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

Heul, H.U. van der

Analysis of the transcriptional regulatory genes of the biosynthetic gene cluster for angucyclines, limamycins and lugdunomycin in *Streptomyces* sp. QL37

Helga U. van der Heul
Chao Du
Lukas Kiefer
Somayah S. Elsayed
Gilles P. van Wezel

ABSTRACT

The *lug* gene cluster in *Streptomyces* sp. QL37 directs the production of non-rearranged angucyclines an C-ring cleaved angucyclines, such as the limamycins, pratensilin A and elmonin. In addition, the BGC is required for the production of lugdunomycin, a highly rearranged angucycline with an unprecedented scaffold. The extremely low production levels have so far prevented studies of mode of action and pharmacological properties of this molecule. To improve its production, more insight into the regulatory networks that controls its expression is needed. Here we present a mutational and functional analysis of four regulatory genes *lugRII-RV* of the angucycline biosynthetic gene cluster (*lug*) that is required for lugdunomycin biosynthesis, and report on the global changes in terms of development, metabolism, and protein expression. Our work shows that *lugRII*, *lugRIV* and *lugRV* are required for the expression of the BGC, and hence for production of angucyclines, limamycins and lugdunomycin. Deletion and overexpression experiments revealed that the orphan response regulator LugRIV and the SARP regulator LugRV are cluster-situated activators of the *lug* gene cluster, as was validated by metabolomics analysis, together with the upregulation of the *lug* proteins in the early stages of growth. Taken together this study shows that LugRIV and LugRV are the main positive regulators of the *lug* gene cluster and can be employed to increase angucycline production and for premature lugdunomycin production by *Streptomyces* sp. QL37.

INTRODUCTION

New antibiotics are desperately required, as the emergence of novel antimicrobial resistance pathogens increases (Lewis, 2013). *Streptomyces* are filamentous soil bacteria with a complex life cycle and are well known antibiotic producers. They have the potential to produce numerous natural products as is indicated by the number of biosynthetic gene clusters (BGCs) located on their genome. An average *Streptomyces* genome can contain more than 50 BGCs (Katz & Baltz, 2016). However, many of these are not or poorly expressed under laboratory conditions, also referred to as cryptic BGCs. Therefore various screening methods are applied to awaken and explore these BGCs (Rutledge & Challis, 2015). One of these is the "One strain-many compounds" strategy (OSMAC) (Romano *et al.*, 2018). Application of this method allowed the discovery of lugdunomycin (Wu *et al.*, 2019). The molecule is produced by *Streptomyces* sp. QL37, a strain derived from the Qinling mountains in China (Wu *et al.*, 2019). Besides lugdunomycin (**1**), also non-rearranged (typical) angucyclines (**2**–**6**) and the rearranged angucyclines pratensilin A (**7**), the limamycins (**8**, **9**) and elmonin (**10**) were detected in the extracts of *Streptomyces* sp. QL37 (Figure S1) (Wu *et al.*, 2019, Xiao *et al.*, 2020). A type II polyketide BGC (*lug*) was identified that encodes the biosynthetic enzymes for these molecules ((Wu *et al.*, 2019); Chapter 3). Rearranged angucyclines are derived from the angucycline backbone (**2**,**3**), following cleavage of the C6a–C7 bond of the canonical angucycline C-ring (Wu *et al.*, 2019). It was proposed that lugdunomycin is the product of a final Diels-Alder reaction between a rearranged angucycline and *iso*-maleimycin (**11**) (Wu *et al.*, 2019, Uiterweerd, 2020). While angucyclines are produced at high levels under various growth conditions, lugdunomycin is hardly produced. To obtain more insight into the activity of lugdunomycin, it is essential to improve its production.

One approach to enhance the expression of a BGC, and thereby induce the production of the respective natural products, is to manipulate the regulatory pathway that controls it (van der Heul *et al.*, 2018). In *Streptomyces* the control of antibiotic production is highly complex and is intertwined with the control of the morphological development. Environmental signals are often sensed by pleiotropic regulators that have a global impact on morphological development and specialised metabolism in *Streptomyces* (Liu *et al.*, 2013, van der Heul *et al.*, 2018). These pleiotropic regulators control the transcriptional activity of regulatory

genes within the BGC, which are referred to as cluster-situated regulators (CSRs). In general, the CSR controls the transcription of the specific BGC it is embedded in, though CSRs that control the activity of multiple BGCs have also been reported (van der Heul *et al.*, 2018, Liu *et al.*, 2013). Different CSR families are known; the best known belongs to the *Streptomyces* antibiotic regulatory protein (SARP) transcriptional regulatory family. These regulators activate the transcription of their target BGC. Famous examples are ActII-ORF4 (actinorhodin), RedD (undecylprodigiosin) and DnrI (daunorubicin) (Fujii *et al.*, 1996, Prija *et al.*, 2017, Hulst *et al.*, 2021). For details see Chapter 2.

The data presented in Chapter 4 suggest that the *lug* cluster specifies angucyclines and limamycins, but not lugdunomycin; indeed, a second BGCs is required for the production of lugdunomycin, namely BGC 23 that is most likely involved in the production of *iso*-maleimycin (**11**). Here, we analyse the transcriptional regulatory genes of the *lug* gene cluster. Comparison of the *lug* gene cluster with other *lug*-type clusters revealed that the regulatory genes *lugRII*, *lugRIII*, *lugRIV* and *lugRV* are typically shared among angucycline BGCs with high similarity to the *lug* gene cluster. We then studied the metabolic, proteomic and morphological changes resulting from the deletion and over-expression of the *lugR* genes. These studies demonstrated that LugRIV and LugRV act as the major transcriptional activators of the *lug* gene cluster.

RESULTS

Bioinformatic analysis of the putative regulators from the *lug* gene cluster

The *lug* gene cluster contains five regulatory genes designated as *lugRI*, *lugRII*, *lugRIII*, *lugRIV* and *lugRV*. The regulatory genes *lugRII–lugRV* are all located in BGCs that are highly similar to the *lug* gene cluster, suggesting that these are required for the control of the BGC (Figure 1, Table S1). Only the regulatory gene *lugRI* is absent in the highly related gene clusters, indicated in Figure 1. To predict the functional role of the *lugR* genes the pfam domains were predicted in the amino acid sequences of the regulatory proteins.

LugRI is a predicted 279 amino acid (aa) protein composed of a C-terminal MmyB-like transcription regulator ligand binding domain (pfam: MLTR-LBD) and an N-terminal helix-turn-helix (HTH) DNA-binding domain that belongs to the

XRE-family of transcriptional regulators (pfam: HTH-31). MmyB-like regulators occur in a wide range of bacteria and are associated with antibiotic and fatty acid metabolism in Actinobacteria (Xu *et al.*, 2012). Genes encoding an MmyB-regulator are often associated with genes for short-chain dehydrogenases/reductases, such as *lugM*, which implies a functional connection between *lugRI* and *lugM* (Xu *et al.*, 2012).

LugRII (222 aa) belongs to the family of NarL-type atypical orphan response regulators (ARRs). It contains a C-terminal LuxR-type HTH DNA-binding domain (pfam: GerE), which is found in NarL/FixJ family of response regulators. RedZ and DnrN, the activators of the undecylprodigiosin BGC in *S. coelicolor* and the daunorubicin BGC in *S. peucetius*, respectively, also belong to this regulatory family (Furuya & Hutchinson, 1996, Wang *et al.*, 2009, Guthrie *et al.*, 1998). RedZ and DnrN activate transcription of the pathway-specific activator genes of their cognate BGC, namely *redD* and *dnrI*, respectively (White & Bibb, 1997, Furuya & Hutchinson, 1996).

LugRIII (238 aa) contains a C-terminal AefR transcriptional regulatory protein domain (pfam: TetR_C_7) and an N-terminal TetR-like DNA-binding domain (pfam: TetR_N). Regulators belonging to the TetR family often repress the transcription of divergently oriented transporter genes (Ramos *et al.*, 2005). The repression is often released due to the binding of the end product of the BGC to the TetR regulator. Also in the *lug* gene cluster *lugRIII* lies next to *lugTI,* which encodes a putative transporter.

LugRIV (266 aa) belongs to the OmpR-type regulators. It contains a C-terminal OmpR-like DNA- binding domain (pfam: Trans_reg_C) and an N-terminal domain response regulator receiver domain (pfam: response _Reg). The protein shares significant homology with JadR1 (52% aa identity) and Aur1P (54% aa identity), which are activators of the jadomycin BGC in *S. venezuelae* ISP5230 and the auricin BGC in *S. aureofaciens* CCM 3239, respectively (Wang *et al.*, 2009, Novakova *et al.*, 2005).

Finally, LugRV (645 aa) is a member of the family of *Streptomyces* antibiotic regulatory proteins (SARPs). SARP regulators often act as cluster-specific activators of antibiotic production (van der Heul *et al.*, 2018, Wietzorrek & Bibb,

1997). They are characterised by an OmpR HTH DNA-binding domain (pfam: Trans_reg_C) and a bacterial transcription activator domain (pfam: BTAD domain) (Liu *et al.*, 2013). Similar to the pleotropic antibiotic regulator AfsR from *S. coelicolor* (Horinouchi *et al.*, 1990), LugRV contains a centrally located NB-ARC domain. SARPs are the most well studied regulators.

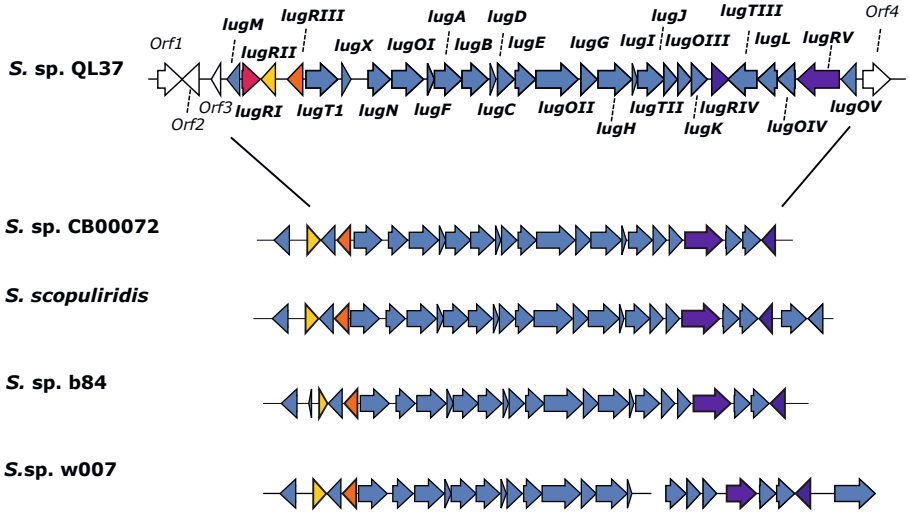

Figure 1 Location of the regulatory genes in the *lug* gene cluster and related type II polyketide gene clusters.
On top the *lug* gene cluster is shown. Below the *lug* gene cluster and the type II polyketide gene clusters are shown from *Streptomyces* sp. CB00072, *S. scopuliridis* RB72, *Streptomyces* sp. b84 and *Streptomyces* sp. w007 (Zhang *et al.*, 2012, Kibret *et al.*, 2018, Cao *et al.*, 2021). All gene clusters contain orthologues of *lugRII* (yellow), *lugRIII* (orange), *lugRIV* (blue) and *lugRV* (purple), but not of *lugRI* (red).

