Analysis of the angucycline biosynthetic gene cluster in Streptomyces sp. QL37 and implications for lugdunomycin production

Heul, H.U. van der

The *lug* gene cluster is sufficient for the biosynthesis of angucyclines and limamycins but not lugdunomycin

Helga U. van der Heul
Michiel Uiterweerd
Isabel Nuñez-Santiago
Guillermo Guerrero Egido
Changsheng Wu
Adriaan J. Minnaard
Somayah S. Elsayed
Gilles P. van Wezel

ABSTRACT

Oxidative ring cleavage of the angucycline C-ring expands the chemical diversity of the angucycline family. An exciting example is the angucycline derivative, lugdunomycin, produced by *Streptomyces* sp. QL37. This molecule is characterised by its unusual structural features, including a heptacyclic ring system, a spiroatom, two all-carbon stereocenters, and a benzaza[4,3,3]propellane motif. The final step in the biosynthesis of lugdunomycin was previously proposed to be catalysed by a Diels-Alder reaction between a C-ring cleaved angucycline and *iso*-maleimycin. Heterologous expression of the angucycline biosynthetic gene cluster (BGC; designated *lug*) of *Streptomyces* sp. QL37 resulted in the production of C-ring rearranged angucyclines and limamycins, but not lugdunomycin. In this study we applied a paired-omics strategy to find the missing factor in the biosynthesis of lugdunomycin, using genomics, metabolomics, and temporal RNA-seq profiling. Molecular networking of the metabolome of the wild-type strain and its *pks* mutant identified all molecules that are not derived from the *lug* cluster. Interestingly, *iso*-maleimycin was detected in a *pks* mutant of the *lug* gene cluster, which implies that the molecule is derived from a BGC different from that for the angucyclines. Transcriptome analysis revealed that transcription of the *lug* gene cluster is coordinated with a BGC homologous to the one for maleimycin biosynthesis in *Streptomyces showdoensis* ATCC 15227, and this BGC is thus the major candidate to produce the dienophile for the final step in lugdunomycin biosynthesis.

INTRODUCTION

Streptomyces species are one of the most important resources for bioactive molecules with high chemical diversity (Barka *et al.*, 2016). One of these chemically diverse structures is the angucycline group, the largest molecular class synthesised by polyketide type II synthases (Kharel *et al.*, 2012, Mikhaylov *et al.*, 2021). They exhibit a wide range of antimicrobial and anti-cancer activities possibly due to their variation in chemical modifications of their typical benz[*a*]anthracene backbone, such as glycosylation, epoxidation and hydroxylation (Kharel *et al.*, 2012, Buttner *et al.*, 2018, Schafer *et al.*, 2015, Lehka *et al.*, 2015, Yoon *et al.*, 2019, Fakhruzzaman *et al.*, 2015). The typical angucycline backbone can be drastically modified by oxidative cleavage of one of the aromatic rings which may be followed by incorporation of various other molecules (Fan & Zhang, 2018, Wang *et al.*, 2015, Rix *et al.*, 2003, Tibrewal *et al.*, 2012, Fan *et al.*, 2012b). Previously we discovered the complex angucycline derived molecule, lugdunomycin, characterised by its benzaza[4,3,3]propellane-6-spiro-2'-2H-naphtho[1,8-bc]furan backbone (Wu *et al.*, 2019). The molecule was discovered by screening *Streptomyces* strains derived from remote areas, and is produced by *Streptomyces* sp. QL37 isolated from soil from the Qinling mountains (Figure 1) (Zhu *et al.*, 2014b). However, lugdunomycin is produced in very low amount under all growth conditions, whereas the angucyclines are typically produced in rather high amounts (Wu *et al.*, 2019). It is crucial to increase the production of the molecule to obtain more knowledge on its bioactivity and functional role. To do so, we use a fundamental approach by obtaining insights in the steps required for the biosynthesis of lugdunomycin. Previously we hypothesised a biosynthesis pathway wherein a typical non-rearranged angucycline is oxygenated by a Baeyer-Villiger oxygenase, leading to a C-ring rearranged angucycline (Wu *et al.*, 2019). Subsequently the C-ring rearranged angucycline (diene) reacts with *iso*-maleimycin (dienophile) in a Diels-Alder reaction. Indeed *iso*-maleimycin was also detected in the extracts of *Streptomyces* sp. QL37 (Uiterweerd, 2020). We hypothesised that *iso*-maleimycin is derived from the limamycins via a highly complex biosynthetic route. In chapter 3 we proposed a biosynthetic gene cluster (BGC) that is required for the production of angucyclines, pratensilin A, limamycin and lugdunomycin, *lug*. The functional role of this BGC was confirmed by deletion of the genes involved in the production of the typical angucycline backbone (*lug-pks* mutant). This polyketide type II BGC contains amongst others five genes for oxygenases, which

likely mediate C-ring cleavage (Chapter 5). One approach to determine whether the complete set of genes required for the biosynthesis of a molecule are present in a BGC is heterologous expression in an optimised host such as *S. coelicolor* A3(2) M1152 (Liu *et al.*, 2018, Gomez-Escribano & Bibb, 2011, Izawa *et al.*, 2014). In this strain the BGCs specifying the production of actinorhodin (Act), undecylprodigiosin (Red), Calcium-dependent antibiotic (CDA) and coelimycin P1 (CPK) have been deleted. This removes the possibility that that energy and nutrients are used by other pathways. In this study we introduced a large clone harbouring the *lug* gene cluster from the *Streptomyces* sp. QL37 genome into *S. coelicolor* M1152, producing angucyclines and C-ring cleaved derivatives, as well as limamycins, but not lugdunomycin. Paired-omics using metabolomics, genomics and temporal RNA-seq, suggests that a second BGC is required to produce the dienophile *iso*-maleimycin required for the final step in lugdunomycin biosynthesis.

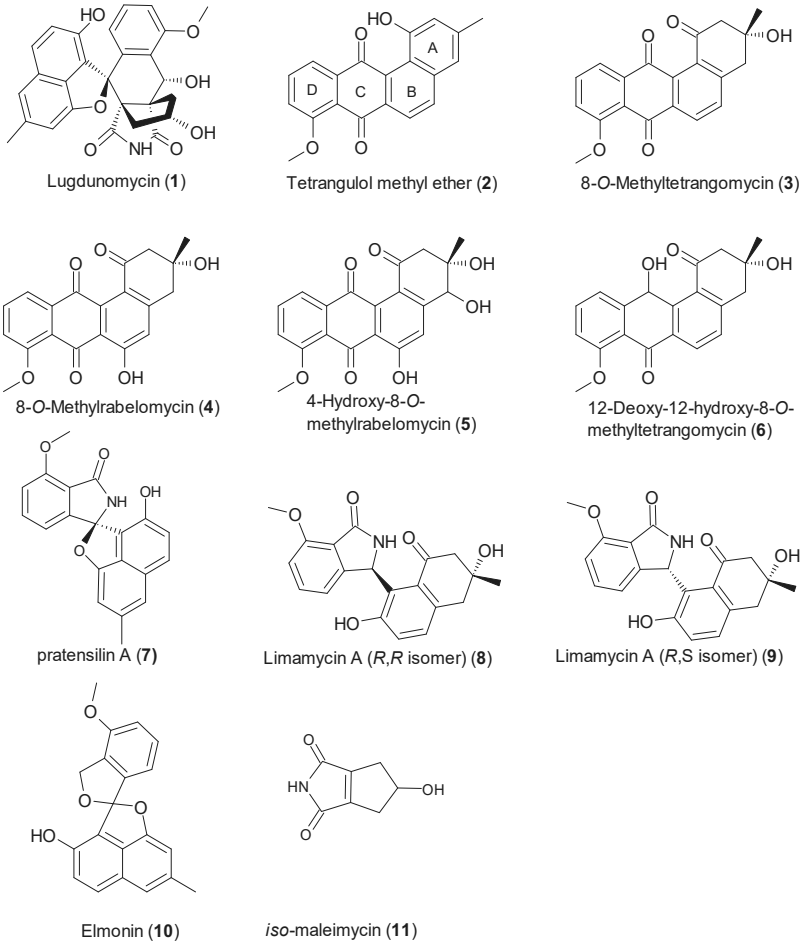


Figure 1 Structures of the compounds discussed in this study.
Lugdunomycin (**1**), the non-rearranged angucyclines (**2**, **3**, **4**, **5**, **6**) and the limamycins (**8**, **9**) were previously isolated from *Streptomyces* sp. QL37 (Wu *et al*., 2019). The structure of compound **7** was later revised to be pratensilin A (Mikhaylov *et al*., 2021). Elmonin (**10**) and *iso*-maleimycin (**11**) were detected in the extracts of *Streptomyces* sp. QL37 by comparison with a standard (Figure S3) (Uiterweerd, 2020).

RESULTS

Heterologous expression of the *lug* gene cluster in *Streptomyces coelicolor* M1152

The *lug* gene cluster of *Streptomyces* sp. QL37 is similar to other angucycline BGCs, but so far, no strain has been reported that produces lugdunomycin and study of several angucycline producers in our laboratory failed to detect the molecule (our unpublished data). A question we therefore sought to resolve is whether the cluster is sufficient to produce lugdunomycin, and if so, if we could enhance its expression. To this end, the cluster was expressed in the angucycline non-producer *S. coelicolor* A3(2) M1152, a derivative of M145 that has been optimised as chassis strain for secondary metabolite production (Gomez-Escribano & Bibb, 2011). In order to do so, a P1-derived artificial chromosome (PAC) library of the genome from *Streptomyces* sp. QL37 was constructed using pESAC-13-A, an *E. coli–Streptomyces* artificial chromosome (Tocchetti *et al.*, 2018, Jones *et al.*, 2013). Three sets of primers were used to define the start (primers 1 and 2), middle (primers 3 and 4) and end (primers 5 and 6) of the BGC, respectively (Table 2). Screening of the library revealed two clones that contained the *lug* gene cluster, one included a 120 kb fragment and the other included a 190 kb fragment of the *Streptomyces* sp. QL37 genome. Both were introduced into *S. coelicolor* M1152. This yielded the strains M1152::*lug*11O and M1152::*lug*3G, respectively. Lugdunomycin was previously isolated from cultures grown on minimal medium (MM) supplemented with 0.5% mannitol and 1% glycerol; pratensilin A (**7**) and the limamycins (**8**, **9**) were isolated from cultures on R5 medium supplemented with 1% mannitol and 0.8% peptone, and angucyclines from both media (Wu *et al.*, 2019). Therefore, the wild-type strain, *S. coelicolor* M1152, strains M1152::*lug*11O and M1152::*lug*3G were grown on both MM and R5 agar plates for seven days at 30 °C.

S. coelicolor M1152::*lug* ex-conjugants produced a brown-yellow pigment on MM and R5 agar plates, whereas *S. coelicolor* M1152 without the cluster did not produce this pigment (Figure 2A). The brown-yellow pigment production by the ex-conjugants containing the *lug* gene cluster suggests angucycline production. Therefore, extracts were prepared from these cultures, followed by LC-MS analysis, where angucyclines production was confirmed in both media (Figure 2B and Figure S1) On R5 agar, angucyclines (**2**, **3**, **4**, **6**), pratensilin A (**7**) and limamycins (**8**,**9**)

were identified in the heterologous host. This indicates that the introduced clone contained all the genes required for angucycline, pratensilin A and limamycin production. Angucyclines were also produced in MM-grown cultures, while lugdunomycin could not be detected under either growth condition. A peak was observed with a monoisotopic mass of 456.14, corresponding to the $[M+H-H_2O]^+$ ion for lugdunomycin, and with the same retention time. However, its MS/MS spectrum did not match that of lugdunomycin, and the mass of the $[M+H]^+$ ion could not be simultaneously observed (data not shown). We therefore conclude that the *lug* gene cluster alone is not sufficient to enable lugdunomycin production by *S. coelicolor* M1152.

