Exploring the chemical space of natural products from Streptomyces using multi-omics approaches

Nuñez Santiago, I.

04

CHAPTER 4

Multi-Omics Analysis of *Streptomyces* Angucycline Producers: Genomic, Metabolomic and Bioactivity Insights

Isabel Nuñez Santiago, Chao Du, Lina M. Bayona Maldonado, Mercedes de la Cruz, Olga Genilloud, Gilles P. van Wezel

Abstract

Actinobacteria are a prolific source of natural products, which find their application as among others as antibiotics, anticancer agents, antifungals or immunosuppressants. Angucyclines form the largest family of type-II polyketide natural products. As prominent natural product producers, streptomycetes produce a large variety of angucyclines, including angucyclinones and limamycins. Lugdunomycin, produced by *Streptomyces* sp. QL37, is a complex rearranged angucycline that incorporates an *iso*-maleimycin moiety (Chapter 3). Here we report the analysis and comparison of angucycline producers and their corresponding BGCs, through a combination of phylogenomics, LC-MS/MS metabolomics and feature-based molecular networking. Limamycin producers *Streptomyces* sp. ICBB8415 and ICBB8309 exhibited metabolomic profiles similar to *Streptomyces* sp. QL37. Metabolomic analysis revealed that the strains *Streptomyces* sp. G0-475 and *Streptomyces* sp. ICBB8415 produced 8-O-methyltetrangomycin and several of its analogues, and an analogue of rubiginone A2. Also, *Streptomyces venezuelae* produced fluostatin A and haloquinone. However, apart from *Streptomyces* sp. QL37 none of the strains contained the BGC for *iso*-maleimycin, which is essential for the biosynthesis of lugdunomycin. Bioactivity assays against a range of clinical pathogenic bacteria and cancer cell lines demonstrated that various strains had significant bioactivity against Gram-positive bacteria, including MRSA, as well as notable anticancer activity.

Introduction

Considering the growing crisis of antibiotic resistance among bacterial pathogens and the rising incidence of cancer, the quest for new, effective, and less toxic drugs is paramount (Wright, 2015, Singh, 2014). Actinobacteria have been the primary source of most antibiotics currently in use (Barka *et al.*, 2016, Bérdy, 2005). Among them, the genus *Streptomyces* is particularly notable, producing over half of all known antibiotics (Barka *et al.*, 2016). However, we have likely explored only a fraction of their chemical diversity.

Natural products exhibit a wide range of chemical structures, including polyketides (PKs), nonribosomal peptides (NRPs), and ribosomally synthesized and post-translationally modified peptides (RiPPs), among others. The biosynthetic processes for these classes of natural products are highly complex. In prokaryotes, the genes responsible for the biosynthesis of specific molecules are often organized into what are called biosynthetic gene clusters (BGCs) (Liu *et al.*, 2021). These clusters can extend up to 100 kilobases and typically comprise core enzymes, tailoring enzymes, resistance genes, and transport genes. They also include regulatory elements that modulate expression levels, as the production of natural products is energetically demanding for the organism (Weber *et al.*, 2015).

Among these classes of NP, type-II polyketides (PKs) are important sources of potential therapeutic agents, particularly as antibiotics and anticancer agents (Hertweck, 2009) and they are produced by type-II polyketide synthases (PKSs). For instance, the best-known type-II PK molecules are the tetracyclines, which are effective against a broad spectrum of both Gram-positive and Gram-negative bacteria. Also, the anthracyclines daunorubicin and doxorubicin, isolated from *Streptomyces peucetius*, are used in first-line chemotherapy for the treatment of solid and haematological tumours (Hulst *et al.*, 2022). While the compounds produced by type-II PKSs are diverse, they share common structural features. Aromatic rings are one of such features, it is the core of type-II PKs, with diversity arising from variations in polyketide chain length and the action of tailoring enzymes that generate a wide array of ring systems and functional groups (Rivers & Lowell, 2024).

Angucyclines are a group of glycosylated molecules belonging to the type-II PKs class. Their aglycones are referred to as angucyclinones (Elsayed *et al.*, 2023). Together, they constitute the largest group of type-II PKS-derived

natural products, featuring over 400 documented members, approximately 45% of which exhibit biological activity (Elsayed *et al.*, 2023). *Streptomyces* sp. QL37 produces a large number of angucyclinones and derivatives, including limamycins and lugdunomycin (Wu *et al.*, 2019b). Lugdunomycin constitutes a new subclass of polyketide antibiotics with a unique chemical scaffold. The lugdunomycin biosynthetic gene cluster (*lug*) was analyzed in detail (Wu *et al.*, 2019b, Xiao *et al.*, 2020). Five genes (*lugOI–OV*) were found which encode putative oxygenases. LugOI clusters with flavoprotein monooxygenases which function as both C12 and C12b hydroxylases. LugOII associates with bifunctional enzymes that possess an N-terminal flavoprotein monooxygenase domain and a C-terminal short-chain dehydrogenase/reductase (SDR) domain. LugOIV contains an SDR domain. Finally, LugOIII and LugOV are part of the antibiotic biosynthesis monooxygenase (ABM) family and they facilitate the cleavage of the angucyclinone C-ring through epoxidation and their sequential enzymatic activities (Elsayed *et al.*, 2023).

To expand our insights into angucycline biosynthesis, we investigated a small collection of known angucycline producers, namely *Streptomyces scopuliridis, Streptomyces* sp. MP131-18 (Paulus *et al.*, 2017), *Streptomyces* sp. CS057 (Malmierca *et al.*, 2018), *Streptomyces* sp. Go-475 (Kibret *et al.*, 2018)*, Streptomyces albus (Jin et al., 2018)*, *Streptomyces ambofaciens (Bunet et al., 2008)*, *Streptomyces venezuelae (Raghavan & Groisman, 2010)*, *Streptomyces* ICBB8309 and *Streptomyces* ICBB8415 (Fotso *et al.*, 2008). Initially, our research on the angucycline cluster across different strains aimed to identify new producers of lugdunomycin. However, as demonstrated in Chapter 3, a second biosynthetic gene cluster (referred to as BGC23 in *Streptomyces* sp. QL37) is responsible for the *iso*-maleimycin moiety of lugdunomycin. None of the strains contained this BGC, but its discovery opens the possibility for heterologous expression of BGC23 in various angucycline producers, potentially leading to the creation of chimeric lugdunomycins. Then, the angucycline BGCs were compared to that of *Streptomyces* sp. QL37, and their metabolomes were analysed for angucyclines and derivatives using LC-MS-based metabolomics and molecular networking using the Global Natural product Social (GNPS). Combining the use of feature based molecular networking (FBMN) and SIRIUS (Duhrkop *et al.*, 2021), we were able to efficiently dereplicate the metabolomic data. Molecules identified included 8-*O*-Methyltetrangomycinand several of its analogues, as well as of fluostatin A, haloquinone, and a rubiginone A2 analogue.

Finally, the extracts were also evaluated for their anti-infective and anticancer profiles against a panel of five pathogens and five cancerous cell lines.

Results and discussion

Bioinformatic comparison of annotated angucycline biosynthetic gene clusters

To characterize angucycline-producing *Streptomyces* strains with angucycline biosynthetic gene clusters (BGCs) similar to the angucycline biosynthetic gene cluster (*lug*) of *Streptomyces* sp. QL37 (Chapter 3), eight Actinobacteria were selected based on previous study. Besides QL37, these strains all contain an angucycline BGC that is similar to the lugdunomycin BGC (*lug*) (van der Heul, 2022). These strains are *Streptomyces* sp. MP131-18, *Streptomyces* sp. CS057, *Streptomyces* sp. Go-475, *S. albus*, *S. ambofaciens*, *S. scopuliridis*, and *S. venezuelae* (Table S1). A review of the literature identified additional *Streptomyces* species reported to produce limamycins, which exhibit close similarity to the limamycins identified in *Streptomyces* sp. QL37. Consequently, we focused our attention on the two limamycin producers, *Streptomyces* sp. ICBB8309 and *Streptomyces* sp. ICBB8415. To better understand the capabilities of these two strains, the genomes were sequenced using PacBio technology and assembled into complete genomes for both strains. The assembly also resulted in the finding of a circular plasmid from ICBB8309. The average coverages were 168 x for ICBB8309 and 164 x for ICBB8415.

To gain insights into the biosynthetic potential of the strains, we conducted a bioinformatic analysis focusing on their angucycline gene clusters. The angucycline clusters from the strains were analyzed using Clinker (Gilchrist & Chooi, 2021) (Figure S1), with the amino acid identity threshold set to 45%. Additionally, a pairwise comparison of each BGC from the angucycline producers was conducted against the *lug* gene cluster (BGC12) of *Streptomyces* sp. QL37 to determine the presence or absence of genes in the *lug* cluster relative to the other strains (Table 1).

All strains contained a highly conserved core PKS region typical of angucycline BGCs. This core includes genes encoding the minimal polyketide synthase (minPKS) components, namely ketosynthase (KS), chain length factor (CLF), and acyl carrier protein (ACP) (Wang *et al.*, 2020). The conservation of these genes across different strains underscores their essential role in the biosynthesis of angucycline and related compounds (Wang *et al.*, 2020).

The angucycline BGC of *S. scopuliridis* demonstrates high degree of identity to BGC12 of *Streptomyces* sp. QL37. This resemblance is particularly evident in the core *minPKS* genes, tailoring enzymes, transporters, and regulatory elements (Table 1, Figure S1). In terms of post-tailoring enzymes and in comparison to *Streptomyces* sp. QL37, the angucycline BGC of *S. scopuliridis* also includes the genes encoding the monooxygenases LugOI (>80% aa identity), LugOII and LugOIV (60-80% aa identity), and LugOIII and LugOV (aa identity 45-60%) (Elsayed *et al.*, 2023, Xiao *et al.*, 2020). This suggests that *S. scopuliridis* has the genetic potential to produce molecules with similar characteristics as *Streptomyces* sp. QL37, including C-ring cleaved rearranged angucyclines (Elsayed *et al.*, 2023). Furthermore, it contains close orthologues for the regulatory genes *lugRIII* and *lugRV* and for the transporter genes *lugTI* and *lugTII* (aa identity between 60-80%). However, the strain lacks an orthologue for *lugTIII*.

