

# Developing Heterologous Expression Platforms for Biosynthetic Pathways of Myxobacterial Natural Products

Dissertation

zur Erlangung des Grades

des Doktors der Naturwissenschaften

der Naturwissenschaftlich-Technischen Fakultät

der Universität des Saarlandes

von

**Jan Schlemmer**

Saarbrücken

2022

Tag des Kolloquiums: 04.08.2022

Dekan: Prof. Dr. Jörn Eric Walter

Berichterstatter: Prof. Dr. Rolf Müller

Prof. Dr. Andriy Luzhetskyy

Vorsitz: Prof. Dr. Alexander Titz

Akad. Mitarbeiter: Dr. Michael Ring

Die vorliegende Arbeit wurde von November 2017 bis Mai 2022 unter der Anleitung von Herrn Prof. Dr. Rolf Müller am Helmholtz-Institut für Pharmazeutische Forschung Saarland (HIPS) angefertigt.

*“When we see the Earth from space, we see ourselves as a whole; we see the unity and not the divisions. It is such a simple image, with a compelling message:*

*“One planet, one human race”.*

Stephen Hawking

*“It's not the size of the dog in fight, it's the size of the fight in the dog.”*

Anonymous American proverb

## Acknowledgments

First, I want to thank Prof. Dr. Rolf Müller for giving me the opportunity and his trust to pursue my PhD in this exceptional work group and supervising my research. With his efforts and hard work he puts into HIPS he created an immaculate research environment for all of us.

I further want to thank Prof. Dr. Andriy Luzhetskyy for accompanying my work as a second supervisor, his valuable input in the thesis committees, and for reviewing my thesis.

I also want to thank my supervisors through the first years, Dr. Katja Gemperlein, Dr. Fabian Panter, and Dr. Domen Pogorevc for always being incredibly helpful, teaching me everything I know about lab work, and always having great advice for me. Further, I want to thank Dr. Chengzhang Fu, Dr. Daniel Krug, and Dr. Joachim Hug for supporting me through the last year with helpful discussions and proofreading my thesis.

My deepest gratitude goes to the great people of HIPS, for making the last four years an unforgettable experience and always being on my side through the ups and downs of my research. Not only did we have fun times during work, but we also made amazing experiences outside of work. Fabienne Wittling who is an indispensable help in the MXV project since her masters work, Dr. Sebastian Groß who was an amazing office buddy and friend, Patrick Haack with his hugs that have the magic to ease every pain, my Beard Brother companions Dr. Jake Haeckl, Dr. Alexander Kiefer (AIKI), and Dr. Sebastian Adam who are like brothers, Nicolas Frank who went through the struggle of writing a dissertation with me and made it so much more bearable, Christine Walt and Dr. Chantal Bader who were not only of great assistance for analytical issues but also great friends, Dr. Kamal Tehrani who was brave enough to join the MXV team and is a great addition ever since, and all the great people of HIPS-MINS that made this time so memorable.

I further want to highlight my thanks to my Chinese friends Jingjun Mo, Dr. Tingting Wang, Jiaqi Liu, Dr. Feng Xie, and Dr. Hu Zeng. You guys always made my lab time entertaining, even on weekends, and always helped me improving my Chinese skills. 谢谢我的中国朋友。

Our group would not work as nearly as smooth as it does without our office team. I want to thank Ellen Merkel and Christina Decker with all my heart for all the work they do for us.

Furthermore, I want to show my deepest gratitude to our lab assistants who always work hard to make our day-to-day work so much easier.

Here, I also want to thank my former supervisors Prof. Dr. Ian Wheeldon, Prof. Dr. Yi Tang, Dr. Cory Schwarz, Dr. Ann-Kathrin Löbs, Dr. John Billingsley, and Dr. Anthony DeNicola. Without the possibilities, knowledge, inspiration, and motivation they gave me throughout my bachelor and master theses research, I would probably never have pursued a PhD.

My best friends from my bachelor and master studies Philipp Altvater and Dan Nguyen Luong: I am incredibly thankful for the friendship we have since day one of the bachelor in 2012. Through the years of the highs and lows of my PhD you always supported me in your own ways, especially with the help of our extraordinary WhatsApp group.

Last but not least, I want to thank my family. I would not be who I am and where I am today without all the support you gave me since I was born. The lengths of this thesis would not be enough to describe what you have done for me. For this, I will always be grateful.

## Abstract

Natural products from myxobacteria gained increasing interest for pharmaceutical applications in the last decades due to their promising bioactivities. This thesis covers the heterologous production and biosynthesis elucidation of three potent myxobacterial compounds. The design, assembly, and heterologous expression of a synthetic myxovalargin (MXV) gene cluster lead to production of competitive yields in the surrogate host *M. xanthus* DK1622. The established heterologous expression platform for MXV was used to initiate in-depth analysis of the involvement of the putative  $\beta$ -hydroxylase MxvH and the single PCP domain MxvB in the MXV biosynthesis. In order to further elucidate the biosynthesis of the disciformycins (DIF), a previously developed heterologous expression platform of the core PKS genes *difBCDEFG* in *M. xanthus* DK1622 was utilized to search for essential tailoring enzymes. With this approach the Fe-S oxidoreductase DifH was identified as the responsible enzyme for the C-C double bond formation at C-3 of the DIF scaffold. The biosynthesis of the maracens and maracins is still hypothetical to this date. The design and assembly of a synthetic putative gene cluster and the heterologous expression thereof in *M. xanthus* DK1622 did not yield production of the target compound so far. Nevertheless, the designed construct could be used for future efforts to successfully produce the maracens and maracins in a suitable surrogate host.

## Zusammenfassung

Da Naturstoffe aus Myxobakterien vielversprechende Bioaktivitäten aufweisen, ist das Interesse an ihnen für pharmazeutische Anwendungen in den letzten Jahrzehnten stetig gestiegen. Diese Thesis befasst sich mit der heterologen Produktion und Aufklärung der Biosynthese von drei Verbindungen mit starken antibiotischen Wirkungen. Design, Assemblierung und heterologe Expression eines synthetischen Myxovalargin (MXV) Genclusters in dem Wirt *M. xanthus* DK1622 haben zu konkurrenzfähigen Produktionsraten geführt. Die etablierte heterologe Expressionsplattform konnte als Anstoß zu einer tiefer greifenden Analyse der zwei beteiligten Enzymen MxvH, einer mutmaßlichen  $\beta$ -hydroxylase, und MxvB, einer alleinstehenden PCP-Domäne, dienen. Um die Biosynthese der Disciformycine weiter aufzuklären, wurde eine bestehende heterologe Expressionsplattform der Kern-PKS Gene in *M. xanthus* DK1622 verwendet, um nach weiteren notwendigen Enzymen zu suchen. Dabei wurde die Fe-S Oxidoreduktase DifH als Enzym identifiziert, das die Doppelbindung an der C-3 Position des DIF Kerngerüsts einbaut. Die Biosynthese von Maracen und Maracin ist bis dato nicht aufgeklärt. Design, Assemblierung und heterologe Expression eines synthetischen Genclusters in dem Wirt *M. xanthus* DK1622 hat bisher nicht zu der gewünschten Produktion geführt. Dennoch könnte das hier entwickelte Konstrukt zukünftigen Bemühungen dienen, Maracen und Maracin in einem passenden Wirt heterolog zu produzieren.

# Table of contents

<b>1</b>	<b>INTRODUCTION.....</b>	<b>1</b>
1.1	NATURAL PRODUCTS.....	1
1.1.1	<i>Antimicrobial resistance (AMR)</i> .....	4
1.2	MYXOBACTERIA AS PROMISING SOURCE FOR NOVEL SECONDARY METABOLITES.....	5
1.3	MULTI-MODULAR MEGASYNTHETASES AS NATURAL PRODUCT “FACTORIES” .....	7
1.3.1	<i>Nonribosomal peptide synthetases (NRPS)</i> .....	8
1.3.2	<i>Polyketide synthases (PKS)</i> .....	10
1.3.3	<i>Structural diversity in NRPS and PKS</i> .....	12
1.3.4	<i>Fatty acid synthases (FAS)</i> .....	13
1.4	HETEROLOGOUS EXPRESSION OF MYXOBACTERIAL SECONDARY METABOLITE PATHWAYS .....	14
1.5	OUTLINE OF THE PRESENT WORK .....	16
<b>2</b>	<b>HETEROLOGOUS EXPRESSION OF A SYNTHETIC MYXOVALARGIN GENE CLUSTER ..</b>	<b>19</b>
2.1	ABSTRACT.....	20
2.2	INTRODUCTION .....	21
2.3	MATERIAL AND METHODS .....	26
2.3.1	<i>Media</i> .....	26
2.3.2	<i>Antibiotic stock solutions</i> .....	26
2.3.3	<i>Reagents and buffer</i> .....	27
2.3.4	<i>Software</i> .....	28
2.3.5	<i>E. coli cultivation</i> .....	28
2.3.6	<i>Preparation of E. coli electro competent cells</i> .....	29
2.3.7	<i>Cultivation of M. xanthus</i> .....	29
2.3.8	<i>Extraction and analysis of MXV</i> .....	30
2.3.9	<i>Plasmid DNA isolation from E. coli</i> .....	31
2.3.10	<i>Polymerase chain reaction (PCR)</i> .....	32
2.3.11	<i>DNA separation and purification</i> .....	33
2.3.12	<i>Determination of DNA concentration</i> .....	34
2.3.13	<i>Molecular cloning</i> .....	34
2.3.14	<i>E. coli transformation</i> .....	37
2.3.15	<i>M. xanthus DK1622 transformation</i> .....	37
2.3.16	<i>Screening for E. coli colonies</i> .....	38
2.3.17	<i>Red/ET – homology-based recombination</i> .....	38
2.3.18	<i>Gibson Assembly</i> .....	39
2.3.19	<i>Design and assembly of the pMYC21_mxvIK construct</i> .....	39
2.3.20	<i>Design of the targeted mxvH deletion using Red/ET homologous recombination</i> .....	40

2.3.21	<i>Functional inactivation of the active site in mxvB</i> .....	41
2.3.22	<i>MIC analysis of M. xanthus DK1622 wild type and mutants</i> .....	42
2.4	RESULTS AND DISCUSSION .....	43
2.4.1	<i>Work towards the heterologous production of MXV</i> .....	43
2.4.2	<i>Heterologous expression of the synthetic MXV BGC in Myxococcus xanthus DK1622</i> .....	51
2.4.3	<i>Investigation of the MXV biosynthesis utilizing the heterologous expression platform</i> .....	61
2.5	CONCLUSION AND OUTLOOK .....	66
2.6	SI INFORMATION .....	68
2.6.1	<i>List of plasmids and strains generated in this study</i> .....	68
2.6.2	<i>Predicted domain specificity of the MXV gene cluster modules</i> .....	73
2.6.3	<i>Organization of the synthetic MXV gene synthesis fragments</i> .....	74
2.6.4	<i>Location of repetitive sequences</i> .....	75
2.6.5	<i>Detailed assembling strategy of the synthetic MXV BGC</i> .....	76
2.6.6	<i>Analysis of plasmids after the final desplitting process</i> .....	77
2.6.7	<i>Integration of the synthetic gene cluster into Myxococcus xanthus DK1622 via Mx8 phage-integrase</i> .....	80
2.6.8	<i>Functional inactivation of the active site in mxvB</i> .....	80
<b>3</b>	<b>INVESTIGATION OF THE DISCIFORMYCIN BIOSYNTHESIS</b> .....	<b>83</b>
3.1	ABSTRACT .....	84
3.2	INTRODUCTION .....	85
3.3	MATERIAL AND METHODS .....	89
3.3.1	<i>Media</i> .....	89
3.3.2	<i>Antibiotic stock solutions</i> .....	89
3.3.3	<i>Reagents and buffer</i> .....	90
3.3.4	<i>Software</i> .....	91
3.3.5	<i>E. coli cultivation</i> .....	91
3.3.6	<i>Preparation of E. coli electro competent cells</i> .....	91
3.3.7	<i>Cultivation of M. xanthus</i> .....	92
3.3.8	<i>Extraction and analysis of disciformycin</i> .....	92
3.3.9	<i>Plasmid DNA isolation from E. coli</i> .....	93
3.3.10	<i>Polymerase chain reaction (PCR)</i> .....	94
3.3.11	<i>DNA separation and purification</i> .....	95
3.3.12	<i>Determination of DNA concentration</i> .....	95
3.3.13	<i>Molecular cloning</i> .....	96
3.3.14	<i>E. coli transformation</i> .....	97
3.3.15	<i>M. xanthus DK1622 transformation</i> .....	98
3.3.16	<i>Screening for E. coli colonies</i> .....	98
3.3.17	<i>Red/ET – homology-based recombination</i> .....	99

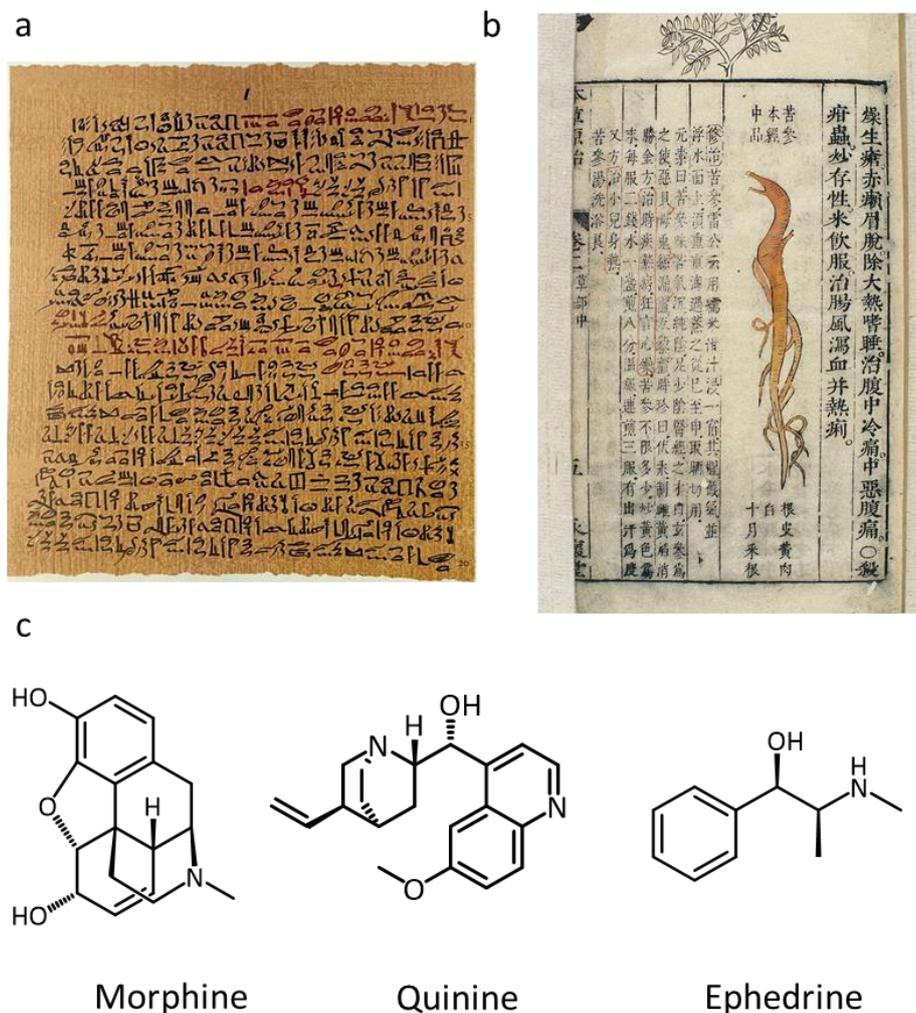
3.3.18	<i>Protein expression and in vitro assays</i> .....	99
3.3.19	<i>Reduction assay for DifA</i> .....	100
3.3.20	<i>In vitro conversion assay with DifA and aglycon DIF294 as substrate</i> .....	101
3.3.21	<i>Generating plasmids TARGT1 and TARGT2</i> .....	101
3.3.22	<i>Replacement of difH by tetR by Red/ET recombineering</i> .....	102
3.4	RESULTS AND DISCUSSION .....	103
3.4.1	<i>In vitro analysis of the cytochrome P450 DifA</i> .....	103
3.4.2	<i>Investigation of two putative glycosyltransferase operons</i> .....	104
3.4.3	<i>Investigation of the putative tailoring enzyme DifH</i> .....	108
3.5	CONCLUSION AND OUTLOOK .....	110
3.6	SUPPORTING INFORMATION .....	111
3.6.1	<i>List of plasmids and strains generated in this study</i> .....	111
3.6.2	<i>Primers used in this work</i> .....	113
<b>4</b>	<b>TOWARDS THE HETEROLOGOUS EXPRESSION OF A SYNTHETIC</b>	
	<b>MARACEN/MARACIN GENE CLUSTER .....</b>	<b>115</b>
4.1	ABSTRACT .....	115
4.2	INTRODUCTION .....	116
4.3	MATERIAL AND METHODS .....	122
4.3.1	<i>Media</i> .....	122
4.3.2	<i>Antibiotic stock solutions</i> .....	122
4.3.3	<i>Reagents and buffer</i> .....	123
4.3.4	<i>Software</i> .....	123
4.3.5	<i>E. coli cultivation</i> .....	124
4.3.6	<i>Preparation of E. coli electro competent cells</i> .....	124
4.3.7	<i>Cultivation of M. xanthus</i> .....	125
4.3.8	<i>Extraction and analysis of maracen/maracin</i> .....	125
4.3.9	<i>Plasmid DNA isolation from E. coli</i> .....	126
4.3.10	<i>Polymerase chain reaction (PCR)</i> .....	126
4.3.11	<i>DNA separation and purification</i> .....	127
4.3.12	<i>Determination of DNA concentration</i> .....	128
4.3.13	<i>Molecular cloning</i> .....	128
4.3.14	<i>E. coli transformation</i> .....	130
4.3.15	<i>M. xanthus DK1622 transformation</i> .....	130
4.3.16	<i>Screening for E. coli colonies</i> .....	131
4.3.17	<i>Gibson Assembly</i> .....	131
4.4	RESULTS AND DISCUSSION .....	132
4.4.1	<i>Design of the synthetic maracen/maracin gene cluster</i> .....	132
4.4.2	<i>Assembly of the synthetic maracen/maracin gene cluster</i> .....	135

4.4.3	<i>Analysis of the heterologous maracen/maracin production with the synthetic gene cluster in the host M. xanthus DK1622::pHybPfa1-Mx9.2</i> .....	135
4.5	CONCLUSION AND OUTLOOK.....	138
4.6	SUPPORTING INFORMATION .....	140
4.6.1	<i>List of plasmids and strains generated or used in this study</i> .....	140
4.6.2	<i>Design and assembly of the synthetic maracen/maracin gene cluster</i> .....	141
4.6.3	<i>Analysis of the integration of the synthetic maracen/maracin gene cluster via Mx8 integrase</i> .....	143
<b>5</b>	<b>DISCUSSION .....</b>	<b>145</b>
5.1	THE IMPORTANCE OF MYXOBACTERIAL NPS AND SUFFICIENT SUPPLY THEREOF .....	145
5.2	THE IMPORTANCE OF HETEROLOGOUS EXPRESSION OF NP BIOSYNTHETIC GENE CLUSTERS (BGC) .....	146
5.3	INVESTIGATION OF THE BIOSYNTHESIS OF MYXOBACTERIAL SECONDARY METABOLITES BY HETEROLOGOUS EXPRESSION .....	147
5.4	CONSIDERATIONS TOWARDS HETEROLOGOUS PRODUCTION OF MYXOBACTERIAL SECONDARY METABOLITES.....	151
5.5	SEMI-SYNTHESIS OF NPS AS COMPLEMENT TO HETEROLOGOUS EXPRESSION.....	153
5.6	FINAL CONCLUSION.....	156
<b>6</b>	<b>REFERENCES.....</b>	<b>157</b>

# 1 Introduction

## 1.1 Natural products

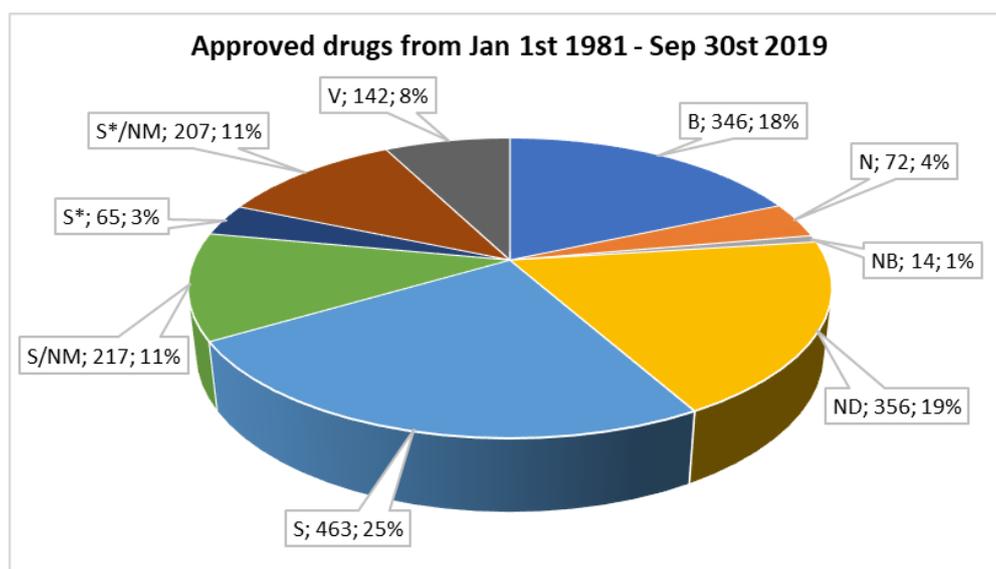
Natural products (NP) can be defined in several perspectives. In a very general attempt to define NPs, they can be regarded as everything that is produced by living systems.<sup>1</sup> Another approach defines NPs as secondary metabolites which are not vital for the producing organisms, *i.e.*, they are not required for their growth.<sup>2</sup> Those secondary metabolites, however, can gain the producer an evolutionary advantage over their competitors.<sup>3</sup> The broad public view on the one hand considers NPs as good and wholesome ‘natural’ or ‘organic’ products by unaltered natural organisms. The scientific view on the other hand regards all compounds as NPs, as long as their chemical identical structure occurs in nature, even if they are produced by genetically modified organisms or chemical synthesis.<sup>4</sup> The production and development of NPs reaches far back to prehuman eras. Millions of years of evolution in plant, fungal, bacterial, and animal sources, created an incredibly vast variety of NPs with a versatile variety of biological activities. The history of humans harvesting the biological activities of NPs for medical treatment, *e.g.*, against cold, parasitic infections and inflammations, can be dated back several millennia.<sup>5</sup> Astonishingly, evidence was found for the use of plant material with considerable therapeutic effect in the Middle Paleolithic age.<sup>6</sup> An extensive list of around 1000 plant-derived substances originates from Mesopotamia and dates to 2600 BC.<sup>5</sup> Another well-known record listing over 700 drugs mostly originating from plants is the Ebers Papyrus, documented around 1500 BC in ancient Egypt.<sup>7</sup> The importance of NPs for medical application in ancient cultures is highlighted by documents listing mostly plant-derived drugs found all over the world. From the Chinese Materia Medica, documented over centuries and containing entries from 1100 BC<sup>8</sup>, over the Indian Ayurvedic system from 1000 BC<sup>9,10</sup> to the Greeks and Romans with drug and formulae records dated 100 – 200 CE. In the Dark and Middle Ages between the 5<sup>th</sup> and the 12<sup>th</sup> centuries, a majority of the Greek and Roman knowledge was preserved by the Arabs, while they also extended it with their own expertise, as well as those of the Chinese and Indians.<sup>11</sup> With the transition into the modern era, the rise of the field of chemistry in the 19<sup>th</sup> century enabled the first isolations of the active compounds of the long-used plant-derived mixtures, *e.g.*, morphine (1816), the antimalarial drug quinine (1820), or the anti-asthma agent ephedrine (1887).<sup>11</sup>



**Figure 1-1 a:** Ebers Papyrus, documenting over 700 drugs from plant origin around 1500 BC in ancient Egypt. 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: 12.03.2022). **b:** Chinese Materia Medica, documented over centuries containing first entries from 1100 BC. Picture modified from [https://commons.wikimedia.org/wiki/File:Chinese\\_Materia\\_medica,\\_C17;\\_Plant\\_drugs,\\_Sophora\\_flavescens\\_Wellcome\\_L0039341.jpg](https://commons.wikimedia.org/wiki/File:Chinese_Materia_medica,_C17;_Plant_drugs,_Sophora_flavescens_Wellcome_L0039341.jpg) (accessed 12.03.2022). **c:** Structure from morphine, quinine, and ephedrine.

Within the early 20<sup>th</sup> century, another turning point in the history of NPs occurred. With the discovery of the first  $\beta$ -lactam antibiotic Penicillin in 1928 by Alexander Fleming, the ‘golden era of antibiotics’ began. The exposition of this new microbial derived group of NPs initiated an extensive investigation of a vast amount of yet undiscovered drugs with anti-infective traits. Until the late 1970s several new microbial produced antibiotic classes were discovered. Microbially derived NPs revealed various activities while many of those secondary metabolites exhibit high potency due to the evolutionary pressures the microorganisms have been exposed to.<sup>12</sup> Some of these biological activities include antibacterial activity (e.g., penicillin, vancomycin, erythromycin, rifamycin),<sup>13,14</sup> cytotoxic activity (e.g., bleomycin and doxorubicin)<sup>13</sup>, antifungal activity (e.g., amphotericin,<sup>14</sup> griseofulvin<sup>13</sup>) as well as

immunosuppressive effects (*e.g.*, cyclosporin<sup>13</sup>). With the discovery of antibiotics like streptomycin, tetracycline or gentamicin, pharmaceutical companies started extensive NP discovery programs.<sup>12</sup> Furthermore, technological advances in the second half of the 20<sup>th</sup> century and ever-developing screening programs lead to a peak of 1600 newly published NPs per year between the 1970s and the 1980s.<sup>15</sup> The development of large synthetic chemical libraries generated by combinatorial chemistry in the 1990s enabled high-throughput screening of pure compounds. Despite several advantages over screening of NP crude extracts<sup>16</sup> a majority of the approved drugs between 1981 and 2018 were NPs or their derivatives, or NP-derived synthetic drugs.<sup>17</sup> The reason for low hit rates of the synthetic chemical library screenings could be their design emphasis for chemical accessibility and maximum achievable size,<sup>18</sup> while NPs were optimized over the millions of years of evolution for their specific biological targets.

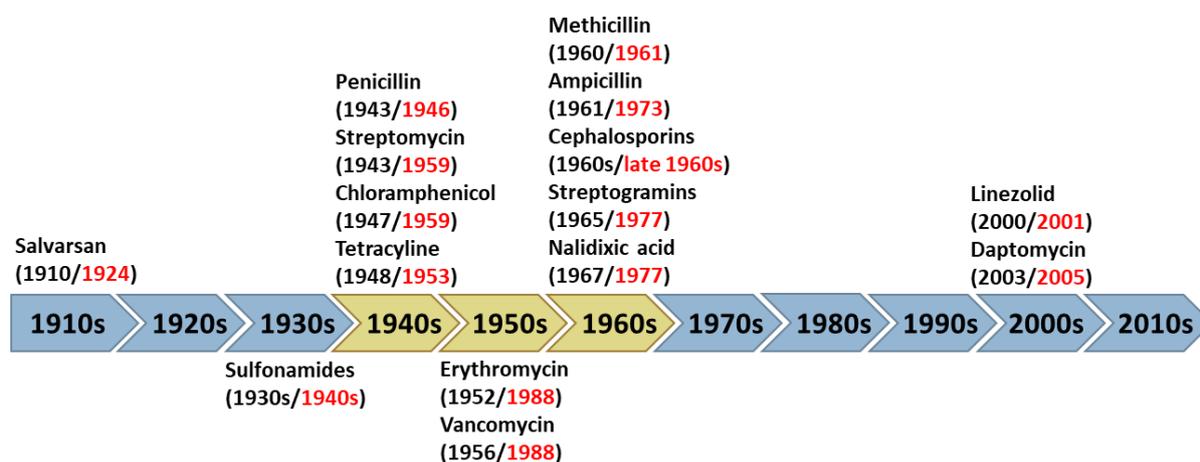


**Figure 1-2** All approved drugs between January 1<sup>st</sup> 1981 and September 30<sup>st</sup> 2019, with a total of 1881 approved drugs. B: biological macromolecule; B: unaltered natural product; NB: botanical drug (defined mixture); ND: natural product derivative; S: synthetic drug; S\*: synthetic drug (NP pharmacophore); V: vaccine; /NM: mimic of natural product. Definitions and values adapted from ref.<sup>17</sup>

However, the abundance of antibiotics entering the market in the 1960s and 70s led to a decline in the need for novel anti-infectives and therefore in interest of the industry to continue the time consuming and expensive research.<sup>19,20</sup> In the following 20 years the only novel broad-spectrum class of antibiotics entering the market were the fluoroquinolones.<sup>20</sup> Since the discovery of new antibiotics is ever decreasing, bacterial antibiotic resistance is on a threatening ascent and prompting the search for novel pharmaceuticals to treat infectious diseases.<sup>21,22</sup>

### 1.1.1 Antimicrobial resistance (AMR)

As stated before, antibiotic resistance or antimicrobial resistance (AMR) of pathogenic bacteria is on a steady rise. For every antibiotic class that entered the market, AMR was observed in the initially susceptible pathogen only a few years later.<sup>23</sup> AMR can be caused by misuse/overuse of antibiotics in humans and animals, leading to the spread of resistance mechanism-carrying pathogenic bacteria obtained by mutations and strong positive selection.<sup>24</sup> However, bacterial strains can also acquire resistance without the anthropogenic use of antibiotics by naturally occurring evolution and horizontal transfer of resistance genes.<sup>24–26</sup> Mutations can lead to modifications of pathways targeted by the antibiotic, or alterations in the affected target site. Further, specific enzymes can inactivate the antibiotic or efflux transporters can render it ineffective.<sup>24,26,27</sup>



**Figure 1-3** Timeline of antibiotics and antibiotic classes entering the market. The year they entered the market and of the first occurrence of clinical resistance (in red) is listed in brackets behind the corresponding antibiotic (class). The ‘golden era of antibiotics’ is highlighted in yellow/golden color. References of the dates: Salvarsan<sup>28</sup>, streptogramins<sup>29,30</sup>, nalidixic acid<sup>31,32</sup>, linezolid<sup>29,33,34</sup>, daptomycin<sup>29,35–37</sup>, and <sup>23</sup> for all others.

Recent statistics attribute approximately 700,000 deaths every year to infections with multidrug-resistant bacteria. Predictions state that, if unstopped, this number might increase to 10 million in 2050.<sup>38</sup> Especially multidrug-resistant Gram-negative bacteria and pathogens of the ESKAPE group (vancomycin-resistant *Enterococci*, *Staphylococcus aureus*, *Klebsiella pneumoniae*, *Acinetobacter baumannii*, *Pseudomonas aeruginosa* and *Enterobacter* species) are a significant threat and therefore ranked with critical priority by the World Health Organization (WHO).<sup>22,39,40</sup> The development of AMR against nearly all approved antibiotics by several pathogens is extremely problematic,<sup>41</sup> especially considering that the last novel class of antibiotics active against Gram-negative bacteria that entered the market were the

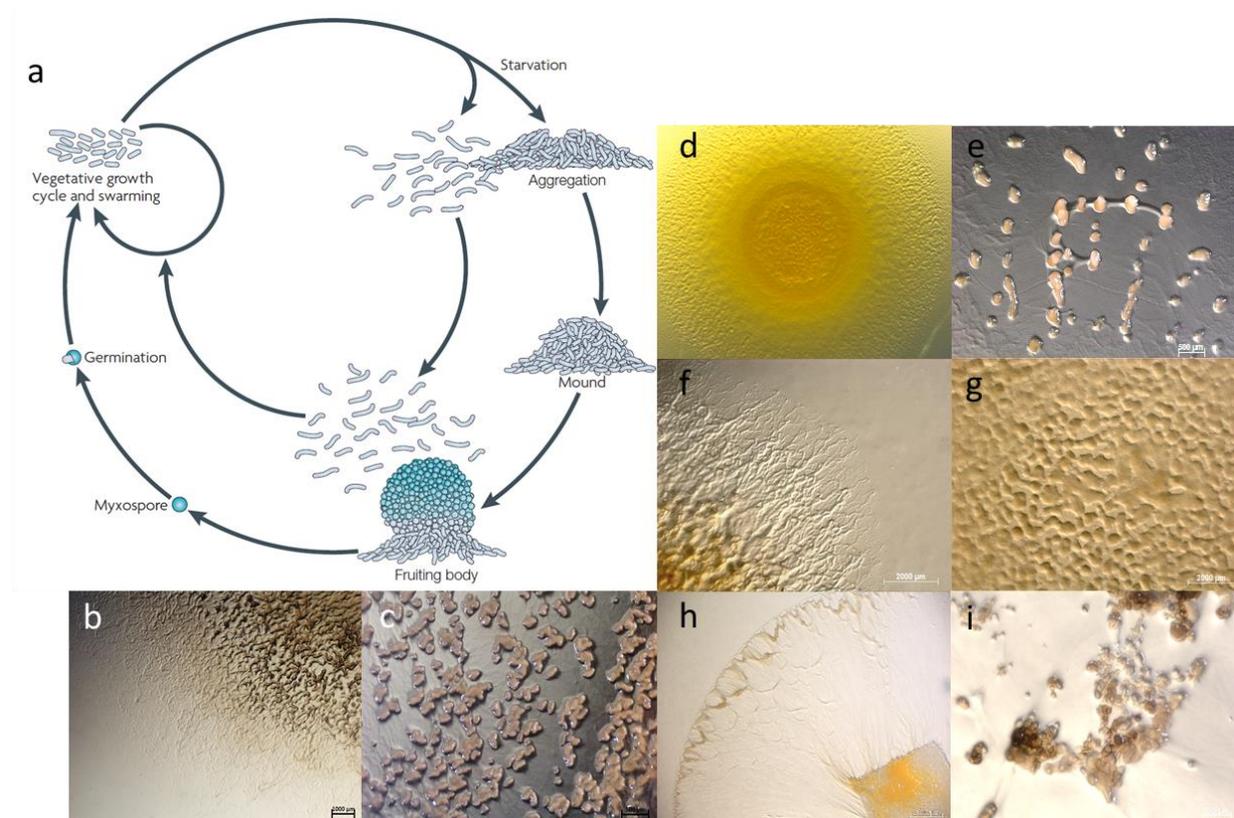
quinolones.<sup>39</sup> With the rise of multidrug resistant pathogens, and the limited time span of antibiotics before AMR towards them emerges, the search for drugs with novel modes of action is increasingly urgent.<sup>42</sup>

## 1.2 Myxobacteria as promising source for novel secondary metabolites

Since the discovery of Penicillin in 1928, microorganisms have been responsible for the majority of discovered antibiotics. Amongst those, actinobacteria played a leading role with over 5000 antibiotics identified in the Actinomycetales order in which 90% are produced by the *Streptomyces* genus alone.<sup>43</sup> The launch of large-scale screenings in the actinobacterial class identified many candidates after the discovery of the two prominent antibiotics streptothricin and streptomycin in the 1940s. However, it is assumed that only a small percentage of the bioactivity potential in this class has been discovered up to recent years due to hampered laboratory amenability of many organisms.<sup>43,44</sup> Further established microbial sources for NPs are fungi (*e.g.*, the cholesterol lowering statins produced by *Aspergillus terreus*)<sup>45</sup> and bacilli (*e.g.*, gramicidin or the lanthionine-containing antibiotics, lantibiotics)<sup>46</sup>.

In the last decades, myxobacteria, a group of Gram-negative  $\delta$ -proteobacteria,<sup>47</sup> arose to prominence in NP discovery. They feature an outstanding ability to produce intriguing secondary metabolites, making them a promising source for novel lead structure development and drug discovery.<sup>48,49</sup> More than 100 unique core structures with fascinating bioactivities and modes of action were found in myxobacteria since the discovery of ambruticin by Warner & Lambert in 1970.<sup>50-53</sup> Myxobacteria are soil residing organisms and known for their rich secondary metabolism, similar to Streptomycetes.<sup>54,55</sup> They exhibit a complex social behavior rarely found anywhere else in the bacterial world.<sup>56</sup> As rod shaped vegetative cells they can glide on surfaces in their search for prey. They feature a sophisticated communication within their swarms, *i.e.*, individual cells can secrete slime forming trails or move as part of cell contact dependent groups.<sup>57</sup> Once the swarm contacts prey, they use hydrolytic exo-enzymes and other secondary metabolites to lyse it and take up its nutrients. Hereby, the entire swarm shares the pool of hydrolytic enzymes to optimize the use of degradation products. In case of nutrient absence, they have the ability to organize their gliding movement to form mounds or even three-dimensional macroscopic fruiting bodies. A small number of cells can turn into resistant myxospores while some become peripheral rods. However, the majority of cells dies,

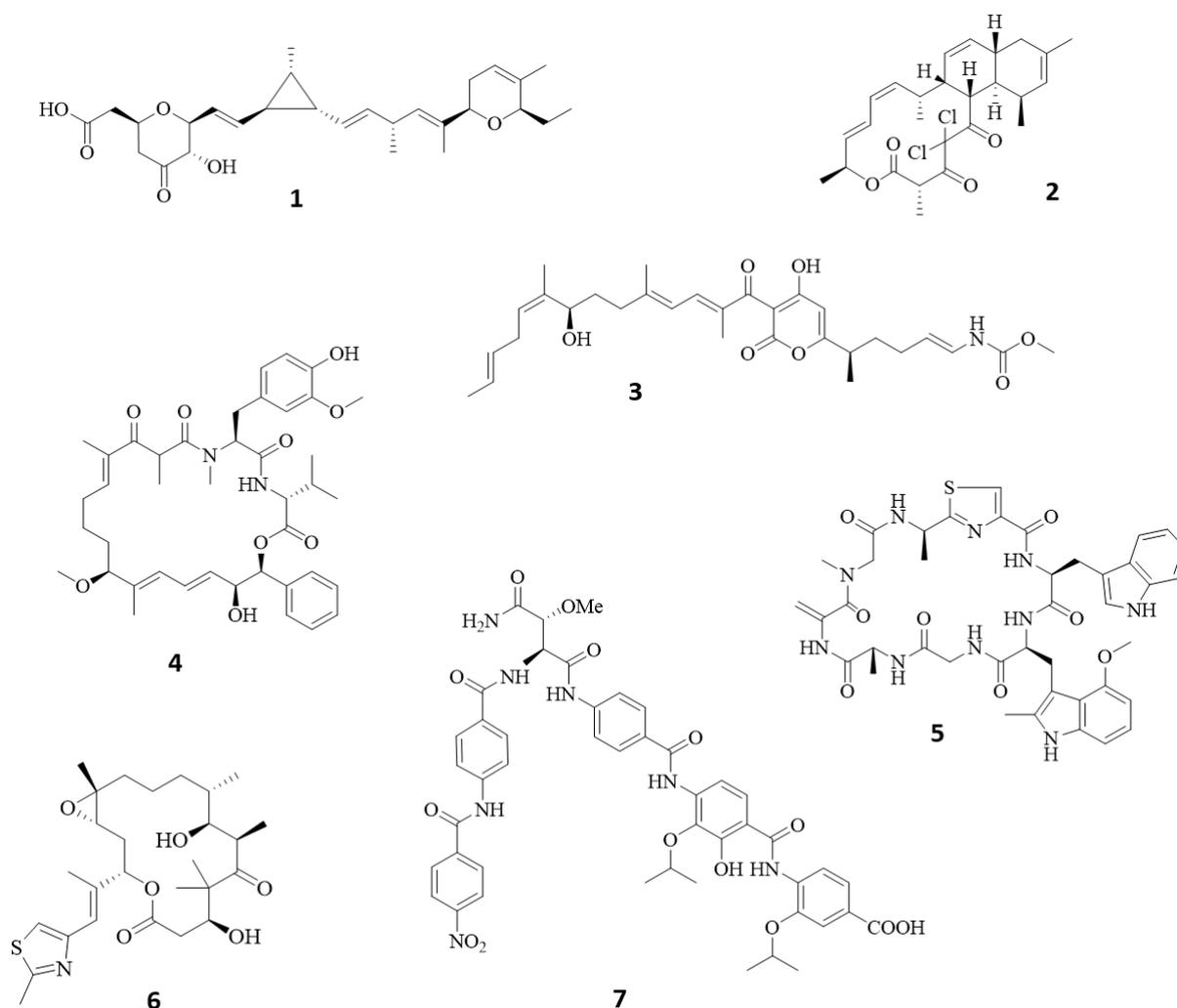
probably to provide nutrients. In this way, myxobacteria are able to survive in hostile environments until they encounter more favorable conditions, which once again causes germination.<sup>57</sup>



**Figure 1-4** a) Lifecycle of *Myxococcus xanthus*.<sup>58</sup> b) *Pyxidicoccus* sp. (*Angiococcus* sp. And48) swarm and c) fruiting bodies. d) *Myxococcus xanthus* DK1622 swarm and e) fruiting bodies. f) *Pyxidicoccus* sp. (*Coralloccoccus* sp.) Ccc1071 swarm and g) fruiting body aggregates. h) *Sorangium cellulosum* swarm and i) fruiting bodies. Microscopic images were kindly provided by Dr. Ronald Garcia.

Their complex way of life is likely the cause for their rich secondary metabolism. With sizes ranging from 9 to 14.8 Mb, myxobacteria possess some of the largest genomes of all known prokaryotes encoding a vast number of biosynthetic gene clusters (BGCs).<sup>55,57,59</sup> Those features and their ability to fight competitors and intruders acquired through evolutionary pressure over millions of years, rendered myxobacteria an intriguing source for bioactive NPs to harvest for drug discovery. The last decades uncovered already a long list of NPs with numerous biological activities like antibacterial (e.g., cystobactamid<sup>60</sup>, disciformycin<sup>61</sup>, myxovalargin<sup>62</sup>, and maracen/maracin<sup>63</sup>), antifungal (e.g., ambruticin<sup>50</sup>), antitumor (e.g., epothilone<sup>64</sup>), antiviral (e.g., aetheramide<sup>65</sup>), immunomodulatory (e.g., argyirin<sup>66</sup>), antimalarial (e.g., chlorotonil<sup>67</sup>), and antifilarial (e.g., corallopyronin<sup>68</sup>). One of the most prominent compounds discovered in myxobacteria is the antitumor agent epothilone, as its

semisynthetic derivative ixabepilone has been approved by the Food and Drug Administration (FDA) for application against breast cancer.<sup>69</sup> Another promising candidate that gained a lot of attention in recent years are the topoisomerase inhibiting cystobactamids, exhibiting highly potent activity against pathogens like *Acinetobacter baumannii*, *Pseudomonas aeruginosa*, and *Escherichia coli*.<sup>60,70,71</sup> Furthermore, the antimalarial macrolactone antibiotic chlorotoniil was isolated from the myxobacterium *Sorangium cellulosum* So ce1525. Besides potent anti-Gram-positive activity, this intriguing candidate exhibited a significant reduction of parasitemia in *Plasmodium berghei* *in vivo* mouse models.<sup>67</sup>



**Figure 1-5** Prominent myxobacterial NPs. **1:** Ambruticin; **2:** Chlorotoniil A; **3:** Corallopyronin A; **4:** Aetheramide A; **5:** Argyrin C; **6:** Epothilone B; **7:** Cystobactamid 919-2

### 1.3 Multi-modular megasynthetases as natural product “factories”

The majority of myxobacterially produced NPs can be classified as nonribosomal peptides (NRPs), polyketides (PKs), or hybrids thereof (PK-NRP hybrids).<sup>72</sup> Other structural types include

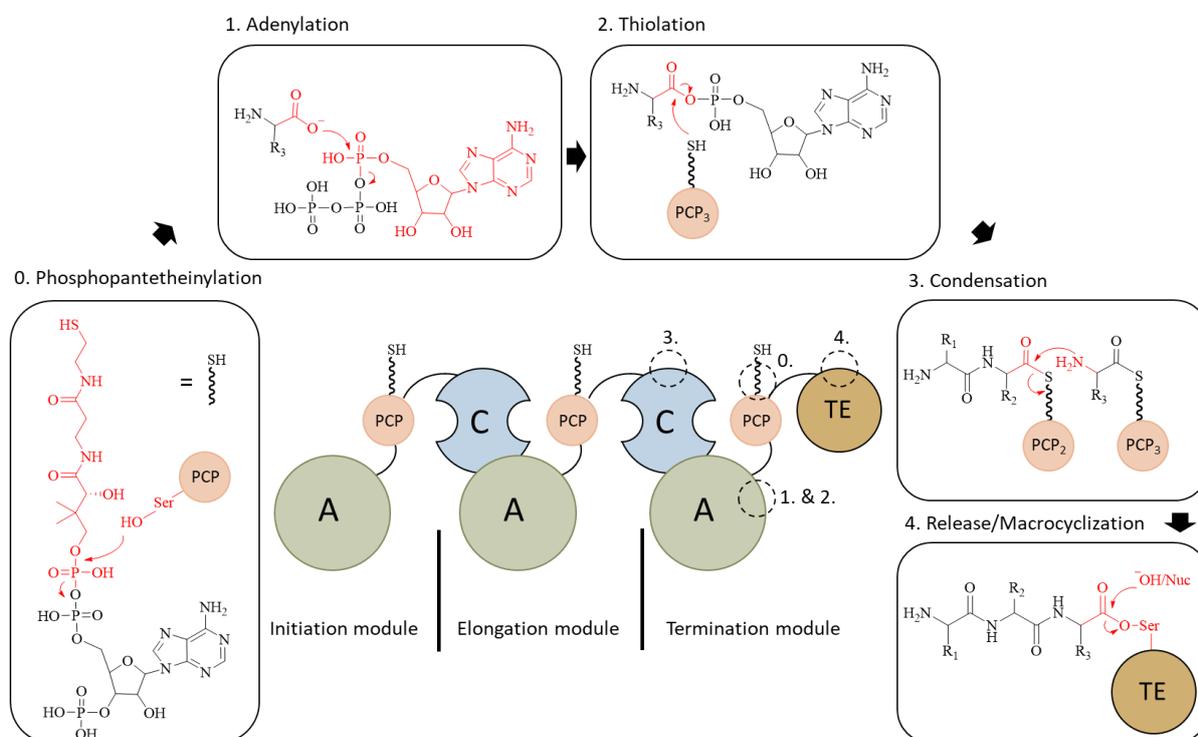
alkaloids, terpenoids, and phenyl-propanoids.<sup>51</sup> NRPs and PKs are both assembled by coupling monomer units into longer chains by large multi-modular enzymes, the so-called megasynthetases (NRPs) and megasynthases (PKs).<sup>72,73</sup> Megasynthetases and megasynthases are organized in a modular manner, with each module constituting a multifunctional subunit composed of various catalytic domains. Each domain has a specific catalytic function necessary for the elongation step within the module. The entirety of those modules forms an assembly line which is responsible for the NP biosynthesis by transferring the growing ketide or peptide chain from module to module. Each module incorporates another monomer unit by iterative chemical condensation steps until the PK/NRP is released at the terminal module of the assembly line. While polyketide synthases (PKS) use acyl-CoA thioesters (*e.g.*, acetyl-CoA, malonyl-CoA, and methylmalonyl-CoA) as monomeric precursors, nonribosomal peptide synthetases have a larger variety of precursors available by utilizing proteinogenic and nonproteinogenic amino acids, as well as other carboxylic acids (*e.g.*, aryl acids).<sup>72</sup> Their structural and catalytic similarities enable interaction between NRPS and PKS to form functional NRPS/PKS hybrid megasynthetases. Interestingly, a high number of myxobacterial NP biosynthetic pathways can be assigned to the NRPS/PKS hybrid type.<sup>74</sup>

### 1.3.1 Nonribosomal peptide synthetases (NRPS)

One of the above-described multi-modular enzyme complexes are the NRPS, which employ a vast pool of precursors consisting of proteinogenic and nonproteinogenic amino acids and other carboxylic acids, to synthesize NRPs. A minimal module for the peptidyl chain elongation in NRPS contains three core domains; condensation domain (C), adenylation domain (A), and peptidyl carrier protein domain (PCP), also referred to as thiolation domain (T). The typical domain order within a NRPS module is C-A-PCP. An elongation step starts with the A domain by selection of the amino acid substrate, activating the carboxyl group with ATP to form aminoacyl-AMP and binding it to the thiol group of the adjacent PCP domain. Subsequently, the PCP domain shuttles the substrate to the following C domain, which forms a peptide bond between two adjacent PCP domain-bound substrates by condensation of the donor carboxyl and the acceptor amine group, resulting in chain elongation.<sup>72</sup>

The PCP domains must be activated by a posttranslational modification step called phosphopantetheinylation to enable functionality. Thereby, they are converted from the inactive *apo* into the active *holo* form by the phosphopantetheinyl transferases (PPTases). This

superfamily of proteins activates the PCP domain by transferring a 4'-phosphopantetheine moiety (PPant) from coenzyme A to the catalytic serine of their active site. Once the PCP domain is activated by phosphopantetheinylation, the chain intermediates can be covalently tethered and shuttled from and to the other catalytic domains.<sup>75</sup>



**Figure 1-6** Basic principle of NRP synthesis, domain arrangement, and their catalytic reactions. **0)** Phosphopantetheinylation of the PCP domain at the conserved serine residue by PPTases. **1)** The A domain mediates selection and adenylation of the amino acid by forming an aminoacyl-AMP and PPi. **2)** Attachment of the activated amino acid to the Ppant moiety of the PCP domain by thiolation and AMP release. **3)** The C domain catalyzes peptide bond formation between the activated amino acid and the upstream amino acid or peptide chain by condensation. **4)** The NRP is released by a terminal TE domain. An intermediate ester bond is formed between the C terminus of the peptide and a conserved serine of the TE domain. Subsequent hydrolysis or intramolecular attack of a nucleophile releases the NRP as linear or macrocyclic product. A = adenylation domain; C = condensation domain; PCP = peptidyl carrier protein domain; TE = thioesterase domain; Nuc = nucleophile. The product of each reaction is highlighted in red. The figure is adapted from Süssmuth & Mainz.<sup>76</sup>

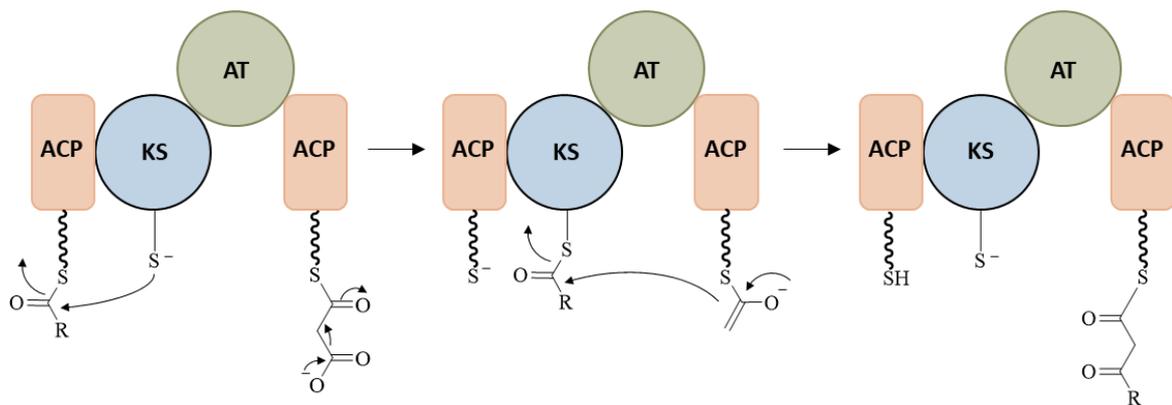
The peptide sequence of the compound is dictated by the substrate specificity of the A domains, and their location among the assembly line.<sup>77</sup> Besides the A domain specificity, C domains with their pseudo dimer organization exhibit a high substrate specificity on the acceptor side and a lower selectivity on the donor side, thus rendering them as a second gate keeper for substrate selection.<sup>78,79</sup> C domains can be classified into several functional subtypes. <sup>L</sup>C<sub>L</sub> domains catalyze peptide bond formation between two L-amino acids, while <sup>D</sup>C<sub>L</sub> domains link a D- with a L-amino acid. Further subtypes are Starter C domains, which acylate the first amino acid with a β-hydroxy-carboxylic acid, Heterocyclization (Cyc) domains,

catalyzing cyclization of cysteine, serine, or threonine residues after peptide bond formation, homologous Epimerization (E) domains, changing the chirality of the last amino acid in the peptide chain, and Dual E/C domains, which catalyze epimerization and condensation.<sup>80</sup> Other domains that modify the PCP domain-bound substrate on the assembly line (*in cis*) are methyltransferase domains (MT), incorporating an additional methyl group,<sup>81</sup> and oxidation domains (Ox), responsible for conversion of thiazoline and oxazoline rings to thiazole and oxazole,<sup>76,77,82</sup> respectively, or for  $\alpha$ -hydroxylation of bound amino acids.<sup>74,83</sup> Modifications of the growing NRP can also occur *in trans*, *i.e.*, during or after release from the NRPS assembly line.<sup>77</sup> Examples for such auxiliary, or tailoring, enzymes are monooxygenases, methyltransferases, *O*-carbamoyl transferases, and glycosyltransferases. Those tailoring enzymes can additionally increase the complexity and variety of NRPs and are sometimes required to act in a strict order to enable the biosynthesis of the final product.<sup>76</sup> The release of the full-length NRP product from the assembly line is usually catalyzed by a terminal thioesterase (TE) domain, which can release the polypeptide either as a linear chain or in cyclic form.<sup>84</sup> The NRPS biosynthesis usually follows the principle of collinearity. This principle claims that each NRPS module activates and couples a single amino acid to the growing peptide chain, *i.e.*, that the number of monomers in the polypeptide correlates with the number of modules in the assembly line and vice versa.<sup>76</sup>

### 1.3.2 Polyketide synthases (PKS)

As in the NRP biosynthesis, the PKS connect monomers into the growing building block chain along a modular, organized assembly line. However, instead of amino acid building blocks, PKS connect activated acyl starter units with malonyl-CoA-derived extender units in a cascade of Claisen condensation reactions.<sup>73</sup> One chain elongation step typically involves three catalytic domains: a  $\beta$ -ketoacylsynthase domain (KS), an acyltransferase domain (AT), and an acyl carrier protein domain (ACP).<sup>73</sup> As for the PCP domains in NRPS, the ACP domain has to be phosphopantetheinylated to convert it from the inactive *apo* into the active *holo* form.<sup>75</sup> Subsequently, after activation, the ACP domain can bind the substrate and shuttle it to the KS domain for C-C bond formation with adjacent monomers or chain intermediates. Once the polyketide passes through the whole assembly line and reaches full chain lengths, a TE domain catalyzes the release of the molecule by hydrolysis or lactonization.<sup>85</sup> PKS are closely related and evolutionarily connected to fatty acid synthases (FAS),<sup>86</sup> however, PKS clearly differ from FAS in providing larger structural diversity. Unlike FAS, they utilize a broader

range of building blocks and can form various chain lengths. Further, the complete reduction of each incorporated monomer unit applied in FAS is only optional in PKS. Those reductive steps can be partly or fully omitted before being handed to the next elongation module, which significantly increases the potential structural diversity.<sup>73</sup> For a full reduction of the initially formed  $\beta$ -keto group, a PKS module requires three additional domains. The ketoreductase domain (KR) forms a  $\beta$ -hydroxyl group which is dehydrated by a dehydratase domain (DH) to form a double bond. The double bond is further reduced by an enoylreductase domain (ER) to complete the full reduction of the  $\beta$ -keto group.<sup>87</sup>



**Figure 1-7** Principle of a chain elongation step by Claisen condensation during PK biosynthesis. The substrate is activated and loaded to the ACP domain by the AT domain. The downstream (methyl)malonyl-S-T is decarboxylated, generating a nucleophilic thioester enolate, which forms a C-C bond by Claisen condensation with the upstream acyl-S-T thioester catalyzed by the KS domain. KS =  $\beta$ -ketoacylsynthase domain; AT = acyltransferase domain; ACP = acyl carrier protein domain. The figure is adapted from Fischbach & Walsh.<sup>72</sup>

PKS can be classified into various types based on their architecture and functionality. Type I PKS are large multifunctional enzymes with linearly organized and covalently fused catalytic domains. Type II PKSs on the other hand are usually a system of dissociable, monofunctional enzymes.<sup>73</sup> Similar to type II PKS, type III PKS consist of dissociable, monofunctional enzymes. They were found in plants, bacteria, and fungi<sup>88–90</sup> and, unlike type II PKS, the proteins of type III PKS form homodimers and use thioester building blocks that are freely activated by CoA without the use of ACP domains.<sup>91</sup> PKS can furthermore be categorized as iterative or noniterative, depending on whether the KS domain catalyzes only one or multiple rounds of elongations.<sup>73</sup> Noniterative type I PKS, as NRPS, follow the principle of collinearity, *i.e.*, the number of extension cycles (and thus the number of building blocks in the structure) typically correlates with the number of modules in the PKS assembly line. Further, the degree of  $\beta$ -keto processing can be inferred from the presence of KR, DH, and ER domains in a module. The principle of collinearity enables the prediction of the polyketide structure based on the

assembly line architecture and vice versa, and further allows rational reprogramming of the polyketide biosynthesis by genetic engineering.<sup>73</sup> It is noteworthy, that some bacterial modular type I PKS do not follow this rule. Modules may be skipped or used more than once, which hinders both the prediction of produced scaffolds as well as reciprocal the conclusions about enzyme organization drawn from a known compound structure. This is particularly challenging for a subclass of PKS systems called *trans*-AT PKS, as the modules lack individual AT domains within the assembly line (*cis*). Here, the substrates are loaded by standalone AT domains (*trans*).<sup>73</sup>

### 1.3.3 Structural diversity in NRPS and PKS

The NPs produced by NRPS, PKS, and their combinations exhibit vast structural diversity. The observed variety of structures that can be synthesized stems from the selection of building blocks, in the case of NRPSs utilizing proteinogenic and nonproteinogenic amino acids as well as aryl acids.<sup>72</sup> The role of the monomer-activating A domains is pivotal for the substrate selectivity. Stachelhaus and coworkers were able to identify the protein sequence of the A domain binding pocket responsible for the substrate selectivity. They developed the so-called nonribosomal code, implying ground rules for the prediction of A domain specificity based on the protein sequence.<sup>92–94</sup> Later, this code was extended by a range of 8 Å to include important residues outside of the binding pocket.<sup>95</sup> With this knowledge, tools for the prediction of the A domain specificity were developed over the years, such as NRPSpredictor2<sup>96</sup> or NRPSsp<sup>97</sup>. This nonribosomal code cannot only be used to facilitate prediction of structures synthesized by NRPS BGCs, in can further be utilized to change A domain specificity. Manipulation of as little as one residue by mutation in the nonribosomal code by genetic engineering can switch the specificity and initiate production of novel derivatives.<sup>92,98,99</sup> The diversity of NRPS can thus be further enhanced by such combinatorial biosynthesis approaches.