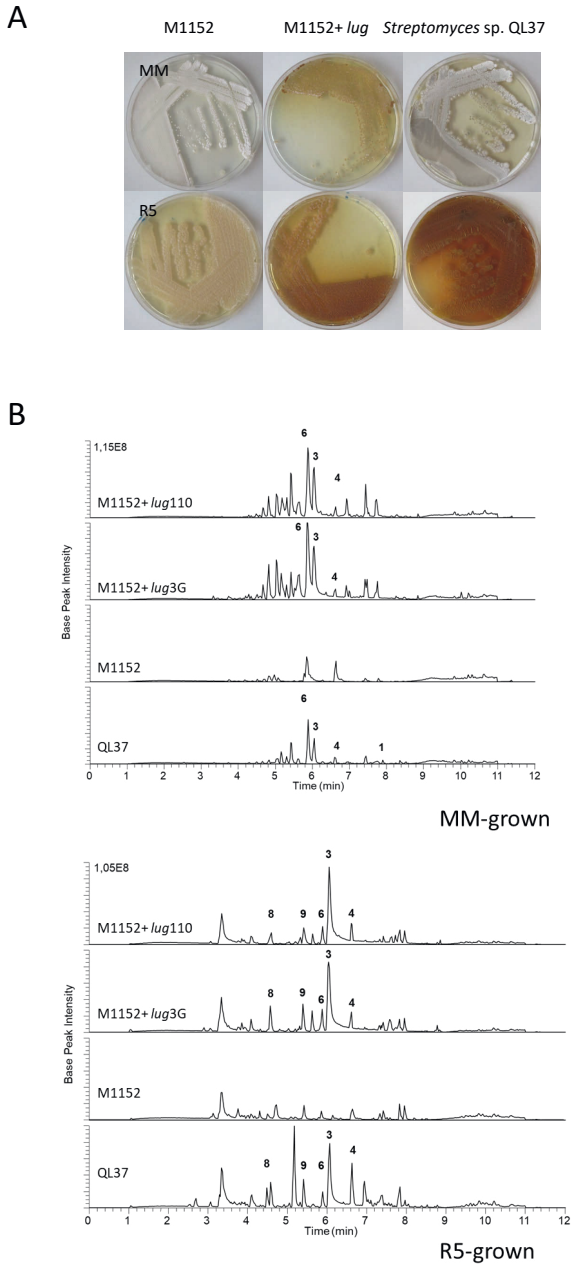


Figure 2 Heterologous expression of the *lug* gene cluster in *S. coelicolor* M1152.
A) *S. coelicolor* M1152 (M1152), *S. coelicolor* M1152::*lug*11O (M1152+*lug*11O), *S. coelicolor* M1152::*lug*3G (M1152+*lug*3G), and wild-type *Streptomyces* sp. QL37 (QL37) grown on MM (top) and R5 (bottom) agar plates. B) LC-MS chromatograms of the crude extracts of the cultures shown in A. Compound numbers are given over their corresponding peaks (Figure 1) As expected M1152 itself did not produce any of the angucycline derived

compounds. *Streptomyces* sp. QL37 produced the non-rearranged angucyclines (**2**,**3**,**4**,**6**), the rearranged angucyclines (**7**,**8**,**9**) and lugdunomycin (**1**). The heterologous expression strains M1152+*lug*11O and M1152+*lug*3G produced the (non)-rearranged angucyclines (**2**–**4**, **6**, **7**–**9**), but did not produce lugdunomycin (**1**) (see also Figure S1).

Metabolome analysis for *Streptomyces* sp. QL37 and its *lug-pks* mutant

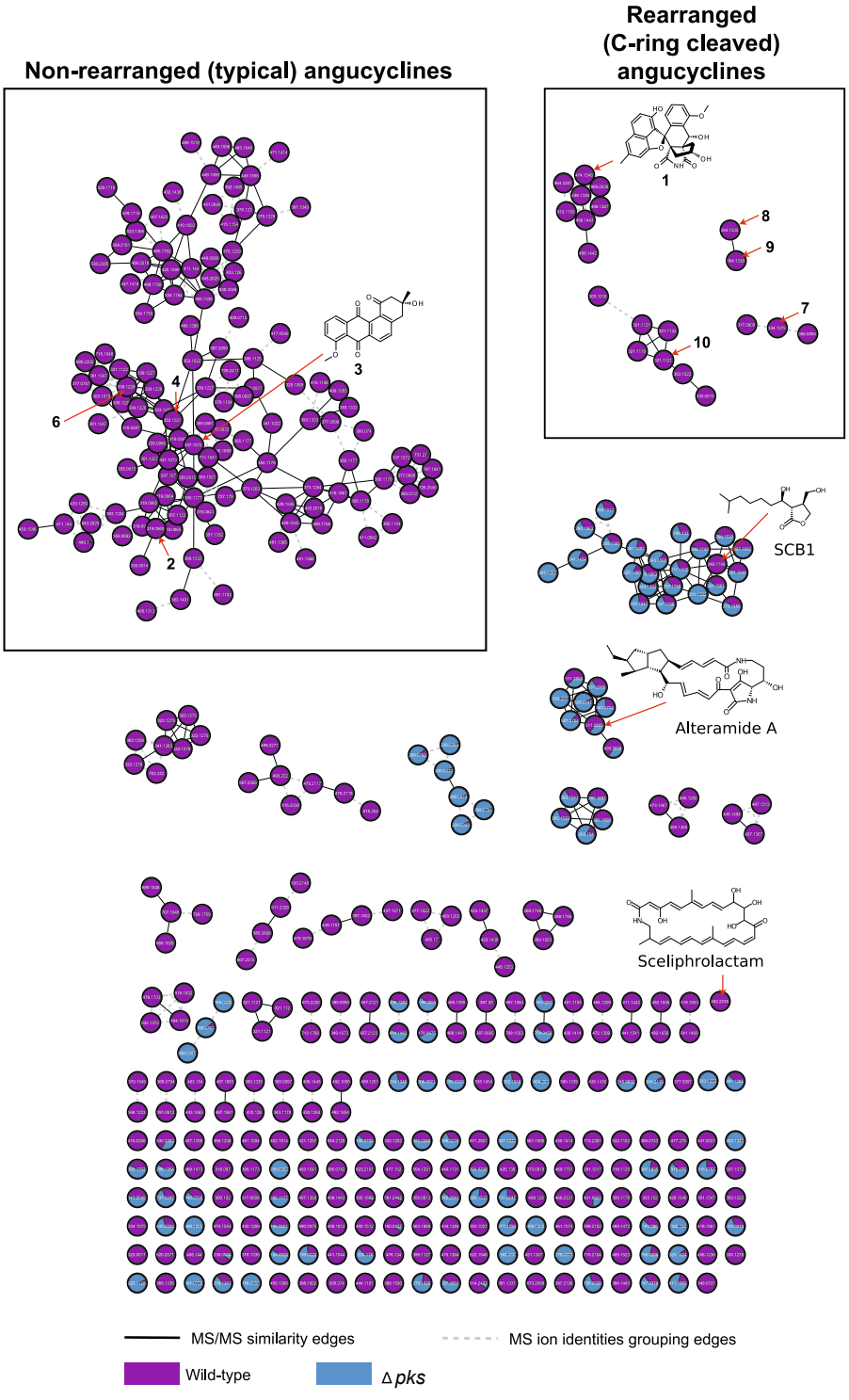
We then wanted to see which metabolites are produced in a *lug-pks* null mutant, which cannot produce angucycline-related molecules (Chapter 3). Molecular networking was employed to analyse the LC-MS data generated from the extracts of the wild-type strain and its *lug-pks* mutant cultures grown for seven days on MM agar plates (Figure 3) This technique enables clustering of structurally related metabolites based on similarities in their MS/MS spectra (Nothias *et al.*, 2020, Schmid, 2020). A molecular network was generated using the Global Natural Products Social Molecular Networking (GNPS) platform, and was visualised using Cytoscape (Kohl *et al.*, 2011, Wang *et al.*, 2016, Nothias *et al.*, 2020, Schmid, 2020). In the wild-type strain, angucyclines (**2**–**4**,**6**), lugdunomycin, pratensilin A (**7**), limamycins (**8**,**9**) and lugdunomycin (**1**) were identified, which clustered as separate molecular families (Figure 3) Additionally, a structurally related isomer of lugdunomycin, with the same mass but different retention time, could be identified for the first time (Figure S2). Possibly *iso*-maleimycin can react in different manners with the angucycline diene, leading to a variety of lugdunomycin isomers (M. Uiterweerd and A.J. Minnaard, unpublished data). We also identified elmonin or its enantiomer oleaceran, a rearranged angucycline, through matching one of the nodes with the chemically synthesised elmonin (**10**) (Figure S3). This molecule is a C-ring cleaved angucycline, which also clustered separately from the non-rearranged angucyclines (Yixizhuoma *et al.*, 2017, Raju *et al.*, 2013). Elmonin may be the direct precursor of the diene that reacts with *iso*-maleimycin in the final Diels-Alder reaction. All the ions in these molecular families were exclusively present in the wild-type strain.

Besides angucycline-related molecules, we detected a molecular family that is related to γ-butyrolactones (GBLs) and one that is related to tetramate macrolactams. The annotation of the latter two was based on manual comparison of measured MS and MS/MS spectra with those previously reported for SCB1 and alteramide A respectively (Figure S4 and Figure S5) (Yang *et al.*, 2005, Shaikh *et al.*, 2021, Moree *et al.*, 2014). Additionally, the genome of *Streptomyces* sp. QL37 contains three predicted BGCs for GBLs (BGC16, -23 and -33) and a predicted

tetramate macrolactams one (BGC6), in agreement with the annotation. The monoisotopic mass of sceliphrolactam was also detected (Figure S6), supported by the presence of a sceliphrolactam BGC (BGC9). For details on the predicted BGCs in *Streptomyces* sp. QL37 I refer to Chapter 3. All the non-angucycline metabolite families were also identified in the *lug-pks* null mutant, with a general tendency of being upregulated in the mutant. This could be due to the enhanced availability of precursors for the other biosynthetic pathways given that the highly abundant angucyclines are not produced anymore.

The production of lugdunomycin was proposed to proceed through a Diels-Alder [4+2] cycloaddition involving a dienophile which is called *iso*-maleimycin (Wu *et al.*, 2019, Uiterweerd, 2020). It was not possible to detect the *iso*-maleimycin moiety using LC-MS, however it could be readily detected using GC-MS (Uiterweerd, 2020). Accordingly, the extracts of *Streptomyces* sp. QL37 wild-type strain and its *lug-pks* mutant were additionally analysed using GC-MS. Interestingly, *iso*-maleimycin could be detected in the extracts of both the wild-type strain (Uiterweerd, 2020) and its *pks* mutant (Figure 4), suggesting that it is not derived from the *lug* gene cluster. This contrasts with the previous hypothesis that *iso*-maleimycin may be derived from limamycins (Wu *et al.*, 2019).

