Metabolic and Genomic Profiling of Actinobacteria Strains: from New Natural Products to Biosynthetic Pathways

Dissertation zu Erlangung des Grades

des Doktors der Naturwissenschaften

der Naturwissenschaftlich-Technischen Fakultät

der Universität des Saarlandes

von

Constanze Paulus

Saarbrücken

2018

Tag des Kolloquiums: 01. Februar 2019

Dekan: Prof. Dr. Guido Kickelbick

Berichterstatter: Prof. Dr. Andriy Luzhetskyy

Prof. Dr. Christian Ducho

Vorsitz: Prof. Dr. Anna Hirsch

Akad. Mitarbeiter: Dr. Angelika Ullrich

Diese Arbeit entstand unter der Anleitung von Prof. Dr. Andriy Luzhetskyy in der Fachrichtung 8.2, Pharmazeutische Biotechnologie der Naturwissenschaftlich-Technischen Fakultät der Universität des Saarlandes von Januar 2015 bis Oktober 2018.

Acknowledgements

My deepest gratitude goes to my supervisor Prof. Dr. Andriy Luzhetskyy for giving me the opportunity to do my PhD in his group. I would like to thank him for all the interesting and challenging projects I was working on and for the great support and encouragement throughout that time.

Furthermore, I would like to express my thanks to my mentor Dr. Yuriy Rebets. I was lucky to have such a good teacher who introduced me into the world of microbiology and who answered every question I had during my PhD.

I also owe my special gratitude to my second supervisor Prof. Dr. Christian Ducho. I thank him for his time and the useful advices to my work in the committee meetings.

Moreover, I would like to thank Birgit Rosenkränzer who facilitated the entry into the new work environment and laboratory work. I owe my thanks also to Dr. Thomas Paululat, Dr. Suvd Nadmid and Dr. Josef Zapp for supporting me in NMR questions and for all useful discussions regarding structure elucidation problems. I would like to thank of course all my colleagues for the pleasant working atmosphere in the laboratory and all discussions that helped to find solutions during difficult steps within a project. And moreover for having a great time and so many nice and fun activities during the last three years.

Mein tiefster Dank gilt meinen Eltern Richard und Ursula Paulus. Dafür, dass sie immer für mich da sind, mich zu jedem Zeitpunkt unterstützen und mir vor allem in schwierigen Zeiten immer Mut zusprechen. Der Dank geht auch an meine drei lieben Geschwister, Julia, Philipp und Catharina Paulus, dafür, dass sie mir immer zur Seit stehen würden und für all die schönen Momente, die wir zusammen verbracht haben trotz der Entfernung.

Publications

Sun, Y.Q.; Busche, T.; Rückert, C.; Paulus, C.; Rebets, Y.; Novakova, R.; Kalinowski, J.; Luzhetskyy, A.; Kormanec, J.; Sekurova, O.N.; Zotchev, S.B. Development of a biosensor concept to detect the production of cluster-specific secondary metabolites, *ACS Synthetic Biology*, **2017**, *6* (6), 1026-1033.

Paulus, C.; Rebets, Y.; Tokovenko, B.; Nadmid, S.; Terekhova, L.P.; Myronovski, M.; Zotchev, S.B.; Rückert, C.; Braig, S.; Zahler, S.; Kalinowski, J.; Luzhetskyy, A. New natural products identified by combined genomics-metabolomics profiling of marine Streptomyces sp. MP 131-18, *Scientific Reports*, **2017**, *7*, 42382.

Paulus, C.; Rebets, Y.; Zapp, J.; Rückert, C.; Kalinowski, J.; Luzhetskyy, A. New Alpiniamides from *Streptomyces* sp. IB2014/011-12 assembled by an unusual hybrid nonribosomal peptide synthetase trans-AT polyketide synthase enzyme, *Frontiers in Microbiology*, **2018**, *9*, 1959.

Conference contributions

Paulus, C.; Rebets, Y.; Tokovenko, B.; Nadmid, S.; Terekhova, L.P.; Myronovski, M.; Zotchev, S.B.; Rückert, C.; Braig, S.; Zahler, S.; Kalinowski, J.; Luzhetskyy, A. New natural products identified by combined genomics-metabolomics profiling of marine Streptomyces sp. MP 131-18, **7th HIPS Symposium, Saarbrücken, Germany, 2016** (poster)

Paulus, C.; Rebets, Y.; Tokovenko, B.; Nadmid, S.; Terekhova, L.P.; Myronovski, M.; Zotchev, S.B.; Rückert, C.; Braig, S.; Zahler, S.; Kalinowski, J.; Luzhetskyy, A. New natural products identified by combined genomics-metabolomics profiling of marine Streptomyces sp. MP 131-18, **VAAM international Workshop, Freiburg, 2016** (poster)

Zusammenfassung

Marine Aktinomyceten stellen eine beliebte Quelle neuer Naturstoffe dar, welche unter Umständen Anwendung als therapeutische Arzneimittel finden. Daher steht die Untersuchung neuer Aktinomyceten Stämme im Fokus der Naturstoffforschung. Die dargelegte Arbeit handelt von der Isolierung und strukturellen Charakterisierung neuer Naturstoffe, welche von Streptomyceten produziert wurden. Desweiteren, wird die Konstruktion eines neuen Biosensors dargestellt, welcher die Produktion von Sekundärmetaboliten detektieren soll.

Es wurden zwei neue spiroindimicine E und F aus der *Streptomyces* sp. MP131-18 isoliert und die chemische Struktur mittels NMR bestätigt. Das Genom von MP131-18 wurde sequenziert und der Gencluster, welcher verantwortlich ist für die Synthese der Bisindole-Substanzen identifiziert und den Lynamicinen/Spiroindmicinen zugeordnet. Weiterhin, wurden vier neue Alpiniamide B-D aus dem Stamm *Streptomyces* sp. IB 2014/11-12 isoliert. Das sequenzierte Genom ermöglichte die Identifizierung des Genclusters, welcher für die Alpiniamide Produktion verantwortlich ist. Die vorgeschlagene Biosyntheseroute wurde bestätigt durch Fütterungsexperimente und Gendeletierungen in einem Hostorganismus.

Abschließend wurde ein Beitrag zur Konstruktion eines Repressor-basierender Biosensors geleistet, welcher die Aktivierung von zuvor inaktiven Gencluster in Streptomyceten detektieren soll. Dieses Konzept wurde erfolgreich an dem aktivierten Coelimycin Gencluster angewendet.

Abstract

Marine actinomycetes are known to be a promising source for new natural products with putative application as therapeutic agents. Thus, the exploitation of novel discovered actinomycetes strains remains in the focus of NP research. The presented thesis deals on the one hand with the isolation and structural characterization of new NPs produced by streptomycetes species. On the other hand, the construction of a novel biosensor concept for the detection of secondary metabolite production is discussed.

New spiroindimicins E and F were isolated from the *Streptomyces* sp. MP131-18 and structures were confirmed by NMR. Additionally, two new lagunapyrones D and E were identified by characteristic MS/MS fragmentation. The genome of MP131-18 has been sequenced and the gene cluster, responsible for the synthesis of bisindole compounds has been identified and connected to lynamicins/spiroindimicin production. Furthermore, four new alpiniamides B-D have been isolated from the *Streptomyces* sp. IB 2014/11-12. The sequenced genome enabled the identification of the gene cluster responsible for alpiniamide production. The predicted biosynthetic pathway was confirmed by feeding experiments and gene deletions in a heterologous host.

Finally, a contribution to the construction of a repressor-based biosensor, which detects the products of awakened silent gene clusters in streptomycetes, has been made. This concept was successfully applied to the activated coelimycin gene cluster.

Table of Contents

1. Introduction

1.1 Impact of Natural Products in Drug Development

Pharmacologically active substances from natural sources play an important role in human life since the earliest times of traditional medicine and until the modern medical practice. The number of new chemical entities (NCEs) with a natural background between 1981 and 2014, recorded by the U.S. Food and Drug Administration (FDA), reflects the significance of natural products (NPs). As it has been summarized by Newman and Cragg, 39% of a total number of 1562 new approved drugs during these 34 years, are unmodified NPs, semisynthetic modified NPs or synthesized substances with a NP pharmacophore.[1] The indications of these drugs in human medicine are versatile. They can be used as antibacterial, anticancer, anti-inflammatory, antiviral or immunosuppressive agents, to name a few. However, many of them seem to be applicable in cancer therapy or in the treatment of infectious diseases which belong to the ten leading causes of death worldwide, according to the World Health Organization.[2] NPs and/or their derivatives account for 52% and 59% of antibacterial and anticancer NCEs respectively. These numbers illustrate that NPs remain an important source to treat life-threatening diseases.[1]

However, the usage of NPs for the antibiotic industry is declining. The difficulties and huge costs, which accompany the process from the initial NP discovery to the final drug caused a decreased NP based antibiotic research by many pharmaceutical companies.[3] Screening of libraries of bacterial extracts using the "Waksman platform" resulted in re-discovery with increased regularity of already known compounds. Moreover, the chemical complexity of the existing NPs aggravates their synthesis or derivatization, which makes them less favorable for the development of suitable drugs. Altogether, the examination of microorganisms, screening of NP extracts libraries for activity and the eventual upscaling for the isolation and purification of compounds are time-consuming and expensive. Instead, the high-throughput screening (HTS) using "screen-friendly", synthetic compounds libraries have been introduced.[4] The benefits of such libraries are reliable supply of compounds and a better, chemical accessibility for the subsequent drug optimization. Although, this approach has been successful for the discovery of many therapeutics, it has only moderate success (hit rate HTS < 0.001%, "Waksman platform" 0.3%) in the finding of medicines against infectious diseases.[3]

Moreover, the fast development of resistance among pathogenic strains against most of the newly introduced antibiotics exacerbates the situation.[5] The decrease in discovery of new antibiotics together with the growing resistance evokes an immense disequilibrium between the number of available antibiotics in contrast to the number of resistant strains, which makes once

treatable diseases again lethal. The most dangerous, multiresistant nosocomial pathogens with currently missing treatment options are summarized as the ESKAPE pathogens (*Enterococcus faecium*, *Staphylococcus aureus*, *Klebsiella pneumoniae*, *Acinetobacter baumanii*, *Pseudomonas aeruginosa* and *Enterobacter species*).[6] The foundation of the "10 x '20 Initiative" in 2010 by the Infectious Disease Society of America (IDSA) was supposed to fill that gap. The initiative was meant to motivate scientists and the pharmaceutical industry in order to develop and approve 10 new, efficient and save drugs against bacterial infections, especially those caused by the ESKAPE-pathogens, until 2020.[7] The success of this program is reflected by the approval of several new antibacterial drugs between 2010-2015, e.g. ceftaroline fosamil acetate (2010), dalbavancin (2014), tedizolid (2014), oritavancin (2014) and ceftolozane/tazobactam (2014), ceftazidime/avibactam (2015), meropenem/vaborbactam (2017) as combined agents against pathogens with extended spectrum ß-lactamases (ESBLs, Fig. 1).[8,9] Similarly, the European Society of Clinical Microbiology and Infectious Diseases (ESCMID) initiates and supports research and educational projects aiming for a better application of antibiotics, an improved treatment and prevention of infectious diseases.[10]

Despite the introduction of new effective antibiotics, which can combat resistant strains, most of them are only improved versions of already existing compounds. The basic effective pharmacophore in most cases is the same as in many other medicines on the market with the same mode of action within the bacterial cell. Hence, to overcome the resistance issue, new compound classes with new core structures and new mode of action are needed. The introduction of drugs with entirely new molecular targets can prevent the fast adaption of bacterial strains to the antibiotic and thus enable a longtime and effective usage.

The discovery of suitable antibiotics against the resistant Gram-negative bacteria (*Acinetobacter baumanii*, *Pseudomonas aeruginosa*, *Enterobacteriaceae*) represents even bigger challenge to overcome.[11] The cell envelope of Gram-negative bacteria is more difficult to penetrate, which is why antibiotics for their treatment are much harder to discover. This is also the main reason for the failure of compounds resulting from HTS of chemical libraries. Even though they are toxicologically and pharmacologically more suitable than NPs they often lack the effective characteristics of typical NP antibiotics such as the ability to penetrate the cell envelope.[12] The likelihood to discover active agents within NPs is also reflected by the fact that 73 % of all antibacterial agents introduced between 1981 and 2014 derive from NPs.[1] As a consequence, it is necessary to maintain NP discovery programs and expand the research by novel approaches with 21st century techniques. Only then, the huge demand for new, effective antibacterial drugs can be fulfilled.

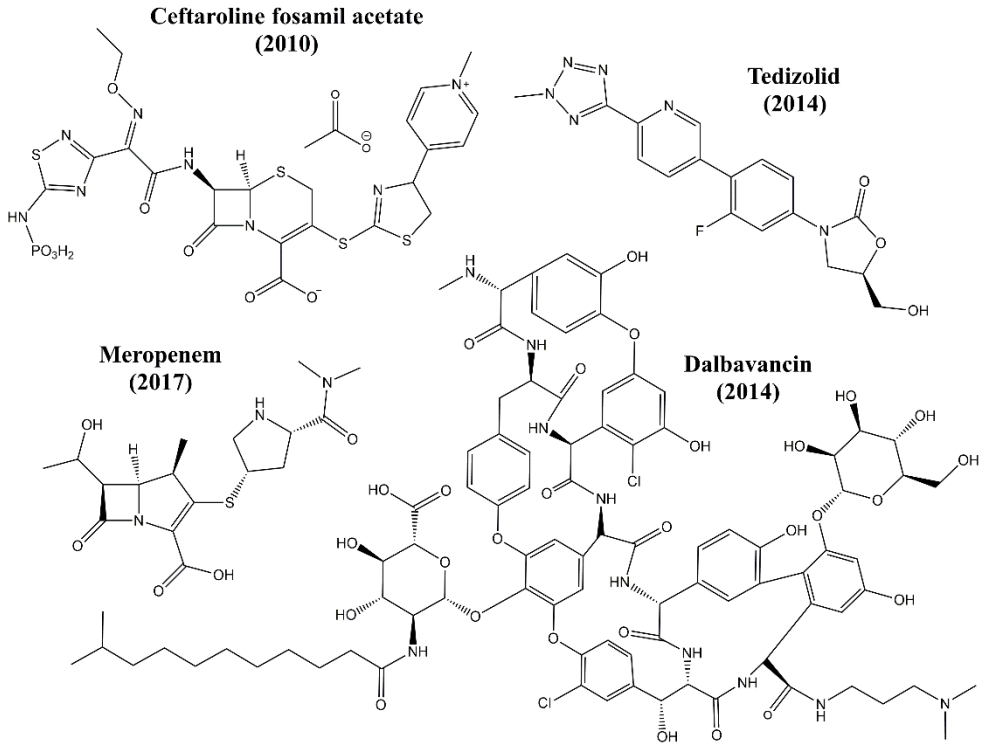


Figure 1. Newly approved antibiotics in clinical use with NP pharmacophore.

1.2 Natural Products from Actinobacteria

Actinomycetes (mainly of *Streptomyces* and related genera) are Gram-positive, mostly soil-dwelling, filamentously growing bacteria and represent a major source of valuable NPs.[13] They occupy different ecological niches but are predominately distributed in soil (population density 10^6-10^9 cells per 1 g), whereas 95% of them belong to the genus *Streptomyces*.[12] Unsurprisingly, representatives of this genus were one of the first producers of medically useful NPs and provide to date more than half of all clinical used antibiotics. The first actinobacterial antibiotic produced in industrial scale was the aminoglycoside streptomycin (Fig. 2), isolated from *Streptomyces griseus* in 1944, used for the tuberculosis treatment.[14] Many other medically valuable antibiotics derived from streptomycetes were subsequently discovered, the latest one being the cyclic lipopeptide daptomycin (Fig. 2) from *S. roseosporus* in 2003 for the treatment of life-threatening infections caused by Gram-positive pathogens such as the methicillin resistant *Staphylococcus aureus* (MRSA).[15] Actinobacterial NPs originate from a divers secondary (or auxiliary) metabolism. In most cases these compounds are not required for bacterial survival and their role is still under debates.[16] Beside the compounds used in human medicine, actinobacteria produce many other NPs some of which have herbicidal and insecticidal properties.[17,18]

Thus, it is not surprising that actinomycetes from terrestrial sources have been the focus in drug discovery research for a long time. However, the likelihood to discover promising secondary metabolites with new core structures from terrestrial actinomycetes has meanwhile considerably decreased. The over-exploitation of soil actinomycetes leads more and more often to a rediscovery of same strains and hence to the isolation of already known metabolites. One of the reasons lies in the specific growth characteristics of different species. The faster growing, and phenotypically more visible streptomycetes are prone to be chosen more easily at the expense of slower growing, less distributed minor groups, which are thus often overlooked. Only recently, scientists put more efforts in the discovery of minor groups of actinobacteria resulting in diversification of NPs.[19] Another reason for the abundant rediscovery of known compounds, is the widespread horizontal gene transfer (HGT) among soil actinomycetes.[20] The conjugative transfer of biosynthetic gene clusters between bacterial cells leads to generation of distant species with a similar NP profile, even though the strains derive from different habitats. As a consequence, researchers started to move towards highly underexplored niches. Especially the marine and freshwater environments with its much higher biological diversity compared to the terrestrial surroundings are believed to provide an auspicious alternative for the discovery of novel drug leads.[21-23] Marine actinobacteria are often exposed to extreme conditions such as low temperatures, high pressures, specific pH-values, depleted nutrient supply and alternating oxygen concentration as opposed to their terrestrial counterparts. In order to survive these living circumstances they have to adapt physiologically and metabolically. This fact encouraged scientists to explore the marine environment assuming that the metabolic adaption of actinomycetes will also lead to production of bioactive secondary metabolites with unique structures. Marine sources including fjords sediments, deep-sea floors, tidal-flats and coral reefs as well as fresh-water lakes are currently under investigation for new actinobacterial strains.[24] Beside the terrestrial counterparts of actinomycetes often found in marine samples, also indigenous, taxonomically distinct actinomycetes exist in the oceans. Besides the well-known genus *Streptomyces*, other members of actinobacteria can be found, e.g. *Marinispora, Rhodococcus, Salinispora* or *Dietzia*.[25] *Nocardiopsis* sp. isolated in a sediment sample from the Trondheim Fjord in Norway produces the thiopeptide TP1161 (Fig. 2) which represents a good example for the rewarding exploitation of such sources.[26] This thiopeptide possesses activity against a panel of Gram-positive strains. A further example is the discovery of *Salinispora arenicola* strain CNT-088 which was isolated from a marine sediment sample collected around the islands of Fiji.[27] The strain was found to produce the cyclohexadepsipeptides arenamides A-C which possess anti-inflammatory properties (Fig. 2). Moreover, the investigation of marine

invertebrates symbiotic living bacteria such as corals, tunicates, sponges or cone snails resulted in identification of promising new compounds.[28] Sponges, sessile multicellular organism, which are found in several marine and fresh-water ecosystems, are known to harbor microorganisms capable of production of bioactive secondary metabolites. An example for this is the strain *Streptomyces axinellae* which lives in symbiosis with the Mediterranean sponge *Axinella polypoides*, collected from Banyuls-sur-Mer, France.[29] This strain was found to produce new tetromycin derivatives with antiparasitic and cytotoxic activities (Fig. 2).[30] Another example of strains living in symbiosis is the *Streptomyces sp.* CP32, which inhabits the cone snail *Conus pulicarius*. New neuroactive thiazoline compounds pulicatins A-E were isolated from this strain (Fig. 2).[31]

However, the majority of microorganisms can't be isolated by conventional methods, since 99% of all bacteria, terrestrial as well as aquatic, are not cultivable under standard laboratory conditions.[32] Also, the applied cultivation conditions do not reflect the stressful natural situations on which bacteria react individually by tuning their metabolic processes, including secondary metabolism. These differences make the majority of secondary metabolism pathways silent under laboratory conditions. Thus, a large part of secondary metabolites is yet waiting to be discovered and alternative approaches are required to unlock the full biosynthetic potential of yet "unculturable" strains. A prominent example of a NP discovered by a new technique is teixobactin. This cyclic depsipeptide was isolated via an iChip in 2015, from a yet-uncultured bacterial strain and shows antibacterial activity against some clinical relevant Gram-positive pathogens including multidrug-resistant strains.[33] Through its specific mode of action during cell-wall synthesis, which is different from the common mechanism, it is believed that treated bacteria might not become resistant as fast as with other antibiotics.[34] Another strategy is the metagenomic approach, a cultivation-independent technique which requires the isolation of DNA directly from the environment (eDNA) and the subsequent cloning into cultivable host strains.[35] An example are the recently discovered calcium-dependent macrocyclic lipopeptides malacidins.[36] The isolation was achieved through a metagenomic-approach using eDNA from soil samples. Noteworthy, is the activity against Gram-positive pathogens including MRSA caused through binding of malacidin to Lipid II during cell wall synthesis.

Despite the growing potential of new techniques, in most cases the method of choice for discovery of new actinobacterial NPs is still based on the classical way. Isolation of metabolites from the culture broth of cultivated strains seems to be preferred in many laboratories since the potential of even well characterized strains appeared to be far not exhausted. Often simple

alteration of cultivation conditions can induce production of new metabolites. In conclusion, the exploration of new ecological niches comprises great possibilities for the discovery of new strains and thus new NPs.

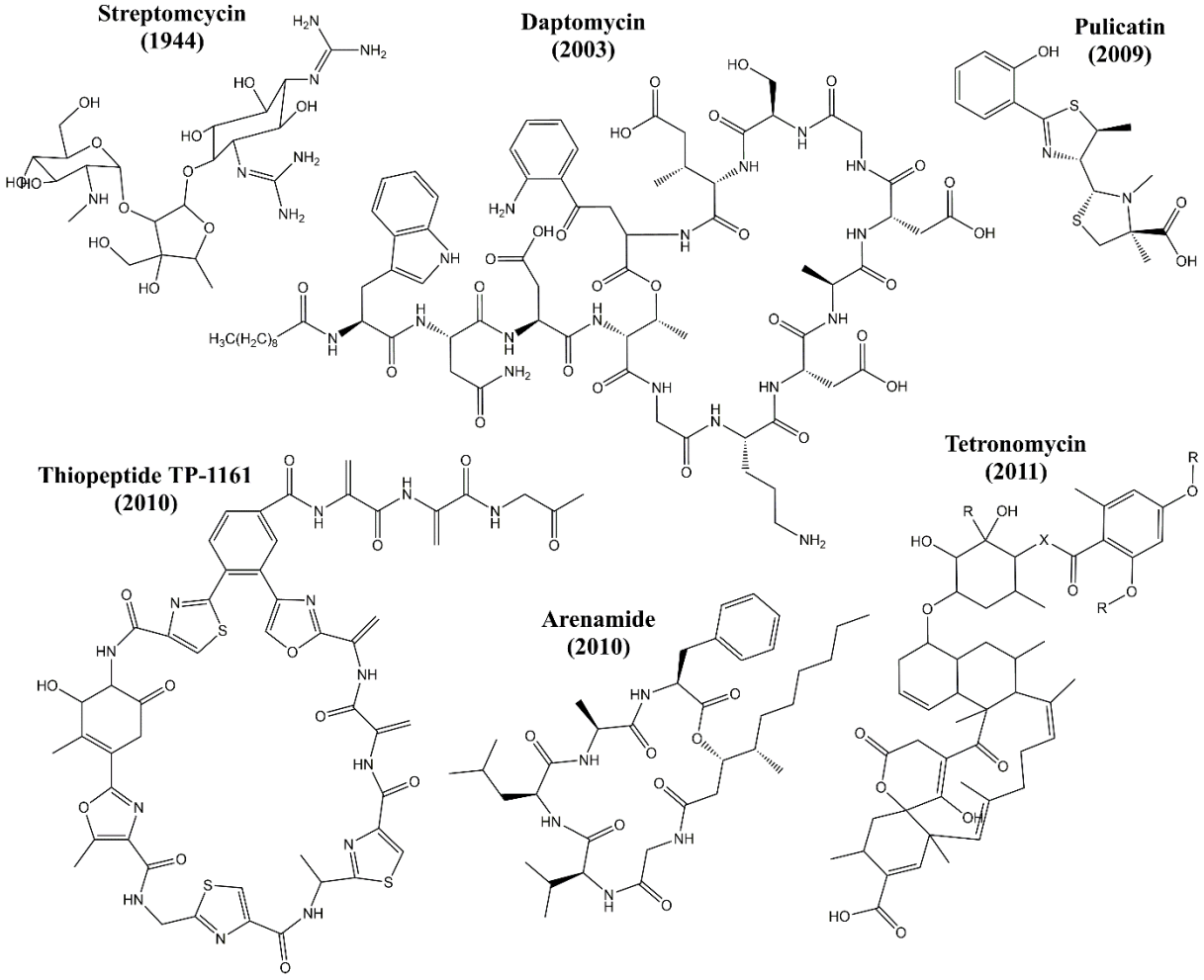


Figure 2. NPs from soil streptomycetes in clinical use (streptomycin, daptomycin) compared to NPs isolated from actinomycetes derived from new ecological niches (pulicatin, thiopeptide TP-1161, arenamide, tetronomycin).

1.3 Dereplication

The abundant rediscovery of the same compounds remains a major concern in NP research. In order to avoid time-consuming work, the reliable discrimination of known and new compounds is necessary. This process, known as dereplication, refers to an effective, preceding discard of futile or ineffectual compounds (Fig. 3).[37] For this purpose, high resolution analytical techniques are required in order to identify the targeted compound in a complex mixture such as the crude extracts of cultivated bacterial strains. A common technique is the analytical separation of the crude material by high performance liquid chromatography (HPLC) and the

subsequent detection of eluting compounds by the use of diode array detectors (DAD), mass spectrometry (MS) or nuclear magnetic resonance spectroscopy (NMR).[38] These techniques yield useful properties e.g. retention time, UV/vis absorption, exact mass, semi-structures, elemental composition, fragmentation patterns, which can be than used for the comparison of data of known NPs. In order to enable efficient dereplication, many NP databases appeared over the last years, providing various types of searching possibilities.[39] The most useful, commercially and publically available databases in terms of the number of included, known NPs are for instance CAS/SciFinder, PubChem and the Chapman & Hall/CRC Dictionary of Natural Products (DNP).[40-42] CAS/Scifinder includes over 283,000 NPs and provides the possibility to search with complete or partial drawn structures or molecular formula. For each result, predicted properties and analytical spectra and also the respective reference to the literature are given. Pubchem, a public domain for searching and comparing substances, provides around 438,000 entries for NPs. Similar, as in other databases, a structure based or molecular formula based search can be performed resulting in hits with listed chemical properties, literature references and derivatives of the given structure, as well as biological properties of compounds including specific activity records. The DNP database contains around 260,000 NPs and provides the broadest list of possibilities to search for instance by an accurate mass, UV maxima, a biological source or a molecular formula. Particularly useful is the information about the source of the described metabolite that greatly narrows down the search results.

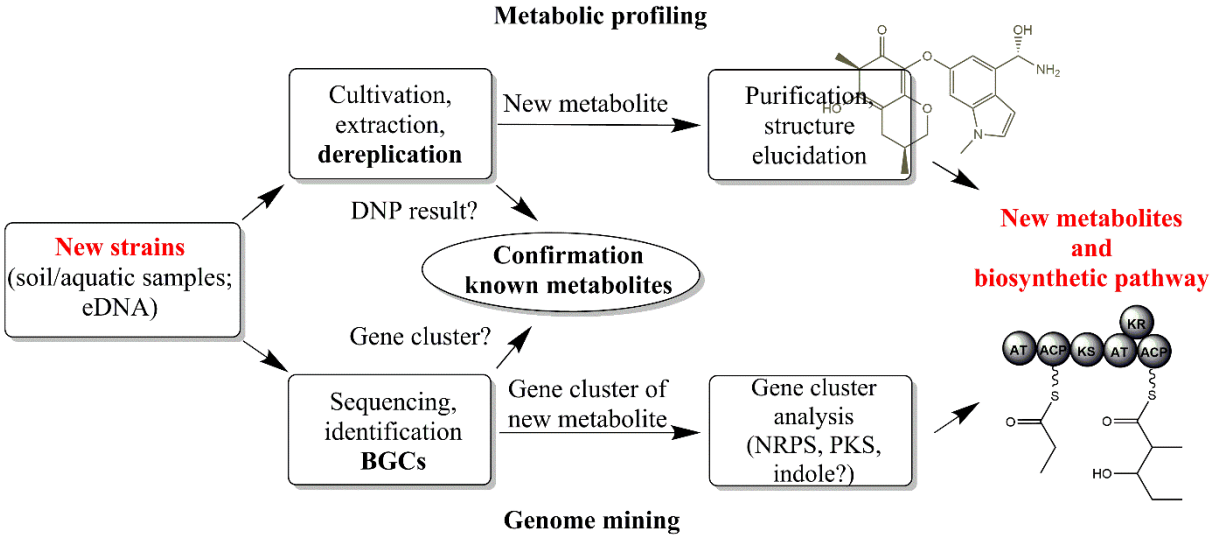


Figure 3. General workflow: metabolic profiling and genomic analysis of new bacterial strains.

The process of dereplication helps to evaluate the novelty of metabolites in a crude extract and saves time if the metabolite can be identified as known compound. In order to increase the

confidence of prediction it is important to have an additional confirmation whether the NP proposed by dereplication is known or new. Nowadays, a genome mining approach creates a strong link between the compound and corresponding biosynthetic genes and can be used to validate the prediction made by analytical methods.

1.4 Genomic Profiling of Actinomycetes

1.4.1 Bacterial Genome Sequencing and Identification of Biosynthetic gene cluster (BGC)

In the early years of antibiotic research, the isolation of the compounds from the culture broth was a precondition to determine the metabolic profile of a specific strain. Nowadays, the sequencing of bacterial genome provides the opportunity to analyze the biosynthetic potential of an organism *in silico*. Through the development of low-cost "Next-Generation Sequencing" techniques (NGS), genome sequencing became mundane, which lead to a sudden rise in the availability of genomic information.[43] In order to deal with the increasing amount of newly obtained data, effective *in silico* methods for the identification of secondary metabolism gene clusters and connection of these clusters to the vast amount of already known metabolites (genome mining) were developed.[44] This way, core biosynthetic genes of already known compounds can be identified, supporting or denying the dereplication results and moreover, pointing on unknown gene clusters within a genome and thus leading to the discovery of novel NPs (Fig.3). Several software for an automated cluster mining were developed to date, e.g. DECEIPHER® (2001), NP.searcher, CLUSEAN (2009), antiSMASH (2011) and PRISM (2015).[45-49] For instance, antiSMASH, first launched in 2011, is a comprehensive platform for the analysis of genomic data. It combines several computing tools for a better comparison of genomic sequences, for the identification of BGCs and prediction of putative secondary metabolites. Approximately 44 different types of gene clusters (including type I-III polyketide synthase (PKS), non-ribosomal-peptidesynthethase (NRPS), lantipeptide, indole, bacteriocin, lassopeptide, phenazine, terpene, siderophore, melanin, ectoine etc.) can be identified by comparison of the genome sequence of a bacterium of interest against a library of common enzyme/protein domains from different biosynthetic pathways. The completed analysis is shown as an overview of all identified BGCs within the genome and additionally the degree of homology to known BGCs is given (Fig. 4 a)). Furthermore, for each cluster detailed information is provided, such as domain annotation (Fig. 4 b)), cluster boarder prediction, regulatory TTA codon annotation and ten closest gene clusters from other bacteria, which are homologous to the query sequence. Moreover, the antiSMASH algorithm provides the detailed

prevision of domain's substrate specificity for NPRS and PKS assembly lines, as well as based on this prediction, the possible core structure of metabolite (Fig. 4 b)).

a)

Overview 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33

Identified secondary metabolite clusters

Cluster	Type	From	To	Most similar known cluster
The following clusters are from record contig000001:				
Cluster 1	T2pks-Terpene	1	45587	Rabelomycin_biosynthetic_gene_cluster (35% of genes show similarity)
Cluster 2	Nrps	183717	237638	Coelibactin_biosynthetic_gene_cluster (90% of genes show similarity)
Cluster 3	T3pks-T1pks	396158	445235	Fosfomycin_biosynthetic_gene_cluster (10% of genes show similarity)
Cluster 4	T1pks-Butyrolactone	520956	590040	Versipelostatin_biosynthetic_gene_cluster (31% of genes show similarity)
Cluster 5	Thiopeptide	620445	652572	Herboxidiene_biosynthetic_gene_cluster (2% of genes show similarity)
Cluster 6	T3pks	750513	791616	Herboxidiene_biosynthetic_gene_cluster (2% of genes show similarity)
Cluster 7	Terpene	811012	832855	Hopene_biosynthetic_gene_cluster (30% of genes show similarity)
Cluster 8	T1pks	891782	945787	Pladienolide_biosynthetic_gene_cluster (37% of genes show similarity)
Cluster 9	T1pks	946080	992517	Tetronasin_biosynthetic_gene_cluster (13% of genes show similarity)
Cluster 10	Nucleoside	1149710	1170075	Toyocamycin_biosynthetic_gene_cluster (40% of genes show similarity)
Cluster 11	Siderophore	1675972	1690283	-

b)

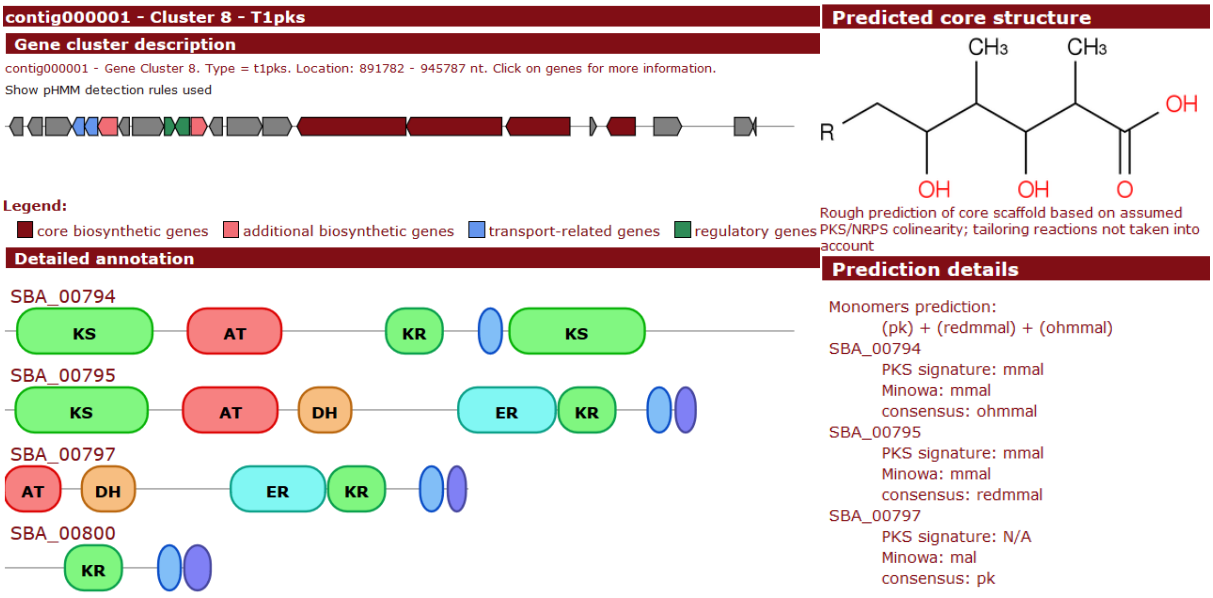


Figure 4. Genome mining software antiSMASH: a) output genome analysis; list of identified secondary BGCs and most similar known cluster b) cluster annotation with identified domains (left) and predicted core structure based on predicted monomer choice (right).[50]

Finally, the external reference to Minimum Information about a Biosynthetic Gene cluster (MIBiG) database for each of the analyzed genes in the predicted gene cluster is accessible, showing the closest experimentally characterized orthologues of the gene, thus giving clues to the function of the respective protein in the biosynthetic pathway.[51] Similar software called PRISM was developed by Magarvey et al. in 2015, also as a tool for a facilitated BGC analysis. Compared to antiSMASH, PRISM provides for example structure prediction for type II

polyketide synthase products and is also suited for a better identification and annotation of cluster-associated genes involved in resistance mechanism. Moreover, structure prediction seems to be more detailed through the addition of foresight of mechanism of cyclization and sugar biosynthesis and attachment. However, PRISM has limitations regarding the type of data submitted, since it cannot process fragmented genomes (multiple contigs) and can work only with the fully assembled genomic data. Also, the wide usage of antiSMASH prompted the development of satellite software (MIBiG, ARTS, etc.) that provide additional possibilities in secondary metabolites gene cluster analysis, like prioritization of compounds of interest based on prediction of their mode of action. The latter is achieved by the identification of self-resistance genes within the gene cluster, pointing on possible mode of action of the produced compound.[52]

Tools, like antiSMASH and PRISM became indispensable for the NP researchers since they help to evaluate the importance of a specific cluster. Identification of biosynthetic genes for a particular compound of interest can thus often verify dereplication results by connecting the identified metabolite to the respective BGC. In general, genome mining has one major limitation. The prediction of chemical structures from the cluster organization is not reliable so far so that the revers approach connecting genes to their products remains difficult. The development of better tools for gene/protein based compound structure prediction is required and is a major focus of antiSMASH like software development.

1.4.2 Enzyme Machinery for Secondary Metabolite Production

Bacteria are able to synthesize a vast number of structurally diverse metabolites. In most cases, genes for the production of secondary metabolites are clustered within the bacterial genome. Despite the clear structural variety of secondary metabolites produced by bacteria, still two types of chemistry are dominating in their biosynthesis: Claisen condensation, similar to assembly of malonyl-CoA units during fatty acid biosynthesis and transpeptidation, a condensation of amino acids forming an amide bond as in the case of ribosomal protein synthesis. In both cases the resulting secondary metabolites, polyketides and non-ribosomal peptides, originate from a simple monomer supplied from the primary metabolism. Due to the obvious simplicity of the primary chemistry behind the polyketides and non-ribosomal peptide synthesis, they represent the major classes of NPs produced by bacteria including actinobacteria. The structural diversity is achieved either by utilization of unusual building blocks or post-assembly modification of the initial product.

Polyketide NPs are assembled by repetitive Claisen condensation events conducted by enzymes called polyketide synthase (PKS). Polyketide synthases are similar to fatty acid synthase (FAS) by both structural and functional features.[53] Both enzymes initiate the biosynthesis with an acetyl-coenzyme A (CoA) as starter unit, which is then elongated with an extender unit (e.g. malonyl-CoA) through a decarboxylative condensation. For this basic reaction, three core activities are required: the β-ketoacylsynthase (KS) which is responsible for covalent junction of the starter/growing polymer and extender unit; the acyltransferase (AT) selecting and loading the extender unit; and the phosphopantetheinylated acyl-carrier protein (ACP), holding the growing chain during the synthesis. Alternatively, it is known that FAS uses also amino acid derived starter units and PKSs can employ beside acetyl-CoA as well malonyl-CoA, propionyl-CoA and the more uncommon aminoacid derived isobutyryl-CoA, isovaleryl-CoA and 2-methylbutyryl-CoA.[54] The completed chain is released by a thioesterase enzyme, through cyclization or hydrolysis depending on the type of the final molecule. Despite the similarity in polymer formation mechanism, the major difference between FAS and PKS lies in the different degree of post-assembly modification. The β-keto group in fatty acids biosynthesis is fully reduced after each elongation cycle, through the combined actions of ketoreductase (KR), dehydratase (DH) and enoylreductase (ER) enzymes. The KR reduces first the ß-keto function to a hydroxyl group, which is further eliminated by DH, leading to a double bond at the respective position and finally the ER reduces this double bond resulting in a fully reduced alkyl chain intermediate. In a similar fashion the growing polyketide chain is modified by the same set of ketoreduction enzymes, but unlike the fatty acid biosynthesis, the ketoreduction steps are rather optional, leading to a high variety of functional groups in the resulting polymer. Additionally, the AT domain in PKSs can select from a broader pool of precursors beside the malonyl-CoA, e.g. methylmalonyl-CoA, ethylmalonyl-CoA, hydroxymalonyl-CoA, etc.[55] Further modifications such as oxidation, glycosylation, hydroxylation, amination, acylation also contribute to the diversity of polyketide NPs.

Polyketide synthases have been classified into subgroups (type I, II and III) depending on their architecture and functional properties. Modular type I PKS consists of linearly arranged obligatory KS, ACP and AT domains, as well as optional KR, DH, ER, MT and TE domains grouped into modules, each performing one elongation and ketoreduction step (Fig.5). The primary structure of type I PKS generated polyketides is dictated by the composition and order of domains within the linearly arranged modules, leading to co-linearity between the architecture of the megaenzyme complex and its product.[56] This feature of type I PKS makes it possible to predict the primary structure of the produced polyketide. Type I PKS are producing

three major classes of compounds (macrolides, polyenes and polyethers) that differ by type of cyclization and degree of ketoreduction.

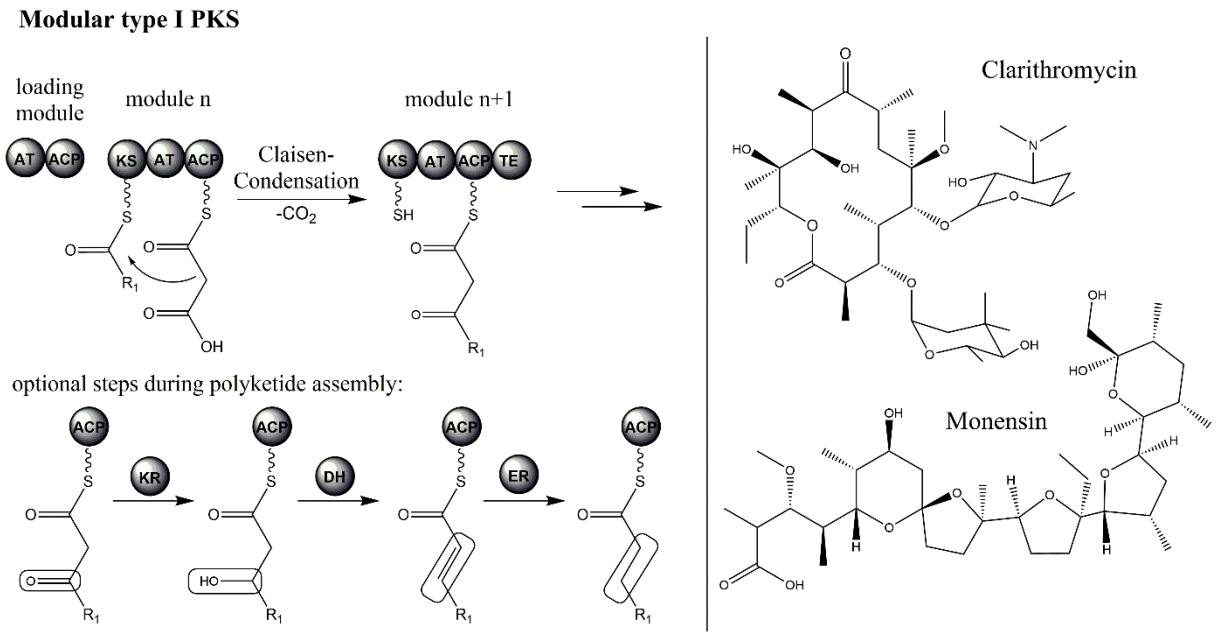


Figure 5. Mechanism of modular type I PKS and typical structures.

Macrolides consist of a large 14 to 16-membered lactone ring with one or more sugars attached. Representatives of this group are for instance erythromycin, bafilomycin or anthracimycin.[57-59] Macrolide antibiotics act as bacteriostatic agent by inhibiting the bacterial protein synthesis.[60] Polyene antibiotics, a subgroup of macrolides, contain a lactone ring with 25-38 carbon atoms including a series of conjugated double bonds. Nystatin, amphotericin and candicidin, the most prominent examples, are used in the treatment of mycotic infections caused by different fungi. The fungicidal properties are attributable to the interaction of the antibiotic with the ergosterol rich region of the fungal plasma membrane.[61-63] Polyethers have multiple tetrahydrofuran and tetrahydropyran rings which derive from internal cyclization events. The inophore properties of these NPs are the key contributors to their antibiotic activity.[64] The major application of compounds such as monensin, narasin and salinomycin are in veterinary medicine due to their activity against bacteria causing coccidiosis.[65] Several variations of type I PKS exist. The classical cis-AT modular type I PKS, described above, are commonly accepted as a structural and activity model for this type of enzymes. They are mostly found in bacteria, but can also be found in fungi and algae.[66] Vice versa, iterative type I PKSs (iPKSs) were believed to be predominant in fungi, but were recently also discovered in bacteria.[67]

The iPKS consists of only one module, which is used in a repetitive manner to conduct a series of acyl-CoA condensations resulting in different degree of reduction in the polyketide chain

(Fig. 6). Typical structures synthesized by this enzyme are 6-methylsalicylic acid (6-MSA) and lovastatine produced by fungi.[68,69] Bacterial iPKS were found to be involved in the biosynthesis of polyketomycin, chlorothricin and avilamycin whereas the iPKS provides 6-methylsalicylic acid for the small aromatic structural component attached to the sugar chain.[70-72]

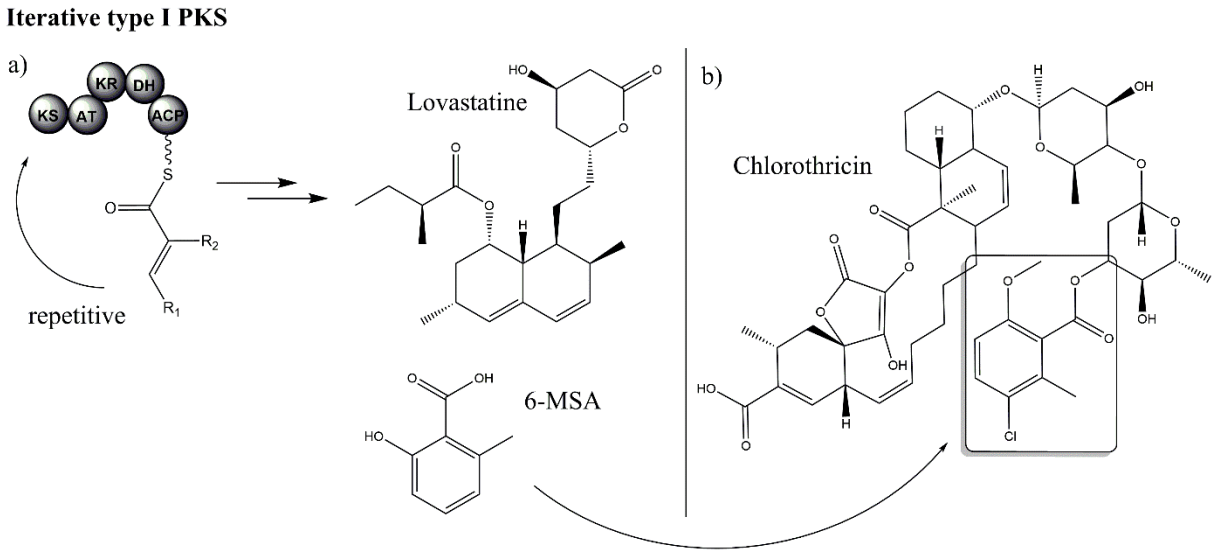


Figure 6. a) Mechanism of iterative type I PKS and typical representatives lovastatine and 6-MSA. b) the structure of chlorothricin, aromatic part in the box is assembled by iPKS.

Another large group of bacterial modular type I polyketide synthases is comprised of enzymes lacking acyltransferase domain.[73] Usually, each module employs its own AT-domain (*cis*-AT PKS) with the defined substrate specificity, introducing the alkyl sidechain diversity into the growing polyketide. However, in so called *trans*-AT PKS, the AT domain is provided by a free-standing enzyme, that typically have a narrow preference to malonyl-CoA as a substrate. *Trans*-AT PKS systems were thought to be rare in streptomycetes, only recently it has been discovered that almost half (38%) of all modular bacterial polyketide synthases belong to the *trans*-AT system.[74] Although that *trans*-AT PKS are more likely to be found in rather unusual bacteria species like those living in symbiosis with higher organisms.[73] Another large difference of *trans*-AT PKS is the lack of correlation between the module architecture and the chemical structure of the resulting polyketide. This absence of co-linearity makes it difficult to predict structures deriving from this type of a PKS enzyme. However, over the years, comparison of many *trans*-AT PKS sequences and the corresponding polyketide structures lead to some regularity in how the specific domains are used and which starter and extender units are preferred. Nonetheless, they are often unpredictable due to the many conjunction possibilities of the domains in contrast to the *cis*-AT-PKSs. *Trans*-AT PKS can harbor domains performing new unusual chemistry, domains acting in different orders, non-functional domains or even

domains which act across modules. The former, is another distinguishable feature of *trans*-AT PKS. These enzymes are described to conduct several biochemical reactions during primary polyketide assembly, that are not found in classical cis-AT type I PKS: in line C- and O-methylation, β-branching, double bond migration, 5- and 6-membered ring formation etc.[74] This unusual type of PKS megaenzymes revealed many extraordinary NPs, e.g. bacillaene, mupirocin, macrolactin or bryostatin.[75-78] Interestingly, recently the iterative *trans*-AT-PKS system was discovered, conducting the biosynthesis of chejuenolides, 17-membered carbocyclic tetraenes isolated from the marine bacterium *Hahella chejuensis* MB-108.[79]

The type I PKS ketoreduction cycle, when resulting in a hydroxy group, and *in li*ne methylations of the polyketide typically lead to formation of a stereocenter. Thus many NPs assembled by type I PKS represent stereoisomers which often define their activity. The stereochemical outcome of the processing enzymes KR, DH as well as C-methyltransferase (C-MT) can be predicted according to their active site residues. KR domains are divided into groups depending on stereospecificity of performed reaction.[80] A-type KRs generate L-β-hydroxy group, whereas B-type KRs generate D-β-hydroxy groups. Starting from a D-α-substituted substrate, L-β-hydroxy groups derive from A1-type KRs and D-β-hydroxy groups from B1-type KRs. The optional epimerization of the α-substituent results in L-α-substituted substrate which is than reduced by A2- or B2-type KRs respectively. C2-type KRs perform solely the epimerization step. The hydroxy group generated by reduction of the keto group can be further dehydrated by the action of the DH domain which generates either a *cis* or a *trans* double bond. The binding of the substrate in the active site cavity has to be specific so that the β-hydroxy group can be protonated by aspartatic acid and L-α-proton can be abstracted by histidine. Due to this specific binding, DHs generate a *cis* double bond from L-β-hydroxy group substrates (A-type KR product) and *trans* double bonds from D-β-hydroxy group substrates (B-type KR product).[80,81] Little is known about the stereospecifity of MT domains which transfer a methyl group to the α-carbon mediated by *S*-adenosylmethionine (SAM). However, the investigation of four different C-MTs from *trans*-AT-PKSs conducting biosynthesis of bongkrekic acid, bacillaene, difficidine and mupirocin, resulted always in D-configuration of the installed methyl group.[82] The fact that KR differ by their sequence around the active site residues can be thus used to predict the configuration by specialized software. Although, at the current state such prediction is still far from being completely reliable, these bioinformatics tools are of great help during analysis of type I PKS megaenzymes and their products.

Iterative type II polyketide synthases (often called aromatic PKS), solely found in prokaryotic organism (mostly actinobacteria with some rare exceptions), are composed of two distinct

ketosynthase enzymes KS_α and KS_β (the latter is lacking active site Cys residue and often called as chain length factor (CLF)), an ACP and a malonyl-CoA:ACP transacylase (MAT), loading the malonyl extender unit on ACP (Fig.7 a)).[83] The first three proteins are forming a complex, called "minimal polyketide synthase", performing a repetitive cycle of malonyl-CoA condensation leading to the formation of a nascent unreduced β-polyketide chain. The growing polymer due to its high reactivity is hidden inside the "minimal PKS" enzymatic complex that also determines the number of condensation events, resulting in the chains of specific length. Unlike the type I PKS, which incorporates the ketoreduction cycle into each step of condensation, type II PKS system employs no or only one ß-ketoreduction event, that is also proposed to dictate the first cyclization position of the polyketide chain. At the same time, type II PKS typically involves multiple post-PKS modification enzymes, such as cyclases (Cyc) for ring system formation and an aromatase/dehydratase for dehydration of cyclic alcohols resulting in an aromatic structure. The resulting rings can be further modified by tailoring enzymes such as oxygenases, hydroxylases, reductases, aminotransferases and O- or C-glycosyltransferase. The attachment of deoxysugars is quite often determining the biological activities of aromatic polyketides. Typically, genes for biosynthesis of these sugars are clustered together with the PKS genes. The products of type II PKS are aromatic polycyclic compounds that are divided into several groups depending on the type of cyclization: anthraquinones (actionrhodin, galvaquinone),[84,85] tetracyclines (chlortetracycline, chelocardin (anti-bacterial)),[86,87] anthracyclines (doxorubicin, daunorubicin, nogalamycin (anticancer)),[88-90] angucyclines (aquayamycin, landomycin, urdamycin, simocyclinone (anticancer/anti-bacterial)).[91-94]

Type III PKS have been predominantly found in plants, but throughout the last decades, numerous type III PKS have been also found in bacterial and fungal genomes. This type of NP biosynthesis enzymes are often referred as chalcone/stilbene synthase superfamily and can be divided into 5 groups according to post-PKS processing enzymes (RppA, PhID, DpgA, ArsB/SarsS, Gcs/ArsC/PKS18).[95] As type II PKS, they operate iteratively but are composed of one ketosynthase enzymes forming a homodimer, elongating the growing chain with a specific number of malonyl units (Fig. 7 b)). The distinctive feature of type III PKS is the lack of ACP and direct utilization of coenzyme A to deliver the extender units and hold the growing polyketide chain during assembly process. Products of these enzymes contain small, hydroxylated aromatic structures with singular aromatic rings or lactones with attached alkyl side chains such as in pyrones or phlorogluycinols (Fig. 7 b)).

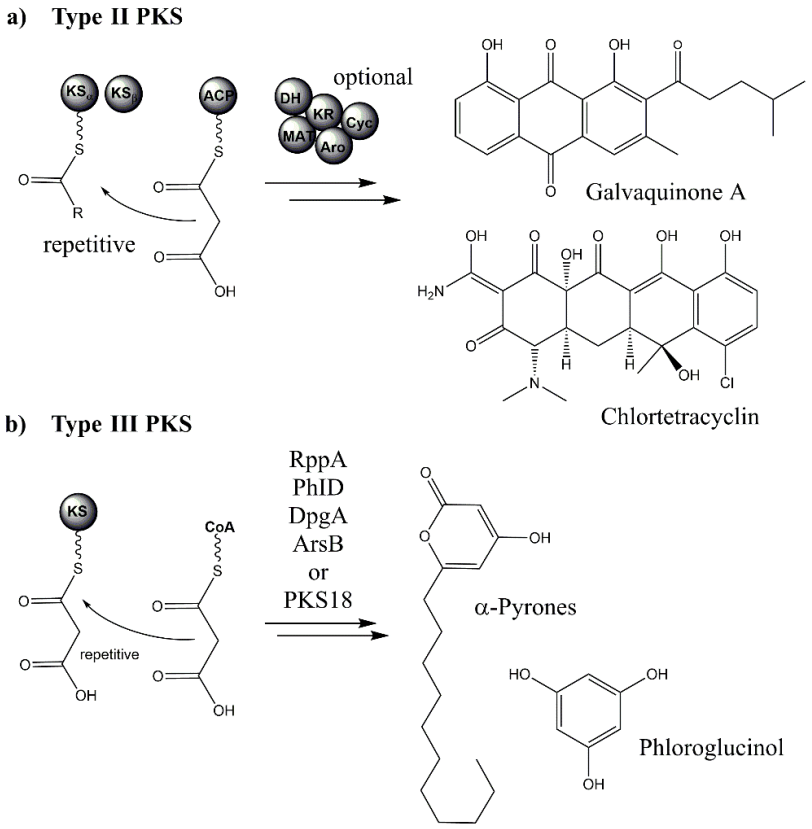


Figure 7. a) Structures and mechanism of a bacterial type II PKS b) Mechanism of the ACP independent type III PKS and possible products.

Nonribosomal peptides represent the second largest group of NPs. These compounds are highly modified oligo- or polypeptides that are produced in ribosome independent way by the enzyme called nonribosomal peptidesynthetase (NRPS).[96] NRPS, similar to type I PKS, have a modular organization, leading to collinearity of the megaenzyme organization and the primary structure of the resulted product. The NRPS module typically consists of three major domains: adenylation (A), condensation (C) and peptide-carrier-protein (PCP) (Fig.8). The A-domain is responsible for activation and loading of a specific amino acid onto the PCP-domain, whereas the C-domain catalyzes the amide-bond formation between two amino acids. The A-domain determines the amino acid to be included into the growing polypeptide chain at each step of an assembly. The specificity of the A-domain to particular amino acid is determined by its primary sequence and is often referred to as nonribosomal code.[97] Some modules might contain additional modification domains, like N- or C- methyltransferase, epimerase, reductase, etc. The final peptide is then released by a TE-domain through aminolysis, hydrolysis or intramolecular cyclization resulting in formation of peptidyl amides, lactones or lactams. The great structural diversity of non-ribosomal peptides is achieved by the incorporation of proteinogenic as well as non-proteinogenic aminoacids and additionally through tailoring

enzymes performing methylation, epimerization, hydroxylation, heterocyclization or glycosylation.[98,99] Similar to PKS, some NRPS enzymes can act iteratively, resulting in structures with repeated amino acids motifs. Prominent examples of nonribosomally synthesized NPs are vancomycin, bacitracin or teixobactin.[33,100,101]

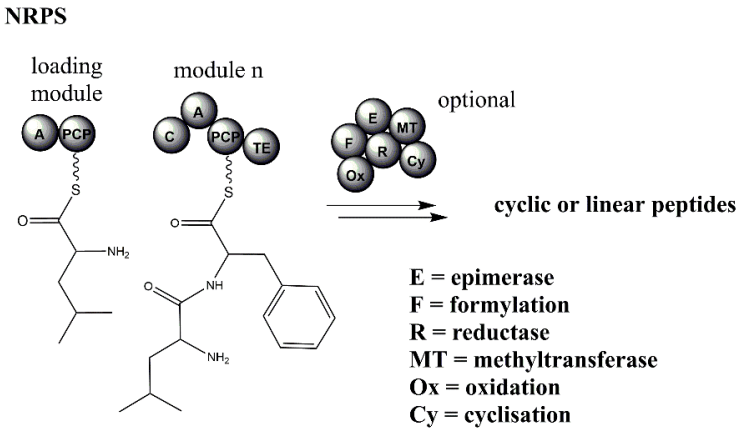


Figure 8. General biosynthetic scheme for nonribosomal peptide synthesis.

