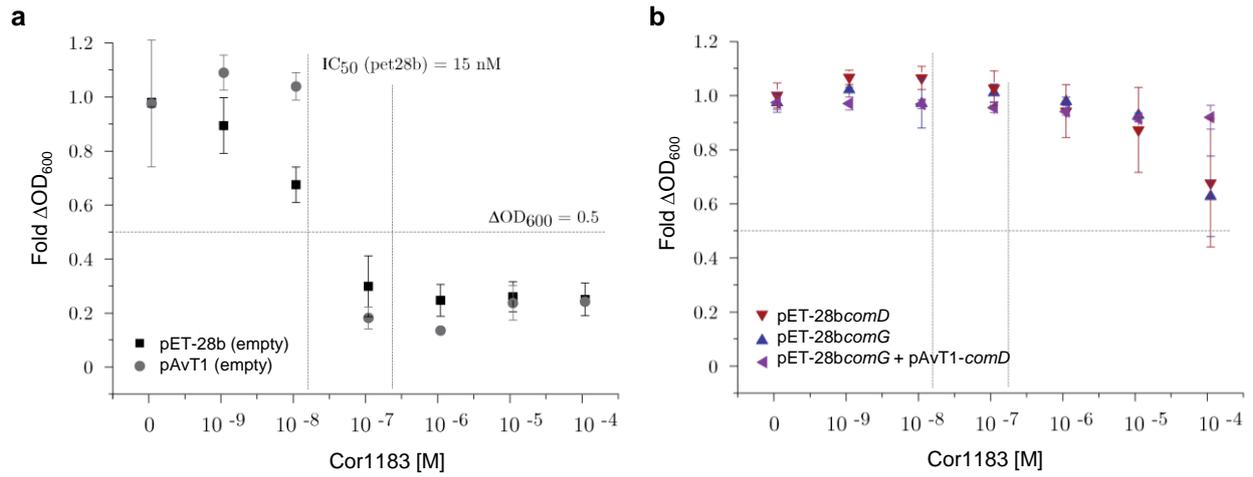




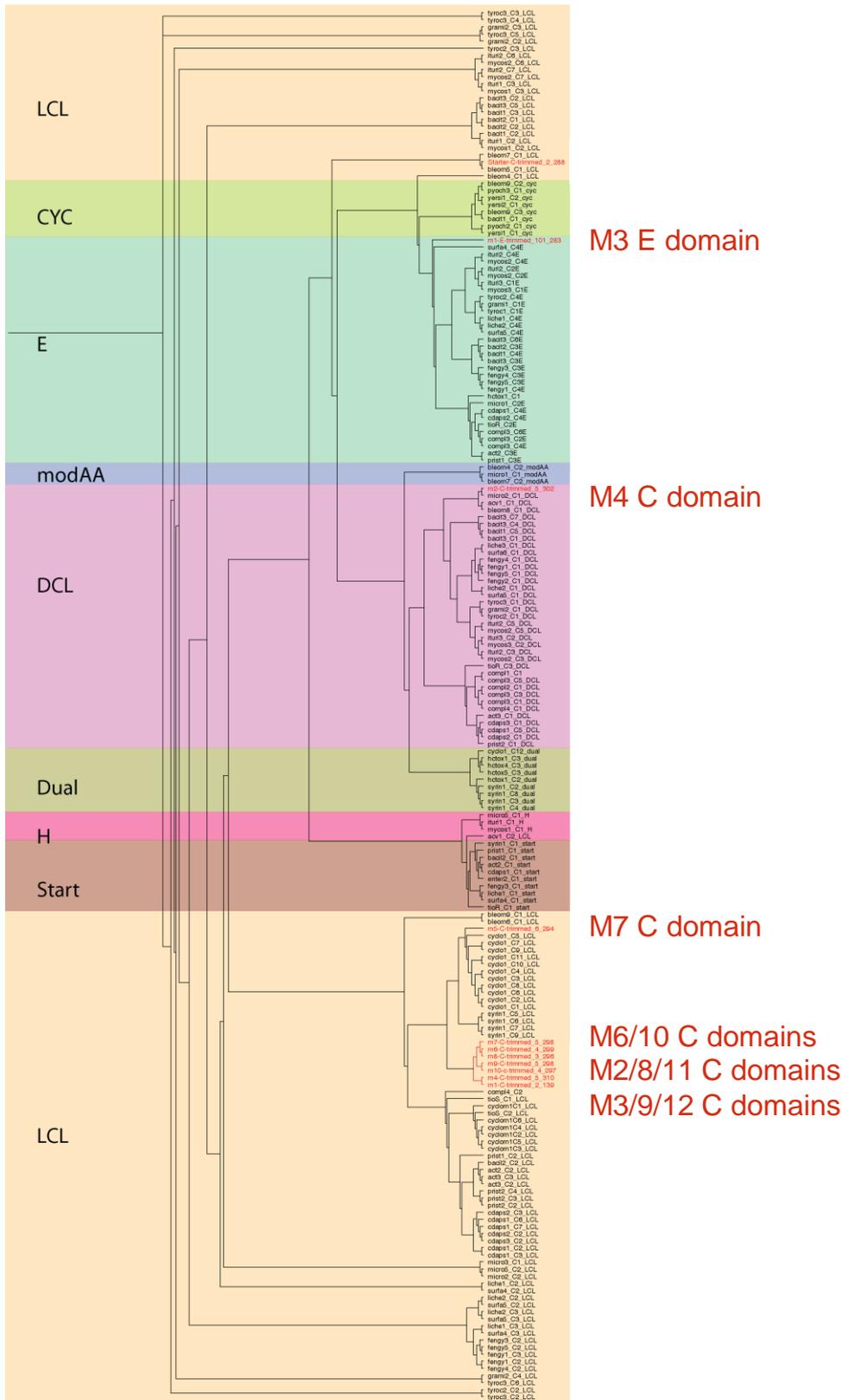
**Supplementary Figure 3 | Determination of the corramycin gene cluster borders. a:** MUSCLE alignments of the BGCs and adjacent genomic regions from *C. coralloides* Mcy10984 and ST201330 with the genome of the non-producer strain *C. coralloides* DSM2259. Arrows show putative genes that belong (blue) or do not belong (grey) to the corramycin BGC based on the respective alignment. The cluster borders were determined by integrating data from all three alignments and are depicted as black lines. **b:** Corramycin BGC and putative functions of the encoded proteins. Parts of the figure were generated with Geneious (version 2020.0 created by Biomatters; available from <https://www.geneious.com>).



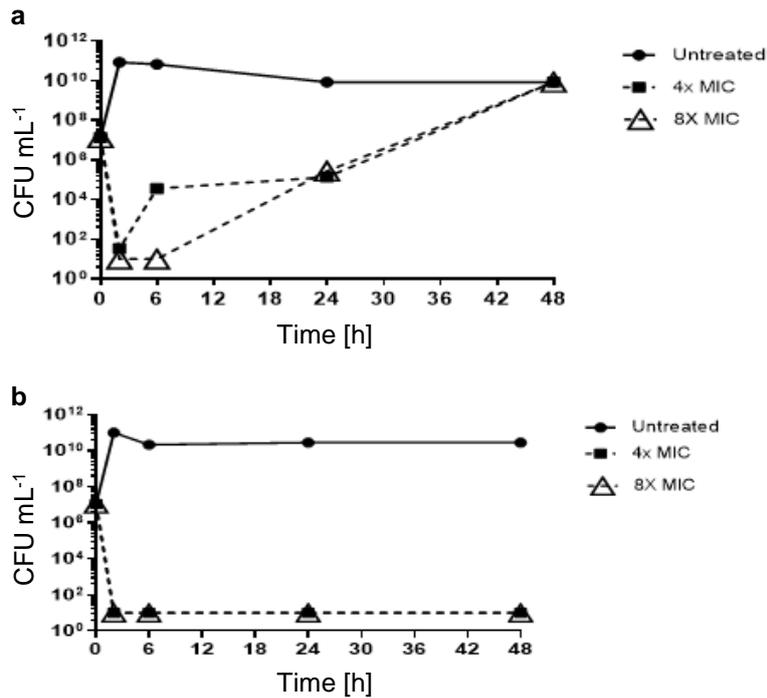
Supplemental Figure 4 | MUSCLE alignment of the corramycin FAAL with other FAALs. Active side residues (red bars) were identified using BLAST. The sequence insertion that is characteristic for FAALs<sup>33</sup> and missing in fatty acid CoA-ligases (FACLs) is shown in blue.



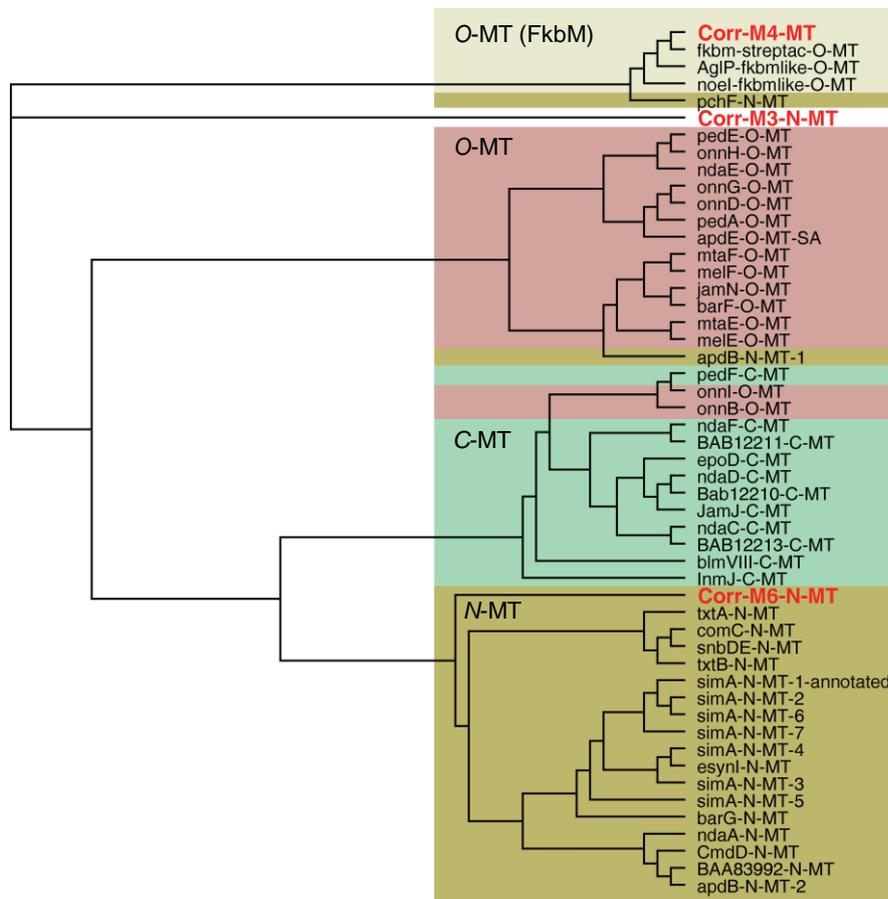
**Supplementary Figure 5 | Heterologous expression of *comD* and *comG* in *E. coli* BL21 (DE3) leads to resistance towards Cor1183. a: Empty vector control. b: Corramycin resistance increased about three magnitudes upon expression of *comD* or *comG*.**



Supplementary Figure 6 | Phylogenetic tree of the corramycin assembly line C and E domains. The phylogenetic tree was generated by NapDos online server.<sup>60</sup>



Supplementary Figure 7 | *E. coli* ATCC25922 time-kill curve of Cor1183 in MHB (a) or M9 medium (b).



Supplementary Figure 8 | Phylogenetic analysis of the corramycin assembly line MT domains. The phylogenetic tree was generated by NapDos online server.<sup>60</sup>

## 3.7.4 Supplementary Table

Supplementary Table 1 | MIC values of Cor1183 against all bacterial strains tested in this study.

Gram-negative species			Gram-positive species		
Strain	MIC ( $\mu\text{g mL}^{-1}$ )		Strain	MIC ( $\mu\text{g mL}^{-1}$ )	
<i>C. freundii</i>	255041	>64	<i>E. faecium</i>	A6349	>64
	ATCC 8090	>64		DSM 17050	64
<i>E. cloacae</i>	17059482	>64	<i>E. faecalis</i>	1069 VanB	16
	DSM 46348	32	<i>S. aureus</i>	11540	>64
<i>E. aerogenes</i>	DSM 12058	>64		ATCC33592	>64
<i>K. pneumoniae</i>	1705966	>64		NRS 643	>64
	1705949	64	<i>S. pneumoniae</i>	02J1175	>64
	ATCC 13883	16-64		ATCC700671	>64
<i>A. baumannii</i>	ATCC 19606	64			
	1705943	>64			
	1705936	>64			
<i>P. aeruginosa</i>	ATCC 9027	>64			
	ATCC 27853	>64			
	DSM 46317	>64			
	ATCC BAA-	>64			
	ATCC27853	>64			
	1705886	>64			
	1705904	>64			
	PAO1	>64			
<i>B. bronchiseptica</i>	NCTC 8344	>64			
<i>B. cepacia</i>	ATCC 25416	>64			
<i>S. maltophilia</i>	ATCC 13637	>64			
	255074	>64			
<i>S. marcescens</i>	255067	>64			
	ATCC 13880	>64			
<i>P. mirabilis</i>	ATCC 29906	>64			
<i>P. vulgaris</i>	DSM 46228	>64			

### 3.8 Supplementary Information References

64. Lando, D. *et al.* FIH-1 is an asparaginyl hydroxylase enzyme that regulates the transcriptional activity of hypoxia-inducible factor. *Genes Dev.* **16**, 1466–1471; 10.1101/gad.991402 (2002).
65. Noma, A. *et al.* Expanding role of the jumonji C domain as an RNA hydroxylase. *J. Biol. Chem.* **285**, 34503–34507; 10.1074/jbc.M110.156398 (2010).
66. Motamedi, H. *et al.* Characterization of methyltransferase and hydroxylase genes involved in the biosynthesis of the immunosuppressants FK506 and FK520. *J. Bacteriol.* **178**, 5243–5248 (1996).
67. 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).
68. Johnson, E. G. *et al.* 4-Nitrotryptophan is a substrate for the non-ribosomal peptide synthetase TxtB in the thaxtomin A biosynthetic pathway. *Mol. Microbiol.*; 10.1111/j.1365-2958.2009.06780.x (2009)

# Chapter 4

## Heterologous production of corramycins and their derivatives to allow for efficient semi-synthesis of improved congeners

### Preliminary author list

Sebastian Groß<sup>1,2,3</sup>, Domen Pogorevc<sup>1,2,3</sup>, Stephane Renard<sup>4</sup> and Rolf Müller<sup>2,3,5</sup>

<sup>1</sup> first author; <sup>2</sup> Helmholtz Center for Infection Research (HZI); <sup>3</sup> Helmholtz-Institute for Pharmaceutical Research Saarland (HIPS); <sup>4</sup> Evotec; <sup>5</sup> corresponding author

This manuscript is in preparation. The work presented herein was performed by HZI/HIPS, Sanofi and Evotec. The author list has to be finalized regarding the number and order of authors. The author list and author contribution section is in a preliminary state and will be finalized soon after announcement of the authors from Sanofi und Evotec. The author contribution section is limited to the authors from the HIPS. All experiments that are not mentioned in the author contribution section was performed by Sanofi and Evotec. Missing data, e.g. the total synthesis route of the C-terminally modified corramycin derivative with improved antibacterial spectrum, will be provided at a later stage. The method section only contains data from the work performed at HIPS.

## **Contributions to the presented work**

### **Author contributions**

Sebastian Groß performed the *in silico* analysis of the native corramycin biosynthetic gene cluster (BGC) and the design, assembly and heterologous expression of the modified BGC. Sebastian Groß and Domen Pogorevc performed *in silico* NRPS engineering experiments, cloned the engineered constructs and did heterologous expression experiments to obtain truncated corramycins. Sebastian Groß and Domen Pogorevc wrote the manuscript.

### **Contributions by others**

The project was supervised by Rolf Müller.

### **Acknowledgement**

We are especially thankful to Silke C. Wenzel for her contributions to the design of the modified corramycin BGC.

## 4 Heterologous production of corramycins and their derivatives to allow for efficient semi-synthesis of improved congeners

### 4.1 Abstract

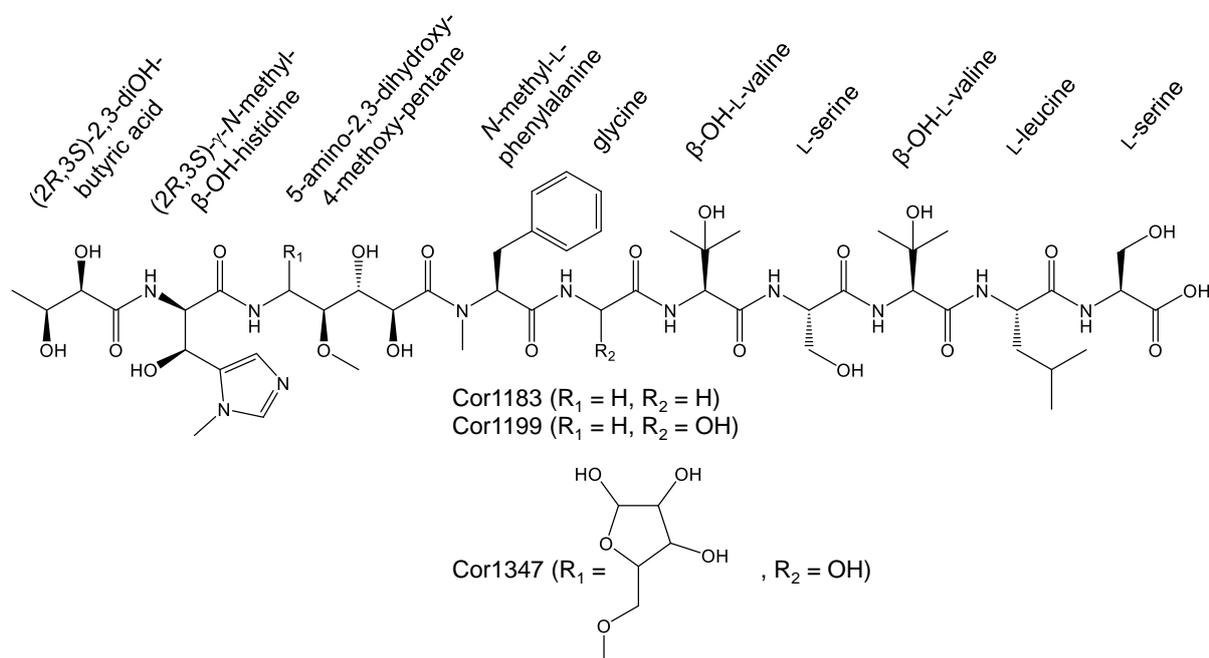
Corramycin is a NRPS-PKS product produced by the myxobacterium *Coralloccoccus coralloides* with promising anti-Gram-negative activity e.g. against *E. coli*. The total synthesis of a new corramycin derivative with modifications on the N-terminus showed improved antibacterial activity in a mouse model of *E. coli* septicemia compared to native corramycin. The limiting factor in corramycin development is the tedious chemical synthesis of the unprecedented (2*R*,3*S*)- $\gamma$ -*N*-methyl- $\beta$ -hydroxy-histidine building block. Herein, we overcame this problem by producing the C-terminal part of corramycin via fermentation. Initially, we established a heterologous production platform for full-length corramycin in *M. xanthus* DK1622. Subsequently, NRPS engineering was applied to partially truncate the 3' end of *comO* at specific positions in the linker regions between C-A, A-T and T-C domains and to test different TE, C and C/E domains as terminal releasing domains. This approach led to heterologous production of three truncated corramycin derivatives at significant yield and applicable for semi-synthesis.

### 4.2 Introduction

Emerging bacterial resistance against all clinically used antibiotic classes has become a growing global health concern in the last decades.<sup>1,2</sup> Especially troublesome are the pathogens from the ESKAPE (*Enterococcus faecium*, *Staphylococcus aureus*, *Klebsiella pneumonia*, *Actinobacter baumannii*, *Pseudomonas aeruginosa* and *Enterobacter* species) panel since they cause severe nosocomial infections with only a few last resort antibiotics left as viable treatment options.<sup>3</sup>

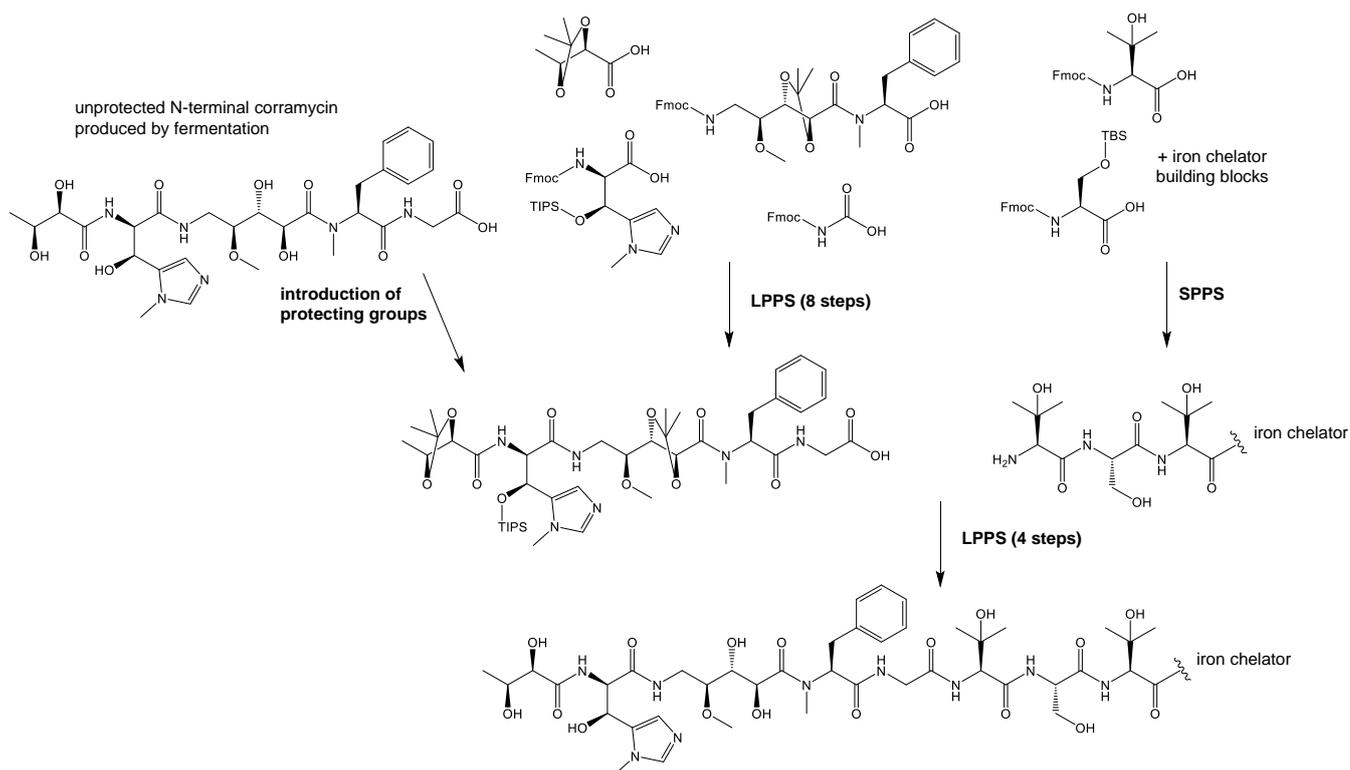
The corramycins (Figure 1) belong to a recently discovered new class of linear peptide antibiotics produced by *Coralloccoccus coralloides* strains (partly unpublished data; Tesmar *et al.*,<sup>4</sup> manuscript in preparation, see chapter 3). One congener, corramycin 1183 (Cor1183), showed strong anti-Gram-negative activity against several *E. coli* strains with minimal inhibitory concentrations (MIC) ranging from 1  $\mu\text{g mL}^{-1}$  to 32  $\mu\text{g mL}^{-1}$ . Furthermore, antibacterial activity against *Salmonella typhimurium* (4  $\mu\text{g mL}^{-1}$ ), *Enterobacter cloacae* (32  $\mu\text{g mL}^{-1}$ ) and *Enterococcus faecium* (16  $\mu\text{g mL}^{-1}$ ) was described. Moreover, the efficacy to treat bacterial infections with Cor1183 was proven in an *in vivo* experiment using a mouse model of *E. coli* septicemia. Structurally, corramycin consists of eight  $\alpha$ -amino acids, some of which are tailored, including two  $\beta$ -hydroxy-L-valines, *N*-methyl-L-phenylalanine and the unprecedented (2*R*,3*S*)- $\gamma$ -*N*-methyl- $\beta$ -hydroxy-histidine. The

N-terminus is capped by a dihydroxy butyric acid and the structure also features an unusual 5-amino-2,3-dihydroxy-4-methoxy-pentanoyl moiety.



**Figure 1 | Structure of corramycins.** Cor1183, Cor1199 and Cor1347 were described previously. Scheme adapted from manuscript in preparation (chapter 3).

Tesmar and coworkers also described a total synthesis route of Cor1183, which turned out to be laborious with the main limitation being the synthesis of the (2*R*,3*S*)- $\gamma$ -*N*-methyl- $\beta$ -hydroxy-histidine building block (unpublished data; manuscript in preparation, see chapter 3). However, despite the tedious chemical synthesis of Cor1183, numerous novel corramycin derivatives have been produced via total synthesis and tested for their bioactivities (unpublished data; personal communication with Stephane Renard). Those experiments showed that the two C-terminal amino acids (*L*-leucine and *L*-serine) can be exchanged without decrease in bioactivity. Derivatives, in which those amino acids were replaced, e.g. by an iron chelator, showed superior anti-Gram-negative activity and a broadened activity spectrum compared to Cor1183. Since the chemical synthesis route for those derivatives allows for the introduction of the unprotected N-terminal part of Cor1183 up to the glycine moiety, one could alternatively imagine the production of this truncated corramycin fragment by fermentation (Figure 2). However, the production titer in the native producer strains was found low with only a few  $\mu\text{g L}^{-1}$ . Poor genetic accessibility of the few producing strains and time-intensive medium optimization makes heterologous production attractive as an alternative approach.



**Figure 2 | Total synthesis and hypothetical semi-synthetic route of iron chelator-linked corramycin derivatives.** The hypothetical semi-synthetic approach using an N-terminal corramycin fragment produced by fermentation is shown on the left. Total synthesis of the protected N-terminal corramycin fragment is achieved by an eight step liquid-phase peptide synthesis (LPPS) and the C-terminal part including the iron chelator is synthesized by solid-phase peptide synthesis (SPPS). Full-length corramycin derivative attached to the iron chelator is obtained by a four steps LPPS. This scheme was generated based on personal communication with Stephane Renard.

*Myxococcus xanthus* was already used for the heterologous expression of numerous myxobacterial biosynthetic gene clusters (BGCs).<sup>5,6</sup> Relatively fast growth with a doubling time of under 6 h, transformation via electroporation and the availability of established inducible and constitutive promoters make *M. xanthus* a suitable candidate for heterologous production of various myxobacterial compounds. Furthermore, growth in production scales up to 1,000 L was achieved.<sup>7</sup> The production titers varied from barely detectable quantities up to more than 1 g L<sup>-1</sup> depending on the product.<sup>8-10</sup> Moreover, genetic manipulations of the BGCs opens up the opportunity to direct the production profile towards certain preferred derivatives or to enable production of new derivatives.

The corramycin BGC includes 15 genes, 5 of which encode a 12-modular non-ribosomal peptide synthetase (NRPS)-polyketide synthase (PKS) hybrid assembly line. The functionality of those multimodular megaenzymes is described in detail elsewhere.<sup>11-13</sup> Shortly, in NRPSs each module is subdivided into independent domains catalysing a single reaction: Adenylation (A) domains activate amino acids using ATP, thiolation (T) domains tether the activated amino acid or the growing peptide and condensation (C) domains catalyse peptide bond formation. A terminal thioesterase (TE)

domain is usually in charge of product release from the assembly line.<sup>14</sup> However, in certain cases C domains have been shown to perform this action in bacterial NRPSs instead.<sup>15</sup>

Various approaches to NRPS engineering have been tested in the past, such as replacement of A or A-T didomains, C-A-T or T-C-A tridomains, deletion and insertion of modules, subdomain swapping in A domains and changing the ‘nonribosomal code’ by changing single amino acid residues in the binding pocket of A domains.<sup>16</sup> However, their success was limited as the outcome of the majority of studies resulted in abolished or substantially reduced production titers.<sup>16–18</sup> One possible explanation for this is the disruption of the protein-protein interactions which were shown in the NRPS structural biology to be essential for proper functioning.<sup>19</sup> Recently, the use of exchange units (XUs)<sup>20</sup> and the exchange unit condensation domain (XUC)<sup>21</sup> concept represented a breakthrough in NRPS engineering. A great variety of novel peptides was produced after combining NRPS building blocks of different origin, using specific fusion points inside C-A linker regions or inside C pseudo-subdomains, respectively.

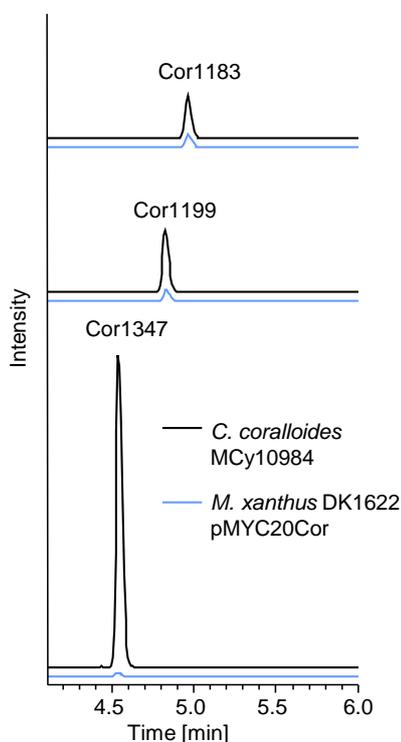
Herein, we describe the design, assembly and heterologous expression of a modified, synthetic BGC for the production of corramycin in *M. xanthus* DK1622. Initially, we observed very low production titers of previously described full-length corramycins. However, since the chemical synthesis of the N-terminal part of corramycin was still not efficient due to the hard-to-synthesize (2*R*,3*S*)- $\gamma$ -*N*-methyl- $\beta$ -hydroxy-histidine building block, we then aimed for the heterologous production of C-terminally truncated corramycins to establish a semi-synthetic synthesis route for corramycins and their semi-synthetic congeners. Therefore, we deleted parts of the NRPS-PKS assembly line using various new engineering sites in the C-A, A-T and T-C didomain linker regions. Finally, we could confirm production of three truncated corramycins, one (Cor666) of which can be used in a semi-synthetic approach to produce corramycin, also providing a powerful platform to generate new lead structures with potent *in vivo* activity in mice.

## 4.3 Results & discussion

### 4.3.1 Design, assembly and heterologous expression of the corramycin biosynthetic gene cluster in *Myxococcus xanthus* DK1622

A modified BGC was designed *in silico* for the heterologous production of corramycin in *M. xanthus* DK1622 (described in Supplementary Information, see Supplementary Figure 1 and Supplementary Table 4). The template sequence originated from *C. coralloides* MCy10984, including fifteen biosynthetic genes *comA-comO* as described previously (manuscript in preparation; see chapter 3).<sup>4</sup> The modified BGC was split into fragments which were chemically synthesized (Supplementary Table 5), because of the large size, high GC content and repetitive

sequence segments in the largest NRPS gene of the cluster, *comO* (Supplementary Figure 2). Assembly of the cluster fragments was done by a combination of *in vivo* transformation-associated recombination (TAR)<sup>22</sup> in yeast and a previously described *in vitro* three-step restriction/ligation cloning strategy using *BsaI*.<sup>23</sup> In order to minimize the risk of unspecific recombination during TAR assembly, the *comO* gene was replaced by a dummy sequence. The fragments of the *comO* harbouring sequence repeats were assembled separately *in vitro* prior to its integration into the remaining part of the cluster. For TAR cloning and heterologous expression in *M. xanthus* DK1622, we used the pMYC vector system that was described elsewhere. (manuscript under revision; see chapter 2) The two operons of the modified BGC, CorOp1 and CorOp2, were first assembled separately on two different pMYC derivatives, yielding pMYC20CorOp1 and pMYC21CorOp2, and subsequently stitched together to generate the final expression construct pMYC20Cor. All cloning steps are summarized in the Supplementary Information (including Supplementary Figure 3 and Supplementary Table 6). The final expression construct was integrated into the *M. xanthus* DK1622 genome via the *Mx8* phage attachment site.<sup>24</sup> Subsequent UPLC-HRMS analysis confirmed the heterologous production of Cor1183, Cor1199 (hydroxylated Cor1183) and Cor1347 (hydroxylated and additionally glykosylated Cor1183) based on retention times and high-resolution masses (Figure 3).



**Figure 3 | UPLC-HRMS analysis of heterologously produced corramycin derivatives in *M. xanthus* DK1622.** EICs 1184.56 [M+H]<sup>+</sup> (top), 1200.56 [M+H]<sup>+</sup> (middle) and 1348.59 [M+H]<sup>+</sup> (bottom) of methanolic crude extracts from native producer *C. coralloides* MCy10984 (shown in black) and heterologous producer *M. xanthus* DK1622 pMYC20Cor (shown in blue).