Figure 3 ***(right)*** Molecular network of the ions detected in the extracts of *Streptomyces* sp. QL37 and its *lug*-pks mutant when grown on MM.
The nodes are labelled by the precursor mass of their ions and pie charts are mapped to the nodes to indicate the relative intensities of the ions in the different samples. The dashed edges in the network connect different ion species of the same molecule and the solid edges connect nodes based on their MS/MS similarity. Arrows highlight metabolites previously identified in *Streptomyces* sp. QL37 (Figure 1).


Non-rearranged (typical) angucyclines
Rearranged
(C-ring cleaved)
angucyclines
1
2
3
4
6
7
8
9
10
SCB1
Alteramide A
Sceliphrolactam
MS/MS similarity edges
MS ion identities grouping edges
Wild-type
Δpks

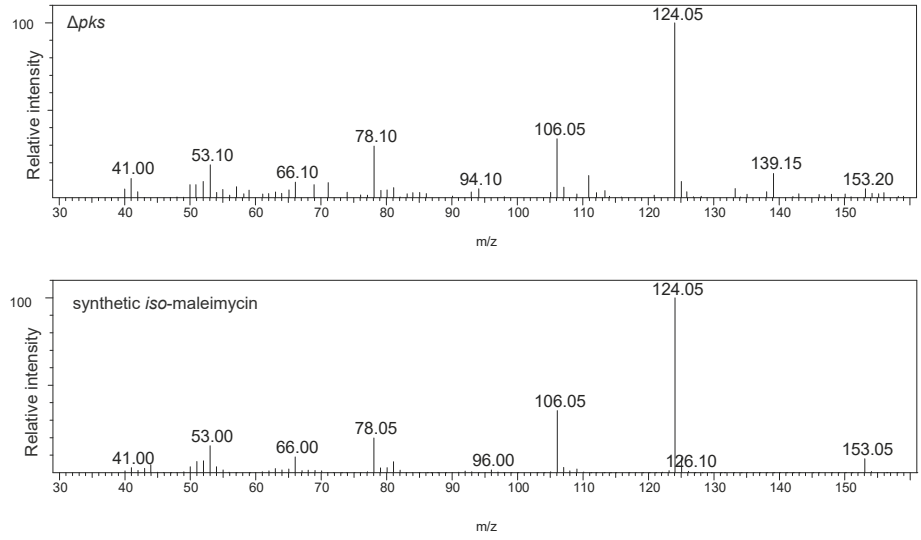


Figure 4 Comparison of the GC-MS spectrum of synthetic iso-maleimycin and the corresponding peak in the extract of the *lug*-pks mutant of *Streptomyces* sp. QL37.
The mutant was grown for seven days on MM at 30°C. Extracts were prepared using ethyl-acetate.

Transcriptional analysis of the BGCs of *Streptomyces* QL37

To study the expression of the different BGCs of *Streptomyces* sp. QL37, and their temporal correlation, RNA-sequencing (RNA-seq) was performed on RNA isolated from *Streptomyces* sp. QL37 grown for different incubation times (24 h, 36 h, 48 h and 60 h) on MM at 30°C. (Schorn *et al.*, 2021). For more details on the RNA-seq data see Chapter 3. The genome of *Streptomyces* sp. QL37 was analysed using antiSMASH, which predicted the presence of in total 35 BGCs. Four hybrid BGCs were recovered, including BGC 6 (Non-ribosomal peptide synthetase (NRPS), T1PKS), BGC9 (hlgE-KS and T1PKS), BGC20 (NRPS, β-lactone and other), BGC23 (β-lactone and γ-butyrolactone) and 32 (T2PKS and terpene). For details on the antiSMASH outcome, see Chapter 3, Figure S1. To define the expression level of each BGC at each time point, the average expression level of its core biosynthetic genes was calculated, such as NRPS, ketosynthase or terpene cyclase genes. Table S1A and S1B show the genes that were assigned as core genes. For some of the hybrid BGCs (BGC 9, BGC 23 and BGC 32) the different BGC types were clearly separated over two fragments of the hybrid BGC, suggesting they could each have a different expression pattern. Therefore, the average expression level

was calculated for each BGC type in these hybrid BGCs (with each type designated as region "a" an "b", Table S1A and B and Figure S7).

A heatmap with hierarchical clustering analysis using Ward's method was generated to analyse the similarities between the expression of each BGC based on the expression level and the expression pattern over time (Figure 5). A corresponding bar plot with error bars is indicated in Figure S8. BGCs that showed a similar expression pattern and level as the *lug* gene cluster (BGC12) were BGC32a (for a spore pigment), BGC6 (for a polycyclic tetramate macrolactam), and BGC23a (for a β-lactone). BGC32a is most likely responsible for the production of the black spore pigment. BGC6 produces polycyclic tetramate macrolactams and these were abundant in the extracts of the wild-type strain and its *pks* mutant (Figure 3).

Other BGCs that were highly expressed in the transcriptome of *Streptomyces* sp. QL37 were BGC7 (for a terpene), BGC9b (for sceliphrolactam), BGC21 (for a nonribosomal peptide), and BGC32b (for a terpene). BGC9b is predicted to produce sceliphrolactam, and was detected in the extracts of the wild-type strain. BGC21 is a likely candidate for the biosynthesis of a novel lipopeptide that was identified in the extracts of the *lugOI* mutant (Chapter 5). The GBLs that were identified in the crude extracts might be derived from BGC16 and/or BGC23b, both of which are predicted GBL BGCs and showed similar expression levels.

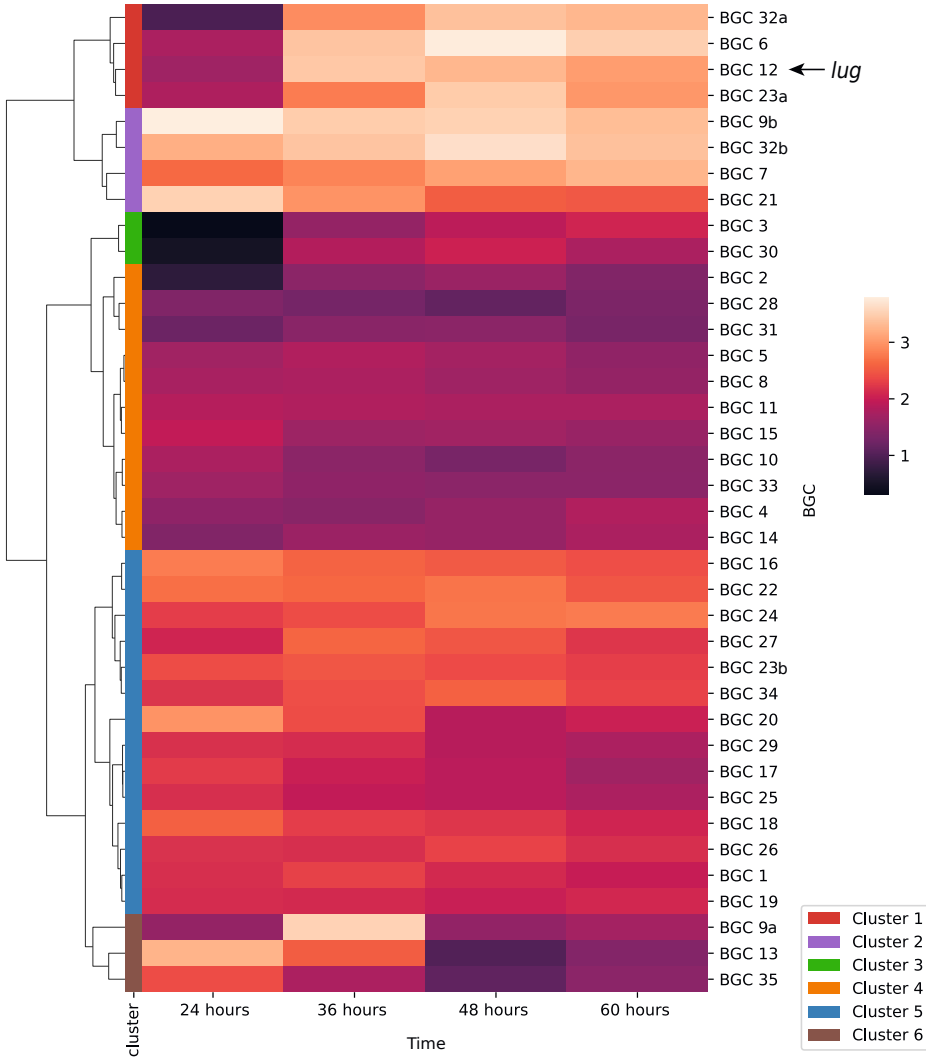


Figure 5 Hierarchical clustering analysis and heatmap of the average expression level of the core genes of the BGCs in the genome of *Streptomyces* sp. QL37.
Heatmap showing the log10 of the absolute expression levels of the core genes in each BGC predicted by antiSMASH at the different measured time points (T_{24h}, T_{36h}, T_{48h} and T_{60h}). The lug gene cluster (BGC12) is relatively highly expressed together with the *iso*-maleimycin gene cluster (BGC23a), the alteramide A BGC (BGC6) and the spore pigment BGC (BGC32a) (Refer to Table S1 for the core genes used in the calculation, and Figure S7 for the annotation of regions “a” and “b”).

Possible role of BGC 23 in lugdunomycin biosynthesis

One of the BGCs co-expressed with the *lug* gene cluster was BGC 23a (Figure 5). This BGC is annotated as a β-lactone BGC by antiSMASH, due to the presence of a gene encoding a 2-isopropylmalate synthase and a gene encoding an AMP-ligase, similar as to the β-lactone BGCs for belactosin and cystargolide ((Wolf *et al.*, 2017a) and Serina Robinson, pers. comm.). In addition, BGC 23 encodes an orthologue of LysW, a carrier protein required for the synthesis of lysine (Matsuda *et al.*, 2017). Recently it was shown that these proteins are also required for the biosynthesis of non-proteogenic building blocks of natural products, such as for vazabitide A (Hasebe *et al.*, 2016). One member of this family of BGCs is required for the biosynthesis of maleimycin ((Makato, 2012-2017, Prima *et al.*, 2017)), produced by *S. showdoensis* ATCC 15227 (Elstner *et al.*, 1973) . BGC23 is highly similar to this likely maleimycin BGC (Figure 6, Table S3). This is highly suggestive, considering that the closely related *iso*-maleimycin was proposed to be involved in lugdunomycin biosynthesis (Uiterweerd, 2020, Wu *et al.*, 2019). Preliminary mutational analysis revealed that neither lugdunomycin nor *iso*-maleimycin could be detected in mutants of BGC23a, whereas the angucyclines, pratensilin A, elmonin and the limamycins were still produced (I. Nuñez-Santiago, S.S. Elsayed and G. P. van Wezel, unpublished data). Taken together, our data show that the *lug* gene cluster is responsible for the biosynthesis of all angucyclines and derivatives as well as limamycins, which was further corroborated by heterologous gene expression. However, the final step for lugdunomycin biosynthesis likely requires a second BGC.