The angucycline BGCs from *Streptomyces* sp. ICBB8309 and *Streptomyces* sp. ICBB8415 are nearly identical (Figure S1), and both produce limamycins (Fotso *et al.*, 2008). Pairwise comparison of the angucycline gene clusters from the two ICBB strains with the angucycline BGC of *Streptomyces* sp. QL37 also shows a high degree of identity (Table 1). *S. scopuliridis* and the ICBB strains are the only strains that contain genes homologous to the five *lugO* oxygenase genes of BGC12 of QL37. This strongly suggests they have the capacity to produce C-ring rearranged angucyclines, similar to for example limamycin A, produced by the ICBB strains (Fotso *et al.*, 2008). It is likely that these compounds are produced through a pathway analogous to the one in *Streptomyces* sp. QL37 (Elsayed *et al.*, 2023). Lastly, *lugX*, a gene of unkown function and with no clear orthologues in the databases, is exclusively present in *Streptomyces* sp. QL37.

LugOI phylogenetically clusters with flavoprotein monooxygenases, a group known for their role in oxidative modifications. It is a key monooxygenase in the BGC12 of QL37, that is responsible for for C12 oxidation and maintaining a keto group at C6 (Elsayed *et al.*, 2023). The importance of LugOI is highlighted by the fact that deletion of *lugOI* in *Streptomyces* sp. QL37 nearly abolished the production of both non-rearranged and rearranged angucyclinones (Elsayed *et al.*, 2023). In our analysis, LugOI orthologues are found in all BGCs with reasonably high confidence (> 45% aa identity). Of the five oxygenases encoded by the *lug* gene cluster, this is the only enzyme that is conserved between all strains tested here, indicating that the reaction catalyzed by it is likely critical for angucycline biosynthesis.

Table 1. Overview of lug genes in BGCs of angucycline-producing compared to *Streptomyces* sp. QL37.

Strain*	LugM	LugRI	LugRII	LugRIII	LugTI	LugX	LugN	LugOI	LugF	LugA	LugB	LugC	LugD	LugE	LugOII	LugG	LugH	LugI	LugTII	LugJ	LugOIII	LugK	LugRIV	LugTIII	LugL	LugOIV	LugRV	LugOV
Streptomyces scopuliridis																												
Streptomyces sp. ICBB8309																												
Streptomyces sp. ICBB8415																												
Streptomyces sp. G0-475																												
Streptomyces venezuelae																												
Streptomyces sp. MP131-18																												
Streptomyces sp. CS057																												
Streptomyces ambofaciens																												
Streptomyces albus																												

*Predicted gene products absent or with less than 45% aa identity relative to the gene products of *Streptomyces* sp. QL37 are indicated in white; those with an aa identity between 45-60% are indicated in light grey; between 60-80% are indicated in dark grey and aa identity above 80% are shown in black.

The genes *lugOIII* and *lugOV* are required for angucycline C-ring cleavage during lugdunomycin biosynthesis. Homologues for both genes are present in *S. scopuliridis* and in the ICBB strains, with between 45-60% aa identity for the predicted gene products. *Streptomyces* sp. G0-475 also has an orthologue of *lugOIII* but lacks one for *lugOV*. This strain produces an analogue of Rubiginone A2, which is a precursor for A-ring rearranged angucyclines (Zhang *et al.*, 2021).

Streptomyces sp. MP131-18, *S. albus*, *S. ambofaciens*, *S. venezuelae*, *Streptomyces* sp. CS057 all share a *minPKS* core with a high degree of sequence identity but lack genes for the characteristic monooxygenases enzymes found in *Streptomyces* sp. QL37. This absence indicates divergence in terms of the angucyclines they produce. The *S. albus* angucycline BGC is identified as the BGC for fluostatins M-Q by antiSMASH. Fluostatins M–Q are atypical angucyclinones featuring a 6-5-6-6 ring skeleton and high oxidized A-rings (Jin *et al.*, 2018), which is a different ring rearrangement as the angucyclines produced by *Streptomyces* sp. QL37. *S. ambofaciens* angucycline BGC also encodes five monooxygenases (SAM23877_RS00645; RS00660; RS00760; RS00765 and RS00805), but they are all FAD dependent, different from the monooxygenases in QL37 angucycline BGC. Study shows that these FAD-dependent monooxygenases are not involved in ring cleavage (Fan & Zhang, 2018).

Although these strains lack BGC23 that is responsible for the production of *iso*-maleimycin (chapter 3) and hence cannot make lugdunomycin, the homologous core PKS machinery and the ability to produce angucyclinones makes them interesting potential hosts for lugdunomycin production. Expressing key enzymes like LugOIII and LugOV may be sufficient to produce the diene elmonin that serves as precursor for lugdunomycin biosynthesis (chapter 3). In particular, by introducing these enzymes in a “toolbox” approach, it may be possible to generate novel angucycline derivatives, expanding the chemical diversity of this class of compounds for potential bioactivity studies.

Metabolomic profiling of the angucycline producers

Metabolomic profiling of the strains was conducted using extracts obtained from strains grown for seven days on minimal medium agar (StrepMM), supplemented with mannitol (0.5% w/v) and glycerol (1% w/v). This medium was selected because it supports the production of both canonical and rearranged angucyclines, as well as lugdunomycin by *Streptomyces* sp. QL37 (Wu *et al.*, 2019b). Media blanks and the *Streptomyces* sp. QL37 Δ*minPKS* mutant, which is unable to produce angucyclines, were included as control. After seven days of incubation, the agar

was extracted three times with ethyl acetate (1:1) and analyzed by LC-MS/MS. The data were pre-processed using MZmine 2.53 (Pluskal *et al.*, 2010), producing an aligned feature list with peak areas for a semi-quantitative analysis of the detected features. The processed data was uploaded to Metaboanalyst (Pang *et al.*, 2024) for statistical analysis, to investigate the metabolic differences between the different angucycline-producing strains.

To obtain insights into the metabolic variation between the strains, unsupervised Principal Component Analysis (PCA) was performed (Fig. 1). The PCA plot demonstrated excellent reproducibility of our sample set, as all replicates from each strain consistently clustered together. PC1 is largely due to the variation between *S. albus* and the rest of the strains (Fig. 1A). To better distinguish between all other strains, PC2 and PC3 were plotted, which shows that while *S. albus* was distinctly separated from the others in the PC1 and PC2 plots, it now clusters more closely with *Streptomyces* sp. G0-475. In contrast, *Streptomyces* sp. CS057 is the most distinct and separated from the other strains in the PC2 and PC3 analysis, suggesting significant metabolic divergence from the other strains. The ICBB strains, along with *Streptomyces* sp. QL37, form a tight cluster, indicating that they have very similar metabolic profiles. This clustering supports the idea that these strains share key biochemical pathways or metabolic characteristics, potentially related to their environmental adaptation or niche-specific functions.

Interestingly, *S. scopuliridis*, *Streptomyces* sp. MP131-11, and *S. venezuelae* cluster together with the QL37 *minPKS* null mutant, which could be due to the poor growth of these strains under the tested conditions. Additionally, a loading plot of PC2 and PC3 was generated highlighting the features related to angucyclines (Fig. 1C). The loading plot revealed that features associated with angucycline production cluster together in the middle-right section, indicating that these features contributed more to the separation of *S. albus* and *Streptomyces* sp. G0-475. Further analysis is required to definitively link these feature groupings with strain-specific production patterns.

Next, the pre-processed data using MZmine 2.53 (Pluskal *et al.*, 2010) were analyzed using Feature-Based Molecular Networking (FBMN) in The Global Natural Product Social (GNPS) platform (Nothias *et al.*, 2020). This molecular networking tool allows automatic detection of structural similarities among features based on MS/MS data in a semi-quantitative fashion. As a result, features with similar scaffolds cluster together. The network produced consisted of 4112 nodes, grouped in 532 spectral families and 1379 singletons and can be publicly accessed via https://gnps.ucsd.edu/ProteoSAFe/status.jsp?task=6d5b8bd74891441e88198dea37e45acc.

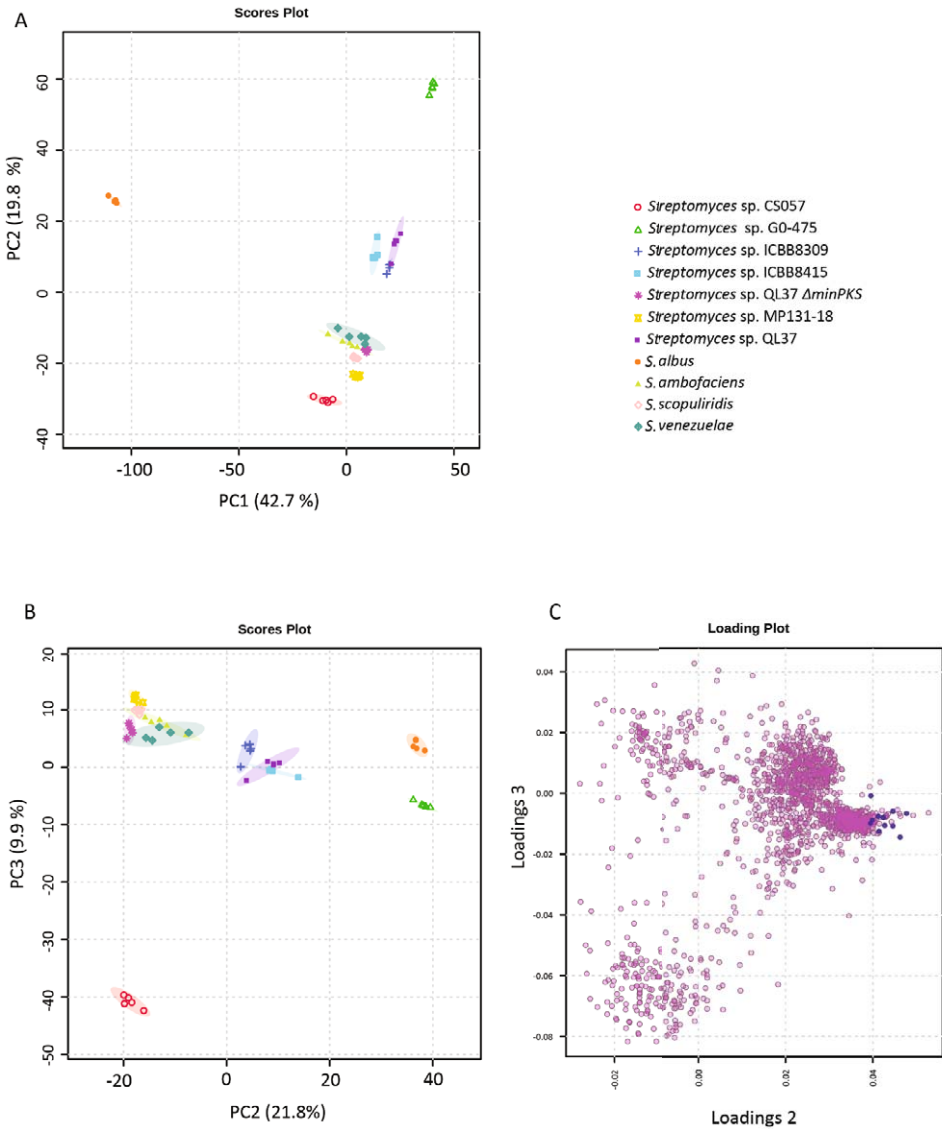

Figure 1. Diversity in chemical profiles among angucycline-producing strains. A. PCA score plot displaying PC1 and PC2 of the crude extracts of the different strains grown in StrepMM agar supplemented with 0.5% mannitol and 1% glycerol. The colored area around the dots represents the 95% confidence interval. B. PCA score plot displaying PC2 and PC3. C. PCA loading plot showing the distribution of metabolic features along the second and third principal components (PC2 and PC3). Features highlighted in blue are those identified as angucyclines based on subsequent analyses (described later in the manuscript).

