Technische Universität München

Fakultät für Chemie

Fachgebiet für Biosystemchemie

The (Bio-)Synthesis of Polycyclic Tetramate Macrolactams (PoTeMs)

Christian Greunke, M.Sc.

Vollständiger Abdruck der von der Fakultät für Chemie der Technischen Universität München zur Erlangung des akademischen Grades eines Doktors der Naturwissenschaften (Dr. rer. nat.) genehmigten Dissertation.

Vorsitzende: Prof. Dr. Corinna Hess

Prüfer der Dissertation:

        1. Prof. Dr. Tobias A.M. Gulder

        2. apl. Prof. Dr. Wolfgang Eisenreich

        3. Prof. Dr. Frank Hahn

Die Dissertation wurde am 13.11.2017 bei der Technischen Universität München eingereicht und durch die Fakultät für Chemie am 11.12.2017 angenommen.

Parts of this thesis have been published in:

- **Greunke, C.**; Glöckle, A.; Antosch, J.; Gulder, T.A.M. Biocatalytic Total Synthesis of Ikarugamycin. *Angew. Chem. Int. Ed. Engl.* **2017**, *56*, 4351-4355, `DOI: 10.1002/anie.201611063`.

- **Greunke, C.***; Antosch, J.*; Gulder, T.A.M. Promiscuous hydroxylases for the functionalization of polycyclic tetramate macrolactams – conversion of ikarugamycin to butremycin. *Chem. Comm.* **2015**, *51*, 5334-5336, `DOI: 10.1039/c5cc00843c`.

  *equally contributing authors

This thesis contains content which will be part of a publication:

- **Greunke, C.**; Duell, E.R.; D'Agostino, P.M.; Glöckle, A.; Lamm, K.; Gulder, T.A.M. Direct Pathway Cloning (DiPaC) to Access Natural Product Biosynthetic Gene Clusters. *Manuscript in preparation.*

Friedrich Naumann STIFTUNG FÜR DIE FREIHEIT

Contents

Contents

Abstract

Polycyclic tetramate macrolactams (PoTeMs) are an intriguing novel class of bacterial secondary metabolites with complex molecular frameworks. Their chemical synthesis is highly challenging, making innovative approaches to access these substances desirable.

This work summarizes novel approaches for the production and modification of the PoTeM ikarugamycin (**27**). By employing the biosynthetic enzymes from the *ika* gene clusters of *Streptomyces* sp. Tü6239 and *Salinispora arenicola* CNS-205, it was possible to conduct a chemoenzymatic synthesis of **27**. This is the first biocatalytic synthesis of a PoTeM, in an one-pot reaction at room temperature, overnight, thereby surpassing yields of classical chemical total synthesis efforts.

By applying PoTeM hydroxylases from either *Lysobacter capsici* DSM 19286 or *Saccharophagus degradans* DSM 17024, the selective hydroxylation of **27** at position C-3 was achieved, resulting in the production of butremycin (**28**) *in vitro* and *in vivo*. This work highlights for the first time that enzymes from different PoTeM pathways can be utilized to generate novel frameworks.

The PKS/NRPS enzyme (Tü6239-IkaA) from the biosynthetic pathway of **27** should furthermore be reprogrammed by site-directed mutagenesis to generate novel PoTeM analogues possessing a L-lysine backbone instead of L-ornithine (**39**). Comparative bioinformatic analysis of adenylation domains specific for the incorporation of **39** and L-lysine (**42**) led to the identification of five relevant amino acid positions within Tü6239-IkaA, potentially altering the substrate selection. By site-directed mutagenesis methods, the A domain sequence was changed according to the *in silico* predictions. A mutant enzyme was generated that, however, still incorporated **39**, but with low catalytic activity.

In addition to the biocatalytic reactions, first trials aiming at the structural analysis of the PoTeM biosynthetic enzymes were started. First crystallization trials were conducted for the alcohol dehydrogenase IkaC from *S. arenicola* CNS-205. Protein crystals were obtained and used for diffraction measurements. So far, the resolution was too low to construct a structural model of this PoTeM enzyme. New sets of improved crystals are available for further measurements to obtain a first structural model of SA-IkaC in the future.

Abstract

Furthermore, a PCR-based methodology to clone biosynthetic gene clusters was developed and termed 'Direct Pathway Cloning' (DiPaC). It has proven feasible to extract gene clusters with sizes from 12 to 54.6 kb and varying GC content from diverse microbial sources, allowing to promote subsequent heterologous expression. By employing DiPaC, seven PoTeM clusters in total were isolated from the respective genomes, with three of them unrelated to any known compound. DiPaC was used to clone the entire erythromycin gene cluster from *Saccharopolyspora erythraea* DSM 40517. The *ery* cluster (54.6 kb) was transferred into the genomes of *Streptomyces coelicolor* M1152 and M1154 and production of erythromycin A (**8**) was proven by liquid-chromatography in combination with mass spectrometry. This is the first heterologous setup to produce **8** in *Streptomyces* to date. The DiPaC method is beneficial to improve the field of natural product chemistry in terms of discovering und studying novel classes of compounds more rapidly.

Zusammenfassung

Polyzyklische Tetramsäure-haltige Makrolactame (PoTeM) stellen eine neuartige Klasse mikrobieller Sekundärmetabolite dar, die sich durch eine beeindruckende strukturelle Komplexität auszeichnet. Die chemische Totalsynthese dieser Verbindungen ist anspruchsvoll und wenig ergiebig, sodass innovative Ansätze wünschenswert wären, die es erlauben, diese Substanzklasse nachhaltig zu erschließen.

Die vorliegende Arbeit befasste sich mit der enzymatischen Synthese und Modifikation von Ikarugamycin (**27**). Durch den Einsatz der Biosyntheseenzyme aus den *ika*-Genclustern von *Streptomyces* sp. Tü6239 sowie *Salinispora arenicola* CNS-205 war es möglich, eine chemoenzymatische Totalsynthese für **27** zu etablieren. Der entwickelte biokatalytische Ansatz besteht aus einer Eintopfreaktion, die bei Raumtemperatur über Nacht durchgeführt werden kann und dabei Ausbeuten liefert, welche die der organisch-chemischen Totalsynthesen übertrifft.

Durch Verwendung von PoTeM-Hydroxylasen aus *Lysobacter capsici* DSM 19286 sowie *Saccharophagus degradans* DSM 17024 gelang es zudem, **27** selektiv am C-3-Kohlenstoffatom durch Einführung einer Alkoholgruppe zu modifizieren. Dadurch wurde Butremycin (**28**) in *in vitro*- und *in vivo*-Ansätzen hergestellt. Diese Arbeiten zeigten erstmals, dass Enzyme unterschiedlichster PoTeM-Biosynthesewege kombiniert werden können und so neue Strukturen generierbar sind.

Ferner sollte das PKS/NRPS-Enzym (Tü6239-IkaA) aus der Biosynthese von **27** durch gezielte Mutagenese so verändert werden, dass PoTeM-Analoga mit L-Lysin (**42**) im Rückgrat anstelle von L-Ornithin (**39**) zugänglich werden. Durch vergleichende, bioinformatische Analysen von Adenylierungsdomänen wurden fünf relevante Aminosäurepositionen innerhalb von Tü6239-IkaA identifiziert, durch deren Austausch ein Substratwechsel erreicht werden sollte. Mittels gezielter Mutagenese gelang es die selektierten Reste entsprechend zu verändern. Allerdings wurde so ein Enzym entwickelt, welches nur noch geringe Aktivität aufwies und weiterhin **39** als Baustein verwendete.

Neben den biokatalytischen Arbeiten sollten auch erste Grundlagen für strukturbiologische Untersuchungen gelegt werden. Erste Kristallisationsversuche wurden für die Alkoholdehydrogenase IkaC aus *S. arenicola* CNS-205 unternommen. Hierbei wurden

Zusammenfassung

Proteinkristalle erzeugt und Beugungsmessungen durchgeführt. Bis jetzt war jedoch die Auflösung, die die Kristalle lieferten, zu gering, um eine Strukturaufklärung damit durchführen zu können. Mit verbesserten Kristallen können neue Messungen durchgeführt werden, um so in der Zukunft ein erstes Strukturmodell von SA-IkaC erstellen zu können.

Weiterhin wurde eine PCR-basierte Methode zur Klonierung von Biosynthesegenclustern entwickelt, das sogenannte „Direct Pathway Cloning" (DiPaC). Damit gelang es, Gencluster von 12 bis 54.6 kb Größe aus unterschiedlichsten mikrobiellen Quellen unabhängig vom GC-Gehalt zu klonieren, um im Anschluss heterologe Expressionsversuche durchzuführen. Durch Verwendung von DiPaC war es innerhalb dieser Arbeit möglich, sieben PoTeM-Gencluster zu isolieren, von denen drei bisher noch keiner chemischen Verbindung zugeordnet werden konnten. Mit DiPaC war es außerdem möglich, das gesamte Erythromycin-Gencluster von *Saccharopolyspora erythraea* DSM 40517 zu klonieren. Das *ery*-Cluster (54.6 kb) wurde in die Streptomyzetenstämme *S. coelicolor* M1152 und M1154 eingebracht und die Produktion von Erythromycin A (**8**) mittels LC-MS nachgewiesen. Dies ist die erste erfolgreiche heterologe Produktion dieses bekannten Sekundärmetaboliten in Streptomyzeten. Dieser Ansatz ist äußerst nützlich und von besonderer Bedeutung für das Forschungsgebiet der Naturstoffbiochemie, um neuartige Naturstoffe (und Stoffklassen) noch schneller zu entdecken und eingehend untersuchen zu können.

1. Introduction

1.1. Secondary metabolites as innovative drug leads

Secondary metabolites play an important role in today's medicine as they cover a broad range of therapeutic applications with often innovative modes of action.[1] Their structures are diverse and complex, making them an ideal source to develop treatment options for a vast range of diseases in the clinic.[2] During the course of evolution, these molecules have developed strong potency to tackle specific targets. For mankind, they significantly contribute to improve life quality, as different examples impressively show (Figure 1): One famous class of fungal secondary metabolites are the penicillins, discovered by Alexander Fleming in 1928, which allowed the treatment of wound infections that have been life-threatening before.[3] These β-lactam compounds, penicillin G (**1**) as one representative,[4] are still among the most important antibiotic drugs. Tetracyclines represent another prominent group of often prescribed natural antibiotics.[5] The glycopeptide vancomycin (**2**) is well known as it is an important antibiotic of last resort, used to fight multi-resistant human-pathogenic bacteria that no other conventional antibiotic is able to tackle.[6] It has become more and more important in the last years, as resistances to conventional antibiotics are continuously evolving. In western industrialized countries, metabolic diseases are increasing, caused by high-caloric, fat-rich diets. For the treatment of atherosclerosis, drugs like lovastatin (**3**) are prescribed, which are able to control endogenous cholesterol production.[7,8] With this drug and analogues, the pharmaceutical industry earned billions of dollars making statin blockbusters.[9] Paclitaxel (**4**), first isolated from the bark of the pacific yew *Taxus brevifolia*, is a potent anti-cancer drug used in cancer chemotherapy.[10] The significance of such secondary metabolites in modern medicine was recently highlighted in 2015: The discovery of avermectins, such as avermectin B_{1a} (**5**), a series of potent anthelmintic and insecticidal compounds by William C. Campbell and Satoshi Ōmura was awarded the Nobel Prize.[11] These drugs have protected millions of people from river blindness and lymphatic filariasis.[11]
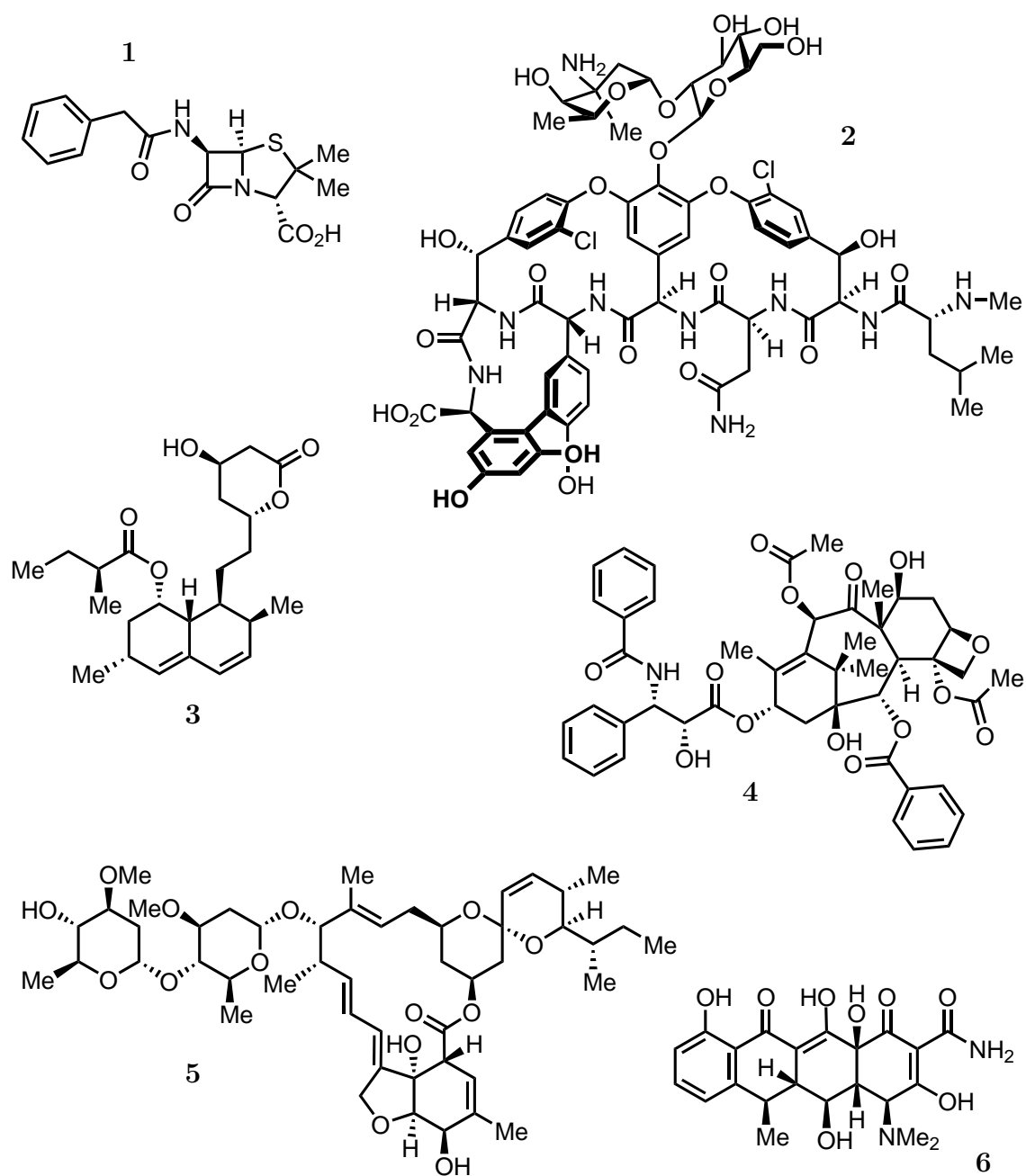
Figure 1. Selection of clinically-important pharmaceuticals derived from secondary metabolites: penicillin G (**1**), the last resort antibiotic vancomycin (**2**), the cholesterol-reducing drug lovastatin (**3**), the cancer chemotherapeutic agent paclitaxel (**4**), the anthelmintic avermectin B_{1a} (**5**) whose discovery was awarded the Nobel Prize in 2015, and the tetracycline antibiotic doxycycline (**6**).

Secondary metabolites are often termed natural products as they are isolated from natural sources, such as plants, fungi and bacteria, which are the producers of these compounds. For example, penicillin G (**1**) is derived from the mold *Penicillium chrysogenum*.[3] Most of the antibiotics known to date are produced by Actinobacteria: Tetracyclines were isolated from *Streptomyces aureofaciens*[5] and vancomycin (**2**) from *Amycolatopsis orientalis*.[6] Although secondary metabolites are not essential for the survival of their producers, it is believed that they provide advantages over opponents in the same living environment.

The complex and intriguing structures found in nature are also a source of inspiration for chemists. Trying to resynthesize natural substances is a laborious and tedious challenge, but allows to gain deep knowledge on nature's creativity. It further opens the door to alter structures and to make more potent analogues. In the case of tetracycline antibiotics, synthetic analogues such as doxycycline (**6**) were generated, having more effective activities than their precursors, e.g., a prolonged half-life in the body.[12,13]

Natural sources are inexhaustibly diverse and the same seems to be true for bacterial secondary metabolites, leading to a virtually unlimited reservoir for novel innovative drug leads.[14] The discovery of unexplored compounds is a key challenge of current research efforts. An unimaginable number of yet unknown molecules possessing exciting and beautiful scaffolds paired with potent pharmacological activities are still awaiting discovery! Likewise, the underlying biosynthetic logic is going to be explored, providing us with lessons from nature, e.g., for the development of new biocatalysts.[15]

1.2. The molecular dogma of natural product research

All the information needed to produce a secondary metabolite is stored on specific locations of the genomic DNA (gDNA) of the natural producer. During evolution, genetic elements known as biosynthetic gene clusters (BGCs) have developed, in particular in bacteria. By sequencing microbial genomes, these gene clusters can be identified and correlated to chemical structures.[16] Based on these findings, a general principle, the molecular dogma of natural product biochemistry, can be proposed (Scheme 1). The gDNA of the natural producer harbors a set of (clustered) genes. If activated, the production of messenger RNA (mRNA) is initiated, which is needed to start the formation of the encoded protein(s) by the ribosome. The corresponding cellular enzymatic machinery provided by this process catalyzes the synthesis of the secondary metabolite(s).

In the last decade, DNA sequencing has promoted the research field of natural product (bio-)chemistry. New powerful sequencing techniques have emerged, which are summarized
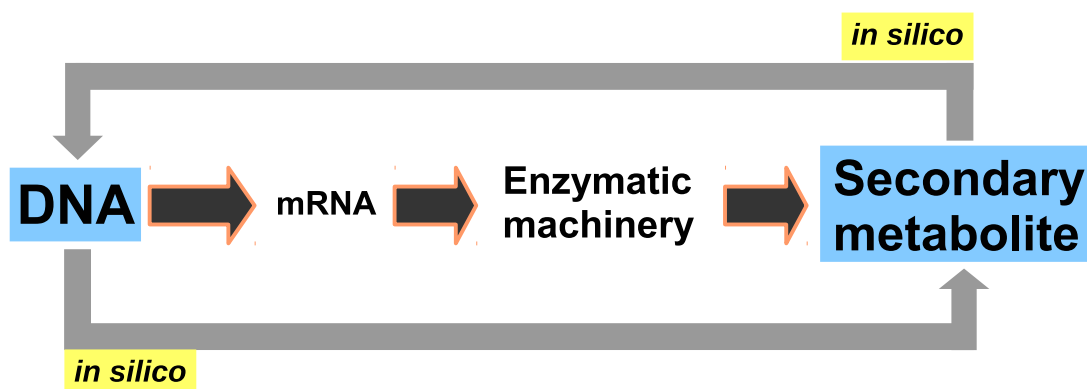
Scheme 1. The dogma of molecular natural product biochemistry states that all the information to synthesize a secondary metabolite is stored on the DNA of the natural producer. By the use of computer programs, *in silico* predictions can be made in both directions.

by the generic term next-generation sequencing (NGS).[17] Since the first genome of the bacterium *Haemophilus influenzae* was published in 1995,[18] many more bacteria from diverse phyla have been sequenced. They often contain multiple gene clusters, of which many cannot be related to a known compound, the so-called 'orphan gene clusters'.[16] Nowadays, it is possible to sequence and assemble whole bacterial genomes within days, thereby leading to a massive increase of genomic data of either partially sequenced or closed genomes.[19,20] Data bases like NCBI's gene bank can be accessed on the internet and sequences downloaded for free, facilitating broad access to the research community. This is a great treasure chest for the discovery of novel chemistry.

By the use of computer programs like antiSMASH,[21] PRISM[22] or SMURF,[23] it is possible to search for secondary metabolite gene clusters *in silico*. This concept is summarized by the term 'genome mining'. The key challenge is to understand the genetic features and to produce their encoded products either in the natural or a heterologous host. Based on predictions, proposals of biosynthetic pathways can be created. Using the nucleotide sequences as a start, open reading frames are identified and genes annotated, in this way allowing the prediction of biosynthetic pathways and thus of the chemical structures of the encoded secondary metabolites. Vice versa, knowing the structure of a natural compound, e.g., cyclic peptides,[24] allows in many cases to come up with an idea of how the biosynthetic machinery should be composed, which enzymes are needed, and how the reaction sequence will be arranged (Scheme 1).

The big challenge is to get functional proof from wet laboratory work. Although *in silico* predictions have evolved and are the first approach to get an idea of the genetic potential of the organism of interest, the real world situation might differ drastically.

1.3. Polyketide biosynthesis

Polyketide synthases (PKSs) produce a broad range of biomedically important secondary metabolites. Well studied examples of this class, presented in the following, are the statin drug lovastatin (**3**), doxorubicin (**7**) and the antibiotic erythromycin A (**8**).[25] The biosynthesis of these compounds involves repetitive, Claisen-like condensation reactions, in which acetyl-CoA (**9**), malonyl-CoA (**10**) or simple acyl-CoA derivatives (**11**–**16**) are utilized to form complex frameworks (Figure 2).[26–28] The process of polyketide formation shares high homology to the biosynthesis of fatty acids.[29] The difference is that the oxidation states of the resulting carbon chain are variable: In most of the cases, lipids are fully reduced, whereas in polyketides, fully-reduced, partially and unreduced states are possible. PKS enzymes can be subdivided into a domain structure that involves a set of essential domains, i.e., the acyltransferase (AT), ketosynthase (KS) and the acyl carrier protein (ACP). Additional, optional domains are the ketoreductase (KR), dehydratase (DH), enoyl reductase (ER) and the thioesterase (TE).
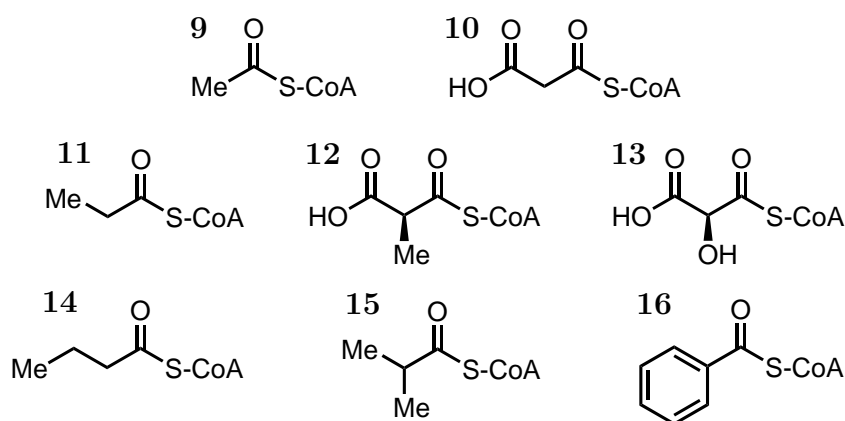
Figure 2. Selection of CoA substrates used as building blocks for PKS biochemistry. Acetyl- (**9**) and malonyl-CoA (**10**) are most frequently used. Further (starter) units are propionyl- (**11**), methylmalonyl- (**12**), hydroxymalonyl- (**13**), butyryl- (**14**), isobutyryl- (**15**) and benzoyl-CoA (**16**).

PKS enzymes, as fatty acid synthases, need to undergo enzymatic activation. After their synthesis by the ribosome, a post-translational modification is required that converts the PKS from the unmodified apo into the catalytically-active holo form.[30] A special class of enzymes, the phosphopantetheinyl transferases (PPTases), e.g., Sfp,[31,32] are able to attach a phosphopantetheine (PPant) residue derived from coenzyme A (CoA) (**17**) onto conserved serine residues of the acyl carrier protein (Scheme 2). This allows the covalent

attachment of CoA-activated substrates as well as the processing of the resulting thioester bound substrates and intermediates during chain elongation.
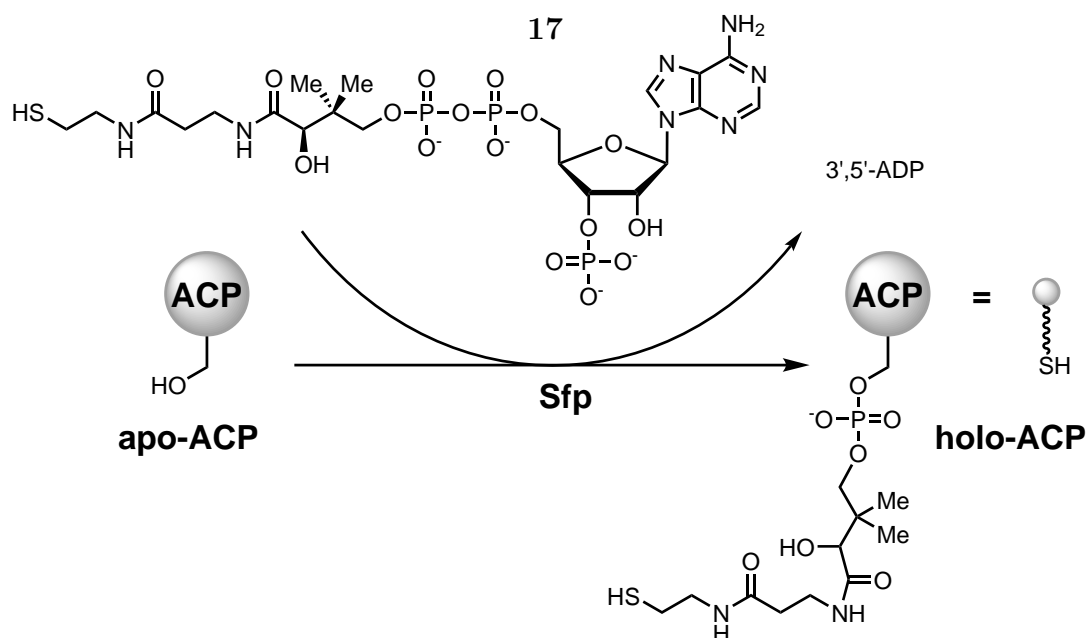
**Scheme 2.** Phosphopantetheinylation is a post-translational modification needed for the activation of PKS enzymes. The PPTase, e.g., Sfp,[31,32] uses a building block of CoA (**17**) to attach the prosthetic group on conserved serine residues, thereby converting the enzyme from the apo form into the catalytically-active holo form.

PKS enzymes are classified according to their domain organization and mode of action: type I, II and III PKS (Scheme 3).[25,29] The formation of the intermediate polyketide chain is either catalyzed by individual modules or in an iterative manner, involving only a single module.[25] Modular type I systems typically use each catalytic domain once, in this way elongating the intermediate chain by addition of one building block (see erythromycin biosynthesis, Scheme 7). The process is comparable to an assembly line and dependent on the sequence of the processing domains. Thus, a prediction on the product structure is often easily possible. This finding is known as the 'collinearity rule'.[33,34]

Type II PKS enzymes work iteratively and have exclusively been found in bacteria. These synthases consist of two detached ketosynthases (KS$_\alpha$, KS$_\beta$) and an ACP (Scheme 4). KS$_\alpha$ is responsible for the synthesis of the polyketide chain by decarboxylative elongation of C-2 building blocks derived from malonyl-CoA (**10**). KS$_\beta$ is not directly involved in the processing of the intermediate polyketide, but controls the length of the produced chain and is therefore also known as the 'chain length factor'.[25] The biosynthesis by such a minimal PKS can be assisted by further modifying enzymes. These include cyclases
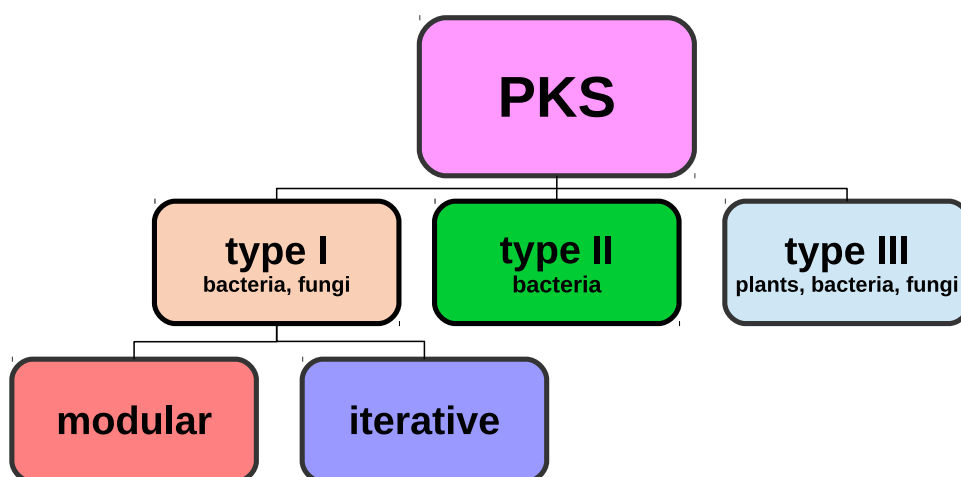
Scheme 3. Classification of polyketide synthases (PKS).

(CYC), ketoreductases (KR) and aromatases (ARO).[29] The biosynthesis of doxorubicin
(**7**) underlies PKS type II principles (Scheme 4).[25]

Type III PKS enzymes are special as they do not require an ACP and can perform the
chain elongation directly on CoA building blocks. They are also known as chalcone/stilbene
synthase-like PKS and work iteratively.[29,35] Similar to type II PKS, they convert malonyl-
CoA (**10**) as elongation substrate, but are flexible concerning CoA activated starter units,
e.g., cinnamic acid.[25] The chain length, as well as the individual cyclization chemistry, lead
to the generation of a huge variety of aromatic compounds. Type III PKS have mainly
been found in plants. One prominent example are the flavonoids, e.g., the plant metabolite
naringenin (**18**) (Scheme 5). Recently, some representatives from bacteria[36] and fungi[37]
have been discovered.
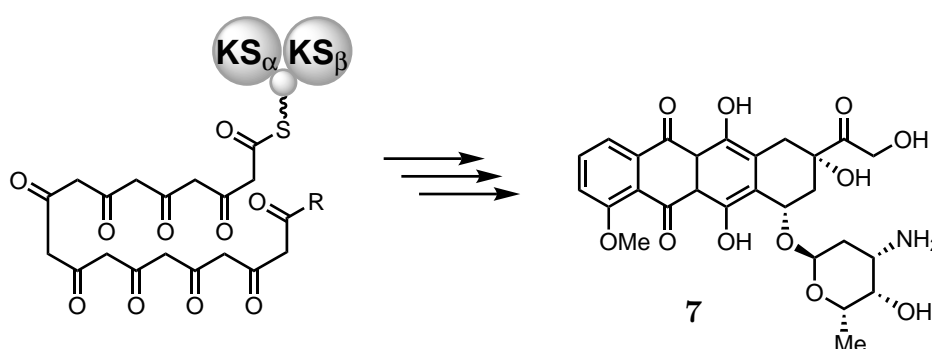
Scheme 4. Doxorubicin (**7**), used in cancer chemotherapy, is produced by a bacterial type II
PKS consisting of two ketosynthase subunits ($KS_{\alpha,\beta}$).[25]

7
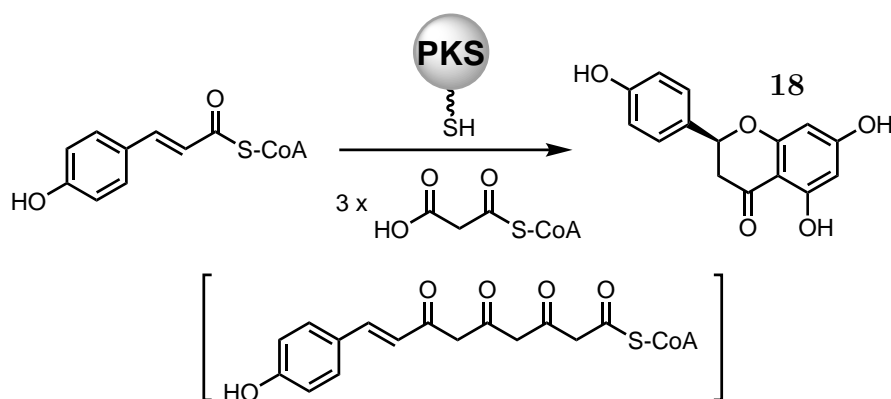
Scheme 5. The plant metabolite naringenin (**18**) is one example of a type III PKS-derived natural product. **18** is formed from hydroxy cinnamic acid and malonyl-CoA (**10**).[25]

Iterative type I PKS (*i*PKS) were identified in fungi first. One prominent example is the biosynthesis of lovastatin (**3**) (Scheme 6).[38,39] The *i*PKS enzymes in general consist of one module which is involved in several repetitive reactions. The domain functions are identical compared to the multimodular domain enzymes (Scheme 7). In the case of **3**, a nonaketide is produced from acetyl-CoA (**9**) and eight equivalents of malonyl-CoA (**10**) by the polyketide synthase LovB. A free-standing enonyl reductase (ER) substitutes for an inactive domain, which is located in between KS-, AT-, DH-, MT-, and the KR- and ACP domains. *S*-Adenosyl methionine (SAM) is the donor of a methyl group, which is introduced by the methyltransferase domain (MT). Another special property is that the polyketide synthase LovB does not employ all of the available modules during each elongation cycle: In the first steps of elongation, partial reductions take place, so that *trans* configured C–C double bonds are formed. After the incorporation of fully reduced C-2 building blocks, another unsaturated position is generated that is in close proximity to the diene system from the initial elongation steps, facilitating a Diels-Alder reaction, which results in the formation of the naphthalene backbone of **3**. Post-PKS-modifications include one oxidation step to introduce the second double bond in **3** and a derivatization by attachment of α-*S*-methylbutyrate. This special building block is produced by a second PKS enzyme, LovF, from two units of malonyl-CoA (**10**) and is introduced by the transesterase LovD at the hydroxy function of the naphthalene moiety.[40]
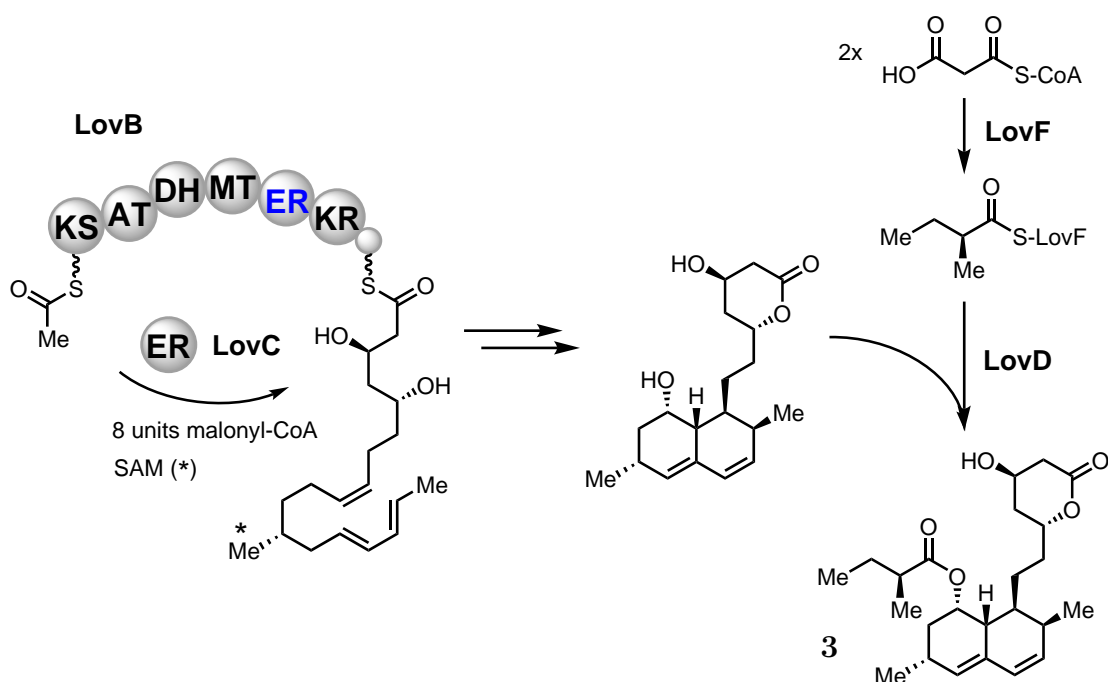
Scheme 6. The cholesterol lowering drug lovastatin (**3**) is produced by a fungal iterative type I PKS.[38,39] Acetyl-CoA (**9**) and eight equivalents of malonyl-CoA (**10**) are condensed by LovB, which possesses an inactive enoyl reductase domain (ER) (blue), working together with free-standing LovC. *S*-adenosyl methionine (SAM) delivers a methyl group (*). The α-*S*-methylbutyrate building block is produced by a second PKS enzyme LovF and attached by the transesterase LovD.[40]

The prototypical example of a modular type I polyketide assembly line is that of erythromycin A (**8**) biosynthesis (Scheme 7).[29,41] **8** is a macrolide antibiotic produced by the soil-dwelling bacterium *Saccharopolyspora erythraea* (formerly *Streptomyces erythreus*)[42] and is of clinical importance as it is used as a broad spectrum antibiotic.[43] Its biosynthesis was intensively studied and the structure of the compound related to an intriguing enzymatic machinery. From the sequenced genome of *S. erythraea*, a gene cluster with a size of about 54.6 kb was identified, comprised of 22 genes.[29,44] All these provide the complete enzymatic machinery to form **8**.[45–47] Three PKS genes (*eryAI*, *eryAII* and *eryAIII*) play a central role, coding for the polyketide synthases DEBS1-3, respectively. These comprise the streamlined machinery for the polyketide macrolactone precursor, 6-deoxyerythronolide B (**19**) (Scheme 7). One building block of propionyl-CoA (**11**), which functions as the starter unit, and six equivalents of methylmalonyl-CoA (**12**), used as elongation substrates, are fused in Claisen condensations. The AT domains of the individual modules are specific for the recruitment of **11** and **12**. The ketosynthase domain (KS) catalyzes the incorporation of **12** by release of carbon dioxide, in this way elongating

the chain. The ketoreductase domains (KR) reduce the β-keto groups to the respective alcohol. When dehydratase (DH) and enoyl reductase (ER) domains are present, the hydroxy function is eliminated and the resulting double bond removed by saturation. In the end, the terminal thioesterase domain (TE) cleaves off the linear precursor and catalyzes the macrolactone formation (Scheme 7).

The biosynthesis includes a set of hydroxylation, glycosylation and methylation events (Scheme 8). A first hydroxylation, catalyzed by the cytochrome P450 monooxygenase EryF, converts 6-deoxyerythronolide B (**19**) into erythronolide B (**20**).[48] In the following, two glycosylation events take place, installing sugar residues at the hydroxy functions of positions C-3 and C-5 using dTDP-ʟ-mycarose (EryBV) and dTDP-ᴅ-desosamine (EryCIII) as substrates.[45] This results in the formation of erythromycin D (**21**). Interestingly, the sugar biosynthesis pathways leading to these special sugar building blocks are also encoded in the gene cluster. Two sets of enzymes are able to convert glucose-1-phosphate into ʟ-mycarose and ᴅ-desosamine.[45] There are two routes in which **21** can be processed to yield the final metabolite **8** (Scheme 8). The favored enzymatic sequence consists of EryG followed by EryK. At first, one hydroxy group within the ʟ-mycarose moiety is methylated, thereby producing erythromycin B (**22**). Another hydroxylation event on the macrolide core at C-12 results in the formation of erythromycin A (**8**). Erythromycin C (**23**) is believed to be a shunt product, occurring when EryK hydroxylates first. This generates **23**, which can also be converted into **8** by EryG.
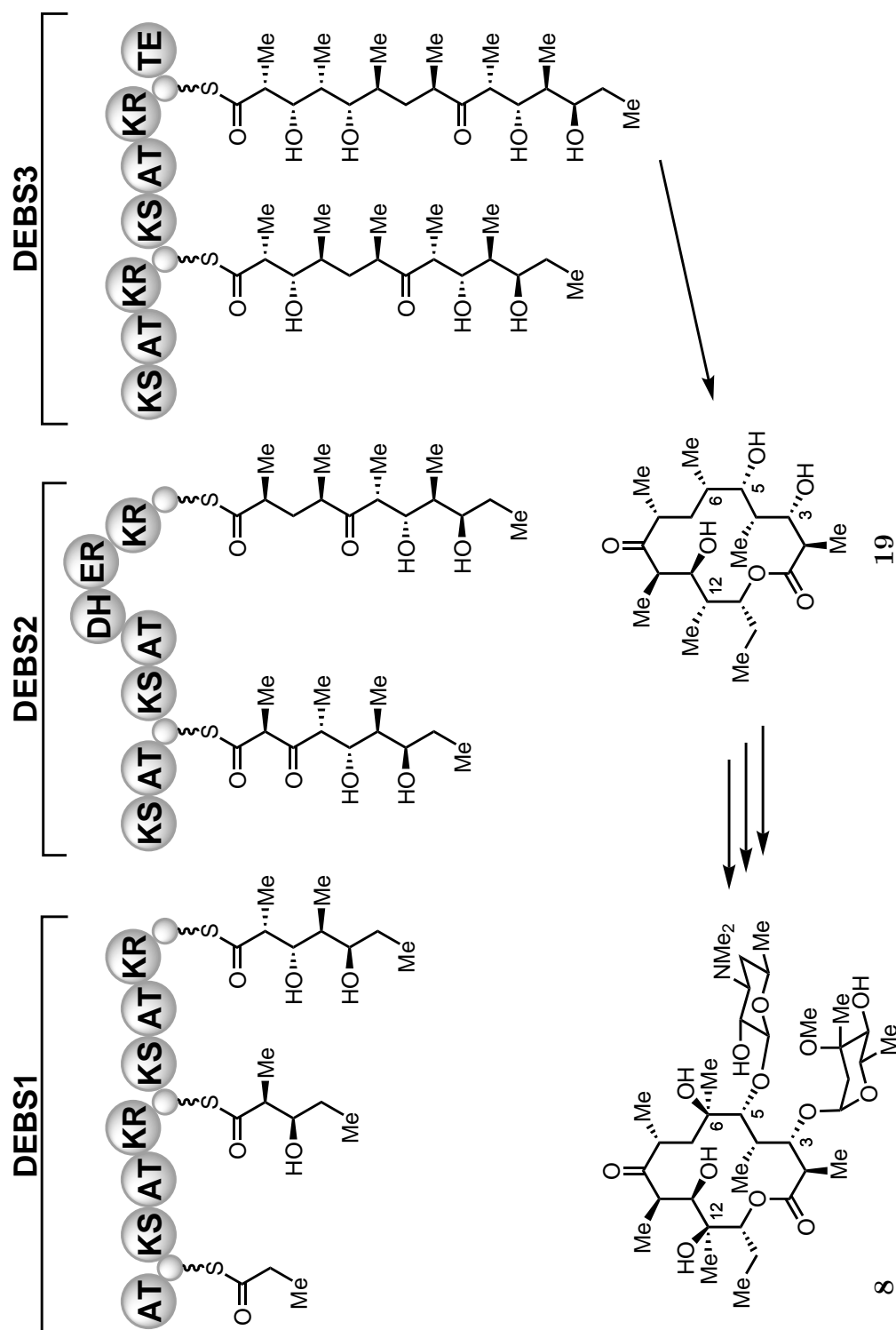
Scheme 7. The erythromycin biosynthetic assembly line. Three PKS enzymes (DEBS1-3) produce the macrolide core structure **19** which is processed to **8**. AT: Acetyltransferase domain, KS: Ketosynthase domain, KR: Ketoreductase domain, DH: Dehydratase domain, ER: Enoyl reductase domain, TE: Thioesterase domain.
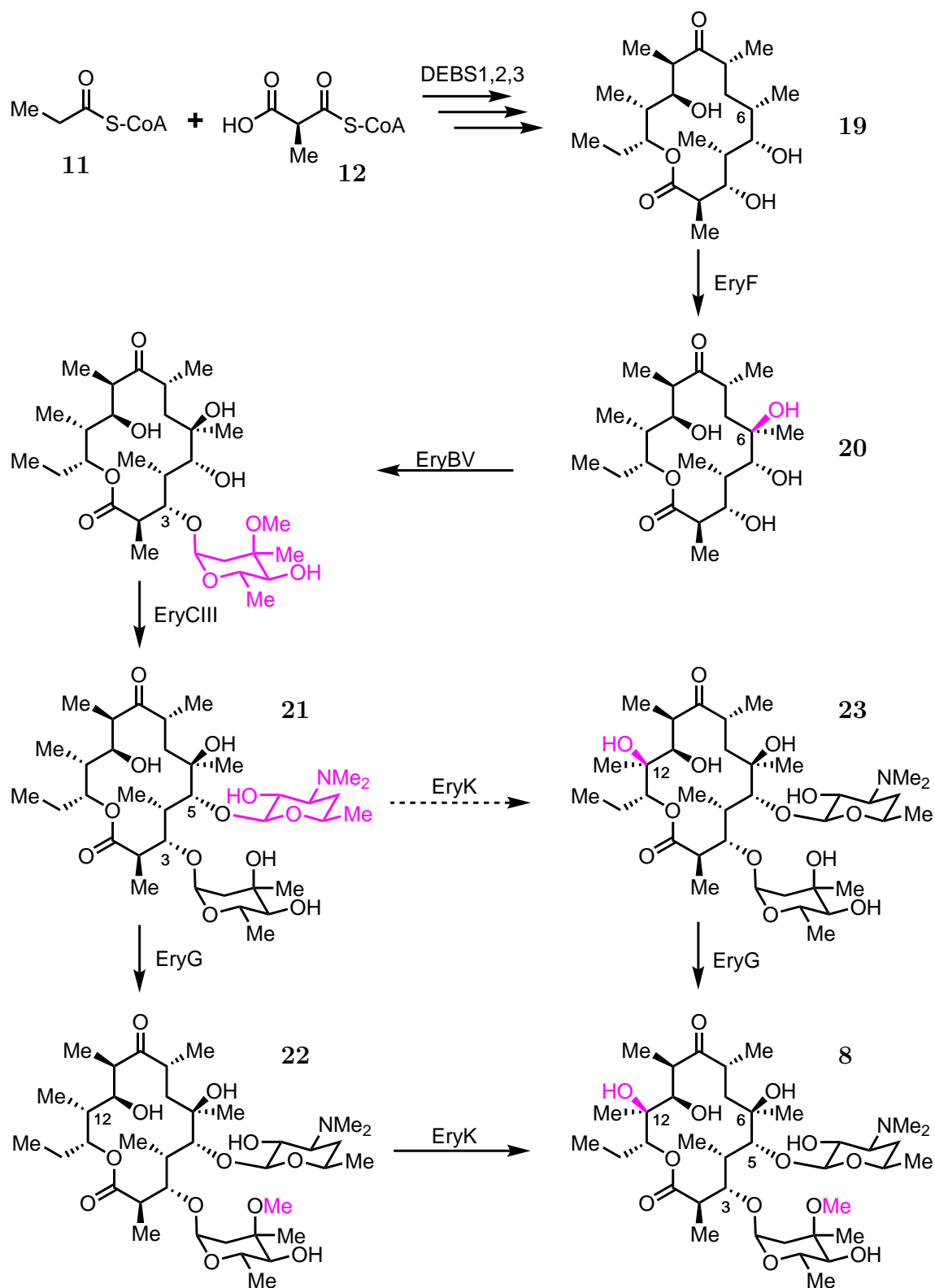
Scheme 8. Summary of erythromycin A (**8**) biosynthesis. The macrolactone core is produced from propionyl- (**11**) and methylmalonyl-CoA (**12**) and is modified further by hydroxylation, glycosylation and methylation, catalyzed by individual enzymes.

The erythromycin PKS system possessed AT domains that are integrated in each of the biosynthetic modules. Such systems are known as *cis*-AT PKS. In addition, *trans*-AT PKS systems do exist, in some cases also termed 'AT-less PKS'.[49] These enzymes lack the AT domains in their modular organization and instead receive the elongation substrate from free-standing ATs. First attempts to analyze genomic data of *Bacillus subtilis* 168, the producer of the antibiotic bacillaene (**24**), assumed sequencing errors as a possible reason for the missing AT domains, but it became evident that such a principle is utilized more frequently.[28] As a consequence, the rule of collinearity cannot be applied in *trans*-AT PKS. In comparison to modular type I PKS systems, predictions on the formed product are thus much more difficult: *trans*-AT modules can be used once or several times for the formation of an intermediate product.[50] Also, skipping of modules within the elongation process is possible. Compared to the *cis*-AT PKS, *trans*-AT PKS possess more variations in their modular composition, thereby achieving even higher structural variability. One feature is the presence of dehydrating bimodules that possess inactive ketosynthase domains (KS_0) to form conspicuous (conjugated) double bonds in *Z*-configuration as found in **24** (Figure 3).[49]
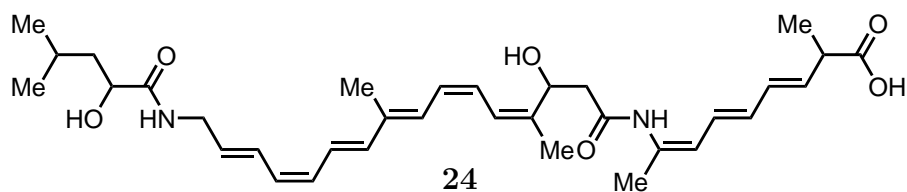

Figure 3. Structure of the antibiotic bacillaene (**24**) produced by *Bacillus subtilis* 168.

In modular PKS systems, domains belonging to non-ribosomal peptide synthetases (NRPSs) are frequently included, thereby contributing to the structural complexity of the related product structures. The biosynthetic principles of NRPS are highlighted in the next chapter.

1.4. **Non-ribosomal peptides**

Non-ribosomal peptides (NRPs) are another important group of secondary metabolites, mainly isolated from bacterial and fungal sources.[24] They often possess useful pharmacological properties, e.g., the antibiotics penicillin G (**1**) and vancomycin (**2**), presented previously, or the immunosuppressant cyclosporine (**25**) administered after organ transplantation (Figure 4).[51] NRPs consist of proteinogenic and non-proteinogenic amino acids. Many of the NRPS biosynthetic gene clusters encode enzymes that can derivatize amino acids from the primary metabolism or generate unusual non-proteinogenic amino acids *de novo*, thereby increasing the substrate pool.[52]

Multimodular enzymes, known as non-ribosomal peptide synthetases (NRPSs), produce the NRPs. The minimal set of catalytic domains consists of the condensation domain (C), which forms peptide bonds, an adenylation domain (A) and a peptidyl carrier protein (PCP), also termed 'thiolation domain' (T). The adenylation domain (A) selects the amino acid and activates it by consumption of ATP to form an energy-rich mixed anhydride with AMP (AMPylation) (Scheme 9). This allows the amino acid to be further processed as a substrate: Loading onto the PCP domain is possible, facilitating initiation of condensation reactions to synthesize a peptide chain.
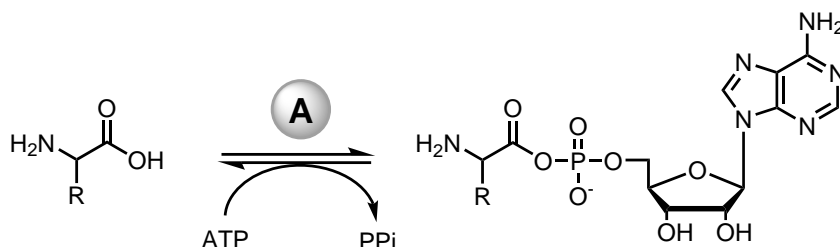
Scheme 9. Activation of an amino acid for NRPS biosynthesis in the form of an energy-rich mixed anhydride is catalyzed by the A domains of the NRPS.

Non-ribosomal peptide synthetases share similar modular principles compared to the type I PKS machineries. The A domain is comparable to a PKS AT domain as both are needed for substrate selection. The task of the PCP domain is identical to the ACP that is providing substrates for elongation reactions. Both serve as anchors for the covalently attached building blocks and biosynthetic intermediates. NRPS carrier proteins likewise undergo post-translational modification by phosphopantetheinylation.[24,30] In addition to the essential domains, a number of optional domains allow further product diversification. Epimerization domains (E) can be present, converting naturally occurring L-amino acids into the corresponding D-epimers. The proton of the α-C atom can be abstracted and

reprotonation initiated from the opposite site, in this way changing the configuration of the stereocenter. Further domains are known, e.g., catalyzing methylation (MT), cyclization (Cy), oxidative events (Ox), reduction (R) and formylation (F) reactions.[24] Together, these domains account for the massive structural diversity of NRPS-derived compounds. Interestingly, NRPs consisting of more than 25 amino acids do not exist.[24] The NRPS assembly line is usually containing a terminal thioesterase domain (TE) that cleaves off the processed peptide chain, in many cases simultaneously inducing macrocyclization.
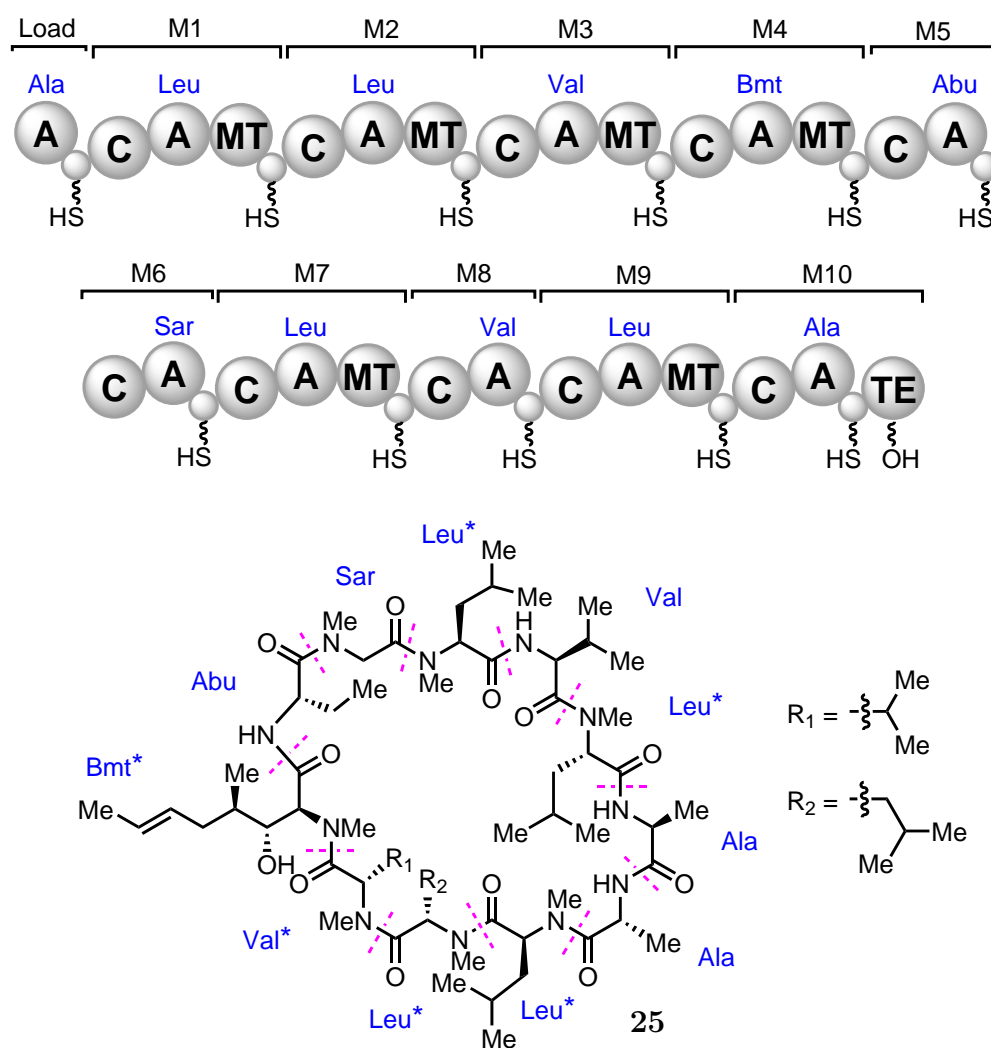
Figure 4. NRPS assembly line for the immunosuppressant cyclosporine (**25**). The individual modules utilize the proteinogenic amino acids alanine (Ala), leucine (Leu), valine (Val) and the non-proteinogenic building blocks sarcosine (Sar), L-α-aminobutyric acid (Abu) and butenyl-methyl-threonine (Bmt, **26**). The terminal thioesterase domain (TE) induces cyclization. *N*-methylations of the amino acids introduced by the methyltransferase domains (MT) are indicated (*).

1. Introduction

The biosynthesis of cyclosporine (**25**) is conducted by a multimodular NRPS (Figure 4).[51,53] It consists of a loading module and ten extension modules that select the proteinogenic amino acids alanine (Ala), leucine (Leu), valine (Val) as well as the non-proteinogenic building blocks sarcosine (Sar), L-α-aminobutyric acid (Abu) and butenyl-methyl-threonine (Bmt, **26**). As a result, a linear, TE-bound peptide chain of eleven amino acids Ala-Leu-Leu-Val-Bmt-Abu-Sar-Leu-Val-Leu-Ala is formed. Some modules possess methylation domains (MT) responsible for the *N*-methylation of the respective amino acids, thereby increasing the structural complexity. The thioesterase domain (TE) induces lactamization and release of **25**.

Due to highly similar biosynthetic principles, hybrid pathways of NRPS and PKS enzymes exist. The catalytic power of both are combinable, in this way accessing the full potential to produce complex and structurally demanding molecules.[54,55] The biosynthesis of the cyclic peptide cyclosporine (**25**) requires the special building block butenyl-methyl-threonine (Bmt, **26**).[56] **26** is derived from a PKS pathway in which acetyl-CoA (**9**) and malonyl-CoA (**10**) are condensed (Scheme 10). In further steps, *S*-adenosyl methionine (SAM) provides a methylgroup, followed by incorporation of another C-2 building block derived from **10**. Reductions involving NADPH, as well as oxidative formation of an α-keto function, lead to a precursor that is converted into **26** by transamination. **26** in its AMPylated form is then processed by the NPRS module to produce **25**.
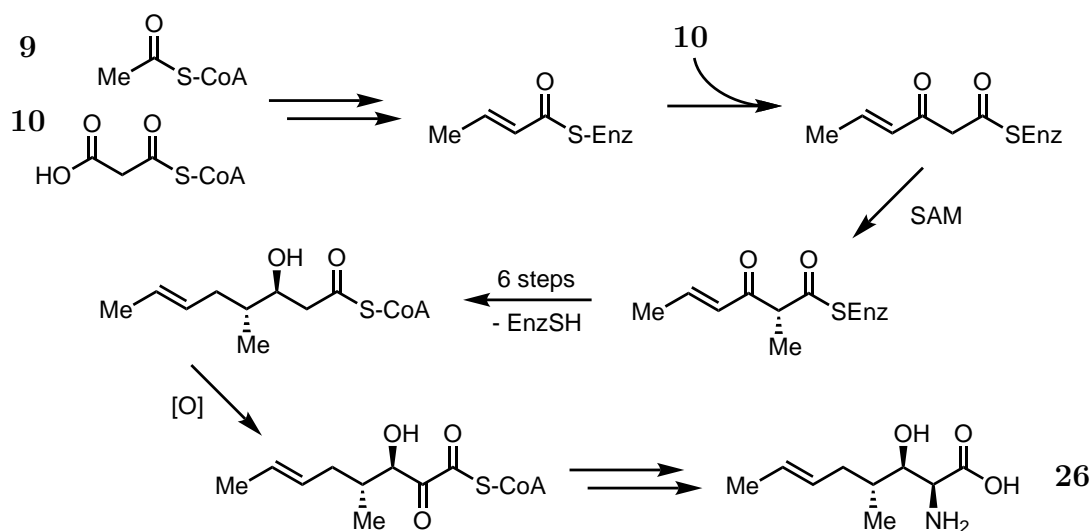
Scheme 10. Biosynthetic pathway of butenyl-methyl-threonine (Bmt, **26**).[56]

In this work, a particularly unusual hybrid PKS/NRPS system will be presented, which uses the catalytical power of both types of modules for the construction of structurally demanding frameworks.

1.5. Polycyclic tetramate macrolactams

Polycyclic tetramate macrolactams (PoTeMs) comprise a relatively new class of bacterial secondary metabolites. They all share a tetramic acid moiety that is fused to a macrolactam ring. Since the discovery of ikarugamycin (**27**) in 1972 by Jomon *et al.*,[57] many more related compounds were isolated from various marine and terrestrial bacteria all over the globe (Figure 5). The most prominent examples apart from **27** and its hydroxyl-derivative butremycin (**28**)[58] are dihydromaltophilin,[59] better known as the 'heat-stable antifungal factor' (HSAF) (**29**),[60] isolated from *Lysobacter enzymogenes* C3,[61] the frontalamides (**30a**,**b**), from *Streptomyces* sp. SPB78 associated with the southern pine beetle *Dendroctonus frontalis*[62] and the alteramides (**31a**,**b**) from sponge-associated *Alteromonas* sp.[63] More representatives that highlight the structural diversity are maltophilin (**32**) from a *Stenotrophomonas maltophilia* strain,[64] the clifednamides (**33a**,**b**),[65] geodin A (**34**) from a southern Australian sponge *Geodia* sp.,[66] xanthobaccin A (**35**) from a *Stenotrophomonas* sp.,[67] and further sponge-accociated compounds like discodermide (**36**),[68] cylindramide (**37**),[69,70] as well as aburatubolactam (**38**), isolated from a marine *Streptomyces* sp.[71–73]

Within the PoTeM structural framework, carbacyclic ring systems are fused to the macrolactam portion. Different arrangements exist, formally consisting of two or three cyclopentane/-hexane units. Tricyclic constitutions with a 5-5-6 pattern, e.g., found in HSAF (**29**) or frontalamides (**30a**,**b**), but also 5-6-5 cyclization, e.g., in ikarugamycin (**27**) or the clifednamides (**33a**,**b**), are occurring. In addition, bicyclic arrangements possessing a 5-5 ring pattern were identified, e.g., in alteramides (**31a**,**b**), geodin A (**34**), cylindramide (**37**) or aburatubolactam (**38**) (Figure 5). The PoTeM scaffolds furthermore vary in their oxygenation and alkylation patterns, as well as in the degree of saturation. The structural diversity results in a broad spectrum of biomedical activities, ranging from antibiotic and anti-fungal to anti-trypanosomale and cytotoxic, making these compounds promising lead structures.[57–65,67–72] However, the mode of action for many of them is yet poorly understood.

Although these compounds are exciting from structural and biomedical perspectives, only little was known about their biosynthetic origin until recently. In the last decade, the underlying principles for the formation of PoTeMs have intensively been studied. The first functional studies on a PoTeM biosynthetic pathway were conducted by Du *et al.* on HSAF (**29**) biosynthesis.[60] The group identified a bimodular PKS/NRPS system composed of nine domains, five of them, KS-AT-DH-KR-ACP, being PKS-related and four, C-A-PCP-TE, typical for NRPS (Figure 6). Owing to the structural similarity of **29** to the spinosyns,[74]
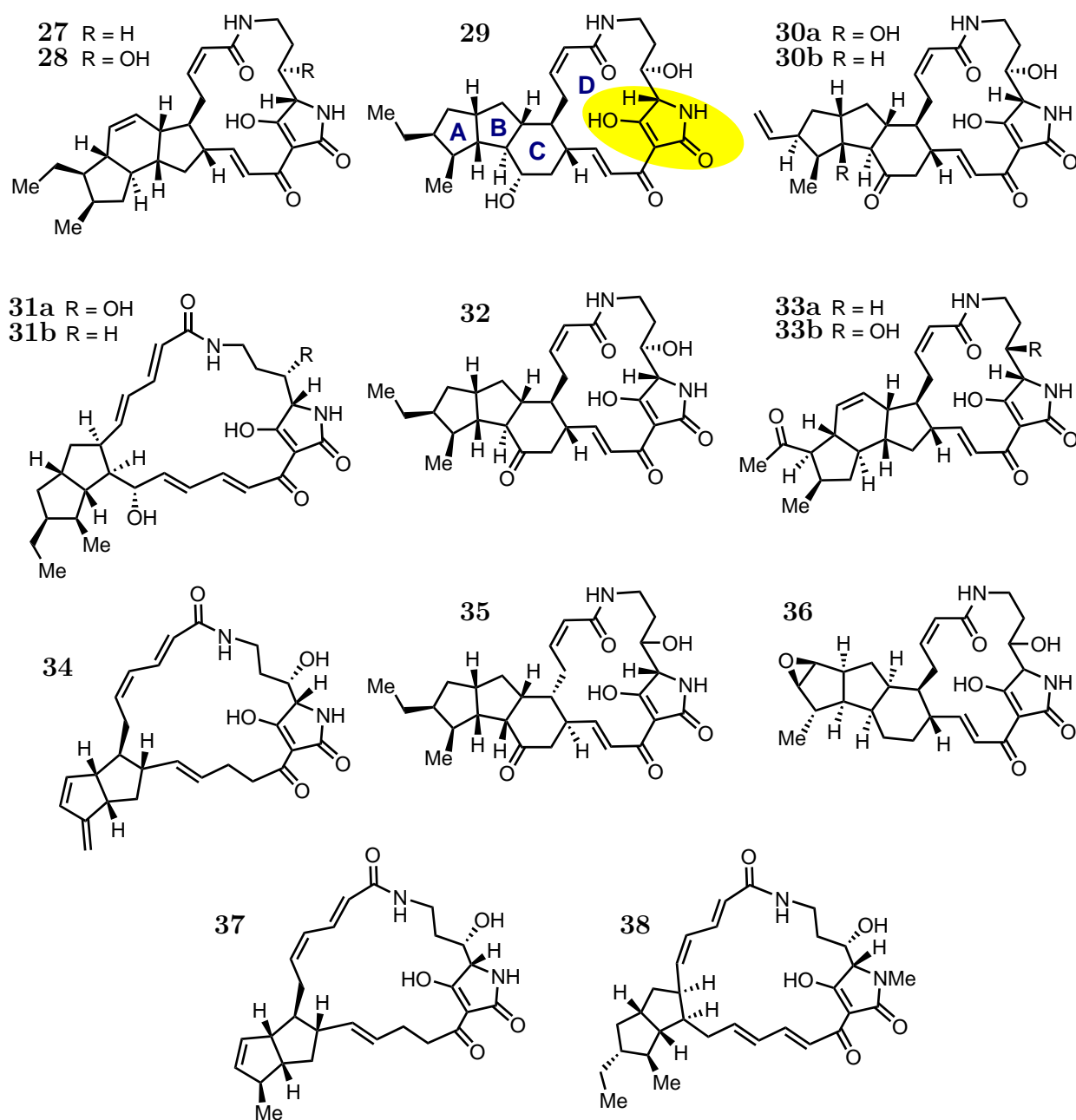
Figure 5. Selection of structures of literature reported PoTeMs: ikarugamycin (**27**), HSAF (**29**), the frontalamides (**30a,b**), the alteramides (**31a,b**), maltophilin (**32**), the clifednamides (**33a,b**), geodin A (**34**), xanthobaccin A (**35**), discodermide (**36**), cylindramide (**37**) and aburatubolactam A (**38**). The A-B-C-D ring system is defined in **29**. The tetramic acid moiety is highlighted in yellow.

a related modular PKS biosynthetic pathway was expected by Du *et al.*[60] The bimodular PKS/NRPS identified in *Lysobacter enzymogenes* C3 was thus assumed to constitute the final two modules of such a PKS/NRPS machinery. However, the remaining modules of the system could not be identified. Disruption of the PKS/NRPS gene indeed abolished production of **29**, thus clearly establishing its importance in HSAF biosynthesis.[60] Based on this hypothesis, Yu *et al.* proposed incorporation of L-ornithine (**39**) and two carbon atoms C-26 and C-27 from malonate by the PKS/NPRS bimodule (Figure 6). They speculated that another PKS gene, in theory encoding eleven more modules, would still be needed to form the remaining carbon-carbon bonds. In addition, the cyclizations were thought to be controlled by the encoded ferredoxin reductase and sterol desaturase, which meanwhile turned out to be wrong. At that time, only a 13 kb partial region of the HSAF gene cluster was regarded, in total comprising four genes, coding for the PKS/NRPS enzyme, an arginase, ferredoxin reductase and sterol desaturase (Figure 6).[60]
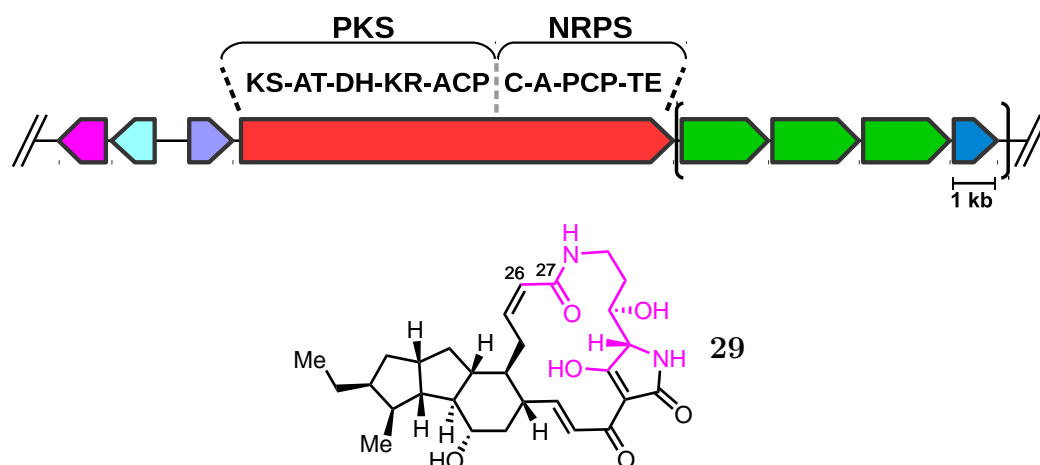
Figure 6. Schematic representation of the HSAF (**29**) biosynthetic gene cluster from *Lysobacter enzymogenes* C3.[60,75] It was postulated that one PKS/NRPS bimodule would generate a minor portion of the skeleton of **29**, highlighted in pink. Genes in brackets encoding downstream enzymes were not yet identified at that time. Color code – red: PKS/NRPS; green: FAD-dependent oxidoreductase; blue: alcohol dehydrogenase; violet: PoTeM hydroxylase; turquoise: ferredoxin reductase; magenta: arginase.

The general picture of PoTeM biosynthesis changed in 2010, when Clardy *et al.* identified the frontalamides (**30a,b**) in *Streptomyces* sp. SPB78, a fully sequenced bacterial strain.[62] Bioinformatic analysis showed that a similar PKS/NRPS gene is the central genetic element, flanked upstream by one gene coding for a sterol desaturase, and downstream by genes for phytoene dehydrogenases, an alcohol dehydrogenase and a cytochrome P450 enzyme (Scheme 11).[62] Focused on the frontalamides, Blodgett *et al.* proved by gene knock-outs

1. Introduction

that the PKS/NRPS gene (*ftdB*) from the frontalamide cluster is responsible to generate the whole carbon framework of **30a**,**b**. By further screening of the *Streptomyces* sp. SPB78 genome, it was verified that no other suitable PKS genes belonging to the *ftd* cluster exist, as previously proposed by the Du lab,[60] indicating an iterative mechanism of the PKS module with the ability to catalyze the production of a second polyketide chain before condensation with L-ornithine (**39**).[62] Such iterative systems are common for secondary metabolites from fungi, but rather unusual for bacterial type I systems. Nevertheless, their proposal for the cyclization chemistry to install the ring systems of PoTeMs was not plausible. Aldol reactions were depicted, but not applicable to explain the reactivity on all positions of the proposed intermediate (Scheme 11). In their model, Clardy *et al.* assumed the formation of β-polyketo-chains, including one reduced position with a double bond in *trans* configuration.[62] The cyclizations would be catalyzed by the redox tailoring enzymes FtdC-FtdF, encoded downstream of the PKS/NRPS gene.[62] For the first time, a proposal was made that included the redox tailoring enzmyes to contribute to the PoTeM cyclization biochemistry.
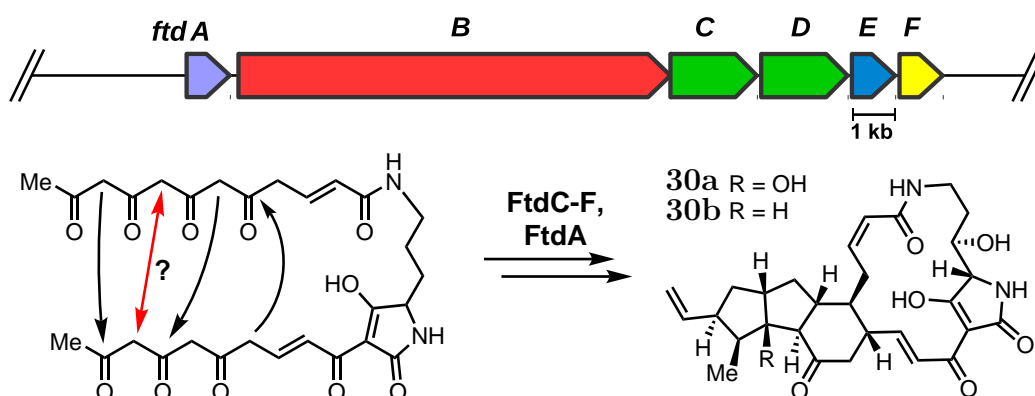
Scheme 11. Postulated biosynthesis by Clardy *et al.* for the frontalamides (**30a**,**b**).[62] An iterative mode of FtdB (PKS/NPRS) was proposed generating two polyketide chains that are fused to L-ornithine (**39**). Further modifications leading to **30a**,**b** are installed by the encoded tailoring enzymes FtdA and FtdC-F. Color code – red: PKS/NRPS; green: FAD-dependent oxidoreductase; blue: alcohol dehydrogenase; yellow: cytochrome P450; violet: PoTeM hydroxylase.