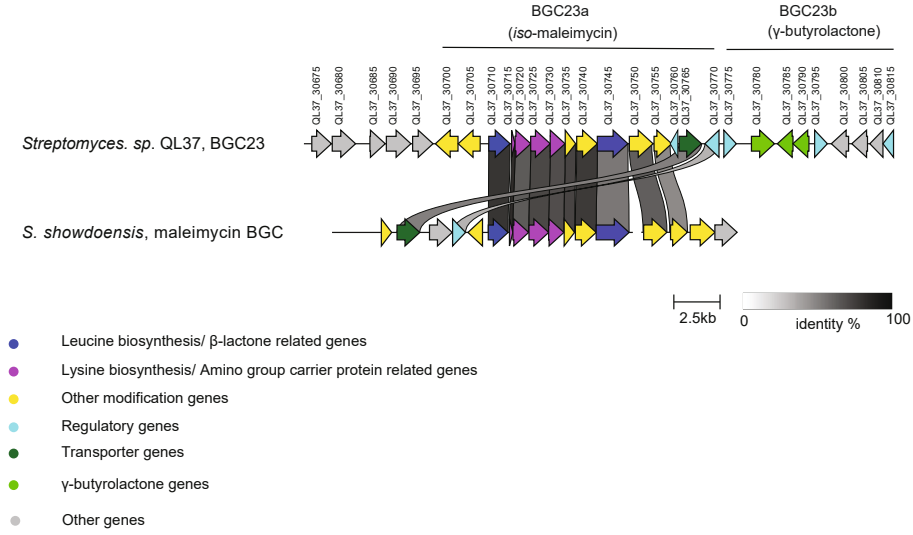


Figure 6 Comparison of BGC23 from *Streptomyces* sp. QL37 with the likely maleimycin BGC cluster in the genome of the maleimycin producer, S. *showdoensis* ATCC 15227. The strokes between the BGCs link genes that share at least 40% identity. Refer to Table S2 for information about the annotation of each gene. The orthologues of the genes from QL37-30750 and QL37-30755 of *Streptomyces showdoensis* were detected on a different scaffold.

DISCUSSION

Streptomyces sp. QL37 produces the highly rearranged angucycline derivative lugdunomycin (Wu *et al.*, 2019). A biosynthesis pathway was proposed wherein a C-ring rearranged angucycline acts as a diene and reacts with *iso*-maleimycin as a dienophile in a Diels-Alder [4+2] cycloaddition to form lugdunomycin. (Uiterweerd, 2020, Wu *et al.*, 2019). We previously hypothesised that *iso*-maleimycin is derived from the limamycins (Wu *et al.*, 2019). Our current study shows that in fact the proposed *lug* gene cluster is an angucycline BGC that only specifies angucycline, pratensilin A, limamycin production, but is not sufficient for lugdunomycin production. An additional BGC is likely required for the biosynthesis of *iso*-maleimycin.

Heterologous expression of the *lug* gene cluster and flanking regions in *S. coelicolor* M1152 led to the biosynthesis of angucyclines and their C-ring rearranged derivatives (pratensilin A and the limamycins), while no lugdunomycin was detected. Therefore, we thoroughly examined the metabolome of *Streptomyces*

sp. QL37 using molecular networking. To do so, we compared the metabolome of the wild type and its *pks* mutant, which does not produce any angucyclines, to deviate between angucycline-related and non-angucycline related molecules (Chapter 3). Besides angucycline-related molecules, *Streptomyces* sp. QL37 also produces alteramide A, sceliphrolactam and γ-butyrolactones (GBLs), which were mainly produced by the *pks* mutant. Interestingly, we found that *iso*-maleimycin is produced by a *lug-pks* mutant that lacks the minimal PKS genes of the lugdunomycin BGC, suggesting the molecule is not derived from the limamycins as previously anticipated, and perhaps from another BGC.

Temporal RNA-seq on the transcriptome of *Streptomyces* sp. QL37 under conditions in which lugdunomycin is produced revealed BGCs that were co-expressed with the *lug* gene cluster, including BGCs for alteramide A, sceliphrolactam and GBLs, in line with the metabolomics data. In addition, BGC23a showed very good co-expression with the *lug* gene cluster; this BGC contains amongst others β-lactone related genes, and AmCP-related genes. Additionally, BGC23 is very similar to a BGC for maleimycin biosynthesis in *S. showdoensis* (Makato, 2012-2017, Wolf *et al.*, 2017a, Prima *et al.*, 2017). This data, together with preliminary mutational studies (I. Nuñez-Santiago, S.S. Elsayed and G. P. van Wezel, unpublished data), makes it highly likely that BGC12 (for lugdunomycin) and BGC23 (for *iso*-maleimycin) are both required to produce lugdunomycin. It is important to note that *S. coelicolor* lacks a cluster homologues to BGC23, and this likely explains why the heterologous expression of the *lug* gene cluster in *S. coelicolor* M1152 did not result in lugdunomycin biosynthesis.

We propose that the final Diels-Alder reaction between *iso*-maleimycin and the C-ring rearranged angucyclines is spontaneous, based on the racemic mixture that is produced (Wu *et al.*, 2019). If our assumptions are correct, introduction of the *lug* gene cluster in an *iso*-maleimycin producer (or vice versa) could facilitate heterologous lugdunomycin production. This is currently being investigate in our laboratory.

We here argue that BGC12 and BGC23a collaborate to produce lugdunomycin, via the joining of metabolites from the angucycline and *iso*-maleimycin pathways. Such joining of two pathways is an exciting new theme, and is something that requires more attention. Recent examples discovered in our laboratory are

the production of endophenasides in *Kitasatospora* sp. MBT66, whereby the glycosylation of phenazines by a methyl-rhamnose moiety is encoded by a BGC for the macrolide leucanicidin, and the biosynthesis of actinomycin L in *Streptomyces* sp. MBT27, whereby an anthranilamide unit, derived from primary metabolism, is incorporated into the actinomycin core (Wu *et al.*, 2016b, Machushynets *et al.*, 2022). As shown in the current work, time- and cost-intensive synthetic biology or heterologous expression approaches on a single BGC will not be successful if the final product is in fact governed by more than one BGC, so this is a concept that needs more attention in the field of microbial drug discovery.

A question that remains to be answered is what the environmental signals are that activate lugdunomycin biosynthesis and how the expression of the BGCs of *Streptomyces* sp. QL37 is coordinated. Perhaps more importantly, it will be important to see whether lugdunomycin offers an ecological advantage to the producer and if so, under which circumstances, and what its bioactivity is. Answers to these questions will shed further light on the biosynthesis, the biological role, and the possible application of this intriguing molecule.

Acknowledgments

We thank Serina Robinson and Marnix Medema for discussions on *iso*-maleimycin biosynthesis. We thank Du Chao for his assistance in creating Figure 5. The work was supported by NACTAR grant 16439 from the Netherlands Organization for Scientific research (NWO).

MATERIALS AND METHODS

Bacterial strains and growth conditions

Streptomyces sp. QL37 was obtained from the MBT strain collection (Zhu *et al.*, 2014b). The strain was isolated from soil in the Qinling mountains (P. R. China) as described previously and was deposited to the collection of the Centraal Bureau voor Schimmelcultures (CBS) in Utrecht, the Netherlands, under deposit number 138593 (Wu *et al.*, 2019). *Streptomyces coelicolor* M1152 was obtained from the John Innes Centre. *S. coelicolor* M1152 lacks genes that are required for prodigiosin, coelimycin P1, calcium dependent antibiotic, and actinorhodin biosynthesis. In additions it carries mutations in the gene *rpoB* for the RNA polymerase β subunit,

which often promotes antibiotic production (Hu *et al.*, 2002, Gomez-Escribano & Bibb, 2011). Strains used in this study are described in Table 1.

Streptomyces were grown on soya flour mannitol (SFM) for seven days at 30 °C, then spores were collected and stored in 20% glycerol at -20 °C (Kieser *et al.*, 2000). For conjugation *Streptomyces* sp. QL37 was grown on SFM containing 60 mM $MgCl_2$ and 60 mM $CaCl_2$ (Wang & Jin, 2014). Ex-conjugants were selected using thiostrepton (20 µg/ml) and apramycin (50 µg/ml) (Wu *et al.*, 2019).

Heterologous expression of the *lug* gene cluster in *Streptomyces coelicolor* M1152

The pESAC13-A library was prepared by the company BioS&T (Montreal, Canada) as previously described (Jones *et al.*, 2013). For the preparation of the genomic library, genomics DNA of *Streptomyces* sp. QL37 was isolated and partially digested with BamHI. These fragments were cloned into the BamHI-digested and dephosphorylated pESAC13-A. This vector is a Phage P1 *E.coli-Streptomyces* Artificial Chromosome containing an *oriT* for conjugation, a phiC31 integrase and an *attP* site for integration into the *attB* recombination locus of the *Streptomyces* genome (Sosio *et al.*, 2000, Jones *et al.*, 2013). This vector is derived from the *Streptomyces* pPAC-S1 (Sosio *et al.*, 2000). The QL37 library contained 3840 clones with an average insert size of 112 kb. PCR primers (1-6) were developed for screening of the PAC library for positive clones containing the *lug* gene cluster (Table 2). PCR screening identified two clones that contain the *lug* cluster (11O of around 120 kb and 3G of around 190 kb).

DH10β cells containing the PAC clones were used to introduce the *lug* cluster into the heterologous strain *S. coelicolor* M1152. First the *lug* cluster was introduced in the methylase deficient strain ET12567 by tri-parental mating including the strains *E. coli* ET12567, ET12567 containing the driver plasmid pUB307 and DH10β containing *lug* (PAC11O and PAC3G) according to A.C. Jones and colleagues (Jones *et al.*, 2013). ET12567 transformants containing both the PAC construct and the driver plasmid pUB307 were selected on LB agar containing kanamycin (50 µg/ml), chloramphenicol (25 µg/ml) and apramycin (50 µg/ml). Colonies was selected and glycerol stocks were prepared of these strains. The three PCR primer pairs (1-6) were used to confirm that the strains contained the vector. These fragments

Table 1 Bacterial strains and plasmids used in this study

Strains		
Name	Description	Reference
***Streptomyces* sp. QL37**	Wild type, CBS 138593	(Wu *et al*., 2019), (Zhu *et al*., 2014b)
Streptomyces* sp. QL37Δ *lug-pks	Deletion mutant of the *lugA–OII* genes (for the minimal PKS), Apra^R.	(Wu *et al*., 2019), Chapter 3
***Streptomyces coelicolor* M1152**	*Streptomyces coelicolor* M145 Δ*act* Δ*red* Δ*cpk* Δ*cda rpoB*(C1298T), conferring resistance to rifampicin	(Gomez-Escribano & Bibb, 2011)
***S. coelicolor* M1152::lug11O**	*S. coelicolor* M1152 containing fragment of 120 kb from genomic DNA of *Streptomyces* sp. QL37 including the *lug* gene cluster; Apra^R, Thio^R.	This study
***S. coelicolor* M1152::lug3G**	*S. coelicolor* M1152 containing fragment of 190 kb from genomic DNA of *Streptomyces* sp. QL37 including the *lug* gene cluster; Apra^R, Thio^R	This study
***E. coli* ET12567**	Methylation deficient host (*dam*-13::Tn9 *dcm*-6 *hsdM hsdS hsdR* Cm^R).	(MacNeil *et al*., 1992)
***E. coli* ET12567/ pUB307**	*E. coli* ET12567 harbouring pUB307.The strain is used for conjugal transfer. It is a derivative of RP1, and contains *oriT*; Cm^R, Tet^R, Km^R.	(Jones *et al*., 2013)
Plasmids		
Name	Description	Reference
pESAC13-A	Used for the creation of *Streptomyces* genomic library by BioS&T. The vector is derived from pPAC-S1 and contains the *oriT* from RK2 for conjugative transfer to *E. coli* and the *phiC31 attP* sites for integration into the *Streptomyces* genome; Thio^{R}and Apra^R.	[34, 43]

were sequenced to verify that the PCR products correspond to the fragments of the *lug* gene cluster.