The diversity of PKS is also immense utilizing a vast variety of starter units such as short-chain (branched) fatty acids, various alicyclic and aromatic acids, amino acids, and extender units derived from malonyl-coenzyme A (CoA).<sup>100</sup> The enormous diversity is further implemented by optional reduction steps, *i.e.*, partly, or fully omitting reductive steps can introduce hydroxyl groups, unsaturated double-bonds, or keto groups, by variations in carbon chain lengths, folding, and termination, as well as by post PKS tailoring steps, *e.g.*, glycosylation, acylation, alkylation, and oxidation to name only a few.<sup>100</sup> It is noteworthy, that

the diversity of polyketides is even more increased by the stereochemical configuration, *e.g.*, (*S*) or (*R*) hydroxyl groups catalyzed by KR domains, *cis* or *trans* double bonds implemented by DH domains, or (*R*) or (*S*) methyl branches generated by ER domains.<sup>73</sup>

### 1.3.4 Fatty acid synthases (FAS)

As previously mentioned, FAS are closely related and evolutionarily connected to PKS.<sup>86</sup> Their primary difference is the full reduction of the incorporated monomers in FAS.<sup>87</sup> FAS are found ubiquitously across all groups of organisms and their biosynthetic mechanism is highly conserved. However, similar to the PKS they can be divided in two types depending on the organization of their catalytic domains.<sup>86</sup> Type I FAS, as type I PKS, consist of large, multifunctional enzymes and are found in fungi and animals.<sup>101</sup> Type II FAS on the other hand, just like type II PKS, use dissociable, monofunctional enzymes and are found in archaea, bacteria, and plants.<sup>102</sup> The biosynthetic pathway of the FAS chain elongation is the same as described previously in PKS: typically using acetyl-CoA as the starter unit and malonyl-CoA as elongation monomers, the substrates are bound to a phosphopantetheinylated acyl carrier protein domain (ACP) by the acyltransferase domain (AT), and subsequently linked to the next monomer or chain intermediate by Claisen condensation catalyzed by the  $\beta$ -ketoacylsynthase domain (KS).<sup>86,102</sup> The incorporated  $\beta$ -keto group is now fully reduced by a NADPH (reduced form of nicotinamide adenine dinucleotide phosphate)-dependent  $\beta$ -ketoreductase domain (KR), a dehydratase domain (DH), and an enoylreductase domain (ER). Unlike in PKS, this full reduction cycle is mandatory in FAS. The saturated acyl product now serves as the starter substrate for the next elongation step, which is repeated until a certain chain length of the fatty acid is reached.<sup>103,104</sup> Finally, after completion of all rounds of chain elongation, the product is either released from the FAS by a TE domain as free fatty acid,<sup>101</sup> or it can be transferred into glycerophospholipids by a variety of acyltransferase systems.<sup>105</sup> While type I FAS usually only produce palmitate, type II FAS are able to produce an enormous diversity of structures utilized in the cellular metabolism due to the diffusible ACP intermediates that can divert into other biosynthetic pathways. Therefore, type II FAS can generate products of different chain length, unsaturated, iso-, and anteiso-branched-chain, and hydroxy fatty acids.<sup>102</sup>

The FAS-produced stearic acid (SA, 18:0) can be utilized as precursor for aerobic biosynthesis of polyunsaturated fatty acids (PUFAs), *e.g.*, arachidonic acid (AA),

eicosapentaenoic acid (EPA), or docosahexaenoic acid (DHA). This PUFA synthesis is catalyzed by a combination of alternating desaturation and elongation reactions in plants and animals.<sup>106,107</sup> However, these PUFAs can also be produced without the presence of fatty acid desaturases and elongases. Iterative type I FAS/PKS-like multi-enzyme complexes, so-called PUFA synthases, have been identified and characterized from several  $\gamma$ -Proteobacteria and exhibited EPA production *via* anaerobic *de novo* synthesis from acetyl-CoA and/or malonyl-CoA substrates and coenzyme NADPH.<sup>108–110</sup>

## 1.4 Heterologous expression of myxobacterial secondary metabolite pathways

Native myxobacterial strains are often difficult to grow under laboratory conditions and the production yield of the target compound remains often too low for further studies, despite significant efforts to optimize the growth conditions and increase production. Low genetic amenability of the native strains further prevents elucidation of the NP biosynthesis by targeted gene deletions or attempts to improve the production yield by genetic modifications, such as introduction of strong promoters<sup>111</sup> or optimization of expression levels of positive and negative regulators.<sup>112–114</sup> Those conditions make the heterologous expression of NP biosynthetic pathways in well-established heterologous hosts an indispensable tool for the discovery, production (optimization), engineering, and characterization of myxobacterial secondary metabolites.<sup>115–117</sup> Various techniques and tools for genetic engineering and molecular biology methods are already available for well-established heterologous hosts. The introduction of biosynthetic gene clusters (BGCs) into said hosts enables the use of these tools and methods to investigate and manipulate the pathways. This allows the elucidation of the NP biosynthesis, production of novel derivatives, or improvement of production yield for further studies, *e.g.*, activity and mode of action investigations. The enormous diversity of microbial NPs as a promising source for novel drug development is gaining increasing interest. However, an estimation that approximately 99.8% of microbes present in many environments are not culturable under laboratory conditions yet, highlights the importance of heterologous expression of NP BGCs to tap into that seemingly endless source of potential active compounds.<sup>115</sup>

Several aspects should be considered during the process of choosing a heterologous host for the expression of the desired BGC. The host must be capable of performing the required

posttranslational modifications for the expressed proteins that need to be transcribed, translated, and correctly folded after transferring the desired BGC. In the case of PKS and NRPS, the host must be able to activate the ACPs or PCPs, respectively, by phosphopantetheinylation.<sup>75</sup> Furthermore, the genomes of myxobacteria are known for their high GC content so the host additionally must be able to express GC-rich genes. In many cases, the desired compound for the heterologous production is antimicrobial or toxic. Therefore, self-resistance genes of the native producer need to be expressed as well, or in the case that self-resistance genes are not available, inducible promoters should be used for a controlled expression.<sup>115</sup> It is noteworthy, that rare codons could result in ribosomal stalling, leading to truncated or degraded proteins, however, it seems that this problem is less relevant in practice than initially believed in theory.<sup>111</sup> The well-established strain *Myxococcus xanthus* DK1622 has proven to be a well-suited host considering these prerequisites. A significant number of BGCs were already successfully expressed in this host,<sup>118</sup> including the recently published cystobactamids.<sup>71</sup>

For a successful integration and expression of the target BGC, it must be obtained from the native producer and adapted for the chosen heterologous host. Traditionally, large cosmid libraries were generated and screened for cosmids with the desired genes. In the last decade, *de novo* DNA synthesis has enabled the *in silico* design of a biosynthetic pathway which can then be produced by gene synthesis companies.<sup>119</sup> The genes usually have to be reorganized and regulatory elements like promoters and terminators have to be added to form regulated operons that facilitate expression in the heterologous host. Subsequently, the modified genes need to be reassembled on vectors that harbor appropriate genetic elements, such as a suitable integration and replication system, and selection markers to enable transformation into hosts for both cloning (*e.g.*, *E. coli*) and genome-integration and expression (*e.g.*, *M. xanthus* DK1622).<sup>116</sup> The assembly of BGCs on respective vectors was traditionally conducted by classical restrictive-hydrolysis and ligation-based methods. This extremely time consuming and laborious approach, however, initiated the development of faster and more versatile techniques.<sup>120</sup> Several advanced *in vitro* (*e.g.*, Gibson assembly<sup>121</sup> or Golden Gate cloning<sup>122</sup>) and *in vivo* (*e.g.*, transformation-associated recombination (TAR)<sup>123</sup> and Red/ET recombination<sup>124</sup>) techniques have been developed. Both *in vivo* methods are based on recombination of overlapping homologous regions. While TAR cloning is carried out in yeast,

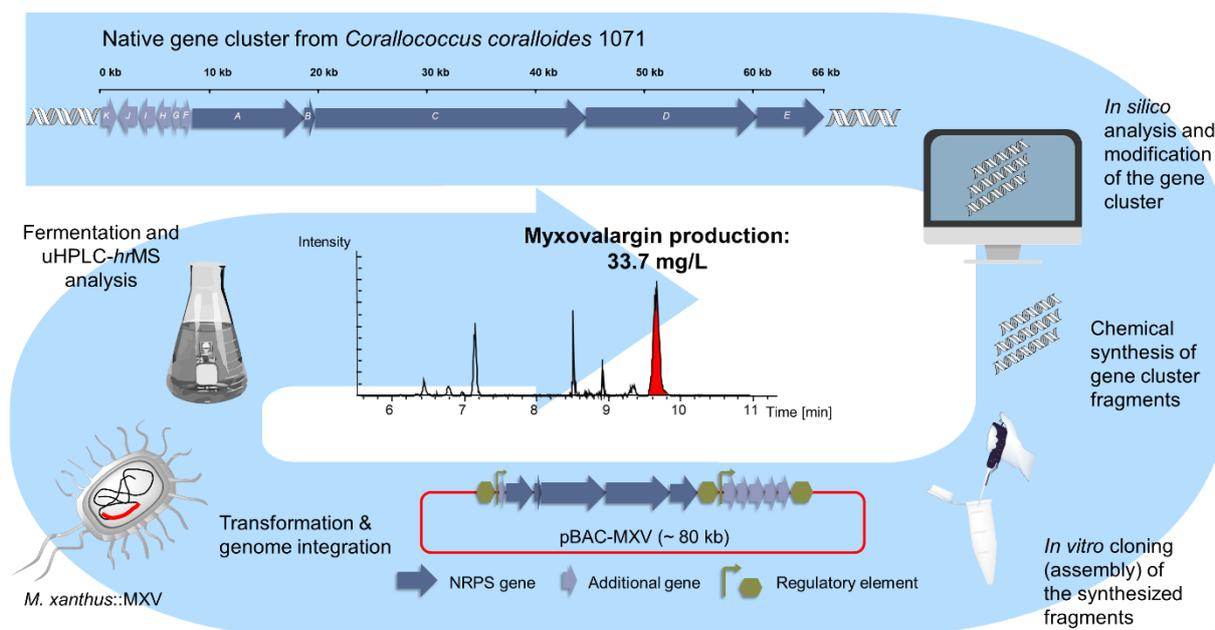
Red/ET recombination is performed in *E. coli* and facilitates straightforward plasmid modifications.

Notably, heterologous expression of NRPS and PKS BGCs and the development of sophisticated techniques to design, reorganize and assemble such pathways, enabled the emergence of the field of combinatorial biosynthesis. This field engages in the investigation of swapping entire modules, domains, or subunits of NRPS and PKS BGCs.<sup>125</sup> One prominent example of this field is the swapping of an erythromycin AT domain with an equivalent of the rapamycin biosynthesis, which yielded 61 new analogues.<sup>126</sup>

## 1.5 Outline of the present work

The work described in this thesis focuses on the heterologous expression of three myxobacterial NPs with intriguing biological activities. The main objective for all three compounds was the development of a heterologous expression platform using the well-established surrogate host *M. xanthus* DK1622. Modified BGCs of these NPs, adapted for the expression and modification in the heterologous host, should enable genetic manipulation for the elucidation of the biosynthetic pathways and potentially the production of novel derivatives.

According to these objectives, the second chapter of this thesis describes the development of a heterologous expression platform for the linear peptide antibiotic myxovalargin (MXV), exhibiting anti-Gram-positive and anti-Gram-negative activity and found during a screening program in 1982 in the myxobacterium *Myxococcus fulvus* Mx f65.<sup>62</sup> The native MXV producer *Corallococcus coralloides* 1071, which was used for initial studies on the underlying biosynthesis, has limitations in amenability for genetic tools and engineering.<sup>127</sup> Objectives of the herein presented work were the design and assembly of an inducible synthetic MXV BGC for heterologous expression in *M. xanthus* DK1622, allowing seamless genetic manipulation of the BGC. A functional expression platform for MXV was supposed to be utilized for targeted gene deletion and/or inactivation to analyze the role of the putative  $\beta$ -hydroxylase MxvH and the single PCP domain MxvB, as head start of in-depth investigations of the biosynthetic pathway. Further, the goal was to increase the production of MXV in the heterologous production system to ensure sufficient supply of the compound.



**Figure 1-8** General workflow of the MXV project. The process started with the identification of the native gene cluster in *Coralloccoccus coralloides* 1071. After *in silico* analysis, modification and reorganization of the gene cluster, gene cluster fragments were synthesized *de novo* and assembled *in vitro*. The final assembled synthetic MXV BGC was transformed into *M. xanthus* DK1622 and integrated into the genome. Fermentation and uHPLC-hrMS analysis showed a maximum MXV production of 33.7 mg/L.

The third chapter describes the work on disciformycin (DIF), a 12-membered macrolide-glycoside antibiotic isolated from the myxobacterium *Pyxidicoccus fallax* AndGT8, exhibiting strong activity against methicillin- and vancomycin-resistant *Staphylococcus aureus* (MRSA/VRSA) strains while showing very low cytotoxicity.<sup>61</sup> The aim was to utilize a heterologous expression platform harboring the PKS genes *difBCDEFG*<sup>128</sup> to further investigate the biosynthesis of DIF and discover necessary tailoring enzymes.

Chapter four describes the work with maracen and maracin, two lipophilic carboxylic acids with promising activity against *P. aeruginosa* (personal communication with Dr. Jennifer Hermann), discovered in the two strains *Sorangium cellulosum* Soce1128 and Soce880 in 1998.<sup>63</sup> The maracen/maracin biosynthesis is only hypothetical to this date. Slow growth and natural resistance against almost all antibiotics of the native *S. cellulosum* producer strains, combined with low genetic amenability, drove the desire for a heterologous expression platform for the maracen/maracin BGC. The objective of this work was to design a new version of a synthetic maracen/maracin gene cluster, reorganized in regulated operons inducible by vanillate and exempted from native promoter and terminator sequences, for the heterologous expression in strain *M. xanthus* DK1622.

Taken together, these chapters highlight critical steps in the heterologous expression of diverse chemical classes across myxobacteria using a single heterologous host. The challenges faced during this work underlined the complexity of myxobacterial natural products and efforts required to understand their biosynthesis. Important steps achieved herein toward improving and understanding the production of these intriguing natural products lays the groundwork for future studies of these compounds and future work based on the heterologous expression systems.

## 2 Heterologous expression of a synthetic myxovalargin gene cluster

### Contributions

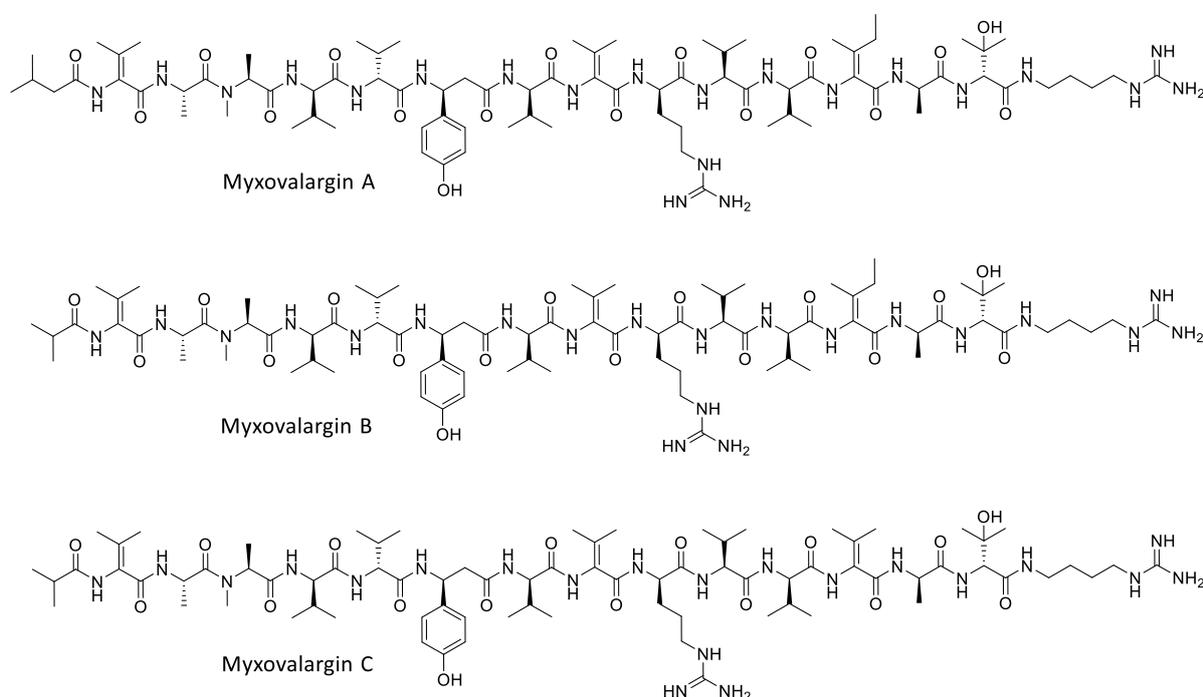
Assembly of the synthetic MXV gene cluster, transformation, and heterologous expression in *M. xanthus* DK1622 was done by the author of this work. Furthermore, design and cloning of the knockout of *mxvH* and gene inactivation of *mxvB* for biosynthesis elucidation, design of the ABC transporter cassette, as well as cultivation experiments if not mentioned otherwise were conducted by the author. *In silico* design of the synthetic MXV gene cluster was done by Dr. Sebastian Groß. Cloning work for the deletion of *mxvH* and inactivation of *mxvB*, as well as cultivations and analysis of production kinetics and multiple induction with vanillate was done with the help of Fabienne Wittling as part of her master thesis. Cloning work to obtain the ABC transporter cassette was done by Dr. Kamal Tehrani. MIC testing was done by Dr. Kamal Tehrani and Fabienne Wittling. The design and experimental procedure for the SNAc synthesis was performed by Dr. Alexander Kiefer.

## 2.1 Abstract

Myxovalargin (MXV) is a linear peptide antibiotic with activity against Gram-positive (MIC 0.3 ~ 5 µg/mL) and Gram-negative (MIC 6 ~ 100 µg/mL) bacteria, found during a screening program from the Gesellschaft für Biotechnologische Forschung (GBF) in 1982.<sup>62</sup> The native MXV producer *Coralloccoccus coralloides* 1071, which was used for initial studies on the underlying biosynthesis, has limitations regarding amenability for genetic tools which presents a hurdle for pathway engineering. Therefore, a heterologous expression platform was highly desirable to gain insights into the biosynthesis of the myxovalargins and allow biosynthetic pathway engineering for the production of new derivatives. Objectives of the work presented here were the design and assembly of an inducible synthetic MXV BGC for heterologous expression in the myxobacterial model strain *Myxococcus xanthus* DK1622. Successful assembly of the synthetic BGC, transformation, and heterologous expression in the DK1622 host led to production of the major derivative MXV A with a maximum yield of 33 mg/L. The first enzymes that were targeted for analysis of their function were the putative  $\beta$ -hydroxylase MxvH, and the single PCP domain MxvB. We assume that MxvH is involved in the formation of the  $\alpha,\beta$ -dehydro- and hydroxy-amino acids. Deletion of *mxvH* abolished the production of MXV and further *in vitro* analysis of expressed genes aims to confirm this hypothesis. MxvB, assumed to participate in the starter unit supply, was inactivated by a single point mutation of the conserved active serine residue. Production analysis is currently in progress and expected to abolish the production of MXV as well. Furthermore, the heterologous expression platform was used for attempts to increase the production yield of MXV and the self-resistance of the DK1622 host was improved from a MIC of 16 mg/L to over 256 mg/L by co-expressing the two MXV ABC transporter proteins MxvI and MxvK. Ongoing studies and cultivation experiments aim to significantly increase the production of MXV A.

## 2.2 Introduction

Myxovalargin (MXV) was initially found during a screening program to find new antibiotics in myxobacteria from the Gesellschaft für Biotechnologische Forschung (GBF) in 1982. The MXV producing strain *Myxococcus fulvus* Mx f65 was isolated from a soil sample collected from the Kaiserstuhl mountains in Germany in 1969.<sup>62</sup> The bioactive compound in the supernatant of *M. fulvus* Mx f65, which showed activity against Gram-positive bacteria (MIC 0.3 ~ 5 µg/mL) and Gram-negative bacteria (MIC 6 ~ 100 µg/mL), was linked to the four closely related peptide antibiotics MXV A, B, C and D, whereas MXV A is the most abundant and active derivative.<sup>62</sup>



**Figure 2-1** Myxovalargin derivatives A, B and C. No structure is available for myxovalargin D.

No inhibitory activity against mold and yeast was detected for any of the MXV derivatives, whereas these antibiotics revealed a relatively significant toxicity in mice (s.c.) with a LD<sub>50</sub> of 10 mg/kg and LD<sub>100</sub> of 30 mg/kg. The native strain *M. fulvus* Mx f65 was found to produce 4 – 6 mg/L MXV A with a MIC of 6 mg/L against the produced antibiotic. Later obtained mutants with a higher MIC (25 – 50 mg/L) were able to produce 30 mg/L in shake flask cultivation and 13 mg/L in bioreactor fermentation.<sup>62</sup> The antibiotic activity of MXV A was attributed to two modes of action by Irschik *et al.* At lower concentrations (1 µg/mL) MXV A inhibits bacterial protein synthesis. *In vitro* experiments suggested that MXV interferes with the binding of aminoacyl-tRNA to the A-site of ribosomes.<sup>129</sup> Concentrations above 5 µg/mL or prolonged

incubation led to destruction of cell membranes which entails secondary effects like decrease in O<sub>2</sub> consumption or instant breakdown of RNA synthesis. The effects on cell membrane were also shown in higher organisms, which could explain the toxicity of MXV in mice.<sup>129</sup> In a recent study, cryo-EM structure confirmed the ribosome as molecular target and revealed binding of MXV which completely occludes the exit tunnel (submitted manuscript)<sup>130</sup>. This study further showed potent antitubercular activity with a MIC of 0.2 µg/mL, rendering MXV a promising candidate for tuberculosis treatment.

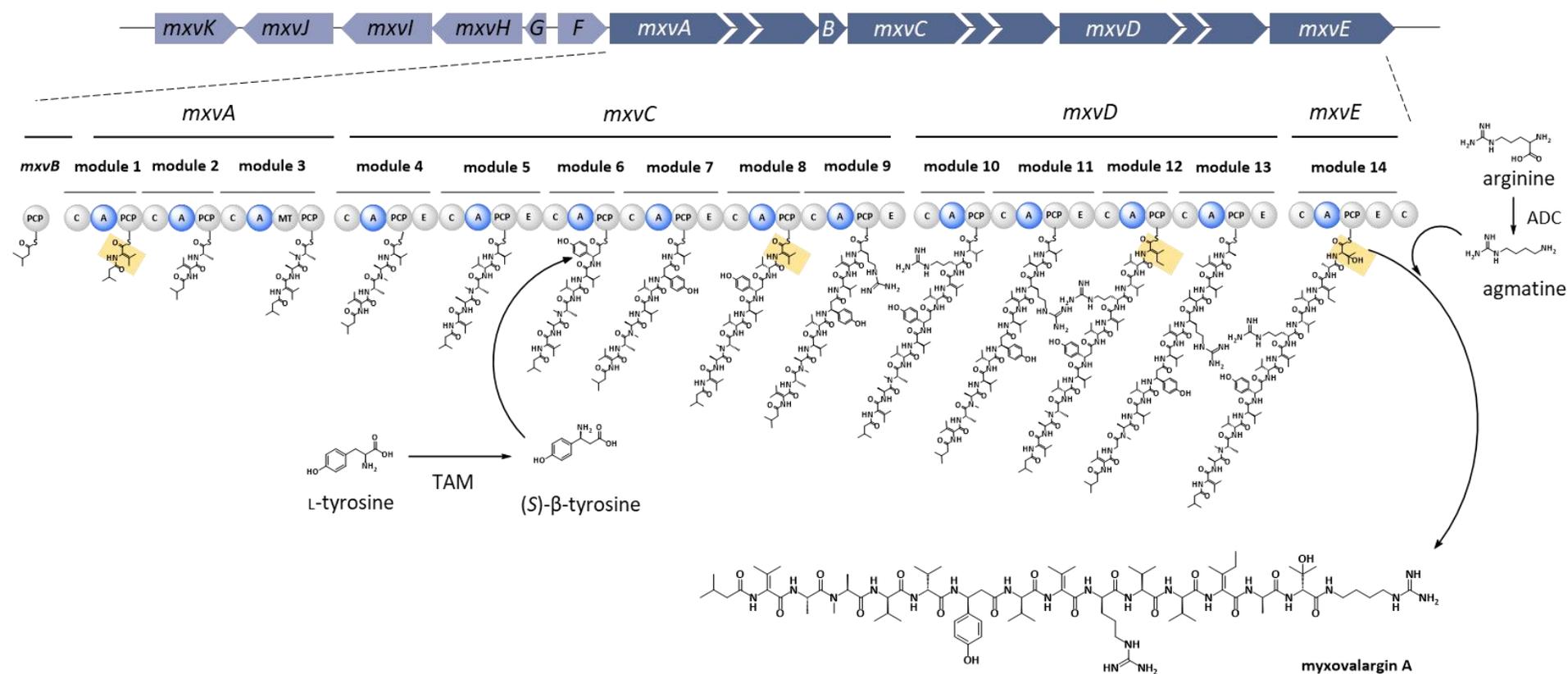
MXV A is a linear peptide antibiotic with a molecular weight of 1676 Da, consisting of 14 amino acids, a 3-methylbutyric acid (isovaleric acid) as starter unit at the *N*-terminal end, and agmatine at the *C*-terminal end. Besides the eight proteinogenic amino acids (or enantiomers thereof) (four *D*-valine, one *L*-valine, *L*-alanine, *D*-alanine and *D*-arginine) there are six unusual amino acids, *N*-methyl alanine,  $\beta$ -tyrosine,  $\beta$ -hydroxyvaline, two  $\alpha,\beta$ -dehydrovalines and one  $\alpha,\beta$ -dehydroisoleucine.<sup>129</sup> The structure elucidation of the MXVs pursued by Steinmetz *et al.* in 1987, identified the components of MXV A by acidic total hydrolysis and H<sup>1</sup> and C<sup>13</sup> nuclear magnetic resonance (NMR) spectroscopy.<sup>131</sup> The MXV biosynthetic gene cluster (BGC) was determined by genome sequence analysis of the known MXV producer strains *Myxococcus fulvus* Mx f65 [MCy8286], *Coralloccoccus coralloides* 1071 (Ccc1071) [MCy6431], *Angiococcus* sp. 983 (Ang983) [MCy5730], *Myxococcus xanthus* 113 (Mxx113) [MCy3592], *Myxococcus xanthus* 136 (Mxx136) [MCy3578], and *Myxococcus* sp. 171 [MCy9171] with bioinformatic tools and *in vivo* single-crossover-based gene disruption experiments.<sup>127,130</sup> Genomic analysis and feeding experiments lead to a structural revision, revealing incorporation of a *D*-valine at module 7 and *L*-valine at module 10, which was further confirmed by total synthesis (submitted manuscript).<sup>130</sup> The MXV BGC comprises a region of approximately 66 kbp encoding the genes which express the biosynthetic machinery responsible for the production of the compound. The 58 kbp core region consists of five genes (*mxvA – E*), encoding the 14-modular NRPS machinery and an upstream region with additional six genes (*mxvF – K*), encoding one tailoring enzyme, two ABC transporter proteins, a type II thioesterase, a MbtH-like protein, and a tyrosine aminomutase.<sup>127</sup> The eleven genes of the MXV BGC, including their individual sequence length and predicted function, are mentioned in Table 2-1 below. A detailed table about the predicted domain specificity of each NRPS module is shown in Table 2-16 in the SI section 2.6.2.

**Table 2-1** Genes of the MXV BGC including their size and putative function. *mxvA* – *mxvE* encoding the 14 NRPS modules while *mxvF* – *mxvK* encode additionally involved enzymes.

Gene	Size [kbp]	Function of homologue / putative function	Accession no. of the homologue	Coverage/Identity [%]
<i>mxvA</i>	10.7	NRPS (modules 1 – 3)	WP_203404054.1	88/46
<i>mxvB</i>	0.5	NRPS (single PCP domain of unknown function)	MCC6603015.1	97/31
<i>mxvC</i>	25.0	NRPS (modules 4 – 9)	WP_159029549.1	99/38
<i>mxvD</i>	15.7	NRPS (modules 10 – 13)	PSK18891.1	100/39
<i>mxvE</i>	5.9	NRPS (modules 14 + terminal C domain)	WP_233097884.1	98/38
<i>mxvF</i>	0.9	Type II thioesterase	MBV9786867.1	85/50
<i>mxvG</i>	0.2	MbtH-like protein	HID38629.1	84/77
<i>mxvH</i>	1.6	Putative $\beta$ -hydroxylase	WP_210613556.1	96/48
<i>mxvI</i>	1.8	ABC transporter permease	RMG93330.1	94/54
<i>mxvJ</i>	1.6	Tyrosine aminomutase <sup>132</sup>	B8ZV93.1	97/94
<i>mxvK</i>	1.5	ABC transporter ATPase	WP_206717488.1	98/58

The architecture of the native MXV BGC of producer strain *Coralloccoccus coralloides* 1071 is shown in Figure 2-2. The core structure of the MXV molecule is assembled by progressive incorporation of the 14 amino acids throughout the NRPS assembly line. The 3-methylbutyric acid starter unit is supplied *via* isovaleryl-CoA, an intermediate from the branched chain amino acid metabolism. It is assumed to be incorporated by a stand-alone PCP-domain expressed by *mxvB*. This only approx. 600 bp long gene *mxvB* is highly conserved in the BGCs between *mxvA* and *mxvC* of all MXV producer strains, underlining its potential importance: The biosynthetic role of the gene *mxvB* was up to date not verified.<sup>127</sup> The terminal agmatine unit is generated from L-arginine by the arginine decarboxylase (ADC).<sup>133,134</sup> The incorporation of this agmatine unit is presumably catalyzed by the terminal C domain of the BGC. Since there is no TEI domain present it is assumed, that the pre-MXV molecule is cleaved from module 14 by the incorporation of agmatine *via* the terminal C domain. The cyanobacterial peptide aeruginoside 126B isolated from *Planktothrix agardhii* CYA126/8 is the only other known compound, which

contains a terminal agmatine moiety. The NRPS assembly line of this natural product harbors a terminal C and PCP domain.<sup>135</sup> In addition to the proteinogenic amino acids or enantiomers thereof, MXV comprises several unusual amino acids.  $\beta$ -tyrosine formation is catalyzed by the tyrosine aminomutase (TAM) from natural L-tyrosine.<sup>132</sup> Further unusual amino acids are the two  $\alpha,\beta$ -dehydrovalines, one  $\alpha,\beta$ -dehydroisoleucine and one  $\beta$ -hydroxyvaline. It is assumed that the putative  $\beta$ -hydroxylase encoded by *mxvH* is involved in their formation. This  $\beta$ -hydroxylase shows similarity to the  $\beta$ -hydroxylase from chloramphenicol synthesis CmLA, which hydroxylates 4-aminophenylalanine in  $\beta$ -position.<sup>136</sup> It was postulated that the PCP-bound L-valine and L-isoleucine are hydroxylated by MxvH, however, the subsequent dehydration of  $\beta$ -hydroxy-L-valine and -L-isoleucine has yet to be elucidated.<sup>127</sup> The proposed biosynthesis is summarized in Figure 2-2.



**Figure 2-2** The architecture of the native MXV BGC of producer strain *C. coralloides* 1071 and the proposed MXV biosynthesis. The five NRPS genes *mxvABCDE* constitute most of the 66 kbp long MXV BGC. *MxvB* is assumed to incorporate the starter unit (isovaleryl in case for MXV A). Module 6 incorporates (S)-β-tyrosine, which is derived from L-tyrosine by a tyrosine aminomutase (TAM/*MxvJ*). The α,β-dehydro-amino acids and the hydroxy valine (orange highlighted squares) are suggested to be formed by *MxvH*. The final agmatine residue is derived from arginine by an arginine decarboxylase (ADC) and hypothesized to be incorporated by the terminal C domain, through which the molecule gets released from the assembly line. PCP = peptidyl carrier protein; C = condensation domain; A = adenylation domain; MT = methyltransferase domain; E = epimerization domain.

## 2.3 Material and methods

### 2.3.1 Media

All media used for liquid cultivations and agar plates were autoclaved for 20 minutes at 121°C before use and kept sterile.

**Table 2-2** Media used in this work with the respective composition

Media	Composition
<b>Lysogeny Broth (LB)</b>	Tryptone 10.0 g/L, yeast extract 5.0 g/L, NaCl 5.0 g/L, agar (for solid media) 15.0 g/L, dH <sub>2</sub> O 1.0 L
<b>2xYT</b>	Tryptone 16.0 g/L, yeast extract 10 g/L, NaCl 5.0 g/L, dH <sub>2</sub> O 1.0 L
<b>CTT</b>	Casitone 10.0 g/L, 1 M Tris-Cl pH 8.0 10.0 mL/L, 1 M K <sub>2</sub> HPO <sub>4</sub> pH 7.6 1.0 mL/L, 1 M MgSO <sub>4</sub> 8.0 mL/L, agar (for solid media) 15.0 g/L and for soft agar 7.5 g/L, dH <sub>2</sub> O 1.0 L
<b>CYE</b>	Casitone 10.0 g/L, yeast extract 5.0 g/L, MgSO <sub>4</sub> 8.0 mM/L, Tris-HCl pH 7.6 10.0 mM/L, dH <sub>2</sub> O 1.0 L
<b>M7/s6</b>	Soy flour 5.0 g/L, corn starch 5.0 g/L, glucose 2.0 g/L, MgSO <sub>4</sub> x 7H <sub>2</sub> O 1.0 g/L, CaCl <sub>2</sub> x 2 H <sub>2</sub> O 1.0 g/L, HEPES 10 g/L, vitamin B12 0.1 mg/L*, FeCl <sub>3</sub> 5 mg/L*, potassium acetate 10 g/L  Adjust pH to 7.4 with NaOH
<b>M7/s4</b>	Soy flour 5.0 g/L, corn starch 5.0 g/L, glucose 2.0 g/L, MgSO <sub>4</sub> x 7H <sub>2</sub> O 1.0 g/L, CaCl <sub>2</sub> x 2 H <sub>2</sub> O 1.0 g/L, HEPES 10 g/L, vitamin B12 0.1 mg/L*, FeCl <sub>3</sub> 5 mg/L*  Adjust pH to 7.4 with NaOH

\*add to culture after autoclaving

### 2.3.2 Antibiotic stock solutions

Several antibiotics were used in liquid media or agar plates for the selection of genetically altered *E. coli* and *M. xanthus* strains: Ampicillin<sup>[a]</sup> (Amp), Chloramphenicol<sup>[a]</sup> (Cm), Zeocin (Zeo), Oxytetracycline (Otc), and Kanamycin (Kan). Table 2-3 depicts the antibiotic concentrations used in this work.

**Table 2-3** Antibiotics used in this work with their stock and working concentration.

Antibiotic	Stock (1,000x) [mg/mL]		Working conc. [ $\mu\text{g/mL}$ ] <sup>[b]</sup>
<b>Ampicillin (Amp)</b>	100	1 g in 10 mL MQ-H <sub>2</sub> O	100
<b>Chloramphenicol (Cm)</b>	25	0.25 g in 10 mL 70 % abs. EtOH	25
<b>Kanamycin (Kan)</b>	50	0.5 g in 10 mL MQ-H <sub>2</sub> O	50
<b>Zeocin (Zeo)<sup>[c]</sup></b>	100		50
<b>Oxytetracycline (Otc)</b>	10	0.1 g in 10 mL 70 % abs. EtOH	10
<b>Spectinomycin (Spec)</b>	100	Dissolved in MQ-H <sub>2</sub> O	100

<sup>[a]</sup> Amp and Cm are not applicable in *M. xanthus* <sup>[b]</sup> For high-copy vectors; low-copy vectors ½ of normal working conc., <sup>[c]</sup> Zeocin is purchased by ThermoFisher Scientific in pre-made ready to use stocks.

Antibiotic stock solutions were stored at -20°C. Oxytetracycline and Zeocin are light sensitive, thus the stock solutions were protected from light.

### 2.3.3 Reagents and buffer

Solutions were prepared with demineralized water (dH<sub>2</sub>O) or ultrapure water (MQ-H<sub>2</sub>O). To obtain MQ-H<sub>2</sub>O, 'Milli-Q Reference A+ System' (Millipore, Merck) was used.

For agarose gel electrophoresis 50x TAE buffer containing TRIS 242 g/L, glacial acetic acid 57.1 mL/L, EDTA (500 mM in dH<sub>2</sub>O pH 8.0) 100 mL/L and dH<sub>2</sub>O *ad* 1 L was used. The TAE buffer (50x) stock solution was diluted with dH<sub>2</sub>O to the 1x working concentration before use. Agarose gel solution was prepared by adding 5 g agarose in 500 mL TAE buffer (1x) for 1% solution. For separation of larger DNA fragments 0.8% agarose gel solution was used (4 g agarose in 500 mL 1x TAE buffer). Agarose was dissolved by microwave heating. The solution was stored at 55 °C to prevent solidification.

Orange G loading dye (10x) was used to stain DNA for visualization in agarose gels on a Transilluminator. The solution contains orange G 10 mg, glycerin 3 mL and TE buffer 2 mL. TE buffer consists of TRIS (0.5 M in dH<sub>2</sub>O) 20 mL, EDTA (0.5 M in dH<sub>2</sub>O, pH 8.0) 2 mL and dH<sub>2</sub>O *ad* 1 L. The 10x loading dye was diluted to 1x in the DNA solution that had to be stained.

Buffers P1, P2, and P3 were used for alkaline lysis and plasmid DNA preparation. The composition of each is shown in the table below.

**Table 2-4** Composition of resuspension buffer P1, lysis buffer P2, and neutralization buffer P3. Buffers were used for alkaline lysis and plasmid DNA preparation.

<b>Resuspension buffer P1</b> Stored at 4° C	TRIS (1 M in dH <sub>2</sub> O) 2.5 mL, EDTA (0.1 M, pH 8.0) 5 mL, Ribonuclease A 25 mg, MQ-H <sub>2</sub> O <i>ad</i> 50 mL
<b>Lysis buffer P2</b>	NaOH (1 M in dH <sub>2</sub> O) 10 mL, SDS (20 % in dH <sub>2</sub> O) 2.5 mL, MQ-H <sub>2</sub> O <i>ad</i> 50 mL
<b>Neutralization buffer P3</b>	K-acetate 14.72 g, MQ-H <sub>2</sub> O <i>ad</i> 50 mL, pH 5.5

### 2.3.4 Software

*Geneious 10.1.3* (Biomatters) was used to analyze native biosynthetic gene clusters, to design synthetic gene clusters, to design oligonucleotides for PCR experiments and sequencing, and to design plasmid maps. Agarose gel pictures from the fluorescence chamber were analyzed with *FusionCapt Advance* Software (Vilber Lourmat). *ChemDraw Ultra 12.0* (CambridgeSoft) was used to create images and chemical formulae. LC chromatograms as well as UV spectra and mass spectrograms were analyzed with *DataAnalysis 4.4* (Bruker Daltonics, Billerica, MA, USA).

### 2.3.5 *E. coli* cultivation

*E. coli* DH10 $\beta$ , NEB 10 $\beta$  or HS996 were used for standard cloning procedures. Except for electroporation for transformation of plasmids, working with *E. coli* was performed under sterile conditions (HeraSafe, Heraeus). For storage at -80°C, 750  $\mu$ L over-night culture was combined with 750  $\mu$ L 50 % glycerin (in MQ-H<sub>2</sub>O) and transferred into a cryo tube.

Cultivation took place in 2xYT or LB medium. Supplementation with antibiotics for selection depended on the plasmid resistance marker. Antibiotic concentrations are shown in Table 2-3. For clone selection, cultures were spread on LB agar (25 mL in a petri dish) after transformation and incubated overnight at 37°C in an incubator. Liquid cultures (5 mL in glass vial or 10-20% v/v liquid media in Erlenmeyer flasks) were incubated overnight at 37°C in a rotary shaker (Multitron, Infors HT) at 200 rpm. Larger constructs of 30 kbp size or longer, were incubated at 30°C for a longer period of time (up to two days of incubation). *E. coli* clones

containing pBeloBAC vector backbone and/or the whole MXV cluster were always cultured at 30° C due to the enormous size of the plasmid.

### **2.3.6 Preparation of *E. coli* electro competent cells**

For preparation of electro competent *E. coli* cells, 5 mL LB medium were inoculated with the respective strain directly from cryo stock and incubated at 37 °C and 200 rpm overnight. 2 mL of overnight culture were used to inoculate 200 mL LB. Cultivation was carried out under the same conditions until an OD<sub>600</sub> (Biophotometer plus, Eppendorf) between 0.4 and 0.6 was reached. The cultures were cooled on ice for 30 min and afterwards transferred into four 50 mL Falcon tubes. All centrifugation steps were performed at 4,000 x g for 10 min at 4 °C (Avanti J-26 XP Centrifuge, Beckman Coulter; Rotor: JLA 10.500), followed by discarding the supernatant. First time, the cell pellets were resuspended in 100 mL (25 mL each) ice cold HEPES (1 mM in dH<sub>2</sub>O; pH7.0), followed by centrifugation. Second time, the cell pellets were resuspended in 50 mL (12.5 mL each) ice cold HEPES (1 mM in dH<sub>2</sub>O; pH 7.0) and combined in one falcon tube before centrifugation. Third time, the cell pellet was resuspended in 50 mL ice cold 10 % glycerin in HEPES (1 mM in dH<sub>2</sub>O; pH 7.0), followed by centrifugation. Finally, the cell pellet was resuspended in 2 mL 10 % glycerin in HEPES (1 mM in dH<sub>2</sub>O; pH7.0) and portioned into 50 µL aliquots. Now the prepared cells were frozen in liquid N<sub>2</sub> and stored at -80°C.

### **2.3.7 Cultivation of *M. xanthus***

*M. xanthus* strains were cultivated either for transformation of heterologous gene constructs or for testing the production of mutant strains. All cultures were incubated at 30°C, either in an incubator (plates) or on shakers at 200 rpm. Production cultures were grown in 25 mL liquid media with corresponding selection marker until sufficient cell density was reached. Then the culture was sub-cultured into 50 mL liquid medium (10% inoculum). After 24h of incubation, vanillate was added to induce the expression of heterologous genes, in case a vanillate promoter was used for induction of a gene cluster. Further, XAD-16 was added to bind produced compounds. The cell cultures were harvested after the culture broth took on a brown color (if not mentioned otherwise).

### 2.3.8 Extraction and analysis of MXV

Culture was centrifuged for 20 min at 8000 rpm, supernatant discarded, and cell/XAD-16 pellet stored at -20°C for a couple of hours/until frozen and then lyophilized overnight. The pellet was then extracted twice with 30 mL MeOH for 20 minutes. After extraction the liquid was filtered and the eluate caught in a round flask. The solvent was then evaporated at a Rotavapor until the extract was completely dry. Now the extract was re-dissolved in 1 mL MeOH and pipetted into a glass vial. Before analysis, sample was centrifuged for at least 10 min at 15000 rpm and 10  $\mu$ L of the supernatant was 10 times diluted in MeOH.

For detection of compound UHPLC-hrMS analysis was performed on a Dionex UltiMate 3000 rapid separation liquid chromatography (RSLC) system (Thermo Fisher Scientific, Waltham, MA, USA) coupled to a Bruker maXis 4G ultra-high-resolution quadrupole time-of-flight (UHR-qTOF) MS equipped with a high-resolution electrospray ionization (HRESI) source (Bruker Daltonics, Billerica, MA, USA). The separation of a 1  $\mu$ L sample was achieved with a linear 5–95% gradient of acetonitrile with 0.1% formic acid in ddH<sub>2</sub>O with 0.1% formic acid on an ACQUITY BEH C18 column (100 mm  $\times$  2.1 mm, 1.7  $\mu$ m dp) (Waters, Eschborn, Germany) equipped with a Waters VanGuard BEH C18 1.7  $\mu$ m guard column at a flow rate of 0.6 mL/min and 45 °C for 18 min with detection by a diode array detector at 200–600 nm. The LC flow was split into 75  $\mu$ L/min before entering the mass spectrometer. Mass spectrograms were acquired in centroid mode ranging from 150–2500 m/z at an acquisition rate of 2 Hz in positive MS mode. Source parameters were set to 500 V end-plate offset; 4000 V capillary voltage; 1 bar nebulizer gas pressure; 5 L/min dry gas flow; and 200 °C dry gas temperature. Ion transfer and quadrupole parameters were set to 350 VPP funnel RF; 400 VPP multipole RF; 5 eV ion energy; and 120 m/z low-mass cut-off. Collision cell was set to 5.0 eV and pre-pulse storage was set to 5  $\mu$ s. Calibration was conducted automatically before every HPLC-MS run by the injection of sodium formate and calibration on the respective clusters formed in the ESI source. All MS analyses were acquired in the presence of the lock masses C<sub>12</sub>H<sub>19</sub>F<sub>12</sub>N<sub>3</sub>O<sub>6</sub>P<sub>3</sub>, C<sub>18</sub>H<sub>19</sub>F<sub>24</sub>N<sub>3</sub>O<sub>6</sub>P<sub>3</sub> and C<sub>24</sub>H<sub>19</sub>F<sub>36</sub>N<sub>3</sub>O<sub>6</sub>P<sub>3</sub>, which generate the [M + H]<sup>+</sup> ions of 622.0289, 922.0098 and 1221.9906. The HPLC-MS system was operated by HyStar 5.1 (Bruker Daltonics, Billerica, MA, USA), and LC chromatograms as well as UV spectra and mass spectrograms were analyzed with DataAnalysis 4.4 (Bruker Daltonics, Billerica, MA, USA).

For quantification of MXV yield an UltiMate 3000 LC System with a Acquity UPLC BEH C-18 column (1.7  $\mu\text{m}$ , 100 x 2 mm) was used, equipped with a VanGuard BEH C-18 (1.7  $\mu\text{m}$ ) guard column, was coupled to an Apollo II ESI source and hyphenated to an amaZon speed 3D ion trap mass spectrometer. Separation was performed at a flow rate of 0.6 mL/min (eluent A: deionised H<sub>2</sub>O + 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  $\mu\text{L}/\text{min}$  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 in positive mode. Source parameters were set to 500 V end plate offset, 4000 V capillary voltage, 1 bar nebulizer gas pressure, 5 L/min dry gas flow and 200 °C dry temperature. Mass spectrograms were analyzed with DataAnalysis 4.4 (Bruker Daltonics, Billerica, MA, USA). Calibration curve was set up with purified MXV A dissolved in MeOH. The trend line of the calibration curve was used to calculate the MXV A concentrations of the samples based on the peak area.

### **2.3.9 Plasmid DNA isolation from *E. coli***

#### **2.3.9.1 “Mini prep”**

For plasmid DNA isolation from *E. coli*, 2 mL overnight culture was centrifuged for 1 min at 15000 rpm. After discarding the supernatant, the cell pellet was resuspended in 250  $\mu\text{L}$  P1 buffer using a bench top shaker. Cell lysis occurred after adding 250  $\mu\text{L}$  P2 buffer and inverting the tubes seven times, and incubation for 3 min. By addition of 250  $\mu\text{L}$  P3 buffer, 10  $\mu\text{L}$  chloroform and inverting seven times, the lysis was stopped. After centrifugation for 10 min at 15000 rpm, the supernatant was transferred into a new 1.5 mL Eppendorf tube containing 500  $\mu\text{L}$  ice cold isopropanol. Another centrifugation step for 5 min at 15000 rpm to precipitate the plasmid DNA was followed by discarding the supernatant and adding 700  $\mu\text{L}$  70 % ethanol (in MQ-H<sub>2</sub>O) for washing. Finally, after centrifugation for 1 min at 15000 rpm, the supernatant was discarded, and the pellet dried before resuspension in 50 - 100  $\mu\text{L}$  MQ-H<sub>2</sub>O.

#### **2.3.9.2 “Midi prep”**

In case larger quantities of isolated plasmid were required for subsequent cloning procedures, 200 mL of *E. coli* overnight culture were grown and harvested. The plasmid was

then isolated from the cells using the Nucleo Bond PC100 (midi-prep) DNA purification kit by Macherey-Nagel according to their protocol.

### **2.3.9.3 Adaptations for working with pBeloBACMXV plasmid**

Since only one copy of plasmid per cell is available, significantly more DNA has to be provided for cloning procedures. Thus, the respective *E. coli* strain containing the pBeloBACMXV plasmid was cultured in 400 mL LB medium for at least 24 hours. The plasmid was then extracted from the cells using the Nucleo Bond PC100 (midi-prep) DNA purification kit by Macherey-Nagel according to their protocol.

### **2.3.10 Polymerase chain reaction (PCR)**

PCR protocols that were commonly used during this work are described below. Depending on the experiment, template size or objective of the PCR, the protocols might slightly differ from below mentioned parameters.

#### **2.3.10.1 Phusion polymerase**

Reaction mix: GC buffer 4.0  $\mu\text{L}$ , dNTPs 5 mM (1.25 mM each) 4.0  $\mu\text{L}$ , Phusion polymerase (Thermo Fisher Scientific) 0.2  $\mu\text{L}$ , primer forward (10  $\mu\text{M}$ ) 0.5  $\mu\text{L}$ , primer reverse (10  $\mu\text{M}$ ) 0.5  $\mu\text{L}$ , DNA template 2  $\mu\text{L}$ \*, MQ-H<sub>2</sub>O 8.8  $\mu\text{L}$ , DMSO 1.0  $\mu\text{L}$ .

\*2  $\mu\text{L}$  genomic DNA, for plasmid DNA 1  $\mu\text{L}$  with a concentration of approximately 10 ng/ $\mu\text{L}$  was used. Volume of MQ-H<sub>2</sub>O had to be adapted accordingly.

Cycle protocol: Initial denaturation 5 min at 98°C; 34 cycles of denature (98°C 20 s), anneal (temperature depends on primers, 25 s), and elongation (72°C 15-30 s/kb); finale elongation at 72°C for 5 min.

#### **2.3.10.2 Taq polymerase**

Reaction mix: Taq buffer (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub> 2.5  $\mu\text{L}$ , dNTPs 5 mM (1.25 mM each) 4.0  $\mu\text{L}$ , primer forward (50  $\mu\text{M}$ ) 0.5  $\mu\text{L}$ , primer reverse (50  $\mu\text{M}$ ) 0.5  $\mu\text{L}$ , Taq polymerase (Thermo Fisher Scientific) 0.2  $\mu\text{L}$ , MgCl<sub>2</sub> (50 mM) 1  $\mu\text{L}$ , DNA template 2  $\mu\text{L}$ , glycerol 50% 4  $\mu\text{L}$ , MQ-H<sub>2</sub>O 5.3  $\mu\text{L}$ .

Cycle protocol: Initial denaturation 5 min at 95°C; 30 cycles of denature (95°C 30 s), anneal (temperature depends on primers, 30 s), and elongation (72°C 1 min/kb); finale elongation at 72°C for 5 min.

### **2.3.10.3 Q5 Polymerase**

Mix for two reactions: 5x Q5 buffer 10  $\mu$ L, dNTPs 5mM (1.25 mM each) 2  $\mu$ L, primer forward (50  $\mu$ M) 0.5  $\mu$ L, primer reverse (50  $\mu$ M) 0.5  $\mu$ L, DNA template 1  $\mu$ L, Q5 polymerase (New England BioLabs) 0.5  $\mu$ L, 5x GC buffer 10  $\mu$ L, MQ-H<sub>2</sub>O 25.5  $\mu$ L.

Cycle protocol: Initial denaturation 30 seconds at 98°C; 30 cycles of denature (98°C 10 s), anneal (temperature depends on primers, 30 s), and elongation (72°C 20-30 seconds/kb); finale elongation at 72°C for 2 min.

### **2.3.11 DNA separation and purification**

Agarose gel electrophoresis was used for separation of DNA fragments based on size. In this work, agarose gel electrophoresis was performed in a Consort EV231 chamber (Sigma-Aldrich). An 1% agarose gel was used with 10  $\mu$ L Roti Safe Gel Stain per 100 mL gel and TAE buffer (1x) as buffer system. After PCR experiments or hydrolyzations with restriction enzymes, OrangeG loading dye (10x) was added to the DNA sample (1/10 of total volume) and the gel was loaded immediately. GeneRuler 1 kb/1 kb plus/100 bp plus DNA ladder (Thermo Fisher Scientific) were used as size standards. The voltage was set between 1 - 5 V/cm until DNA fragments were sufficiently separated. DNA detection took place in a Fusion FX chamber with UV fluorescence detector (Vilber Lourmat).

For DNA purification, agarose gel was placed on a blue light transilluminator, desired DNA fragments were cut out of the agarose gel with a scalpel and transferred into an Eppendorf tube. The proper DNA purification was done with the Macherey-Nagel NucleoSpin Gel and PCR Clean-up Kit following the manufacturer information. Finally, purified DNA was eluted in 30  $\mu$ L of MQ-H<sub>2</sub>O/elution buffer (from NucleoSpin Gel and PCR Clean-up Kit). DNA fragments of larger size were purified using the peqGOLD Gel Extraction Kit (peqlab).

#### **2.3.11.1 Adaptations for working with pBeloBACMXV plasmid**

Due to the size of the plasmid of almost 80 kb DNA purification could not be done with the previous described Macherey-Nagel NucleoSpin kit anymore (cut-off from 50bp-20kbp). Purification was either performed with Agarose gel DNA extraction kit by Roche (using silica beads, with a maximum cut-off of up to 100kbp) or doing a simple precipitation of DNA (detailed protocol in the next paragraph).

### **2.3.11.2 DNA precipitation**

NaAcetate solution (pH 5.2) 15  $\mu$ L was added to 100  $\mu$ L digestion reaction and inverted 1-2 times. Then 300  $\mu$ L of 100% ice cold ethanol was added and inverted 2-3 times. The reaction was incubated at -20° C for 30-60 minutes and spun down at 4° C at 15.000 rpm for 20-30 minutes. The DNA was then washed with 70% ethanol and dried.

### **2.3.12 Determination of DNA concentration**

NanoDrop 2000c (Thermo Fisher Scientific) was used for determination of DNA concentrations of solutions. For that purpose, 1  $\mu$ L of DNA solution was put on the pedestal and measured at 260 - 280 nm wavelength. MQ-H<sub>2</sub>O served as a reference (blank).

Furthermore, the sample DNA concentration was determined visually by loading 1  $\mu$ L DNA sample on an agarose gel and comparing the DNA band intensity with the intensity of a DNA ladder band of similar size. 6  $\mu$ L of the DNA ladder were loaded as standard (using recommended standard concentrations).