To reduce network complexity and better focus on the type and distribution of PKS-II molecules produced by the strains, we analyzed the angucycline-related features using the software SIRIUS (Duhrkop *et al.*, 2021), which allows obtaining structural information about metabolites using their tandem mass spectrometry data. It merges the analysis of isotope patterns in MS spectra with the fragmentation pattern analysis in MS/MS spectra and utilizes CSI:FingerID (Duhrkop *et al.*, 2015) to build fragmentation trees and predict and search for molecular structure in databases. Most importantly, it incorporates CANOPUS (Duhrkop *et al.*, 2021, Djoumbou Feunang *et al.*, 2016, Kim *et al.*, 2021) for compound class prediction. In this way we could mapped the features from the molecular network predicted to be angucyclines (Fig. 2).

In total 211 features have been reported as being related to angucyclines by NPClassifier (Kim *et al.*, 2021) (NPC), integrated in SIRIUS. These features, alongside the nodes linked by GNPS within the spectral families are shown in Figure 2. Each node represents a feature with a pie chart in the interior. Each segment of the pie chart represents the relative abundance or presence of a compound in a specific strain. The nodes are connected if the cosine score is above 0.6. Some angucycline spectral families were formed by features exclusively produced by *S. venezuelae*. This was particularly evident in subnetworks B, J, and O (Fig. 2). Spectral family B is particularly interesting, as it is the second largest subnetwork, comprising 59 nodes that are predominantly composed of features identified solely in *S. venezuelae*.

Amongst the features identified in *Streptomyces* sp. QL37, most were shared with those of *Streptomyces* sp. G0-475. These include spectral families B1, K and some smaller networks (Figure 2). This overlap is not observed with the other strains. *S. scopuliridis*, which has a BGC that is very similar to the *lug* cluster, produced only a few angucyclines, likely due to the low expression achieved under the chosen growth conditions. Spectral family S is formed by features exclusive to strain *Streptomyces* sp. CS057, suggesting that these angucyclines are unique to this strain.

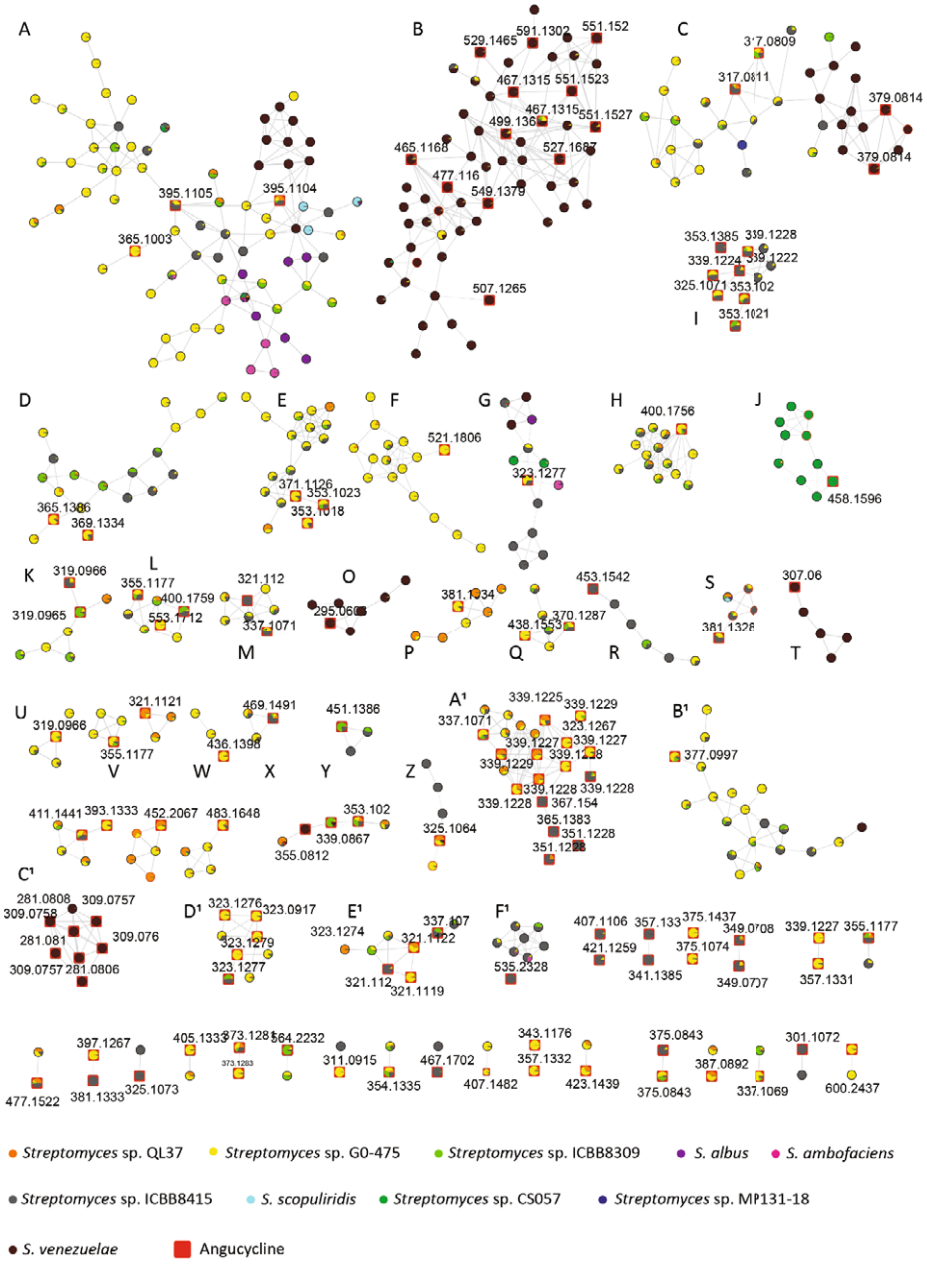

Figure 2. Molecular network representing the molecules identified in the crude extracts of the angucycline-producer strains by GNPS-molecular networking and SIRIUS in the extracts. Colors of the pie charts correspond to the relative abundance of the chromatographic peaks in each strain. Red square represents features identified as angucyclines by CANOPUS. The edge connects two nodes when the cosine similarity score is above 0.6.

More detailed examination of the molecular network was conducted to screen library hits identified through the GNPS platform. Using both the GNPS library and the analogue search functionalities on the platform, the hits were systematically mapped within the network (Fig. 3). The focus was particularly on the dereplication of the angucyclines features reported by SIRIUS. Following the guidelines of the Metabolomics Standards Initiative (MSI) (Salek *et al.*, 2013), several compounds are annotated at Level 2 confidence; which indicates that their identification is based on matching experimental data, mass spectrometry, to those of known compounds in publicly available or curated databases (GNPS library), providing strong but not definitive evidence of their identity. The scope of the identification was expanded to include closely related chemical variants by incorporating analogue searches, enhancing our understanding of the molecular diversity present within the sample.

In *Streptomyces*, 8-*O*-methyltetrangomycin has been identified not only as a prominent angucycline but also as a precursor for other rearranged angucyclines (Elsayed *et al.*, 2023, Nuutila *et al.*, 2024). This compound, also referred to as 6-deoxy-8-*O*-methylrabelomycin, or MM 47755, is a well-known antibiotic with significant antibacterial activity against several Gram-positive bacteria (Shigihara, 1988). 8-O-Methyltetrangomycin was the hit that was identified most within the datasets. The feature *m/z* 337.1071 in the spectral family A^1 (Figure S2) was a hit for 8-O-Methyltetrangomycin and several analogues of 8-*O*-methyltetrangomycin were identified across multiple subnetworks, specifically I, A^1, and L (Figure 3). These compounds were predominantly detected in extracts from *Streptomyces* sp. QL37, *Streptomyces* sp. G0-475, and *Streptomyces* sp. ICBB8415. This metabolic versatility underscores the significance of 8-*O*-Methyltetrangomycin in the biosynthetic pathways of *Streptomyces* species, where it serves as a foundational scaffold for the generation of structurally diverse and potentially bioactive metabolites.

Spectral families T and O are formed by features produced by *S. venezuelae*. In the spectral family T, fluostatin A was annotated, while in the spectral family O, haloquinone was annotated. Fluostatin A is a potent inhibitor of dipeptidyl peptidase III (Akiyama *et al.*, 1998). Haloquinone was first described in *S. venezuelae* extracts during a screening of molecules with activity against *Halobacteria*, a group classified as archaebacteria (Wersmeyer-Wenk *et al.*, 1981). These organisms have a unique cell-wall structure, composed of glycoproteins similar to those found in eukaryotic cells, resulting in have lower sensitivity to many antibiotics.

Lastly, the spectral family D encompasses an analogue of rubiginone A2, a type of angucyclinones previously characterized from *S. griseorubiginosus* for its activity as enhancer of the cytotoxicity of vincristine (Oka *et al.*, 1990). The features of this spectral family were found in extracts of *Streptomyces* strains ICBB8415, *Streptomyces* sp. ICBB8309, *Streptomyces* sp. G0-475 and *Streptomyces* sp. QL37.