By expanded bioinformatic analysis of microbial genomes, it was possible to retrieve putative PoTeM BGCs in a huge variety of phylogenetically diverse bacteria for the first time.[62] These ranged from Gram-positive actinomycetes to Gram-negative γ-proteobacteria that were not known to be potential producers of PoTeM compounds before, indicating PoTeM gene clusters to be widespread in bacteria. Blodgett *et al.* were able to relate the genes from the frontalamide gene cluster *ftd* of *Streptomyces* sp. SPB78 to homologs

of other putative clusters from sequenced microbes, thereby identifying the conserved synteny of PoTeM gene clusters among species.[62] The sequence conservation allowed cluster screenings by PCR resulting in the discovery of the clifednamides (**33a**,**b**) from an environmental *Streptomyces* strain.[65]

In 2011, the complete sequence of the HSAF (**29**) biosynthetic gene cluster with a size of about 20 kb was published.[75] Eight genes in total were annotated that are positioned next to each other (Figure 6). Disruption mutations placed in the FAD-dependent oxidoreductase genes abolished the production of **29**, confirming the importance of these genes that were disregared by Yu *et al.* at the beginning. Further, Lou *et al.* tested that the A domain of the NRPS specifically activates L-ornithine (**39**),[75] which is functionalized at its α- and δ-amine groups with the polyketide precursor chains. Experiments were conducted to understand the formation of the tetramic acid moiety. ACP-bound stearoyl fatty acids as simplified substrate mimics (acyl-*S*-ACP) were used together with **39**, proving the ability of the NRPS module to generate tetramic acid products *in vitro*.[75] The reaction was confirmed by comparison to a synthetic standard of the product. From these data, it was concluded that the tetramic acid portion, an unifying structural motive of all PoTeMs, is formed by the thioesterase domain (TE) in a Dieckmann-like reaction.[75]

β-Hydroxyornithine was speculated to be the building block explaining the hydroxy moiety in HSAF (**29**).[60] Blodgett *et al.* were able to relate the first gene, *ftdA*, of the frontalamide gene cluster to a possible hydroxylating function, derivatizing the C-3 position of the PoTeM core, as found in the frontalamides (**30a**,**b**). But they could not ascertain the timing during biosynthesis.[62] Evidence in this regard was provided by a study from Li *et al.*, who confirmed for **29** that the so-called 'sterol desaturases' encoded in the HSAF gene cluster introduce the hydroxy group at the L-ornithine (**39**) moiety of **29** as the last step in its biosynthesis.[76] Feeding experiments using the non-hydroxylated precursor of HSAF were conducted and enzyme extracts prepared, in both ways achieving a conversion into **29**, thus validating the *in vitro* results.[76]

For a long time, it was unclear which genes are essential to build up a functional PoTeM biosynthetic machinery. On the basis of studies conducted in the Gulder lab focusing on ikarugamycin (**27**), a first functional proof was made that only three genes are needed to produce this PoTeM.[77] These include the PKS/NRPS hybrid enzyme (*i*PKS/NRPS), a FAD-dependent oxidoreductase and an alcohol dehydrogenase. The ikarugamycin gene cluster *ika* of *Streptomyces* sp. Tü6239 was identified by the central PoTeM-PKS/NRPS gene (*ikaA*) with its size of approximately 10 kb (Scheme 12). Adjacent to this, the two reductase genes (*ikaB*, *ikaC*) are positioned on the same operon. Transfer of the whole

cluster *ikaABC* into *Escherichia coli* BAP1,[78] under the control of a T7 promoter, made the cells capable of producing **27**.[77] The borders of the cluster were for the first time firmly established by this experiment and demonstrated that PoTeM clusters are simply structured. In the case of **27** only three genes are involved, which is in sharp contrast to the rather complex molecular framework of the final product **27**.[77] This result was also important in clearly showing that the PKS module must work iteratively to form the PoTeM carbon frame.

On the basis of these new results, the present model for the biosynthesis of PoTeMs is conceived (Scheme 12): The *i*PKS part of the PKS/NRPS megaenzyme (IkaA) with its domains KS-AT-DH-KR-ACP generates two polyene chains. They are linked to L-ornithine (**39**) by the C domain of the NRPS module, with the final installation of the tetramic acid moiety to form **40**, a central precursor molecule, likely of all PoTeMs. The reductive tailoring enzymes are responsible to fully construct the molecular framework of **27**. They introduce the carbacyclic portion and make changes to the oxidative state of the carbon skeleton, thereby installing the complex molecular frameworks of PoTeMs. The FAD-dependent oxidoreductase IkaB and the alcohol dehydrogenase IkaC control the cyclization reactions in a sequential manner, by hydride induced ring-closing reactions. In the case of **27**, the FAD-dependent oxidoreductase converts **40** into the stable intermediate **41**, involving a spontaneous Diels-Alder reaction, thereby generating ring systems A and B. At last, the alcohol dehydrogenase closes ring C, forming the 5-6-5 carbacycle to produce **27**.[77]

Shortly after the report of this work, Zhang *et al.* reconstituted the biosynthesis of **27** in *S. lividans* TK64 validating the above results. An ikarugaymcin cluster of *Streptomyces* sp. ZJ306[79] was identified by PCR-based screening from a genomic library, similar to Antosch *et al.* for *Streptomyces* sp. Tü6239,[77] targeting conserved domain sequences of the PKS/NPRS gene. In feeding experiments employing sodium [1-^{13}C]acetate, it was confirmed that a PKS pathway is involved in the production of **27** in the natural producer *Streptomyces* sp. ZJ306. By sequential knockout of the genes *ikaB* and/or *ikaC*, they were able to detect the biosynthetic intermediates **40** and **41**. The stable intermediate **41** with its 5-6 carbacycle, produced by the FAD-dependent oxidoreductase (IkaB), was isolated and confirmed by high resolution mass spectrometry and NMR spectroscopy.[79] Furthermore, Zhang *et al.* purified and characterized the respective alcohol dehydrogenase IkaC and demonstrated its role in the biosynthetic sequence *in vitro*. IkaC catalyzes a Michael addition-like reaction, in which ring C is installed by formal incorporation of H$_2$, thereby converting **41** into **27**.[79] Also, the co-substrate selection of the alcohol
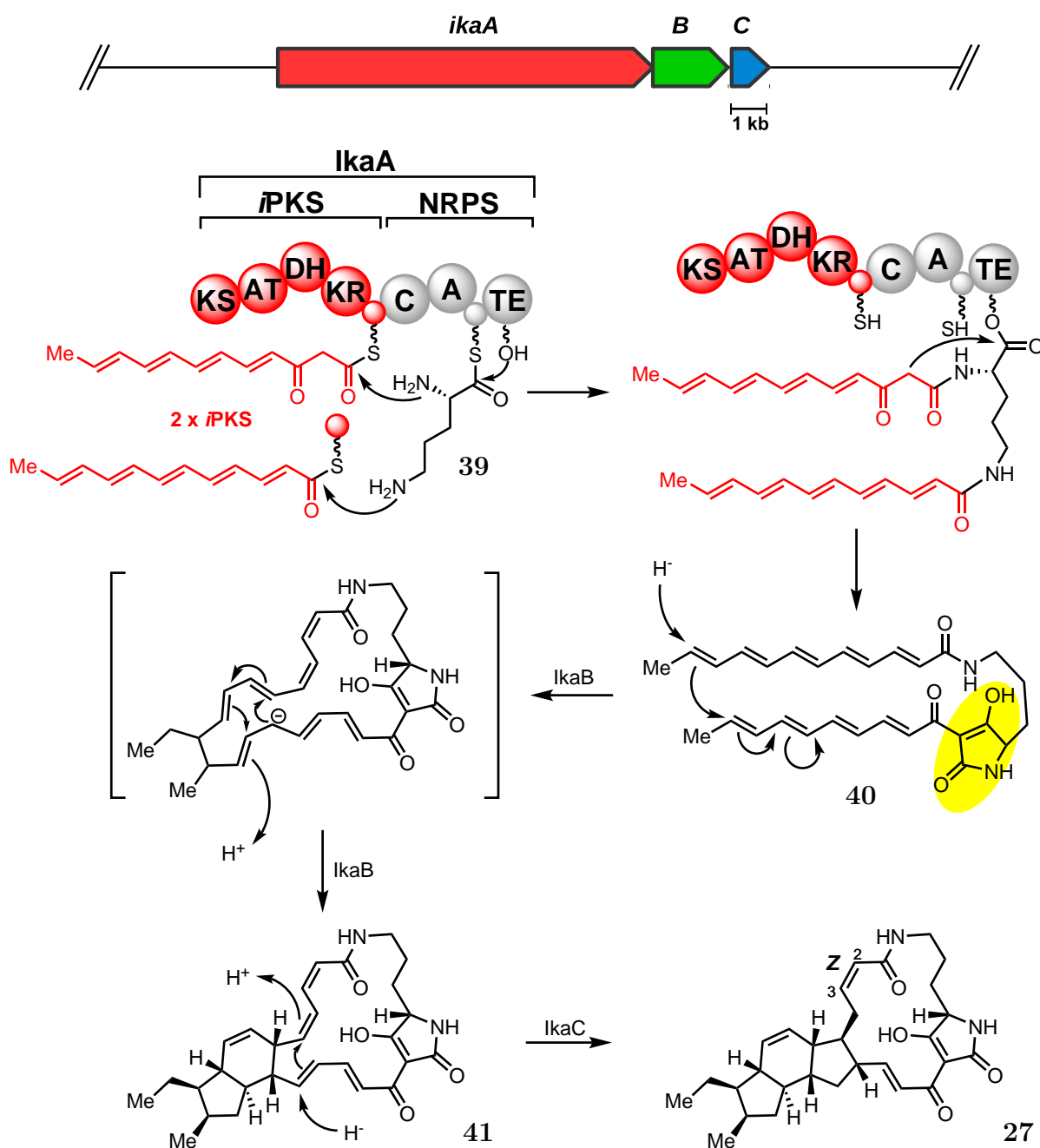
Scheme 12. Ikarugamycin (**27**) biosynthesis.[77,79] The three gene cluster *ikaA/B/C* for **27** provides the *i*PKS/NRPS (IkaA), the FAD-depentent oxidoreductase (IkaB) and an alcohol dehydrogenase (IkaC). Acetyl-CoA (**9**) and malonyl-CoA (**10**), together with L-ornithine (**39**) are condensed to the unsaturated precursor **40**, catalyzed by the *i*PKS/NRPS enzyme (IkaA). The *i*PKS portion (red) produces two unsaturated polyenes in repetitive Claisen condensations. The NRPS module (gray) recruits **39** and the thioesterase (TE) domain catalyzes the formation of the tetramic acid moiety (yellow) in a Dieckmann-like reaction. According to Antosch *et al.*, ikarugamycin (**27**) is formed from **40** in sequential ring closure reactions, including the stable precursor **41**, catalyzed by the reductases IkaB and C.[77]

dehydrogenase was experimentally validated. It accepts NADPH and NADH as reduction equivalents.[79] This study was another proof for the sequential action of IkaB and IkaC, supporting the present knowledge on PoTeM biosynthesis.

The early steps in PoTeM biosynthesis, the formation of the dodecaene acid chains from acetyl-CoA (**9**) and malonyl-CoA (**10**) and their attachment to L-ornithine (**39**) to form **40** are still not well understood (Scheme 13). The existence of the early polyene precursor **40** was proven by LC-MS experiments, but its isolation and characterization is challenging because this compound easily decomposes due to heat and light exposure.[79,80] A study from the Du lab described the generation of the polyene acid chains on purified PKS enzyme, measured after tryptic digest.[80]
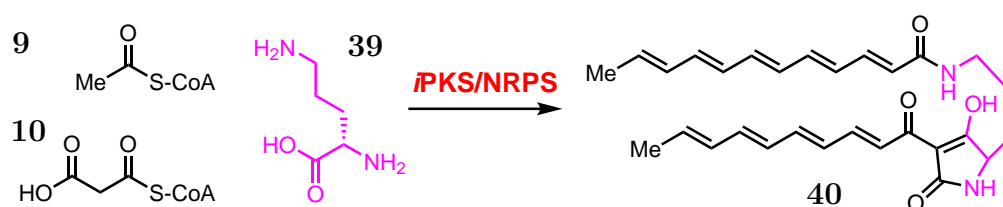
Scheme 13. Origin of PoTeMs. The basic building blocks acetyl-CoA (**9**) and malonyl-CoA (**10**), together with L-ornithine (**39**) are required to form the unsaturated precursor **40** by the *i*PKS/NRPS enzyme. **40** is believed to be common in all biosynthetic pathways of PoTeM compounds known to date, but difficult to isolate.

Although a dozen of PoTeM gene clusters were meanwhile identified in diverse bacteria,[62] many of them have not been related to a chemical structure. The functions of the encoded PoTeM enzymes are still awaiting detailed elucidation. The studies of Antosch *et al.* and Zhang *et al.* highlighted the importance of the accessories enzymes, in particular the FAD-dependent oxidoreductases for PoTeM scaffold formation. However, an explanation on how they control the installation of the carbacyclic portions, e.g., to generate 5-6-5 carbacycles as in **27** or 5-5-6 in **29** and **30a**,**b**, is still lacking. The mechanistic importance of the variable number of FAD-dependent oxidoreductase genes in diverse PoTeM gene clusters needs to be evaluated. In the case of the clusters responsible for the formation of **27**, only one gene is involved, but in others pathways, two or at most three copies are present (Figure 6 and Scheme 11).[62] It will be exciting to know whether all of these are functional, and, in particular, what the individual functions are. Detailed research on PoTeM enzymes is thus needed to fully understand the underlying molecular principles.

2. Aims of this thesis

Based on the successful heterologous production of ikarugamycin (**27**) in the Gulder lab, it should be tested whether such a structurally complex secondary metabolite can be synthesized in an *in vitro* setup, only applying the purified biosynthetic enzymes. Are the three biocatalysts (IkaABC) provided by the gene cluster of *Streptomyces* sp. Tü6239 sufficient to do so? Can the biosynthetic intermediates **40** and **41** be produced by omitting the reductive enzymes IkaB and/or IkaC?

To answer these questions, the relevant genes from the ikarugamycin biosynthetic gene clusters of *Streptomyces* sp. Tü6239[77,81] and *Salinispora arenicola* CNS-205 need to be cloned. Suitable plasmid systems, pHis8, pGS-21a (both *E. coli*) and pXY200 (*Streptomyces*) were chosen, facilitating to heterologously produce the biocatalysts as His-tagged fusion proteins.

Once enzyme production is successfully established, cellular lysates should be prepared and mixed to confirm the catalytic activity of a triplet IkaABC, to set the stage for an *in vitro* reaction setup. This requires the building blocks acetyl-CoA (**9**), malonyl-CoA (**10**), as well as L-ornithine (**39**) and additional cofactors such as NADPH and ATP. A phosphopantetheinyl transferase (PPTase) Sfp together with CoA will be needed to activate IkaA by installing the PPant residues. In addition, a suitable buffer system has to be identified. The use of full-length hybrid *i*PKS/NRPS could conclusively prove the iterative mode of the PKS part, as it has to convert one building block **9** and five units **10** in repetitive Claisen condensation reactions to make a single hexaketide polyene chain. By introducing L-ornithine (**39**), the PoTeM biosynthetic precursor **40** will be formed, which is converted into **41** and the final product **27**.

Having established the *in vitro* production system for **27**, different possibilities arise. It could be used to explore the substrate scope. Altered starting materials, if accepted, might lead to the genesis of unnatural compounds. Enzymes from putative PoTeM gene clusters, previously identified by 'genome mining' and available in the Gulder lab, could be tested. The biocatalytic potential of these needs to be unraveled. More structurally exciting PoTeM representatives are thought to be discovered, that are encoded in putative (silent) gene clusters. Cloning of the respective genes from fosmid libraries has to be conducted

for subsequent heterologous production of the respective enzymes. By application of new enzyme combinations, the *in vitro* system for the total synthesis of **27** could be used to make novel PoTeM species.

In a related project, hydroxylating enzymes encoded in many PoTeM clusters upstream of the PKS/NRPS genes should be investigated. Blodgett *et al.* and Li *et al.* were able to make mutants lacking the putative 'sterol desaturase', resulting in the genesis of C-3 non-hydroxylated HSAF (**29**).[62,76] Li *et al.* were able to convert 3-deOH-HSAF into **29** by the use of 'sterol desaturase' enzyme extracts and feeding of the non-hydroxylated precursor to enzyme-producing *E. coli*, in this way confirming the proposed biochemistry.[76] In the same manner, ikarugamycin (**27**) should be converted into its hydroxyl derivative, butremycin (**28**). For this, PoTeM clusters have to be analyzed for genes coding for 'sterol desaturases' or 'fatty acid hydroxylases'. Subsequently, after successful cloning of promising candidates, the respective enzymes need to be prepared to perform the hydroxylations *in vivo* and *in vitro*.

Another focus is to set the stage for the crystallization of PoTeM enzymes. So far, there is no crystal structure available to shed light on the underlying molecular mechanisms. The early steps in PoTeM biosynthesis, the formation of the unsaturated precursors and the following cyclization reactions are of high interest. This might provide a more detailed knowledge on the molecular mechanisms of the reductive cyclization steps which are to date still not well understood.

For the cloning of large genes and entire PoTeM biosynthetic gene clusters, a PCR-based method has been developed and termed 'Direct Pathway Cloning' (DiPaC). It should be investigated if DiPaC can be expanded for the targeting of medium- and large-sized gene clusters. As a prominent example, the erythromycin cluster *ery* (54.6 kb) from *Saccharopolyspora erythraea* was chosen. The heterologous production of erythromycin A (**8**) or macrolide precursors will complete this thesis.

3. Results and Discussion

## 3.1. Promiscuous hydroxylases for the functionalization of polycyclic tetramate macrolactams – conversion of ikarugamycin to butremycin

In nature, many PoTeMs exist that possess a hydroxy group at the former L-ornithine moiety (C-3 atom). The most prominent examples are HSAF (**29**) and the frontalamides (**30a**,**b**).[60,62] This functional group is introduced by cluster-encoded enzymes coding for PoTeM hydroxylases that are found upstream of the central PKS/NRPS gene.[60,62,82] NCBI BLAST similarity analysis annotates these as 'sterol desaturases' or 'fatty acid hydroxylases'. Previous work showed that knockouts on the respective gene in the HSAF cluster resulted in the isolation of unhydroxylated compounds,[62] concluding the hydroxylation event to take place after the formation of the PoTeM carbon skeleton.[76] Li *et al.* were able to use a homologue enzyme from *Lysobacter enzymogenes* C3 to convert the HSAF precursor lacking the hydroxy group (3-deOH-HSAF) into **29**.[76]

In 2014, Kyeremeh *et al.* described the isolation of butremycin (**28**) from the actinomycete strain *Micromonospora* sp. K310, found in Ghanaian mangrove river sediments.[58] Butremycin (**28**) is the C-3 hydroxyl derivative of ikarugamycin (**27**) (Scheme 14). Similar to **27**, **28** shares a 5-6-5 ring pattern, but different, compared to **29** and **30a**,**b** with a 5-5-6 carbacycle.

In the current work, it should be tested whether an enzymatic conversion of **27** into **28** is possible by applying the PoTeM hydroxylases from *Saccharophagus degradans* DSM 17024 and *Lysobacter capsici* DSM 19286. The genomes of these bacteria contain PoTeM clusters and *L. capsici* is a producer of HSAF (**29**), showing that a functional

3. Results and Discussion
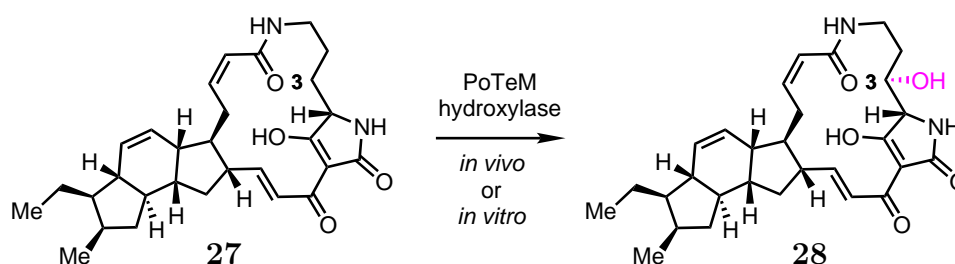
Scheme 14. Selective enzymatic hydroxylation of ikarugamycin (**27**) results in the production of butremycin (**28**).

hydroxylating gene product must be present. The gene cluster from *S. degradans* has not been related to a PoTeM structure so far, but it possesses a corresponding hydroxylase gene. The respective genes were bioinformatically analyzed and open reading frames (ORFs) determined. Cloning of these genes into *E. coli* compatible expression plasmids pHis8/pGS-21a facilitated heterologous production of the corresponding proteins. These vectors were used to upgrade the *E. coli* BAP1 pJA::*ikaABC* system, developed by Antosch *et al.* for the heterologous production of **27**.[77]

In addition, lysates containing PoTeM hydroxylases were prepared to test whether the biocatalytic reaction to produce **28** can be conducted *in vitro*. Together with standards of **27** and **28**, the conversion reactions were validated for both approaches, *in vivo* and *in vitro*.

The results indicate promiscuity of the investigated enzymes towards the applied substrate. Although the hydroxylase from *L. capsici* is derived from a HSAF (**29**) production pathway, it is capable to catalyze the desired reaction on the skeleton of **27** with an *as*-indacene (5-6-5) carbacycle. The hydroxylase from *S. degradans* performed in the same way. This led to the conclusion that enzymes from different PoTeM producers can be combined to produce functionalized metabolites. This opens the door to apply novel biocatalytic strategies for the derivatization of secondary metabolites of this class. On the basis of this knowledge, the hidden biosynthetic potential of (putative) PoTeM gene clusters will be unraveled in future studies. PoTeM hydroxylases are exciting tools, as they manage to selectively introduce last-stage modifications, for which no comparable organic-synthetic approach exists to date.

ChemComm

Promiscuous hydroxylases for the functionalization of polycyclic tetramate macrolactams – conversion of ikarugamycin to butremycin†‡

Christian Greunke,§ Janine Antosch§ and Tobias A. M. Gulder*

Polycyclic tetramate macrolactams (PTMs) are a structurally, biomedically and biosynthetically intriguing class of bacterial metabolites. By combining parts of the machineries of different PTM biosynthetic pathways, we demonstrate for the first time the substrate promiscuity of a class of PTM tailoring enzymes, thereby facilitating the (bio)synthesis of butremycin.

In the course of evolution, Nature has developed a large set of diverse natural products with unique structures and biological functions. An interesting group of such metabolites are the PTMs that are produced by bacteria. Depending on their individual structures, PTMs exhibit a number of valuable biological activities, ranging from antimicrobial to cytotoxic.[1–6] All PTMs contain a tetramic acid moiety incorporated into a macrolactam ring that is further modified with a fused carbocyclic system. Both overall size of the macrolactam and oxidation levels of the individual carbon atoms vary in different PTMs, as does the number of rings and their exact cyclization pattern in the carbocyclic portion. Overall, this leads to the structural diversity of this natural product class, with HSAF (**1**)[1] and the frontalamides (**2**),[2] alteramide A (**3**),[3] and ikarugamycin (**4**)[4] being typical examples (Fig. 1).

Triggered by initial work on the structure and biosynthesis of **1** by Du *et al.*[7a] and of **2** by the Clardy lab,[2] PTM biosynthetic pathways have meanwhile attracted significant attention. In particular the pioneering work by Du on HSAF assembly in *Lysobacter*[7] and our contribution towards understanding the biosynthesis of **4**[8] that was later validated by Zhang *et al.*[9] have revealed a number of interesting features of PTM biosynthesis. These include the unusual catalytic activities of the thioesterase domain in the mixed polyketide–non-ribosomal peptide (PKS–NRPS) biosynthetic machinery,[7b,c]

Biosystems Chemistry, Department of Chemistry and Center for Integrated Protein Science Munich (CIPSM), Technische Universität München, Lichtenbergstraße 4, 85747 Garching, Germany. E-mail: tobias.gulder@ch.tum.de

† This work is dedicated to our mentor and distinguished colleague Prof. Dr Dr h.c. mult. Gerhard Bringmann.
‡ Electronic supplementary information (ESI) available. See DOI: 10.1039/c5cc00843c
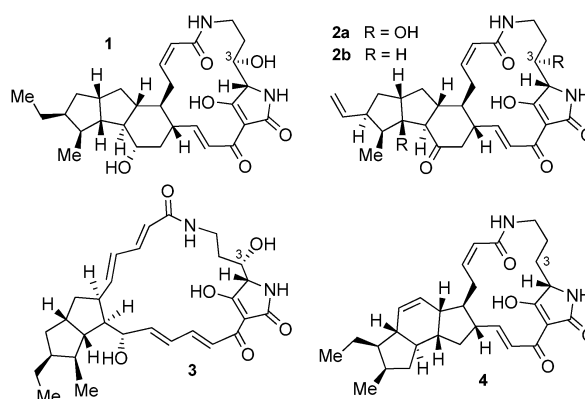§ These authors contributed equally to this work.

Fig. 1 Structures of the PTMs HSAF (**1**), the frontalamides (**2**), alteramide (**3**), and ikarugamycin (**4**).

the origin of the C-3 hydroxyl group found in many PTMs that was shown to be introduced into **1** by late-stage hydroxylation,[7d] the unprecedented iterative use of a single PKS module for incorporation of twelve C_2-units into a bacterial metabolite[7e,8,9] and the remarkably low number of enzymes overall involved in the biosynthesis of the highly complex cyclic structures by reductive ring formation, only three in case of **4**.[8,9] In addition, Zhao *et al.* showed the applicability of their DNA assembler methodology towards unlocking a cryptic PTM biosynthetic pathway from *Streptomyces griseus*.[10]

Owing to the broad range of biological activities of different PTMs and the simplicity of PTM biosynthetic pathways in terms of the low number of tailoring enzymes needed to shape these complex molecules, the engineered biosynthesis of PTMs by swapping PTM genes between different pathways would be highly desirable. However, such an approach has not yet been reported. The success of such experiments is highly dependent on the level of substrate promiscuity of the employed enzyme. Within this work, we set out to test this by utilizing PTM-pathway derived sterol desaturases (SD), the catalysts responsible for C-3 hydroxylation in the biosynthesis of HSAF (**1**),[7d] to functionalize

ikarugamycin (**4**), a compound with a significantly altered cyclization pattern when compared to **1** (*cf.* Fig. 1).

For our experiments we decided to use the SD gene of a bacterial strain with a putative PTM biosynthetic gene cluster resembling that of the HSAF pathway in *Lysobacter enzymogenes* C3.[7] The SD gene of the commercially available *Lysobacter capsici* DSM 19286 was selected as a good candidate (see Fig. 2A/B). The production of **1** by *L. capsici* was confirmed by its cultivation and chemical analysis of organic extracts of the culture broth. Having thus indirectly proven the presence of a biocatalytically active SD protein that hydroxylates the HSAF skeleton in *L. capsici*, we set out to clone the respective candidate gene into a suitable over-expression vector. To allow later purification of the recombinant SD protein, the vectors pHIS8[11] and pGS-21a[12] were used, both encoding N-terminal poly(His)-tags facilitating affinity chromatography, with pGS-21a additionally encoding a GST tag to enhance protein solubility. The putative SD gene of the PTM biosynthetic gene cluster in *L. capsici* was amplified from genomic DNA by PCR using two different sets of primers (see ESI‡). This led to the production of a 810 bp copy of the SD gene as predicted by GLIMMER[13] and to a longer 1002 bp copy that includes an alternative, more upstream start codon of the original sequence and better resembles the length of other annotated SDs in published genomic data. The thus obtained constructs were transferred into *E. coli* BL21(DE3) for recombinant protein production studies. Unfortunately, all expression experiments with the four constructs revealed only very low overproduction levels and exclusive formation of insoluble inclusion bodies. Best overproduction levels were still achieved with the 810 bp construct of SD in pGS-21a (pSDs*Lcap*).

We next examined the utility of pSDs*Lcap* to provide the hydroxylating SD biocatalyst *in vivo*. The plasmid was transferred into our recently reported *E. coli* BAP1[14] based ikarugamycin (**4**) heterologous production system[8] by electroporation. Both the transcription of the ikarugamycin (**4**) pathway genes as well as of the SD gene from *L. capsici* were controlled by *lacI*. Fermentation of the resulting strain in ZY autoinduction media[15] revealed the exclusive production of the 3-hydroxylated derivative of **4** (see Fig. 3D), a compound very recently isolated as the natural product butremycin (**5**) from *Micromonospora* sp. K310.[16] Ikarugamycin (**4**) was not detectable in these experiments at all, revealing its highly effective, complete conversion to **5** catalysed by SD *in vivo*.[17] Because of the identical chromatographic behaviour of **5** produced by our approach compared to **5**
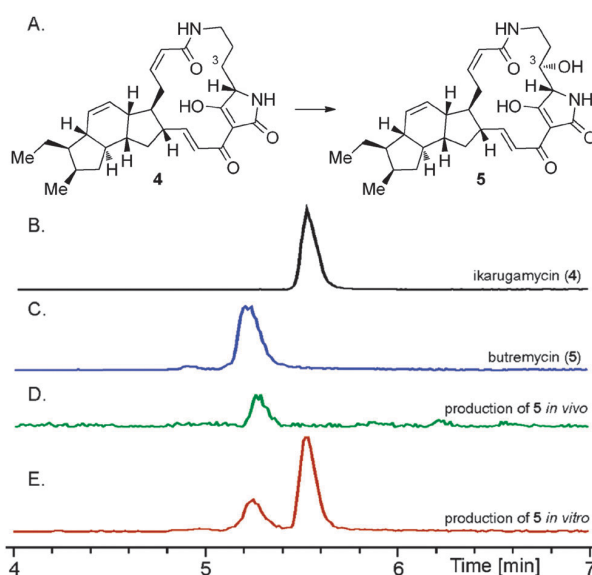

Fig. 3 (A) Conversion of ikarugamycin (**4**) to butremycin (**5**). HPLC-MS-analyses of (B) a standard of ikarugamycin (**4**), (C) a standard of butremycin (**5**), (D) the coexpression of Ika_ABC[8] in *E. coli* BAP1[14] together with SD from *L. capsici*, (E) the *in vitro* transformation of **4** using the supernatant derived from a culture of an *E. coli* BL21(DE3) SD (*S. degradans*) overexpression strain. EICs at 479.3 (**4**) and 495.2 (**5**) are shown.

from *Micromonospora* sp., formation of the identical diastereomer, *i.e.*, with 3*S*-configuration, can be concluded.[18]

We next turned our attention to the possibility to hydroxyl-ate purified ikarugamycin (**4**) using recombinant SD, as this methodology may easily be expanded to functionalize other PTM molecules without the need of a complete, heterologously reconstituted PTM pathway or a genetically transformable natural producer. Indeed, when feeding **4** to *E. coli* BL21(DE3) carrying pSDs*Lcap*, a small portion of **4** was converted into **5** (with approx. 90% of **4** remaining, data not shown).

To improve turnover rates, which might have been hampered in the *E. coli* feeding experiments due to problems of **4** passing the bacterial cell wall, the *in vitro* modification of **4** was re-investigated. As the SD encoded by *L. capsici* did not yield sufficient quantities of soluble protein for purification (see above), the SD gene of the putative PTM biosynthetic gene cluster of *Saccharophagus degradans* DSM 17024, a close homolog of the respective *Lysobacter* genes, was selected as an alternative (*cf.* Fig. 2C). It was readily cloned into pHIS8 and pGS-21a (see ESI‡) and the resulting constructs tested in protein overproduction assays in *E. coli* BL21(DE3). Using the respective pGS-21a construct (pSD*Sde*), overproduction of soluble protein was indeed achieved. Purification of the recombinant SD was conducted using Ni-affinity chromatography. However, significant degradation of SD during the isolation procedure was observed. The SD enriched fraction was nevertheless assayed for its hydroxylating activity. In these experiments, **4** was incubated with SD for 12 hours, leading to approx. 30% conversion to **5** (Fig. 3E). The overall turnover was not increased when extending the reaction time, hinting at a complete degradation of SD under the assay conditions. To potentially circumvent
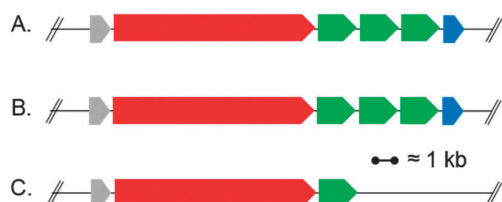

Fig. 2 Schematic representation of (A) the PTM biosynthetic pathway in *Lysobacter enzymogenes* C3 encoding HSAF (**1**)[7] and the putative PTM pathways in (B) *L. capsici* DSM 19286 and (C) *Saccharophagus degradans* DSM 17024. Genes putatively coding for: red: iterative PKS–NRPS system; green: oxidoreductase; blue: alcohol dehydrogenase; grey: sterol desaturase (SD).

the problems associated with decomposition of recombinant SD during purification, the *in vitro* hydroxylation of **4** using cell lysates of the *E. coli* BL21(DE3)::pSD*Sde* expression system was tested. These experiments gave the best results so far for the *in vitro* preparation of **5**, leading to overall approx. 50% C-3 hydroxylation of **4** to **5**. This hydroxylation step could be forced to completion by simple extractive work-up of the mixture of **4** and **5** and its repeated metabolization with fresh cell lysate.

The members of the PTM class of natural products are characterized by highly complex molecular structures that arise from remarkably streamlined biosynthetic machineries. Besides the unusual iterative PKS–NRPS system there are only a very small number of tailoring enzymes involved in PTM biosynthesis. Generating new molecules by adding or swapping individual biosynthetic genes between pathways thus seems to be a rewarding task. A precondition for success of such experiments, however, is a relaxed substrate specificity of the respective encoded enzymes. We herein prove for the first time that a class of PTM tailoring enzymes, usually annotated as sterol desaturases (SD), is capable of transforming "unnatural" substrates derived from a different PTM pathway. This led to the development of a heterologous production system of butremycin (**5**) utilizing the ikarugamycin biosynthetic genes together with a SD from *L. capsici*. In addition, even purified **4** can be converted into **5** utilizing SD alone, both *in vivo* and *in vitro*. As this class of SDs from PTM pathways seems to have highly specific activity for C-3 hydroxylation of PTMs, as also shown for the previously uncharacterized SD from *Saccharophagus degradans*, we propose to rename these enzymes according to their function as PTM hydroxylases. The results presented here provide diverse methods for late-stage C-3 hydroxylation of PTM molecules. The utility of the PTM hydroxylases for the functionalization of other substrates is currently being tested in our laboratory. Our work should furthermore set the stage for other PTM pathway gene swapping experiments that are likewise being investigated by our group.

Notes and references

1 P. R. Graupner, S. Thornburgh, J. T. Mathieson, E. L. Chapin, G. M. Kemmitt, J. M. Brown and C. E. Snipes, *J. Antibiot.*, 1997, **50**, 1014.
2 J. A. V. Blodgett, D.-C. Oh, S. Cao, C. R. Currie, R. Kolter and J. Clardy, *Proc. Natl. Acad. Sci. U. S. A.*, 2010, **107**, 11692.
3 H. Shigemori, M.-A. Bae, K. Yazawa, T. Saaki and J. Kobayashi, *J. Org. Chem.*, 1992, **57**, 4317.
4 K. Jomon, Y. Kuroda, M. Ajisaka and H. Sakai, *J. Antibiot.*, 1972, **25**, 271.
5 S. Cao, J. A. V. Blodgett and J. Clardy, *Org. Lett.*, 2010, **12**, 4652.
6 S. Kanazawa, N. Fusetani and S. Matsunaga, *Tetrahedron Lett.*, 1993, **34**, 1065.
7 (*a*) F. Yu, K. Zaleta-Rivera, X. Zhu, J. Huffman, J. C. Millet, S. D. Harris, G. Yuen, X.-C. Li and L. Du, *Antimicrob. Agents Chemother.*, 2007, **51**, 64; (*b*) L. Lou, G. Qian, Y. Xia, J. Hang, H. Chen, K. Zaleta-Rivera, Y. Li, Y. Shen, P. H. Dussault, F. Liu and L. Du, *J. Am. Chem. Soc.*, 2011, **133**, 643; (*c*) L. Lou, H. Chen, R. L. Cerny, Y. Li, Y. Shen and L. Du, *Biochemistry*, 2012, **51**, 4; (*d*) Y. Li, J. Huffman, Y. Li, L. Du and Y. Shen, *MedChemComm*, 2012, **3**, 982; (*e*) Y. Li, H. Chen, Y. Ding, Y. Xie, H. Wang, R. L. Cerny, Y. Shen and L. Du, *Angew. Chem., Int. Ed.*, 2014, **53**, 7524.
8 J. Antosch, F. Schaefers and T. A. M. Gulder, *Angew. Chem., Int. Ed.*, 2014, **53**, 3011.
9 G. Zhang, W. Zhang, Q. Zhang, T. Shi, L. Ma, Y. Zhu, S. Li, H. Zhang, Y.-L. Zhao, R. Shi and C. Zhang, *Angew. Chem., Int. Ed.*, 2014, **53**, 4840.
10 Y. Luo, H. Huang, J. Liang, M. Wang, L. Lu, Z. Shao, R. E. Cobb and H. Zhao, *Nat. Commun.*, 2013, **4**, 2894.
11 J. M. Jez, J. L. Ferrer, M. E. Bowman, R. A. Dixon and J. P. Noel, *Biochemistry*, 2000, **39**, 890.
12 The expression vector pGS-21a is commercially available from GenScript.
13 A. L. Delcher, K. A. Bratke, E. C. Powers and S. L. Salzberg, *Bioinformatics*, 2007, **23**, 673.
14 B. A. Pfeifer, S. J. Admiraal, H. Gramajo, D. E. Cane and C. Khosla, *Science*, 2001, **291**, 1790.
15 F. W. Studier, *Protein Expression Purif.*, 2005, **41**, 207.
16 K. Kyeremeh, K. S. Acquah, A. Sazak, W. Houssen, J. Tabudravu, H. Deng and M. Jaspers, *Mar. Drugs*, 2014, **12**, 999.
17 It is important to note that the heterologous ikarugamycin production system and our natural producer of **4**, *Streptomyces* sp. Tü 6239, do not produce any detectable amounts of butremycin (**5**).
18 The characterization of **5** by coelution experiments was performed on a variety of different HPLC phases (XBridge C18, Chromolith RP18e, Eurospher II C18A and C8, and the chiral phase Nucleocell Delta-RP), with no chromatographic differences detectable.

3.2. Biocatalytic total synthesis of ikarugamycin

Inspired by the heterologous production of ikarugamycin (**27**) using only three biosynthetic genes *in vivo*,[77] a novel *in vitro* approach for the biocatalytic total synthesis of **27** was to be developed. The ikarugamycin clusters of *Streptomyces* sp. Tü6239 and *Salinispora arenicola* CNS-205, both consisting of three genes, were available and coding for the needed enzymes. The cloning of the respective genes into *E. coli* suitable expression plasmids (pHis8, pGS-21a) and the purification of an enzymatic triplet IkaABC were the initial steps. The PoTeM *i*PKS/NRPS enzyme IkaA and the FAD-dependent oxidoreductase IkaB from *Streptomyces* sp. Tü6239 were successfully obtained using this approach, as well as the alcohol dehydrogenase IkaC from *S. arenicola*.

With these biocatalysts in hand, the successful production of **27** and the precursor **41** was achieved in a one-pot setup at room temperature overnight (Scheme 15). As substrates, acetyl-CoA (**9**), malonyl-CoA (**10**) and L-ornithine (**39**) were provided. By LC-MS analysis, the compounds **27** and **41** were identified. The retention time of **27** was in perfect agreement with a commercial standard. In addition, ^{1}H- and ^{13}C-NMR spectra of biocatalytically-produced **27** and an authentic standard, isolated from a designed *Streptomyces* strain in the Gulder lab, were recorded. Identical spectra were obtained validating the integrity of enzymatically produced **27**. The biocatalytic setup allowed to construct 15 carbon–carbon and two nitrogen–carbon bonds, generating eight stereogenic centers with perfect stereocontrol. Upscaling the reaction was possible, thereby beating the efficacy of existing organic-synthetic approaches.

In addition, regeneration of the CoA substrates **9** and **10** was established. The malonyl-CoA synthetase MatB from *Streptomyces coelicolor* M1152[83] was applied to produce **10** from malonate, ATP and CoA, and an enzymatic tandem consisting of AckA and Pta, both from *E. coli*, for **9** from acetate, respectively.[84–86]
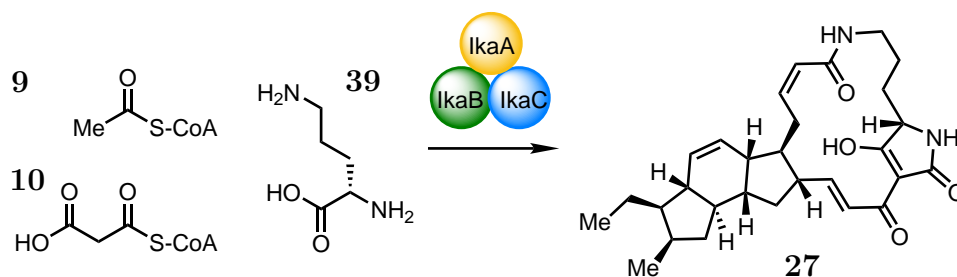
Scheme 15. Preparation of an enzymatic triplet IkaABC for the total synthesis of **27**.

The alcohol dehydrogenase IkaC from *S. arenicola* was able to substitute for insoluble IkaC from *Streptomyces* sp. Tü6239, and turned out to be active in the biocatalytic reactions. In this way, concurrently, one more demonstration was made showing the applicability of phylogenetically different PoTeM enzymes in one reaction.

This *in vitro* approach has the advantage of easily applying varying enzymes. Different substrates can now be tested, e.g., analogues of L-ornithine (**39**) without having to override the cellular pool of natural substrates. The biosynthetic intermediates of other pathways could be produced and compared, altogether allowing the monitoring of biosynthetic reaction steps. The established strategy has the big potential to apply uncharacterized PoTeM enzymes from diverse (putative) clusters, which have been cloned within the project (Table A.1). The *in vitro* setup for the production of **27** is thus also applicable to discover novel, yet unknown PoTeM compounds in the future.

Biocatalytic Total Synthesis of Ikarugamycin

Christian Greunke, Anna Glöckle, Janine Antosch, and Tobias A. M. Gulder*

Dedicated to Professor Bradley S. Moore on the occasion of his 50th birthday

Abstract: *Nature provides an inexhaustible diversity of small organic molecules with beautiful molecular architectures that have strong and selective inhibitory activities. However, this tremendous biomedical potential often remains inaccessible, as the structural complexity of natural products can render their synthetic preparation extremely challenging. This problem is addressable by harnessing the biocatalytic procedures evolved by nature. In this work, we present an enzymatic total synthesis of ikarugamycin. The use of an iterative PKS/NRPS machinery and two reductases has allowed the construction of 15 carbon–carbon and 2 carbon–nitrogen bonds in a biocatalytic one-pot reaction. By scaling-up this method we demonstrate the applicability of biocatalytic approaches for the ex vivo synthesis of complex natural products.*

*I*karugamycin (**1**) is the parent compound of the growing family of bacterial polycyclic tetramate macrolactams (PoTeMs).[1] It was discovered in 1972 from *Streptomyces* sp. by Jomon et al.[2] Since then, several other bacterial producers of **1** and structural analogues thereof have been identified, which predominantly possess antibacterial and antifungal properties.[3] Compound **1**, moreover, shows anti-leukemic[4] and anti-inflammatory[5] activity (Figure 1).
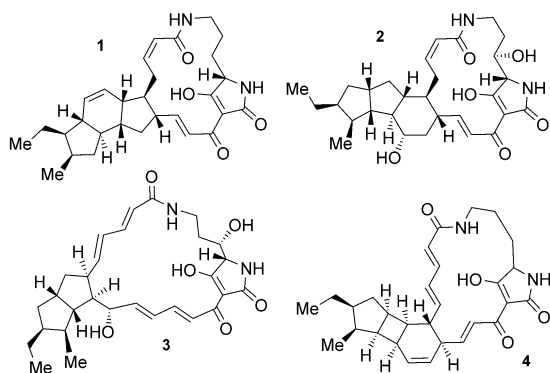
Figure 1. Selection of the structural diversity of PoTeMs: ikarugamycin (**1**), HSAF (**2**), alteramide A (**3**), and compound d (**4**).
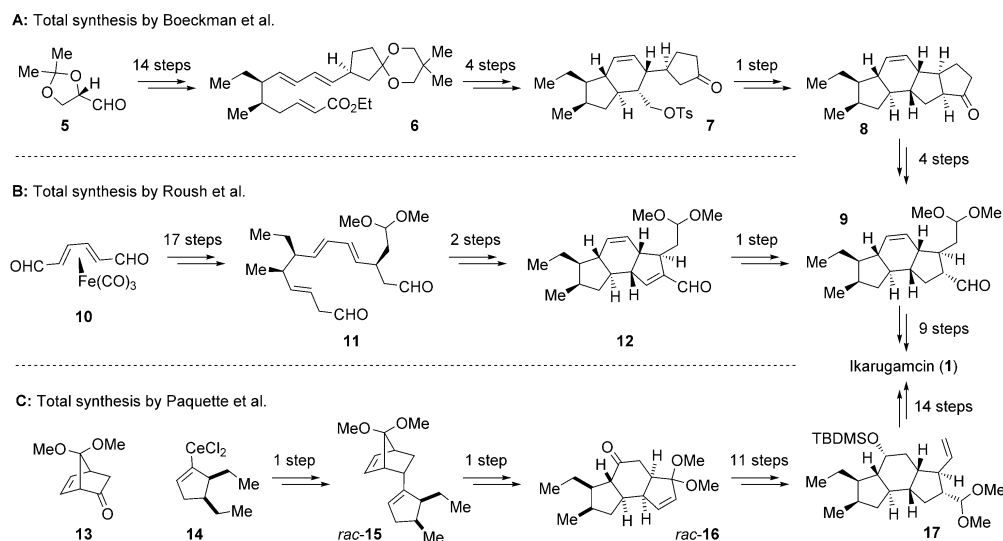
[*] C. Greunke, A. Glöckle, J. Antosch, Prof. Dr. T. A. M. Gulder
    Biosystems Chemistry, Department of Chemistry and Center for
    Integrated Protein Science Munich (CIPSM)
    Lichtenbergstrasse 4, 85748 Garching (Germany)
    E-mail: tobias.gulder@ch.tum.de

   Supporting information and the ORCID identification number(s) for
   the author(s) of this article can be found under:
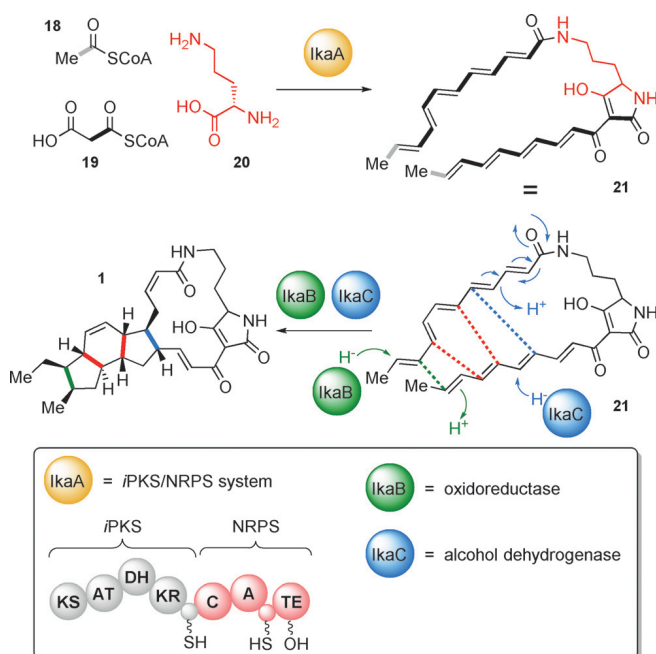   http://dx.doi.org/10.1002/anie.201611063.

Ikarugamycin (**1**) inhibits the uptake of oxidized low-density lipoprotein in macrophages[6] as well as HIV type 1 Nef-induced downregulation of the CD4 receptors on the cell surface.[7] Therefore, **1** is increasingly used as a tool to study endocytotic processes.[8] Ikarugamycin (**1**) is also of high interest from a purely structural point of view. As in all PoTeMs, the definitive structural features are a tetramic acid embedded in a macrolactam that is fused to a stereochemically demanding carbacyclic ring system. Variations in the cyclization pattern of PoTeMs give rise to the large diversity of their core structures, as exemplified by HSAF (**2**),[9] alteramide A (**3**),[10] and compound d (**4**).[11]

As a consequence of its promising biomedical properties and the complex structure, ikarugamycin (**1**) is an attractive, but highly challenging target for total synthesis. So far, three synthetic routes to **1** exist. Boeckman et al.[12] transformed acetonide **5** (via **6–9**) in 32 linear steps into the target molecule **1** (Scheme 1, route A). In a formal total synthesis of **1**, Roush and Wada established an improved route from **10** via **11** and **12** to Boeckman's intermediate **9** (Scheme 1, route B).[13] Both syntheses employ a Diels–Alder cycloaddition as the key step for the construction of the carbacyclic system. In the approach developed by Paquette et al.,[14] the construction of the *as*-indacene core commences by addition of the vinyl organocerium reagent **14** to bicyclic ketone **13** to give intermediate **15** (Scheme 1, route C). An oxy-Cope rearrangement affords the tricyclic ketone **16**, which is tranformed into the target molecule in 25 additional steps. These syntheses are impressive examples of the power of modern organic chemistry and the creativity of synthetic chemists. However, they also showcase that the construction of complex molecules often requires a large number of functional-group manipulations, such as protective-group operations or reduction/oxidation chemistry. In the case of the three total syntheses of **1**, this leads to an accumulation of 27–32 individual synthetic steps in the longest linear sequences, with the total yield of each method consequently being < 1%.

Investigations into the biosynthesis of ikarugamycin (**1**) have shown that the highly unusual iterative PKS (iPKS)/NRPS machinery IkaA catalyzes the formation of tetramic acid **21** from acetyl-CoA (**18**), malonyl-CoA (**19**), and L-ornithine (**20**; Scheme 2). The unstable biosynthetic intermediate **21** is cyclized to the natural product **1** by the action of the two reductases IkaB and IkaC.[15,16] These enzymes catalyze the formation of the outer (IkaB, green) and inner bond (IkaC, blue) of the carbacyle by formal addition of H_2. The central cyclohexenyl unit is likely formed by an additional, non-enzymatic Diels–Alder reaction (red). This

Scheme 1. Comparison of the total syntheses of **1** by A) Boeckman, B) Roush, and C) Paquette.

Scheme 2. Biosynthesis of ikarugamycin (**1**).

ing Information). Initial expression tests with *ikaC* in *E. coli* BL21(DE3), however, revealed the exclusive formation of insoluble inclusion bodies. To circumvent this problem we aimed to express homologous *ikaC* genes identified by BLAST homology searches, in analogy to previous work by Pahari et al.[19] This led to the discovery of a putative ikarugamycin biosynthetic gene cluster in *Salinispora arenicola* CNS-205. The ability of this cluster to produce **1** was verified by heterologous reconstitution using our previously established system in *E. coli* (data not shown).[15]
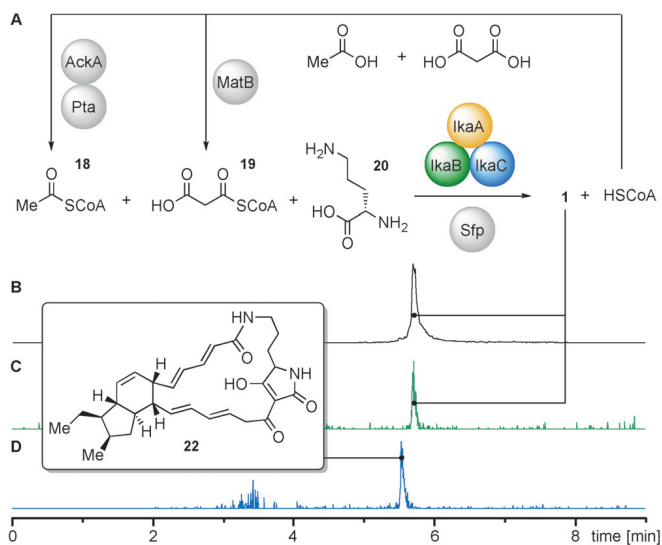
Recombinant expression of the *ikaC* homologue from *S. arenicola* [*ikaC(SA)*] cloned into pHis8 indeed provided soluble IkaC.

The isolation and purification of all recombinant proteins was performed by affinity chromatography on Ni-NTA beads followed by desalting on Sephadex G-25 and sample concentration with centrifugal concentrators. In addition to the biosynthetic enzymes IkaABC, the phosphopantetheinyl transferase (PPTase) Sfp from *Bacillus subtilis* was required and, thus, likewise produced in a recombinant form.[20] This promiscuous enzyme posttranslationally activates the active-site serine residue of PKS/NRPS carrier proteins (CPs) by attachment of a prosthetic phosphopantetheinyl group derived from coenzyme A.[21] As *E. coli* BL21(DE3) does not encode a homologous enzyme, catalytic amounts of Sfp needed to be added to the enzymatic reactions to obtain functional holo-CPs. Alternatively, IkaA was also accessible from *E. coli* BAP1. This strain possesses a chromosomal copy of Sfp[22] and, thus, facilitates the direct recombinant production of holo-IkaA (see lysate assays in the Supporting Information).

With all the required enzymes in hand, we approached the enzymatic total synthesis of **1** in the test tube (Scheme 3 A). Significant amounts of **1** were formed in reactions with purified samples of IkaABC and Sfp in HEPES buffer (pH 7.8) containing 1.5% glycerol as well as the required cofactors (ATP, NADPH, FAD, MgCl$_2$) and synthetic building blocks acetyl-CoA (**18**), malonyl-CoA (**19**), and L-ornithine (**20**; Scheme 3 C). Omission of IkaC in the enzymatic reactions led to the selective production of the known intermediate **22** with a bicyclic carbacyclic portion (Scheme 3 D).[15,16] The application of IkaA alone, however, did not result in the formation of an early biosynthetic intermediate (cf. with lysate assays in the Supporting Information).

To show the applicability of this method for the preparation of quantities of **1** comparable to those of the existing

extremely efficient biosynthetic route to **1**, with only three biocatalysts involved, is in sharp contrast to its high structural complexity and the resulting tedious total synthetic procedures. We were, therefore, interested in the development of a one-pot total synthesis of **1** that harnesses the catalytic power of the enzymatic triplet IkaABC.

For the production of the required biocatalysts, the genes *ikaA*, *ikaB*, and *ikaC* from the ikarugamycin producer *Streptomyces* sp. Tü6239[3a,15] were individually cloned into pHis8-[17] and pGS-21a-based[18] expression vectors. The recombinant production of IkaB and the large iPKS/NRPS system IkaA (ca. 340 kDa) was achieved by utilizing the native genes from *Streptomyces* sp. Tü6239 (see the Support-

Scheme 3. A) Developed method for the biocatalytic total synthesis of **1**. Colored circles: essential enzymes IkaA (orange), IkaB (green), and IkaC (blue). Gray circles: optional enzymes for activation of IkaA (Sfp) and for the in situ synthesis of **18** (AckA, Pta) and **19** (MatB). HPLC-MS analyses of the biocatalytic total synthesis of ikarugamycin (**1**): B) authentic standard of **1**; C) synthesis of **1** with IkaABC; D) reaction with IkaAB to give intermediate **22**.

total syntheses and to also facilitate characterization of the product by NMR spectroscopy, the enzymatic reaction was scaled to preparative amounts (3.7 mL reaction volume). After an overnight reaction, **1** was extracted from the aqueous phase and the raw material directly subjected to purification by preparative HPLC (see the Supporting Information). Ikarugamycin (**1**) was thus isolated in 9.0% yield (with respect to all employed building blocks **18**–**20**). Comparison of the NMR data of synthetic **1** with an authentic standard derived by isolation from an ikarugamycin-producing *Streptomyces* unambiguously validated the identity of the synthetic material (Figures S26 and S27).

Having established a first biocatalytic access to **1**, we next aimed at replacing the extremely expensive precursors **18** and **19** (both > 1000 € per 0.1 mmol). This was to be achieved by coupling the enzymatic triplet IkaABC with an in situ biocatalytic formation of **18** and **19** from acetic and malonic acid, respectively. For the synthesis of **18** we employed an enzymatic system consisting of the acetate kinase AckA and the phosphotransacetylase Pta, both from *E. coli*.[23] Formation of **19** was envisioned by utilizing the malonyl-CoA synthetase MatB from *Streptomyces coelicolor* M1152.[24] The respective genes were amplified by PCR from gDNA of *E. coli* and *S. coelicolor*, cloned into pHis8, and heterologously produced in *E. coli* BL21(DE3). The recombinant enzymes were isolated and purified as IkaABC. The catalytic activity of all the enzymes was initially validated in individual test assays (Figures S28 and S29). The addition of the three enzymes directly to IkaABC indeed led to an in situ biocatalytic supply of the precursors. This permitted the application of the expensive coenzyme A in catalytic amounts, as this thiol gets released again upon loading of the building blocks onto the iPKS/NRPS system. Overall, this

approach facilitated the total synthesis of **1** exclusively from cheap starting materials.[25]

A number of interesting metabolic pathways has already been reconstituted in vitro to study the interplay of enzymes in complex biosynthetic pathways.[26] Examples include the seminal studies on 6-deoxyerythronolide B,[27] tetracenomycin,[28] actinorhodin,[29] tetronate RK-682,[30] norsolorinic acid,[31] and dihydromonacolin L.[32] In the last decades, such studies that predominantly aim at the elucidation of complex biosynthetic mechanisms has led to spectacular insights into the chemical creativity of nature. However, the application of biosynthetic enzymes aimed at the biocatalytic total synthesis of polyketides is still very rare. The first landmark examples were the total syntheses of enterocin (**23**) by Moore and co-workers[33] and of defucogilvocarcin M (**24**) by Rohr and co-workers[34] (Figure 2). For the preparation of **23** and **24**, 12 and
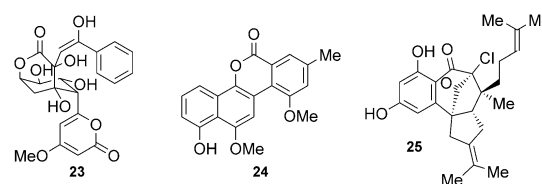
Figure 2. Complex polyketides prepared by biocatalytic total synthesis with detection by HPLC-MS: enterocin (**23**), defucogilvocarcin M (**24**), and merochlorin A (**25**).

15 individual biocatalysts, respectively, in a single reaction vessel were required. In particular, the large number of these, in some case delicate, biosynthetic enzymes that have to be produced, isolated, and employed in an active form makes the application of these impressive methods difficult. A first example of the use of a significantly simpler biocatalytic production system was very recently published by Moore and co-workers. They succeeded in the development of an enzymatic synthesis of the merochlorins A (**25**) and B by harnessing four recombinant enzymes of the *mcl* biosynthetic machinery.[35]

In the current study, we employ the three recombinant enzymes IkaABC in a biocatalytic total synthesis of ikarugamycin (**1**). Most of the synthetic work is performed by the bimodular iPKS/NRPS system IkaA. This enzyme stitches together two hexaketidic polyene precursors from acetyl- and malonyl-CoA, catalyzes their attachment to the N-termini of the L-ornithine loaded on the NRPS module, and the final cyclization of the tetramic acid moiety with concomitant product release. IkaA alone thus performs 13 individual, bond-forming reactions! The *as*-indacene structural element is formed by IkaA and IkaB, in combination with a Diels–Alder cycloaddition (see Scheme 2). These two reductive enzymes thereby selectively define all the stereocenters in the carbacyclic molecular portion of **1**. All in all, the three enzymes thus allow the construction of 15 carbon–carbon and 2 carbon–nitrogen bonds, with concomitant setting of 8 stereogenic centers.

In conclusion, the highly efficient biosynthetic pathway to **1** enables us to perform an enzymatic synthesis of this

structurally and stereochemically complex molecule with only three recombinant enzymes. This leads to a comparably simple biocatalytic reaction setup towards **1**: in a single pot, over night, at ambient temperatures. The expensive coenzyme-A-bound building block can, furthermore, optionally be supplied in situ starting from acetate, malonate, catalytic amounts of coenzyme A, and three additional recombinant enzymes.

The biosynthetic intermediate **21** produced by IkA is a common biosynthetic precursor of the PoTeMs. By exchanging enzymes catalyzing the cyclization of **21** by homologues from other biosynthetic pathways, novel PoTeMs should be accessible by the described biocatalytic system. These investigations are currently underway in our laboratory.

Conflict of interest

The authors declare no conflict of interest.

Keywords: biocatalysis · enzyme catalysis · ikarugamycin · natural products · total synthesis

[1] Polycyclic tetramate macrolactams have previously been abbreviated as "PTMs". As a consequence of the ambiguous nature of the term "PTM"—resulting from the much more common, identical abbreviation for posttranslational modifications—we herein propose to use "PoTeMs" instead.

[2] K. Jomon, Y. Kuroda, M. Ajisaka, H. Sasaki, *J. Antibiot.* **1972**, *25*, 271–280.

[3] a) M. Bertasso, M. Holzenkämpfer, A. Zeeck, E. Stackebrandt, W. Bei, H.-P. Fiedler, *J. Antibiot.* **2003**, *56*, 364–371; b) M. Pérez, C. Crespo, C. Schleissner, P. Rodríguez, P. Zúñiga, F. Reyes, *J. Nat. Prod.* **2009**, *72*, 2192–2194; c) C. Liu, X. Wang, J. Zhao, Q. Liu, L. Wang, X. Guan, H. He, W. Xiang, *Int. J. Syst. Evol. Microbiol.* **2013**, *63*, 3579–3584; d) K. Kyeremeh, K. S. Acquah, A. Sazak, W. Houssen, J. Tabudravu, H. Deng, M. Jaspars, *Mar. Drugs* **2014**, *12*, 999–1012; e) R. Lacret, D. Oves-Costales, C. Gómez, C. Díaz, M. de la Cruz, I. Pérez-Victoria, F. Vicente, O. Genilloud, F. Reyes, *Mar. Drugs* **2014**, *13*, 128–140; f) M. J. Xu, J. H. Wang, X. L. Bu, H. L. Yu, P. Li, H. Y. Ou, Y. He, F. D. Xu, X. Y. Hu, X. M. Zhu, P. Ao, J. Xu, *Sci. Rep.* **2016**, *6*, 18977.

[4] R. Popescu, E. H. Heiss, F. Ferk, A. Peschel, S. Knasmüller, V. M. Dirsch, G. Krupitza, B. Kopp, *Mutat. Res.* **2011**, *709–710*, 60–66.

[5] B. Malcomson, H. Wilson, E. Veglia, G. Thillaiyampalam, R. Barsden, S. Donegan, A. El Banna, J. S. Elborn, M. Ennis, C. Kelly, S. D. Zhang, B. C. Schock, *Proc. Natl. Acad. Sci. USA* **2016**, *113*, E3725–E3734.

[6] K. Hasumi, C. Shinohara, S. Naganuma, A. Endo, *Eur. J. Biochem.* **1992**, *205*, 841–846.

[7] T. Luo, B. L. Fredericksen, K. Hasumi, A. Endo, J. V. Garcia, *J. Virol.* **2001**, *75*, 2488–2492.

[8] a) A. Moscatelli, F. Ciampolini, S. Rodighiero, E. Onelli, M. Cresti, N. Santo, A. Idilli, *J. Cell Sci.* **2007**, *120*, 3804–3819; b) E. Onelli, C. Prescianotto-Baschong, M. Caccianiga, A. Moscatelli, *J. Exp. Bot.* **2008**, *59*, 3051–3068; c) V. Bandmann, U. Homann, *Plant J.* **2012**, *70*, 578–584; d) V. Bandmann, J. D. Müller, T. Köhler, U. Homann, *FEBS Lett.* **2012**, *586*, 3626–3632; e) H. Li, X. Ye, X. Guo, Z. Geng, G. Wang, *J. Hazard. Mater.* **2016**, *314*, 188–196; f) S. R. Elkin, N. W. Oswald, D. K. Reed, M. Mettlen, J. B. MacMillan, S. L. Schmid, *Traffic* **2016**, *17*, 1139–1149.

[9] P. R. Graupner, S. Thornburgh, J. T. Mathieson, E. L. Chapin, G. M. Kemmitt, J. M. Brown, C. E. Snipes, *J. Antibiot.* **1997**, *50*, 1014–1019.

[10] H. Shigemori, M.-A. Bae, K. Yazawa, T. Sasaki, J. Kobayashi, *J. Org. Chem.* **1992**, *57*, 4317–4320.

[11] Y. Luo, H. Huang, J. Liang, M. Wang, L. Lu, Z. Shao, R. E. Cobb, H. Zhao, *Nat. Commun.* **2013**, *4*, 2894.

[12] a) R. K. Boeckman, J. J. Napier, E. W. Thomas, R. I. Sato, *J. Org. Chem.* **1983**, *48*, 4152–4154; b) R. K. Boeckman, C. H. Weidner, R. B. Perni, J. J. Napier, *J. Am. Chem. Soc.* **1989**, *111*, 8036–8037.

[13] W. R. Roush, C. K. Wada, *J. Am. Chem. Soc.* **1994**, *116*, 2151–2152.

[14] a) L. A. Paquette, D. Macdonald, L. G. Anderson, J. A. Wright, *J. Am. Chem. Soc.* **1989**, *111*, 8037–8039; b) L. A. Paquette, J. L. Romine, H.-S. Lin, J. Wright, *J. Am. Chem. Soc.* **1990**, *112*, 9284–9292; c) L. A. Paquette, D. Macdonald, L. G. Anderson, *J. Am. Chem. Soc.* **1990**, *112*, 9292–9299.

[15] J. Antosch, F. Schaefers, T. A. M. Gulder, *Angew. Chem. Int. Ed.* **2014**, *53*, 3011–3014; *Angew. Chem.* **2014**, *126*, 3055–3058.

[16] G. Zhang, W. Zhang, Q. Zhang, T. Shi, L. Ma, Y. Zhu, S. Li, H. Zhang, Y. L. Zhao, R. Shi, C. Zhang, *Angew. Chem. Int. Ed.* **2014**, *53*, 4840–4844; *Angew. Chem.* **2014**, *126*, 4940–4944.

[17] J. M. Jez, J. L. Ferrer, M. E. Bowman, R. A. Dixon, J. P. Noel, *Biochemistry* **2000**, *39*, 890–902.

[18] The expression vector pGS-21a is commercially available from GenScript.

[19] P. Pahari, M. K. Kharel, M. D. Shepherd, S. G. van Lanen, J. Rohr, *Angew. Chem. Int. Ed.* **2012**, *51*, 1216–1220; *Angew. Chem.* **2012**, *124*, 1242–1246.

[20] L. E. N. Quadri, P. H. Weinreb, M. Lei, M. M. Nakano, P. Zuber, C. T. Walsh, *Biochemistry* **1998**, *37*, 1585–1595.

[21] For an excellent review on polyketide biosynthesis, see C. Hertweck, *Angew. Chem. Int. Ed.* **2009**, *48*, 4688–4716; *Angew. Chem.* **2009**, *121*, 4782–4811.

[22] B. A. Pfeifer, S. J. Admiraal, H. Gramajo, D. E. Cane, C. Khosla, *Science* **2001**, *291*, 1790–1792.

[23] M. C. Walker, B. W. Thuronyi, L. K. Charkoudian, B. Lowry, C. Khosla, M. C. Y. Chang, *Science* **2013**, *341*, 1089–1094.

[24] A. J. Hughes, A. Keatinge-Clay, *Chem. Biol.* **2011**, *18*, 165–176.

[25] The enzymatic production of acetyl- and malonyl-CoA in situ leads to a reduction in the total synthetic yield to about 25% compared with the respective reactions with stoichiometric amounts of **18** and **19**. This effect was previously observed when employing MatB for the production of malonyl-CoA in the enzymatic total synthesis of enterocin (see Ref. [33a]).

[26] a) A. A. Roberts, K. S. Ryan, B. S. Moore, T. A. M. Gulder, *Top. Curr. Chem.* **2010**, *297*, 149–203; b) E. S. Sattely, M. A. Fisch-

bach, C. T. Walsh, *Nat. Prod. Rep.* **2008**, *25*, 757–793; c) N. Tibrewal, Y. Tang, *Annu. Rev. Chem. Biomol. Eng.* **2014**, *5*, 347–366; d) B. Lowry, C. T. Walsh, C. Khosla, *Synlett* **2015**, *26*, 1008–1025.

[27] S. Kapur, B. Lowry, S. Yuzawa, S. Kenthirapalan, A. Y. Chen, D. E. Cane, C. Khosla, *Proc. Natl. Acad. Sci. USA* **2012**, *109*, 4110–4115.

[28] B. Shen, R. Hutchinson, *Science* **1993**, *262*, 1535–1540.

[29] C. W. Carreras, C. Khosla, *Biochemistry* **1998**, *37*, 2084–2088.

[30] Y. Sun, F. Hahn, Y. Demydchuk, J. Chettle, M. Tosin, H. Osada, P. F. Leadley, *Nat. Chem. Biol.* **2010**, *6*, 99–101.

[31] J. M. Crawford, P. M. Thomas, J. R. Scheerer, A. L. Vagstad, N. L. Kelleher, C. A. Townsend, *Science* **2008**, *320*, 243–246.

[32] S. Ma, J. W.-H. Li, J. W. Choi, H. Zhou, K. K. M. Lee, V. A. Moorthie, X. Xie, J. T. Kealey, N. A. Da Silva, J. C. Vederas, Y. Tang, *Science* **2009**, *326*, 589–592.

[33] a) Q. Cheng, L. Xiang, M. Izumikawa, D. Meluzzi, B. S. Moore, *Nat. Chem. Biol.* **2007**, *3*, 557–558; b) R. Teufel, A. Miyanaga, Q. Michaudel, F. Stull, G. Louie, J. P. Noel, P. S. Baran, B. Palfey, B. S. Moore, *Nature* **2013**, *503*, 552–556.

[34] M. K. Kharel, P. Pahari, H. Lian, J. Rohr, *Org. Lett.* **2010**, *12*, 2814–2817.

[35] a) R. Teufel, L. Kaysser, M. T. Villaume, S. Diethelm, M. K. Carbullido, P. S. Baran, B. S. Moore, *Angew. Chem. Int. Ed.* **2014**, *53*, 11019–11022; *Angew. Chem.* **2014**, *126*, 11199–11202; b) S. Diethelm, R. Teufel, L. Kaysser, B. S. Moore, *Angew. Chem. Int. Ed.* **2014**, *53*, 11023–11026; *Angew. Chem.* **2014**, *126*, 11203–11206.

3.3. Unpublished results

3.3.1. Synthetic PoTeMs by A domain mutagenesis

The carbon core of PoTeMs is formed by the *i*PKS/NPRS. Two hexaketidic precursors, which are produced from acetyl-CoA (**9**) and malonyl-CoA (**10**), are attached to L-ornithine (**39**) by formation of amide bonds at the α- and δ-amino groups. The A domain of the NPRS part selectively recruits L-ornithine (**39**). This is the basis for all yet known PoTeMs.

Reprogramming of the A domain of Tü6239-IkaA might in theory lead to the genesis of a novel series of molecules. The amino acid sequences of PoTeM A domains are highly conserved, as observed in a multiple sequence alignment with 18 different (partly putative) PoTeM gene clusters (Figure 7). By expansion of this alignment by 26 additional A domain sequences[87–93] with known substrate specificity for either L-lysine (**42**), L-ornithine (**39**) or hydroxyornithine, five amino acids were identified from statistical comparison that might potentially be relevant to alter the substrate specificity (Table 1 and Figure 7). All these considerations are based on the Stachelhaus code,[94] which is derived from the crystal structure of the gramicidin S synthetase 1 (GrsA).[95] It is used to assign substrate specificity based on a defined set of amino acids within A domains.[94]

Table 1. Overview of the A domain specificity code according to Stachelhaus.[94] Amino acids of the Tü6239-IkaA A domain, which were altered by site-directed mutagenesis, are shown in bold and italic.

