Exploring chemical diversity of angucycline antibiotics: molecular basis of simocyclinone and grecocycline biosynthesis

Dissertation

zur Erlangung des Grades

des Doktors der Naturwissenschaften

der Naturwissenschaftlich-Technischen Fakultät III

Chemie, Pharmazie, Bio- und Werkstoffwissenschaften

der Universität des Saarlandes

von

Oksana Bilyk

Saarbrücken

2015

Tag des Kolloquiums:	11. Dezember 2015
Dekan:	Prof. Dr. Dirk Bähre
Berichterstatter:	Prof. Dr. Andriy Luzhetskyy
	Prof. Dr. Rita Bernhardt
Vorsitz:	Prof. Dr. Gert-Wieland Kohring
Akad. Mitarbeiter:	Dr. Michael Kohlstedt

“The important thing is not to stop questioning.”
Albert Einstein

ACKNOWLEDGMENT

I would like to express my deepest gratitude to my supervisor Prof. Andriy Luzhetskyy for the continuous support of my PhD studies, patient guidance, enthusiastic encouragement, and challenging my thinking by interesting discussions. I always could find his advice and inspiration when needed.

Besides my advisor I would like to thank Prof. Rita Bernhardt for being the second reviewer of this thesis and Prof. Andreas Bechthold in whose group I have started doing my PhD.

I owe my special thank you to Dr. Thomas Paululat for collaboration on NMR analysis and structure elucidation. My thank you goes also to Dr. Elke Brötz for her help with metabolites purification and occasional help on chemical questions, Dr. Bogdan Tokovenko for bioinformatical support, Dr. Sascha Baumann for his help with the topoisomerase activity assays and Prof. Sergey Zotchev and Dr. Olga Sekurova for their collaboration on TAR cloning.

I would like to express thanks to all the members, former and present, of the AMEG research group. I would like to thank Yuriy for his insightful comments and advises, Lilya and Maria for being cheerful and supportive. I also want to thank Bohdan, Yousra, Maksym, Birgit, Tatjana, Natja, Niko, Stephie, Suvd and Constanze (including those already mentioned) for all the wonderful moments.

At the same time I would like to thank Prof. Rolf Müller and his group for tight collaboration and all the discussions.

Ця дисертація ніколи б не побачила світ, як би я не вчилась в Львівському фізико-математичному ліцеї. Михайло Михайлович, завдяки Вам я зрозуміла, що біологія може бути цікавою. Лілія Тадеївна, Ви дали нам знання з хімії, які не можливо було б отримати де-інде. Валентина Олексіївна, безмежне дякую Вам за прекрасні уроки культорології, за те що прививали нам любов до Берніні і Мікеланджело, Мане і Ренуара, Кнута і Рембо. Я хочу подякувати і всім іншим вчителям, які вкладали в нас свою душу.

Також я хочу сказати дякую викладачам в університеті, зокрема кафедрі біохімії, а також Гнатуш Світлані Олексіївній. І звичайно ж співробітникам Інституту біології клітини, де мене навчили всіх практичних основ молекулярної біології.

Однак, найбільше я є вдячна своїй сім‘ї. Богдан, Ти є, був і завжди будеш моєю найбільшою підтримкою. Я хочу подякувати Тобі за кожну хвилину, що ми провели разом від нашого знайомства і до сьогоднішнього дня. Дякую що Ти завжди був поруч зі мною, допомагав і завжди чекав з теплою вечерею, коли я поверталась пізно з роботи. Я вдячна Тобі за всі наші подорожі, за Твою любов і за те що подарував мені нашого Данилка.

Данчику, Ти є моїм світлим промінчиком. Та любов і щастя, які Ти даруєш мені однією посмішкою, неможливо описати словами. Ти навчив мене цінувати час і змінив мої пріорітети.

Я дуже хочу подякувати моїм мамі і татові. Ви подарували мені життя, дозволяли мені бути самостійною, приймати важливі рішення і йти своїм шляхом. Дякую, що і зараз, коли ми є так далеко одне від одного, Ви знаходите час, силу і безмежне бажання допомогти нам. Ви відкрили для мене музику, мистецтво і світ літератури. Дякую Вам за все. Ляночка, дякую за Твою підтримку, розуміння і за те що ти є саме тією людиною до якої я йду коли мені найважче.

Ще я хочу подякувати Андрію, але тепер як другу, а не науковому керівнику. Дякую за Твою підтримку, мотивацію і в те що вірив в мене навіть тоді, коли я перестала. Дякую, що не дозволив мені забути ким я є і знайти свою дорогу.

Without any of you, mentioned above, this thesis would have not been possible.

Oksana Bilyk
September, 2015

PUBLICATIONS

1. **Bilyk, O.**, Brötz, E., Tokovenko, B., Bechthold, A., Paululat, T., Luzhetskyy, A. (2015) New simocyclinones: surprising evolutionary and biosynthetic insights. Submitted to *ACS Chemical Biology*

2. **Bilyk, O**., Sekurova, O., Zotchev, S., Luzhetskyy, A. (2015) Cloning and heterologous expression of the grecocycline biosynthetic gene cluster. *to be submitted*

3. **Bilyk, O**., Tokovenko, B., Luzhetskyy, A. (2015) Draft genome sequence of *Kitasatospora nimpheas*. *to be submitted*

4. Brötz, E., **Bilyk, O.**, Kröger, S., Paululat, T., Bechthold, A., Luzhetskyy, A. (2014). Amycomycins C and D, new angucyclines from *Kitasatospora* sp. Tetrahedron Lett. 2014 Oct; 55(42):5771-3.

SELECTED CONFERENCES

Bilyk, O., Brötz, E., Tokovenko, B., Bechthold, A., Paululat, T., Luzhetskyy, A. "Molecular genetics of simocyclinones biosynthesis: surprising assembly of the tetraene chain", oral presentation at: (a) 6th Congress of European Microbiologists (FEMS 2015) June 7-11, 2015; (b) International VAAM Workshop, Technische Universität Dresden, October 2-5, 2014

Bilyk, O., Sekurova, O., Zotchev, S., Luzhetskyy, A. "Reconstruction of the grecocycline biosynthetic pathway from *Streptomyces* Acta 1362 using transformation-associated recombination in yeast", poster presentation at 17th International Symposium on the Biology of Actinomycetes (ISBA XVII), October 8-12, 2014

Bilyk, O., Paululat, T., Luzhetskyy, A. "Novel simocyclinones from *Kitasatospora* sp.", poster presentation at 1st European Conference on Natural Products (ECNP), September 22-25, 2013

ABSTRACT

Members of the angucycline group of antibiotics are produced by different *Streptomyces* species and share common structural feature – the four-ring aglycon frame organized in a planner manner. Simocyclinone D8 is a potent inhibitor of the DNA gyrase supercoiling activity. It is the highly complex hybrid antibiotic, which comprises four parts that are coming from three pools of primary metabolism: angucyclic and linear polyketides, deoxysugar moiety and aminocoumarin. Although a hypothesis explaining simocyclinone biosynthesis has been previously proposed, little was proven *in vivo* due to the genetic inaccessibility of the producer strain. Three new D-type simocyclinones (D9, D10 and D11) produced by *Kitasatospora* sp. and *Streptomyces* sp. NRRL-24484, as well as their biosynthetic gene clusters, have been identified. The gene inactivation and expression studies have disproven the role of a modular polyketide synthase (PKS) system in the assembly of the linear dicarboxylic acid. Instead, the new stand-alone ketosynthase genes were shown to be involved in the biosynthesis of the tetraene chain. Additionally, the gene responsible for the conversion of simocyclinone D9 into D8 had been identified. Also biosynthetic gene cluster involved in the biosynthesis of another angucyclic polyketides grecocyclines from *Streptomyces* sp. Acta 13-62 was identified and cloned. Grecocyclines have unique structural moieties such as a dissacharide side chain, an additional amino sugar at the C-5 position and a thiol group. Enzymes from this pathway may be used for the derivatization of known active angucyclines in order to improve their desired biological properties.

ZUSAMMENFASSUNG

Zugehörige zur Antibiotika Gruppe der Angucycline werden von verschiedenen *Streptomyces* Arten produziert und weisen strukturelle Gemeinsamkeiten, wie den planaren 4-Ring Aglykon Rahmen auf. Simocyclinon D8 ist ein starker Gyrasehemmer und verhindert damit das Supercoiling der DNA. Es ist ein hochkomplexes Hybrid Antibiotikum, das aus 4 Teilen, die aus 3 Gruppen des Primärstoffwechsels kommen, besteht: angucyclische und lineare Polyketide, Desoxyzuckerteile und Aminocoumarin. Obwohl kürzlich eine Hypothese aufgestellt wurde, die die Biosynthese von Simocyclinonen erklärt, ist *in vivo* aufgrund der genetischen Unzugänglichkeit des produzierenden Stammes nur wenig nachgewiesen. Drei neue D-Typ Simocyclinone (D9, D10, D11), produziert von *Kitasatospora* sp. und *Streptomyces* sp. NRRL-24484, sowie deren biosynthetische Gencluster wurden identifiziert. Die Gen Inaktivierungs- und Expressionsstudien haben die Rolle eines modularen Polyketidsynthasesystems (PKS) beim Zusammenfügen der linearen Dicarboxylsäure widerlegt. Stattdessen wurde gezeigt, dass die neuen eigenständigen Ketosynthase Gene an der Biosynthese der Tetraenkette beteiligt sind. Darüber hinaus wurde das für die Umwandlung von D9 in D8 verantwortliche Gen identifiziert. Auch das biosynthetische Gencluster, das an der Biosynthese eines anderen angucyclischen Polyketids, Grecocyclin von *Streptomyces* sp. Acta 13-62, beteiligt ist, wurde identifiziert und geklont. Grecocycline haben einzigartige strukturelle Reste wie die Disaccharid-Seitenkette, einen zusätzlichen Aminozucker am C5 Atom und eine Thiolgruppe. Enzyme dieses Wegs können für die Derivatisierung bekannter aktiver Angucycline verwendet werden um deren gewünschte biologische Eigenschaften zu verbessern.

TABLE OF CONTENTS

I. INTRODUCTION

1. Polyketides as a source for new therapeutics

The polyketides are one of the most remarkable classes of natural products, distributed among bacteria, fungi, plants and animals. Few thousands of polyketides have been discovered till date from different sources. They exhibit vast structural and functional diversity, in addition to broad spectrum of biological and pharmacological activities. These secondary metabolites have been successfully implemented in medicine and agriculture, as they demonstrate antibiotic (tetracycline, erythromycin), anticancer (daunorubicin), anticholesterol (lovastatin), antifungal (griseofulvin, nistatin), antiparasitic (avermectin), insecticide (pamamycin, nonactin) and immunosuppressive (rapamycin) properties [1] (Figure 1).

Today many polyketides and their derivatives are available on the market. Total sales of the more than 40 polyketide medicines reach more than 20 billion US $ a year, close to the total sales of protein therapeutics [2]. Because of their industrial potential and importance, intensive efforts have been made in finding novel polyketides since the middle of the 20th century.

In order to deal with increasing development of the multidrug-resistant microorganism and the lack of the effective therapy for many severe diseases, discovering of the new natural products with improved properties is extremely urgent. The polyketides are one of the most promising niches due to their fascinating structural diversity that gives rise to the broad spectrum of biological activities.

2. Discovery of the polyketide biosynthetic pathways

Original function of polyketides in their native environment is so far not clear in all cases. They are believed to function as pigments, virulence factors, as molecules of defense or signaling molecules [3]. However, all of them share one common feature – the way how they are synthesized. Despite their structural differences, they are mainly derived from the acetic acid – one of the simplest building blocks in the nature. Intriguingly, the vast structural and functional diversity of polyketides results from the controlled assembly of some of the simplest building blocks in nature – acetate and propionate [3].

In 1893 Collie and Myers isolated a first polyketide – orcinol, while attempting to prove the structure of dehydroacetic acid. Later that year James N. Collie coined the term “polyketide”

and proposed that orcinol and related compounds most probably are synthesized by repetitive condensation or polymerization reactions [4]. Later this hypothesis was proven experimentally by A. Birch and coworkers. Birch was studying biosynthesis of the aromatic polyketide 6-methylsalicylic acid (6-MSA) in fungi *Penicilum patulum*. He used isotopically labeled acetate and showed that 6-MSA is formed from four acetate units. With the advent of the modern molecular biology techniques in the second half of the 20th century, it became possible to analyze how polyketides are made within the living cell.

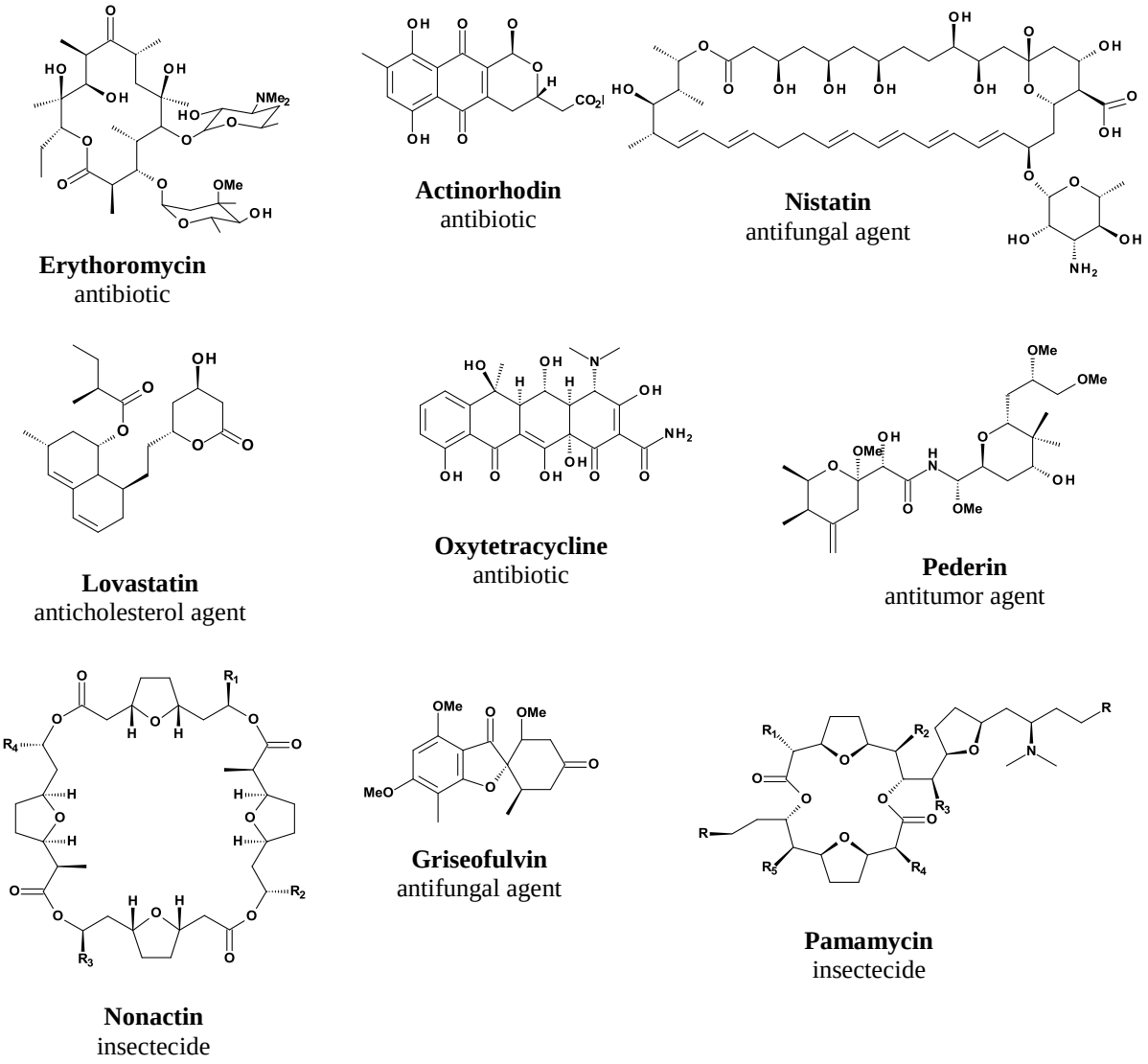


Figure 1. Examples of polyketide secondary metabolites

Actinorhodin, a blue pigment produced by *S. coelicolor*, was the first polyketide understood in "genetic and molecular biological" terms [5]. Through series of experiments, the genes responsible for the actinorhodin biosynthesis were identified and later sequenced. Using DNA

recombinant technology it was possible to clone DNA fragment, able to complement all classes of actinorhodin non-producing mutants of *S. coelicolor*. Furthermore, a heterologous expression of the cloned DNA in *S. parvulus*, directed the biosynthesis of the blue pigment [6]. The sequences of established actinorhodin proteins showed high homology with the enzymes from fatty acid biosynthesis.

3. Biochemistry of polyketides biosynthesis

3.1. Polyketide biosynthesis vs. fatty acid biosynthesis

The two biosynthetic pathways – fatty acid biosynthesis and polyketide biosynthesis are closely related and share common features – pool of simple precursors that are used to build the growing polyketide chain, such as acetyl-CoA and malonyl-CoA and the type of enzymes that are employed for chain assembly [3, 4]. In general, both, polyketides and fatty acids are assembled by successive rounds of decarboxylative Claisen condensation between an activated acyl starter unit and malonyl-CoA derived extender unit [3]. The simplest set of enzymes that is sufficient to build the polyketide chain involves β-ketoacylsynthase (KS), which catalyses condensation and phosphopantethienylated acyl carrier protein (ACP) or coenzyme A (CoA), which serve as an anchor for the growing chain. Optionally, malonyl-acetyl transferase (MAT) might transfer the building blocks from the coenzyme A pools to the ACP. The result of the condensation is a generation of ACP-bound β-ketoacylthioester. The phosphopantethiene functions in ACP as a flexible arm that carriers the growing chain and delivers it to the enzymes that are modifying nascent polyketide. After every chain elongation the keto ester can be reduced to an alcohol by ketoreductase (KR), dehydrated to alkine by dehydratase (DH) and alkine can be hydrogenated by an enoyl reductase (ER) to yield fully saturated acyl backbone. During the fatty acid biosynthesis (performed by FAS – fatty acid synthase) successive cycles lead to a chain of the required length (14, 16 or 18 carbon atoms). The chain is released by the thioesterase enzyme (TE). All FASs have the same set of components – KS, ACP, KR, DH, ER, MAT and TE and typically catalyze a full reductive cycle after each elongation. In contrast, during polyketide chain assembly, modification of the keto group is optional and can be completely excluded. If the reduction of the keto group is skipped, it will be present in the polyketide; when no functional DH is present, alcohol is incorporated into the chain and skipping of the enoyl reductase reaction will yield unsaturated chain. Polyketide biosynthesis differs from fatty acid biosynthesis in many other ways. For example, polyketide synthase can utilize much broader spectrum of the biosynthetic building blocks (e.g methylmalonyl-CoA, hydroxymalonyl-CoA, ethylmalonyl-CoA and others),

whereas FASs uses only – malonyl-CoA. The utilization ofvarious starter and extender units, the optionality of the keto group downstream modifications,the introduction of a chirality and variations in the cyclisation of the polyketide chain can yield a fascinating complex polyketide [3].

Based on the architecture pattern and mode of action PKSs are classified on different types. Type I PKSs are gigantic polypeptides, that comprise a certain number of linear arrange and covalently fused functional modules. Type II PKS refers to a dissociable of discrete and typically monofunctional domains. Chalcone synthase, or type III PKS, is a homodimeric protein that is not depended on ACP and acts directly on CoA-activated substrate. In addition, PKSs are categorized as iterative or noniterative, depending on whether each module catalyzes one or more rounds of elongation (Table 1).

Table 1. Types of PKSs

PKS	ACP	Strater unit	Extender unit	Organisms
Type I modular				
• noniterative cis-AT	Yes	various	various	Bacteria, protists
• noniterative trans-AT	Yes	various	various	Bacteria, protists
• Iterative	Yes	various	Malonyl-CoA	Fungi, some bacteria
Type II (iterative)	Yes	various	Malonyl-CoA	Bacteria
Type III (iterative)	No	various	Malonyl-CoA	Plants, bacteria, fungi
PKS-NRPS hybrids	Yes	Various, amino acids	Malonyl-CoA, amino acids	Bacteria, fungi

3.2. Modular polyketide synthases

The erythromycin synthase (6-deoxyerythronolide B synthase, DEBS) from *Saccharopolyspora erythraea* is the prototypical PKS for polyketide biosynthesis. Since the DEBS genes were cloned, there was an explosive growth of PKS sequences discovery. The approx. number of assembly line polyketides sequences available in NCBI database is more than 1000 [7], nevertheless, DEBS is still one of the most favorable model for scientists to study the details and mechanisms of polyketide biosynthesis. The first DNA sequences and analyses of the erythromycin biosynthetic gene cluster have been published in 1990th, simultaneously by two groups [8-11]. The first macrolide intermediate during erythromycin

biosynthesis is 6-deoxyerythronolide B (6-dEB). Three gigantic genes - *eryAI*, *eryAII* and *eryAIII* are responsible for the biosynthesis of 6-dEB. They encode three multienzyme polypeptides –DEBS 1, 2 and 3, respectively (~ 350 kDa each). Interestingly, the DEBS proteins are relatively small, compared to those, involved in biosynthesis of, for example, rapamycin, rifamycin and nystatin [4].

Each DEBS is made of two modules (DEBS 1 – modules 1 and 2, DEBS 2 – modules 3 and 4 and DEBS 3 – modules 5 and 6). In addition, DEBS 1 harbors loading domain (LD) and DEBS 3 – TE. Loading domain of DEBS 1 consists of AT and ACP and accepts the starter unit – propionate, whereas TE is responsible for the release of the polyketide chain and its' subsequent cyclisation into the lactone ring. Each module contains three domains for catalyzing elongation reaction – KS, AT and ACP and a variable set of domains involved in the reduction steps. The growing polyketide chain is bound to the synthases throughout the entire biosynthetic process (Figure 2).

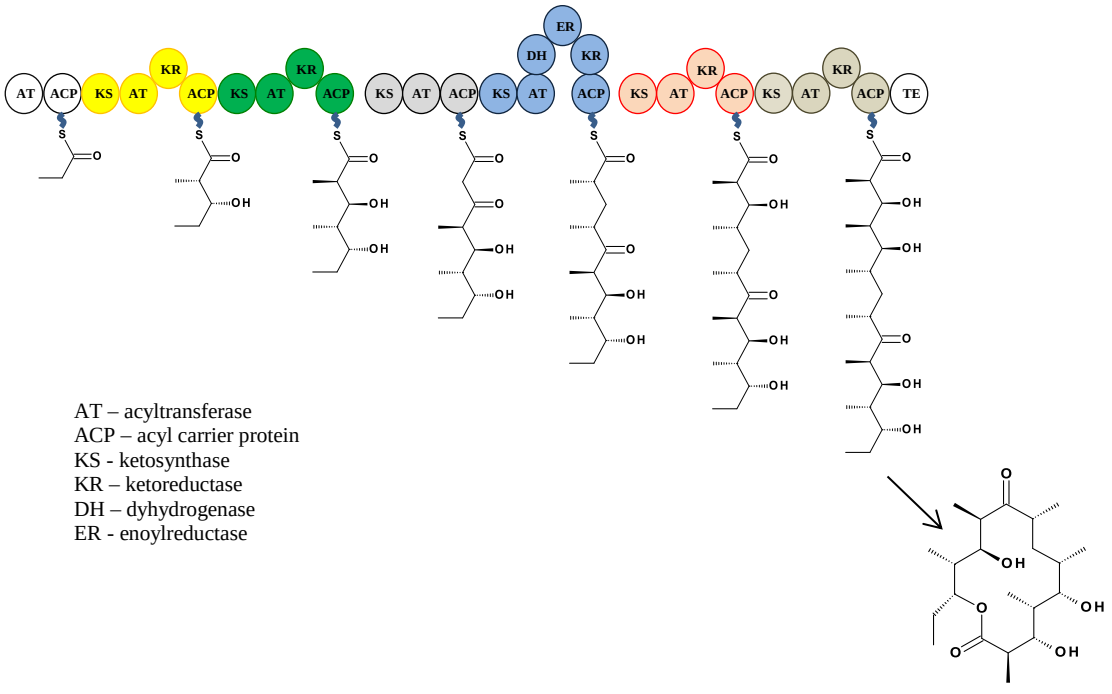


Figure 2. Domain organisation of the erythromycin polyketide synthase. Putative domains are represented as circles.

To generate 6-dEB, one propionyl-CoA, six methylmalonyl-CoA units and six NADPH are required. The erythromycin biosynthesis is initiated by a propionyl-CoA uptake from the

primary metabolism pool and transfering it onto the loading module.The propionyl moiety is passed to the first module that simultaneously picks up (2*S*)-methylmalonyl-CoA from the cell environment. Aftewards, methylmalonyl-CoA is decarboxylated and attached to the propionyl anchored on the ACP of the module 1. This chain is then passed downstream with the incremental addition of precursors at each module of the assembly line, resulting in the linear heptaketide intermediate that will be off-loaded and cyclized. 6-dEB should be further decorated to form the mature erythromycin molecule. Usually, the polyketides are prepared in a "skeleton" form and later are derivatized through various kinds of modification reactions (hydroxylation and sugar incorporation for 6-dEB) to give the final product.

Sometimes the order of domains in the modular PKSs may not correspond to the polyketide pattern. The rapamycin polyketide synthase (RAPS) represents one of the first example [12]. RAPS comprises three large polypeptides – RAPS 1 (~900 kDA, 4 modules), RAPS 2 (~1 MDa, 6 modules) and RAPS 3 (660 kDa, 4 modules). Modules 3 and 6 of RAPS contain KR and DH domains (plus ER domain in module 3) that are nominally active, but functionally redundant. There is a chance that these domains were disabled in the way that is not visible from the primary sequence, or, perhaps they are functional but sites of their action are modified in the post-translational events. Mismatch between PKS sequence and the polyketide pattern was observed in other cases as well [4].

3.3. Type I iterative PKS

Type I iterative polyketide synthases are close "relatives" of the modular PKSs. Type I iterative PKSs have thesimilar protein architecture, as modular PKSs, however, the single module in this case acts iteratively – reusing domains in a cyclic fashion [13]. AviM, involved in biosynthesis of avilomycin, was the first type I iterative PKS discovered in bacteria [14]. It comprises KS, AT, DH and ACP domains typical for type I PKS. At first, AviM was treated not as iterative type I PKS from bacteria, but rather as exception from type II PKS paradigm [13].

3.4. Type II PKSs

Type II PKS are involved in a biosynthesis of natural products that give rise to a number of clinical useful drugs and promising drug candidates. Type II PKS – derived polyketides are structurally diverse and are classified as anthracyclines, angucyclines, aureolic acids, tetracyclines, tetracenomycins, pradimicin-type polyphenols, and benzoisochro-manequinones [15]. Type II PKS, or so-called "minimal" PKS, are comprised of several individual enzymes,

each encoded by a separate gene. A standard set of enzymes includes two ketosynthase units (KS_α and KS_β) and an ACP [16-19]. Almost in all gene clusters "minimal" PKS genes are grouped into a typical KS_α/KS_β/ACP operon [15]. Genes encoding enzymes that are involved in further modification of the nascent linear polyketide chain (cyclases, aromatases and ketoreductases), usually, are clustered in a close proximity to the PKS genes. Further modifications of the poly-β-keto intermediate usually include hydroxylation, methylation and glycosylation.

In contrast to type I PKSs which may employ a great variety of starter and extender units, type II PKSs usually use acetyl-CoA and malonyl-CoA, respectively. KS_α and KS_β are highly similar, however only KS_α catalyze C-C bond formation through Claisen condensation. KS_β (or chain length factor, CLF) is deficient for cysteine in its active center, which is crucial for the polyketide assembly. Despite the mutation in the active center, it was shown that KS_β plays an important role in the polyketide chain elongation, as it is a primary determinant of a polyketide chain length and is generating acetyl KS after the decarboxylation of malonyl-ACP [15, 20]. Since the genes encoding MCATs are absent in most of the type II PKS gene clusters, there has been an ongoing debate regarding the transfer of the malonyl extander units to the PKS. It was proposed that an endogenous MCAT is recruited from fatty acid biosynthesis. An alternative model is a self-malonylation of ACP, as it has been demonstrated *in vitro*. During the biosynthesis KS and CLF form a heterodimer and the protein cleft keeps the nascent poly-β-keto intermediate extended in the way, where highly reactive keto groups are separated and probably contact KS_α-KS_β mainly as enols [15]. A typical length of polyketides synthesized by type II PKS is either 16, 20 or 24 C atoms. The length of the polyketide chain is controlled by the protein cavity formed by KS and CLF. Khosla and coworkers [21] have demonstrated that regions defining chain length are located at the interface of the dimer. It was suggested that the residues 109, 112 and 116 in the actinorhodin CLF serve as a gatekeepers at the KS/CLF tunnel and reducing the length of this residues allows two more extension cycles. Additionally, interactions of KS/CLF with other proteins have effect on the polyketide chain length [21]. Other components of type II PKS complex have influence on the chain length as well, for example disruption of an aromatase/cyclase from the oxytetracycline (*otc*) gene cluster lead to the production of novel, shorter, polyketides. Similar effect was shown for *whiE*, *act*, and hedamycin (*hed*) [22-24]. Also KRs from actinorhodin and hedamycin biosynthetic gene clusters are able to modulate chain length specificity of KS-CLF [22]. In conclusion, it seems that not only CLF but the entire type II PKS complex determines the polyketide chain length.

3.5. Priming for type II PKS

In most cases aromatic PKS are primed by decarboxylation of malonyl-CoA to yield an acetyl-KS intermediate, which is subsequently elongated by the minimal PKS [15]. However, many type II PKS deviate from decarboxylation mechanism and are primed with non-acetate units [25]. Two alternative pathways are involved: presence of the monofunctional acyl-CoA-ligase (ACL) and AT, which are responsible for the activation and transfer of the starter unit to the "minimal" PKS and involvement of the additional KSIII (and ACP) for biosynthesis and attachment of the short-chain fatty acids [25].

The first mechanism of the alternative mechanism for type II PKS priming was intensively studied for enterocin, where biosynthesis is primed with benzoyl-CoA that is formed from phenylalanine via β-oxidative route and oxytetracycline, where malonamate is activated and transferred to the minimal PKS [25-27].

Scenario, where KSIII is implicated in the biosynthesis of primer unit was discovered for daunorubicin, frenolicin, R1128 and others. Daunorubicin and doxorubicin gene clusters contain two additional PKS elements – FabH-like KS and AT [28, 29]. It was suggested that KSIII catalyzes first condensation between malonyl-CoA and propionyl-CoA to yield β-keto ester, as in fatty acid biosynthesis. Later the diketide would be loaded on the minimal PKS by AT. Hutchinson and Strohl groups, in independent studies, showed that *dps* PKS exhibit relaxed starter unit specificity to acetyl-CoA in the absence of the KSIII [30, 31].

In general, presence of KSIII (and additional ACP) in the type II PKS gene cluster is a good indication for priming with non-acetate unit. However, there are exceptions as well, for example during the biosynthesis of the polyketide with the anticancer activity – hedamycin [22].

3.6. Pamamycin and nonactin – polyketide biosynthesis beyond the paradigm

There are several exception from the PKSI/PKSII paradigm. The most striking are ketosynthases involved in the biosynthesis of nonactins and pamamycins. Typically, succinate is not used as a building block by PKSs as it cannot directly participate in Claisen condensation, however feeding experiments clearly demonstrate that there are at least two natural products – pamamycin and nonactin, known to utilize succinate as a building block [32]. Pamamycins and nonactin are a group of macrodiolide antibiotics of the streptomycetes origin with a number of biological activities [32, 33]. Pamamycins are composed of two asymmetrical parts – hydroxy acid small and hydroxy acid large, while nonactin of four

molecules of enantiomeric nonactic acid [32, 34]. Studies using the ^{13}C labeled precursors showed a direct incorporation of 3-oxoadipate into the nonactin and utilization of acetate and succinate as building blocks for its assembly [35]. Despite the structural evidence, molecular mechanisms of their biosynthesis remained unclear for a long time. The gene cluster for the biosynthesis of nonactin (*non*) contains five unusual KS that were classified as non-iterative type II PKS based on sequence similarity despite the lack of ACP [33, 34]. Pamamycin gene cluster (*pam*) share a high degree of homology with the *non* gene cluster, it harbors genes encoding seven discrete KS enzymes and one ACP (*pamC*). Deletion of *pamC* shifted pamamycin biosynthesis toward the accumulation of lower molecular weight compounds, indicating that PamC takes part in the delivery of starter units, as described above. All KSs in the *pam* cluster, except for PamA, exhibit active site triad typical for type III PKS. PamA was shown to participate in the first condensation step between succinate and malonyl- or methylmalonyl-CoA to yield 3-oxoadipyl-CoA, which represents an intermediate between the secondary and primary metabolism. Interestingly, 3-oxoadipyl-CoA and 2-methyl-3-oxoadypil-CoA have to be rotated by acyltransferase PamB to further participate in the biosynthetic pathway. The resulting 4-oxoadipyl-CoA and 5-methyl-4-oxoadipyl-CoA are used as extenders for a Claisen condensation facilitating the incorporation of succinate and the starter units most probably are supplied as ACP-esters. Small and large hydroxy acids, which are synthesized in the following steps, later are re-activated by an acyl-CoA ligase, PamL and the closure of the pamamycin macrodiolide ring is perfomed by PamJ and PamK KSs that catalyze an unusual C-O condensation reaction (Figure 3). The function of PamA was proven in *in vitro* and *in vivo* studies and so far is a first studied enzyme responsible for the incorporation of succinate into a polyketide molecule [32].

4. Simocyclinone D8

Simocyclinone D8 (SD8) was isolated in 2000 from the mycelium extract of *S. antibioticus* Tü6040, collected in Iguaçu, Argentina [36]. SD8 is a unique natural hybrid antibiotic, as it consists of four different moeities that are comming from at least three primary metabolites pools [37]. Angucyclic polyketide is joined with the aminocoumarin part through the D-olivose sugar and the polyunsaturated linear polyketide (Figure 4). The aromatic polyketide core is characterised by a large number of unusually placed hydroxyl groups and an epoxy group placed between C12a and C6a. D-olivose is acetylated at 4-OH position and linked to the tetraene chain is attached to its 3-OH group. The aminocoumarin moeity is halogenated at C17'' position and linked to the tetraene side chain via an amid bond.

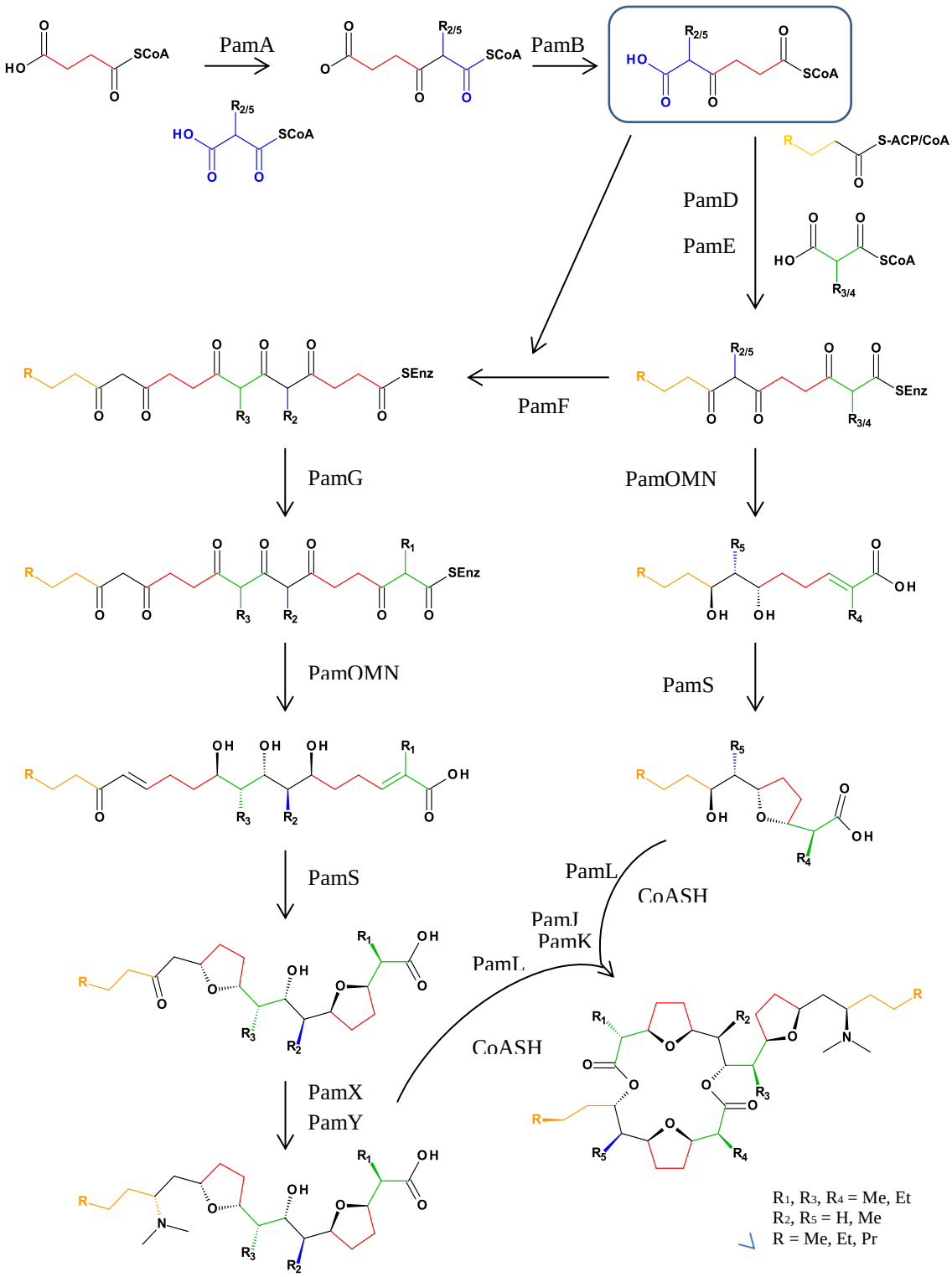


Figure 3. Proposed pathway for the biosynthesis of pamamycin

Beside SD8, *S. antibioticus* also produces other simocyclinones that can be considered as intermediates in the SD8 biosynthesis. These simocyclinones are divided into four types: 1) A-type simocyclinones consist of an aromatic polyketide moiety; 2) B-simocyclinones with

D-olivose is attached to the angucycline; 3) C-type simocyclinones - angucyclic polyketide plus deoxysugar plus tetraene chain and 4) D-type is a full molecule with the four functional parts.

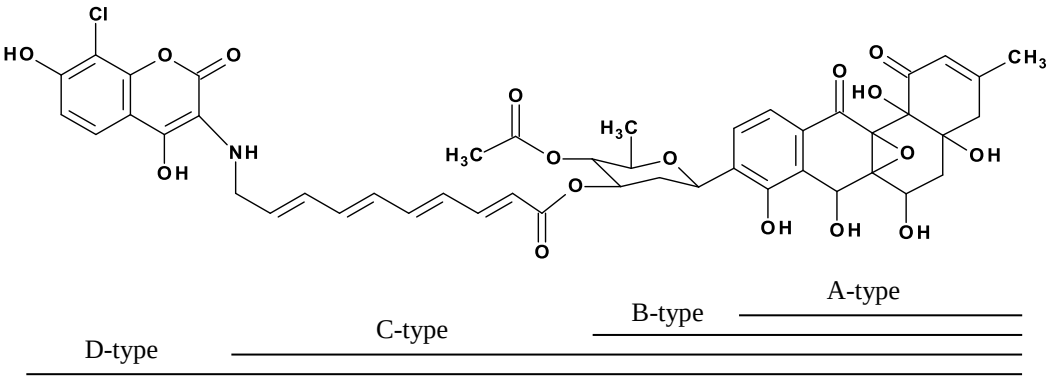


Figure 4. Simocyclinone D8

4.1. Biosynthesis of simocyclinone

SD8 biosynthetic gene cluster from *S. antibioticus* Tü6040 was reported by the groups of Bechthold [38] and Heide [39] and revealed 49 ORFs. The genes responsible for the biosynthesis of the aromatic polyketide (*simA* genes) show high similarity to the angucycline encoding genes from various gene clusters. A set of the minimal PKS genes and "auxiliary" genes (*simA1-simA5, simA11*) has been identified as well as genes for ketoreductases and oxygenases (*simA6-simA10, simA12*) that are involved in the modification of the nascent polyketide chain. The deduced amino acid sequence of SimA13 shares high similarity with P450 hydroxylases and is one of the candidates to be involved in the formation of the epoxy group in the angucyclic polyketide (personal communications). All *simA* genes are clustered together, except *simA13*, which is located in the different locus. The *simB*-series genes are involved in the formation and the attachment of the D-olivose moiety. The deoxyhexose biosynthesis starts from the activation of glucose-1-phosphate by NDP-D-glucose synthase (SimB1) with the subsequent dehydratation (SimB2 and SimB3) and ketoreduction (SimB4 and SimB5). SimB7 is the glycosyltransferase attaching the D-olivose moiety to the angucycline and SimB6 is the acetyltransferase involved in the acetylating of the glycosylated angucycine. Several simocyclinones from B-, C- and D-series carry not acetylated D-olivose, indicating that incorporation of an acetyl group take place after the attachment of the deoxysugar to the aglycon. Notebly, the *simB* genes are not placed together – *simB1*, *simB2*,

simB3 and *simB7* are located on the one end and *simB4*, *simB5* and *simB6* on the other end of the cluster. The *simD1-simD6* genes encode enzymes involved in converting of L-tyrosine to aminocoumarin. It was suggested that L-tyrosine is activated by covalent binding to a 4-phospho-pantetheinyl cofactor of SimD6 (SimH) similar way as during the novobiocin biosynthesis [39, 40]. Subsequently L-tyrosine is modified to the β-keto intermediate by SimD1 and SimD2, and afterwards is oxidatively cyclased by SimD3 yielding the aminocoumarin part. An amid bond between the aminocoumarin and the octatetraene dicarboxylic acid moiety in SD8 is formed by SimD5 (SimL). *In vitro* studies on SimL done by the groups of Heide [41] and Walsh [42] demonstrated the SimL relaxed specificity to structurally diverse substrates what was successfully applied to generate a novel novobiocin analogue. *simD4* (the only *simD* gene located elsewhere in the *sim* cluster) is highly similar to the halogenases and presumably preforms the formation of the chlorinated SD8. *simX1* (*simY*), located between *simD1* and *simD2* in the *sim* gene cluster, is a MbtH-like protein, found in many secondary metabolites gene clusters [43]. When the *sim* cluster was published *simX1* was annotated as the gene of an unknown function [38]. However, later Boll et al., [43] demonstrated that SimH (SimD6) and its analogues from other aminocoumarin biosynthetic gene clusters, require MbtH-like proteins for their catalytic activity in a molar ratio 1:1. In the absence of MbtH-like proteins activity of the tyrosine-adenylating enzymes was lowered by 99.0 – 99.8%.

Three ORFs of the *sim* cluster encode large, multifunctional type I PKSs [38]. SimC1A contains three modules (one loading module and two extension modules), while SimC1B and SimC1C have only one extension module each (Figure 5). The domains number and order in SimC1A-SimC1C PKSs do not perfectly correspond to the observed polyketide pattern. Activities of the KR and DH domains are predicted to be required for the biosynthesis of the tetraene chain, but the module II does not contain a DH domain, and there are no DH and KR domains in the module IV. The ER domain from the module III is not needed for the octatetraene dicarboxylic acid formation and the AT-specificity to the methylmalonyl substrates in modules I and IV is not consistent with the polyketide side chain structure [38]. However, as mentioned above, similar observations have been made for other natural product clusters as well. A number of genes encoding enzymes, which might be involved in the tailoring reactions on the assembled polyketide chain, have been identified in the *sim* cluster.

Feeding experiments with [1,3-$^{13}C_2$]- and [2-^{13}C]malonic acid confirmed that angucyclic polyketide is built from 10 acetate units and the octatetraene dicarboxylic acid derives from

five acetate units, starting from the left side (C-10") [37]. Also, biosynthesis of simocyclinone in an $^{18}O_2$ rich atmosphere showed signal shifts for C-6a, C-12b, C10"and C17a" positions, in accordance to the proposed ways of the aminocoumarin and the tetraene chain formation [44].

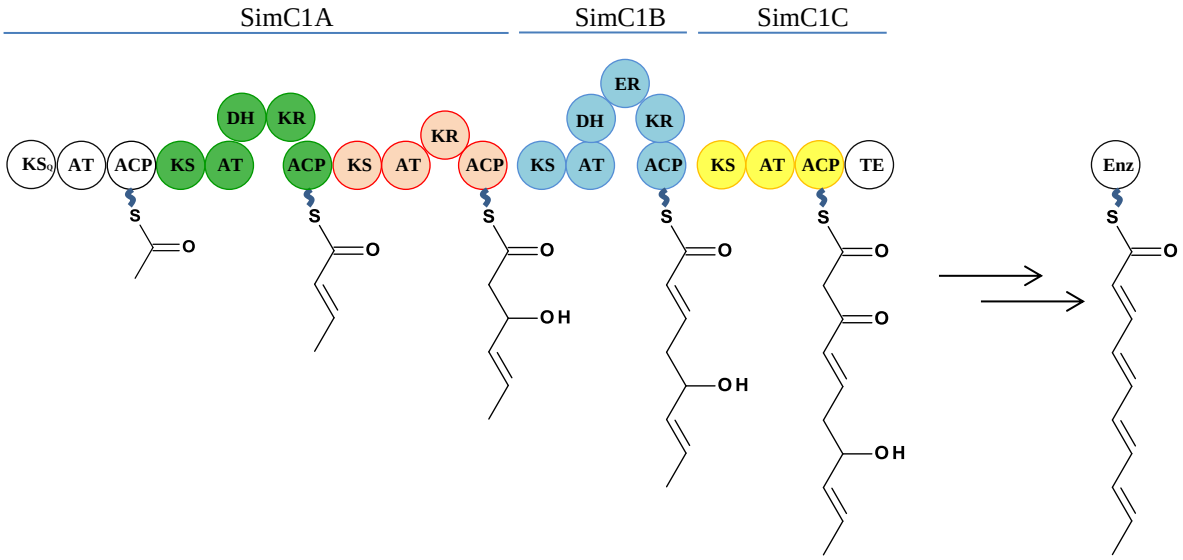


Figure 5. Proposed model for the tetraene chain biosynthesis in *S. antibioticus* Tü6040

Two transporter (*simEx1*, *simEx2*) and three regulatory (*simReg1*, *simReg2* and *simReg3*) genes have been identified in the *sim* cluster [38]. SimReg1 is a key regulator of the simocyclinone biosynthesis, as it activates the transcription of the simocyclinone structural genes, transporter genes, regulatory gene *simReg3* and its own transcription [45]. SimReg2 (SimR) and SimEx1 (SimX) resemble the TetR/TetA repressor-efflux pump pair and are involved in simocyclinone efflux and resistance. The *simX* promoter is directly repressed by SimR, which binds to two operator sites in the *simR-simX* intergenic region [46]. SimR was shown to dissociate from the intergenic regions in the presence of the SD8 [46-49]. The operator of SimReg2 partially overlaps with the DNA-binding region of SimReg1upstream to the start codon of *simX*, indicating the competition between SimReg1 and SimReg2 for binding to the *simX* promoter (P_{SEx1}) [45]. Bechthold and co-workers suggested the following preliminary model for the regulation of simocyclinone biosynthesis and export. When the concentration of simocyclinone or its intermediates is low, SimR binds to the *simX* regulatory region and its transcription is repressed. At the same time SimReg1 activates expression of the

structural *sim* genes and thus the simocyclinone biosynthesis. SimR is released from the simX promoter when the concentration of simocyclinone increases, allowing SimReg1 to bind to the P_{SEx1}. This activates the *simX* expression and efflux of simocyclinone from the cell. This mechanism couples the biosynthesis of simocyclinone to its export [45].

4.2. Simocyclinone D8 – DNA gyrase inhibitor

SD8 is a potent inhibitor of the DNA gyrase supercoiling and has activity against gram-positive bacteria, certain clinical isolates of *Escherichia coli* and distinct cytostatic activities against human tumor cell lines and [36, 50].

DNA topoisomerases are enzymes that modulate DNA topology to maintain a chromosome superstructure and integrity [51]. These enzymes are divided in two types depending on their mechanism of action which can be implemented through braking one (type I) or two (type II) strands of DNA. The DNA gyrase belongs to the type II topoisomerases and is the only one among this type capable of catalyzing DNA supercoiling, in the reaction driven by ATP hydrolysis [52].

The DNA gyrase consists of two subunits, GyrA (97 kDa in *E.coli*) and GyrB (90 kDa in *E.coli*) that form a heterodimer. The GyrA subunit is involved in the interaction with DNA, and contains tyrosine in the active-site responsible for DNA cleavage, while the ATPase active site is located in the GyrB. There are two mechanisms of DNA gyrase inhibition by the antibacterial drugs: gyrase poisoning, or stabilization of the covalent enzyme-DNA complex and inhibition of the gyrase enzymatic activity [52]. Ciprofloxacin is an example of the cleavage-complex stabilizing agent, and novobiocin is a compound that inhibits the ATPase activity of the gyrase is an example of catalytic inhibitors. Novobiocin was the first aminocoumarin, which was introduced into the human anti-infective therapy in 1964 under the name Albamycin ® and used for the treatment of *Staphylococcus aureus* infections, including MRSA [53]. Four of five natural aminocoumarins discovered, novobiocin, clorobiocin, coumermycin and simocyclinone are antibiotics that target DNA gyrase, but they do so in the entirely different ways [53]. The “classical” aminocoumarins – novobiocin, clorobiocin, coumermycin and their derivatives inhibit the gyrase ATPase reaction by competing with ATP for binding to GyrB. Their binding site lies in the 24-kDa terminal sub-domain of GyrB and overlaps with the ATP-binding site, preventing ATP binding. Specifically, the binding site of the adenine ring of ATP overlap the binding site of the noviose sugar of aminocoumarins [52].

Surprisingly, despite carrying the aminocoumarin moiety, SD8 does not target the ATP-binding site of GyrB, but binds to the N-terminal domain of GyrA, and prevents its interaction with DNA [54, 55]. The interaction between DNA gyrase and SD8 have been well studied, including two crystal structures of a complex formed between a 59-kDa [54] and 55-kDa [56] N-terminal domain fragment of the *E. coli* GyrA subunit. Analysis of the GyrA59-SD8 complex revealed two binding pockets that separately accommodate the aminocoumarin and the angucycline parts of SD8, forming a cross-linked GyrA59 tetramer in the crystal. Additional analyses were performed to prove that it was not a crystallographic artifact [54]. A new crystal structure of SD8 with GyrA55 was generated to further explain the mechanism of the DNA gyrase inhibition by this antibiotic. The conformation of SD8 in GyrA55-SD8 was significantly different to that observed in GyrA59-SD8, first of all because SD8 was bound entirely within one GyrA55 homodimer. The angucycline part was shifted to a position where it spans the interface between the two GyrA monomers, whereas the aminocoumarin binding pocket is essentially the same as observed before, but its orientation is different. Also, no additional electron density adjacent to the angucycline, which was modeled as Mg^{2+} in the GyrA59-SD8 complex, has been identified. The observed position of SD8 across the dimer interface "staples" it closed, preventing conformational changes that need to occur to allow DNA binding. Through the evaluation of point mutants, which were generated earlier to probe GyrA-SD8 interactions, Maxwell and co-workers have shown, that the new structure is more likely to be representative of the mode of action of SD8 on DNA gyrase [56].

Interestingly, another binding site for SD8 in GyrB was proposed by Sissi et al., [57]. Additional binding was seen on the GyrB47 domain lacking ATP- and coumarin-binding site, probably affecting the enzyme's interface and impairing DNA binding. Later, Maxwell and cowerkers [56], based on the different affinity of SD8 to GyrA and GyrB have suggested that the interaction between SD8 and GyrB may be an in vitro artifact, manifesting in the absence of GyrA.

SD8 show potency against both the *E. coli* and *S. aureus* gyrase, but not topoisomerases IV [52]. IC_{50} value for SD8 was calculated as at least two times lower than that for novobiocin and approximately seven times lower than that for ciprofloxacin [55]. SD8 also exhibit potent antiproliferative activity against certain cancer lines, where its target appears to be human topoisomerase II (Topo II) [52]. Interestingly, SD8 was shown to inhibit the relaxation activity of human topoisomerase I (Topo I) [58]. This observation was unexpected as unlike human Topo II, human Topo I doesn't have any structural similarities to the gyrase. The

inhibition effect of SD8 on human Topo I and II are essentially identical in the relaxation assays. The IC_{50} value of SD8 on human Topo I was approximately 2-fold higher than that of clinical drug camptothecin, whereas the IC50 value of SD8 for human Topo II was approximately 2-fold lower than that of etoposide [58].

Fluoroquinolones and aminocoumarins are the only two classes of the gyrase inhibitors, which have been introduced into clinical anti-infective therapy. Over the years fluoroquinolones like ciproflaxicin have become one of the most successful antibacterial agents. However, there usefulness is becoming limited by the spread of fluoroquinolone-resistant bacteria [53]. Coumarins are potent inhibitors of gyrase with the limited resistance, as mutations in GyrB subunit show loss of enzymatic activity due to their proximity to the ATPase active site. On the other hand, their clinical success is limited by eukaryotic toxicity, poor activity against Gram-negative bacteria and low solubility. Therefore, searching for a new gyrase inhibitors is of a great interest. SD8 as a potent gyrase inhibitor with the novel mode of action may serve as a useful lead compound for the development of new antibacterial and anticancer drugs.

5. Transformation-associated recombination in yeast

Nowadays more than 11000 secondary metabolite gene clusters from Actinobacteria are available in the databases, the vast majority of which are “orphans” that synthesize products of yet unknown structure [59]. Heterologous expression of the orphan biosynthetic gene clusters is a possibility to explore the Actinobacteria biosynthetic potential. Heterologous expression has been used to discover new compounds by screening genomic libraries, to analyze natural product biosynthesis using genetic engineering and to confirm the clustering of secondary metabolite biosynthetic genes [60]. Natural product gene clusters may reach up to 100kb in size and even more, but cosmid and fosmid library inserts are restricted in size of approximately 40 kb and further manipulations are needed to assemble the entire biosynthetic pathway on a single vector from the multiple cosmids [61].

Transformation associated recombination (TAR) approach is an alternative, based on the homologous recombination to selectively capture a locus of interest from the DNA pool [62]. In the TAR protocols the “capture” vector with the short homology arms to the sequence of interest is cotransformed with the genomic DNA into the yeast *Saccharomyces cerevisiae*. TAR cloning in yeast takes advantage of its natural very efficient recombination activity – the capture vector arms and the homologous target DNA undergo recombination to yield a stable

plasmid containing the genomic region of interest [63-65]. Such plasmids are maintained as circular yeast artificial chromosomes (YACs) and are capable of carrying inserts up to 1 Mb [66]. Originally TAR was developed to facilitate cloning of large genomic DNA fragments without necessity to construct and screen genomic libraries. However, the scope of this methodology was extended, showing that it could be used for direct cloning of the biosynthetic pathways, reassembling eDNA-derived clones into a single construct and even for the assembly of a 1,077,947 bp functional synthetic genome from 1078 1-kb cassettes [63, 64, 66, 67].

Moore and coworkers [63] in 2014 have successfully adapted the TAR-based approach for the capture of a silent 67-kb biosynthetic pathway from the genomic DNA of a marine actinomycete. The followed heterologous expression of this gene cluster led to discovery of the novel antibiotic taromycin A. TAR enables straightforward and specific capture of a target biosynthetic gene cluster, without the size limitations or time consuming and laborious generation and screening of genomic libraries. Furthermore, it is a simple but efficient strategy for the discovery, characterization and modification of the natural products from bacteria that are hard to manipulate under the regular laboratory conditions.

6. Outline of this work

The purpose of this work was to explore molecular basis of biosynthesis of two angucyclic antibiotics: simocyclinone and grecocycline.

Two new producer strains for simocyclinones, *Kitasatospora* sp. and *Streptomyces* sp. NRRL B-24484, have been identified using genome mining, leading to purification and structure elucidation of novel D-type simocyclinones. The genomes of the producing strains were sequenced and the simocyclinone biosynthetic gene clusters identified. A comparison of the newly identified simocyclinone gene clusters with the previously published one provides several important and surprising insights into the biosynthesis of these complex natural products. Moreover, the development of genetic tools for the producing *Kitasatospora* strain allowed characterization of genes involved in the biosynthesis of the tetraene.

In addition, transformation-associated recombination in yeast had been successfully employed for cloning and heterologous expression of the grecocycline biosynthetic gene cluster. Unique glycosyltransferase, oxygenase and thioesterase involved in grecocycline biosynthesis had been identified.

II. RESULTS

1. New simocyclinones: surprising evolutionary and biosynthetic insights

Abstract

Simocyclinone D8 has attracted attention due to its highly complex hybrid structure and the unusual way it inhibits bacterial DNA gyrase by preventing DNA binding to the enzyme. Although a hypothesis explaining simocyclinone biosynthesis has been previously proposed, little was proven *in vivo* due to the genetic inaccessibility of the producer strain. Herein, we report discovery of three new D-type simocyclinones (D9, D10 and D11) produced by *Kitasatospora* sp. and *Streptomyces* sp. NRRL B-24484, as well as the identification and annotation of their biosynthetic gene cluster. Unexpectedly, the arrangement of the newly discovered biosynthetic gene clusters is starkly different from the previously published one, despite nearly the identical structures of D8 and D9 simocyclinones. The gene inactivation and expression studies have disproven the role of a modular polyketide synthase (PKS) system in the assembly of the linear dicarboxylic acid. Instead, the new stand-alone ketosynthase genes were shown to be involved in the biosynthesis of the tetraene chain. Additionally, we identified the gene responsible for the conversion of simocyclinone D9 into D8.

Introduction

Simocyclinones are a family of complex hybrid natural products from *Streptomyces antibioticus* Tü6040 (*1*). Simocyclinone D8 (**1, SD8**) shows the strongest activity against Gram-positive and Gram-negative bacteria and also has the ability to stop the proliferation of tumor cells (*2*, *3*). It comprises a chlorinated aminocoumarin linked to an angucyclic polyketide via a tetraene linker and an acetylated D-olivose sugar. SD8 (**1**) inhibits DNA gyrase by preventing DNA binding to the enzyme (*4-6*). The crystal structure of simocyclinone bound to the enzyme revealed two binding pockets, one for aminocoumarin and one for the angucycline moiety. This unique mode of action on the validated drug target makes SD8 (**1**) an interesting candidate for drug development. However, the structural complexity of the molecule represents a major problem for medicinal chemistry's ability to generate a library of simocyclinones derivatives with improved properties. Alternatively, biosynthetic pathway engineering can provide a platform for simocyclinone derivatization. This approach, however, requires the investigation and detailed elucidation of the simocyclinone biosynthetic pathway. The biosynthetic gene cluster responsible for SD8 (**1**) has been cloned and annotated by the Bechthold and Heide groups (*7, 8*). However, despite the extensive efforts of several laboratories, the simocyclinone biosynthetic pathway could not be engineered because of the genetic inaccessibility of the production strain. This did not allow researchers to use the power of metabolic engineering to decipher the simocyclinone biosynthetic pathway and to date only two enzymes have been characterizedSimD5 (SimL) is an amid-forming ligase that attaches the aminocoumarin to the tetraene chain (*9-11*)and SimC7 is an angucyclinone ketoreductase (*12*).

In the present study, we report the genomic-driven identification and structure elucidation of novel D-type simocyclinones from *Kitasatospora* sp. and *Streptomyces* sp. NRRL B-24484. The genomes of the producing strains were sequenced and the simocyclinone biosynthetic gene clusters identified. A comparison of the newly identified simocyclinone gene clusters with the previously published one provides several important and surprising insights into the biosynthesis of these complex natural products. Moreover, the development of genetic tools for the producing *Kitasatospora* strain allowed us to characterize genes involved in the biosynthesis of the tetraene.

Results and Discussion

Genomic based identification of the new simocyclinone-producing strains

Access to the large and steadily increasing volume of genomic data enables rapid identification of bacteria species that harbor biosynthetic gene clusters that are responsible for the biosynthesis of a certain class of compounds. The genome mining/alignment of natural producers significantly helps in prioritizing which strain has the potential to synthesize the compound of interest or its derivatives. We have focused on the screening of simocyclinones producers due to their remarkable antibacterial activities and puzzlingly complex chemical structure. Two strains, *K.* sp. and *S.* sp NRRL B-24484, have been identified that carry sub-clusters of genes in their genomes that are highly similar to those of the simocyclinone biosynthetic gene cluster from *S. antibioticus* Tü6040. Indeed, cultivation of *K.* sp. and *S.* sp NRRL B-24484 allowed us to isolate simocyclinone D9 (**2,** SD9), simocyclinone D10 (**3,** SD10) and simocyclinone D11 (**4,** SD11) (Figure 6, Figure S1).

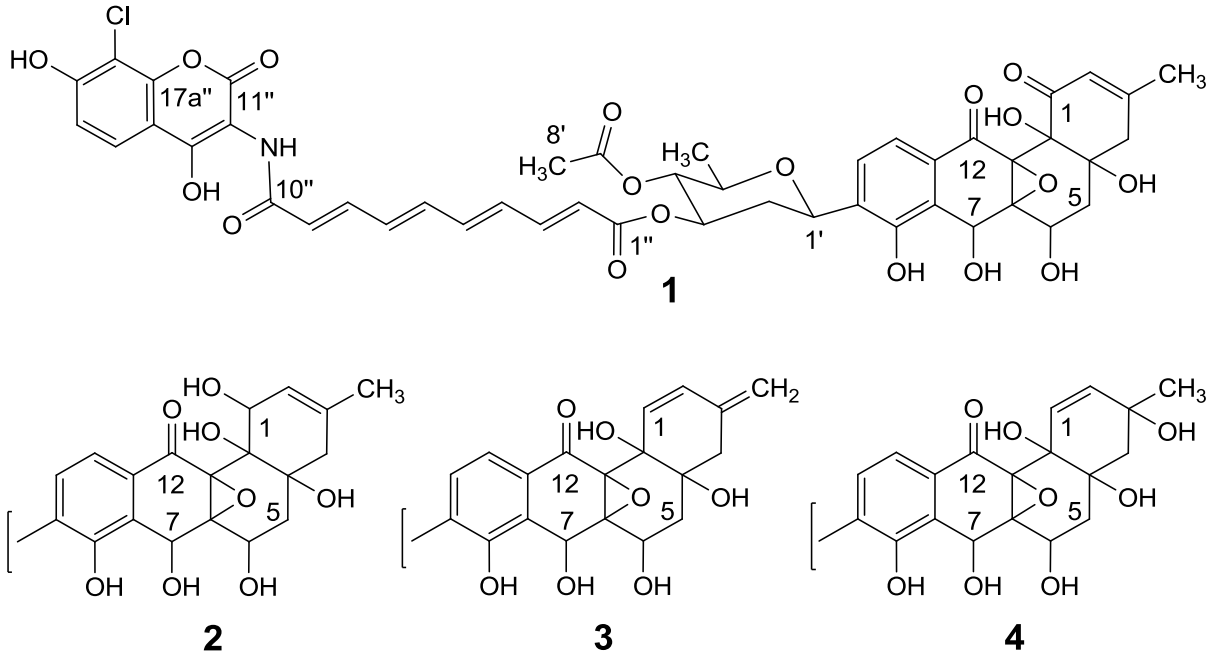


Figure 6. Structures of SD8 (**1**), SD9 (**2**), SD10 (**3**) and SD11 (**4**)

The structure elucidation of the detected simocyclinones shows the minor differences in the A-ring of the angucycline sub-structure between the previously described SD8 (**1**) and newly isolated molecules. The arrangement of the other sub-structures (aminocoumarin ring, tetraene part, deoxysugar and acetyl moiety) in these hybrid molecules are identical (Chapter S1, Figures S1-S5c, Tables S1a-S3b)

Similarly to SD8 newly identified simocyclinones SD9 (**2**) and SD11(**4)** are able to inhibit the process of DNA gyrase supercoiling with 4 (**2,** determined IC50 value is 0.41 ± 0.031 μM) and 50 (**4**, determined IC50 value is 4.59 ± 0.5 μM) times lower activities than SD8 (**1**, determined IC50 value is 0.09 ± 0.018 μM) (Figure 7).

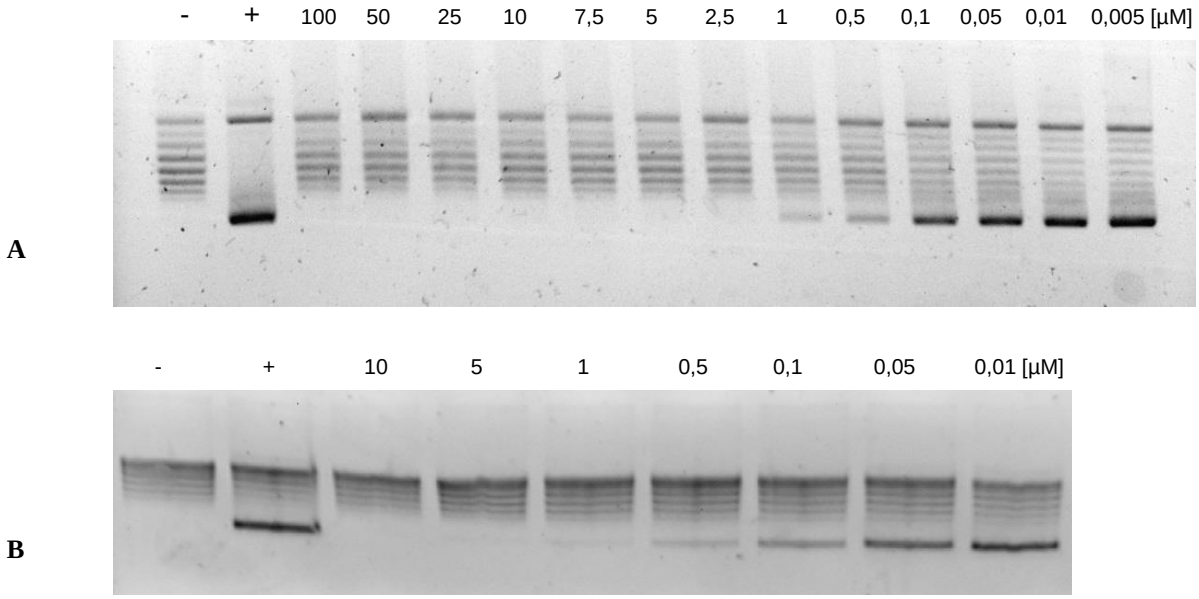


Figure 7. Supercoiling assay with E. coli DNA gyrase in the presence of (**A**) SD9 (**2**) or (**B**) SD8 (**1**). Controls: (-) without enzyme; (+) untreated

Comparison of the simocyclinone D9 (*smc* and *sml*) and simocyclinone D8 (*sim*) biosynthetic gene clusters

Given the high similarity of the chemical pattern of SD8 (**1**) and SD9(**2**)/SD10(**3**)/SD11 (**4**), the corresponding gene clusters would be expected to be nearly identical, with few deviations.

The identified *smc* (in *K.* sp.) and *sml* (in *S.* sp. NRRL B-24484) gene clusters are around of 47,000 bp long and comprise 43 open reading frames (ORF) (Figure 8; Table 2). To simplify the description process we will refer mostly to the experimentally characterized *smc* biosynthetic gene cluster. Despite the high similarity between the *smc* and the *sim* genes, the overall architecture of both gene clusters has the significant differences. The most remarkably, three genes from the *sim* gene cluster encoding polyketide synthase type I (*simC1A*-*simC1C*) are missing from the *smc* gene cluster. Moreover, their homologs were also not found within the genome outside of the *smc* gene cluster, which again opened the question about the biosynthetic mechanism of the tetraene chain formation.

