Technische Universität München

Fakultät für Chemie

Unlocking the Biosynthetic Potential of *Streptomyces* sp. Tü6314

Zhengyi Qian

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.

Vorsitzender: Prof. Dr. Matthias Feige

Prüfer der Dissertation: 1. Prof. Dr. Tobias A.M. Gulder

2. TUM Junior Fellow Dr. Stephan Hacker

Die Dissertation wurde am 26.11.2019 bei der Technischen Universität München eingereicht und durch die Fakultät für Chemie am 16.12.2019 angenommen.

Parts of this thesis have been published in:

- **Z. Qian,** J. Antosch,* J. Wiese, J.F. Imhoff, H.-P. Fiedler, A. Pöthig, T.A.M. Gulder. Structures and biological activities of cycloheptamycins A and B. *Org. Biomol. Chem.* **2019**, *17*, 6595-6600. *equal contribution

- **Z. Qian**, T. Bruhn, P.M. D'Agostino, A. Herrmann, M. Haslbeck, N. Antal, H.-P. Fiedler, R. Brack-Werner, T.A.M. Gulder. Discovery of the Streptoketides by Direct Cloning and Rapid Heterologous Expression of a Cryptic PKS II Gene Cluster from *Streptomyces* sp. Tü6314. *J. Org. Chem.* **2019**, DOI:10.1021/acs.joc.9b02741.

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

- **Z. Qian**, J. Antosch, P.M. D'Agostino, T. Liu, M. Fottner, R. Zhu, A. Pöthig, T.A.M. Gulder, et al.. Functional characterization of the biosynthesis of the antibiotic cycloheptamycins. Manuscript in preparation.

II

IV

Contents

Abstract

Natural products have played important roles in the history of drug discovery. Many natural products and/or their derivatives are used clinically, for example as antibacterial, antifungal, anticancer, antiparasitic and immunosuppressive agents. In microbes, these natural products are usually encoded by genes clustered on the genome, termed biosynthetic gene clusters (BGCs). Among these BGCs, polyketide synthase (PKS) and non-ribosomal peptide synthetase (NRPS) gene clusters are of great importance and are responsible for the production of polyketides and non-ribosomal peptides, respectively. *Streptomyces* is the largest genus of Actinobacteria and it is characterized by a complex secondary metabolism, which produced numerous clinically important drugs, especially antibiotics. Recent advances in genome sequencing and bioinformatic analysis have revealed that the genomes of *Streptomyces* contain substantially more potential BGCs than compounds that have been identified, leaving *Streptomyces* as underexplored reservoir for natural products discovery.

In the first study within this thesis, we isolated cycloheptamycins A and B from the culture broth of *Streptomyces* sp. Tü6314, which is a terrestrial strain isolated from a soil sample from Egerpatak, Ardeal, Romania in 2001. Cycloheptamycin A was first reported in the 1970s as an antibiotic and its structure was examined by mass spectrometry at that time. In our study, we thoroughly validated its structure based on NMR and mass spectrometric analysis, and for the first time elucidated its stereostructure by combining peptide hydrolysis and amino acid analysis with X-ray crystal structure determination. The structure of cycloheptamycin B was assigned using comparative MS/MS experiments and NMR analysis. Our study showed that the cycloheptamycins have selective antibiotic activity against *Propionibacterium acnes* and have no cytotoxicity, highlighting their potential as selectively anti-infective drugs or drug leads.

In the second study, we sequenced the genome of *Streptomyces* sp. Tü6314 and identified a cryptic type II PKS gene cluster (the *skt* cluster) by bioinformatic analysis. This cluster was directly cloned using the *Streptomyces* site-specific integration vector pSET152 by the linear plus linear homologous recombination-mediated recombineering (LLHR), followed by its heterologous expression in *Streptomyces coelicolor* M1152 and M1154. This allowed us to isolate six polyketides from the fermentation broth of the heterologous host strain, of which three are known and three are new compounds, with the latter named streptoketides. Our study also showed that four of the six compounds have anti-HIV activities.

In the third study, we identified the BGC encoding the cycloheptamycins by bioinformatic analysis from the genome of *Streptomyces* sp. Tü6314. This BGC was confirmed by gene deletion, direct cloning and heterologous expression in *Streptomyces coelicolor* M1152 and M1154. The cluster was then engineered by gene disruption and two more cycloheptamycin analogs

(cycloheptamycin C and D) were isolated from the engineered strains. Besides, *in vivo* gene disruption and *in vitro* biochemistry showed that an isopropylmalate synthase homolog is involved in the biosynthesis of the norvaline building block, highlighting the evolvement of a specific enzyme for a specific reaction in nature. In addition, we also observed an unusual transformation of cycloheptamycin into a diketopiperazine containing linear depsipeptide during the fermentation or at high pH conditions.

   In summary, these studies have led to the production of 12 bioactive compounds (of which 8 are new) from either the *Streptomyces* sp. Tü6314 wild type strain or after gene cluster heterologous expression and engineering. These findings show that *Streptomyces* sp. Tü6314 is a rich producer of natural products and reveal new biosynthetic pathways and enzymatic transformations. All these studied lead us to a better understanding of the full biosynthetic potential of underexplored *Streptomyces*.

Zusammenfassung

Naturstoffe spielen in der Geschichte der Arzneimittelentwicklung eine wichtige Rolle. Viele Naturstoffe und / oder deren Derivate werden klinisch eingesetzt, beispielsweise als antibakterielle, antimykotische, antitumorale, antiparasitäre und immunsuppressive Wirkstoffe. In Mikroben werden diese Verbindungen üblicherweise von Genen codiert, die im in Form von Clustern organisiert sind, den sogenannten Biosynthesegenclustern (BGCs). Unter diesen BGCs sind Cluster der Polyketidsynthasen (PKS) und der nicht-ribosomalen Peptidsynthetasen (NRPS) von großer Bedeutung und für die Produktion von Polyketiden bzw. nicht-ribosomalen Peptiden verantwortlich. *Streptomyces* ist die größte Gattung von Actinobakterien, die sich durch einen komplexen Sekundärstoffwechsel auszeichnet, der zahlreiche klinisch wichtige Medikamente, insbesondere Antibiotika, hervorbrachte. Jüngste Fortschritte bei der Genomsequenzierung und der bioinformatischen Analyse haben gezeigt, dass die Genome von Streptomyceten wesentlich mehr potenzielle BGCs enthalten als aus ihnen identifizierte Verbindungen vorliegen, so dass Streptomyceten als unerforschtes Reservoir für die Entdeckung von Naturstoffen bezeichnet werden können.

In der ersten Studie dieser Arbeit haben wir Cycloheptamycin A und B aus der Kulturbrühe von *Streptomyces* sp. Tü6314 isoliert, ein terrestrischer Stamm, der 2001 aus einer Bodenprobe aus Egerpatak, Ardeal, Rumänien, gewonnen wurde. Cycloheptamycin A wurde erstmals in den 1970er Jahren als Antibiotikum beschrieben und seine Struktur zu diesem Zeitpunkt massenspektrometrisch untersucht. In unserer Studie haben wir die Struktur von Cycloheptamycin Aanhand von NMR- und massenspektrometrischen Analysen validiert und zum ersten Mal dessen Stereostruktur durch Kombination von Peptidhydrolyse und Aminosäureanalyse mit Röntgenkristallstrukturbestimmung aufgeklärt. Die Struktur von Cycloheptamycin B wurde unter Verwendung von vergleichenden MS/MS-Experimenten und NMR-Analysen zugeordnet. Unsere Studie hat gezeigt, dass die Cycloheptamycine eine selektive antibiotische Aktivität gegen *Propionibacterium acnes* aufweisen und dabei keine Zytotoxizität besitzen. Die hebt ihr Potenzial als mögliche antiinfektive Arzneimittel oder Leitstrukturen hervor.

In der zweiten Studie sequenzierten wir das Genom von *Streptomyces* sp. Tü6314 und identifizierten durch bioinformatische Analyse den BGC, der die Cycloheptamycine codiert. Die Funktion dieses BGCs wurde durch Gendeletion, direkte Klonierung und heterologe Expression in *Streptomyces coelicolor* M1152 und M1154 bestätigt. Das Cluster wurde dann durch Gendeletionen gezielt verändert und so zwei weitere Cycloheptamycin-Analoga (Cycloheptamycin C und D) aus den konstruierten Mutanten isoliert. In-vivo-Gendeletion und in vitro biochemische Analysen zeigten, dass ein Isopropylmalatsynthase-Homolog an der Biosynthese des Norvalin-Bausteins beteiligt ist, was die Entwicklung eines bestimmten Enzyms für eine bestimmte Reaktion

in der Natur zeigt. Darüber hinaus beobachteten wir während der Fermentation oder bei Bedingungen mit hohem pH auch eine ungewöhnliche Umwandlung von Cycloheptamycin in ein lineares Diketopiperazin, das ein lineares Depsipeptid enthält.

In der dritten Studie identifizierten wir einen kryptischen Typ-II-PKS-Gencluster (*skt*) aus *Streptomyces* sp. Tü6314 durch bioinformatische Analyse. Dieses Cluster wurde direkt unter Verwendung des ortsspezifischen *Streptomyces*-Integrationsvektors pSET152 durch linear plus linear homologous recombination-mediated recombineering (LLHR) kloniert, gefolgt von seiner heterologen Expression in *S. coelicolor* M1152 und M1154. Dies ermöglichte es uns, sechs Polyketide aus der Fermentationsbrühe des heterologen Wirtsstamms zu isolieren, von denen drei bekannte und drei neue Verbindungen sind. Letztere bezeichneten wir als Streptoketide. Unsere Studie zeigte auch, dass vier der sechs Verbindungen Anti-HIV-Aktivitäten aufweisen.

Zusammenfassend haben unsere Studien zur Produktion von 12 bioaktiven Verbindungen (von denen 8 neu sind) aus dem Wildtyp-Stamm *Streptomyces* sp. Tü6314 oder nach heterologer Gencluster-Expression und Veränaderung. Diese Befunde zeigen, dass *Streptomyces* sp. Tü6314 ein reichhaltiger Produzent von Naturstoffen ist und enthüllt neue Biosynthesewege und enzymatische Transformationen. Alle diese Studien tragen dazu bei ein besseres Verständnis des gesamten Biosynthesepotenzials von Streptomyceten zu generieren.

1. Introduction

1.1 Microbial Natural Products and Drug Discovery

Natural products are small molecules produced from primary or secondary metabolism by living organisms such as plants, animals or microorganisms.[1] They often possess complex structures and are characterized by a huge chemical diversity. Natural products have played highly significant roles in the drug discovery and development process over the last several decades.[2] For example, the 2015 Nobel Prize in Physiology or Medicine was awarded to Youyou Tu for the discovery of the plant natural product artemisinin (**1**), and to William C. Campbell and Satoshi Omura for the discovery of the microbial natural product avermectins such as **2** (Figure 1). Artemisinin (**1**) is used as an efficient drug against malaria. A derivative of the avermectins, ivermectin, has radically lowered the incidence of onchocerciasis (river blindness) and lymphatic filariasis (elephantiasis).[3]
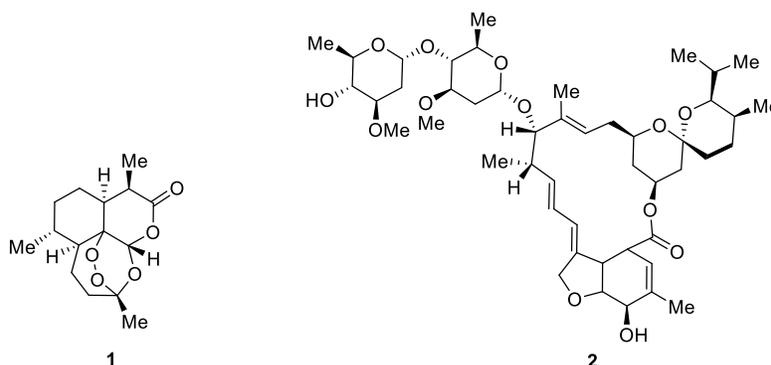

Figure 1. Structures of the famous molecules of the 2015 Nobel Prize in Physiology or Medicine. Artemisinin (**1**) was isolated from *Artemisia annua* L. and avermectins, for example **2**, were discovered from *Streptomyces avermitilis*.

From all the natural product producers, microbes are noticeable for their ability to biosynthesize natural products with a diverse range of useful functions, e.g., antibiotics, anticancer agents, insecticides and immunosuppressants (Figure 2). One of the most important drugs ever discovered from microbial natural products is penicillin (**3**) isolated from *Penicillium* sp.[4] The discovery of **3** by Alexander Fleming in 1929 and its application starting in the 1940s opened the door for humans to utilize microbial natural products as antibiotics to fight against pathogenic microorganisms.[4-5] The use of penicillin (**3**) as an antibiotic ultimately changed the course of human civilization, saving hundreds of thousands of lives from infections that would have been fatal previously.[5]

In the 1940s, the work of Selman Waksman initiated the systematic exploration of microbial sources for novel bioactive natural products, with the discovery of streptomycin (**4**) in 1943 as the culmination.[6] This antibiotic was used as the first curative therapy for tuberculosis. The discovery of clinically important antibiotics from microbes greatly stimulated drug discovery, particularly of

antibiotics. The era from the 1940s to 1960s is regarded as the 'Golden age' of antibiotic discovery.[7] Many novel antibiotics from microbial sources were discovered during this period, including tetracycline (**5**, isolated from *Streptomyces aureofaciens*, 1945); chloramphenicol (**6**, isolated from *Streptomyces venezulae*, 1947); erythromycin (**7**, isolated from *Streptomyces erythraea*, 1949); vancomycin (**8**, isolated from *Amycolatopsis orientalis*, 1953); and rifamycin (**9**, isolated from *Streptomyces mediterranei*, 1957).[8-9] In addition to these successful achievements in antibiotic discovery, anticancer agents were also developed from microbial natural products during this period, including actinomycins (isolated from *Actinomyces antibioticus*, 1940);[10] mitomycin (**10**, isolated from *Streptomyces caespitosus* or *Streptomyces lavendulae*, 1956);[11] daunorubicin (**11**, isolated from *Streptomyces peucetius*, 1964);[12] and bleomycin (**12**, isolated from *Streptomyces verticillus*, 1966).[13] Beyond that, clinically important immunosuppressive drugs were further developed from microbial natural products in the following decades, including rapamycin (**13**, isolated from *Streptomyces hygroscopicus*, 1975)[14-15] and cyclosporin (**14**, isolated from *Tolypocladium* inflatum *Gams*, 1976)[16] (Figure 2).

Despite these successful stories in the middle of the 20th century, the late 20th century has seen a considerable decline in natural product drug discovery, especially by pharmaceutical companies.[17] One reason for this decline was the advances in both high throughput screening (HTS) and combinatorial synthesis.[3, 17] The high rediscovery rate of known compounds in the natural product field and the low amount of molecules isolated from native producers have further led to the decreased application of small molecules from nature in drug discovery.[18] Meanwhile, the emergence of multi-drug resistant human pathogens become more and more serious, revitalizing the importance of novel natural product discovery.[2, 19-21]

Natural products can be classified into different families according to their chemical structures and biosynthetic origins. The major families of natural products include polyketides (PKs), non-ribosomal peptides (NRPs), ribosomally synthesized and post-translationally modified peptides (RiPPs), isoprenoids, alkaloids, aminoglycosides and nucleosides.[22] Within this thesis, modular biosynthetic systems of the PKS- and NRPS-type were studied, which will thus be introduced in the following chapters.
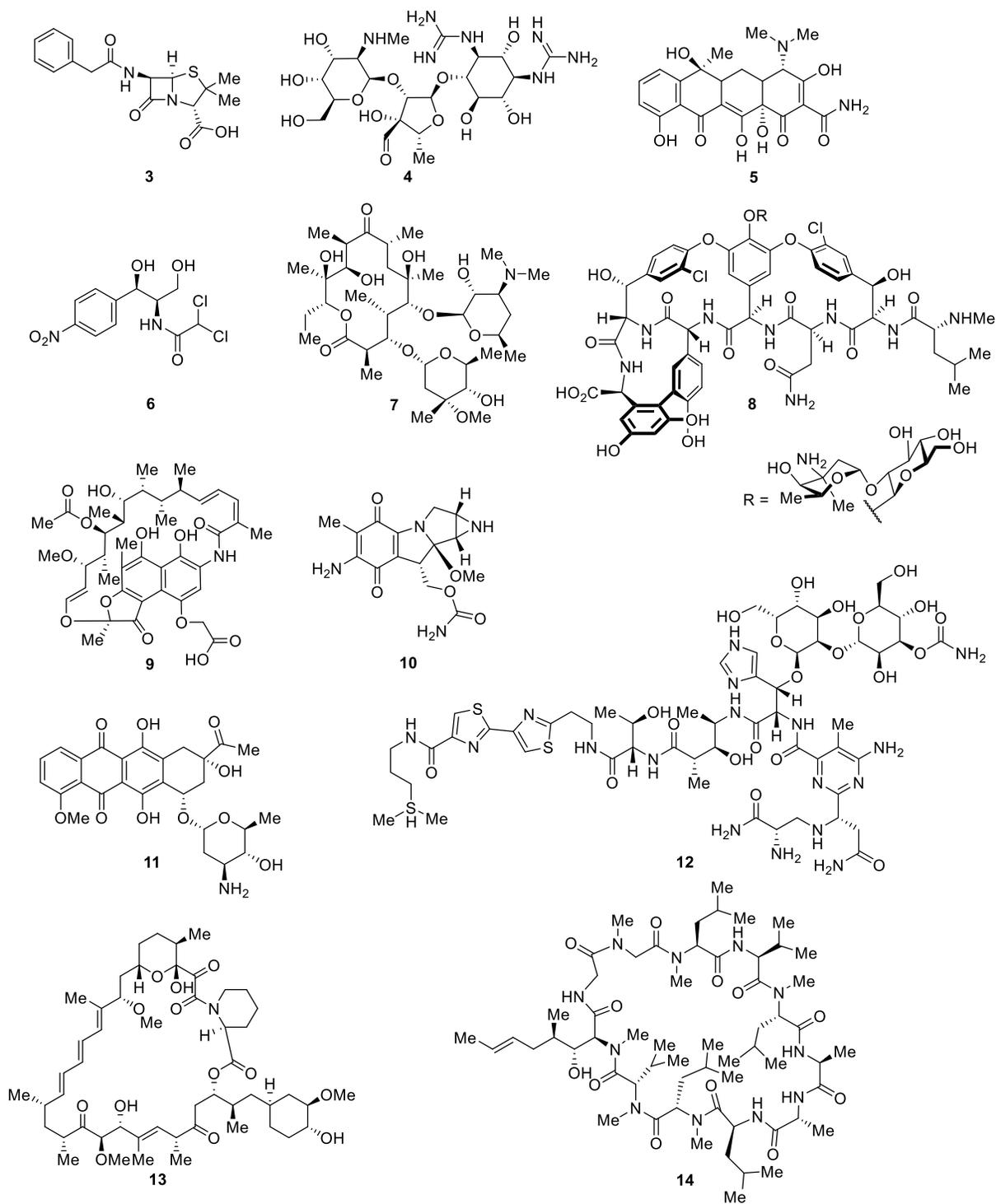
Figure 2. Selected antibiotics, anticancer agents and immunosuppressants from microbial sources isolated during the 'Golden age'. Penicillin G (**3**), streptomycin (**4**), tetracycline (**5**), chloramphenicol (**6**), erythromycin (**7**), vancomycin (**8**), rifamycin B (**9**), mitomycin (**10**), daunorubicin (**11**), bleomycin (**12**), rapamycin (**13**), and cyclosporin (**14**).

3

1.2 Polyketides and Polyketide Synthases

The polyketide natural products are a remarkable class of compounds that play important roles in drug development and discovery.[23] The above-mentioned avermectins (**2**), tetracycline (**5**), erythromycin (**7**), rifamycin (**9**), daunorubicin (**11**) and rapamycin (**13**) are all polyketide type natural products. The polyketide skeletons are biosynthesized by core enzymes, the polyketide synthases (PKSs), by successive condensation of activated short carboxylic acid precursors (e.g., acetyl-CoA, malonyl-CoA, and methylmalonyl-CoA, Figure 3).[24]
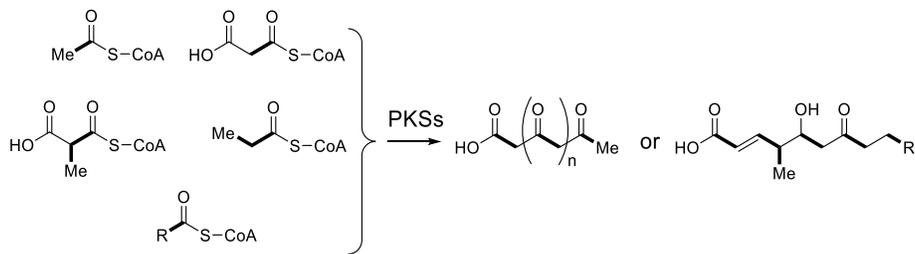
Figure 3. Generic scheme for the biosynthesis of polyketides from different precursors. The short carboxylic acids serve as precursors in activated CoA form. The bold bonds indicate the extension units derived from the building blocks.

PKSs are composed of several catalytic domains, which usually include acyltransferase domains (ATs), ketosynthase domains (KSs), acyl carrier protein domains (ACPs), ketoreductase domains (KRs), dehydratase domains (DHs), and enoyl reductase domain (ERs). These domains can be stand-alone proteins or they can be integrated into a single giant protein (Figure 4A).[25] Detailed domain functions will be discussed in the following chapter. Specially, ACPs are initially expressed as *apo*-ACP with a conserved active-site of serine residue. After translation, the ACPs need to be activated from *apo*-ACP to *holo*-ACP by phosphopantetheinyl transferases (PPTases). This activation leads to attachment of a phosphopantetheine (PPant) residue derived from coenzyme A (CoA) (**15**) onto the conserved serine residue (Figure 4B).[26] After activation, the ACPs are able to accept the CoA-activated substrates as well as the elongated intermediates.

PKSs are typically divided into three different types: type I, II, III PKSs, according to their enzyme architectures and molecular catalytic mechanisms.[24, 27]
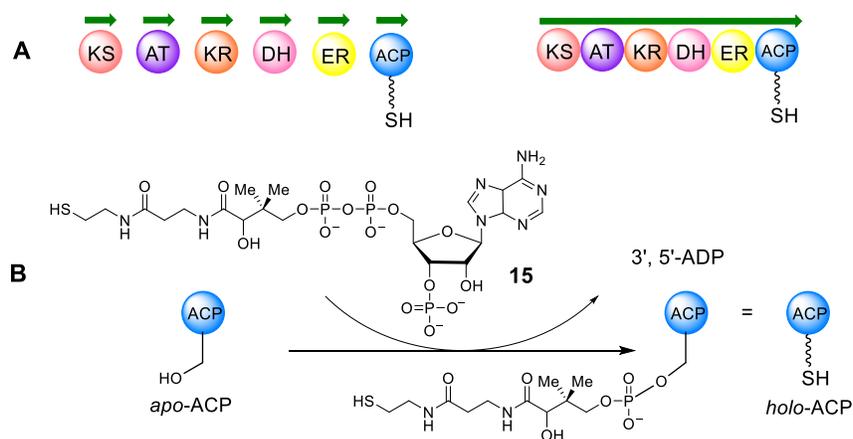
Figure 4. (**A**) Scheme for canonical PKS domains and their organizations. (**B**) Scheme for the activation of PKS ACPs. The proteins are shown by green arrows. Domain abbreviations: ACP, acyl carrier protein; AT, acyltransferase; DH, dehydratase; ER, enoyltransferase; KR, β-ketoreductase; KS, β-acyl ACP synthase; TE, thioesterase.

1.2.1 Type I PKSs

Type I PKSs are multifunctional enzymes that can be divided into individual biosynthetic modules. A canonical module, which is responsible for the catalysis of one cycle of polyketide chain elongation, minimally consists of an AT domain, a KS domain and an ACP domain.[24, 27] The modules can also include KR domains, DH domains, and ER domains within a so-called reductive loop that performs successive oxidative state adjustment by reduction. The AT domain of each module is specific for the recruitment of a CoA-activated short carboxylic acid precursor, transferring it to the downstream ACP. The ACPs incorporate the CoA-activated substrates as well as the elongated intermediates via covalent thioester bounds. The following KS domain takes over the polyketide chain from the previous module and catalyzes the decarboxylative Claisen condensation between the KS-bound polyketide chain and the ACP-bound elongation group, elongating the polyketide chain and leaving the chain bound to the ACP. Additionally, if a KR domain is present, the β-keto group will be reduced to a β-hydroxy group. An additional DH domain will eliminate the hydroxy group, resulting in an α-β-unsaturated alkene. An additional ER domain will reduce the α-β-double-bond to a single-bond. Type I PKS biosynthesis usually is terminated by a thioesterase domain (TE) that cleaves the final product off the biosynthetic enzyme by a hydrolysis or macrocyclization reaction (Figure 5).[25, 28-30] All the modules and domains typically work stepwise, which is reminiscent of industrial assembly line processes, to ultimately form the polyketide products.[31] Each module usually incorporates a single elongation building block, thus the number of modules and the individual domain organizations within the PKS enzymes can be directly translated to the structures of the final products. In this way, the order of modules usually corresponds to the sequence of the building blocks in the polyketide chain. This mechanism is often

referred to as the 'co-linearity rule'.[31] Complex post tailoring reactions can follow PKS product assembly and can be catalyzed by enzymes such as oxidoreductases, methyl or glycosyl transferases, halogenases, and deoxysugar biosynthetic enzymes to generate the final, fully functionalized polyketides. Polyketides derived of type I PKSs are usually macrolides, polyethers or polyenes, such as avermectins (**2**), erythromycin (**7**), and rifamycin (**9**).
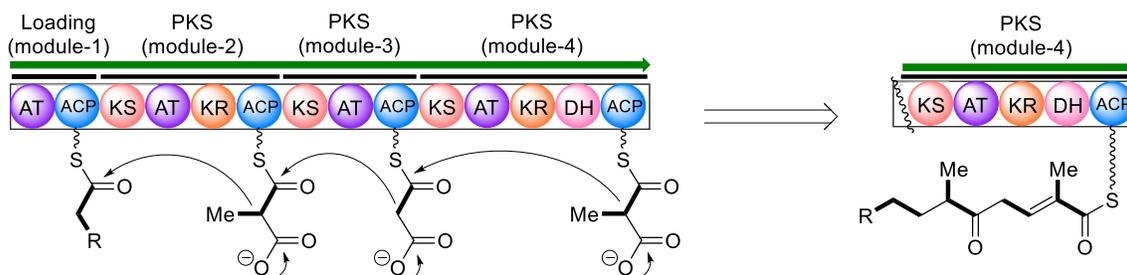

Figure 5. Generic scheme for type I PKS assembly: the loading module primes the PKS with the starter unit, followed by stepwise condensation and modification catalyzed by each module to form the polyketide chain bounded to the ACPs. The individual PKS proteins are indicated by green arrows. Each PKS protein consists of several modules which are delineated by solid black lines.

1.2.1.1 Non-iterative type I PKSs

Modular type I PKSs are among the most well studied PKS systems. Non-iterative modular type I PKSs use one module to incorporate one polyketide unit and the module is used only once in each round of polyketide assembly.[25] The prototypical non-iterative modular type I PKS is represented by the 6-deoxyerythromycin B (**16**) synthases (DEBSs) for the biosynthesis of reduced polyketides (Figure 6).[24, 27] The biosynthesis of **16** starts from loading of a propionyl-CoA unit by the loading module. Then the following module recruits one unit of methylmalonyl-CoA as elongation unit and fuses it to the former chain by decarboxylative Claisen condensation to afford a new polyketide chain. The KR, DH, ER domains in each module will control the reductive state of the polyketide chain. The chain elongation is performed on each module, finally furnishing the full-length polyketide chain bound to the last ACP. The terminal TE domain cleaves off the linear precursor and catalyzes macrolactone formation to afford **16**.[25, 28-29] After several tailoring modifications, the bioactive erythromycin (**7**) is biosynthesized.
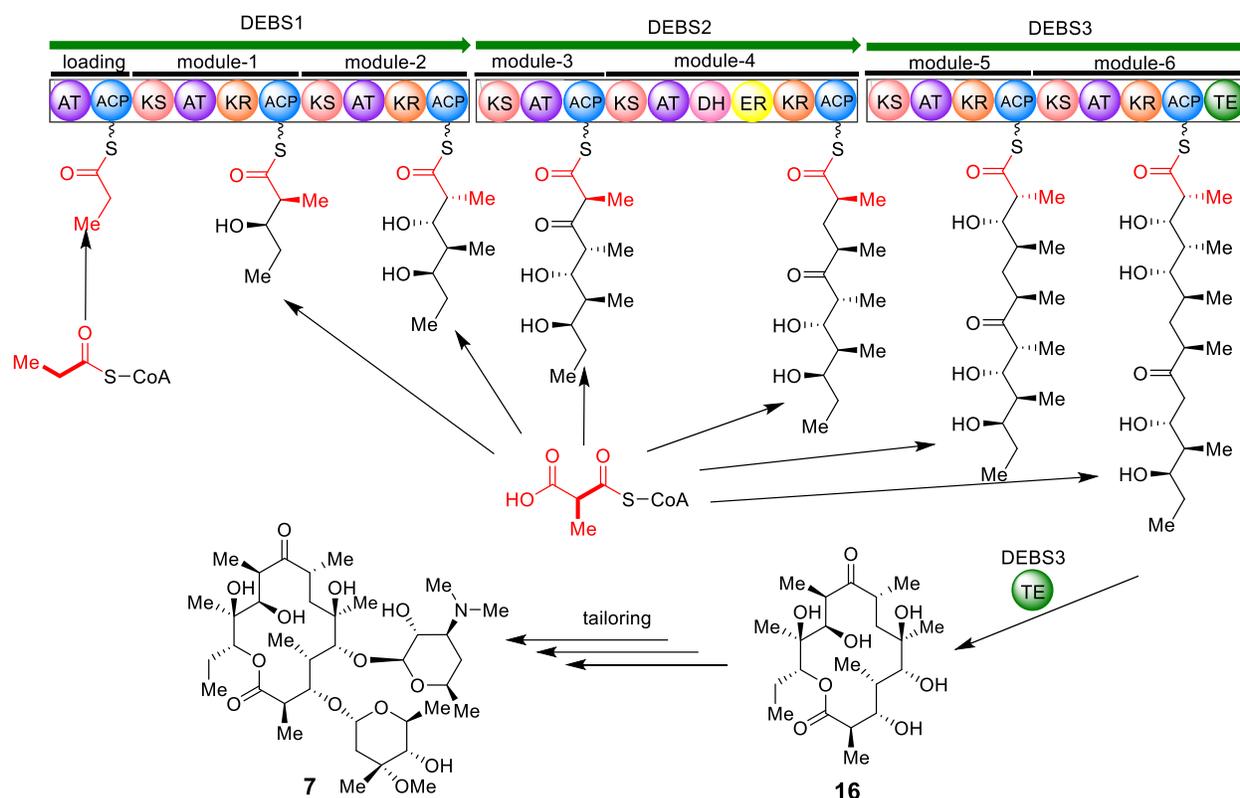
Figure 6. Domain organization of the 6-deoxyerythronolide B (**16**) synthases (DEBSs) and the biosynthesis of erythromycin (**7**). DEBSs work stepwise to grow the polyketide chain attached to the ACP domains of each modules. The TE domain from the last module cleaves off the linear precursor after full modification by modular functions. The individual PKS proteins are indicated by green arrows. Each PKS protein consists of several modules which are delineated by solid black lines.

1.2.1.2 Iterative type I PKSs

Another kind of type I PKSs is the iterative type I PKSs (*i*PKSs), which use modules or individual domains more than once during biosynthesis of the encoded polyketide. *i*PKSs are commonly found in fungi. Fungal *i*PKSs are further classified into highly reducing (HR-), nonreducing (NR-) and partially reducing (PR-) PKSs, according to the degree of reduction within the polyketide chain (Figure 7A-C).[32] HR-PKSs that contain a fully reducing modifying region are able to reduce the primary β-carbonyl condensation product to form a fully saturated carbon–carbon bond using KR, DH, and ER activities, as exemplified by the biosynthesis of the cholesterol lowering agent lovastatin (**17**), in which the *i*PKS LovB and a free-standing ER domain protein LovC work iteratively to build up a fully reduced polyketide chain (Figure 7A).[32] In contrast to HR-PKSs, NR-PKRs lack all reducing/modifying domains and directly use the primary polyketide chain, typically for cyclization reactions resulting in products with aromatic rings, as exemplified by the biosynthesis of the 6-methylsalicylic acid (**18**) (Figure 7B).[32] The PR-PKSs will partially reduce

the polyketide chain during extension, as exemplified by the biosynthesis of the noranthrone (**19**) (Figure 7C).[32] However, the degree of reduction can vary and the factors governing this variability are still largely unknown.[32-33]
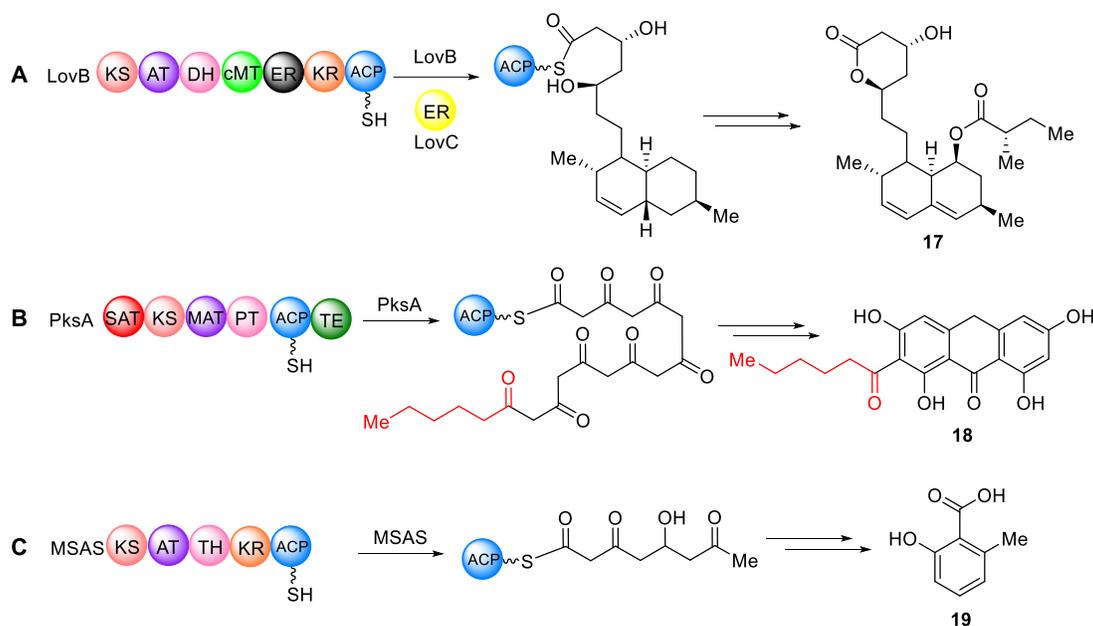

Figure 7. Examples of fungal iterative type I PKSs (*i*PKS) and their products.[32] (**A**) HR-PKSs (highly reducing PKSs), exemplified by the lovastatin (**17**) pathway. (**B**) NR-PKSs (nonreducing PKSs), exemplified by the noranthrone (**19**) pathway. (**C**) PR-PKSs (partially reducing PKSs), exemplified by the 6-methylsalicylic acid (**18**) pathway. CMT, C-methyltransferase. The black colored ER domain in (**A**) is supposed to be inactive. The red colored fragment in **19** is derived from hexaonoic acid, a fatty acid precursor.

Iterative type I PKSs are also found in bacteria. One type of bacterial iterative type I PKSs can produce aromatic polyketides, in a manner very similar to the fungal *i*PKSs (Figure 8A-C). These iterative type I PKSs have the characteristic modular type I PKS domains of KS, AT, DH and ACP. The DH domain was afterwards renamed to TH domain (thioester hydrolase), as it is not involved in dehydration of the β-hydroxyketide intermediate tethered on the acyl carrier protein, but instead catalyzes thioester hydrolysis to release the product from the ACP.[34-35] AviM from the avilamycin biosynthetic pathway was the first reported bacterial iterative type I PKS (Figure 8A).[36] Heterologous expression of *aviM* in *Streptomyces lividans* TK24 and *Streptomyces coelicolor* CH999 led to the production of orsellinic acid (**20**), which is a structural element in avilamycin (**21**).[36] Some bacterial iterative type I PKSs also carry a KR domain, which will reduce the polyketide chain to some degree and thus produce different compounds. It is interesting that even with very similar domain organizations and sequence homology, these PKSs may produce significant structural diversity. For example, the type I PKSs in the biosynthesis of chlorothricin

8

(**23**) via **22** and neocarzinostatin (**25**) via **24** have very similar structures, but their products are totally different (Figure 8B-C).[35, 37-39]
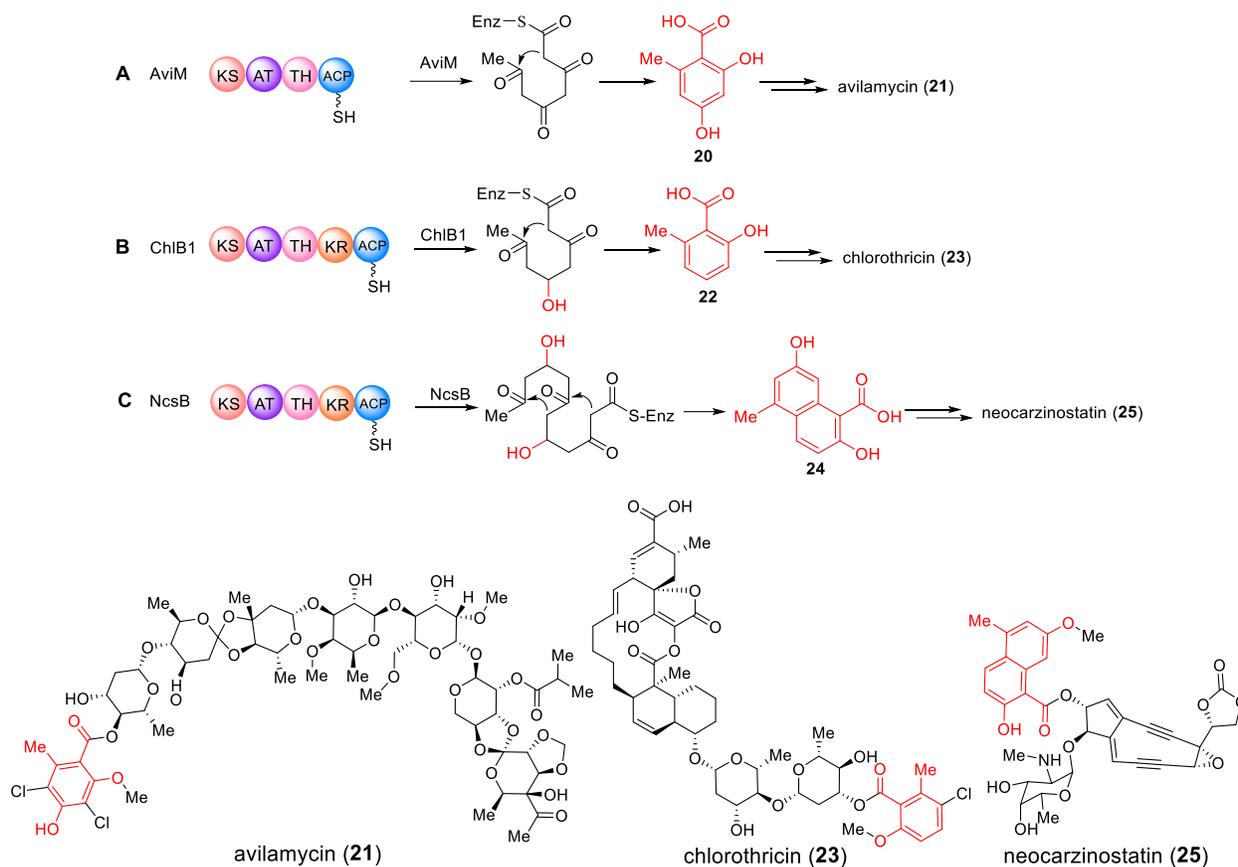

Figure 8. Examples of bacterial iterative type I PKSs that produce aromatic polyketides. (**A**) AviM for orsellinic acid (**20**) biosynthesis in the avilamycin (**21**) pathway.[36] (**B**) ChlB1 for 6-methyl-salicyclic acid (**22**) biosynthesis in the chlorothricin (**23**) pathway.[37] (**C**) NcsB for 2-hydroxy-5-methyl-NPA (NPA = 1-naphthoic acid) (**24**) biosynthesis in the neocarzinostatin (**25**) pathway.[39]

Another type of bacterial *i*PKS is responsible for the biosynthesis of the polycyclic tetramate macrolactams (PoTeMs). Typical PoTeMs share a tetramate-embedding macrocyclic lactam ring that is fused to a subset of diverse carbocyclic ring systems,[40] such as HSAF (heat-stable antifungal factor, **26**)[41] and ikarugamycin (**27**) (Figure 9).[42] Although the PoTeMs family of compounds comprises complex structural diversity, heterologous expression and combinational engineering have shown that their skeletons are biosynthesized by a single-module *i*PKS-NRPS (Non-ribosomal Peptide Synthetase, see Chapter 1.3) hybrid gene (Figure 9).[40, 43-46] In a proposed PoTeMs biosynthetic pathway, the single-module PTM *i*PKS is responsible for the synthesis of the two different polyketide moieties of the PoTeMs skeleton.[40, 43-47] A similar single-module PKS is also found in the biosynthesis of the enediyne type natural products, in which a polyene chain is biosynthesized first, followed by formation of the enediyne core (cf. **25** in Figure 8).[48-49]
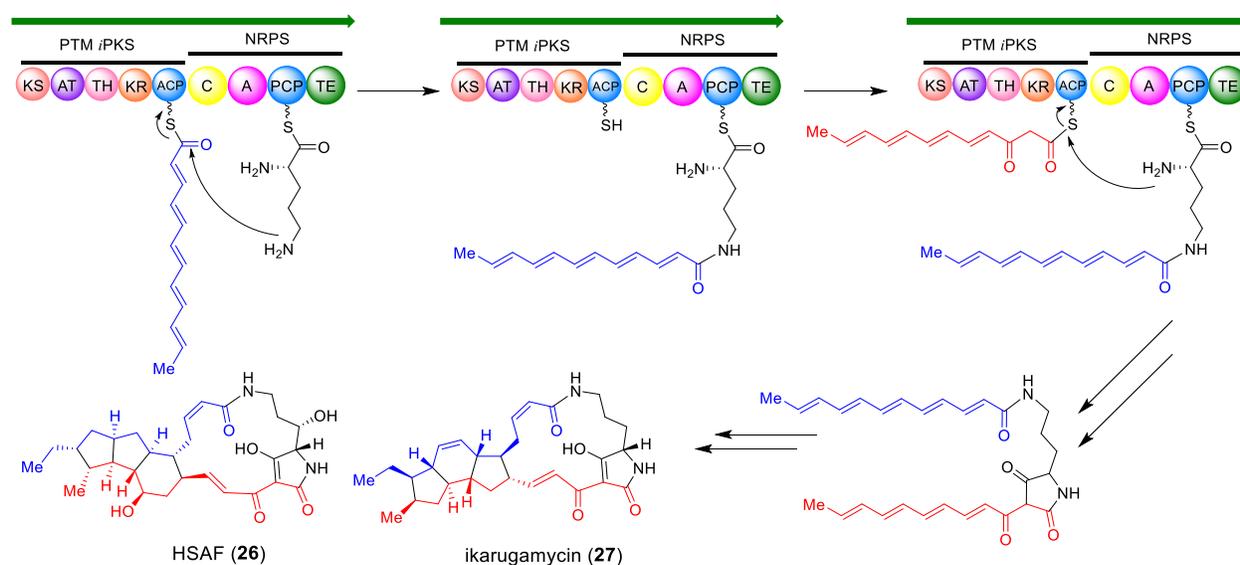
Figure 9. Biosynthesis of the PoTeM family compounds by the *i*PKS-NRPS system. The single-module type I PKS is hybridized with an NRPS module and is supposed to synthesizes two separate hexaketide chains in the biosynthesis of PoTeMs. For NRPS biosynthetic logic, see below Chapter 1.3.

1.2.1.3 *Trans*-AT PKSs

In addition to the above-mentioned type I PKSs, studies have shown that PKS systems have much more diversity in both mechanism and structure.[27, 33, 50] One significant finding was that some modular PKSs completely lack the cognate AT domain, whose missing activity instead is provided *in trans* by one or a few free-standing proteins with AT functions (Figure 10).[27, 33, 50] Thus, this kind of modular PKSs was termed *trans*-AT (or AT-less) PKS, while the typical modular PKS was named *cis*-AT PKS.[25] Another difference is that for *cis*-AT PKS, methyl branches typically stem from incorporation of methylmalonyl-CoA extender units selected by the corresponding AT domain, while for the *trans*-AT PKS, the methyl branch is introduced by an MT domain. There also can be one or more stand-alone AT proteins. Different other domains can also be stand-alone domains, such as the ER domain.[27, 33, 50] Even though the *cis*- and *trans*-AT PKS significantly differ in their structures, their product can in principle be identical (Figure 10).

A recent survey of PKS genes in sequenced bacterial genomes showed that almost 38% of bacterial modular PKSs are *trans*-AT type, suggesting that this kind of PKS is responsible for the biosynthesis of a major natural product class. This kind of PKS is particularly common in chemically less well-studied bacteria, such as unusual taxa, uncultivated bacteria, mutualists, and pathogens. In addition to the AT architecture, *trans*-AT PKSs also possess numerous unique peculiarities when compared to *cis*-PKS. An important characteristic is their great architectural diversity, which includes modules with unusual domain orders or unique domains, such as non-

elongating modules (KS^0) or *O*-methyltransferase (OMT) domains. The KS^0 differs from the normal KS in lacking a conserved histidine that is necessary for decarboxylative condensation. Usually, this domain will modify the polyketide chain jointly with an upstream module.[50] OMT domains introduce *O*-methyl groups to the polyketide backbones. Usually, OMTs occur in bimodules, with the OMT being localized behind the non-elongating module.[50] Other domains with unusual functions can also be found in *trans*-AT PKS, such as dehydrating bimodules, modules inserting oxygen into polyketide chains, modules introducing β,γ-double bonds, pyran synthase (PS) modules, and Michael-branching modules.[50] The first reported extended *trans*-AT PKS system with an attributed metabolite was the pederin PKS, whose function was proposed based on the enzymatic architecture with the hypothesis that transacting ATs perform acylations of entire PKS systems.[51]
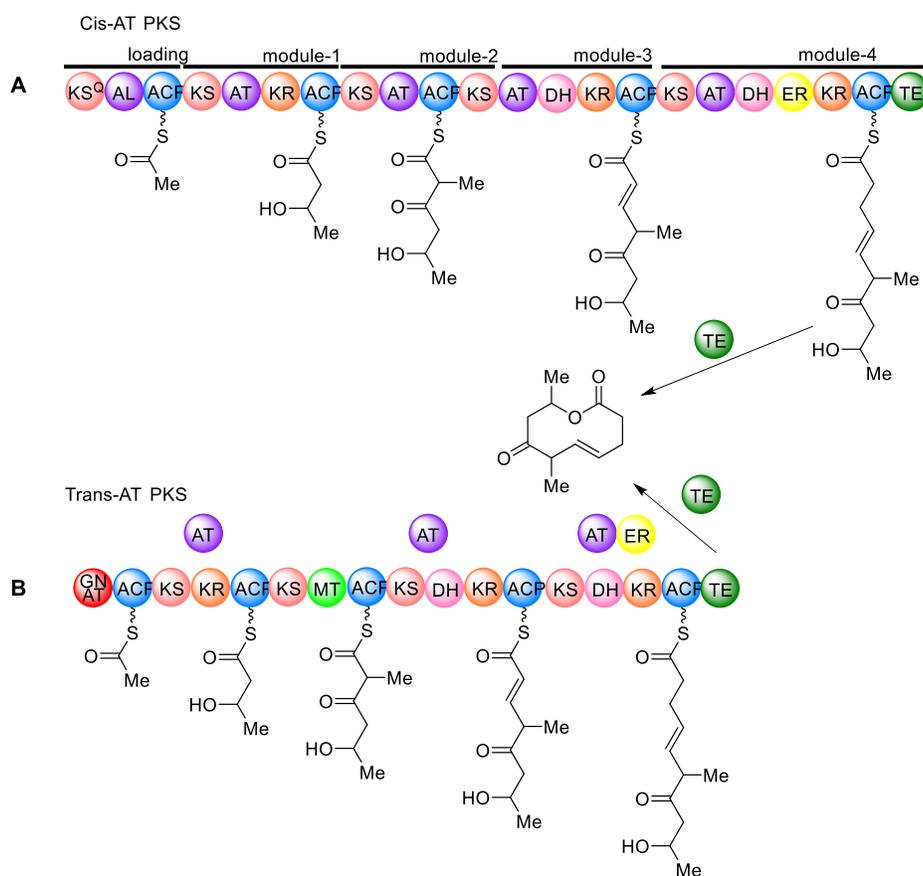

Figure 10. Schematic comparison of a theoretical *cis*-AT and a *trans*-AT PKS.[50] (**A**) Typical *cis*-AT PKS with AT domains integrated in each module. (**B**) Modular *trans*-AT PKS contains free-standing AT protein. Domain abbreviations: AL, acyl-ligase; KS^Q, decarboxylating KS present in many loading modules; GNAT, acetyl-loading AT of the GCN5-related N-acetyl transferase superfamily.

11

1.2.2 Type II PKSs

In contrast to the linear multimodular type I PKSs, type II PKSs are composed of dissociated enzymes. They all have core regions called 'minimal PKSs' consisting of three kinds of proteins: ketosynthase alpha (KS_α); ketosynthase beta, (KS_β, or chain length factor); and acyl carrier protein (ACP).[52] The KS_α and KS_β subunits show high sequence similarities, while their functions are totally different. KS_α catalyzes the Claisen-type C–C bond formations from CoA activated building blocks. KS_β was shown to be involved in the polyketide chain initiation by generating acetyl KS by decarboxylation of malonyl-ACP.[53] In addition, the KS_β subunit is the primary determinant of the polyketide chain length and it has thus also been termed 'chain length factor'. The minimal PKS works iteratively to assemble a nascent polyketide chain, exclusively using malonyl-CoA as elongation units. Furthermore, it partially controls the regiochemistry of the first ring cyclization (Figure 11A).[27, 52] Additional PKS subunits, including ketoreductases (KRs), cyclases (CYCs) and aromatases (AROs), work together to control the folding pattern of the poly-β-keto intermediate to form the aromatic cores. The KRs selectively reduce one keto group to stereo-selectively form a secondary alcohol, which can either influence the orientation of the poly-β-keto chain for a favored aldol condensation or define the configuration of a persisting carbinol moiety.[54] The CYCs function in a 'chaperone-like' manner and thus help in directing nascent polyketide intermediates into particular reaction channels to afford the desired cyclized rings.[54] The AROs also work during ring cyclization by dehydrating alcohol functions within the cyclic systems to yield aromatic rings.[54] Complex chemical modifications, including oxidation, dimerization, reduction, methylation or glycosylation will follow to ultimately make products with extraordinary structural diversity, as exemplified by the tetracenomycin (**28**) biosynthesis (Figure 11B).[27, 52] The products of type II PKSs are usually aromatic metabolites, such as tetracycline (**5**) and daunorubicin (**12**).
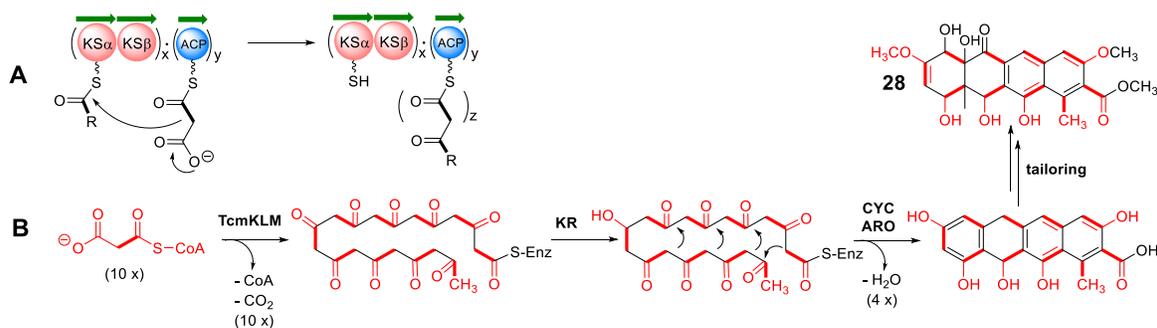

Figure 11. (**A**) Generic scheme for type II PKS assembly: the minimal PKS consists of KS_α, KS_β and ACP and works iteratively to assemble a nascent polyketide chain. The PKS proteins are shown by green arrows. (**B**) The tetracenomycin minimal PKS TcmKLM uses 10 units of malonyl-CoA to assemble a polyketide chain. After cyclization and post tailoring, tetracenomycin (**28**) is produced.