	Pos 1	Pos 2	Pos 3	Pos 4	Pos 5	Pos 6	Pos 7	Pos 8
GrsA (Ref)[95] Phenylalanine	D	A	W	T	I	A	A	I
Wild type IkaA L-ornithine (**39**)	D	*V*	*G*	*E*	L	V	*S*	*I*
Mutant IkaA L-lysine (**42**)	D	*A*	*E*	*D*	L	V	*T*	*V*

For the production of novel PoTeMs, it was to be tested whether selective introduction of mutations might alter the substrate specificity from L-ornithine (**39**) to L-lysine (**42**) (Figure 8). To achieve this goal, the relevant partial DNA sequence of the A domain of Tü6239-*ikaA* was subcloned into pBluescript II SK(–) using a T4 DNA ligase-based cloning approach and restriction sites *Kpn*I and *Sac*I. On this basis, mutations were introduced in a sequential manner using PCR and primers incorporating the sequence

Figure 7. Aligment of sets of A domain amino acid sequences with GrsA (reference, yellow). The Stachelhaus amino acid positions marked in turquoise were compared. The PoTeM sequences (lines 27-44) are highly conserved and expected to incorporate L-ornithine (**39**). Sequences selective for L-lysine (**42**) (lines 3-16) and **39** (lines 17-26) were included. From the sequence comparison, amino acids were deduced that might alter the amino acid selection towards **42** for the case of Tü6239-lkaA (lines 2 and 27).

alterations (Scheme 16). Approaches used were SLIM[96] and HiFi DNA Assembly site-directed mutagenesis (NEB). Based on the amino acid sequence of Tü6239-IkaA (3124 AA), alterations were made at positions V2495A, G2498E (both SLIM) and E2537D, S2589T, I2597V (HiFi DNA Assembly) (Table A.3). For the codon changes, *E. coli* optimized triplets were chosen, as listed in the codon usage table of SnapGene (dataset: '*Escherichia coli* K12').

After validation of the sequence alterations by Sanger sequencing, the mutated A domain DNA fragment was PCR amplified from the pBluescript template and reinserted into *ikaA*. Thereby, the relevant part was replaced with the altered codons. For the generation of *C*-His8-TEV-Tü6239-PKS/NRPS_5mut, the mutated fragment has to replace the original sequence of the pHis8C-TEV construct. The resulting construct could be directly used for biocatalytic experiments, as described previously (Chapter 3.2). To heterologously produce the L-lysine-based analogue of **27**, an ikarugamycin cluster (*ikaABC*) bearing the mutations in the A domain of *ikaA* would be desirable. For this purpose, the plasmid pSET152-*ermE*::IKA (provided by A. Glöckle) was selected as it allows production of **27** in *Streptomyces* hosts in good yields and could potentially be used in the same way to produce the L-lysine-containing PoTeMs. The reintroduction was performed as follows: The mutated A domain fragment was inserted into pSET152-*ermE*::IKA in a three fragment HiFi DNA Assembly approach (Figure A.3). The backbone was divided into two parts of 8816 bp and 8677 bp which were PCR amplified using Q5 polymerase in combination with highGC Enhancer supplement. The same was done for the DNA fragment possessing the five specificity changing mutations (670 bp, 73 % GC). Restrictions analyses using *Mlu*I proved the presence of correctly assembled plasmids, as the codon change to introduce V2495A creates an additional cutting site. The resulting pSET152 plasmid was used as a template to amplify *ikaA_5mut* and subclone the PKS/NRPS gene (9375 bp) into pHis8C-TEV conventionally using restriction sites *Eco*RI and *Hin*dIII, as described previously (Chapter 3.2).

The inserts of the two plasmids, either for overexpression of mutated PKS/NRPS enzyme in *E. coli* BL21(DE3) (pHis8C-TEV) or for compound production in *Streptomyces* (pSET152-*ermE*) were sequenced to ensure that cloning was successfully achieved (Chapter A.7).

Production of mutant IkaA as a *C*-terminal octahistidyl-tagged fusion protein was successful in *E. coli* BL21(DE3) (Figure 9). Amounts and purity were equivalent compared to the wild type protein (Chapter 3.2). A total synthesis approach was conducted using the established enzymatic triplet IkaABC with the exchange with the generated mutant
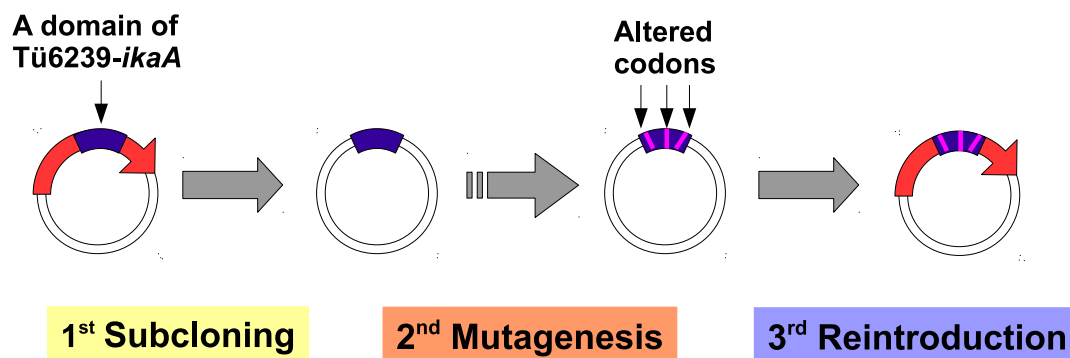
Scheme 16. Mutagenesis strategy for the alteration of the A domain sequence of Tü6239-*ikaA*. The portion of the A domain that should be mutated was subcloned into pBluescript. Afterwards, five selected codons were altered by site-directed mutagenesis. Sequence replacement of the respective DNA part was done by HiFi DNA Assembly.
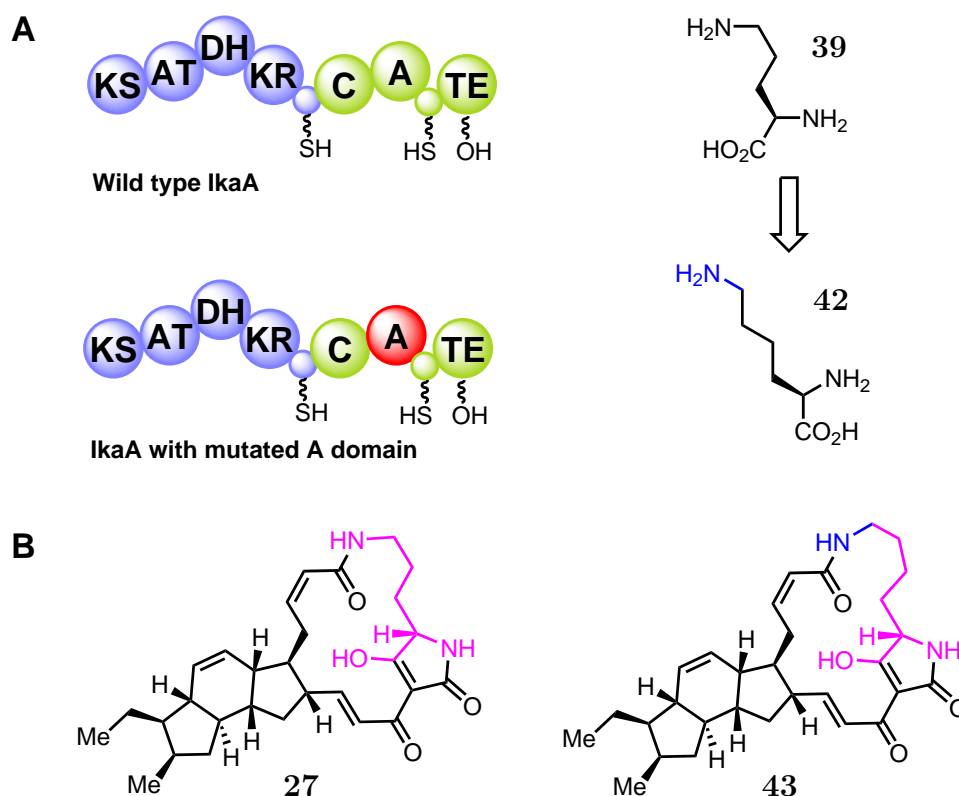
Figure 8. (A) By site-directed mutagenesis within the A domain of Tü6239-IkaA, substrate specificity should be altered from L-ornithine (**39**) to L-lysine (**42**). As a result, a class of novel synthetic PoTeMs is in theory accessible. (B) Structures of ikarugamycin (**27**) and an analogue **43** based on **42**.

PKS/NRPS. In addition, L-ornithine (**39**) was replaced with L-lysine (**42**) to test for incorporation. A reaction product with an expected mass of $m/z = 493.30$ ([M+H]$^+$) should be obtained in theory. Unfortunately, for biocatalytic reaction batches, no corresponding mass peak was found during analysis by LC-MS. Next, *in vivo* production using a *Streptomyces* sp. was tested. *Streptomyces* spores were used for conjugation to stably-integrate the mutated *ika* cluster. Extracts of supernatants as well as bacterial pellets did not contain any mass fitting the theoretical L-lysine derivative **43**. Instead, traces of **27** were detectable (personal communications, Anna Glöckle).
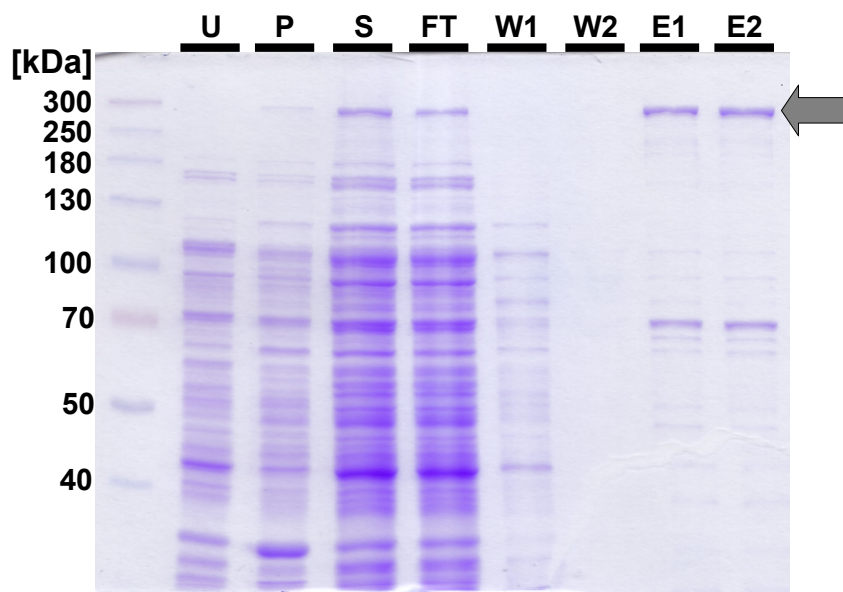
Figure 9. SDS-PAGE analysis of a purification attempt for *C*-His8-TEV-Tü6239-IkaA_5mut (340 kDa).

The data indicate that mutagenesis resulted in a weakly active enzyme that is still able to recruit **39**. Multiple explanations arise from this finding. The first is that the reprogramming might not be effective so that no substrate change towards **42** took place. Further amino acid substitutions (at different positions) within the A domain pocket would in this case be needed to achieve the goal of a reprogrammed enzyme. From a chemical point of view, a different situation arises when **42** is present. The elongation by a methylene unit might cause problems during downstream chemistry. The cyclization enzyme IkaB might not accept the elongated substrate. So far, it is difficult to detect early intermediates in PoTeM biosynthesis and it can thus not be clearly elucidated at which step the formation of expected **43** stops. The A domain sequences of PoTeM-*i*PKS/NRPS enzymes and underlying mechanisms are highly conserved, potentially not tolerating the introduced structural changes.

The *in vivo* situation gives a hint that reprogramming might not be complete, as traces of **27** are found. However, it is important to ensure that **42** is accepted as the major substrate. This could be achieved applying biochemical assays using radioactively-labeled pyrophosphates, as done by Lou *et al.*[75] It has to be ensured that the redesigned enzyme is specific for the unnatural substrate. The NRPS module must recruit **42** and incorporate it. In the cell, the competition between **39** and **42** cannot be suppressed. The more information on A domains are obtained and L-lysine-specific modules are found, the exchange of the whole A domain will be another approach to test.

The aim of producing a L-lysine-based PoTeM analogues is charming, as novel series of synthetic molecules might be accessible. The biocatalytic PoTeM plug-and-play system would be further developed by relatively small changes and the door opened for unnatural PoTeMs with hopefully exciting biomedical and structural properties.

3.3.2. Crystallization attempts for SA-AlcD (IkaC)

For a deep understanding of the underlying molecular mechanism of PoTeM enzymes, a structural characterization of these biocatalysts would be desirable. So far, there is no crystal structure available for any PoTeM enzyme. As a set of biocatalysts was prepared for the total synthesis of **27** (Chapter 3.2), initial trials were conducted to contribute to this field.

From the purity and producible amounts, the alcohol dehydrogenase (IkaC) from *Salinispora arenicola* CNS-205 was selected as the most promising first candidate. SA-AlcD was produced as an *N*-terminal fusion protein in *E. coli* BL21(DE3) (constructs: pHis8/pHis8-TEV::SA-AlcD). The fermentation conditions followed standard protocols: A temperature of 16 °C and IPTG at a concentration of 0.5 mM led to successful production (Figure 10).[97]

FPLC-based attempts, including gel filtration approaches using a HiLoad 16/600 Superdex 200 pg column (GE Healthcare), were conducted for further purification. However, the protein purity did not improve by this procedure. MALDI peptide fragment analysis revealed the presence of minor impurities consisting of a 16S rRNA pseudouridine synthase, the 30S ribosomal protein S3, a Crp/Fnr family transcriptional regulator and a transcriptional repressor, all from *E. coli*. Only the molecular chaperone GroEL with a size of 60 kDa was removed by FPLC. Gel filtration procedures were thus skipped after these tests.

The Nickel-NTA purified alcohol dehydrogenase was desalted on Sephadex G-25 and stored in 20 mM Tris-HCl, pH 7.5. The hanging drop vapour diffusion method was conducted at 20 °C. Successful growth of cubic crystals was observed within three days
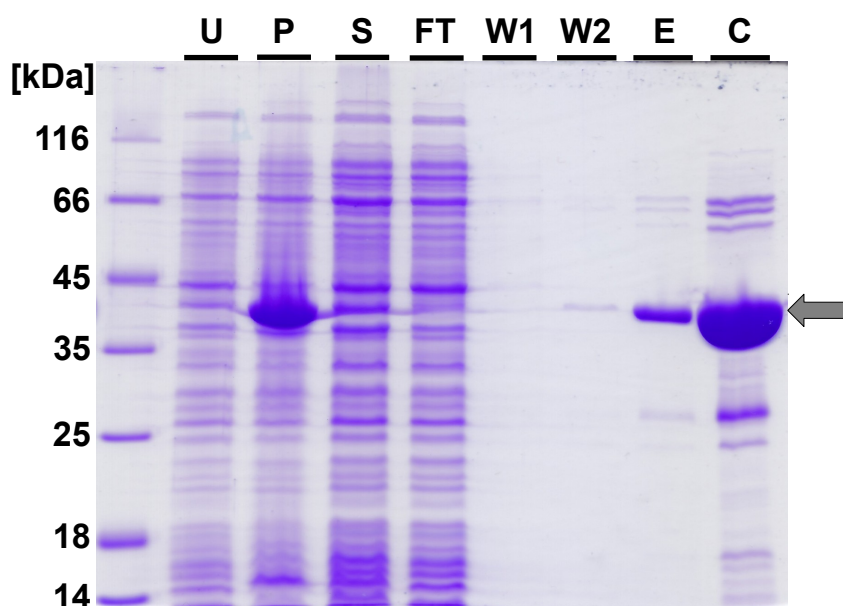
Figure 10. Production and purification of SA-AlcD. The enzyme was produced as an *N*-terminal His-tagged fusion protein in *E. coli* BL21(DE3) and stored in 20 mM Tris-HCl, pH 7.5 for crystallization experiments.

(Figure 11). Suitable buffers consisted of 1.0 to 1.2 M sodium citrate, pH 7.2 and 1.0 M succinic acid, pH 7.0, optionally containing 0.1 M HEPES, pH 7.0 and 0.5 % PEG 2000MME as additives. The protein content of the crystals was verified by their exposure to UV light at a wavelength of $\lambda = 280$ nm. Crystals were then sent for measurement to the synchrotron particle accelerator Swiss Light Source (SLS) (Paul Scherrer Institute, Villigen, Switzerland) to conduct X-ray diffraction experiments (TU Munich, Chair of Biochemistry, Prof. Dr. Michael Groll). Unfortunately, the measurements gave only a resolution of about 7 Å. The datasets were thus not yet suitable to determine the atom positions in the crystal. Further trials will be conducted as new crystals were prepared. If it is possible to obtain structural models of PoTeM enzymes, the catalytic mechanisms of the biosynthesis could be clarified in more detail, e.g., how cyclizations are controlled to establish the unique carbacyclic ring systems of PoTeMs. This would be a milestone in PoTeM research and a big contribution to the field.
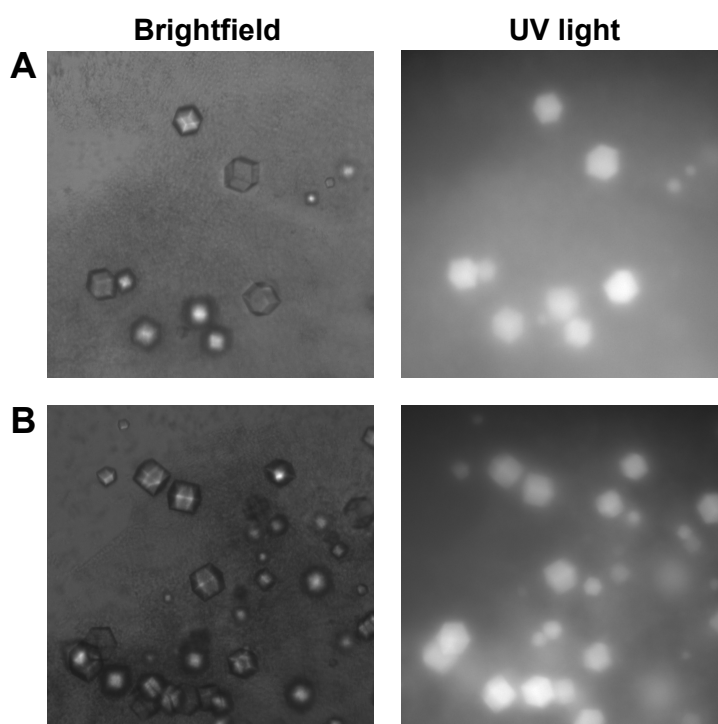
Figure 11. Crystallization attempts for SA-AlcD. Crystals were grown using the hanging drop method at 20 °C in buffers (A) 1.0 M sodium citrate, pH 7.2 and (B) 1.0 M succinic acid, 0.1 M HEPES and 0.5 % PEG 2000MME, pH 7.0. UV light ($\lambda = 280$ nm) was used to verify the protein content.

3.3.3. Purification trials for Tü6239-AlcD (IkaC) using *N*-lauroyl sarcosine

For the total synthesis of ikarugamycin (**27**), an enzymatic triplet IkaABC had to be provided. The corresponding enzymes from *Streptomyces* sp. Tü6239 should be heterologously produced in *E. coli*, which was successful for the *i*PKS/NRPS enzyme IkaA as well as for the FAD-dependent oxidoreductase IkaB. The alcohol dehydrogenase IkaC was not purifiable using standard laboratory protocols for native purification (Chapter 3.2). Production in *E. coli* resulted in an aggregation of insoluble protein after cell lysis. Tests using the detergent polyethylene glycol sorbitan monolaurate (known as Tween-20) did not improve the situation (Lab report, Michael Reinke). Expression tests in *Streptomyces lividans* TK24, as the heterologous host (expression construct pXY211::Tü-AlcD), did not help to solve the problem. Overproduction of completely insoluble protein was observed, too. Compared to the *E. coli* system, overproduction levels were less strong. The use of Nickel-NTA beads to enrich potential traces of soluble protein failed, proving exclusive localization in the inclusion bodies.

In comparison, the IkaC homolog from *Streptomyces* sp. ZJ306 described by Zhang *et al.* gave recombinant protein, which was purified from *E. coli* BL21(DE3).[79] In analogy, the enzyme from *Salinispora arenicola* CNS-205 was used in this work as a substitute in the biocatalytic total synthesis assays, although a massive aggregation in the pellet fraction was still observed. This problem seems to be common to most PoTeM accessory enzymes. From the available organisms in the Gulder lab possessing (putative) PoTeM BGCs, the accessory genes were cloned into a variety of expression plasmids (Chapter A.2). Test expressions of 14 FAD-dependent oxidoreductases and six alcohol dehydrogenases from a set of phylogenetically diverse bacteria in *E. coli* revealed insolubility of nearly all candidates. Sharing the characteristic to tend to the pellet fraction after cell lysis impeded their purification and functional study (Bachelor's Thesis, Florian Klein).

The phenomenon of inclusion body formation is well known.[98–100] It restricts direct access to the protein of interest, making purification a lot more challenging. Protein overproduction utilizing plasmids for the *E. coli* system containing strong promoters, like the commonly used T7 from bacteriophage,[101] can negatively influence the performance of the host to make soluble protein.[102] As a consequence, protein recovery from inclusion bodies is desired. One option can be the use of denaturing protocols applying chaotropic agents such as urea or guanidinium chloride.[103] These will unfold the protein completely, thereby making refolding necessary. It is possible that the protein of interest loses its catalytic activity during this process. Another strategy is the use of mild solubilization agents, e.g., alkaline conditions, high concentrations of *n*-propanol or β-mercaptoethanol, or high pressure, converting the misfolded protein from the inclusion bodies into folding intermediates, which need to be refolded at last.[103]

One promising approach, utilizing non-denaturing solubilization agents, was described by Peternel *et al.* using *N*-lauroyl sarcosine.[104] A significant amount of protein can thereby often be extracted from inclusion bodies while retaining the biochemical activity.[104] The underlying protocol was tested for the alcohol dehydrogenase (IkaC) from *Streptomyces* sp. Tü6239. After an overnight fermentation of *E. coli* BL21(DE3) pHis8-TEV::Tü6239-AlcD at 16 °C in the presence of IPTG in a concentration of 0.5 mM, the target protein was overproduced (Figure 12). After sonication, resuspension of the debris pellet in the Peternel buffer (50 mM Tris-HCl, 150 mM NaCl, 0.2 % *N*-lauroyl sarcosine, pH = 8.0) and incubation on ice with moderate shaking for at least six hours allowed solubilization of respectable amounts of target protein. Slight enrichment by His-tag affinity chromatography using High Density Nickel Agarose (Jena Bioscience) was possible. MALDI peptide fragment

analysis supported that the respective alcohol dehydrogenase is indeed solubilized, with sequence coverage matches from 33 to 47 % to the prediction.
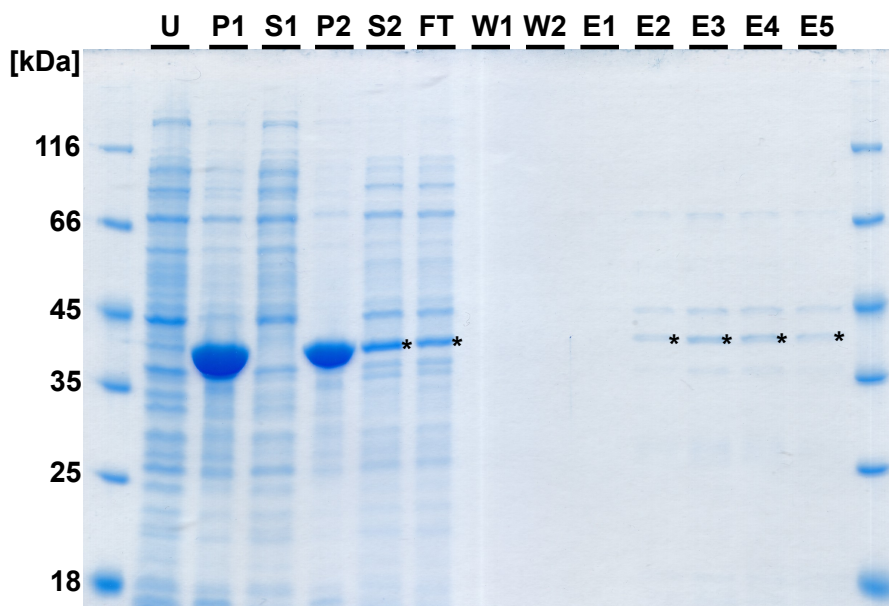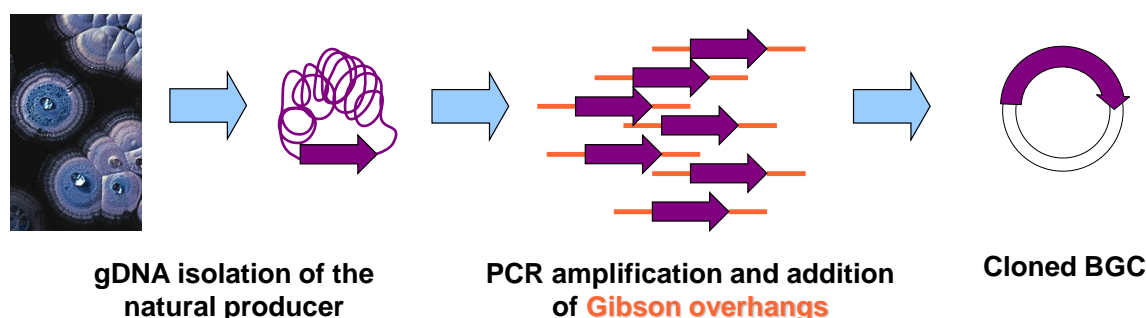

Figure 12. SDS-PAGE analysis of a purification attempt for *N*-His8-TEV-Tü6239-AlcD (Tü6239-IkaC, *) using *N*-lauroyl sarcosine. The fusion protein (40.0 kDa) was expressed in *E. coli* BL21(DE3) at 16 °C overnight.

This result is very promising as it is the first successful experiment allowing solubilization of Tü6239-IkaC. The tested approach might thus be useful to access all other PoTeM enzymes that have not been purifiable before. Furthermore, it should be feasible to improve the yields in the case of the FAD-dependent oxidoreductase (IkaB) from *Streptomyces* sp. Tü6239. Once confirmed that the biocatalysts are still possessing their activity, they can be used to explore their role *in vitro*. Novel PoTeM compounds could thus become accessible in the test tube.

3.4. Direct Pathway Cloning

A PCR-based method to access biosynthetic gene clusters (BGCs) has been developed in the course of this work and termed Direct Pathway Cloning (DiPaC). The basic idea is to use a polymerase to amplify the cluster region of interest only employing the (purified) gDNA of the natural producer (Scheme 17). As the gene clusters are the source for (novel) secondary metabolites, it is the key challenge to capture them in a rapid and highly reliable manner. Having the clusters in hand allows their heterologous expression to produce the encoded compounds.

gDNA isolation of the   **PCR amplification and addition**   **Cloned BGC**
natural producer   **of Gibson overhangs**

Scheme 17. The DiPaC method utilizes the power of high-fidelity DNA polymerases to amplify entire gene clusters (purple arrow) and reconstitute the generated DNA fragment(s) into any desired plasmid by Gibson Assembly *in vitro*. In this way, DiPaC promotes heterologous expression to test BGCs for the encoded molecules.

The DiPaC method consists of two steps: first, the polymerase-based amplification of the gene cluster of interest, and second, the insertion of the amplification product by Gibson Assembly.[105] Both processes exclusively require commercially available reagents, a high-fidelity DNA polymerase, such as Q5 polymerase, and the HiFi DNA Assembly Master Mix (both from NEB). One advantage of DiPaC is the minimal experimental setup: A PCR cycler and standard cloning reagents, such as competent *E. coli* bacteria. The only two things that have to be provided are the gDNA of the natural producer of interest and a set of suitable primers. Further advantages are that the complete cloning process takes place *in vitro* allowing to access toxic clusters that might kill the passaging host. Plasmids can be freely chosen, allowing to select any expression host of interest.

Compared to state of the art cloning methods for gene clusters that are based on homologous recombination, 'transformation assisted recombination in yeast' (TAR)[106,107] and 'linear-plus-linear homologous recombination' (LLHR),[108,109] and that require tremendous screening efforts, DiPaC efficiencies are in the same range as standard cloning procedures. Genomic libraries, an alternative to the previously described methods, have the same weakness. To cover a usual *Streptomyces* genome, approximately 1000 to 2000 *E. coli*

clones have to be individually picked and analyzed. In addition, it is often likely that only parts of the BGC of interest are obtained on a single library fosmid, making laborious efforts to join the relevant pieces of DNA necessary. The homologous recombination based methods were demonstrated to be capable of capturing huge gene clusters, 73 kb by TAR[107] and 106 kb by LLHR,[110] overcoming the size-limit of the pCC1FOS-based libraries, which is about 40 kb. A problem for homologous recombination might be repetitive sequences, negatively influencing the overall efficiencies. DiPaC directly allows to access the DNA regions of interest. Small gene clusters are amplifiable in only one PCR. Also, the Gibson Assembly driven process makes cluster refactoring possible, e.g., gene swapping or promoter addition/exchange and further combination of diverse genes and cluster parts, thereby constituting a powerful complementary method next to currently established ones, mentioned above.

The DiPaC process starts with the *in silico* simulation, involving primer design. Dependent on the size of the cluster, subdivision into several fragments is necessary, that are amplified in parallel. Computer simulation is crucial for the whole process. Using software like SnapGene or Geneious, the cloning strategy can be validated. This helps avoiding errors during primer design, e.g., multiple primer binding or too low annealing temperatures of the Gibson homology overhangs, to mention the most frequent ones. A sequential approach, meaning the insertion of one part after another, might be the most reliable way to reconstitute the cluster sequence. It is a conservative strategy, but the advantage is that intermediate plasmids harboring parts of the cluster can be isolated and analyzed. Concentration ranges, needed for an effective assembly reaction, are reachable more easily. In principle, a one pot assembly should also be applicable, as the HiFi DNA Assembly reaction is known to be suitable for up to eleven fragments. The fragment concentration is the limiting factor in the setup and becomes more challenging with increasing size as the molecular amounts will drop.

A successful PCR, the key challenge for DiPaC, highly depends on the amount of gDNA and its purity. Annealing temperatures for the primer pair might differ in the experiment. For this reason, the most suitable annealing temperature should be determined experimentally in the beginning. This does influence the amount of formed product and the formation of side-products. Clean-up of produced fragments is required to obtain a high efficiency in the HiFi DNA Assembly reaction. By the use of gel extraction, a loss of material has to be considered, making increased PCR batches necessary. Alternatively, the DNA fragments can be concentrated, e.g., by elution in small volumes of elution buffer.

A suitable vector system for DiPaC should be chosen in the beginning. For capturing and storage of the clusters, a broad set of plasmids is available. It is important to consider plasmid stability with increasing cluster size. Low copy plasmids might be more reliable, as these are less prone to incorporate mutations or to get rearranged or lost during cultivation. The vector system should be compatible to the expression host which again should be as closely related as possible to the native host. For example, as many secondary metabolites originate from Actinobacteria like *Streptomyces*, expression in a laboratory *Streptomyces* strain is for sure one of the most promising approaches.[111] There are a lot of other expression platforms known from literature, starting with *E. coli*,[77] which can be genetically manipulated quickly and has a fast growth rate. But also *Bacillus*,[112,113] *Pseudomonas*,[114,115] cyanobacteria[116] and yeast[117,118] have been used. Depending on the host, plasmid elements, like promoters, ribosomal binding sites, replication origins and transfer elements (for *Streptomyces* conjugation) have to fit the individual requirements. In the Gulder lab, a set of plasmids is available and comprising pET-28b-SUMO,[119] pCC1FOS[120] and pSET152.[121] Derivatives of these have been constructed, e.g., pSET152-*ermE* (Anna Glöckle) possessing the *ermE* promoter and a ribosomal binding site in front of the multiple cloning site. Recently, PtetO promoter[122] derivatives have been cloned, resulting in pCC1FOS-*ptetO* (Paul D'Agostino and Johanna Brüggenthies). Vector systems for cyanobacteria are being established (Elke Duell), thereby affording the opportunity to access this interesting bacterial phylum in the future.

In this thesis, DiPaC was applied to clone a set of PoTeM BGCs as well as the whole biosynthetic pathway leading to erythromycin A (**8**), as presented in the following.

3.4.1. Utilization of DiPaC for PoTeM gene clusters

Initial successful experiments to clone the full-length PKS/NRPS gene *ikaA* from *Streptomyces* sp. Tü6239 (9.4 kb) inspired trials to expand this approach for whole gene clusters. The PoTeM BGCs were ideal targets to start with, because of their relatively small sizes ranging from approximately 12 to 20 kb. By the use of Q5 DNA polymerase, it was possible to amplify and clone in total seven PoTeM clusters (Figure 13).

At first, the clusters coding for ikarugamycin (**27**) from *Streptomyces* sp. Tü6239 and *Salinispora arenicola* CNS-205 were cloned into pET-28b-SUMO. The respective sizes were about 12.3 kb (Tü6239) and 12.2 kb (*S. arenicola*), with corresponding GC contents of 72 and 69 %, respectively. The PCR template was available in two forms in the Gulder lab, as fosmid DNA[77] extracted from a genomic library and as gDNA, isolated from the

respective natural producers. In PCR reactions with Q5 polymerase and in combination with High GC Enhancer, it was possible to amplify both clusters from both templates.
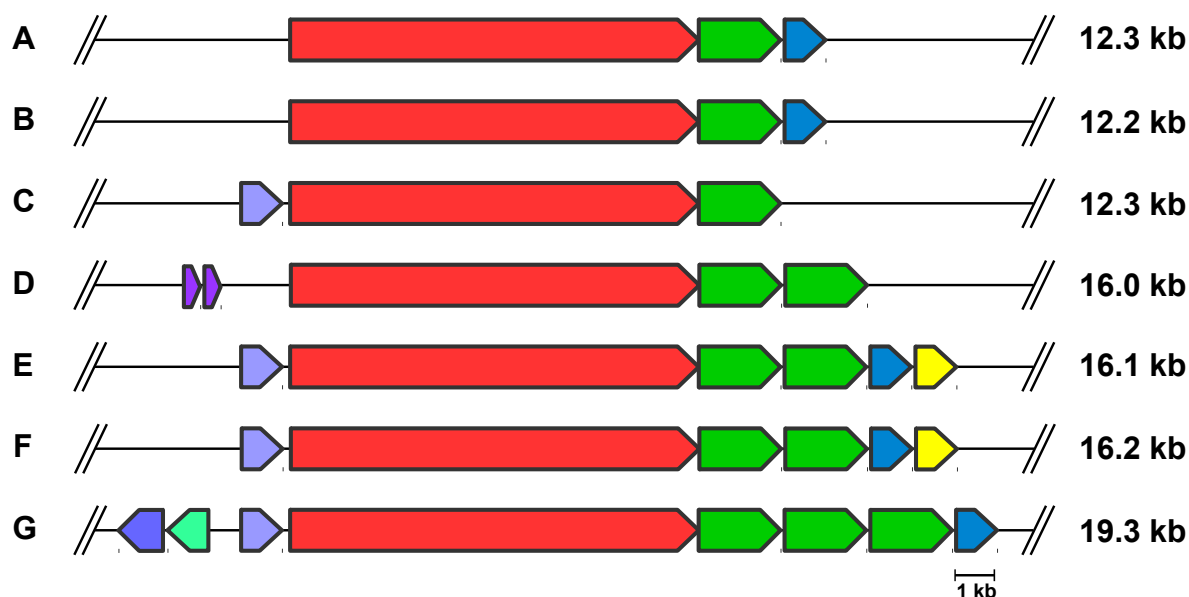
Figure 13. Schematic overview of PoTeM gene clusters that were captured using the DiPaC method: The ikarugamycin (**27**) BGCs from *Streptomyces* sp. Tü6239 (A) and *Salinispora arenicola* CNS-205 (B), putative clusters from *Saccharophagus degradans* DSM 17024 (C), *Saccharopolyspora spinosa* NRRL 18395 (D) and *Streptomyces* sp. Tü6314 (E), as well as from *Streptomyces griseus* DSM 40236 (F). The largest PoTeM cluster was from *Lysobacter capsici* DSM 19286 (G).

In addition to the BGCs coding for **27** from *Streptomyces* sp. Tü6239 and *S. arenicola*, putative clusters from *Saccharophagus degradans* DSM 17024 (12.3 kb, 47 % GC), *Saccharopolyspora spinosa* NRRL 18395 (16.0 kb, 69 % GC) and *Streptomyces* sp. Tü6314 (16.1 kb, 72 % GC), as well as from *Streptomyces griseus* DSM 40236 (16.2 kb, 72 % GC) were captured. The largest accessed PoTeM cluster, which was amplified in one PCR, was from *Lysobacter capsici* DSM 19286 (19.3 kb, 66 % GC) (Figure 13). It is similarly structured compared to the one from *L. enzymogenes* C3[75] and was proven to be responsible for HSAF (**29**) production (personal communications, Dr. Janine Antosch). The product spectra of the other clusters are currently under investigation in the Gulder lab.

3.4.2. DiPaC to access the erythromycin gene cluster

Small gene clusters, like the PoTeM BGCs, are quickly accessible by DiPaC. Most of them were amplified in one PCR reaction. As a model to extend the method to larger gene clusters, the erythromycin BGC *ery* from *Saccharopolyspora erythraea* DSM 40517 was

chosen. The size of the cluster is about 54.6 kb and has a GC content of 72.5 %.[44] It is comprised of 22 genes that are dividable into four groups. Three genes of about 10 kb, *eryAI-AIII*, are coding for the polyketide synthases that are needed for the formation of the macrolide aglycon core structure (Figure 14A). There are genes for oxidative modifications and methylations included, as well as for the production of the sugar building blocks dTDP-ʟ-mycarose and dTDP-ᴅ-desosamine.[45] The cluster includes a self-resistance gene (*ermE*) which is found in 3' proximity of the PKS genes.[45]

DiPaC should be used to clone the gene cluster *ery* in its wild type form. In bioinformatic preparation, the cluster was split into five parts and a sequential, five fragment assembly strategy was developed (Figure 14B). This included the use of unique restriction sites to open intermediate plasmids after each extension of the cluster to further insert PCR amplified fragments. The cutting sites were removed during HiFi-DNA Assembly, so that the wild type sequence of the cluster was maintained and the restriction site at the end of the fragments was kept unique. The *in silico* strategy was validated in the lab and *ery* was reconstructed according to the preparatory work. The plasmid backbone was derived from pET-28b-SUMO and cluster parts 1 to 4 were assembled successfully. Due to stability reasons in *E. coli*, a backbone exchange by homologous recombination (LCHR) was conducted for pCC1FOS to insert the fifth part needed to yield complete *ery*.

Homologous recombination was also used to transfer *ery* from pCC1FOS to pLK01.[120] pLK01 is a derivative of pCC1FOS and possesses elements for conjugational transfer. This allows stable cluster integration into the genomes of *Streptomyces*. Compatible homologous regions can be used to conduct LCHR for backbone exchange. The neomycin/kanamycin selection cassette is another difference compared to pCC1FOS (chloramphenicol), so that the backbones are separable during this process. After successful integration of *ery* into the genomes of *S. coelicolor* M1152 and M1154, these bacteria were able to produce erythromycin A (**8**) (Figure 14D and Figure A.4). Fermentation was conducted on mannitol soya flour agar[123] plates at 30 °C for one and a half week. The extracts of cultures of *Streptomyces coelicolor* conjugated with pLK01::*ery* were compared to a commercially-available standard and a control of unconjugated bacteria after HPLC-MS analysis (Figure 14C-E). In addition, MS^2 spectra were recorded, confirming the presence of **8** in the extracts of the designed *S. coelicolor* (Figure A.5).

This example is of importance as it shows that large gene clusters are accessible using DiPaC. A parallel PCR setup to amplify all cluster fragments at the same time accelerates the whole process and can be utilized to clone comparable clusters.
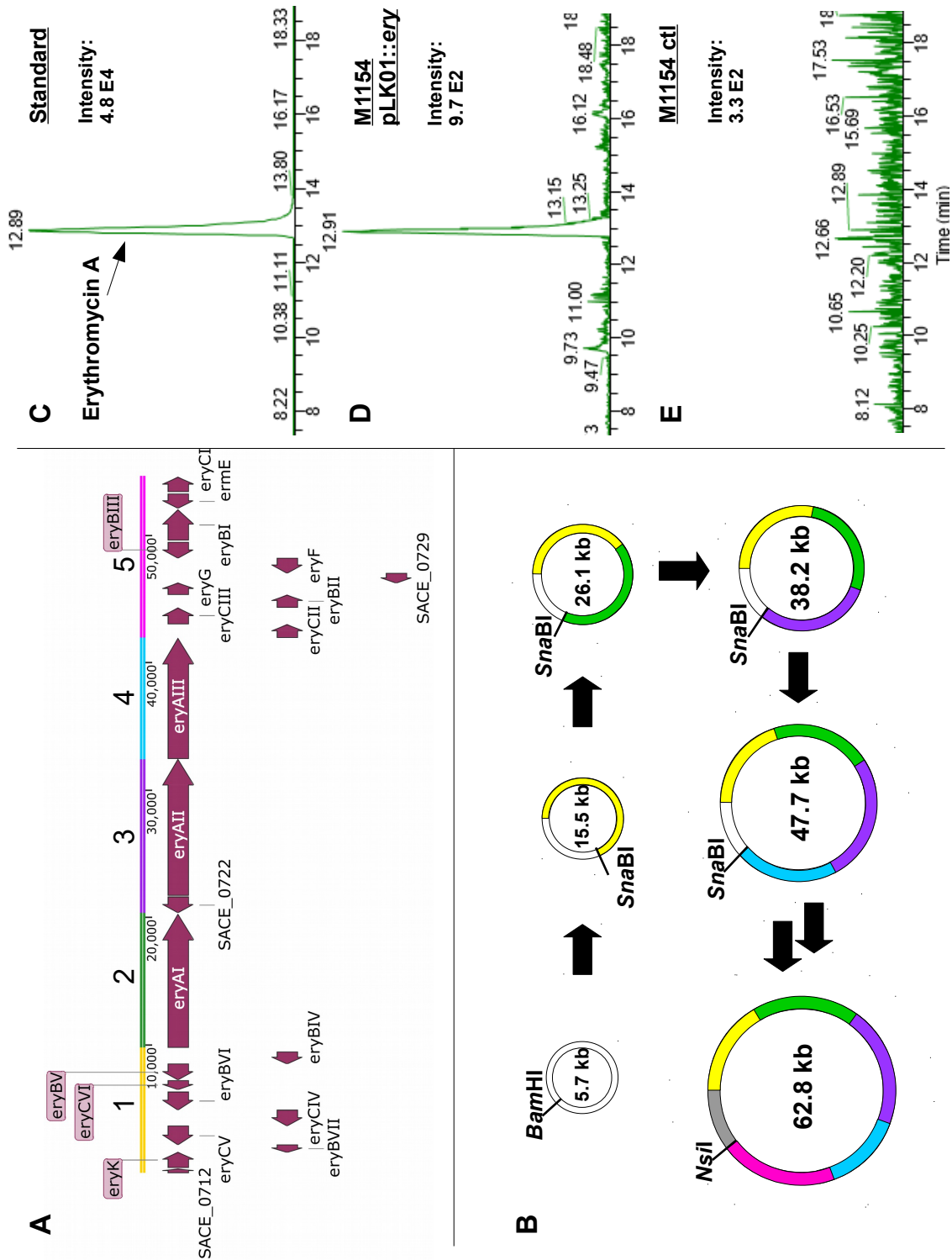
Figure 14. DiPaC strategy to access the erythromycin gene cluster *ery* (54.6 kb) in a sequential manner. (A) Map of the erythromycin gene cluster from *Saccharopolyspora erythraea* NRRL 2338 and (B) the assembly strategy using a five fragment approach in combination with unique restriction sites for backbone linearization. HPLC-MS analyses (mass trace m/z = 734.46 [M+H]$^+$) of an erythromycin A (**8**) standard (C), in comparison with heterologous expression derived **8** from *S. coelicolor* M1154 pLK01::*ery* (D) and an extract derived from a non-conjugated control culture (E).

4. Summary and Outlook

4.1. Exploration of the chemical space of PoTeMs

One future goal is to identify novel PoTeM clusters from newly sequenced bacteria by bioinformatics. Recent NCBI BLAST searches, targeting thioesterase sequences of PoTeM PKS/NPRS enzymes, lead to a collection of a total of 34 putative PoTeM BGCs, a majority of which has not been investigated yet (Chapter A.6). A lot of biosynthetic potential is thus still hidden and awaits its exploration. The ongoing sequencing of microbial genomes will promote the identification of further PoTeM and PoTeM-like gene clusters, encoding not only known, but also completely novel compounds. The biosynthetic potential looks promising as lots of exciting discoveries can be made in this field of research.

The growing knowledge of PoTeM gene clusters now starts to enable their engineering. A typical PoTeM-like gene sequence was recently identified in *Streptomyces griseus*,[124] but these Gram-positive bacteria do not produce any related compound under standard laboratory conditions.[125] The cluster was active when inserting promoters in front of each single gene.[125] After reintegration of the modified cluster into *Streptomyces lividans* 66, compounds A-C (**44**, **45**, **46**), possessing a 5-5 ring system were found (Figure 15).[125] In addition, compound D (**47**), which possesses a remarkable 5-6-4 carbacycle, was described.[125] It is explained to be a shunt product which is formed in a spontaneous Diels-Alder reaction when the second oxidoreductase (SGR813) is absent.[125] To date, no other PoTeM with a similar ring structure is known.

Recently, a new family of PoTeMs was discovered, the so-called pactamides (**48a**–**f**).[126] These compounds possess potent cytotoxic activities against human breast cancer cell lines.[126] The gene cluster from *Streptomyces pactum* SCSIO 02999 is silent and Saha *et al.* (Zhang lab) discovered these compounds by promoter engineering and subsequent heterologous expression in *S. lividans* TK64 (Figure 16).[126] The examples of Saha *et al.* and Luo *et al.* demonstrate the activation of silent PoTeM gene clusters.[125,126] From observations in the Gulder lab, it seems as if most of the orphan PoTeM gene clusters are in a silent state. Promoter engineering will be the key to discover the encoded molecules.

4. Summary and Outlook
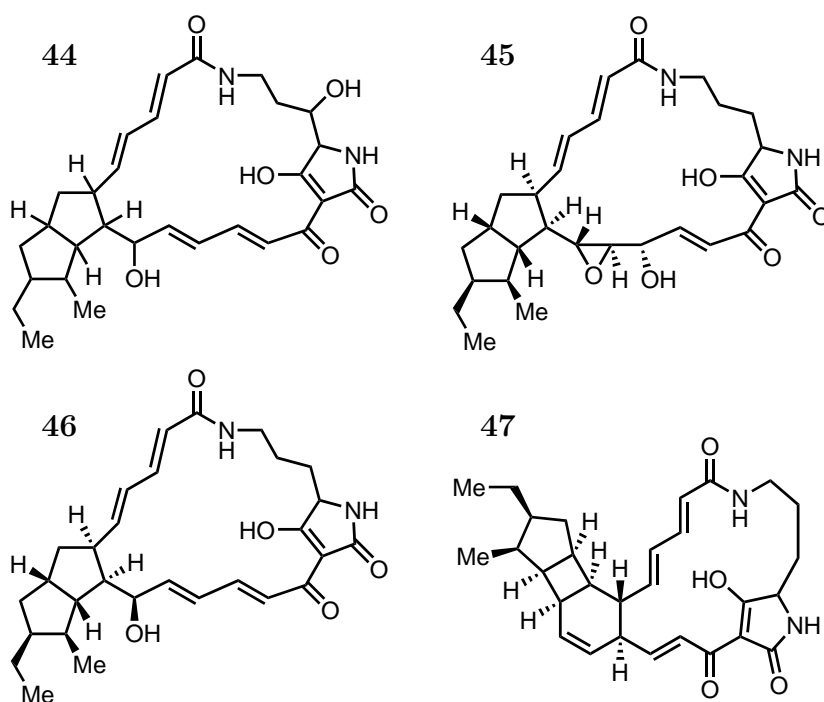
Figure 15. Structures of PoTeMs isolated from *S. griseus* **44**, **45**, **46** and **47**. Compound d (**47**) possesses an unusual 5-6-4 carbacyclic ring system.
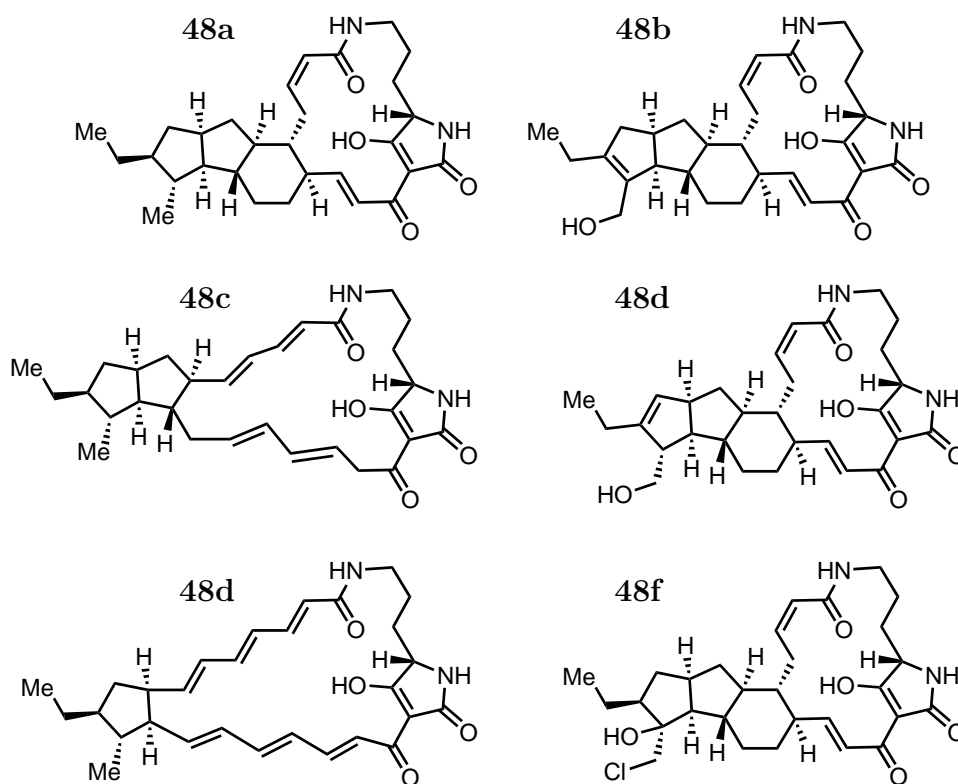
Figure 16. Structures of pactamides A-F (**48a–f**) of *S. pactum* SCSIO 02999.

In 2017, the isolation of novel capsimycins from *Streptomyces xiamenensis* 318 was reported.[127] These compounds are structurally similar to ikarugamycin (**27**), sharing the same carbon skeleton. Alterations in the oxidative state are observed at three carbon positions (Figure 17), on the double bond of the six-membered ring and on the ethyl-residue at the western part of the molecule. As expected, a three gene ikarugamycin-type cluster was identified together with a fourth gene *ikaD*, which was found eleven genes downstream of *ikaABC*. From a biosynthetic point of view, this finding is interesting as it shows that genes that are needed for PoTeM production do not need to be positioned next to the core genes.
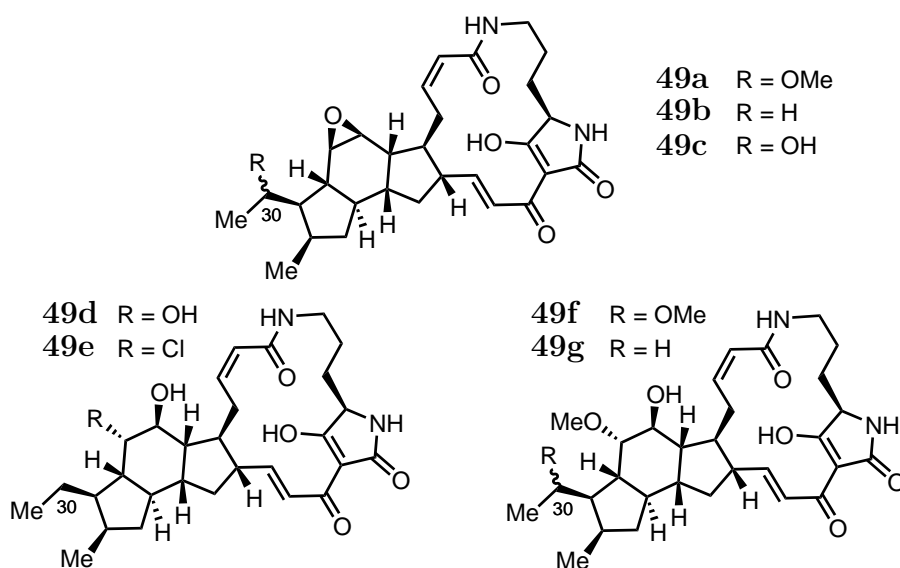
Figure 17.   Overview of the capsimycin family from *Streptomyces xiamenensis* 318.[127] Representatives containing a stable epoxide are capsimycin (**49a**), capsimycins B and G (**49b,c**). Monohydroxylated capsimycins C and D (**49d,e**), as well as chemically produced derivatives E and F (**49f,g**), were reported.

Interestingly, the cytochrome P450 monooxygenase IkaD can install a stable epoxide on the six-membered ring of the *as*-indacene system, functionalizing the double bond as found in capsimycin, capsimycin B and G (**49a**–**c**) (Figure 17).[127] The epoxide can be opened to form the dihydroxy compound capsimycin C (**49d**), and by addition of chlorine capsimycin D (**49e**). The methoxy-derived compounds capsimycins E and F (**49f,g**) were reported to be produced chemically by heating under acidic conditions in the presence of methanol from **49a** and **49b**. The members of the capsimycin family also vary in the oxidative state at position C-30, a hydroxy- and methoxy group can be present or hydrogen. In addition to the exciting structural flexibility, good bioactivities against a set of pancreatic cell lines were determined.[127] The capsimycins are a good example for PoTeM diversity.

4. Summary and Outlook

More such examples as the capsimycins[127] and pactamides[126] are expected to be reported in the future.

By using the DiPaC method, annotated clusters are quickly accessible to promote heterologous expression. In addition, it allows to conduct promoter additions and/or replacements by sequential assembly of cluster parts. The enigmatic process of heterologous expression will hopefully become a minor problem in PoTeM research, as an increasing number of examples will provide a basis to overcome this issue. The most exciting field of research will be the elucidation of the catalytic power of PoTeM accessory enzymes. They are encoded in the gene clusters next to the *i*PKS/NRPS gene. From the available clusters in the Gulder lab, which were cloned using DiPaC (Chapter 3.4.1), many genes were already subcloned (Table A.1). The FAD-dependent oxidoreductases will be the most exciting enzymes as they control the ring cycle formation. Changes to the structural frame work of the PoTeMs are directly derived from the mode of action of these biocatalysts.

Fermentation conditions have to be carefully chosen as well, as shown by the example of lysobacteramide B (**50**). **50** is thought to be a side product of HSAF (**29**) biosynthesis and was isolated from cultures of *L. enzymogenes* C3 grown in nutrient-rich yeast glycerol medium (Figure 18).[128] Unfortunately, no good explanation exists for the methyl-derivatization on the tetramic acid residue. *N*-methyl modifications have been described by Lacret *et al.*,[129] but a responsible gene of the natural host has not yet been discovered.
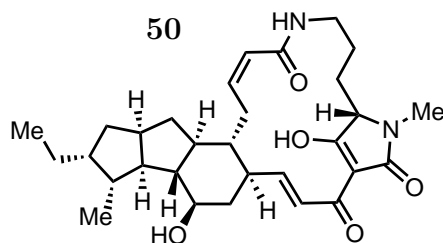
Figure 18. Structure of a newly identified HSAF pathway side product, lysobacteramide B (**50**), from *Lysobacter enzymogenes* C3.[128]

A contribution to further explore new PoTeM enzymes is the work on the *in vivo* and *in vitro* setups for the production of ikarugamycin (**27**) and butremycin (**28**) (Chapters 3.1 and 3.2). A future approach will be to use these as a platform to test unexplored genes. Exciting outcomes towards the generation of novel compound(s) are to be expected. The plug-and-play system in hand allows to test a set of different enzymes. This could not be started yet as most of the enzymes were found to be insoluble in *E. coli* lysates after heterologous expression. Once this problem is solved, exciting discoveries are in prospective. First experiments indicating a solution to overcome the solubility issue were

conducted in the course of this work, using *N*-lauroyl sarcosine according to the protocol of Peternel *et al.* (Chapter 3.3.3).[104]

The lists of PoTeM compounds known to date will undoubtedly further increase. A good classification of the natural product class can be made according to the carbacyclic ring patterns, i.e., if tri-/bi- or monocyclic frameworks are present (Table 2). It will be interesting to see how frequent each of the cyclization patterns will occur. Most of the structures isolated so far possess a tricyclic 5-5-6 or bicyclic 5-5 ring system. With a growing number of compounds, it is exciting to see whether they fit this scheme.

Table 2. PoTeMs listed according to the cyclization pattern of the carbacycles.

Ring pattern	PoTeM compound	Reference(s)
5-5-6	HSAF (Dihydromaltophilin) (**29**)	59–61,75
	Maltophilin (**32**)	64
	Frontalamides (**30a,b**)	62
	Lysobacteramide B (**50**)	128
	Discodermide (**36**)	68
	Xanthobaccin A (**35**)	67
	Pactamides A, B, D, F (**48a,b,d,f**)	126
5-6-5	Ikarugamycin (**27**)	57,77,79
	Butremycin (**28**)	58,82
	Capsimycins (**49a–g**)	127
	Clifednamides (**33a,b**)	65
5-4-6	Compound d (**47**)	125
5-5	Alteramides (**31a,b**)	63
	Aburatubolactam (**38**)	71–73
	Cylindramide (**37**)	69,70
	Geodin A (**34**)	66
	Compound a-c (**44,45,46**)	125
	Pactamide C (**48c**)	126
5	Pactamide E (**48e**)	126

Also, PoTeMs sharing (artificial) ring systems might exist. Compound d (**47**),[125] with its so far unique 5-4-6 pattern, might be an indication of yet undiscovered representatives.

4.2. Synthetic PoTeMs by introduction of novel substrates

As a L-ornithine (**39**) building block is present in all PoTeM backbones discovered so far, further experiments regarding extension of the substrate scope at this position will be promising. The total biocatalytic setup for the production of ikarugamycin (**27**) (Chapter 3.2) is an ideal system for this purpose, as it allows to test novel substrates without competition with the native subtrate pool, as occurring in the cellular setting in an *in vivo* experiment. The idea would be to substitute **39** with slightly-modified analogues that will probably be incorporated by the *i*PKS/NRPS enzyme. From theory, there are several possibilities to derivatize **39** (Figure 19). On carbon positions C-4 and C-5, a methyl group or short alkyl-residue could be attached. If accepted by the enzyme and further processed by the downstream biocatalysts, this would directly lead to a new series of PoTeM analogues.
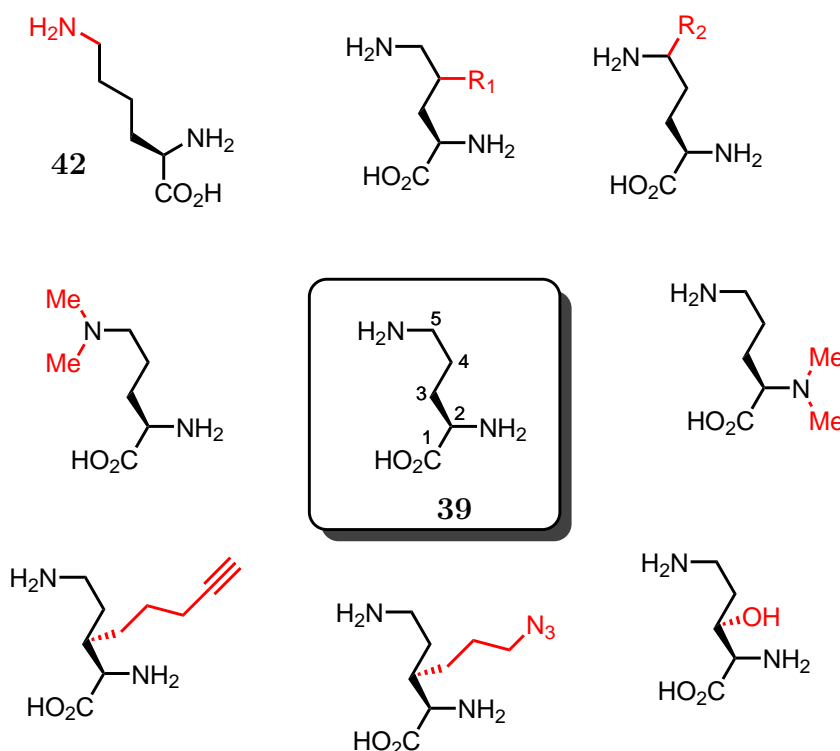
Figure 19. Overview of L-ornithine (**39**) related compounds. L-Lysine (**42**), as the most closely related naturally occurring amino acid, was tested in biocatalytic setups employing the enzymatic triplet IkaABC, but a related PoTeM analogue could not get generated. Besides **39** and **42**, all the structures proposed are awaited to be synthesized and tested in future studies.

There are many prominent examples for C-3-hydroxylated PoTeMs in nature, e.g., HSAF (**29**) or the frontalamides (**30a,b**). In this thesis, the application of PoTeM hydroxylases was described to make butremycin (**28**) from ikaruamycin (**27**) (Chapter 3.1). In theory, the same could be achieved by the use of C-3-hydroxylated derivatives of **39** to generate **28**. It is unclear whether the *i*PKS/NRPS enzyme would tolerate this building block, but if so, another elegant route can be explored to produce **28**. Modified building blocks might also allow to study the early steps in PoTeM biosynthesis, taking place on the *i*PKS/NRPS enzyme. As there are two amino groups present in **39** at α- and δ-position, selective dimethylations could be installed on either *N*-function as protecting groups. In this way, one amino group would be blocked so that the linkage reaction with the unsaturated polyene chain can only occur at the respective other position. In theory, a polyene-L-ornithine intermediate is to be expected. The attachment order of the polyenes is not proven yet. An attack on the δ-amino group at first and on the α-position at second is stated, but also the opposite is possible, and could be investigated by this experiment.[79,80] Concluding experimental data are not available yet.

Not much is known about the targets in the cell that are engaged by the PoTeM compounds. Ikarugamycin (**27**) is known to specifically inhibit clathrin-mediated endocytosis.[130] Azide or alkine linkers on **39** could be used as warheads for activity-based proteom profiling (ABPP) studies to find the target protein.[131] Modifications that covalently link a probe of **27** to the target protein upon a trigger event are the prerequisites and could be installed via a modified L-ornithine building block. Unfortunately, these compounds have to be synthesized *de novo* as no commercial source is available.

L-Lysine (**42**), as an alternative substrate, has been tested within this thesis, but it was not yet accepted as a building block. Total biocatalytic reactions with wild type and mutant Tü6239-*i*PKS/NRPS were conducted, but no expected analogue product was found. The A domain seems to be highly specific and does not accept **42** so far, even if **42** is provided exclusively in the *in vitro* setup, as the most promising substrate. One additional methylene unit might not be tolerated. The chemical properties, e.g., unfavorable tensions on the macrolactam ring or no recognition of the theoretical biosynthetic intermediates by the downstream enzymes, might be reasons impeding to access PoTeM series based on **42**. The field of substrate specificity has been completely disregarded so far. It would be of great interest to test novel substrates which could in theory substitute **39**. It will be exciting to explore how flexible the system is towards unnatural derivatives. In the best case, a series of novel PoTeM compounds will be generated. If successful, all efforts will be rewarded as this would be an outstanding example for combinatorial biochemistry.

Bibliography

(1) Cragg, G. M.; Newman, D. J. *Biochim. Biophys. Acta* **2013**, *1830*, 3670–3695, DOI: `10.1016/j.bbagen.2013.02.008`.

(2) Cragg, G. M.; Newman, D. J. *Pharm. Biol.* **2001**, *39 Suppl 1*, 8–17, DOI: `10.1076/phbi.39.s1.8.0009`.

(3) Fleming, A. *Bull. World Health Organ.* **2001**, *79*, 780–790.

(4) Libby, R. L.; Holmberg, N. L. *Science* **1945**, *102*, 303–304, DOI: `10.1126/science.102.2647.303`.

(5) Schäfer, B. *Chemie in unserer Zeit* **2017**, *51*, 238–253, DOI: `10.1002/ciuz.201700710`.

(6) Hubbard, B. K.; Walsh, C. T. *Angew. Chem. Int. Ed. Engl.* **2003**, *42*, 730–765, DOI: `10.1002/anie.200390202`.

(7) Goldstein, J. L.; Brown, M. S. *Nature* **1990**, *343*, 425–430, DOI: `10.1038/343425a0`.

(8) Biasucci, L. M.; Biasillo, G.; Stefanelli, A. *Clin. Chem. Lab. Med.* **2010**, *48*, 1685–1691, DOI: `10.1515/cclm.2010.277`.

(9) Zanin, V.; Marcuzzi, A.; Kleiner, G.; Piscianz, E.; Monasta, L.; Zacchigna, S.; Crovella, S.; Zauli, G. *J. Cardiovasc. Transl. Res.* **2013**, *6*, 981–988, DOI: `10.1007/s12265-013-9506-8`.

(10) Wani, M. C.; Taylor, H. L.; Wall, M. E.; Coggon, P.; McPhail, A. T. *J. Am. Chem. Soc.* **1971**, *93*, 2325–2327, DOI: `10.1021/ja00738a045`.

(11) Callaway, E.; Cyranoski, D. *Nature* **2015**, *526*, 174–175, DOI: `10.1038/nature.2015.18507`.

(12) Blackwood, R. K.; Beereboom, J. J.; Rennhard, H. H.; von Wittenau, M. S.; Stephens, C. R. *J. Am. Chem. Soc.* **1963**, *85*, 3943–3953, DOI: `10.1021/ja00907a010`.

(13) Newton, P. N.; Chaulet, J. F.; Brockman, A.; Chierakul, W.; Dondorp, A.; Ruangveerayuth, R.; Looareesuwan, S.; Mounier, C.; White, N. J. *Antimicrob. Agents Chemother.* **2005**, *49*, 1622–1625, DOI: `10.1128/AAC.49.4.1622-1625.2005`.

(14) Harvey, A. L.; Edrada-Ebel, R.; Quinn, R. J. *Nat Rev Drug Discov* **2015**, *14*, 111–129, DOI: `10.1038/nrd4510`.

Bibliography

(15) Larson, C. B.; Crüsemann, M.; Moore, B. S. *J. Nat. Prod.* **2017**, *80*, 1200–1204, DOI: `10.1021/acs.jnatprod.6b01121`.

(16) Bentley, S. D. *et al.* *Nature* **2002**, *417*, 141–147, DOI: `10.1038/417141a`.

(17) van Dijk, E. L.; Auger, H.; Jaszczyszyn, Y.; Thermes, C. *Trends Genet.* **2014**, *30*, 418–426, DOI: `10.1016/j.tig.2014.07.001`.

(18) Fleischmann, R. D.; Adams, M. D.; White, O.; Clayton, R. A.; Kirkness, E. F.; Kerlavage, A. R.; Bult, C. J.; Tomb, J. F.; Dougherty, B. A.; Merrick, J. M.; et al., *Science* **1995**, *269*, 496–512, DOI: `10.1126/science.7542800`.

(19) Land, M.; Hauser, L.; Jun, S. R.; Nookaew, I.; Leuze, M. R.; Ahn, T. H.; Karpinets, T.; Lund, O.; Kora, G.; Wassenaar, T.; Poudel, S.; Ussery, D. W. *Funct Integr Genomics* **2015**, *15*, 141–161, DOI: `10.1007/s10142-015-0433-4`.

(20) Gomez-Escribano, J. P.; Alt, S.; Bibb, M. J. *Mar. Drugs* **2016**, *14*, DOI: `10.3390/md14040078`.

(21) Weber, T.; Blin, K.; Duddela, S.; Krug, D.; Kim, H. U.; Bruccoleri, R.; Lee, S. Y.; Fischbach, M. A.; Müller, R.; Wohlleben, W.; Breitling, R.; Takano, E.; Medema, M. H. *Nucleic Acids Res.* **2015**, *43*, W237–243, DOI: `10.1093/nar/gkv437`.

(22) Skinnider, M. A.; Dejong, C. A.; Rees, P. N.; Johnston, C. W.; Li, H.; Webster, A. L.; Wyatt, M. A.; Magarvey, N. A. *Nucleic Acids Res.* **2015**, *43*, 9645–9662, DOI: `10.1093/nar/gkv1012`.

(23) Khaldi, N.; Seifuddin, F. T.; Turner, G.; Haft, D.; Nierman, W. C.; Wolfe, K. H.; Fedorova, N. D. *Fungal Genet. Biol.* **2010**, *47*, 736–741, DOI: `10.1016/j.fgb.2010.06.003`.

(24) Süssmuth, R. D.; Mainz, A. *Angew. Chem. Int. Ed. Engl.* **2017**, *56*, 3770–3821, DOI: `10.1002/anie.201609079`.

(25) Hertweck, C. *Angew. Chem. Int. Ed. Engl.* **2009**, *48*, 4688–4716, DOI: `10.1002/anie.200806121`.

(26) Wilson, M. C.; Moore, B. S. *Nat. Prod. Rep.* **2012**, *29*, 72–86, DOI: `10.1039/c1np00082a`.

(27) Ray, L.; Moore, B. S. *Nat. Prod. Rep.* **2016**, *33*, 150–161, DOI: `10.1039/c5np00112a`.

(28) Piel, J.; Hertweck, C.; Shipley, P. R.; Hunt, D. M.; Newman, M. S.; Moore, B. S. *Chem. Biol.* **2000**, *7*, 943–955, DOI: `10.1016/S1074-5521(00)00044-2`.

(29) Staunton, J.; Weissman, K. J. *Nat. Prod. Rep.* **2001**, *18*, 380–416, DOI: `10.1039/A909079G`.

(30) Walsh, C. T.; Gehring, A. M.; Weinreb, P. H.; Quadri, L. E.; Flugel, R. S. *Curr. Opin. Chem. Biol.* **1997**, *1*, 309–315, DOI: `10.1016/S1367-5931(97)80067-1`.

(31) Nakano, M. M.; Corbell, N.; Besson, J.; Zuber, P. *Mol. Gen. Genet.* **1992**, *232*, 313–321, DOI: `10.1007/BF00280011`.

(32) Quadri, L. E.; Weinreb, P. H.; Lei, M.; Nakano, M. M.; Zuber, P.; Walsh, C. T. *Biochemistry* **1998**, *37*, 1585–1595, DOI: `10.1021/bi9719861`.

(33) Fischbach, M. A.; Walsh, C. T. *Chem. Rev.* **2006**, *106*, 3468–3496, DOI: `10.1021/cr0503097`.

(34) Callahan, B.; Thattai, M.; Shraiman, B. I. *Proc. Natl. Acad. Sci. U. S. A.* **2009**, *106*, 19410–19415, DOI: `10.1073/pnas.0902364106`.

(35) Schröder, J. *Nat. Struct. Biol.* **1999**, *6*, 714–716, DOI: `10.1038/11472`.

(36) Funa, N.; Ohnishi, Y.; Ebizuka, Y.; Horinouchi, S. *J. Biol. Chem.* **2002**, *277*, 4628–4635, DOI: `10.1074/jbc.M110357200`.

(37) Funa, N.; Awakawa, T.; Horinouchi, S. *J. Biol. Chem.* **2007**, *282*, 14476–14481, DOI: `10.1074/jbc.M701239200`.

(38) Schümann, J.; Hertweck, C. *J. Biotechnol.* **2006**, *124*, 690–703, DOI: `10.1016/j.jbiotec.2006.03.046`.

(39) Cox, R. J. *Org. Biomol. Chem.* **2007**, *5*, 2010–2026, DOI: `10.1039/b704420h`.

(40) Campbell, C. D.; Vederas, J. C. *Biopolymers* **2010**, *93*, 755–763, DOI: `10.1002/bip.21428`.

(41) Staunton, J.; Wilkinson, B. *Chem. Rev.* **1997**, *97*, 2611–2630, DOI: `10.1021/cr9600316`.

(42) Mcguire, J. M.; Bunch, R. L.; Anderson, R. C.; Boaz, H. E.; Flynn, E. H.; Powell, H. M.; Smith, J. W. *Antibiot. Chemother.* **1952**, *2*, 281–283.

(43) Schönfeld, W.; Kirst, H. A. *Macrolide Antibiotics*, 1st ed.; Birkhäuser Verlag, Basel, 2002; DOI: `10.1007/978-3-0348-8105-0`.