### **2.3.13 Molecular cloning**

For conventional, preparative restriction/ligation reactions, both vector and insert DNA were treated with type II or type IIS (*Bsa*I) restriction endonucleases (REs). Thereby, digested DNA is linearized, meaning that vector and insert DNA with suitable ends can be ligated by T4 DNA ligase. In the following sections, restriction hydrolyzation of DNA will sometimes be referred to as 'digest' of DNA, which is colloquial use in daily lab environment. All restriction endonucleases, T4 DNA ligase, and FastAP were purchased from Thermo Fisher Scientific.

#### **2.3.13.1 Standard protocol for preparative restriction hydrolyzation of DNA**

Preparative digest was used to hydrolyze sufficient amount of DNA for follow up ligation of insert and vector.

**Table 2-5** Composition of standard preparative restriction hydrolyzation reactions for single digest (with one enzyme) and double digest (with two enzymes, respectively). DNA solution was usually obtained by DNA isolation (see section 2.3.9) or in some cases by PCR (see section 2.3.10).

Single digest	Double digest
25 µL plasmid DNA solution	24 µL plasmid DNA solution
3 µL 10x restriction endonuclease buffer	3 µL 10x restriction endonuclease buffer
2 µL restriction endonuclease	1.5 µL restriction endonuclease 1
-	1.5 µL restriction endonuclease 2

Reactions were incubated for several hours or overnight at the required temperature (for most enzymes 37°C). To prevent vector religation, FastAP enzyme was added after  $t_{1/2}$  of the reaction (1 µL in 30 µL reactions). This leads to dephosphorylation of the 5' ends of the linearized vector DNA and therefore improves the overall cloning efficacy.

### 2.3.13.2 Standard protocol for analytical restriction hydrolyzation of plasmids containing insert

Analytical digests were used to verify plasmid size and screen for clones harboring plasmids with the right insert after a cloning process.

**Table 2-6** Composition of standard analytical restriction hydrolyzation reactions for single digest (with one enzyme) and double digest (with two enzymes, respectively). DNA solution was usually obtained by DNA isolation (see section 2.3.9) or in some cases by PCR (see section 2.3.10).

Single digest	Double digest
8.5 µL plasmid DNA solution	8.0 µL plasmid DNA solution
1.0 µL 10x restriction endonuclease buffer	1.0 µL 10x restriction endonuclease buffer
0.5 µL restriction endonuclease	0.5 µL restriction endonuclease 1
-	0.5 µL restriction endonuclease 2

Reactions were incubated for approximately 2 hours at the required temperature.

### 2.3.13.3 Standard ligation reaction

Standard ligation protocol was used to assemble prior linearized DNA fragments. T4 DNA ligase catalyzes bond formation between a 5' phosphate group and a 3' hydroxy group and thereby leads to the ligation of vector and insert DNA. The composition of ligation reaction performed in this work is shown in Table 2-7.

**Table 2-7** Composition of standard ligation reaction, used to assemble linearized DNA

Component	Volume [ $\mu$ L]
Vector DNA	0.5 – 2.0 <sup>[a]</sup>
Insert DNA	6.5 – 8.0 <sup>[a]</sup>
T4 DNA ligase buffer	1.0
T4 DNA ligase	0.5

<sup>[a]</sup> Vector and insert DNA volume depended on determined DNA concentration and varied between 0.5-2  $\mu$ L for vector DNA solution and 6.5-8  $\mu$ L for insert DNA solution. Total reaction volume was always 10  $\mu$ L with a vector/insert DNA molar ratio of around 1/10.

Ligation reaction was incubated for 5 h at room temperature or overnight at 16°C.

#### 2.3.13.4 Adaptations for working with pBeloBACMXV plasmid

When working with the BAC backbone containing plasmid the reactions were always vortexed briefly instead of pipetting.

In order to prepare higher amounts of DNA for subsequent ligation and achieve a more efficient digestion of DNA, 100  $\mu$ L reactions were set up. Those reactions were incubated for 4-6 hours. Addition of restriction endonuclease enzymes was split into  $\frac{1}{2}$  of the volume at the start of the reaction and the second  $\frac{1}{2}$  after  $t\frac{1}{2}$  of the reaction.

**Table 2-8** Composition of 100  $\mu$ L digestion reactions for single and double digest

DNA	3-4 $\mu$ g
10x restriction endonuclease buffer	10 $\mu$ L
Restriction endonuclease	8 $\mu$ L <sup>[a]</sup>
FastAP	2 $\mu$ L <sup>[b]</sup>
H <sub>2</sub> O	<i>ad</i> 100 $\mu$ L

<sup>[a]</sup>Maximum volume of 8  $\mu$ L of enzyme was used, in case FastAP was needed too, the total enzyme volume (restriction endonuclease and FastAP) had a maximum of 10  $\mu$ L. For double digest reactions with two restriction endonucleases, 4  $\mu$ L of each enzyme was used. However, the ratio between enzymes 1 and 2 could vary depending on the buffer used. The buffer was chosen according to DoubleDigest Calculator by ThermoFisher Scientific.<sup>137</sup> In case enzyme 1 had only 50-100% efficiency in the chosen buffer, while enzyme 2 had 100%, a ratio of enzyme

1/enzyme 2 of 4.6 $\mu$ L/3.4 $\mu$ L was chosen. <sup>[b]</sup> FastAP was only added to vector DNA, to prevent vector religation. FastAP enzyme was added after  $t_{1/2}$  of the reaction.

Ligations were set up in 20  $\mu$ L reactions using 2  $\mu$ L T4 DNA ligase buffer and 1  $\mu$ L T4 DNA ligase. Vector and insert concentrations were tried in different ratios and MQ-H<sub>2</sub>O was added to a volume of 20  $\mu$ L.

### **2.3.14 *E. coli* transformation**

Transformation of *E. coli* was carried out by electroporation, which shortly increases cell membrane permeability and enables the uptake of plasmid DNA. For electroporation, 50  $\mu$ L electrocompetent cells were mixed with 5  $\mu$ L of ligation mix or 0.5  $\mu$ L of gene synthesis fragment DNA solution and transferred into an electroporation cuvette (electrode distance 1 mm). The electroporation was performed at 1,300 V/cm, 10  $\mu$ F, and 600  $\Omega$  (Eporator V1.01, Eppendorf). Thereafter, the cells were resuspended in 1 mL of LB medium and incubated at 37 °C and 900 rpm (Titramax 1000, Heidolph) for 60 min. After centrifugation at 8000 rpm for 2 min, most of the supernatant was discarded and the cells were resuspended in the remaining medium. In case of using gene synthesis fragments, the centrifugation step was skipped and 50  $\mu$ L were directly used for the following step. The cell suspension was plated out on LB agar with appropriate antibiotic concentration for selection and incubated overnight at 37 °C.

When working with pBeloBAC vector backbone or the pBeloBACMVX plasmid, the transformation into competent *E. coli* cells was done by using the whole ligation reaction. Therefore, the 20  $\mu$ L of ligation reaction was pipetted on a dialysis membrane floating on MQ H<sub>2</sub>O for 10-20 minutes before using the remaining mix for transformation.

### **2.3.15 *M. xanthus* DK1622 transformation**

*M. xanthus* DK1622 was cultured overnight in 25 mL CTT medium until an appropriate cell density was reached (OD<sub>600</sub> of 0.6). 2 mL of the culture was pipetted into an Eppendorf tube and centrifuged for 2 min at 6000 rpm and room temperature. Supernatant was discarded. The cell pellet was resuspended in 1 mL sterile MQ water and centrifuged again for 2 min at 6000 rpm, supernatant was again discarded using a pipette. This washing step was repeated once with 800  $\mu$ L. A small hole was poked into the lid using a needle, and 35  $\mu$ L of sterile MQ water plus plasmid DNA (5-10  $\mu$ L DNA/plasmid with a concentration of 100-

500 ng/ $\mu$ L DNA) was added. Cell pellet was resuspended, and the suspension was used for electroporation at 650 V, 25  $\mu$ F, and 400  $\Omega$  (Bio Rad GenePulser Xcell). After electroporation, the culture was resuspended in 1 mL CTT liquid medium and incubated on a shaker for 6 hours at 30° C and 1000 rpm. A glass vial with foam plug was used to mix 3 mL of CTT soft agar and the respective selection marker with the 1 mL transformation cultures. The suspension was thoroughly shaken and plated on a CTT agar plate containing the respective selection marker as well. Plates were incubated on 30° C for several days until clones appeared.

For constructs with the pBeloBac vector backbone, cultures were incubated at room temperature for 24 hours after transformation. After plating the cultures, the plates were incubated at room temperature for 1-2 days before storing in the 30° C incubator.

### **2.3.16 Screening for *E. coli* colonies**

To verify a successful transformation and isolate the constructed plasmid, 3 to 24 single *E. coli* colonies were picked and used for inoculation of 5 mL LB medium with appropriate antibiotic. Incubation took place at 37 °C and 200 rpm overnight. Plasmid DNA isolation from *E. coli* was carried out by the protocol described in 2.3.9. Afterwards, an analytical digest was performed (2.3.13.2) and the presence of correct DNA fragments was examined by agarose gel electrophoresis according to 2.3.11.

### **2.3.17 Red/ET – homology-based recombination**

Red/ET homology-based recombination<sup>124</sup> was performed according to following protocol: overnight culture of *E. coli* GB05 with the according plasmid was cultured with the respective selection marker. This overnight culture was then used to inoculate 12x2 mL Eppendorf tubes with a hole in the lid and 1.4 mL LB medium containing the respective selection marker (inoculum volume: 4x15  $\mu$ L, 4x25 $\mu$ L and 4x30-100  $\mu$ L). Those tubes were incubated on a shaker at 30° C until an OD<sub>600</sub> of roughly 0.2 was reached (duration approximately 2-3 hours). Three of the cultures that reached the desired OD<sub>600</sub> first were used for the Red/ET recombination: First 40  $\mu$ L of 10% L-Arabinose (dissolved in MQ-H<sub>2</sub>O and sterile filtered) were added to the cultures, which were then incubated at 37 °C for 45 minutes on a shaker. Afterwards, cultures were centrifuged for 1 minute at 9000 rpm at 2° C and supernatant was discarded using a pipette. All following steps had to be carried out on ice. The cell pellet was re-suspended in 1 mL ice cold sterile MQ-H<sub>2</sub>O, centrifuged for 1 minute at 9000 rpm and the supernatant discarded with a pipette again. This step was repeated once,

except not all of the supernatant was discarded (20-50  $\mu\text{L}$  leftover liquid). These tubes with cell pellet were now used for transformation with 1  $\mu\text{L}$ , 2  $\mu\text{L}$  and 2.5  $\mu\text{L}$ , respectively, of template DNA. The transformation was carried by electroporation as described before (see section 2.3.14). After the electroporation, the cultures were incubated for 75 minutes at 37° C on a shaker. Subsequently, the whole mL of each culture was plated on a LB Agar plate with the respective selection marker and incubated on 30° C. After 1-2 days, when colonies were grown, the whole cell mass was harvested by pipetting 2x1 mL of MQ-H<sub>2</sub>O onto the plate and scraping it off, then pipetting the suspension into new Eppendorf tubes. The plasmids of these cultures were then isolated by using the before described mini-prep protocol (see section 2.3.9.1) and the isolated plasmid was transformed into competent *E. coli* strains NEB 10 $\beta$ , HS996 or DH10 $\beta$  again to increase the chance of only having one plasmid per cell, *i.e.* no false positive clones that contain a mixture of right and wrong plasmids after the Red/ET process.

### 2.3.18 Gibson Assembly

Gibson assembly<sup>121</sup> reaction mix (New England Biolabs) was set up with 1  $\mu\text{L}$  of vector (approx. 40-60ng/ $\mu\text{L}$ ), 0.2  $\mu\text{L}$  of insert (approx. 100 ng/ $\mu\text{L}$ ), 1.3  $\mu\text{L}$  of MQ-H<sub>2</sub>O, and 2.5  $\mu\text{L}$  of NEBuilder® HiFi DNA Assembly Master Mix (New England BioLabs). Subsequently, the Gibson reaction mix was incubated in the PCR cycler for 15 minutes at 60°C and transformed into *E. coli* HS996 competent cells to screen for right plasmids.

### 2.3.19 Design and assembly of the pMYC21\_*mxvIK* construct

The plasmid backbone was assembled from one part of plasmid pMYC20CysOp1\_v2 (containing P<sub>van</sub>, p15A ori, chloramphenicol resistance gene as well as yeast replication genes, obtained from Dr. Sebastian Groß) and one part from plasmid pMYC21 (containing the Mx9 integrase, kanamycin resistance gene as well as the tD2 terminator, obtained from Dr. Sebastian Groß)<sup>71</sup>. Genes *mxvI* and *mxvK* were obtained by PCR with standard Phusion polymerase protocol as described in material and methods (section 2.3.10.1). Gene *mxvI* contained *NdeI* and *KspAI* RE recognition site overhangs, and gene *mxvK* *KspAI* and *XmaI* overhangs for cloning. Template for PCR amplification of *mxvI* and *mxvK* was pBeloBacMXV. Primer used for amplification: *mxvI* fwd and *mxvI* rev for *mxvI*, and *mxvK* fwd and *mxvK* rev for *mxvK*. Assembly was performed in a one-pot golden gate cloning approach by restrictive hydrolysis and ligation. pMYC20CysOp1\_v2 was hydrolyzed by *NdeI* and *PacI*, pMYC21 by *PacI* and *KspAI*, *mxvI* by *XmaI* and *NdeI*, and *mxvK* by *XmaI* and *KspAI*. All fragments were ligated

in one reaction, clones analyzed by analytical restrictive hydrolysis and one correct plasmid was sequenced with the primers listed in Table 2-9.

**Table 2-9** Primers and their sequence used for design and sequencing of the pMYC21\_*mxvK* plasmid

Primer name	Primer sequence 5' – 3'
<b>mxvI fwd</b>	GGAGTCATATGTCCGAGCCCCTCCCAC
<b>mxvI rev</b>	ACTCCGTTAACCTAGGTCAGCCGACGTTTCCCGTCT
<b>mxvK fwd</b>	GGAGTCCTAGGCACACCTTCCGGGGGAATGA
<b>mxvK rev</b>	ACTCCGTTAACGCTAGTCCTCGTCGAGGGGCT
<b>KT mxvI seq fwd</b>	CCACCTTCACCCTCTTCTCC
<b>KT mxvI seq rev</b>	CGGAAGCTCACGTCCTCGAA
<b>KT mxvK seq fwd</b>	GGATGCGCAACAAGCACGAC
<b>KT mxvK seq rev</b>	GCTTGTCTCGGGAATGCGG

### 2.3.20 Design of the targeted *mxvH* deletion using Red/ET homologous recombination

Gene *mxvH* was replaced on pBeloBacMXV by *kanR* using Red/ET. *KanR* was amplified by PCR using plasmid pDPO-mxn116 from Dr. Domen Pogorevc as template and *mxvH* KO *kanR* fwd and *mxvH* KO *kanR* rev as primer pair. Both primers had 50 bp overhangs homologous to the *mxvH* flanking regions, as well as a *Xma*I RE recognition site flanking the *kanR* gene after amplification. PCR and Red/ET recombination were performed according to protocols described in section 2.3.10.1 and 2.3.17, respectively. After Red/ET recombineering, plasmid pBeloBacMXV:: $\Delta$ *mxvH*\_kanR was hydrolyzed with *Xma*I to remove *kanR* and subsequently re-ligated. Restrictive hydrolyzation and ligation were conducted according to protocols described in sections 2.3.13.1 and 2.3.13.3.

**Table 2-10** Primers used to amplify the template used to replace *mxvH* in pBeloBacMXV by Red/ET.

Primer name	Primer sequence 5' – 3'
<b>mxvH KO kanR fwd</b>	CCGGGCTCTGCCGGCGGGGTGGGAAGGGGCTCGGACATGGGGTTCCTGGGCCTAGGTC AGAAGAACTCGTCAAGAAGGC
<b>mxvH KO kanR rev</b>	TCTGGACGGACATGAGGCCGAGAGCCTGCGCATCAAGATGGGGAACTGACCTAGGTG GACAGCAAGCGAACCG

### 2.3.21 Functional inactivation of the active site in *mxvB*

The conserved serine residue was detected by finding the core motif of PCP domains (I/L)GG(D/H)SL.<sup>138</sup> Therefore, all PCP domains of the MXV gene cluster were extracted, translated, and aligned *in silico* using Geneious software. The consensus of the core motif of all 14 PCP domains was found to be LGGDSI, the conserved Ser residue at this location was chosen to be the target for inactivation by point mutation, changing the Ser residue to Ala (see supplementary Figure 2-13). A PCR fragment was designed switching the sequence of Ser at this position from GGA to GGC. Primers were designed to obtain the mutated gene *mxvB\** by three separated PCRs. Two PCRs amplified the parts upstream and downstream of the targeted Ser, with overlaps at this region (Primer pair P<sub>1</sub> and P<sub>2\*</sub>, and primer pair P<sub>3\*</sub> and P<sub>4</sub>, respectively). The primers of this overlapping region contained the point mutated base pair C instead of A. Overlapping extension PCR, using the obtained two overlapping fragments and primers flanking the whole gene *mxvB\**, amplified the final construct (Primer pair P<sub>1</sub> and P<sub>4</sub>). The final construct was further flanked by the RE recognition sites *PvuI* and *BspTI*. Simultaneously, a kanamycin resistance gene *kanR* was amplified by PCR with overhang primers containing 50 bp homology regions flanking *mxvB*, as well as *PvuI* and *BspTI* RE recognition sites. This *kanR* construct was used to replace *mxvB* in pBeloBacMXV by Red/ET to form pBeloBacMXVΔ*mxvB\_kanR*. Subsequently, *kanR* was removed by restrictive hydrolysis with *PvuI* and *BspTI*, and the PCR construct *mxvB\**, hydrolyzed with the same REs, was ligated into this position forming pBeloBacMXVΔ*mxvB\_mxvB\** (additional information see section 2.6.8 Figure 2-14 and Table 2-23). The final construct was sequenced by Illumina sequencing to confirm the correct sequence.

### **2.3.22 MIC analysis of *M. xanthus* DK1622 wild type and mutants**

A small piece of bacterial mass from freshly plated *M. xanthus* DK1622 strains was suspended in CTT media and incubated at 30° C and 200 rpm. After 3 days, multiple dilutions of the bacterial suspension were made and incubated overnight with the same conditions as before. Next morning, the dilution that grew to a OD<sub>600</sub> of 0.5 was used to further dilute to OD<sub>600</sub> = 0.1, then added (75 µL/well) to a 96-well plate already containing 75 µL/well of MXV A serially diluted from 256 to 2 µg/mL (final concentration). The 1% DMSO solution in CTT with and without bacterial suspension was used as positive and negative control, respectively. To evaluate the effect of vanillate induction, the same MIC assay was performed in the presence of 1 mM sodium vanillate. The plates were sealed and incubated at 30° C with shaking. The bacterial growth was investigated after 72 h. MIC was defined as the lowest concentration of MXV A that inhibited visible growth of bacteria.

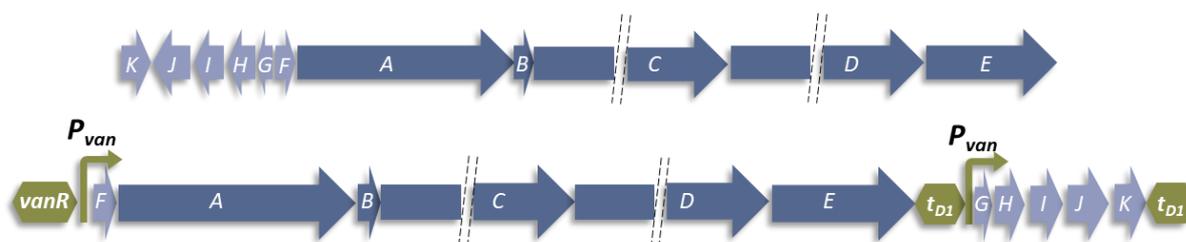
## 2.4 Results and discussion

### 2.4.1 Work towards the heterologous production of MXV

#### 2.4.1.1 Design of the MXV biosynthetic gene cluster

The synthetic MXV BGC was designed *in silico* for heterologous expression in *M. xanthus* DK1622. Since the produced antibiotic is toxic, the native promoter/terminator sequences were replaced by the established P<sub>van</sub> promoter and tD1 terminator system to allow functional gene expression.<sup>139,140</sup> Rearranging the native gene organization enabled the expression of the whole BGC by only two promoters and terminators. In this way the number of repetitive promoter and terminator sequences is limited to two, which facilitates *in vitro* cloning as well as chemical synthesis of terminators and repetitive sequences, which can be difficult to achieve. While reorganizing genes from the native BGC, intergenic regions that putatively contain native promoters and/or terminators were deleted. In case the intergenic regions including native promoters and terminators are deleted, it is vital to identify the exact starts of the respective genes to avoid accidentally truncating or elongating them, which would potentially result in the expression of non-functional proteins. The putative exact gene starts were reviewed including information from 'FramePlot 4.0beta analysis', manually comparison of potential ribosomal binding sites (RBSs), analyzing the genes with SwissProt<sup>141</sup> and UniProt<sup>142</sup> databases, and sequence alignment between the different producer strains.

The P<sub>van</sub> promoter and the t<sub>D1</sub> terminator were chosen for the regulation of the synthetic MXV BGC in the heterologous host, since the functionality of both genetic regulators have been proven previously in *M. xanthus* DK1622.<sup>71</sup> The genes were re-organized in a way that all biosynthetic genes can be regulated in two operons: the first operon comprises the genes *mxvFABCDE*, while the second operon consists of the genes *mxvGHIJK* which is directly located downstream of the first operon. The coding strand of *mxvK* was reversed, so that all genes are located on the same coding strand. The reorganization of gene encoding sequences (CDS) of the BGC from the native producer strain *Corallococcus coralloides* 1071 was performed as minimalistic as possible to decrease the risk of potential negative effects on the pathway expression. In the figure below, the structure of the synthetic BGC is shown compared to the native *C. coralloides* 1071 cluster:



**Figure 2-3 Top:** Structure of the native MXV cluster from *Coralococcus coralloides* 1071. **Bottom:** Structure of the synthetic MXV BGC. The first operon contains the genes *mxvFABCDE* and the second operon contains the genes *mxvGHIJK*. The expression of both operons is regulated by the vanillate-inducible promoter ( $P_{van}$ ) system and the  $t_{D1}$  terminator. The vanillate repressor gene (*vanR*) is located upstream of  $P_{van}$  of the first operon. Genes *mxvC* and *mxvD* were shortened in the figure due to the lengths.

#### 2.4.1.2 Assembling strategy and synthesis of MXV gene fragments

An overview of the assembling process of the synthetic MXV BGC is shown in Figure 2-4. Three large repetitive sequences were found in the NRPS genes *mxvC* and *mxvD* by dotplot analysis with *Geneious* (detailed information in SI section 2.6.4). Their presence is likely attributed to the incorporation of similar building blocks performed by structurally similar NRPS modules, which can lead to similarity on protein level and also on DNA level. Considering these three long repetitive sequence segments, homologous recombination-based methods for the assembly were excluded to prevent unwanted homologous recombination events leading to deletions of essential sequences. Instead, the assembly of the synthetic DNA fragments was based on a 3-step cloning strategy using type IIS restriction endonucleases (RE) previously described by Yan, Burgard *et al.*<sup>143</sup> with classical restriction hydrolysis/ligation-based cloning techniques. The stepwise assembly was enabled by dividing the MXV BGC in artificial fragments flanked by selected unique conventional type II RE recognition sites on the 5' and 3' ends. These unique type II RE sites were flanked by recognition sites for the type IIS RE *BsaI*, forming splitter elements (SE) located between the catalytic domain-encoding regions. Type IIS REs hydrolyze DNA outside of their recognition sequence, which enables the possibility to exchange DNA segments within these SEs after the BGC assembly. A final desplitting step, *i.e.*, hydrolyzing the vector with *BsaI* and subsequent religation, removes these elements and leads to the final 'scarless' MXV BGC, *i.e.*, no leftover DNA sequences from RE recognition sites (see scheme in Figure 2-4). The described 3-step cloning strategy with SEs was facilitated by virtually removing all *BsaI* and used unique conventional type II sites from the MXV BGC, by silent point mutations. Additionally, all *NdeI* RE sites were removed in the same way to ensure the possibility of replacing genes downstream of the  $P_{van}$  promoter. The terminal three bases of the  $P_{van}$  promoter CAT and the start codon ATG of any gene encoding

sequence result in the RE recognition sequence of *Nde*I 5' CATATG 3', allowing the replacement of genes in that location *via Nde*I. The silent point mutations were introduced by synonymous codon substitution, *i.e.*, the deliberate usage of codons with similar abundance in the whole BGC to prevent bottlenecks in the translation.

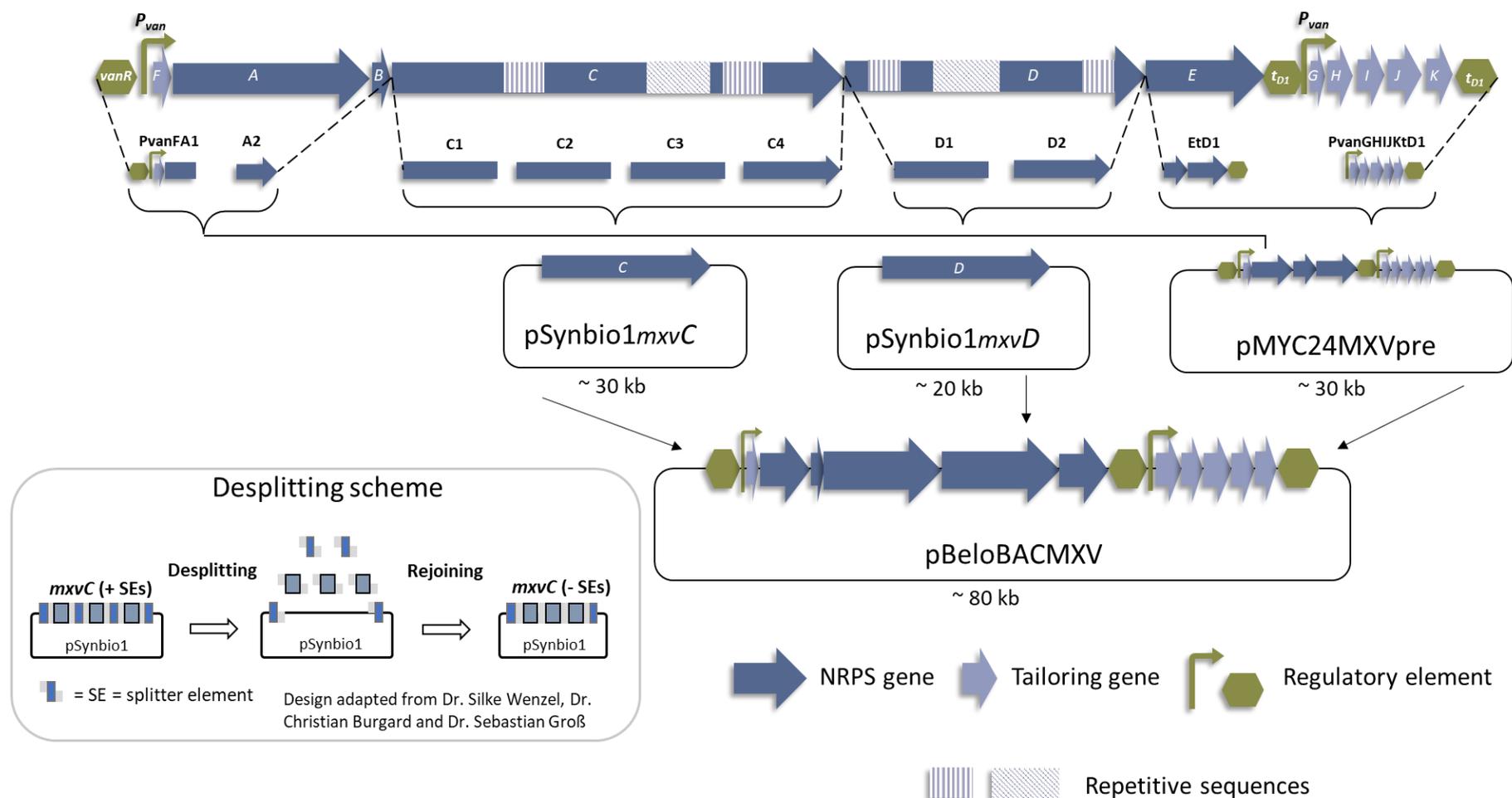
The enormous size of the MXV cluster (66,514 bp) required the gene synthesis of several fragments with a size maximum of 7.9 kbp (to reduce the duration and cost of the gene synthesis). The cluster was divided into ten gene synthesis fragments (see SI Table 2-17) which were synthesized by *ATG:biosynthetics*, designed for a stepwise assembly into pSynbio1 cloning vector (high-copy vector backbone, designed for amplification and selection in *E. coli*<sup>143</sup>) and pMYC24 shuttle vector. The pMYC24 shuttle vector for cloning in *E. coli* and *M. xanthus* was obtained by replacing URA3 with *ampR* in the pMYC20 vector, which is described by Groß *et al.*<sup>71</sup> The gene synthesis fragments were delivered in pUC57 standard cloning vectors, transformed into *E. coli* DH10 $\beta$  and the sequence of the gene synthesis fragment-containing constructs was verified by restriction hydrolysis with several enzymes and subsequent agarose gel electrophoresis after isolation from *E. coli* *via* alkaline lysis (see section 2.3.13.2, 2.3.11, and 2.3.9.1). The release of the gene synthesis fragments from the pUC57 vectors was achieved by restrictive hydrolysis *via* the flanking SEs for the assembling process.

Considering their lengths, the genes *mxvC* and *mxvD* were designed to be independently assembled from four and two gene synthesis fragments, respectively, using the pSynbio1 cloning vector. One hundred bp of the start and end, respectively, of *mxvC* and *mxvD* were located within separate SEs to preserve the option to modify the gene start region in case the gene starts were falsely interpreted. The remaining fragments (PvanFA1, A2, frag-EtD1, and PvanGHIJKtD1) had to be assembled in the pMYC24 vector, resulting in pMYC24MXVpre, which would later serve as shuttle vector for transformation into *M. xanthus* DK1622. Afterwards, the completely assembled and desplitted genes *mxvC* and *mxvD* were introduced into the SEs in between their respective 100 bp 5' and 3' ends in the frag-E part of pMYC24MXVpre. The pMYC24 vector backbone (p15A low copy ori) was designed to be exchangeable by a pBeloBAC vector backbone. BAC vector backbones lead to a replication of only 1-2 plasmid copies per cell, which allows stable maintenance of large DNA constructs. All

restriction hydrolysis/ligation-based cloning steps, including the respective inserts/vectors used, are summarized in supplementary Table 2-18.

#### **2.4.1.3 Assembly of the modified MXV BGC**

The assembly of the ten synthesized DNA fragments (see Table 2-17) was conducted as previously described (see chapter 2.4.1.2) by step-by-step hydrolysis and ligation as summarized in Table 2-18. An overview of the whole assembly process is shown in the figure below.



**Figure 2-4** An overview of the MXV synthetic BGC assembly process. The design of the synthetic BGC was based on the 66 kbp long native MXV BGC of the producer strain *C. coralloides* Ccc1071. The designed BGC was divided into ten fragments and chemically synthesized. Fragments C1-C4 were subcloned on pSynbio1 vector generating pSynbio1mxvC; fragments D1-D2 were subcloned on pSynbio1 vector generating pSynbio1mxvD; and fragments PvanFA1, A2, EtD1 and PvanGHIJKtD1 were subcloned on pMYC24 vector to form the shuttle vector pMYC24MXVpre. *mxvC* and *mxvD* were removed from their pSynbio1 cloning vectors and recovered for insertion in the final MXVpre vector. The pMYC24 vector backbone was exchanged to a pBeloBac vector backbone prior to inserting *mxvC* and *mxvD*. All final constructs had to be desplitted to remove all splitter elements (desplitting scheme shown on the left). The figure further highlights the repetitive sequences located on *mxvC* and *mxvD* by the white/grey shaded boxes in the blue arrows.

After successful assembly of the vectors pSynbio1mxvC, pSynbio1mxvD and pMYC24MXVpre, the recovery of the genes *mxvC* and *mxvD* and insertion into the shuttle vector pMYC24MXVpre turned out to be inefficient and time consuming, potentially due to the size of the genes *mxvC* and *mxvD* (*mxvC* 25 kbp, and *mxvD* 15.5 kbp, respectively) and the repetitive sequences. No clones containing a correct plasmid with inserted *mxvC* or *mxvD* were obtained. Therefore, the pMYC24 vector backbone (containing a p15A low copy ori) was replaced by the pBeloBac vector backbone (single copy ori, 1-2 plasmids per cell). BAC vector backbones are capable of maintaining large DNA fragments of over 300 kbp.<sup>144</sup> An additional decrease of the culturing temperature to 30° C, instead of 37° C applied for smaller plasmid constructs, ensured slower doubling time of the *E. coli* cells harboring the plasmid (details of adapted parameters regarding cloning protocols with pBeloBac vector backbone described in material and methods section 2.3.9.3, 2.3.11.1, and 2.3.13.4, respectively). Slower doubling times of the cells reduce the replicating rate of plasmids and thus reducing the chance for mutations and facilitate handling of large constructs. After exchanging the vector backbone, first the insertion of *mxvC* and subsequently *mxvD* into the pBeloBacMXVpre vector was successful, generating the pBeloBacMXVpreCD gene construct.

The size of the expression construct (approx. 80 kbp) including the three repetitive sequences complicated the cloning process. With increasing plasmid construct sizes, as soon as *mxvC* and *mxvD* were tried to be inserted into pMYC24MXVpre, the cloning efficiency significantly decreased. All screened transformants harbored incorrect plasmids, the band patterns shown on agarose gels concluded that the transformants were neither harboring the re-ligated vector nor the insert. One of the bottlenecks of the process might have been the *in vitro* ligation of insert into vector (*e.g.*, *mxvC* and *mxvD* into MXVpre and desplitting). It was observed, that with increasing size of DNA construct the concentration of DNA after hydrolysis and purification decreased significantly, which in turn lead to low concentrations of vector and insert for the follow up cloning procedure, leading to low cloning efficiency. To ensure sufficient amount of DNA for the follow up cloning, the amount of initial DNA used for hydrolysis was increased from 2–3 µg to 4 µg. Additionally, the hydrolyzed DNA was purified with a Roche DNA extraction kit which uses silica beads as DNA binding material that can bind DNA fragments of up to 150 kbp in size. Further, the transformation and ligation efficiency drop significantly with increasing construct size.<sup>145</sup> Multiple ligation reactions were set up with varying molar ratios and concentrations of vector and insert. It was attempted to increase the

transformation efficiency by increasing the DNA used for transformation, de-salting the ligation mix, lowering the incubation temperature after transformation, and decreasing the concentration of antibiotic as selection marker. These adaptations on the cloning and transformation protocol seemed to have a positive impact on the overall cloning efficiency and are described in detail in material and methods section 2.3.11.1, 2.3.13.4, and 2.3.14.

#### **2.4.1.4 Desplitting the final MXV synthetic BGC**

The last step of the MXV synthetic BGC assembly was the desplitting of the pBeloBacMXVpreCD plasmid to obtain the SE-free pBeloBacMXV plasmid. Therefore, the plasmid was hydrolyzed with the restriction endonuclease *BsaI* and the resulting DNA fragments subsequently separated by a DNA extraction and purification kit. The small SEs (24-26 bp) were excluded, and the remaining DNA fragments could be recovered and re-ligated. The *BsaI* RE introduced sticky ends were all designed to be unique, *i.e.*, in theory the fragments could only ligate in a way to form the correct construct harboring the full MXV synthetic BGC again. The DNA recovery, especially to obtain high concentrations of all fragments, was tricky due to the large size differences of the generated fragments after hydrolyzing the construct with *BsaI*. Using the peqGOLD Gel Extraction Kit (peqlab) (see section 2.3.11) enabled purification of all fragment sizes between 200 bp and 24.8 kbp while at the same time excluding the SEs (cut-off of 50 bp). The first correctly ligated/assembled plasmid (according to its restriction digestion patterns) was sent for Illumina sequencing. This revealed, that the 200 bp fragment located between two SEs and harboring the 3' end of *mxvC* and 5' end of *mxvD* (see SI Table 2-17 fragment frag-EtD1) was inserted in reversed direction. As it turned out a mistake in designing the sticky ends enabled this 200 bp fragment to be inserted in both directions. Since there is a theoretical 50% chance of this short fragment to be ligated in the right direction, an additional 144 clones of the agar plates containing clones from the previous desplitting procedure were screened for plasmids of the right size. These plasmids were not only analyzed with REs to confirm the right size and the presence of remaining SEs by analytical digests, but further also analyzed by PCR to investigate the direction of the mentioned 200 bp fragment. Therefore, two primer pairs were designed, one that would bind on the plasmid in case the fragment is inserted in forward direction, and another that would bind on the plasmid in case the fragment is inserted in reverse direction, respectively. Out of the 144 screened clones, nine clones harbored a plasmid with the correct size. Seven of those showed a band for the primer pair for forward inserted fragment by PCR. Out of those seven, four clones

revealed to have all SEs removed. Two of those four plasmids were sent for Illumina sequencing, which showed that one plasmid (clone 76) was completely free of mutations (whereas clone 70 featured one point mutation inside the gene cluster). Detailed information about the screening process, including SE locations, REs used for analysis and PCR to check for the 200 bp fragment, are shown in the SI (see section 2.6.6 Table 2-19 and Table 2-21). Subsequent analytical restriction digestions and *in silico* analysis of the final plasmid construct revealed, that two *BsaI* restriction endonuclease recognition sites remain in the sequence flanking the pBeloBac vector backbone, since the recognition sites of *BsaI* read in the wrong direction due to an error in design. Unfortunately, they cannot be removed by the desplitting procedure, which restricts Red/ET homologous based recombination and subsequent replacement of the selection marker with *BsaI*. Therefore, a future replacement of the pBeloBac vector backbone by an *in silico* designed and synthesized pBeloBac vector backbone removing these *BsaI* recognition sites is in progress.

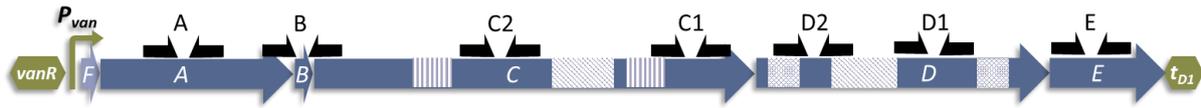
The last desplitting step of plasmid pBeloBacMXVpreCD to obtain the final plasmid with the entire MXV BGC pBeloBacMXV was supposedly to be straight forward and efficient, due to the design with unique sticky ends after restrictive hydrolysis with *BsaI*. Unfortunately, the only 200 bp long fragment located on the 3' end of *mxvC* and 5' end of *mxvD* could be ligated into the construct forward and reversed. Since this design flaw was discovered in advance, it was rather simple to overcome this obstacle by intensive screening of further 144 clones and analyzing the area containing that 200 bp fragment by PCR. However, this problem extended the cloning procedure of the synthetic MXV BGC. Nonetheless, the assembly process was successfully completed, yielding one plasmid harboring the whole MXV BGC without any gene mutations according to Illumina sequencing. The whole procedure was time consuming and tedious, which emphasizes how challenging working with GC-rich myxobacterial BGCs of such an enormous size including repetitive sequences can be, especially when applying classical digestion/ligation approaches such as 'Golden-Gate' cloning.

The lessons learned from this work, including multiple steps of troubleshooting, and optimizing the parameters, can be helpful for future work and projects containing design and assembly of similar composed gene clusters and are discussed in detail in chapter 5.

## 2.4.2 Heterologous expression of the synthetic MXV BGC in *Myxococcus xanthus* DK1622

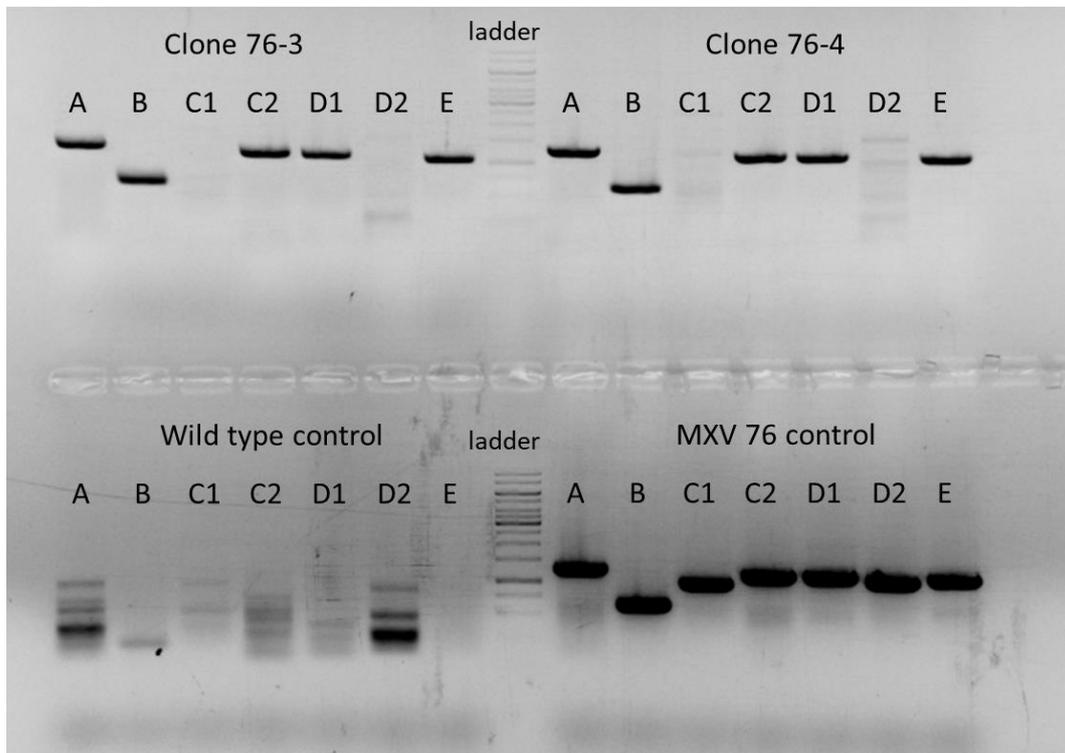
### 2.4.2.1 Integration of the MXV BGC into *Myxococcus xanthus* DK1622 via Mx8 phage-integrase

The final MXV BGC construct pBeloBacMXV was integrated into the genome of wild-type *M. xanthus* DK1622 via the Mx8 phage-integrase according to transformation protocol described in the material and methods part (see section 2.3.15). To confirm correct transformants with successfully integrated MXV gene cluster, colonies were selected and some of the cell mass used for colony PCR (see section 2.3.10.1). Using four specifically designed primers to bind to the flanking regions of the Mx8 *attB* site and Mx8 phage-integrase, respectively, this colony PCR can confirm the successful integration of a gene construct harboring the Mx8 phage-integrase. These four specific Mx8 primers however can only confirm the successful integration of the Mx8 phage-integrase and adjacent genes. They cannot confirm the full integration and integrity of the whole construct, in this case the complete synthetic MXV BGC, and thus risking proceeding with false positive clones. The problematic nature of this screening approach was revealed when the first two genotypically verified (by colony PCR with Mx8 primers) *M. xanthus* DK1622 transformants were cultured but did not yield in production of any MXV derivatives, although the colony PCR with the Mx8 primers showed a successful integration. A similar issue was experienced during the transformation and heterologous expression of a synthetic argyirin BGC in *M. xanthus* DK1622, where further colony PCRs with specific designed primers binding to multiple regions in the argyirin BGC revealed that only a part of the BGC was integrated into the genome (personal communication with Dr. Domen Pogorevc). In order to further analyze the integrated gene construct and the presence of all genes of the MXV BGC, seven primer pairs were designed binding to each of the five NRPS genes of operon 1 (*mxvA* – *mxvE*) as shown in Figure 2-5. Due to their size and the presence of the repetitive sequences in their coding strand, two primer pairs were designed for *mxvC* and *mxvD*. The sequences of the primer pairs are listed in the SI section 2.6.7. The small genes located in operon 2 were not regarded for this first approach, since it was assumed that possible deletions were most likely to happen within the repetitive sequences.



**Figure 2-5** The seven primer pairs located on the five NRPS genes *mxvA* – *mxvE*, used to investigate their presence by colony PCR.

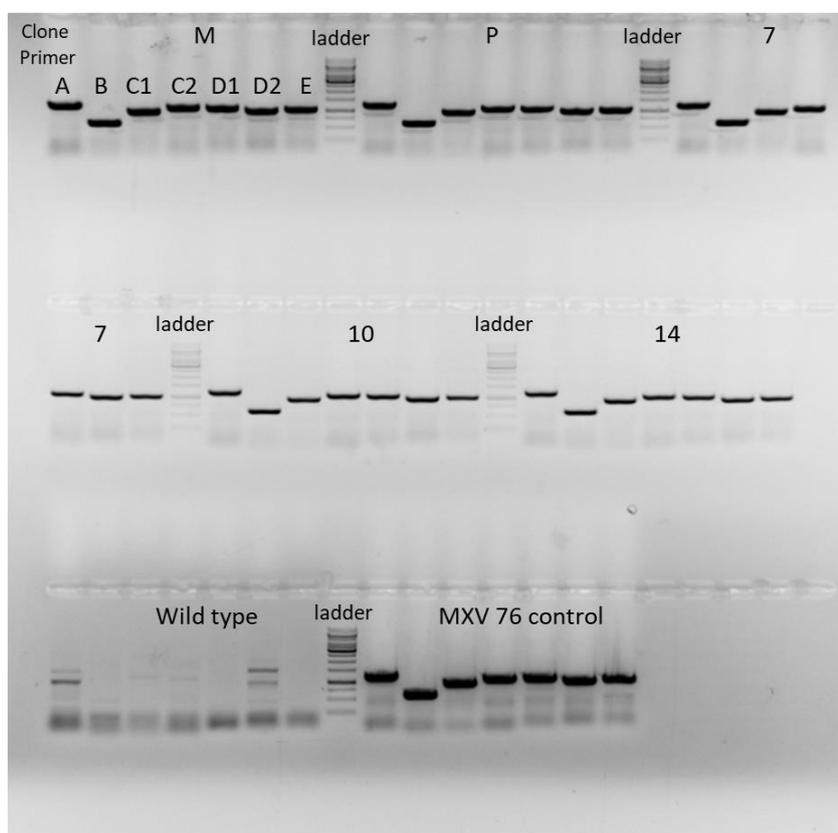
Analysis by colony PCR with those seven additional primer pairs of the previous two cultured transformants, which did not show any MXV production, revealed that parts of the BGC located within the coding strands of *mxvC* and *mxvD* were not present in the integrated gene construct. Primer pairs C1 and D2 did not amplify the respective DNA fragment and therefore showed no visible bands after visualizing the PCR reactions on an agarose gel (Figure 2-6).



**Figure 2-6** Colony PCR with the seven designed primer pairs binding to the five NRPS genes *mxvA* – *mxvE* located in operon 1. Clones 76-3 and 76-4 are the analyzed *M. xanthus* DK1622 mutants that showed positive colony PCR results with the Mx8 primers after transformation with pBeloBacMXV. Controls were made with wildtype *M. xanthus* DK1622 and pBeloBacMXV (Illumina sequenced and confirmed clone 76). Primer pairs C1 and D2 did clearly not show a band, except some faded bands that indicate unspecific primer binding. Ladder= GeneRuler 1 kbp DNA ladder.

Those parts of the MXV BGC between the two largest repetitive sequences (see Figure 2-5) were deleted either during or after integration into the genome. The two ATP-dependent recombinases *recA1* and *recA2* enable homologous-based recombination in *M. xanthus* DK1622 as a self-repair mechanism.<sup>146,147</sup> It is possible that gene sequences with high

similarity, *e.g.*, the repetitive sequences located in *mxvC* and *mxvD* within the MXV NRPS coding strands can be recombined and deleted when located in proximity. Therefore, the transformation process was repeated with adapted parameters. The cells were incubated for 24 hours at room temperature instead of 6 hours at 30° C after transformation. Subsequently, after plating the cultures on agar, they were stored at room temperature for several days before incubating at 30° C. Newly detected transformants were screened by colony PCR with primer pairs C1 and D2 instead of the Mx8 primers, which showed bands for five clones. They were now further analyzed with the remaining five primer pairs which gave five positive transformants, resulting in an integration efficiency of approximately 25%.



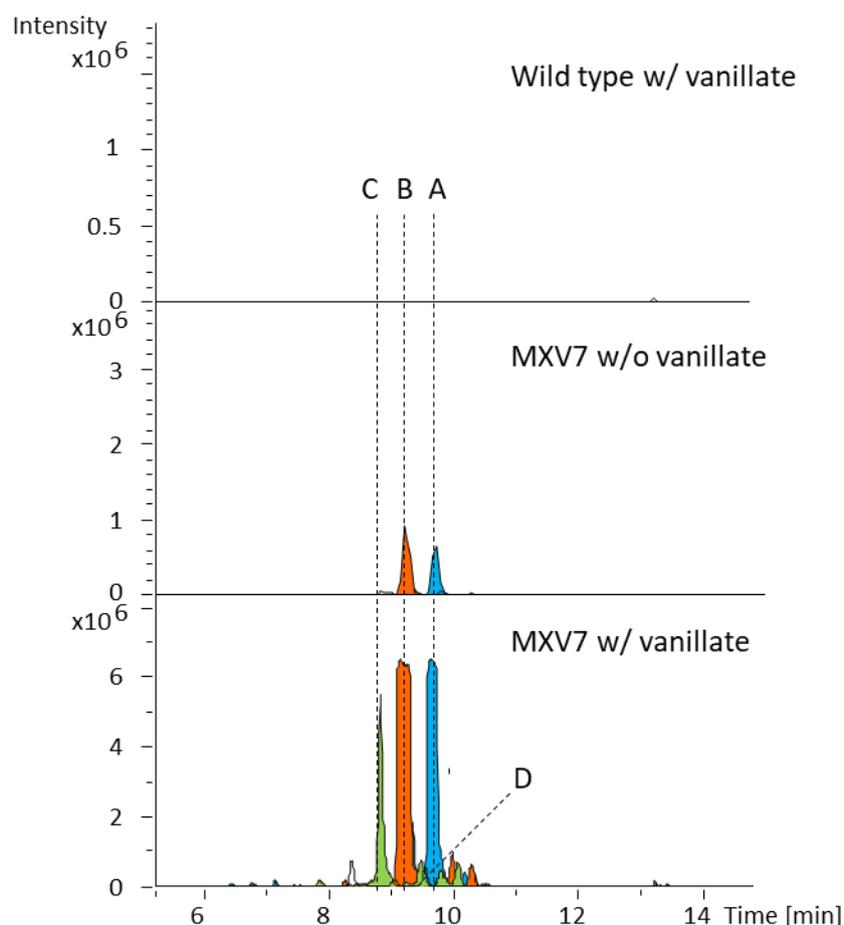
**Figure 2-7** The five analyzed mutants *M. xanthus* DK1622::pBeloBacMXV-M, P, 7, 10 and 14 now showing bands for all primer pairs. As controls a DK1622 wildtype and pBeloBacMXV plasmid were used. Ladder= GeneRuler 1 kbp DNA ladder.

The positive colony PCRs with these five mutants indicate that all five NRPS genes *mxvA* – *mxvE* were successfully integrated. However, the used primer pairs give only certainty about the analyzed locations. Deletions outside those locations, or in operon 2, can theoretically only be ruled out by designing multiple additional primers spanning the whole BGC or by genome sequencing. The here described risk that parts of the integrated BGC can be deleted after transformation into *M. xanthus* DK1622, as also experienced in the argyris heterologous

production, should also be considered in other heterologous expression projects when mutants do not show production of the desired compound. In that case, the integrated gene construct should be further analyzed by multiple colony PCRs investigating wide-spread areas of the construct. Additionally, the described adaptations in the transformation process seem to improve the efficiency by lowering the temperature and thus increasing the doubling time of the cells.

#### 2.4.2.2 Analysis of the heterologous production of MXV derivatives in the DK1622 host

The five mutants with positive PCR results were grown in liquid culture and the presence of MXV derivatives in the culture broth was analyzed according to protocol (see section 2.3.8). Four of these five clones showed production for the four major MXV derivatives MXV A, B, C, and D, respectively.



**Figure 2-8** shows extracted ion chromatograms (EIC) from uHPLC-hrMS analysis of the DK1622::pBeloBacMXV culture broth extracts. The top chromatogram shows wild type control induced with vanillate, the middle chromatogram clone DK1622::pBeloBacMXV-7 with no vanillate induction, and the bottom chromatogram shows clone 7 with vanillate induction. All four major MXV derivatives can only be seen in the induced mutant culture extract and are labelled respectively. Small peaks for MXV A and B were also detected for the uninduced control.

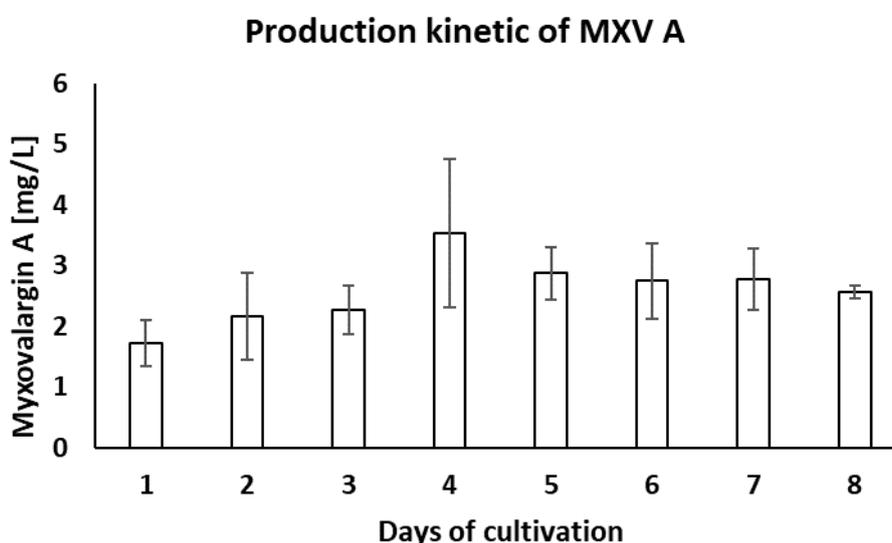
**Table 2-11** Initially determined single values, exemplifying the production titer of MXV A measured for the five tested DK1622::pBeloBacMXV mutant cultures.

Clone	Myxovalargin A
<b>MXV-7</b>	33.7 mg/L
<b>MXV-10</b>	12.7 mg/L
<b>MXV-14</b>	8.4 mg/L
<b>MXV-M</b>	-
<b>MXV-P</b>	13.2 mg/L

The production titers for the most active and abundant derivative MXV A are shown in Table 2-11. It is noteworthy, that MXV A and MXV B were also detected in the uninduced control culture. This is explainable by the background expression of the leaky vanillate promoter.<sup>148</sup> Although all four major derivatives were visible in the extracts of the four producing mutants, MXV A is the most active and abundant, and the only one for which pure compound was available for analysis. Hence, all quantifications conducted with the heterologous expression construct refer to MXV A. Mutant DK1622::pBeloBacMXV-M did not show production of any MXV derivative. It is possible, that parts of the gene cluster that were not investigated by the designed primers were deleted in that clone or that one or multiple mutations occurred within the gene cluster. Incomplete or false DNA sequence can lead to translation of false or truncated proteins, which in turn would hinder production of the desired compound. The only possibility to confirm such deletions or mutations would be to sequence the whole genome of the non-producing transformant. However, since four clones successfully produced MXV, the non-producing one was disregarded for future cultivation experiments at this point. The highest quantified production in this first heterologous production approach was 33.7 mg/L by *M. xanthus* DK1622::pBeloBacMXV clone 7. Later MIC experiments with mutant and wild type strain showed a MIC for MXV A of approximately 32 mg/L. This indicated that the limit of possible production was already reached, since higher MXV concentrations would lead to self-toxicity. The measured production of 33.7 mg/L could however not be replicated anymore in all later culture experiments. Instead, concentrations ranged between 3.5 and 13 mg/L, similar to the yields measured for mutants 10, 14 and P in the first cultivation approach (see Table 2-11). The MXV production titer of the heterologously

expressed synthetic MXV BGC in *M. xanthus* DK1622 is comparable to the titer of the native producer *Myxococcus fulvus* Mx f65. In cultivations with the wildtype strain 4–6 mg/L were obtained, while mutants that were treated with *N*-methyl-*N'*-nitro-*N*-nitrosoguanidine had a higher resistance to the antibiotic and produced 30 mg/L in shake flasks and 13 mg/L in a bioreactor.<sup>62</sup> Successful heterologous production and the measured yields confirm that *M. xanthus* DK1622 is a suitable host for the heterologous expression of the MXV BGC, which enables a simplified engineering of the BGC for a detailed analysis of the biosynthesis.