However, the production yield was too low for quantification, as only trace amounts of Cor1183 could be detected. Medium optimization and various substrate feeding efforts did not result in improvement of production titer (data not shown). Furthermore, we tested the MIC of Cor1183 against *M. xanthus* DK1622 wild-type and the heterologous producer strain, which harbors the corramycin BGC including the resistance genes *comD* and *comG*. The highest tested concentration was 64  $\mu\text{g mL}^{-1}$ . By doing this we wanted to exclude the possibility of a potential self-toxicity effect leading to a growth inhibition of the host strain. Even after supplementation of polymyxin B to permeabilize the cell membrane and to potentially increase intracellular Cor1183 concentration, no growth inhibition was observed in the wild-type strain or the heterologous host. However, we were not able to precisely determine the intracellular concentration of Cor1183 in this experiment and thus a potential risk that the production is limited by a negative effect on cell metabolism via the intracellularly produced Cor1183 can not be excluded.

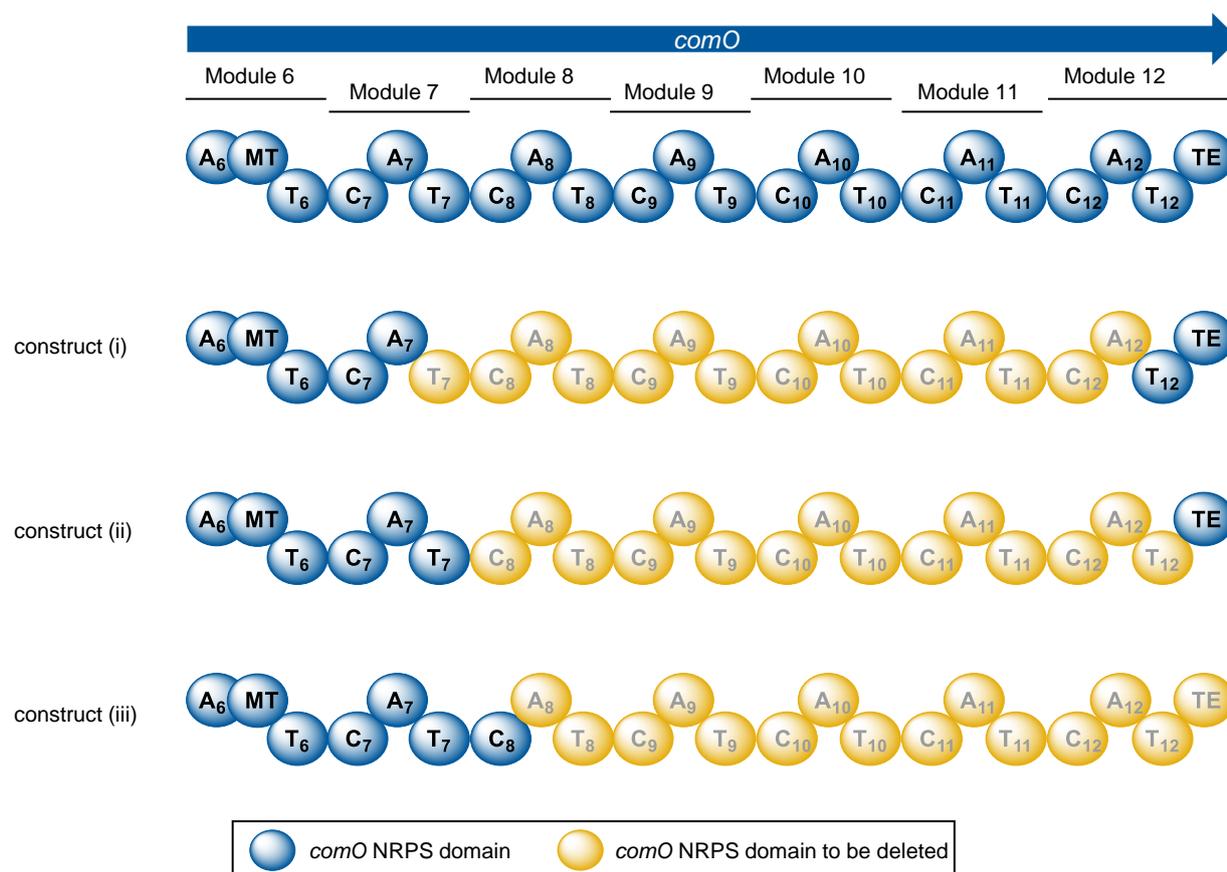
Low supply of unusual or modified amino acids (e.g.  $\beta$ -hydroxy-L-valine) required for biosynthesis of Cor1183 could be another potential reason for the low production in *M. xanthus* DK1622. Feeding experiments performed by von Tesmar and co-workers showed that modules 8 and 10 initially incorporate L-valine which is then hydroxylated, either while being tethered to the assembly line or after release of the corramycin precursor.<sup>4</sup> Gene(s) responsible for this modification have thus far not been identified, nevertheless we cannot exclude that  $\beta$ -hydroxy-L-valine is produced by one of the putative hydroxylases present in the corramycin BGC. The same  $\beta$ -hydroxy-L-valine moiety was also found in myxoprincomide,<sup>25</sup> a class of nonribosomal peptides natively produced by *M. xanthus* DK1622. Interestingly, putative genes for the biosynthesis of the  $\beta$ -hydroxy-L-valine moiety have neither been identified in the BGC of myxoprincomide,<sup>25</sup> indicating that the required genetic information could be located elsewhere in the genome. If the same proteins are indeed involved in the biosynthesis of  $\beta$ -hydroxy-L-valine in corramycin and myxoprincomide, then the corresponding genes are already present in *M. xanthus* DK1622. However, since only low amounts of myxoprincomide were produced by *M. xanthus* DK1622,<sup>25</sup> the expression level of those genes and thus the production of  $\beta$ -hydroxy-L-valine might be a limiting factor for the biosynthesis of corramycin and myxoprincomide.

Another potential reason for the low production titer of corramycin might be linked to a previously hypothesized pre-drug mechanism. Based on the experiment of a fatty acid activation by the FAAL domain of module 1, von Tesmar and co-workers assumed the intracellular presence of acylated corramycin derivatives.<sup>4</sup> However, even though the FAAL domain was shown to be evidently active *in vitro*, no acylated compound was identified. Thus, they stated that the acyl moiety is probably removed prior to or during compound secretion (Supplementary Figure 4b).<sup>4</sup> A similar observation was made in the biosynthesis of the myxobacterial lipopeptide vioprolide, in which acylated

precursors are produced and the bioactive compound is biosynthesized by hydrolysis of the fatty acid residue (Supplementary Figure 4a).<sup>8</sup> In contrast to the hypothesized acylated corramycin precursor, acylated vioprolides could be identified experimentally. We also did not find acylated corramycin derivatives produced by the heterologous host either; however, we speculate that another potential reason for the low production titer of Cor1183 could be a defective pre-drug release mechanism in the heterologous host.

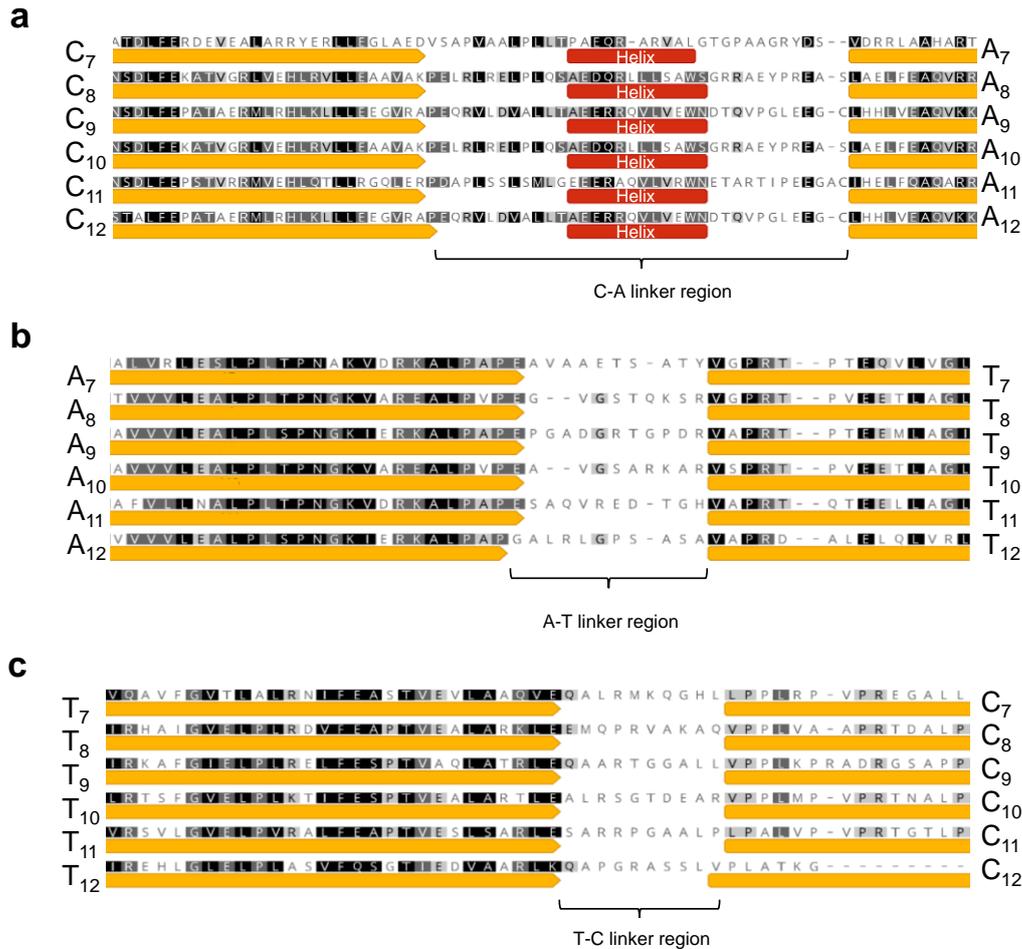
### **4.3.2 Truncation of the assembly line to produce the N-terminal pharmacophore of corramycin**

The chemical synthesis of corramycin has been described and used to determine the absolute stereochemistry of the compound (manuscript in preparation; see chapter 3). Although successful, the synthesis required over 20 steps and especially the (2*R*,3*S*)- $\gamma$ -*N*-methyl- $\beta$ -hydroxy-histidine fragment (5 % yield over 7 steps) turned out to be the main obstacle currently preventing a total synthesis route of Cor1183 or derivatives at a reasonable yield and cost. Therefore, we aimed to develop a semi-synthesis approach for which the N-terminal part of Cor1183 was intended to be generated by fermentation, including the (2*R*,3*S*)- $\gamma$ -*N*-methyl- $\beta$ -hydroxy-histidine moiety. To achieve this goal, we first developed a strategy to produce the N-terminal corramycin part up to the glycine residue introduced by module 7, because this fragment could be integrated directly into the established chemical synthesis route (as depicted in Figure 2). For this, we planned to delete modules 8-11 and connect the A domain of module 7 to the T domain of module 12 and the downstream TE (Figure 4 and Supplementary Figure 3). In a second construct we planned to directly fuse the T domain of module 7 to the C-terminal TE. In a third construct we planned to delete the entire C-terminal assembly line from the A domain of module 8 on, thus having the C domain of module 8 as terminal releasing domain.



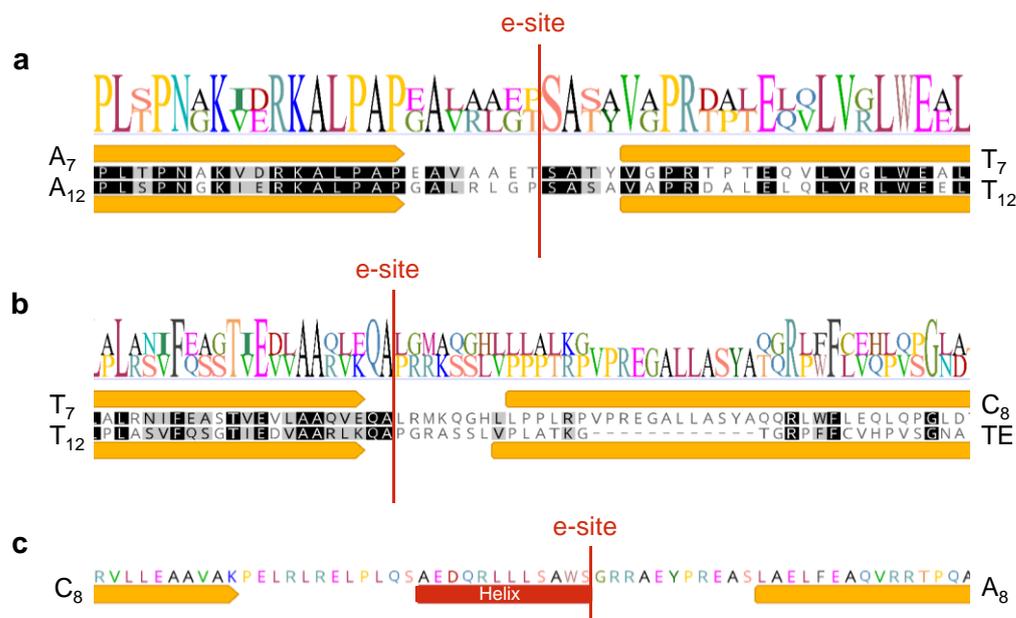
**Figure 4 | NRPS engineering strategy to truncate the corramycin assembly line.** Constructs i-iii with different parts of the assembly line deleted (shown in yellow).

Since the linkers between the domains are critical to maintain domain-domain interactions and to ensure structural integrity and functionality of the NRPS,<sup>19</sup> we searched for suitable engineering sites in the linkers to fuse domains, which are natively not adjacent in the corramycin assembly line. Bozhüyük *et al.* have shown that an intact conserved helix structure in the C-A linker is crucial to maintain domain interactions and identified an engineering site downstream thereof. Here, we used XtalPred<sup>26</sup> to search for secondary structures in the respective domain linkers in ComO. Furthermore, we aligned primary amino acid sequences of all ComO linkers using MUSCLE alignment<sup>27</sup> to find putative conserved amino acids in the linker regions where definite secondary structures could not be identified (Figure 5).



**Figure 5 | MUSCLE protein sequence alignment of all ComO linker regions.** The sequence logos (on top) show amino acid conservation. Fully conserved amino acids are labelled black and partly conserved or non-conserved amino acids are shown in dark grey, light grey or white in each linker sequence, respectively. **a:** Sequence alignment of the C-A linker regions. The helix structure (red annotation) prediction was performed using XtalPred (<http://xtalpred.godziklab.org/XtalPred-cgi/xtal.pl>). No conserved sequence motif downstream of the predicted helix structure was found. **b, c:** Sequence alignment of the A-T and T-C linker regions, respectively. XtalPred analysis identified disordered regions in all A-T and T-C linkers explaining their low similarity. Created using Geneious version 2020.0 (Biomatters).

Notably, we did not find the previously described consensus motif WNATE that was located downstream of the helix structure in the C-A interface (Figure 5a).<sup>20</sup> A potential reason for this might be that Bozhüyük *et al.* focused on NRPSs from *Photorhabdus* and *Xenorhabdus* strains, which are phylogenetically only distantly related to myxobacteria. Thus, we expected that the consensus motif in myxobacteria would differ from the one identified by Bozhüyük and coworkers. No reliable secondary structures could be identified in the corramycin A-T and T-C linker regions of ComO (Figure 5b and c), respectively, as analysis by XtalPred identified those linkers as disordered regions. A primary sequence alignment was therefore used to identify the last conserved amino acid after the upstream domain (e.g. A domain in case of A-T and T domain in case of T-C interface). The engineering site was therefore defined directly after this last conserved residue (Figure 6).



**Figure 6 | Engineering sites (e-sites) used for the deletion of parts in ComO.** MUSCLE protein sequence alignment of the A-T linkers from modules 7 and 12 (a) and the T-C linker from modules 7/8 and T-TE linker from module 12 (b). c: Primary protein sequence of the C-A linker from module 8 with e-site downstream of the helix structure. Created using Geneious version 2020.0 (Biomatters).

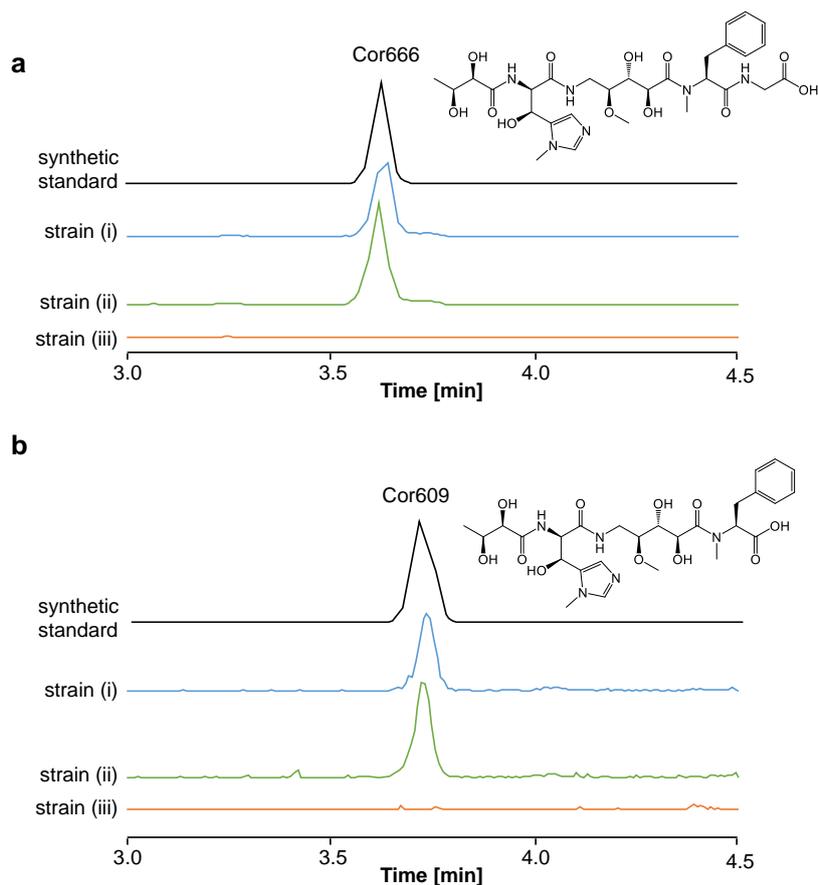
After identification of suitable engineering sites, we subsequently cloned constructs (i-iii) using pMYC20CorOp1 as template. This template harbors the first synthetic operon including the genes encoding the entire NRPS-PKS assembly line (Supplementary Figure 3). We used *in vivo* recombineering in *E. coli* to replace a specific part of *comO* by a kanamycin resistance cassette, which was flanked by *BsaI* restriction sites (R-sites). Since *BsaI* R-sites were removed from the synthetic cluster during gene cluster design (see the Supplementary Information), this enzyme was subsequently used to excise the resistance gene. The generated compatible sticky ends allowed re-ligation of the expression vector without extant R-sites (Supplementary Figure 5). We choose homology sequences for *in vivo* recombineering in a way that they enabled recombination at the specific engineering positions in the linker regions between C-A, A-T and T-C didomains as shown in Figure 6. Next, we generated the strain *M. xanthus* DK1622 pMYC21CorOp2 by integration of the second synthetic operon into the genome via the *Mx9* phage integration site.<sup>28</sup> Constructs (i), (ii) and (iii) were afterwards also independently transformed into this strain, resulting in strain (i): *M. xanthus* DK1622 pMYC21CorOp2 pMYC20CorOp1(i), strain (ii): *M. xanthus* DK1622 pMYC21CorOp2 pMYC20CorOp1(ii) and strain (iii): *M. xanthus* DK1622 pMYC21CorOp2 pMYC20CorOp1(iii), respectively.

Strain (i) and (ii) indeed produced the target mass of 667.29 Da  $[M+H]^+$ , which matches the corramycin fragment with glycine as C-terminal amino acid (Cor666) (Figure 7), whereas strain (iii) did not produce any corramycin-related compounds, which most likely shows that the C domain of

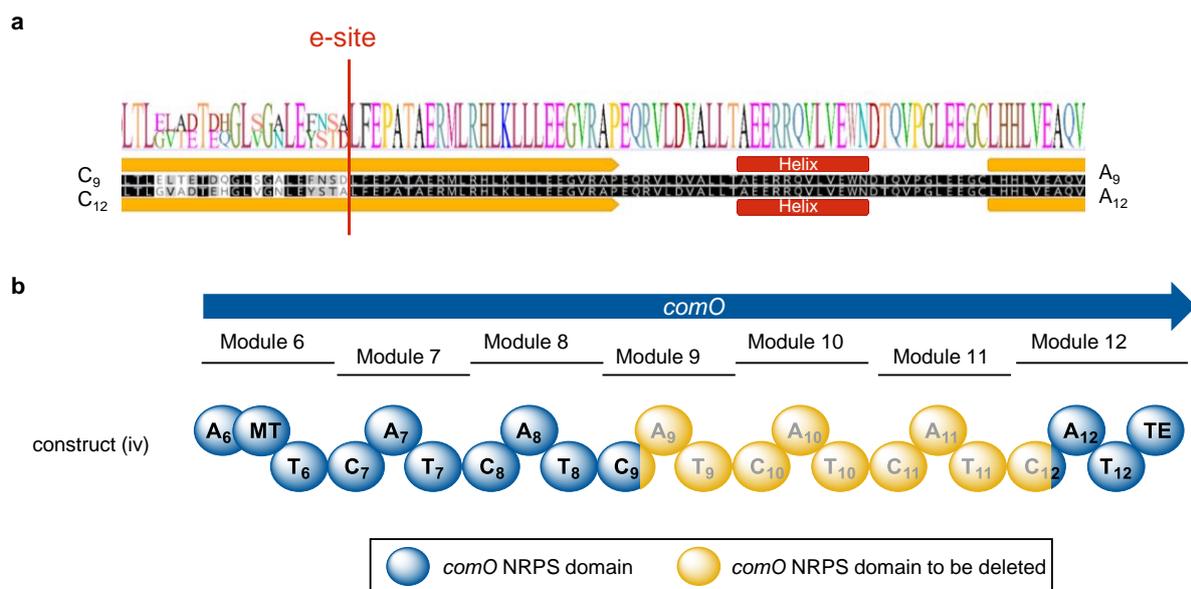
module 8 does not operate as a terminal releasing domain. Surprisingly, we also identified a corramycin fragment with phenylalanine as C-terminal amino acid (610.27 Da  $[M+H]^+$ ; Cor609) in strain (i) and (ii). Retention times, high-resolution masses and MS<sup>2</sup> fragmentation patterns of Cor609 and Cor666 were confirmed by comparison with synthetic standards (Figure 7, Supplementary Figure 6 and Supplementary Figure 7). The fact that Cor609 was produced either means that the C domain of module 7 acts as a releasing domain or the TE is rather unspecific and prone to perform precursor release at various processing states. To our knowledge, a release of the precursor peptide by a C domain located in the middle of the assembly line was not described yet. Another possible scenario may include the spontaneous release of the Cor609 intermediate without the action of any catalytic domain. A similar phenomenon was observed, for example, in the biosynthesis of the type I PKS product bacillaene, where the deletion of the terminal TE domain lead to the spontaneous hydrolysis of 13 bacillaene intermediates.<sup>29</sup> Similar observations were made in the biosynthesis of rifamycin, in which premature chain termination lead to the production of intermediates after deletion of the terminal lactamization enzyme RiffF.<sup>30</sup> However, we assume that both scenarios, the release by the C domain of module 7 and the spontaneous release, are highly unlikely because Cor609 was also not found in strain (iii), which lacks the TE but still harbors the module 7 C domain.

Furthermore, the corramycin BGC does not contain a gene encoding a type II TE, which could catalyze the release of peptide intermediates from a stalled NRPS.<sup>31</sup> NRPS stalling usually occurs when false building blocks that are not accepted by the downstream modules are incorporated into the growing peptide chain. Since Cor609 and Cor666 only contain building blocks that are also present in full-length corramycins, blocking of the assembly line by those derivatives appears unlikely in any case. However, there is still an option that another type II TE encoded somewhere else in the genome of the heterologous producer strain catalyzes the release of corramycin intermediates.

Consequently, we decided to verify if the TE has a relaxed substrate specificity as it was e.g. shown for the TE in tyrocidin biosynthesis.<sup>32,33</sup> To do so, we cloned another construct (iv), in which we connected the C domain of module 9 with the A domain of module 12 (Figure 8).

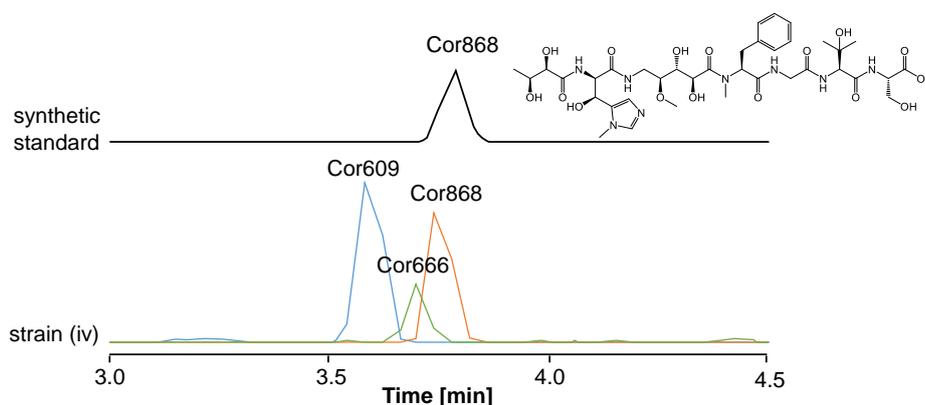


**Figure 7 | UPLC-HRMS analysis of truncated corramycins in the methanolic extracts of *M. xanthus* DK1622 strains (i, ii and iii).** EICs 667.30 [M+H]<sup>+</sup> (a) and 610.26 [M+H]<sup>+</sup> (b) of strains (i) (black), (ii) (blue) and (iii) (green). The structures of Cor666 and Cor609 are shown. The production of Cor666 and Cor609 was confirmed by comparison of the MS<sup>2</sup> fragmentation pattern with synthetic standards (Supplementary Figure 6 and Supplementary Figure 7).



**Figure 8 | Engineering site to generate construct (iv).** a: MUSCLE protein sequence alignment of the C-A linkers from modules 9 and 12. Sequence logo is shown on top. The engineering site was upstream of the conserved helix structure. Created using Geneious version 2020.0 (Biomatters). b: Scheme of the assembly line with the yellow part being deleted to obtain construct (iv).

Both modules are responsible for serine incorporation and show 99.8 % identity on the nucleotide level from the 3' end of the C domain to the end of the A domain, which is why we did not rely on the previously described C-A linker engineering site downstream of the helix structure.<sup>20</sup> We thus decided to place the engineering site in the C domains of modules 9 and 12 where the sequences started being identical (see Figure 8a). After deletion of modules 10, 11 and parts of module 9 using the same strategy as described above (Supplementary Figure 5), the resulting sequence was almost identical to module 9 fused with terminal T-TE didomain. We here expected the production of two new truncated corramycin derivatives with  $\beta$ -hydroxy-L-valine and serine as C-terminal amino acids, respectively. Expression of construct (iv; pMYC20CorOp1(iv)) in *M. xanthus* DK1622 pMYC21CorOp2 (strain (iv)) lead to the production of the expected corramycin derivative Cor868 (869.38 Da  $[M+H]^+$ ), which harbours serine as C-terminal amino acid and is missing a  $\beta$ -hydroxy-L-valine, L-leucine and L-serine (Figure 9 and Supplementary Figure 8). Indeed, Cor666 and Cor609 were also detected implying that the TE is unspecific and facilitates early release of intermediates from the assembly line. However, we failed to identify the truncated corramycin derivative with  $\beta$ -hydroxy-L-valine as C-terminal amino acid. We thus speculate that the TE exhibits a certain degree of specificity and is only able to release specific peptide intermediates from the assembly line.



**Figure 9 | UPLC-HRMS analysis of truncated corramycins in the methanolic extracts of *M. xanthus* DK1622 strain (iv).** EICs 869.38  $[M+H]^+$  (orange), 667.30  $[M+H]^+$  (green) and 610.26  $[M+H]^+$  (blue) of strains (iv) and EIC 869.38  $[M+H]^+$  of the synthetic Cor868 standard (black; shown on top). The structure of Cor868 is shown. Production of Cor868 was confirmed by comparing retention time and MS<sup>2</sup> fragmentation pattern with a synthetic standard (Supplementary Figure 8).

Next, we analyzed if the unspecific release of peptide intermediates also happens in the native producer strain or in the heterologous producer harboring the full cluster. Surprisingly, we found production of Cor609 in the heterologous producer and Cor609, Cor666 and Cor868 in the native producer strain as well (Supplementary Figure 9). This makes the previously discussed presence of a type II TE catalyzing the release of corramycin intermediates unlikely, because the same type II TE would have to be encoded in both the genome of the native and the heterologous producer strain. Nevertheless, partial deletions of the assembly line, significantly directed the production towards

truncated products. The production yields of the truncated corramycins in strains (i-iv), *C. coralloides* MCy10984 and *M. xanthus* DK1622 pMYC20Cor were relatively quantified based on MS by comparing the peak surface areas. Table 1 summarizes the strains, the corresponding constructs and the architecture of the terminal NRPS module. Table 2 shows the relative production yield of truncated corramycins of the best producing clone from each strain. The complete relative quantification including all tested clones is shown in Supplementary Figure 10 and discussed in detail in the Supplementary Information.

**Table 1 | Summary of the strains tested for the production of truncated corramycins.**

Description	Strain	Construct	Terminal NRPS module
strain (i)	<i>M. xanthus</i> DK1622	pMYC21CorOp2 pMYC20CorOp1(i)	
strain (ii)	<i>M. xanthus</i> DK1622	pMYC21CorOp2 pMYC20CorOp1(ii)	
strain (iii)	<i>M. xanthus</i> DK1622	pMYC21CorOp2 pMYC20CorOp1(iii)	
strain (iv)	<i>M. xanthus</i> DK1622	pMYC21CorOp2 pMYC20CorOp1(iv)	
MCy10984	<i>C. coralloides</i> MCy10984	-	
-	<i>M. xanthus</i> DK1622	pMYC20Cor	

**Table 2 | Relative production yields of truncated corramycins in strains (i-iv), *M. xanthus* DK1622 pMYC20Cor and MCy10984.** The given percentage values originate only from the best producing clone of each strain. Each clone was measured in technical duplicates and the calculated percentages are the mean values from the technical duplicates compared to the strain with the highest production (shown in bold). A detailed discussion about the production yields including all tested clones was added in the Supplementary Information (see Supplementary Figure 10).

Strain	(i)	(ii)	(iii)	(iv)	MCy10984	DK1622 pMYC20Cor
Cor609	7 %	7 %	-	11 %	<b>100 %</b>	5 %
Cor666	96 %	<b>100 %</b>	-	67 %	19 %	-
Cor868	-	-	-	<b>100 %</b>	11 %	-

The production of Cor666 was substantially increased in strains (i), (ii) and (iv) compared to the native producer strain, with the highest production in strains (i) and (ii). As Cor666 is the fragment that can eventually be introduced into a semi-synthesis route to generate modified corramycins, we were interested in the absolute production yield of Cor666 in the best producing strain (ii), which was 0.5 mg L<sup>-1</sup>. Although this production yield is not sufficient for industrial application yet, which requires yields in the gram-scale, the basis for the supply of the hard-to-synthesize N-terminal corramycin part by fermentation was set. However, in future experiments the production of Cor666 has to be improved, which could e.g. be obtained by medium optimization including the supplementation of precursors such as D-3-phosphoglycerate, L-histidine, β-alanine, acetate, L-phenylalanine and glycine (those precursors were previously identified by Tesmar and coworkers<sup>4</sup>). Furthermore, for *M. xanthus* DK1622 fermentation procedures with scales up to several hundred liters have already been established,<sup>25</sup> whereas the native corramycin producer strain *C. coralloides* was so far only cultivated in shake flasks. The use of large-scale fermentors instead of shake flasks has the great advantage that cultivation conditions like pH value and aeration can be monitored and kept stable over a longer time period, which bears the potential to increase the heterologous production yield of Cor666. Another option to potentially increase the Cor666 production is testing different promoter systems to achieve optimal gene expression of the modified corramycin BGC, which does not necessarily mean the highest possible gene expression as it was shown in several studies.<sup>34–37</sup> Induction of the heterologous gene expression by the addition of vanillate (described in the method section) at a later time point when higher cell densities are reached is another way to potentially improve Cor666 production. If Cor666 is toxic for the producer strain, growth might be inhibited by early induction of the gene expression, finally leading to decreased overall production. A potential self-toxicity issue might also be addressed by overexpression of genes that mediate resistance towards corramycin, such as *comG* and *comD* (Tesmar *et al.*,<sup>4</sup> manuscript in preparation, see chapter 3).