(44) Oliynyk, M.; Samborskyy, M.; Lester, J. B.; Mironenko, T.; Scott, N.; Dickens, S.; Haydock, S. F.; Leadlay, P. F. *Nat. Biotechnol.* **2007**, *25*, 447–453, DOI: `10.1038/nbt1297`.

(45) Zhang, H.; Wang, Y.; Wu, J.; Skalina, K.; Pfeifer, B. A. *Chem. Biol.* **2010**, *17*, 1232–1240, DOI: `10.1016/j.chembiol.2010.09.013`.

(46) Jiang, M.; Fang, L.; Pfeifer, B. A. *Biotechnol. Prog.* **2013**, *29*, 862–869, DOI: `10.1002/btpr.1759`.

Bibliography

(47) Jiang, M.; Zhang, H.; Pfeifer, B. A. *J Vis Exp* **2013**, e4346, DOI: `10.3791/4346`.

(48) Zotchev, S. B.; Hutchinson, C. R. *Gene* **1995**, *156*, 101–106, DOI: `10.1016/0378-1119(95)00012-U`.

(49) Piel, J. *Nat. Prod. Rep.* **2010**, *27*, 996–1047, DOI: `10.1039/b816430b`.

(50) Helfrich, E. J.; Reiter, S.; Piel, J. *Curr. Opin. Biotechnol.* **2014**, *29*, 107–115, DOI: `10.1016/j.copbio.2014.03.004`.

(51) Bushley, K. E.; Raja, R.; Jaiswal, P.; Cumbie, J. S.; Nonogaki, M.; Boyd, A. E.; Owensby, C. A.; Knaus, B. J.; Elser, J.; Miller, D.; Di, Y.; McPhail, K. L.; Spatafora, J. W. *PLoS Genet* **2013**, *9*, e1003496, DOI: `10.1371/journal.pgen.1003496`.

(52) Walsh, C. T.; O'Brien, R. V.; Khosla, C. *Angew. Chem. Int. Ed. Engl.* **2013**, *52*, 7098–7124, DOI: `10.1002/anie.201208344`.

(53) Felnagle, E. A.; Jackson, E. E.; Chan, Y. A.; Podevels, A. M.; Berti, A. D.; McMahon, M. D.; Thomas, M. G. *Mol. Pharm.* **2008**, *5*, 191–211, DOI: `10.1021/mp700137g`.

(54) Qiao, K.; Chooi, Y. H.; Tang, Y. *Metab Eng* **2011**, *13*, 723–732, DOI: `10.1016/j.ymben.2011.09.008`.

(55) Trautman, E. P.; Healy, A. R.; Shine, E. E.; Herzon, S. B.; Crawford, J. M. *J. Am. Chem. Soc.* **2017**, DOI: `10.1021/jacs.7b00659`.

(56) Offenzeller, M.; Santer, G.; Totschnig, K.; Su, Z.; Moser, H.; Traber, R.; Schneider-Scherzer, E. *Biochemistry* **1996**, *35*, 8401–8412, DOI: `10.1021/bi960224n`.

(57) Jomon, K.; Kuroda, Y.; Ajisaka, M.; Sakai, H. *J. Antibiot. (Tokyo)* **1972**, *25*, 271–280, DOI: `10.7164/antibiotics.25.271`.

(58) Kyeremeh, K.; Acquah, K. S.; Sazak, A.; Houssen, W.; Tabudravu, J.; Deng, H.; Jaspars, M. *Mar. Drugs* **2014**, *12*, 999–1012, DOI: `10.3390/md12020999`.

(59) Graupner, P. R.; Thornburgh, S.; Mathieson, J. T.; Chapin, E. L.; Kemmitt, G. M.; Brown, J. M.; Snipes, C. E. *J. Antibiot. (Tokyo)* **1997**, *50*, 1014–1019, DOI: `10.7164/antibiotics.50.1014`.

(60) Yu, F.; Zaleta-Rivera, K.; Zhu, X.; Huffman, J.; Millet, J. C.; Harris, S. D.; Yuen, G.; Li, X. C.; Du, L. *Antimicrob. Agents Chemother.* **2007**, *51*, 64–72, DOI: `10.1128/AAC.00931-06`.

(61) Li, S.; Jochum, C. C.; Yu, F.; Zaleta-Rivera, K.; Du, L.; Harris, S. D.; Yuen, G. Y. *Phytopathology* **2008**, *98*, 695–701, DOI: `10.1094/PHYTO-98-6-0695`.

(62) Blodgett, J. A.; Oh, D. C.; Cao, S.; Currie, C. R.; Kolter, R.; Clardy, J. *Proc. Natl. Acad. Sci. U. S. A.* **2010**, *107*, 11692–11697, DOI: `10.1073/pnas.1001513107`.

(63) Shigemori, H.; Bae, M. A.; Yazawa, K.; Sasaki, T.; Kobayashi, J. *J. Org. Chem.* **1992**, *57*, 4317–4320, DOI: `10.1021/jo00041a053`.

(64) Jakobi, M.; Winkelmann, G.; Kaiser, D.; Kempler, C.; Jung, G.; Berg, G.; Bahl, H. *J. Antibiot. (Tokyo)* **1996**, *49*, 1101–1104, DOI: `10.7164/antibiotics.49.1101`.

(65) Cao, S.; Blodgett, J. A.; Clardy, J. *Org Lett* **2010**, *12*, 4652–4654, DOI: `10.1021/ol1020064`.

(66) Capon, R. J.; Skene, C.; Lacey, E.; Gill, J. H.; Wadsworth, D.; Friedel, T. *J. Nat. Prod.* **1999**, *62*, 1256–1259, DOI: `10.1021/np990144v`.

(67) Hashidoko, Y.; Nakayama, T.; Homma, Y.; Tahara, S. *Tetrahedron Lett.* **1999**, *40*, 2957–2960, DOI: `10.1016/S0040-4039(99)00336-6`.

(68) Gunasekera, S. P.; Gunasekera, M.; Mccarthy, P. *J. Org. Chem.* **1991**, *56*, 4830–4833, DOI: `10.1021/jo00016a006`.

(69) Kanazawa, S.; Fusetani, N.; Matsunaga, S. *Tetrahedron Lett.* **1993**, *34*, 1065–1068, DOI: `10.1016/S0040-4039(00)77493-4`.

(70) Hart, A. C.; Phillips, A. J. *J. Am. Chem. Soc.* **2006**, *128*, 1094–1095, DOI: `10.1021/ja057899a`.

(71) Bae, M. A.; Kang, H. S.; Rue, S. W.; Seu, J. H.; Kim, Y. H. *Biochem. Biophys. Res. Commun.* **1998**, *246*, 276–281, DOI: `10.1006/bbrc.1998.8598`.

(72) Bae, M. A.; Yamada, K.; Uemura, D.; Seu, J. H.; Kim, Y. H. *J. Microbiol. Biotechnol.* **1998**, *8*, 455–460.

(73) Henderson, J. A.; Phillips, A. J. *Angew. Chem. Int. Ed. Engl.* **2008**, *47*, 8499–8501, DOI: `10.1002/anie.200803593`.

(74) Kirst, H. A.; Michel, K. H.; Martin, J. W.; Creemer, L. C.; Chio, E. H.; Yao, R. C.; Nakatsukasa, W. M.; Boeck, L.; Occolowitz, J. L.; Paschal, J. W.; Deeter, J. B.; Jones, N. D.; Thompson, G. D. *Tetrahedron Lett.* **1991**, *32*, 4839–4842, DOI: `10.1016/S0040-4039(00)93474-9`.

(75) Lou, L.; Qian, G.; Xie, Y.; Hang, J.; Chen, H.; Zaleta-Rivera, K.; Li, Y.; Shen, Y.; Dussault, P. H.; Liu, F.; Du, L. *J. Am. Chem. Soc.* **2011**, *133*, 643–645, DOI: `10.1021/ja105732c`.

(76) Li, Y. Y.; Huffman, J.; Li, Y.; Du, L. C.; Shen, Y. M. *Medchemcomm* **2012**, *3*, 982–986, DOI: `10.1039/c2md20026k`.

Bibliography

(77) Antosch, J.; Schaefers, F.; Gulder, T. A. M. *Angew. Chem. Int. Ed. Engl.* **2014**, *53*, 3011–3014, DOI: 10.1002/anie.201310641.

(78) Pfeifer, B. A.; Admiraal, S. J.; Gramajo, H.; Cane, D. E.; Khosla, C. *Science* **2001**, *291*, 1790–1792, DOI: 10.1126/science.1058092.

(79) Zhang, G.; Zhang, W.; Zhang, Q.; Shi, T.; Ma, L.; Zhu, Y.; Li, S.; Zhang, H.; Zhao, Y. L.; Shi, R.; Zhang, C. *Angew. Chem. Int. Ed. Engl.* **2014**, *53*, 4840–4844, DOI: 10.1002/anie.201402078.

(80) Li, Y.; Chen, H.; Ding, Y.; Xie, Y.; Wang, H.; Cerny, R. L.; Shen, Y.; Du, L. *Angew. Chem. Int. Ed. Engl.* **2014**, *53*, 7524–7530, DOI: 10.1002/anie.201403500.

(81) Bertasso, M.; Holzenkämpfer, M.; Zeeck, A.; Stackebrandt, E.; Beil, W.; Fiedler, H. P. *J. Antibiot. (Tokyo)* **2003**, *56*, 364–371, DOI: 10.7164/antibiotics.56.364.

(82) Greunke, C.; Antosch, J.; Gulder, T. A. M. *Chem. Commun. (Camb.)* **2015**, *51*, 5334–5336, DOI: 10.1039/c5cc00843c.

(83) Hughes, A. J.; Keatinge-Clay, A. *Chem. Biol.* **2011**, *18*, 165–176, DOI: 10.1016/j.chembiol.2010.12.014.

(84) Walker, M. C.; Thuronyi, B. W.; Charkoudian, L. K.; Lowry, B.; Khosla, C.; Chang, M. C. *Science* **2013**, *341*, 1089–1094, DOI: 10.1126/science.1242345.

(85) Sanchez, L. B.; Müller, M. *FEBS Lett.* **1996**, *378*, 240–244, DOI: 10.1016/0014-5793(95)01463-2.

(86) Brown, T. D.; Jones-Mortimer, M. C.; Kornberg, H. L. *J. Gen. Microbiol.* **1977**, *102*, 327–336, DOI: 10.1099/00221287-102-2-327.

(87) Uytterhoeven, B.; Appermans, K.; Song, L.; Masschelein, J.; Lathouwers, T.; Michiels, C. W.; Lavigne, R. *Microbiologyopen* **2016**, *5*, 279–286, DOI: 10.1002/mbo3.326.

(88) Rottig, M.; Medema, M. H.; Blin, K.; Weber, T.; Rausch, C.; Kohlbacher, O. *Nucleic Acids Res.* **2011**, *39*, W362–367, DOI: 10.1093/nar/gkr323.

(89) Christiansen, G.; Philmus, B.; Hemscheidt, T.; Kurmayer, R. *J. Bacteriol.* **2011**, *193*, 3822–3831, DOI: 10.1128/JB.00360-11.

(90) Yamanaka, K.; Maruyama, C.; Takagi, H.; Hamano, Y. *Nat. Chem. Biol.* **2008**, *4*, 766–772, DOI: 10.1038/nchembio.125.

(91) Calcott, M. J.; Owen, J. G.; Lamont, I. L.; Ackerley, D. F. *Appl. Environ. Microbiol.* **2014**, *80*, 5723–5731, DOI: 10.1128/AEM.01453-14.

(92) Bachmann, B. O.; Ravel, J. *Methods Enzymol.* **2009**, *458*, 181–217, DOI: 10.1016/s0076-6879(09)04808-3.

(93) Grammel, N.; Pankevych, K.; Demydchuk, J.; Lambrecht, K.; Saluz, H. P.; Krügel, H. *Eur. J. Biochem.* **2002**, *269*, 347–357, DOI: `10.1046/j.0014-2956.2001.02657.x`.

(94) Stachelhaus, T.; Mootz, H. D.; Marahiel, M. A. *Chem. Biol.* **1999**, *6*, 493–505, DOI: `10.1016/S1074-5521(99)80082-9`.

(95) Conti, E.; Stachelhaus, T.; Marahiel, M. A.; Brick, P. *EMBO J.* **1997**, *16*, 4174–4183, DOI: `10.1093/emboj/16.14.4174`.

(96) Chiu, J.; Tillett, D.; Dawes, I. W.; March, P. E. *J. Microbiol. Methods* **2008**, *73*, 195–198, DOI: `10.1016/j.mimet.2008.02.013`.

(97) Greunke, C.; Glöckle, A.; Antosch, J.; Gulder, T. A. M. *Angew. Chem. Int. Ed. Engl.* **2017**, *56*, 4351–4355, DOI: `10.1002/anie.201611063`.

(98) Williams, D. C.; Van Frank, R. M.; Muth, W. L.; Burnett, J. P. *Science* **1982**, *215*, 687–689, DOI: `10.1126/science.7036343`.

(99) Freedman, R. B.; Wetzel, R. *Curr. Opin. Biotechnol.* **1992**, *3*, 323–325, DOI: `doi.org/10.1016/0958-1669(92)90158-F`.

(100) Chrunyk, B. A.; Evans, J.; Lillquist, J.; Young, P.; Wetzel, R. *J. Biol. Chem.* **1993**, *268*, 18053–18061.

(101) Bräutigam, A. R.; Sauerbier, W. *J. Virol.* **1973**, *12*, 882–886.

(102) Borkowski, O.; Ceroni, F.; Stan, G. B.; Ellis, T. *Curr. Opin. Microbiol.* **2016**, *33*, 123–130, DOI: `10.1016/j.mib.2016.07.009`.

(103) Singh, A.; Upadhyay, V.; Upadhyay, A. K.; Singh, S. M.; Panda, A. K. *Microb Cell Fact* **2015**, *14*, 41, DOI: `10.1186/s12934-015-0222-8`.

(104) Peternel, S.; Grdadolnik, J.; Gaberc-Porekar, V.; Komel, R. *Microb Cell Fact* **2008**, *7*, 34, DOI: `10.1186/1475-2859-7-34`.

(105) Gibson, D. G.; Young, L.; Chuang, R. Y.; Venter, J. C.; Hutchison, r., C. A.; Smith, H. O. *Nat Methods* **2009**, *6*, 343–345, DOI: `10.1038/nmeth.1318`.

(106) Kouprina, N.; Larionov, V. *Nat. Protoc.* **2008**, *3*, 371–377, DOI: `10.1038/nprot.2008.5`.

(107) Yamanaka, K.; Reynolds, K. A.; Kersten, R. D.; Ryan, K. S.; Gonzalez, D. J.; Nizet, V.; Dorrestein, P. C.; Moore, B. S. *Proc. Natl. Acad. Sci. U. S. A.* **2014**, *111*, 1957–1962, DOI: `10.1073/pnas.1319584111`.

(108) Bian, X.; Huang, F.; Stewart, F. A.; Xia, L.; Zhang, Y.; Müller, R. *Chembiochem* **2012**, *13*, 1946–1952, DOI: `10.1002/cbic.201200310`.

(109) Wang, H.; Li, Z.; Jia, R.; Hou, Y.; Yin, J.; Bian, X.; Li, A.; Müller, R.; Stewart, A. F.; Fu, J.; Zhang, Y. *Nat. Protoc.* **2016**, *11*, 1175–1190, DOI: 10.1038/nprot.2016.054.

(110) Yin, J.; Hoffmann, M.; Bian, X.; Tu, Q.; Yan, F.; Xia, L.; Ding, X.; Stewart, A. F.; Müller, R.; Fu, J.; Zhang, Y. *Sci. Rep.* **2015**, *5*, 15081, DOI: 10.1038/srep15081.

(111) Nah, H. J.; Pyeon, H. R.; Kang, S. H.; Choi, S. S.; Kim, E. S. *Front. Microbiol.* **2017**, *8*, 394, DOI: 10.3389/fmicb.2017.00394.

(112) Li, Y.; Li, Z.; Yamanaka, K.; Xu, Y.; Zhang, W.; Vlamakis, H.; Kolter, R.; Moore, B. S.; Qian, P. Y. *Sci. Rep.* **2015**, *5*, 9383, DOI: 10.1038/srep09383.

(113) Kumpfmüller, J.; Methling, K.; Fang, L.; Pfeifer, B. A.; Lalk, M.; Schweder, T. *Appl. Microbiol. Biotechnol.* **2016**, *100*, 1209–1220, DOI: 10.1007/s00253-015-6990-6.

(114) Wenzel, S. C.; Gross, F.; Zhang, Y.; Fu, J.; Stewart, A. F.; Müller, R. *Chem. Biol.* **2005**, *12*, 349–356, DOI: 10.1016/j.chembiol.2004.12.012.

(115) Gross, F.; Ring, M. W.; Perlova, O.; Fu, J.; Schneider, S.; Gerth, K.; Kuhlmann, S.; Stewart, A. F.; Zhang, Y.; Müller, R. *Chem. Biol.* **2006**, *13*, 1253–1264, DOI: 10.1016/j.chembiol.2006.09.014.

(116) Videau, P.; Wells, K. N.; Singh, A. J.; Gerwick, W. H.; Philmus, B. *ACS Synth Biol* **2016**, *5*, 978–988, DOI: 10.1021/acssynbio.6b00038.

(117) Jackson, B. E.; Hart-Wells, E. A.; Matsuda, S. P. *Org Lett* **2003**, *5*, 1629–1632, DOI: 10.1021/ol034231x.

(118) Kishimoto, S.; Tsunematsu, Y.; Sato, M.; Watanabe, K. *Chem. Rec.* **2017**, DOI: 10.1002/tcr.201700015.

(119) Baer, P.; Rabe, P.; Fischer, K.; Citron, C. A.; Klapschinski, T. A.; Groll, M.; Dickschat, J. S. *Angew. Chem. Int. Ed. Engl.* **2014**, *53*, 7652–7656, DOI: 10.1002/anie.201403648.

(120) Kaysser, L.; Bernhardt, P.; Nam, S. J.; Loesgen, S.; Ruby, J. G.; Skewes-Cox, P.; Jensen, P. R.; Fenical, W.; Moore, B. S. *J. Am. Chem. Soc.* **2012**, *134*, 11988–11991, DOI: 10.1021/ja305665f.

(121) Bierman, M.; Logan, R.; O'Brien, K.; Seno, E. T.; Rao, R. N.; Schoner, B. E. *Gene* **1992**, *116*, 43–49, DOI: 10.1016/0378-1119(92)90627-2.

(122) Chang, Y.; Mead, D.; Dhodda, V.; Brumm, P.; Fox, B. G. *Protein Sci.* **2009**, *18*, 2316–2325, DOI: 10.1002/pro.242.

(123) Wang, X. K.; Jin, J. L. *FEMS Microbiol. Lett.* **2014**, *357*, 99–103, DOI: 10.1111/1574-6968.12507.

(124) Ohnishi, Y.; Ishikawa, J.; Hara, H.; Suzuki, H.; Ikenoya, M.; Ikeda, H.; Yamashita, A.; Hattori, M.; Horinouchi, S. *J. Bacteriol.* **2008**, *190*, 4050–4060, DOI: `10.1128/JB.00204-08`.

(125) Luo, Y.; Huang, H.; Liang, J.; Wang, M.; Lu, L.; Shao, Z.; Cobb, R. E.; Zhao, H. *Nat Commun* **2013**, *4*, 2894, DOI: `10.1038/ncomms3894`.

(126) Saha, S.; Zhang, W.; Zhang, G.; Zhu, Y.; Chen, Y.; Liu, W.; Yuan, C.; Zhang, Q.; Zhang, H.; Zhang, L.; Zhang, W.; Zhang, C. *Chem. Sci.* **2017**, *8*, 1607–1612, DOI: `10.1039/c6sc03875a`.

(127) Yu, H. L.; Jiang, S. H.; Bu, X. L.; Wang, J. H.; Weng, J. Y.; Yang, X. M.; He, K. Y.; Zhang, Z. G.; Ao, P.; Xu, J.; Xu, M. J. *Sci. Rep.* **2017**, *7*, 40689, DOI: `10.1038/srep40689`.

(128) Xu, L.; Wu, P.; Wright, S. J.; Du, L.; Wei, X. *J. Nat. Prod.* **2015**, *78*, 1841–1847, DOI: `10.1021/acs.jnatprod.5b00099`.

(129) Lacret, R.; Oves-Costales, D.; Goméz, C.; Díaz, C.; de la Cruz, M.; Pérez-Victoria, I.; Vicente, F.; Genilloud, O.; Reyes, F. *Mar. Drugs* **2014**, *13*, 128–140, DOI: `10.3390/md13010128`.

(130) Hoepflinger, M. C.; Hoeftberger, M.; Sommer, A.; Hametner, C.; Foissner, I. *Front Plant Sci* **2017**, *8*, 20, DOI: `10.3389/fpls.2017.00020`.

(131) Speers, A. E.; Adam, G. C.; Cravatt, B. F. *J. Am. Chem. Soc.* **2003**, *125*, 4686–4687, DOI: `10.1021/ja034490h`.

List of Abbreviations

AT   Acyltransferase domain

ATP   Adenosine triphosphate

BGC   Biosynthetic gene cluster

CoA   Coenzyme A

DiPaC   Direct pathway cloning

FAD   Flavin adenine dinucleotide

gDNA   Genomic deoxyribonucleic acid

HPLC   High-performance liquid chromatography

HSAF   Heat-stable anti-fungal factor

IPTG   Isopropyl β-D-1-thiogalactopyranoside

LCHR   Linear plus circular homologous recombination

LC-MS   Liquid chromatography coupled to mass spectrometry

MALDI   Matrix-assisted laser desorption/ionization

NADPH   Nicotinamide adenine dinucleotide phosphate

NMR   Nuclear magnetic resonance spectroscopy

NRP   Non-ribosomal peptide

NRPS   Non-ribosomal peptide synthetase

PCR   Polymerase chain reaction

PKS   Polyketide synthase

PoTeM   Polycyclic tetramate macrolactam

PPTase   Phosphopantetheinyl transferase

SAM   *S*-Adenosyl methionine

TE   Thioesterase domain

List of Figures

List of Schemes

List of Tables

Attachments

A.1. Supplemental materials of publications

A.1.1. PoTeM hydroxylases – Greunke *et al.* 2015

The supplemental information is related to the following publication which was highlighted in Chapter 3.1:

- **Greunke, C.***; Antosch, J.*; Gulder, T.A.M. Promiscuous hydroxylases for the functionalization of polycyclic tetramate macrolactams – conversion of ikarugamycin to butremycin. *Chem. Comm.* **2015**, *51*, 5334-5336, DOI: 10.1039/c5cc00843c.

  *equally contributing authors

Supporting Information

Promiscuous Hydroxylases for the Functionalization of Polycyclic Tetramate Macrolactams – Conversion of Ikarugamycin to Butremycin

Christian Greunke[‡]**, Janine Antosch**[‡] **and Tobias A. M. Gulder**

Biosystems Chemistry, Department of Chemistry and Center for Integrated Protein Science Munich (CIPSM), Technische Universität München, Lichtenbergstraße 4, 85747 Garching.

E-Mail: tobias.gulder@ch.tum.de

‡*: These authors contributed equally to this work.*

Contents

1.) General methods

1.1 Bacterial strains and cultivation

E. coli DH5a was used as the standard cloning host. *E. coli* BAP1[1] was used for heterologous expression experiments. *E. coli* BL21(DE3) was used for protein expression. All work with *E. coli* was carried out on LB agar or LB liquid media (Roth) with the appropriate antibiotics for selection. Heterologous expression experiments were either carried out in LB with induction of secondary metabolite production using 1 mM IPTG. Alternatively, auto-induction media was used. Per 50 mL, this was composed of 46 mL ZY media (10 g tryptone, 5 g yeast extract, 925 mL H_2O), 0,5 mg arginine, 50 µL magnesium sulfate solution (24.65 g $MgSO_4$ x 7 H_2O, 13.84 g $MgSO_4$ x H_2O, 100 mL H_2O), 1 mL 50 x 5052 solution (250 g glycerol, 100 g α-lactose, 25 g glucose, 730 mL H_2O), and 2.5 mL 20 x NPS solution (142 g Na_2HPO_4, 136 g KH_2PO_4, 66 g $(NH_4)_2SO_4$, 900 mL H_2O).[2] *E. coli* BAP1 pHis8::Tü6239_*ikaABC* was used for in vivo experiments.[3]

Lysobacter capsici (DSM 19286) was grown in 1/10-fold concentrated TSB-Medium at 30 °C while shaking (200 rpm) for three days, until gDNA was prepared following the instructions of the Bacteria DNA Preparation Kit (Jena Bioscience).

Saccharophagus degradans (DSM 17024) was obtained from DSMZ. A pCC1Fos-based genomic library was prepared and a clone, covering the whole PTM cluster, was identified by PCR screening. This material was used as template DNA for PCR.

1.2 Transformation of electrocompetent *E. coli*

Electrocompetent cells were generated as follows: 200 mL of SOB media (20 g tryptone, 5 g yeast extract, 0.6 g NaCl, 0.2 g KCl, 10 mM $MgCl_2$, 10 mM $MgSO_4$) were inoculated with 5 mL of a preculture of the respective *E. coli* strain (grown overnight at 37 °C) and incubated at 37 °C at 200 rpm until the OD_{600} reached 0.8. The culture was centrifuged at 4 °C for 5 min at 5000 rpm. The supernatant was discarded, the pellet re-suspended in 25 mL of icecold 10 % glycerol and centrifuged at 4 °C for 5 min at 5,000 rpm. This procedure was repeated 4 times with different amounts of ice-cold 10 % glycerol (20 mL, 15 mL, 10 mL, 5 mL). After the washing steps, the pellet was re-suspended in 1 mL 10 % glycerol and aliquoted in 75 µL portions. The aliquots were flash frozen in liquid nitrogen and stored at -80 °C until use.

For electroporation, 5 µL of the respective plasmid to be introduced were added to an aliquot of thawed electrocompetent cells. The electroporation was carried out in a 0.2 cm ice-cold electroporation cuvette using a BioRad GenePulser II set to: 200 Ω, 25 µF and 2.5 kV. The expected time constant was 4.5-5.5 ms. Immediately after the electric pulse, 700 µL of ice-cold LB broth was added and the cells were incubated for 45 minutes at 37 °C. 150 µL solution was plated on LB agar supplemented with appropriate antibiotics and incubated at 37 °C overnight.

1.3 Isolation and purification of plasmids and PCR amplicons

Plasmids were isolated from single clones incubated overnight in LB broth with appropriate antibiotics using the peqGOLD Plasmid Miniprep Kit I, C-Line (Peqlab) according to the protocol provided by the manufacturer. PCR products were either isolated using the PCR Purification Kit (Jena Bioscience) or by gel electrophoresis with re-isolation from the excised gel sample using the peqGOLD Gel Extraction Kit, S-Line (Peqlab).

1.4 Restriction analysis

A plasmid restriction (10 µL volume) was composed of: 1 µL compatible NEB buffer (10 fold); 1 µL BSA (10 fold); 0,125 µL per each enzyme, 5 µL plasmid solution containing 200 ng plasmid DNA and pure water. This mixture was incubated for 2 h at 37 °C and analyzed by gel electrophoresis. All enzymes and buffers used were purchased from NEB.

1.5 Chemical analysis

HPLC-MS analyses were carried out using an UltiMate 3000 LC System system coupled to a LCQ Fleet Ion Trap Mass Spectrometer (Thermo Scientific). HPLC-HR/MS analyses were performed on a Thermo Finnigan LTQ FT-ICR equipped with a Dionex Ultimate 3000 separation module. The Xcalibur software was used to control the system. Interpretation of the recorded data was carried out using the Thermo Xcalibur Qual Browser 2.2 SP1.48 software.

The chromatographic separation for HPLC-MS was carried out on a XBridge (Waters) C-18 HPLC column (30 x 4,6 mm, 3,5 µm particle size) using the following gradient with water (A) and acetonitrile (B), both buffered with 0,1 % formic acid, as the eluents: linear gradient from 20 % B (0 min), 95 % B (6 min), 95 % B (7.5 min), 20 % B (8 min), 20 % B (9 min), all at a flow rate of 1,1 mL/min. The chromatographic separation for HPLC-HR/MS was carried out on a XBridge (Waters) C-18 HPLC column (100 x 4,6 mm, 3,5 µm particle size) using the following gradient with water (A) and acetonitrile (B), both buffered with 0,1 % formic acid, as the eluents: linear gradient from 5 % B (2 min), 5 % B (2 min), 20 % B (2.5 min), 95 % B (5 min), 95 % B (6 min), 20 % B (9 min), 20 % B at 1,1 mL/min flow rate.

2.) Bioinformatic data
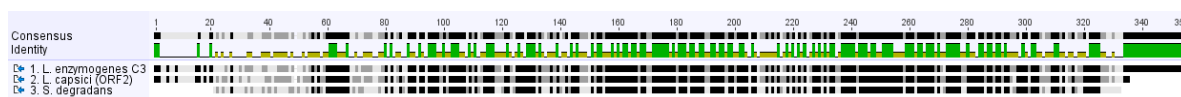
Figure S1. Alignment of SD-like protein sequences from *Lysobacter enzymogenes* C3 (1), *Lysobacter capsici* [ORF2 – Extended construct] (2) and *Saccharophagus degradans* SD-like protein (3) using Geneious default settings with Blosum 62 cost matrix.

3.) Cloning of SD genes

3.1 Screening of gDNA samples for SD (*L. capsici*) – Taq-Polymerase

A total 25 µL PCR reaction was composed as follows: 2.5 µL of tenfold Taq buffer, 1 µL DMSO, 0.25 µL deoxynucleotide triphosphates (10 mM stock), 0.125 µL forward primer (20 µM), 0.125 µL reverse primer (20 µM) (see chapter 3, table S6, row 1 and 2), 19.875 µL pure water and 1 µL Template DNA (*L. capsici* gDNA, c = 305 ng/µL) were mixed with 0.125 µL Taq polymerase. Optimal cycling conditions are depicted in table S1.

Table S1.

| Step	Temperature [°C]	Time [sec]	Cycle nr.
1	95	120	
2	95	45	
3	52	30	x 34
4	72	45	
5	72	300	
6	16	∞	
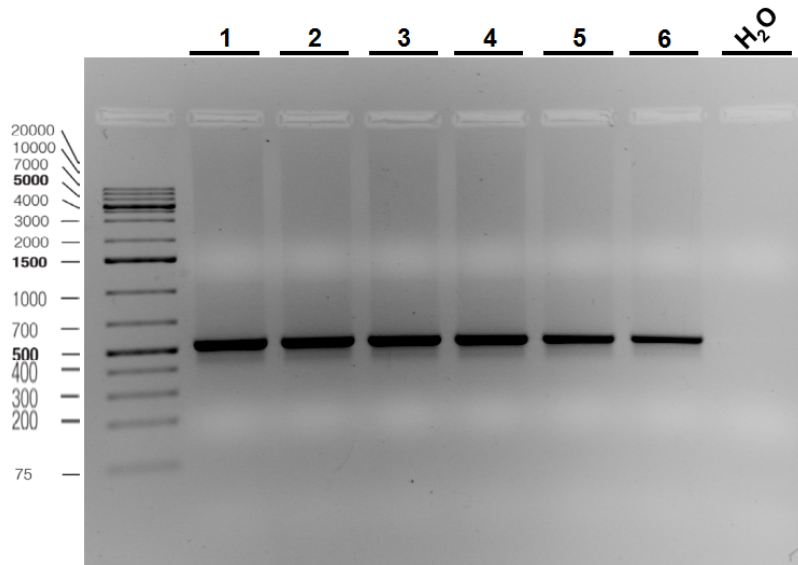
Figure S2. PCR screening of 6 samples for the presence of *L. capsici* SD with SD-Capsici For/Rev. The expected amplicon was 526 bp.

3.2 Amplification of SD (*L. capsici*) (ORF1, 810 bp) – Q5 polymerase

A total 25 µL PCR reaction was composed as follows: 5 µL of fivefold Q5 Reaction Buffer, 0.5 µL deoxynucleotide triphosphates (10 mM stock), 0.625 µL forward primer (20 µM), 0.625 µL reverse primer (20 µM) (see chapter 3, table S6, row 3 and 5), 17.125 µL pure water and 1 µL template DNA (*L. capsici* gDNA, c = 305 ng/µL) were mixed with 0.125 µL Q5 polymerase (2000 U/mL, NEB). Optimal cycling conditions are depicted in table S2.

Table S2.

Step	Temperature [°C]	Time [sec]	Cycle nr.
1	98	60	
2	98	10	
3	72	20	x 30
4	72	35	
5	72	120	
6	10	∞	
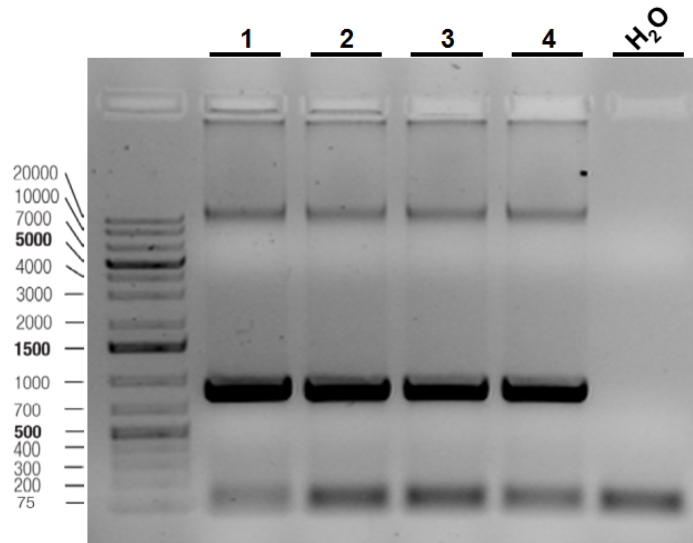
Figure S3. PCR amplification of *L. capsici* SD (ORF1, GLIMMER prediction) with *Eco*RI / *Hin*dIII overhangs (844 bp).

3.3 Amplification of SD (*L. capsici*) – (ORF2, 1002bp) – Q5 polymerase

A total 25 µL PCR reaction was composed as follows: 5 µL of fivefold Q5 Reaction Buffer, 0.5 µL deoxynucleotide triphosphates (10 mM stock), 0.625 µL forward primer (20 µM), 0.625 µL reverse primer (20 µM) (see chapter 3, table S6, row 4 and 5), 17.125 µL pure water and 1 µL template DNA (*L. capsici* gDNA, c = 305 ng/µL) were mixed with 0.125 µL Q5 polymerase (2000 U/mL, NEB). Optimal cycling conditions are depicted in table S3.

Table S3.

Step	Temperature [°C]	Time [sec]	Cycle nr.
1	98	60	
2	98	10	
3	72	20	x 30
4	72	45	
5	72	120	
6	10	∞	
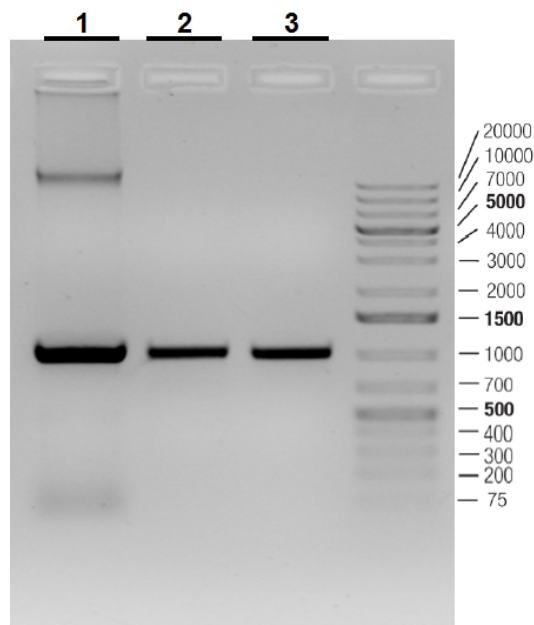
Figure S4. PCR amplification of *L. capsici* SD (ORF2). Lane 1: Crude PCR mix. Lane 2: Gel purified PCR product containing *Eco*RI / *Hin*dIII overhangs (1035 bp). Lane 3: *Eco*RI + *Hin*dIII digested PCR product after clean-up (1014 bp).

3.4 Colony screening PCR

Clones were picked and examined via colony PCR (for conditions, see Table S4) using specific primers (see Table S6) for the amplification of the sterol desaturase gene expected to be located in the respective construct (see Figure S5). Positive hits were further characterized by plasmid isolation followed by end-sequencing to verify the recombined plasmid.

A total 25 µL PCR reaction was composed as follows: 2.5 µL of tenfold Taq buffer, 1 µL DMSO, 0.25 µL deoxynucleotide triphosphates (10 mM stock), 0.125 µL forward primer (20 µM), 0.125 µL pET-RP (reverse primer) (20 µM) (see chapter 3, table S6, row 1 (*L. capsici*) / 6 (*S. degradans*) and 8), 19.875 µL pure water and 1 µL Template (bacterial material) were mixed with 0.125 µL Taq polymerase. Optimal cycling conditions are depicted in table S4.

Table S4. PCR conditions for colony screening

Step	Temperature [°C]	Time [sec]	Cycle nr.
1	95	300	
2	95	45	
3	47	30	x 34
4	72	60	
5	72	300	
6	16	∞	
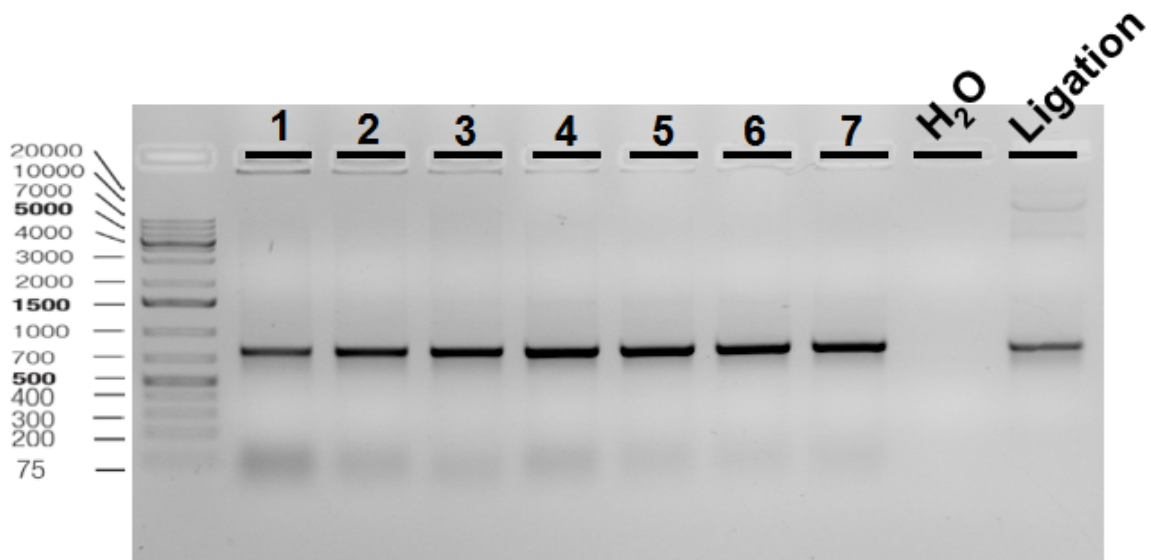
Figure S5. *Example for a screening PCR of seven* E. coli *DH5α clones with SD-Capsici for / pET-RP rev. An amplicon of 770 bp was obtained for clones harboring pHis8 / pGS-21a::*L-capsici*_SD (ORF1/2).*

### 3.4.1 Restriction analysis *L. capsici SD* [ORF2] cloned in pHis8
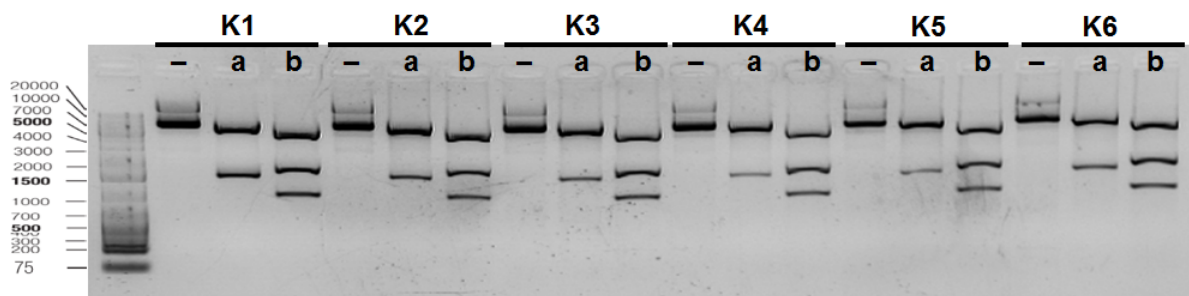
Figure S6. Restriction analysis of pHis8::*L-capsici*_SD_ORF2 (6328 bp) cut with a) *Eco*RI + *Hin*dIII (5314 + 1014 bp) and b) *Pvu*I (4415 + 1232 + 664 [+17] bp). Undigested plasmid is shown in line (-).

3.4.2 Restriction analysis *L. capsici SD* [ORF2] cloned in pGS-21a
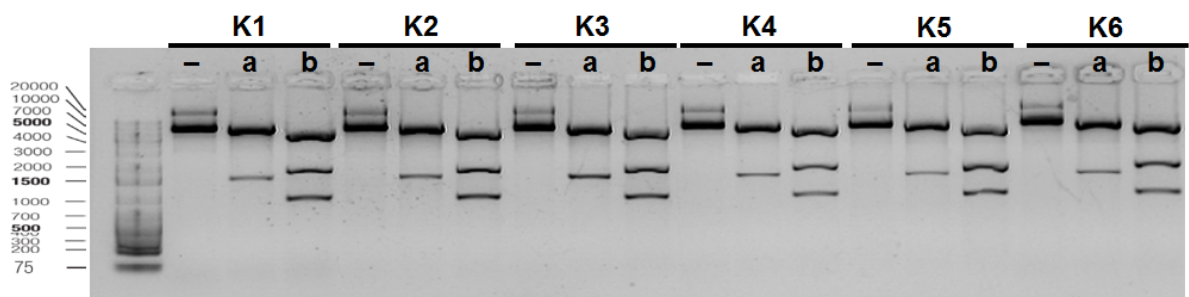
Figure S7. Restriction analysis of pGS-21a::*L-capsici*_SD_ORF2 (7165 bp) cut with a) *Eco*RI + *Hin*dIII (6151 + 1014 bp) and b) *Pvu*I (5180 + 1304 + 664 [+17] bp). Undigested plasmid is shown in line (-).

3.5 Amplification of SD (*S. degradans*) – Q5 polymerase (NEB)

A total 25 μL PCR reaction was composed as follows: 5 μL of fivefold Q5 Reaction Buffer, 0.5 μL deoxynucleotide triphosphates (10 mM stock), 0.625 μL forward primer (20 μM), 0.625 μL reverse primer (20 μM) (see chapter 3, table S6, row 6 and 7), 16.125 μL pure water and 2 μL template DNA (fosmid DNA, c = 17.4 ng/μL) were mixed with 0.125 μL Q5 polymerase (2000 U/mL, NEB). Optimal cycling conditions are depicted in table S5.

Table S5.

Step	Temperature [°C]	Time [sec]	Cycle nr.
1	98	60	
2	98	10	
3	65	20	x 30
4	72	35	
5	72	120	
6	10	∞	
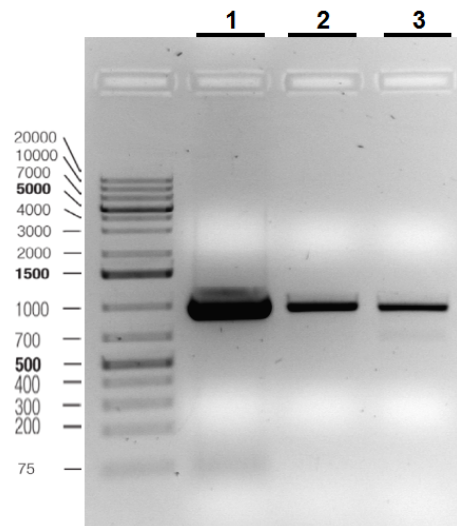
Figure S8. PCR amplification of *S. degradans* SD. Lane 1: Crude PCR mix. Lane 2: Gel purified PCR product containing *Bam*HI / *Xho*I overhangs (961 bp). Lane 3: *Bam*HI + *Xho*I digested PCR product after clean-up (945 bp).

3.5.1 Restriction analysis SD (*S. degradans*) cloned in pHis8
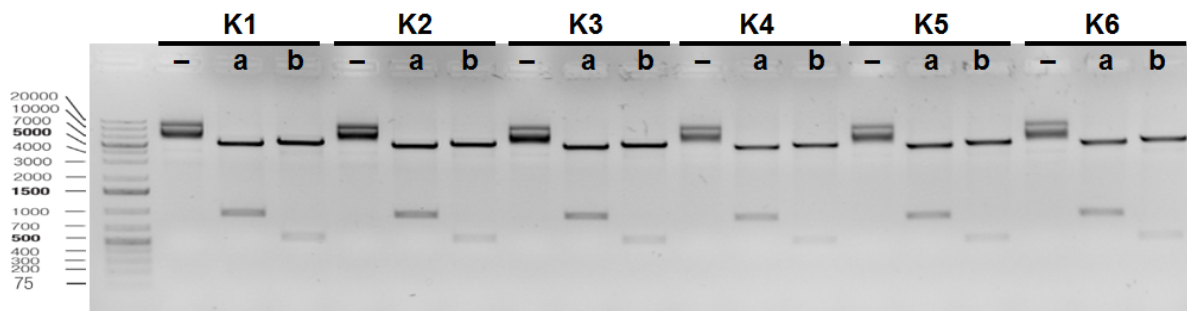
Figure S9. Restriction analysis of pHis8::*S-degradans*_SD (6238 bp) cut with a) *Bam*HI + *Xho*I (5239 + 945 bp) and b) *Dra*III (5679 + 559 bp). Undigested plasmid is shown in line (-).

3.5.2 Restriction analysis SD (*S. degradans*) cloned in pGS-21a
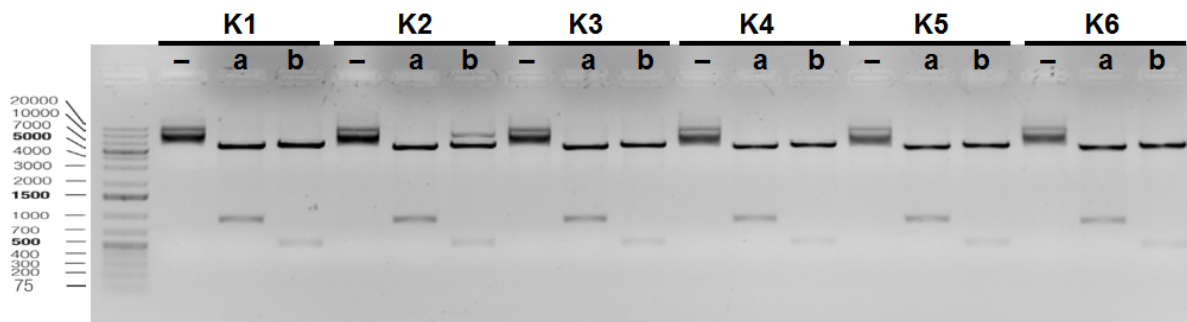
Figure S10. Restriction analysis of pHis8::*S-degradans*_SD (7075 bp) cut with a) *Bam*HI + *Xho*I (6130 + 945 bp) and b) *Dra*III (6516 + 559 bp). Undigested plasmid is shown in line (-).

Table S6. List of used primers in this study.

Entry	Primer name	Sequence	Comments	Tm [°C]
1	SD-Capsici For	5'-GTACGCGGTCTACTTCGTGA-3'	Screening primer for *L. capsici's* SD-like protein	56.5
2	SD-Capsici Rev	5'-CCAGGTATAGCTGCCGAACA-3'	Screening primer for *L. capsici's* SD-like protein	56.9
3	SD-Capsici_ORF1_Eco RI For	5'-CGGAACTTAGAATTCGTGGTTATCTACGTAATTGCCGCGA-3'	Cloning primer containing an *Eco*RI restriction site, GLIMMER predicted frame (ORF1) 810 bp	64.8
4	SD-Capsici_ORF2_Eco RI For	5'-AGAACTTAGAATTCAACACGGACGCCGGTCGCA-3'	Cloning primer containing an *Eco*RI restriction site, Extended frame (ORF2) 1002 bp	66.5
5	SD-Capsici_HindIII Rev	5'-ATTCAACAAGCTTAAAGGCTCAGGGCGGCGT-3'	Cloning primer containing a *Hin*dIII restriction site, GLIMMER predicted frame	66.5
6	SD_*S-degradans*_BamHI For	5'-AACTTGGATCCGTGGATACCTTAACTCCCTCAG-3'	Cloning primer containing a *Bam*HI restriction site	64.8
7	SD_*S-degradans*_XhoI Rev	5'-ACCTTCTCGAGTTACTGGCGGTAGGTAATGTTG-3'	Cloning primer containing a *Xho*I restriction site	66.5
8	pET-RP	5'-CTAGTTATTGCTCAGCGG-3'	Screening / Sequencing primer	50.2

4.) *In vivo* and *in vitro* assays

4.1 Heterologous expression experiments - Fermentation of *E. coli* BAP1

The sequenced sterol desaturase pGS-21a-constructs were transformed into electrocompetent *E. coli* BAP1 cells via electroporation. These *E. coli* BAP1 contained the pHis8::Tü6239_*ikaABC* construct which incorporates the whole ikarugamycin cluster. The transformed cells were fermented and automatically induced in auto-induction media with 0.05 mg/mL arginine (kanamycin 50 µg/mL and ampicillin 100 µg/mL). The fermentation was carried out for 72 h at 16 °C and 200 rpm in a 250 mL Erlenmeyer flask containing 150 mL of auto-induction medium. For the isolation and physico-chemical characterization, the pH of the main culture was adjusted to pH 5.0 with 1N HCl and centrifuged for 10 min at 3,800 rpm. The supernatant was extracted three times with 150 mL ethylacetate. The combined organic phases were concentrated in a rotary evaporator. The pellet was dissolved in 50 mL of a methanol/acetone solution and sonicated in an ultrasonic bath for 15 min. The solvents were removed in vacuo. The raw product was dissolved in 1 mL methanol for the further LC-MS measurement.

4.2 Feeding experiments

100 µL *E. coli* BL21(DE3) pGS-21a::SD (after 15 h incubation with IPTG 1 mM, 16 °C, 200 rpm), 10 µL ikarugamycin (1 mg/mL in methanol, Santa Cruz Biotechnology, Texas), 1 µL ATP (100 mM) and 1 µL NADPH (100 mM) were combined and incubated for 6 h at 30 °C while shaking (350 rpm). Afterwards, 100 µL of pure methanol was added to the cellular mixture. The revealed supernatant after centrifugation for 1 min at 13,000 rpm was used for LC-MS measurement.

4.3 *In vitro* conversion assays

100 µL of the cell lysate supernatant, prepared from induced *E. coli* BL21(DE3), were spiked with 1 µL ATP (100 mM) and 1 µL NADPH (100 mM). In control reactions, the supernatants were incubated at 95 °C for 10 minutes. Adjacently, 10 µL ikarugamycin from a 1 mg/mL concentrated methanolic stock solution were added. After an incubation period of 6 hours at 30 °C while shaking (350 rpm), the reactions were quenched by the addition of 110 µL methanol. After centrifugation for 1 min at 13,000 rpm, the supernatants were carefully transferred and used for the LC-MS measurements.

The enzymatic reaction with N-GST-SD (*S. degradans*) was composed as follows: 20 µL enriched enzyme, 10 µL sodium phosphate buffer (1 M, pH 8.0), 1 µL ATP (100 mM), 1 µL NADPH (100 mM) and 10 µL ikarugamycin (c = 1 mg/mL, methanol stock solution) filled up with water to a volume of 100 µL. The reaction was incubated at 30 °C while shaking (350 rpm) for 12 hours. The reaction was then quenched by the addition of 100 µL methanol and subsequent centrifugation for 1 min at 13,000 rpm. The supernatant was used for LC-MS analysis.

4.4 Heterologous protein production in *E. coli* and enrichment of N-GST-SD (*S. degradans*)

Starter cultures of *E. coli* BL21(DE3) pGS-21a::SD_*S-degradans* (5 mL) were grown overnight in Luria broth (Roth) and used to inoculate Terrific broth (Roth), supplemented with 100 µg/mL ampicillin. When an OD_{600} value of 0.6 was reached, cultures were cooled and protein synthesis was induced by adding 0.5 mM IPTG following an overnight incubation at 16 °C and moderate shaking (200 rpm). Alternatively, starter culteres were mixed 1:100 (v/v) with ZYM-5052 auto-induction medium.[2] After cultivation for around 3 hours at 37 °C while shaking (200 rpm), cultures were cooled and further incubated at 16 °C overnight.

In the following, cells were harvested by centrifucation and resuspended in lysis buffer (50 mM NaH_2PO_4 x 2 H_2O, 300 mM NaCl, 10 % Glycerol, 10 mM imidazol, pH 8.0). After sonification on ice and a subsequent centrifugation (10,000 rpm, 4 °C, 30 min), the supernatants were taken and incubated with High Density Nickel Agarose (Jena Bioscience) and moderate shaking for at least 1 hour. This mixture was then applied onto a column, which was washed with lysis buffer (40 mM imidazole, pH 7.5). Finally, the protein was eluted with lysis buffer (250 mM imidazole, pH 7.5).

PD-10 columns (GE Healthcare) were used to exchange the eluate's buffer to 50 mM Tris-HCl buffer, pH 7.5 supplemented with 10 % glycerol. This protein solution was then concentrated using Vivaspin 2 columns (Sartorius).

Protein concentrations were determined photometrically by the use of the Implen Nanophotometer 330.

Bacterial induction and protein enrichment were monitored by SDS-PAGE analysis (Figure S11).
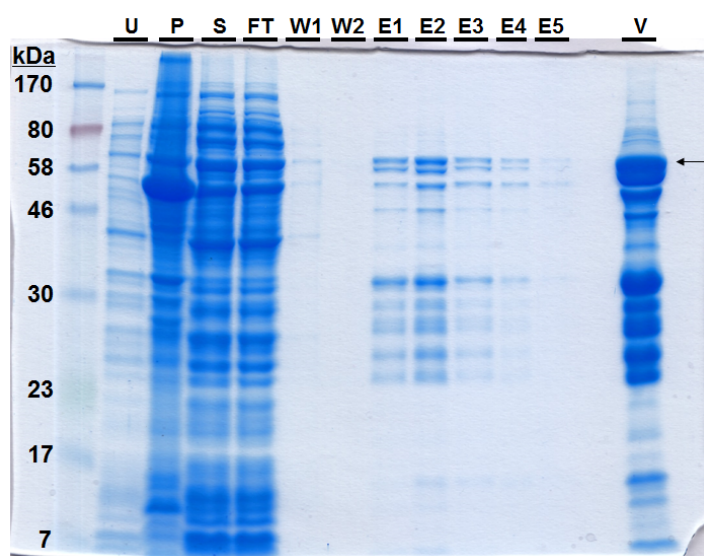
Figure S11. SDS-PAGE analysis for the expression and enrichment of N-GST-SD (*S. degradans*). Samples: uninduced bacteria (U), pellet (P), supernatant (S), flow-through (FT), wash fractions (W1, W2), elution fractions (E1-E5), desalted and concentrated eluate (V).

5.) Sequences of investigated PTM hydroxylase genes and corresponding translations

5.1 *Lysobacter capsici* DSM 19286

ORF1 – 810 bases (GLIMMER prediction):
```
GTGGTTATCTACGTAATTGCCGCGATCGGCCTGGACTGGGATCCGGGCGCCGCCAGCCTCGGCTTCGTGCTCGCCACCGTCTT
CTATCTGGCGATGCTCGAATTAATCATTCCGTACGAGCGCGCATGGCTGCCCTCGCGGCGCGAGTGGGGCCTGTACGCGGTCT
ACTTCGTGATCACGATGATCGGCGGCGCGCTGGCGCAGCTGCCGGTGATGGCGGTGGTGTCGATGGTCGCCCCGCTGCATCCG
CTGTTGCCGCTGTGGGCGCAGATCCCGCTCGCGCTGCTGCTGAGCTCGCTGGCGAGCTATGCGGTGCATCGCGCCGGCCACGA
CATCCCGCTGCTGTGGCGGCTGCACGGCGTGCATCACGTGCCCGACAAGGTCAACGTCGGCAACAACGGCGTCAACCACGTGT
TCGACATCGTCCTGGCCCAGTTCCTGGTGCAGGTGTCGCTGGCGCTGATCGGCTTTTCCGAGCACGCGGTGTTCGCCGTCGGC
ATCTTCATCATCGCCCAGGGCTATTTCGTCCACGCCAACATCGACGTGCGCCTGGGCTGGCTCAACCACATCCTCGCCAGCCC
CGAGCTGCACCGCATGCATCACAGCGCCGACAAGGCCGAGGCCGGGCATTTCGGCTCGGACCTGTCGATCTGGGACCGCCTGT
TCGGCAGCTATACCTGGCGGCCCGATCGCAAGCCGCGCGCGATCGGCCTGTTCGATCCGGGCTCGTTCCCGCGCAACGGCGCG
ATCTTCTCGACCCTCGTCCACCCGCTGCGGCCGCGCAAGAAACCCGCGCCGACGCCGCCCTGA
```

Translation ORF1 – 269 amino acids:
```
MVIYVIAAIGLDWDPGAASLGFVLATVFYLAMLELIIPYERAWLPSRREWGLYAVYFVITMIGGALAQLPVMAVVSMVAPLHP
LLPLWAQIPLALLLSSLASYAVHRAGHDIPLLWRLHGVHHVPDKVNVGNNGVNHVFDIVLAQFLVQVSLALIGFSEHAVFAVG
IFIIAQGYFVHANIDVRLGWLNHILASPELHRMHHSADKAEAGHFGSDLSIWDRLFGSYTWRPDRKPRAIGLFDPGSFPRNGA
IFSTLVHPLRPRKKPAPTPP
```

ORF2 – 1005 bases (Extended construct):
```
ATGAACACGGACGCCGGTCGCAGTTCAAACCTGACGGACACTGCTCAATCCGACCGTGACCCCGACCCGCTTAATATCGCCGG
ATTATCTTTCGGCGCCGCCGACGGGGTTGAAAACCCAACCCAGGTCACGCCCGCCCGGATAAATCCATTGAGCTTTATTAGTT
ATATCGCGCACCCGCTGTTATTAATTTCCGTGGTTATCTACGTAATTGCCGCGATCGGCCTGGACTGGGATCCGGGCGCCGCC
AGCCTCGGCTTCGTGCTCGCCACCGTCTTCTATCTGGCGATGCTCGAATTAATCATTCCGTACGAGCGCGCATGGCTGCCCTC
GCGGCGCGAGTGGGGCCTGTACGCGGTCTACTTCGTGATCACGATGATCGGCGGCGCGCTGGCGCAGCTGCCGGTGATGGCGG
TGGTGTCGATGGTCGCCCCGCTGCATCCGCTGTTGCCGCTGTGGGCGCAGATCCCGCTCGCGCTGCTGCTGAGCTCGCTGGCG
AGCTATGCGGTGCATCGCGCCGGCCACGACATCCCGCTGCTGTGGCGGCTGCACGGCGTGCATCACGTGCCCGACAAGGTCAA
CGTCGGCAACAACGGCGTCAACCACGTGTTCGACATCGTCCTGGCCCAGTTCCTGGTGCAGGTGTCGCTGGCGCTGATCGGCT
TTTCCGAGCACGCGGTGTTCGCCGTCGGCATCTTCATCATCGCCCAGGGCTATTTCGTCCACGCCAACATCGACGTGCGCCTG
GGCTGGCTCAACCACATCCTCGCCAGCCCCGAGCTGCACCGCATGCATCACAGCGCCGACAAGGCCGAGGCCGGGCATTTCGG
CTCGGACCTGTCGATCTGGGACCGCCTGTTCGGCAGCTATACCTGGCGGCCCGATCGCAAGCCGCGCGCGATCGGCCTGTTCG
ATCCGGGCTCGTTCCCGCGCAACGGCGCGATCTTCTCGACCCTCGTCCACCCGCTGCGGCCGCGCAAGAAACCCGCGCCGACG
CCGCCCTGA
```

Translation ORF2 – 334 amino acids (Extended construct):
```
MNTDAGRSSNLTDTAQSDRDPDPLNIAGLSFGAADGVENPTQVTPARINPLSFISYIAHPLLLISVVIYVIAAIGLDWDPGAA
SLGFVLATVFYLAMLELIIPYERAWLPSRREWGLYAVYFVITMIGGALAQLPVMAVVSMVAPLHPLLPLWAQIPLALLLSSLA
SYAVHRAGHDIPLLWRLHGVHHVPDKVNVGNNGVNHVFDIVLAQFLVQVSLALIGFSEHAVFAVGIFIIAQGYFVHANIDVRL
GWLNHILASPELHRMHHSADKAEAGHFGSDLSIWDRLFGSYTWRPDRKPRAIGLFDPGSFPRNGAIFSTLVHPLRPRKKPAPT
PP
```

5.2 *Saccharophagus degradans* DSM 17024

ORF – 939 bases:
```
GTGGATACCTTAACTCCCTCAGATACGCTCGATTCAGGAATAGCTCAAGAACCAACTCCCCCGCAAACAAAAAACACCAAGAC
GTTTAGCTTTGCCACTTTGCTACGCCACCTTTGTTACCCGCTACTTATGCTGGCCGTGTGCTTAACGTTATTTATTGGCTTTA
CGTTTGATATCGATTTTGGCTTGTGCAACATGGTGTTTCTGGTAGGCACTATTGCCTACTTAGCGCTGTGCGAGCGGTTAATC
CCCTATAAAAAGCAATGGCACCCAACTGCTAAAGAGTGGGGCCGAGACGCTATTTATTTAATTCTCACTATGATGGGTGGTGC
CTCTGCGGTAGCAGTGGTTTTTGCAATAGCAGCTGTAGTATCGCCAAAGGAATCTACTTTAGGACTATGGGTAGAAGTGCCCA
TTGCGATTTTGCTCACGTCGCTTGGTTCCTATATTTTTCATCGCGCGGGGCACGACATACCTTTTTTGTGGCGCTTTCATGGT
ATTCACCACGCGGCAGATAAAATAAACGTCAGTAACAACGCCCTAAATCACATAGCAGATGTATTCGGACGCAGGCTGCTCGC
ACAATTACCCTTAATACTATTAGGCATTTCTACCCCAGCCCTTTTTATTGTGAGTATTTTTAATACAGCGCAAGGTTATTTTT
CTCACGCAAACGTCGACGTAAAAATTGGTTGGTTGAACTATATAGTAGGCAGCCCAGAGCAACATCGTTTACACCATAGTAAA
GATCTCGCCGAGGCCGGCCATTTCAGCGTAGATATCACCCTGTGGGATTTCTTGTTTAAAAGCTATTTTTGGAAGAAAGGAAA
ACAACCAAAGGAAATTGGTGTTACTAATCGCGCTTCGTTTCCGCCCTCTAATAACATTTGGAAAAATATTATTCACCCACTGC
GTCGCAACATTACCTACCGCCAGTAA
```

Translation – 312 amino acids:
```
MDTLTPSDTLDSGIAQEPTPPQTKNTKTFSFATLLRHLCYPLLMLAVCLTLFIGFTFDIDFGLCNMVFLVGTIAYLALCERLI
PYKKQWHPTAKEWGRDAIYLILTMMGGASAVAVVFAIAAVVSPKESTLGLWVEVPIAILLTSLGSYIFHRAGHDIPFLWRFHG
IHHAADKINVSNNALNHIADVFGRRLLAQLPLILLGISTPALFIVSIFNTAQGYFSHANVDVKIGWLNYIVGSPEQHRLHHSK
DLAEAGHFSVDITLWDFLFKSYFWKKGKQPKEIGVTNRASFPPSNNIWKNIIHPLRRNITYRQ
```

6.) Sequencing data of Sterol Desaturase expression vectors

The plasmids' inserts were sequenced by GATC Biotech (Konstanz, Germany). The results, obtained as files of the .ab1 type, were analysed using Geneious software version 8.0.x (Biomatters). Alignments to the according reference sequences were performed using Geneious Aligment default settings.
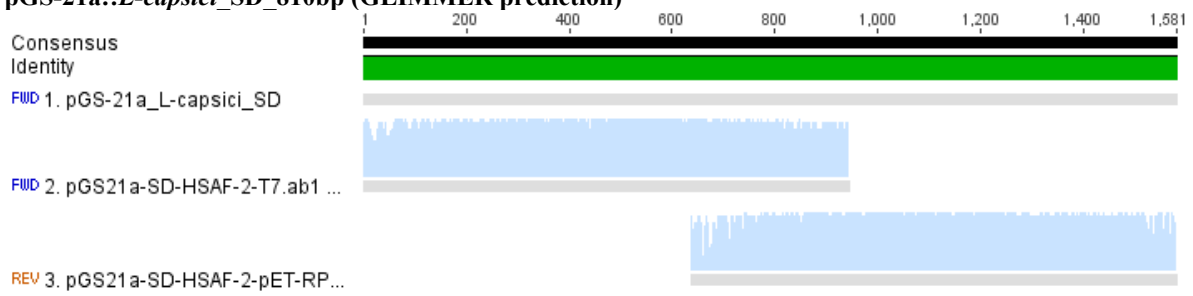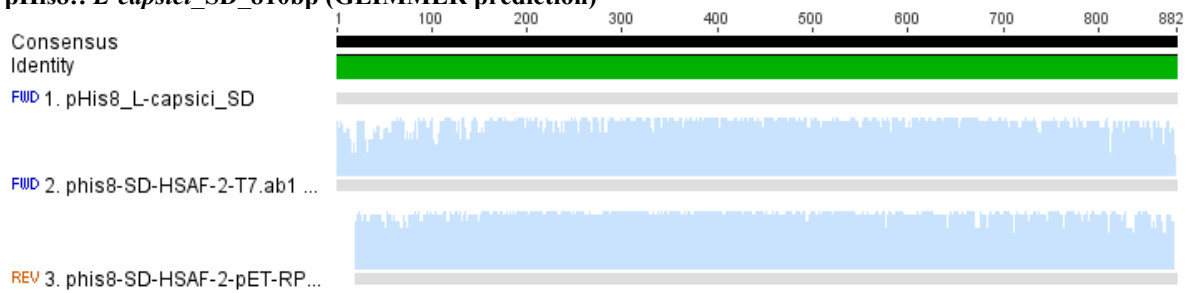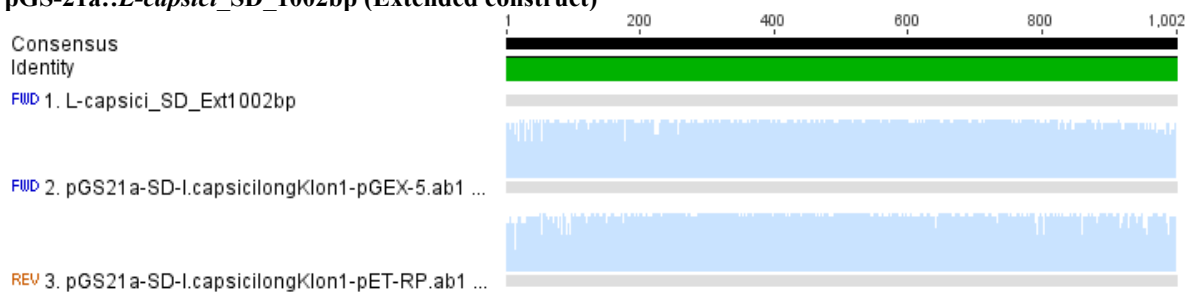
pGS-21a::SD_*S-degradans*
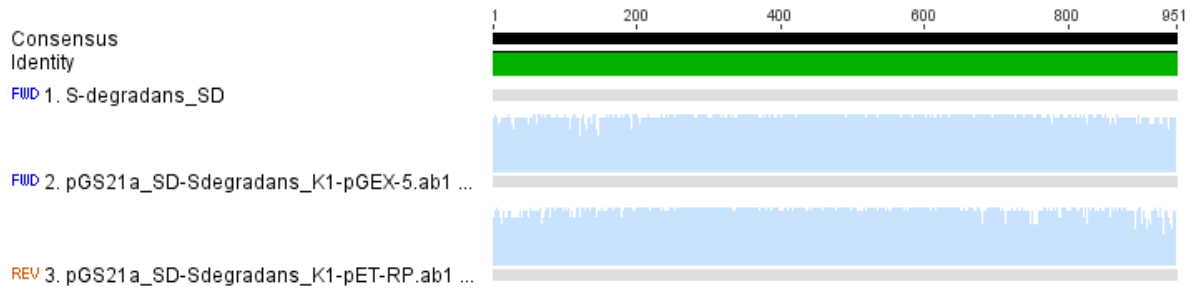
pHis8::SD_*S-degradans*
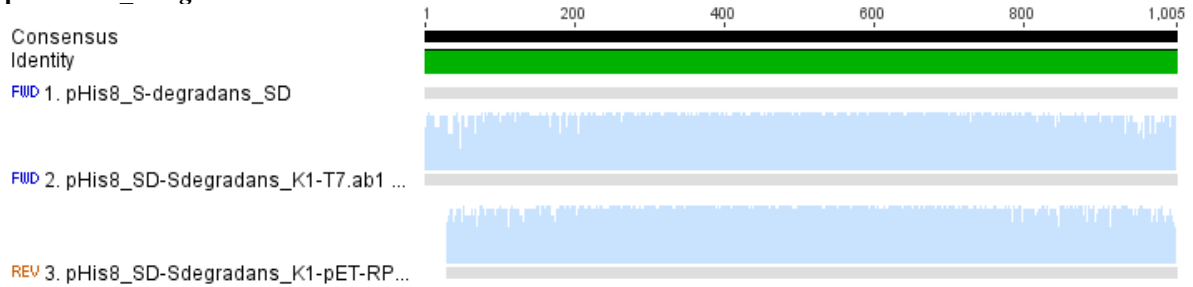
pGS-21a::*L-capsici*_SD_810bp (GLIMMER prediction)

pHis8:: *L-capsici*_SD_810bp (GLIMMER prediction)

pGS-21a::*L-capsici*_SD_1002bp (Extended construct)

pHis8::*L-capsici*_SD_1002bp (Extended construct)

95
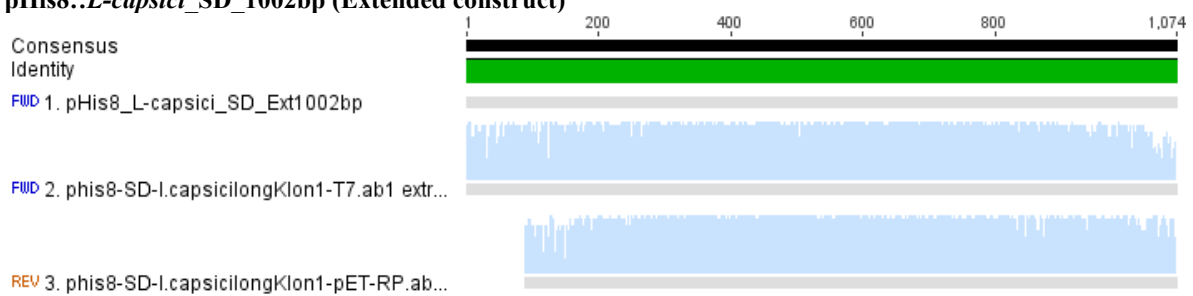
7.) LC-MS traces and MS data

7.1) LC-MS traces of ikarugamycin standard (black), butremycin standard (blue), production of butremycin in vitro (orange) and in vivo (green)
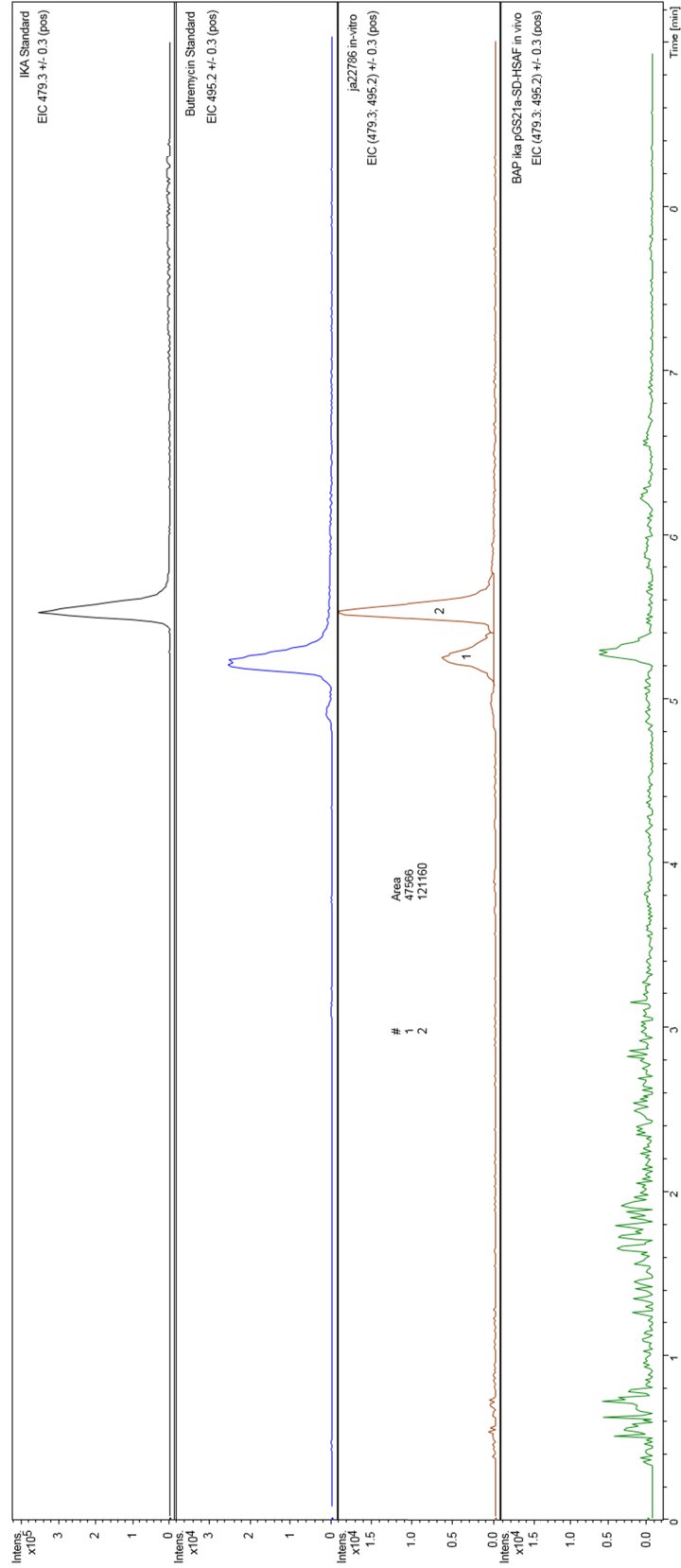
7.2) HR/MS data of ikarugamycin
(calculated values: $[M+H]^+ = 479.2904$; $[M+Na]^+ = 501.2724$; $[2M+H]^+ = 957.5736$; $[2M+Na]^+ = 979.5555$).
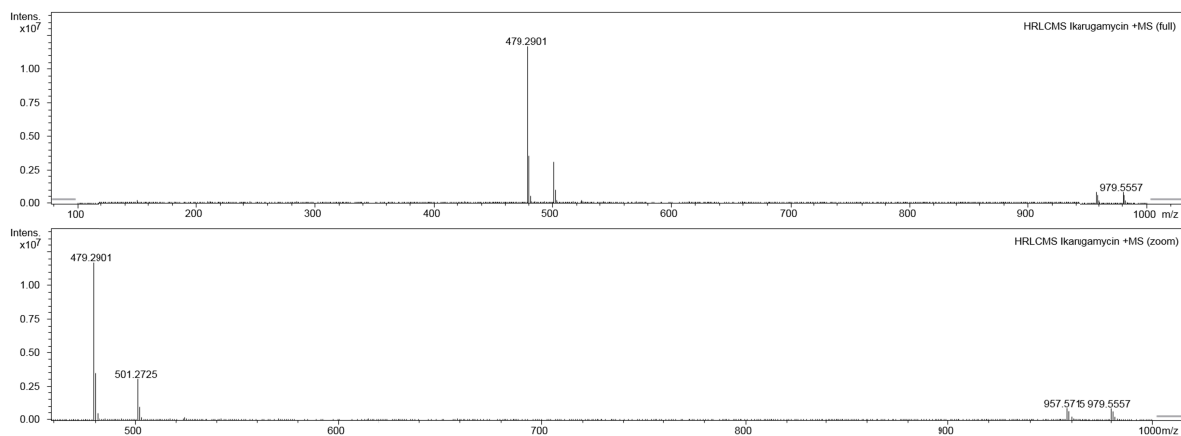
7.3) HR/MS data of butremycin
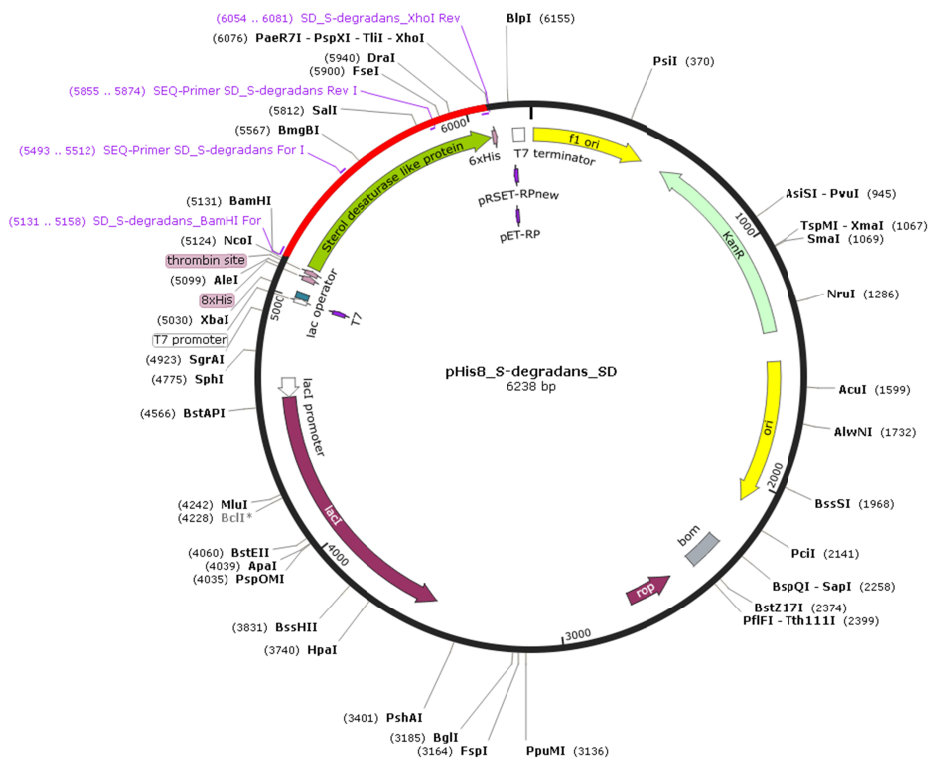(calculated values: $[M+H]^+ = 495.2853$; $[M+Na]^+ = 517.2673$; $[M-H+2Na]^+ = 539.2498$).