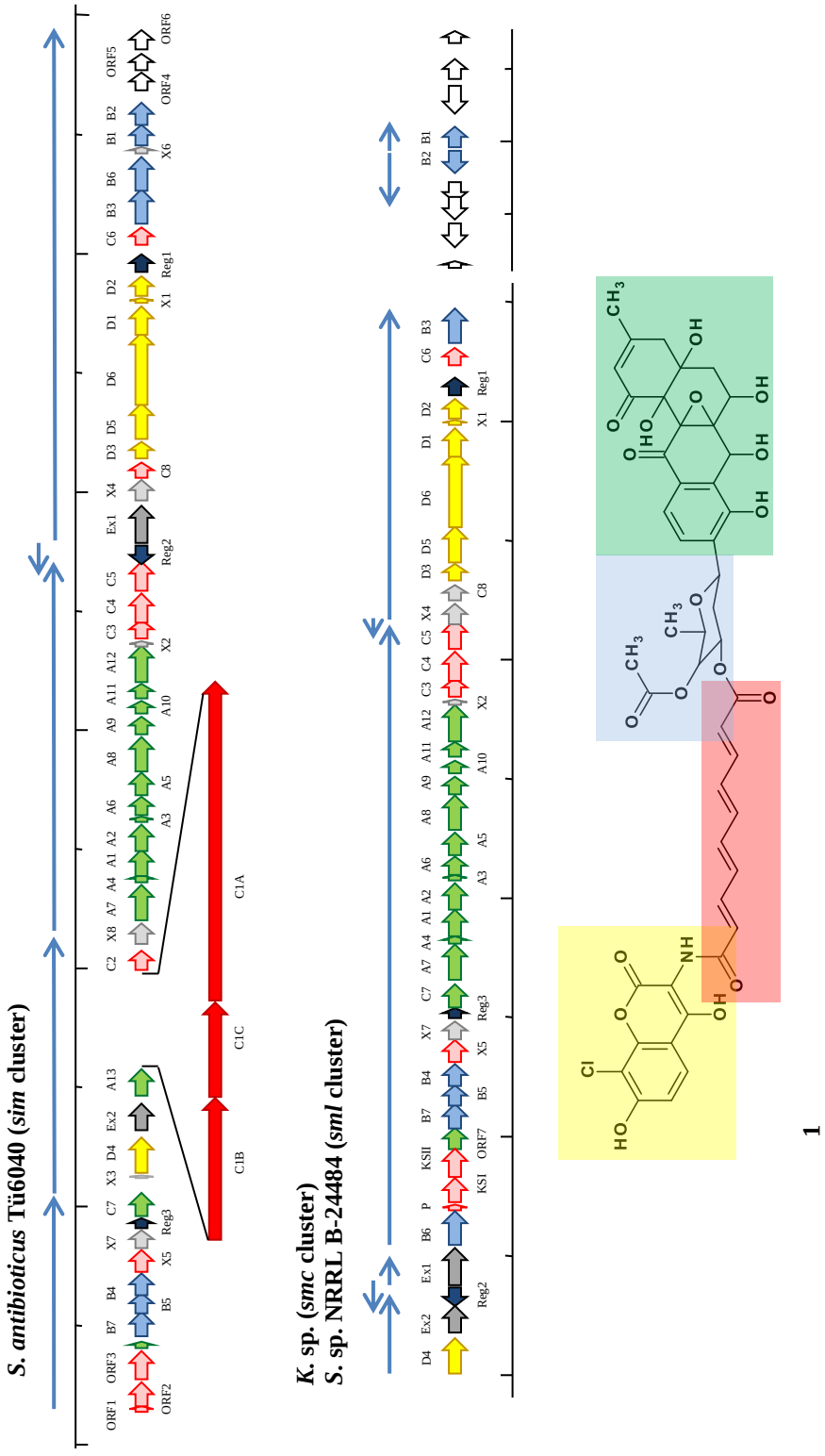


Figure 8. Simocyclinone biosynthetic gene clusters from *S. antibioticus* Tü6040 (*sim*), *K.* sp. (*smc*) and *S.* sp. NRRL B-24484 (*sml*). In yellow are labeled genes involved in aminocoumarin biosynthesis, blue – genes involved in deoxysugar formation, green – genes

responsible for the angucyclic core synthesis, pink – genes putatively involved in modification of the polyketide chain precursor, red – type I polyketide synthase genes, dark grey – regulatory genes, dark blue – efflux, pale grey – genes of unknown function, white – genes previously annotated as flanking in the *sim* cluster. Blue arrows indicate operons.

The second remarkable difference between the *sim* and the *smc* gene clusters is found within the operons responsible for dTDP-D-olivose production. In *S. antibioticus* Tü6040, seven genes (*simB1-simB7*) involved in the biosynthesis of activated D-olivose have been identified. In contrast, two genes (*smcB1* and *smcB2* encoding dTDP-glucose synthase and dTDP-glucose-4,6-dehydratase, respectively) are located approximately 900 kb downstream from the *smc* gene cluster.

Interestingly, three genes (*ORF123*) that were reported in the *sim* cluster as flanking and not involved in the biosynthesis of simocyclinone are located inside the *smc* cluster. A 1044 bp gene that shows high similarity to different ketoreductases is located downstream from *ORF3* in the *smc* cluster. Furthermore, closer look at *sim* cluster, revealed a 289-bp truncation at the end of *ORF3*.

The gene operon involved in angucycline biosynthesis is very similar to one within the *sim* gene cluster and consists of the following: 1) a minimal PKS gene set represented by *smcA1, smcA2* and *smcA3*, which encodes alpha and beta ketosynthases (KS_α and KS_β) and acyl carrier protein (ACP), respectively; 2) two cyclase encoding genes - *smcA4* and *smcA5*; 3) three ketoreductase encoding genes – *smcA6, smcA9 and smcA10*; 4) two oxygenase encoding genes – *smcA7* and *smcA8*; 5) one gene, *smcA11*, which is responsible for the activation of ACP with its phosphopantetheine arm; and 6) the gene encoding a decarboxylase, *smcA12*. We have not identified any homologues of the putative hydroxylase P450 encoding gene *simA13* within the *smc* gene cluster. As mentioned above, an additional gene, annotated as *ORF7*, putatively involved in the modification of the angucycline moiety, has been identified. Its truncated form is present in the genome of *S. antibioticus* Tü6040 and flanks the *sim* gene cluster. The full-length ORF for this gene from the *smc* gene cluster shows 73% homology to the Aln4 ketoreductase encoding gene from the alnumycin biosynthetic gene cluster (*13*). Aminocoumarin is another characteristic part of the simocyclinone molecules.

The genes from the *smc* gene cluster responsible for the biosynthesis of aminocoumarin are represented by *smcD1, smcD2, smcD3, smcD4, smcD5, smcD6* and *smcX1* and are nearly

identical to those from the *sim* gene cluster (*7, 8*). SD8 is halogenated at the C-8 position of the aminocoumarin ring, a reaction performed by the halogenase SimD4. The corresponding gene, *smcD4*, encoding the putative halogenase, has also been found within the *smc* gene cluster.

K. sp. and *S. antibioticus* Tü6040 simocyclinone clustrome description (transcriptome of a gene cluster)

To analyse the expression levels and the operon architecture of the *smc* gene cluster we performed RNA sequencing of *K.* sp. samples grown in two media: DNPM and NL-5.

All the genes from the *smc* cluster showed good expression (within the top 16% of all genes, when dividing mean DNPM/NL5 read counts by gene length). Transcription of the *smc* cluster appears to start approximately 192 bp upstream from the annotated coding sequence (CDS) for *smcD4* and comprises the following transcription units: 1) *smcD4-smcEx2*; 2*) smcB6-smcC5*; 3) *smcX4-smcB3* and 4) reverse-strand *smcReg2*. Two more genes, *smcB1* and *smcB2*, are not inside the cluster. *smcB2* represents an atypical transcript with a very short leader sequence and *smcB1* is a leaderless transcript (Figure 8).

The transcriptional start of *orf1* in *sim* cluster is located between 50 and 41 bp upstream of the coding region and appear to have the transcription read-through till *simC7*. The following operons are represented by *simX3-simC1A*, *simC2-simC5* and *simEx1-orf6*. The only gene transcribed from the reverse-strand is *simReg2* (Figure 8).

On average, *smc* cluster gene expression increased 8-fold in NL5 media compared to DNPM (Figure S10). Interestingly, these two media do not differ purely in the amount of synthesized simocyclinones. While *K.* sp. in NL5 does produce more simocyclinones by weight, only a few simocyclinone derivatives could be identified. On the other hand, in the DNPM medium, more type D simocyclinones (and also C precursors) are produced, though their total amount is lower than that in the NL5 medium.

The RNA sequencing and analysis provides the exact and clear expression profiles of both the *smc* and *sim* gene clusters. The elucidation of the number of operons and their level of expression will enable the rational design and engineering of both clusters. Biosynthetic engineering of complex gene clusters is significantly simplified when the complete information about the promoters, terminators, transcriptional and translational starts is available. By describing and characterizing clustromes of the *smc* and *sim* gene clusters, we

Table 2. Deduced functions of ORFs in the simocyclinone cluster

№	polypeptide	aa	Most similar protein (identity, %)	acc. number	predicted function
1	SmcD4	488	SimC6; *S. antibioticus* (91)	AEU17894.1	putative halogenase
2	SmcEx2	415	SimEx2; *S. antibioticus* (90)	AEU17895.1	putative Na/H antiporter
3	SmcReg2	246	SimReg2; *S. antibioticus* (82)	AF324838_17	TetR family transcriptional regulator
4	SmcEx	542	Sim17; *S. antibioticus* (93)	AAL15595.1	EmrB/QacA family drug resistance transporter
5	SmcB6	488	SimB6; *S. antibioticus* (61)	AAK06811.1	putative acetyl transferase
6	SmcP	83	ORF1; *S. antibioticus* (82)	AEU17883.1	putative acyl carrier protein
7	SmcKSI	360	ORF2; *S. antibioticus* (71)	AEU17884.1	putative 3-oxoacyl-ACP synthase I
8	SmcKSII	433	ORF3; *S. antibioticus* (89)	AEU17885.1	putative 3-oxoacyl-ACP synthase II
9	ORF7	348	ChaN; *S.chattanoogensis* (76)	AIU99206.1	aldo-keto reductase
10	SmcB7	387	SaqGT5; *M.* sp. Tu 6368 (58)	ACP19370.1	C-glycosyltransferase
11	SmcB5	253	SimB5; *S. antibioticus* (72)	AEU17887.1	putative 4-ketoreductase
12	SmcB4	324	SimB4; *S. antibioticus* (72)	AEU17888.1	putative 3-ketoreductase
13	SmcX5	318	SimX5; *S. antibioticus* (89)	AEU17889.1	putative 3-oxo-ACP synthase
14	SmcX7	258	SimX7; *S. antibioticus*] (55)	AEU17890.1	conserved hypothetical protein
15	SmcReg3	155	SimReg3; *S. antibioticus* (73)	AEU17891.1	putative regulator MarR/SlyA family
16	SmcC7	279	SimC7; *S. antibioticus* (72)	AEU17892.1	putative hydroxylase/dehydratase
17	SmcA7	510	SimA7; *S. antibioticus* (85)	AAK06782.1	putative oxygenase
18	SmcA4	109	SimA4; *S. antibioticus* (89)	AAK06783.1	putative cyclase
19	SmcA1	423	SimA1; *S. antibioticus* (90)	AAK06784.1	putative ketosynthase
20	SmcA2	405	SimA2; *S. antibioticus* (86)	AAK06785.1	putative ketosynthase
21	SmcA3	90	SchP6; *S.* sp. SCC 2136 (71)	CAH10115.1	putative acyl carrier protein
22	SmcA6	262	SimA6; *S. antibioticus* (87)	AAK06787.1	putative ketoreductase
23	SmcA5	316	SimA5; *S. antibioticus* (83)	AAK06788.1	putative cyclase/dehydrase
24	SmcA8	497	Sim7; *S. antibioticus* (76)	AAL15585.1	oxygenase
25	SmcA9	255	Sim7; *S. antibioticus* (92)	AAL15585.1	ketoreductase
26	SmcA10	191	SimA10; *S. antibioticus* (80)	AAK06791.1	putative reductase
27	SmcA11	222	Sim10; *S. antibioticus* (69)	AAL15588.1	Ppan-transferase
28	SmcA12	529	SimA12; *S. antibioticus* (91)	AAK06793.1	putative decarboxylase
29	SmcX2	72	Sim12; *S. antibioticus* (64)	AAL15590.1	hypothetical protein
30	SmcC3	258	SimC3; *S. antibioticus* (68)	AAK06795.1	thioesterase
31	SmcC4	457	Sim14; *S. antibioticus* (86)	AAL15592.1	dioxygenase
32	SmcC5	455	Sim15; *S. antibioticus* (80%)	AAL15593.1	putative alcohol dehydrogenase
33	SmcX4	318	Sim18; *S. antibioticus* (82)	AAL15596.1	predicted phosphohydrolase
34	SmcC8	226	Sim19; *S.antibioticus* (72)	AAL15597.1	putative Ppan-transferase
35	SmcD3	236	SimK; *S.antibioticus* (61)	AAG34182.1	reductase
36	SmcD5	523	SimL; *S. antibioticus* (76)	AAG34183.1	aminocumarin ligase
37	SmcD6	101	SimD6; *S. antibioticus* (68)	AAK06804.1	putative peptide synthase
38	SmcD1	437	SimD6; *S. antibioticus* (68)	AAK06804.1	cytochrome P450
39	SmcX1	71	SimY; *S. antibioticus* (70)	AAG34186.1	putative conserved protein MbtH
40	SmcD2	274	SimJ1; *S. antibioticus* (82)	AAG34187.1	3-ketoacyl-(ACP)-reductase
41	SmcReg1	251	SimReg1; *S. antibioticus* (65)	AAK06808.1	OmpR/PhoB subfamily response regulator
42	SmcC6	247	SimJ2; *S. antibioticus* (79)	AAG34189.1	3-ketoacyl-(ACP)-reductase
43	SmcB3	483	Sim20; *S. antibioticus* (82)	AAL15606.1	dNDP-4-keto-6-deoxy-glucose-2,3-dehydratase
44	SmcB1	292	RmlA; *K. kifunensis* (89)	ADO32599.1	glucose-1-phosphate dTDP-transferase
45	SmcB2	326	RfbB; *K. kifunensis* (92)	ADO32598.1	dTDP-D-glucose-4,6-dehydratase

also aim to initiate a new standard in biosynthetic gene clusters characterization. Availability of a clustrome description provides an excellent starting point for further engineering of biosynthetic gene clusters.

Functional analysis of the genes/enzymes putatively involved in tetraene side chain biosynthesis

We decided to focus on the surprising absence of the *simC1A, simC1B* and *simC1C* genes, which have been proposed to be involved in tetraene chain formation, from the *K.* sp and *S.* sp. NRRL B-24484 genomes. We screened the *K.* sp. genome for the presence of the other PKSI encoding genes and identified a single putative type I PKS encoding gene. It includes one module, which contains an acyltransferase (AT), dehydratase (DH), enoyl reductase (ER), two ketoreductase (KR) domains and ACP, but lacks a β-ketosynthase (KS) domain. Directly upstream of this gene, and translationally coupled to type I PKS, a gene encoding a separate KS domain is located. Interestingly, the corresponding protein is approx. 200 amino acid residues longer than regular KS domains and 18 amino acids that are located in the C-terminus show high similarity to the AT domain. The third gene in this complex encodes FabB, which is similar to β-ketoacyl synthase. Together these genes may provide a polyketide structure, although the order of the domains do not correspond to the tetraene pattern: 1) the ER domain would yield the fully saturated polyketide; 2) the KS domain encoding fragment is located outside the gene; 3) the number of domains is not sufficient for formation of the tetraene chain, unless this synthase acts iteratively.

To determine whether this unusual type I PKS is involved in the biosynthesis of simocyclinones, we inactivated the corresponding gene via a single crossover. No change in the production of simocyclinones was observed in the ΔPKSI mutant, thus excluding its involvement in the biosynthesis of simocyclinones (Figure S6.). Obviously, the polyketide chain substructure of simocyclinones must be synthesized through the different route not involving type I PKS enzymes.

Intrigued by these result, we checked whether the SimC1A-SimC1C encoding genes are involved in the biosynthesis of the tetraene chain of simocyclinones in *S. antibioticus* Tü6040. In the *sim* gene cluster, the *simC1A, simC1B, simC1C* genes, were proposed to be responsible for tetraene chain formation (*7*). However, the structure and order of the domains in SimC1A, SimC1B and SimC1C do not match the pattern of the polyketide chain within the SD8 molecule (*7*). A *simC1B* mutant, in which KS domain encoding region was replaced by the hygromycin resistance cassette was generated (*S. antibioticus* ΔKS). To our surprise, inactivation of the KS domain of SimC1B ketosynthase did not influence SD8 (**1**) biosynthesis (Figures S8a, S8b), indicating that the expression of the SimC1B KS domain and likely the entire PKSI sub-cluster is not necessary for biosynthesis of SD8 (**1**) in *S.*

antibioticus Tü6040. Therefore, the hypothesis of tetraene biosynthesis that is based on PKSI involvement should be revised.

To identify genes responsible for biosynthesis of the tetraene chain, the *smc* cluster was screened for new candidates. Three ketosynthase encoding genes (*smcKSI*, *smcKSII* and *smcX5*), which might take part in a polyketide biosynthesis, have been detected after careful re-annotation of both, *sim* and *smc* gene clusters. Their orthologs in the *sim* gene cluster were marked as *ORF2*, *ORF3*, and *simX5* and suggested to not be involved in simocyclinone biosynthesis. However, the transcriptome of the *smc* and *sim* biosynthetic gene clusters clearly shows that the expression of the *smcKSI*, *smcKSII* and *smcX5* (*ORF2*, *ORF3* and *simX5*) genes is coordinated with the expression of other genes in both clusters.

smcKSI and *smcKSII* encode proteins with high similarities to 3-oxoacyl-ACP-synthase I (57-62%) and 3-oxoacyl-ACP-synthase II (79-82%), respectively. SmcKSII contains a typical Cys(191)-His(322)-His(357) amino acid motif in its active site, however, we could not identify this motif in SmcKSI. SmcX5 is highly similar to 3-oxoacyl-ACP-synthase III (KAS III, FabH), a key enzyme in fatty acid biosynthesis in plants and bacteria, which initiates the condensing process and uses acetyl-CoA as primer. Additionally, acyl carrier protein (ACP) homolog (*smcP*, *ORF1*) has been identified upstream from the *smcKSI*. Together, SmcP, SmcKSI, SmcKSII, and SmcX5 may form an enzyme complex that is necessary for octatetraene dicarboxylate biosynthesis. However, subsequent ketoreduction and dehydration reactions are required for the maturation of the tetraene chain.

To check the involvement of SmcKSII in tetraene chain formation an in-frame deletion mutant was generated. No D-type molecules of simocyclinone have been detected in the crude extract of the obtained mutant, although several ions exhibiting the mass, UV/Vis spectrum and retention time typical for B-type simocyclinones (angucycline with the attached sugar unit) were observed (Figure S7.).

SD9 (**2**) was easily detected after complementation of the Δ*smcKSII* mutant with the intact *smcKSII*. Albeit, expression of the mutated gene encoding SmcKSII with amino acid substitutions in the active center didn't restore simocyclinone biosynthesis, undoubtedly proving the importance of the presence of the catalytically active form of the SmcKSII protein.

Based on the collected data, a proposed model for tetraene chain biosynthesis was created (Figure 9). To generate the decatetraene side chain with three ketosynthases, the assembly

should most probably start with a four-carbon unit (e.g., crotonyl-CoA, acetoacetyl-CoA) condensed to malonyl CoA with the help of SmcX5, the protein similar to the type KSIII. A number of ketosynthases employ non-acetate starter units for the polyketide biosynthesis (*14, 15*). Very often, incorporation of an alternative to acetate primer units involves KSIII and ACP components as described for frenolicin, daunorubicin and others (*15-17*).

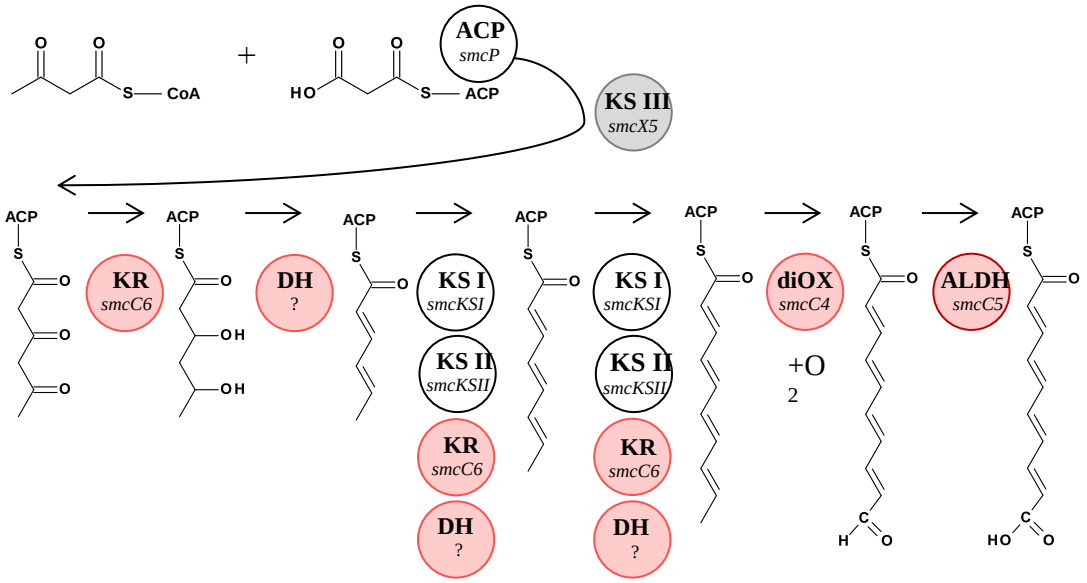


Figure 9. Proposed model for tetraene chain biosynthesis. In circles are marked genes and the function of their products that are involved in a particular biosynthetic step.

SmcX5 contains cysteine, proline and histidine in its active center and thus may be a candidate for the initiation of tetraene chain formation. Two other ketosynthases SmcKSI and SmcKSII, together with ACP, should be involved in the extension of the linear polyketide. The SmcKSI/KSII pair may act as a heterodimer catalyzing a polyketide chain extension similarly to the KS/CLF type II PKS system. Several more reactions are necessary to form the fully maturated dicarboxylic tetraene acids – ketoreduction, dehydratation and incorporation of the second carboxy group. Most of the gene encoding proteins with the necessary enzymatic activities can be found within both *sim* and *smc* gene clusters. SimC6/SmcC6 is homologous to 3-oxoacyl-ACP-reductases and a higher identity with FabG4 from *S. avermitilis MA-4680*. The PKS KR domains also belong to the same family. Most likely, this enzyme is involved in the subsequent ketoreduction of the polyketide chain. We could not identify the gene candidate responsible for the dehydration reaction of the hydroxylated polyketide intermediate, therefore the involvement of dehydratase from fatty acid biosynthesis cannot be excluded. Finally, the second carboxy group could be incorporated by two

enzymes, SimC4/SmcC4 and SimC5/SmcC5, to the tetraene chain. The deduced amino acid sequence of the *smcC4* encoding protein is similar to the carotenoid dioxygenases that catalyze the oxidative cleavage of carotenoids at various chain positions, thus leading to the formation of an aldehyde (*18*). The oxygen at the C-10 position of the tetraene indeed derives from molecular oxygen, as previously shown by feeding experiments (*19*). Newly generated aldehyde is oxidized to the carboxylic acid by SimC5/SmcC5 (aldehyde dehydrogenase) and later is attached to the sugar. The inactivation of both genes, *simC4* and *simC5*, in *S. antibioticus* Tü6040 led to the complete cessation of SD8 production, with the corresponding accumulation of the B-type simocyclinones (angucycline with the deoxysugar attached) (*20*). Although the exact mechanism of the decatetraene dicarboxylic acid formation remains unclear at the moment and requires further deciphering, the scheme and enzymes involved seems to be plausible.

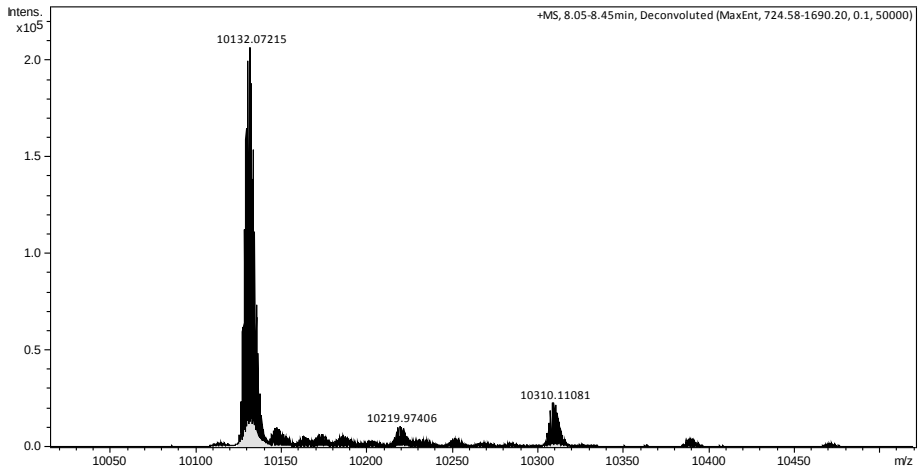
Heterologous expression of *smcP, smcKSI and smcKSII* in *S.albus*

Deduced amino acid sequences of SmcP, SmcKSI and SmcKSII are highly similar to ACP and two 3-oxoacyl-ACP-synthases, respectively. Together with KSIII homolog SmcX5 they may be involved in tetraene chain formation. To check their function *smcP*, *smcKSI*, *smcKSII* (pSmcTet) and *smcX5* (pSmcX5) have been expressed in *S. albus* del10. Additionally, *smcC8* (encoding putative Ppan-transferase) or *smcC8* + *smcC3* (encoding thioesterase) were introduced into *S. albus* del10 carrying pSmcTet and pSmcX5. However, we couldn't detect any precursors of octatetraene dicarboxylate biosynthesis. It may be because of (a) absence of the appropriate starter unit utilized by SmcX5 in the CoA-pool of *S. albus*; (b) difference in genes expression from different plasmids.

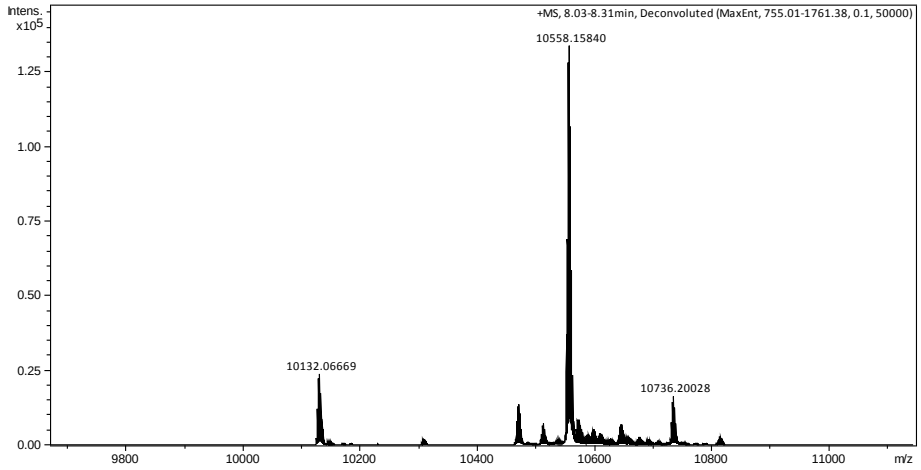
In vitro assay with SmcX5 and SmcP

To check if SmcX5 can perform first step of condensation between acetoacetyl-CoA and malonyl-ACP, appropriate proteins (KSIII – SmcX5 and SmcP - ACP) have been overexpressed in *E. coli*. In vitro assay was performed as following: 2.5µM ACP was incubated with 200µM Malonyl-CoA or Acetyl-CoA or Acetoacetyl-CoA in presence of Sfp-synthase (New England Biolabs), 75mM HEPES and 10mM $MgCl_2$ at 37 ^oC. After 1 hour of incubation 200µM substrate (malonyl-CoA, acetyl-CoA or acetoacetyl-CoA) and 0.1µM KSIII were added and incubated at 28 ^oC for another hour and analyzed with MS. Ions that correspond to apo-ACP (10.132Da) and holo-ACP with loaded malonyl (10.558Da), acetyl (10.514Da) or acetoacetyl (10.556Da) were detected (Figure 10). However, no ions that may indicate elongation step performed by SmcX5 have been identified. One of the reasons why

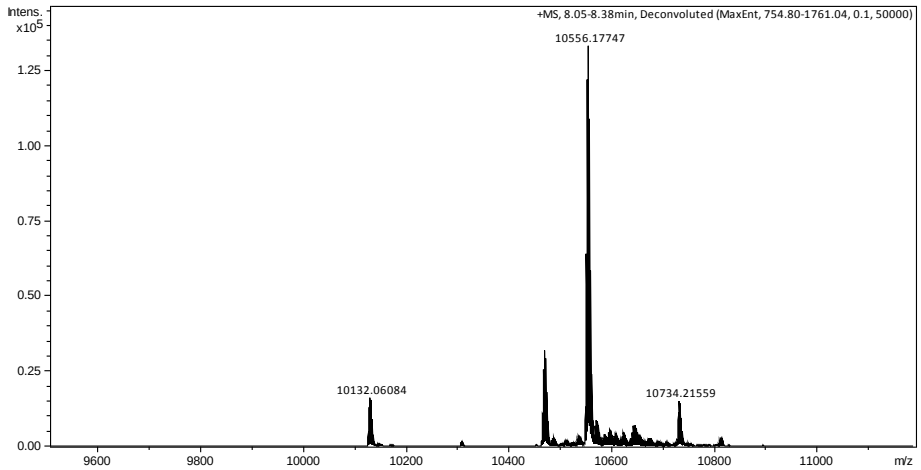
there was no reaction made by ketosynthase may be the low concentration of the protein. To solve this problem SmcX5 was concentrated ten times using Amicon microcon filter columns 10k and used for another in vitro assay. However, no expected product was detected, even after change to phosphate buffer (0.7M K_2HPO_4, 0.3M KH_2PO_4) and adding 1mM TCEP.



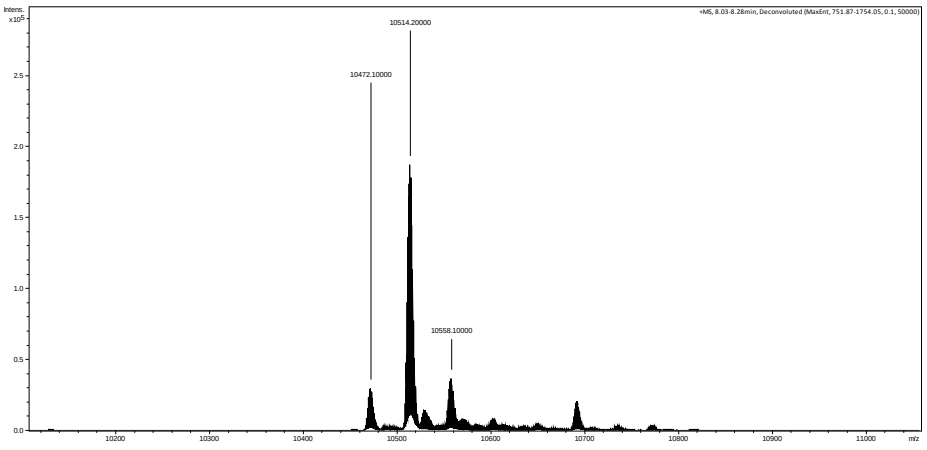
A



B



C



D

Figure 10. A – ACP (SmcP), B – ACP loaded with malonyl-CoA, C – ACP loaded with acetoacetyl-CoA, D – ACP loaded with acetate-CoA.

Inactivation of *smcA7*, *smcA8* and *smcA9*

Proteins encoded by *smcA7*, *smcA8* and *smcA9* show high similarity to two oxygenases and oxygenase-reductase, respectively. Homologs of these genes from other species take part in oxidizing of the nascent angucyclic core prior to its' further decoration with sugars. SmcA7 is highly similar (70% identity) to UrdE from the urdamycin biosynthetic pathway, which performs C-12/C-12b hydroxylation of the angucycline. SmcA8 is shares 63% identity with ox domain of LndM, whereas SmcA9 is a putative 6-ketoreductase (70 % similarity with UrdM red domain). Three deletion mutants were generated via double-crossover – *K.* sp. Δ*smcA7h*, *K.* sp. Δ*smcA8h* and *K.* sp. Δ*smcA9h* (Figure 11).

The functions of oxygenases in the decoration of the angucycline moiety

The only difference between SD8 (**1**) and the newly identified SD9 (**2**) was in the substitution of the C1 position. In the SD8 (**1**) molecule the ketogroup is present, while in SD9 (**2**) it is reduced to hydroxyl, and in SD10 (**3**) it is dehydrated to form a double bond (Figure 1). In the *smc* gene cluster, we identified *ORF7* as showing a high similarity to ketoreductases. The orthologous gene within the *sim* gene cluster is present in a truncated non-functional form. To determine if the structural difference between SD8 (**1**) and SD9 (**2**) is due to the function of this gene, *ORF7* was heterologously expressed in *S. antibioticus* Tü6040. Unfortunately, we could not identify new derivatives of simocyclinones in the recombinant strain. For the next step, we replaced *ORF7* with the hygromycin-resistance gene. However, the biosynthesis of SD9 (**1**) was completely abolished in this mutant.

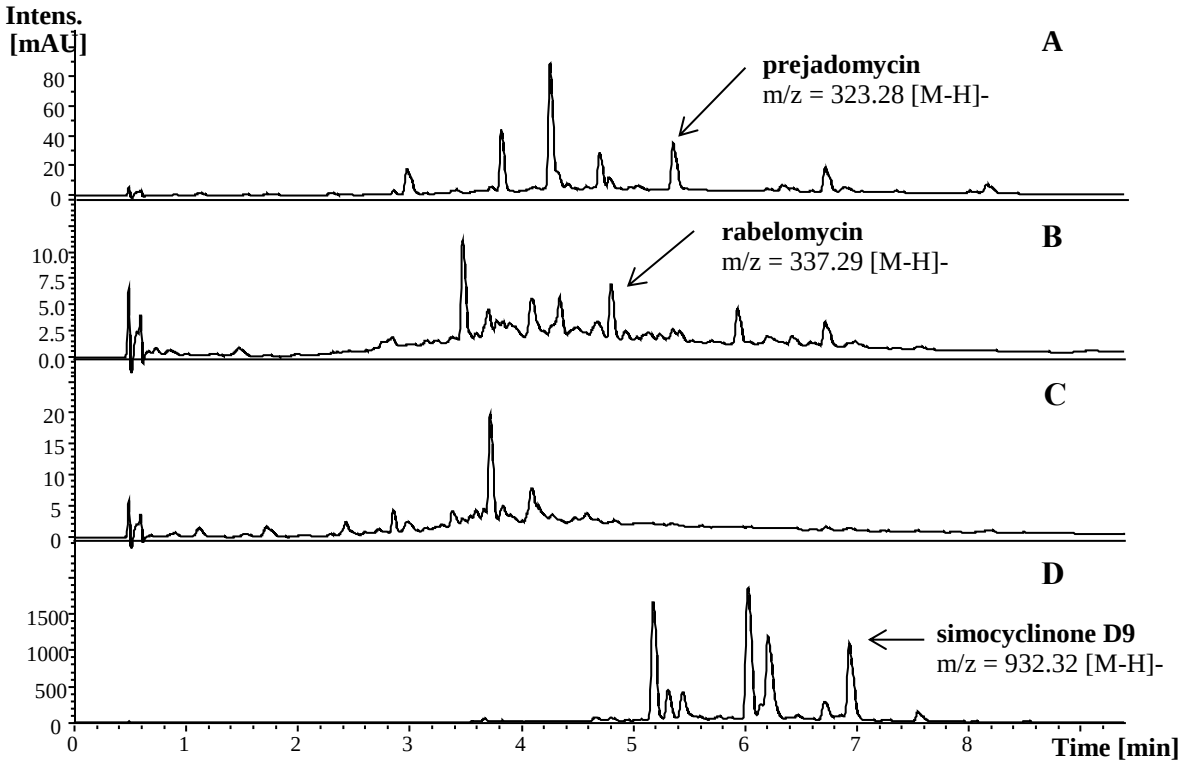


Figure 11. Chromatograms of the HPLC-ESI/MS analysis of *K.* sp. mutants: A - ΔsmcA7, B - ΔsmcA8, C - ΔsmcA8 and the wild type – D.

This demonstrated that the integration of the hygromycin resistance gene into the *smc* gene cluster disrupts the transcriptional/translational process of the operon and thus leads to the complete abolishment of the simocyclinone production. To avoid such an effect, we then generated a point mutation in *ORF7* while leaving the rest of the cluster intact (Figure S9). This mutant was named *K.* sp. *ORF7**. A characteristic feature of this mutant is the point mutations in the active center of the enzyme that result in a catalytically inactive protein. The *K.* sp. *ORF7** mutant produces D-type simocyclinones. At least seven new peaks were identified in the crude extract obtained for this mutant. Four of the peaks are presumably angucyclines, and three peaks exhibit UV/Vis spectrum that are typical for D-type simocyclinones and show molecular ions in ESI-MS m/z = 874.17 $[M+H]^+$, m/z = 914.33 $[M-H]^-$ and m/z = 930.39 $[M-H]^-$ (Figure 12). Based on the MS/MS analysis and the retention time, the molecule with mass 931 was assigned to SD8 (Figure 13), suggesting that ORF7 is indeed involved in reduction of the ketogroup to a hydroxyl group in the C1 position of the angucycline moiety during simocyclinone biosynthesis. In addition, the compound with the molecular mass M = 873 was semipurified. The amount and purity do not allow for the generation of a high-quality 1D NMR spectra, so the structure determination was conducted from this semipurified fraction based on 2D NMR signals, which led to the structure of

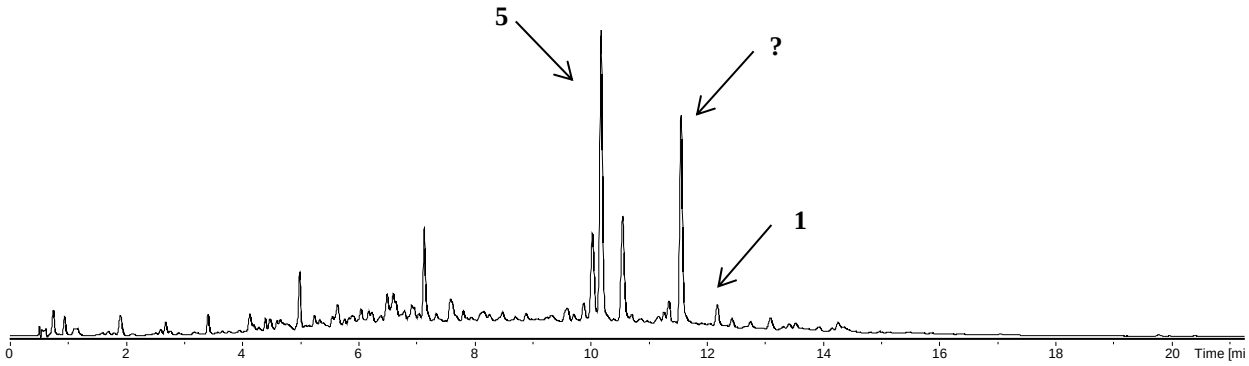


Figure 12. Chromatograms of the HPLC-ESI/MS analysis of *K.* sp. *ORF7** mutant crude extract. Arrows indicate peaks corresponding to simocyclinones. **5** – simocyclinone D5 (**5,** SD5), m/z = 874.17 $[M+H]^+$; **?** – D-type simocyclinone of unknown structure, m/z = 914.33 $[M-H]^-$; **1** – simocyclinone D8 (**1,** SD8), m/z = 930.25 $[M-H]^-$

simocyclinone D5 (**5**), which have been predicted before by Schimana et al. (*21*), but had not been detected experimentaly. SD5 (**5**) has the same angucycline as SD8 (**1**) with the intact carbonyl group at C-1 (Tables S4a, S4b; Figures S5a-S5c). The presence of the carbonyl group at C-1 in SD5 (**5**) undoubtedly demonstrated the function of ORF7 (Figure 14).

We speculate that the enzymes most likely involved in the simocyclinone biosynthesis form the protein complex, which is then disturbed when one of the structural units is deleted. When the mutated form of the enzyme is physically present within the complex, the metabolic intermediates are passed from one enzyme to another through the intramolecular tunnel without its release into the solution. Such substrate channeling prevents the release of unstable intermediates and also makes the process faster and more efficient (*22*, *23*). Therefore, deletions or substitutions of genes from biosynthetic gene clusters may lead to much lower production and accumulation of shunt products. Thus, by changing the gene inactivation strategy from very rough to more precise we were able to generate the first engineered D-type simocyclinone in the native producer. Our finding may partially explain the previous failures to generate D-type simocyclinones via biosynthetic engineering. In addition, the RNA sequencing and analysis provides the exact and clear expression profiles of both the *smc* and *sim* gene clusters. The elucidation of the number of operons and their level of expression will enable the rational design and engineering of both clusters. Biosynthetic engineering of complex gene clusters is significantly simplified when the complete information about the promoters, terminators, transcriptional and translational starts is available. By describing and

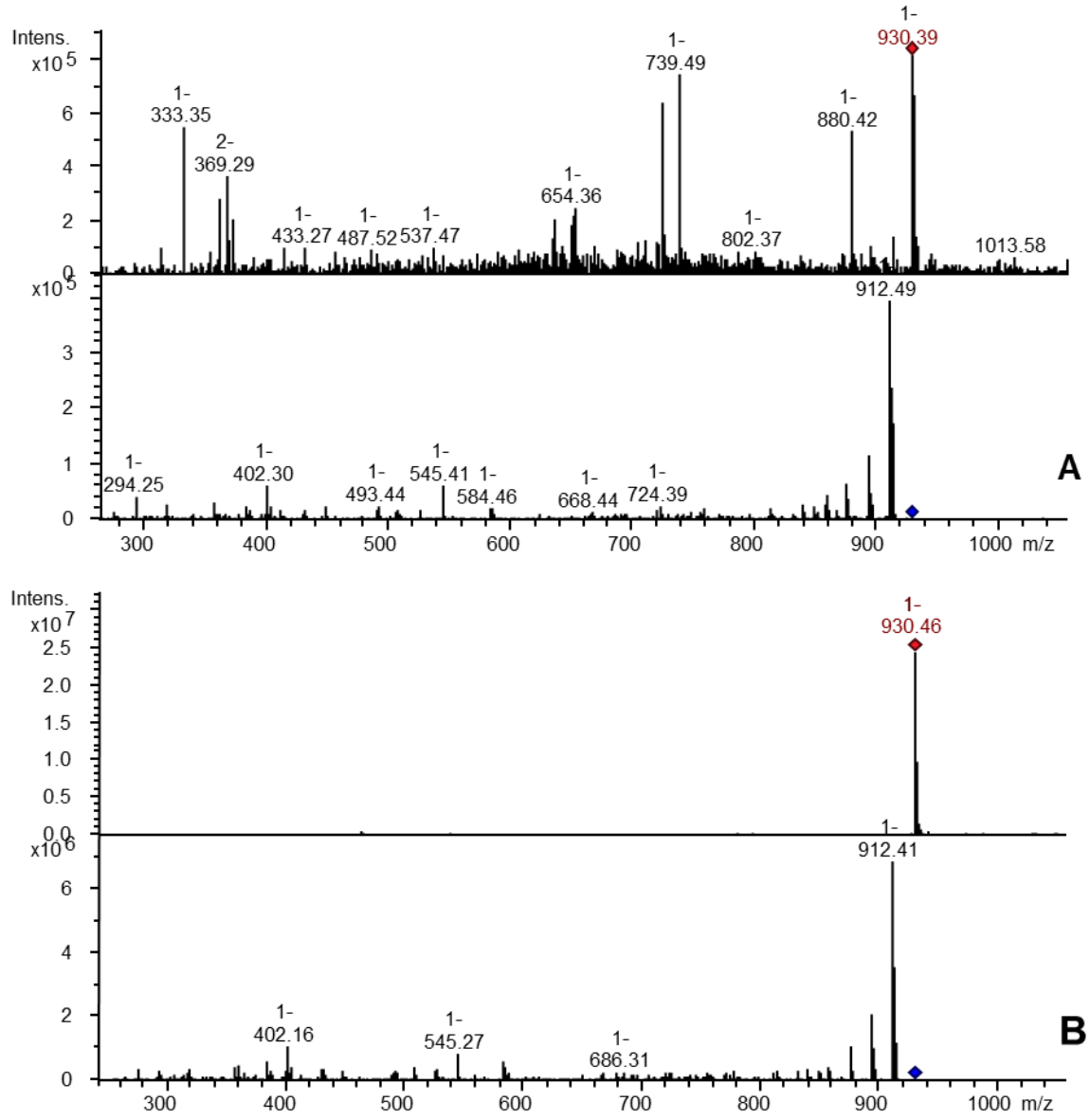
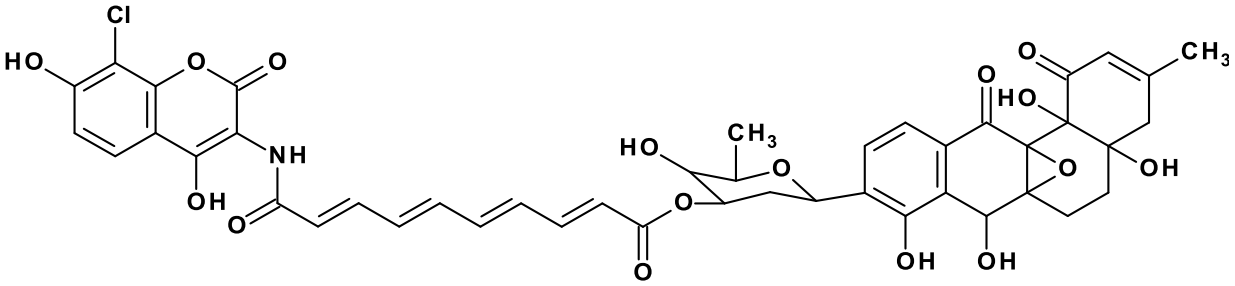


Figure 13. Comparison of the MS and MS/MS of D-type simocyclinone produced by *K.* sp. *ORF7**, m/z = 930.39 $[M-H]^-$ (**A**) and **1** (**B**).

characterizing clustromes of the *smc* and *sim* gene clusters, we also aim to initiate a new standard in biosynthetic gene clusters characterization. Availability of a clustrome description provides an excellent starting point for further engineering of biosynthetic gene clusters. The three identified simocyclinone gene clusters (*sim*, *smc* and *sml*) are typical examples of the "bricks and mortar" evolution events postulated by Marnix et al. (*24*). These hybrid gene clusters are composed of several sub-clusters (bricks) and individual genes (the "mortar").



5

Figure 14. Structure of simocyclinone D5 (**5,** SD5)

During the evolution of the *sim* gene cluster from the *smc* gene cluster, substantial gene reorganization took place (deletion of the ORF7 function, gaining the deoxysugar biosynthetic genes and reorganizing the gene order). These cluster rearrangements led to the generation of SD8 (**1**), which is four times more active than SD9 (**2**) *in vitro*, which could have brought some selective advantages to its producer. The acquisition of two additional deoxysugar biosynthetic genes might help in the coordination of the gene expression within the *sim* gene cluster. It is not clear which selective advantages to *S. antibioticus* Tü6040 are brought by PKS I encoding genes, which are expressed, but are not necessary for SD8 (**1**) biosynthesis. In summary, the comparative and functional analysis of the *smc, sml* and *sim* gene clusters led to the decifering of the simocyclinone biosynthetic route (Figure 15).

Conclusions

The identification of several new simocyclinones and the cloning and functional analysis of the corresponding gene clusters provided new insights into the biosynthesis and evolution of these hybrid bioactive molecules. In particular, a surprising and new biosynthetic route for the biosynthesis of the decatetraene dicarboxylic acid has been detected. Based on bioinformatics analysis and the experimental data, the involvement of several stand-alone ketosynthases, but not the PKSI system as suggested previously, in the biosynthesis of the tetraene chain has been proposed. The genetic basis of the structural differences of SD8 (**1**) and the newly identified SD9 (**2**), SD10 (**3**) and SD11 (**4**) has been found. A structure-activity relationship study of simocyclinones clearly demonstrates the importance of the A-ring modifications for the activity of the entire molecule. The understanding of the biosynthetic pathway and the genetic accessibility of this new producer opens up the important possibility of generating novel analogs of these unique molecules.

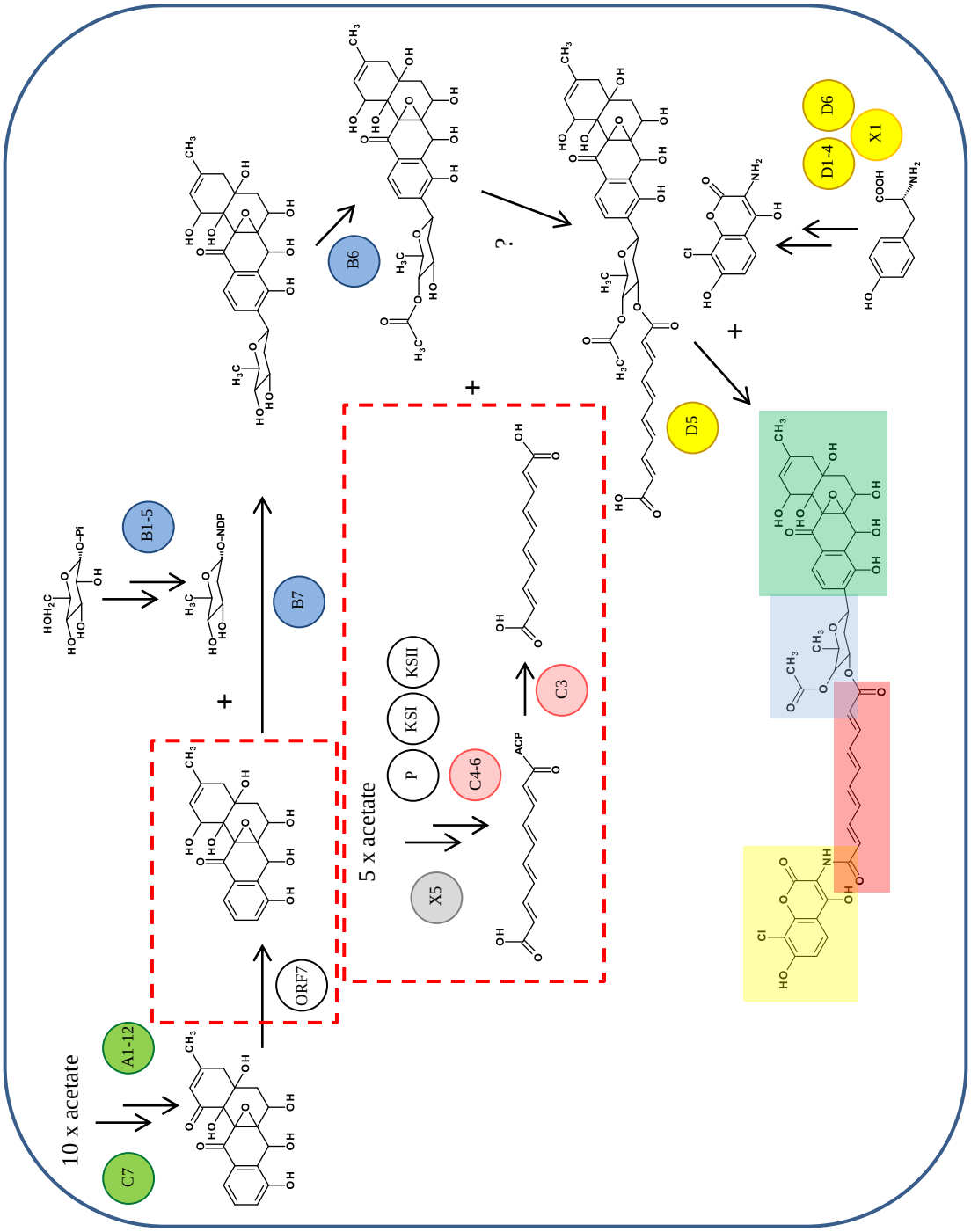


Figure 15. Proposed biosynthesis of simocyclinone D9 (**2,** SD9). A1-12, C7, ORF7, B1-5, B7, B6, X5, P, KSI, KSII, C4-6, C3, D1-4, D6, X1, D5 – correspond to the enzymes involved in a particular biosynthetic step. Biosynthetic steps highlighted in the dotted boxes have been investigated during the current study.

Methods

Genus affiliation

To determine whether the new simocyclinone producing strain belongs to the genus of *Streptomyces*, 16S rDNA was sequenced and analyzed. It shows the highest similarity to *Kitasatospora setae*. One of the phenotypic criteria to distinguish between *Streptomyces* and *Kitasatospora* is the chemical composition of cell wall peptidoglycan (*25*). Diaminopimelic acid (DAP) in *Streptomyces* is present only in the LL-isoform, in contrast to the peptidoglycan of *Kitasatospora* that contains both LL- (in spores) and *meso* (in mycelia) isomers of DAP. Analysis of DAP from our new strain has revealed the presence of both LL- and *meso* isomers in the cell wall peptidoglycan, which proves the affiliation of the new strain with the *Kitasatospora* genus.

Strains, growth conditions and media

For standard purposes *S. antibioticus* Tü6040 and *K.* sp. and *S. albus* wild types and all mutants were grown on 2% manitol and 2% soy bean meal, pH 7.5, prepared as solid medium and tryptone soy broth (TS broth), prepared as liquid medium, at 28 ^{0}C. For the maintenance of the mutants, hygromycin or apramycin was added to a final concentration of 50 μg/ml. For simocyclinone production a liquid medium NL-5 (NaCl, K_2HPO_4, $MgSO_4 \times 7H_2O$, glycerin and trace element solution, pH 7.2) and DNPM (4% dextrin, 0,75% soytone, 0,5% baking yeasts and 2,1% MOPS, pH 6.8) were used. DNA manipulation was carried out in *E. coli* XL-1 Blue MRF (Stratagene) and the non-methylating *E. coli* ET12567/pUZ8002 strain was used to drive conjugative transfer of non-methylated DNA to *Actinobacteria* as described previously(*26*). *E. coli* strains were grown on Luria-Bertani agar or liquid medium containing the appropriate antibiotic for selection. Vector pJET1.2 was from ThermoScientific (Darmstadt, Germany) and pKC1132, which carried the apramycin-resistance gene used for gene disruption, was from Eli Lilly and Company (Indianapolis, IN, USA). pKG1139, which contains *gusA* and apramycin-resistance genes and was used for gene disruption, was present in our laboratory (*27*).

Simocyclinone production analysis and NMR

For simocyclinone production 3 ml of 24-h old pre-culture was inoculated into 50 ml of NL-5 or DNPM media and grown for 4 days at 30°C with agitation at 200 rpms. The culture broth

was extracted with ethylacetate, samples were evaporated, dissolved in 500 µl of methanol and subjected to LC-MS analysis.

Standard measurements were performed on a Dionex Ultimate 3000 RSLC system using a BEH C18, 100 x 2.1 mm, 1.7 µm d_p column (Waters, Germany). Separation was achieved by a linear gradient from (A) H_2O + 0.1 % FA to (B) ACN + 0.1 % FA at a flow rate of 600 µL/min at 45 °C. The gradient was initiated by a 0.5 min isocratic step at 5 % B, followed by an increase to 95 % B in 18 min to end up with a 2 min step at 95 % B before re-equilibration under the initial conditions. UV spectra were recorded by a DAD in the range from 200 to 600 nm. MS data was acquired with an Amazon Speed 3D ion trap mass spectrometer (Bruker Daltonics, Germany) using the Apollo ESI source. Mass spectra were acquired in centroid mode ranging from 150 to 1500 m/z.

High-resolution measurements were performed on a Dionex Ultimate 3000 RSLC system using a BEH C18, 100 x 2.1 mm, 1.7 µm d_p column (Waters, Germany). Separation of a 1 µl sample was achieved by a linear gradient from (A) H_2O + 0.1 % FA to (B) ACN + 0.1 % FA at a flow rate of 600 µL/min and 45 °C. The gradient was initiated by a 0.5 min isocratic step at 5 % B, followed by an increase to 95 % B in 18 min to end up with a 2 min step at 95 % B before re-equilibration under the initial conditions. UV spectra were recorded by a DAD in the range from 200 to 600 nm. The LC flow was split to 75 µL/min before entering the maXis 4G hr-ToF mass spectrometer (Bruker Daltonics, Germany) using the Apollo ESI source. Mass spectra were acquired in centroid mode ranging from 150 – 2500 m/z at a 2 Hz scan rate.

NMR spectra were measured using a Varian VNMR-S 600 MHz spectrometer equipped with 3 mm triple resonance inverse and 3 mm dual broadband probes (^{1}H, ^{13}C, 2D NMR) or a Bruker Avance III equipped with a 5 mm TCI cryoprobe (^{1}H and 2D NMR). Solvent signals were used as internal standard (DMSO-d_6: δ_H = 2.50 ppm, δ_C = 39.5 ppm; CD_3OD: δ_H = 3.31 ppm, δ_C = 49.0 ppm, CD_3CN: δ_H = 1.93 ppm, δ_C = 1.28 ppm). NMR spectra in DMSO-d_6 were recorded atT = 35 °C and in CD_3OD or CD_3CN at T = 25 °C.

Purification of SD5

K. sp. *ORF7** was cultivated in DNPM medium in shaking flasks, total volume 6 L at 30°C, 200 rpm for 5 days. Mixture of culture broth and biomass was extracted with 6 L of ethylacetate (EtAc) and led to 615 mg of crude extract. The extract was purified on SPE cartridge (ACN 20%, 40%, 60%, 80%, 100%). The 60% ACN fraction (8 mg) contained the

compounds of interest and was further purified by HPLC (HTec C18 column, 250 × 10 mm, 5 µm, A = H_2O + 0.1% FA, B = ACN + 0.1% FA) with gradient starting from 5% B to 95% B in 28.2 min. Flow 5.8 ml/min, injection volume 40 µl of an MeOH extract (*c* = 5.3 mg/ml), detection wavelength λ = 420 nm. The isolate (0.2 mg) could be identified as new simocyclinone D 12. High resolution ESI mass spectra were measured using a Bruker Impact II mass spectrometer equipped with a Dionex RSLC UHPLC.

Purification of SD9 and SD10

K. sp. WT was cultivated in NL5 medium in shaking flasks with a total volume of 2 L at 30°C, 200 rpm for 5 days. The biomass was decomposed in MeOH and later extracted with EtAc, which yielded 160 mg of the crude extract. The crude extract later was purified by the Hypha company using preparative HPLC-MS. One fraction (N1) contained SD9 and was used for NMR measurement. To obtain sufficient SD10 for structure elucidation, fraction N3 was further purified by HPLC (conditions identical to SD11 purification), which yielded 0.7 mg of the compound that was used for further measurements.

General genetic manipulation and PCR

Standard molecular biology procedures were performed as described previously (*28*). Isolation of plasmid DNA from *E. coli* and DNA restriction/ligation were performed by following the protocols of the manufacturers of the kits, enzymes, and reagents, Qiagen, Promega, NEB and ThermoScientific. PCR reactions were performed by using Phusion High-Fidelity DNA polymerase (ThermoScientific) for complementation and expression experiments and DreamTaq polymerase (ThermoScientific) to verify mutants. Primers were purchased from Eurofins MWG Operon. The oligonucleotide primers that were used are listed in Table S5.

Genome sequencing

For the genome sequencing, genomic DNA was isolated from *K.* sp. as described before (*29*). Genome sequencing of the cultivated strain was performed using Illumina MiSEQ technology. High-molecular-mass DNA was extracted from the selected strain, and two libraries were created: an 8-kb mate-pair library, as well as a shorter-insert (~600bp) shotgun paired-end library. Reads of an approximate length 2x250 (before trimming) were obtained for both libraries. Newbler software version 2.8 was used for genome assembly.

RNA sequencing and analysis

RNA sequencing was performed by Vertis Biotechnologie AG. Briefly, ribosomal RNA depletion of isolated total RNA was performed with Ribo-Zero Bacteria kit from Epicentre. The rRNA depleted RNA was fragmented with RNase III and 5'PPP structures were removed using RNA 5' polyphosphatase (Epicentre). Afterwards, RNA fragments were poly(A)-tailed using poly(A) polymerase. First-strand cDNA synthesis was performed using an oligo(dT)-adapter primer and M-MLV reverse transcriptase. The resulting cDNA was amplified and purified, and fragments in the length range 250-350 bp were sequenced on an Illumina HiSeq 2000 machine with 100-bp read length. For DNPM, we had 8987253 reads, and for NL5 we had 7433408 reads (only mapped using Novoalign V3.00.04 non-rRNA reads were counted).

Differential expression analysis was performed with DeSeq package for the R statistical environment (*30*).

We used the EDASeq R package for gene GC content and length bias control in *K.* sp. samples (*31*). Neither the bias plots nor the log-fold-change plots revealed significant dependencies between GC/length and read counts, and therefore no EDASeq normalization was applied. The ArrayQualityMetrics (*32*) quality report has not revealed any problems.

DNA-sequencing and computer-assisted sequence analysis

Nucleotide sequences were determined at GATC Biotech AG (Cologne, Germany) by using either standard primers (pJET1.2 Forw. and pJET1.2 Rev.) or customized, internal primers. Computer-assisted sequence analysis was performed with the Geneious software (Geneious, version 7.1.7). Database comparison was performed with the BLAST search tools on the server of the National Center for Bio-technology Information, National Library of Medicine, NIH (http://www.ncbi.nlm.nih.gov/).

Intergenic conjugation between *E .coli* and *K.* sp., *S. antibioticus* Tü6040 or *S. albus* J1074

Plasmids were transferred to *K.* sp., *S. antibioticus* Tü6040 or *S. albus* J1074 by intergeneric conjugation with *E. coli.* Spores from one plate were recovered by centrifugation, washed twice in fresh TS broth, and resuspended in TS broth (recipient culture). The *E. coli* donor ET 12567 (pUZ8002) was grown at 37 °C for 16–18 h on LB agar plus apramycin (50 mg/ml) and kanamycin (30 mg/ml). These donor cells were detached from the plate with a loop and resuspended in the recipient culture. Samples of this combination (500 µl) were plated on MS

medium (20 g/L, d-mannitol, 20 g/L soy flour, 18 g/L agar, pH 7.2). Plates were incubated at 28 °C for 8–10 h and then covered with water (1 ml) that contained phosphomycin and apramycin at a final concentration of 200 mg/ml and 50 mg/ml, respectively, for selection of exconjugants. The incubation at 28 °C was continued for 3–5 days until exconjugants appeared.

Construction of gene inactivation and complementation constructs

For the generation of chromosomal mutants of *K. sp* and *S. antibioticus* Tü6040 by homologous recombination, the gene disruption plasmids were constructed as described below. To inactivate *smcKSII*, two flanking homologous sequences, 2.5 kb each, were appropriately inserted into the EcoRI-EcoRV and MunI-XbaI digest pKG1139 to yield plasmid pKG*KSII*FR. Primers were used to introduce *EcoRI*, *EcoRV* and *MunI*, *StuI*, *XbaI* restriction sites to both homologous fragments. A unique *StuI* restriction site was inserted into the forward primer of the second homologous shoulder. The *hph* gene was amplified by PCR, using pIJ10700 as a template, and cloned into ploxLERE plasmid, which was previously digested with BamHI/MfeI and blunted with Klenow Fragment. Primers were used to introduce *BamHI* and *MfeI* into the *hph* gene. Here, *hph*, flanked by two *loxP* sites, was cut out from ploxLERE*hph* after restriction enzyme treatment with *EcoRV* and ligated into *StuI* pKG*KSII*FR to generate plasmid **pKG*KSII*Δ**.

To inactivate PKSI-encoding gene, the internal fragment (3 kb) was inserted into the *HindIII–XbaI* sites of pKC1132 to yield inactivation vector **pKC*PKS*Δ**. To generate the cassette for *ORF7* inactivation, 3 kb-long DNA regions upstream and downstream from *ORF7* were amplified with PCR. DNA restriction sites were artificially introduced into homologous regions using primers. A downstream homologous fragment (ORF7r) was treated with EcoRV and EcoRI and ligated with the pKG1139 vector prior digested with EcoRI and EcoRV, to yield plasmid pKGORF7r. *MfeI*, *SspI* and *StuI* restriction sites were introduced into the 3' end of ORF7r. The upstream homologous region was treated with XbaI and StuI and ligated into pKG*ORF7*R digested with XbaI/StuI to generate pKG*ORF7*RF. *hph* flanked by two *loxP* sites was cut out from ploxLERE*hph* after restriction treatment with EcoRV and ligated into SspI treated pKG*ORF7*RF to generate plasmid **pKG*ORF7*Δ**. To generate the plasmid for site-directed mutagenesis, *ORF7** was cloned into pKG*ORF7*rf, to give pKG*ORF7**rf. Primers were used to introduce nucleotide mutations into *ORF7* and the *MfeI* restriction site into the 5' and *StuI* into 3' ends of the gene. Later, the *hph* gene with two loxP sites was ligated into pKG*ORF7**rf, which resulted in plasmid **pKG*ORF7***. To delete the KS-domain of SimC1B

in *S. antibioticus* Tü6040 **pKG*KS*Δ** was constructed. Two homology shoulders (2.5 kb each) were fused using overlapping PCR and cloned into pJET1.2 (pJET1.2*KS*FR). Primers were used to insert a unique *SspI* restriction site between the two shoulders and *StuI* restriction sites at the ends of the fused construct. As in the case of pKG*KSII*, the *hph* gene flanked by two *loxP* sites was cloned into SspI pJET1.2*KS*FR. The resulting inactivation construct (two homology sequences and hygromycin-resistance gene) was excised from pJET1.2*KS*FR using StuI and cloned into EcoRV-treated pKG1139. For the generation of the plasmid that was used for *ΔsmcKSII* strain mutant complementation, *smcKSII* was amplified by PCR using Phusion polymerase. Suitable restriction sites (*XbaI* and *EcoRV*) were artificially introduced upstream and downstream of the gene. In addition to the *XbaI* restriction site, the forward primer also includes a synthetic 21p promoter (*33*) and a ribosomal binding site. After digestion with XbaI and EcoRV, the fragment was ligated into the pKC1139 (*34*) plasmid. This yielded the complementation plasmid **pKC*KSII*com**. To generate the expression vector **pKG*KSII*****, *smcKSII* was amplified by PCR with two sets of primers that included point mutations. Suitable restriction sites (*XbaI* and *EcoRV*) were introduced upstream and downstream from the gene, using primers. The forward primer also contained the 21p promoter. The 1.1 kb PCR-fragment was digested with XbaI and EcoRV and ligated into a pKG1139 vector treated with EcoRV/XbaI, to generate **pKG*KSII***com**.

pKGsmcA7Δ. To inactivate *smcA7*, two flanking homologous sequences, 2.5 kb each, were appropriately inserted into the EcoRI-EcoRV and MunI-XbaI digest pKG1139 to yield plasmid pKG*A7*FR. Primers were used to introduce *EcoRI*, *EcoRV* and *MunI*, *SspI*, *XbaI* restriction sites to both homologous fragments. A unique *SspI* restriction site was inserted into the forward primer of the second homologous shoulder. The *hph* gene was amplified by PCR, using pIJ10700 as a template, and cloned into ploxLERE plasmid, which was previously digested with BamHI/MunI. Primers were used to introduce *BamHI* and *MunI* into the *hph* gene. Here, *hph,* flanked by two *loxP* sites, was cut out from ploxLERE*hph* after restriction enzyme treatment with *EcoRV* and ligated into *SspI* pKG*A7*FR to generate plasmid **pKGsmcA7Δ**.

pKGsmcA9Δ. To inactivate *smcA9*, pKGsmcA9Δ vector was constructed as pKGsmcA7Δ.

pKGsmcA8Δ. Left part of the *simA8* gene and its flanking region were amplified from the chromosome of *K.* sp. using primer pair simA8-11F and simA811R (Table S5). Obtained 2.3 kb DNA fragment digested with EcoRI and EcoRV, which cloning sites were incorporated into primers, was cloned into respective sites of the pKG1139 vector. This generated

pKG*A8*F. Right part of the *simA8* gene and region flanking it were amplified using simA8-21F and simA8-21R primers (Table S5) from the chromosome of *K*. sp. The amplified 2.2 kb DNA fragment was digested with EcoRV and HindIII, which recognition sites were introduced into primers, and cloned into respective sites of pKG*A8*F. As a result the pKG*A8*FR plasmid with the unique *EcoRV* site in the coding region of the *simA8* gene was obtained. The plasmid pKG*A8*FR was hydrolyzed with EcoRV and ligated with hygromycin resistance cassette *hyg* that was obtained as a PvuII fragment from pIJ10700 (Table S5). This yielded pKGsimA8Δ.