1.2.3 Type III PKS

Type III PKSs also work iteratively to build polyketide chains. However, unlike type I and type II PKSs, type III PKSs are ACP-independent systems.[27, 55] With the resulting simple structure, they are self-contained enzymes that form homodimers. Their single active site in each monomer catalyzes the priming, extension, and cyclization reactions iteratively to form polyketide products.[56] Type III PKSs do not require the presence of ACPs, as they can perform polyketide chain elongation directly using CoA building blocks (Figure 12A).[27, 55-56] Type III PKSs were once believed to be plant specific, but the first bacterial type III PKS, RppA, was reported in 1999 from *Streptomyces griseus*.[55] After that, studies have shown that type III PKSs are widely present in bacteria and fungi.[56-59] Typical type III PKSs are exemplified by chalcone synthases (CHS) from the plant aringenin chalcone (**29**) biosynthetic pathway and RppA from the bacterial flavolin (**31**) biosynthetic pathway (Figure 12B-C).[55]
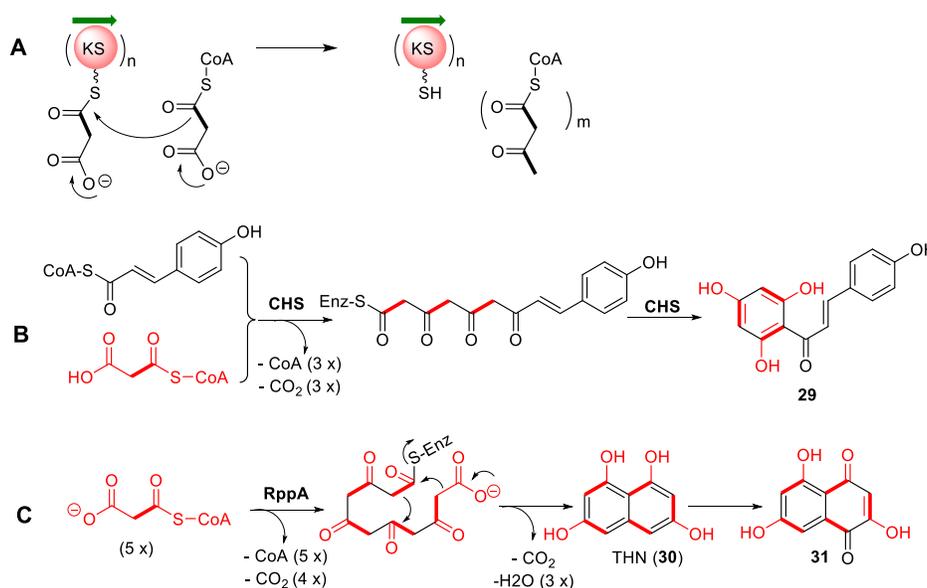

Figure 12. (**A**) Generic scheme for type III PKS assembly: a single subunit of a type III PKS catalyzes the priming, extension, and cyclization reactions iteratively to form polyketide products independent of an ACP domain. (**B**) Typical plant type III chalcone synthase (CHS) catalyzes the sequential condensation of three acetate units from malonyl-CoA to a 4-coumaroyl-CoA starter molecule. This is followed by a cyclization reaction, leading to the formation of an aromatic product **29**. (**C**) RppA catalyzes the synthesis of 1,3,6,8-tetrahydroxynaphthalene (THN, **30**) from 5 units malonyl-CoA.

1.3 Non-ribosomal Peptides and Non-ribosomal Peptide Synthetases

The non-ribosomal peptide natural products are another important class of compounds with diverse properties in drug discovery.[60] The above mentioned penicillin (**3**), vancomycin (**8**), actinomycin (**10**), bleomycin (**13**) and cyclosporin (**15**) are all non-ribosomal peptide natural products. Non-ribosomal peptides are synthesized by large multifunctional mega-proteins called non-ribosomal peptide synthetases (NRPSs) in a way very similar to the modular type I PKSs.[61] NRPSs also consist of diverse domains with different catalytic functions and the domains usually are grouped together in modules.[60-61] All the modules and domains work stepwise as assembly lines to produce the peptide natural products, just like the modular type I PKSs. The 'co-linearity rule' also applies to the NRPS in most cases.[31] Core domains of NRPSs are the condensation domains (C), adenylation domains (A) and peptidyl carrier protein domains (PCP). The PCP domain is also referred to as thiolation domain (T domain) and it also needs to be post-translationally activated from its *apo*-form to its *holo*-form by phosphopantetheinyl transferases (PPTases), just as in ACP activation (see chapter 1.2).[26] Additional domains with different functions can also be present, including methylation domains (MT), epimerization domains (E), heterocyclization domains (Cy), oxidative domains (Ox), reduction domains (R), formylation domains (F) or thioesterase domains (TE) (Figure 13A).[60-61]
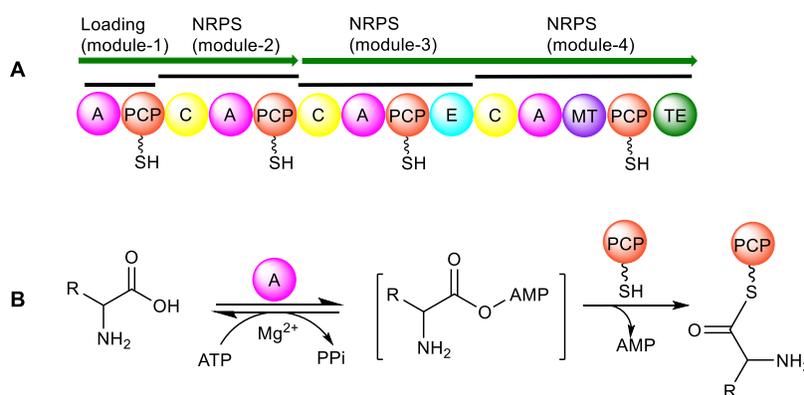

Figure 13. (**A**) Scheme representing a typical NRPS domain modular organization. Individual NRPS proteins are indicated by green arrows. Each module is delineated by solid black lines. (**B**) Activation of the amino acid building block by an A-domain catalyzed adenylation reaction in NRPS. Abbreviations: ATP, adenosine triphosphate; AMP, adenosine monophosphate; PPi, pyrophosphate; A, adenylation domain; PCP, peptidyl carrier protein domain; C, condensation domain

NRPSs typically use amino acids as building blocks, including both proteinogenic and non-proteinogenic amino acids, which is different to the PKSs that use CoA activated short carboxylic acids as precursors.[61-62] The A domain selectively activates a single amino acid building block and incorporates it in the respective module by converting it into an aminoacyl adenylate intermediate

at the expense of ATP. Then, the PCP domain incorporates the activated substrates by forming a covalent thioester bond to the phosphopantetheinyl tether (Figure 13B).[60-61] The C domain catalyzes amide bond formation between the PCP-bound substrate of a given module and the peptide chain attached by the previous module, thereby forming an elongated peptide chain at the respective PCP-domain.[60-62] The TE domain from the last module finally cleaves off the extended peptide chain from the biosynthetic enzyme by a hydrolysis or macrocyclization reaction, which results either in a linear peptide or macrocyclic product, respectively.[63] The additional domains are responsible for further modifications at the growing peptide chain. The MT domain catalyzes the transfer of the *S*-methyl group of *S*-adenosylmethionine (SAM) to the α-amino group of the PCP-bound peptide, thus making the peptide less susceptible to proteolytic breakdown.[63] The E domain promotes epimerization of the Cα-carbon of the PCP-tethered aminoacyl substrate to afford a D/L equilibrium, followed by specific incorporation of the D-amino acids into the growing peptide chain by action of the enantioselective donor site of the downstream C domain.[63] The Cy domain is the only domain that can replace the core C domain of the NRPS and perform both condensation and cyclodehydration of a cysteine, serine, or threonine to form a five-membered ring in the peptide backbone.[64] The oxidative state of the resulting oxazoline and thiazoline rings can be altered by additional oxidation (Ox) or reduction (R) domains.[63] F domains were initially found in the linear gramicidin nonribosomal peptide synthetase as the first domain in the first module. It transfers a formyl group from formyltetrahydrofolate (fH₄F) onto the first amino acid valine.[65] Complex post tailoring reactions can be catalyzed by further enzymes, such as cytochrome P450s, oxidoreductases, glycosyl transferases, halogenases, and deoxysugar biosynthetic enzymes to generate the fully modified natural products. NRPSs can be classified into three groups according to their biosynthetic logic: linear, iterative, and nonlinear NRPSs.[66-67]

1.3.1 Linear NRPSs

Linear NRPSs share similar domain organization and building block assembly principles when compared to the type I PKSs. They all perform as assembly line and adhere to the co-linearity rule. In linear NRPSs, the three core domains are arranged in the order of C-A-PCP in an elongation module.[66-67] Each module introduces a single amino acid to the PCP-bound peptide chain by peptide-bond formation catalyzed by the C domain. Thus, the number and order of modules in typical linear NRPSs matches that of the amino acids found in the corresponding products.[67-68] *Vise versa*, the amino acids can be used to predicted the organization of the NRPS modules. The additional tailoring domains usually are integrated in the modules. The terminal module in most cases contains a TE domain that can release the full-length peptide chain from the PCP by simple hydrolysis or macrolactonization.[69] Linear NRPSs are exemplified by vancomycin (**8**) biosynthesis (Figure 14). In vancomycin biosynthesis, each module recruits one amino acid as elongation unit

and fuses it to the former chain by peptide bond formation. The E domains in the respective modules change the stereochemistry of the corresponding amino acid into D in the peptide chain.[63] The chain elongation is performed step by step, finally furnishing the full-length peptide chain bound to the last PCP. After oxidative peptide tailoring by P450s, which introduce the prototypical biaryl bonds into the peptide chain, the terminal TE domain finally cleaves off the linear precursor by hydrolysis. Post NRPS tailoring by glycosylation is achieved by glycosyl transferases to afford the bioactive **8**.[63]
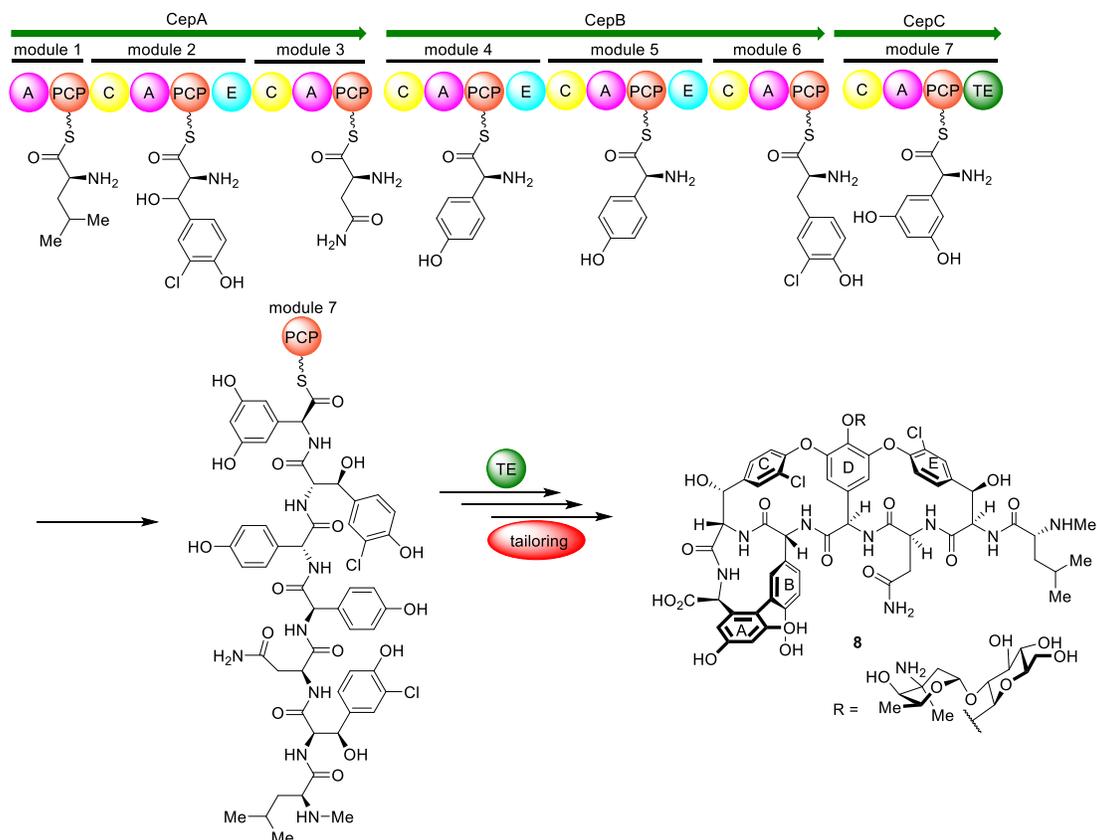
Figure 14. Linear NRPSs assembly line as exemplified by vancomycin (**8**) biosynthesis. In linear NRPS biosynthesis, each module is used once to add one building block to the peptide chain.

1.3.2 Iterative NRPSs

Iterative NRPSs reuse dedicated modules more than once (iterative) to build up the final product. Thus, the number of modules in iterative NRPSs does not reflect the number of amino-acid building blocks in the corresponding products.[66-67] As illustrated in gramicidin S (**32**) biosynthesis, the NRPSs first make a pentapeptide chain similar to linear NRPSs. Then, two identical pentapeptide halves are assembled in a head-to-tail manner by unusual TE-activity to make the gramicidin S dimeric structure (Figure 15).[70] Iterative NRPSs share the same domain organization and building block assembly principles with linear NRPSs. It is thus not yet possible to distinguish them by

16

analysis of their primary sequence.[67] Studies on gramicidin S biosynthesis showed that the PCP-TE of the GrsB catalyzes the dimerization and subsequent formation of the decapeptide lactam gramicidin S (**32**).[71] However, even here multiple sequence alignments with other TE domains do not allow the prediction of such an iterative function.[67] This high degree of specialization is additionally reflected in the low sequence identity (10%–15%) among TE-domains.[60]
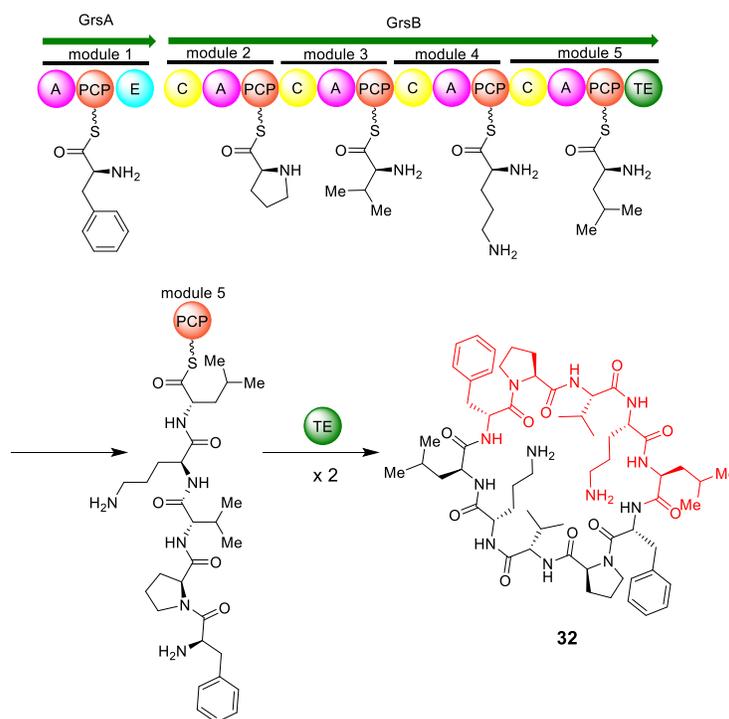

Figure 15. Iterative NRPSs assembly line as exampled by gramicidin S (**32**) biosynthesis. Iterative NRPSs reuse the entire assembly line or certain modules.

1.3.3 Non-linear NRPSs

Non-linear NRPSs assemble peptide chains using the domains in a non-linear way. Thus, the sequence of amino acids does not correlate to the arrangement of modules on the synthetase template.[66] Unusual arrangement of the domains are characteristic for this type of NRPSs, which often causes unusual cyclization or branching chemistry.[67] As illustrated in the vibriobactin (**33**) biosynthesis (Figure 16), the NRPSs use one molecule of norspermidine (NS), two molecules of threonine, and three molecules of dihydroxybenzoyl (DHB) to make the final product. The free soluble NS is used by the NRPSs as a nucleophile to be incorporated into the product without prior binding to the enzyme as a thioester.[67] In contrast to the linear or iterative NRPSs, there is little resemblance between the NRPS architecture of the biosynthetic machinery and the final peptide product in the non-linear NRPSs.

17
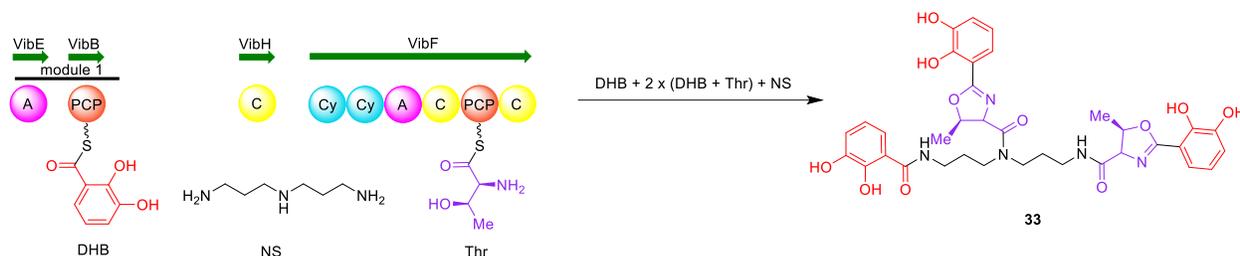
Figure 16. Non-linear NRPSs assembly line as exemplified by vibriobactin (**33**) biosynthesis. Non-linear NRPSs use the domains in a non-linear way.

1.3.4 Hybrid pathways

In addition to the above mentioned PKS and NRPS pathways, nature also uses hybrid pathways to produce different kinds of natural products. These hybrids can be a blend of different types of PKSs, NRPSs, or combinations of PKS-NRPSs.[72] The bleomycin (**12**) and rapamycin (**13**) natural products both belong to the PKS-NRPS hybrid products. This pathway hybrid is characterized by a mix of both PKS and NRPS modules, as also exemplified in PoTeM biosynthesis in chapter 1.2.1.2.[73-74] Single or several groups of domains are also frequently found in hybrid systems, mostly with unusual functions.[68] For example, a free-standing C domain (SgcC5) in C-1027 (**34**) biosynthesis catalyzes an ester bond formation (Figure 17).[75] Fatty acid precursors are also seen in PKS or NRPS products.[76] The previously mentioned noranthrone (**19**) uses a fatty acid to prime PKS assembly (Figure 7C).[32] Fatty acid precursors are also used by NRPS pathways, leading to the lipopeptide family of natural products, such as daptomycin.[77]
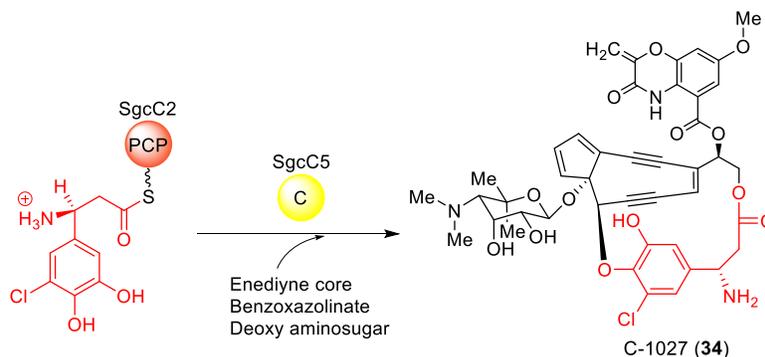

Figure 17. Attachment of (*S*)-3-chloro-5-hydroxy-β-tyrosine onto the enediyne core by SgcC5, leading to the C-1027 (**34**) chromophore. The timing for each of the coupling steps is unknown.

1.4 Genome Mining for Natural Products Discovery

Whole-genome sequencing technology has developed rapidly since the beginning of the 21th century, highlighted by the whole-genome sequencing of *Streptomyces coelicolor* and *Streptomyces avermitilis*.[78-79] The whole-genome sequence of these organisms showed that there were many more potential biosynthetic gene clusters (BGCs) in the genome than the number of compounds that had been identified, despite many years of intensive natural product research on these model organisms.[78-79] For example, analysis of the complete genome sequence of *Streptomyces coelicolor* A3(2) identified 16 more BGCs than the already identified natural products.[80] Meanwhile, the cost of genome sequencing further dropped considerably, allowing greater accessibility to genomic data.[81] Thus, researchers can focus on the microbes that they are interested in according to BGC predictions based on the genomic information prior to compound isolation, leading to modern natural product discovery by so-called genome mining.[82-85] Automated *in silico* bioinformatics platforms have been developed to facilitate the prediction of natural products encoded by BGCs, which benefit natural product discovery greatly.[86] One of the most comprehensive platforms currently available is the 'antibiotics and secondary metabolite analysis shell' (antiSMASH), which can identify up to 44 different gene cluster types based on hits against a library of enzymes/protein domains commonly observed in secondary metabolite biosynthetic pathways.[87-91] However, it is estimated that less than 10 % of the predicted BGCs are expressed in sufficient quantities to be observed under conventional laboratory growth conditions, leaving the rest as cryptic gene clusters that probably encode novel or unknown metabolites.[5]

One major target of genome mining is to unlock the full chemical potential of these cryptic BGCs. Currently, the methods used to achieve this goal can be summarized into two major categories: homologous expression and heterologous expression. Approaches for homologous expression use native hosts as producer, which include (a) optimization of the growth conditions, (b) co-culture with other microorganisms, (c) supplementation of the cultures with signaling molecules, (d) ribosome engineering, and (e) manipulation of the target genome.[84] Approaches (a) to (d) need no genetic manipulation. Therefore, they are technically rather simple ways to induce the expression of BGCs even when only little is known about the cryptic BGCs. Another advantage of these methods is that they can induce pleiotropic changes in the expression of the BGCs, leading to the simultaneous expression of multiple BGCs. However, these methods are intrinsically empirical, making it difficult to predict the outcomes and hard to directly target the most interesting BGCs. Simultaneous expression of multiple BGCs can also be a drawback: it can lead to a complex mixture of dozens of compounds that make the identification and isolation of individual compounds difficult.[18] Compared to approaches (a) to (d), approach (e) needs genetic manipulation. This approach can also be pleiotropic, such as transposon mutagenesis and manipulation of global regulators. Some genetic concepts for genetic manipulations focus on single BGCs, thus making

this method pathway specific. These methods include overexpression of positive regulatory genes, disruption of negative regulatory genes, whole gene cluster duplication, and refactoring of promoters / ribosome binding sites.[92]

Homologous expression can be limited by the native producers when they are slow growing, not suited for large-scale laboratory growth conditions, genetically intractable or currently not culturable under standard laboratory conditions.[93] Indeed, it is estimated that 99% of the bacterial species from soil have not been cultivated in the laboratory.[94-95] Heterologous expression uses a surrogate host as producer that can bypass these limitations. Two major factors need to be considered for heterologous expression are (a) how to obtain the desired BGCs and (b) which suitable host system to choose. Sources of the desired BGCs can be genomic DNA from the culturable organism of the native producers, or culture-independent metagenomic materials. One challenge for obtaining the desired BGCs is how to transfer it into suitable expression vectors. Traditionally, the most common method has been constructing genomic/metagenomic libraries, followed by screening colonies to identify the desired BGCs. Screening such libraries is time-consuming and for this methodology it is generally difficult to cope with gene clusters >40 kb.[18] Recently, several direct cloning systems have been developed to bypass genomic library construction and to deal with large gene clusters. These methods include the transformation-associated recombination cloning (TAR),[96] RecET-mediated linear-plus-linear homologous recombination (LLHR),[97] Cas9-assisted targeting of chromosome segments (CATCH),[98] direct pathway cloning (DiPaC),[99-100] and site-specific recombination relied reactions (Figure 18A-E).[101-103] TAR cloning relies on the *in vivo* homologous recombination ability of yeast (Figure 18A).[96] It can be used to clone DNA fragments up to 250 kb from multiple complex genomes.[104] This approach was also used to clone a 67 kb NRPS gene cluster, leading to the discovery of taromycin after gene cluster refactoring.[105] LLHR relies on the homologous recombination ability of phage recombination systems in *E. coli* (Figure 18B). It was shown to be able to capture BGCs up to 52 kb in size from *Photorhabdus luminescens*.[97] This method was also used to clone the 106 kb salinomycin gene cluster from *Streptomyces albus* with downstream three-piece assembly.[106] Extension of this method by combination of LLHR with exonuclease *in vitro* assembly (ExoCET) even facilitated to clone the complete 106 kb salinomycin gene cluster in a single step.[107] CATCH uses the programable CRIPSR/Cas9 system to cleave the target DNA fragments from intact genome *in vitro* and then uses Gibson assembly to ligate the DNA fragment with the cloning vector (Figure 18C). This method was shown to be able to clone bacterial genomic sequences of up to 100 kb in a single step.[98] DiPaC depends on the ability to amplify long DNA fragments using high-fidelity PCR polymerases, such as the Q5 polymerase (Figure 18D). After PCR amplification of the target gene clusters, the PCR products can be ligated to a vector of choice using Gibson assembly or Sequence- and ligation-independent cloning (SLIC). This method has been used to

clone several BGCs from different sources.[99-100, 108] Site-specific recombination methods rely on site-specific recombination systems, such as Cre/loxP or Φ BT1 integrase/*att*B/*att*P systems (Figure 18E).[101-103] This method was used to clone several biosynthesis gene cluster from *Streptomyces* at a frequency higher than 80%.[103]
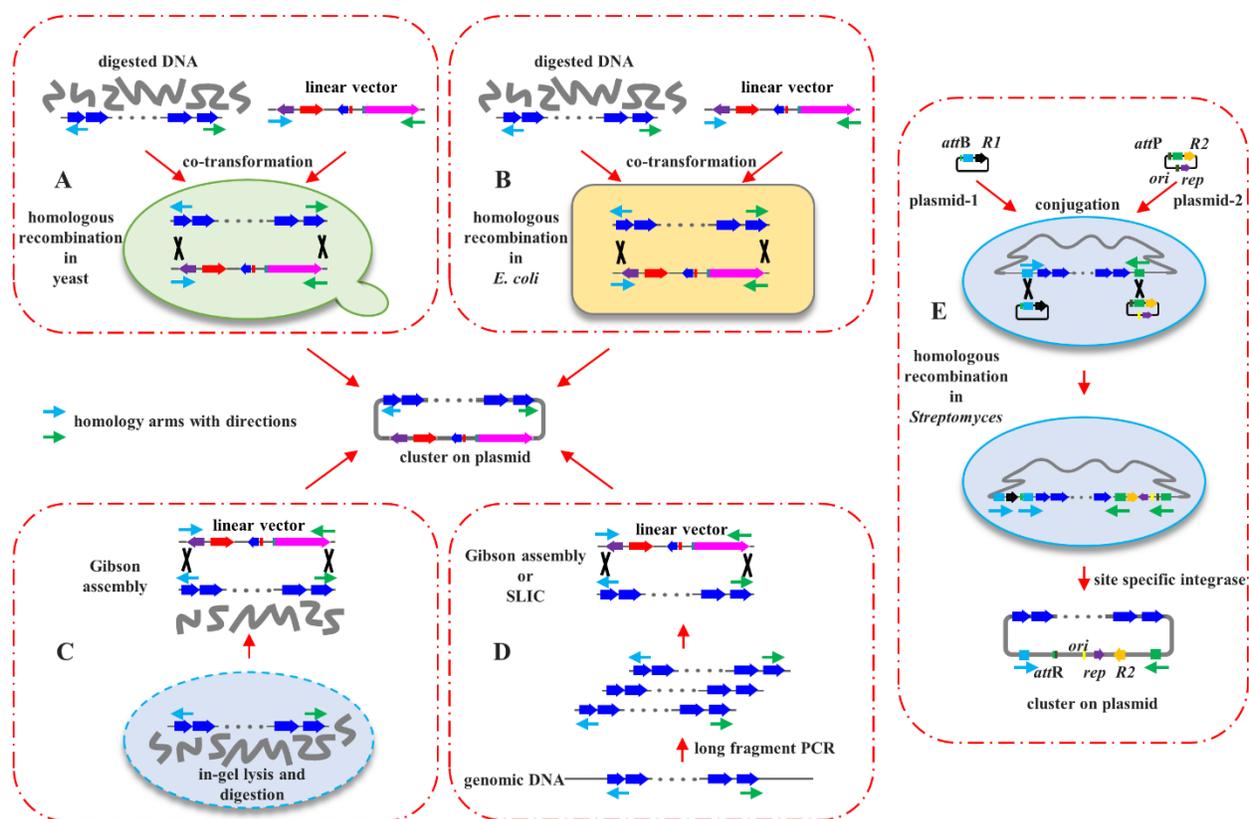

Figure 18. Recently developed methods to clone BGCs. (A) TAR cloning in yeast. After co-transforming the digested DNA and the linear vector into yeast cells, homologous recombination occurs between the designed homology arms. (B) LLHR cloning in *E. coli*, specifically in *E. coli* GB05-dir. After co-transforming the digested DNA and the linear vector into induced *E. coli* GB05-dir cells, homologous recombination occurs between the designed homology arms. (C) *In vitro* CATCH cloning. The cells carrying the target BGC are lysed in gel and the chromosomes are cleaved by RNA-guided Cas9 at the designated target sites. Afterwards, Gibson assembly is applied to assemble the target DNA fragment into the linear vector. (D) *In vitro* DiPaC cloning. Target DNA fragments are PCR amplified from the genomic DNA. Afterwards, Gibson assembly is applied to assemble the amplified DNA fragments into the linear vector. (E) Site-specific recombination method. Plasmid-1 is a suicide plasmid designed with *att*B site and a region homologous to one end of the cluster. Plasmid-2 is a temperature sensitive plasmid designed with *att*P site and a region homologous to the other end of the cluster. These two plasmids are transferred into *Streptomyces* separately and homologous recombination occurs to integrate them into the chromosome. Site specific integrase is applied *in vivo* or *in vitro* to induce specific recombination between the *att*B/*att*P sites, resulting in a plasmid backbone from the plasmid-2 carrying the target gene cluster. R1 and R2 indicate different resistant genes.

After cloning the BGCs into vectors, they need to be expressed in a suitable surrogate host for heterologous compound production. Commonly used heterologous hosts are model strains or genetically modified strains, which can transcriptionally activate and produce molecules encoded by diverse clusters, such as *Streptomyces* and *Escherichia coli*. *Streptomyces* strains have been used for heterologous expression for decades, such as *S. albus*, *S. coelicolor* and *S. lividans*.[109] Facilitated by the advancement of DNA manipulation methods, genetically modified *Streptomyces* hosts were recently constructed for better heterologous expression. *S. albus* Del14 was constructed by deleting 15 clusters encoding secondary metabolite biosynthetic pathways from *S. albus* J1074.[110] *S. coelicolor* M1152 and *S. coelicolor* M1154 were constructed by deleting four antibiotic gene clusters from *S. coelicolor* M145 and introducing *rpoB* or *rpoB* + *rpsL* mutations, respectively.[111] These genetically modified hosts have reduced genomes and can offer a 'clean' natural product background, which can benefit heterologous expression and particularly downstream product identification and isolation. Different types of natural products have been successfully characterized using *Streptomyces* as hosts, including PKS, NRPS and PKS-NRPS hybrid products.[112] *Streptomyces* hosts are often used when the BGCs are from the same or related species. Different BGCs can have different expression performance when introduced into different hosts, showing complex cluster-host interactions.[110] *Escherichia coli* is the most common bacterial strain for biolabs, which is an ideal host for heterologous expression due to its fast growth rate, gentle culture conditions and versatile DNA manipulation tools.[113] *E. coli* BAP1, a strain derived from *E. coli* BL21(DE3) by genomic integration of a single copy of the *sfp* gene (encoding a PPtase from *Bacillus subtilis*), was used to produce the PKS type products 6-deoxyerythronolide B (**16**) from *Saccharopolyspora erythraea*.[114] This strain can also be used to produce several NRPS type natural products from cyanobacteria and *Serratia*, or PKS-NRPS hybrid type ikarugamycin from *Streptomyces*.[100, 108, 115-116]

In addition, the genetical manipulation methods used in homologous hosts can also be applied to heterologous expression, making heterologous expression an important way for genome mining. The genome sequence analysis, gene cluster cloning, and heterologous expression thus paves the way to novel natural products production.

2. Aim of the Thesis

As evident from the numerous important natural products presented above, *Streptomyces* is an excellent source of novel and biomedically interesting small molecules. One example is *Streptomyces* sp. Tü6314, which was isolated by Prof. Dr. Hans-Peter Fiedler (University of Tübingen, retired) in Egerpatak, Ardeal, Romania in 2001 and supplied to our laboratory for further studies. Dr. Janine Antosch, a former PhD student from the Gulder lab, had previously shown that this strain produces cycloheptamycin A, a cyclodepsipeptide natural product which was first discovered in the 1970s.[117] Within her work, cycloheptamycin A was found to have strong and selective inhibitory potential against *Propionibacterium acnes*, a human pathogen causing severe cosmetic and other health damages. Despite this interesting biomedical property, the absolute stereostructure of cycloheptamycin A as well as its biosynthetic origin remained elusive. Given the biological potential of this compound and the yet underinvestigated metabolic profile of *Streptomyces* sp. Tü6314, the resulting aims of this thesis were:

(1) Isolation and full stereochemical characterization of cycloheptamycin A and natural derivatives from *Streptomyces* sp. Tü6314.

(2) Genome analysis of *Streptomyces* sp. Tü6314 for the identification of the cycloheptamycin BGC, its recombinant expression and functional characterization, including the production of altered structural derivatives.

(3) Analysis of the full metabolic potential of *Streptomyces* sp. Tü6314 by *in silico* genome analysis and development of a methodology to rapidly access natural products encoded by silent BGCs from this strain.

Altogether, this thesis therefore focused on natural product isolation and biosynthetic studies from *Streptomyces* sp. Tü6314, paving the way for a better understanding of the full biosynthetic potential of this *Streptomyces*.

3. Results and Discussion

This thesis is submitted as a publication-based dissertation. The content of the individual publications will briefly be summarized in the following chapters.

3.1 Structures and biological activities of cycloheptamycins A and B

The following chapter is based on the publication:

Z. Qian, * J. Antosch,* J. Wiese, J.F. Imhoff, H.-P. Fiedler, A. Pöthig, T.A.M. Gulder. Structures and biological activities of cycloheptamycins A and B, *Org. Biomol. Chem.* **2019**, *17*, 6595-6600.[118] *equal contribution (Highlighted in OBC HOT article collection.)
This paper is available online:

https://pubs.rsc.org/en/content/articlelanding/2019/ob/c9ob01261c/unauth#!divAbstract

Reproduced from Ref. 118 with permission from The Royal Society of Chemistry.

Within this work, we established *Streptomyces* sp. Tü6314 as a new producer of the literature-known cycloheptamycin A (**35**) along with a new analog, cycloheptamycin B (**36**). The planar structure of these compounds was for the first time established by in-depth NMR structure elucidation combined with MS/MS analysis. In addition, full peptide hydrolysis and chemical modification applying Marfey's method for amino acid analysis were applied to elucidate the absolute configuration of several amino acid building blocks. This information was further utilized in combination with single-crystal X-ray analysis to determine the full stereostructure of the cycloheptamycin peptide family (Figure 19).
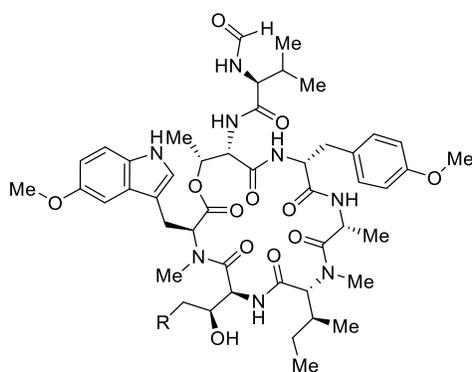
Cycloheptamycin A (**35**), R = CH$_3$
Cycloheptamycin B (**36**), R = H

Figure 19. Structures of cycloheptamycin A (**35**) and B (**36**).

In collaboration with J. Wiese and J.F. Imhoff, the biomedical potential of this compound family was evaluated. The results revealed a strong and selective antibiotic potential of cycloheptamycin A (**35**) against *Propionibacterium acnes* (IC$_{50}$ = 4.22 ± 0.41 µM), combined with no toxic effects

against tested human cells. The cycloheptamycins thus represent promising lead structures for the development of selective antibiotics against this human pathogen.

PAPER

Check for updates

Structures and biological activities of cycloheptamycins A and B†

Zhengyi Qian,‡[a] Janine Antosch,‡[a] Jutta Wiese,[b] Johannes F. Imhoff,[b] Hans-Peter Fiedler,[c] Alexander Pöthig [ID] *[d] and Tobias A. M. Gulder [ID] *[a,e]

The heptadepsipeptide cycloheptamycin A was isolated from the terrestrial *Streptomyces* sp. Tü 6314. Its constitution was elucidated on the basis of NMR spectroscopic experiments and mass spectrometric analysis. Its stereostructure was investigated by peptide hydrolysis and derivatization and firmly established by X-ray structure analysis. In addition to the parent compound, a new cycloheptamycin analog, cycloheptamycin B, was discovered and structurally assigned using comparative MS/MS experiments and NMR. The biological profile of both compounds was investigated, revealing a selective inhibitory potential of cycloheptamycins against *Propionibacterium acnes*.

1. Introduction

Owing to their densely populated and thus highly challenging natural environment, soil-dwelling microorganisms have evolved a virtually inexhaustible arsenal of small molecules with important biological functions that confer competitive advantages to their producers. The biological functions of such natural products are diverse, ranging from facilitators of intra and inter species communication to cytotoxic or antimicrobial properties.[1] Actinobacteria are particularly talented producers of such specialized metabolites.[2] It is thus not surprising that some of the most important antibiotics in medical use originate from these organisms, including β-lactams, glycopeptides, tetracyclins and macrolides.[3] Due to this tremendous metabolic potential, actinobacteria have extensively been screened for the production of bioactive small molecules, both in academia and industry. As a consequence, rediscovery rates of known compounds have continuously increased, ultimately leading to a decline of classical natural product discovery programs, in particular in the chemical industry.[4,5] With the emergence of multi-drug resistant human pathogens, however, revitalizing natural product discovery efforts becomes increasingly important.[6–8] Given their impressive track record as a prolific source of drug leads and the astonishing phylogenetic diversity of actinobacteria, combined with state-of-the-art analytical and biomolecular methodology, new small molecules and/or important new biological properties of known metabolites will certainly continue to be discovered and fuel current drug development pipelines.[3] Within our interest in antibacterial compounds from bacterial sources we herein investigated the metabolic potential of *Streptomyces* sp. Tü 6314.

2. Results and discussion

Strain Tü 6314 was isolated from soil from Egerpatak, Ardeal, Romania in 2001. Analysis of the 16*S* rRNA gene sequence by BLAST showed 99% identity with a large number of *Streptomyces* strains, thus evidencing that this strain belongs to the genus *Streptomyces*. A coalescent tree of single copy genes using the ARTS program package[9] revealed it to be most closely related to *Streptomyces pratensis* ATCC 33331. To investigate the metabolic potential of Tü 6314, the strain was cultivated in oatmeal medium (ISP medium 3),[10] extracted with ethyl acetate and the organic extract further processed by stepwise purification using size exclusion followed by semi-preparative high-performance liquid chromatography (HPLC). The main product thus obtained possessed a molecular mass of 949.5022 or 971.4848 units, as determined by high resolution positive ESI-MS spectrometry, best fitting a molecular compo-

[a]Biosystems Chemistry, Department of Chemistry and Center for Integrated Protein Sciences Munich, Technical University of Munich, Lichtenbergstraße 4, 85748 Garching, Germany
[b]GEOMAR Helmholtz Center for Ocean Research Kiel, RD3 Marine Microbiology, Düsternbrooker Weg 20, 24105 Kiel, Germany
[c]Institute of Microbiology, University of Tübingen, Auf der Morgenstelle 28, D-72076 Tübingen, Germany
[d]Department of Chemistry and Catalysis Research Center (CRC), Technical University of Munich, Ernst-Otto-Fischer-Straße 1, 85748 Garching, Germany.
E-mail: alexander.poethig@tum.de
[e]Chair of Technical Biochemistry, Technical University of Dresden, Bergstraße 66, 01069 Dresden, Germany. E-mail: tobias.gulder@tu-dresden.de
†Electronic supplementary information (ESI) available. CCDC 1900571. For ESI and crystallographic data in CIF or other electronic format see DOI: 10.1039/c9ob01261c
‡These authors contributed equally to this work.

This journal is © The Royal Society of Chemistry 2019

Org. Biomol. Chem., 2019, **17**, 6595–6600 | 6595

sition of $C_{48}H_{68}N_8O_{12}$ in its protonated form (calc. 949.5029) or as sodium adduct (calc. 971.4849), respectively (see ESI, Fig. S1†). This resulted in a calculated 19 degrees of unsaturation. The ^{1}H-NMR spectrum clearly suggested the analyte to be of peptidic nature, bearing 6–9 amino acid residues, some of which equipped with aromatic protons, as well as methoxy and N-methyl functionalities. In depth analysis of ^{13}C,

2D-COSY, NOESY, HSQC and HMBC data allowed for the unambiguous identification of 7 amino acid building blocks, namely N-formyl valine, threonine, O-methyl tyrosine, alanine, N-methyl isoleucine, β-hydroxy norvaline, as well as N-methyl-5-methoxy tryptophan (Table 1 and ESI, Fig. S4–S11†).

These building blocks accounted for a total of 18 double bond equivalents, thus pointing at a macrocyclic overall struc-

Table 1 NMR data of cycloheptamycin (**1**) at 400 MHz (^{1}H) and 100 MHz (^{13}C) in DMSO-d_6

Amino acid	Signal	^{1}H [ppm]	COSY, Mult. [J in Hz]	^{13}C [ppm]	HMBC	NOESY
Tryptophan	1			171.7 (C)		
	2	4.55	3, dd [12.1, 5.1]	61.0 (CH)	1, 3, 4, 15	3, 6, 12
	3a	3.42	2, 3b, dd, [12.8, 5.1]	24.7 (CH$_2$)	1, 2, 4, 5, 12	2, 6, 12
	b	3.13	2, 3a, pt [12.6]			
	4			108.0 (C)		
	5			127.5 (C)		
	6	7.12	8, d [2.4]	99.8 (CH)	4, 7, 8, 10	2, 3, 13
	7			153.2 (C)		
	8	6.74	6, 9, dd [8.8, 2.4]	111.5 (CH)	6, 7, 10	9, 13, 45
	9	7.27	8, d [8.8]	112.0 (CH)	4, 5, 6, 7, 10, 12	8, 13, 45
	10			131.4 (C)		
	11	10.80	12, d [2.4]	(NH)	4, 5, 10, 12	n.d.
	12	7.08	11, d [2.4]	124.4 (CH)	3, 4, 5, 7, 10	2, 3, 45
	13	3.68	s	55.3 (CH$_3$)	7	6, 8, 9
	14	3.34	s	31.8 (CH$_3$)	2, 15	^{b}
β-Hydroxy norvaline	15			170.8 (C)		
	16	5.17	17, 21, m	55.3 (CH)	15, 17, 18, 22	14, 17, 18, 19, 21
	17	3.71	16, 18, 20, m	72.3 (CH)	15, 18, 19	^{b}
	18	1.31	17, 19, m	24.0 (CH$_2$)	17, 19	^{b}
	19	0.88^a	18, t [7.3]	10.3 (CH$_3$)	17, 18	^{b}
	20	4.90	17, bs	(OH)		14
	21	8.37	16, d [8.7]	(NH)	15, 16, 22	16, 18, 19, 23, 45
Isoleucine	22			168.0 (C)		
	23	4.67	24, d [11.0]	63.8 (CH)	22, 24, 28, 29	21, 25, 26, 27, 28, 30, 31
	24	2.08	23, 27, m	32.7 (CH)	22, 27	25, 26, 27, 28
	25a	1.49	25b, 26, m	24.6 (CH$_2$)	23, 24, 26, 27	23, 24, 28
	25b	1.16	25a, 26, m			
	26	0.89^a	25, t [7.3]	10.8 (CH$_3$)	24, 25	^{b}
	27	0.73	24, d [7.3]	14.2 (CH$_3$)	23, 24, 25	^{b}
	28	2.73	s	28.6 (CH$_3$)	23, 29	23, 24, 25, 26, 27, 31
Alanine	29			172.9 (C)		
	30	5.00	31, 32, m	43.4 (CH)	29, 31, 33	23, 27, 31, 32
	31	1.28	30, d [6.6]	17.1 (CH$_3$)	29, 30	(17), 23, 28, 30, 32
	32	8.83	30, d [7.4]	(NH)	33	30, 31, 34, 35
Tyrosine	33			170.8 (C)		
	34	4.59	35a/b, 41, m	53.2 (CH)	33, 35, 36, 42	32, 35, 37, 38, 41
	35a	2.99	34, 35b, m	38.6 (CH$_2$)	33, 34, 36, 37	32, 34, 35, 38, 41
	35b	2.37	34, 35a, m			
	36			129.2 (C)		
	37	6.91	38, d [8.5]	130.5 (CH)	35, 36, 38, 40	34, 35, 38, 41, 45
	38	6.22	37, d [8.5]	113.0 (CH)	36, 40	34, 35, 37, 45
	39			157.4 (C)		
	40	3.56	s	54.9 (CH$_3$)	39	^{b}
	41	8.01	34, d [9.2]	(NH)	33, 42	34, 35, 37, 43, 44, 45
Threonine	42			165.3 (C)		
	43	4.33	44, 46, dd [8.0, 4.1]	52.7 (CH)	42, 44, 45, 47, 48	41, 45, 46
	44	4.26	43, 45, m	68.8 (CH)	-	41, 46
	45	-0.21	44, bs	12.7 (CH$_3$)	n.d.	41, 46, 48
	46	7.47	43, d [7.9]	(NH)	42, 43, 47	43, 44, 45, 48, 49, 52, 53
N-Formyl valine	47			169.6 (C)		
	48	4.22	49, 52, dd [9.2, 6.0]	55.8 (CH)	47, 49, 50, 51, 53	45, 46, 49, 52
	49	1.90	48, 50, 51, m	30.4 (CH)	47, 48, 50, 51	46, 48, 53
	50	0.76	49, d [6.8]	19.2 (CH$_3$)	48, 49, 51	^{b}
	51	0.71	49, d [7.1]	17.5 (CH$_3$)	48, 49, 50	^{b}
	52	8.11	48, d [9.2]	(NH)	48, 53	46, 48, 51
	53	8.01	s	160.9 (COH)	48	46, 49, 50, 51

^{a} Assignment might be inverse. ^{b} Correlations not unambiguously assignable due to signal overlap/noise.

27
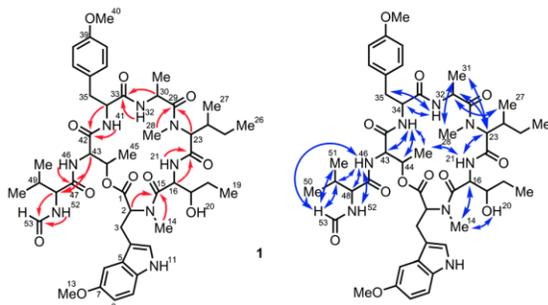
Fig. 1 Key HMBC (left, red arrows) and 2D NOESY (right, blue double arrows) NMR correlations used to assemble the planar structure of cycloheptamycin A (**1**).

ture. The relative position of the amino acid units within the heptapeptide backbone was evident from a series of decisive HMBC correlations, in particular from NH-, and *N*-methyl functionalities, as well as from the amino acid α protons (Fig. 1, left). These included correlations from H-2 and *N*-Me-14 of tryptophan to the β-hydroxy norvaline carbonyl C-15, of H-16 and NH-21 to the carbonyl C-22 of *N*-Me isoleucine, of H-23 and N-Me-28 to the alanine carboxy C-29, of H-30 and NH-32 to the *O*-methyl tyrosine keto function C-33, of H-34 and NH-41 to the threonine amide C-42, from H-43 and NH-46 to carbonyl C-47 of valine, and also from H-48 and NH-52 to the *N*-formyl group C-53. The peptide arrangement obtained by this data thus only left the threonine hydroxyl function and the tryptophan carboxylic acid as macrocyclization sites, leading to the overall depsipeptide structure **1**. This amino acid arrangement was further corroborated by a series of inter amino acid NOESY interactions (Fig. 1, right) between tryptophan and the β-hydroxy norvaline (*N*-Me-14 ↔ H-16 and OH-20), β-hydroxy norvaline and *N*-Me-isoleucine (NH-21 ↔ H-23), *N*-Me-isoleucine and alanine (H-23 ↔ H-30 and Me-31, Me-27 ↔ H-30, NMe-28 ↔ Me-31), alanine and *O*-Me-tyrosine (NH-32 ↔ H-34 and H-35), *O*-Me-tyrosine and threonine (NH-35 ↔ H-43, H-44 and Me-45), threonine and valine (NH-46 ↔ H-48, H-49 and NH-52), as well as from threonine (NH-46 ↔ H-53) and valine (H-49, Me-50, Me-51 ↔ H-53) to the terminal formyl substituent.

Literature search with the structure thus obtained revealed that **1** had already been reported in 1970 by Godtfredsen *et al.*[11] In the corresponding manuscript, however, only the planar structure of **1** was established by in-depth mass spectrometric analyses combined with a series of chemical derivatization reactions. The resulting molecular structure derived of this impressive but tedious experimental achievement is now verified by our results. We thereby also report the high-resolution NMR spectroscopic data of **1** for the first time.