Mutational analysis of the *lugR* genes

The bioinformatic analysis presented above showed that genes highly similar to *lugRII–lugRV* are conserved in BGCs that are highly similar to the *lug* gene cluster. To evaluate the role of these genes in lugdunomycin biosynthesis, we sought to inactivate the genes independently in *Streptomyces* sp. QL37 via homologous recombination (Materials and Methods). To this end, knock-out constructs pΔ*lugRII*, pΔ*lugRIII*, pΔ*lugRIV*, pΔ*lugRV* were created, based on a method published previously (Swiatek *et al.*, 2012). The generated constructs were composed of the conjugatable pWHM3-*oriT* harbouring approximately 1.5 kb of the upstream and downstream regions of the respective *lugR* genes, with an apramycin resistance cassette (*aac*(3)*IV*) in-between (Vara *et al.*, 1989, Garg

& Parry, 2010). The constructs were conjugated to *Streptomyces* sp. QL37 using the methylation-deficient *E. coli* ET12567/pUZ8002 (MacNeil *et al.*, 1992). Through homologous recombination either one of the *lugR* regulatory genes was replaced by the apramycin cassette. The presence of *loxP* sites flanking the apramycin cassette allowed the efficient removal of the resistance cassette by introduction of the pUWLcre construct expressing Cre recombinase (Khodakaramian *et al.*, 2006, Swiatek *et al.*, 2012). The generated in-frame single gene deletion mutants for *lugRII, lugRIII, lugRIV* and *lugRV* were designated QL37Δ*lugRII*, QL37Δ*lugRIII*, QL37Δ*lugRIV*, and QL37Δ*lugRV*, respectively.

To assess the possible role of the *lugR* genes in the control of development and angucycline production, the respective mutants were grown on different media for seven days at 30 °C (Figure 2). Angucyclines are brown or yellowish diffusible pigments, and therefore their production can be seen by visual inspection (Wu *et al.*, 2019). However, lugdunomycin is colourless and its production can only be followed based on its spectral properties (Wu *et al.*, 2019). Wild-type *Streptomyces* sp. QL37 and its *lugR* mutants were grown on minimal media (MM) agar plates supplemented with 0.5% mannitol and 1% glycerol, on R5 agar supplemented with 0.8% peptone and 1% mannitol (R5), or on soya flour mannitol (SFM) agar. Strain QL37Δ*lug-pks* that lacks the minimal polyketide synthase (PKS) genes was included as a control for the absence of angucycline production (Wu *et al.*, 2019) (described in Chapters 3, 4 and 5). These media were selected because limamycins were previously isolated from R5, lugdunomycin from MM and angucyclines from both media. SFM was selected as it is routinely used for promoting sporulation. On MM agar plates, the wild-type *Streptomyces* sp. QL37 produced spores and a yellow pigment, on R5 agar it only produced a sparse aerial mycelium and a brown pigment, while on SFM agar it produced abundant spores and was also brown-pigmented. Small differences were seen between QL37Δ*lug-pks* and the wild-type strain on MM and SFM, but on R5 medium differences were significant; on this media, the wild-type strain hardly developed, while the mutant developed well.

On R5 media, QL37Δ*lugRIII* had a similar phenotype as the wild-type strain. Conversely, QL37Δ*lugRII*, QL37Δ*lugRIV*, and QL37Δ*lugRV* resembled the *lug-pks* mutant, as it was less pigmented and showed better development than the wild-type strain. On MM the phenotypes were similar to those of the wild-type strain, although QL37Δ*lugRII* sporulated better, while sporulation was reduced

in QL37Δ*lugRV*. Mutants QL37Δ*lugRII,* QL37Δ*lugRIV* and QL37Δ*lugRV* hardly produced brown pigments on SFM agar plates. We anticipate that the brown pigment reflects the level of angucycline production, and that lack of sporulation on R5 is due to high production of angucyclines. Following this logic, it is likely that angucyclines were not produced in significant quantities in cultures of the *lugRII, lugRIV* or *lugRV* mutants, while the *lugRIII* mutant still produced abundant angucyclines.

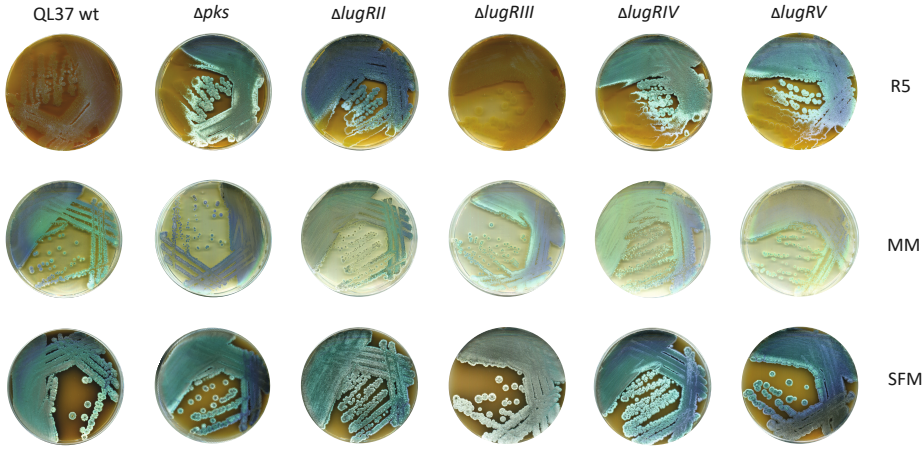

Figure 2 Phenotypes of *lugR* mutants of *Streptomyces* sp. QL37.
The strains were grown on MM, on R5 and on SFM agar plates for seven days at 30 °C. The phenotype of the *lugRII*, *lugRIV* and *lugRV* mutants is similar to the *pks* mutant, implying that these strains do not produce angucyclines, while the *lugRIII* mutant is similar to the wild type implying that this strain still produced angucyclines.

Metabolomic analysis of the *lugR* mutants

To analyse the effect of the deletion of the *lugR* genes on angucycline production in more detail, ethyl acetate extracts were prepared from the wild-type strain, the *lug-pks* mutant and the different *lugR* mutants and analysed using liquid chromatography coupled to mass spectrometry (LC-MS). For this, the strains were grown in triplicate on MM and R5 agar plates (Wu *et al.*, 2019). Natural products were extracted from both the mycelium and the agar. MZmine was used to process the LC-MS data. Principle component analysis (PCA) (Figure 3A and B), and a heatmap with hierarchical clustering analysis were generated based on the processed peak lists (Figure 3C and D). The PCA showed clear separation of the metabolome of the mutants QL37Δ*lugRII*, QL37Δ*lugRIV*, QL37Δ*lugRV* and QL37Δ*pks* from that of the wild-type strain and its *lugRIII* mutant. This is also

reflected in the heatmap. Analysis of the extracted ion chromatograms (Figure 4) and the respective peak areas of the molecules that were previously isolated confirmed that QL37Δ*lugRIII* still produced the typical angucyclines (**2**–**6**), the rearranged angucyclines (**7**–**10**), and lugdunomycin (**1**). On MM the production level of the molecules was lower in QL37Δ*lugRIII* as compared to the levels produced by the wild-type strain. No angucyclines or derivatives thereof were produced by QL37Δ*lugRII*, QL37Δ*lugRIV* and QL37Δ*lugRV*.

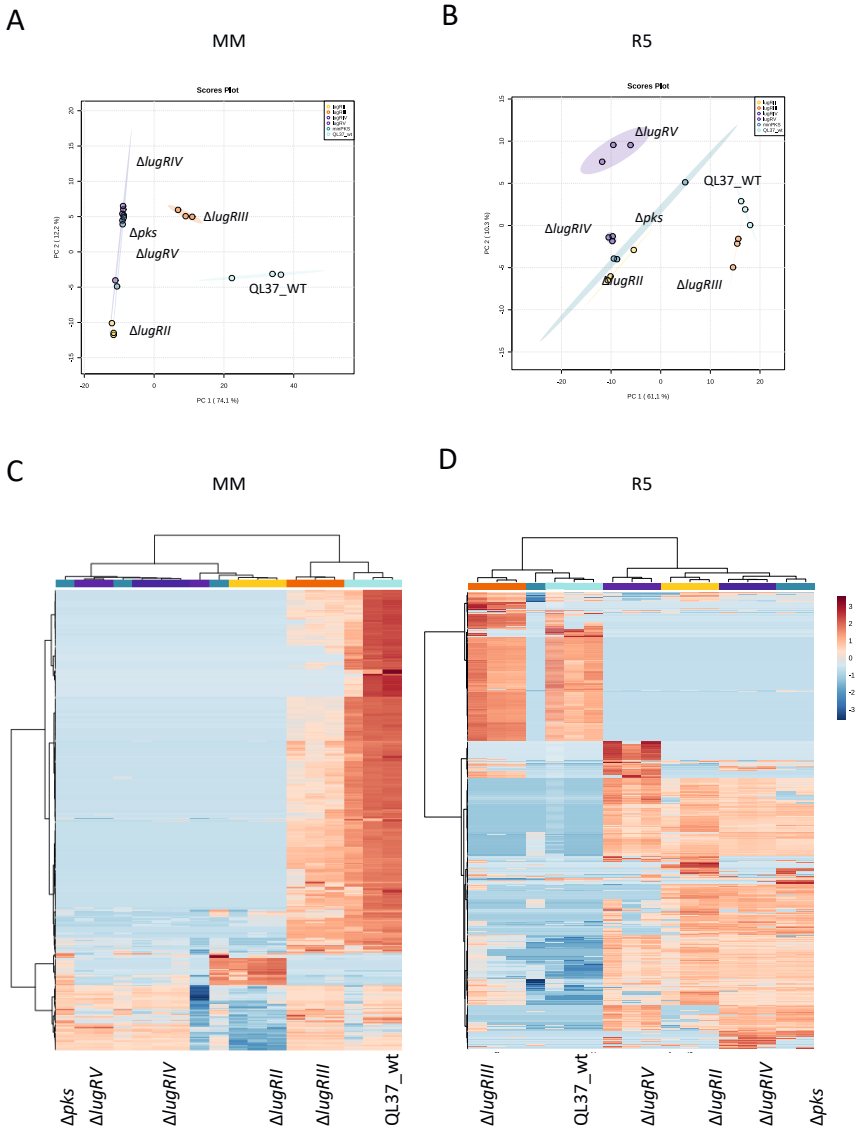

Figure 3 Exploratory statistical analysis of the metabolomic profiles of *Streptomyces* sp. QL37, its *lugRII–RV* and *lug-pks* mutants grown on MM and R5.
PCA score plots of the extracts from the strains grown on MM (A) and R5 (B). Heatmaps with additional hierarchical clustering of the metabolomic profile of the extracts from the analysed strains grown on MM (C) and R5 (D). Mass features that have relatively high peak aeras are displayed in red and mass features that have relatively low peak areas are displayed in blue. Note that mutants lacking *lugRII, lugRIV* or *lugRV* failed to produce angucyclines under these conditions.