Plasmids used for the heterologous expression were constructed as described below

pSmcX5. To generate the expression vector pSmcX5, *smcX5* was amplified by PCR. Suitable restriction sites (*HindIII* and *XbaI*) were introduced upstream and downstream from the gene, using primers. Also forward primer was carrying 21p promoter. To obtain suitable shuttle vector, pALHimar was treated with HindIII/XbaI and 8.5kb-fragment was eluted from the agarose gel (done according to the manufacturer, Promega). 1 kb PCR-fragment was digested with HindIII and KpnI and ligated into pAL vector, to generate pSmcX5.

pSmcC8 and pSmcC8C3. Expression vector pSmcC8 was obtained by cloning 750-bp PCR fragment containing *smcC8*-coding region from the *K*. sp. chromosome into pTOS vector treated with HindIII/XbaI. Forward primer was used to introduce 21p promoter, RBS (ribosome binding site) and *XbaI* restriction site. Restriction site *HindIII* was introduced into the 3' end of the reverse primer. To generate pSmcC8C3, pSmcC8 was digested with BglII/XbaI and 1-kb PCR fragment that contains *smcC3* placed under the control of A9 promoter (introduced by forward primer together with RBS and *XbaI* restriction site) was cloned.

pSmcTet. To obtain suitable shuttle vector, pUWLCre was treated with KpnI/XbaI and 8kb-fragment was eluted from the agarose gel (done according to the manufacturer, Promega). 2.7 kb DNA fragment containing *smcP*, *smcKSI* and *smcKSII* was amplified in PCR reaction under the standard conditions. Forward primer was used to introduce 21 promoter and *KpnI* restriction site. *XbaI* restriction site was introduced in reverse primer. Obtained PCR-product was treated with XbaI and KpnI and subsequently ligated into pUWL vector digested with KpnI and XbaI to yield pSmcTet.

Generation of chromosomal mutant strains of *K.* sp. and *S. antibioticus* Tü6040

For the generation of all deletion mutants, single crossover mutants were screened for loss of vector-resistance as a consequence of a double crossover event. Deletions within the genes were confirmed by PCR.

Topoisomerase assays

Assays were performed as described before (*35*). Commercial kits from Inspiralis (Norwich, UK) were used to test anti-gyrase activities of compounds. For standard reactions, 0.5 µg of supercoiled plasmid incubated with 1 unit of enzyme in the presence of different concentrations of compounds at 37°C for 30 minutes. Reactions with no enzyme and a standard reaction in presence of 5% (v/v) DMSO served as controls. Reactions were quenched by the addition of DNA gel loading buffer with 10% SDS (w/v). The samples were separated on 0.7% (w/v) agarose gel and DNA was visualized using ethidium bromide. All reactions were repeated at least three times.

References

1. Theobald, U., Schimana, J., and Fiedler, H. P. (2000) Microbial growth and production kinetics of Streptomyces antibioticus Tu 6040, *Antonie Van Leeuwenhoek 78*, 307-313.
2. Schimana, J., Fiedler, H. P., Groth, I., Sussmuth, R., Beil, W., Walker, M., and Zeeck, A. (2000) Simocyclinones, novel cytostatic angucyclinone antibiotics produced by Streptomyces antibioticus Tu 6040. I. Taxonomy, fermentation, isolation and biological activities, *The Journal of antibiotics 53*, 779-787.
3. Richter, S. N., Frasson, I., Palumbo, M., Sissi, C., and Palu, G. (2010) Simocyclinone D8 turns on against Gram-negative bacteria in a clinical setting, *Bioorganic & medicinal chemistry letters 20*, 1202-1204.
4. Edwards, M. J., Flatman, R. H., Mitchenall, L. A., Stevenson, C. E., Le, T. B., Clarke, T. A., McKay, A. R., Fiedler, H. P., Buttner, M. J., Lawson, D. M., and Maxwell, A. (2009) A crystal structure of the bifunctional antibiotic simocyclinone D8, bound to DNA gyrase, *Science 326*, 1415-1418.
5. Sissi, C., Vazquez, E., Chemello, A., Mitchenall, L. A., Maxwell, A., and Palumbo, M. (2010) Mapping simocyclinone D8 interaction with DNA gyrase: evidence for a new binding site on GyrB, *Antimicrob Agents Chemother 54*, 213-220.
6. Hearnshaw, S. J., Edwards, M. J., Stevenson, C. E., Lawson, D. M., and Maxwell, A. (2014) A new crystal structure of the bifunctional antibiotic simocyclinone D8 bound to DNA gyrase gives fresh insight into the mechanism of inhibition, *J Mol Biol 426*, 2023-2033.
7. Trefzer, A., Pelzer, S., Schimana, J., Stockert, S., Bihlmaier, C., Fiedler, H. P., Welzel, K., Vente, A., and Bechthold, A. (2002) Biosynthetic gene cluster of simocyclinone, a natural multihybrid antibiotic, *Antimicrob Agents Chemother 46*, 1174-1182.
8. Galm, U., Schimana, J., Fiedler, H. P., Schmidt, J., Li, S. M., and Heide, L. (2002) Cloning and analysis of the simocyclinone biosynthetic gene cluster of Streptomyces antibioticus Tu 6040, *Arch Microbiol 178*, 102-114.
9. Luft, T., Li, S. M., Scheible, H., Kammerer, B., and Heide, L. (2005) Overexpression, purification and characterization of SimL, an amide synthetase involved in simocyclinone biosynthesis, *Arch Microbiol 183*, 277-285.
10. Pacholec, M., Freel Meyers, C. L., Oberthur, M., Kahne, D., and Walsh, C. T. (2005) Characterization of the aminocoumarin ligase SimL from the simocyclinone pathway and tandem incubation with NovM,P,N from the novobiocin pathway, *Biochemistry 44*, 4949-4956.
11. Anderle, C., Hennig, S., Kammerer, B., Li, S. M., Wessjohann, L., Gust, B., and Heide, L. (2007) Improved mutasynthetic approaches for the production of modified aminocoumarin antibiotics, *Chemistry & biology 14*, 955-967.
12. Schafer, M., Le, T. B., Hearnshaw, S. J., Maxwell, A., Challis, G. L., Wilkinson, B., and Buttner, M. J. (2015) SimC7 Is a Novel NAD(P)H-Dependent Ketoreductase Essential for the Antibiotic Activity of the DNA Gyrase Inhibitor Simocyclinone, *J Mol Biol 427*, 2192-2204.
13. Oja, T., Palmu, K., Lehmussola, H., Lepparanta, O., Hannikainen, K., Niemi, J., Mantsala, P., and Metsa-Ketela, M. (2008) Characterization of the alnumycin gene cluster reveals unusual gene products for pyran ring formation and dioxan biosynthesis, *Chemistry & biology 15*, 1046-1057.

14. Hertweck, C., Luzhetskyy, A., Rebets, Y., and Bechthold, A. (2007) Type II polyketide synthases: gaining a deeper insight into enzymatic teamwork, *Nat Prod Rep 24*, 162-190.
15. Moore, B. S., and Hertweck, C. (2002) Biosynthesis and attachment of novel bacterial polyketide synthase starter units, *Nat Prod Rep 19*, 70-99.
16. Bibb, M. J., Sherman, D. H., Omura, S., and Hopwood, D. A. (1994) Cloning, sequencing and deduced functions of a cluster of Streptomyces genes probably encoding biosynthesis of the polyketide antibiotic frenolicin, *Gene 142*, 31-39.
17. Hutchinson, C. R. (1997) Biosynthetic Studies of Daunorubicin and Tetracenomycin C, *Chemical reviews 97*, 2525-2536.
18. Harrison, P. J., and Bugg, T. D. (2014) Enzymology of the carotenoid cleavage dioxygenases: reaction mechanisms, inhibition and biochemical roles, *Arch Biochem Biophys 544*, 105-111.
19. Holzenkampfer, M., and Zeeck, A. (2002) Biosynthesis of simocyclinone D8 in an 18O2-rich atmosphere, *The Journal of antibiotics 55*, 341-342.
20. Bihlmaier, C. (2005) Polyenantibiotika aus Streptomyceten : molekularbiologische Untersuchungen zur Biosynthese von Simocyclinon und alpha-Lipomycin.
21. Schimana, J., Walker, M., Zeeck, A., and Fiedler, P. (2001) Simocyclinones: diversity of metabolites is dependent on fermentation conditions, *J Ind Microbiol Biotechnol 27*, 144-148.
22. Huang, X., Holden, H. M., and Raushel, F. M. (2001) Channeling of substrates and intermediates in enzyme-catalyzed reactions, *Annu Rev Biochem 70*, 149-180.
23. Miles, E. W., Rhee, S., and Davies, D. R. (1999) The molecular basis of substrate channeling, *J Biol Chem 274*, 12193-12196.
24. Medema, M. H., Cimermancic, P., Sali, A., Takano, E., and Fischbach, M. A. (2014) A systematic computational analysis of biosynthetic gene cluster evolution: lessons for engineering biosynthesis, *PLoS Comput Biol 10*, e1004016.
25. Ichikawa, N., Oguchi, A., Ikeda, H., Ishikawa, J., Kitani, S., Watanabe, Y., Nakamura, S., Katano, Y., Kishi, E., Sasagawa, M., Ankai, A., Fukui, S., Hashimoto, Y., Kamata, S., Otoguro, M., Tanikawa, S., Nihira, T., Horinouchi, S., Ohnishi, Y., Hayakawa, M., Kuzuyama, T., Arisawa, A., Nomoto, F., Miura, H., Takahashi, Y., and Fujita, N. (2010) Genome sequence of Kitasatospora setae NBRC 14216T: an evolutionary snapshot of the family Streptomycetaceae, *DNA Res 17*, 393-406.
26. Luzhetskyy, A., Fedoryshyn, M., Gromyko, O., Ostash, B., Rebets, Y., Bechthold, A., and Fedorenko, V. (2006) IncP plasmids are most effective in mediating conjugation between Escherichia coli and streptomycetes, *Genetika 42*, 595-601.
27. Myronovskyi, M., Welle, E., Fedorenko, V., and Luzhetskyy, A. (2011) Beta-glucuronidase as a sensitive and versatile reporter in actinomycetes, *Appl Environ Microbiol 77*, 5370-5383.
28. Green, M. R., and Sambrook, J. (2012) *Molecular cloning : a laboratory manual*, 4th ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.
29. Pospiech, A., and Neumann, B. (1995) A versatile quick-prep of genomic DNA from gram-positive bacteria, *Trends Genet 11*, 217-218.
30. Anders, S., and Huber, W. (2010) Differential expression analysis for sequence count data, *Genome Biol 11*, R106.
31. Risso, D., Schwartz, K., Sherlock, G., and Dudoit, S. (2011) GC-content normalization for RNA-Seq data, *BMC Bioinformatics 12*, 480.

32. Kauffmann, A., Gentleman, R., and Huber, W. (2009) arrayQualityMetrics--a bioconductor package for quality assessment of microarray data, *Bioinformatics 25*, 415-416.
33. Siegl, T., Tokovenko, B., Myronovskyi, M., and Luzhetskyy, A. (2013) Design, construction and characterisation of a synthetic promoter library for fine-tuned gene expression in actinomycetes, *Metab Eng 19*, 98-106.
34. Kieser, T. B., M.J. Buttner, M.J. Chater, K.F. Hopwood, D.A. (2000) *Practical Streptomyces Genetics*.
35. Baumann, S., Herrmann, J., Raju, R., Steinmetz, H., Mohr, K. I., Huttel, S., Harmrolfs, K., Stadler, M., and Muller, R. (2014) Cystobactamids: myxobacterial topoisomerase inhibitors exhibiting potent antibacterial activity, *Angew Chem Int Ed Engl 53*, 14605-14609.

2. Amycomycins C and D, new angucyclines from *Kitasatospora* sp.

Abstract

Two new angucycline compounds containing O-glycosylated 6-deoxy-a-L-talose were isolated from *Kitasatospora* sp. The compounds were elucidated based on spectroscopic methods including UV, HRESIMS, and NMR. A feeding experiment using alizarin led to the conversion of alizarin to a monoglycosylated product with high efficiency, which allows isolation of the sugar after hydrolysis and determination of the absolute configuration of the sugar.

Introduction

While screening for new secondary metabolites two new angucyclines with an unusual O-glycosylation pattern were isolated from *K.* sp. and named amycomycin C (**6**) and D (**7**) due to their structural relationship to amycomycin B [1]. NMR studies revealed that the attached 6-deoxyhexose was the rare 6-deoxy-a-talose which is also a part of the lipopolysaccharides and polysaccharides found in the outer cell membrane of Gram-negative bacteria [2-4]. Only a few natural secondary metabolites from bacteria are known to contain 6-deoxy-a-talose as the substructure. These natural products are talosin A and B [5], phenazoviridin [6], and some acyl- and aryl-a-glycosides from *Streptomyces* sp. GöM1 [7]. Moreover to the best of our knowledge C-1-O-glycosylation has been described only for benzanthrin A and B as well as for amycomycin B in the literature [1, 8], while the glycosylation of angucyclines at position C-6 is reported only for the antibiotic TAN-1085 [9]. A feeding experiment using alizarin led to the O-glycosylation of alizarin (**8**), from which the sugar could be hydrolyzed, isolated, and determined as 6-deoxy-a-L-talose (Figure 16).

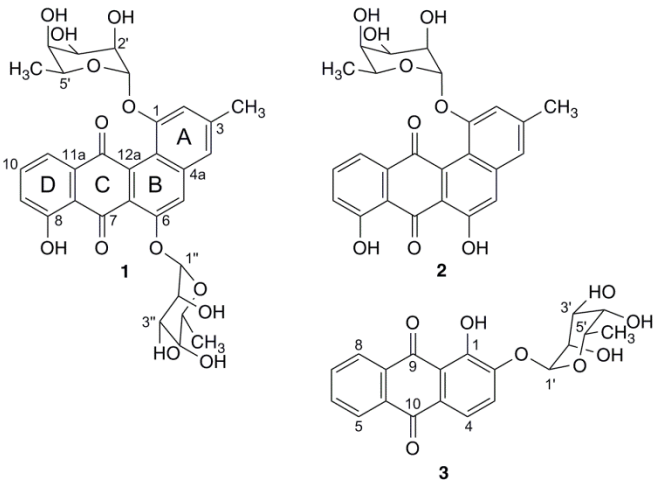


Figure 16. 6-Deoxy-α-L-talopyrranosides from *K.* sp.

Results and discussion

Amycomycins

HRESI(+)MS revealed m/z = 613.1918 [M+H]+ for **6** and m/z = 467.1335 [M+H]+ for **7**, which led to the molecular formula C31H32O13 (**6**) and C25H22O9 (**7**). The chemical shifts of the protons and carbons in the 1D NMR spectra for **6** and **7** show high similarity. Therefore the structure elucidation is discussed in detail for 1 and only differences for **7** are highlighted.

The ESI(+) mass spectrum of **6** showed two fragments m/z = 467.1 $[M–(6\text{-deoxyhexose})+H]^+$ and 321.1 $[M–(2*(6\text{-deoxyhexose}))+H]^+$ which correspond to the splitting of two sugar units. The NMR spectra of **6** show 3 methyl groups, 16 methin groups, and 12 quaternary carbons (Table 3). Moreover two sugars are visible and the aglycone was identified as an angucyclinone based on NMR data.

Table 3. NMR data for **6-7** (500 MHz, 25 °C, MeOD)

	1		2	
Pos.	δ_C	δ_H (J in Hz) [a]	δ_C [b]	δ_H (J in Hz) [a]
1	155.2		155.9	
2	113.6	7.26 d (1.1)	113.2	7.15 s
3	143.4		144.2	
3a	22.2	2.52 s	22.9	2.48 s
4	121.2	7.32 s (br)	121.2	7.22 s (br)
4a	140.4		142.9	
5	118.5	7.75 s (br)	119.0	7.42 s
6	153.6		158.6	
6a	125.7		139.6	
7	189.2		194.7	
7a	117.3		117.3	
8	162.8		163.8	
9	124.2	7.24 dd (8.4, 1.0)	125.0	7.27 d (8.3)
10	137.7	7.69 dd (8.4, 7.6)	139.6	7.75 dd (8.3, 7.6)
11	118.3	7.46 dd (7.6, 1.0)	119.8	7.50 d (7.6)
11a	137.0		138.9	
12	187.9		188.1	
12a	141.4		141.4	
12b	117.5		117.5	
1’	101.3	5.59 d (1.5)	101.8	5.57 s (br)
2’	71.8	3.86 ddd (3.1, 1.5, 1.5)	72.4	3.87 ddd (3.3, 1.5, 1.5)
3’	67.5	3.94 dd (3.2, 3.1)	68.2	3.94 dd (3.3, 3.3)
4’	74.2	3.71 ddd (3.2, 1.7, 1.5)	74.8	3.72 dd (3.3, 1.5)
5’	69.9	4.15 d (6.7)	70.5	4.16 q (6.6)
6’	17.0	1.28 d (6.7)	17.6	1.28 d (6.6)
1’’	100.8	5.81 d (1.5)		
2’’	71.9	4.12 ddd (3.2, 1.7, 1.5)		
3’’	67.2	4.34 dd (3.4, 3.2)		
4’’	74.3	3.75 ddd (3.4, 1.7, 1.8)		
5’’	70.0	4.06 q (6.7)		
6’’	16.9	1.21 d (6.7)		

[a] coupling constants for 1’-6’’are extracted from 1D-TOCSY (600 MHz, CD_3OD), [b] assignment based on HSQC and HMBC.

The angucycline ring A is substituted by a methyl group at C-3 and flanked by aromatic protons, as shown by the small coupling constant $J_{2\text{-H},4\text{-H}}$ = 1.1 Hz and the HMBC signals C-2/3a-H_3, C-4/3a-H_3, C-3a/2-H, and C-3a/4-H. The connectivity between ring A and B is indicated by the quaternary carbons δ_C = 117.5 (C-12b) and 140.4 ppm (C-4a) with HMBC signals C-4a/4-H and strong m-couplings C-12b/2-H, C-12b/4-H, and C-12b/5H which established the assignment of C-12b. Ring B showed connectivity with ring A through the HMBC signal C-5/4-H. The quinone structure of ring C was identified from the significant

carbonyl chemical shifts (δ_C = 187.9 and 189.2 ppm). The connection of ring C with ring B was represented by the HMBC signal C-6a/5-H. Ring D showed a phenolic hydroxyl group and an AMX spin system through the COSY cross peaks (9-H/10-H and 10-H/11-H) and coupling constants $J_{9\text{-H},10\text{-H}}$ = 8.5 Hz and $J_{10\text{-H},11\text{-H}}$ = 7.6 Hz. The NOE-signals (3´-H/11-H, 11-H/3´-H, and 10-H/2-H) between the aromatic proton of AMX spin system at δ_H = 7.46 ppm in ring D and the sugar unit attached at ring A defined the orientation of ring D with the OH group at C-8 (δ_C = 162.8 ppm). HMBC signals C-8/9-H and C-8/10-H supported the phenolic hydroxyl group at C-8. The HMBC signals C-7a/H-11, C-11a/H10, and C-12/11-H highlight the connection of ring C and D. One of the remaining quaternary carbons (δ_C = 141.4 ppm) was signed to be C-12a due to comparison with literature values [10, 11]. Thus, the aglycone structure was determined as 6-hydroxytetrangulol [12]. The angucycline core is di-O-glycosylated at C-1 and C-6 as shown by the HMBC signals C-1/1´-H and C-6/1´´-H in combination to the chemical shifts of C-1 (δ_C = 155.2 ppm) and C-6 (δ_C = 153.6 ppm). The 1H NMR signals at δ_H = 5.82 and 5.59 ppm corresponding to δ_C = 100.8 and δ_C = 101.3 ppm, respectively indicated two anomeric protons from O-glycosylated saccharides (Table 3). Their small coupling constants $J_{1´\text{-H},2´\text{-H}}$ and $J_{1´´\text{-H},2´´\text{-H}}$ = 1.5 Hz indicate α-sugars. This was supported by the carbon–proton coupling constant $^1J_{\text{C-10/10-H}}$ = 171.8 Hz [13]. 2D NMR data (COSY, HSQC, and HMBC) identify both saccharides as 6-deoxyhexoses (Table S3). All coupling constants in the sugar units (determined from 1D-TOCSY experiments) were smaller than $^3J_{H,H}$ = 6.7 Hz; thus no diaxial 3JH,H couplings were observed, so these sugars could not be either rhamnose or fucose [14, 15]. In addition sugar proton chemical shifts and *w*-coupling 2´-H/4´-H ($J_{2´\text{-H},4´\text{-H}}$ = 1.5 Hz) and 2´´-H/4´´-H ($J_{2´´\text{-H},4´´\text{-H}}$ = 1.7 Hz) indicated these sugars to be 6-deoxy-a-talose (Figure 17) [6]. This assumption was supported by the NOE signals between the 3´-H/5´-H which determined the chair conformation of the saccharides (Figure 17) [7]. A boat conformation could be excluded through the NOE signals between the 3´-H/5´-H and 3´-H/11-H.

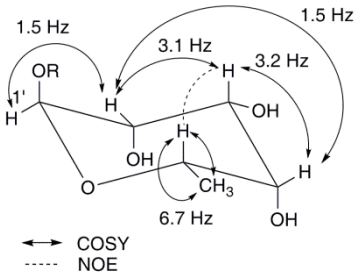


Figure 17: Selected coupling constants and NOE signal determine the chair conformation of the 6-deoxy-α-talopyranoside unit

NOE signals for 3´-H/11-H, 11-H/3´-H, and 1´-H/2-H indicated the orientation favored by saccharide A and the ROESY signal 5- H/1´´-H indicated the favored orientation of α-saccharide B (Figure 18).

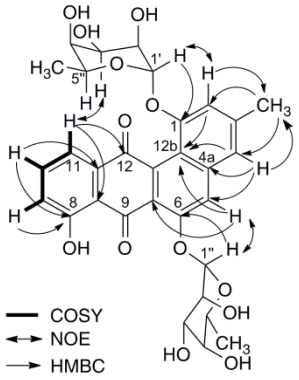


Figure 18. Selected COSY, HMBC, and NOE signals for **6**

HRESI(+)MS showed a m/z = 467.1330 $[M+H]^+$ for **7** leading to the molecular formula $C_{25}H_{22}O_9$. The molecular formula and the MS fragment m/z = 321.1 [M–(6-deoxyhexose)+H]$^+$ indicated only one sugar to be present in **7**. Comparison of NMR data of **6** and **2** led to the assumption, that the same aglycone and the same sugar are present in **6** and **7**. The HMBC signal C-1/10-H (δ_C = 155.9 ppm/δ_H = 5.57 ppm) and ROESY 2-H/1´-H signal (δ_H = 7.15 ppm/5.57 ppm) determined C-1-O-glycosylation of the angucycline. Moreover the deep-field shift of carbon C-6 in **7** (δ_C = 158.6 ppm) in comparison with **6** (δ_C = 153.6 ppm) indicates the free C-6-OH-group.

2-O-α-L-Talosyl-1,2-hydroxyanthraquinone

To prove the glycosylation potential of *K.* sp. a feeding experiment with alizarin was done which led to the new compound **8**. HRESI(+)MS of **8** shows a molecular ion m/z = 387.1085 for $[M+H]^+$ that fits to the assumed molecular formula $C_{20}H_{18}O_8$ for a monoglycosylated alizarin. The fragment m/z = 241.0514 [M–(6-deoxyhexose)+H]$^+$ indicates the splitting of the attached sugar.

NMR data of the aglycone are in good agreement with alizarin as the substructure (Table 4). O-glycosylation of alizarin at C-2 was proved by the HMBC signal C-2/1´-H and ROESY cross peak 3-H/1´-H. The aglycone free hydroxyl group must be in position C-1 showing a typical chemical shift of δH = 12.66 ppm. The NMR data determined that the 6-deoxysugar was 6-deoxy-α-talose like in **6** and **7**. Unfortunately the superposition of protons 2´-H, 3´-H,

and 5´-H in a multiplet at δH = 3.85 ppm, does not allow to determine sugar conformation based on proton–proton coupling or via NOESY/ROESY. Therefore, additional NMR measurement was performed in $CDCl_3$ with drops of CF_3COOD, which led to sufficient separation of 2´-H, 3´-H, and 5´-H (Table S6) allowing further NMR measurements (Table S6) and proving the sugar to be 6-deoxy-α-talose.

Table 4. NMR data for **3** (500 MHz, 25 °C, DMSO-d_6)

Pos.	δ_C	δ_H (*J* in Hz)	COSY[a]	HMBC[a]
1	152.4			3-H
2	149.8			1'-H, 3-H, 4'-H
3	121.8	7.63 d (8.5)	4-H	^1J
4	120.1	7.72 d (8.5)	3-H	3-H
4a	126.2			3-H,
5	126.8	8.18 m	6-H	^1J
6	135.2	7.93 m	5-H	^1J, 8-H
7	134.3	7.93 m	8-H	5-H, ^1J
8	126.6	8.23 m	7-H	^1J
8a	133.3			
9	188.6			8-H
9a	116.5			4'-H
10	180.9			5-H, 4-H
10a	132.9			6-H, 5-H, 8-H
1'	99.1	5.73 s (br)		(2'-H, 3'-H), ^1J, (2'-OH)
2'	70.0	3.85 m	2'-OH	3'-OH, 2'-OH, 1'-H, 4'-H,
3'	65.0	3.85 m	3'-OH (4'-H)	(2'-H), 4'-H, 4'-OH, 3'-OH, 2'-OH, 1'-H
4'	72.0	3.59 d (6.5)	4'-OH, (3'-H, 4'-H)	3'-H, 5'-H, 6'-H_3, 4'-OH, 3'-OH
5'	68.5	3.85 m	6'-H_3, (4'-H)	6'-H_3, 4'-OH, 1'-H, ^1J
6'	16.6	1.09 d (6.4)	5'-H	^1J, 5'-H, (4'-H)
1-OH		12.66 s (br)	/	
2'-OH		5.39 d (7.0)	2'-H	
3'-OH		5.12 d (6.0)	3'-H	
4'-OH		5.02 d (6.5)	4'-H	

[a] weak signals in brackets.

After hydrolysis of 3 in methanolic HCl, methyl 6-deoxy-a-talopyranoside (3.9 mg) was isolated. Comparison of the optical rotation value in H_2O $[\alpha]_D$ = – 87 (*c* 2 mg/ml, 20 °C) and MeOH $[\alpha]_D$ = – 101 (*c* 2 mg/ml, 20 °C) with the literature values confirmed L-configuration of the 6-deoxy-α-talose [6, 7]. From this experiment we could conclude that the sugars in **1** and **2** are 6-deoxy-α-L-talose, too.

Amycomycins C and D and monoglycosylated alizarin in comparison with alizarin were screened against a small panel of Gram-positive and Gram-negative bacteria (Table S2). Amycomycin C (**6**) displayed no remarkable activity, whereas the monoglycosylated amycomycin D (**7**) revealed a slight increase in activity against *Staphylococcus aureus* Newman, *Pichia anomala*, *Mucor hiemalis* (MIC = 9.21 µM), and *E. coli* ToIC (MIC = 14.6 µM). The monoglycosylated alizarin is slightly more active than its aglycone, to mention here is the activity against *E. coli* ToIC (MIC = 9.11 lM). Unusual is the fact that the genes

responsible for glycosylation could not be found within the gene cluster for angucycline biosynthesis.

In summary we isolated and characterized three new 6-deoxy-α-L-talopyranosides from *K*. sp.

References

1. Guo, Z.K., et al., *Angucyclines from an insect-derived actinobacterium Amycolatopsis sp. HCa1 and their cytotoxic activity.* Bioorg Med Chem Lett, 2012. **22**(24): p. 7490-3.
2. Muldoon, J., et al., *Structure of an acidic polysaccharide from the marine bacterium Pseudoalteromonas flavipulchra NCIMB 2033(T).* Carbohydr Res, 2003. **338**(5): p. 459-62.
3. Zahringer, U., et al., *Structure of a new 6-deoxy-alpha-D-talan from Burkholderia (Pseudomonas) plantarii strain DSM 6535, which is different from the O-chain of the lipopolysaccharide.* Carbohydr Res, 1997. **300**(2): p. 143-51.
4. Knirel, Y.A., et al., *Structure of the O-polysaccharide of Pseudomonas putida FERM P-18867.* Carbohydr Res, 2002. **337**(17): p. 1589-91.
5. Kim, W.G., et al., *Talosins A and B: new isoflavonol glycosides with potent antifungal activity from Kitasatospora kifunensis MJM341. II. Physicochemical properties and structure determination.* J Antibiot (Tokyo), 2006. **59**(10): p. 640-5.
6. Kato, S., et al., *Phenazoviridin, a novel free radical scavenger from Streptomyces sp. Taxonomy, fermentation, isolation, structure elucidation and biological properties.* J Antibiot (Tokyo), 1993. **46**(10): p. 1485-93.
7. Bitzer, J.Z.A., *6-Deoxy-α-L-talopyranosides from Streptomyces sp.* Eur. J. Org. Chem, 2006.
8. Rasmussen, R.R., et al., *Benzanthrins A and B, a new class of quinone antibiotics. II. Isolation, elucidation of structure and potential antitumor activity.* J Antibiot (Tokyo), 1986. **39**(11): p. 1515-26.
9. Ohmori, K., et al., *Concise total synthesis and structure assignment of TAN-1085.* Angew Chem Int Ed Engl, 2004. **43**(24): p. 3167-71.
10. Shaaban, K.A., et al., *Landomycins P-W, cytotoxic angucyclines from Streptomyces cyanogenus S-136.* J Nat Prod, 2011. **74**(1): p. 2-11.
11. Schneemann, I., et al., *Mayamycin, a cytotoxic polyketide from a Streptomyces strain isolated from the marine sponge Halichondria panicea.* J Nat Prod, 2010. **73**(7): p. 1309-12.
12. Yamashita, N., et al., *6-Hydroxytetrangulol, a new CPP32 protease inducer produced by Streptomyces sp.* J Antibiot (Tokyo), 1998. **51**(1): p. 79-81.
13. Shibuya, N., et al., *6-Deoxy-D-talan and 6-deoxy-L-talan. Novel serotype-specific polysaccharide antigens from Actinobacillus actinomycetemcomitans.* J Biol Chem, 1991. **266**(25): p. 16318-23.
14. Kuroda, M., Y. Mimaki, and Y. Sashida, *Saundersiosides C-H, rearranged cholestane glycosides from the bulbs of Ornithogalum saundersiae and their cytostatic activity on HL-60 cells.* Phytochemistry, 1999. **52**(3): p. 435-43.
15. Borbone, N.D.M., S. Iorizzi, M. Zollo, F. Debitus, C. Ianaro, A. and Pisano B., *New Glycosphingolipids from the Marine Sponge Aplysinella rhax and Their Potential as Nitric Oxide Release Inhibitors.* European Journal of Organic Chemistry, 2001.

3. Draft genome sequence of *Kitasatospora nimpheas*

Actinobacteria are of great importance for biotechnology as the source of diverse natural products, including anticancer agents, antibiotics and others [1]. *Kitasatospora* genus together with *Streptacidiphilus* and *Streptomyces* constitute *Streptomycetaceae* family within the phylum of *Actinobacteria* [2]. For a long time the status of genus *Kitasatospora* has been a matter of a debate [2-5]. Based on distinct clades of sequences of both 16S rRNA genes and16S-23S gene spacers Zhang et al. could strongly suggest that *Kitasatospora* and *Streptomyces* are separate genera [5]. Also, there are clear phenotypic differences between *Kitasatospora* and *Streptomyces*. Cell wall peptidoglycan of *Kitasatospora* contains both *LL*- and *meso*-diaminopimecil acid (DAP); in contrast, the peptidoglycan of streptomycetes contains only *LL* isomer of DAP. More recently, it was shown that *Kitasatospora* strains do not contain genes for the developmental proteins BldB and WhiJ, and for the actin-like cytoskeletal protein Mbl. Another feature of kitasatosporae is the loss of cell division activator SsgA and its transcriptional activator SsgR in some species [2]. Furthermore, tetracycline producer "*Streptomyces viridifaciens*" DSM 40239 merits species status within *Kitasatospora* taxon [2].

Eight genomes of *Kitasatospora* strains are available on the web site of NCBI (http://www.ncbi.nlm.nih.gov/genome) today [2, 6, 7]. Here we report draft genome sequence of new *Kitasatospora nimpheas*, which was found recently to produce new angucycline antibiotics amicomycins C and D [8].

Genomic DNA of the strain was isolated and sequenced using two Illumina MiSEQ libraries – 4 914 332 short-insert (paired-end, PE) and 6 576 880 long-insert (mate pair, MP) read pairs. PE insert size was 597.4 with a standard deviation of 179, MP insert size was 8 841.7 with a standard deviation of 2 210.4 (as reported by the Newbler v2.8 genome assembly software). A total of 843 contigs (8 398 185 bases) were assembled, including 612 large contigs (8 353 677 bases, N50 contig size 26 kbp). The final assembly resulted in 27 scaffolds (8 511 455 bases total), with most of the genome being in scaffold 1 (8 353 987 bases). The genome was then annotated using prokka [9]. The *K. nimpheas* genome has average GC content of 73.5%. The analysis of *K. nimpheas* genome revealed that its chromosome contains 7536 predicted protein coding sequences, and 98 predicted tRNA genes.

Using antiSMASH tool [10] we were able to estimate the potential of *K. nimpheas* to produce secondary metabolites. Preliminary data indicates the presence of 27 putative gene clusters for

the biosynthesis of different secondary metabolites on the main scaffold. Among predicted gene clusters, 6 encode terpenes, 2 – NRPS, 2 are involved in biosynthesis of siderophores, 2 are type II PKS (one of them have been identified to encode amicomycin biosynthetic machinery), 1 is type III PKS and 1 is a hybrid type I PKS-hlgks-teprene cluster.

References

1. *Antibiotics: Current Innovations and Future Trends*2015: Caister Academic Press. xii + 430 (plus colour plates).
2. Girard, G., et al., *Analysis of novel kitasatosporae reveals significant evolutionary changes in conserved developmental genes between Kitasatospora and Streptomyces.* Antonie Van Leeuwenhoek, 2014. **106**(2): p. 365-80.
3. Omura, S., et al., *Kitasatosporia, a new genus of the order Actinomycetales.* J Antibiot (Tokyo), 1982. **35**(8): p. 1013-9.
4. Wellington, E.M., et al., *Taxonomic status of Kitasatosporia, and proposed unification with Streptomyces on the basis of phenotypic and 16S rRNA analysis and emendation of Streptomyces Waksman and Henrici 1943, 339AL.* Int J Syst Bacteriol, 1992. **42**(1): p. 156-60.
5. Zhang, Z., Y. Wang, and J. Ruan, *A proposal to revive the genus Kitasatospora (Omura, Takahashi, Iwai, and Tanaka 1982).* Int J Syst Bacteriol, 1997. **47**(4): p. 1048-54.
6. Hwang, J.Y., et al., *Draft Genome Sequence of Kitasatospora cheerisanensis KCTC 2395, Which Produces Plecomacrolide against Phytopathogenic Fungi.* Genome Announc, 2014. **2**(3).
7. Ichikawa, N., et al., *Genome sequence of Kitasatospora setae NBRC 14216T: an evolutionary snapshot of the family Streptomycetaceae.* DNA Res, 2010. **17**(6): p. 393-406.
8. Brotz, E., et al., *Amycomycins C and D, new angucyclines from Kitasatospora sp.* Tetrahedron Letters, 2014. **55**(42): p. 5771-5773.
9. Seemann, T., *Prokka: rapid prokaryotic genome annotation.* Bioinformatics, 2014. **30**(14): p. 2068-9.
10. Blin, K., et al., *antiSMASH 2.0--a versatile platform for genome mining of secondary metabolite producers.* Nucleic Acids Res, 2013. **41**(Web Server issue): p. W204-12.

4. Cloning and heterologous expression of the grecocycline biosynthetic gene cluster

Introduction

Actinobacteria is a potential source of secondary metabolites with diverse chemical scaffolds and interesting biological activities. 43% of microbial biologically active compounds were isolated from Actinobacteria, especially the excellent producers of the genus – *Streptomyces* [1]. The implementation of Actinobacteria biosynthetic potential involves the expression of orphan secondary metabolite gene clusters in heterologous host. Therefore, construction of vectors carrying entire biosynthetic gene clusters is of high interest.

Nowadays cosmid vectors are routinely used for genomic libraries constructions, they can accommodate between 31kb and 44 kb of foreign DNA. However, gene clusters encoding secondary metabolites may have size over 100kb i and cosmid vectors capacity is simply not great enough to deal with such large DNA segments. There are several alternatives of replicons capable to accommodate larger DNA insertions: P15a, RK2 and F1 (BAC). P15a is a low copy (about 15 copies of vector per cell) replicon from *E. coli* [2]. Its utility in maintaining large DNA constructs was demonstrated in the assembly of 62.4 kb large epothilone biosynthetic gene cluster [3]. RK2 replicons belong to the IncP incompatibility group and have been used for metagenomics library constructions [4]. Vectors constructed from this replicon function in numerous Gram-negative bacterial species and have been transferred to Gram-positive bacteria, yeast and mammalian cells [5, 6]. Bacterial artificial chromosome (BACs) and P1 artificial chromosomes (PACs)-derived libraries are good alternative to cosmid libraries when handling of large DNA fragments is needed. PACs that combine features of BACs and bacteriophage P1 vectors can carry inserts in size from 60kb to 150kb, whereas BACs are capable of cloning and propagating of large DNA fragments up to 700 kb (with an average insert size 150 kb). Though, construction of BAC-derived libraries requires special equipment, is time-consuming and expensive.

During the past years transformation-associated recombination (TAR) in yeast was successfully applied to assemble and clone large DNA fragments, including secondary metabolite gene clusters [7-10]. Originally TAR method was developed for cloning of large genomic fragments without having to construct and screen genomic libraries [11]. This approach relies on homologous recombination between DNA of interest and short (~ 60 bp) “capture arms” of the TAR-vector. Advantage of the method is elimination of in vitro

enzymatic reactions such as restriction and ligation and reducing the amount of DNA handling.

Herein, we report successful assembly and heterologous expression of 35-kb long grecocycline biosynthetic gene cluster (*gre*) using TAR method. Fragments of the gene cluster were obtained using PCR and transformed into *S. cerevisiae* together with the "capture" vector, resulting in the construct, carrying the entire gene cluster. The obtained construct was successfully expressed in *S. albus*.

Results

Grecocycline biosynthetic gene cluster identification and description

The *Streptomyces* sp. Acta1362 draft genome contains 8,710,318 base pairs with an average GC content of 71% and includes 8,026 open reading frames. Using antiSMASH we have predicted 41 potential biosynthetic gene clusters, containing two type II PKS clusters [12]. As grecocyclines contain sugar units (Figure 19), the presence of genes encoding for glycosyltransferases and deoxysugar biosynthesis are expected within the biosynthetic gene cluster. Indeed, three glycosyltransferases were identified in one of type II PKS gene cluster, along with genes required for the deoxysugars biosynthesis, genes encoding post-PKS tailoring enzymes, regulatory and transporter genes. Altogether they compose the putative grecocycline biosynthetic gene cluster (*gre* cluster) that comprises 32 predicted open reading frames (ORF) (Figure 20, Table 5).

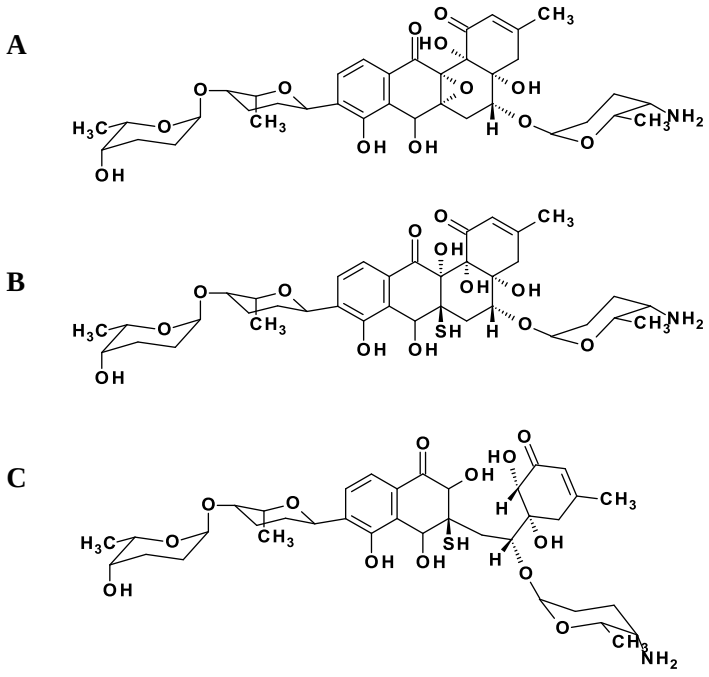


Figure 19. Structures of grecocylines A (A), B (B) and C (C).

There are several genes involved in the angucycline formation and maturation: 1) the type II PKS gene set is represented by *greA1*, *greA2* and *greA3* encoding alpha and beta ketosynthases (KSα and KSβ) and acyl carrier protein (ACP), respectively; 2) two cyclase encoding genes - *greA4* and *greA5*; 3) three ketoreductase encoding genes – *greA6*, *greV* and *greO*; 4) four oxygenase genes – *greM2*, *greE*, *greL* and *greD*; 5) one gene, *greA7*, which is responsible for the activation of ACP with its phosphopantetheine arm; and 6) the gene encoding a decarboxylase, *greN*. Interestingly, *greD* is the only one transcribed on the antisense strand. An additional gene that is encoding enzyme presumably acting on the angucyclic core of the grecocyclines is *greTh*. The deduced amino acid sequence of GreTh is highly similar to thioesterases and may be involved in formation of the thiol group in grecocycline B.

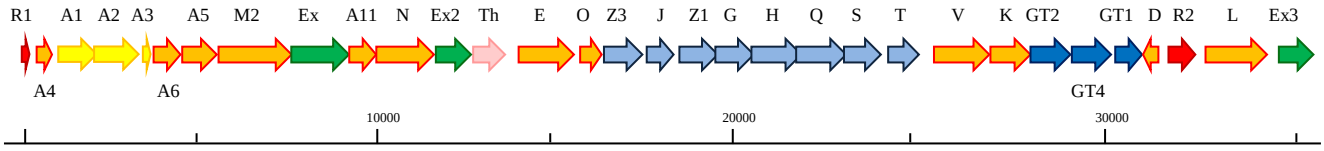


Figure 20. Grecocycline biosynthetic gene cluster (gre) from *S.* sp. Acta 13-62. In yellow are labeled genes encoding type II PKS, orange – genes involved in angucycline formation, blue – genes involved in biosynthesis of rhodinose and α-tolyposamine, dark blue – genes encoding glycosyltransferases, pink – gene encoding putative thioesterase, green – efflux, red – regulatory genes. Dashed arrow indicates formation of shunt product grecocycline C.

Two different deoxysugars are attached to the angucycline: L-rhodinose and L-tolyposamine. Eleven of thirty-two genes are encoding enzymes that are involved in biosynthesis and transfer of sugars to the polyketide. The genes *greG*, *greH*, *greS*, *greT*, *greQ*, *greZ1* and *greZ3* encode enzymes that are involved in the formation of NDP-rhodinose from D-glucose-1-phosphate. Additionally, greJ encodes NDP-hexose-4-aminotransferase that probably transfers aminogroup to the hexose to form L-tolyposamine.

Grecocyclines contain three sugar units and an equal number of glycosyltransferase encoding genes was identified in the gre cluster – GreGT1, GreGT2 and GreGT4. greGT2 encodes a protein that is homologous to SaqGT5 and UrdGT2 [13, 14] and may be involved in

Table 5. Deduced function of ORFs in the grecocycline biosynthetic gene cluster

Polypeptide	aa	Similar protein	Acc. number	Identity	Proposed function
GreR1	216	PgaR1; *Streptomyces* sp. PGA64	AHW57766.1	68%	transcriptional regulator
GreA4	110	LndF; *Streptomyces globisporus*	AAU04837.1	82%	polyketide cyclase
		JadI; *Streptomyces venezuelae*	AAO65345.1	81%	
GreA1	354	PgaA; *Streptomyces* sp. PGA64	AAK57525.1	85%	ketosynthase alfa
GreA2	405	PgaB; *Streptomyces* sp. PGA64	AAK57526.1	73%	ketosynthase beta
GreA3	92	PgaC; *Streptomyces* sp. PGA64	AAK57527.1	72%	acyl carrier protein
GreA6	262	PgaD; *Streptomyces* sp. PGA64	AAK57528.1	87%	ketoreductase
GreA5	315	UrdL; *Streptomyces* fradiae	AAF00205.1	79%	aromatase
		PgaL; *Streptomyces* sp. PGA64	AAK57529.1	74%	
GreM2	794	PgaM; *Streptomyces* sp. PGA64	AAK57530.1	67%	two-domain mono-oxygenase
GreEx	503	PgaJ; *Streptomyces* sp. PGA64	AAK57531.1	53%	transporter
GreA11	236	OvmF; *Streptomyces antibioticus*	CAG14972.1	63%	4'-phosphopantetheinyl transferase
GreN	515	PgaI; *Streptomyces* sp. PGA64	AAK57534.1	84%	acyl-CoA carboxylase, beta-subunit
GreEx2	391	uncultured soil bacterium V167	ACX83629.1	59%	putative major facilitator transporter
GreTH	305	*Streptomyces roseoverticillatus*	WP_030366371.1	63%	thioesterase
GreE	498	PgaE; *Streptomyces* sp. PGA64	AAK57522.1	97%	monooxygenase
GreO	199	UrdO; *Streptomyces fradiae*	AAF00220.1	62%	reductase
GreZ3	341	PgaZ3; *Streptomyces* sp. PGA64	AHW57779.1	45%	TDP-hexose-4-ketoreductase
GreJ	384	PgaC1; *Streptomyces* sp. PGA64	AHW57776.1	80%	TDP-hexose-4-aminotransferase
GreZ1	200	PgaZ1; *Streptomyces* sp. PGA64	AHW57777.1	74%	TDP-hexose-3,5-epimerase
GreG	356	PgaG; *Streptomyces* sp. PGA64	AHW57786.1	75%	TDP-hexose synthetase
GreH	337	PgaH1; *Streptomyces* sp. PGA64	AHW57787.1	83%	TDP-hexose-4,6-dehydratase
GreQ	435	PgaQ; *Streptomyces* sp. PGA64	AHW57788.1	86%	TDP-hexose-3,4-dehydratase
GreS	465	PgaS; *Streptomyces* sp. PGA64	AHW57789.1	71%	TDP-hexose-2,3-dehydratase
GreT	328	PgaT; *Streptomyces* sp. PGA64	AHW57790.1	66%	TDP-hexose-3-ketoreductase
GreV	254	LanV; *Streptomyces cyanogenus*	AAD13552.1	62%	Reductase homolog
GreK	497	SaqE; *Micromonospora* sp. Tu 6368	ACP19351.1	68%	putative oxygenase
GreGT2	379	SaqGT5; *Micromonospora* sp. Tu 6368	ACP19370.1	59%	glycosyltransferase
GreGT4	425	Lcz36; *Streptomyces sanglieri*	ABX71153.1	54%	glycosyltransferase
GreGT1	392	SaqGT3; *Micromonospora* sp. Tu 6368	ACP19364.1	59%	glycosyltransferase
GreD	217	FrnE; *Streptomyces roseofulvus*	AAC18100.1	56%	DSBA oxidoreductase
GreR2	118	*Streptomyces* sp. CNH287	WP_027750658.1	59%	HxlR family transcriptional regulator
GreL	227	*Actinoplanes missouriensis*	WP_014442785.1	40%	putative monooxygenase
GreEx3	542	*Streptomyces* sp. W007	WP_007448654.1	95%	MFS transporter

attachment of L-rhodinose to the C-9 position of the aglycon similar to urdamycins and saquayamycins. Probably, GreGT1 is responsible for introduction of second L-rhodinose unit, as it is highly similar to SaqGT3, SaqGT4, LanGT1 which are responsible for the extention of an oligosaccharide chain in angucyclines [13-15]. To our knowledge, grecocyclines are the first angucyclines that carry the sugar moiety in C-5 position of the angucyclic core. Furthermore, α-tolyposamine is a very rare deoxysugar in natural products, so far it was found only in BU-4514N from Microtetraspora sp. [16]. We presume that GreGT4 is transferring aminosugar to the C-5 position of grecocyclines. The deduced amino sequence of GreGT4 share 54% with Lcz3 from the lactanomycin gene cluster and 50% identity with UrdGT1a involved in urdamycin biosynthesis [17, 18]. As mentioned above, *greL* encodes

monooxygenase that doesn't have a predicted function, and might be a potential candidate for hydroxylation of the aglycon at C-5 prior to the sugar attachment (Figure 21).

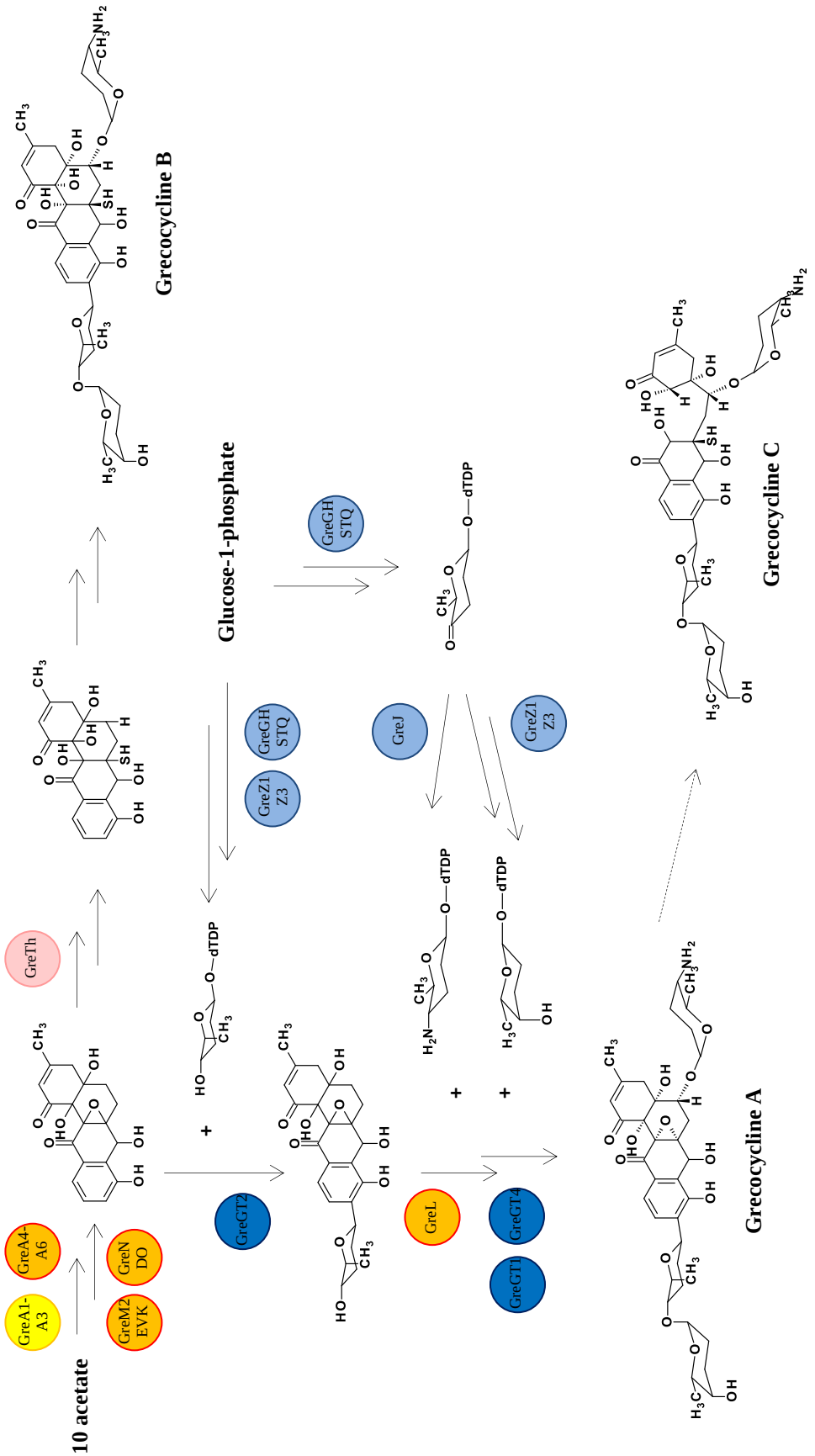


Figure 21. Proposed biosynthesis of grecocyclines. In circles are indicated enzymes putatively involved in particular biosynthetic steps.

Two regulatory – *greR1*, *greR2* and three transporter genes – *greEx*, *greEx2* and *greEx3* have been identified in the gre cluster. greR1 is located on the 5' end of the gene cluster, and encodes protein with the similarity to transcriptional regulators from the OmpR family [19]. GreR2 is homologous to the HxlR family of transcriptional regulators. GreEx, GreEx2 and GreEx3 are similar to MFS transporters and probably are exporting grecocyclines out of the cell.

All the attempts to generate the deletion mutants in *S.* sp. Acta 13-62 failed and therefore we decided to assemble the gene cluster for heterologous expression

Grecocylcine gene cluster assembly in yeast

To express the gene cluster in the heterologous host, we targeted appropriate DNA region for TAR cloning (Figure 22). Previous reports showed that short 40- to 70-bp DNA-specific targeting sequences were enough for successful TAR cloning [20]. The grecocycline biosynthetic gene cluster was split on three regions, R1, R2 and R3, overlapping with each other and with the shuttle vector (pCLY10) in 38 bp. The capture arms to the gre cluster were introduced into pCLY10 via PCR to generate pCLY10ol. Primers were designed to overlap 38 bp of the 500 bp long region upstream from the gre cluster (forward primer) and 38 bp of the 130 bp long region downstream from the cluster (reverse primer). Assembly of the gre cluster was done in two steps: 1) capture of R1 and R3 on the shuttle vector to give pR1R3; 2) capture of R2 on pR1R3 to give pGRE. R1, R2 and R3 were amplified in PCR reaction with the primers carrying homologous arms to the adjacent DNA regions. After transformation of S. cerevisiae BY4742 cells with the mixture of obtained PCR-products (R1, R3 and pCLY10ol), positive clones were identified by PCR and confirmed by restriction mapping to give pR1R3. Later, linearized pR1R3 and R2 (R2.1 + R2.2) were transformed into yeast cells and resulted in pGRE – a construct carrying the entire gre biosynthetic gene cluster. The 44-kb pGRE construct was stably maintained in *E. coli*.

Heterologous expression of pGRE

For the heterologous production, we chose *S. albus*, a model actinomycete commonly used as a biosynthetic host. pGRE was introduced by conjugative transfer from *E. coli* into the genome of *S. albus* J1074. *S. albus* strain carrying pGRE (*S. albus* pGRE) produced a series of angucyclines, as revealed by ESI-MS analysis (Figure 23).

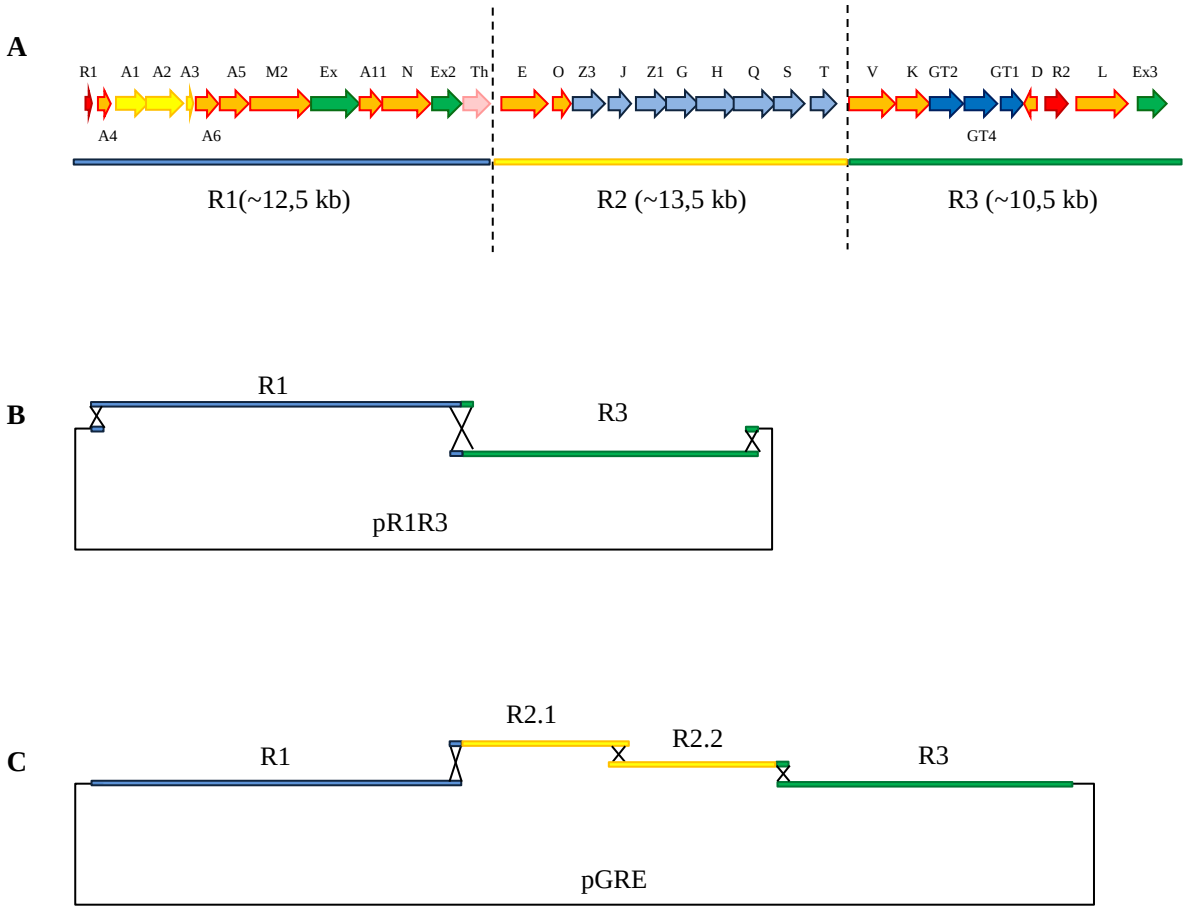


Figure 22. Strategy for assembling the grecocycline biosynthetic gene cluster using TAR. A – gre cluster divided on R1, R2 and R3; B – assembly of pR1.R3; C – assembly of pGRE.

To determine the chemical structure of produced compounds, we cultivated S. *albus* pGRE in 5 L to yield ~130mg of crude extract, from which we purified ten fractions which were analyzed by LC MS. Three compounds from fractions N4, N9 and N10 were assigned to known angucycline intermediates or shunt products, based on the UV/Vis spectrum shape and the accurate mass. The compound from fraction N4 is rabelomycin (m/z = 338.07863), fractions N9 and N10 contain dehydrorabelomycin (m/z = 320.06741) and tetrangulol (m/z = 304.07257), respectively. No glycosylated products have been detected after the heterologous expression of pGRE. Althought no grecocyclines have been detected in the crude extracts via LC MS analysis it isobvious that the cloned biosynthetic gene cluster is responsible for grecocycline biosynthesis, since several known angucyclines were observed. To identify the reason why after the heterologous expression of cloned gre gene cluster no grecocyclines were produced we have sequenced the pGRE plasmid.pGRE was sequenced via primer walking, covering approximately 97,1% of the construct. Analysis of the sequenced data revealed twenty six nucleotide substitutions.

Eight nucleotide substitutions were identified in the DNA region encoding GreM2. Inactivation of *greM2* homologs in urdamycin and landomycin biosynthetic pathways led to accumulation of rabelomycin, tetrangulol and other shunt products [21, 22], preventing subsequent conversion of aglycon into the natural product.

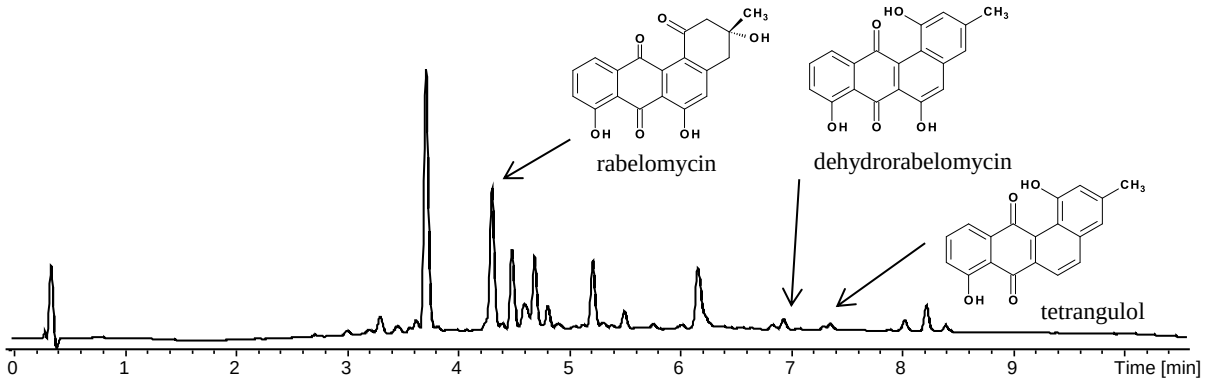


Figure 23. Chromatogram of the HPLC-ESI/MS analysis of S. albus pGRE. Arrows indicate ions corresponding to rabelomycin, dehydrorabelomycin and tetrangulol.

To complement mutated region encoding *greM2*, native *greM2* was placed under the control of the strong synthetic 21p promoter. The introduction of this construct into the *S. albus* pGRE strain did not change the spectrum of the compounds observed after the expression of only pGRE. It may be because of an unbalanced expression of *greM2* and the rest of the *gre* cluster from different plasmids.

Discussion

During the decades actinomycetes have been the source or inspiration for the numerous pharmaceutically useful compounds, including antibiotics, immunosuppressive, antiproliferative and antiparasitic drugs [23]. Also, the rise of DNA sequencing technologies revealed the wealth of natural product biosynthetic gene clusters in the genomes of actinomycetes, most of which remain cryptic [24]. One approach to address cryptic/silent biosynthetic gene cluster is heterologous expression, however cloning of large natural product clusters still represent a challenge. Therefore, technics for large DNA fragments cloning are of great interest.

Several techniques have been developed to get access to entire gene clusters for their subsequent expression e.g. sequence-independent (In-FusionTM, SLIC, Gibson isothermal assembly) and standardized restriction enzyme assembly protocols (BioBricks TM,

BglBricks, Golden Gate) [25]. For example, one-step in vitro recombination method (Gibson isothermal assembly) has a capacity to assemble and repair overlapping DNA molecules in a single isothermal step. This approach was used to assemble the complete synthetic 583-kb genome of Mycoplasma genitalium. The Gibson isothermal assembly is based on combination on 5' T5 exonuclease activity combined with DNA polymerase and Taq DNA ligase. This method is simple, rapid and may be used for scarless assembly of a large number of DNA fragments in one step [26]. However, this approach is hardly applicable for DNA with high G/C content, due to the secondary structures that may occur in the single-stranded DNA overhangs, as it would compete with the required single-stranded annealing of neighboring assembly fragments.

Another routine method widely used in laboratories is a Golden Gate Assembly, which offers seamless, multi-part DNA assembly [27]. This method uses single type II restriction endonucleases (for example BsaI) that cut outside their recognition sequence, allowing 4-bp overhangs that can be freely chose. Thus, assembly can be seamless between neighboring fragments, containing appropriate 4-bp termini. A major limitation in Golden Gate assembly may be presence of the selected type II recognition site (BsaI) in the internal sequence of DNA fragments to be assembled, additionally rather short 4-bp overhangs may limit a single multi-part assembly step when number of fragments to assemble becomes larger (greater than 10) or if the fragment termini is characterized by extremely high or law GC% content. Furthermore, base-pair mismatching at similar overhangs had been reported [25].

TAR method has been a common protocol for manipulating DNA in yeast for over a decade, and like methods described above relies on overlapping sequences between the fragments to be cloned or assembled.

In this study we used TAR method to assemble and heterologously express grecocycline biosynthetic gene cluster. The entire 36-kb gene cluster was successfully captured on the shuttle vector carrying short (38-bp) homologous shoulders. We did cluster assembly in two steps. The advantage of step by step method is very fast and effortless obtaining of all parts and their joining together. The excessive DNA fragments, which are coming when a genomic DNA mixture is used for transformation, may sufficiently reduce the recombination frequency. Hence, yeast transformation only with essential fragments may increase frequency of the correct assembly in a fold range.

On the other hand, using polymerases for generation of the building blocks gave rise to numerous mutations.

The re-assembled gre gene cluster was expressed in *S. albus*. Undoubtedly, the angucycline compounds, which we have isolated, are produced after the expression of gre biosynthetic cluster from pGRE, since the host lacks type II PKS-encoding genes.

Although no grecocycline have been detected after heterologous expression of the gre gene cluster we have isolated rabelomycin, dehydrorabelomycin and tetrangulol. This was expected, as they represent intermediates and shunt products during biosynthesis of angucycline polyketides [22, 28-30]. Also this data is in agreement with the mutations in *greM2*. The deduced amino acid sequence of GreM2 shows significant similarity to the bifunctional oxygynase-reductases from the landomycin (LndM2, LanM2, 59%), gaudimycin (PgaM, 67%) and jadomycin (JadF, 63%, partially) biosynthetic pathways. LanM2 and PgaM belong to short-chain alcohol dehydrogenases/reductases (SDRs) family and are translated as the two-domain flavoprotein oxygenase fusion with a separate C-terminal SDR [21, 31, 32]. LanM2 was shown to catalyze a thioesterase-like decarboxylative 2,3-dehydration of the ACP-tethered nascent angucycline generated by "minimal" PKS [32]. This reaction gives prejadomycin, which is recognized by further enzymes involved in the landomycins biosynthesis. The same reaction was proven in in vitro studies for JadF during one-pot defucogilvocarcin enzymatic biosynthesis, where its removal led to accumulation of rabelomycin [33]. Thus, accumulation of rabelomycin may be due to the lack of GreM2 N-terminal oxidative activity. On the other hand, conversion of UWM6 into rabelomycin was observed after incubation with LanE [32]. LanE homolog in grecocycline biosynthetic gene cluster is represented by GreE (74% identity with LanE and 76% identity with PgaE). As a result rabelomycin may originate from UWM6, which is recognized as a substrate by GreE. Only the oxygenase domain of GreM2 is required to release angucyclic intermediate tethered to ACP. One amino acid substitution (A2T) was identified in sequence of greM2 encoding N-terminal domain after sequencing of pGRE. This mutation occurred in the first amino acid after methionine and may not change enzymatic activity and folding of the GreM2 oxygenase domain. In contrast, three amino acid substitutions were identified in the reductase domain encoding sequence of greM2, would likely impair its catalytic activity. Therefore, we presume that mutated GreM2 may keep oxygenase activity and perform 2,3-dehydration of the ACP-tethered angucycline to give prejadomycin. Later prejadomycin is recognized by GreE and (a) converted to dehydrorabelomycin or (b) in cooperation with GreV, produce tetrangulol.

Table 6. List of mutations introduced into the gre genes during amplification*.