Godtfredsen *et al.* also conducted first experiments towards the elucidation of the absolute configuration of **1**.[11] To address this problem, they fully hydrolyzed **1** by treatment with 6 N HCl and the resulting hydrolyzate was used in two separate biocatalytic transformations using selective ᴅ- and

ᴌ-amino acid oxidases. Using this methodology, ᴌ-valine, ᴅ-OMe-tyrosine, ᴅ-alanine and ᴌ-*erythro*-β-hydroxy norvaline were identified, thus establishing the absolute configuration of 4 out of the 7 amino acid residues of their isolated cycloheptamycin (**1**). To corroborate these results for our material, we used Marfey's method.[12,13] This led to the verification of the results from the amino acid oxidase treatments by Godtfredsen *et al.* for ᴅ-alanine and ᴌ-*erythro*-β-hydroxy norvaline and additionally revealed the threonine residue to be ᴌ-configured (see ESI, Fig. S17–S19†). The stereochemical identity of the isoleucine and the *O*-Me-tryptophan building blocks remained elusive due to the *N*-Me-groups attached to these moieties and, in addition, the instability of tryptophan under the harsh reaction conditions needed for complete peptide hydrolysis. Having firmly elucidated the absolute configuration of most of the amino acid building blocks in **1**, we intended to solve its full stereostructure by X-ray structure analysis. After considerable screening efforts to identify suitable crystallization conditions for **1**, crystals suitable for diffraction measurements were ultimately obtained by dissolving **1** in methanol with 5 mM ammonium acetate with slow evaporation of the solvent at room temperature. The obtained crystals enabled the collection of crystallographic data up to a resolution of 0.83 Å, facilitating to validate the connectivity in **1** as elucidated based on the NMR data above.[14] The details concerning the measurement, data processing and model refinement are provided in the ESI.† Fig. 2 shows the molecular structure of **1** in the solid state, confirming the 19-membered ring structure. Although the position of the respective hydrogen atoms could not be refined freely, the conformation of the ring is most likely determined by two intramolecular hydrogen bonds: the amide nitrogen atoms N-21 and N-41 are perfectly located and oriented towards the opposing carbonyl oxygen atoms, which can act as acceptor atoms (dotted lines in Fig. 2). The diffraction data did not allow for the direct determination of the absolute stereostructure of **1** using anomalous dispersion. However, the absolute configurations of the amino acid building blocks deduced after peptide hydrolysis (see above) in combination with the relative configuration of **1** unambiguously derived from the diffraction experiment facilitated the elucidation of the full stereostructure of **1** as shown in Fig. 2. The conformation of **1** adopted in the crystal structure also helps understand the observed NOESY cross-peaks of the methyl group of the threonine unit (position 45) to the aromatic protons H-8, H-9 and H-12 of the tryptophan and to H-37 and H-38 of the tyrosine moiety (*cf.* Table 1 and structure of **1** in Fig. 2B, left), which were not expected from the alternative representation of **1** (Fig. 3B, right) that more clearly depicts its overall connectivity but ignores the apparent conformation. In addition, the crystal structure suggests methyl C-45 to be located between the aromatic side-chains of both tyrosine and tryptophane, which might explain its unusual chemical shift of −0.21 ppm as a result of ring current effects.

In addition to the parent compound **1**, the isolation of a minor component **2** of the extract with a slightly decreased

This journal is © The Royal Society of Chemistry 2019

Org. Biomol. Chem., 2019, **17**, 6595–6600 | **6597**
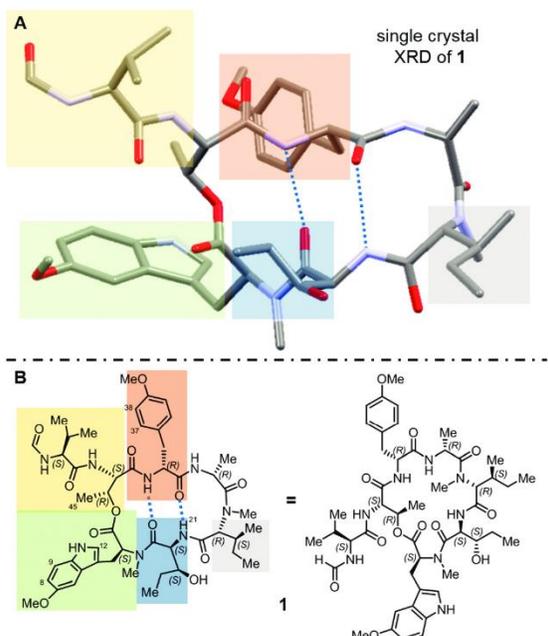
Fig. 2 (A) Molecular structure of cycloheptamycin A (**1**) in the solid state determined by single crystal X-ray diffraction. Atoms are shown as capped sticks (carbon: grey, nitrogen: blue, oxygen: red) and likely intramolecular hydrogen bonds are shown as blue dotted lines. For reasons of clarity, only one independent molecule of the asymmetric unit is shown, hydrogen atoms are omitted. (B) Connectivity scheme (left) and alternative graphical representation (right) of **1**.
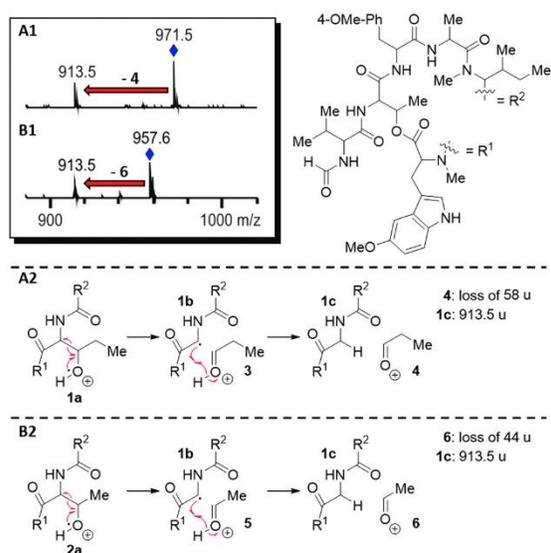

Fig. 3 Structure elucidation of cycloheptamycin B (**2**) based on comparative MS/MS analysis with **1**.

retention time and an identical UV spectrum as revealed by online HPLC-DAD analysis was likewise possible. The molecular mass of this compound was determined as 957.4692 u, which perfectly matched a molecular composition of $C_{47}H_{66}N_8O_{12}Na$ (calc. 957.4692). The mass difference of 14 u was strongly indicative of a formal loss of CH_2 when compared to cycloheptamycin A (**1**). We thus speculated that the new analog cycloheptamycin B (**2**) is most likely a result of incomplete O- (at the Trp or Tyr residue) or N-methylation (at the Trp or Ile unit) during biosynthesis, or might alternatively be a product of imperfect substrate selection during peptide assembly, for example replacing Val versus Ile or Thr versus β-hydroxy norvaline in the peptide backbone. The method of choice to identify the location of the missing methyl group was comparative MS/MS of **1** and **2**. Importantly, the MS/MS data of both compounds revealed an identical product ion peak at m/z 913.5. This corresponds to a formal loss of C_3H_6O from **1** or of C_2H_4O from **2** and is thus consistent with the mass difference of 14 u between these two molecules (Fig. 3, A1 and B1). In cycloheptamycin A (**1**), the formation of the product ion at 913.5 u can be explained by a cleavage of the β-hydroxy norvaline side chain. Upon ionization of **1** the radical cation **1a** can be formed (Fig. 3, A2). This promotes cleavage of the C,C-bond in relative α-position to the secondary alcohol and thus leads to a loss of **3** and formation of radical **1b**. The latter can abstract a hydrogen radical from **3** under formation of **1c**, the compound detected at 913.5 u, and radical cation **4**. The identical mechanism can be proposed for cycloheptamycin B (**2**), in which the β-hydroxy norvaline is replaced by a threonine. In this situation, α-cleavage of **2a** gives **1b** along with **5**, which after hydrogen transfer again results in the observed **1c** and radical cation **6**. The thus identified structural differences of **1** versus **2** at the side-chain of the hydroxylated amino acid building block is further corroborated by a number of additional product ions. Most importantly, upon loss of the functionalized tryptophan unit alone a mass difference of 14 u remains between the two product ions, while loss of the dipeptide consisting of tryptophan and the subsequent hydroxylated amino acid building block leads to a product ion with identical m/z (see ESI, Fig. S2 and S3† for full MS/MS data).

The structure of **2** as determined by MS/MS was further corroborated by analysis of the full NMR data set of the compound, clearly validating threonine to have replaced β-hydroxy norvaline in **2** versus **1** (Fig. 4; see Fig. S12 to S16 and Table S1 in ESI† for full NMR data of **2**).

Compared to other cyclodepsipeptides described in the literature, the cycloheptamycins are structurally most closely related to the cyclodepsipeptide marformycin A from Streptomyces drozdowiczii SCSIO 10141,[15] sharing the 19-membered ring system and the first three amino acid building blocks N-formyl-L-valine-L-threonine-D-OMe-tyrosine.

To evaluate the biomedical potential of cycloheptamycins A (**1**) and B (**2**), we conducted a series of antimicrobial (against Bacillus subtilis DSM 347, E. coli K12 DSM 498, Staphylococcus epidermidis DSM 20044, Propionibacterium acnes DSM 1897, Xanthomonas campestris DSM 2405, Candida albicans DSM

6598 | Org. Biomol. Chem., 2019, **17**, 6595–6600

This journal is © The Royal Society of Chemistry 2019

29
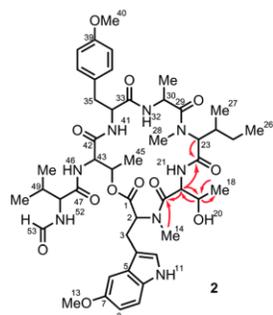
Fig. 4 Key HMBC (red arrows) NMR correlations used to validate the structure of **2**.

1386), enzyme inhibition (phosphodiesterase PDE-4B2, human acetylcholinesterase, glycogen synthase kinase-3β) and cell-based cytotoxicity (human hepatocellular liver carcinoma cell line HepG2, mouse fibroblast cell line NIH 3T3) assays (for experimental procedures, see below). Both compounds showed no enzyme inhibition or cytotoxic activity, nor antibacterial effects against *B. subtilis*, *E. coli* or *C. albicans*. When tested at a concentration of 100 μM, compounds **1** and **2** exhibited weak inhibition of *X. camestris* (31% and 22%, respectively), while *S. epidermidis* was exclusively inhibited by **2** (27%). Most interestingly, *P. acnes* was efficiently inhibited by cycloheptamycin A (**1**) with an IC$_{50}$ value of 4.22 ± 0.41 μM. Cycloheptamycin B (**2**) was considerably less active (IC$_{50}$ = 17.32 ± 2.78 μM). Tetracycline was used as a positive control in the *P. acnes* assays and reached an IC$_{50}$ value of 0.74 ± 0.11 μM. While the inhibitory activity of tetracycline is thus about 6 to 23 fold stronger when compared to **1** and **2**, the rather selective effect of the cycloheptamycins against *P. acnes* indicate that this compound class might serve as an interesting starting point for the development of treatment options against *P. acnes*. This is a significant finding as there is a strong interest in the improvement of therapies to cure diseases caused by *P. acnes*.[16] This strain is not only the causative agent of acne, a chronic inflammatory disease,[17] but also induces a range of postoperative and device-related infections.[16] Oral administration of isotretinoin (marketed as *Accutane* by Roche) is an effective treatment option,[17] but the compound exhibits a broad range of side effects, including damages to the skin, anemia, anaphylactic reactions, depression and others. The cycloheptamycin scaffold might thus serve as a platform to enhance our knowledge on *P. acnes* and to develop new lead structures for the future treatment of *P. acnes* infections.

Experimental section

General experimental procedures

Optical rotations were measured on a Rudolph Research Autopol III automatic polarimeter. NMR spectra were recorded using Bruker AV-400, AV-500 and AV-500C instruments. Chemical shift

values were referenced to the residual solvent signal of DMSO-d$_6$ (^{1}H: 2.50 ppm; ^{13}C: 39.5). HRESIMS data were recorded using a Waters Q-TOF Premier mass spectrometer.

Cultivation, extraction and compound isolation

The strain was isolated from a garden soil collected at Egerpatak, Ardeal, Romania, in 2001 using Raffinose-Histidine agar which was supplemented with cycloheximide, nalidixic acid and nystatin each at 50 μg mL^{-1}.[18] The agar plates were incubated three weeks at 27 °C. For production of cycloheptamycins, *Streptomyces* sp. Tü 6314 was cultivated in a 1 L-Erlenmeyer flask with four baffles using oatmeal medium (ISP medium 3; 250 mL for each flask) composed of 20 g oatmeal and 5 mL mineral solution in tap water (pH 7.8, adjusted with 1 N KOH).[10] The main culture was inoculated with 4 vol% of pre-culture, grown for 72 hours at 27 °C in NL-410 medium[19] in 500 mL-Erlenmeyer-flasks at 120 rpm. The fermentation was carried out at 27 °C for 144 hours on a rotary shaker at 120 rpm. The production of cycloheptamycins was monitored by reversed phase HPLC. For the isolation the pH of the main culture was adjusted to pH 5.0 with 1N HCl and the cells pelleted by centrifugation at 3800 rpm for 10 min. The supernatant was extracted three times with 250 mL ethyl acetate in a separating funnel. The combined organic phases were concentrated under reduced pressure. The resulting solid raw extract was extracted with 50 mL methanol/acetone solution in an ultrasonic bath for 15 min. The methanol extract was separated from the remaining solid and dried under reduced pressure. The solid thus obtained was dissolved in methanol, subjected to a Sephadex LH-20 column (90 × 2.5 cm) and separated using MeOH as the eluent. Final purification of the cycloheptamycins was achieved by semi-preparative reversed-phase HPLC using a stainless steel column (250 × 16 mm; Maisch) packed with 10 μm Nucleosil-100 C-18. A linear gradient of acetonitrile (ACN) and water was employed, starting with 30% ACN and reaching 70% ACN within 25 min at a flow rate of 16 mL min^{-1}. Cycloheptamycin A (**1**) eluted at 21.4 min (11.3 mg L^{-1} fermentation broth) and cyclohpetamycin B (**2**) at 16.7 min (0.7 mg L^{-1} fermentation broth). Both compounds were obtained as pale-yellow to white amorphous solids. The identical UV spectra of **1** and **2** were in agreement with those published for **1** by Godtfredsen *et al.*[11] The determined optical rotations of **1** ([α]$_d^{20}$ = +30°, *c* = 1.0, CHCl$_3$) and **2** ([α]$_d^{20}$ = +29°, *c* = 1.0, CHCl$_3$) were slightly lower than those published for **1** ([α]$_d^{20}$ = +37°, *c* = 1.0, CHCl$_3$).[11]

X-ray crystallographic analysis of compound 1

The data were collected on a Bruker D8 Venture TXS rotating anode diffractometer using monochromatized Mo Kα (λ = 0.71073 Å) radiation. The crystal was kept at 100.0(2) K during the data collection process. The structure was solved in space group *P* 1 using SHELXT and refinement was carried out using the SHELXL-2018 program. Crystallographic data were deposited at the Cambridge Crystallographic Data Centre (CCDC 1900571†). Crystal data of **1** (as solved and refined in *P* 1):

$C_{48}H_{68}N_8O_{12}$ (M = 949.50 g mol^{-1}), triclinic, P 1 (no. 1), a = 20.847(1) Å, b = 21.2683(11) Å, c = 23.8537(11) Å, α = 90.0400(16)°, β = 90.0581(14)°, γ = 89.9976(16)°, V = 10576.3(9) Å^3, Z = 8, T = 100.00(10) K, μ = 0.086 mm^{-1}, D_{calc} = 1.192 g cm^{-3}, 170 060 reflections measured (2.18° $\leq$ θ $\leq$ 21.97°), 50 951 unique (R_{int} = 0.0343, R_{sigma} = 0.0319). The final R_1 was 0.1161 ($I > 2\sigma(I)$) and wR_2 was 0.3711 (all data). For detailed discussion, see ESI.†

Biological assays

The antimicrobial activity of the cycloheptamycins **1** and **2** against the bacteria *Bacillus subtilis* DSM 347, *E. coli* K12 DSM 498, the phytopathogenic strain *Xanthomonas campestris* DSM 2405, the yeast *Candida albicans* DSM 1386, the clinically relevant strain *Staphylococcus epidermidis* DSM 20044 as well as the inhibitory activity against phosphodiesterase (PDE-4B2) and the cytotoxic activity against HepG2 (human hepatocellular liver carcinoma cell line) and NIH 3T3 (mouse fibroblasts cell line) were determined according to Schulz *et al.*[20] The determination of the acetylcholinesterase inhibitory activity was performed according to Ohlendorf *et al.*[21] Glycogen synthase kinase-3β inhibition by **1** and **2** was determined as described by Wiese *et al.*[22] In addition, the assay against *Propionibacterium acnes* DSM 1897 was carried out according the method of Schneemann *et al.*[23] The concentration of the compounds in the preliminary bioassays was 100 μM (antibiotic tests), 50 μM (cytotoxic tests), and 10 μM (enzymatic tests).

Conflicts of interest

There are no conflicts to declare.

Acknowledgements

Notes and references

1 J. Bérdy, *J. Antibiot.*, 2005, **58**, 1–26.

2 E. A. Barka, P. Vatsa, L. Sanchez, N. Gaveau-Vaillant, C. Jacquard, H.-P. Klenk, C. Clément, Y. Ouhdouch and G. P. van Wezel, *Microbiol. Mol. Biol. Rev.*, 2016, **80**, 1–43.

3 O. Genilloud, *Nat. Prod. Rep.*, 2017, **34**, 1203–1232.

4 L. Katz and R. H. Baltz, *J. Ind. Microbiol. Biotechnol.*, 2016, **43**, 155–176.

5 J. Bérdy, *J. Antibiot.*, 2012, **65**, 395–395.

6 J. Clardy, M. A. Fischbach and C. T. Walsh, *Nat. Biotechnol.*, 2006, **24**, 1541–1550.

7 F. von Nussbaum, M. Brands, B. Hinzen, S. Weigand and D. Häbich, *Angew. Chem., Int. Ed.*, 2006, **45**, 5072–5129.

8 S. Donadio, S. Maffioli, P. Monciardini, M. Sosio and D. Jabes, *J. Antibiot.*, 2010, **63**, 423–430.

9 M. Alanjary, B. Kronmiller, M. Adamek, K. Blin, T. Weber, D. Huson, B. Philmus and N. Ziemert, *Nucleic Acids Res.*, 2017, **45**, W42–W48.

10 E. B. Shirling and D. Gottlieb, *Int. J. Syst. Bacteriol.*, 1966, **16**, 313–340.

11 W. O. Godtfredsen, S. Vangedal and D. W. Thomas, *Tetrahedron*, 1970, **26**, 4931–4946.

12 P. Marfey, *Carlsberg Res. Commun.*, 1984, **49**, 591–596.

13 R. Bhushan and H. Brückner, *Amino Acids*, 2004, **27**, 231–247.

14 Crystallographic details are provided in the ESI, as is crystallographic data in cif format. CCDC 1900571 contains the supplementary data for this paper.†

15 X. Zhou, H. B. Huang, J. Li, Y. X. Song, R. W. Jiang, J. Liu, S. Zhang, Y. Hua and J. H. Ju, *Tetrahedron*, 2014, **70**, 7795–7801.

16 A. Perry and P. Lambert, *Expert Rev. Anti-Infect. Ther.*, 2011, **9**, 1149–1156.

17 H. C. Williams, R. P. Dellavalle and S. Garner, *Lancet*, 2012, **379**, 361–372.

18 E. Küster and S. T. Williams, *Nature*, 1964, **202**, 928–929.

19 M. Goodfellow and H.-P. Fiedler, *Antonie van Leeuwenhoek*, 2010, **98**, 119–142.

20 D. Schulz, P. Beese, B. Ohlendorf, A. Erhard, H. Zinecker, C. Dorador and J. F. Imhoff, *J. Antibiot.*, 2011, **64**, 763–768.

21 B. Ohlendorf, D. Schulz, A. Erhard, K. Nagel and J. F. Imhoff, *J. Nat. Prod.*, 2012, **75**, 1400–1404.

22 J. Wiese, J. F. Imhoff, T. A. M. Gulder, A. Labes and R. Schmaljohann, *Mar. Drugs*, 2016, **14**, 200.

23 I. Schneemann, I. Kajahn, B. Ohlendorf, H. Zinecker, A. Erhard, K. Nagel, J. Wiese and J. F. Imhoff, *J. Nat. Prod.*, 2010, **73**, 1309–1312.

6600 | *Org. Biomol. Chem.*, 2019, **17**, 6595–6600

This journal is © The Royal Society of Chemistry 2019

31

3.2 Discovery of the streptoketides by direct cloning and rapid heterologous expression of a cryptic PKS II gene cluster from *Streptomyces* sp. Tü6314

The following chapter is based on the publication:

In this work, we identified a cryptic type II PKS gene cluster (*skt*) from the genome data of *Streptomyces* sp. Tü6314. The *skt* cluster spans ~21 kb and consists of 19 open reading frames (ORFs) (Figure 20A). This DNA region was cloned directly from the Tü6314 genome by the LLHR method using the *Streptomyces* site-specific integration vector pSET152, followed by its rapid heterologous expression in *Streptomyces coelicolor*. HPLC analysis of the extracts from the culture broth of the heterologous hosts revealed the production of new compounds.
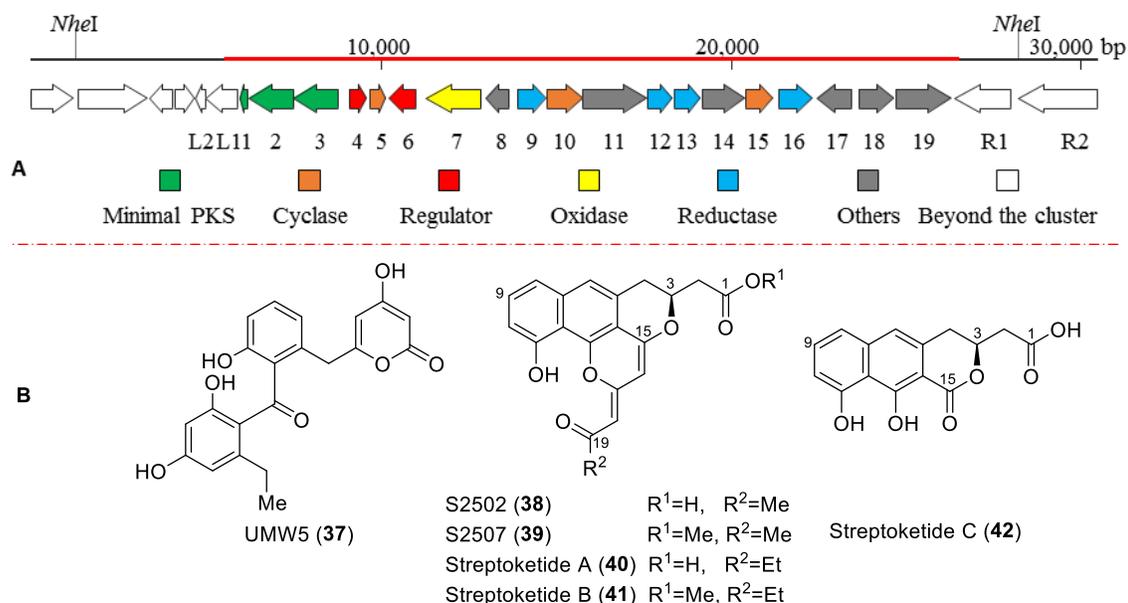
Figure 20. (**A**) Organization of the streptoketide BGC (*skt*) from *Streptomyces* sp. Tü 6314. (**B**) Compounds isolated from the *skt* heterologous expression strain.

Compound isolation and structure elucidation led to the identification of six polyketides, of which UMW5 (**37**), S2502 (**38**) and S2507 (**39**) were already known and streptoketides A-C (**40-42**) were identified as new natural products (Figure 20B). The structures of these polyketides were established by HR-MS and NMR structure elucidation. In collaboration with T. Bruhn and M.

Haslbeck, we compared the experimental ECD spectra to the corresponding calculated spectra and determined the sterostructure of these compounds as *S*-configured. In collaboration with A. Herrmann and R. Brack-Werner, we showed that four of the six polyketides exhibited anti-HIV activities, with **42** having the most pronounced effects with an EC_{50} value of 17.3 µM. In addition, no negative cellular impacts were detected during our tests, making these structures interesting starting points for further investigations.

JOC
The Journal of Organic Chemistry

Discovery of the Streptoketides by Direct Cloning and Rapid Heterologous Expression of a Cryptic PKS II Gene Cluster from Streptomyces sp. Tü6314

Zhengyi Qian, Torsten Bruhn, Paul M D'Agostino, Alexander Herrmann, Martin Haslbeck, Noemi Antal, Hans-Peter Fiedler, Ruth Brack-Werner, and Tobias A. M. Gulder

Just Accepted

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Discovery of the Streptoketides by Direct Cloning and Rapid Heterologous Expression of a

Cryptic PKS II Gene Cluster from *Streptomyces* sp. Tü6314

Zhengyi Qian,[a] Torsten Bruhn,[b] Paul M. D'Agostino,[a,c] Alexander Herrmann,[d] Martin Haslbeck,[e]

Noémi Antal,[f] Hans-Peter Fiedler,[f] Ruth Brack-Werner,[d] and Tobias A. M. Gulder[a,c*]

[a]Biosystems Chemistry, Department of Chemistry and Center for Integrated Protein Science Munich (CIPSM), Technical University of Munich, Lichtenbergstraße 4, 85748 Garching bei München, Germany.

[b]Bundesinstitut für Risikobewertung, Max-Dohrn-Str. 8-10, 10789 Berlin, Germany.

[c]Technische Universität Dresden, Chair of Technical Biochemistry, Bergstraße 66, 01602 Dresden, Germany.

[d]Helmholtz Zentrum München, German Research Center for Environmental Health, Institute of Virology, Ingolstädter Landstraße 1, 85764 Neuherberg, Germany.

[e]Department of Chemistry, Technical University of Munich, Lichtenbergstraße 4, 85748 Garching bei München, Germany.

[f]Institute of Microbiology, University of Tübingen, Auf der Morgenstelle 28, D-72076 Tübingen, Germany

*Correspondence: tobias.gulder@ch.tum.de, tobias.gulder@tu-dresden.de

1

35

ABSTRACT

Genome sequencing and bioinformatic analysis have identified numerous cryptic gene clusters that have the potential to produce novel natural products. Within this work, we identified a cryptic type II PKS gene cluster (*skt*) from *Streptomyces* sp. Tü6314. Facilitated by linear plus linear homologous recombination-mediated recombineering (LLHR), we directly cloned the *skt* gene cluster using the *Streptomyces* site-specific integration vector pSET152. The direct cloning allowed for rapid heterologous expression in *Streptomyces coelicolor*, leading to the identification and structural characterization of six polyketides (three known compounds and the new streptoketides), four of which exhibiting anti-HIV activities. Our study shows that the pSET152 vector can be directly used for LLHR, expanding the Rec/ET direct cloning toolbox and providing the possibility for rapid heterologous expression of gene clusters from *Streptomyces*.

2

36

Introduction

Natural products have played a highly significant role in the drug discovery and development process over the last decades.[1] The important drugs penicillin (anti-biotic), avermectins (anti-helmintic and insecticidal) and artemisinin (anti-malaria) are instructive examples of natural products heavily used in current medical applications. Out of all the known microbial natural product producers, actinomycetes produce over half of the antibiotics that exhibit selective biological activities against pathogenic bacteria and fungi, and about 75% of these are produced by *Streptomyces*.[2] However, after the Golden Age of natural product drug discovery in the middle of the 20th century, the late 20th century has seen a sharp decline in such discovery programs from pharmaceutical companies, in part because of the advances in both high throughput screening (HTS) and combinatorial synthesis.[3] The high rediscovery rate of known molecules and the often low quantities of compounds isolated from native producers have further promoted this decrease.[4] Whole-genome sequencing has shown that many microbes have far greater potential to produce secondary metabolites when compared to the chemical diversity already identified.[5-6] Bioinformatic methods and computational tools have been developed to screen this new wealth of genomic data for the identification of biosynthetic gene clusters (BGCs) in a process known as genome mining.[7-9] Genome-based structure prediction then paves the way for the directed discovery of natural products that are likely to yield yet unidentified chemical scaffolds. However, many BGCs are not sufficiently expressed under standard laboratory culture conditions (so-called silent/cryptic pathways) in their natural hosts to allow compound detection, making the discovery of their encoded metabolites a difficult task.[4, 10] Heterologous expression of BGCs is an emerging method in genome mining to overcome this limitation. If successful, metabolites can easily be tracked when a heterologous BGC is expressed in a background-clear host with significantly

3

simplified target identification, for example by comparative HPLC-MS analysis.[4, 10] *Streptomyces coelicolor* M1152 and *Streptomyces coelicolor* M1154 have specifically been engineered for such secondary metabolites production and are commonly used as heterologous expression hosts for the discovery of novel natural products.[11]

A precondition for the heterologous expression of BGCs is the ability to intercept the respective DNA fragment from the genomic DNA of the natural host. Traditional heterologous expression relied on construction of genomic DNA large-insert clone libraries to screen for colonies harbouring the desired BGCs. Generating and screening such libraries is time-consuming and it is often difficult to find a colony that harbors the entire BGC, especially for gene clusters >40 kb.[4] Recently, new cloning systems have been developed that bypass genomic DNA library construction, thereby largely improving the efficiency of BGCs interception. Important methods include RecET-mediated linear-plus-linear homologous recombination (LLHR), transformation-associated recombination (TAR) and Direct Pathway Cloning (DiPaC).[12-15] LLHR in *E. coli* is suitable for capturing BGCs up to 52 kb in size from *Photorhabdus luminescens*.[12] This method was also used to clone the 106 kb salinomycin gene cluster from *Streptomyces albus* from multiple fragments with downstream whole construct assembly.[16] Extension of the methodology by combination with exonuclease *in vitro* assembly (ExoCET) even facilitated capture of the complete 106 kb gene cluster in a single step.[17] However, the vectors used in LLHR need to be modified by Redαβ recombineering to make them suitable for heterologous expression.[18] Generating a vector system that allows omitting this recombineering step thus has the potential to further streamline LLHR. The pSET152 vector is an *E. coli-Streptomyces* shuttle vector with φC31 integration system, enabling its maintenance in *E. coli* and site-specific integration into the *Streptomyces* genome.[18-19] In principle, the pSET152 vector should be applicable for direct cloning

4

of BGCs from *Streptomyces* genomic DNA for downstream heterologous expression within another *Streptomyces* host.

Type II polyketide synthases (PKSs) produce structurally diverse aromatic metabolites that often possess important biological activities, such as the antibiotic tetracyclines and anticancer drug doxorubicin.[20-22] Type II PKS BGCs have a minimal PKS core consisting of three proteins: ketosynthase alpha (KS$_\alpha$); ketosynthase beta (KS$_\beta$, or chain length factor); and acyl carrier protein (ACP).[23] The minimal PKS is responsible for assembly of the nascent polyketide chain, which is then cyclized to form the aromatic core structure. Complex chemical modifications including oxidation, reduction, methylation and/or glycosylation result in a broad array of structural complexity.[23] Herein, we report the one step capture of a cryptic type II PKS gene cluster from *Streptomyces* sp. Tü6314 by LLHR using the vector pSET152 and its heterologous expression in the host *Streptomyces coelicolor*, leading to the successful recombinant production of six aromatic polyketides, including three new compounds termed streptoketides.

RESULTS AND DISCUSSION

Identification and bioinformatic analysis of the *skt* gene cluster from *Streptomyces* sp. Tü6314

The complete genome of *Streptomyces* sp. Tü6314 was sequenced by PacBio sequencing technology. This resulted in a large (7.76 Mbp) and a small (12.4 kbp) contig, corresponding to a genome size of approx. 7.8 Mbp with a GC content of 71%. Bioinformatic analysis of the genome data using antiSMASH 4.0[24] revealed a cryptic type II PKS gene cluster (*skt*) with ≤ 28% of genes showing similarity to previously characterized pathways in the MIBiG (Minimum Information about a Biosynthetic Gene cluster) database (Figure 1A).[25] The sequence of the *skt* gene cluster was annotated and submitted to GenBank and can be accessed using accession number

5

MK424349. Detailed BLASTp analysis showed that the *skt* gene cluster has many genes encoding proteins with high homology to oxytetracycline[20] and SF2575[26] biosynthesis (Table 1).

Table 1. Deduced functions of ORFs in the *skt* gene cluster.

ORF	Size[a]	Predicted function	*Streptomyces rimosus* homolog[b]	Accession number	SF2575 homolog[b]	Accession number
L2	77		no hit	-	-	-
L1	300		NmrA/HSCARG family protein (32/45%)	WP_030635220.1	-	-
skt1	81	ACP	OxyC (55/75%)	AAZ78327.1	SsfC (57/77%)	ADE34520.1
skt2	422	KSβ/CLF	OxyB (58/69%)	AAZ78326.1	SsfB (55/67%)	ADE34519.1
skt3	423	KSα	OxyA (66/79%)	AAZ78325.1	SsfA (65/79%)	ADE34518.1
skt4	167	regulator	MarR family transcriptional regulator (30/47%)	WP_030371052.1	-	
skt5	151	cyclase	OxyI (54/67%)	AAZ78332.2	SsfY4 (53/65%)	ADE34486.1
skt6	267	regulator	OtcR (42/61%)	AJO26937.1	SsfT1 (41/61%)	ADE34517.1
skt7	518	oxidase	GMC family oxidoreductase (35/50%)	WP_030663369.1	-	
skt8	224	methyl-transferase	SAM-dependent methyltransferase (54/69%)	WP_030682298.1	-	
skt9	267	ketoacyl reductase	OxyJ (66/78%)	AAZ78333.1	SsfU (70/79%)	ADE34491.1
skt10	347	Aromatase / Cyclase	OxyK (44/58%)	AAZ78334.2	SsfY1 (46/61%)	ADE34490.1
skt11 *	615	CoA ligase	OxyH?		SsfL2 (45/58%)	ADE34493.1
skt12	246	reductase	SDR family oxidoreductase (35/46%)	WP_030668664.1	SsfU (35/47%)	ADE34491.1
skt13	263	reductase	OxyM (46/59%)	AAZ78336.1	SsfU (28/42%)	ADE34491.1
skt14	417	mono-oxygenase	OxyE (56/68%)	AAZ78329.1	SsfO2 (26/40%)	ADE34483.1
skt15	257	cyclase	OxyN (61/71%)	AAZ78337.1	-	
skt16	328	reductase	aldo/keto reductase (50/65%)	KOT97719.1	SsfF (60/72%)	ADE34525.1
skt17	343	KSIII	3-oxoacyl-ACP synthase (57/68%)	WP_050512192.1	SsfG (35/49%)	ADE34509.1
skt18	333	acyl-transferase	acyltransferase domain-containing protein (59/69%)	WP_030372723.1	SsfV (48/58%)	ADE34485.1
skt19	536	carboxylase	methylmalonyl-CoA carboxyltransferase (81/88%)	GCD42787.1	SsfE (78/84%)	ADE34513.1
R1	556		Membrane protein (66/81%)	KEF22117.1	-	
R2	741		Elongation factor G (75/85%)	GCD47180.1	-	

a. Size in number of amino acids. b. *Streptomyces rimosus* is the producer of oxytetracycline, *Streptomyces* sp. SF2575 is the producer of SF2575. Homolog shows protein sequences identity/similarity%. * OxyH is not correctly annotated in the database.

6

40

It has previously been shown that both KS_α and KS_β proteins will group phylogenetically into clades that correlate well with the chain length and initial cyclization pattern of the polyketide precursor produced by a Type II PKS.[27-28] Here, a phylogenetic tree was constructed to compare the *skt* KS_α/KS_β (Skt3/Skt2) with the KS_αs/KS_βs from functionally characterized type II gene clusters.[28] The *skt* KS_α (Skt3) and KS_β (Skt2) grouped well into parallel clades, with a proposed chain length of 21 carbons (Figure S1). Attempts to detect and isolate the respective PKS II products by fermentation of *Streptomyces* sp. Tü6314 were unsuccessful (>10 L fermentation, data not shown). This suggested that the *skt* gene cluster was silent or expressed at low levels in the native host under the used culture conditions. We therefore aimed to activate and express the *skt* BGC in a suitable recombinant host system.

Figure 1. (A) Gene organization of the *skt* gene cluster. The red line indicates the part of the cloned *skt* cluster and was used for heterologous expression within this study. *NheI* was used to excise the cluster from genomic DNA. (B) General workflow for the direct cloning of *skt* by LLHR. pSET152 was used as the expression vector and integrated into the *S. coelicolor* genome for heterologous expression. (C) Left: simulated restriction enzyme analysis of the captured cluster by pSET152 digested with *BlpI* (1), *PstI* (2) and *PvuII* (3). Right: authentic restriction enzyme analysis of the captured cluster digested with above restriction enzymes. For details on DNA fragment sizes, please see Figure S4 and Table S3. (D) HPLC-UV analysis of culture extracts from *S. coelicolor*

8

42

transformed with emtpy pSET152 and pSET152-*skt*. Putative signals of compounds **3a** (t_R approx. 19.3 min) and 3b (t_R approx. 20.4 min) are labeled with *.

LLHR capture and heterologous expression of the *skt* gene cluster in *S. coelicolor*

For heterologous expression, we selected *S. coelicolor* M1152 and M1154 as hosts. Both M1152 and M1154 are genetically engineered overexpression hosts with a reduced metabolic background.[11] M1152 and M1154 have the ability to produce 20-40 times higher titers of recombinant products when compared to their parental strain M145.[11] They are thus ideally suited for the recombinant production of natural products from other *Streptomyces*. LLHR is a powerful tool to capture BGCs from a broad range of diverse microbes.[12, 16, 18, 29] In this study, we used the pSET152 vector to directly capture the 21.7 kb *skt* BGC via LLHR (Figure 1B). After LLHR, subsequent colony screening PCR resulted in one positive colony out of twelve (Figure S2). Restriction digest analysis showed that the desired *skt* gene cluster was successfully cloned into pSET152 (Figure 1C, Figures S3-S4). The subsequently isolated pSET152-*skt* plasmid was directly introduced into *S. coelicolor* M1152 and M1154 via conjugation for heterologous expression without further modification. The exconjugants with *skt* integrated in the M1154 genome developed strongly red colonies, while the engineered M1152 strain was only slightly red and the strain harboring the empty pSET152 vector developed white to pale grey colonies. Comparative analysis of the metabolite profiles showed that additional compounds were produced in both M1152/*skt* and M1154/*skt* when compared to the empty pSET152 control. The production titers of M1154/*skt* were significantly higher than those of M1152/*skt* (Figure 1D). Consequently, M1154/*skt* was used for downstream compounds isolation.

9

Isolation and structural elucidation of the natural products

For compound production, strain M1154/*skt* was cultivated in ISP Medium 4. The combined organic extracts of the supernatant and the cell pellet were fractionated by Sephadex LH-20 size exclusion chromatography. Fractions containing compounds absent in the control strain M1154/pSET152 were further purified by semi-preparative HPLC on C18 material. This yielded a total of 6 pure compounds exclusively present in M1154/*skt* (Figure 2). Compound **1** (3.6 mg, HPLC retention time at 14.2 min) possessed a molecular mass of 383.1122 units (Figures S5-S6) which best fits a molecular formula of $C_{21}H_{18}O_7$ in its protonated form (calcd. 383.1125 u), leading to a calculated 13 degrees of unsaturation. Literature search with this and the ^{1}H and ^{13}C NMR data revealed this compound to be UWM5 (**1**), a well-known polyketide shunt product (Table S4, Figures S7-S9). Compound **1** was first isolated from the doxorubicin producing strain with a gene encoding a crucial cyclase disrupted.[30] Compound **2a** (0.7 mg, 16.8 min) had a molecular mass of 353.1017 u (Figures S10-S11) corresponding to a molecular formula of $C_{20}H_{16}O_6$ in its protonated form (calcd. 353.1020 u), likewise possessing 13 degrees of unsaturation. The molecular mass of **3a** (2.8 mg, 19.3 min) was 367.1174 u (Figure S15-S16), resulting in a molecular composition of $C_{21}H_{18}O_6$ in its protonated form, again containing 13 degrees of unsaturation. The observed mass difference of 14 units between **2a** and **3a** therefore suggested **3a** to be a methylated analog (calcd. 367.1176 u) of **2a**. This was also consistent with the differences in the ^{1}H and ^{13}C NMR spectra of **2a** versus **3a**, differing in the additional presence of a methyl ester in **3a** (Tables S5-S8, Figures S12-S14 and S17-S19). Using this combined data, **2a** and **3a** were readily identified as S2502 and S2507, respectively, two molecules previously obtained by heterologous expression of the nogalamycin anthraquinone aglycone genes in *Streptomyces lividans* TK24.[31]

44

Two additional natural products that seemed to be highly related to **2a/3a** were isolated: compounds **2b/3b**. Compound **2b** (2.6 mg, 17.9 min) possessed an identical molecular formula of $C_{21}H_{18}O_6$ (MS data of the protonated form 367.1174 u, calcd. 367.1176 u) when compared to **3a** (Figures S20-S21). The significantly reduced retention time under identical RP-HPLC isolation conditions (Δt_R = 1.4 min) suggested **2b** to bear a free acid function, with the additional methyl group likely being present in form of a phenolic methyl ester. However, inspection of the ^{1}H and ^{13}C NMR data showed that no methyl ester was present in **2b** (Figure S24). Instead, the compound contained an additional methylene unit. ^{1}H NMR data showed this methylene group to be directly connected to an electron-withdrawing group and a methyl function, based on chemical shifts and signal multiplicity (2H, q with J = 7.4 Hz at 2.35 ppm and 3H, t, with J = 7.4 Hz at 1.02 ppm, also observable in **3b**) (Table S9). 2D NMR (COSY, HMBC) clearly proved the methylene unit to be inserted at the terminal ketone C19 (Figures S25-S26), leading to an ethyl ketone in **2b** versus the methyl ketone in **2a/3a** (Figure 2, box). This suggests the PKS system to be alternatively primed with a propionyl-CoA starter unit (see discussion below), leading to the observed new polyektide named streptoketide A (**2b**). Compound **3b** (6.9 mg, 20.4 min) possessed a molecular mass of 381.1329 u (Figures S22-S23) which is thus 14 u higher than that of **2b**, again suggesting the presence of an additional methyl group (calcd. 381.1333). With the differences in retention times between **2a** and **3a** being identical to those of **2b** and **3b** (Δt_R = 2.5 min), the presence of a methyl ester in **3b** was most likely. This assumption was corroborated by its NMR data (Table S9, Figures S27-S29), revealing **3b** to likewise be a new natural product, which was termed streptoketide B (**3b**).

11

45
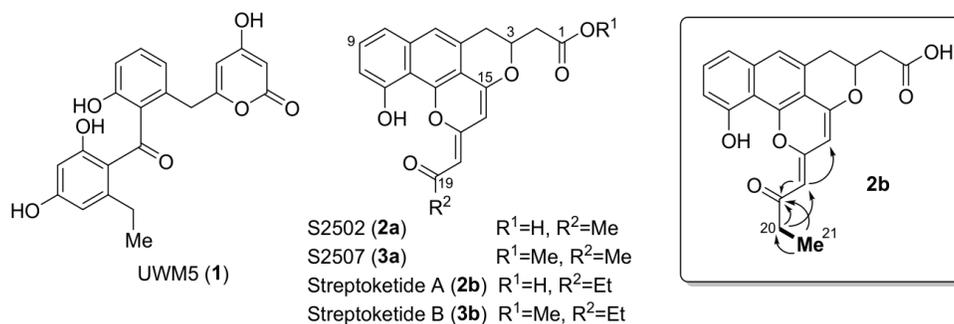
Figure 2. Structures of the known isolated compounds UWM5 (**1**), S2502 (**2a**), S2507 (**3a**) and of the new natural products streptoketide A (**2b**) and B (**3b**). Box: Selection of key COSY (bold bond) and HMBC (arrows) correlations used to elucidate the position of the ethyl substitution, exemplarily shown for **2b**.

Compound **4** (3.6 mg, 16.6. min) had a molecular mass of 287.0562 u in ESI negative mode, corresponding to a composition of $C_{15}H_{12}O_6$ in its deprotonated form (calcd. 287.0561 u) with 10 degrees of unsaturation (Figures S30-S31). The ^1H and ^{13}C chemical shifts and ^1H signal multiplicities were mostly highly similar to those of C1 to C15 of **2a/b** and **3a/b**, with largest chemical shift differences at C13 (e.g. 149.9 in **2b** versus 162.2 in **4**) and C15 (e.g. 157.4 in **2b** versus 169.7 in **4**). The typical singlet ^1H NMR signals of C16 and C18 of **2a/b** and **3a/b** were not present in the spectrum of **4** (Table 2). This suggested atoms C16 to C20/C21 to be absent in **4**. In addition, the retention time of **4** was similar to that of the free acids **2a** and **3a**, pointing at C1 to be a carboxylic acid function. The increase in the chemical shifts at C13 and C15 suggested C15 to be a cyclic ester function. Evaluation of the HMBC NMR data indeed corroborated these assumptions (Figures S32-S33), leading to the overall structure of the new polyketide streptoketide C (**4**) depicted in Figure 3.

Table 2. NMR spectroscopic data of **4** recorded in DMSO-d_6 at 500 MHz (^1H) and 125 MHz (^{13}C).

12

46

Position	Group	^{1}H (mult., J [Hz])	^{13}C	HMBC
1	COOH	13.04 (bs)	171.0	
2	CH$_2$	2.85 (dd, 16.3, 5.1) 2.77 (dd, 16.3, 7.8)	38.7*	1, 3, 4
3	CH	4.99 (m)	76.2	-
4	CH$_2$	3.14 (dd, 16.6, 3.3) 3.08 (dd, 16.6, 10.9)	31.8	2, 3, 5, 6, 14
5	C		133.1	
6	CH	7.16 (s)	116.3	4, 7, 8, 12, 13, 14, 15
7	C		138.9	
8	CH	7.25 (d, 7.9)	118.2	6, 7, 9, 10, 11, 13
9	CH	7.49 (t, 7.9)	131.7	7, 8, 10, 11, 12
10	CH	6.84 (d, 7.9)	110.7	8, 11, 12
11	COH	10.06 (bs)	156.5	
12	CH		112.8	
13	COH	12.57 (bs)	162.2	
14	CH		101.3	
15	CO$_2$R		169.7	

*: extracted from HMBC due to signal overlap with DMSO.
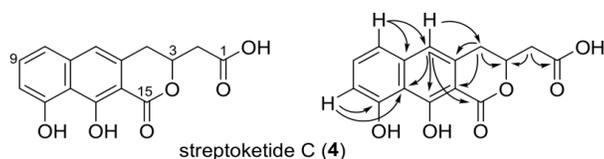
Figure 3. Left: structure of streptoketide C (**4**). Right: selected key HMBC correlations (arrows) used to assemble the structure of **4**.

Compounds **2-4** all bear a stereocenter at C3 with unknown configuration, also for the literature known **2a/3a**. The method of choice to elucidate the configuration at this position was the comparison of experimental ECD spectra to the corresponding calculated spectra for both possible enantiomers. These investigations were initiated with streptoketide C (**4**) with the smallest molecular structure. A comparison of the experimental ECD spectrum of **4** with that calculated for its *S*-configured enantiomer using CAM-B3LYP/def2-TZVP[32-33] revealed a satisfying match between the two curves (Figure 4, left). It has to be kept in mind that the ECD effect of **4** is comparably small and that it is a carboxylic acid. This means that the ECD can be highly influenced

13

47

by the solvent and pH. It is thus not surprising that especially the calculated rotational strengths of

the first excited states do not fully fit the experiment. Nonetheless, the more pronounced ECD in

the smaller wavelength region are very well reproduced and thus allow the robust determination

of the absolute configuration of the stereocenter of **4** as *S*.

Above results were further corroborated by also investigating the configuration of S2507 (**3a**)

following the identical approach. Comparison of the experimental ECD spectrum with that

calculated for the corresponding *S*-enantiomer using CAM-B3LYP/def2-TZVP calculated again

revealed *S* configuration at C3 of **3a**. Owing to the close biosynthetic relationship of **2-4**, the

identical configuration can be conveyed to all other compounds reported on this study.
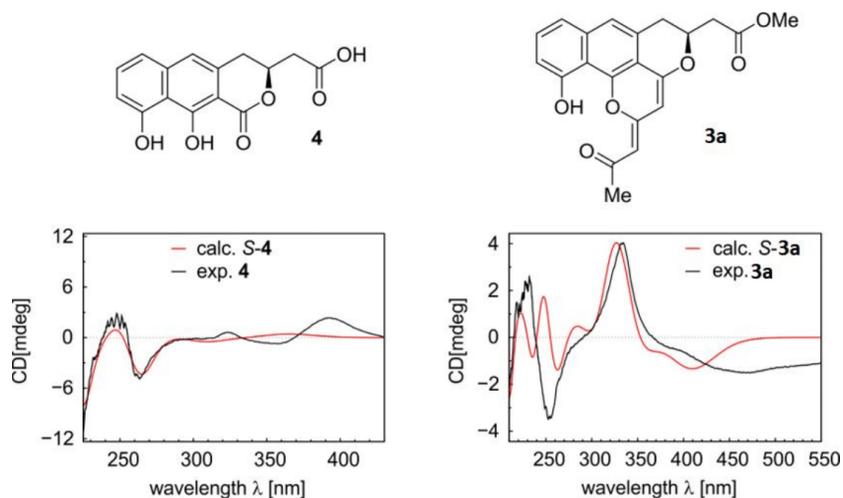
Figure 4. Comparison of experimental and calculated ECD spectra of **4** and **3a** revealing C3
configuration to be *S*.