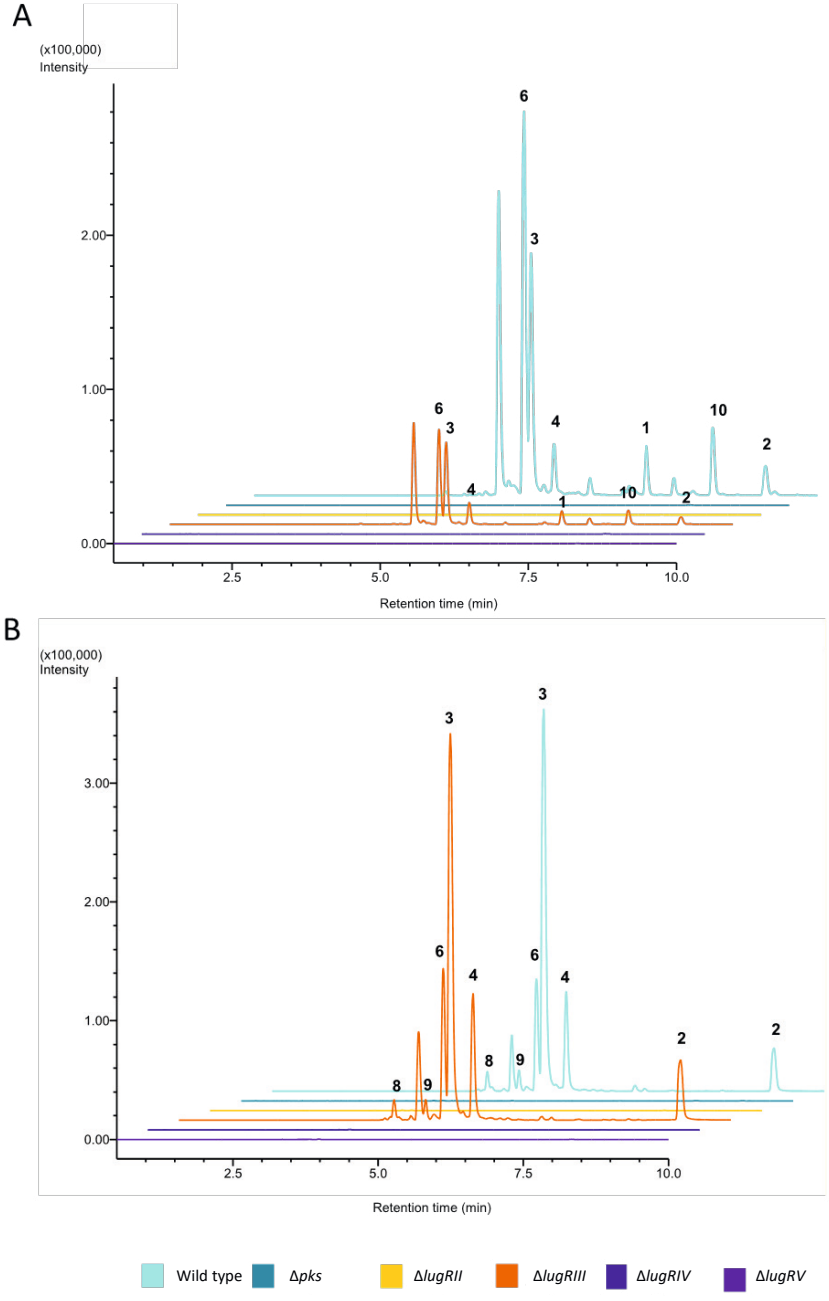

Figure 4 Overlay of extracted ion chromatograms of compounds 1–10 in the extracts from *Streptomyces* sp. QL37, its *lugR* mutants and its *pks* mutant grown on MM (A) or R5 (B). The wild-type strain and its *lugRIII* mutant produced angucyclines, rearranged angucyclines and lugdunomycin, while *lugRII*, *lugRIV* and *lugRV* mutants did not, similar to the *lug-pks* null mutant. See also Figure 3.

Complementation of the *lugRII*, *lugRIV* and *lugRV* mutants and hierarchical order of their regulatory activity

To ascertain that the observed effects of the deletion of *lugRII*, *lugRIV* and *lugRV* were indeed solely due to the deletion of these genes, and to assess whether the regulators are subject to a certain hierarchy in terms of activation of the *lug* gene cluster, the different *lugR* mutants were complemented by introduction of the overexpression constructs for the regulatory genes (Figure 5). For this, the *lugR* genes were each cloned behind the constitutive *ermE* promoter (*ermE**p) in the conjugative vector pSET152, which integrates at the øC31 attachment site in the genome (Motamedi *et al.*, 1995). This resulted in the overexpression constructs pO*lugRII*, pO*lugRIII*, pO*lugRIV*, and pO*lugRV*. Each construct was conjugated to all mutants that were unable to produce angucyclines, namely to QL37Δ*lugRII*, QL37Δ*lugRIV* and QL37Δ*lugRV*. The empty vector pSET152 was used as the control.

Introduction of the *lugRV* construct into the *lugRII*, *lugRIV* and *lugRV* mutants restored the production of the brown diffusible pigments and abolished development on R5 agar, indicative of restoration of angucycline production in all three mutants. However, overexpression of *lugRIV* restored angucycline production in only *lugRII* and *lugRIV* but not in *lugRV* mutants, indicating a hierarchical relationship between *lugRIV* and *lugRV*, with *lugRV* dominant over *lugRIV*. In turn, *lugRII* failed to complement any of the *lugR* mutants, including the *lugRII* mutant itself. Still, the observation that angucycline production was restored to the *lugRII* mutant by introduction of constructs expressing *lugRIV* or *lugRV* from the *ermE**p, strongly suggests that the *lugRII* mutant is not defective in angucycline biosynthesis due to second site mutations and thus that *lugRII* is indeed a positive regulator of the *lug* gene cluster. Why p*lugRII* could not complement the *lugRII* mutant remains to be elucidated. Introduction of *lugRIII* did not affect the production of angucyclines in the mutants *lugRII*, *lugRIV* and *lugRV*, which is in line with the observation that deletion of *lugRIII* only has a minor effect on angucycline production. As for cultures on MM and SFM agar, only introduction of the *lugRV* expression construct had an effect, in line with the minor phenotypic differences seen for the *lugR* mutants on these media.

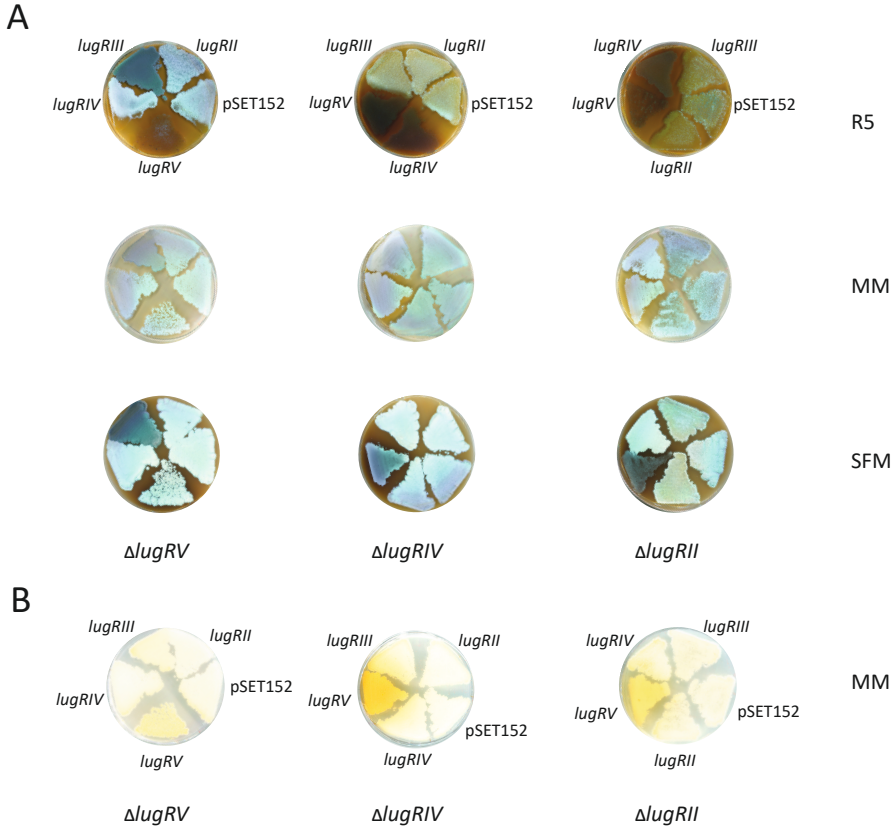

Figure 5 Complementation of *lugR* mutants via expression of the *lugR* genes from the *erm*E*p promoter.
Integrative plasmids expressing either *lugRII*, *lugRIII*, *lugRIV* or *lugRV* from the *erm*E* promoter were introduced into the respective mutants, and restoration of angucycline production assessed. As a control, transformants harbouring the empty vector (pSET152) were included in the experiment.

Effect of the enhanced expression of the *lugR* genes on development and angucycline production

We then wondered what the effect would be of the enhanced expression of the *lugRIV* and *lugRV* genes in the wild-type strain. These genes were chosen for more detailed analysis, considering their major impact on the control of the *lug* gene cluster. Overexpression strains O*lugRIV* and O*lugRV* and the control strain *Streptomyces* sp. QL37::pSET152 were grown on MM, R5 and SFM agar plates (Figure 6). When grown on SFM, the *Streptomyces* sp. QL37::pSET152 produced

aerial hyphae that were largely devoid of spores, indicating that the introduction of pSET152 already affected the morphological differentiation. Reduced sporulation was also observed for ex-conjugants O*lugRIV and* O*lugRV*. On this medium, O*lugRV* stood out from the other strains by producing large amounts of brown pigment, together with a bald phenotype with only sparse aerial mycelium that is most likely due to the DNA-targeting activity of angucyclines (Kharel *et al.*, 2012). On MM agar, the control strain and O*lugRIV* produced spores and yellow pigments, while O*lugRV* showed production of yellow pigment and slightly less sporulation than the control strain. Finally, when the strains were grown on R5 medium all the exconjugants produced abundant brown pigments, associated with reduced development. Strains O*lugRIV* and O*lugRV* produced more brown pigments than the control strain, which indicates that these strains produced more angucyclines.

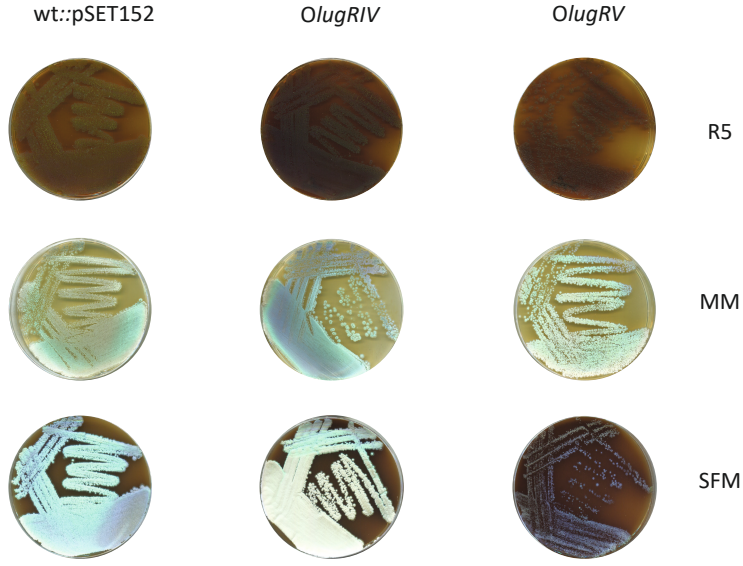

Figure 6 Effects of enhanced expression of the regulatory genes *lugRIV* and *lugRV* on morphogenesis and antibiotic production.
Strains O*lugRIV* and O*lugRV* produced more brown pigments on R5 than the control strain, while O*lugRV* was inhibited in sporulation on SFM after growth for seven days at 30 °C

We then aimed to verify whether the enhanced expression of the regulatory genes from the *lug* gene cluster affected the respective secondary metabolites and subsequently the expression of Lug proteins. To do so, we followed a paired omics strategy by combining metabolomics with quantitative proteomics (Schorn *et al.*, 2021). We analysed cultures of the overexpression strains O*lugRIV* and O*lugRV*,

which overexpress *lugRIV* and *lugRV*, respectively, from the strong *ermE** promoter. The plasmid pSET152 without insert was used as the control. To perform the combined metabolomics/proteomics experiment, all strains were grown in triplicate on MM and R5 covered with cellophane for two days. Mycelia and spores were harvested to prepare proteins extracts for quantitative proteomics, while the secondary metabolites were extracted from the remaining solid media, for analysis via LC-MS. In addition, the strains were grown for seven days on MM, the time point at which lugdunomycin is produced, and secondary metabolites were subsequently extracted from mycelium and agar.