Protein	Mutation	Proposed function
GreA5	A217T	aromatase
GreM	A2T; R339C; G407C; P665L; R734C; E747G	bifunctional oxygynase-reductase
GreN	I353V	carboxylase
GreK	V79V**; K101K**; L387P;	putrative oxygenase
GreGT4	W22R	glycosyltransferase
GreGT1	A291A**	glycosyltransferase
GreD	E147V	DSBA oxidoreductase
GreL	P104T; R106Q	putative monooxygenase
GreEx3	I322F; A500V	MFS transporter

* - mutations in the intergenic regions are not included
** - nucleotide substitution do not influence amino acid sequence

Other nucleotide substitutions were located in regions encoding GreA5, GreN, GreGT1, GreGT4, GreD, GreL and GreEx3 (Table 6). Two times nucleotide change did not influence amino acid sequence (in GreGT1 and GreK). Clearly, the mutation in greA5 did not have impact on the protein function; otherwise aromatic polyketides would not be formed. Most probably mutation in GreN (I353V) also does not influence the enzyme activity significantly. GreN is highly similar (~80%) to different acetyl-CoA carboxyltransferases. The orthologs of this enzyme have been identified in many gene clusters of secondary metabolites [14, 34, 35]. For example, acyl-CoA carboxylase JadJ, from the jadomycin biosynthetic pathway, is supplying malonyl-CoA for the polyketide biosynthesis. Its inactivation severely reduced jadomycin B production [34]. Isoleucine and valine are highly similar amino acids and their swapping may not influence folding and activity of the enzyme, additionally heterologous expression of pGRE resulted in a significant yield of angucyclines. Therefore, it was concluded that mutation in *greN* is not essential. As all other mutated enzymes/proteins are acting on the later stages of biosynthesis, it is impossible to predict if they had any impact on the protein function.

In summary, TAR method is an efficient and rapid approach for assembling of large DNA molecules into one biosynthetic gene cluster and its heterologous expression, although the PCR derived mutations should be taken into account. Despite this, we believe that the cloned gene cluster is indeed involved in the biosynthesis of grecocycline in the native producer and several unique enzymes (specially GreGT4 and the sugar biosynthesis genes) will find their application in the derivatization of known active angucyclines via combinatorial biosynthesis.

Methods and Materials

Strains, growth conditions and media

For standard purposes S. sp. Acta 1362 and *S. albus* d10 wild types and mutant were grown on 2% manitol and 2% soy bean meal, pH 7.5, prepared as solid medium and tryptone soy broth (TS broth), prepared as liquid medium, at 28 0C. For maintenance of *S. albus* d10 pGRE apramycin was added to a final concentration of 50 µg/ml. For production a liquid medium DNPM (4% dextrin, 0,75% soytone, 0,5% baking yeasts and 2,1% MOPS, pH 6.8) was used. DNA manipulations were carried out in *E. coli* BG2005 and the non-methylating *E. coli* ET12567/pUZ8002 was used to drive conjugative transfer of non-methylated DNA to Actinobacteria as described previously [36]. *E. coli* strains were grown on Luria-Bertani agar or liquid medium containing appropriate antibiotic for selection. pKG1139, which contains *gusA* and apramycin-resistance genes was used for heterologous gene expression, was present in our laboratory [37].

General genetic manipulation and standard PCR

Standard molecular biology procedures were performed as described previously [38]. Isolation of plasmid DNA from *E. coli* and DNA restriction/ligation were performed by following the protocols of the manufacturers of the kits, enzymes, and reagents, Qiagen, Promega, NEB and ThermoScientific. PCR reactions were performed by using Phusion High-Fidelity DNA polymerase (Thermo Fisher Scientific) for complementation and expression experiments and DreamTaq polymerase (Thermo Fisher Scientific) to verify mutants. Primers were purchased from Eurofins MWG Operon. The oligonucleotide primers that were used are listed in Table S1. The oligonucleotide primers used to sequence pGRE are listed in Table S2.

Gene complementation cassettes

pUWLgreM2. To generate the expression vector pUWLgreM2, *greM2* was amplified by PCR. Suitable restriction sites (*XbaI* and *KpnI*) were introduced upstream and downstream from the gene, using primers. Also forward primer was carrying 21p promoter. To obtain suitable shuttle vector, pUWLCre was treated with KpnI/XbaI and 8kb-fragment was eluted from the agarose gel (done according to the manufacturer, Promega). 2.4 kb PCR-fragment was digested with XbaI and KpnI and ligated into pUWL vector, to generate pUWLgreM2.

gre gene cluster assembly using TAR

Homologous sequences to the gre cluster were introduced into pCLY10 via PCR. Primers were designed to overlap 38 bp of the 500 bp long region upstream from the gre cluster (forward primer) and 38 bp of the 130 bp long region downstream from the cluster (reverse primer). Prior to the PCR pCLY10 was treated with HindIII, obtained PCR-product was treated with DpnI for 3 hours. gre cluster was split into 3 parts: region 1, region 2 and region 3 (R1, R2 and R3), each size in range from 10 to 14 kb (Fig. N.). R1 and R3 were chosen for the first step of the cluster assembly. Primers for R1 were designed to overlap with 38bp pCLY10 (forward primer) and R3 (reverse primer); R3-primers had 38-bp long overlaps with R1 (forward primer) and pCLY10 (reverse primer). Mixture of obtained PCR-products (R1, R3 and pCLY10) was transformed into Saccharomyces cerevisiae BG4742 (LEU) cells in concentration 100 ng each. After 4 days of incubation at 30 oC, transformants were picked from the plates, plasmid DNA was isolated and transformed into *E. coli* GB2005 cells. Verification of obtained constructs revealed pCLY10 carrying R1 and R3 regions (pCR1.R3). ScaI restriction site was introduced between R1 and R3 artificially using primers, thus giving possibility to introduce R2 into pCR1.R3. Due to the problems with PCR, R2 was split into two parts: R2.1 (5743 bp) and R2.2 (7586 bp), overlap between R2.1 and R1 was 110 bp, R2.2 and R3 49 bp, R2.1 and R2.2 41 bp. To amplify pCLY10, R1 and R3 MasterAmpTM Extra-Long PCR Kit from Epicentre was used, to amplify R2.1 and R2.3 polymerase from the kit was substituted with Phusion High-Fidelity DNA Polymerase from ThermoScientific. Mixture of R2.1, R2.2 PCR-products and pCR1.R3, treated with ScaI, was transformed into yeast cells. Transformants were picked from the plates after 4 days incubation at 30 °C, prior to plasmid DNA isolation, colony-PCR was done. Primers were specific to the DNA sequence that can be present only in a reassembled gene cluster construct. 22 positive clones were picked from 95 after PCR, 4 clones were selected randomly and plasmid DNA was isolated. After verification all four clones were positive for containing pR1.R3 reassembled with R2.1 and R2.2 (pGRE).

Conditions for amplification of R1, R2.1, R2.2, R3 and pCLY10

To amplify pCLY10, R1, R2.1, R2.2 and R3 Master-AmpTM Extra-Long PCR Kit (Epicentre) was used: PreMix 4 for amplification R3 and R2.2, PreMix 8 for amplification R1 and R2.1. Phusion High - Fidelity DNA (Thermo Fisher Scientific) polymerase was used to amplify R2.1 and R2.2, and Master-AmpTM Extra-Long Polymerase Mix (Epicentre) for R1 and R3 amplification.

Transformation of yeast

Yeast transformation was done as described before [39].

Isolation of plasmid DNA from yeast

Plasmid DNA was isolated from the yeast cells using Wizard Plus SV Miniprep kit from Promega. 37,5 Units of Longlife Zymolyase (G-Biosciences, USA) were added to P1 buffer, samples were incubated at 37 C for 30-60 min. Further steps were done according to the manufacturer.

Purification of compounds

S. albus pGRE was cultivated in DNPM medium in shaking flasks, total volume 5 L at 30°C, 200 rpm for 5 days. Mixture of culture broth and biomass was extracted with 5 L of ethylacetate (EtAc) and led to 120 mg of crude extract. The extract was purified by HPLC (HTec C18 column, 250 × 10 mm, 5 µm, A = H2O + 0.1% FA, B = ACN + 0.1% FA) with gradient starting from 5% B to 95% B in 28.2 min. Flow 5.8 ml/min, injection volume 40 µl of an MeOH extract (c = 5.3 mg/ml), detection wavelength λ = 420 nm.

References

1. *Antibiotics: Current Innovations and Future Trends*2015: Caister Academic Press. xii + 430 (plus colour plates).
2. Sambrook, J., Fritsch, E.F., and Maniatis, T. , *Molecular Cloning: A Laboratory Manual*. 2 ed1989, Cold Spring Harbor, New York.: Cold Spring Harbor Laboratory Press
3. Zhang, L., G. Zhao, and X. Ding, *Tandem assembly of the epothilone biosynthetic gene cluster by in vitro site-specific recombination.* Sci Rep, 2011. **1**: p. 141.
4. Wexler, M., et al., *A wide host-range metagenomic library from a waste water treatment plant yields a novel alcohol/aldehyde dehydrogenase.* Environ Microbiol, 2005. **7**(12): p. 1917-26.
5. Aakvik, T., et al., *A plasmid RK2-based broad-host-range cloning vector useful for transfer of metagenomic libraries to a variety of bacterial species.* FEMS Microbiol Lett, 2009. **296**(2): p. 149-58.
6. Bates, S., A.M. Cashmore, and B.M. Wilkins, *IncP plasmids are unusually effective in mediating conjugation of Escherichia coli and Saccharomyces cerevisiae: involvement of the tra2 mating system.* J Bacteriol, 1998. **180**(24): p. 6538-43.
7. Yamanaka, K., et al., *Direct cloning and refactoring of a silent lipopeptide biosynthetic gene cluster yields the antibiotic taromycin A.* Proc Natl Acad Sci U S A, 2014. **111**(5): p. 1957-62.
8. Bonet, B., et al., *Direct Capture and Heterologous Expression of Salinispora Natural Product Genes for the Biosynthesis of Enterocin.* J Nat Prod, 2014.
9. Ross, A.C., et al., *Targeted Capture and Heterologous Expression of the Pseudoalteromonas Alterochromide Gene Cluster in Escherichia coli Represents a Promising Natural Product Exploratory Platform.* ACS Synth Biol, 2014.
10. Kim, J.H., et al., *Cloning large natural product gene clusters from the environment: piecing environmental DNA gene clusters back together with TAR.* Biopolymers, 2010. **93**(9): p. 833-44.
11. Larionov, V., et al., *Highly selective isolation of human DNAs from rodent-human hybrid cells as circular yeast artificial chromosomes by transformation-associated recombination cloning.* Proc Natl Acad Sci U S A, 1996. **93**(24): p. 13925-30.
12. Blin, K., et al., *antiSMASH 2.0--a versatile platform for genome mining of secondary metabolite producers.* Nucleic Acids Res, 2013. **41**(Web Server issue): p. W204-12.
13. Erb, A., et al., *Cloning and sequencing of the biosynthetic gene cluster for saquayamycin Z and galtamycin B and the elucidation of the assembly of their saccharide chains.* Chembiochem, 2009. **10**(8): p. 1392-401.
14. Westrich, L., et al., *Cloning and characterization of a gene cluster from Streptomyces cyanogenus S136 probably involved in landomycin biosynthesis.* FEMS Microbiol Lett, 1999. **170**(2): p. 381-7.
15. Luzhetskyy, A., et al., *Iteratively acting glycosyltransferases involved in the hexasaccharide biosynthesis of landomycin A.* Chem Biol, 2005. **12**(7): p. 725-9.
16. Toda, S., et al., *A new neuritogenetic compound BU-4514N produced by Microtetraspora sp.* J Antibiot (Tokyo), 1993. **46**(6): p. 875-83.
17. Trefzer, A., et al., *Function of glycosyltransferase genes involved in urdamycin A biosynthesis.* Chem Biol, 2000. **7**(2): p. 133-42.

18. Zhang, X., et al., *Biosynthetic investigations of lactonamycin and lactonamycin z: cloning of the biosynthetic gene clusters and discovery of an unusual starter unit.* Antimicrob Agents Chemother, 2008. **52**(2): p. 574-85.
19. Itou, H. and I. Tanaka, *The OmpR-family of proteins: Insight into the tertiary structure and functions of two-component regulator proteins.* Journal of Biochemistry, 2001. **129**(3): p. 343-350.
20. Kouprina, N. and V. Larionov, *Selective isolation of genomic loci from complex genomes by transformation-associated recombination cloning in the yeast Saccharomyces cerevisiae.* Nat Protoc, 2008. **3**(3): p. 371-7.
21. Zhu, L., et al., *Identification of the function of gene lndM2 encoding a bifunctional oxygenase-reductase involved in the biosynthesis of the antitumor antibiotic landomycin E by Streptomyces globisporus 1912 supports the originally assigned structure for landomycinone.* J Org Chem, 2005. **70**(2): p. 631-8.
22. Faust, B., et al., *Two new tailoring enzymes, a glycosyltransferase and an oxygenase, involved in biosynthesis of the angucycline antibiotic urdamycin A in Streptomyces fradiae Tu2717.* Microbiology, 2000. **146 (Pt 1)**: p. 147-54.
23. Berdy, J., *Bioactive microbial metabolites.* J Antibiot (Tokyo), 2005. **58**(1): p. 1-26.
24. Bentley, S.D., et al., *Complete genome sequence of the model actinomycete Streptomyces coelicolor A3(2).* Nature, 2002. **417**(6885): p. 141-7.
25. Ellis, T., T. Adie, and G.S. Baldwin, *DNA assembly for synthetic biology: from parts to pathways and beyond.* Integr Biol (Camb), 2011. **3**(2): p. 109-18.
26. Gibson, D.G., et al., *Enzymatic assembly of DNA molecules up to several hundred kilobases.* Nat Methods, 2009. **6**(5): p. 343-5.
27. Engler, C., R. Kandzia, and S. Marillonnet, *A one pot, one step, precision cloning method with high throughput capability.* PLoS One, 2008. **3**(11): p. e3647.
28. Luzhetskyy, A., et al., *LanGT2 Catalyzes the First Glycosylation Step during landomycin A biosynthesis.* Chembiochem, 2005. **6**(8): p. 1406-10.
29. Rix, U., et al., *The oxidative ring cleavage in jadomycin biosynthesis: a multistep oxygenation cascade in a biosynthetic black box.* Chembiochem, 2005. **6**(5): p. 838-45.
30. Lombo, F., et al., *Elucidation of oxygenation steps during oviedomycin biosynthesis and generation of derivatives with increased antitumor activity.* Chembiochem, 2009. **10**(2): p. 296-303.
31. Palmu, K., et al., *Artificial reconstruction of two cryptic angucycline antibiotic biosynthetic pathways.* Chembiochem, 2007. **8**(13): p. 1577-84.
32. Kharel, M.K., et al., *Elucidation of post-PKS tailoring steps involved in landomycin biosynthesis.* Org Biomol Chem, 2012. **10**(21): p. 4256-65.
33. Kharel, M.K. and J. Rohr, *Delineation of gilvocarcin, jadomycin, and landomycin pathways through combinatorial biosynthetic enzymology.* Curr Opin Chem Biol, 2012. **16**(1-2): p. 150-61.
34. Han, L., et al., *An acyl-coenzyme A carboxylase encoding gene associated with jadomycin biosynthesis in Streptomyces venezuelae ISP5230.* Microbiology, 2000. **146 (Pt 4)**: p. 903-10.
35. Metsa-Ketela, M., K. Ylihonko, and P. Mantsala, *Partial activation of a silent angucycline-type gene cluster from a rubromycin beta producing Streptomyces sp. PGA64.* J Antibiot (Tokyo), 2004. **57**(8): p. 502-10.
36. Luzhetskyy, A., et al., *IncP plasmids are most effective in mediating conjugation between Escherichia coli and streptomycetes.* Genetika, 2006. **42**(5): p. 595-601.

37. Myronovskyi, M., et al., *Beta-glucuronidase as a sensitive and versatile reporter in actinomycetes.* Appl Environ Microbiol, 2011. **77**(15): p. 5370-83.
38. Green, M.R. and J. Sambrook, *Molecular cloning : a laboratory manual*. 4th ed2012, Cold Spring Harbor, N.Y.: Cold Spring Harbor Laboratory Press.
39. Gietz, R.D. and R.A. Woods, *Transformation of yeast by lithium acetate/single-stranded carrier DNA/polyethylene glycol method.* Methods Enzymol, 2002. **350**: p. 87-96.

III. DISCUSSION

1. *sim* biosynthetic gene cluster

Given the redundancy of biosynthetic gene clusters within actinobacteria genomes, we have screened the genomic database for availability of simocyclinone biosynthetic genes. Indeed, two strains - *K.* sp. and *S.* sp. NRRL B-24484 have been identified carrying the sub-clusters of genes in their genomes highly similar to those of simocyclinone biosynthetic gene cluster from *S. antibioticus* Tü6040. When both strains are cultivated in the production media, novel derivatives of the D-type simocyclinone can be detected in their production profiles. SD9, the major and the most active metabolite of *K.* sp. is usually produced along with minor SD10, SD11 and other D-type simocyclinones of unknown structure. Newly detected simocyclinones show structural deviation in the A-ring of the angucycline moiety. The arrangement of the other sub-structures – acetylated D-olivose, tetraene chain and aminocoumarin ring is identical between described before SD8 and newly isolated molecules. Based on the high similarity of the chemical pattern of SD8 and SD9/SD10/SD11, one can expect that corresponding gene clusters would be nearly identical with the few deviations. Surprisingly, despite the high similarity between the individual genes, the overall architecture of SD8 and SD9 gene clusters is organized in the substantially different ways.

1.1 Analysis of the genes/enzymes putatively involved in the angucycline biosynthesis and modification

Thirteen genes (*smcC7, smcA7-smcA12*) involved in the angucycline formation are organized in one operon within the *smc* cluster (*smcA* subcluster).

The gene products of *smcA1, smcA2* and *smcA3* show significant homology to the type II iterative PKSs from different species. SmcA1 and SmcA2 resemble ketoacylsynthases and SmcA3 is the acyl carrier protein homologue; together they compose the expected “minimal PKS” set. SmcA1 contains typical catalytic triad Cys(169)-His(309)-His(346) and obviously catalyzes Claisen condensation; SmcA2 with the conserved glutamine residue in the active center is a CLF. Also, a typical LGyeS (39-43) phosphopantetheinyl binding sequence [70] has been identified within this protein. Phosphopantetheinyl transferase, necessary for ACP activation is encoded by *smcA11* and is located downstream in the *smcA* subcluster. SmcA4 and SmcA5 are putative cyclases showing 92% and 90% homology to the SimA4 and SimA5, respectively. Together with the product of *smcA6* (putative ketoreductase with 90% and 87%

homology to *pgaD* and *aurI*), these proteins should generate UMW6, a common precursor of many angucycline polyketides.

Angucyclic core of simocyclinones is highly oxidized with an epoxide group at the positions C-12a and C-6a and several other unusually placed hydroxyl groups. The enzymes responsible for thosedecoration of the polyketide aglycon belong to two families: NADPH-dependent flavoprotein mono-oxagenases and short-chain alcohol dehydrogenases/reductases (SDR). Because of the significant sequences similarity of their proteins, function prediction based on the primary amino acid sequencing is challenging, thereby heterologous expression, gene-inactivation experiments or in vitro assays are required. Typically, the genes encoding oxygenases and reductases follow conserved order in the angucycline biosynthetic gene clusters and are forming a subcluster [71].

Deduced amino acid sequence of SmcA7 has the high identity to UrdE (70%) and LanE (67%) from urdamycin (*urd*) and landomycin (*lan*) biosynthetic pathways, respectively. Recent studies have shown that purified UrdE and LanE can utilize prejadomycin (2,3-dehydro-UWM6) as a substrate. UrdE catalyzed two consecutive NADPH/O2-dependent reactions – C-12 and C-12b hydroxylations, followed by 6-ketoreduction by UrdM to produce gaudimycin. In contrast, landomycins do not contain a hydroxyl group at C-12b and an incubation of LanE and LanV together with prejadomycin led to the accumulation of 11-deoxylandomycinone – an intermediate in landomycin biosynthesis. Thus, despite substantial sequence identity, LanE is making C-4a/C-12b dehydration that is promoted by LanV. Based on the structure of angucycline in simocyclinones, SmcA7 should follow "UrdE-pathway" during C-12b modification and introduce hydroxygroup (Figure 24). The Δ*smcA7*h mutant was accumulating 2,3-dehydro-UWM6 as a main product, confirming this hypothesis. Similar results were obtained *in vitro* for GilOI, which is the UrdE homologue in gilvocarcin (*gil*) biosynthesis. The removal of GilOI from the enzyme cocktail produced prejadomycin [72]. Most probably SmcA7 is performing two consecutive steps – C-12 and C-12b hydroxylations of prejadomycin, where the second intermediate is highly unstable and should be directly further modified by the following enzymes. In *urd* and *pga/cab* pathways these reactions are performed by UrdM and PgaM/CabV, respectively.

UrdM and PgaM belong to the SDR family and are translated as the two-domain flavoprotein oxygenase fusion (ox domain) with a separate C-terminal SDR (red domain) [73, 74]. In contrast, CabV is an independent reductase not linked with oxygenase domain. In *in vitro* studies, the C-6 ketoreduction was catalyzed by UrdMred, whereas UrdMox domain was not

required for C-12b hydroxylation, despite the enzyme has been linked with the reaction before [71, 73]. Additionally, coupling of the second hydroxylation reaction performed by UrdE (or PgaE) to that of the SDR enzyme is a strict requirement, otherwise this reaction invariably leads to the product degradation [71, 72, 75]. SmcA9 shares 72% identity with PgaM and 70% with UrdM and is a candidate for 6-ketoreductase reaction. However, SmcA9 doesn't contain an ox domain and is approx. half-shorter from the N-terminal side. Inactivation of *smcA9* did not lead to accumulation of prejadomycin, rabelomycin or any other intermediates or shunt products of the angucycline biosynthesis, indicating that biosynthesis diverted onto a degradation pathway and additionally proving dependence of the SmcA7 function on SmcA9. An accumulation of prejadomycin in the Δ*smcA7* mutant and the change to the degradation pathway in the ΔsmcA9 mutant suggest that the reaction sequence in simocyclinone biosynthesis is SmcA7-SmcA9 and two last reactions cannot be separated from each other similarly to the *urd* and *pga/cab* pathways.

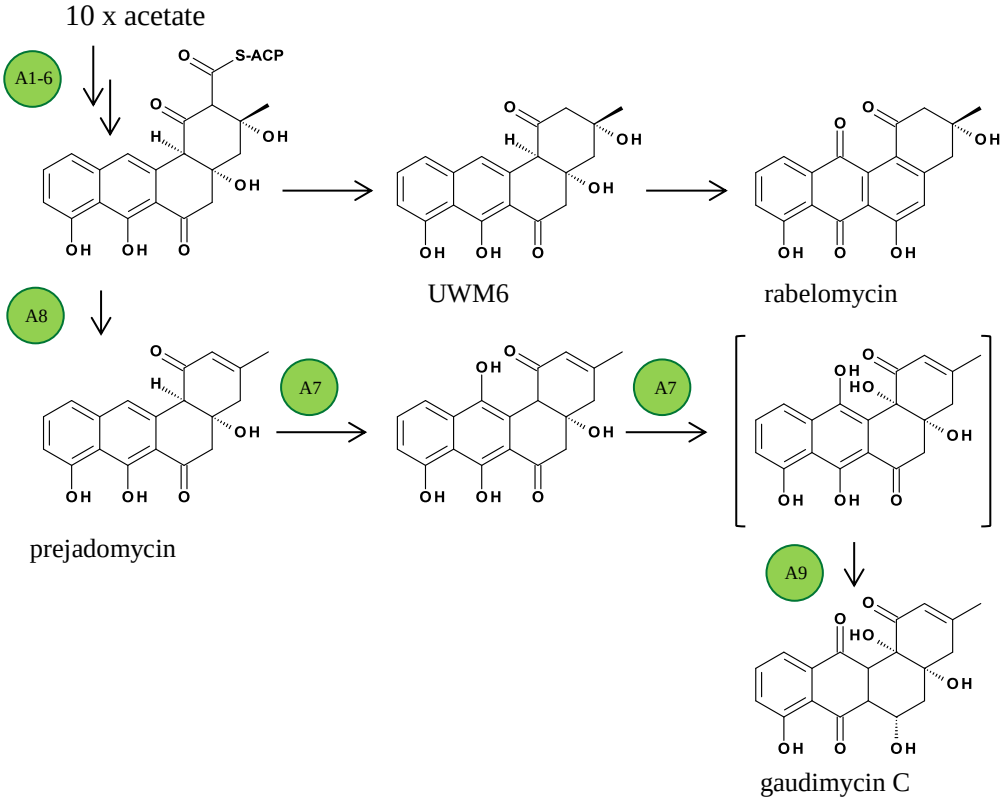


Figure 24. Proposed pathway for the biosynthesis of an angucyclic part of simocyclinones.

Upstream from *smcA9* is located *smcA8* encoding an oxygenase highly similar to JadF (*jad*, jadomycin biosynthetic pathway), and N-terminal domains of LanM2 and LndM2 from the landomycin biosynthesis in *S. globisporus*. JadF was shown to catalyze the 2,3-dehydratation on the ACP-tethered intermediate during prejadomycin biosynthesis in vitro, moreover it

catalyzes decarboxylation of the angucyclic intermediate and its hydrolysis from ACP. Removal of JadF from the enzyme cocktail during one-pot defucogilvocarcin enzymatic biosynthesis resulted in accumulation of rabelomycin. Rabelomycin is a shunt product in this pathway and was produced from UWM6 by aerial oxidation and spontaneous degradation [72]. Analysis of the *K*.sp. Δ*smcA8h* mutant revealed peak with the LC-MS and UV/Vis properties that are in agreement with rabelomycin, therefore, SmcA8 may play the same role during simocyclinone biosynthesis as JadF for jadomycin. The same was hypothesized for the ox domain of UrdM by Patrikainen et al [71].

Comparison of the putative product of *smcA10* to proteins in the database suggests that it belongs to the NADF-dependent FMN reductases (62% identity to UrdO and 63% to SaqO). The exact function of SmcA10 remains unknown; presumably it may act as a reductase during the aromatic polyketide biosynthesis. SmcA12 is a putative methylmalonyl-CoA carboxyltransferase with amino acids identity 91% to SimA12 and may take part in providing acetyl-CoA building blocks for the angucycline biosynthesis.

Any homologues of *simA13*, which encodes putative P450 hydroxylase, haven't been identified in the *smc* gene cluster. This is interesting, because SimA13 could have been a candidate for introducing of the epoxy group into the angucycline moiety. Epoxy groups have been found in many bioactive secondary metabolites and may play a key role in their pharmacological properties [76]. Epoxides are usually introduced by enzymes of cytochrome P450 family, for instance, EpoK in epothilone (epo) [77] and HedR in *hed* [78] biosynthetic pathways. Flavoenzymes are also known to create an epoxide via oxygen transfer to a C–C double bond [76]. However, BLAST search against the *smc* cluster didn't reveal any homologues of both groups of enzymes that may be involved in the epoxide formation. Thus, the origin of the epoxy group in the simocyclinones remains unclear.

An additional gene, *orf7*, encoding oxidoreductase, has been identified in the *smc* cluster. The activity of the ORF7 in *K.* sp. determines the only structural difference between SD8 and SD9. The precise analysis revealed a truncated form of the corresponding gene in *S. antibioticus* Tü6040 genome. The *orf7* homologue flanks the *sim* cluster and has not been annotated before. The full-length ORF for this gene from the *smc* cluster is highly similar to aldo-ketoreductases that reduce aldehydes and ketones to primary and secondary alcohols [79]. In order to check ORF7 involvement in reduction of the ketogroup to hydroxygroup in SD9 several attempts have been made. An interesting phenomenon was observed after several attempts to inactivate *orf7*. If *orf7* was completely replaced with the hygromycin-resistance

gene, the strain did not produce simocyclinones of D-type. The mutant with the expected phenotype was observed after an inactivation of the catalytic center of the protein. The ORF7 protein was present, but not able to catalyze reduction in the C-1 position. Probably, the enzymes involved in the simocyclinone biosynthesis form the protein complex, which is disturbed when one of the structural units is deleted. When the mutated form of the enzyme is physically present within the complex, the metabolic intermediates are passed from one enzyme to another through the intramolecular tunnel without its release into solution. Such substrate channeling prevents the release of unstable intermediates, also it makes the process faster and more efficient [80, 81]. Therefore, deletions or substitutions of genes from biosynthetic gene clusters may lead to much lower production and accumulation of shunt products. Furthermore, substrate channeling may explain why no new derivatives were detected after heterologous expression of the gene.

Recently, SimC7 was proven to be angucyclinone ketoreductase, essential for antibiotic activity of simocyclinone [82]. In *S. antibioticus* Tü6040, *simC7* is located separately from all other genes involved in angucycline biosynthesis. In contrast, *simC7* homologue in *K.* sp. is flanking *simA* subcluster and is transcriptionally coupled to it.

According to the transcriptome data, all genes involved in angucycline biosynthesis, including *orf7*, are located within one transcription unit.

1.2. Analysis of the genes/enzymes putatively involved in D-olivose biosynthesis

One of the remarkable differences between the *sim* and the *smc* gene clusters is found within the operons responsible for the dTDP-D-olivose production. In *S. antibioticus* Tü6040 seven genes (*simB1-simB7*) involved in biosynthesis of the activated D-olivose form two sub-clusters, in *K.* sp. their homologs are scattered throughout the entire genome. *smcB1* and *smcB2* have been identified approximately 900 kb downstream from the *smc* gene cluster. SimB1 and SimB2 share 47,3% and 55% of identity with their homologs in *K.* sp., respectively. Downstream from *smcB1* are located genes putatively encoding dTDP-4-dehydrorhamnose-3,5-epimerase, glycosyltransferase, dTDP-4-hexulose reductase and acetyltransferase, together they constitute a set of genes required for L-rhamnose biosynthesis [83]. It is not unique example of recruiting of the sugar biosynthesis genes from the outside for the natural product formation [84-87]. For example, during the biosynthesis of elloramycin in *S. olivaceus* [84] and caprazamycin in *Streptomyces* sp. MK730-62F2 [85] four genes (NDP-glucose synthase, dNDP-glucose-dehydratase, 4-ketoreducatse and sugar

epimerase) involved in the biosynthesis of L-rhamnose are located in different chromosomal loci.

All other genes requiredid for D-olivose biosynthesis are located within the *smc* gene cluster, but in the different order in comparison to their orthologs from the *sim* gene cluster. *smcB7*, *smcB4* and *smcB5* are forming a sub-cluster as in the *sim* cluster; *smcB3* is located at the *3'* end and is the last gene in the *smc* cluster, whereas *smcB6* is located separately. Products of these genes share 61% (SmcB6) to 82% (SmcB3) identity with their homologs from *S. antibioticus* Tü6040. Deduced amino acid sequence of SmcB7 has high similarity with the glycosyltransferases from saquayamycin (*saq*) and *urd* biosynthetic pathways. As only one glycosyltransferase has been identified within the *smc* cluster (as expected), SmcB7 should catalyze attachment of the deoxysugar to angucycline. SmcB6 is the acetyltransferase involved in the sugar decoration at 4-OH position (Figure 25).

1.3. Analysis of the genes/enzymes putatively involved in aminocoumarin moiety biosynthesis

smcD1, *smcD2*, *smcD3*, *smcD4*, *smcD5*, *smcD6*, *smcD7* and *smcX1* genes from the *smc* biosynthetic gene cluster are responsible for the biosynthesis of aminocoumarin and are nearly identical to those from the *sim* gene cluster [38, 39].

SmcD6 shows 57% similarity to NovH from the novobiocin (*nov*) biosynthetic pathway [88] and activates aminocoumarin precursor – L-tyrosine. Like SimD6, SmcD6 has an extra 400 amino acids at the N terminus, which are highly similar to the non-ribosomal peptide synthase (NRPS) condensation domain [39]. This feature may point to evolutionary origin of SmcD6 from NRPS or their common ancestor. SD9 is halogenated at the C-8 position of aminocoumarin ring; this reaction is performed by the putative halogenase SmcD4 (91% identity with SimD4) (Figure 24). The genes *smcD3*, *smcD5*, *smcD6*, *smcD1*, *smcX1* and *smcD2* are forming a sub-cluster close to the 3' end of the *smc* cluster, whereas *smcD4* is the first gene in the *smc* cluster from the 5' end.

1.4. Analysis of the genes/enzymes putatively involved in tetraene side chain biosynthesis

The most pronounced difference between the two simocyclinone biosynthetic gene clusters (*sim* and *smc*) is the absence of the type I PKS encoding machinery in the *smc* cluster. Furthermore, no homologs of *simC1A*, *simC1B* and *simC1C* have been detected elsewhere in the genome of *K.* sp.

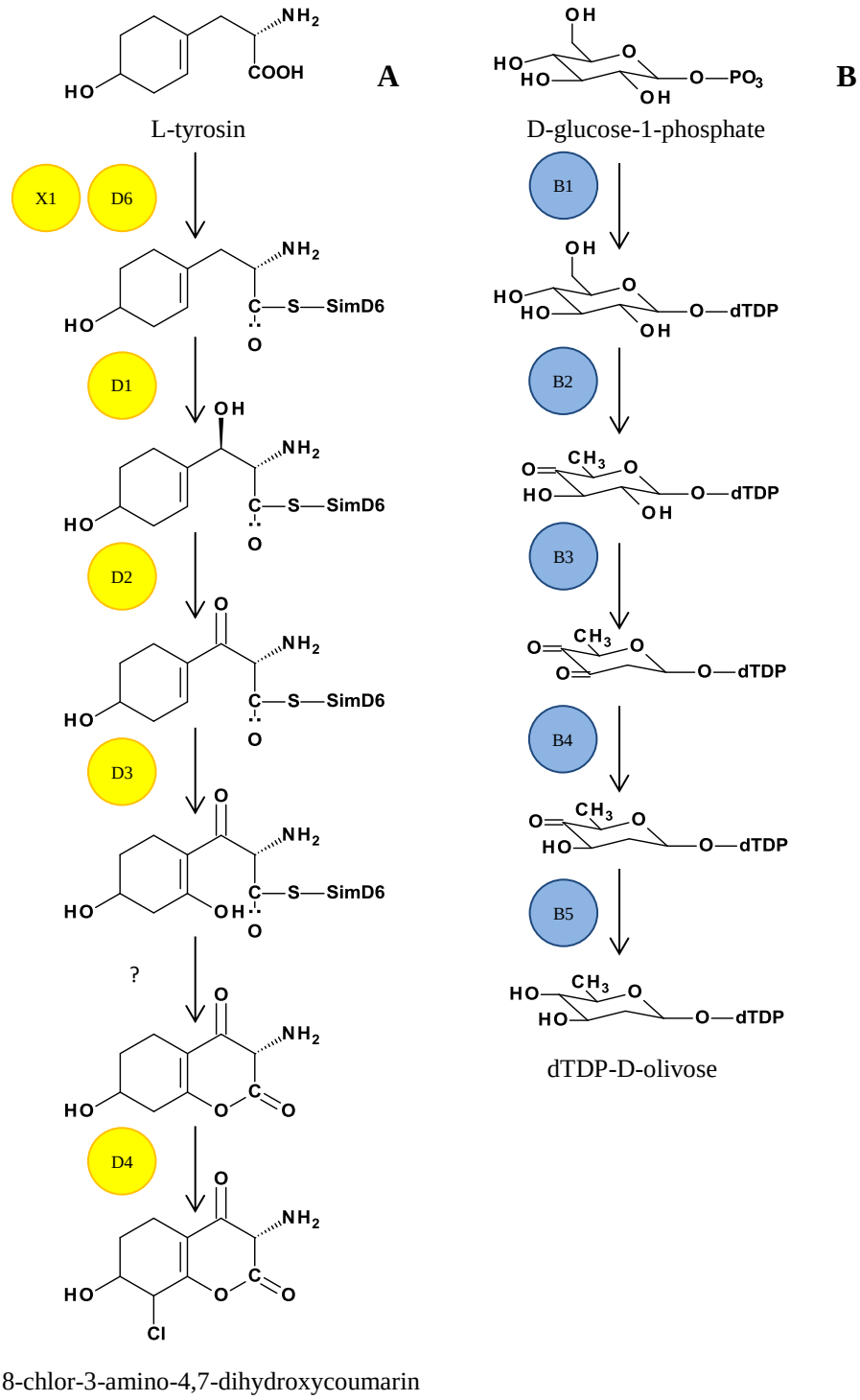


Figure 25. Proposed pathway for the biosynthesis of aminocoumarin (A) and D-olivose (B)

A close analysis of the *K.* sp. genome revealed type I PKS encoding gene, however it's inactivation didn't have any impact on the simocyclinone production. Furthermore, the inactivation of *simC1B* gene does not influence SD8 biosynthesis supporting the idea of the alternative tetraene biosynthetic pathway. Therefore, the hypothesis of tetraene biosynthesis

that is based on type I PKS involvement should be revised. Careful re-annotation of the *smc* and *sim* gene clusters revealed three ketosynthase encoding genes (*smcKSI*, *smcKSII* and *smcX5*), which might take part in a linear polyketide biosynthesis. Interestingly, their orthologs in *S. antibioticus* Tü6040 were suggested to be not involved in simocyclinone biosynthesis and marked as flanking genes (*orf2* and *orf3*) or genes of unknown function (*simX5*). However, the transcriptome of the *smc* and *sim* biosynthetic gene clusters clearly shows that the expression of the *smcKSI*, *smcKSII* and *smcX5* (*orf2*, *orf3* and *simX5*) genes is coordinated with the expression of other genes in both clusters.

Indeed, the deletion of SmcKSII from this complex completely abolished the formation of D-type simocyclinones and led to the synthesis of glycosylated angucyclines. To generate the decatetraene side chain with three ketosynthases, the assembly should most probably start with a four-carbon unit (e.g., crotonyl-CoA, acetoacetyl-CoA) unless one of the KS does not act iteratively. A number of ketosynthases employ non-acetate starter units for the polyketide biosynthesis [15, 25]. Very often, incorporation of an alternative to acetate primer units involves KSIII and ACP components as described for frenolicin, daunorubicin and others [25, 89, 90]. 3-oxoacyl-ACP-synthase III (KAS III, FabH) homologs have been identified in both *sim* (*simX5*) and *smc* (*smcX5*) clusters. KAS III is a key enzyme in fatty acid biosynthesis in plants and bacteria; it initiates the condensing process and uses acetyl-CoA as a primer and malonyl-ACP as an extender unit. Nevertheless, some KAS III-like enzymes from different organisms can prime polyketide biosynthesis with different starter units [91]. FabH homologs have been found in some type II PKS gene clusters [28, 78, 89, 92-94]. In all cases, they are involved in the generation of the non-acetate priming unit for polyketide chain biosynthesis. SmcX5 contains cysteine, proline and histidine in its active center and thus may be a candidate for the initiation of tetraene chain formation. Interestingly, the feeding experiment with the [1-^{13}C] labelled acetate did not show perfect incorporation of C-9'' into the tetraene chain, which indirectly supports the idea of a non-acetate priming of this moiety [95]. Two other ketosynthases SmcKSI and SmcKSII, together with ACP, should be involved in the extension of the linear polyketide. Several more reactions are necessary to form the fully maturated dicarboxylic tetraene acids – ketoreduction, dehydratation and incorporation of the second carboxy group. Most of the gene encoding proteins with the necessary enzymatic activities can be found within both *sim* and *smc* gene clusters. SimC6/SmcC6 is homologous to 3-oxoacyl-ACP-reductases and a higher identity with FabG4 from *S. avermitilis MA-4680*. The PKS KR domains also belong to the same family. Most likely, this enzyme is involved in the subsequent ketoreduction of the polyketide chain. We could not identify the gene

candidate responsible for the dehydration reaction of the hydroxylated polyketide intermediate, therefore the involvement of dehydratase from fatty acid biosynthesis cannot be excluded. Finally, the second carboxy group could be incorporated by two enzymes, SimC4/SmcC4 and SimC5/SmcC5, to the tetraene chain. The deduced amino acid sequence of the *smcC4* encoding protein is similar to the carotenoid dioxygenases that catalyze the oxidative cleavage of carotenoids at various chain positions, thus leading to the formation of an aldehyde [96]. The oxygen at the C-10 position of the tetraene indeed derives from molecular oxygen, as previously shown by feeding experiments [44]. Newly generated aldehyde is oxidized to the carboxylic acid by SimC5/SmcC5 (aldehyde dehydrogenase) and later is attached to the sugar. The inactivation of both genes, *simC4* and *simC5*, in *S. antibioticus* Tü6040 led to the complete cessation of SD8 production, with the corresponding accumulation of the B-type simocyclinones (angucycline with the deoxysugar attached) [97]. Although the exact mechanism of the decatetraene dicarboxylic acid formation remains unclear at the moment and requires further deciphering, the scheme and enzymes involved seems to be plausible.

1.5. Genes of regulation and self-resistance

Three regulatory and two efflux genes were annotated in the *smc* gene cluster. SmcReg1 shares 65% identity with SimReg1 and presumably is a master switch of the simocyclinone biosynthesis in *K.* sp. SmcReg2 and SmcEx1 form a TetR/TetA-like resistance pair, like SimR/SimX and show 82%/79% identity, respectively. SmcReg3 is highly similar to the MarR/SlyA transcription regulators involved in development of the resistance to multiple antibiotics, oxidative stress agents, oxidative stress, etc [98]. SmcEx2 is 90% identical to SimEx2 and is a putative Na^+/H^+ antiporter.

1.6. Genes of unknown function

Three genes of unknown function have been identified in the *smc* cluster – *smcX2*, *smcX4* and *smcX7*. The gene *smcX2* is located directly downstream of simA12 and both genes are translationally coupled . SmcX2 is a hypothetical protein with the acyl-CoA carboxylase epsilon subunit domain. Together with SmcA12 it may be involved in generation of malonyl-CoA building blocks, used for the angucycline and tetraene biosynthesis. SmcX4 is a predicted metallophosphoesterase and SmcX7 is a hypothetical protein found in different *Streptomyces* species.

2. *sml* biosynthetic gene cluster

The simocyclinone biosynthetic gene cluster from *S.* sp. NRRL B-24484 (*sml* gene cluster) has the same set of genes and architecture as the *smc* cluster. Both, the *smc* and *sml* cluster share 97% pairwise identity.

2.1. Summary

The three identified simocyclinone gene clusters (*sim*, *smc* and *sml*) are typical examples of the "bricks and mortar" evolution events postulated by Marnix et al. [99]. These hybrid gene clusters are composed of several sub-clusters (bricks) and individual genes (the "mortar"). During the evolution of the *sim* gene cluster from the *smc* gene cluster, substantial gene reorganization took place (deletion of the ORF7 function, gaining the deoxysugar biosynthetic genes and reorganizing the gene order). These cluster rearrangements led to the generation of SD8 (**1**), which is four times more active than SD9 (**2**) *in vitro*, which could have brought some selective advantages to its producer. The acquisition of two additional deoxysugar biosynthetic genes might help in the coordination of the gene expression within the *sim* gene cluster. It is not clear which selective advantages to *S. antibioticus* Tü6040 are brought by PKS I encoding genes, which are expressed, but are not necessary for SD8 (**1**) biosynthesis.

3. Amicomycins

One strain-many compounds (OSMAC) approach was described in 2000s by Zeeck and co-workers to trigger productivity of a single microorganism through turning on cryptic or silent biosynthetic genes by varying accessible cultivation parameters (media composition, pH, temperature, oxygen supply, quantity and quality of light, bioreactor platform, addition of precusors or enzyme inhibitors) [100].

Indeed, a fermentation of *K.* sp. under certain conditions led to isolation of two hitherto unknown polyketides – amicomycin C and D. Angucyclic nature of amicomycins suggests that their biosynthesis is directed by type II PKS genes. Preliminary data after the analysis of the *K.* sp. genome sequence with the secondary metabolites search tool antiSMASH 2.0 [101] indicates the presence of two type II PKS gene clusters. One of them is the simocyclinone biosynthetic gene cluster. Another type II PKS gene cluster shares 83% homology with spore pigment biosynthetic genes from different streptomycetes and is not involved in the biosynthesis of angucyclic polyketides. This observation suggests that amicomycins share biosynthetic route with the angucyclic core of simocyclinones.

Unusual is the fact that genes responsible for the glycosylation could not be identified within the cluster. Only one glycosyltransferase encoding gene (*smcB7*) has been identified in the *smc* cluster. SmcB7 is accepting D-olivose as a substrate and is forming C-C bond at C-9 position of the aglycon. From the whole broad range of glycosyltransferases targeting aromatic polyketides, only UrdGT2 was shown to establish both C- and O-glycosidic bonds [102]. Furthermore, C-1-O-glycosylation has been described only for benzanthrin A and B as well as for amycomycin B in the literature [103, 104], while the glycosylation of angucyclines at position C-6 is reported only for the antibiotic TAN-1085 and the antitumor compound J-124131 [105, 106]. It is rather unlikely that SmcB7 is accepting talose as a substrate and performs glycosylation at these unusual positions, since it shares the high homology to several glycosyltransferases attaching D-olivose via C-C bond. Additionally, 6-deoxy-α-talose is rather untypical sugar for secondary metabolites. 6-deoxy-talose has been identified in isoflavonol glycosides talosins A and B produced by *K. kifunensis*. It has been suggested that talosins are produced from exogenous genistein that is included in fermentation medium and *K. kifunensis* is decorating it with 6-deoxy-L-talose. Talosins A and B are antifungal agents and deoxy-sugar plays an import role in their biological activity [107]. The genes of talose biosynthesis in *K. kifunensis* have been cloned and their function proved in *in vitro* assay [108]. In *K. kifunensis* biosynthesis of 6-deoxy-talose is encoded within the biosynthetic gene cluster of the extracellular glycan. Four proteins (RmlA$_{Kkf}$, RmlB$_{kkf}$, RmlC$_{Kkf}$ and Tal) from this cluster are completing the biosynthesis of dTDP-6-deoxy-talose from dTTP, α-D-glucose-1-phosphate and NAD(P)H. The homologs of the genes encoding these proteins have been identified in the genome of *K.* sp. RmlA in *K.* sp. is a glucose-1-phosphate thymidyltransferase with 89% identity to RmlA$_{Kkf}$ and probably is activating α-D-glucose-1-phosphate with dTTP. The next step – conversion of dTDP-D-glucose into dTDP-6-deoxy-D-glucose-*xylo*-4-hexulose is performed by RmlB (dTDP-D-glucose-4,6-dehydratase, 92% identity to RmlB$_{Kkf}$), later it is recognized by dTDP-6-deoxy-D-xylo-4-hexulose-3,5-epimerase RmlC (83% homology with RmlC$_{Kkf}$) and is modified to dTDP-6-deoxy-L-glucose-*lyxo*-4-hexulose. At this step the sugar biosynthetic pathway is branching to give either dTDP-6-deoxy-L-talose or dTDP-L-rhamnose. Tal from *K. kifunensis* is a NAD(P)H-dependent dTDP-6-deoxy-L-talose 4-dehydrogenase it is performing the last step in talose biosynthesis. Its homolog was identified in *K.* sp. and presumably is involved in the same reaction. The order of the talose biosynthetic genes in *K.* sp is the same as in *K. kifunensis*, however the overall architecture of the cluster is slightly different. In *K. kifunensis* three glycosyltransferase-encoding genes have been identified in the gene cluster of the

extracellular glycan biosynthesis – *orf1*, *orf8* and *orf11* (N-terminally truncated), in *K.* sp. two homologs of these genes (*orf1*and *orf8*) are present. The proteins encoded by *orf1*and *orf8* might be candidates for attaching of 6-deoxy-talose to the angucycline core of amicomycins.

Production of individual secondary metabolites by Actinomycetes is sensitive to growth and media conditions. Changes in the microbial environment result in differential expression and transcription of certain genes and finally in a different metabolome. Thus, the fine control over genes expression can affect the chemical structure of the compounds produced from the same biosynthetic gene cluster, and a single compound should not be regarded as the sole product of a biosynthetic gene cluster.

4. *gre* biosynthetic gene cluster

Novel angucyclines - grecocyclines A, B and C have been isolated from *S.* sp. Acta1362 by Paululat et al [109]. Grecocyclines have unique structural moieties such as a dissacharide side chain, an additional amino sugar at the C-5 position and a thiol group. Grecocyclines are biologically active compounds – grecocycline A shows cytotoxic activity and grecocycline B inhibits protein tyrosine phosphatase 1B [109].

4.1. Analysis of the genes/enzymes putatively involved in the angucycline biosynthesis and modification

The genome scanning of the strain Acta1362 revealed a gene cluster (*gre*) responsible for the grecocycline biosynthesis that is approximately 36 kb long and comprises 32 ORFs. The genes that are required for the biosynthesis of the glycosylated angucycline have been identified within the *gre* cluster. The genes *greA1*, *greA2* and *greA3* encode KS, CLF and ACP, respectively. Three deduced proteins from the *gre* cluster, GreA4, GreA5 and GreA6, show similarity to cyclases and ketoreductase involved in UWM6 formation. GreA4 (82% identical with LndF) and GreA5 (79% identical with UrdL) are putative cyclases and GreA6 encodes a putative ketoreductase (87% similarity to PgaD). The cluster also harbors genes encoding oxygenases and reductases that are probably involved in the biosynthesis of the grecocycline aglycon. Two deduced proteins of the *gre* cluster GreE and GreM2 have high similarity to oxygenases involved in angucycline formation [71]. GreE shares 76% identity with PgaE monooxygenase from the rubromycin gene cluster [110], whereas GreM2 is identical to bifunctional oxygenase-reductases – PgaM, SaqM, Sim7 [19, 39, 110]. Further genes involved in modifying steps on the polyketide are *greV* and *greO* (putatively involved

in reduction steps) (Figure 26). GreK shows high similarity to the putative oxygenase SaqE [19]. The grecocycline gene cluster harbors additionally two other oxygenases – *greL* and *greD*. Blast search didn't reveal any conserved domains in GreL, it has 37-40% of identity with monooxygynases from different species without precise function identified. Deduced amino acid sequence of GreD shows high similarity to proteins from DsbA-FrnE family, which are presumed to be a thiol oxidoreductases and are involved in biosynthesis of frenolicin and nanaomycin [111]. Intrestingly, greD is the only one gene that is transcribed on the antisense strand.

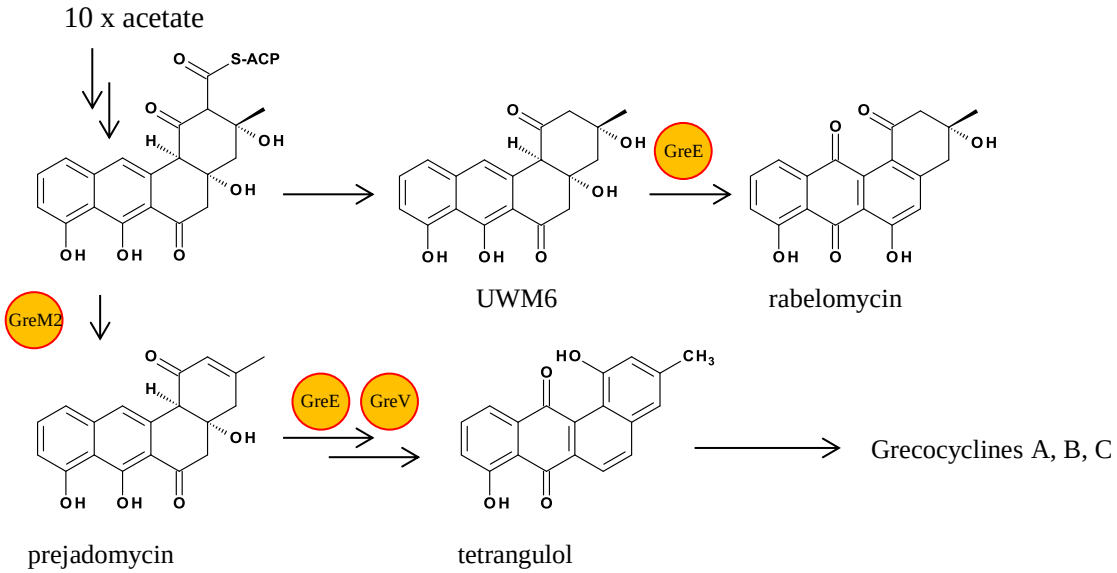


Figure 26. Proposed pathway for the biosynthesis of the polyketide part of grecocycline

One of the unique features of the grecocycline B is a thiol group located at the C-6a position. The introduction of sulfur into the angucycline core was reported for WS009A and B, a potent endothelin receptor antagonists from *S.* sp. no. 89009. In this compound, C6a was substituted with N-acetylcysteine moiety [112], however no further studies on the biosynthetical origin of this group have been done. In the gre cluster, thioesterase-encoding gene (*greTH*), which might be involved in formation of the thiol group in grecocycline B has been identified.

4.2. Analysis of the genes/enzymes putatively involved in biosynthesis of deoxysugars

Two different deoxy sugars are attached to the angucycline: L-rhodinose and L-tolyposamine. Eleven of thirty-two genes are involved in the modification and transfer of the sugars to the polyketide moiety. First, D-glucose-1-phosphate is converted to NDP-D-glucose (GreG), then

4,6- and 2,3-dehydratation (GreH and GreS) and 3-ketoreduction (GreT) reactions yield NDP-4-keto-2,6-dideoxy-D-glucose intermediate. The NDP-rhodinose is formed by subsequent 3,4-dehydratation (GreQ), 3,5-epimerisation (GreZ1) and 4-ketoreduction (GreZ3). GreJ encodes NDP-hexose-4-aminotransferase that transfer an amino group to the hexose to form L-tolyposamine (Figure 27). Grecocyclines contain three sugar units and equal number of glycosyltransferase encoding genes has been identified in the *gre* cluster – *greGT1*, *greGT2* and *greGT4*. All of them have significant similarity to different glycosyltransferases from saquayamycin, landomycin and urdamycin gene clusters [17, 19, 113]. *greGT2* encodes a protein that is homologous to SaqGT5 and LanGT2 [17, 19]. SaqGT5 and LanGT2 where proven to catalyze first glycosylation step during biosynthesis of saquayamycin Z and landomycin A, respectively [19, 114]. LanGT2 transfers D-olivose to O-8 of the angucycline, in contrast SaqGT5 is involved in attachment of D-olivose to C-9. In grecocyclines, L-rhodinose is attached to C-9 of aglycon, probably this reaction is mediated by GreGT2. GreGT1 may be responsible for introduction of the second L-rhodinose unit, as it is highly similar to SaqGT3, SaqGT4, LanGT1 and other O-glycosyltransferases [17, 19, 115]. SaqGT3 and LanGT1 are iteratively acting glycosyltransferases, they attach two D-olivose units to the polysaccharide chains, whereas SaqGT4 is believed to transfer two L-rhodinoses to the pentasaccharide chain during biosynthesis of saquayamicn Z [19]. Deduced amino sequence of GreGT4 share 54% with Lcz3 from lactanomycin and 50% identity with UrdGT1a from urdamycin biosynthetic pathways [113, 116] and presumably is transferring the aminosugar to the polyketide. As mentioned above *greL* encodes monooxygenase that doesn't have close homologs in other angucycline gene clusters, perhaps GreL is oxidizing aglycon in C-5 position prior to sugar attachment.

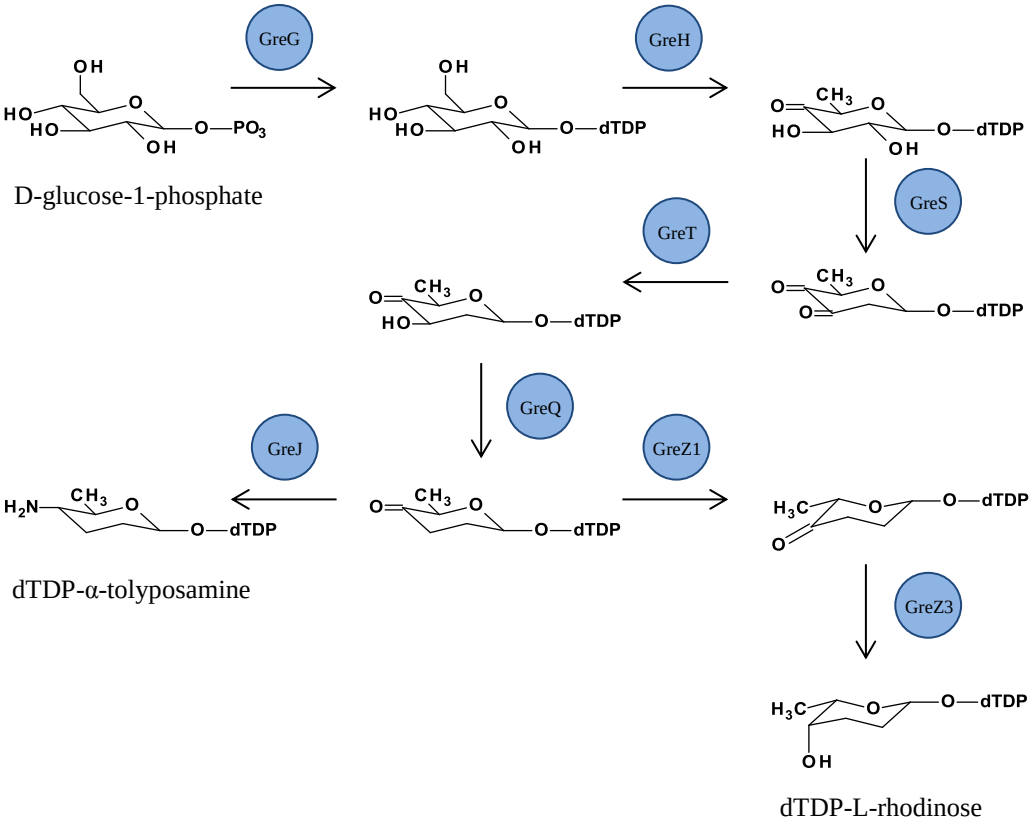


Figure 27. Proposed pathway for the biosynthesis of D-rhodinose and α-tolyposamine

4.3. Genes of regulation and self-resistance

Two regulatory – *greR1*, *greR2* and three transporter genes – *greEx*, *greEx2* and *greEx3* have been identified in the *gre* cluster. *greR1* is located on the 5' terminus of the gene cluster, and encodes protein with similarity to transcriptional regulators from the OmpR family [117]. Deduced amino acid sequence of GreR1 is very similar to JadR1 and LndI. Disruption of *lndI* caused complete loss of landomycin E production in *S. globisporus*, while its overexpression led to an increase in the antibiotic production [118, 119]. LndI is an autoregulator as it can bind to its own promoter region as well as to the promoters of the *lnd* structural genes [120]. Presumably, GreR1 plays the same role in the grecocycline biosynthesis. GreR2 is homologous to HxlR transcriptional regulators. GreEx, GreEx2 and GreEx3 are similar to MFS transporters and probably are transporting secondary metabolites through the cell membrane.

4.4. Summary

As mention above, presence of the aminosugar and a thiol group are unique features of grecocyclines. Incorporated into the natural product, sugars may sufficiently influence pharmacokinetic properties of the metabolite. For instance, L-desosamine of erythromycin is

essential for its binding with a target – 50S RNA [121]. Typically secondary metabolires are glycosylated with deoxysugars like D-olivose, L-rhodinose, L-mycarose and L-digitoxose [102]. α-tolyposamine, attached to the grecocycline aglycon, is a rare structural moiety, previously reported only for BU-4514N from *Microtetraspora* sp. [122]. It may become a unique building block for the derivatization of natural products via glycosylation. Furthermore, grecocyclines are the first angucyclines carrying the deoxysugar unit in the C-5 position of the angucyclic core. Therefore, glycosyltransferases that are attaching sugar moieties into C-5 position may have a great potential for natural product combinatorial biosynthesis.

During the evolution microbial secondary metabolites have been optimized for the biological functions useful to the producing organism. On the other hand, they have not necessarily been optimized for desirable drug properties and this can be improved by structural modifications. Combinatorial biosynthesis has been accepted as a useful tool to increase chemical functionality of natural products. Glycosylation, oxidation and halogenation are only few examples of the "decorating" reactions suitable for genetic engineering of polyketides. Enzymes encoded within the grecocycline biosynthetic pathway (GreTh, GreGT4, GreL) are a promising addition to the existing toolbox used for combinatorial biosynthesis. Their employment opens a possibility for generation of novel derivatives of known active polyketides, perhaps, with improved biological properties.

IV. REFERENCES

1. Hertweck, C., *Decoding and reprogramming complex polyketide assembly lines: prospects for synthetic biology.* Trends Biochem Sci, 2015.
2. Butler, M.S., *Natural products to drugs: natural product-derived compounds in clinical trials.* Nat Prod Rep, 2008. **25**(3): p. 475-516.
3. Hertweck, C., *The biosynthetic logic of polyketide diversity.* Angew Chem Int Ed Engl, 2009. **48**(26): p. 4688-716.
4. Staunton, J. and K.J. Weissman, *Polyketide biosynthesis: a millennium review.* Nat Prod Rep, 2001. **18**(4): p. 380-416.
5. Rudd, B.A. and D.A. Hopwood, *Genetics of actinorhodin biosynthesis by Streptomyces coelicolor A3(2).* J Gen Microbiol, 1979. **114**(1): p. 35-43.
6. Malpartida, F. and D.A. Hopwood, *Molecular cloning of the whole biosynthetic pathway of a Streptomyces antibiotic and its expression in a heterologous host.* Nature, 1984. **309**(5967): p. 462-4.
7. O'Brien, R.V., et al., *Computational identification and analysis of orphan assembly-line polyketide synthases.* J Antibiot (Tokyo), 2014. **67**(1): p. 89-97.
8. Cortes, J., et al., *An unusually large multifunctional polypeptide in the erythromycin-producing polyketide synthase of Saccharopolyspora erythraea.* Nature, 1990. **348**(6297): p. 176-8.
9. Bevitt, D.J., et al., *6-Deoxyerythronolide-B synthase 2 from Saccharopolyspora erythraea. Cloning of the structural gene, sequence analysis and inferred domain structure of the multifunctional enzyme.* Eur J Biochem, 1992. **204**(1): p. 39-49.
10. Tuan, J.S., et al., *Cloning of genes involved in erythromycin biosynthesis from Saccharopolyspora erythraea using a novel actinomycete-Escherichia coli cosmid.* Gene, 1990. **90**(1): p. 21-9.
11. Donadio, S., et al., *Modular organization of genes required for complex polyketide biosynthesis.* Science, 1991. **252**(5006): p. 675-9.
12. Schwecke, T., et al., *The biosynthetic gene cluster for the polyketide immunosuppressant rapamycin.* Proc Natl Acad Sci U S A, 1995. **92**(17): p. 7839-43.
13. Shen, B., *Polyketide biosynthesis beyond the type I, II and III polyketide synthase paradigms.* Curr Opin Chem Biol, 2003. **7**(2): p. 285-95.
14. Gaisser, S., et al., *Cloning of an avilamycin biosynthetic gene cluster from Streptomyces viridochromogenes Tu57.* J Bacteriol, 1997. **179**(20): p. 6271-8.
15. Hertweck, C., et al., *Type II polyketide synthases: gaining a deeper insight into enzymatic teamwork.* Nat Prod Rep, 2007. **24**(1): p. 162-90.
16. Han, L., et al., *Cloning and characterization of polyketide synthase genes for jadomycin B biosynthesis in Streptomyces venezuelae ISP5230.* Microbiology, 1994. **140 (Pt 12)**: p. 3379-89.
17. Westrich, L., et al., *Cloning and characterization of a gene cluster from Streptomyces cyanogenus S136 probably involved in landomycin biosynthesis.* FEMS Microbiol Lett, 1999. **170**(2): p. 381-7.
18. Decker, H. and S. Haag, *Cloning and characterization of a polyketide synthase gene from Streptomyces fradiae Tu2717, which carries the genes for biosynthesis of the angucycline antibiotic urdamycin A and a gene probably involved in its oxygenation.* J Bacteriol, 1995. **177**(21): p. 6126-36.
19. Erb, A., et al., *Cloning and sequencing of the biosynthetic gene cluster for saquayamycin Z and galtamycin B and the elucidation of the assembly of their saccharide chains.* Chembiochem, 2009. **10**(8): p. 1392-401.

20. Bisang, C., et al., *A chain initiation factor common to both modular and aromatic polyketide synthases.* Nature, 1999. **401**(6752): p. 502-5.
21. Tang, Y., S.C. Tsai, and C. Khosla, *Polyketide chain length control by chain length factor.* J Am Chem Soc, 2003. **125**(42): p. 12708-9.
22. Das, A. and C. Khosla, *In vivo and in vitro analysis of the hedamycin polyketide synthase.* Chem Biol, 2009. **16**(11): p. 1197-207.
23. Nicholson, T.P., et al., *First in vitro directed biosynthesis of new compounds by a minimal type II polyketide synthase: evidence for the mechanism of chain length determination.* Chem Commun (Camb), 2003(6): p. 686-7.
24. Shen, Y., et al., *Ectopic expression of the minimal whiE polyketide synthase generates a library of aromatic polyketides of diverse sizes and shapes.* Proc Natl Acad Sci U S A, 1999. **96**(7): p. 3622-7.
25. Moore, B.S. and C. Hertweck, *Biosynthesis and attachment of novel bacterial polyketide synthase starter units.* Nat Prod Rep, 2002. **19**(1): p. 70-99.
26. Xiang, L. and B.S. Moore, *Characterization of benzoyl coenzyme A biosynthesis genes in the enterocin-producing bacterium "Streptomyces maritimus".* J Bacteriol, 2003. **185**(2): p. 399-404.
27. Piel, J., et al., *Cloning, sequencing and analysis of the enterocin biosynthesis gene cluster from the marine isolate 'Streptomyces maritimus': evidence for the derailment of an aromatic polyketide synthase.* Chem Biol, 2000. **7**(12): p. 943-55.
28. Grimm, A., et al., *Characterization of the Streptomyces peucetius ATCC 29050 genes encoding doxorubicin polyketide synthase.* Gene, 1994. **151**(1-2): p. 1-10.
29. Ye, J., et al., *Isolation and sequence analysis of polyketide synthase genes from the daunomycin-producing Streptomyces sp. strain C5.* J Bacteriol, 1994. **176**(20): p. 6270-80.
30. Meurer, G. and C.R. Hutchinson, *Functional analysis of putative beta-ketoacyl:acyl carrier protein synthase and acyltransferase active site motifs in a type II polyketide synthase of Streptomyces glaucescens.* J Bacteriol, 1995. **177**(2): p. 477-81.
31. Rajgarhia, V.B. and W.R. Strohl, *Minimal Streptomyces sp. strain C5 daunorubicin polyketide biosynthesis genes required for aklanonic acid biosynthesis.* J Bacteriol, 1997. **179**(8): p. 2690-6.
32. Rebets, Y., et al., *Insights into the pamamycin biosynthesis.* Angew Chem Int Ed Engl, 2015. **54**(7): p. 2280-4.
33. Walczak, R.J., et al., *Nonactin biosynthesis: the potential nonactin biosynthesis gene cluster contains type II polyketide synthase-like genes.* FEMS Microbiol Lett, 2000. **183**(1): p. 171-5.
34. Kwon, H.J., et al., *Cloning and heterologous expression of the macrotetrolide biosynthetic gene cluster revealed a novel polyketide synthase that lacks an acyl carrier protein.* J Am Chem Soc, 2001. **123**(14): p. 3385-6.
35. Nelson, M.E. and N.D. Priestley, *Nonactin biosynthesis: the initial committed step is the condensation of acetate (malonate) and succinate.* J Am Chem Soc, 2002. **124**(12): p. 2894-902.
36. Schimana, J., et al., *Simocyclinones, novel cytostatic angucyclinone antibiotics produced by Streptomyces antibioticus Tu 6040. I. Taxonomy, fermentation, isolation and biological activities.* J Antibiot (Tokyo), 2000. **53**(8): p. 779-87.
37. Holzenkampfer, M., et al., *Simocyclinones, novel cytostatic angucyclinone antibiotics produced by Streptomyces antibioticus Tu 6040 II. Structure elucidation and biosynthesis.* J Antibiot (Tokyo), 2002. **55**(3): p. 301-7.
38. Trefzer, A., et al., *Biosynthetic gene cluster of simocyclinone, a natural multihybrid antibiotic.* Antimicrob Agents Chemother, 2002. **46**(5): p. 1174-82.