Further experiments should facilitate a better insight into the heterologous production of MXV and find possible ways to increase the production titer. Therefore, the following fermentation experiments focused on the production kinetics, the influence of supplemented adsorber resin XAD-16, multiple inductions with vanillate, and co-expressing an additional construct with the two ABC transporter genes *mxvI* and *mxvK*. First, the production kinetic was analyzed to understand at which time points of the cultivation MXV starts to be produced, when the production titer peaks, and when the titer stagnates. Therefore, 24 production cultures were inoculated from the same pre-culture, enabling to take triplicates for production analysis of each day for a period of eight days.



**Figure 2-9** Concentration of MXV A measured over the duration of eight days. Samples were taken in triplicates.

The production of MXV started as soon as the cultures were induced with vanillate, *i.e.*, the concentration reached around 2 mg/L after 24 hours already which was approximately 50% of the maximum reached during this experiment. Peak production was reached on day 4

with an average titer of around 3.5 mg/L. The following four days, concentration of MXV dropped slightly and stagnated at a value of approx. 3 mg/L, indicating that the production of MXV peaks relatively early after four days of cultivation. These results are consistent with production kinetics found in cultivations with the native producer *M. fulvus* Mx f65, where MXV was detected soon after inoculation. The titer was increasing until stationary phase was reached at which point it remained constant.<sup>62</sup> Future cultivation experiments can therefore be ended after 4 days to harvest at the time point of maximum concentration as well as to save time. It is noticeable, that the concentration of MXV is significantly lower compared to the previous cultivation. The fluctuations in the production yields between different sets of cultivation experiments hamper the comparability thereof. Those fluctuations can possibly be caused by the complex regulation system affecting the metabolome of *M. xanthus* DK1622, for which only a few studies are available to date.<sup>149,150</sup> The ability of *M. xanthus* to form fruiting bodies, the appearance of the tan-phenotype,<sup>151</sup> as well as the relatively unexplored regulatory system<sup>112–114,152</sup> with various competing BGCs, can lead to variations in the metabolome and therefore also in the production yields of certain compounds. Fruiting body formation for example leads to production of life-cycle related compounds. Analysis thereof led to the discovery of the novel compound homospermidine lipid.<sup>153</sup> High fluctuations in yields of heterologously produced target compounds in *M. xanthus* were also exhibited before, *e.g.*, in the heterologous production of the argyrins.<sup>148</sup>

In order to analyze the presence of MXV in the supernatant, the adsorber resin XAD-16, and cells, mutant DK1622::MXV-10 was inoculated in eight production cultures. Of those eight cultures, two cultures were supplemented with XAD-16 and six did not contain any XAD-16. The cultures without XAD-16 were used to test the presence of MXV in the supernatant and cell pellet. Additionally, two cultures without XAD-16 were used to test the influence on yield using a liquid-liquid extraction procedure with methanol of the whole culture broth. The results are shown in the table below.

**Table 2-12** Production titers in [mg/L] of different methods to extract MXV A. SN = supernatant, L/L = liquid-liquid extraction with methanol. Analyzed mutant was *M. xanthus* DK1622::pBeloBacMXV clone 10.

Sample	Myxovalargin A [mg/L]
MXV-10 w/o XAD-16 SN 1	3.43
MXV-10 w/o XAD-16 SN 2	4.64
MXV-10 w/o XAD-16 cell pellet 1	1.75
MXV-10 w/o XAD-16 cell pellet 2	2.65
MXV-10 w/ XAD-16 1	6.8
MXV-10 w/ XAD-16 2	8.56
MXV-10 w/o XAD-16 L/L extraction 1	8.94
MXV-10 w/o XAD-16 L/L extraction 2	9.61

Production titers for supernatant and cell pellet combined were 5.18 mg/L for culture 1 and 7.29 mg/L for culture 2, respectively. The same cultures cultivated with XAD-16 as adsorber resin for extracellular compound, *i.e.*, only XAD-16 and cells were extracted as one pellet, produced 6.8 mg/L MXV for culture 1 and 8.56 mg/L for culture 2, respectively. The production for the cultures with XAD-16 was only slightly higher (1.6 and 1.26 mg/L more, respectively), while the two cultures cultivated without XAD-16 and extracted with methanol *via* liquid-liquid extraction showed the highest concentrations of MXV. Nevertheless, the variations in the yield between the different extraction methods are neglectable, especially considering the naturally occurring yield variations in different DK1622 cultures as described in the previous paragraph. A significant difference apparent in this experiment, however, was the growth or viability of the cells. The cultures containing XAD-16 were viable several days longer (depictable by the yellow color) compared to those without the resin, whose color turned brown indicating the death phase, about three days earlier. The extended growth when using XAD-16, could be caused by binding of the antibiotic compound to the resin, however, a control experiment should be conducted to confirm this. Nonetheless, for the reason of extended growth phase, future cultivation experiments were decided to be performed using the adsorber resin XAD-16.

The adsorber resin XAD-16 can decrease the toxicity in the culture by binding the produced antibiotic, thus facilitating cell growth, and assist in the stability of the produced compound.<sup>154</sup> Binding of the inducer vanillate to XAD-16 however, could decrease the protein expression of the BGC and lead to a lower production of MXV. To counterbalance this effect, the influence of multiple inductions with vanillate on the MXV production was investigated in another cultivation approach. An induction of *M. xanthus* DK1622::pBeloBacMXV mutant P every two days over a duration of eight days of cultivation, lead to a production of 13 mg/L MXV, approx. 1.5 fold higher compared to only 8.5 mg/l in the single induced control culture. These results suggest that a sufficiently high concentration of vanillate in the culture medium is important for the MXV production. Later cultivation experiments should also investigate the addition of XAD-16 to the culture medium 1–2 days after induction with vanillate. Both, the inducer concentration as well as the presence of adsorber resin to prevent high antibiotic concentrations in the medium, can apparently have a positive influence on the MXV yield. However, a high concentration of inducer could also lead to an overexpression of the proteins, leading to protein misfolding and aggregation, and metabolic overburdening of the cells,<sup>155</sup> thus making a good balance between resin and inducer is vital for an increased production titer.

In summary, the heterologous production of MXV in the host *M. xanthus* DK1622 was successful and production titers between 3.5 and 33 mg/L are comparable to the native producer.<sup>62</sup> These findings indicate that the DK1622::pBeloBacMXV mutants can be used as a future MXV producer strain as well as well as expression platform to further investigate and elucidate the MXV biosynthesis. Since the initial production of 33 mg/L was not achieved in follow-up cultivations anymore, and titers fluctuated around 10 mg/L, more experiments on increasing the production were planned. The findings showed that in future cultivation experiments XAD-16 should be used, combined with multiple inductions using vanillate. In future approaches to improve the yield, the influence of supplementation with precursors should be investigated. Supply of precursor, such as amino acids with high abundance in the target compound, can help to prevent metabolic bottlenecks. Repetitive addition of necessary precursors for the argyrimycin production achieved a 20-fold increase in yield compared to the native producer to a value of 160 mg/L.<sup>156</sup> In case of MXV, several amino acids such as valine, arginine, and alanine, as well as the unusual amino acids  $\beta$ -tyrosine (or the precursor tyrosine) and agmatine, could be supplied externally during cultivation. It remains to be seen, whether

the production of MXV can be boosted by amino acid and precursor feeding, or an excess of precursors would lead to substrate saturation and overburdening of the NRPS assembly line.

#### **2.4.2.3 Increasing the MXV self-resistance of the heterologous expression platform**

Irschik *et al.* described in their first publication about the discovery of MXV, that the native strain produced 4 – 6 mg/L MXV while the MIC against the antibiotic was only 6 mg/L. They treated the native *M. fulvus* strain Mxf65 with *N*-methyl-*N'*-nitro-*N*-nitrosoguanidine, leading to mutants with a MIC of 25 – 50 mg/L, which were able to yield 13 mg/L MXV in bioreactors and up to 30 mg/L in shake flasks.<sup>62</sup> These findings indicate that the self-resistance of the host plays an important role in the production yield of MXV and could explain why it seems that the limit of MXV concentration produced by the heterologous host was already reached. MIC testing of the wild type *M. xanthus* DK1622 against MXV showed a resistance of up to 16 – 32 mg/L. At the same time, mutants harboring the MXV BGC still grew distinctly at a MXV concentration of 32 mg/L. However, this value is not comparable since the potential resistance genes (ABC transporter proteins encoded by *mxvI* and *mxvK*) are only expressed when the medium is supplemented with vanillate, which would also induce production of MXV, changing the concentration and thus the MIC. Therefore, a new expression construct was designed, harboring only the two ABC transporter expressing genes *mxvI* and *mxvK*, regulated by a vanillate promoter and tD2 terminator.<sup>139,140</sup> The construct was designed for transformation into the DK1622 wild type in order to clearly see the influence on the MIC of the two resistance genes alone. It was further designed to be compatible for transformation into the MXV BGC harboring mutants. Therefore, the plasmid backbone of the construct contained a kanamycin resistance marker, compatible with tetracycline on pBeloBacMXV, as well as the Mx9 integrase, compatible with the Mx8 integrase on the pBeloBacMXV backbone. The detailed cloning procedure of the construct pMYC21\_*mxvIK* is described in section 2.3.19. After transforming the construct into *M. xanthus* DK1622 wild type, MIC testing was performed to analyze the influence of the ABC transporter genes on the resistance against MXV. Wild type strain and the newly obtained DK1622::pMYC21\_*mxvIK* were grown in 96-well plates with increasing MXV concentrations. The wild type control showed a MIC of 16–32 mg/L again, while both the induced DK1622::pMYC21\_*mxvIK* and the uninduced mutant control culture revealed a MIC above 256 mg/L. An explanation for the equal MIC for both induced and uninduced cultures could be the leakage of the vanillate promoter.<sup>148</sup> It seems that the background expression of the ABC transporter genes is already sufficient for a high resistance

against MXV. However, the highest tested MXV concentration was 256 mg/L. Higher concentrations could reveal differences in MIC between induced and uninduced control. Nevertheless, the growth of the cultures in MXV concentrations higher than 64 mg/L was significantly hampered, *i.e.*, that even though the strains are still able to survive in high MXV concentrations, their growth might not be sufficient for a relevant production of MXV.

Transformation and integration of the ABC transporter cassette into the Mx9 site of DK1622::pBeloBacMXV mutants, cultivation and analysis of extracts (according to protocol described in section 2.3.8) did not lead to a significant increase in the production. Three mutants with the co-expressed ABC transporter construct, each cultivated in triplicates, yielded in 8.6, 8.6, and 5.6 mg/L respectively. The average of those three mutants of 7.6 mg/L was about 1.2 mg/L higher compared to the control mutant not harboring the additional ABC transporter genes (6.45 mg/L). The production yield was in all mutants lower compared to previous measured titers. Possible reasons for alternating production yields were discussed before in section 2.4.2.2. However, the only slightly increased yields of the DK1622::pBeloBacMXV::*mxvIK* mutants, harboring an additional set of the two self-resistance genes *mxvI* and *mxvK*, leads to the conclusion of other bottlenecks for higher production of MXV A. The expression of the ABC transporter genes proved to significantly improve the self-resistance of the wild type DK1622 against MXV A, nevertheless, other bottlenecks like deficiency of certain precursors could prevent the mutants of producing high amounts of MXV A and thus making use of the increased resistance. Therefore, as discussed in the previous section, future experiments should analyze feeding of the cultures with precursors like valine, alanine, arginine and the unusual amino acids agmatine and  $\beta$ -tyrosine.

## **2.4.3 Investigation of the MXV biosynthesis utilizing the heterologous expression platform**

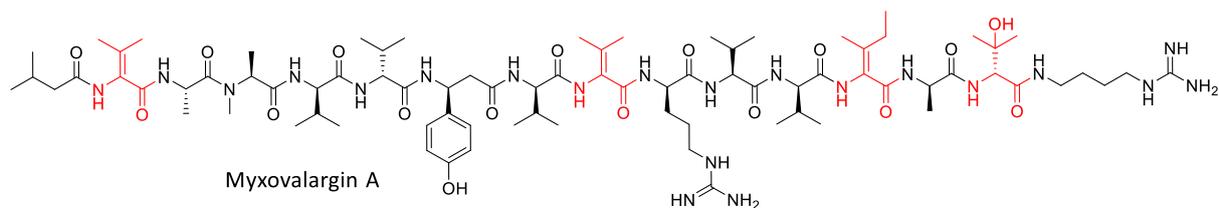
### **2.4.3.1 Targeted deletion of the putative $\beta$ -hydroxylase MxvH**

The heterologous expression platform for MXV enabled in-depth investigation of the putative  $\beta$ -hydroxylase MxvH. Understanding the role of this enzyme is desirable, since it is suggested to catalyze the hydroxylation and dehydrogenation of L-valine and L-isoleucine in the biosynthesis, but the function of MxvH is yet unknown.<sup>127</sup> The gene *mxvH* encoding this putative  $\beta$ -hydroxylase was deleted in a previous experiment in the native producer *Coralloccoccus coralloides* 1071 by Ullrich Scheid, which extinguished the production of MXV.

However, polar effects on neighboring genes could not be ruled out. Therefore a clean deletion, *i.e.*, without affecting neighboring genes, of *mxvH* by Red/ET homologous recombination using the bioynthetic MXV BGC was desirable. Primer design and cloning strategy are described in detail in the SI section 2.3.20. After replacing *mxvH* in the plasmid pBeloBacMXV by *kanR*, subsequent deletion of *kanR*, and religation of the plasmid forming pBeloBacMXV $\Delta$ *mxvH*, the plasmid was sequenced by Illumina to confirm the correct sequence and transformed into *M. xanthus* DK1622 wild type strain. Cultivation, extraction, and analysis was performed according to protocols in section 2.3.7, and 2.3.8, respectively. The extracts were analyzed for the presence of MXV A, as well as derivatives potentially occurring due to the missing  $\beta$ -hydroxylase. Since MxvH is presumably performing the hydroxylation and/or the dehydrogenation on the  $\alpha$ ,  $\beta$ -dehydrovaline and -isoleucine and the  $\beta$ -hydroxyvaline, possible derivatives with one to three present  $\alpha,\beta$ -dehydro-amino acids combined with hydroxylated or non-hydroxylated valine were in the focus of the search (see Table 2-13 below).

**Table 2-13** Potential MXV A derivatives searched for in the MxvH deletion mutant DK1622 pBeloBacMXV $\Delta$ *mxvH*.

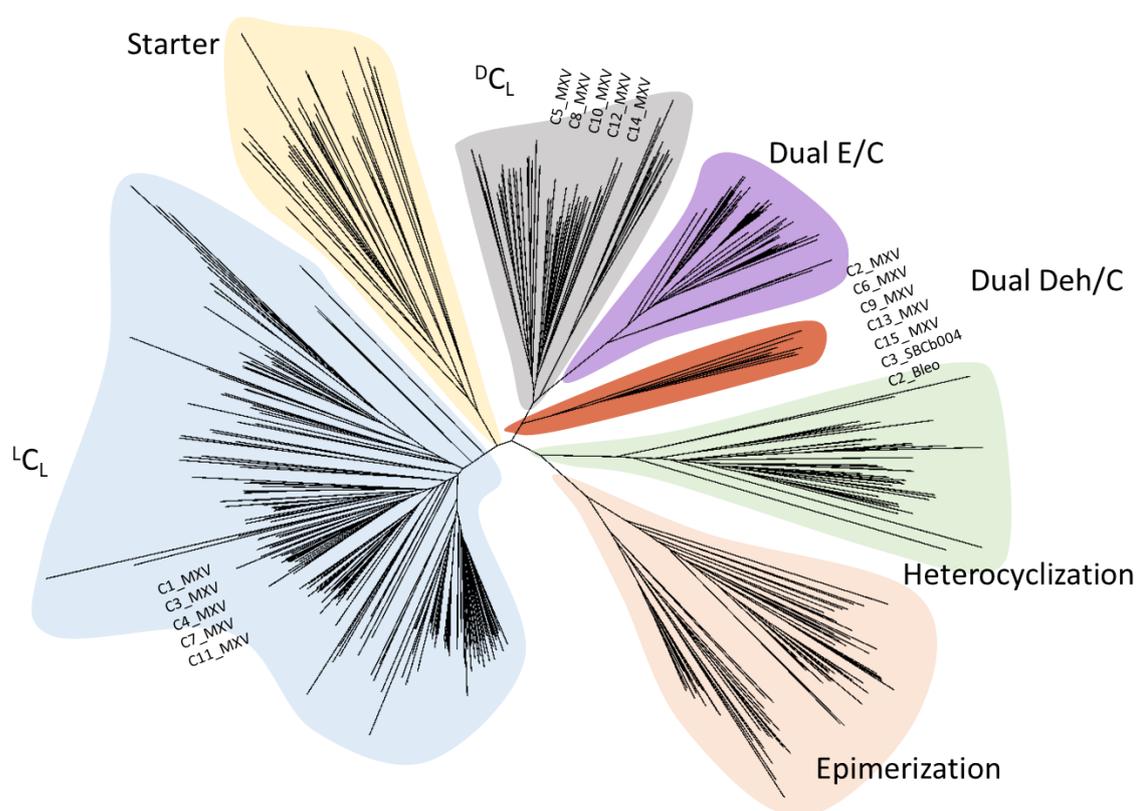
Number of dehydrogenated AAs	Hydroxylated valine	[M + 2H] <sup>2+</sup>	[M + 3H] <sup>3+</sup>
3	Yes	839.03	559.69
	No	831.03	554.36
2	Yes	840.04	560.36
	No	832.04	555.03
1	Yes	841.05	561.03
	No	833.05	555.70
0	Yes	842.06	561.71
	No	834.06	556.37



**Figure 2-10** Structure of MXV A; amino acids potentially altered in the mutant lacking MxvH are marked in red color. The search focused on derivatives where one or multiple of the  $\alpha,\beta$ -dehydro-amino acids are not dehydrogenated, in combination with hydroxylated and non-hydroxylated valine.

The extracts of cultivated *M. xanthus* DK1622::pBeloBac $\Delta$ *mxvH* showed no presence of either MXV A or any of the in Table 2-13 and Figure 2-10 shown potential derivatives. These findings show the importance of MxvH in the MXV production as well as indicate that the putative  $\beta$ -hydroxylase is catalyzing the hydroxylation and/or dehydrogenation on the NRPS assembly line or on the amino acids prior to binding to the PCP domain, *i.e.* the involved C and/or A domains in the corresponding modules seem to have a specificity for hydroxylated and/or dehydrogenated valine and isoleucine. These findings, and new insights in the current literature<sup>157,158</sup> lead to the suggestion, that the corresponding valines and isoleucine are hydroxylated by MxvH and dehydrated by the C domain of the following module. In 2018 Patteson *et al.* described the 2,3-dehydration by a unique group of C domains in the oxyvinylglycine biosynthesis<sup>157</sup> and in 2020 Wang *et al.* found two unusual C domains in the allopeptide biosynthesis, dehydrating serine and threonine residues, respectively.<sup>158</sup> Furthermore, Pogorevc *et al.* describes a new proposed dual Deh/C (dehydration/condensation) domain subtype, forming its own clade in a phylogenetic tree of the different C domain subtypes. This new subtype clade contained a C domain of the argyirin biosynthesis described in this study, as well as C domains from several other  $\alpha,\beta$ -dehydro amino acid forming pathways.<sup>156</sup> Inclusion of the MXV C domains of modules following the  $\alpha,\beta$ -dehydro-amino acids (modules 2, 9, and 13, respectively) in this phylogenetic tree showed that those respective C domains align within the same clade (Figure 2-11). Interestingly, the terminal C domain downstream of module 14 as well as the C domain of module 6 also align in this clade, although the incorporated building block in module 14 is a hydroxy-D-valine and in module 5 a D-valine. It seems that these C domains are closely related to the other dual Deh/C domains. However, the dehydration of the hydroxyl-D-valine at module 14 by the downstream C domain seems to be disabled by its functionality to incorporate agmatine. The C domain of module 6 might have a dual Deh/C functionality as well, but it occurs that the D-valine is not hydroxylated by MxvH, potentially because of missing protein-protein

interactions at this position. Nevertheless, in order to prove the suggested hydroxylation by MxvH and following dehydration by C domains, *in vitro* analysis should be implemented. The plan is to express *mxvH* and two consecutive modules involved in the dehydro-amino acid formation, or in case the expression of whole modules is not successful, standalone domains thereof. The substrates of the modules can be provided as synthetic *N*-acetylcysteamine(SNAc)-esters for binding to the respective PCP domains. Fortunately, the first  $\alpha,\beta$ -dehydro-valine is already incorporated in module 1, *i.e.*, the required substrates that had to be synthesized as SNAcs are the isovaleric acid-valine for module 1 and the isovaleric acid-valine-alanine precursor for module 2



**Figure 2-11** Phylogenetic tree showing the different C domain subtypes including the new proposed Dual Deh/C domain subtype. The phylogenetic tree includes 525 C domains from a phylogenetic study<sup>80</sup>, the C domains of the MXV gene cluster, as well as the Dual Deh/C domains of the argyirin (C3\_SBCb004)<sup>156</sup> and the bleomycin biosynthesis (C2\_Bleo).<sup>159</sup>

#### 2.4.3.2 Functional inactivation of the single PCP domain MxvB

The gene *mxvB* is highly conserved in all MXV producers, however, the function of the single PCP domain expressed by it is yet unknown. It is assumed that MxvB is responsible for the precursor supply to the NRPS assembly line.<sup>127</sup> A length of 513 bp of the gene, while the annotated PCP domain is only 202 bp long, suggests that the gene potentially harbors another

domain of unknown function. Recent bioinformatics analysis of the gene using blastp<sup>160</sup> suggests that the gene contains at least parts of an A-domain (homologue with the best coverage and identity is an amino acid adenylation domain-containing protein, see accession no. Table 2-1). Due to the short lengths, attempts to knockout *mxvB* in the native strain were not successful. With the recent accomplishment of establishing a heterologous expression system with a synthetic MXV BGC, targeted inactivation of the PCP domain of *mxvB* was enabled. The design aimed towards replacing *mxvB* in the synthetic BGC by a point mutated version *mxvB\** using Red/ET. In *mxvB\**, the active site is inactivated by replacing the conserved serine in the core motif<sup>138</sup> by an alanine. It was decided to rather inactivate the gene by a point mutation than deleting it completely, to prevent disrupting potential protein-protein interactions between MxvB and the NRPS assembly line. The detailed design and cloning strategy is described in section 2.3.21 and SI section 2.6.8. At the stage of writing this thesis, the cloning process of the pBeloBACMXV\_*mxvB\** construct was successful and confirmed to be mutation free by Illumina sequencing. Future work involves obtaining DK1622 mutants with that construct and testing their production of MXV. It is expected that the production of MXV would be abolished. In order to confirm the proposed function of MxvB to be involved in the precursor supply, cultures with pBeloBACMXV\_*mxvB\** mutants will be fed with an isovaleric acid SNAc, which should re-establish the MXV production, at least partially.

## 2.5 Conclusion and outlook

The successful development of a heterologous production platform for MXV enabled further investigation of the biosynthesis due to convenient genetic amenability of the heterologous host *M. xanthus* DK1622 and well-developed tools and protocols for micro- and molecular biological work with this strain. Increased yields compared to the native host, with reliable and easy production, facilitates preparation of sufficient MXV for further research on the compound. The production titers varied between 3.5 and 13 mg/L in most cultivation experiments with a peak of 33 mg/L. It was able to be improved by 1.2 mg/L to 7.6 mg/L compared to 6.4 mg/L in the control culture, by increasing of the resistance through additional expression of the ABC transporter proteins. Potential yield limitations could be caused by the assembly rate of the NRPS or substrate saturation, or by metabolic bottlenecks such as precursor supply, which has to be tested in the future with feeding experiments. This newly gained opportunity to engineer the MXV gene cluster in the synthetic BGC and analyze it in the heterologous host also paves the way for elucidating the biosynthesis in more detail. Targeted deletion of the putative  $\beta$ -hydroxylase MxvH abolished production and no other MXV derivatives lacking the dehydro- or hydroxy-residues were found. This finding underlines the vital role of MxvH in the MXV biosynthesis. It is assumed, that MxvH is functioning on the assembly line or on the amino acids prior to incorporation, while the respective modules appear to be specific for the hydroxylated and/or dehydrogenated amino acids. Further *in vitro* analysis with the expressed proteins MxvH, module 1, and module 2, should shed light on the reaction forming the  $\alpha,\beta$ -dehydro- and hydroxy-amino acids. Current hypothesis suggests that MxvH hydroxylates and then the downstream C domain dehydrates the amino acids. In addition, the function of *mxvB* in the MXV biosynthesis was also scrutinized, since *mxvB* is highly conserved in all MXV producers expressing a single PCP domain. Currently it is assumed to be involved in the precursor supply. The gene was inactivated by a point mutation rather than deleting it to prevent effects on protein-protein interactions with the assembly line. To this date, the cloning work to obtain a MXV BGC construct with inactivated *mxvB*\* was successful, however the transformation and production testing has yet to be accomplished. It can be speculated that the production will be abolished, and subsequent feeding of an isovaleryl SNAc substrate would re-establish the production, therefore confirming the hypothesis that MxvB is responsible for the precursor supply.

Additionally, the heterologous production platform for MXV enables engineering of the NRPS assembly line itself. Future experiments aim to analyze the incorporation of the agmatine unit, which is currently assumed to be catalyzed by the terminal C domain of module 14. The platform can also be used for attempts to obtain new MXV derivatives with improved pharmacokinetic properties and lower cytotoxicity. Current ideas include truncated versions of MXV, lacking the agmatine or several terminal residues, which would enable coupling of new residues to the truncated structure by semi synthetic approaches.

## 2.6 SI Information

### 2.6.1 List of plasmids and strains generated in this study

**Table 2-14** List of strains generated or used in this study.

Bacterial strain	Genotype	Reference or source
<i>E. coli</i> DH10 $\beta$	F <sup>-</sup> , <i>mcrA</i> , $\Delta(mrr-hsdRMS-mcrBC)$ , $\Phi 80lacZ\Delta M15$ , $\Delta lacX74$ , <i>recA1</i> , <i>araD139</i> , $\Delta(ara-leu)7697$ , <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 <sup>-</sup> , <i>mcrA</i> , $\Delta(mrr-hsdRMS-mcrBC)$ , $\Phi 80lacZ\Delta M15$ , $\Delta lacX74$ , <i>recA1</i> , <i>araD139</i> , $\Delta(ara-leu)7697$ , <i>galU</i> , <i>galK</i> , <i>rpsL</i> (Str <sup>R</sup> ), <i>endA1</i> , <i>nupG</i> , <i>fhuA::IS2</i>	Invitrogen
<i>E. coli</i> NEB 10 $\beta$	<i>mcrA</i> , <i>spoT1</i> $\Delta(mrr-hsdRMS-mcrBC)$ , $\Phi 80d(lacZ\Delta M15)$ , <i>recA1</i> , <i>relA1</i> , $\Delta lacX74$ , <i>recA1</i> , <i>araD139</i> , $\Delta(ara-leu)7697$ , <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 <sup>-</sup> , <i>mcrA</i> , $\Delta(mrr-hsdRMS-mcrBC)$ , $\Phi 80lacZ\Delta M15$ , $\Delta lacX74$ , <i>recA1</i> , <i>araD139</i> , $\Delta(ara-leu)7697$ , <i>galU</i> , <i>galK</i> , <i>rpsL</i> (Str <sup>R</sup> ), <i>endA1</i> , <i>nupG</i> , $\lambda^{-}$ , $\Delta fhuA$ , P <sub>BAD</sub> - <i>gbaA</i> $\Delta ybcC$ , $\Delta recET19$	Gene Bridges
<i>E. coli</i> DH10 $\beta$ pSynbio1_C1	<i>E. coli</i> DH10 $\beta$ pSynbio1_C1, Amp <sup>R</sup>	This work
<i>E. coli</i> DH10 $\beta$ pSynbio1_C12	<i>E. coli</i> DH10 $\beta$ pSynbio1_C12, Amp <sup>R</sup>	This work
<i>E. coli</i> DH10 $\beta$ pSynbio1_C123	<i>E. coli</i> DH10 $\beta$ pSynbio1_C123, Amp <sup>R</sup>	This work
<i>E. coli</i> DH10 $\beta$ pSynbio1_C1234	<i>E. coli</i> DH10 $\beta$ pSynbio1_C1234, Amp <sup>R</sup>	This work
<i>E. coli</i> NEB10 $\beta$ pSynbio1_mxC	<i>E. coli</i> NEB10 $\beta$ pSynbio1_mxC, Amp <sup>R</sup>	This work
<i>E. coli</i> DH10 $\beta$ pSynbio1_D1	<i>E. coli</i> DH10 $\beta$ pSynbio1_D1, Amp <sup>R</sup>	This work
<i>E. coli</i> DH10 $\beta$ pSynbio1_D12	<i>E. coli</i> DH10 $\beta$ pSynbio1_D12, Amp <sup>R</sup>	This work
<i>E. coli</i> NEB10 $\beta$ pSynbio1_mxD	<i>E. coli</i> NEB10 $\beta$ pSynbio1_mxD, Amp <sup>R</sup>	This work
<i>E. coli</i> DH10 $\beta$ pMYC24	<i>E. coli</i> DH10 $\beta$ pMYC24, Amp <sup>R</sup> , Tet <sup>R</sup> , Cm <sup>R</sup>	This work

<i>E. coli</i> DH10 $\beta$ pMYC24PvanFA1	<i>E. coli</i> DH10 $\beta$ pMYC24PvanFA1, Amp <sup>R</sup> , Tet <sup>R</sup> , Cm <sup>R</sup>	This work
<i>E. coli</i> NEB10 $\beta$ pMYC24PvanFA12	<i>E. coli</i> DH10 $\beta$ pMYC24PvanFA12, Amp <sup>R</sup> , Tet <sup>R</sup> , Cm <sup>R</sup>	This work
<i>E. coli</i> NEB10 $\beta$ pMYC24PvanFAEtD1	<i>E. coli</i> DH10 $\beta$ pMYC24PvanFAEtD1, Amp <sup>R</sup> , Tet <sup>R</sup> , Cm <sup>R</sup>	This work
<i>E. coli</i> DH10 $\beta$ pJET Amp <sup>R</sup>	<i>E. coli</i> DH10 $\beta$ pJET Amp <sup>R</sup>	This work
<i>E. coli</i> DH10 $\beta$ pMYC20	<i>E. coli</i> DH10 $\beta$ pMYC20, Tet <sup>R</sup> , Cm <sup>R</sup>	This work
<i>E. coli</i> NEB10 $\beta$ pMYC24MXVpre	<i>E. coli</i> NEB10 $\beta$ pMYC24MXVpre, Amp <sup>R</sup> , Tet <sup>R</sup> , Cm <sup>R</sup>	This work
<i>E. coli</i> HS996 pBeloBacMXVpre	<i>E. coli</i> HS996 pBeloBacMXVpre, Tet <sup>R</sup> , Cm <sup>R</sup>	This work
<i>E. coli</i> HS996 pBeloBacMXVpreC	<i>E. coli</i> HS996 pBeloBacMXVpreC, Tet <sup>R</sup> , Cm <sup>R</sup>	This work
<i>E. coli</i> HS996 pBeloBacMXVpreCD	<i>E. coli</i> HS996 pBeloBacMXVpreCD, Tet <sup>R</sup> , Cm <sup>R</sup>	This work
<i>E. coli</i> HS996 pBeloBacMXV	<i>E. coli</i> HS996 pBeloBacMXV, Tet <sup>R</sup> , Cm <sup>R</sup>	This work
<i>E. coli</i> HS996 pMYC21 <i>mxvIK</i>	<i>E. coli</i> HS996 pMYC21 <i>mxvIK</i> , Kan <sup>R</sup> , Cm <sup>R</sup>	This work
<i>E. coli</i> HS996 pBeloBacMXV $\Delta$ <i>mxvH_kanR</i>	<i>E. coli</i> HS996 pBeloBacMXV $\Delta$ <i>mxvH_kanR</i> , Tet <sup>R</sup> , Cm <sup>R</sup> , Kan <sup>R</sup>	This work
<i>E. coli</i> HS996 pBeloBacMXV $\Delta$ <i>mxvH</i>	<i>E. coli</i> HS996 pBeloBacMXV $\Delta$ <i>mxvH</i> , Tet <sup>R</sup> , Cm <sup>R</sup>	This work
<i>E. coli</i> BL21 (DE3) pHis-TEV_ <i>mxv</i> AMod1 pCDF-1b_ <i>mxvG</i>	<i>E. coli</i> DH10 $\beta$ pHis-TEV_ <i>mxv</i> AMod1 pCDF-1b_ <i>mxvG</i> , Kan <sup>R</sup>	This work
<i>E. coli</i> BL21 (DE3) pHis-TEV_ <i>mxv</i> AMod2 pCDF-1b_ <i>mxvG</i>	<i>E. coli</i> DH10 $\beta$ pHis-TEV_ <i>mxv</i> AMod2 pCDF-1b_ <i>mxvG</i> , Kan <sup>R</sup>	This work
<i>E. coli</i> BL21 (DE3) pHis-TEV_ <i>mxvH</i>	<i>E. coli</i> DH10 $\beta$ pHis-TEV_ <i>mxvH</i> , Kan <sup>R</sup>	This work

<b><i>E. coli</i> HS996 pBeloBacMXVΔmxvB_ka nR</b>	<i>E. coli</i> HS996 pBeloBacMXVΔmxvB_ka <sup>R</sup> , Tet <sup>R</sup> , Cm <sup>R</sup> , Kan <sup>R</sup>	This work
<b><i>E. coli</i> HS996 pBeloBacMXVΔmxvB_m xvB*</b>	<i>E. coli</i> HS996 pBeloBacMXVΔmxvB_mxvB*, Tet <sup>R</sup> , Cm <sup>R</sup> , Kan <sup>R</sup>	This work
<b><i>M. xanthus</i> DK1622</b>	-	HIPS/MINS
<b><i>M. xanthus</i> DK1622 pBeloBacMXV</b>	<i>M. xanthus</i> DK1622 pBeloBacMXV, Otc <sup>R</sup>	This work
<b><i>M. xanthus</i> DK1622 pBeloBacMXV ΔmxvH</b>	<i>M. xanthus</i> DK1622 pBeloBacMXV ΔmxvH, Otc <sup>R</sup>	This work
<b><i>M. xanthus</i> DK1622 pBeloBacMXV pMYC21 mxvIK</b>	<i>M. xanthus</i> DK1622 pBeloBacMXV pMYC21 mxvIK, Otc <sup>R</sup> , Kan <sup>R</sup>	This work

**Table 2-15** List of plasmids generated or used in this study.

Plasmid	Genotype	Reference
<b>pSynbio1</b>	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	<sup>143</sup>
<b>pMYC20</b>	TetR-mx8 cloned into pMYC; <i>tetR</i> (Otc <sup>R</sup> ) from pALTER(R)-1, Mx8 integrase from <i>Myxococcus</i> phage Mx8	<sup>71</sup>
<b>pMYC21</b>	KanR-mx9 cloned into pMYC; <i>kanR</i> (Kan <sup>R</sup> ) from pACYC177, Mx9 integrase from <i>Myxococcus</i> phage Mx9	<sup>71</sup>
<b>pMYC24</b>	Replacement of <i>ura3</i> by Amp <sup>R</sup>	This work
<b>pJET1.2</b>	pUC ori and <i>rep</i> from pMB1, P <sub>lacUV5-eco471</sub> (endonuclease)/T7 promoter-MCS, <i>bla</i> (Amp <sup>R</sup> )	Thermo Fisher Scientific
<b>pBeloBac</b>	<i>tetR</i> (Tet <sup>R</sup> ), Mx8 integrase from <i>Myxococcus</i> phage Mx8, CEN6/ARS4, <i>traJ/oriT</i> , <i>CmR</i> (Cm <sup>R</sup> ), <i>sopA</i> , <i>sopB</i> , <i>sopC</i> , <i>repE</i>	<sup>144</sup> , adapted in HIPS/MINS

<b>pHis-TEV</b>	<i>ori</i> (pBR322), <i>lac</i> operator, <i>neo</i> , T7 promoter, ribosome binding site, N-HisTag	Novagen
<b>pCDF-1b</b>	<i>ori</i> (CloDF13), <i>lac</i> operator, <i>spect(variant)</i> , T7 promoter, ribosome binding site, HisTag	Novagen
<b>pSynbio1_C1</b>	Gene synthesis product C1 cloned into pSynbio1, C1 is the first part of <i>mxvC</i>	This work
<b>pSynbio1_C12</b>	Gene synthesis product C2 cloned into pSynbio1_C1, C2 is the second part of <i>mxvC</i>	This work
<b>pSynbio1_C123</b>	Gene synthesis product C3 cloned into pSynbio1_C12, C3 is the third part of <i>mxvC</i>	This work
<b>pSynbio1_C + SE</b>	Gene synthesis product C4 cloned into pSynbio1_C123, C4 is the fourth part of <i>mxvC</i> , complete <i>mxvC</i> including SEs	This work
<b>pSynbio1_<i>mxvC</i></b>	pSynbio1_C + SE without SEs	This work
<b>pSynbio1_D1</b>	Gene synthesis product D1 cloned into pSynbio1, D1 is the first part of <i>mxvD</i>	This work
<b>pSynbio1_D + SE</b>	Gene synthesis product D2 cloned into pSynbio1_D1, D2 is the second part of <i>mxvD</i> , complete <i>mxvD</i> including SEs	This work
<b>pSynbio1_<i>mxvD</i></b>	pSynbio1_D + SE without SEs	This work
<b>pMYC24PvanFA1</b>	Gene synthesis product PvanFA1 cloned into pMYC24, containing P <sub>van</sub> , <i>mxvF</i> and first part of <i>mxvA</i>	This work
<b>pMYC24PvanFA</b>	Gene synthesis product A2 cloned into pMYC24PvanFA1, A2 is the second part of <i>mxvA</i> , complete <i>mxvA</i>	This work

<b>pMYC24PvanFAfrag-EtD1</b>	Gene synthesis product frag-EtD1 cloned into pMYC24PvanFA, <i>mxvE</i> and tD1 terminator	This work
<b>pMYC24MXVpre + SE</b>	Gene synthesis product PvanGHIJKtD1 cloned into pMYC24PvanFAfrag-EtD1, P <sub>van</sub> , <i>mxvGHIJK</i> , tD1, including SEs	This work
<b>pBeloBacMXVpre + SE</b>	pMYC24 vector backbone replaced by pBeloBac vector backbone in pMYC24MXVpre + SE	This work
<b>pBeloBacMXVpreC + SE</b>	<i>mxvC</i> cloned into pBeloBacMXVpre + SE	This work
<b>pBeloBacMXVpreCD + SE</b>	<i>mxvD</i> cloned into pBeloBacMXVpreC + SE	This work
<b>pBeloBacMXV</b>	SE free MXV gene cluster on pBeloBac vector backbone	This work
<b>pBeloBacMXV <math>\Delta</math><i>mxvH_kanR</i></b>	<i>mxvH</i> replaced by <i>kanR</i> in pBeloBacMXV	This work
<b>pBeloBacMXV <math>\Delta</math><i>mxvH</i></b>	<i>kanR</i> deleted in pBeloBacMXV $\Delta$ <i>mxvH_kanR</i>	This work
<b>pBeloBacMXV<math>\Delta</math><i>mxvB_kanR</i></b>	<i>mxvB</i> replaced by <i>kanR</i> in pBeloBacMXV	This work
<b>pBeloBacMXV<math>\Delta</math><i>mxvB_mxvB*</i></b>	<i>kanR</i> replaced by <i>mxvB*</i> in pBeloBacMXV $\Delta$ <i>mxvB_kanR</i>	This work
<b>pHis-TEV_<i>mxvA</i>Mod1</b>	PCR fragment module 1 von <i>mxvA</i> cloned into pHis-TEV	This work
<b>pHis-TEV_<i>mxvA</i>Mod2</b>	PCR fragment module 2 von <i>mxvA</i> cloned into pHis-TEV	This work
<b>pHis-TEV_<i>mxvH</i></b>	PCR fragment <i>mxvH</i> cloned into pHis-TEV	This work
<b>pCDF-1b_<i>mxvG</i></b>	PCR fragment <i>mxvG</i> cloned into pCDF-1b	This work

## 2.6.2 Predicted domain specificity of the MXV gene cluster modules

The domain specificities were analyzed before by Ullrich Scheid.<sup>127</sup> Since the analysis was performed approx. 10 years ago, the specificities were re-evaluated by analyzing the native MXV gene cluster of *M. fulvus* Mx65 again (re-evaluation performed on November 15<sup>th</sup> 2021 using antiSMASH version 6.0.1). The re-evaluation revealed no discrepancies to the previous collected data and is shown in Table 2-16 below.

**Table 2-16** Predicted domain specificity in the MXV gene cluster. C = condensation domain; A = adenylation domain; PCP = peptide carrier protein; MT = methyltransferase; DcL = condensation of upstream D amino acid with downstream L amino acid; LcL = condensation of upstream L amino acid with downstream L amino acid. Table adapted from Ullrich Scheid.

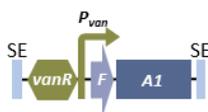
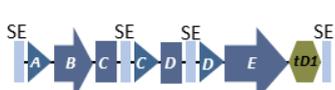
Gene	Module	Domain order	Domain specificity	Phenotype in Myxovalargin A
<b><i>mxvA</i></b>	Module 1	C-A-PCP	C domain: LcL A domain: Val	dehydro - valine
	Module 2	C-A-PCP	C domain: DcL A domain: Ala	L-alanine
	Module 3	C-A-MT-PCP	C domain: LcL A domain: Ala	N-methyl-L-alanine
<b><i>mxvB</i></b>	-	PCP	-	-
<b><i>mxvC</i></b>	Module 4	C-A-PCP-E	C domain: LcL A domain: Val	D-valine
	Module 5	C-A-PCP-E	C domain: DcL A domain: Val	D-valine
	Module 6	C-A-PCP	C domain: DcL A domain: $\beta$ -tyrosine	(S)- $\beta$ -tyrosine
	Module 7	C-A-PCP-E	C domain: LcL A domain: Val	D-valine
	Module 8	C-A-PCP	C domain: DcL A domain: Val	dehydro - valine
	Module 9	C-A-PCP-E	C domain: DcL A domain: (Gln)	D-arginine
	<b><i>mxvD</i></b>	Module 10	C-A-PCP	C domain: DcL

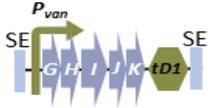
		A domain: Val	
Module 11	C-A-PCP-E	C domain: LcL	D-valine
		A domain: Val	
Module 12	C-A-PCP	C domain: DcL	dehydro-isoleucine
		A domain: Ile	
Module 13	C-A-PCP-E	C domain: DcL	D-alanine
		A domain: Ala	
<b>mxvE</b>	Module 14	C-A-PCP-E	C domain: DcL
		A domain: Val	
-	C	C domain: DcL	-

### 2.6.3 Organization of the synthetic MXV gene synthesis fragments

The table below gives detailed information about the ten synthesized MXV gene fragments including their size, organization of the gene construct, the respective SEs, and corresponding restriction endonucleases for their release from the gene synthesis vector (pUC57).

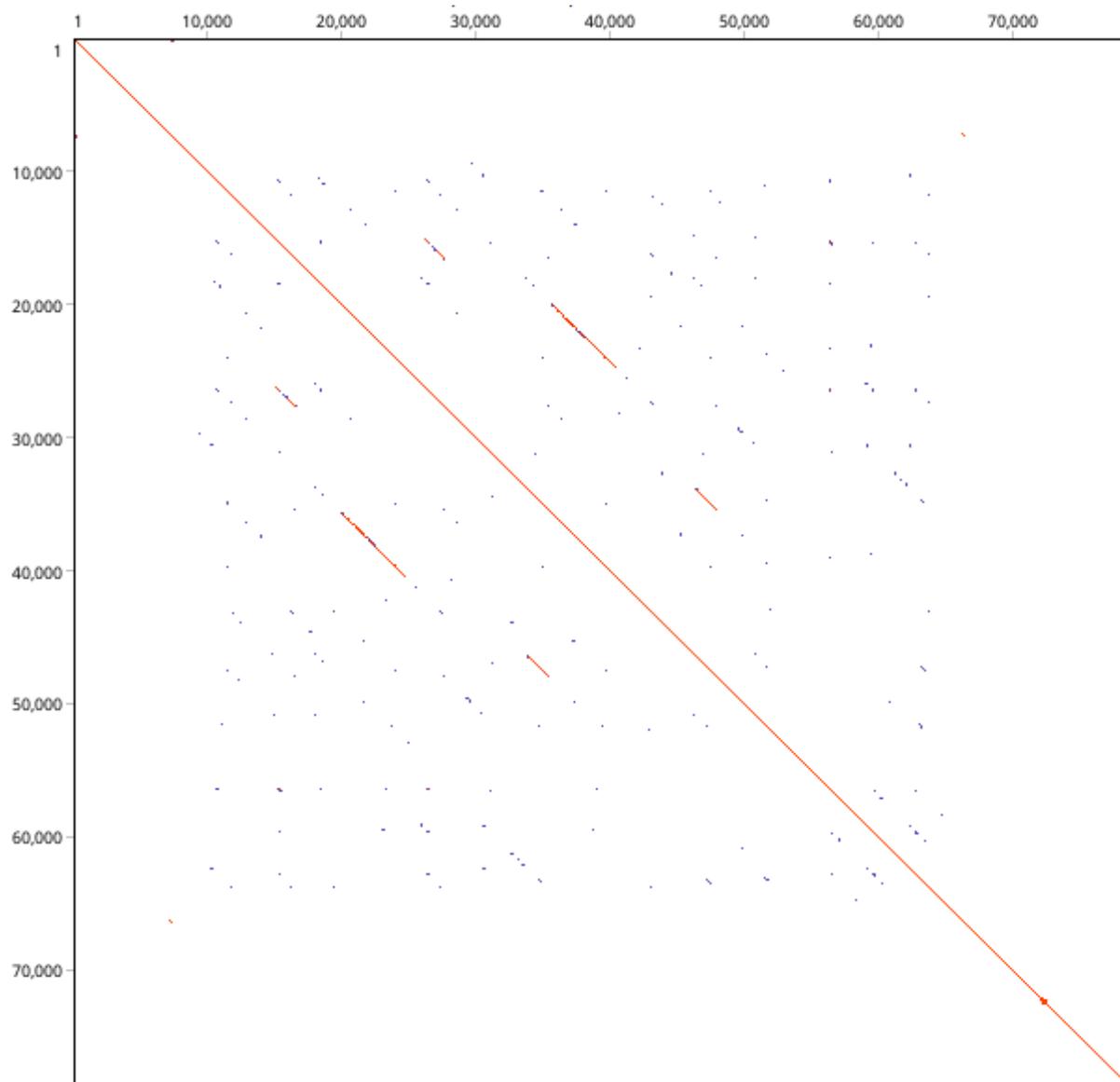
**Table 2-17** The ten synthetic MXV fragments with their respective size, gene organization, SEs and RE sites for release. Sp = spacer; SE = splitter element; RE = restriction endonuclease;  $P_{van}$  = vanillate promoter; *vanR* = vanillate repressor;  $t_{D1}$  =  $t_{D1}$  terminator.

Synthetic fragment	Size [kbp]	Organization	SEs with containing REs	REs for release
<b>PvanFA1</b>	6.338		5'-EcoRV-Bsal- ...-Bsal-HindIII- sp-Nsil-3'	EcoRV/Nsil
<b>A2</b>	6.453		5'-HindIII-Bsal- ...-Bsal-BglII-sp- Nsil-3'	HindIII/Nsil
<b>Frag-EtD1</b>	7.166		5'-BglII-Bsal-...- Bsal-XbaI-sp- PvuI-Bsal-...- Bsal-AvrII-sp- BspTI-Bsal-...- Bsal-SgrDI-sp- Nsil-3'	BglII/Nsil

<b>PvanGHIJKtD1</b>	7.156		5'-SgrDI-Bsal- ...-Bsal-Nsil-3'	SgrDI/Nsil
<b>C1</b>	6.296		5'-PvuI-XbaI-Bsal-...-Bsal-EcoRI-sp-PmeI-3'	PvuI/PmeI
<b>C2</b>	6.294		5'-EcoRI-Bsal-...-Bsal-HindIII-sp-PmeI-3'	EcoRI/PmeI
<b>C3</b>	5.944		5'-HindIII-Bsal-...-Bsal-AvrII-sp-PmeI-3'	HindIII/PmeI
<b>C4</b>	6.484		5'-AvrII-Bsal-...-Bsal-PvuI-sp-PmeI-3'	AvrII/PmeI
<b>D1</b>	7.851		5'-PvuI-AvrII-Bsal-...-Bsal-HindIII-sp-PmeI-3'	PvuI/PmeI
<b>D2</b>	7.711		5'-HindIII-Bsal-...-Bsal-BspTI-PmeI-3'	HindIII/PmeI

#### 2.6.4 Location of repetitive sequences

The locations of the three repetitive sequences in the final MXV BGC: first repetitive sequence on *mxvC* base pairs 33449 – 35530 and base pairs 45933 – 48173; second repetitive sequence on *mxvC* base pairs 35370 – 40331 and *mxvD* base pairs 19366 – 24647; third repetitive sequence on *mxvD* base pairs 14885 – 16989 and base pairs 25679 – 27689.



**Figure 2-12** Dotplot display of the MXV gene cluster against itself on Geneious software. The three repetitive sequences (short red lines) are clearly visible.

### 2.6.5 Detailed assembling strategy of the synthetic MXV BGC

The ten synthesized MXV DNA fragments (listed in the table below in the column “insert”) were assembled on three separate plasmids (pMYC24MXVpre, pSynbio1*mxvC* and pSynbio1*mxvD*). The cloning procedure, including desplitting steps and final cloning of *mxvC* and *mxvD* into the pMYC24MXVpre shuttle vector, is summarized in the table below.

**Table 2-18** Detailed assembly of the plasmids pMYC24MXVpre, pSynbio1C, pSynbio1D and the final construct pBeloBacMXV.

Vector	Insert	Restriction enzymes	Formed plasmid
<b>pMYC24</b>	PvanFA1	<i>EcoRV/Nsil</i>	pMYC24PvanFA1
<b>pMYC24PvanFA1</b>	A2	<i>HindIII/Nsil</i>	pMYC24PvanFA
<b>pMYC24PvanFA</b>	frag-EtD1	<i>BglII/Nsil</i>	pMYC24PvanFAfrag-EtD1
<b>pMYC24PvanFAfrag-EtD1</b>	PvanGHIJKtD1	<i>SgrDI/Nsil</i>	pMYC24MXVpre
<b>pSynbio1</b>	C1	<i>PvuI/PmeI</i>	pSynbio1C1
<b>pSynbio1C1</b>	C2	<i>EcoRI/PmeI</i>	pSynbio1C12
<b>pSynbio1C12</b>	C3	<i>HindIII/PmeI</i>	pSynbio1C123
<b>pSynbio1C123</b>	C4	<i>AvrII/PmeI</i>	pSynbio1C1234 = pSynbio1C + SE
<b>pSynbio1C + SE</b>	-	<i>BsaI</i>	pSynbio1mxvC
<b>pSynbio1</b>	D1	<i>PvuI/PmeI</i>	pSynbio1D1
<b>pSynbio1D1</b>	D2	<i>HindIII/PmeI</i>	pSynbio1D + SE
<b>pSynbio1D + SE</b>	-	<i>BsaI</i>	pSynbiomxvD
<b>pMYC24MXVpre</b>	pBeloBac	<i>EcoRV/EcoRI</i>	pBeloBacMXVpre
<b>pBeloBacMXVpre</b>	<i>mxvC</i>	<i>XbaI/PvuI</i>	pBeloBacMXVpreC
<b>pBeloBacMXVpreC</b>	<i>mxvD</i>	<i>AvrII/BspTI</i>	pBeloBacMXVpreCD + SE
<b>pBeloBacMXVpreCD + SE</b>	-	<i>BsaI</i>	pBeloBacMXV

### 2.6.6 Analysis of plasmids after the final desplitting process

After the final desplitting step, an extensive analysis of the plasmids from the screened clones had to be pursued. Therefore, the plasmids were first digested with *KpnI* to confirm the right size, and further digested with multiple enzymes investigating the presence or absence of each splitter element. Since one of the DNA fragments that had to be ligated could be inserted in forward or reverse direction, an additional PCR was done to confirm the inserted direction. The details of this screening process are shown in the three tables below.

**Table 2-19** This table shows nine out of 144 screened clones with analytical digest for right plasmid size (*KpnI*) and each splitter element (SE1-SE6), as well as PCR with primer pair “R” and “F” (see Table 2-20) to confirm the correct direction of the 200 bp fragment located between NRPS genes *mxvC* and *mxvD*.

Clone	<i>KpnI</i>	SE 1	SE 2	SE 3	SE 4	SE 5	SE 6	Primer pair “R”	Primer pair “F”
31	✓	✓	✓	✓	✓	✓	✓		✓
70	✓	✓	✓	✓	✓	✓	✓	✓	
76	✓	✓	✓	✓	✓	✓	✓	✓	
85	✓	✓	✓	✓	✓	✓	✓	✓	
89	✓	✓	✓	✓	✓	✗	✓	✓	
95	✓	✓	✓	✗	✓	✗	✓	✓	
117	✓	✓	✓	✗	✓	✗	✗	✓	
120	✓	✓	✓	✓	✓	✓	✓	✓	
122	✓	✓	✓	-	-	-	-		✓

**Table 2-20** Sequences of the two primer pairs used to detect the direction of the 200 bp fragment after religation by PCR. R=right=forward direction; F=false=reverse direction. Primer sequences are written in 5’ to 3’ direction.

<b>Primer pair “R” (right position of fragment)</b>	Forward primer	“fragCD R fwd”	CATGAGCTGCTCGACATAGG
	Reverse primer	“frag CD R rev”	CGACCTACAACGACGGCG
<b>Primer pair “F” (false position of fragment)</b>	Forward primer	“fragCD F fwd”	GCATGAGCTGCTCGACG
	Reverse primer	“fragCD F rev”	CCTACAACGACATAGGC

**Table 2-21** Additional information about the splitter elements of pBeloBacMXVpreCD: location on the plasmid, the restriction endonuclease enzyme pair used for analytical digest (RE pair; referring to SE pair in Table 2-19) including the used buffer, and the fragment sizes that would appear in the agarose gel in case the respective SE is still present or not.

Splitter element location in plasmid (bp) (Restriction Enzyme (RE) pair)	Enzyme(s) and buffer	Fragment sizes with SE (in kbp)	Fragments sizes w/o SE (in kbp)
49730 - 49751 (RE pair 1)	<i>PvuI</i> and <i>EcoRI</i> , buffer red	37.5 25.5 9.0 6.7	37.5 25.4 15.7
49953 - 49972 (RE pair 2)	<i>AvrII</i> and <i>EcoRI</i> , buffer red	37.5 25.5 8.8 6.9	37.5 25.4 15.7
65450 - 65469 (RE pair 3)	<i>BspTI</i> and <i>NdeI</i> , buffer orange	53.0 19.5 6.3	59.1 19.5
71583 - 71603 (RE 4)	<i>SgrDI</i> , buffer red	34.1 16.6 15.8 12.3	50.6 15.7 12.3
17651 - 17670 and 24086 - 24105 (RE pair 5)	<i>HindIII</i> and <i>BglII</i> , buffer red	63.0 7.2 6.4 2.0	76.5 2.0
24871 - 24890 (RE pair 6)	<i>XbaI</i> and <i>NdeI</i> , buffer 2xTango	46.9 19.5 12.5	59.1 19.5

## 2.6.7 Integration of the synthetic gene cluster into *Myxococcus xanthus* DK1622 via Mx8 phage-integrase

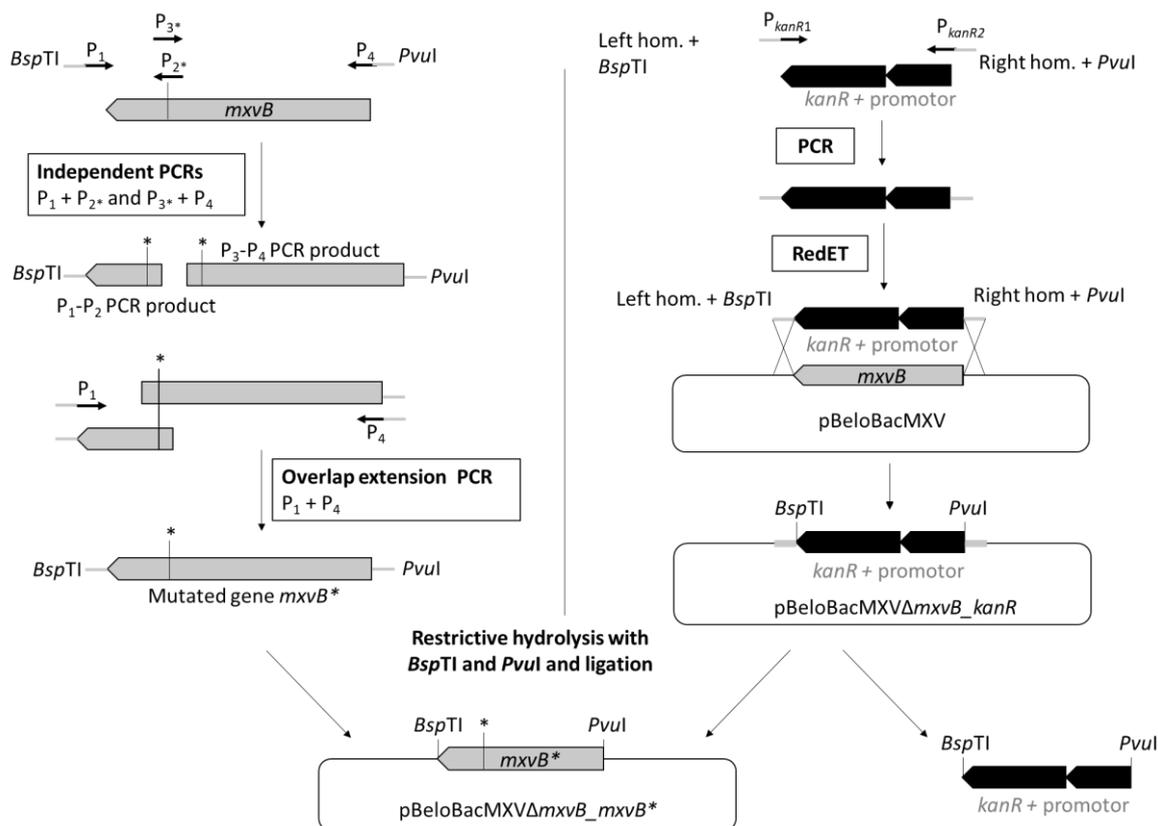
**Table 2-22** Primers used to confirm the integration of the MXV BGC into the Mx8 site of *M. xanthus* DK1622.