Summarized, rational NRPS engineering in combination with heterologous expression was successfully applied to produce truncated corramycin derivatives, one of which can be used for semi-synthetic generation of novel corramycin lead structures in an already existing synthesis route. Thereby, the inefficient chemical synthesis of the (2*R*,3*S*)-γ-*N*-methyl-β-hydroxy-histidine building block was circumvented. However, as previously discussed, the low production titer of Cor666 even in the best producing strain (ii) has to be improved in order to achieve industrial applicability. Nonetheless, the hitherto achieved production titer might be sufficient, especially in large scale fermentations, to provide enough material to be used for the semi-synthesis of small amounts of new corramycin analogs. In general, this approach is applicable as an alternative to complex total syntheses of drug lead structures harboring an unaltered natural product-derived pharmacophore.

Furthermore, our results showed that also NRPS engineering sites in the A-T linkers and between T and TE domains can yield functional NRPS assembly lines. However, we cannot draw general conclusions about those engineering sites as no reliable secondary structures were identified within the A-T and T-C/T-TE linkers, which is why we relied on sequence alignments. Nevertheless, our strategy to choose the position of the engineering sites based on conserved or highly similar residues, may also be successful in other cases. The alignment of a variety of linker sequences originating from phylogenetically close (e.g. from one organism) or distant (e.g. from the order of myxococcales) NRPSs might give further information about suitable engineering sites and their general application in NRPS engineering. Moreover, we used the engineering site downstream of the previously described helix structure within C-A linkers<sup>20</sup> (strain (iii)), but we could not detect any corramycin derivatives produced by this strain. Admittedly, we did not use this engineering site to connect the C domain (of module 8) to another domain, but we intended to use the C domain as terminal releasing domain. We speculate that the C domain simply does not work as releasing domain and that the part of the C-A linker that is attached is not the reason for this, because no domain-domain interactions to a downstream domain have to be ensured. Since we used another engineering site inside the C domains of modules 9 and 12 (strain (iv)), we cannot draw conclusions about the general applicability of the previously described engineering site by Bozhüyük and coworkers in the case of myxobacterial NRPSs. The engineering site we used is not transferable to other cases, because it only based on the sequence alignment of two almost identical modules originating from the same NRPS. This sequence similarity is not found when comparing other C-A linker regions as seen when comparing Figure 5 and Figure 8. Thus, to confirm if the previously described C-A linker engineering site is also applicable for engineering of myxobacterial NRPSs, further experiments have to be performed in the future.

## 4.4 Methods

### 4.4.1 Cultivation of strains

*E. coli* DH10 $\beta$ , HS996 and NEB10 $\beta$  strains were used for cloning and cultivated in LB medium (10 g L<sup>-1</sup> tryptone, 5 g L<sup>-1</sup> NaCl, 5 g L<sup>-1</sup> yeast extract, pH 7.6) at 37 °C or 30 °C (handling plasmids larger than 15 kb). Ampicillin (100  $\mu$ g mL<sup>-1</sup>), chloramphenicol (34  $\mu$ g mL<sup>-1</sup>), kanamycin (50  $\mu$ g mL<sup>-1</sup>) and oxytetracyclin (10  $\mu$ g mL<sup>-1</sup>) were used as selection markers. *M. xanthus* DK1622 was used as heterologous expression host. *C. coralloides* MCy10984 is a native corramycin producer and was used as control in production screening experiments. Cultivation on solid medium was performed using M7 medium (5 g L<sup>-1</sup> probion (ME069), 5 g L<sup>-1</sup> corn starch, 2 g L<sup>-1</sup> glucose, 1 g L<sup>-1</sup> yeast extract, 1 g L<sup>-1</sup>, MgSO<sub>4</sub> × 7H<sub>2</sub>O, 1 g L<sup>-1</sup>, CaCl<sub>2</sub> × 2H<sub>2</sub>O, 10 g L<sup>-1</sup> HEPES, pH 7.4) or solid CTT medium. Liquid cultivation was performed in CTT medium (10 g L<sup>-1</sup> casitone, 1.21 g L<sup>-1</sup> TRIS, 8 mM MgSO<sub>4</sub>, 1 mM KH<sub>2</sub>PO<sub>4</sub> (pH 7.6), pH 7.6) to grow *M. xanthus* DK1622 cultures for plasmid transformations. M7/s4 medium (5 g L<sup>-1</sup> soy flour, 5 g L<sup>-1</sup> corn starch, 2 g L<sup>-1</sup> glucose, 1 g L<sup>-1</sup> yeast extract, 1 g L<sup>-1</sup> MgSO<sub>4</sub> × 7H<sub>2</sub>O, 1 g L<sup>-1</sup> CaCl<sub>2</sub> × 2H<sub>2</sub>O, 10 g L<sup>-1</sup> HEPES, pH 7.4 and supplemented with 0.1 mg L<sup>-1</sup> of vitamin B<sub>12</sub> and 5 mg L<sup>-1</sup> of FeCl<sub>3</sub> after autoclaving) was used for production screening cultures (50 mL scale). M7/s4 starting cultures (without supplements) were inoculated from solid CTT or M7 agar starting cultures. Production screening cultures were inoculated from 1-3 d old M7/s4 pre cultures (10 % (v/v) inoculation volume) and cultivated for 5 d. After 1 d, heterologous gene expression in *M. xanthus* DK1622 was induced by adding vanillate (1 mM final concentration). XAD16 absorber resin was added after 2 d (2 % (v/v)). All liquid cultivations were performed in 300 mL baffled Erlenmeyer flasks on an orbital shaker at 160 rpm at 30 °C. Kanamycin (50  $\mu$ g mL<sup>-1</sup>) and oxytetracyclin (10  $\mu$ g mL<sup>-1</sup>) were used as selection markers when cultivating heterologous *M. xanthus* DK1622 strains. *S. cerevisiae* ATCC4004247 was used for TAR cloning. Cultivations were performed at 30 °C in YPAD medium (20 g L<sup>-1</sup> glucose, 10 g L<sup>-1</sup> peptone, 10 g L<sup>-1</sup> yeast extract, 100 mg L<sup>-1</sup> adenine-hemisulfate, pH 7.0). YNB medium (20 g L<sup>-1</sup> glucose, 8 g L<sup>-1</sup> YNB base w/o leucine, 2 g L<sup>-1</sup> amino acid mix w/o leucine, 100 mg L<sup>-1</sup> adenine-hemisulfate, pH 7.0) was used for selection of transformants.

### 4.4.2 *In silico* experiments

*Geneious* v10.1.3 (Biomatters Ltd.) was used to analyze the native corramycin BGC sequence of MCy10984 and to design the modified BGC *in silico*. Repetitive sequence segments in *comO* were analyzed using dotplot (*EMBOSS* 6.5.7 tool *dottup*: <http://emboss.sourceforge.net/>). All *in silico* experiments are described in detail in the Supplementary Information. GenBank accession numbers will be provided upon acceptance of the manuscript.

### 4.4.3 DNA synthesis and BGC assembly

The modified BGC was synthesized in fragments (fragment description listed in Supplementary Table 5). DNA synthesis was carried out by ATG: biosynthetics GmbH. Sequence-verified DNA synthesis fragments were delivered in pGH standard vector harboring an *ampR* (*bla*) gene for selection on ampicillin. Restriction endonuclease hydrolysis, DNA ligation, *E. coli* transformation and plasmid DNA isolation were done according to standard protocols.<sup>38</sup> Both operons of the modified BGC (CorOp1 and CorOp2) were assembled in two separate TAR cloning reactions using the vectors pMYC20 or pMYC21 (manuscript under revision; see chapter 2), respectively. TAR cloning was performed according to standard protocols.<sup>39</sup> The *comO* gene was replaced by a dummy sequence in order to minimize the risk of unspecific recombination during TAR assembly. The fragments of *comO* harboring repetitive sequence segments were assembled separately *in vitro* by a three-step restriction/ligation cloning strategy using *BsaI* (Supplementary Information in manuscript under revision; see chapter 2) prior to its assembly with the rest of the cluster.<sup>23</sup> Supplementary Figure 3 schematically depicts all cloning steps performed to obtain the final expression construct pMYC20Cor. Supplementary Table 6 summarizes all *in vitro* cloning steps performed in this work.

### 4.4.4 Genetic manipulation of expression constructs

Red/ET recombineering<sup>40</sup> in combination with restriction hydrolysis and re-ligation was used to delete part of *comO* from the plasmids and thus truncate the assembly line for the production of shorter corramycin derivatives. Amplification of *kanR* gene from pDPO-mxn116-Pvan-Tpase<sup>41</sup> was done via PCR. Apart from template binding sites, primers contained *BsaI* R-sites and 50 bp sequences that are homologous to the sequence part in *comO*, which was adjacent to the part that was deleted. Supplementary Table 3 lists all primers used for this experiment. After Red/ET recombineering, we selected clones harboring the correct recombination products on kanamycin and oxytetracycline. After the plasmid isolation, we verified the clones by restriction analysis. Next, the recombination product was hydrolysed with *BsaI* and re-ligated to remove *kanR* from the construct. Clones, which lost their resistance towards kanamycin, were selected for plasmid isolation and restriction analysis. Supplementary Table 7 lists all manipulated plasmids that were generated in this study. Supplementary Table 1 and Supplementary Table 2 list all strains and plasmids generated during the cloning process, respectively.

#### 4.4.5 Transformation of *M. xanthus* DK1622 and verification by colony PCR

Expression constructs were transformed into *M. xanthus* DK1622 via electroporation using established standard protocols.<sup>42</sup> Integration of the constructs pMYC20Cor and pMYC20CorOp1 into the chromosome occurred by site-specific phage recombination in Mx8 attachment site.<sup>24</sup> The construct pMYC21CorOp2 was integrated into the chromosome by site-specific phage recombination in Mx9 attachment site.<sup>28</sup> Integration was verified by colony PCR using different combinations of primers. The primers Mx8-attB-up2, Mx8-attB-down, Mx8-attP-up2 and Mx8-attP-down were used to verify Mx8 attachment site integration and Mx9attB1\_up, Mx9attB1\_down, Mx9attP\_up and Mx9attP\_down to verify Mx9 attachment site integration (Supplementary Table 3). DNA preparation, suitable primer combinations, reaction conditions and PCR product sizes are described by Pogorevc *et al.*<sup>41</sup> and Gemperlein *et al.*<sup>43</sup> Supplementary Table 1 lists all expression strains generated in this work.

#### 4.4.6 Sample preparation and UPLC-ESI-HRMS analysis

Cells and XAD16 absorber resin of 50 mL screening cultures were harvested by centrifugation at  $3,200 \times g$  for 15 min at 4 °C. Extraction was done two times for 60 min with 30 ml methanol under stirring at room temperature (RT). Extracts were filtered using folded filter paper (8-12  $\mu\text{m}$  pore size) and dried using a rotary evaporator. Dried extract was dissolved in 3 mL methanol and analyzed using UPLC-HRMS. An UltiMate 3000 LC System (Dionex) with a Acquity UPLC BEH C-18 column (1.7  $\mu\text{m}$ , 100 x 2 mm; Waters), equipped with a VanGuard BEH C-18 (1.7  $\mu\text{m}$ ; Waters) guard column, was coupled to an Apollo II ESI source (Bruker) and hyphenated to maXis 4G ToF mass spectrometer (Bruker). Separation was performed at a flow rate of 0.6 mL min<sup>-1</sup> (eluent A: deionized water + 0.1 % formic acid (FA), eluent B: acetonitrile + 0.1 % FA) at 45 °C using the following gradient: 5 % B for 30 s, followed by a linear gradient up to 95 % B in 18 min and a constant percentage of 95 % B for further 2 min. Original conditions were adjusted with 5 % B within 30 s and kept constant for 1.5 min. The LC flow was split to 75  $\mu\text{L}$  min<sup>-1</sup> before entering the mass spectrometer. Mass spectra were acquired in centroid mode ranging from 150–2,500  $m/z$  at a 2 Hz full scan rate. Mass spectrometry source parameters were set to 500 V as end plate offset, 4,000 V as capillary voltage, 1 bar nebulizer gas pressure, 5 L min<sup>-1</sup> dry gas flow and 200 °C dry temperature. For MS<sup>2</sup> experiments, CID (collision-induced dissociation) energy was ramped from 35 eV for 500  $m/z$  to 45 eV for 1,000  $m/z$ . MS full scan acquisition rate was set to 2 Hz and MS/MS spectra acquisition rates were ramped from 1 to 4 Hz for precursor ion intensities of 10 kcts to 1000 kcts. We used *Compass DataAnalysis* version 4.4 (Bruker) to interpret MS data.

#### 4.4.7 Relative quantification of truncated corramycins

To quantify the relative production of truncated corramycins Cor609, Cor666 and Cor868 in the heterologous producers (strains (i-iv) and *M. xanthus* DK1622 pMYC20Cor) and native producer strain (MCy10984), we analyzed the methanolic extracts (see previous section) of 2-3 different clones from each strain (originating from 50 mL screening cultures). The extract of each clone was measured twice using the maXis 4G ToF mass spectrometer as described above. The MS peak surface of the target compounds (EICs 610.27 [M+H]<sup>+</sup>, 667.29 [M+H]<sup>+</sup> and 869.38 [M+H]<sup>+</sup>) were integrated manually using *DataAnalysis* version 4.4 (Bruker). The best producing clone of each strain was selected for calculation of the relative production of truncated corramycins normalized to the clone showing highest production in all strains (Table 2). A detailed analysis including all clones was described in the Supplementary Information (Supplementary Figure 10).

#### 4.4.8 Absolute quantification of the Cor666 production

Quantification of Cor666 in the heterologous producer strains *M. xanthus* DK1622 pMYC21CorOp2 pMYC20CorOp1(ii) was done using an amaZon speed 3D ion trap MS system (Bruker) with an Apollo II ESI source. ESI source settings were identical as described above. We measured Cor666 standard solutions with concentrations of 0.001 mg mL<sup>-1</sup>, 0.01 mg mL<sup>-1</sup>, 0.1 mg mL<sup>-1</sup>, 1 mg L<sup>-1</sup> and 10 mg L<sup>-1</sup>. Solutions for each concentration were prepared three times and measured two times. The peak surface areas of Cor666 EICs were integrated and the calculation of the quantities in the crude extracts was done using *QuantAnalysis* version 4.4 (Bruker), respectively.

#### 4.4.9 MIC determination of Cor1183 against *M. xanthus* DK1622

For MIC determination of Cor1183, *M. xanthus* DK1622 wild-type or *M. xanthus* DK1622 pMYC20Cor cultures were grown in CTT medium until an optical density at 600 nm wavelength (OD<sub>600 nm</sub>) of 1.0 was reached. The cultures were diluted 1/10 by adding fresh CTT medium. 150 µL of the diluted culture were transferred into the wells of a 96-well plate with doubling Cor1183 concentrations from 0.125 µg mL<sup>-1</sup> to the highest concentration of 64 µg mL<sup>-1</sup> and optionally 3 µg mL<sup>-1</sup> polymyxin B. A culture without Cor1183 supplementation in CTT was cultivated as control. The cultures were grown for 2 days at 30 °C under shaking and subsequently the OD<sub>600 nm</sub> was determined using an Infinite F200 Pro microplate reader (Tecan) and *Tecan i-control* version 1.10.4.0 software.

## 4.5 References

1. Friedman, N. D., Temkin, E. & Carmeli, Y. The negative impact of antibiotic resistance. *Clin. Microbiol. Infect.* **22**, 416–422; 10.1016/j.cmi.2015.12.002 (2016).
2. Levy, S. B. & Marshall, B. Antibacterial resistance worldwide: causes, challenges and responses. *Nat. Med.* **10**, S122-9; 10.1038/nm1145 (2004).
3. Tacconelli, E. *et al.* Discovery, research, and development of new antibiotics: the WHO priority list of antibiotic-resistant bacteria and tuberculosis. *Lancet Infect. Dis.* **18**, 318–327; 10.1016/S1473-3099(17)30753-3 (2018).
4. von Tesmar, A. Dissertation. Saarland University (2017).
5. Wenzel, S. C. & Müller, R. in *Industrial biotechnology, Microorganisms Volume 3a and 3b*, edited by C. Wittmann & J. Liao (Wiley-VCH, Weinheim, Germany, 2017), pp. 453–485.
6. 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**; 10.1016/B978-0-12-409547-2.14818-8 (2020).
7. Arslanian, R. L. *et al.* Large-scale isolation and crystallization of epothilone D from *Myxococcus xanthus* cultures. *J. Nat. Prod.* **65**, 570–572; 10.1021/np010438s (2002).
8. 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; 10.1002/anie.201802479 (2018).
9. Fu, J. *et al.* Efficient transfer of two large secondary metabolite pathway gene clusters into heterologous hosts by transposition. *Nucleic Acids Res.* **36**, e113; 10.1093/nar/gkn499 (2008).
10. Pogorevc, D. *et al.* Biosynthesis and Heterologous Production of Argyrins. *ACS Synth. Biol.* **8**, 1121–1133; 10.1021/acssynbio.9b00023 (2019).
11. Marahiel, M. A. Working outside the protein-synthesis rules: insights into non-ribosomal peptide synthesis. *J. Pept. Sci.* **15**, 799–807; 10.1002/psc.1183 (2009).
12. Weissman, K. J. The structural biology of biosynthetic megaenzymes. *Nat. Chem. Biol.* **11**, 660–670; 10.1038/nchembio.1883 (2015).
13. Staunton, J. & Weissman, K. J. Polyketide biosynthesis: a millennium review. *Nat. Prod. Rep.* **18**, 380–416 (2001).

14. Du, L. & Lou, L. PKS and NRPS release mechanisms. *Nat. Prod. Rep.* **27**, 255–278; 10.1039/b912037h (2010).
15. Müller, S. *et al.* Biosynthesis of crocacin involves an unusual hydrolytic release domain showing similarity to condensation domains. *Chem. Biol.* **21**, 855–865; 10.1016/j.chembiol.2014.05.012 (2014).
16. Winn, M., Fyans, J. K., Zhuo, Y. & Micklefield, J. Recent advances in engineering nonribosomal peptide assembly lines. *Nat. Prod. Rep.* **33**, 317–347; 10.1039/c5np00099h (2016).
17. Nguyen, K. T. *et al.* Combinatorial biosynthesis of novel antibiotics related to daptomycin. *Proc. Natl. Acad. Sci. USA* **103**, 17462–17467; 10.1073/pnas.0608589103 (2006).
18. Uguru, G. C. *et al.* Active-site modifications of adenylation domains lead to hydrolysis of upstream nonribosomal peptidyl thioester intermediates. *J. Am. Chem. Soc.* **126**, 5032–5033; 10.1021/ja048778y (2004).
19. Marahiel, M. A. A structural model for multimodular NRPS assembly lines. *Nat. Prod. Rep.* **33**, 136–140; 10.1039/c5np00082c (2016).
20. Bozhüyük, K. A. J. *et al.* *De novo* design and engineering of non-ribosomal peptide synthetases. *Nat. Chem.* **10**, 275–281; 10.1038/nchem.2890 (2018).
21. Bozhüyük, K. A. J. *et al.* Modification and *de novo* design of non-ribosomal peptide synthetases using specific assembly points within condensation domains. *Nat. Chem.* **11**, 653–661; 10.1038/s41557-019-0276-z (2019).
22. Kouprina, N. & Larionov, V. Transformation-associated recombination (TAR) cloning for genomics studies and synthetic biology. *Chromosoma* **125**, 621–632; 10.1007/s00412-016-0588-3 (2016).
23. Yan, F. *et al.* Synthetic biology approaches and combinatorial biosynthesis towards heterologous lipopeptide production. *Chem. Sci.* **9**, 7510–7519; 10.1039/c8sc02046a (2018).
24. Magrini, V., Creighton, C. & Youderian, P. Site-specific recombination of temperate *Mycococcus xanthus* phage Mx8: Genetic elements required for integration. *J. Bacteriol.* **181**, 4050–4061 (1999).

25. 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; 10.1002/anie.201106305 (2012).
26. Slabinski, L. *et al.* XtalPred: a web server for prediction of protein crystallizability. *Bioinformatics* **23**, 3403–3405; 10.1093/bioinformatics/btm477 (2007).
27. Madeira, F. *et al.* The EMBL-EBI search and sequence analysis tools APIs in 2019. *Nucleic Acids Res.* **47**, W636–W641; 10.1093/nar/gkz268 (2019).
28. Julien, B. Characterization of the integrase gene and attachment site for the *Myxococcus xanthus* bacteriophage Mx9. *J. Bacteriol.* **185**, 6325–6330; 10.1128/JB.185.21.6325–6330.2003 (2003).
29. Moldenhauer, J., Chen, X. H., Borriss, R. & Piel, J. Biosynthesis of the antibiotic bacillaene, the product of a giant polyketide synthase complex of the trans-AT family. *Angew. Chem. Int. Ed. Engl.* **46**, 8195–8197 (2007).
30. Yu, T. W. *et al.* Direct evidence that the rifamycin polyketide synthase assembles polyketide chains processively. *Proc. Natl. Acad. Sci. U.S.A.* **96**, 9051–9056 (1999).
31. Kotowska, M. & Pawlik, K. Roles of type II thioesterases and their application for secondary metabolite yield improvement. *Appl. Microbiol. Biotechnol.* **98**, 7735–7746; 10.1007/s00253-014-5952-8 (2014).
32. Kohli, R. M., Walsh, C. T. & Burkart, M. D. Biomimetic synthesis and optimization of cyclic peptide antibiotics. *Nature* **418**, 658–661 (2002).
33. 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).
34. Ajikumar, P. K. *et al.* Isoprenoid pathway optimization for Taxol precursor overproduction in *Escherichia coli*. *Science* **330**; 10.1126/science.1191652 (2010).
35. Farasat, I. *et al.* Efficient search, mapping, and optimization of multi-protein genetic systems in diverse bacteria. *Mol. Syst. Biol.* **10**; 10.15252/msb.20134955 (2014).
36. Anthony, J. R. *et al.* Optimization of the mevalonate-based isoprenoid biosynthetic pathway in *Escherichia coli* for production of the anti-malarial drug precursor amorpha-4,11-diene. *Metab. Eng.* **11**; 10.1016/j.ymben.2008.07.007 (2009).

37. Myronovskyi, M. & Luzhetskyy, A. Native and engineered promoters in natural product discovery. *Nat. Prod. Rep.*; 10.1039/c6np00002a (2016).
38. Sambrook, J. & Russell, D. W. *Molecular cloning: A laboratory manual* (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 2001).
39. Agatep, R., Kirkpatrick, R. D., Parchaliuk, D. L., Woods, R. A. & Gietz, R. D. Transformation of *Saccharomyces cerevisiae* by the lithium acetate/single-stranded carrier DNA/polyethylene glycol protocol. *Technical Tips Online* **3**, 133–137; 10.1016/S1366-2120(08)70121-1 (1998).
40. Zhang, Y., Buchholz, F., Muyrers, J. P.P. & Stewart, A. F. A new logic for DNA engineering using recombination in *Escherichia coli*. *Nature Genetics* **20**, 123; 10.1038/2417 (1998).
41. Pogorevc, D. *et al.* Production optimization and biosynthesis revision of coralopyronin A, a potent anti-filarial antibiotic. *Metab. Eng.* **55**, 201–211; 10.1016/j.ymben.2019.07.010 (2019).
42. Kashefi, K. & Hartzell, P. L. Genetic suppression and phenotypic masking of a *Myxococcus xanthus* *frzF*- defect. *Mol. Microbiol.* **15**, 483–494 (1995).
43. 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; 10.1039/C3SC53163E (2014).

## 4.6 Supplementary Information

### 4.6.1 *In silico* analysis of the native corramycin BGC sequence

*In silico* analysis of the native corramycin BGC sequence from MCy10984 revealed a GC content of 71.8 %. By looking at the size of intergenic regions and the coding strand on which the genes are located, five putative transcriptional units were identified within the corramycin BGC: *comA-C*, *comF-D*, *comG*, *comH* and *comI-O* and one additional gene (*comP*) downstream of *comO* encoding a methyltransferase. Dotplot analysis (*EMBOSS 6.5.7* tool *dottup*: <http://emboss.sourceforge.net/>) of the entire BGC revealed several repetitive sequence segments in *comO* spanning over 15 kb. One large sequence repeat was identified between modules 6 (C6, A6, T6) and 8 (C6, A6, T6), which catalyze the incorporation of  $\beta$ -hydroxy-L-valine, respectively. Another large sequence repeat was identified between modules 7 (C7, A7) and 10 (C10, A10), which catalyze the incorporation of serine. Furthermore, numerous shorter sequence repeats were identified in *comO*.

#### 4.6.2 *In silico* design of the modified BGC

The modified BGC sequence was designed *in silico* based on the sequence of MCy10984. We included all genes from *comA* to *comP*. The modified BGC was organized in two transcriptional units CorOp1 and CorOp2. CorOp1 contains the putative native operon *comI-O* (without native promoter, RBS of *comI* and terminator). CorOp2 combines all other genes in the following order: *comB-C*, *comP*, *comA*, *comH*, *comF*, *comE-D*, *comG* (native promoters, the RBSs of *comB* and terminators excluded). Vanillate-inducible promoter system<sup>44</sup> was used for inducible gene expression in *M. xanthus* DK1622.

$P_{van}$ , including *vanR* encoding repressor gene, was engineered upstream of CorOp1 and only  $P_{van}$  (without *vanR*) was added upstream of CorOp2. *tD1* terminator sequence from *M. xanthus* bacteriophage Mx8<sup>24</sup> was engineered downstream of both modified operons, respectively. For TAR assembly, we engineered *LEU2* gene encoding  $\beta$ -isopropylmalate dehydrogenase auxotrophy marker downstream and 100 bp sequences originating from *URA3* gene downstream and upstream of both operons, respectively. We added unique R-sites for cloning purposes. Supplementary Table 4 lists all genetic elements including sequence origin and unique R-sites used for the design of CorOp1 and CorOp2. We removed 27 R-sites (e.g. *BsaI*) by synonymous codon substitutions while keeping the codon usage bias for a single amino acid as similar as possible to the native codon usage bias in the BGC. To reduce costs and turnaround time for DNA synthesis, we divided the modified BGC into twelve fragments (Supplementary Table 5). We added 100 bp homologous sequences to all adjacent cluster fragments used for TAR cloning. R-sites were engineered at the 5' and 3' ends of all fragments for DNA synthesis vector release or for step-wise assembly in cloning vectors. We flanked some fragments by splitter elements (SEs) if they were used for three-step assembly in cloning vectors.<sup>23</sup> SEs consist of unique 'conventional' type II R-sites flanked by two *BsaI* recognition sequences extended with 5 bp sequences. *BsaI* is a type IIS restriction endonuclease cutting outside of the recognition sequence. This allowed generation of variable and unique 5 bp sticky ends for ligation. Unique 'conventional' R-sites allowed stepwise cloning of several fragments into a cloning vector.

#### 4.6.3 Assembly of the modified BGC

*comO* fragments 1-4 were assembled by restriction/ligation-based *in vitro* cloning techniques using pSynbio1 as cloning vector. The SEs were removed by restriction hydrolysis with *BsaI* and re-ligation using T4 ligase. After rejoining of the fragments, the generated construct did not contain additional R-sites. TAR cloning was used to assemble pMYC20preCorOp1 and pMYC21CorOp2. For pMYC20preCorOp1 assembly, DNA fragments a-e, dummy and *EcoRV*-linearized pMYC20

were used. The dummy was used instead of *comO* fragments to avoid unspecific recombination caused by repetitive sequence segments. pMYC21CorOp2 was cloned from fragments f-h and *EcoRV*-linearized pMYC21. During TAR assembly, the counter selection marker *URA3* in the vectors was disrupted whereas *LEU2* was introduced into the plasmid together with the cluster fragments. Transformation of *S. cerevisiae* ATCC4004247 was done according to the *standard high-efficiency transformation protocol I* described by Agatep et al.<sup>39</sup> However, centrifuging was done at  $3,200 \times g$  and heat shock lasted for 45 min. Constructed plasmids were isolated according to the protocol from Kouprina & Larionov<sup>45</sup> (*Identification of gene-positive pools*) and transformed into *E. coli* DH10 $\beta$ . Clones harboring the correct construct were verified by restriction analysis.

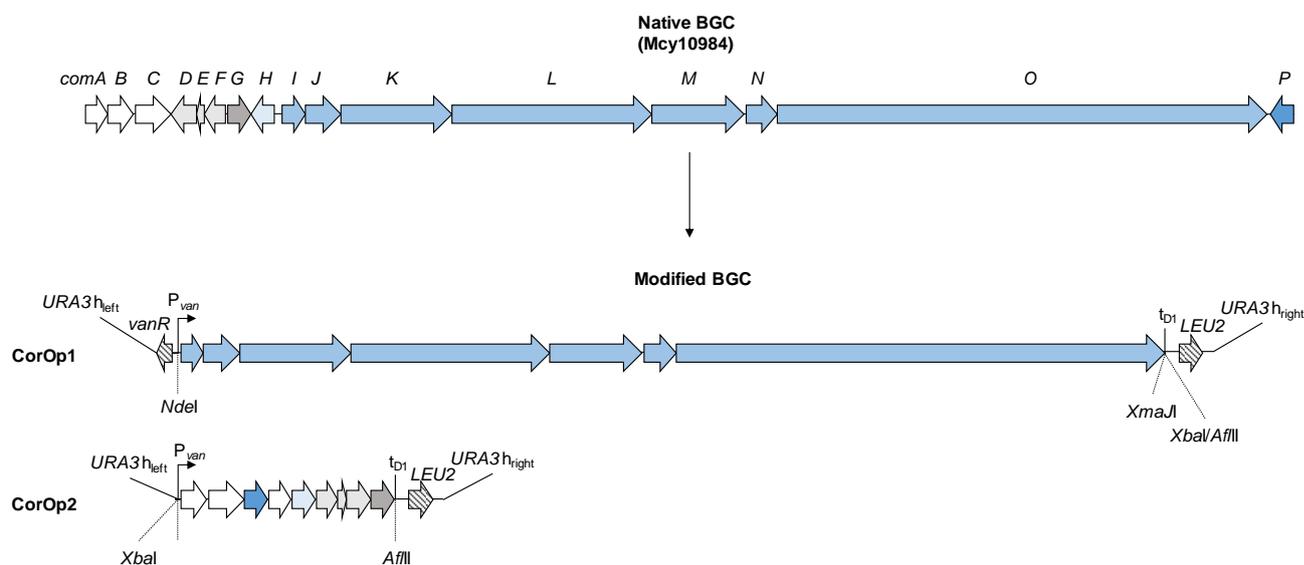
13Dummy was replaced by *comO* (fragment 1-4) in pMYC20preCorOp1 via restriction/ligation to generate pMYC20CorOp1. To generate final expression construct pMYC20Cor, we ligated CorOp2 into pMYC20CorOp1. Supplementary Figure 3 schematically depicts all cloning steps performed to obtain the final expression construct pMYC20Cor. Supplementary Table 6 lists all conventional cloning steps for the generation of pMYC20Cor. We verified all generated constructs by restriction analysis. Additionally, the Illumina paired-end technology on a MiSeq PE300 platform (in-house) was used to verify the sequences of pMYC20CorOp1 (2,646-fold mean sequencing coverage), pMYC21CorOp2 (7,537-fold) and pMYC20Cor (681-fold). Furthermore the sequences of modified expression constructs were sequenced: pMYC20CorOp1(i) (1,517-fold), pMYC20CorOp1(ii) (1,313-fold), pMYC20CorOp1(iii) (1,539-fold), pMYC20CorOp1(iv) (1,693-fold). Supplementary Table 1 and Supplementary Table 2 list all strains and plasmids generated during cloning process.