Considering our specific interest in lugdunomycin, extracts from all strains were scrutinized for the possible production of lugdunomycin-like compounds. However, two separate BGCs are required for lugdunomycin biosynthesis (Chapter 3) and none of the strains have a BGC similar to BGC23. Within the lugdunomycin spectral family, three nodes were features with the same retention time (7.1 min), and represent different lugdunomycin adducts: *m/z* 474.1537 $[M+H]^+$, *m/z* 456.1537 $[M-H_2O+H]^+$ and *m/z* 438.1333 $[M-2H_2O+H]^+$. Expectedly, these features were only identified in samples of *Streptomyces* sp. QL37 (Figure S3-S5). The last feature of this network, *m/z* 456.1441 (retention time 6.8 min), is predominantly found in *Streptomyces* sp. QL37 extracts, with partial presence in *Streptomyces* sp. ICBB8309 and *Streptomyces* sp. 8415. The raw data was then analyzed to compare the MS/MS fragmentation pattern of the feature in *S*treptomyces sp. QL37 (Figure S2, S3). However, they only share some fragments corresponding probably to the loss of water molecules, but lacked the characteristic fragments present in the lugdunomycin isomer from *Streptomyces* sp. QL37.

Bioactivity testing of the angucycline-producer strains

Angucyclines are known for their diverse bioactivities, ranging from antimicrobial to anticancer effects (Hertweck, 2009). The antimicrobial activity was assessed of extracts from, *Streptomyces* sp. QL37, *Streptomyces* sp. CS057, *Streptomyces* sp. Go-475, *S. albus* and *S. venezuelae*. The strains *Streptomyces* sp. MP131-18, *S. scopuliridis*, and *S. ambofaciens* were not included due to insufficient growth under the tested conditions, while *Streptomyces* sp. ICBB8309 and *Streptomyces* sp. ICBB8415 were not available at the time when the bioactivity assays were conducted. Extracts from each strain were tested in five replicates, with each strain grown and extracted independently on separate agar plates. Additionally, a media blank was included as a control. It needs to be noted that the observed activity may also be influenced, at least in part, by other natural products produced by these strains.

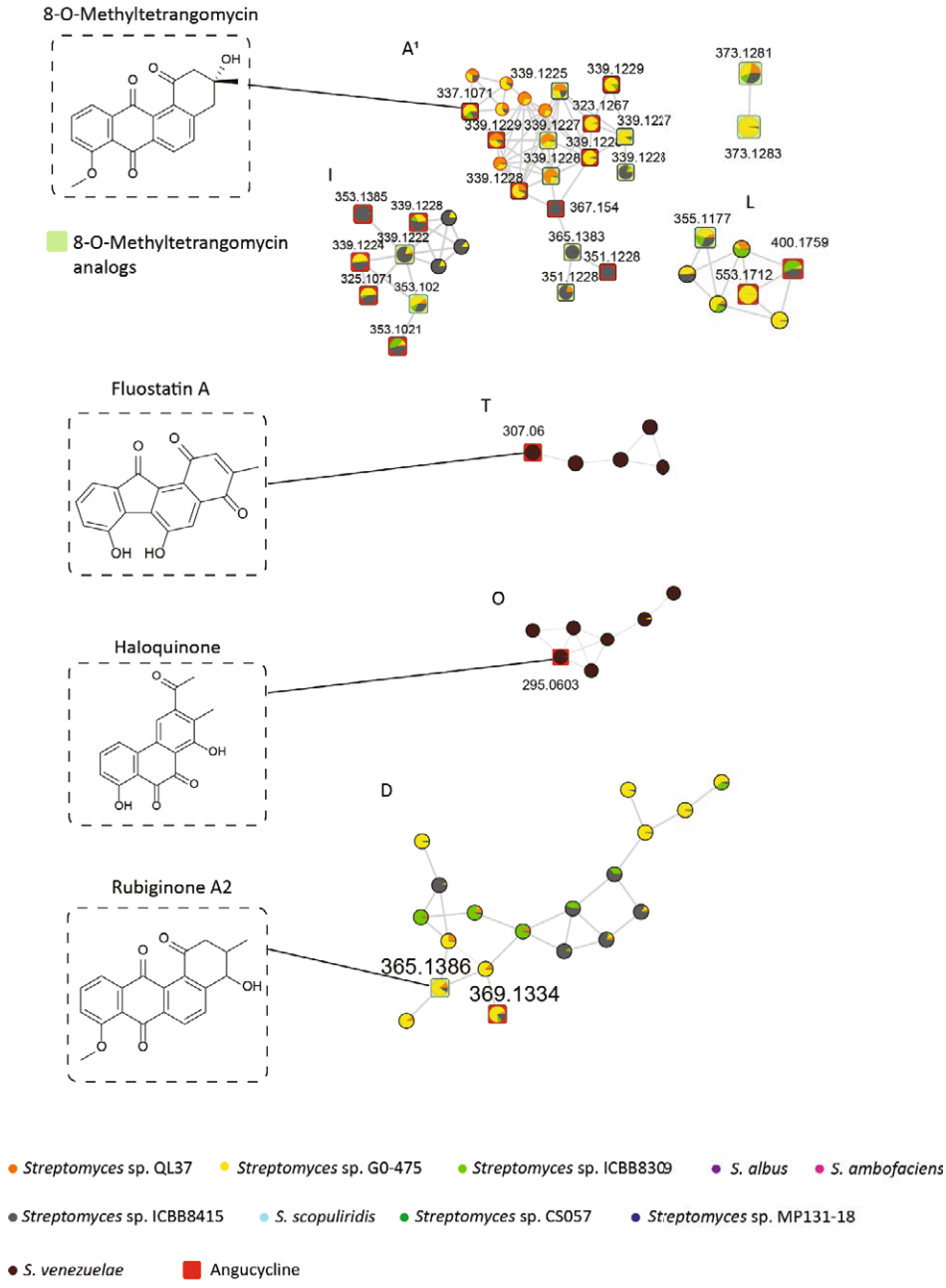

Figure 3. Spectral families containing the dereplicated angucyclines and analogues identified by GNPS-molecular networking of culture extracts. Colours of the pie charts withing the nodes correspond to the relative abundance of the ions in the different strains. Red square represents features identified as angucyclines by CANOPUS. Green squares represent features identified as analogues.

A panel of pathogens were used as target strains, each with significant clinical concern: (i) the yeast *Candida albicans* (*C. albicans*), a major cause of invasive fungal infections, which affects 17% of patients in intensive care units and is associated with high morbidity and mortality (Soulountsi *et al.*, 2021, Filioti *et al.*, 2007); (ii) *Staphylococcus aureus*, a Gram-positive bacterium responsible for severe wound infections, including both methicillin-sensitive (MSSA) and methicillin-resistant (MRSA) strains (Kali, 2015); (iii) *Escherichia coli*, a Gram-negative bacterium commonly associated with opportunistic infections (Braz *et al.*, 2020); and (iv) *Klebsiella pneumoniae*, a Gram-negative bacterium known for its high resistance to many antibiotics (Navon-Venezia *et al.*, 2017). The extracts were tested in five replicates for each strain, each grown and extracted independently on separate agar plates. Technical replicates of the antimicrobial assays were also performed to ensure the reliability of the results. The antibiotic activity results are displayed in Table 2.

Bioactivity was observed for most of the extracts against Gram-positive bacteria, while none showed significant antibiotic activity against Gram-negative indicator strains. Extracts of *Streptomyces* sp. QL37 and *Streptomyces sp.* G0-475 had a similar activity profile, with moderate antibiotic activity against *S. aureus* MRSA MB5393. The similarity between the activities of their extracts is consistent with the close resemblance of the angucyclines identified through FBMN analysis. *Streptomyces* sp. CS057, known for the production of warkamycin (Malmierca *et al.*, 2018), showed high activity against all Gram-positive pathogens tested, but failed to inhibit growth of *C. albicans*. Extracts from *Streptomyces albus* showed high activity against Gram-positive bacteria, while they also inhibited *C. albicans*.

Cytotoxicity assays

Angucyclines are well-documented for their potent anticancer properties, which led us to evaluate the cytotoxicity of extracts from the same angucycline-producing strains against five different cancer cell lines, namely MiaPaca-2, A2058, A549, MCF7, and HepG2. MiaPaca-2 is a human pancreatic cancer cell line commonly used in studies of pancreatic ductal adenocarcinoma. This cell line is known for its resistance to many conventional chemotherapeutics, making it an important model for testing the efficacy of novel anticancer compounds (Rückert *et al.*, 2012). A2058 is a human melanoma cell line, derived from metastatic melanoma and is often used to evaluate the cytotoxicity of drugs targeting melanoma cells.

Table 2. Antimicrobial test result of angucycline-producer strains extracts.