The further diversity of bacterial NPs is achieved by combining type I PKS and NRPS enzyme machineries to produce a hybrid polyketide-peptide compounds. An interesting feature of hybrid PKS/NRPS systems is that at the interface between PKS and NRPS enzymes the growing chain has to be translocated between them. This means that the KS domain has to accept rather a peptidyl chain than a polyketide chain and vice versa the C domain has to accept a polyketide chain rather than a peptidyl substrate.[102] The hybrid enzymes employ all the features of both components, leading to formation of highly complex metabolites. Examples for such complex hybrid compounds are bleomycin and epothilone.[103,104]

Non-ribosomal peptides and polyketides are by far the largest groups of bacteria derived NPs. Biosynthetic gene clusters encoding terpene, lantipeptide, lassopeptide, RiPP (ribosomally synthesized and post-translationally modified peptide) or indole NPs occur less frequent within the bacterial genomes, however they contribute to great variety of bacterial NPs. Only indoles biosynthesis will be discussed in more details. The bisindole NPs consist of two tryptophan units fused together, often forming an additional ring-system such as in spiroindimicin or borregomycin and in the indolocarbazoles erdasporine and rebeccamycin. The biosynthesis of bisindoles requires four core enzymes involved in modification and processing of amino acid L-tryptophan precursor.[105] At first, L-tryptophan is converted into indolpyruvate imine by an amino acid oxidase. Its active species, indolpyruvate enamine undergoes than a radical coupling catalyzed by chromopyrrolic acid synthase to give the intermediate chromopyrrolic acid (CPA). The combined action of cytochrome P450 and a flavin-dependent enzyme generates the final

indolocarbazole core structure (Fig.9). However, the last step is sometimes skipped, like in case of the violacein or lynamicin. The indolocarbazoles are often modified by halogenation and genes encoding corresponding enzymes are coupled with the core genes. The halogenation of the tryptophan occurs usually before synthesis of the core structure.[106] The indolocarbazoles have different biological activities, however the most prominent of them are rebeccamycin and staurosproine which are used as drug leads in anticancer therapy due to their cytotoxic properties.[107]

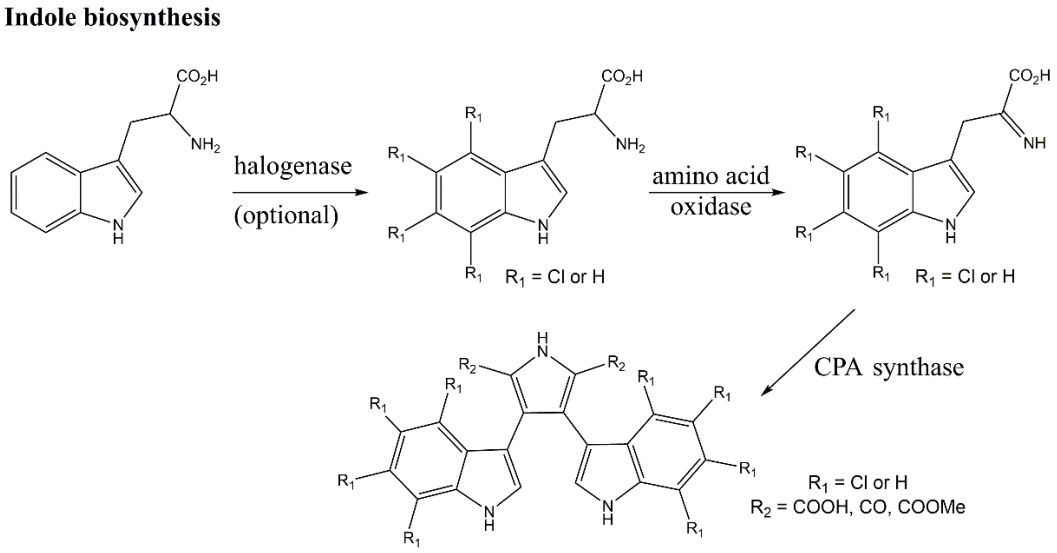


Figure 9. General biosynthetic scheme for indole biosynthesis.

The understanding of the logic behind the biosynthetic machinery of NPs enables scientists to predict the secondary metabolism potential of an examined organism when genomic data is available. In combination with metabolic profiling (test cultivation, analytics and dereplication) the detailed examination of the predicted biosynthetic gene clusters (genome sequencing, identification of BGC and biosynthesis pathway prediction) are a basis for identification of known compounds and provide hints about putatively new metabolites produced by the strain of interest. Such a pre-analysis prevents the needless isolation of irrelevant compounds and helps to prioritize the NPs produced by the isolated actinobacteria using the analytical methods before proceeding to their purification and characterization.

1.5 Isolation and Characterization of Secondary Metabolites

If the metabolic and genomic assessments result in an annotation of an identified NP as potentially new, the next steps involve isolation and structure elucidation. Furthermore, the isolation of a compound provides a possibility for characterization of its biological activity. For an efficient isolation and purification procedure the combination of different extraction-, pre-purification- and final purification methods are applied.

1.5.1 Extraction and Purification

Crude extracts usually consist of a large number of chemically diverse substances which differ in size, molecular weight and lipophilicity/hydrophilicity. According to these physicochemical properties, the optimal purification procedure has to be chosen individually. The initial step after harvesting a large cultivation batch is the appropriate separation of the compounds from the cellular matrix by an extraction. A liquid-liquid extraction with low or high polarity solvents yields either more lipophilic or more hydrophilic compounds, which decreases simultaneously the amount of unwanted materials. Subsequent pre-purification steps such as solid phase extraction (SPE), size exclusions chromatography (SEC) or flash chromatography (medium pressure chromatography) can give manageable fractions containing compounds of interest. As a final purification step, often HPLC with normal phase or reversed phase columns and equipped with an automated fraction collection and detectors such as DAD (diode array detection), FD (fluorescence detection), RID (refractive index detection), NIR (near-infrared detection), CD (circular dichroism), MS (mass spectrometry) or MS-MS (tandem mass spectrometry) is used.[108-110] Overall, there is no general procedure for efficient extraction, separation and purification. The procedure includes always multiple steps and has to be adjusted to the certain NP. Nowadays, the isolation is often activity-guided or target-guided (focus on specific compounds selected during dereplication) in order to facilitate and accelerate the process.[111]

1.5.2 Structure Elucidation and Assignment of Stereochemistry

1.5.2.1 Nuclear Magnetic Resonance Spectroscopy (NMR)

In the early years of NP research, the elucidation of the chemical structure was a complicated and time-consuming task due to the lack of an appropriate instrumentation. The elucidation of a molecule's structure required the partial degradation of the compound into smaller fragments and often only a total synthesis could give the final proof. The general improvement of

spectroscopic techniques during the last 50 years facilitated the NP structure elucidation. Particularly NMR spectroscopy became a powerful and reliable tool for this purpose. In the beginning, the low sensitivity of the first NMR spectrometers (60 MHz magnets) and the absence of multiple pulse sequences for the acquisition of two-dimensional NMR spectra limited utilization of this technology. However, NMR spectroscopy improved to such an extent that it became the method of choice in NP research. The high sensitivity nowadays, provided by spectrometers with up to 1 GHz magnets, allows the measurement of less than 1 mg samples in a reasonable timeframe. The principle of NMR spectroscopy is the fact that NMR active nuclei (e.g. ^{1}H, ^{13}C, ^{15}N) when exposed to a strong magnetic field, possess a characteristic precession. If radiofrequency (*Rf*), which matches the precessional frequency (Larmor frequency v_0) of the nucleus, is applied perpendicular to the external magnetic field, absorption occurs and the nucleus changes from a lower to a higher energy level. The relaxation process (return to the lower energy level) causes a change in the impedance of the oscillator coils which is detected as signal, known as free induction decay (FID). Fourier transformation of this time-dependent signal gives the conventional NMR spectrum. Depending on the structural surroundings, a nucleus is shielded or deshielded and has thus a different precessional frequency. The chemical shift value indicates the extent of this shielding/deshielding effect and can be thus used to distinguish between chemical inequivalent groups. The scalar coupling (*J* in Hz) and the resulting splitting pattern and dipolar coupling are used for constitutional and configurational assignment of the molecule.

1.5.2.2 Determination of Stereochemistry

Many NPs carry in their structure stereocenters resulting from specific steps during biosynthesis. The assignment of the configuration of each stereocenter in a particular molecule is a relevant step during structure elucidation of NPs. The stereochemistry influences the chemical and physical properties of a given compound and hence the pharmacological properties including biological mode of action. This fact is evident for many enantiomeric pairs of compounds. For instance, (+)-nanaomycin A possesses antibacterial activity, however its enantiomeric counterpart (-)-nanaomycin A exhibits additionally antifungal and antimycoplasma activities.[112] Due to the different 3-dimensional structures of enantiomers, the mode of action of the two chiral drugs could significantly differ due to specifically targeting different binding sites of the inhibited enzyme or even inactivating different processes within the cell. Thus unambiguous determination of the stereochemistry of NP is very important, since different enantiomers in a racemic mixture can be more or less active, inactive, provoke undesired side effects or be even toxic.

a) Relative Configuration

The relative stereochemistry of a molecule describes the position of two chiral centers in relation to each other. The NOESY (nuclear Overhauser effect spectroscopy) experiment, which is based on the observation of interproton nuclear Overhauser effects (nOe), provides one possibility to reveal the relative configuration of two neighbored stereocenters. Cross-peaks appearing in the NOESY spectrum connect protons that are spatially close to each other. In general, a strong signal can be seen for a pair of protons with a distance of 1.8-2.5 Å, a medium or weak signal for distances of 2.5-6 Å and no signal for distances larger than 6 Å.[113] However, in NOESY experiments, nOe signals can become very close to zero in the case of molecules with size of 1000-3000 Da.[114] In such cases the ROESY experiment (rotating frame nuclear Overhauser effect spectroscopy) is used to circumvent the signal loss. Although the nOe based assignment is very reliable for cyclic compounds, it has its limits when it comes to non-cyclic, highly flexible molecules, which can coexist in multiple conformers. In some circumstances, the large contribution of minor conformers to the nOe signal can lead to inconsistent results. NOESY or ROESY experiments are mainly used for the determination of the cis/trans configuration of double bonds or cyclic structures (Fig. 10).

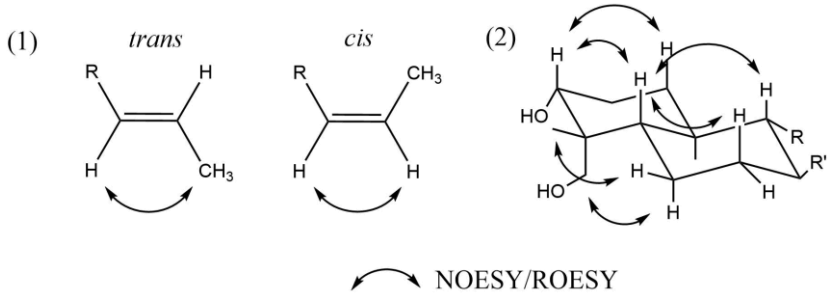


Figure 10. NOESY/ROESY correlation for (1) cis/trans geometry of a double bond and (2) a cyclic system.[115]

The determination of the relative stereochemistry in acyclic systems is also possible via the *J*-based (*J* = spin coupling constant) configurational analysis.[116] The vicinal coupling constants of two protons ($^3J_{HH}$) or a proton and a carbon ($^3J_{HC}$) are strongly dependent on the dihedral angle ϕ (Karplus equation) and can be thus used for the relative configurational assignment in a polysubstituted acyclic molecule.[90] The geminal carbon-proton coupling ($^2J_{HC}$) can be taken into account as well, if the carbon (C_3) bears an electronegative substituent. The size of the $^2J_{HC}$ value then depends on the angle between the proton and the electronegative substituent Z (Fig. 11). Considering the Newman projection of a 2,3-disubstituted butane system, the vicinal coupling constant ranges from 0-16 Hz for $^3J_{H2H3}$ and from 0-9 Hz for $^3J_{H2C4}$. The geminal

coupling constant $^{2}J_{H2C3}$ values ranges from -6 to 8 Hz. In all cases the value additionally strongly depends on the substitution pattern. Based on these empirical values and all $^{2,3}J_{HX}$ measured values extracted from the respective NMR experiment, the relative conformation of two or more adjoining or alternating chiral centers in a linear molecule can be estimated.

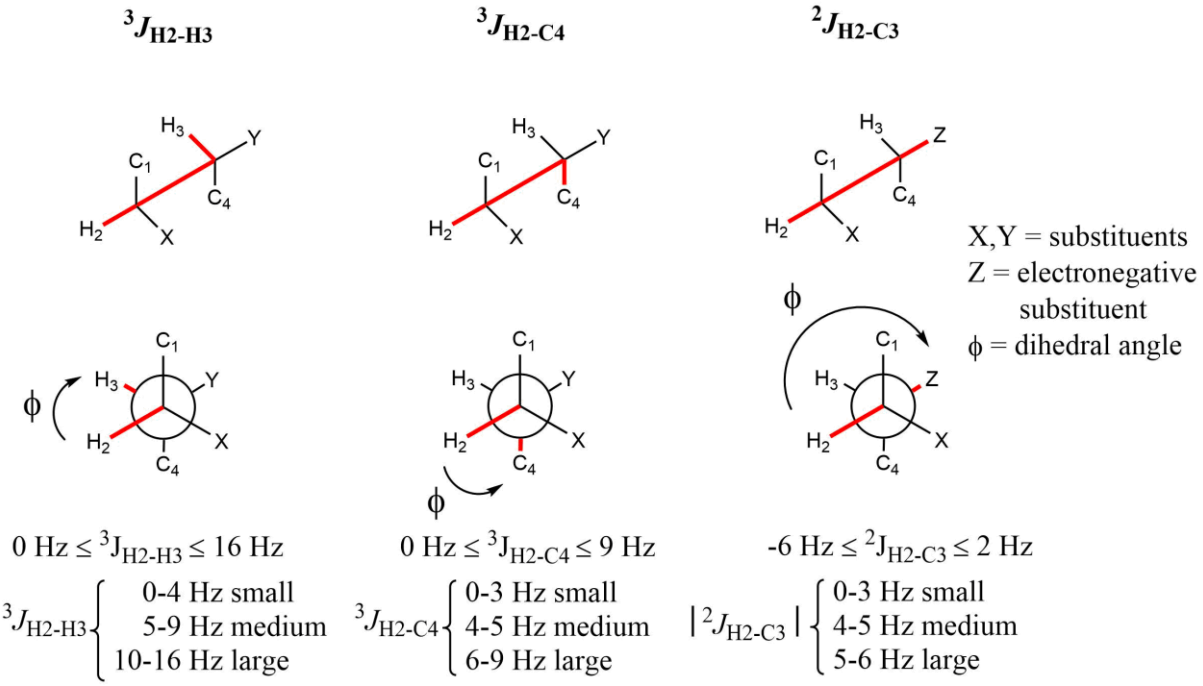


Figure 11. Estimation of vicinal and geminal coupling constants of a linear system dependent on the dihedral angle ϕ.[117]

Although the *J*-based configurational assignment became a powerful and reliable tool for the determination of the relative stereochemistry in acyclic systems, some possible limitations should be mentioned. In principle, the method considers only staggered rotamers of the two possible configurations *threo* and *erythro*. Thus, if a given conformer differs more than 15° from the staggered arrangement, the analysis of the *J*-values may lead to false results.[118] Also, the range of variability of $^{2,3}J_{HX}$ depends strongly on the electronegativity of the substituents attached to the stereogenic center. Thus, a correct prediction of *J* values in advance may be difficult.[119] Lastly, the *J*-based configurational analysis considers only the most populated staggered rotamer while the remaining rotamers are neglected.[116] Therefore, the unambiguous assignment of the relative stereochemistry can be impossible in some cases. Nevertheless, the *J*-based method became an indispensable tool and facilitated the assignment of the relative stereochemistry in many NPs.

b) Absolute configuration

The relative stereochemistry describes solely the arrangement of two stereocenters in relation to each other. However, it is necessary to determine the absolute configuration of the compound of interest. A chiral center is assigned as *R* or *S* in accordance to the spatial arrangement of the substituents following the Cahn-Ingold-Prelog (CIP) convention. Before NMR became the method of choice for the assignment of the absolute configuration, it was more common to use for instance chiroptical methods like circular dichroism (CD) and optical rotatory dispersion (ORD), crystallographic X-ray diffraction (XRD) or rotation dependent methods like infrared vibrational CD or Raman-spectroscopy.[120-123] Although, these methods yield reliable results they require large amounts of the compound and even crystallized samples in the case of XRD. These reasons were decisive for the development of suitable NMR techniques for the determination of the absolute configuration. The use of NMR requires comparatively small amount of sample which can be recovered, the method is applicable to both solid and liquid samples and NMR instruments are usually more often available and easier to use. In general, there are two approaches for the assignment of the absolute configuration using NMR.[124] The first is based on the fact that the addition of a chiral solvating agent (CSA) to the non-chiral standard NMR solvent provokes a difference in the chemical shifts compared to the sample without CSA. The second approach includes the derivatization of the sample of interest with two enantiomers of a chiral derivatization agent (CDA), whereas the differences of the chemical shifts of the two resulting diastereomers can be used for the determination of the absolute configuration. Although many CDAs have been developed so far, the "Mosher's method" first described in 1973 which uses α-methoxy-α-trifluoro-methylphenylacetic acid (MTPA) as chiral derivatization agent, has been used most often.[125] The “Mosher’s method” takes advantage of the fact that the phenyl group of the derivatization agent exerts an anisotropic effect on both residues R1 and R2 of the stereocenter. The preparation of the sample with both diastereomers of MTPA results in two (R)- and (S)-MTPA ester derivatives which differ in their chemical shift values when compared to each other (Fig. 12).[126] This chemical shift difference is expressed in the value $\Delta\delta^{SR}$. If this value becomes positive for R2 (shielded by phenyl group) and negative for R1 (unaffected by phenylgroup), the residues have to be arranged as it is shown in Fig. 12 (left) and vice versa if the signals for R1 and R2 are opposite (Fig. 12 right). Mosher’s method is a powerful tool for the assignment of the absolute configuration and has been successfully applied in many cases. Nevertheless, it has its limitations. For a reliable assignment, the $\Delta\delta^{SR}$ should lie above the experimental error and the sign distribution within R1 and R2 has to be uniform. These criteria are not always met by MTPA, which engendered

the development of CDAs adjusted to specialized cases resulting in much cleaner chemical shifts.[124]

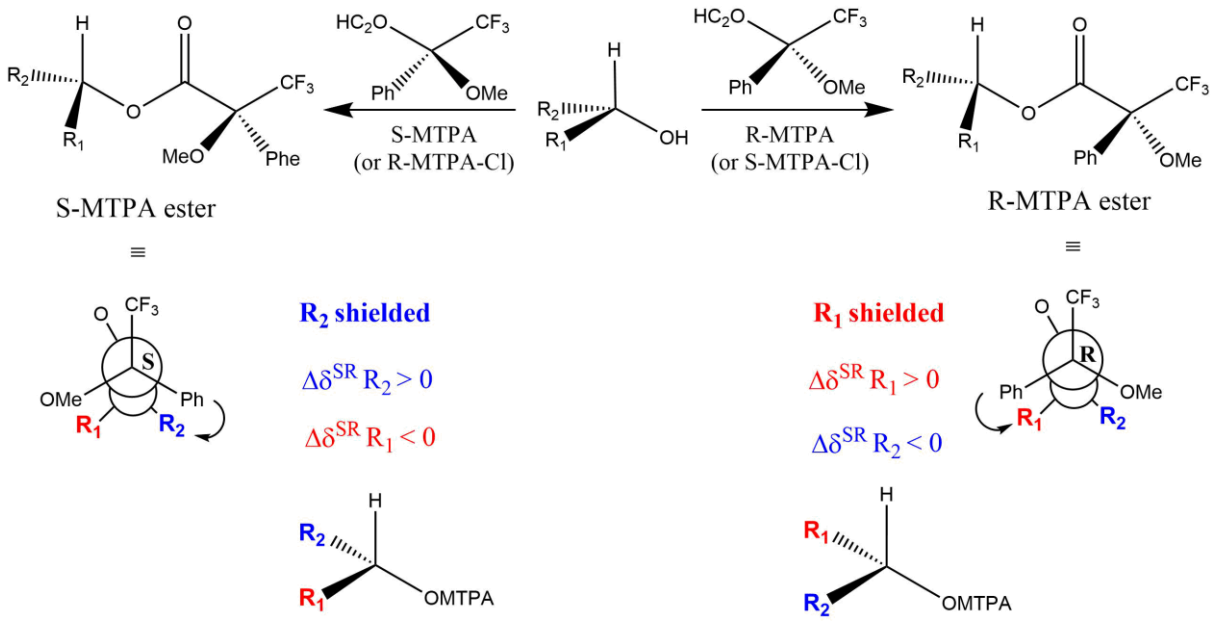


Figure 12. Mosher reaction of a secondary alcohol with R/S-MTPA yielding R/S-MTPA ester. Shielding effects are indicated by arrows in the respective Newman projection. Expected chemical shift differences $\Delta\delta^{SR}$ and subsequent arrangement of R_1 and R_2.[124,126]

In conclusion, there is no universal applicable procedure for the determination of a 3-dimensional structure, since the success of the assignment is highly dependent on the structural properties of a compound. NPs have often highly complex chemical structures which complicates the determination of the absolute configuration. A combination of different methods is thus often needed for the unambiguous assignment of stereocenters. Nevertheless, NMR based methods are often the first choice due to the rather simple sample preparation procedure and the possibility to recover the sample in most cases.

1.6 Outline of the work

The aim of the presented work was the isolation and characterization of novel NPs aided by a metabolic and genomic evaluation of the actinobacteria strains isolated from unusual environmental samples. The combination of analysis steps, such as dereplication of produced secondary metabolites, activity tests against Gram-negative and Gram-positive bacteria and genome mining/BGC analysis lead to pre-selection of two promising streptomycetes species, chosen for further examination. As complement to this, a different approach for the detection of novel secondary metabolites has been used during the study presented in Chapter 4.

The genome of the marine *Streptomyces* sp. MP131-18 was sequenced and 36 gene clusters involved in the biosynthesis of secondary metabolites were identified using antiSMASH. The metabolic analysis revealed the production of the two compound classes: bisindoles and pyrones. Isolation and structure elucidation using spectroscopic methods (MS and NMR) lead to the identification of two new spiroindimicins E and F. Furthermore, two new α-pyrone lagunapyrones D and E were predicted by analysis of respective gene cluster and identified by tandem MS fragmentation approach. The biosynthetic pathways for both groups of compounds were proposed. As described in Chapter 2, the results show that marine actinomycetes represent a promising source of new NPs and that a combined genomic-metabolic approach leads indeed to new compounds and enable the identification of their biosynthetic origin.

Chapter 3 describes the examination of *Streptomyces* sp. IB2014/011-12 strain isolated from samples collected at Lake Baikal with potent activity against Gram-positive bacteria. Activity-guided isolation leads to the identification of the linear polyketide alpiniamide A and its new derivatives B-D. The structures were established by NMR and relative and absolute configurations were assigned for alpiniamide A and B. The alpiniamides biosynthesis gene cluster was identified by genome mining, feeding of labeled precursors and gene deletion experiments. The heterologous expression of the cloned gene cluster confirmed these findings, so that the biosynthesis of alpiniamides could be clearly ascribed to a hybrid trans AT- PKS-NRPS megaenzyme with an unusual domain arrangement.

The last chapter deals with the development of a new transcriptional repressor-based biosensor concept, which allows the detection of a specific NP resulted from activation of a secondary metabolism gene cluster in streptomycetes. The biosynthetic gene cluster for the metabolites undecylprodigiosin (*red*) and coelimycin P1 (*cpk*) present in the genome of *Streptomyces lividans* TK24 have been used for the pivotal study. The biosensor construct was developed by placing the *bpsA* reporter gene for indigoidine synthetase under the control of the promoter/operator region, which is controlled by the TetR-like repressor originated from either *red* or *cpk* gene cluster. The successful application of the coelimycin biosensor was shown, since the biosynthesis of indigoidine was induced in response to the production of coelimycin. Such reporter/biosensor systems can be used in broader applications for the detection of activated cryptic gene cluster, which encode metabolite-sensing repressors proteins.

References

1 Newman, D. J. & Cragg, G. M. Natural Products as Sources of New Drugs from 1981 to 2014. *Journal of natural products* **79**, 629–661, doi:10.1021/acs.jnatprod.5b01055 (2016).

2 *The top 10 causes of death*, <http://www.who.int/news-room/fact-sheets/detail/the-top-10-causes-of-death> (06.10.2018)

3 Li, J. W.-H. & Vederas, J. C. Drug discovery and natural products: End of an era or an endless frontier? *Science (New York, N.Y.)* **325**, 161–165, doi:10.1126/science.1168243 (2009).

4 Pereira, D. A. & Williams, J. A. Origin and evolution of high throughput screening. *British journal of pharmacology* **152**, 53–61, doi:10.1038/sj.bjp.0707373 (2007).

5 Davies, J. & Davies, D. Origins and evolution of antibiotic resistance. *Microbiology and molecular biology reviews : MMBR* **74**, 417–433, doi:10.1128/mmbr.00016-10 (2010).

6 Santajit, S. & Indrawattana, N. Mechanisms of Antimicrobial Resistance in ESKAPE Pathogens. *BioMed research international* **2016**, 2475067, doi:10.1155/2016/2475067 (2016).

7 The 10 x '20 Initiative: Pursuing a global commitment to develop 10 new antibacterial drugs by 2020. *Clinical infectious diseases : an official publication of the Infectious Diseases Society of America* **50**, 1081–1083, doi:10.1086/652237 (2010).

8 Deak, D., Outterson, K., Powers, J. H. & Kesselheim, A. S. Progress in the Fight Against Multidrug-Resistant Bacteria? A Review of U.S. Food and Drug Administration-Approved Antibiotics, 2010-2015. *Annals of internal medicine* **165**, 363–372, doi:10.7326/m16-0291 (2016).

9 Cho, J. C., Zmarlicka, M. T., Shaeer, K. M. & Pardo, J. Meropenem/Vaborbactam, the First Carbapenem/β-Lactamase Inhibitor Combination. *The Annals of pharmacotherapy* **52**, 769–779, doi:10.1177/1060028018763288 (2018).

10 Microbiology, E.-E. S. o. C. & Diseases, I. *ESCMID: Research & Projects*, <https://www.escmid.org/research_projects/> (06.10.2018)

11 Bassetti, M., Ginocchio, F. & Mikulska, M. New treatment options against gram-negative organisms. *Critical care (London, England)* **15**, 215, doi:10.1186/cc9997 (2011).

12 Baltz, R. H. Antimicrobials from actinomycetes: Back to the future. *Microbe-American Society For Microbiology* **2**, 125 (2007).

13 Genilloud, O. Actinomycetes: Still a source of novel antibiotics. *Natural product reports* **34**, 1203–1232, doi:10.1039/c7np00026j (2017).

14 Schatz, A., Bugle, E. & Waksman, S. A. Streptomycin, a Substance Exhibiting Antibiotic Activity Against Gram-Positive and Gram-Negative Bacteria.*. *Experimental Biology and Medicine* **55**, 66–69, doi:10.3181/00379727-55-14461 (1944).

15 Procópio, R. E. d. L., Silva, I. R. d., Martins, M. K., Azevedo, J. L. d. & Araújo, J. M. d. Antibiotics produced by Streptomyces. *The Brazilian journal of infectious diseases : an official publication of the Brazilian Society of Infectious Diseases* **16**, 466–471, doi:10.1016/j.bjid.2012.08.014 (2012).

16 Clardy, J., Fischbach, M. A. & Currie, C. R. The natural history of antibiotics. *Current biology : CB* **19**, R437-441, doi:10.1016/j.cub.2009.04.001 (2009).

17 Sousa, C. d. S., Soares, A. C. F. & Garrido, M. d. S. Characterization of streptomycetes with potential to promote plant growth and biocontrol. *Scientia Agricola* **65**, 50–55, doi:10.1590/s0103-90162008000100007 (2008).

18 Pimentel-Elardo, S. M. *et al.* Anti-parasitic compounds from Streptomyces sp. strains isolated from Mediterranean sponges. *Marine drugs* **8**, 373–380, doi:10.3390/md8020373 (2010).

19 Takahashi, Y. & Nakashima, T. Actinomycetes, an Inexhaustible Source of Naturally Occurring Antibiotics. **7**, doi:10.3390/antibiotics7020045 (2018).

20 McDonald, B. R. & Currie, C. R. Lateral Gene Transfer Dynamics in the Ancient Bacterial Genus Streptomyces. *mBio* **8**, doi:10.1128/mBio.00644-17 (2017).

21 Fenical, W. & Jensen, P. R. Developing a new resource for drug discovery: Marine actinomycete bacteria. *Nature chemical biology* **2**, 666–673, doi:10.1038/nchembio841 (2006).

22 Fiedler, H.-P. *et al.* Marine actinomycetes as a source of novel secondary metabolites. *Antonie van Leeuwenhoek* **87**, 37–42, doi:10.1007/s10482-004-6538-8 (2005).

23 Olano, C., Méndez, C. & Salas, J. A. Antitumor compounds from marine actinomycetes. *Marine drugs* **7**, 210–248, doi:10.3390/md7020210 (2009).

24 Lam, K. S. Discovery of novel metabolites from marine actinomycetes. *Current opinion in microbiology* **9**, 245–251, doi:10.1016/j.mib.2006.03.004 (2006).

25 Subramani, R. & Aalbersberg, W. Marine actinomycetes: An ongoing source of novel bioactive metabolites. *Microbiological research* **167**, 571–580, doi:10.1016/j.micres.2012.06.005 (2012).

26 Engelhardt, K. *et al.* Production of a new thiopeptide antibiotic, TP-1161, by a marine Nocardiopsis species. *Applied and environmental microbiology* **76**, 4969–4976, doi:10.1128/aem.00741-10 (2010).

27 Asolkar, R. N. *et al.* Arenamides A-C, cytotoxic NFkappaB inhibitors from the marine actinomycete Salinispora arenicola. *Journal of natural products* **72**, 396–402, doi:10.1021/np800617a (2009).

28 Seipke, R. F., Kaltenpoth, M. & Hutchings, M. I. Streptomyces as symbionts: An emerging and widespread theme? *FEMS microbiology reviews* **36**, 862–876, doi:10.1111/j.1574-6976.2011.00313.x (2012).

29 Pimentel-Elardo, S. M., Scheuermayer, M., Kozytska, S. & Hentschel, U. Streptomyces axinellae sp. nov., isolated from the Mediterranean sponge Axinella polypoides (Porifera). *International journal of systematic and evolutionary microbiology* **59**, 1433–1437, doi:10.1099/ijs.0.007856-0 (2009).

30 Pimentel-Elardo, S. M. *et al.* New tetromycin derivatives with anti-trypanosomal and protease inhibitory activities. *Marine drugs* **9**, 1682–1697, doi:10.3390/md9101682 (2011).

31 Lin, Z. *et al.* Pulicatins A-E, neuroactive thiazoline metabolites from cone snail-associated bacteria. *Journal of natural products* **73**, 1922–1926, doi:10.1021/np100588c (2010).

32 Vartoukian, S. R., Palmer, R. M. & Wade, W. G. Strategies for culture of 'unculturable' bacteria. *FEMS microbiology letters* **309**, 1–7, doi:10.1111/j.1574-6968.2010.02000.x (2010).

33 Ling, L. L. *et al.* A new antibiotic kills pathogens without detectable resistance. *Nature* **517**, 455–459, doi:10.1038/nature14098 (2015).

34 Fiers, W. D., Craighead, M. & Singh, I. Teixobactin and Its Analogues: A New Hope in Antibiotic Discovery. *ACS infectious diseases* **3**, 688–690, doi:10.1021/acsinfecdis.7b00108 (2017).

35 Brady, S. F., Simmons, L., Kim, J. H. & Schmidt, E. W. Metagenomic approaches to natural products from free-living and symbiotic organisms. *Natural product reports* **26**, 1488–1503, doi:10.1039/b817078a (2009).

36 Hover, B. M. *et al.* Culture-independent discovery of the malacidins as calcium-dependent antibiotics with activity against multidrug-resistant Gram-positive pathogens. *Nature Microbiology* **3**, 415–422, doi:10.1038/s41564-018-0110-1 (2018).

37 Beutler, J. A., Alvarado, A. B., Schaufelberger, D. E., Andrews, P. & McCloud, T. G. Dereplication of Phorbol Bioactives: Lyngbya majuscula and Croton cuneatus. *Journal of natural products* **53**, 867–874, doi:10.1021/np50070a014 (1990).

38 Sticher, O. Natural product isolation. *Natural product reports* **25**, 517–554, doi:10.1039/b700306b (2008).

39 Gaudêncio, S. P. & Pereira, F. Dereplication: Racing to speed up the natural products discovery process. *Natural product reports* **32**, 779–810, doi:10.1039/c4np00134f (2015).

40 *SciFinder | CAS*, <https://www.cas.org/products/scifinder> (06.10.2018)

41 *The PubChem Project*, <https://pubchem.ncbi.nlm.nih.gov/> (06.10.2018)

42 *Dictionary of natural products on DVD*. (CRC Press, 2012).

43 Land, M. *et al.* Insights from 20 years of bacterial genome sequencing. *Functional & integrative genomics* **15**, 141–161, doi:10.1007/s10142-015-0433-4 (2015).

44 Ziemert, N., Alanjary, M. & Weber, T. The evolution of genome mining in microbes - a review. *Natural product reports* **33**, 988–1005, doi:10.1039/c6np00025h (2016).

45 *Ecopia Announces the Discovery of Two New Compounds*, <https://www.businesswire.com/news/home/20040413005220/en/Ecopia-Announces-Discovery-New-Compounds> (06.10.2018)

46 Li, M. H. T., Ung, P. M. U., Zajkowski, J., Garneau-Tsodikova, S. & Sherman, D. H. Automated genome mining for natural products. *BMC bioinformatics* **10**, 185, doi:10.1186/1471-2105-10-185 (2009).

47 Weber, T. *et al.* CLUSEAN: A computer-based framework for the automated analysis of bacterial secondary metabolite biosynthetic gene clusters. *Journal of Biotechnology* **140**, 13–17, doi:10.1016/j.jbiotec.2009.01.007 (2009).

48 Blin, K. *et al.* antiSMASH 2.0--a versatile platform for genome mining of secondary metabolite producers. *Nucleic acids research* **41**, W204-212, doi:10.1093/nar/gkt449 (2013).

49 Skinnider, M. A., Merwin, N. J., Johnston, C. W. & Magarvey, N. A. PRISM 3: Expanded prediction of natural product chemical structures from microbial genomes. *Nucleic acids research* **45**, W49-W54, doi:10.1093/nar/gkx320 (2017).

50 *antibiotics & Secondary Metabolite Analysis SHell Version 4.2.0*, <https://antismash.secondarymetabolites.org/#!/start> (06.10.2018)

51 Medema, M. H. *et al.* Minimum Information about a Biosynthetic Gene cluster. *Nature chemical biology* **11**, 625 EP -, doi:10.1038/nchembio.1890 (2015).

52 Alanjary, M. *et al.* The Antibiotic Resistant Target Seeker (ARTS), an exploration engine for antibiotic cluster prioritization and novel drug target discovery. *Nucleic acids research* **45**, W42-W48, doi:10.1093/nar/gkx360 (2017).

53 Hertweck, C. The biosynthetic logic of polyketide diversity. *Angew Chem Int Ed Engl* **48**, 4688-4716, doi:10.1002/anie.200806121 (2009).

54 Moore, B. S. & Hertweck, C. Biosynthesis and attachment of novel bacterial polyketide synthase starter units. *Natural Product Reports* **19**, 70-99, doi:10.1039/B003939J (2002).

55 Chan, Y. A., Podevels, A. M., Kevany, B. M. & Thomas, M. G. Biosynthesis of polyketide synthase extender units. *Nat Prod Rep* **26**, 90-114 (2009).

56 Keatinge-Clay, A. T. The structures of type I polyketide synthases. *Nat Prod Rep* **29**, 1050-1073, doi:10.1039/c2np20019h (2012).

57 Wiley, P. F., Gerzon, K., Flynn, E. H., Sigal, M. V. & Quarck, U. C. Erythromycin. V. Isolation and Structure of Degradation Products. *Journal of the American Chemical Society* **77**, 3677-3678, doi:10.1021/ja01618a092 (1955).

58 Carr, G. *et al.* Bafilomycins Produced in Culture by Streptomyces spp. Isolated from Marine Habitats Are Potent Inhibitors of Autophagy. *Journal of Natural Products* **73**, 422-427, doi:10.1021/np900632r (2010).

59 Alt, S. & Wilkinson, B. Biosynthesis of the Novel Macrolide Antibiotic Anthracimycin. *ACS Chemical Biology* **10**, 2468-2479, doi:10.1021/acschembio.5b00525 (2015).

60 Fohner, A. E., Sparreboom, A., Altman, R. B. & Klein, T. E. PharmGKB summary: Macrolide antibiotic pathway, pharmacokinetics/pharmacodynamics. *Pharmacogenetics and genomics* **27**, 164–167, doi:10.1097/fpc.0000000000000270 (2017).

61 Serhan, G., Stack, C. M., Perrone, G. G. & Morton, C. O. The polyene antifungals, amphotericin B and nystatin, cause cell death in Saccharomyces cerevisiae by a distinct mechanism to amphibian-derived antimicrobial peptides. *Annals of clinical microbiology and antimicrobials* **13**, 18, doi:10.1186/1476-0711-13-18 (2014).

62 Kotler-Brajtburg, J. *et al.* Classification of polyene antibiotics according to chemical structure and biological effects. *Antimicrobial Agents and Chemotherapy* **15**, 716–722 (1979).

63 Hammond, S. M. & Kliger, B. N. Mode of action of the polyene antibiotic candicidin: Binding factors in the wall of Candida albicans. *Antimicrobial Agents and Chemotherapy* **9**, 561–568 (1976).

64 Kevin Ii, D. A., Meujo, D. A. & Hamann, M. T. Polyether ionophores: Broad-spectrum and promising biologically active molecules for the control of drug-resistant bacteria and parasites. *Expert opinion on drug discovery* **4**, 109–146, doi:10.1517/17460440802661443 (2009).

65 Rutkowski, J. & Brzezinski, B. Structures and properties of naturally occurring polyether antibiotics. *BioMed research international* **2013**, 162513, doi:10.1155/2013/162513 (2013).

66 Kohli, G. S., John, U., Van Dolah, F. M. & Murray, S. A. Evolutionary distinctiveness of fatty acid and polyketide synthesis in eukaryotes. *The Isme Journal* **10**, 1877, doi:10.1038/ismej.2015.263 (2016).

67 Zhang, Q., Pang, B., Ding, W. & Liu, W. Aromatic Polyketides Produced by Bacterial Iterative Type I Polyketide Synthases. *ACS Catalysis* **3**, 1439-1447, doi:10.1021/cs400211x (2013).

68 Dimroth, P., Walter, H. & Lynen, F. Biosynthese von 6-Methylsalicylsäure. *European journal of biochemistry* **13**, 98–110 (1970).

69 Dimroth, P., Ringelmann, E. & Lynen, F. 6-Methylsalicylic acid synthetase from Penicillium patulum. Some catalytic properties of the enzyme and its relation to fatty acid synthetase. *European journal of biochemistry* **68**, 591–596 (1976).

70 Daum, M. *et al.* Organisation of the biosynthetic gene cluster and tailoring enzymes in the biosynthesis of the tetracyclic quinone glycoside antibiotic polyketomycin. *Chembiochem : a European journal of chemical biology* **10**, 1073–1083, doi:10.1002/cbic.200800823 (2009).

71 Weitnauer, G. *et al.* Biosynthesis of the orthosomycin antibiotic avilamycin A: Deductions from the molecular analysis of the avi biosynthetic gene cluster of Streptomyces viridochromogenes Tü57 and production of new antibiotics. *Chemistry & Biology* **8**, 569–581, doi:10.1016/s1074-5521(01)00040-0 (2001).

72 Mascaretti, O. A. *et al.* Biosynthesis of the macrolide antibiotic chlorothricin. *Biochemistry* **20**, 919–924, doi:10.1021/bi00507a042 (2002).

73 Piel, J. Biosynthesis of polyketides by trans-AT polyketide synthases. *Natural Product Reports* **27**, 996-1047, doi:10.1039/B816430B (2010).

74 Helfrich, E. J. N. & Piel, J. Biosynthesis of polyketides by trans-AT polyketide synthases. *Natural Product Reports* **33**, 231-316, doi:10.1039/C5NP00125K (2016).

75 Moldenhauer, J., Chen, X.-H., Borriss, R. & Piel, J. Biosynthesis of the antibiotic bacillaene, the product of a giant polyketide synthase complex of the trans-AT family. *Angewandte Chemie (International ed. in English)* **46**, 8195–8197, doi:10.1002/anie.200703386 (2007).

76 Gurney, R. & Thomas, C. M. Mupirocin: Biosynthesis, special features and applications of an antibiotic from a gram-negative bacterium. *Applied microbiology and biotechnology* **90**, 11–21, doi:10.1007/s00253-011-3128-3 (2011).

77 Schneider, K. *et al.* Macrolactin is the polyketide biosynthesis product of the pks2 cluster of Bacillus amyloliquefaciens FZB42. *Journal of natural products* **70**, 1417–1423, doi:10.1021/np070070k (2007).

78 Trindade-Silva, A. E., Lim-Fong, G. E., Sharp, K. H. & Haygood, M. G. Bryostatins: Biological context and biotechnological prospects. *Current opinion in biotechnology* **21**, 834–842, doi:10.1016/j.copbio.2010.09.018 (2010).

79 Ng, B. G., Han, J. W., Lee, D. W., Choi, G. J. & Kim, B. S. The chejuenolide biosynthetic gene cluster harboring an iterative trans-AT PKS system in Hahella chejuensis strain MB-1084. *The Journal of antibiotics* **71**, 495–505, doi:10.1038/s41429-017-0023-x (2018).

80 Keatinge-Clay, A. T. Stereocontrol within polyketide assembly lines. *Natural product reports* **33**, 141–149, doi:10.1039/c5np00092k (2016).

81 Valenzano, C. R. *et al.* Stereospecificity of the dehydratase domain of the erythromycin polyketide synthase. *Journal of the American Chemical Society* **132**, 14697–14699, doi:10.1021/ja107344h (2010).

82 Xie, X., Khosla, C. & Cane, D. E. Elucidation of the Stereospecificity of C-Methyltransferases from trans-AT Polyketide Synthases. *Journal of the American Chemical Society* **139**, 6102–6105, doi:10.1021/jacs.7b02911 (2017).

83 Das, A. & Khosla, C. Biosynthesis of aromatic polyketides in bacteria. *Accounts of chemical research* **42**, 631-639, doi:10.1021/ar8002249 (2009).

84 Brockmann, H. & Loeschcke, V. Actinomycetenfarbstoffe, IV. Mitteil.: über die Konstitution des Actinorhodins und die Isolierung des Proto-actinorhodins. *Chemische Berichte* **88**, 778-788, doi:doi:10.1002/cber.19550880607 (1955).

85 Hu, Y., Martinez, E. D. & MacMillan, J. B. Anthraquinones from a marine-derived Streptomyces spinoverrucosus. *Journal of natural products* **75**, 1759–1764, doi:10.1021/np3004326 (2012).

86 Duggar, B. M. Aureomycin: A Product of the continuing Search for new Antibiotics. *Annals of the New York Academy of Sciences* **51**, 177–181, doi:10.1111/j.1749-6632.1948.tb27262.x (1948).

87 Lechevalier, M. P., Prauser, H., Labeda, D. P. & Ruan, J.-S. Two New Genera of Nocardioform Actinomycetes: Amycolata gen. nov. and Amycolatopsis gen. nov. *International Journal of Systematic Bacteriology* **36**, 29–37, doi:10.1099/00207713-36-1-29 (1986).

88 Arcamone, F. *et al.* New developments in antitumor anthracyclines. *Pharmacology & therapeutics* **76**, 117–124 (1997).

89 Arcamone, F. *et al.* Adriamycin, 14-hydroxydaunomycin, a new antitumor antibiotic from S. peucetius var. caesius. *Biotechnology and bioengineering* **11**, 1101–1110, doi:10.1002/bit.260110607 (1969).

90 Wiley, P. F., MacKellar, F. A., Caron, E. L. & Kelly, R. B. Isolation, characterization and degradation of nogalamycin. *Tetrahedron Letters* **9**, 663-668, doi:https://doi.org/10.1016/S0040-4039(00)75607-3 (1968).

91 Henkel, T., Ciesiolka, T., Rohr, J. & Zeeck, A. Urdamycins, new angucycline antibiotics from Streptomyces fradiae. V. Derivatives of urdamycin A. *J Antibiot (Tokyo)* **42**, 299-311 (1989).

92 Sezaki, M., Kondo, S., Maeda, K., Umezawa, H. & Ohno, M. The structure of aquayamycin. *Tetrahedron* **26**, 5171–5190, doi:10.1016/s0040-4020(01)98726-5 (1970).

93 Henkel, T., Rohr, J., Beale, J. M. & Schwenen, L. Landomycins, new angucycline antibiotics from Streptomyces sp. I. Structural studies on landomycins A-D. *The Journal of antibiotics* **43**, 492–503, doi:10.7164/antibiotics.43.492 (1990).

94 Schimana, J. *et al.* Simocyclinones, novel cytostatic angucyclinone antibiotics produced by Streptomyces antibioticus Tü 6040. *The Journal of antibiotics* **53**, 779-787 (2000).

95 Shimizu, Y. & Ogata, H. Type III Polyketide Synthases: Functional Classification and Phylogenomics. **18**, 50-65, doi:10.1002/cbic.201600522 (2017).

96 Strieker, M., Tanovic, A. & Marahiel, M. A. Nonribosomal peptide synthetases: structures and dynamics. *Current opinion in structural biology* **20**, 234-240, doi:10.1016/j.sbi.2010.01.009 (2010).

97 Stachelhaus, T., Mootz, H. D. & Marahiel, M. A. The specificity-conferring code of adenylation domains in nonribosomal peptide synthetases. *Chem Biol* **6**, 493-505, doi:10.1016/s1074-5521(99)80082-9 (1999).

98 Hur, G. H., Vickery, C. R. & Burkart, M. D. Explorations of catalytic domains in non-ribosomal peptide synthetase enzymology. *Nat Prod Rep* **29**, 1074-1098, doi:10.1039/c2np20025b (2012).

99 Walsh, C. T., O'Brien, R. V. & Khosla, C. Nonproteinogenic amino acid building blocks for nonribosomal peptide and hybrid polyketide scaffolds. *Angew Chem Int Ed Engl* **52**, 7098-7124, doi:10.1002/anie.201208344 (2013).

100 McCormick, M. H., McGuire, J. M., Pittenger, G. E., Pittenger, R. C. & Stark, W. M. Vancomycin, a new antibiotic. I. Chemical and biologic properties. *Antibiot Annu* **3**, 606-611 (1955).

101 Johnson, B. A., Anker, H. & Meleney, F. L. Bacitracin: a new antibiotic produced by a member of the B. subtilis group. *Science* **102**, 376-377 (1945).

102 Du, L., Sánchez, C. & Shen, B. Hybrid peptide-polyketide natural products: Biosynthesis and prospects toward engineering novel molecules. *Metabolic engineering* **3**, 78–95, doi:10.1006/mben.2000.0171 (2001).

103 Du, L., Sánchez, C., Chen, M., Edwards, D. J. & Shen, B. The biosynthetic gene cluster for the antitumor drug bleomycin from Streptomyces verticillus ATCC15003 supporting functional interactions between nonribosomal peptide synthetases and a polyketide synthase. *Chemistry & Biology* **7**, 623-642, doi:https://doi.org/10.1016/S1074-5521(00)00011-9 (2000).

104 Molnár, I. *et al.* The biosynthetic gene cluster for the microtubule-stabilizing agents epothilones A and B from Sorangium cellulosum So ce90. *Chemistry & Biology* **7**, 97-109, doi:https://doi.org/10.1016/S1074-5521(00)00075-2 (2000).

105 Alkhalaf, Lona M. & Ryan, Katherine S. Biosynthetic Manipulation of Tryptophan in Bacteria: Pathways and Mechanisms. *Chemistry & Biology* **22**, 317-328, doi:https://doi.org/10.1016/j.chembiol.2015.02.005 (2015).

106 Ryan, K. S. & Drennan, C. L. Divergent pathways in the biosynthesis of bisindole natural products. *Chemistry & biology* **16**, 351-364 (2009).

107 Onaka, H. Biosynthesis of heterocyclic antibiotics in actinomycetes and an approach to synthesize the natural compounds. *Actinomycetologica* **20**, 62–71 (2006).

108 Sarker, S. D. & Nahar, L. An introduction to natural products isolation. *Methods in molecular biology (Clifton, N.J.)* **864**, 1–25, doi:10.1007/978-1-61779-624-1_1 (2012).

109 Cozzolino, D. Near infrared spectroscopy in natural products analysis. *Planta medica* **75**, 746–756, doi:10.1055/s-0028-1112220 (2009).

110 Wolfender, J.-L. HPLC in natural product analysis: The detection issue. *Planta medica* **75**, 719–734, doi:10.1055/s-0028-1088393 (2009).

111 Kaufmann, K. *et al.* Activity-guided screening of bioactive natural compounds implementing a new glucocorticoid-receptor-translocation assay and detection of new anti-inflammatory steroids from bacteria. *Biotechnology letters* **35**, 11–20, doi:10.1007/s10529-012-1042-0 (2013).

112 Finefield, J. M., Sherman, D. H., Kreitman, M. & Williams, R. M. Enantiomeric natural products: Occurrence and biogenesis. *Angewandte Chemie (International ed. in English)* **51**, 4802–4836, doi:10.1002/anie.201107204. (2012).

113 Wüthrich, K. NMR studies of structure and function of biological macromolecules (Nobel lecture). *Angewandte Chemie (International ed. in English)* **42**, 3340–3363, doi:10.1002/anie.200300595 (2003).

114 Atta-ur-Rahman. *Solving Problems with NMR Spectroscopy (Second Edition)*. (Academic Press, 2016).

115 Claridge, T. D. W. *High-Resolution NMR Techniques in Organic Chemistry*. 3. Aufl. edn, (Elsevier Reference Monographs, 2016).

116 Matsumori, N., Kaneno, D., Murata, M., Nakamura, H. & Tachibana, K. Stereochemical Determination of Acyclic Structures Based on Carbon−Proton Spin-Coupling Constants. A Method of Configuration Analysis for Natural Products. *The Journal of Organic Chemistry* **64**, 866–876, doi:10.1021/jo981810k (1999).

117 Bifulco, G., Dambruoso, P., Gomez-Paloma, L. & Riccio, R. Determination of relative configuration in organic compounds by NMR spectroscopy and computational methods. *Chemical reviews* **107**, 3744–3779, doi:10.1021/cr030733c (2007).

118 Matsumori, N., Murata, M. & Tachibana, K. Conformational analysis of natural products using long-range carbon-proton coupling constants: Three-dimensional structure of okadaic acid in solution. *Tetrahedron* **51**, 12229–12238, doi:10.1016/0040-4020(95)00790-f (1995).

119 Contreras, R. H. & Peralta, J. E. Angular dependence of spin-spin coupling constants. *Progress in Nuclear Magnetic Resonance Spectroscopy* **4**, 321–425 (2000).

120 Berova, N., Nakanishi, K., Woody, R. W. & Woody, R. *Circular dichroism: Principles and applications*. (John Wiley & Sons, 2000).

121 Costante, J., Hecht, L., Polavarapu, P. L., Collet, A. & Barron, L. D. Absolute Configuration of Bromochlorofluoromethane from Experimental and Ab Initio Theoretical Vibrational Raman Optical Activity. *Angewandte Chemie International Edition in English* **36**, 885–887, doi:10.1002/anie.199708851 (1997).

122 Polavarapu, P. L. Optical rotation: Recent advances in determining the absolute configuration. *Chirality* **14**, 768–781, doi:10.1002/chir.10145 (2002).

123 Wang, F. *et al.* Absolute Configuration of tert -Butyl-1-(2-methylnaphthyl)phosphine Oxide. *The Journal of Organic Chemistry* **67**, 6539–6541, doi:10.1021/jo025908h (2002).

124 Seco, J. M., Quiñoá, E. & Riguera, R. Assignment of the absolute configuration of polyfunctional compounds by NMR using chiral derivatizing agents. *Chemical reviews* **112**, 4603–4641, doi:10.1021/cr2003344 (2012).

125 Dale, J. A. & Mosher, H. S. Nuclear magnetic resonance enantiomer regents. Configurational correlations via nuclear magnetic resonance chemical shifts of diastereomeric mandelate, O-methylmandelate, and. alpha.-methoxy-. alpha.-trifluoromethylphenylacetate (MTPA) esters. *Journal of the American Chemical Society* **95**, 512–519 (1973).

126 Hoye, T. R., Jeffrey, C. S. & Shao, F. Mosher ester analysis for the determination of absolute configuration of stereogenic (chiral) carbinol carbons. *Nature protocols* **2**, 2451–2458, doi:10.1038/nprot.2007.354 (2007).

2. Publications

I

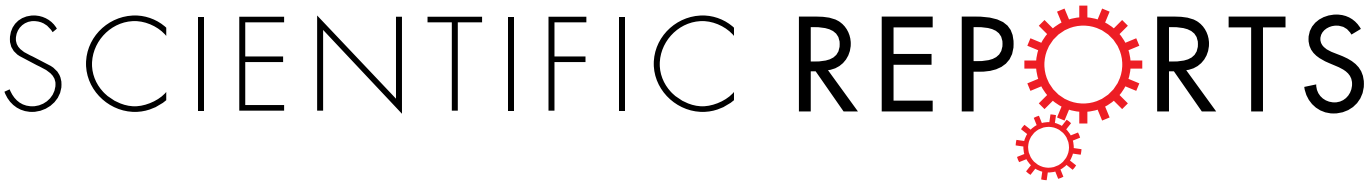
New natural products identified by combined genomics-metabolomics profiling of marine *Streptomyces* sp. MP 131-18

Constanze Paulus, Yuriy Rebets, Bogdan Tokovenko, Suvd Nadmid, Larisa P. Terekhova, Maksym Myronovskyi, Sergey B. Zotchev, Christian Rückert, Simone Braig, Stefan Zahler, Jörn Kalinowski & Andriy Luzhetskyy

Scientific Reports, **2017**, *7*, 42382

DOI: 10.1038/srep42382

Published online 10th February 2017



OPEN

New natural products identified by combined genomics-metabolomics profiling of marine *Streptomyces* sp. MP131-18

Received: 12 June 2016
Accepted: 10 January 2017
Published: 10 February 2017

Constanze Paulus[1], Yuriy Rebets[1], Bogdan Tokovenko[1], Suvd Nadmid[1], Larisa P. Terekhova[2], Maksym Myronovskyi[1], Sergey B. Zotchev[3,4], Christian Rückert[5,†], Simone Braig[6], Stefan Zahler[6], Jörn Kalinowski[5] & Andriy Luzhetskyy[1,7]

Marine actinobacteria are drawing more and more attention as a promising source of new natural products. Here we report isolation, genome sequencing and metabolic profiling of new strain *Streptomyces* sp. MP131-18 isolated from marine sediment sample collected in the Trondheim Fjord, Norway. The 16S rRNA and multilocus phylogenetic analysis showed that MP131-18 belongs to the genus *Streptomyces*. The genome of MP131-18 isolate was sequenced, and 36 gene clusters involved in the biosynthesis of 18 different types of secondary metabolites were predicted using antiSMASH analysis. The combined genomics-metabolics profiling of the strain led to the identification of several new biologically active compounds. As a result, the family of bisindole pyrroles spiroindimicins was extended with two new members, spiroindimicins E and F. Furthermore, prediction of the biosynthetic pathway for unusual α-pyrone lagunapyrone isolated from MP131-18 resulted in foresight and identification of two new compounds of this family – lagunapyrones D and E. The diversity of identified and predicted compounds from *Streptomyces* sp. MP131-18 demonstrates that marine-derived actinomycetes are not only a promising source of new natural products, but also represent a valuable pool of genes for combinatorial biosynthesis of secondary metabolites.

The discovery of new antibiotics remains one of the most important tasks of modern biotechnology due to the rapid emergence of antibiotic resistance among pathogenic bacteria[1]. The latter leads to an increasing number of untreatable or poorly treatable bacterial infections, which can potentially become one of the leading causes of mortality[2]. This makes the search for new antimicrobial compounds of vital importance for modern medicine.