39. Galm, U., et al., *Cloning and analysis of the simocyclinone biosynthetic gene cluster of Streptomyces antibioticus Tu 6040.* Arch Microbiol, 2002. **178**(2): p. 102-14.
40. Chen, H. and C.T. Walsh, *Coumarin formation in novobiocin biosynthesis: beta-hydroxylation of the aminoacyl enzyme tyrosyl-S-NovH by a cytochrome P450 NovI.* Chem Biol, 2001. **8**(4): p. 301-12.
41. Luft, T., et al., *Overexpression, purification and characterization of SimL, an amide synthetase involved in simocyclinone biosynthesis.* Arch Microbiol, 2005. **183**(4): p. 277-85.
42. Pacholec, M., et al., *Characterization of the aminocoumarin ligase SimL from the simocyclinone pathway and tandem incubation with NovM,P,N from the novobiocin pathway.* Biochemistry, 2005. **44**(12): p. 4949-56.
43. Boll, B., T. Taubitz, and L. Heide, *Role of MbtH-like proteins in the adenylation of tyrosine during aminocoumarin and vancomycin biosynthesis.* J Biol Chem, 2011. **286**(42): p. 36281-90.
44. Holzenkampfer, M. and A. Zeeck, *Biosynthesis of simocyclinone D8 in an 18O2-rich atmosphere.* J Antibiot (Tokyo), 2002. **55**(3): p. 341-2.
45. Horbal, L., et al., *SimReg1 is a master switch for biosynthesis and export of simocyclinone D8 and its precursors.* AMB Express, 2012. **2**(1): p. 1.
46. Le, T.B., et al., *Coupling of the biosynthesis and export of the DNA gyrase inhibitor simocyclinone in Streptomyces antibioticus.* Mol Microbiol, 2009. **72**(6): p. 1462-74.
47. Le, T.B., et al., *Structures of the TetR-like simocyclinone efflux pump repressor, SimR, and the mechanism of ligand-mediated derepression.* J Mol Biol, 2011. **408**(1): p. 40-56.
48. Le, T.B., et al., *Crystallization and preliminary X-ray analysis of the TetR-like efflux pump regulator SimR.* Acta Crystallogr Sect F Struct Biol Cryst Commun, 2011. **67**(Pt 3): p. 307-9.
49. Le, T.B., et al., *The crystal structure of the TetR family transcriptional repressor SimR bound to DNA and the role of a flexible N-terminal extension in minor groove binding.* Nucleic Acids Res, 2011. **39**(21): p. 9433-47.
50. Richter, S.N., et al., *Simocyclinone D8 turns on against Gram-negative bacteria in a clinical setting.* Bioorg Med Chem Lett, 2010. **20**(3): p. 1202-4.
51. Vos, S.M., et al., *All tangled up: how cells direct, manage and exploit topoisomerase function.* Nat Rev Mol Cell Biol, 2011. **12**(12): p. 827-41.
52. Collin, F., S. Karkare, and A. Maxwell, *Exploiting bacterial DNA gyrase as a drug target: current state and perspectives.* Appl Microbiol Biotechnol, 2011. **92**(3): p. 479-97.
53. Heide, L., *New aminocoumarin antibiotics as gyrase inhibitors.* Int J Med Microbiol, 2014. **304**(1): p. 31-6.
54. Edwards, M.J., et al., *A crystal structure of the bifunctional antibiotic simocyclinone D8, bound to DNA gyrase.* Science, 2009. **326**(5958): p. 1415-8.
55. Flatman, R.H., et al., *Simocyclinone D8, an inhibitor of DNA gyrase with a novel mode of action.* Antimicrob Agents Chemother, 2005. **49**(3): p. 1093-100.
56. Hearnshaw, S.J., et al., *A new crystal structure of the bifunctional antibiotic simocyclinone D8 bound to DNA gyrase gives fresh insight into the mechanism of inhibition.* J Mol Biol, 2014. **426**(10): p. 2023-33.
57. Sissi, C., et al., *Mapping simocyclinone D8 interaction with DNA gyrase: evidence for a new binding site on GyrB.* Antimicrob Agents Chemother, 2010. **54**(1): p. 213-20.
58. Oppegard, L.M., et al., *Inhibition of human topoisomerases I and II by simocyclinone D8.* J Nat Prod, 2012. **75**(8): p. 1485-9.
59. Doroghazi, J.R., et al., *A roadmap for natural product discovery based on large-scale genomics and metabolomics.* Nat Chem Biol, 2014. **10**(11): p. 963-8.

60. Gomez-Escribano, J.P. and M.J. Bibb, *Streptomyces coelicolor as an expression host for heterologous gene clusters.* Methods Enzymol, 2012. **517**: p. 279-300.
61. Green, M.R.S., J., *Molecular Cloning: A Laboratory Manual* fourth ed. Vol. 1. 2012: John Inglis.
62. Kouprina, N. and V. Larionov, *Selective isolation of genomic loci from complex genomes by transformation-associated recombination cloning in the yeast Saccharomyces cerevisiae.* Nat Protoc, 2008. **3**(3): p. 371-7.
63. Yamanaka, K., et al., *Direct cloning and refactoring of a silent lipopeptide biosynthetic gene cluster yields the antibiotic taromycin A.* Proc Natl Acad Sci U S A, 2014. **111**(5): p. 1957-62.
64. Kim, J.H., et al., *Cloning large natural product gene clusters from the environment: piecing environmental DNA gene clusters back together with TAR.* Biopolymers, 2010. **93**(9): p. 833-44.
65. Larionov, V., et al., *Highly selective isolation of human DNAs from rodent-human hybrid cells as circular yeast artificial chromosomes by transformation-associated recombination cloning.* Proc Natl Acad Sci U S A, 1996. **93**(24): p. 13925-30.
66. Gibson, D.G., et al., *Creation of a bacterial cell controlled by a chemically synthesized genome.* Science, 2010. **329**(5987): p. 52-6.
67. Ross, A.C., et al., *Targeted Capture and Heterologous Expression of the Pseudoalteromonas Alterochromide Gene Cluster in Escherichia coli Represents a Promising Natural Product Exploratory Platform.* ACS Synth Biol, 2014.
68. Luzhetskyy, A., et al., *IncP plasmids are most effective in mediating conjugation between Escherichia coli and streptomycetes.* Genetika, 2006. **42**(5): p. 595-601.
69. Green, M.R. and J. Sambrook, *Molecular cloning : a laboratory manual*. 4th ed2012, Cold Spring Harbor, N.Y.: Cold Spring Harbor Laboratory Press.
70. *Comprehensive Natural Products II: Chemistry and Biology*. 1 ed. Vol. 1. 2010.
71. Patrikainen, P., et al., *Tailoring enzymes involved in the biosynthesis of angucyclines contain latent context-dependent catalytic activities.* Chem Biol, 2012. **19**(5): p. 647-55.
72. Pahari, P., et al., *Enzymatic total synthesis of defucogilvocarcin M and its implications for gilvocarcin biosynthesis.* Angew Chem Int Ed Engl, 2012. **51**(5): p. 1216-20.
73. Faust, B., et al., *Two new tailoring enzymes, a glycosyltransferase and an oxygenase, involved in biosynthesis of the angucycline antibiotic urdamycin A in Streptomyces fradiae Tu2717*. Microbiology, 2000. **146 (Pt 1)**: p. 147-54.
74. Kallio, P., et al., *A nested gene in Streptomyces bacteria encodes a protein involved in quaternary complex formation.* J Mol Biol, 2008. **375**(5): p. 1212-21.
75. Kallio, P., et al., *Flavoprotein hydroxylase PgaE catalyzes two consecutive oxygen-dependent tailoring reactions in angucycline biosynthesis.* Biochemistry, 2011. **50**(24): p. 5535-43.
76. Walsh, C.T. and T.A. Wencewicz, *Flavoenzymes: versatile catalysts in biosynthetic pathways.* Nat Prod Rep, 2013. **30**(1): p. 175-200.
77. Julien, B., et al., *Isolation and characterization of the epothilone biosynthetic gene cluster from Sorangium cellulosum.* Gene, 2000. **249**(1-2): p. 153-60.
78. Bililign, T., et al., *The hedamycin locus implicates a novel aromatic PKS priming mechanism.* Chem Biol, 2004. **11**(7): p. 959-69.
79. Mindnich, R.D. and T.M. Penning, *Aldo-keto reductase (AKR) superfamily: genomics and annotation.* Hum Genomics, 2009. **3**(4): p. 362-70.
80. Huang, X., H.M. Holden, and F.M. Raushel, *Channeling of substrates and intermediates in enzyme-catalyzed reactions.* Annu Rev Biochem, 2001. **70**: p. 149-80.
81. Miles, E.W., S. Rhee, and D.R. Davies, *The molecular basis of substrate channeling.* J Biol Chem, 1999. **274**(18): p. 12193-6.

82. Schafer, M., et al., *SimC7 Is a Novel NAD(P)H-Dependent Ketoreductase Essential for the Antibiotic Activity of the DNA Gyrase Inhibitor Simocyclinone.* J Mol Biol, 2015. **427**(12): p. 2192-204.
83. Giraud, M.F. and J.H. Naismith, *The rhamnose pathway.* Curr Opin Struct Biol, 2000. **10**(6): p. 687-96.
84. Ramos, A., et al., *Biosynthesis of elloramycin in Streptomyces olivaceus requires glycosylation by enzymes encoded outside the aglycon cluster.* Microbiology, 2008. **154**(Pt 3): p. 781-8.
85. Kaysser, L., et al., *Formation and attachment of the deoxysugar moiety and assembly of the gene cluster for caprazamycin biosynthesis.* Appl Environ Microbiol, 2010. **76**(12): p. 4008-18.
86. Madduri, K., C. Waldron, and D.J. Merlo, *Rhamnose biosynthesis pathway supplies precursors for primary and secondary metabolism in Saccharopolyspora spinosa.* Journal of Bacteriology, 2001. **183**(19): p. 5632-5638.
87. Luzhetskyy, A., et al., *Cloning and heterologous expression of the aranciamycin biosynthetic gene cluster revealed a new flexible glycosyltransferase.* Chembiochem, 2007. **8**(6): p. 599-602.
88. Steffensky, M., et al., *Identification of the novobiocin biosynthetic gene cluster of Streptomyces spheroides NCIB 11891.* Antimicrob Agents Chemother, 2000. **44**(5): p. 1214-22.
89. Bibb, M.J., et al., *Cloning, sequencing and deduced functions of a cluster of Streptomyces genes probably encoding biosynthesis of the polyketide antibiotic frenolicin.* Gene, 1994. **142**(1): p. 31-9.
90. Hutchinson, C.R., *Biosynthetic Studies of Daunorubicin and Tetracenomycin C.* Chem Rev, 1997. **97**(7): p. 2525-2536.
91. Gonzalez-Mellado, D., et al., *The role of beta-ketoacyl-acyl carrier protein synthase III in the condensation steps of fatty acid biosynthesis in sunflower.* Planta, 2010. **231**(6): p. 1277-89.
92. Bao, W., P.J. Sheldon, and C.R. Hutchinson, *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**(30): p. 9752-7.
93. Meadows, E.S. and C. Khosla, *In vitro reconstitution and analysis of the chain initiating enzymes of the R1128 polyketide synthase.* Biochemistry, 2001. **40**(49): p. 14855-61.
94. Xu, Z., A. Schenk, and C. Hertweck, *Molecular analysis of the benastatin biosynthetic pathway and genetic engineering of altered fatty acid-polyketide hybrids.* J Am Chem Soc, 2007. **129**(18): p. 6022-30.
95. Walker, M., *Strukturaufklärung und Biosynthese der Simocyclinone und anderer Sekundärmetabolite im Rahmen der Wirkstoffsuche bei endosymbiotischen und anderen Mikroorganismen sowie Röntgenstrukturanalyse ausgewählter Naturstoffe*, 1999.
96. Harrison, P.J. and T.D. Bugg, *Enzymology of the carotenoid cleavage dioxygenases: reaction mechanisms, inhibition and biochemical roles.* Arch Biochem Biophys, 2014. **544**: p. 105-11.
97. Bihlmaier, C., *Polyenantibiotika aus Streptomyceten : molekularbiologische Untersuchungen zur Biosynthese von Simocyclinon und alpha-Lipomycin*, 2005.
98. Wilkinson, S.P. and A. Grove, *Ligand-responsive transcriptional regulation by members of the MarR family of winged helix proteins.* Curr Issues Mol Biol, 2006. **8**(1): p. 51-62.

99. Medema, M.H., et al., *A systematic computational analysis of biosynthetic gene cluster evolution: lessons for engineering biosynthesis.* PLoS Comput Biol, 2014. **10**(12): p. e1004016.
100. Bode, H.B., et al., *Big effects from small changes: possible ways to explore nature's chemical diversity.* Chembiochem, 2002. **3**(7): p. 619-27.
101. Blin, K., et al., *antiSMASH 2.0--a versatile platform for genome mining of secondary metabolite producers.* Nucleic Acids Res, 2013. **41**(Web Server issue): p. W204-12.
102. Erb, A., et al., *A bacterial glycosyltransferase gene toolbox: generation and applications.* Phytochemistry, 2009. **70**(15-16): p. 1812-21.
103. Guo, Z.K., et al., *Angucyclines from an insect-derived actinobacterium Amycolatopsis sp. HCa1 and their cytotoxic activity.* Bioorg Med Chem Lett, 2012. **22**(24): p. 7490-3.
104. Rasmussen, R.R., et al., *Benzanthrins A and B, a new class of quinone antibiotics. II. Isolation, elucidation of structure and potential antitumor activity.* J Antibiot (Tokyo), 1986. **39**(11): p. 1515-26.
105. Ohmori, K., et al., *Concise total synthesis and structure assignment of TAN-1085.* Angew Chem Int Ed Engl, 2004. **43**(24): p. 3167-71.
106. Shigeru, N.K., N. Masao, T. Hiroshi, N. Nakamjima, H. Mioko, H. Shigemi, Y., *Antitumor substance j-124131 and method for preparing the same*, 2001.
107. Yoon, T.M., et al., *Talosins A and B: new isoflavonol glycosides with potent antifungal activity from Kitasatospora kifunensis MJM341. I. Taxonomy, fermentation, isolation, and biological activities.* J Antibiot (Tokyo), 2006. **59**(10): p. 633-9.
108. Karki, S., et al., *Cloning and in vitro characterization of dTDP-6-deoxy-L-talose biosynthetic genes from Kitasatospora kifunensis featuring the dTDP-6-deoxy-L-lyxo-4-hexulose reductase that synthesizes dTDP-6-deoxy-L-talose.* Carbohydr Res, 2010. **345**(13): p. 1958-62.
109. Paululat, T.K., A. Hausmann, H. Karagouni, A. D. Zinecker, H., Imhoff, J. F. Fiedler, H.-P., *Grecocyclines: New Angucyclines from Streptomyces sp. Acta 1362.* European Journal of Organic Chemistry, 2010. **2010**(12): p. 2344-2350.
110. Metsa-Ketela, M., K. Ylihonko, and P. Mantsala, *Partial activation of a silent angucycline-type gene cluster from a rubromycin beta producing Streptomyces sp. PGA64.* J Antibiot (Tokyo), 2004. **57**(8): p. 502-10.
111. Bibb, M.J., et al., *Cloning, Sequencing and Deduced Functions of a Cluster of Streptomyces Genes Probably Encoding Biosynthesis of the Polyketide Antibiotic Frenolicin.* Gene, 1994. **142**(1): p. 31-39.
112. Miyata, S., et al., *WS009 A and B, new endothelin receptor antagonists isolated from Streptomyces sp. no. 89009. I. Taxonomy, fermentation, isolation, physico-chemical properties and biological activities.* J Antibiot (Tokyo), 1992. **45**(7): p. 1029-40.
113. Trefzer, A., et al., *Function of glycosyltransferase genes involved in urdamycin A biosynthesis.* Chem Biol, 2000. **7**(2): p. 133-42.
114. Luzhetskyy, A., et al., *LanGT2 Catalyzes the First Glycosylation Step during landomycin A biosynthesis.* Chembiochem, 2005. **6**(8): p. 1406-10.
115. Luzhetskyy, A., et al., *Iteratively acting glycosyltransferases involved in the hexasaccharide biosynthesis of landomycin A.* Chem Biol, 2005. **12**(7): p. 725-9.
116. Zhang, X., et al., *Biosynthetic investigations of lactonamycin and lactonamycin z: cloning of the biosynthetic gene clusters and discovery of an unusual starter unit.* Antimicrob Agents Chemother, 2008. **52**(2): p. 574-85.
117. Itou, H. and I. Tanaka, *The OmpR-family of proteins: Insight into the tertiary structure and functions of two-component regulator proteins.* Journal of Biochemistry, 2001. **129**(3): p. 343-350.

118. Rebets, Y., et al., *Production of landomycins in Streptomyces globisporus 1912 and S cyanogenus S136 is regulated by genes encoding putative transcriptional activators.* FEMS Microbiol Lett, 2003. **222**(1): p. 149-53.
119. Gromyko, O., et al., *Generation of Streptomyces globisporus SMY622 strain with increased landomycin E production and it's initial characterization.* J Antibiot (Tokyo), 2004. **57**(6): p. 383-9.
120. Rebets, Y.V., et al., *Expression of the regulatory protein LndI for landomycin E production in Streptomyces globisporus 1912 is controlled by the availability of tRNA for the rare UUA codon.* FEMS Microbiol Lett, 2006. **256**(1): p. 30-7.
121. Katz, L. and G.W. Ashley, *Translation and protein synthesis: macrolides.* Chem Rev, 2005. **105**(2): p. 499-528.
122. Toda, S., et al., *A new neuritogenetic compound BU-4514N produced by Microtetraspora sp.* J Antibiot (Tokyo), 1993. **46**(6): p. 875-83.
123. Wang, L., et al., *Autoregulation of antibiotic biosynthesis by binding of the end product to an atypical response regulator.* Proc Natl Acad Sci U S A, 2009. **106**(21): p. 8617-22.
124. Kitani, S., et al., *Characterization of a regulatory gene, aveR, for the biosynthesis of avermectin in Streptomyces avermitilis.* Appl Microbiol Biotechnol, 2009. **82**(6): p. 1089-96.

V. APPENDIX

CURRICULUM VITAE

Name, Surname:	Oksana Bilyk
Date of birth:	2 July 1988
Place of birth:	Lviv, Ukraine
Nationality:	Ukrainian
E-mail:	o.protsiv@gmail.com

EDUCATION

2011 – present	PhD at Department of Pharmaceutical Biotechnology, University of Saarland, Saarbrücken, Germany
2009 – 2010	MSc at Department of Biochemistry, Ivan Franko National University of Lviv, Ukraine
2005 – 2009	BSc at Department of Biochemistry, Ivan Franko National University of Lviv, Ukaine
2002 – 2005	Lviv Physics and Mathematics Lyceum
1995 – 2002	Lviv Secondary School №2

SUPPORTING INFORMATION related to part II. RESULTS

SUPPLEMENTAL INFORMATION

New simocyclinones: surprising evolutionary and biosynthetic insights

Table of Contents

Chapter S1. Identification of new D-type simocyclinones in *K.* sp. and *S.* sp. NRRL B-24484

Information derived from the genome mining of the *K.* sp. and *S.* sp. NRRL B-24484 strains promised the production of simocyclinones similar to those produced by *S. antibioticus* Tü6040. Indeed, after HPLC-MS analysis of extracts from both strains, several peaks were identified with a UV spectra typical of D-type simocyclinones, as reported in the literature [5].

HR ESI-MS of SD9 (**2**) showed that the molecular ion m/z = 932.2176 $[M-H]^-$ (calculated mass: 932.2169 for $[M-H]^-$), which indicated that **2** had the molecular formula $C_{46}H_{44}NO_{18}Cl$. This formula and a molecular mass of M = 933 $gmol^{-1}$ were not previously reported for any known simocyclinone. Compared to SD8 (**1**), the molecular weight of **2** is two mass units higher, thus leading to the assumption that it has the same basic structure as D8, where the hydration of one double bond or reduction the of one carbonyl group has occurred at the corresponding alcohol. Structure elucidation allowed us to determine that D9 (**2**) contained the same chlorinated aminocoumarin, the same tetraene unit and the same sugar, but with a modified angucycline unit, as compared to SD8 (**1**) (Figures S2a – S2e, Tables S1a, S1b).

The angucycline part of **2** contains one carbonyl group, six methine groups (two aromatic, one double-bond and three aliphatic methines), one methylene group and nine quaternary carbons substituted with one methyl group, one epoxide, six hydroxy groups and one sugar. Ring A of the angucycline shows a double-bond in position C2-C3 substituted with a methyl group at C-3, neighboring a hydroxy methine at C-1 (δ_H = 4.63 / δ_C = 73.0) and a methylene group at C-4 assigned by COSY 1-H / 2-H and HMBC signals C-12b / 2-H, C-2 / 3-H_3, C-2 / 4-H_2, and 3-C / 4-H_2, C-4 / 2-H. This substitution pattern in ring A is described for SF2315B, which only differs from SD8 (**1**) by the presence of a proton in the C-12b position instead of a hydroxyl group. Thus, our NMR data are in good agreement with that reported for SF2315B [6]. Ring B contains a CH_2-CH-OH unit in the C-5-C6-position assigned by HMBC signals C-4 / 5-H_2, C-6a / 5-H_2, C-4a / 5-H_2 and C-12b / 5-H_2, the angular carbons bear hydroxyl groups based on their carbon chemical shifts ($\delta_{C\text{-}4a}$ = 70.8 and $\delta_{C\text{-}12b}$ = 74.3). Ring C includes the one carbonyl group at C-12 (δ_C = 196.0), as proven by HMBC signal C-12 / 11-H, and one hydroxy-methine at C-7 indicated by HMBC crosspeaks C-11a / 7-H and C-7a / 7-H. The C-6a-C12a-epoxide between rings B and C is proven by the chemical shifts δ_C = 64.2 (C-6a) and δ_C = 65.6 (C-12a). Ring D is an aromatic ring with two protons, 10-H and 11-H, which are in ortho position to each other (J = 8.2 Hz). Moreover, a phenol group is at C-8 (δ_C = 152.2) and the sugar unit is attached at C-9, which deduced from HMBC analysis. The sugar is determined as an α-sugar by the proton-proton coupling constant of $J_{1'\text{-}H,2'\text{-}Hax}$ = 11.2 Hz, which is a C-glycoside due to the carbon chemical shift of $\delta_{C\text{-}1'}$ = 70.6 ppm. The protons 3'-H and 4'-H, 5'-H must be in axial positions, as shown by the coupling constants $J_{2'\text{-}Hax,\ 3'\text{-}H}$ = 10.4 Hz, $J_{3'\text{-}H,\ 4'\text{-}H}$ = 9.5 Hz and $J_{4'\text{-}H,\ 5'\text{-}H}$ = 9.5 Hz, and supported by ROESY crosspeaks 1'-H / 3'-H, 1'-H / 5'-H, 3'-H / 5'-H. An acetyl group is attached at position C'-4, as proven by the HMBC signals C-7' / 4'-H. Therefore, the sugar is 4-acetyl-α-

olivose, as in SD8 (**1**). At position C-3' a tetraene chain is attached via an ester group C-1'', which was established from the HMBC correlation C-1'' / 3'-H and C-1'' / 2''-H and C-1'' / 3''-H. The configuration is proven to be *E* for the tetraene double bonds C-2'' / C-3'', C-4'' / C-5'', C-6'' / C-7'', C-8'' / C-9'' due to coupling constants $J \sim 15$ Hz (2''-H / 3''-H, 8''-H / 9''-H), $J \sim 13.5$ Hz (4''-H / 5''-H, 6''-H / 7''-H), and $J \sim 11.6$ Hz (3''-H / 4''-H, 5''-H / 6''-H, 7''-H / 8''-H) and ROESY signals (2''-H / 4''-H, 3''-H / 5''-H, 4''-H / 6''-H, 5''-H / 7''-H, 6''-H / 8''-H, 7''-H / 9''-H). An aminocoumarin moiety is attached via an amide group C-10'' as proven by HMBC signals C-10'' / 8''-H, C-10'' / 9''-H, C-10'' / NH and C-12'' / NH. The constitution is proven to be an aminocoumarin unit by comparison of NMR data to that of **1** which is supported by HMBC signals C-13a'' / 13''-OH, C-13a''/ 14''-H, C-14'' / 13''-OH, C-16'' / 14''-H, C-16'' / 15''-H, C-16'' / 16''-OH. The chloride is proven by the isotopic pattern of the molecular ion of **2** in the ESI mass spectrum. Finally, attached to the angucycline is the same substituent containing olivose, a tetraene chain and an aminocoumarin unit, as in SD8 (**1**), the new SD9 (**2**) only differs in ring A of the angucycline part of the structure

The second isolated compound SD10 (**3**) shows the same molecular mass, 915 $gmol^{-1}$, and formula, $C_{46}H_{42}NO_{17}Cl$, as simocyclinone D7, but the structure determination proves that **3** is a new compound. Structure elucidation of **3** was conducted based on NMR measurements (^1H, COSY, HSQC and HMBC) and again displays a modification of ring A in the angucycline moiety of the simocyclinone structure. Ring A consists of one endocyclic and one exocyclic double bond and a methylene group. The C-13 methylene group shows NMR signals $\delta_C = 114.4$ corresponding to $\delta_H =$ 4.98, 4.95, which is typical for an exocyclic double bond. This exocyclic double bond could be assigned to C-3-C-13 in conjugation with the second double bond C-1-C-2, which shows the protons H-1 and H-2 to be in *Z* configuration ($J = 10.3$ Hz). The position of the endocyclic double bond C-1-C-2 could be deduced from HMBC signals C-2 / H-1, C-3 / H-1, C-5 / H-1, C-12b / H-1 and C-1 / H-2, C-3 / H-2, C-13 / H-2. The aliphatic methylene group is determined to be C-4 via HMBC crosspeaks C-3 / 4-H_2, C-4 / 13-H_3, C-4 / 5-H_2 and C-5 / 4-H_2. As this is the only difference in the structure between SD9 (**2**) and SD10 (**3**), the final structure elucidation is displayed in the Supplemental Information (Tables S2a, S2b; Figures S3a-S3c).

The third isolated compound, SD11 (**4**), shows the same molecular mass M = 933 as D9 (**1**) in HR ESI MS m/z = 932.2103 $[M-H]^-$, which results in the same molecular formula, $C_{46}H_{44}NO_{18}Cl$. Detailed NMR analysis shows that only ring A is different in SD11 (**4**) (Tables S3a, S3b; Figures S4a-

S4c). Ring A consists of an endocyclic double bond at C1-C2 showing two doublets (J = 10.3 Hz) with typical double bond proton chemical shifts 6.24 (H-1) and 5.67 (H-2), a C-4 methylene group and a quaternary carbon C-4 substituted by a hydroxyl and a methyl group. The angucyclic carbons bear hydroxy groups deduced from carbon chemical shifts δ_C = 70.5 (C-4a) and δ_C = 69.1 (C-12b). The constitution of ring A is determined from HMBC signals C-4a/H-1, C-4/4-H_2, C-12b/H-2, C-2/4-H_2, C-3/H-1, C-3/ 4-H_2, C-13/2-H and C-13/ 4-H_2. The rest of the structure is proven by COSY and HMBC spectra and is shown to be the same as SD9 (**2**) and SD10 (**3**).

Figure S1: Fermentation and isolation of new simocyclinones from *K.* sp.

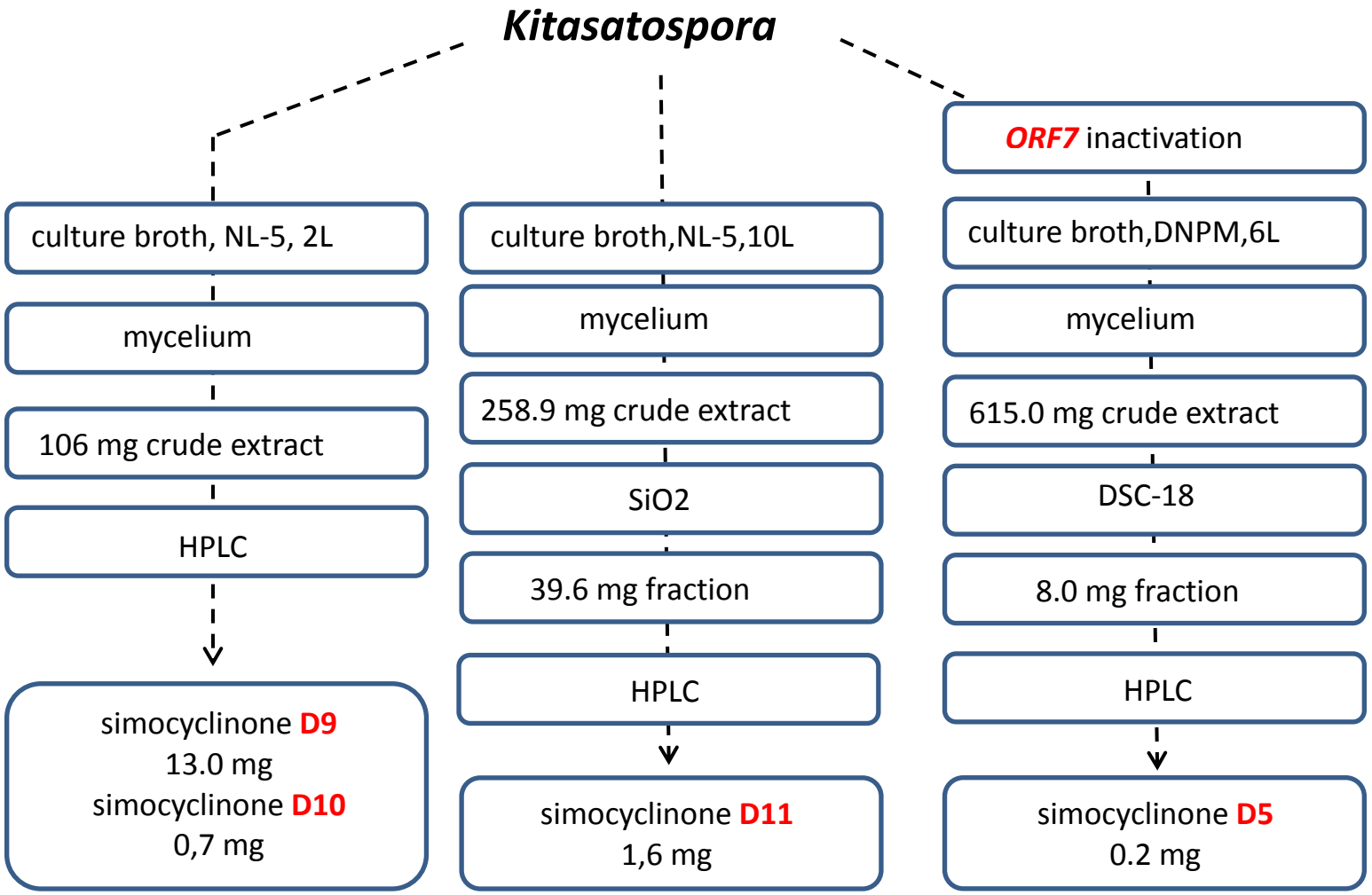


Table S1a

Pos.	δ_C (J Hz)	δ_H (J Hz)	COSY	TOCSY	HMBC	H2BC	ROESY
1	73.0	4.63 s	2-H, (4-H_a, 4-H_b,13-H_3)	2-H, (4-H_a, 4-H_b),13-H_3			2-H
2	123.2	5.18 s	1-H, (4-H_a, 4-H_b,13-H_3)	2-H, 4-H_a, 4-H_b,13-H_3	4-H_a, 4-H_b, 13-H_3	1-H, 13-H_3	1-H, 13-H_3
3	131.5				4-H_a, 4-H_b, 13-H_3		
4	42.0	2.17 d (17.8, H_a)	(1-H, 2-H), 4-H_b, 13-H_3	1-H, 2-H, 4-H_b, 13-H_3	2-H, 5-H_2, 13-H_3		4-H_b, 5-H_2, 6-H, (13-H_3),
		1.97 d (17.8, H_b)	(1-H, 2-H), 4-H_a, 13-H_3	1-H, 2-H, 4-H_b, 13-H_3			4-H_b, (13-H_3)
4a	70.8				4-H_b, 5-H_2		
5	38.5	1.93	6-H	6-H		6-H	6-H
6	63.2	4.33 br s	5-H_2	5-H_2	5-H_2	5-H_2	5-H_2, 4-H_a
6a	64.2				5-H_2		
7	65.3	5.70 d (6.3)	7-OH	7-OH		7-OH	7-OH, 8-OH
7-OH		7.00 d (6.3)	7-H	7-H			8-OH
7a	123.1				7-H, 8-OH, (10-H), 11-H		
8	152.2				(7-H), 8-OH, 10-H, (1'-H)		
8-OH		9.88					7-H
9	133.1				8-OH, 11-H, 1'-H, 2'-H_{ax}		
10	125.6	7.46 d (8.2)	11-H	11-H	11-H, 1'-H	11-H	11-H, (2'-H_{ax})
11	117.8	7.19 d (8.2)	10-H	10-H		10-H	10-H
11a	130.1				7-H, 10-H		
12	196.0				11-H		
12a	65.6						
12b	74.3				2-H, 5-H_2		
13	22.3	1.56 s	1-H, 2-H	1-H, 2-H, 4-H_a, 4-H_b	2-H	1-H, 2-H	2-H
1'	70.6	4.97 d (11.2)	2'-H_{ax}	2'-H_{ax}, 3'-H, (4-H), 6-H_3	10-H, 2'-H_{ax}, 5'-H	2'-H_{ax}	2'-H_{eq}, 3'-H, 5'-H
2'	36.6	eq: 2.42 dd (11.0, 5.3)	(1'-H), 2'-H_{ax}, 3'-H	2'-H_{ax}, 3'-H, 4'-H, 5'-H, 6'-H_3		1'-H, 3'-H	1'-H, 2'-H_{ax}, 3'-H
		ax: 1.57 ddd (11.2, 11.0, 10.4)	1'-H, 2'-H_{eq}	1'-H2'-H_{ax}, 3'-H, 4'-H, 6'-H_3			10-H, 1'-H, 2'-H_{eq}
3'	71.7	5.18 ddd (10.4, 9.5, 5.3)	2'-H_{ax}, 2'-H_{eq}, 4'-H	1'-H, 2'-H_{ax}, 2'-H_{eq}, 4'-H, 5'-H, 6'-H_3	1'-H, 2'-H_{ax}, 2'-H_{eq}	2'-H_{eq}, 4'-H	1'-H, 2'-H_{eq}, 5'-H
4'	74.2	4.77 dd (9.5, 9.5)	3'-H, 5'-H	2'-H_{ax}, 2'-H_{eq}, 3'-H, 5'-H, 6'-H_3	2'-H_{ax}, 2'-H_{eq}, 5'-H, 6'-H_3	3'-H, 5'-H	2'-Hax, (5'-H), 6'-H_3
5'	73.2	3.80 dd (9.5, 6.2)	4'-H, 6'-H_3	3'-H, 4'-H, 6'-H_3	1'-H, 6'-H_3	4'-H, 6'-H_3	1'-H, 3'-H, (4'-H), 6'-H_3
6'	17.8	1.18 d (6.2)	5'-H	1'-H, 2'-H_{ax}, 2'-H_{eq}, 3'-H, 4'-H, 5'-H	4'-H	5'-H	4'-H, 5'-H, (8'-H_3)
7'	169.7				4'-H, 8'-H		
8'	20.5	2.02 s			^{1}J		

*assignment not possible

2D signals in brackets are week

Table S1a: NMR data of simocyclinone D9 (**2**), angucycline and sugar part. (600 / 150MHz, DMSO-d_6, Temperature=35°C, solvent as internal reference)

Table S1b

Pos.	δ_C (J Hz)	δ_H (J Hz)	COSY	TOCSY	HMBC	H2BC	ROESY
1''	165.3				3'-H, 2''-H, 3''-H		
2''	121.5	6.00 d (15.5)	3''-H	3''-H, 4''-H, 5''-H, 6''-H 7''-H	4''-H	3''-H	4''-H
3''	144.5	7.28 dd (15.5, 11.5)	2''-H, 4''-H	2''-H, 4''-H, 5''-H, 6''-H, 7''-H, 9''-H	4''-H, 5''-H	2''-H, 4''-H	5''-H
4''	132.9	6.65 dd (13.3, 11.7)	3''-H, 5''-H	3''-H, 5''-H, 6''-H, 7''-H,8''-H	2''-H, 5''-H	3''-H, 5''-H	2''-H, 6''-H
5''	140.3	6.87 dd (13.3, 11.7)	4''-H	2''-H, 3''-H, 4''-H, 7''-H, 8''-H	3''-H, 7''-H	4''-H	3''-H
6''	139.1	6.87 dd (13.6, 11.7)	7''-H	2''-H, 3''-H, 4''-H, 7''-H, 8''-H	4''-H, 5''-H, 7''-H, 8''-H	5''-H, ^{1}J	4''-H
7''	134.4	6.72 dd (13.6, 11.7)	6''-H, 8''-H	2''-H, 3''-H, 5''-H, 6''-H, 8''-H, 9''-H	5''-H, 6''-H, 8''-H, 9''-H	6''-H, 8''-H	9''-H
8''	141.1	7.31 dd (15.2, 11.6)	7''-H, 9''-H	2''-H, 4''-H, 5''-H, 6''-H, 7''-H, 9''-H	6''-H, 7''-H	7''-H, 9''-H	6''-H
9''	124.9	6.63 d (15.5)	8''-H	2''-H, 3''-H, 4''-H, 5''-H, 6''-H, 7''-H, 8''-H	7''-H, 8''-H	8''-H	7''-H, NH
10''	166.2				8''-H, 9''-H, NH		
11''	158.8				NH		
12''	105.2						
13''	157.9				NH, 14''-H		
13''-OH		10.27					14''-H
13a''	115.4				13''-OH, 14''-H, 15''-H		
14''	119.6	7.25 s			13''-OH	15''-H, ^{1}J	13''-OH
15''	115.9	7.25 s			^{1}J	14''-H, ^{1}J	16''-OH
16''	151.0				13''-OH, 14''-H		
16''-OH		10.26					15''-H
17''	114.5				15''-H		
17a''	145.6				14''-H		
NH		9.84					9''-H
OH*		12.65, 9.84, 6.23, 5.32					

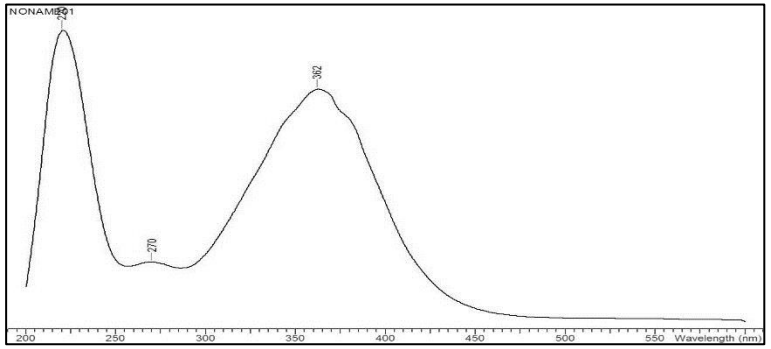
*assignment not possible

2D signals in brackets are week

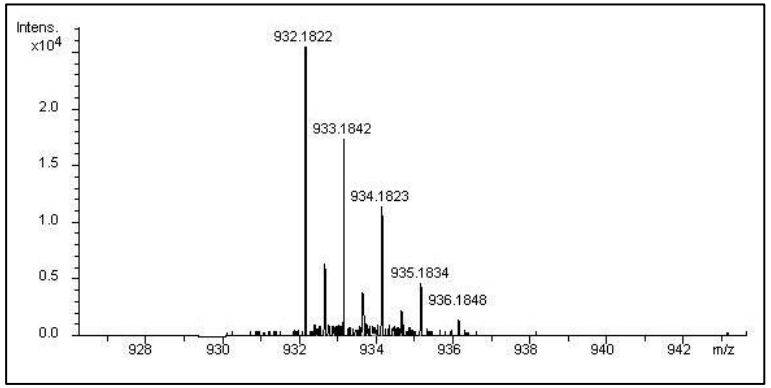
Table S1b: NMR data of simocyclinone D9 (**2**), angucycline and sugar part. (600 / 150MHz, DMSO-d_6, Temperature=35°C, solvent as internal reference)

Figure S2a, related to Figure 2: UV, MS and 1D NMR spectra of simocyclinone D9 (**2**). **(A)** UV spectrum of **2. (B)** ESI-MS (neg. mode) of **2. (C)** ^{1}H NMR spectrum of **2** (600 MHz, DMSO-d_6, 35°C). **(D)** ^{13}C NMR spectrum of **2** (150 MHz, DMSO-d_6, 35°C).

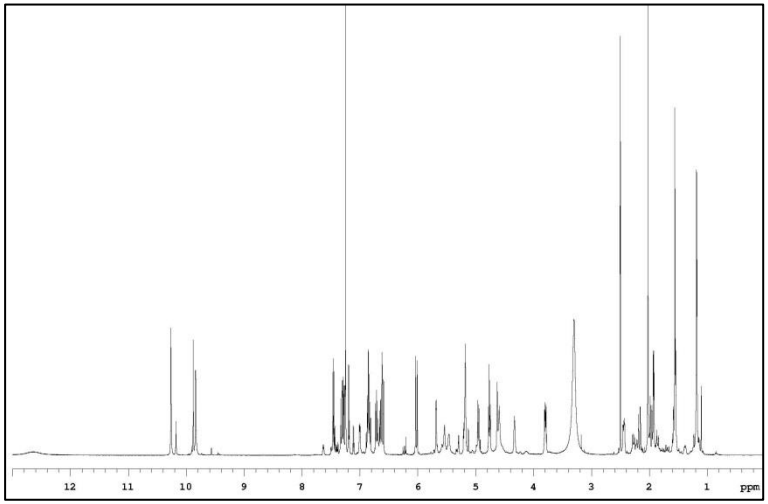
A



B



C



D

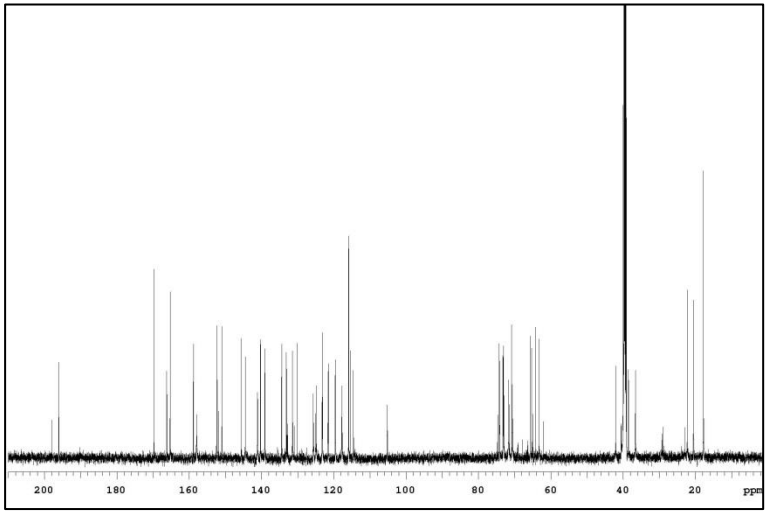
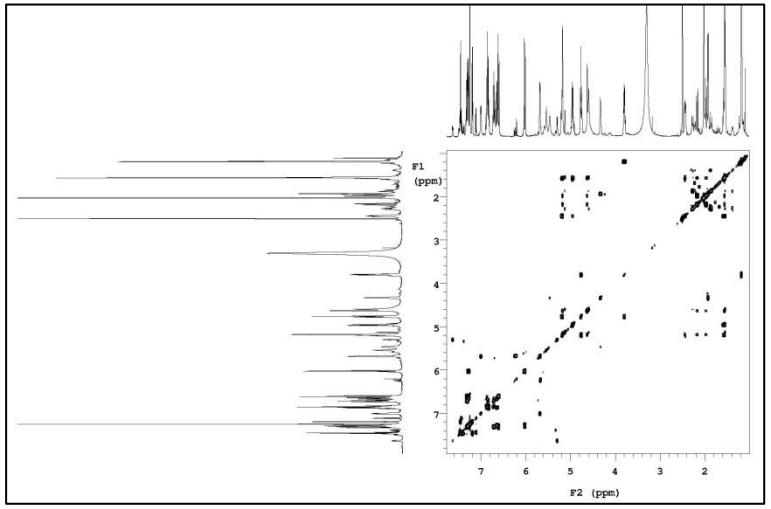
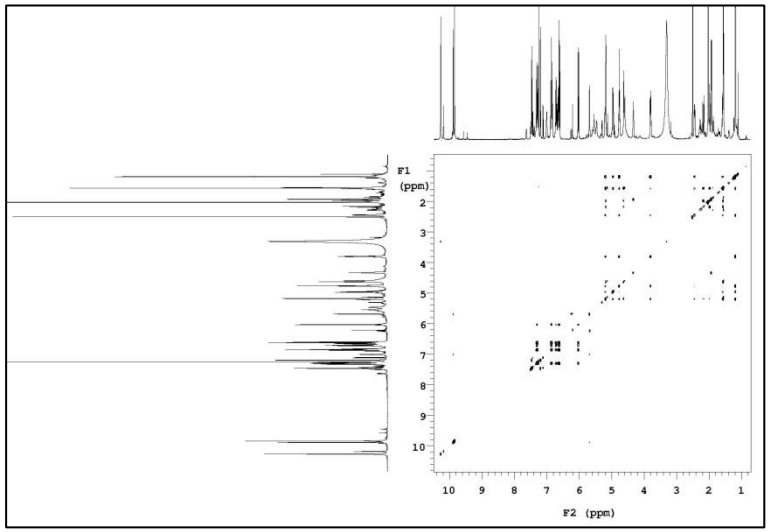


Figure S2b, related to Figure 2: 2D NMR spectra of simocyclinone D9 (**2**). **(A)** COSY spectrum of **2** (600 MHz, DMSO-d_6, 35 °C). **(B)** TOCSY spectrum of **2** (600 MHz, DMSO-d_6, 35°C). **(C)** ROESY spectrum of **2**. **(D)** HSQC spectrum of **2** (600 MHz, DMSO-d_6, 35°C).

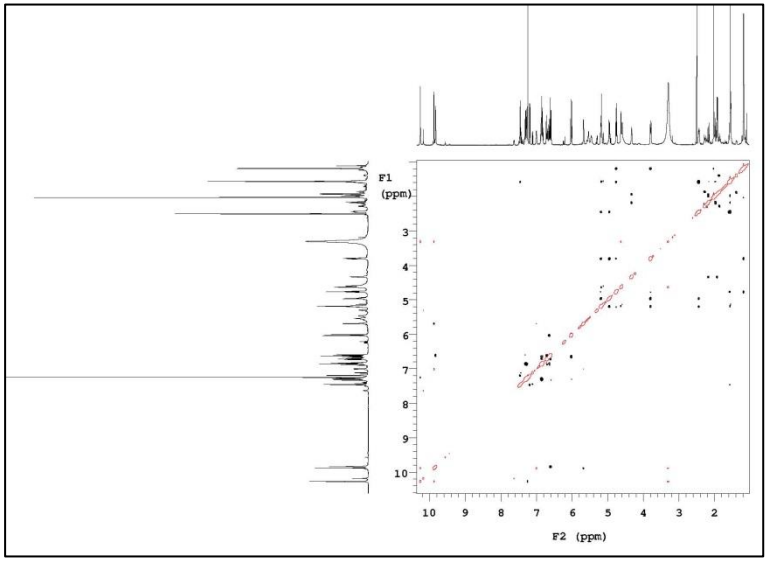
A



B



C



D

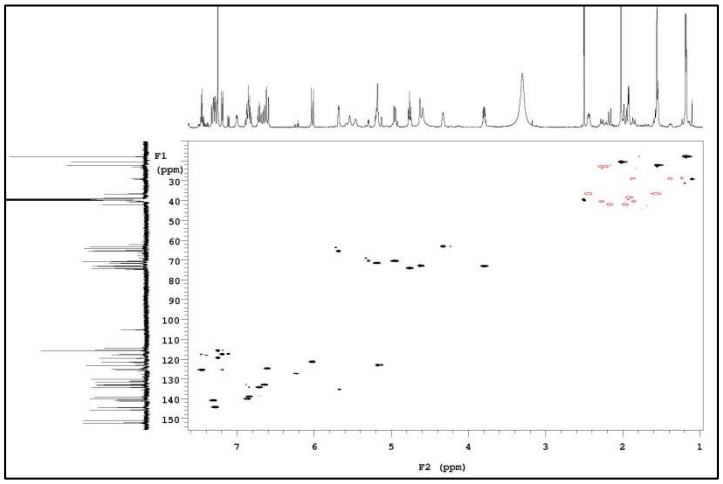
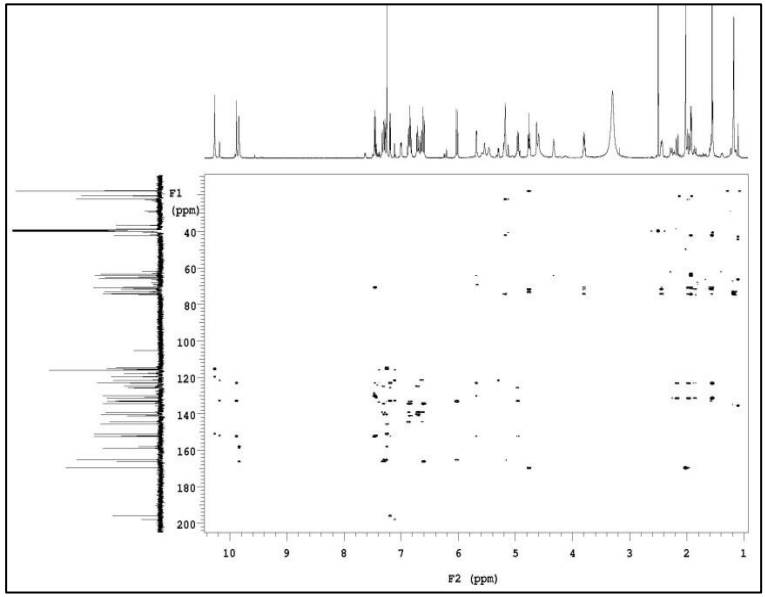


Figure S2c, related to Figure 2: 2D NMR spectra and structure elucidation of simocyclinone D9 (**2**) **. (A)** HMBC spectrum of **2** (600 MHz, DMSO-d_6, 35 °C). **(B)** H2BC spectrum of **2** (600 MHz, DMSO-d_6, 35°C). **(C)** 2D NMR correlations in aminocoumarin part of **2**. (D) 2D NMR correlations in tetraene part of **2**.

A



B

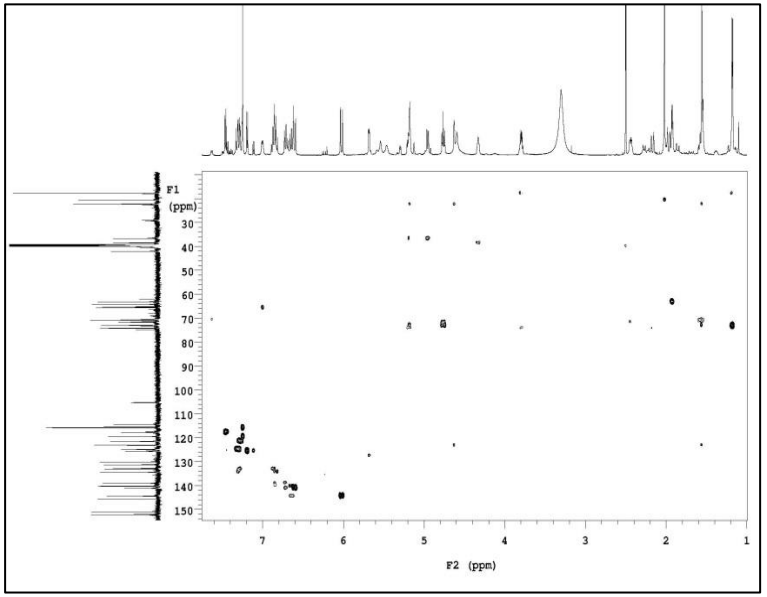
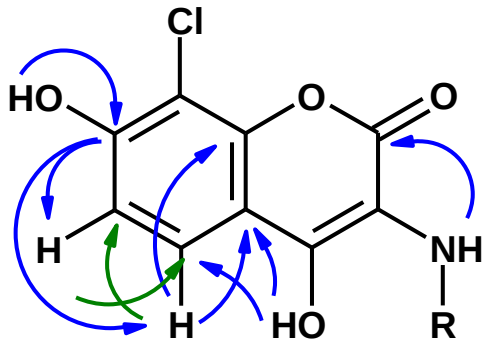
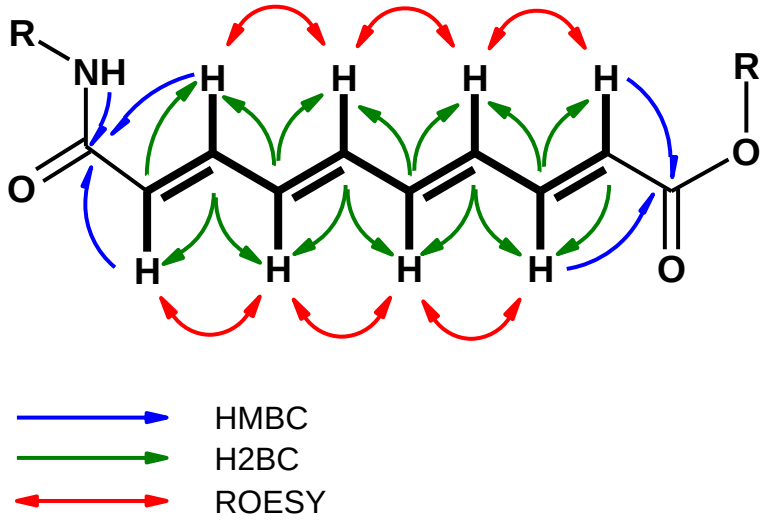


Figure S2d: Structure elucidation of simocyclinone D9 (**2**). **(A)** 2D NMR correlations in amiocoumarin part of **2. (B)** 2D NMR correlations in tetraene part of **2. (C)** 2D NMR correlations in sugar part of **2. (D)** 2D NMR correlations in angucycline part of **2.**

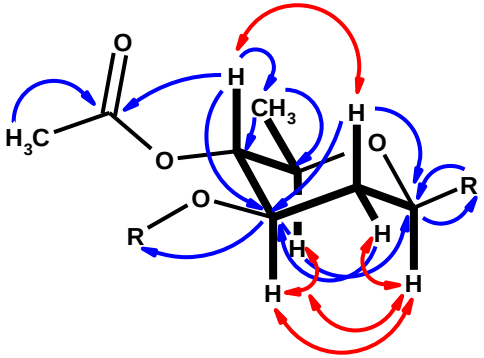
A



B



C



D

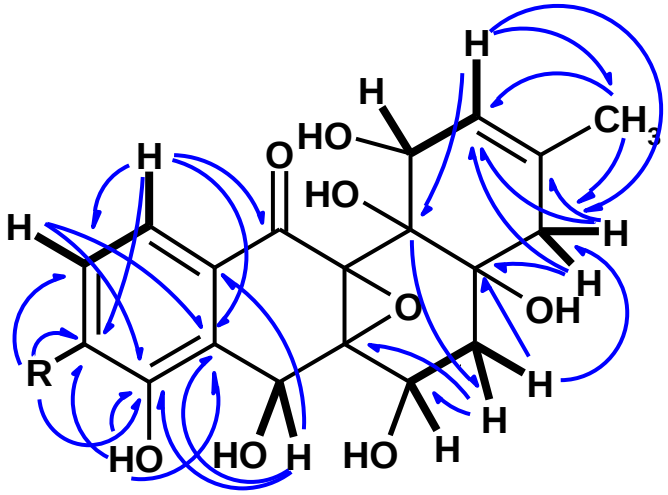


Figure S2e: Structure elucidation of simocyclinone D9 (**2**).

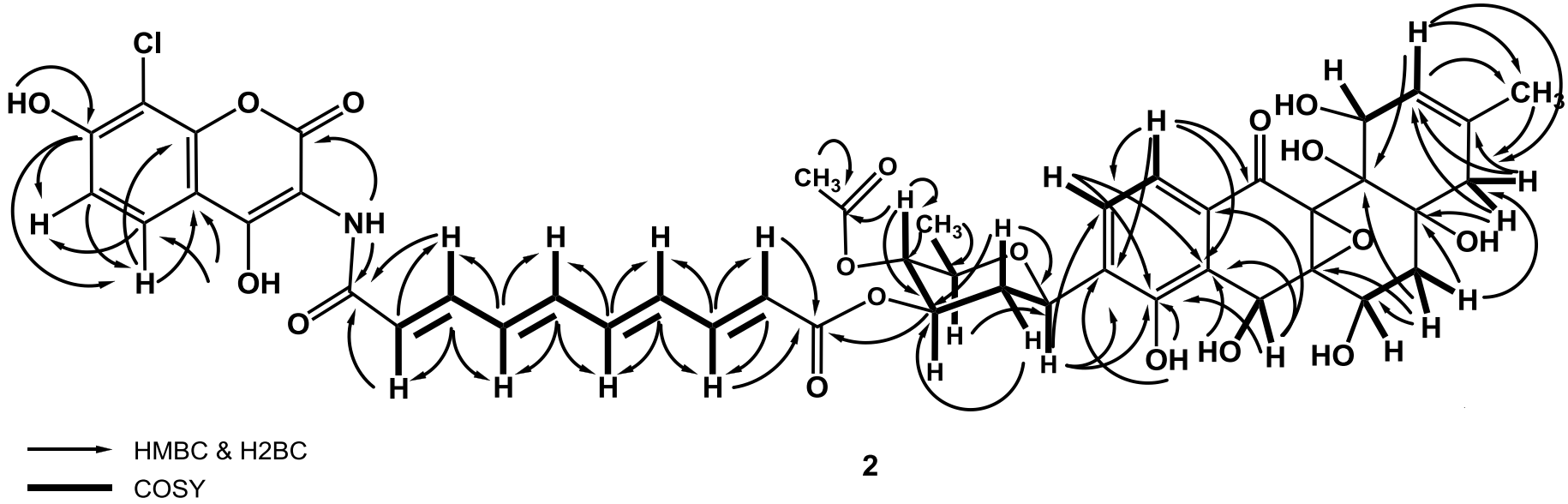


Table S2a

Pos.	δ_C*	δ_H (J Hz)	COSY	HMBC
1	130.7	6.10 d (10.3)	2-H	2-H
2	130.8	6.22 d (10.3)	1-H	1-H, 4-H
3	141.7			1-H, 2-H
4	41.9	a: 2.75 d (14.2)	4-H_b	2-H, 13-H_2
		b: 2.30 d(14.2)	4-H_a	
4a	70.2			1-H, 2-H, 4-H_b, 6-H
5	33.5	a: 1.82 dd (15.0, 3.1)	5-H_b, 6-H	4-H_b, 6-H
		b: 1.75 d (15.0)	5-H_a, 6-H	
6	65.0	4.49 d (3.1)	5-H_a, 5-H_b	5-H_b
6a	67.3			6-H
7	64.5	5.83 s br		6-H
7a	125.7			7-H, 11-H
8	154.1			7-H, 1'-H
9	135.3			1'-H
10	126.7	7.53 d (8.4)	11-H	
11	119.8	7.62 d (8.4)	10-H	
11a	129.5			7-H, 10-H
12	197.7			11-H
12a	68.6			5-H_b, 6-H, 7-H
12b	73.0			1-H
13	114.4	4.98, 4.95		2-H, 4-H_a, 4-H_b
1'	72.7	4.99 d (11.2)	2'-H_{ax}	10-H
2'	37.7	eq: 2.59 dd (12.6, 5.3)	2'-H_{ax}	
		ax: 1.59 dd (12.3, 11.2)	1'-H, 2'-H_{eq}, 3'-H	
3'	73.4	5.22 m	2'-H_{ax}, 2'-H_{eq}, 4'-H	2'-H_{eq}, 4'-H
4'	75.7	4.85***	3'-H, 5'-H	2'-Heq, 3'-H, 5'-H
5'	75.2	3.76 m	4'-H, 6'-H_3	4'-H
6'	17.9	1.27 d (6.1)	5'-H	4'-H, 5'-H
7'	171.1			4'-H, 8'-H_3
8'	20.5	2.03 s		

1.)*from HMBC/HSQC 2.) **no HMBC-Signal 3.)***covered from water 4.) 2D signals in brackets are week

Table S2a: NMR data of simocyclinone D10 (**3**), angucycline and sugar part. (700 / 175MHz, CD_3OD, Temperature = 25°C, solvent as internal reference)

Table S2b

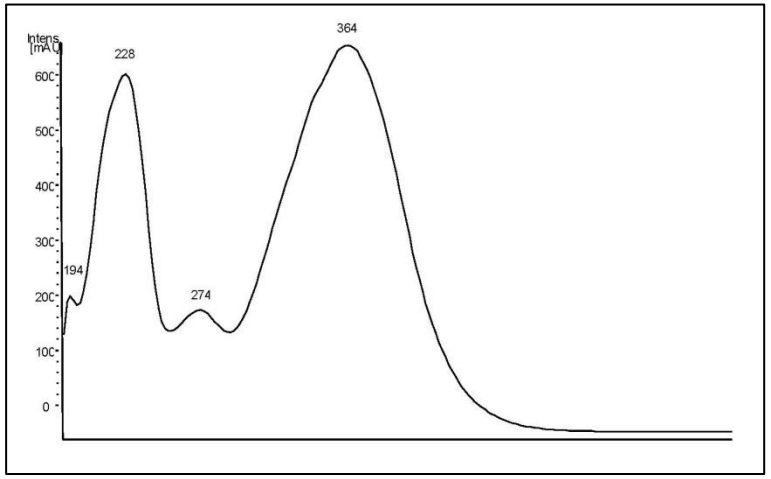
Pos.	δ_C*	δ_H (J Hz)	COSY	HMBC
1''	168.3			2''-H, 3''-H
2''	124.3	6.52 d (15.0)	3''-H	
3''	143.8	7.44 dd (15.0, 11.8)	2''-H, 4''-H	5''-H
4''	134.9	6.68 dd (13.6, 11.8)	3''-H, 5''-H	2''-H
5''	140.9	6.79 m	4''-H	6''-H
6''	134.8	6.79 m	7''-H	5''-H
7''	134.6	6.62 dd (13.2, 11.8)	6''-H, 8''-H	9''-H
8''	145.7	7.34 dd (15.0, 11.8)	7''-H, 9''-H	6''-H
9''	122.7	5.95 d (15.0)	8''-H	
10''	167.0			8''-H, 9''-H
11''	**			
12''	**			
13''	**			
13a''	115.7			14''-H, 15''-H
14''	120.5	7.19 s br		15''-H
15''	116.7	7.19 s br		14''-H
16''	152.2			14''-H, 15''-H
17''	115.7			15''-H
17a''	143.1			14''-H

1.)*from HMBC/HSQC 2.) **no HMBC-Signal 3.)***covered from water 4.) 2D signals in brackets are week

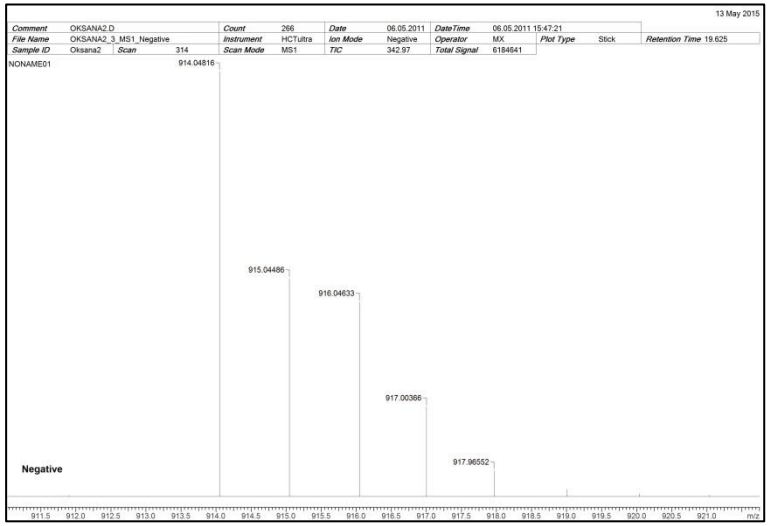
Table S2b: NMR data of simocyclinone D10 (**3**), tetrane and aminocoumarin part. (700 / 175MHz, CD_3OD, Temperature = 25°C, solvent as internal reference)

Figure S3a: UV, MS and ^{1}H NMR spectra of simocyclinone D10 (**3**). **(A)** UV spectrum of **3**. **(B)** ^{1}H NMR spectrum of **3** (700 MHz, CD_3OD, 25°C). **(C)** ESI-MS (neg. mode) of **3**.

A



B



C

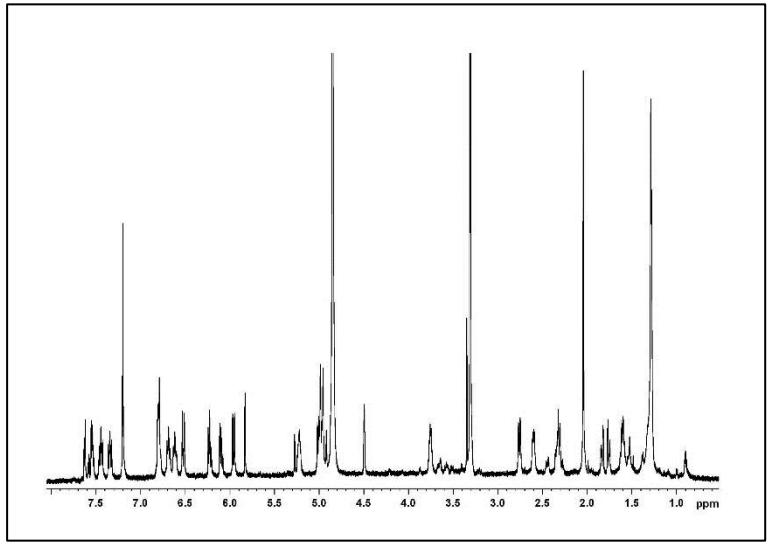
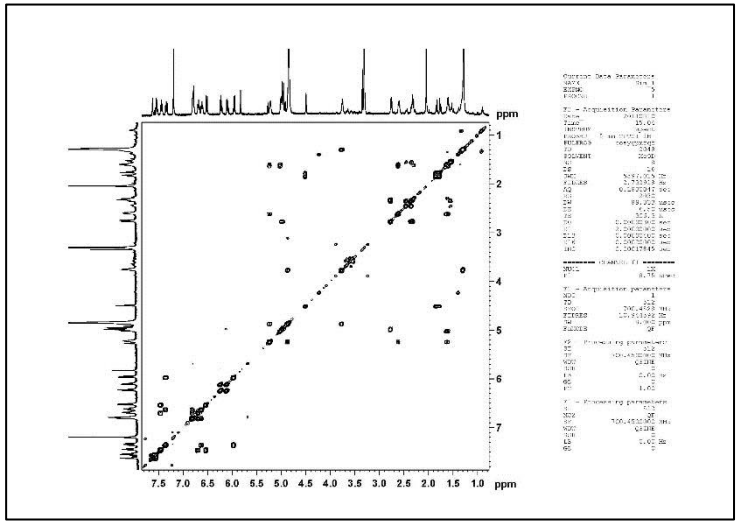
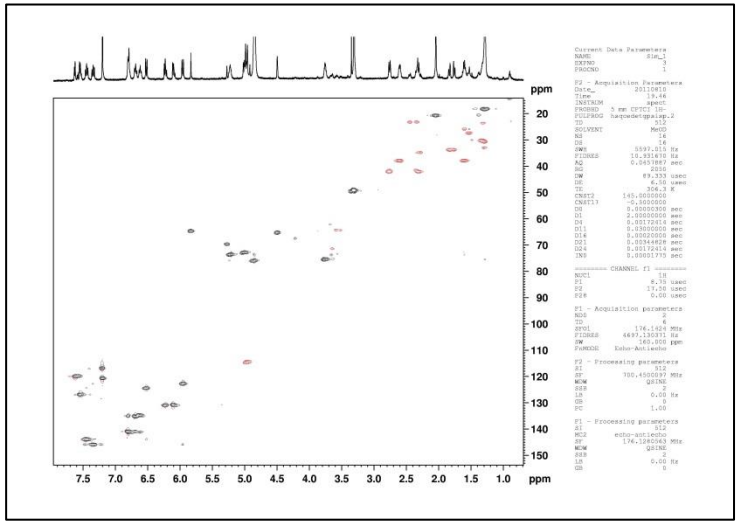


Figure S3b: 2D NMR spectra of simocyclinone D10 (**3**). **(A)** COSY spectrum of **3** (700 MHz, CD_3OD, 25 °C). **(B)** HSQC spectrum of **3** (700 MHz, CD_3OD, 25 °C). **(C)** HMBC spectrum of **3 (**700 MHz, CD_3OD, 25 °C).