	*# Gram-negative pathogen				Gram-positive pathogen				Yeast pathogen	
	K. pneumoniae ATCC 700603		*E. coli* ATCC 25922		*MRSA* _MB5393		*MSSA* ATCC 29213		*C.albicans ATCC* 64124	
	%INH: (day 1)	%INH: (day 2)	%INH: (day 1)	%INH: (day 2)	%INH: (day 1)	%INH: (day 2)	%INH:(day 1)	%INH: (day 2)	%INH: (day 1)	% INH:(day 2)
Streptomyces sp. QL37	-25	-25	-16	-14	-6	-27	-55	-89	-51	-44
	-6	-15	-14	-16	-46	-42	-64	-90	-55	-46
	-24	-38	-30	-22	-39	-40	-66	-90	-50	-42
	-19	-41	-20	-27	-54	-55	-72	-94	-58	-66
	-20	-19	-17	-18	-64	-59	-94	-97	-57	-58
Streptomyces sp. Go-475	-20	6	-31	-32	-87	-88	-75	-86	-77	-78
	-2	3	-32	-13	-78	-79	-68	-83	-77	-67
	-21	2	-23	-9	-80	-82	-67	-79	-85	-89
	1	-2	-24	-22	-80	-84	-51	-72	-78	-88
	-21	7	-25	-20	-82	-84	-43	-80	-84	-90
Streptomyces sp. CS057	-22	-23	-21	-17	-102	-101	-104	-102	6	11
	-28	-24	-23	-19	-102	-103	-104	-103	2	19
	-24	-20	-18	-16	-104	-101	-105	-105	-5	23
	-25	-19	-16	-19	-103	-101	-106	-103	-2	13
	-28	-20	-20	-15	-103	-101	-107	-103	-4	9
S. albus	-24	-34	-44	-35	-97	-99	-98	-100	-95	-93
	-22	-36	-37	-34	-96	-99	-98	-98	-94	-92
	-25	-30	-42	-34	-98	-98	-102	-97	-96	-95
	-12	-18	-47	-36	-94	-94	-93	-98	-93	-90
	-13	-31	-50	-47	-94	-95	-98	-99	-95	-98
S. venezuelae	-25	-19	-31	-20	-100	-98	-99	-94	-89	-95
	-64	-4	-1	-22	-101	-103	-96	-95	-95	-96
	-42	-22	-32	-20	-101	-99	-98	-97	-97	-97
	-13	13	-24	-18	-101	-100	-104	-98	-96	-97
	-39	-15	-28	-17	-101	-100	-108	-101	-92	-99
Media Blank	-15	-11	-12	-8	-6	-14	-35	-28	9	29
	-15	-6	-12	-7	12	0	-15	14	5	32
	-20	-12	-11	-7	9	-2	6	6	8	36
	-16	-6	-12	-11	5	-6	39	0	9	28
	-28	-24	-18	-13	5	-3	-54	0	11	17

***** Negative values indicate reduced % of growth compared to the control. Pathogens are indicated in five different colours (yellow, green, orange, pink and purple). In bright color, inhibition of over 75% is displayed, in pale color, inhibition of over 50%.

Each strain extract was measured in five replicates, and two time-point measurements per pathogen (INH% day 1 and INH% day 2)

A549 is a human lung carcinoma cell line widely used in cancer research due to its relevance to non-small cell lung cancer (NSCLC), which accounts for the majority of lung cancer cases. The A549 cells are particularly valuable for studying the effects of potential therapeutics on lung cancer. MCF7 is a well-known human breast cancer cell line key model for studying hormone-responsive breast cancers and for evaluating the efficacy of anti-estrogen therapies and other anticancer agents. HepG2 is a human liver cancer cell line that has retained many of the biochemical functions of normal liver cells, making it a useful model for studying liver cancer biology, drug metabolism, and hepatotoxicity, as well as for testing potential anticancer agents targeting liver cancer. By utilizing this diverse panel of cancer cell lines, we aimed to assess the broad-spectrum anticancer potential of our angucycline-producing strains, as well as their efficacy across different types of cancer.

The results of the assays are summarized in Table 3. Extracts from a, b, and c, demonstrated selective cytotoxicity towards cancer cell lines, although these strains did not inhibit growth of *C. albicans*. This selective activity suggests that these extracts may target some cancer cell lines specifically, rather than exerting a broad toxic effect on all eukaryotic cells. To see if compounds may be used in cancer therapy, first extensive analysis of selectivity and safety of the compounds in the extracts is required. *S. albus* and *Streptomyces* sp. G0-475 exhibited cytotoxic activity against all tested cancer cell lines. Interestingly, *Streptomyces* sp. G0-475 displayed a bioactivity profile distinct from that of *Streptomyces* sp. QL37, which showed cytotoxicity only against cell lines A2058, MCF7 and HepG2. This suggests that either the two strains produce different angucyclines, or the observed bioactivity may (in part) come from BGCs for other compounds with anticancer activity.

Streptomyces sp. QL37 is very potent against a wide variety of cancer cell lines, achieving complete or near-complete inhibition in most of the tested cell lines, particularly those representing pancreatic, breast, liver, and melanoma cancers. However, lung cancer cells (A549) appeared slightly more resistant, showing partial inhibition in some cases. This strain could potentially be a strong candidate for developing anti-cancer compounds due to its broad-spectrum activity. However, several *Streptomyces* sp. QL37 mutants displayed even stronger cytotoxic effects (T., 2019). Extracts of the QL37Δ*lugOV* mutant showed particularly high anticancer activity, especially those obtained from cultures grown in rich media (ISP4). Thus, while wildtype *Streptomyces* sp. QL37 already has considerable anticancer potential, certain mutants show enhanced activity, most likely due to the accumulation of intermediates that are otherwise metabolized further.

Streptomyces sp. CS057 demonstrated potent anticancer activity across several cancer cell lines. It achieved 100% inhibition against pancreatic (MiaPaca), breast (MCF7), and liver (HepG2) cancer cells, indicating complete effectiveness. For melanoma (A2058), the inhibition ranged from 85% to 99%, showing slightly reduced effectiveness compared to other strains. In contrast, lung cancer cells (A549) displayed partial resistance, with inhibition values ranging from 56% to 78%, suggesting moderate efficacy without full inhibition of these cells.

The results of the extracts from *S. venezuelae* gave 100% growth inhibition for the MiaPaca (pancreatic), MCF7 (breast), and HepG2 (liver) cell lines. Against

A2058 melanoma cells near-complete inhibition was seen, with inhibition of 82% to 100%. In contrast, extracts showed only moderate effectiveness against A549 lung cancer cell line, with values between 53% and 82%.

Table 3. Summary of results of the anticancer assays.

	*MiaPaca	A2058	A549	MCF7	HepG2
Streptomyces sp. QL37	1	-59	-21	-100	-100
	1	-76	-15	-100	-100
	1	-74	-18	-100	-100
	0	-95	-34	-100	-100
	-95	-100	-88	-100	-100
Streptomyces sp. Go-475	-98	-99	-98	-99	-99
	-98	-98	-94	-99	-99
	-97	-96	-93	-96	-97
	-97	-98	-93	-97	-99
	-95	-98	-93	-98	-99
Streptomyces sp. CS057	-100	-99	-82	-99	-91
	-100	-98	-70	-99	-92
	-100	-98	-73	-100	-92
	-100	-87	-57	-96	-86
	-100	-92	-64	-98	-88
S. albus	-100	-100	-103	-99	-100
	-100	-100	-102	-98	-100
	-100	-100	-102	-100	-100
	-100	-100	-102	-100	-100
	-100	-100	-102	-100	-100
S. venezuelae	-2	-100	-82	-100	-100
	-100	-100	-102	-100	-99
	-20	-100	-98	-100	-100
	-17	-100	-99	-100	-100
	-98	-100	-102	-100	-100
Media Blank	-13	2	-12	2	4
	-7	2	-8	-2	4
	-8	3	-9	9	3
	-8	3	-6	4	3
	-5	-1	-16	-31	-38

*Negative values indicate reduced % of growth compared to the control. Orange color indicate a growth inhibition over 75%. Pink color indicate growth inhibition between 50-75%.

Conclusions

In our research we identified key genomic, metabolomic, and bioactivity characteristics of the small collection of angucycline-producer strains. Through the targeted collection and comparative analysis of the BGCs from these strains, we identified significant similarities, but also major differences. Notably, the angucycline BGC from *S. scopuliridis* exhibited high identity to that of *Streptomyces* sp. QL37.

Metabolomic profiling using LC-MS/MS and the molecular networking and annotation tools FBMN and SIRIUS, revealed that while there is considerable overlap in the angucycline metabolites produced by different strains, unique profiles also emerged, with certain strains like *S. venezuelae* and *Streptomyces* sp. G0-475 displaying distinct spectral families. These findings were further corroborated by the bioactivity assays, which demonstrated that the antimicrobial and anticancer activities of the strains are mediated by diverse sets of metabolites. The angucycline family comprises a diverse group of natural products, with members exhibiting varying modes and intensities of bioactivity. However, they are predominantly biosynthesized by a single BGC. Typically, *Streptomyces* strains possess between 25–70 distinct BGCs (Belknap *et al.*, 2020), which specify a broad range of natural products. Therefore, we cannot solely attribute the observed bioactivity to angucyclines. For instance, despite their close genomic and metabolomic resemblance, *Streptomyces* sp. QL37 and *Streptomyces* sp. G0-475 exhibited different bioactivity profiles, suggesting the presence of unique bioactive compounds in each strain, produced from the angucycline BGCs and/or from other BGCs that may specify bioactive molecules. The distinct bioactivity profiles observed in our study suggest that the antimicrobial and anticancer properties of angucycline-producing strains are influenced by a combination of shared and unique metabolites, offering promising avenues for the discovery of novel therapeutics. However, as stated above, the observed bioactivities are likely a combination of natural products that are derived from different BGCs, and not only to angucyclines and angucyclinones.

Thus, we presented a genomic and metabolomic study of known angucycline producers, highlighting their potential as sources of antibiotics and anticancer compounds. Future research should focus on optimizing the culture conditions and exploring the regulatory mechanisms that govern the production of these metabolites to fully harness their biosynthetic potential. Moreover, the selective anticancer activity observed in some extracts warrants further investigation to evaluate their therapeutic potential and specificity, with the goal of identifying new leads for anticancer drug development.

Our initial research on the angucycline BGCs across different strains aimed to identify new producers of lugdunomycin. As demonstrated in Chapter 3, a second biosynthetic gene cluster (BGC23) in *Streptomyces* sp. QL37 is responsible for the *iso*-maleimycin moiety of lugdunomycin. Although the strains

we examined do not contain BGC23 and thus are not capable of producing lugdunomycin, their homologous core PKS machinery and ability to produce angucyclinone compounds make them interesting candidate hosts for the production of lugdunomycin or derivatives. Introducing key enzymes such as LugOIII and LugOV may be sufficient to produce precursors for lugdunomycin-like compounds. By leveraging a "toolbox" approach to introduce these enzymes into suitable hosts, novel angucycline derivatives may be generated, thereby expanding the chemical diversity of this class of compounds for potential bioactivity studies.

Materials and Methods

Microorganisms and culturing conditions

All media and routine *Streptomyces* techniques were done following routine protocols (Kieser, 2000). *Streptomyces* sp. QL37 was obtained from soil collected in the Qinling mountains of the People's Republic of China, as previously described (Zhu *et al.*, 2014, Wu *et al.*, 2019b). The strain was deposited in the Centraal Bureau voor Schimmelcultures (CBS) in Utrecht, The Netherlands, with the accession number 138593. The strains included in our study were *Streptomyces albus (Jin et al., 2018)*, *Streptomyces ambofaciens (Bunet et al., 2008)*, and *Streptomyces venezuelae (Raghavan & Groisman, 2010)* (all of them purchase from DSMZ), and the strains *Streptomyces* sp. MP131-18 (Paulus *et al.*, 2017), *Streptomyces* sp. CS057 (Malmierca *et al.*, 2018), *Streptomyces* sp. Go-475 (Kibret *et al.*, 2018) *Streptomyces* ICBB8309 and *Streptomyces* ICBB8415 (Fotso *et al.*, 2008) were kindly gifted by collaborators. More information about these strains can be found in table S1.