Conjugation of the PAC constructs was executed according to the protocol of T. Kieser and colleagues(Kieser *et al.*, 2000). After mixing *E. coli* and *Streptomyces* the strains were spread on an SFM plate supplemented with 10 mM $MgCl_2$ and incubated at 30 °C. After 20 hours the plates were overlayed with nalidixic acid (10 µg/ml) and apramycin (50 µg/ml) and incubated for another 6 days. Ex-conjugants were selected streaked onto new plates to remove *E. coli* and to obtain pure isolates. To obtain spore suspensions the strains were grown on SFM medium containing the selection marker apramycin. The mycelium was scraped off the plate and re-suspended into 20% glycerol. The integration was verified using the primers 1-6 (Table 2).

Table 2 Primers used in this study

Number	Name	Sequence 5' --> 3'
1	StartLug+1_Fw	ACGGACAACGCCTTCTCCGCTGAAC
2	StartLug+336_Rv	CTGCCTACGCCGACGTTGTCACCG
3	MiddleLug+1_Fw	CCGACCGTACCGAAGGCATGATCTC
4	MiddleLug+477_Rv	CGGTAGCCCTCGCAGTGGGTGATG
5	EndLug+1 _Fw2	TATCCGGGTCAAGCGGATCAAGCTG
6	EndLug+425_Rv2	TGGCGGCGTGTGCCTTTCGGAGAG

Natural products extraction

Streptomyces strains were confluently grown on minimal medium (MM) agar plates (25 ml) containing 0.5% mannitol + 1% glycerol and on R5 containing 1% mannitol and 0.8% peptone (Wu *et al.*, 2019). The same agars as described in Chapter 3 were used for the preparation of the media. The experiment was conducted in triplicate. 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 12 hours. Subsequently the ethyl acetate was decanted and evaporated at room temperature. This process was repeated twice. The dried extract was re-dissolved in methanol (MeOH) and centrifugated to remove any undissolved material. Subsequently the MeOH solutions were transferred to new pre-weighed glass vials, and the solvent dried under ambient conditions (heterologous expression samples) or under nitrogen

(GNPS networking). The crude extracts were weighed and dissolved in methanol to a fixed concentration for LC-MS analysis. The prepared solutions were centrifuged again for 20 min at 4 °C to remove any suspended matters. For the detection of *iso*-maleimycin using GC-MS in the metabolome of the *pks* mutant, extracts were prepared in a similar manner as for the detection of angucyclines, limamycins and lugdunomycin (Uiterweerd, 2020). However, extracts were not re-dissolved in methanol, because *iso*-maleimycin degrades in this solvent.

LC-MS methods

Extracts from S. coelicolor ***(heterologous gene expression)***
LC-DAD-HRESIMS spectra were obtained using a Waters Acquity UPLC system, equipped with Waters Acquity PDA, and coupled to a Thermo Instruments MS system (LTQ XL/LTQ Orbitrap XL). The UPLC system was run using Acquity UPLC HSS T3 C_{18} column (1.8 µm, 100 Å, 2.1 × 100 mm). Solvent A was 0.1% formic acid, 95% H_2O and 5% ACN. Solvent B was 0.1% formic acid, 95% ACN and 5% H_2O. The gradient used was 2% B for 0.5 min, 2-40% for 5.5 min, 40-100% for 2 min, and 100% for 3 min. The flow rate used was 0.5 ml/min. The MS conditions used were: capillary voltage 5 V, capillary temperature 300 °C, auxiliary gas flow rate 5 arbitrary units, sheath gas flow rate 50 arbitrary units, spray voltage 3.5 kV, mass range 100-2000 amu, FT resolution 30000. Spectra were analysed using Thermo Scientific Xcalibur.

Extracts for molecular networking
LC-MS/MS acquisition was performed using Shimadzu Nexera X2 UHPLC system, with attached PDA, coupled to Shimadzu 9030 QTOF mass spectrometer, equipped with a standard ESI source unit, in which a calibrant delivery system (CDS) is installed. The dry extracts were dissolved in MeOH to a final concentration of 1 mg/ml or 0.5 mg/ml, and 2 µL were injected into a Waters Acquity HSS C_{18} column (1.8 µm, 100 Å, 2.1 × 100 mm). The column was maintained at 30 °C, and run at a flow rate of 0.5 ml/min, using 0.1% formic acid in H_2O as solvent A, and 0.1% formic acid in acetonitrile as solvent B. A gradient was employed for chromatographic separation starting at 5% B for 1 min, then 5–85% B for 9 min, 85–100% B for 1 min, and finally held at 100% B for 3 min. The column was re-equilibrated to 5% B for 3 min before the next run was started.

All samples were analysed in positive polarity, using data dependent acquisition mode. In this regard, full scan MS spectra (*m/z* 100–1700, scan rate 10 Hz, ID enabled) were followed by two data dependent MS/MS spectra (*m/z* 100–1700, scan rate 10 Hz, ID disabled) for the two most intense ions per scan. The ions were selected when they reach an intensity threshold of 1500, isolated at the tuning file Q1 resolution, fragmented using collision induced dissociation (CID) with fixed collision energy (CE 20 eV), and excluded for 1 s before being re-selected for fragmentation. The parameters used for the ESI source were: interface voltage 4 kV, interface temperature 300 °C, nebulizing gas flow 3 L/min, and drying gas flow 10 L/min.

GNPS molecular networking

For molecular networking using GNPS, the Feature Based Molecular Networking (FBMN) platform was used in combination with ion identity networking (IIMN), where the data were first processed using MZmine version 2.53 and subsequently further processed in MZmine version 2.37.1.corr17.7 (Nothias *et al.*, 2020, Schmid, 2020). LC-MS data were processed similarly as described by D. A. van Bergeijk and colleagues (van Bergeijk *et al.*, 2022).

Pre-processing LC-MS data in MZmine 2.53 for feature -based molecular networking

Raw data obtained from LC-MS analysis were converted to mzXML centroid files using Shimadzu LabSolutions Postrun analysis. The files were imported into MZmine 2.53 for data processing (Pluskal *et al.*, 2010). Unless stated otherwise, m/z tolerance was set to 0.002 m/z or 10.0 ppm, RT tolerance was set to 0.05, noise level was set to 2.0E2 and the minimum absolute intensity to 4.0E2. Raw data were cropped to RT 0.5-10.0min. Mass ion peaks were detected (positive polarity, mass detector: centroid). Additionally mass ion peaks at an MS level 2 were detected using the centroid algorithm with noise level of 0. Their chromatograms were built using ADAP chromatogram builder (minimum group size of scans: 10, group intensity threshold: 2.0E2). To enhance the deconvolution step, the chromatograms were smoothed (filter width: 9), and subsequently deconvoluted (algorithm: local minimum search; chromatographic threshold: 90%; search minimum in RT range: 0.05; minimum relative height: 1%, minimum ratio of peak top/edge: 2, peak duration range: 0.03–3.00) For chromatogram deconvolution, additional MS2 scan pairing was performed with *m/z* range of 0.02 Da. The detected peaks were

deisotoped (monoisotopic shape; maximum charge: 2; representative isotope: most intense). The peak lists of each extract were aligned (weight for RT= weight for m/z =20); compare isotopic pattern with a minimum score of 50%). Missing peaks that were detected in at least one of the samples were filled with gap-filling algorithm.

Identification of complexes, adducts, and fragments in the aligned peak list was done using the identification options complex search (Ionization method: $[M+H]^+$, Maximum complex peak height: 50%), Adduct Search (Adducts: $[M+Na]^+$, $[M+K]^+$ and $[M+NH_3]^+$; Maximum relative adduct peak height: 3000%), and fragment search (Maximum fragment peak height: 50%, Minimum MS2 peak height: 0.0E0). After identification of the mass features and filtering to remove duplicate peak row and removal of mass features due to detected ringing (*m/z* tolerance: 1.0 *m/z* or 1000.0 ppm), only the mass features with associated MS/MS spectra were selected with peak list rows filter.

Pre-processing LC-MS data in MZmine 2.37.1 corr17.7 for ion identity networking
The data were further processed in in MZmine version 2.37.1.corr17.7 to detect the ion species and to generate an ion identity family (Schmid, 2020). MetaCorrelate module was used, an algorithm that groups ions derived from the same molecule, based on the retention time and the peak shape (RT tolerance 0.1 min, min height 4.0E2, noise level 2.0E2, min sample filter (default setting) min samples in all: abs=1, rel= 0.0%; min samples in group: abs= 0; rel = 0.0%, Min % intensity overlap: 60%, exclude estimated features (gap-filled)), correlation grouping (min data points: 5; min data points on edge: 2; measure: Pearson; min feature shape correlation: 5), Feature height correlation ((default settings) min data points: 2; measure: Pearson; min Correlation: 70%)). Next, ion identity networking was applied to annotate the different ion adducts (check all features, min height 4.0E2, adducts were selected in the ion library ($[M+H]^+$, $[M+Na]^+$, $[M+K]^+$ and $[M+NH_4]^+$ and modifications were selected ($[M-H_2O]$, $[M-2H_2O]$). The MS mode was set on positive; the maximum charge was 2 and the maximum molecules/clusters on 3. The data were subsequently exported for GNPS (select spectra to merge across samples; *m/z* merge mode: most intensive; intensity merge mode: sum intensities; cosine threshold is 70%, peak count threshold 20% and isolation window offset *m/z* 0.00; isolation window width *m/z* 3.0.

Data filtering

The mass feature was kept when its peak aera was higher than 3000 in two out of three replicates of each group. All other mass features were considered as noise and removed from the dataset. Afterwards the mass features that were also present in the media blanks were removed by a filtering step: the mean peak area of a mass feature in each group should be higher than 50 times the peak area of the corresponding mass feature in the blank.

Generation feature-based ion identity molecular network

The data were submitted to the GNPS web tool and a network was generated using feature networking. To generate the network the data were filtered, where MS/MS fragment ions within +/-17 Da of the precursor *m/z* were removed. Only the top 6 fragment ions in the +/- 50 Da were selected for window filtering of the MS/MS spectra. Both the precursor ion mass tolerance and the MS/MS ion mass tolerance were set to 0.02 Da. A molecular network was generated where the edges were filtered to have a cosine score above 0.7 and more than 7 matched peaks. Spectral library search was performed using the default settings. The molecular network was visualised using the software Cytoscape (Kohl *et al.*, 2011).