A



B



C

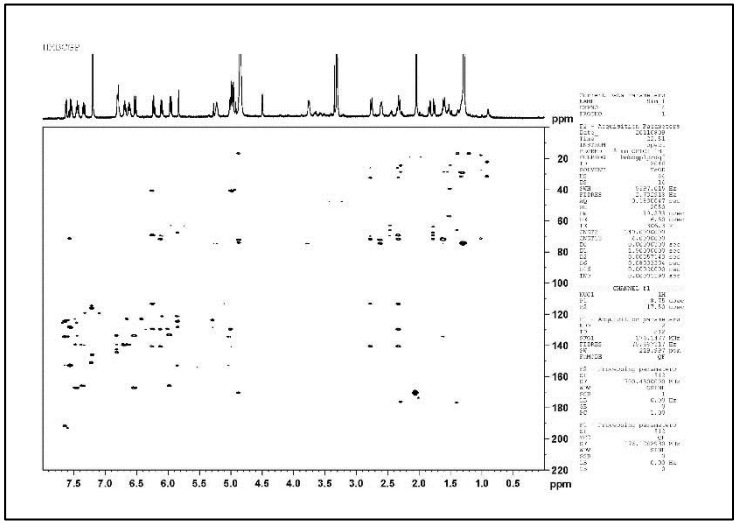


Figure S3c: Structure elucidation of simocyclinone D10 **(3):** 2D NMR correlations of **3.**

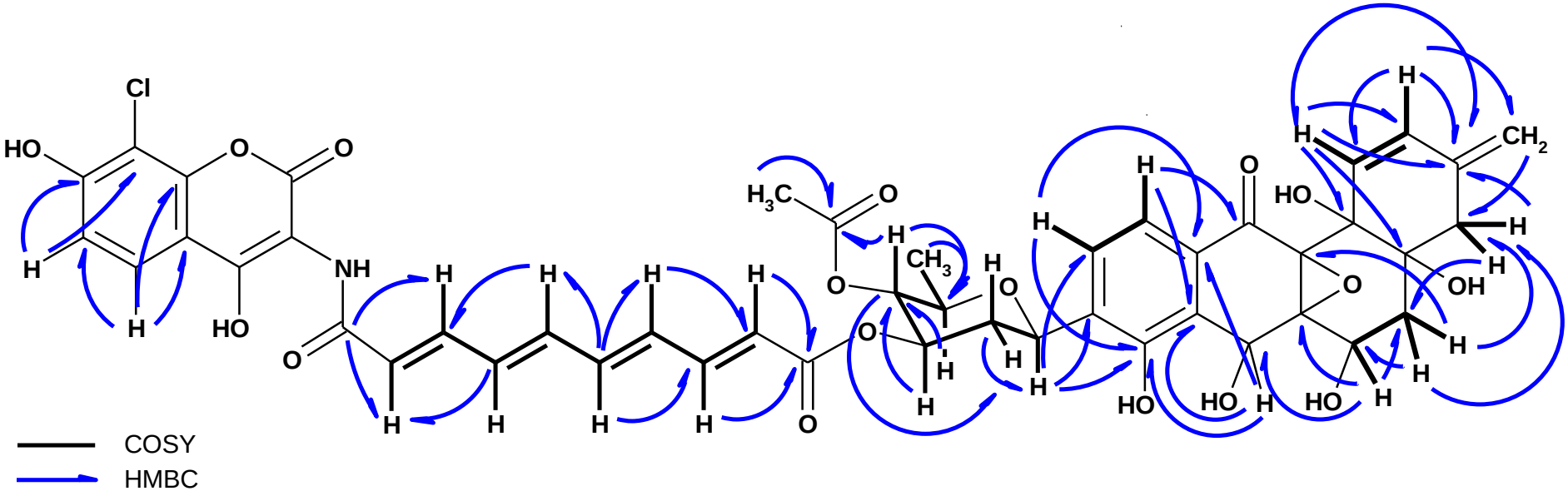


Table S3a

Pos.	δ_C*	δ_H (J Hz)	COSY	HMBC
1	127.4	6.24 d (10.3)	2-H	
2	135.4	5.67 d (10.3)	1-H	13-H_3, 4-H_b
3	66.3			1-H, 4-H_a, 4-H_b
4	44.2	H_a: 2.12 d (13.6)	4-H_b	2-H, 5-H_a, 5-H_b, 13-H_3
		H_b: 1.76 d (13.6)	4-H_a	
4a	70.5			1-H, 4-H_b, 5-H_a, 5-H_b
5	39.6	H_a: 1.94 dd (12.9, 9.6)	5-H_b, 6-H	4-H_b
		H_b: 1.80 dd (13.1, 6.0)	5-H_a, 6-H	
6	63.2	4.22 m	5-H_a, 5-H_b, 6-OH	5-H_a, 5-H_b
6-OH		5.56 br d (6.7)	6-H	
6a	67.8			5-H_a, 5-H_b
7	63.7	5.72 d (7.4)	7-OH	
7-OH		6.70 **		
7a	124.2			7-H, 8-OH, 11-H
8	152.4			8-OH
8-OH		9.56 s		
9	133.6			8-OH, 1‘-H, 2‘-H_{ax}
10	125.6	7.48 d (8.2)	11-H	
11	118.2	7.40 d (8.2)	10-H	
11a	128.5			10-H
12	190.1			11-H
12a	67.1			
12b	69.1			2-H, 5-H_a
13	29.2	1.11 s		2-H, 4-H_a
1’	70.6	4.97 d (11.6)	2’-H_{ax}	10-H, 2’-H_{ax}
2’	36.5	H_{eq}: 2.44 dd (14.7, 5.1)	2’-H_{ax}, 3’-H	
		H_{ax}: 1.58 dd (14.7, 11.7)	1’-H, 2’-H_{eq}	
3’	71.6	5.19 ddd (11.3, 9.4, 5.1)	2’-H_{eq}, 2’-H_{ax}, 4’-H	2’-H_{ax}, 4’-H
4’	74.1	4.76 dd (9.6, 9.6)	3’-H, 5’-H	3’-H, 5’-H, 6’-H_3
5’	73.1	3.80 dq (9.6, 6.1)	4’-H, 6’-H_3	4’-H, 6’-H_3
6’	17.7	1.17 d (5.9)	5’-H	
7’	169.7			4’-H, 8’-H_3
8’	20.5	2.02 s		

*missing signal **signal covered from others

Table S3a: NMR data of simocyclinone D11 (**4**), angucycline and sugar part. (600 / 150MHz, DMSO-d_6, Temperature = 35°C, solvent as internal reference)

Table S3b

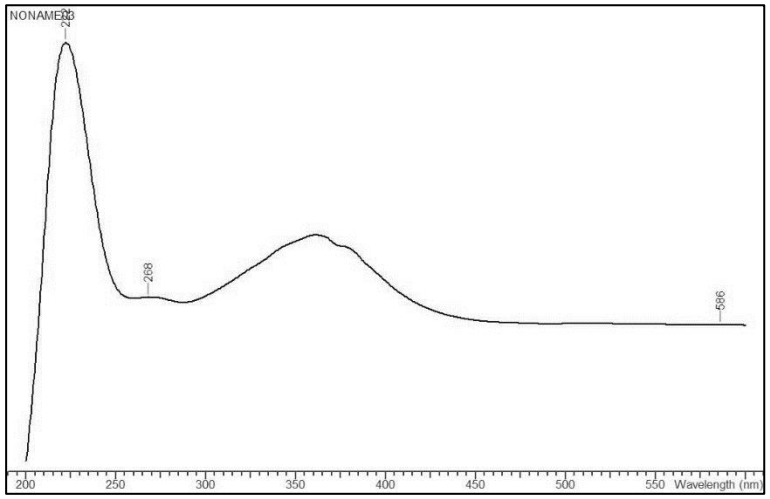
Pos.	δ_C*	δ_H (J Hz)	COSY	HMBC
1''	165.2			2''-H, 3''-H
2''	121.1	6.02 d (14.4)	3''-H	
3''	144.4	7.29 dd (14.4, 13.0)	2''-H, 4''-H	
4''	133.1	6.65 dd (13.0, 12.2)	3''-H, 5''-H	2''-H
5''	140.3	6.85 dd (12.2, 11.1)	4''-H, 6''-H	3''-H, 7''-H
6''	139.0	6.86 dd (14.1, 11.1)	5''-H, 7''-H	4''-H
7''	134.2	6.72 dd (14.1, 11.6)	6''-H, 8''-H	9''-H
8''	140.9	7.31 dd (15.3, 11.5)	7''-H, 9''-H	6''-H
9''	124.8	6.60 d (15.3)	8''-H	
10''	166.0			8''-H, 10''-H
11''	158.8			
12''	*			
13''	*			
13''-OH		10.2 s		
13a''	115.4			13''-OH, 14''-H, 15''-H
14''	119.5	7.25		
15''	115.8	7.25		14''-H
16''	150.9			14''-H, 15''-H
17''	114.6			
17a''	145.6			14''-H

*missing signal **signal covered from others

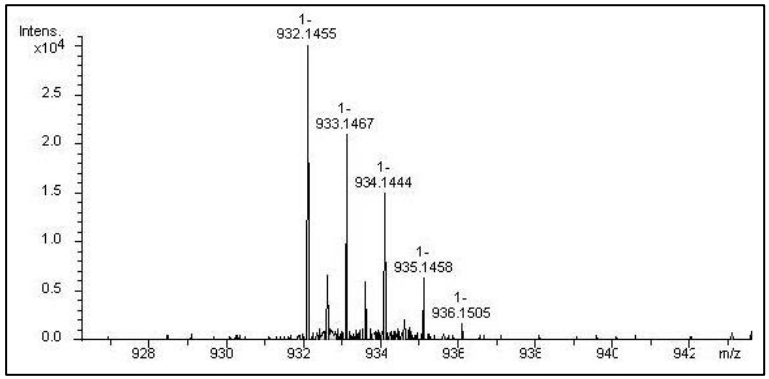
Table S3b: NMR data of simocyclinone D11 (**4**), aminocoumarine and tetraene part. (600 / 150MHz, DMSO-d_6, Temperature = 35°C, solvent as internal reference)

Figure S4a: UV, MS and ^{1}H and ^{13}C NMR spectra of simocyclinone D11 (**4**). **(A)** UV spectrum of **4. (B)** ESI-MS (neg. mode) of **4**. **(C)** ^{1}H NMR spectrum of **4** (600 MHz, DMSO-d_6, 35°C) **(D)** ^{13}C NMR spectrum of **4** (600 MHz, DMSO-d_6, 35°C).

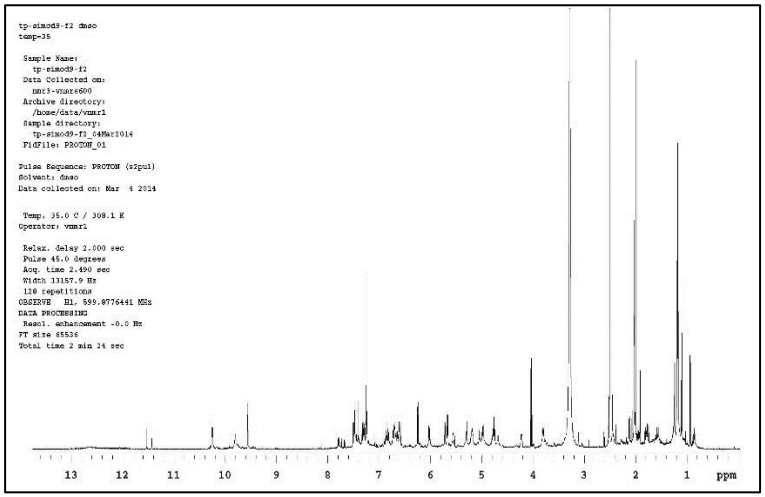
A



B



C



D

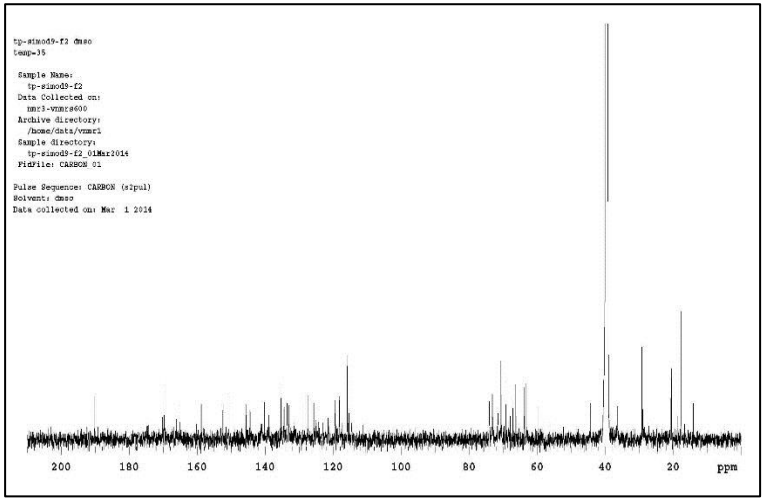
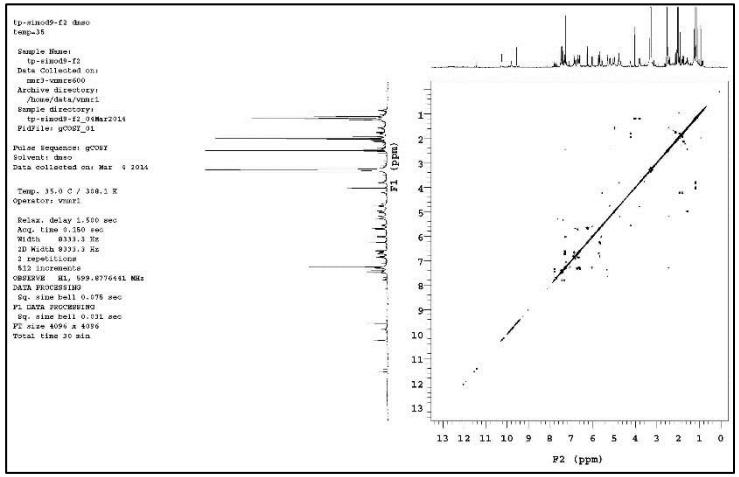
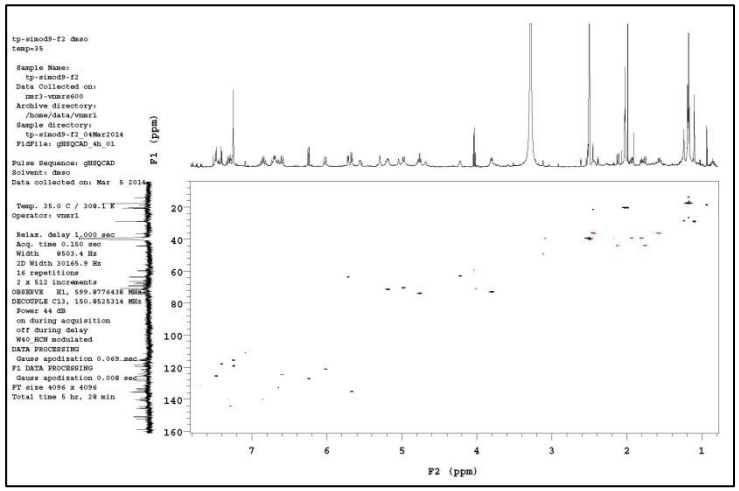


Figure S4b: 2D NMR spectra of simocyclinone D11 (**4**). **(A)** COSY spectrum of **4** (600 MHz, DMSO-d_6, 35 °C). **(B)** HSQC spectrum of **4** (600 MHz, DMSO-d_6, 35°C). **(D)** HMBC spectrum of **4** (600 MHz, DMSO-d_6, 35°C).

A



B



C

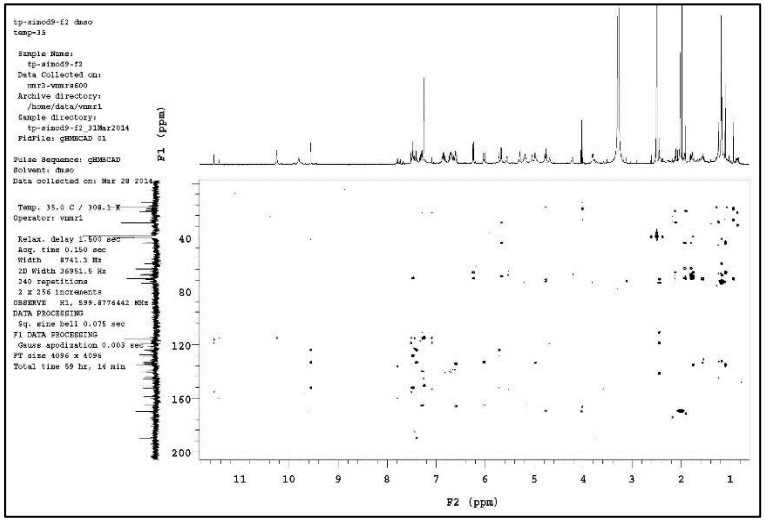


Figure S4c: Structure elucidation of simocyclinone D11 (**4**). 2D NMR correlations of **4.**

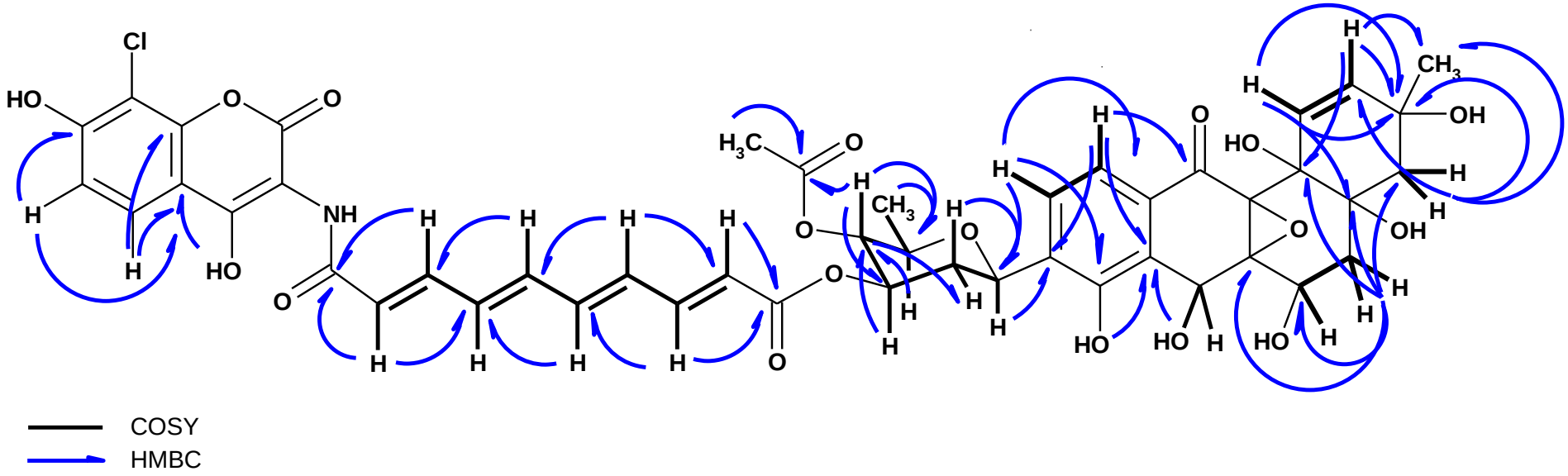


Table S4a

Pos.	δ_C [a]	δ_H [a]	COSY[b]	HMBC[b]
1	196.8			(2-H)[d], 13-H_3
2	122.4	5.89	13-H_3	13-H_3
3	161.3			13-H_3
4	41.7	H_a: 2.69	4-H_b	2-H, 13-H_3
		H_b: 2.39	4-H_a	
4a	75.2			4-H_b
5	28.5	H_a: 1.80	5-H_b, 6-H_a, 6-H_b	6-H_a, 6-H_b
		H_b: 1.61	5-H_a, 6-H_a, 6-H_b	
6	23.4	H_a: 2.53	5-H_a, 5-H_b, 6-H_b	5-H_a
		H_b: 2.36	5-H_a, 5-H_b, 6-H_a	
6a	73.4			5-H_a, 6-H_b
7	70.5	5.34		
7a	125.7			7-H
8	153.4			7-H, 1'-H
9	135.8			1'-H
10	127.9	7.43	11-H	1'-H
11	119.4	7.38	10-H	
11a	129.4			7-H, 10-H
12	194.3			11-H
12a	64.7			5-H_a, 6-H_a, (7-H)
12b	76.0			2-H, 5-H_a
13	23.8	1.90		2-H4-H_a
1'	72.3	4.90	2'-H_{ax}	10-H, 2'-H_{ax}
2'	37.5	H_{ax}: 1.48	1'-H, 2'-H_{eq}, 3'-H	1'-H
		H_{eq}: 2.41	2'-H_{ax}	
3'	75.7	4.96	2'-H_{ax}, 4'-H	1'-H, 2'-H_{ax}, 4'-H
4'	74.8	3.30	3'-H, 5'-H	2'-H_{eq}, 3'-H, 6'-H_3
5'	77.0	3.53	4'-H, 6'-H_3	4'-H, 6-H_3
6'	18.3	1.31	5'-H	4'-H

[a]values taken from 2D-Spectra

[b]2D signals in brackets are week

[c]not visible in HSQC/HMBC, no ^{13}C NMR available

[d]only visible at very high scale

Table S4a: NMR data of simocyclinone D12 (**5**), angucycline and sugar part. (700 / 175MHz, CD_3CN / HCOOH, Temperature = 25°C, solvent as internal reference)

Table S4b

Pos.	δ_C [a]	δ_H [a]	COSY[b]	HMBC[b]
1''	166.9			3'-H, 2''-H, 3''-H
2''	123.6	6.01	3''-H	4''-H
3''	144.6	7.34	2''-H, 4''-H	4''-H
4''	134.7	6.65	3''-H, 5''-H	2''-H, 5''-H
5''	140.2	6.74	4''-H, 6''-H	3''-H, 4''-H, 7''-H
6''	144.4	7.42	5''-H, 7''-H	7''-H
7''	134.3	6.65	6''-H, 8''-H	5''-H
8''	144.0	7.42	7''-H, 9''-H	7''-H
9''	123.4	6.51	8''-H	7''-H
10''	167.8			9''-H
11''	[c]			
12''	[c]			
13''	155.5			14''-H
13a''	116.6			14''-H
14''	120.0	7.22	15''-H	
15''	112.5	7.14	14''-H	
16''	154.3			15''-H
17''	121.0			15''-H
17a''	141.3			14''-H

[a]values taken from 2D-Spectra

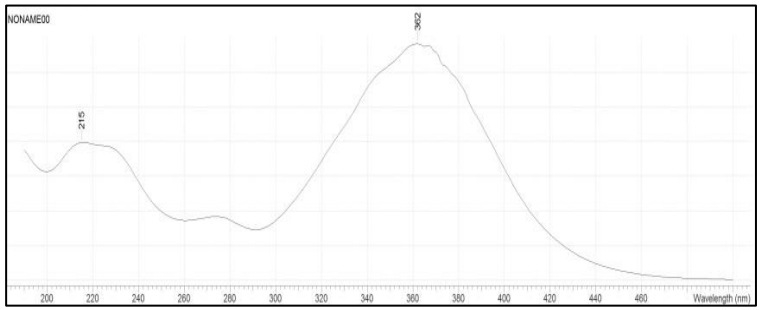
[b]2D signals in brackets are week

[c]not visible in HSQC/HMBC, no ^{13}C NMR available

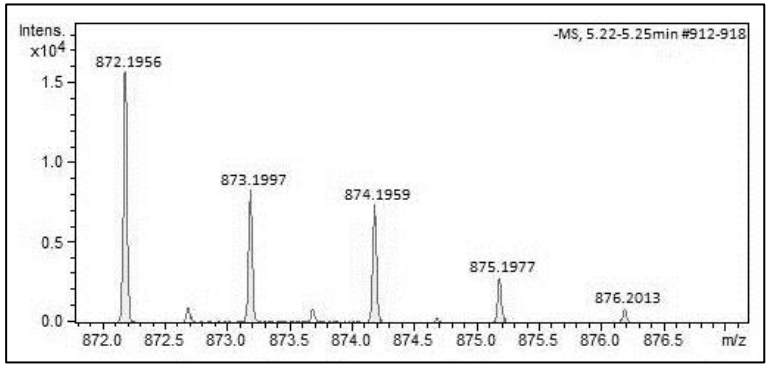
Table S4b: NMR data of simocyclinone D5 (**5**), aminocoumarine and tetraene part. (700 / 175MHz, CD_3CN / HCOOH, Temperature = 25°C, solvent as internal reference)

Figure S5a: UV, MS and ^{1}H NMR spectra of simocyclinone D5 (**5**)**. (A)** UV spectrum of **5. (B)** ESI-MS (neg. mode) of **5**. **(C)** ^{1}H NMR spectrum of **5** (700 MHz, CD3CN/HCOOH, 25°C).

A



B



C

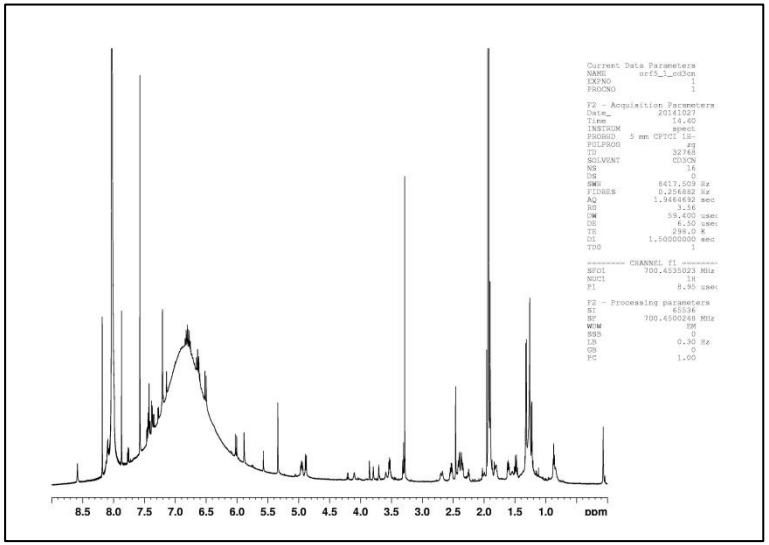
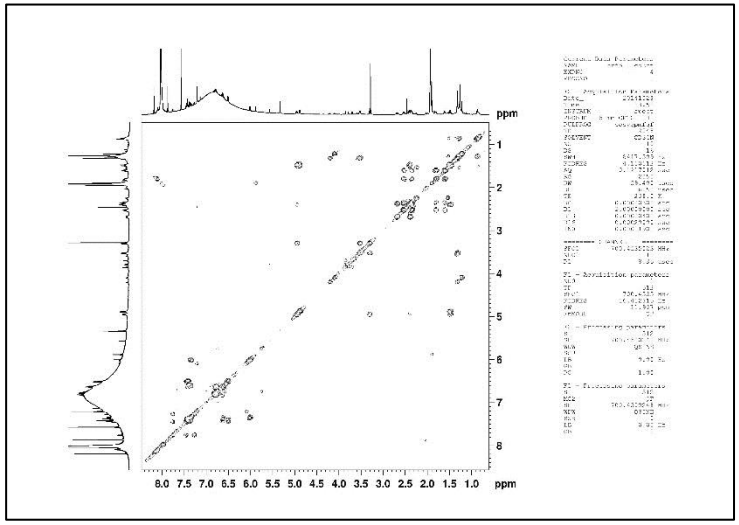
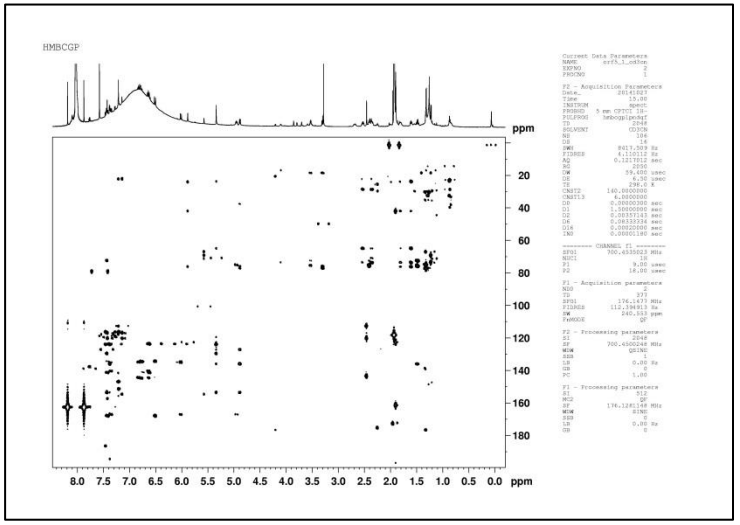


Figure S5b: 2D NMR spectra of simocyclinone D5 (**5**)**. (A)** COSY spectrum of **5** (700 MHz, CD3CN/HCOOH, 25°C)**. (B)** HSQC spectrum of **5** (700 MHz, CD3CN/HCOOH, 25°C) **(C)** HMBC spectrum of **5** (700 MHz, CD3CN/HCOOH, 25°C).

A



C



B

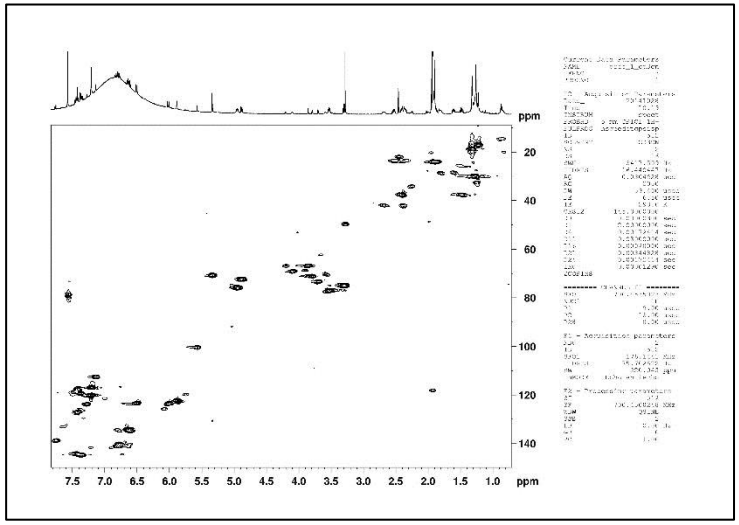


Figure S5c: Structure elucidation of simocyclinone D5 (**5**)**.** 2D NMR correlations of **5.**

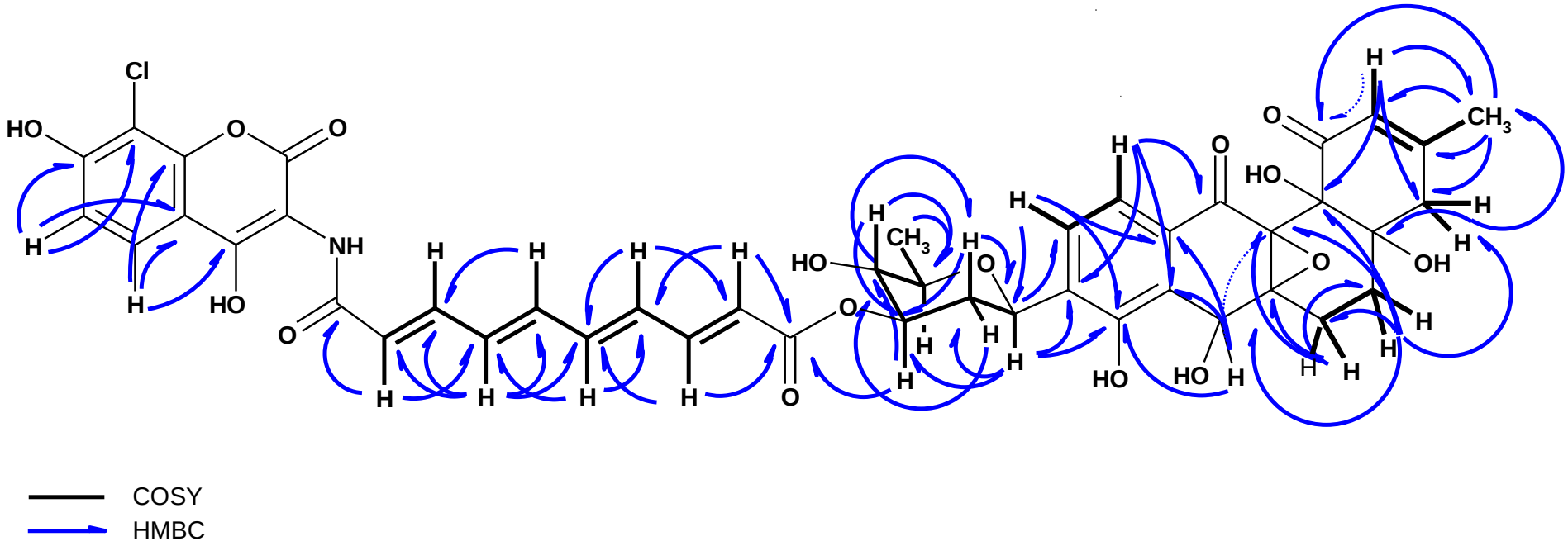


Figure S6: Chromatogram of the HPLC-ESI/MS analysis of *K.* sp. *ΔPKS* mutant crude extract. Arrows indicate peaks corresponding to simocyclinones.

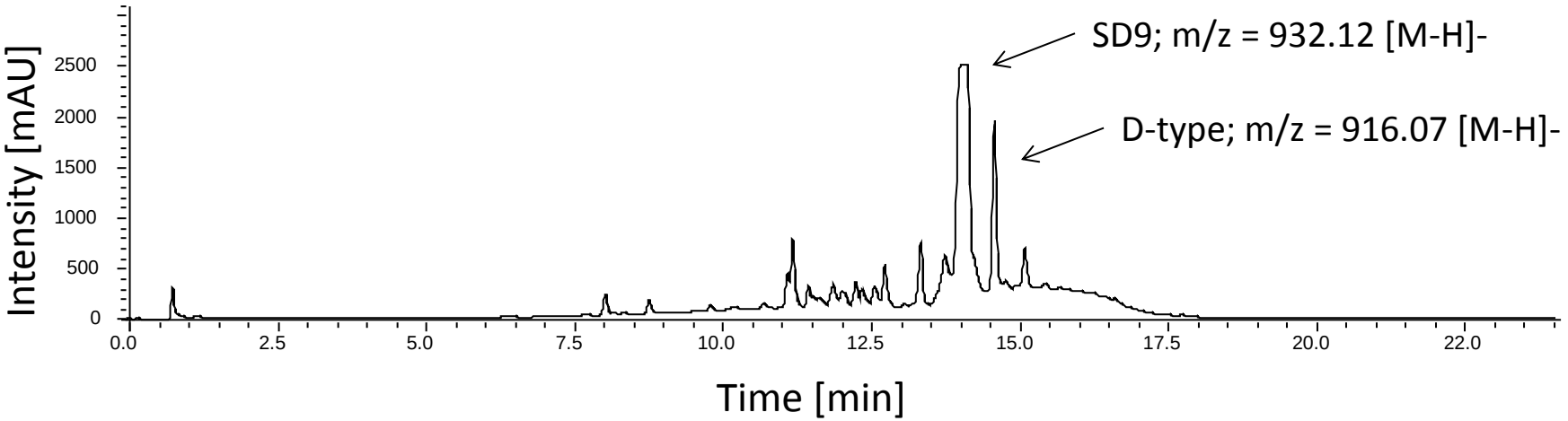


Figure S7: Chromatograms of the HPLC-ESI/MS analysis of: **(A)** crude extract of *K.* sp. with deleted *smcKSII* ; **(B)** crude extract of *K.* sp. with deleted *smcKSII* complemented with intact *smcKSII*; **(C)** crude extract of *K.* sp. with deleted *smcKSII* complemented with modified *smcKSII*; **(D)** crude extract of *K.* sp. wild type strain. Arrows indicate ions corresponding to simocyclinones.

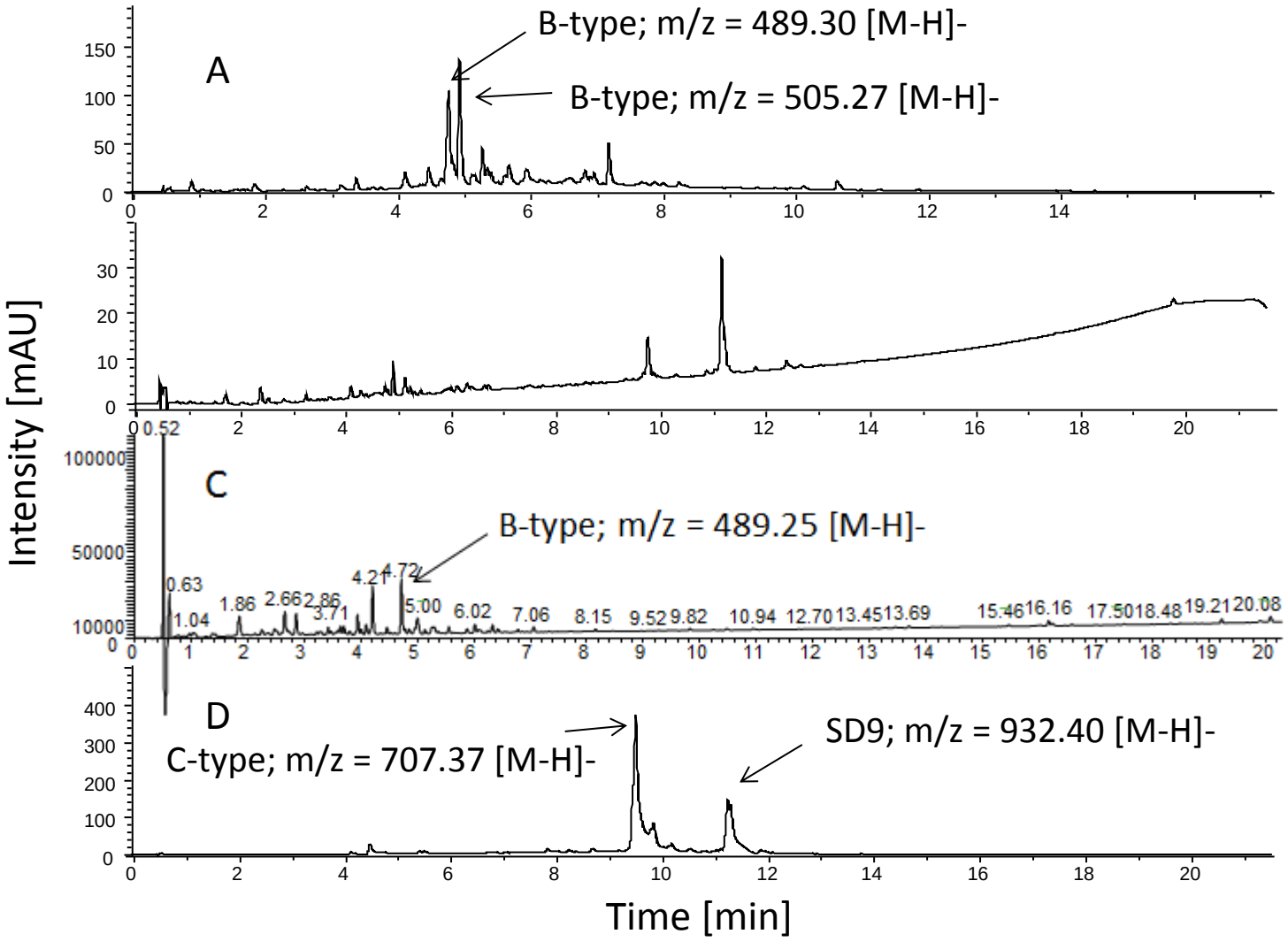
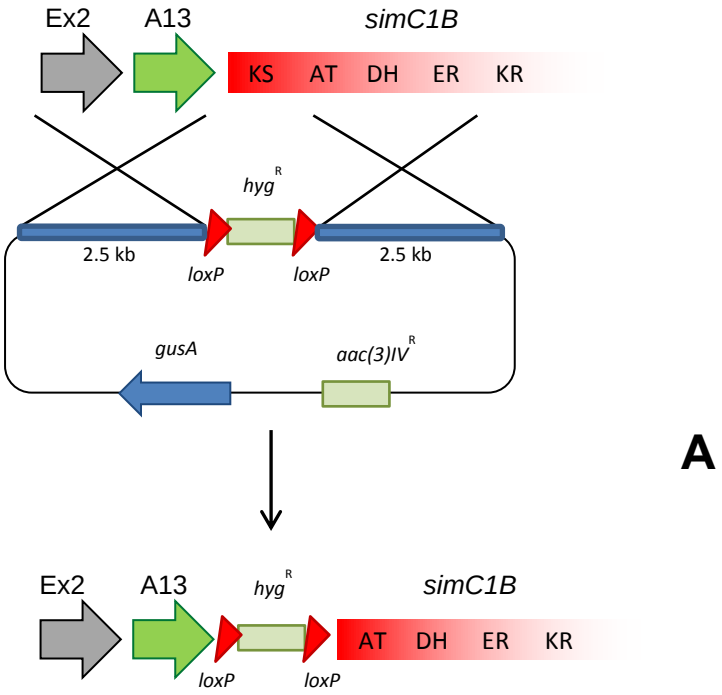
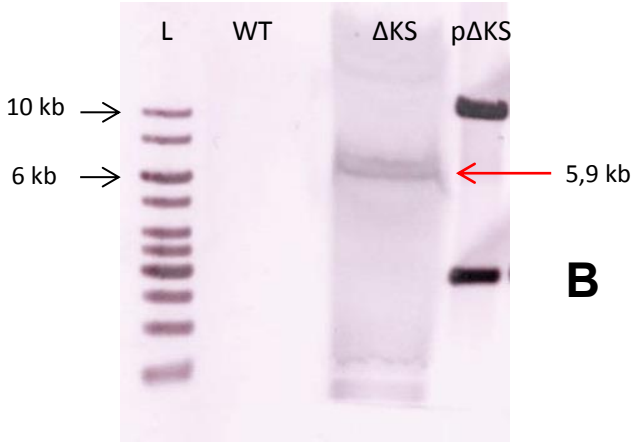


Figure S8a: *S. antibioticus* Tü6040 Δ*KS* mutant. **(A)** Generation of *S. antibioticus* Tü6040 ΔKS mutant via double crossover (DNA region encoding KS domain was replaced with a hygromycin-resistance gene). *hyg*^{R} – hygromycin resistance gene, *aac(3)IV*^{R} - apramycin resistance gene, *gusA* – β-glucoronidase gene, *loxP* – target site for Cre recombinase. **(B)** Nylon membrane after hybridization of *S. antibioticus* Tü6040 wild type and *S. antibioticus* ΔKS chromosomal DNA with *hph*-zond. Red arrow indicates predicted 5,9 kb fragment in *S. antibioticus* Tü6040 ΔKS mutant. L – DNA ladder, WT – wild type of *S. antibioticus* Tü6040, ΔKS - *S. antibioticus* Tü6040 ΔKS, pΔKS – pKGKSΔ cassette used for KS domen disruption.



A



B

Figure S8b: Chromatogram of the HPLC-ESI/MS analysis of *S. antibioticus* mutant crude extract. Arrows indicate ions corresponding to simocyclinones.

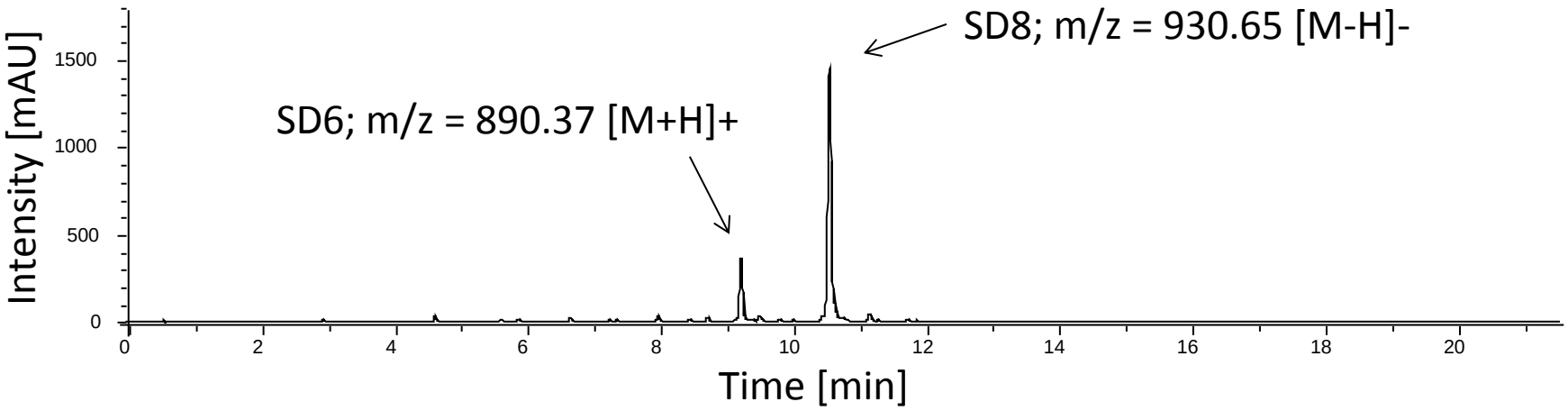


Figure S9: Generation of *K.sp. ORF7** mutant via double crossover. *aac(3)IV*^{R} – apramycin resistance gene; *hyg*^{R} – hygromacin resistance gene; *gusA* – β-glucoronidase gene; *loxP* – target site for Cre recombinase

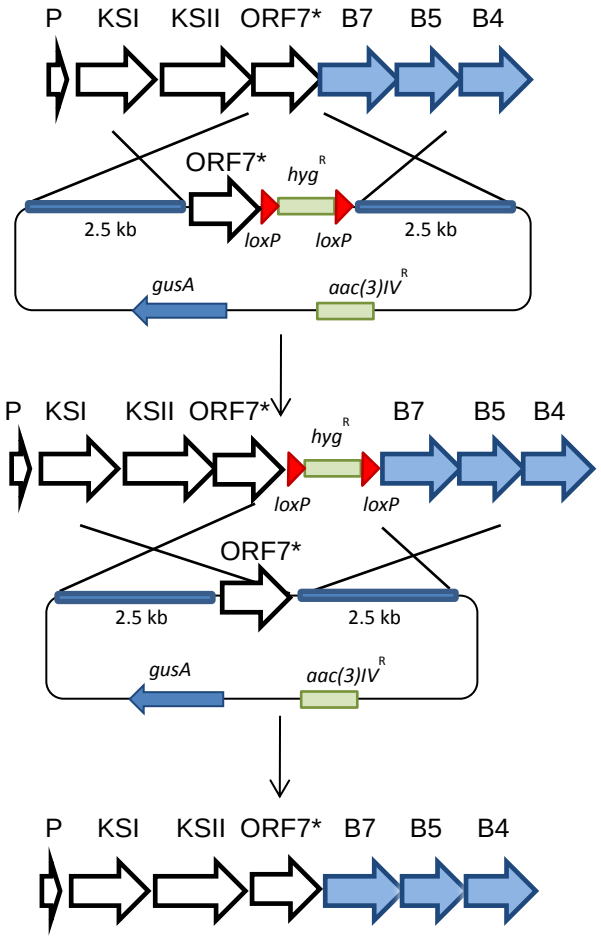


Figure S10: Level of transcription of *smc* genes in NL5 and DNPM production media. In yellow are labeled genes involved in aminocoumarin biosynthesis, blue – genes involved in sugar unit formation, green – genes responsible for the angucyclic core synthesis, pink – genes putatively involved in modification of the polyketide chain precursor, dark grey – regulatory genes, dark blue – efflux, pale grey – genes of unknown function, white – *smcP, smcKSI, smcKSII, orf7*. 1-43 – number of the gene (see Table 1).

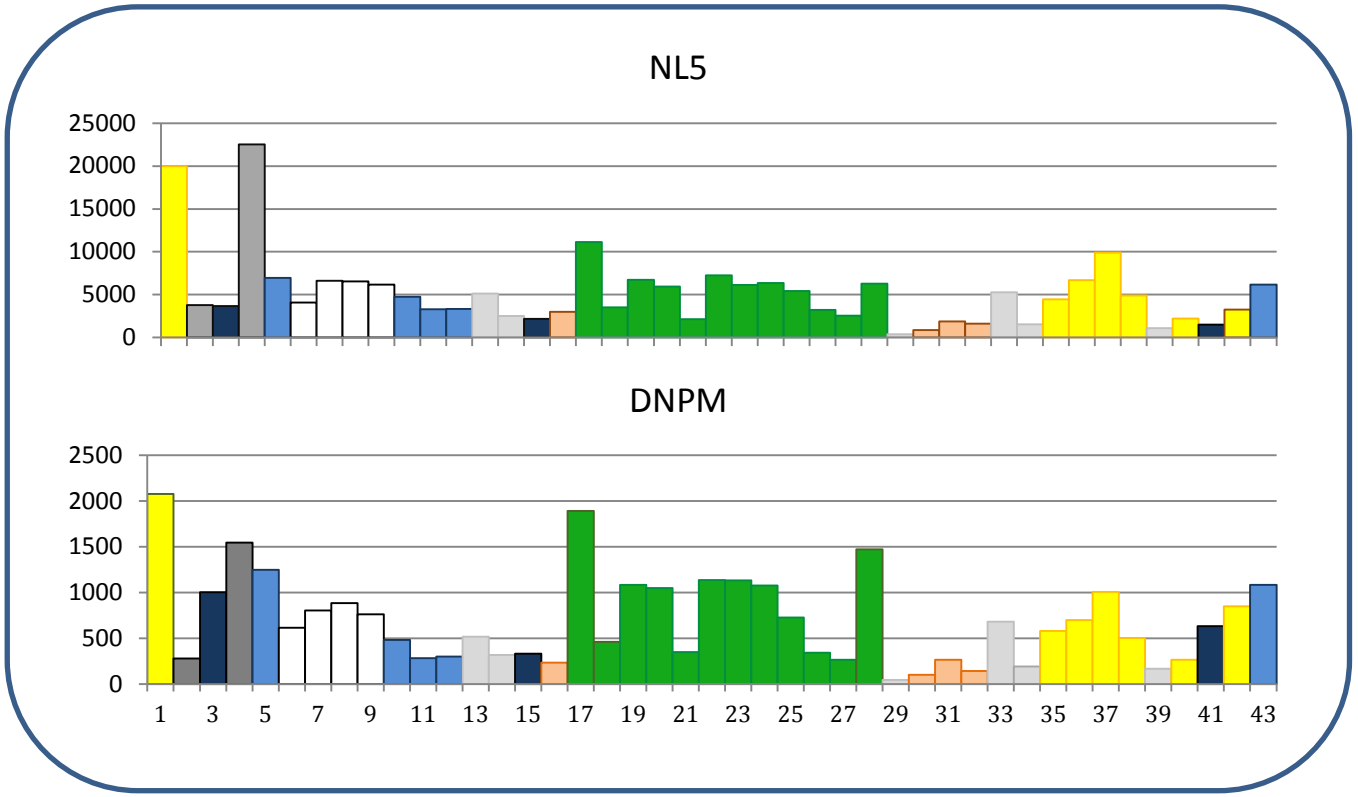


Table S5

Primer	Sequence (5'→ 3')	Restriction site
KS_R_fus-f	TTCGGCAGTATCTGAAGAAG**GATATC**GCGCAGCCCGACGCACCGGG[a]	*EcoRV*
KS_F_fus-r	CCCGGTGCGTCGGGCTGCGC**GATATC**CTTCTTCAGATACTGCCGAA[a]	*EcoRV*
KS_R-SspI-r	ACT**AATATT**CCGCCCCATGGGTCAGGACGA[a]	*SspI*
KS_F-SspI-f	ACT**AATATT**GGTCTGGCCGGGACGACGCTC[a]	*SspI*
KSII_ff-EcoRI-f	CCT**GAATTC**AAGTCCGGCCCGACGCTCGGCAG[a]	*EcoRI*
KSII_ff-MCS -r	CCC**GATATC**AA**CAATTG**AA**AGGCCT**AATGTCCGCCGACGGGCTTCCGG[a]	*EcoRV*, *MunI*, *StuI*
KSII_fr-MunI-f	CCC**CAATTG**AATCTGAACTACGCGATACAGCA[a]	*MunI*
KSII_fr-XbaI-r	CCC**TCTAGA**AATGCGCGTCTGTCGGCTGGCCG[a]	*XbaI*
putPKS-562-f	ATC**AAGCTT**CGACCCCACTGGACCGACACC[a]	*HindIII*
putPKS-562-r	ACG**TCTAGA**CGCGGTGGCCTGGTGCAGCCG[a]	*XbaI*
KSII-com-f	ATA**TCTAGA**A*ATGTGCGGGCTCTAACACGTCCTAGTATGGTAGGATGAGCAA*GGTACCAAGCTTATTGGCACTAGTCGAGCAACGGAGGTACGGACATGCCCGAGGCCCCGGGTCGG[a,b]	*XbaI*
KSII-com-r	ATA**GATATC**ATCACGCCGAAGGCCGATGCTT[a]	*EcoRV*
KSII*21p-f.1	ATA**TCTAGA**A*ATGTGCGGGCTCTAACACGTCCTAGTATGGTAGGATGAGCAA*CGAGCAACGGAGGTACGGACATGCCCGAGGCCCCGGGTCGG[a,b]	*XbaI*

KSII*-r.1	TGGCGTAGTTCCCCGCCGACGCCGCCGTGGCCAGGGTCAG	-
KSII*-f.2	CTGACCCTGGCCACGGCGGCGTCGGCGGGGAACTACGCCA	-
KSII*21p-r.2	ATA**GATATC**ATCACGCCGAAGGCCGATGCTT[a]	*EcoRV*
ORF7-rev -f	CCC**GAATTC**GTGACCGTGCGGTCGGGGAAG	*EcoRI*
ORF7_rev- -r	CCC**GATATC**C**AATATT**C**CAATTG**C**AGGCCT**TGGAGCAGAACGCGAAAGCCGT[a]	*EcoRV, SspI, MfeI, StuI*
ORF7_for -f	CCC**AATATT**CGGGCATGCCCTCCTCTACCA[a]	*SspI*
ORF7_for-r	CCC**TCTAGA**GTGGTCCCGGAACGGCTCGCC[a]	*XbaI*
ORF7* –for1	CCCC**CAATTG**TATGAAGTATTTCGTCCTGGG[a]	*MunI*
ORF7*-rev1	ACGTCTCCGACTGGCCGCCCCCCGCCACATTGGCGGTCGCCA	-
ORF7*-for2	GGCGGCCAGTCGGAGACGT	-
ORF7*-rev2-	CCCT**AGGCCT**ATCAGCGGAGATCGGCCAGCCC[a]	*StuI*
hph-MfeI-f	CCC**CAATTG**AGAATAGGAACTTCGGAATAGG[a]	*MfeI*
hph-BamHI-r	CCC**GGATCC**TCCGTATTTGCAGTACCAGCGT[a]	*BamHI*
smcX5-21-XbaI	CCC**TCTAGA***ATGTGCGGGCTCTAACACGTCCTAGTATGGTAGGATGAGCAA*GGAGGTACGGAATGATCGAGCAGCTGCTGGCG[a,b]	*XbaI*
smcX5-HindIII	CCC**AAGCTT**CTAAATCTCCGCTGTGCCGCC[a]	*HindIII*
smcC3-A9-XbaI	CCC**TCTAGA***CCGGGCGGCTTCCTCATGCTTGACTTGACTAGGATAAAGGG*GGAGGTACGGAATGCATTCACCGCCCGTC	*XbaI*

	TGG [a,b]	
smcC3-BglII	CCC**AGATCT**TCACCGGACAACTCCTCTCGT[a]	*BglII*
SmcTet-KpnI-f	CCC**GGTACC***TGTGCGGGCTCTAACACGTCCTAGTATGGTA GGATGAGCAA*CTGGAGGCGGTCATGGCGGTTGTGACGCC AGACGAA[a,b]	*KpnI*
SmcTet-rev	CCC**TCTAGA**CGGGCATGCCCTCCTCTACCA[a]	*XbaI*
SmcA7_ff-EcoRI-f	cca**GAATTC**TGTCAGGTCGGCGTCCAGGAGA	*EcoRI*
SmcA7_ff-MCS-EcoRV-r	CCC**GATATC**A**CAATTG**CC**AATATT**ACGAGCTCAGCCCC GGCAAGGC	*EcoRV*, *MunI*, *SspI*
SmcA7_fr-MunI-f	CCC**CAATTG**TTCCCATGTCCTTCCGTTGTTC	*MunI*
SmcA7_fr-XbaI-r	CCC**TCTAGA**AAGCGCTTCGAACCGGACCGTC	*XbaI*
SmcA9_ff-EcoRI-f	CCA**GAATTC**TCGACCTGCACGGACCCCAGCA	*EcoRI*
SmcA9_ff-MCS-EcoRV-r	CCC**GATATC**A**CAATTG**CC**AATATT**AGCGTCGCACATCT GGACGAAG	*EcoRV*, *MunI*, *SspI*
SmcA9_fr-MunI-f	CCC**CAATTG**TGCCGGGTGGAACACACCGTCG	*MunI*
SmcA9_fr-XbaI-r	CCC**TCTAGA**AAGGAGTTCTCCTCGTCGAATG	*XbaI*
Hph-MunI-f	CCC**CAATTG**AGAATAGGAACTTCGGAATAGG	*MunI*
Hph-BamHI-r	CCC**GGATCC**TCCGTATTTGCAGTACCAGCGT	*BamHI*
simA8-11F	TT**GAATTC**CACTCCTCTCTCCGTC	*EcoRI*

simA8-11R	TT**GATATC**AACTCGTTCGGTGACG	*EcoRV*
simA8-21F	AA**GATATC**GTTGGTCGACGATCTCCA	*EcoRV*
simA8-21R	AA**AAGCTT**CACCATCGACGACCTGAG	*HindIII*
simA8Forw	ACAGGAGTTCTCCTCGTCGA	
simA8Rev	CAAGGACCACGCCGAGCGTC	

[a] – restriction sites introduced artificially are marked in bold; [b] – promoter sequences introduced artificially are marked in italic

Table S5: Oligonucleotides used in this study.

SUPPORTING INFORMATION

Amycomycin C and D, new angucyclines from Kitasatospora sp.

Table of Contents

Experimental section

General

NMR spectra were obtained on a Bruker Avance 500 MHz spectrometer equipped with a 5mm TCI cryoprobe (Bruker Biospin GmbH, Germany) and a Varian VNMR-S 600 MHz spectrometer (Varian, USA) equipped with 3mm triple resonance inverse probe. The spectra were indicated and referenced to residual 1H signals in deuterated solvents (MeOD: δH = 3.30 ppm, δC = 49.0 ppm, DMSO-d6: δH = 2.50 ppm, δC = 39.5 ppm). HR-ESI-MS spectra were acquired using an UltiMate 3000TM rapid separation liquid chromatography system (Dionex RSLC) equipped with an UV/Vis diode array detector module and connected to an Orbitrap (Thermofisher scientific) or an UHR-TOF mass spectrometer (Bruker Daltronics, MaXis Q-Tof) in positive ion mode under the following conditions (Waters BEH C18 column, 100 × 2.1mm, 1.7 µm, eluting with 0.55 mL/min 95% H2O/MeCN to 5% H2O/MeCN (with isocratic 0.01% FA) over 18 min. For HPLC separation UltiMate 3000TM rapid separation liquid chromatography system (Dionex RSLC) equipped with an UV/Vis diode array detector module and fraction collection unit was used (Thermofisher Scientific). Optical rotation values ([α]D) were obtained on a Jasco J720 spectro-polarimeter with a 1 cm cuvette at 20 °C. IR-spectra were obtained with Bruker IFS 66V. Melting points were detected on a Mettler FP80 with heating table Mettler FP82 Hot Stage under a Nikon microscope 511501, objective 10/0.21. CD-Spectra were measured on Chirascan spectrometer in 0.2 cm quartz-cuvette. Absorptions spectra were measured on Varian Cary 100 double beam spectrometer. As spraying reagents on TLC (KG 60 F254 0.20 mm, Merck) vanillin sulfuric acid and Orcin solution were used: Orcin solution: Solution A: iron(III)chloride (1 g) solved in conc. H2SO4 (100 ml). Solution B: orcin (6 g) solved in ethanol (100 ml). Before usage 10 ml of A were mixed with 1 ml of B. Spraying and heating the TLC with a heat gun at min. 100 °C. Brown-black spot indicate sugars. Vanillin sulfuric acid: The spraying reagent is made from 1.0 g vanillin diluted in 100 ml concentrated sulfuric acid.

Fermentation and isolation of amycomycin C and D in NL5 medium

Batch fermentation of strain *Kitasatospora* sp. was done in a 3-l fermenter, equipped with a turbine impeller system in NL 5 medium (1 liter NL-5 medium consists of NaCl 1.00 g, K2HPO4 1.00 g, MgSO4 × 7 H2O 0.50 g, glycerin 25.0 g, l-glutamine 5.84 g, trace-element-solution 2.00 ml, pH 7.3; trace-element-solution consists of ZnCl 40.0 mg, FeCl3 × 6 H2O 200 mg, CuCl2 × 2 H2O 10.0 mg, MnCl2 × 4 H2O, Na2B4O7 × 10 H2O 10.0 mg, (NH4)Mo7O24 × 4 H2O). The fermenter was inoculated with 5% by volume of a shake flask culture grown in a TSB medium at 28° C in 500 ml Erlenmeyer flasks with 4 baffles for 24 h on a rotary shaker at 250 rpm. The fermentation broth was cultivated for 5 days (30° C, 180 rpm). The culture broth was centrifuged (30 min, 20° C, 7000 rpm, Beckmann-Coulter; Rotor JLA 8.1000) for separation of supernatant and biomass. The biomass was stirred with MeOH (20 min) and evaporated. The resulting residue was mixed with ethyl acetate (400 ml) centrifuged (15 min, 20° C, 7000 rpm, Beckmann-Coulter; Rotor JLA 8.1000) and the supernatant was taken and evaporated. The biomass extract yielded 3.00 g of an oily residue, which was chromatographed in portions of two times 1.50 g on a SiO2 column (KG 60, 3 cm × 22 cm, CHCl3/MeOH 9:1). Fractions were analyzed with spraying reagent vanillin sulfuric acid on SiO2-TLC (CHCl3/MeOH 9:1, fraction 4: Rf: 0.43, fraction 6: Rf: 0.21). Fraction 4 (17.7 mg) was purified via repeated size exclusion chromatography on sephadex LH 20 column (1 × 1 cm × 16 cm, MeOH, 2 × 0.5 cm × 13 cm, MeOH) and led to 200 µg amycomycin D (2). Fraction 6 (7.65 mg) yielded after size exclusion chromatography on sephadex LH20 in 3.42 mg. Final semi preparative workup on HPLC (C18 column, Phenomenex Luna C18(2), 100 A, 250 × 10 mm, 5µm) with a 20 min gradient of 10%

B to 90% B (A: H2O, B: acetonitrile, retention time 12.2 min, detection-λ = 328 nm) resulted in 1.00 mg of amycomycin C (1).

Cultivation and isolation of amycomycin D in DNPM medium

A second cultivation of *K*. sp. (8.8 l) was done in 500 ml flasks containing 100 ml DNPM-medium (1 liter DNPM medium consists of dextrin 40.0 g, soy tone 7.50 g, fresh yeast 5.00 g, 4-Morpholinopropanesulfonic acid (MOPS) 21.0 g, pH 6.8). The cultures were inoculated with 7 ml pre-culture and cultivated for 5 days (28° C, 180 rpm). The pre-cultures were cultivated in 300 ml Erlenmeyer flasks (2 days, 28 °C, 180 rpm) filled with 50 ml TSB medium and 100 µl of sucrose-culture (1 l TSB consists of tryptic soybroth 30.0 g). The culture broth was centrifuged (30 min, 20° C, 7000 rpm, Beckmann-Coulter; Rotor JLA 8.1000) for separation of supernatant and biomass. The biomass was stirred with MeOH (20 min) and evaporated. The resulting residue was mixed with ethyl acetate (400 ml) centrifuged (15 min, 20° C, 7000 rpm, Beckmann-Coulter; Rotor JLA 8.1000) and the supernatant was taken and evaporated. The biomass extract, yielded 184 mg which were diluted in 2.4 ml (DMSO/MeOH 1:4) and pre-separated through prep-HPLC (Waters C18 column, 150 × 19 mm, flow 25 ml/min; 200-300µl per injection) with a gradient from 10%B to 90% B in 30 min (A: H2O, B: acetonitrile, detection-λ = 328 nm) Fraction 14 (1.6 mg, retention time = 21.8 min) led after semi-prep HPLC-isolation to 0.6 mg of amycomycin D (Phenomenex Luna C18, 250 × 10 mm, 5µm, flow: 3 ml/min, gradient 5% B to 100% B in 20 min, A: H2O + 0.1% FA, B: CH3CN + 0.1% FA, detection- λ = 328 nm, RT: 18.5 min, injection 20-50 µl, c = 4 mg/ml) – HR-ESI-MS: pos. 467.1330 [M+H] +, 321.0757 [M-sugar+H]+.

Feeding experiment with alizarin and isolation of O-α-l-talosyl-1,2-dihydroxyanthraquinone

The pre-culture was cultivated in 300 ml Erlenmeyer flasks filled with 50 ml TSB-medium (1 l TSB consists of tryptic soybroth 30.0 g) and 100 µl of *K*. sp. containing sucrose-solution (2 days, 30 °C, 180 rpm). The pre-culture was platted and grown on MS-medium plates (3 days, 30 °C). The grown spores were covered with TSB-medium (4 ml) and scratched off. The spore-mixture was transferred into 300 ml Erlenmeyer flasks filled with TSB-medium (46 ml) and grown for 1 day (28 °C, 180 rpm). In each case 1 ml of this second pre-culture was transferred into 100 ml DNPM-medium in 500 ml Erlenmeyer flasks (42 ×) and cultivated (30 °C, 180 rpm). The feeding started on cultivation-day 2 with the addition of 1 ml alizarin stock-solution (c = 0.25 mmolL-1) and went on during third and fourth day with addition of 500 µl of alizarin stock-solution (c = 0.25 mmolL-1). In total the alizarin concentration per sample reached c = 0.5 mmol/l. The harvest started at day 5. The culture crude was separated into biomass and supernatant via centrifugation (30 min, 7000 rpm). The biomass was decomposed in MeOH (500 ml) evaporated and extracted with EtAc (300 ml). Evaporation of EtAc led to 94.2 mg biomass extract. The supernatant was extracted with EtAc (8 l) and resulted after evaporation in vacuum in 240 mg of an oily residue. The extracts from the feeding experiment were combined and loaded on Amberlite XAD16 column (500 × 3 cm, water). The column was washed with water, water/MeOH 80:20, water/MeOH 60:40 and water/MeOH 40:60 before elution of monoglycosylated alizarin started. The main fraction of monoglycosylated alizarin was eluted with pure MeOH and led to 209 mg pre-fraction which was purified further via semi-preparative HPLC (Nucleodur HTec, 250 × 10 mm, 5µm, A: CH3CN + 0.1 % FA; B: H2O + 0.1 % FA, detection λ = 290 + 328 nm) using a 20 min gradient from 5% A to 95% of A, flow 3 ml/min, injection 100 µl/run (retention time: 16.8 min) The fraction was concentrated and extracted with ethyl acetate. The organic layer was washed with ammonium-formiat solution (50 mM), water (2 ×) and dried with Na2SO4. Fraction 4 led to 54.2 mg of a yellow powder related to glycosylated alizarin. The yellow powder was heated up in EtOH

(10 ml) and hot filtered through a sintered-glass funnel (porosity 4) to get rid of a kind of white insoluble compound. After removing the EtOH the yellow powder was recrystallized in MeOH and the solution was stored in the fridge ~ 2 weeks at 4 °C. Orange needles were achieved (28.4 mg). – Mp. 193-195 °C. – HR-ESI-MS: pos. 387.1085 [M+H] +, 241.0495 [Alizarin+H]+.