For solid cultivation of the angucycline producers, strains were grown confluently with approximately 2×10^6 spores on *Streptomyces* minimal media (StrepMM) agar plates containing 1% glycerol and 0.5 % mannitol (w/v) as the carbon sources.

For the antimicrobial assays, five different pathogens: *E. coli* ATCC 25922 (EC), *Klebsiella pneumoniae* ATCC 700603, Methicillin-resistant *S. aureus* MB 5393 (MRSA), *S. aureus* ATCC 29213 (Methicillin-sensitive) and *Candida albicans* ATCC 64124, were tested. For the anti-cancer assays, the following cell lines were tested: human melanoma (A2058), human lung carcinoma (A549), human hepatocellular carcinoma (HepG2), human colon cancer (HT29),

human breast cancer (MCF7) and human pancreatic cancer (MIA PaCa-2). For high throughput screenings (HTS), targets were grown in different media. *S*. *aureus* ATCC 29213 and *K*. *pneumoniae* ATCC 700603 were grown in Mueller–Hinton II (beef extract 2 g/L, acid hydrolysate of casein 17.5 g/L and starch 1.5 g/L). *E*. *coli* ATCC 25922, Methicillin-resistant *S*. *aureus* MB 5393 and *A*. *baumannii* MB 5973 were grown in Luria–Bertani broth (10 g/L peptone, 5 g/L yeast extract and 5 g/L sodium chloride) (Audoin *et al.*, 2013). *C*. *albicans* ATCC 64124 were grown in RPMI-1640 modified medium [10.4 g/L of RPMI-1640 medium (R8755; Sigma, St. Louis, MO), 6.7 g/L of yeast nitrogen base, 1.8% (w/v) glucose, and 40 mM HEPES (pH = 7.1)] (Audoin *et al.*, 2013, Monteiro *et al.*, 2012, Vicente *et al.*, 2009, Santos *et al.*, 2020)

Genomic and Bioinformatic Analysis

The genomes of the angucycline-producing strains were obtained from the National Center for Biotechnology Information (NCBI) database. To identify polyketide synthase type II (PKS-II) biosynthetic gene clusters (BGCs), the genomes were analyzed using antiSMASH (Blin *et al.*, 2023). The identified PKS-II BGCs were then downloaded and further analyzed. These BGCs were uploaded to Clinker (Gilchrist & Chooi, 2021), where they were compared with a identity threshold set at 45% to assess the conservation and variations among the gene clusters.

Preparation of crude extracts and metabolite profiling

The crude extracts from the solid-grown cultures were prepared as followed. After seven days of growth at 30 °C the agar plates were cut into small pieces and soaked in 25 ml of ethyl acetate for 24h. This process was repeated three times. After evaporation, extracts were dissolved in methanol (MeOH) to a final concentration of 1 mg/mL.

For metabolite profiling of the crude extracts, liquid chromatography-tandem mass spectrometry (LC-MS/MS) analysis was performed using a Shimadzu Nexera X2 ultra high-performance liquid chromatography (UPLC) system, equipped with a photodiode array detector (PDA), coupled to a Shimadzu 9030 QTOF mass spectrometer, equipped with an electrospray ionization (ESI) source unit, which included a calibrant delivery system (CDS). For the LC separation, 2 µL of the sample was injected into a Waters Acquity HSS C18 column (1.8 µm, 100 Å, 2.1 × 100 mm). The column temperature was

maintained at 30 °C, and the separation was carried out at a flow rate of 0.5 mL/min. Solvent A consisted of 0.1% formic acid in H2O, while solvent B comprised 0.1% formic acid in acetonitrile (ACN). The gradient elution profile started with 5% B for 1 min, followed by a linear increase from 5% to 85% B over 9 min, then a steep increase to 100% B over 1 min, and finally, an isocratic hold at 100% B for 4 min. To re-equilibrate the column, it was set to 5% B for 3 min before the subsequent run.

The PDA acquisition was conducted within the wavelength range of 200–600 nm, with a scan rate of 4.2 Hz and a slit width of 1.2 nm. The temperature of the flow cell was maintained at 40 °C throughout the analysis. All samples were analyzed in positive polarity mode using a data-dependent acquisition method. This involved acquiring full scan MS spectra within the m/z range of 100–1700 at a scan rate of 10 Hz, with intelligent data (ID) enabled. Subsequently, two data-dependent MS/MS spectra within the same m/z range and scan rate were obtained for the two most intense ions detected in each scan. Collision-induced dissociation (CID) was employed for ion fragmentation with a fixed collision energy (CE) of 20 eV. To avoid re-analysis of previously fragmented ions, they were excluded from selection for 1 s before being eligible for fragmentation again. The parameters used for the ESI source were as follows: interface voltage of 4 kV, interface temperature of 300 °C, nebulizing gas flow rate of 3 L/min, and drying gas flow rate of 10 L/min42.

Statistical analysis and Feature-based molecular network analysis

Prior to statistical analysis, mzXML files, which were converted using Shimadzu LabSolutions Postrun Analysis, were imported into Mzmine 2.53 (Pluskal *et al.*, 2010) and processed as previously described (Machushynets *et al.*, 2019). The *m/z* tolerance was set to 0.002 *m/z* or 10.0 ppm, RT tolerance was set to 0.05 min, noise level was set to 2.0E2 and the minimum absolute intensity was set to 5.0E2 by default unless specified otherwise. Raw data were cropped to RT 0.5–12 min. Mass ion peaks were detected for MS level 1 (positive polarity, mass detector: centroid) and their chromatograms were built using ADAP chromatogram builder (Myers *et al.*, 2017) (MS level 1, minimum group size in number of scans: 10; group intensity threshold: 2.0E2). The detected peaks were smoothed (filter width: 9), and the chromatograms were deconvoluted (algorithm: local minimum search; chromatographic threshold: 85%; search minimum in RT range: 0.05; minimum relative height: 1%; minimum ratio of

peak top/edge: 2; peak duration: 0.03–3.00 min). The detected peaks were deisotoped (monotonic shape; maximum charge: 2; representative isotope: most intense). Peak lists from different extracts were aligned (weight for RT = weight for *m/z* = 20; compare isotopic pattern with a minimum score of 50%). Missing peaks detected in at least one of the samples were filled with the gap-filling algorithm (RT tolerance: 0.1 min). Among the peaks, we identified fragments (maximum fragment peak height: 50%), adducts ($[M+Na]^+$, $[M+K]^+$, $[M+NH_4]^+$, maximum relative adduct peak height: 3000%) and complexes (Ionization method: $[M+H]^+$, maximum complex height: 50%). Duplicate peaks were filtered. The aligned peak list was exported as a comma-separated file for statistical analysis. Only features with peak area values four times higher than in the media blank were kept. Statistical analysis was performed using MetaboAnalyst (Pang *et al.*, 2024), where log transformation and pareto scaling was initially applied to the data. The processed data were subjected to principal components analysis (PCA).

For feature-based molecular networking, the converted files were processed using MZmine 2.53 (Pluskal *et al.*, 2010) with the previously mentioned parameters, along with two additional steps. First, an extra mass detection step was performed for MS level 2 using positive polarity, with the mass detector set to centroid mode and a noise level threshold of 2.0E0. After processing, the peak list was filtered to retain only features with corresponding MS/MS scans. The filtered peak list was then exported for use in FB-GNPS. The molecular networks were visualized using Cytoscape 3.7.2.

Compound classes prediction from MS^2 spectra was done using SIRIUS and other tools from the SIRIUS (Duhrkop *et al.*, 2021) data analysis package, such as ZODIAC (Ludwig *et al.*, 2019) and CANOPUS (Kim *et al.*, 2021, Duhrkop *et al.*, 2021, Djoumbou Feunang *et al.*, 2016) compound class prediction. For SIRIUS, instrument was set as Q-TOF, the MS2 mass accuracy was set to 10 ppm, the default parameters were used for the rest of the options. The possible ionizations were set for $[M+H]^+$, $[M+K]^+$ and $[M+Na]^+$, and the selected databases for molecular formula search were COCONUT, GNPS, NPD PubMed and PubChem. For ZODIAC, all default parameters were used.

Bioactivity Assays

The antibacterial and antifungal properties were assessed using previously established methods with pathogenic microorganisms from Fundación MEDINA's

collection, as previously described (Audoin *et al.*, 2013, Monteiro *et al.*, 2012, Santos *et al.*, 2020). In summary, single colonies of each microorganism were cultured overnight in their respective media at 37 °C and 220 rpm. The cultures were then diluted to achieve assay inoculum concentrations of approximately 1.1×10^6 CFU/mL for methicillin-resistant *Staphylococcus aureus* MB 5393, 5.0×10^5 CFU/mL for *Acinetobacter baumannii* MB 5973, *Escherichia coli* ATCC 25922, *Klebsiella pneumoniae* ATCC 700603, and *Staphylococcus aureus* ATCC 29213, and 2.5×10^4 CFU/mL for *Aspergillus fumigatus* ATCC 46645. For *Candida albicans*, the optical density at 660 nm of the liquid culture was adjusted to 0.25 and then diluted 1:100 to prepare the assay inoculum.

In the assays, 90 µL of the diluted inoculum were combined with 10 µL of extract per well. Positive and negative internal controls were included on the plates following the methodologies previously described (Santos *et al.*, 2020, Monteiro *et al.*, 2012, Audoin *et al.*, 2013). Absorbance (at 600 nm) was measured using an Envision plate reader. Data analysis was conducted using Genedata Screener software (Genedata, Inc., Basel, Switzerland), which calculated the percentage of growth inhibition induced by the extracts and the RZ′ factor to assess assay robustness (Zhang *et al.*, 1999). In the experiments conducted, RZ′ factor values ranged from 0.87 to 0.92.

The percentage of growth inhibition was determined using the following formula:

$$\%\ Inhibition = 100 - \left(100x\frac{(T_{FE} - T_{0E}) - (T_{FB} - T_{0B})}{(T_{FG} - T_{0G}) - (T_{FB} - T_{0B})}\right)$$

where T0 represents the absorbance at 0 hours, TF is the absorbance at 24 hours, E is the absorbance of the extract wells, B represents the blank wells, and G denotes the control growth wells.