Primer name	Primer sequence 5' – 3'
MXVA fwd	GAAGTTCGTCTGCTCGAAGATGCGC
MXVA rev	AGCTTCACCATCCAGCAGCTCTTCA
MXVB fwd	GACGAACTCCGCGACGTTCA
MXVB rev	GCGTCGTTGAGACGTTGAGTTC
MXVC1 fwd	TCTCCGTCGACTGTCAGGAG
MXVC1 rev	GAAGGTGCTCAGCGTGCTGT
MXVC2 fwd	CACGTAGTCCAGCTCACCCGTCG
MXVC2 rev	CGCATGCCGACGCATTGGCTGTC
MXVD1 fwd	TGGACGAGCTGCAGCGGAATGGA
MXVD1 rev	GTGCCTATCGGTCTGCCATTGG
MXVD2 fwd	TAGCTGCATTGCGCTCCGCTCTG
MXVD2 rev	TTCCCCATGTCCTTCGCGCAGGA
MXVE fwd	GTGAGCTGTTCCGTCCAGTACCC
MXVE rev	CATGGACTGCTGCGCTATCTGTC

## 2.6.8 Functional inactivation of the active site in *mxvB*



**Figure 2-13** Alignment of the fourteen PCP domains of the MXV gene cluster. The alignment shows the core motif LGGDSI and reveals the conserved active Ser residue.



**Figure 2-14** Detailed scheme of the design to amplify the mutated gene *mxvB\** by PCR and replacement of *mxvB*. Left side: Amplification of *mxvB\** by three separate PCRs; two PCRs with primer pairs  $P_1$  and  $P_2^*$ , and primer pairs  $P_3^*$  and  $P_4$  to form two fragments, each containing the point mutation and overlapping ends. The third overlapping extension PCR using the two overlapping fragments including the point mutation and primer pair  $P_1$  and  $P_4$  to amplify the final construct *mxvB\**. Right side: Amplification of *kanR* with flanking homologous regions to replace *mxvB* by Red/ET. Both fragments were flanked by RE recognition sites *BspTI* and *PvuI* for replacement of *kanR* by *mxvB\** to form the final construct *pBeloBacMXVΔmxvB\_mxvB\**. Figure adapted by Fabienne Wittling.

**Table 2-23** Primers used to form *pBeloBacMXVΔmxvB\_mxvB\**.

Primer name	Primer sequence 5'–3'
<b><math>P_1</math> = <i>mxvB</i> PM fwd</b>	GTGTCTTAAGTCAGCGACCCAGGGGGCGCG
<b><math>P_2^*</math> = <i>mxvB</i> PM down rev</b>	CGGTGACGCCGCGCGCGC
<b><math>P_3^*</math> = <i>mxvB</i> PM up fwd</b>	GCGCGCGCGGCCTCACCG
<b><math>P_4</math> = <i>mxvB</i> PM rev</b>	ACACCGATCGGGGAATGGCCATGTCCGTATCCG
<b><math>P_{kanR1}</math> = <i>mxvB</i> <i>kanR</i> fwd</b>	GACGTTTCATGGCTTTCCGCTCTCCAGTGCCTACGTCGGTACCGTTTCCGCTTAAGTCA GAAGAACTCGTCAAGAAGGC
<b><math>P_{kanR2}</math> = <i>mxvB</i> <i>kanR</i> rev</b>	CCGAGCAGGTCGAGTCCGCGTTGGCGTCGTTCGAGACGTTTCGAGTTCTAGCGATCGTG GACAGCAAGCGAACC



### 3 Investigation of the disciformycin biosynthesis

#### Contributions

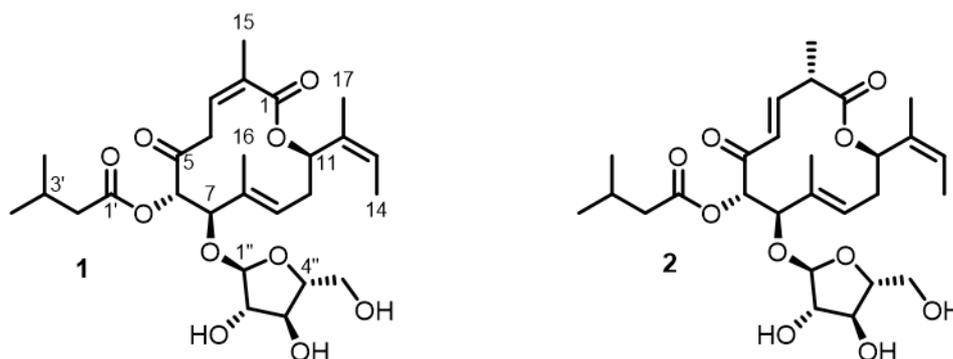
The investigation of the two putative glycosyltransferase operons and the investigation of the putative tailoring enzyme DifH by heterologous expression in *M. xanthus* DK1622 was performed by the author of this thesis. Expression, purification, and *in vitro* analysis of the cytochrome P450 DifA, AndGT8 ferredoxin and ferredoxin reductase was previously carried out by Dr. Konrad Viehrig. The genes for glycosyltransferase operon 1 were discovered by Dr. Konrad Viehrig and the genes for glycosyltransferase operon 2 were discovered by Dr. Fabian Panter. RhFRED expression plasmid (pet28RhFRED and pet28RhFRED-fdx with His-tag) was obtained from the laboratory of Prof. David Sherman. CO difference spectrometry was performed by Dr. Konrad Viehrig and Yogan Khatri in the laboratory of Prof. Bernhard, Biochemie, Universität des Saarlandes. The Pdx protein was obtained from Yogan Khatri (AG Prof. Bernhard).

### 3.1 Abstract

Disciformycins (DIF) are a novel macrolide antibiotic compound class with promising potential to treat MRSA/VRSA pathogens. Since no cross-resistance to other antibiotics was found, it is suggested that this molecule represents a novel mechanism and/or antibiotic class.<sup>61</sup> Intense efforts, initiated by high interest in this molecule, led to the successful total synthesis of DIF.<sup>161,162</sup> The biosynthesis of DIF is not fully elucidated to this date. Bioinformatic analysis revealed a polyketide synthase (PKS) biosynthetic gene cluster (BGC) encoding six PKS genes *difBCDEFG*, with adjacent tailoring genes *difA*, *difH* and *difI*, suggested to be responsible for the DIF production in the native producer *Pyxidicoccus fallax* AndGT8.<sup>61</sup> The proposed biosynthesis by Surup *et al.*, however, is only partially confirmed so far. Since the native producer was not amenable to genetic manipulation, and thus hinders biosynthesis elucidation by targeted gene deletions, heterologous expression in the model strain *Myxococcus xanthus* DK1622 of the DIF BGC was chosen to enable further investigation. Heterologous expression of the PKS genes *difBCDEFG* in *Myxococcus xanthus* DK1622 lead to the production of the DIF polyketide scaffold, aglycon DIF294, and therefore confirming their proposed function in the DIF biosynthesis.<sup>128</sup> *In vitro* analysis of DifA revealed its potential functionality as a P450 monooxygenase, consistent with its predicted role in hydroxylation at C-6,<sup>128</sup> but no hydroxylation of the aglycon DIF294 as substrate was detected *in vitro*. Most likely, the polyketide scaffold has to be glycosylated prior to recognition as substrate by DifA, in a mechanism shown similar to the Pikromycin biosynthesis.<sup>163</sup> Furthermore, DifH, a putative Fe-S oxidoreductase, was confirmed as a tailoring enzyme in the disciformycin biosynthesis, forming the C-C double bond of DIF at C-3. Two major tailoring steps of the biosynthesis, the glycosylation with a  $\alpha$ -arabinofuranose at C-7 and the acylation of the hydroxy group at C-6, remain to be elucidated. The here discovered results, however, are already beneficial for future efforts on investigating the complete biosynthesis of DIF.

### 3.2 Introduction

Disciformycin (DIF) A (**1**) and B (**2**) are two 12-membered macrolide-glycoside antibiotics isolated from the myxobacterium *Pyxidicoccus fallax* AndGT8.<sup>61</sup> They exhibit strong activity against methicillin- and vancomycin-resistant *Staphylococcus aureus* (MRSA/VRSA) strains while showing very low cytotoxicity, making them promising candidates for development. Their structures and absolute configurations are shown in Figure 3-1.



**Figure 3-1** Structures of disciformycin A (**1**) and B (**2**)

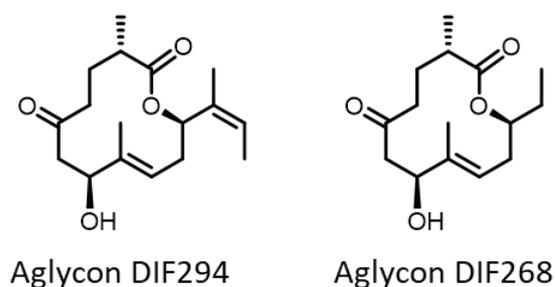
Disciformycins, with **2** being the more active derivative, show strong activity against Gram-positive bacteria, but more importantly the compound revealed to be especially active against several MRSA/VRSA strains. Table 3-1 shows that the MICs of **2** range around values comparable to vancomycin, which is currently used as last resort antibiotic in human therapy. Furthermore, low cytotoxicity of **1** and **2** was observed at concentrations even higher than 10  $\mu\text{M}$ , which increases its potential towards development of a new antibiotic. In addition to the promising MICs and low cytotoxicity, no cross-resistance with other antibiotic classes was found, which indicates a putative novel mechanism of action for this compound.<sup>61</sup> This caused high interest in further investigation of the DIFs, which eventually led to the successful total synthesis of **2** by Waser *et al.* and Kwon *et al.*<sup>161,162</sup>

**Table 3-1** MIC values [ $\mu\text{g}/\text{mL}$ ] for selected bacteria and  $\text{IC}_{50}$  values [ $\mu\text{M}$ ] for mammalian cell lines of disciformycin A (1), disciformycin B (2), and vancomycin. Adapted from Surup *et al.*

	1	2	Vancomycin
<i>Bacillus subtilis</i> DSM-10	4.2	0.8	0.25
<i>Nocardioides simplex</i> DSM-20130	33.3	16.6	0.42
<i>Paenibacillus polymyxa</i> DSM-36	16.6	16.6	-
<i>Staphylococcus camosus</i> DSM-20501	7.8	2.4	0.25
<i>Staphylococcus aureus</i> Newman	8.0	1.2	0.5
<i>Staphylococcus aureus</i> DSM-11822 (MRSA)	4.0	0.6	1.0
<i>Staphylococcus aureus</i> N315 (MRSA)	8.0	1.2	1.0
<i>Staphylococcus aureus</i> Mu50 (MRSA/VRSA)	2.0	0.6	16.0
Colon carcinoma cells HCT-116	>10	>10	-
Murine fibroblast-like cells L929	>10	>10	-
Chinese hamster ovary cells CHO-K1	>10	>10	-

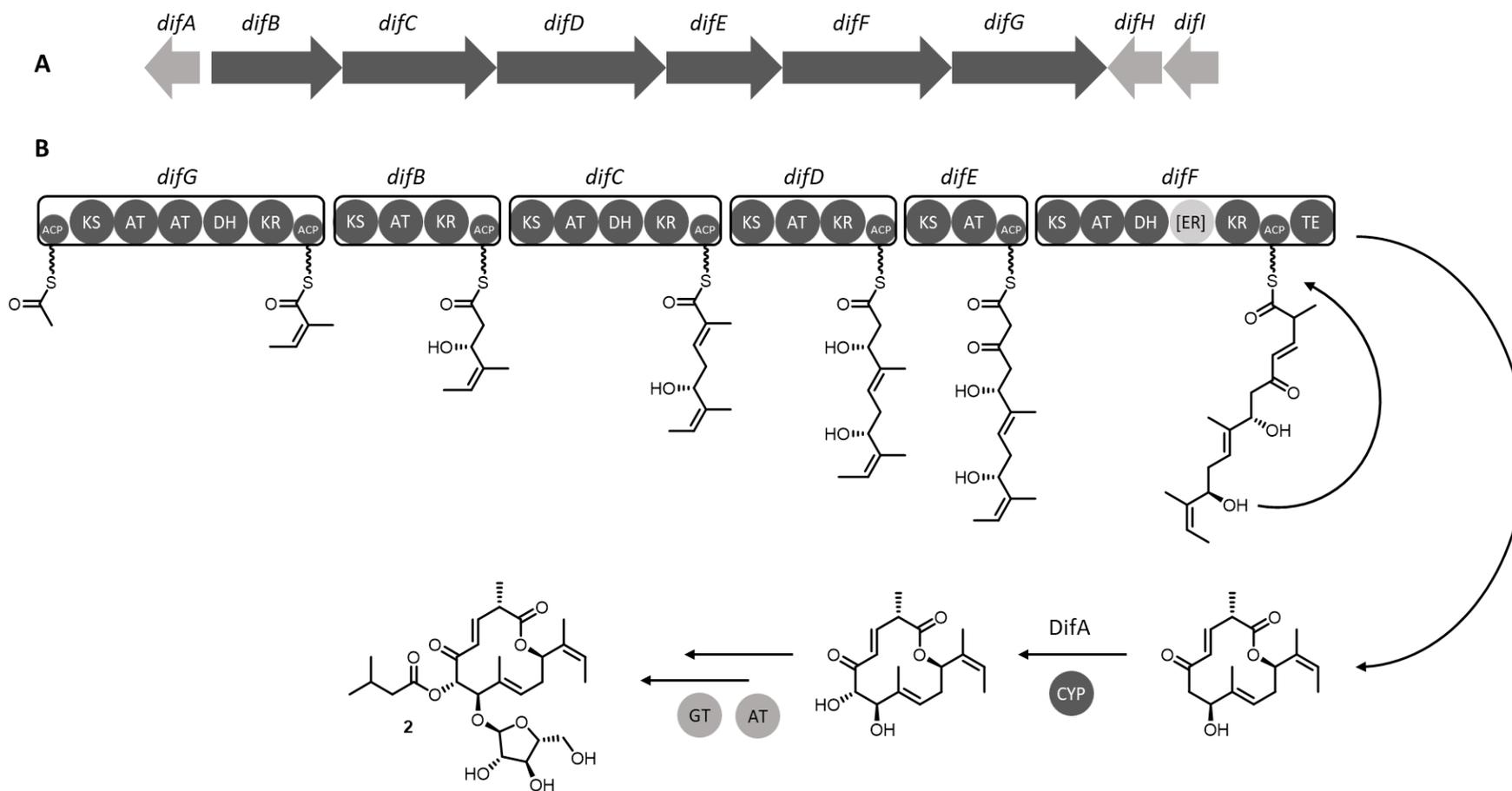
Elucidation of the structure and absolute configuration was performed by Surup *et al.* using a combination of NMR, MS, degradation, and molecular modeling techniques.<sup>61</sup> The molecules with the sum form of  $\text{C}_{27}\text{H}_{40}\text{O}_{10}$  are composed of a C-12 membered core aglycon with an isovaleric acid linked to C-6 and a  $\alpha$ -arabinofuranose linked to C-7. Sequencing of the AndGT8 genome and subsequent bioinformatic genome analysis using antiSMASH identified a linear type I PKS responsible for the biosynthesis of DIF. The biosynthetic gene cluster (BGC) consists of nine genes, six of which are the streamlined PKS genes *difBCDEFG*, responsible for assembly of the polyketide scaffold. Upstream of *difBCDEFG*, *difA* is located as a separate operon, encoding a cytochrome P450 tailoring enzyme, whereas two open reading frames (ORFs), *difH* and *difI*, encoding proteins with unknown function are located downstream of *difBCDEFG*. Surup *et al.* proposed a model for the biosynthesis based on the *in silico* analysis of DIF, starting with the fused starter and extender module DifG, followed by extension of the polyketide at modules DifB to DifF and release by macrocyclization (see Figure 3-3). Interestingly, one domain with an unknown function is located between the DH and KR domain in DifF, which is usually occupied by an ER domain. BLAST analysis of the DifF sequence

identified a Rossmann fold, which includes a NADP(H)-binding site. However, no other motifs typical for ER domains or indications for its function were found.<sup>61</sup> Dr. Konrad Viehriig later confirmed, that the core PKS cluster *difBCDEFG* is responsible for the synthesis of the DIF precursor aglycon DIF294 and the shunt or degradation product aglycon DIF268, identified by heterologous expression in *M. xanthus* DK1622.<sup>128</sup> Furthermore, *in vitro* analysis of purified DifA confirmed a cytochrome P450 holoenzyme that can be reduced to its ferric form by purified AndGT8 ferredoxin and ferredoxin reductase.<sup>128</sup> However, DifA did not oxidize either of the isolated aglycons. So far, no candidate enzymes responsible for the acylation, the glycosylation, and insertion of the double bond at C-3 required in DIF biosynthesis were identified, neither within the BGC nor somewhere else in the genome.



**Figure 3-2** DIF precursor aglycon DIF294, and shunt or degradation product aglycon DIF268, both produced by heterologous expression of the PKS genes *difBCDEFG*.

The Objective of this work was to further analyze the biosynthesis of the DIFs in the heterologous expression host *M. xanthus* DK1622. The goal was to identify the glycosyl- and acyltransferases responsible for the glycosylation and acylation of the DIF aglycon and to proof hydroxylation of C-6 by DifA. Furthermore, the involvement of the putative Fe-S oxidoreductase DifH as a potential candidate for the double bond formation at C-3 should be investigated.



**Figure 3-3** Proposed biosynthesis of DIF. **A)** BGC of DIF consisting of the nine genes *difA-l*. The six PKS genes *difBCDEFG* are shown in dark grey and putative tailoring enzymes are depicted in light grey. **B)** Proposed biosynthesis of DIF. While the hydroxy group at C-6 is potentially introduced by *DifA*, the enzymes *GT* (glycosyltransferase) and *AT* (acyl transferase) are responsible for acylation and glycosylation and are currently unknown. ACP = acyl carrier protein domain; AT = acyl transferase (domain); CYP = cytochrome P450; DH = dehydrogenase domain; ER = enoyl reductase domain; GT = glycosyl transferase; KS = ketosynthetase domain; KR = ketoreductase domain; TE = thioesterase domain. Figure adapted from Surup *et al.*<sup>61</sup>

### 3.3 Material and methods

#### 3.3.1 Media

All media used for liquid cultivations and agar plates were autoclaved for 20 minutes at 121°C before use and kept sterile.

**Table 3-2** Media used in this work with the respective composition

Media	Composition
<b>Lysogeny Broth (LB)</b>	Tryptone 10.0 g/L, yeast extract 5.0 g/L, NaCl 5.0 g/L, agar (for solid media) 15.0 g/L, dH <sub>2</sub> O 1.0 L
<b>2xYT</b>	Tryptone 16.0 g/L, yeast extract 10 g/L, NaCl 5.0 g/L, dH <sub>2</sub> O 1.0 L
<b>CTT</b>	Casitone 10.0 g/L, 1 M Tris-Cl pH 8.0 10.0 mL/L, 1 M K <sub>2</sub> HPO <sub>4</sub> pH 7.6 1.0 mL/L, 1 M MgSO <sub>4</sub> 8.0 mL/L, agar (for solid media) 15.0 g/L and for soft agar 7.5 g/L, dH <sub>2</sub> O 1.0 L
<b>CYE</b>	Casitone 10.0 g/L, yeast extract 5.0 g/L, MgSO <sub>4</sub> 8.0 mM/L, Tris-HCl pH 7.6 10.0 mM/L, dH <sub>2</sub> O 1.0 L

#### 3.3.2 Antibiotic stock solutions

Several antibiotics were used in liquid media or agar plates for the selection of genetically altered *E. coli* and *M. xanthus* strains: Ampicillin<sup>[a]</sup> (Amp), Chloramphenicol<sup>[a]</sup> (Cm), Zeocin (Zeo), Oxytetracycline (Otc), and Kanamycin (Kan). Table 3-3 depicts the antibiotic concentrations used in this work.

**Table 3-3** Antibiotics used in this work with their stock and working concentration.

Antibiotic	Stock (1,000x) [mg/ml]		Working conc. [µg/ml] <sup>[b]</sup>
<b>Ampicillin (Amp)</b>	100	1 g in 10 mL MQ-H <sub>2</sub> O	100
<b>Chloramphenicol (Cm)</b>	25	0.25 g in 10 mL 70 % abs. EtOH	25
<b>Kanamycin (Kan)</b>	50	0.5 g in 10 mL MQ-H <sub>2</sub> O	50
<b>Zeocin (Zeo)<sup>[c]</sup></b>	100		50
<b>Oxytetracycline (Otc)</b>	10	0.1 g in 10 mL 70 % abs. EtOH	10

<sup>[a]</sup> Amp and Cm are not applicable in *M. xanthus* <sup>[b]</sup> For high-copy vectors; low-copy vectors ½ of normal working conc., <sup>[c]</sup> Zeocin is purchased by ThermoFisher Scientific in pre-made ready to use stocks.

Antibiotic stock solutions were stored at -20°C. Oxytetracycline and Zeocin are light sensitive, thus the stock solutions were protected from light.

### 3.3.3 Reagents and buffer

Solutions were prepared with demineralized water (dH<sub>2</sub>O) or ultrapure water (MQ-H<sub>2</sub>O). To obtain MQ-H<sub>2</sub>O, 'Milli-Q Reference A+ System' (Millipore, Merck) was used.

For agarose gel electrophoresis 50x TAE buffer containing TRIS 242 g/L, glacial acetic acid 57.1 mL/L, EDTA (500 mM in dH<sub>2</sub>O pH 8.0) 100 mL/L and dH<sub>2</sub>O *ad* 1 L was used. The TAE buffer (50x) stock solution was diluted with dH<sub>2</sub>O to the 1x working concentration before use. Agarose gel solution was prepared by adding 5 g agarose in 500 mL TAE buffer (1x) for 1% solution. For separation of larger DNA fragments 0.8% agarose gel solution was used (4 g agarose in 500 mL 1x TAE buffer). Agarose was dissolved by microwave heating. The solution was stored at 55 °C to prevent solidification.

Orange G loading dye (10x) was used to stain DNA for visualization in agarose gels on a Transilluminator. The solution contains orange G 10 mg, glycerin 3 mL and TE buffer 2 mL. TE buffer consists of TRIS (0.5 M in dH<sub>2</sub>O) 20 mL, EDTA (0.5 M in dH<sub>2</sub>O, pH 8.0) 2 mL and dH<sub>2</sub>O *ad* 1 L. The 10x loading dye was diluted to 1x in the DNA solution that had to be stained.

Buffers P1, P2, and P3 were used for alkaline lysis and plasmid DNA preparation. The composition of each is shown in the table below.

**Table 3-4** Composition of resuspension buffer P1, lysis buffer P2, and neutralization buffer P3. Buffers were used for alkaline lysis and plasmid DNA preparation.

<b>Resuspension buffer P1</b> Stored at 4° C	TRIS (1 M in dH <sub>2</sub> O) 2.5 mL, EDTA (0.1 M, pH 8.0) 5 mL, Ribonuclease A 25 mg, MQ-H <sub>2</sub> O <i>ad</i> 50 mL
<b>Lysis buffer P2</b>	NaOH (1 M in dH <sub>2</sub> O) 10 mL, SDS (20 % in dH <sub>2</sub> O) 2.5 mL, MQ-H <sub>2</sub> O <i>ad</i> 50 mL
<b>Neutralization buffer P3</b>	K-acetate 14.72 g, MQ-H <sub>2</sub> O <i>ad</i> 50 ml, pH 5.5

### 3.3.4 Software

*Geneious 10.1.3* (Biomatters) was used to analyze native biosynthetic gene clusters, to design synthetic gene clusters, to design oligonucleotides for PCR experiments and sequencing, and to design plasmid maps. Agarose gel pictures from the fluorescence chamber were analyzed with *FusionCapt Advance* Software (Vilber Lourmat). *ChemDraw Ultra 12.0* (CambridgeSoft) was used to create images and chemical formulae. LC chromatograms as well as UV spectra and mass spectrograms were analyzed with *DataAnalysis 4.4* (Bruker Daltonics, Billerica, MA, USA).

### 3.3.5 *E. coli* cultivation

*E. coli* DH10 $\beta$ , NEB 10 $\beta$  or HS996 were used for standard cloning procedures. Except for electroporation for transformation of plasmids, working with *E. coli* was performed under sterile conditions (HeraSafe, Heraeus). For storage at -80°C, 750  $\mu$ L over-night culture was combined with 750  $\mu$ L 50 % glycerin (in MQ-H<sub>2</sub>O) and transferred into a cryo tube.

Cultivation took place in 2xYT or LB medium. Supplementation with antibiotics for selection depended on the plasmid resistance marker. Antibiotic concentrations are shown in Table 3-3. For clone selection, cultures were spread on LB agar (25 mL in a petri dish) after transformation and incubated overnight at 37°C in an incubator. Liquid cultures (5 mL in glass vial or 10-20% v/v liquid media in Erlenmeyer flasks) were incubated overnight at 37°C in a rotary shaker (Multitron, Infors HT) at 200 rpm. Larger constructs of 30 kb size or longer, were incubated at 30°C for a longer period of time (up to two days of incubation).

### 3.3.6 Preparation of *E. coli* electro competent cells

For preparation of electro competent *E. coli* cells, 5 mL LB medium were inoculated with the respective strain directly from cryo stock and incubated at 37 °C and 200 rpm overnight. 2 mL of overnight culture were used to inoculate 200 mL LB. Cultivation was carried out under the same conditions until an OD<sub>600</sub> (Biophotometer plus, Eppendorf) between 0.4 and 0.6 was reached. The cultures were cooled on ice for 30 min and afterwards transferred into four 50 mL Falcon tubes. All centrifugation steps were performed at 4,000  $\times$  g for 10 min at 4 °C (Avanti J-26 XP Centrifuge, Beckman Coulter; Rotor: JLA 10.500), followed by discarding the supernatant. First time, the cell pellets were resuspended in 100 mL (25 mL each) ice cold HEPES (1 mM in dH<sub>2</sub>O; pH7.0), followed by centrifugation. Second time, the cell pellets were

resuspended in 50 mL (12.5 mL each) ice cold HEPES (1 mM in dH<sub>2</sub>O; pH 7.0) and combined in one falcon tube before centrifugation. Third time, the cell pellet was resuspended in 50 mL ice cold 10 % glycerin in HEPES (1 mM in dH<sub>2</sub>O; pH 7.0), followed by centrifugation. Finally, the cell pellet was resuspended in 2 mL 10 % glycerin in HEPES (1 mM in dH<sub>2</sub>O; pH7.0) and portioned into 50 µl aliquots. Now the prepared cells were frozen in liquid N<sub>2</sub> and stored at - 80° C.

### **3.3.7 Cultivation of *M. xanthus***

*M. xanthus* strains were cultivated either for transformation of heterologous gene constructs or for testing production titers of target compounds. All cultures were incubated at 30° C, either in an incubator (plates) or on shakers at 200 rpm (Erlenmeyer flasks). Production cultures were grown in 25 mL liquid media with corresponding selection marker until sufficient cell density was reached. Subsequently, 5 mL of the culture was sub-cultured into 50 mL liquid medium (10% inoculum). After 24h of incubation, vanillate was added to induce the expression of heterologous genes, in case a vanillate promoter was used for induction of a gene cluster. Further, XAD-16 was added to bind produced compounds. The cell cultures were harvested after the culture broth took on a brown color (if not mentioned otherwise).

### **3.3.8 Extraction and analysis of disciformycin**

Culture was centrifuged for 20 min at 8000 rpm, supernatant discarded, and cell/XAD-16 pellet stored at -20°C for a couple of hours/until frozen and then lyophilized overnight. The pellet was then extracted twice with 30 mL ethyl acetate for 20 minutes. After extraction the liquid was filtered and the eluate caught in a round flask. The solvent was then evaporated at a Rotavapor until the extract was completely dry. Now the extract was re-dissolved in 1 mL MeOH and pipetted into a glass vial. Before analysis, sample was centrifuged for at least 10 min at 15000 rpm and 10 µL of the supernatant was 10 times diluted in MeOH.

For detection of compound UHPLC-hrMS analysis was performed on a Dionex UltiMate 3000 rapid separation liquid chromatography (RSLC) system (Thermo Fisher Scientific, Waltham, MA, USA) coupled to a Bruker maXis 4G ultra-high-resolution quadrupole time-of-flight (UHR-qTOF) MS equipped with a high-resolution electrospray ionization (HRESI) source (Bruker Daltonics, Billerica, MA, USA). The separation of a 1 µL sample was achieved with a linear 5–95% gradient of acetonitrile with 0.1% formic acid in ddH<sub>2</sub>O with 0.1% formic acid on

an ACQUITY BEH C18 column (100 mm × 2.1 mm, 1.7 μm dp) (Waters, Eschborn, Germany) equipped with a Waters VanGuard BEH C18 1.7 μm guard column at a flow rate of 0.6 mL/min and 45 °C for 18 min with detection by a diode array detector at 200–600 nm. The LC flow was split into 75 μL/min before entering the mass spectrometer. Mass spectrograms were acquired in centroid mode ranging from 150–2500 m/z at an acquisition rate of 2 Hz in positive MS mode. Source parameters were set to 500 V end-plate offset; 4000 V capillary voltage; 1 bar nebulizer gas pressure; 5 L/min dry gas flow; and 200 °C dry gas temperature. Ion transfer and quadrupole parameters were set to 350 VPP funnel RF; 400 VPP multipole RF; 5 eV ion energy; and 120 m/z low-mass cut-off. Collision cell was set to 5.0 eV and pre-pulse storage was set to 5 μs. Calibration was conducted automatically before every HPLC-MS run by the injection of sodium formate and calibration on the respective clusters formed in the ESI source. All MS analyses were acquired in the presence of the lock masses C<sub>12</sub>H<sub>19</sub>F<sub>12</sub>N<sub>3</sub>O<sub>6</sub>P<sub>3</sub>, C<sub>18</sub>H<sub>19</sub>F<sub>24</sub>N<sub>3</sub>O<sub>6</sub>P<sub>3</sub> and C<sub>24</sub>H<sub>19</sub>F<sub>36</sub>N<sub>3</sub>O<sub>6</sub>P<sub>3</sub>, which generate the [M + H]<sup>+</sup> ions of 622.0289, 922.0098 and 1221.9906. The HPLC-MS system was operated by HyStar 5.1 (Bruker Daltonics, Billerica, MA, USA), and LC chromatograms as well as UV spectra and mass spectrograms were analyzed with DataAnalysis 4.4 (Bruker Daltonics, Billerica, MA, USA).

### 3.3.9 Plasmid DNA isolation from *E. coli*

#### 3.3.9.1 “Mini prep”

For plasmid DNA isolation from *E. coli*, 2 mL overnight culture was centrifuged for 1 min at 15000 rpm. After discarding the supernatant, the cell pellet was resuspended in 250 μL P1 buffer using a bench top shaker. Cell lysis occurred after adding 250 μL P2 buffer and inverting the tubes seven times, and incubation for 3 min. By addition of 250 μL P3 buffer, 10 μL chloroform and inverting seven times, the lysis was stopped. After centrifugation for 10 min at 15000 rpm, the supernatant was transferred into a new 1.5 mL Eppendorf tube containing 500 μL ice cold isopropanol. Another centrifugation step for 5 min at 15000 rpm to precipitate the plasmid DNA was followed by discarding the supernatant and adding 700 μL 70 % ethanol (in MQ-H<sub>2</sub>O) for washing. Finally, after centrifugation for 1 min at 15000 rpm, the supernatant was discarded and the pellet dried before resuspension in 50 - 100 μL MQ-H<sub>2</sub>O.

#### 3.3.9.2 “Midi prep”

In case larger quantities of isolated plasmid were required for subsequent cloning procedures, 200 mL of *E. coli* overnight culture were grown and harvested. The plasmid was

then isolated from the cells using the Nucleo Bond PC100 (midi-prep) DNA purification kit by Macherey-Nagel according to their protocol.

### **3.3.10 Polymerase chain reaction (PCR)**

PCR protocols that were commonly used during this work are described below. Depending on the experiment, template size or objective of the PCR, the protocols might slightly differ from below mentioned parameters.

#### **3.3.10.1 Phusion polymerase**

Reaction mix: GC buffer 4.0  $\mu\text{L}$ , dNTPs 5 mM (1.25 mM each) 4.0  $\mu\text{L}$ , Phusion polymerase (Thermo Fisher Scientific) 0.2  $\mu\text{L}$ , primer forward (10  $\mu\text{M}$ ) 0.5  $\mu\text{L}$ , primer reverse (10  $\mu\text{M}$ ) 0.5  $\mu\text{L}$ , DNA template 2  $\mu\text{L}$ \*, MQ-H<sub>2</sub>O 8.8  $\mu\text{L}$ , DMSO 1.0  $\mu\text{L}$ .

\*2  $\mu\text{L}$  genomic DNA, for plasmid DNA 1  $\mu\text{L}$  with a concentration of approximately 10 ng/ $\mu\text{L}$  was used. Volume of MQ-H<sub>2</sub>O had to be adapted accordingly.

Cycle protocol: Initial denaturation 5 min at 98°C; 34 cycles of denature (98°C 20 s), anneal (temperature depends on primers, 25 s), and elongation (72°C 15-30 s/kb); finale elongation at 72°C for 5 min.

#### **3.3.10.2 Taq polymerase**

Reaction mix: Taq buffer (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub> 2.5  $\mu\text{L}$ , dNTPs 5 mM (1.25 mM each) 4.0  $\mu\text{L}$ , primer forward (50  $\mu\text{M}$ ) 0.5  $\mu\text{L}$ , primer reverse (50  $\mu\text{M}$ ) 0.5  $\mu\text{L}$ , Taq polymerase (Thermo Fisher Scientific) 0.2  $\mu\text{L}$ , MgCl<sub>2</sub> (50 mM) 1  $\mu\text{L}$ , DNA template 2  $\mu\text{L}$ , glycerol 50% 4  $\mu\text{L}$ , MQ-H<sub>2</sub>O 5.3  $\mu\text{L}$ .

Cycle protocol: Initial denaturation 5 min at 95°C; 30 cycles of denature (95°C 30 s), anneal (temperature depends on primers, 30 s), and elongation (72°C 1 min/kb); finale elongation at 72°C for 5 min.

#### **3.3.10.3 Q5 Polymerase**

Mix for two reactions: 5x Q5 buffer 10  $\mu\text{L}$ , dNTPs 5mM (1.25 mM each) 2  $\mu\text{L}$ , primer forward (50  $\mu\text{M}$ ) 0.5  $\mu\text{L}$ , primer reverse (50  $\mu\text{M}$ ) 0.5  $\mu\text{L}$ , DNA template 1  $\mu\text{L}$ , Q5 polymerase (New England BioLabs) 0.5  $\mu\text{L}$ , 5x GC buffer 10  $\mu\text{L}$ , MQ-H<sub>2</sub>O 25.5  $\mu\text{L}$ .

Cycle protocol: Initial denaturation 30 seconds at 98°C; 30 cycles of denature (98°C 10 s), anneal (temperature depends on primers, 30 s), and elongation (72°C 20-30 seconds/kb); finale elongation at 72°C for 2 min.

### **3.3.11 DNA separation and purification**

Agarose gel electrophoresis was used for separation of DNA fragments based on size. In this work, agarose gel electrophoresis was performed in a Consort EV231 chamber (Sigma-Aldrich). An 1% agarose gel was used with 10  $\mu$ l Roti Safe Gel Stain per 100 mL gel and TAE buffer (1x) as buffer system. After PCR experiments or hydrolyzations with restriction enzymes, OrangeG loading dye (10x) was added to the DNA sample (1/10 of total volume) and the gel was loaded immediately. GeneRuler 1 kb/1 kb plus/100 bp plus DNA ladder (Thermo Fisher Scientific) were used as size standards. The voltage was set between 1 - 5 V/cm until DNA fragments were sufficiently separated. DNA detection took place in a Fusion FX chamber with UV fluorescence detector (Vilber Lourmat).

For DNA purification, agarose gel was placed on a blue light transilluminator, desired DNA fragments were cut out of the agarose gel with a scalpel and transferred into an Eppendorf tube. The proper DNA purification was done with the Macherey-Nagel NucleoSpin Gel and PCR Clean-up Kit following the manufacturer information. Finally, purified DNA was eluted in 30  $\mu$ l of MQ-H<sub>2</sub>O/elution buffer (from NucleoSpin Gel and PCR Clean-up Kit).

#### **3.3.11.1 DNA precipitation**

NaAcetate solution (pH 5.2) 15  $\mu$ l was added to 100  $\mu$ l digestion reaction and inverted 1-2 times. Then 300  $\mu$ l of 100% ice cold ethanol was added and inverted 2-3 times. The reaction was incubated at -20° C for 30-60 minutes and spun down at 4° C at 15.000 rpm for 20-30 minutes. The DNA was then washed with 70% ethanol and dried.

### **3.3.12 Determination of DNA concentration**

NanoDrop 2000c (Thermo Fisher Scientific) was used for determination of DNA concentrations of solutions. For that purpose, 1  $\mu$ l of DNA solution was put on the pedestal and measured at 260 - 280 nm wavelength. MQ-H<sub>2</sub>O served as a reference (blank).

Furthermore, the sample DNA concentration was determined visually by loading 1  $\mu$ l DNA sample on an agarose gel and comparing the DNA band intensity with the intensity of a DNA ladder band of similar size. 6  $\mu$ l of the DNA ladder were loaded as standard (using recommended standard concentrations).

### 3.3.13 Molecular cloning

For conventional, preparative hydrolyzation/ligation reactions, both vector and insert DNA were treated with type II or type IIS (*Bsal*) restriction endonucleases (REs). Thereby, DNA is linearized, meaning that vector and insert DNA with matching overhangs can be ligated by T4 DNA ligase. In the following sections, restrictive hydrolyzation of DNA will sometimes be referred to as ‘digest’ of DNA, which is colloquial use in daily lab environment. All restriction endonucleases, T4 DNA ligase, and FastAP were purchased from Thermo Fisher Scientific.

#### 3.3.13.1 Standard protocol for preparative restrictive hydrolyzation of DNA

Preparative digest was used to hydrolyze sufficient amount of DNA for follow up ligation of insert and vector.

**Table 3-5** Composition of standard preparative restrictive hydrolyzation reactions for single digest (with one enzyme) and double digest (with two enzymes, respectively). DNA solution was usually obtained by DNA isolation (see section 3.3.9) or in some cases by PCR (see section 3.3.10).

Single digest	Double digest
25 µL plasmid DNA solution	24 µL plasmid DNA solution
3 µL 10x restriction endonuclease buffer	3 µL 10x restriction endonuclease buffer
2 µL restriction endonuclease	1.5 µL restriction endonuclease 1
	1.5 µL restriction endonuclease 2

Reactions were incubated for several hours or overnight at the required temperature (for most enzymes 37°C). To prevent vector religation, FastAP enzyme was added after  $t_{1/2}$  of the reaction (1 µl in 30 µl reactions). This leads to dephosphorylation of the 5' ends of the linearized vector DNA and therefore improves the overall cloning efficacy.

#### 3.3.13.2 Standard protocol for analytical restrictive hydrolyzation of plasmids containing insert

Analytical digests were used to verify plasmid size and screen for clones harboring plasmids with the right insert after a cloning process.

**Table 3-6** Composition of standard analytical restrictive hydrolyzation reactions for single digest (with one enzyme) and double digest (with two enzymes, respectively). DNA solution was usually obtained by DNA isolation (see section 3.3.9) or in some cases by PCR (see section 3.3.10).

Single digest	Double digest
8.5 µL plasmid DNA solution	8.0 µL plasmid DNA solution
1.0 µL 10x restriction endonuclease buffer	1.0 µL 10x restriction endonuclease buffer
0.5 µL restriction endonuclease	0.5 µL restriction endonuclease 1
	0.5 µL restriction endonuclease 2

Reactions were incubated for approximately 2 hours at the required temperature.

### 3.3.13.3 Standard ligation reaction

Standard ligation protocol was used to assemble prior linearized DNA fragments. T4 DNA ligase catalyzes bond formation between a 5' phosphate group and a 3' hydroxy group and thereby leads to the ligation of vector and insert DNA. The composition of ligation reaction performed in this work is shown in Table 3-7.

**Table 3-7** Composition of standard ligation reaction, used to assemble linearized DNA.

Component	Volume [µL]
<b>Vector DNA</b>	0.5 – 2.0 <sup>[a]</sup>
<b>Insert DNA</b>	6.5 – 8.0 <sup>[a]</sup>
<b>T4 DNA ligase buffer</b>	1.0
<b>T4 DNA ligase</b>	0.5

<sup>[a]</sup> Vector and insert DNA volume depended on determined DNA concentration and varied between 0.5-2 µl for vector DNA solution and 6.5-8 µl for insert DNA solution. Total reaction volume was always 10 µl with a vector/insert DNA molar ratio of around 1/10.

Ligation reaction was incubated for 5 h at room temperature or overnight at 16°C.

### 3.3.14 *E. coli* transformation

Transformation of *E. coli* was carried out by electroporation, which shortly increases cell membrane permeability and enables the uptake of plasmid DNA. For electroporation, 50 µl electrocompetent cells were mixed with 5 µl of ligation mix or 0.5 µl of gene synthesis fragment DNA solution and transferred into an electroporation cuvette (electrode distance 1 mm). The electroporation was performed at 1,300 V/cm, 10 µF, and 600 Ω (Eporator V1.01,

Eppendorf). Thereafter, the cells were resuspended in 1 mL of LB medium and incubated at 37 °C and 900 rpm (Titramax 1000, Heidolph) for 60 min. After centrifugation at 8000 rpm for 2 min, most of the supernatant was discarded and the cells were resuspended in the remaining medium. In case of gene synthesis fragments, the centrifugation step was skipped and 50 µl were directly used for the following step. The cell suspension was plated out on LB agar with appropriate antibiotic for selection and incubated overnight at 37 °C.

### **3.3.15 *M. xanthus* DK1622 transformation**

*M. xanthus* DK1622 was cultured overnight in 25 mL CTT medium until an appropriate cell density was reached (OD<sub>600</sub> of 0.6). 2 mL of the culture was pipetted into an Eppendorf tube and centrifuged for 2 min at 6000 rpm and room temperature. Supernatant was discarded. The cell pellet was resuspended in 1 ml sterile MQ water and centrifuged again for 2 min at 6000 rpm, supernatant was again discarded using a pipette. This washing step was repeated once with 800 µL. A small hole was poked into the lid using a needle, and 35 µL of sterile MQ water plus plasmid DNA (5-10 µL DNA/plasmid with a concentration of 100-500 ng/µL DNA) was added. Cell pellet was resuspended and the suspension was used for electroporation at 650 V, 25 µF, and 400 Ω (Bio Rad GenePulser Xcell). After electroporation, the culture was resuspended in 1 mL CTT liquid medium and incubated on a shaker for 6 hours at 30° C and 1000 rpm. A glass vial with foam plug was used to mix 3 mL of CTT soft agar and the respective selection marker with the 1 mL transformation culture. The suspension was thoroughly shaken and plated on a CTT agar plate containing the respective selection marker as well. Plates were incubated on 30° C for several days until clones appeared.

### **3.3.16 Screening for *E. coli* colonies**

To verify a successful transformation and isolate the constructed plasmid, 3 to 24 single *E. coli* colonies were picked and used for inoculation of 5 mL LB medium with appropriate antibiotic. Incubation took place at 37 °C and 200 rpm overnight. Plasmid DNA isolation from *E. coli* was carried out by the protocol described in 3.3.9. Afterwards, an analytical digest was performed (3.3.13.2) and the presence of correct DNA fragments was examined by agarose gel electrophoresis according to 3.3.11.

### 3.3.17 Red/ET – homology-based recombination

Red/ET homology-based recombination<sup>124</sup> was performed according to following protocol: overnight culture of *E. coli* GB05 with the according plasmid was cultured with the respective selection marker. This overnight culture was then used to inoculate 12x2 mL Eppendorf tubes with a hole in the lid and 1.4 mL LB medium containing the respective selection marker (inoculum volume: 4x15  $\mu$ L, 4x25 $\mu$ L and 4x30-100  $\mu$ L). Those tubes were incubated on a shaker at 30° C until an OD<sub>600</sub> of roughly 0.2 was reached (duration approximately 2-3 hours). Three of the cultures that reached the desired OD<sub>600</sub> first were used for the Red/ET recombination: First 40  $\mu$ L of 10% L-Arabinose (dissolved in MQ-H<sub>2</sub>O and sterile filtered) were added to the cultures, which were then incubated at 37 °C for 45 minutes on a shaker. Afterwards, cultures were centrifuged for 1 minute at 9000 rpm at 2° C and supernatant was discarded using a pipette. All following steps had to be carried out on ice. The cell pellet was resuspended in 1 mL ice cold sterile MQ-H<sub>2</sub>O, centrifuged for 1 minute at 9000 rpm and the supernatant discarded with a pipette again. This step was repeated once, except not all of the supernatant was discarded (20-50  $\mu$ L leftover liquid). These tubes with cell pellet were now used for transformation with 1  $\mu$ L, 2  $\mu$ L and 2.5  $\mu$ L, respectively, of template DNA. The transformation was carried by electroporation as described before (see section 3.3.14). After the electroporation, the cultures were incubated for 75 minutes at 37° C on a shaker. Subsequently, the whole mL of each culture was plated on a LB Agar plate with the respective selection marker and incubated on 30° C. After 1-2 days, when colonies were grown, the whole cell mass was harvested by pipetting 2x1 mL of MQ-H<sub>2</sub>O onto the plate and scraping it off, then pipetting the suspension into new Eppendorf tubes. The plasmids of these cultures were then isolated by using the before described mini-prep protocol (see section 3.3.9.1) and the isolated plasmid was transformed into competent *E. coli* strains NEB 10 $\beta$ , HS996 or DH10 $\beta$  again to increase the chance of only having one plasmid per cell, *i.e.* no false positive clones that contain a mixture of right and wrong plasmids after the Red/ET process.

### 3.3.18 Protein expression and *in vitro* assays

Expression and characterization of recombinant DifA: Cytochrome P450 gene *difA* was cloned in expression plasmid pETM44 and expressed as MBP-tagged protein according to protocols adapted for optimized heterologous expression of cytochromes. *E. coli* BL21 DE3/pETM44\_*difA* was grown in 200 mL of TB medium supplemented with 50 mg/L kanamycin. As the culture reached an OD<sub>600</sub> of 2.0, 5-amino-levulinic acid was added to a final

concentration of 1 mM, FeCl<sub>3</sub> was added to 0.5 mM, and the culture was incubated for 30 min at 25°C before addition of IPTG to 0.25 mM. The culture was then further incubated for 60 h (25 °C, 200 rpm) until the pH of the culture was 7.1, then harvested and washed. Cleared *E. coli* lysate was separated with the ÄKTA HPLC system using the MBP-Trap column. A one-step elution with maltose buffer was applied according to the manufacturers protocol. All buffers used contained 5 mM DTT to prevent oxidation of the protein. During the purification, absorption was monitored at 280 and 430 nm. Fractions containing the absorption peak at 430 nm were pooled, concentrated, and cleaved with TEV-protease. The cleaved MBP-tag was separated from DifA in a second chromatography step. After buffer exchange, the protein was concentrated to 0.5 mL final volume. Protein concentration was determined 10 mg/mL using Bradford measurement and 80 μM according to spectrometric analysis. SDS-PAGE of the protein revealed a strong band at 40 kDa, with protein impurities at above 80 kDa. Recording of CO-spectra and difference spectra with test substrates DIF294 and DIF268 was performed according to standard protocols.

Native electron carrier proteins Fdx and FdR from *P. fallax* AndGT8 were cloned into pETM44 as described for DifA. *E. coli* BL21 FdR/Fdx were grown in TB medium with 50 mg/L kanamycin until the culture reached an OD<sub>600</sub> of 1.0, then IPTG was added to a concentration of 0.5 mM and cultivation was continued until the pH of the culture was 7.1. Cell harvest and protein purification as with DifA.

RhFRED and RhFRED-fdx (ferredoxin-domain replaced with spinach-ferredoxin for less stringent acceptor specificity) were expressed in TB-medium as Fdx and FdR. Protein was isolated with HisTrap column using standard protocols. His-tag was not removed.

### 3.3.19 Reduction assay for DifA

8 μL DifA were used in a reaction volume of 800 μL (final concentration 0.8 μM). Fully reduced DifA (DTT treatment for 5 minutes) was used (in both cuvettes) to obtain a baseline blank, then the sample cuvette was treated with CO gas. The Absorption maximum at 450 nm was OD = 0.074, extinction coefficient for P450 at 450 nm is  $\epsilon = 0.099$ , meaning that the actual concentration of active P450 was calculated 0.8 μM. Different redox partners were then tested with DifA in the molar ratio of CYP:Fdx:FdR = 1:10:3 for separate proteins, or 1:10 in the case of fused Fdx-fdR systems, and in presence of 1 mM NADPH. For Fdx (8 μM) and FdR (2.4 μM), 10% reduction of DifA were observed after 5 min incubation at room temperature.

### 3.3.20 *In vitro* conversion assay with DifA and aglycon DIF294 as substrate

*In vitro* conversion was performed in 10 mM potassium phosphate buffer at pH 7.4 or 50 mM Tris/HCl buffer at pH 8.0 in 0.5 mL or 0.25 mL volumes. The following conditions were used: 100  $\mu$ M substrate (DIF294), 20 mM Fdx, 2  $\mu$ M FdR, 0.5  $\mu$ M DifA. All ingredients were mixed on ice; the reaction was started by adding 1mM NADPH and incubated at 30°C for 2h. The reaction was stopped by adding 500  $\mu$ L of ethyl acetate and strong mixing. After brief centrifugation, the organic phase was dried *in vacuo*, dissolved in MeOH and subjected to ESI-MS for detection of the reaction products.

### 3.3.21 Generating plasmids TARGT1 and TARGT2

#### 3.3.21.1 TARGT1

The two native mannosyltransferases from AndGT8 were obtained from plasmid TAR4 by PCR with primer pair Dif\_Pvan\_fwd and Dif\_GT\_rev. After separation on an agarose gel and purification, the 3.16 kb DNA fragment was cloned into a pJET plasmid with a subsequent transformation into *E. coli* strain DH10 $\beta$  for replication of the plasmid. Clones harboring the pJET::GT1 plasmid were sent for sequencing using primers GT TAR4 seq fwd, GT TAR4 seq rev, GT TAR4 seq 3, pJET1.2\_F and pJET1.2\_R (the latter two are LGC primer). Clone 3 showed the best sequencing results and was stored as glycerol stock at -80°C. Glycosyltransferase operon GT1 was now inserted into plasmid TAR2 by hydrolyzing the pJET plasmid containing GT1 and TAR2 with *Swa*I and *Acl*I and subsequent ligation. Resulting clones were analyzed by restrictive hydrolyzation with *Bst*EII and clones with the correct construct were sequenced for verification, stored as glycerol stocks, and used for further work. Primers are listed in SI Table 3-13.

#### 3.3.21.2 TAR2GT2

The second glycosyltransferase operon GT2 was obtained from genomic DNA of AndGT8 *via* four separate PCR reactions. PCR 1 was performed with primer pair Dif\_Pvan\_fwd and Dif\_Pvan\_rev, and TAR4 as template to obtain the vanillate promoter. PCR 2 was performed with primer pair Dif\_MOX\_fwd and Dif\_MOX\_rev, and genomic AndGT8 DNA as template, PCR 3 with Dif\_MT\_ManT\_fwd and Dif\_MT\_ManT\_rev and genomic DNA as template as well. The primers were designed to introduce overlapping regions between PCR products 1 – 3. A fourth PCR, combined the three overlapping fragments, using primer pair Dif\_Pvan\_fwd and Dif\_MT\_ManT\_rev. As described for plasmid TAR2GT1 the PCR product was cloned into a pJET

plasmid for replication in *E. coli* and subsequent sequencing. The construct GT2 was ligated into TAR2 via *Swa*I and *Acl*I. Primers Mod GT operon seq 1, Mod GT operon seq 2, Mod GT operon seq 3, Mod GT operon seq4, pJET1.2F and pJET1.2R were used for sequencing. The sequence of one plasmid was verified by sequencing and was stored as glycerol stock and used for further work. Primers are listed in SI Table 3-13.

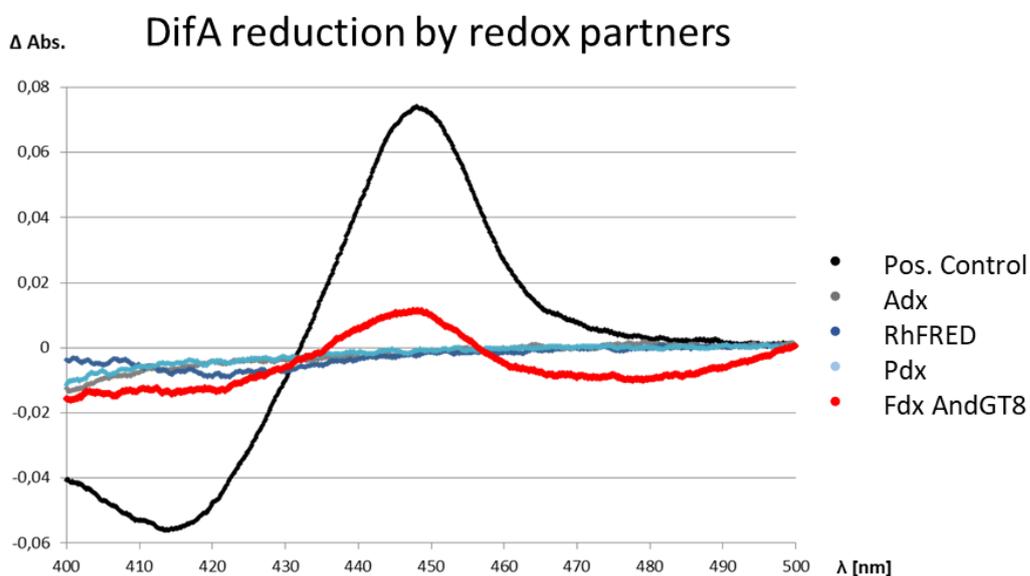
### **3.3.22 Replacement of *difH* by *tetR* by Red/ET recombineering**

*DifH* was replaced by a tetracycline resistance gene on plasmid TAR2 by Red/ET homologous linear to circular recombination. *TetR* gene was obtained from plasmid pMYC24 (described in section 2.4.1.2) by PCR using primers DifHko fwd and DifHko rev. Both primers contained 50 bp overhangs homologous to the direct adjacent flanking regions of *difH*. Red/ET homologous recombination with *tetR* as donor DNA and TAR2 as receiving plasmid was performed according to protocol described in section 3.3.17. Plasmids were analyzed after Red/ET recombineering by restrictive hydrolysis and separation of the DNA fragments by gel electrophoresis (see section 3.3.13.2 and 3.3.11).