Two thirds of all antibiotics originate from biological sources or are the semi-synthetic derivatives of biologically produced natural compounds[3]. Actinomycete bacteria, in particular those of the genus *Streptomyces*, are one of the most promising biological sources of new natural products, and will remain so at least for the near future[4]. Due to the diverse secondary metabolism these bacteria accumulate a large number of compounds, which may become drug leads for the development of antibacterials, antivirals, immunosuppressants, antifungals, insecticides, and antitumorals. The intensive exploration of terrestrial actinomycetes in 1950–70 s has led to frequent re-discovery of bioactive compounds, thus drawing interest to new ecological niches, which may become sources of new actinomycetes[5]. Given that oceans cover more than 70% of the Earth's surface and host approximately 87% of global biodiversity, they appear largely underexplored in terms of discovery of new microorganisms, including actinomycetes. Taking into consideration unprecedented diversity of marine organisms and comparatively little work done so far, the reported identification of more than 20.000 new marine natural products is astounding[6].

[1]Helmholtz-Institute for Pharmaceutical Research Saarland, Actinobacteria Metabolic Engineering Group, Saarbrücken, Germany. [2]Gause Institute of New Antibiotics, Russian Academy of Medical Sciences, Moscow, Russia. [3]Department of Biotechnology, Norwegian University of Science and Technology, Trondheim, Norway. [4]Department of Pharmacognosy, University of Vienna, Vienna, Austria. [5]Center for Biotechnology, Bielefeld University, Bielefeld, Germany. [6]Department of Pharmacy - Center for Drug Research, University of Munich, Munich, Germany. [7]Universität des Saarlandes, Pharmaceutical Biotechnology, Saarbrücken, Germany. [†]Present address: Department of Biology, Massachusetts Institute of Technology, Cambridge, MA, USA. Correspondence and requests for materials should be addressed to A.L. (email: a.luzhetskyy@mx.uni-saarland.de)

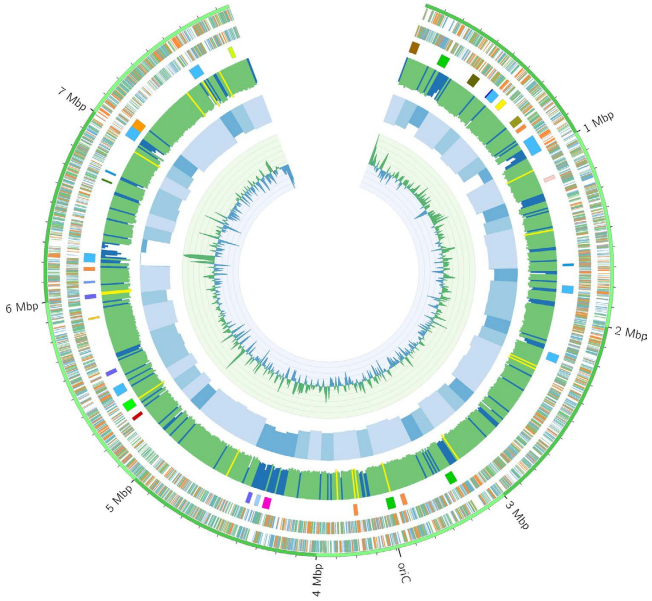


Figure 1. Schematic representation of the *Streptomyces* sp. MP131-18 genome (scaffold 1 only), created with the help of Circos[58]. Megabases are labeled; smaller ticks correspond to 100 kbp segments. From outside: genes on the forward and the reverse strands (blue: shorter than 900 bp, green: between 900 and 1500 bp long, orange: longer than 1500 bp); 35 secondary metabolite clusters colored by predicted type; G+C content, 10 kbp window (blue color highlights segments with G+C content <69%, yellow highlights correspond to G+C content over 76%); G+C content, 100 kbp window (lighter blue is higher G+C, darker blue is lower G+C content); G+C skew (green: positive; blue: negative).

Many of these compounds are produced by marine actinomycetes isolated from the deep sea sediments, coral reefs, marine invertebrates, and plants[7,8]. The taxonomic diversity of these actinomycetes ranges from typical representatives of *Streptomyces* to more exotic and rare *Dietzia*, *Rhodococcus, Salinispora, Marinophilus, Solwaraspora, Salinibacterium, Aeromicrobium, Williamsia* and *Verrucosispora* species, increasing the chances for the discovery of new bioactive natural products[8]. Several compounds produced by marine actinomycetes have a strong potential to be developed into pharmaceutical drugs. Diazepinomicin, a dibenzodiazepine alkaloid from a marine *Micromonospora* strain, possessed antibacterial and antitumor activities, and had been in phase II clinical trials for the treatment of glioblastoma[9,10]. Salinosporamide A, a β-lactone-ɣ-lactam from *Salinispora tropica*, entered phase I clinical trials as a drug for treatment of multiple myeloma just three years after its discovery[11].

The post-genomic era in actinomycete research is outlined by the discovery of multiple secondary metabolite biosynthesis gene clusters in the strains thought to be producing only few compounds[12,13]. This finding led to a rapid development of genomics-based approaches for the discovery of new natural products, resulting in isolation of new secondary metabolites from well-studied strains[14,15]. The application of genomics-based approaches has also broadened the diversity of natural products discovered from the marine actinomycetes[16,17].

Here, we report the genome sequencing of the marine-derived actinomycete MP131-18. The detailed phylogenetic classification identified the strain as *Streptomyces*. The secondary metabolite biosynthesis gene clusters analysis and complete dereplication of secondary metabolites profile of the strain led to identification of several known and new natural products, their association with the corresponding gene clusters, and prediction of the biosynthetic pathways.

Results

General properties of the *Streptomyces* sp. MP131-18 genome. Strain MP131-18 was isolated from a deep-water marine sediment sample collected in the Trondheim fjord, Norway. The sediment suspension from which this isolate was obtained was treated with extremely high frequency radiation (EHF) that was shown to selectively promote growth of various rare actinomycete bacteria[18]. The MP131-18 isolate was recovered after plating of the diluted EHF-treated suspension on oatmeal agar. On this medium, the isolate had a cream-colored aerial mycelium, while its substrate mycelium and spores were brown with olive tinge. Its cell wall was shown to contain LL-diaminopimelic acid, which is characteristic for *Streptomyces* spp. Still, phenotypic characteristics of the isolate MP131-18 were not typical for streptomycetes, prompting phylogenetic analysis of its 16S rRNA gene. A 1421 bp PCR fragment obtained from the genomic DNA of MP131-18, representing the almost complete 16S rRNA gene, was sequenced and analysed using Ribosomal Database Project Classifier, which strongly suggested it belonging to the genus *Streptomyces*.

Sequencing of the *Streptomyces* sp. MP131-18 genome was performed using two Illumina MiSEQ libraries – short-insert (paired-end, PE) and long-insert (mate pair, MP). The genome was assembled in a total of 10 scaffolds. The chromosome of *Streptomyces* sp. MP131-18 appears to be linear with most of the genome (7,861,428 bp) in scaffold 1 (Fig. 1). The other 9 scaffolds cover in total 95 kbp of genomic information with the largest being

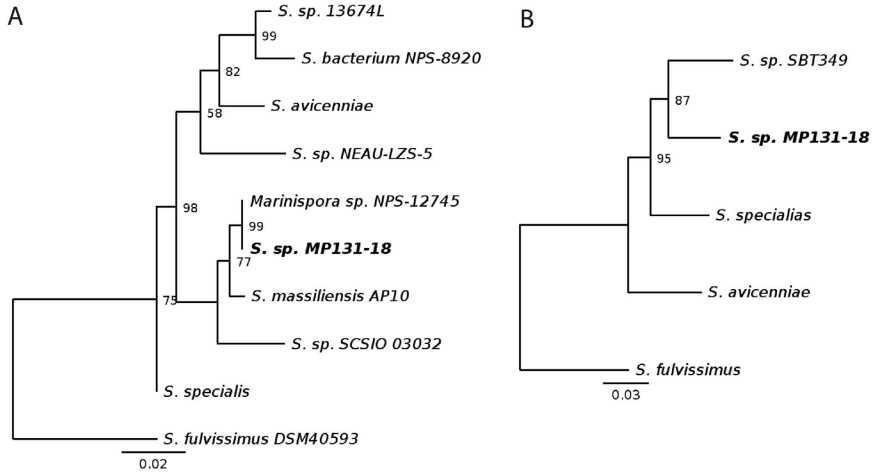


Figure 2. (**A**) Relation of *Streptomyces* sp. MP131-18 to the most 16S rRNA similar *Streptomyces*. *S. fulvissimus* 16S rRNA was used as outgroup. B. Relation of *Streptomyces* sp. MP131-18 to three other *Streptomyces*, as estimated by a 5-protein concatemer phylogenetic analysis. *S. fulvissimus* concatenated protein was used as outgroup.

48 kbp in size, and do not carry any housekeeping or secondary metabolism genes. Based on both PE and MP linking evidence (Figs 1S and 2S), the 9 extra scaffolds could represent unplaced parts of the scaffold 1 – that is, scaffold 1 is the complete bacterial chromosome, where gaps could be filled by the other 9 scaffolds. No plasmids were identified in *Streptomyces* sp. MP131-18. The G+C content (72.4%), the number of protein (7,054) and tRNA (78) encoding genes are comparable with other streptomycetes[12]. The genes for chromosomal replication initiation factor, *dnaA* (SBA_03553), and β-subunit of the DNA polymerase III, *dnaN* (SBA_03552), are located in the central part of the scaffold 1. The *oriC* is located between the *dnaN* and *dnaA* genes and contains three DnaA boxes (TTGTGCACAGG) conserved in *Streptomyces* species[19]. Both ends of scaffold 1 contain telomere-like ~100 bp inverted repeats with 4 possible hairpin structures.

The analysis of the genome sequence using the secondary metabolites biosynthesis genes search tool antiSMASH[20] revealed the presence of 36 putative gene clusters (Fig. 1). In comparison, the genomes of *S. coelicolor*[12], *S. averitilis*[13], and *S. albus* J1074[21] contain 20, 25 and 22 secondary metabolism gene clusters, respectively. The location of many of the secondary metabolism gene clusters in the *Streptomyces* sp. MP131-18 chromosome coincides with regions of low G+C content (Fig. 1).

Phylogenetic analysis. To taxonomically delineate *Streptomyces* sp. MP131-18, we performed phylogenetic analysis by two complementary approaches using 16S ribosomal RNA sequences (Table 1S) and 5 genes coding for housekeeping proteins (RpoB, DnaK1, RecA, SsgB, and SsgA; Table 2S)[22,23]. A dendrogram based on the 16S rDNA gene (Fig. 2A) clearly shows that *Streptomyces* sp. MP131-18 is closely related to the unclassified actinobacterium NPS-12745 (GenBank accession EF551062.1), previously proposed to be a member of the new genus *Marinispora*[24]. There is only a single ambiguity in the aligned part of both 16S rRNAs: where *Streptomyces* sp. MP131-18 has a cytosine (1516 bp), NPS-12745 has an IUPAC ambiguity code for pyrimidine. The second closest hit that showed similarity to 16S rRNA from both *Streptomyces* sp. MP131-18 and *Streptomyces* sp. NPS-12745 is strain *Streptomyces massiliensis*, isolated from human gut[25].

To perform an alternative protein-based phylogenetic analysis, the amino acid sequences of five highly conserved proteins (RpoB, DnaK1, RecA, SsgB, and SsgA; Table 2S)[22,23] from MP131-18 were used for searches (via BlastP) against the NCBI NR database to identify the most similar proteins from other species. Three *Streptomyces* species were occurring the most often in search results for DnaK1, RecA, RpoB and SsgB: marine actinomycete *Streptomyces sp*. SBT349[26], as well as the soil isolates *S. specialis*[27], and *S. avicenniae*, both found in the rhizosphere of the mangrove plant[28]. SsgA search results had protein identities under 70% and thus were not considered. The five protein sequences from 4 genomes (MP131-18 and the three aforementioned species) and *S. fulvissimus* as an outgroup were concatenated, multiple-aligned using MAFFT, and used for RaxML dendrogram construction (Fig. 2B).

The phylogeny based on the five proteins suggests (and 16S rDNA based phylogeny supports) that *Streptomyces* sp. MP131-18 is most similar to *S. specialis* and *S. avicenniae*. Both protein- and 16S-based phylogenies identified representatives of the *Streptomyces* genus as neighbours of MP131-18. Thus, we conclude with a high degree of confidence that MP131-18 as well as NPS-12745 and the unclassified *Streptomycetaceae* bacterium NPS-8920 belong to the genus *Streptomyces*.

Functional gene annotation. Functional annotation of *Streptomyces* sp. MP131-18 proteins with the bactNOG subset of the eggNOG v4 database (using protein BLAST with an expectation value cut-off of 0.001)[29] resulted in identification of possible orthologues with some biological function assigned for 4,533 (64.3%) out of 7,054 proteins (Table 3S), with some of the genes assigned to more than one category. Of the remainder, 986 CDSs (14%) had no hits against bactNOGs, and 1,535 proteins (21.8%) had hits against proteins without functional

category (*function unknown*). Among the proteins with functional assignment, 2,018 (28.6%) are implicated in metabolism, including 170 (2.4%) participating in secondary metabolism.

Analysis of secondary metabolism gene clusters. The antiSMASH3.0 analysis of *Streptomyces* sp. MP131-18 genome revealed 36 gene clusters predicted to be involved in the secondary metabolism of the strain, occupying 8.4% of the chromosome. However, the re-evaluation of antiSMASH3.0 results revealed that cluster 1 is most probably encoding two distinct pathways for terpene and polyketide biosynthesis (designated as cluster 1a and 1b respectively) (Table 4S).

Within the genome of *Streptomyces* sp. MP131-18, six gene clusters with type I polyketide synthase (PKS) genes were predicted, most of them coding for small mono- or bimodular PKSs. Cluster 35 is probably involved in production of a compound containing an α-pyridone ring, based on high similarity to PKS and post-PKS enzymes from the piericidin A1 biosynthesis gene cluster[30]. Other type I PKS gene clusters (9, 10 and 33) have no homologues in sequenced bacterial genomes available in the public databases, and their products could be just partially predicted from the genes' organization. Gene cluster 1b is coding for type II PKS most probably synthesizing compounds of the angucycline group. The presence of glycosyltransferase and sugar aminotransferase encoding genes in the 1b cluster suggests glycosylation of the produced polyketide. Gene cluster 7 is coding for the type III PKS that resembles the α-pyrone type PKS from fungi and plants[31]. A similar type III PKS is also found in the hybrid gene cluster 3, together with a monomodular type I PKS encoding gene. Cluster 19 is predicted to encode a biosynthetic pathway for a polyunsaturated aryl polyene-like compound. Only three non-ribosomal peptide synthase (NRPS) gene clusters were found within the genome of *Streptomyces* sp. MP131-18 (2, 15, 17). Cluster 2 is similar to the coelibactin gene cluster from *S. coelicolor*[12]. The structure of this compound is not known, but it is thought to act as zincophore. Four gene clusters - 13, 14, 30, and 34 - combine type I PKS and NRPS-encoding genes. Clusters 13, 14, and 30 include genes coding for discrete adenylation domain proteins that most probably are supplying the amino acid starter units for the PKS components.

Five gene clusters (1a, 8, 16, 18, 29) within the genome of *Streptomyces* sp. MP131-18 are devoted to terpenes biosynthesis, including cluster 18 for the sesquiterpene geosmin. Cluster 23 is coding for a type I PKS and a terpene synthase/cyclase. It also contains a glycosyltransferase suggesting that the putative hybrid product might be glycosylated.

Another large group of secondary metabolism gene clusters are coding for ribosomally synthesized and post-translationally modified peptides (RiPPs)[32]. Clusters 6, 21, 25, and 27 are coding for lantipeptide type compounds. Cluster 6, beside the *lanM* homologue, also harbours a gene for thiazole ring formation. Cluster 20 is similar to a recently discovered actinobacterial gene cluster for lasso-peptide biosynthesis[33].

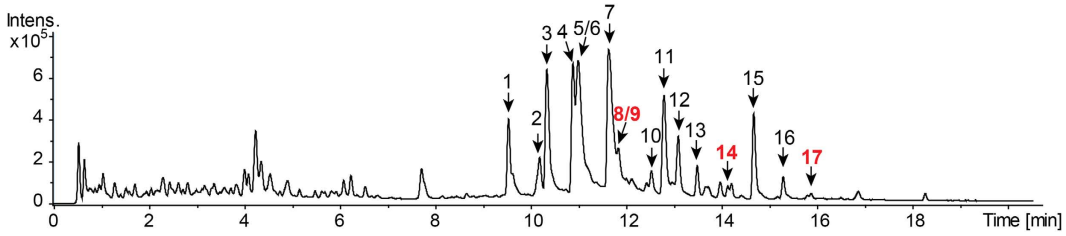
Several other secondary metabolites could be potentially produced by *Streptomyces* sp. MP131-18. Among them, two siderophore molecules encoded by clusters 12 and 32 (cluster 32 showed high similarity to desferrioxamine B biosynthesis genes), melanine (cluster 26), ectoine (cluster 31), bacteriocin (cluster 28), and A-factor like butyrolactone (cluster 4). Cluster 11 is presumably governing biosynthesis of a pyrrolopyrimidine nucleoside antibiotic. Two other gene clusters, 22 and 36, are predicted to be responsible for phenazine and indolocarbazole type compounds biosynthesis, respectively.

***Streptomyces* sp. MP131-18 produces a group of bisindole pyrrole antibiotics.** Based on the genome analysis, *Streptomyces* sp. MP131-18 appeared to be a strain with prolific potential for production of various, potentially new, secondary metabolites. In order to assess this, MP131-18 was grown in two different production media, and extracted metabolites were found to be active against *Bacillus subtilis* but not *Pseudomonas putida*. The extracts were further analysed by high-resolution LC-QTOF mass spectrometry and compounds were dereplicated using the Dictionary of Natural Products (DNP) database[34].

The dominant metabolites produced by *Streptomyces* sp. MP131-18 were two groups of closely related bisindole pyrrole antibiotics, lynamicins and spiroindimicins (Fig. 3). Based on the exact masses and absorption spectra, we were able to identify lynamicins A-G (peaks 2-4, 6, 7, 10, 12) and spiroindimicins B and C (peaks 11, 5) (Figs 3 and 4). These compounds were previously isolated from the marine actinomycete *Streptomyces* sp. SCSIO 03032 and a strain designated as *Marinispora* NPS-12745[24,35,36]. Both lynamicins and spiroindimicins were shown to have antibacterial activity. Thus, the ability of the *Streptomyces* sp. MP131-18 extracts to inhibit growth of *B. subtilis* is most probably caused by accumulation of these compounds. Lynamicin E and spiroindimicin B were purified (4.8 mg and 2.9 mg from 5L of culture, respectively) and their structures confirmed by NMR. 1D ^{1}H, ^{13}C NMR spectra and 2D correlations of ^{1}H-^{1}H COSY, HSQC and HMBC NMR spectra are summarized in Supplementary Tables 5S and 6S. Moreover, the compound **1** was identified as lycogarubin C[37], also known as chromopyrrolic acid, the key intermediate in bisindole pyrroles biosynthesis (Figs 3 and 3S)[38].

The compound with RT of 11.84 min showed the absorption spectra typical for identified bisindole pyrroles, but the detected mass did not correspond to any known lynamicins or spiroindimicins (Fig. 3)[24,35,36]. During the HPLC, the peak was found to separate into two peaks (Fig. 4S), and both of these compounds **8** and **9** were purified and their structures elucidated by NMR.

The compounds **8** and **9** were obtained as white solids (yield: 0.3 mg and 0.6 mg from 5 L of culture, respectively). The HRESIMS of compound **9** gave a quasimolecular ion peak at *m/z* 404.1202 $[M+H]^+$ corresponding to a molecular formula of $C_{23}H_{18}ClN_3O_2$ (Fig. 3). The UV spectrum of **9** with maxima at 248 and 286 nm was similar to other spiroindimicins. Compound **9** displayed the same molecular ion peak at HRESIMS (*m/z* 404.1202 $[M+H^+]$), UV absorption spectra (λ_{max} 248 and 286 nm) and molecular formula $C_{23}H_{18}ClN_3O_2$ as **8,** but had slightly different retention time. The new compounds were named spiroindimicin E (**8**) and spiroindimicin F (**9**). The 35Da mass difference of both new compounds to spiroindimicin B let us assume that the disparity is caused by the lack of one chlorine atom. This was confirmed by the analysis of the isotopic pattern of mass peaks of



Peak number	Compounds	Retention time	m/z $[M+H]^+$/M	Calculated mass	UV [nm]
1	Lycogarubin C	9.54	414.1485/413.1407	413.1375	228, 268
2	**Lynamicin H**	10.19	390.1014/389.0936	-	224, 288
3	Lynamicin E	10.34	448.1077/447.0999	447.0985	244, 268, 274
4	Lynamicin A	10.86	424.0622/423.0544	423.0541	244, 276, 300
5	Spiroindimicin C	10.88	424.0622/423.0544	423.0541	244, 254, 300
6	Lynamicin D	10.99	482.0669/481.0591	481.05961	244, 290
7	Lynamicin B	11.36	458.0201/457.0123	457.0151	244, 306
8	**Spiroindimicin E**	11.84	404.1202/403.1124	-	248, 286
9	**Spiroindimicin F**	11.84	404.1202/403.1124	-	248, 286
10	Lynamicin C	12.35	433.9779/432.9701	432.9707	228, 294
11	Spiroindimicin B	12.79	438.0802/437.0724	437.0697	254, 320
12	Lynamicin G/F	13.08	438.0787/437.0709	437.0673	244, 290
13	Lagunapyrone A	13.48	513.3591/512.3513	512.3501	224, 300
14	**Lagunapyrone D**	14.12	527.3735/526.3657	-	224, 300
15	Lagunapyrone B	14.66	541.3909/540.3831	540.3814	238, 302
16	Lagunapyrone C	15.28	555.4050/554.3972	554.3971	224, 302
17	**Lagunapyrone E**	15.86	569.4193/568.4115	-	224, 302

Figure 3. DAD chromatogram of extracts from *Streptomyces* sp. MP131-18. The identified compounds are indicated by numbers and their features are listed in the table. Compounds identified during this work are highlighted in red.

spiroindimicins B, E and F (data not shown). Furthermore the same *m/z* at HRESIMS and UV spectra for the new compounds suggest that they are constitutional isomers (Figs 3 and 4).

2D NMR data of both spiroindimicins E and F showed high similarity to those of spiroindimicin B (Tables 7S and 8S). Careful analysis of protons and carbons in the aromatic region allowed us to assign the position of chlorine atoms. In fact, the ^{1}H NMR spectrum of compound **8** exhibited eight aromatic signals. Proton splitting pattern of four of them and their HMBC correlations suggested compound **8** lacks a chlorine atom at C-6" in comparison to spiroindimicin B (Table 6S). Furthermore, the remaining two meta coupled doublets revealed the chlorine atom is at C-6' in spiroindimicin E (Table 7S). This finding was further supported by the HMBC correlations. Sequential COSY correlations of protons H5'-H8' and HMBC correlations of spiroindimicin F suggested the indol ring comprising protons H5'-H8' is not substituted, while the remaining indol ring has chlorine substitution at C6" due to proton splitting patterns as well as HMBC correlations (Table 8S).

Another minor compound with RT of 10.19 min was obtained in scarce amounts and it exhibited a quasimolecular ion peak at m/z 390.1014 $[M+H]^+$. This compound is predicted to be monochlorinated lynamicin F-like bisindole pyrrole based on its HRESIMS and isotopic pattern (Fig. 3). However, further investigation is required to establish the exact structure of this metabolite, tentatively designated as lynamicin H.

Isolated compounds (lycogarubin C, lynamicin E, spiroindimicin B and spiroindimicin E) were tested against *Bacillus subtilis, Escherichia coli* and *Pseudomonas putida* (Table 9S). None of them were found to be active against Gram negative test cultures and only spiroindimicin B demonstrated moderate activity against *Bacillus subtilis.* The antibacterial properties of spiroindimicins, similar to lynamicins[24], seem to correlate with the degree of

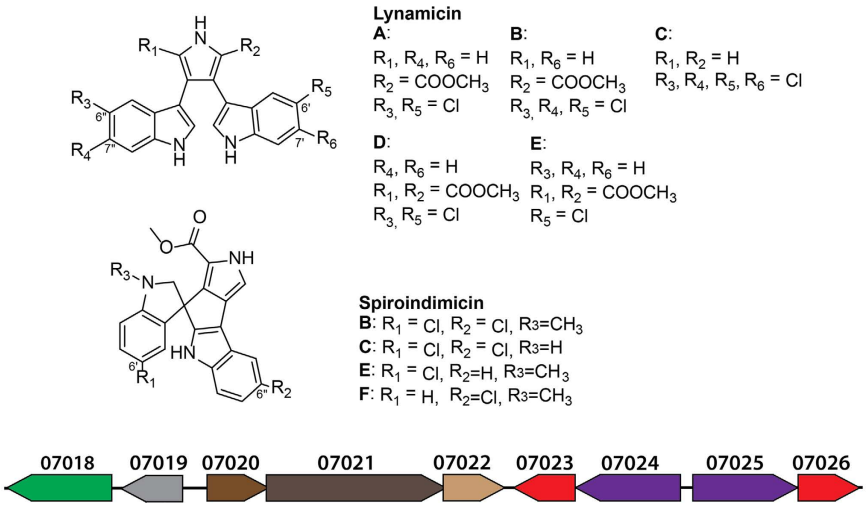


Figure 4. Structures of bisindole pyrrole compounds identified in the extract from *Streptomyces* sp. MP131-18 and schematic representation of gene cluster 36. Predicted genes functions: SBA_07018 – transcriptional regulator; SBA_07019 – putative transporter protein; SBA_07020 - tryptophan 2-monooxygenase; SBA_07021 – chromopyrrolic acid synthase, VioB homologue; SBA_07022 – RebP-like cytochrome P450; SBA_07023, SBA_07026 - flavin reductases; SBA_07024, SBA_07025 – halogenases.

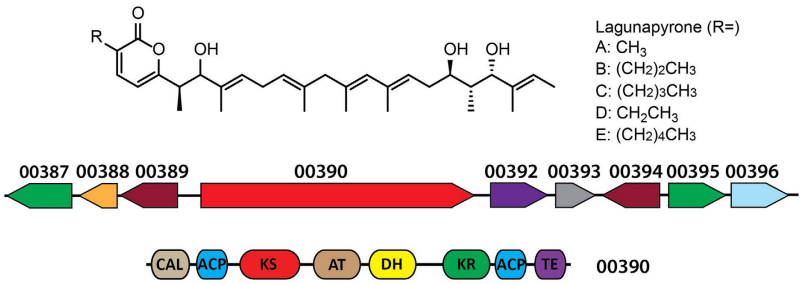


Figure 5. Structure of lagunapyrones A-E and schematic representation of gene cluster 3. The domain organization of putative lagunapyrone iPKS (SBA_00390) is predicted by antiSMASH 3.0[20].

chlorination. The monochlorinated lynamicins and spiroindimicins are weak inhibitors of bacterial growth, while antibacterial activity of both families of compounds is increasing with the increase of number of halogen atoms in their structures.

One of the new compounds, spiroindimicin E, as well as previously described lynamicin E and lycogarubin C were tested for cytotoxic activity (Fig. 5S). Spiroindimicin E shows a very slight effect on cellular growth of T24 bladder carcinoma cells, whereas treatment with 10 μM lynamicin E led to more than 50% reduction of proliferation capacity. At the same time, as expected, lycogarubin C almost completely inhibited proliferation of the tested cancer cell line at this concentration.

Among the secondary metabolite biosynthesis gene clusters, the only candidate identified for bisindole pyrroles biosynthesis is cluster 36, thereafter called *lyn* (Fig. 4, Table 4S). The core genes of the cluster, designated SBA_07020, SBA_07021 and SBA_07022, are coding for tryptophan 2-monooxygenase, chromopyrrolic acid synthase, and cytochrome P450, respectively. These genes are known to be involved in assembly of bisindole pyrrole core and probably its closing into spiroindimicins structures. Genes SBA_07024 and SBA_07025 are encoding two halogenases that presumably attach chlorine atoms at two different positions on the bisindole rings (6'/6" and 7'/7") (Fig. 4). The halogenase genes are forming operons with the genes SBA_07022 and SBA_07026 coding for flavin reductases participating in halogenation reaction. Lastly, SBA_07018 is encoding a transcriptional regulator of the NarP family, probably controlling the expression of the structural genes, and SBA_07019 is encoding a putative transporter protein probably involved in export of metabolites from the cell.

***Streptomyces* sp. MP131-18 produces unusual α-pyrones.** Beside the bisindole pyrroles, *Streptomyces* sp. MP131-18 was found to accumulate other compounds belonging to the α-pyrone group. Three lagunapyrones A-C were identified by exact mass and absorption spectra in the extracts from *Streptomyces* sp. MP131-18 culture (Figs 3 and 5). These compounds were originally isolated from the unclassified marine actinomycete CNB-984, and comprise a group of closely related α-pyrones functionalized by a highly methyl-branched C19 side chain[39]. Despite the structural similarity of α-pyrones, they can be assembled by PKS enzymes of either type I, II or III. The best candidate for lagunapyrones biosynthesis within the genome of *Streptomyces* sp.

MP131-18 is cluster 3, which contains genes coding for a single module type I PKS (SBA_00390) and α-pyrone type III PKS enzymes (SBA_00389). The possible implication of the type III PKS allowed us to predict the high flexibility of the choice of acyl-CoA esters that are utilized by the pathway. Based on this prediction, we were able to identify two new lagunapyrones D and E produced by *Streptomyces* sp. MP131-18. The newly identified metabolites have MS2 fragmentation patterns similar to the previously identified compounds A-C (Fig. 6S). The characteristic daughter ion peak at *m/z* 321.3 $[M+H]^+$ was observed for all lagunapyrones, which is ascribable to the side chain cleaved between C6 and C7 after loss of three water molecules. Lagunapyrones D and E differ from the described A and C congeners, respectively, by 14Da, most probably indicating a differences in the alkyl chain length at C2 position of α-pyrone ring (Fig. 5).

Discussion

The discovery of new antibiotics faced the stagnation phase when scientists realized that the same compounds are more and more often isolated from different actinomycete bacteria obtained from soil. Realizing this phenomenon caused a rapid change in the strategies used for natural product discovery by switching to new sources in the search of producing organisms. Undoubtedly, the oceans are the largest source of biodiversity and are relatively poorly studied. The growing information on the biosynthetic potential of marine microorganisms further invigorated the research in this area[6,8].

Here we report the use of a combined genomics-metabolomics approach in order to identify and isolate new compounds from a marine-derived actinomycete. Strain MP131-18 was isolated from a deep-water sediment sample and its genome was sequenced. The genome size, structure, G+C content and other features are typical for other streptomycetes genomes (Fig. 1). Both 16S rRNA and multilocus marker phylogenetic analyses clearly showed that the strain belongs to the genus *Streptomyces* (Fig. 2). The strain was found to be closely related to another marine actinomycete previously classified as *Marinispora* sp. NPS-12745[24]. Interestingly, several regions in the *Streptomyces* sp. MP131-18 genome have lower than average G+C content (Fig. 1). These regions comprise putative genomic islands, the recently restructured parts of the chromosome, most likely acquired by horizontal gene transfer[40]. The location of the secondary metabolism gene clusters within these putative genomic islands suggests that their origin is most likely other microorganisms, including *Proteobacteria* and fungi.

For a streptomycete, *Streptomyces* sp. MP131-18 possesses an average number of gene clusters dedicated to secondary metabolism: 36 clusters, occupying 8.4% of the genome. This number is slightly higher than in *S. coelicolor* A3(2) (20 gene clusters)[12] or *S. avermitilis* MA-4680 (25 gene clusters)[13] with 5% and 6.6% of the genes governing secondary metabolism biosynthesis, respectively, but still lower than in the case of *Kutzneria albida* with 14% of the chromosome devoted to secondary metabolism[41]. The *Streptomyces* sp. MP131-18 genome is encoding clusters for the synthesis of at least 18 different types of secondary metabolites, proving the high potential of marine actinomycetes to produce chemically diverse compounds. These include genes for a lassopeptide that is just a second example among actinobacteria[33] and aryl polyenes that were found so far only in *Proteobacteria*[42]. The latter compounds are thought to act as photoprotecting and antioxidant agents.

The gene cluster 36, due to similarity to rebeccamycin biosynthesis genes, was predicted to be involved in the biosynthesis of an indolocarbazole type compound. Indeed, two groups of highly halogenated bisindole pyrroles, lynamicins and spiroindimicins, were identified in the extracts of *Streptomyces* sp. MP131-18 (Fig. 3). Beside lynamicins A-G and spiroindimicins A-D, previously isolated from the marine actinomycetes *Streptomyces* sp. SCSIO 03032 and *Marinispora* NPS-12745[24,35,36], we have purified two new compounds designated spiroindimicin E and F (Figs 3 and 4). Tentatively, a new lynamicin H was also identified, but was not structurally characterized. Besides that, lycogarubin C can be found in the *Streptomyces* sp. MP131-18 extract[37]. This metabolite was originally discovered in the extract of the slime mould *Lacogala epidendrum*, and later as an intermediate in staurosporine and rebeccamycin biosynthesis in *S. longisporoflavus* and *Lechevalieria aerocolonigenes* A, respectively[43,44]. The presence of lycogarubin C in the extracts of *Streptomyces* sp. MP131-18 and its structural resemblance to the core of the identified antibiotics suggest that it might act as a common precursor or a shunt product in the biosynthesis of lynamicins and spiroindimicins. The fact that halogenation occurs before the assembly of the bisindole pyrrole ring system supports the shunt product hypothesis (Figs 3 and 7S)[45]. This indicates the high flexibility in substrate selection by the bisindole pyrrole assembly enzymes. Because of the structural diversity of lynamicins and spiroindimicins, the identified gene cluster might comprise a source of individual genes for the combinatorial biosynthesis of new bisindole pyrrole and indolocarbazole antibiotics, especially the ones encoding halogenation steps and formation of spirane structure.

A group of α-pyrones called lagunapyrones were also identified in the extract of MP131-18 (Fig. 5)[39]. These compounds comprise a long methyl-branched polyketide chain assembled into the final molecule by the linkage to the short fatty acid chain through the pyrone ring. The only gene cluster that could be involved in the biosynthesis of such compounds (cluster 3) contains type I and type III PKS genes. The type I PKS resembles the iterative highly reductive PKS enzymes from the fungal plant pathogen *Alternaria solani*, involved in alternapyrone biosynthesis, and is thought to synthesise the C20 polyketide chain (Fig. 6)[46]. Furthermore, the predicted lagunapyrone iPKS does not contain the enoylreductase domain. This activity is thought to be supplied by separate enzymes similar to those involved in the biosynthesis of the fungal metabolite lovastatin[47]. The acyl-transferase domains of PKSI enzymes are predicted to be specific for malonate extender units. The highly methyl-branched C19 polyketide side chain of lagunapyrones is assumed to be decorated as a post-PKS modification similar of the corresponding steps in alternapyrone or nafuredin biosynthesis[46,48]. However, the lagunapyrone PKSI is lacking the C-methyltransferase domain, while a gene coding for standalone C-methyltransferases is located in close proximity to the PKSIII gene. The type III PKS is presumably further extending the C20 polyketide chain assembled by PKSI with a single acetate unit. The same enzyme is thought to perform the intermolecular closing of α-pyrone ring with the short chain fatty acid-CoA esters[49]. Taking into account the flexibility of type III PKS enzyme in choosing the acyl-CoAs, we predicted the existence of at least two other lagunapyrones with C2 and

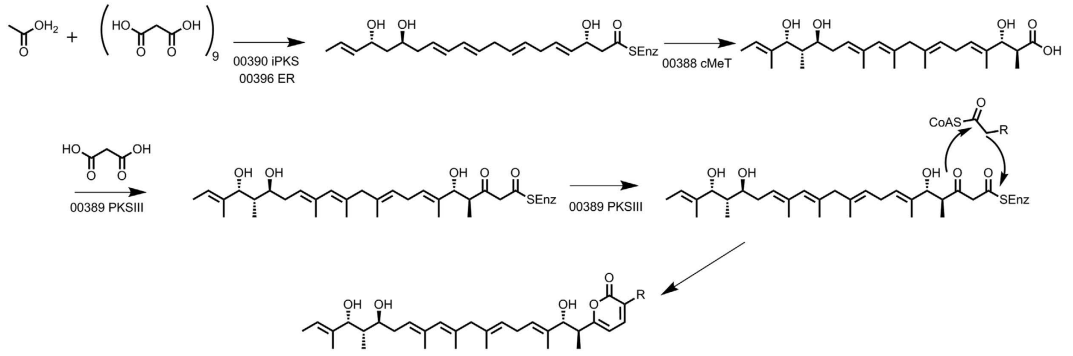


Figure 6. Proposed scheme for lagunapyrones A-D biosynthesis.

C5 alkyl chains. Indeed, the corresponding compounds, named as lagunapyrone B and E, were identified in the extract of *Streptomyces* sp. MP131-18 (Figs 3 and 5). This finding proves that the combination of detailed metabolomics and genomics is a powerful tool in finding new natural products. The prediction and understanding of biosynthetic processes leading to formation of particular metabolites helps to identify these compounds.

Methods

Bacterial strains, cultures conditions and general procedures. *Streptomyces* sp. MP131-18 was isolated from a marine sediment collected at the depth of 450 m at the Tautra ridge in the Trondheim fjord, Norway (63°55,909N, 010°61,846 E). The sediment sample was treated with extremely high frequency radiation and plated in dilutions on the oatmeal agar as described in[18]. *Streptomyces* strains were grown on mannitol soy flour agar (MS agar) and in liquid TSB medium[50]. NL19 (MS medium without agar)[50] and SG[51] mediums were used for secondary metabolite production.

Genome sequencing. For DNA isolation, *Streptomyces* sp. MP131-18 strain was inoculated into TSB medium and grown at 28 °C with shaking (200 rpm) for 3 days. High quality total cellular DNA was isolated using salting out procedure[50]. The purity and concentration of the genomic DNA was determined using a Nanodrop 2000 spectrophotometer (Thermo Fisher Scientific).

The obtained genomic DNA was sequenced using two MiSEQ libraries (Illumina) – short-insert (paired-end) and long-insert (mate pair). Assembly of the shotgun reads was performed with the Newbler v2.8 assembler (Roche). Genome annotation was performed using Prokka[52]. All predicted ORFs were manually re-inspected to correct start codon position. For the identification of secondary metabolites clusters antiSMASH 3.0 was used[20].

Nucleotide sequence accession number. The *Streptomyces* sp. MP131-18 Whole Genome Shotgun project has been deposited at DDBJ/ENA/GenBank under the accession LZNS00000000. The version described in this paper is version LZNS01000000.

Phylogenetic analysis. rRNA delineation was initially performed using the ARB-SILVA database[53], which suggested a number of neighbour (most similar) species. 16S rRNA sequences of these species were downloaded from ARB-SILVA (Table 1S), and multiple-aligned using MAFFT v7.222[54] (algorithm: auto, scoring matrix: 200PAM/k = 2, gap open penalty 1.53, offset value 0.123). The dendrogram was constructed using RaxML 7.2.8[55] (nucleotide model: GTR GAMMA, algorithm: rapid bootstrapping and search for best-scoring ML tree, bootstrap replicates: 1000). The tree was formatted in Geneious 9.0.4[56].

MAFFT and RAxML were also used for 5-gene alignments, with the same options as above. Accession numbers of the used gene sequences are listed in Table 2S.

Strain cultivation and metabolites extraction. The pre-cultures were grown in 10 ml of TSB in 100 mL flasks with glass beads for 3 days at 28 °C. Subsequently, the main cultures (50 ml of SG or NL19 medium in 500 ml flasks) were inoculated with 4 ml of pre-culture and cultivated at 28 °C on a rotary shaker at 180 rpm for 7 days. Metabolites were extracted with acetone:methanol 1:1 from biomass and with ethyl acetate from cultural liquid, evaporated and dissolved in 250 μl of methanol. The LC-HRMS data were collected on a Dionex Ultimate 3000 RSLC system using a BEH C18, 100 × 2.1 mm, 1.7 μm d_p column (Waters, Germany). Separation of 1 μl sample was achieved by a linear gradient of solvent B (acetonitrile with 0.1% of formic acid) against solvent A (water with 0.1% of formic acid) at a flow rate of 600 μl/min and 45 °C. The gradient started by a 0.5 min isocratic step at 5% B then increased to 95% B over 18 min to end up with a 2 min step at 95% B before re-equilibration under the initial conditions. UV spectra were acquired by a DAD in the range of 200 to 600 nm. The mass-spec data was collected on a maXis 4 G hr-ToF ultrahigh resolution mass spectrometer (Bruker Daltonics, Germany) using the Apollo II ESI source. Mass spectra were acquired in centroid mode ranging from 200 to 2500 m/z at a 2 Hz scan rate. Extracts from biomass and supernatant were analysed separately.

Data were collected and analysed with the Bruker Compass Data Analysis software, version 4.1 (Bruker, Billerica, USA). The screening for known compounds was performed using the Dictionary of Natural Products

database version 6.1 (CRC Press, Boca Raton, USA), using the following parameters: accurate molecular mass (mass match tolerance 10 ppm), absorption spectra and source of compounds isolation[34].

Isolation and purification of compounds. *Streptomyces* sp. MP131-18 was grown at 28 °C for 3 days in 6 × 500 ml flasks containing 50 ml of TSB medium and pre-cultures were used to inoculate 100 × 500 ml flasks containing 50 ml of SG media. Cultures were incubated at 28 °C for 7 days on rotary shakers at 180 rpm. The biomass and cultural liquid were separated by centrifugation. Metabolites were extracted from supernatant with 3L of ethyl acetate and from biomass with 500 ml of a 1:1 mixture of acetone:methanol. Extracts were evaporated and resuspended in 20 ml of methanol. Extracts from biomass and supernatant were combined.

Extracts were fractionated in two steps: size-exclusion chromatography and preparative HPLC. The extracts were loaded on a 1 m long column packed with LH 20 Sephadex (Sigma-Aldrich) and eluted with methanol as solvent. The fractions were collected every 15 min with a flowrate of 1 ml per minute. The obtained fractions were evaporated and dissolved in 0.5 ml of MeOH.

Samples were further separated by preparative HPLC (Agilent 1100 Series, Agilent Technologies and Dionex UltiMate 3000, Thermoscientific) using NUCLEODUR® C18 HTec column (250 × 10 mm, 5 μm) (Macherey-Nagel) with a linear gradient of solvent B (acetonitrile with 0.1% of formic acid) against solvent A (water with 0.1% of formic acid) at a flow rate of 4.5 ml/min and 45 °C. New compounds were separated using a gradient starting from 30% and increasing to 70% of B over 30 min. UV spectra were recorded with DAD detector at 280 nm. Individual peaks were collected and analysed with a LC-MS amaZon system (Bruker, Daltonics) using a BEH C18, 50 × 2.1 mm, 1.7 μm d_p column (Waters, Germany) and a linear gradient 5-95% of B over 9 min.

NMR spectra were acquired on a Bruker Ascend 700 MHz NMR spectrometer equipped with a 5 mm TXI cryoprobe. As solvent, deuterated $MeOD_4$ and DMSO-d_6 were used and HSQC, HMBC, ^{1}H-^{1}H COSY spectra were recorded using standard pulse programs.

Antibacterial assay. Antimicrobial activities of the extracted metabolites were assayed by a disc diffusion method, loading 40 μl of each extract on 6 mm diameter paper discs. Test cultures of *Bacillus subtilis* ATCC 6633 and *Pseudomonas putida* KT 2440 were plated from the liquid cultures on solid LB medium, dried for 20 minutes prior applying the discs. Plates were incubated overnight at 37 °C. The zones of inhibition were measured manually with accuracy ±1 mm. The minimum inhibitory concentration (MIC) determination was conducted as described[57], except that LB medium was used to grow the test cultures. Solution of thiazolyl blue tetrazolium bromide (Sigma-Aldrich) was used for growth visualization.

Cell proliferation assays. T24 bladder carcinoma cells were stimulated for 72 h and stained with crystal violet solution (0.5% crystal violet (w/v), 20% methanol (v/v) in H_2O). Next, unbound dye was removed by rinsing with distilled water. Following air-drying, crystal violet was dissolved in ethanol/sodium-citrate solution (50% ethanol (v/v), 0.1M sodium-citrate) and the absorbance was measured at 540 nm at day 0 to obtain the initial amount of living cells and after 72 h in order to calculate the proliferation rate. Data are expressed as mean ± SEM and analyzed using two-way ANOVA followed by Bonferroni's multiple comparisons test. $^*p < 0.05$, $^{**}p < 0.001$, $^{***}p < 0.001$.

References

1. Thabit, A. K., Crandon, J. L. & Nicolau, D. P. Antimicrobial resistance: impact on clinical and economic outcomes and the need for new antimicrobials. *Expert Opinion on Pharmacotherapy* **16,** 159–177 (2015).
2. Solomon, S. L. & Oliver, K. B. Antibiotic resistance threats in the United States: Stepping back from the brink. *American Family Physician* **89,** 938–941 (2014).
3. Berdy, J. Thoughts and facts about antibiotics: Where we are now and where we are heading. *Journal of Antibiotics* **65,** 385–395 (2012).
4. Solecka, J., Zajko, J., Postek, M. & Rajnisz, A. Biologically active secondary metabolites from Actinomycetes. *Central European Journal of Biology* **7,** 373–390 (2012).
5. Monciardini, P., Iorio, M., Maffioli, S., Sosio, M. & Donadio, S. Discovering new bioactive molecules from microbial sources. *Microbial Biotechnology* **7,** 209–220 (2014).
6. Hu, G. P. *et al.* Statistical research on marine natural products based on data obtained between 1985 and 2008. *Marine Drugs* **9,** 514–525 (2011).
7. Valli, S. *et al.* Antimicrobial potential of Actinomycetes species isolated from marine environment. *Asian Pac J Trop Biomed* **2,** 469–473 (2012).
8. Valliappan, K., Sun, W. & Li, Z. Y. Marine actinobacteria associated with marine organisms and their potentials in producing pharmaceutical natural products. *Applied Microbiology and Biotechnology* **98,** 7365–7377 (2014).
9. Charan, R. D. *et al.* Diazepinomicin, a new antimicrobial alkaloid from a marine *Micromonospora* sp. *Journal of Natural Products* **67,** 1431–1433 (2004).
10. Mason, W. P. *et al.* A phase II study of the Ras-MAPK signaling pathway inhibitor TLN-4601 in patients with glioblastoma at first progression. *Journal of Neuro-Oncology* **107,** 343–349 (2012).
11. Gulder, T. A. & Moore, B. S. Salinosporamide natural products: Potent 20 S proteasome inhibitors as promising cancer chemotherapeutics. *Angew Chem Int Ed Engl* **49,** 9346–9367 (2010).
12. Bentley, S. D. *et al.* Complete genome sequence of the model actinomycete *Streptomyces coelicolor* A3(2). *Nature* **417,** 141–147 (2002).
13. Omura, S. *et al.* Genome sequence of an industrial microorganism *Streptomyces avermitilis*: Deducing the ability of producing secondary metabolites. *Proceedings of the National Academy of Sciences of the United States of America* **98,** 12215–12220 (2001).
14. Rebets, Y., Brotz, E., Tokovenko, B. & Luzhetskyy, A. Actinomycetes biosynthetic potential: how to bridge *in silico* and *in vivo*? *Journal of Industrial Microbiology & Biotechnology* **41,** 387–402 (2014).
15. Zerikly, M. & Challis, G. L. Strategies for the discovery of new natural products by genome mining. *Chembiochem* **10,** 625–633 (2009).
16. Zhao, X. Q. Genome-based studies of marine microorganisms to maximize the diversity of natural products discovery for medical treatments. *Evidence-Based Complementary and Alternative Medicine* (2011).

17. Udwary, D. W. *et al.* Genome sequencing reveals complex secondary metabolome in the marine actinomycete *Salinispora tropica*. *Proceedings of the National Academy of Sciences of the United States of America* **104,** 10376–10381 (2007).
18. Terekhova, L. *Selective isolation of rare actinomycetes* (ed. D. I. Kurtboke) 82–101 (Nambour, University of the Sunshine Coast, 2003).
19. Smuczyk-Krawczyszyn, A. *et al.* Cluster of DnaA boxes involved in regulation of *Streptomyces* chromosome replication: from *in silico* to *in vivo* studies. *Journal of Bacteriology* **188,** 6184–6194 (2006).
20. Weber, T. *et al.* antiSMASH 3.0-a comprehensive resource for the genome mining of biosynthetic gene clusters. *Nucleic Acids Research* **43,** 237–243 (2015).
21. Zaburannyi, N., Rabyk, M., Ostash, B., Fedorenko, V. & Luzhetskyy, A. Insights into naturally minimised *Streptomyces albus* J1074 genome. *Bmc Genomics* **15,** 10.1186/1471-2164-15-97 (2014).
22. Gao, B. L. & Gupta, R. S. Phylogenetic framework and molecular signatures for the main clades of the Phylum *Actinobacteria*. *Microbiology and Molecular Biology Reviews* **76,** 66–112 (2012).
23. Girard, G. *et al.* A novel taxonomic marker that discriminates between morphologically complex actinomycetes. *Open Biology* **3,** 10.1098/Rsob.130073 (2013).
24. McArthur, K. A. *et al.* Lynamicins A-E, chlorinated bisindole pyrrole antibiotics from a novel marine actinomycete. *Journal of Natural Products* **71,** 1732–1737 (2008).
25. Pfleiderer, A. *et al.* Culturomics identified 11 new bacterial species from a single anorexia nervosa stool sample. *European Journal of Clinical Microbiology & Infectious Diseases* **32,** 1471–1481 (2013).
26. Horn, H., Cheng, C., Edrada-Ebel, R., Hentschel, U. & Abdelmohsen, U. R. Draft genome sequences of three chemically rich actinomycetes isolated from Mediterranean sponges. *Marine Genomics* **24,** 285–287 (2015).
27. Kampfer, P. *et al.* Streptomyces specialis sp nov. *International Journal of Systematic and Evolutionary Microbiology* **58,** 2602–2606 (2008).
28. Xiao, J. *et al.* Streptomyces avicenniae sp nov., a novel actinomycete isolated from the rhizosphere of the mangrove plant Avicennia mariana. *International Journal of Systematic and Evolutionary Microbiology* **59,** 2624–2628 (2009).
29. Powell, S. *et al.* eggNOG v4.0: nested orthology inference across 3686 organisms. *Nucleic Acids Research* **42,** 231–239 (2014).
30. Liu, Q. *et al.* Elucidation of piericidin A1 biosynthetic locus revealed a thioesterase-dependent mechanism of alpha-pyridone ring formation. *Chemistry & Biology* **19,** 243–253 (2012).
31. Lin, Z. J. *et al.* A bacterial source for mollusk pyrone polyketides. *Chemistry & Biology* **20,** 73–81 (2013).
32. Ortega, M. A. & van der Donk, W. A. New Insights into the biosynthetic logic of ribosomally synthesized and post-translationally modified peptide natural products. *Cell Chem Biol* **23,** 31–44 (2016).
33. Elsayed, S. S. *et al.* Chaxapeptin, a lasso peptide from extremotolerant *Streptomyces leeuwenhoekii* strain C58 from the hyperarid Atacama desert. *J Org Chem* **80,** 10252–10260 (2015).
34. Whittle, M., Willett, P., Klaffke, W. & van Noort, P. Evaluation of similarity measures for searching the Dictionary of Natural Products database. *Journal of Chemical Information and Computer Sciences* **43,** 449–457 (2003).
35. Zhang, W. J. *et al.* Indimicins A-E, bisindole alkaloids from the deep-sea-derived *Streptomyces* sp. SCSIO 03032. *Journal of Natural Products* **77,** 1887–1892 (2014).
36. Zhang, W. J. *et al.* Spiroindimicins A-D: new bisindole alkaloids from a deep-sea-derived actinomycete. *Organic Letters* **14,** 3364–3367 (2012).
37. Frode, R. *et al.* Isolation and synthesis of 3,4-bis(Indol-3-Yl)pyrrole-2,5-dicarboxylic acid-derivatives from the slime-mold *Lycogala epidendrum*. *Tetrahedron Letters* **35,** 1689–1690 (1994).
38. Howard-Jones, A. R. & Walsh, C. T. Staurosporine and rebeccamycin aglycones are assembled by the oxidative action of StaP, StaC, and RebC on chromopyrrolic acid. *Journal of the American Chemical Society* **128,** 12289–12298 (2006).
39. Lindel, T., Jensen, P. R. & Fenical, W. Lagunapyrones A-C: Cytotoxic acetogenins of a new skeletal class from a marine sediment bacterium. *Tetrahedron Letters* **37,** 1327–1330 (1996).
40. Dobrindt, U., Hochhut, B., Hentschel, U. & Hacker, J. Genomic islands in pathogenic and environmental microorganisms. *Nature Reviews Microbiology* **2,** 414–424 (2004).
41. Rebets, Y. *et al.* Complete genome sequence of producer of the glycopeptide antibiotic Aculeximycin *Kutzneria albida* DSM 43870(T), a representative of minor genus of *Pseudonocardiaceae*. *Bmc Genomics* **15,** 10.1186/1471-2164-15-885 (2014).
42. Schoner, T. A. *et al.* Aryl polyenes, a highly abundant class of bacterial natural products, are functionally related to antioxidative carotenoids. *Chembiochem* **17,** 247–253 (2016).
43. Howard-Jones, A. R. & Walsh, C. T. Enzymatic generation of the chromopyrrolic acid scaffold of rebeccamycin by the tandem action of RebO and RebD. *Biochemistry* **44,** 15652–15663 (2005).
44. Nishizawa, T., Gruschow, S., Jayamaha, D. H. E., Nishizawa-Harada, C. & Sherman, D. H. Enzymatic assembly of the bis-indole core of rebeccamycin. *Journal of the American Chemical Society* **128,** 724–725 (2006).
45. Sanchez, C. *et al.* The biosynthetic gene cluster for the antitumor rebeccamycin: Characterization and generation of indolocarbazole derivatives. *Chemistry & Biology* **9,** 519–531 (2002).
46. Fujii, I., Yoshida, N., Shimomaki, S., Oikawa, H. & Ebizuka, Y. An iterative type I polyketide synthase PKSN catalyzes synthesis of the decaketide alternapyrone with regio-specific octa-methylation. *Chemistry & Biology* **12,** 1301–1309 (2005).
47. Ma, S. M. *et al.* Complete reconstitution of a highly reducing iterative polyketide synthase. *Science* **326,** 589–592 (2009).
48. Ui, H. *et al.* Nafuredin, a novel inhibitor of NADH-fumarate reductase, produced by Aspergillus niger FT-0554. *Journal of Antibiotics* **54,** 234–238 (2001).
49. Schäberle, T. F. Biosynthesis of α-pyrones. *Beilstein J. Org. Chem.* **12,** 571–588 (2016).
50. Kieser, T. *et al. Practical Streptomyces genetics*. (John Innes Foundation, 2000).
51. Rebets, Y. *et al.* Production of landomycins in *Streptomyces globisporus* 1912 and *S. cyanogenus* S136 is regulated by genes encoding putative transcriptional activators. *Fems Microbiology Letters* **222,** 149–153 (2003).
52. Seemann, T. Prokka: rapid prokaryotic genome annotation. *Bioinformatics* **30,** 2068–2069 (2014).
53. Quast, C. *et al.* The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Research* **41,** 590–596 (2013).
54. Katoh, K., Misawa, K., Kuma, K. & Miyata, T. MAFFT: A novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Research* **30,** 3059–3066 (2002).
55. Stamatakis, A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30,** 1312–1313 (2014).
56. Kearse, M. *et al.* Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* **28,** 1647–1649 (2012).
57. Andrews, J. M. Determination of minimum inhibitory concentrations. *Journal of Antimicrobial Chemotherapy* **48,** 5–16 (2001).
58. Krzywinski, M. *et al.* Circos: An information aesthetic for comparative genomics. *Genome Research* **19,** 1639–1645 (2009).

Acknowledgements
This work was supported through the ERC starting grant EXPLOGEN No. 281623 and from DZIF to AL.

Author Contributions
Y.R., S.Z., J.K. and A.L. designed the experiments, C.P., Y.R., B.T., S.N., S.B., M.M. and S.Z. performed the experiments, C.P. and S.N. developed the analytical tools, C.P., C.R., Y.R., B.T., S.Z. and A.L. analysed the data, and C.P., Y.R., B.T. and S.Z. wrote the manuscript, all authors reviewed the manuscript.

Additional Information
Supplementary information accompanies this paper at http://www.nature.com/srep

Competing financial interests: The authors declare no competing financial interests.

How to cite this article: Paulus, C. *et al.* New natural products identified by combined genomics-metabolomics profiling of marine *Streptomyces* sp. MP131-18. *Sci. Rep.* **7**, 42382; doi: 10.1038/srep42382 (2017).

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Supplementary information.

New natural products identified by combined genomics-metabolomics profiling of marine *Streptomyces* sp. MP131-18.

Constanze Paulus[1], Yuriy Rebets[1], Bogdan Tokovenko[1], Suvd Nadmid[1], Larisa P. Terekhova[2], Maksym Myronovskyi[1], Sergey B. Zotchev[3,4], Christian Rückert[5#], Simone Braig[6], Stefan Zahler[6], Jörn Kalinowski[5], Andriy Luzhetskyy[1,7]*

[1]Helmholtz-Institute for Pharmaceutical Research Saarland, Actinobacteria Metabolic Engineering Group, Saarbrücken, Germany

[2]Gause Institute of New Antibiotics, Russian Academy of Medical Sciences, Moscow, Russia

[3]Department of Biotechnology, Norwegian University of Science and Technology, Trondheim, Norway

[4]Department of Pharmacognosy, University of Vienna, Vienna, Austria

[5]Center for Biotechnology, Bielefeld University, Bielefeld, Germany

[6]Department of Pharmacy - Center for Drug Research, University of Munich, Munich, Germany

[7]Universität des Saarlandes, Pharmaceutical Biotechnology, Saarbrücken, Germany

#present address: Department of Biology, Massachusetts Institute of Technology, Cambridge, MA, USA

*Corresponding author:

Prof. Dr. Andriy Luzhetskyy

E-mail: a.luzhetskyy@mx.uni-saarland.de

Universität des Saarlandes, Pharmaceutical Biotechnology, Building C2.3, 66123 Saarbrücken, Germany.