Anticancer assays

The cytotoxic effects of the extracts were evaluated using the high-throughput screening (HTS) assays portfolio from MEDINA (Cautain et al., 2015). Cytotoxicity was assessed using the MTT [3- (4,5-dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide] assay, where absorbance was measured at 570 nm. The extracts were tested against human hepatocellular carcinoma (HepG2) cells, human melanoma (A2058), human lung carcinoma (A549), human

hepatocellular carcinoma (HepG2), human colon cancer (HT29), human breast cancer (MCF7), and human pancreatic cancer (MIA PaCa-2) cells.

Aknowledgements

We are grateful to Andriy Luzhetskyy, a.luzhetskyy@mx.uni-saarland.de; Jose A salas, jasalas@uniovi.es; Philip J Protea, phil.proteau@oregonstate.edu and Sergey B. Zotchev, sergey.zotchev@univie.ac.at for providing the angucycline-producer strains found in this manuscript. Also, we are grateful to Fundación Medina, for the bioactivity testing of the extracts performed in their facilities.

Supplementary information

Table S1. Overview of strains included in the study and corresponding angucyclines documented in the literature.

Strain	BGC	Reference
Streptomyces sp. QL37	Lugdunomycin	(Wu *et al.*, 2019b)
Streptomyces scopuliridis	Thioangucycline	https://mibig.secondarymetabolites.org/repository/BGC0002477/index.html#r1c1 https://www.dsmz.de/collection/catalogue/details/culture/DSM-41917
Streptomyces sp. ICBB8390	Limamycin	(Fotso *et al.*, 2008)
Streptomyces sp. ICBB8415	Limamycin	(Fotso *et al.*, 2008)
Streptomyces sp. Go-475	8-O-methyltetrangomycin	(Kibret *et al.*, 2018)
Streptomyces sp. MP131-18	Prejadomycin	https://antismash.secondarymetabolites.org/upload/bacteria-d62b6964-203f-4f75-9b02-ff6fec535ef6/index.html (Paulus *et al.*, 2017)
Streptomyces sp.CS057	Warkmycin	(Malmierca *et al.*, 2018)
Streptomyces venezuelae	Jadomycin	(Raghavan & Groisman, 2010) https://www.dsmz.de/collection/catalogue/details/culture/DSM-40230
Streptomyces ambofaciens	Alpomycin	(Bunet *et al.*, 2008) https://www.dsmz.de/collection/catalogue/details/culture/DSM-40053
Streptomyces albus	Fluostatin	https://www.mdpi.com/1660-3397/16/3/87 https://www.dsmz.de/collection/catalogue/details/culture/DSM-41398

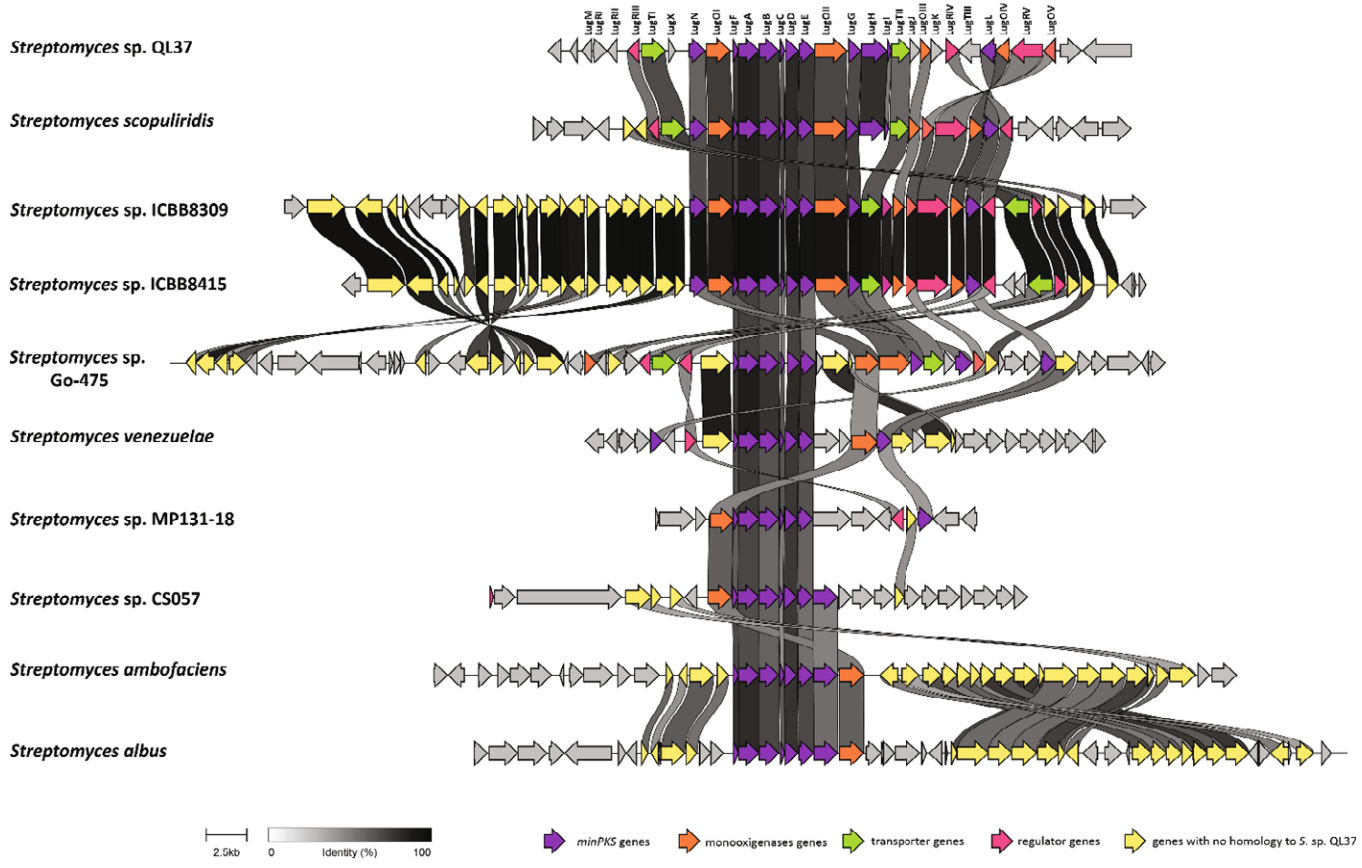

Figure S1. Comparison of annotated angucycline biosynthetic gene clusters to the angucycline BGC (*lug*) of *Streptomyces* sp. QL37. Genes are represented by arrows and colored based on function annotation. The % of identity is displayed by the color intensity of the strokes between the different BGCs. Comparison was done using Clinker (Gilchrist & Chooi, 2021).

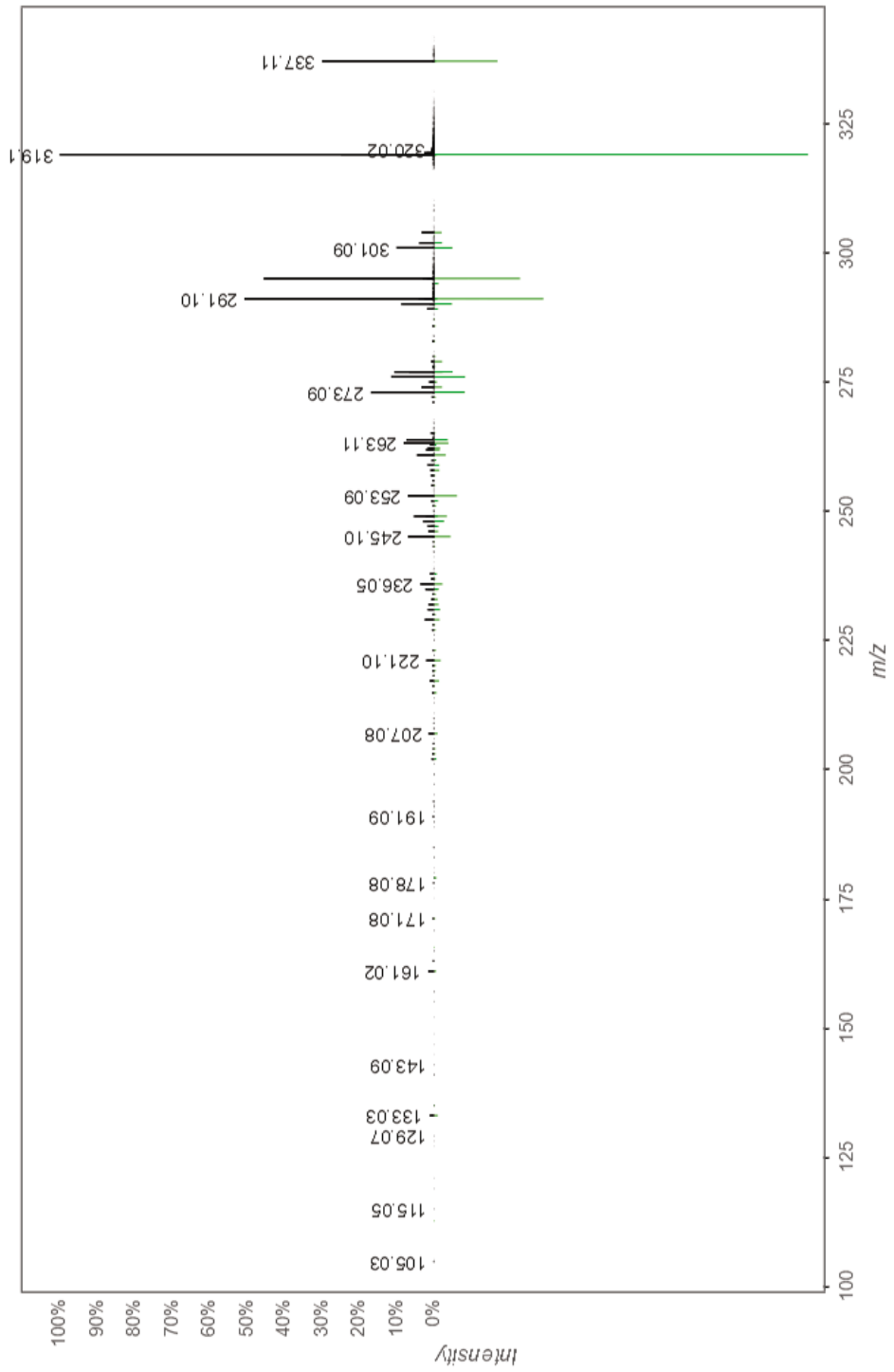

Figure S2. Mirror plot for the GNPS library hit 8-O-Methyltetrangomycin, *m/z* 337.1071.