## 3.4 Results and discussion

### 3.4.1 *In vitro* analysis of the cytochrome P450 DifA

The hydroxylation of aglycons DIF294 and DIF268 by DifA was tested before *in vitro* by Dr. Konrad Viehrig using recombinant Adrenodoxin 4-108 (Adx) and its native Adrenodoxin-Reductase (AdR) as electron delivery system.<sup>128</sup> However, UHPLC-HRMS analysis showed no conversion of either of the two substrates. Since the structural integrity of the Adx/AdR proteins was confirmed in a control reaction, it was assumed that Adx/AdR are probably no suitable electron donors for DifA. Consequently, the other surrogate electron transport proteins RhFRED and RhFRED-Fdx, a reductase domain of a naturally occurring P450-Fdx-FdR fusion protein from *Rhodococcus*, as well as Pdx, a natural fused Fdx-FdR system from *M. xanthus*, were tested. Furthermore, genome sequence analysis of the native Disciformycin producer AndGT8 revealed one ferredoxin and two ferredoxin reductase genes, which were consequently cloned. AndGT8-Fdx and one AndGT8-Reductase were successfully expressed in *E. coli* BL21 and purified for investigation with CO difference spectrometry (detailed protocol see section 3.3.19).



**Figure 3-4** Reduction assay to test electron donors for DifA. Successful reduction of DifA is indicated by a peak at 450 nm. DTT was used as positive control. Reduction was successful using the native Fdx and FdR proteins from AndGT8 (red line).

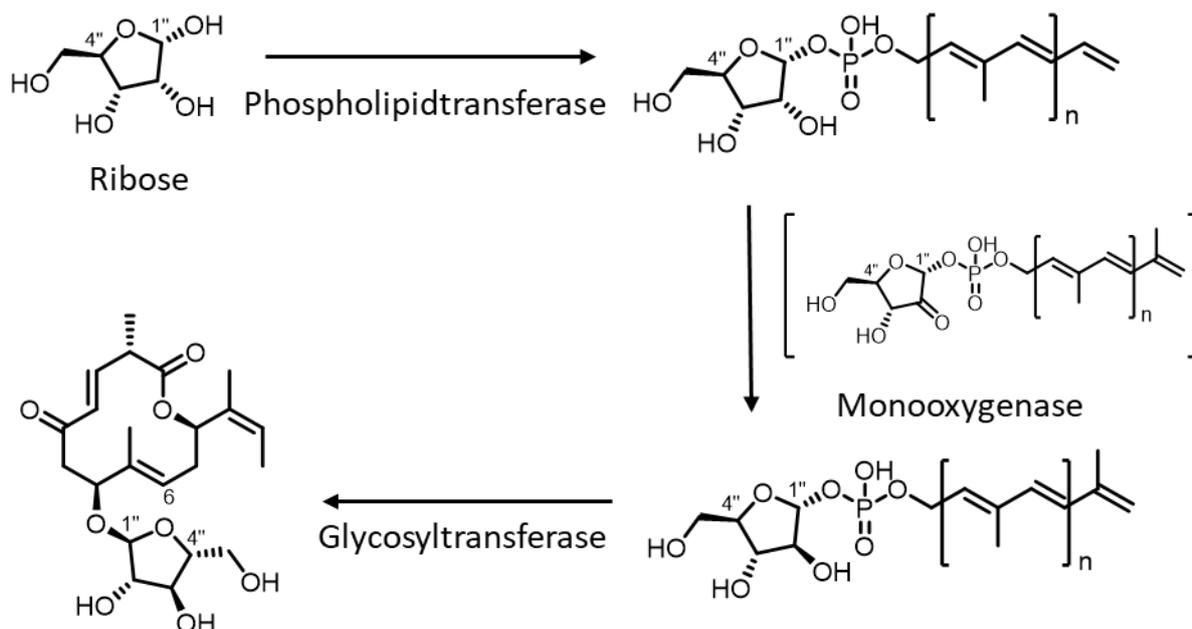
As shown in Figure 3-4, the reduction of the recombinant cytochrome P450 enzyme DifA was achieved only when using the recombinant Fdx/FdR enzymes from *P. fallax* AndGT8,

indicated by the peak at 450 nm after CO-treatment, which only occurs when the iron cofactor is reduced to Fe(II).

Even though different reaction conditions using DifA, Fdx/FdR, NADPH and the potential substrate aglycon DIF294 were evaluated, no conversion of DIF294 was observed *in vitro*. Co-expression of DifA and AndGT8-fdx/fdR in the *M. xanthus* host with the PKS resulted in production of the unmodified polyketide identical to expression of the PKS only. These findings show that DIF294 is not the actual substrate of DifA. We hypothesize that the aglycon might have to be glycosylated prior to hydroxylation by DifA, a mechanism similarly described in the Pikromycin biosynthesis.<sup>163</sup> Publications describing glycosylation with arabinofuranose sugars point towards a sugar lipid as exclusive donor.<sup>164</sup> Since no glycosylated aglycon DIF294 is detectable in the wild type or mutant strains, the only option to obtain a suitable substrate for DifA is a chemical glycosylation or the discovery of the responsible glycosyltransferase for glycosylation of the aglycon. Co-expressing the glycosyltransferase involved in the DIF biosynthesis with the PKS genes could lead to production of the glycosylated precursor scaffold.

### 3.4.2 Investigation of two putative glycosyltransferase operons

Since no glycosyltransferase encoding genes are associated with the DIF BGC, we conducted a bioinformatic analysis of the And GT8 genome in order to identify genes potentially involved in the glycosylation reaction. Several potential glycosyltransferases were discovered and tested in this work by arranging them into two operons and co-transforming them separately into *M. xanthus* DK1622 hosts harboring the core PKS cluster. Operon one (GT1) consists of two mannosyltransferases, and operon two (GT2) of a phospholipid transferase, a monooxygenase, and a glycosyltransferase. We proposed that the phospholipid transferase connects ribose to a dolichol chain. Afterwards, the monooxygenase catalyzes an epimerization of the 2' OH group to form arabinofuranose, which is barely available as free-standing molecule due to the chemical equilibrium with its isomers but is stable in complex with dolichol.<sup>165</sup> Finally, the glycosyltransferase connects the arabinofuranose to the DIF aglycon (see Figure 3-5).



**Figure 3-5** Proposed DIF glycosylation reaction pathway by the enzymes encoded by glycosyltransferase operon 2, discovered by Dr. Fabian Panter and examined in this work

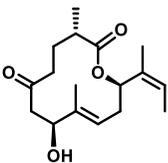
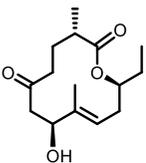
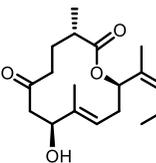
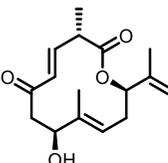
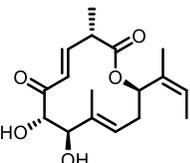
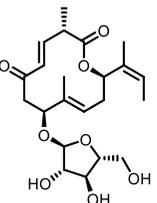
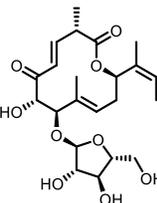
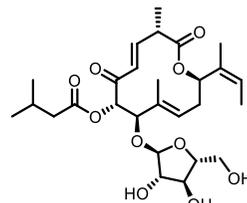
All genes were amplified by PCR from genomic DNA of AndGT8 and combined by overlapping extension PCR to form the two putative glycosyltransferase operons (detailed cloning procedure described in section 3.3.21). The expression of both operons was regulated using a vanillate-inducible promoter system.<sup>139</sup> Each operon was cloned into plasmid pTAR2 which harbors *difA*, Fdx/FdR, as well as the putative tailoring enzymes *difHIJ* to form plasmids pTAR2GT1 and pTAR2GT2, respectively. The three plasmids pTAR2, pTAR2GT1 and pTAR2GT2 were transformed independently into *M. xanthus* DK1622::pMyxZeo\_dif (which harbors the PKS genes *difBCDEFG*) via electroporation and integrated into the genome via the Mx8 integrase system. The obtained strains are shown in Table 3-8.

**Table 3-8** DK1622 mutants obtained and analyzed in this work to investigate the putative glycosyltransferases. GT1 = Glycosyltransferase operon 1; GT2 = Glycosyltransferase operon 2

Strain	Dif genes included
DK1622 wildtype	No modification
DK1622 pMyxZeo_dif	PKS core cluster <i>difB-difG</i>
DK1622 pMyxZeo_dif pTAR2	<i>difB-difG</i> + <i>difH/difI/difJ/difA</i>
DK1622 pMyxZeo_dif pTAR2GT1	<i>difB-difG</i> + <i>difH/difI/difJ/difA</i> + GT1
DK1622 pMyxZeo_dif pTAR2GT2	<i>difB-difG</i> + <i>difH/difI/difJ/difA</i> + GT2

After transformation, the obtained mutants were cultured and extracts analyzed as described in material and methods sections 3.3.7 and 3.3.8 for their production of disciformycin or its intermediates (see Table 3-9).

**Table 3-9** DIF production profile of the generated heterologous DIF producer strains containing the PKS core cluster and tailoring enzymes. Cultures of DK1622 wildtype (WT), DK1622 pMyxZeo\_dif (Dif\_PKS), and the corresponding double mutants (with plasmid pTAR2, pTAR2GT1, and pTAR2GT2, respectively) were investigated for their production of these intermediates. N.d. = not detected.

	Aglycon DIF294	Aglycon DIF268	Aglycon DIF308	Aglycon DIF292	Aglycon DIF292 hydroxylated	Disciformycin intermediate 424	Disciformycin intermediate 440	Disciformycin B
<b>Structure (protonated)</b>								
<b>Retention time [min]</b>	8.9	7.5	9.9	8.5	n.d.	n.d.	n.d.	9.8
<b>Mass [M+H]<sup>+</sup></b>	295.19	269.17	309.21	293.17	309.17	425.22	441.21	525.27
<b>WT</b>	✗	✗	✗	✗	✗	✗	✗	✗
<b>Dif_PKS</b>	✓	✓	✓	✗	✗	✗	✗	✗
<b>Dif_PKS pTAR2</b>	✓	✓	✓	✓	✗	✗	✗	✗
<b>Dif_PKS pTAR2GT1</b>	✓	✓	✓	✓	✗	✗	✗	✗
<b>Dif_PKS pTAR2GT2</b>	✓	✓	✓	✓	✗	✗	✗	✗

As seen in Table 3-9 the co-expression of the two putative glycosyltransferases together with *difA*, *difH*, *difI*, and *difJ* in the DK1622 mutants containing the six DIF PKS genes did not result in production of any hydroxylated or O-glycosylated intermediates. Aglycons DIF294, DIF268, and DIF308 were produced in all mutants harboring the DIF PKS genes. The only further intermediate that could be detected was the dehydrogenated aglycon DIF292 with a double bond at position C-3, indicating that the dehydrogenation is catalyzed by one of the tailoring enzymes expressed by *difH*, *difI* or *difJ*. We currently assume, that the putative glycosyltransferase candidate genes are either not involved in the glycosylation of DIF, or not functional in the DK1622 host strain. Another potential approach to discover potential candidate genes for the glycosylation (and/or the acylation) is a transcriptomics analysis comparing DIF producing and non-producing AndGT8 cultures. We observed media-dependent production of DIF in AndGT8. Thus, a transcriptomics analysis of the AndGT8 DIF producer strain cultivated in different media, might give a hint about the glycosyltransferase genes required for DIF formation. In another approach, cell-protein extracts could be loaded on a column to which the core aglycon of DIF is bound for affinity chromatography. Proteins that bind to the DIF core aglycon substrate are potential tailoring enzymes and should be further analyzed to identify the corresponding genes.

### 3.4.3 Investigation of the putative tailoring enzyme *DifH*

The previous heterologous expression of the tailoring genes *difA*, *difH*, *difI*, and *difJ* together with the DIF core PKS cluster showed that the DIF precursor with double bond at C-3 (aglycon DIF292) is only produced in presence of these tailoring genes. *DifA* was already analyzed in depths and confirmed to be a cytochrome P450.<sup>128</sup> Bioinformatic analysis revealed homology of *difH* to a Fe-S oxidoreductase while *difI* and *difJ* showed homology to a CRP-like DNA-binding protein, and a phosphotransferase, respectively. These findings made *difH* the most promising candidate for a targeted gene deletion to investigate its involvement in the dehydration at C-3 of the aglycon DIF292. Therefore, *difH* on plasmid pTAR2 was replaced with a tetracycline resistance gene (*tetR*) by Red/ET recombineering<sup>124</sup> to form pTAR2 $\Delta$ *difH*::*tetR* (detailed description of the cloning process in section 3.3.22). After the Red/ET process, correct constructs were transformed into DK1622::pMyxZeo\_dif and integrated into the Mx8 *attB* site. Mutants were analyzed by colony PCR to verify successful integration of pTAR2 $\Delta$ *difH*::*tetR*. Subsequently, mutants DK1622::pMyxZeo\_dif-Mx9\_pTAR2 $\Delta$ *difH*::*tetR*-

Mx8 were cultured and the extracts analyzed using uHPLC-HRMS (according to protocol in section 3.3.7 and 3.3.8).

**Table 3-10** Strains obtained in this work including the DIF genes they harbor, tested for their production of aglycon DIF292. All strains containing *difH* produced aglycon DIF292 while the strains lacking *difH*, including the newly obtained *difH* deletion strain, showed no production thereof.

Strain	DIF genes included	DIF292
<b>DK1622 wildtype</b>	No modification	×
<b>DK1622 pMyxZeo_dif</b>	PKS core cluster <i>difB-difG</i>	×
<b>DK1622 pMyxZeo_dif pTAR2</b>	<i>difB-difG</i> + <i>difH/difI/difJ/difA</i>	✓
<b>DK1622 pMyxZeo_dif pTAR2GT1</b>	<i>difB-difG</i> + <i>difH/difI/difJ/difA</i> + GT1	✓
<b>DK1622 pMyxZeo_dif pTAR2GT2</b>	<i>difB-difG</i> + <i>difH/difI/difJ/difA</i> + GT2	✓
<b>DK1622 pMyxZeo_dif pTAR2Δ<i>difH</i> tetR</b>	<i>difB-difG</i> + <i>difI/difJ/difA</i>	×

The deletion of *difH* confirmed that DifH, a putative Fe-S oxidoreductase, is introducing the C-C double bond at C-3. The function of *difI* and *difJ* however are still unknown, but it seems they are not involved in the biosynthesis of DIF. Theoretically, all three or two of the genes *difHJI* could be required for the C-C double bond formation and deleting *difH* lead to disruption of protein-protein interaction. To make sure that DifH is the only required enzyme, it should be expressed alone together with the DIF PKS genes in future analysis. However, due to the homologues of *difJ* and *difI* found in bioinformatic analysis, this is considered very unlikely. Revealing DifH as responsible enzyme to form the double bond in the DIF molecule was a further important step in elucidating the biosynthesis of DIF and only leaves the glycosylation and acylation of the core structure to be uncovered.

### 3.5 Conclusion and outlook

In the here described work the biosynthesis of DIF was investigated. The core PKS cluster *difBCDEFG* was shown before by Dr. Konrad Viehriig to be responsible for the synthesis of the DIF precursor aglycon DIF294, identified by heterologous expression in *M. xanthus* DK1622.<sup>128</sup> *In vitro* analysis of DifA confirmed its functionality and identified the native Fdx/FdR enzymes from *P. fallax* AndGT8 as electron donors. However, no hydroxylation of aglycon DIF294 as substrate for DifA was found, leading to the hypothesis that the glycosylated intermediate is the actual substrate. Further analysis of the tailoring genes adjacent to the PKS core cluster with the heterologous expression system exposed the putative Fe-S oxidoreductase DifH as enzyme introducing the double bond at C-3, leading to precursor aglycon DIF292. Unfortunately, the herein examined potential glycosyltransferases did not lead to a glycosylation of the DIF aglycon. The search for the required enzymes responsible for the glycosylation, as well as the involved acyltransferase, is the main objective of future studies connected to this project. An RNA sequencing approach or pull-down experiment, as described in section 3.4.2, are the most promising approaches to find putative candidate enzymes at this point. A semi-synthetic attempt to bind the  $\alpha$ -arabinofuranose to aglycon DIF294 or DIF292 could also be considered to test the glycosylated intermediate as substrate for DifA. In case of successful hydroxylation at position C-6, the acylation at this position with the isovaleric acid unit could also be performed synthetically. A successful heterologous expression of DIF, and understanding of the biosynthesis, would be essential to improve the production yield and to enable structure engineering. Targeted engineering of the DIF BGC in the heterologous expression platform could lead to production of novel DIF derivatives with improved activities. An overview of successful generation of novel analogs by biosynthetic engineering and/or combinatorial biosynthesis of heterologously expressed biosynthetic pathways is reviewed by Huo, Hug *et al.*<sup>117</sup> These examples show, that it is highly desirable to put efforts in the further investigation of the biosynthesis.

## 3.6 Supporting information

### 3.6.1 List of plasmids and strains generated in this study

**Table 3-11** List of strains generated or used in this study.

Bacterial strain	Genotype	Reference or source
<i>E. coli</i> DH10β	F <sup>-</sup> , <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 <sup>-</sup>	Invitrogen
<i>E. coli</i> HS996	F <sup>-</sup> , <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> NEB 10β	<i>mcrA</i> , <i>spoT1</i> Δ( <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 <sup>-</sup> , <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 <sup>-</sup> , Δ <i>fhuA</i> , P <sub>BAD</sub> - <i>gbaA</i> Δ <i>ybcC</i> , Δ <i>recET19</i>	Gene Bridges
<i>E. coli</i> DH10β pTAR1	<i>E. coli</i> DH10β pTAR1, Kan <sup>R</sup>	Dr. Konrad Viehrig
<i>E. coli</i> DH10β pTAR2	<i>E. coli</i> DH10β pTAR2, Kan <sup>R</sup>	Dr. Konrad Viehrig
<i>E. coli</i> DH10β pTAR4	<i>E. coli</i> DH10β pTAR4, Kan <sup>R</sup>	Dr. Konrad Viehrig
<i>E. coli</i> DH10β pJETGTTAR4	<i>E. coli</i> DH10β pJETGTTAR4, Amp <sup>R</sup>	This work
<i>E. coli</i> DH10β pTAR2GT1	<i>E. coli</i> DH10β pTAR2GT1, Kan <sup>R</sup>	This work
<i>E. coli</i> DH10β pJETMGO(GT2)	<i>E. coli</i> DH10β pJETMGO(GT2), Amp <sup>R</sup>	This work
<i>E. coli</i> DH10β pTAR2GT2	<i>E. coli</i> DH10β pTAR2GT2, Kan <sup>R</sup>	This work
<i>E. coli</i> GB05-red pTAR2	<i>E. coli</i> GB05-red pTAR2, Kan <sup>R</sup>	This work
<i>E. coli</i> GB05-red pTAR2 <i>difHko</i>	<i>E. coli</i> GB05-red pTAR2 <i>difHko</i> , Kan <sup>R</sup> , Tet <sup>R</sup>	This work
<i>E. coli</i> NEB 10β pTAR2 <i>difHko</i>	<i>E. coli</i> NEB 10β pTAR2 <i>difHko</i> , Kan <sup>R</sup> , Tet <sup>R</sup>	This work

<i>M. xanthus</i> DK1622 pMyxZeo_dif	<i>M. xanthus</i> DK1622 pMyxZeo_dif, Amp <sup>R</sup> , Zeo <sup>R</sup>	Dr. Konrad Viehrig
<i>M. xanthus</i> DK1622 pMyxZeo_dif pTAR2	<i>M. xanthus</i> DK1622 pMyxZeo_dif pTAR2, Zeo <sup>R</sup> , Kan <sup>R</sup>	This work
<i>M. xanthus</i> DK1622 pMyxZeo_dif pTAR2GT1	<i>M. xanthus</i> DK1622 pMyxZeo_dif pTAR2GT1, Zeo <sup>R</sup> , Kan <sup>R</sup>	This work
<i>M. xanthus</i> DK1622 pMyxZeo_dif pTAR2GT2	<i>M. xanthus</i> DK1622 pMyxZeo_dif pTAR2GT2, Zeo <sup>R</sup> , Kan <sup>R</sup>	This work
<i>M. xanthus</i> DK1622 pMyxZeo_dif pTAR2ΔdifH	<i>M. xanthus</i> DK1622 pMyxZeo_dif pTAR2ΔdifH, Zeo <sup>R</sup> , Kan <sup>R</sup> , Tet <sup>R</sup>	This work

Table 3-12 List of plasmids generated or used in this study.

Plasmid	Genotype	Reference
pTAR1	CEN6/ARS4, LEU2, p15A ori, P <sub>van</sub> , vanR, difHIJ, Mx8 integrase from <i>Myxococcus</i> phage Mx8, Kan <sup>R</sup>	Dr. Konrad Viehrig
pTAR2	CEN6/ARS4, LEU2, p15A ori, P <sub>van</sub> , vanR, difAHIJ, Fdx-Fdr system, Mx8 integrase from <i>Myxococcus</i> phage Mx8, Kan <sup>R</sup>	Dr. Konrad Viehrig
pTAR4	CEN6/ARS4, LEU2, p15A ori, P <sub>van</sub> , vanR, orf13535 gene and orf11077 gene (both glycosyltransferases from <i>P. fallax</i> AndGT8), Mx8 integrase from <i>Myxococcus</i> phage Mx8, Kan <sup>R</sup>	Dr. Konrad Viehrig
pJET1.2	pUC ori and rep from pMB1, P <sub>lacUV5-eco47I</sub> (endonuclease)/T7 promoter-MCS, bla (Amp <sup>R</sup> )	Thermo Fisher Scientific
pJETGTAR4	pJET1.2, two glycosyltransferases from TAR4	This work introduced
pJETMGO(GT2)	pJET1.2, putative glycosyltransferase genes of operon 2 (section 3.3.21.2) introduced	This work
pTAR2GT1	pTAR2 including glycosyltransferase operon 1 (glycosyltransferases from TAR4)	This work
pTAR2GT2	pTAR2 including glycosyltransferase operon 2 (obtained from pJETMGO(GT2))	This work

<b>pTAR2difHko</b>	pTAR2, <i>difH</i> replaced by Tet <sup>R</sup>	This work
<b>pMyxZeo_dif</b>	<i>Ori rep</i> (pMB1), Amp <sup>R</sup> , Zeo <sup>R</sup> , Mx9 integrase from <i>Myxococcus</i> phage Mx9, P <sub>van</sub> , <i>vanR</i> , <i>difBCDEFG</i>	Dr. Konrad Viehrig

### 3.6.2 Primers used in this work

**Table 3-13** List of additional primers used in this work

Primer name	Primer sequence 5' – 3'
<b>Dif_Pvan_fwd</b>	ATTTAAATGGCTGGACTCTAGCCGACCGACTGAGACG
<b>Dif_Pvan_rev</b>	GCAATATCCCTGACCCGTTGATTTCCGTTTCATATGCGTTTCCTCGCATCGTGGTTCCG
<b>Dif_GT_rev</b>	GCAATTCACGGAAGCCGCCACGCG
<b>Dif_MOX_fwd</b>	CCGAACCACGATGCGAGGAAACGCATATGAACGAAATCAACCGGGTCAGGGATATT GC
<b>Dif_MOX_rev</b>	CGGTCGCCTGAGTCATTTCTGGGGCTTCAGCTCAGCTTCCGGAGGTGGGCC
<b>Dif_MT_ManT_fwd</b>	GGCCACCTCCGGAAGCTGAGCTGAAGCCCCAGAAATGACTCAGGCGACCG
<b>Dif_MT_ManT_rev</b>	AACGTTGCAAGAACAGGGAGGCGGAGCTGAGG
<b>Mod GT operon seq 1</b>	TGTCAAAGGCAATCATCG
<b>Mod GT operon seq 2</b>	GCGAGGAACGGAATGC
<b>Mod GT operon seq 3</b>	GACGTTGTCCTTGTATTCATCC
<b>Mod GT operon seq 4</b>	GAACGTGTCGAGTGAGTTCCG
<b>GT TAR4 seq fwd</b>	GGAGCGACCTTTTGAGGTGTCC
<b>GT TAR4 seq rev</b>	GGACACCTCAAAGGTCGCTCC
<b>GT TAR4 seq 3</b>	CCTGCTCATCGTCAACG
<b>DifHko fwd</b>	GGCTCTTCACGCAGAACCTCGCCGCGCGTGCGGGCCCTGAGCACCTGAATTAATTC TCATGTTTGACA
<b>DifHko rev</b>	ACTCTGCCTTGGGCCACTGTGCCGTGTCGCGCTCGACAAGCGCGGCCGTCTCAGGTTCG AGGTGGCCCGGC



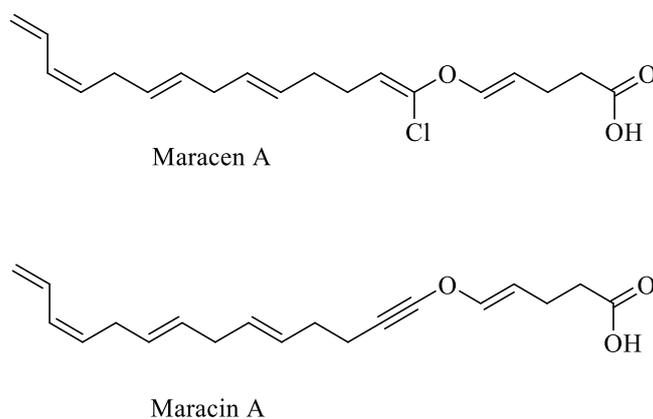
## 4 Towards the heterologous expression of a synthetic maracen/maracin gene cluster

### 4.1 Abstract

Maracen A and Maracin A, first described in 1998 by Herrmann *et al.*,<sup>63</sup> show promising activity against two *Pseudomonas aeruginosa* indicator strains (personal communication with Dr. Jennifer Herrmann). Unfortunately, the elucidation of the biosynthesis is only hypothetical to this date and leaves important questions open.<sup>166</sup> Slow growth and natural resistance against almost all antibiotics of the native *S. cellulosum* producer strains, combined with low genetic amenability, throve the desire for a heterologous expression platform for the maracen/maracin BGC. Previous attempts for a heterologous expression of the native BGC were not successful, however.<sup>166,167</sup> The objective of this work was to design a new version of a synthetic maracen/maracin gene cluster, reorganized in regulated operons inducible by vanillate and exempted from native promoter and terminator sequences, for the heterologous expression in the well-established host strain *M. xanthus* DK1622. Therefore, the 18 genes *mrc1* to *mrc18* of native BGC were reorganized into three operons regulated by a vanillate promoter and tD1 terminator.<sup>139,168</sup> Intergenic regions potentially containing native promoter and terminator sequences were deleted to prevent interference with the new regulation system. Subsequently, the newly designed BGC was synthesized in five fragments and assembled by stepwise restrictive hydrolyzation and ligation. Since the proposed precursor for maracen/maracin biosynthesis is the PUFA derived EPA, the assembled gene construct was transformed into the PUFA gene cluster containing host strain DK1622::pHybPfa1-Mx9.2 *via* the Mx8 integrase. Cultivation of the host strain prior to integration of the new maracen/maracin gene construct revealed sufficient production of the putative precursor EPA. However, after the transformation of the synthetic BGC, no production of maracen A or maracin A was detectable to this date. Future experiments should further investigate the successful integration of the complete construct, re-evaluate the design and restructuring of the *mrc* genes, as well as considering alternative heterologous hosts.

## 4.2 Introduction

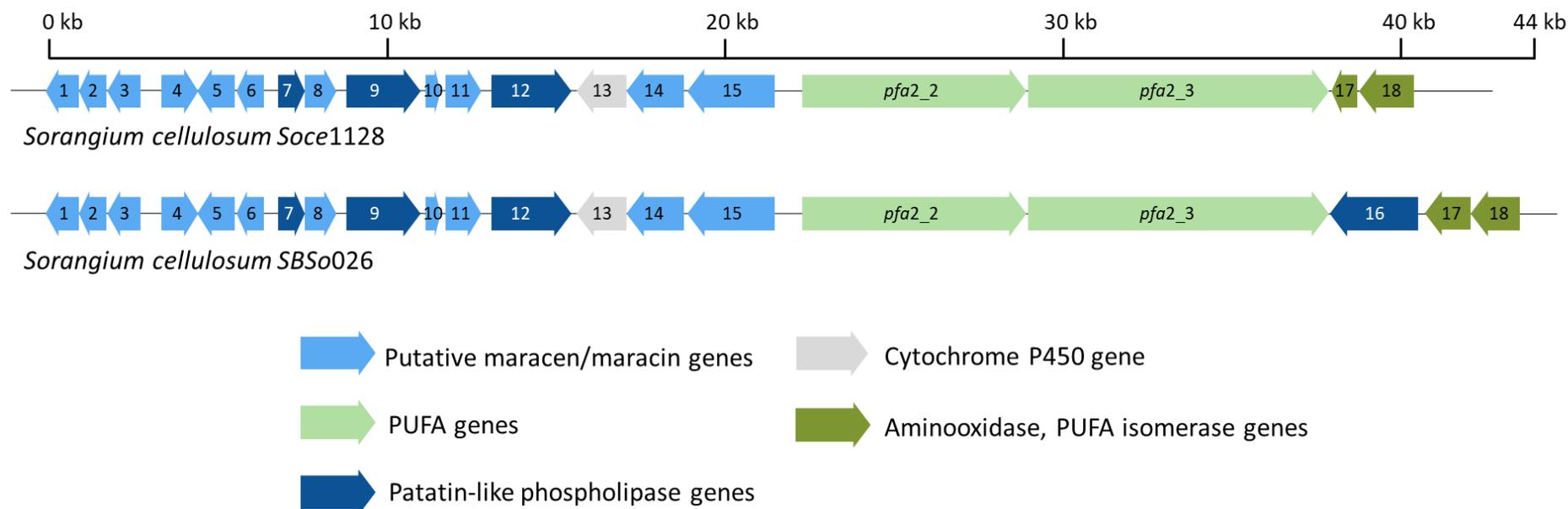
Maracen A and maracin A, two lipophilic carboxylic acids precipitating as colorless oils, are new types of  $\alpha$ -chloro divinyl ether and ethynyl vinyl ether Antibiotics, respectively. They were discovered during a screening program against the nonpathogenic *Mycobacterium phlei* in the two strains *Sorangium cellulosum* Soce1128 and Soce880 in 1998.<sup>63</sup> Hermann *et al.* elucidated the structure of maracin A and maracen A by NMR, revealing that the latter differs from maracin only by the presence of an  $\alpha$ -chlorovinyl instead of the alkynyl group.



**Figure 4-1** Structures of maracen A and maracin A.

*In vitro* activity against the tuberculosis causing pathogen showed an  $IC_{99}$  of  $< 12.5 \mu\text{g/mL}$ . At the same time, a low *in vitro* activity against the mouse fibroblast line L929 indicates a low cytotoxicity.<sup>63</sup> More recent data revealed a promising activity against the two *P. aeruginosa* strains PAO1 (MIC  $5.4 \mu\text{g/mL}$ ), and PA14 (MIC  $4 \mu\text{g/mL}$ ), respectively (personal communication with Dr. Jennifer Herrmann).

Dr. Katrin Jungmann elaborated a biosynthesis hypothesis in her dissertation based on *in silico* analysis of the 18 *mrc* genes located up- and downstream of the *pfa* gene cluster in the maracen/maracin producer strain *Sorangium sp.* SBS026.<sup>166</sup> The native cluster of SBS026 is shown in Figure 4-2 in comparison to native producer strain Soce1128.



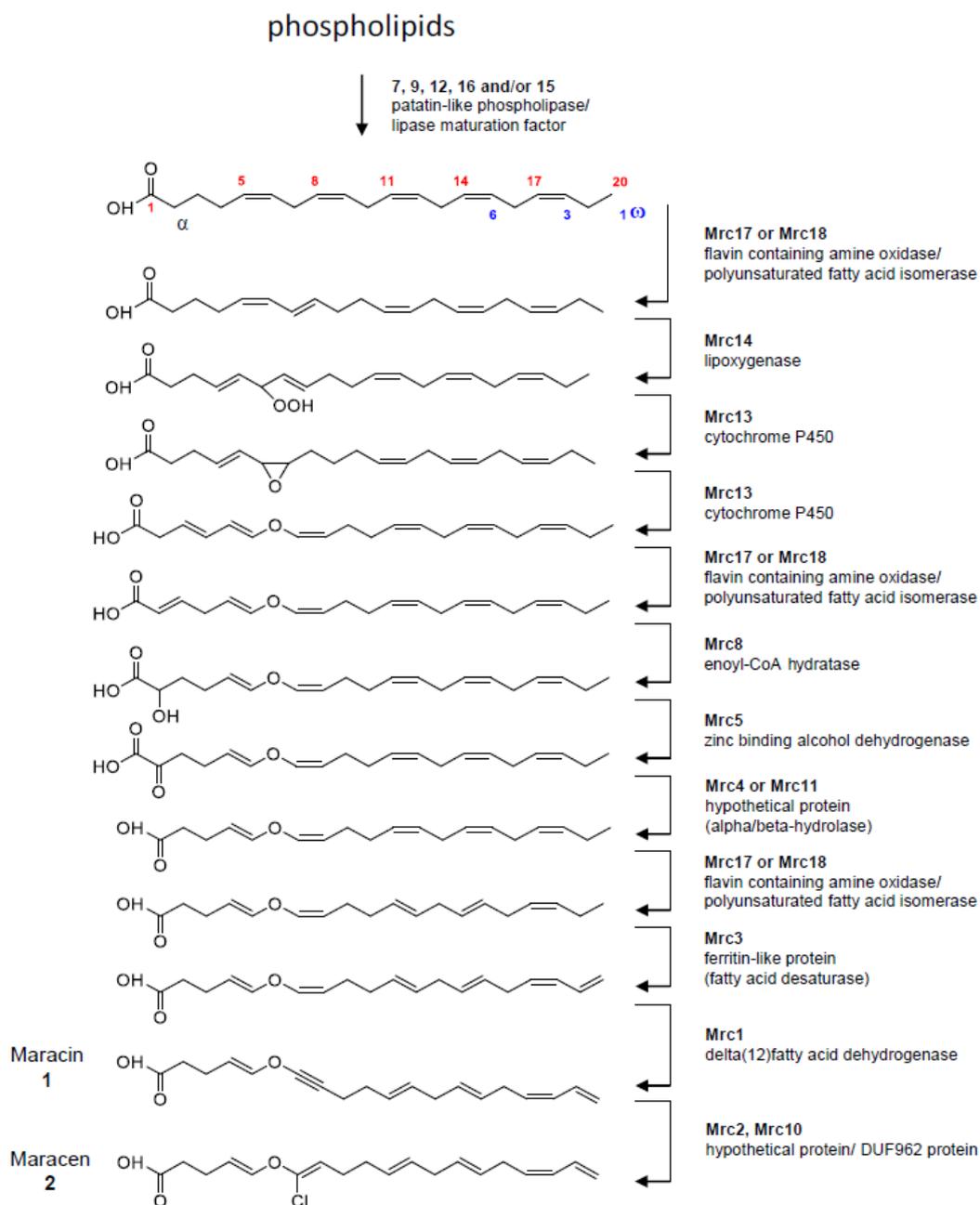
**Figure 4-2** The two putative maracen/maracin gene clusters from producer strains *Soce1128* and *SBSO026*. Genes *mrc1* – *mrc15*, *mrc18*, and the two PUFA genes *pfa2\_2* and *pfa2\_3* are identical in both clusters. The major difference lies in the absence of *mrc16* in the *Soce1128* cluster, as well as the length of *mrc17*, which is significantly shorter in *Soce1128*.

The proposed biosynthesis postulates, that maracen and maracin are derived from the PUFA (polyunsaturated fatty acid) precursor EPA (eicosapentaenoic acid). The hypothetical biosynthesis starts with the release of EPA. This reaction is catalyzed by phospholipases, which could potentially be performed by one of the four patatin-like phospholipases encoded in the putative maracen/maracin BGC (*mrc7*, *9*, *12*, and *16*). In the transition from EPA to maracen/maracin, the fatty acid undergoes several steps of isomerization, which are likely to be catalyzed by one of the two putative isomerases encoded by *mrc17* and *mrc18*. Further steps include oxidization to a 6-hydroperoxide by Mrc14 followed by conversion to an epoxy fatty acid and subsequent transformation into the corresponding divinyl-ether (both by a putative cytochrome P450 encoded by *mrc13*). A later  $\alpha$ -oxidation of the fatty acid and subsequent elimination of the terminal CO<sub>2</sub> moiety are suggested to be catalyzed by the zinc-dependent alcohol dehydrogenase Mrc5 and two candidate genes of unknown function *mrc4* and *mrc11*, respectively. The final steps of the proposed maracen and maracin biosynthesis include a double bond formation at the  $\omega$ -1 position of the fatty acid intermediate, which is suggested to be catalyzed by Mrc3, and the triple bond formation at C6-C7 by Mrc1. The halogenation of the triple bond by chlorine might be catalyzed by one of the enzymes encoded by *mrc2* or *mrc10*, two candidate genes whose functions are still unknown. Recent bioinformatic analysis of *mrc2* and *mrc10* using HHpred<sup>169</sup> indicates similarities to membrane/transport proteins, which would contradict the current biosynthetic hypothesis. The detailed postulated biosynthesis for maracen and maracin is described in the dissertation of Dr. Katrin Jungmann.<sup>166</sup> A figure summarizing the biosynthesis (Figure 4-3) and a table with the putative genes of the maracen/maracin BGC (Table 4-1) are shown below.

**Table 4-1** Putative genes involved in the formation of maracen/maracin. The table shows length, proposed function of the homologous protein, source of the homologous protein, identity and similarity as well as the accession number in GenBank. Table adapted from Dr. Katrin Jungmann.

Gene	Length (bp/aa)	Proposed function of homologous protein	Source of homologous protein	Identity/similarity, %	Accession number (GenBank)
<i>mrc0</i>	1656/552	protein kinase	<i>Sorangium cellulosum</i>	83/85	WP_015351249
<i>mrc1</i>	960/320	delta(12)-fatty dehydrogenase	acid <i>Sorangium cellulosum</i>	39/60	WP_044969276

<b>mrc2</b>	678/226	hypothetical protein				
<b>mrc3</b>	903/301	ferritin, fatty acid desaturase		<i>Burkholderia sp.</i>	52/52	WP_007180466
<b>mrc4</b>	1056/352	hypothetical protein, $\alpha,\beta$ -hydrolase (DUF 2048)		<i>Myxococcus xanthus</i>	46/47	AAO22902
<b>mrc5</b>	1041/347	zinc-dependent dehydrogenase	alcohol	<i>Streptomyces scabiei</i>	41/42	WP_037705470
<b>mrc6</b>	816/272	phospholipid/glycerol acyltransferase		<i>Anaeromyxobacter sp.</i>	43/43	ABS25954
<b>mrc7</b>	768/256	patatin-like phospholipase		<i>Stigmatella aurantiaca</i>	32/33	WP_002618026
<b>mrc8</b>	801/267	enoyl-CoA hydratase		<i>Myxococcus fulvus</i>	56/57	WP_046711249
<b>mrc9</b>	2250/750	patatin-like phospholipase		<i>Stigmatella aurantiaca</i>	27/28	EAU67493
<b>mrc10</b>	417/139	hypothetical protein		<i>Chondromyces apiculatus</i>	58/58	WP_044234628
<b>mrc11</b>	1023/341	hypothetical protein, $\alpha,\beta$ -hydrolase (DUF 2048)		<i>Myxococcus xanthus</i>	46/47	AAO22902
<b>mrc12</b>	2250/750	patatin-like phospholipase		<i>Stigmatella aurantiaca</i>	29/29	EAU67493
<b>mrc13</b>	1431/477	cytochrome P450		<i>Cyanotheca sp.</i>	41/41	WP_012596349
<b>mrc14</b>	1644/548	lipoxygenase		<i>Cystobacter violaceus</i>	46/46	WP_043397944
<b>mrc15</b>	2538/846	lipase maturation factor 1		<i>Parachlamydia acanthamoebae</i>	26/26	KIA77111
<b>mrc16</b>	2250/750	cyclic nucleotidebinding/patatin-like phospholipase domain containing protein		<i>Coralloccoccus coralloides</i>	34/34	AFE10789
<b>mrc17</b>	1341/447	aminoxidase, isomerase	PUFA	<i>Propionibacterium acnes</i>	22/22	WP_002513679
<b>mrc18</b>	1392/464	aminoxidase, isomerase	PUFA	<i>Propionibacterium acnes</i>	26/26	WP_002513679



**Figure 4-3** Hypothetical maracen/maracin biosynthesis as proposed by Dr. Katrin Jungmann. Interestingly, gene *mrcO* was not considered to be involved in the biosynthesis. Figure created by Dr. Katrin Jungmann.<sup>166</sup>

Although the hypothetical biosynthesis seems conclusive, there are some inconsistencies described by Dr. Katrin Jungmann between predicted gene functions and the supposed catalyzed reaction, *e.g.*, the  $\alpha$ -hydroxylation step. Furthermore, several genes are encoded multiple times in the BGC and/or highly identical like the patatin-like phospholipases and the PUFA isomerases. It remains to be investigated if all genes are required for the biosynthesis or whether they can replace each other, *e.g.*, the patatin-like phospholipase domain-containing protein *mrc16*, which is present in producer strain *SBSO026* but missing in

*Soce1128*. For the purpose of further elucidating the biosynthesis and confirming the proposed pathway for maracen/maracin formation, a heterologous expression of the putative maracen/maracin BGC is highly desirable. Two attempts to express a synthetic maracen/maracin gene cluster in the host strains *M. xanthus* DK1622 and *S. cellulosum* strains *Soce1525*, *Soce10*, and *SoceGT47* failed before.<sup>166,167</sup> Transformation attempts into *S. cellulosum* strains were not successful, and heterologous expression in *M. xanthus* DK1622 did not yield production of maracen or maracin. In a new approach, the objective was to reorganize the native cluster to enable the regulation of the genes by the established inducible vanillate promoter and tD1 terminator, a system that was shown to be successful several times before.<sup>71,170</sup> Deletion of all native promoters and terminators and encoding the genes in regulated operons increase the chance of a successful *mrc* gene expression, as it is not known whether the native promoters of the maracen/maracin BGC function in the *M. xanthus* DK1622 host.

## 4.3 Material and methods

### 4.3.1 Media

All media used for liquid cultivations and agar plates were autoclaved for 20 minutes at 121°C before use and kept sterile.

**Table 4-2** Media used in this work with the respective composition

Media	Composition
<b>Lysogeny Broth (LB)</b>	Tryptone 10.0 g/L, yeast extract 5.0 g/L, NaCl 5.0 g/L, agar (for solid media) 15.0 g/L, dH <sub>2</sub> O 1.0 L
<b>2xYT</b>	Tryptone 16.0 g/L, yeast extract 10 g/L, NaCl 5.0 g/L, dH <sub>2</sub> O 1.0 L
<b>CTT</b>	Casitone 10.0 g/L, 1 M Tris-Cl pH 8.0 10.0 mL/L, 1 M K <sub>2</sub> HPO <sub>4</sub> pH 7.6 1.0 mL/L, 1 M MgSO <sub>4</sub> 8.0 mL/L, agar (for solid media) 15.0 g/L and for soft agar 7.5 g/L, dH <sub>2</sub> O 1.0 L
<b>M7/s4</b>	Soy flour 5.0 g/L, corn starch 5.0 g/L, glucose 2.0 g/L, MgSO <sub>4</sub> x 7H <sub>2</sub> O 1.0 g/L, CaCl <sub>2</sub> x 2 H <sub>2</sub> O 1.0 g/L, HEPES 10 g/L, vitamin B12 0.1 mg/L*, FeCl <sub>3</sub> 5 mg/L*  Adjust pH to 7.4 with NaOH

### 4.3.2 Antibiotic stock solutions

Several antibiotics were used in liquid media or agar plates for the selection of genetically altered *E. coli* and *M. xanthus* strains: Ampicillin<sup>[a]</sup> (Amp), Chloramphenicol<sup>[a]</sup> (Cm), Zeocin (Zeo), Oxytetracycline (Otc), and Kanamycin (Kan). Table 4-3 depicts the antibiotic concentrations used in this work.

**Table 4-3** Antibiotics used in this work with their stock and working concentration.

Antibiotic	Stock (1,000x) [mg/ml]	Working conc. [µg/ml] <sup>[b]</sup>
<b>Ampicillin (Amp)</b>	100 1 g in 10 mL MQ-H <sub>2</sub> O	100
<b>Chloramphenicol (Cm)</b>	25 0.25 g in 10 mL 70 % abs. EtOH	25
<b>Kanamycin (Kan)</b>	50 0.5 g in 10 mL MQ-H <sub>2</sub> O	50
<b>Oxytetracycline (Otc)</b>	10 0.1 g in 10 mL 70 % abs. EtOH	10

<sup>[a]</sup> Amp and Cm are not applicable in *M. xanthus* <sup>[b]</sup> For high-copy vectors; low-copy vectors ½ of normal working conc.

Antibiotic stock solutions were stored at -20°C. Oxytetracycline is light sensitive, thus the stock solution was protected from light.

### 4.3.3 Reagents and buffer

Solutions were prepared with demineralized water (dH<sub>2</sub>O) or ultrapure water (MQ-H<sub>2</sub>O). To obtain MQ-H<sub>2</sub>O, 'Milli-Q Reference A+ System' (Millipore, Merck) was used.

For agarose gel electrophoresis 50x TAE buffer containing TRIS 242 g/L, glacial acetic acid 57.1 mL/L, EDTA (500 mM in dH<sub>2</sub>O pH 8.0) 100 mL/L and dH<sub>2</sub>O *ad* 1 L was used. The TAE buffer (50x) stock solution was diluted with dH<sub>2</sub>O to the 1x working concentration before use. Agarose gel solution was prepared by adding 5 g agarose in 500 mL TAE buffer (1x) for 1% solution. For separation of larger DNA fragments 0.8% agarose gel solution was used (4 g agarose in 500 mL 1x TAE buffer). Agarose was dissolved by microwave heating. The solution was stored at 55 °C to prevent solidification.

Orange G loading dye (10x) was used to stain DNA for visualization in agarose gels on a Transilluminator. The solution contains orange G 10 mg, glycerin 3 mL and TE buffer 2 mL. TE buffer consists of TRIS (0.5 M in dH<sub>2</sub>O) 20 mL, EDTA (0.5 M in dH<sub>2</sub>O, pH 8.0) 2 mL and dH<sub>2</sub>O *ad* 1 L. The 10x loading dye was diluted to 1x in the DNA solution that had to be stained.

Buffers P1, P2, and P3 were used for alkaline lysis and plasmid DNA preparation. The composition of each is shown in the table below.

**Table 4-4** Composition of resuspension buffer P1, lysis buffer P2, and neutralization buffer P3. Buffers were used for alkaline lysis and plasmid DNA preparation.

<b>Resuspension buffer P1</b> Stored at 4° C	TRIS (1 M in dH <sub>2</sub> O) 2.5 mL, EDTA (0.1 M, pH 8.0) 5 mL, Ribonuclease A 25 mg, MQ-H <sub>2</sub> O <i>ad</i> 50 mL
<b>Lysis buffer P2</b>	NaOH (1 M in dH <sub>2</sub> O) 10 mL, SDS (20 % in dH <sub>2</sub> O) 2.5 mL, MQ-H <sub>2</sub> O <i>ad</i> 50 mL
<b>Neutralization buffer P3</b>	K-acetate 14.72 g, MQ-H <sub>2</sub> O <i>ad</i> 50 ml, pH 5.5

### 4.3.4 Software

*Geneious 10.1.3* (Biomatters) was used to analyze native biosynthetic gene clusters, to design synthetic gene clusters, to design oligonucleotides for PCR experiments and sequencing, and to design plasmid maps. Agarose gel pictures from the fluorescence chamber

were analyzed with *FusionCapt Advance* Software (Vilber Lourmat). *ChemDraw Ultra 12.0* (CambridgeSoft) was used to create images and chemical formulae. LC chromatograms as well as UV spectra and mass spectrograms were analyzed with DataAnalysis 4.4 (Bruker Daltonics, Billerica, MA, USA).

#### 4.3.5 *E. coli* cultivation

*E. coli* DH10 $\beta$ , NEB 10 $\beta$  or HS996 were used for standard cloning procedures. Except for electroporation for transformation of plasmids, working with *E. coli* was performed under sterile conditions (HeraSafe, Heraeus). For storage at -80°C, 750  $\mu$ L over-night culture was combined with 750  $\mu$ L 50 % glycerin (in MQ-H<sub>2</sub>O) and transferred into a cryo tube.

Cultivation took place in 2xYT or LB medium. Supplementation with antibiotics for selection depended on the plasmid resistance marker. Antibiotic concentrations are shown in Table 4-3. For clone selection, cultures were spread on LB agar (25 mL in a petri dish) after transformation and incubated overnight at 37°C in an incubator. Liquid cultures (5 mL in glass vial or 10-20% v/v liquid media in Erlenmeyer flasks) were incubated overnight at 37°C in a rotary shaker (Multitron, Infors HT) at 200 rpm. Larger constructs of 30 kb size or longer, were incubated at 30°C for a longer period of time (up to two days of incubation).

#### 4.3.6 Preparation of *E. coli* electro competent cells

For preparation of electro competent *E. coli* cells, 5 mL LB medium were inoculated with the respective strain directly from cryo stock and incubated at 37 °C and 200 rpm overnight. 2 mL of overnight culture were used to inoculate 200 mL LB. Cultivation was carried out under the same conditions until an OD<sub>600</sub> (Biophotometer plus, Eppendorf) between 0.4 and 0.6 was reached. The cultures were cooled on ice for 30 min and afterwards transferred into four 50 mL Falcon tubes. All centrifugation steps were performed at 4,000  $\times$  g for 10 min at 4 °C (Avanti J-26 XP Centrifuge, Beckman Coulter; Rotor: JLA 10.500), followed by discarding the supernatant. First time, the cell pellets were resuspended in 100 mL (25 mL each) ice cold HEPES (1 mM in dH<sub>2</sub>O; pH7.0), followed by centrifugation. Second time, the cell pellets were resuspended in 50 mL (12.5 mL each) ice cold HEPES (1 mM in dH<sub>2</sub>O; pH 7.0) and combined in one falcon tube before centrifugation. Third time, the cell pellet was resuspended in 50 mL ice cold 10 % glycerin in HEPES (1 mM in dH<sub>2</sub>O; pH 7.0), followed by centrifugation. Finally, the cell pellet was resuspended in 2 mL 10 % glycerin in HEPES (1 mM in dH<sub>2</sub>O; pH7.0) and

portioned into 50  $\mu$ l aliquots. Now the prepared cells were frozen in liquid N<sub>2</sub> and stored at - 80° C.

#### **4.3.7 Cultivation of *M. xanthus***

*M. xanthus* strains were cultivated either for transformation of heterologous gene constructs or for testing production titers of target compounds. All cultures were incubated at 30° C, either in an incubator (plates) or on shakers at 200 rpm (Erlenmeyer flasks). Production cultures were grown in 25 mL liquid media with corresponding selection marker until sufficient cell density was reached. Subsequently, 5 mL of the culture was sub-cultured into 50 mL liquid medium (10% inoculum). After 24h of incubation, vanillate was added to induce the expression of heterologous genes, in case a vanillate promoter was used for induction of a gene cluster. The cell cultures were harvested after the culture broth took on a brown color (if not mentioned otherwise).

#### **4.3.8 Extraction and analysis of maracen/maracin**

Cultures were extracted twice with 30 mL ethyl acetate for 20 minutes. After extraction, the liquid was centrifuged for 10 minutes and 100  $\mu$ L were transferred to an Eppendorf tube. Before analysis, samples were centrifuged for at least 10 min at 15000 rpm and 50  $\mu$ L of the supernatant were pipetted into a LC-MS vial and submitted for analysis.

For detection of compound UHPLC-hrMS analysis was performed on a Dionex UltiMate 3000 rapid separation liquid chromatography (RSLC) system (Thermo Fisher Scientific, Waltham, MA, USA) coupled to a Bruker maXis 4G ultra-high-resolution quadrupole time-of-flight (UHR-qTOF) MS equipped with a high-resolution electrospray ionization (HRESI) source (Bruker Daltonics, Billerica, MA, USA). The separation of a 1  $\mu$ L sample was achieved with a linear 5–95% gradient of acetonitrile with 0.1% formic acid in ddH<sub>2</sub>O with 0.1% formic acid on an ACQUITY BEH C18 column (100 mm  $\times$  2.1 mm, 1.7  $\mu$ m dp) (Waters, Eschborn, Germany) equipped with a Waters VanGuard BEH C18 1.7  $\mu$ m guard column at a flow rate of 0.6 mL/min and 45 °C for 18 min with detection by a diode array detector at 200–600 nm. The LC flow was split into 75  $\mu$ L/min before entering the mass spectrometer. Mass spectrograms were acquired in centroid mode ranging from 150–2500 m/z at an acquisition rate of 2 Hz in positive MS mode. Source parameters were set to 500 V end-plate offset; 4000 V capillary voltage; 1 bar nebulizer gas pressure; 5 L/min dry gas flow; and 200 °C dry gas temperature. Ion transfer and quadrupole parameters were set to 350 VPP funnel RF; 400 VPP multipole RF; 5 eV ion

energy; and 120 m/z low-mass cut-off. Collision cell was set to 5.0 eV and pre-pulse storage was set to 5  $\mu$ s. Calibration was conducted automatically before every HPLC-MS run by the injection of sodium formate and calibration on the respective clusters formed in the ESI source. All MS analyses were acquired in the presence of the lock masses C<sub>12</sub>H<sub>19</sub>F<sub>12</sub>N<sub>3</sub>O<sub>6</sub>P<sub>3</sub>, C<sub>18</sub>H<sub>19</sub>F<sub>24</sub>N<sub>3</sub>O<sub>6</sub>P<sub>3</sub> and C<sub>24</sub>H<sub>19</sub>F<sub>36</sub>N<sub>3</sub>O<sub>6</sub>P<sub>3</sub>, which generate the [M + H]<sup>+</sup> ions of 622.0289, 922.0098 and 1221.9906. The HPLC-MS system was operated by HyStar 5.1 (Bruker Daltonics, Billerica, MA, USA), and LC chromatograms as well as UV spectra and mass spectrograms were analyzed with DataAnalysis 4.4 (Bruker Daltonics, Billerica, MA, USA).

### **4.3.9 Plasmid DNA isolation from *E. coli***

#### **4.3.9.1 “Mini prep”**

For plasmid DNA isolation from *E. coli*, 2 mL overnight culture was centrifuged for 1 min at 15000 rpm. After discarding the supernatant, the cell pellet was resuspended in 250  $\mu$ l P1 buffer using a bench top shaker. Cell lysis occurred after adding 250  $\mu$ l P2 buffer and inverting the tubes seven times, and incubation for 3 min. By addition of 250  $\mu$ l P3 buffer, 10  $\mu$ l chloroform and inverting seven times, the lysis was stopped. After centrifugation for 10 min at 15000 rpm, the supernatant was transferred into a new 1.5 mL Eppendorf tube containing 500  $\mu$ l ice cold isopropanol. Another centrifugation step for 5 min at 15000 rpm to precipitate the plasmid DNA was followed by discarding the supernatant and adding 700  $\mu$ l 70 % ethanol (in MQ-H<sub>2</sub>O) for washing. Finally, after centrifugation for 1 min at 15000 rpm, the supernatant was discarded and the pellet dried before resuspension in 50 - 100  $\mu$ l MQ-H<sub>2</sub>O.

#### **4.3.9.2 “Midi prep”**

In case larger quantities of isolated plasmid were required for subsequent cloning procedures, 200 mL of *E. coli* overnight culture were grown and harvested. The plasmid was then isolated from the cells using the Nucleo Bond PC100 (midi-prep) DNA purification kit by Macherey-Nagel according to their protocol.

### **4.3.10 Polymerase chain reaction (PCR)**

PCR protocols that were commonly used during this work are described below. Depending on the experiment, template size or objective of the PCR, the protocols might slightly differ from below mentioned parameters.

#### 4.3.10.1 Phusion polymerase

Reaction mix: GC buffer 4.0  $\mu\text{L}$ , dNTPs 5 mM (1.25 mM each) 4.0  $\mu\text{L}$ , Phusion polymerase (Thermo Fisher Scientific) 0.2  $\mu\text{L}$ , primer forward (10  $\mu\text{M}$ ) 0.5  $\mu\text{L}$ , primer reverse (10  $\mu\text{M}$ ) 0.5  $\mu\text{L}$ , DNA template 2  $\mu\text{L}$ \*, MQ-H<sub>2</sub>O 8.8  $\mu\text{L}$ , DMSO 1.0  $\mu\text{L}$ .

\*2  $\mu\text{L}$  genomic DNA, for plasmid DNA 1  $\mu\text{L}$  with a concentration of approximately 10 ng/ $\mu\text{L}$  was used. Volume of MQ-H<sub>2</sub>O had to be adapted accordingly.

Cycle protocol: Initial denaturation 5 min at 98°C; 34 cycles of denature (98°C 20 s), anneal (temperature depends on primers, 25 s), and elongation (72°C 15-30 s/kb); finale elongation at 72°C for 5 min.