Table 1S. 16S ribosomal DNA sequences used for 16S phylogenie.

Species	**ARB-SILVA accession / rRNA locus_tag**	**Identity with the SBA_07180, %**
Streptomyces sp. MP131-18	SBA_07180	-
Streptomyces fulvissimus DSM40593	SFUL_2616	92.53
Streptomyces sp. 13674L	EU741227.1.1494	96.67
Streptomycetaceae bacterium NPS-8920	EF470589.1.1523	96.43
Streptomyces avicenniae	EU399234.1.1491	96.4
Streptomyces sp. NEAU-LZS-5	KC304791.1.1538	96.09
Marinispora sp. NPS-12745	EF551062.1.1533	99.97
Streptomyces massiliensis AP10	JX101691.1.1502	99
Streptomyces specialis	AM934703.1.1500	97.35
Streptomyces. sp. *SCSIO 03032*	JN798514.1	93.88

For *S. fulvissimus* and *S.* sp. MP131-18, locus_tags are provided; for *Streptomyces.* sp. *SCSIO 03032* GenBank accession number; other IDs are from ARB-SILVA database.

Table 2S. Table of compared gene sequences (locus_tags).

Protein	***S.* sp. MP131-18**	***S. fulvissimus***	***S. specialis***	***S.* sp. SBT349**	***S. avicenniae***
RpoB	SBA_02824	SFUL_RS21975	BN2279_RS23020	ADY01_RS15195	IF27_RS10465
DnaK1	SBA_03307	SFUL_RS16290	BN2279_RS14960	ADY01_RS04030	IF27_RS03445
RecA	SBA_01513	SFUL_RS28035	BN2279_RS19870	ADY01_RS00075	IF27_RS21960
SsgB	SBA_03814	SFUL_RS15590	BN2279_RS17305	ADY01_RS23410	IF27_RS03120
SsgA	SBA_04743	SFUL_RS12465	BN2279_RS22425	ADY01_RS28310	IF27_RS06350

Table 3S. Functional annotation of *Streptomyces* sp. MP131-18 genes.

INFORMATION STORAGE AND PROCESSING	**1068**
Translation, ribosomal structure and biogenesis	170
RNA processing and modification	1
Transcription	677
Replication, recombination and repair	219
Chromatin structure and dynamics	1
CELLULAR PROCESSES AND SIGNALING	**846**
Cell cycle control, cell division, chromosome partitioning	38
Defense mechanisms	141
Signal transduction mechanisms	303
Cell wall/membrane/envelope biogenesis	204
Intracellular trafficking, secretion, and vesicular transport	28
Posttranslational modification, protein turnover, chaperones	132
METABOLISM	**2,018**
Energy production and conversion	315
Carbohydrate transport and metabolism	483
Amino acid transport and metabolism	355
Nucleotide transport and metabolism	95
Coenzyme transport and metabolism	151
Lipid transport and metabolism	162
Inorganic ion transport and metabolism	287
Secondary metabolites biosynthesis, transport and catabolism	170
POORLY CHARACTERIZED	**2,227**
General function prediction only	692
Function unknown	1,535

Table 4S. Secondary metabolism gene clusters classified by the type.

Cluster type	**Number of clusters**	**Clusters ID**
PKSI	6	5, 9, 10, 24, 33, 35
PKSII	1	1b
PKSIII	1	7
PKSI-PKSIII	1	3
NRPS	3	2, 15, 17
PKS-NRPS	4	13, 14, 30, 34
Terpene	5	1a, 8, 16, 18, 29
PKSI-Terpene	1	23
RiPPs	5	6, 20, 21, 25, 27
Nuceloside	1	11
Butyrolactone	1	4
Siderophore	2	12, 32
Arylpolyene (PKS)	1	19
Phenazine	1	22
Melanin	1	26
Bacteriocin	1	28
Ectoine	1	31
Indole	1	36

Table 5S. NMR data of Lynamicin E.

position	δ_H (J in Hz)	δ_C	Cosy	HMBC
1	NH	-	-	-
2	-	122.9	-	-
3				
4				
5	-	122.8	-	-
6	-	161.3	-	-
7	3.69[a] s	50.6		
8	3.66[a] s	50.4	-	-
9	-	161.7	-	-
1'	NH	-	-	-
2'	6.90 s	126.3	-	3'
3'	-	108.0	-	-
4'	-	128.8		-
5'	7.02 d (1.7)	118.6	-	9'-6'-7'
6'	-	124.4	-	-
7'	6.92 dd (2.0, 8.6)	120.5	7.18	5'
8'	7.18 dd (0.4, 8.6)	111.7	6.92	4'-6'
9'	-	134.2	-	-
1''	NH	-	-	-
2''	6.87 s	124.4	-	3''-4''-9''
3''	-	107.9	-	-
4''	-	127.7	-	-
5''	7.07 d (7.9)	119.5	6.80	3''-9''
6''	6.80 t (0.7, 7.6)	118.3	6.97-7.07-7.25	8''-4''
7''	6.97 ddd (1.0, 7.7)	118.9	6.80-7.25	5''-9''
8''	7.25 d (8.1)	110.6	6.97	6''-4''
9''	-	136.1	-	-

Sample was measured in MeOH. Data were recorded on Bruker Ascend 700 MHz. Spectra referenced to internal solvent for MeOH at 3.31 ppm (^{1}H) and 49.0 ppm (^{13}C). [a] Chemical shifts may be interchangeable.

Table 6S. NMR data of Spiroindimicin B.

position	δ_H (J in Hz)	δ_C	Cosy	HMBC
1	NH	-	-	-
2	6.98 s	111.7	-	5-3
3	-	127.9	-	-
4	-	142.6	-	-
5	-	117.4	-	-
6	-	162.2	-	-
7	3.59 s	51.1	-	5-6
1'	-	-	-	-
2'	4.09 d (8.8)	65.2	-	10'-3'-4'-4-2''-9'
	3.67 d (8.8)			
3'	-	53.1	-	-
4'	-	135.8	-	-
5'	6.34 d (2.1)	123.3	7.05	3'-7'-9'
6'	-	124.1	-	-
7'	7.04 (2.1, 8.5)	128.6	6.34-6.64	5'-9'
8'	6.64 d (8.5)	109.4	7.05	5'-4'
9'	-	153.8	-	-
10'	2.95 s	36.3	-	2'-9'
1''	NH	-	-	-
2''	-	156.6	-	-
3''	-	112.8	-	-
4''	-	122.9	-	-
5''	7.58 d (2.0)	119.2	6.99	3''-7''-4''
6''	-	126.1	-	5''-7''-6''-9''
7''	6.99 dd (2.0, 8.6)	121.5	7.25-7.58	5''-4''-8''
8''	7.25 d (8.6)	114.1	6.99	4''-6''
9''	-	139.9	-	-

Sample was measured in MeOH. Data were recorded on Bruker Ascend 700 MHz. Spectra referenced to internal solvent for MeOH at 3.31 ppm (^{1}H) and 49.0 ppm (^{13}C).

Table 7S. NMR data of Spiroindimicin E.

position	δ_H (J in Hz)	δ_C	HMBC
1	NH	-	-
2	6.95 s	111.4	5-4
3	-	125.7	-
4	-	142.9	-
5	-	117.0	-
6	-	162.3	-
7	3.58 s	52.1	6
1'	-	-	-
2'	4.09 d (8.8)	65.0	10'-7-4-9'
	3.69 d (8.8)		
3'	-	52.8	-
4'	-	136.4	-
5'	6.34 d (2.0)	123.7	3'-7'-9'
6'	-	123.1	-
7'	7.044 (6.06)	128.1	5'-9'
8'	6.64 d (8.5)	109.7	5'-4'
9'	-	153.7	-
10'	2.95 s	36.2	2'-9'
1''	NH	-	-
2''	-	152.3	-
3''	-	118.4	-
4''	-	122.0	-
5''	7.28 d (0.7, 7.3)	113.5	7''-4''
6''	7.03 (1.72)	121.0	5''-7''-4''-9''-8''
7''	7.045 (0.81)	120.5	5''-4''-8''
8''	7.58 (0.7, 7.4)	119.7	5''-4''-9''
9''	-	141.7	-

Sample was measured in MeOH. Data were recorded on Bruker Ascend 700 MHz. Spectra referenced to internal solvent for MeOH at 3.31 ppm (^{1}H) and 49.0 ppm (^{13}C).

Table 8S. NMR data of Spiroindimicin F

position	δ_H (J in Hz)	δ_C	Cosy	HMBC
1	11.35 s	-	7.09	5-3-4
2	7.09 (2.7)	111.0	11.35	5-3-4
3	-	125.5	-	-
4	-	141.5	-	-
5	-	115.2	-	-
6	-	159.9	-	-
7	3.47 s	50.1	-	6
1'	-	-		-
2'	3.64/3.90 d	63.1	-	10'-3'-4'-4-9'-2''
3'	-	51.4	-	-
4'	-	132.2	-	-
5'	6.33 d (7.3)	121.6	6.44	7'-9'
6'	6.44 t (7.3)	117.1	6.33-7.06	8'-7'-4'
7'	7.06 t (1.2, 8.1)	127.6	6.44-6.71	8'-5'-4'-9'
8'	6.71 d (8.0)	107.2	7.06	6'-4'
9'	-	153.3	-	-
10'	2.9 s	35.7	-	2'-9'
1''	11.4 s	-	-	3''-4''-9''-2''
2''	-	155.7	-	-
3''	-	110.6	-	-
4''	-	120.8	-	-
5''	7.72 s	117.7	-	-
6''	-	123.1	6.99	4''-8''
7''	6.99 dd (2.04, 8.6)	119.7	7.25	5''-8''-9''
8''	7.25 (8.6)	123.5	6.99-11.4	3''-7''-8''-9''
9''	-	138.0	-	-

Sample was measured in MeOH. Data were recorded on Bruker Ascend 700 MHz. Spectra referenced to internal solvent for MeOH at 3.31 ppm (^{1}H) and 49.0 ppm (^{13}C).

Table 9S. Minimal inhibitory concentrations (µg/ml [µM]) of isolated compounds against bacterial test cultures.

Organism	lycogarubin C	lynamicin E	spiroindimicin B	spiroindimicin E
Bacillus subtilis	>100 [>240]	>100 [>223]	25 [57]	>100 [>247]
Escherichia coli	>100 [>240]	>100 [>223]	>100 [>228]	>100 [>247]
Pseudomonas putida	>100 [>240]	>100 [>223]	>100 [>228]	>100 [>247]

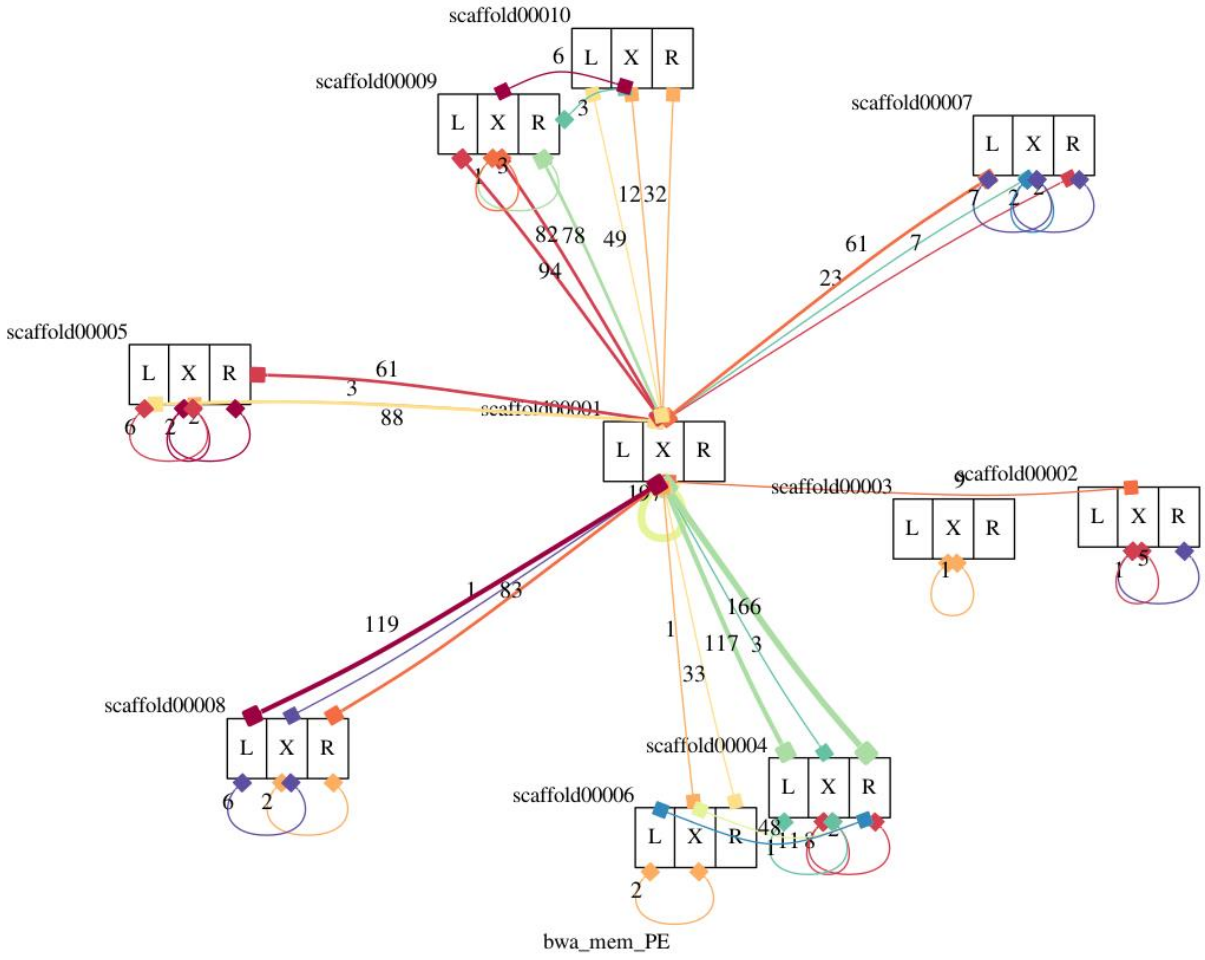


Figure 1S. Paired-end links between scaffolds: each link represents a read pair. L, left edge; R, right edge; X, neither left nor right edge (i.e. somewhere else in the scaffold). Mean PE insert size is 470 with standard deviation ~200; thus scaffold edge was estimated at 670bp. Loops depict self-links. Line weight is proportional (within fixed min/max limits) to the number of links, which is also shown next to each connecting line. Scaffold 1 (in the center) has no PE links from its L/R edges.

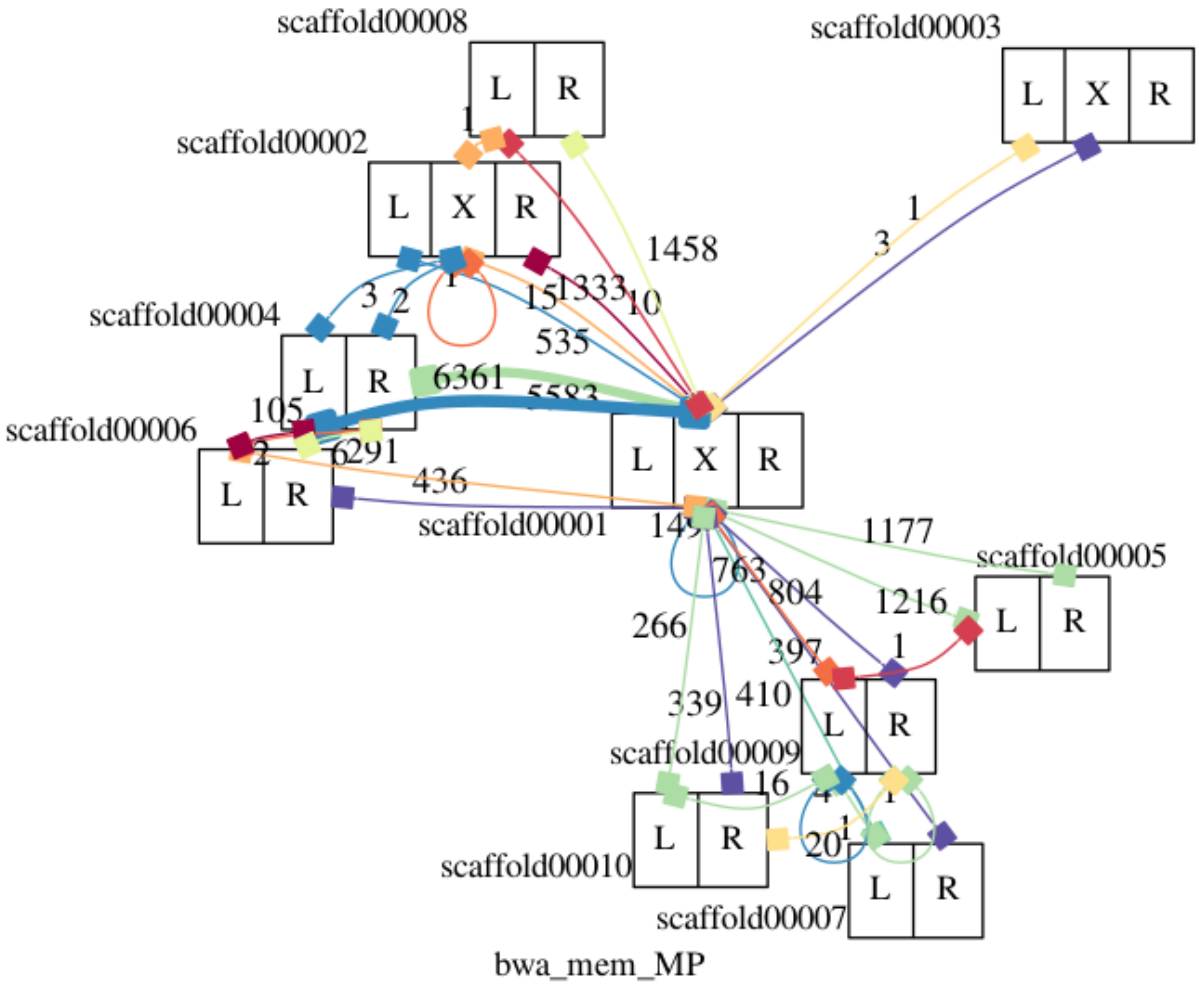


Figure 2S. Mate-pair links between scaffolds: each link represents a read pair. Mean MP insert size is 8100 with standard deviation ~1000; thus scaffold edge was estimated at 9100bp. Scaffold 1 (in the center) has no MP links from its L/R edges.

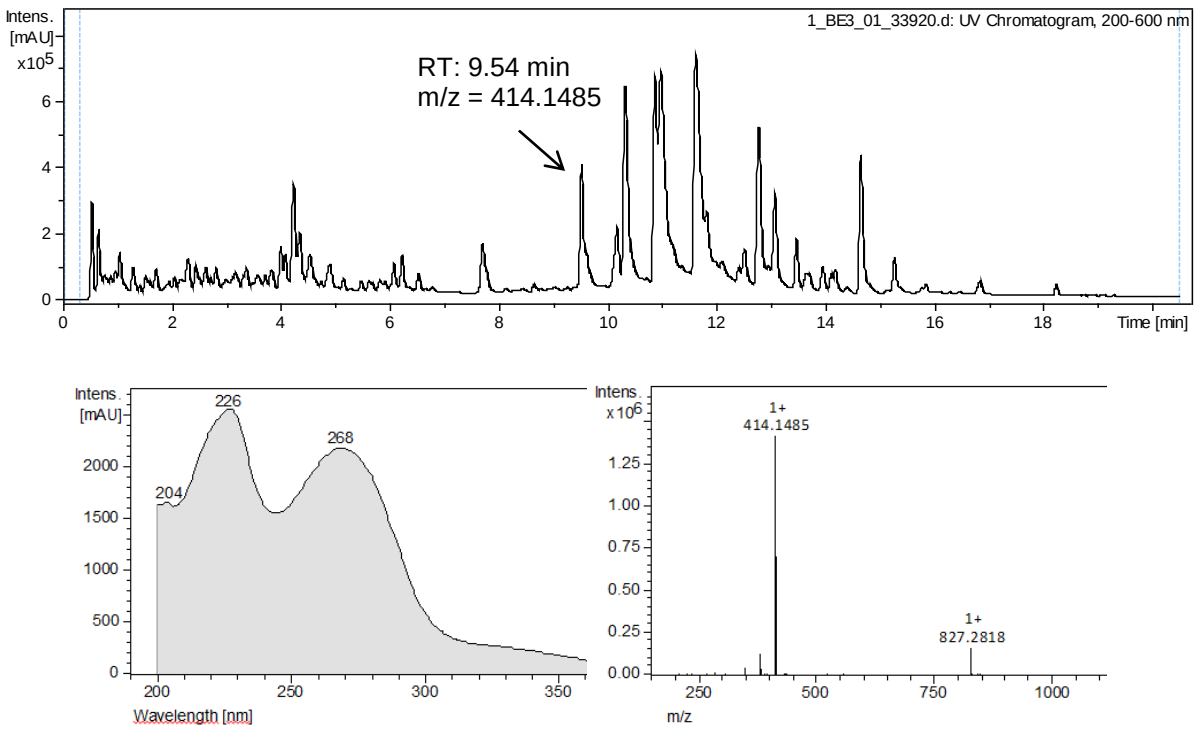


Figure 3S. MS chromatogram of lycogarubin C

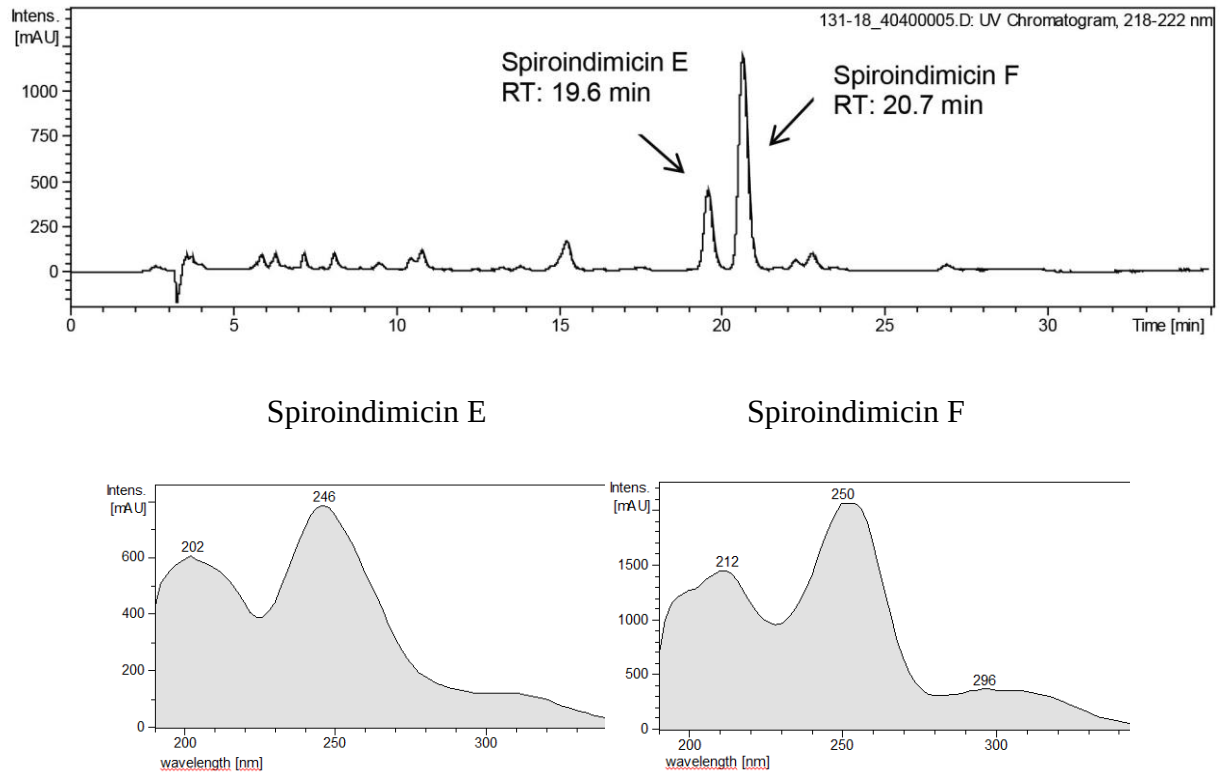


Figure 4S. Semipreparative HPLC chromatogram for the isolation of spiroindimicin E and F.

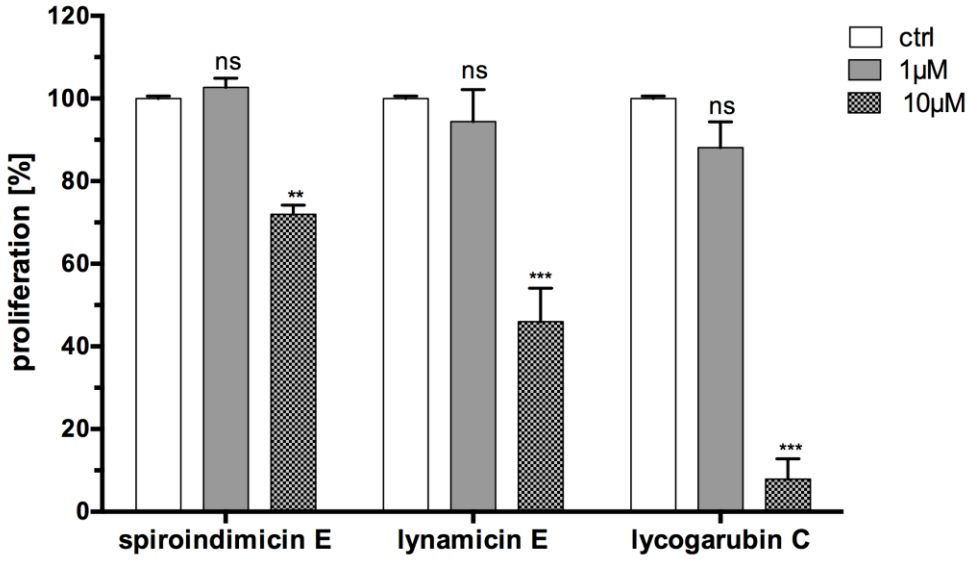
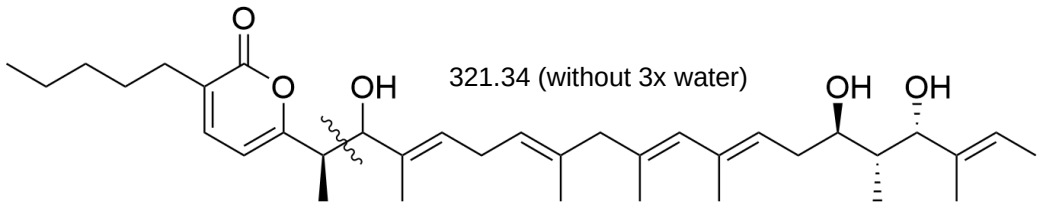
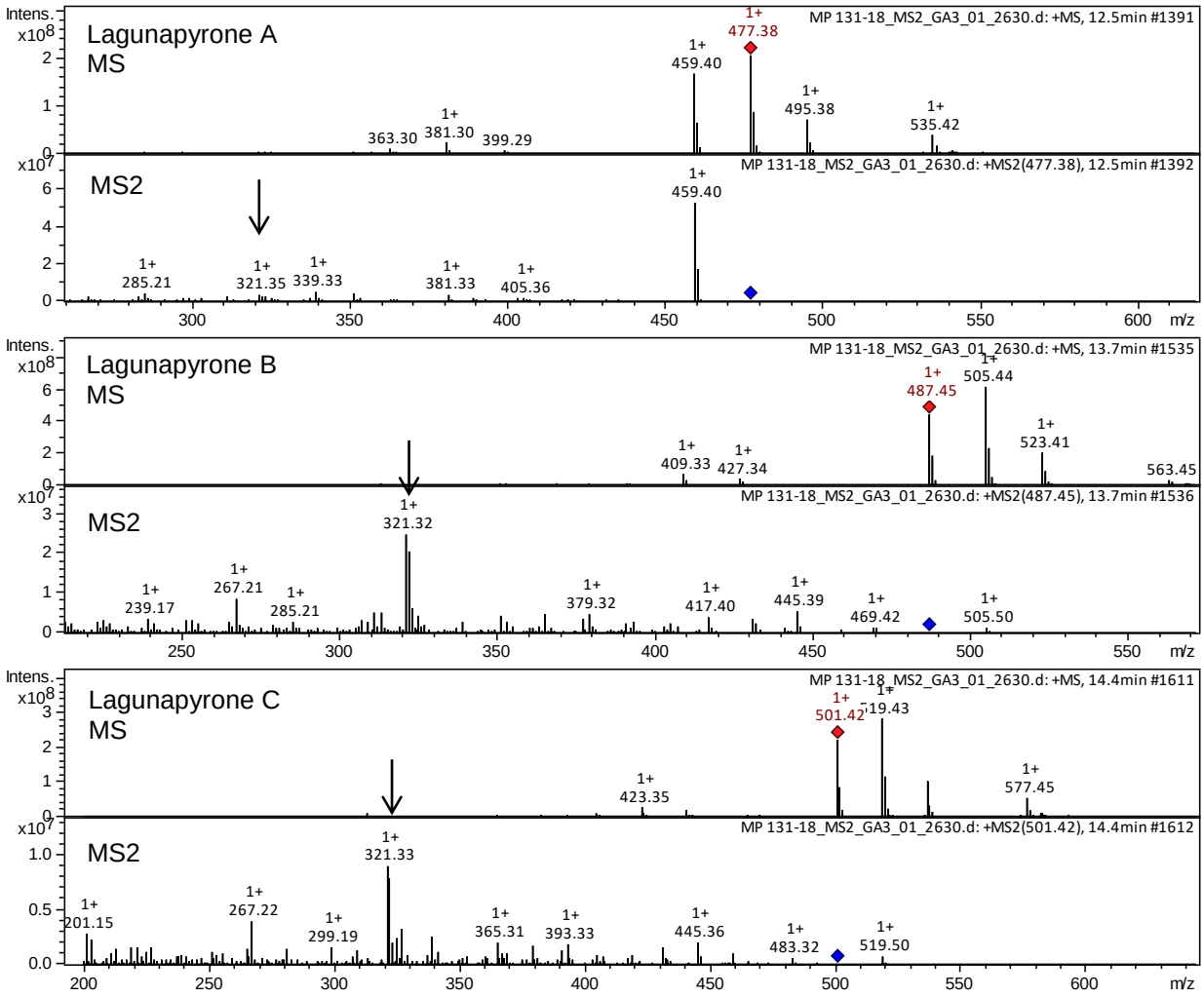


Figure 5S. Growth-inhibitory effects on T24 bladder carcinoma cells. Cancer cells were treated for 72h with increasing concentrations of the indicated substances and proliferation rate was analysed by staining with crystal violet.


O
O
OH
321.34 (without 3x water)
OH
OH


Intens.
Lagunapyrone A
MS
MP 131-18_MS2_GA3_01_2630.d: +MS, 12.5min #1391
1+ 477.38
1+ 459.40
1+ 495.38
1+ 535.42
1+ 381.30
363.30
399.29
MS2
MP 131-18_MS2_GA3_01_2630.d: +MS2(477.38), 12.5min #1392
1+ 459.40
1+ 285.21
1+ 321.35
1+ 339.33
1+ 381.33
1+ 405.36
m/z
Lagunapyrone B
MS
MP 131-18_MS2_GA3_01_2630.d: +MS, 13.7min #1535
1+ 487.45
1+ 505.44
1+ 523.41
1+ 409.33
1+ 427.34
563.45
MS2
MP 131-18_MS2_GA3_01_2630.d: +MS2(487.45), 13.7min #1536
1+ 321.32
1+ 267.21
1+ 239.17
1+ 285.21
1+ 379.32
1+ 417.40
1+ 445.39
1+ 469.42
1+ 505.50
Lagunapyrone C
MS
MP 131-18_MS2_GA3_01_2630.d: +MS, 14.4min #1611
1+ 501.42
1+ 519.43
1+ 577.45
1+ 423.35
MS2
MP 131-18_MS2_GA3_01_2630.d: +MS2(501.42), 14.4min #1612
1+ 321.33
1+ 267.22
1+ 201.15
1+ 299.19
1+ 365.31
1+ 393.33
1+ 445.36
1+ 483.32
1+ 519.50

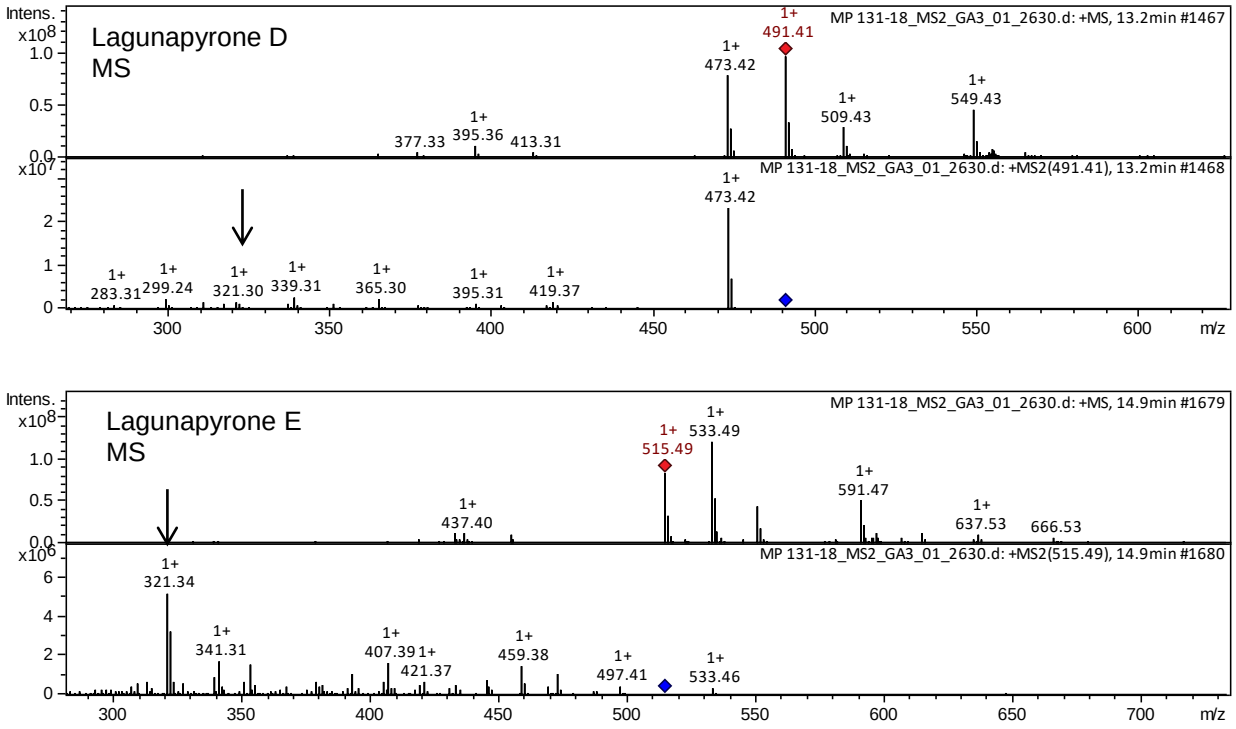


Figure 6S. MS2 for lagunapyrones (continuation).

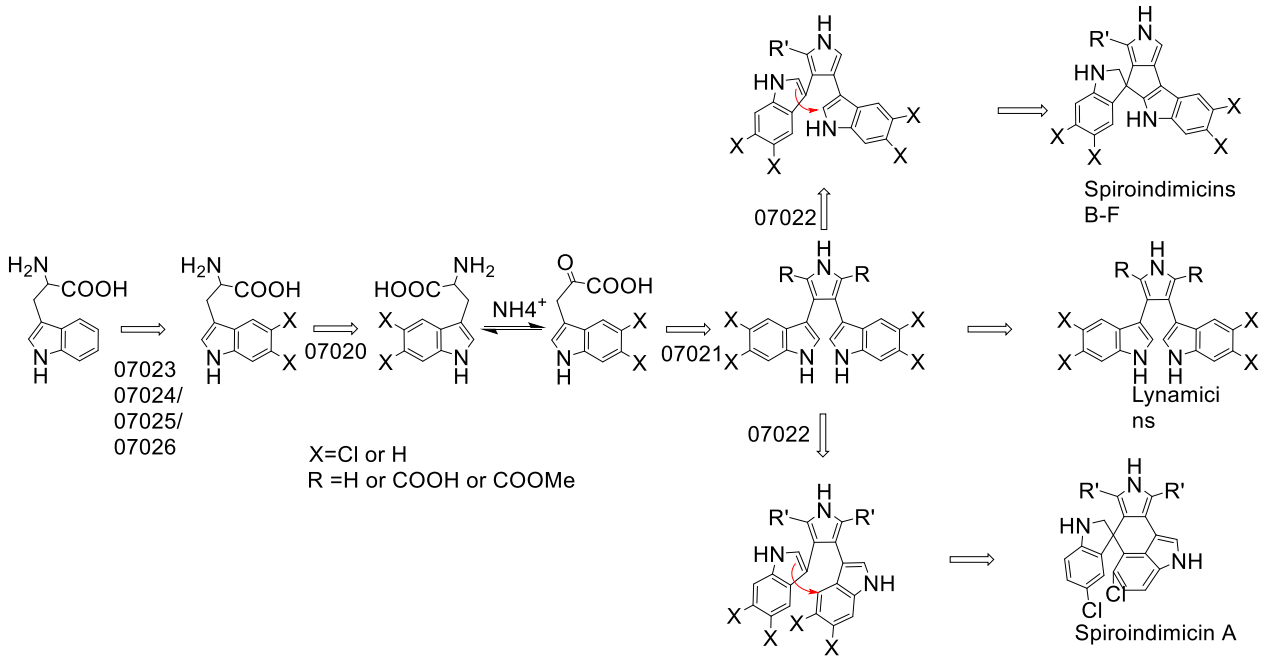


Figure 7S. Proposed biosynthesis pathway for lynamicins and spiroindimicins.

II

New Alpiniamides from *Streptomyces* sp. IB 2014/011-12 Assembled by an Unusual Hybrid Non-ribosomal Peptide Synthetase *Trans*-AT Polyketide Synthase Enzyme

Constanze Paulus, Yuriy Rebets, Josef Zapp, Christian Rückert, Jörn Kalinowski and Andriy Luzhetskyy

Frontiers in Microbiology, **2018**, *9*, 1959

DOI: 10.3389/fmicb.2018.01959

Published online 22nd August 2018

frontiers in Microbiology

ORIGINAL RESEARCH
published: 22 August 2018
doi: 10.3389/fmicb.2018.01959

New Alpiniamides From *Streptomyces* sp. IB2014/011-12 Assembled by an Unusual Hybrid Non-ribosomal Peptide Synthetase *Trans*-AT Polyketide Synthase Enzyme

*Constanze Paulus[1,2], Yuriy Rebets[2], Josef Zapp[2], Christian Rückert[3], Jörn Kalinowski[3] and Andriy Luzhetskyy[1,2]**

[1] Helmholtz-Institute for Pharmaceutical Research Saarland, Saarbrücken, Germany, [2] Department for Pharmaceutical Biotechnology, University of Saarland, Saarbrücken, Germany, [3] Center for Biotechnology (CeBiTec), Bielefeld University, Bielefeld, Germany

OPEN ACCESS

Edited by:
Marie-Joelle Virolle,
Centre National de la Recherche Scientifique (CNRS), France

Reviewed by:
Martin Schmeing,
McGill University, Canada
Stefano Donadio,
Naicons Srl, Italy
Yanyan Li,
Centre National de la Recherche Scientifique (CNRS), France

***Correspondence:**
Andriy Luzhetskyy
a.luzhetskyy@mx.uni-saarland.de;
andriy.luzhetskyy@helmholtz-hzi.de

Specialty section:
This article was submitted to Microbial Physiology and Metabolism, a section of the journal Frontiers in Microbiology

Received: 04 May 2018
Accepted: 02 August 2018
Published: 22 August 2018

Citation:
Paulus C, Rebets Y, Zapp J, Rückert C, Kalinowski J and Luzhetskyy A (2018) New Alpiniamides From Streptomyces sp. IB2014/011-12 Assembled by an Unusual Hybrid Non-ribosomal Peptide Synthetase Trans-AT Polyketide Synthase Enzyme. Front. Microbiol. 9:1959. doi: 10.3389/fmicb.2018.01959

The environment of Lake Baikal is a well-known source of microbial diversity. The strain *Streptomyces* sp. IB2014/011-12, isolated from samples collected at Lake Baikal, was found to exhibit potent activity against Gram-positive bacteria. Here, we report isolation and characterization of linear polyketide alpiniamide A **(1)** and its new derivatives B–D **(2–5)**. The structures of alpiniamides A–D were established and their relative configuration was determined by combination of partial Murata's method and ROESY experiment. The absolute configuration of alpiniamide A was established through Mosher's method. The gene cluster, responsible for the biosynthesis of alpiniamides (*alp*) has been identified by genome mining and gene deletion experiments. The successful expression of the cloned *alp* gene cluster in a heterologous host supports these findings. Analysis of the architecture of the *alp* gene cluster and the feeding of labeled precursors elucidated the alpiniamide biosynthetic pathway. The biosynthesis of alpiniamides is an example of a rather simple polyketide assembly line generating unusual chemical diversity through the combination of domain/module skipping and double bond migration events.

Keywords: *Streptomyces*, secondary metabolites, NRPS-trans-AT-polyketide synthase, stereochemistry, bioactivity

INTRODUCTION

The consistent development of antibiotic resistance in life-threatening pathogens diminishes the availability of effective medications for the treatment of infectious diseases. Natural products originating from plants and microorganisms have inspired medicinal drug research for decades (Ventola, 2015). They are produced as secondary metabolites and represent a major source of drug leads and serve as templates for semisynthetic derivatives. A vast number of secondary metabolites from microorganism, particularly from actinobacteria have been described to date. *Actinobacteria* represent one of the most thoroughly examined group of bacteria in terms of

natural product research, which is reflected in the continuing discovery of important antibiotics, e.g., teicoplanin, daptomycin, and fidaxomicin over the years (Procópio et al., 2012). Many natural products not only serve as antibiotics but also as antiviral, immunosuppressive and even as anticancer agents (Vaishnav and Demain, 2011). Hence, the discovery of new bioactive natural products is still indispensable for sustaining the rapid progress of medicinal research (Dias et al., 2012).

A large proportion of biologically active secondary metabolites are polyketide derivatives. These compounds are assembled by polyketide synthases (PKSs), enzymes that conduct a simple repetitive condensation of acyl units to form poly-ß-ketide chains followed by conversion of the chain into structurally diverse metabolites. Depending on the structure and the enzymatic pathway, PKSs can be divided into three major types. Type I PKSs have modular architectures consisting of several modules each harboring ketosynthase (KS) and acyl carrier protein (ACP) domains accompanied by an acyltransferase domain (AT), which is responsible for the selection of the extender unit, as well as a set of processing enzymes (KR – ketoreductase, DH – dehydratase, and ER – enolreductase) (Khosla, 2009). Recently, a new type of modular PKS was discovered that lacked the acyltransferase domains within the modular proteins (Helfrich and Piel, 2016). The acyltransferase activity is a standalone protein that is typically encoded within the respective gene cluster. PKSs of this type are known as trans-AT and they are believed to be the most abundant type of secondary metabolite assembly lines.

Bacterial strains of the genus *Streptomyces* inhabit different ecological niches. They are found not only in soil but also in fresh water systems, marine habitats, isolated eco-systems, and in symbiosis with insects and plants (Hasani et al., 2014). Lake Baikal is a unique eco-system, rich in endemic species of living organisms. We have recently isolated a variety of actinobacteria strains from Lake Baikal (Axenov-Gribanov et al., 2016). Several of these strains were found to be active against Gram-positive and Gram-negative bacteria. The strain *Streptomyces* sp. IB2014/011-12 showed the most promising results when tested against *Bacillus subtilis*. This finding motivated us to analyze the strain in its entirety to identify the metabolites responsible for the observed activity.

Here, we report the activity-guided screening of metabolites produced by *Streptomyces* sp. IB2014/011-12 that resulted in the isolation of alpiniamide A **(1)**, which was previously described as a product of the endophytic *Streptomyces* sp. YIM66017 (Golinska et al., 2015), and its novel derivatives. The genome of this strain has been sequenced and analyzed, which has enabled us to identify the gene cluster responsible for the biosynthesis of alpiniamides.

MATERIALS AND METHODS

Bacterial Strains, Culture Conditions, and General Procedures

The isolation and phylogenetic characterization of *Streptomyces* sp. IB2014/011-12 were reported in (Axenov-Gribanov et al., 2016). *Streptomyces* strains were grown on solid nutrient medium MS (mannitol soy flour agar) and in liquid TSB medium (Kieser, 2000). For secondary metabolite production, NL19 (MS medium without agar) and SG (glucose, yeast, Bacto Soytone, and calcium carbonate) medium have been used. *Escherichia coli* XL1Blue (Agilent, United States) was used for routine cloning, and *E. coli* MW 6026 was used as a donor in the intergenic conjugation (Blodgett et al., 2007). *E. coli* strains were grown in Luria-Bertani (LB) broth. For MW 6026, diaminopimelic acid was added. When required, antibiotics were added to the cultures at the following concentrations: 50 $\mu g\ ml^{-1}$ apramycin, 100 $\mu g\ ml^{-1}$ spectinomycin, 100 $\mu g\ ml^{-1}$ phosphomycin, and 100 $\mu g\ ml^{-1}$ carbenicillin (Sigma, United States; Roth, Germany).

Recombinant DNA Techniques

Chromosomal DNA from *Streptomyces* strains and plasmid DNA from *E. coli* were isolated using standard protocols (Makar et al., 1975; Kieser, 2000). Restriction enzymes and molecular biology reagents were used according to the recommendations of the supplier (Thermo Fisher Scientific, Germany; NEB, **United States**).

Genome Sequencing, Assembly, and Annotation

For DNA isolation, *Streptomyces* sp. IB2014/011-12 was inoculated into TSB medium and grown at 28°C with shaking (200 rpm) for 3 days. High quality total cellular DNA was isolated using salting out procedure. The purity and concentration of the genomic DNA was determined using a Nanodrop 2000 spectrophotometer (Thermo Fisher Scientific).

For sequencing of the *Streptomyces* sp. IB2014/011-12 genome, an Illumina paired-end sequencing library (TruSeq sample preparation kit; Illumina, United States) was constructed according to the manufacturer's protocol. The *Streptomyces* sp. IB2014/011-12 draft genome sequence was established on an Illumina HiSeq system in rapid run mode (2 × 250 nt) with a pair distance of about 500 bp. Upon sequencing and processing of the obtained data, a *de novo* assembly was performed using the GS De Novo Assembler (version 2.8.) (Roche Diagnostics, Mannheim, Germany) with default settings. Annotation of the genome was performed by means of prokka v1.11 and the GenDB 2.0 platform (Meyer et al., 2003; Seemann, 2014). For the identification of secondary metabolites clusters antiSMASH 3.0 was used (Weber et al., 2015). The assembled and annotated draft sequence of the *Streptomyces* sp. IB2014/011-12 genome was deposited in the GenBank database under accession number QEIK00000000.

Generation of the Construct for Gene Cluster Inactivation

Two DNA fragments, C9-1 and C9-2, flanking the first 2000 bp of the gene *alpA1* have been amplified using the primer pairs C9-2REcV and C9-2FEcI and C9-1FEcV and C9-1RXba (**Supplementary Table S1**). The obtained 2 kb fragments were cloned into a pST Blue-I AccepTor^{TM} vector (Novagen, United States). The construct containing fragment C9-1 was digested with *Eco*RV (primer) and *Xba*I (vector MCS) and ligated with the C9-2 fragment, which was retrieved with the

same restriction enzymes. The resulting plasmid was digested with *Eco*RV and ligated with the spectinomycin-resistance cassette. The final construct was transformed into pKG 1132 (Myronovskyi et al., 2011). This construct was transferred into *Streptomyces* sp. IB2014/011-12 via an intergeneric conjugation. The exconjugants were grown under non-selective conditions and screened for white spectinomycin-resistant colonies when grown on MS supplemented with X-gluc (X-Gluc DIRECT, United States). The deletion of part of the *alpA1* gene was confirmed by PCR using the 11-12DelCheckF and 11-12DelCheckR primer pair (**Supplementary Table S1**).

Cloning of the *alp*-Gene Cluster and Red/ET Mediated Gene Deletion

The two 2.5 kb DNA fragments, Tar1 and Tar2, flanking a 46.7 kb region of the *Streptomyces* sp. IB2014/011-12 chromosome that includes the entire *alp*-gene cluster were amplified with the primer pairs 11-12C9TarF1Not and 11-12C9TarR1Nhe and 11-12C9TarF2NheI and 11-12C9TarR2HindIII, respectively (**Supplementary Table S1**), using Phusion DNA polymerase and cloned into the pJET1.2 vector (Thermo Fisher Scientific, **United States**). Fragments were assembled by digesting the Tar1-containing plasmid with *Nhe*I (primer) and *Hind*III (vector MCS) and ligating with the *Nhe*I/*Hind*III-retrieved Tar2 fragment. The resulting construct was sub-cloned into a *Nhe*I/*Hind*III-digested pCLY10 vector (Bilyk et al., 2016). The final construct was linearized with *Nhe*I and mixed with *Streptomyces* sp. IB2014/011-12 chromosomal DNA in a 1:5 ratio. The mixture was transformed into *S. cerevisiae* BY4742 (Winston et al., 1995) with the standard LiAc protocol (Gietz and Schiestl, 2007). Transformants were selected on YNB medium supplemented with yeast synthetic drop-out medium supplements without leucine (Sigma-Aldrich, United States). Colonies were plated in patches of 100, washed, and analyzed by PCR for the presence of clones harboring the desired construct using the primers 11-12C9CheckHind (**Supplementary Table S1**) annealed to the cloned region outside of the homology fragment and pCLYCheckHind (**Supplementary Table S1**) annealed to the vector. The positive clone was further pooled out, the total DNA was purified using a standard protocol (Green and Sambrook, 2012) and transformed into *E. coli* XL1Blue to give 011-12p1-49 clone carrying the desired region of the *Streptomyces* sp. IB2014/011-12 chromosome. 011-12p1-49 plasmid DNA was purified and sequenced by MinION (Oxford Nanopore, United Kingdom). 011-12p1-49 was introduced into *S. lividans* TK24 and *S. albus* Del14 (unpublished data, Dr. M. Myronovskyi, personal communication) by intergeneric conjugation. The resulting strains were grown, and the production of alpiniamides was analyzed as described below. *alpD*, *alpR*, and *alpE* genes were inactivated by replacement with a hygromycin resistance cassette from plasmids patt-shyg (Myronovskyi et al., 2011) within the 011-12p1-49 construct. λ-Red recombineering was performed as described (Gust et al., 2004). Primers used to amplify the cassette and to verify the deletions are listed in **Supplementary Table S1**. The resulting constructs were introduced into the host *S. albus* Del 14 via conjugation and clones were selected with hygromycin. The mutants were cultivated in NL19 and production was analyzed using LC-MS as described below.

Production, Extraction and LC-MS Analysis of Alpiniamides

The *Streptomyces* sp. IB2014/011-12 was grown in 10 ml of TSB for 2 days at 28°C on a rotary shaker. The main culture (100 ml of NL19 in 500 mL flasks with glass beads) was inoculated with 1 ml of the pre-culture and cultivated at 28°C and 180 rpm for 7 days. The metabolites in the cultural liquid were extracted with ethyl acetate and from the biomass with a mixture of acetone and methanol (1:1). The solvents were evaporated, and the residue was dissolved in 300 μl of methanol.

The LC-MS data were collected on a Dionex Ultimate 3000 RSLC system using a BEH C18, 100 × 2.1 mm, 1.7 μm d_p column (Waters, Germany). Separation of a 1 μl sample was achieved by a linear gradient of solvent B (acetonitrile with 0.1% of formic acid) against solvent A (water with 0.1% of formic acid) at a flow rate of 600 μl/min and 45°C. The gradient started with a 0.5 min isocratic step of 5% B then increased to 95% B over 18 min and ended with a 2 min step of 95% B before re-equilibration under the initial conditions. UV spectra were acquired by a DAD in the range of 200 to 600 nm. The mass spectrometry data were collected on an amazon SL speed mass spectrometer (Bruker Daltonics, Germany) using an Apollo II ESI source. Mass spectra were acquired in centroid mode ranging from 200 to 2000 *m/z* at a scan rate of 2 Hz. The HRMS data were collected on a Dionex Ultimate 3000 RSLC system using a BEH C18, 100 × 2.1 mm, 1.7 μm d_p column (Waters, Germany). Separation of a 1 μl sample was achieved by a linear gradient of solvent B (acetonitrile with 0.1% of formic acid) against solvent A (water with 0.1% of formic acid) at a flow rate of 480 μl/min and 45°C. The gradient started with a 0.5 min isocratic step of 5% B then increased to 95% B over 20 min and ended with a 2 min step of 95% B before re-equilibration under the initial conditions. UV spectra were acquired by a DAD in the range of 200 to 600 nm. High-resolution mass spectrometric data were collected on an LTQ Orbitrap mass spectrometer (Thermo Fischer Scientific, United States).

Data were collected and analyzed with the Bruker Compass Data Analysis software, version 4.2 (Bruker, Billerica, MA, United States) and the Thermo Xcalibur software, version 3.0. The screening for known compounds was performed using the Dictionary of Natural Products Database, version 10.0 (CRC Press, Boca Raton, FL, United States), using the following parameters: accurate molecular mass, absorption spectra and biological source (Running, 1993). Compounds were considered to be similar when the difference in accurate mass was less than 2 ppm and the absorption spectra were identical.

Feeding Experiments

Thirty milliliter of liquid NL19 medium was inoculated with 300 μl of a 22 h old seed culture of the strain *Streptomyces* sp. IB2014/011-12. After 10 h of growth on a rotary shaker at 28°C, the culture was supplemented with 200 μl of

Glycine-2-^{13}C or 200 µl methionine(-methyl-^{13}C) respectively, solved in water (5.2 mg/ml). This procedure was repeated four times every 10–12 h. The final concentration of compounds in the flask was 2.31 mM of Glycine-2-^{13}C and 1.16 mM of methionine(-methyl-^{13}C). After 7 days of cultivation, the culture was extracted separately, biomass with acetone/methanol (1:1) and the supernatant with ethyl acetate. The solvent was evaporated and the obtained residues solved in MeOH. The extracts were subjected to LC-MS analysis.

Isolation and Structure Elucidation of Alpiniamides

For the isolation of the metabolites, the strain was cultivated as described above in 8 L of NL19. Metabolites were extracted from cultural liquid with equal volume of ethyl acetate, solvent was evaporated and resulting 1.56 g of crude extract was dissolved in 8.5 ml of methanol. The crude extract was purified through size-exclusion chromatography using Sephadex® LH 20 (Sigma-Aldrich) and MeOH as eluent (1 m long column with 700 ml volume of Sephadex). Fractions were collected every 15 min with a speed of 1-2 drops per second (approximately 30 ml per hour). The obtained fractions with antibacterial activity were further purified by preparative and subsequent semipreparative high-performance-liquid-chromatography (HPLC) using the following equipment: Dionex Ultimate 3000 from ThermoScientific (preparative) and Agilent 1260 Series and 1100 Series from Agilent Technologies (semipreparative). For preparative HPLC, a Nucleodur C18 HTEC column (150 × 21 mm, 5 µm) was used with a multistep gradient from 15–20% B (B: acetonitrile with 0.1% formic acid; A: water with 0.1% formic acid) over 2 min and 20–60% B over 20 min at a flow rate of 20 ml/min and 45°C. Semipreparative HPLC was performed using a Synergi Phusion RP-Column (250 × 10 mm, 4.6 µm; Phenomenex) with a gradient elution from 5 to 95% B (B: methanol with 0.1% of formic acid; A: water with 0.1% formic acid) over 20 min at a flow rate of 4.5 ml/min and 45°C. UV spectra were recorded with a DAD detector at 200–600 nm.

NMR spectra were acquired in deuterated methanol (CD_3OD) and deuterated chloroform ($CDCl_3$) at 298 K on a Bruker Avance III 700 or 500 MHz spectrometer, both equipped with a 5 mm TXI cryoprobe. NMR shifts were relative to the residual solvent signal CH_3OD at δ 3.30 and $CDCl_3$ δ 7.24 for ^{1}H, or to the solvent itself at δ 49.00 (CD_3OD) and 77.00 ($CDCl_3$) for ^{13}C measurements. NMR data were analyzed using Topspin, version 3.5 pl7 (Bruker, United States).

Preparation of the S- and R-MTPA-Ester of Alpiniamide A (6S and 6R), Alpiniamide B_1 (7S and 7R), and Alpiniamide B_2 (8S and 8R) by the Modified Mosher's Method

To a solution of alpiniamide A (**1**, 0.5 mg, 1.45 µmol) and dry pyridine (20 µl) in dry deuterated chloroform (100 µl) at room temperature, α-methoxy-α-(trifluoromethyl)phenylacetyl chloride (R-MTPA-Cl) (20 µl, 0.10 mmol) was added. The reaction was mixed at room temperature for 4 h. Another 200 µl of $CDCl_3$ was added, and the sample now containing the S-MTPA-ester of alpiniamide A (**6S**) was directly subjected to NMR measurements (**Supplementary Figure S1**). The R-MTPA-ester of alpiniamide A (**6R**) was prepared in the same way using S-MTPA-Cl instead of its R-isomer. The same procedure was carried out with only a fifth of all ingredients to get **7S** and **7R** from alpiniamide B_1 **(2)** and **8S** and **8R** from alpiniamide B_2 **(3)**, respectively (**Supplementary Figure S1**).