97
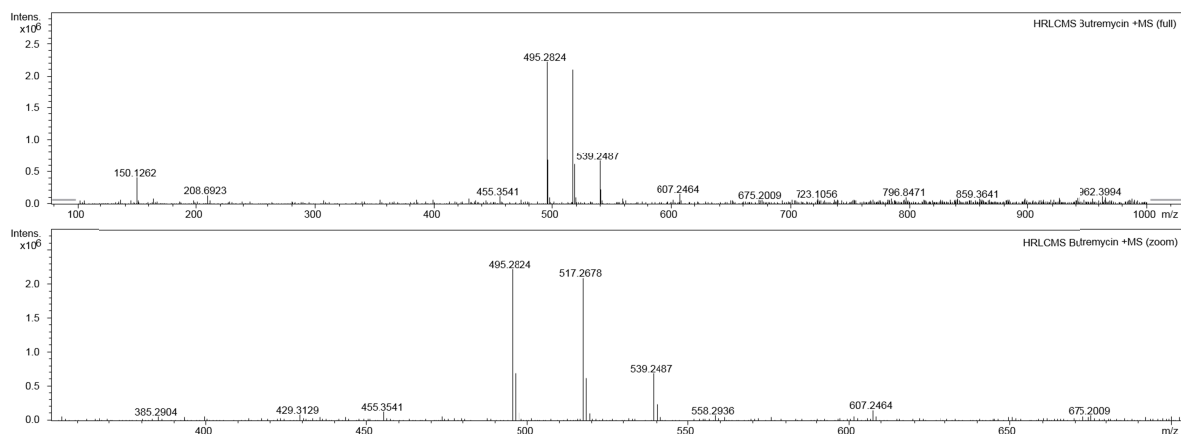
8.) Vector maps
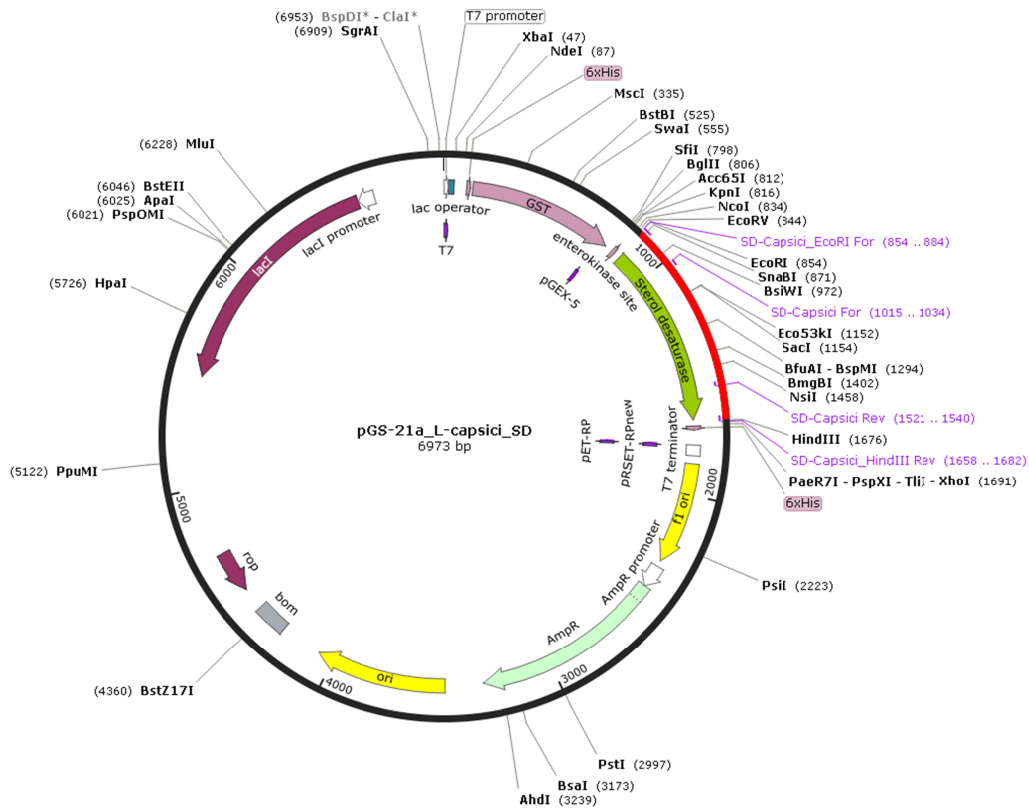
pGS-21a::SD_*S-degradans* [7075 bp]

pGS-21a_S-degradans_SD
7075 bp
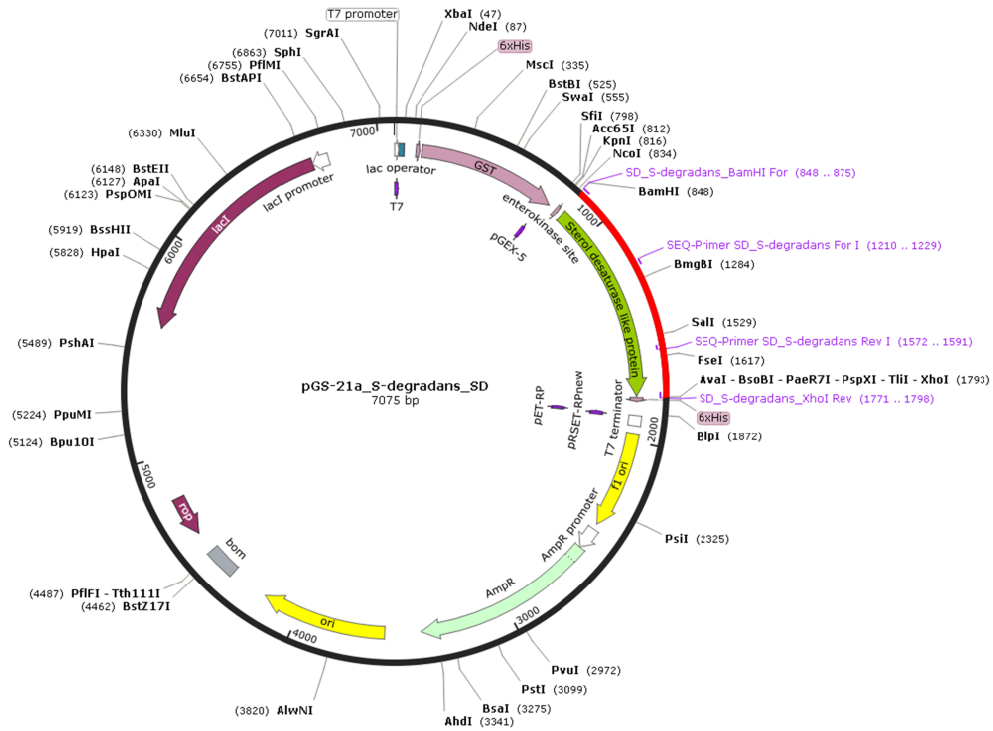
pHis8::SD_*S-degradans* [6238 bp]

pHis8_S-degradans_SD
6238 bp
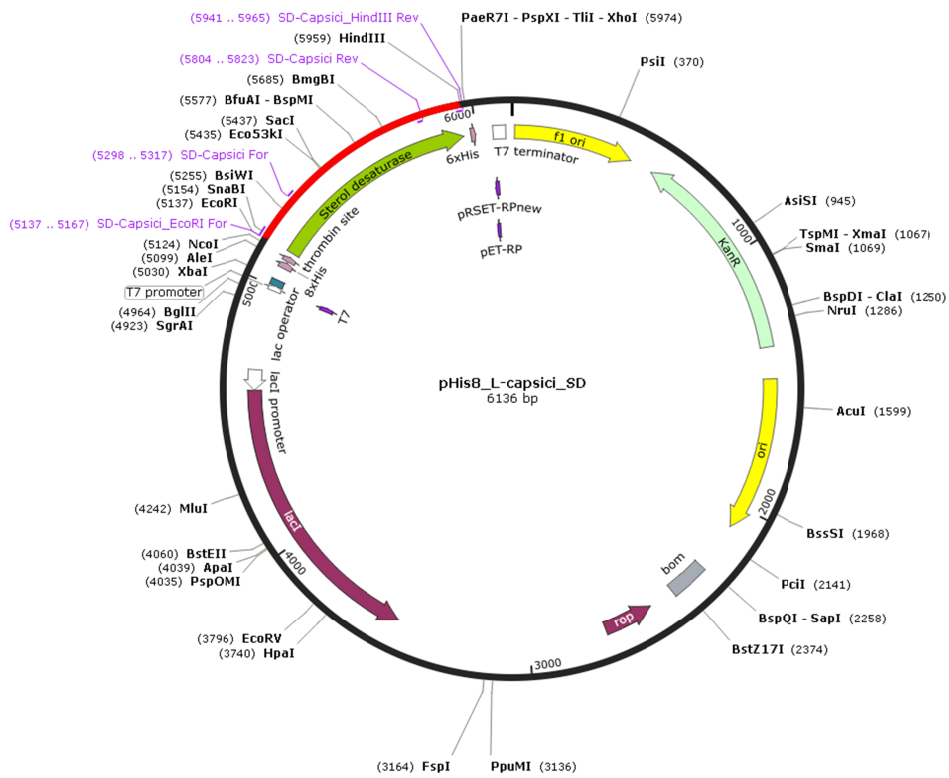
pGS-21a::*L-capsici*_SD_810bp (GLIMMER prediction) [6973 bp]
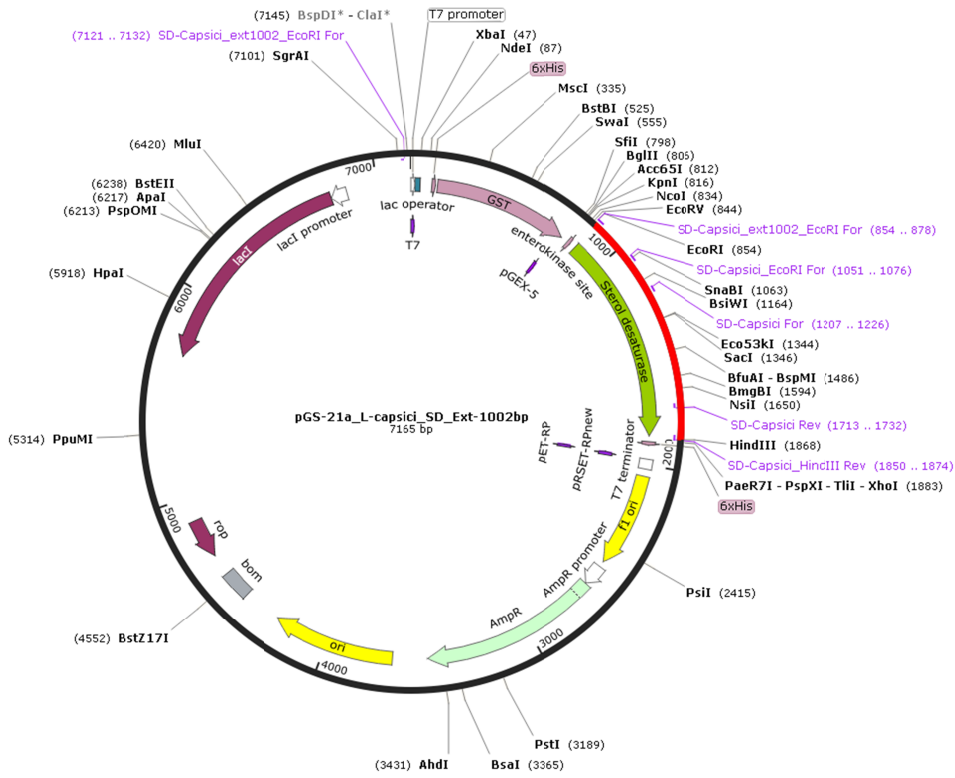

pHis8:: *L-capsici*_SD_810bp (GLIMMER prediction) [6136 bp]

pGS-21a::*L-capsici*_SD_1002bp (Extended construct) [7165 bp]

pHis8::*L-capsici*_SD_1002bp (Extended construct) [6328 bp]
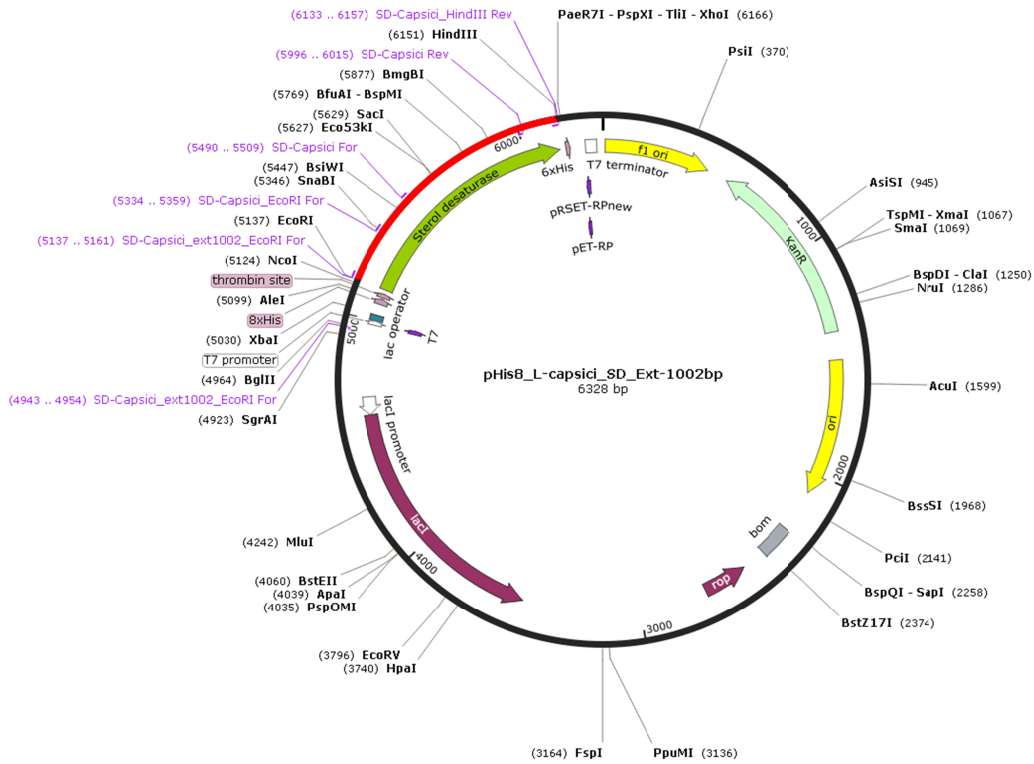
9.) References

1. Pfeifer BA, Admiraal SJ, Gramajo H, Cane DE, Khosla C. Biosynthesis of complex polyketides in a metabolically engineered strain of E-coli. *Science* 2001, **291**(5509)**:** 1790-1792.

2. Studier FW. Protein production by auto-induction in high-density shaking cultures. *Protein Expres Purif* 2005, **41**(1)**:** 207-234.

3. Antosch J, Schaefers F, Gulder TA. Heterologous reconstitution of ikarugamycin biosynthesis in E. coli. *Angewandte Chemie (International ed in English)* 2014, **53**(11)**:** 3011-3014.

A.1.2. Ikarugamycin total synthesis – Greunke *et al.* 2017

The supplemental information is related to the following publication which was highlighted in Chapter 3.2:

- **Greunke, C.**; Glöckle, A.; Antosch, J.; Gulder, T.A.M. Biocatalytic Total Synthesis of Ikarugamycin. *Angew. Chem. Int. Ed. Engl.* **2017**, *56*, 4351-4355, DOI: `10.1002/anie.201611063`.

Biocatalytic Total Synthesis of Ikarugamycin

Christian Greunke, Anna Glöckle, Janine Antosch and Tobias A. M. Gulder[*]

Biosystems Chemistry, Department of Chemistry and Center for Integrated Protein Science Munich (CIPSM), Technische Universität München, Lichtenbergstraße 4, 85748 Garching bei München, Germany. *E-mail: tobias.gulder@tum.de.

Supporting Information

Table of contents

1. Materials and Methods

Commercial materials. The listed commercial materials were purchased from the following manufacturers: Luria-Bertani (LB) Broth Miller, LB Agar Miller, Terrific Broth modified (TB), tris(hydroxymethyl)aminomethane (Tris), 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES), kanamycin sulfate (Kan), ampicillin sodium salt (Amp), Rotiphorese 10x SDS-PAGE buffer concentrate, Rotiphorese 50x TAE buffer concentrate, ammonium persulfate, Rotiphorese Gel 40 (37.5:1) from Carl Roth (Karlsruhe, Germany). Imidazole from BASF (Ludwigshafen, Germany). Magnesium chloride hexahydrate (MgCl$_2 \cdot$ 6 H$_2$O) and sodium chloride (NaCl) from Grüssing (Filsum, Germany). Acetyl- and malonyl-CoA lithium salts, glycerol, malonic acid and sodium acetate from Sigma-Aldrich (Taufkirchen, Germany). Deoxynucleotides (dNTPs), Q5 High-Fidelity DNA polymerase, Antarctic phosphatase, ColorPlus™ Prestained Protein Marker (7-175 kDa) and restriction enzymes *Hin*dIII, *Xho*I from New England Biolabs (Frankfurt am Main, Germany). Zinc chloride (ZnCl$_2$), sodium dihydrogenphosphate dihydrate (NaH$_2$PO$_4 \cdot$ 2 H$_2$O) and *N,N,N',N'*-tetramethylethylendiamine (TEMED) from Merck (Darmstadt, Germany). T4 DNA ligase, High Density Nickel Agarose, Bacterial DNA Preparation Kit and restriction enzymes *Bam*HI, *Eco*RI from Jena Bioscience (Jena, Germany). Isopropyl β-D-1-thiogalactopyranoside (IPTG), nicotinamide adenine dinucleotide phosphate tetrasodium salt (NADPH), flavin adenine dinucleotide disodium salt hydrate (FAD), coenzyme A trilithium salt (85 %) and L-ornithine hydrochloride from Carbolution Chemicals (Saarbrücken, Germany). Adenosine 5'-triphosphate acid disodium salt (ATP) from PanReac AppliChem (Darmstadt, Germany). Oligonucleotides from Sigma Aldrich were resuspended at a stock concentration of 100 µM in pure water and stored on ice for immediate use or at -20 °C for long term storage. Kits for plasmid isolation and DNA extraction from agarose gels, peqGOLD Plasmid Miniprep Kit I (C-Line) and peqGOLD Gel Extraction Kit (S-Line), centrifugal filters 3 KDa (500 µl), PD-10 columns, peqGOLD Protein-Marker VII 'prestained' (40-300 kDa), Vivaspin 2 Hydrosart membrane columns (10,000 and 30,000 MWCO) and PES 3kD centrifuge columns from VWR (Darmstadt, Germany). Pierce™ Unstained Protein Molecular Weight Marker (14.4-116 kDa) from Thermo Fisher Scientific (Schwerte, Germany).

Bacterial strains. Escherichia coli DH5α was used as the host for cloning procedures. *E. coli* strains were grown in LB medium or on LB agar. After transformation of chemically competent *E. coli* cells, SOC medium was added to the cellular suspensions. *E. coli* BL21(DE3) was used as heterologous host for the recombinant production of the enzymes MatB, Pta, AckA, Tü6239-PKS/NRPS (IkaA), Tü6239-FAD (IkaB) and SA-AlcD (IkaC) as octahistidyl-tagged fusion proteins from the pHis8 vector system. For lysate reactions, *E. coli* BAP1[1] was used to produce *N*-His8-Tü6239-PKS/NRPS in its pantotheinylated form. Additionally, Tü6239-FAD (IkaB) was expressed as a *N*-GST fusion protein (construct: pGS-21a::Tü6239-ikaB).

Preparation of bacterial cells for chemical transformation.[2] E. coli DH5α or BL21(DE3) LB overnight cultures (DH5α: 10 mL; BL21(DE3): 1 mL) were prepared to inoculate 250 ml of SOC medium and cultivated overnight at 18 °C while shaking (180 rpm) until an OD$_{600}$ value of 0.4 to 0.6 was reached. Cells were then stored on ice and pelleted afterwards for ten minutes at 4000 rpm, 4 °C. The supernatant was discarded and cells were resuspended in ice-cold, sterile transformation buffer (10 mM HEPES, 15 mM CaCl$_2 \cdot$ 2 H$_2$O, 250 mM KCl, 55 mM MnCl$_2 \cdot$ 2 H$_2$O, pH 6.7). Cells were again incubated on ice and pelleted by centrifugation for ten minutes at 4000 rpm, 4 °C. The supernatant was discarded and cells were resuspended in 18.4 mL ice-cold transformation buffer, supplemented with 1.4 mL DMSO (final volume content 7 %). Aliquots of 100 µl were prepared, flash-frozen in liquid nitrogen and stored at -80 °C until use.

DNA workup. PCR products were purified either by the use of the PCR Purification Kit (Jena Bioscience) or by re-isolation after agarose gel electrophoresis using the peqGOLD Gel Extraction Kit, S-Line (Peqlab) following the instructions of the manufacturer's protocols. The preparation of plasmids was conducted using the peqGOLD Plasmid Isolation Kit, S-Line (Peqlab).

gDNA preparation. All procedures were conducted by use of the Bacterial DNA Preparation Kit (Jena Bioscience), according to the manufacturer's protocol.

Cloning. The ikarugamycin biosynthetic genes required for this study were amplified from fosmid DNA template covering the ikarugamycin clusters of either *Streptomyces* sp. Tü6239 or *Salinispora arenicola* CNS-205. The Tü6239-PKS/NRPS gene was amplified using Q5 polymerase (NEB) in combination with High GC Enhancer supplement. The gene coding for the malonyl-CoA ligase MatB was amplified using gDNA prepared from *Streptomyces coelicolor* stain M1152 as PCR template.[3] To amplify this GC-rich gene (GC content ~ 76 %), addition of High GC Enhancer was necessary. Genes coding for AckA and Pta were amplified using gDNA prepared from *E. coli* BW25113 as PCR template.[4]

Purified PCR products were digested with appropriate restriction enzymes to generate sticky ends compatible for a ligation with linearized and dephosphorylated vector backbones of pHis8, pHis8-TEV, pHis8C-TEV or pGS-21a using T4 DNA ligase (Jena Bioscience). Restriction enzymes were purchased from NEB and Jena Bioscience and were heat-inactivated before ligation was started. Antarctic phosphatase for dephosphorylation of linearized plasmids was purchased from NEB.

The integrity of plasmids was first verified by restriction analysis. Inserts of positive clones were sequenced at GATC Biotech (Konstanz, Germany). Results were retrieved as ab1-type files and were analyzed using Geneious software version 8.x (Biomatters). Alignments to the according reference sequences were performed using Geneious Aligment default settings. Maps of plasmids used in this study were constructed using the SnapGene software (from GSL Biotech; available at snapgene.com).

Protein purification from E. coli. Plasmids pHis8::Tü6239-ikaA, Tü6239-ikaB and SA-ikaC, as well as pHis8C-TEV::Tü6239-ikaA were transformed into *E. coli* BL21(DE3). Transformants were grown at 37 °C in LB medium containing 50 µg/ml kanamycin sulfate overnight. These starter cultures were used to inoculate TB supplemented with antibiotic. When an OD_{600} value of 0.6 was reached, cultures were cooled on ice for 20 min and protein synthesis was induced by adding 0.5 mM IPTG, followed by an overnight incubation at 16 °C with shaking (200 rpm). *E. coli* BL21(DE3) pHis8/pHis8C-TEV::Tü6239-ikaA cultures were incubated at 20 °C overnight while shaking.

Cells were harvested by centrifugation and resuspended in lysis buffer (50 mM $NaH_2PO_4 \cdot 2 H_2O$, 300 mM NaCl, 10 % glycerol, 10 mM imidazole, pH 8.0). After sonication on ice and subsequent centrifugation (10,000 rpm, 4 °C, 30 min), the supernatants were incubated with High Density Nickel Agarose (Jena Bioscience) and moderate shaking for one hour on ice. This mixture was then applied onto a column, which was washed with lysis buffer (40 mM imidazole, pH 7.5). Finally, the protein was eluted with lysis buffer (250 mM imidazole, pH 7.5). PD-10 columns (GE Healthcare) were used to exchange the eluate's buffer, for IkaA/C: 50 mM Tris-HCl, 10 % glycerol, pH 7.5; IkaB: 50 mM HEPES, 100 mM NaCl, 20 % glycerol, pH 7.5; MatB, AckA, Pta: 50 mM HEPES, 100 mM sodium chloride, 2.5 mM EDTA, 20 % glycerol, pH 7.8. The protein solutions were then concentrated using Vivaspin 2 Hydrosart membrane columns (Sartorius), resulting in recombinant enzymes with an approximate purity of 80 % (IkaA) up to > 90 % (IkaBC, Sfp, MatB, AckA, Pta).

Protein concentrations were determined photometrically using the Implen Nanophotometer 330. Bacterial induction and protein enrichment were monitored by SDS-PAGE analysis. Final protein concentrations after Vivaspin treatment were measured in a spreadsheet using absorbance values (wavelength = 280 nm) from http://protcalc.sourceforge.net/: *N*-His8-IkaA: 22.4 mg ml^{-1}, $\varepsilon_{280\,nm}$ = 347,820 M^{-1} cm^{-1}; *C*-His8-IkaA: 8.09 mg ml^{-1}, $\varepsilon_{280\,nm}$ = 349,100 M^{-1} cm^{-1}; *N*-His8-IkaB: 6.7 mg ml^{-1}, $\varepsilon_{280\,nm}$ = 89,830 M^{-1} cm^{-1}; *N*-His8-IkaC(SA): 10.0 mg ml^{-1}, $\varepsilon_{280\,nm}$ = 57,750 M^{-1} cm^{-1}; *C*-His6-Sfp: 30.5 mg ml^{-1}, $\varepsilon_{280\,nm}$ = 26,980 M^{-1} cm^{-1}; *N*-His8-MatB: 23.8 mg ml^{-1}, $\varepsilon_{280\,nm}$ = 33,000 M^{-1} cm^{-1}; *N*-His8-AckA: 22.0 mg ml^{-1}, $\varepsilon_{280\,nm}$ = 23,970 M^{-1} cm^{-1}; *N*-His8-Pta: 17.2 mg ml^{-1}, $\varepsilon_{280\,nm}$ = 36,630 M^{-1} cm^{-1}.

Enzymatic reactions

Enzymatic reactions to test in vitro production of acetyl-CoA.

Reactions were set up in PBS pH 7.4 in a total volume of 100 µl containing 10 mM sodium acetate, 1 mM CoA, 2.5 mM ATP, 10 mM $MgCl_2$, 1 µM AckA and 10 µM Pta. After an incubation period at 30 °C for one hour, the

reactions were centrifuged at 16,000 × g for two minutes and supernatants were analysed by HPLC (λ = 254 nm) (Eurospher II 100-3 C18 A, 150 x 4.6 mm) using the following gradient with water (A) and acetonitrile (B), both buffered with 0.1 % formic acid, as the eluents: linear gradient from 5 % B (3 min), 50 % B (15 min), 70 % B (20 min), 5 % B (24 min), all at a flow rate of 1.0 mL/min.

Enzymatic reactions to test in vitro production of malonyl-CoA.

Reactions were set up in PBS pH 7.4 in a total volume of 100 µl containing 22 mM malonic acid (dissolved in PBS pH 7.4), 0.55 mM CoA, 1.3 mM ATP, 9 mM $MgCl_2$, 5 µM MatB. After an incubation period at 30 °C overnight, the reactions were centrifuged at 16,000 × g for two minutes and supernatants were analysed using the identical conditions as for acetyl-CoA above.

Enzymatic reactions to produce ikarugamycin and the IkaB intermediate. Reactions were set up in HEPES buffer (100 mM, 1.5 % glycerol, pH 7.8) with a volume of 100 µL containing 2 mM acetyl-CoA, 10 mM malonyl-CoA, 2 mM L-ornithine, 2 mM coenzyme A, 10 mM ATP, 10 mM NADPH, 2 mM FAD, 7.5 mM $MgCl_2$, 10 µM *N*-His8-IkaA or 5 µM *C*-His8-IkaA, 10 µM IkaB, 10 µM IkaC, and 5 µM Sfp. The employed enzymes were freshly produced and isolated. The reactions were incubated overnight at 30 °C with shaking (350 rpm). Reactions were quenched by addition of 100 µL MeOH and vigorous vortexing. After centrifugation at 16,000 x g for two minutes, supernatants were carefully transferred onto PES 3kD centrifuge columns. The resulting flow-through after centrifugation at 10,000 x g for 30 minutes was used for LC-MS analysis. Samples were analyzed using an UltiMate 3000 LC System system coupled to a LCQ Fleet Ion Trap Mass Spectrometer (Thermo Scientific). Interpretation of the recorded data was performed using the Thermo Xcalibur Qual Browser 2.2 SP1.48 software. The chromatographic HPLC separation was carried out on a XBridge (Waters) C-18 HPLC column (30 x 4.6 mm, 3.5 µm particle size). The following gradient with water (A) and acetonitrile (B) as the eluents, both buffered with 0.1 % formic acid, was applied: 20 % B (0 min) → 95 % B (6 min) → 95 % B (7.5 min) → 20 % B (8 min) → 20 % B (9 min). The flow rate was kept constant at 1.1 mL/min. For preparative production of ikarugamycin (**1**), the PES filtration step was omitted. The enzymatic reaction was conducted as above in a total of 3.7 mL reaction volume. Ikarugamycin (**1**) was extracted from the aqueous enzyme buffer solution by exhaustive extraction with hexane. The combined organic phases were washed with brine, dried over $MgSO_4$, filtered, and the solvent evaporated. The crude extract was directly submitted to isolation of **1** by preparative HPLC using a Eurospher II (100-5, C8A, 125 x 8 mm) column (Knauer, Berlin). The following gradient with water (A) and acetonitrile (B) as the eluents, both buffered with 0.01 % TFA, was applied: 60 % B (0 min) → 60 % B (2 min) → 100 % B (12 min) → 100 % B (16 min) → 60 % B (16.5 min) → 60 % B (20 min). The flow rate was kept constant at 3 mL/min. Ikarugamycin eluted from the column after 10 min. Acetonitrile was removed under reduced pressure from the combined fractions containing **1**. Residual water was removed by freeze-drying overnight. This led to the isolation of 0.32 mg (9 %) of pure ikarugamycin (**1**) as white powder. NMR spectra of **1** were recorded on a Bruker AV500 system equipped with a cryo probe.

Enzyme assays using cellular lysate supernatants. LB overnight cultures of *E. coli* BAP1 pHis8::Tü6239-ikaA, *E. coli* BL21(DE3) pGS-21a::Tü6239-ikaB and pHis8::SA-ikaC were used to inoculate TB medium, supplemented with appropriate antibiotics. Three cultures with a volume of 0.6 L (IkaA) and 0.3 L (IkaB and C) were prepared and grown at 37 °C while shaking at 200 rpm to an OD_{600} value of about 0.6 to 0.8. Cultures were cooled on ice for 20 min and cells were induced by adding IPTG to a final concentration of 0.5 mM. After an overnight incubation at 16 °C while shaking (200 rpm), cells were harvested by centrifugation and were resuspended in lysis buffer (50 mM $NaH_2PO_4 \cdot 2\ H_2O$, 300 mM NaCl, 10 % glycerol, 10 mM imidazole, pH 8.0). Per gram of cell pellet, 4 ml of buffer were added. The *E. coli* cells were lysed on ice by sonication. After centrifugation (10,000 rpm, 4 °C, 30 min), the supernatants were combined in an Erlenmeyer flask. After a four hour incubation at 30 °C while shaking (150 rpm), the resulting mixture was extracted with ethyl acetate twice. The solvents were evaporated in vacuo and the residue was dissolved in 500 µl methanol:water (1:1). After centrifugation at 16,000 × g for two minutes, the supernatant was used for LC-MS analysis as above.
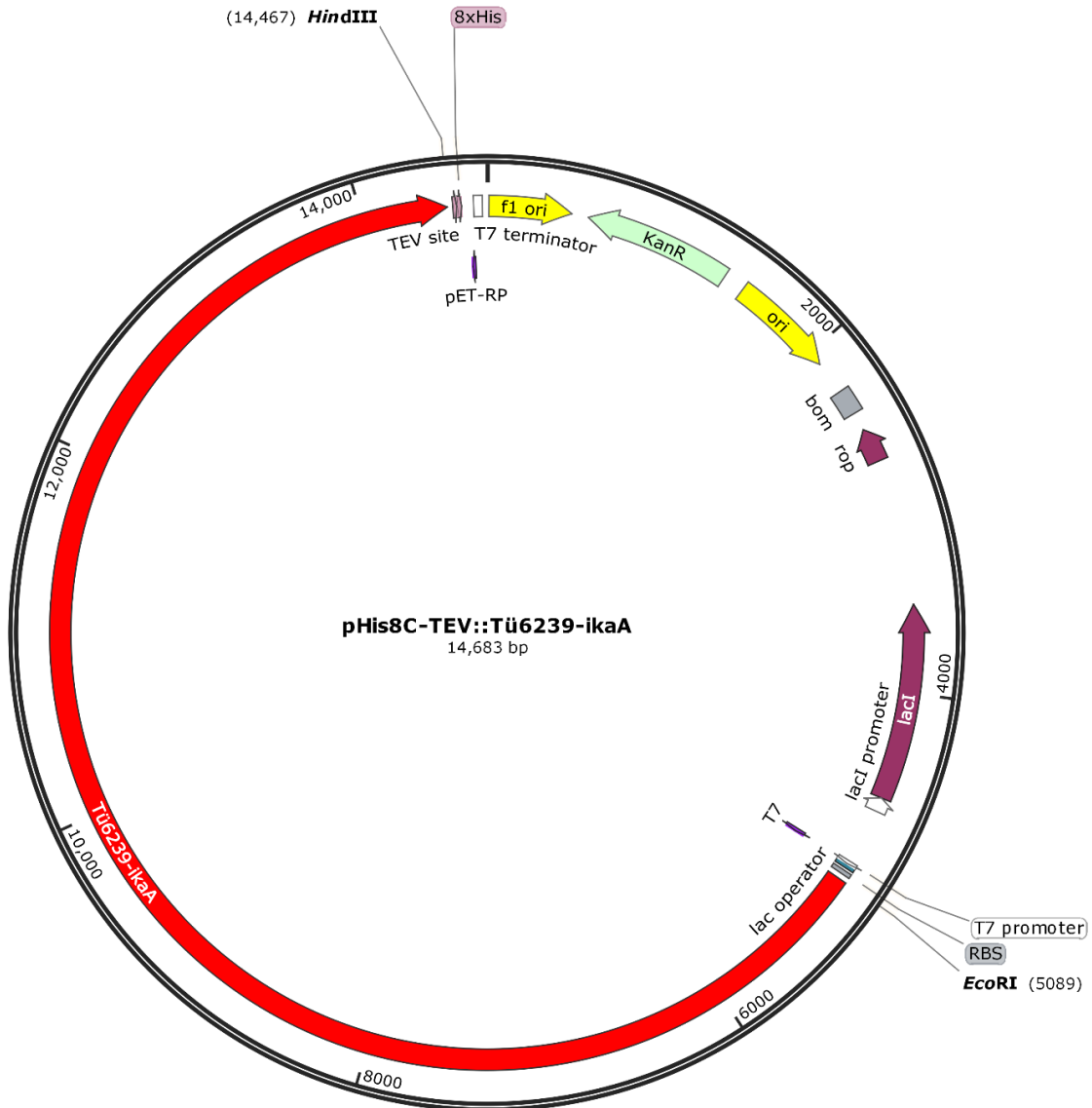
2. Characterization of the employed expression constructs and proteins

a. IkaA

- vector map

Figure S1. Vector map of the expression construct of *ikaA* in pHis8.

Figure S2. Vector map of the expression construct of *ikaA* in pHis8C-TEV.

- DNA sequence of Tü6239-PKS/NRPS (*ikaA*) [9375 bp]; GenBank: KY402202:

```
ATGGATTCCATGCACCACCCTGCCCCCGTCCCCGTACCCGAAGTCCCCGCGCCCGTCCCGTCCCAGGACGACGCGTTCGC
CATCGTCGGCATCGGCTGCCGGCTGCCCGGCGGCGCCAGCGACTACCGGACCTTCTGGCGCAACCTCCTCGACGGCAAGG
ACTGCATCACCGACACCCCCGCCGACCGCTACGACACCCGCACCCTGGGCAGCGGCGACAAGGCCAAGCCCGGCCGGCTG
GTCGGCGGACGCGGTGGATACATCGACGGCTTCGACGAGTTCGACCCCGCCTTCTTCGGCATCAGCCCGCGCGAGGCCGA
GCACATGGACCCCCAGCAGCGGAAGCTCCTGGAGGTCGCCTGGGAGGCGCTGGAGGACGGCGGCCTCAAGCCCGCCGAGC
TGGCCGGCAGCGATGTCGGGGTGTACGTCGGGGCGTTCACCCTCGACTACAAGATCCTGCAGTTCGCCGACCTCGGCTTC
GAGACCCTGGCCGCGCACACCGCCACCGGCACCATGATGACGATGGTGTCCAACCGGATCTCGTACTGCTTCGACTTCCG
CGGACCCTCGGTCTCCGTCGACACCGCGTGCAGCGGCTCCCTGGTCGCCGTCCACCTCGCCTGCCAGAGCCTGCGCCGCG
GCGAGACCTCCGTCGCCCTGGCCGGCGGCACCCTGCTGCACATGGCGCCGCAGTACACCATCGCCGAGACCAAGGGCGGG
TTCCTCTCCCCCGACGGCCGCTCCCGCGCCCTGGACGCCTCCGCCAACGGCTACGTGCGCGCCGAGGGCGTCGGCATGGT
CGCCATCAAGCGCCTCGCGGACGCGCAGCGCGACGGCGATCCCATCCACGCCGTCATCATCGGCAGCGGCGTCAACCAGG
ACGGCCGCACCAACGGCATCACCGTGCCCAACCCCGACGCGCAGGTCGCCCTGATCGAGCGGGTCTGCGCCGCCGCCGGC
```

```
GTCACCCCCGGCAGCCTCCAGTACGTCGAGGCGCACGGCACCTCCACCCCCGTCGGCGACCCGCTGGAGGCCAACGCCCT
CGGCCGCGCGCTCTCCATCGGCCGCGAGCCGGGCGCCCGGACGTACGTCGGCTCGGTCAAGACCAACATCGGGCACACCG
AGTCCGCCGCCGGCATCGCCGGGCTGATCAAGACGGTGCTCAGCCTCAAGCACAAGGTCATCCCGCCGCACATCAACCTG
GAGAAGCTCAACCCGCAGATCGACGAGGCGTCCCTGCCGTACGAGATCCCCCGCGAGCCCACCCCCTGGCCCGAGCACAG
CGGGCCGGCCCGGGCCGGCGTCAACTCCTTCGGCTTCGGCGGGACCAACGCCCACGTCCTGCTCCAGGAGGCACCGCCGA
CCGTCGGGGAGCCCGCGCCACCGGCCACCGACGGGTACTCCGTGCTGCCGCTCAGCGCCCGCGACCCCGAAGCCTTTCCC
GCCATCGCCACCGGCCTGCGCGAACGGCTCGCCGAGGGACTGCCGGTGGGCGACGCCGCCTACACCCTCGCCCACCGGCG
GCAGCATCTGGAGCAGCGGCTGTCCGTCGTGTACGACTCCCCCGAGGCCCTCGACGAGGTGCTCGGCGCCGTCGCCCGCG
GCGAGAGCCACCCGCGTGCCGTCGCCGGCACCCAGCGGGAGGGCCTGGACCGCAGGCTGGTGTGGGTGTTCACCGGCATG
GGCCCGCAGTGGTGGGCCATGGGCCGCCAGTTGTACGCGAGCGAGCCCGTCTACCGGGAGGTCATCGACCGCTGCGACCA
GGAGATCGCCGCGCTCACCGGCTGGTCCCTCACCCAGGAGCTGAACGCCGACGAGGCCGACTCCCGGATGAGCGAGACCT
GGCTCGCCCAGCCCGCCAACTTCGCCGTCCAGATCGCCCTGGCCGCCCTGTGGCGCAGCAAGGGGATCCAGCCCGACGCC
GTCACCGGGCACAGCACCGGTGAGGTCGCCGCGTTCTACGAGGCCGGGGTGTACACCCTCCCCGAGGCCGTGAAGATCGT
GGTGCACCGCAGCCGGCTCCAGCAGAAGCTCATCGGCACCGGCTCCATGCTCGCCGTCAGCCTCACCGAGGCGGAGGCCG
CCCGCCGGGTGCGCCCGCACGGCGACCGGGTCTCCATCGCCGCCGTCAACAGCCCCACCTCCATCACCCTGGCCGGGGAC
ACCGAGGCGCTGGAGGTGATCGCCGCCGAGCTGGGCGCCGAGGACATCTTCGCCCGCTTCCTGGAGGTCGGCGTCCCGTA
CCACAGCCCCCGCATGGAGCTGATCAAGGACGAGCTGCTGACCTCGCTCGCCGATCTCAAGCCGCAGCAGGCGAAGTTGC
CGCTGTACCTCACCGCGCTGCCGGGCACCGTCGCGCAGGGCACGGAGCTGGACGCCGACTACTGGTGGCGCAATGTGCGC
GAGGCCGTGCACTTCCGGGCCGCCGTGGACCGGCTGCTGGACGACGGCTACGGCGTCTTCCTGGAGATCGGCCCGCACCC
CGTGCTCGCCCACTCCCTGCGCGAGTGCTGCGAGGCCCGCGACGCGCACAGCGTCACCCTGGCCTCCATCCGCCGCAAGG
CGGACGAGCGCGAACGCCTCACCCTGTCGCTCGCCGCGCTGCACAGCCTCGGCTTCGCCGTGGACTGGCACGCCCTGCAC
CCCGCCGGGCGGCCGGCCGAACTGCCGCGCTACCCGTTCCGGCGCGACCGGTACTGGGTCGAGCCGGCCCCGGTCGCGCA
GATCCGGCTCGGCCACCGCGACCACCCGCTGCTGGGCCGCCGCACCGCGAGCGCCGAGCCGGTGTGGGAGGTGAAGCTGG
ACGCGGAGGCCGCCCCGTACCTGGAGGACCACCGCATCCAGGGCACCGTGCTGTTCCCGGCCGCCGGCTATCTGGAGATG
GCCGCGCAGGCCATGCGGGCGCTGACCGGTGATGAGCACAGCACCGCCGCGCTGGCCGGCATCGAGCTGCGCAAGGCGCT
GTTCCTGCCGGACGGCGAGCCGCAGACGGTGCAGCTGTCCTTCTCCTCCGACGCCGCCGCGTTCTCCATCGCCACCGTGG
GCGCCGCCGGCGCCGAGCCGACCGTGCACGCCACCGGTACGGTACGGGCCGCCCAGCGCCGCCGGCTGACCGCGCCGCTG
GACACCGTCGCCGTCCGGGCCCGCGCCGCCCGCCACCTGAGCGGCCCCGACTGCTACGCCGAACTGGCCGCGCTCGGCTA
CCACTACGGCCCCGCCTTCCAGGGCATCGAGGAGGTGTGGATCGGCGAGGGCGAGGCCCTGGCCCGGATCCGTCCGCCGC
AGGGGCTCACCCCGGACGCGGCGGCGCACCACATGCATCCGGTGCTGCTCGACTCCTGCTTCCAGTCGCTGCTGACCCCG
CAGCTGCTCACCGCGCCCGCCGGGCCCGGGGGCACCGGCATCCGGCTGCCGCTGTCCATCGCCGAGGTACGGCTGGACCC
GGTCGGCGACCGCGAACTGTGGGTGCACGCCACCGTCACCGGCGACGACGAGGACGAACTCACCGGTGACATCGCCGTGT
ACGACGGCGCCGACGGTACGCCGCTGGGCCGCGTCGCCGGCTTCCGCGCCGCCGATGTGGAGAAGGCCGCCACCACCGTG
GGGCTGTCCACCATCGACAGCTGGCTCACCGAACCGAGCTGGGTGCCGTGCCCGCTGCCCGAGGCGGCGTCCGCCGCGCC
GGCGGCCGGGCGGCACGTACTGTTCGCCGACGCGGGCGGGGTCGCGCAGCGGCTGGCCGCGCTGATCGGCGAGGCCGGCG
GGGAGGCCCATCTGGTCCGGCCCGGTGCCGCGTACGGCCTGGACCGCACGGCGAGGACCGCCACCGTCGTCCCCGGATCC
GCGGATGACCTGCGGCGGTTGCTCACCGATCTCGGGCAGGTGGACGGCGTCGTCCACCTGTGGAACCTGGACCGGCCGGC
GCTGGCCGACGCCCCGCGCGGACGGTTCGCGGACATCGCCTCCACCGGCGCGTACGCCCTGATCGCCCTCACTCAGGCCC
TGCTCGCCGACCCGGAGCGGCACGGCGGCACCCCGGTGCACATCGTCACCAGAGCCGCCCAGTGCGTGGTCCCCGGTGAG
CCGGTGGAGCCGCTGGGCGCGCCCGCCTGGGGCATCGGCCGGGTGCTGTGGCAGCAGGAACTGGCCGGGCGCGGCGGCAA
GCTGATCGACCTGGCGGCCGACGGCGGCGTCGAGGAGGACGCGTACGCGCTGCTGCGCGAGCTGGCCGACCCCACCGGCG
CGGCCGAGCGCGAGGACGAGATCGCGCTGCGCGCCGGGGAGCGGCACACCAGCCGGCTGGTGGCCGCCGAGGGGCTGAGC
AGGCCGCTGCCCCTGCGGCTGCGCCCGGACGGCAGCTATCTGGTGACCGGCGCGTTCGGCGCGCTCGGCAGGCTGCTGTG
CCGCACGCTGGTCAGGCGCGGGGCGCGGCGGCTGATCCTGGTGGGCCGCACCCGGCTGCCGGAGCGCGAGCGCTGGGCCG
ACCAGGACCCGAACTCGCCGGCCGGGCGGCACGTGGCCTTCCTCAAGGAGCTGGAGGCGCTGGGCGCGCAGCCGATTCTC
GCGCCGCTGGACATCACCGACGAGGACGCGCTGGCCGGCTGGCTCGCCGGGTACCGGCGCGCCCAGGGGCCGCCGATCCG
CGGGGTGTTCCATCTGGCGGGGCAGGTGCGCGACACCCTGGTGCCGGAGATGGACCGGGAGGTGTTCGACGCCGTCCACG
ACCCGAAGGTGGTGGGCGCGGCGCTGCTGCACCGGCAGCTGAGCGGCGAACCGCTGGAGCACTTCGTGCTGTTCGCCTCG
GTCGCGGCCTGGCTGACGACGGCCGGACAGACCAACTACGCGGCGGGGAACGCCTTCCTGGACGCGCTGGCGCACCACCG
CCGCGCGCAGGGGCTGCCGGCGCTGGCGCTGGACTGGGGCCCGTGGGCCACCGGCATGATCGAGGAACTGGGCCTGATCG
ACCACTACCGCAACAGCCGGGGCATGTCCTCGCTGGCGCCCGAGGCGGGCATGGCGGTGCTGGAGCGGGTCATCGGGCAG
GACCGGGCACAGCTGCTGGTGGCCACGGTCGTGGACTGGCCGGTGTTCATGTCCTGGTACGCGGCGCCGCCGCGGCTGGT
CACGGAGCTGGCGGCCACCGCCCAGGGACCGGGGTCCGAGGGCGACGGCAGTTTCCTGGACGCGTTCCGGGAGGCCACCG
```

```
CGGACAAGCGGCGGCTGCTGCTGACCGAGCGGTTCACGACGCTGGTGGCGGGTGTGCTGCGGGTGCGGGCCGAGCAGGTG
GATCCGGCGGTCAGCCTGAATCTGCTGGGGCTCGACTCGCTGCTGGCGATGGAGCTGCGAGCGCGGGTGGTGGCCGAGGT
GGGCATCGCGCTGCCGGTGGTGGCGCTGCTGTCCAGCGCGCCGGCCGGGGACCTGATCACCCAGCTGCACGAGGGCCTGG
AGGAGTTGCTGGCCGAGGAGGGCAGCGGCGCCGCGGTGACGGCGGTGGAGCGCCTTCGAGGACGAGGCCGAGTTCCCGCTG
ACGCAGAACCAGAAGGCGCTGTGGTTCCTGAAGCAGCTGAACCCGGACGGCTTCGCGTACAACATCGGCGGCGCCGTCGA
GGTGCGGGTCGAGCTGGACCCGGACCTGATGTTCGAGGCGTTTCGCCGGCTGCTGGCCCGGCATCCCGTGCTGCGGGCGA
ACTTCCTGCTGGTGGAGGGGCAGGCGGTGCAGCGGATCTCCCCGGAGATCAAGGAGGACATCGCGCTCTTCGACGTCGAG
GACCGCGCGTGGGACGACATCTACCGGATGATCATCGAGGAGTACCGCAAGCCGTACGACCTGGCGACCGATCCGCTGAT
CCGGTTCCGCCTCTTCCGGCGCGGCCCGGACCGCTGGGTCATCACCAAGGCCGTCCACCACATCATCTCGGACGCCATCT
CCACCTTCACCTTCATCGAGGAACTGCTGTCCCTGTACGAGGGGCTGCGGCAGGGCCACGACGTCGAACTGCCGCCGGTG
TCCGCCCGCTATCTGGACTTCCTCAACTGGCAGAACGCGTTCCTGGCCGGCCGCGAGGCGCAGAAGATGCTCGCGTACTG
GCGGGGGCAGCTGCCGGACGAGGTGCCGGTGCTGGCGCTGCCCACCGACAAGCCGCGCCCGGCGGTGCTCACCCACAACG
GGGCGTCCGAGTTCTTCGCCCTGGACGCGGAGTTGAGCGCCCGGGTGCACGCGCTGGCGCGGGAGCACAACGTCACCGTC
TTCATGGTGCTGCTGAGCGCGTACTACCTGCTGCTGCACCGCTATGCGGGGCAGGACGACATCATCGTCGGCTCCCCCGT
CACCGGCCGCACCCAGGAGGAGTTCGGCGCCGTCTACGGGTACTTCGTGAACCCGCTGCCGCTGCACGCCTCGCTGGCCG
GTGACCCCACGGTCGCCGAGCTGCTGGACCAGGTGCGCACCACGGTGCTGGGCGGCCTGGACCACCAGGAGTACCCGTTC
ACGCTGCTGGTGGAGCAGCTGGGGCTGGCCCACGACCCGAGCCGGTCGGCGGTCTTCCAGGCGATGTTCATCCTGCTGCA
CCACAAGGTGGCCACCGAGAAGTACGGCTACAAGCTGGAGTACATCGAGCTGCCCGAGGAGGAGGGCCAGTTCGACCTGA
CGCTGTCCGCGTACGAGGAGGAGGCGGACGGGCGGTTCCACTGCGTCTTCAAGTACAACACCGACCTCTTCGAGGCGGAG
ACGATCCGGCGGCTCGCCGGGCACTACACGCAGCTCCTGGAGTCGCTGACCGCGGCGCCCGCCGACGCCGCCACCGGTGG
ACTGCGGATGCTGTCGGGCGGCGAGCGGGAGCGGATCCTCACCGAGTGGAGCGGGGCCGGGCAGGGCGCGCAGGACGCGC
CGGTGCCGGTGCACCGGCTGATCGCCGAGGCGGCGCACCGTACCCCGCAGGCGATCGCGGTGGCCGCGCCCGCCGAGAGC
GGGGAGACCCGGCGGCTGACGTACGGCGAACTGGAGGAGCGCGCCGGCGAACTGGCCGGGCGGCTGCGGGCGCGCGGCGT
GCGCGAGGGCACCGTCGTCGCGCTGTGCCTGGAGAAGTCGCCCGAGCTGATCACCGCCCTGCTGGCGGTCCTCAAGGCGG
GCGGCGCCTATCTGCCGCTGGACCCGGACTATCCGGCCGACCGGCTCGCGTACATGGTGCGCAACGCCGGGGCCACGCTG
GTGATCGGCGGGACGGGCGGCGCGGCCGAGGGGCTGCCGGGCACCGTGGTCACCCTGGAGGAACTGCTCGCGGGCGAGGC
CGGCGAAGCGGGGCCGGACGCCGAGCCGGGGCCCGACTCCCCCGCCTACGTCATCTACACCTCGGGCTCCACCGGGCGCC
CCAAGGCGGTCGCGGTCAGCCACCGCAATCTGGCCTCGGTGTACGCCGGATGGCGCGACGCCTACCGCCTGGAGGAGGGC
GGCATCCGGGTCCATCTCCAGATGGCCAGCCCCTCCTTCGACGTCTTCACCGGCGACCTGACCCGAGCCCTGTGCTCGGG
CGGCACGCTGGTGCTGGTCGGCCGGGAGCTGCTGTTCAACACCGCCCGGCTGTACGAGACGATGCGCGCCGAACGGGTGG
ACTGCGGCGAGTTCGTGCCCGCCGTGGTGCGCACCCTGGTGCGGCACTGCGAGGACACCGGCGCCCGGCTGGACTTCCTG
CGGCTGCTGATCGTGGGCTCGGACTCCTGGAAGGCCGAGGAGTACGAGCGGCTGCGCGCGCTGGGCGCACAGCGCCTGGT
GAACTCGTACGGGCTCACCGAGGCCACCATCGACAGCGCCTGGTTCGAGGGTCCCGCGGATGACCTGGAGGGCGGCCGGA
TGGTGCCCATCGGGCGGCCGTTCCCGGGCAGCGCGCTGTACATCCTGGACTCGCGCGGCGAGCCGGTGCCGCCCGGTGTC
CCCGGCGAGCTGTGGATCGGCGGCACCGGGGTGGCGCTCGGCTACCTCGGCGACGAGGCGCTGACCGGGGAGCGGTTCCT
CACCCGCGCCCTGGCCGGCGACGCTCCGGTACGGCTGTACCGCACCGGTGACCTCGCGCGCTGGGACGCGGCCGGCACCG
TCCATCTGCTGGGCCGGGCCGACTCGCAGATCAAGGTGCGCGGGCACCGCATCGAGATCGGGGAGATCGAGTCGCACCTG
GCGGCCTGCCCCGAGCTGGCCCAGGCGCAGGTCACCGTGCGGCCGGACGCGGGCGGCGAGAACGTGCTGTGCGCGTACGG
GGTGGCGGCCCCGGGCGCCGTGCTGGACTGGCGCGAGGTGCGCCGGCGCCTGGCGGACTATCTGCCGACGTTCATGATCC
CCACCCACTTCACCGAGCTGCCCGCCCTGCCGCTCACCCCGAACGGCAAGGTGGACGTGGCGGCGCTGCCCGCCCCGCGC
ACCGGCGACGGCGCGGACGGGCCGGTGTACGAGGCCCCCGTCACGCTGTACGAGACCCGGATGGCCGAGCACTGGCAGCG
GCTGCTGGGCATCGAGGCCCCCGGGCCCGGTCTGGGCCACGACTTCTTCGAGACCGGTGGCAGCTCCATCCGGCTGATCG
AGCTGATCTACCACCTGCAGGCCGAGTTCGGGATCTCCATCCCGGTCAGCCGGCTGTTCCAGGTGACGACGCTGCACGGC
ATGGCCAAGACGGTCGAGCGGATCGTCACCGGGGAGATCGAGGGGTCGCTGCCGTATCTGCGGTTCAACGAGAACGCCGC
GGCGGGCACGGTGTTCTGCTTCCCGCCGGCCGGTGGCCACGGCCTGGTCTACCGGGAGTTCGCGGCGCGGCTGCCGGAGT
TCGAGTTCCTCGCCTTCAACTACCTGATGGGCGAGGACAAGGTAAGCGGGTACGCCGACCTGGTGGCCGGGCACCGGCCG
GAGGGCGAGATCGACCTGCTCGGCTACTCGCTGGGCGGCAACCTCGCCTTCGAGGTGGCCAAGGAGCTGGAGCGGCGCGG
CCGCACCGTGCGCCACGTCGTCATCATGGACTCGCTGCGGGTGACGGAGTCCTACGAGCTGGGCCCGGAGCACCTGGCCG
TCTTCGAGCGCGAGCTGGCCGAGCATCTGCGCAAGCACACCGGCTCGGCGCTGGTCGCGGAGAAGACGCGCGAACAGGCC
AAGGACTACCTGGAGTTCACCGGCCGCACCGCCAACCCCGGCACCACCGGGGCCCGGATCGCGGTGATCAGTGACGAGGA
GAACGCGGCCGCGTACGACAGCGGCGCCGAGGGCAGCTGGCACGGCGCCTCCCGTACCGGAACCGACGTGCTGCGCGGGG
TGGGCCGGCACGCCGACATGCTCGATCCGGGGACGGTCGAGCACAACGCGCGCCTGGCGCGCGGCATTCTCACCGGCGGT
GATGGCGAGGTATGA
```

110

Translation for Tü6239-PKS/NRPS [3124 AA]:

```
MDSMHHPAPVPVPEVPAPVPSQDDAFAIVGIGCRLPGGASDYRTFWRNLLDGKDCITDTPADRYDTRTLGSGDKAKPGRL
VGGRGGYIDGFDEFDPAFFGISPREAEHMDPQQRKLLEVAWEALEDGGLKPAELAGSDVGVYVGAFTLDYKILQFADLGF
ETLAAHTATGTMMTMVSNRISYCFDFRGPSVSVDTACSGSLVAVHLACQSLRRGETSVALAGGTLLHMAPQYTIAETKGG
FLSPDGRSRALDASANGYVRAEGVGMVAIKRLADAQRDGDPIHAVIIGSGVNQDGRTNGITVPNPDAQVALIERVCAAAG
VTPGSLQYVEAHGTSTPVGDPLEANALGRALSIGREPGARTYVGSVKTNIGHTESAAGIAGLIKTVLSLKHKVIPPHINL
EKLNPQIDEASLPYEIPREPTPWPEHSGPARAGVNSFGFGGTNAHVLLQEAPPTVGEPAPPATDGYSVLPLSARDPEAFP
AIATGLRERLAEGLPVGDAAYTLAHRRQHLEQRLSVVYDSPEALDEVLGAVARGESHPRAVAGTQREGLDRRLVWVFTGM
GPQWWAMGRQLYASEPVYREVIDRCDQEIAALTGWSLTQELNADEADSRMSETWLAQPANFAVQIALAALWRSKGIQPDA
VTGHSTGEVAAFYEAGVYTLPEAVKIVVHRSRLQQKLIGTGSMLAVSLTEAEAARRVRPHGDRVSIAAVNSPTSITLAGD
TEALEVIAAELGAEDIFARFLEVGVPYHSPRMELIKDELLTSLADLKPQQAKLPLYLTALPGTVAQGTELDADYWWRNVR
EAVHFRAAVDRLLDDGYGVFLEIGPHPVLAHSLRECCEARDAHSVTLASIRRKADERERLTLSLAALHSLGFAVDWHALH
PAGRPAELPRYPFRRDRYWVEPAPVAQIRLGHRDHPLLGRRTASAEPVWEVKLDAEAAPYLEDHRIQGTVLFPAAGYLEM
AAQAMRALTGDEHSTAALAGIELRKALFLPDGEPQTVQLSFSSDAAAFSIATVGAAGAEPTVHATGTVRAAQRRRLTAPL
DTVAVRARAARHLSGPDCYAELAALGYHYGPAFQGIEEVWIGEGEALARIRPPQGLTPDAAAHHMHPVLLDSCFQSLLTP
QLLTAPAGPGGTGIRLPLSIAEVRLDPVGDRELWVHATVTGDDEDELTGDIAVYDGADGTPLGRVAGFRAADVEKAATTV
GLSTIDSWLTEPSWVPCPLPEAASAAPAAGRHVLFADAGGVAQRLAALIGEAGGEAHLVRPGAAYGLDRTARTATVVPGS
ADDLRRLLTDLGQVDGVVHLWNLDRPALADAPRGRFADIASTGAYALIALTQALLADPERHGGTPVHIVTRAAQCVVPGE
PVEPLGAPAWGIGRVLWQQELAGRGGKLIDLAADGGVEEDAYALLRELADPTGAAEREDEIALRAGERHTSRLVAAEGLS
RPLPLRLRPDGSYLVTGAFGALGRLLCRTLVRRGARRLILVGRTRLPERERWADQDPNSPAGRHVAFLKELEALGAQPIL
APLDITDEDALAGWLAGYRRAQGPPIRGVFHLAGQVRDTLVPEMDREVFDAVHDPKVVGAALLHRQLSGEPLEHFVLFAS
VAAWLTTAGQTNYAAGNAFLDALAHHRRAQGLPALALDWGPWATGMIEELGLIDHYRNSRGMSSLAPEAGMAVLERVIGQ
DRAQLLVATVVDWPVFMSWYAAPPRLVTELAATAQGPGSEGDGSFLDAFREATADKRRLLLTERFTTLVAGVLRVRAEQV
DPAVSLNLLGLDSLLAMELRARVVAEVGIALPVVALLSSAPAGDLITQLHEGLEELLAEEGSGAAVTAVERFEDEAEFPL
TQNQKALWFLKQLNPDGFAYNIGGAVEVRVELDPDLMFEAFRRLLARHPVLRANFLLVEGQAVQRISPEIKEDIALFDVE
DRAWDDIYRMIIEEYRKPYDLATDPLIRFRLFRRGPDRWVITKAVHHIISDAISTFTFIEELLSLYEGLRQGHDVELPPV
SARYLDFLNWQNAFLAGREAQKMLAYWRGQLPDEVPVLALPTDKPRPAVLTHNGASEFFALDAELSARVHALAREHNVTV
FMVLLSAYYLLLHRYAGQDDIIVGSPVTGRTQEEFGAVYGYFVNPLPLHASLAGDPTVAELLDQVRTTVLGGLDHQEYPF
TLLVEQLGLAHDPSRSAVFQAMFILLHHKVATEKYGYKLEYIELPEEEGQFDLTLSAYEEEADGRFHCVFKYNTDLFEAE
TIRRLAGHYTQLLESLTAAPADAATGGLRMLSGGERERILTEWSGAGQGAQDAPVPVHRLIAEAAHRTPQAIAVAAPAES
GETRRLTYGELEERAGELAGRLRARGVREGTVVALCLEKSPELITALLAVLKAGGAYLPLDPDYPADRLAYMVRNAGATL
VIGGTGGAAEGLPGTVVVTLEELLAGEAGEAGPDAEPGPDSPAYVIYTSGSTGRPKAVAVSHRNLASVYAGWRDAYRLEEG
GIRVHLQMASPSFDVFTGDLTRALCSGGTLVLVGRELLFNTARLYETMRAERVDCGEFVPAVVRTLVRHCEDTGARLDFL
RLLIVGSDSWKAEEYERLRALGAQRLVNSYGLTEATIDSAWFEGPADDLEGGRMVPIGRPFPGSALYILDSRGEPVPPGV
PGELWIGGTGVALGYLGDEALTGERFLTRALAGDAPVRLYRTGDLARWDAAGTVHLLGRADSQIKVRGHRIEIGEIESHL
AACPELAQAQVTVRPDAGGENVLCAYGVAAPGAVLDWREVRRRLADYLPTFMIPTHFTELPALPLTPNGKVDVAALPAPR
TGDGADGPVYEAPVTLYETRMAEHWQRLLGIEAPGPGLGHDFFETGGSSIRLIELIYHLQAEFGISIPVSRLFQVTTLHG
MAKTVERIVTGEIEGSLPYLRFNENAAAGTVFCFPPAGGHGLVYREFAARLPEFEFLAFNYLMGEDKVSGYADLVAGHRP
EGEIDLLGYSLGGNLAFEVAKELERRGRTVRHVVIMDSLRVTESYELGPEHLAVFERELAEHLRKHTGSALVAEKTREQA
KDYLEFTGRTANPGTTGARIAVISDEENAAAYDSGAEGSWHGASRTGTDVLRGVGRHADMLDPGTVEHNARLARGILTGG
DGEV
```

- Sequencing results
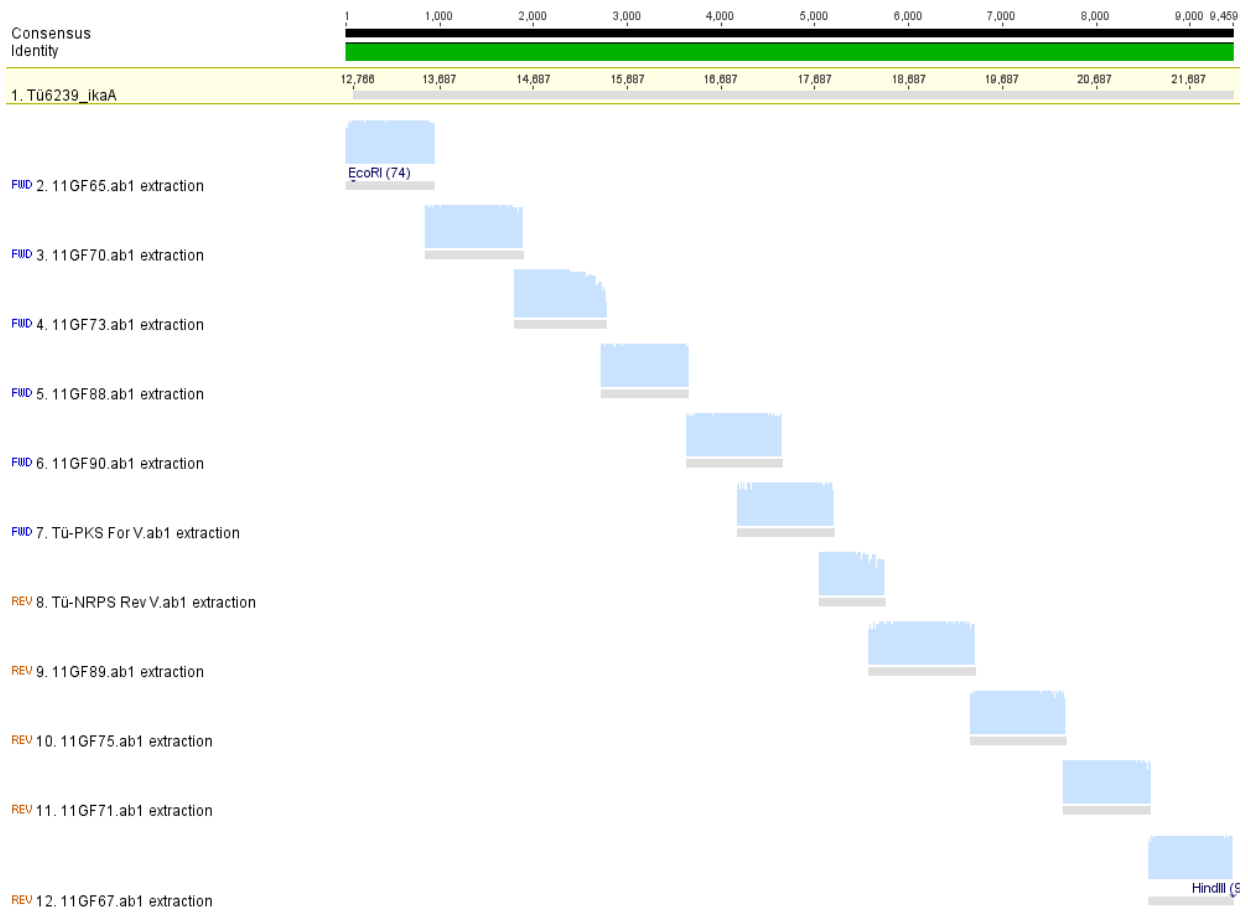
Figure S3. Schematic representation of the sequencing results for the *Eco*RI/*Hin*dIII-flanked insert *ikaA* of pHis8::Tü6239-ikaA. Lines 2 to 12 graphically show the perfect agreement of the individual overlapping sequencing runs with the original sequence of *ikaA* (top line).