GC-MS analysis for the identification of *iso*-maleimycin

GC-MS analysis was executed as previously described (Uiterweerd, 2020). For analysis and identification, synthetic samples of *iso*-maleimycin (0.23 mg/ml), and a sample of the extract (1 mg/ml) were injected on a Shimadzu GC-2010 gas chromatograph coupled to a Shimadzu GC-MS-QP2010 gas chromatograph mass spectrometer. For the synthetic samples, split ratio 20.0, for the extract, split ratio 10.0. The injection temperature was 200 °C, detector temperature 250 °C. Separation was carried out on an HP-1 column, 100 % dimethylsiloxane, (30 m L × 0.25 mm Ø × 0.25 µm thickness). The temperature program was set as follows: 150 °C hold for 5 min, then increase 150 °C to 200 °C with 20 °C/min, finally 3 minutes hold at 200 °C. The carrier gas was He, with a column flow of 0.9 ml/min, pressure 82.7 kPa. Ionization by means of EI, ionization energy 70 eV, mass range from m/z 50–225. The identification was done by comparing retention times and mass fragmentation patterns of the synthetic standards with those obtained from the extracts.

Bioinformatics

For the prediction of the BGCs in the genome sequence of *Streptomyces* sp. QL37 (NZ_PTJS00000000.1) and other *Streptomyces* strains, the bioinformatic tool antiSMASH 5.0 was used (Blin *et al.*, 2019). BGC23 was compared with the maleimycin BGCs using the genome sequence of *S. showdoensis* ATCC 15227 (NZ_LAQS01000000).

Statistics

Heatmaps were generated using Matplotlib package (version 3.5.0)(Hunter, 2007). Ward's hierarchical clustering was applied using "ward" method in Scipy package (version 1.8.0) (Virtanen *et al.*, 2020).

SUPPLEMENTARY TABLES

Table S1A Genes used for the calculation of the average expression of a BGC shown in Figure 5

BGC Region	Class	Gene	locus_tag (NCBI, NZ_PTJS01000000)	Prokka annotation
BGC 1	NRPS	QL37_01200	C5F59_RS01200	Linear gramicidin synthase subunit D
BGC 2	Terpene	QL37_01665	C5F59_RS01660	tRNA 5-methylaminomethyl-2-thiouridine biosynthesis bifunctional protein MnmC
		QL37_01670	C5F59_RS01665	Demethylmenaquinone methyltransferase
		QL37_01675	C5F59_RS01670	15-cis-phytoene desaturase
		QL37_01680	C5F59_RS01675	Aminopyrrolnitrin oxygenase PrnD
		QL37_01685	C5F59_RS01680	All-trans-phytoene synthase
BGC 3	Bacteriocin	QL37_02695	C5F59_RS02730	Linocin-M18
BGC 4	T3PKS	QL37_02885	C5F59_RS02920	type III polyketide synthase
BGC 5	NRPS	QL37_03200	C5F59_RS03230	Dimodular nonribosomal peptide synthase
BGC 6	NRPS,T1PKS	QL37_03460	C5F59_RS03495	Gramicidin S synthase 2
BGC 7	Terpene	QL37_03860	C5F59_RS03900	Squalene--hopene cyclase
BGC 8	Linaridin	QL37_06275	C5F59_RS06355	hypothetical protein
		QL37_06280	C5F59_RS06360	hypothetical protein

Table S1A Genes used for the calculation of the average expression of a BGC shown in Figure 5 *(continued)*

BGC Region	Class	Gene	locus_tag (NCBI, NZ_PTJS01000000)	Prokka annotation
BGC 9a	T1PKS,hglE-KS	QL37_07220	C5F59_RS07315	Phthiocerol/phenolphthiocerol synthesis polyketide synthase type I PpsA
		QL37_07225	C5F59_RS07320	Phenolphthiocerol synthesis polyketide synthase type I Pks15/1
BGC 9b	T1PKS	QL37_07410	C5F59_RS07500	Phenolphthiocerol synthesis polyketide synthase type I Pks15/1
		QL37_07420	C5F59_RS07505	Erythronolide synthase, modules 3 and 4
		QL37_07440	C5F59_RS07510	Erythronolide synthase, modules 1 and 2
		QL37_07450	C5F59_RS07515	Phenolphthiocerol synthesis polyketide synthase type I Pks15/1
		QL37_07460	C5F59_RS07520	Oxygen regulatory protein NreC
		QL37_07465	C5F59_RS07525	Erythronolide synthase, modules 3 and 4
		QL37_07480	C5F59_RS07530	Phenolphthiocerol synthesis polyketide synthase type I Pks15/1
BGC 10	Bacteriocin	QL37_08865	C5F59_RS08915	hypothetical protein
BGC 11	Terpene	QL37_10150	C5F59_RS10245	Pentalenene synthase
BGC 12	T2PKS	QL37_12915	C5F59_RS12910	Actinorhodin polyketide putative beta-ketoacyl synthase 1
		QL37_12920	C5F59_RS12915	Actinorhodin polyketide putative beta-ketoacyl synthase 2

Table S1A Genes used for the calculation of the average expression of a BGC shown in Figure 5 *(continued)*

BGC Region	Class	Gene	locus_tag (NCBI, NZ_PTJS01000000)	Prokka annotation
BGC 13	PKS-like	QL37_13225	C5F59_RS13225	3-oxoacyl-[acyl-carrier-protein] synthase 3 protein 1
BGC 14	Siderophore	QL37_13755	C5F59_RS13750	Aerobactin synthase
BGC 15	Lanthipeptide-classIII	QL37_14765	C5F59_RS14775	Serine/threonine-protein kinase D
BGC 16	Butyrolactone	QL37_18390	C5F59_RS18460	hypothetical protein
BGC 17	Terpene	QL37_27840	C5F59_RS27970	(+)-T-muurolol synthase
BGC 18	Siderophore	QL37_28170	C5F59_RS28300	N(2)-citryl-N(6)-acetyl-N(6)-hydroxylysine synthase
		QL37_28175	C5F59_RS28305	N(6)-hydroxylysine O-acetyltransferase

Table S1B Genes used for the calculation of the average expression of a BGC shown in Figure 5

BGC Region	Class	Gene	locus_tag (NCBI, NZ_PTJS01000000)	Prokka annotation
BGC 19	NRPS-like	QL37_28445	C5F59_RS28575	Tyrocidine synthase 3
BGC 20	NRPS,Betalactone, other	QL37_28675	C5F59_RS28800	2-isopropylmalate synthase
		QL37_28680	C5F59_RS28805	Succinate-semialdehyde dehydrogenase (acetylating)
		QL37_28685	Not identified	hypothetical protein
		QL37_28690	C5F59_RS28815	Linear gramicidin synthase subunit B
		QL37_28695	C5F59_RS28820	Dimodular nonribosomal peptide synthase
		QL37_28700	C5F59_RS28825	Linear gramicidin synthase subunit B
		QL37_28705	C5F59_RS28830	Linear gramicidin synthase subunit D
BGC 21	NRPS	QL37_29280	C5F59_RS29410	Tyrocidine synthase 3
		QL37_29285	C5F59_RS29415	Linear gramicidin synthase subunit D
BGC 22	RiPP-like	QL37_29740	C5F59_RS29875	hypothetical protein

Table S1B Genes used for the calculation of the average expression of a BGC shown in Figure 5 *(continued)*

BGC Region	Class	Gene	locus_tag (NCBI, NZ_PTJS01000000)	Prokka annotation
BGC 23a	Betalactone	QL37_30710	C5F59_RS30845	2-isopropylmalate synthase
		QL37_30715	C5F59_RS30850	hypothetical protein
		QL37_30720	C5F59_RS30855	Alpha-aminoadipate--LysW ligase LysX
		QL37_30725	C5F59_RS30860	N-acetyl-gamma-glutamyl-phosphate reductase
		QL37_30730	C5F59_RS30865	Acetylglutamate kinase
		QL37_30735	C5F59_RS30870	GMP synthase [glutamine-hydrolyzing]
		QL37_30740	C5F59_RS30875	Catabolic NAD-specific glutamate dehydrogenase RocG
		QL37_30745	C5F59_RS30880	Short-chain-fatty-acid--CoA ligase
BGC 23b	Butyrolactone	QL37_30790	C5F59_RS30925	hypothetical protein
BGC 24	Ectoine	QL37_30990	C5F59_RS31125	L-ectoine synthase
BGC 25	Terpene	QL37_33360	C5F59_RS33500	Germacrene A synthase
BGC 26	NRPS	QL37_34050	C5F59_RS34210	Dimodular nonribosomal peptide synthase
BGC 27	hglE-KS	QL37_34940	C5F59_RS35105	Mycocerosic acid synthase
		QL37_34950	C5F59_RS35115	Polyketide biosynthesis protein PksE

Table S1B Genes used for the calculation of the average expression of a BGC shown in Figure 5 *(continued)*

BGC Region	Class	Gene	locus_tag (NCBI, NZ_PTJS01000000)	Prokka annotation
BGC 28	Bacteriocin	QL37_35390	C5F59_RS35545	hypothetical protein
BGC 29	NRPS	QL37_36325	C5F59_RS36490	Linear gramicidin synthase subunit D
BGC 30	T3PKS	QL37_36750	C5F59_RS36920	1,3,6,8-tetrahydroxy-naphthalene synthase
BGC 31	Melanin	QL37_37320	C5F59_RS37485	hypothetical protein
BGC 32a	T2PKS	QL37_37710	C5F59_RS37875	Actinorhodin polyketide putative beta-ketoacyl synthase 1
		QL37_37715	C5F59_RS37880	Actinorhodin polyketide putative beta-ketoacyl synthase 2
BGC 32b	Terpene	QL37_37755	C5F59_RS37920	Germacradienol/geosmin synthase
BGC 33	butyrolactone	QL37_38805	C5F59_RS38975	hypothetical protein
BGC 34	Terpene	QL37_38925	C5F59_RS39090	Pentalenene synthase
C2 BGC 1	NRPS	QL37_39520	C5F59_RS39750	hypothetical protein
		QL37_39525	C5F59_RS39755	Dimodular nonribosomal peptide synthase
		QL37_39530	C5F59_RS39760	Chondramide synthase cmdD
		QL37_39535	C5F59_RS39765	hypothetical protein
		QL37_39540	C5F59_RS39770	hypothetical protein

Table S2 Organization of BGC 23 and its comparison with the maleimycin cluster in *S. showdoensis* ATCC 15227

BGC 23		
Gene *S. sp.* QL37	**Protein**	**aa identity (%)with Mal protein in *S. showdoensis***
QL37-30710	Isopropylmalate synthase	88.7
QL37-30715	LysW	90,2
QL37-30720	RimK /LysX	70.6
QL37-30725	Semialdehyde dehydrogenase/LysY/ArgC	77.1
QL37-30730	Acetylglutamate kinase LysZ/ArgB	78.6
QL37-30735	Glutamine aminotransferase class I	77.5
QL37-30740	Glutamine /leucine/phenylalanine/valine dehydrogenase	83.4
QL37-30745	AMP-ligase	67.7
QL37-30750	Aminotransferase class I and II, Pyridoxal phosphate dependent	70.9
QL37-30755	Fumyrylacetoacetate hydrolase	58.1
QL37-30760	MerR family transcriptional regulator	Not in BGC
QL37-30765	MFS transporter	59.8
QL37-30770	DNA binding response regulator	51.9
Biosynthetic genes detected in *mal* gene cluster *S. showdoensis* and absent in BGC 23 of *S.* sp. QL37		
Gene in *S. showdoensis*	**Protein**	**aa identity (%) with protein in *S.*sp. QL37 outside BGC 23**
VO63_RS35620	Short-chain dehydrogenase	41.5
VO63_RS35610	Transposase	34.7
VO63_RS35600	Arylmalonate decarboxylase	No BLAST hit in genome
VO63_RS31285	Flavin dependent oxidoreductase	65.7
VO63_RS31290	Flavin dependent oxidoreductase	43