#### 4.3.10.2 Taq polymerase

Reaction mix: Taq buffer (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub> 2.5  $\mu\text{L}$ , dNTPs 5 mM (1.25 mM each) 4.0  $\mu\text{L}$ , primer forward (50  $\mu\text{M}$ ) 0.5  $\mu\text{L}$ , primer reverse (50  $\mu\text{M}$ ) 0.5  $\mu\text{L}$ , Taq polymerase (Thermo Fisher Scientific) 0.2  $\mu\text{L}$ , MgCl<sub>2</sub> (50 mM) 1  $\mu\text{L}$ , DNA template 2  $\mu\text{L}$ , glycerol 50% 4  $\mu\text{L}$ , MQ-H<sub>2</sub>O 5.3  $\mu\text{L}$ .

Cycle protocol: Initial denaturation 5 min at 95°C; 30 cycles of denature (95°C 30 s), anneal (temperature depends on primers, 30 s), and elongation (72°C 1 min/kb); finale elongation at 72°C for 5 min.

#### 4.3.11 DNA separation and purification

Agarose gel electrophoresis was used for separation of DNA fragments based on size. In this work, agarose gel electrophoresis was performed in a Consort EV231 chamber (Sigma-Aldrich). An 1% agarose gel was used with 10  $\mu\text{l}$  Roti Safe Gel Stain per 100 mL gel and TAE buffer (1x) as buffer system. After PCR experiments or hydrolyzations with restriction enzymes, OrangeG loading dye (10x) was added to the DNA sample (1/10 of total volume) and the gel was loaded immediately. GeneRuler 1 kb/1 kb plus/100 bp plus DNA ladder (Thermo Fisher Scientific) were used as size standards. The voltage was set between 1 - 5 V/cm until DNA fragments were sufficiently separated. DNA detection took place in a Fusion FX chamber with UV fluorescence detector (Vilber Lourmat).

For DNA purification, agarose gel was placed on a blue light transilluminator, desired DNA fragments were cut out of the agarose gel with a scalpel and transferred into an Eppendorf tube. The proper DNA purification was done with the Macherey-Nagel NucleoSpin Gel and

PCR Clean-up Kit following the manufacturer information. Finally, purified DNA was eluted in 30  $\mu$ l of MQ-H<sub>2</sub>O/elution buffer (from NucleoSpin Gel and PCR Clean-up Kit).

#### 4.3.12 Determination of DNA concentration

NanoDrop 2000c (Thermo Fisher Scientific) was used for determination of DNA concentrations of solutions. For that purpose, 1  $\mu$ l of DNA solution was put on the pedestal and measured at 260 - 280 nm wavelength. MQ-H<sub>2</sub>O served as a reference (blank).

Furthermore, the sample DNA concentration was determined visually by loading 1  $\mu$ l DNA sample on an agarose gel and comparing the DNA band intensity with the intensity of a DNA ladder band of similar size. 6  $\mu$ l of the DNA ladder were loaded as standard (using recommended standard concentrations).

#### 4.3.13 Molecular cloning

For conventional, preparative hydrolyzation/ligation reactions, both vector and insert DNA were treated with type II or type IIS (*Bsa*I) restriction endonucleases (REs). Thereby, DNA is linearized, meaning that vector and insert DNA with matching overhangs can be ligated by T4 DNA ligase. In the following sections, restrictive hydrolyzation of DNA will sometimes be referred to as 'digest' of DNA, which is colloquial use in daily lab environment. All restriction endonucleases, T4 DNA ligase, and FastAP were purchased from Thermo Fisher Scientific.

##### 4.3.13.1 Standard protocol for preparative restrictive hydrolyzation of DNA

Preparative digest was used to hydrolyze sufficient amount of DNA for follow up ligation of insert and vector.

**Table 4-5** Composition of standard preparative restrictive hydrolyzation reactions for single digest (with one enzyme) and double digest (with two enzymes, respectively). DNA solution was usually obtained by DNA isolation (see section 4.3.9) or in some cases by PCR (see section 4.3.10).

Single digest	Double digest
25 $\mu$ l plasmid DNA solution	24 $\mu$ l plasmid DNA solution
3 $\mu$ l 10x restriction endonuclease buffer	3 $\mu$ l 10x restriction endonuclease buffer
2 $\mu$ l restriction endonuclease	1.5 $\mu$ l restriction endonuclease 1
	1.5 $\mu$ l restriction endonuclease 2

Reactions were incubated for several hours or overnight at the required temperature (for most enzymes 37°C). To prevent vector religation, FastAP enzyme was added after t $\frac{1}{2}$  of the

reaction (1  $\mu\text{L}$  in 30  $\mu\text{L}$  reactions). This leads to dephosphorylation of the 5' ends of the linearized vector DNA and therefore improves the overall cloning efficacy.

#### 4.3.13.2 Standard protocol for analytical restrictive hydrolyzation of plasmids containing insert

Analytical digests were used to verify plasmid size and screen for clones harboring plasmids with the right insert after a cloning process.

**Table 4-6** Composition of standard analytical restrictive hydrolyzation reactions for single digest (with one enzyme) and double digest (with two enzymes, respectively). DNA solution was usually obtained by DNA isolation (see section 4.3.9) or in some cases by PCR (see section 4.3.10).

Single digest	Double digest
8.5 $\mu\text{L}$ plasmid DNA solution	8.0 $\mu\text{L}$ plasmid DNA solution
1.0 $\mu\text{L}$ 10x restriction endonuclease buffer	1.0 $\mu\text{L}$ 10x restriction endonuclease buffer
0.5 $\mu\text{L}$ restriction endonuclease	0.5 $\mu\text{L}$ restriction endonuclease 1
	0.5 $\mu\text{L}$ restriction endonuclease 2

Reactions were incubated for approximately 2 hours at the required temperature.

#### 4.3.13.3 Standard ligation reaction

Standard ligation protocol was used to assemble prior linearized DNA fragments. T4 DNA ligase catalyzes bond formation between a 5' phosphate group and a 3' hydroxy group and thereby leads to the ligation of vector and insert DNA. The composition of ligation reaction performed in this work is shown in Table 4-7.

**Table 4-7** Composition of standard ligation reaction, used to assemble linearized DNA.

Component	Volume [ $\mu\text{L}$ ]
<b>Vector DNA</b>	0.5 – 2.0 <sup>[a]</sup>
<b>Insert DNA</b>	6.5 – 8.0 <sup>[a]</sup>
<b>T4 DNA ligase buffer</b>	1.0
<b>T4 DNA ligase</b>	0.5

<sup>[a]</sup> Vector and insert DNA volume depended on determined DNA concentration and varied between 0.5-2  $\mu\text{L}$  for vector DNA solution and 6.5-8  $\mu\text{L}$  for insert DNA solution. Total reaction volume was always 10  $\mu\text{L}$  with a vector/insert DNA molar ratio of around 1/10.

Ligation reaction was incubated for 5 h at room temperature or overnight at 16°C.

#### **4.3.14 *E. coli* transformation**

Transformation of *E. coli* was carried out by electroporation, which shortly increases cell membrane permeability and enables the uptake of plasmid DNA. For electroporation, 50 µl electrocompetent cells were mixed with 5 µl of ligation mix or 0.5 µl of gene synthesis fragment DNA solution and transferred into an electroporation cuvette (electrode distance 1 mm). The electroporation was performed at 1,300 V/cm, 10 µF, and 600 Ω (Eporator V1.01, Eppendorf). Thereafter, the cells were resuspended in 1 mL of LB medium and incubated at 37 °C and 900 rpm (Titramax 1000, Heidolph) for 60 min. After centrifugation at 8000 rpm for 2 min, most of the supernatant was discarded and the cells were resuspended in the remaining medium. In case of gene synthesis fragments, the centrifugation step was skipped and 50 µl were directly used for the following step. The cell suspension was plated out on LB agar with appropriate antibiotic for selection and incubated overnight at 37 °C.

#### **4.3.15 *M. xanthus* DK1622 transformation**

*M. xanthus* DK1622 was cultured overnight in 25 mL CTT medium until an appropriate cell density was reached (OD<sub>600</sub> of 0.6). 2 mL of the culture was pipetted into an Eppendorf tube and centrifuged for 2 min at 6000 rpm and room temperature. Supernatant was discarded. The cell pellet was resuspended in 1 ml sterile MQ water and centrifuged again for 2 min at 6000 rpm, supernatant was again discarded using a pipette. This washing step was repeated once with 800 µL. A small hole was poked into the lid using a needle, and 35 µL of sterile MQ water plus plasmid DNA (5-10 µL DNA/plasmid with a concentration of 100-500 ng/µL DNA) was added. Cell pellet was resuspended and the suspension was used for electroporation at 650 V, 25 µF, and 400 Ω (Bio Rad GenePulser Xcell). After electroporation, the culture was resuspended in 1 mL CTT liquid medium and incubated on a shaker for 6 hours at 30° C and 1000 rpm. A glass vial with foam plug was used to mix 3 mL of CTT soft agar and the respective selection marker with the 1 mL transformation culture. The suspension was thoroughly shaken and plated on a CTT agar plate containing the respective selection marker as well. Plates were incubated on 30° C for several days until clones appeared.

### **4.3.16 Screening for *E. coli* colonies**

To verify a successful transformation and isolate the constructed plasmid, 3 to 24 single *E. coli* colonies were picked and used for inoculation of 5 mL LB medium with appropriate antibiotic. Incubation took place at 37 °C and 200 rpm overnight. Plasmid DNA isolation from *E. coli* was carried out by the protocol described in 4.3.9. Afterwards, an analytical digest was performed (4.3.13.2) and the presence of correct DNA fragments was examined by agarose gel electrophoresis according to 4.3.11.

### **4.3.17 Gibson Assembly**

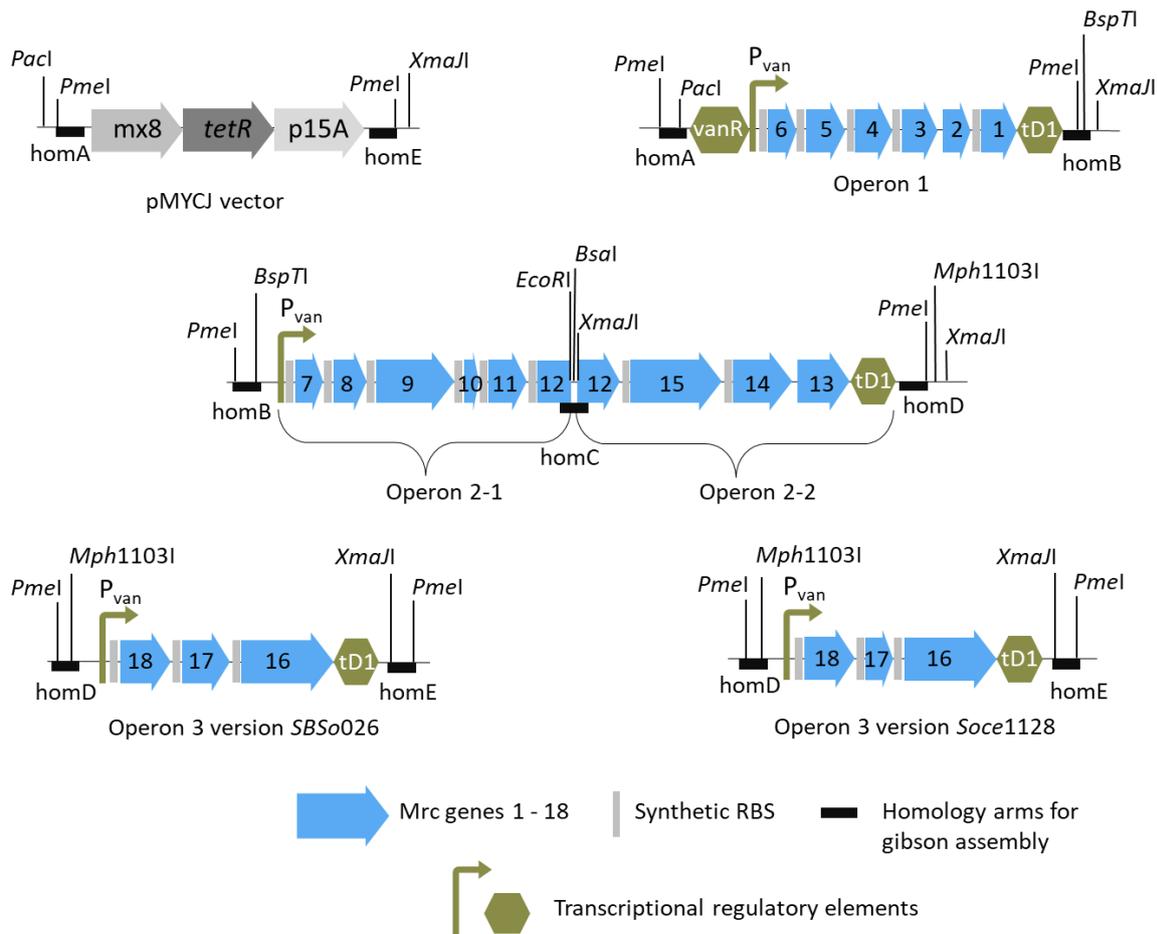
Gibson assembly<sup>121</sup> reaction mix (New England BioLabs) was set up with 1 µL of vector (approx. 40-60ng/µL), 0.2 µL of insert (approx. 100 ng/µL), 1.3 µL of MQ-H<sub>2</sub>O, and 2.5 µL of NEBuilder® HiFi DNA Assembly Master Mix (New England BioLabs). Subsequently, the Gibson reaction mix was incubated in the PCR cycler for 15 minutes at 60°C and transformed into *E. coli* HS996 competent cells to screen for right plasmids.

## 4.4 Results and discussion

### 4.4.1 Design of the synthetic maracen/maracin gene cluster

The design of the synthetic maracen/maracin gene cluster was based on the native BGCs of the two producer strains *Soce1128* and *SBSO026*.<sup>63,171</sup> The proposed native gene cluster consists of the genes *mrc1* – *mrc18* as well as the two polyunsaturated fatty acid metabolism involved genes *pfa2\_2* and *pfa2\_3* and can be seen in Figure 4-2. While *mrc16* was used from producer strain *SBSO026*, all other genes were based on *Soce1128*.

The two gene clusters are almost identical. They only differ in the presence of gene *mrc16*, a putative patatin-like phospholipase domain-containing protein, which is not present in strain *Soce1128*, as well as the length of *mrc17* which is significantly shorter in *Soce1128*. Since the necessity of *mrc16* is not known yet, two versions of the gene cluster were designed. Therefore, the synthetic cluster was split into three operons with the goal to change as little as possible from the native coding sequence. Each operon is regulated by a vanillate promoter and tD1 terminator.<sup>139,168</sup> Operon 1 contains *mrc1* – *mrc6* and operon 2 contains genes *mrc7* – *mrc15*, both operons are identical for both BGC versions. Two versions of operon 3 were designed to enable an analysis of the necessity and function of *mrc16*. Operon 3 *Soce1128* contains genes *mrc16* from *SBSO026* and genes *mrc17* and *mrc18* from *Soce1128*, while operon 3 *SBSO026* contains genes *mrc16* – *mrc18* from *SBSO026*. Furthermore, gene *mrc16* was flanked by unique RE recognition sites to enable removal at a later time point. Reorganizing the genes in this way into the three operons ensured as little change in the native organization as possible. For operon 1 only the direction of *mrc4* had to be inverted and for operon 2 only the CDS of genes *mrc13* – *mrc15* as one unit. The organization of the two synthetic maracen/maracin BGC versions is shown in Figure 4-4 below.



**Figure 4-4** Overview of the five synthetic fragments and the assembling strategy. Each operon is regulated by a vanillate promoter and a tD1 terminator, the vanillate repressor unit is only present in operon 1. Operon 2 was split into two fragments because of the size. Each fragment is flanked by unique RE recognition sites for assembly in the pMYCJ vector, as well as by 20 bp long homology arms (homA – homE) for alternative Gibson assembly. P<sub>van</sub> = vanillate promoter; tD1 = terminator.

While reorganizing genes from the native BGC, intergenic regions that putatively contain native promoters and/or terminators were deleted to prevent interaction with P<sub>van</sub> and tD1. In most cases, those intergenic regions were deleted and replaced by a synthetic RBS (ribosomal binding site) which was used before in the argyris synthetic genes cluster and designed by Dr. Domen Pogorevc.<sup>172</sup> Only the 5'UTRs (untranslated regions) upstream of the two genes *mrc2* and *mrc13* were kept native as they had a clear RBS and were short enough to exclude the possibility of the presence of potential terminators or promoters. In case the intergenic regions including native promoters and terminators are deleted, it is vital to identify the exact starts of the respective genes to avoid accidentally truncating or elongating them, which would potentially result in the expression of non-functional proteins. Each gene was first analyzed with *FramePlot* 4.0beta to find potential gene starts. The translated sequence from the longest version from start to stop codon was then analyzed with *XtalPred* for the

secondary structure.<sup>173</sup> Disordered regions are likely not included in the amino acid sequence, while helices and strands are probably included. The amino acid sequence was further analyzed with Pfam.<sup>174</sup> This protein analysis tool can indicate if the analyzed protein shows truncations compared to related proteins in the data bank, which would mean that the analyzed sequence might be too short. In cases of uncertainty about the gene lengths, the amino acid sequence was further searched in the data bank of HHpred.<sup>169</sup> This tool shows a list of highly related proteins. In case the most related proteins show similar lengths, the analyzed amino acid sequence is more likely to be correct compared to when all related proteins are much shorter/longer.

After replacing intergenic regions with a synthetic RBS, and reorganizing the genes into three operons, *Bsa*I RE recognition sites were removed from the maracen/maracin BGC to enable scarless Red/ET<sup>124</sup> engineering. Therefore, silent point mutations in the *Bsa*I RE recognition sequences were introduced using a codon of the same amino acid with a similar percentage of codon usage. In this way, a change in the translation rate of the protein and therefore risking negative effects on the protein expression should be prevented. Furthermore, the sequence of every start codon after a vanillate promoter was changed to ATG to form a *Nde*I RE recognition site enabling the insertion of genes downstream of the promoter after cluster assembly (see also section 2.4.1.2).

A new backbone pMYCJ was designed containing a tetracycline resistance gene, a p15A ori, and an Mx8 integrase for integration into a *M. xanthus* strain containing the PUFA genes in the Mx9 site.

The synthetic maracen/maracin cluster was divided into six fragments between 4.3 kbp and 7.5 kbp length for synthesis. They were synthesized by GenScript and delivered in pUC57 plasmids for storage and recovery. Each fragment was flanked by RE recognition sites for the release of the pUC57 plasmid and assembly. The RE recognition sites enabled either a one-pot golden gate assembly or a step-by-step assembly, both by restrictive hydrolysis and ligation. Further, release with the blunt end RE *Pme*I enables homology-based Gibson assembly. Operon 2 was split into two fragments for synthesis due to its size. A detailed table about the synthesized fragments including their size, organization, and flanking RE recognition sites for the different assembling strategies is shown in SI section 4.6.2 Table 4-10. Further, Figure 4-4 shows an overview of the synthetic fragments as well as the general assembly strategy.

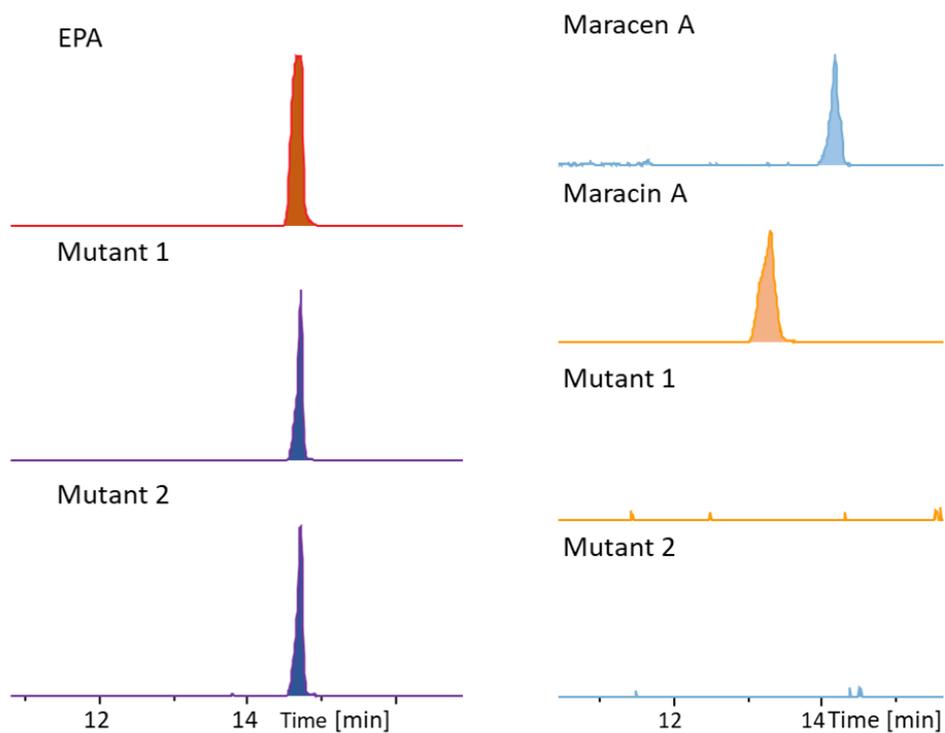
#### 4.4.2 Assembly of the synthetic maracen/maracin gene cluster

The assembly of the synthetic maracen/maracin cluster succeeded with the stepwise cloning procedure according to protocols described in section 4.3.13, 4.3.14, and 4.3.16 (molecular cloning, transformation in *E. coli*, and screening for clones). The detailed assembly process can be seen in Figure 4-1 and SI section 4.6.2. Due to the relatively small size of the gene cluster of 30 kbp, the stepwise assembly process was very efficient. Correct constructs were found after each cloning step and therefore the one-pot assembly, *i.e.*, attempting to assemble all synthetic fragments in one step, was not tried since the chance to succeed might be rather small. Gibson assembly of the synthetic gene cluster was installed as a backup assembly plan in case the cloning would be unsuccessful. This method for assembly of the cluster was tried in parallel to the stepwise cloning two times, but the assembly failed. The two acquired plasmid constructs harboring the two versions of the maracen/maracin synthetic gene cluster pMYCJ\_Mrc\_Full\_Soce1128 and pMYCJ\_Mrc\_Full\_SBS026, respectively, were sequenced by Illumina. Illumina sequencing confirmed the correct and mutation-free sequence of pMYCJ\_Mrc\_Full\_SBS026, while the construct with the Soce1128 version of the gene cluster shows mutations in parts of the cluster and a very low coverage of sequence readouts. It was speculated, that there might be a mix of plasmids harboring different constructs (right and wrong versions) in the screened *E. coli* host. Most likely, the last cloning step of pMYCJ\_Mrc\_Full\_Soce1128 has to be repeated and sent for Illumina sequencing again, to get the correct final construct of this version. Therefore, the following described integration into *M. xanthus* DK1622 and production testing was performed with the correct SBS026 construct version.

#### 4.4.3 Analysis of the heterologous maracen/maracin production with the synthetic gene cluster in the host *M. xanthus* DK1622::pHybPfa1-Mx9.2

The synthetic maracen/maracin gene construct pMYCJ\_Mrc\_Full\_SBS026 was transformed into the *M. xanthus* host containing the PUFA genes DK1622::pHybPfa1-Mx9.2. Colony PCR analysis with Mx8 primers showed positive clones. Mutants were cultured in CTT medium adding vanillate for induction. After a few days of cultivation, when the cultures turned brownish, they were extracted with ethyl acetate and extracts analyzed on LC-MS. However, in the extracts no signals indicative of maracen or maracin were found (see Figure 4-5). To verify the integration of the whole gene construct, multiple primer pairs were

designed spanning the whole gene cluster (primer details in SI section 4.6.3). The colony PCR results could not confirm the successful integration. In most cases, no bands appeared for the used primer pairs for all mutants that showed positive integration with the Mx8 primers. In some cases, no band was visible for the plasmid control while sometimes a band appeared for the WT control. At this point, it cannot be confirmed whether the integration did not work, or the PCR itself is problematic. Possibly some primer pairs are not binding correctly to the target sequence. Further reasons could be insufficient optimized PCR conditions or unspecific primer binding, which would explain bands appearing for the WT control. In many cases several weak bands appeared for the mutants, which also indicates unspecific primer binding. Either way, in future efforts the PCR conditions and primers used have to be revised, and the colony PCR repeated to confirm the right integration. Only after confirming the right integration of the whole gene cluster, further troubleshooting regarding the functionality of the gene cluster or the validity of the extraction and analysis method can be attempted. One could speculate that the gene cluster is not complete, *e.g.*, important genes are missing, or that mistakes in the re-organization of the genes were made, *e.g.*, truncation of crucial proteins. Further, the host could be unable to produce the maracens/maracins. However, at this point, the most relevant suggestion is that the full integration was not successful. The genes *mrc9* and *mrc12* are almost identical, which could lead to homologous recombination and deletions during the integration process (as described in the MXV chapter section 2.4.2.1, dotplot visualization showing the repetitive genes *mrc9* and *mrc12* in SI section 4.6.3).



**Figure 4-5** UHPLC-hrMS analysis of mutants *M. xanthus* DK1622::pHybPfa1-Mx9.2::pMYCJ\_MrcFull\_SBS026, harboring the PUFA genes and putative maracen/maracin gene cluster. The left three chromatograms show a EPA standard and the EPA peaks found in the mutant cultures. The right four chromatograms show a maracen A and maracin A standard, and the two mutant extracts in comparison.

## 4.5 Conclusion and outlook

In the here described project, a synthetic maracen/maracin gene cluster was designed based on the BGCs from the native producer strains *S. cellulosum* Soce1128 and SBS026. The direction of the coding sequence of some genes were inverted to form three operons, all regulated by the inducible vanillate promoter system and tD1 terminators. Intergenic regions that potentially contain native promoter and terminator sequences were deleted and replaced by a synthetic RBS. Due to the uncertainty about the necessity of *mrc16*, two versions of the synthetic BGC were designed. Operons 1 and 2 are both identical and contain genes *mrc1* – *mrc15* from Soce1128, operon 3 version SBS026 contains *mrc16* – *mrc18* from SBS026, while operon 3 version Soce1128 contains *mrc16* from SBS026 and *mrc17* and *mrc18* from Soce1128.

Stepwise assembly of the synthesized fragments for the SBS026 version was successful, while the final construct of version Soce1128 showed some inconsistencies in the Illumina sequencing results. The final construct version pMYCJ\_Mrc\_Full\_SBS026, however, showed a mutation-free sequence by Illumina sequencing and was transformed into the PUFA gene cluster containing *M. xanthus* host DK1622::pHybPfa1-Mx9.2, which is able to produce a sufficient amount of the putative maracen/maracin precursor EPA. Unfortunately, no production of maracen A or maracin A was detected for all obtained mutants. The reasons why the heterologous production did not work can only be speculated to this point. One possible reason could be an incomplete integration of the construct into the host strain. This phenomenon was seen before in the MXV heterologous expression (see section 2.4.2.1). Analysis of the mutants using colony PCR and multiple primer pairs spanning the whole synthetic BGC was inconclusive to this date. In order to confirm the complete integration of the synthetic maracen/maracin construct, the used primers as well as the PCR conditions have to be revised, or genomic DNA of the mutants can be isolated to use as template for the PCR. Once the successful and complete integration is confirmed, but still no compound is produced, further investigations on the design of the gene cluster as well as on the involved genes can be conducted. In future attempts, it is also worth thinking about codon optimization of the gene cluster for optimized expression in the heterologous *M. xanthus* DK1622 host, as well as attempting to integrate the synthetic gene cluster in phylogenetically closer *Sorangium cellulosum* hosts, which however is difficult since no well-established surrogate host of this species is available. Another reason which cannot be ruled out is the genetic completeness of

the biosynthetic pathway. The whole maracen/maracin biosynthetic pathway is currently completely hypothetical without any experimental confirmation, a risk with regards to the plausibility of the biosynthesis model. Notably, the biosynthesis is not module-based so it does not adhere to co-linearity rules, making it difficult to interpret the potential genes.

## 4.6 Supporting information

### 4.6.1 List of plasmids and strains generated or used in this study

**Table 4-8** List of strains generated or used in this study.

Bacterial strain	Genotype	Reference or source
<i>E. coli</i> DH10 $\beta$	F <sup>-</sup> , <i>mcrA</i> , $\Delta(mrr-hsdRMS-mcrBC)$ , $\Phi80lacZ\Delta M15$ , $\Delta lacX74$ , <i>recA1</i> , <i>araD139</i> , $\Delta(ara-leu)7697$ , <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 <sup>-</sup> , <i>mcrA</i> , $\Delta(mrr-hsdRMS-mcrBC)$ , $\Phi80lacZ\Delta M15$ , $\Delta lacX74$ , <i>recA1</i> , <i>araD139</i> , $\Delta(ara-leu)7697$ , <i>galU</i> , <i>galK</i> , <i>rpsL</i> (Str <sup>R</sup> ), <i>endA1</i> , <i>nupG</i> , <i>fhuA::IS2</i>	Invitrogen
<i>E. coli</i> NEB 10 $\beta$	<i>mcrA</i> , <i>spoT1</i> $\Delta(mrr-hsdRMS-mcrBC)$ , $\Phi80d(lacZ\Delta M15)recA1$ , <i>relA1</i> , $\Delta lacX74$ , <i>recA1</i> , <i>araD139</i> , $\Delta(ara-leu)7697$ , <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> NEB10 $\beta$ ::pUC57_Operon1	<i>E. coli</i> NEB 10 $\beta$ pUC57_Operon1, Amp <sup>R</sup>	This work
<i>E. coli</i> NEB10 $\beta$ ::pUC57_Operon2-1	<i>E. coli</i> NEB 10 $\beta$ pUC57_Operon2-1, Amp <sup>R</sup>	This work
<i>E. coli</i> NEB10 $\beta$ ::pUC57_Operon2-2	<i>E. coli</i> NEB 10 $\beta$ pUC57_Operon2-2, Amp <sup>R</sup>	This work
<i>E. coli</i> NEB10 $\beta$ ::pUC57_Operon3 Soce1128	<i>E. coli</i> NEB10 $\beta$ pUC57_Operon3 Soce1128, Amp <sup>R</sup>	This work
<i>E. coli</i> NEB10 $\beta$ ::pUC57_Operon3 SBS026	<i>E. coli</i> NEB10 $\beta$ pUC57_Operon3 SBS026, Amp <sup>R</sup>	This work
<i>E. coli</i> NEB10 $\beta$ ::pMYCJ	<i>E. coli</i> NEB10 $\beta$ pMYCJ, Tet <sup>R</sup>	This work
<i>E. coli</i> NEB10 $\beta$ ::pMYCJ_Op1	<i>E. coli</i> NEB10 $\beta$ pMYCJ_Operon1, Tet <sup>R</sup>	This work
<i>E. coli</i> NEB10 $\beta$ ::pMYCJ_Op1_2-1	<i>E. coli</i> NEB10 $\beta$ pMYCJ_Operon1 and Operon 2-1, Tet <sup>R</sup>	This work
<i>E. coli</i> NEB10 $\beta$ ::pMYCJ_Op1_2	<i>E. coli</i> NEB10 $\beta$ pMYCJ_Operon1 and Operon 2-1 + Operon 2-2 (complete Operon 2), Tet <sup>R</sup>	This work
<i>E. coli</i> NEB10 $\beta$ ::pMYCJ_MrcFull_Soce1128	<i>E. coli</i> NEB10 $\beta$ pMYCJ_MrcFull (Mrc Operon 1, 2, and 3 Soce1128 version), Tet <sup>R</sup>	This work
<i>E. coli</i> NEB10 $\beta$ ::pMYCJ_MrcFull_SBS026	<i>E. coli</i> NEB10 $\beta$ pMYCJ_MrcFull (Mrc Operon 1, 2, and 3 SBS026 version), Tet <sup>R</sup>	This work
<i>M. xanthus</i> DK1622		HIPS/MINS

<b><i>M. xanthus</i> DK1622::pHybPfaMx9.2</b>	<i>M. xanthus</i> DK1622 pHybPfaMx9.2 (PUFA genes), Zeo <sup>R</sup> , Kan <sup>R</sup>	Dr. Katja Gemperlein
<b><i>M. xanthus</i> DK1622::pHybPfaMx9.2::pMYCJ_ MrcFullSoce1128</b>	<i>M. xanthus</i> DK1622 pHybPfaMx9.2 pMYCJ_MrcFull_Soce1128 Mx8, Zeo <sup>R</sup> , KanR, TetR	This work
<b><i>M. xanthus</i> DK1622::pHybPfaMx9.2::pMYCJ_ MrcFullSBSO026</b>	<i>M. xanthus</i> DK1622 pHybPfaMx9.2 pMYCJ_MrcFull_SBSO026 Mx8, Zeo <sup>R</sup> , KanR, TetR	This work

**Table 4-9** List of plasmids generated or used in this work.

Plasmid	Genotype	Reference
<b>pMYCJ</b>	<i>p15A</i> ori, Mx8 integrase from <i>Myxococcus</i> phage Mx8, <i>tetR</i> (Otc <sup>R</sup> )	This work
<b>pMYCJ_Op1</b>	Operon 1 synthetic fragment cloned into pMYCJ	This work
<b>pMYCJ_Op1_2-1</b>	Operon 2-1 synthetic fragment (first part of Operon 2) cloned into pMYCJ_Op1	This work
<b>pMYCJ_Op1_2</b>	Operon 2-2 synthetic fragment (second part of Operon 2) cloned into pMYCJ_Op1_2-1	This work
<b>pMYCJ_MrcFull_Soce1128</b>	Operon 3 version Soce1128 synthetic fragment cloned into pMYCJ_Op1_2, complete synthetic BGC	This work
<b>pMYCJ_MrcFull_SBSO026</b>	Operon 3 version SBSO026 synthetic fragment cloned into pMYCJ_Op1_2, complete synthetic BGC	This work

#### 4.6.2 Design and assembly of the synthetic maracen/maracin gene cluster

The synthesized fragments of the synthetic maracen/maracin gene cluster, including their contained genes, regulatory elements, and the flanking RE recognition sites for the release of synthesis vector and assembly are shown in Table 4-10 below.

**Table 4-10** Organization of the maracen/maracin synthetic gene cluster fragments including genes, regulatory elements, and RE sites for release and assembly.

Operon (size in kbp)	Genes and regulatory elements	Flanking RE recognition sites
<b>Operon 1 (6.9)</b>	<i>mrc1-6</i> (Soce1128), <i>vanR</i> , P <sub>van</sub> , tD1	Stepwise: <i>PacI</i> , <i>XmaI</i> Gibson: <i>PmeI</i> (5' and 3')
<b>Operon 2-1 (6.2)</b>	<i>mrc7-11</i> (Soce1128), first part <i>mrc12</i> (Soce1128), P <sub>van</sub>	Stepwise: <i>BspTI</i> , <i>XmaI</i> Gibson: <i>PmeI</i> 5', <i>EcoRI</i> 3'
<b>Operon 2-2 (7.5)</b>	Second part <i>mrc12</i> (Soce1128), <i>mrc 13-15</i> (Soce1128), tD1	Stepwise: <i>EcoRI</i> , <i>XmaI</i> Gibson: <i>BsaI</i> 5', <i>PmeI</i> 3'
<b>Operon 3 SBS026 (5.2)</b>	<i>mrc16-18</i> (SBS026), P <sub>van</sub> , tD1	Stepwise: <i>Mph1103I</i> , <i>XmaI</i> Gibson: <i>PmeI</i> (5' and 3')
<b>Operon 3 Soce1128 (4.7)</b>	<i>mrc16</i> (SBS026), <i>mrc17-18</i> (Soce1128), P <sub>van</sub> , tD1	Stepwise: <i>Mph1103I</i> , <i>XmaI</i> Gibson: <i>PmeI</i> (5' and 3')
<b>pMYCJ (4.3)</b>	<i>tetR</i> , p15A, Mx8 integrase	Stepwise: <i>PacI</i> , <i>XmaI</i> Gibson: <i>PmeI</i> (5' and 3')

The assembling strategy of the synthetic gene cluster by stepwise restrictive hydrolysis and ligation is shown in Table 4-11 below.

**Table 4-11** Assembling strategy by stepwise restrictive hydrolysis and ligation. The table lists the vector and insert for each step, the RE recognition sites for insertion and the formed plasmid.

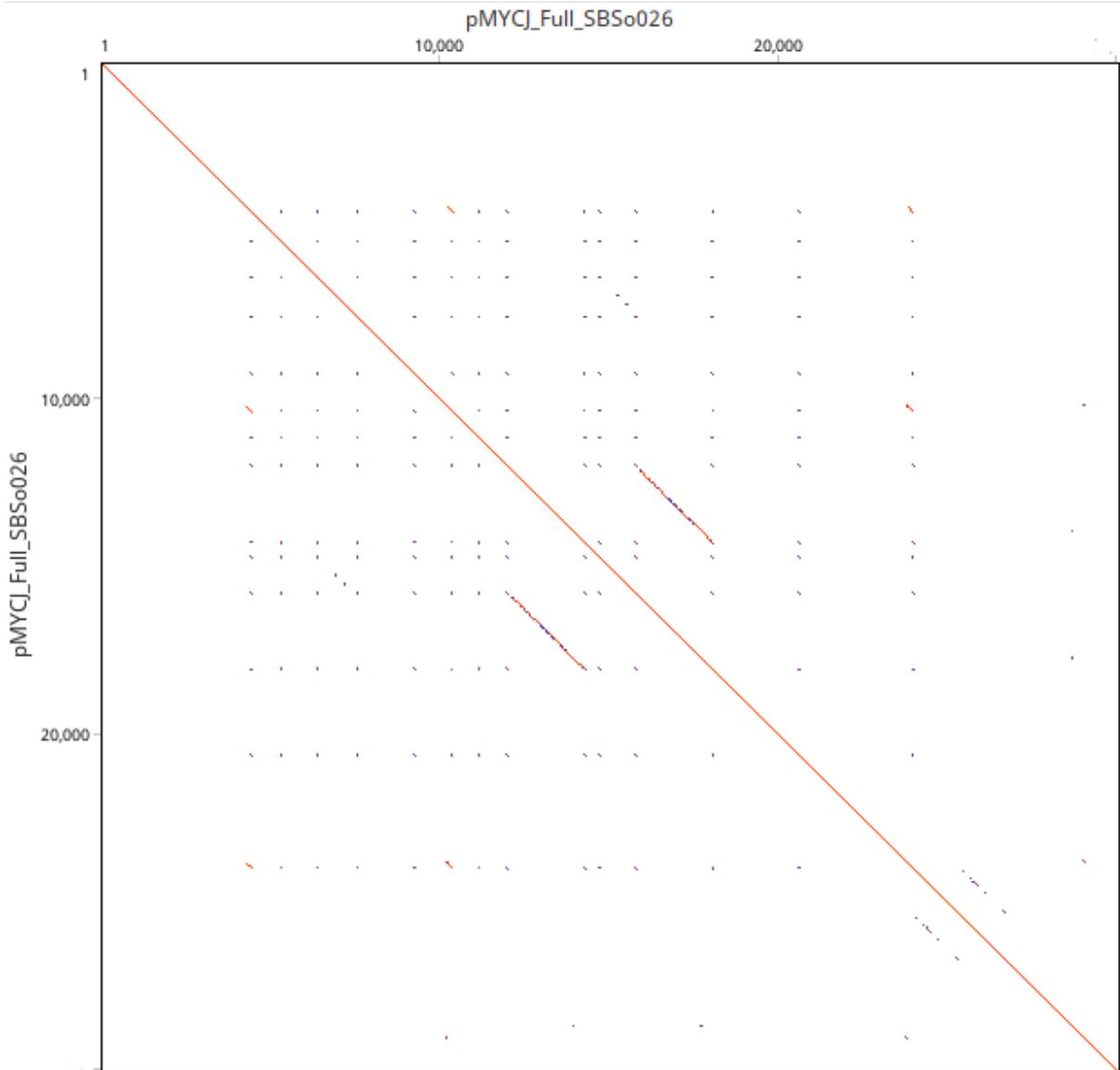
Vector	Insert	REs	Formed plasmid
<b>pMYCJ</b>	Operon 1	<i>XmaI</i> , <i>PacI</i>	pMYCJ_Op1
<b>pMYCJ_Op1</b>	Operon 2-1	<i>XmaI</i> , <i>BspTI</i>	pMYCJ_Op1_2-1
<b>pMYCJ_Op1_2-1</b>	Operon 2-2	<i>XmaI</i> , <i>EcoRI</i>	pMYCJ_Op1-2
<b>pMYCJ_Op1-2</b>	Operon 3 (SBS026)	<i>XmaI</i> , <i>Mph1103I</i>	pMYCJ_Full_SBS026
<b>pMYCJ_Op1-2</b>	Operon 3 (Soce1128)	<i>XmaI</i> , <i>Mph1103I</i>	pMYCJ_Full_Soce1128

### 4.6.3 Analysis of the integration of the synthetic maracen/maracin gene cluster *via* Mx8 integrase

**Table 4-12** Primers used to analyze the integration of the maracen/maracin gene cluster by colony PCR.

Primer name	Primer sequence 5' – 3'
<b>Mrc6 fwd</b>	GCTCATCGACCTGGGGTATT
<b>Mrc6 rev</b>	GCTTCAGGAGCTTCGAGAGG
<b>Mrc2-3 fwd</b>	CGATGTACCGGTCCTGCTAT
<b>Mrc2-3 rev</b>	CCAGGAACAAGGCGCAGGTC
<b>Mrc9 fwd</b>	GCTGATGACCACCGAGGACGCGCAG
<b>Mrc9 rev</b>	CGTGTCGTCTTGTGTGCGTGCTGCC
<b>Mrc10-11 fwd</b>	GGTCGCTCCGGGCGGATTC
<b>Mrc10-11 rev</b>	CCTCGATGTACCGGGCGAGC
<b>Mrc12 fwd</b>	CCGCGGAAGAGGTGCAAAGGCTCC
<b>Mrc12 rev</b>	GCGAGATAGACGAGCCCGACGCC
<b>Mrc13-14 fwd</b>	GGAGCTCTGCGACCTTGTC
<b>Mrc13-14 rev</b>	GGTGCTTCGCCATGTACGAG
<b>Mrc16 fwd</b>	GCGGAACACGAGCTCGGTGC
<b>Mrc16 rev</b>	CGATCTCCGGGACGATCGCC

The figure below shows the two repetitive sequences in the maracen/maracin gene cluster caused by the almost identical genes *mrc9* and *mrc12*.



**Figure 4-6** Dotplot alignment of the gene construct pMYCJ\_Full\_SBS026 against itself visualized by Geneious.

## 5 Discussion

This thesis focused on aspects of myxobacterial natural product (NP) research, with an emphasis on heterologous production of bioactive myxobacterial secondary metabolites and the elucidation of their biosynthesis. The projects of this thesis are split in three chapters. The second chapter describes the *in silico* design of a synthetic myxovalargin (MXV) gene cluster and its heterologous expression in the surrogate host *Myxococcus xanthus* DK1622. The successful heterologous expression initiated in-depth analysis of the MXV biosynthesis by biosynthetic pathway engineering of the expression construct as well as efforts to increase the production yield. The third chapter deals with the macrolide-glycoside antibiotic disciformycin (DIF), isolated from the myxobacterium *Pyxidicoccus fallax* AndGT8 and exhibiting strong activity against methicillin- and vancomycin-resistant *Staphylococcus aureus* (MRSA/VRSA) strains.<sup>61</sup> The  $\alpha$ -chloro divinyl ether and ethynyl vinyl ether antibiotics maracen/maracin, discovered in the two strains *Sorangium cellulosum* Soce1128 and Soce880,<sup>63</sup> are described in the fourth chapter. Both chapters focus on the investigation of the biosynthesis and efforts to discover all involved genes by heterologous expression of the PKS machinery and putative tailoring enzymes (DIF), as well as an *in silico* designed synthetic gene cluster (maracen/maracin), in the heterologous host *M. xanthus* DK1622.

### 5.1 The importance of myxobacterial NPs and sufficient supply thereof

NPs produced by microorganisms are promising candidates for pharmaceutical applications regarding their bioactivity. Through optimization of secondary metabolites over millions of years of evolution to utilize them for their advantage over competitors, those microbial NPs obtained fascinating activities for a large selection of targets.<sup>12</sup> Considering those properties, NP discovery gained increasing attention and importance to harvest those bioactive secondary metabolites for pharmaceutical applications.<sup>12</sup> NPs with antibacterial activity received a special interest during the last decades due to the increasing threat of antimicrobial resistance. Myxobacteria became a promising source for drug discovery and novel lead structure development<sup>48,49</sup> with more than 100 discovered unique core structures exhibiting fascinating bioactivities and modes of action.<sup>51,53,175</sup> However, the journey from the discovery of a NP to the entrance in the market for pharmaceutical application is long, expensive, and requires extensive research such as bioactivity studies, structure analysis,

pharmacokinetic studies, *in vivo* models etc.. In order to conduct these studies a large supply of the desired NP is required, which is in many cases not given by the native producer strain that only produces trace amounts thereof. For this reason, it is necessary to improve the production yield of the desired NP, which is often one of the first objectives after discovery of the compound. Unfortunately, such native strains are often difficult to cultivate under laboratory conditions, which leads to low yields of the associated NP.

## **5.2 The importance of heterologous expression of NP biosynthetic gene clusters (BGC)**

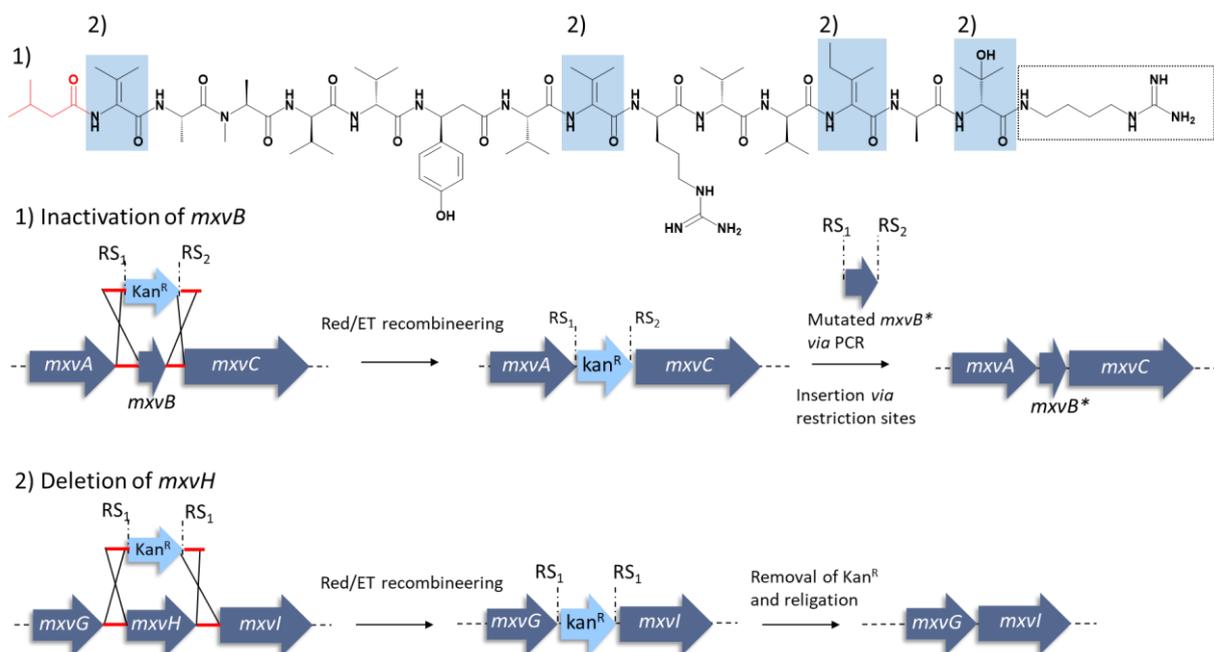
The previous section described the importance of having a reliable supply and sufficient amount of NPs with intriguing bioactivity for in-depth research thereof. Traditional methods like optimization of media and culture conditions can significantly increase the yield of the desired NP. These methods, however, are unspecific regarding targeted yield improvement and can be time consuming with little to no success. More recent and more specific methods applied to improve the supply of NPs are metabolic engineering or heterologous expression of their BGCs. Metabolic engineering can lead to a targeted increase of precursor supply by changing the metabolic pathway of primary and secondary metabolism in favor of the desired NP. Further methods applied in native producers include *e.g.*, induction of spontaneous mutations,<sup>176,177</sup> insertion of more suitable promoters,<sup>178,179</sup> overexpression of positive regulators,<sup>114,180</sup> or disruption of negative regulators.<sup>112,113,181</sup> On the downside, these methods require knowledge about the metabolism and amenability for genetic manipulation. This can be a difficult requirement to meet, especially for myxobacteria, for which only a few studies are available on the investigation of secondary metabolism<sup>149,150,182</sup> and many native producers are not genetically amendable. In this regard, the heterologous expression of NP BGCs in well-established and investigated surrogate hosts can be very beneficial. Such heterologous expression platforms can not only be utilized to optimize the medium composition and cultivation conditions, *e.g.*, optimization of media for the heterologous production of the myxopyronins (MXN) and coralopyronins (COR) in *M. xanthus* DK1622 resulted in a 41-fold increased production titer of MXN A, and a 25-fold increased production of COR A, respectively, compared to the standard CTT medium.<sup>183</sup> The production yield can further be improved by deleting competing BGCs in the host genome that share precursors with the desired NP,<sup>184</sup> or by improving the self-resistance of the host to the NP by

overexpression of the resistance genes. The additional co-expression of the two ABC transporter protein encoding genes *mxvI* and *mxvK* in the mutants *M. xanthus* DK1622::pBeloBacMXV harboring the MXV BGC led to an increased MIC from 32 to >256 mg/L in those producer strains, as described in section 2.4.2.3. Toxicity of bioactive NPs can significantly hamper the production as reaching the MIC limit of the heterologous host will lead to cell death. Furthermore, enabling regulated expression of the target BGCs by inducible or constitutive promoters can lead to improvement of the production,<sup>185</sup> and more importantly, inducible regulation can be very beneficial for production of antibacterial compounds toxic to the surrogate host. Beside applications to increase the production yield of desired NPs, heterologous expression is also an important tool for investigation of the biosynthesis by targeted and seamless deletions, and for the generation of novel derivatives. The latter can be realized by *e.g.*, precursor feeding, mutasynthesis, deletion/insertion of tailoring enzymes or deletion or swapping of modules or domains of the NRPS/PKS megasynthetases and megasynthases. However, to conduct engineering towards production of novel derivatives, it is crucial to understand the biosynthesis of the compound. Furthermore, heterologous expression is pivotal for the emerging field of metagenomics. Functional metagenomics aims to discover NPs of uncultured bacteria by extracting DNA directly from soil samples, clone it into well-established heterologous hosts and screen the mutants for novel chemicals.<sup>186</sup> A significant number of novel compounds has already been discovered by this metagenomics approach, of which a selection are reviewed by Huo, Hug *et al.*<sup>117</sup> For the reasons discussed above, heterologous expression is an indispensable tool for NP discovery and research with potential for future applications.

### **5.3 Investigation of the biosynthesis of myxobacterial secondary metabolites by heterologous expression**

As discussed above, heterologous expression of NP BGCs can be a helpful tool to *e.g.*, elucidate the NP biosynthesis, engineer the structure, or to improve the production yield thereof. These points were the main objectives of the research conducted described in chapter 2–4 of this thesis. The second chapter deals with the heterologous expression of the myxovalgins in the surrogate host *M. xanthus* DK1622. A synthetic BGC was designed based on the native producer strain *Myxococcus fulvus* Mx f65. Genetic disruption experiments in this native strain were previously performed to elucidate the borders of the gene cluster and

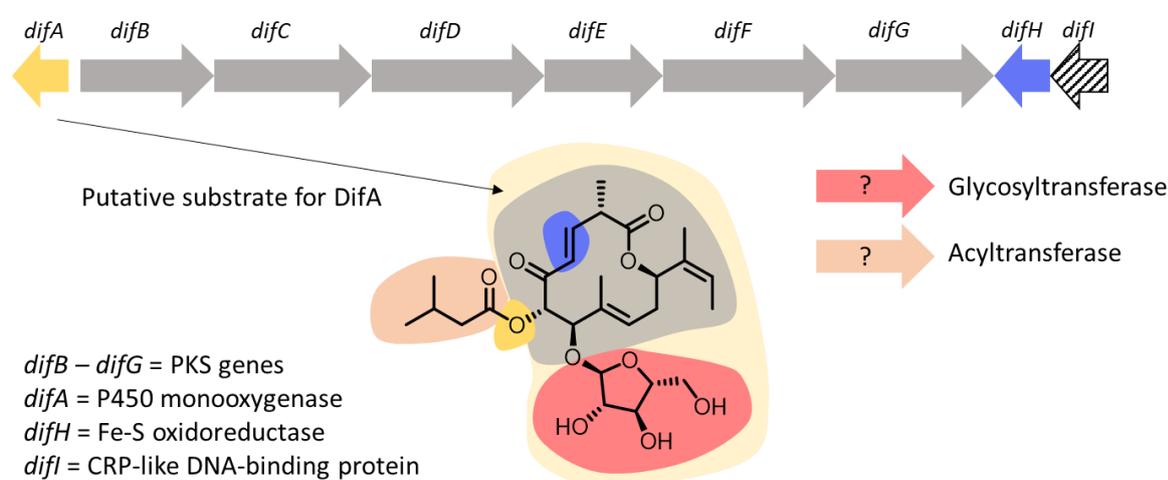
narrow down the MXV BGC to a 66 kbp area including five NRPS and six adjacent genes.<sup>127</sup> Although these knockout experiments located the BGC responsible for the MXV production, the biosynthesis was still hypothetical as additional tailoring enzymes might be involved expressed at distant loci. The successful heterologous expression of the synthetic BGC in this work confirmed the assigned correlation of these genes with the production of MXV and enabled further investigations on the biosynthesis, as not all gene functions could be elucidated previously due to hampered genetic amenability in the native producer strain. As described in section 2.4.3, first efforts were taken into action to investigate the functions of a putative  $\beta$ -hydroxylase expressed by the gene *mxvH* and a single PCP domain expressed by *mxvB*, which features putative portions of an A domain. Previous attempts to investigate *mxvH* and *mxvB* in the native strain provided limited results or were unsuccessful.<sup>127</sup> The terminal agmatine moiety is another intriguing step of the biosynthesis, which can be investigated with the heterologous expression platform. It is currently suggested to be incorporated by the terminal C domain and thereby releasing MXV. The cyanobacterial peptide aeruginoside 126B is the only other known compound with a terminal agmatine moiety and the NRPS assembly line harbors a terminal C and PCP domain.<sup>135</sup> Usually the growing polypeptide is released by a thioesterase I (TE) domain at the C-terminus of a NRPS assembly line. However, many PK and NRP biosynthetic pathways feature a different chain release mechanism, which has been recently reviewed by Little and Hertweck.<sup>187</sup> For example, chain release by condensation-like C<sub>T</sub> domains located at the C-terminus (like TE I domains) *via* macrolactamisation, amidation, Dieckmann condensation, hydrolysis, transesterification, or lactonisation have been described.<sup>187</sup> One noteworthy example is the terminal C<sub>T</sub>-domain-catalyzed chain release in the biosynthesis of wortmanamide B *via* amidation by the achiral amino acids 5-aminopentanoic acid (5PA) or  $\beta$ -alanine.<sup>188</sup> A similar chain release mechanism *via* amidation with agmatine at the terminal C domain could be catalyzed in the MXV biosynthesis. Future approaches should aim to investigate the chain release and incorporation of agmatine by *e.g.*, inactivation of the terminal C domain or replacing it with a TE I domain. Elucidating and understanding the biosynthesis is of great importance for attempts to optimize the structure of MXV to *e.g.*, decrease the cytotoxicity or to improve the pharmacokinetic properties of the molecule. Current approaches aim to investigate the substitution of the terminal agmatine unit by structurally related moieties (personal communication with Dr. Kamaledin Tehrani).



**Figure 5-1** Investigations on the biosynthesis of MXV A by biosynthetic pathway engineering, initiated in this work. **1)** Inactivation of *mxvB* by single point mutation to confirm its function for supply of the isovaleric acid starter unit. **2)** Deletion of *mxvH* to analyze its involvement in the formation of  $\alpha$ ,  $\beta$ -dehydrovaline and –isoleucine and  $\beta$ -hydroxyvaline. Dotted rectangle: The incorporation of the agmatine unit is an intriguing step of the biosynthesis that should be investigated in future experiments. The conducted experiments are discussed in detail in sections 2.4.3.1 and 2.4.3.2.