Figure S4. Schematic representation of the sequencing results for the *Eco*RI/*Hin*dIII-flanked insert *ikaA* of pHis8C-TEV::Tü6239-ikaA. Lines 2 to 14 graphically show the perfect agreement of the individual overlapping sequencing runs with the original sequence of *ikaA* (top line).

- SDS page analysis
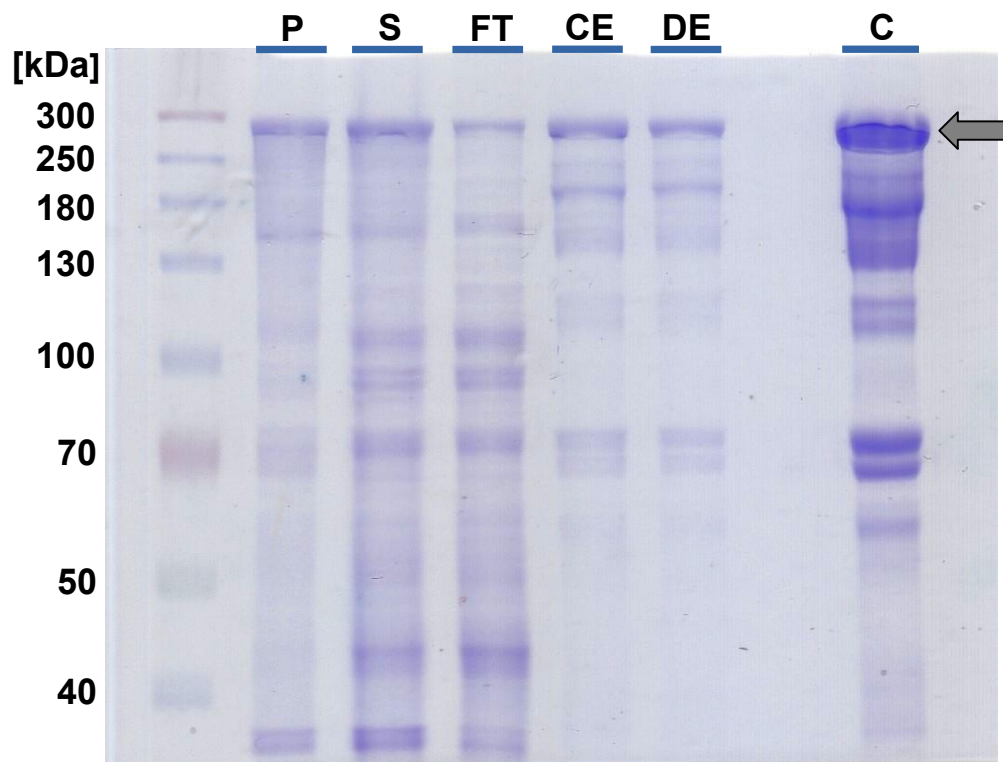
Figure S5. SDS-PAGE analysis monitoring the expression and enrichment of *N*-His8-Tü6239-PKS/NRPS (339.4 kDa). The debris pellet (P), supernatant (S), flow-through (FT), combined eluates (CE), PD-10 column eluate (DE) and Vivaspin concentrate (C) are shown.
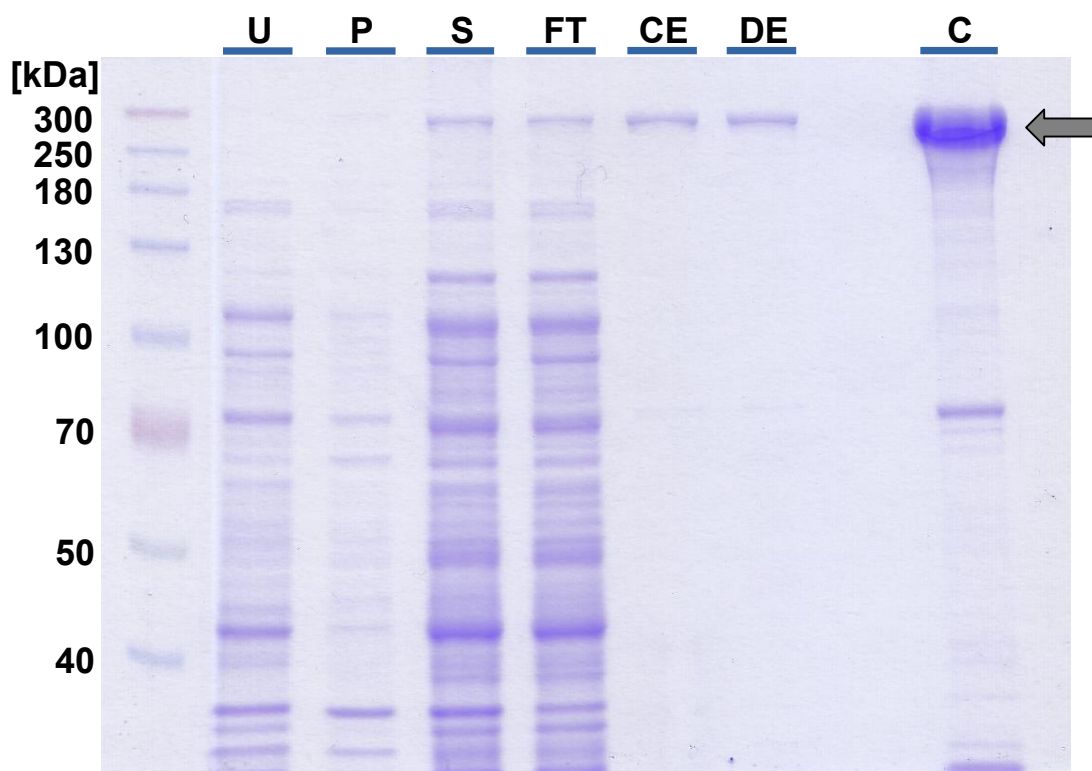
Figure S6. SDS-PAGE analysis monitoring the expression and enrichment of *C*-His8-TEV-Tü6239-PKS/NRPS (340.4 kDa). Fractions of uninduced bacteria (U), debris pellet (P), supernatant (S), flow-through (FT), combined eluates (CE), PD-10 column eluate (DE) and Vivaspin concentrate (C) are shown. The minor impurities visible in the concentrated fraction of IkaA were analyzed by MALDI-TOF after tryptic digest and shown to originate from *E. coli*. Upper band at approx. 70 kDa: putative bifunctional UDP-glucuronic acid oxidase/UDP-4-amino-4-deoxy-L-arabinose formyltransferase; lower band: glutamine-fructose-6-phosphate aminotransferase.

b. IkaB

- vector map
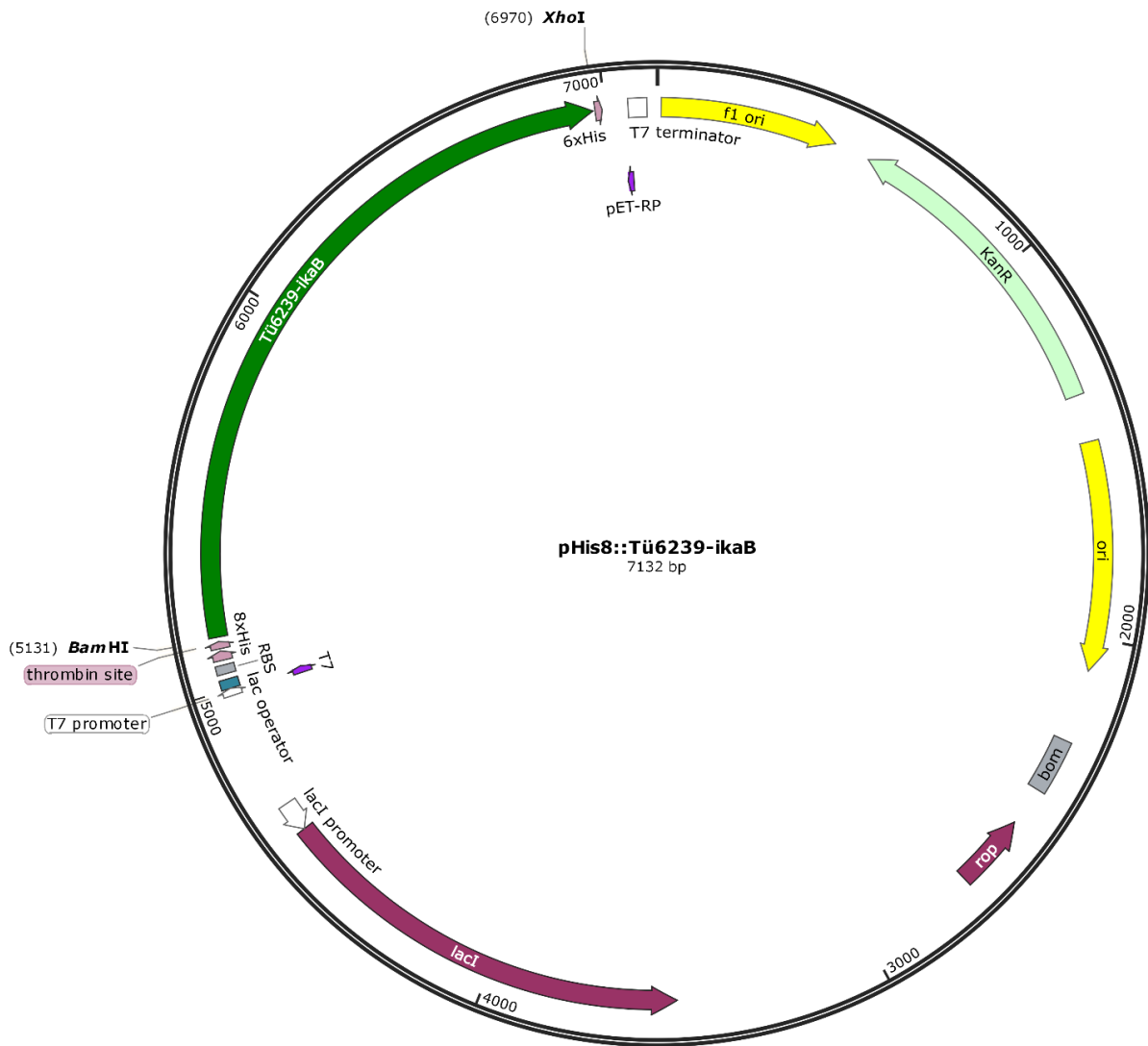
Figure S7. Vector map of the expression construct of *ikaB* in pHis8.

Figure S8. Vector map of the expression construct of *ikaB* in pGS-21a.

- DNA sequence of Tü6239-FAD (*ikaB*) [1833 bp]; GenBank: KY402203:

```
ATGACGCCTTTCGTTCAGCCGGCGGTCGACACCAAGGAGCACAGCGCCATGTCATCCCCCACCACCTCCGGCACCCCGGG
CAGGCAGTCGATGATCATCATCGGCGGCGGCCTGGGGGGCCTGTCCACCGGCTGCTACGCGCAGATGAACGGCTACGCGA
CGCGGGTCTTCGAGATGCACGAGATCCCGGGCGGTTCCTGCACCGCCTGGGAGCGCGGGGACTTCACCTTCGACTGGTGC
GTCAGCTGGCTGCTGGGCAGCGGTCCCGGCAACGAGATGTACCAGATCTGGATGGAACTGGGGGCGTTGCAGGGCAAGGA
GATGCGCCAGTTCGACGTCTTCAACATCGTGCGGGTGCGCGGCGGCCAGCCGGTGTACTTCTACTCCGACCCGGACCGGC
TCCAGGCGCACCTGCTGGAGATCTCCCCGGCCGACGCCCGCCGCATCAAGAACTTCTGCGAGGGGGTGCGCACCTTCCAG
AAGGCGCTGTCGGTCTACCCGTTCCTCAAGCCGGTGGGGCTGATGGGGCGGTGGGAACGGTGGAAGATGCTGGCCTCGTT
CCTGCCGTACTTCAACGCCATCCGCAAGTCCATCACCGAGCTGATGACGGACTACGCGGAGAAGTTCCAGCACCCGGTGC
TGCGCGAGGCCTTCAACTACGTGCTGTACGAGAAGCACGCCGACTTCCCCGTCCTGCCGTTCTGGTTCCAGCTGGCCTCG
CACGCCAACGGCTCGGCGGGGGTGCCCGAGGGCGGCTCGCTGGAGCTGGCCCGGTCCGTGGAGCGGCGCTACCTGGGGCT
CGGCGGGGAGATCACCTACAACGCCAAGGTGGAGAAGATCCTCGTCGAGCACGACAAGGCGGTGGGAGTGCGGCTCACCG
```

```
ACGGCCGCGAGTTCCGCGCGGACATCGTGGTGTCGGCGGCCGATCTGCACACCACCGCCATGGAGATGCTCGGCGGCCGG
TATCTCAACGACACCTGGCGCAAGCTGCTCACCGAGACGATCGACGAGGTGGGCACGATCTCCCCCGGCTATGTCTCGCT
GTTCCTGGGGCTGCGCCGGCCGTTCCCCGAGGGCGAGCCGTGCACCACGTACGTGCTGGAGGACAGCATGGCGGAGAAGC
TCACCGGCATGCGGCATCCCAGCATGAACGTGCAGTTCCGCAGCTGCCACTACCCGGAGCTGTCGCCGCGCGAGACCACG
GTCATCTTCGCCACGTACTTCTCGGAGGCCGAGCCGTGGCGGGCGCTGCGCGACGACGTGCCGGAACAGGCGGGCCGGGT
GCGGCGCGGTCAGGTGCTGCACACCCTGCCGGTGAAGCACGGCAAGGCGTACACCCAGGCCAAGCGGCAGGCGCGGATCA
CCATCGAGAACTTCCTGGACGAGCGGTTCCCCGGTCTCAAGGACGCGGTCGCCGTGCGGGACGTGTCCACGCCGCTGACG
CAGGTGCGCTACACGGGCACCTACAACGGCGGGTTCCCCGGCTGGCAGCCGTTCGTGGACGGCGGGGAGACCGTGGAGGT
GGAGATCAACAAGAACGGCCCGGTGCTGCCGGGGCTCTCCAACTTCTATCTGGCCGGGGTGTGGGTCACCGTCGGCGGGC
TGATCCGGGCGGTGGCCTCGGGCCGGCAGGTCACGCAGGTGATCTGCCGGGACGACGGGCGGGAGTTCACGGCGAGCGTG
GACGAGAGCGCGCCGCCGCCCACCCAGGTCGCCATCCCGGTGGGCAAGCAGCCGGGCGTGCCGGATCTGGCGGCCGGGTT
CCCCGCCCAGACCGCCGGCGGCGAGACCGCCACCGGCGCCAACAACACCGTCACATCGTCGAGGAGCGTGTAG
```

- Translation for Tü6239-FAD [610 AA]:

```
MTPFVQPAVDTKEHSAMSSPTTSGTPGRQSMIIIGGGLGGLSTGCYAQMNGYATRVFEMHEIPGGSCTAWERGDFTFDWC
VSWLLGSGPGNEMYQIWMELGALQGKEMRQFDVFNIVRVRGGQPVYFYSDPDRLQAHLLEISPADARRIKNFCEGVRTFQ
KALSVYPFLKPVGLMGRWERWKMLASFLPYFNAIRKSITELMTDYAEKFQHPVLREAFNYVLYEKHADFPVLPFWFQLAS
HANGSAGVPEGGSLELARSVERRYLGLGGEITYNAKVEKILVEHDKAVGVRLTDGREFRADIVVSAADLHTTAMEMLGGR
YLNDTWRKLLTETIDEVGTISPGYVSLFLGLRRPFPEGEPCTTYVLEDSMAEKLTGMRHPSMNVQFRSCHYPELSPRETT
VIFATYFSEAEPWRALRDDVPEQAGRVRRGQVLHTLPVKHGKAYTQAKRQARITIENFLDERFPGLKDAVAVRDVSTPLT
QVRYTGTYNGGFPGWQPFVDGGETVEVEINKNGPVLPGLSNFYLAGVWVTVGGLIRAVASGRQVTQVICRDDGREFTASV
DESAPPPTQVAIPVGKQPGVPDLAAGFPAQTAGGETATGANNTVTSSRSV
```

- Sequencing results
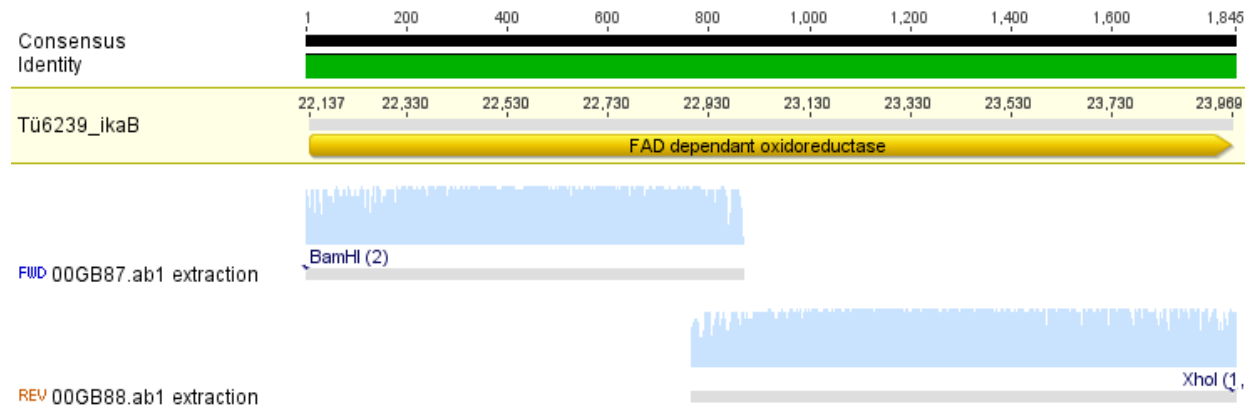
Figure S9. Schematic representation of sequencing results for the *Bam*HI/*Xho*I-flanked insert *ikaB* of pHis8::Tü6239-ikaB. Lines 2 and 3 graphically show the perfect agreement of the individual overlapping sequencing runs with the original sequence of *ikaB* (top line).
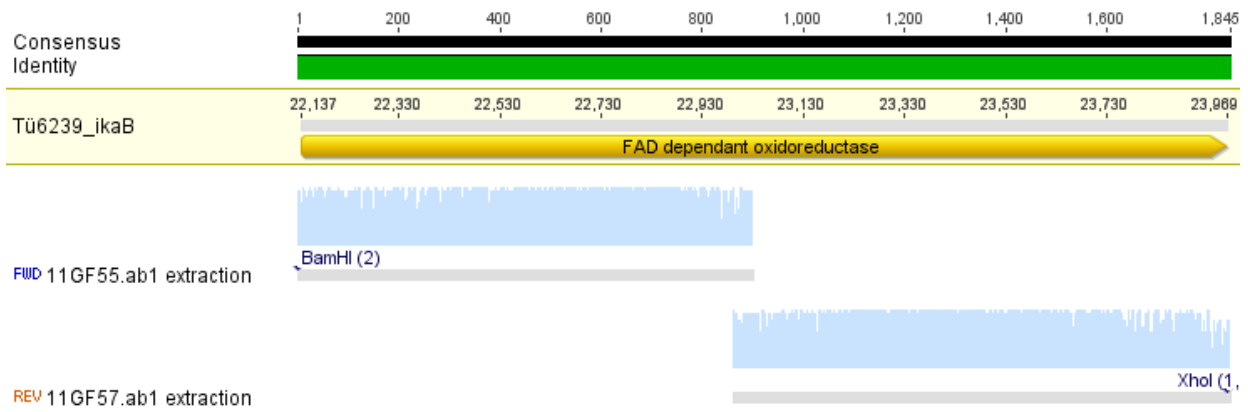
Figure S10. Schematic representation of sequencing results for the *Bam*HI/*Xho*I-flanked insert *ikaB* of pGS-21a::Tü6239-ikaB. Lines 2 and 3 graphically show the perfect agreement of the individual overlapping sequencing runs with the original sequence of *ikaB* (top line).

- SDS page analysis
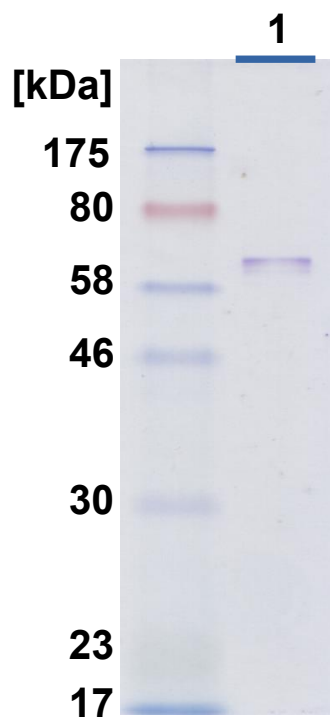
Figure S11. SDS-PAGE analysis of purified IkaB, *N*-His8-Tü6239-FAD (69.9 kDa).

c. IkaC (SA)

- vector map
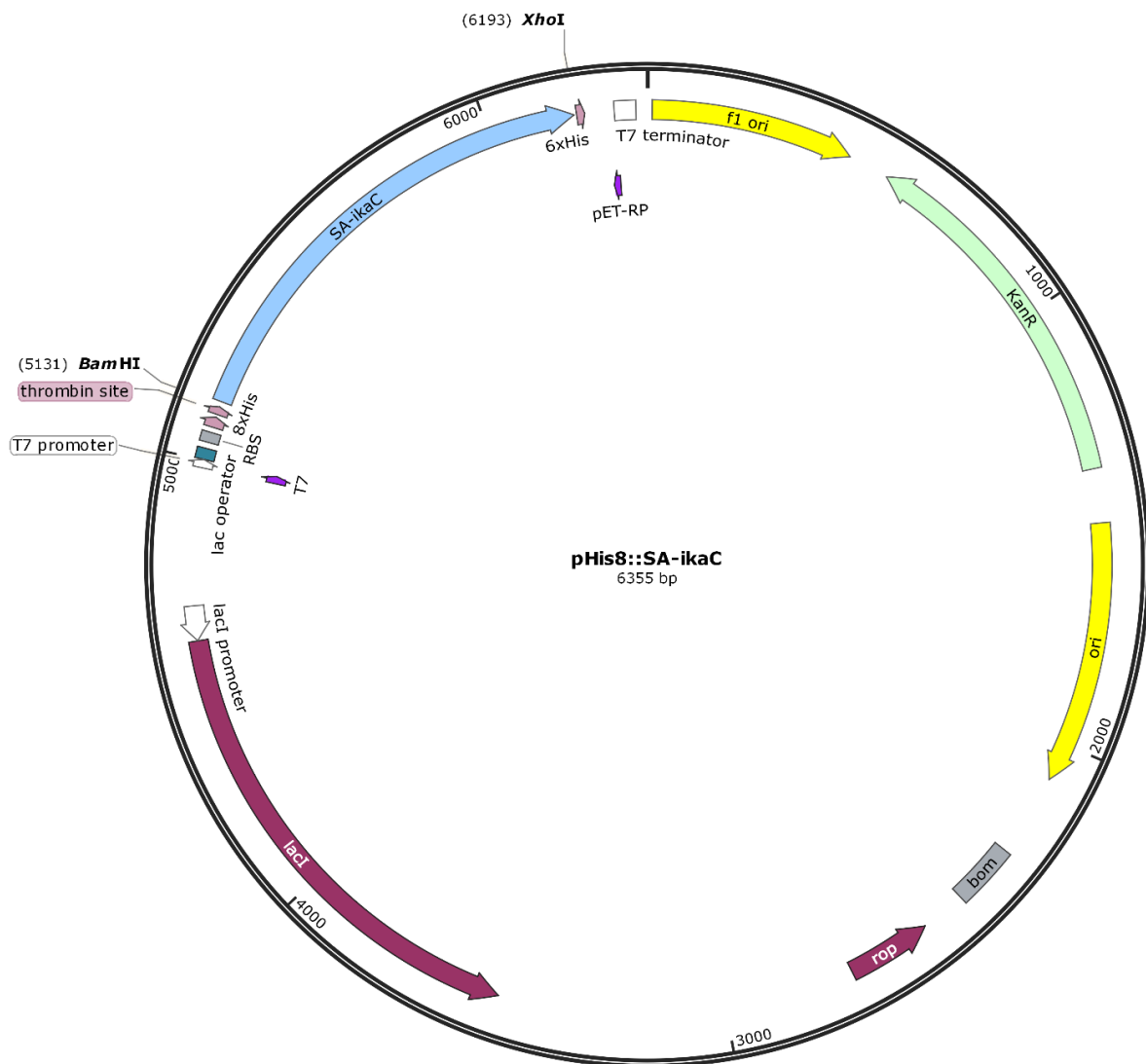
Figure S12. Vector map of the expression construct of *ikaC(SA)* in pHis8.

- DNA sequence of SA-AlcD (*ikaC*) [1056 bp]; GenBank: CP000850.1:

```
ATGAAAATCGACAAGTGGGTCGTCCGCGCGCACCTCGACGGGCTGCCGGACGCCGAACGAGTTTATGAGAAGGTCACCGA
GGACGTGCCGGTCGCACTCGGAGACGACGAGATGCTGTTCCGGACCAGGTACGTGTCGGTCGACCCGTACCTGCACGGGC
TCGCCCTGGAGACCCCAATCGGCGAGCACATGGTCGCCGACTCGATCATGGAGGTGGTCCAGGCCGGCCCGGGCGCTGCC
TTCCGGGTCGGCGACCTGGTCCAGGGCTACGGCGGCTGGCGCAGCCACATCATCGGCACTGGCGGCGAGGTGCGGTGGCA
GGGCGAGTCGTTCTCCATGGTATTCCCGGCATACCGCCACCTGAACCCGGACCACTACGACGACGTGCTGCCGGTGACGA
CCGCGCTCGGCATTATGGGCGGGCCGGGCATCACCGCCTGGGGCACGCTGGCACACTTCATGGCGATTCGGCCGGGCGAC
ACTCTGGTGATCAGCGGTGCGTCGGGCGCGGTCGGCACGCTGGTCGGCCAGCTCGCGAAGCGGGCCGGGGCGCGGGTGGT
GGGCACCACGTCGTCGCCGGCGAAGATCGGGTTCCTGACCGAGCTCGGCTTCGACGCCGTCGTTCCGTACCGGCTCGGAG
ACGACCCGGACCGGGTGCGCGAGGAGTTGCTCAGGGCCGCACCGGACGGCGTCGACCGATACTTCGACAACCTGGGCGGG
ACGATCACCGATGTGGTCTTCAGCATGCTCACGGTGCACAGCCAGGTTGCCGTCTGCTGGCAGTGGGCGACCCAGGTGAA
CCGCGACTACACCGGCCCGCGCCTGCTGCCGTACATCATGCAGTCCCGCACCACGATCCGCGGGATCTTCGCCTACGAGT
```

120

```
GGTTCACCGAGGAGAACCTCACGGCGATGCGCGACGAACTCGGCGGCATGATCCGCCGCGGCGAGCTGAGCTTCCGGCAG
ACCGTCTATCGCGGGTTCGACGAGATACCGGCCGCGTATCAGAGCCTCTATGTGGACCGCTCCGACAACCGCGGCAAGGT
CCTGGTGGAGATCTGA
```

- Translation for SA-AlcD [351 AA]:

```
MKIDKWVVRAHLDGLPDAERVYEKVTEDVPVALGDDEMLFRTRYVSVDPYLHGLALETPIGEHMVADSIMEVVQAGPGAA
FRVGDLVQGYGGWRSHIIGTGGEVRWQGESFSMVFPAYRHLNPDHYDDVLPVTTALGIMGGPGITAWGTLAHFMAIRPGD
TLVISGASGAVGTLVGQLAKRAGARVVGTTSSPAKIGFLTELGFDAVVPYRLGDDPDRVREELLRAAPDGVDRYFDNLGG
TITDVVFSMLTVHSQVAVCWQWATQVNRDYTGPRLLPYIMQSRTTIRGIFAYEWFTEENLTAMRDELGGMIRRGELSFRQ
TVYRGFDEIPAAYQSLYVDRSDNRGKVLVEI
```

- Sequencing results
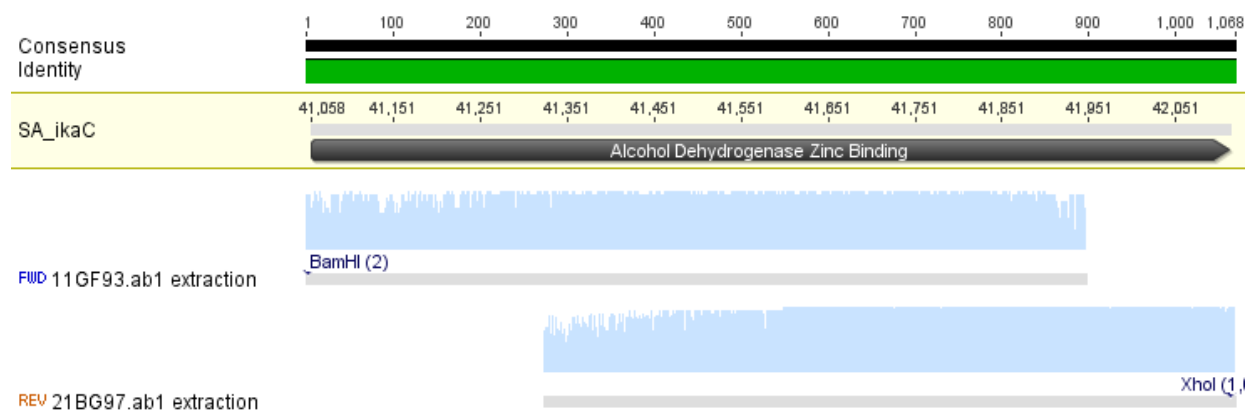
Figure S13. Schematic representation of sequencing results for the *Bam*HI/*Xho*I-flanked insert *ikaC(SA)* of pHis8::SA-ikaC. Lines 2 and 3 graphically show the perfect agreement of the individual overlapping sequencing runs with the original sequence of *ikaC* (top line).
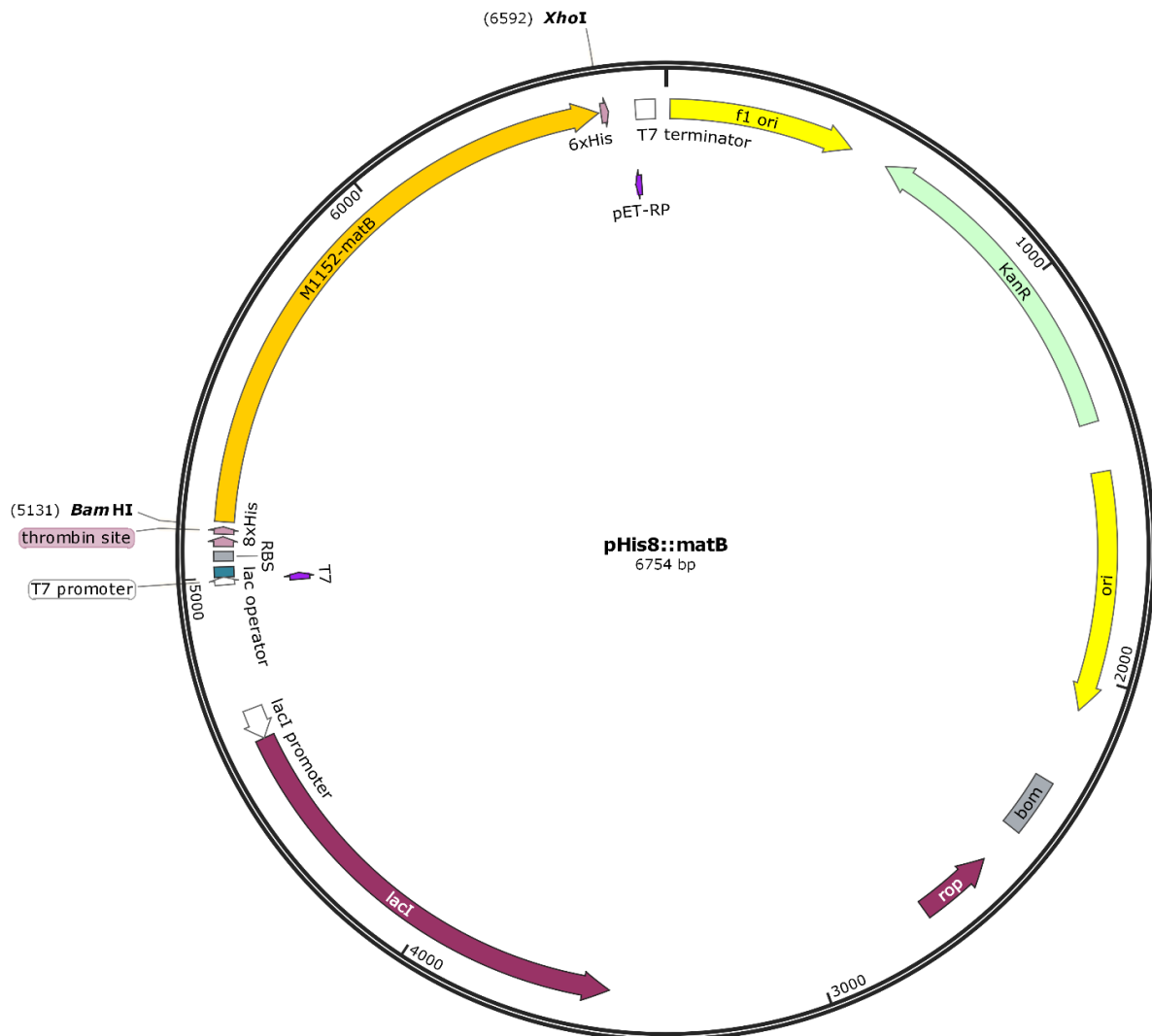
- SDS page analysis
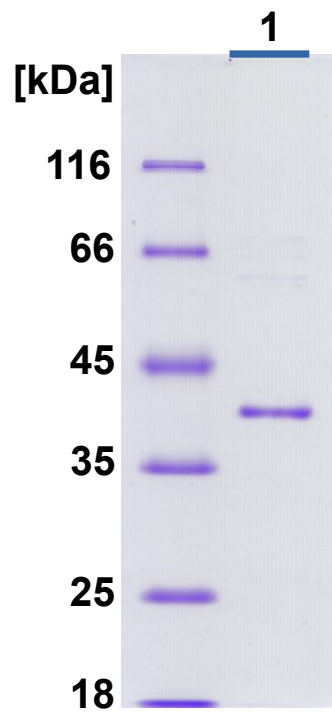
Figure S14. SDS-PAGE analysis of purified IkaC, *N*-His8-SA-AlcD (41.1 kDa).

d. Optional enzymes

- **MatB**

- vector map

Figure S15. Vector map of the expression construct of *matB* in pHis8.

- DNA sequence of *matB* [1455 bp]:

```
TCCTCTCTCTTCCCGGCCCTCTCCCCGGCCCCGACCGGCGCCCCGGCCGACCGGCCCGCGCTGCGGTTCGGCGAGCGCTC
CCTGACCTACGCGGAACTCGCCGCGGCGGCGGGCGCCACGGCCGGGCGGATCGGCGGCGCCGGCCGGGTCGCGGTCTGGG
CCACCCCGGCGATGGAGACCGGCGTCGCCGTGGTGGCGGCGCTGCTGGCCGGGGTCGCCGCCGTACCGCTCAACCCGAAG
TCCGGCGACAAGGAACTCGCGCACATCCTCTCCGACAGCGCGCCCTCGCTCGTCCTGGCGCCCCCGGACGCGGAACTCCC
GCCCGCCCTCGGGGCCCTGGAGCGCGTCGACGTCGACGTGCGGGCCCGCGGGGCGGTCCCCGAGGACGGTGCCGACGACG
GCGACCCCGCGCTCGTCGTCTACACCTCGGGCACCACGGGACCGCCGAAGGGCGCCGTCATCCCCCGGCGGGCGCTCGCC
ACGACCCTGGACGCGCTCGCCGACGCGTGGCAGTGGACCGGCGAGGACGTGCTGGTGCAGGGGCTGCCGCTGTTCCACGT
GCACGGGCTGGTCCTCGGCATCCTCGGCCCCGCTGCGCCGGGGCGGGTCCGTGCGGCACCTGGGCCGGTTCTCCACCGAGG
GTGCGGCGCGGGAGCTGAACGACGGCGCGACCATGCTGTTCGGGGTGCCGACGATGTACCACCGGATCGCCGAGACGCTC
```

```
CCCGCCGACCCGGAGCTGGCGAAGGCGCTCGCCGGGGCCCGGCTGCTGGTGTCGGGGTCGGCCGCGCTGCCGGTGCACGA
CCACGAGCGCATCGCCGCCGCCACCGGGCGCCGGGTGATCGAGCGGTACGGCATGACCGAGACGCTGATGAACACCAGCG
TGCGCGCCGACGGCGAGCCGCGCGCCGGGACGGTGGGCGTGCCGCTGCCCGGTGTGGAGCTGCGGCTGGTGGAGGAGGAC
GGCACGCCGATCGCGGCGCTCGACGGGGAGAGCGTCGGCGAGATCCAGGTTCGCGGCCCCGAACCTGTTCACCGAGTACCT
GAACCGCCCCGACGCCACCGCCGCCGCCTTCACCGAGGACGGCTTCTTCCGCACCGGCGACATGGCGGTGCGCGACCCCG
ACGGCTATGTCCGCATCGTCGGCCGCAAGGCCACCGACCTGATCAAGAGCGGCGGTTACAAGATCGGGGCCGGGGAGATC
GAGAACGCCCTGCTCGAACACCCGGAGGTCCGGGAGGCCGCCGTCACCGGCGAACCCGACCCCGACCTCGGGGAACGGAT
CGTGGCCTGGATCGTCCCGGCCGACCCCGCCGCCCCGCCCGCCCTCGGCACGCTGGCCGACCACGTCGCCGCCCGGCTCG
CCCCGCACAAGCGGCCGCGCGTCGTCCGGTACCTCGACGCGGTGCCCCGCAACGACATGGGGAAGATCATGAAGCGGGCG
CTGAACCGTGACTGA
```

- Translation for MatB [484 AA]:

```
SSLFPALSPAPTGAPADRPALRFGERSLTYAELAAAAGATAGRIGGAGRVAVWATPAMETGVAVVAALLAGVAAVPLNPK
SGDKELAHILSDSAPSLVLAPPDAELPPALGALERVDVDVRARGAVPEDGADDGDPALVVYTSGTTGPPKGAVIPRRALA
TTLDALADAWQWTGEDVLVQGLPLFHVHGLVLGILGPLRRGGSVRHLGRFSTEGAARELNDGATMLFGVPTMYHRIAETL
PADPELAKALAGARLLVSGSAALPVHDHERIAAATGRRVIERYGMTETLMNTSVRADGEPRAGTVGVPLPGVELRLVEED
GTPIAALDGESVGEIQVRGPNLFTEYLNRPDATAAAFTEDGFFRTGDMAVRDPDGYVRIVGRKATDLIKSGGYKIGAGEI
ENALLEHPEVREAAVTGEPDPDLGERIVAWIVPADPAAPPALGTLADHVAARLAPHKRPRVVRYLDAVPRNDMGKIMKRA
LNRD
```

- Sequencing results
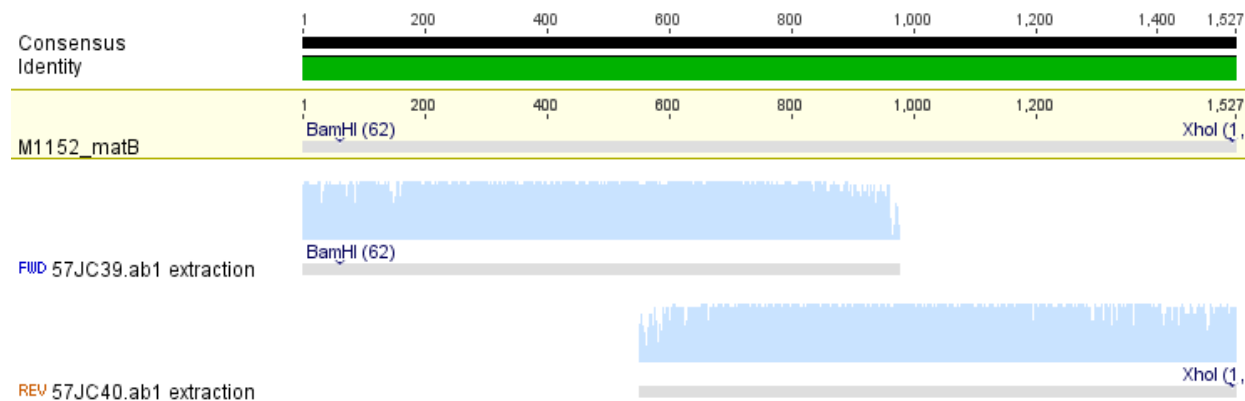
Figure S16. Schematic representation of sequencing results for the *Bam*HI/*Xho*I-flanked insert *matB* of pHis8::matB. Lines 2 and 3 graphically show the perfect agreement of the individual overlapping sequencing runs with the original sequence of *matB* (top line).

- **AckA**

- vector map
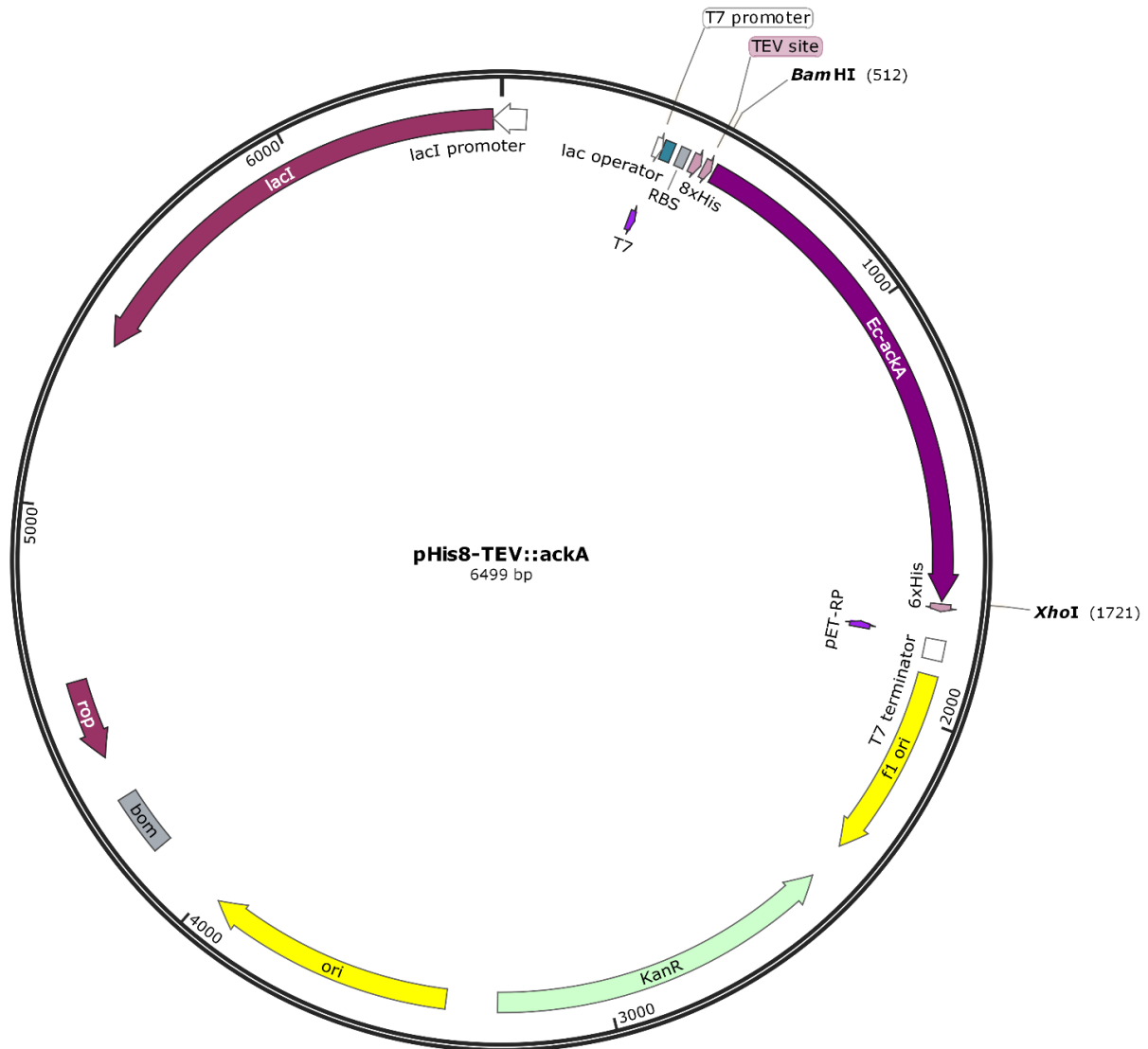
Figure S17. Vector map of the expression construct of *ackA* in pHis8-TEV.

- DNA sequence of *ackA* [1203 bp]:

```
ATGTCGAGTAAGTTAGTACTGGTTCTGAACTGCGGTAGTTCTTCACTGAAATTTGCCATCATCGATGCAGTAAATGGTGA
AGAGTACCTTTCTGGTTTAGCCGAATGTTTCCACCTGCCCGAAGCACGTATCAAATGGAAAATGGACGGCAATAAACAGG
AAGCGGCTTTAGGTGCAGGCGCCGCTCACAGCGAAGCGCTCAACTTTATCGTTAATACTATTCTGGCACAAAAACCAGAA
CTGTCTGCGCAGCTGACTGCTATCGGTCACCGTATCGTACACGGCGGCGAAAAGTATACCAGCTCCGTAGTGATCGATGA
GTCTGTTATTCAGGGTATCAAAGATGCAGCTTCTTTTGCACCGCTGCACAACCCGGCTCACCTGATCGGTATCGAAGAAG
CTCTGAAATCTTTCCCACAGCTGAAAGACAAAAACGTTGCTGTATTTGACACCGCGTTCCACCAGACTATGCCGGAAGAG
TCTTACCTCTACGCCCTGCCTTACAACCTGTACAAAGAGCACGGCATCCGTCGTTACGGCGCGCACGGCACCAGCCACTT
CTATGTAACCCAGGAAGCGGCAAAAATGCTGAACAAACCGGTAGAAGAACTGAACATCATCACCTGCCACCTGGGCAACG
GTGGTTCCGTTTCTGCTATCCGCAACGGTAAATGCGTTGACACCTCTATGGGCCTGACCCCGCTGGAAGGTCTGGTCATG
GGTACCCGTTCTGGTGATATCGATCCGGCGATCATCTTCCACCTGCACGACACCCTGGGCATGAGCGTTGACGCAATCAA
```

```
CAAACTGCTGACCAAAGAGTCTGGCCTGCTGGGTCTGACCGAAGTGACCAGCGACTGCCGCTATGTTGAAGACAACTACG
CGACGAAAGAAGACGCGAAGCGCGCAATGGACGTTTACTGCCACCGCCTGGCGAAATACATCGGTGCCTACACTGCGCTG
ATGGATGGTCGTCTGGACGCTGTTGTATTCACTGGTGGTATCGGTGAAAATGCCGCAATGGTTCGTGAACTGTCTCTGGG
CAAACTGGGCGTGCTGGGCTTTGAAGTTGATCATGAACGCAACCTGGCTGCACGTTTCGGCAAATCTGGTTTCATCAACA
AAGAAGGTACCCGTCCTGCGGTGGTTATCCCAACCAACGAAGAACTGGTTATCGCGCAAGACGCGAGCCGCCTGACTGCC
TGA
```

- Translation for AckA [400 AA]:

```
MSSKLVLVLNCGSSSLKFAIIDAVNGEEYLSGLAECFHLPEARIKWKMDGNKQEAALGAGAAHSEALNFIVNTILAQKPE
LSAQLTAIGHRIVHGGEKYTSSVVIDESVIQGIKDAASFAPLHNPAHLIGIEEALKSFPQLKDKNVAVFDTAFHQTMPEE
SYLYALPYNLYKEHGIRRYGAHGTSHFYVTQEAAKMLNKPVEELNIITCHLGNGGSVSAIRNGKCVDTSMGLTPLEGLVM
GTRSGDIDPAIIFHLHDTLGMSVDAINKLLTKESGLLGLTEVTSDCRYVEDNYATKEDAKRAMDVYCHRLAKYIGAYTAL
MDGRLDAVVFTGGIGENAAMVRELSLGKLGVLGFEVDHERNLAARFGKSGFINKEGTRPAVVIPTNEELVIAQDASRLTA
*
```

- Sequencing results
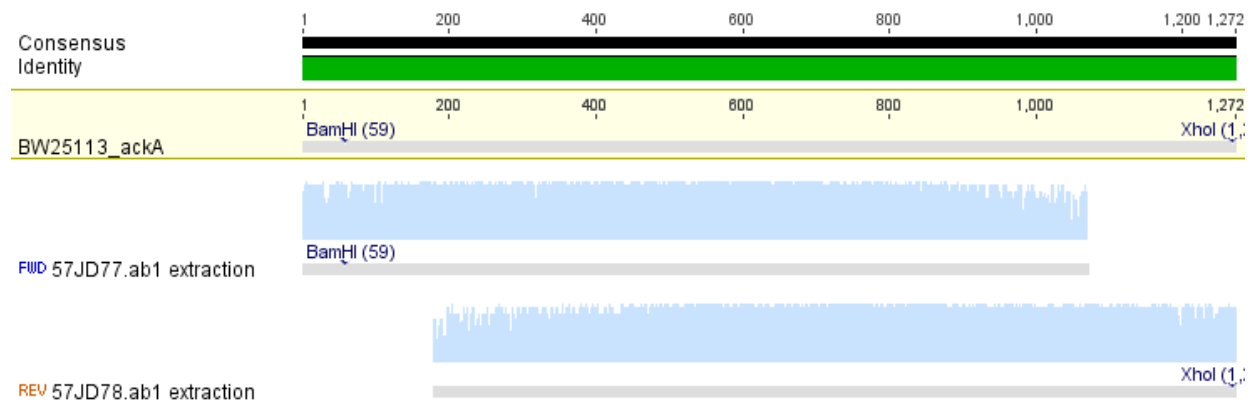
Figure S18. Schematic representation of sequencing results for the *Bam*HI/*Xho*I-flanked insert *ackA* of pHis8-TEV::ackA. Lines 2 and 3 graphically show the perfect agreement of the individual overlapping sequencing runs with the original sequence of *ackA* (top line).

- **Pta**

- vector map

Figure S19. Vector map of the expression construct of *pta* in pHis8-TEV.

- DNA sequence of *pta* [2142 bp]:

```
TCCCGTATTATTATGCTGATCCCTACCGGAACCAGCGTCGGTCTGACCAGCGTCAGCCTTGGCGTGATCCGTGCAATGGA
ACGCAAAGGCGTTCGTCTGAGCGTTTTTCAAACCTATCGCTCAGCCGCGTACCGGTGGCGATGCGCCCGATCAGACTACGA
CTATCGTGCGTGCGAACTCTTCCACCACGACGGCCGCTGAACCGCTGAAAATGAGCTACGTTGAAGGTCTGCTTTCCAGC
AATCAGAAAGATGTGCTGATGGAAGAGATCGTCGCAAACTACCACGCTAACACCAAAGACGCTGAAGTCGTTCTGGTTGA
AGGTCTGGTCCCGACACGTAAGCACCAGTTTGCCCAGTCTCTGAACTACGAAATCGCTAAAACGCTGAATGCGGAAATCG
TCTTCGTTATGTCTCAGGGCACTGACACCCCCGGAACAGCTGAAAGAGCGTATCGAACTGACCCGCAACAGCTTCGGCGGT
```

127

```
GCCAAAAACACCAACATCACCGGCGTTATCGTTAACAAACTGAACGCACCGGTTGATGAACAGGGTCGTACTCGCCCGGA
TCTGTCCGAGATTTTCGACGACTCTTCCAAAGCTAAAGTAAACAATGTTGATCCGGCGAAGCTGCAAGAATCCAGCCCGC
TGCCGGTTCTCGGCGCTGTGCCGTGGAGCTTTGACCTGATCGCGACTCGTGCGATCGATATGGCTCGCCACCTGAATGCG
ACCATCATCAACGAAGGCGACATCAATACTCGCCGCGTTAAATCCGTCACTTTCTGCGCACGCAGCATTCCGCACATGCT
GGAGCACTTCCGTGCCGGTTCTCTGCTGGTGACTTCCGCAGACCGTCCTGACGTGCTGGTGGCCGCTTGCCTGGCAGCCA
TGAACGGCGTAGAAATCGGTGCCCTGCTGCTGACTGGCGGTTACGAAATGGACGCGCGCATTTCTAAACTGTGCGAACGT
GCTTTCGCTACCGGCCTGCCGGTATTTATGGTGAACACCAACACCTGGCAGACCTCTCTGAGCCTGCAGAGCTTCAACCT
GGAAGTTCCGGTTGACGATCACGAACGTATCGAGAAAGTTCAGGAATACGTTGCTAACTACATCAACGCTGACTGGATCG
AATCTCTGACTGCCACTTCTGAGCGCAGCCGTCGTCTGTCTCCGCCTGCGTTCCGTTATCAGCTGACTGAACTTGCGCGC
AAAGCGGGCAAACGTATCGTACTGCCGGAAGGTGACGAACCGCGTACCGTTAAAGCAGCCGCTATCTGTGCTGAACGTGG
TATCGCAACTTGCGTACTGCTGGGTAATCCGGCAGAGATCAACCGTGTTGCAGCGTCTCAGGGTGTAGAACTGGGTGCAG
GGATTGAAATCGTTGATCCAGAAGTGGTTCGCGAAAGCTATGTTGGTCGTCTGGTCGAACTGCGTAAGAACAAAGGCATG
ACCGAAACCGTTGCCCGCGAACAGCTGGAAGACAACGTGGTGCTCGGTACGCTGATGCTGGAACAGGATGAAGTTGATGG
TCTGGTTTCCGGTGCTGTTCACACTACCGCAAACACCATCCGTCCGCCGCTGCAGCTGATCAAAACTGCACCGGGCAGCT
CCCTGGTATCTTCCGTGTTCTTCATGCTGCTGCCGGAACAGGTTTACGTTTACGGTGACTGTGCGATCAACCCGGATCCG
ACCGCTGAACAGCTGGCAGAAATCGCGATTCAGTCCGCTGATTCCGCTGCGGCCTTCGGTATCGAACCGCGCGTTGCTAT
GCTCTCCTACTCCACCGGTACTTCTGGTGCAGGTAGCGACGTAGAAAAAGTTCGCGAAGCAACTCGTCTGGCGCAGGAAA
AACGTCCTGACCTGATGATCGACGGTCCGCTGCAGTACGACGCTGCGGTAATGGCTGACGTTGCGAAATCCAAAGCGCCG
AACTCTCCGGTTGCAGGTCGCGCTACCGTGTTCATCTTCCCGGATCTGAACACCGGTAACACCACCTACAAAGCGGTACA
GCGTTCTGCCGACCTGATCTCCATCGGGCCGATGCTGCAGGGTATGCGCAAGCCGGTTAACGACCTGTCCCGTGGCGCAC
TGGTTGACGATATCGTCTACACCATCGCGCTGACTGCGATTCAGTCTGCACAGCAGCAGTAA
```

- Translation for Pta [713 AA]:

```
MSSKLVLVLNCGSSSLKFAIIDAVNGEEYLSGLAECFHLPEARIKWKMDGNKQEAALGAGAAHSEALNFIVNTILAQKPE
LSAQLTAIGHRIVHGGEKYTSSVVIDESVIQGIKDAASFAPLHNPAHLIGIEEALKSFPQLKDKNVAVFDTAFHQTMPEE
SYLYALPYNLYKEHGIRRYGAHGTSHFYVTQEAAKMLNKPVEELNIITCHLGNGGSVSAIRNGKCVDTSMGLTPLEGLVM
GTRSGDIDPAIIFHLHDTLGMSVDAINKLLTKESGLLGLTEVTSDCRYVEDNYATKEDAKRAMDVYCHRLAKYIGAYTAL
MDGRLDAVVFTGGIGENAAMVRELSLGKLGVLGFEVDHERNLAARFGKSGFINKEGTRPAVVIPTNEELVIAQDASRLTA
```

- Sequencing results

Figure S20. Schematic representation of sequencing results for the *Eco*RI/*Hin*dIII-flanked insert *pta* of pHis8-TEV::pta. Lines 2 to 4 graphically show the perfect agreement of the individual overlapping sequencing runs with the original sequence of *pta* (top line).

128

- **Sfp**

- DNA sequence of *sfp* [675 bp]:

```
ATGAAGATTTACGGAATTTATATGGACCGCCCGCTTTCACAGGAAGAAAATGAACGGTTCATGACTTTCATATCACCTGA
AAAACGGGAGAAATGCCGGAGATTTTATCATAAAGAAGATGCTCACCGCACCCTGCTGGGAGATGTGCTCGTTCGCTCAG
TCATAAGCAGGCAGTATCAGTTGGACAAATCCGATATCCGCTTTAGCACGCAGGAATACGGGAAGCCGTGCATCCCTGAT
CTTCCCGACGCTCATTTCAACATTTCTCACTCCGGCCGCTGGGTCATTGGTGCGTTTGATTCACAGCCGATCGGCATAGA
TATCGAAAAAACGAAACCGATCAGCCTTGAGATCGCCAAGCGCTTCTTTTCAAAAACAGAGTACAGCGACCTTTTAGCAA
AAGACAAGGACGAGCAGACAGACTATTTTTATCATCTATGGTCAATGAAAGAAAGCTTTATCAAACAGGAAGGCAAAGGC
TTATCGCTTCCGCTTGATTCCTTTTCAGTGCGCCTGCATCAGGACGGACAAGTATCCATTGAGCTTCCGGACAGCCATTC
CCCATGCTATATCAAAACGTATGAGGTCGATCCCGGCTACAAAATGGCTGTATGCGCCGCACACCCTGATTTCCCCGAGG
ATATCACAATGGTCTCGTACGAAGAGCTTTTATAA
```

- Translation for Sfp [224 AA]:

```
MKIYGIYMDRPLSQEENERFMTFISPEKREKCRRFYHKEDAHRTLLGDVLVRSVISRQYQLDKSDIRFSTQEYGKPCIPD
LPDAHFNISHSGRWVIGAFDSQPIGIDIEKTKPISLEIAKRFFSKTEYSDLLAKDKDEQTDYFYHLWSMKESFIKQEGKG
LSLPLDSFSVRLHQDGQVSIELPDSHSPCYIKTYEVDPGYKMAVCAAHPDFPEDITMVSYEELL
```

- **SDS page analysis for all optional enzymes**
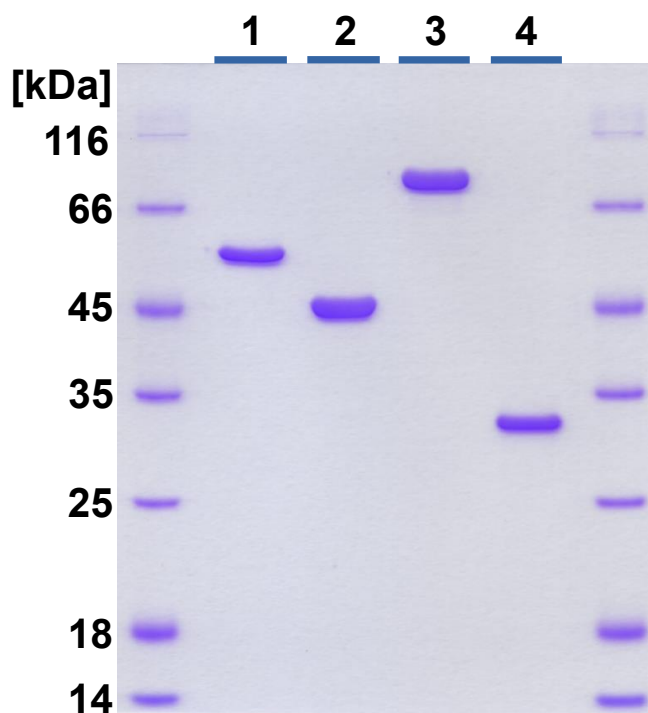
Figure S21. SDS-PAGE analysis of purified proteins MatB, AckA, Pta and Sfp. (**1**) *N*-His8-MatB (52.9 kDa), (**2**) *N*-His8-AckA (45.8 kDa), (**3**) *N*-His8-Pta (79.8 kDa), (**4**) *C*-His6-Sfp (27.2 kDa).

e. pHis8-TEV

- vector map

Figure S22. Vector map of pHis8-TEV.

- DNA sequence of pHis8-TEV [5330 bp]:

```
CTCTCTTCCGGGCGCTATCATGCCATACCGCGAAAGGTTTTGCGCCATTCGATGGTGTCCGGGATCTCGACGCTCTCCCT
TATGCGACTCCTGCATTAGGAAGCAGCCCAGTAGTAGGTTGAGGCCGTTGAGCACCGCCGCCGCAAGGAATGGTGCATGC
AAGGAGATGGCGCCCAACAGTCCCCCGGCCACGGGGCCTGCCACCATACCCACGCCGAAACAAGCGCTCATGAGCCCGAA
GTGGCGAGCCCGATCTTCCCCATCGGTGATGTCGGCGATATAGGCGCCAGCAACCGCACCTGTGGCGCCGGTGATGCCGG
CCACGATGCGTCCGGCGTAGAGGATCGAGATCTCGATCCCGCGAAATTAATACGACTCACTATAGGGGAATTGTGAGCGG
ATAACAATTCCCCTCTAGAAATAATTTTGTTTAACTTTAAGAAGGAGATATACCATGAAACACCACCACCACCACCACCA
CCACGGTGGTGAGAATCTTTATTTTCAGGGAGGATCCGAATTCGAGCTCCGTCGACAAGCTTGCGGCCGCACTCGAGCAC
CACCACCACCACCACTGAGATCCGGCTGCTAACAAAGCCCGAAAGGAAGCTGAGTTGGCTGCTGCCACCGCTGAGCAATA
ACTAGCATAACCCCTTGGGGCCTCTAAACGGGTCTTGAGGGGTTTTTTGCTGAAAGGAGGAACTATATCCGGATTGGCGA
ATGGGACGCGCCCTGTAGCGGCGCATTAAGCGCGGCGGGTGTGGTGGTTACGCGCAGCGTGACCGCTACACTTGCCAGCG
CCCTAGCGCCCGCTCCTTTCGCTTTCTTCCCTTCCTTTCTCGCCACGTTCGCCGGCTTTCCCCGTCAAGCTCTAAATCGG
GGGCTCCCTTTAGGGTTCCGATTTAGTGCTTTACGGCACCTCGACCCCAAAAAACTTGATTAGGGTGATGGTTCACGTAG
TGGGCCATCGCCCTGATAGACGGTTTTTCGCCCTTTGACGTTGGAGTCCACGTTCTTTAATAGTGGACTCTTGTTCCAAA
CTGGAACAACACTCAACCCTATCTCGGTCTATTCTTTTGATTTATAAGGGATTTTGCCGATTTCGGCCTATTGGTTAAAA
AATGAGCTGATTTAACAAAAATTTAACGCGAATTTTAACAAAATATTAACGTTTACAATTTCAGGTGGCACTTTTCGGGG
AAATGTGCGCGGAACCCCTATTTGTTTATTTTTCTAAATACATTCAAATATGTATCCGCTCATGAATTAATTCTTAGAAA
AACTCATCGAGCATCAAATGAAACTGCAATTTATTCATATCAGGATTATCAATACCATATTTTTGAAAAAGCCGTTTCTG
TAATGAAGGAGAAAACTCACCGAGGCAGTTCCATAGGATGGCAAGATCCTGGTATCGGTCTGCGATTCCGACTCGTCCAA
```

```
CATCAATACAACCTATTAATTTCCCCTCGTCAAAAATAAGGTTATCAAGTGAGAAATCACCATGAGTGACGACTGAATCC
GGTGAGAATGGCAAAAGTTTATGCATTTCTTTCCAGACTTGTTCAACAGGCCAGCCATTACGCTCGTCATCAAAATCACT
CGCATCAACCAAACCGTTATTCATTCGTGATTGCGCCTGAGCGAGACGAAATACGCGATCGCTGTTAAAAGGACAATTAC
AAACAGGAATCGAATGCAACCGGCGCAGGAACACTGCCAGCGCATCAACAATATTTTCACCTGAATCAGGATATTCTTCT
AATACCTGGAATGCTGTTTTCCCGGGGATCGCAGTGGTGAGTAACCATGCATCATCAGGAGTACGGATAAAATGCTTGAT
GGTCGGAAGAGGCATAAATTCCGTCAGCCAGTTTAGTCTGACCATCTCATCTGTAACATCATTGGCAACGCTACCTTTGC
CATGTTTCAGAAACAACTCTGGCGCATCGGGCTTCCCATACAATCGATAGATTGTCGCACCTGATTGCCCGACATTATCG
CGAGCCCATTTATACCCATATAAATCAGCATCCATGTTGGAATTTAATCGCGGCCTAGAGCAAGACGTTTCCCGTTGAAT
ATGGCTCATAACACCCCTTGTATTACTGTTTATGTAAGCAGACAGTTTTATTGTTCATGACCAAAATCCCTTAACGTGAG
TTTTCGTTCCACTGAGCGTCAGACCCCGTAGAAAAGATCAAAGGATCTTCTTGAGATCCTTTTTTTCTGCGCGTAATCTG
CTGCTTGCAAACAAAAAAACCACCGCTACCAGCGGTGGTTTGTTTGCCGGATCAAGAGCTACCAACTCTTTTTCCGAAGG
TAACTGGCTTCAGCAGAGCGCAGATACCAAATACTGTCCTTCTAGTGTAGCCGTAGTTAGGCCACCACTTCAAGAACTCT
GTAGCACCGCCTACATACCTCGCTCTGCTAATCCTGTTACCAGTGGCTGCTGCCAGTGGCGATAAGTCGTGTCTTACCGG
GTTGGACTCAAGACGATAGTTACCGGATAAGGCGCAGCGGTCGGGCTGAACGGGGGGTTCGTGCACACAGCCCAGCTTGG
AGCGAACGACCTACACCGAACTGAGATACCTACAGCGTGAGCTATGAGAAAGCGCCACGCTTCCCGAAGGGAGAAAGGCG
GACAGGTATCCGGTAAGCGGCAGGGTCGGAACAGGAGAGCGCACGAGGGAGCTTCCAGGGGGAAACGCCTGGTATCTTTA
TAGTCCTGTCGGGTTTCGCCACCTCTGACTTGAGCGTCGATTTTTGTGATGCTCGTCAGGGGGGCGGAGCCTATGGAAAA
ACGCCAGCAACGCGGCCTTTTTACGGTTCCTGGCCTTTTGCTGGCCTTTTGCTCACATGTTCTTTCCTGCGTTATCCCCT
GATTCTGTGGATAACCGTATTACCGCCTTTGAGTGAGCTGATACCGCTCGCCGCAGCCGAACGACCGAGCGCAGCGAGTC
AGTGAGCGAGGAAGCGGAAGAGCGCCTGATGCGGTATTTTCTCCTTACGCATCTGTGCGGTATTTCACACCGCATATATG
GTGCACTCTCAGTACAATCTGCTCTGATGCCGCATAGTTAAGCCAGTATACACTCCGCTATCGCTACGTGACTGGGTCAT
GGCTGCGCCCCGACACCCGCCAACACCCGCTGACGCGCCCTGACGGGCTTGTCTGCTCCCGGCATCCGCTTACAGACAAG
CTGTGACCGTCTCCGGGAGCTGCATGTGTCAGAGGTTTTCACCGTCATCACCGAAACGCGCGAGGCAGCTGCGGTAAAGC
TCATCAGCGTGGTCGTGAAGCGATTCACAGATGTCTGCCTGTTCATCCGCGTCCAGCTCGTTGAGTTTCTCCAGAAGCGT
TAATGTCTGGCTTCTGATAAAGCGGGCCATGTTAAGGGCGGTTTTTTCCTGTTTGGTCACTGATGCCTCCGTGTAAGGGG
GATTTCTGTTCATGGGGGTAATGATACCGATGAAACGAGAGAGGATGCTCACGATACGGGTTACTGATGATGAACATGCC
CGGTTACTGGAACGTTGTGAGGGTAAACAACTGGCGGTATGGATGCGGCGGGACCAGAGAAAAATCACTCAGGGTCAATG
CCAGCGCTTCGTTAATACAGATGTAGGTGTTCCACAGGGTAGCCAGCAGCATCCTGCGATGCAGATCCGGAACATAATGG
TGCAGGGCGCTGACTTCCGCGTTTCCAGACTTTACGAAACACGGAAACCGAAGACCATTCATGTTGTTGCTCAGGTCGCA
GACGTTTTGCAGCAGCAGTCGCTTCACGTTCGCTCGCGTATCGGTGATTCATTCTGCTAACCAGTAAGGCAACCCCGCCA
GCCTAGCCGGGTCCTCAACGACAGGAGCACGATCATGCGCACCCGTGGGGCCGCCATGCCGGCGATAATGGCCTGCTTCT
CGCCGAAACGTTTGGTGGCGGGACCAGTGACGAAGGCTTGAGCGAGGGCGTGCAAGATTCCGAATACCGCAAGCGACAGG
CCGATCATCGTCGCGCTCCAGCGAAAGCGGTCCTCGCCGAAAATGACCCAGAGCGCTGCCGGCACCTGTCCTACGAGTTG
CATGATAAAGAAGACAGTCATAAGTGCGGCGACGATAGTCATGCCCCGCGCCCACCGGAAGGAGCTGACTGGGTTGAAGG
CTCTCAAGGGCATCGGTCGAGATCCCGGTGCCTAATGAGTGAGCTAACTTACATTAATTGCGTTGCGCTCACTGCCCGCT
TTCCAGTCGGGAAACCTGTCGTGCCAGCTGCATTAATGAATCGGCCAACGCGCGGGGAGAGGCGGTTTGCGTATTGGGCG
CCAGGGTGGTTTTTCTTTTCACCAGTGAGACGGGCAACAGCTGATTGCCCTTCACCGCCTGGCCCTGAGAGAGTTGCAGC
AAGCGGTCCACGCTGGTTTGCCCCAGCAGGCGAAAATCCTGTTTGATGGTGGTTAACGGCGGGATATAACATGAGCTGTC
TTCGGTATCGTCGTATCCCACTACCGAGATATCCGCACCAACGCGCAGCCCGGACTCGGTAATGGCGCGCATTGCGCCCA
GCGCCATCTGATCGTTGGCAACCAGCATCGCAGTGGGAACGATGCCCTCATTCAGCATTTGCATGGTTTGTTGAAAACCG
GACATGGCACTCCAGTCGCCTTCCCGTTCCGCTATCGGCTGAATTTGATTGCGAGTGAGATATTTATGCCAGCCAGCCAG
ACGCAGACGCGCCGAGACAGAACTTAATGGGCCCGCTAACAGCGCGATTTGCTGGTGACCCAATGCGACCAGATGCTCCA
CGCCCAGTCGCGTACCGTCTTCATGGGAGAAAATAATACTGTTGATGGGTGTCTGGTCAGAGACATCAAGAAATAACGCC
GGAACATTAGTGCAGGCAGCTTCCACAGCAATGGCATCCTGGTCATCCAGCGGATAGTTAATGATCAGCCCACTGACGCG
TTGCGCGAGAAGATTGTGCACCGCCGCTTTACAGGCTTCGACGCCGCTTCGTTCTACCATCGACACCACCACGCTGGCAC
CCAGTTGATCGGCGCGAGATTTAATCGCCGCGACAATTTGCGACGGCGCGTGCAGGGCCAGACTGGAGGTGGCAACGCCA
ATCAGCAACGACTGTTTGCCCGCCAGTTGTTGTGCCACGCGGTTGGGAATGTAATTCAGCTCCGCCATCGCCGCTTCCAC
TTTTTCCCGCGTTTTCGCAGAAACGTGGCTGGCCTGGTTCACCACGCGGGAAACGGTCTGATAAGAGACACCGGCATACT
CTGCGACATCGTATAACGTTACTGGTTTCACATTCACCACCCTGAATTGA
```

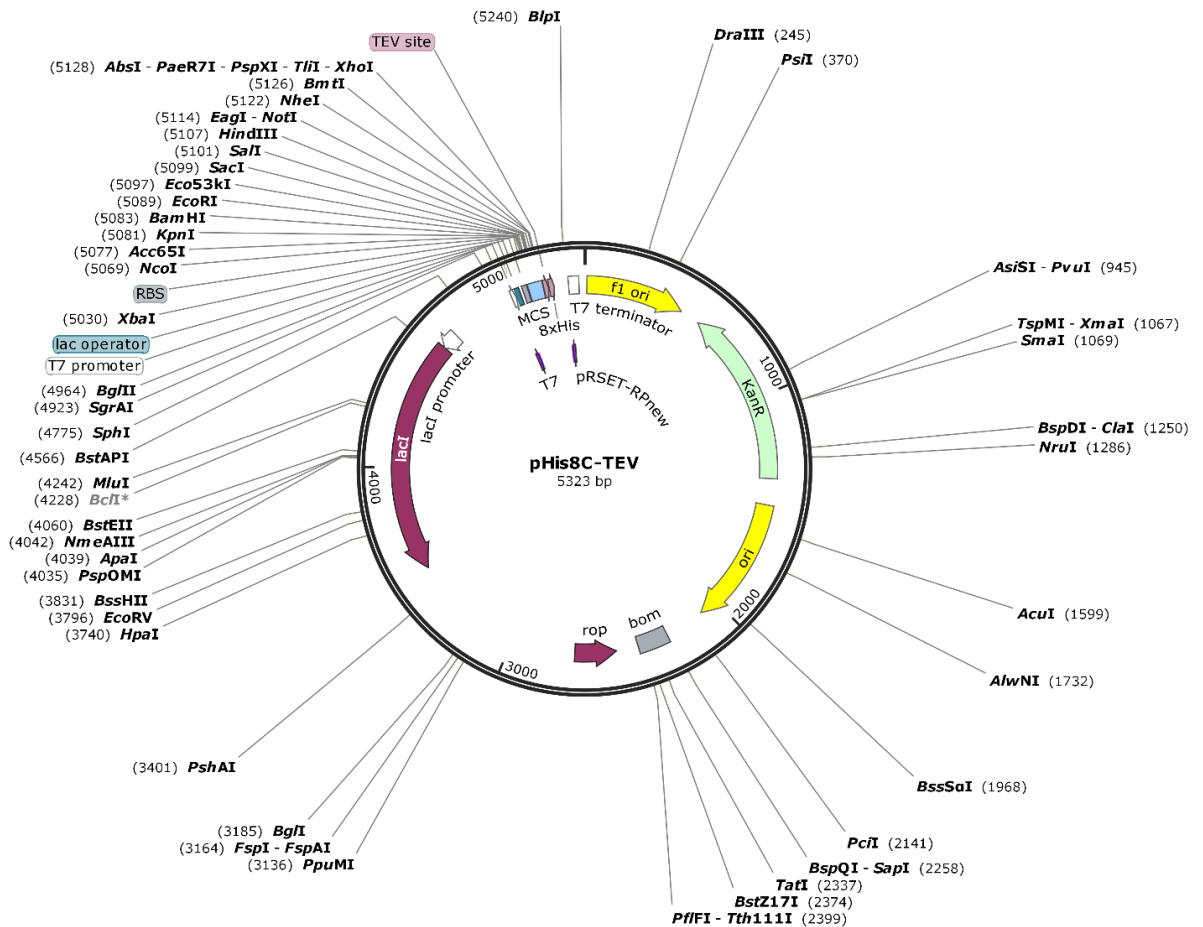### f. pHis8C-TEV

- vector map



Figure S23. Vector map of pHis8C-TEV.