Table S2 Organization of BGC 23 and its comparison with the maleimycin cluster in *S. showdoensis* ATCC 15227 *(continued)*

BGC 23		
Biosynthetic genes detected in BGC 23 *S.*sp. QL37 absent in maleimycin BGC of *S. showdoensis*		
Gene *S. sp.* QL37	**Protein**	**aa identity with protein in *S.showdoensis***
QL37-30705	Acyl-CoA dehydrogenase	27.1
QL37-30700	Acyl-CoA dehydrogenase	28.6

SUPPLEMENTARY FIGURES

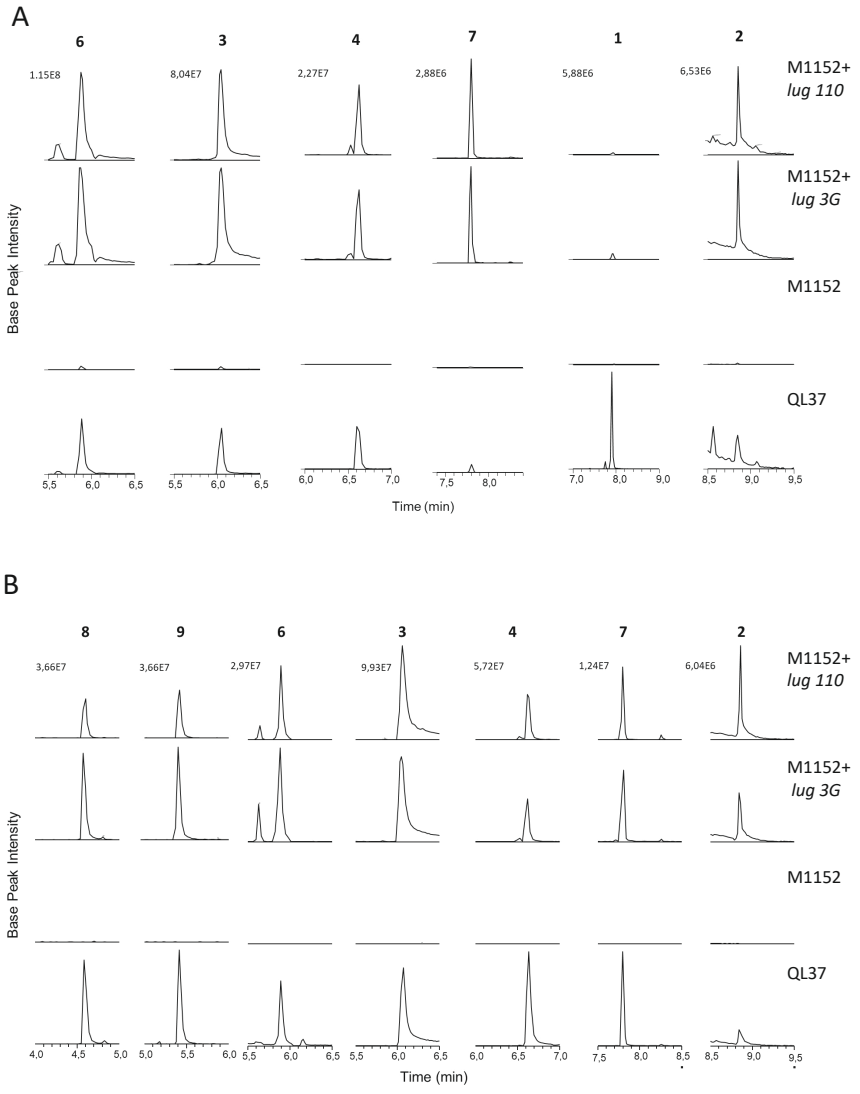


Figure S1 Extracted ion chromatogram of compounds **1–9** in the extracts of *Streptomyces* sp. QL37 and S. *coelicolor* carrying the *lug* gene cluster.
Extracted chromatogram (XIC) of the angucyclines (**2**,**3**,**4**,**6**) pratensilin A (**7**) limamycins (**8**,**9**) and lugdunomycin (**1**) in extracts derived from *S coelicolor.*M1152 carrying the *lug* gene cluster (MM152+lug 11O and MM152+lug 3G), *S. coelicolor* M1152 (MM152) and *Streptomyces* sp. QL37 (QL37). The strains were grown on MM (A) and R5 (B)

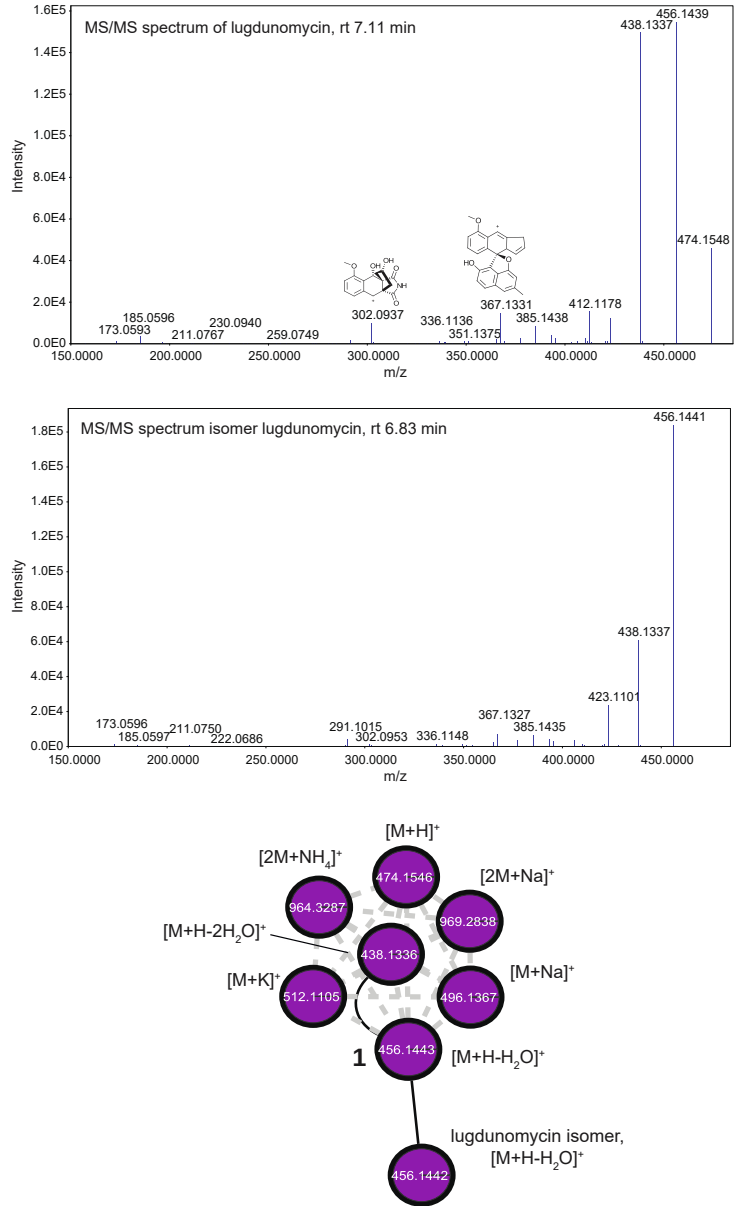


Figure S2 Subnetwork for lugdunomycin reveals lugdunomycin isomer.
Comparison of the MS/MS spectra of the different lugdunomycin isomers. These different molecules were detected at different retention times, but share the characteristic fragments of lugdunomycin, and thus could be analogues structures.

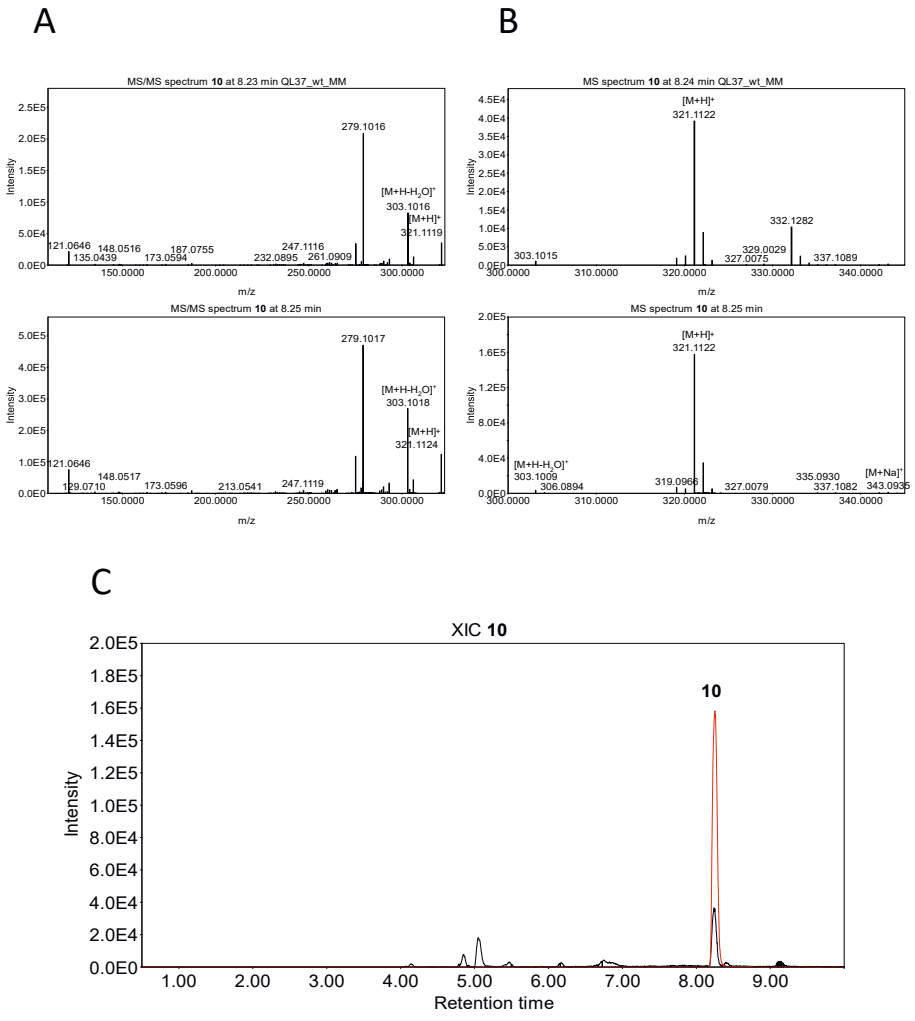


Figure S3 Identification of compound 10 in the extract of *Streptomyces* sp. QL37. Comparison of the MS/MS (A) and MS spectra (B) of compound **10** and its corresponding peak in the crude extract of *Streptomyces* sp. QL37. The extracted ion chromatograms of the two peaks are shown in (C) (Elmonin was a kind gift from Michiel Uiterweerd and Prof. Dr. A. Minnaard (university of Groningen).