In line with the observed changes in pigmentation and development, PCA and heatmap hierarchical clustering analysis revealed major changes in the metabolome of MM-grown O*lugRIV* and O*lugRV* (Figure 7A and C), while on R5 agar the differences were less significant (Figure 7B and D). The effect of the overexpression of the regulatory genes on the different structural classes produced by QL37 was then investigated in more detail using molecular networking. Using Feature-Based molecular networking (FBMN) in the Global Natural Product Social Networking (GNPS) platform, molecular networks of structurally related metabolites were generated in which isomers having different retention times can be identified, and the relative quantification of the mass features could be mapped (Nothias *et al.*, 2020). Ion Identity Molecular networking was then used to connect ion adducts of the same mass feature based on several criteria (Nothias *et al.*, 2020, Schmid, 2020). In this way, two molecular networks were generated. The ions detected in the media blanks were removed before generation of the networks (Figure 8 and Figure S2).

Analysis of the molecular network from MM-grown cultures obtained after 2 days of growth revealed different molecular families for the non-rearranged angucyclines, the limamycins pratensilin A and lugdunomycin (Figure 8). Elmonin (**10**) was identified as a single node in the network. In addition, a molecular family of γ-butyrolactones was detected (Chapter 4) (Yang *et al.*, 2005). Both the rearranged and non-rearranged angucycline molecular families were predominantly present in O*lugRIV* and O*lugRV*, while lugdunomycin was mainly detected in O*lugRV*. The vast majority of the mass features in the angucycline-related molecular families were detected only in O*lugRIV* and O*lugRV* and not or hardly in the parental strain; this could be due to the early extraction time point (two days), when angucyclines

are not yet or hardly produced in wild-type cells, while production is accelerated in the overexpression strains. The control strain produced more angucyclines and limamycins on R5 agar, albeit with lower concentrations as compared to the overexpression strains, particularly in comparison with O*lugRIV* (Figure S2).

When the metabolites were extracted from a seven days culture on MM (the time point which was previously used for the isolation of lugdunomycin (Wu *et al.*, 2019)), the angucycline-related metabolites could be detected in cultures of the control strain and of the overexpression strains (Figure S3). Again, concentrations were higher in the overexpression strains than in the control, although the difference was somewhat less pronounced as compared to the early time points. Interestingly, lugdunomycin production was not increased at this time point and was even slightly decreased compared to the control strain. This suggests uncoupling of lugdunomycin and angucycline biosynthesis, most likely because BGC 23 is required for the biosynthesis of lugdunomycin, but not for other angucyclines and derivatives.

Analysis of the proteome of the strains overexpressing the *lugR* genes

Quantitative proteomics was executed on the proteins isolated from the biomass of the O*lugRIV*, O*lugRV* and pSET152 control strains grown for two days on MM and R5 at 30 °C. First PCA analysis was executed to assess the reproducibility of the replicates and visualise the similarity of the proteome among the three recombinant strains (Figure S4). PCA analysis showed that the replicates of the proteome of the control and O*lugRIV* were separated in both MM and R5 media, with a relatively small variation within each set, while variation in O*lugRV* samples was relatively large in both cultures. Still, these replicates were also clearly separated from the control strain and to a lesser extent from O*lugRIV* in MM samples, which is in line with the previously observed difference in the metabolome of these strains. Bioinformatics analysis showed that the *lug* gene cluster most likely contains 28 genes (Chapter 3). Of the predicted gene products, 17 were detected in both the MM and R5-grown cultures by proteomics. To visualise the differential expression of the Lug proteins in each of the overexpression strains relative to the control, heatmaps and violin plots were generated (Figure 9A and B; Figure S5A and B). Violin plots were generated to show the distribution of the overall expression of the Lug proteins. When grown on MM, the Lug proteins were upregulated in O*lugRIV* and O*lugRV* compared to the control strain (Figure 9A and B). This correlated

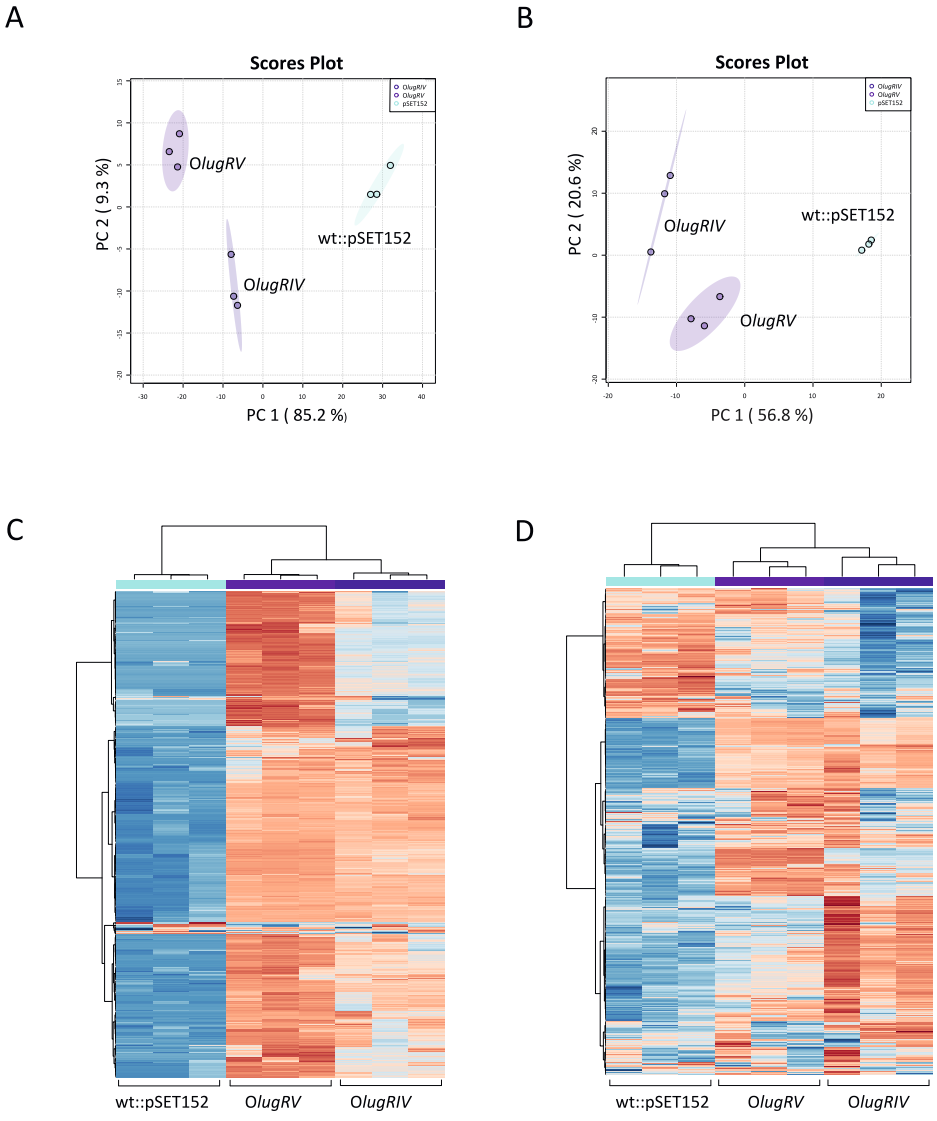

Figure 7 Statistical analysis of the metabolomic profiles of *Streptomyces* sp. QL37 over-expressing *lugRIV* and *lugRV*.
PCA score plots of the extracts from the analysed strains grown on MM (A) and R5 (B) for 48 hours. Heatmaps with additional hierarchical clustering of the metabolomic profile of the extracts from the analysed strains grown on MM (C) and R5 (D). Mass features that have relatively high peak areas are displayed in red and mass features that have relatively low peak areas are displayed in blue.

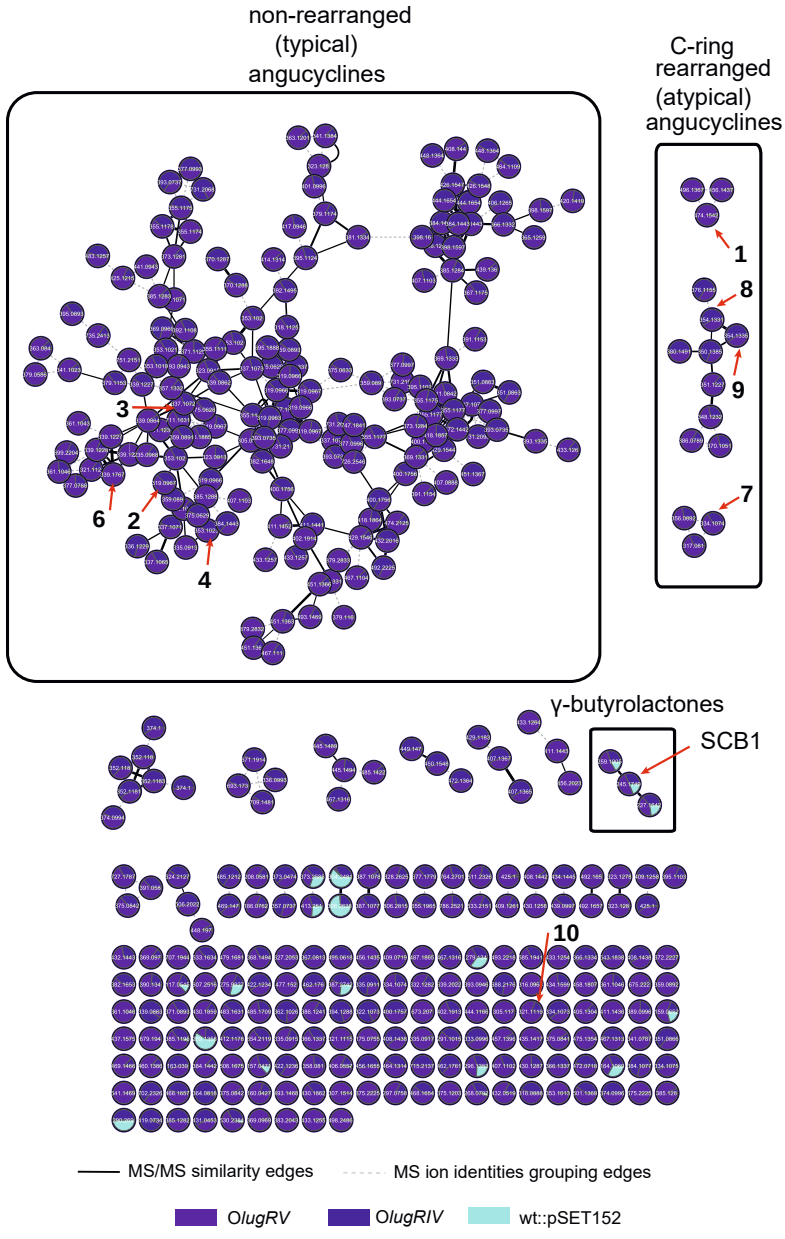

Figure 8 Molecular network of the ions detected in the extracts of *Streptomyces* sp. QL37 O*lugRIV*, O*lugRV* and the control strain grown on MM.
The nodes are labelled by the precursor mass of their ions and pie charts are mapped to the nodes to indicate the relative intensities of the ions in the different samples. Arrows are used to highlight the metabolites which were previously identified in *Streptomyces* sp. QL37. The molecular families, which were assigned based on the previously known metabolites or based on analysing their MS/MS spectra, are highlighted.

well with the enhanced or precocious production of the angucycline-related metabolites. The change in the Lug proteins was less significant on R5 medium, again in line with the metabolomics data (Figure S4). As lugdunomycin is derived from two BGCs, namely the angucycline BGC (*lug*; BGC 12) and the *iso*-maleimycin BGC (BGC 23) (Chapter 4), the proteins encoded by the *iso*-maleimycin BGC were also investigated. In samples from MM-grown cultures ten proteins encoded by the *iso*-maleimycin BGC were detected (Figure S6), whereas in samples from R5-grown cultures only three proteins were detected (Figure S5). Although the expression of the *iso*-maleimycin BGC appeared increased in the heatmaps in O*lugRV* as compared to the control strain, this increase was not statistically significant (*p*-value = 0.21733) according to the Mann-Whitney test (Figure 9 and Figure S6). This is in accordance with the observed failure to detect a significant increase in lugdunomycin production over time, despite the constitutive expression of the activators of the *lug* gene cluster (Figure S3).