Bioactivity Assay

The antimicrobial activities of isolated compounds were measured by a disk diffusion assay. A 30 µl aliquot of each compound was loaded on a paper disk with a diameter of 6 mm, allowed to dry, and set on an LB agar plate coated with *B. subtilis* from liquid overnight culture. The plates were incubated at 37°C for 12 h. The zones of inhibition were measured manually. The minimal inhibitory concentrations were estimated by standard serial dilutions protocol in 200 µL in 96 well plates. Briefly, serial dilutions (1:1) of alpiniamides were prepared using DMSO as a solvent prior to aliquoting 10 µL of each solution into 96 well plate. Kanamycin was used as a positive control (full inhibition of growth), and DMSO was used for the negative control (full growth). 190 µL of bacterial test cultures in appropriate media (1:500 dilution of overnight culture) were added to each well, and the plates were shaken at 30°C for 16-20 h. Then, 5 µL of thiazolyl blue tetrazolium bromide (10 mg mL^{-1}) solution was added to each well, and the plates were incubated at 30°C for an additional hour. MICs were determined as the concentration of antibiotic in the well where the additive did not change the color of the solution from yellow to dark blue.

RESULTS

Genome Analysis

Streptomyces sp. IB2014/011-12 has been isolated from the net-spinning caddisfly *Trichoptera* sp. (larvae) (Axenov-Gribanov et al., 2016). Phylogenetic analysis based on the 16 sRNA showed that the strain IB2014/011-12 belongs to the genus *Streptomyces*. The strain grows as a typical actinobacteria forming colonies with aerial mycelia and gray spores (**Supplementary Figure S2**). The strain accumulates a brownish pigment during growth on a solid medium. When grown in liquid media (NL19) the strain demonstrated activity against Gram-positive bacteria. The genome of *Streptomyces* sp. IB2014/011-12 has been sequenced and assembled into 71 contigs (**Table 1**). The largest contig is 956.9 kbp. The overall size amounts to 8099.23 kbp, which is in the normal range for streptomyces. The genome consists of a single chromosome and has no extra chromosomal DNA based on the sequence coverage. The genome has a G + C content of 71.5%. The chromosome is predicted to contain 7.282 coding sequences, 5 rRNA clusters, 80 tRNA genes and one tmRNA gene (Palecková et al., 2009).

TABLE 1 | Features of *Streptomyces* sp. IB2014/011-12 genome.

Genome size	8099.23 bp
Contigs	71
GC content	71.5%
CDS	7.282
tRNA	80
mRNA	1
rRNA	5
Cluster	29

The analysis of the genome using antiSMASH software revealed that 29 gene clusters are involved in the biosynthesis of diverse secondary metabolites including terpenes, lanthipeptides, non-ribosomal peptides, and polyketides (**Supplementary Table S2**).

Based on the gene cluster analysis with antiSMASH the production of several secondary metabolites by *Streptomyces* sp. IB2014/011-12 can be predicted (**Supplementary Table S2**). Like most of actinobacteria strains, the genome contains gene clusters for desferrioxamine, ectoine, and melanin production. Desferrioxamine B, a high-affinity iron chelator (siderophore), is important for scavenging ferric iron as a nutrient for the cells (Barona-Gómez et al., 2004). Ectoine, another common metabolite, is an osmoprotectant (Zhu et al., 2014). The function of melanin, a black pigment, that is produced by many types of bacteria, is still under discussion (Woo et al., 2010). The compound is most likely involved in protection against chemical and biological stresses, such as exposure to heavy metals, oxidizing agents and UV radiation (Allam, 2012), and it might be the reason why the MS medium is colored by that strain (**Supplementary Figure S2**). In addition, the gene cluster that produces the class III lantipeptide AmfS is present in the genome of *Streptomyces* sp. IB2014/011-12. The predicted amino acid sequence of the precursor peptide from the gene cluster coincides with the sequence of AmfS isolated from *Streptomyces griseus* (Ueda et al., 2002). AmfS-like compounds are known to be positive regulators of the formation of aerial-mycelium in streptomyces. Another secondary metabolite that could be predicted from the *Streptomyces* sp. IB2014/011-12 genome analysis is roseoflavin. This compound belongs to the riboflavin antibiotic family, which targets riboswitches that affect bacterial growth (Mansjö and Johansson, 2011; Schwarz et al., 2016). Furthermore, the gene cluster no. 3 is probably coding for the class II lasso peptide SRO15-2005 since it shows 100% sequence homology to the corresponding gene cluster of *S. roseosporus* NRRL 15998 (Kersten et al., 2013).

Dereplication of Secondary Metabolites Produced by *Streptomyces* sp. IB2014/011-12

Dereplication is a quick and simple method of analyzing metabolites based on LC-MS data. It allows known compounds to be distinguished from potentially new metabolites to avoid their purification and analysis. With the help of UV/Vis spectra, high-resolution mass spectrometric data, biological source and other criteria, one can compare entries in data banks such as the "Dictionary of Natural Products (DNP)" in order to identify metabolites in the extract of interest (Whittle et al., 2003). At first, to estimate the potential novelty of the compounds produced by *Streptomyces* sp. IB2014/011-12, we analyzed the LC-MS data of the extracts of the strain grown in NL 19 media. The obtained exact masses from MS data were further compared to the DNP. This led to the identification of some known compounds and prediction of putative new metabolites produced by this strain (**Figure 1**). The major class of compounds produced by this strain is polycyclic tetramate macrolactams (PTM) (**Supplementary Figures S3**, **S4**) (Xu et al., 2015). We identified alteramide A **(9)** (Shigemori et al., 1992) and several isomers that were not distinguishable since they all have *m/z* 511.28207 $[M+H]^+$ (510.27427 $[M]$, calculated 510.2730) ions and UV curves typical for PTMs but different retention times. Under the used conditions, the isomers elute at RTs of 17.7, 18.6 and 19.0 min. Alteramide B **(10)** and several isomers with *m/z* 495.28549 $[M+H]^+$ (494.27769 $[M]$, calculated 494.2781) were present in the extract as well. They eluted at RT 18.3, 18.9, 19.3, and 19.5 min. In addition to these, the strain also produces clifednamide A **(11)** (*m/z* 509.26395 $[M+H]^+$, 508.25615 $[M]$ calculated 508.2573, RT at 18.4) and clifednamide B **(12)** (*m/z* = 493.27017 $[M+H]^+$, 492.26237 $[M]$ calculated 492.2624, RT at 18.7) (Cao et al., 2010) and possible isomers of these compounds that eluted at different retention times (RT at 18.8 and 18.9 min). Furthermore, the compound eluting at RT 18.2 min with *m/z* 513.29565 $[M+H]^+$ (512.28785 $[M]$, calculated 512.2886) corresponds to dihydromaltophylin **(13)** (heat-stable antifungal factor, HSAF) (Li et al., 2008).

Alteramide A and B, isolated from marine bacterium *Alteromonas species,* are macrocyclic lactams that contain a dienone and a dienoyl tetramic acid moiety (Shigemori et al., 1992; Ding et al., 2016). They exhibit activity against leukemia cells but are not active against bacteria. Dihydromaltophilin is similar to alteramide in that it also has three five-membered rings. The compound was isolated from *S. maltophilia* R3089 and exhibits activity against a broad spectrum of fungi but is not active against Gram-positive or Gram-negative bacteria (Jakobi et al., 1996). The clifednamides, another group of PTMs, were first isolated from *Streptomyces* sp. JV178. The biological activity of these compounds has not yet been published. In the genome of *Streptomyces* sp. IB2014/011-12, a hybrid type I PKS-NRPS cluster no. 13 has been identified, and it is similar to the known gene clusters for PTMs biosynthesis (**Supplementary Figure S5**). The biosynthetic gene cluster for alteramide, clifednamide and dihydromaltophilin differ slightly in their structures. Cluster no. 13, with the core gene coding for iterative hybrid NRPS-type I PKS, provide all necessary enzymatic activities for PTM production (CDCs: 14650–14675, **Supplementary Figure S5**). The gene cluster encodes for a putative hydroxylase (14675), an iterative type 1 PKS-NRPS (14670), two oxidoreductases (14665–14660), alcohol dehydrogenase (14655), and cytochrome P450 hydroxylase (14650) in the same arrangement as in the

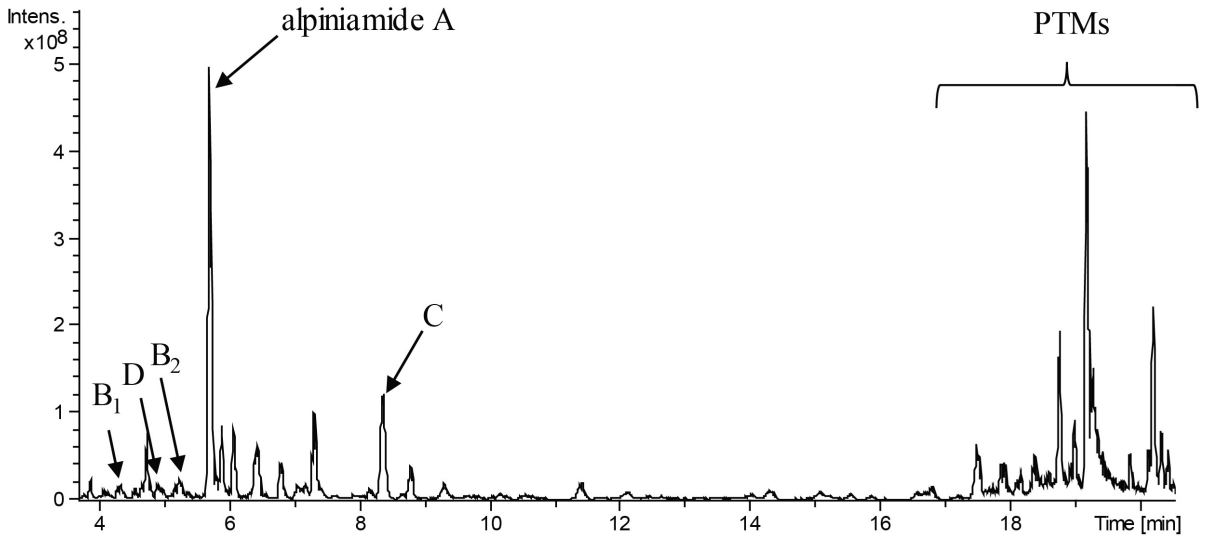


FIGURE 1 | LC-MS chromatogram (minute 4 to 20 from an overall 20 min run) of extract of *Streptomyces* sp. IB2014/011-12 cultivated in NL19 medium. The peaks for different alpiniamides are marked with letters. Peaks for polycyclic tetramate macrolactams are highlighted with the bracket.

biosynthetic gene cluster for HSAF and frontalamides (Lou et al., 2011).

Isolation and Structure Elucidation of Alpiniamide A–D (1–5)

Streptomyces sp. IB2014/011-12 produced additional metabolites (**Figure 1**) that we could not dereplicate against known products. For this reason, we assumed that the activity was caused by potentially new metabolite(s). We grew the strain in 8 L of NL19 and extracted the produced metabolites. Subsequently, a bioactivity-guided purification was carried out by tracking down the bioactive compounds using the disk diffusion assay at each step of purification. The crude extract was first fractionated using size-exclusion chromatography, and the obtained fractions were tested against *B. subtilis*. The active fractions were combined and further purified with preparative HPLC and then with semipreparative HPLC. The fractions collected after each separation step were tested for antibacterial activity. This resulted in the isolation of five compounds. The structure elucidation showed that they are alpiniamides (**Figure 2**).

Alpiniamide A (1) was obtained as a slightly red solid (4 mg), with a molecular formula of $C_{17}H_{29}NO_6$ as determined by high-resolution electrospray ionization mass spectrometry (HRESIMS) at *m/z* 344.20596 $[M+H]^+$, indicating 4 degrees of unsaturation. Its NMR spectra, acquired in CD_3OD (**Supplementary Table S3**) revealed 17 carbons, five methyls (one adjacent to a double bond), two methylenes, six methines (two secondary alcohols and one as part of a double bond), and four quarternary carbons (three carbonyls and one as part of a trisubstituted double bond). The chemical shifts of the methylene group at δ_H 4.13 and 4.25 and δ_C 49.0 indicated the presence of an amide nitrogen atom nearby. These data together with the results of 2D ^{1}H-^{1}H-COSY, HSQCED, HMBC, and ROESY led to the structure of alpiniamide, which was previously isolated from the *Streptomyces* sp. YIM660107 (Zhou et al., 2013), with NMR data recorded in $CDCl_3$. In order to compare the literature data with our data, we reran the 1D NMR spectra of **1** in $CDCl_3$ (**Supplementary Table S3** and **Supplementary Figures S6**, **S7**). The obtained data were close to those from the reported structure. But the remarks to the relative stereochemistry of alpiniamide were confusing. C-2/C-3 was found to be *threo*, but the structure of alpiniamide showed the *erythro*-form. C-2′/C-3′/C-4′ was announced as *threo* as well, which is ambiguous and most likely a wrong phrasing for a system with three chiral centers. However, the structure showed C-2′/C-3′ and C-3′/C-4′ both in *erythro*-configuration. Therefore, we were motivated to investigate the stereochemistry of **1** in more detail.

Due to the large vicinal coupling constant of 9.5 Hz for J_{H2H3}, the configuration for C-2/C-3 in part A (C-1 to C-9) was found to be *threo*. We are aware, that a single coupling constant cannot usually distinguish which of two diastereomers might be present since there are three possible staggered conformations for each diastereomer, two of which will typically have very similar predicted coupling constants for a pair of vicinal protons. But assignments become possible when one can make some reliable predictions on which conformation predominates. Such a situation is given for part A of **1**, where the hydroxyl at C-3 can form an intramolecular hydrogen bond with carbonyl C-1 when measured in the non-polar solvent $CDCl_3$. Then, the $^3J_{HH}$ coupling constant of two vicinal protons are considerably large when they are in the *threo* configuration. Small values for $^3J_{HH}$ indicate the *erythro* form (Stiles et al., 1964; House et al., 1973). A similar situation predominates in part B between the hydroxyl at C-3′ and the carbonyls C-1′ and C-5′. Careful analysis of $^3J_{HH}$ coupling constants led to 7.0 Hz for $J_{H2'H3'}$ and 4.5 Hz for $J_{H3'H4'}$. The small coupling constant for H-3′/H-4′ give rise

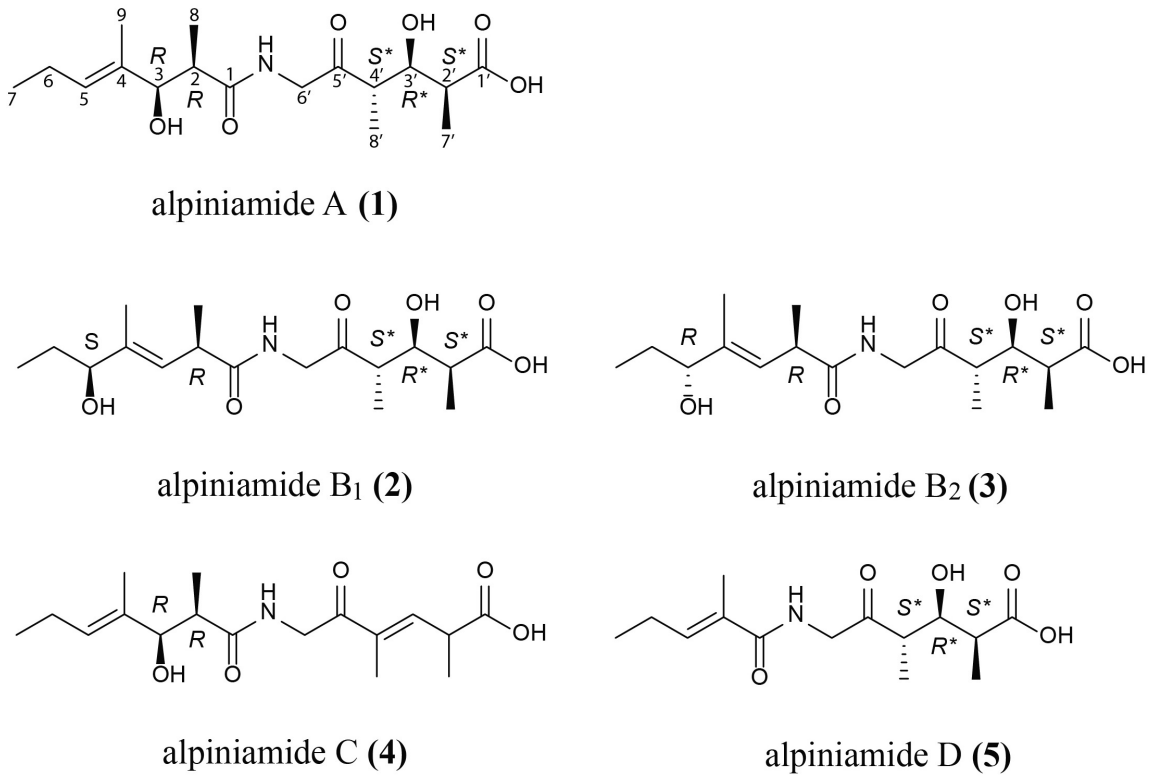


FIGURE 2 | Revised structure of alpiniamide A **(1)** and structures of new alpiniamides B-D **(2–5)** from *Streptomyces* sp. IB2014/11-12.

to an *erythro* configuration, whereas the moderate high *J* value for H-2′/H-3′ is a hint for the *threo*-form (**Supplementary Table S3**) (Xu et al., 2017). Nevertheless, a final proof of the relative stereochemistry of **1** requires a complete set of $^{3}J_{HH}$ and $^{2,3}J_{HC}$ coupling constants as shown in Murata's method (Matsumori et al., 1999; Bifulco et al., 2007). Due to the limited amount of **1** and its insufficient stability during storing, we were not able to perform it.

In order to determine the absolute configuration as well, we applied Mosher's method (Dale and Mosher, 1973; Hoye et al., 2007). Portions of **1** were separately treated with (R)-MTPA-Cl and (S)-MTPA-Cl to yield the Mosher ester **6S** and **6R**, respectively (**Supplementary Figure S1**). The differences in the proton chemical shifts of **6S** and **6R** ($\Delta\delta_{(S-R)}$) should give positive or negative values from which the configuration can be established. For part A of the molecule we obtained negative values for H-2 and H-8 and positive values for H-5, H-6, H-7, and H-9 (**Supplementary Table S4** and **Supplementary Figure S8**). On the basis of these results, the absolute configuration can be determined as *R* at C-3. From the relative configuration, the neighboring positions can also be assigned, resulting in *R*-configuration at C-2. The results for part B are less clear. We obtained negative values for H-2,' and positive values for H-4,' H-7,' and H-8'. For an appropriate interpretation H-2' and H-7' should have one sign and H-4' and H-8' the opposite sign. However, H-7' does not coincide with this rule and makes therefore a distinct statement about the absolute configuration for H-3′ impossible. Such issues with the assignment of the absolute configuration with MTPA are known for linear secondary alcohols (Seco J.M. et al., 2004). The reason for inconsistent sign distribution and small $\Delta\delta_{(S-R)}$ values often lies in the presence of different conformers of the MTPA ester. Theoretical calculations revealed that a rotation about the C_α – CO and C_α – Ph bond generates three conformers (Latypov et al., 1996). All three conformers are present in similar populations and each conformer has a different shielding/deshielding effect, and therefore it influences the final spectrum in different ways. This information together with the inconsistent results from Mosher's method for the identical parts of the close related compounds **2** and **3** (see below) hinders the assignment of the absolute configuration for C-3′. Alternative reagents, e.g., methoxyphenylacetic acid (MPA) or α-(9-anthryl)-α-methoxy-acetic acid (9-AMA) should give more precise values (Seco J.M. et al., 2004). Due to the little amount of **1** we had to refrain from further efforts to determine the absolute configuration for this part of the molecule.

Therefore, the structure of **1** was established as shown in **Figure 3**. We named it alpiniamide A, due to the close relationship to alpiniamide from a *Streptomyces* species mentioned above. It is not excluded that alpiniamide and alpiniamide A have identical structures. But due to the confusion concerning the relative configuration of alpiniamide in the literature and the lack of coupling constants especially for H-2′-H-4′, we were not able to determine it.

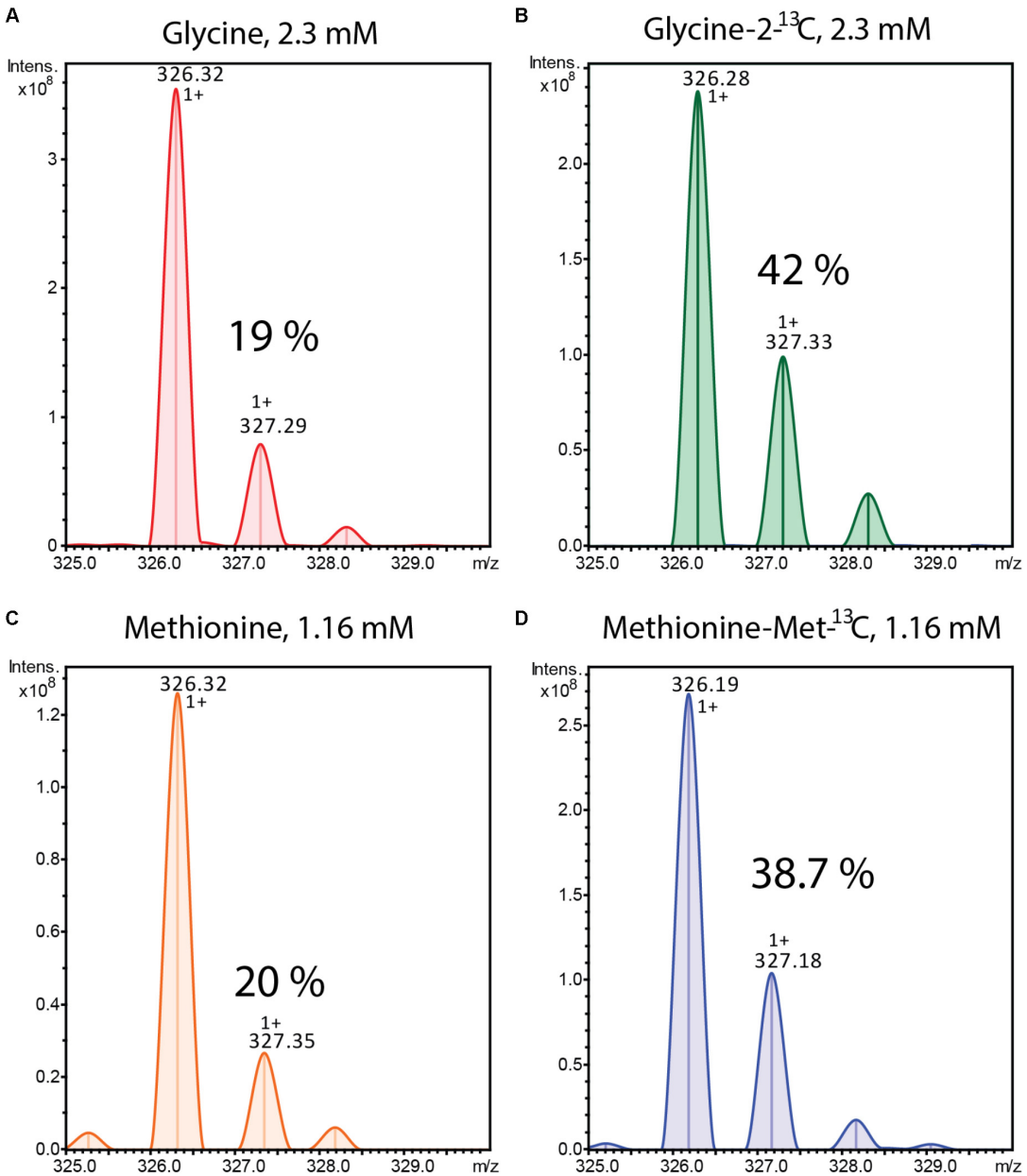


FIGURE 3 | Mass spectra of alpiniamide A (*m/z* 326.32 $[M\text{-}H_2O+H]^+$) extracted from *Streptomyces* sp. IB2014/011-12 culture cultivated in the presence of unlabelled **(A)** and ^{13}C labeled **(B)** glycine and unlabelled **(C)** and ^{13}C labeled **(D)** methionine. The % of isotope containing ions is indicated.

Alpiniamide B_1 and B_2 (2 and 3), were both isolated as red solids (0.5 mg), with a molecular formula of $C_{17}H_{29}NO_6$ as determined by their HRESIMS data (*m/z* 344.20596 $[M+H]^+$). The ^1H NMR data (**Supplementary Table S3** and **Supplementary Figures S9**, **S10**) of both epimers were very close to each other and to those found for **1**. But in contrast to it, the double bond in part A of **2** and **3** was shifted from C-4 to C-3 and the secondary alcohol from C-3 to C-5. The only difference between alpiniamide B_1 and B_2 was found in the stereochemistry of the secondary alcohol at C-5. When applying Mosher's method to both isomers, the stereocenter at C-5 was assigned as *S* in alpiniamide B_1 and *R* in alpiniamide B_2 (**Figure 2**, **Supplementary Table S4**, and **Supplementary Figure S8**). Due to the high conformity in the chemical shifts of H-2 and its neighboring atoms the configuration at C-2 is the same in both molecules and most likely *R* as it is given for alpiniamide A. The *E*-geometry of the double bond C-3/C-4 was confirmed by ROESY measurements with key correlation between double bond proton H-3 and the H-5 of the secondary alcohol. Part A is identical in all three alpiniamides. Even the coupling constants for H-2′, H-3,′ and H-4′ of the three compounds were identical indicating the same stereochemistry

for this substructure in all three molecules. As for alpiniamide A, the determination of the absolute stereochemistry via Mosher's method failed for **2** and **3** due to the anomaly of the detected $\Delta\delta_{(S-R)}$ values (**Supplementary Figure S8**).

Alpiniamide C (4) was obtained as a white solid (0.8 mg) with a molecular formula of $C_{17}H_{27}NO_5$ as determined from its HRESIMS at *m/z* 326.19589 $[M+H]^+$. The mass difference of 18 units compared to **1–3** indicated the loss of a water molecule. The ^1H and ^{13}C resonances (**Supplementary Table S3** and **Supplementary Figure S11**) lacked one methine and a secondary alcohol function. Instead, resonances for an additional trisubstituted double bond (δ_H 6.87; δ_C 134.8 and 147.0) appeared in the spectra. The chemical shifts of part A were close to those of **1**. Therefore, the additional double bond must be located in part B. Its position between C-3′ and C-4′ was established by HHCOSY correlations starting from methyl H-7′ (δ_H 1.29 d, 7.0 Hz) via H-2′ (δ_H 3.40 m) to H-3′ (δ_H 6.87 d, 9.5 Hz).

Alpiniamide D (5) is the derivative with the lowest molecular mass. It is a colorless solid (0.4 mg) with a molecular formula of $C_{14}H_{23}NO_5$ as determined from its HRESIMS at *m/z* 286.16458 $[M+H]^+$. Due to its NMR data, part B of **5** is identical to that of **1**. However, signals for methines C-2 and C-3 and methyl C-8 in part A were missing in the spectra of **5** (**Supplementary Table S3** and **Supplementary Figure S12**) and gave therefore a hint of their deletion in part A. Vicinal HMBC correlations from carbonyl C-1 (172.2) to the double bond proton at δ_H 6.39 and the adjacent methyl protons at δ_H 1.84 supported this observation and led to structure **5** for alpiniamide D.

Alpiniamide A was previously isolated from the *Streptomyces* sp. YIM 66017. The crude extract of that strain was tested against the *Bacillus anthracis* and fungi *Fusarium solani* (Zhou et al., 2013). In both cases inhibition zones were observed. Hence, we have tested the known alpiniamide A and its derivatives C and D on a panel of bacterial and yeast test cultures. The minimal inhibitory concentration (MIC) was determined against the Gram positive bacteria *Staphylococcus carnosus* DSMZ 20501, *Kocuria rhizophila* DSMZ 348, *Enterococcus mundtii* DSMZ 4840, *Micrococcus luteus* DSMZ 1790, *Mycobacterium smegmatis* DSMZ 43286, and *B. subtilis* DSMZ 10, against the Gram negative bacteria *Erwinia persicina* DSMZ 19328 and *Pseudomonas putida* KT2440 and against yeast *Candida glabrata* DSMZ 11226. In all cases we did not observe growth inhibitory activity in the concentration range up to 100 μg/ml. Also, pure compounds were found to be not active against *B. subtilis* in disk diffusion test. Since the mixture of all alpiniamides in the last stage of the purification was clearly active against *B. subtilis* we assume that the inhibitory concentration of alpiniamides lies above 100 μg/ml or the observed antibacterial activity is caused by synergistic effect with some minor compound(s).

Feeding Experiments

From the structures of the isolated compounds we can predict that they are synthesized by the combined action of NRPS and PKS enzymes. The left and the right sides of the amide bond are typical products for a type I PK, whereas the peptide bond in the middle most likely arises from the amino acid glycine introduced by an NRPS enzyme. The methyl groups of the alpiniamides could originate from the direct use of methylmalonate during polyketide assembly or from a secondary methylation event with S-adenosyl methionine (SAM) as a donor. We grew two cultures of *Streptomyces* sp. IB2014/011-12 fed with ^{13}C-labeled glycine and with ^{13}C-labeled methionine (methyl-^{13}C), respectively. The glycine was fed for the first time at 10 h post-inoculation. Methionine was fed for the first time at 36 h after inoculation. For both compounds, the feeding was repeated four more times every 10 h resulting in final concentration 2.31 mM of Glycine-2-^{13}C and 1.16 mM of methionine(-methyl-^{13}C). As a control, the culture supplemented with the corresponding unlabeled compound was used. The metabolites were extracted and analyzed by LC-MS. As result, the +1 isotopic peaks of all alpiniamides in the ^{13}C-labeled glycine culture increased in intensity (42%) indicating the successful incorporation of the compound (**Figures 3A,B**). In contrast, the +1 isotopic peaks of alpiniamides in the culture supplemented with unlabeled glycine were observed at their normal intensity (19%). In the case of methionine, the results demonstrated the successful incorporation as well. The +1 isotopic peak of alpiniamides in the ^{13}C-labeled methionine culture reached an intensity of 38.7% (**Figures 3C,D**). The +1 isotopic peaks of alpiniamide in the unlabeled culture appeared at their normal intensity of 20%. These results suggest that glycine is incorporated intact into alpiniamides and is the source of the amino group. On the other hand, we confirmed that at least some methyl groups originate from the SAM-dependent C-methyltransferase activity rather than methylmalonate.

Alpiniamide Gene Cluster Inactivation and Heterologous Expression

The labeled substrate feeding tests suggest that the alpiniamides probably arise from the action of a hybrid PKS/NRPS assembly line that facilitates the incorporation of glycine. The only gene cluster in the genome of *Streptomyces* sp. IB2014/011-12 that corresponds to the hypothesized biosynthetic scheme is no. 9 (**Figure 6A** and **Table 2**). This cluster encodes a hybrid NRPS-trans-AT-PKS that consists of four ketosynthase and one NRPS modules and a single NRPS module with an adenylation domain predicted to be specific for glycine as substrate. Two PK modules contain SAM-dependent C-methyltansferase domains. Overall, the architecture of the PKS-NRPS enzymes of cluster no. 9 suggests that it might be involved in the biosynthesis of alpiniamides. To verify this assumption, we replaced 2 kb of gene 12750 encoding the first PKS megaenzyme and its promoter with the spectinomycin-resistance cassette. The deletion was achieved through targeted gene disruption via double homologous recombination. The utilization of the pKG1132 suicide vector simplified the selection of colonies with double crossover by blue-white phenotypes due to the presence of the *gusA* gene in the vector backbone (Myronovskyi et al., 2011). The mutant strain and the wild type were cultivated in NL19 production medium, and the produced metabolites were analyzed by LC-MS. Compared to the wild-type strain, the mutant IB2014/011-12ΔalpA1 completely lacks the ability to

TABLE 2 | Description of alpiniamide biosynthesis gene cluster and deduced function of *alp*-genes.

ID	Name	Predicted function	Closest characterized homologs (% of identical amino acids)	GenBank accession number
12715	*alpI*	fatty acyl-CoA reductase	putative short chain dehydrogenase, hitachimycin, 46% (*Streptomyces scabrisporus*)	LC008143
12720	*alpH*	TetR family transcriptional regulator	*TetR/AcrR family; *Streptomyces* sp. IB2014 011-1 (100%)	-
12725	*alpW*	transporter, ATP binding protein	UvrA-like protein, quinomycin, 76% (*Streptomyces griseovariabilis*)	JN852959
12730	*alpG*	glyoxalase protein	putative lyase, pentalenolactone, 82% (*Streptomyces avermitilis*)	BA000030
12735	*alpF*	transcriptional activator/DNA repair enzyme	putative AraC-family transcriptional regulator, pentalenolactone, 86% (*Streptomyces avermitilis*)	BA000030
12740	*alpT*	malonyl-CoA ACP transacylase	acyltransferase/oxidoreductase, chivosazole, 50% (*Sorangium cellulosum*)	DQ065771
12745	*alpE*	3-oxoacyl-ACP synthase 3	3-oxoacyl-ACP synthase, cosmomycin D, 48% (*Streptomyces olindensis*)	JJOHO1000002
12750	*alpA3*	PKS: KS, AT, ACP; KS, DH, KR, ACP	polyketide synthase, kijamimicin 48% (*Actinomadura kijaniata*)	EU301739
12755	*alpA2*	PKS: KS, ACP, cMT, TE	polyketide synthase, bacillaene, 38% (Bacillus amyloliquefaciens)	AJ634060
12760	*alpA1*	PKS/NRPS: C, A, PCP; KS, ACP, cMT; KS, DH, tDH, KR, ACP	OnnI, Onnamide, 44% (*Candidatus Entotheonella* sp.)	AY688304
12765	*alpD*	acyl-CoA dehydrogenase	*acyl-CoA dehydrogenase, dutomycin, 88% (*Streptomyces minoensis*)	KP710956
12770	*alpR*	ribosomal RNA large subunit methyltransferase G	methyltransferase containing protein; *Streptomyces* sp. IB2014 011-1 (100%)	-

*shown are closest homologs.

produce alpiniamides (**Figure 4**). This proves that gene cluster no. 9, designated *alp*-cluster (**Table 2**), is indeed responsible for the biosynthesis of alpiniamides.

In addition, we have cloned the 46.7 kb region of the chromosome of *Streptomyces* sp. IB2014/011-12 harboring the entire *alp* gene cluster and surrounding regions using transformation associated recombination in yeast. The cloning of the right region of *Streptomyces* sp. IB2014/011-12 chromosome and overall architecture of the cluster was verified through sequencing with MinION (Oxford Nanopore, United Kingdom). The construct was introduced into *S. lividans* TK24 and *S. albus* Del14. The recombinant strains bearing the cloned *alp*-gene cluster were found to produce alpiniamides (**Figure 5**). However, we were able to identify only alpiniamides A, B, and C in the extracts of generated strains. The lack of alpiniamide D could be caused either by a change in the behavior of the enzymatic assembly line or by the relatively low production of this minor derivative, hindering its identification. Moreover, a set of mutant constructs with deletion of *alpD*, *alpR*, and *alpE* genes were created. The resulting plasmids were introduced into *S. albus* Del14 and analyzed for production of alpiniamides. The deletion of *alpD* encoding an acyl-CoA dehydrogenase and *alpR* encoding a ribosomal RNA methyltransferase (**Table 2**) had no significant effect on the production of compounds

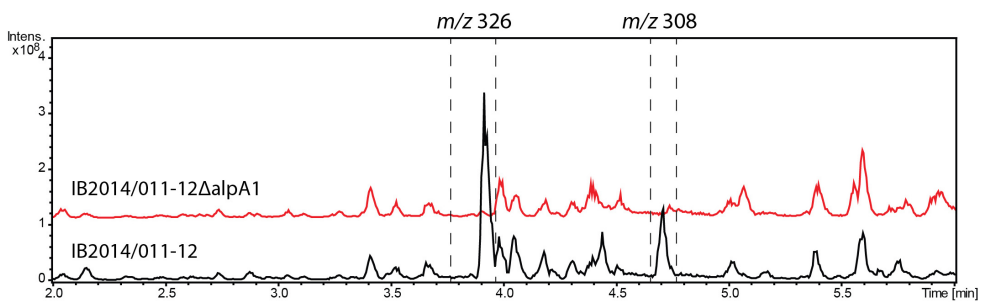


FIGURE 4 | Production of alpiniamides A-B (*m/z* 326 $[M\text{-}H_2O+H]^+$) and alpiniamide C (*m/z* 308 $[M\text{-}H_2O+H]^+$) by *Streptomyces* sp. IB2014/011-12 and its mutant strain IB2014/011-12ΔalpA1 lacking *alpA1* gene. LC-MS chromatogram RT 2–6 min out of an overall 9 min gradient run is shown. Black line – extract of wild type *Streptomyces* sp. IB2014/011-12 and red line – extract of mutant strain lacking part of *alpA1* gene and its promoter. Strains were cultivated for 7 days in NL19 medium. The peaks for alpiniamides are highlighted. Alpiniamide D elutes under these conditions together with main alpiniamides A and B.

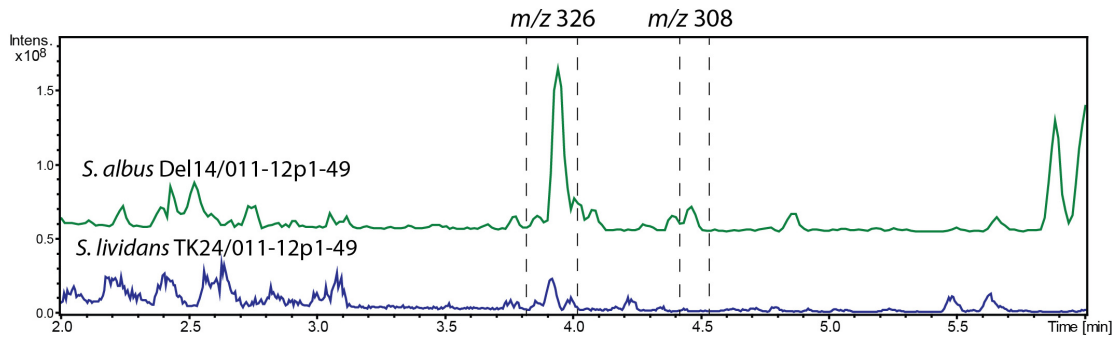


FIGURE 5 | LC-MS chromatogram RT 2–6 min out of an overall 9 min gradient run of extracts of *S. albus* Del14/011-12p1-49 and *S. lividans* TK24/011-12p1-49 strains carrying the *alp*-gene cluster construct. The strains were cultivated in NL19 media and compounds were extracted from the biomass. Peaks that correspond to alpiniamides are highlighted. Alpiniamide A-B (*m/z* 326 $[M\text{-}H_2O+H]^+$) and alpiniamide C (*m/z* 308 $[M\text{-}H_2O+H]^+$).

(**Supplementary Figure S13**). On the contrary, *S. albus* Del14 carrying the construct with deleted *alpE* gene was not able to produce alpiniamides (**Supplementary Figure S13**). Lastly, in order to determine the boundaries of *alp* gene cluster we performed the *blastn* search for the actinobacterial genome lacking the *alp* gene cluster but preserving the regions flanking the cluster in the genomes of *Streptomyces* sp. IB2014/011-12. As such, the *Kitasatospora albolonga* YIM 101047 and *Streptomyces* sp. IB2014/011-12 whole genome alignment have revealed that the *alp* gene cluster include gene between *alpA1* and *alpG*, when *alpR,D* and *alpW-I* are most probably not involved in alpiniamides biosynthesis (**Supplementary Figure S14**).

Deduction of Alpiniamide Biosynthesis

The predicted architecture of a hybrid NRPS-trans-AT-PKS mega-enzyme encoded by the *alp* gene cluster made it possible to propose the biosynthetic steps for the assembly of alpiniamides (**Figure 6B**). Polyketide synthases lacking the acyltransferase (trans-AT-PKS) were thought to be rare systems in actinomycetes. Only recently it has been discovered that they are present in several actinobacterial strains (Helfrich and Piel, 2016). However, in general, this type of PKS enzymes is thought to be the major class of polyketide biosynthesis machineries. In these systems the usual in-line acyltransferase domains are missing, and instead, free-standing enzymes take their place (Jenner et al., 2013). Apparently, the orientation and location of the core genes in the *alp*-gene cluster do not reflect the order of the biosynthetic steps in alpiniamide assembly. The organization of the *alp* gene cluster lacks the co-linearity due to its unusual domain orders and domains acting across modules, which is a well-known feature of trans-AT-PKSs (Helfrich and Piel, 2016).

The synthesis is initiated by the loading module 0, which is part of the bimodular protein encoded by gene *alpA1* (12750) (**Table 2**). Unlike the others, this module possesses its own AT domain that is predicted to be specific for methylmalonate. Since the starter unit in alpiniamide biosynthesis is thought to be a propionate, we believe that loading module 0 loads the methylmalonate and then decarboxylates it to produce propionate. Such a decarboxylative loading process is known for several *cis*-AT-PKS systems, including the ones involved in the biosynthesis of tylosin, niddamycin, pikromycin, spinosyn, and monensin (Fouces et al., 1999; Xue and Sherman, 2001). However, unlike the cis-AT systems, which typically harbor a KS^Q decarboxylative domain with the active site Cys mutated to a Gln, the Cys in the canonical CHH motif of AlpA1 KS0 is substituted with a Ser (KS^S). Similar substitutions were found in the loading KS of trans-AT oocydin synthase (Matilla et al., 2012) and several cis-AT PKSs from the nystatin, rimocidin, and pimaricin biosynthesis (Aparicio et al., 2000; Brautaset et al., 2003; Seco E.M. et al., 2004). These KS^S domains are proposed to participate in decarboxylation or synthesis of the starter substrates. At least in the case of the nystatin loading module, it was shown that the Ser residue of KS^S0 is not crucial for proper initiation of the biosynthesis (Brautaset et al., 2003). However, in all of the cases described above, the polyketide chain is initiated by acetate rather than propionate. At the same time, another scenario cannot be excluded. The *alpE* gene located close to $alpA_1$ is predicted to encode the KSIII enzyme. This enzyme is similar to the daunorubicin loading KS as well as to the ACP-shuttle-type KSs from several other anthracycline polyketides that are initiated with a propionate starter unit, such as cosmomycin D and cinerubin B (Garrido et al., 2006; Kersten et al., 2013). The high similarity between AlpE and these KSs point on alternative scenario at which the decarboxylative function of the loading module in alpiniamides assembly might be provided by the standalone decarboxylative KSIII, similar to daunorubicin/aklavinone biosynthesis (Tsukamoto et al., 1992). The essentiality of AlpE for the production of alpiniamides in heterologous hosts supports this idea.

Subsequently, two elongation steps conducted by module 1 and 2 occur and incorporate two units of malonate. The substrate is passed on to the PKS by the trans-acting acyltransferase encoded by gene 12740. This protein is a bidomain AT-ACP enzyme that is predicted to be specific for malonate. The function of the ACP domain cannot be deduced from the available data, but it harbors a Ser residue for attachment of phosphopantetheine. Module 1 possesses KR and DH domains for ketoreduction and dehydration reactions, respectively. As expected, a double bond is present in alpiniamide A, C, and D between C-4 and C-5. However, in the case of alpiniamides

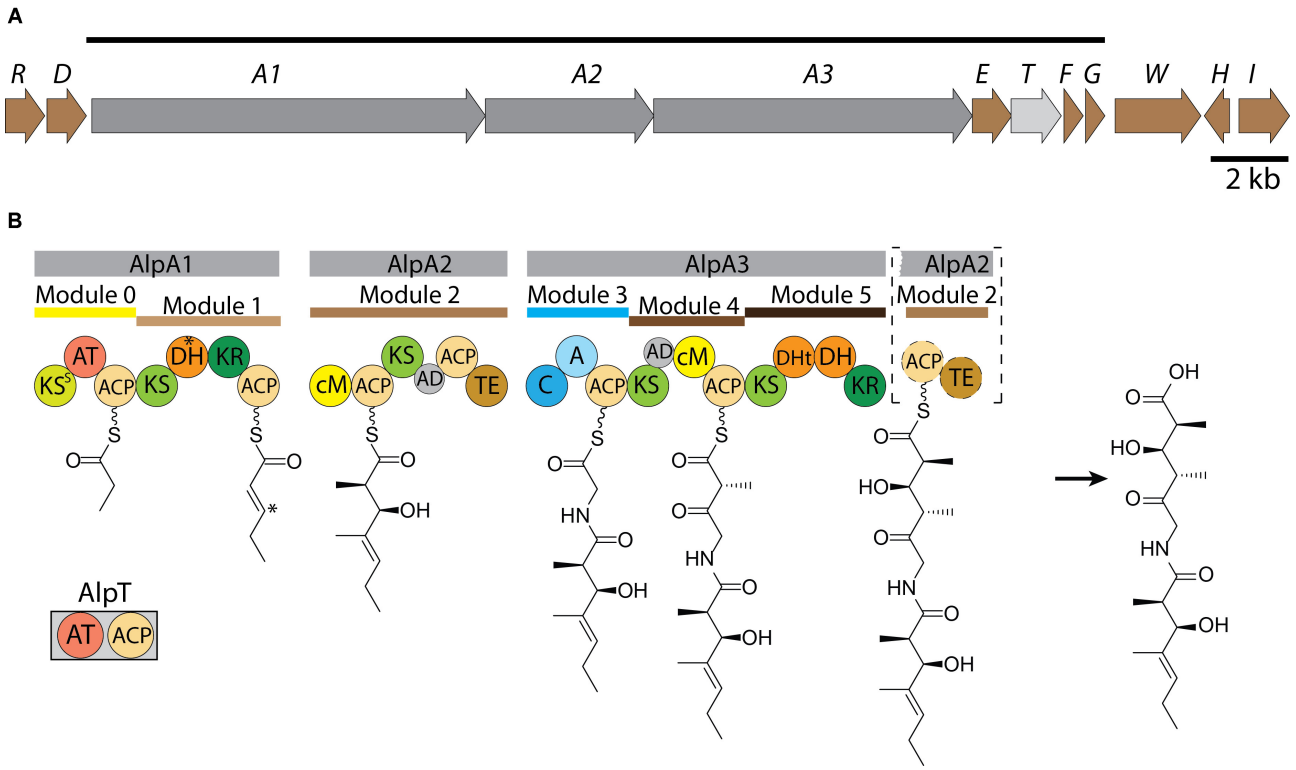


FIGURE 6 | Scheme of the organization of *alp*-gene cluster **(A)**. The line above shows the deduced boundaries of the cluster. The scale bar corresponds to 2 kb. Proposed pathway for the biosynthesis of alpiniamide **(B)**. Domains: KS - ketosynthase, AT, acyltransferase; DH, dehydratase; KR, ketoreductase; ACP, acyl carrier protein; cM, C-methyltransferase; AD, acyl-transferase docking domain; TE, thioesterase; C, condensation; A, adenylation; DHt, truncated dehydratase. *Indicates the DH presumably responsible for double bond migration and corresponding position at growing alpiniamide chain.

B_1 and B_2, the DH domain is apparently skipped, resulting in a hydroxyl group at C-5. The ketoreductase domain is missing in module 2; although, alpiniamides have the hydroxyl group at the corresponding position. A possible explanation for this would be that the activity of the KR domain from module 1 is shared between modules 1 and 2. Interestingly B_1 and B_2 differ in the stereochemistry at C-5 position. Ketoreductases are highly stereospecific and are divided into groups depending on which stereochemistry they provide (Keatinge-Clay, 2016). The *in silico* analysis of module 1 KR place it in B1 stereospecificity type performing D-α-reduction that correspond to the configuration observed in alpiniamide B_2 (Kitsche and Kalesse, 2013). Therefore, it is unclear how the different stereochemistry at position C-5 as well as the *threo* configuration at C-2/C-3 is generated. Module 2 has a C-methyltransferase domain. The acyltransferases in trans-AT-PKS are known to solely utilize malonyl-CoA as opposed to branched substrates as elongation units, and the side chain methyl groups are typically incorporated by C-methyltransferases. This seems to also be true in the case of the alpiniamides. From the labeled methionine feeding experiment, we clearly see that the methyl groups of the alpiniamides are introduced by the S-adenosyl-methionine (SAM) dependent methyltransferases present in module 2 and 4. Since these compounds each contain four methyl groups, we assume that both methyltransferases need to act twice during the biosynthesis. Thus, modules 2 and 4 provide the C-methyltransferase activity also to module 1 and 5. Another key feature of module 2 is the presence of a thioesterase domain. This TE is predicted to be a class II TE of the α/β hydrolase family. These enzymes hydrolyze the residues attached to the phosphopantetheine group of ACPs. The exact function of the TE domain of AlpA2 is not clear, especially considering its location. A similar architecture was observed in the case of the burkholderic acid trans-AT PKS, which harbors internal TE domains with unknown functions (Franke et al., 2012). TE II domains are often not essential for biosynthesis, but can be involved in removing aberrant substrates from stalled PKS megasynthases and control the loading of the starter unit (Kotowska and Pawlik, 2014). The TE domain of AlpA2 has the Cys in place of Ser residue of catalytic triad Ser/Asp/His (**Supplementary Figure S15**). However, such substitutions are not rare and can be found in TE14 family of thioesterases (Cantu et al., 2010).

Another interesting feature of module 2 is the possible migration of the double bond. The double bond in alpiniamides B_1 and B_2 is shifted from the canonical position between C-2 and C-3 to the position between C-3 and C-4. A so-called double bond migration is described for several trans-AT produced polyketides such as bacillaene, corallopyronin A, and rhizoxin (Butcher et al., 2007; Partida-Martinez and Hertweck, 2007; Lohr et al., 2013),

and this process is thought to be mediated by a dehydratase domain either by direct β-γ dehydration or α-β dehydration and subsequent isomerization (Gay et al., 2014). However, it is not clear if alpiniamides B_1 and B_2 arise from alpiniamide A through the catalytic action of one of the DH domains (most likely by the DHt of module 5 that is predicted to be non-functional) or from an acid-induced double bond migration.

At the next step, a non-ribosomal peptide synthetase module within $AlpA_3$ extend the polyketide intermediate with a glycine. Interestingly, glycine is often found to act as a bridge between two polyketide chains in hybrid NRPS-trans-AT-PKS assembled natural products, such as bacillaene, batumin, calyculin, corallopyronin, misakinolide and oxazolomycin (Helfrich and Piel, 2016).

Modules 4 and 5 further extend the molecule with two malonate units. Typically, ketosynthases ensuing a NRPS module are highly substrate specific and do rarely accept aberrant substrates with bulkier amino acid residues (Kohlhaas et al., 2013). As in module 2, module 4 lacks the ketoreductive function even though the hydroxy group at C-3′ is present. This activity is most likely provided by module 5, which contains KR and DH domains. However, based on the structures of alpiniamide A, B and D, we assume that the DH domain is occasionally skipped. Only alpiniamide C seems to be dehydrated at this step, which resulted in the double bond between C-3′ and C-4′. Module 5 is missing the ACP and TE domains, which are instead located in module 2. The mechanism by which alpiniamides are released from the biosynthetic enzyme is not clear. Most likely, the TE activity is provided by the respective domain of module 2. However, the location of the TE domain is unusual as well as the composition of catalytic triad (Cys/Asp/His). The fact that it is placed right before the NRPS module 3 makes us think that it also might be involved in the cleavage of aberrant substrates than in the release of the final product.

In the case of alpiniamide D, module 2 is probably skipped entirely during the biosynthesis since it lacks the C-1 and C-2 part of the molecule. Such a case is very rare but was previously described for epothilone K (Moss et al., 2004) and for the trans-AT-PKS-derived compound albicidin (Huang et al., 2001).

Even though alpiniamides are quite small molecules they provide a fascinating example of high chemical diversity introduced by the PKS assembly line. The alpiniamides NRPS-trans-AT-PKS utilizes a combination of domain and module skipping, in-line methylation and double bond migration events to produce a variety of chemical structures from a simple initial building block.

CONCLUSION

We have isolated five new alpiniamide derivatives and identified the gene cluster responsible for their biosynthesis in the new *Streptomyces* sp. IB2014/011-12. This shows that there are still great opportunities to discover promising natural products, especially from unique and largely unexplored ecological niches such as Lake Baikal.

AUTHOR CONTRIBUTIONS

CP, YR, and JZ performed the experiments and analyzed the data. CR and JK sequenced and assembled the genome. AL conceived and designed the experiments. All authors participated in the manuscript preparation and discussion.

FUNDING

The research leading to these results has received funding from the European Commission's Seventh Framework Program (FP7/2007–2013, KBBE.2013.3.6-02: Synthetic Biology toward applications) under the grant agreement STREPSYNTH (Project No. 613877).

ACKNOWLEDGMENTS

We are grateful to Dr. S. Wenzel and Dr. B. Tokovenko for useful discussions.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2018.01959/full#supplementary-material

REFERENCES

Allam, N. G. (2012). Protective role of *Aspergillus fumigatus* melanin against ultraviolet (UV) irradiation and *Bjerkandera adusta* melanin as a candidate vaccine against systemic candidiasis. *Afr. J. Biotechnol.* 11, 6566–6577. doi: 10.5897/AJB11.4136

Aparicio, J. F., Fouces, R., Mendes, M. V., Olivera, N., and Martín, J. F. (2000). A complex multienzyme system encoded by five polyketide synthase genes is involved in the biosynthesis of the 26-membered polyene macrolide pimaricin in *Streptomyces natalensis*. *Chem. Biol.* 7, 895–905. doi: 10.1016/S1074-5521(00)00038-7

Axenov-Gribanov, D., Rebets, Y., Tokovenko, B., Voytsekhovskaya, I., Timofeyev, M., and Luzhetskyy, A. (2016). The isolation and characterization of actinobacteria from dominant benthic macroinvertebrates endemic to Lake Baikal. *Folia Microbiol.* 61, 159–168. doi: 10.1007/s12223-015-0421-z

Barona-Gómez, F., Wong, U., Giannakopulos, A. E., Derrick, P. J., and Challis, G. L. (2004). Identification of a cluster of genes that directs desferrioxamine biosynthesis in *Streptomyces coelicolor* M145. *J. Am. Chem. Soc.* 126, 16282–16283. doi: 10.1021/ja045774k

Bifulco, G., Dambruoso, P., Gomez-Paloma, L., and Riccio, R. (2007). Determination of relative configuration in organic compounds by NMR spectroscopy and computational methods. *Chem. Rev.* 107, 3744–3779. doi: 10.1021/cr030733c

Bilyk, O., Sekurova, O. N., Zotchev, S. B., and Luzhetskyy, A. (2016). Cloning and heterologous expression of the grecocycline biosynthetic gene cluster. *PLoS One* 11:e0158682. doi: 10.1371/journal.pone.0158682

Blodgett, J. A. V., Thomas, P. M., Li, G., Velasquez, J. E., van der Donk, W. A., Kelleher, N. L., et al. (2007). Unusual transformations in the biosynthesis of the antibiotic phosphinothricin tripeptide. *Nat. Chem. Biol.* 3, 480–485. doi: 10.1038/nchembio.2007.9

Brautaset, T., Borgos, S. E. F., Sletta, H., Ellingsen, T. E., and Zotchev, S. B. (2003). Site-specific mutagenesis and domain substitutions in the loading module of the nystatin polyketide synthase, and their effects on nystatin biosynthesis in *Streptomyces noursei*. *J. Biol. Chem.* 278, 14913–14919. doi: 10.1074/jbc.M212611200

Butcher, R. A., Schroeder, F. C., Fischbach, M. A., Straight, P. D., Kolter, R., Walsh, C. T., et al. (2007). The identification of bacillaene, the product of the PksX megacomplex in *Bacillus subtilis*. *Proc. Natl. Acad. Sci. U.S.A.* 104, 1506–1509. doi: 10.1073/pnas.0610503104

Cantu, D. C., Chen, Y., and Reilly, P. J. (2010). Thioesterases: a new perspective based on their primary and tertiary structures. *Protein Sci.* 19, 1281–1295. doi: 10.1002/pro.417

Cao, S., Blodgett, J. A. V., and Clardy, J. (2010). Targeted discovery of polycyclic tetramate macrolactams from an environmental *Streptomyces* strain. *Org. Lett.* 12, 4652–4654. doi: 10.1021/ol1020064

Dale, J. A., and Mosher, H. S. (1973). Nuclear magnetic resonance enantiomer regents. Configurational correlations via nuclear magnetic resonance chemical shifts of diastereomeric mandelate, O-methylmandelate, and.alpha.-methoxy-.alpha.-trifluoromethylphenylacetate (MTPA) esters. *J. Am. Chem. Soc.* 95, 512–519. doi: 10.1021/ja00783a034

Dias, D. A., Urban, S., and Roessner, U. (2012). A historical overview of natural products in drug discovery. *Metabolites* 2, 303–336. doi: 10.3390/metabo2020303

Ding, Y., Li, Y., Li, Z., Zhang, J., Lu, C., Wang, H., et al. (2016). Alteramide B is a microtubule antagonist of inhibiting *Candida albicans*. *Biochim. Biophys. Acta* 1860, 2097–2106. doi: 10.1016/j.bbagen.2016.06.025

Fouces, R., Mellado, E., Díez, B., and Barredo, J. L. (1999). The tylosin biosynthetic cluster from *Streptomyces fradiae*: genetic organization of the left region. *Microbiology* 145(Pt 4), 855–868. doi: 10.1099/13500872-145-4-855

Franke, J., Ishida, K., and Hertweck, C. (2012). Genomics-driven discovery of burkholderic acid, a noncanonical, cryptic polyketide from human pathogenic *Burkholderia* species. *Angew. Chem.* 51, 11611–11615. doi: 10.1002/anie.201205566

Garrido, L. M., Lombó, F., Baig, I., Nur-E-Alam, M., Furlan, R. L. A., Borda, C. C., et al. (2006). Insights in the glycosylation steps during biosynthesis of the antitumor anthracycline cosmomycin: characterization of two glycosyltransferase genes. *Appl. Microbiol. Biotechnol.* 73, 122–131. doi: 10.1007/s00253-006-0453-z

Gay, D. C., Spear, P. J., and Keatinge-Clay, A. T. (2014). A double-hotdog with a new trick: structure and mechanism of the trans-acyltransferase polyketide synthase enoyl-isomerase. *ACS Chem. Biol.* 9, 2374–2381. doi: 10.1021/cb500459b

Gietz, R. D., and Schiestl, R. H. (2007). High-efficiency yeast transformation using the LiAc/SS carrier DNA/PEG method. *Nat. Protoc.* 2, 31–34. doi: 10.1038/nprot.2007.13

Golinska, P., Wypij, M., Agarkar, G., Rathod, D., Dahm, H., and Rai, M. (2015). Endophytic actinobacteria of medicinal plants: diversity and bioactivity. *Antonie Van Leeuwenhoek* 108, 267–289. doi: 10.1007/s10482-015-0502-7

Green, M. R., and Sambrook, J. (2012). *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press.