#### 4.6.4 Relative quantification of truncated corramycin production

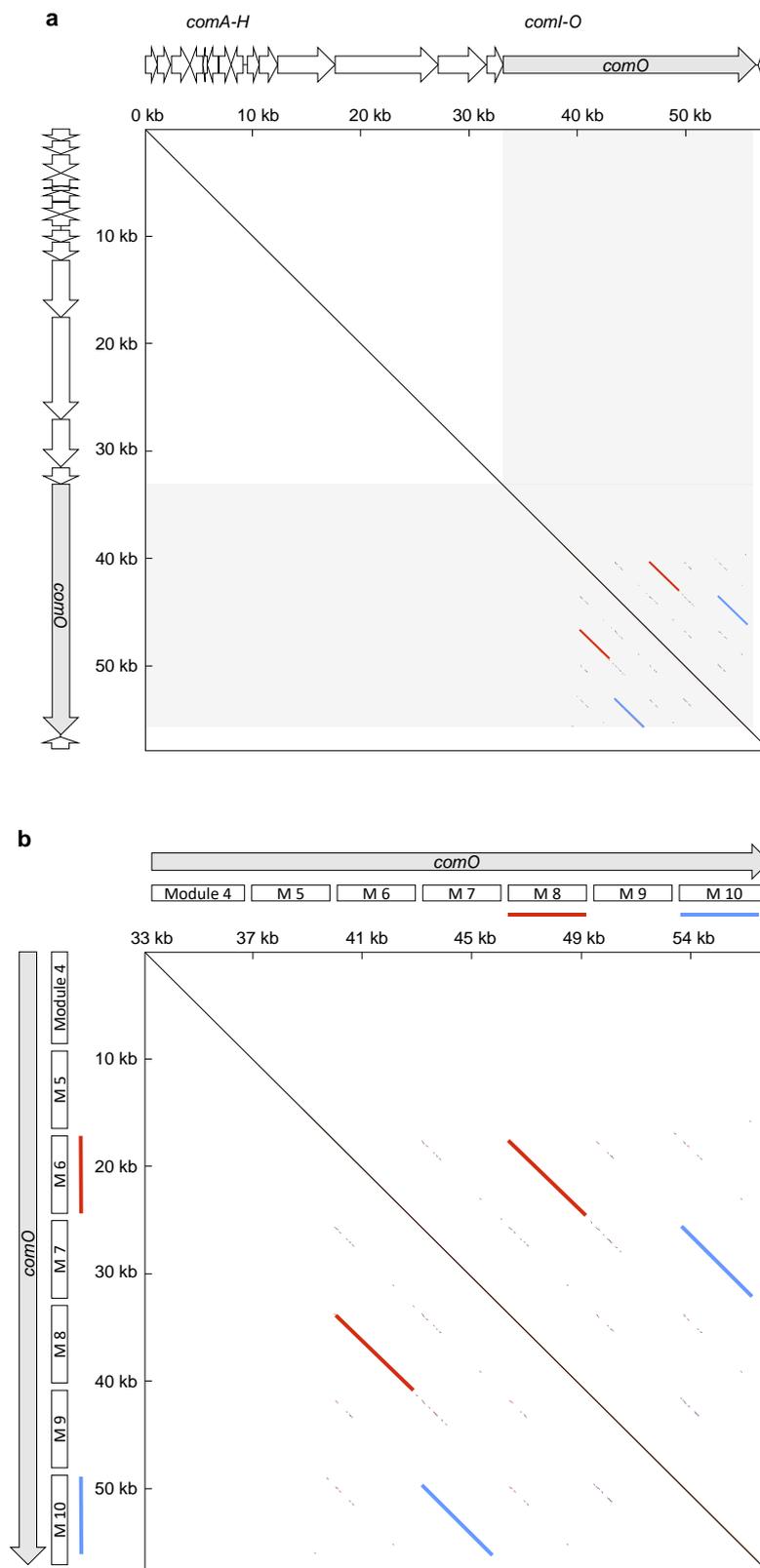
The relative production of truncated corramycins was determined as described in the method section. Supplementary Figure 10 shows that the production of Cor609 is around 20 times higher in the native producer strain *C. coralloides* MCy10984 as compared to strains (i), (ii), (iv) and *M. xanthus* DK1622 pMYC20Cor (Supplementary Figure 10a). On the contrary, the production of Cor666 was only 20 % in the native producer strain compared to the production in strains (i) and (ii) (Supplementary Figure 10b). Strain (iv) showed a production of Cor666 of around 40 % compared to strain (i), but the standard deviation was high meaning that the tested clones produced strongly differing amounts of the target compound. This high variation was also observed in the production of Cor868 (Supplementary Figure 10c). Nevertheless, the average production of Cor868 in strain (iv) was around 10 times higher compared to MCy10984. Huge differences in the corramycin production between different clones of one strain can e.g. be explained with different cell densities that were achieved during cultivation. At this time we cannot make a statement about the achieved cell densities as we did not monitor the optical density (OD) of the cultures during cultivation. For

more accurate results, the cell dry weight (CDW) should be determined and the relative production should be set in ratio with the CDW. However, the use of clumpy M7/s4 medium makes the determination of the CDW and OD difficult. The use of a different, clearer medium might solve this problem, but only if the low production of corrAMYCINS is not further reduced, which could potentially result in compound peaks being downsized on noise level, thus preventing any quantification. Another reason for the high variation between different clones of one strain can be mutations, which lead to improved or decreased growth and overall fitness. In future experiments, more clones have to be screened for the production of truncated corrAMYCINS and the best producing clone(s) have to be cultivated in greater numbers to get more accurate results.

#### 4.6.5 Supplementary Figures

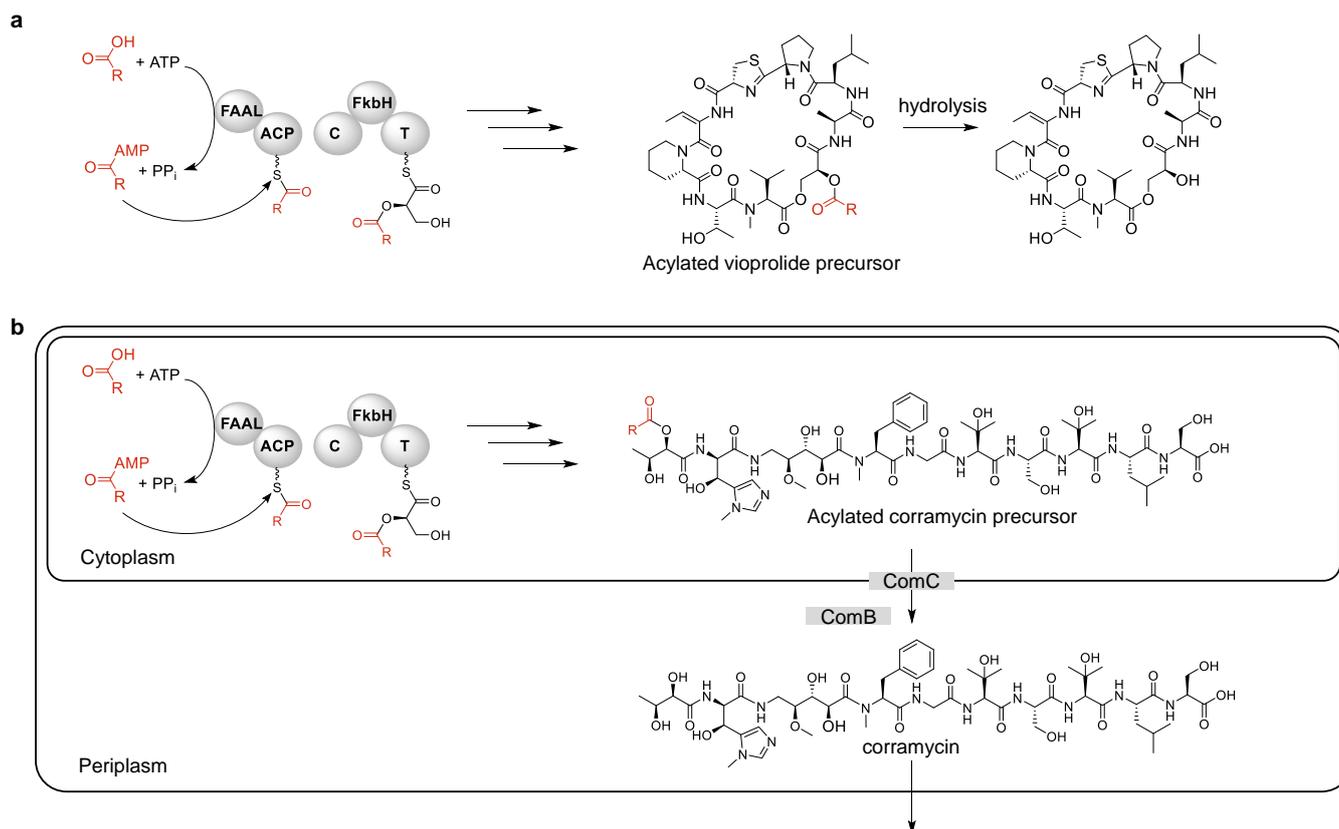


**Supplementary Figure 1 | Design of the modified corrAMYCIN BGC based on the native BGC from MCy10984.** The modified BGC consists of two operons: CorOp1: URA3 homology left for TAR assembly, vanR repressor, vanillate inducible promoter P<sub>van</sub>, biosynthetic genes comI-O, tD1 terminator, LEU2 auxotrophy marker and URA3 homology right for TAR assembly. CorOp2: URA3 homology left, P<sub>van</sub>, biosynthetic genes comB-C, comP, comA, comH, comF, comD-E and comG, tD1, LEU2 and URA3 homology right. Location of important R-sites are shown by dashed lines.

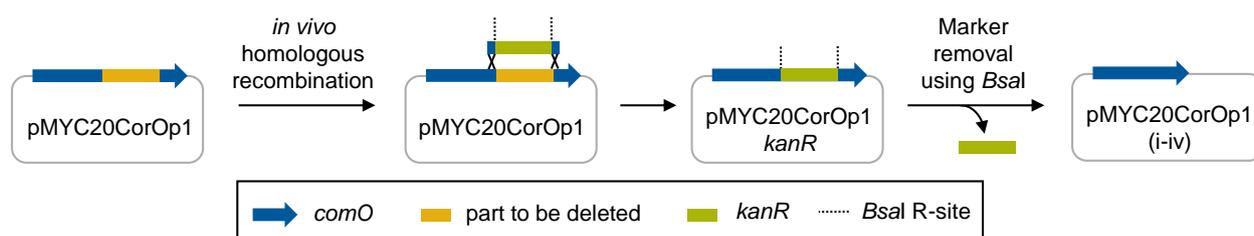


**Supplementary Figure 2 | Dotplots of entire gene cluster and *comO*.** **a:** Dotplot of entire gene cluster. Two large repetitive sequence segments (red and blue bars) and numerous smaller repeats (red/blue dots) span over 15 kb in *comO* (highlighted in light grey). **b:** Dotplot of *comO*. Magnification of the grey area in a. The Dotplot was created using the EMBOSS 6.5.7 tool *dottup* in Geneious (Geneious version 2020.0 created by Biomatters. Available from <https://www.geneious.com>).

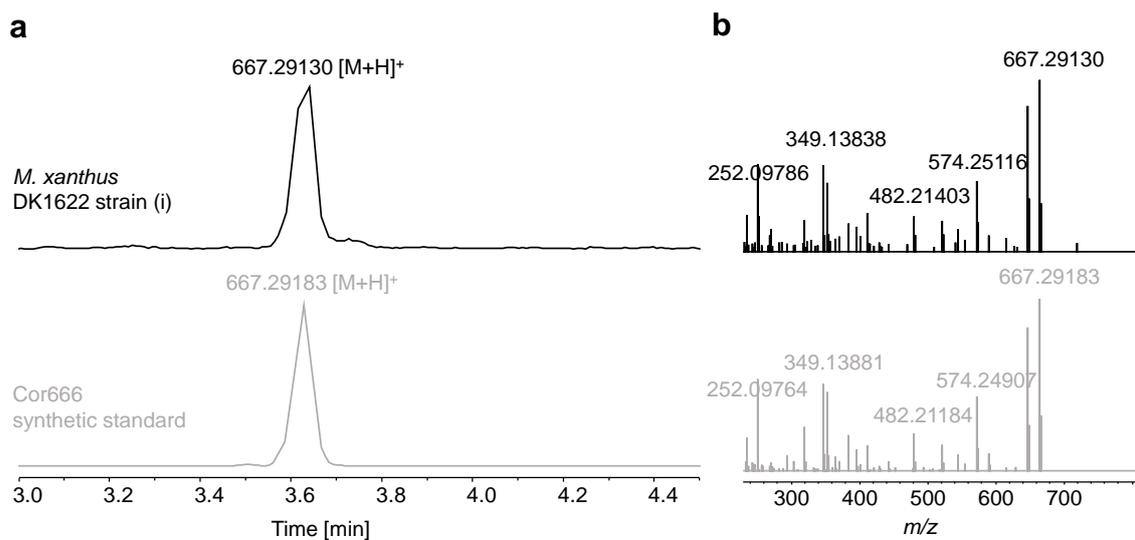




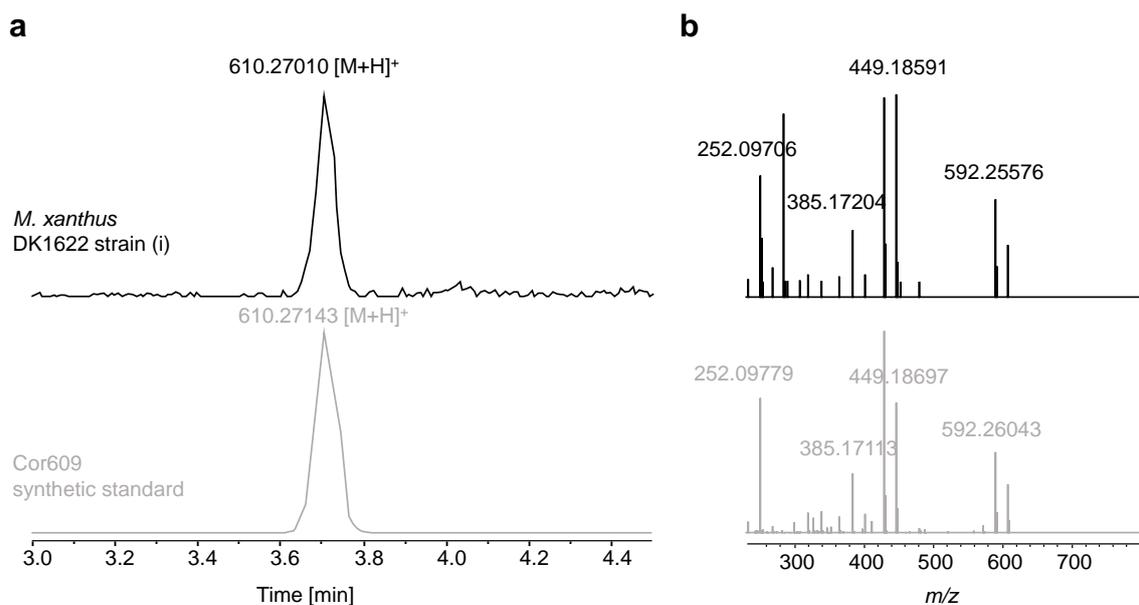
**Supplementary Figure 4 | Biosynthesis of acylated vioprolide and corramycin precursors. a:** ATP-dependent fatty acid activation by the FAAL domain leads to the production of an acylated vioprolide precursor. Maturation of the active product is achieved by hydrolysis of the fatty acid residue. Scheme was modified from Ref. <sup>8</sup>. **b:** It was hypothesized by Tesmar and coworkers that the production of acylated corramycin precursor is followed by export from the cytoplasm into the periplasm by ComC and subsequent hydrolysis by ComB.<sup>4</sup> Scheme was modified from Ref. <sup>4</sup>.



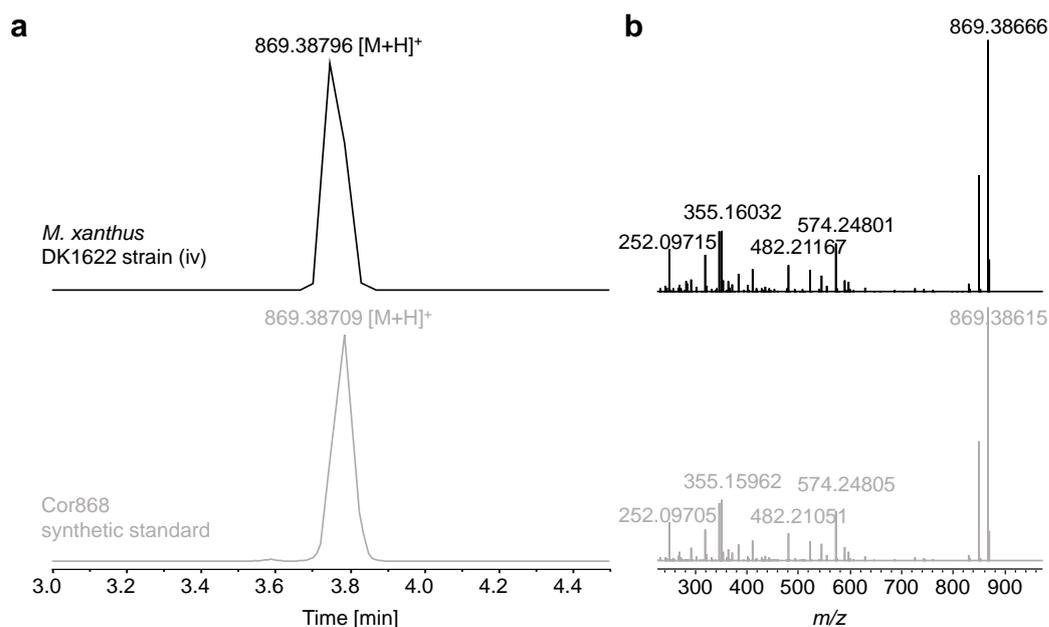
**Supplementary Figure 5 | Cloning strategy to delete parts of *comO*.** *In vivo* homologous recombination was used to replace the part to be deleted (yellow) in *comO* on pMYC20CorOp1 by a kanamycin resistance marker (green). Hydrolysis with *BsaI* was used to remove the resistance marker. Re-ligation of the plasmid lead to generation of the desired deletion constructs.



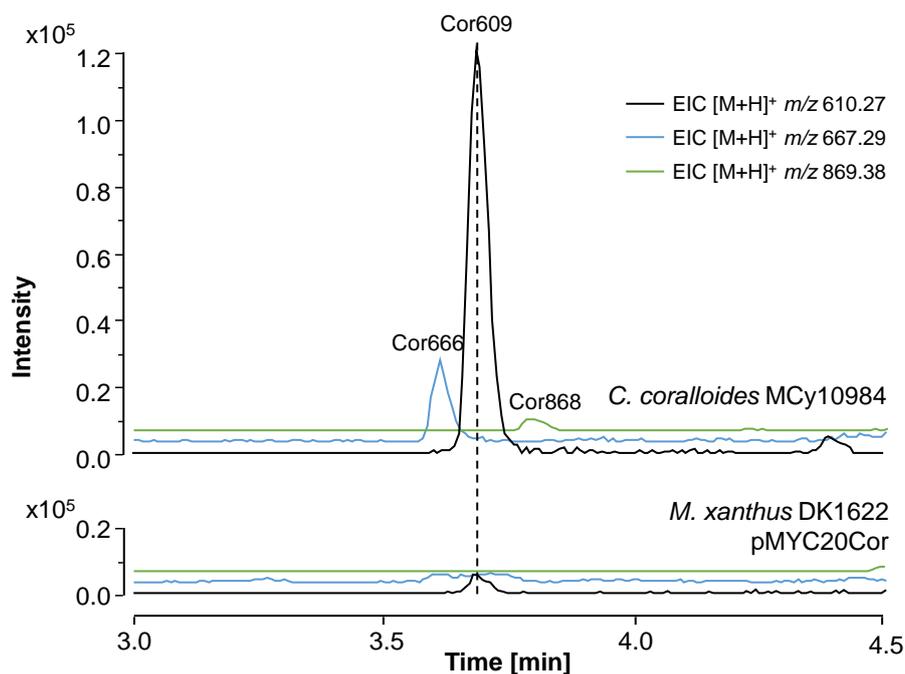
**Supplementary Figure 6 | UPLC-HRMS and MS<sup>2</sup> data of Cor666 produced in strain (i) compared to synthetic standard. a: EIC 667.30 [M+H]<sup>+</sup>. b: MS<sup>2</sup> fragmentation pattern.**



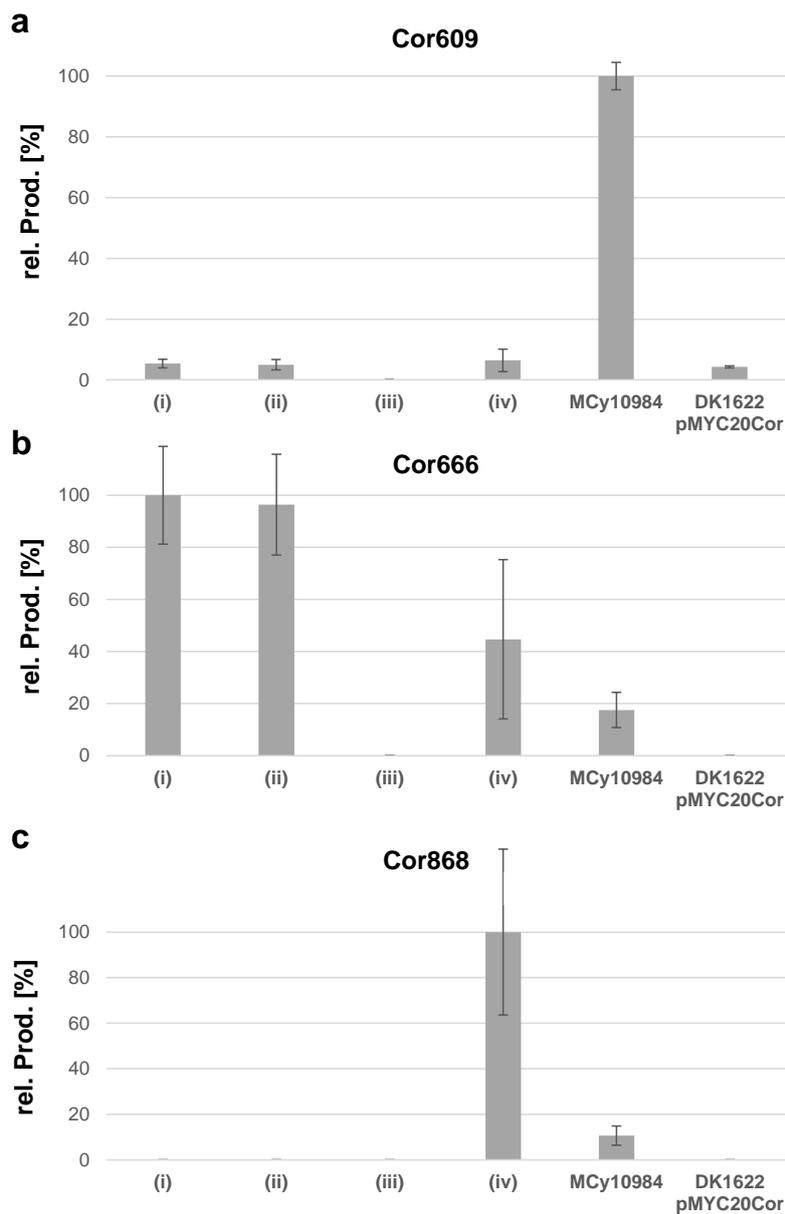
**Supplementary Figure 7 | UPLC-HRMS and MS<sup>2</sup> data of Cor609 produced in strain (ii) compared to synthetic standard. a: EIC 610.30 [M+H]<sup>+</sup>. b: MS<sup>2</sup> fragmentation pattern.**



**Supplementary Figure 8 | UPLC-HRMS and MS<sup>2</sup> data of Cor868 produced in strain (iv) compared to synthetic standard. a: EIC 869.30 [M+H]<sup>+</sup>. b: MS<sup>2</sup> fragmentation pattern.**



**Supplementary Figure 9 | UPLC-HRMS analysis of truncated corramycins in the methanolic extracts of *C. coralloides* MCy10984 and *M. xanthus* DK1622 pMYC20Cor. EICs 610.26 [M+H]<sup>+</sup> (black), 667.29 [M+H]<sup>+</sup> (blue) and 869.38 [M+H]<sup>+</sup> (green) are shown for the extracts of *C. coralloides* MCy10984 (top) and *M. xanthus* DK1622 pMYC20Cor (bottom). Cor609 was produced in both strains, whereas Cor666 and Cor868 was only detected in the extract of MCy10984.**



**Supplementary Figure 10 | Relative production yields of truncated corramycins in strains (i-iv), *M. xanthus* DK1622 pMYC20Cor and MCy10984.** Relative production levels of truncated corramycins Cor609 (a), Cor666 (b) and Cor868 (c) in strains (i-iv), MCy10984 and DK1622 pMYC20Cor. The samples were analyzed using HPLC-HRMS and the peak surface area was integrated manually for yield comparison. The mean values from biological duplicates or triplicates (each measured in duplicates) were calculated and normalized to the mean value of the highest producing strain. Error bar: standard deviation.

## 4.6.6 Supplementary Tables

Supplementary Table 1 | Strains used and generated in this work.

Bacterial strain	Genotype	Reference
<b>Native cystobactamid producer strains</b>		
<i>C. coralloides</i> MCy10984	-	HIPS/MINS
<b>Cloning strains</b>		
<i>E. coli</i> DH10 $\beta$	F, <i>mcrA</i> , $\Delta$ ( <i>mrr-hsdRMS-mcrBC</i> ), $\Phi$ 80 <i>lacZ</i> $\Delta$ M15, $\Delta$ <i>lacX74</i> , <i>recA1</i> , <i>araD139</i> , $\Delta$ ( <i>ara-leu</i> )7697, <i>galU</i> , <i>galK</i> , <i>rpsL</i> (Str <sup>R</sup> ), <i>endA1</i> , <i>nupG</i> , $\lambda$	Invitrogen
<i>E. coli</i> HS996	F, <i>mcrA</i> , $\Delta$ ( <i>mrr-hsdRMS-mcrBC</i> ), $\Phi$ 80 <i>lacZ</i> $\Delta$ M15, $\Delta$ <i>lacX74</i> , <i>recA1</i> , <i>araD139</i> , $\Delta$ ( <i>ara-leu</i> )7697, <i>galU</i> , <i>galK</i> , <i>rpsL</i> (Str <sup>R</sup> ), <i>endA1</i> , <i>nupG</i> , <i>fhuA</i> ::IS2	Invitrogen
<i>E. coli</i> NEB10 $\beta$	<i>mcrA</i> , $\Delta$ ( <i>mrr-hsdRMS-mcrBC</i> ), $\Phi$ 80d( <i>lacZ</i> $\Delta$ M15) <i>recA1</i> , <i>relA1</i> , $\Delta$ <i>lacX74</i> , <i>recA1</i> , <i>araD139</i> , $\Delta$ ( <i>ara-leu</i> )7697, <i>galK16</i> , <i>galE15</i> , <i>rpsL</i> (Str <sup>R</sup> ), <i>endA1</i> , <i>nupG</i> , <i>fhuA</i>	New England Biolabs
<i>E. coli</i> GB05-red	F, <i>mcrA</i> , $\Delta$ ( <i>mrr-hsdRMS-mcrBC</i> ), $\Phi$ 80 <i>lacZ</i> $\Delta$ M15, $\Delta$ <i>lacX74</i> , <i>recA1</i> , <i>araD139</i> , $\Delta$ ( <i>ara-leu</i> )7697, <i>galU</i> , <i>galK</i> , <i>rpsL</i> (Str <sup>R</sup> ), <i>endA1</i> , <i>nupG</i> , $\lambda$ , $\Delta$ <i>fhuA</i> , P <sub>BAD</sub> - <i>gbaA</i> $\Delta$ <i>ybcC</i> , $\Delta$ <i>recE19</i>	Gene Bridges
<i>S. cerevisiae</i> ATCC4004247	MATa, <i>his3</i> $\Delta$ 1, <i>leu2</i> $\Delta$ 0, <i>met15</i> $\Delta$ 0, <i>ura3</i> $\Delta$ 0, <i>ydr411c</i> ::KanMX4	ATCC
<i>E. coli</i> HS996 pUC57-a	<i>E. coli</i> HS996 pUC57-a, Amp <sup>R</sup>	This work
<i>E. coli</i> HS996 pUC57-b	<i>E. coli</i> HS996 pUC57-b, Amp <sup>R</sup>	This work
<i>E. coli</i> HS996 pUC57-c	<i>E. coli</i> HS996 pUC57-c, Amp <sup>R</sup>	This work
<i>E. coli</i> HS996 pUC57-d	<i>E. coli</i> HS996 pUC57-d, Amp <sup>R</sup>	This work
<i>E. coli</i> HS996 pUC57-e	<i>E. coli</i> HS996 pUC57-e, Amp <sup>R</sup>	This work
<i>E. coli</i> HS996 pUC57-f	<i>E. coli</i> HS996 pUC57-f, Amp <sup>R</sup>	This work
<i>E. coli</i> HS996 pUC57-g	<i>E. coli</i> HS996 pUC57-g, Amp <sup>R</sup>	This work
<i>E. coli</i> HS996 pUC57-h	<i>E. coli</i> HS996 pUC57-h, Amp <sup>R</sup>	This work
<i>E. coli</i> HS996 pUC57-1	<i>E. coli</i> HS996 pUC57-1, Amp <sup>R</sup>	This work
<i>E. coli</i> HS996 pUC57-2	<i>E. coli</i> HS996 pUC57-2, Amp <sup>R</sup>	This work
<i>E. coli</i> HS996 pUC57-3	<i>E. coli</i> HS996 pUC57-3, Amp <sup>R</sup>	This work
<i>E. coli</i> HS996 pUC57-4	<i>E. coli</i> HS996 pUC57-4, Amp <sup>R</sup>	This work
<i>E. coli</i> HS996 pUC57-Dummy	<i>E. coli</i> HS996 pUC57-Dummy, Amp <sup>R</sup>	This work
<i>E. coli</i> DH10 $\beta$ pSynbio1-1	<i>E. coli</i> DH10 $\beta$ pSynbio1-1, Amp <sup>R</sup>	This work
<i>E. coli</i> DH10 $\beta$ pSynbio1-12	<i>E. coli</i> DH10 $\beta$ pSynbio1-12, Amp <sup>R</sup>	This work
<i>E. coli</i> DH10 $\beta$ pSynbio1-123	<i>E. coli</i> DH10 $\beta$ pSynbio1-123, Amp <sup>R</sup>	This work
<i>E. coli</i> DH10 $\beta$ pSynbio1-1234 (+SEs)	<i>E. coli</i> DH10 $\beta$ pSynbio1-1234 (+SEs), Amp <sup>R</sup>	This work
<i>E. coli</i> DH10 $\beta$ pSynbio1-1234	<i>E. coli</i> DH10 $\beta$ pSynbio1-1234, Amp <sup>R</sup>	This work
<i>S. cerevisiae</i> ATCC4004247 pMYC20preCorOp1	<i>S. cerevisiae</i> ATCC4004247 pMYC20preCorOp1, <i>LEU2</i>	This work
<i>S. cerevisiae</i> ATCC4004247 pMYC21CorOp2	<i>S. cerevisiae</i> ATCC4004247 pMYC21CorOp2, <i>LEU2</i>	This work
<i>E. coli</i> DH10 $\beta$ pMYC20preCorOp1	<i>E. coli</i> DH10 $\beta$ pMYC20preCorOp1, Otc <sup>R</sup> , Cm <sup>R</sup>	This work
<i>E. coli</i> DH10 $\beta$ pMYC21CorOp2	<i>E. coli</i> DH10 $\beta$ pMYC21CorOp2, Otc <sup>R</sup> , Cm <sup>R</sup>	This work
<i>E. coli</i> DH10 $\beta$ pMYC20CorOp1	<i>E. coli</i> DH10 $\beta$ pMYC20CorOp1, Otc <sup>R</sup> , Cm <sup>R</sup>	This work
<i>E. coli</i> NEB10 $\beta$ pMYC20Cor	<i>E. coli</i> DH10 $\beta$ pMYC20Cor, Otc <sup>R</sup> , Cm <sup>R</sup>	This work
<i>E. coli</i> NEB10 $\beta$ pMYC20CorOp1(i)kanR	<i>E. coli</i> NEB10 $\beta$ pMYC20CorOp1(i)kanR, Otc <sup>R</sup> , Cm <sup>R</sup> , Kan <sup>R</sup>	This work
<i>E. coli</i> NEB10 $\beta$ pMYC20CorOp1(ii)kanR	<i>E. coli</i> NEB10 $\beta$ pMYC20CorOp1(ii)kanR, Otc <sup>R</sup> , Cm <sup>R</sup> , Kan <sup>R</sup>	This work
<i>E. coli</i> NEB10 $\beta$ pMYC20CorOp1(iii)kanR	<i>E. coli</i> NEB10 $\beta$ pMYC20CorOp1(iii)kanR, Otc <sup>R</sup> , Cm <sup>R</sup> , Kan <sup>R</sup>	This work

Bacterial strain	Genotype	Reference
<b>Cloning strains</b>		
<i>E. coli</i> NEB10β pMYC20CorOp1(iv)kanR	<i>E. coli</i> NEB10β pMYC20CorOp1(i)kanR, Otc <sup>R</sup> , Cm <sup>R</sup> , Kan <sup>R</sup>	This work
<i>E. coli</i> NEB10β pMYC20CorOp1(i)	<i>E. coli</i> NEB10β pMYC20CorOp1(i)kanR, Otc <sup>R</sup> , Cm <sup>R</sup>	This work
<i>E. coli</i> NEB10β pMYC20CorOp1(ii)	<i>E. coli</i> NEB10β pMYC20CorOp1(i)kanR, Otc <sup>R</sup> , Cm <sup>R</sup>	This work
<i>E. coli</i> NEB10β pMYC20CorOp1(iii)	<i>E. coli</i> NEB10β pMYC20CorOp1(i)kanR, Otc <sup>R</sup> , Cm <sup>R</sup>	This work
<i>E. coli</i> NEB10β pMYC20CorOp1(iv)	<i>E. coli</i> NEB10β pMYC20CorOp1(i)kanR, Otc <sup>R</sup> , Cm <sup>R</sup>	This work
<b>Heterologous host and production strains</b>		
<i>M. xanthus</i> DK1622	-	HIPS/MINS
<i>M. xanthus</i> DK1622 pMYC20Cor	<i>M. xanthus</i> DK1622 pMYC20Cor, Otc <sup>R</sup>	This work
<i>M. xanthus</i> DK1622 pMYC20CorOp1(i) pMYC21CorOp2	<i>M. xanthus</i> DK1622 pMYC20CorOp1(i) pMYC21CorOp2, Otc <sup>R</sup>	This work
<i>M. xanthus</i> DK1622 pMYC20CorOp1(ii) pMYC21CorOp2	<i>M. xanthus</i> DK1622 pMYC20CorOp1(ii) pMYC21CorOp2, Otc <sup>R</sup>	This work
<i>M. xanthus</i> DK1622 pMYC20CorOp1(iii) pMYC21CorOp2	<i>M. xanthus</i> DK1622 pMYC20CorOp1(iii) pMYC21CorOp2, Otc <sup>R</sup>	This work
<i>M. xanthus</i> DK1622 pMYC20CorOp1(iv) pMYC21CorOp2	<i>M. xanthus</i> DK1622 pMYC20CorOp1(iv) pMYC21CorOp2, Otc <sup>R</sup>	This work