DISCUSSION

The aim of this study was to obtain insight into the regulatory network of the *lug* gene cluster, as a prelude to improve the production of lugdunomycin. Based on bioinformatic predictions, the *lug* gene cluster encodes five regulatory proteins belonging to five different families, including an XRE-type regulator (LugRI), an atypical response regulator (ARR) of the LuxR-type (LugRII), a TetR-like regulator (LugRIII), an ARR of the OmpR-type (LugRIV) and a SARP regulator (LugRV). In this study the role of *lugRII–V* in angucyclines and lugdunomycin production was examined. These regulatory genes are conserved in *lug*-type gene clusters highly similar to *lug* (Chapter 4).

Targeted deletion combined with phenotypic and metabolomics analysis, showed that LugRII, LugRIV and LugRV are positive regulators of the *lug* gene cluster, while LugRIII played only a minor role under the conditions tested. The *lugRIV* and *lugRV* mutants could be complemented with the introduction of constructs expressing the respective genes from the *ermE* promoter, which showed that there were no polar effects caused by the generated mutations. However, the *lugRII* mutants could not be complemented by the introduction of *lugRII*; instead, angucycline production could be restored by expressing either *lugRIV* and *lugRV*. This suggests that the angucycline biosynthetic genes were not affected in the

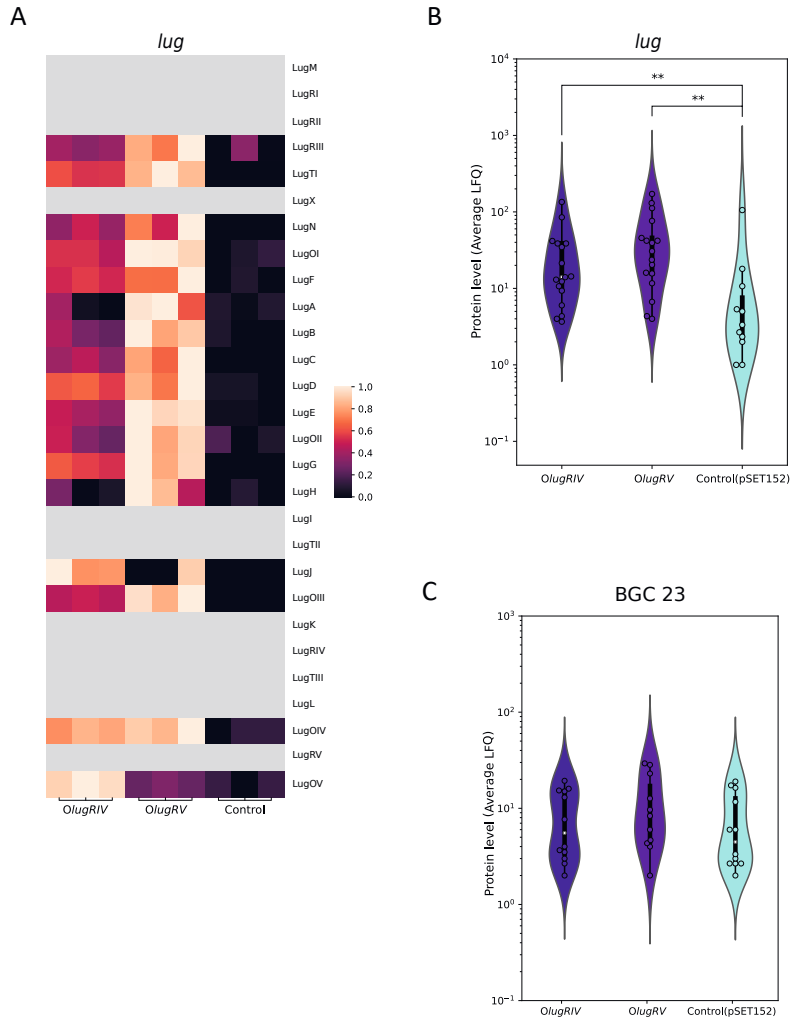

Figure 9 Statistical analysis of the proteins encoded by the gene clusters involved in lugdunomycin biosynthesis in O*lugRIV*, O*lugRV* and the control strain (QL37::pSET152) grown on MM for two days.

A) Heatmap showing the levels of the proteins encoded by the lug gene cluster in each replicate of O*lugRIV*, O*lugRV* and the control strain, normalised by row to the length of 1. The grey tiles indicate that the protein was not detected. B) The distribution of the overall expression level of the Lug proteins represented in a violin plot. The dots indicate the average expression level of each protein. ** indicates a p-value less than 0.01 (O*lugRIV* vs control: *p*-value =0.00901, O*lugRV* vs control: *p*-value =0.00190) C) The distribution of the overall expression level of the proteins from BGC 23. The expression of BGC 23 was not significantly different among the tested strains (OlugRIV vs control: *p*-value = 0.60230; O*lugRV* vs control: *p*-value = 0.21733). The p-values were calculated using the Mann-Whitney test.

lugRII mutant, and that perhaps insufficient expression of *lugRII* was achieved from plasmid O*lugRII*. The function of *lugRII* and its role in the control for angucycline biosynthesis requires further studies.

Cross-complementation experiments suggested a hierarchical order between the positive regulators LugRIV and LugRV, whereby the ARR LugRIV activates the SARP LugRV which in turn activates the structural genes of the *lug* gene cluster, resulting in the biosynthesis of the angucyclines. A similar hierarchical system wherein an ARR activates a SARP regulator that in turn activates the BGC, was observed for the undecylprodigiosin and daunorubicin BGCs in the angucycline non-producers *S. coelicolor* and *S. peucetius*, respectively (Hulst *et al.*, 2021, Fujii *et al.*, 1996, Guthrie *et al.*, 1998, Furuya & Hutchinson, 1996). The ARR-family regulators RedZ and DnrN belong to the NarL/FixJ superfamily of LuxR-type regulators, and show more similarity to LugRII, whereas LugRV belongs to the ARRs from the OmpR-type, such as JadR1, Aur1P and LanI (Wang *et al.*, 2009, Novakova *et al.*, 2005, Yushchuk *et al.*, 2019). The latter directly activate the expression of the structural genes of the angucycline BGCs governing the production of jadomycin, auricin and landomycin in *S. venezuelae*, *S. aureofaciens* and *S. cyanogenus*, respectively, without the involvement of a SARP regulator (Novakova *et al.*, 2005, Wang *et al.*, 2009, Yushchuk *et al.*, 2019). This highlights interesting regulatory differences between angucycline BGCs.

Angucyclines are toxic molecules and a mechanism to inhibit their production was expected. Various angucycline BGCs (kinamycin, jadomycin, auricin and landomycin) contain genes coding for activators and for repressors (Aigle *et al.*, 2005, Kormanec *et al.*, 2014, Zou *et al.*, 2014, Yushchuk *et al.*, 2019), whereas our experiments on the *lugR* genes only identified activatory roles. Similar results were observed in the study of the thioangucycline BGC of *Streptomyces* sp. CB00072 (Cao *et al.*, 2021). Repression of angucycline production may be executed by an autoregulatory mechanism involving the ARRs LugRII and LugRIV, since the binding of ARRs and thereby the expression of the target BGC is modulated by the binding of the intermediates and final products of the BGCs (Wang *et al.*, 2009). The function of LugRI in angucycline biosynthesis, if any, was not elucidated and needs more attention. Distant orthologues of *lugRI* (≤34% between the gene products) were found in the type II polyketide BGC of *Streptomyces* sp. CB02414, the producer of rubiginones, which belong to the angucyclines (Zhang *et al.*, 2021).

However, the role of these regulators in the control of the production of rubiginones has not been elucidated.

Overexpression of LugRIV and LugRV resulted in precocious angucycline and lugdunomycin production in the early stages of growth. A simultaneous increase in the expression of the Lug proteins in the O*lugRIV* and O*lugRV* strains relative to the control was also observed. However, prologues growth of the strains (seven days instead of two) did not result in higher levels of lugdunomycin in the overexpression mutants relative to the control strain, even though the angucyclines were still produced at a higher level. Lugdunomycin biosynthesis requires two substrates: a C-ring rearranged angucycline and *iso*-maleimycin (Wu *et al.*, 2019), which are probably produced from two different BGCs (Chapter 4). Investigation of the proteomics data revealed no statistically significant changes in the expression of the proteins encoded by the hypothesised *iso*-maleimycin BGC (BGC 23) between the control strains and either O*lugRIV* or O*lugRV*, which could explain the failure of O*lugRIV* and O*lugRV* to overproduce lugdunomycin. In MM-grown cultures more proteins encoded by the *iso*-maleimycin BGC were detected compared to R5-grown cultures, which may explain the absence of lugdunomycin in R5-grown cultures. Based on that, constitutive expression of the *iso*-maleimycin gene cluster can possibly help to increase lugdunomycin production.

In conclusion, this study revealed that the *lug* gene cluster that is required for angucycline, limamycin and lugdunomycin production is positively controlled by the three regulators LugRII, LugRIV and LugRV. A hierarchical relationship could be identified between LugRIV and LugRV, where LugRV is the ultimate cluster-situated regulator. This contrasts with the regulatory networks controlling other known angucycline BGCs, several of which lack a SARP regulator and instead fully depend on ARR-family regulators. Overexpression of these regulators led to increased angucyclines production, but not of lugdunomycin. This again underlines the complexity of the metabolic and regulatory pathways that together govern the biosynthesis of this cryptic molecule.

MATERIALS AND METHODS

Bacterial strains and growth conditions

Bacterial strains used in this study are indicated in Table 1. *Streptomyces* sp. QL37 was obtained from the soil in the Qinling mountains (P. R. China) (Zhu *et al.*, 2014b). The strain is deposited to the collection of the Centraal Bureau voor Schimmelcultures (CBS) in Utrecht, The Netherlands, under deposit number 138593. The metabolite production profile of each *Streptomyces* strain used in this study is indicated in Table 1 and Table 2. *Escherichia coli* JM109 was used for general cloning and grown on Luria Broth without antibiotics. *E. coli* ET12567/pUZ8002 was used for conjugation of plasmids to *Streptomyces* sp. QL37 (Wang & Jin, 2014, MacNeil *et al.*, 1992, MacNeil, 1988). Strains containing plasmids were selected on LB containing ampicillin (100 µg/mL), apramycin (50 µg/mL), chloramphenicol (25 µg/mL) and kanamycin (50 µg/mL). Spore stocks were prepared according to (Kieser *et al.*, 2000) from cultures of *Streptomyces* sp. QL37 grown on SFM agar for seven days at 30°C. For conjugation *Streptomyces* sp. QL37 was grown on SFM containing 60 mM $MgCl_2$ and 60 mM $CaCl_2$ (Wang & Jin, 2014, Kieser *et al.*, 2000). Exconjugants were selected using thiostrepton (20 µg/mL) and apramycin (50 µg/mL).

Construction of knock-out mutants

The in-frame deletion mutants were created according to the method previously described (Swiatek *et al.*, 2012) and described in Chapter 5. The correct mutants were selected on SFM containing either apramycin (50 µg/mL) or thiostrepton (10 µg/mL), where they should only grow in the presence of apramycin. In order to remove the apramycin cassette by recombination of the *lox*P sites the pUWL-Cre construct was conjugated to the strains. The removal of the apramycin cassette was verified by PCR with the primers given in Table 2 and Sanger sequencing of the PCR products (Swiatek *et al.*, 2012).