Gust, B., Chandra, G., Jakimowicz, D., Yuqing, T., Bruton, C. J., and Chater, K. F. (2004). λ *Red-Mediated Genetic Manipulation of Antibiotic-Producing Streptomyces*. New York, NY: Elsevier, 107–128. doi: 10.1016/S0065-2164(04)54004-2

Hasani, A., Kariminik, A., and Issazadeh, K. (2014). Streptomycetes: characteristics and their antimicrobial activities. *IJABBR* 2, 63–75.

Helfrich, E. J. N., and Piel, J. (2016). Biosynthesis of polyketides by trans-AT polyketide synthases. *Nat. Prod. Rep.* 33, 231–316. doi: 10.1039/c5np00125k

House, H. O., Crumrine, D. S., Teranishi, A. Y., and Olmstead, H. D. (1973). Chemistry of carbanions. XXIII. Use of metal complexes to control the aldol condensation. *J. Am. Chem. Soc.* 95, 3310–3324. doi: 10.1021/ja00791a039

Hoye, T. R., Jeffrey, C. S., and Shao, F. (2007). Mosher ester analysis for the determination of absolute configuration of stereogenic (chiral) carbinol carbons. *Nat. Protoc.* 2, 2451–2458. doi: 10.1038/nprot.2007.354

Huang, G., Zhang, L., and Birch, R. G. (2001). A multifunctional polyketide-peptide synthetase essential for albicidin biosynthesis in *Xanthomonas albilineans*. *Microbiology* 147, 631–642. doi: 10.1099/00221287-147-3-631

Jakobi, M., Winkelmann, G., Kaiser, D., Kempter, C., Jung, G., Berg, G., et al. (1996). Maltophilin: a new antifungal compound produced by *Stenotrophomonas maltophilia* R3089. *J. Antibiot.* 49, 1101–1104. doi: 10.7164/antibiotics.49.1101

Jenner, M., Frank, S., Kampa, A., Kohlhaas, C., Pöplau, P., Briggs, G. S., et al. (2013). Substrate specificity in ketosynthase domains from trans-AT polyketide synthases. *Angew. Chem.* 52, 1143–1147. doi: 10.1002/anie.201207690

Keatinge-Clay, A. T. (2016). Stereocontrol within polyketide assembly lines. *Nat. Prod. Rep.* 33, 141–149. doi: 10.1039/c5np00092k

Kersten, R. D., Ziemert, N., Gonzalez, D. J., Duggan, B. M., Nizet, V., Dorrestein, P. C., et al. (2013). Glycogenomics as a mass spectrometry-guided genome-mining method for microbial glycosylated molecules. *Proc. Natl. Acad. Sci. U.S.A.* 110, E4407–E4416. doi: 10.1073/pnas.1315492110

Khosla, C. (2009). Structures and mechanisms of polyketide synthases. *J. Org. Chem.* 74, 6416–6420. doi: 10.1021/jo9012089

Kieser, T. (2000). *Practical Streptomyces Genetics*. Norwich: Innes.

Kitsche, A., and Kalesse, M. (2013). Configurational assignment of secondary hydroxyl groups and methyl branches in polyketide natural products through bioinformatic analysis of the ketoreductase domain. *Chembiochem* 14, 851–861. doi: 10.1002/cbic.201300063

Kohlhaas, C., Jenner, M., Kampa, A., Briggs, G. S., Afonso, J. P., Piel, J., et al. (2013). Amino acid-accepting ketosynthase domain from a trans-AT polyketide synthase exhibits high selectivity for predicted intermediate. *Chem. Sci.* 4, 3212–3217. doi: 10.1039/c3sc50540e

Kotowska, M., and Pawlik, K. (2014). Roles of type II thioesterases and their application for secondary metabolite yield improvement. *Appl. Microbiol. Biotechnol.* 98, 7735–7746. doi: 10.1007/s00253-014-5952-8

Latypov, S. K., Seco, J. M., Quiñoá, E., and Riguera, R. (1996). MTPA vs MPA in the determination of the absolute configuration of chiral alcohols by 1 H NMR. *J. Org. Chem.* 61, 8569–8577. doi: 10.1021/jo960719i

Li, S., Jochum, C. C., Yu, F., Zaleta-Rivera, K., Du, L., Harris, S. D., et al. (2008). An antibiotic complex from *Lysobacter enzymogenes* strain C3: antimicrobial activity and role in plant disease control. *Phytopathology* 98, 695–701. doi: 10.1094/PHYTO-98-6-0695

Lohr, F., Jenniches, I., Frizler, M., Meehan, M. J., Sylvester, M., Schmitz, A., et al. (2013). $\alpha,\beta \rightarrow \beta,\gamma$ double bond migration in corallopyronin A biosynthesis. *Chem. Sci.* 4, 4175–4180. doi: 10.1039/C3SC51854J

Lou, L., Qian, G., Xie, Y., Hang, J., Chen, H., Zaleta-Rivera, K., et al. (2011). Biosynthesis of HSAF, a tetramic acid-containing macrolactam from *Lysobacter enzymogenes*. *J. Am. Chem. Soc.* 133, 643–645. doi: 10.1021/ja105732c

Makar, A. B., McMartin, K. E., Palese, M., and Tephly, T. R. (1975). Formate assay in body fluids: application in methanol poisoning. *Biochem. Med.* 13, 117–126. doi: 10.1016/0006-2944(75)90147-7

Mansjö, M., and Johansson, J. (2011). The riboflavin analog roseoflavin targets an FMN-riboswitch and blocks *Listeria monocytogenes* growth, but also stimulates virulence gene-expression and infection. *RNA Biol.* 8, 674–680. doi: 10.4161/rna.8.4.15586

Matilla, M. A., Stöckmann, H., Leeper, F. J., and Salmond, G. P. C. (2012). Bacterial biosynthetic gene clusters encoding the anti-cancer haterumalide class of molecules: biogenesis of the broad spectrum antifungal and anti-oomycete compound, oocydin A. *J. Biol. Chem.* 287, 39125–39138. doi: 10.1074/jbc.M112.401026

Matsumori, N., Kaneno, D., Murata, M., Nakamura, H., and Tachibana, K. (1999). Stereochemical determination of acyclic structures based on carbon-proton spin-coupling constants. A method of configuration analysis for natural products. *J. Org. Chem.* 64, 866–876. doi: 10.1021/jo981810k

Meyer, F., Goesmann, A., McHardy, A. C., Bartels, D., Bekel, T., Clausen, J., et al. (2003). GenDB–an open source genome annotation system for prokaryote genomes. *Nucleic Acids Res.* 31, 2187–2195. doi: 10.1093/nar/gkg312

Moss, S. J., Martin, C. J., and Wilkinson, B. (2004). Loss of co-linearity by modular polyketide synthases: a mechanism for the evolution of chemical diversity. *Nat. Prod. Rep.* 21, 575–593. doi: 10.1039/b315020h

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. doi: 10.1128/AEM.00434-11

Palecková, P., Bobek, J., and Mikulík, K. (2009). tmRNA of *Streptomyces collinus* and *Streptomyces griseus* during the growth and in the presence of antibiotics. *Microb. Biotechnol.* 2, 114–122. doi: 10.1111/j.1751-7915.2008.00066.x

Partida-Martinez, L. P., and Hertweck, C. (2007). A gene cluster encoding rhizoxin biosynthesis in "*Burkholderia rhizoxina*", the bacterial endosymbiont of the fungus *Rhizopus microsporus*. *Chembiochem* 8, 41–45. doi: 10.1002/cbic.200600393

Procópio, R. E., Silva, I. R., Martins, M. K., Azevedo, J. L., and Araújo, J. M. (2012). Antibiotics produced by Streptomyces. *Braz. J. Infect. Dis.* 16, 466–471. doi: 10.1016/j.bjid.2012.08.014

Running, W. (1993). Computer software reviews. Chapman and hall dictionary of natural products on CD-ROM. *J. Chem. Inf. Model.* 33, 934–935. doi: 10.1021/ci00016a603

Schwarz, J., Konjik, V., Jankowitsch, F., Sandhoff, R., and Mack, M. (2016). Identification of the key enzyme of roseoflavin biosynthesis. *Angew. Chem.* 55, 6103–6106. doi: 10.1002/anie.201600581

Seco, E. M., Pérez-Zúñiga, F. J., Rolón, M. S., and Malpartida, F. (2004). Starter unit choice determines the production of two tetraene macrolides, rimocidin and CE-108, in Streptomyces diastaticus var. 108. *Chem. Biol.* 11, 357–366. doi: 10.1016/j.chembiol.2004.02.017

Seco, J. M., Quiñoá, E., and Riguera, R. (2004). The assignment of absolute configuration by NMR †. *Chem. Rev.* 104, 17–118. doi: 10.1021/cr000665j

Seemann, T. (2014). Prokka: rapid prokaryotic genome annotation. *Bioinformatics* 30, 2068–2069. doi: 10.1093/bioinformatics/btu153

Shigemori, H., Bae, M. A., Yazawa, K., Sasaki, T., and Kobayashi, J. (1992). Alteramide A, a new tetracyclic alkaloid from a bacterium *Alteromonas* sp. associated with the marine sponge *Halichondria okadai*. *J. Org. Chem.* 57, 4317–4320. doi: 10.1021/jo00041a053

Stiles, M., Winkler, R. R., Chang, Y.-L., and Traynor, L. (1964). Stereochemical assignments for β-Ketols formed by aldol addition of three simple ketones to p -nitrobenzaldehyde. *J. Am. Chem. Soc.* 86, 3337–3342. doi: 10.1021/ja01070a027

Tsukamoto, N., Fujii, I., Ebizuka, Y., and Sankawa, U. (1992). Cloning of aklavinone biosynthesis genes from *Streptomyces galilaeus*. *J. Antibiot.* 45, 1286–1294. doi: 10.7164/antibiotics.45.1286

Ueda, K., Oinuma, K.-I., Ikeda, G., Hosono, K., Ohnishi, Y., Horinouchi, S., et al. (2002). AmfS, an extracellular peptidic morphogen in *Streptomyces griseus*. *J. Bacteriol.* 184, 1488–1492. doi: 10.1128/JB.184.5.1488-1492.2002

Vaishnav, P., and Demain, A. L. (2011). Unexpected applications of secondary metabolites. *Biotechnol. Adv.* 29, 223–229. doi: 10.1016/j.biotechadv.2010.11.006

Ventola, C. L. (2015). The antibiotic resistance crisis: part 1: causes and threats. *Pharm. Ther.* 40, 277–283.

Weber, T., Blin, K., Duddela, S., Krug, D., Kim, H. U., Bruccoleri, R., et al. (2015). antiSMASH 3.0-a comprehensive resource for the genome mining of biosynthetic gene clusters. *Nucleic Acids Res.* 43, W237–W243. doi: 10.1093/nar/gkv437

Whittle, M., Willett, P., Klaffke, W., and van Noort, P. (2003). Evaluation of similarity measures for searching the dictionary of natural products database. *J. Chem. Inf. Comput. Sci.* 43, 449–457. doi: 10.1021/ci025591m

Winston, F., Dollard, C., and Ricupero-Hovasse, S. L. (1995). Construction of a set of convenient *Saccharomyces cerevisiae* strains that are isogenic to S288C. *Yeast* 11, 53–55. doi: 10.1002/yea.320110107

Woo, P. C. Y., Tam, E. W. T., Chong, K. T. K., Cai, J. J., Tung, E. T. K., Ngan, A. H. Y., et al. (2010). High diversity of polyketide synthase genes and the melanin biosynthesis gene cluster in *Penicillium marneffei*. *FEBS J.* 277, 3750–3758. doi: 10.1111/j.1742-4658.2010.07776.x

Xu, K., Yang, P.-F., Yang, Y.-N., Feng, Z.-M., Jiang, J.-S., and Zhang, P.-C. (2017). Direct assignment of the *Threo* and *Erythro* configurations in polyacetylene glycosides by 1H NMR spectroscopy. *Org. Lett.* 19, 686–689. doi: 10.1021/acs.orglett.6b03855

Xu, L., Wu, P., Wright, S. J., Du, L., and Wei, X. (2015). Bioactive polycyclic tetramate macrolactams from *Lysobacter enzymogenes* and their absolute configurations by theoretical ECD calculations. *J. Nat. Prod.* 78, 1841–1847. doi: 10.1021/acs.jnatprod.5b00099

Xue, Y., and Sherman, D. H. (2001). Biosynthesis and combinatorial biosynthesis of pikromycin-related macrolides in *Streptomyces venezuelae*. *Metab. Eng.* 3, 15–26. doi: 10.1006/mben.2000.0167

Zhou, H., Yang, Y., Zhang, J., Peng, T., Zhao, L., Xu, L., et al. (2013). Alkaloids from an endophytic *Streptomyces* sp. YIM66017. *Nat. Prod. Commun.* 8, 1393–1396.

Zhu, D., Liu, J., Han, R., Shen, G., Long, Q., Wei, X., et al. (2014). Identification and characterization of ectoine biosynthesis genes and heterologous expression of the ectABC gene cluster from *Halomonas* sp. QHL1, a moderately halophilic bacterium isolated from Qinghai Lake. *J. Microbiol.* 52, 139–147. doi: 10.1007/s12275-014-3389-5

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The reviewer YL and handling Editor declared their shared affiliation.

frontiers

Supplementary Material

New Alpiniamides from *Streptomycetes sp.* 11-12 assembled by an unusual hybrid-nonribosomal peptide synthetase-transAT-polyketidesynthase enzyme

Constanze Paulus [1], Yuriy Rebets [2], Josef Zapp[2], Christian Rückert[3], Jörn Kalinowski[3], Andriy Luzhetskyy [1, 2,*]

[1] Helmholtz-Institute for Pharmaceutical Research, Saarbrücken, Germany

[2] Department for pharmaceutical Biotechnology, University of Saarland, Saarbrücken, Germany

[3] Center for Biotechnology (CeBiTec), Bielefeld University, Bielefeld, Germany

1 Supplementary Data

Supplementary Material should be uploaded separately on submission. Please include any supplementary data, figures and/or tables.

2 Supplementary Figures and Tables

2.1 Supplementary tables

Table S1. Primer used in this work.

Name	sequence
11-12_12720ChF	TCGACAGACAGCTGCCCTG
11-12_12720ChR	GCACACAGTAGAACGTGTTG
11-12_12715ChF	ACCCGAAGTTCGTGGTCCTG
11-12_12715ChR	CGGACCGTTGCGTGTTCACA
11-12_12695ChF	CACGCTGGTGTTCGACTACC
11-12_12695ChR	AGAGCTCTTCGATGGGGTAG
11-12_12720DF	TGGGGCGCCCCCGGTTCCTCGGTTAGCCTGAGTGACATATGAAACTACGCCCCCAACTGAGAG
11-12_12720DR	GGAAGGGCCGGGAACCCTGATCTCCGCCGTCAGCGCTTGACTCGACCCGGTACCGGAGTA
11-12_12715DF	CACGGGCGACGCAGACGCGAGACTCTGGAGGGGCCGTGCAACTACGCCCCCAACTGAGAG
11-12_12715DR	GACGCGTCGCGCCACCGCGAGCCAGCGCCTCACCGCCGCACTCGACCCGGTACCGGAGTA
11-12_12695DF	TGAGAGGGAACGAGAAGTGAAGACAGAGGCCCTGTTCATCGACTACGCCCCCAACTGAGAG
11-12_12695DR	CCCTGTCCCGGGAACACGTAGCAGCGCATCAGAGCCGCCCGCTCGACCCGGTACCGGAGTA
C9-2REcV	atGATATCATCCAGGAGATGATCTCGCG
C9-2FEcI	atGAATTCAATCTGGACCGCAACGGCCT
C9-1FEcV	atGATATCG CCTGAAGGACCAACTCGT
C9-1RXba	atTCTAGAACTTCGATGGGAT CGCCGA
11-12DelCheckF	AAGTGCTGGTCGAGGACGTA
11-12DelCheckR	TACGCCTGGCTGATGG AGAT
11-12C9TarF1Not	ATGCGGCCGCAACGGCGTGAAGCCGAAGAT
11-12C9TarR1Nhe	ATGCTAGCGAATGCTGCTCATGGGGTGT
11-12C9TarF2NheI	ATGCTAGCCTGCACACGATCATCAGCGA
11-12C9TarR2HindIII	ATAAGCTTGATTGGACGGAATGTCCGTG
11-12C9CheckHind	GACGAC ACGTACTACCAGGA
pCLYCheckHind	CTCTTCGGTGAACTTCGGG

Sequence highlighted in yellow anneal with the priming sites of hygromycin resistance cassette from patt-shyg-oriT.

Table S2. Secondary metabolism gene clusters identified within the genome of *Streptomyces sp.* IB2014/011-12 with the use of the antiSMASH software.

BGC	Type	Most similar known cluster.
1	Lantipeptide	AmfS antibiotic (Ueda et al., 2002)
2	Terpene	-
3	Lassopeptide	Lassopetide similar to SRO15-2005 (Kersten et al., 2011)
4	Lantipeptide	-
5	Siderophore	Desferrioxamine B (Barona-Gómez et al., 2004)
6	Bacteriocin	-
7	Butyrolactone	-
8	Terpene	-
9	NRPS-Trans-at-type 1-PKS	-
10	Melanin-NRPS	Peptide similar to Coelichelin (Challis and Ravel, 2000)
11	Type 3 PKS	-
12	NRPS	-
13	type 1 PKS-NRPS	Similar to polycyclic tetramate macrolactams (Luo et al., 2013)
14	Terpene	Squalene antibiotic similar to Hopene (Pan et al., 2015)
16	Type 1 PKS	-
17	Other	Riboflavin antibiotic similar to Roseoflavin (Schwarz et al., 2016)
18	Terpene	-
19	thiopeptide	-
20	Melanin	Phenoxazinone antibiotic Grixazone (Suzuki et al., 2007)
21	Ectoine	Ectoine (Zhu et al., 2014)
22	NRPS	-
23	Type 3 PKS	Alkylresorcinol (Cook et al., 2010)
24	Melanin	Dihydroxynaphthalene Melanin (Woo et al., 2010)
25	NRPS	-
26	Siderophore	-
27	Type 2 PKS	-
28	Terpene	-
29	NRPS	-

Table S3. NMR spectroscopic data ($MeOD_4$) for alpiniamides 1-5.

Position	(1)		(1)[1]		(2)		(3)		(4)		(5)	
	δ_C [a]	δ_H [b]	δ_C [a]	**δ_H [b]**	**δ_C [a]**	δ_H [b]	δ_C [a]	δ_H [b]	**δ_C [a]**	**δ_H [b]**	**δ_C [a]**	**δ_H [b]**
1	178.3, C		176.75, C		178.0, C		177.2, C		178.2, C		172.2, C	
2	45.1, CH	2.54, dq (9.5, 7.0)	44.28, CH	2.52, dq (9.5, 7.0)	40.4, CH	3.36, dq (9.5, 7.0)	40.5, CH	3.36, dq (9.5, 7.0)	45.5, CH	2.56, dq (9.5, 7.1)	130.7, C	
3	81.0, CH	3.98, d (9.5)	80.40, CH	4.03, d (9.5)	126.6, CH	5.43, br d (9.3)	127.2, CH	5.43, br d (9.3)	81.0, CH	3.98, d (9.5)	139.4, CH	6.39, tq (7.2, 1.5)
4	135.1, C		133.13, C		140.6, C		139.8, C		135.0, C		22.3, C	2.20, quint (7.5, 1.0)
5	131.8, CH	5.41, t (7.2)	131.94, CH	5.40, t (7.2)	79.3, CH	3.85, t (6.7)	79.8, CH	3.85, t (6.8)	132.0, CH	5.42, t (7.3)	13.2, CH_3	1.05, t (7.5)
6	21.5, CH_2	2.06, quint (7.5)	20.85, CH_2	2.02, quint (7.5)	28.5, CH_2	1.54, m	28.1, CH_2	1.55, m	21.5, CH_2	2.06, quint (7.5)	12.2, CH_3	1.85, br s
7	14.1, CH_3	0.97, t (7.5)	13.88, CH_3	0.94, t (7.5)	10.1, CH_3	0.85, t (7.5)	10.3, CH_3	0.85, t (7.5)	14.2, CH_3	0.97, t (7.5)		
8	14.8, CH_3	0.93, d (7.0)	14.38, CH_3	0.95, d (7.0)	18.5, CH_3	1.22, d (7.0)	18.6, CH_3	1.20, d (7.0)	14.7, CH_3	0.94, d (7.0)		
9	10.3, CH_3	1.60, s	10.41, CH_3	1.57, s	11.6, CH_3	1.66, d (1.5)	11.1, CH_3	1.66, d (1.5)	10.5, CH_3	1.61, s		
1’	178.4, C		179.03, C		178.7, C		178.3, C		180.8, C		178.1, C	
2’	44.6, CH	2.50, quint (7.0)	42.71, CH	2.54, quint (7.0)	44.60, CH	2.48, quint (7.0)	44.7, CH	2.47, quint (7.0)	44.3, CH	3.41, dq (9.5, 7.1)	44.4, CH	2.55, quint (7.1)
3’	74.2, CH	4.01, ddd (7.0, 5.3, 1.5)	73.54; CH	4.02, m	74.4, CH	3.96, ddd (7.0, 5.3, 1.5)	74.4, CH	3.95, ddd (7.0, 5.3, 1.5)	146.7, CH	6.87, dq (9.5, 1.5)	74.1, CH	4.03, dd (7.1, 5.3)
4’	47.7, CH	2.92, qd (7.0, 5.3)	47.23, CH	2.79, d (7.0, 4.5)	47.9, CH	2.87, qd (7.0, 5.5)	48.1, CH	2.88, qd (7.0, 5.5)	135.0, CH		47.7, CH	2.93, qd (7.1, 5.3)
5’	209.9, C		209.27, C		209.5, C		209.1, C		197.4, C		209.9, C	
6’	49.0, CH_2	4.13, d (18.5) 4.25, d (18.5)	48.81, CH_2	4.19, dd (19.0, 5.0) 4.29, dd (19.0, 5.5)	48.8 CH_2	4.12, d (18.5) 4.16, d (18.5)	48.8, CH_2	4.13, d (18.5) 4.17, d (18.5)	46.0, CH_2	4.44, d (18.5) 4.48, d (18.5)	49.0, CH_2	4.18, d (18.5) 4.22, d (18.5)
7’	15.0, CH_3	1.16, d (7.0)	14.49, CH_3	1.17, d (7.0)	14.9, CH_3	1.15, d (7.0)	15.0, CH_3	1.15, d (7.0)	18.4, CH_3	1.30, d (7.0)	14.8, CH_3	1.15, d (7.0)
8’	10.5, CH_3	1.13, d (7.0)	9.79, CH_3	1.14, d (7.0)	10.8, CH_3	1.12, d (7.0)	10.8, CH_3	1.12, d (7.0)	11.2, CH_3	1.82, d (1.5)	10.6, CH_3	1.14, d (7.0)
NH				7.09, t (5.3)								

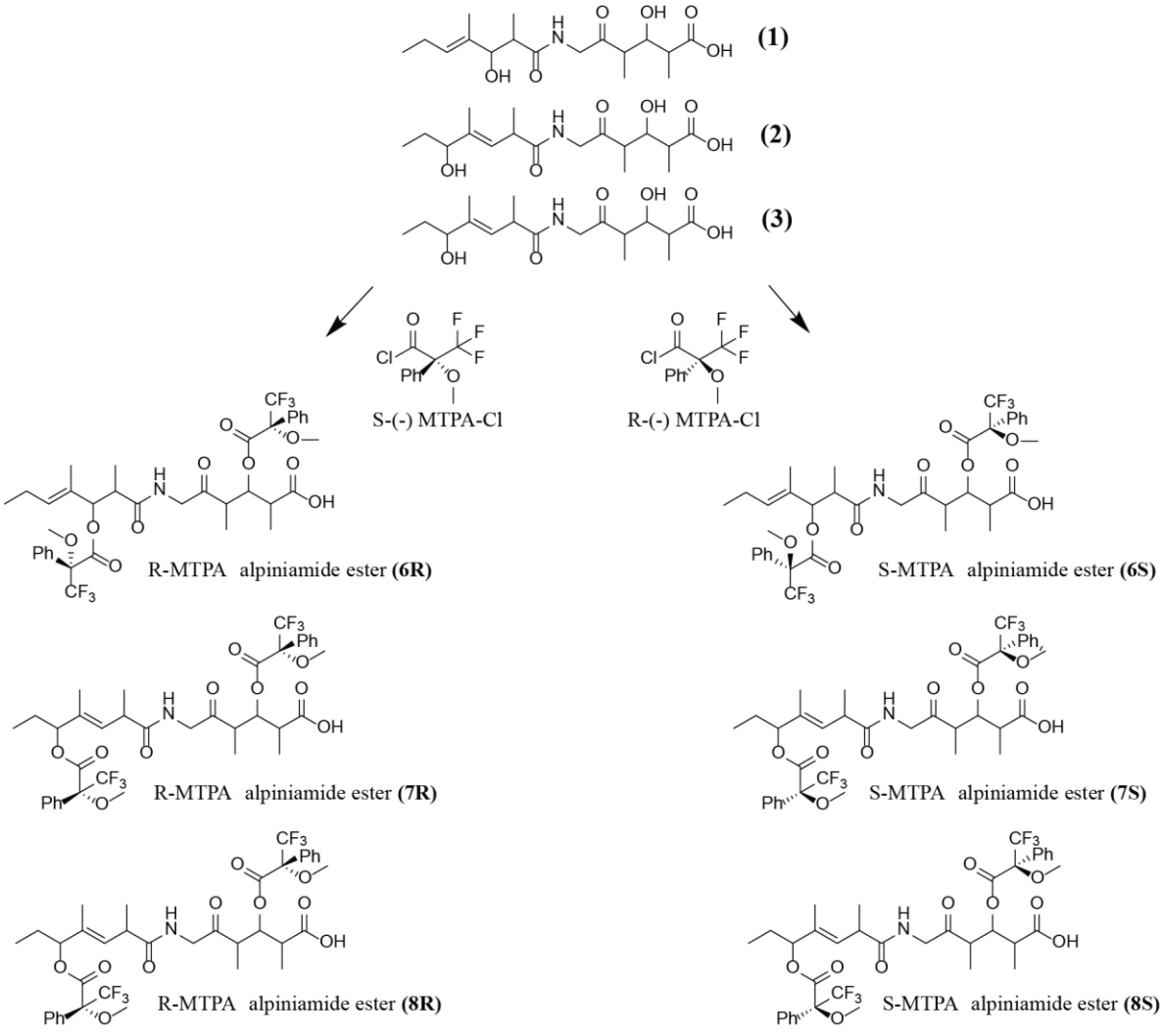
[1] = NMR data aquired in $CDCl_3$; [a] = followed by multiplicity; [b] = followed by coupling constant J in Hz; all ^{13}C chemical shifts were taken from 2D spectra HSQC/HMBC except for (1) in $CDCl_3$.

Table S4. ^{1}H chemical shifts (in ppm, recorded in $CDCl_3$) and $\Delta\delta_{(S-R)}$ values for Mosher derivatives 6R and 6S of alpiniamide A (**1**), 7R and 7S of alpiniamide B_1 (**2**) and 8R and 8S of alpiniamide B_2 (**3**).

Position	**(1)**			**(2)**			**(3)**		
	6S $\boldsymbol{\delta_H}$	6R $\boldsymbol{\delta_H}$	$\Delta\delta_{(S-R)}$	7S $\boldsymbol{\delta_H}$	7R $\boldsymbol{\delta_H}$	$\Delta\delta_{(S-R)}$	8S $\boldsymbol{\delta_H}$	8R $\boldsymbol{\delta_H}$	$\Delta\delta_{(S-R)}$
1	-	-	-	-	-	-	-	-	-
2	2.69	2.72	-0.03	3.32	3.36	-0.04	3.40	3.36	+0.04
3	5.53	5.48	+0.05	5.60	5.67	-0.07	5.64	5.57	+0.07
4	-	-	-	-	-	-	-	-	-
5	5.75	5.69	+0.06	5.17	5.20	-0.03	5.31	5.26	+0.04
6	2.04	2.00	+0.04	1.68	1.64	+0.04	1.70	1.74	-0.04
7	0.95	0.93	+0.02	0.90	0.82	+0.08	0.83	0.89	-0.06
8	1.01	1.02	-0.01	1.25	1.30	-0.05	1.26	1.25	+0.01
9	1.58	1.41	+0.17	1.54	1.69	-0.15	1.70	1.53	+0.17
1’	-	-	-	-	-	-	-	-	-
2’	2.96	3.00	-0.04	3.42	3.46	-0.04	3.46	3.45	+0.01
3’	4.21	4.20	+0.01	4.29	4.28	+0.01	4.39	4.38	+0.01
4’	3.46	3.42	+0.04	3.02	3.00	+0.02	3.07	3.05	+0.02
5’	-	-	-	-	-	-	-	-	-
6’	4.11	4.26	-0.15	n.d.	n.d.	-	n.d.	n.d.	-
	3.78	4.01	-0.23	n.d.	n.d.	-	n.d.	n.d.	-
7’	1.27	1.22	+0.05	1.37	1.36	+0.01	1.38	1.37	+0.01
8’	1.37	1.35	+0.02	1.29	1.27	+0.02	1.30	1.30	0

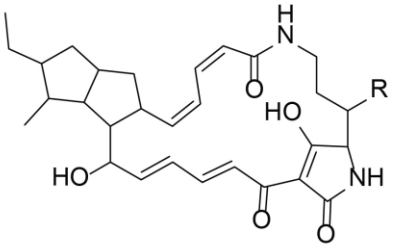
n.d.: not detected due to overlap with MTPA signals.

2.2 Supplementary Figures

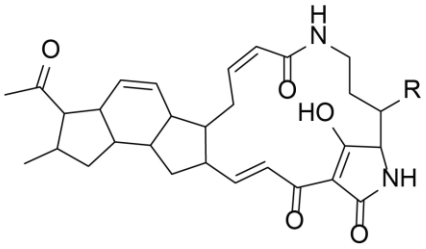


Supplementary figure S1. Reaction of alpiniamides with R-MTPA-Cl and S-MTPA-Cl in order to determine the absolute configuration.

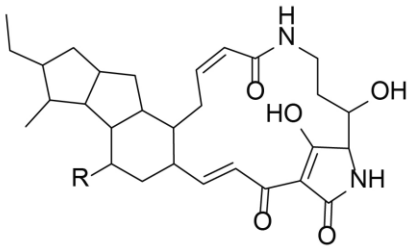
Supplementary figure S2. *Streptomycetes sp.* IB2014/011-12 grown on MS medium for 7 days.



alteramide A **(9)** R=OH
alteramide B **(10)** R=H

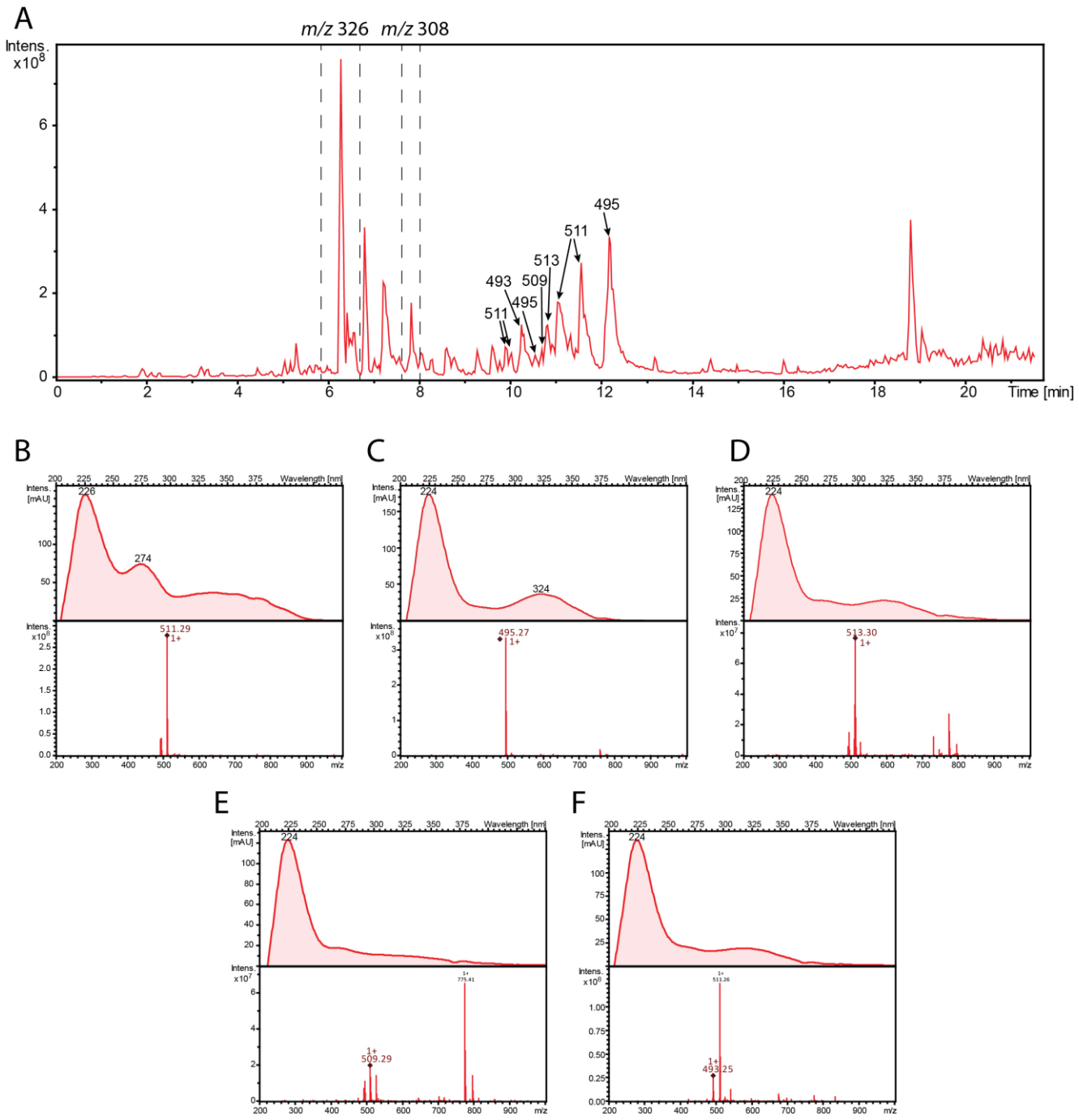


clifednamide A **(11)** R=OH
clifednamide B **(12)** R=H

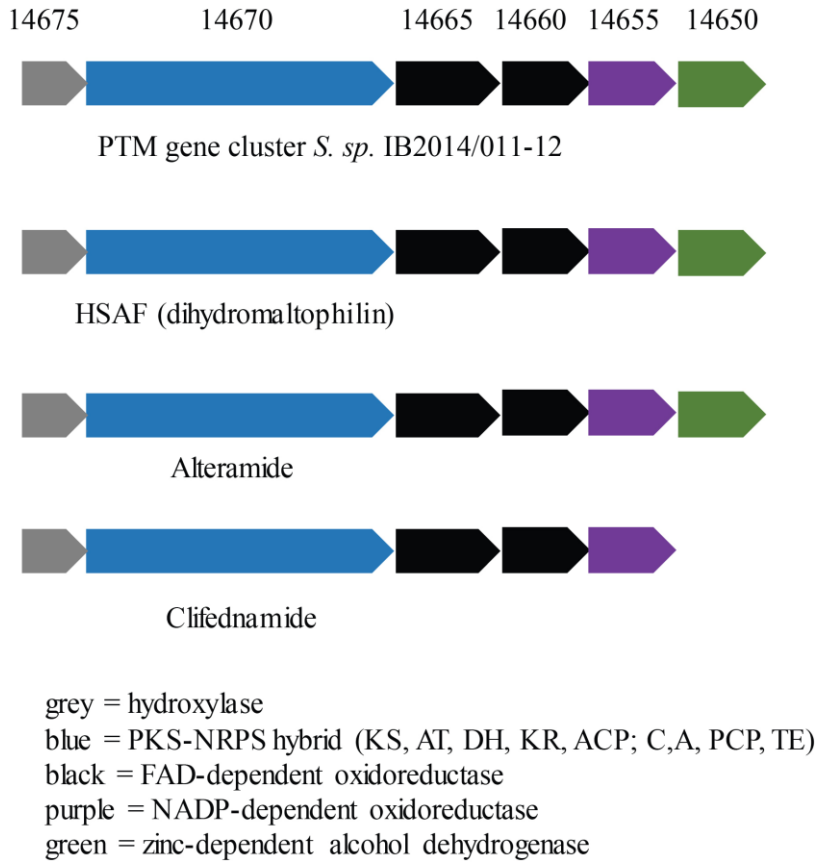


dihydromaltophylin **(13)** (HSAF); R=OH

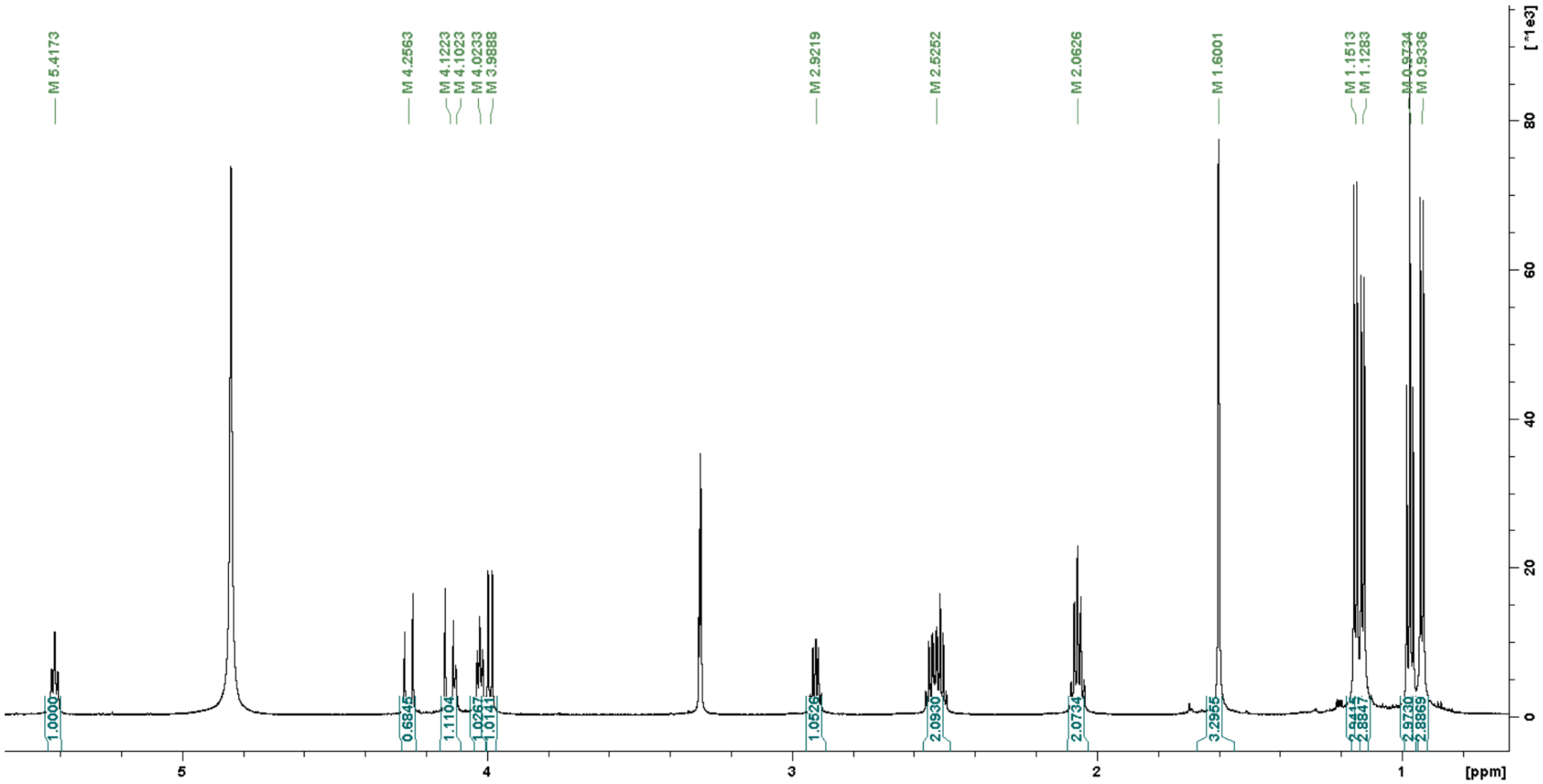
Supplementary figure S3. Structures of polycyclic tetramate macrolactams **(9-13)** identified in the extract of *Streptomyces sp.* IB2014/011-12.



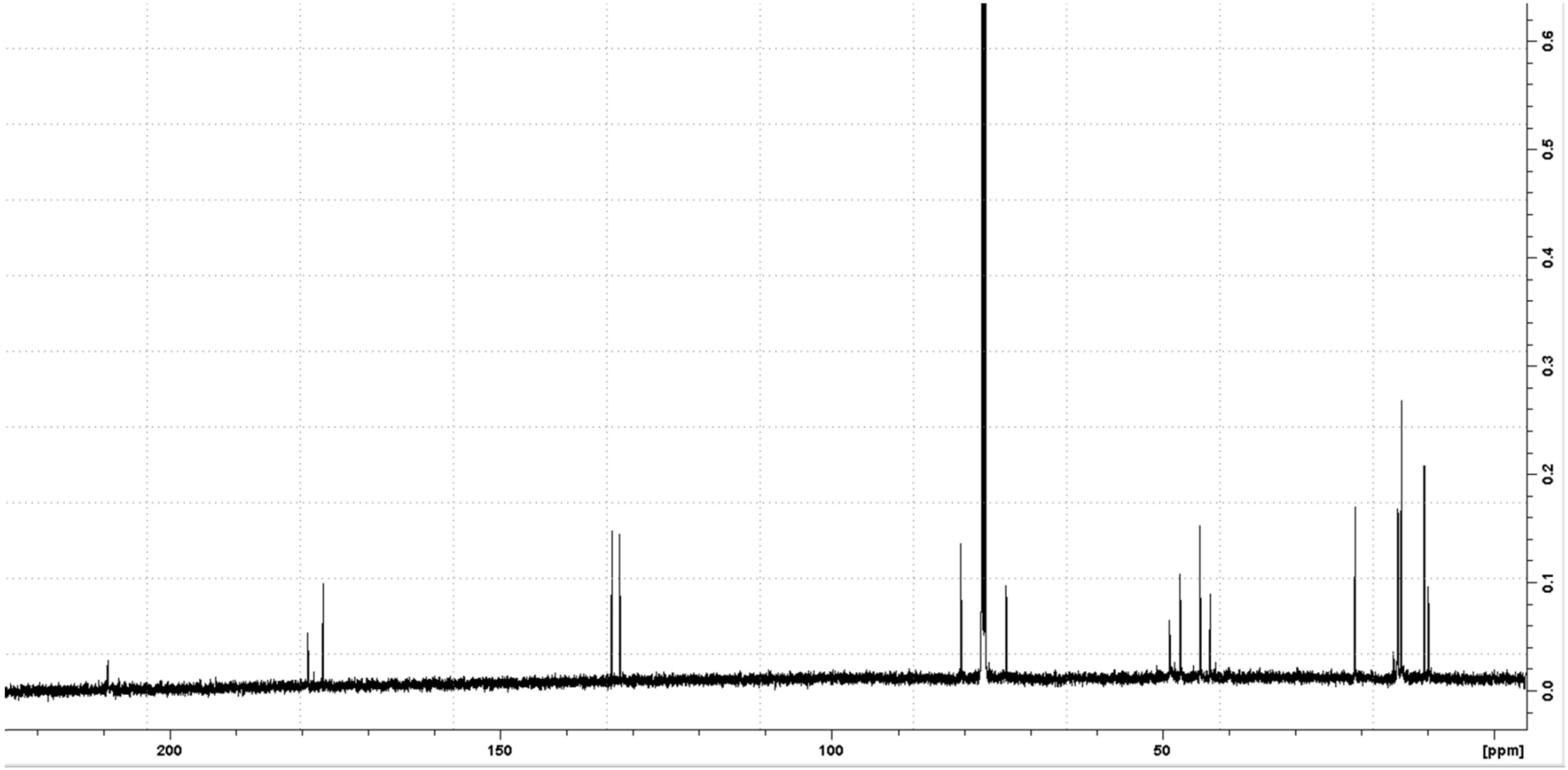
Supplementary figure S4. (A) MS and UV spectra of polycyclic tetramate macrolactams identified in the extract of *Streptomycetes sp.* IB2014/011-12. **(B)** alteramide A **(9)**; **(C)** alteramide B **(10)**; **(D)** clifednamide A **(11)**; **(E)** clifednamide B **(12)**; **(F)** dihydromaltophylin **(13)** (HSAF).



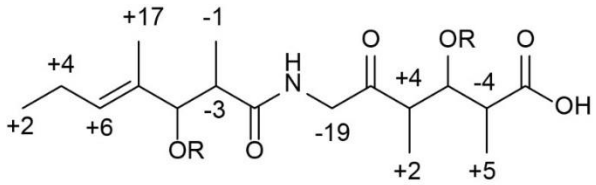
Supplementary figure S5. Genetic organization of gene cluster for PTMs production from the genome of *Streptomyces sp.* IB2014/011-12 and gene cluster of dihydromaltophylin, alteramide and clifednamide.



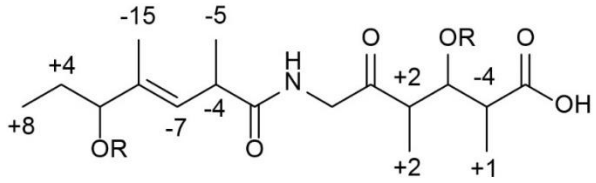
Supplementary figure S6. ^{1}H NMR-spectra of alpiniamide A in $MeOD_4$ 500 MHz.



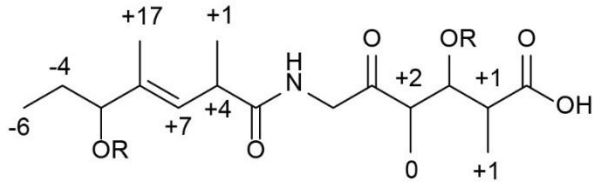
Supplementary figure S7. ^{13}C NMR-spectra of alpiniamide A in $CDCl_3$ 500 MHz.



alpiniamide A **(1)** R = MTPA

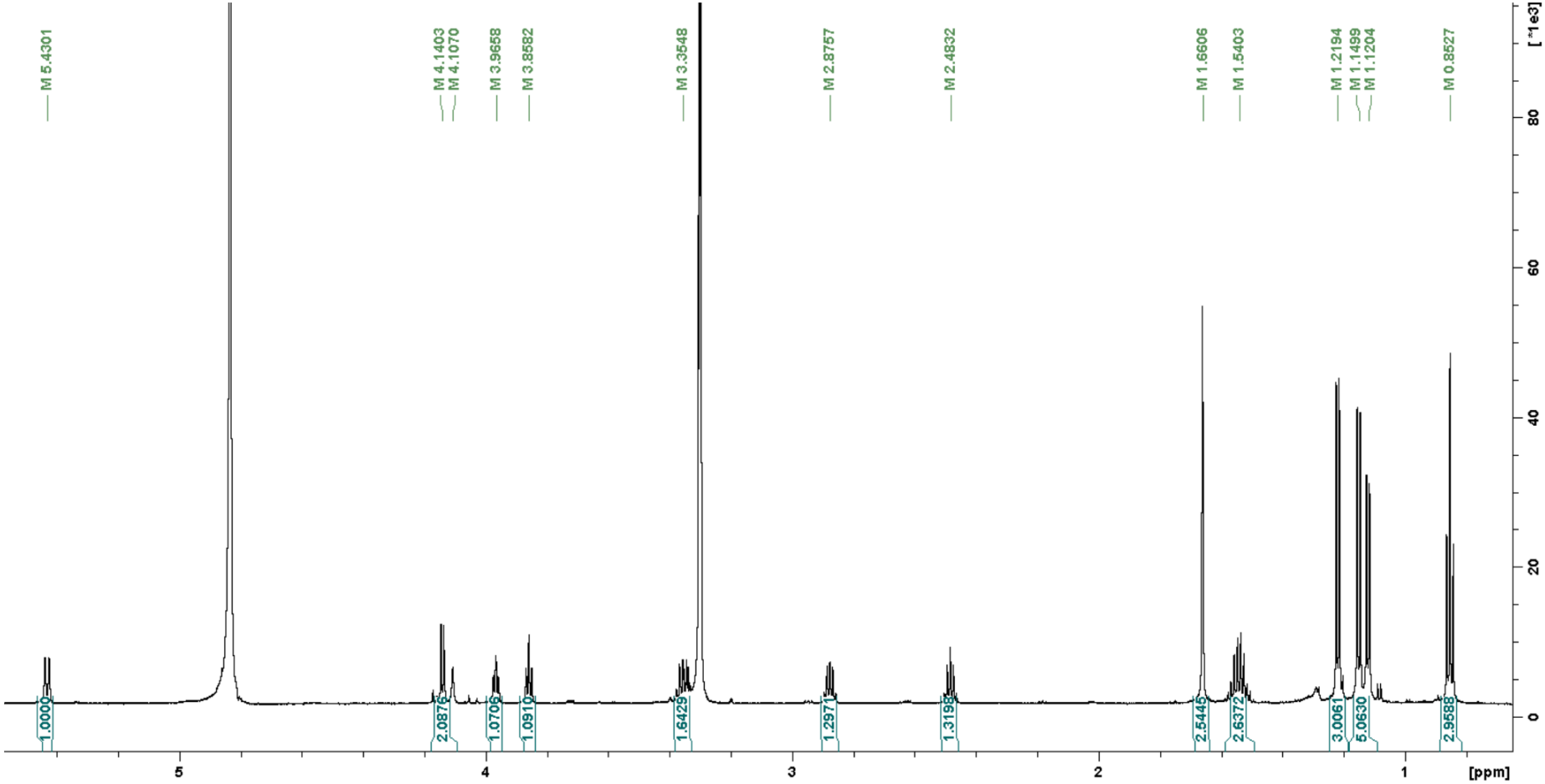


alpiniamide B_1 **(2)**

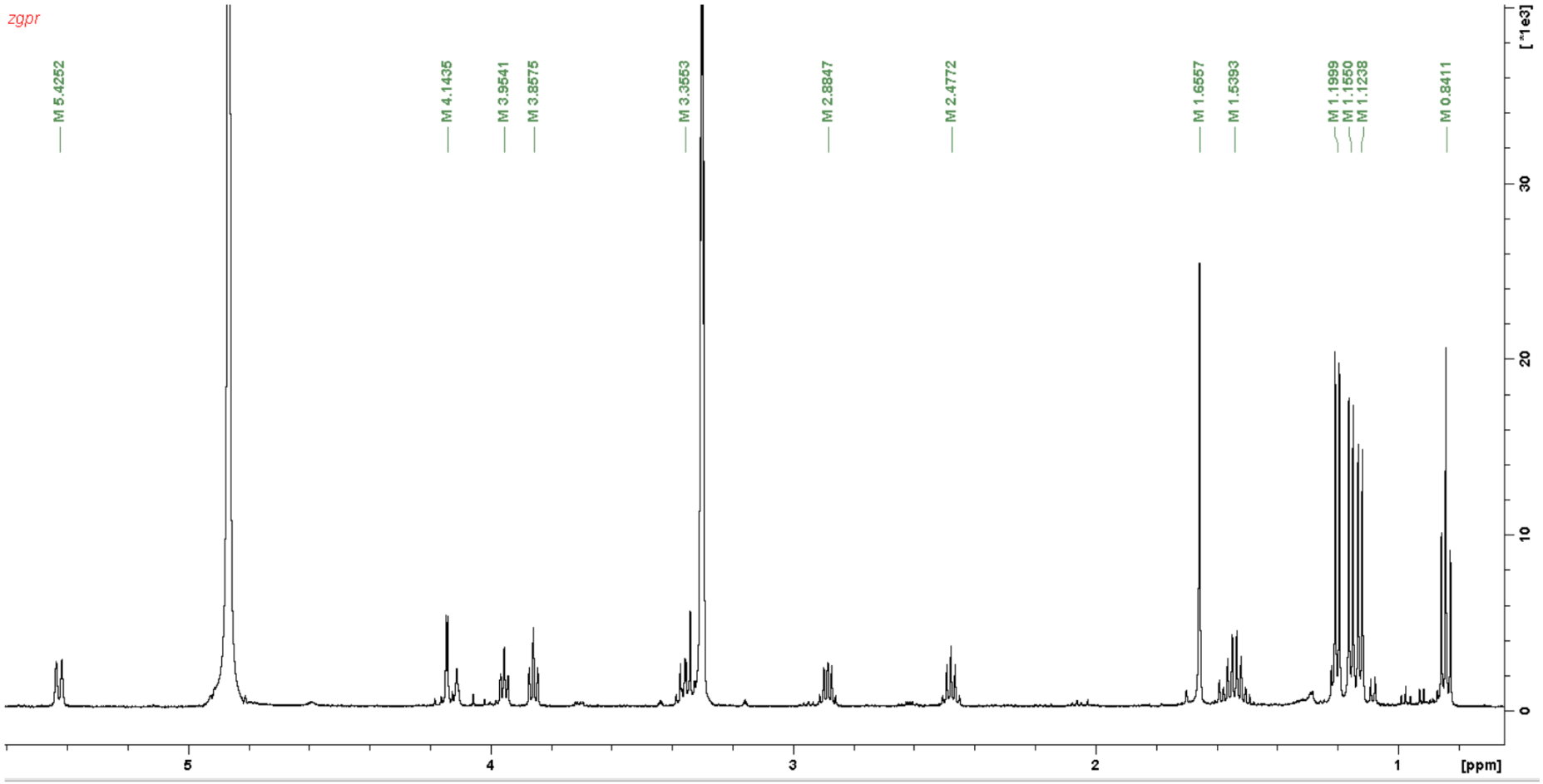


alpiniamide B_2 **(3)**

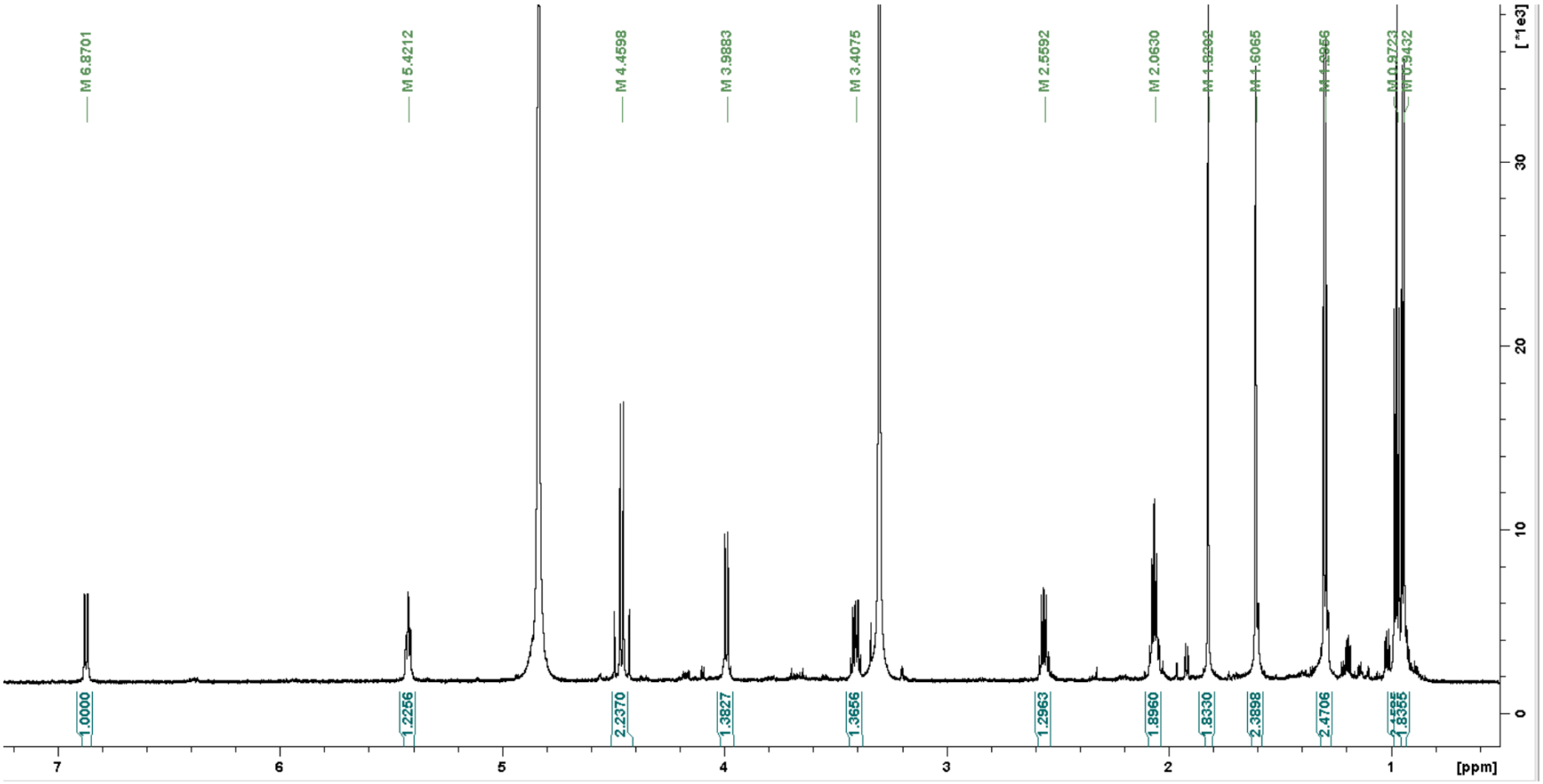
Supplementary figure S8. $\Delta\delta_{(S-R)}$ values ($x10^{-2}$ ppm) of the MTPA esters of **1-3**.