Proposed *skt* biosynthetic pathway

The *skt* gene cluster encodes many proteins homologous to the oxytetracycline (*oxy*) and SF2575

(*ssf*) gene clusters, including genes for the minimal PKSs: KS$_\alpha$ (*skt3*), KS$_\beta$ (*skt2*) and ACP (*skt1*);

14

48

the cyclases (*skt5*, *skt10*, *skt15*), the C-9 reductase (*skt9*) and other reductases or oxygenases (Table 1).[20, 26] However, the gene encoding the amidotransferase OxyD in the biosynthetic pathways of tetracyclines is not present. The amidotransferase is responsible for producing the malonamyl starter unit unique of the tetracyclines. In the absence of OxyD, the *oxy* PKS is initiated by an acetate primer.[20] The absence of this gene in *skt* is thus consistent with the structures of all products isolated in this study, which all derive from either acetyl- or propionyl-CoA starter untis.[34] The ability to incorporate both C2 and C3 starter units is well studied in the daunorubicin biosynthetic pathway, in which a PKS III type enzyme DpsC confers starter unit fidelity and the acyltransferase DpsD orchestrates propionate selection.[35-38] Encoded within the *skt* gene cluster, Skt17 indeed shows high similarity to DpsC (58% sequence identity) and Skt18 has high similarity to DpsD (58% sequence identity). Skt18 possesses 48% identity to the acyltransferase SsfV, which is proposed to be involved in initiation of SF2575 biosynthesis with a malonamyl starter unit.[26] Combining the sequence identity and structure, it is thus likely that Skt17 and Skt18 are responsible for priming the PKS system with a propionyl-CoA starter unit to yield the **1/2b/3b** backbone (21-carbon). The occurrence of **2a/3a** can thus be explained by a relaxed substrate selectivity of Skt17/Skt18 enabling alternative initiation with an acetyl-CoA starter unit, resulting in a 20-carbon chain product. This is also consistent with the isolation of UWM5 (**1**, 21-carbon) and SEK43 (20-carbon) from the doxorubicin pathway engineered strains.[34] Skt19 is predicted to be a methylmalonyl-CoA carboxyltransferase, with a possible role of providing propionyl-CoA from methylmalonyl-CoA. Compounds **2a** and **3a** were also previously isolated from an engineered *S. lividans* TK24 by combination of anthracycline and actinorhodin biosynthetic genes.[31, 39] Total synthesis of **2a/3a** indicated that the second pyran ring can be formed easily by a non-enzymatic cyclization/dehydration.[40] Based on this analysis, we propose the following overall

15

biosynthetic pathway to the streptoketides (Figure 5). The *skt* minimal PKSs (Skt1, Skt2, Skt3) together with Skt17, Skt18, Skt19 initially affords the nascent polyketide chain **5a/b**. Then, Skt9 - which is 66% and 70% identical to the reductases OxyJ and SsfU, respectively - regioselectively reduces the backbone at C-9 to yield **6a/b**. Initial cyclization to **7a/b** is likely directed by Skt10, which shares high similarity to OxyK that catalyzes the first cyclization in oxytetracyclin biosynthesis.[20] **7b** can undergo spontaneous non-enzymatic cyclization to **1**. Controlled cyclization and reduction chemistry catalyzed by cyclases Skt5/Skt15 and reductases Skt12/Skt13/Skt16, respectively, paves the way to **8a/b** and **9a/b**. Hemiacetal formation in **9a/b** delivers intermediate **10a/b**, which upon oxidative *C,C*-bond cleavage yields streptoketide C (**4**). Alternatively, dehydration gives **11a/b**, which upon cyclization/dehydration yield **2a/2b**. The latter can be further transformed to **3a/3b**. The BGC encodes a putative methyl transferase, Skt8, which could be involved in this transformation. However, only very limited amounts of **3a**/**3b** can be detected in the raw extract of the recombinant production strain (cf. Figure 1D). In addition, these compounds can be formed from **2a**/**2b** upon dissolving in MeOH during isolation(cf. Figure S34). As the raw extract was dissolved in MeOH and fractionated on Sephadex using MeOH as the eluent, the majority of the isolated **3a**/**3b** is hence likely formed by non-enzymatic methyl ester formation.
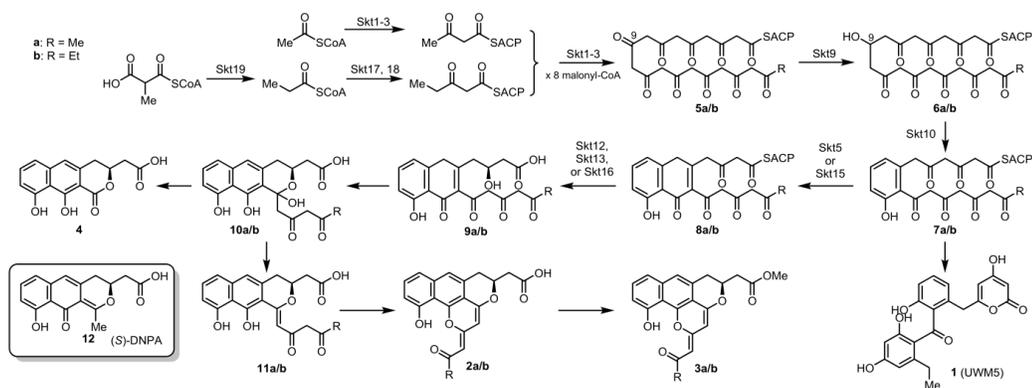
50

Figure 5. Proposed biosynthetic pathway leading to the isolated polyketides **1-4**.

Anti-viral tests of compounds 1-4

As the previously known compounds also isolated within this study, S2502 (**2a**) and S2507 (**3a**) were reported to have biological activity against adeno-, cytomegalo-, herpes simplex and influenza B viruses at 1 µM test concentration,[40] we were interested in evaluating the inhibitory potential of **1-4** against the human immunodeficiency virus type 1 (HIV-1). Therefore, we performed an EASY-HIT assay[41] using the reporter cell line LC5-RIC and a wild-type HIV-1$_{LAI}$ strain. The LC5-RIC cell line contains a reporter construct encoding the fluorochrome DsRed1 that is expressed upon HIV infection. Antiviral activity of compounds is determined by measuring the intensities of fluorescent signals of HIV-inoculated cultures.[41] We treated HIV inoculated cells with serial dilutions of **1-4** up to a concentration of 50 µM and checked for effects on cell viability conducting a CellTiter Blue® Assay.[42] **2a**, **3a**, **3b** and **4** showed anti-HIV activity, with **4** as the most potent compound, inhibiting HIV infection at a 50% effective concentration (EC$_{50}$) of 17.3 µM. Neither **1** nor **2b** inhibited HIV infection at concentrations up to 50 µM (Figure 6A). There were no signs of negative cellular impacts during those assays (Figure 6B). Both **2a** and **3a** have anti-viral activities against HIV, but these are weaker than the antiviral effects reported for these compounds against other viruses.[40] Parallel testing of emtricitabine in the EASY-HIT assay yielded an EC$_{50}$ value of 0.7 µM. The difference between the antiviral activities of compounds **2a**, **3a**, **3b**, **4** and this clinical drug is not surprising, since emtricitabine is the result of multiple rounds of optimization.[43]
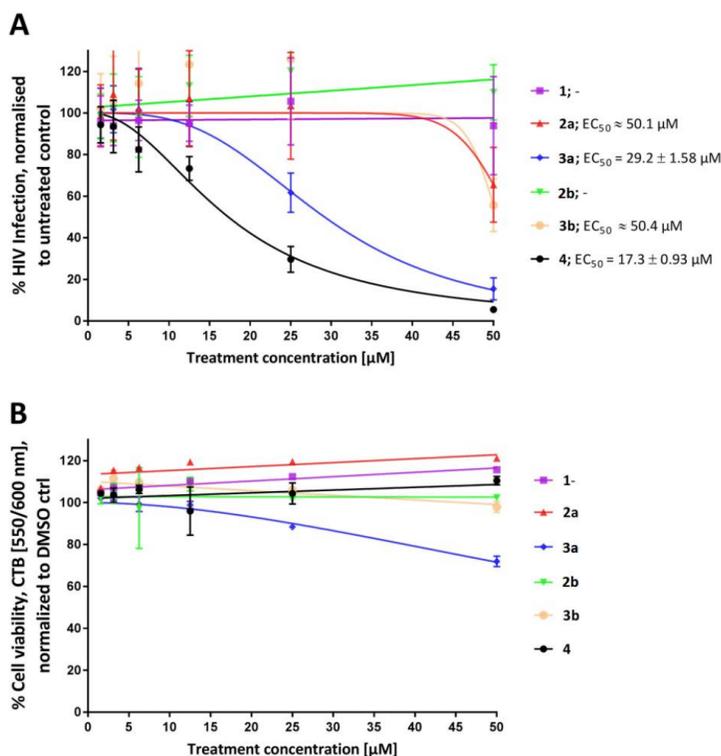
Figure 6. Inhibitory activities of **1-4** against HIV-1 infection. HIV inoculated cells were treated with serial dilutions of **1-4** up to a concentration of 50 µM and checked for effects on (**A**) anti-HIV activity and (**B**) cell viability. Compound effects were evaluated in LC5-RIC cells exposed to HIV-1_{LAI}. Shown are the means and standard deviations (±SD) of three independent experiments with triplicates (n=3; m=3).

In conclusion, we have directly captured the type II PKS gene cluster *skt* from *Streptomyces* sp. Tü6314 using the pSET152 vector using the LLHR method. After the rapid heterologous expression in *Streptomyces coelicolor*, we isolated and characterized three known (**1**, **2a**, **3a**) and three new natural products streptoketides A-C (**2b**, **3b**, **4**) and determined the absolute configurations of **2-4** for the first time. Compounds **2a**, **3a**, **3b** and **4** showed anti-HIV activity,

18

52

with **4** having the most pronounced effects with an EC_{50} value of 17.3 µM. While this activity is >10 fold weaker than that of the established antiviral drug emtricitabine, there have been no deleterious effects of the compounds observed within our test system, making their structures interesting starting points for further investigations.

Interestingly, during the course of this work, Liu et al. reported a strategy of direct cloning and heterologous expression of natural product biosynthetic gene cluster in *Bacillus subtilis* via Red/ET recombineering.[29] Both their work and ours share the same idea of using a strain specific vector to bypass the requirement of further genetic modifications after LLHR. In this way, heterologous expression of BGCs in a host from the same or closely related species can be simplified.[29] Our study thus contributes to the application of heterologous expression techniques for the efficient production of new natural products encoded by cryptic biosynthetic gene cluster.[4,10]

EXPERIMENTAL SECTION

Materials and Methods

Strain *Streptomyces* sp. Tü6314

Streptomyces sp. Tü6314 was isolated from a garden soil collected at Egerpatak, Romania. The bacterium was examined for a number of key properties known to be of value in streptomycete systematics. The presence of LL-diaminopimelic acid in the peptidoglycan together with its colonial characteristics and partial sequencing of the 16S rRNA gene allowed its assignment to the genus *Streptomyces*.

19

53

Bacterial strains and plasmids

Bacterial strains and plasmids that were used are listed in Table S1. All *E. coli* strains were cultured in Luria-Bertani (LB) liquid or on LB agar media at 37°C. *Streptomyces* strains were maintained on MS agar (20.0 g/L mannitol, 20.0 g/L soya flour, 20.0 g/L agar) at 30°C or cultivated in YMG liquid medium (4.0 g/L yeast extract, 10.0 g/L malt extract, 4.0 g/L glucose, adjusted to pH 7.2 using 1 M KOH) at 28°C with constant shaking at 200 rpm. Liquid ISP Medium 4 (10.0 g/L soluble starch, 1.0 g/L K_2HPO_4, 1.0 g/L $MgSO_4 \cdot 7H_2O$, 1.0 g/L NaCl, 2.0 g/L $(NH_4)_2SO_4$, 2.0 g/L $CaCO_3$, 1.0 mL trace salt solution) was used for large-scale fermentation. The trace salt solution (1.0 g/L $FeSO_4 \cdot 7H_2O$, 1.0 g/L $MnCl_2 \cdot 4H_2O$, 1 g/L $ZnSO_4 \cdot 7H_2O$) was sterile filtered. Apramycin (apr, 30 μg/mL), kanamycin (kan, 50 μg/mL) and chloramphenicol (cm, 15 μg/mL) were added to the media as required. For conjugation between *E. coli* and *Streptomyces*, apr (30 μg/mL) and nalidixic acid (NA, 50 μg/mL) were added.

Bacterial genomic DNA isolation.

The *Streptomyces* genomic DNA isolation was slightly modified from the method described by Wang et al..[18] In brief, *Streptomyces* sp. Tü6314 was cultured in 50 mL YMG medium at 28°C for three days. After centrifugation, the cells were resuspended in 4.5 mL of Solution 1 (10% sucrose [w/v], 50 mM Tris-HCl, pH 8.0, 10 mM EDTA). Then 500 μL of 30 mg/mL lysozyme were added and the mixture was incubated for 1 h at 37°C with occasional inverting. After adding 1.5 mL of 3.3% SDS and 100 μL of Proteinase K (10 mg/mL), the tube was mixed by gentle inversion and incubated at 37°C for at least 1 hour until the solution became clear. The solution was first combined with 2 mL 6 M NaCl and then 1 volume of phenol-chloroform-isoamyl alcohol (25:24:1) was added to create an emulsion. After centrifugation, the aqueous phase was carefully transferred

20

54

to a new tube using a wide-bore pipette tip. The DNA was precipitated by adding 0.1 volume of 3 M sodium acetate (pH 5.4), followed by adding 2.5 volumes of absolute ethanol and gently inverting to precipitate DNA. The DNA was dissolved in 4 mL TE buffer and 50 µL RNase A (4 mg/mL) was added. The DNA was precipitated again with ethanol and washed twice with 70% (v/v) ethanol, dried and dissolved in 500 µL 10 mM Tris-HCl, pH 8.0.

Phylogenic tree construction

The phylogenic tree was constructed using the MEGA X program based on ClustalW alignment using the Neighbor-Joining method.[44] The protein sequences were downloaded from the MIBiG database.[25]

Capture of the *skt* gene cluster using LLHR

*Nhe*I restriction sites flanking the proposed *skt* cluster were identified and chosen to release the gene cluster from the genomic DNA (Figure 1A). Approximately 50 μg of the extracted genomic DNA was digested with 50 U of *Nhe*I at 37°C overnight. The digested genomic DNA was purified by ethanol precipitation. The resulting DNA pellet was dissolved in 25 µL ddH$_2$O.

The linear capturing vector was PCR amplified from the pSET152 backbone using Q5 High-Fidelity DNA polymerase with primer pairs pSET152-cap_cluster21-F/R. These primers were designed to include 50-bp homology arms targeting the flanking region of the *skt* gene cluster, slightly inside the *Nhe*I digested position (Figure 1A-B). The online program Oligo Analyzer (http://www.bioinformatics.nl/cgi-bin/primer3plus/primer3plus.cgi) was used to find optimal primers with minimum hair pin/dimer formation. The PCR annealing temperature was estimated using the NEB Tm calculator (http://tmcalculator.neb.com/#!/main). PCR conditions were as

21

follows: 98°C 1 min; 32 cycles of 98°C 10 s, 58°C 20 s, 72°C 5 min; 72°C 10 min. After PCR amplification, *Dpn*I was used to digest the template DNA, followed by agarose gel purification.

LLHR was performed as described in the literature with several modifications.[12, 18] In brief, 300 µL of overnight cultured *E. coli* GB05-dir was inoculated into 15 mL fresh LB medium without antibiotics. The cells were grown at 37°C, 200 rpm for ~1.5 h until OD_{600} reached ~0.3. After adding 300 µL of 1 M L-(+)-arabinose, the cells were grown at 37°C at 200 rpm for 45 min until the OD_{600} reached 0.6-0.8. Then, 1.5 mL cells were transferred into 1.5 mL ice-cold Eppendorf tubes and centrifuged for 30 s at 8,000 *g* at 2°C. The supernatant was discarded and the cell pellet was washed using 1 mL ice-cold water. The washing step was repeated once. Then the cells were resuspended in 30 µL ice-cold water. A total of 5 µg of purified *Nhe*I digested *Streptomyces* sp. Tü6314 genomic DNA and ~500 ng of PCR amplified pSET152 capturing vector were added. Transformation into *E. coli* was achieved by electroporation using ice-cold cuvettes and a BIO-RAD MicroPulser (program E2 for bacteria). Electroporated cells were incubated at 37°C for 120 min while shaking and then spread on LB agar plates supplemented with 30 µg/mL apramycin. Colonies were screened by PCR with the primer pairs pSET152_cap_seq-F/cap_cluster21_verification_L-R. The primers used in this study are listed in Table S2.

Heterologous expression of the *skt* gene cluster

The captured *skt* gene cluster was transformed into the donor strain *E. coli* ET12567/pUZ8002 and then transferred to *S. coelicolor* M1152 or *S. coelicolor* M1154 by intergeneric conjugation as described previously, with minor modifications.[45] A single colony of the donor strain was inoculated into 3 mL LB supplemented with kan/cm/apr and grown at 37°C overnight. A 100 µL

22

56

overnight culture was used to inoculate 10 mL fresh LB with kan/cm/apr and grown at 37°C until

the OD_{600} reached 0.6-0.8. The grown *E. coli* cells were centrifuged and washed twice with an

equal volume of fresh LB. Spores from the host strains were heat activated at 50°C for 10 min in

1 mL YMG medium followed by cooling using tap water and then mixed with the washed donor

cells. The mating mixture was then spread on MS agar plates and incubated at 30°C. After 20 h,

the MS agar plates were overlaid with 1 mg/plate of NA and apr. Five days later, the Apr[R]

exconjugants were picked and grown on MS agar containing NA and apr, resulting in single

colonies. The empty pSET152 vector was also introduced into the host strains and used as a

control. One well-grown single colony was inoculated 100 mL YMG medium and grown at 28°C

200 rpm for 7 days before High Performance Liquid Chromatography (HPLC) analysis. The

culture was centrifuged and the supernatant was extracted with ethyl acetate. The crude extracts

were dissolved in methanol for HPLC analyses. HPLC was performed on a Jasco HPLC system

(UV-1575 Intelligent UV/VIS Detector, DG-2080-53 3-Line Degaser, two PU-1580 Intelligent

HPLC Pumps, AS-1550 Intelligent Sampler, HG-1580-32 Dynamic Mixer, Galaxie-

Chromatography-Software) with a Eurospher II 100-3 C18 A (150 × 4.6 mm) column purchased

from Knauer (Germany). The eluent system consisted of: A = H_2O + 0.05% TFA and B =

acetonitrile + 0.05% TFA. The analytical method consisted of a linear gradient: 0-2 min 5% B, 2-

30 min 5% to 95% B, 30-35 min 95% B, 35.2-38 min 5% B with a flow rate at 1 mL/min, 25°C.

Compound extraction and isolation.

For product isolation, one well-grown *S. coelicolor* M1154/*skt* single colony was inoculated 100

mL YMG medium and grown at 28°C 200 rpm for 5 days to make a seed culture. The seed culture

was used to inoculate 4 × 1 L ISP Medium-4 in four 3-L flasks to a final concentration of 1% (v/v)

23

and grown at 28°C 200 rpm for 7 days. After fermentation, the cultures were filtered using filter paper and the liquid fraction was adjusted to pH 4 using HCl. The supernatant was extracted three times with an equal volume of ethyl acetate. The cell pellet was extracted four times with 1 L acetone until the solvent became colorless. The organic extracts were combined and dried under reduced pressure to obtain the crude extracts. The crude extracts were dissolved in 50 mL methanol and subjected to a 90 × 3 cm Sephadex LH-20 column using methanol for elution. The eluted fractions were checked by HPLC and fractions containing products not present in the unmodified host strain were collected, followed by purification using a Jasco semi-preparative HPLC system (UV-1575 Intelligent UV/VIS Detector, two PU-2068 Intelligent prep Pumps, a MIKA 1000 Dynamic Mixing Chamber, 1000 μL Portmann Instruments AG Biel-Benken, a LC-NetII / ADC, a Rheodyne injection valve, Galaxie-Chromatography-Software), with a Eurospher II 100-5 C18 A (250 × 16 mm) column with precolumn (30 × 16 mm) purchased from Knauer (Germany). The eluent system consisted of: A = H_2O + 0.05% TFA and B = acetonitrile + 0.05% TFA, with a linear gradient: 0-43 min 25-50% B with a flow rate at 12 mL/min at room temperature. After preparative separation, the fractions containing the desired product were combined and the acetonitrile was removed under reduced pressure. The remaining aqueous phases were frozen in liquid nitrogen and the water was removed by lyophilization (Alpha 2-4 Christ with Chemistry-Hybrid-Pump-RC6 pump). UWM5 (**1**): light-brown amorphous solid (yield 0.9 mg/L, t_R = 14.2 min). ^{1}H and ^{13}C NMR data, see Table S4. HPLC and UV data, see Figure S5. HRMS (ESI-TOF) *m/z*: [M+H]+ calcd. for $C_{21}H_{19}O_7$ 383.1125 (see Figure S6); found 383.1122. The spectroscopic data is in agreement with the literatue.[30] S2502 (**2a**): dark-brown amorphous solid (yield 0.2 mg/L, t_R = 16.8 min). ^{1}H and ^{13}C NMR data, see Table S5 and S6. HPLC and UV data, see Figure S10. HRMS (ESI-TOF) *m/z*: [M+H]+ calcd. for $C_{20}H_{17}O_6$ 353.1020 (see Figure S11); found 353.1017. The

24

spectroscopic data is in agreement with the literatue.[31, 40] S2507 (**3a**): dark-brown amorphous solid (yield 0.7 mg/L, t_R = 19.3 min). ^{1}H and ^{13}C NMR data, see Table S5 and S6. HPLC and UV data, see Figure S15. HRMS (ESI-TOF) *m/z*: [M+H]$^+$ calcd. for $C_{21}H_{19}O_6$ 367.1176 (see Figure S16); found 353.1174. The spectroscopic data is in agreement with the literatue. [31, 40] Streptoketide A (**2b**): dark-brown amorphous solid (yield 0.7 mg/L, t_R = 17.9 min). ^{1}H and ^{13}C NMR data, see Table S9. HPLC and UV data, see Figure S20. HRMS (ESI-TOF) *m/z*: [M+H]$^+$ calcd. for $C_{21}H_{19}O_6$ 367.1176 (see Figure S21); found 353.1174. Streptoketide B (**3b**): dark-brown white amorphous solid (yield 1.7 mg/L, t_R = 20.4 min). ^{1}H and ^{13}C NMR data, see Table S9. HPLC and UV data, see Figure S22. HRMS (ESI-TOF) *m/z*: [M+H]$^+$ calcd. for $C_{22}H_{21}O_6$ 381.1333 (see Figure S23); found 353.1329. Streptoketide C (**4**): pale beige color white amorphous solid (yield 0.9 mg/L, t_R = 16.6 min). ^{1}H and ^{13}C NMR data, see Table 2. HPLC and UV data, see Figure S30. HRMS (ESI-TOF) *m/z*: [M-H]$^-$ calcd. for $C_{15}H_{11}O_6$ 287.0562 (see Figure S31); found 287.0561.

NMR data collection and CD spectra measurement

For NMR data collection, all purified compounds were dissolved in deuterated DMSO. The NMR spectra were recorded on a Bruker AVHD500 or a Bruker AV500-cryo spectrometer. The chemical shifts δ are listed as parts per million [ppm] and refer to δ(TMS) = 0. The spectra were calibrated using residual undeuterated solvent as internal reference.

CD spectra were measured using a Jasco J-715 spectropolarimeter (Jasco, Gross-Umstadt, Germany). Experiments were performed in quartz cuvettes with 0.1 cm path length at a substance concentration of 0.1 mg/ml and spectra were recorded from 200 to 600 nm in acetonitrile at 22°C. Eight to twelve spectra were accumulated for each substance and subsequently baseline corrected by subtracting the pure solvent spectrum.

25

Calculation of ECD spectra

The ECD spectra of **4** and **2a** were computed[46] using ORCA[47] for the optimizations and frequency calculations and Gaussian09[48] for the excited states investigations. The conformational analysis was simplified in terms of the rotational freedom of the alkyl side chains. As it is known that these do have a negligible effect on the ECD only the lowest conformation of the side chain was considered for the conformational analysis. More important is the orientation of the side chain relative to the ring (equatorial or axial). The ECD spectra of **4** and **3a** in the equatorial conformation do have mirror image like spectra when compared to the axial one. B3LYP-D3/def2-TZVP[49] calculations including the chain-of-spheres approximation yield very similar energies for both conformers of **4**, while that of **3a** do have a strong difference. According to Gibbs energies **4**-eq and **4**-ax differ only by 0.5 kcal/mol (in favour of eq). In case of **2a**, only the equatorial conformation is populated at RT (energy difference >8 kcal/mol). ECD calculations were performed with CAM-B3LYP/def2-TZVP, tests with wB97XD and B3LYP gave comparable results. To further process the results SpecDis[50] was used. For the Boltzmann weighting Gibbs free energies were utilized. Gaussians were prepared with a value of 0.26 eV and to compensate systematic errors in the calculation of the excited states a UV shift of 25 nm for **4** and of 39 nm for **3a** was applied.

HIV screening assay (EASY-HIT). The EASY-HIT assay is based on the HIV-1 susceptible reporter cell line LC5-RIC that contains a stably integrated reporter construct encoding the fluorochrome DsRed1 that will be expressed upon HIV infection and expression of the early viral

26

60

proteins Tat and Rev.[41] Briefly, 10,000 LC5-RIC cells were seeded into each well of black 96-well

plates and incubated over night at 37°C, 5% CO_2. Compound stocks (50 mM in DMSO) were

screened in a 1:2 series dilution at concentrations from 1.5 to 50 µM at a final DMSO concentration

of 0.1% to determine IC_{50} curves. After addition of the compound, LC5-RIC cells were inoculated

with HIV-1$_{LAI}$ at a MOI of 0.5 and incubated for 48 hours at 37°C, 5% CO_2. Finally, DsRed1

reporter expression was measured using a fluorescence microplate reader at an excitation filter

wavelength of 552 nm and an emission filter wavelength of 596 nm.

Cell viability assay. A CellTiter-Blue® cell viability assay from Promega was performed to check

on cell viability of HIV-1$_{LAI}$ inoculated and test compound treated LC5-RIC cultures. This assay

measures the ability of metabolically active cells to convert the redox dye resazurin into resorufin,

which can be detected by fluorescence spectroscopy. 10,000 LC5-RIC cells were seeded into each

well of black 96-well plates and incubated over night at 37°C, 5% CO_2. Compound stocks (50 mM

in DMSO) were screened in a 1:2 series dilution at concentrations from 1.5 to 50 µM at a final

DMSO concentration of 0.1%. After 48 hours of incubation at 37°C, 5% CO_2 a 1:5 CTB

reagent:cell culture medium mix was added to each well and incubated for another hour. Finally,

fluorescence signal of resorufin was detected using a fluorescence microplate reader at an

excitation filter wavelength of 550 nm and an emission filter wavelength of 600 nm.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS publication website at DOI:XX.

- Plasmids, bacterial strains and primers

27

- Phylogenetic analysis of ketosynthases

- Restriction analysis of constructs

- HPLC, UV, MS and NMR data of **1-4**

- Heat of formation (B3LYP/def2-TZVP), number of imaginary imaginery frequencies, Cartesian coordinates of **3a** and **4**.

Author Information.

Corresponding Author

*Tobias A. M. Gulder

E-mail: tobias.gulder@ch.tum.de, tobias.gulder@tu-dresden.de

Telephone number: +49-(0)89-289-13833

ORCID

P. M. D'Agostino: 0000-0002-8323-5416

T. Bruhn: 0000-0002-9604-1004

T. A. M. Gulder: 0000-0001-6013-3161

Postal Address

Technische Universität München

Department of Chemistry and Center for Integrated Protein Science Munich (CIPSM)

Biosystems Chemistry

Lichtenbergstraße 4

85748 Garching

28

62

References

1. Newman, D. J.; Cragg, G. M., Natural Products as Sources of New Drugs from 1981 to 2014. *J. Nat. Prod.* **2016,** *79*, 629-661.

2. Demain, A. L., Importance of microbial natural products and the need to revitalize their discovery. *J. Ind. Microbiol. Biot.* **2014,** *41*, 185-201.

3. Shen, B., A New Golden Age of Natural Products Drug Discovery. *Cell* **2015,** *163*, 1297-1300.

4. Rutledge, P. J.; Challis, G. L., Discovery of microbial natural products by activation of silent biosynthetic gene clusters. *Nat. Rev. Microbiol.* **2015,** *13*, 509-23.

5. Bentley, S. D.; Chater, K. F.; Cerdeno-Tarraga, A. M.; Challis, G. L.; Thomson, N. R.; James, K. D.; Harris, D. E.; Quail, M. A.; Kieser, H.; Harper, D.; Bateman, A.; Brown, S.; Chandra, G.; Chen, C. W.; Collins, M.; Cronin, A.; Fraser, A.; Goble, A.; Hidalgo, J.; Hornsby, T.; Howarth, S.; Huang, C. H.; Kieser, T.; Larke, L.; Murphy, L.; Oliver, K.; O'Neil, S.; Rabbinowitsch, E.; Rajandream, M. A.; Rutherford, K.; Rutter, S.; Seeger, K.; Saunders, D.; Sharp, S.; Squares, R.; Squares, S.; Taylor, K.; Warren, T.; Wietzorrek, A.; Woodward, J.; Barrell, B. G.; Parkhill, J.; Hopwood, D. A., Complete genome sequence of the model actinomycete *Streptomyces* coelicolor A3(2). *Nature* **2002,** *417*, 141-147.

6. Mukherjee, S.; Seshadri, R.; Varghese, N. J.; Eloe-Fadrosh, E. A.; Meier-Kolthoff, J. P.; Goker, M.; Coates, R. C.; Hadjithomas, M.; Pavlopoulos, G. A.; Paez-Espino, D.; Yoshikuni, Y.; Visel, A.; Whitman, W. B.; Garrity, G. M.; Eisen, J. A.; Hugenholtz, P.; Pati, A.; Ivanova, N. N.; Woyke, T.; Klenk, H. P.; Kyrpides, N. C., 1,003 reference genomes of bacterial and archaeal isolates expand coverage of the tree of life. *Nat. Biotechnol.* **2017,** *35*, 676-683.

7. Kwiatkowska, M.; Parker, D.; Wiltsche, C., PRISM-Games 2.0: A Tool for Multi-objective Strategy Synthesis for Stochastic Games. *Lect. Notes Comput. Sci.* **2016,** *9636*, 560-566.

8. Blin, K.; Wolf, T.; Chevrette, M. G.; Lu, X.; Schwalen, C. J.; Kautsar, S. A.; Suarez Duran, H. G.; de Los Santos, E. L. C.; Kim, H. U.; Nave, M.; Dickschat, J. S.; Mitchell, D. A.; Shelest, E.;

29

Breitling, R.; Takano, E.; Lee, S. Y.; Weber, T.; Medema, M. H., antiSMASH 4.0-improvements in chemistry prediction and gene cluster boundary identification. *Nucleic Acids Res.* **2017,** *45*, W36-W41.

9. Ziemert, N.; Alanjary, M.; Weber, T., The evolution of genome mining in microbes - a review. *Nat. Prod. Rep.* **2016,** *33*, 988-1005.

10. Scherlach, K.; Hertweck, C., Triggering cryptic natural product biosynthesis in microorganisms. *Org. Biomol. Chem.* **2009,** *7*, 1753-1760.

11. Gomez-Escribano, J. P.; Bibb, M. J., Engineering *Streptomyces* coelicolor for heterologous expression of secondary metabolite gene clusters. *Microb. Biotechnol.* **2011,** *4*, 207-15.

12. Fu, J.; Bian, X.; Hu, S.; Wang, H.; Huang, F.; Seibert, P. M.; Plaza, A.; Xia, L.; Muller, R.; Stewart, A. F.; Zhang, Y., Full-length RecE enhances linear-linear homologous recombination and facilitates direct cloning for bioprospecting. *Nat. Biotechnol.* **2012,** *30*, 440-6.

13. Yamanaka, K.; Reynolds, K. A.; Kersten, R. D.; Ryan, K. S.; Gonzalez, D. J.; Nizet, V.; Dorrestein, P. C.; Moore, B. S., Direct cloning and refactoring of a silent lipopeptide biosynthetic gene cluster yields the antibiotic taromycin A. *Proc. Natl. Acad. Sci.* **2014,** *111*, 1957-62.

14. Greunke, C.; Duell, E. R.; D'Agostino, P. M.; Glockle, A.; Lamm, K.; Gulder, T. A. M., Direct Pathway Cloning (DiPaC) to unlock natural product biosynthetic potential. *Metab. Eng.* **2018,** *47*, 334-345.

15. D'Agostino, P. M.; Gulder, T. A. M., Direct Pathway Cloning Combined with Sequence- and Ligation-Independent Cloning for Fast Biosynthetic Gene Cluster Refactoring and Heterologous Expression. *ACS Synth. Biol.* **2018,** *7*, 1702-1708.

16. Yin, J.; Hoffmann, M.; Bian, X.; Tu, Q.; Yan, F.; Xia, L.; Ding, X.; Stewart, A. F.; Muller, R.; Fu, J.; Zhang, Y., Direct cloning and heterologous expression of the salinomycin biosynthetic gene cluster from *Streptomyces* albus DSM41398 in *Streptomyces* coelicolor A3(2). *Sci. Rep.* **2015,** *5*, 15081.

17. Wang, H.; Li, Z.; Jia, R.; Yin, J.; Li, A.; Xia, L.; Yin, Y.; Müller, R.; Fu, J.; Stewart, A. F., ExoCET: exonuclease in vitro assembly combined with RecET recombination for highly efficient direct DNA cloning from complex genomes. *Nucleic Acids Res.* **2017,** *46*, e28-e28.

18. Wang, H.; Li, Z.; Jia, R.; Hou, Y.; Yin, J.; Bian, X.; Li, A.; Muller, R.; Stewart, A. F.; Fu, J.; Zhang, Y., RecET direct cloning and Redalphabeta recombineering of biosynthetic gene clusters, large operons or single genes for heterologous expression. *Nat. Protoc.* **2016,** *11*, 1175-90.

30

64

19. Bierman, M.; Logan, R.; Obrien, K.; Seno, E. T.; Rao, R. N.; Schoner, B. E., Plasmid Cloning Vectors for the Conjugal Transfer of DNA from Escherichia-Coli to *Streptomyces* Spp. *Gene* **1992,** *116*, 43-49.

20. Pickens, L. B.; Tang, Y., Oxytetracycline biosynthesis. *J. Biol. Chem.* **2010,** *285*, 27509-15.

21. Zhu, T.; Cheng, X.; Liu, Y.; Deng, Z.; You, D., Deciphering and engineering of the final step halogenase for improved chlortetracycline biosynthesis in industrial *Streptomyces* aureofaciens. *Metab. Eng.* **2013,** *19*, 69-78.

22. Aubelsadron, G.; Londosgagliardi, D., Daunorubicin and Doxorubicin, Anthracycline Antibiotics, a Physicochemical and Biological Review. *Biochimie* **1984,** *66*, 333-352.

23. Shen, B., Biosynthesis of aromatic polyketides. *Top. Curr. Chem.* **2000,** *209*, 1-51.

24. Blin, K.; Wolf, T.; Chevrette, M. G.; Lu, X. W.; Schwalen, C. J.; Kautsar, S. A.; Duran, H. G. S.; Santos, E. L. C. D. L.; Kim, H. U.; Nave, M.; Dickschat, J. S.; Mitchell, D. A.; Shelest, E.; Breitling, R.; Takano, E.; Lee, S. Y.; Weber, T.; Medema, M. H., antiSMASH 4.0-improvements in chemistry prediction and gene cluster boundary identification. *Nucleic Acids Res.* **2017,** *45*, W36-W41.

25. Medema, M. H.; Kottmann, R.; Yilmaz, P.; Cummings, M.; Biggins, J. B.; Blin, K.; de Bruijn, I.; Chooi, Y. H.; Claesen, J.; Coates, R. C.; Cruz-Morales, P.; Duddela, S.; Dusterhus, S.; Edwards, D. J.; Fewer, D. P.; Garg, N.; Geiger, C.; Gomez-Escribano, J. P.; Greule, A.; Hadjithomas, M.; Haines, A. S.; Helfrich, E. J. N.; Hillwig, M. L.; Ishida, K.; Jones, A. C.; Jones, C. S.; Jungmann, K.; Kegler, C.; Kim, H. U.; Kotter, P.; Krug, D.; Masschelein, J.; Melnik, A. V.; Mantovani, S. M.; Monroe, E. A.; Moore, M.; Moss, N.; Nutzmann, H. W.; Pan, G. H.; Pati, A.; Petras, D.; Reen, F. J.; Rosconi, F.; Rui, Z.; Tian, Z. H.; Tobias, N. J.; Tsunematsu, Y.; Wiemann, P.; Wyckoff, E.; Yan, X. H.; Yim, G.; Yu, F. G.; Xie, Y. C.; Aigle, B.; Apel, A. K.; Balibar, C. J.; Balskus, E. P.; Barona-Gomez, F.; Bechthold, A.; Bode, H. B.; Borriss, R.; Brady, S. F.; Brakhage, A. A.; Caffrey, P.; Cheng, Y. Q.; Clardy, J.; Cox, R. J.; De Mot, R.; Donadio, S.; Donia, M. S.; van der Donk, W. A.; Dorrestein, P. C.; Doyle, S.; Driessen, A. J. M.; Ehling-Schulz, M.; Entian, K. D.; Fischbach, M. A.; Gerwick, L.; Gerwick, W. H.; Gross, H.; Gust, B.; Hertweck, C.; Hofte, M.; Jensen, S. E.; Ju, J. H.; Katz, L.; Kaysser, L.; Klassen, J. L.; Keller, N. P.; Kormanec, J.; Kuipers, O. P.; Kuzuyama, T.; Kyrpides, N. C.; Kwon, H. J.; Lautru, S.; Lavigne, R.; Lee, C. Y.; Linquan, B.; Liu, X. Y.; Liu, W.; Luzhetskyy, A.; Mahmud, T.; Mast, Y.; Mendez, C.; Metsa-Ketela, M.; Micklefield, J.; Mitchell, D. A.; Moore, B. S.; Moreira, L. M.; Muller, R.; Neilan, B. A.; Nett, M.;

31

65

Nielsen, J.; O'Gara, F.; Oikawa, H.; Osbourn, A.; Osburne, M. S.; Ostash, B.; Payne, S. M.; Pernodet, J. L.; Petricek, M.; Piel, J.; Ploux, O.; Raaijmakers, J. M.; Salas, J. A.; Schmitt, E. K.; Scott, B.; Seipke, R. F.; Shen, B.; Sherman, D. H.; Sivonen, K.; Smanski, M. J.; Sosio, M.; Stegmann, E.; Sussmuth, R. D.; Tahlan, K.; Thomas, C. M.; Tang, Y.; Truman, A. W.; Viaud, M.; Walton, J. D.; Walsh, C. T.; Weber, T.; van Wezel, G. P.; Wilkinson, B.; Willey, J. M.; Wohlleben, W.; Wright, G. D.; Ziemert, N.; Zhang, C. S.; Zotchev, S. B.; Breitling, R.; Takano, E.; Glockner, F. O., Minimum Information about a Biosynthetic Gene cluster. *Nat. Chem. Biol.* **2015,** *11*, 625-631.

26. Pickens, L. B.; Kim, W.; Wang, P.; Zhou, H.; Watanabe, K.; Gomi, S.; Tang, Y., Biochemical analysis of the biosynthetic pathway of an anticancer tetracycline SF2575. *J. Am. Chem. Soc.* **2009,** *131*, 17677-89.

27. Metsa-Ketela, M.; Halo, L.; Munukka, E.; Hakala, J.; Mantsala, P.; Ylihonko, K., Molecular evolution of aromatic polyketides and comparative sequence analysis of polyketide ketosynthase and 16S ribosomal DNA genes from various *Streptomyces* species. *Appl. Environ. Microb.* **2002,** *68*, 4472-4479.

28. Feng, Z.; Kallifidas, D.; Brady, S. F., Functional analysis of environmental DNA-derived type II polyketide synthases reveals structurally diverse secondary metabolites. *Proc. Natl. Acad. Sci.* **2011,** *108*, 12629-12634.

29. Liu, Q.; Shen, Q.; Bian, X.; Chen, H.; Fu, J.; Wang, H.; Lei, P.; Guo, Z.; Chen, W.; Li, D.; Zhang, Y., Simple and rapid direct cloning and heterologous expression of natural product biosynthetic gene cluster in Bacillus subtilis via Red/ET recombineering. *Sci. Rep.* **2016,** *6*, 34623.

30. Lomovskaya, N.; Doi-Katayama, Y.; Filippini, S.; Nastro, C.; Fonstein, L.; Gallo, M.; Colombo, A. L.; Hutchinson, C. R., The *Streptomyces* peucetius dpsY and dnrX genes govern early and late steps of daunorubicin and doxorubicin biosynthesis. *J. Bacteriol.* **1998,** *180*, 2379-2386.

31. Kunnari, T.; Kantola, J.; Ylihonko, K.; Klika, K. D.; Mantsala, P.; Hakala, J., Hybrid compounds derived from the combination of anthracycline and actinorhodin biosynthetic pathways. *J. Chem. Soc. Perk. Trans. 2* **1999**, 1649-1652.

32. Yanai, T.; Tew, D. P.; Handy, N. C., A new hybrid exchange–correlation functional using the Coulomb-attenuating method (CAM-B3LYP). *Chem. Phy. Lett.* **2004,** *393*, 51-57.

32

33. Weigend, F.; Ahlrichs, R., Balanced basis sets of split valence, triple zeta valence and quadruple zeta valence quality for H to Rn: Design and assessment of accuracy. *Phys. Chem. Chem. Phys.* **2005,** *7*, 3297-3305.

34. Moore, B. S.; Hertweck, C., Biosynthesis and attachment of novel bacterial polyketide synthase starter units. *Nat. Prod. Rep.* **2002,** *19*, 70-99.

35. Jackson, D. R.; Shakya, G.; Patel, A. B.; Mohammed, L. Y.; Vasilakis, K.; Wattana-Amorn, P.; Valentic, T. R.; Milligan, J. C.; Crump, M. P.; Crosby, J.; Tsai, S. C., Structural and Functional Studies of the Daunorubicin Priming Ketosynthase DpsC. *ACS Chem. Biol.* **2018,** *13*, 141-151.

36. Rajgarhia, V. B.; Priestley, N. D.; Strohl, W. R., The product of dpsC confers starter unit fidelity upon the daunorubicin polyketide synthase of *Streptomyces* sp. strain C5. *Metab. Eng.* **2001,** *3*, 49-63.

37. Bao, W.; Sheldon, P. J.; Hutchinson, C. R., Purification and properties of the *Streptomyces* peucetius DpsC beta-ketoacyl:acyl carrier protein synthase III that specifies the propionate-starter unit for type II polyketide biosynthesis. *Biochemistry* **1999,** *38*, 9752-7.

38. Bao, W.; Sheldon, P. J.; Wendt-Pienkowski, E.; Hutchinson, C. R., The *Streptomyces* peucetius dpsC gene determines the choice of starter unit in biosynthesis of the daunorubicin polyketide. *J. Bacteriol.* **1999,** *181*, 4690-5.

39. Kantola, J.; Kunnari, T.; Hautala, A.; Hakala, J.; Ylihonko, K.; Mantsala, P., Elucidation of anthracyclinone biosynthesis by stepwise cloning of genes for anthracyclines from three different *Streptomyces* spp. *Microbiology* **2000,** *146 (Pt 1)*, 155-63.

40. Krohn, K.; Vukics, K., First chemical synthesis of the antiviral agents S2502 and S2507. *Synthesis-Stuttgart* **2007**, 2894-2900.

41. Kremb, S.; Helfer, M.; Heller, W.; Hoffmann, D.; Wolff, H.; Kleinschmidt, A.; Cepok, S.; Hemmer, B.; Durner, J.; Brack-Werner, R., EASY-HIT: HIV full-replication technology for broad discovery of multiple classes of HIV inhibitors. *Antimicrob. Agents Chemother.* **2010,** *54*, 5257-5268.

42. Riss, T. L.; Moravec, R. A.; Niles, A. L.; Duellman, S.; Benink, H. A.; Worzella, T. J.; Minor, L., Cell viability assays. In *Assay Guidance Manual [Internet]*, Eli Lilly & Company and the National Center for Advancing Translational Sciences: **2016**.

43. Liotta, D. C.; Painter, G. R., Discovery and development of the anti-human immunodeficiency virus drug, emtricitabine (Emtriva, FTC). *Accounts Chem. Res.* **2016,** *49*, 2091-2098.

33

67

44. Kumar, S.; Stecher, G.; Li, M.; Knyaz, C.; Tamura, K., MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms. *Mol. Biol. Evol.* **2018,** *35*, 1547-1549.

45. Wang, X. K.; Jin, J. L., Crucial factor for increasing the conjugation frequency in *Streptomyces* netropsis SD-07 and other strains. *FEMS Microbiol. Lett.* **2014,** *357*, 99-103.

46. Pescitelli, G.; Bruhn, T., Good computational practice in the assignment of absolute configurations by TDDFT calculations of ECD spectra. *Chirality* **2016,** *28*, 466-474.

47. Neese, F., Software update: the ORCA program system, version 4.0. *Wiley Interdiscip. Rev.-Comput. Mol. Sci.* **2018,** *8*, e1327.

48. Frisch, M.; Trucks, G.; Schlegel, H.; Scuseria, G.; Robb, M.; Cheeseman, J.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G., Gaussian09 Revision D. 01, Gaussian Inc. Wallingford CT. *URL:* http://www.*gaussian. com* **2009**.

49. Grimme, S.; Ehrlich, S.; Goerigk, L., Effect of the damping function in dispersion corrected density functional theory. *J. Comput. Chem.* **2011,** *32*, 1456-1465.

50. Bruhn, T.; Schaumlöffel, A.; Hemberger, Y.; Pescitelli, G., SpecDis version 1.71; Berlin, Germany. 2017. *URL:* https://specdis-software.jimdo.com **2017**.

34

68

3.3 Biosynthesis of cycloheptamycins

The following chapter is based on the manuscript:

Z. Qian, J. Antosch, P.M. D'Agostino, T. Liu, M. Fottner, R. Zhu, A. Pöthig, T.A.M. Gulder, et al.. Functional characterization of the biosynthesis of the antibiotic cycloheptamycins. Manuscript in preparation.

Within this work, we identified an NRPS-type BGC responsible for the biosynthesis of cycloheptamycins from bioinformatic analysis of the genome data of *Streptomyces* sp. Tü6314. The cluster spans ~49 kb and consists of 23 open reading frames (ORFs) (Figure 21A). This DNA region was cloned directly from the Tü6314 genome by the LLHR method. Gene disruption in the native Tü6314 strain and heterologous expression of the cluster in *Streptomyces coelicolor* confirmed that it is responsible for cycloheptamycin biosynthesis.
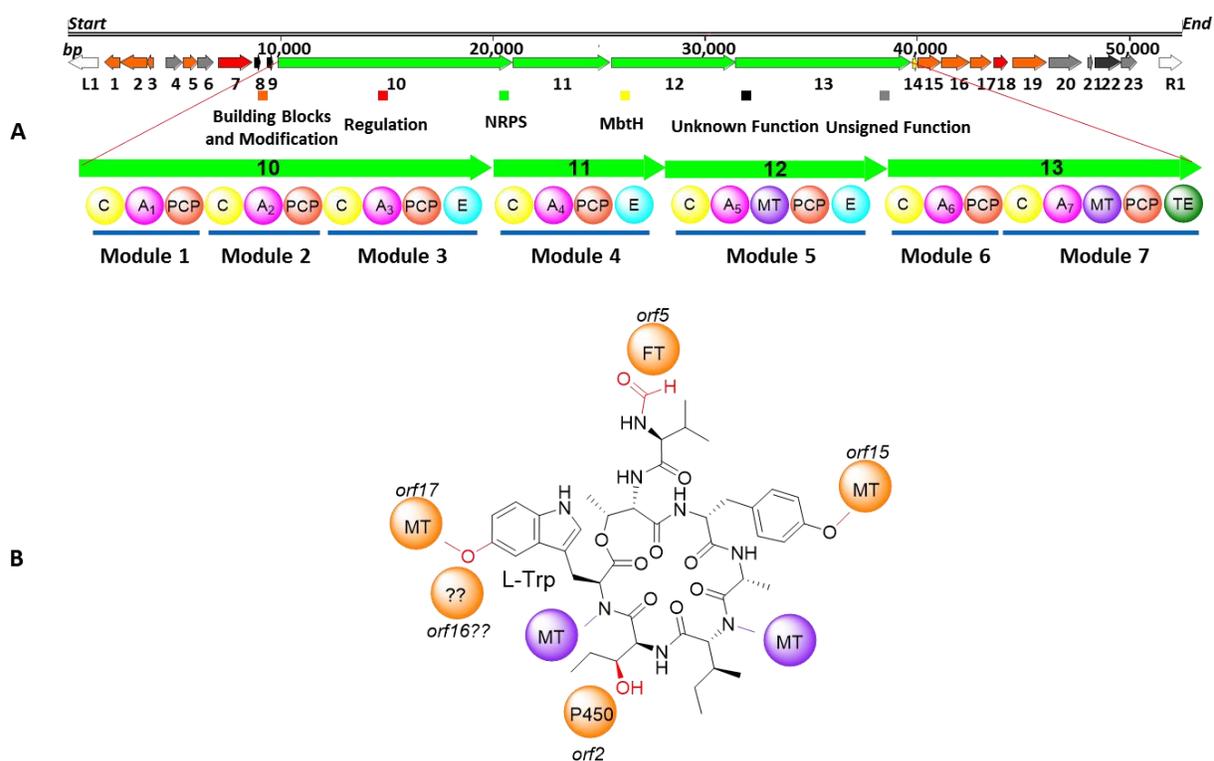
Figure 21. (**A**) Organization of the cycloheptamycin BGC from *Streptomyces* sp. Tü 6314. (**B**) Structure of cycloheptamycin A and the proposed modifying enzymes or NRPS domains.

Bioinformatic analysis of the functions of the genes combined with gene disruptions revealed further information about this gene cluster. More information will be published soon as a paper work.

4. Summary and outlook

4.1 Cycloheptamycins

In summary, six compounds related to cycloheptamycins were isolated from *Streptomyces* sp. Tü6314 in this study. The BGC was cloned and the biosynthetic pathway was studied by heterologous expression combined with gene disruption. An evolved isopropylmalate synthase was shown to be involved in biosynthesis of the norvaline building block of cycloheptamycin A (**35**). With the selective antibiotic activity against *Propionibacterium acnes* of cycloheptamycin A (**35**) and B (**36**), these findings may lead to potent anti-infective drug or drug lead. The elucidation of the biosynthetic pathway enriched our knowledge about NRPS biosynthesis and provided insight into enzyme evolution. Some remaining questions, for example the biosynthesis of the 5-methoxy-tryptophan or the regulation of the biosynthesis, are still to be elucidated. Further studies thus may focus on these questions. Whole gene cluster cloning and heterologous expression also provided possibilities for gene cluster engineering to make new compounds, including domain exchange or pathway recombination.

4.2 Streptoketides

In this study, we isolated six polyketides by heterologous expression of a cryptic type II PKS gene cluster (**37-42,** Figure 20B). Four of the six compounds exhibited anti-HIV activities, which may lead to potent anti-viral drugs or drug leads. The stereostructures were determined by comparing their calculated ECD spectra with the experimental spectra. Further verification could be obtained by getting crystals of these compounds and doing single-crystal X-ray analysis. We also noticed that these compounds are produced in rather limited yield. To solve this problem, a series of other host strains could be tested, the regulator genes could be manipulated or the promoters could be exchanged.

4.3 *Streptomyces* sp. Tü6314 genome mining

Genome analysis showed that Tü6314 genome contains 28 predicted BGCs (Table 1). Only 10 of which show similarity greater than 50% to already known BGCs. Our studies have shown that two of these BGCs are responsible to produce cycloheptamycins (cluster 5) and streptoketides (cluster 20), respectively. During our study, literature has reported that a cluster similar to cluster 8 is responsible for the production of the polyene macrolactam sceliphrolactam[120] and a cluster similar to cluster 27 for the production of a NRPS-PKS hybrid product detoxin.[121] All these finding have made it clearer to understand the full biosynthetic potential of *Streptomyces* sp. Tü6314. Nevertheless, the remaining uncharacterized gene clusters are still a rich wealth for genome mining, which may need more efforts to achieve the full biosynthesis potential of this underexplored strain.

Table 1. Cluster prediction results of *Streptomyces* sp. Tü6314 genome analyzed by antiSMASH 5.0.[122]

No.	Predicted type	Size/bp	Most similar known cluster/type		Similarity
1	NRPS, T1PKS, blactam	146,260	carbapenem MM4550	Non-NRP beta-lactam	65%
2	NRPS	49,062	coelichelin	NRP	90%
3	terpene	24,583	isorenieratene	Terpene	100%
4	bacteriocin	8,918			
5	blactam, NRPS, betalactone	84,586	marformycin	NRP	33%
6	lanthipeptide	22,854			
7	terpene	25,483	hopene	Terpene	69%
8	T1PKS	89,660	sceliphrolactam	Polyketide	88%
9	bacteriocin	10,265			
10	NRPS	61,091	cadaside	NRP	19%
11	siderophore	13,122	ficellomycin	NRP	3%
12	terpene	19,715			
13	bacteriocin	9,546			
14	butyrolactone	8,310	lactonamycin	Polyketide	3%
15	NRPS, T1PKS	56,274	istamycin	Saccharide	11%
16	siderophore	11,781	desferrioxamin	Other	83%
17	lanthipeptide	23,068	azalomycin	Polyketide	8%
18	terpene	20,592			
19	ectoine	8,610	ectoine	Other	100%
20	T2PKS, PKS-like	71,509	cinerubin B	Polyketide:Type II	25%
21	terpene	20,542	steffimycin D	Polyketide:Type II + Saccharide:Hybrid	16%
22	terpene,ectoine	20,927	ectoine	Other	100%
23	bacteriocin	10,227			
24	T3PKS	41,058	tetronasin	Polyketide	11%
25	melanin	10,464	melanin	Other	100%
26	T2PKS, terpene	72,527	spore pigment	Polyketide	83%
27	NRPS	53,462	rimosamide	NRP	21%
28	butyrolactone	10,935			

References

1. Demain, A. L.; Sanchez, S., Microbial drug discovery: 80 years of progress. *J. Antibiot.* **2009,** *62*, 5-16.

2. Newman, D. J.; Cragg, G. M., Natural Products as Sources of New Drugs from 1981 to 2014. *J. Nat. Prod.* **2016,** *79*, 629-661.