Generation of strains for enhanced gene expression of the *lugR* genes

For constitutive expression of the *lugR* genes, the genes were amplified from the genome using the primers indicated in Table 3 and cloned downstream of the *ermE** promoter. For this, the genes *lugRIII* and *lugRV* were first sub-cloned in the pHM10a vector using the restriction enzymes NdeI and HindIIII. The constructs were then digested with BglII and EcoRI, resulting in a fragment that contained the

*ermE** promoter, the gene, and a part of the pHM10a plasmid. This fragment was subsequently cloned into pSET152 digested with BamHI and EcoRI. A three-way ligation including the *ermE** promoter, the plasmid pSET152, and the gene resulted in constructs with *lugRII* or *lugRIV* behind the *ermE** promoter in pSET152. The integrity of the constructs was verified using Sanger sequencing and restrictions enzyme analysis. The constructs were isolated from JM109 and transformed in the methylase deficient strain *E. coli* ET12567/pUZ8002 for conjugation to *Streptomyces* sp. QL37 (MacNeil *et al.*, 1992).

Table 1 Bacterial strains used in this work

Strain	Genotype	Reference
***Streptomyces* sp. QL37**	Wild type lugdunomycin producer	(Zhu *et al.*, 2014b, Wu *et al.*, 2019, Wu, 2016)
***E. coli* JM109**	*end*A1, *rec*A1, *gyr*A96, *thi, hsd*R17 (r_k^-, m_k^+), *rel*A1, *sup*E44, λ–, Δ(*lac-pro*AB), [F´, *tra*D36, *pro*AB, *lac*I^{q}ZΔM15], IDE3.	(Sambrook J., 1989)
***E. coli* ET12567::pUZ8002**	Methylation-deficient host (*dam*$^-$ *dcm*$^-$*hsdM*$^-$) for intergeneric conjugation. Contains the nontransmissible, *oriT*$^-$ mobilizing plasmid pUZ8002.	(MacNeil *et al.*, 1992)
QL37Δ*lug-pks*	*Streptomyces* sp. QL37 Δ*pks*::apra^R	(Wu *et al.*, 2019)
QL37Δ*lugRII*	*Streptomyces* sp. QL37 Δ*lugRII*	This work
QL37Δ*lugRIII*	*Streptomyces* sp. QL37 Δ*lugRIII*	This work
QL37Δ*lugRIV*	*Streptomyces* sp. QL37Δ*lugRIV*	This work
QL37Δ*lugRV*	*Streptomyces* sp. QL37Δ*lugRV*	This work
WT::pSET152	*Streptomyces* sp. QL37 carrying pSET152	
OE-*lugRII*	*Streptomyces* sp. QL37::pSET152-*ermE* -lugRII*	This work
OE-*lugRIII*	*Streptomyces* sp. QL37::pSET152-*ermE* -lugRIII*	This work
OE-*lugRIV*	*Streptomyces* sp. QL37::pSET152-*ermE* -lugRIV*	This work
OE-*lugRV*	*Streptomyces* sp. QL37::pSET152-*ermE*-lugRV*	This work

Table 2 Oligonucleotides used in this study for the creation of knock-out constructs

Gene	Application	Sequence (5' -> 3') ^	Position #	Gene length (nt)
lugRII	Left flank	GTCAGAATTCCTTCAGCGATGCCGGATAAC	+19/-694	669
		GAAGTTATCCATCACCTCTAGACAGGGATGCTCTACGAGATG		
	Right flank	GAAGTTATCGCGCATCTCTAGAACCGTTGTCGACCTGCATGAATC		
		GTCAAAGCTTGGCGGTAGGATGAGTTGGTGACTG		
	Mutant verification	TGACCGAGGTTGCCGAAACAGG		
		CGACTTCACGGGCCGCCTGAAC		
lugRIII	Left flank	GTCAGAATTCGCGTAGGTGGCGAGTCCGACAAG	+63/-794	717
		GTCATCTAGAGTTGCACAACGCCGAGCGAGG		
	Right flank	GTCATCTAGAGTCCGTCGAGACGGCACTGAC		
		GTCAAAGCTTCGGAGTGCAGGTGCCGTTTCAC		
	Mutant verification	AGCAGGACCACACCGATCAG		
		TGACCGGCTCGCCAAGATTTC		

Table 2 Oligonucleotides used in this study for the creation of knock-out constructs *(continued)*

Gene	Application	Sequence (5' -> 3') ^	Position #	Gene length (nt)
lugRIV	Left flank	GTCAGAATTCGTCCTCAACTACGCGGAGTG	+24/-698	801
		GTCATCTAGAATGGGTATGGGCAGGCCAATC		
	Right flank	GTCATCTAGATAAGCTCGGTTCGAGCCAGTG		
		GTCAAAGCTTGCTACGACAAGCGCAAGCTG		
	Mutant verification	GCCGTCTACAAGACCTGGTTCC		
		CCGATCCGTACGACGGACTGC		
lugRV	Left flank	GTCAGAATTCACCTTGCCGCCCAGTGAATG	-18/-1917	1939
		GAAGTTATCCATCACCTCTAGACCGTGCGGTTCATTCCAGTAG		
	Right flank	GAAGTTATCGCGCATCTCTAGAGAGAACGGCGAAGTGCATGAG		
		GTCAAAGCTTGGACGATCTCGTCGCTGATG		
	Mutant verification	ACTTCCATGTCAGCCTGGACG		
		GATCTCCTCGACGACCCGGTC		

position relative to the translational start site (+1) of the respective genes.
^ restriction sites underlined. GAATTC, EcoRI; AAGCTT, HindIII; TCTAGA, XbaI;

Table 3 Oligonucleotides used in this study for the creation of overexpression constructs

Gene	Sequence (5' -> 3') ^	Position #	Gene length (nt)
lugRII	CGATAAGCTTTCCCTGGCCACGCCTTACGC	+1/-700	669
	GCGCCCATCCATATGGTTTTCGCAG		
lugRIII	CGATAAGCTTTCGTCCGGGTGGCGGTCATC	+1/-729	717
	CGATGAATTCATATGACTGAAGCCAGAACGTCCGAGG		
lugRIV	CGATTCTAGAGAGGTCAGGGACAGCGCATGG	+1/-805	801
	CGATGGATCCATATGTCGCACGACCTCGATCTG		
lugRV	CGATAAGCTTGGATGAGCCGGCATGAGATGAACCG	+1/-2009	1938
	CGATGAATTCATATGCACTTCGCCGTTCTCGGCCC		

^ restriction sites underlined. GGATCC, BamHI; GAATTC, EcoRI; AAGCTT, HindIII; CATATG, NdeI; TCTAGA, XbaI.
# position relative to the translational start site (+1) of the respective genes.

Table 4 Plasmids and constructs used in this study

Plasmid	Description	Reference
pWHM3-oriT.	*E. coli*/*Streptomyces* shuttle vector, high-copy number in *E. coli.* For conjugative transfer *oriT* was inserted into unique NdeI site. Amp^R, $Thio^R$.	(Wu *et al.*, 2019, Vara *et al.*, 1989)
pSET152	*E. coli*/*Streptomyces* shuttle vector carrying *attP* site and integrase gene of ϕC31 phage for stable integration into chromosomal *attB* site of *Streptomyces*, $Apra^R$	(Bierman *et al.*, 1992)
pHM10a	*E. coli*/*Streptomyces* shuttle vector, harbouring the *ermE** promoter and engineered ribosome binding site (RBS), Amp^R, Hyg^R	(Motamedi *et al.*, 1995)
pUWL-Cre	Cre-recombinase expression plasmid, $Thio^R$	(Khodakaramian *et al.*, 2006)
pOE-*lugRII*	pSET152 with coding region of *lugRII* under the control of the *ermE** promoter	This study
pOE-*lugRIII*	pSET152 with coding region of *lugRIII* downstream *ermE** promoter	This study
pOE-*lugRIV*	pSET152 with coding region of *lugRIV* downstream *ermE** promoter	This study
pOE-*lugRV*	pSET152 with coding region of *lugRV* downstream *ermE**promoter	This study
pΔ*lugRII*	pWHM3 containing flanking regions of *lugRII* interspaced with the *apra*^{R}-loxP cassette	This study

Table 4 Plasmids and constructs used in this study *(continued)*

Plasmid	Description	Reference
pΔ*lugRIII*	pWHM3 containing flanking regions of *lugRIII* interspaced with the *apra*^{R}-loxP cassette	This study
pΔ*lugRIV*	pWHM3 containing flanking regions of *lugRIV* interspaced with the *apra*^{R}-loxP cassette	This study
pΔ*lugRV*	pWHM3 containing flanking regions of *lugRV* interspaced with the *apra*^{R}-loxP cassette	This study

Natural products extraction

Spores of the mutants and the wild type strains were confluently grown on minimal medium (MM) agar plates (25 mL) containing 0.5% mannitol and 1% glycerol and on R5 Difco containing 1% mannitol and 0.8% peptone for seven days at 30 °C (Wu *et al.*, 2019). Agar added to MM was Iberian agar derived from TM Duche & Sons Ltd (batch from 2017). The same agars as described in Chapter 3 were used for the preparation of the media. Spores of the strains for enhanced expression of the *lugR* genes and the control strain (QL37::pSET152) were grown for two days on agar plates (MM and R5) covered with cellophane. Natural products were extracted from the agar and the cellophane. Before extraction, the mycelium was removed from the cellophane and used for proteomics. In addition, the spores for enhanced expression of the *lugR* genes were grown for seven days on MM agar plates. Natural products were extracted from both the spores and the agar. Natural products extractions were further executed as described in Chapter 5, with only one exception; evaporation of organic solvents was executed at room temperature.

LC-MS methods

Mutants of the regulatory genes, the wild wild-type strain and strains for enhanced expression (OlugRIV and OlugRV and control strain) grown on MM for seven days.

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 extracts 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 4 min. The column was re-equilibrated to 5% B for 3 min before the next run was started.

All the samples were analysed in positive polarity, using data dependent acquisition mode. In this regard, full scan MS spectra (m/z 100 – 2000, scan rate 20 Hz) were followed by three data dependent MS/MS spectra (m/z 100 – 2000, scan rate 20 Hz) for the three most intense ions per scan. The ions were selected when they reach an intensity threshold of 1000, isolated at the tuning file Q1 resolution, fragmented using collision induced dissociation (CID) with collision energy ramp (CE 20–50 eV), and excluded for 0.05 s (one MS scan) 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.

LC-MS analysis for the strains with enhanced expression of the regulatory genes (metabolomics combined with proteomics)

The natural product extracts for combined metabolomics and proteomics experiment were analysed with a similar method as described above. The LC-method was identical to the method described for the analysis of the extracts of the mutants, only a few changes were made in the MS acquisition method.

All the 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.

Comparative Metabolomics

Before statistical analysis the data obtained from the LC-MS-runs were processed using MZmine version 2.53 (Pluskal *et al.*, 2010). The data derived from MM-grown cultures and R5- grown cultures were processed separately. Data processing in MZmine for statistical analysis in MetaboAnalyst was executed as in Chapter 4 (Chong *et al.*, 2019)(See Pre-processing LC-MS data in MZmine 2.53) with a few exceptions. With the detection of the mass ion peaks only mass ion peaks at an MS level 1 were detected. Thus, in the chromatogram deconvolution step no additional MS2 scan pairing was performed. Furthermore, within the modules isotopic peak grouper, join aligner and gap-filling an RT tolerance of 0.1 min was applied, instead of 0.05 min for cultures grown on MM. The final step was removing mass features detected before 1 min and after 10 min with peak list row filter (RT 1.0–10.0). Peaks with and without MS/MS data were kept. The data were exported to MetaboAnalyst file and were filtered as described in Chapter 4.