**Supplementary Table 2 | Plasmids used and generated in this work.**

Plasmid	Genotype	Reference
pSynbio1	Non-integrative plasmid for cloning in <i>E. coli</i> ; <i>oriV</i> and <i>trfA</i> from RK2 plasmid, <i>bla</i> (Amp <sup>R</sup> ), MCS	3
pMYC	pMYC vector backbone; <i>p15A</i> ori, <i>cat</i> (Cm <sup>R</sup> ) from pACYC184, <i>traJ</i> , <i>oriT</i> , <i>CEN6/ARS4</i> , <i>URA3</i> , t <sub>D2</sub> terminator from <i>Myxococcus</i> phage Mx8	This work
pMYC20	TetR-mx8 cloned into pMYC; <i>tetR</i> (Otc <sup>R</sup> ) from pALTER(R)-1, <i>mx8</i> integrase from <i>Myxococcus</i> phage Mx8	This work
pMYC21	KanR-mx9 cloned into pMYC; <i>kanR</i> (Kan <sup>R</sup> ) from pACYC177, <i>mx9</i> integrase from <i>Myxococcus</i> phage Mx9	This work
pSynbio1-1	Gene synthesis product 1 (fragment 1 of <i>comO</i> ) cloned into pSynbio1	This work
pSynbio1-12	Gene synthesis product 2 (fragment 2 of <i>comO</i> ) cloned into pSynbio1-1	This work
pSynbio1-123	Gene synthesis product 3 (fragment 3 of <i>comO</i> ) cloned into pSynbio1-1	This work
pSynbio1-1234 (+SEs)	Gene synthesis product 4 (fragment 4 of <i>comO</i> ) cloned into pSynbio1-1; contains SEs	This work
pSynbio1-1234	Gene synthesis product 4 (fragment 4 of <i>comO</i> ) cloned into pSynbio1-1; SEs removed	This work
pMYC20preCorOp1	Gene synthesis product a ( <i>ura3</i> homology, vanillate promoter and repressor <i>vanR</i> , <i>comI</i> , <i>comJ</i> and 5' end of <i>comK</i> ), gene synthesis product b (3' end of <i>comK</i> and 5' end of <i>comL</i> ), gene synthesis product c ( <i>comL</i> ), gene synthesis product d (3' end of <i>comL</i> and 5' end of <i>comM</i> ), gene synthesis product e (3' end of <i>comM</i> and <i>comN</i> ) and Dummy sequence ( <i>rpsL</i> , <i>LEU2</i> and <i>ura3</i> homology) cloned into pMYC20 by TAR	This work
pMYC21CorOp2	Gene synthesis product f ( <i>ura3</i> homology, vanillate promoter, <i>comB</i> , <i>comE</i> and <i>comP</i> ), gene synthesis product g ( <i>comA</i> , <i>comH</i> , <i>comF</i> , <i>comE</i> ) and gene synthesis product h ( <i>comD</i> , <i>comG</i> , <i>LEU2</i> and <i>ura3</i> homology)	This work
pMYC20CorOp1	<i>comO</i> fragments 1-4 (without SEs) cloned into pMYC20preCorOp1	This work
pMYC20Cor	CorOp2 cloned into pMYC20CorOp1	This work

Plasmid	Genotype	Reference
pMYC20CorOp1(i)kanR	<i>kanR</i> cloned into pMYC20CorOp1 via RedET recombineering	This work
pMYC20CorOp1(ii)kanR	<i>kanR</i> cloned into pMYC20CorOp1 via RedET recombineering	This work
pMYC20CorOp1(iii)kanR	<i>kanR</i> cloned into pMYC20CorOp1 via RedET recombineering	This work
pMYC20CorOp1(iv)kanR	<i>kanR</i> cloned into pMYC20CorOp1 via RedET recombineering	This work
pMYC20CorOp1(i)	<i>kanR</i> removed from pMYC20CorOp1(i)kanR	This work
pMYC20CorOp1(ii)	<i>kanR</i> removed from pMYC20CorOp1(ii)kanR	This work
pMYC20CorOp1(iii)	<i>kanR</i> removed from pMYC20CorOp1(iii)kanR	This work
pMYC20CorOp1(iv)	<i>kanR</i> removed from pMYC20CorOp1(iv)kanR	This work

**Supplementary Table 3 | Oligonucleotides used in this work.** Restriction sites (R-sites) are underlined and binding regions are marked in bold.

Oligonucleotide	Sequence (5'-3')
Mx8-attP-up2	CGACGGTGCCGACAAATAC
Mx8-attB-up2	GCGCACTGGACCATCACGTC
Mx8-attP-down	GGCTTGTGCCAGTCAACTGCG
Mx8-attB-down	CGGATAGCTCAGCGGTAGAG
Mx9attB1_up	TGCCAGGGCTTACGGCTTC
Mx9attB1_down	CAGCACGGGTGCAGCAAC
Mx9attP_up	GCGCCGAACTTAACAAGTTG
Mx9attP_down	TCCAGGTCCTCACGCTTGAC
KanR-1-F	AGGTGGACCGCAAGGCGCTGCCGGCTCCGGAGGCGGTGGCCGCCGAG ACGTGAGACCTTGGACAGCAAGCGAACCGG
KanR-1-R	CGAACGAGCTGCAGCTCCAATGCGTCTCGCGGGCCACCGCCGAGGCC GACGTCTGAGACCTCAGAAGAACTCGTCAAGAAG TCTTGAAGCGTCCACCGTGGAGGTGCTGGCCGCGCAGGTGGAGCAGG
KanR-2-F	CATGAGACCTTGGACAGCAAGCGAACCGG
KanR-2-R	CGGCCTGTTCCCTTCGTGGCGAGGGGAACCAGCGACGAGGCGCGCCCC GGTGCTGAGACCTCAGAAGAACTCGTCAAGAAG
KanR-3-F	AGTTGCCCTGCAGAGCGCGGAGGATCAGCGACTGCTGCTCTCCGCGT GGTGAGACCTTGGACAGCAAGCGAACCGG
KanR-3-R	TGCCTGAGACCTCAGAAGAACTCGTCAAGAAG
KanR-4-F	TGACCGAGACGGATCAGGGGCTCTCAGGTGCGTGGAGTTAACAGCG ACTGAGACCTTGGACAGCAAGCGAACCGG
KanR-4-R	AGCAGCTTCAGGTGCCTCAGCATCCGCTCCGCTGTCGCCGGCTCGAAC AGGTCGTGAGACCTCAGAAGAACTCGTCAAGAAG

**Supplementary Table 4 | Genetic elements used in the design of the modified gene cluster.** Nucleotide position numbering in modified cluster refers to CorOp1 and CorOp2 separately. Genetic elements which are not part of the cluster but necessary for cloning are labelled in light blue.

Genetic element	Nucleotide position in modified cluster	Sequence origin (GenBank accession)	Nucleotide position in original sequence
<b>CysOp1</b>			
<i>URA3</i> homology left	0 – 100	pRS416 (U03450)	606 - 705
Pvan (+ <i>vanR</i> )	101 – 1,180	pMR3679 <sup>44</sup>	1862 – 2941
<i>comI</i> - <i>comO</i>	1,181 – 46,943	MCy10984 BGC	unpublished (chapter 3)
<i>Xma</i> II	36,817 – 36,822	-	-
tD1 terminator	48,130 – 48,178	<i>Myxococcus xanthus</i> phage Mx8 <sup>24</sup>	-
<i>Xba</i> I, spacer, <i>Afl</i> III	48,179 – 48,196	-	-
<i>LEU2</i>	48,197 – 50,431	pRS415 (U03449)	3498 - 5732
<i>URA3</i> homology right	50,432 – 50,531	pRS416 (U03450)	506 - 605
<b>CysOp2</b>			
<i>URA3</i> homology left	0 – 100	pRS416 (U03450)	606 - 705
<i>Xba</i> I	101 – 106	-	-
Pvan	107 – 236	pMR3679 <sup>44</sup>	2,812 – 2,941
<i>comB</i> , <i>E</i> , <i>P</i> , <i>A</i> , <i>H</i> , <i>F</i> , <i>E</i> , <i>D</i> , <i>G</i>	237 – 10,192	MCy10984 BGC	unpublished (chapter 3)
tD1 terminator	10,193 – 10,241	<i>Myxococcus xanthus</i> phage Mx8 <sup>24</sup>	-
<i>Afl</i> III	10,242 – 10,247	-	-
<i>LEU2</i>	10,248 – 12,482	pRS415 (U03449)	3498 - 5732
<i>URA3</i> homology right	10,483 – 10,582	pRS416 (U03450)	506 - 605

**Supplementary Table 5 | Name, size and restriction sites or splitter elements (SEs) of the thirteen gene synthesis fragments.** Spacer sequence (sp) was introduced between SE restriction sites.

Fragment	Description	Size [bp]	Flanking restriction sites or SEs
a	<i>URA3</i> homology, <i>vanR</i> , P <sub>van</sub> , <i>comI</i> , <i>comJ</i> , (5') 1273 bp of <i>comK</i>	5,245	5'- <i>BsaI</i> , 3'- <i>BsaI</i>
b	(3') 4143 bp of <i>comK</i> , (5') 1037 bp of <i>comL</i>	5,216	5'- <i>BsaI</i> , 3'- <i>BsaI</i>
c	<i>comL</i>	5,216	5'- <i>BsaI</i> , 3'- <i>BsaI</i>
d	(3') 3,485 bp of <i>comL</i> , (5') 1726 bp of <i>comM</i>	5,216	5'- <i>BsaI</i> , 3'- <i>BsaI</i>
e	(3') 2,838 bp of <i>comM</i> , <i>comN</i>	4,415	5'- <i>BsaI</i> , 3'- <i>BsaI</i>
f	<i>comB</i> , <i>comE</i> , <i>comP</i>	4,176	5'- <i>BsaI</i> , 3'- <i>BsaI</i>
g	<i>comA</i> , <i>comH</i> , <i>comF</i> , <i>comE</i> , (5') 425 bp of <i>comD</i>	4,216	5'- <i>BsaI</i> , 3'- <i>BsaI</i>
h	(3') 872 bp of <i>comD</i> , <i>comG</i>	4,399	5'- <i>BsaI</i> , 3'- <i>BsaI</i>
1	<i>comO</i> fragment 1	6,240	SE (5'): <i>KpnI-BsaI</i> SE (3'): <i>BsaI-HindIII-sp(GACCTA)-PmeI</i>
2	<i>comO</i> fragment 2	6,244	SE (5'): <i>HindIII-BsaI</i> SE (3'): <i>BsaI-NdeI-sp(GACCTA)-PmeI</i>
3	<i>comO</i> fragment 3	5,844	SE (5'): <i>NdeI-BsaI</i> SE (3'): <i>BsaI-XbaI-sp(GGCCTA)-PmeI</i>
4	<i>comO</i> fragment 4	5,057	SE (5'): <i>XbaI-BsaI</i> SE (3'): <i>BsaI-PmeI</i>
Dummy	<i>comO</i> homology, <i>rpsL</i> , <i>comO</i> homology, LEU2, <i>URA3</i> homology	3,177	-

**Supplementary Table 6 | Source of vector and insert DNA and restriction endonucleases used for construction of plasmids in this work.**

Product generated	Vector	Insert	Insert source	Restriction enzyme
pSynbio1-1	pSynbio1	1	Gene synthesis fragment	<i>KpnI/PmeI</i>
pSynbio1-12	pSynbio1-1	2	Gene synthesis fragment	<i>HindIII/PmeI</i>
pSynbio1-123	pSynbio1-12	3	Gene synthesis fragment	<i>NdeI/PmeI</i>
pSynbio1-1234 (+SEs)	pSynbio1-123	4	Gene synthesis fragment	<i>XbaI/PmeI</i>
pSynbio1-1234	pSynbio1-1234	-	-	<i>BsaI</i>
pMYC20CorOp1	pMYC20preCorOp1	1234 ( <i>comO</i> )	pSynbio1-1234	<i>BsaI</i>
pMYC20Cor	pMYC20CorOp1	CorOp2	pMYC21CorOp2	<i>XbaI/AflIII</i>

**Supplementary Table 7 | Modified plasmids generated via Red/ET recombineering and restriction hydrolysis/re-ligation.**

Product generated	Description	Vector used for Red/ET	Red/ET product
pMYC20CorOp1(i)	Deletion of <i>comO</i> module 7 T domain to module 12 A domain	pMYC20CorOp1	pMYC20CorOp1(i)kanR
pMYC20CorOp1(ii)	Deletion of <i>comO</i> module 8 C domain to module 12 T domain	pMYC20CorOp1	pMYC20CorOp1(ii)kanR
pMYC20CorOp1(iii)	Deletion from <i>comO</i> module 8 A domain on	pMYC20CorOp1	pMYC20CorOp1(iii)kanR
pMYC20CorOp1(iv)	Deletion of <i>comO</i> module 9 A domain to module 12 C domain	pMYC20CorOp1	pMYC20CorOp1(iv)kanR

#### 4.7 Supplementary Information References

44. 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; 10.1128/JB.01110-12 (2012).
45. Kouprina, N. & Larionov, V. Selective isolation of genomic loci from complex genomes by transformation-associated recombination cloning in the yeast *Saccharomyces cerevisiae*. *Nat. Protoc.* **3**, 371–377; 10.1038/nprot.2008.5 (2008)



## 5 Discussion

The central topic of this thesis is the heterologous expression of bacterial secondary metabolite pathways for the production of pharmaceutically relevant NPs. The major focus was placed on the elucidation of the biosynthesis pathways and the production of structurally altered NP congeners. Therefore, not only heterologous expression but also *in vitro* reconstitution of enzyme activities and *in silico* analysis of the respective BGCs proved to be powerful tools to shed light on complex biosynthesis steps.

The second and fourth chapter focus on the *in silico* design, assembly and heterologous expression of modified BGCs in *M. xanthus* DK1622 for the production of the antibiotic classes of cystobactamids and corramycins. In both cases, targeted deletions of entire genes or part of genes lead to the production of entirely new or truncated congeners of the NPs. More specifically, chapter 2 discusses the elucidation of the cystobactamid biosynthesis based on the heterologous expression experiments in combination with *in vitro* experiments, whereas chapter 4 discusses the heterologous production of truncated corramycins for the implementation into a semi-synthetic approach to obtain pharmaceutically improved congeners. Chapter 3 discusses the first isolation of corramycins from two *C. coralloides* strains, the identification of the respective BGC and proposal of a putative biosynthesis pathway. The work in chapter 3, which was mainly performed by Dr. Alexander von Tesmar (from HIPS side),<sup>1</sup> therefore set the basis for the establishment of the heterologous production platform described in chapter 4.

### 5.1 Major drawbacks in antibiotic development

The development of an antibiotic into a marketable state is a laborious process with a failure rate of 95 %, <sup>2</sup> costs of hundreds of millions of US dollars and time scales up to 10 and more years. <sup>3</sup> The discovery and early phase development of antibiotics is often undertaken by small and medium-sized companies. A recent study investigated costs and durations of the early phase antibiotic development by conducting a survey with 25 companies participating. <sup>4</sup> It was estimated that lead compound identification costs from 100 thousand to over a million US dollars taking 6 months up to 4 years. A lead compound is defined as a compound with biological activity and pharmacological properties bearing the potential to be developed to a therapeutic after refinement of suboptimal structures. Lead-like compounds serve as starting point for chemical modifications for the generation of congeners with more drug-like properties. Apart from a promising antibacterial activity (ideally against clinically relevant pathogenic strains), a lead-like antibacterial has to fulfill specific selection criteria such as structural novelty, a new target or mode-of-action, a minimal number of diversification points and low cytotoxicity. <sup>5</sup> Physicochemical properties such as aqueous solubility,

microsomal stability, CYP450 inhibition and cell permeability of lead-like compounds may be suboptimal and have to be targeted in optimization processes to attain drug-like properties.<sup>6</sup> For lead compound optimization the companies spent (or plan to spend) around 1 to 5 million US dollar over a time period of another 6 months to 4 years.<sup>4</sup> Preclinical testing takes 6 months to 2 years and costs around 1 to 10 million US dollar, whereas phase II clinical trials take 1 to 4 years costing 1 to 20 million US dollars. Reasons for the high failure rate in antibiotic development are the overall profitability,<sup>3</sup> the discovery process to identify structures exhibiting new targets, and the clinical trials, where only 25 % of the antibiotic candidates succeed.<sup>2</sup> As antimicrobial resistance development positively correlates with the prevalence of antibiotic use in human health care,<sup>7</sup> the thoughtful use of established antibiotic classes is still sufficient to treat most of the infections.<sup>8</sup> As a result, new antibiotics are often only used as choice of last resort to treat multidrug-resistant pathogens, finally resulting in low unit sales.

## **5.2 Limitations and solution approaches in antibiotic drug discovery from microbial sources**

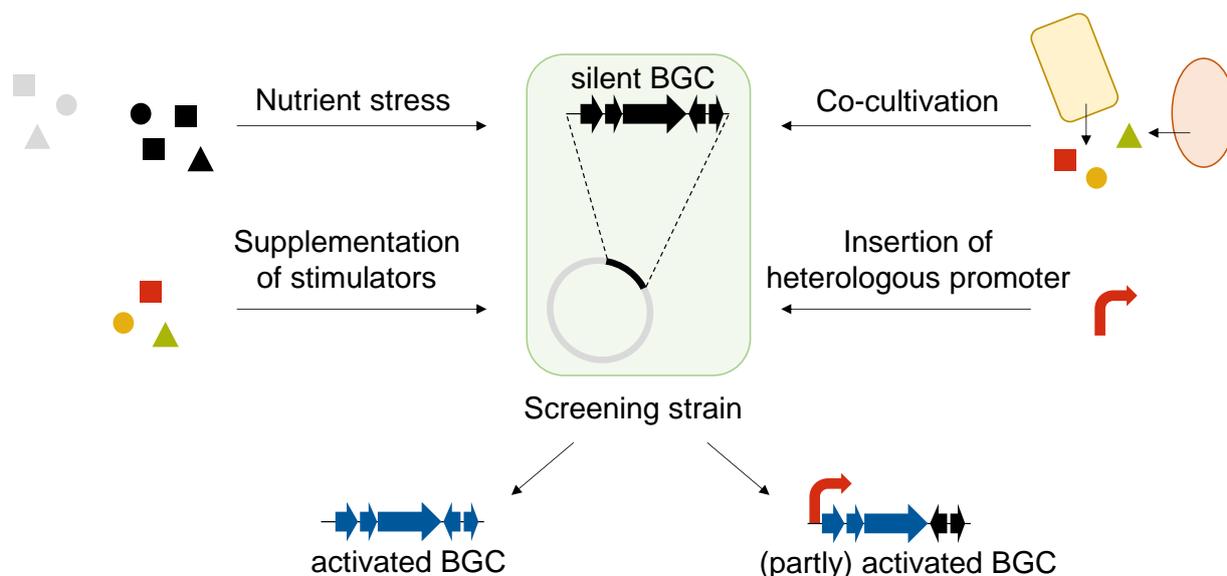
Most of the major antibiotic classes discovered in the “golden age of antibiotics” derived from actinobacteria,<sup>9</sup> which nowadays increases the probability to rediscover known compound classes with known targets when continuing the same screening approaches and source organisms. For instance, this led to a decrease in the discovery rate from 70 to 100 new antibacterial compound classes per year in the 1970s to early 1980s to only around 20 per year in the late 1980s to 1990s deriving from actinomycetes.<sup>10</sup> Without a paradigm change in the traditional screening methodology, the discovery of new compound classes in actinomycetes becomes more and more unlikely. Nevertheless, this does not mean that the time of NP-based screening is over, because in general most of the FDA-approved drugs, and antibiotics in particular, are still based on the structures Nature provided.<sup>11</sup> It means that new NP screening systems have to be developed. One recent example is the identification of darobactin.<sup>12</sup> The chosen source organism was a *Photorhabdus* strain, which is a gut commensal in entomopathogenic nematodes and was thus hypothesized to share similar requirements for antibiotics with human commensals. Darobactin showed superior antimicrobial activity against a large number of Gram-negative human pathogens, whereas almost no toxicity in human cell lines or antimicrobial activity against human gut commensals, including Gram-negative symbionts of the genus *Bacteroides*, were observed. This activity was attributed to the new target BamA, which is an outer-membrane chaperone in Gram-negative bacteria. Notably, the authors did not discuss why Gram-negative human gut commensals are not affected by darobactin. Possible explanations might be e.g. an altered darobactin target site in BamA or the presence of an analog compensating the functional loss of BamA.

Switching the source organisms from highly explored to underexplored bacterial orders is a promising alternative that allows discovery of new compound classes even with old screening methods. The main reason for this is that the biosynthetic machineries in underexplored bacteria generally differ from those in highly explored bacteria with higher probability. The growing number of NRPS and PKS systems, which differ in their structure and functionality from previously classified systems (mainly deriving from actinobacteria), is due to a reinforced exploration of other bacterial sources. The different chemistry executed by these uncharacterized systems make the discovery of completely new structures with unprecedented modes of action more likely. As described before, myxobacteria were shown to have a large potential to produce NPs with various biological activities.<sup>13</sup> Interestingly, only a minor fraction of the myxobacterial strains living in various habitats were isolated and cultivated under laboratory conditions so far. It was estimated that 1 g of soil can harbor  $10^{8-11}$  bacterial cells and  $10^{3-6}$  different species,<sup>14,15</sup> whereas, only 5 to 10 different myxobacterial species and less than  $10^5$  myxobacterial cells are typically yielded from this amount of soil.<sup>16-18</sup> This displays less than 0.1 % of the total cells and species estimated in 1 g of soil. However, based on a 16S rRNA pyrosequencing experiment of a soil sample from Shandong University campus, the order of myxococcales accounted for 4.1 % of the total bacterial sequences,<sup>18</sup> indicating that the vast majority of myxobacterial species was not accessed yet. Although the proportion of myxobacterial species compared to total bacterial species might vary tremendously in different soil samples, this study gives a clear indication of the biosynthetic potential that lies hidden behind the huge number of yet uncultured myxobacteria.

Furthermore, myxobacteria harbor a great genetic potential for the production of NPs as they have the largest genomes in the kingdom of bacteria. However, the production titers of promising antimicrobials are often below the desired threshold that is required for further lead development. Naturally, bacteria produce antibiotics to gain a competitive advantage over rival strains that are living in the same habitat and fighting for the same resources. In contrast to the conditions in the natural environment, under laboratory conditions bacteria are often isolated and cultivated without nutritional limitations. These cultivation conditions eliminate the necessity for the bacterium to produce antibiotics, resulting in downregulation of the gene expression of BGCs (cryptic or silent BGCs) encoding the biosynthetic machinery for antibiotic production. Thus, the antibiotic production that scientists observe may be only the result of basal gene expression by so-called leaky promoters. On the other hand, the lack of knowledge about newly discovered bacterial strains and appropriate cultivation media may also lead to unwanted nutritional limitations, which result in slow growth and thus low production titers. It was shown that nutrient stress or the accumulation of specific stimulator molecules such as *N*-acetylglucosamine (part of the bacterial cell wall),<sup>19</sup> desferrioxamines (siderophores),<sup>20</sup> promomycin (ionophore)<sup>21</sup> or  $\gamma$ -butyrolactone<sup>22</sup> may trigger the

production of antibiotics by inducing a signal cascade leading to the activation of global antibiotic biosynthesis transcription regulators in *Streptomyces*. The knowledge about global antibiotic biosynthesis activators in actinobacteria was for example used for high-throughput screening of silent BGCs.<sup>23,24</sup> However, if only limited information about nutritional requirements and global activators are available, finding optimal cultivation conditions for a certain production strain is a ridge walk between sufficient nutrients and enough stress to achieve acceptable yields. Especially most of the myxobacteria are not yet cultivable with an efficacy that enables high-throughput screening of global activators for antibiotic production. Nevertheless, in a few cases the regulation of secondary metabolite production was investigated, e.g. for *Sorangium cellulosum* and *Cystobacter fuscus*.<sup>25,26</sup>

Another approach to trigger antibiotic production is the microbial co-cultivation including the cultivation of bacteria with other bacteria but also with microbes like fungi. The secretion of small molecules by one strain thereby induces the expression of formerly silent BGCs leading to increased production titers of the respective secondary metabolites. Interestingly, not only antagonistic but also mutualistic interactions between the cocultivated species can lead to stimulation of antibiotic production.<sup>27</sup> Another method to activate silent BGCs is the insertion of a heterologous promoter upstream of the BGC, thereby forcing the strain to express the genes and produce the desired metabolite. This method was for example used to activate a type II PKS BGC in the myxobacterium *Pyxidicoccus fallax* An d48, leading to the production of the new topoisomerase inhibitors called pyxidicyclines.<sup>28</sup> However, this method has some major drawbacks. The target strain has to be genetically accessible, e.g. in terms of transformability or the availability of usable selection markers and heterologous promoters. Furthermore, complex BGCs consisting of numerous transcriptional units cannot be fully activated by the insertion of one heterologous promoter, resulting in unbalanced or incomplete gene expression of only a few biosynthetic genes, probably leading to the production of precursor molecules missing their biological activity. Thus, this method is more applicable for BGCs with a simple organization level. Figure 1 shows some of the above-mentioned strategies that are used to activate silent BGCs.



**Figure 1 | Example strategies to induce gene expression of silent BGCs.** The cultivation of bacterial strains in nutrient-poor medium might lead to nutrient stress inducing gene expression. The direct supplementation of stimulators or the cocultivation with competitive or mutualistic bacteria or fungi can also induce gene expression. The introduction of a heterologous promoter upstream of the BGC might lead to (partial) activation of the gene expression.

In this thesis, the discovery of the new myxobacterial antibiotic corramycin was described (chapter 3),<sup>1</sup> following a traditional screening approach in which a huge library of nearly 4,000 myxobacterial crude extracts was tested in a high-throughput manner for their activity against *E. coli*. In the underexplored bacterial order of myxococcales this screening method may still be the simplest way to find new interesting molecules. However, one may expect a decline in myxobacterial antibiotic discovery comparable to that after the “golden age of antibiotics” when the screening methodology for actinomycetes had to be improved to further yield novel antibiotic classes. One advantage for the screening process in myxobacteria is that for actinomycetes many innovative screening approaches have already been established, which only have to be adapted. This may quench the decline in the myxobacterial compound screening process in the future.

After the initial discovery process, the low production titers of the target compounds often tempt pharmaceutical companies to develop chemical synthesis routes. Total synthesis (in many cases) does not only provide sufficient amounts of the target molecule, but also opens up nearly unlimited options for structure engineering to improve the antibacterial activity spectrum or other pharmacokinetic properties. However, due to the complex structure of some NP building blocks, total synthesis sometimes reaches its limits. Another elegant way to circumvent the low production titer issue, e.g. caused by slow growth and genetic inaccessibility of native producer strains, is the heterologous expression of the entire secondary metabolite pathway in a well-described host strain. This method also offers the structure engineering of the target compound; however, within the limitation of natural building blocks. Nevertheless, the (heterologous) production of potentially

structure-engineered NPs also allows the combination with chemical synthesis approaches (semi-synthesis), thus opening up even more possibilities for structure engineering.

### 5.3 Heterologous expression of myxobacterial secondary metabolite pathways

As described before, due to a lack of knowledge many myxobacterial strains are difficult to cultivate or genetically manipulate, if at all. Consequently, many interesting secondary metabolites remain undiscovered behind silent BGCs. Heterologous expression of the respective BGCs in well characterized host strains has not only been proven to enable production of the desired metabolites, but also to increase the production titer, perform structure engineering and enlighten the crucial biosynthetic steps. For myxobacterial NRPS/PKS pathways a number of heterologous host strains have already been employed such as several *M. xanthus* strains, *Coralloccoccus macrosporus*, *Burkholderia*, *Streptomyces* sp., *Pseudomonas* sp. and *E. coli* (Table 1).

The size of the expressed BGCs ranged from a small 1 kb cluster, exemplified by the flaviolin BGC from *Sorangium cellulosum* So ce 56,<sup>39</sup> up to large clusters over 60 kb such as the soraphen BGC from *S. cellulosum* So ce 26<sup>49</sup> or the coralloyronin BGC from *C. coralloides*.<sup>32</sup> The production yields vary significantly from 1  $\mu\text{g L}^{-1}$  of epothilone heterologously produced in *E. coli*<sup>38</sup> up to over 500  $\text{mg L}^{-1}$  of myxochromide A, myxochromide S and vioprolide in *M. xanthus* DK1622 and *C. macrosporus* GT-2.<sup>40,41,43,50</sup> Even though there is one example of a heterologous *Pseudomonas putida* KT2440 strain producing 8  $\text{mg L}^{-1}$  of myxochromide S,<sup>42</sup> myxobacterial strains seem to be the better choice as heterologous host strains for the expression of myxobacterial BGCs. To stick to the example of heterologous production of myxochromide S, *M. xanthus* DK1622 and *C. macrosporus* GT-2 produced more than 60-fold up to 75-fold of the target compound compared to *P. putida* KT2440. Consequently, phylogenetic relatedness seems to play an important role for the heterologous expression as GC content, codon usage bias and the overall transcriptional and translational protein machinery are influencing factors.

**Table 1 | Selected examples of heterologously produced compounds from myxobacteria and in myxobacteria.** The given yields may refer to one or several analogs of a certain compound class. Mm: Minimal medium; CDW: cell dry weight. <sup>[a]</sup> additionally supplemented with amino acid mixture. <sup>[b]</sup> additionally supplemented with L-leucine and vitamin B<sub>12</sub>. <sup>[c]</sup> additionally supplemented with D,L-pipecolic acid. This table was modified based on Ref. <sup>29</sup>.