3. Shen, B., A New Golden Age of Natural Products Drug Discovery. *Cell* **2015,** *163*, 1297-1300.

4. Fleming, A., On the antibacterial action of cultures of a penicillium, with special reference to their use in the isolation of B. influenzae. *B. World Health Organ.* **2001,** *79*, 780-790.

5. Katz, L.; Baltz, R. H., Natural product discovery: past, present, and future. *J. Ind. Microbiol. Biot.* **2016,** *43*, 155-176.

6. Schatz, A.; Bugle, E.; Waksman, S. A., Streptomycin, a Substance Exhibiting Antibiotic Activity Against Gram-Positive and Gram-Negative Bacteria.∗. *Proc. Soc. Exp. Biol. Med.* **1944,** *55*, 66-69.

7. Singh, S. B.; Barrett, J. F., Empirical antibacterial drug discovery - Foundation in natural products. *Biochem. Pharmacol.* **2006,** *71*, 1006-1015.

8. Milshteyn, A.; Schneider, J. S.; Brady, S. F., Mining the Metabiome: Identifying Novel Natural Products from Microbial Communities. *Chem. Biol.* **2014,** *21*, 1211-1223.

9. Aminov, R., History of antimicrobial drug discovery: Major classes and health impact. *Biochem. Pharmacol.* **2017,** *133*, 4-19.

10. Waksman, S. A.; Woodruff, H. B., Bacteriostatic and bactericidal substances produced by a soil Actinomyces. *Proc. Soc. Exp. Biol. Med.* **1940,** *45*, 609-614.

11. Hata, T.; Sano, Y.; Sugawara, R.; Matsumae, A.; Kanamori, K.; Shima, T.; Hoshi, T., Mitomycin, a New Antibiotic from *Streptomyces* .1. *J. Antibiot.* **1956,** *9*, 141-146.

12. Dimarco, A.; Gaetani, M.; Dorigotti, L.; Soldati, M.; Bellini, O., Daunomycin - a New Antibiotic with Antitumor Activity. *Cancer Chemoth. Rep.* **1964,** *38*, 31-38.

13. Umezawa, H.; Maeda, K.; Takeuchi, T.; Okami, Y., New Antibiotics Bleomycin A and B. *J. Antibiot.* **1966,** *19*, 200-209.

14. Sehgal, S. N.; Baker, H.; Vezina, C., Rapamycin (Ay-22,989), a New Antifungal Antibiotic .2. Fermentation, Isolation and Characterization. *J. Antibiot.* **1975,** *28*, 727-732.

15. Vezina, C.; Kudelski, A.; Sehgal, S. N., Rapamycin (Ay-22,989), a New Antifungal Antibiotic .1. Taxonomy of Producing Streptomycete and Isolation of Active Principle. *J. Antibiot.* **1975,** *28*, 721-726.

16. Ruegger, A.; Kuhn, M.; Lichti, H.; Loosli, H. R.; Huguenin, R.; Quiquerez, C.; Wartburg, A. V., Cyclosporin-a, a Peptide Metabolite from Trichoderma-Polysporum (Link Ex Pers) Rifai, with a Remarkable Immunosuppressive Activity. *Helv. Chim. Acta* **1976,** *59*, 1075-1092.

17. Li, J. W. H.; Vederas, J. C., Drug Discovery and Natural Products: End of an Era or an Endless Frontier? *Science* **2009,** *325*, 161-165.

18. Rutledge, P. J.; Challis, G. L., Discovery of microbial natural products by activation of silent biosynthetic gene clusters. *Nat. Rev. Microbiol.* **2015,** *13*, 509-23.

19. Koehn, F. E.; Carter, G. T., The evolving role of natural products in drug discovery. *Nat. Rev. Drug. Discov.* **2005,** *4*, 206-220.

20. Lam, K. S., New aspects of natural products in drug discovery. *Trends Microbiol.* **2007,** *15*, 279-289.

21. Cragg, G. M.; Newman, D. J., Natural products: A continuing source of novel drug leads. *BBA-Gen. Subjects* **2013,** *1830*, 3670-3695.

22. Tang, M. C.; Zou, Y.; Watanabe, K.; Walsh, C. T.; Tang, Y., Oxidative Cyclization in Natural Product Biosynthesis. *Chem. Rev.* **2017,** *117*, 5226-5333.

23. Staunton, J.; Weissman, K. J., Polyketide biosynthesis: a millennium review. *Nat. Prod. Rep.* **2001,** *18*, 380-416.

24. Weissman, K. J., Introduction to polyketide biosynthesis. *Method. Enzym.* **2009,** *459*, 3-16.

25. Donadio, S.; Staver, M. J.; Mcalpine, J. B.; Swanson, S. J.; Katz, L., Modular Organization of Genes Required for Complex Polyketide Biosynthesis. *Science* **1991,** *252*, 675-679.

26. Walsh, C. T.; Gehring, A. M.; Weinreb, P. H.; Quadri, L. E.; Flugel, R. S., Post-translational modification of polyketide and nonribosomal peptide synthases. *Curr. Opin. Chem. Biol.* **1997,** *1*, 309-315.

27. Shen, B., Polyketide biosynthesis beyond the type I, II and III polyketide synthase paradigms. *Curr. Opin. Chem. Biol.* **2003,** *7*, 285-295.

28. Cortes, J.; Haydock, S. F.; Roberts, G. A.; Bevitt, D. J.; Leadlay, P. F., An Unusually Large Multifunctional Polypeptide in the Erythromycin-Producing Polyketide Synthase of Saccharopolyspora-Erythraea. *Nature* **1990,** *348*, 176-178.

29. Keatinge-Clay, A. T., The structures of type I polyketide synthases. *Nat. Prod. Rep.* **2012,** *29*, 1050-1073.

30. Keatinge-Clay, A. T., Polyketide synthase modules redefined. *Angew. Chem. Int. Edit.* **2017,** *56*, 4658-4660.

31. Fischbach, M. A.; Walsh, C. T., Assembly-line enzymology for polyketide and nonribosomal peptide antibiotics: logic, machinery, and mechanisms. *Chem. Rev.* **2006,** *106*, 3468-3496.

32. Herbst, D. A.; Townsend, C. A.; Maier, T., The architectures of iterative type I PKS and FAS. *Nat. Prod. Rep.* **2018,** *35*, 1046-1069.

33. Hertweck, C., The biosynthetic logic of polyketide diversity. *Angew. Chem. Int. Edit.* **2009,** *48*, 4688-716.

34. Moriguchi, T.; Kezuka, Y.; Nonaka, T.; Ebizuka, Y.; Fujii, I., Hidden Function of Catalytic Domain in 6-Methylsalicylic Acid Synthase for Product Release. *J. Biol. Chem.* **2010,** *285*, 15637-15643.

35. Sun, H. H.; Ho, C. L.; Ding, F. Q.; Soehano, I.; Liu, X. W.; Liang, Z. X., Synthesis of (R)-Mellein by a Partially Reducing Iterative Polyketide Synthase. *J. Am. Chem. Soc.* **2012,** *134*, 11924-11927.

36. Gaisser, S.; Trefzer, A.; Stockert, S.; Kirschning, A.; Bechthold, A., Cloning of an avilamycin biosynthetic gene cluster from *Streptomyces viridochromogenes* Tu57. *J. Bacteriol.* **1997,** *179*, 6271-6278.

37. Jia, X. Y.; Tian, Z. H.; Shao, L.; Qu, X. D.; Zhao, Q. F.; Tang, J.; Tang, G. L.; Liu, W., Genetic characterization of the chlorothricin gene cluster as a model for spirotetronate antibiotic biosynthesis. *Chem. Biol.* **2006,** *13*, 575-585.

38. Zhao, Q. F.; He, Q. L.; Ding, W.; Tang, M. C.; Kang, Q. J.; Yu, Y.; Deng, W.; Zhang, Q.; Fang, J.; Tang, G. L.; Liu, W., Characterization of the azinomycin B biosynthetic gene cluster revealing a different iterative type I polyketide synthase for naphthoate biosynthesis. *Chem. Biol.* **2008,** *15*, 693-705.

39. Liu, W.; Nonaka, K.; Nie, L. P.; Zhang, J.; Christenson, S. D.; Bae, J.; Van Lanen, S. G.; Zazopoulos, E.; Farnet, C. M.; Yang, C. F.; Shen, B., The neocarzinostatin biosynthetic gene cluster from *Streptomyces carzinostaticus* ATCC 15944 involving two iterative type I polyketide synthases. *Chem. Biol.* **2005,** *12*, 293-302.

40. Saha, S.; Zhang, W.; Zhang, G.; Zhu, Y.; Chen, Y.; Liu, W.; Yuan, C.; Zhang, Q.; Zhang, H.; Zhang, L., Activation and characterization of a cryptic gene cluster reveals a cyclization cascade for polycyclic tetramate macrolactams. *Chem. Sci.* **2017,** *8*, 1607-1612.

41. Yu, F.; Zaleta-Rivera, K.; Zhu, X.; Huffman, J.; Millet, J. C.; Harris, S. D.; Yuen, G.; Li, X.-C.; Du, L., Structure and biosynthesis of heat-stable antifungal factor (HSAF), a broad-spectrum antimycotic with a novel mode of action. *Antimicrob. Agents Chemother.* **2007,** *51*, 64-72.

42. Jomon, K.; Kuroda, Y.; AJISAKA, M.; SAKAI, H., A new antibiotic, ikarugamycin. *J. Antibiot.* **1972,** *25*, 271-280.

43. Luo, Y.; Huang, H.; Liang, J.; Wang, M.; Lu, L.; Shao, Z.; Cobb, R. E.; Zhao, H., Activation and characterization of a cryptic polycyclic tetramate macrolactam biosynthetic gene cluster. *Nat. Comm.* **2013,** *4*, 2894.

44. Antosch, J.; Schaefers, F.; Gulder, T. A., Heterologous reconstitution of ikarugamycin biosynthesis in E. coli. *Angew. Chem. Int. Edit.* **2014,** *53*, 3011-3014.

45. Li, Y.; Wang, H.; Liu, Y.; Jiao, Y.; Li, S.; Shen, Y.; Du, L., Biosynthesis of the Polycyclic System in the Antifungal HSAF and Analogues from Lysobacter enzymogenes. *Angewandte. Chemie.* **2018,** *130*, 6329-6333.

46. Liu, Y.; Wang, H.; Song, R.; Chen, J.; Li, T.; Li, Y.; Du, L.; Shen, Y., Targeted Discovery and Combinatorial Biosynthesis of Polycyclic Tetramate Macrolactam Combamides A–E. *Org. Lett.* **2018**.

47. Chen, H. T.; Du, L. C., Iterative polyketide biosynthesis by modular polyketide synthases in bacteria. *Appl. Microbiol. Biot.* **2016,** *100*, 541-557.

48. Van Lanen, S. G.; Shen, B., Biosynthesis of enediyne antitumor antibiotics. *Curr. Top. Med. Chem.* **2008,** *8*, 448-459.

49. Liang, Z.-X., Complexity and simplicity in the biosynthesis of enediyne natural products. *Nat. Prod. Rep.* **2010,** *27*, 499-528.

50. Helfrich, E. J. N.; Piel, J., Biosynthesis of polyketides by trans-AT polyketide synthases. *Nat. Prod. Rep.* **2016,** *33*, 231-316.

51. Piel, J., A polyketide synthase-peptide synthetase gene cluster from an uncultured bacterial symbiont of Paederus beetles. *Proc. Natl. Acad. Sci.* **2002,** *99*, 14002-14007.

52. Shen, B., Biosynthesis of aromatic polyketides. *Top. Curr. Chem.* **2000,** *209*, 1-51.

53. Bisang, C.; Long, P. F.; Corte, J.; Westcott, J.; Crosby, J.; Matharu, A.-L.; Cox, R. J.; Simpson, T. J.; Staunton, J.; Leadlay, P. F., A chain initiation factor common to both modular and aromatic polyketide synthases. *Nature* **1999,** *401*, 502.

54. Hertweck, C.; Luzhetskyy, A.; Rebets, Y.; Bechthold, A., Type II polyketide synthases: gaining a deeper insight into enzymatic teamwork. *Nat. Prod. Rep.* **2007,** *24*, 162-190.

55. Funa, N.; Ohnishi, Y.; Fujii, I.; Shibuya, M.; Ebizuka, Y.; Horinouchi, S., A new pathway for polyketide synthesis in microorganisms. *Nature* **1999,** *400*, 897-899.

56. Yu, D. Y.; Xu, F. C.; Zeng, J.; Zhan, J. X., Type III polyketide synthases in natural product biosynthesis. *IUBMB Life* **2012,** *64*, 285-295.

57. Seshime, Y.; Juvvadi, P. R.; Fujii, I.; Kitamoto, K., Discovery of a novel superfamily of type III polyketide synthases in Aspergillus oryzae. *Biochem. Bioph. Res. Co.* **2005,** *331*, 253-260.

58. Gross, F.; Luniak, N.; Perlova, O.; Gaitatzis, N.; Jenke-Kodama, H.; Gerth, K.; Gottschalk, D.; Dittmann, E.; Muller, R., Bacterial type III polyketide synthases: phylogenetic analysis and potential for the production

of novel secondary metabolites by heterologous expression in pseudomonads. *Arch. Microbiol.* **2006,** *185*, 28-38.

59. Funa, N.; Awakawa, T.; Horinouchi, S., Pentaketide resorcylic acid synthesis by type III polyketide synthase from Neurospora crassa. *J. Biol. Chem.* **2007,** *282*, 14476-14481.

60. Finking, R.; Marahiel, M. A., Biosynthesis of nonribosomal peptides. *Annu. Rev. Microbiol.* **2004,** *58*, 453-488.

61. Marahiel, M. A.; Stachelhaus, T.; Mootz, H. D., Modular peptide synthetases involved in nonribosomal peptide synthesis. *Chem. Rev.* **1997,** *97*, 2651-2673.

62. Konz, D.; Marahiel, M. A., How do peptide synthetases generate structural diversity? *Chem. Biol.* **1999,** *6*, R39-R48.

63. Sieber, S. A.; Marahiel, M. A., Molecular mechanisms underlying nonribosomal peptide synthesis: approaches to new antibiotics. *Chem. Rev.* **2005,** *105*, 715-738.

64. Bloudoff, K.; Fage, C. D.; Marahiel, M. A.; Schmeing, T. M., Structural and mutational analysis of the nonribosomal peptide synthetase heterocyclization domain provides insight into catalysis. *Proc. Natl. Acad. Sci.* **2017,** *114*, 95-100.

65. Schoenafinger, G.; Schracke, N.; Linne, U.; Marahiel, M. A., Formylation domain: an essential modifying enzyme for the nonribosomal biosynthesis of linear gramicidin. *J. Am. Chem. Soc.* **2006,** *128*, 7406-7407.

66. Hur, G. H.; Vickery, C. R.; Burkart, M. D., Explorations of catalytic domains in non-ribosomal peptide synthetase enzymology. *Nat. Prod. Rep.* **2012,** *29*, 1074-1098.

67. Mootz, H. D.; Schwarzer, D.; Marahiel, M. A., Ways of assembling complex natural products on modular nonribosomal peptide synthetases. *Chembiochem* **2002,** *3*, 490-504.

68. Sussmuth, R. D.; Mainz, A., Nonribosomal Peptide Synthesis-Principles and Prospects. *Angew. Chem. Int. Edit.* **2017,** *56*, 3770-3821.

69. Keating, T. A.; Ehmann, D. E.; Kohli, R. M.; Marshall, C. G.; Trauger, J. W.; Walsh, C. T., Chain termination steps in nonribosomal peptide synthetase assembly lines: directed acyl-S-enzyme breakdown in antibiotic and siderophore biosynthesis. *Chembiochem* **2001,** *2*, 99-107.

70. Kohli, R. M.; Trauger, J. W.; Schwarzer, D.; Marahiel, M. A.; Walsh, C. T., Generality of peptide cyclization catalyzed by isolated thioesterase domains of nonribosomal peptide synthetases. *Biochemistry* **2001,** *40*, 7099-7108.

71. Hoyer, K. M.; Mahlert, C.; Marahiel, M. A., The iterative gramicidin S thioesterase catalyzes peptide ligation and cyclization. *Chem. Biol.* **2007,** *14*, 13-22.

72. Masschelein, J.; Mattheus, W.; Gao, L. J.; Moons, P.; Van Houdt, R.; Uytterhoeven, B.; Lamberigts, C.; Lescrinier, E.; Rozenski, J.; Herdewijn, P.; Aertsen, A.; Michiels, C.; Lavigne, R., A PKS/NRPS/FAS Hybrid Gene Cluster from Serratia plymuthica RVH1 Encoding the Biosynthesis of Three Broad Spectrum, Zeamine-Related Antibiotics. *Plos One* **2013,** *8*, e54143.

73. Fisch, K. M., Biosynthesis of natural products by microbial iterative hybrid PKS-NRPS. *Rsc. Adv.* **2013,** *3*, 18228-18247.

74. Li, Y. Y.; Chen, H. T.; Ding, Y. J.; Xie, Y. X.; Wang, H. X.; Cerny, R. L.; Shen, Y. M.; Du, L. C., Iterative Assembly of Two Separate Polyketide Chains by the Same Single-Module Bacterial Polyketide Synthase in the Biosynthesis of HSAF. *Angew. Chem. Int. Edit.* **2014,** *53*, 7524-7530.

75. Lin, S. J.; Van Lanen, S. G.; Shen, B., A free-standing condensation enzyme catalyzing ester bond formation in C-1027 biosynthesis. *Proc. Natl. Acad. Sci.* **2009,** *106*, 4183-4188.

76. Austin, M. B.; Saito, T.; Bowman, M. E.; Haydock, S.; Kato, A.; Moore, B. S.; Kay, R. R.; Noel, J. P., Biosynthesis of Dictyostelium discoideum differentiation-inducing factor by a hybrid type I fatty acid - type III polyketide synthase. *Nat. Chem. Biol.* **2006,** *2*, 494-502.

77. Miao, V.; Coeffet-LeGal, M.-F.; Brian, P.; Brost, R.; Penn, J.; Whiting, A.; Martin, S.; Ford, R.; Parr, I.; Bouchard, M., Daptomycin biosynthesis in Streptomyces *roseosporus*: cloning and analysis of the gene cluster and revision of peptide stereochemistry. *Microbiology* **2005,** *151*, 1507-1523.

78. Bentley, S. D.; Chater, K. F.; Cerdeno-Tarraga, A. M.; Challis, G. L.; Thomson, N. R.; James, K. D.; Harris, D. E.; Quail, M. A.; Kieser, H.; Harper, D.; Bateman, A.; Brown, S.; Chandra, G.; Chen, C. W.; Collins, M.; Cronin, A.; Fraser, A.; Goble, A.; Hidalgo, J.; Hornsby, T.; Howarth, S.; Huang, C. H.; Kieser, T.; Larke, L.; Murphy, L.; Oliver, K.; O'Neil, S.; Rabbinowitsch, E.; Rajandream, M. A.; Rutherford, K.; Rutter, S.; Seeger, K.; Saunders, D.; Sharp, S.; Squares, R.; Squares, S.; Taylor, K.; Warren, T.; Wietzorrek, A.; Woodward, J.; Barrell, B. G.; Parkhill, J.; Hopwood, D. A., Complete genome sequence of the model actinomycete *Streptomyces coelicolor* A3(2). *Nature* **2002,** *417*, 141-147.

79. Ikeda, H.; Ishikawa, J.; Hanamoto, A.; Shinose, M.; Kikuchi, H.; Shiba, T.; Sakaki, Y.; Hattori, M.; Omura, S., Complete genome sequence and comparative analysis of the industrial microorganism *Streptomyces avermitilis*. *Nat. Biotechnol.* **2003,** *21*, 526-531.

80. Challis, G. L., Exploitation of the *Streptomyces coelicolor* A3(2) genome sequence for discovery of new natural products and biosynthetic pathways. *J. Ind. Microbiol. Biot.* **2014,** *41*, 219-232.

81. Muir, P.; Li, S. T.; Lou, S. K.; Wang, D. F.; Spakowicz, D. J.; Salichos, L.; Zhang, J.; Weinstock, G. M.; Isaacs, F.; Rozowsky, J.; Gerstein, M., The real cost of sequencing: scaling computation to keep pace with data generation. *Genome Biol.* **2016,** *17*, 53.

82. Crawford, J. M.; Clardy, J., Microbial genome mining answers longstanding biosynthetic questions. *Proc. Natl. Acad. Sci.* **2012,** *109*, 7589-7590.

83. Oves-Costales, D.; Challis, G. L., Mining Microbial Genomes for Metabolic Products of Cryptic Pathways. *Drug Discov.* **2012,** *25*, 140-158.

84. Bachmann, B. O.; Van Lanen, S. G.; Baltz, R. H., Microbial genome mining for accelerated natural products discovery: is a renaissance in the making? *J. Ind. Microbiol. Biot.* **2014,** *41*, 175-184.

85. Kang, H. S., Phylogeny-guided (meta)genome mining approach for the targeted discovery of new microbial natural products. *J. Ind. Microbiol. Biot.* **2017,** *44*, 285-293.

86. Weber, T.; Kim, H. U., The secondary metabolite bioinformatics portal: Computational tools to facilitate synthetic biology of secondary metabolite production. *Synth. Syst. Biotechnol.* **2016,** *1*, 69-79.

87. Medema, M. H.; Blin, K.; Cimermancic, P.; de Jager, V.; Zakrzewski, P.; Fischbach, M. A.; Weber, T.; Takano, E.; Breitling, R., antiSMASH: rapid identification, annotation and analysis of secondary metabolite biosynthesis gene clusters in bacterial and fungal genome sequences. *Nucleic Acids Res.* **2011,** *39*, W339-W346.

88. Blin, K.; Medema, M. H.; Kazempour, D.; Fischbach, M. A.; Breitling, R.; Takano, E.; Weber, T., antiSMASH 2.0-a versatile platform for genome mining of secondary metabolite producers. *Nucleic Acids Res.* **2013,** *41*, W204-W212.

89. Weber, T.; Blin, K.; Duddela, S.; Krug, D.; Kim, H. U.; Bruccoleri, R.; Lee, S. Y.; Fischbach, M. A.; Muller, R.; Wohlleben, W.; Breitling, R.; Takano, E.; Medema, M. H., antiSMASH 3.0-a comprehensive resource for the genome mining of biosynthetic gene clusters. *Nucleic Acids Res.* **2015,** *43*, W237-W243.

90. Blin, K.; Wolf, T.; Chevrette, M. G.; Lu, X. W.; Schwalen, C. J.; Kautsar, S. A.; Duran, H. G. S.; Santos, E. L. C. D. L.; Kim, H. U.; Nave, M.; Dickschat, J. S.; Mitchell, D. A.; Shelest, E.; Breitling, R.; Takano,

E.; Lee, S. Y.; Weber, T.; Medema, M. H., antiSMASH 4.0-improvements in chemistry prediction and gene cluster boundary identification. *Nucleic Acids Res.* **2017,** *45*, W36-W41.

91. Ziemert, N.; Alanjary, M.; Weber, T., The evolution of genome mining in microbes - a review. *Nat. Prod. Rep.* **2016,** *33*, 988-1005.

92. Baltz, R. H., Genetic manipulation of secondary metabolite biosynthesis for improved production in *Streptomyces* and other actinomycetes. *J. Ind. Microbiol. Biot.* **2016,** *43*, 343-370.

93. Ongley, S. E.; Bian, X. Y.; Neilan, B. A.; Muller, R., Recent advances in the heterologous expression of microbial natural product biosynthetic pathways. *Nat. Prod. Rep.* **2013,** *30*, 1121-1138.

94. Tringe, S. G.; von Mering, C.; Kobayashi, A.; Salamov, A. A.; Chen, K.; Chang, H. W.; Podar, M.; Short, J. M.; Mathur, E. J.; Detter, J. C.; Bork, P.; Hugenholtz, P.; Rubin, E. M., Comparative metagenomics of microbial communities. *Science* **2005,** *308*, 554-557.

95. Roesch, L. F.; Fulthorpe, R. R.; Riva, A.; Casella, G.; Hadwin, A. K. M.; Kent, A. D.; Daroub, S. H.; Camargo, F. A. O.; Farmerie, W. G.; Triplett, E. W., Pyrosequencing enumerates and contrasts soil microbial diversity. *ISME J.* **2007,** *1*, 283-290.

96. Yamanaka, K.; Reynolds, K. A.; Kersten, R. D.; Ryan, K. S.; Gonzalez, D. J.; Nizet, V.; Dorrestein, P. C.; Moore, B. S., Direct cloning and refactoring of a silent lipopeptide biosynthetic gene cluster yields the antibiotic taromycin A. *Proc. Natl. Acad. Sci.* **2014,** *111*, 1957-62.

97. Fu, J.; Bian, X.; Hu, S.; Wang, H.; Huang, F.; Seibert, P. M.; Plaza, A.; Xia, L.; Muller, R.; Stewart, A. F.; Zhang, Y., Full-length RecE enhances linear-linear homologous recombination and facilitates direct cloning for bioprospecting. *Nat. Biotechnol.* **2012,** *30*, 440-6.

98. Jiang, W. J.; Zhao, X. J.; Gabrieli, T.; Lou, C. B.; Ebenstein, Y.; Zhu, T. F., Cas9-Assisted Targeting of CHromosome segments CATCH enables one-step targeted cloning of large gene clusters. *Nat. Comm.* **2015,** *6*, 8101.

99. Greunke, C.; Duell, E. R.; D'Agostino, P. M.; Glockle, A.; Lamm, K.; Gulder, T. A. M., Direct Pathway Cloning (DiPaC) to unlock natural product biosynthetic potential. *Metab. Eng.* **2018,** *47*, 334-345.

100. D'Agostino, P. M.; Gulder, T. A. M., Direct Pathway Cloning Combined with Sequence- and Ligation-Independent Cloning for Fast Biosynthetic Gene Cluster Refactoring and Heterologous Expression. *ACS Synth. Biol.* **2018,** *7*, 1702-1708.

101. Hu, S. B.; Liu, Z. Q.; Zhang, X.; Zhang, G. Y.; Xie, Y. L.; Ding, X. Z.; Mo, X. T.; Stewart, A. F.; Fu, J.; Zhang, Y. M.; Xia, L. Q., "Cre/loxP plus BAC": a strategy for direct cloning of large DNA fragment and its applications in Photorhabdus luminescens and Agrobacterium tumefaciens. *Sci. Rep.* **2016,** *6*, 29087.

102. Dai, R. X.; Zhang, B.; Zhao, G. P.; Ding, X. M., Site-specific recombination for cloning of large DNA fragments in vitro. *Eng. Life Sci.* **2015,** *15*, 655-659.

103. Du, D.; Wang, L.; Tian, Y.; Liu, H.; Tan, H.; Niu, G., Genome engineering and direct cloning of antibiotic gene clusters via phage φBT1 integrase-mediated site-specific recombination in Streptomyces. *Sci. Rep.* **2015,** *5*, 8740.

104. Kouprina, N.; Larionov, V., Selective isolation of genomic loci from complex genomes by transformation-associated recombination cloning in the yeast Saccharomyces cerevisiae. *Nat. Protoc.* **2008,** *3*, 371-377.

105. Yamanaka, K.; Reynolds, K. A.; Kersten, R. D.; Ryan, K. S.; Gonzalez, D. J.; Nizet, V.; Dorrestein, P. C.; Moore, B. S., Direct cloning and refactoring of a silent lipopeptide biosynthetic gene cluster yields the antibiotic taromycin A. *Proc. Natl. Acad. Sci.* **2014,** *111*, 1957-1962.

106. Yin, J.; Hoffmann, M.; Bian, X.; Tu, Q.; Yan, F.; Xia, L.; Ding, X.; Stewart, A. F.; Muller, R.; Fu, J.; Zhang, Y., Direct cloning and heterologous expression of the salinomycin biosynthetic gene cluster from *Streptomyces albus* DSM41398 in *Streptomyces coelicolor* A3(2). *Sci. Rep.* **2015,** *5*, 15081.

107. Wang, H.; Li, Z.; Jia, R.; Yin, J.; Li, A.; Xia, L.; Yin, Y.; Muller, R.; Fu, J.; Stewart, A. F.; Zhang, Y., ExoCET: exonuclease in vitro assembly combined with RecET recombination for highly efficient direct DNA cloning from complex genomes. *Nucleic Acids Res.* **2017,** *46*, e28-e28.

108. Duell, E. R.; D'Agostino, P. M.; Shapiro, N.; Woyke, T.; Fuchs, T. M.; Gulder, T. A., Direct pathway cloning of the sodorifen biosynthetic gene cluster and recombinant generation of its product in E. coli. *Microb. Cell Fact.* **2019,** *18*, 32.

109. Baltz, R. H., *Streptomyces* and *Saccharopolyspora* hosts for heterologous expression of secondary metabolite gene clusters. *J. Ind. Microbiol. Biot.* **2010,** *37*, 759-772.

110. Myronovskyi, M.; Rosenkranzer, B.; Nadmid, S.; Pujic, P.; Normand, P.; Luzhetskyy, A., Generation of a cluster-free *Streptomyces albus* chassis strains for improved heterologous expression of secondary metabolite clusters. *Metab. Eng.* **2018,** *49*, 316-324.

111. Gomez-Escribano, J. P.; Bibb, M. J., Engineering *Streptomyces coelicolor* for heterologous expression of secondary metabolite gene clusters. *Microb. Biotechnol.* **2011,** *4*, 207-215.

112. Galm, U.; Shen, B., Expression of biosynthetic gene clusters in heterologous hosts for natural product production and combinatorial biosynthesis. *Expert Opin. Drug Dis.* **2006,** *1*, 409-437.

113. Ahmadi, M. K.; Pfeifer, B. A., Recent progress in therapeutic natural product biosynthesis using Escherichia coli. *Curr. Opin. Biotech.* **2016,** *42*, 7-12.

114. Pfeifer, B. A.; Admiraal, S. J.; Gramajo, H.; Cane, D. E.; Khosla, C., Biosynthesis of complex polyketides in a metabolically engineered strain of E-coli. *Science* **2001,** *291*, 1790-1792.

115. Schmidt, E. W.; Nelson, J. T.; Rasko, D. A.; Sudek, S.; Eisen, J. A.; Haygood, M. G.; Ravel, J., Patellamide A and C biosynthesis by a microcin-like pathway in Prochloron didemni, the cyanobacterial symbiont of Lissoclinum patella. *Proc. Natl. Acad. Sci.* **2005,** *102*, 7315-7320.

116. Antosch, J.; Schaefers, F.; Gulder, T. A. M., Heterologous Reconstitution of Ikarugamycin Biosynthesis in E. coli. *Angew. Chem. Int. Edit.* **2014,** *53*, 3011-3014.

117. Godtfred.Wo; Vangedal, S.; Thomas, D. W., Cycloheptamycin, a New Peptide Antibiotic - Structure Determination by Mass Spectrometry. *Tetrahedron* **1970,** *26*, 4931-4946.

118. Qian, Z.; Antosch, J.; Wiese, J.; Imhoff, J.; Fiedler, H.-P.; Gulder, T. A.; Pöthig, A., Structures and Biological Activities of Cycloheptamycins A and B. *Org. Biomol. Chem.* **2019,** *17*, 6595-6600.

119. Qian, Z.; Bruhn, T.; M D'Agostino, P.; Herrmann, A.; Haslbeck, M.; Antal, N.; Fiedler, H.-P.; Brack-Werner, R.; Gulder, T. A. M., Discovery of the Streptoketides by Direct Cloning and Rapid Heterologous Expression of a Cryptic PKS II Gene Cluster from *Streptomyces* sp. Tü6314. *J. Org. Chem.* **2019,** DOI:10.1021/acs.joc.9b02741.

120. Low, Z. J.; Pang, L. M.; Ding, Y.; Cheang, Q. W.; Hoang, K. L. M.; Tran, H. T.; Li, J.; Liu, X.-W.; Kanagasundaram, Y.; Yang, L., Identification of a biosynthetic gene cluster for the polyene macrolactam sceliphrolactam in a *Streptomyces* strain isolated from mangrove sediment. *Sci. Rep.* **2018,** *8*, 1594.

121. McClure, R. A.; Goering, A. W.; Ju, K.-S.; Baccile, J. A.; Schroeder, F. C.; Metcalf, W. W.; Thomson, R. J.; Kelleher, N. L., Elucidating the rimosamide-detoxin natural product families and their biosynthesis using metabolite/gene cluster correlations. *ACS Chem. Biol.* **2016,** *11*, 3452-3460.

122. Blin, K.; Shaw, S.; Steinke, K.; Villebro, R.; Ziemert, N.; Lee, S. Y.; Medema, M. H.; Weber, T., antiSMASH 5.0: updates to the secondary metabolite genome mining pipeline. *Nucleic Acids Res.* **2019**, *47*, W81–W87.

List of Abbreviations

A domain	adenylation domain
ACP	acyl carrier protein
ARO	aromatase
AT	acyltransferase
ATP	Adenosine triphosphate
BGC	biosynthetic gene cluster
C domain	condensation domain
CoA	Coenzyme A
Cy domain	heterocyclization domain
CYC	cyclase
DH	dehydratase
E domain	epimerization domain
ER	enoyl reductase
F domain	formylation domain
HPLC	high-performance liquid chromatography
KR	ketoreductase
KS	ketosynthase
KSα	ketosynthase alpha
KSβ	ketosynthase beta
LC-MS	liquid chromatography coupled to mass spectrometry
LCHR	linear plus circular homologous recombination
LLHR	linear plus linear homologous recombination
MT domain	methyltransferase domain
NMR	nuclear magnetic resonance spectroscopy
NRPS	non-ribosomal peptide synthetase
OMT domain	O-methyltransferase
Ox domain	oxidative domain
PCP domain	peptidyl carrier protein domain
PCR	Polymerase chain reaction
PKS	Polyketide synthase
PPTase	Phosphopantetheinyl transferase
R domain	reduction domain
SAM	*S*-adenosylmethionine
TE	thioesterase

Appendix

Supplemental materials of publications

S I. Supplemental information for cycloheptamycins A and B

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

Z. Qian,* J. Antosch,* J. Wiese, J.F. Imhoff, H.-P. Fiedler, A. Pöthig, T.A.M. Gulder. Structures and biological activities of cycloheptamycins A and B, *Org. Biomol. Chem.* **2019**, *17*, 6595-6600.

*equal contribution

Available online:

https://pubs.rsc.org/en/content/articlelanding/2019/ob/c9ob01261c/unauth#!divAbstract

Structures and Biological Activities of Cycloheptamycins A and B

Zhengyi Qian,[a,†] Janine Antosch,[a,†] Jutta Wiese,[b] Johannes F. Imhoff,[b] Hans-Peter-Fiedler,[c] Alexander Pöthig,[d,*] and Tobias A. M. Gulder[a,e,*]

a. Biosystems Chemistry, Department of Chemistry and Center for Integrated Protein Sciences Munich, Technical University of Munich, Lichtenbergstraße 4, 85748 Garching, Germany.

b. GEOMAR Helmholtz Center for Ocean Research Kiel, RD3 Marine Microbiology, Düsternbrooker Weg 20, 24105 Kiel, Germany.

c. Institute of Microbiology, University of Tübingen, Auf der Morgenstelle 28, D-72076 Tübingen, Germany.

d. Department of Chemistry and Catalysis Research Center (CRC), Technical University of Munich, Ernst-Otto-Fischer-Straße 1, 85748 Garching, Germany

e. Chair of Technical Biochemistry, Technical University of Dresden, Bergstraße 66, 01069 Dresden, Germany

* Corresponding author: tobias.gulder@tu-dresden.de; alexander.poethig@tum.de

† These authors contributed equally to this work.

Supporting Information

Contents

Crystal Structure Determination of Cycloheptamycin A (1)

Data acquisition

After a significant number of crystallization attempts employing different conditions a crystalline sample of compound **1** suitable for single crystal X-ray diffraction was obtained by slow evaporation of a sample in a 5 mM ammonium acetate solution in methanol. Intensities were collected on a Bruker D8-Venture diffractometer equipped with a CMOS detector (Bruker Photon 100), an TXS rotating anode source with Mo Kα radiation (λ = 0.71073 Å) and a Helios mirror optic. The crystals were coated in perfluoropolyether and mounted in the cooled nitrogen stream (100 K) of the diffractometer on a microsampler. Diffraction data was processed with APEX III [1] and the implemented SAINT and SADABS programs. [2a, 2b] Molecular structures were solved within APEX III using SHELXT [3] and refined with SHELXL-2017 [4] in conjunction with SHELXLE [5]. Peptide residues were placed with help of the DSR program [6] refined using various restraints for geometric and displacement parameters.

Space group determination and model refinement

First, the compound was determined to crystallize in orthorhombic space group $P\,2_1\,2_1\,2_1$ with a unit cell of a = 20.8469(10) Å, b = 21.2672(11) Å, c = 23.8512(11) Å, V = 10575 Å^3. Data reduction and correction yielded satisfying R(int) = 0.0514 and R(sigma) = 0.0226. Two independent molecules were identified in the asymmetric unit and could be refined until no significant residual electron density was present in the difference Fourier map. However, the model only yielded a comparably high R1 = 0.1322 and a wR2 = 0.3793. We therefore checked for symmetry reduction in combination with application of the twin law [-1 0 0 0 -1 0 0 0 1]. After according data reduction and correction (triclinic), yielding respective R(int) = 0.0586 and R(sigma) = 0.0552, and the structure was also be solved in spacegroup $P\,1$ and refined with 8 independent molecules in the asymmetric unit. Additionally, the PLATON SQUEEZE procedure [7] was applied, to treat residual electron density deriving from disordered solvent molecules which could not be refined. The respective R1 amounted to a value of 0.1004 and wR2 = 0.3330, additionally exhibiting convergence issues. Similar results were obtained when trying to refine the structure in spacegroup $P\,2_1$.

A detailed inspection of the reflections in the reciprocal space revealed very low intensities of every second reflection in **b*** direction. Furthermore, these reflection were only observed at lower resolution, so we suspected $\lambda/2$ contributions. We therefore reindexed only the strong reflections, yielding a smaller unit cell (a = 20.872(5) Å, b = 10.637(3) Å, c=23.884(5) Å, b = 89.915(5)°) and respective R(int) = 0.0881 and R(sigma) = 0.0336 after data reduction and correction. We were able to refine a model of the crystal structure in spacegroup $P\,2$ (with application of a found twin law [-1 0 0 0 -1 0 0 0 1]) with two independent molecules in the asymmetric unit, however, also exhibiting convergence issues and a comparably high R1 = 0.1242 and wR2 = 0.3522.

Since the main information, which we aimed for by determining the crystal structure – namely the proof of connectivity and relative configuration of the oligopeptide – could already been derived from all the solutions we obtained in all tested spacegroups, we refrained from further more sophisticated modelling attempts (e.g. finding and refining possible modulated

structures). The data provided is that of the solution in spacegroup $P\,1$ and we only use the molecular model as structural proof without discussing geometrical parameters in detail.

Details for structure determination and refinement in spacegroup P 1:

A total of 1497 frames were collected. The total exposure time was 11.03 hours. The frames were integrated with the Bruker SAINT software package using a narrow-frame algorithm. The integration of the data using a triclinic unit cell yielded a total of 170060 reflections to a maximum θ angle of 21.97°, of which 50951 were independent (average redundancy 3.597, completeness = 99.9%, R_{int} = 3.43%, R_{sig} = 3.19%) and 39534 were greater than $2\sigma(F^2)$. The final cell constants were determined as: $\underline{a}$ = 20.847(1) Å, $\underline{b}$ = 21.2683(11) Å, $\underline{c}$ = 23.8537(11) Å, α = 90.0400(16), β = 90.0581(14), γ = 89.9976(16),volume = 10576.3(9) Å^3. Data were corrected for absorption effects using the Multi-Scan method (SADABS). The calculated minimum and maximum transmission coefficients (based on SADABS) are 0.7085 and 0.7452.

The structure was solved using SHELXT and refined using SHELXL in conjunction with SHELXLE, for the space group $P\,1$. The final anisotropic full-matrix least-squares refinement on F^2 with 5025 variables converged at R1 = 0.1161%, for the observed data and wR2 = 0.3711% for all data. The goodness-of-fit was 1.585. The largest peak in the final difference electron density synthesis was 0.870 e$^-$/Å^3 and the largest hole was -0.337 e$^-$/Å^3 with an RMS deviation of 0.074 e$^-$/Å^3. On the basis of the final model, the calculated density was 1.152 g/cm^3 and F(000), 4064 e$^-$. PLATON SQUEEZE was used to remove residual electron density originating from co-crystallizing water molecules (see CIF).

Chemical formula	$C_{48}H_{68}N_8O_{12}$	
Formula weight	949.50 g/mol	
Temperature	100(2) K	
Wavelength	0.71073 Å	
Crystal size	0.096 x 0.150 x 0.330 mm	
Crystal habit	clear pale yellow fragment	
Crystal system	triclinic	
Space group	$P\,1$	
Unit cell dimensions	a = 20.847(1) Å	α = 90.0400(16)°
	b = 21.2683(11) Å	β = 90.0581(14)°
	c = 23.8537(11) Å	γ = 89.9976(16)°
Volume	10576.3(9) Å^3	
Z	8	
Density (calculated)	1.192 g/cm^3	
Absorption coefficient	0.086 mm^{-1}	
F(000)	4064	

Theta range for data collection	2.18 to 21.97° (as integrated)
Index ranges	-21<=h<=21, -22<=k<=22, -25<=l<=25
Reflections collected	170060
Independent reflections	50951 [R(int) = 0.0343]
Coverage of independent reflections	99.9%
Absorption correction	Multi-Scan
Max. and min. transmission	0.7085 and 0.7452
Structure solution technique	Intrinsic phasing
Structure solution program	XT, VERSION 2014/4
Refinement method	Full-matrix least-squares on F^2
Refinement program	SHELXL-2014/7 (Sheldrick, 2014)

84

Function minimized	$\Sigma\, w(F_o^2 - F_c^2)^2$
Data / restraints / parameters	50951/ 12323/ 5025
Goodness-of-fit on F^2	1.585
Final R indices	39534 data; I>2σ(I) R1 = 0.1161, wR2 = 0.3425
	all data R1 = 0.1380, wR2 = 0.3711
Weighting scheme	w=1/[σ^2(F$_o^2$)+(0.2000P)^2] where P=(F$_o^2$+2F$_c^2$)/3
Largest diff. peak and hole	1.152 and -0.485 eÅ^{-3}
R.M.S. deviation from mean	0.088 eÅ^{-3}

Crystallographic details are provided in the supporting information, as is crystallographic data in cif format. CCDC 1900571 contain the supplementary data for this paper. These data can be obtained free of charge from the Cambridge Crystallographic Data Centre via http://www.ccdc.cam.ac.uk/data_request/cif.