Sugar-cleavage of O-α-l-talosyl-1,2-dihydroxyanthraquinone and determination of d- and l-configuration The O-α-l-talosyl-1,2-dihydroxyanthraquinone (23.7 mg) was solved in dry MeOH (17.4 ml). Then Acetyl chloride (4.93 ml) was added and the solution was stirred at room temperature (16 h). The reaction control was done on TLC (SiO2, CHCl3/MeOH 4:1) using Orcin solution as spraying reagent. The finished reaction was evaporated under N2-stream and led to 31 mg of crude product. Column chromatography (SiO2, 1 × 16 cm, CH2Cl2/MeOH 7:4) led to three fractions. Fraction 1: mixture of Alizarin and sugar A, fraction 2: mixture of sugar A and B, fraction: 3 sugar B. Further column chromatography (SiO2, 1 × 16 cm, CH2Cl2/MeOH 92:8) of fraction 1 and 2 with subsequent size exclusion chromatography on sephadex LH20 (1 × 16 cm, MeOH) led in total to 3.9 mg of sugar A, methyl-6-deoxy-α-l-talopyranosid. Fraction 3 was purified via sephadex LH20 (1 × 16 cm, MeOH) and gave 0.4 mg of sugar B, methyl-6-deoxy-β-l-talopyranosid (not pure). – Methyl-6-deoxy-α-l-talopyranosid: [α]20 = -100.85 (c = 2 mg/ml, MeOH), [α]20 = -87.35 (c = 2 mg/ml, H2O). – HR-ESI-MS: pos.179.0911 [M+H]+, 201.0731 [M+Na]+, 379.1571 [2M+Na]+.

Table S1: Physico-chemical properties of amycomycin C (1) and D (2)

	1	2
Appearance (solid) in MeOH (diluted)	orange yellow	violett red
Molecular formula	C31H32O13	C25H22O9
Melting point °C	232	not determined
HR-ESI-MS (m/z) Found [(M+H)+] Calcd [(M+H)+]	 613.1918 613.1916	 467.1335 467.1370
UV λmax (ε) nm (MeOH)	224, 291 (1533), 312 (1727), 420 (612)	not determined
CD: λmax (θ) (MeOH)	213 (+124), 223 (+118), 242 (+77), 275 (+7), 311 (+41), 358 (-303), 406 (+146), 451 (+234), 510 (+100), 567 (+17)	not determined
IR υmax cm-1	3404, 2924, 2856, 2369, 1633, 1459, 1363, 1284, 1235, 1102, 1027, 974, 868, 812	not determined

Figure S1: UV-spectrum of 1 in MeOH (c = 25 mg/ml; 4.083·10-4 mol/l)

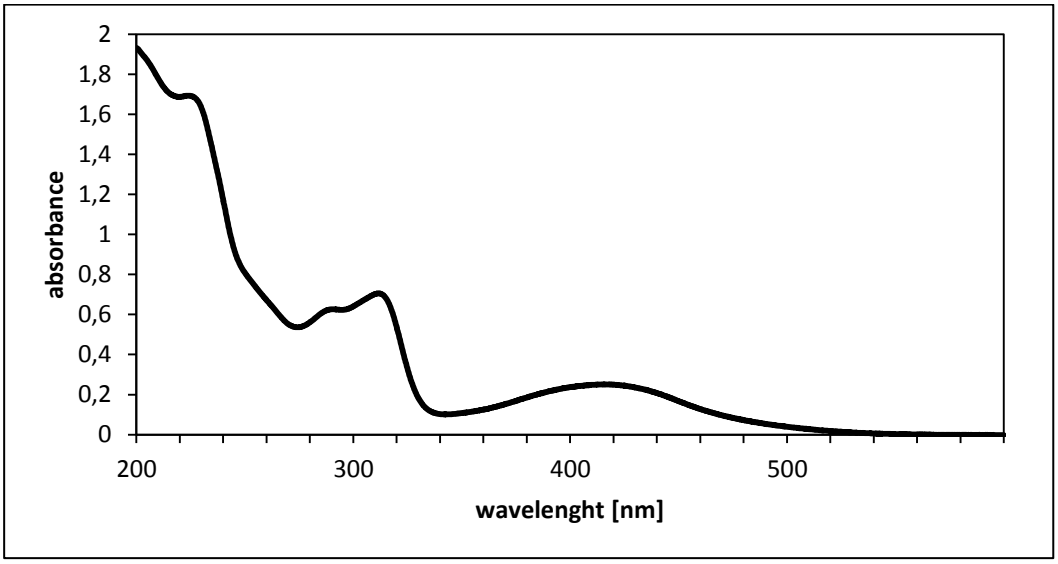


Figure S2: CD-spectrum of 1 in MeOH (c = 2 mg/ml; 3.27·10-3 mol/l)

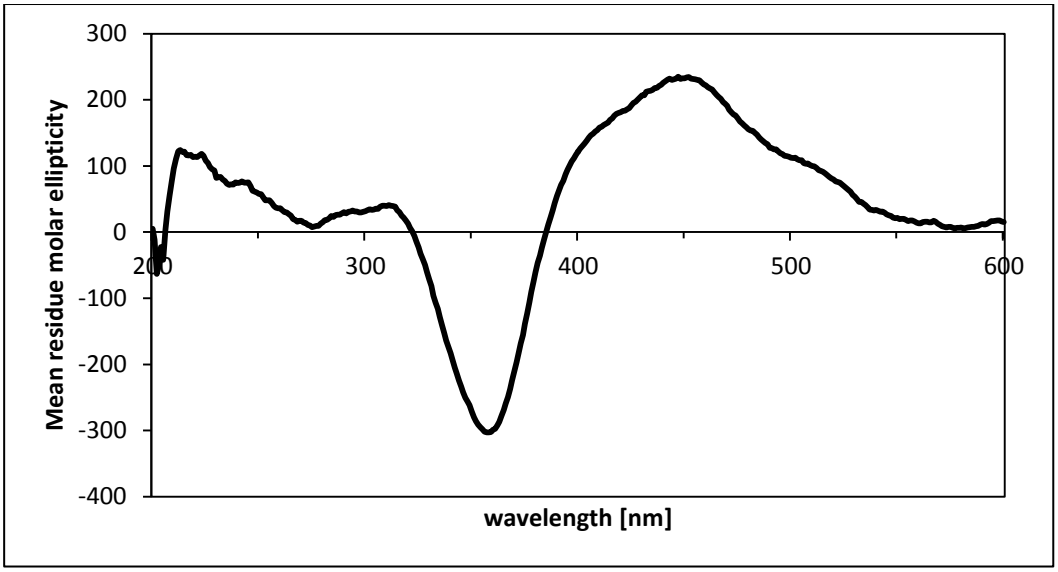
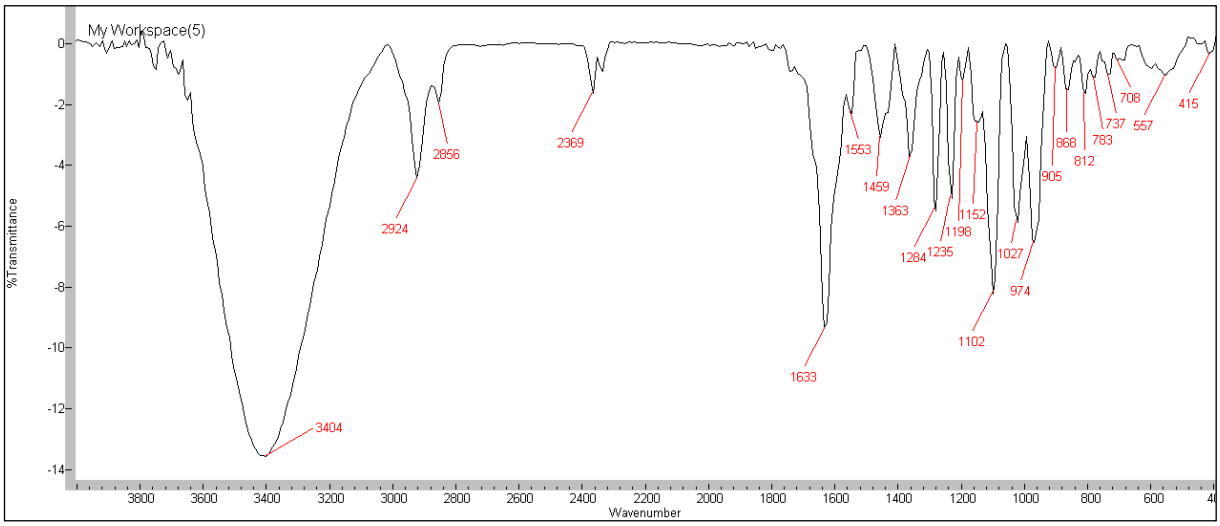


Figure S3: IR-spectrum of 1



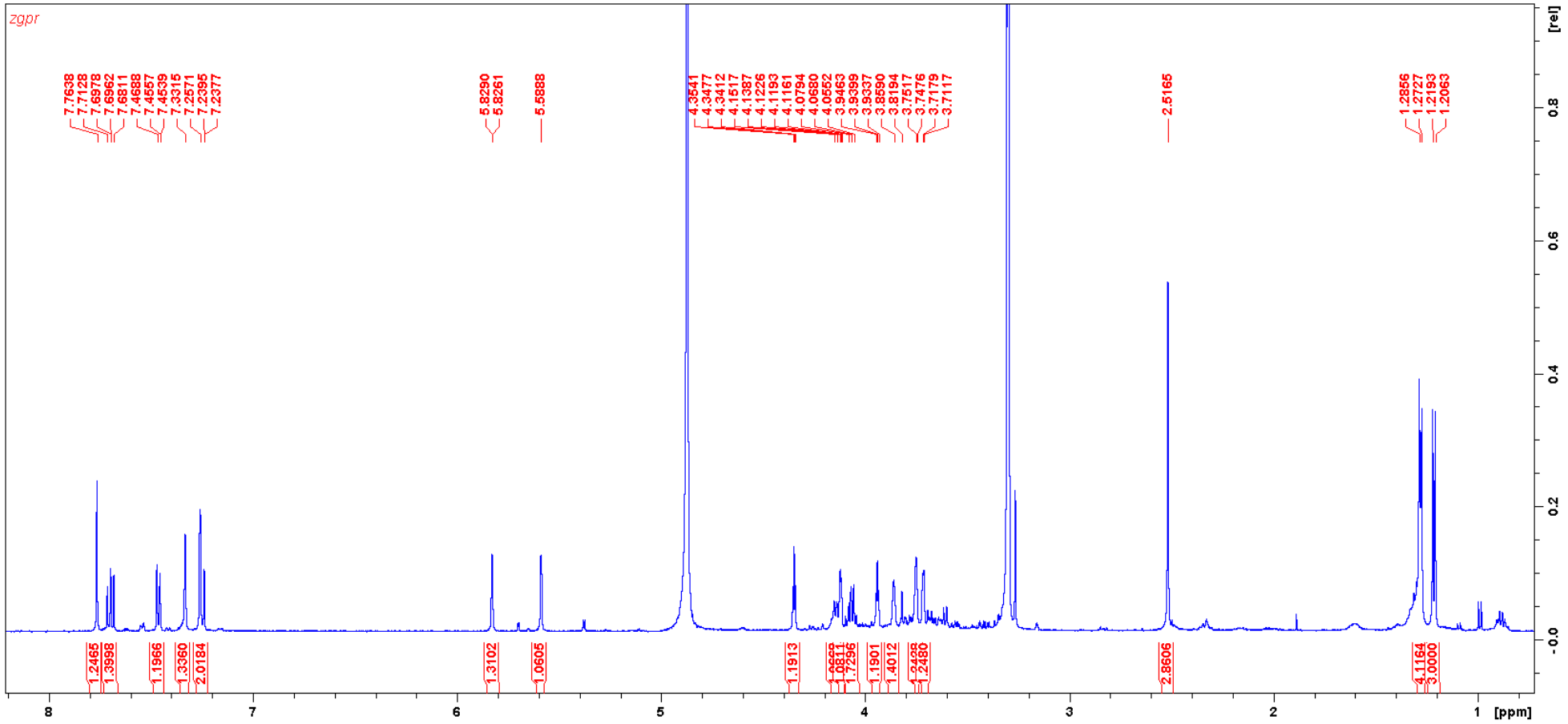


Figure S4. 1H NMR (500 MHz, 25 ºC, MeOD) spectrum of amycomycin C (1)

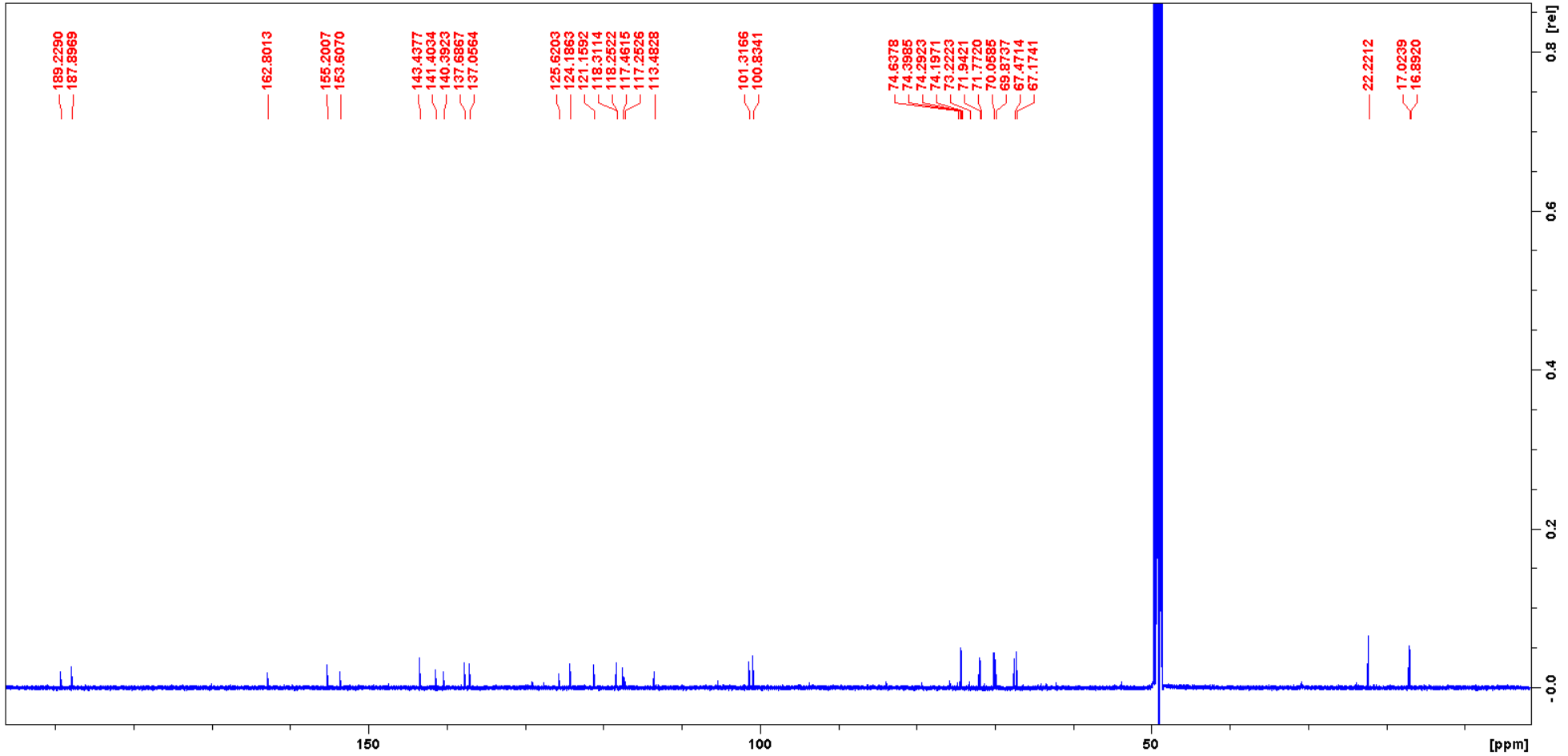


Figure S5. 13C NMR (500 MHz, 25 ºC, MeOD) spectrum of amycomycin C (1)

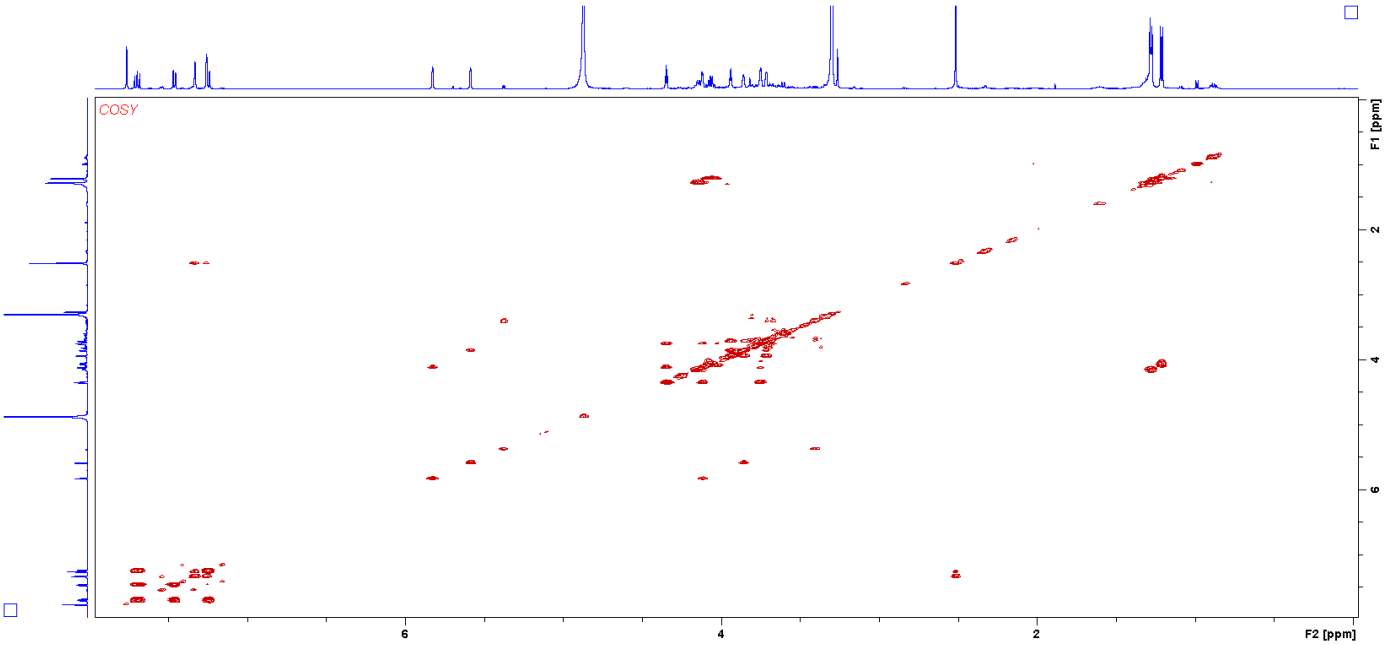


Figure S6. COSY (500 MHz, 25 ºC, MeOD) spectrum of amycomycin C (1)

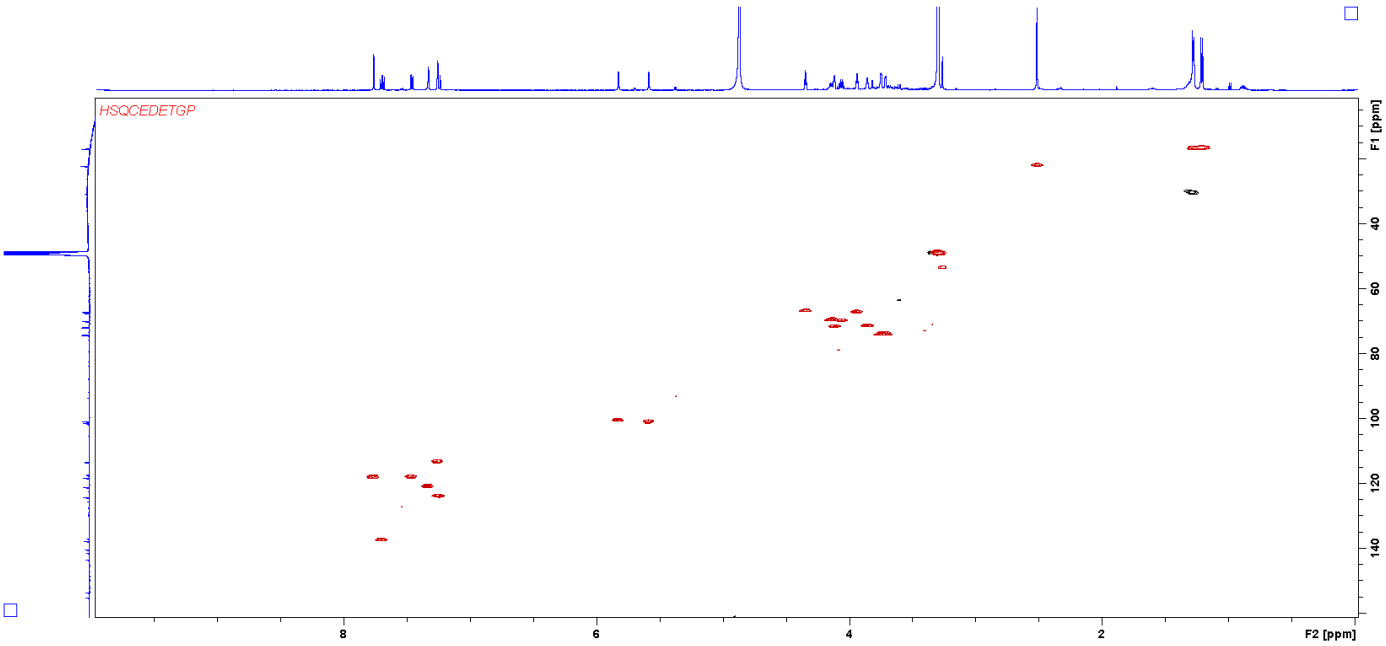


Figure S7. HSQC (500 MHz, 25 ºC, MeOD) spectrum of amycomycin C (1)

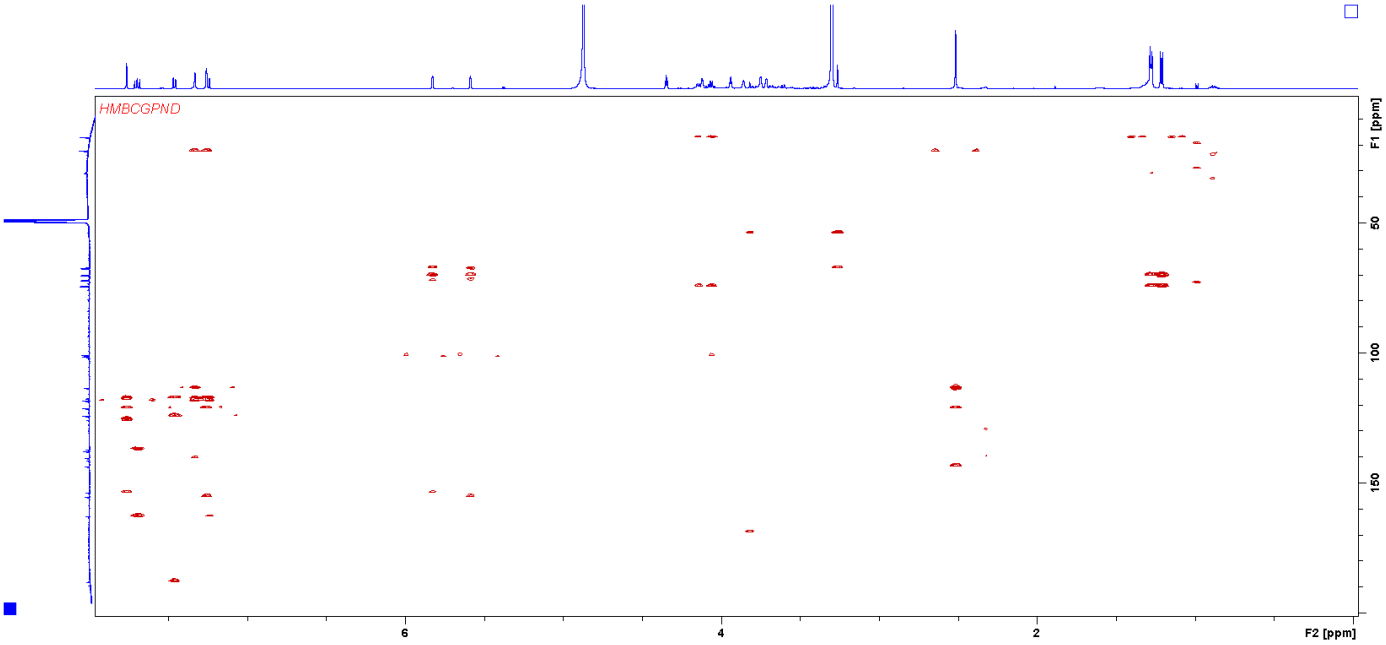


Figure S8. HMBC (500 MHz, 25 ºC, MeOD) spectrum of amycomycin C (1)

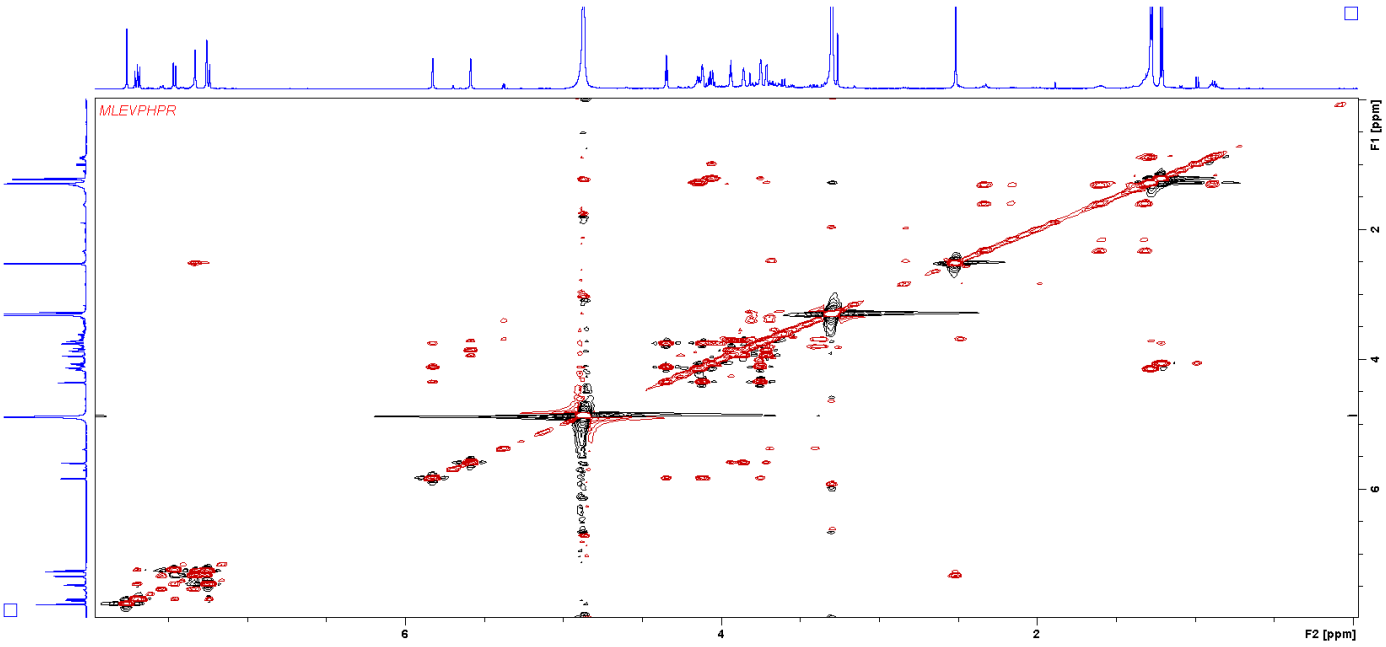


Figure S9. TOCSY (500 MHz, 25 ºC, MeOD) spectrum of amycomycin C (1)

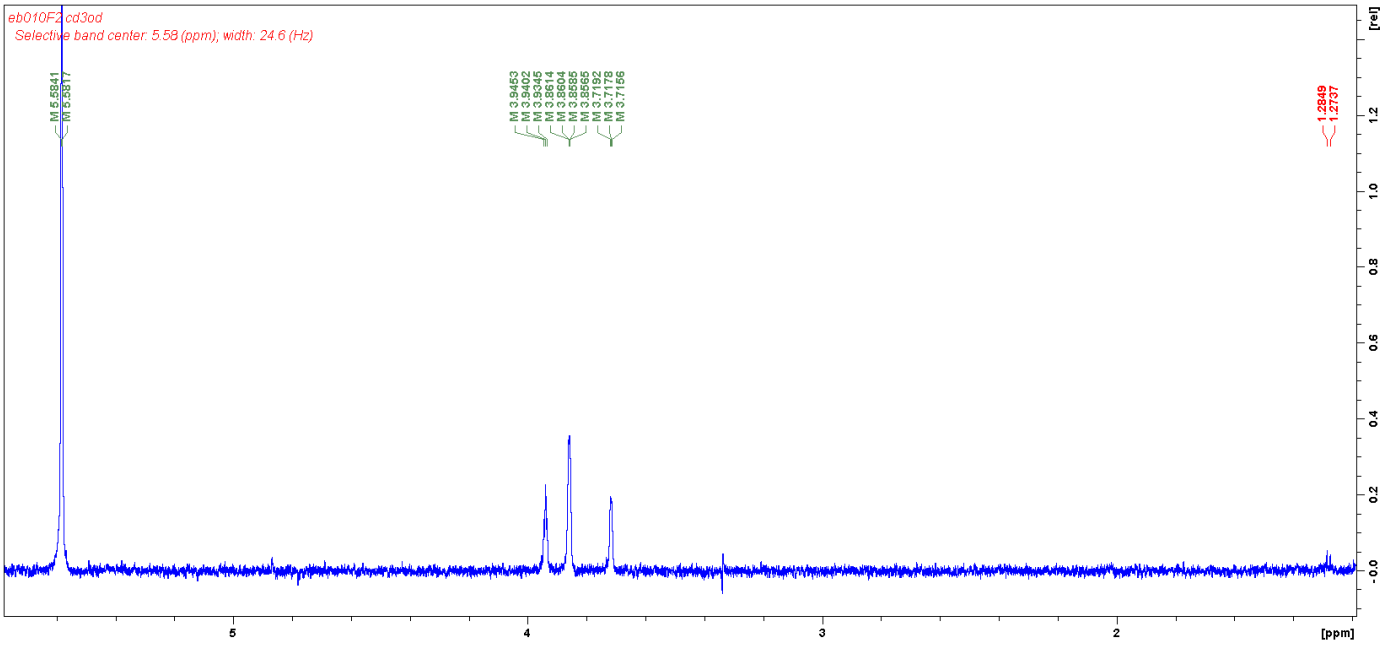


Figure S10. 1D TOCSY (600 MHz, 35 ºC, MeOD) spectrum of amycomycin C (1), excitation of 5.58 ppm

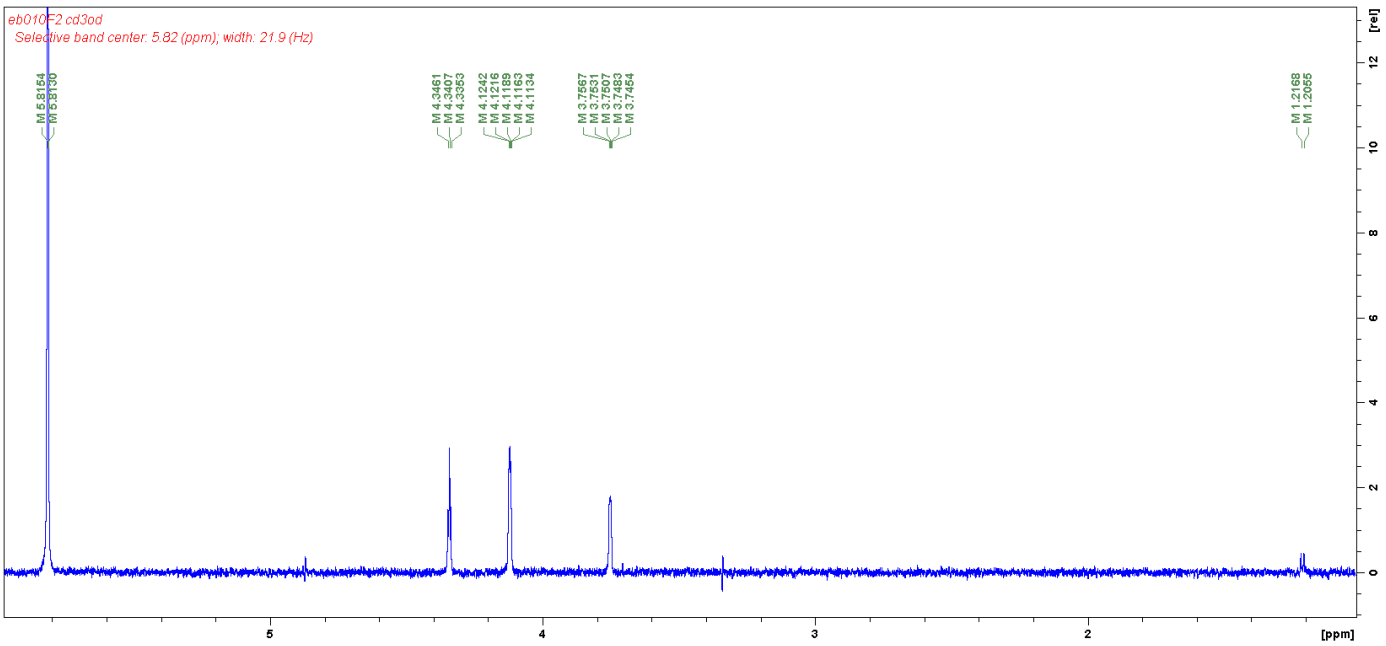


Figure S11. 1D TOCSY (600 MHz, 35 ºC, MeOD) spectrum of amycomycin C (1), excitation of 5.82 ppm

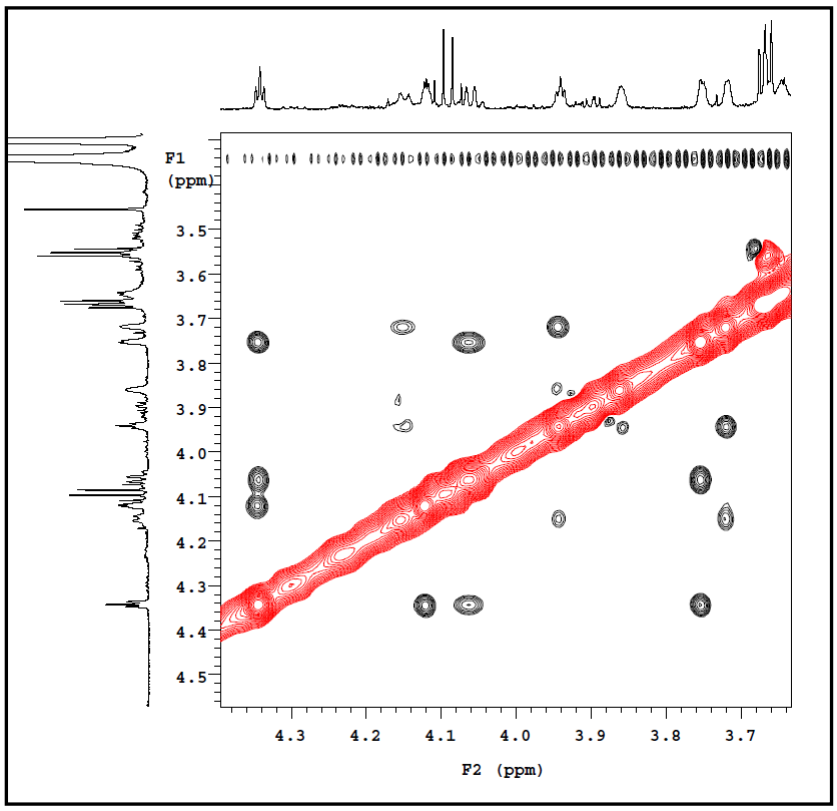


Figure S12. band selective ROESY (600 MHz, 35 ºC, MeOD) spectrum of amycomycin C (1)

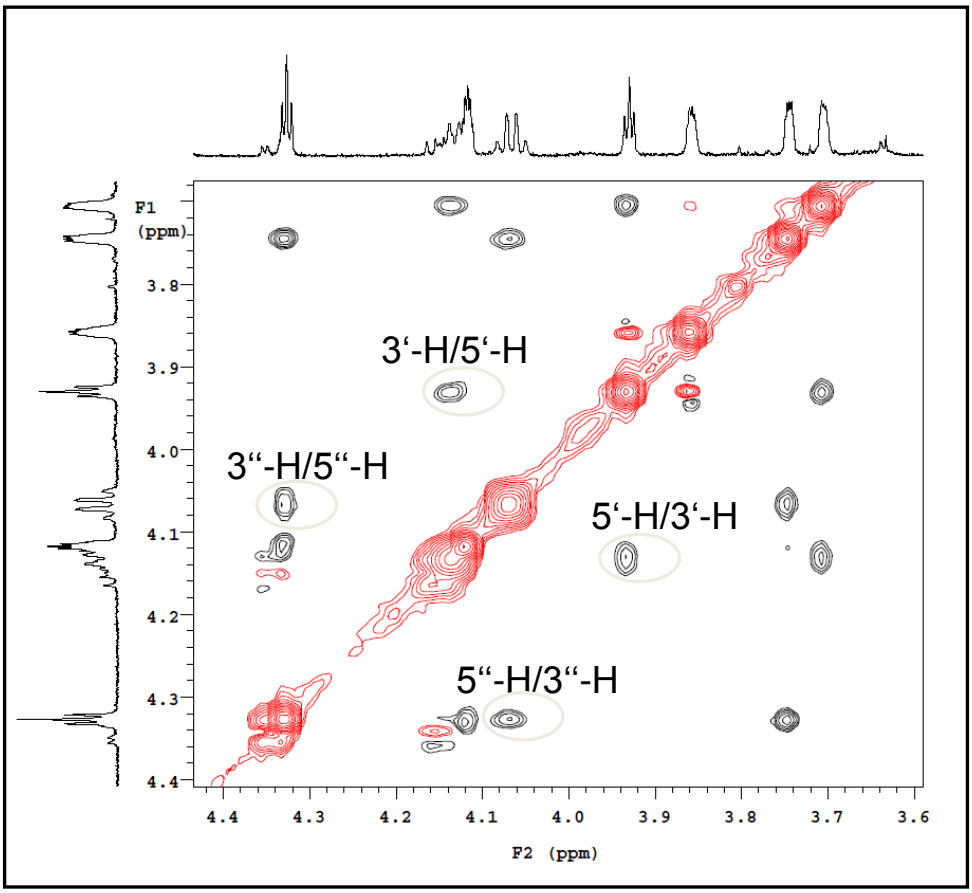


Figure S13. band selective NOESY (600 MHz, 35 ºC, MeOD) spectrum of amycomycin C (1)

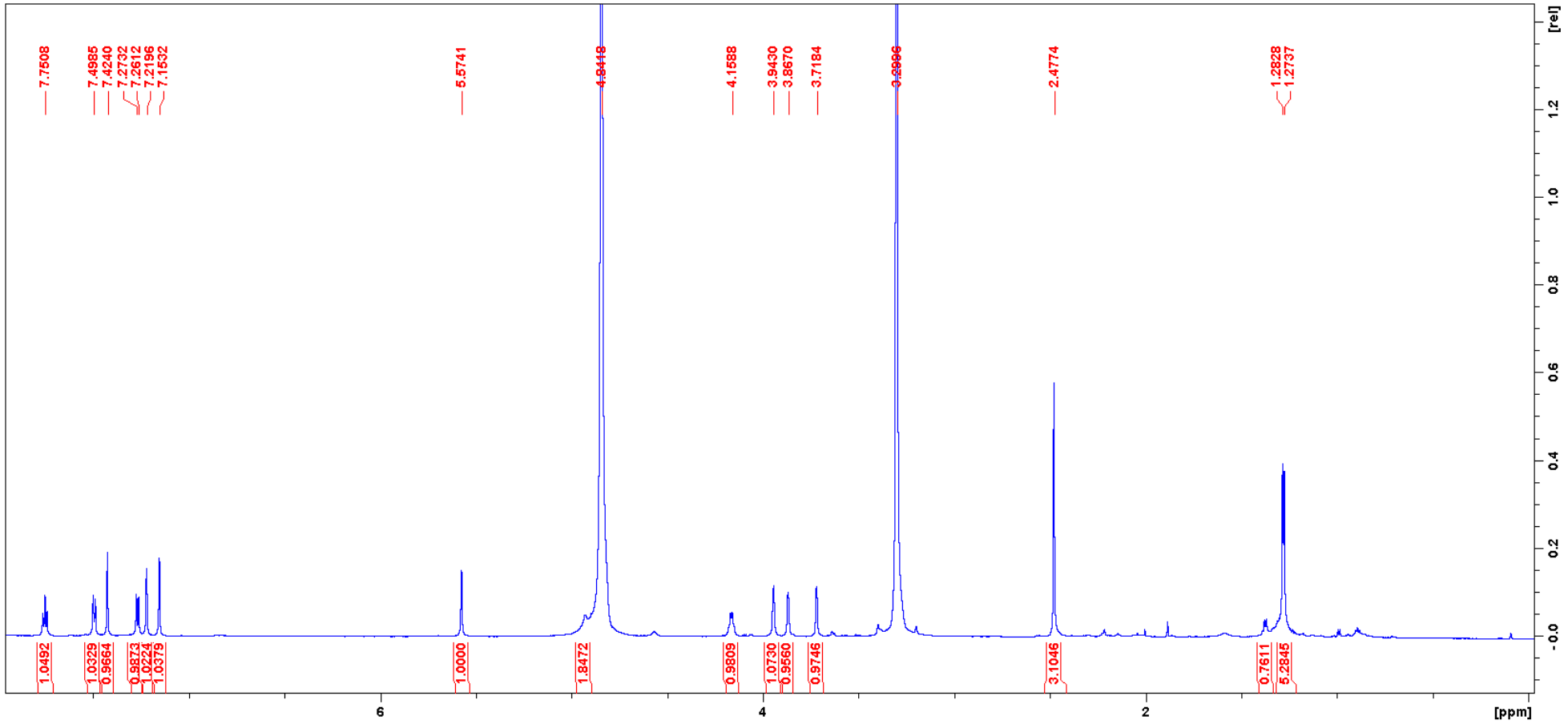


Figure S14. 1H NMR (500 MHz, 25 ºC, MeOD) spectrum of amycomycin D (2)

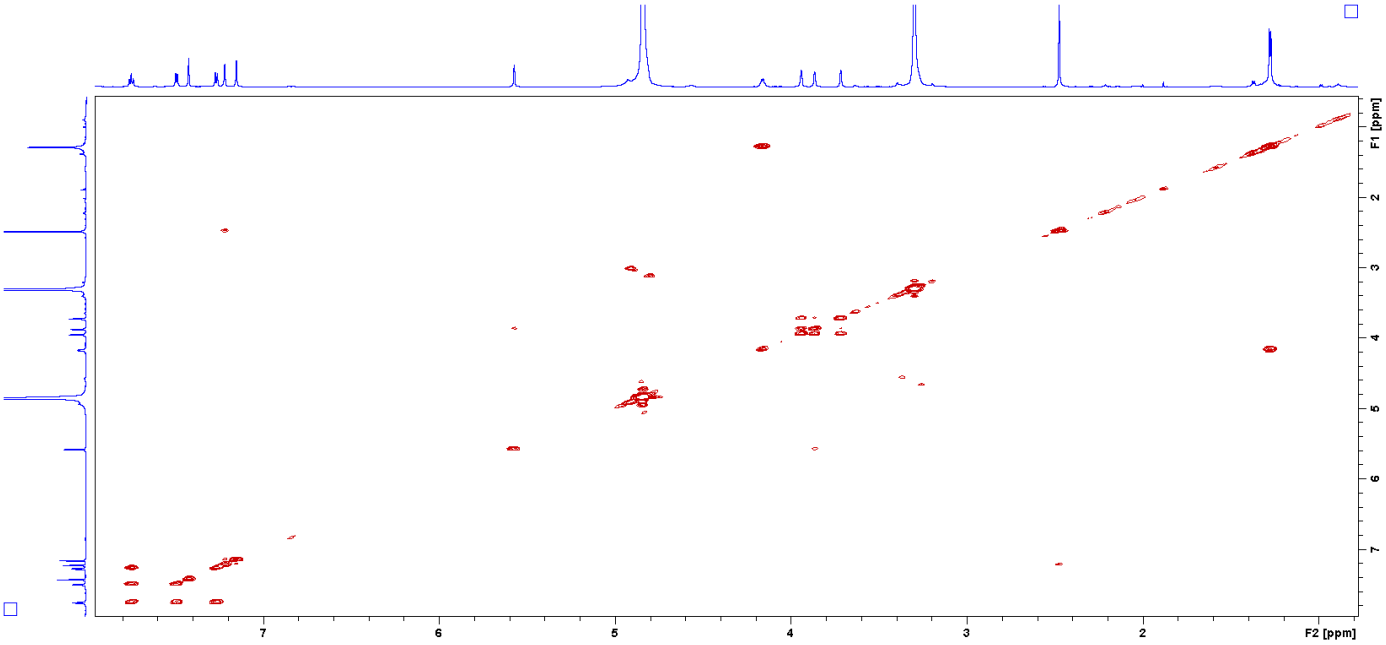


Figure S15. COSY (500 MHz, 25 ºC, MeOD) spectrum of amycomycin D (2)

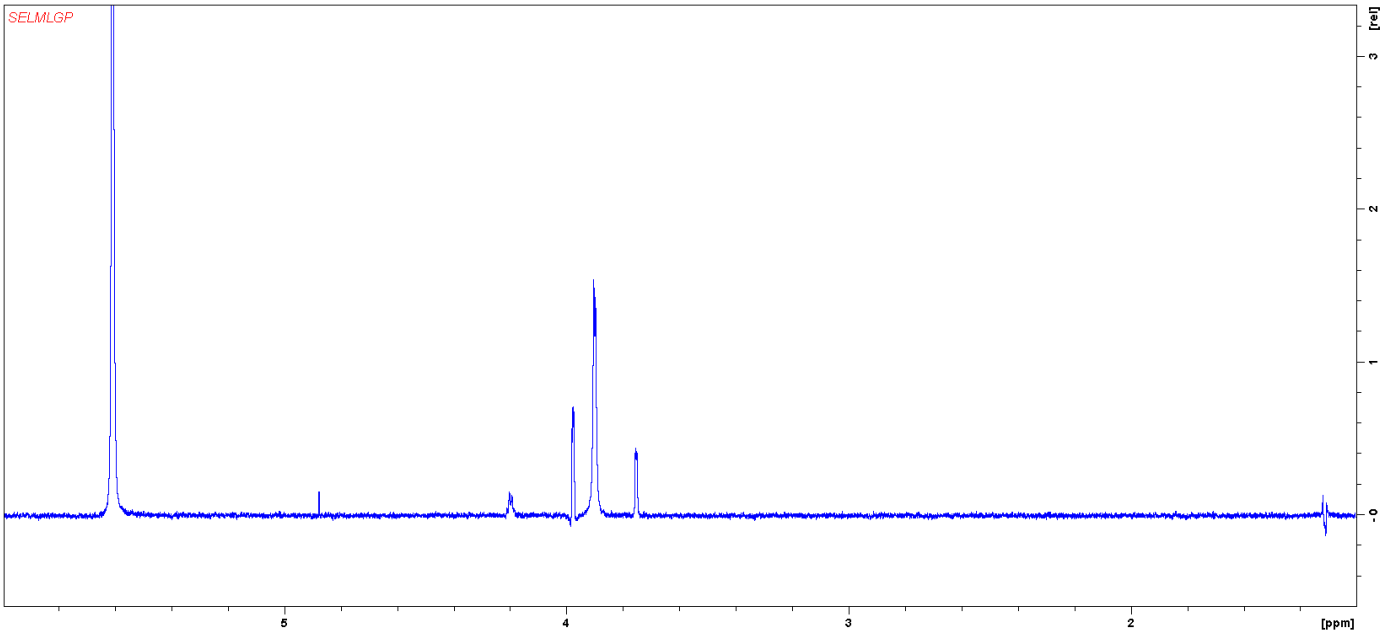


Figure S16. 1D-TOCSY (500 MHz, 25 ºC, MeOD) spectrum of amycomycin D (2), excitation of 5.57 ppm

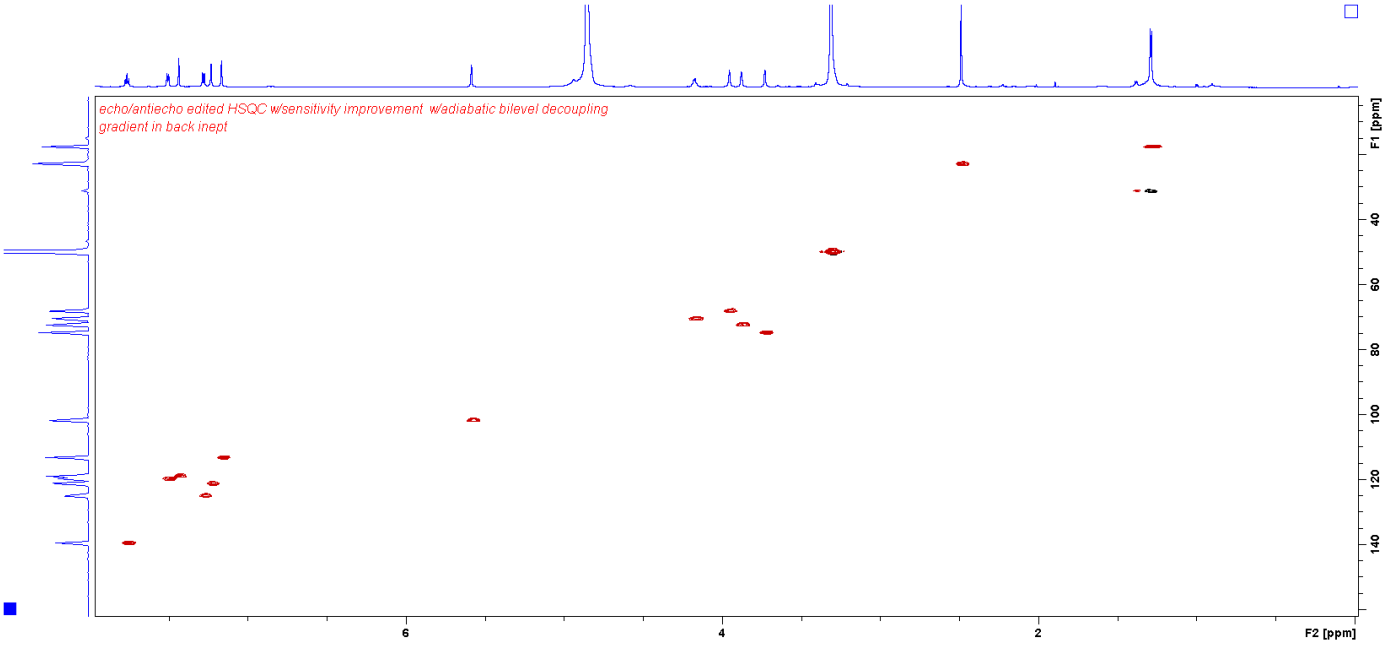


Figure S17. HSQC (500 MHz, 25 ºC, MeOD) spectrum of amycomycin D (2)

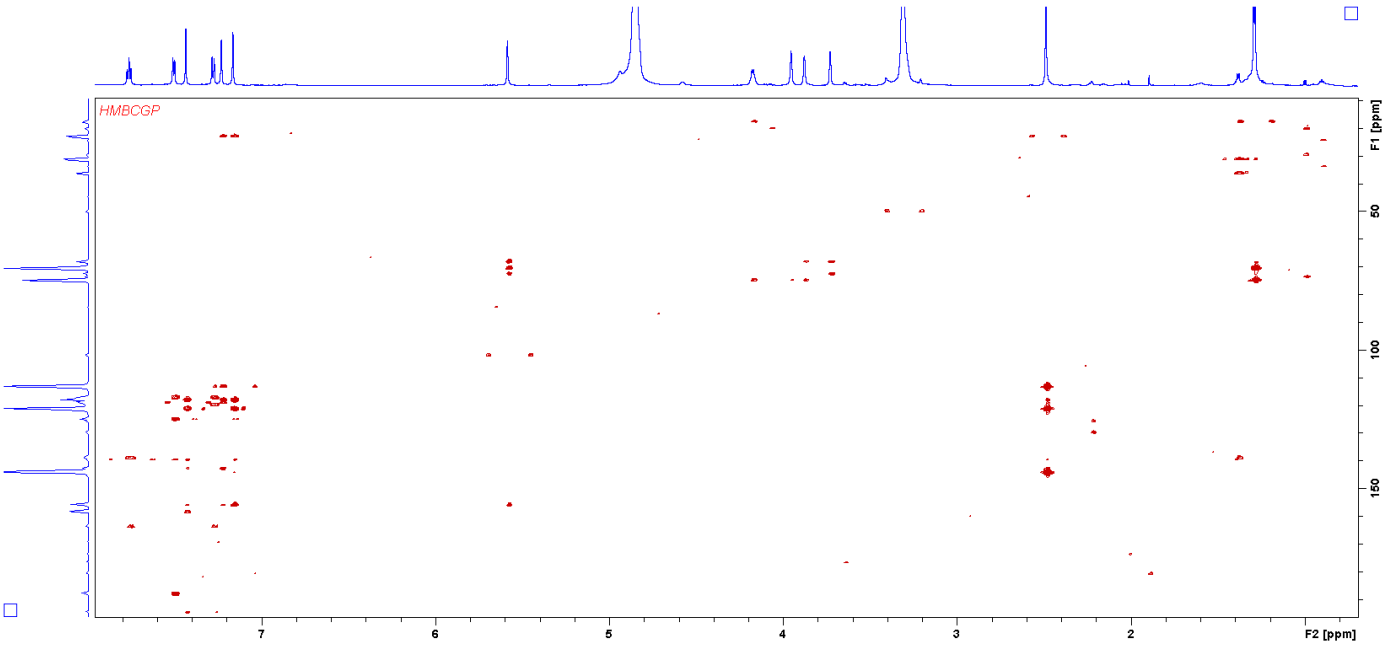


Figure S18. HMBC (500 MHz, 25 ºC, MeOD) spectrum of amycomycin D (2)

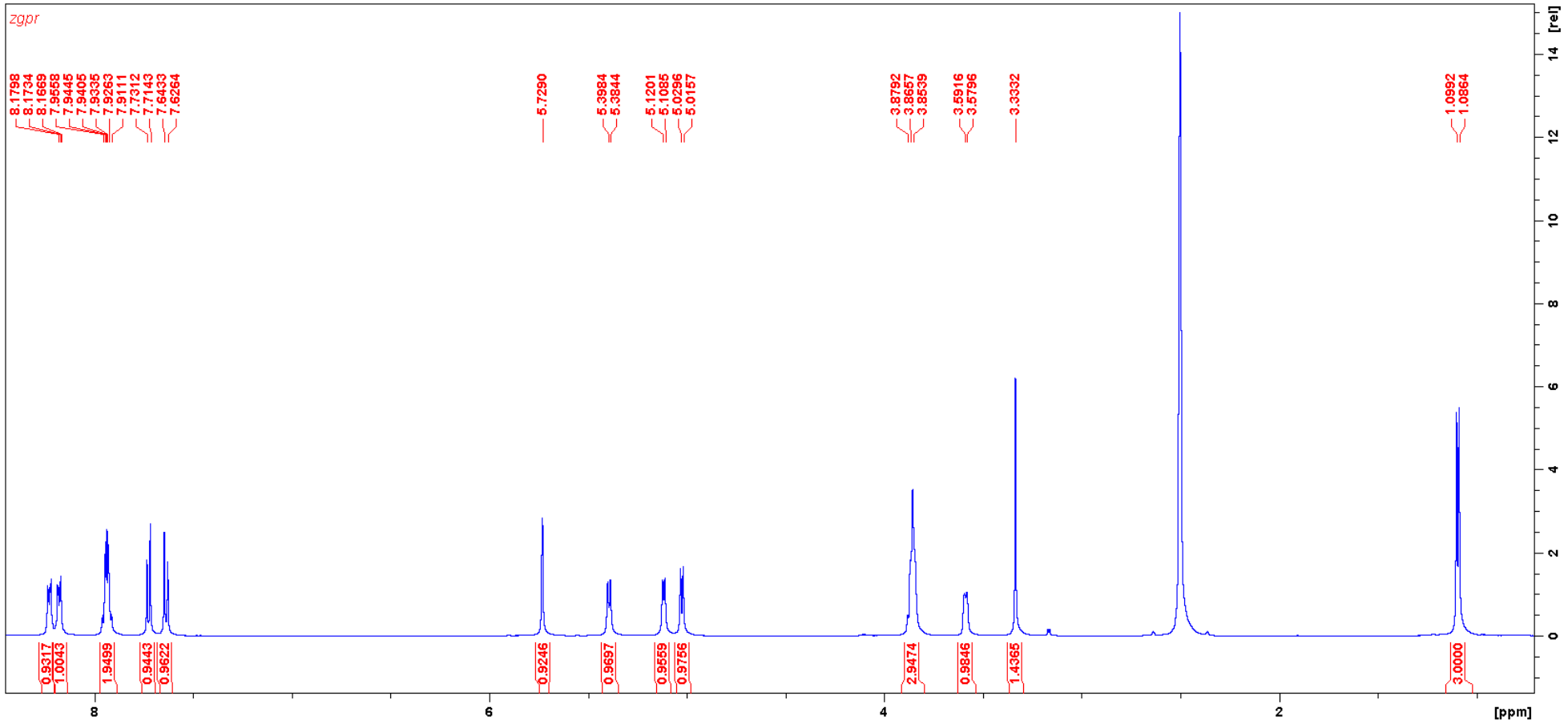


Figure S19. 1H NMR (500 MHz, 25 ºC, MeOD) spectrum of O-α-l-talosyl-1,2-dihydroxyantharquinone (3)

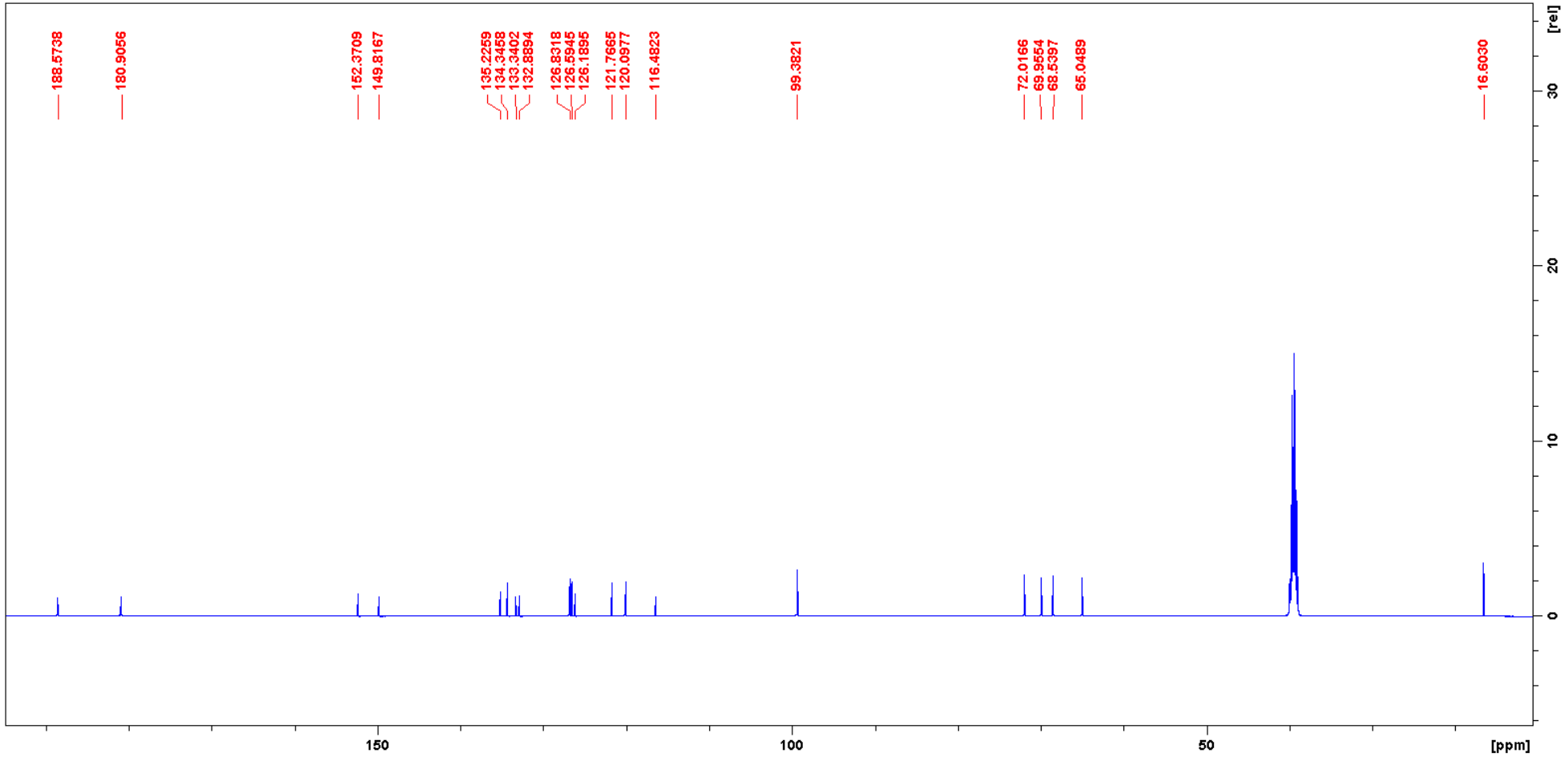


Figure S20. 13C NMR (500 MHz, 25 ºC, MeOD) spectrum of O-α-l-talosyl-1,2-dihydroxyantharquinone (3)

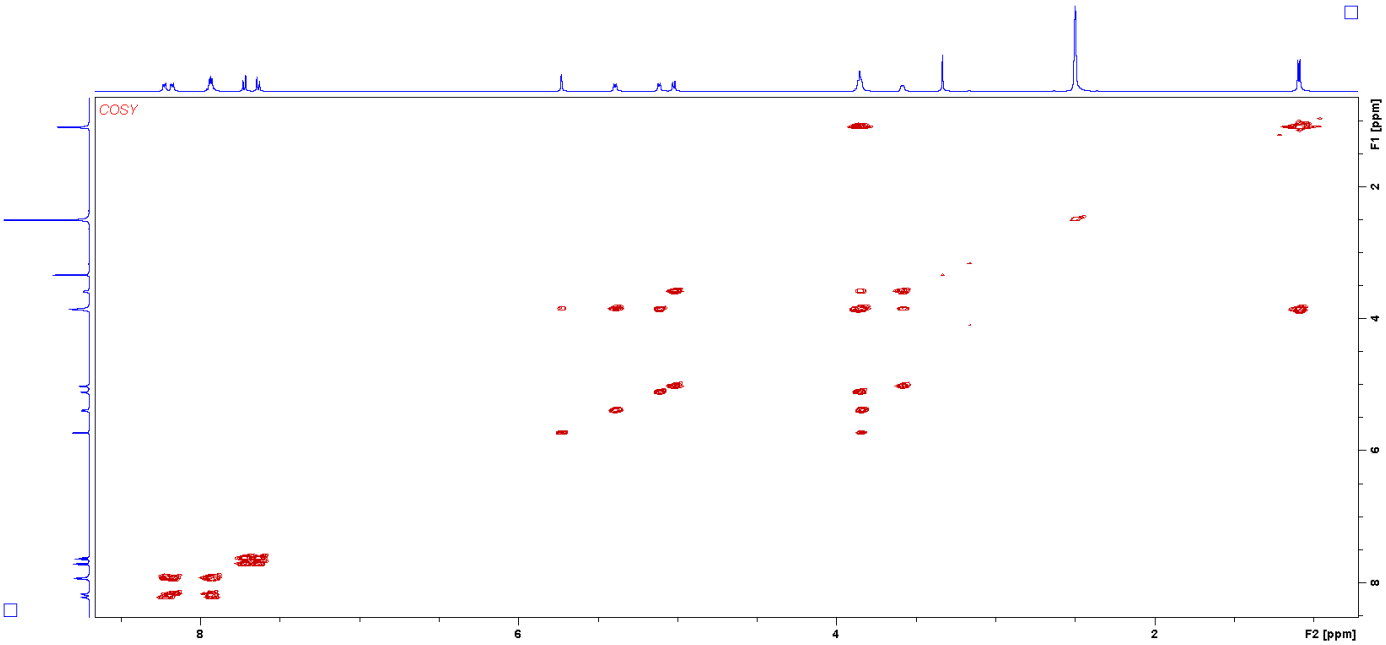


Figure S21. COSY (500 MHz, 25 ºC, MeOD) spectrum of O-α-l-talosyl-1,2-dihydroxyantharquinone (3)

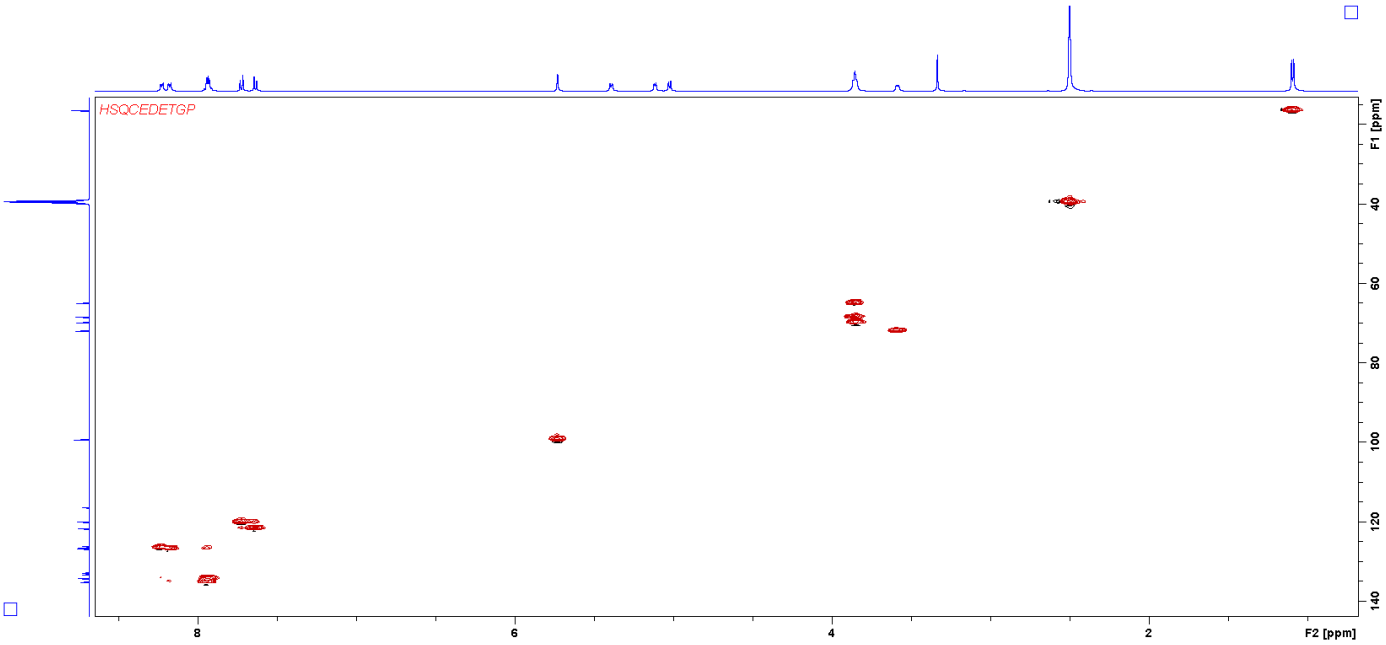


Figure S22. HSQC (500 MHz, 25 ºC, MeOD) spectrum of O-α-l-talosyl-1,2-dihydroxyantharquinone (3)