The completeness of the BGC is crucial for the heterologous expression in a surrogate host. On the other hand, heterologous expression can be used to find the genes involved in the biosynthesis of a target compound, which was the primary goal of the chapters 3 and 4 described in this thesis. The proposed biosynthesis of DIF is based on a PKS cluster found in the native producer *Pyxidicoccus fallax* AndGT8.<sup>61</sup> Heterologous expression of the hypothetical BGC did not deliver production of DIF, it did however produce the DIF core aglycon produced by the PKS genes *difBCDEFG*.<sup>128</sup> The established heterologous expression platform in *M. xanthus* DK1622 was further utilized to search for the additional required tailoring enzymes to form the final DIF structure. Earlier, Dr. Konrad Viehrig had revealed the functionality of DifA as a P450 monooxygenase, consistent with its predicted role in hydroxylation at C-6.<sup>128</sup> In this work, we were able to identify a putative Fe-S oxidoreductase expressed by *difH* which most likely catalyzes the double-bond formation at C-3 of the aglycon (see section 3.4.3). The DIF biosynthesis further requires a glycosylation and acylation, for which the heterologous expression platform should be used for further efforts to discover the responsible tailoring enzymes. Heterologous expression of the complete DIF scaffold would be highly desirable to conduct further research, as this compound shows strong activity

against several Gram-positive and Gram-negative pathogens, including MRSA/VRSA indicator strains.<sup>61</sup>



**Figure 5-2** Overview of the involved genes in the DIF biosynthesis and the parts of the molecule they form. The PKS assembly line expressed by *difBCDEFG* forms the core scaffold (grey) and DifH the double bond at C-3 (blue). DifA is assumed to introduce the hydroxygroup at C-6 (dark yellow) while the glycosylated core scaffold (light yellow) is suggested to be the substrate of DifA at this point. A glycosyltransferase and acyltransferase are still missing for the hypothetical biosynthesis (light and dark red), while *difI* does not seem to be involved.

Problems on a different stage of biosynthesis elucidation and heterologous production were encountered in the attempt to heterologously express the maracen/maracin gene cluster. The biosynthetic pathway is only hypothetical to this point and the heterologous expression of an *in silico* designed and *de novo* synthesized maracen/maracin BGC containing the putative genes *mrc1–18* in *M. xanthus* DK1622 did not yield any production of target molecules. There are several reasons for the unsuccessful heterologous expression, which are discussed in detail in section 4.4.3. One cause could be an incomplete BGC missing important genes or including wrong genes. This reason would be particularly difficult to investigate, since the native producer strains are not amenable to genetic manipulation.

A juxtaposition of the three chapters of this work again underlines the importance of a heterologous production of NPs in well-established surrogate hosts. The possibilities enabled in the MXV project with the successful heterologous expression platform stand in strong contrast to the limitations that were faced in the DIF and maracen/maracin projects. Biosynthesis elucidation, structure optimization by *e.g.*, BGC engineering, and production yield improvement are important constituents in the development of a NP towards its applications as a new drug. Further efforts and troubleshooting in the DIF and maracen/maracin heterologous expressions are therefore highly recommended.

## 5.4 Considerations towards heterologous production of myxobacterial secondary metabolites

Several aspects need to be considered before designing a heterologous expression platform. They include evaluation of a suitable surrogate host, adaption of the regulatory system of the target BGC to suite the new host, as well as an efficient cloning system to mobilize the BGC from the native strain and transfer it to the heterologous host.

The chosen surrogate host should be well-established regarding laboratory cultivation and techniques for genetic manipulation. Methods for a stable genome integration or maintenance of plasmid constructs harboring the non-native BGC, and a suitable regulatory system for the gene expression have to be available. For the expression of myxobacterial BGCs, especially polyketides, *M. xanthus* currently seems to be the best heterologous host.<sup>189</sup> *Pseudomonas putida* was also shown before to be a suitable surrogate host for NPs, e.g., for the production of myxochromide S and pretubulysin.<sup>190–192</sup> For actinomycete-derived BGCs on the other hand, engineered streptomycetes such as *S. coelicolor*<sup>193</sup>, *S. avermitilis*<sup>194</sup>, *S. lividans*<sup>195</sup>, or *S. albus* J1074<sup>184</sup>, have been revealed as advantageous hosts. Several other surrogate hosts such as *E. coli*<sup>196</sup> or *Bacillus subtilis*<sup>197</sup> have been shown to successfully express BGCs of various descends. However, the model host *M. xanthus* is still the best choice for myxobacterial secondary metabolites. BGCs from the myxobacterial suborder *Sorangineae* are less likely to be expressed successfully in *M. xanthus*.<sup>118</sup> It is generally assumed that heterologous expression in closely related hosts is more efficient than in phylogenetically distant hosts because the functionality of transcriptional elements of a BGC such as promoters and ribosomal binding sites is more likely, as well as a similar codon usage renders the translation more efficient.<sup>118</sup> As discussed in section 4.4.3, this could also be a reason for the yet unsuccessful heterologous expression of the *Sorangium cellulosum* descendant maracen/maracin BGC in *M. xanthus* DK1622. Unfortunately, no well-developed *Sorangium cellulosum* host for heterologous expression was established to date, although some strains of that suborder are accessible to genetic manipulation.<sup>118</sup> The establishment of a *S. cellulosum* surrogate host is highly desirable for expression of phylogenetically related *Sorangineae* BGCs. It could be crucial for a successful heterologous expression of the maracen/maracin BGC and enable elucidation of the biosynthesis thereof. However, since no such host is available at the moment, codon optimization of the synthetic BGC for an adaption

of the translation rate in the *M. xanthus* DK1622 host could be a viable attempt to make the heterologous expression work. The regulatory system was already adapted for expression in *M. xanthus* by utilizing a vanillate inducible promoter and a tD1 terminator.<sup>139,168</sup> This system was shown to be successful in *M. xanthus* multiple times before, as just recently shown in the here described heterologous expression of a synthetic MXV BGC. To ensure no interference of native promoters and terminators with the Pvan/tD1 system, intergenic regions were deleted when designing the synthetic maracen/maracin BGC.

Besides deciding for a suitable surrogate host, the desired BGC has to be mobilized into vectors for transformation, integration, and heterologous expression. Traditionally, the BGCs of interest were often obtained from the native producer strain by cosmid, fosmid or bacterial artificial chromosome (BAC) libraries. Recent developments in genome accessibility and genetic engineering enabled to bypass construction and screening of genomic libraries.<sup>117</sup> Multiple methods emerged that showed successful mobilization of NP BGCs into vectors for cloning and heterologous expression, *e.g.*, Gibson assembly<sup>121</sup>, transformation-associated recombination (TAR)<sup>198–200</sup>, linear plus linear homologous recombination (LLHR)<sup>201,202</sup>, and Cas9-assisted targeting of chromosome segments (CATCH)<sup>121</sup> among others. The ExoCET method (Exonuclease Combined with RecET recombination), which combines *in vitro* exonuclease with LLHR enabled direct cloning of the 106 kbp salinomycin BGC into a BAC vector.<sup>203</sup> Methods that enable mobilization of large myxobacterial BGCs are of great importance, as these BGCs typically range from 10 to 120 kbp<sup>117</sup> (*e.g.*, MXV comprises a 66 kbp BGC).

The recently emerged *de novo* synthesis of DNA in combination with *in silico* design of BGCs comes along with the possibility to completely re-organize the BGCs regarding operon structure, regulatory system, and strategy for assembly and cloning. The cloning strategy, however, should be chosen carefully as assembly of such large BGCs containing megasynthetases and megasynthases is not trivial. Some problems that can appear are described in the MXV chapter (sections 2.4.1.3 and 2.4.2). Since the MXV NRPS assembly line contains two large repetitive sequences, likely caused by the incorporation of repeating amino acid building blocks, homologous-based assembly techniques were excluded as possible strategy. TAR cloning for example, based on recombination of homologous regions, could lead to recombination of the said repetitive sequences and cause deletions of essential cluster

parts in between. The classical digestion and ligation approach was chosen to bypass this issue. This strategy however turned out to be very time-consuming, especially since transformation efficiency drastically decreases with increasing size of the inserted BGC<sup>145</sup> and multiple transformation steps in *E. coli* were required for the cloning process. For small BGCs on the other hand, this strategy can be reliable and efficient. The ~30 kbp maracen/maracin BGC was assembled with this technique within five weeks. Additionally, the repetitive sequences seemed to have impeded the transformation and genome-integration process. Several mutants were lacking parts of the cluster between those repetitive sequences that were likely deleted during the integration into the genome (discussed in detail in section 2.4.2.1). Adaptions on the *M. xanthus* transformation protocol seemed to have a positive impact on the integration efficiency, or increasing the chance of integration without deletions, respectively. The same issue was exhibited before in the heterologous expression of the argyrins (personal communication with Dr. Domen Pogorevc) and could have happened in the transformation and integration process of the synthetic maracen/maracin BGC, as this gene cluster contains two genes with almost identical sequences. Further investigations with PCR analysis or sequencing of genomic DNA have to be conducted in the future to confirm the complete integration of the maracen/maracin BGC. Nevertheless, for future approaches to clone BGCs with similar properties as the MXV gene cluster, Gibson assembly could be considered as an alternative cloning strategy. The homologous recombination is based on small regions (~20 bp) on the 5' and 3' ends of the DNA fragments that are processed with exonucleases. Larger repetitive sequences within the DNA fragments should in theory not be problematic. For example, Li *et al.* were able to assemble the 67 kbp long GC-rich pristinamycin II biosynthetic gene cluster from *Streptomyces pristinaespiralis* with a modified Gibson assembly strategy.<sup>204</sup>

## 5.5 Semi-synthesis of NPs as complement to heterologous expression

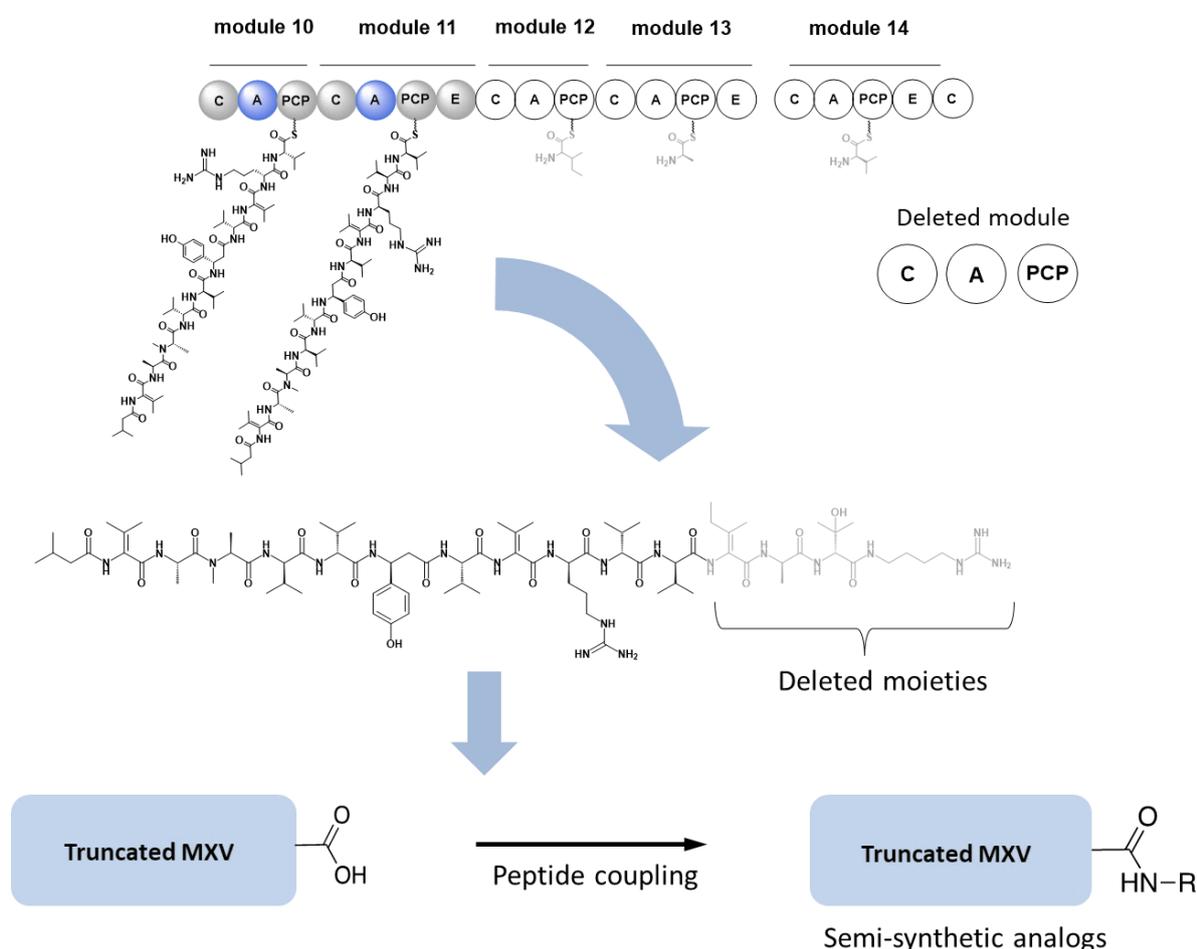
The total synthesis of microbial NPs can be a viable alternative to the heterologous expression in some cases and is often pursued in parallel. A significant number of NPs was already successfully synthesized, *e.g.*, griselimycin<sup>205</sup>, chloromyxamid<sup>206</sup>, and bottromycin<sup>207</sup>, amongst others. Chu *et al.* were even able to manufacture novel antibiotics active against MRSA strains by total synthesis based on bioinformatic prediction of primary sequences,

completely bypassing the need for bacterial cultivation, and gene expression<sup>208</sup>, underlining the benefit total synthesis combined with genome mining can accomplish. However, total synthesis of NPs also features some drawbacks. The large and complex structures of many PKS or NRPS derived NPs require multiple synthesis steps leading to minimal product yields and the presence of stereo centers in the molecules often significantly hamper the synthesis. Moreover, the optimization of the compound by modifying the structure to obtain novel analogues can be difficult to achieve. Heterologous expression of NP BGCs on the other hand, if successful, can lead to superior yields compared to total synthesis and production titers can further be optimized or scaled-up to large fermentation volumes. Additionally, structure optimization by genetic engineering of the BGC can lead to novel analogues with improved bioactivities. Biosynthetic pathway engineering in the darobactin heterologous expression platform resulted in the production of 13 new ‘non-natural’ and four previously hypothetical natural darobactin derivatives.<sup>209</sup> One of the new ‘non-natural’ derivatives exhibited significantly improved activities compared to darobactin A.

Total synthesis of MXV was only recently accomplished<sup>130</sup> (submitted manuscript) after significant efforts over years of a whole dissertation project.<sup>210</sup> Total synthesis of the large NRP MXV containing 14 amino acids (including unusual dehydro- and hydroxy amino acids) is a time consuming and difficult process. Although the heterologous expression of the synthetic MXV BGC was time consuming and difficult as well, the established heterologous expression platform with reasonable yields in this case is superior to the total synthesis considering MXV production and possible structure engineering approaches.

The total synthesis of DIF, which was achieved before by Waser *et al.* and Kwon *et al.*,<sup>162,211</sup> is to date the only possibility to produce DIF aside from cultivation of the native producer. However, the total synthesis could potentially be beneficial for biosynthesis elucidation by analyzing hydroxylation by DifA of a synthetically glycosylated DIF aglycon. This so-called semi-synthetic approach is an advantageous combination of heterologous expression and chemical synthesis. Since the core aglycon of DIF including the C-C double-bond can be produced heterologously in *M. xanthus* DK1622, the missing glycosylation, hydroxylation, and acylation could be catalyzed synthetically, which would bypass multiple synthesis steps. A semi-synthetic approach can also be applied for structure optimization of MXV. Currently, the terminal agmatine unit prevents linkage of moieties to the MXV structure

since no terminal carboxylic group is available. The heterologous expression platform, however, enabled pathway engineering approaches to delete one or multiple terminal modules of the NRPS assembly line. A premature release of truncated myxovalargin derivatives lacking the agmatine unit (or multiple peptides), would amount to a free terminal carboxylic group, enabling linkage of different peptides. These new analogues can be examined for potentially improved physico-chemical properties or antibacterial activity. A popular example is the semi-synthesis of the anti-cancer drug Paclitaxel from the precursor 10-deacetyl baccatin III.<sup>212</sup> Further, the semi-synthesis of the novel tetracycline class antibiotics like doxycycline and tigecycline from NP antibiotic scaffolds obtained by fermentation are noteworthy.<sup>213,214</sup>



**Figure 5-3** Deletion of terminal NRPS modules to enforce the premature release of truncated MXV derivatives. Truncated MXV with a free terminal carboxylic group can serve as template for semi-synthetic modification reactions in order to produce derivatives with improved physico-chemical properties and antibacterial activity profile.

The here discussed comparison of total synthesis and heterologous expression of myxobacterial NPs again underline the benefits of a functioning heterologous expression

platform. However, combining the advantages of both methods in semi-synthetic approaches can be beneficial to obtain novel analogues or to complete an inchoate or partial biosynthesis.

## 5.6 Final conclusion

This thesis focused on the heterologous expression and biosynthesis elucidation of the peptide antibiotic MXV, the macrolide-glycoside antibiotic DIF, and the  $\alpha$ -chloro divinyl ether and ethynyl vinyl ether antibiotics maracen/maracin. The *in silico* designed MXV BGC lead to a successful heterologous expression with promising production yields, opening possibilities for biosynthetic pathway engineering and analysis of important biosynthetic steps. Initial investigations described in this work deal with the putative  $\beta$ -hydroxylase MxvH and the single PCP domain MxvB, and future experiments should confirm their hypothetical functions. Furthermore, the intensive troubleshooting efforts and cloning and transformation protocol adaptations should be beneficial for future projects attempting to mobilize and express large NP BGCs with similar features in the surrogate host *M. xanthus* DK1622. The attempt to fully elucidate the biosynthesis of DIF with a heterologous expression platform in *M. xanthus* DK1622 was only partly successful. In addition to the DIF core scaffold produced by the core PKS genes *difBCDEFG* achieved before, the double-bond formation by a putative Fe-S oxidoreductase expressed by *difH* was revealed. Nevertheless, these are important findings as basis for future attempts to discover the missing tailoring enzymes for the glycosylation and acylation. *In silico* design and heterologous expression of a putative maracen/maracin gene cluster in *M. xanthus* DK1622 did ultimately not result in production of the desired compound or derivatives thereof. Several troubleshooting trials, including lessons learned from the MXV project, as well as efforts like establishing a *Sorangium cellulosum* surrogate host or optimize the codon usage, are some of the aspects highly recommended for future approaches to successfully produce the maracens/maracins heterologously. The efforts of this thesis provided new insights to establish unprecedented heterologous production systems in myxobacteria and can be seen as a compass for future engineering and heterologous expression endeavors.

## 6 References

1. Hanson, J. R. *The classes of natural product and their isolation*. (2003).
2. Pickens, L. B., Tang, Y. & Chooi, Y.-H. Metabolic engineering for the production of natural products. *Annual review of chemical and biomolecular engineering* **2**, 211–236; 10.1146/annurev-chembioeng-061010-114209 (2011).
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. All natural. *Nat Chem Biol* **3**, 351; 10.1038/nchembio0707-351 (2007).
5. Cragg, G. M. & Newman, D. J. Natural products: a continuing source of novel drug leads. *Biochim. Biophys. Acta* **1830**, 3670–3695; 10.1016/j.bbagen.2013.02.008 (2013).
6. Lietava, J. Medicinal plants in a Middle Paleolithic grave Shanidar IV? *Journal of Ethnopharmacology* **35**, 263–266; 10.1016/0378-8741(92)90023-k (1992).
7. Borchardt, J. K. The Beginnings of Drug Therapy: Ancient Mesopotamian Medicine. *Drug news & perspectives* **15**, 187–192 (2002).
8. Huang, K. C. *The pharmacology of Chinese herbs*. 2nd ed. (CRC Press, Boca Raton, London, 1999).
9. Kapoor, L. D. *Handbook of Ayurvedic Medicinal Plants. Herbal Reference Library* (CRC Press, 2000).
10. Dev, S. Ancient-modern concordance in Ayurvedic plants: some examples. *Environ Health Perspect* **107**, 783–789 (1999).
11. Cragg, G. M. & Newman, D. J. Biodiversity: A continuing source of novel drug leads. *Pure and Applied Chemistry* **77**; 10.1351/pac200577010007 (2005).
12. Baker, D. D., Chu, M., Oza, U. & Rajgarhia, V. The value of natural products to future pharmaceutical discovery. *Nat. Prod. Rep.* **24**, 1225–1244; 10.1039/b602241n (2007).

13. Dewick, P. M. *Medicinal Natural Products* (Wiley, 2009).
14. Procópio, R. E. d. L., Silva, I. R. d., Martins, M. K., Azevedo, J. L. d. & Araújo, J. M. d. Antibiotics produced by *Streptomyces*. *Braz. J. Infect. Dis.* **16**, 466–471; 10.1016/j.bjid.2012.08.014 (2012).
15. Pye, C. R., Bertin, M. J., Lokey, R. S., Gerwick, W. H. & Linington, R. G. Retrospective analysis of natural products provides insights for future discovery trends. *Proc. Natl. Acad. Sci. USA* **114**, 5601–5606; 10.1073/pnas.1614680114 (2017).
16. Koehn, F. E. & Carter, G. T. The evolving role of natural products in drug discovery. *Nat. Rev. Drug Discov.* **4**, 206–220; 10.1038/nrd1657 (2005).
17. 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).
18. 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).
19. Bush, K. *et al.* Tackling antibiotic resistance. *Nature reviews. Microbiology* **9**, 894–896; 10.1038/nrmicro2693 (2011).
20. Lewis, K. Antibiotics. Recover the lost art of drug discovery. *Nature* **485**, 439–440; 10.1038/485439a (2012).
21. Gould, I. M. Antibiotic resistance: the perfect storm. *Int. J. Antimicrob. Agents* **34**, S2–S5; 10.1016/S0924-8579(09)70549-7 (2009).
22. World Health Organization. Global Antimicrobial Resistance Surveillance System (GLASS) Report. Early implementation. Available at <http://apps.who.int/iris/bitstream/handle/10665/259744/9789241513449-eng.pdf;jsessionid=97E045EBFE3B29F286CBA3EE8360F0D1?sequence=1> (2018).
23. 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).

24. Aminov, R. I. & Mackie, R. I. Evolution and ecology of antibiotic resistance genes. *FEMS microbiology letters* **271**, 147–161; 10.1111/j.1574-6968.2007.00757.x (2007).
25. Bhullar, K. *et al.* Antibiotic resistance is prevalent in an isolated cave microbiome. *PLoS ONE* **7**, e34953; 10.1371/journal.pone.0034953 (2012).
26. Blair, J. M. A., Webber, M. A., Baylay, A. J., Ogbolu, D. O. & Piddock, L. J. V. Molecular mechanisms of antibiotic resistance. *Nature Reviews Microbiology* **13**, 42–51; 10.1038/nrmicro3380 (2014).
27. Li, X.-Z. & Nikaido, H. Efflux-Mediated Drug Resistance in Bacteria. *Drugs* **69**, 1555–1623; 10.2165/11317030-000000000-00000 (2009).
28. Stekel, D. First report of antimicrobial resistance pre-dates penicillin. *Nature* **562**; 10.1038/d41586-018-06983-0 (2018).
29. Hutchings, M. I., Truman, A. W. & Wilkinson, B. Antibiotics: past, present and future. *Current Opinion in Microbiology* **51**, 72–80; 10.1016/j.mib.2019.10.008 (2019).
30. Dublanchet, A., Soussy, C. J., Squinazi, F. & Duval, J. Résistance de *Staphylococcus aureus* aux streptogramines. *Ann Inst Pasteur Microbiol* **128A**, 277–287 (1977).
31. Emmerson, A. M. & Jones, A. M. The quinolones: decades of development and use. *J Antimicrob Chemother* **51**, 13–20; 10.1093/jac/dkg208 (2003).
32. Martin Gellert, Kiyoshi Mizuuchi, Mary H. O'Dea, Tateshi Itoh & Jun-Ichi Tomizawa. Nalidixic acid resistance: A second genetic character involved in DNA gyrase activity. *PNAS* **74**, 4772–4776; 10.1073/pnas.74.11.4772 (1977).
33. 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).
34. 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).

35. 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).
36. M. K. Hayden *et al.* Development of Daptomycin Resistance In Vivo in Methicillin-Resistant *Staphylococcus aureus*. *Journal of clinical microbiology* **43**, 5285–5287; 10.1128/JCM.43.10.5285-5287.2005 (2005).
37. Kathryn Sabol *et al.* Emergence of Daptomycin Resistance in *Enterococcus faecium* during Daptomycin Therapy. *Antimicrobial agents and chemotherapy* **49**, 1664–1665; 10.1128/AAC.49.4.1664-1665.2005 (2005).
38. O'Neil, J. *Antimicrobial Resistance: Tackling a crisis for the health and wealth of nations* (2014).
39. 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).
40. Boucher, H. W. *et al.* Bad Bugs, No Drugs: No ESKAPE! An Update from the Infectious Diseases Society of America. *Clinical Infectious Diseases* **48**, 1–12; 10.1086/595011 (2009).
41. Kupferschmidt, K. Resistance fighters. *Science* **352**, 758–761; 10.1126/science.352.6287.758 (2016).
42. Kostyanov, 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).
43. 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).

- 
44. Piel, J. Approaches to capturing and designing biologically active small molecules produced by uncultured microbes. *Annu. Rev. Microbiol.* **65**, 431–453; 10.1146/annurev-micro-090110-102805 (2011).
45. Subhan, M., Faryal, R. & Macreadie, I. Exploitation of *Aspergillus terreus* for the Production of Natural Statins. *Journal of fungi (Basel, Switzerland)* **2**; 10.3390/jof2020013 (2016).
46. Mannanov, R. N. Antibiotics produced by *Bacillus* bacteria **37**, 117–123 (2001).
47. Garcia, R., Gerth, K., Stadler, M., Dogma Jr., I. J. & Müller, R. Expanded phylogeny of myxobacteria and evidence for cultivation of the 'unculturables'. *Mol. Phylogenet. Evol.* **57**, 878–887; 10.1016/j.ympev.2010.08.028 (2010).
48. Reichenbach, H. Myxobacteria, producers of novel bioactive substances. *J. Ind. Microbiol. Biotechnol.* **27**, 149–156 (2001).
49. Gerth, K., Pradella, S., Perlova, O., Beyer, S. & Müller, R. Myxobacteria: proficient producers of novel natural products with various biological activities - past and future biotechnological aspects with the focus on the genus *Sorangium*. *J. Biotechnol.* **106**, 233–253; 10.1016/j.jbiotec.2003.07.015 (2003).
50. Ringel, S. M. *et al.* Ambruticin (W7783), a new antifungal antibiotic. *J. Antibiot.* **30**, 371–375; 10.7164/antibiotics.30.371 (1977).
51. Weissman, K. J. & Müller, R. Myxobacterial secondary metabolites: bioactivities and modes-of-action. *Nat. Prod. Rep.* **27**, 1276–1295; 10.1039/c001260m (2010).
52. 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).
53. Schäberle, T. F., Lohr, F., Schmitz, A. & König, G. M. Antibiotics from myxobacteria. *Nat. Prod. Rep.* **31**, 953–972; 10.1039/c4np00011k (2014).
54. Dawid, W. Biology and global distribution of myxobacteria in soils. *FEMS Microbiol. Rev.* **24**, 403–427; 10.1111/j.1574-6976.2000.tb00548.x (2000).
-

55. 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).
56. David Edward Whitworth. *Myxobacteria. Multicellularity and differentiation* (ASM Press, 2008).
57. 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).
58. 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).
59. Zaburanyi, N., Bunk, B., Maier, J., Overmann, J. & Muller, 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).
60. Baumann, S. *et al.* Cystobactamids: myxobacterial topoisomerase inhibitors exhibiting potent antibacterial activity. *ANGEWANDTE CHEMIE-INTERNATIONAL EDITION* **53**, 14605–14609; 10.1002/anie.201409964 (2014).
61. Surup, F. *et al.* Disciformycins A and B: 12-membered macrolide glycoside antibiotics from the myxobacterium *Pyxidicoccus fallax* active against multiresistant staphylococci. *Angew. Chem. Int. Ed. Engl.* **49**, 13588–13591; 10.1002/anie.201406973 (2014).
62. Irschik, H., Gerth, K., Kemmer, T., Steinmetz, H. & Reichenbach, H. The Myxovalargins, new peptide antibiotics from *Myxococcus fulvus* (Myxobacterales) I. Cultivation, isolation, and some chemical and biological properties. *J. Antibiot.* **36**, 6–12 (1983).
63. Herrmann, M., Bohlendorf, B., Irschik, H., Reichenbach, H. & Hofle, G. Maracin and maracen: New types of ethynyl vinyl ether and alpha-chloro divinyl ether antibiotics from *Sorangium cellulosum* with specific activity against mycobacteria. *Angew. Chem. Int. Ed. Engl.* **37**, 1253–1255 (1998).

- 
64. Mulzer, J., Altmann, K. H., Hofle, G., Müller, R. & Prantz, K. Epothilones - a fascinating family of microtubule stabilizing antitumor agents. *Comptes Rendus Chimie* **11**, 1336–1368; 10.1016/j.crci.2008.02.005 (2008).
65. Plaza, A. *et al.* Aetheramides A and B, potent HIV-inhibitory depsipeptides from a myxobacterium of the new genus "Aetherobacter". *Organic letters* **14**; 10.1021/ol3011002 (2012).
66. Sasse, F. *et al.* Argyrins, immunosuppressive cyclic peptides from myxobacteria. I. Production, isolation, physico-chemical and biological properties. *J. Antibiot.* **55**, 543–551 (2002).
67. Held, J. *et al.* Antimalarial activity of the myxobacterial macrolide chlorotonil A. *Antimicrob. Agents Chemother.* **58**, 6378–6384; 10.1128/AAC.03326-14 (2014).
68. 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).
69. Mulzer, J. (ed.). *The Epothilones, an Outstanding Family of Anti-Tumor Agents* (Springer, New York, 2009).
70. 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).
71. Groß, S., Schnell, B., Haack, P. A., Auerbach, D. & Müller, R. In vivo and in vitro reconstitution of unique key steps in cystobactamid antibiotic biosynthesis. *Nat. Commun.* **12**, 1696; 10.1038/s41467-021-21848-3 (2021).
72. 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).
73. Hertweck, C. The Biosynthetic Logic of Polyketide Diversity. *Angew. Chem. Int. Ed. Engl.* **48**, 4688–4716; 10.1002/anie.200806121 (2009).
-

74. Du, L. H., Sanchez, C. & Shen, B. Hybrid peptide-polyketide natural products: biosynthesis and prospects toward engineering novel molecules. *Metab. Eng.* **3**, 78–95 (2001).
75. Beld, J., Sonnenschein, E. C., Vickery, C. R., Noel, J. P. & Burkart, M. D. The phosphopantetheinyl transferases: catalysis of a post-translational modification crucial for life. *Nat. Prod. Rep.* **31**, 61–108; 10.1039/c3np70054b (2013).
76. Süssmuth, R. D. & Mainz, A. Nonribosomal peptide synthesis - Principles and prospects. *Angew. Chem. Int. Ed.* **56**, 3770–3821; 10.1002/anie.201609079 (2017).
77. 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).
78. 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).
79. 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).
80. 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).
81. Weckwerth, W. *et al.* Biosynthesis of PF1022A and related cyclooctadepsipeptides. *J. Biol. Chem.* **275**, 17909–17915; 10.1074/jbc.M001084200 (2000).
82. Labby, K. J., Watsula, S. G. & Garneau-Tsodikova, S. Interrupted adenylation domains: unique bifunctional enzymes involved in nonribosomal peptide biosynthesis. *Nat. Prod. Rep.* **32**, 641–653; 10.1039/c4np00120f (2015).
83. Silakowski, B. *et al.* New lessons for combinatorial biosynthesis from myxobacteria. The myxothiazol biosynthetic gene cluster of *Stigmatella aurantiaca* DW4/3-1. *J. Biol. Chem.* **274**, 37391–37399; 10.1074/jbc.274.52.37391 (1999).

- 
84. Wenzel, S. C. & Müller, R. Myxobacterial natural product assembly lines: fascinating examples of curious biochemistry. *Nat. Prod. Rep.* **24**, 1211–1224; 10.1039/b706416k (2007).
85. Kohli, R. M. & Walsh, C. T. Enzymology of acyl chain macrocyclization in natural product biosynthesis. *Chem. Commun.*, 297–307; 10.1039/b208333g (2003).
86. 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).
87. Staunton, J. & Weissman, K. J. Polyketide biosynthesis: a millennium review. *Natural product reports* **18**, 380–416 (2001).
88. Moore, B. S. & Höpke, J. N. Discovery of a new bacterial polyketide biosynthetic pathway. *ChemBioChem* **2**, 35–38; 10.1002/1439-7633(20010105)2:1<35::AID-CBIC35>3.0.CO;2-1 (2001).
89. 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).
90. 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).
91. 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).
92. 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).
93. 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).
-

94. Döhren, H. von, Dieckmann, R. & Pavela-Vrancic, M. The nonribosomal code. *Chem. Biol.* **6**, R273-R279 (1999).
95. 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).
96. 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).
97. Prieto, C., García-Estrada, C., Lorenzana, D. & Martín, J. F. NRPSp: non-ribosomal peptide synthase substrate predictor. *Bioinformatics (Oxford, England)* **28**, 426–427; 10.1093/bioinformatics/btr659 (2012).
98. Bian, X., Plaza, A., Yan, F., Zhang, Y. & Müller, R. Rational and efficient site-directed mutagenesis of adenylation domain alters relative yields of luminide derivatives in vivo. *Biotechnol. Bioeng.* **112**, 1343–1353; 10.1002/bit.25560 (2015).
99. 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).
100. Bradley S. Moore & Christian Hertweck. Biosynthesis and attachment of novel bacterial polyketide synthase starter units. *Nat. Prod. Rep.* **19**, 70–99; 10.1039/B003939J (2002).
101. Stuart Smith & Shiou-Chuan Tsai. The type I fatty acid and polyketide synthases: a tale of two megasynthases. *Nat. Prod. Rep.* **24**, 1041–1072; 10.1039/B603600G (2007).
102. White, S. W., Zheng, J., Zhang, Y. M. & Rock. The structural biology of type II fatty acid biosynthesis. *Annu Rev Biochem* **74**, 791–831; 10.1146/annurev.biochem.74.082803.133524 (2005).
103. Maier, T., Leibundgut, M. & Ban, N. The crystal structure of a mammalian fatty acid synthase. *Science (New York, N.Y.)* **321**, 1315–1322; 10.1126/science.1161269 (2008).

104. Cronan, J. E. & Thomas, J. Bacterial fatty acid synthesis and its relationships with polyketide synthetic pathways. *Complex Enzymes in Microbial Natural Product Biosynthesis, Part B: Polyketides, Aminocoumarins and Carbohydrates* **459**, 395–433; 10.1016/S0076-6879(09)04617-5 (2009).
105. Zhang, Y. M. & Rock, C. O. Thematic review series: Glycerolipids - Acyltransferases in bacterial glycerophospholipid synthesis. *J. Lipid Res.* **49**, 1867–1874; 10.1194/jlr.R800005-JLR200 (2008).
106. Warude, D., Joshi, K. & Harsulkar, A. Polyunsaturated fatty acids: Biotechnology. *Crit. Rev. Biotechnol.* **26**, 83–93 (2006).
107. Jiao, J. & Zhang, Y. Transgenic biosynthesis of polyunsaturated fatty acids: a sustainable biochemical engineering approach for making essential fatty acids in plants and animals. *Chem. Rev.* **2013**, 3799–3814; 10.1021/cr300007p (2013).
108. Yazawa, K. Production of eicosapentaenoic acid from marine bacteria. *Lipids* **31 Suppl**, S297-S300 (1996).
109. Metz, J. G. *et al.* Production of polyunsaturated fatty acids by polyketide synthases in both prokaryotes and eukaryotes. *Science* **293**, 290–293 (2001).
110. Allen, E. E. & Bartlett, D. H. Structure and regulation of the omega-3 polyunsaturated fatty acid synthase genes from the deep-sea bacterium *Photobacterium profundum* strain SS9. *Microbiology* **148**, 1903–1913 (2002).
111. Stevens, D. C., Hari, T. P. A. & Boddy, C. N. The role of transcription in heterologous expression of polyketides in bacterial hosts. *Nat. Prod. Rep.* **30**, 1391–1411; 10.1039/c3np70060g (2013).
112. Sandmann, A., Frank, B. & Müller, R. A transposon-based strategy to scale up myxothiazol production in myxobacterial cell factories. *J. Biotechnol.* **135**, 255–261; 10.1016/j.jbiotec.2008.05.001 (2008).

113. Rachid, S., Gerth, K. & Müller, R. NtcA-A negative regulator of secondary metabolite biosynthesis in *Sorangium cellulosum*. *J. Biotechnol.* **140**, 135–142; 10.1016/j.jbiotec.2008.10.010 (2008).
114. 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).
115. Wenzel, S. C. & Müller, R. Recent developments towards the heterologous expression of complex bacterial natural product biosynthetic pathways. *Curr. Opin. Biotechnol.* **16**, 594–606; 10.1016/j.copbio.2005.10.001 (2005).
116. 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).
117. Huo, L. *et al.* Heterologous expression of bacterial natural product biosynthetic pathways. *Nat. Prod. Rep.* **36**, 1412–1436; 10.1039/C8NP00091C (2019).
118. Hug, J. J. & Müller, R. Host Development for Heterologous Expression and Biosynthetic Studies of Myxobacterial Natural Products. 6.09. In *Comprehensive Natural Products III*, edited by H.-W. (B.) Liu & T. P. Begley (Elsevier, Oxford, 2020), Vol. 6, pp. 149–216.
119. Kosuri, S. & Church, G. M. Large-scale de novo DNA synthesis: technologies and applications. *Nat. Methods* **11**, 499–507; 10.1038/nmeth.2918 (2014).
120. Cobb, R. E., Ning, J. C. & Zhao, H. DNA assembly techniques for next-generation combinatorial biosynthesis of natural products. *J Ind Microbiol Biotechnol* **41**, 469–477; 10.1007/s10295-013-1358-3 (2014).
121. Gibson, D. G. *et al.* Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nature Methods* **6**, 343–345; 10.1038/nmeth.1318 (2009).
122. 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).

123. 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).
124. Zhang, Y., Buchholz, F., Muyrers, J. P. & Stewart, F. A. A new logic for DNA engineering using recombination in *Escherichia coli*. *Nat. Genet.* **20**, 123–128; 10.1038/2417 (1998).
125. Sun, H., Liu, Z., Zhao, H. & Ang, E. L. Recent advances in combinatorial biosynthesis for drug discovery. *Drug Des. Devel. Ther.* **9**, 823–833; 10.2147/DDDT.S63023 (2015).
126. McDaniel, R. *et al.* Multiple genetic modifications of the erythromycin polyketide synthase to produce a library of novel "unnatural" natural products. *Proc. Natl. Acad. Sci. USA* **96**, 1846–1851 (1999).
127. Scheid, U. Studies on the biosynthesis of myxobacterial natural products. Untersuchung der Biosynthese myxobakterieller Naturstoffe. Dissertation. Saarland University, 2021.
128. Viehrig, K. Genome Mining in the Myxobacterium *Chondromyces crocatus* Cm c5 for the discovery of novel secondary metabolites. Dissertation. Saarland University, 2015.
129. Irschik, H. & Reichenbach, H. The mechanism of action of myxovalargin A, a peptide antibiotic from *Myxococcus fulvus*. *J. Antibiot.* **38**, 1237–1245 (1985).
130. Koller, T. O. *et al.* Broad-spectrum antibiotic myxovalargin: Biosynthesis, structural revision, total synthesis and molecular characterization of ribosomal inhibition. *Nature Chem submitted manuscript* (2022).
131. Steinmetz, H., Irschik, H., Reichenbach, H. & Höfle, G. Structure elucidation of the peptide antibiotics myxovalargin A-D. In *Chemistry of Peptides and Proteins - Proceedings of the Sixth UssR-FRG Symposium on Chemistry of Peptides and Proteins (Hamburg, FRG, Sept 1-5, 1987)*, edited by W.A. König & W. Voelter (Attempo Verlag, Tübingen, 1987), pp. 13–18.
132. 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).

133. Wu, W. H. & Morris, D. R. Biosynthetic Arginine Decarboxylase from *Escherichia coli*. *J. Biol. Chem.* **248**, 1687–1695; 10.1016/S0021-9258(19)44245-2 (1973).
134. Rosenfeld, H. J. & Roberts, J. Arginine decarboxylase from a *Pseudomonas* species. *Journal of bacteriology* **125**, 601–607; 10.1128/jb.125.2.601-607.1976 (1976).
135. Ishida, K. *et al.* Biosynthesis and structure of aeruginoside 126A and 126B, cyanobacterial peptide glycosides bearing a 2-carboxy-6-hydroxyoctahydroindole moiety. *Chem. Biol.* **14**, 565–576; 10.1016/j.chembiol.2007.04.006 (2007).
136. Makris, T. M., Knoot, C. J., Wilmot, C. M. & Lipscomb, J. D. Structure of a dinuclear iron cluster-containing  $\beta$ -hydroxylase active in antibiotic biosynthesis. *Biochemistry* **52**, 6662–6671; 10.1021/bi400845b (2013).
137. ThermoFisher Scientific. DoubleDigest Calculator. Available at <https://www.thermofisher.com/de/de/home/brands/thermo-scientific/molecular-biology/thermo-scientific-restriction-modifying-enzymes/restriction-enzymes-thermo-scientific/double-digest-calculator-thermo-scientific.html>.
138. Weber, T., Baumgartner, R., Renner, C., Marahiel, M. A. & Holak, T. A. Solution structure of PCP, a prototype for the peptidyl carrier domains of modular peptide synthetases. *Structure* **8**, 407–418; 10.1016/S0969-2126(00)00120-9 (2000).
139. 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).
140. 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).
141. Bairoch, A. & Apweiler, R. The SWISS-PROT protein sequence database and its supplement TrEMBL in 2000. *Nucleic Acids Res* **28**, 45–48; 10.1093/nar/28.1.45 (2000).

142. UniProt: the universal protein knowledgebase in 2021. *Nucleic acids research* **49**, D480-D489; 10.1093/nar/gkaa1100 (2021).
143. Yan, F. *et al.* Synthetic biology approaches and combinatorial biosynthesis towards heterologous lipopeptide production. *Chemical Science* **9**, 7510–7519; 10.1039/c8sc02046a (2018).
144. Shizuya, H. *et al.* Cloning and stable maintenance of 300-kilobase-pair fragments of human DNA in *Escherichia coli* using an F-factor-based vector. *Proc. Natl. Acad. Sci. USA* **89**, 8794–8797; 10.1073/pnas.89.18.8794 (1992).
145. 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).
146. Lusetti, S. L. & Cox, M. M. The bacterial RecA protein and the recombinational DNA repair of stalled replication forks. *Annu Rev Biochem* **71**, 71–100; 10.1146/annurev.biochem.71.083101.133940 (2002).
147. Sheng, D.-H. *et al.* Functional Division Between the RecA1 and RecA2 Proteins in *Myxococcus xanthus*. *Front. Microbiol.* **11**, 140; 10.3389/fmicb.2020.00140 (2020).
148. Pogorevc, D. & Müller, R. Biotechnological production optimization of argyrins – a potent immunomodulatory natural product class. *Microb. Biotechnol.*; 10.1111/1751-7915.13959 (2021).
149. Bock, T. *et al.* The AibR-isovaleryl coenzyme A regulator and its DNA binding site - a model for the regulation of alternative de novo isovaleryl coenzyme A biosynthesis in *Myxococcus xanthus*. *Nucleic Acids Res.* **45**, 2166–2178; 10.1093/nar/gkw1238 (2017).
150. Mahmud, T. *et al.* A novel biosynthetic pathway providing precursors for fatty acid biosynthesis and secondary metabolite formation in myxobacteria. *J. Biol. Chem.* **277**, 32768–32774; 10.1074/jbc.M205222200 (2002).
151. Burchard, R. P., Burchard, A. C. & Parish, J. H. Pigmentation phenotype instability in *Myxococcus xanthus*. *Can. J. Microbiol.* **23**, 1657–1662; 10.1139/m77-238 (1977).

152. 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).
153. Hoffmann, M. *et al.* Homospermidine Lipids: A compound class specifically formed during fruiting body formation of *Myxococcus xanthus* DK1622. *ACS Chem. Biol.* **13**, 273–280; 10.1021/acscchembio.7b00816 (2018).
154. Marshall, V. P., McWethy, S. J., Sirotti, J. M. & Cialdella, J. I. The effect of neutral resins on the fermentation production of rubradirin. *J. Ind. Microbiol.* **5**, 283–287; 10.1007/BF01578202 (1990).
155. Zhang, Z. *et al.* High-level production of membrane proteins in *E. coli* BL21(DE3) by omitting the inducer IPTG. *Microb Cell Fact* **14**, 142; 10.1186/s12934-015-0328-z (2015).
156. Pogorevc, D. *et al.* Biosynthesis and Heterologous Production of Argyrins. *ACS Synth. Biol.* **8**, 1121–1133; 10.1021/acssynbio.9b00023 (2019).
157. Patteson, J. B., Dunn, Z. D. & Li, B. In vitro Biosynthesis of the Nonproteinogenic Amino Acid Methoxyvinylglycine. *Angew. Chem. Int. Ed.*; 10.1002/anie.201713419 (2018).
158. Wang, S. *et al.* Discovery and Biosynthetic Investigation of a New Antibacterial Dehydrated Non-Ribosomal Tripeptide. *Angew. Chem. Int. Ed.* **60**, 3229–3237; 10.1002/anie.202012902 (2021).
159. 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).
160. Gish, W. & States, D. J. Identification of protein coding regions by database similarity search. *Nat Genet* **3**, 266–272; 10.1038/ng0393-266 (1993).
161. Waser, P. & Altmann, K.-H. Die Totalsynthese des Antibiotikums Disciformycin B durch Ringschlussmetathese. *Angew. Chem.* **132**, 17546–17550; 10.1002/ange.202004589 (2020).

162. Kwon, Y., Schulthoff, S., Dao, Q. M., Wirtz, C. & Fürstner, A. Total Synthesis of Disciformycin A and B: Unusually Exigent Targets of Biological Significance. *Chemistry (Weinheim an der Bergstrasse, Germany)* **24**, 109–114; 10.1002/chem.201705550 (2018).
163. Sherman, D. H. *et al.* The structural basis for substrate anchoring, active site selectivity, and product formation by P450 PikC from *Streptomyces venezuelae*. *The Journal of biological chemistry* **281**; 10.1074/jbc.M605478200 (2006).
164. Wolucka, B. A. Biosynthesis of D-arabinose in mycobacteria - a novel bacterial pathway with implications for antimycobacterial therapy. *FEBS J* **275**, 2691–2711; 10.1111/j.1742-4658.2008.06395.x (2008).
165. Breitmaier, E. & Jung, G. *Organische Chemie. Grundlagen, Stoffklassen, Reaktionen, Konzepte, Molekülstruktur ; zahlreiche Formeln, Tabellen*. 5th ed. (Thieme, Stuttgart, 2005).
166. Jungmann, K. Investigation of bacterial secondary metabolite pathways from *Sorangium cellulosum*. Dissertation. Saarland University, 2015.
167. Feickert, M. Synthetische Biotechnologie zur Heterologen Produktion von Maracen. Diploma Thesis (Pharmacy). Saarland University, 2017.
168. Magrini, V., Creighton, C. & Youderian, P. Site-specific recombination of temperate *Myxococcus xanthus* phage Mx8: genetic elements required for integration. *Journal of bacteriology* **181**, 4050–4061; 10.1128/jb.181.13.4050-4061.1999 (1999).
169. Gabler, F. *et al.* Protein Sequence Analysis Using the MPI Bioinformatics Toolkit. *Current Protocols in Bioinformatics* **72**, e108; 10.1002/cpbi.108 (2020).
170. 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).
171. Gemperlein, K. Biosynthesis and heterologous production of polyunsaturated fatty acids from mycobacteria. Dissertation. Saarland University, 2014.

172. Pogorevc, D. Establishing and engineering heterologous production systems for argyrimycin and  $\alpha$ -pyrone antibiotics. Doctoral Thesis. Saarland University, Jan 2019.
173. Slabinski, L. *et al.* XtalPred: a web server for prediction of protein crystallizability. *Bioinformatics* **23**, 3403–3405; 10.1093/bioinformatics/btm477 (2007).
174. Mistry, J. *et al.* Pfam: The protein families database in 2021. *Nucleic acids research* **49**, D412–D419; 10.1093/nar/gkaa913 (2021).
175. Herrmann, J., Fayad, A. A. & Müller, R. Natural products from myxobacteria: novel metabolites and bioactivities. *Natural product reports*; 10.1039/c6np00106h (2016).
176. Inaoka, T., Takahashi, K., Ohnishi-Kameyama, M., Yoshida, M. & Ochi, K. Guanine nucleotides guanosine 5'-diphosphate 3'-diphosphate and GTP co-operatively regulate the production of an antibiotic bacilysin in *Bacillus subtilis*. *Journal of Biological Chemistry* **278**, 2169–2176; 10.1074/jbc.M208722200 (2003).
177. Shima, J., Hesketh, A., Okamoto, S., Kawamoto, S. & Ochi, K. Induction of actinorhodin production by rpsL (encoding ribosomal protein S12) mutations that confer streptomycin resistance in *Streptomyces lividans* and *Streptomyces coelicolor* A3(2). *J. Bacteriol.* **178**, 7276–7284 (1996).
178. Buntin, K. *et al.* Biosynthesis of thuggacins in myxobacteria: comparative cluster analysis reveals basis for natural product structural diversity. *Chem. Biol.* **17**, 342–356; 10.1016/j.chembiol.2010.02.013 (2010).
179. 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).
180. Bergmann, S. *et al.* Activation of a silent fungal polyketide biosynthesis pathway through regulatory cross talk with a cryptic nonribosomal peptide synthetase gene cluster. *Appl. Environ. Microbiol.* **76**, 8143–8149 (2011).

181. Mao, D., Bushin, L. B., Moon, K., Wu, Y. & Seyedsayamdost. Discovery of scmR as a global regulator of secondary metabolism and virulence in *Burkholderia thailandensis* E264. *Proc. Natl. Acad. Sci. USA* **114**, E2920-E2928; 10.1073/pnas.1619529114 (2017).
182. Bolten, C. J., Heinzle, E., Müller, R. & Wittmann, C. Investigation of the central carbon metabolism of *Sorangium cellulosum*: metabolic network reconstruction and quantification of pathway fluxes. *J. Microbiol. Biotechnol.* **19**, 23–36; 10.4014/jmb.0803.213 (2009).
183. 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).
184. 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).
185. Siegl, T., Tokovenko, B., Myronovskyi, M. & Luzhetskyy, A. Design, construction and characterisation of a synthetic promoter library for fine-tuned gene expression in actinomycetes. *Metab. Eng.* **19**, 98–106; 10.1016/j.ymben.2013.07.006 (2013).
186. Handelsman, J., Rondon, M. R., Brady, S. F., Clardy, J. & Goodman, R. M. Molecular biological access to the chemistry of unknown soil microbes. A new frontier for natural products. *Chemistry & biology* **5**, R245-R249; 10.1016/S1074-5521(98)90108-9 (1998).
187. Little, R. F. & Hertweck, C. Chain release mechanisms in polyketide and non-ribosomal peptide biosynthesis. *Nat. Prod. Rep.*; 10.1039/d1np00035g (2021).
188. Hai, Y. & Tang, Y. Biosynthesis of Long-Chain N-Acyl Amide by a Truncated Polyketide Synthase-Nonribosomal Peptide Synthetase Hybrid Megasyntase in Fungi. *J. Am. Chem. Soc.* **140**, 1271–1274; 10.1021/jacs.7b13350 (2018).
189. Wenzel, S. C. & Müller, R. Host Organisms: Myxobacterium. In *Industrial biotechnology, Microorganisms Volume 3a and 3b*, edited by C. Wittmann & J. Liao (Wiley-VCH, Weinheim, Germany, 2017), pp. 453–485.

190. Loeschcke, A. & Thies, S. *Pseudomonas putida*-a versatile host for the production of natural products. *Appl. Microbiol. Biotechnol.* **99**, 6197–6214; 10.1007/s00253-015-6745-4 (2015).
191. 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).
192. 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).
193. Gomez-Escribano, J. P. & Bibb, M. J. Engineering *Streptomyces coelicolor* for heterologous expression of secondary metabolite gene clusters. *Microb. Biotechnol.* **4**, 207–215; 10.1111/j.1751-7915.2010.00219.x (2011).
194. Komatsu, M., Uchiyama, T., Omura, S., Cane, D. E. & Ikeda, H. Genome-minimized *Streptomyces* host for the heterologous expression of secondary metabolism. *Proc. Natl. Acad. Sci. USA* **107**, 2646–2651 (2010).
195. Penn, J. *et al.* Heterologous production of daptomycin in *Streptomyces lividans*. *J Ind Microbiol Biotechnol* **33**, 121–128; 10.1007/s10295-005-0033-8 (2006).
196. Zhang, H., Fang, L., Osburne, M. S. & Pfeifer, B. A. The Continuing Development of *E. coli* as a Heterologous Host for Complex Natural Product Biosynthesis. In *Nonribosomal Peptide and Polyketide Biosynthesis. The continuing development of E. coli as a heterologous host for complex natural product biosynthesis*, edited by B. S. Evans (Springer New York, New York, NY, 2016), Vol. 1401, pp. 121–134.
197. Kumpfmüller, J. *et al.* Production of the polyketide 6-deoxyerythronolide B in the heterologous host *Bacillus subtilis*. *Applied Microbiology and Biotechnology* **100**, 1209–1220; 10.1007/s00253-015-6990-6 (2016).
198. Larionov, V. *et al.* Specific cloning of human DNA as yeast artificial chromosomes by transformation-associated recombination. *Proc. Natl. Acad. Sci. USA* **93**, 491–496 (1996).

199. 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).
200. Kouprina, N. & Larionov, V. TAR cloning: insights into gene function, long-range haplotypes and genome structure and evolution. *Nat Rev Genet* **7**, 805–812; 10.1038/nrg1943 (2006).
201. 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).
202. 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).
203. Wang, H. *et al.* ExoCET. Exonuclease in vitro assembly combined with RecET recombination for highly efficient direct DNA cloning from complex genomes. *Nucleic Acids Res.* **46**, e28; 10.1093/nar/gkx1249 (2018).
204. Li, L. *et al.* A stepwise increase in pristinamycin II biosynthesis by *Streptomyces pristinaespiralis* through combinatorial metabolic engineering. *Metab. Eng.* **29**, 12–25; 10.1016/j.ymben.2015.02.001 (2015).
205. Kling, A. *et al.* Antibiotics. Targeting DnaN for tuberculosis therapy using novel griselimycins. *Science* **348**, 1106–1112; 10.1126/science.aaa4690 (2015).
206. Gorges, J. *et al.* Structure, Total Synthesis, and Biosynthesis of Chloromyxamides: Myxobacterial Tetrapeptides Featuring an Uncommon 6-Chloromethyl-5-methoxypipelic Acid Building Block. *Angew. Chem. Int. Ed. Engl.* **57**, 14270–14275; 10.1002/anie.201808028 (2018).
207. Shimamura, H. *et al.* Structure determination and total synthesis of bottromycin A2: a potent antibiotic against MRSA and VRE. *Angew. Chem. Int. Ed. Engl.* **48**, 914–917; 10.1002/anie.200804138 (2009).

208. Chu, J. *et al.* Discovery of MRSA active antibiotics using primary sequence from the human microbiome. *Nature chemical biology* **12**, 1004–1006; 10.1038/nchembio.2207 (2016).
209. Groß, S. *et al.* Improved broad-spectrum antibiotics against Gram-negative pathogens via darobactin biosynthetic pathway engineering. *Chem. Sci.* **12**, 11882–11893; 10.1039/d1sc02725e (2021).
210. Siebke, M. Studien zur Totalsynthese von Myxovalargin und Darstellung einer Myxovalargin-Bibliothek, 2021.
211. Waser, P. & Altmann, K.-H. An RCM-Based Total Synthesis of the Antibiotic Disciformycin B. *Angew. Chem. Int. Ed.* **59**, 17393–17397; 10.1002/anie.202004589 (2020).
212. Baloglu, E. & Kingston, D. G. A new semisynthesis of paclitaxel from baccatin III. *J. Nat. Prod.* **62**, 1068–1071; 10.1021/np990040k (1999).
213. Thaker, M., Spanogiannopoulos, P. & Wright, G. D. The tetracycline resistome. *Cellular and molecular life sciences : CMLS* **67**, 419–431; 10.1007/s00018-009-0172-6 (2010).
214. Kasbekar, N. Tigecycline: a new glycylicycline antimicrobial agent. *Am J Health Syst Pharm* **63**, 1235–1243; 10.2146/ajhp050487 (2006).