85
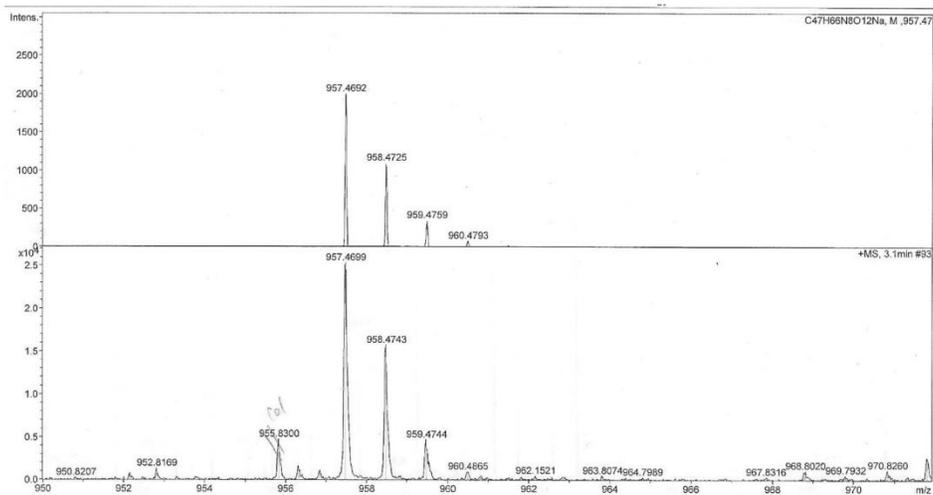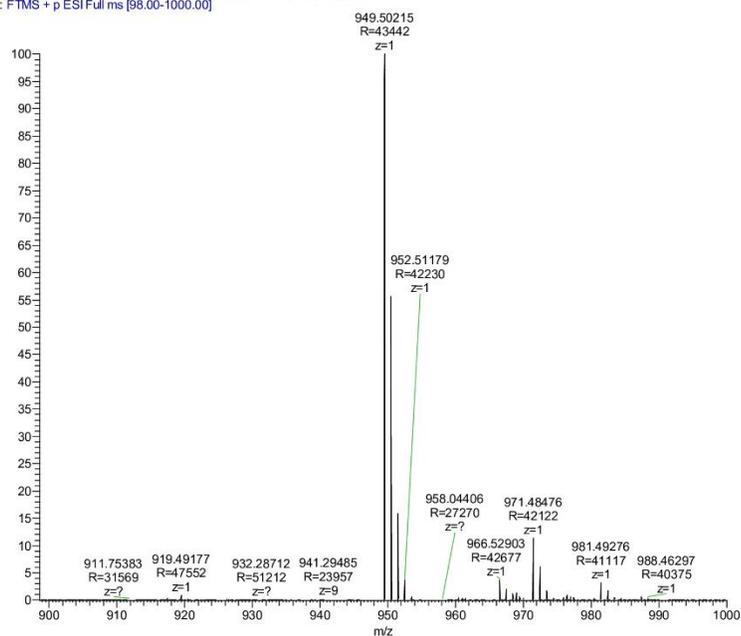
Figure S1. HRMS data of **1** (top; FT-ESI-MS, positive) and **2** (bottom; micrOTOF-Q ESI-MS, positive).
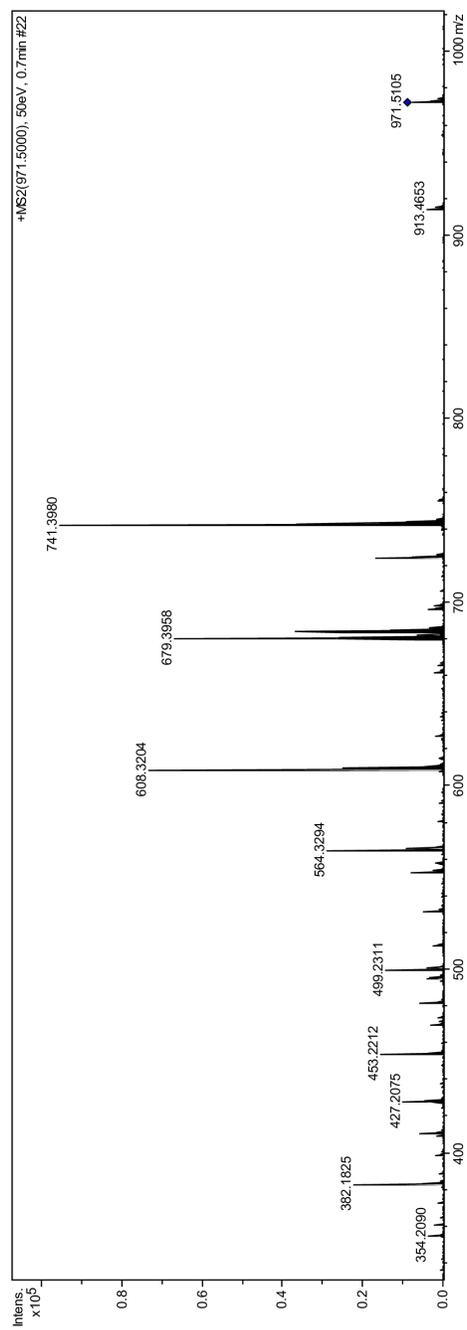
Figure S2. MS/MS data of **1**.
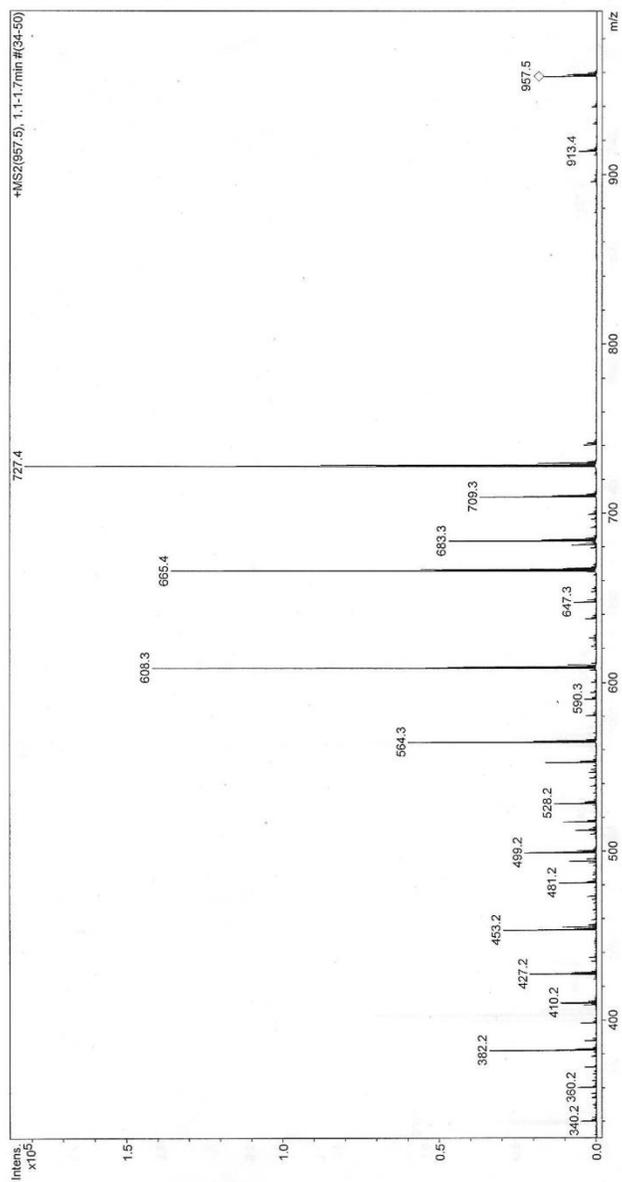
Figure S3. MS/MS data of **2**.
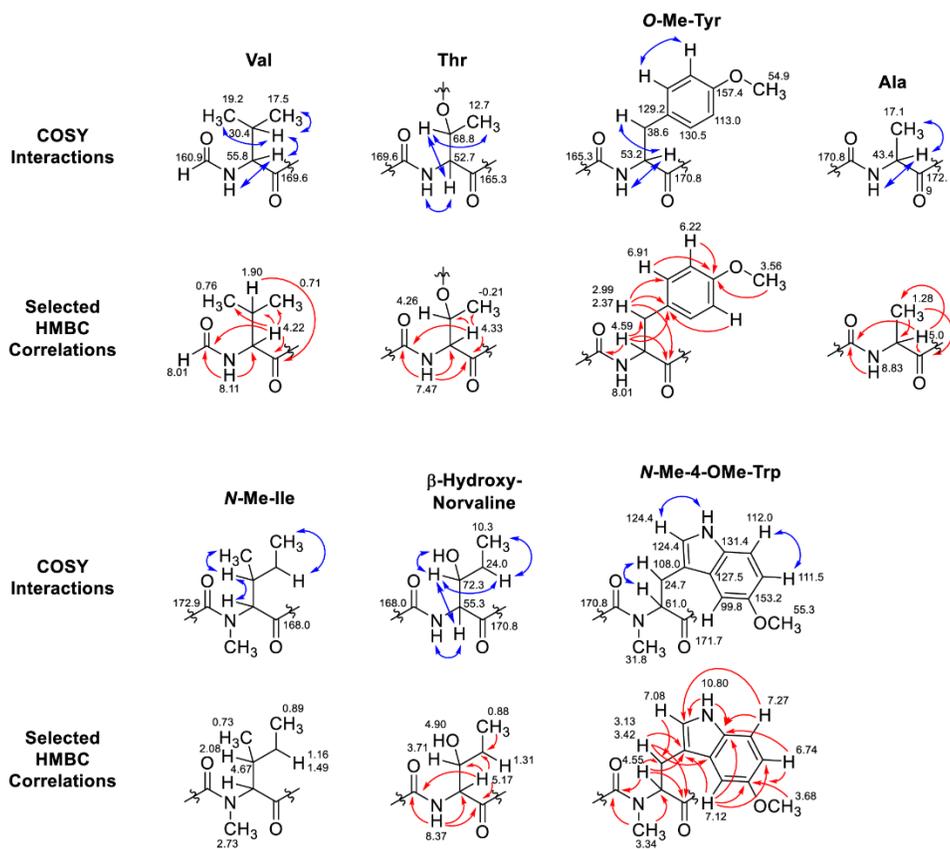
Figure S4. NMR structure determination of the individual amino acid building blocks in cycloheptamycin A (**1**).
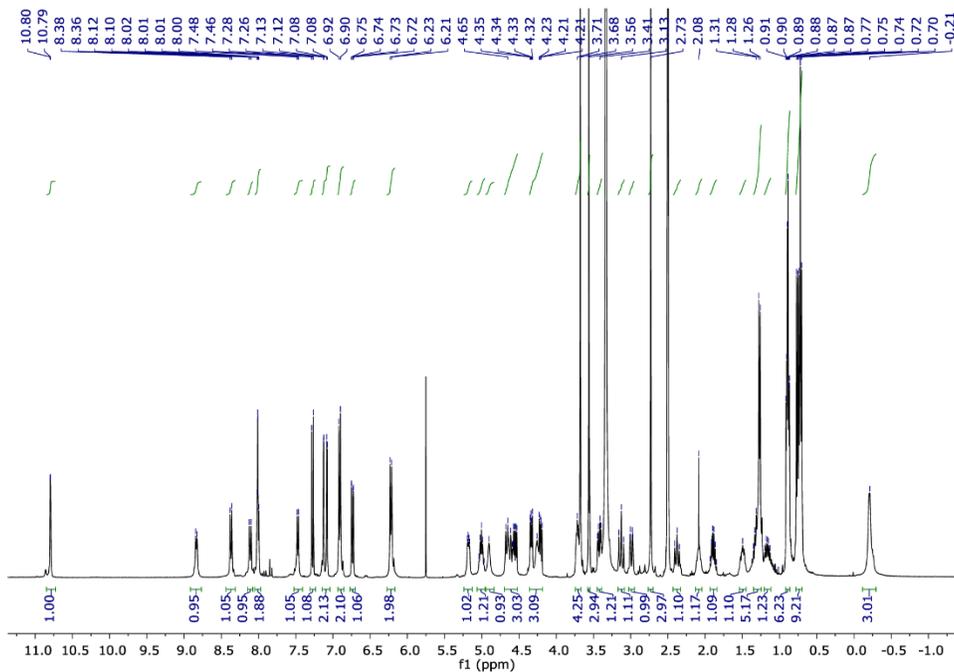
Figure S5. ^{1}H NMR spectrum (400 MHz) of cycloheptamycin A (**1**) in DMSO-d$_6$.
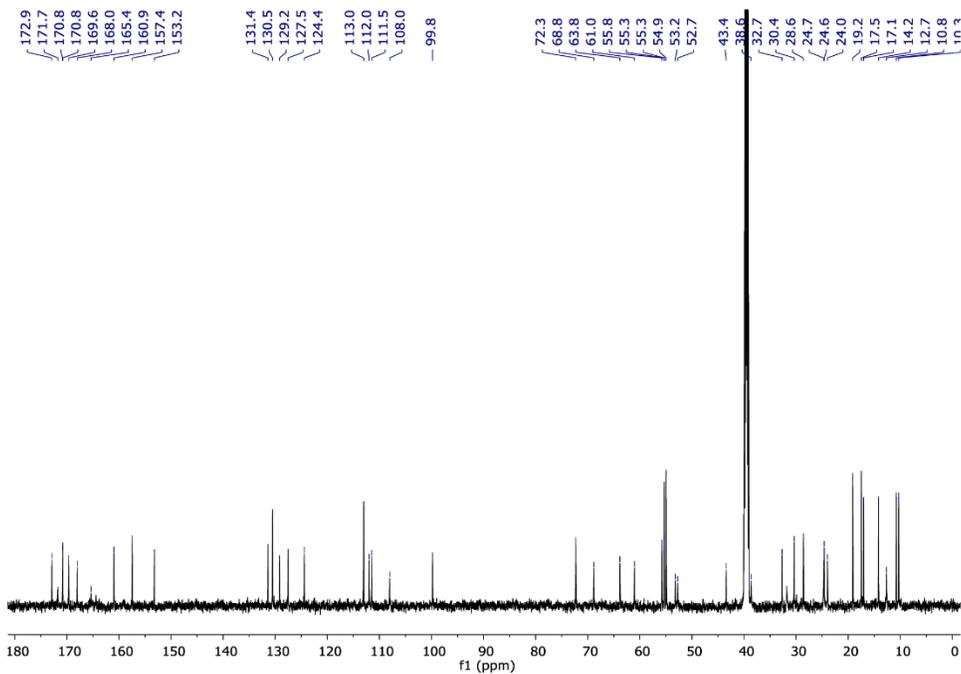
Figure S6. ^{13}C NMR spectrum (100 MHz) of cycloheptamycin A (**1**) in DMSO-d$_6$.
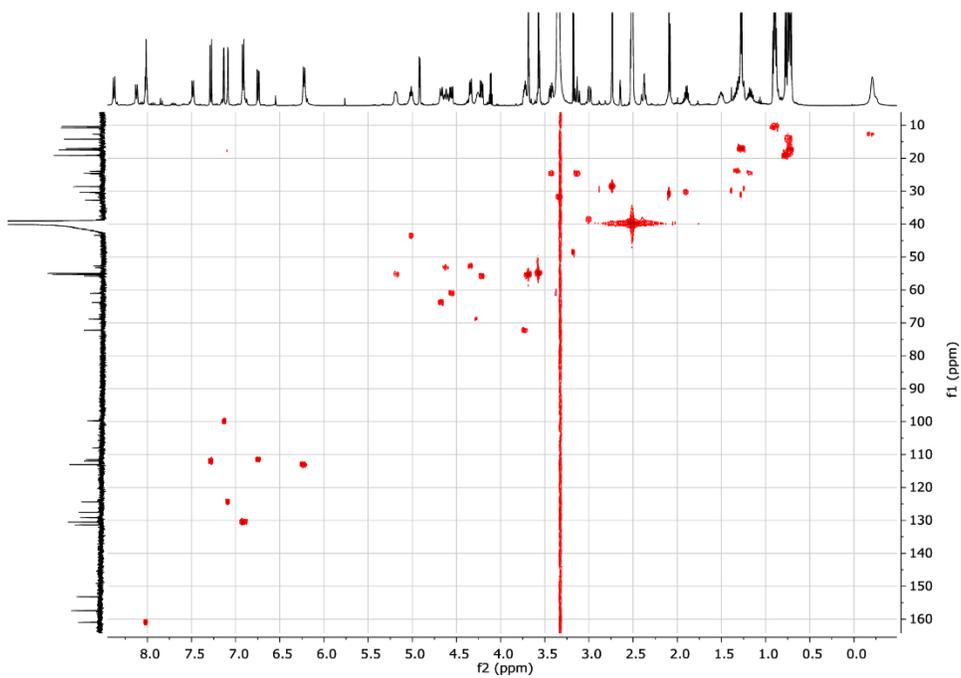
Figure S7. 2D-HSQC NMR spectrum of cycloheptamycin A (**1**) in DMSO-d$_6$.
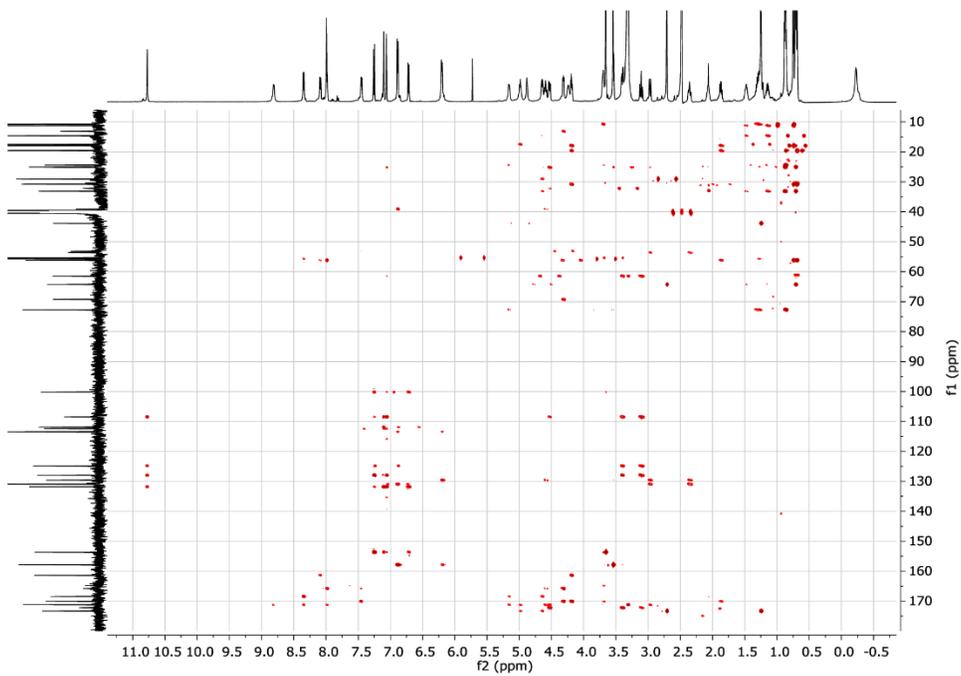
Figure S8. 2D-HMBC NMR spectrum of cycloheptamycin A (**1**) in DMSO-d$_6$.

91
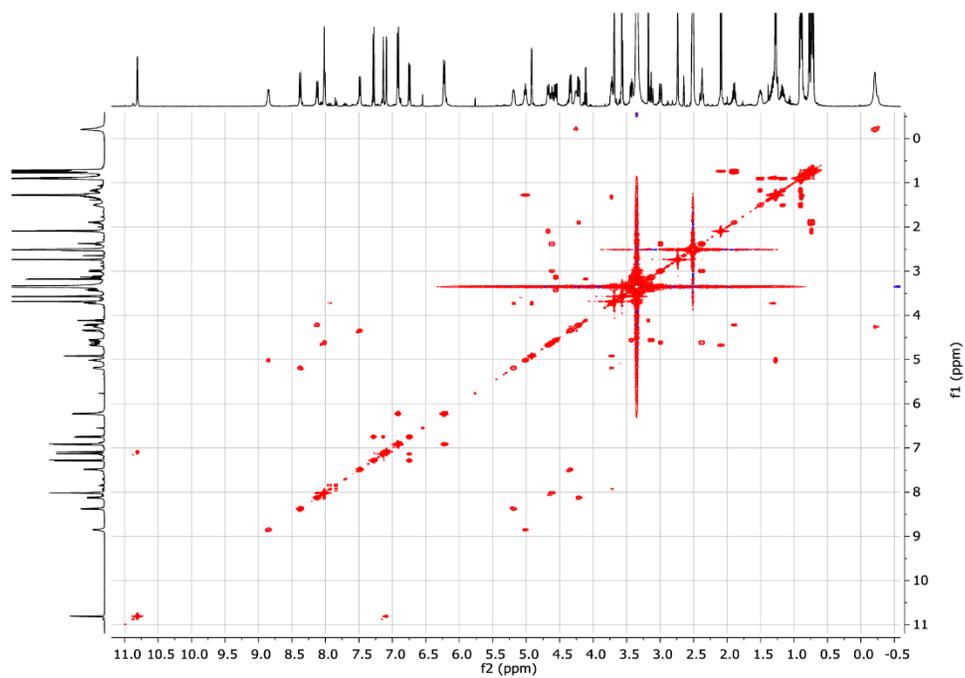
Figure S9. 2D-COSY NMR spectrum of cycloheptamycin A (**1**) in DMSO-d$_6$.
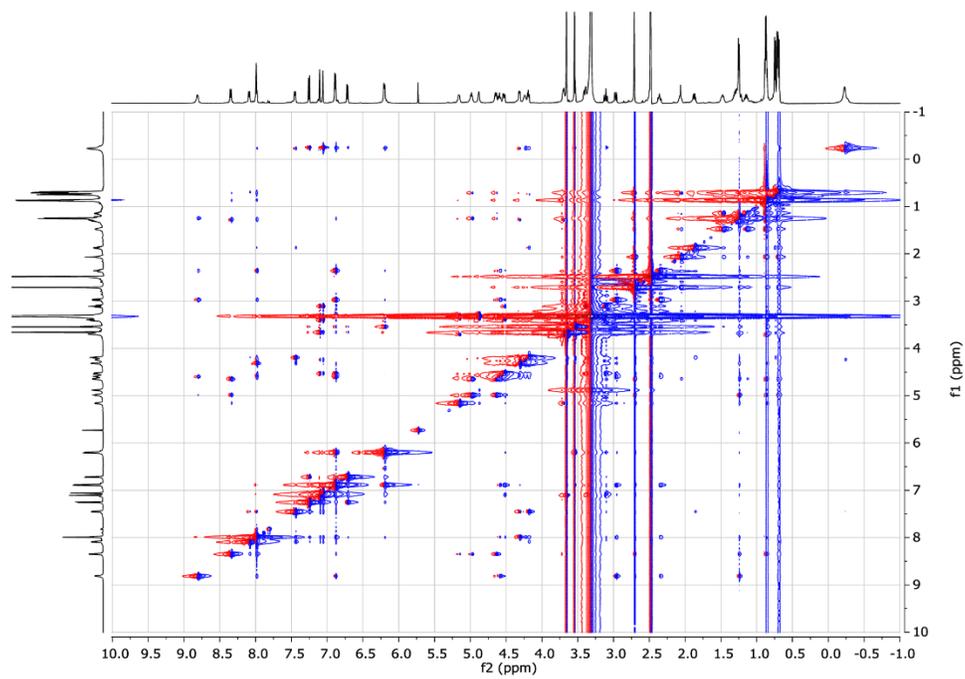
Figure S10. 2D-NOESY NMR spectrum of cycloheptamycin A (**1**) in DMSO-d$_6$.
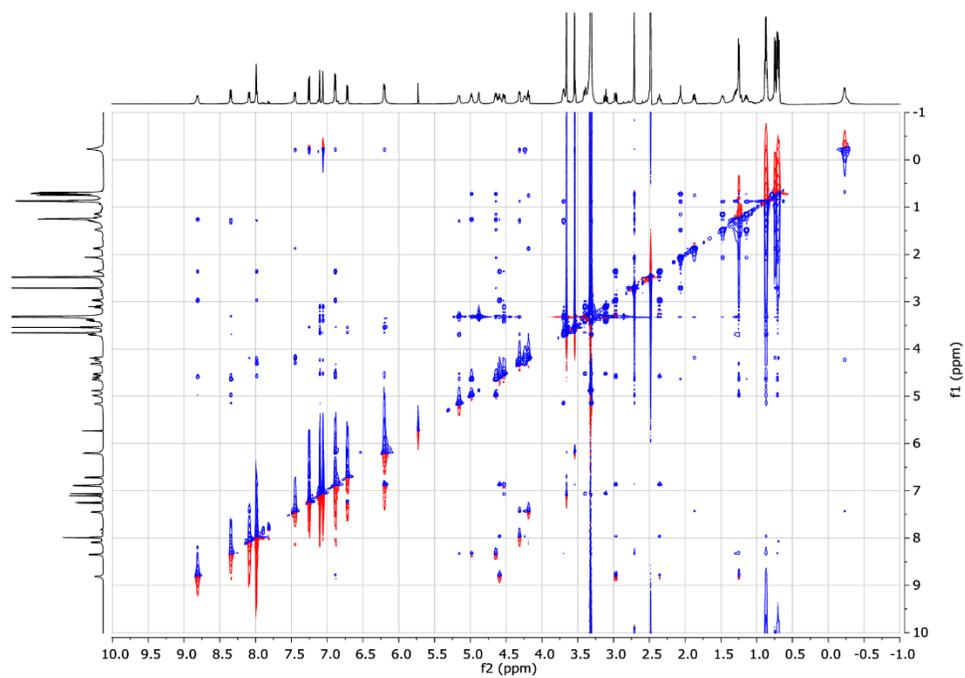
Figure S11. 2D-NOESY NMR (adjusted processing) spectrum of cycloheptamycin A (**1**) in DMSO-d$_6$.

Table S1. NMR data of cycloheptamycin B (**2**) at 500 MHz (^{1}H) and 125 MHz (^{13}C) in DMSO-d_6. The position numbering was kept identical to Cycloheptamycin A (**1**) for better comparability. Compound **2** thus does not contain C-19.

Amino acid	Signal	^{1}H [ppm]	COSY, Mult. [J in Hz]	^{13}C [ppm]	HMBC
tryptophane	1			171.9 (C)	
	2	4.52	3, dd [12.3, 5.1]	61.1 (CH)	1, 3, 4
	3a	3.43	2, 3b, dd, [13.0, 5.1]	24.7 (CH$_2$)	1, 2, 4, 5, 12
	b	3.10	2, 3a, pt [12.6]		
	4			108.0 (C)	
	5			127.5 (C)	
	6	7.13	8, d [2.4]	99.7 (CH)	4, 7, 8, 10
	7			153.2 (C)	
	8	6.74	6, 9, dd [8.7, 2.4]	111.5 (CH)	6, 7, 10
	9	7.27	8, d [8.7]	112.0 (CH)	5, 6, 7, 10, 12
	10			131.4 (C)	
	11	10.79	12, d [2.4]	(NH)	4, 5, 10, 12
	12	7.07	11, d [2.4]	124.4 (CH)	3, 4, 5, 7, 10
	13	3.68	s	55.3 (CH$_3$)	7
	14	3.34	s	31.7 (CH$_3$)	2, 15
threonine	15			170.4 (C)	
	16	5.21	17, 21, m	55.5 (CH)	15, 17, 18, 22
	17	4.04	16, 18, m	66.6 (CH)	16, 18
	18	0.95	17, d [6.2]	16.9 (CH$_3$)	17
	(19)				
	20	4.88	17, bs	(OH)	
	21	8.36	16, d [8.6]	(NH)	15, 16, 22
isoleucine	22			168.0 (C)	
	23	4.71	24, d [11.1]	63.8 (CH)	22, 24, 25, 28, 29
	24	2.07	23, 27, m	32.7 (CH)	23, 25, 26
	25a	1.53	25b, 26, m	24.6 (CH$_2$)	23, 24, 26, 27
	25b	1.17	25a, 26, m		
	26	0.90	25, t [7.3]	10.8 (CH$_3$)	24, 25
	27	0.73	24, d [7.3]	14.1 (CH$_3$)	23, 24, 25
	28	2.73	s	28.6 (CH$_3$)	23, 29
alanine	29			172.9 (C)	
	30	5.02	31, 32, m	43.3 (CH)	29, 31, 33
	31	1.27	30, d [6.5]	17.1 (CH$_3$)	29, 30
	32	8.86	30, d [7.8]	(NH)	30, 31, 33
tyrosine	33			170.8 (C)	
	34	4.61	35a/b, 41, m	53.1 (CH)	33, 35, 36, 42
	35a	2.98	34, 35b, m	38.7 (CH$_2$)	33, 34, 36, 37
	35b	2.36	34, 35a, m		
	36			129.1 (C)	
	37	6.91	38, d [8.3]	130.5 (CH)	35, 36, 38, 39
	38	6.21	37, d [8.3]	113.0 (CH)	36, 40
	39			157.4 (C)	
	40	3.56	s	54.9 (CH$_3$)	39
	41	8.06	34, d [9.5]	(NH)	33, 42
threonine	42			165.3 (C)	
	43	4.35	46, dd [8.1, 4.0]	52.6 (CH)	42, 44, 45, 47
	44	4.20	45, m	69.0 (CH)	43
	45	-0.24	44, bs	12.6 (CH$_3$)	44
	46	7.47	43, d [8.1]	(NH)	42, 43, 47
N-formyl valine	47			169.6 (C)	
	48	4.19	49, 52, dd [9.1, 5.8]	55.7 (CH)	47, 49, 51, 53
	49	1.88	48, 50, 51, m	30.3 (CH)	47, 48, 50, 51
	50	0.76	49, d [6.8]	19.1 (CH$_3$)	48, 49, 51
	51	0.70	49, d [6.8]	17.5 (CH$_3$)	48, 49, 50
	52	8.12	48, d [9.1]	(NH)	48, 53
	53	8.00	s	160.9 (COH)	48
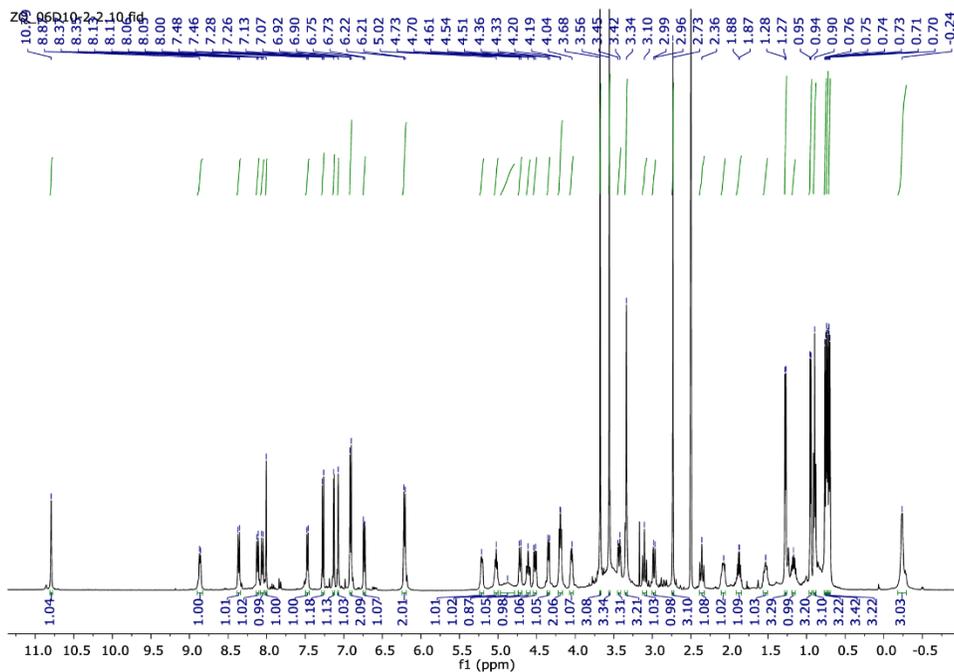
Figure S12. ^{1}H NMR spectrum (500 MHz) of cycloheptamycin B (**2**) in DMSO-d$_6$.
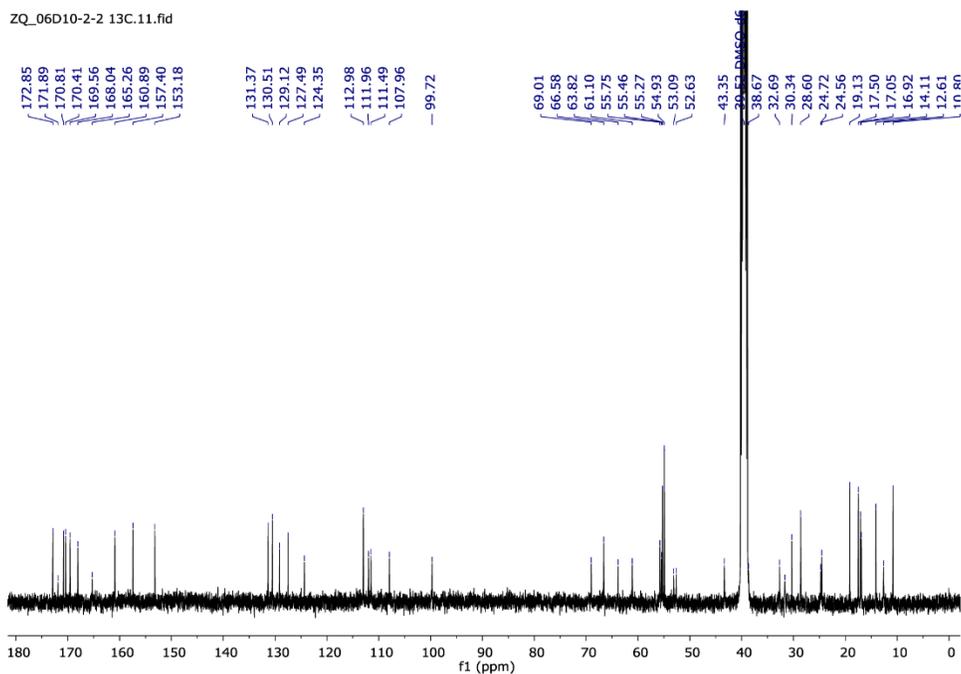
Figure S13. ^{13}C NMR spectrum (125 MHz) of cycloheptamycin B (**2**) in DMSO-d$_6$.
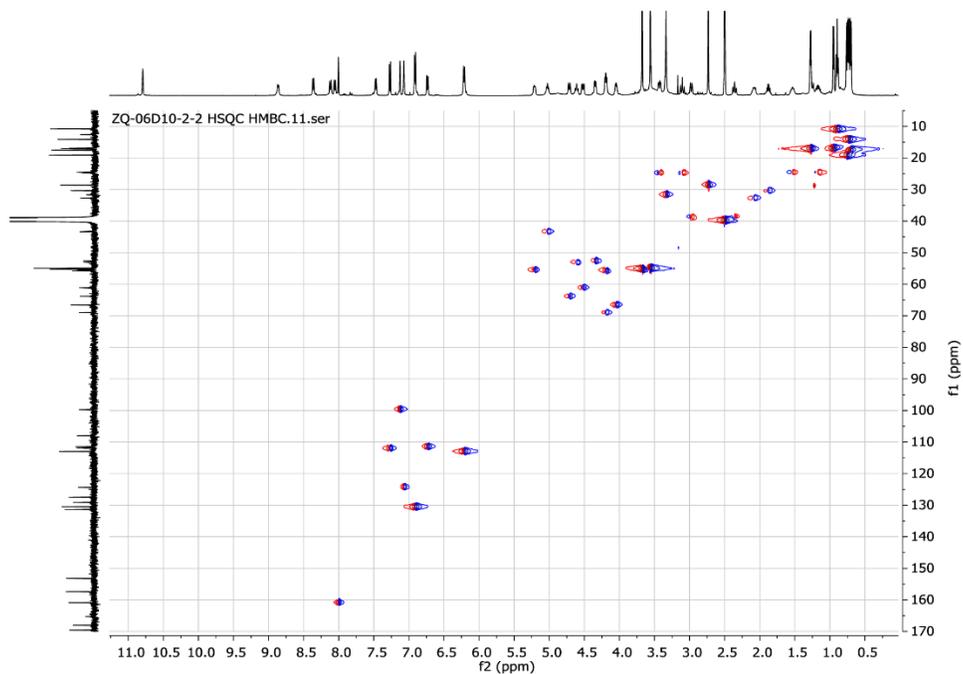
Figure S14. 2D-HSQC NMR spectrum of cycloheptamycin B (**2**) in DMSO-d$_6$.
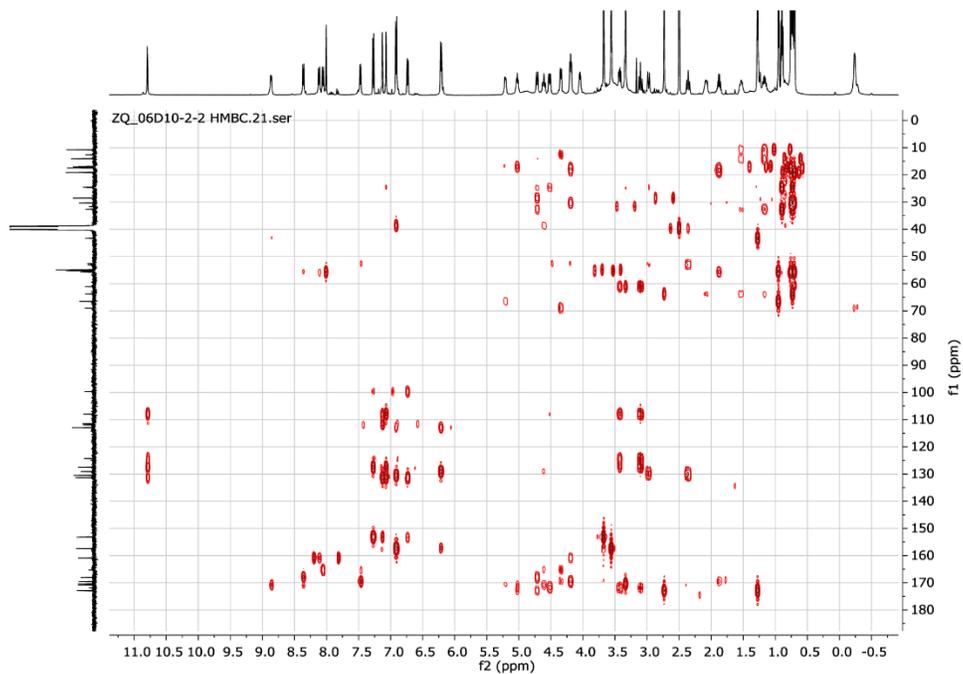
Figure S15. 2D-HMBC NMR spectrum of cycloheptamycin B (**2**) in DMSO-d$_6$.
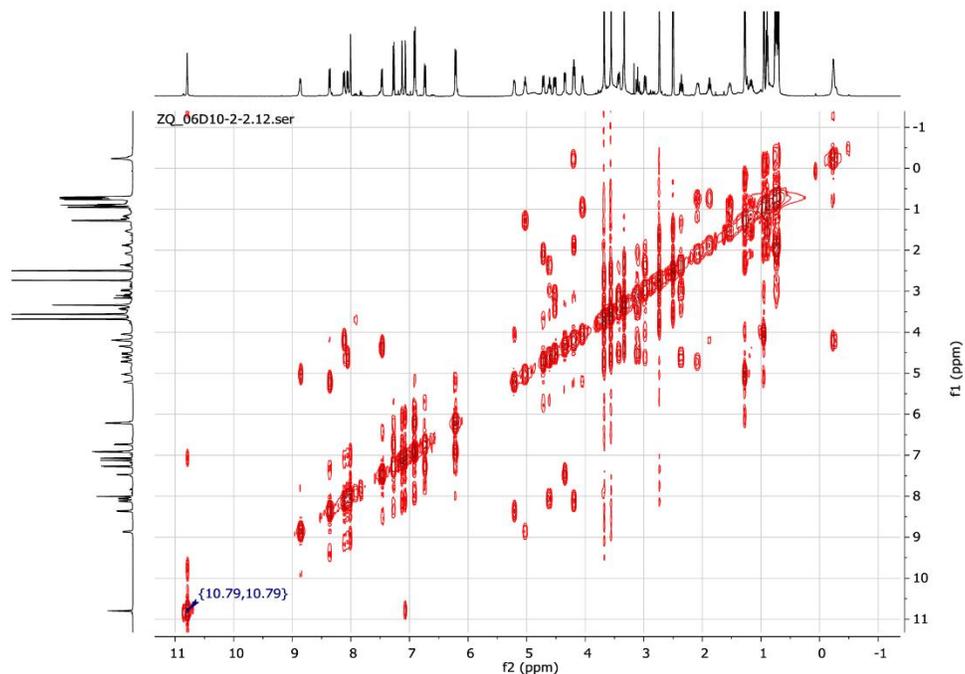
Figure S16. 2D-COSY NMR spectrum of cycloheptamycin B (**2**) in DMSO-d_6.
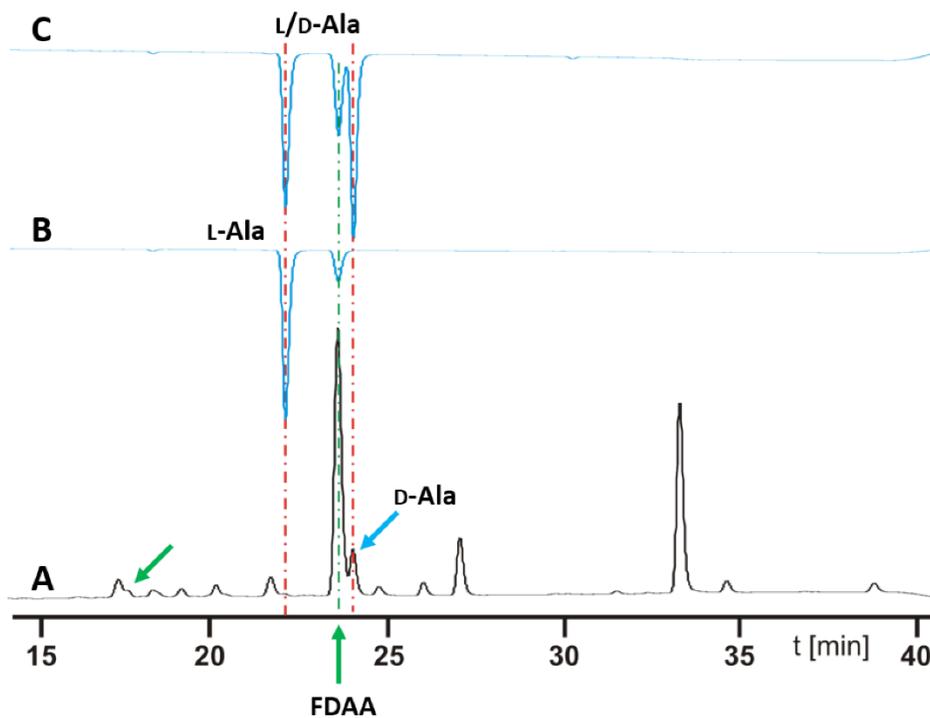
Figure S17. Determination of the absolute configuration of the Ala building block in **1** by HPLC-ESIMS analysis of the FDAA-functionalized peptide hydrolysate (**A**) in comparison to FDAA-functionalized synthetic standards of L-Ala (**B**) and D/L-Ala (**C**) using Marfey's method[8,9] revealing the Ala building block to be L-configured.

97
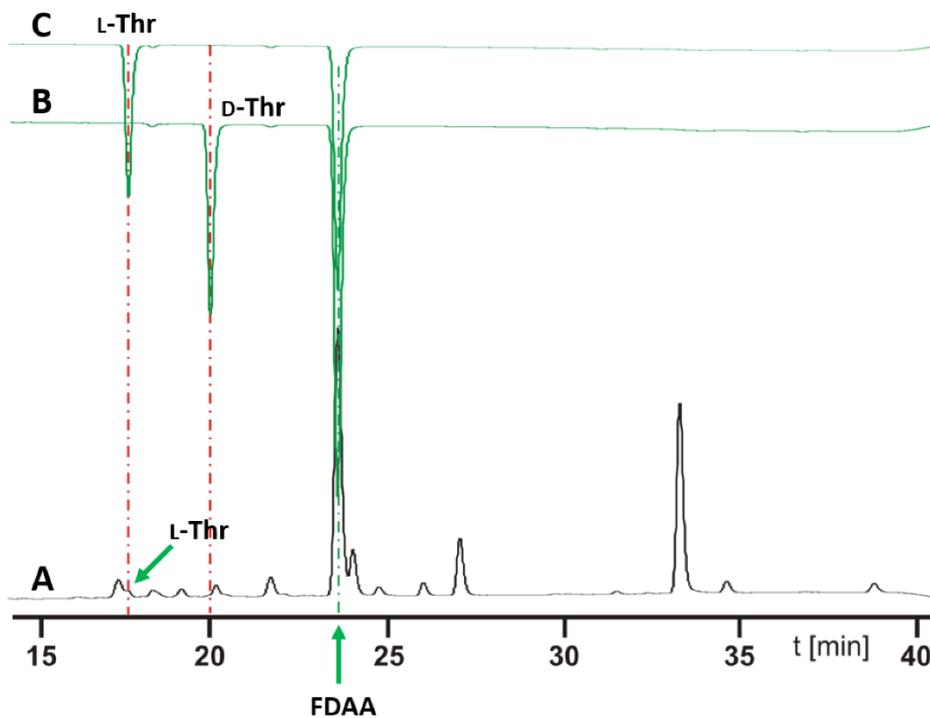
Figure S18. Determination of the absolute configuration of the Thr building block in **1** by HPLC-ESIMS analysis of the FDAA-functionalized peptide hydrolysate (**A**) in comparison to FDAA-functionalized synthetic standards of D-Thr (**B**) and L-Thr (**C**) using Marfey's method[8,9] revealing the Thr building block to be L-configured.
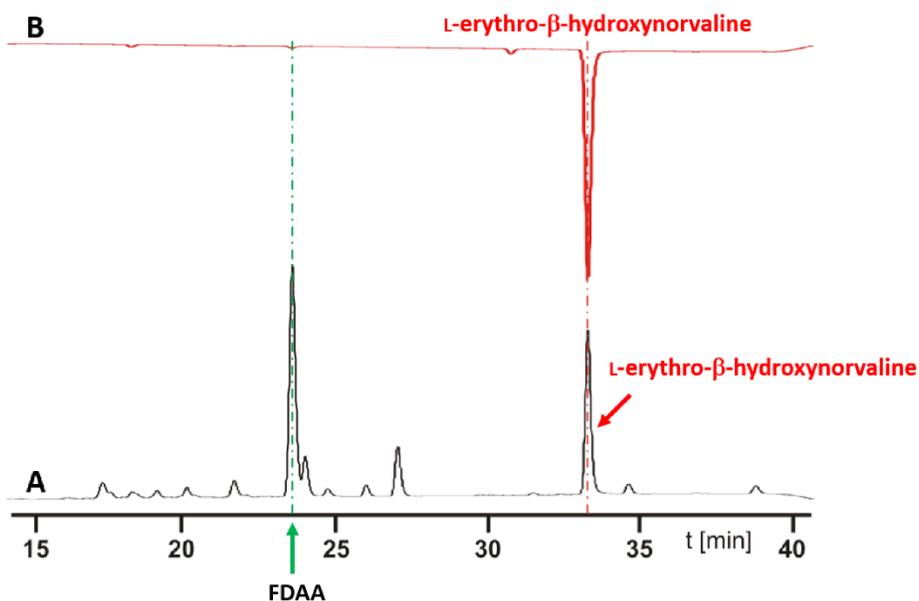
Figure S19. Determination of the absolute configuration of the β-hydroxynorvaline building block in **1** by HPLC-ESIMS analysis of the FDAA-functionalized peptide hydrolysate (**A**) in comparison to FDAA-functionalized synthetic standards of L-erythro-β-hydroxynorvaline (**B**) using Marfey's method[8,9] revealing this building block to be L-configured.

98
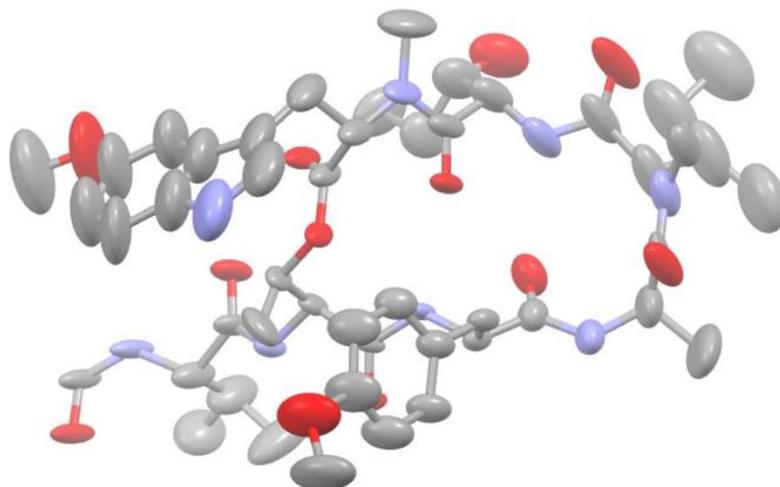
Figure S20. Molecular structure of compound **1** in the solid state (one independent molecule is shown). Ellipsoids are drawn at 50% probability. Hydrogen atoms are omitted for clarity.

References:

1. Bruker, APEX suite of crystallographic software, APEX 3 Version 2015-5.2, Bruker AXS Inc., Madison, Wisconsin, USA, 2015.
2. Bruker, SAINT, Version 8.27b and Version 8.34A and SADABS, Version 2012/1 and Version 2014/5, Bruker AXS Inc., Madison, Wisconsin, USA, 2014.
3. Sheldrick, G.M. *Acta Cryst. A* **2015**, *71*, 3-8.
4. Sheldrick, G.M. *Acta Cryst. C* **2015**, *71*, 3-8.
5. Huebschle, C.B.; Sheldrick, G.M.; Dittrich, B. *J. Appl. Cryst.* **2011**, *44*, 1281-1284.
6. Kratzert, D; Krossing, I. *J. Appl. Cryst.* **2018**, *51*, 928-934.
7. Spek, A.L. *Acta Cryst.* **2009**, *D65*, 148-155.
8. Marfey, P. *Carlsberg Res. Commun.* **1984**, *49*, 591–596.
9. Bhushan, R.; Brückner, H. *Amino Acids* **2004**, *27*, 231–247.

S II. Supplemental information for streptoketides

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

Z. Qian, T. Bruhn, P.M. D'Agostino, A. Herrmann, M. Haslbeck, N. Antal, H.-P. Fiedler, R. Brack-Werner, T.A.M. Gulder. Discovery of the Streptoketides by Direct Cloning and Rapid Heterologous Expression of a Cryptic PKS II Gene Cluster from *Streptomyces* sp. Tü6314, *J. Org. Chem.* **2019**, DOI:10.1021/acs.joc.9b02741

Available online: https://pubs.acs.org/doi/10.1021/acs.joc.9b02741

SUPPORTING INFORMATION

Discovery of the Streptoketides by Direct Cloning and Rapid Heterologous Expression of a Cryptic PKS II Gene Cluster from *Streptomyces* sp. Tü6314

Zhengyi Qian,[a] Torsten Bruhn,[b] Paul M. D'Agostino,[a,c] Alexander Herrmann,[d] Martin Haslbeck,[e] Noémi Antal,[f] Hans-Peter Fiedler,[f] Ruth Brack-Werner,[d] and Tobias A. M. Gulder[a,c*]

[a]Biosystems Chemistry, Department of Chemistry and Center for Integrated Protein Science Munich (CIPSM), Technical University of Munich, Lichtenbergstraße 4, 85748 Garching bei München, Germany.

[b] Bundesinstitut für Risikobewertung, Max-Dohrn-Str. 8-10, 10789 Berlin, Germany.
[c]Technische Universität Dresden, Chair of Technical Biochemistry, Bergstraße 66, 01602 Dresden, Germany.

[d]Helmholtz Zentrum München, German Research Center for Environmental Health, Institute of Virology, Ingolstädter Landstraße 1, 85764 Neuherberg, Germany.

[e]Department of Chemistry, Technical University of Munich, Lichtenbergstraße 4, 85748 Garching bei München, Germany.

[f]Institute of Microbiology, University of Tübingen, Auf der Morgenstelle 28, D-72076 Tübingen, Germany

[*]Correspondence: tobias.gulder@ch.tum.de, tobias.gulder@tu-dresden.de

101

Table of contents

Commercial materials

The primers were synthesized by Sigma-Aldrich (Taufkirchen, Germany). The restriction enzymes and polymerase (Q5 High-Fidelity DNA polymerase) were purchased from New England Biolabs (Frankfurt am Main, Germany). Plasmid isolation kit (peqGOLD Plasmid Miniprep Kit I, C-Line) and DNA agarose gel extraction kit (peqGOLD Gel Extraction Kit, S-Line) were purchased from VWR (Darmstadt, Germany). The deuterated DMSO was purchased from EurisoTop (Saarbrücken, Germany).

Table S1. Plasmids and strains used in this study

Plasmid or Strain	Characteristics	References
Plasmids		
pSET152	ϕC31 *attP-int*, *aph(3)II*, *oriT* (RP4), Apr^R	1
pSET152-*skt*	pSET152 derivative with *skt* gene cluster inserted	This study
E. coli		
DH5α	general cloning host strain	Invitrogen
GB05-dir	GB2005, *araC*-BAD-ETγA	2
ET12567/pUZ8002	*recF*, *dam-*, *dcm-*, Cm^R, Kan^R, carrying plasmid pUZ8002	3
Streptomyces		
Tü6314	Wild-type *Streptomyces* isolate	This study
M1152	derived from *S. coelicolor* M145: Δ*act*, Δ*red*, Δ*cpk*, Δ*cda*, *rpoB*(C1298T), SCP1$^-$, SCP2$^-$	4
M1154	derived from *S. coelicolor* M145: Δ*act*, Δ*red*, Δ*cpk*, Δ*cda*, *rpoB*(C1298T), *rpsL*(A262G), SCP1$^-$, SCP2$^-$	4
M1152/pSET152	empty pSET152 vector integrated in the M1152 chromosome, Apr^R	This study
M1154/pSET152	empty pSET152 vector integrated in the M1154 chromosome, Apr^R	This study
M1152/*skt*	*skt* gene cluster integrated in the M1152 chromosome, Apr^R	This study
M1154/*skt*	*skt* gene cluster integrated in the M1154 chromosome, Apr^R	This study

Table S2. Primers used in this study

Primer	Sequence 5'-3'*	Application
pSET152-cap_cluster21-F	gcgttctcgccgaccggggagacgatccagtgcttgacggacggagtgggGTCATAGCTGTTTCCTG	Capture vector PCR
pSET152-cap_cluster21-R	gcggctcgcccgtcgtcaccgtccatctgtccacctactacgagaactggACTGGCCGTCGTTTTAC	Capture vector PCR
pSET152_cap_seq-F	TGCTGCAAGGCGATTAAG	PCR screening
cap_cluster21_verification_L-R	CGACCCGAAGGTGAGCAACC	PCR screening

*Capitalized letters represent the primer binding regions. Lowercase letters represent the homology arms.
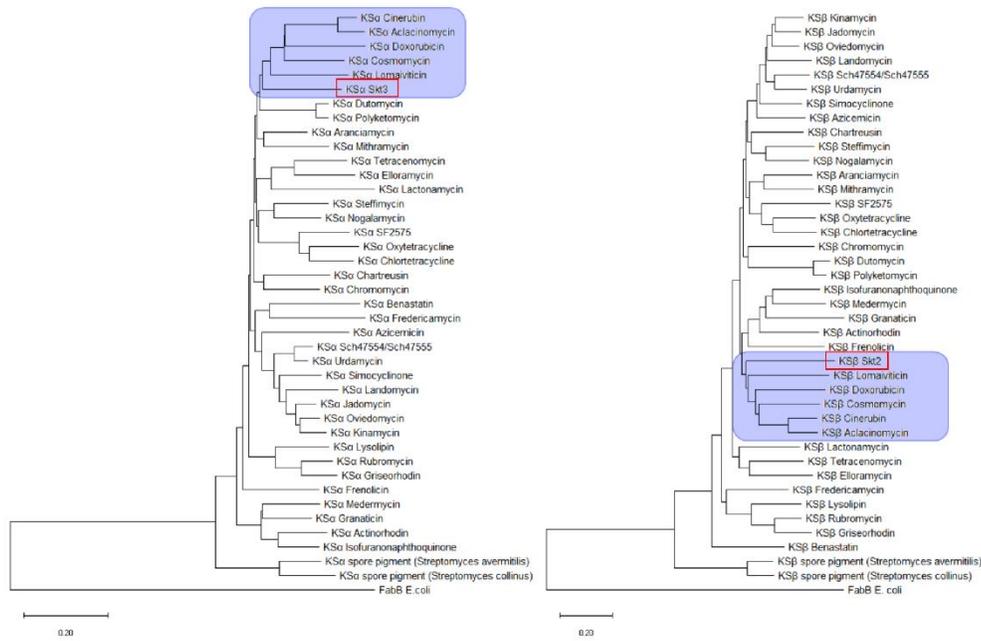
Figure S1. ClustalW-based phylogenetic trees containing KS_α (left) and KS_β (right) from sequenced PKS gene clusters. The *skt* gene cluster KS_α (Skt3) and KS_β (Skt2) are marked by a red rectangle. The branches with colored background have a chain length of 21 carbons. The *fabB* gene from *E. coli* was used as an outgroup. The sequences were downloaded from the MIBiG (Minimum Information about a Biosynthetic Gene cluster) database.[5]
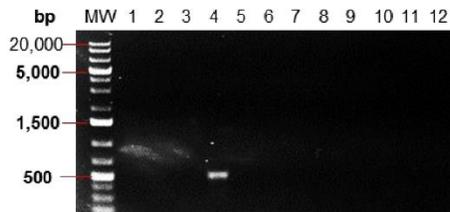
Figure S2. Colony screening PCR after LLHR using primer pairs pSET152_cap_seq-F and cap_cluster21_ verification_L-R. Target band has a product size of 508 bp.
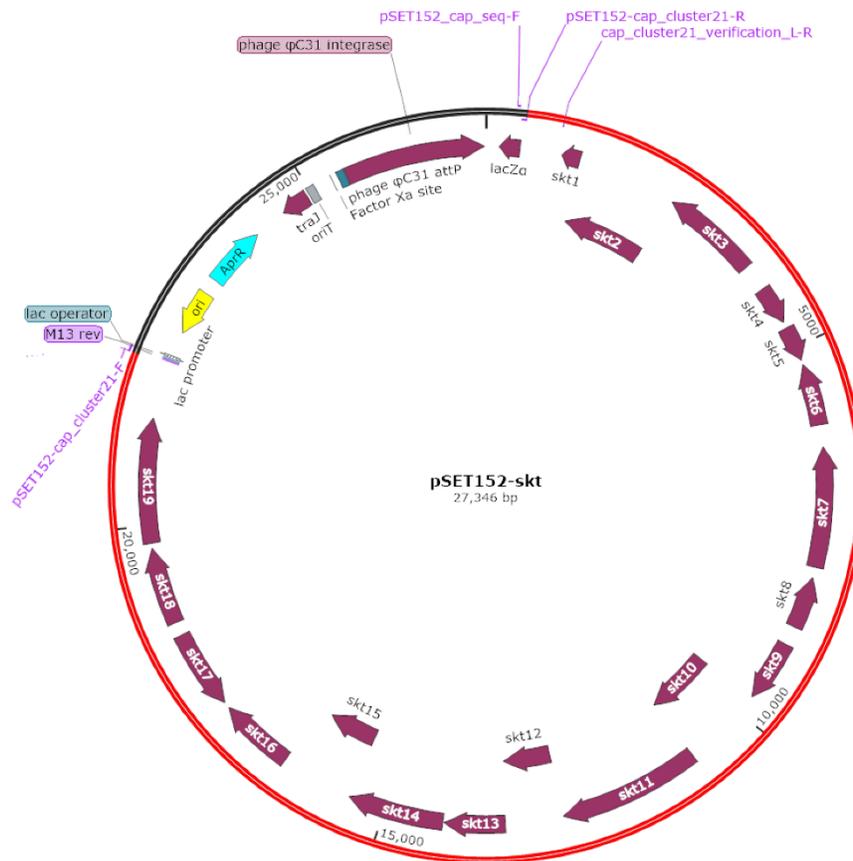
Figure S3. Plasmid map of pSET152-*skt* with positions of primers used in this study. The part labeled with red indicates the cloned *skt* gene cluster. This picture was generated by SnapGene software (GSL Biotech; available at snapgene.com).
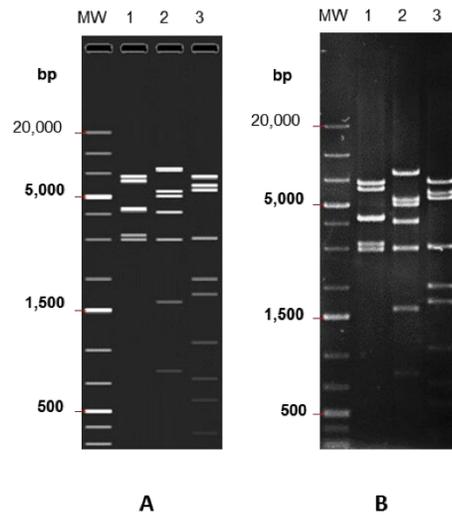
Figure S4. (A) Simulated restriction enzyme analysis of the pSET152-*skt* plasmid digested with *Blp*I (1), *Pst*I (2) and *Pvu*II (3). See Table S3 for details on expected DNA fragments sizes. This picture was generated by the SnapGene software (GSL Biotech; available at snapgene.com) (B) Authentic restriction enzyme analysis of the captured cluster digested with *Blp*I (1), *Pst*I (2) and *Pvu*II (3).