- DNA sequence of pHis8C-TEV [5323 bp]:

```
TGGCGAATGGGACGCGCCCTGTAGCGGCGCATTAAGCGCGGCGGGTGTGGTGGTTACGCGCAGCGTGACCGCTACACTTG
CCAGCGCCCTAGCGCCCGCTCCTTTCGCTTTCTTCCCTTCCTTTCTCGCCACGTTCGCCGGCTTTCCCCGTCAAGCTCTA
AATCGGGGGCTCCCTTTAGGGTTCCGATTTAGTGCTTTACGGCACCTCGACCCCAAAAAACTTGATTAGGGTGATGGTTC
ACGTAGTGGGCCATCGCCCTGATAGACGGTTTTTCGCCCTTTGACGTTGGAGTCCACGTTCTTTAATAGTGGACTCTTGT
TCCAAACTGGAACAACACTCAACCCTATCTCGGTCTATTCTTTTGATTTATAAGGGATTTTGCCGATTTCGGCCTATTGG
TTAAAAAATGAGCTGATTTAACAAAAATTTAACGCGAATTTTAACAAAATATTAACGTTTACAATTTCAGGTGGCACTTT
TCGGGGAAATGTGCGCGGAACCCCTATTTGTTTATTTTTCTAAATACATTCAAATATGTATCCGCTCATGAATTAATTCT
TAGAAAAACTCATCGAGCATCAAATGAAACTGCAATTTATTCATATCAGGATTATCAATACCATATTTTTGAAAAAGCCG
TTTCTGTAATGAAGGAGAAAACTCACCGAGGCAGTTCCATAGGATGGCAAGATCCTGGTATCGGTCTGCGATTCCGACTC
GTCCAACATCAATACAACCTATTAATTTCCCCTCGTCAAAAATAAGGTTATCAAGTGAGAAATCACCATGAGTGACGACT
GAATCCGGTGAGAATGGCAAAAGTTTATGCATTTCTTTCCAGACTTGTTCAACAGGCCAGCCATTACGCTCGTCATCAAA
ATCACTCGCATCAACCAAACCGTTATTCATTCGTGATTGCGCCTGAGCGAGACGAAATACGCGATCGCTGTTAAAAGGAC
AATTACAAACAGGAATCGAATGCAACCGGCGCAGGAACACTGCCAGCGCATCAACAATATTTTCACCTGAATCAGGATAT
CTTCTAATACCTGGAATGCTGTTTTCCCGGGGATCGCAGTGGTGAGTAACCATGCATCATCAGGAGTACGGATAAAATG
CTTGATGGTCGGAAGAGGCATAAATTCCGTCAGCCAGTTTAGTCTGACCATCTCATCTGTAACATCATTGGCAACGCTAC
```

```
CTTTGCCATGTTTCAGAAACAACTCTGGCGCATCGGGCTTCCCATACAATCGATAGATTGTCGCACCTGATTGCCCGACA
TTATCGCGAGCCCATTTATACCCATATAAATCAGCATCCATGTTGGAATTTAATCGCGGCCTAGAGCAAGACGTTTCCCG
TTGAATATGGCTCATAACACCCCTTGTATTACTGTTTATGTAAGCAGACAGTTTTATTGTTCATGACCAAAATCCCTTAA
CGTGAGTTTTCGTTCCACTGAGCGTCAGACCCCGTAGAAAAGATCAAAGGATCTTCTTGAGATCCTTTTTTTCTGCGCGT
AATCTGCTGCTTGCAAACAAAAAAACCACCGCTACCAGCGGTGGTTTGTTTGCCGGATCAAGAGCTACCAACTCTTTTTC
CGAAGGTAACTGGCTTCAGCAGAGCGCAGATACCAAATACTGTCCTTCTAGTGTAGCCGTAGTTAGGCCACCACTTCAAG
AACTCTGTAGCACCGCCTACATACCTCGCTCTGCTAATCCTGTTACCAGTGGCTGCTGCCAGTGGCGATAAGTCGTGTCT
TACCGGGTTGGACTCAAGACGATAGTTACCGGATAAGGCGCAGCGGTCGGGCTGAACGGGGGGTTCGTGCACACAGCCCA
GCTTGGAGCGAACGACCTACACCGAACTGAGATACCTACAGCGTGAGCTATGAGAAAGCGCCACGCTTCCCGAAGGGAGA
AAGGCGGACAGGTATCCGGTAAGCGGCAGGGTCGGAACAGGAGAGCGCACGAGGGAGCTTCCAGGGGGAAACGCCTGGTA
TCTTTATAGTCCTGTCGGGTTTCGCCACCTCTGACTTGAGCGTCGATTTTTGTGATGCTCGTCAGGGGGGCGGAGCCTAT
GGAAAAACGCCAGCAACGCGGCCTTTTTACGGTTCCTGGCCTTTTGCTGGCCTTTTGCTCACATGTTCTTTCCTGCGTTA
TCCCCTGATTCTGTGGATAACCGTATTACCGCCTTTGAGTGAGCTGATACCGCTCGCCGCAGCCGAACGACCGAGCGCAG
CGAGTCAGTGAGCGAGGAAGCGGAAGAGCGCCTGATGCGGTATTTTCTCCTTACGCATCTGTGCGGTATTTCACACCGCA
TATATGGTGCACTCTCAGTACAATCTGCTCTGATGCCGCATAGTTAAGCCAGTATACACTCCGCTATCGCTACGTGACTG
GGTCATGGCTGCGCCCCGACACCCGCCAACACCCGCTGACGCGCCCTGACGGGCTTGTCTGCTCCCGGCATCCGCTTACA
GACAAGCTGTGACCGTCTCCGGGAGCTGCATGTGTCAGAGGTTTTCACCGTCATCACCGAAACGCGCGAGGCAGCTGCGG
TAAAGCTCATCAGCGTGGTCGTGAAGCGATTCACAGATGTCTGCCTGTTCATCCGCGTCCAGCTCGTTGAGTTTCTCCAG
AAGCGTTAATGTCTGGCTTCTGATAAAGCGGGCCATGTTAAGGGCGGTTTTTTCCTGTTTGGTCACTGATGCCTCCGTGT
AAGGGGGATTTCTGTTCATGGGGGTAATGATACCGATGAAACGAGAGAGGATGCTCACGATACGGGTTACTGATGATGAA
CATGCCCGGTTACTGGAACGTTGTGAGGGTAAACAACTGGCGGTATGGATGCGGCGGGACCAGAGAAAAATCACTCAGGG
TCAATGCCAGCGCTTCGTTAATACAGATGTAGGTGTTCCACAGGGTAGCCAGCAGCATCCTGCGATGCAGATCCGGAACA
TAATGGTGCAGGGCGCTGACTTCCGCGTTTCCAGACTTTACGAAACACGGAAACCGAAGACCATTCATGTTGTTGCTCAG
GTCGCAGACGTTTTGCAGCAGCAGTCGCTTCACGTTCGCTCGCGTATCGGTGATTCATTCTGCTAACCAGTAAGGCAACC
CCGCCAGCCTAGCCGGGTCCTCAACGACAGGAGCACGATCATGCGCACCCGTGGGGCCGCCATGCCGGCGATAATGGCCT
GCTTCTCGCCGAAACGTTTGGTGGCGGGACCAGTGACGAAGGCTTGAGCGAGGGCGTGCAAGATTCCGAATACCGCAAGC
GACAGGCCGATCATCGTCGCGCTCCAGCGAAAGCGGTCCTCGCCGAAAATGACCCAGAGCGCTGCCGGCACCTGTCCTAC
GAGTTGCATGATAAAGAAGACAGTCATAAGTGCGGCGACGATAGTCATGCCCCGCGCCCACCGGAAGGAGCTGACTGGGT
TGAAGGCTCTCAAGGGCATCGGTCGAGATCCCGGTGCCTAATGAGTGAGCTAACTTACATTAATTGCGTTGCGCTCACTG
CCCGCTTTCCAGTCGGGAAACCTGTCGTGCCAGCTGCATTAATGAATCGGCCAACGCGCGGGGAGAGGCGGTTTGCGTAT
TGGGCGCCAGGGTGGTTTTTCTTTTCACCAGTGAGACGGGCAACAGCTGATTGCCCTTCACCGCCTGGCCCTGAGAGAGT
TGCAGCAAGCGGTCCACGCTGGTTTGCCCCAGCAGGCGAAAATCCTGTTTGATGGTGGTTAACGGCGGGATATAACATGA
GCTGTCTTCGGTATCGTCGTATCCCACTACCGAGATATCCGCACCAACGCGCAGCCCGGACTCGGTAATGGCGCGCATTG
CGCCCAGCGCCATCTGATCGTTGGCAACCAGCATCGCAGTGGGAACGATGCCCTCATTCAGCATTTGCATGGTTTGTTGA
AAACCGGACATGGCACTCCAGTCGCCTTCCCGTTCCGCTATCGGCTGAATTTGATTGCGAGTGAGATATTTATGCCAGCC
AGCCAGACGCAGACGCGCCGAGACAGAACTTAATGGGCCCGCTAACAGCGCGATTTGCTGGTGACCCAATGCGACCAGAT
GCTCCACGCCCAGTCGCGTACCGTCTTCATGGGAGAAAATAATACTGTTGATGGGTGTCTGGTCAGAGACATCAAGAAAT
AACGCCGGAACATTAGTGCAGGCAGCTTCCACAGCAATGGCATCCTGGTCATCCAGCGGATAGTTAATGATCAGCCCACT
GACGCGTTGCGCGAGAAGATTGTGCACCGCCGCTTTACAGGCTTCGACGCCGCTTCGTTCTACCATCGACACCACCACGC
TGGCACCCAGTTGATCGGCGCGAGATTTAATCGCCGCGACAATTTGCGACGGCGCGTGCAGGGCCAGACTGGAGGTGGCA
ACGCCAATCAGCAACGACTGTTTGCCCGCCAGTTGTTGTGCCACGCGGTTGGGAATGTAATTCAGCTCCGCCATCGCCGC
TTCCACTTTTTCCCGCGTTTTCGCAGAAACGTGGCTGGCCTGGTTCACCACGCGGGAAACGGTCTGATAAGAGACACCGG
CATACTCTGCGACATCGTATAACGTTACTGGTTTCACATTCACCACCCTGAATTGACTCTCTTCCGGGCGCTATCATGCC
ATACCGCGAAAGGTTTTGCGCCATTCGATGGTGTCCGGGATCTCGACGCTCTCCCTTATGCGACTCCTGCATTAGGAAGC
AGCCCAGTAGTAGGTTGAGGCCGTTGAGCACCGCCGCCGCAAGGAATGGTGCATGCAAGGAGATGGCGCCCAACAGTCCC
CCGGCCACGGGGCCTGCCACCATACCCACGCCGAAACAAGCGCTCATGAGCCCGAAGTGGCGAGCCCGATCTTCCCCATC
GGTGATGTCGGCGATATAGGCGCCAGCAACCGCACCTGTGGCGCCGGTGATGCCGGCCACGATGCGTCCGGCGTAGAGGA
TCGAGATCTCGATCCCGCGAAATTAATACGACTCACTATAGGGGAATTGTGAGCGGATAACAATTCCCCTCTAGAAATAA
TTTTGTTTAACTTTAAGAAGGAGATATACCATGGCGGGTACCGGATCCGAATTCGAGCTCGTCGACAAGCTTGCGGCCGC
AGCTAGCCTCGAGGGTGAGAATCTTTATTTTCAGGGAGGTCATCACCATCACCATCACCATCACTAAGATCCGGCTGCTA
ACAAAGCCCGAAAGGAAGCTGAGTTGGCTGCTGCCACCGCTGAGCAATAACTAGCATAACCCCTTGGGGCCTCTAAACGG
GTCTTGAGGGGTTTTTTGCTGAAAGGAGGAACTATATCCGGAT
```
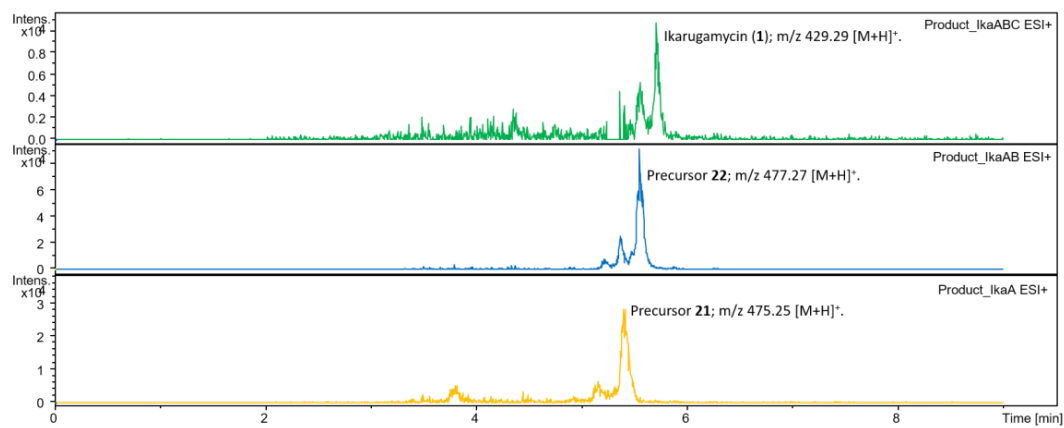
3. HPLC-MS traces of the enzymatic ikarugamycin production

a. using purified enzymes
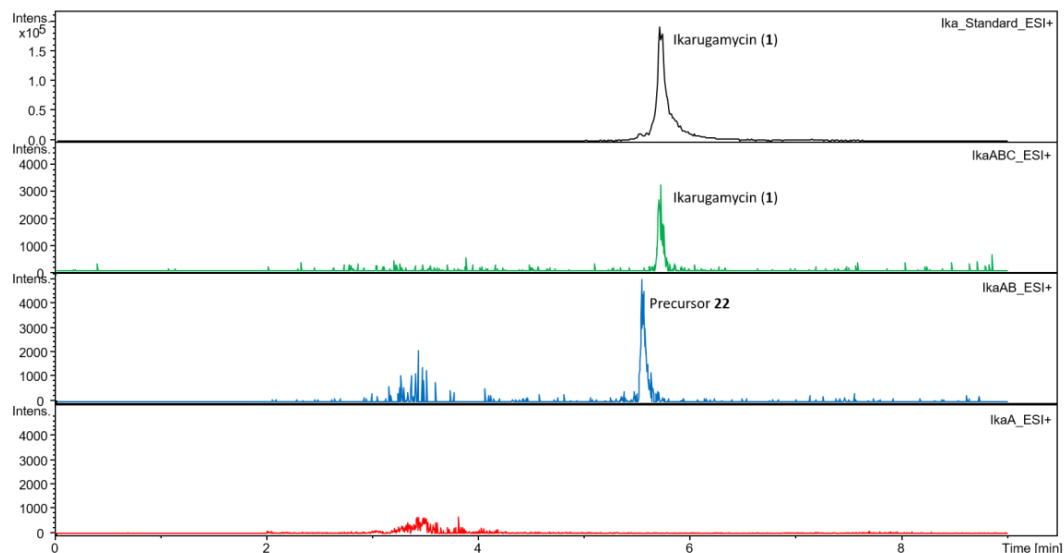
Figure S24. HPLC-MS analyses of enzymatic reactions for the synthesis of ikarugamycin (**1**).

### b. using cell lysates of the expression host

**Figure S25.** HPLC-MS analyses of enzymatic reaction using raw lysate extracts.

134

Interpretation of results with purified enzymes versus lysate assays. Omission of IkaC in the enzymatic reactions using purified enzymes led to the selective production of intermediate **22**. The application of IkaA alone, however, did not result in the expected formation of detectable amounts of the tetramic acid precursor **21** or its respective analog derived of a likewise conceivable early Diels-Alder cycloaddition. This was surprising, as such an IkaA product had been detected before in *ΔikaBC* knock-out mutants of ikarugamycin producing strains.[5],[6] When using lysed cell extracts of the *E. coli* recombinant hosts instead of purified IkaABC, the enzymatic reaction also furnished a compound whose mass data was consistent with **21** or a primary Diels-Alder product. Isolation of this ikarugamycin precursor failed due to decomposition, indicating that it is only stable within the stabilizing complex matrix of a bacterial raw extract, not under the defined conditions of our enzymatic setup. This suggests that the direct product of IkaA rather corresponds to the polyene **21**, with Diels-Alder cyclization only taking place during or after reductive bond formation catalyzed by IkaB. The observable differences in the product spectrum of in vitro (only **1**) versus lysate (**1**, **21**, **22**) assays with IkaABC is likely a result of the different stoichiometry (IkaA:B:C: in vitro 1:1:1; lysate 2:1:1) and/or an enhanced stability of IkaA in lysate versus in vitro reaction conditions.

4. NMR spectra of ikarugamycin (1**): synthetic** 1 **compared to authentic standard**
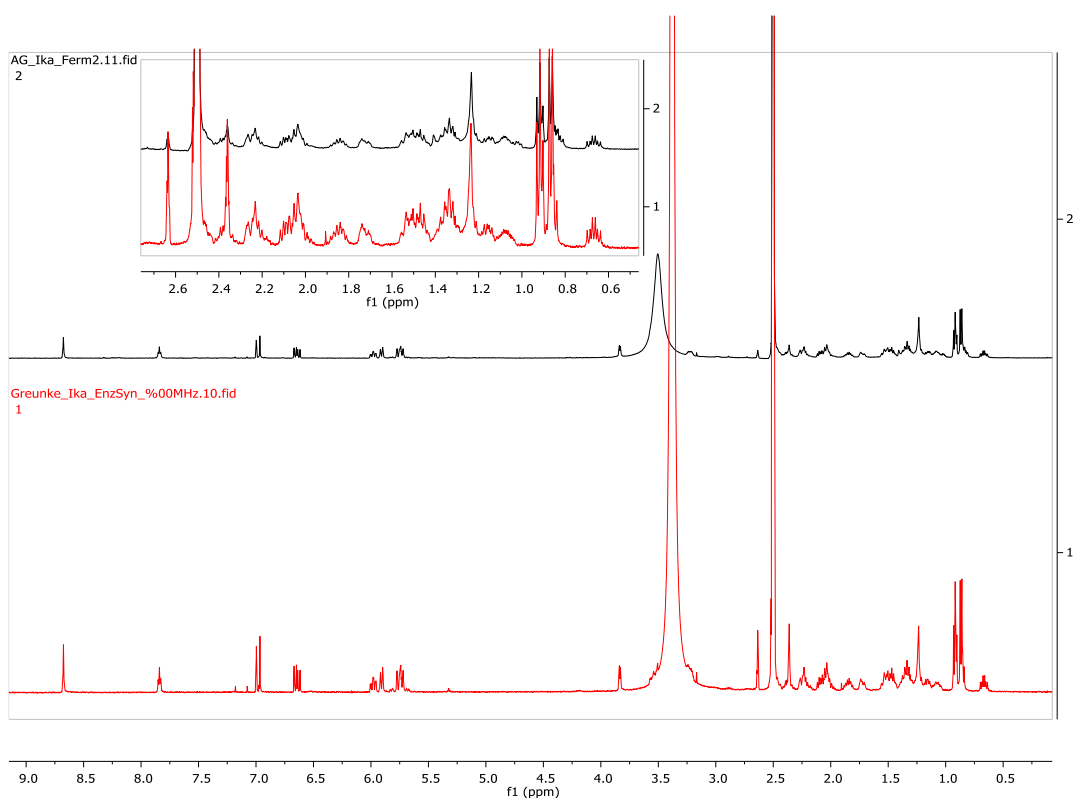
Figure S26. ^{1}H-NMR spectra (500 MHz, DMSO) of ikarugamycin (**1**) derived of enzymatic total synthesis (red) compared to an authentic standard isolated from *Streptomyces* sp. (black).
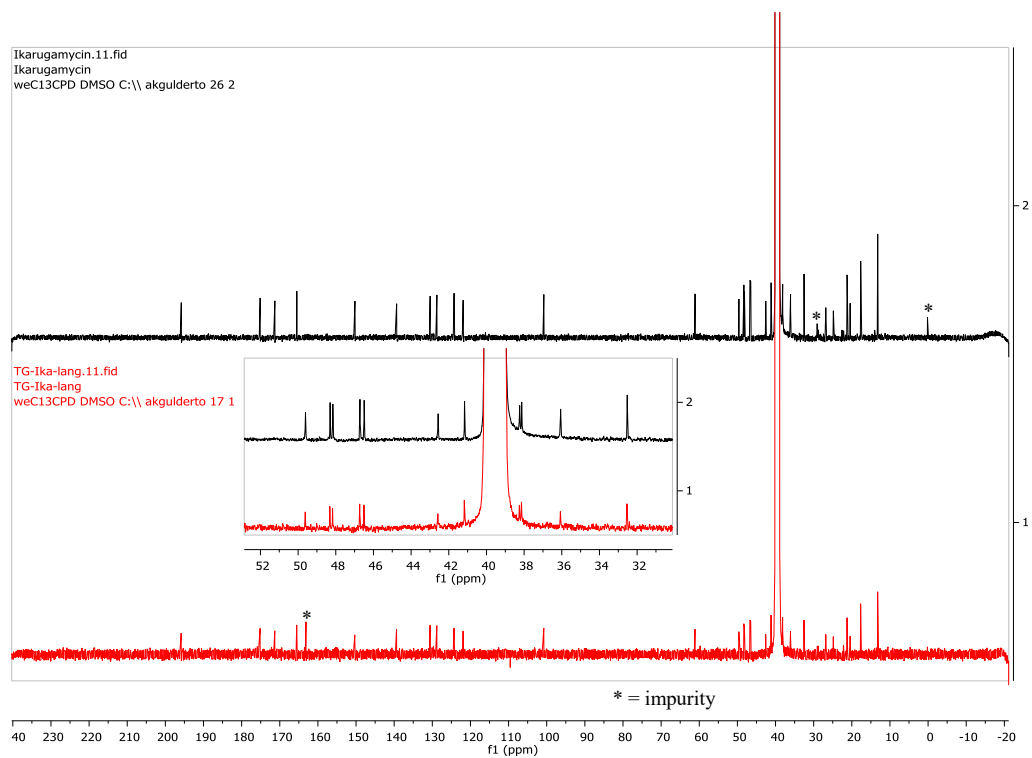
Figure S27. ^{13}C-NMR spectra (125 MHz, DMSO) of ikarugamycin (**1**) derived of enzymatic total synthesis (red) compared to an authentic standard isolated from *Streptomyces* sp. (black).

136

5. HPLC analyses of the biocatalytic precursor synthesis

a. Acetyl-CoA
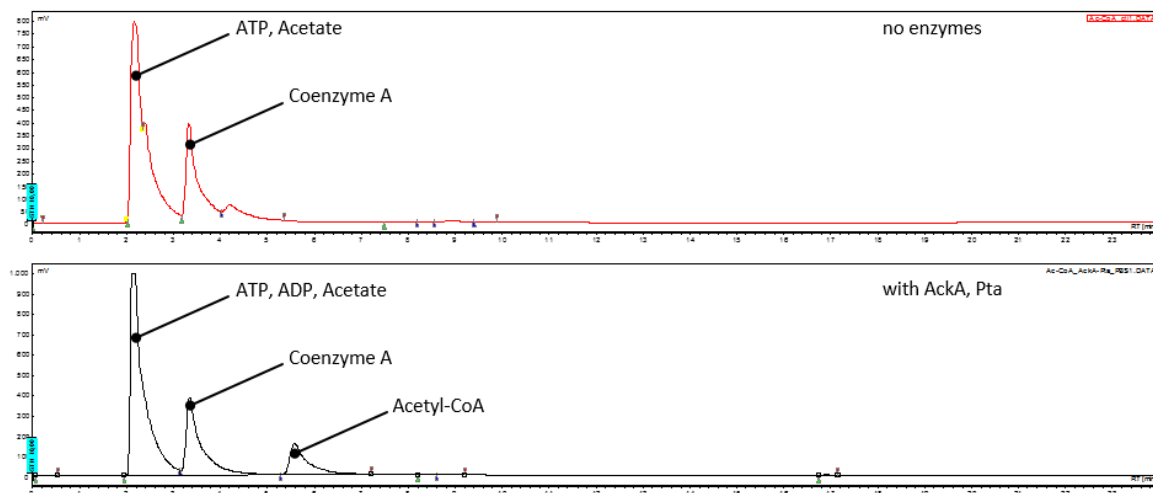
Figure S28. HPLC analyses of enzymatic reactions generating acetyl-CoA.

b. Malonyl-CoA
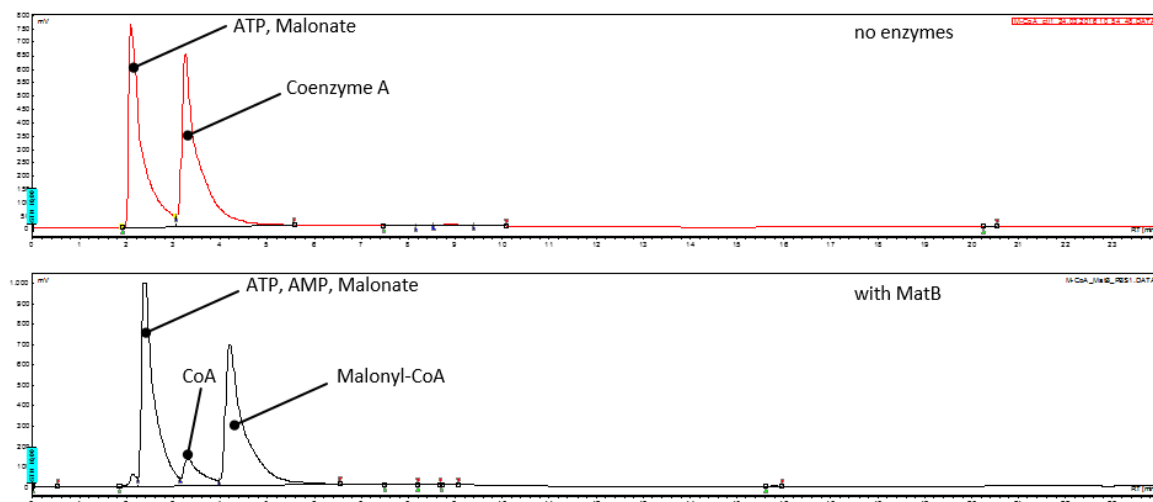
Figure S29. HPLC analyses of enzymatic reactions generating malonyl-CoA.

6. Table of primers

Table S1. List of oligonucleotides used for cloning in this study. Introduced sequence for restriction digest underlined.

Oligonucleotide	Sequence (5' → 3')
Tü6239_PKS-NRPS_EcoRI For	AGCT<u>GAATTC</u>ATGGATTCCATGCACCAC
Tü6239_PKS-NRPS_HindIII Rev	AGCT<u>AAGCTT</u>TCATACCTCGCCATCACC
Tü6239_PKS-NRPS_wo_TGA Rev	AGCT<u>AAGCTT</u>TACCTCGCCATCACCGCC
Tü6239_FAD_BamHI For	AGCC<u>TGGATCC</u>ATGACGCCTTTCGTTCAG
Tü6239_FAD_XhoI Rev	ATCTT<u>CTCGAG</u>CTACACGCTCCTCGACGA
Tü6239_AlcD_BamHI For	ACCT<u>GGATCC</u>GTGATCGCCGAGCACATCC
Tü6239_AlcD_XhoI Rev	ACTT<u>CTCGAG</u>CTACAGGGCGACCAGGAC
SA_AlcD_BamHI For	AGCC<u>TGGATCC</u>ATGAAAATCGACAAGTGG
SA_AlcD_XhoI Rev	ATCTT<u>CTCGAG</u>TCAGATCTCCACCAGGAC
Scoel_MatB_BamHI For	ATCAATT<u>GGATCC</u>TCCTCTCTCTTCCCGGCCCTCTCC
Scoel_MatB_XhoI Rev	CATCTAT<u>CTCGAG</u>TCAGTCACGGTTCAGCGCCCGC
Ecoli_AckA_BamHI For	ACCGTGT<u>GGATCC</u>ATGTCGAGTAAGTTAGTACTG
Ecoli_AckA_XhoI Rev	ATATT<u>CTCGAG</u>TCAGGCAGTCAGGCGGCT
Ecoli_Pta_EcoRI For	ACAGCTT<u>GAATTC</u>TCCCGTATTATTATGCTGATCC
Ecoli_Pta_HindIII Rev	TTCT<u>AAGCTT</u>TTACTGCTGCTGTGCAGA

7. References

[1]   B. A. Pfeifer, S. J. Admiraal, H. Gramajo, D. E. Cane, C. Khosla, *Science* **2001**, *291*, 1790-1792.

[2]   H. Inoue, H. Nojima, H. Okayama, *Gene* **1990**, *96*, 23-28.

[3]   A. J. Hughes, A. Keatinge-Clay, *Chem. Biol.* **2011**, *18*, 165-176.

[4]   M. C. Walker, B. W. Thuronyi, L. K. Charkoudian, B. Lowry, C. Khosla, M. C. Chang, *Science* **2013**, *341*, 1089-1094.

[5]   J. Antosch, F. Schaefers, T. A. M. Gulder, *Angew. Chem. Int. Ed. Engl.* **2014**, *53*, 3011-3014.

[6]   G. Zhang, W. Zhang, Q. Zhang, T. Shi, L. Ma, Y. Zhu, S. Li, H. Zhang, Y. L. Zhao, R. Shi, C. Zhang, *Angew. Chem. Int. Ed. Engl.* **2014**, *53*, 4840-4844.

A.1.3. Direct Pathway Cloning of the erythromycin gene cluster – Greunke *et al.* (unpublished)

The supplemental information is related to Chapter 3.4.2:

- **Greunke, C.**; Duell, E.R.; D'Agostino, P.M.; Glöckle, A.; Lamm, K.; Gulder, T.A.M. Direct Pathway Cloning (DiPaC) to Access Natural Product Biosynthetic Gene Clusters. *Manuscript in preparation.*

Direct Pathway Cloning (DiPaC) to Access Natural Product Biosynthetic Gene Clusters

Christian Greunke[a], Elke R. Duell[a], Paul M. D'Agostino[a], Anna Glöckle, Katharina Lamm and Tobias A. M. Gulder[a,b]

Biosystems Chemistry, Department of Chemistry and Center for Integrated Protein Science Munich (CIPSM), Technische Universität München, Lichtenbergstraße 4, 85748 Garching bei München, Germany.
[b]E-mail: tobias.gulder@tum.de.

Supporting Information

Table of contents

1. Materials and Methods

1.1. General Materials and Methods

Bacterial strains and plasmids.

Bacterial strains and plasmids used and generated in this study are listed in Table S1. Selection antibiotics were applied as follows: Ampicillin sodium salt (Amp) at 100 µg/mL; Kanamycin sulfate (Kan) at 50 µg/mL; Chloramphenicol (Cam) at 12.5 µg/mL; Nalidixic acid (NA) at 25 µg/mL.

DiPaC in silico simulations.

For the development of cloning strategies and primer design, SnapGene (GSL Biotech; Version 4.0.4) or the Geneious (1) software package (Version 8.1.9) and the NEBuilder Assembly web tool (New England Biolabs; http://nebuilder.neb.com) were used. Maps of plasmids were constructed using the SnapGene software. The sequenced genome of *Saccharopolyspora erythraea* NRRL 2338/DSM 40517 (erythromycin A; *ery* gene cluster) (2) was downloaded from NCBI database using the accession numbers NC_009142.

Primer design and PCR.

Primers for Gibson Assembly were designed using the NEBuilder Assembly tool. Gibson homology arms were composed of at least 18 nt with a calculated melting temperature of ≥ 50 °C. HiFi DNA assemblies were simulated using SnapGene, in this way excluding multiple binding sites of the primers. Also, to avoid severe secondary hairpin structures and primer dimer formation, evaluations were performed using the bioinformatic tools OligoAnalyzer (Integrated DNA Technologies; https://eu.idtdna.com/calc/analyzer) and the OligoEvaluator (Sigma-Aldrich; http://www.oligoevaluator.com).

A standard 25 µL PCR reaction batch for long-amplicon cycling reactions consisted of: 1X Q5 reaction buffer, 100-200 µM deoxynucleotide triphosphates, 500 nM of forward and reverse primer, highGC Enhancer (only for *ery* gene cluster), DNA template and 0.01-0.02 U/µL Q5 High-Fidelity DNA polymerase (NEB) were mixed. Template DNA amounts were 50 ng for *S. fonticola*, 15 ng for *N. punctiforme* and 100 ng for *S. erythraea*. Cycling was conducted using a T100 Thermal Cycler (Biorad) or a LifeECO Thermal Cycler (Biozym) as follows: 1.) Initial denaturation, 98 °C for 1 min; 2.) Denaturation, 98 °C for 10 sec; 3.) Annealing, T_A for 20 sec; 4.) Extension, 72 °C for t_{Ext}; steps 2.) to 4.) were repeated in total for 30 cycles; 5.) Final extension, 72 °C for t_{Ext}, and 6.) End phase, 16 °C. T_A was estimated using NEBs Tm Calculator tool (http://tmcalculator.neb.com) for the initial binding parts of the primers and t_{Ext} was approximately determined for 30 sec per 1,000 bp for low GC organisms or 45 sec per 1,000 bp for high GC organisms.

Colony screening PCRs were performed using Taq DNA polymerase (NEB). Clones were picked and resuspended in 40 µL of pure water and examined in a 25 µL PCR batch composed as follows: Taq buffer, 4 % DMSO, 100 µM deoxynucleotide triphosphates, 200 nM of forward and reverse primer, DNA template (bacterial suspension in water) and Taq DNA polymerase (0.025 U/µL, NEB) were mixed. Optimal cycling parameters were set as follows: 1.) Initial denaturation, 95 °C for 5 min; 2.) Denaturation, 95 °C for 45 sec; 3.)

Annealing, T_A for 30 sec; 4.) Extension, 72 °C for t_{Ext}; steps 2.) to 4.) were repeated in total 34 times; 5.) Final extension, 72 °C for t_{Ext}, at least 2 min, and 6.) End phase, 16 °C. T_A was estimated using NEBs Tm Calculator tool (http://tmcalculator.neb.com) and a minimal T_A of 47 °C at least. t_{Ext} was calculated using 1 min per 1,000 bp.

Cloning reagents.

Q5 High-Fidelity DNA polymerase, deoxynucleotide triphosphates, NEBuilder HiFi DNA Assembly Master Mix, various restriction endonucleases, if possible as High-Fidelity (HF) versions, and Antarctic phosphatase were purchased from NEB. T4 DNA ligase was purchased from Jena Biosciences. Oligonucleotides were from Sigma Aldrich, synthesis option "deprotected and desalted", resuspended at a stock concentration of 100 μM in pure water and stored on ice for immediate use or at -20 °C for long term storage. The peqGOLD Plasmid Miniprep Kit I (C-Line) from VWR was purchased to isolate plasmids from *E. coli* LB overnight cultures. For DNA extractions from agarose gels, the Monarch DNA Gel Extraction Kit (NEB) was purchased. PCR products or linearized vector backbones were purified using the PCR Purification Kit from Jena Bioscience. Agarose was purchased from Sigma-Aldrich. SERVA DNA Stain Clear G (Serva) was applied according to the manufacturer's instructions in prestained agarose gels. Rotiphorese TAE buffer was used for agarose gel electrophoresis and was purchased from Carl Roth as a 50-fold concentrate.

Setup of HiFi DNA Assembly reactions.

The HiFi DNA assemblies were composed as recommended by the NEB guidelines: For a two-fragment assembly, at least 0.02 pmol of prepared vector backbone together with a one- to twofold amount of insert were used in a total reaction batch of 20 μL. For three fragment assemblies, DNA components were applied in equimolar ratios in the range of 0.03 up to 0.2 pmol. In all cases, the isothermal reaction was performed at 50 °C for one hour.

Sanger sequencing of plasmid samples.

Sanger sequencing was performed at GATC Biotech (Konstanz, Germany). Results were retrieved as ab1-type files and were analyzed using Geneious software. Sequence alignments to the related reference sequences were performed using Geneious Alignment default settings.

1.2. Table of primers

Table S1. List of oligonucleotides used for cloning and screening procedures.

Oligonucleotide	Sequence (5' → 3')¥	Description
Vector primers		
T7 primer	TAATACGACTCACTATAGGG	Colony screening and sequencing
pET-RP	CTAGTTATTGCTCAGCGG	Colony screening and sequencing of (pET-28b backbone)
lac_prom_pSET152	CTTCCGGCTCGTATGTTG	Colony screening and sequencing of (pCC1FOS backbone)
Cloning of ery		
screen_ery_for	GTCATCGAAGGCATGACGAC	Screening primer for *ery*
screen_ery_rev	GTTCTACCGCAACGAGAACC	Screening primer for *ery*
ery-p1_for	ctcacagagaacagattggtggatccgGTGGTCTTCGACGGGGAGCT	Amplification of *ery*-p1 for insertion into pET-28b-SUMO
ery-p1_rev	tggcggccgttactagtgtacgtaTCGATCGAGATGTGCACG	Amplification of *ery*-p1 for insertion into pET-28b-SUMO
screen_ery-p1_for	TGATCCCATTCACCGGAG	Colony screening and sequencing of
ery-p2_for	tgcgtgcacatctcgatcgaTTTCCCGGATGCCGGTGTC	Amplification of *ery*-p2 for insertion into pET-28b-SUMO::*ery*-p1
ery-p2_rev	actggcggccgttactagtgtacgtagCAAATCCCGCGCCTTTAG	Amplification of *ery*-p2 for insertion into pET-28b-SUMO::*ery*-p1
screen_ery-p2_for	CTTCGCCGAGTCCTACAC	Colony screening and sequencing of
ery-p3_for	CCCTAAAGGCGGCGGGATTGC	Amplification of *ery*-p3 for insertion into pET-28b-SUMO::*ery*-p12
ery-p3_rev	actggcggccgttactagtgtacgtactACAGGTCCTCTCTCCCCC	Amplification of *ery*-p3 for insertion into pET-28b-SUMO::*ery*-p12
screen_ery-p3_for	TGGACGAGCTGGAGAAG	Colony screening and sequencing of
ery-p4_for	gcggggggaggacctgtagATGAGCGGGTGACAACGGC	Amplification of *ery*-p4 for insertion into pET-28b-SUMO::*ery*-p123
ery-p4_rev	actggcggccgttactagtgtacgtaTCATGAATTCCCTCCGCC	Amplification of *ery*-p4 for insertion into pET-28b-SUMO::*ery*-p123
screen_ery-p4_for	CCTGATCGACGTCTACCC	Colony screening and sequencing of
SUMO_ery-p5_for	ctcacagagaacagattggtgggatccGACCACTTCACGATGGTG	Amplification of *ery*-p5 for subcloning into pET-28b-SUMO
SUMO_ery-p5_rev	tggcggccgttactagtgtCAAGCCCCAGCCTTGAGG	Amplification of *ery*-p5 for subcloning into pET-28b-SUMO
screen_ery-p5_for	CCTTCAGCTTCTACCCGG	Colony screening and sequencing of
ery-p1234_HR_LP_pCC1FOS_for	gagtcgacctgcaggcatgcGCTGGTCTTCGACGGGGAGCT	Preparation of a LCHR-compatible helper plasmid of pCC1FOS, introduction of CAP_site_LEFT
ery-p1234_HR_LP_rev	atgaccgcataagcttCAGCCAGTCGAGGAGTGC	Preparation of a LCHR-compatible helper plasmid of pCC1FOS, introduction of CAP_site_LEFT
ery-p1234_HR_RP_for	cgactggctgaagcttATGCGGTCATCAGGACAC	Preparation of a LCHR-compatible helper plasmid of pCC1FOS, introduction of CAP_site_RIGHT

Primer	Sequence	Description
ery-p1234_HR_RP_pCC1FOS_rev	ggtgagactatagaatactcatgcat<u>TCATGA</u>ATTCCCTCCGCC	Preparation of a LCHR-compatible helper plasmid of pCC1FOS, introduction of CAP_site_RIGHT
ery-p5_for	tgggcgagggaattcatga<u>CCACGACCGATCGCGCC</u>	Amplification of *ery*-p5 for insertion into pCC1FOS::*ery*-p1234
pCC1FOS_ery-p5_rev	ggtgagactatagaatactcatgcat<u>TCAAGCCCCAGCCTTGAGG</u>	Amplification of *ery*-p5 for insertion into pCC1FOS::*ery*-p1234
pLK01_lin_for	ACCAGCATGGATAAAGGC	Preparation of a linearized fragment of pLK01 used for LCHR
pLK01_lin_rev	TGAGTATTCTATAGTCTCACCTAAATAG	Preparation of a linearized fragment of pLK01 used for LCHR

*Underlined sequences indicate restriction site. Specific nucleotide targets indicated in uppercase. Gibson homology arm sequences are indicated by lowercase.

1.3. Table of strains

Table S2. Strains used in this study.

Strains	Description	Reference or Source
E. coli		
DH5α	Host strain for cloning	NEB
TOP10	Host strain for cloning	Invitrogen
EPI300-T1	Host strain for cloning	Epicentre
GB08-red	Host strain for LCHR	(3, 4)
ET12567	Donor strain for conjugation	(5)
ET12567/pUB307	Helper strain for triparental conjugation	(6)
BAP1	Heterologous expression strain	(7)
Streptomyces		
Streptomyces coelicolor M1152	Host strain for heterologous expression	(8)
Streptomyces coelicolor M1154	Host strain for heterologous expression	(8)
Others		
Saccharopolyspora erythraea DSM 40517	Natural producer of erythromycin A	DSMZ

1.4. Table of plasmids

Table S3. Plasmids used and constructed in this study.

Plasmids	Description	Reference or Source
pET-28b-SUMO	Modified derivative of pET-28b used for assembly operations, Kan^R	(9)
pCC1FOS	Cloning vector used for assembly operations, Cam^R	Epicentre
pLK01	*Streptomyces* integrative vector used for the expression of *ery*; Kan^R	(10)
pET-28b-SUMO::*ery*-p1	Part 1 of the *ery* cluster of *S. erythraea*	This study
pET-28b-SUMO::*ery*-p12	Parts 1 and 2 of the *ery* cluster of *S. erythraea*	This study
pET-28b-SUMO::*ery*-p123	Parts 1, 2 and 3 of the *ery* cluster of *S. erythraea*	This study
pET-28b-SUMO::*ery*-p1234	Parts 1, 2, 3 and 4 of the *ery* cluster of *S. erythraea*	This study
pET-28b-SUMO::*ery*-p5	Part 5 of the *ery* cluster of *S. erythraea*	This study
pCC1FOS::HR_*ery*-p1234	Helper plasmid used for backbone exchange by LCHR; needs to be linearized by *Hin*dIII	This study
pCC1FOS::*ery*-p1234	Parts 1, 2, 3 and 4 of the *ery* cluster of *S. erythraea*	This study
pCC1FOS::*ery*	Complete *ery* cluster of *S. erythraea*	This study
pLK01::*ery*	Complete *ery* cluster of *S. erythraea*	This study

2. Cluster specific Material and Methods

2.1 Erythromycin

gDNA isolation.

S. erythraea DSM 40517 was purchased from the Leibniz Institute DSMZ, the German collection of microorganisms and cell cultures GmbH. *S. erythraea* was cultivated in GYM liquid medium (0.4 % glucose, 0.4 % yeast extract, 1.0 % malt extract, pH 7.2) at 28 °C while moderate shaking (200 rpm) for 7 days. The gDNA was isolated using the Bacterial DNA Preparation Kit (Jena Bioscience) according to the manufacturer's protocol. After hydration, the gDNA was diluted with 5 mM TE buffer (pH 8.5) to a stock concentration of 100 ng/µl and stored at -20 °C until use.

DiPaC strategy for ery.

Q5 polymerase (NEB) was used to amplify the *ery* cluster as five sections. The addition of HighGC Enhancer was mandatory. Each PCR batch with a volume of 25 µL contained 100 ng of gDNA of *S. erythraea* DSM 40517 as template. The vector backbone of pET-28b-SUMO was initially used as the cloning vehicle and fragments 1-4 were assembled in a sequential manner (Figure S1). At first, pET-28b-SUMO was linearized using *Bam*HI and dephosphorylated. With the insertion of parts *ery*-p1, 2, 3 and 4, a unique *Sna*BI restriction site on the 3′-end was added for backbone linearization (Table S4). This site was attached in the reverse primer and got artificially inserted within the fragments during the amplification process. *Sna*BI was removed during assembly so that the wild type sequence between the cluster fragments was maintained and the restriction site itself on the 3′-end was kept unique. All the pET-28b-SUMO constructs (see Table S3) were screened by the use of a suitable screening primer pair binding in the end of the insert in combination with the backbone primer pET-RP (Table S1). Due to the large size of pET-28b-SUMO::*ery*-p1234, the vector backbone was exchanged to a pCC1FOS based plasmid by LCHR using *E. coli* GB08-red to construct pCC1FOS::*ery*-p1234 (Figure 2). For successful LCHR the helper plasmid pCC1FOS::HR_*ery*-p1234 was firstly constructed. pCC1FOS::HR_*ery*-p1234 was built to contain one 500 bp capturing arm homologous to the beginning of *ery*-p1 and a second 479 bp capturing arm homologous to the end-sequence of *ery*-p4, was generated in a three fragment HiFi DNA Assembly reaction out of *Hin*dIII-linearized pCC1FOS and the PCR amplified capturing fragments. This construct was *Hin*dIII-linearized and dephosphorylated and then used for the LCHR experiment. To complete the pCC1FOS::*ery* vector, section 5 was subcloned to create pET-28b-SUMO::*ery*-p5, amplified from plasmid template and inserted in a last assembly reaction into pCC1FOS::*ery*-p1234 to generate pCC1FOS::*ery*. *E. coli* EPI300-T1^R colonies were screened using primers screen_ery-p5_for and lac_prom_pSET152. The expression construct for *Streptomyces*, pLK01::*ery*, was constructed by LCHR using pCC1FOS::*ery* as circular template and a compatible linear fragment of pLK01 (6617 bp) generated by PCR using primers pLK01_Lin_For/Rev (use of HighGC Enhancer mandatory). The pLK01 fragment can in principle be used to transfer every pCC1FOS-captured cluster, as it possesses homologous arms, in this work 380 and 1758 bp (10).

Homologous recombination to conduct vector backbone exchanges.

Backbone exchanges were performed by homologous recombination in *E. coli* GB08-red (LCHR). The bacteria were made electrocompetent by repeated washing steps with 10 % glycerol. The template plasmid, from which *ery* or parts of it should be transferred, was used for transformation. Cells were PCR-verified to ensure the successful uptake and were then used to prepare LB overnight cultures, grown under antibiotic pressure at 37 °C while shaking (200 rpm). 15 mL of LB medium were inoculated in the ratio 1:100 and cultures were grown to an OD_{600} value of about 0.2 to 0.3. Next, L-arabinose was added at a final concentration of 0.1 % to induce the recombinase system. Cultivation was proceeded until the OD_{600} was between 0.6 and 0.8. Cells were then made electrocompetent by repeated steps of washing with sterile, ice-cold 10 % glycerol. For this, 20, 15 and at last 10 mL of glycerol were applied. Centrifugation to pellet bacteria in between was performed at $5{,}000 \times g$ for 5 min at 4 °C. The pellet was resuspended in 300 µL of 10 % (v/v) glycerol solution to prepare three aliquots with a volume of 100 µL each and were stored on ice. Cells were then electroporated with the linearized capturing backbone, 100 to 200 ng of DNA were used, as well as a water control. The electroporation was carried out in a 0.2 cm ice-cold electroporation cuvette using a BioRad GenePulser II set to: 200 Ω, 25 µF and 2.5 kV. Directly after pulsing, 700 µL of SOC medium were added to the cells and were incubated at 37 °C while moderate shaking (200 rpm) for one hour. The bacterial suspension was streaked on LB agar plates with the corresponding antibiotic of the target backbone and cultivated overnight at 37 °C. Colonies of *E. coli* GB08-red were checked by PCR using suitable primer pairs. The minipreparation eluates from *E. coli* GB08-red overnight cultures were taken to transform plasmid storage strains (*E. coli* DH5α, TOP10, EPI300-T1^R). For LB overnight cultures of *E. coli* EPI300-T1^R containing pCC1FOS/pLK01, L-arabinose was added to a final concentration of 0.01 % to induce high-copy replication of plasmids. After isolation of the target plasmids, the repertoire of Screening-PCR and analytical restriction digest was used to identify the exchanged plasmid. End-sequencing of vector/insert border regions could be used in addition to proof correct integration.

Streptomyces conjugation.

Plasmid pLK01::*ery* was introduced into *S. coelicolor* M1152 and M1154 by triparental intergeneric conjugation with *E. coli* ET12567/pUB307 (6). In short, pLK01::*ery* was applied to transform electrocompetent *E. coli* ET12567. PCR-validated clones were selected and used to inoculate LB supplemented with Kan. Similarly, LB cultures of *E. coli* ET12567/pUB307 were grown. Overnight cultures of *E. coli* ET12567 pLK01::*ery* and *E. coli* ET12567/pUB307 were used to prepare LB cultures which were cultivated at 37 °C while moderate shaking (200 rpm). When reaching an OD_{600} value of about 0.5-0.6, bacteria were pelleted using the laboratory centrifuge ($5{,}000 \times g$, 5 min). The supernatants were discarded and cells washed twice under sterile conditions with LB medium. Pellets were resuspended in 500 µL LB medium and dense cellular suspensions were obtained. In forehand, spores of *S. coelicolor* M1152/M1154 were mixed with

TES buffer (50 mM, pH 8.0) and incubated at 50 °C for 10 minutes. Cells were then incubated further at 28 °C until use. For conjugation, *E. coli* ET12567 pLK01::*ery*, *E. coli* ET12567/pUB307 and activated spores were mixed. After pelleting by centrifugation (4,000 × *g*, 2 min), the supernatants were carefully removed and the cells resuspended in the remaining medium. The cellular suspension was streaked onto MS agar plates (2 % mannitol, 2 % soya flour, 10 mM MgCl$_2$, 60 mM CaCl$_2$, 2 % agar) (11) and cultivated at 30 °C. After 16 to 20 hours, agar plates were overlaid with an aqueous solution containing 1 mg NA and 1 mg Kan. Exconjugates were obtained after four to seven days of cultivation. A secondary selection was performed streaking single colonies on MS agar plates containing selection antibiotics (Kan, NA). Kan-resistant clones were selected, PCR verified and used to inoculate CASO medium (Carl Roth), supplemented with the appropriate antibiotics.

Heterologous production of erythromycin A.

As the expression host, *S. coelicolor* strains M1152 and M1154 were chosen. pLK01::*ery* was transferred by conjugation and exconjugates after two passages of antibiotic selection were taken and screened by for the presence of a 485 bp portion of *ery* (Figure 7). Positive clones were streaked on MS agar plates (2 % mannitol, 2 % soya flour, 10 mM MgCl$_2$, 60 mM CaCl$_2$, 2 % agar) and cultivated for 12 days at 30 °C. The agar was cut into small pieces and incubated with a solvent mixture consisting of EtOAc::MeOH (80 / 20 % [v/v]). After a 10 min incubation in the sonicating bath, the mixture was filtrated and evaporated *in vacuo*. The residues were resuspended in a 500 µL of MeOH, insoluble materials were removed by centrifugation (17,000 x *g*, 2 min) using a benchtop centrifuge and a cold-filtration step using syringe filters (PTFE; 0.2 µm particle filter). The samples were then subjected to LC-MS analysis. Commercially-available erythromycin (Sigma-Aldrich, catalogue no.: E5389-1G) was used to validate the retention times during HPLC-MS analysis.

HPLC-MS(/MS) analyses of extracts for erythromycin A.

Samples were analyzed using an UltiMate 3000 LC System system coupled to a LCQ Fleet Ion Trap Mass Spectrometer (Thermo Scientific). Interpretation of the recorded data was performed using the Thermo Xcalibur Qual Browser 2.2 SP1.48 software. The chromatographic HPLC separation was carried out on a Hypersil Gold aQ C18 column (150 × 2.1 mm, 3 µm particle size). The following liquid chromatography methods were applied with water (A) and acetonitrile (B) as the eluents, both buffered with 0.1 % formic acid. For fast extract screening (method 1), the gradient was set as follows: Preconditioning 5 % B (2.5 min); 5 % B (0 min) → 95 % B (8 min) → 100 % B (8.4 min) → 100 % B (10.8 min) → 5 % B (11.2 min) → 5 % B (12 min), at a constant flow rate of 0.7 mL/min. For a broad separation of components (method 2), the gradient was set as follows: Preconditioning 5 % B (4 min); 5 % B (0 min) → 50 % B (20 min) → 100 % B (20.5 min) → 100 % B (24.5 min) → 5 % B (25 min) → 5 % B (26 min), at a constant flow rate of 0.7 mL/min. Mass detection was recorded in the positive ionization mode. In the MS^2 measurements, 35 % normalized collision energy (cid) was applied to fragment the mass m/z = 734.5 (± 3.0 Da), and SRM was adjusted to fragment m/z = 576.22 (± 2.0 Da).

3. Results

3.1. Results for Erythromycin

Table S4. Restriction site strategy used for vector backbone linearizations to sequentially assemble *ery*.

Plasmid name	Restriction site	Description
pET-28b-SUMO (5650 bp)	*Bam*HI	Modified derivative of pET-28b used to insert *ery*-p1 (9.8 kb | 70 % GC) and subcloning of ery-p5 (12,7 kb | 71 % GC)
pET-28b-SUMO::*ery*-p1 (15,522 bp)	*Sna*BI	Intermediate fragment to insert *ery*-p2 (10.6 kb | 74 % GC)
pET-28b-SUMO::*ery*-p12 (26,099 bp)	*Sna*BI	Intermediate fragment to insert *ery*-p3 (12.1 kb | 73 % GC)
pET-28b-SUMO::*ery*-p123 (38,170 bp)	*Sna*BI	Intermediate fragment to insert *ery*-p4 (9.5 kb | 74 % GC)
pET-28b-SUMO::*ery*-p1234 (47,686 bp)	*Sna*BI	Insertion of *ery*-p5 (12.6 kb | 71 %), failed due to stability issues
pCC1FOS::*ery*-p1234 (50,168 bp)	*Nsi*I	Insertion of *ery*-p5 (12.6 kb | 71 %) yielding complete *ery*
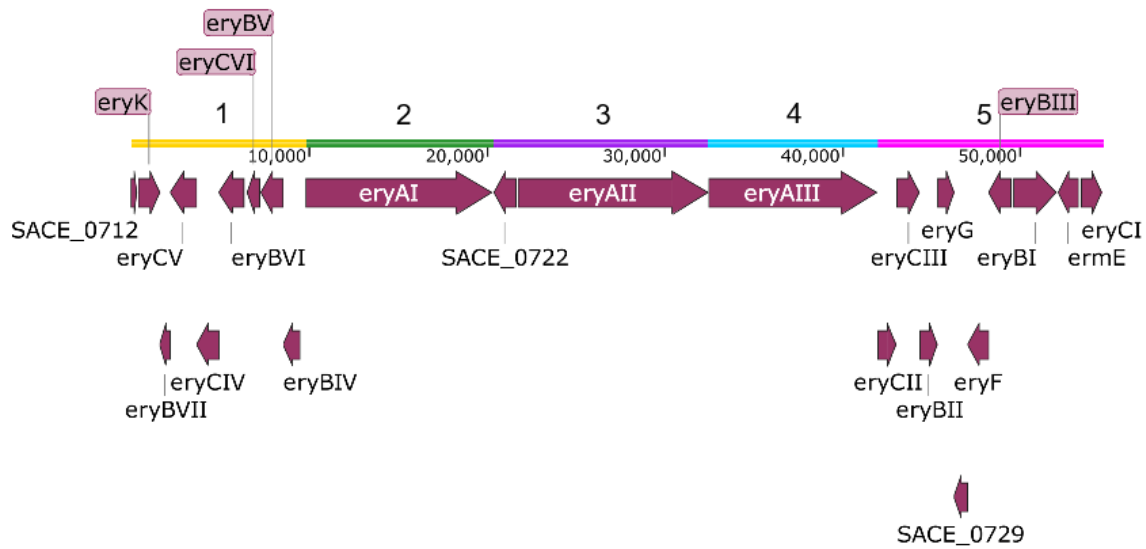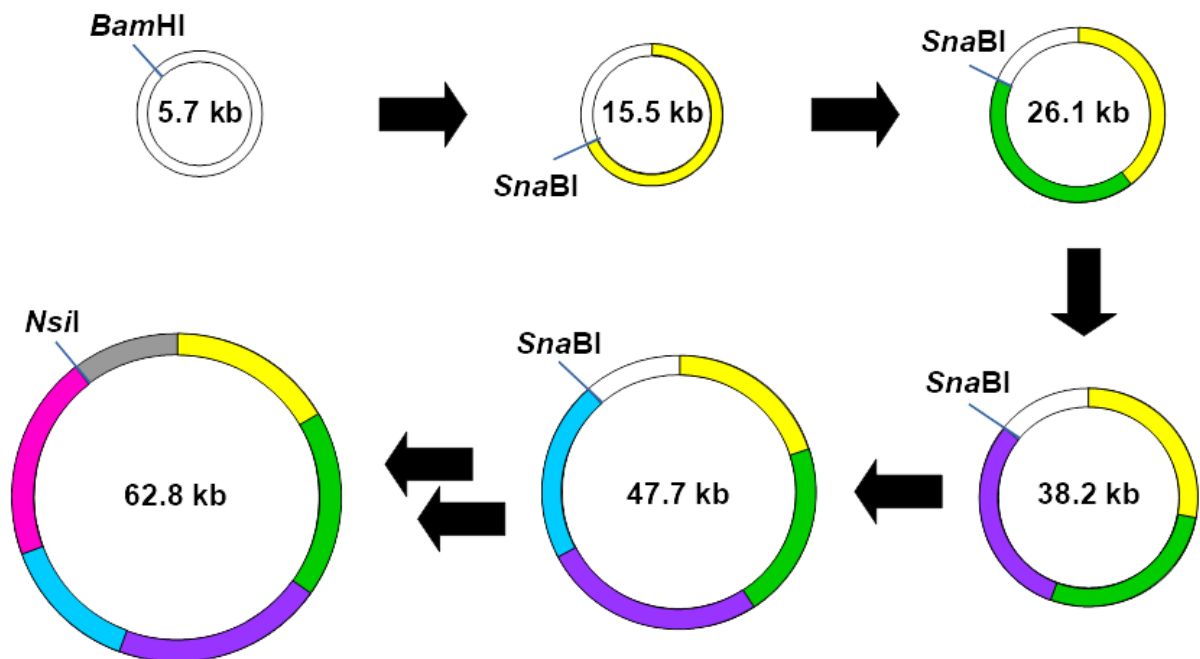
Figure S1. Partitioning of the *ery* cluster (54.6 kb, 73 % GC) for amplification and assembly. (A) The cluster was divided into five parts, *ery*-p1 (9.8 kb, 70 % GC, yellow), *ery*-p2 (10.6 kb, 74 % GC, green), *ery*-p3 (12.1 kb, 73 % GC, purple), *ery*-p4 (9.5 kb, 74 % GC, turquoise) and *ery*-p5 (12.6 kb, 71 % GC, magenta) which were amplified by PCR. (B) Sequential assembly strategy to reconstitute complete *ery*.
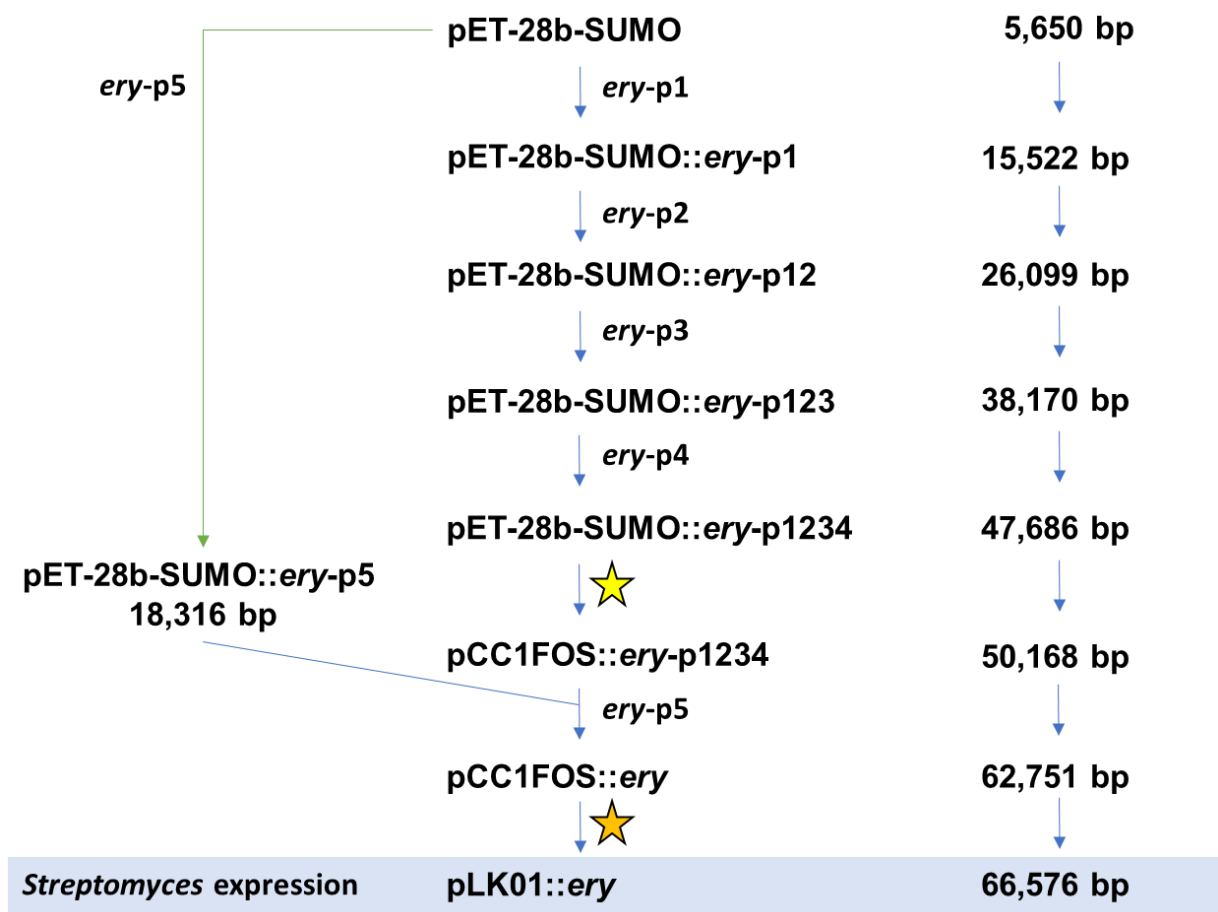
Figure S2. Sequential assembly strategy to access *ery* (54.6 kb). The BGC from *S. erythraea* DSM 40517 was split into five parts (*ery*-p1 to *ery*-p5) which were amplified by PCR. pET-28b-SUMO was the vehicle to assemble the clusters parts *ery*-p1 to -4. A backbone exchange for pCC1FOS was conducted by homologous recombination (LCHR), because of stability issues of pET-28b-SUMO::*ery*-p1234 in *E. coli* DH5α (yellow star). *ery*-p5 was subcloned first and inserted within pCC1FOS::*ery*-p1234 after amplification from plasmid template. LCHR was applied to transfer *ery* into pLK01 for heterologous expression in *Streptomyces* (orange star).
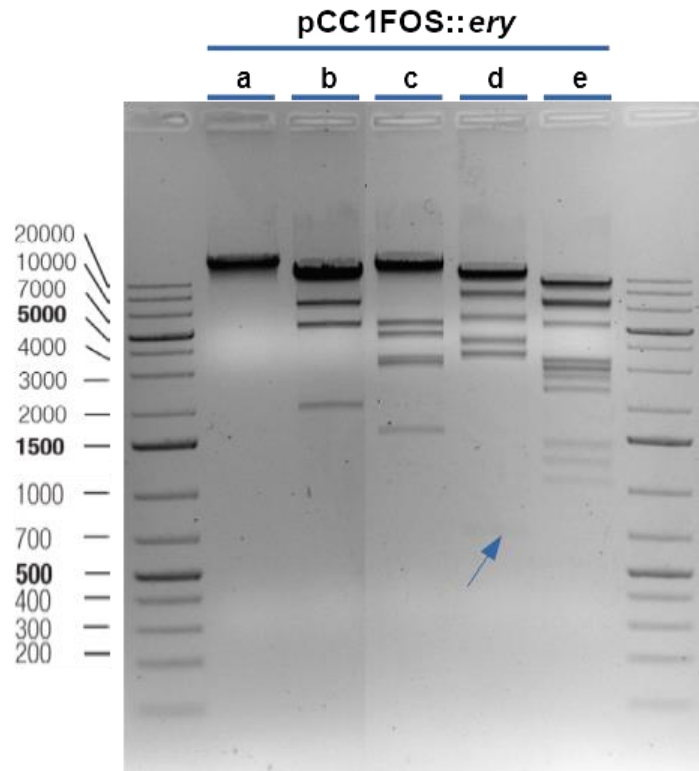
Figure S3. Agarose gel electrophoresis to analyze restriction enzyme digests of pCC1FOS::*ery*: (a) *Nsi*I-HF (62,751 bp), (b) *Bmt*I + *Xba*I (25,763; 22,514; 7317; 5318; 1839 bp), (c) *Hin*dIII-HF + *Sbf*I-HF (45,645; 5226; 4501; 3087; 2873; 1419 bp), (d) *Psp*XI (18,099; 16,391; 8107; 5456; 3922; 3840; 3183; 3135; 618 bp) and (e) *Not*I-HF (12,166; 11,072; 7498; 6956; 5226; 3129; 3075; 2808; 2785; 2503; 2175; 1293; 1119; 946 bp)
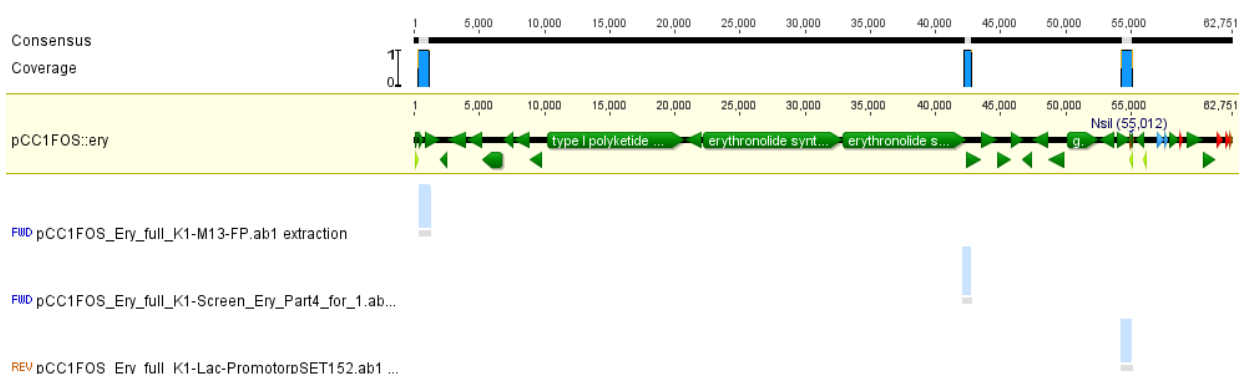
Figure S4. End-sequencing of pCC1FOS::*ery* to confirm correct integration of *ery*-p5. The reference sequence (top line, yellow) was aligned with three sequencing traces covering the overlap of pCC1FOS to the beginning of *ery* (line 2) and *ery-p5* (lines 3, 4).