Compound	Cluster origin	Heterologous producer	Pathway type	Size [kb]	Promoter	Medium	Yield	Ref.
Argyrin A, B	<i>Cystobacter</i> sp. SBCb004	<i>M. xanthus</i> DK1622 $\Delta$ <i>mchA-tet</i>	NRPS	33	<i>P<sub>nptII</sub></i>	M7/s4 <sup>[a]</sup>	160 mg L <sup>-1</sup>	30
				33	<i>P<sub>nptII</sub></i> / <i>P<sub>van</sub></i>	M7/s4 <sup>[a]</sup>	250 mg L <sup>-1</sup>	29
Bengamide	<i>Myxococcus virescens</i> ST200611	<i>M. xanthus</i> DK1622	PKS/NRPS	25	<i>P<sub>nptII</sub></i>	CTT	> 10 mg L <sup>-1</sup>	31
Corallopyronin A	<i>Coralloccoccus coralloides</i> B035	<i>M. xanthus</i> DK1622 $\Delta$ <i>mchA-tet</i>	PKS/NRPS	65	<i>P<sub>nptII</sub></i>	M7/s6	37 mg L <sup>-1</sup> ~	32
		<i>M. xanthus</i> DK1622		65	<i>P<sub>van</sub></i>	M7/s6	100 mg L <sup>-1</sup>	33
Corramycin	<i>Coralloccoccus coralloides</i> Mcy10984	<i>M. xanthus</i> DK1622	PKS/NRPS	58	<i>P<sub>van</sub></i>	M7/s4	0.5 mg L <sup>-1</sup>	this work
Cystobactamid	<i>Cystobacter velatus</i> Cbv34	<i>M. xanthus</i> DK1622	NRPS	55	<i>P<sub>van</sub></i>	M7/s4	> 10 mg L <sup>-1</sup>	this work
Epothilone	<i>Sorangium cellulosum</i> So ce90	<i>M. xanthus</i> DK1622	PKS/NRPS	56	native	CMM	0.1 – 0.4 mg L <sup>-1</sup>	34,35
		<i>Streptomyces coelicolor</i> CH999		56	<i>actI</i>	R2YE	50 – 100 µg L <sup>-1</sup>	36
		<i>Streptomyces venezuelae</i> DHS2001		56	<i>pikAI</i>	R2YE	0.4 µg L <sup>-1</sup>	37
		<i>Escherichia coli</i> K486-62-1		54	<i>P<sub>BAD</sub></i>	2xYT	1 µg L <sup>-1</sup>	38
Flaviolin	<i>S. cellulosum</i> So ce56	<i>Pseudomonas putida</i> KT2440	PKS	1.1	<i>P<sub>m</sub></i>	LB	6 mg L <sup>-1</sup>	39
Myxochromide A	<i>Myxococcus</i> sp.	<i>M. xanthus</i> DK1622 $\Delta$ <i>mchA-tet</i>	PKS/NRPS	29	<i>P<sub>nptII</sub></i>	CTT	~ 500 mg L <sup>-1</sup>	40
Myxochromide S	<i>Stigmatella aurantiaca</i> DW4/3-1	<i>M. xanthus</i> DK1622	PKS/NRPS	29	<i>P<sub>nptII</sub></i>	CTT	> 500 mg L <sup>-1</sup>	41
		<i>P. putida</i> KT2440		29	<i>P<sub>m</sub></i>	LB	8 mg L <sup>-1</sup>	42
		<i>Coralloccoccus macrosporus</i> GT-2		29	<i>P<sub>aphII</sub></i>	M	600 mg L <sup>-1</sup>	43
		<i>M. xanthus</i> DK1622 $\Delta$ <i>mchA-tet</i>		53	<i>P<sub>nptII</sub></i>	M7/s6	156 mg L <sup>-1</sup>	32
Myxothiazol	<i>S. aurantiaca</i> DW4/3-1	<i>M. xanthus</i> DZF1	PKS/NRPS	57	<i>P<sub>m</sub></i>	CTT	20 mg L <sup>-1</sup>	44
Oxytetracycline	<i>Streptomyces rimosus</i>	<i>P. putida</i> FG2005	PKS	57	<i>P<sub>m</sub></i>	Mm <sup>[b]</sup>	0.6 mg L <sup>-1</sup>	45
		<i>M. xanthus</i> DK1622		32	native	CTTYE	10 mg L <sup>-1</sup>	46
Pretubulysin	<i>Cystobacter</i> sp. CBCb004	<i>M. xanthus</i> DK1622	PKS/NRPS	40	<i>P<sub>tet</sub></i>	CTT <sup>[c]</sup>	0.2 mg L <sup>-1</sup>	47
		<i>P. putida</i>		40	<i>P<sub>tet</sub></i>	LB	1.76 µg L <sup>-1</sup>	47
PUFAs	<i>Aetherobacter fasciculatus</i> SBSr002	<i>M. xanthus</i> DK1622	PKS/FAS	18	<i>P<sub>tet</sub></i>	CTT	~ 1 mg/CDW	48
Soraphen A	<i>S. cellulosum</i> So ce26	<i>Streptomyces lividans</i> ZX7	PKS	67	<i>tipA</i>	YEME	0.3 mg L <sup>-1</sup>	49
Vioprolide	<i>Cystobacter violaceus</i> Cb vi35	<i>M. xanthus</i> DK1622	NRPS	56	<i>P<sub>tet</sub></i>	CTT	500 mg L <sup>-1</sup>	50

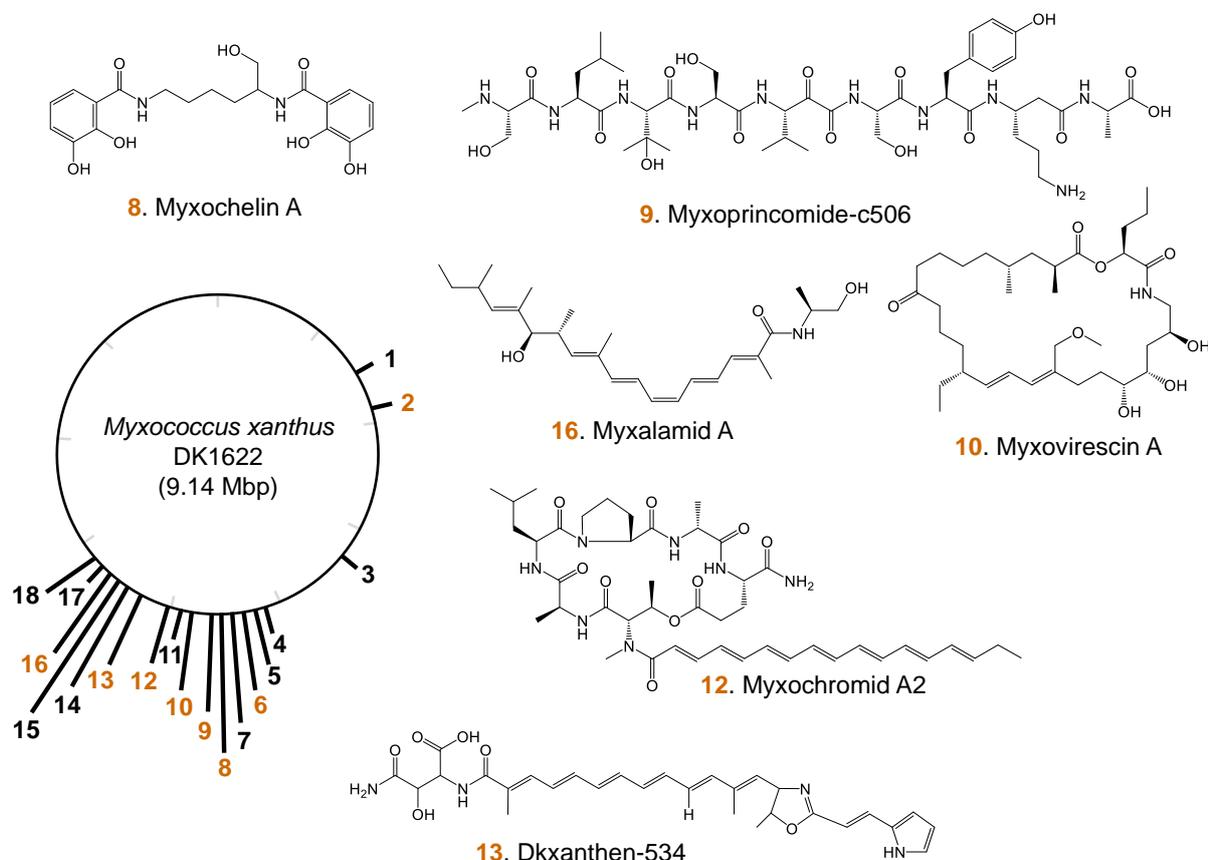
Interestingly, *C. macrosporus* GT-2 showed the higher production yield of myxochromide S with 600 mg L<sup>-1</sup> than the well-established and more commonly used strain *M. xanthus* DK1622.<sup>41,43</sup> *C. macrosporus* GT-2 is a thermophilic strain with a faster generation time than *M. xanthus* DK1622. However, since the heterologous gene expression was regulated by different promoter systems, a direct classification as a better or worse heterologous expression strain does not make sense. *M. xanthus* DK1622 was already used for the expression of numerous diverse myxobacterial BGCs with relatively high yields. Furthermore, in contrast to other myxobacteria (including *Coralloccoccus* species), for *M. xanthus* DK1622 a relatively large number of genetic engineering tools are available. The broad applicability of *M. xanthus* DK1622 as heterologous host was e.g. shown by successfully expressing the actinobacterial oxytetracycline BGC from *Streptomyces rimosus*, resulting in the production of 10 mg L<sup>-1</sup> of oxytetracycline.<sup>46</sup> However, similar to the heterologous production of myxochromide S in *P. putida* KT2440, the production yield of oxytetracycline in *M. xanthus* DK1622 was much lower compared to other heterologous host strains that were phylogenetically closer related to *S. rimosus*.<sup>51</sup> Nevertheless, the oxytetracycline BGC was expressed using the native promoter from *S. rimosus*, thus leaving room for production optimization by using established promoter systems. Furthermore, it shows that *M. xanthus* DK1622 has a sufficient potential to heterologously express BGCs from phylogenetically distant origin.

The heterologous expression of the cystobactamid BGC (52 kb; described in chapter 2) and the corramycin BGC (59 kb; described in chapter 4) in *M. xanthus* DK1622 resulted in the production of over 8 mg L<sup>-1</sup> of the major derivative Cys919-1 and only trace amounts of Cor1183, respectively. The reasons for high, low or no production of heterologously produced metabolites can be complex and often depend on several factors. In case of the heterologous production of cystobactamids, the production titer decreased significantly (< 0.1 mg L<sup>-1</sup>) when we scaled the cultivation volume up from 50 mL to > 1 L. Potential reasons for this decrease in production could be worse aeration or different cell densities in bigger shaking flasks. The reached cell density as well as the supply with essential precursors play an important role in production cultures. For example, medium optimization experiments led to the development of acetate-rich M7/s6 medium and therefore to a significant yield improvement of heterologous myxopyronin and coralloyronin production (two PKS products deriving from acetate building blocks) in *M. xanthus* DK1622 compared to the standard CTT medium.<sup>32,33</sup> For the heterologous production of argyriins in *M. xanthus* DK1622, a 3.5-fold increase in yield could be achieved by the addition of specific amino acids to M7/s4 medium.<sup>30</sup> Furthermore, the supply of cofactors that are essential for the activity of biosynthetic enzymes has to be ensured. For instance, vitamin B<sub>12</sub> is a vital cofactor for the radical SAM (S-adenosyl methionine) biosynthesis, which was shown in the biosynthesis of vioprolides.<sup>50</sup> We observed the same in the biosynthesis of cystobactamids, where the lack of vitamin B<sub>12</sub> led to

complete abolishment of cystobactamid production in the heterologous producer (data not shown), probably because the vitamin B<sub>12</sub>-dependent radical SAM methyltransferase CysS lost activity.<sup>52</sup>

As stated before, the choice of the heterologous host strain for the expression of a certain BGC is the first and probably most critical decision one has to make. In contrast to myxobacteria, actinobacteria have been extensively studied for nearly 40 years and thus several improved strains have been generated with deletion of multiple BGCs.<sup>53</sup> One example is the sequential deletion of all of the 10 PKS and NRPS BGCs in *Streptomyces coelicolor*.<sup>54</sup> A total of 1.22 Mbp, representing 14 % of the entire genome, were deleted. Another more recent example is a cluster-free *Streptomyces albus* strain, which is used for the heterologous expression of cryptic BGCs, e.g. leading to the isolation of the previously unknown fralnimycin.<sup>55</sup> Moreover, genome-minimized strains that do not belong to actinobacteria were also described, such as the *Bacillus subtilis* 168 strain PG10 that has a 36 % reduction in genome size and was used for the production of lantibiotic peptides.<sup>56,57</sup> Another engineered *Bacillus* strain, *B. amyloliquefaciens* CH12, was used for the production of bacillaene.<sup>58</sup> In this strain all PKS pathways were inactivated. Notably, also an *E. coli* genome-minimized strain was generated harboring a 4 Mb synthetic genome with genome-wide replacement of synonymous two sense codons and a stop codon.<sup>59</sup> This strain was hypothesized to be an optimal heterologous host for the production of non-canonical biopolymers.<sup>53,60</sup> The above mentioned strains may serve as perfect chassis for the heterologous expression, especially of silent BGCs, because only a minimal number of endogenous pathways compete for the same pool of precursor building blocks. Furthermore, genome-minimized strains often show significantly reduced metabolic background, which simplifies the identification and isolation of target metabolites. For myxobacteria such a chassis strain does not exist yet; however, a first step was made by deleting the myxochromide A BGC in *M. xanthus* DK1622, generating *M. xanthus* DK1622  $\Delta$ *mchA-tet*.<sup>32</sup> Myxochromide A is one of the major products in the extract of *M. xanthus* DK1622 when cultivated under laboratory conditions. The deletion of the BGC was expected to reduce metabolic background and increase amino acid supply for potential heterologously produced peptides. Furthermore, the insertion of a tetracycline resistance gene into the genomic locus of the deleted myxochromide A BGC was assumed an optimal genomic locus to serve as target site for chromosomal integration of potential heterologous BGCs. *M. xanthus* DK1622  $\Delta$ *mchA-tet* was e.g. used for the heterologous production and production optimization of the myxobacterial compounds myxopyronin and corallopyronin.<sup>32,33</sup> Nevertheless, *M. xanthus* DK1622  $\Delta$ *mchA-tet* is still far from being a heterologous chassis strain as *M. xanthus* DK1622 possesses further 23 BGCs, 18 of which are PKS/NRPS gene clusters, that would have to be deleted to get a cluster-free strain (Figure 2).<sup>61</sup> Importantly, one has to keep in mind that the deletion of genomic regions may lead e.g. to impaired

or unstable growth even if the deleted regions do not seem to be involved in this process at first glance.<sup>62</sup>



**Figure 2 | PKS/NRPS BGCs in the genome of *M. xanthus* DK1622.** BGCs of known metabolites with elucidated structure are labelled in orange. Exemplary structures of one representative from each compound class are shown. This figure was modified based on Ref. <sup>29</sup>.

The transformation of the cystobactamid and corramycin BGCs into *M. xanthus* DK1622  $\Delta$ *mchA-tet* would be the next step to potentially achieve higher production of the target compounds. Furthermore, the integration of multiple cluster copies into the *M. xanthus* DK1622  $\Delta$ *mchA-tet* genome, e.g. into the  $\Delta$ *mchA-tet* integration site and the mx8<sup>63</sup> and mx9<sup>64</sup> attachment sites, or by using transposon-mediated genome integration<sup>41</sup> may lead to yield enhancement. The positive influence of the copy number of BGCs on the antibiotic production level was e.g. shown for heterologous pamamycin and demethoxyaranciamycinone production in *S. albus* J1074 and *S. albus* T11, respectively, resulting in a 3.5-fold production increase with three BGC copies (pamamycin) and a 7.7-fold increase with four BGC copies (demethoxyaranciamycinone).<sup>65</sup> However, a similar approach was not successful in *M. xanthus* DK1622 where two coralopyronin BGC copies were introduced into the genome without beneficial effects on the coralopyronin production.<sup>29</sup> It was hypothesized, and later proven, that the production bottleneck was the insufficient self-resistance of the heterologous strain.

The self-toxicity of potent antibacterial compounds produced in bacterial host strains is a common issue, especially in heterologous expression. Bacterial producers of antibiotics have developed certain self-protection mechanisms such as antibiotic efflux, antibiotic modification, target modification, antibiotic degradation, sequestration and target bypass.<sup>66,67</sup> The resistance-mediating proteins are often encoded within the respective BGCs that encode the protein machinery for antibiotic production. Issues with self-toxicity of the produced compounds were e.g. overcome by overexpression of those resistance genes.<sup>68</sup> In chapter 4, the heterologous production of full-length and truncated corramycins was discussed. Although the determination of the Cor1183 MIC revealed insensitivity of the heterologous producer strain *M. xanthus* DK1622 pMYC20Cor up to a concentration of 64  $\mu\text{g mL}^{-1}$ , we only detected trace amounts of Cor1183 in the crude extract. In the cystobactamid project (chapter 2) the self-toxicity towards the major product Cys919-1 was also tested in a MIC experiment (not discussed in chapter 2) according to the protocol for MIC determination in chapter 4. Here, we also observed insensitivity up to 64  $\mu\text{g mL}^{-1}$  in the heterologous producer *M. xanthus* DK1622 pMYC20Cys\_v2. At first this may lead to the interpretation that self-resistance is not the limitation and reason for low production titers, which was especially observed in case of corramycin. After deleting parts of the assembly line, truncated corramycins, such as Cor666, were produced in sufficient amounts for quantification implying that either self-toxicity is reduced or the hypothesized pre-drug mechanism (see chapter 3) is somehow influenced. One can also question if the MIC determination experiment is appropriate to test the self-toxicity limitation of strains to produce a certain compound, because the compound has to be taken up by the bacterium in sufficient amounts to achieve intracellular concentrations similar to those achieved by the biosynthesis. The solution would be measuring the intracellular concentration, e.g. by attaching a fluorophore to the antibiotic;<sup>69</sup> however, it is not unlikely that the uptake of the antibiotic is influenced in a negative way, especially when the fluorophore-attached molecule exceeds 600 Da, which is the cutoff for compounds to penetrate the outer membrane of Gram-negative bacteria.<sup>70</sup> Even if the antibiotic is actively imported via a membrane transporter as it is the case for corramycin, which is clearly exceeding a size of 600 Da, an attached fluorophore might prevent the active transport process. Recently, a LC-MS/MS method was evaluated to measure intracellular accumulation of compounds within bacteria.<sup>71</sup> In this study, the authors exposed *E. coli* and efflux-deficient *E. coli*  $\Delta\text{tolC}$  to different antibiotics and compared the shift in MIC with their LC-MS/MS analysis obtained after extraction of the bacterial lysates.<sup>71</sup> As the different amounts of compound detected by LC-MS/MS correlated with the shift in MIC, this method might be an option to measure intracellular antibiotic concentration. However, in case of corramycin even the precise determination of the intracellular Cor1183 concentration might not be

meaningful because the hypothesized pre-drug mechanism (see chapter 3) may lead to the presence of intracellular acylated corramycin precursors, which potentially exhibit a different self-toxicity in *M. xanthus* DK1622 compared to Cor1183.

Another common method to improve the expression of BGCs, potentially leading to increased production titers, is the use of various heterologous promoter systems. In general, promoters regulate the first stage of gene expression. The use of native promoter systems harbor the danger of natural regulatory networks influencing the gene expression in a negative way, whereas the use of synthetic regulatory elements that are not depending on cellular control mechanisms allow to bypass it.<sup>72</sup> Promoters are commonly subdivided into two groups: The constitutive promoters, which exhibit a well-defined and usually constant transcriptional activity, and the inducible promoters, which can be used to turn on the gene expression at a certain time point under the addition of a specific inducer molecule. However, the level of gene expression of certain constitutive promoters can vary, e.g. depending on growth conditions and stage, and the addition of inducers using inducible promoter systems may have (negative) impact on the global gene expression level.<sup>72</sup> Notably, the level of gene expression is not necessarily proportional with the NP production yield. Several studies have shown that an optimal enzyme expression for yield maximization lies between very low and very high expression.<sup>72-75</sup> Consequently, the choice of an appropriate promoter system is thus of vital importance for the outcome in heterologous expression experiments.

For *M. xanthus* DK1622 so far a small number of different promoters has been evaluated for the heterologous expression of NP BGCs, such as the constitutive  $P_{\text{nptII}}$ , the  $P_{\text{tet}}$ , the  $P_{\text{m}}$  and the inducible  $P_{\text{van}}$  and  $P_{\text{IPTG}}$  promoter systems (Table 1). However, there are only a few studies on systematic comparison of the different promoter systems and their impact on the production level of certain NPs. For instance, the exchange of the constitutive  $P_{\text{nptII}}$  promoter by the inducible  $P_{\text{van}}$  promoter lead to a twofold increase in corallopyronin A production.<sup>33</sup> Another more extensive study compared certain available promoter sequences using the same 5' UTRs (untranslated region) and their impact on argyirin production level, showing that the  $P_{\text{nptII}}$  and  $P_{\text{van}}$  are superior to other tested promoters.<sup>29</sup> For the heterologous production of cystobactamids and corramycins (see chapters 2 and 4) we used the inducible  $P_{\text{van}}$  promoter, because of its superiority in the previously mentioned studies and because we wanted to ensure the heterologous producer strains to reach high cell densities before inducing the gene expression and subsequent production of potentially toxic compounds.

Summarized, the heterologous expression is a versatile tool enabling to achieve high production titers of the respective compounds, to isolate novel compounds by expressing cryptic BGCs or to investigate the complex biosynthetic routes of NPs. However, compared e.g. to actinobacterial heterologous expression systems, which provide numerous options for modifications, myxobacterial

heterologous expression systems are still in their infancy with a lot of room for development. The following sections discuss the possibilities that synthetic biology approaches in combination with genetic engineering of NP biosynthetic pathways and heterologous expression open up.

#### 5.4 Heterologous expression in the field of synthetic biology

By broad definition, synthetic biology is a multidisciplinary research area focusing on the engineering of man-made lifeforms from standardized building blocks that can perform a predefined function.<sup>76</sup> Two general approaches in the field of synthetic biology are discussed: First, the “top-down” approach, in which existing natural systems are modified towards reduced but more specialized systems that are e.g. capable of producing drugs, biofuels and fine chemicals. Second, the “bottom-up” approach, in which novel biological systems are designed from scratch, independently from existing systems, being able to synthesize chemical entities that are not found in nature. The second approach is a more knowledge-driven approach, not infrequently with the goal to design minimal biological systems.<sup>76</sup> The advances in the field of synthetic biology, e.g. in *de novo* synthesis of DNA<sup>77</sup> and the development of bioinformatics tools (discussed in chapter 1), even go so far that entire synthetic bacterial genomes are not an utopian dream anymore.<sup>59</sup> Furthermore, today nearly unlimited options to modify BGCs *in silico* for the heterologous expression exist.

In the cystobactamid (chapter 2) and corrAMYCIN (chapter 4) projects, the design of the heterologous expression constructs followed the top-down approach, in which only a minimal amount of modifications was made to leave the BGCs as native as possible. As the native producer strains of the compound classes (*Cystobacter* sp., *Myxococcus* sp. and *Coralloccoccus* sp.) are phylogenetically closely related with the heterologous host *M. xanthus* DK1622, we speculated that essential requirements such as the ability for functional expression, proper translation and post-translational modifications of the biosynthetic enzymes were given. Since there is only very little knowledge about gene expression and the necessary regulatory systems in myxobacteria at present, we assessed a *de novo* design of a BGC in a bottom-up approach too likely to fail. To uncouple the expression of the BGCs from their native regulatory constraints, and because we had literally no information about the native promoter systems from the BGCs, we introduced the P<sub>van</sub> promoter (discussed in the previous section) and tried to force the gene expression of the entire BGCs under its control. Therefore, we had to re-organize the structure of the transcriptional units to a certain extent, because we wanted to keep the number of promoters limited ( $\leq 2$ ) as they could serve as unwanted interfering repetitive sequence segments during BGC assembly. Furthermore, we did not completely reduce the number of operons to only one, because the native clusters also harbor several operons and we tried to avoid the production of too large transcripts, which may not be stable or lead to an inefficient

translation process. Another important modification we made was the removal of several R-sites by synonymous codon substitution. The removal of *BsaI* R-sites enabled the assembly of the BGCs parts harboring repetitive sequence segments or the deletion of (part of) genes using sequence homology-independent techniques, such as conventional restriction hydrolysis/ligation, without leftover sequences that could disrupt protein functionality. Consequently, this modification was essential in both projects, not only for BGC assembly but also as basis for the manipulation of the pathways resulting in the production of new derivatives and providing insights into underlying biosynthetic processes.

In the time when *de novo* DNA synthesis was still in early development, capturing and cloning of entire secondary metabolite pathways for heterologous expression relied on certain techniques that were discussed in chapter 1. Recently, an attempt to heterologously express the corramycin BGC after PCR-based cluster fragment amplification and assembly via TAR failed, because unwanted homologous recombination events lead to the deletion of parts of the assembly line.<sup>1</sup> Thus, no production of corramycin or related shunt-products was achieved. However, not only assembly of complex BGCs harboring repetitive sequence segments, but also their refactoring, e.g. regarding operon structure, or manipulation of the pathway, e.g. scarless gene deletions or removal of R-sites, is much more difficult compared to the opportunities synthetic biology approaches offer. However, despite all the advantages, major disadvantages of *de novo* DNA synthesis with subsequent BGC assembly for the heterologous expression is that it is time consuming, expensive (compared to classical techniques) and the success rate heavily relies on the quality of the sequencing data. For instance, in the cystobactamid project (chapter 2) an entire NRPS-encoding gene had to be re-synthesized because the template sequence was wrong due to sequence read miss-assembly, finally delaying the progress in BGC assembly. Furthermore, also the *de novo synthesis* of entire BGCs is still challenging and is often complemented by traditional cloning techniques, a problem that was already addressed in the early 2000s.<sup>78</sup> Consequently, instead of directly spending the resources in a synthetic biology approach, the use of classical approaches has to be considered beforehand. E.g. if a NP BGC is not of the highest complexity regarding size and repetitive sequence segments, direct cloning techniques and evaluation of a heterologous host strain may be used to gather necessary information as basis for the design of a modified and more flexible synthetic BGC.

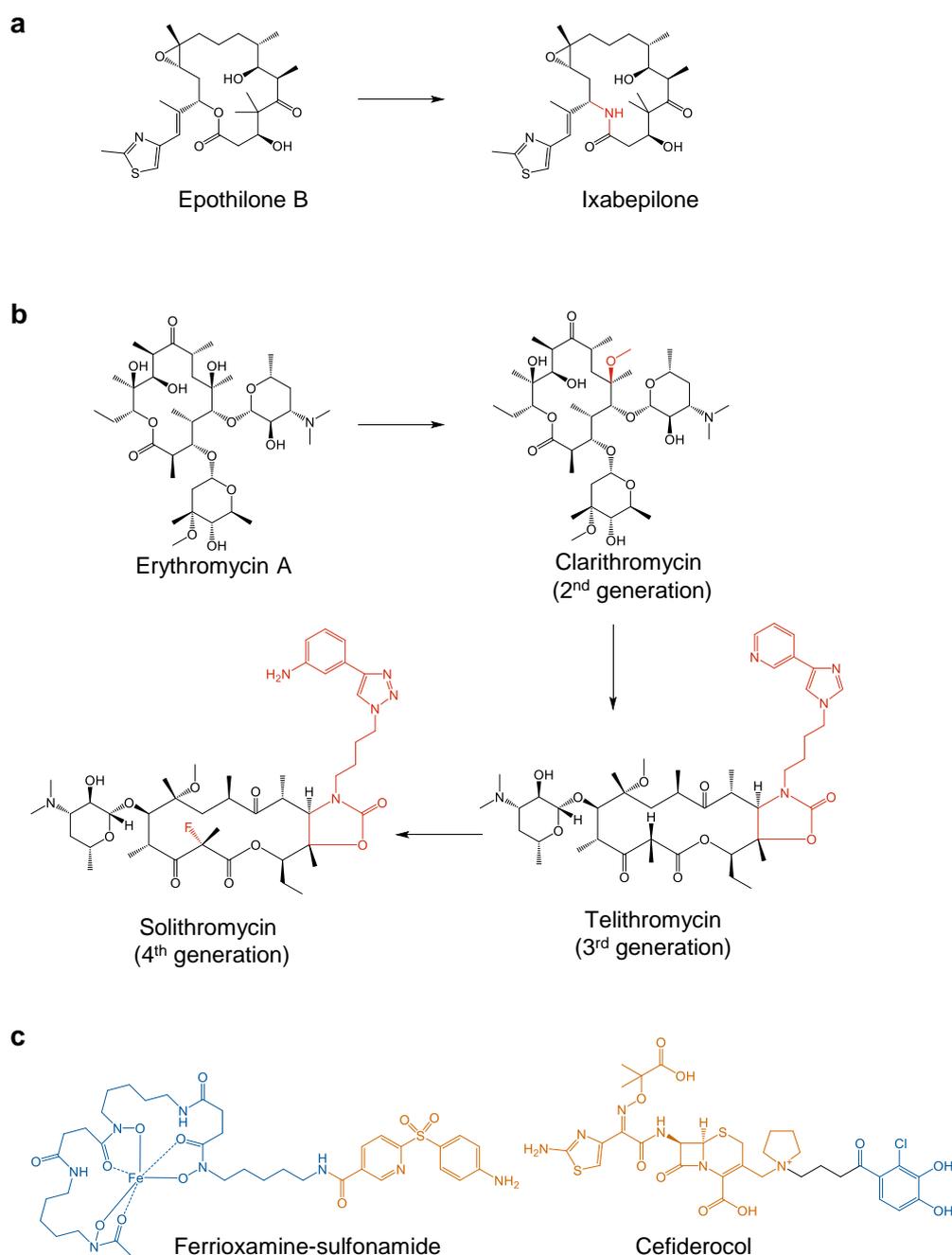
## 5.5 Structure engineering of bacterial NPs

In the period from 1981 to 2019 only 7 % of the antibacterial drugs approved by the FDA were unaltered NPs, whereas 48 % were NP derivatives.<sup>11</sup> The reason why only such a minor percentage of unaltered NPs reached the market is that NPs, despite their potent bioactivities, often lack the pharmaceutical properties required for medical applications, such as solubility, low toxicity and

metabolizability. Furthermore, to overcome rising AMR, the pharmaceutical industry is forced to constantly develop new antibiotics. Consequently, various methods were developed to modify the structure of NP antibiotics, including supplementation of certain building blocks during fermentation in precursor-directed biosynthesis and mutasynthesis, genetic engineering of the producer strains, semi-synthetic methods, combinations from those methods and even establishment of total chemical synthesis routes (see chapter 1). The development of the latter is often challenging and inefficient in the context of NPs, because typically many synthesis steps are required finally resulting in low overall yields. The reason for this lies within the complex chemical structure of NPs. Sometimes only one complex building block is sufficient to limit a whole total synthesis approach. One option to tackle this problem is the semi-synthesis, in which a natural product that was produced by fermentation is chemically modified after its purification. Even if the possible options for structure modification are limited compared to the options total synthesis offers, the structural modifications are not restricted to building blocks that are accepted by the biosynthetic machineries of the producer strains.

### 5.5.1 Semi-synthesis for structure diversification of NPs

One prominent example of a semi-synthetically optimized myxobacterial compound that was approved by the FDA in 2007 as anticancer drug is ixabepilone.<sup>79</sup> It is synthesized by chemically substituting a lactone with a lactam in the natural derivative epothilone B, thus making the generated derivative more stable compared to the natural compound (Figure 3a). Another example are semi-synthetically produced erythromycin analogs to overcome acid lability and macrolide-resistance in certain bacterial pathogens (Figure 3b). The selective *O*-methylation at one position in erythromycin lead to the development the analog clarithromycin, which is acid-stable and exhibits an improved antibacterial spectrum.<sup>80,81</sup> Later, the introduction of alkyl-aryl side chains and the removal of the cladinose sugar moiety with subsequent introduction of a keto group resulted in a new subclass and the 3<sup>rd</sup> generation of macrolides, the ketolides.<sup>80</sup> One member of this new subclass, telithromycin, showed 4-8-fold greater potency, activity against macrolide-resistant strains and finally received market authorization; however, it is no longer widely used due to adverse side effects.<sup>80,81</sup> In a 4<sup>th</sup> generation of macrolides, the fluoroketolides, the imidazolyl-pyridine side chain was removed, because it was hypothesized to cause the previously mentioned side effects. Furthermore, the first member of this class, solithromycin, contains a fluorine on the macrocyclic ring which was shown to prevent the group from enolizing as seen with telithromycin.<sup>80,82</sup> The generation of semi-synthetic analogs has also been performed in other antibiotic classes such as the  $\beta$ -lactams, cephalosporins, glycopeptides and aminoglycosides.<sup>83</sup>



**Figure 3 | NP structure diversification using semi-synthesis. a:** The semi-synthetic modification (shown in red) of epothilone B lead to the generation of the marketed anticancer drug ixabepilone. **b:** Erythromycin macrolide antibiotics from first generation to fourth generation via semi-synthesis. Structural modifications compared to the previous generation are shown in red. **c:** Two examples of semi-synthetically produced siderophore-conjugated antibiotics. The warhead is shown in orange and the siderophore part is shown in blue.

One proven approach to tackle the issue with permeability-related resistance in Gram-negative bacteria is the conjugation of antibiotics to iron-chelators.<sup>84</sup> Since the human immune response maintains an iron poor environment and iron availability is crucial for the virulence of the pathogenic bacteria, up-regulation of the iron uptake systems in the bacterium is the consequence, finally promoting the uptake of siderophore-conjugated antibiotics.<sup>85</sup> Therefore, naturally occurring microbial siderophores such as hydroxamic acids, catechols, hydroxyl carboxylic acids or

*ortho*-hydroxy phenyl-substituted oxazoles are used to exploit iron uptake pathways for antibiotic delivery through the bacterial cell envelope. This strategy was inspired by natural siderophore-antibiotic conjugates, the sideromycins,<sup>86</sup> and is known as the “Trojan Horse” approach.<sup>84</sup> There are numerous decisions to make prior to (semi-)synthesis of those conjugates, e.g. which type of warhead (antibiotic) has to be combined with which type of siderophore and also which type of linker between those components will be used, depending on whether the warhead needs to be released after uptake or not. An early example of a semi-synthetically produced siderophore-antibiotic is a ferrioxamine-sulfonamide and was already reported in 1977 (Figure 3c).<sup>87</sup> Since the chemical synthesis of siderophores from the scratch is often impractical, most semi-synthetic approaches start from siderophores isolated by fermentation with subsequent conjugation of the antibiotic warhead by chemical synthesis.<sup>85</sup> In contrast to many other antibiotics,  $\beta$ -lactams tolerate extensive peripheral substitution and are thus a common choice for the design of sideromycins. For instance, the semi-synthesis of pyoverdine-ampicillin conjugates resulted in a compound that is selectively active against *Pseudomonas*, because *Pseudomonas* naturally utilizes pyoverdine for iron sequestration.<sup>85,88</sup> On the contrary, a recent example of a cephalosporin-catechol, cefiderocol (Figure 3c), showed broad-spectrum activity primarily against multi-drug resistant Gram-negative pathogens and is now in clinical phase III trials.<sup>84,89</sup> Furthermore, the conjugation of siderophores to antibiotics, which are naturally solely active against Gram-positive bacteria, resulted in an extended activity spectrum including Gram-negative strains. For example, the large size of daptomycin is disadvantageous for the penetration of the outer membrane of Gram-negative bacteria and daptomycin is thus only used to treat infections caused by Gram-positive pathogens; however, recently reported siderophore conjugates showed remarkable potency in the low  $\mu\text{M}$  range against multi-drug resistant *A. baumannii* strains *in vitro* and *in vivo*.<sup>85,90,91</sup> However, although there is great potential to modify the antibacterial activity spectrum in terms of potency and selectivity by conjugating siderophores to antibiotics, no such compound has yet been approved for clinical use.<sup>84</sup> The major reason for this lies within the rapidly developing bacterial resistance towards the siderophore-antibiotics, which was mainly observed in *in vitro* experiments.<sup>85</sup> The resistance mechanism was often the deletion of siderophore receptors that are specific for a certain siderophore. As bacteria possess several of those uptake systems, they can afford to inactivate one of them. However, despite a high FoR observed *in vitro*, numerous siderophore-antibiotics showed promising *in vivo* activity with a lower FoR in animal models or in humans, e.g. in the case of albomycins, salmycins and ferrimycins.<sup>85,92–96</sup> The authors explain this incongruence with different conditions in *in vitro* laboratory assays compared to the situation *in vivo* where the loss of one essential iron uptake system might not be compensated by up-regulation of another system. The reserve system, if available, taking over the iron uptake may be less efficient at the site of infection, thus giving the immune system precious time to react. Notably, one solution to circumvent fast resistance

development was the combination of more than one iron-chelating moiety attached to an antibiotic warhead. A so-called mixed ligand conjugate combining bis-catechol and mono-hydroxamate siderophores attached to carbacephalosporin minimized the development of resistance by using multiple transport processes in *E. coli*.<sup>85,97</sup> Reduced resistance development was even observed in *E. coli* strains, which were defective in one type of siderophore uptake system and showed resistance towards bis-catechol conjugates or mono-hydroxamate conjugates.