Table S3. Simulated restriction enzyme analysis

| Enzyme	*Blp*I (1)	*Pst*I (2)	*Pvu*II (3)
Band 1	6650 bp	7488 bp	6665 bp
Band 2	6143 bp	5293 bp	5763 bp
Band 3	4245 bp	4973 bp	5394 bp
Band 4	4216 bp	4033 bp	3031 bp
Band 5	3136 bp	2972 bp	2012 bp
Band 6	2956 bp	1596 bp	1723 bp
Band 7		802 bp	1086 bp
Band 8		189 bp	733 bp
Band 9			578 bp
Band 10			361 bp
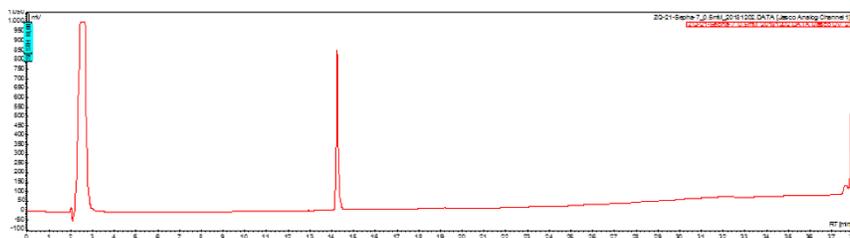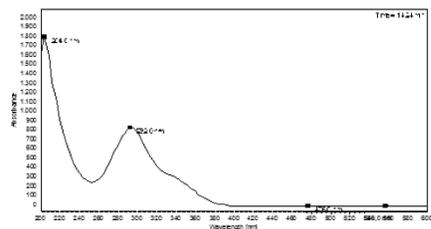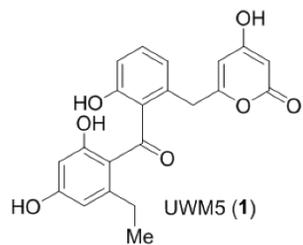
Figure S5. HPLC-UV trace of purified **1** (bottom) and its UV absorption spectrum (top right).
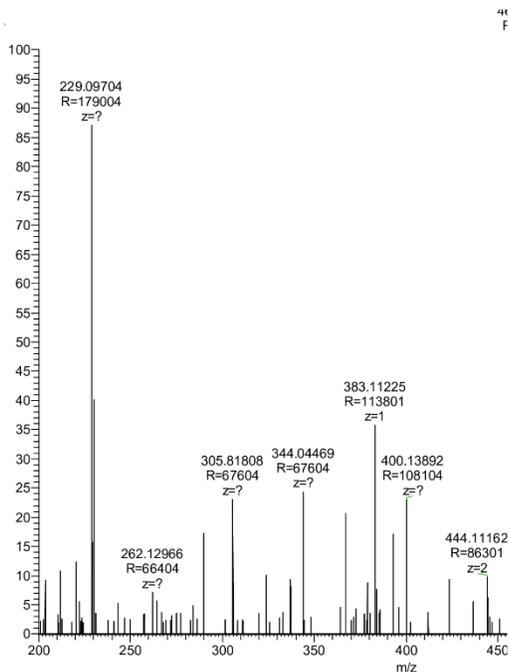
Figure S6. High-resolution ESI-(+) mass spectrum of compound **1**.

Table S4. ^{1}H (500 MHz) and ^{13}C (125 MHz) NMR data of UMW5 from the literature[6] and compound **1**, both recorded in DMSO-d$_6$.

Position	UMW5 ^{1}H (mult, J [Hz])	UMW5 ^{13}C	Compound **1** ^{1}H (mult, J [Hz])	Compound **1** ^{13}C
1		172.1*1		163.7
2	5.03 (d, 1.8)	87.7	5.16 (d, 2.1)	88.4
3		166.9*1		170.3
4	5.58 (brs)	102.2	5.65 (d, 2.1)	100.9
5		162.2*3		164.5
6	3.56 (s)	36.5	3.60 (s)	36.6
7		133.6		133.4
8	6.72 (d, 7.9)	131.5*2	6.73 (d, 7.9)	121.0
9	7.19 (t, 7.9)	120.7*2	7.20 (t, 7.9)	130.0
10	6.77 (d, 7.9)	114.5	6.76 (d, 7.9)	114.7
11		154.4		154.6
12	-	130.7	-	130.9
13	-	199.3	-	199.3
14	-	117.2	-	117.5
15	-	164.1*3	-	162.0
16	6.11 (d, 2.2)	100.2	6.08 (d, 2.3)	100.3
17	-	163.5	-	161.9
18	6.16 (d, 2.2)	108.5	6.14 (d, 2.3)	108.5
19	-	148.0	-	148.0
20	2.34 (dt, 7.4, 7.4)	25.9	2.34 (q, 7.4)	26.0
21	1.00 (t, 7.4)	15.1	0.97 (t, 7.4)	15.3
3-OH	not reported	-	11.54 (s)	-
11-OH	not reported	-	9.69 (s)	-
15-OH	not reported	-	11.27 (s)	-
17-OH	not reported	-	10.07 (s)	-

*1, *2, *3: Assignments seem to be interchanged in the original literature.
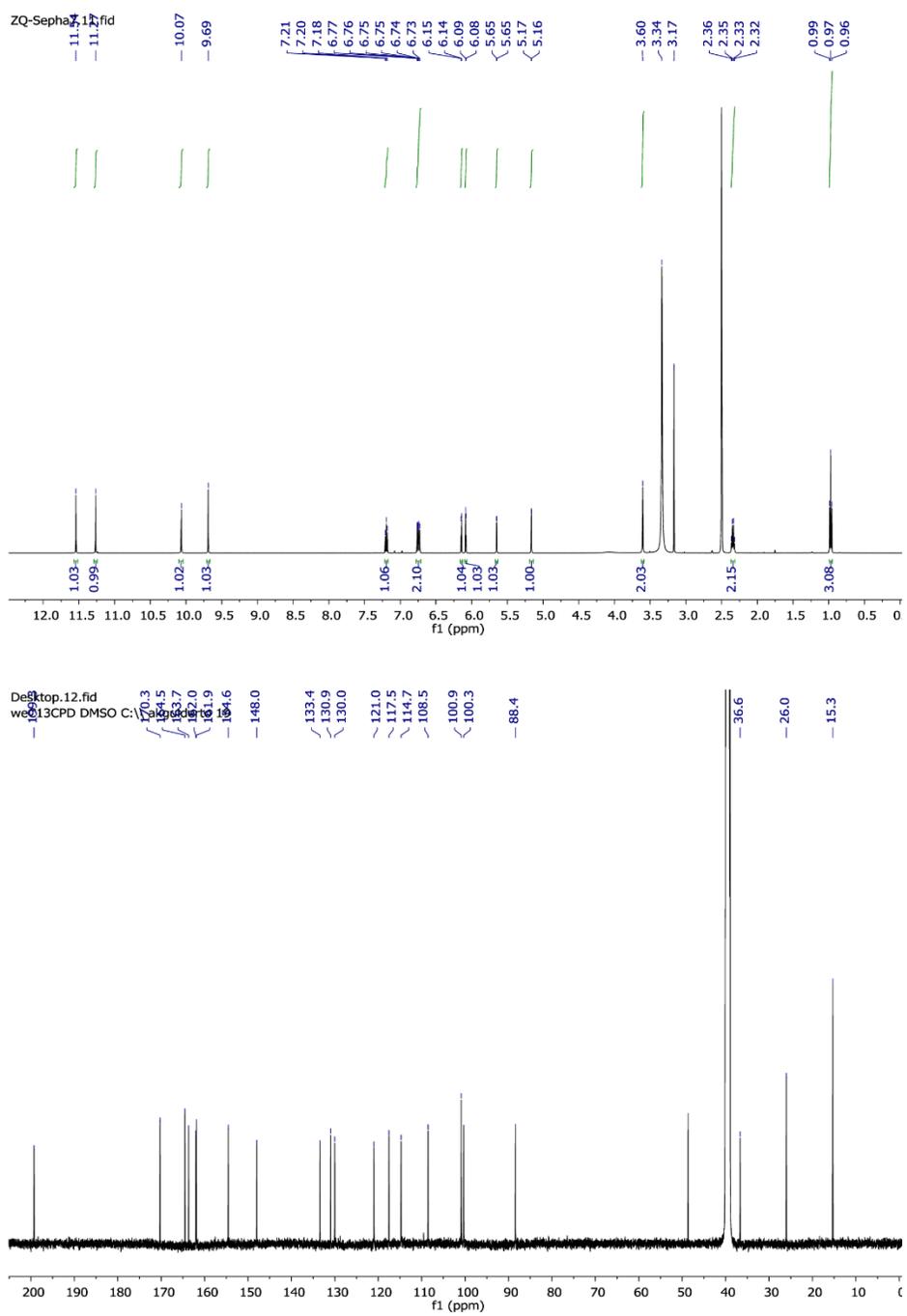
Figure S7. ^1H (top) and ^{13}C-NMR spectra (bottom) of compound **1**.
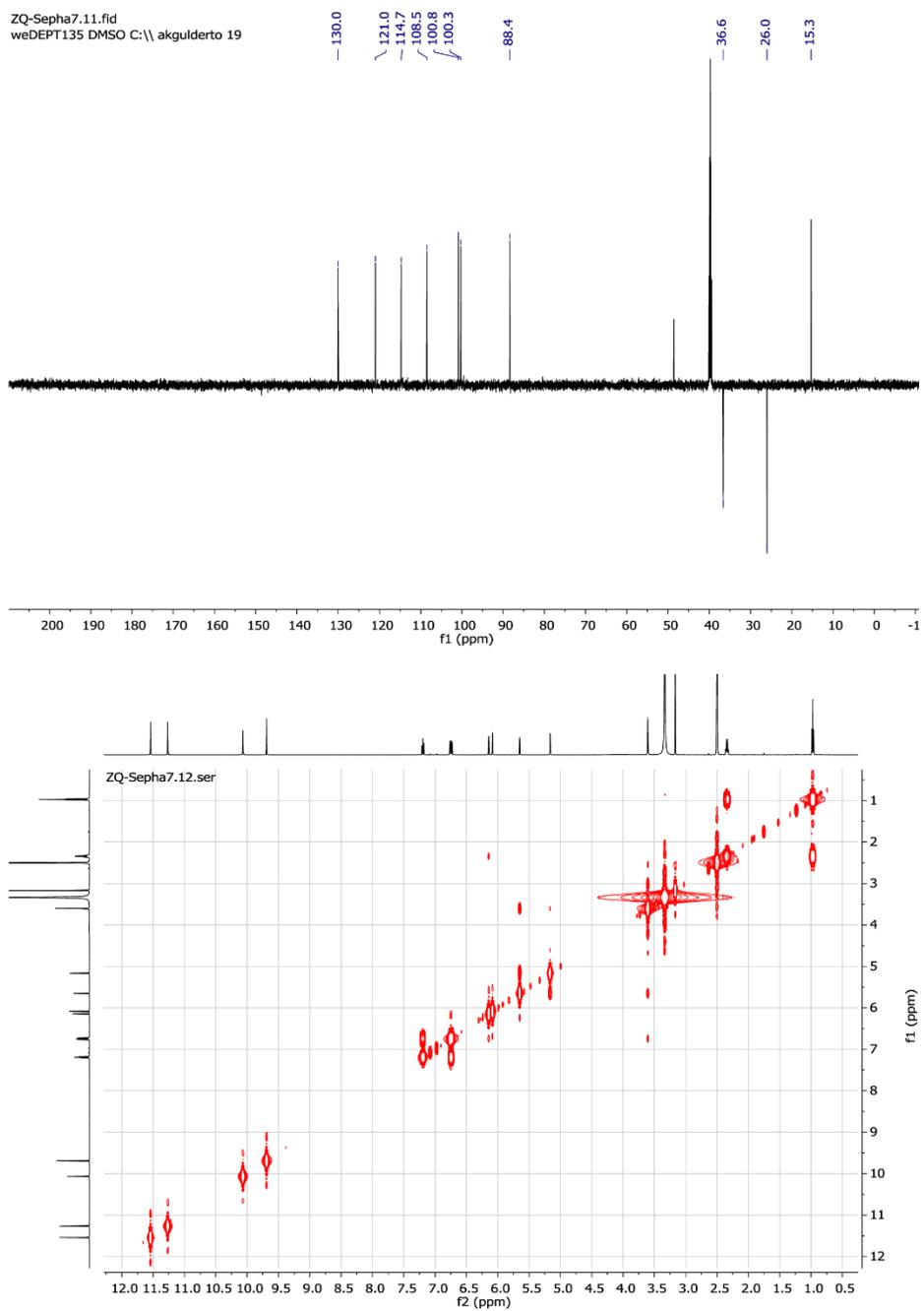
Figure S8. DEPT135 (top) and COSY-NMR spectra (bottom) of compound **1**.
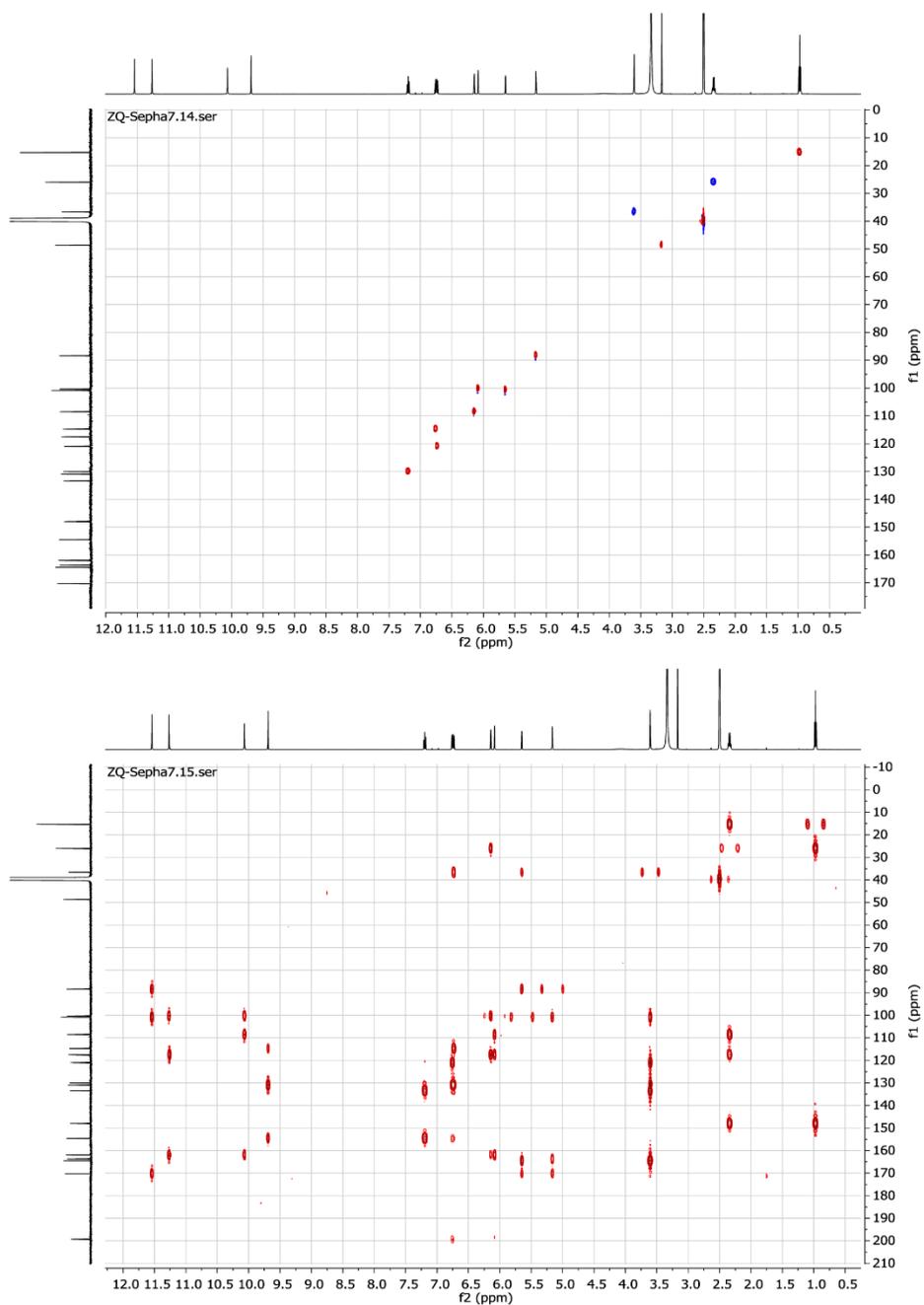
**Figure S9.** HSQC (top) and HMBC-NMR spectra (bottom) of compound **1**.
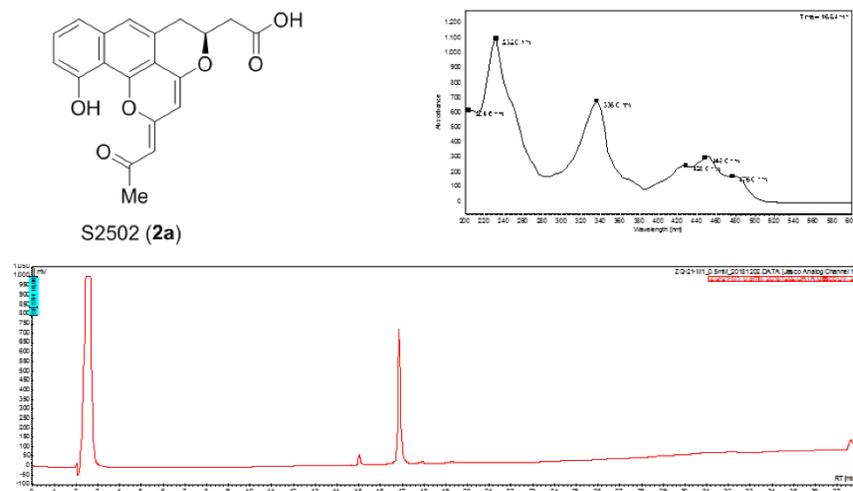
**Figure S10.** HPLC-UV trace of purified **2a** (bottom) and its UV absorption spectrum (top right).
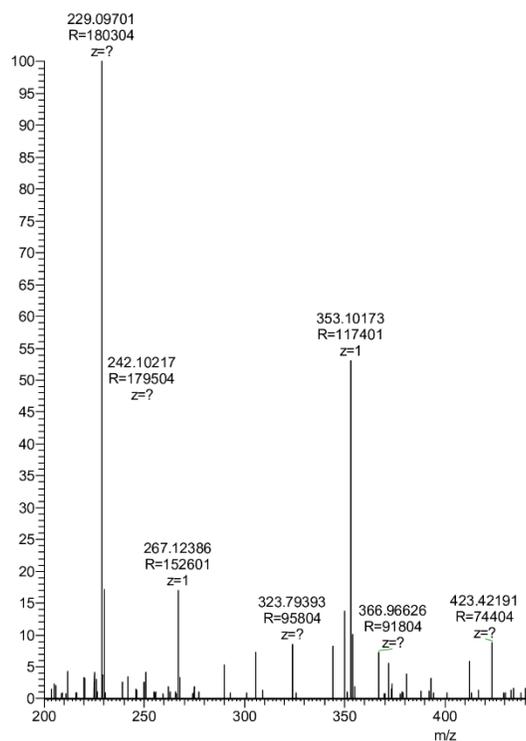
**Figure S11.** High-resolution ESI-(+) mass spectrum of compound **2a**.

Table S5. ^{13}C NMR data of chemically synthetic S2502,[7] naturally isolated S2502[8] from the literature and compound **2a**, all recorded in DMSO-d_6 at 125 MHz (synthetic 2502 and **2a**) or 100 MHz (isolated S2502).

Position	S2502 (synthetic)	S2502 (isolated)	Compound 2a
1	171.9	170.8	171.5
2	39.6	39.2	*1
3	70.3	74.2	75.2
4	32.0	31.2	31.3
5	127.4	125.5	126.7
6	122.4	121.7	122.0
7	137.3	136.5	136.8
8	118.9	118.1	118.5
9	130.9	130.1	130.6
10	113.2	112.5	112.8
11	154.9	154.3	154.4
12	112.0	111.5	111.5
13	150.5	149.9	150.0
14	107.9	106.8	107.3
15	158.6	157.4	157.9
16	97.3*2	96.8*2	97.0*2
17	162.6	161.7	162.1
18	97.1*2	96.6*2	96.8*2
19	193.6	192.8	193.2
20	30.7	30.0	30.3
21			

*1: Signal invisible due to overlap with NMR solvent signal. *2: signal assignment might be interchanged.

Table S6. ^{1}H NMR data of chemically synthesized S2502,[7] naturally isolated S2502[8] from the literature and compound **2a** recorded in DMSO-d_6, all recorded in DMSO-d_6 at 500 MHz (synthetic 2502 and **2a**) or 400 MHz (isolated S2502).

| Position	S2502 (synthetic)	S2502 (isolated)	Compound **2a**
1			
2	3.04 (m)	2.91 (dd, 16.1, 4.9) 2.74 (dd, 16.1, 7.8)	2.70 (m)
3	4.70 (m)	4.69 (m)	4.71 (m)
4	3.30 (m)	3.18 (dd, 16.4, 3.1) 3.00 (dd, 16.4, 10.6)	3.24 (m) 3.05 (m)
5			
6	7.48 (s)	7.32 (s)	7.50 (s)
7			
8	7.33 (d, 8.0)	7.18 (dd, 7.8, 1.0)	7.33 (d, 7.9)
9	7.48 (t, 8.0)	7.38 (dd, 7.9, 7.8)	7.48 (t, 7.9)
10	6.90 (d, 8.0)	6.82 (dd, 8.4, 1.0)	6.90 (d, 7.9)
11			
12			
13			
14			
15			
16	6.06 (s)	5.90 (s)	6.09 (s)
17			
18	5.54 (s)	5.44 (s)	5.55 (s)
19			
20	2.05 (s)	2.03 (s)	2.05 (s)
21			
OH	11.26 (s)	11.16 (s)	11.26 (s)

S-14

114
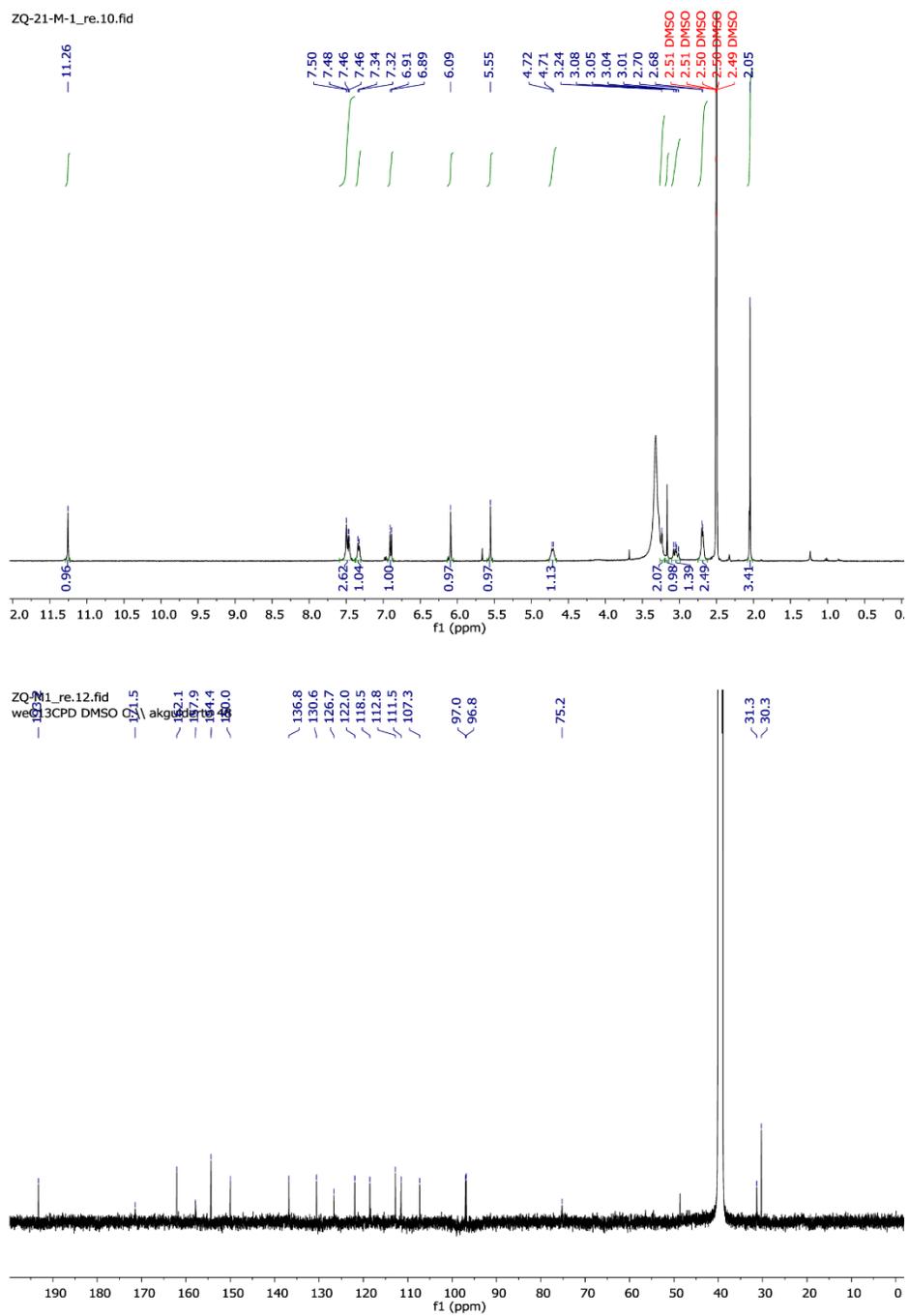
**Figure S12.** ¹H (top) and ¹³C-NMR spectra (bottom) of compound **2a**.
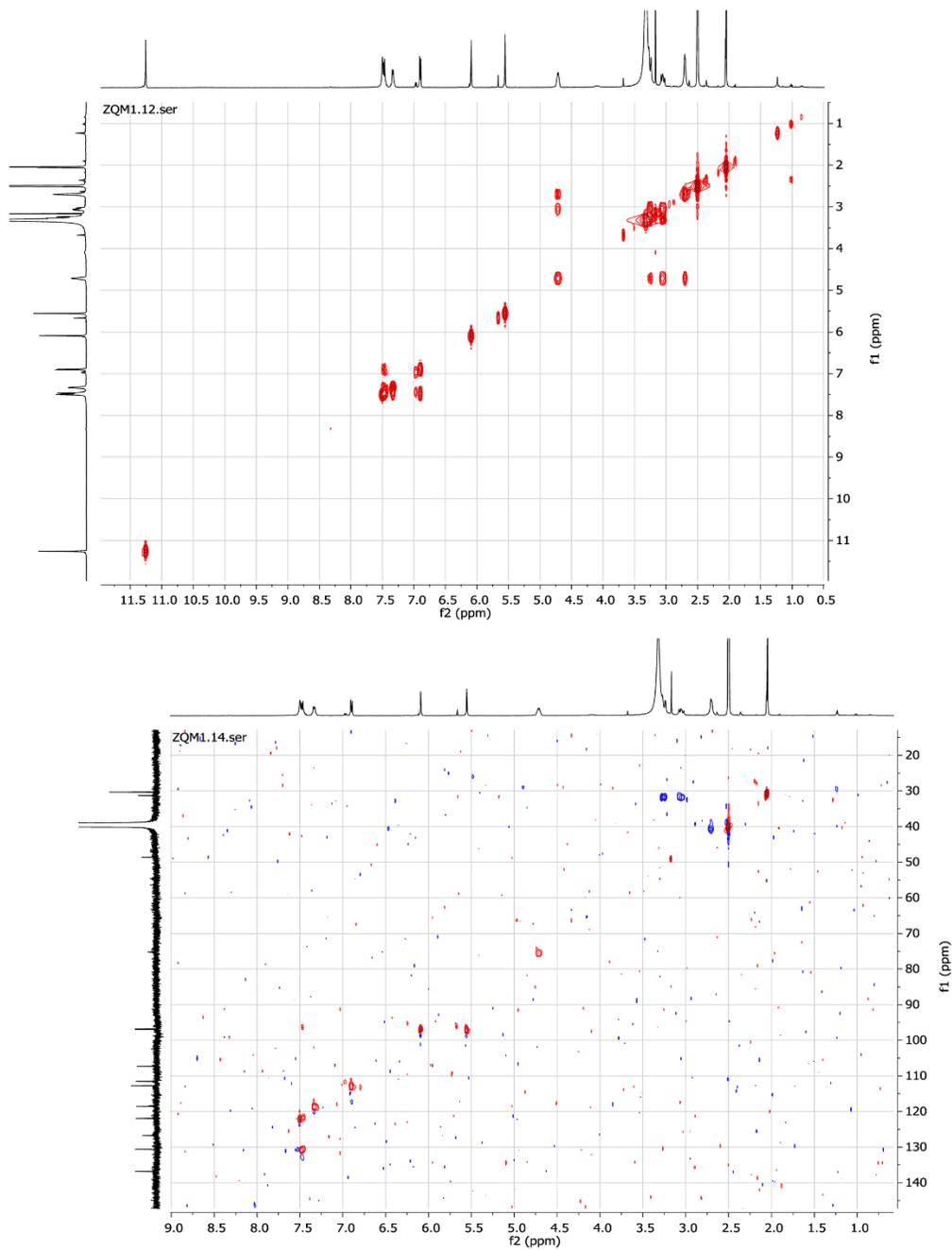
Figure S13. COSY- (top) and HSQC-NMR spectra (bottom) of compound **2a**.
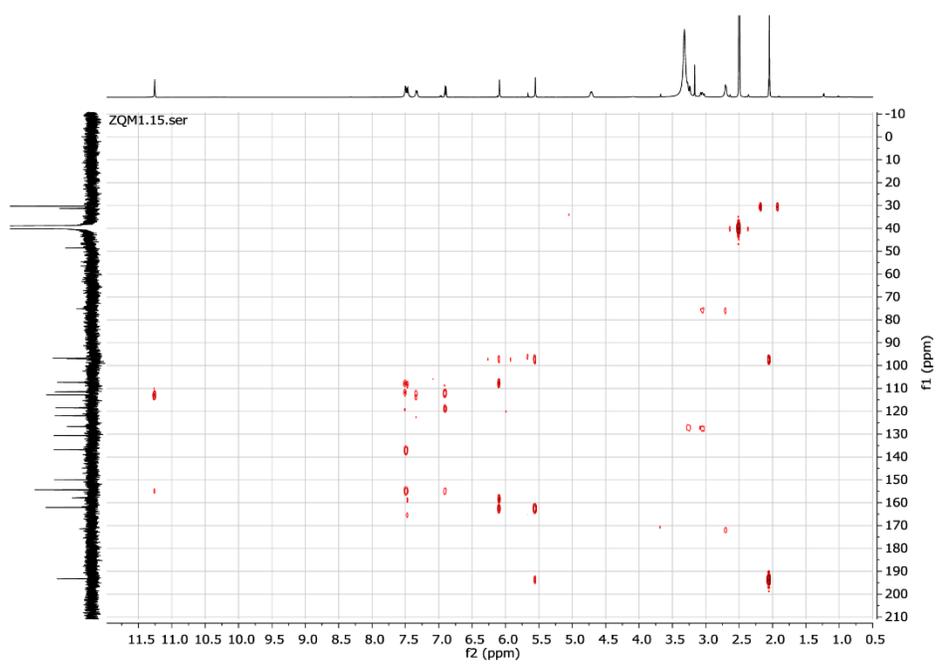
**Figure S14.** HMBC-NMR spectrum of compound **2a**.
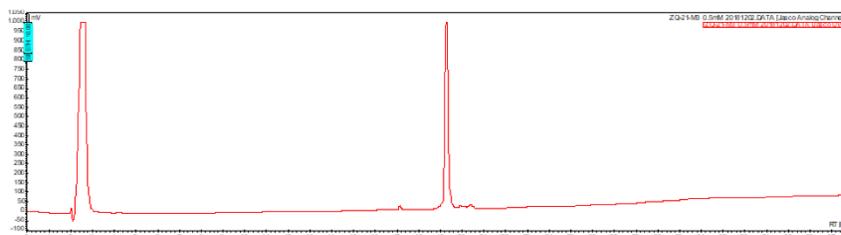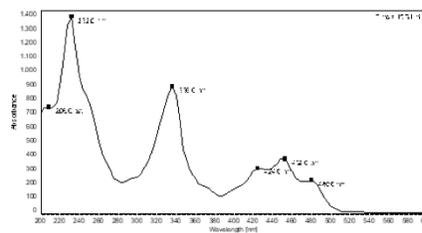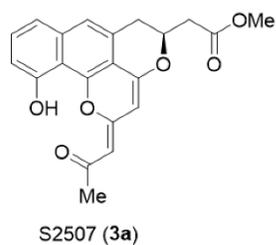
Figure S15. HPLC-UV trace of purified **3a** (bottom) and its UV absorption spectrum (top right).
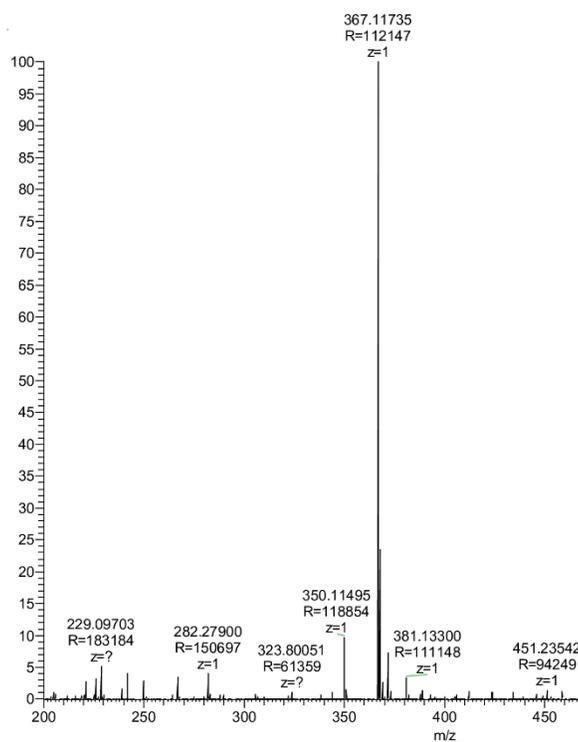
Figure S16. High-resolution ESI-(+) mass spectrum of compound **3a**.

Table S7. ^{13}C NMR data of chemically synthesized S2507 (125 MHz, CDCl$_3$),[7] naturally isolated S2507 (100 MHz, CDCl$_3$)[8] from the literature and compound **3a** recorded in DMSO-d$_6$ at 125 MHz.

Position	S2507 (synthetic)	S2507 (isolated)	Compound 3a
1	170.0	170.0	170.1
2	39.6	39.6	38.8
3	74.0	74.0	74.4
4	32.2	32.2	31.1
5	124.7	124.7	126.3
6	122.1	122.1	122.0
7	137.1	137.1	136.8
8	118.2	118.2	118.6
9	130.9	130.9	130.6
10	112.6	112.6	111.5
11	155.5	155.5	154.3
12	113.7	113.7	112.8
13	151.1	150.0	150.0
14	106.9	106.9	107.2
15	157.6	157.6	157.4
16	97.6*	97.6*	97.2*
17	162.2	162.2	161.9
18	97.1*	97.1*	97.0*
19	194.2	194.3	193.3
20	30.4	30.4	30.3
21	52.2	52.2	51.8

*[2]: signal assignment might be interchanged.

Table S8. ^{1}H NMR data of chemically synthesized S2507 (500 MHz, CDCl$_3$),[7] naturally isolated S2507 (400 MHz, CDCl$_3$)[8] from the literature and compound **3a** recorded in DMSO-d$_6$ at 500 MHz.

Position	S2507 (synthetic)	S2507 (isolated)	Compound **3a**
1			
2	2.90 (dd, 16.0, 7.0)	2.91 (dd, 15.8, 7.2)	2.96 (dd, 16.3, 4.6)
	2.74 (dd, 16.0, 6.0)	2.75 (dd, 15.8, 5.8)	2.86 (dd, 16.4, 8.1)
3	4.76 (m)	4.73 (m)	4.77 (m)
4	3.18 (dd, 16.0, 3.0)	3.20 (dd, 16.0, 3.0)	3.23 (dd, 16.5, 3.0)
	3.03 (dd, 16.0, 10.0)	2.91 (dd, 15.8, 7.2)	3.08 (dd, 16.5, 10.5)
5			
6	7.26 (s)	7.28 (s)	7.51 (s)
7			
8	7.17 (d, 8.0)	7.19 (dd, 7.4, 1.0)	7.34 (d, 7.8)
9	7.43 (t, 8.0)	7.45 (dd, 7.8, 7.3)	7.49 (t, 7.8)
10	7.02 (d, 8.0)	7.03 (dd, 7.8, 1.0)	6.90 (d, 7.8)
11			
12			
13			
14			
15			
16	5.73 (s)	5.75 (s)	6.11 (s)
17			
18	5.37 (s)	5.38 (s)	5.56 (s)
19			
20	2.15 (s)	2.15 (s)	2.05 (s)
21	3.77 (s)	3.77 (s)	3.68 (s)
OH	11.38 (s)	11.43 (s)	11.24 (s)
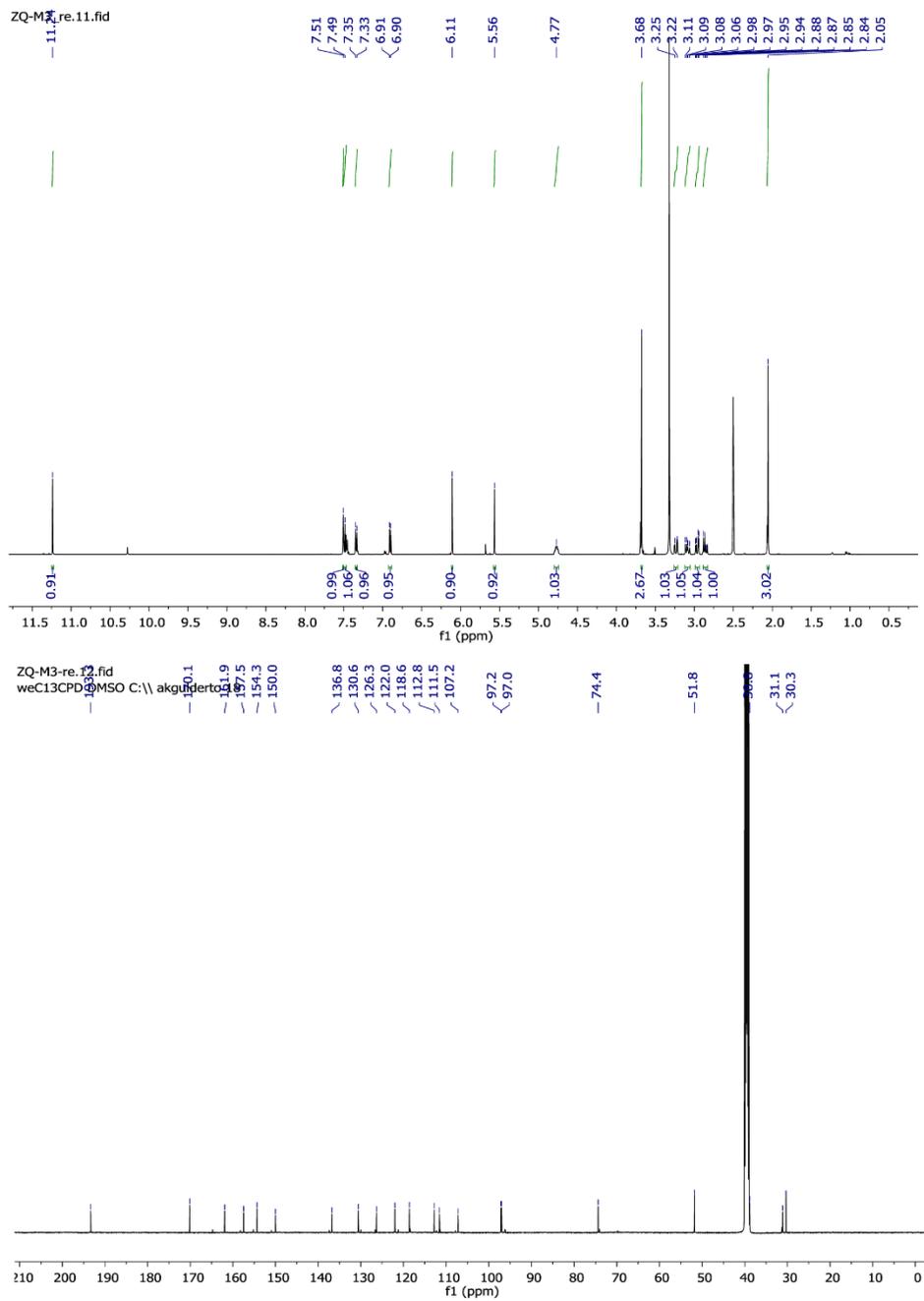
Figure S17. ^{1}H (top) and ^{13}C-NMR spectra (bottom) of compound **3a**.
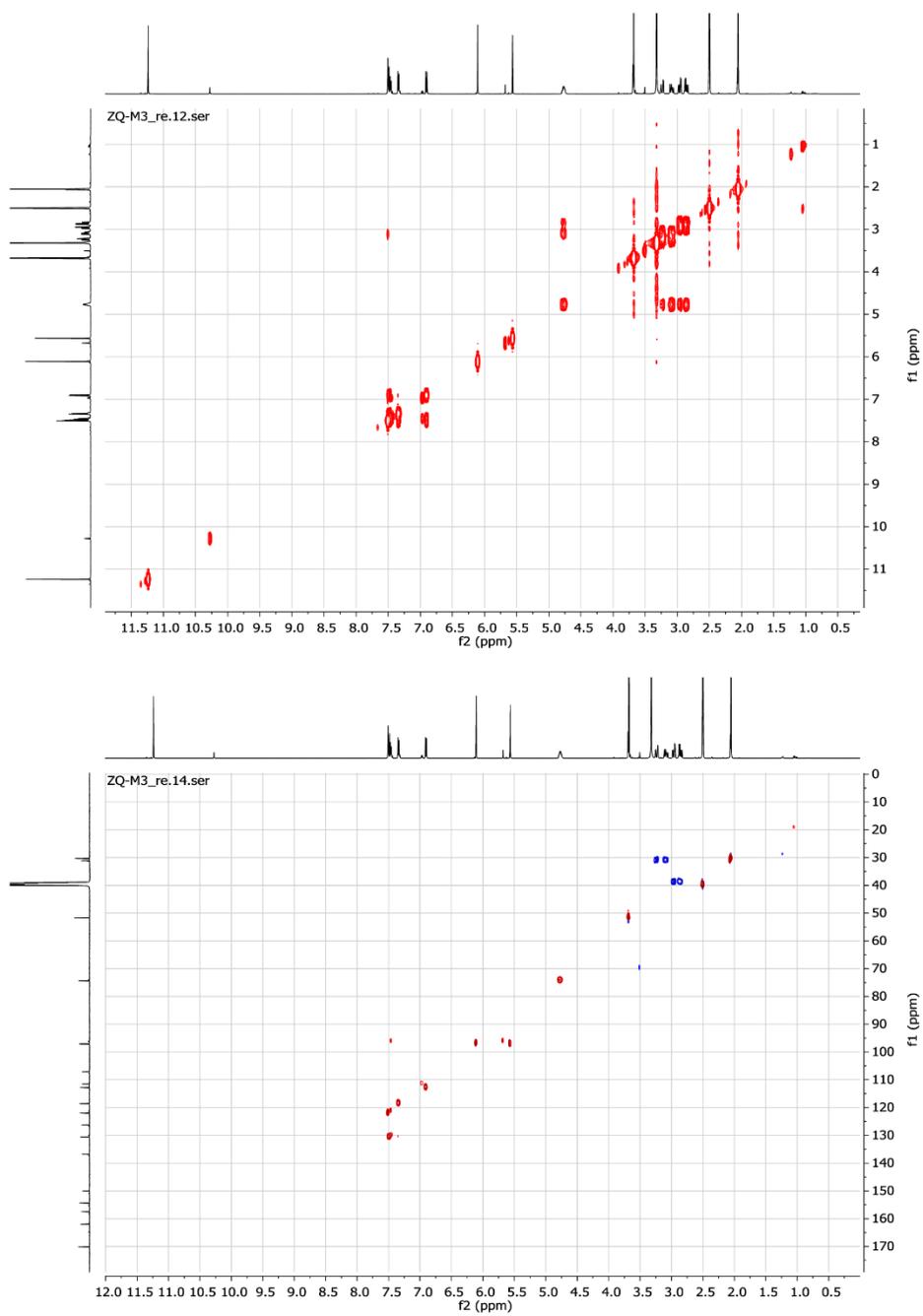
Figure S18. COSY- (top) and HSQC-NMR spectra (bottom) of compound **3a**.
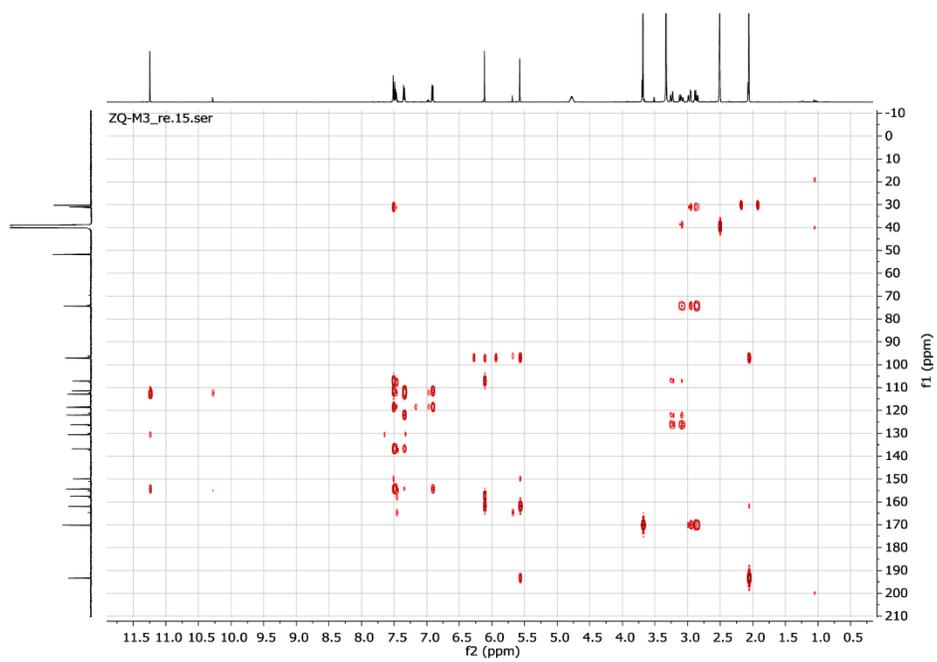
Figure S19. HMBC-NMR spectrum of compound **3a**.

streptoketide A (**2b**)
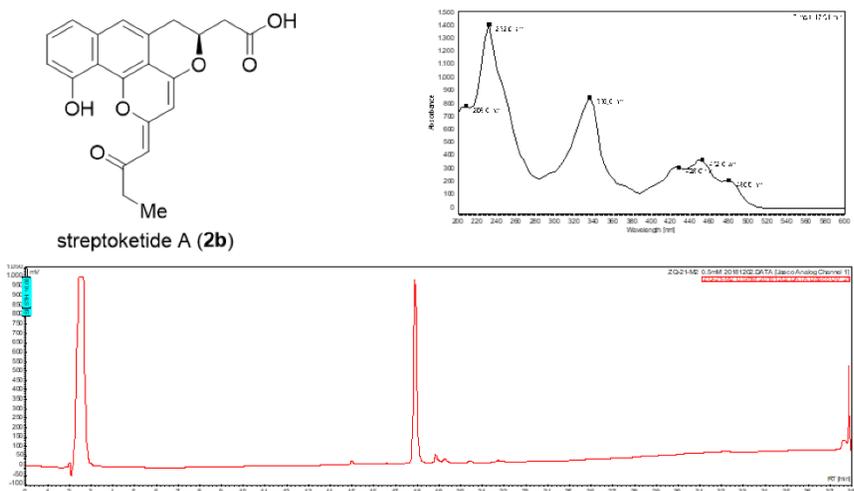
Figure S20. HPLC-UV trace of purified **2b** (bottom) and its UV absorption spectrum (top right).
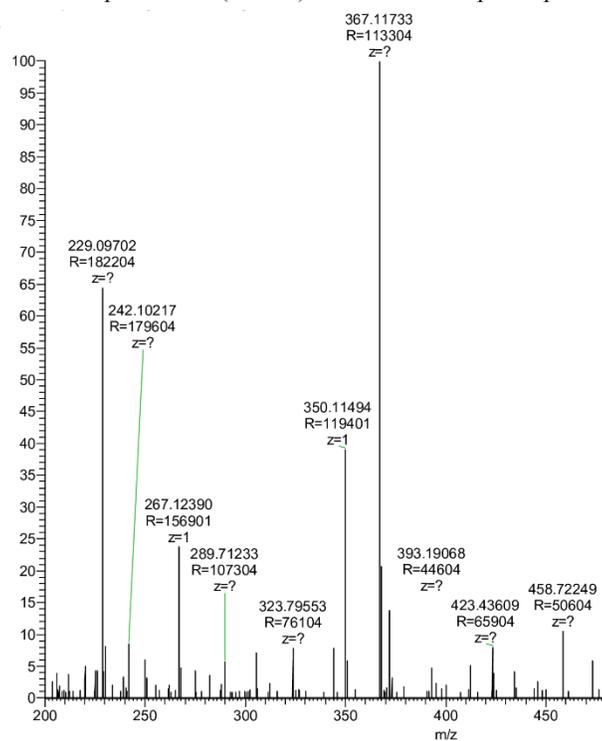
Figure S21. High-resolution ESI-(+) mass spectrum of compound **2b**.
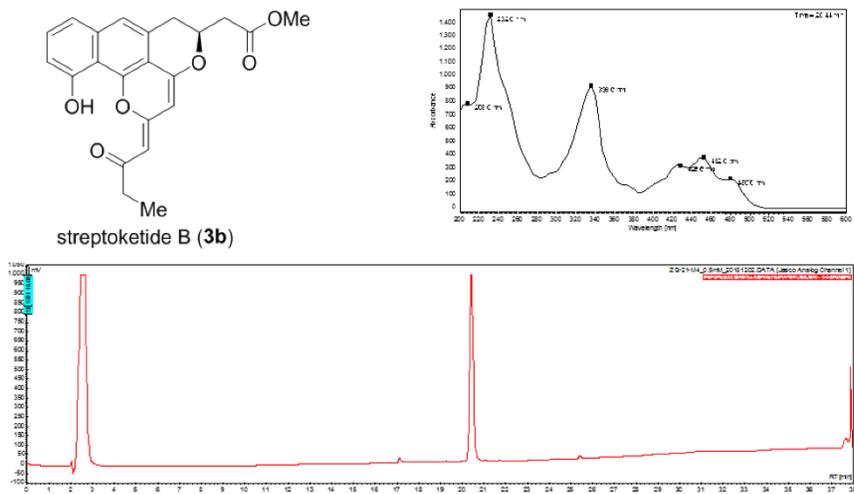
Figure S22. HPLC-UV trace of purified **3b** (bottom) and its UV absorption spectrum (top right).
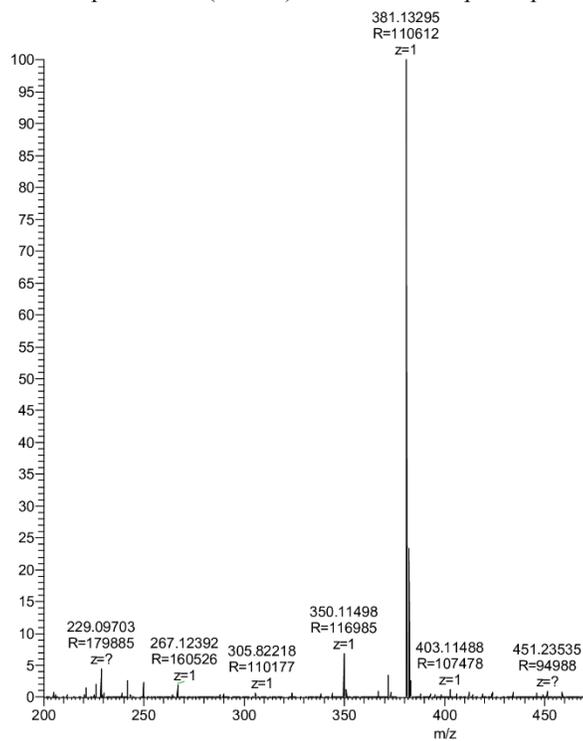
Figure S23. High-resolution ESI-(+) mass spectrum of compound **3b.**

Table S9. ^{1}H (500 MHz) and ^{13}C (125 MHz) NMR data of streptoketides A (**2b**) and B (**3b**) recorded in DMSO-d$_6$.