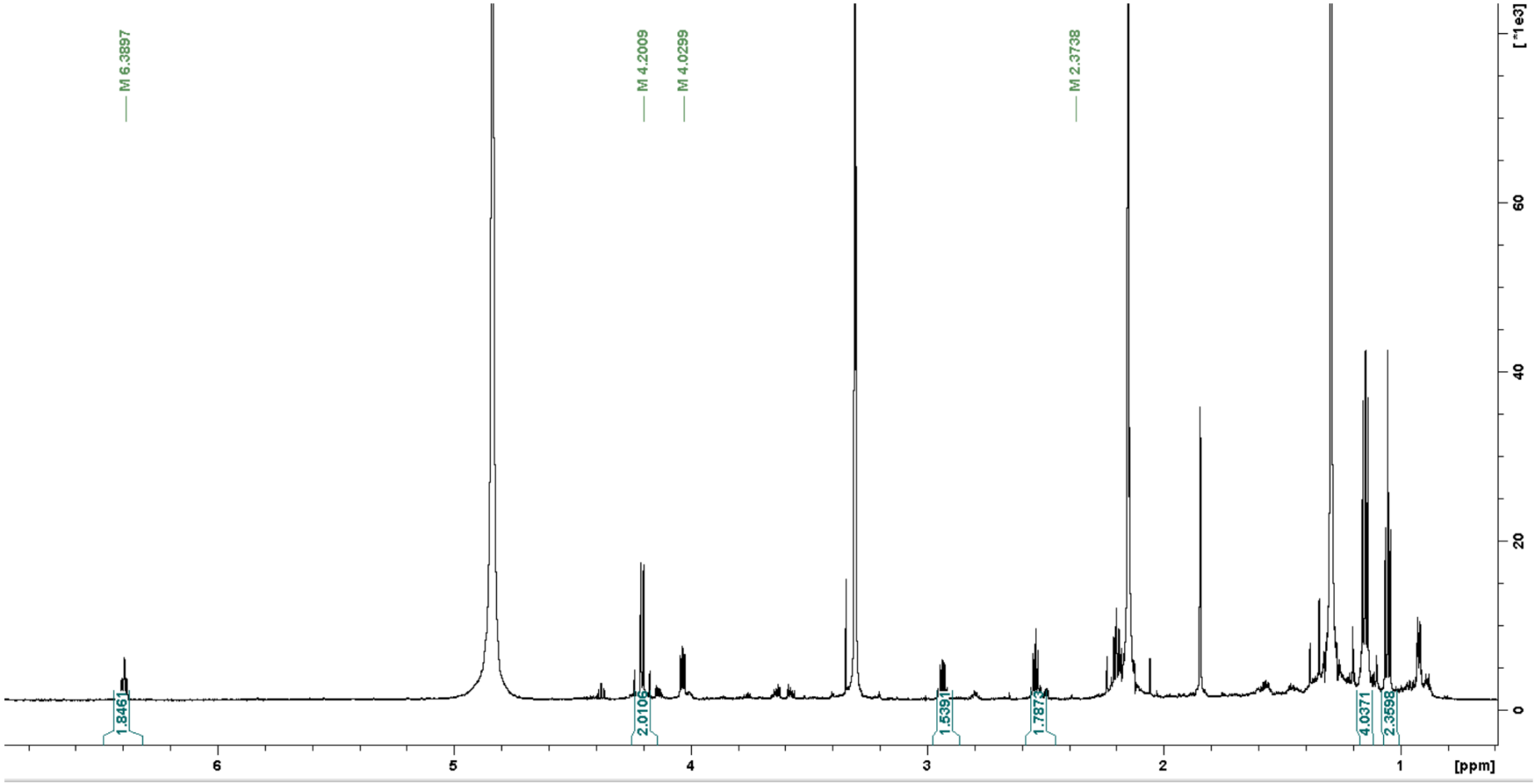
Supplementary figure S9. ^{1}H NMR-spectra of alpiniamide B_1 in $MeOD_4$ 500 MHz.



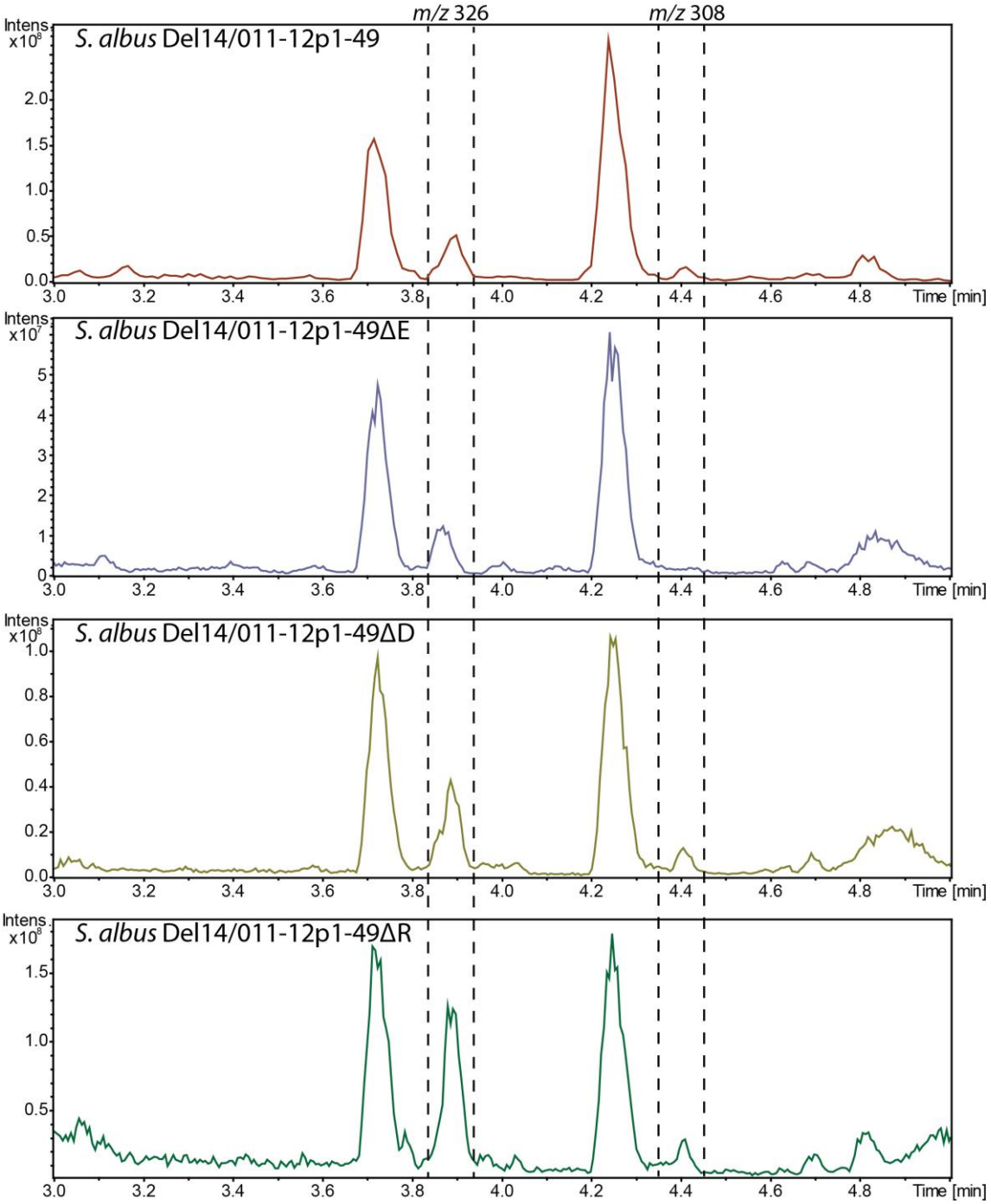
Supplementary figure S10. ^{1}H NMR-spectra of alpiniamide B_2 in $MeOD_4$ 500 MHz.



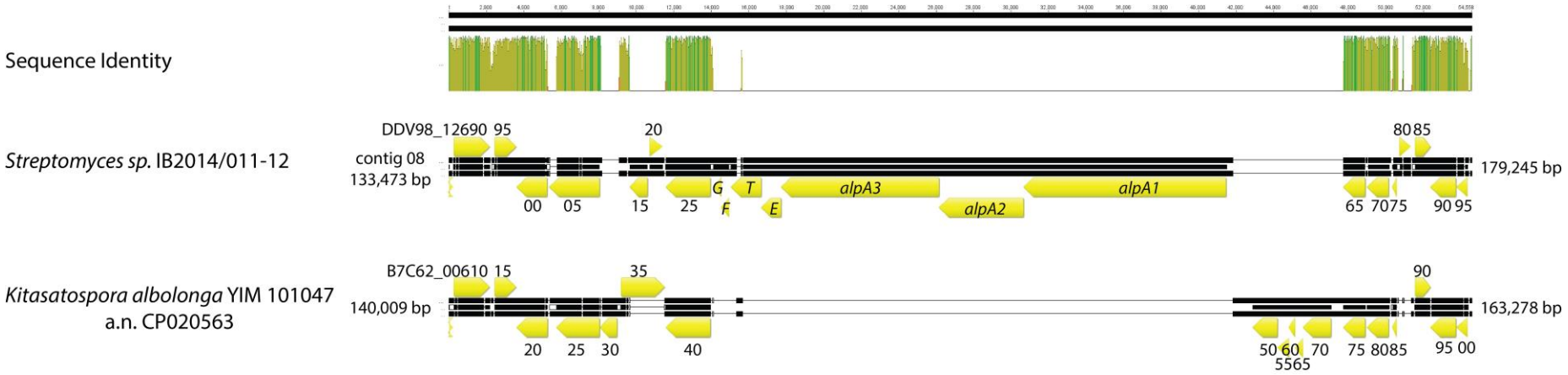
Supplementary figure S11. ^{1}H NMR-spectra of alpiniamide C in $MeOD_4$ 500 MHz.



Supplementary figure S12. ^{1}H NMR-spectra of alpiniamide D in $MeOD_4$ 500 MHz.



Supplementary figure S13. LC-MS chromatogram of extracts of *S. albus* Del14 carrying 011-12p1-49 plasmid and it's variants with the deletion of individual *alp* genes: *S. albus* Del14/011-12p1-49ΔE – *alpE*; *S. albus* Del14/011-12p1-49ΔD – *alpD; S. albus* Del14/011-12p1-49ΔR – *alpR.* Peak that corresponds to alpiniamides are highlighted: *m/z* 326 $[M\text{-}H_2O+H]^+$ for alpiniamides A and B, m/z 308 $[M\text{-}H_2O+H]^+$ for alpiniamide C. In the case of *S. albus* Del14/011-12p1-49ΔE small peak eluting around same RT as alpiniamide A has *m/z* of 285 $[M+H]^+$.



Supplementary figure S14. Alignment of whole genome of *Kitasatospora albolonga* YIM 101047 (GenBank accession number CP020563) and contig 08 of *Streptomyces sp.* IB2014/011-12 genome containing *alp* gene cluster and regions surrounding it. The alignment was performed with Geneious v. 8.1.7 software (Biomatters Ltd., New Zealand) using progressive Mauve algorithm.

Supplementary figure S15. Multiple sequence alignment of TE domain of AlpA2 using the Conserved Domain BLAST with the closest representatives of COG3319 super-family. The proposed active site residues (catalytic triad) are marked with red frames. The catalytic triad is deduced based on structural data for thioeserase domain of surfactin synthetase subunit 3 SrfA-C (730829, highlighted in yellow) (Bruner et al., 2002).

3 REFERENCES

Barona-Gómez, F., Wong, U., Giannakopulos, A. E., Derrick, P. J., and Challis, G. L. (2004). Identification of a cluster of genes that directs desferrioxamine biosynthesis in Streptomyces coelicolor M145. *Journal of the American Chemical Society* 126, 16282–16283. doi: 10.1021/ja045774k

Bruner, S. D., Weber, T., Kohli, R. M., Schwarzer, D., Marahiel, M. A., Walsh, C. T., et al. (2002). Structural basis for the cyclization of the lipopeptide antibiotic surfactin by the thioesterase domain SrfTE. *Structure (London, England : 1993)* 10, 301–310.

Challis, G. L., and Ravel, J. (2000). Coelichelin, a new peptide siderophore encoded by the Streptomyces coelicolor genome: Structure prediction from the sequence of its non-ribosomal peptide synthetase. *FEMS Microbiology Letters* 187, 111–114. doi: 10.1111/j.1574-6968.2000.tb09145.x

Cook, D., Rimando, A. M., Clemente, T. E., Schröder, J., Dayan, F. E., Nanayakkara, N. P. D., et al. (2010). Alkylresorcinol synthases expressed in Sorghum bicolor root hairs play an essential role in the biosynthesis of the allelopathic benzoquinone sorgoleone. *The Plant cell* 22, 867–887. doi: 10.1105/tpc.109.072397.

Kersten, R. D., Yang, Y.-L., Xu, Y., Cimermancic, P., Nam, S.-J., Fenical, W., et al. (2011). A mass spectrometry-guided genome mining approach for natural product peptidogenomics. *Nature chemical biology* 7, 794–802. doi: 10.1038/nchembio.684.

Luo, Y., Huang, H., Liang, J., Wang, M., Lu, L., Shao, Z., et al. (2013). Activation and characterization of a cryptic polycyclic tetramate macrolactam biosynthetic gene cluster. *Nature communications* 4, 2894. doi: 10.1038/ncomms3894

Pan, J.-J., Solbiati, J. O., Ramamoorthy, G., Hillerich, B. S., Seidel, R. D., Cronan, J. E., et al. (2015). Biosynthesis of Squalene from Farnesyl Diphosphate in Bacteria: Three Steps Catalyzed by Three Enzymes. *ACS central science* 1, 77–82. doi: 10.1021/acscentsci.5b00115

Schwarz, J., Konjik, V., Jankowitsch, F., Sandhoff, R., and Mack, M. (2016). Identification of the Key Enzyme of Roseoflavin Biosynthesis. *Angewandte Chemie (International ed. in English)* 55, 6103–6106. doi: 10.1002/anie.201600581

Suzuki, H., Ohnishi, Y., and Horinouchi, S. (2007). GriC and GriD constitute a carboxylic acid reductase involved in grixazone biosynthesis in Streptomyces griseus. *J. Antibiot.* 60, 380–387. doi: 10.1038/ja.2007.52

Ueda, K., Oinuma, K.-I., Ikeda, G., Hosono, K., Ohnishi, Y., Horinouchi, S., et al. (2002). AmfS, an Extracellular Peptidic Morphogen in Streptomyces griseus. *Journal of Bacteriology* 184, 1488–1492. doi: 10.1128/JB.184.5.1488–1492.2002

Woo, P. C. Y., Tam, E. W. T., Chong, K. T. K., Cai, J. J., Tung, E. T. K., Ngan, A. H. Y., et al. (2010). High diversity of polyketide synthase genes and the melanin biosynthesis gene cluster in Penicillium marneffei. *The FEBS journal* 277, 3750–3758. doi: 10.1111/j.1742-4658.2010.07776.x

Zhu, D., Liu, J., Han, R., Shen, G., Long, Q., Wei, X., et al. (2014). Identification and characterization of ectoine biosynthesis genes and heterologous expression of the ectABC gene cluster from Halomonas sp. QHL1, a moderately halophilic bacterium isolated from Qinghai Lake. *Journal of microbiology (Seoul, Korea)* 52, 139–147. doi: 10.1007/s12275-014-3389-5.

III

Development of a Biosensor Concept to Detect the Production of Cluster-Specific Secondary Metabolites

Yi-Qian Sun, Tobias Busche, Christian Rückert, Constanze Paulus, Yuriy Rebets, Renata Novakova, Jörn Kalinowski, Andriy Luzhetskyy, Jan Kormanec, Olga N. Sekurova and Sergey B. Zotchev

ACS Synthetic Biology, **2017**, *6* (6), 1026-1033

DOI: 10.1021/acssynbio.6b00353

Published online 21st February 2017

ACS SyntheticBiology

Research Article

pubs.acs.org/synthbio

Development of a Biosensor Concept to Detect the Production of Cluster-Specific Secondary Metabolites

Yi-Qian Sun,[†,‡] Tobias Busche,[§] Christian Rückert,[§] Constanze Paulus,[∥,#] Yuriy Rebets,[∥] Renata Novakova,[⊥] Jörn Kalinowski,[§] Andriy Luzhetskyy,[∥,#] Jan Kormanec,[⊥] Olga N. Sekurova,[▽] and Sergey B. Zotchev*[,▽]

[†]Department of Biotechnology, Norwegian University of Science and Technology, 7491 Trondheim, Norway

[‡]The Department of Laboratory Medicine, Children's and Women's Health (LBK), Norwegian University of Science and Technology, 7491 Trondheim, Norway

[§]Center for Biotechnology, Bielefeld University, Universitätsstraße 27, 33615 Bielefeld, Germany

[∥]Helmholtz Institute for Pharmaceutical Research Saarland, Actinobacteria Metabolic Engineering Group, 66123 Saarbrücken, Germany

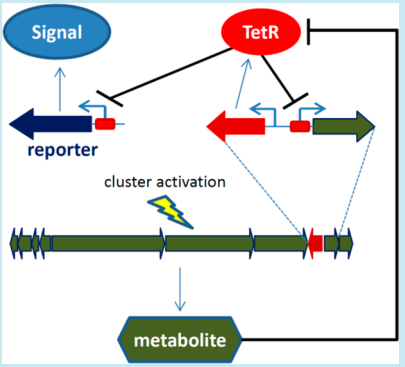
[⊥]Institute of Molecular Biology, Slovak Academy of Sciences, 845 51 Bratislava, Slovak Republic

[#]Universität des Saarlandes, Pharmaceutical Biotechnology, 66123 Saarbrücken, Germany

[▽]Department of Pharmacognosy, University of Vienna, 1090 Vienna, Austria

Supporting Information

ABSTRACT: Genome mining of actinomycete bacteria aims at the discovery of novel bioactive secondary metabolites that can be developed into drugs. A new repressor-based biosensor to detect activated secondary metabolite biosynthesis gene clusters in *Streptomyces* was developed. Biosynthetic gene clusters for undecylprodigiosin and coelimycin in the genome of *Streptomyces lividans* TK24, which encoded TetR-like repressors and appeared to be almost "silent" based on the RNA-seq data, were chosen for the proof-of-principle studies. The *bpsA* reporter gene for indigoidine synthetase was placed under control of the promotor/operator regions presumed to be controlled by the cluster-associated TetR-like repressors. While the biosensor for undecylprodigiosin turned out to be nonfunctional, the coelimycin biosensor was shown to perform as expected, turning on biosynthesis of indigoidine in response to the concomitant production of coelimycin. The developed reporter system concept can be applied to those cryptic gene clusters that encode metabolite-sensing repressors to speed up discovery of novel bioactive compounds in *Streptomyces*.



KEYWORDS: *Streptomyces, secondary metabolite biosynthesis, orphan gene clusters, TetR repressor, biosensor*

Natural products from *Streptomyces* and other actinomycete bacteria remain a rich source for the discovery of new bioactive compounds that may have a potential to be developed into human medicines. Recent advances in genomics of these important bacteria revealed their unprecedented capacity to synthesize structurally diverse secondary metabolites, while only a few of those are actually produced in standard laboratory conditions.[1] Consequently, a concept of genome mining was developed, that utilizes information on secondary metabolite biosynthesis gene clusters in order to either activate their expression in native producers, or express them heterologously in engineered bacterial hosts.[2] Despite some success stories,[3] genome mining still appears to be unpredictable in terms of outcomes and may yield unexpected results.[4] The most laborious part of genome mining remains analytics aimed at detection and identification of a compound expected to be produced as a result of activation or heterologous expression of a targeted gene cluster. Development of a robust biosensor that could be used to directly assess actual production of a cluster-specific compound would be of great advantage, also with respect to high-throughput screening after random mutagenesis.

Certain types of proteins provide unique opportunities for development of such biosensors. For example, resistance to tetracycline in Gram-negative bacteria is due to the export of this antibiotic by the TetA efflux pump. Expression of *tetA* is under tight transcriptional control of the repressor TetR, which is encoded by a divergently transcribed gene. Upon binding to tetracycline, TetR dissociates from its operator upstream of *tetA*, thus allowing its transcription, and consequently, efflux of tetracycline by TetA.[5] The first biosensor capable of detecting oxytetracycline production by *Streptomyces rimosus* was constructed in *Escherichia coli* based on the original TetR

Received: November 22, 2016
Published: February 21, 2017

DOI: 10.1021/acssynbio.6b00353
ACS Synth. Biol. 2017, 6, 1026−1033

Table 1. Secondary Metabolite Biosynthesis Gene Clusters in the Genome of *S. lividans* TK24 Identified Using AntiSMASH 3.0, Followed by Manual Curation[a]

no	cluster type	TetR-like repressor	product	scaffold-building gene expression, max TPM
1	Terpene	−	Terpene	SLIV_00150 (1.7)
2[a]	Terpene	+	2-methylisoborneol	SLIV_00800 (324.1)
3	NRPS	+	Coelibactin	SLIV_00885 (179.9)
4[a]	PKSIII	−	Polyketide	SLIV_00945 (0.28)
5	Indole	−	Indole (putative)	SLIV_01965 (16.0)
6[a]	Other	−	Unknown, likely false-positive	N/A
7	Terpene	−	Hopanoids	SLIV_04695 (109.3)
8	Lantipeptide	−	Lanthionone-containing peptide SapB	SLIV_05105 (35.4)
9	NRPS	−	Nonribosomally synthesized dipeptide	SLIV_06265 (20.7)
10	PKS I-butyrolactone	+	Coelimycin	SLIV_06770 (0.4)
11	Siderophore	−	Siderophore (putative)	SLIV_07005 (4.1)
12[a]	Terpene	+	Geosmin	SLIV_07775 (417.9)
13	Bacteriocin	−	Bacteriocin (putative)	SLIV_07915 (955.7)
14	PKS I	+	Undecylprodigiosin	SLIV_09145 (2.0)
15	Siderophore	−	Siderophore (putative)	SLIV_09615 (7.5)
16	PKS II	−	Spore pigment	SLIV_11780 (5.9)
17	Terpene	−	Albaflavenone	SLIV_12255 (63.5)
18	PKSII	+	Actinorhodin	SLIV_12950 (192.9)
19	NRPS	−	Calcium-dependent antibiotic	SLIV_21540 (3.4)
20	Siderophore	−	Desferroxamine B	SLIV_23760 (1035.8)
21	Melanin	−	Melanin	SLIV_24170 (19.8)
22	Ectoine	−	Ectoine	SLIV_28385 (1011.2)
23[a]	PKS III	+	1,3,6,8-tetrahydroxynaphthalene	SLIV_31765 (27.7)
24	Bacteriocin	−	Bacteriocin (putative)	SLIV_34165 (71.5)
25	NRPS	+	Coelichelin	SLIV_35480 (217.1)
26	Lantipeptide	−	Class I lantipeptide (putative)	SLIV_36570 (38.7)
27	Terpene	−	Isorenieratene	SLIV_36960 (32.3)
28[a]	PKS I	−	Polyunsaturated fatty acid (putative)	SLIV_37285 (10.6)
29	Terpene	−	Terpene	SLIV_37475 (1.8)

[a]Clusters misannotated or missed by the software *via* artificial fusion of several clusters.

repressor.[6] TetR-like repressor-coding genes have been identified within or in the vicinity of secondary metabolite clusters in many *Streptomyces*, which implied their potential roles in sensing cluster-specific molecules. Some of such repressors were indeed confirmed to bind both the end products of cluster-specific biosynthetic pathways, and/or its precursors. One of the first examples that concerned TetR-like repressors responding to secondary metabolites other than tetracyclines was that of LanK, the repressor controlling landomycin biosynthetic pathway in *Streptomyces cyanogenus*.[7] This protein was shown to bind various landomycins with altered aglycon structure, as well as extremely low concentration of the final product, landomycin A. Consequently, the authors have suggested that LanK can be used to develop a biosensor for detection of landomycin-related compounds. The Repressor function of SymR, a TetR-like protein encoded by the simocyclinone biosynthesis gene cluster of *S. antibioticus*, was shown to be abrogated upon its binding to either simocyclinone D8 or its precursor C4.[8] Recently, chlorothricin and its biosynthetic intermediates were shown to modulate binding of a TetR-like repressor ChlF1 to its target genes in the chlorothricin gene cluster of *Streptomyces antibioticus*.[9]

These findings prompted us to investigate if TetR-like repressors in *Streptomyces* can be used for development of robust biosensors, which would couple easily detectable expression of the reporter with the production of specific secondary metabolites. Such biosensor would represent a clear advantage over the earlier described reporter system that detects expression of the secondary metabolite biosynthesis genes, as described by Guo *et al.* for gaudimycin biosynthesis.[10] Indeed, our design would allow direct monitoring of the cluster-specific product, not just expression of its biosynthetic genes. Here, we present a proof-of-principle for such a biosensor, and argue that similar design can greatly assist in the genome-based bioprospecting for novel bioactive compounds in *Streptomyces* and other actinomycete bacteria.

RESULTS AND DISCUSSION

Identification of Poorly Expressed Secondary Metabolite Biosynthesis Gene Clusters in *Streptomyces lividans* TK24. Analysis of the *S. lividans* TK24 genome[11] using online version of the software antiSMASH 3.0[12] revealed 27 gene clusters for the biosynthesis of secondary metabolites, but this number was amended to 29 after manual curation using BLAST searches (Table 1). Out of those, 17 gene clusters for known compounds were identified, governing biosynthesis of actinorhodin, undecylprodegiosin, coelibactin, albaflavenone, coelimycin, hopanoids, desferroxamine, lanthionine-containing peptide SapB, 2-methylisoborneol, calcium-dependent antibiotic, melanin, spore pigment, ectoine, coelichelin, 1,3,6,8-tetrahydroxynaphthalene, isorenieratene and geosmin. It is worth noting that *S. lividans* TK24 has two identical gene clusters for the biosynthesis of albaflavenone, apparently located within the terminal inverted repeats of the linear chromosome. Next, we decided to explore the transcriptional landscape of the secondary metabolite biosynthesis gene

DOI: 10.1021/acssynbio.6b00353
ACS Synth. Biol. 2017, 6, 1026−1033

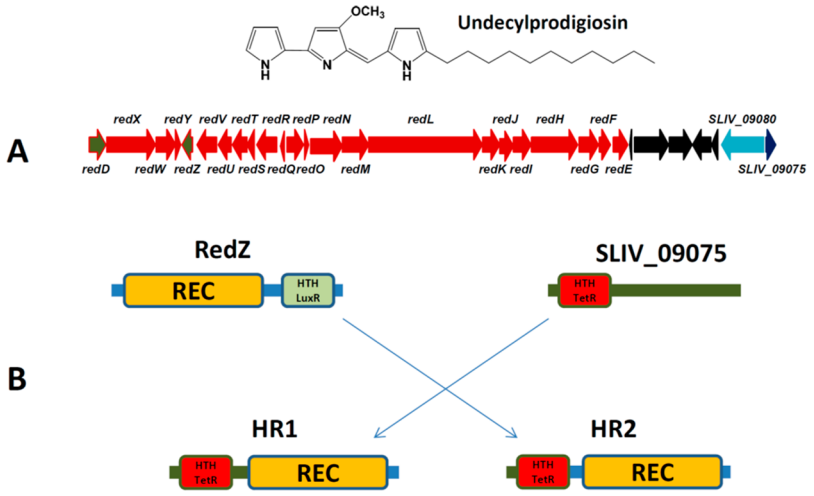


Figure 1. (A) Organization of the undecylprodegiosin biosynthesis gene cluster and its right flank in *S. lividans* TK24. Regulatory genes *redD* and *redZ* are shown in dark green, while TetR-like repressor- and its putative cognate transporter-encoding genes are shown in blue and light blue, respectively. (B) Scheme of hybrid repressor construction. HTH, helix-turn-helix DNA binding domain; REC, putative ligand binding receiver domain.

clusters using RNA-seq technology.[13] Total RNA was isolated from *S. lividans* TK24 cultures growing in liquid minimal media with or without addition of casamino acids at early and late logarithmic phases, as well as at stationary phase. The mRNA fractions were enriched by depletion of stable RNA and then converted to strand-specific transcriptome libraries. After sequencing by Illumina technology, gene expression analysis was performed for every scaffold-building gene of a secondary metabolite gene cluster by calculating the transcripts per million reads (TPM) in order to find transcriptionally silent and therefore putatively repressed secondary metabolite gene clusters (a detailed description of the RNA-seq experiments will be published elsewhere). We specifically examined the data for the gene clusters encoding TetR-like repressors, namely those for 2-methylisoborneol (cluster 2), coelimycin (cluster 10), geosmin (cluster 12), undecylprodigiosine (cluster 14), spore pigment (cluster 16), actinorhodin (cluster 18), 1,3,6,8-tetrahydroxynaphthalene (cluster 23), and coelichelin (cluster 25). Full overview of the gene expression profiles for the *S. lividans* TK24 secondary metabolite biosynthesis gene clusters is presented in Supporting Information, Table S1. We have focused our examination on the genes from the targeted clusters, which encode enzymes responsible for the assembly of secondary metabolite scaffolds, reasoning that such scaffolds would represent a minimal structure that TetR repressors could bind. Out of examined gene clusters, only those for coelimycin and undecylprodigiosine fulfilled the criteria of being expressed at very low levels and encoding TetR-like repressors.

Attempts on Construction of the Undecylprodegiosin Biosensor. First, we have focused on the undecylprodigiosin biosynthetic gene cluster (*red*, Figure 1A), which has been characterized in quite some details.[14] These earlier studies, however, did not address the function of the divergently transcribed genes SLIV_09075 and SLIV_09080 located *ca.* 9 kb downstream of *redE*. SLIV_09075 encodes a TetR-like repressor, and the SLIV_09080 gene product is similar to MMPL domain-containing transporters. Since the *red* cluster does not encode any transporters, it seemed logical to suggest that the SLIV_09075+SLIV_09080 pair is responsible for undecylprodigiosin efflux in response to the production of this compound, with the TetR-like repressor being the receiver of the signal. Considering this, we PCR-amplified the putative promoter/operator region upstream of SLIV_09080, and placed it upstream of the *bpsA* reporter in an integrative plasmid. BpsA is a monomodular nonribosomal peptide synthetase responsible for the biosynthesis of the blue pigment indigoidine, which is easily detectable both visually and spectroscopically.[15] When the resulting putative undecylprodigiosin reporter plasmid pYQS042 (Table S1) was introduced into *S. lividans* TK24, no indigoidine production was observed in the recombinant strain, suggesting that the SLIV_09075-encoded repressor blocks expression of *bpsA*.

Protein RedD encoded by the SLIV_09220 gene in the *red* cluster of *S. lividans* was shown to be a positive activator of the undecylprodigiosin gene cluster in *S. coelicolor*.[16] Therefore, overexpression of RedD was expected to trigger biosynthesis of undecylprodigiosin in *S. lividans* as well, to which the constructed biosensor would react. Consequently, genes encoding RedD (SLIV_09220) and another putative regulator of the LuxR family, RedZ (SLIV_09200)[14] in the *red* cluster were overexpressed in the recombinant *S. lividans* TK24 harboring reporter plasmid pYQS042. Overexpression of both *redD* and *redZ* led to apparent activation of undecylprodigiosin production (activation was more efficient upon overexpression of *redD*), while no production of indigoidine could be observed in the engineered strains. These results implied that the TetR-like repressor encoded by SLIV_09075 does not bind undecylprodigiosin, and its gene may not belong to the *red* cluster.

In the next attempt to construct undecylprodigiosin-specific biosensor, we decided to combine the features of TetR-like repressor and LuxR-type activator in one hybrid protein. The N-terminal part of LuxR activator proteins are known to contain autoinducer-binding domains that have affinity to N-acyl homoserine lactones (AHLs).[17] Taking into account the fact that no AHL synthase is encoded by the *red* cluster, and recent reports on the LuxR "solos" that respond to diverse small molecules instead of AHLs,[18] we considered it possible that an N-terminal domain in the RedZ response regulator may sense undecylprodigiosin or its precursors. Therefore, we decided to generate hybrid proteins that would combine the

DOI: 10.1021/acssynbio.6b00353
ACS Synth. Biol. 2017, 6, 1026−1033

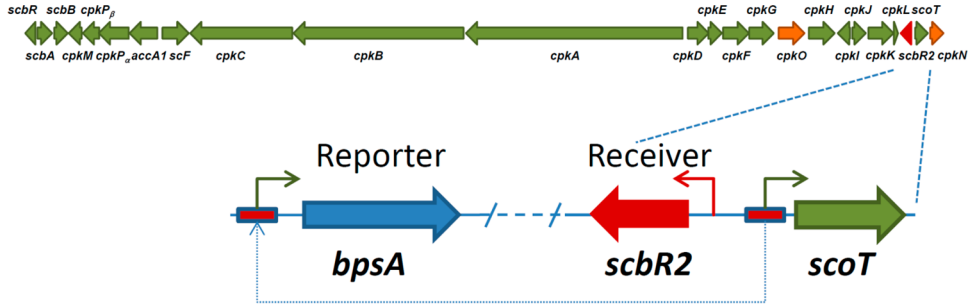


Figure 2. Organization of the coelimycin biosynthetic gene cluster and its parts used for the biosensor construction. Genes encoding putative positive activators are shown in orange, and a TetR-like repressor-encoding gene is shown in red. To construct a reporter, indigoidin synthetase gene *bpsA* was cloned downstream of the putative promoter/operator region controlling expression of type II thioesterase ScoT.

DNA binding domain of the TetR-like repressor encoded by SLIV_09075 and a putative metabolite-sensing domain of RedZ. The rationale behind this attempt to construct hybrid repressors stems from the earlier reported success in creating a hybrid transcription factor that combines the regulatory and activation domains of cAMP-responsive element binding protein with the DNA binding domain of the TetR repressor.[19] Functional metabolite-sensing hybrid repressors have also been reporter by Keasling group, which combined domains of metabolic enzymes to the DNA-binding domain of TetR.[20] Consequently, we constructed two versions of a hybrid repressor that, providing our assumption noted above was correct, could respond to the production of undecylprodigiosin. These versions were made by linking a putative N-terminal receiver domain of RedZ with the helix-turn-helix (HTH) DNA binding domain of the SLIV_09075-encoded TetR-like repressor. The domain boundaries of both proteins were predicted using online ProDom software tool, corresponding parts of the genes were amplified by PCR from the genomic DNA of *S. lividans* TK64 and fused using Gibson ligation (see Methods section for details).

The two versions of hybrid repressor were made with different interdomain linkers originating either from RedZ (for HR1 version) or TetR-like (for HR2 version) regulators (Figure 1B, Table S1: constructs pYQS065 and pYQS066, respectively). The wild type copy of the SLIV_09075 gene was replaced by either of the SLIV_09075-*redZ* hybrid genes in *S. lividans* TK24 using double crossover recombination. The reporter construct pYQS042 was then integrated into the chromosomes of the strains carrying hybrid genes. Both resulting recombinant strains showed production of indigoidine without overexpression of *redD*, suggesting that the hybrid proteins lost the repressor function characteristic of SLIV_09075. Most likely, the HTH DNA-binding domain of SLIV_09075 in hybrid proteins could no longer bind to the operator upstream of the reporter gene, probably due to the perturbations in their tertiary structures. Yet another possible explanation could be that the hybrid proteins acquired an activator function and, in the absence of the intact SLIV_09075, can activate expression of the *bpsA* reporter gene. In order to simplify visual detection of indigoidine production, which could be masked by actinorhodin and undcylprodigiosin, a recombinant strain *S. lividans* RedStrep1 was constructed, which lacks *red* and *act* clusters, but retains SLIV_09075. To delete both clusters in *S. lividans* TK24, an efficient deletion system based on the positive selection of double crossover events with *bpsA* reporter gene was used[14] (see Methods for details). Two RedStrep1-based recombinant strains were constructed, which harbored both the integrated reporter pYQS042, and a low-copy number autonomously replicating plasmid expressing HR1 or HR2 proteins under control of the *ermE**p promoter (Table S2). No indigoidine production was observed by these recombinant strains, which excluded the possibility that hybrid proteins had direct activator function and could outcompete the SLIV_09075-encoded repressor. In summary, hybrid proteins constructed here proved to be nonfunctional, prompting further studies toward robust design of the hybrid TetR-like repressors.

Construction of the Coelimycin-Specific Biosensor. Our failure to construct undecylprodigiosin-specific biosensor was mainly due to the fact that TetR-like repressor encoding gene located at the border of the *red* cluster apparently had no function in the biosynthesis of this antibiotic. To establish a proof of concept, we next focused on the coelimycin gene cluster (*cpk*) for biosensor development, since this compound has never before been identified in *S. lividans*, while methods for its detection were available from the earlier reports on the closely related strain *S. coelicolor* M145.[21] Coelimycin is a polyketide alkaloid, the biosynthetic gene cluster for which in *S. coelicolor* is silent, but can be activated upon deletion of the *scbR2* repressor gene.[22] In *S. coelicolor*, *scbR2* is located proximal (but transcribed in opposite direction) to the gene *scoT* encoding a type II thioesterase that proved to be important for coelimycin biosynthesis in the latter organism.[23] In *S. lividans* TK24, the arrangement of genes is similar (Figure 2), and it was safe to assume that the SLIV_06715-encoded repressor would control the expression of the divergently transcribed thioesterase II-encoding gene SLIV_06710. We PCR-amplified putative promoter/operator region upstream of SLIV_06710, and placed the *bpsA* reporter gene under its control in an integrative vector pSET152.[24] This reporter construct pYQS040 was integrated into the chromosome of *S. lividans* RedStrep1. The reporter strain RedStrep1 [pYQS040] showed no indigoidine production in a range of conditions, indicating that the SLIV_06715-encoding repressor acts *in trans*, and blocks the transcription of the reporter gene.

Testing of the Coelimycin Biosensor Using Artificial Activation of the *cpk* Gene Cluster. In order to obtain proof-of-principle for our biosensor system, we decided to force coelimycin production in the reporter strain RedStrep1 [pYQS040] using overexpression of a pathway-specific regulator. The latter strategy was shown to be useful in genome mining aimed at the discovery of novel compounds.[25] However, this approach may lead to unexpected results, and detection of the products of potentially activated clusters may prove difficult, especially in case of complex and unusual

DOI: 10.1021/acssynbio.6b00353
ACS Synth. Biol. 2017, 6, 1026−1033

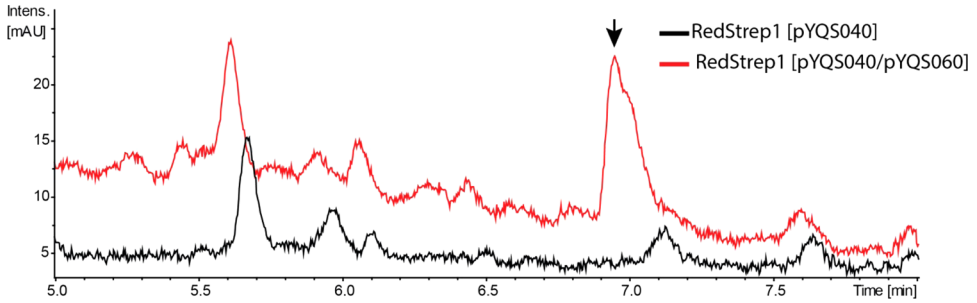


Figure 3. UV chromatogram (at 360 nm) of metabolites produced by *S. lividans* RedStrep1 [pYQS040] and RedStrep1 [pYQS040/pYQS060] when grown in SG medium. The peak presumably corresponding to coelimycin P1 is indicated by arrow. More detailed analysis, including MS/MS-based identification of coelimycin P1, was done as reflected in Figure S2 (Supporting Information).

clusters.[4] The *cpk* gene cluster in *S. lividans* encodes two putative positive regulators of the SARP family, SLIV_06705 (CpkN) and SLIV_06745 (CpkO) (Figure 2). The *cpkO* gene product was previously shown in *S. coelicolor* to positively regulate the *cpk* gene cluster expression,[26] while the function of *cpkN* gene was unclear. These putative regulatory genes were expressed under control of the P*ermE* promoter from the SCP2*-based replicative plasmids pYQS060 (*cpkN*) and pYQS061 (*cpkO*) in strain RedStrep1 [pYQS040]. Indigoidine production was observed in both RedStrep1 [pYQS040/pYQS060] and RedStrep1 [pYQS040/pYQS061] recombinant strains, which indicated activation of the *cpk* cluster. The recombinant strain overexpressing SLIV_06705 also showed more pronounced indigoidine production, suggesting that this SARP regulator is a stronger activator for the *cpk* cluster. In order to confirm that positive regulator-induced reporter expression was due to the relief of the SLIV_06715-encoded repressor upon coelimycin binding, gene SLIV_06770 encoding a scaffold-building PKS in the *cpk* cluster was inactivated by insertion of a suicide vector pYQS070. The resulting recombinant colonies of RedStrep1 [pYQS040/pYQS060/pYQS070] were no longer producing indigoidine, thus confirming that the developed biosensor system indeed responds to the production of coelimycin (Figure 3).

To obtain a final proof of coelimycin production, we compared metabolic profiles of the recombinant strains *S. lividans* RedStrep1 [pYQS040] carrying just the biosensor, and RedStrep1 [pYQS040/pYQS060] that in addition expressed the SARP activator RedD. The recombinant strains were cultivated in several commonly used liquid media (TSB, SG, NL19, phage medium and R5A) and extracted metabolites were analyzed by LC-MS. When grown in R5A or SG media strain RedStrep1 [pYQS040/pYQS060] was found to produce both coelimycin P1 and blue pigment indigoidine. Coelimycin P1 was by its absorption spectrum (characteristic peak at 360 nm), its exact mass (detected m/z 349.1213 $[M + H]^+$, calculated 349.1215 $[M + H]^+$), and by its MS/MS fragmentation pattern (Figure S2). Interestingly, trace amounts of coelimycin P1 were also detected in the extract of *S. lividans* RedStrep1 [pYQS040] strain when incubated in the R5A media, but not in SG. In these conditions, the biosensor did not react to the production of minute amounts of coelimycin P1, probably due to its efficient and rapid efflux from the cells. The accumulation of coelimycin P1 was not detected when strains were grown either in TSB, phage media or NL19 media. Under these conditions production of indigoidine by *S. lividans* RedStrep1 [pYQS040] was also not observed.

CONCLUSIONS

The genome mining approach has shown a considerable potential in discovery of novel microbial secondary metabolites.[25] Still, the metabolomes of prolific secondary metabolite producers are often very complex, precluding identification of cluster-specified compounds. Biosensors coupled to the actual production of target compounds would be very helpful in streamlining the discovery process, eliminating nonproducers from the screening. In a recent report Guo *et al.* constructed a reporter system linking the expression of genes for gaudimycin biosynthesis to a reporter system in *Streptomyces*.[10] However, authors have reported high percentage of false-positives which, while expressing a reporter, did not actually produced a target metabolite. Also, in some cases expression of biosynthetic genes does not correlate with the compound production, probably due to some obstacles affecting translation or post-translational modification of biosynthetic enzymes.[27] In our study, TetR-like repressors and *bpsA* reporter gene were used to construct biosensors that would directly detect production of secondary metabolites. Apparently, success in construction of a functional biosensor depends on whether identified TetR-like repressor gene actually belongs to the biosynthetic cluster, and can bind the cluster-specified compound. In the case of coelimycin, we have obtained a proof-of-principle, demonstrating the utility of such a biosensor, although we cannot exclude the possibility that biosensor may also react to some intermediates in the biosynthetic pathway. Publicly available genome sequences, especially those for actinomycete bacteria, provide access to hundreds of biosynthetic gene clusters encoding TetR-like repressors, for which biosynthetic products remain unknown. Moreover, there is a plethora of other regulator families, in which the binding of a metabolite leads to the release of the repressor from its operator (*e.g.*, LacI family regulators),[28] which are associated with the secondary metabolite biosynthesis gene clusters. Using the metabolite-sensing repressor-based biosensor principle, it will become possible to assess functional expression of these clusters either after random mutagenesis, under various cultivation conditions, or through genetic manipulation. On the basis of the promising results reported by Chou and Keasling on metabolite-sensing hybrid repressors,[20] it seems plausible that a more versatile and modular biosensor system can be developed. Indeed, fusion of a metabolite sensor represented by a domain of a biosynthetic enzyme predicted to bind the final product (or a critical intermediate) to a standardized part representing HTH domain that binds specific operator controlling a reporter can be done quickly and efficiently. Replacement of the receiver domain in a standardized hybrid repressor with a domain of a biosynthetic

DOI: 10.1021/acssynbio.6b00353
ACS Synth. Biol. 2017, 6, 1026−1033

enzyme predicted to bind the final product (or a critical intermediate) would only require minimal engineering in order to tune the biosensor to a particular gene cluster.

METHODS

Bacterial Strains, Plasmids, and Growth Conditions. Description of plasmids and bacterial strains used or generated during this study is provided in Tables S1 and S2 (Supporting Information), respectively. *S. lividans* TK24 was used as the wild-type strain. RedStrep 1 (Δ*act* Δ*red*) was constructed as shown in Figure S1 (Supporting Information) with the deletions of nt 2961047−2982325 (*act* cluster) and nt 2081767−2113444 (*red* cluster) in *S. lividans* TK24 genome.[11] The deletions were validated by genome sequencing. RedStrep 1 was used in this study to avoid the potential color disturbance due to production of pigmented antibiotics specified by the *act* and *red* clusters in *S. lividans* TK24.

Escherichia coli strains were grown in Luria−Bertani (LB) broth or on LB agar. XL1-blue was used for general cloning. ET12567 (pUZ8002) was used for intergeneric transfer of plasmids to *Streptomyces* with the standard procedure of conjugation.[28] *Streptomyces* spore preparation was obtained by growing on soy flour mannitol medium (SFM) at 30 °C for 4−7 days. Antibiotics were supplemented to growth medium at the following concentrations when applicable: ampicillin 100 μg/mL; apramycin 50 or 100 μg/mL; chloramphenicol 20 μg/mL; hygromycin 100 μg/mL; kanamycin 25 μg/mL; nalidixic acid 30 μg/mL, and thiostrepton 30 μg/mL.

RNA Isolation, Library Preparation and cDNA Sequencing. *S. lividans* TK24 cells were grown in MM media supplemented with and without casamino acids. Samples were taken at early and late logarithmic phases, as well as at stationary phase. Harvesting and RNA isolation were performed as previously described.[29] RNA quality and quantity was checked by an Agilent 2100 Bioanalyzer (Agilent Technologies, Böblingen, Germany) and Trinean Xpose system (Gentbrugge, Belgium) prior and after rRNA depletion by using the Ribo-Zero rRNA Removal Kit (Bacteria) (Illumina, San Diego, CA, USA). The TruSeq Stranded mRNA Library Prep Kit (Illumina, San Diego, CA, USA) was used to prepare the cDNA libraries, which were then sequenced in paired-end mode on an Illumina HiSeq 1500 system (San Diego, CA, USA) with 2 × 27 nt read length.

Read Mapping, Data Visualization and Analysis of Gene Expression. Reads were trimmed to a minimum of 20 bases in length using Trimmomatic V0.33.[30] Mapping of the trimmed reads to the *S. lividans* TK24 genome sequence[10] was performed with Bowtie 2 using standard settings.[31] ReadXplorer 2.2.0[32] was used for visualization of short read alignments and calculation of transcripts per million reads (TPM) of every scaffold-building gene of a secondary metabolite gene cluster.[33]

DNA Manipulation. General DNA cloning, DNA digestion with restriction enzymes and plasmid transformation into *Escherichia coli* were performed as described in Sambrook *et al.*[34] PCR fragments were amplified using MasterAmp Extra-Long DNA Polymerase Mix and buffers (Epicenter, USA). The primers used in this study were described in Table S1. Primers for PCR were designed using Clone Manager Professional Version 9 (Sci-Ed Software, USA). To construct new vectors, PCR and vector fragments were assembled either by T4 DNA ligase or Gibson ligation.[35] Recombinant constructs were verified by DNA sequencing by GATC Biotech. *Streptomyces* genomic DNA was isolated using Wizard Genomic DNA Purification Kit (Promega, USA).

Construction of *S. lividans* RedStrep1 Strain. Two DNA fragments flanking the *act* cluster were PCR amplified from the genomic DNA of *S. lividans* TK24. A 2.7-kb DNA fragment flanking the 3′ end of SLIV_12925 (Figure S1A) was amplified with the primers 12925Dir and 12925Rev (Table S4). This DNA fragment was digested using *Mfe*I and *Cla*I and cloned into pAMR4[15] resulting in pAct1. A 3-kb DNA fragment flanking the 3′ end of SLIV_13030 (Figure S1) was amplified with the primers 13030Dir and 130130Rev (Table S4). This DNA fragment was digested with *Afl*II and *Hin*dIII and cloned in pAcl1, resulting in pAct2. The plasmid was introduced into *S. lividans* TK24 by conjugation, and after several rounds of propagation *bpsA*-negative apramycin-resistant and kanamycin-sensitive colonies (indicating a double crossover event) were selected. The replacement of the *act* cluster by apramycin-resistance cassette was confirmed by Southern-blot hybridization (data not shown). One clone (*S. lividans* Δ*act*) was used for further deletion of the *red* cluster. Likewise, two DNA fragments flanking the *red* cluster were PCR amplified from the genomic DNA of *S. lividans* TK24. A 2.9-kb DNA fragment flanking the 5′ end of SLIV_09220 (Figure S1A) was amplified with the primers 9220Dir and 9220Rev (Table S4), digested with *Mfe*I and *Bam*HI and cloned into pSPR11, resulting in pRed1. The cloning vector pSPR11 was prepared by replacing the apramycin-resistance gene *aac(3)IV* in pAMR4 with synthetic spectinomycin-resistance gene *aadA* from the plasmid paat-saadA-oriT, containing *oriT* and *P-GG* and *G-CC* recombination sites for *phi*C31 recombinase.[36] A 3-kb DNA fragment flanking the 3′ end of SLIV_13030 (Figure S1A) was amplified with the primers 9115Dir and 9115Rev (Table S4), digested with *Afl*II and *Hin*dIII and cloned in pRed1 resulting in pRed2. The plasmid was introduced into *S. lividans* Δ*act* by conjugation with selection for spectinomycin resistance. After several rounds of sporulation, four *bpsA*-negative spectinomycin-resistant and kanamycin-sensitive colonies were selected, and the correct double crossover event in the *red* cluster was confirmed by Southern-blot hybridization (data not shown). One *S. lividans* Δ*act* Δ*red* clone was used for further studies. Both antibiotic resistance markers, the apramycin resistance marker with FRT sites for recombination,[37] and the spectinomycin resistance marker with *B-CC* and *P-GG* sites for recombination, were removed from the *S. lividans* Δ*act* Δ*red* strain by successive transformation of the plasmid pUWLHFLP, containing Flp recombinase[38] and pKHint31, containing *phi*C31 recombinase.[39] The removal of both markers was confirmed by antibiotic sensitivity tests after five rounds of nonselective sporulation. Elimination of the plasmids carrying recombinase genes was achieved after two rounds of nonselective propagation. The representative clone from the resulting strain, *S. lividans* Δ*act-FRT* Δ*red-BCC*, later termed RedStrep1, was verified by Southern blot hybridization (Figure S1B) using probes for both clusters.

Construction of Genes Encoding Hybrid Proteins. The products of genes SLIV_09075 and SLIV_09200 (*redZ*) were analyzed using ProDom online tool (http://prodom.prabi.fr/prodom/current/html/home.php) in order to predict protein domain boundaries. Accordingly, hybrid proteins of the following compositions were designed: (i) HR1 would contain C-terminal HTH domain of the SLIV_09075 gene product with a C-terminal linker (aa 1−104) fused to the putative receiver domain of RedZ (aa 1−115); (HR2 would contain C-

DOI: 10.1021/acssynbio.6b00353
ACS Synth. Biol. 2017, 6, 1026−1033

terminal HTH domain of the SLIV_09075 gene product (aa 1−78) fused to the putative receiver domain of RedZ followed by its C-terminal linker (aa 1−145). Parts of the corresponding genes encoding respective domains were amplified by PCR from the genomic DNA of *S. lividans* TK24 (primers given in Table S3, Supporting Information) and joined using Gibson ligation.[36] The resulting fragments were cloned into the pKC1218HygR vector under control of *ermE**p promoter, and verified by DNA sequencing.

Analysis of Secondary Metabolites Produced by the Recombinant Strains. Strains were grown for 3 days at 28 °C in 10 mL of TSB medium. Subsequently, 4 mL of the preculture was used to inoculate 50 mL of production media in 500 mL flasks: NL19, SG, R5A or phage medium. The main cultures were cultivated at 28 °C on a rotary shaker at 180 rpm for 5 days. Metabolites were extracted with acetone:methanol 1:1 from biomass and with ethyl acetate from cultural liquid, solvents were evaporated and extracts were dissolved in 300 μL of methanol.

Samples were separated on a Dionex Ultimate 3000 RSLC system (Thermoscientific) using a BEH C18, 100 × 2.1 mm, 1.7 μm d_p column (Waters) with linear gradient of solvent B (acetonitrile with 0.1% of formic acid) against solvent A (water with 0.1% of formic acid) at a flow rate of 600 μL/min and 45 °C. The gradient started by a 0.5 min isocratic step at 5% B then increased to 95% B over 9 min to end up with a 2 min isocratic step at 95% B before re-equilibration under the initial conditions. UV spectra were acquired by a DAD in the range of 200 to 600 nm. The mass-spec data was collected on an amaZON speed resolution mass spectrometer (Bruker Daltonics, Germany) using ESI source. Mass spectra were acquired in centroid mode ranging from 100 to 2000 m/z at a 2 Hz scan rate. Data was collected and analyzed with the Bruker Compass Data Analysis software, version 4.2 (Bruker, Billerica, USA).

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acssynbio.6b00353.

Figure S1 (construction of *S. lividans* RedStrep1 strain), Figure S2 (identification of coelimycin P1), Tables S1−S4 (descriptions of bacterial strains, plasmids, oligonucleotides, supporting main text) (PDF)

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: sergey.zotchev@univie.ac.at. Tel: +43 1 4277 55296. Fax: +43 1 4277 855296.

ORCID

Sergey B. Zotchev: 0000-0002-9324-245X

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

This work was supported by the European Commission under the 7th Framework Program through the "Collaborative Project" action "STREPSYNTH" Grant No. 613877, the University of Vienna and the Slovak Research and Development Agency under the contract No. DO7RP-0037-12. JK was also supported by the Slovak Research and Development Agency under the contract APVV-15-0410. We are thankful to Kristel Bernaerts and Wouter Daniels (KU Leuven, Belgium) for growing *S. lividans* TK24 for the RNA-seq experiments.

■ REFERENCES

(1) Hopwood, D. A. (2006) Soil to genomics: the *Streptomyces* chromosome. *Annu. Rev. Genet. 40*, 1−23.

(2) Zarins-Tutt, J. S., Barberi, T. T., Gao, H., Mearns-Spragg, A., Zhang, L., Newman, D. J., and Goss, R. J. (2016) Prospecting for new bacterial metabolites: a glossary of approaches for inducing, activating and upregulating the biosynthesis of bacterial cryptic or silent natural products. *Nat. Prod. Rep. 33*, 54−72.

(3) Katz, L., and Baltz, R. H. (2016) Natural product discovery: past, present, and future. *J. Ind. Microbiol. Biotechnol. 43*, 155−176.

(4) Sekurova, O. N., Pérez-Victoria, I., Martín, J., Degnes, K. F., Sletta, H., Reyes, F., and Zotchev, S. B. (2016) New Deferoxamine Glycoconjugates Produced upon Overexpression of Pathway-Specific Regulatory Gene in the Marine Sponge-Derived *Streptomyces albus* PVA94−07. *Molecules 21*, 1131.

(5) Cuthbertson, L., and Nodwell, J. R. (2013) The TetR family of regulators. *Microbiol. Mol. Biol. Rev. 77*, 440−475.

(6) Hansen, L. H., Ferrari, B., Sørensen, A. H., Veal, D., and Sørensen, S. J. (2001) Detection of oxytetracycline production by *Streptomyces rimosus* in soil microcosms by combining whole-cell biosensors and flow cytometry. *Appl. Environ. Microbiol. 67*, 239−244.