In the corramycin project, a high FoR was observed *in vitro* when using MHB medium, because the uptake of Cor1183 was only mediated by one transporter, SbmA (chapter 3). The use of M9 minimal medium, which reflects an *in vivo* infection in a more realistic way, resulted in a substantially lower FoR. This result is congruent with the results observed when using mixed ligand conjugates, which also showed reduced resistance development when uptake relied on multiple transport systems. Furthermore, native corramycin exhibited a rather narrow activity spectrum, which makes it an interesting candidate for structure engineering. It was shown that conjugation with an iron-chelating moiety via semi-synthesis lead to a broadened activity spectrum (chapter 4; unpublished data; personal communication with Stephane Renard). The basis for a semi-synthetic approach was set by producing truncated derivatives, which can be introduced into an existing but inefficient total synthesis route (chapter 4). This was achieved by engineering the NRPS-PKS assembly line and thus the targeted approach for corramycin structure diversification is a combination of production by fermentation using a heterologous strain with an engineered biosynthesis pathway and chemical synthesis.

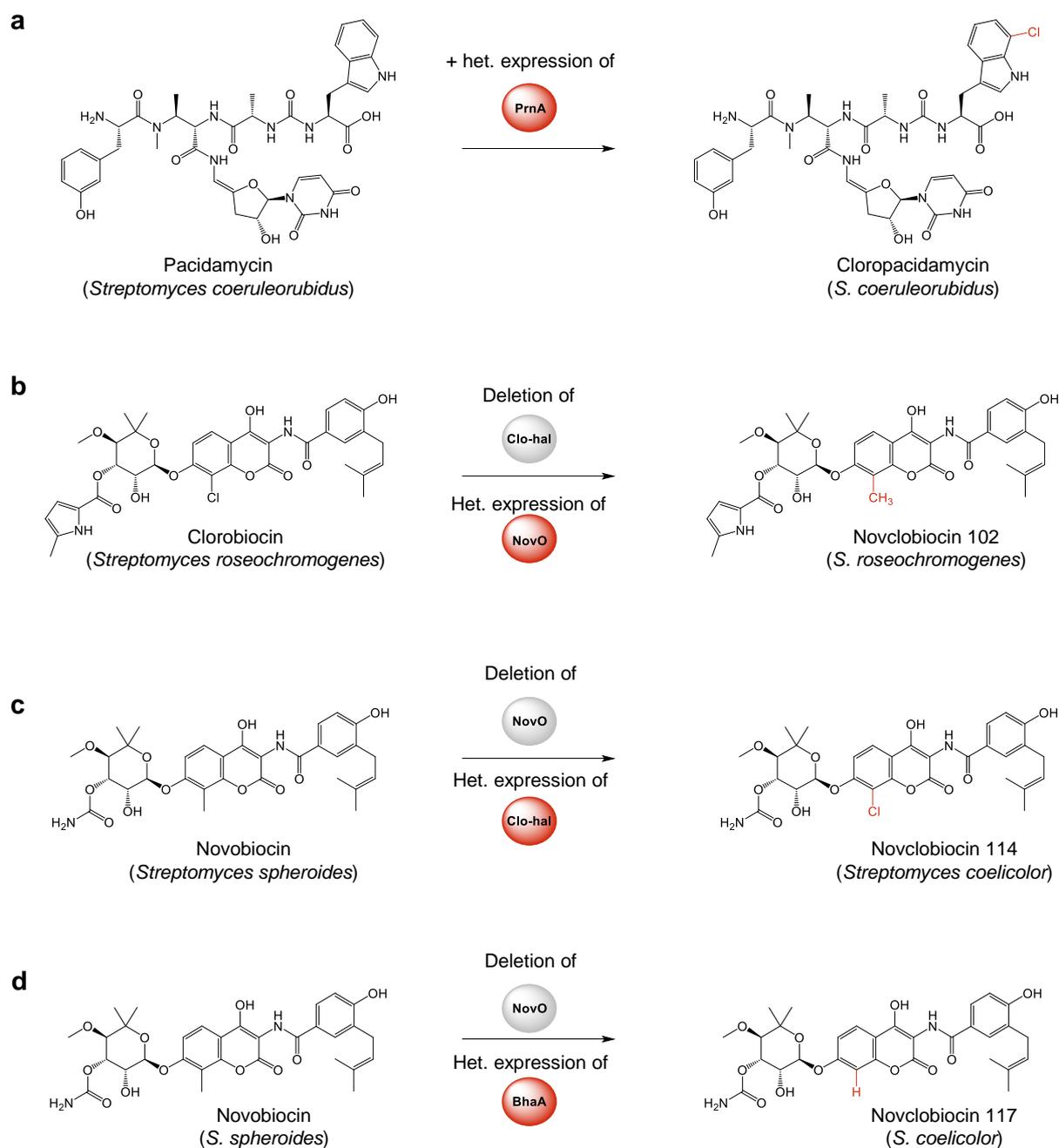
### 5.5.2 Engineering of tailoring enzymes

Tailoring enzymes play an important role in the biosynthesis of NPs as they lead to greater structural diversity. The different types of tailoring enzymes and domains involved in the biosynthesis of NRPs are described in chapter 1. The structural diversity influences the bioactivity and pharmaceutical properties within a compound class. Native producers often produce a whole cocktail of derivatives, which all together help a certain strain to defend against a variety of competitors. However, from a pharmaceutical point of view the partitioned production of multiple derivatives results in decreased production of a specific analog of interest. Thus, the demand to redirect the production profile towards a specific analog or to modify a certain compound derivative for improved pharmaceutical properties made tailoring enzymes good candidates for engineering approaches.

There are several ways tailoring enzymes can take influence on the final structure of the NP. Some independent tailoring enzymes modify the simple biosynthetic precursors that are later utilized by the assembly line. An advantage of those enzymes in terms of structure engineering is that they do not have to be specific for the final compound produced, but only specific for the precursor molecule.

Furthermore, no or fewer protein-protein interactions between the precursor tailoring enzymes and the assembly line enzymes have to be considered. This allows introduction of such tailoring enzymes into unrelated biosynthetic pathways leading to a broader precursor substrate pool. However, the major disadvantage of this approach is that the biosynthetic machinery has to accept the modified precursor prior to incorporation. One example is the expression of the tryptophane-7-halogenase PrnA from *Pseudomonas fluorescens* Pf-5, natively involved in pyrrolnitril biosynthesis, in *Streptomyces coeruleorubidus* aiming for the production of halogenated pacidamycin (Figure 4a).<sup>98</sup> Indeed, halogenated tryptophane was accepted and incorporated by the pacidamycin assembly line; however, the new derivative was only a minor derivative alongside the native compound, clearly showing the limitation of this approach. In the cystobactamid project the precursor tailoring enzyme encoding gene *cysC* was deleted leading to complete abolishment of cystobactamid production (chapter 2). CysC catalyzes the hydroxylation of the precursor building block *pABA* prior to further tailoring steps involving an *O*- and iterative *C*-methylations. However, the hydroxylation step seems to be critical, because the deletion of *cysC* from the BGC lead to complete abolishment of the cystobactamid production, showing how sensitive the biosynthetic machineries can be when it comes to changes in the availability of precursors. In future experiments it has to be investigated whether the supplementation of 3-hydroxy-*pABA* and 2,3-dihydroxy-*pABA* during cultivation of the *cysC* deletion strain may lead to complementation of the production. In the case of a successful complementation, a mutasynthesis approach with feeding of *pABA* derivatives that naturally do not occur in cystobactamid biosynthesis would be a follow-up experiment, potentially leading to the production of novel cystobactamids.

Other tailoring enzymes that can be addressed by structure engineering can act independently, *in trans*, during or after biosynthesis of the NP, which is attached to the assembly line or released from it, respectively. As those enzymes are usually much more specific towards the compound class they natively modify, the success of heterologous expression of the respective gene in different hosts for the modification of other compound classes is limited by the enzymes substrate specificity. Nevertheless, there are some examples of the heterologous expression of tailoring enzymes from similar metabolic pathways, which lead to successful modification of natural analogs. For example, clorobiocin and novobiocin are potent inhibitors of the bacterial DNA gyrase which are structurally very similar; however, among further structural differences, clorobiocin contains a chlorine at position 8 of its aminocoumarin group whereas novobiocin harbors a methyl group at this position.<sup>99</sup> Inactivation of the gene *clo-hal* encoding the halogenase in the clorobiocin BGC of *Streptomyces roseochromogenes* var. *oscitans* and subsequent heterologous expression of *novO*, encoding the *C*-methyltransferase from the novobiocin BGC (*Streptomyces spheroides*), resulted in the production of the new hybrid compound novclorobiocin 102.

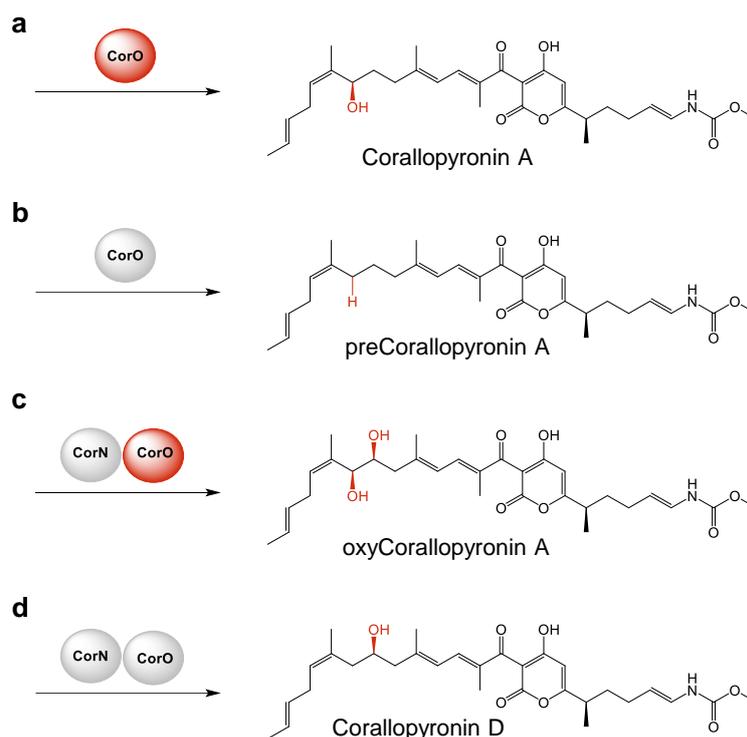


**Figure 4 | NP structure modification via tailoring enzyme engineering.** Deleted enzymes are shown in grey spheres, heterologously expressed enzymes are shown in red spheres and structural differences of the modified compound are highlighted in red. **a:** Modification of precursor building blocks. Heterologous expression of the tryptophane-7-halogenase PrnA from *Pseudomonas fluorescens* Pf-5 in *Streptomyces coeruleorubidus* resulted in the production of cloropacidamycin. **b:** Structure engineering using compound-specific *trans*-acting tailoring enzymes. Deletion of the halogenase Clo-hal from clorobiocin biosynthesis and heterologous expression of the methyltransferase NovO from novobiocin biosynthesis in *Streptomyces roseochromogenes* leads to the production of novclobiocin 102. **c:** Deletion of NovO and heterologous expression of Clo-hal in *Streptomyces spheroides* results in novclobiocin 114 production. **d:** Deletion of *novO* from the novobiocin BGC and heterologous expression of the BGC with the halogenase BhaA in *Streptomyces coelicolor* yielded novclobiocin 117, which was also observed in the NovO deletion mutant without BhaA.

Structurally it differs from clorobiocin by having a methyl instead of a chlorine group at position 8 of the aminocoumarin moiety (Figure 4b).<sup>99</sup> Moreover, the heterologous coexpression of the entire novobiocin BGC and Clo-hal in *Streptomyces coelicolor* lead to the production of novclobiocin 114, a novobiocin analog harboring a chlorine group instead of a methyl group at position 8 of the aminocoumarin moiety (Figure 4c).<sup>100</sup> However, an attempt to functionally replace Clo-hal by BhaA, a halogenase involved in the biosynthesis of the glycopeptide antibiotic balhimycin, did not result in the production of chlorinated novobiocin analogs in *S. coelicolor* (Figure 4d).<sup>100</sup>

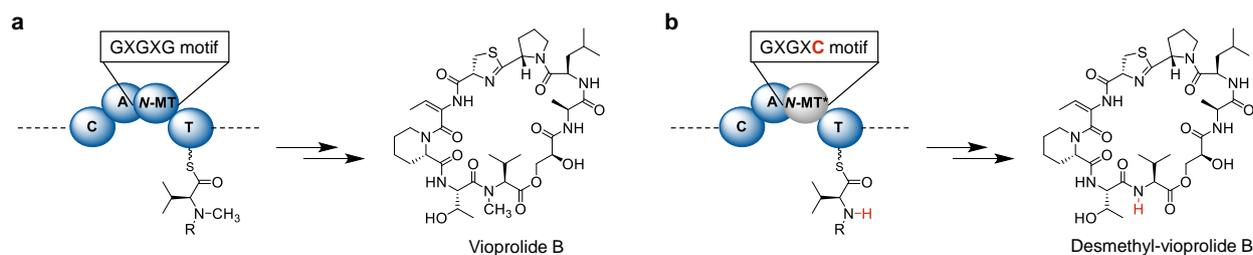
Another example is the heterologous expression of *ram29*, a gene encoding a transmembrane-spanning mannosyl transferase from the ramoplanin BGC,<sup>101</sup> in *Streptomyces fungicidicus*, the producer of the lipopeptide enduracidin,<sup>102</sup> which is closely related to ramoplanin.<sup>103–105</sup> Among other structural differences, ramoplanin is di-mannosylated whereas enduracidin exhibits no mannosylation. The mannosyl substituents in ramoplanin were shown to enhance aqueous solubility, which was a critical property in the development as a potential drug for *Clostridium difficile* infections.<sup>106</sup> Enduracidin, however, shows only poor aqueous solubility preventing its development as a drug for human infections. Initially, the expression of *ram29* did not result in mannosylation of enduracidin, but after exchanging the native Shine-Dalgarno and the GTG start codon mannosylated enduracidin was produced. Surprisingly, enduracidin was only mannosylated once, not twice.<sup>105</sup> The authors hypothesized that either another enzyme, which is not encoded within the ramoplanin BGC, catalyzes the second mannosylation or that *S. fungicidicus* expresses an  $\alpha$ -mannosidase that removes the second mannosyl group.<sup>105</sup> Another explanation could be that the substrate specificity of Ram29 is too narrow to catalyze a di-mannosylation, since the mono-mannosylated enduracidin derivative was only a minor product alongside the native enduracidin. Those examples underline that the substrate specificities of tailoring enzymes acting *in trans* may limit their broad applicability in structure engineering.

The deletion of *trans*-acting tailoring enzymes is another method used for structure engineering. For instance, the independent or combined deletions of *corO*, encoding a cytochrome P450, and *corN*, encoding an ECH-like enzyme, from the coralopyronin BGC resulted in heterologous production of the non-hydroxylated pre-coralopyronin A and two new derivatives, coralopyrinin D and oxyCoralopyrinin A, bearing a hydroxyl group at a different position in *M. xanthus*, respectively (Figure 5a-c).<sup>33</sup> The simultaneous deletion of *corO* and *corN* resulted in the exclusive production of coralopyronin D (Figure 5d). Similarly, endogenous tailoring enzymes were deleted from the cystobactamid BGC, which lead to the heterologous production of novel cystobactamids (see chapter 2). The independent deletions of *cysQ*, encoding an *O*-methyltransferase, and *cysJ*, encoding a hydroxylase, yielded the production of cystobactamids with desmethylated and desmethoxylated linker moieties, respectively.



**Figure 5 | NP structure engineering by deletion of tailoring enzymes.** Deleted enzymes are shown in grey spheres, active enzymes are shown in red spheres and positions in the structures that are affected by the (deleted) enzymes are highlighted in red. **a:** Final hydroxylation step in the coralloyronin biosynthesis mediated by CorO. **b:** Deletion of CorO prevents formation of coralloyronin A. **c:** Deletion of CorN leads to a different hydroxylation pattern and the production of coralloyronin D and oxyCoralloyronin A. **d:** Deletion of CorN and CorO resulted in the production of coralloyronin D but not oxyCoralloyronin A.

There are only a few examples of deletions or inactivations of *cis*-acting NRPS tailoring domains that lead to the production of novel NP derivatives. For example, the daptomycin-related antibiotic A54145 has an *N*-methylated glycine residue at amino acid position 5 and the respective NRPS module has a C-A-MT-T architecture. The deletion of the MT domain yielded a functional C-A-T module that incorporated glycine instead of *N*-methylated glycine, leading to new derivatives with improved MIC against *S. aureus*.<sup>107</sup> Another example, in which an *cis*-acting NRPS tailoring domain was inactivated by rational mutation of conserved residues, is the inactivation of the *N*-MT domain of module 10 in the vioprolide biosynthesis (Figure 6).<sup>50</sup> The mutation of the conserved GXGXXG motif to GXGXC finally resulted in the production of *N*-desmethyl vioprolides.



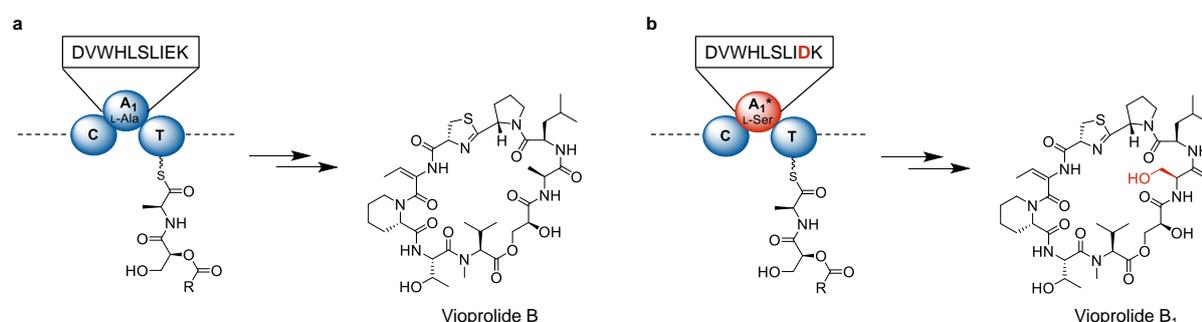
**Figure 6 | Inactivation of an *in cis* NRPS tailoring domain for NP structure diversification.** **a:** Biosynthesis of vioprolide B by the native vioprolide NRPS megasynthetase with an active *N*-MT domain. **b:** Targeted mutation of the native GXXGX motif to a GXXGX motif in the *N*-MT domain (shown in red) yielded desmethylated vioprolides, exemplified by desmethyl-vioprolide B. Up- and downstream modules of the vioprolide megasynthetase are not shown (indicated by dashed line) for reasons of simplicity. The structural difference between vioprolide B and the desmethyl-analog is highlighted in red.

In this thesis another example of a successful domain deletion was given. The newly characterized bifunctional AMDH domain was deleted independently or in combination with CysJ to achieve redirection of the production profile towards one series of cystobactamids with L-asparagine instead of L-isoasparagine linkers and to produce novel derivatives, respectively (see chapter 2). One critical point when deleting NRPS tailoring domains are the engineering sites, because choosing suboptimal engineering sites may negatively affect folding of the entire protein or influence protein-protein interactions. This is especially important for tailoring domains that are embedded into other NRPS domains such as MT domains or the AMDH domain, which are integrated into A domains. When we deleted the AMDH domain from the stand-alone NRPS module CysH, we had no information about the 3D structure of the protein, because previous attempts to crystalize CysH failed and no known structure template was available in public databases. The engineering sites we used were based only on sequence alignments of the A domain (and AMDH domain) in CysH with other A domains, which clearly showed the insertion borders of the AMDH domain. However, the deletion of the AMDH domain also lead to disruption of potential protein-protein interactions, because the *O*-methyltransferase CysQ was not functional in the AMDH deletion strain. In future experiments, targeted point mutations in certain conserved catalytic residues in the AMDH domain may lead to functional inactivation but maintained protein-protein interactions.

### 5.5.3 Engineering of NRPS/PKS megasynthetases

As an alternative and more complex approach to engineering of tailoring enzymes, direct engineering of NRPS/PKS assembly lines can be used to increase the structural diversity of NPs. With this approach more profound changes, e.g. in the core structure of a NRP, are possible as compared to engineering of tailoring enzymes. Several possibilities to engineer NRPS/PKSs were described, such as active site modifications of A domains or deletions, insertions and exchanges of NRPS subunits, domains and modules.<sup>104</sup>

As discussed in chapter 1, A domains were used for numerous manipulation attempts, e.g. to change their substrate specificity or to narrow down the substrate tolerance in case they were promiscuous.<sup>50,108–111</sup> One advantage of A domain engineering is that the introduction of single or few point mutations can lead to an altered substrate specificity but does not affect the overall domain structure and thus also not interactions with adjacent domains or modules. Furthermore, this method is independent from domain and module linker regions. For example, in the myxobacterial vioprolide NRPS megasynthetase, the first A domain activates alanine; however, the specificity-conferring code of the A domain<sup>112</sup> is highly similar to A domains activating serine.<sup>50</sup> The exchange of a single residue lead to the incorporation of serine and the production of new vioprolide derivatives, one of which showed a 650-fold decreased cytotoxicity (Figure 7).<sup>50</sup> Interestingly, a broad tolerance for alanine was retained in the A domain, resulting in the production of native vioprolides harboring alanine as minor component alongside the new derivatives.

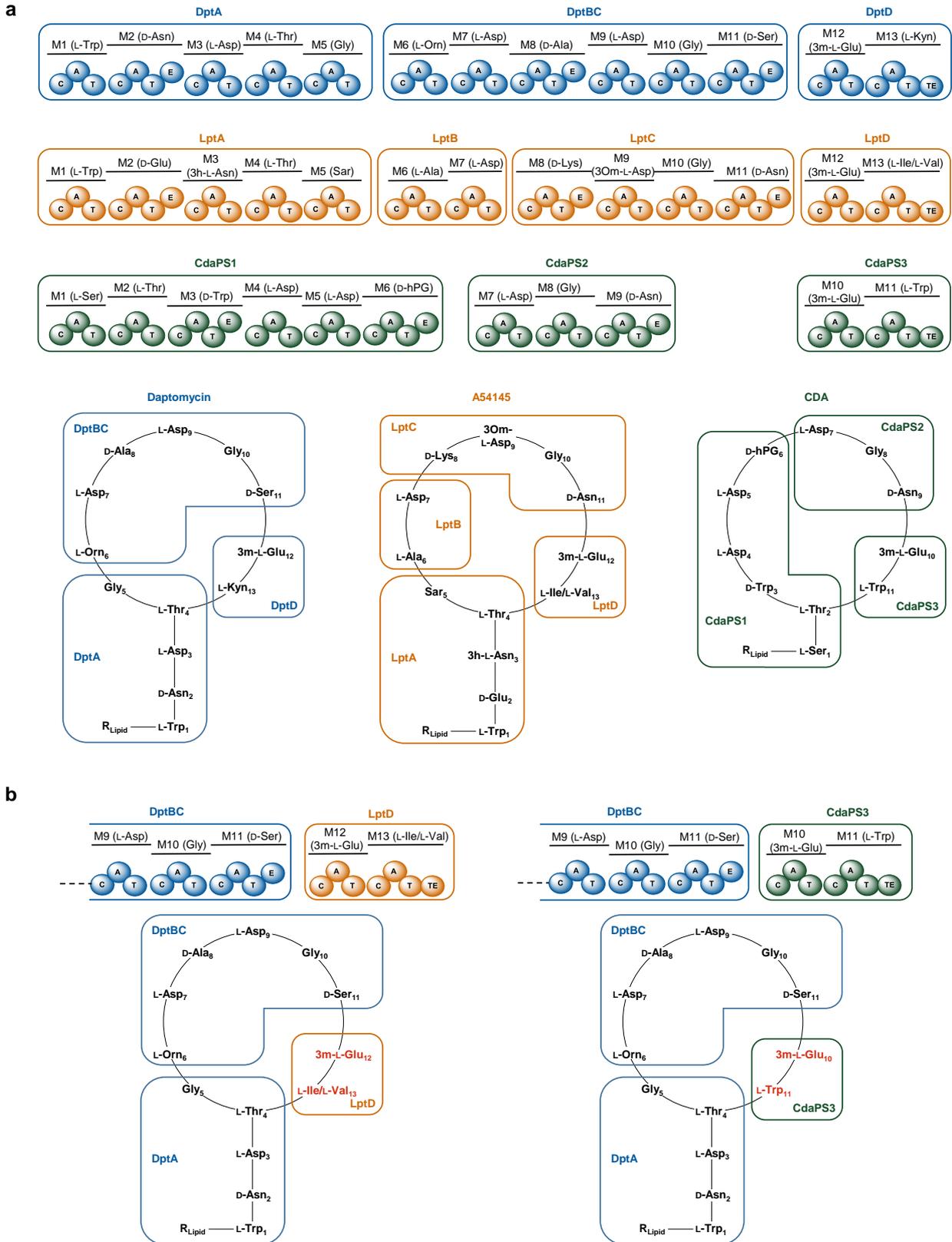


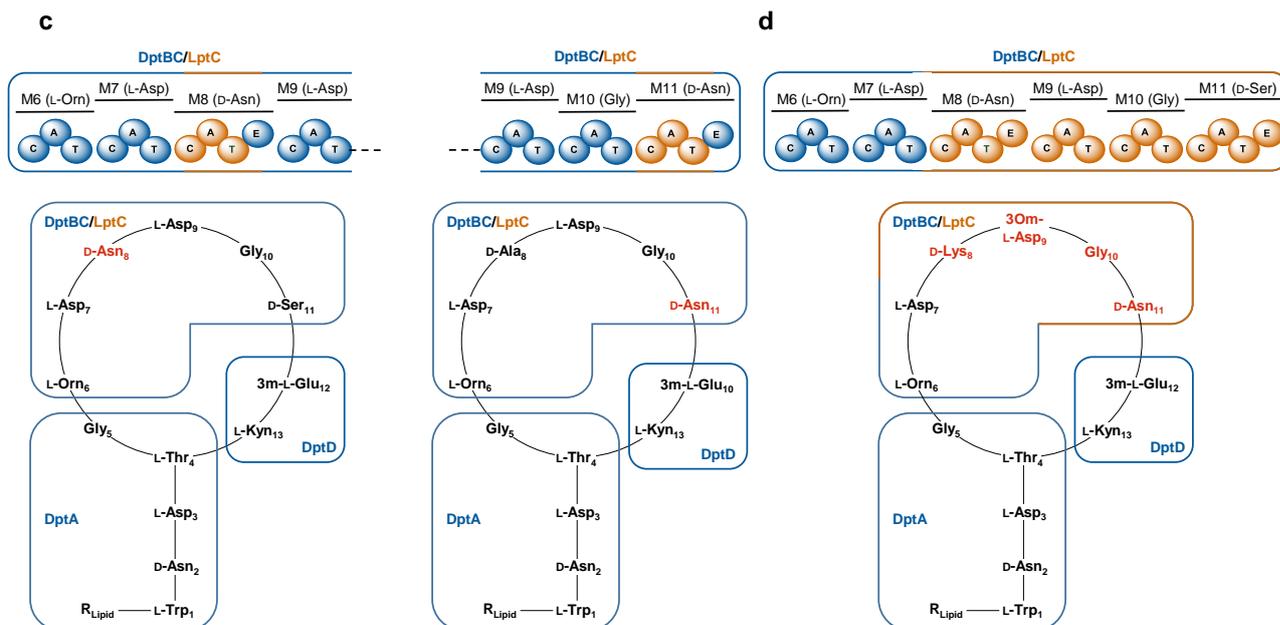
**Figure 7 | Alteration of the A domain-specificity by targeted mutation of a critical binding pocket residue.** Essential binding pocket residues by Stachelhaus *et al.*<sup>112</sup> are shown in the box. **a:** Biosynthesis of vioprolide B by the native vioprolide NRPS megasynthetase with an A<sub>1</sub> domain-specificity for L-alanine. **b:** Targeted mutation of one critical residue (shown in red) lead to the production of hydroxylated vioprolides (by the incorporation of L-serine instead of L-alanine), exemplified by desmethyl-vioprolide B<sub>1</sub>. Up- and downstream modules of the vioprolide megasynthetase are not shown (indicated by dashed line) for reasons of simplicity. The structural difference between vioprolide B and the hydroxylated analog is highlighted in red.

One disadvantage and limiting factor of this method are the substrate specificities of the downstream A and C domains. If those downstream domains do not, or only barely, accept the new substrate, further processing is hampered leading to no or very low production of a potential new product. For instance, the engineering attempt of the promiscuous A domain of the first module in the myxobacterial argyirin NRPS aimed for increasing the substrate specificity to drive the production profile towards more desirable analogs.<sup>30</sup> Although the production profile shifted towards analogs with the desired alanine residue, total argyirin production was reduced substantially. Another attempt was performed with the goal to modify the substrate specificity of the fourth module of the argyirin NRPS including the extended region of the A domain binding pocket.<sup>30,113</sup> However, this approach resulted in complete abolishment of argyirin production, clearly showing that rational A domain engineering is a challenging task. In another example, the well-described L-phenylalanine-specific

A domain of GrsA was chosen to create a library in which the eight variable specificity-conferring residues<sup>112</sup> contain single point mutations. It was shown that the mutation of one specific residue within the binding pocket changed the substrate specificity towards L-tyrosine and non-natural *para*-substituted phenylalanine derivatives like *p*-azido-L-phenylalanine and *O*-propargyl-L-tyrosine.<sup>104,114</sup>

Apart from the manipulation of A domains, the deletion, insertion and exchange of NRPS domains, modules or entire subunits is another strategy to engineer the structure of NRPs. The key initial work in this field was done at Cubist Pharmaceuticals with the development of daptomycin (Figure 8).<sup>104</sup> The exchange of entire NRPS subunits such as the terminal two modules of the daptomycin pathway, encoded by *dptD*, with the terminal modules of the highly similar A54145 and CDA pathways, encoded by *cdaPS3* and *lptD*, respectively, was already discussed in chapter 1 (Figure 8b).<sup>115</sup> After successful NRPS subunit exchanges, Cubist Pharmaceuticals aimed for the exchange of modules. The replacement of the C-A-T tridomain from the D-alanine-specific module 8 with the highly homologous C-A-T tridomain of the D-serine-specific module 11 resulted in the production of the expected daptomycin analog harboring a D-serine at position 8.<sup>104,116</sup> Both modules, 8 and 11, have a C-A-T-E architecture, but the E domains were not included in the replacement to preserve inter-module downstream associations. The replacement of the module 11 C-A-T tridomain by the module 8 C-A-T domain yielded a new analog harboring a D-alanine at position 11.<sup>104,116</sup> However, for both new analogs reduced production levels, between 15 % and 45 % of the wild-type daptomycin production level, have been observed. The independent exchange of the modules 8 and 11 C-A-T tridomains with the D-asparagine-specific C-A-T tridomain from module 11 in the A54145 pathway also lead to the production of new daptomycin analogs harboring D-asparagine at the specific position, however with further reduced production titers (Figure 8c).<sup>104,116</sup> Furthermore, it was shown that the exchange of entire C-A-T-E modules is possible, but the production titers were significantly decreased, which underlines the importance of maintaining module-module interactions. The exchange of the four modules 8 to 11 from the daptomycin pathway with modules 8 to 11 from the A54145 pathway also yielded novel daptomycin analogs, albeit with only 0.5 % production level compared to the wild-type control (Figure 8d).<sup>104,116</sup>





**Figure 8 | NRPS engineering on the example of the daptomycin, A54145 and CDA megasynthetases. a:** The daptomycin megasynthetase is shown in blue, the A54145 megasynthetase is shown in orange and the CDA megasynthetase is depicted in dark green. The corresponding structures are shown below. **b:** Terminal NRPS subunit exchanges lead to the production of new daptomycin analogs. **c:** C-A-T tridomain exchanges yielded new daptomycin derivatives. **d:** Exchange of four modules from the daptomycin megasynthetase by four modules of the A54145 megasynthetase resulted in the production of a hybrid daptomycin/A54145 analog. For reasons of simplicity only the modified parts of the entire assembly lines were shown.