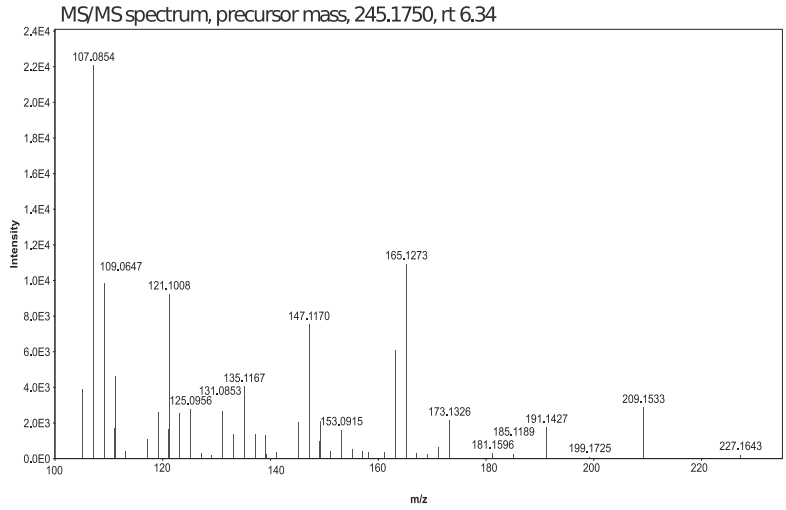


Figure S4 MS/MS spectrum of the precursor mass related to SCB1.
MS/MS spectrum of a node in one of the molecular families in the GNPS network indicated in Figure 3. This node showed a fragmentation pattern consistent with that of SCB1 as previously described (Yang *et al.*, 2005).

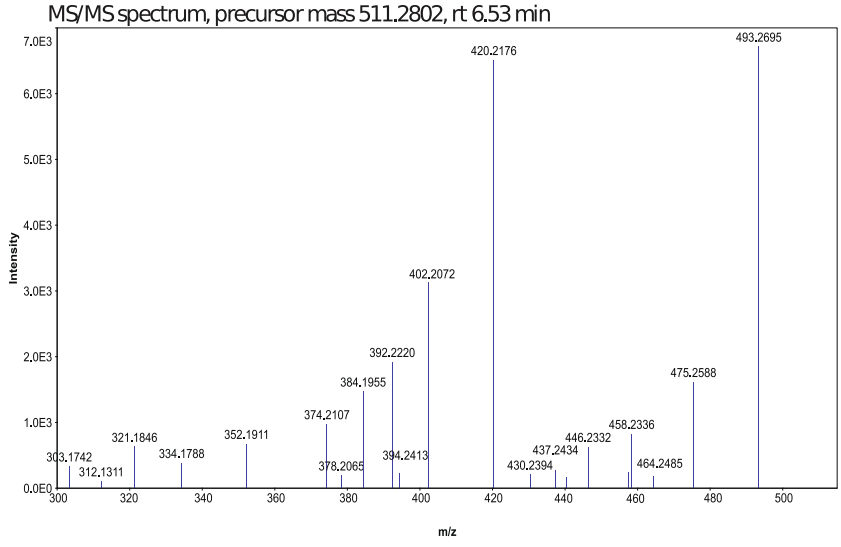


Figure S5 MS/MS spectrum of the precursor mass related to alteramide A.
MS/MS spectrum of a node in one of the molecular families in the GNPS network indicated in Figure 3. This node showed a fragmentation pattern consistent with that of alteramide A as previously described (Shaikh *et al.*, 2021, Moree *et al.*, 2014).

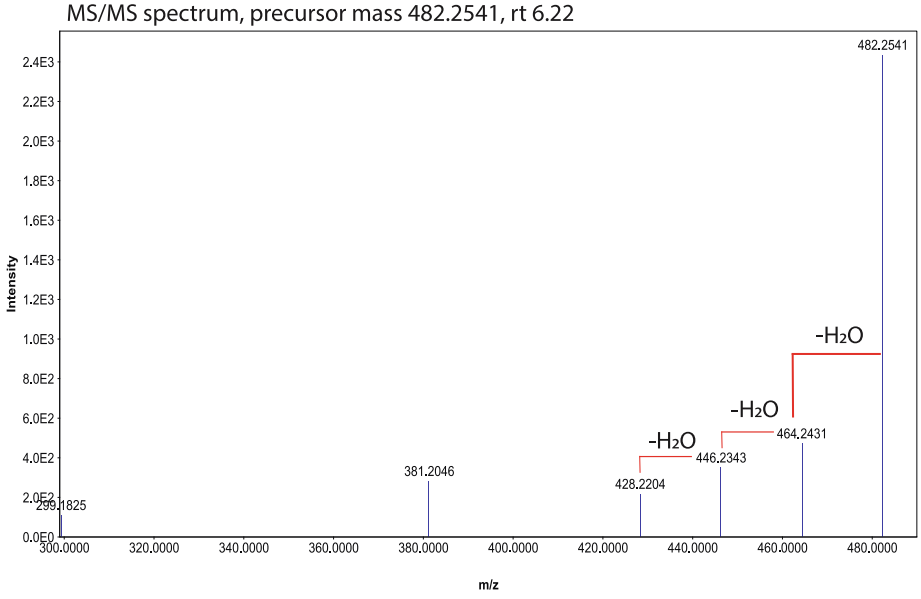


Figure S6 MS/MS spectrum of the precursor mass related to sceliphrolactam.
Detection of the monoisotopic mass of sceliphrolactam and the detected MS/MS spectrum in the extracts of *Streptomyces* sp. QL37 grown on MM. However, the MS/MS of sceliphrolactam has never been reported (Low *et al.*, 2018).

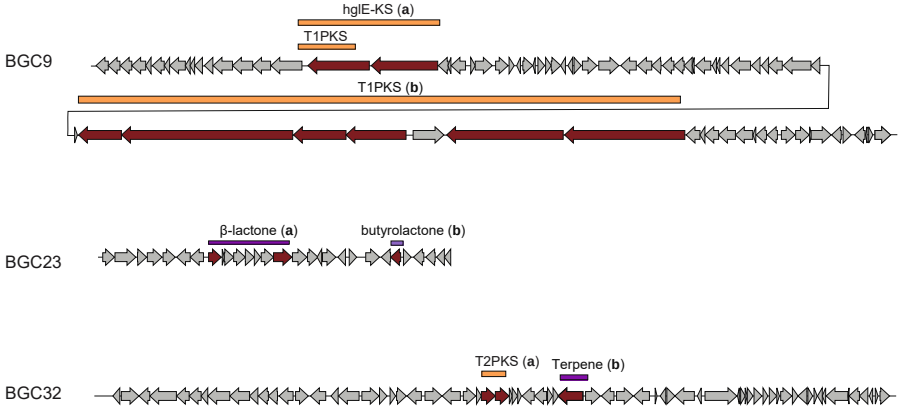


Figure S7 Organization of BGC9, BGC23 and BGC32 and annotation of BGC types.
Illustration was adapted from antiSMASH 5.0 (Blin *et al.*, 2019). Bars above the BGCs indicate the core regions (protocluster) that were assigned to each BGC and the BGC type, based on the predicted core genes (dark red genes). The expression of the genes included in the core regions was used to predict the average gene expression of the BGC. The different BGC types are indicated as “a” and “b” in Figure 5.

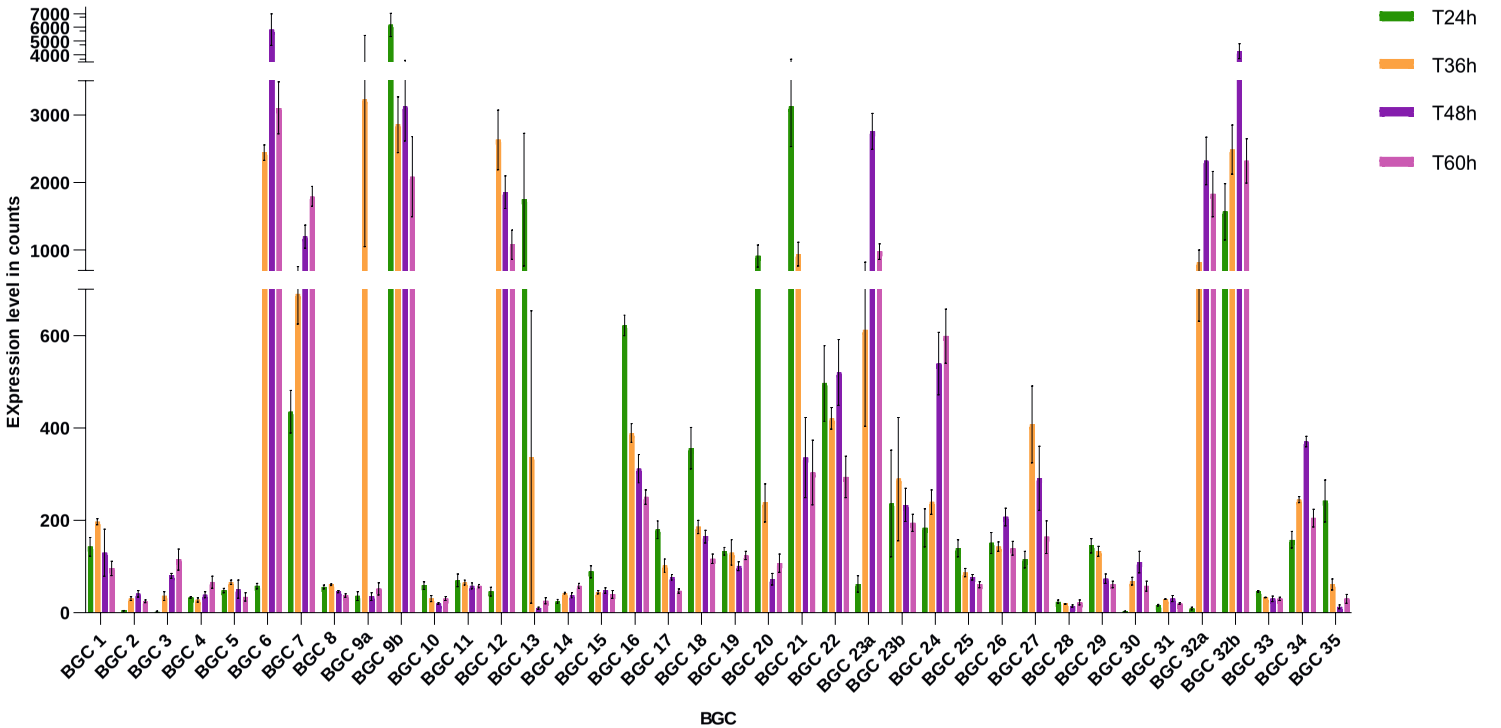


Figure S8 Average expression level of the core genes of the BGCs predicted in the genome of *Streptomyces* sp. QL37.
The graph corresponds to the heatmap indicated in Figure 5. The expression level is indicated in normalised counts. The RNA-seq data was normalized using DESeq2(Love *et al.*, 2014). The core biosynthetic genes encoding for a nonribosomal peptide synthase (NRPS), polyketide synthase (PKS), terpene cyclase or ribosomally-synthesised and post-translationally modified peptides (RiPP) were predicted by antiSMASH. In case the BGC contains multiple core genes, the average gene expression of multiple core genes was calculated. The error bars indicate the SEM.