(7) Ostash, B., Ostash, I., Zhu, L., Kharel, M. K., Luzhetskiĭ, A., Bechthold, A., Walker, S., Rohr, J., and Fedorenko, V. (2010) Properties of lanK-based regulatory circuit involved in landomycin biosynthesis in *Streptomyces cyanogenus* S136. *Russ. J. Genet. 46*, 604−609.

(8) Le, T. B., Fiedler, H. P., den Hengst, C. D., Ahn, S. K., Maxwell, A., and Buttner, M. J. (2009) Coupling of the biosynthesis and export of the DNA gyrase inhibitor simocyclinone in *Streptomyces antibioticus*. *Mol. Microbiol. 72*, 1462−1474.

(9) Li, Y., Li, J., Tian, Z., Xu, Y., Zhang, J., Liu, W., and Tan, H. (2016) Coordinative Modulation of chlorothricin biosynthesis by binding of the glycosylated intermediates and end product to a responsive regulator ChlF1. *J. Biol. Chem. 291*, 5406−5417.

(10) Guo, F., Xiang, S., Li, L., Wang, B., Rajasarkka, J., Grondahl-Yli-Hannuksela, K., Ai, G., Metsa-Ketela, M., and Yang, K. (2015) Targeted activation of silent natural product biosynthesis pathways by reporter-guided mutant selection. *Metab. Eng. 28*, 134−142.

(11) Rückert, C., Albersmeier, A., Busche, T., Jaenicke, S., Winkler, A., Friethjonsson, O. H., Hreggviethsson, G. O., Lambert, C., Badcock, D., Bernaerts, K., Anne, J., Economou, A., and Kalinowski, J. (2015) Complete genome sequence of *Streptomyces lividans* TK24. *J. Biotechnol. 199*, 21−22.

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

(13) Creecy, J. P., and Conway, T. (2015) Quantitative bacterial transcriptomics with RNA-seq. *Curr. Opin. Microbiol. 23*, 133−140.

(14) Williamson, N. R., Fineran, P. C., Leeper, F. J., and Salmond, G. P. (2006) The biosynthesis and regulation of bacterial prodiginines. *Nat. Rev. Microbiol. 4*, 887−899.

(15) Knirschova, R., Novakova, R., Mingyar, E., Bekeova, C., Homerova, D., and Kormanec, J. (2015) Utilization of a reporter system based on the blue pigment indigoidine biosynthetic gene *bpsA* for detection of promoter activity and deletion of genes in *Streptomyces*. *J. Microbiol. Methods 113*, 1−3.

(16) Takano, E., Gramajo, H. C., Strauch, E., Andres, N., White, J., and Bibb, M. J. (1992) Transcriptional regulation of the redD transcriptional activator gene accounts for growth-phase-dependent production of the antibiotic undecylprodigiosin in Streptomyces coelicolor A3(2). *Mol. Microbiol. 6*, 2797−2804.

(17) Chen, J., and Xie, J. (2011) Role and regulation of bacterial LuxR-like regulators. *J. Cell. Biochem. 112*, 2694−2702.
(18) Patel, H. K., Suárez-Moreno, Z. R., Degrassi, G., Subramoni, S., González, J. F., and Venturi, V. (2013) Bacterial LuxR solos have evolved to respond to different molecules including signals from plants. *Front. Plant Sci. 4*, 447.
(19) Keeley, M. B., Busch, J., Singh, R., and Abel, T. (2005) TetR hybrid transcription factors report cell signaling and are inhibited by doxycycline. *BioTechniques 39*, 529−536.
(20) Chou, H. H., and Keasling, J. D. (2013) Programming adaptive control to evolve increased metabolite production. *Nat. Commun. 4*, 2595.
(21) Gomez-Escribano, J. P., Song, L., Fox, D. J., Yeo, V., Bibb, M. J., and Challis, G. L. (2012) Structure and biosynthesis of the unusual polyketide alkaloid coelimycin P1, a metabolic product of the *cpk* gene cluster of *Streptomyces coelicolor* M145. *Chem. Sci. 3*, 2716−2720.
(22) Gottelt, M., Kol, S., Gomez-Escribano, J. P., Bibb, M., and Takano, E. (2010) Deletion of a regulatory gene within the *cpk* gene cluster reveals novel antibacterial activity in *Streptomyces coelicolor* A3(2). *Microbiology 156*, 2343−2353.
(23) Kotowska, M., Ciekot, J., and Pawlik, K. (2014) Type II thioesterase ScoT is required for coelimycin production by the modular polyketide synthase Cpk of *Streptomyces coelicolor* A3(2). *Acta Biochim. Pol. 61*, 141−147. (https://www.ncbi.nlm.nih.gov/pubmed/24660171)
(24) Bierman, M., Logan, R., O'Brien, K., Seno, E. T., Rao, R. N., and Schoner, B. E. (1992) Plasmid cloning vectors for the conjugal transfer of DNA from *Escherichia coli* to *Streptomyces* spp. *Gene 116*, 43−49.
(25) Rutledge, P. J., and Challis, G. L. (2015) Discovery of microbial natural products by activation of silent biosynthetic gene clusters. *Nat. Rev. Microbiol. 13*, 509−523.
(26) Takano, E., Kinoshita, H., Mersinias, V., Bucca, G., Hotchkiss, G., Nihira, T., Smith, C. P., Bibb, M., Wohlleben, W., and Chater, K. (2005) A bacterial hormone (the SCB1) directly controls the expression of a pathway-specific regulatory gene in the cryptic type I polyketide biosynthetic gene cluster of *Streptomyces coelicolor*. *Mol. Microbiol. 56*, 465−479.
(27) Engelhardt, K., Degnes, K. F., and Zotchev, S. B. (2010) Isolation and characterization of the gene cluster for biosynthesis of the thiopeptide antibiotic TP-1161. *Appl. Environ. Microbiol. 76*, 7093−7101.
(28) Swint-Kruse, L., and Matthews, K. S. (2009) Allostery in the LacI/GalR family: variations on a theme. *Curr. Opin. Microbiol. 12*, 129−137.
(29) Busche, T., Silar, R., Picmanova, M., Patek, M., and Kalinowski, J. (2012) Transcriptional regulation of the operon encoding stress-responsive ECF sigma factor SigH and its anti-sigma factor RshA, and control of its regulatory network in Corynebacterium glutamicum. *BMC Genomics 13*, 445.
(30) Kieser, T., Bibb, M. J., Buttner, M. J., Chater, K. F., and Hopwood, D. A. (2000) *Practical Streptomyces Genetics*, John Innes Foundation, Norwich.
(31) Bolger, A. M., Lohse, M., and Usadel, B. (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics 30*, 2114−2120.
(32) Langmead, B., and Salzberg, S. L. (2012) Fast gapped-read alignment with Bowtie 2. *Nat. Methods 9*, 357−359.
(33) Hilker, R., Stadermann, K. B., Doppmeier, D., Kalinowski, J., Stoye, J., Straube, J., Winnebald, J., and Goesmann, A. (2014) ReadXplorer−visualization and analysis of mapped sequences. *Bioinformatics 30*, 2247−2254.
(34) Wagner, G. P., Kin, K., and Lynch, V. J. (2012) Measurement of mRNA abundance using RNA-seq data: RPKM measure is inconsistent among samples. *Theory Biosci. 131*, 281−285.
(35) Sambrook, J., Fritsch, E., and Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual*, 2nd ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor.
(36) Gibson, D. G., Young, L., Chuang, R. Y., Venter, J. C., Hutchison, C. A., 3rd, and Smith, H. O. (2009) Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat. Methods 6*, 343−345.
(37) Myronovskyi, M., Rosenkranzer, B., and Luzhetskyy, A. (2014) Iterative marker excision system. *Appl. Microbiol. Biotechnol. 98*, 4557−4570.
(38) Gust, B., Challis, G. L., Fowler, K., Kieser, T., and Chater, K. F. (2003) PCR-targeted *Streptomyces* gene replacement identifies a protein domain needed for biosynthesis of the sesquiterpene soil odor geosmin. *Proc. Natl. Acad. Sci. U. S. A. 18*, 1541−1548.
(39) Fedoryshyn, M., Petzke, L., Welle, E., Bechthold, A., and Luzhetskyy, A. (2008) Marker removal from actinomycetes genome using Flp recombinase. *Gene 419*, 43−47.

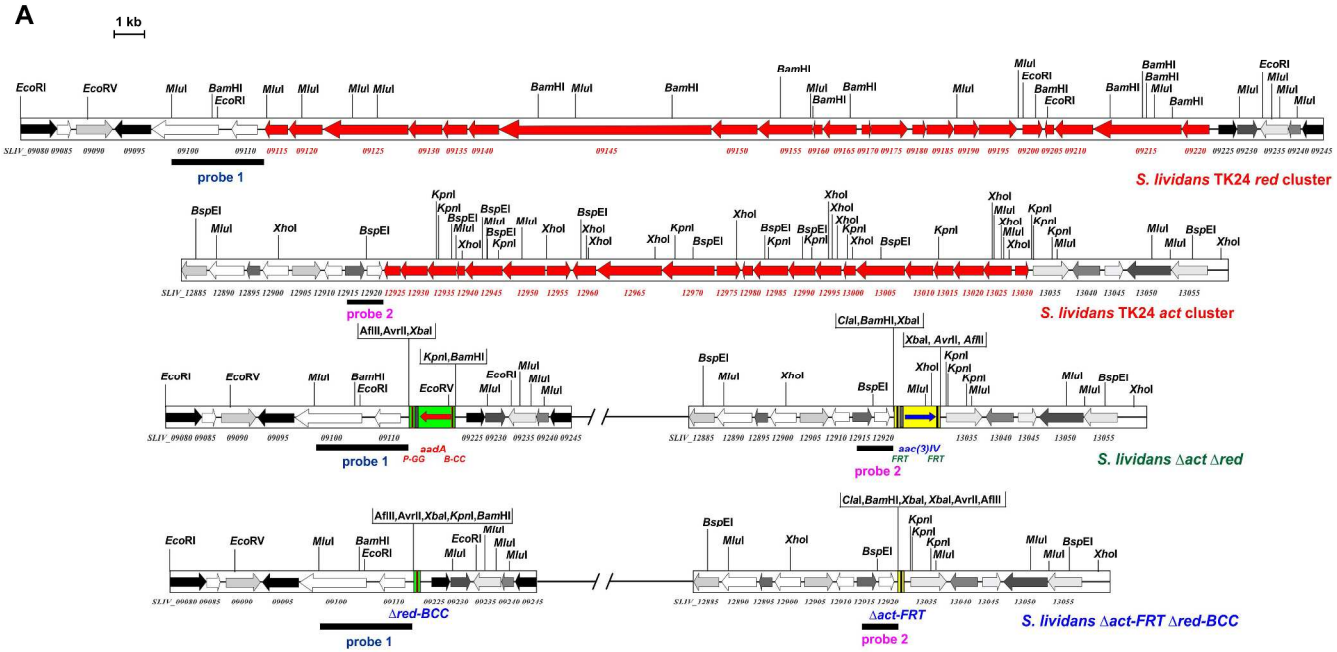
DOI: 10.1021/acssynbio.6b00353
ACS Synth. Biol. 2017, 6, 1026−1033

Supporting Information

Sun et al. **Development of a Biosensor Concept to Detect Production of Cluster-Specific Secondary Metabolites**

Figure S1. (A) Physical maps of chromosomal DNA containing the wild-type *S. lividans* TK24 *act* cluster for actinorhodin and *red* cluster for undecylprodigiosin, and the deleted alleles of the *S. lividans*, Δ*act* Δ*red* and *S. lividans*, Δ*act-FRT* Δ*red-BCC* strains. Thick arrows denote the direction and size of genes. Red arrows represent the *act* and *red* genes. Gene labelling is based on the genomic sequence of *S. lividans* TK24 (GenBank Acc. No. CP009124)[1]. The black bars below the maps represent the probes used for Southern hybridization analysis. Relevant restriction sites are indicated.

(**B)** Southern hybridization analysis of chromosomal DNA from the indicated strains verifying the correct integration and removing of the antibiotic cassette. 1 μg of gDNA from the corresponding strain was digested with the restriction enzymes indicated, separated by electrophoresis in 0.8% (w/v) agarose gel and transferred on Hybond N (Amersham) as described in Ausubel et al.[2] Hybridization followed the standard DIG protocol (Roche, Mannheim, Germany) using the DIG-labelled probes 1 and 2 for verification of correct *red* and *act* deletion, respectively. Lambda DNA-*Bst*EII digest was used as the size standard.


A
1 kb
probe 1
S. lividans TK24 red cluster
probe 2
S. lividans TK24 act cluster
probe 1
probe 2
S. lividans Δact Δred
probe 1
probe 2
S. lividans Δact-FRT Δred-BCC

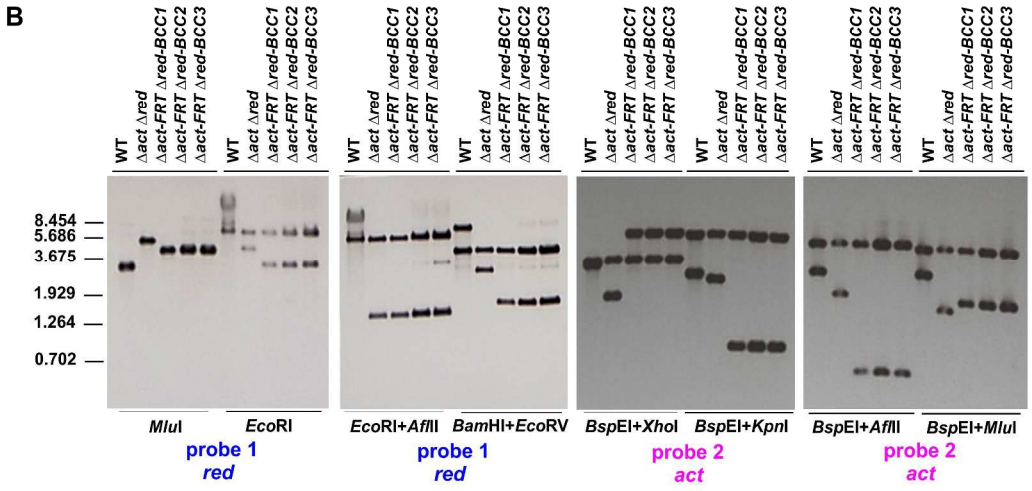

B
WT
Δact Δred
Δact-FRT Δred-BCC1
Δact-FRT Δred-BCC2
Δact-FRT Δred-BCC3
8.454
5.686
3.675
1.929
1.264
0.702
MluI
EcoRI
probe 1
red
EcoRI+AflII
BamHI+EcoRV
probe 1
red
BspEI+XhoI
BspEI+KpnI
probe 2
act
BspEI+AflII
BspEI+MluI
probe 2
act

Figure S2. Absorptions, accurate mass and MS/MS profile of metabolite detected in the extract of *S. lividans* RedStrep1 [pYQS040/pYQS060] strain. Detected compound has m/z of 349.1213 $[M+H]^+$ (calculated for coelimycin P1 (C17H20N2)4S) 349.1215 $[M+H]^+$) and three UV maximums at 225, 275 and 360 nm with peak at 360 nm being characteristic for coelimycin P1[3]. The MS/MS fragmentation pattern is the same as described for coelimycin P1[4].

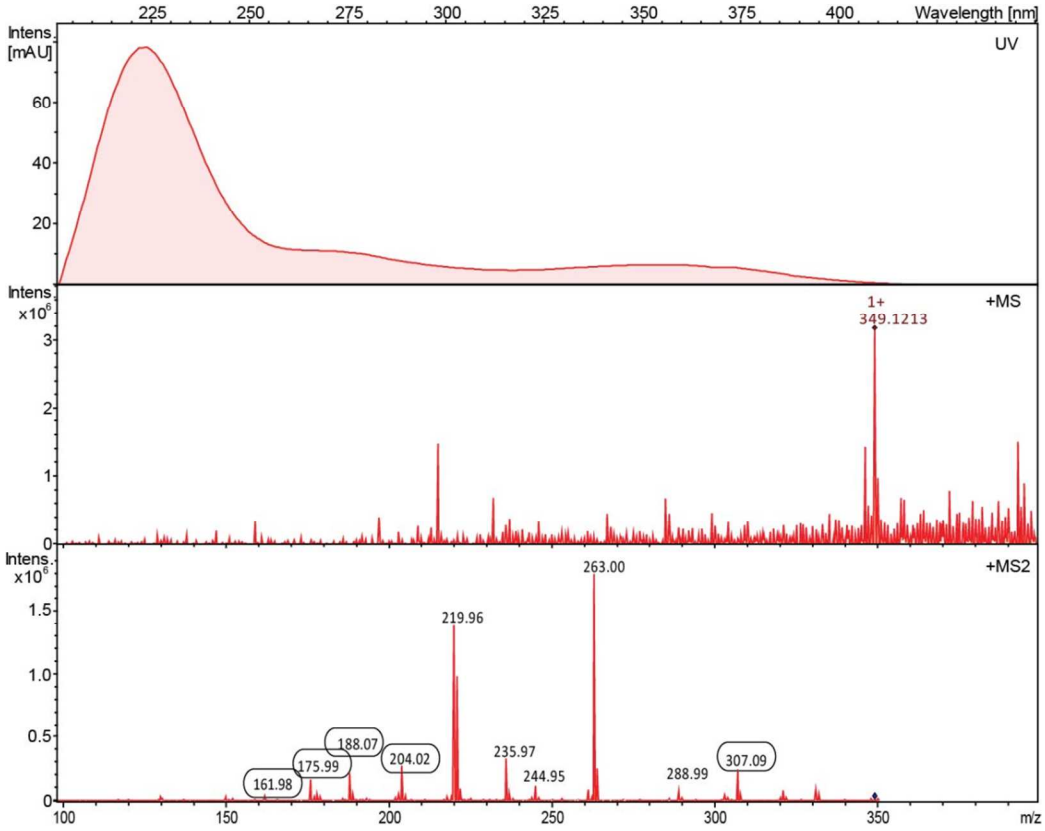


Table S1. Plasmids used in this study.

Plasmid name	Description	Construction details	Source
pSET152	Integrative *E.coli-Streptomyces* vector, *ϕC31* attachment site, Am^R		5
pUWL-oriT	*E.coli- Streptomyces* shuttle vector, Amp^R, Tsr^R		6
pBpsA	Integrative *E.coli-Streptomyces* vector, *ϕBT1* attachment site, Am^R		7
pKC1218 hyg_mut	*E.coli- Streptomyces* shuttle vector, Hyg^R	Am^R marker replaced with HygR, one *Eco*RI site eliminated	This study
pTSA101	Intermediate/construction vector	pKC1218hyg_mut backbone with *bpsA* gene under control of P21 promoter and CymR operator	This study
pYQS031	Intermediate/construction vector	Ligation of PCR PermE-*dnaQ2* with pKC1218 hyg_mut (HindIII/EcoRI)	This study
pYQS040	SLIV_06710 PO - BpsA in pSET152 backbone	Ligation: pSET152 (BamHI/EcoRI) + PCR SLIV_06710 PO (BamHI/SpeI) + pTSA101 (SpeI/EcoRI, *bpsA*)	This study
pYQS041	SLIV_06815 PO - BpsA in pSET152 backbone	Ligation: pSET152 (BamHI/EcoRI) + PCR SLIV_06815 PO (BamHI/SpeI) + pTSA101 (SpeI/EcoRI, *bpsA*)	This study
pYQS042	SLIV_09080 PO - BpsA in pSET152 backbone	Ligation: pSET152 (BamHI/EcoRI) + PCR SLIV_09080 PO (BamHI/SpeI) + pTSA101 (SpeI/EcoRI, *bpsA*)	This study
pYQS046	Backbone was used for double crossover to knockout target gene	Ligation: PCR dnaQ1 flanking A (AgeI/BamHI) + PCR *dnaQ1* flanking B (BamHI/HindIII) + pTSA101 (HindIII/AgeI)	This study
pYQS056	SLIV_31825 PO - *bpsA* in pSET152 backbone	Ligation: pSET152 (BamHI/EcoRI) + PCR SLIV_31825 PO (BamHI/SpeI) + pTSA101 (SpeI/EcoRI, *bpsA*)	This study
pYQS060	Cluster specific positive regulator: PermE-SLIV_06705 in pKC1218, Hyg^R	Ligation of pYQS031 with PCR SLIV_06705 (EcoRI/XbaI)	This study

pYQS061	Cluster specific positive regulator: PermE-SLIV_06745 in pKC1218, Hyg^R	Ligation of pYQS031 with PCR SLIV_06745 (EcoRI/XbaI)	This study
pYQS062	Cluster specific positive regulator: PermE-SLIV_09200 in pKC1218, Hyg^R	Ligation of pYQS031 (EcoRI/XbaI) with PCR SLIV_09200 (MfeI/XbaI)	This study
pYQS063	Cluster specific positive regulator: PermE-SLIV_09220 *redD* in pKC1218, Hyg^R	Ligation of pYQS031 with PCR SLIV_09220 RedD (EcoRI/XbaI)	This study
pYQS064	Cluster specific positive regulator: PermE-SLIV_31695 in pKC1218, Hyg^R	Ligation of pYQS031 with PCR SLIV_31695 (EcoRI/XbaI)	This study
pYQS065	Used to construct hybrid repressor HR1: REC SLIV_09200 - linker SLIV_09200 - HTH SLIV_09075	Gibson assembly: Reverse PCR pYQS046 + PCR SLIV_09075 flanking A1 + PCR HR1-frag4 + PCR HR1-frag1 + PCR SLIV_09075 flanking B1	This study
pYQS066	Used to construct hybrid repressor HR2: HTH SLIV_09075 - linker SLIV_09075 - REC SLIV_09200	Gibson assembly: Reverse PCR pYQS046 + PCR SLIV_09075 flanking A2 HR2-frag2 + PCR HR2-frag3 + PCR SLIV_09075 flanking B2	This study
pYQS070	single crossover plasmid for insertion knockout of SLIV_06770, Amp^R and Tsr^R	Ligation PCR SLIV_06770 middle with pUWL (NdeI/BamHI)	This study
pYQS079	PermE-HR1, pKC1218 backbone, Hyg^R	Ligation of PCR Hybrid Repressor 1 (XbaI/MfeI) with pYQS060 (XbaI/EcoRI)	This study
pYQS080	PermE-HR2, pKC1218 backbone, Hyg^R	Ligation PCR Hybrid Repressor 2 (XbaI/MfeI) with pYQS060 (XbaI/EcoRI)	This study

PO: promoter-operator. HR: hybrid repressor. PR: putative positive regulator

Table S2. *Escherichia coli* and *Streptomyces lividans* strains used in this study.

Strain name	Description	Plasmid(s) used for conjugation	Source/reference
Escherichia coli XL1-blue	General cloning host: *recA1 endA1 gyrA96 thi-1 hsdR17 supE44 relA1 lac* [F′ *proAB lacI*^{q} ZΔ*M15* Tn*10* (Tet^{R})]		Stratagene
Escherichia coli ET125671 (pUZ8002)	Mediates conjugative DNA transfer from RP4 *oriT* with helper plasmid pUZ8002 (Kan^{R}, Cm^{R}). Methylation deficient (*dam*$^{-}$, *dcm*$^{-}$, *hsdM*$^{-}$)		8
S. lividans TK24	*S. lividans* wild-type strain		2
RedStrep1	TK24 *Δact Δred*	Deleted clusters for actinorhodine and undecylprodigiosin	This study
TK24 + Reporter 1	TK24 *attB$_{C31}$ :: SLIV_06710$_{PO}$-bpsA*	pYQS040, *attB$_{C31}$* integrative	This study
TK24 + Reporter 2	TK24 *attB$_{C31}$ :: SLIV_06815$_{PO}$-bpsA*	pYQS041, *attB$_{C31}$* integrative	This study
TK24 + Reporter 3	TK24 *attB$_{C31}$ :: SLIV_09080$_{PO}$-bpsA*	pYQS042, *attB$_{C31}$* integrative	This study
RedStrep 1 + Reporter 1	RedStrep 1 *attB$_{C31}$ :: SLIV_06710$_{PO}$-bpsA*	pYQS040, *attB$_{C31}$* integrative	This study
RedStrep 1 + Reporter 2	RedStrep 1 *attB$_{C31}$ :: SLIV_06815$_{PO}$-bpsA*	pYQS041, *attB$_{C31}$* integrative	This study
RedStrep 1 + Reporter 3	RedStrep 1 *attB$_{C31}$ :: SLIV_09080$_{PO}$-bpsA*	pYQS042, *attB$_{C31}$* integrative	This study
RedStrep 1 + Reporter 4	RedStrep 1 *attB$_{C31}$ :: SLIV_31825$_{PO}$-bpsA*	pYQS056, *attB$_{C31}$* integrative	This study
RedStrep 1 + Reporter 1+ PR1	RedStrep 1 SLIV_06710$_{PO}$-BpsA + ermEp-SLIV_06705	pYQS060, replicative	This study
RedStrep 1 + Reporter 1+ PR2	RedStrep 1 SLIV_06710$_{PO}$-BpsA + ermEp-SLIV_06745	pYQS061, replicative	This study
RedStrep 1 + Reporter 4+ PR5	RedStrep 1 SLIV_31825$_{PO}$-BpsA + ermEp-SLIV_31695	pYQS064, replicative	This study
TK24 + Reporter 3 + PR3	RedStrep 1 SLIV_09080$_{PO}$-BpsA + ermEp-SLIV_09200	pYQS062, replicative	This study

TK24 + Reporter 3 + PR4	RedStrep 1 SLIV_09080$_{PO}$-BpsA + ermEp-SLIV_09220 RedD	pYQS063, replicative	This study
TK24 *ΔSLIV_09075 :: HR1*	TK24 *ΔSLIV_09075 :: Hybrid repressor_1*	pYQS065, for deletion knockout by double crossover	This study
TK24 *ΔSLIV_09075 :: HR2*	TK24 *ΔSLIV_09075 :: Hybrid repressor_2*	pYQS066, for deletion knockout by double crossover	This study
TK24 *ΔSLIV_09075 :: HR1* + Reporter 3	TK24 *ΔSLIV_09075 :: Hybrid repressor_1* + *SLIV_09080$_{PO}$-BpsA*	pYQS042, *attB$_{C31}$* integrative	This study
TK24 *ΔSLIV_09075 :: HR2* + Reporter 3	TK24 *ΔSLIV_09075 :: Hybrid repressor_1* + *SLIV_09080$_{PO}$-BpsA*	pYQS042, *attB$_{C31}$* integrative	This study
RedStrep 1 + Reporter 1 + PR1 *ΔSLIV_06770*	SLIV_06770 was knocked out in "RedStrep 1 + Reporter 1 + PR1"	pYQS070, for insertion knockout by single crossover	This study
RedStrep 1 + Reporter 3+ *HR1*	RedStrep 1 *SLIV_09080$_{PO}$-BpsA* + *ermEp-HR1*	pYQS079, replicative	This study
RedStrep 1 + Reporter 3 + *HR2*	RedStrep 1 *SLIV_09080$_{PO}$-BpsA* + *ermEp-HR2*	pYQS080, replicative	This study

$_{PO}$: promoter-operator. *HR*: hybrid repressor. PR: putative Positive Regulator

HR1: REC SLIV_09200 - linker SLIV_09200 - HTH SLIV_09075

HR2: HTH SLIV_09075 - linker SLIV_09075 - REC SLIV_09200

Table S3. Primers for plasmid construction used in this study.

Primer name	5'-3' sequence (*Italic*: endonuclease restriction enzyme site)	PCR product (size bps, template)	Enzyme/paired primer
YQ30_SE	TAAT*GGATCC*GTCGTGCTCCGTGGTCGC	PCR SLIV_06710 PO (258, TK24)	*BamHI*
YQ30_AS	ACACC*ACTAGT*CGCCTTTTCCCCTTACCGTTC		*SpeI*
YQ31_SE	TATA*GGATCC*GCCTGCCTCCTTGTTCAT	PCR SLIV_06815 PO (137, TK24)	*BamHI*
YQ31_AS	CAGT*ACTAGT*GGGTCCCCCCCAGGAATC		*SpeI*
YQ32_SE	AGAT*GGATCC*TTTCTCCGCTTCACCTCG	PCR SLIV_09080 PO (164, TK24)	*BamHI*
YQ32_AS	AATT*ACTAGT*CGGGCAGCCTTCGGCAGGA		*SpeI*
YQ42_SE	CAAT*GGATCC*GACGCCTCCTTCCAGTG	PCR SLIV_31825 PO (120, TK24)	*BamHI*
YQ42_AS	ACACC*ACTAGT*GAACCCCCACCTCCTTAAG		*SpeI*
YQ47_SE	ATCA*TCTAGA*CATCCGCGGCACGGCAAAC	PCR SLIV_06705 (927, TK24)	*XbaI*
YQ47_AS	GTACT*GAATTC*CGAGATCGGGCGCGTCAC		*EcoRI*
YQ48_SE	CGCA*TCTAGA*AGTCGTGGCCAGGAGAATAC	PCR SLIV_06745 (1743, TK24)	*XbaI*
YQ48_AS	ATACT*GAATTC*GACGGCGGAGACGGTGGG		*EcoRI*
YQ49_SE	ATCA*TCTAGA*GCCCGCACGCCACGTACGGT	PCR SLIV_09200 (739, TK24)	*XbaI*
YQ49_AS	ATACT*CAATTG*GCCGGACCGGGGCATTCAGC		*MfeI*
YQ50_SE	ATCA*TCTAGA*CCGATCGTTCGGTGGATGAC	PCR SLIV_09220 *redD* (1097, TK24)	*XbaI*
YQ50_AS	ATACT*GAATTC*CCGGGTCGTCAGGCGCTGAG		*EcoRI*
YQ51_SE	GGCA*TCTAGA*ATGCGGGCACTGTATCCAGG	PCR SLIV_31695 (863, TK24)	*XbaI*
YQ51_AS	ATACT*GAATTC*CTCACGGCCTCCGCATGAGC		*EcoRI*
YQ54_SE	GATAAGCTTGGGCTGCAGGTC	Reverse PCR pYQS046 (7884, pYQS046)	
YQ54_AS	ACCGAGTCCCACCGGTCGAAAC		

YQ55_SE	TTCGACCGGTGGGACTCGGTGCTGGAGCGCAGGTTGTC	PCR SLIV_09075 flanking A1 (1530, TK24)	
YQ55_AS	GGGTCGTCATTTTCTCCGCTTCACCTCGG		
YQ56_SE	AGCGGAGAAAATGACGACCCGTGTCCTG	PCR HR1-frag4 (455, TK24)	
YQ56_AS	CAATCTTGGGTTCGGGGCGGAGGTTCTTC		
YQ57_SE	CCGCCCCGAACCCAAGATTGAAGCCGGC	PCR HR1-frag1 (218, TK24)	
YQ57_AS	GGGTGCGTCAGACGACGAGTTCGAGCAGG		
YQ58_SE	ACTCGTCGTCTGACGCACCCCGTCCAGG	PCR SLIV_09075 flanking B1 (1530, TK24)	
YQ58_AS	ACCTGCAGCCCAAGCTTATCCGGCGGTCCTCTCGTCGG		
YQ59_AS	CACGGGTCGTGCGAGCTGCCTGTTCGAG	PCR SLIV_09075 flanking A2 HR2-frag2 (1845, TK24)	YQ55_SE
YQ60_SE	GGCAGCTCGCACGACCCGTGTCCTGGTG	PCR HR2-frag3 (359, TK24)	
YQ60_AS	GGGTGCGTCAGATGGTTCTGATGACGTGCAC		
YQ61_SE	CAGAACCATCTGACGCACCCCGTCCAGG	PCR SLIV_09075 flanking B2 (1530, TK24)	YQ58_AS
YQ68_SE	ATCA*CATATG*AACTTCCCGCCAGCCTGAAC	PCR SLIV_06770 middle (2142, TK24)	*NdeI*
YQ68_AS	ATAA*GGATCC*AAGAAGTCGGCGTCGAACTC		*BamHI*
YQ75_SE	ACGGTGTCTCGTACACCTTC	Amplify SLIV_06770 to identify insertion knockout (9894)	
YQ75_AS	TTCCGTAGGTGGCCATGAGG		
YQ80_SE	GCCA*TCTAGA*GGTGAAGCGGAGAAAATGACGA	PCR Hybrid Repressor 1 (678, pYQS065)	*XbaI*
YQ80_AS	CTACT*CAATTG*GGTGCGTCAGACGACGAGTT		*MfeI*
YQ81_SE	GCCA*TCTAGA*AGGTGAAGCGGAGAAAATGC	PCR Hybrid Repressor 2 (696, pYQS066)	*XbaI*
YQ81_AS	CCGCT*CAATTG*CGTCAGATGGTTCTGATGAC		*MfeI*

Table S4. Primers used in construction of *S. lividans* RedStrep1.

Oligonucleotide	Sequence (5′ − 3′)
12925Dir	CCCCCAATTGCGCTGAGCAAGCAGATCCGGGCCC
129215Rev	CCCCATCGATGGAAGGCCGTCAGCGGCCCGTGGC
13030Dir	CCCCAAGCTTCGAAGACCGCCAGGGACGTCAGCC
13013Rev	CCCCCTTAAGTGACGCGCACCCTCGCCTTCGCGC
9220Dir	CCCCCAATTGCCGGTACGCTTCAGATATGAGTGC
9220Rev	CCCCGGATCCACCGTCAATGTGTTGAGGCCG
9115Dir	CCCCCTTAAGCGTGTACTGACCCCGCCCGTCACG
9115Rev	CCCCAAGCTTCGATCTCCTCGGCGACCCGGTCG

References

(1) Ausubel, F.M., Brent, R., Kingston, R.E., Moore, D.O., Seidman, J.S., Smith, J.A., and Struhl, K. (1995) Current Protocols in Molecular Biology. Wiley, New York.

(2) Ruckert, C., Albersmeier, A., Busche, T., Jaenicke, S., Winkler, A., Friethjonsson, O. H., Hreggviethsson, G. O., Lambert, C., Badcock, D., Bernaerts, K., Anne, J., Economou, A., and Kalinowski, J. (2015) Complete genome sequence of Streptomyces lividans TK24, *J Biotechnol 199*, 21-22.

(3) Whittle, M., Willett, P., Klaffke, W. and van Noort, P. (2003) Evaluation of similarity measures for searching the Dictionary of Natural Products database. *J. Chem. Inform. Comp. Sci. 43*, 449-457.

(4) Gomez-Escribano, J.P., Song, L., Fox, D.J., Yeo, V., Bibb, M.J., and Challis, G.L. (2012) Structure and biosynthesis of the unusual polyketide alkaloid coelimycin P1, a metabolic product of the *cpk* gene cluster of *Streptomyces coelicolor* M145. *Chem. Sci. 3*, 2716–2720.
(5) Bierman, M., Logan, R., O'Brien, K., Seno, E. T., Rao, R. N., and Schoner, B. E. (1992) Plasmid cloning vectors for the conjugal transfer of DNA from Escherichia coli to *Streptomyces* spp. *Gene 116*, 43-49.

(6) Luzhetskyy, A., Zhu, L., Gibson, M., Fedoryshyn, M., Durr, C., Hofmann, C., Hoffmeister, D., Ostash, B., Mattingly, C., Adams, V., Fedorenko, V., Rohr, J., and Bechthold, A. (2005) Generation of novel landomycins M and O through targeted gene disruption, *Chembiochem 6*, 675-678.
(7) Knirschova, R., Novakova, R., Mingyar, E., Bekeova, C., Homerova, D., and Kormanec, J. (2015) Utilization of a reporter system based on the blue pigment indigoidine biosynthetic gene *bpsA* for detection of promoter activity and deletion of genes in *Streptomyces. J. Microbiol. Methods 113*, 1-3.
(8) Flett F., Mersinias, V., and Smith, C.P. (1997) High efficiency intergeneric conjugal transfer of plasmid DNA from *Escherichia coli* to methyl DNA-restricting streptomycetes. *FEMS Microbiol. Lett. 155*, 223-229.

5. Discussion

New pharmaceuticals for the treatment of infectious diseases and especially those caused by multi-resistant pathogens are of urgent need. Even carbapenems, long considered as invulnerable, are of risk to loose effectiveness due to the spread of new types of β-lactamases which makes some of the Gram-negative bacteria untreatable.[1] Microorganisms such as actinomycetes are a rich source of secondary metabolites whereof many possess antibacterial, antiviral, anticancer or immunosuppressive properties. The decades of exploitation of these bacteria species lead to the discovery of many pharmaceutical and medicinally important drugs. NPs remain one of the major sources of biologically active compounds with potential to combat antimicrobials resistance. However, to find a long-term solution for the issue of fast developing resistance, novel NP scaffolds introducing an unusual mode of action would be needed. The discovery of such compound classes became unfortunately difficult. The overexploitation of soil actinomycetes leads meanwhile to a frequent re-isolation of already known NPs. Therefore, the exploration of unusual sources of new bacteria is necessary in order to increase the chances to find novel secondary metabolites with unique chemical scaffolds and thus enhanced biological activities.

In the attempt to discover novel compounds, we have examined several actinobacteria strains obtained from different sources. The first screening covered 15 strains from DSMZ (Deutsche Sammlung von Mikroorganismen und Zellkulturen) and 9 strains isolated from sediment samples collected at Trondheim Fjord.[2,3] Genomes of these actinobacteria were sequenced and based on phylogenetic analysis 10 of them belong to the genus *Streptomyces* and 14 are representatives of other minor genera of actinomycetes (*Dactylsporangium*, *Nocardiopsis*, *Lechevaleria*, *Actinopolyspora*, *Actinomadura*, *Micromonospora*, *Kribella*, *Actinoalluteichus*, *Verrucosisspora* and *Glycomyces)*. Dereplication of extracts from these strains has revealed that several of them produce known metabolites e.g. nonactins, neocarzilin, matlystatin, flavofungin, lynamicins, napyradiomycin, rapamycin, azalomycin, rosamicin as well as putatively new NPs. One strain, *Streptomyces sp.* MP131-18 was considered to be the most promising based on the metabolic profiling and genome mining. The potent anticancer activity of bisindole compounds lynamicins found in the extracts of *Streptomyces sp.* MP131-18 prompted us to investigate the strain in more detail. In a second activity-guided screening program, 36 strains obtained from macroinvertebrates collected from Lake Baikal[4] were tested against Gram-positive *(Bacillus subtilis)* and Gram-negative (*Pseudomonas putida and Escherichia Coli*) bacteria. Due to the strong inhibition of *B. subtilis* by *Streptomyces sp.*

IB2014/011-12, we decided that this strain would be a good candidate for discovery of new NPs. The thorough examination of the two strains MP131-18 and IB 2014/011-12 lead to the discovery of new derivatives of spiroindimicins, lagunapyrones and alpiniamides. Spiroindimicins, belonging to the well-examined compound class indoles, were previously identified in *Streptomyces sp.* SCSIO 03032 isolated from the Bay of Bengal in Indian Ocean.[5] Similarly, alpiniamide A, an amino acid bearing linear polyketide has been discovered before in the endophytic *Streptomycetes sp*. YIM66017.[6] Although, both streptomycetes strains MP131-18 and IB 2014/011-12 were indeed new, the isolated metabolites were previously identified from the extracts of very distinct strains sampled from distant locations. Apparently, even the shift from soil actinomycetes to their aquatic counterparts does not prevent the frequent re-discovery of known NP scaffolds. One of the most possible reasons for the re-discovery of the same secondary metabolites from different actinobacteria species might be an active exchange of genomic material among them. As such, we found alpiniamide gene cluster in 8 other *Strepomyces* sp. isolated from different regions e.g. *Streptomyces* sp. CFMR 7 (Penang, Malaysia)[7], *S. cavourensis* TJ430 (Yunnan province, China)[8], *Streptomyces* sp. DUT11 (marine sediment, Xinghai Bay, China).[9] At the same time, four strains *Streptomyces* sp. Tü6075, *Streptomyces globisporus* TFH56, *Kitasatospora albolonga YIM101047* and *Streptomyces* sp. ZFG47 are missing the *alp* cluster at the same highly conserved genomic locus (Suppl. Fig. S14, chapter 3). This finding is another support of the hypothesis that secondary metabolism genes are widespread among actinobacteria due to lateral gene transfer. Hence, in order to disclose the hidden potential of not yet discovered metabolites a deeper look into the microbiome is required. Not only exploitation of unusual niches but the application of new biotechnological procedures is inevitable. The use of metagenomics, an approach mediated by the direct recovery of genomic DNA from the environment, is thus a solution to access the diversity of secondary metabolites derived from yet uncultivated strains. Prominent examples of the success of metagenomic approaches are the aforementioned teixobactin and malacidins which exhibit both considerable biological activities.

In order to avoid the frequent rediscovery, dereplication became an important step during NP discovery. The well advanced analytical techniques and the growing databases for data alignment ensures in many cases the preceding discard of known compounds. However, not always an unambiguous identification of known metabolites can be achieved. In the extract of IB2014/11-12, several PTMs have been identified by alignment of exact mass and UV absorption with DNP entries. Unfortunately, closely related derivatives were not distinguish-

able solely by these parameters. For instance, the compound with *m/z* 511.28207 $[M+H]^+$ coincides with alteramide A but as well with the its isomers maltophylin and xanthobaccin A. MS/MS fragmentation patterns can be taken into account as additional parameter for metabolite identification. However, depending on the chemical structure, MS/MS fragmentations are not always easy to interpret and are not a clear sign for a specific compound. Databases, like Global Natural Products Social Molecular Networking (GNPS)[10], that contain fragmentation patters for many NPs, can be of a great help in such situations. Although it might occur that the searched compound is not included. In contrast, we were able to identify unambiguously the bisindoles lynamicins/spiroindimicins and lagunapyrones in the extract of MP131-18 solely through comparison of exact mass and UV absorption with hits from DNP. Surely, the characteristic isotopic pattern of the mono-, di-, tri- and tetra-chlorinated lynamicins facilitated their identification. Likewise, the different isotopic pattern of the new spiroindimicins derivatives indicating monochlorination compared to the dichlorinated spiroindimicins B and C, gave a hint about their structural novelty. In case of lagunapyrones, the specific MS/MS fragmentation pattern with the multiple water loss helped eventually to clearly assign the compound to a DNP entry. Additionally, two new lagunapyrones were identified by the mass difference of 14 Da originating from different length of their alkyl side chain, predicted on the basis of proposed biosynthetic pathway. Dereplication plays an important role during NPs discovery and further advances of hyphenated analytical techniques and improvement of natural products databases will greatly speed up and simplify research in this direction.

The genomic era, expanding rapidly in the last 20 years, gave rise to sequencing of many bacterial genomes. The analysis of sequenced genomes using bioinformatics tools like antiSMASH or PRISM, reveals the secondary metabolism potential through prediction of the respective biosynthetic gene clusters. Within the genome of MP131-18 34 BGCs were identified whereas for IB2014/0111-12 analysis revealed 29 gene clusters. A majority of these clusters encode known secondary metabolites present in many different actinomycetes. For instance, BGCs responsible for production of coelibactin, ectoine, piericidin, roseoflavin, melanin or desferrioxamin are often present in many actinobacterial species due to their role in maintaining cell survival and adaptation to the environmental conditions. The desferrioxamin, a high-affinity iron chelator ensures for example the strain growth under iron limiting conditions. The deletion of desferrioxamin gene cluster in *S. albus* was lethal to the strain.[11] Beside these adaptive clusters, remaining BGCs open up the possibility to discover novel compounds since many of them have not been assigned to a putative metabolite yet. However,

often the identification of a respective compound is challenging. Most of the BGCs remain silent under the standard laboratory conditions or the metabolite can't be identified in the extract due to low production, compound stability or extraction issues. For example, the coelimycin P1 is produced by *S. coelicolor* only for a very short period of time between 24 and 60 hours of cultivations and is quickly degraded afterwards. The structural studies of this compound were strongly complicated by the fact of its ectopic production.[12] Thus, tools for activation of silent gene clusters are of great demand. Many strategies and approaches are indeed available for induction of silent gene cluster expression. However, the detection of corresponding metabolite often remains quite challenging. Since many strategies are based on influencing the global regulatory networks it is difficult to assign the activated gene cluster to a particular compound found in the extract of the manipulated strain. One such way is the use of a reporter guided approach.[13,14] In this case, the promoter of a structural gene from the BGC of interest is fused to the reporter gene allowing to monitor changes in its transcription and thus expression of the entire gene cluster. However, this technique still detects the changes in expression level of a gene rather than directly responds to the presence of a compound, providing a space for a high rate of false-positive results. As described in chapter four, biosensors represent a promising tool for the fast direct detection of produced secondary metabolites in actinomycetes. By simply putting a reporter gene under the control of a TetR-type repressor, originating from the BGC of interest, makes it possible to monitor the accumulation of a compound that is an outcome of the corresponding BGC activation. We have used this strategy to detect the induced production of coelimycin P1 in *S. lividans* TK24, which normally does not accumulate this metabolite. The successful application of the aforementioned technique makes us believe that it can be successfully used in combination with many other approaches (Tn-mutagenesis, random mutagenesis, promoter insertion, cluster remodeling, etc.). Besides that, it obviously can be expanded to applications, such as adjustment of fermentation conditions, overproducer selection and search for new producing strains.

Despite the well-established extraction and separation techniques, the isolation of secondary metabolites is still a challenging task. Although the amount of a compound needed for acquisition of NMR spectra and structure elucidation has significantly decreased, the yield is often still not sufficient for downstream applications and testing. Quite high amounts of metabolites are needed if stereochemistry has to be determined or if degradation experiments are needed to elucidate more complex structures. The activity tests against microorganisms or cell lines require additional amounts of the isolated NP. Whereas the known lynamicins/spiro-

indimicins were easy to isolate in high quantity, their new derivatives were only sparsely present in the extract of the strain. Nevertheless, we managed to obtain 0.6 mg and 0.3 mg of two new spiroindimicins E and F, respectively. Thereof, spiroindimicins E were subjected to activity test against T24 bladder carcinoma cells, in addition to the antibacterial test. Spiroindimicin E showed only slight effect on cellular growth compared to lynamicin E (50% inhibition of proliferation capacity) and lycogarubin C, which almost completely inhibited cell proliferation at the concentration of 10 µM. Furthermore, four new alpiniamides have been successfully isolated. Even though the amount was limited, alpiniamides B_1 and B_2, as well as the main product alpiniamide A, were used for determination of stereochemistry. Although, alpiniamide A had been isolated before, the stereochemistry seemed to be determined incorrectly. Therefore, we attempted to elucidate the correct configuration via NMR experiments. As result, we have assigned the left part of alpiniamide A as *threo* configuration, whereas the right side of the molecule was more complicated due to inconsistent results of the Mosher experiment. Such issues can be solved by the use of different derivatization agents, which would increase the chemical shift differences between the examined compounds, and thus yield more reliable result. Examples for such compounds would be for instance α-methoxyphenylacetic acid (MPA) or 9-anthrylmethoxyacetic acid (9-AMA). These and other compounds were extensively studied for a variety of polyfunctional derivatization agents by Riguera *et al.*.[15] The right part of alpiniamides B_1 and B_2 was equally problematic while the left part could be unambiguously assigned as *S* and *R*, respectively. In addition to this, genome-mining tools are meanwhile able to predict the stereochemical outcome of the ketoreductase mediated reaction. In the case of alpiniamide PKS the KR domain of module 1 is predicted to be a B1 type by antiSMASH, meaning the resulting hydroxy group is supposed to be in D configuration. This bioinformatics prediction coincides with the structures of alpiniamides A and B_2, since the respective stereocenter is in D-configuration in both compounds, but is wrong for alpiniamide B_1. It is not clear how two different stereochemistry can be present in the position resulting from the activity of the same KR domain. The situation is even more complicated by the fact that this KR domain is shared between module 1 and 2. Specialized tools for the prediction of stereocenter configuration within polyketides such as the pHHM web application can be even of a greater help.[16] The stereocenter resulting from KR domain in module 1/2 is predicted to be D by this tool as well. Prediction of stereochemistry using a bioinformatics approach is useful and might be already applicable to certain metabolites, although it still needs to be further developed and tuned to get reliable results. Hence, so far the method of choice for stereochemistry determination is still the chemical way.

With both genome sequences and chemical structures on hand, we were able to predict the biosynthetic pathways of spiroindimicins, lagunapyrones and alpiniamides. The structural information gave a first hint about the type of the enzymatic machinery responsible for the assembly of corresponding metabolite. In the case of spiroindimicin the identification of the BGC was simple due to the very characteristic core enzymes needed for construction of the indole skeleton. Halogenase genes for chlorination, tryptophan monooxygenase for formation of indolpyruvatimin, chromopyrrolic acid synthase for pyrrole formation and cytochrome P450 for indole formation are typically involved in assembly of these NPs. The antiSMASH mediated *Streptomyces* sp. Mp 131-18 genome analysis revealed a gene cluster similar to the one responsible for biosynthesis of the indole compound rebeccamycin.[17] A detailed analysis of all genes, present in this cluster and their products allowed us to predict the steps in lynamicin/spiroindimicin biosynthesis.

The identification of the alpiniamide BGC was more challenging. Within the genome of *Streptomyces* sp. IB2014/11-12, two BGCs are present, which could be potentially involved in the biosynthesis of these NPs. We hypothesized that alpiniamide, a linear polyketide, bearing an amino group in the middle, can be synthesized either by type I PKS and additional enzyme(s) for insertion of the amino group or by the combined action of NRPS and PKS. Thus, cluster no. 16 (type I PKS, harboring aminotransferase gene) and cluster no. 9 (hybrid NRPS trans-AT-PKS) were worth considering. However, we could not rule out one of the clusters solely based on analysis of the present genes. Cluster no. 9 seemed to be more likely candidate due to the presence of NRPS module with predicted glycine specificity of A domain, and in line methyltransferase domains within the PKS part for generating the methyl branches. However, the order of domains of this unusual PKS-NRPS megaenzyme did not correspond to the structure of alpiniamide. Therefore, an experimental proof has been required for the unambiguous identification of the *alp* gene cluster. We have used two different approaches, feeding with labeled precursors and classical gene disruption experiments, to prove that gene cluster no. 9 is indeed responsible for the biosynthesis of alpiniamides. The successful feeding experiments with labeled glycine and methionine gave a first hint that cluster no. 9 might be involved in biosynthesis of NPs of interest. Furthermore, as a second proof we aimed to delete putative gene cluster in the native produced strain. This was achieved by disruption of gene 12750 through replacing a 2 kb region with the spectinomycin resistance cassette. The mutant completely lacked production of alpiniamides. This experiment confirmed that cluster no. 9 is responsible for alpiniamide biosynthesis. However, genetic manipulation in the wild type strain

is not straightforward due to many reasons e.g. low conjugation efficiency. Thus, a heterologous expression of the gene cluster in a model strain greatly facilitates further genetic experiments in order to establish the particular steps and gene functions in the biosynthesis of respective NPs.

As mentioned before, the production of secondary metabolites is often low in native strain and in some cases not stable. Therefore, increase and stabilization of the production yield by manipulating the gene cluster, preferably in a heterologous host, is of high importance. In addition, the genetic manipulations in order to decipher biosynthetic pathways are easier to perform in a well-established and genetically tractable model strain with defined or even clean metabolic background, rather than in a native producer. Despite the visual simplicity of the heterologous expression approach it is not always successful. In many cases, the expression of the gene cluster of interest cannot be achieved. For instance, we have cloned the spiroindimicin gene cluster from MP 131-18 by TAR-cloning procedure and successfully introduced into *S. lividans* TK24 and *S. albus* J1074. However, the cluster was not expressed under multiple tested conditions. The attempt to induce production of lynamicins/spiroindimicins in the heterologous hosts by insertion of a strong promoter in front of the halogenase and regulatory genes also failed. Zhang et al. also attempted to express the spiroindimicin cluster from *Streptomyces* sp. SCSIO 03032 in *S. coelicolor* YF11 resulting in production of small amounts of chromopyrolic acid but not the final chlorinated and cyclized products.[18] Deletion of cytochrome P450 gene (*spmP*) in the *Streptomyces* sp. SCSIO 03032, which did not affect production of spiroindimicins, confirmed that this gene is not relevant for synthesis. The authors suggest that this function might be provided by any of the 15 other cytochromes P450 located elsewhere in the genome of the native strain. In contrast, the heterologous expression of the alpiniamide gene cluster was successful. This allowed us to perform deletions of biosynthetic genes using Red/ET recombineering technique. As result, we were able to examine the functions of several genes in the biosynthesis of alpiniamides. We constructed mutants lacking *alpD* (acyl-CoA dehydrogenase), *alpR* (ribosomal RNA methyltransferase) and *alpE* (3-oxoacyl-ACP synthase 3) respectively and analyzed their ability to produce alpiniamides. Deletion of *alpD* and *alpR* did not affect the alpiniamide production suggesting that these genes do not participate in alpiniamide biosynthesis. Whereas, deletion of *alpE* lead to complete abolishment of alpiniamides production.

In conclusion, two new members of the spiroindimicin family and two new lagunapyrones have been identified in the extract of the marine derived *Streptomyces sp.* MP131-18, as well as four new alpiniamides were discovered from the *Streptomyces sp*. IB2014/11-12. These findings show clearly that screening of streptomycetes species from new ecological niches represent a great source for valuable natural products. The combined metabolic and genomics approach is a powerful tool for their identification and elucidation of the biosynthetic origin. Additionally, the application of biotechnological tools such as biosensors will speed up the natural product research even further.

References

1 Khan, A. U., Maryam, L. & Zarrilli, R. Structure, Genetics and Worldwide Spread of New Delhi Metallo-beta-lactamase (NDM): a threat to public health. **17**, 101, doi:10.1186/s12866-017-1012-8 (2017).

2 Bredholdt, H. *et al.* Rare actinomycete bacteria from the shallow water sediments of the Trondheim fjord, Norway: isolation, diversity and biological activity. *Environmental Microbiology* **9**, 2756-2764, doi:doi:10.1111/j.1462-2920.2007.01387.x (2007).

3 Bredholt, H., Fjærvik, E., Johnsen, G. & Zotchev, S. B. Actinomycetes from sediments in the Trondheim fjord, Norway: diversity and biological activity. *Marine drugs* **6**, 12-24 (2008).

4 Axenov-Gribanov, D. *et al.* The isolation and characterization of actinobacteria from dominant benthic macroinvertebrates endemic to Lake Baikal. *Folia microbiologica* **61**, 159-168 (2016).

5 Zhang, W. *et al.* Spiroindimicins A–D: New Bisindole Alkaloids from a Deep-Sea-Derived Actinomycete. *Organic Letters* **14**, 3364-3367, doi:10.1021/ol301343n (2012).

6 Zhou, H. *et al.* Alkaloids from an endophytic streptomyces sp. YIM66017. *Natural product communications* **8**, 1393-1396 (2013).

7 Nanthini, J. *et al.* Complete genome sequence of Streptomyces sp. strain CFMR 7, a natural rubber degrading actinomycete isolated from Penang, Malaysia. *Journal of Biotechnology* **214**, 47-48, doi:https://doi.org/10.1016/j.jbiotec.2015.09.007 (2015).

8 Yang, S., Huang, Y., Wang, C., Zhao, P. & Zhang, W. [Isolation, identification of TJ430 strain and characterization of its antifungal metabolite]. *Wei sheng wu xue bao = Acta microbiologica Sinica* **54**, 624-634 (2014).

9 Xu, X.-N. *et al.* Genome Mining of the Marine Actinomycete Streptomyces sp. DUT11 and Discovery of Tunicamycins as Anti-complement Agents. *Frontiers in Microbiology* **9**, 1318, doi:10.3389/fmicb.2018.01318 (2018).

10 Wang, M. *et al.* Sharing and community curation of mass spectrometry data with Global Natural Products Social Molecular Networking. *Nature biotechnology* **34**, 828-837, doi:10.1038/nbt.3597 (2016).

11 Myronovskyi, M., Rosenkränzer, B. & Luzhetskyy, A. Iterative marker excision system. *Applied Microbiology and Biotechnology* **98**, 4557-4570, doi:10.1007/s00253-014-5523-z (2014).

12 Gottelt, M., Kol, S., Gomez-Escribano, J. P., Bibb, M. & Takano, E. Deletion of a regulatory gene within the cpk gene cluster reveals novel antibacterial activity in Streptomyces coelicolor A3(2). *Microbiology* **156**, 2343-2353, doi: 10.1099/mic.0.038281-0 (2010).

13 Guo, F. *et al.* Targeted activation of silent natural product biosynthesis pathways by reporter-guided mutant selection. *Metab Eng* **28**, 134-142, doi:10.1016/j.ymben. 2014.12.006 (2015).

14 Ahmed, Y., Rebets, Y., Tokovenko, B., Brötz, E. & Luzhetskyy, A. Identification of butenolide regulatory system controlling secondary metabolism in Streptomyces albus J1074. *Scientific Reports* **7**, 9784, doi:10.1038/s41598-017-10316-y (2017).

15 Seco, J. M., Quiñoá, E. & Riguera, R. Assignment of the absolute configuration of polyfunctional compounds by NMR using chiral derivatizing agents. *Chemical reviews* **112**, 4603–4641, doi:10.1021/cr2003344 (2012).

16 Kitsche, A. & Kalesse, M. Configurational Assignment of Secondary Hydroxyl Groups and Methyl Branches in Polyketide Natural Products through Bioinformatic Analysis of the Ketoreductase Domain. *ChemBioChem* **14**, 851-861, doi:10.1002/cbic.201300063 (2013).

17 Sánchez, C. *et al.* The Biosynthetic Gene Cluster for the Antitumor Rebeccamycin: Characterization and Generation of Indolocarbazole Derivatives. *Chemistry & Biology* **9**, 519-531, doi:https://doi.org/10.1016/S1074-5521(02)00126-6 (2002).

18 Ma, L. *et al.* Identification and characterization of a biosynthetic gene cluster for tryptophan dimers in deep sea-derived Streptomyces sp. SCSIO 03032. *Applied Microbiology and Biotechnology* **101**, 6123-6136, doi:10.1007/s00253-017-8375-5 (2017).