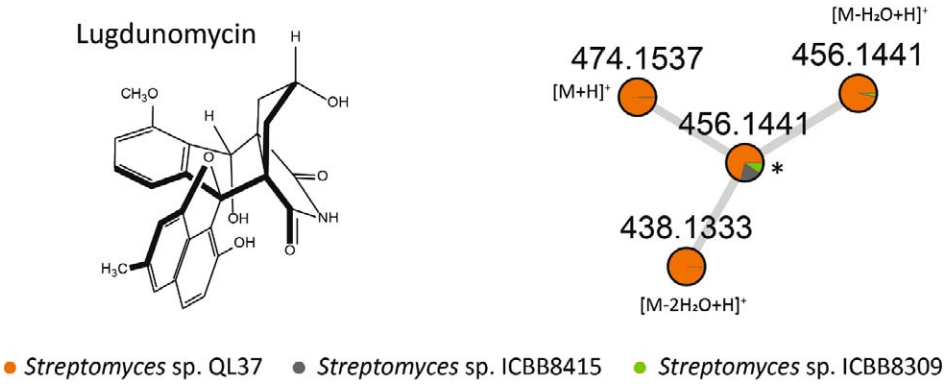

Figure S3. Identification of lugdunomycin-like molecules using molecular networking. The structure of lugdunomycin is depicted on the left. On the right, the lugdunomycin network is shown, consisting of four nodes. The three nodes with the 7.1-min retention time are exclusively present in the *Streptomyces sp.* QL37 extract (orange). The fourth node (*m/z* 456.1441, retention time 6.8 min, highlighted with a star) is predominantly found in *Streptomyces* sp. QL37 and corresponds to a lugdunomycin isomer.

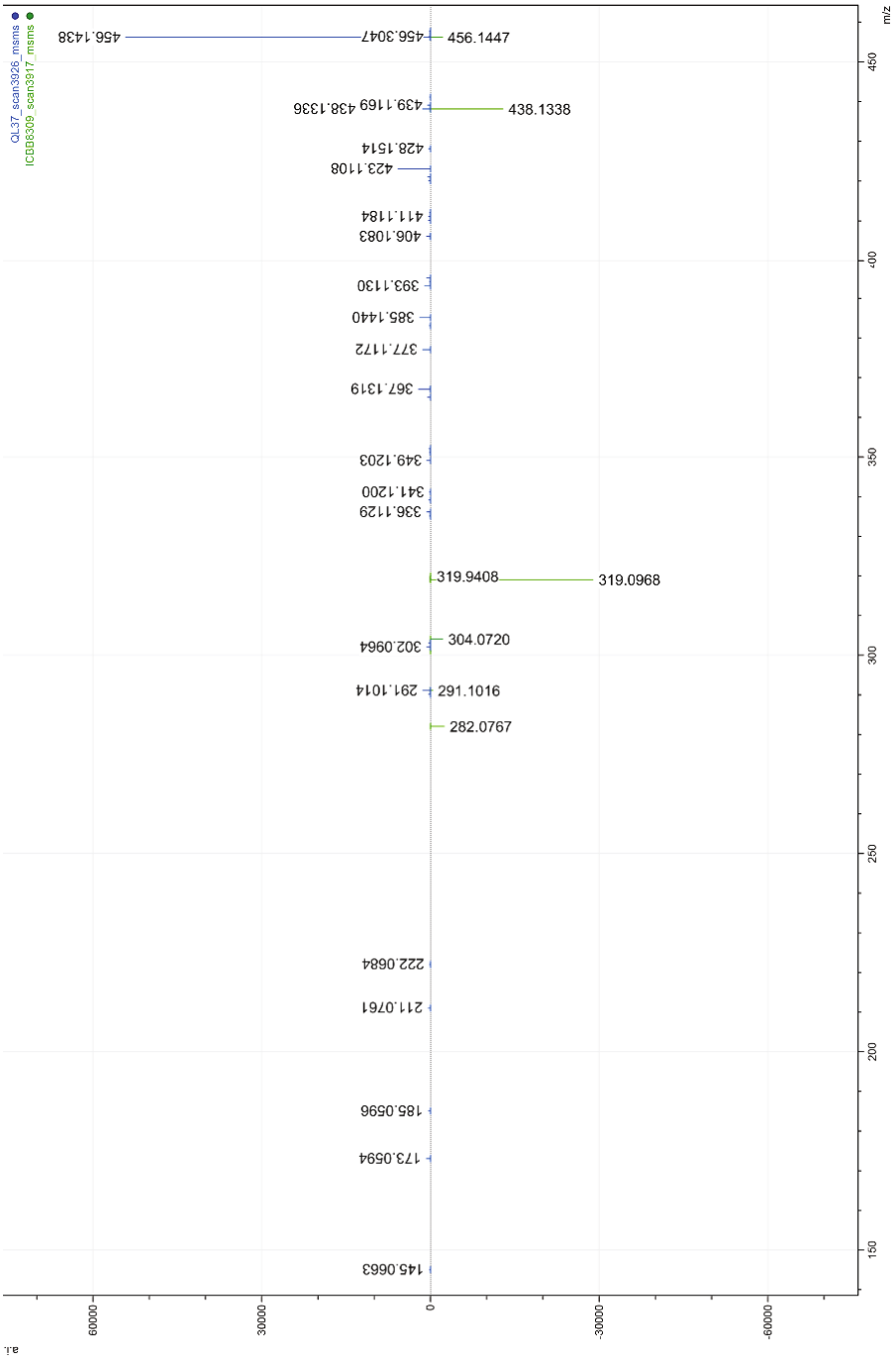

Figure S4. Direct MS/MS comparison of scan 3926 belonging to the feature *m/z* 456.1441 RT 6.8 mins in extract of *Streptomyces* sp. QL37 and scan 3917 of the same feature from *Streptomyces* sp. ICBB8309.

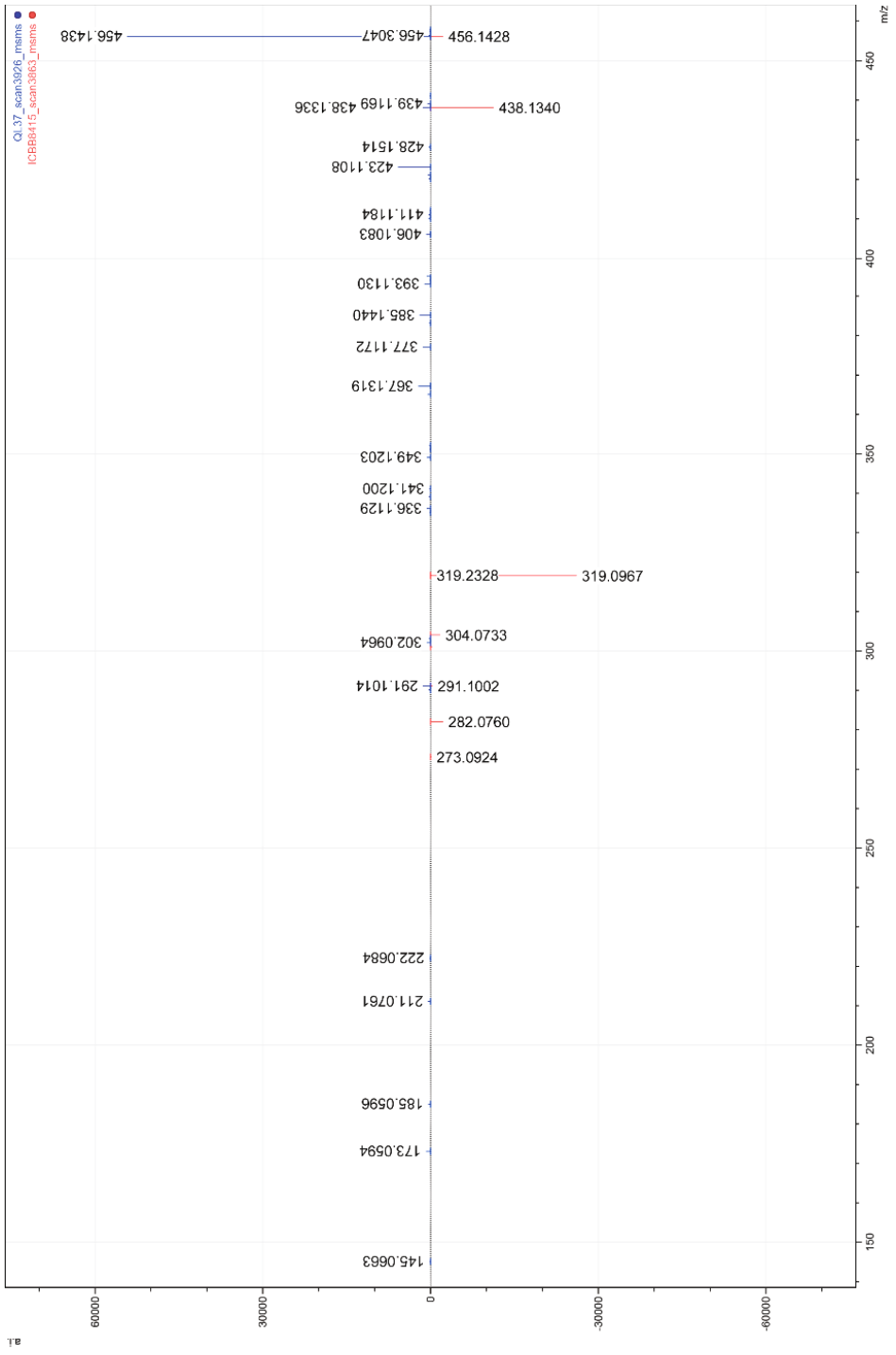

Figure S5. Direct MS/MS comparison of scan 3926 belonging to the feature *m/z* 456.1441 RT 6.8 mins in *Streptomyces* sp. QL37 extract and scan 3863 of the same feature from *Streptomyces* sp. ICBB8415.