153
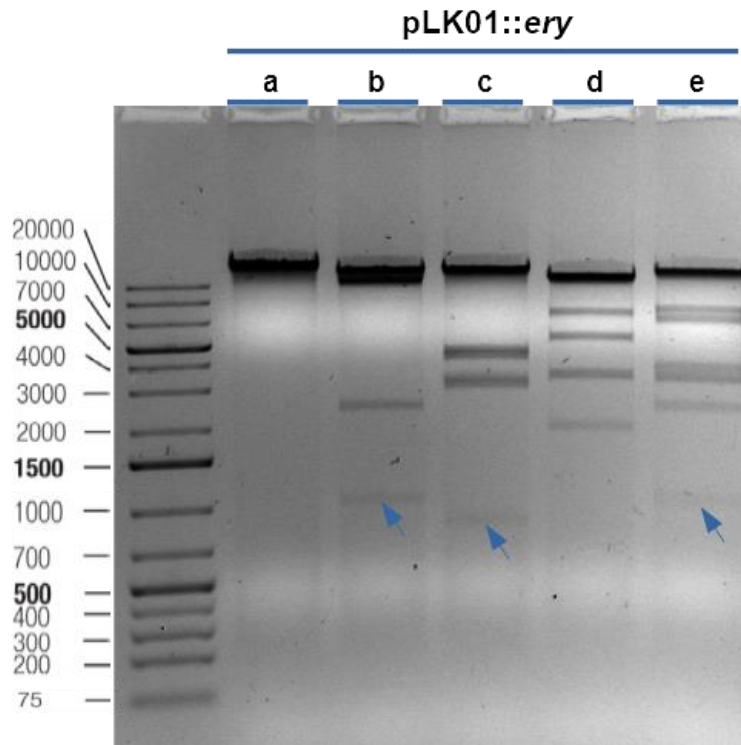
Figure S5. Agarose gel electrophoresis to analyze restriction enzyme digests of pLK01::*ery*: (a) *Nsi*I (66,576 bp), (b) *Hin*dIII (45,645; 17,676; 2286; 969 bp), (c) *Sbf*I (51,194; 4501; 4119; 3087; 2873; 802 bp), (d) *Bmt*I + *Xba*I (26,318; 22,514; 7317; 5318; 3270; 1839 bp) and (e) *Nde*I + *Hin*dIII (42,101; 7672; 6776; 3544; 3228; 2286; 969 bp).
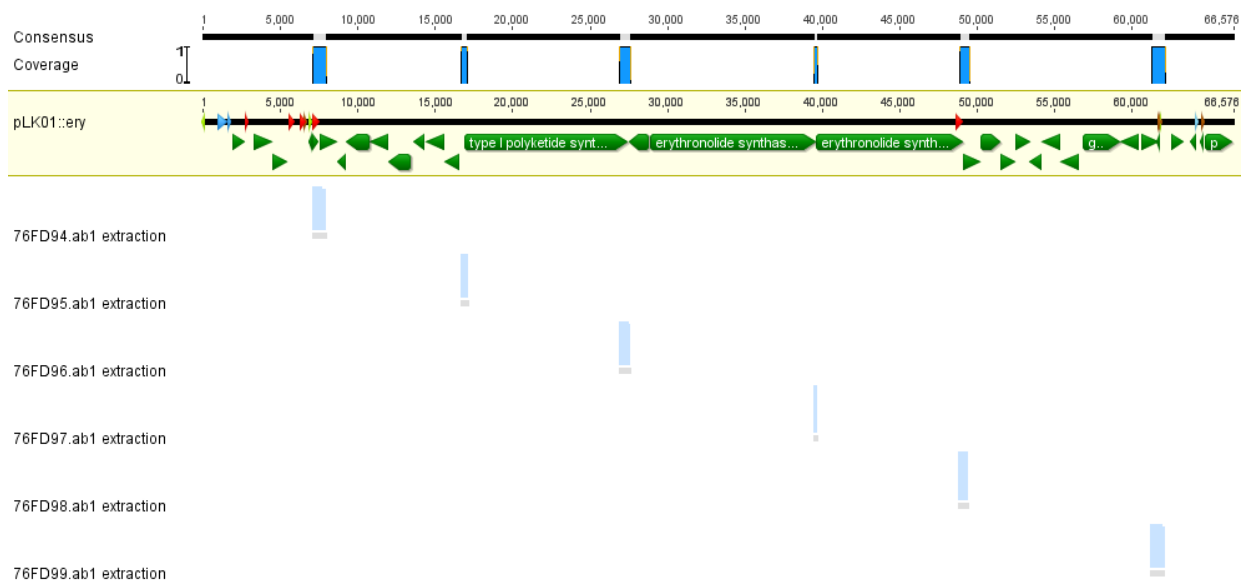
Figure S6. End-sequencing of pLK01::*ery* to confirm the presence of *ery*. The reference sequence (top line, yellow) was aligned with six sequencing traces covering the overlap of the pLK01 backbone to *ery*, and the overlapping regions of the assembled fragments *ery*-p1 to -p5 (lines 2 to 7).
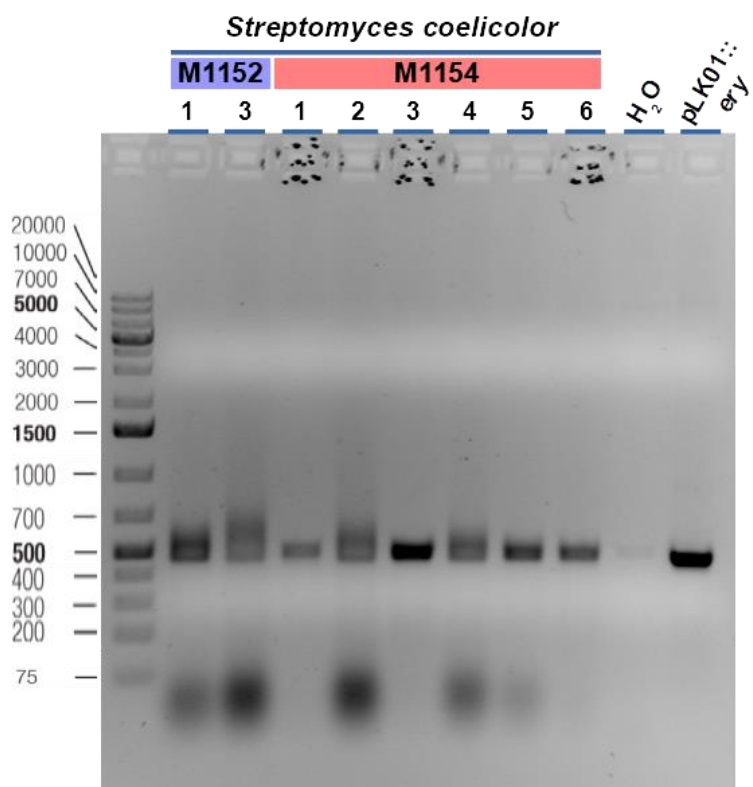
Figure S7. Agarose gel electrophoresis to analyze PCR screenings of CASO precultures of *S. coelicolor* for the presence of *ery*. An amplicon of expected size (485 bp) could be found in all cases, validating a successful conjugational transfer. As controls, reactions with water and plasmid template (pLK01::*ery*) were included.
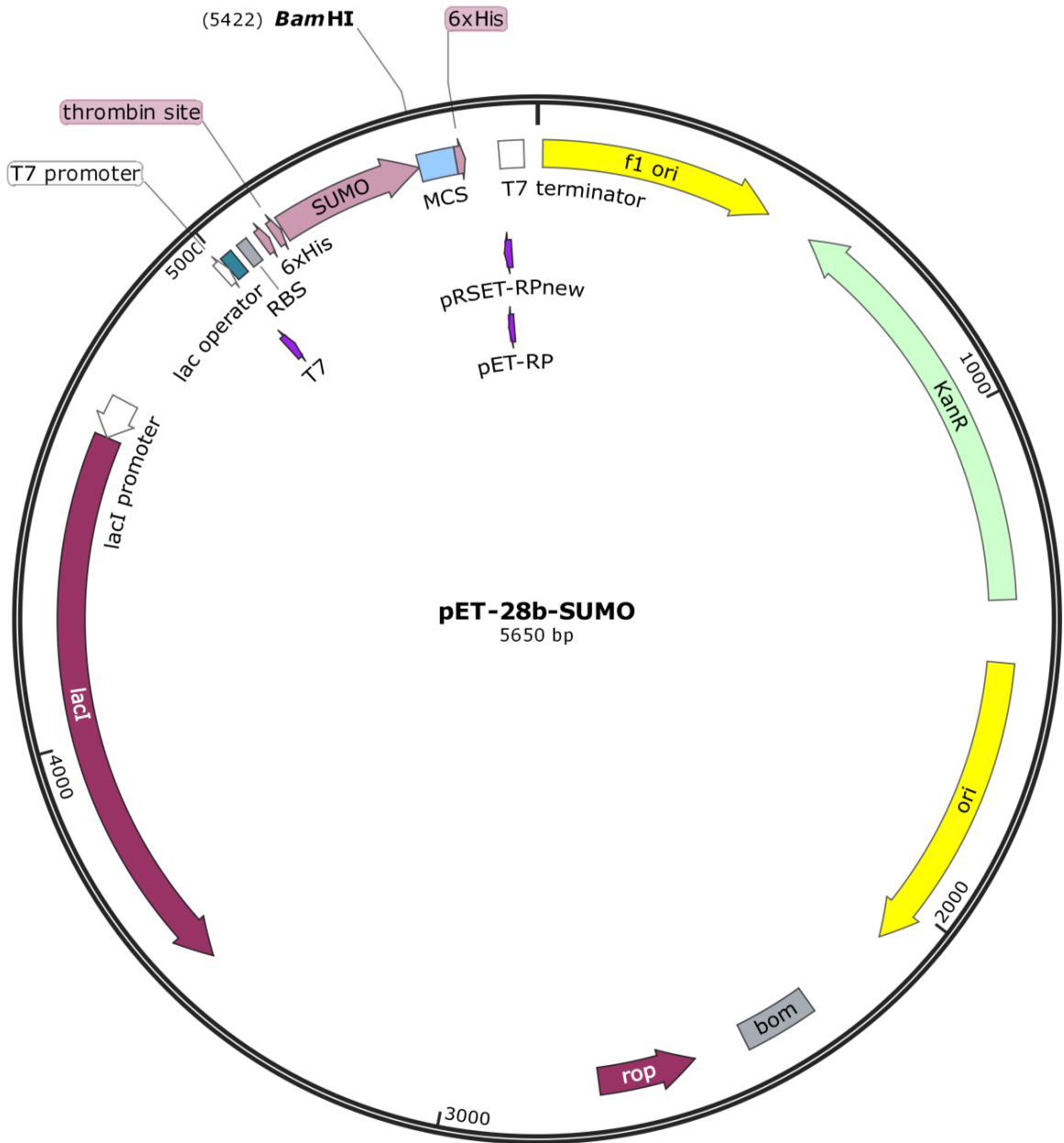
Figure S8. Plasmid map of pET-28b-SUMO (5650 bp).

Figure S9. Plasmid map of pET-28b-SUMO::*ery*-p1234 (47,686 bp).
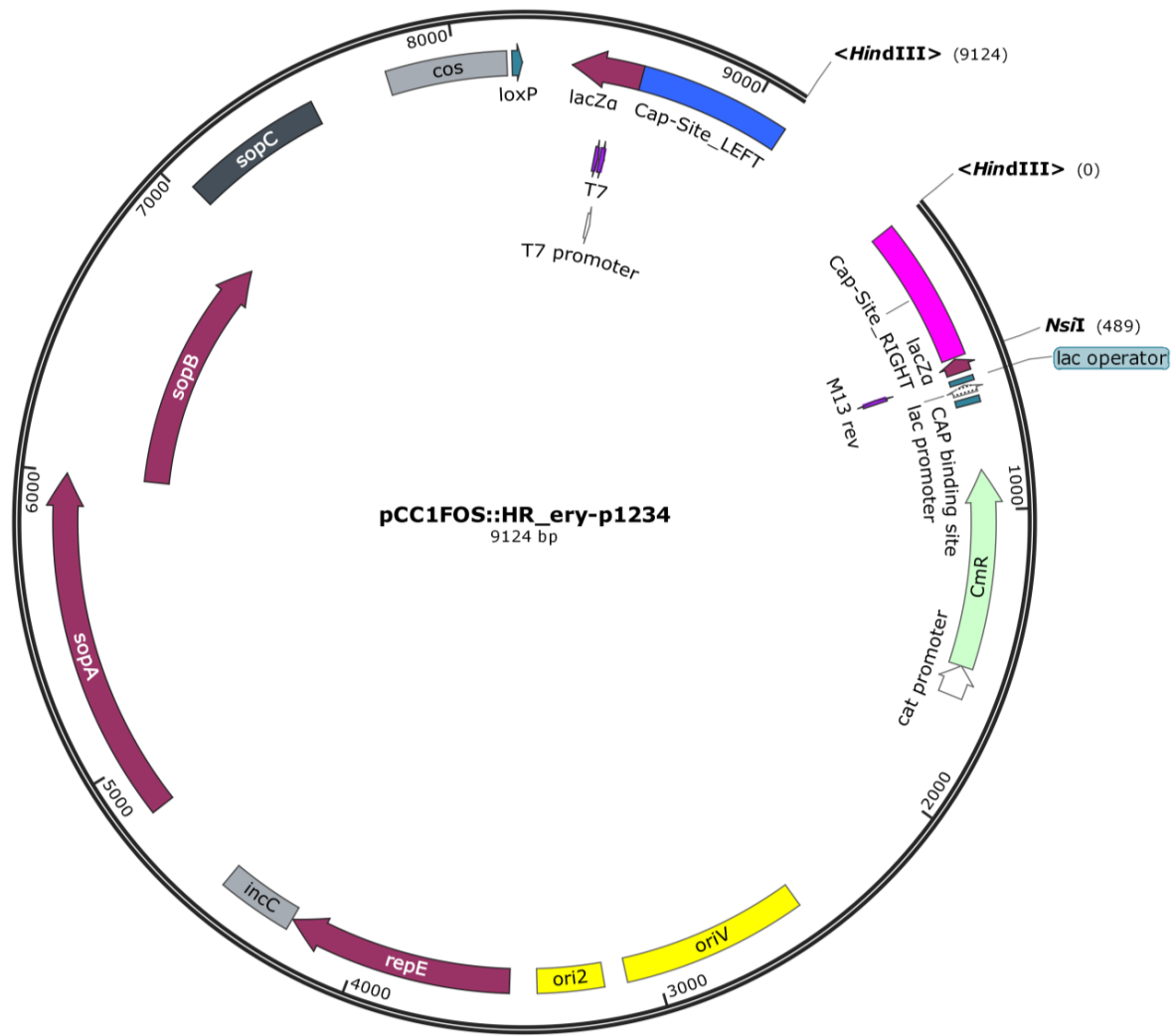
Figure S10. Plasmid map of *Hin*dIII-linearized pCC1FOS::HR_*ery*-p1234 (9124 bp) used to conduct a backbone exchange (pET-28b-SUMO → pCC1FOS) for part *ery*-p1234.
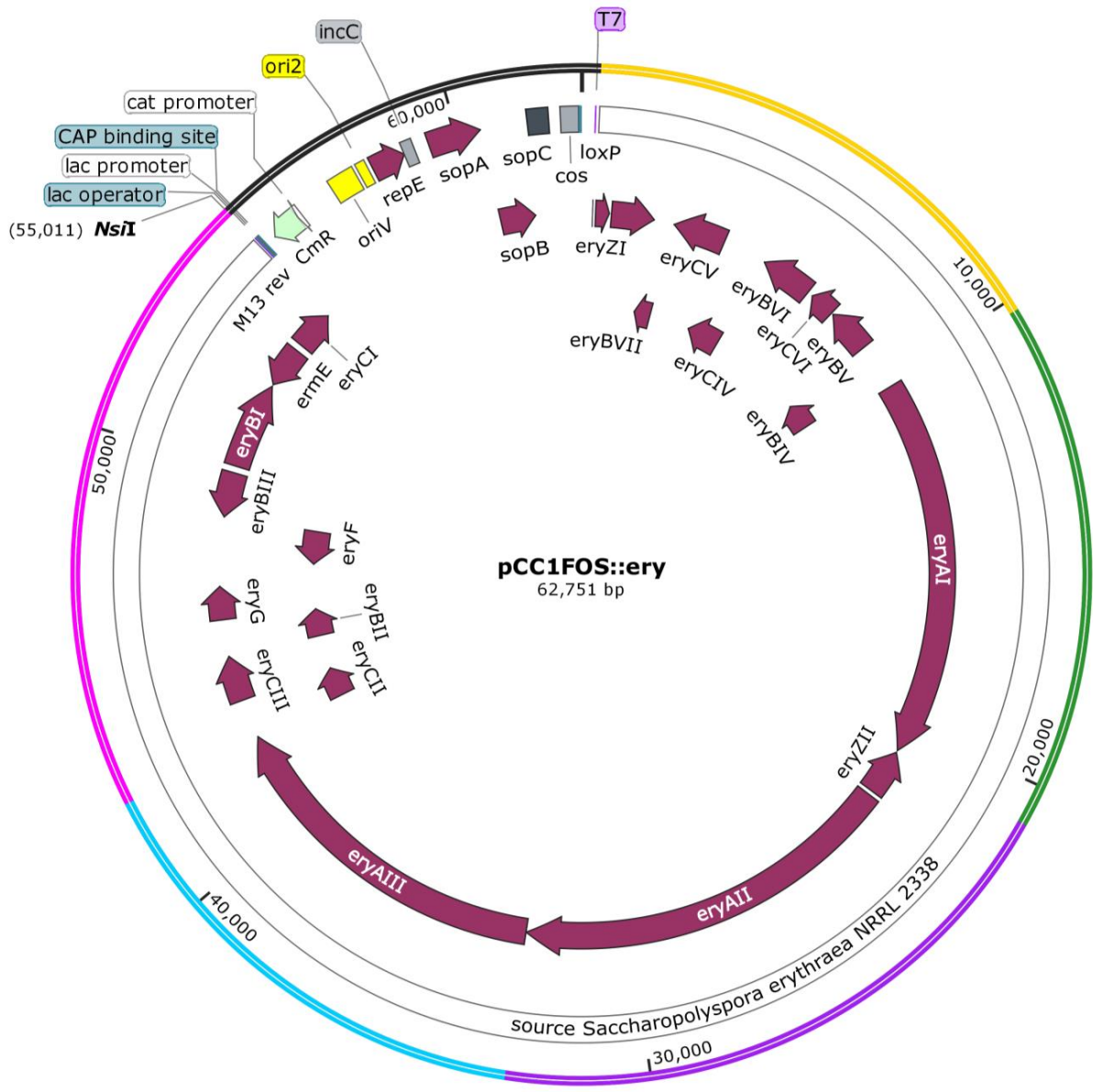
Figure S11. Plasmid map of pCC1FOS::*ery* (62,751 bp).

Figure S12. Plasmid map of pLK01::*ery* (66,576 bp) used for heterologous expression in *S. coelicolor*.

4. References

1. Kearse M, *et al.* (2012) Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* 28(12):1647-1649.

2. Oliynyk M, *et al.* (2007) Complete genome sequence of the erythromycin-producing bacterium *Saccharopolyspora erythraea* NRRL23338. *Nat. Biotechnol.* 25(4):447-453.

3. Fu J, *et al.* (2012) Full-length RecE enhances linear-linear homologous recombination and facilitates direct cloning for bioprospecting. *Nat. Biotechnol.* 30(5):440-446.

4. Wang H, *et al.* (2016) RecET direct cloning and Redαβ recombineering of biosynthetic gene clusters, large operons or single genes for heterologous expression. *Nat. Protoc.* 11(7):1175-1190.

5. MacNeil DJ, *et al.* (1992) Analysis of *Streptomyces avermitilis* genes required for avermectin biosynthesis utilizing a novel integration vector. *Gene* 111(1):61-68.

6. Flett F, Mersinias V, & Smith CP (1997) High efficiency intergeneric conjugal transfer of plasmid DNA from *Escherichia coli* to methyl DNA-restricting streptomycetes. *FEMS Microbiol. Lett.* 155(2):223-229.

7. Pfeifer BA, Admiraal SJ, Gramajo H, Cane DE, & Khosla C (2001) Biosynthesis of complex polyketides in a metabolically engineered strain of E. coli. *Science* 291(5509):1790-1792.

8. Gomez-Escribano JP & Bibb MJ (2011) Engineering *Streptomyces coelicolor* for heterologous expression of secondary metabolite gene clusters. *Microb Biotechnol* 4(2):207-215.

9. Baer P, *et al.* (2014) Induced-fit mechanism in class I terpene cyclases. *Angew. Chem. Int. Ed. Engl.* 53(29):7652-7656.

10. Kaysser L, *et al.* (2012) Merochlorins A-D, cyclic meroterpenoid antibiotics biosynthesized in divergent pathways with vanadium-dependent chloroperoxidases. *J. Am. Chem. Soc.* 134(29):11988-11991.

11. Wang XK & Jin JL (2014) Crucial factor for increasing the conjugation frequency in *Streptomyces netropsis* SD-07 and other strains. *FEMS Microbiol. Lett.* 357(1):99-103.

A.2. Constructed plasmids – PoTeM (accessory) enzymes

Within the PoTeM project, genes from various available PoTeM biosynthetic gene clusters were subcloned (Table A.1). All constructs were verified by restriction analysis and Sanger sequencing.

Table A.1.   List of cloned PoTeM (accessory) genes.

Source	Gene	Plasmid(s)
L. capsici DSM 19286	*sd*	pHis8, pGS-21a
	sd_orf2	pHis8, pGS-21a
S. degradans DSM 17024	*fad*	pHis8, pGS-21a
	sd	pHis8, pGS-21a, pXY200
	pks-nrps	pET-28b-SUMO
S. spinosa NRRL 18395	*fad1*	pHis8-TEV
	fad2	pHis8-TEV, pGS-21a-TEV
S. albus DSM 40313	*fad1*	pHis8-TEV, pGS-21a-TEV
	fad2	pHis8-TEV, pGS-21a-TEV
	alcd	pHis8-TEV, pGS-21a-TEV
S. arenicola CNS-205	*ikaA*	pHis8, pET-28b-SUMO
	ikaB	pHis8, pGS-21a, pET-28b-SUMO
	ikaC	pHis8(-TEV), pHis8C-TEV, pET-28b-SUMO, pRSET_A_MRGS_His6-TEV
S. griseus DSM 40236	*fad1*	pHis8-TEV, pGS-21a-TEV
	fad2	pHis8-TEV, pGS-21a-TEV
	alcd	pHis8-TEV, pGS-21a-TEV
S. roseosporus NRRL 15998	*fad1*	pET-28b-SUMO
	fad2	pHis8-TEV, pGS-21a-TEV
	alcd	pHis8-TEV, pGS-21a-TEV
Streptomyces sp. Tü6239	*ikaA*	pHis8, pHis8C-TEV, pET-28b-SUMO
	ikaB	pHis8(-TEV), pHis8C-TEV, pGS-21a(-TEV), pXY200/211, pET-28b-SUMO
	ikaC	pHis8(-TEV), pHis8C-TEV, pGS-21a, pXY200/201/211, pET-28b-SUMO
	ikaA_Δnrps	pHis8, pBluescript II SK(–)
	ikaA_Δpks	pHis8, pBluescript II SK(–)
	ikaA-TE	pHis8-TEV, pGS-21a-TEV
Streptomyces sp. Tü6314	*fad1*	pHis8-TEV, pGS-21a
	fad2	pHis8-TEV, pGS-21a
	alcd	pHis8-TEV, pGS-21a
		pAEM35, pAEM35-oriT

A.3. Cloned PoTeM gene clusters

Table A.2.   List of cloned (putative) PoTeM biosynthetic gene clusters.

Source	Size and GC content	Plasmid(s)
Streptomyces sp. Tü6239	12.3 kb, 72 %	pET-28b-SUMO
S. arenicola CNS-205	12.2 kb, 69 %	pET-28b-SUMO
S. degradans DSM 17024	12.3 kb, 47 %	pET-28b-SUMO
S. spinosa NRRL 18395	16.0 kb, 69 %	pET-28b-SUMO
Streptomyces sp. Tü6314	16.1 kb, 72 %	pET-28b-SUMO
S. griseus DSM 40236	16.2 kb, 72 %	pET-28b-SUMO, pSET152-*ermE*
L. capsici DSM 19286	19.3 kb, 66 %	pET-28b-SUMO

A.4. Constructed plasmids – Tü6239-IkaA A domain mutagenesis

Table A.3.   List of plasmids constructed for A domain mutagenesis within Tü6239-PKS/NRPS (IkaA). The respective codons to achieve the amino acid substitutions in full length IkaA were introduced by either SLIM[a] according to Chiu *et al.*[96] or HiFi-DNA Assembly site-directed mutagenesis[b].

Plasmid	Description
pBluescript II SK(−)::Tü6239-*ikaA*_A-domain-spez	Subcloning
mut2: V2495A, G2498E	Mutagenesis[a]
mut3: V2495A, G2498E, E2537D	Mutagenesis[b]
mut4: V2495A, G2498E, E2537D, S2589T	Mutagenesis[b]
mut5: V2495A, G2498E, E2537D, S2589T, I2597V	Mutagenesis[b]
pHis8C-TEV::Tü6239-*ikaA*_5mut	Protein synthesis in *E. coli*
pSET152-*ermE*::Tü6239-*ikaABC*_5mut	*Streptomyces* expression

A.5. Constructed plasmids – Erythromycin project

Table A.4.   List of plasmids constructed in the erythromycin project.

Plasmid	Description
pET-28b-SUMO::Ery-Part1	Intermediate plasmid
pET-28b-SUMO::Ery-Part12	Intermediate plasmid *eryAI*
pET-28b-SUMO::Ery-Part123	Intermediate plasmid *eryAI*, *eryAII*
pET-28b-SUMO::Ery-Part1234	Intermediate plasmid *eryAI*, *eryAII*, *eryAIII*
pET-28b-SUMO::Ery-Part5	Subcloning of Ery-Part5
pET-28b-SUMO::DEBS1+TE	*E. coli* plasmid for pyrone production
pCC1FOS::Ery-Part1234	Intermediate plasmid
pCC1FOS::Ery_full_GA	Plasmid harboring complete *ery*
pLK01::Ery_full_DSM40517	*Streptomyces* expression plasmid
pETDuet-1::*prpE*::*pccB*_rbs-*accA2*	*E. coli* helper plasmid for the production of methylmalonyl-CoA (**12**)
pETDuet-1-RBS2M::*prpE*::*pccB* _rbs-*accA2*	*E. coli* helper plasmid for the production of methylmalonyl-CoA (**12**)

A.6. NCBI BLAST search for PoTeM gene clusters

The analysis was conducted using the 'Discontiguous Megablast' program provided by Geneious 8.1.9 and NCBI's database 'nr'. As query sequences, the TE parts of the PKS/NRPS genes of *Streptomyces* sp. Tü6239[77] and *L. enzymogenes* C3[60] were used.

- *Actinoalloteichus hymeniacidonis* strain HPA177(T) (=DSM 45092(T))
- *Actinoalloteichus* sp. ADI127-7
- *Actinoalloteichus* sp. GBA129-24
- *Gynuella sunshinyii* YC6258
- ***Lysobacter capsici* DSM 19286**
- *Lysobacter capsici* strain 55
- *Lysobacter enzymogenes* strain C3 genome [HSAF]
- *Lysobacter enzymogenes* strain C3 HSAF biosynthetic gene cluster
- *Lysobacter gummosus* strain 3.2.11
- ***Saccharophagus degradans* 2-40 (DSM 17024)**
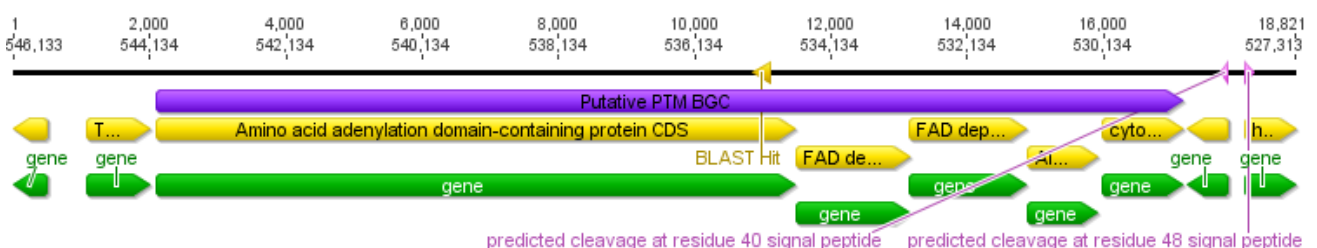- ***Saccharopolyspora spinosa* NRRL 18395**
- ***Salinispora arenicola* CNS-205**
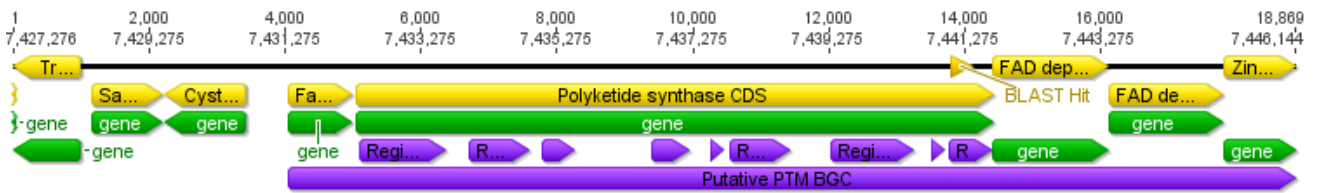- *Streptomyces albus* J1074 ($\neq$ **available DSM 40313**)
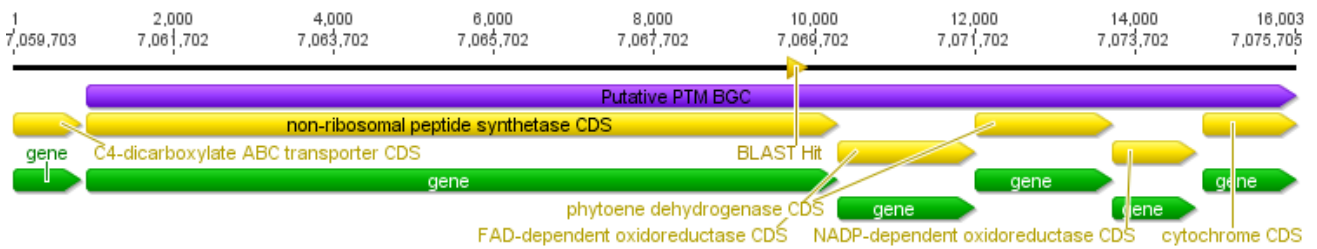- *Streptomyces albus* strain SM254
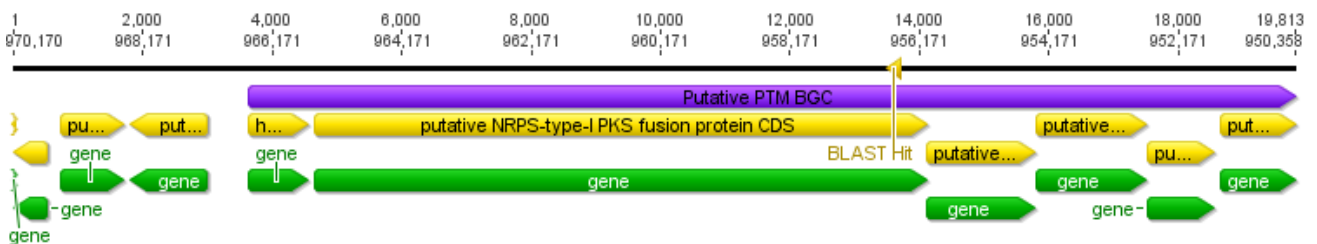- *Streptomyces clavuligerus* strain F613-1
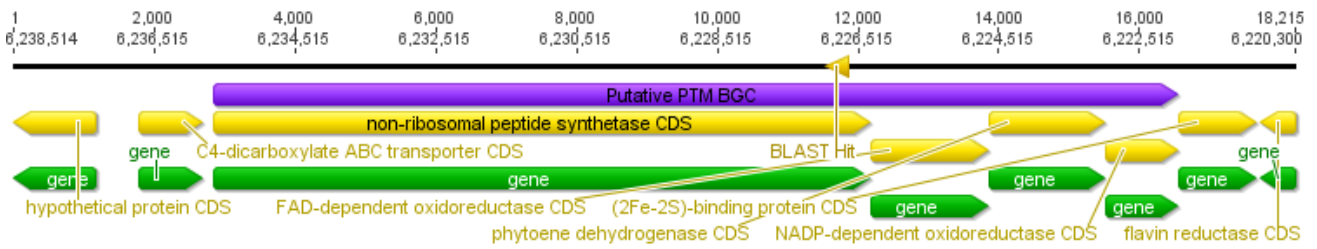- *Streptomyces fulvissimus* DSM 40593
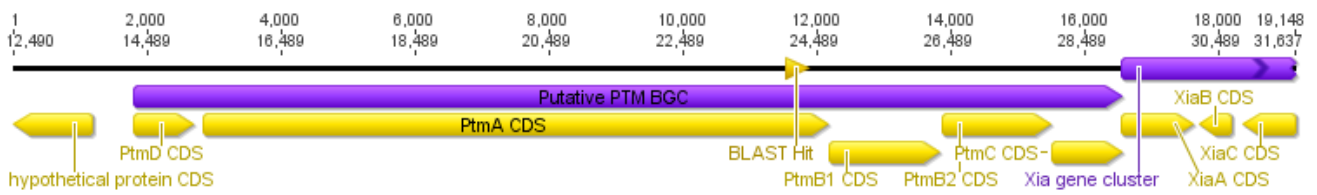- *Streptomyces globisporus* C-1027
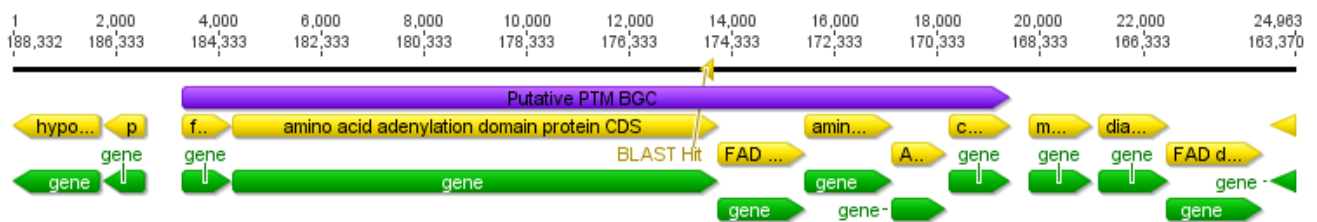- ***Streptomyces griseus subsp. griseus* NBRC 13350 (DSM 40236)**
- *Streptomyces pactum* strain KLBMP 5084
- *Streptomyces pactum* strain SCSIO 02999
- *Streptomyces pratensis* ATCC 33331
- ***Streptomyces roseosporus* NRRL 15998**
- *Streptomyces* sp. CFMR 7 strain CFMR-7
- *Streptomyces* sp. FR-008
- *Streptomyces* sp. Mg1
- *Streptomyces* sp. PAMC26508
- *Streptomyces* sp. SirexAA-E
- *Streptomyces* sp. TLI_053
- *Streptomyces* sp. Tü6075
- ***Streptomyces* sp. Tü6329**
- ***Streptomyces* sp. Tü6314**
- *Streptomyces* sp. ZJ306
- *Streptomyces vietnamensis* strain GIM4.0001
- *Streptomyces xiamenensis* strain 318

    Bold: available in the Gulder lab

Actinoalloteichus hymeniacidonis strain HPA177(T) (=DSM 45092(T))
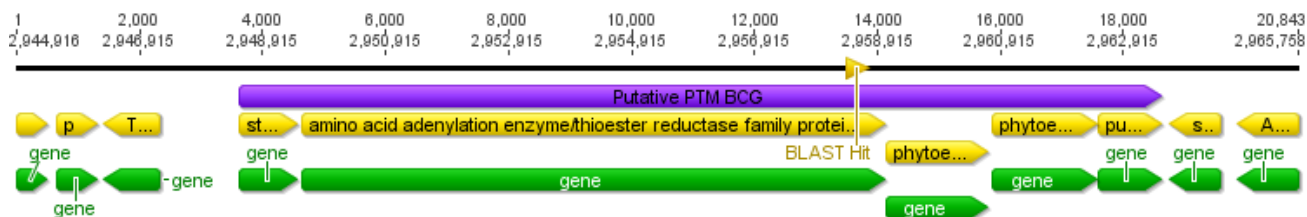
Actinoalloteichus sp. ADI127-7
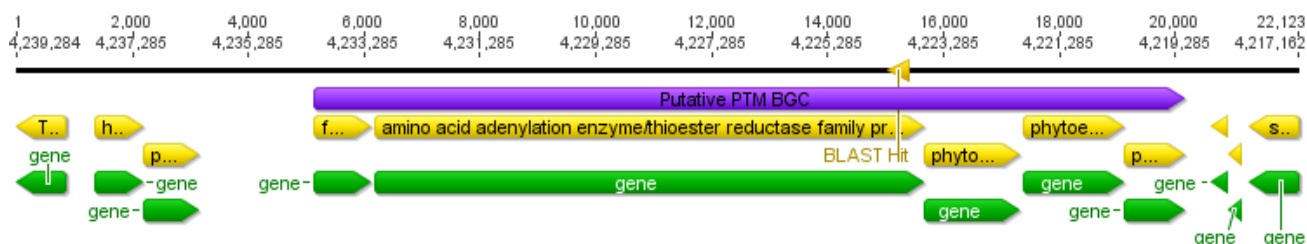
Actinoalloteichus sp. GBA129-24
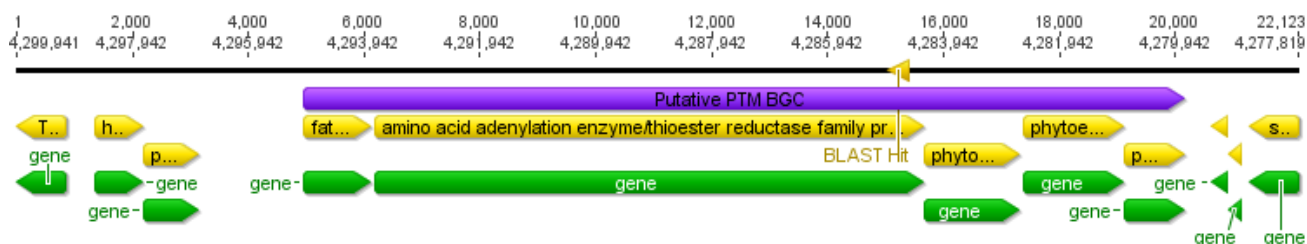
Gynuella sunshinyii YC6258
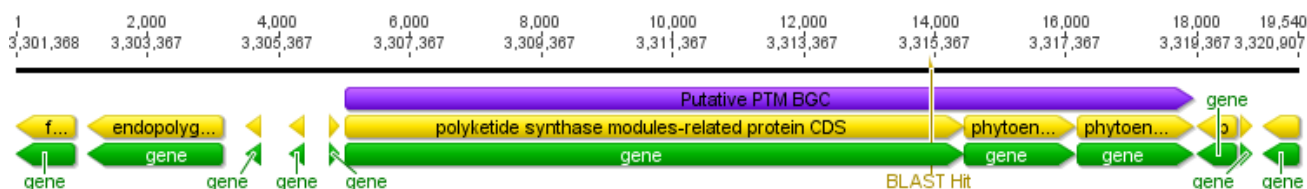
Lysobacter capsici DSM 19286
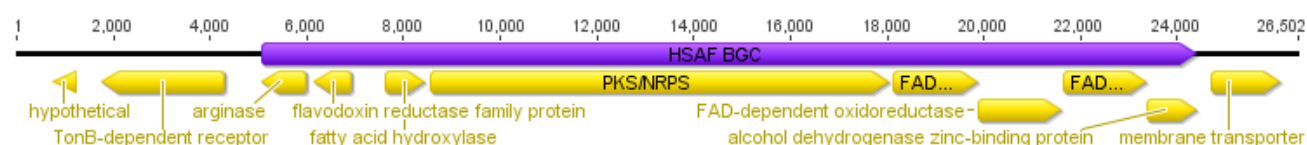
Lysobacter capsici strain 55
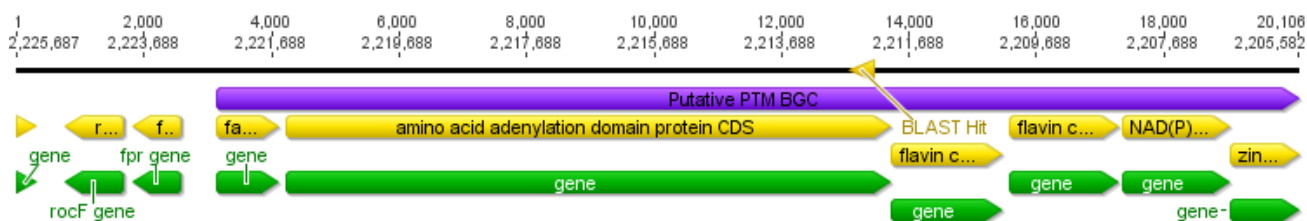
Lysobacter enzymogenes strain C3 genome [HSAF]
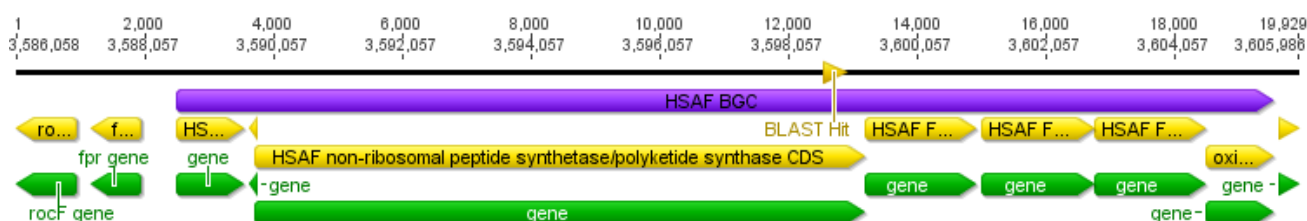
Ref.: F. Yu, K. Zaleta-Rivera, X. Zhu, J. Huffman, J. C. Millet, S. D. Harris, G. Yuen, X. C. Li, L. Du, Antimicrob. *Agents Chemother.* **2007**, 51, 64-72.

Lysobacter enzymogenes strain C3 HSAF biosynthetic gene cluster
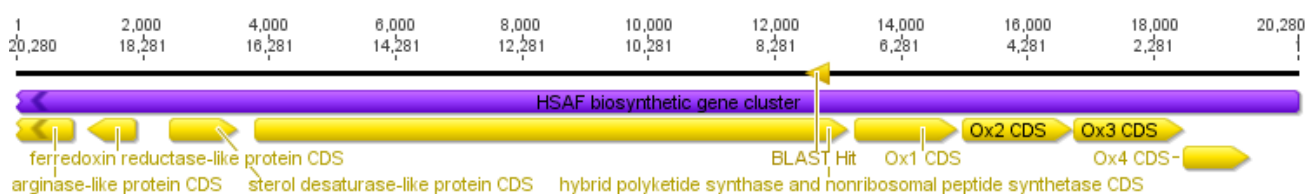
Ref.: L. Lou, G. Qian, Y. Xie, J. Hang, H. Chen, K. Zaleta-Rivera, Y. Li, Y. Shen, P. H. Dussault, F. Liu, L. Du, *J. Am. Chem. Soc.* **2011**, 133, 643-645.

Lysobacter gummosus strain 3.2.11
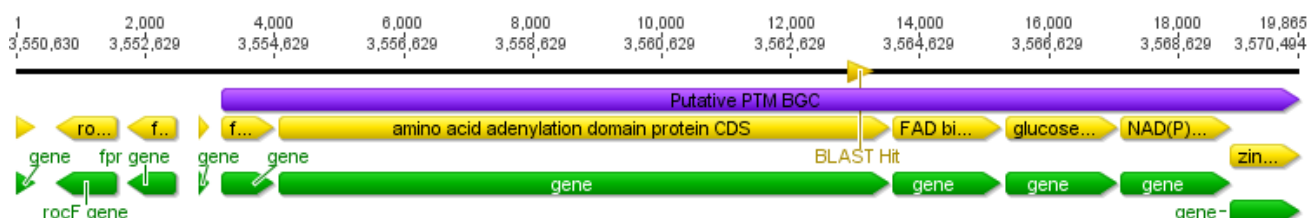
168

Saccharophagus degradans 2-40 (DSM 17024)

Saccharopolyspora spinosa NRRL 18395

Salinispora arenicola CNS-205

Ref.: J. Antosch, PhD thesis, **2015**.

Streptomyces albus J1074 (≠ DSM 40313)

Streptomyces albus strain SM254

Streptomyces clavuligerus strain F613-1

Streptomyces fulvissimus DSM 40593

Streptomyces globisporus C-1027

Streptomyces griseus subsp. griseus NBRC 13350 (DSM 40236)

Streptomyces pactum strain KLBMP 5084

Putative PTM BGC
non-ribosomal peptide synthetase CDS
gene
C4-dicarboxylate ABC transporter CDS
BLAST Hit
gene
hypothetical protein CDS
FAD-dependent oxidoreductase CDS
(2Fe-2S)-binding protein CDS
gene
phytoene dehydrogenase CDS
NADP-dependent oxidoreductase CDS
flavin reductase CDS
gene
gene
gene
gene

Streptomyces pactum strain SCSIO 02999

Putative PTM BGC
PtmA CDS
XiaB CDS
PtmD CDS
BLAST Hit
PtmC CDS
XiaC CDS
hypothetical protein CDS
PtmB1 CDS
PtmB2 CDS
Xia gene cluster
XiaA CDS

Streptomyces pratensis ATCC 33331

Putative PTM BGC
amino acid adenylation domain protein CDS
hypo...
p
f..
amin...
c...
m...
dia...
gene
gene
BLAST Hit
FAD ...
A...
gene
gene
gene
FAD d...
gene
gene
gene
gene
gene
gene

Streptomyces roseosporus NRRL 15998

Iterative PKS/NRPS-System
Putative PTM BGC
Alcohole Dehydrogenase Zind Binding
Sterole Desaturase-Like Protein
FAD-dependent Oxidoreductase
FAD-dependent Oxidoreductase

Streptomyces sp. CFMR 7 strain CFMR-7

Putative PTM BGC
cys...
C...
polyketide synthase CDS
phytoe...
cyt...
h.
gene
gene
BLAST Hit
FAD-de...
alc...
gene
gene
gene
gene
gene-
gene
gene

Streptomyces sp. FR-008
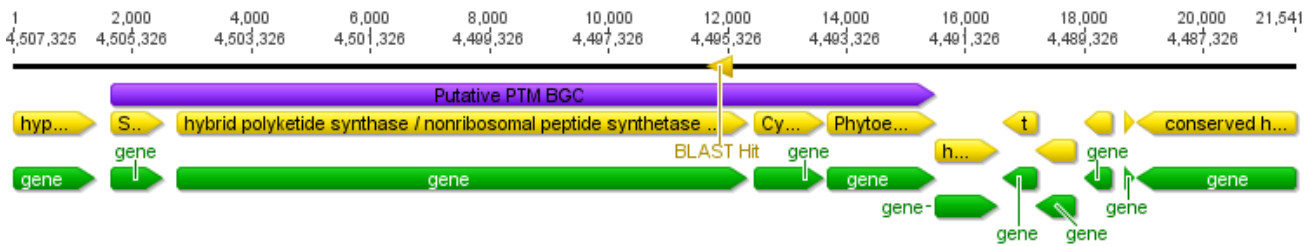

Streptomyces sp. Mg1
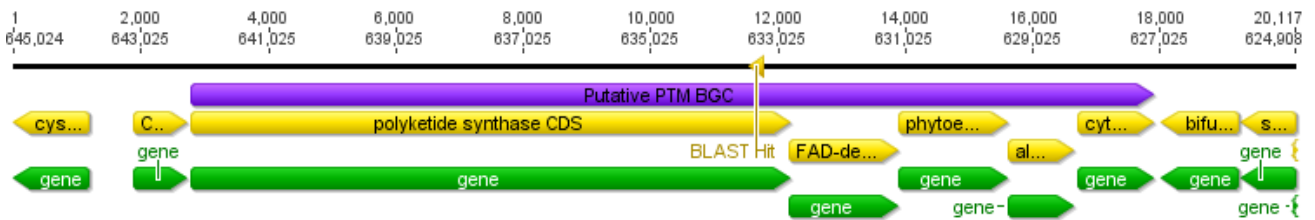

Streptomyces sp. PAMC26508
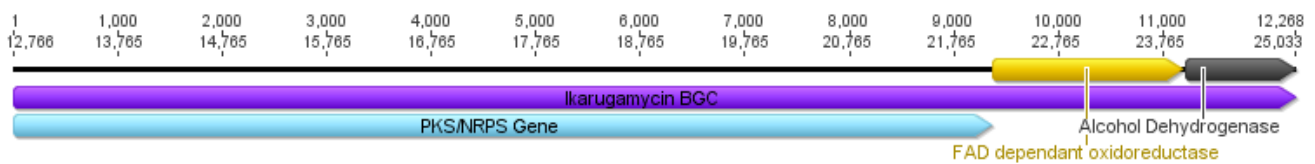

Streptomyces sp. SirexAA-E

Streptomyces sp. TLI_053
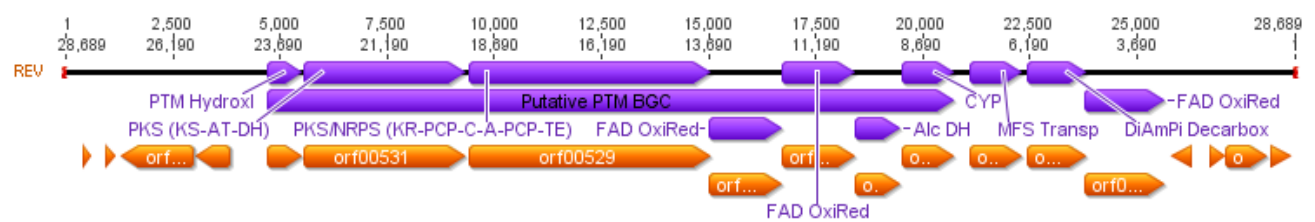

Streptomyces sp. Tue6075
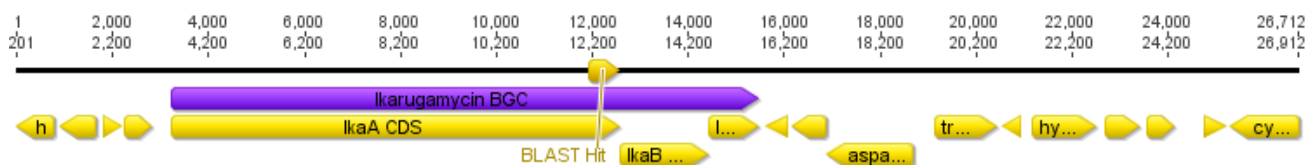

Streptomyces sp. Tue6239 ✺

Ref.: J. Antosch, F. Schaefers, T.A.M. Gulder, *Angew. Chem. Int. Ed. Engl.* **2014**, 53, 3011-3014.
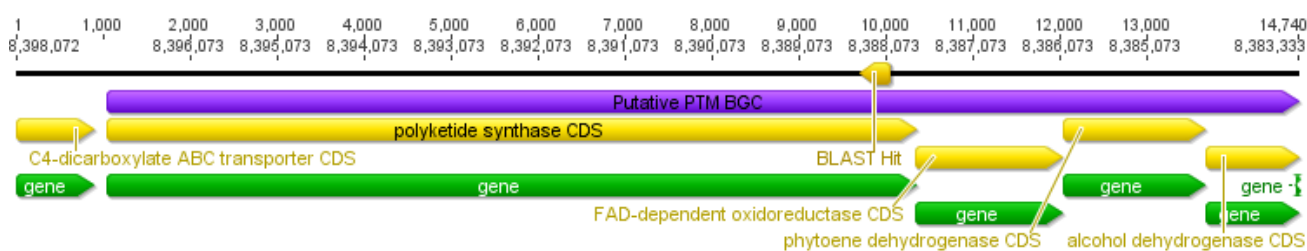
Streptomyces sp. Tue6314 ✺
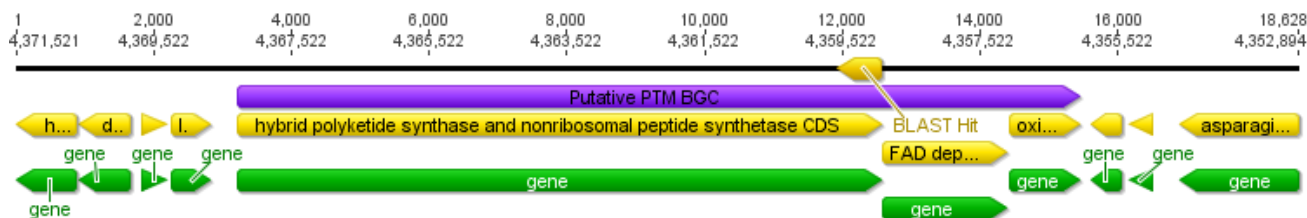
Streptomyces sp. ZJ306

Ref.: G. Zhang, W. Zhang, Q. Zhang, T. Shi, L. Ma, Y. Zhu, S. Li, H. Zhang, Y. L. Zhao, R. Shi, C. Zhang, *Angew. Chem. Int. Ed. Engl.* **2014**, 53, 4840-4844.

Streptomyces vietnamensis strain GIM4.0001
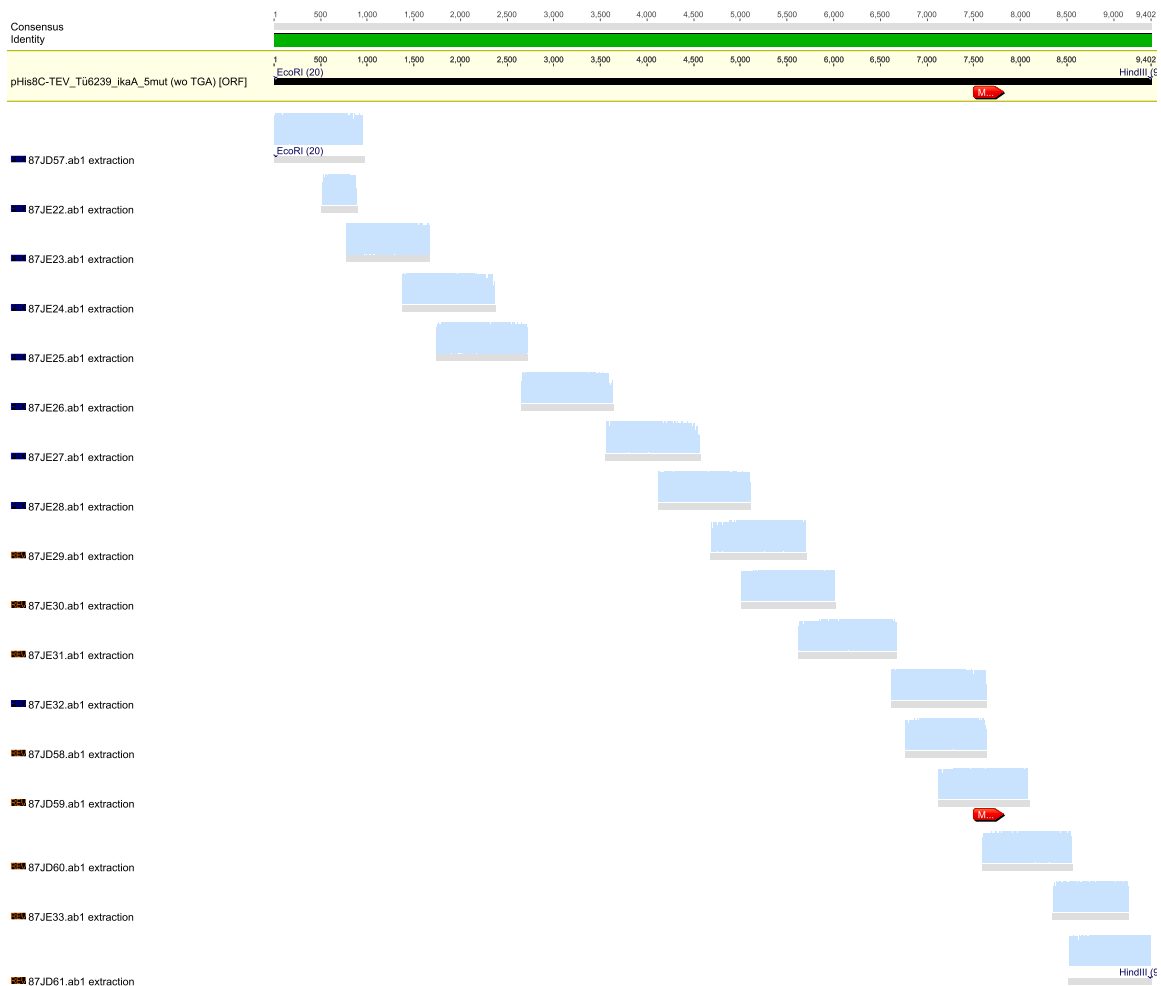

Streptomyces xiamenensis strain 318

Ref.: M. J. Xu, J. H. Wang, X. L. Bu, H. L. Yu, P. Li, H. Y. Ou, Y. He, F. D. Xu, X. Y. Hu, X. M. Zhu, P. Ao, J. Xu, *Sci. Rep.* **2016**, 6, 18977.
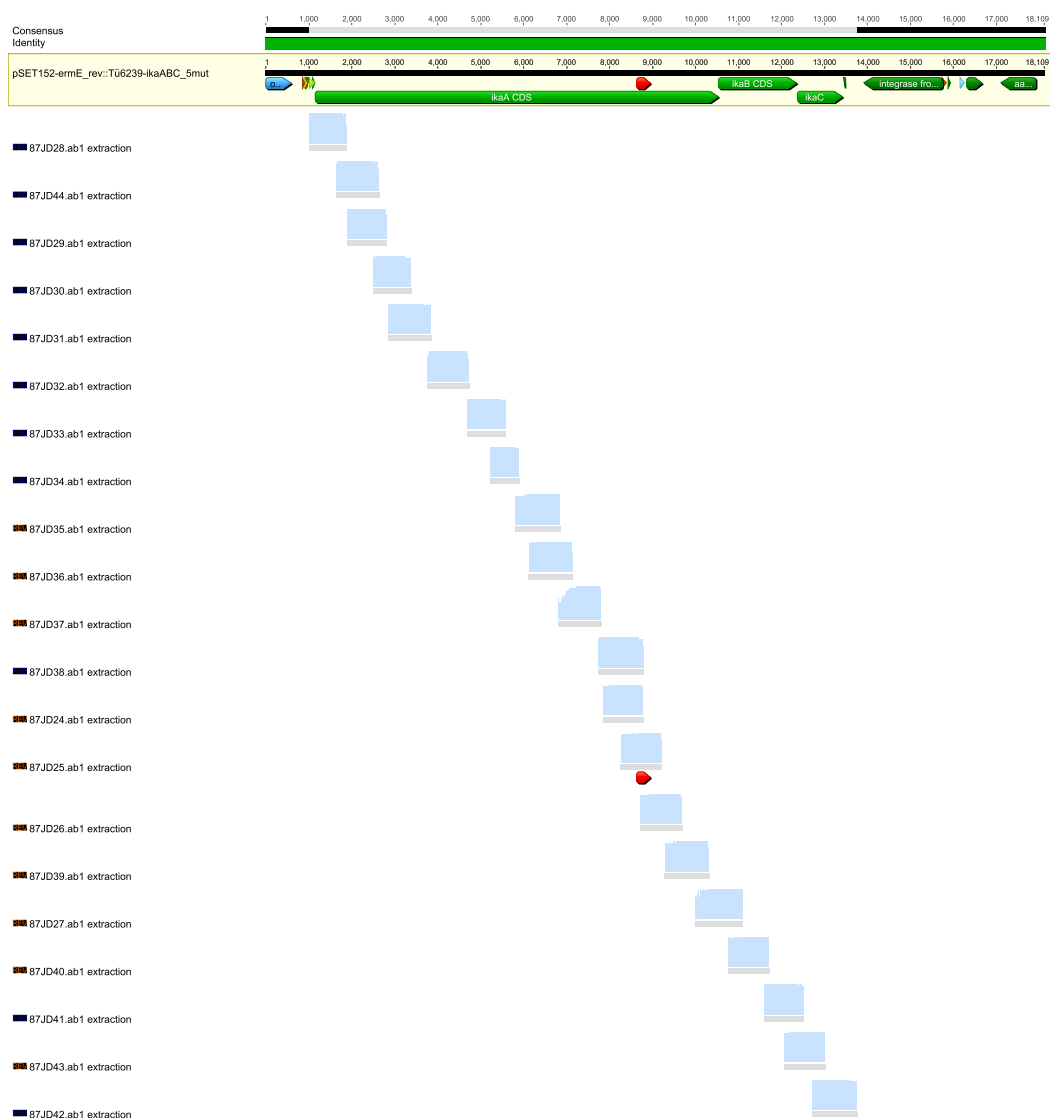
A.7. Sequencing results

A.7.1. pHis8C-TEV::Tü6239-*ikaA_*5mut

Figure A.1. Schematic representation of the sequencing results for the *Eco*RI/*Hin*dIII-flanked insert of pHis8C-TEV::Tü6239-*ikaA_*5mut. The A domain part (321 bp), containing the five altered codons, is depicted in red. Lines 2 to 18 graphically show the perfect agreement of the individual overlapping sequencing runs with the reference sequence of *ikaA_5mut* (top line, yellow).
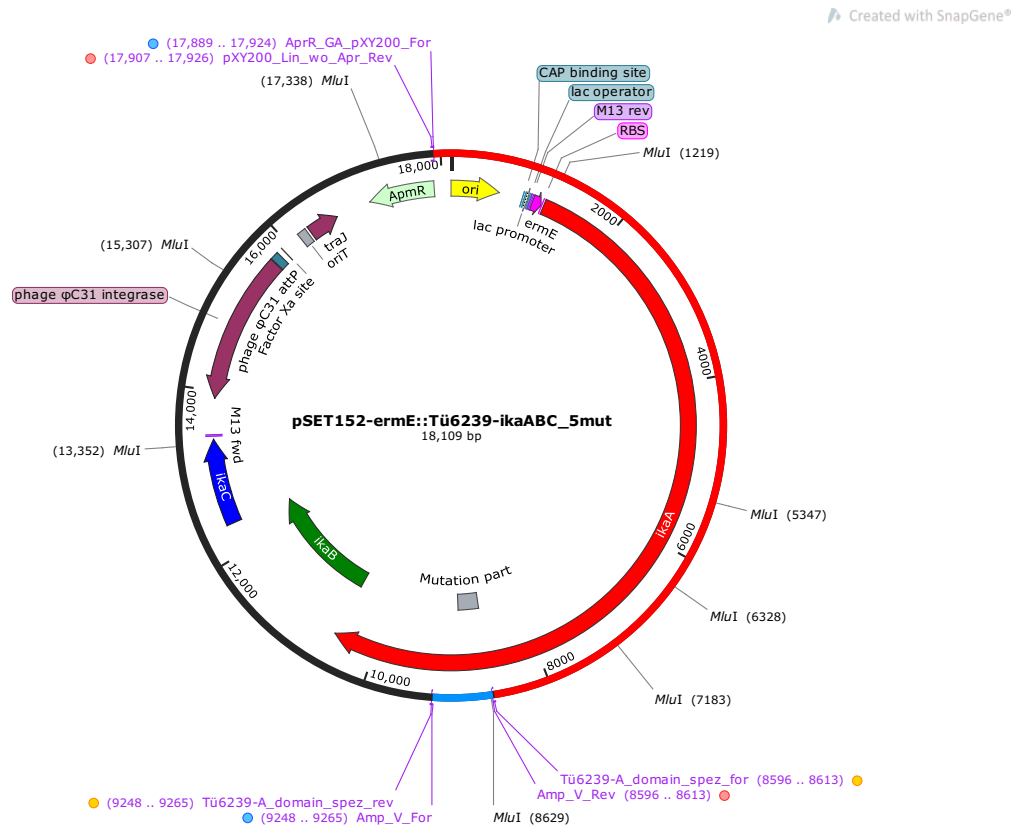
A.7.2. pSET152-*ermE*::Tü6239-*ikaABC*_5mut

Figure A.2. Schematic representation of the sequencing results for *ikaABC* of pSET152-*ermE*::Tü6239-*ikaABC*_5mut. The A domain part (321 bp), containing the five altered codons, is depicted in red. Lines 2 to 22 graphically show the perfect agreement of the individual overlapping sequencing runs with the reference sequence of *ikaABC* in pSET152-*ermE* (top line, yellow).

A.8. Plasmid maps

pSET152-*ermE*::Tü6239-*ikaABC*_5mut

Figure A.3. Plasmid map of pSET152-*ermE*::Tü6239-*ikaABC*_5mut. The A domain part (blue, 670 bp) was introduced in a three fragment HiFi DNA Assembly reaction. The remaining two parts (8816 bp [red] and 8677 bp [black]) were PCR amplified using Q5 polymerase in combination with High GC Enhancer supplement.

A.9. LC-MS data

Supporting information for the heterologous production of erythromycin A (**8**) in *Streptomyces coelicolor*:

- HPLC-MS screening of *S. coelicolor* extracts for **8** – Figure A.4, page 179.

- MS^2 spectra of **8** produced by *S. coelicolor* compared to commercial **8** – Figure A.5, page 180.
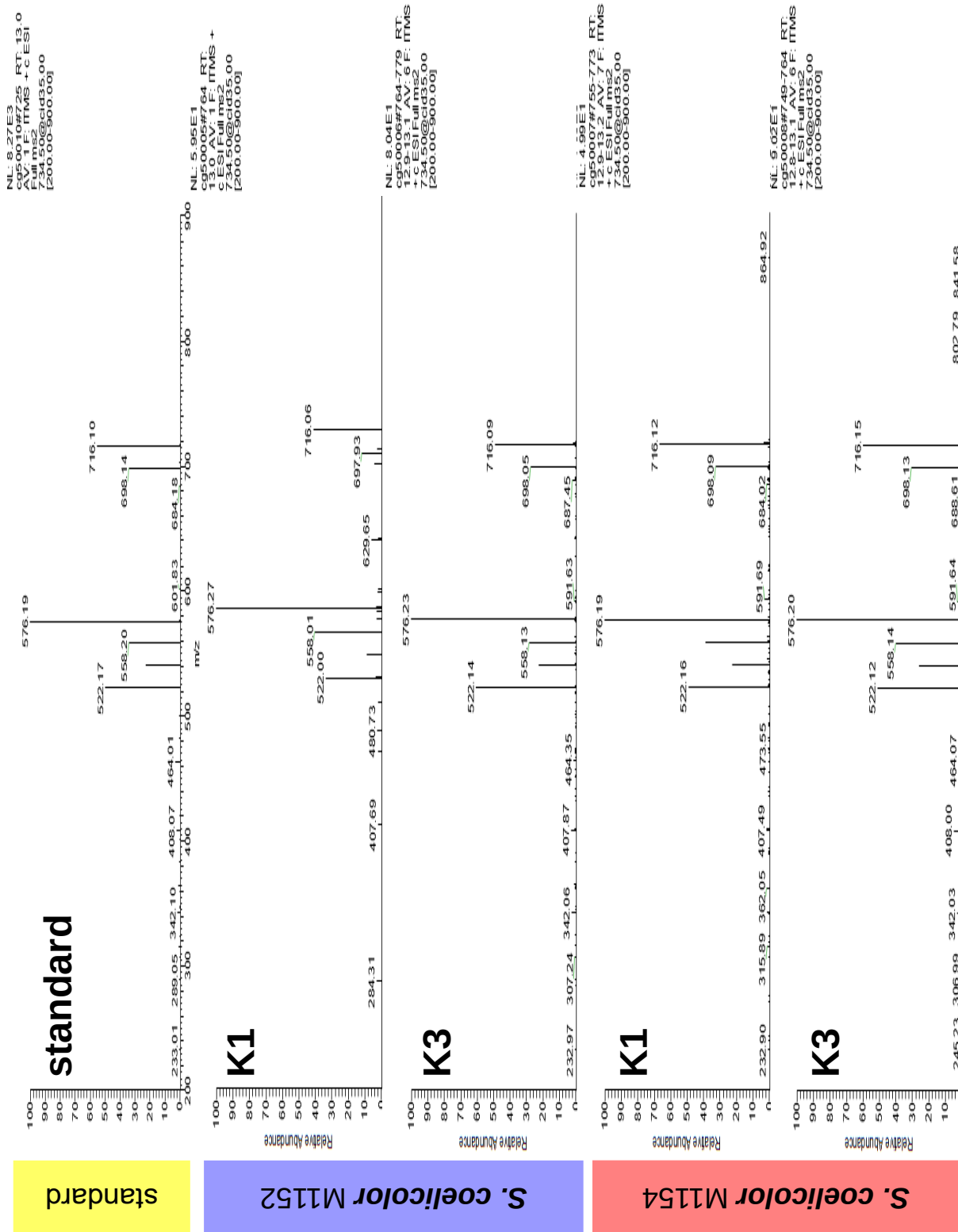
Figure A.4. HPLC-MS screening of *S. coelicolor* extracts for the production of erythromycin A (**8**). The extracted mass trace corresponding to **8** (m/z = 734.46, [M+H]$^+$) is shown in the chromatograms.

Figure A.5. MS^2 spectra of erythromycin A (**8**) produced by *S. coelicolor* compared to commercial **8**.

Approval letter from publisher

Approval letter from publisher of the publication which was highlighted in Chapter 3.2:

- **Greunke, C.**; Glöckle, A.; Antosch, J.; Gulder, T.A.M. Biocatalytic Total Synthesis of Ikarugamycin. *Angew. Chem. Int. Ed. Engl.* **2017**, *56*, 4351-4355, DOI: 10.1002/anie.201611063.

Thank you for your order!

Dear Mr. Christian Greunke,

Thank you for placing your order through Copyright Clearance Center's RightsLink® service.

Order Summary

Licensee:	Mr. Christian Greunke
Order Date:	Nov 7, 2017
Order Number:	4223500649807
Publication:	Angewandte Chemie International Edition
Title:	Biocatalytic Total Synthesis of Ikarugamycin
Type of Use:	Dissertation/Thesis
Order Total:	0.00 EUR

View or print complete details of your order and the publisher's terms and conditions.

Sincerely,

Copyright Clearance Center

How was your experience? Fill out this survey to let us know.

Erklärung

Hiermit versichere ich, dass ich, Christian Greunke, die vorliegende Arbeit selbstständig verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel benutzt habe, dass alle Stellen der Arbeit, die wörtlich oder sinngemäß aus anderen Quellen übernommen wurden, als solche kenntlich gemacht und dass die Arbeit in gleicher oder ähnlicher Form noch keiner Prüfungsbehörde vorgelegt wurde.
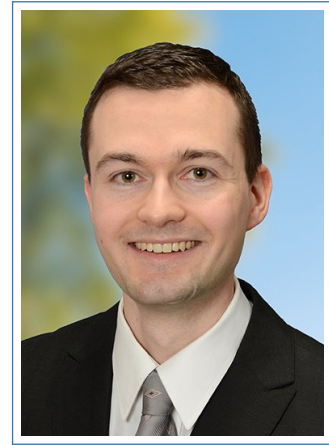
Ort, Datum   Unterschrift

Christian
Greunke

Curriculum Vitae

Münchener Str. 24a
85748 Garching b. München
Germany
+49(0)157 5743438
✉ christian.greunke@gmx.de
orcid.org/0000-0001-7787-4281

PRIVATE DETAILS

Birthday/-place	March 2nd, 1988 in Troisdorf, North Rhine-Westphalia, Germany.
Citizenship	German

CIVILIAN SERVICE

07/2007 – 03/2008	**Nursing service**, *St. Johannes hospital*, Sieglar, Germany.

EDUCATION

04/2014 – (11/2017)	**PhD thesis in Chemistry, Dr. rer. nat.**, *Technical University of Munich*, Garching b. München, Germany.
Focus	Biochemistry, protein science, molecular biology, microbiology.
Dissertation	The (Bio-)Synthesis of Polycyclic Tetramate Macrolactams (PoTeMs).
04/2014 – 03/2017	**PhD scholarship** by the Friedrich Naumann Foundation for Freedom.
10/2011 – 09/2013	**Master of Science, Life and Medical Sciences**, *University of Bonn*, Bonn, Germany, *grade: 1.4*.
Focus	Chemical biology, biochemistry, life sciences.
Master's thesis	Chlorin e6-modified aptamers for tumour-targeted phototherapy.
10/2008 – 09/2011	**Bachelor of Science, Molecular Biomedicine**, *University of Bonn*, Bonn, Germany, *grade: 2.0*.
Bachelor's thesis	Ein Beitrag zur Charakterisierung von wirksamkeitsmitbestimmenden Inhaltsstoffen aus Efeublättertrockenextrakt.
08/1998 – 06/2007	**Gymnasium**, *Heinrich-Böll-Gymnasium*, Sieglar, Germany, *Abitur grade: 1.4*.

PUBLICATIONS
(*: equal contribution/corresponding authors)

2017 **C. Greunke**, A. Glöckle, J. Antosch, <u>T.A.M. Gulder</u>, Biocatalytic Total Synthesis of Ikarugamycin. *Angew. Chem. Int. Ed. Engl.* **2017**, *56*, 4351-4355, DOI: 10.1002/anie.201611063.

2015 **C. Greunke***, J. Antosch*, <u>T.A.M. Gulder</u>, Promiscuous hydroxylases for the functionalization of polycyclic tetramate macrolactams – conversion of ikarugamycin to butremycin. *Chem. Comm.* **2015**, *51*, 5334-5336, DOI: 10.1039/c5cc00843c.

C. Greunke, A. Hage-Hülsmann, T. Sorkalla, N. Keksel, F. Häberlein, <u>H. Häberlein</u>, A systematic study on the influence of the main ingredients of an ivy leaves dry extract on the β_2-adrenergic responsiveness of human airway smooth muscle cells. *Pulm. Pharmacol. Ther.* **2015**, *31*, 92-98, DOI: 10.1016/j.pupt.2014.09.0020.

SCIENTIFIC TALKS

2017 **53rd Natural Product Meeting**, *Chemistry, Biology and Ecology*, Garching b. München, Germany.
"Direct Pathway Cloning of Natural Product Biosynthetic Gene Clusters"

2015 **49th Natural Product Meeting**, *Chemistry, Biology and Ecology*, Jena, Germany.
"Understanding and Applying PTM Biosynthetic Logic"

POSTERS

2015 **GDCh Scientific Forum Chemistry**, Dresden, Germany.

J. Antosch*, **C. Greunke***, T.A.M. Gulder, Heterologous Expression and Functionalization of Polycyclic Tetramate Macrolactams.

AWARDS

2015 Poster prize in the section "Chemical Biology" by the German Chemical Society (GDCh).

LANGUAGES

English Fluent, C2
French   Basic, A2

Garching b. München, 2017-11-03