Another set of experiments using only A domains or C-A didomains for exchange was performed using the pyoverdine pathway in *Pseudomonas aeruginosa* PAO1.<sup>104,117</sup> The A domain or C-A didomain of the terminal L-threonine-specific module 11 was replaced by the respective (di)domain with the same or different substrate specificity from the same BGC or from homologous pyoverdine BGCs in related *Pseudomonas* strains. Production levels of 29 % to 100 % compared to wild-type level were observed when A domain or C-A didomains which are also specific for L-threonine were introduced into the pathway. Furthermore, also the production of pyoverdine analogs with L-lysine or L-serine at position 11 was observed with 76 % and 18 % of the wild-type production level when C-A didomains were exchanged. However, many exchanges lead to complete abolishment of the pyoverdine production and the exchange of C-A didomains resulted in the production of truncated products in the majority of the cases.

As discussed in chapter 1, the insertion of a NRPS module has for example been performed in the assembly line of the glycopeptide antibiotic balhimycin.<sup>118</sup> In this case, a chimera module of modules 4 and 5, both specific for D-hydroxyphenyl glycine, was inserted between the fourth and fifth module. The chimera module consisted of the C-A of module 5 and the T-E of module 4. Except of the desired balhimycin analog harboring three instead of two D-hydroxyphenyl glycine residues, also truncated products were detected. In contrast to module insertions, also the deletion of modules

have been performed, leading to products with reduced length. For instance, the deletion of the L-leucine-specific module 2 (C-A-T architecture) in the NRPS subunit SrfA-A resulted in the production of a surfactin analog with reduced ring size.<sup>119</sup> Another more recent example is the deletion of domains and modules in the plipastatin NRPS assembly line in *B. subtilis*.<sup>120</sup> Plipastatin is an antifungal lipopeptide consisting of a cyclic 10 amino acid core with a  $\beta$ -hydroxy fatty acid chain attached. The independent deletion of the entire modules 6 and 7 lead to complete abolishment of the plipastatin production. However, the deletion of only the A domains of the respective modules or the T domain of module 7 resulted in the production of penta- and hexapeptide derivatives.

As described in the previous examples and numerous other studies, NRPS engineering attempts often result in non-functional assembly lines going along with complete abolishment of compound production or in negatively affected assembly lines leading to decreased productivity and shunt products. The main reason for this is that alterations in the NRPS domain and module structure disrupt the domain-domain and module-module interactions. This also highlights the importance of the linker regions. Two recent studies from Bozhüyük and coworkers described two new engineering sites between C and A domains as well as between C subdomains (chapter 1).<sup>121,122</sup> Those engineering sites allowed for the combination of NRPS A-T-C tridomains (exchange units – XUs) and C<sub>Dsub</sub>-A-T-C<sub>Asub</sub> units (exchange unit condensation – XUC) from different assembly lines yielding various unnatural peptides. However, all successful engineering attempts were performed with NRPS assembly lines from closely related *Photobacterium* and *Xenorhabdus* strains. The combination of XUs and XUCs from phylogenetically distant organisms failed. Thus, the broad applicability of these approaches has to be proven in future experiments.

In this thesis NRPS engineering was performed to achieve the production of truncated corrAMYCIN derivatives (chapter 4). Therefore, protein sequence alignments from C-A linkers, A-T linkers and T-C linkers as well as a protein secondary structure prediction tool were used to identify conserved regions and potential secondary structures in the linkers, respectively. Notably, in the C-A linkers of the corrAMYCIN pathway we did not find the conserved motif WNATE downstream of the helix structure in the C-A linker that was described by Bozhüyük and coworkers.<sup>121</sup> In our case, we did not identify any conserved motif downstream of the potential helix structures, but a potential helix structure was predicted for every C-A linker from modules 7 to 12. Deletion of the assembly line from the module 8 C-A linker downstream of the potential helix structure did not result in the production of truncated corrAMYCINs, which either means that the terminal C domain did not act as releasing domain or that the engineering site does not work for this specific myxobacterial case. In another attempt, we instead chose an engineering site inside the C<sub>Asub</sub> domain, because modules 9 and 12 were highly similar starting from this specific position in the C domain. Thus, we did not rely on an engineering site inside the C-A linker region. In this case, truncated corrAMYCINs were

produced showing that the domain-domain interactions were not disrupted. However, the question remains if this engineering site will find broad applicability, because both modules were specific for L-serine thus explaining their high similarity. Furthermore, we identified and used two new engineering sites in A-T and T-C linkers resulting in the production of truncated corramycins. As the A-T and T-C linkers were predicted as disordered regions with low similarity even though they originate from the same pathway, our engineering sites may not find broad applicability in other systems. In future experiments, alignments of C-A, A-T and T-C linkers from other pathways and other myxobacterial strains may help to identify conserved regions in those linkers that may serve as engineering sites with broad applicability.

## 5.6 Conclusion and Outlook

This thesis focused on two myxobacterial antibiotic compound classes, the cystobactamids and the corramycins. For both compound classes heterologous expression systems were established. Studies on the biosynthesis of cystobactamids provided insights into the biosynthesis of the unique cystobactamid linker moiety, involving a newly described bifunctional aminomutase dehydratase (AMDH) NRPS domain, and the shuttling of the linker between two NRPS subunits. Furthermore, various novel natural and unnatural cystobactamids were identified after targeted deletion of biosynthetic genes. Future experiments will focus on further characterization of the AMDH domain, e.g. by targeted mutations in conserved core motifs, and the isolation and characterization of the novel cystobactamids described herein. The studies on the AMDH domain provide new knowledge about *cis*-acting NRPS tailoring domains and may serve as basis for further investigation of the unique biochemistry. In the corramycin project, new NRPS engineering sites were identified and their application in engineering lead to the production of truncated corramycins, which was desired to yield the pharmacophore building block of corramycin for semi-synthetic optimization. The engineering sites may be applied in future experiments on other myxobacterial NRPS pathways to achieve production of new compound derivatives.

## 5.7 References

1. von Tesmar, A. Dissertation. Saarland University, 2017.
2. Payne, D. J., Gwynn, M. N., Holmes, D. J. & Pompliano, D. L. Drugs for bad bugs: confronting the challenges of antibacterial discovery. *Nat. Rev. Drug Discov.* **6**, 29–40; 10.1038/nrd2201 (2007).
3. Årdal, C. *et al.* Antibiotic development — economic, regulatory and societal challenges. *Nat. Rev. Microbiol.* **18**, 267–274; 10.1038/s41579-019-0293-3 (2020).
4. Årdal, C. *et al.* Insights into early stage of antibiotic development in small- and medium-sized enterprises: a survey of targets, costs, and durations. *J. Pharm. Policy Pract.* **11**, 1–10; 10.1186/s40545-018-0135-0 (2018).
5. Karawajczyk, A. *et al.* Expansion of chemical space for collaborative lead generation and drug discovery: the European Lead Factory Perspective. *Drug Discovery Today* **20**, 1310–1316; 10.1016/j.drudis.2015.09.009 (2015).
6. Hughes, J. P., Rees, S., Kalindjian, S. B. & Philpott, K. L. Principles of early drug discovery. *Br. J. Pharmacol.* **162**, 1239–1249; 10.1111/j.1476-5381.2010.01127.x (2011).
7. Holmes, A. H. *et al.* Understanding the mechanisms and drivers of antimicrobial resistance. *The Lancet* **387**, 176–187; 10.1016/s0140-6736(15)00473-0 (2016).
8. World Health Organization. Global Antimicrobial Surveillance System (GLASS) Report 2016-2017 (2018). Available at <https://www.who.int/glass/resources/publications/early-implementation-report/en/> (accessed 14.09.2020).
9. Nicolaou, K. C. & Rigol, S.. A brief history of antibiotics and select advances in their synthesis. *J. Antibiot.* **71**, 153–184; 10.1038/ja.2017.62 (2018).
10. Watve, M. G., Tickoo, R., Jog, M. M. & Bhole, B. D. How many antibiotics are produced by the genus *Streptomyces*? *Arch. Microbiol.* **176**, 386–390; 10.1007/s002030100345 (2001).
11. 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; 10.1021/acs.jnatprod.9b01285 (2020).
12. Imai, Y. *et al.* A new antibiotic selectively kills Gram-negative pathogens. *Nature* **576**, 459–464; 10.1038/s41586-019-1791-1 (2019).

13. 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).
14. Gans, J., Wolinsky, M. & Dunbar, J. Computational Improvements Reveal Great Bacterial Diversity and High Metal Toxicity in Soil. *Science* **309**, 1387–1390; 10.1126/science.1112665 (2005).
15. Schloss, P. D. & Handelsman, J. Toward a Census of Bacteria in Soil. *PLOS Computational Biology* **2**, e92; 10.1371/journal.pcbi.0020092 (2006).
16. Reichenbach, H. The ecology of the myxobacteria. *Environ. Microbiol.* **1**, 15–21 (1999).
17. Dawid, W. Biology and global distribution of myxobacteria in soils. *FEMS Microbiol. Rev.* **24**, 403–427 (2000).
18. Zhou, X. *et al.* Myxobacterial community is a predominant and highly diverse bacterial group in soil niches. *Environ. Microbiol. Rep.* **6**, 45–56; 10.1111/1758-2229.12107 (2014).
19. Rigali, S. *et al.* Feast or famine: the global regulator DasR links nutrient stress to antibiotic production by *Streptomyces*. *EMBO Rep.* **9**, 670–675; 10.1038/embor.2008.83 (2008).
20. Yamanaka, K. *et al.* Desferrioxamine E produced by *Streptomyces griseus* stimulates growth and development of *Streptomyces tanashiensis*. *Microbiology* **151**, 2899–2905; 10.1099/mic.0.28139-0 (2005).
21. Amano, S. *et al.* Promomycin, a polyether promoting antibiotic production in *Streptomyces* spp. *J. Antibiot.* **63**, 486–491; 10.1038/ja.2010.68 (2010).
22. Horinouchi, S. & Beppu, T. Hormonal control by A-factor of morphological development and secondary metabolism in *Streptomyces*. *Proc. Jpn. Acad., Ser. B* **83**, 277–295; 10.2183/pjab.83.277 (2007).
23. Zhu, H., Sandiford, S. K. & van Wezel, G. P. Triggers and cues that activate antibiotic production by actinomycetes. *J. Ind. Microbiol. Biotechnol.* **41**, 371–386; 10.1007/s10295-013-1309-z (2014).
24. Chen, L. *et al.* High-Throughput Screening for *Streptomyces* Antibiotic Biosynthesis Activators. *Appl. Environ. Microbiol.* **78**, 4526–4528; 10.1128/AEM.00348-12 (2012).
25. Rachid, S., Gerth, K., Kochems, I. & Müller, R. Deciphering regulatory mechanisms for secondary metabolite production in the myxobacterium *Sorangium cellulosum* So ce56. *Mol. Microbiol.* **63**, 1783–1796; 10.1111/j.1365-2958.2007.05627.x (2007).

26. Rachid, S., Sasse, F., Beyer, S. & Müller, R. Identification of StiR, the first regulator of secondary metabolite formation in the myxobacterium *Cystobacter fuscus* Cb f17.1. *J. Biotechnol.* **121**, 429–441; 10.1016/j.jbiotec.2005.08.014 (2006).
27. Ueda, K. & Beppu, T. Antibiotics in microbial coculture. *J. Antibiot.* **70**, 361–365; 10.1038/ja.2016.127 (2017).
28. Panter, F., Krug, D., Baumann, S. & Müller, R. Self-resistance guided genome mining uncovers new topoisomerase inhibitors from myxobacteria. *Chem. Sci.* **9**, 4898–4908; 10.1039/C8SC01325J (2018).
29. Pogorevc, D. Thesis. Saarland University, Jan 2019.
30. Pogorevc, D. *et al.* Biosynthesis and Heterologous Production of Argyrins. *ACS Synth. Biol.* **8**, 1121–1133; 10.1021/acssynbio.9b00023 (2019).
31. 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; 10.1002/anie.201508277 (2015).
32. 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; 10.1016/j.ymben.2017.10.004 (2017).
33. Pogorevc, D. *et al.* Production optimization and biosynthesis revision of coralopyronin A, a potent anti-filarial antibiotic. *Metab. Eng.* **55**, 201–211; 10.1016/j.ymben.2019.07.010 (2019).
34. Julien, B. & Shah, S. Heterologous expression of epothilone biosynthetic genes in *Myxococcus xanthus*. *Antimicrob. Agents Chemother.* **46**, 2772–2778; 10.1128/AAC.46.9.2772–2778.2002 (2002).
35. Lau, J. *et al.* Optimizing the heterologous production of epothilone D in *Myxococcus xanthus*. *Biotechnol. Bioeng.* **78**, 280–288 (2002).
36. Tang, L. *et al.* Cloning and heterologous expression of the epothilone gene cluster. *Science* **287**, 640–642 (2000).
37. Park, S. R. *et al.* Heterologous production of epothilones B and D in *Streptomyces venezuelae*. *Appl. Microbiol. Biotechnol.* **81**, 109–117; 10.1007/s00253-008-1674-0 (2008).
38. Mutka, S. C., Carney, J. R., Liu, Y. & Kennedy, J. Heterologous Production of Epothilone C and D in *Escherichia coli*. *Biochemistry* **45**, 1321–1330; 10.1021/bi052075r (2006).

- 
39. 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; 10.1007/s00203-005-0059-3 (2006).
40. Yan, F. *et al.* Synthetic biology approaches and combinatorial biosynthesis towards heterologous lipopeptide production. *Chem. Sci.* **9**, 7510–7519; 10.1039/c8sc02046a (2018).
41. Fu, J. *et al.* Efficient transfer of two large secondary metabolite pathway gene clusters into heterologous hosts by transposition. *Nucleic Acids Res.* **36**, e113; 10.1093/nar/gkn499 (2008).
42. Wenzel, S. C. *et al.* Heterologous expression of a myxobacterial natural products assembly line in pseudomonads via red/ET recombineering. *Chem. Biol.* **12**, 349–356; 10.1016/j.chembiol.2004.12.012 (2005).
43. Perlova, O., Gerth, K., Kuhlmann, S., Zhang, Y. & Müller, R. Novel expression hosts for complex secondary metabolite megasynthetases: Production of myxochromide in the thermophilic isolate *Corallocooccus macrosporus* GT-2. *Microb. Cell Fact.* **8**; 10.1186/1475-2859-8-1 (2009).
44. 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; 10.1128/AEM.01503-06 (2006).
45. Gross, F. *et al.* Metabolic engineering of *Pseudomonas putida* for methylmalonyl-CoA biosynthesis to enable complex heterologous secondary metabolite formation. *Chem. Biol.* **13**, 1253–1264; 10.1016/j.chembiol.2006.09.014 (2006).
46. 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; 10.1128/AEM.02841-09 (2010).
47. 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; 10.1016/j.chembiol.2012.01.007 (2012).
48. 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; 10.1039/C3SC53163E (2014).

49. Zirkle, R., Ligon, J. M. & Molnar, I. Heterologous production of the antifungal polyketide antibiotic soraphen A of *Sorangium cellulosum* So ce26 in *Streptomyces lividans*. *Microbiology* **150**, 2761–2774; 10.1099/mic.0.27138-0 (2004).
50. 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; 10.1002/anie.201802479 (2018).
51. Yin, S. *et al.* Heterologous expression of oxytetracycline biosynthetic gene cluster in *Streptomyces venezuelae* WVR2006 to improve production level and to alter fermentation process. *Appl. Microbiol. Biotechnol.* **100**, 10563–10572; 10.1007/s00253-016-7873-1 (2016).
52. Wang, Y., Schnell, B., Müller, R. & Begley, T. P. Iterative Methylations Resulting in the Biosynthesis of the t-Butyl Group Catalyzed by a B12-Dependent Radical SAM Enzyme in Cystobactamid Biosynthesis. *Methods Enzymol.* **606**, 199–216; 10.1016/bs.mie.2018.05.018 (2018).
53. Ke, J. & Yoshikuni, Y. Multi-chassis engineering for heterologous production of microbial natural products. *Curr. Opin. Biotechnol.* **62**, 88–97; 10.1016/j.copbio.2019.09.005 (2020).
54. Zhou, M. *et al.* Sequential deletion of all the polyketide synthase and nonribosomal peptide synthetase biosynthetic gene clusters and a 900-kb subtelomeric sequence of the linear chromosome of *Streptomyces coelicolor*. *FEMS Microbiol. Lett.* **333**, 169–179; 10.1111/j.1574-6968.2012.02609.x (2012).
55. 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; 10.1016/j.ymben.2018.09.004 (2018).
56. Reuß, D. R. *et al.* Large-scale reduction of the *Bacillus subtilis* genome: consequences for the transcriptional network, resource allocation, and metabolism. *Genome Res.* **27**, 289–299; 10.1101/gr.215293.116 (2017).
57. van Tilburg, A. Y. *et al.* MiniBacillus PG10 as a Convenient and Effective Production Host for Lantibiotics. *ACS Synth. Biol.* **9**, 1833–1842; 10.1021/acssynbio.0c00194 (2020).
58. Moldenhauer, J., Chen, X. H., Borriss, R. & Piel, J. Biosynthesis of the antibiotic bacillaene, the product of a giant polyketide synthase complex of the trans-AT family. *Angew. Chem. Int. Ed. Engl.* **46**, 8195–8197 (2007).

- 
59. Fredens, J. *et al.* Total synthesis of *Escherichia coli* with a recoded genome. *Nature* **569**, 514–518; 10.1038/s41586-019-1192-5 (2019).
60. Chin, J. W. Expanding and reprogramming the genetic code. *Nature* **550**, 53–60; 10.1038/nature24031 (2017).
61. Wenzel, S. C. & Müller, R. in *Industrial biotechnology, Microorganisms Volume 3a and 3b*, edited by C. Wittmann & J. Liao (Wiley-VCH, Weinheim, Germany, 2017), pp. 453–485.
62. Ara, K. *et al.* *Bacillus* minimum genome factory: effective utilization of microbial genome information. *Biotechnol. Appl. Biochem.* **46**, 169–178; 10.1042/BA20060111 (2007).
63. Magrini, V., Creighton, C. & Youderian, P. Site-specific recombination of temperate *Myxococcus xanthus* phage Mx8: Genetic elements required for integration. *J. Bacteriol.* **181**, 4050–4061 (1999).
64. Julien, B. Characterization of the integrase gene and attachment site for the *Myxococcus xanthus* bacteriophage Mx9. *J. Bacteriol.* **185**, 6325–6330; 10.1128/JB.185.21.6325–6330.2003 (2003).
65. Manderscheid, N. *et al.* An influence of the copy number of biosynthetic gene clusters on the production level of antibiotics in a heterologous host. *J. Biotechnol.* **232**, 110–117; 10.1016/j.jbiotec.2016.05.038 (2016).
66. 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; 10.3389/fmicb.2018.02928 (2018).
67. Hopwood, D. A. How do antibiotic-producing bacteria ensure their self-resistance before antibiotic biosynthesis incapacitates them? *Mol. Microbiol.* **63**, 937–940; 10.1111/j.1365-2958.2006.05584.x (2007).
68. Ongley, S., Bian, X., Neilan, B. A. & Müller, R. Recent advances in the heterologous expression of microbial natural product biosynthetic pathways. *Nat. Prod. Rep.* **30**, 1121–1138; 10.1039/c3np70034h (2013).
69. M. Rhia L. Stone, Wanida Phetsang, Matthew A. Cooper & Mark A. T. Blaskovich. Visualization of Bacterial Resistance using Fluorescent Antibiotic Probes. *JoVE (Journal of Visualized Experiments)*, e60743; 10.3791/60743 (2020).

70. O'Shea, R. & Moser, H. E. Physicochemical properties of antibacterial compounds: implications for drug discovery. *J. Med. Chem.* **51**, 2871–2878; 10.1021/jm700967e (2008).
71. Iyer, R. *et al.* Evaluating LC–MS/MS To Measure Accumulation of Compounds within Bacteria. *ACS Infect. Dis.* **4**, 1336–1345; 10.1021/acsinfecdis.8b00083 (2018).
72. Myronovskiy, M. & Luzhetskyy, A. Native and engineered promoters in natural product discovery. *Nat. Prod. Rep.*; 10.1039/c6np00002a (2016).
73. Anthony, J. R. *et al.* Optimization of the mevalonate-based isoprenoid biosynthetic pathway in *Escherichia coli* for production of the anti-malarial drug precursor amorpha-4,11-diene. *Metab. Eng.* **11**; 10.1016/j.ymben.2008.07.007 (2009).
74. Farasat, I. *et al.* Efficient search, mapping, and optimization of multi-protein genetic systems in diverse bacteria. *Mol. Syst. Biol.* **10**; 10.15252/msb.20134955 (2014).
75. Ajikumar, P. K. *et al.* Isoprenoid pathway optimization for Taxol precursor overproduction in *Escherichia coli*. *Science* **330**, 70–74; 10.1126/science.1191652 (2010).
76. Ausländer, S., Ausländer, D. & Fussenegger, M. Synthetic Biology—The Synthesis of Biology. *Angew. Chem. Int. Ed. Engl.* **56**, 6396–6419; 10.1002/anie.201609229 (2017).
77. Kosuri, S. & Church, G. M. Large-scale de novo DNA synthesis: technologies and applications. *Nat. Methods* **11**, 499–507; 10.1038/nmeth.2918 (2014).
78. Kodumal, S. J. *et al.* Total synthesis of long DNA sequences: Synthesis of a contiguous 32-kb polyketide synthase gene cluster. *Proc. Natl. Acad. Sci. U.S.A.* **101**, 15573–15578 (2004).
79. Conlin, A., Fournier, M., Hudis, C., Kar, S. & Kirkpatrick, P. Ixabepilone. *Nat. Rev. Drug Discov.* **6**, 953–954; 10.1038/nrd2469 (2007).
80. Fernandes, P., Martens, E. & Pereira, D. Nature nurtures the design of new semi-synthetic macrolide antibiotics. *J. Antibiot.* **70**, 527–533; 10.1038/ja.2016.137 (2017).
81. Zuckerman, J. M. Macrolides and ketolides: azithromycin, clarithromycin, telithromycin. *Infect. Dis. Clin. North Am.* **18**; 10.1016/j.idc.2004.04.010 (2004).
82. Fernandes, P., Martens, E., Bertrand, D. & Pereira, D. The solithromycin journey—It is all in the chemistry. *Bioorg. Med. Chem.* **24**, 6420–6428; 10.1016/j.bmc.2016.08.035 (2016).

- 
83. Pawlowski, A. C., Johnson, J. W. & Wright, G. D. Evolving medicinal chemistry strategies in antibiotic discovery. *Curr. Opin. Biotechnol.* **42**, 108–117; 10.1016/j.copbio.2016.04.006 (2016).
84. Negash, K. H., Norris, J. K.S. & Hodgkinson, J. T. Siderophore–Antibiotic Conjugate Design: New Drugs for Bad Bugs? *Molecules* **24**, 3314; 10.3390/molecules24183314 (2019).
85. Lin, Y.-M., Ghosh, M., Miller, P. A., Möllmann, U. & Miller, M. J. Synthetic sideromycins (skepticism and optimism): selective generation of either broad or narrow spectrum Gram-negative antibiotics. *Biometals* **32**, 425–451; 10.1007/s10534-019-00192-6 (2019).
86. Wencewicz, T. A. & Miller, M. J. in *Antibacterials. Volume II*, edited by J. F. Fisher, S. Mobashery & M. J. Miller (Springer, Cham, Switzerland, 2018), pp. 151–183.
87. Zähler, H., Diddens, H., Keller-Schierlein, W. & Nägeli, H. U. Some experiments with semisynthetic sideromycins. *Jpn. J. Antibiot.* **30**, Suppl:201-6 (1977).
88. Kinzel, O., Tappe, R., Gerus, I. & Budzikiewicz, H. The Synthesis and Antibacterial Activity of two Pyoverdin-ampicillin Conjugates, Entering *Pseudomonas aeruginosa* via the Pyoverdin-mediated Iron Uptake Pathway. *J. Antibiot.* **51**, 499–507; 10.7164/antibiotics.51.499 (1998).
89. Zhanel, G. G. *et al.* Cefiderocol: A Siderophore Cephalosporin with Activity Against Carbapenem-Resistant and Multidrug-Resistant Gram-Negative Bacilli. *Drugs* **79**, 271–289; 10.1007/s40265-019-1055-2 (2019).
90. Ghosh, M. *et al.* Siderophore Conjugates of Daptomycin are Potent Inhibitors of Carbapenem Resistant Strains of *Acinetobacter baumannii*. *ACS Infect. Dis.* **4**, 1529–1535; 10.1021/acsinfecdis.8b00150 (2018).
91. Ghosh, M. *et al.* Targeted Antibiotic Delivery: Selective Siderophore Conjugation with Daptomycin Confers Potent Activity against Multidrug Resistant *Acinetobacter baumannii* Both *in Vitro* and *in Vivo*. *J. Med. Chem.* **60**, 4577–4583; 10.1021/acs.jmedchem.7b00102 (2017).
92. Braun, V., Pramanik, A., Gwinner, T., Köberle, M. & Bohn, E. Sideromycins: tools and antibiotics. *Biometals* **22**, 3–13; 10.1007/s10534-008-9199-7 (2009).
93. Pramanik, A. & Braun, V. Albomycin uptake via a ferric hydroxamate transport system of *Streptococcus pneumoniae* R6. *J. Bacteriol.* **188**, 3878–3886; 10.1128/JB.00205-06 (2006).

94. Sackmann, W., Reusser, P., Neipp, L., Kradolfer, F. & Gross, F. Ferrimycin A, a New Iron-Containing Antibiotic. *Antibiot. Chemother.* **12**, 34–45 (1962).
95. Knüsel, F. & Nüesch, J. Mechanism of Action of Sideromycins. *Nature* **206**, 674–676; 10.1038/206674a0 (1965).
96. Gause, G. F. Recent studies on albomycin, a new antibiotic. *Br. Med. J.* **2**, 1177–1179; 10.1136/bmj.2.4949.1177 (1955).
97. Ghosh, A. *et al.* Iron transport-mediated drug delivery using mixed-ligand siderophore- $\beta$ -lactam conjugates. *Chem. Biol.* **3**, 1011–1019; 10.1016/s1074-5521(96)90167-2 (1996).
98. Roy, A. D., Gruschow, S., Cairns, N. & Goss, R. J. M. Gene expression enabling synthetic diversification of natural products: chemogenetic generation of pacidamycin analogs. *J. Am. Chem. Soc.* **132**, 12243–12245; 10.1021/ja1060406 (2010).
99. Eustaquio, A. S. *et al.* Clorobiocin Biosynthesis in *Streptomyces*: Identification of the Halogenase and Generation of Structural Analogs. *Chem. Biol.* **10**, 279–288; 10.1016/S1074-5521(03)00051-6 (2003).
100. Eustaquio, A. S. *et al.* Production of 8'-halogenated and 8'-unsubstituted novobiocin derivatives in genetically engineered *Streptomyces coelicolor* strains. *Chem. Biol.* **11**, 1561–1572; 10.1016/j.chembiol (2004).
101. Chen, J. S. *et al.* Functional identification of the gene encoding the enzyme involved in mannosylation in ramoplanin biosynthesis in *Actinoplanes* sp. *Biotechnol. Lett.* **35**, 1501-1508; 10.1007/s10529-013-1233-3 (2013).
102. Higashide, E., Hatano, K., Shibata, M. & Nakazawa, K. Enduracin, a New Antibiotic. I. *J. Antibiot.* **21**, 126–137; 10.7164/antibiotics.21.126 (1968).
103. Cavalleri, B., Pagani, H., Volpe, G., Selva, E. & Parenti, F. A-16686, a New Antibiotic from *Actinoplanes*. *J. Antibiot.* **37**, 309–317; 10.7164/antibiotics.37.309 (1984).
104. Winn, M., Fyans, J. K., Zhuo, Y. & Micklefield, J. Recent advances in engineering nonribosomal peptide assembly lines. *Nat. Prod. Rep.* **33**, 317–347; 10.1039/C5NP00099H (2016).
105. Wu, M.-C. *et al.* Engineered biosynthesis of enduracidin lipoglycopeptide antibiotics using the ramoplanin mannosyltransferase Ram29. *Microbiology* **161**, 1338–1347; 10.1099/mic.0.000095 (2015).

- 
106. Cudic, P. *et al.* Functional Analysis of the Lipoglycopeptide Antibiotic Ramoplanin. *Chem. Biol.* **9**, 897–906; 10.1016/s1074-5521(02)00191-6 (2002).
107. Baltz, R. H. Combinatorial Biosynthesis of Cyclic Lipopeptide Antibiotics: A Model for Synthetic Biology To Accelerate the Evolution of Secondary Metabolite Biosynthetic Pathways. *ACS Synth. Biol.* **3**, 748–758; 10.1021/sb3000673 (2014).
108. Chen, C.-Y., Georgiev, I., Anderson, A. C. & Donald, B. R. Computational structure-based redesign of enzyme activity. *Proc. Natl. Acad. Sci. USA* **106**, 3764–3769; 10.1073/pnas.0900266106 (2009).
109. Eppelmann, K., Stachelhaus, T. & Marahiel, M. A. Exploitation of the selectivity-conferring code of nonribosomal peptide synthetases for the rational design of novel peptide antibiotics. *Biochemistry* **41**, 9718–9726; 10.1021/bi0259406 (2002).
110. Stevens, B. W., Lilien, R. H., Georgiev, I., Donald, B. R. & Anderson, A. C. Redesigning the PheA domain of gramicidin synthetase leads to a new understanding of the enzyme's mechanism and selectivity. *Biochemistry* **45**, 15495–15504; 10.1021/bi061788m (2006).
111. Kaljunen, H. *et al.* Structural Elucidation of the Bispecificity of A Domains as a Basis for Activating Non-natural Amino Acids. *Angew. Chem. Int. Ed. Engl.* **54**, 8833–8836; 10.1002/anie.201503275 (2015).
112. Stachelhaus, T., Mootz, H. D. & Marahiel, M. A. The specificity-conferring code of adenylation domains in nonribosomal peptide synthetases. *Chem. Biol.* **6**, 493–505; 10.1016/S1074-5521(99)80082-9 (1999).
113. 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; 10.1093/nar/gki885 (2005).
114. Kries, H. *et al.* Reprogramming nonribosomal peptide synthetases for "clickable" amino acids. *Angew. Chem. Int. Ed. Engl.* **53**, 10105–10108; 10.1002/anie.201405281 (2014).
115. Miao, V. *et al.* Genetic engineering in *Streptomyces roseosporus* to produce hybrid lipopeptide antibiotics. *Chem. Biol.* **13**, 269–276; 10.1016/j.chembiol.2005.12.012 (2006).
116. Nguyen, K. T. *et al.* Combinatorial biosynthesis of novel antibiotics related to daptomycin. *Proc. Natl. Acad. Sci. USA* **103**, 17462–17467; 10.1073/pnas.0608589103 (2006).

117. Calcott, M. J., Owen, J. G., Lamont, I. L. & Ackerley, D. F. Biosynthesis of Novel Pyoverdines by Domain Substitution in a Nonribosomal Peptide Synthetase of *Pseudomonas aeruginosa*. *Appl. Environ. Microbiol.* **80**, 5723–5731; 10.1128/AEM.01453-14 (2014).
118. Butz, D. *et al.* Module extension of a non-ribosomal peptide synthetase of the glycopeptide antibiotic balhimycin produced by *Amycolatopsis balhimycina*. *ChemBioChem* **9**, 1195–1200; 10.1002/cbic.200800068 (2008).
119. Mootz, H. D. *et al.* Decreasing the Ring Size of a Cyclic Nonribosomal Peptide Antibiotic by In-Frame Module Deletion in the Biosynthetic Genes. *J. Am. Chem. Soc.* **124**, 10980–10981; 10.1021/ja027276m (2002).
120. Gao, L. *et al.* Module and individual domain deletions of NRPS to produce plipastatin derivatives in *Bacillus subtilis*. *Microb. Cell Fact.* **17**, 1–13; 10.1186/s12934-018-0929-4 (2018).
121. Bozhüyük, K. A. J. *et al.* *De novo* design and engineering of non-ribosomal peptide synthetases. *Nat. Chem.* **10**, 275–281; 10.1038/nchem.2890 (2018).
122. Bozhüyük, K. A. J. *et al.* Modification and *de novo* design of non-ribosomal peptide synthetases using specific assembly points within condensation domains. *Nat. Chem.* **11**, 653–661; 10.1038/s41557-019-0276-z (2019).