Position	^{13}C of **2b** (acid)	^{13}C of **3b** (Me)	^{1}H of **2b** (acid)	^{1}H of **3b** (Me)
1	171.2	170.1		
2	hidden	38.8	2.83 (dd, 16.3, 5.0)	2.95 (dd, 16.3, 4.5)
			2.75 (dd, 16.3, 7.8)	2.85 (dd, 16.3, 8.2)
3	74.6	74.3	4.73 (m)	4.76 (m)
4	31.2	31.1	3.24 (dd, 16.6, 2.8)	3.22 (dd, 16.7, 3.1)
			3.07 (dd, 16.6, 10.4)	3.07 (dd, 16.2, 10.3)
5	126.5	126.3		
6	121.9	121.9	7.51 (s)	7.48 (m)
7	136.8	136.7		
8	118.6	118.5	7.34 (d, 7.9)	7.32 (d, 8.0)
9	130.6	130.5	7.49 (t, 7.9)	7.48 (t, 8.0)
10	112.7	112.7	6.91 (d, 7.9)	6.90 (d, 8.0)
11	154.3	154.3		
12	111.5	111.5		
13	150.0	149.9		
14	107.2	107.1		
15	157.4	157.2		
16	97.0	97.1	6.11 (s)	6.09 (s)
17	161.9	161.8		
18	96.4	96.4	5.57 (s)	5.56 (s)
19	196.6	196.6		
20	35.6	35.6	2.35 (q, 7.4)	2.35 (q, 7.4)
21	9.3	9.3	1.02 (t, 7.4)	1.02 (t, 7.4)
22 (OMe)	-	51.8	-	3.68 (s)
OH	-	-	11.29	11.27
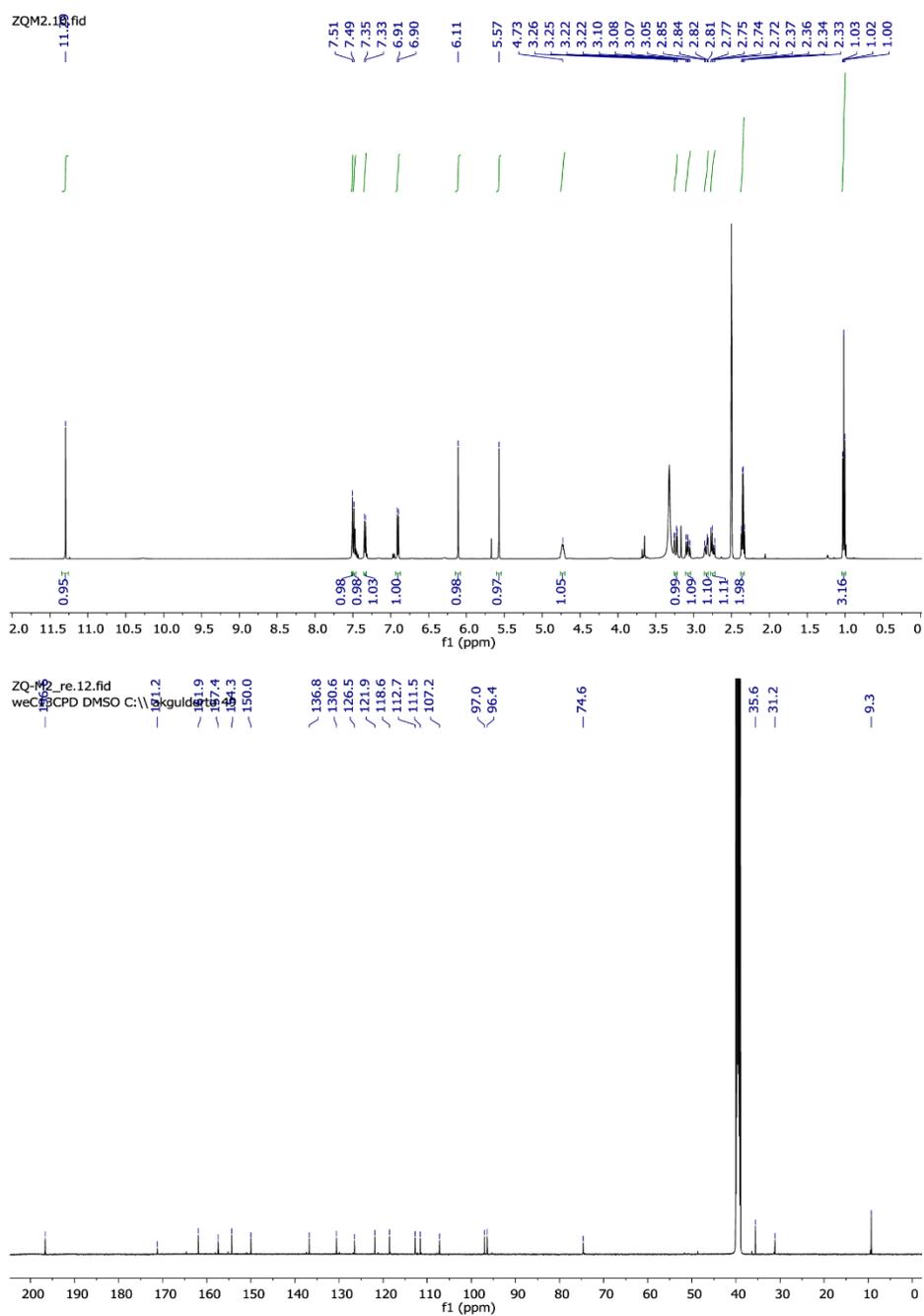
Figure S24. ^{1}H (top) and ^{13}C-NMR spectra (bottom) of compound **2b**.
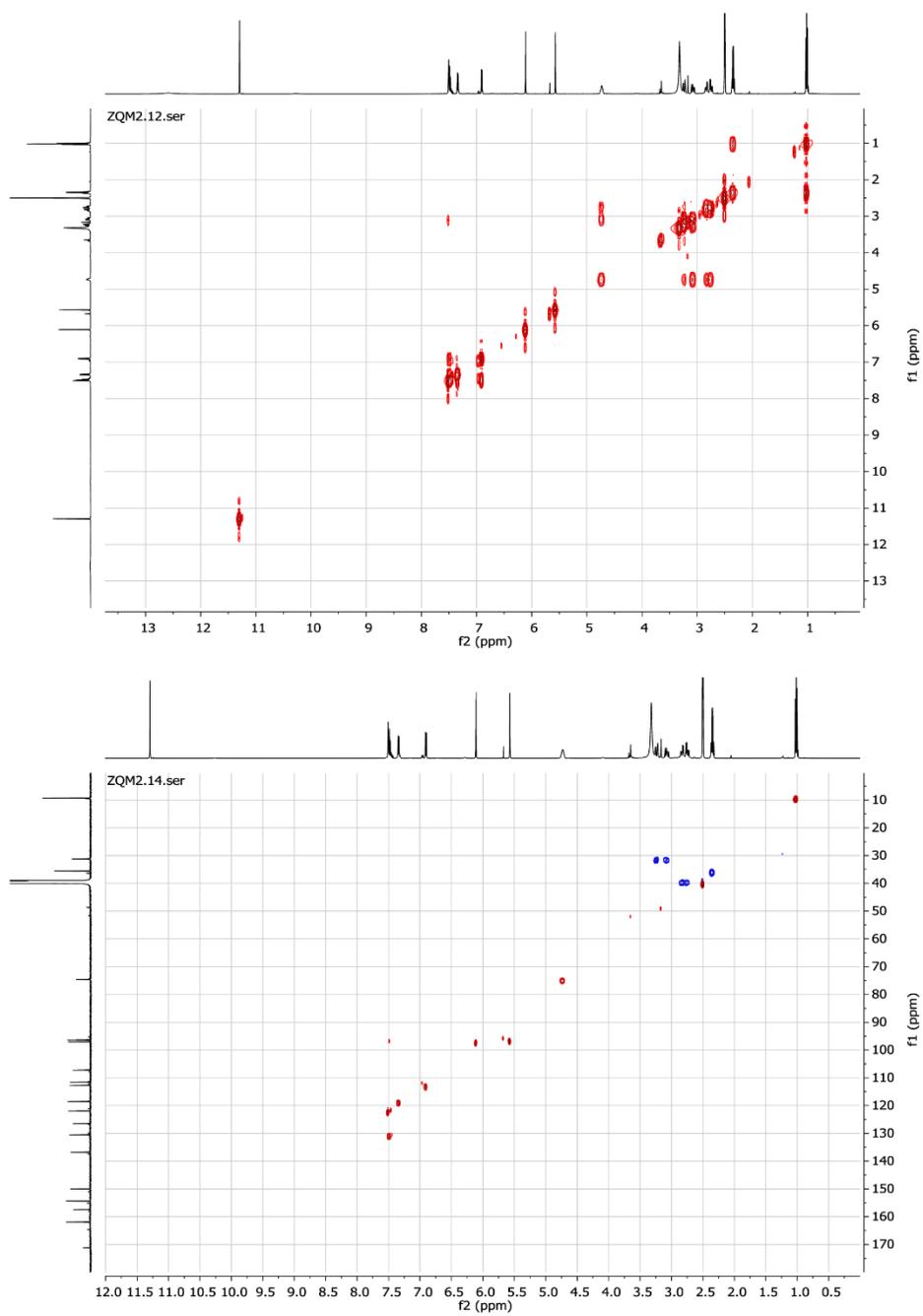
Figure S25. COSY- (top) and HSQC-NMR spectra (bottom) of compound **2b**.
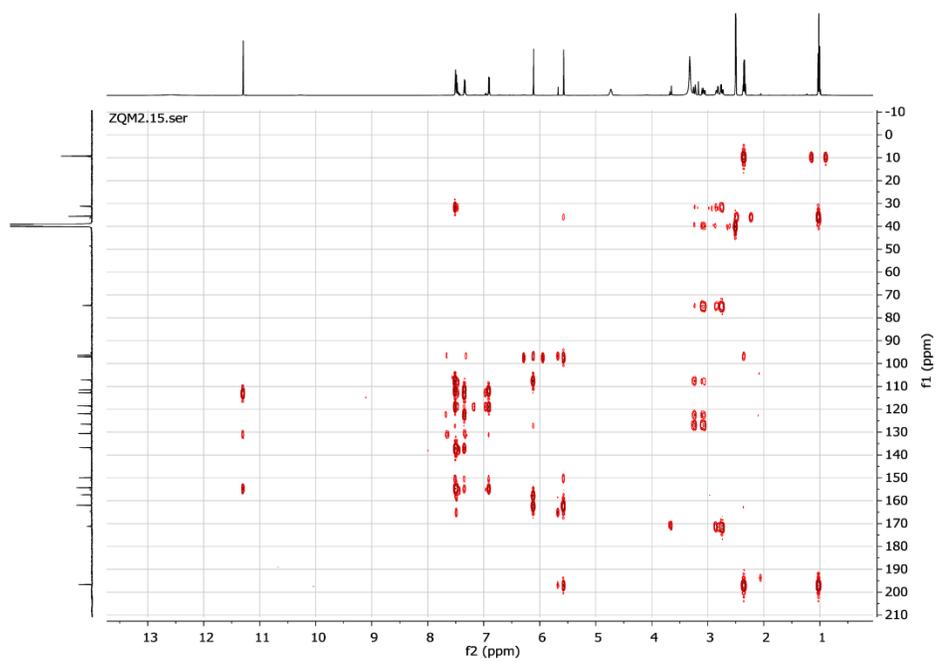
Figure S26. HMBC-NMR spectrum of compound **2b**.
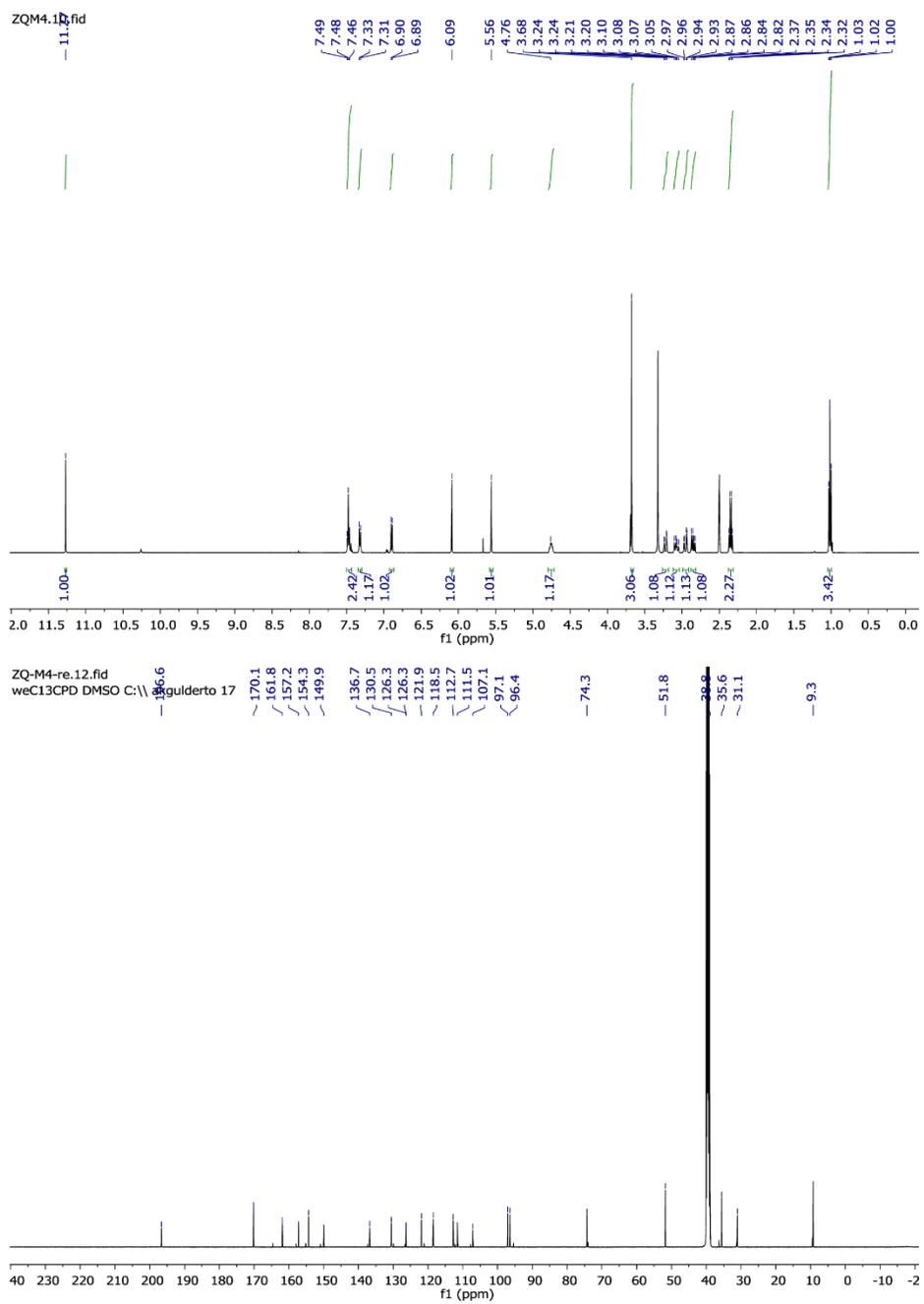
Figure S27. ^{1}H (top) and ^{13}C-NMR spectra (bottom) of compound **3b**.
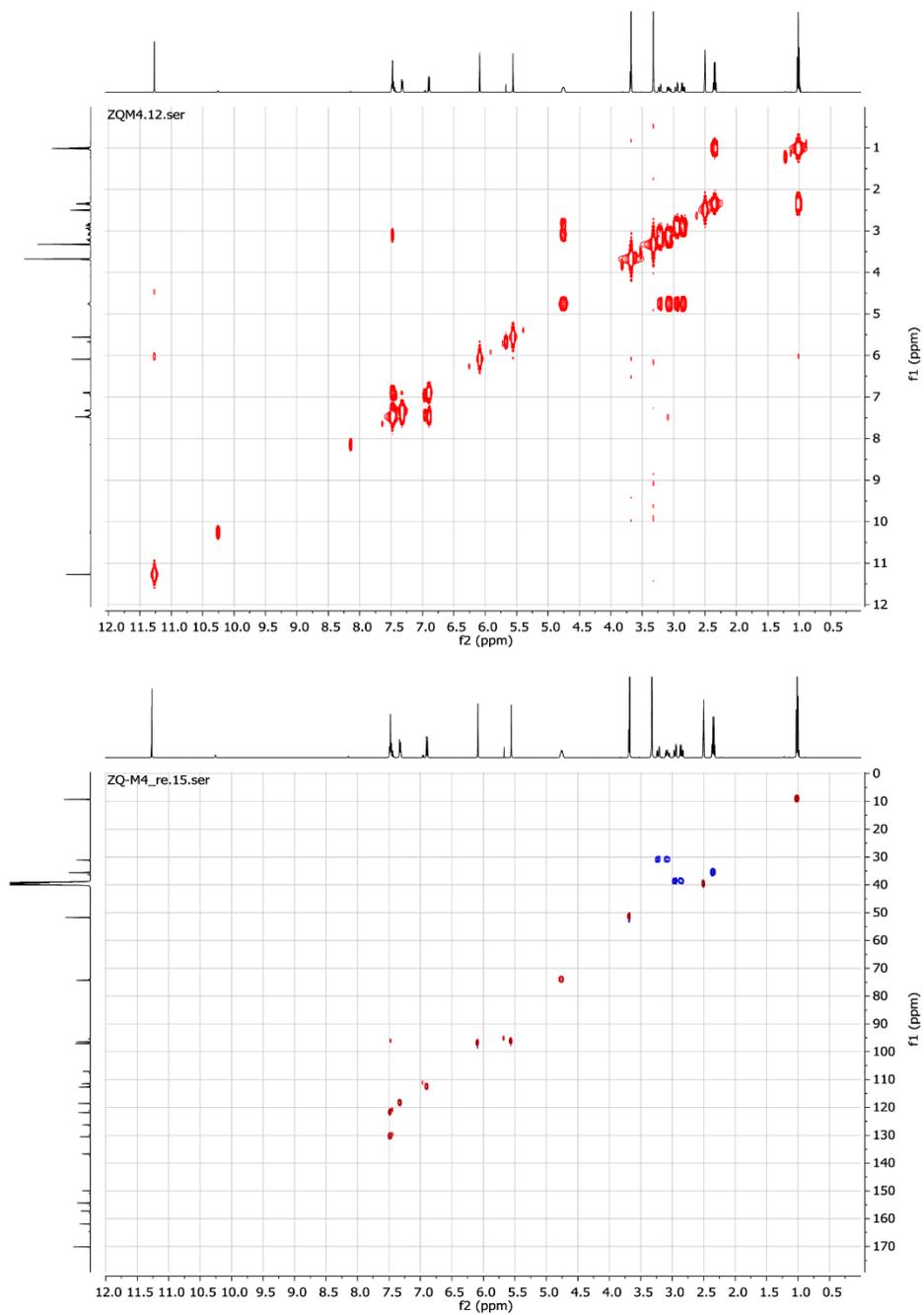
Figure S28. COSY- (top) and HSQC-NMR spectra (bottom) of compound **3b**.
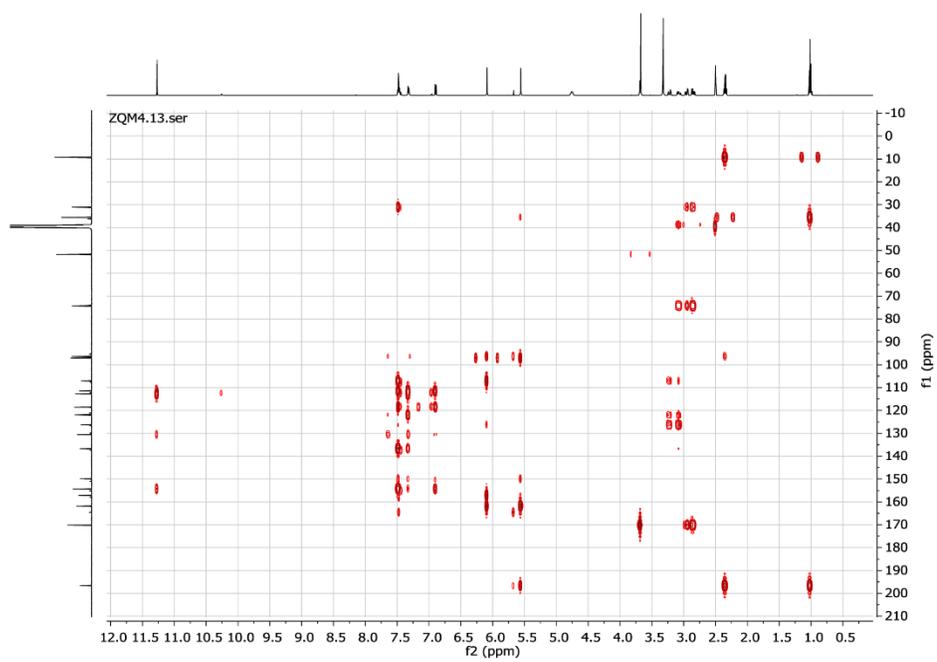
Figure S29. HMBC-NMR spectrum of compound **3b**.
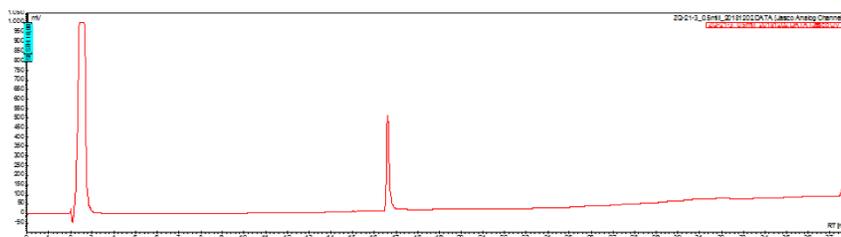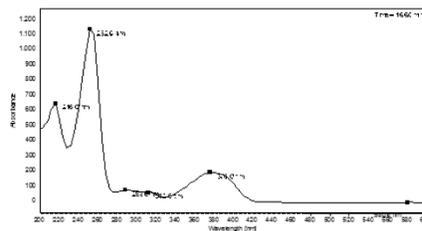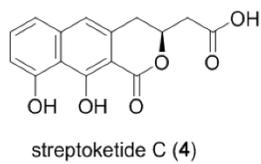
Figure S30. HPLC-UV trace of purified **4** (bottom) and its UV absorption spectrum (top right).
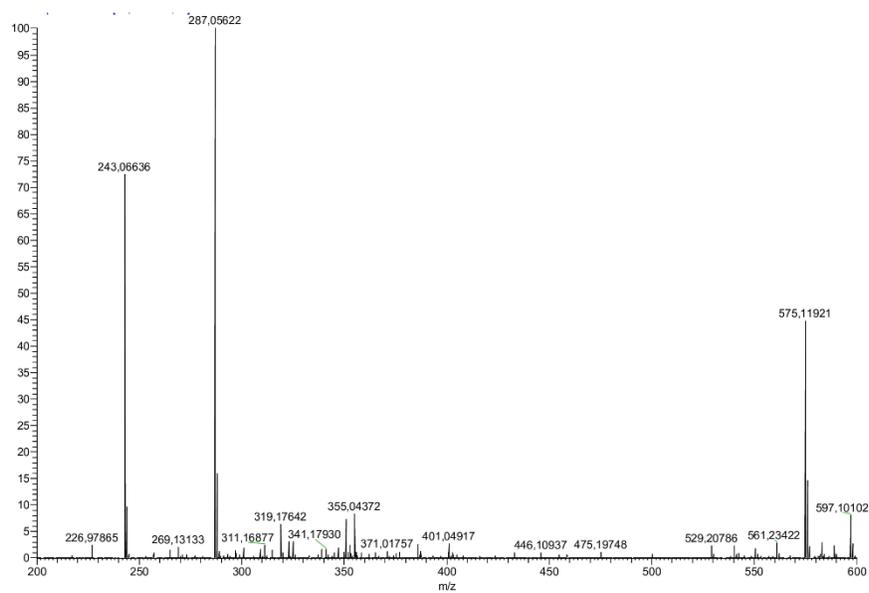
Figure S31. High-resolution ESI-(+) mass spectrum of compound **4**.
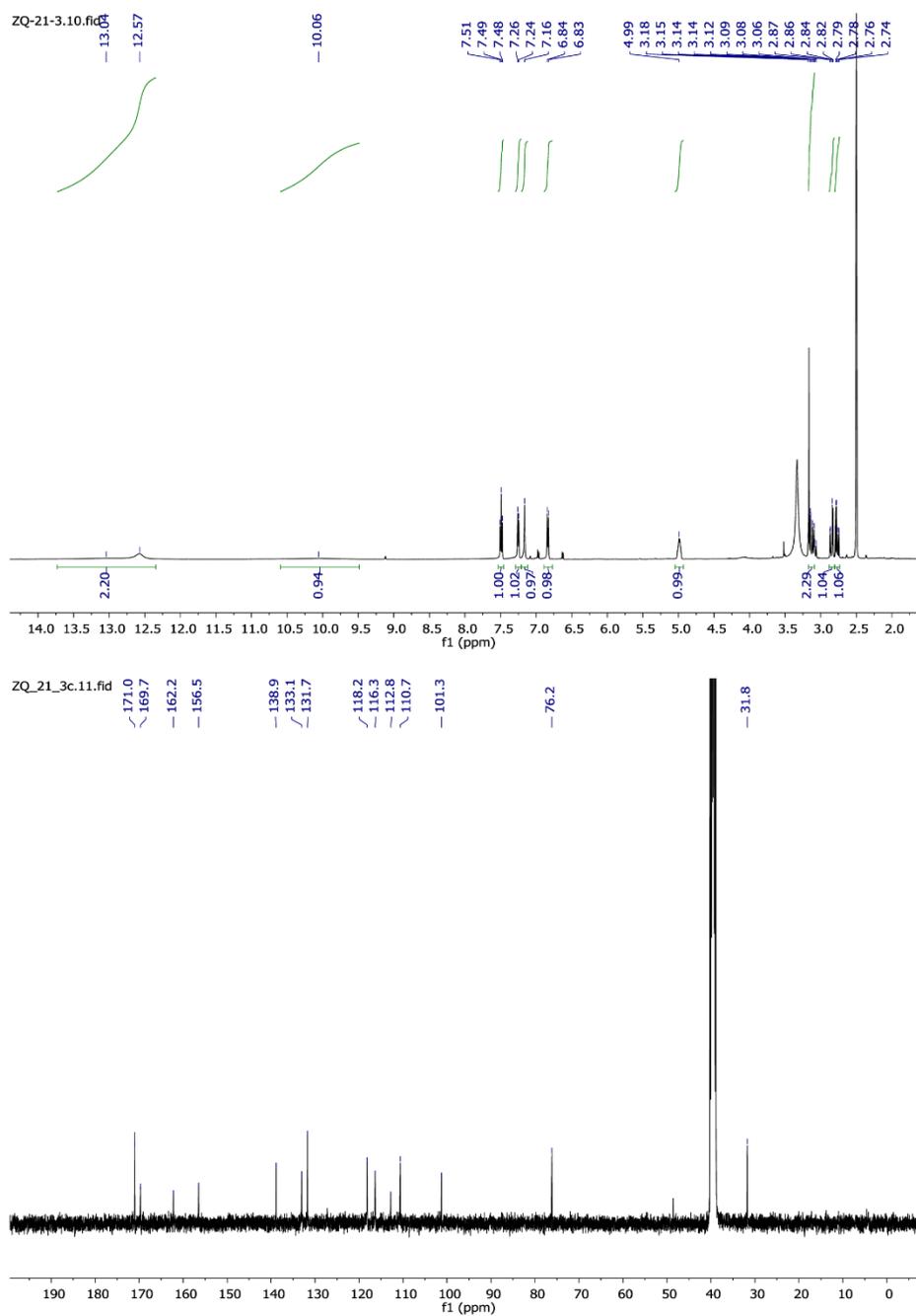
Figure S32. ^{1}H (top) and ^{13}C-NMR spectra (bottom) of compound **4**.
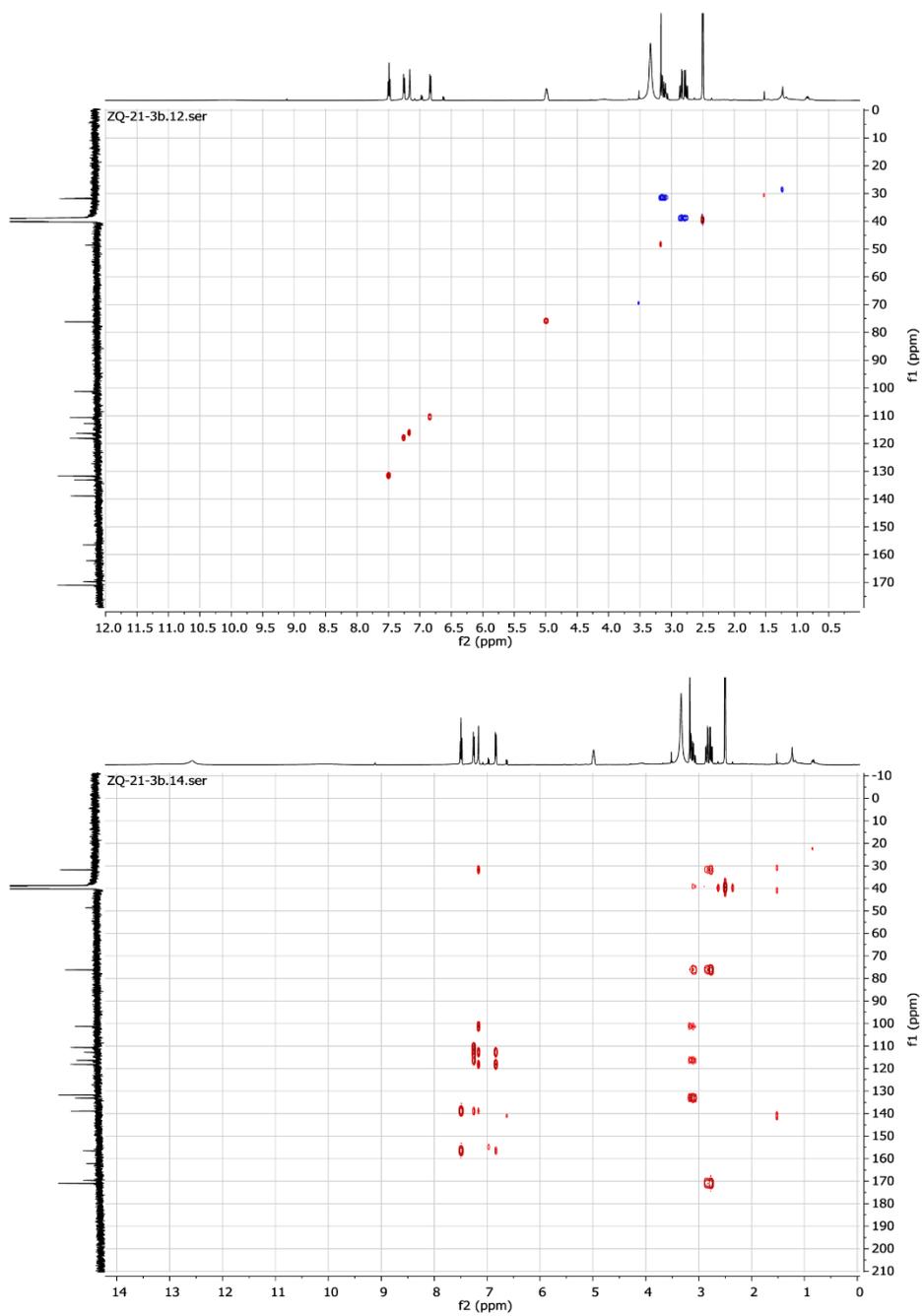
Figure S33. HSQC (top) and HMBC-NMR spectra (bottom) of compound **4**.
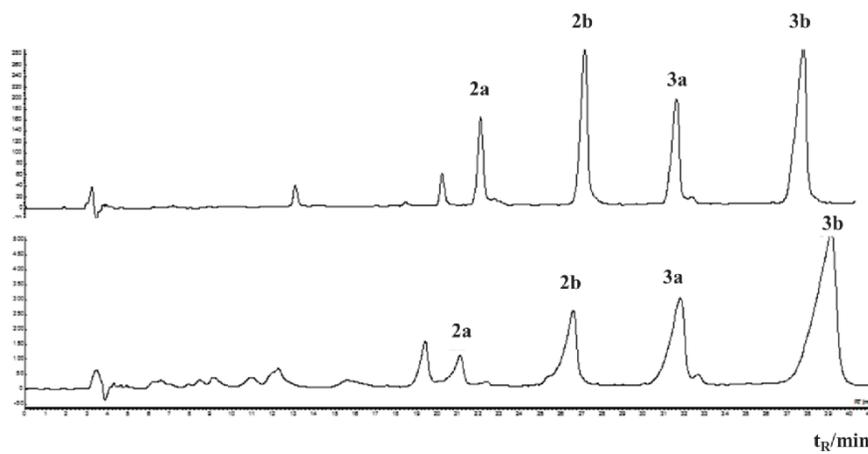
Figure S34. The compound ratio between acids **2a/b** and methyl esters **3a/b** changes over time upon standing in MeOH. Top: HPLC-UV trace of compound mixture directly after injection. Bottom: HPLC-UV trace several hours after dissolving extract in MeOH. The amount of **3a/b** increases, while **2a/b** content drops, pointing at a non-enzymatic esterification reaction.

Heat of formation (B3LYP/def2-TZVP), number of imaginary frequencies, and cartesian coordinates of the equatorial and the axial conformation of **3a**.

3a (equatorial), Heat of Formation: -1262.298918310823 Eh

No imaginary frequencies

C	-18.69634544813687	1.24675014577728	0.35813574796113
C	-18.71779973718725	-0.14162668565488	0.18859895556447
C	-17.54471810005603	-0.86455477618494	0.13881271785282
C	-16.28411511680621	-0.20922359140401	0.28357330770858
C	-16.28344607402373	1.21834412041186	0.44136697438836
C	-17.50530083446073	1.91938314421800	0.47698136246379
C	-15.02129472070770	-0.86067867481514	0.26390558594929
C	-13.84005969534935	-0.14681954398220	0.33924600344363
C	-13.85826035737579	1.26092601704586	0.50874563401713
C	-15.05324005846412	1.91299053435621	0.55992530861224
C	-12.59473712671422	-0.86880760367963	0.25098571308097
O	-11.42422178254779	-0.19372138989737	0.29047209803886
C	-11.46507463292313	1.18835999028420	-0.13288991493086
C	-12.52776288588565	1.94507897674530	0.64594546095264
O	-14.99861639469285	-2.23244748225670	0.16622277162647
C	-13.81466984419974	-2.95887433368056	0.14454028789850
C	-12.59329875895881	-2.21543472494498	0.15888396024912
O	-14.60713160972048	-6.47821768247045	-0.18781533063667
C	-14.88709662855022	-5.31893805878886	0.08702862813045
C	-13.82850254862824	-4.31666300131180	0.07611300269442
O	-17.65579906238748	-2.19302559745171	-0.06917265808034
C	-10.05489651769326	1.71983939563819	0.07297237759322
C	-9.90710312716273	3.16468346547049	-0.34756905484143
O	-9.46105970810992	4.04476070838979	0.34050941281935
O	-10.34313284636511	3.35954780430226	-1.60893120374241
C	-10.25036818247227	4.71302912190165	-2.09030151760930
C	-16.30909633700252	-4.97412701871073	0.47081474303768
H	-19.63186140474308	1.79009451568959	0.39207725337828
H	-19.64947780723379	-0.67935953257774	0.07867895232711
H	-17.48543063776575	2.99379717687787	0.60556730165185
H	-15.08059380926272	2.98767460182471	0.69618408081049
H	-11.71264924832122	1.20355412213313	-1.19773236753235
H	-12.22991180283231	1.98770964383350	1.69918678568373
H	-12.58743957754365	2.97147065098359	0.28069985898812
H	-11.66476727400965	-2.76291970052018	0.09836405599668
H	-12.84737284540813	-4.76226964706750	-0.02802833553959
H	-16.77910924892884	-2.60223241913439	-0.07038540339537
H	-9.77187679032851	1.64290978134096	1.12158887164483
H	-9.36315638096415	1.11171690929246	-0.51406678031148
H	-10.67103766863385	4.69888457301967	-3.09123165726548
H	-10.81613684588805	5.38259407938047	-1.44432141575618
H	-9.21066077040664	5.03629597968605	-2.11091812981207
H	-16.79160096784116	-5.88005969716339	0.83241433758977
H	-16.36650946371728	-4.20124111645357	1.23600507616244
H	-16.86075931958897	-4.63895318045235	-0.41138285886287

3a (axial), not used for the ECD calculation, Heat of Formation: -1262.299055406826 Eh

No imaginary frequencies

C	-19.06350660626086	1.00219264149527	0.60708977062272
C	-19.08157144146503	-0.38983419285334	0.46687190833326
C	-17.91007030067547	-1.09610755957650	0.29430809616351
C	-16.65415543957643	-0.41653247281669	0.26698129267402
C	-16.65859869374659	1.01489985511192	0.38685936865277
C	-17.87968573456393	1.69623594953923	0.56589919029250
C	-15.39103678687371	-1.04682464561119	0.10595495167376
C	-14.22752063110107	-0.31061184050073	-0.01852941057215
C	-14.25177713607346	1.10515228546673	0.06727784655544
C	-15.43975915205710	1.73478650720068	0.29777775091484
C	-12.99376557300051	-1.01706417523933	-0.26028173807424
O	-11.85308146681830	-0.32106949514169	-0.46455596408068
C	-11.75342660074146	0.96444192592982	0.19835900323292
C	-12.95328766182520	1.83298037457156	-0.15395453538385
O	-15.35376190801211	-2.42194495326998	0.05761451105583
C	-14.16956635423957	-3.13023217054735	-0.09495148653536
C	-12.98006085534861	-2.36660347121158	-0.30930015011837
O	-14.80622186300747	-6.68476802345686	0.29109384489091
C	-15.13463246603299	-5.50515662178398	0.32082097422235
C	-14.14608458273283	-4.48896157348711	-0.02044514865627
O	-18.02117326976268	-2.43174334927577	0.13358419426791
C	-11.54294788404168	0.71233150574619	1.69029775307034
C	-10.96267494166881	1.90045277208588	2.42033758039769
O	-11.01405926728564	3.04700895376338	2.04821176152887
O	-10.36989429688913	1.51004949095068	3.55959774434184
C	-9.77258399811374	2.54331473812893	4.36215565268919
C	-16.54264217625260	-5.14145693135655	0.73359664054501
H	-19.99809598486560	1.53132433512597	0.74203284092799
H	-20.01009213132999	-0.94412559217157	0.48084633200859
H	-17.86520612295478	2.77399110217995	0.66364001655273
H	-15.47319276039046	2.81384694338182	0.39201363783740
H	-10.85041615544600	1.40329746671064	-0.22127788648991
H	-12.90941251035991	2.75598133401473	0.42153489820521
H	-12.87369962410396	2.10758407876253	-1.21036342553946
H	-12.05477353821407	-2.90006209728544	-0.46605067506256
H	-13.17266677604097	-4.92042454521555	-0.21598884623397
H	-17.14314047899950	-2.82742022227366	0.04224449146697
H	-12.48360769117846	0.43933350809080	2.17807051937138
H	-10.87162118369212	-0.13704075335411	1.81695575002949
H	-9.07223291009682	3.12753053442127	3.76695216078675
H	-10.54149967573671	3.20308331890625	4.76318171413521
H	-9.25745860129506	2.02392107830257	5.16515054856048
H	-16.99787017509141	-6.01315399202089	1.19878784349143
H	-16.56850254648506	-4.30349819694674	1.42945272547673
H	-17.13571404555213	-4.88236382449021	-0.14714404822922

Heat of formation and Gibbs free enthalpy (B3LYP/def2-TZVP), number of imaginary frequencies, and cartesian coordinates of the equatorial and the axial conformation of **4**.

4equ

Heat of Formation: -1030.156703967375 Eh

Gibbs Free Enthalpy: -1029.94901907 Eh

No imaginary frequencies

```
C   -11.69238114677195    0.76783776601595    0.44952621214811
C   -11.81388881287684   -0.64335443308506    0.24743784449737
C   -13.10675466117885   -1.21861262218924    0.09295380660277
C   -14.23625397221701   -0.45797421381730    0.13111939587375
C   -14.13482591245121    0.95016461978019    0.31274681584131
C   -12.88226739037940    1.54346798980030    0.48418356285828
C   -10.39152416790787    1.33260209511106    0.61717547435456
C   -9.27540065169274     0.51947977366222    0.57369987552591
C   -9.41164377765927    -0.85670277564119    0.37761228976196
C   -10.65098953822003   -1.43546234984335    0.21794072708181
C   -15.62230160129236   -1.03501251974324    0.05396806807260
C   -16.59082166084035   -0.04283128877416   -0.56766764067399
O   -16.51866510759997    1.24500924257967    0.10055969331722
C   -15.32230466796599    1.79075508581556    0.36602882979822
O   -12.78063640447147    2.86744775290244    0.69289482084836
O   -10.19306087175647    2.64871988828369    0.82181602074326
O   -15.29403971122785    2.98420817914017    0.65998810231222
C   -18.05362603583495   -0.43887024876225   -0.45787013646610
C   -18.46535702268164   -1.53528001034748   -1.40368843102152
O   -19.70118920096277   -1.99313296130243   -1.10378226116094
O   -17.82202851065966   -1.95286556262032   -2.33126322534395
H   -13.17726276430009   -2.29016643399344   -0.05022820501076
H   -8.30354953491013     0.97638717867203    0.69995095801388
H   -8.52195922826786    -1.47319299312004    0.35134239485267
H   -10.74437205015585   -2.50273642777314    0.06555672316088
H   -15.62903444889205   -1.95392711294706   -0.53207700470506
H   -15.96927131151539   -1.28343799238095    1.06354823680255
H   -16.32883031262025    0.11894325584501   -1.61623989301897
H   -13.71849867324404    3.22608302352236    0.74227965878118
H   -11.05191809079193    3.10575579012331    0.83322609359210
H   -18.30479286077483   -0.73019934968322    0.56416539218170
H   -18.67223046309672    0.43240485034344   -0.69074206449158
H   -19.92691943478221   -2.66570719557357   -1.76616213512977
```

4ₐₓ

Heat of Formation: -1030.156778912124 Eh

Gibbs Free Enthalpy: -1029.94815384 Eh

No imaginary frequencies

C	-11.35166996645877	0.84554248268484	0.52304935484320
C	-11.49660644988310	-0.54552475358073	0.22156561170467
C	-12.77679109837172	-1.05631601113911	-0.13160541000942
C	-13.87110923178822	-0.25055929998513	-0.19545711464368
C	-13.75365331684890	1.13288637197815	0.12206175012811
C	-12.51034505508576	1.66778105109200	0.45993016242887
C	-10.05792433065013	1.34604606985485	0.86006054676783
C	-8.97083955254676	0.49272658453832	0.87672416565795
C	-9.13110163384813	-0.86253810459097	0.58048461454712
C	-10.36687685304766	-1.38157912199162	0.26528713388047
C	-15.23027843103165	-0.72207094303192	-0.61978569334214
C	-16.30677772731975	0.03805580708484	0.13572643822397
O	-16.13860085281633	1.46831196025139	-0.05927990987623
C	-14.91950076539385	2.00874559850244	0.10188128872055
O	-12.38802151780709	2.97728026240635	0.73986573496899
O	-9.83772191732707	2.63832265229450	1.17173119489709
O	-14.84738508938222	3.22922964177535	0.22560071358145
C	-16.33772990672268	-0.22197139733647	1.63839809445338
C	-16.77671500633294	-1.61782283760325	1.99284242113543
O	-16.68882507677949	-1.83344299955823	3.32180167814751
O	-17.16583651807675	-2.45579283023429	1.21956043567568
H	-12.86069935832380	-2.10833509173444	-0.37539255951630
H	-8.00201189291869	0.90203622575714	1.12697400490381
H	-8.26413696599720	-1.51089767143074	0.60124516380301
H	-10.48104726511470	-2.43205784436686	0.03219321256041
H	-15.36912851866659	-0.53034099679699	-1.68898855806997
H	-15.35575614553362	-1.79230882447232	-0.46162833026572
H	-17.28419719244209	-0.17556780273223	-0.29058846099130
H	-13.28867696194289	3.39205388628916	0.58000802440056
H	-10.67181740410713	3.13314293439646	1.09670652202896
H	-15.36138400698725	-0.04123285344894	2.09266356948091
H	-17.02839193136990	0.47999940491129	2.11142778814153
H	-17.00048105907717	-2.73695154978283	3.48983641163328

References

1.	Bierman, M.; Logan, R.; Obrien, K.; Seno, E. T.; Rao, R. N.; Schoner, B. E., Plasmid Cloning Vectors for the Conjugal Transfer of DNA from Escherichia-Coli to Streptomyces Spp. *Gene* **1992,** *116* (1), 43-49.

2.	Fu, J.; Bian, X.; Hu, S.; Wang, H.; Huang, F.; Seibert, P. M.; Plaza, A.; Xia, L.; Muller, R.; Stewart, A. F.; Zhang, Y., Full-length RecE enhances linear-linear homologous recombination and facilitates direct cloning for bioprospecting. *Nat. Biotechnol.* **2012,** *30*, 440-6.

3.	Paget, M. S.; Chamberlin, L.; Atrih, A.; Foster, S. J.; Buttner, M. J., Evidence that the extracytoplasmic function sigma factor ςE is required for normal cell wall structure in Streptomyces coelicolor A3 (2). *J. Bacteriol.* **1999,** *181*, 204-211.

4.	Gomez-Escribano, J. P.; Bibb, M. J., Engineering Streptomyces coelicolor for heterologous expression of secondary metabolite gene clusters. *Microb. Biotechnol.* **2011,** *4*, 207-15.

5.	Epstein, S. C.; Charkoudian, L. K.; Medema, M. H., A standardized workflow for submitting data to the Minimum Information about a Biosynthetic Gene cluster (MIBiG) repository: prospects for research-based educational experiences. *Stand Genomic Sci* **2018,** *13*.

6.	Lomovskaya, M.; Doi-Katayama, Y.; Filippini, S.; Nastro, C.; Fonstein, L.; Gallo, M.; Colombo, A. L.; Hutchinson, C. R., The *Streptomyces peucetius* dpsY and dnrX Genes Govern Early and Late Steps of Daunorubicin and Doxorubicin Biosynthesis. *J. Bacteriol.* **1998,** *180*, 2379-86.

7.	Krohn, K.; Vukics, K., First Chemical Synthesis of the Antiviral Agents S2502 and S2507. *Synthesis* **2007,** 2894-900.

8.	Kunnari, T.; Kantola, J.; Ylihonko, K.; Klika, K. D.; Mäntsälä, P.; Hakala, J., Hybrid compounds derived from the combination of anthracycline and actinorhodin biosynthetic pathways. *J. Chem. Soc., Perkin Trans 2*, **1999**, 1649-52.

S III. Supplemental information for cycloheptamycin biosynthesis

The following supplemental information is related to the following publication which was highlighted in Chapter 3.3:

Z. Qian, J. Antosch, P.M. D'Agostino, T. Liu, M. Fottner, R. Zhu, A. Pöthig, T.A.M. Gulder, et al.. Functional characterization of the biosynthesis of the antibiotic cycloheptamycins. Manuscript in preparation.

Approval letter from publisher

Permission for reproduction of OBC article

Structures and biological activities of cycloheptamycins A and B

Z. Qian, J. Antosch, J. Wiese, J.F. Imhoff, H. Fiedler, A. Pöthig and T.A.M. Gulder, *Org. Biomol. Chem.*, **2019**, *17*, 6595, **DOI:** 10.1039/C9OB01261C

This paper is available online:
https://pubs.rsc.org/en/content/articlelanding/2019/ob/c9ob01261c/unauth#!divAbstract

Permission for reproduction of JOC article

Copyright Clearance Center

RightsLink®

🏠 Home   ? Help   📞 Live Chat   👤 Zhengyi Qian ⌄

ACS Publications
Most Trusted. Most Cited. Most Read.

Discovery of the Streptoketides by Direct Cloning and Rapid Heterologous Expression of a Cryptic PKS II Gene Cluster from Streptomyces sp. Tü6314

Author: Zhengyi Qian, Torsten Bruhn, Paul M D'Agostino, et al

Publication: The Journal of Organic Chemistry

Publisher: American Chemical Society

Date: Nov 1, 2019

Copyright © 2019, American Chemical Society

PERMISSION/LICENSE IS GRANTED FOR YOUR ORDER AT NO CHARGE

This type of permission/license, instead of the standard Terms & Conditions, is sent to you because no fee is being charged for your order. Please note the following:

- Permission is granted for your request in both print and electronic formats, and translations.
- If figures and/or tables were requested, they may be adapted or used in part.
- Please print this page for your records and send a copy of it to your publisher/graduate school.
- Appropriate credit for the requested material should be given as follows: "Reprinted (adapted) with permission from (COMPLETE REFERENCE CITATION). Copyright (YEAR) American Chemical Society." Insert appropriate information in place of the capitalized words.
- One-time permission is granted only for the use specified in your request. No additional uses are granted (such as derivative works or other editions). For any other uses, please submit a new request.

BACK   CLOSE WINDOW

Erklärung

Hiermit versichere ich, dass ich, Zhengyi Qian, 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