Statistical analysis was performed using MetaboAnalyst V4.0 (Chong *et al.*, 2019) as described in Chapter 5.

GNPS molecular networking

Pre-processing LC-MS data in MZmine 2.53 and MZmine 2.37.1 corr17.7 and data filtering

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 (Phelan, 2020, Xie *et al.*, 2020). Data processing in MZmine for statistical analysis in MetaboAnalyst was executed as in Chapter 4, with a few exceptions: within the modules isotopic peak grouper, join aligner and gap-filling an RT tolerance of 0.1 min was applied, instead of 0.05 min for cultures grown on MM. Data filtering executed as in Chapter 4

Feature-based ion identity molecular networks

The data were submitted to the GNPS web tool and a network was generated using FBMN. The same filtration method was used as described in Chapter 4, with one exception: For R5-grown cultures the edges were filtered to have a cosine score

above 0.8. The molecular network was visualised using the software Cytoscape (Kohl *et al.*, 2011).

Proteomics

Spores were confluently grown on MM and R5 agar plates covered with cellophane. After two days of growth at 30 °C the mycelium was scraped off the plate and subsequently transferred to eppendorfs containing metal beads. These were subsequently frozen in liquid nitrogen. The cells were disrupted using a pre-cooled TissueLyser (Qiagen, the Netherlands) for two cycles of 30 s at 30 Hz. The disrupted cells were subsequently treated with a lysis buffer (4% SDS, 0.06 M DTT, 100 mM Tris-HCl pH 7.6, 50 mM EDTA) in order to dissolve the proteins. The proteins were precipitated using chloroform-methanol (Wessel & Flugge, 1984). The precipitated proteins were dissolved in 0.1% RapiGest SF surfactant (Waters, USA) at 95 °C. Proteins were digested with trypsin according to Rooden *et.al.* (van Rooden *et al.*, 2018). Protein digestion was followed by addition of trifluoroacetic acid for the complete removal and degradation of RapiGest SF. Using the STAGE-Tipping technique peptide solutions containing 8 µg of peptide were cleaned and desalted (Rappsilber *et al.*, 2007, Gubbens *et al.*, 2014). Protein concentrations were determined with the BCA (bicinchoninic acid assay) assay (Walker, 1994). Final peptide concentration was adjusted to 100 ng/µL using sample solution (3% acetonitrile, 0.5% formic acid) before analysis.

UPLC and mass spectrometry measurement and data processing for proteomics

200 ng (2 µL) of the digested peptide was injected and analysed by reversed-phase liquid chromatography on a nanoAcquity UPLC system (Waters, Massachusetts, U.S.) equipped with HSS-T3 C_{18} 1.8 µm, 75 µm X 250 mm column (Waters, Massachusetts, U.S.). A gradient from 1% to 40% acetonitrile in 110 min (ending with a brief regeneration step to 90% for 3 min) was applied. [Glu^1]-fibrinopeptide B was used as lock mass compound and sampled every 30 s. Online MS/MS analysis was done using Synapt G2-Si HDMS mass spectrometer (Waters, Massachusetts, U.S.) with an $UDMS^E$ method set up as described in (Distler *et al.*, 2014).

Raw data from all samples were first analysed using the vendor software ProteinLynx Global SERVER (PLGS) version 3.0.3. Generally, mass spectrum data were generated using an MS^E processing parameter with charge 2 lock

mass 785.8426, and default energy thresholds. For protein identification, default workflow parameters except an additional acetyl in N-terminal variable modification were used. A reference protein database was made from predicted 7476 coding sequences from the NCBI database (NZ_PTJS00000000). The resulted dataset was imported to ISOQuant (version 1.8, Distler *et al.*, 2014) for label-free quantification. ISOQuant processing includes spectrum alignment, data normalization at spectrum level, peptide quantification by top3 method and subsequent protein quantification. Default high identification parameters were used in the quantification process.

Of the 7476 proteins from the database, a total of 2104 proteins were identified across all the samples in MM-derived protein extracts. For each sample, on average 1069 proteins were identified. In R5-derived protein extracts 1306 proteins were identified across all samples. For each sample, on average 662 proteins were identified.

Statistical analysis of the proteomics data

Heatmaps were generated using Matplotlib package (version 3.5.0) (Hunter, 2007). Violin plots were generated using Seaborn package (version 0.11.2) (Waskom, 2021). Mann-Witney tests were performed using "mannwhitneyu" method in Scipy package (version 1.8.0) (Virtanen *et al.*, 2020).

SUPPLEMENTARY TABLES

Table S1 Shared amino acid identity of LugR proteins (in %) with their orthologues encoded by *lug*-type gene clusters

Strain/Protein	LugRII	LugRIII	LugRIV	LugRV
S. sp. CB00072	53.85	64.86	64.73	58.41
S. scopuliridis	52.75	66.22	66.82	59.72
S. sp. b84	53.85	65.32	61.13	58.57
S. sp. w007	53.85	65.77	61.13	55.53

SUPPLEMENTARY FIGURES

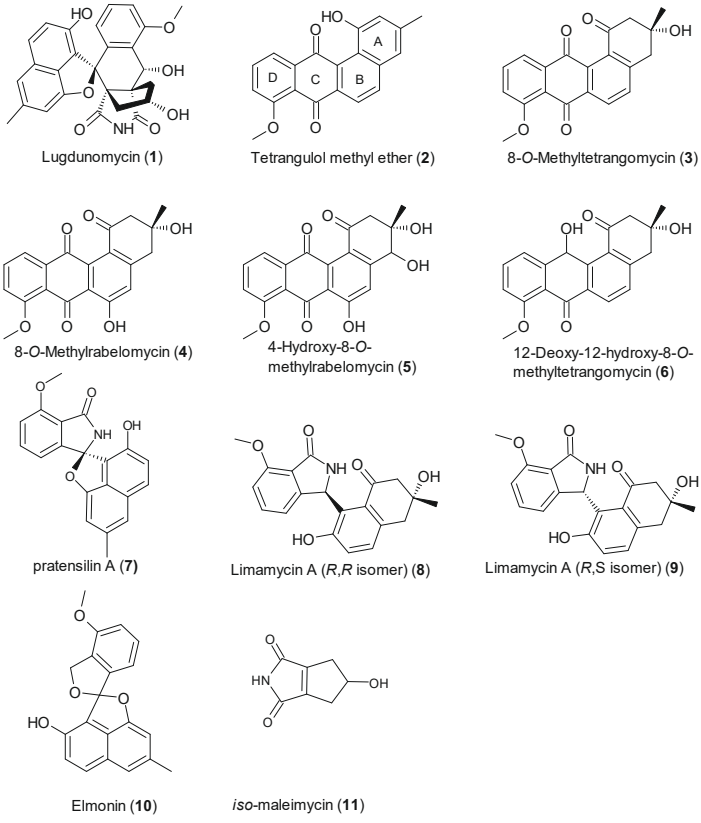

Figure S1 Structures of the compounds discussed in this study
Structures of molecules produced by *Streptomyces* sp. QL37 (**1**–**10)**. Lugdunomycin (**1**), the non-rearranged angucyclines (**2**–**6**) and the rearranged angucyclines (**7**–**9**) were previously isolated from *Streptomyces* sp. QL37 (Wu *et al.*, 2019). Elmonin (**10**) was detected in the extracts of *Streptomyces* sp. QL37 by comparison with a standard (Chapter 4)

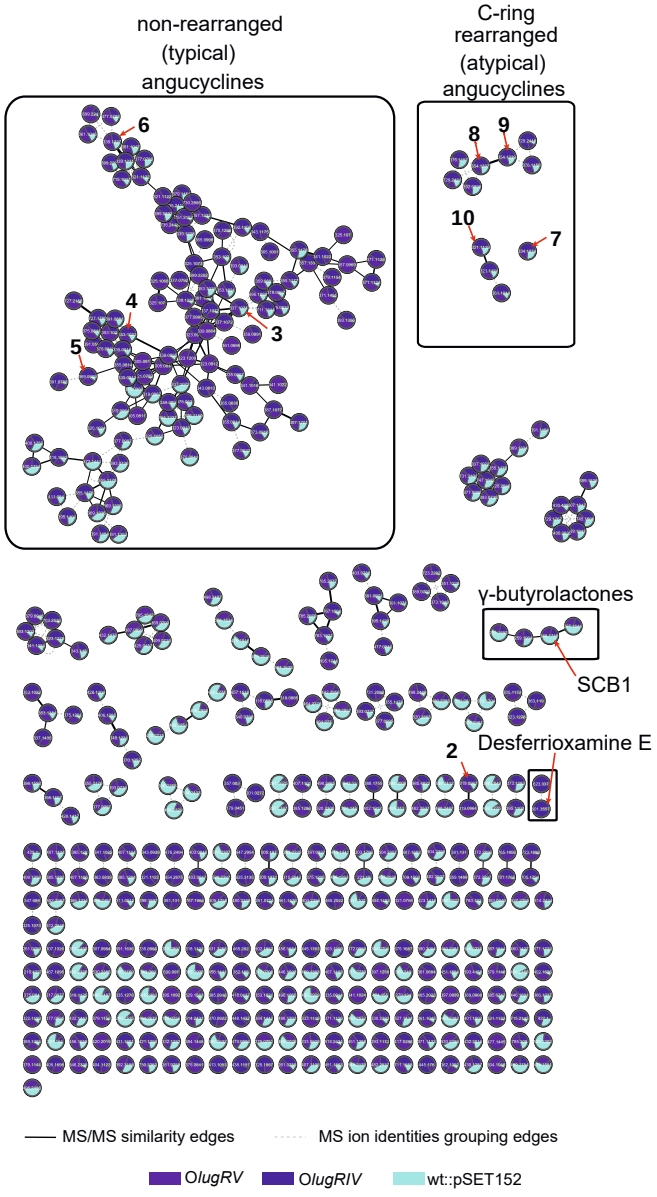

Figure S2 Molecular network of the ions detected in the extracts of *Streptomyces* sp. QL37 O*lugRIV*, O*lugRV* and the control strain grown on R5.
Molecular network of the ions detected in extracts of *Streptomyces*. sp. QL37 overexpressing *lugRIV* and *lugRV* and the control strain grown on R5. The 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. Arrows highlight metabolites that were previously identified in *Streptomyces* sp. QL37. Molecular families that could be dereplicated are indicated.