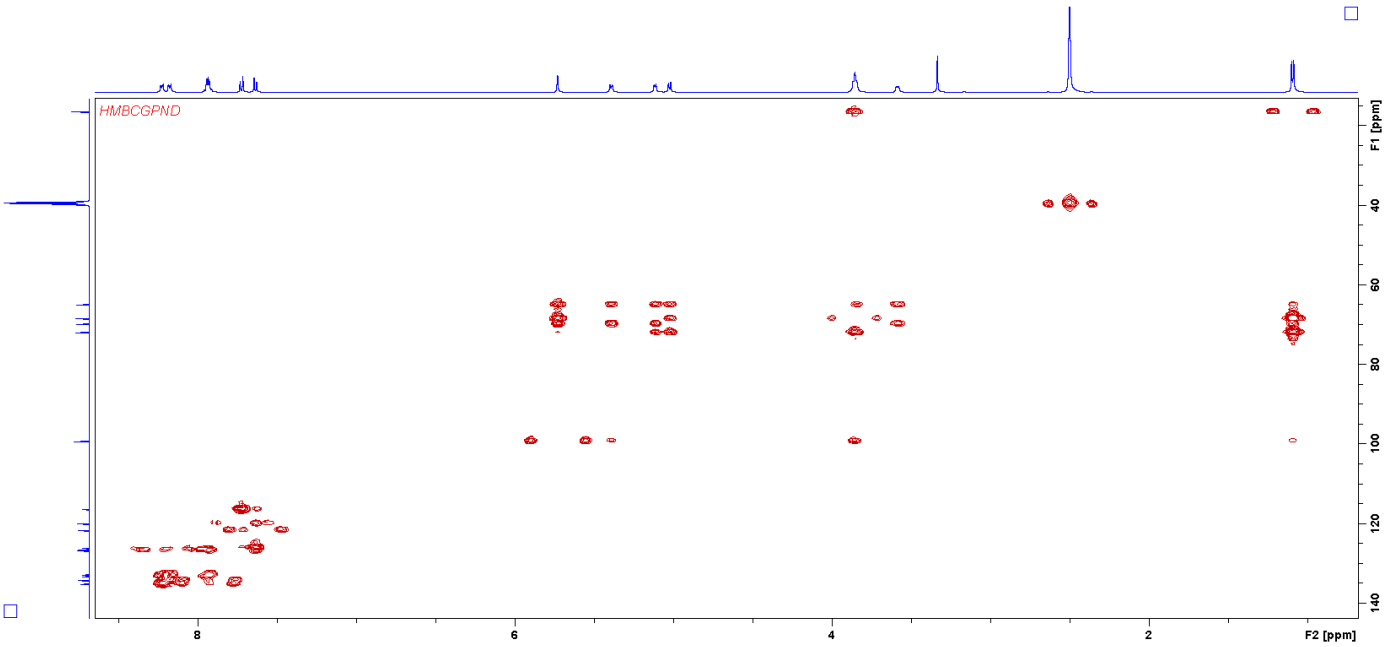


Figure S23. HMBC (500 MHz, 25 ºC, MeOD) spectrum of O-α-l-talosyl-1,2-dihydroxyantharquinone (3)

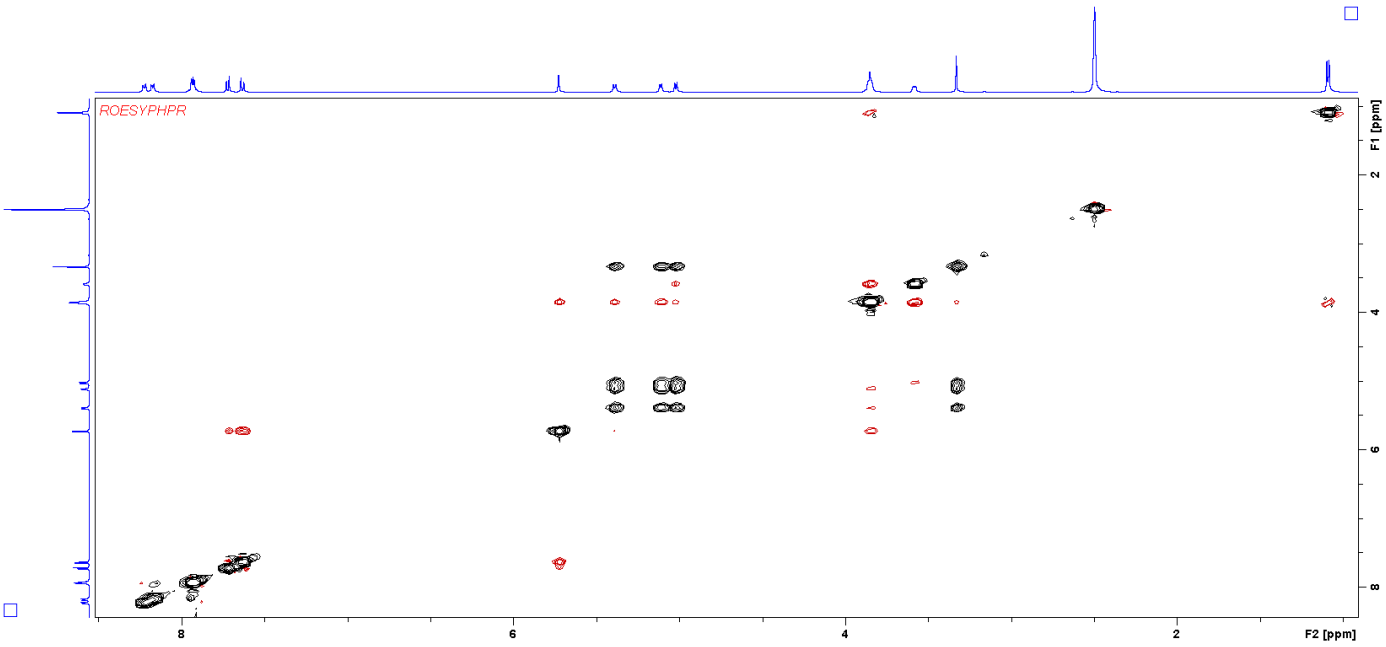


Figure S24. ROESY (500 MHz, 25 ºC, MeOD) spectrum of O-α-l-talosyl-1,2-dihydroxyantharquinone (3)

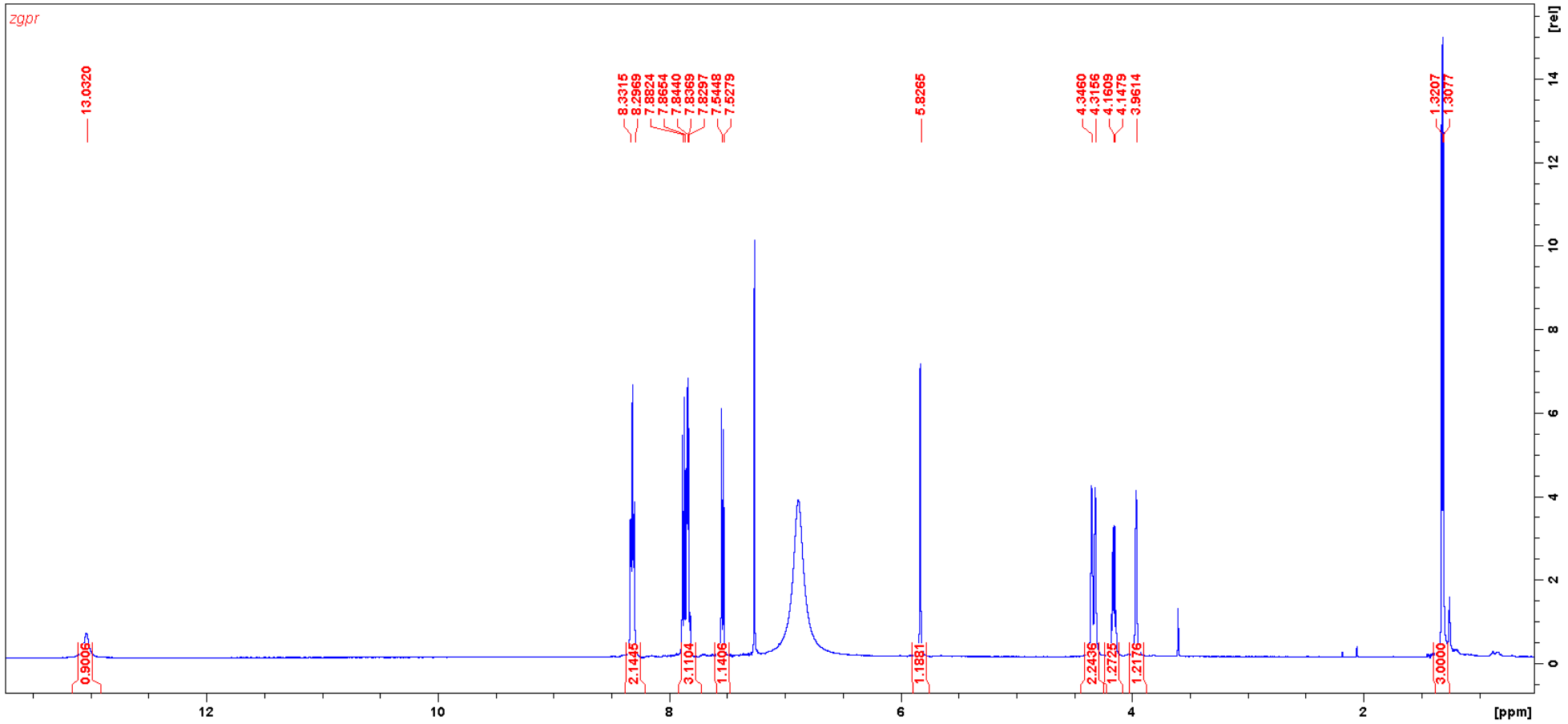


Figure S25. 1H NMR (500 MHz, 25 ºC, CDCl3+CF3COOD) spectrum O-α-l-talosyl-1,2-dihydroxyantharquinone (3)

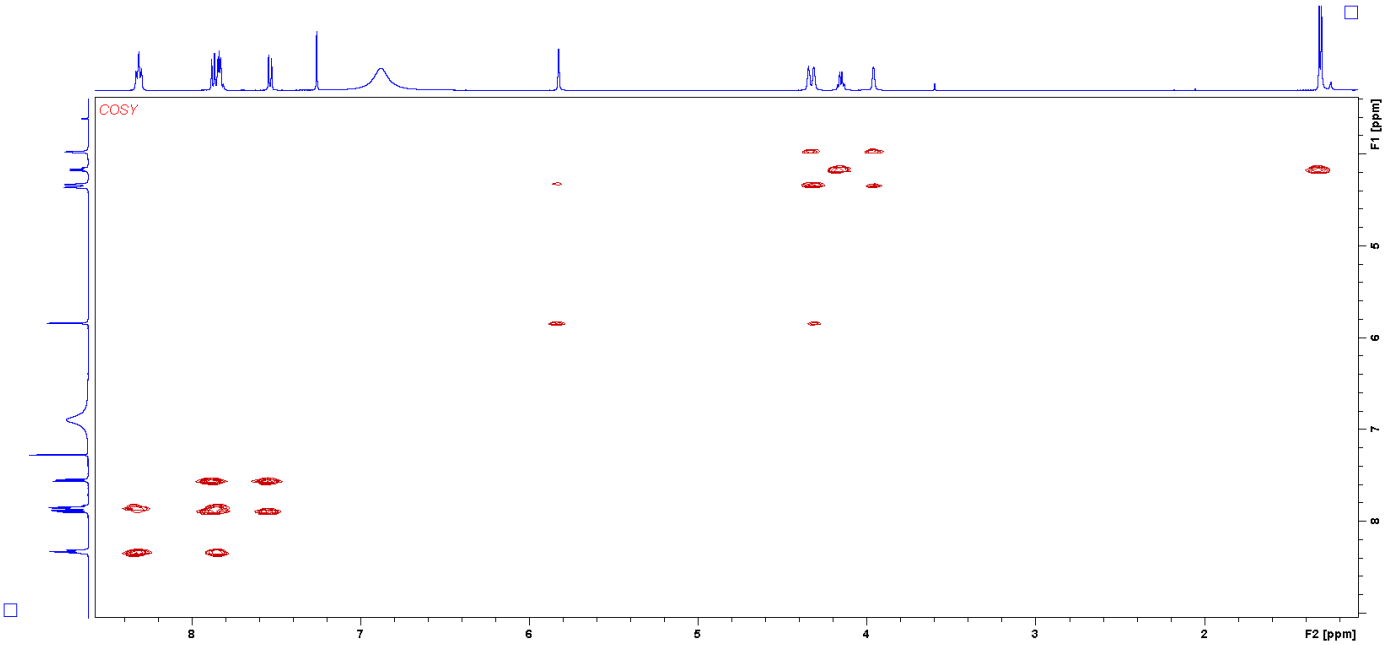


Figure S26. COSY (500 MHz, 25 ºC, CDCl3+CF3COOD) spectrum of O-α-l-talosyl-1,2-dihydroxyantharquinone (3)

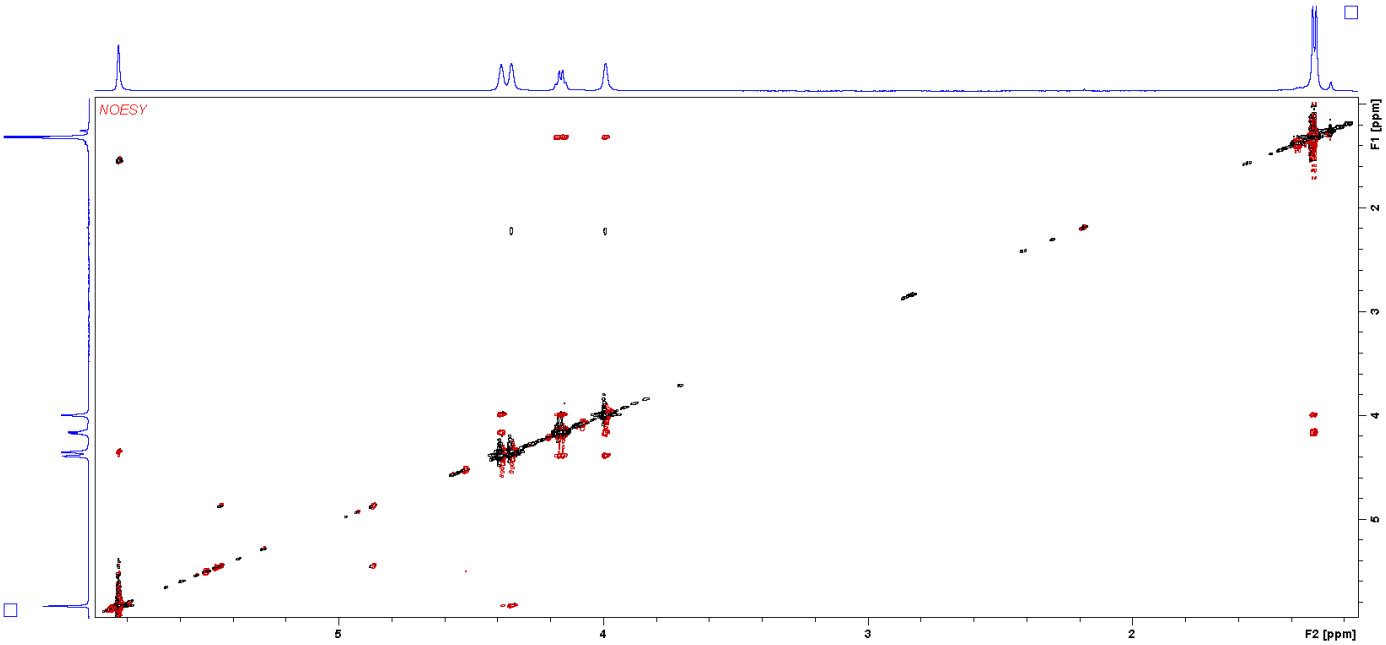


Figure S27. NOESY (500 MHz, 25 ºC, CDCl3+CF3COOD) spectrum of O-α-l-talosyl-1,2-dihydroxyantharquinone (3)

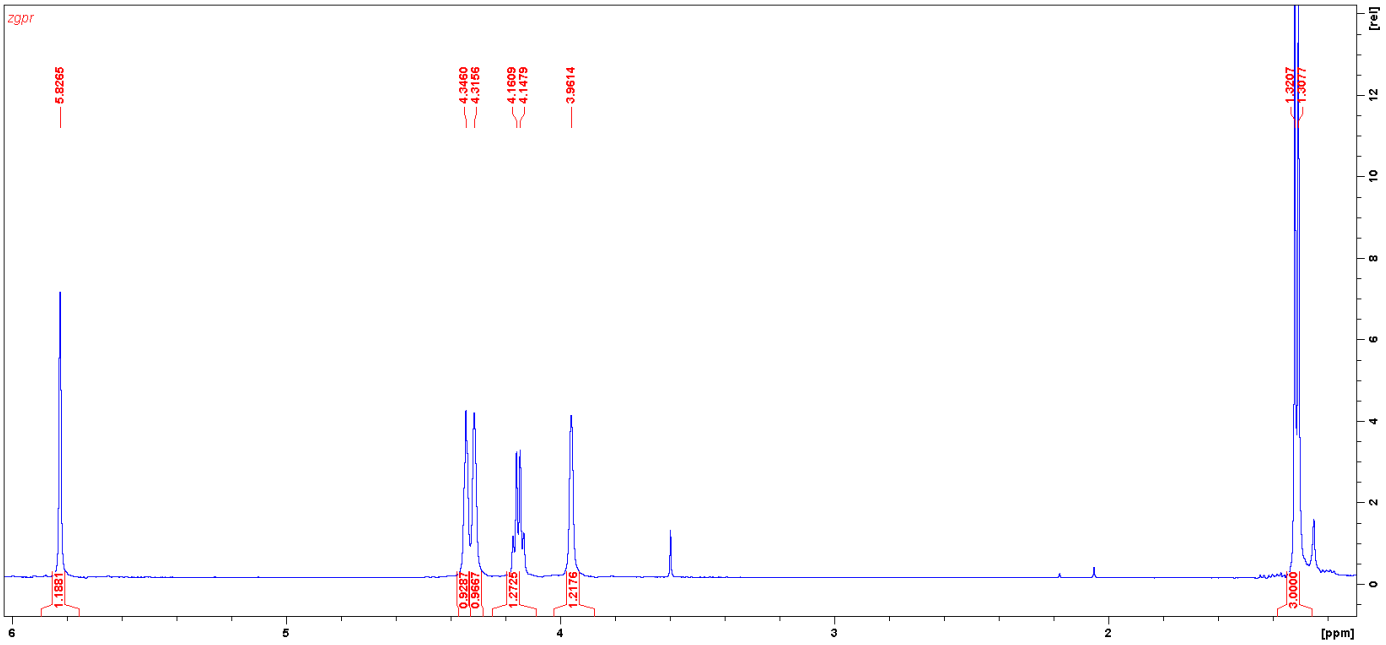


Figure S28. Expanded 1H NMR (500 MHz, 25 ºC, CDCl3+CF3COOD) spectrum O-α-l-talosyl-1,2-dihydroxyantharquinone (3)

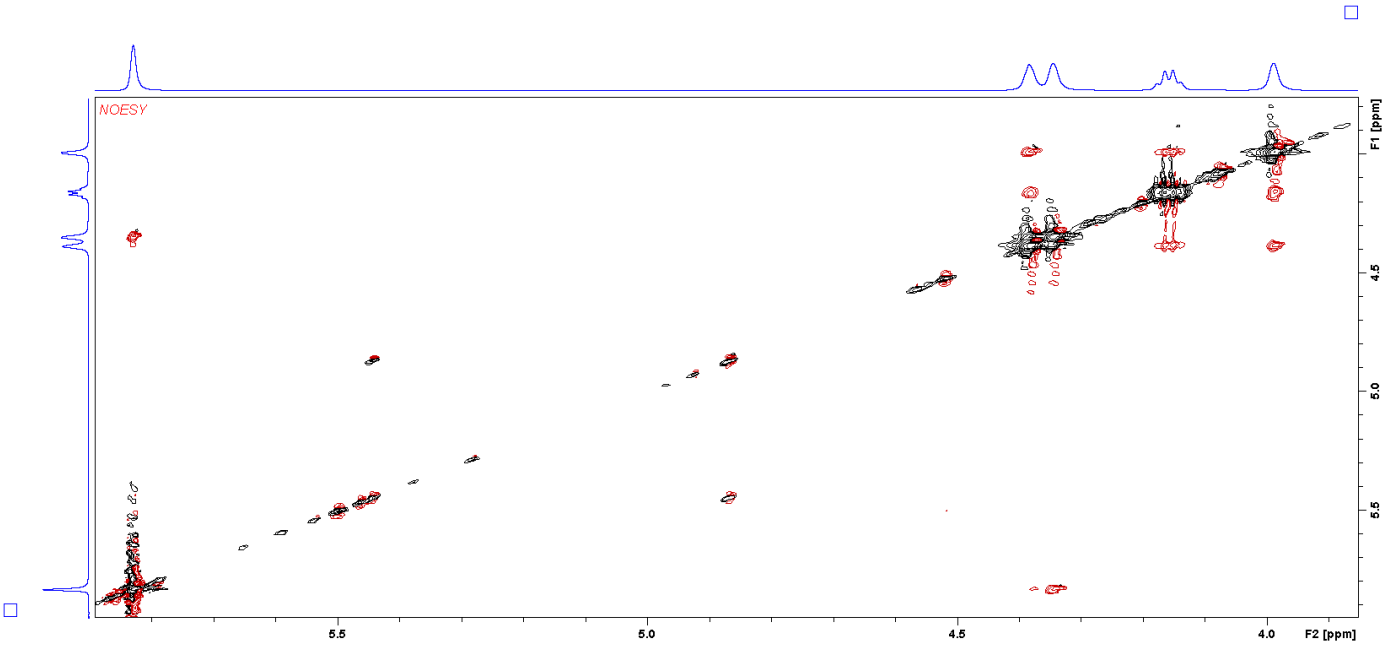


Figure S29. Expanded NOESY (500 MHz, 25 ºC, CDCl3+CF3COOD) spectrum of O-α-l-talosyl-1,2-dihydroxyantharquinone (3)

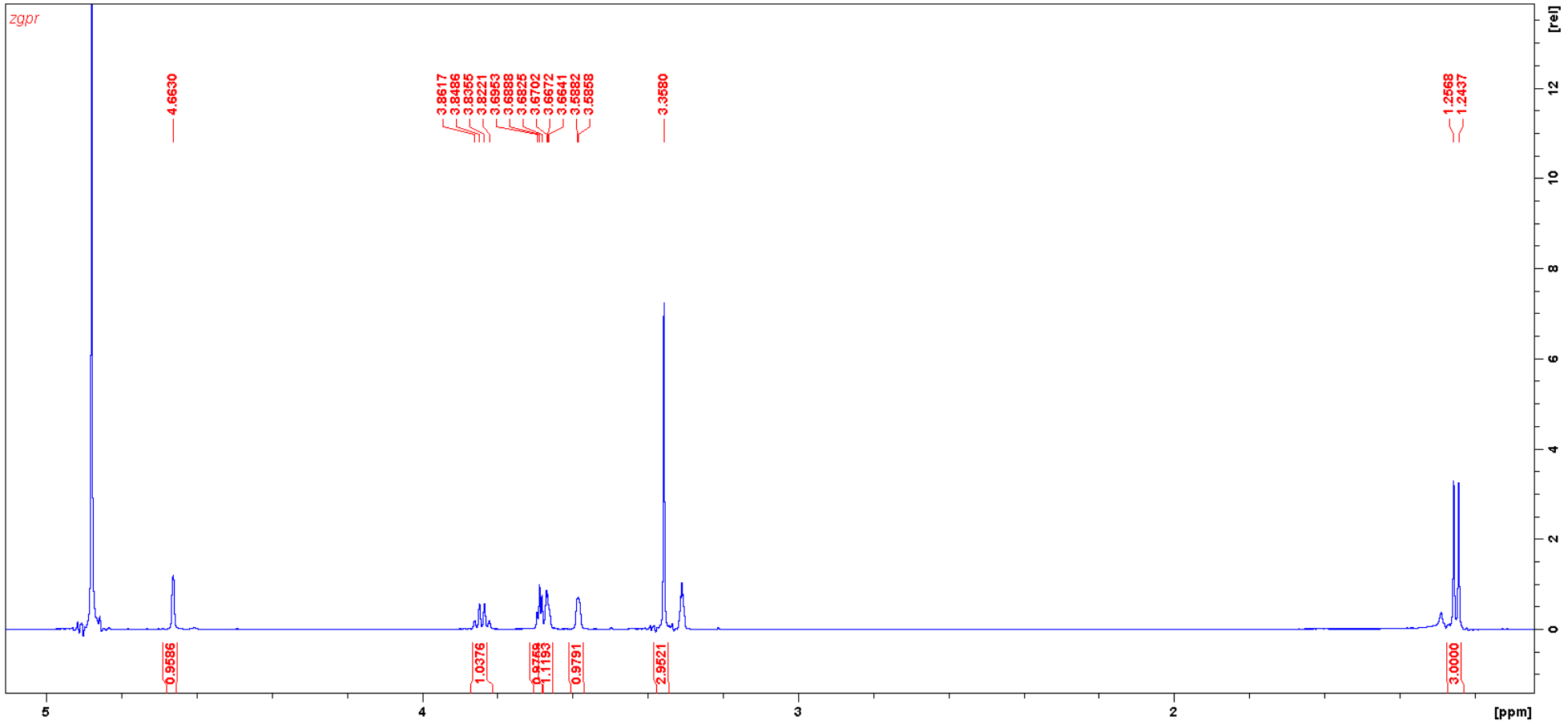


Figure S30. 1H NMR (500 MHz, 25 ºC, MeOD) spectrum of methyl-6-deoxy-α-l-talopyranosid

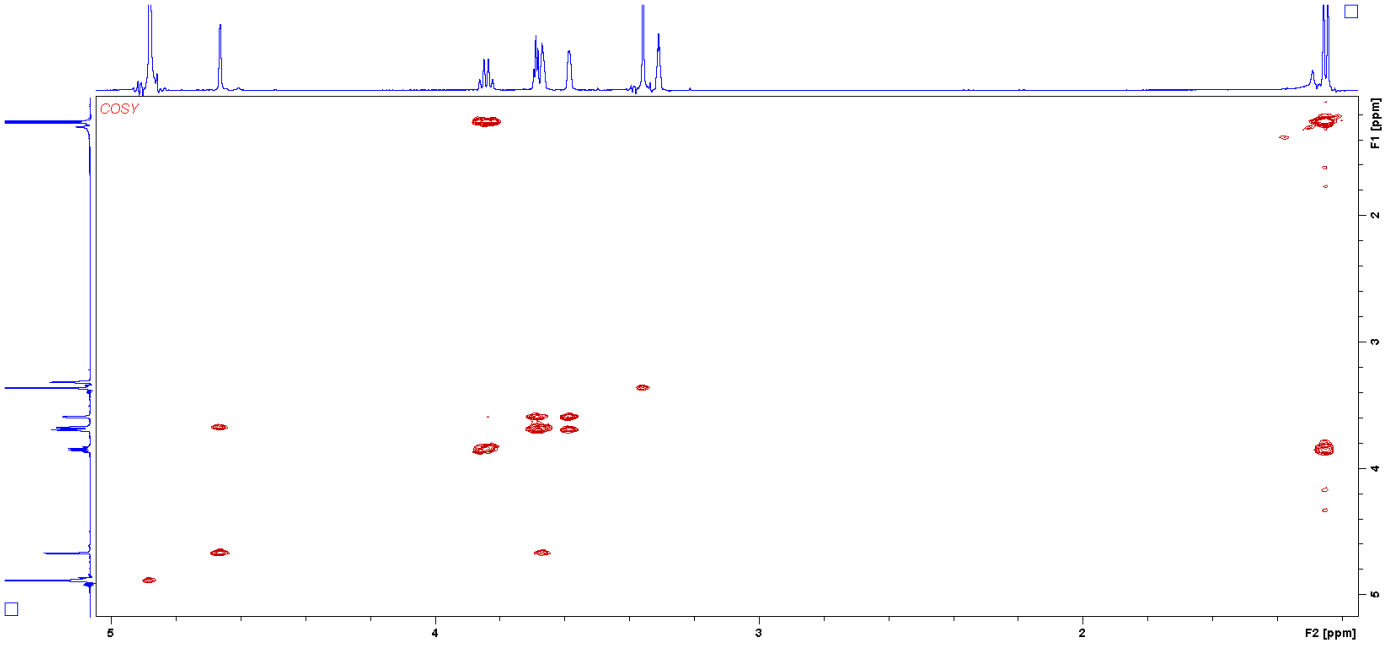


Figure S31. COSY (500 MHz, 25 ºC, MeOD) spectrum of methyl-6-deoxy-α-l-talopyranosid

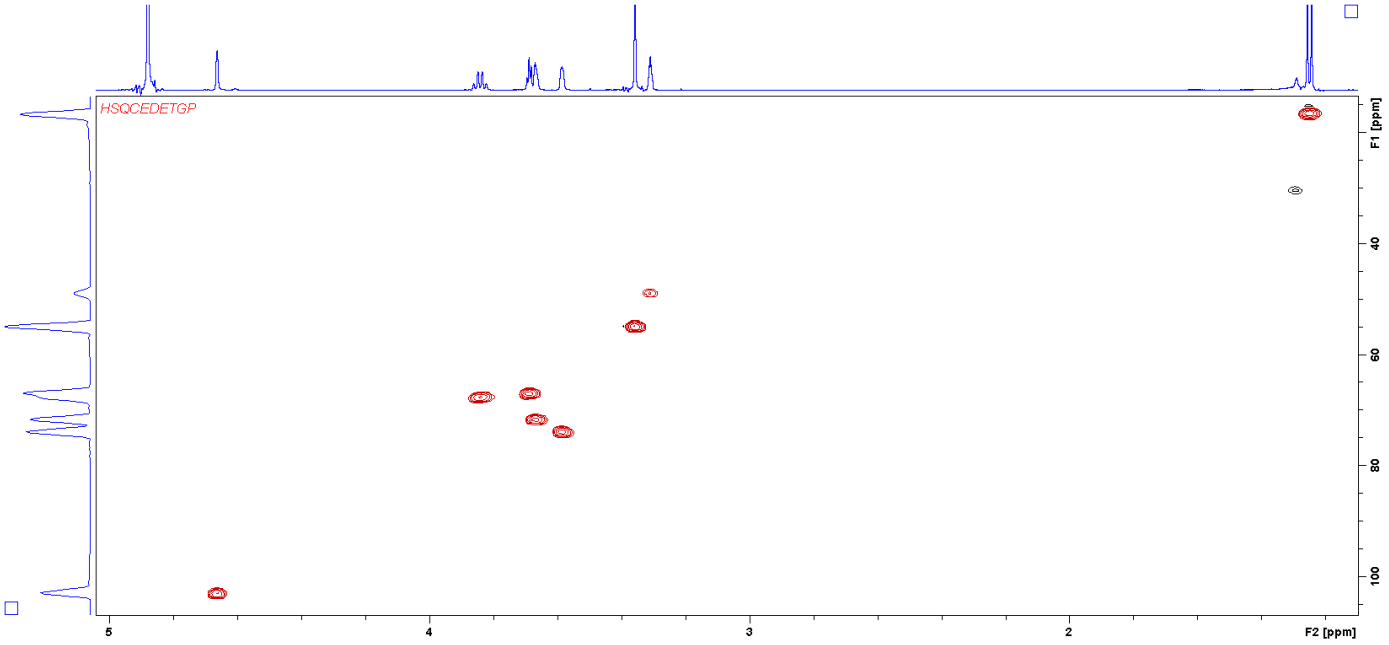


Figure S32. HSQC (500 MHz, 25 ºC, MeOD) spectrum of methyl-6-deoxy-α-l-talopyranosid

Figure S33. Coupled HSQC (600 MHz, 25 ºC, MeOD) spectrum of methyl-6-deoxy-α-l-talopyranosid, distance between crosspeaks result in 1J1'-C/1'-H = 171.8 Hz.

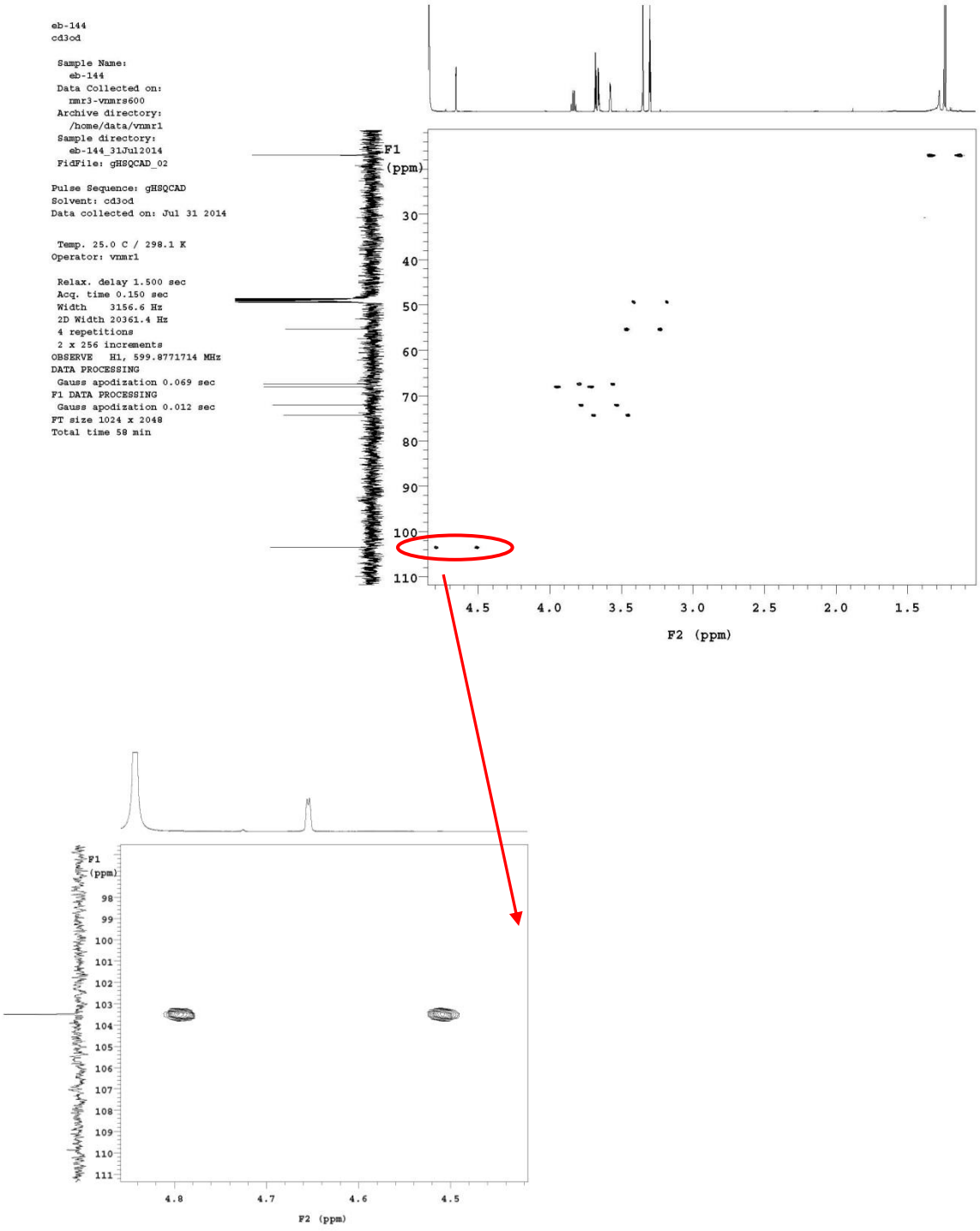


Table S2. MIC50 values (µM) against different pathogens and cell lines

7	Amycomycin C (1)	Amycomycin D (2)	Monoglyc. Alizarin (3)	Alizarin (reference)
Bacillus subtilis	>105	n.d.	n.d.	n.d.
Micrococcus luteus	>105	n.d.	13.2	86.7
Staphylococcus aureus Newman	>105	41.0	>105	No activity
Mycobacterium diernhoferi	>105	n.d.	n.d.	n.d.
Candida albicans	>105	n.d.	>105	>105
P. anomala	>105	32.0	>105	No activity
Mucor hiemalis	94.2	9.21	>105	No activity0
Chromobacterium violaceum	>105	n.d.	>105	>105
E. coli	48.0	14.6	9.11	17.4
ToIC				
E. coli DH5α	>105	n.d.	n.d.	n.d.
E. coli DSM116	n.d.	n.d.	>105	>105
Pseudomonas aeruginosa PA14	>105	n.d.	>105	No activity
Mycobacterium phlei	n.d.	n.d.	>105	No activity
HCT-116	>18.0	~214	n.d.	n.d.

[a] Gram-positiv bacteria: *M. luteus, St. aureus* Newman; [b] fungi; [c] Gram-negative bacteria: *E. coli*; [d] human colon Carcinoma. n.d.: not determined.

Biological Assay

Bacterial Cultures. All microorganisms were handled under standard conditions recommended by the depositor. Overnight cultures of bacteria were prepared in EBS medium (0.5% peptone casein, 0.5% proteose peptone, 0.1% peptone meat, 0.1% yeast extract; pH 7.0) or TSB medium

(1.7% peptone casein, 0.3% peptone soymeal, 0.25% glucose, 0.5% NaCl, 0.25% K2HPO4; pH 7.3). The latter medium was used for E. faecalis and S. pneumonia cultures. Yeast and fungi were grown in Myc medium (1% phytone peptone, 1% glucose, 50 mM HEPES, pH 7.0).

Microbial Susceptibility Assay (MIC). Overnight cultures of microorganisms were diluted to OD600 0.01 (bacteria) or 0.05 (yeast-fungi) in the respective growth medium. Serial dilutions of compounds were prepared as duplicates in sterile 96-well plates. The cell suspension was added and microorganisms were grown overnight on a microplate shaker (750 rpm, 30°C or 37°C). Growth inhibition was assessed by measuring the OD600 on a plate reader. MIC50 values were determined as average relative to respective control samples by sigmoidal curve fitting.

Cell Cultures. Cell lines were obtained from the German Collection of Microorganisms and Cell Cultures (Deutsche Sammlung für Mikroorganismen und Zellkulturen, DSMZ). All cell lines were cultured under conditions recommended by the depositor. Media were purchased from Sigma, fetal bovine serum (FBS Gold) from PAA, and other reagents from GIBCO (Invitrogen).

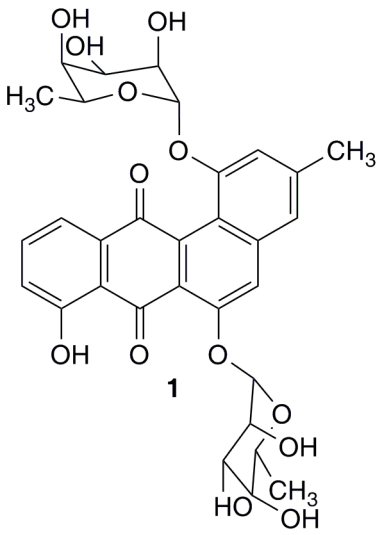
Cytotoxic activity (IC50). Cells were seeded at 6 x 103 cells per well of 96-well plates in 180 µl complete medium and treated with compounds in serial dilution after 2 h of equilibration. Each compound was tested in duplicate as well as the internal methanol control. After 5 d incubation, 20 µl of 5 mg-ml MTT (Thiazolyl blue tetrazolium bromide) [1] in PBS was added per well and it was further incubated for 2 h at 37°C. The medium was then discarded and cells were washed with 100 µl PBS before adding 100 µl 2-propanol-10 N HCl (250:1) in order to dissolve formazan granules. The absorbance at 570 nm was measured using a microplate reader (EL808, Bio-Tek Instruments Inc.), and cell viability was expressed as percentage relative to the respective methanol control. IC50 values were determined by sigmoidal curve fitting and values represent the average ± SD of two independent measurements.

[1] Mosmann T (1983): Rapid Colorimetric Assay for Cellular Growth and Survival: Application to Proliferation and Cytotoxicity Assays. Journal of lmmunological Methods 65:55-63.

Table S3. NMR spectroscopic data (500 MHz, 25 °C, MeOD) for amycomycin C (1)

Pos.	δC	δH, mult, (J in Hz)	COSY[a]	HMBC[a]	ROESY[a]	bashdROESY[b]	bashdNOESY[a,b]
1	155.2	-	-	2-H, 1'-H	-		
2	113.6	7.26 d (1.1)	4-H	4-H, 3a-H3	(4-H), 1'-H, 3a-H3		
3	143.4	-	-	3a-H3	-		
3a	22.2	2.52 s	4-H	2-H, 4-H	2-H, 4-H		
4	121.2	7.32 s (br)	3a-H3, 2-H	2-H, 3a-H3, 5-H	3a-H3, 5-H		
4a	140.4	-	-	4-H	-		
5	118.5	7.75 s (br)	-	4-H	2''-H, 5-H		
6	153.6	-	-	5-H	2''-H		
6a	125.7	-	-	5-H	-		
7	189.2[a]	-	-	-	-		
7a	117.3	-	-	11-H	-		
8	162.8	-	-	9-H, 10-H	-		
9	124.2	7.24 dd (8.4, 1.0)	10-H	11-H	10-H		
10	137.7	7.69 dd (8.4, 7.6)	9-H, 11-H	-	9-H, (11-H)		
11	118.3	7.46 dd (7.6, 1.0)	10-H	9-H	3'-H, 10-H		
11a	137.0	-	-	10-H	-		
12	187.9	-	-	11-H	-		
12a	141.4[a]	-	-	-	-		
12b	117.5	-	-	2-H, 4-H, 5-H	-		
1'	101.3	5.59 d (1.5)	2'-H	1J	2-H, 2'-H		
2'	71.8	3.86 ddd (3.1, 1.5, 1.5)	1'-H, 3'-H, (4'-H)	(1'-H)	(3'-H)	1'-H	1'-H
3'	67.5	3.94 dd (3.2, 3.1)	2'-H, 4'-H	1'-H	2'-H, 4'-H, (5'-H)	11-H, 4'-H	11-H, 4'-H
4'	74.2	3.71 ddd (3.2, 1.7, 1.5)	(2'-H), 3'-H	5'-H, 6'-H	3'-H, 5'-H, 6'-H	5'-H, 3'-H, 6'-H	5'-H, 3'-H, (6'-H)
5'	69.9	4.15 d (6.7)	6'-H	1'-H, 6'-H	6'-H	3'-H, 4'-H, 6'-H	3'-H, 4'-H, 6'-H
6'	17.0	1.28 d (6.7)	5'-H	1J, 5'-H	4'-H, 5'-H		
1''	100.8	5.81 d (1.5)	2''-H	5''-H	4-H, 2''-H		
2''	71.9	4.12 ddd (3.2, 1.7, 1.5)	1''-H, 3''-H, (4''-H)	1''-H	(1'-H), 3'-H, 4'-H	1''-H, 3''-H	1''-H, 3''-H
3''	67.2	4.34 dd (3.4, 3.2)	2''-H, 4''-H	1''-H	2''-H, 4'-H, 5''-H	5''-H, 4''-H	5''-H, 4''-H,
4''	74.3	3.75 ddd (3.4, 1.8, 1.7)	(2''-H), 3''-H	(5''-H), 6''-H	3'-H, 5'-H, 6'-H	3''-H, 5''-H, 6''-H	3''-H, 5''-H, (6''-H)
5''	70.0	4.06 q (6.7)	6''-H	1''-H, 6''-H	3''-H, 4'-H, 6'-H	4''-H, 3''-H, 6''-H	4''-H, 3''-H, 6''-H
6''	16.9	1.21 d (6.7)	5''-H	1J, 5''-H	4''-H, 5''-H		

[a]weak signals are shown in brackets. [b] 600 MHz, 35 °C.

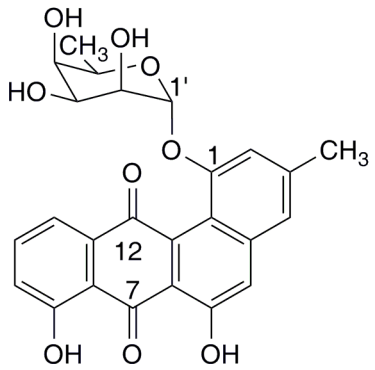


1

Table S4. NMR spectroscopic data (500 MHz, 25 °C, MeOD) for amycomycin D (2)

Pos.	δC	δH, mult, (J in Hz)	COSY[a]	HMBC[a]
1	155.9	-	-	2-H, 1'-H
2	113.2	7.15 s	4-H	1J, 4-H, 3a-H3
3	144.2	-	-	3a-H3
3a	22.9	2.48 s	2-H, 4-H	1J
4	121.2	7.22 s (br)	3a-H3, 2-H	2-H, 3a-H3, 1J
4a	142.9	-	-	4-H
5	119.0	7.42 s	-	4-H
6	158.6	-	-	5-H
6a	139.6	-	-	5-H
7	194.7	-	-	-
7a	117.3	-	-	9-H, 11-H
8	163.8	-	-	9-H, 10-H
9	125.0	7.27 d (8.3)	10-H, 11-H	1J, 11-H
10	139.6	7.75 dd (8.3, 7.6)	9-H, 11-H	-
11	119.8	7.50 d (7.6)	(9-H), 10-H	9-H
11a	138.9	-	-	10-H
12	188.1	-	-	11-H
12a	[b]	-	-	-
12b	118.1	-	-	2-H, 4-H, 5-H
1'	101.8	5.57 s (br)	2'-H	1J
2'	72.4	3.87 ddd (3.1, 1.5, 1.5)	1'-H, 3'-H, 4'-H	1'-H, 4'-H
3'	68.2	3.94 dd (3.3, 3.1)	2'-H, 4'-H	1'-H
4'	74.2	3.72 s (br)	(2'-H), 3'-H	5'-H, 6'-H
5'	69.9	4.16 q (6.7)	6'-H	1'-H, 6'-H
6'	17.0	1.28 d (6.7)	5'-H	1J, 5'-H

[a]weak signals are shown in brackets. [b] not detected from 2D-NMR HMBC spectra.

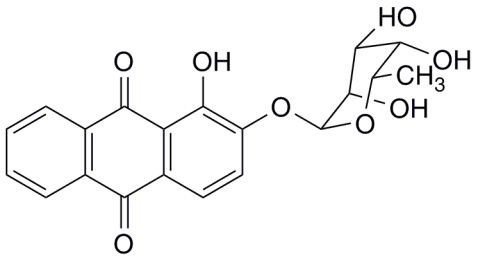


2

Table S5. NMR spectroscopic data (500 MHz, 25 °C, DMSO-d6) of 2-O-α-l-talosyl-1,2-hydroxyanthraquinone (3)

Pos.	δC	δH mult, (J in Hz)	COSY[a]	HMBC[a]	ROESY[a]
1	152.4			3-H	
2	149.8			1’-H, 3-H, 4’-H	
3	121.8	7.63 d (8.5)	4-H	1J	1’-H, (4-H)
4	120.1	7.72 d (8.5)	3-H	3-H	(3-H)
4a	126.2			3-H,	
5	126.8	8.18 m	6-H	1J	
6	135.2	7.93 m	5-H	1J, 8-H	
7	134.3	7.93 m	8-H	5-H, 1J	8-H
8	126.6	8.23 m	7-H	1J	7-H
8a	133.3				
9	188.6			8-H	
9a	116.5			4’-H	
10	180.9			5-H, 4-H	
10a	132.9			6-H, 5-H, 8-H	
1’	99.1	5.73 s (br)		(2’-H, 3’-H), 1J, (2’-OH)	2’-H, (2’-OH), 3-H, 4-H
2’	70.0	3.85 m	2’-OH	3’-OH, 2’-OH, 1’-H, 4’-H,	3’-OH, 2’-OH, 1’-H
3’	65.0	3.85 m	3’-OH (4’-H)	(2’-H), 4’-H, 4’-OH, 3’-OH, 2’-OH, 1’-H	2’-H, 3’-OH
4’	72.0	3.59 d (6.5)	4’-OH, (3’-H, 4’-H)	3’-H, 5’-H, 6’-H3, 4’-OH, 3’-OH	3’-H, 4’-H, 4’-OH, (3’-OH)
5’	68.5	3.85 m	6’-H3, (4’-H)	6’-H3, 4’-OH, 1’-H, 1J	4’-H, 4’-OH
6’	16.6	1.09 d (6.4)	5’-H	1J, 5’-H, (4’-H)	5’-H, (4’-H)
1-OH		12.66 s (br)			
2’-OH		5.39 d (7.0)	2’-H		3’-H
3’-OH		5.12 d (6.0)	3’-H		3’-H
4’-OH		5.02 d (6.5)	4’-H		4’-H

[a]weak signals are shown in brackets



3

Table S6. Band selective NMR data (500 MHz per 25 °C, CDCl3 + F3CCOOD) for 2-O-α-l-talosyl-1,2-hydroxyanthraquinone (3)

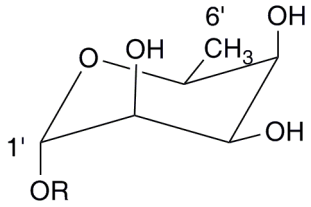
Pos.	δH mult, (J in Hz)	COSY[a]	NOESY
1’	5.83 s	2’-H	2’-H
2’	4.35 s (br)	1’-H	1’-H
3’	4.39 s (br)	4’-H	5’-H, 4’-H
4’	3.99 s (bs)	(4.16), 4.39	5’-H, 3’-H, 6’-H3
5’	4.16 q (6.5)	1.31	4’-H, 3’-H
6’	1.31 d (6.5)	4.16	5’-H, 4’-H

[a]weak signals in brackets.

Table S7. NMR data (500 MHz per 25 °C, MeOD) for methyl-6-deoxy-α-L-talose gained from methanolic hydrolysis of 3.

Pos.	δC	δH mult, (J in Hz)	COSY[a]
1’	103.1	4.66 s (br)	2’-H
1’-OCH3	55.0	3.36 s	
2’	71.7	3.67 dd (3.3, 3.2)	1’-H
3’	67.1	3.69 m	4’-H
4’	74.0	3.59 dd (2.7, 1.5)	3’-H,
5’	67.7	3.84 q (6.6)	6’-H3, (4’-H)
6’	16.6	1.25 d (6.6)	5’-H

[a]weak signals in brakets. R = OCH3



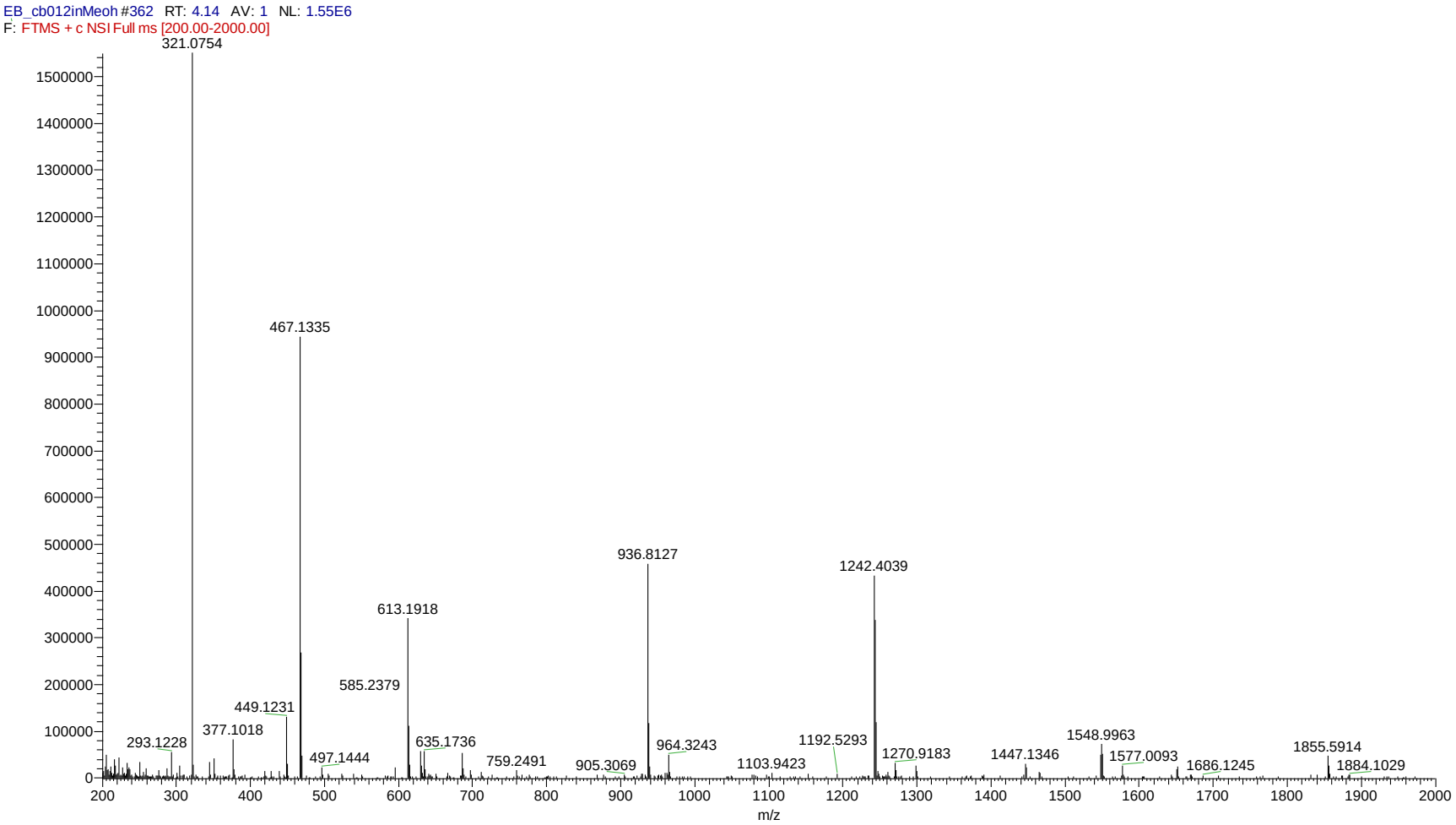


Figure S33. HR-ESI-MS spectrum of amycomycin C (1)

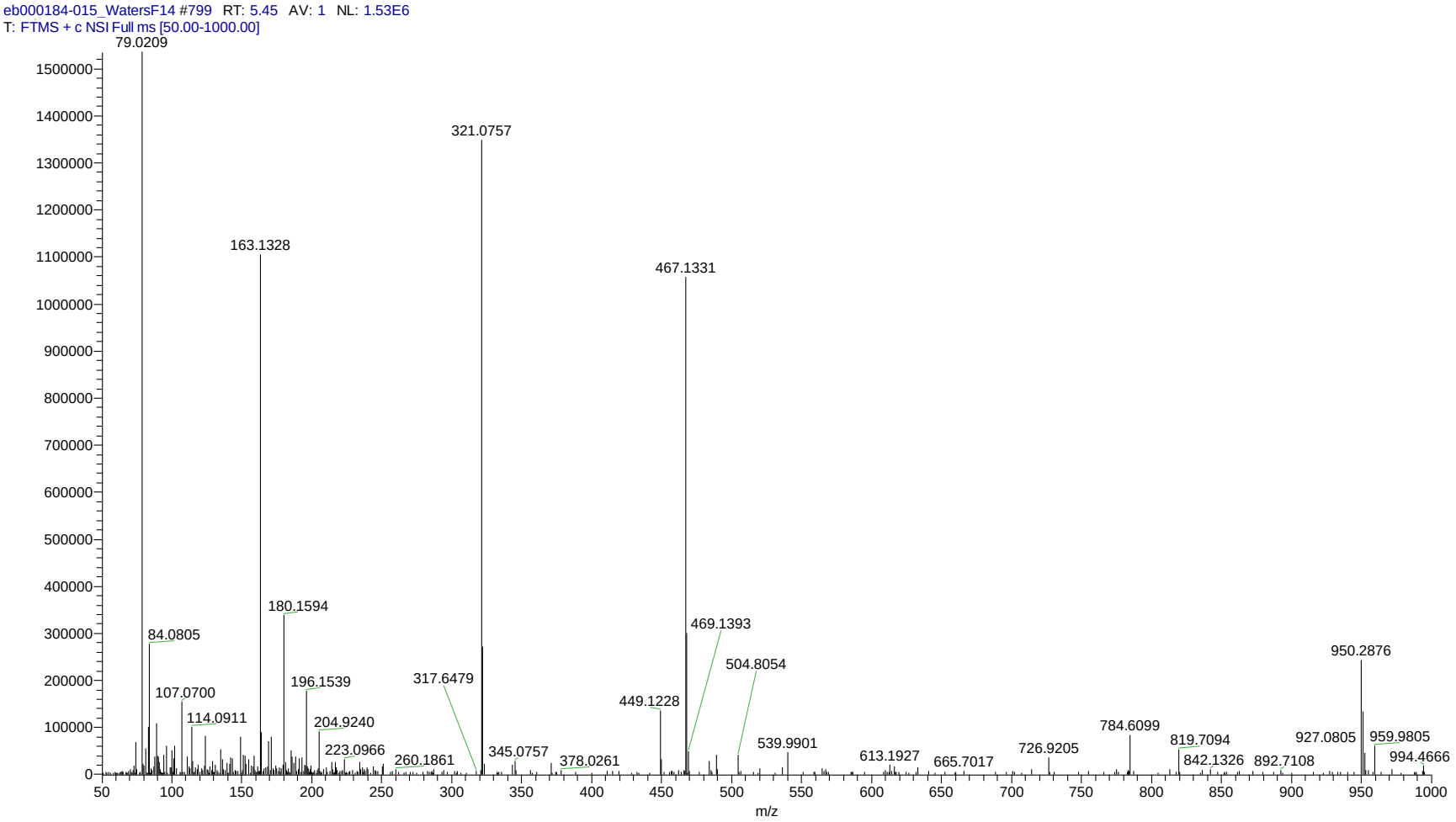


Figure S34. HR-ESI-MS spectrum of amycomycin D (2)

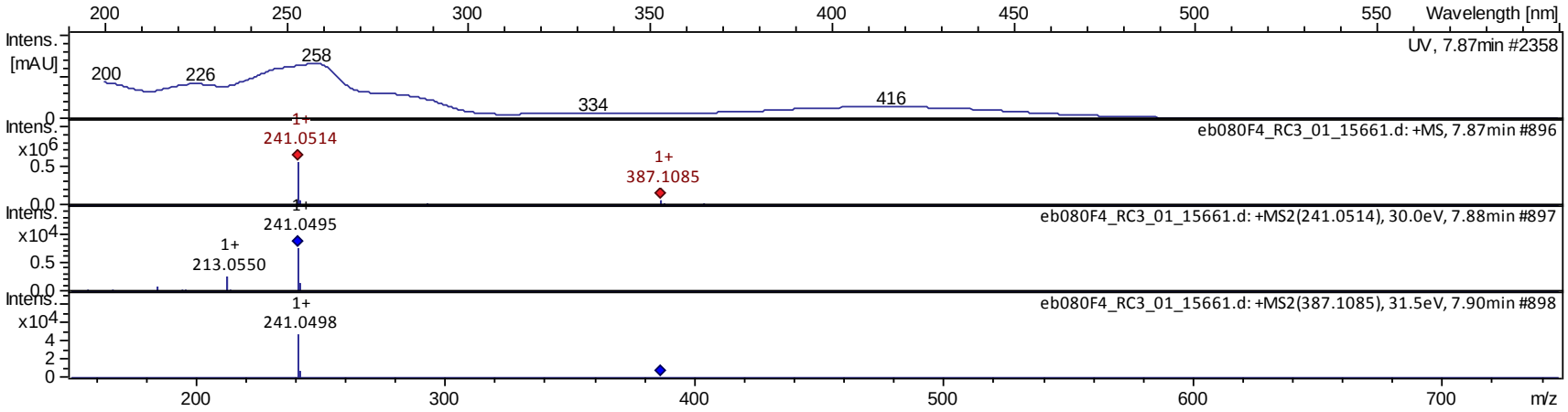


Figure S35. HR-ESI-MS spectrum of O-α-l-talosyl-1,2-dihydroxyantharquinone (3)

SUPPLEMENTAL INFORMATION

Cloning and heterologous expression of the grecocycline biosynthetic gene cluster

Table of Contents

Table 1. Oligonucleotides used in this study

Primer	Sequence (5'→ 3')	Restriction site
greTH-fF-SspI-f	TTT**AATATT**CGTGCACGGGCGGACGGTGTT[a]	SspI
greTH-fR-MCS-r	TTT**AATATT**AAGCCGCCGCCGCCGAGGACG[a]	SspI
greTH-fF-fus-r	CGGCGGGGGAAAGGCTTATG**GATATC**TTCCGAGCC TCCGTTGGGTG[a]	EcoRV
greTH-fR-fus-f	CACCCAACGGAGGCTCGGAA**GATATC**CATAAGCCT TTCCCCCGCCG[a]	EcoRV
greGT1-fR-f (fuz)	CAGCACCGTACGGAGGCCTGA**GATATC**GCGGCACG CCCGGACGGCCGG[a]	EcoRV
greGT1-fF-f	CCC**AATATT**CCGGCCCGTTCCCCCTGACCG[a]	SspI
greGT1-fF-r (fuz)	CCGGCCGTCCGGGCGTGCCGC**GATATC**TCAGGCCTC CGTACGGTGCTG[a]	EcoRV
greGT1-fR-r	CCC**AATATT**GATCCGGAAGGCTCCAGGTCC[a]	SspI
pCLY10-Gre-st	TGGGCTGCAGGTCGACTCTAGAGGATCCGCGGCCGC GCCCCCCGGATGGCGCGGCTGTGTGGTGCTTCAAAG TGGC	-
pCLY10-Gre-end	TCGCGCCCTCCATGAGGCGTACCCGAAGTTCACCGA AGGTCGGACGCCGGGCCCCGGCCCGTCGATGCGCG GCGGG	-
greR1-f	GTGTGGTGCTTCAAAGTGGC	-
greR1-r	GCCCCGGGCGGTGCTCACTG	-
greR3-f	AGGGACCACCCCGTGAAAGC	-
greR3-r	AAGCTTTATGTGCTCATGCC	-
greR2.1-f	GGAGACCCGCCGGGTCCTGAT	-
greR2.1-r	GGAACACGTTGATCATTTCGT	-
greR2.2-f	CGGACTCCCACACGAGAGGAA	-
greR2.2-r	GCCCTTCTCCACACGGATGGT	-
greR1-R2.1-f	GTGGCCAACAACCCCGGCGCC	-
greR1-R2.1-r	TAGAGGGTCAGCAGCGCCACC	-
greR2.1.-2.2-f	CGGCTCGCCGAAGGAGCCCGA	-
greR2.1-2.2-r	GGTGCCTTCGACGGTGGAGGC	-
greR2.2-R3-f	ACTGGGCGGCGGAGCCCTGTG	-
greR2.2-R3-r	GGCGCCGGTGCGTTCCCGCAG	-

[a] – restriction sites introduced artificially are marked in bold

Table S2. Oligonucleotides used for sequencing pGRE

Primer	Sequence (5'→ 3')
gre-seq-51.2	AGACGGTCACGGACAACCTGC
gre-seq-50.2	ATTCTGCTGGTCGTCTTCGGT
gre-seq-49.2	TAACTCATGGGTAAGCCCTTA
gre-seq-46.2	GGTGTACGTCGAAGGTGTTGA
gre-seq-43.2	TATGTGCAGATGCAGCAGGTC
gre-seq-41.2	AACCGTGGATGTACGAGAAGG
gre-seq-39.2	AGCTGTACTTGGCCGACATCC
gre-seq-36.2	AACTCTGTTCACTGCAAGCCG

gre-seq-35.2	TCAACTCCCTCACCCAGCACG
gre-seq-34.2	AACTACACCAAGGTCCACAAA
gre-seq-33.2	AACTCACCAACAGCGACATCA
gre-seq-27.2	GGGTGATCGGCAACGACATCA
gre-seq-24.2	AAGAACGCCATCGACTGGTTC
gre-seq-23.2	AGAACTCCAAGGTCAGCAGCA
gre-seq-18.2	GCTCATCAGCAGGTTCCTCTT
gre-seq-12.2	AACATCACCATCAACAGCGTG
gre-seq-7.1	TCATCGACACGAACCTCAACA
gre-seq-4.1	TGGCCATGAAGCACCACGTGG
gre-seq-1	ATCCGGTTGGTGTGTGATGTT
gre-seq-2	ATAGCTTGCCAGGAGCAGTTC
gre-seq-3	CGACGATGGGGCTCGACCAGG
gre-seq-4	GGTCGGGCACTCTCTGGGCGC
gre-seq-5	CCAGCGGAGTGGTCGTCAGCG
gre-seq-6	CCGCGGCCACGGCGGCTTCAA
gre-seq-7	CTGGTCAACAATGCGGGCCGC
gre-seq-8	GTGAGAAGGACGAACGCATCC
gre-seq-9	GCCGGTGTCGCCGAGGCCAGG
gre-seq-10	CCGGCGCTTCGAACGGCTGGA
gre-seq-11	GGCCTGCTGCTGAACATCTCC
gre-seq-12	TCTTCATCATCCAGAGGGCGC
gre-seq-13	GCTTCGGCGGACTGCTGCTGT
gre-seq-14	GATCCGCTCGGCGCTGGGGGC
gre-seq-15	GGAACGACGGCGGGCCGCTTC
gre-seq-16	CGCCACCGAGATCCAGCACCA
gre-seq-17	GACCGCTCGGACGACCCCGCC
gre-seq-18	TCCTGATCCGCTCCCTGGCCAT
gre-seq-19	GCACGGGCACCCGGCGGCACC
gre-seq-20	GGCGGGATGCGAGCGGGCTAT
gre-seq-21	CATTCCTGGAGGAGTGCCTCA
gre-seq-22	CAGGCGCGGCTGCGAGGTCCT
gre-seq-23	CCCCTGCGGGACGTGCTGTCC
gre-seq-24	ACGCCGATCCGGCGACGGAGG
gre-seq-25	GAGATGGAGGACGCCAACGCC
gre-seq-26	CGCCCCGGCGCGACCCGGACG
gre-seq-27	ACGCCGAACTCGCCGCGAGGG
gre-seq-28	CCCGGCAGATCAACTACTCCG
gre-seq-29	TCACCTACATCCCCCAGGACA
gre-seq-30	CGTCCTGGACGGTGCCATGAT
gre-seq-31	CCGAGAAGCTGATCCCGCTCT
gre-seq-32	GGCTCCTCGGCCAACCTGCTC
gre-seq-33	GGCTGCGGGAGGGACTGGACG
gre-seq-34	GCCAAGGAGTTCGGCGGCGTC
gre-seq-35	TCCTCGACACCGTGGAGCTGG
gre-seq-36	CTCATGGAGAACGTGATGTTC
gre-seq-37	CGGCGCTCGCCGCTGTCCTGC
gre-seq-38	GCTGCACTACGCCAAGTACCT

gre-seq-39	GCCACCGGCCCCACCGCGCCC
gre-seq-40	CGGCATGGTCTCCGGCCTCGA
gre-seq-41	GCTGGCGCCCGGACGTCGTGG
gre-seq-42	CCAGGAACTCCTGTCGAATCC
gre-seq-43	GCCCCGGTGCTCGCGAAGTTC
gre-seq-44	CGGAGGCCTGATGCGCGTCCT
gre-seq-45	ACCCTCGCCCTCAACGGCGTA
gre-seq-46	CCCGGAGACGGCCCGCCGTCC
gre-seq-47	ACGCCTTCACGGCGGACTGCC
gre-seq-48	CCGTGGCCGACGGGCCTCCTG
gre-seq-49	AGTCTGTCGGGCTGTGCAGCC
gre-seq-50	AAGGAGTTCGCCAAGGAGGGC
gre-seq-51	GCCGGCTGCCGGCGGAGGCGG