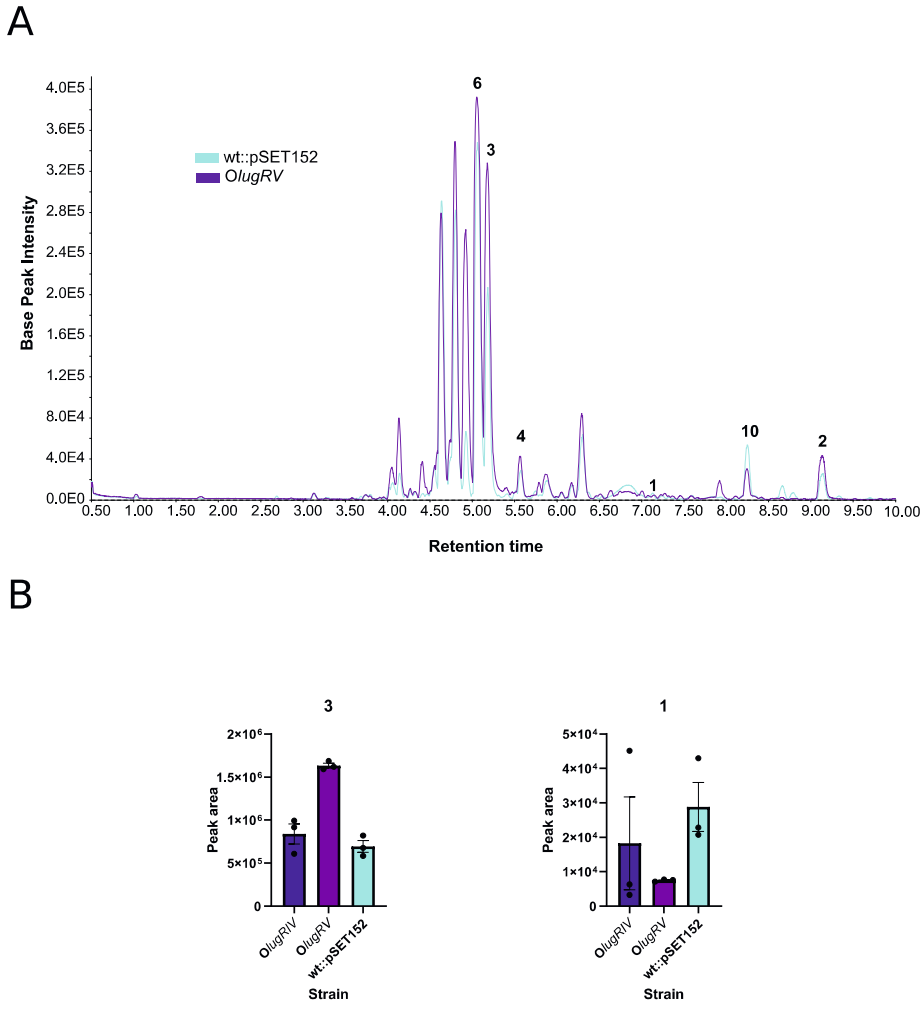

Figure S3 Production of angucyclines and lugdunomycin by of *Streptomyces* sp. QL37 overexpression *lugRIV* and *lugRV* after seven days of growth at 30 °C.
A) Overlay of LC-MS chromatograms of the control strain and O*lugRV*. B) Bar plots representing the peak areas of 8-*O*-methytetrangomycin (3) and lugdunomycin (1) with the error bar indicating the standard error of the mean (SEM).

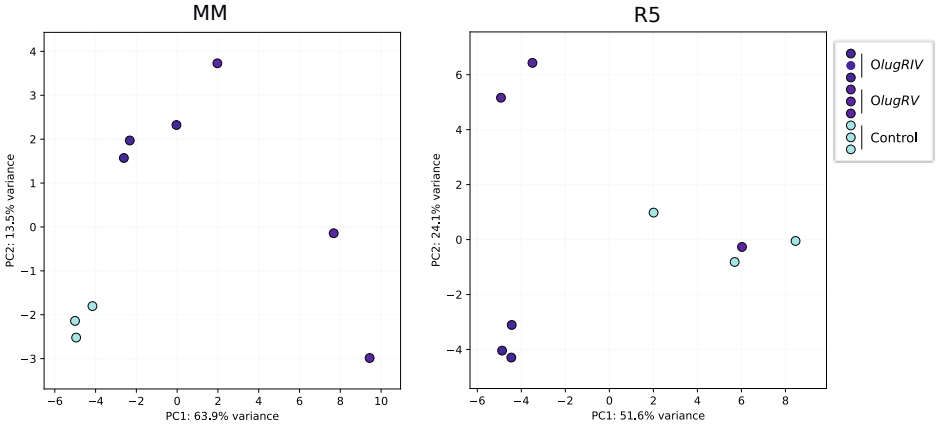

Figure S4 Exploratory statistical analysis of the proteome of *Streptomyces* sp. QL37 with enhanced expression of *lugRIV*, *lugRV* and the control strain.
PCA score plots for the extracts of the strains grown on MM (A) and R5 (B) for two days at 30 °C.

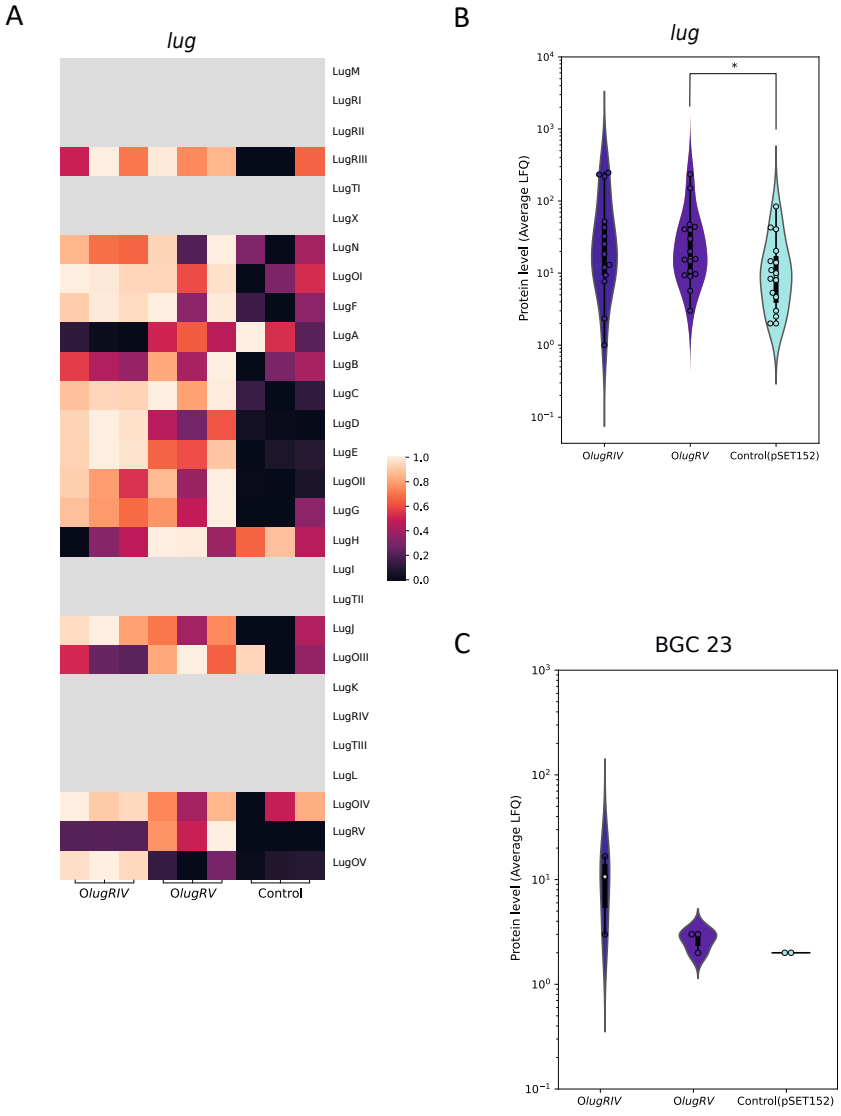

Figure S5 Statistical analysis of the proteins encoded by the gene clusters involved in lugdunomycin biosynthesis in O*lugRIV*, O*lugRV* and the control strain (QL37::pSET152). The strains were grown on R5 for two days at 30 °C. A) Heatmap shows the protein level in each replicate of O*lugRIV*, O*lugRV* and the control strain, normalised by row to the length of 1. Grey tiles indicate that the protein was not detected. B) Distribution of the overall expression level of the Lug proteins represented in a violin plot. Dots indicate the average expression level of each Lug protein; asterisk (*) indicates that the *p*-value was below 0.05 (O*lugRIV* vs control: *p*-value = 0.08372 and O*lugRV* vs control = 0.04749) C) Distribution of the overall expression level of the proteins from BGC 23. The expression of BGC 23 was not significantly different among the strains, and only three proteins were detected (O*lugRIV* vs control: *p*-value = 0.13864; O*lugRV* vs control: *p*-value = 0.31731). The *p*-values were calculated using a Mann-Witney test.

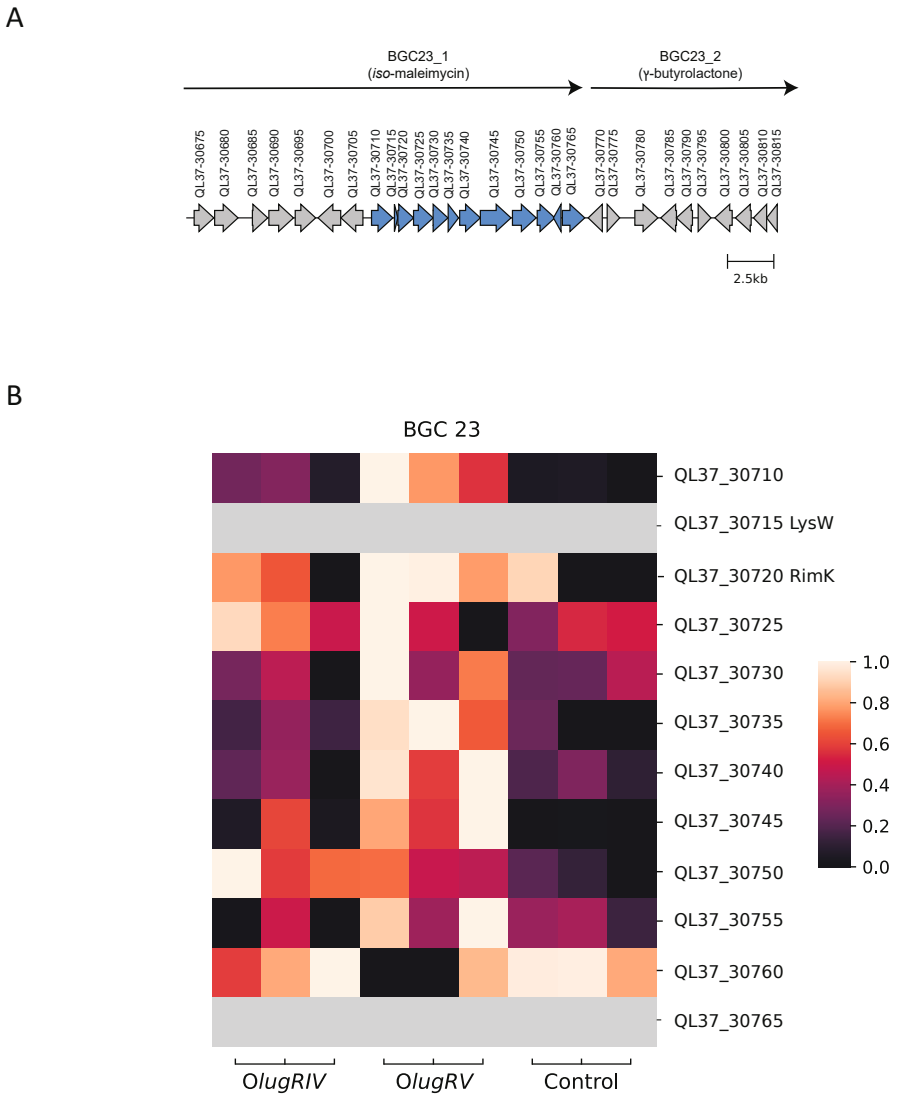

Figure S6 Organization of BGC 23 and the level of proteins encoded by BGC 23 represented in a heatmap.

A) Organization of the predicted *iso*-maleimycin BGC. Blue colours indicate the core biosynthetic genes (QL37_30710-QL37_30755) that are putatively required for the synthesis of *iso*-maleimycin (**11**), shared with the maleimycin BGC and one predicted regulator (QL37_30760) and one predicted transporter gene QL37_30765. For more details see Chapter 4. B) Heatmap shows the protein level of each protein of BGC 23 indicated in blue (A) in each replicate of *OlugRIV*, *OlugRV* and the control strain. Strains were grown on MM for two days at 30 °C. Levels were normalised by row to the length of 1. Grey tiles indicate that the protein was not detected.