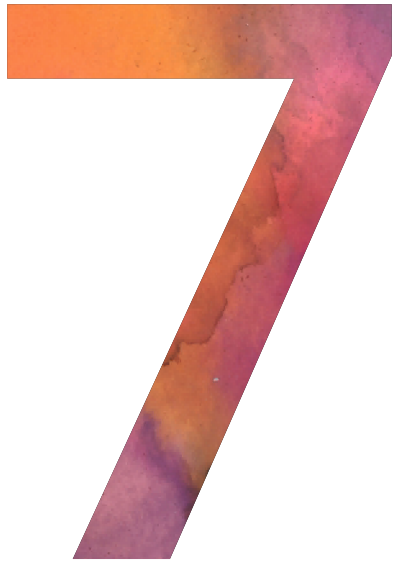
Analysis of the angucycline biosynthetic gene cluster in Streptomyces sp. QL37 and implications for lugdunomycin production

Heul, H.U. van der



General Discussion

GENERAL DISCUSSION

The increase in resistance phenomena, such as antibiotic resistant pathogenic bacteria and drug resistant tumour cells, demands the search for novel drug leads (Wright, 2017, Huang *et al.*, 2021). *Streptomyces* species are a rich source of bioactive natural products with great chemical diversity (Baltz, 2008). These filamentous soil bacteria are responsible for the production of almost half of all antibiotics used in the clinic as well as natural products applied for other medical, biotechnological and agricultural purposes (Berdy, 2005, Hopwood, 2007). Advancements in genome mining strategies revealed that genomes of streptomycetes contain many more biosynthetic gene clusters (BGCs) for natural products than initially thought, many of which are not actively transcribed under routine screening conditions (Bentley *et al.*, 2002, Medema *et al.*, 2011) This offers the perspective of a vast chemical space that may still await discovery. Approaches to activate these BGCs are discussed in Chapter 2 (van der Heul *et al.*, 2018).

One group of compounds that is well known for their antibacterial and anticancer activities are the angucyclines. The biosynthesis pathways of the typical angucyclines have been studied extensively, such as those from gaudimycin, urdamycin and landomycin (Kharel *et al.*, 2012). The initial angucycline framework, characterised by its benz[a]anthracene structure is constructed by the minimal polyketide synthase (PKS) followed by modifications by post-PKS enzymes, leading to a wide array of interesting chemical scaffolds (Kharel *et al.*, 2012). Particularly the opening of one of the aromatic rings offers a drastic structural diversification. This has been demonstrated by the discovery of the jadomycins, kinamycins and gilvocarcins, which are derived from a typical angucycline intermediate of which the B-ring has been opened by a monooxygenase (Fan & Zhang, 2018).

The discovery of the angucycline-derived molecule lugdunomycin shows that even in a well-known compound family a novel structure can be found. This molecule, produced by *Streptomyces* sp. QL37, is characterised by its unique benzaza[4,3,3] propellane-6-spiro-2'-2H-naphtho[1,8-bc]furan backbone and represents a novel subclass of polyketides (Wu *et al.*, 2019). The hypothesised biosynthesis pathway includes a Baeyer-Villiger oxidative cleavage, expansion of the C-ring followed by a Diels-Alder reaction between a C-ring rearranged angucycline as a diene and iso-

maleimycin as a dienophile. Although many different C-ring cleaved angucyclines have been discovered since the identification of lugdunomycin, the reaction mechanism for the fragmentation of the angucycline quinone moiety (the C-ring) is still unknown Thus, this molecule offers great possibilities for the discovery of novel organic chemistry and enzymology that may be applied in combinatorial (bio)synthesis of new natural products or the discovery of new BGCs. (Mikhaylov *et al.*, 2021). Lugdunomycin has antibiotic activity against Gram-positive bacteria (Wu *et al.*, 2019). The molecule is produced in very low quantities by the natural host; 0.5 mg of compound was obtained from 7.5 L of medium, which limits the opportunity to do pharmacological and pharmacokinetic tests (Wu *et al.*, 2019).

In this thesis a fundamental approach was applied to obtain knowledge on the biosynthetic reactions and the control of lugdunomycin biosynthesis, using a combination of bioinformatic analysis, gene deletion experiments, heterologous expression, metabolomics, proteomics, and RNA-seq. These approaches led to increased understanding of the enzymes encoded by the *lug* gene cluster, its regulation and the biosynthesis of lugdunomycin. One of the foreseen applications is to improve the yield of lugdunomycin and expand the angucyclines chemical space.

Characterisation of the *lug* gene cluster

Angucycline production is directed by a type II polyketide synthase (PKS). Analysis of the *Streptomyces* sp. QL37 genome using antiSMASH (Medema *et al.*, 2011) identified a polyketide type II gene cluster (designated *lug*) with high similarity to the BGC of kiamycin, an angucycline derivative produced by *Streptomyces* sp. W007 (Zhang *et al.*, 2012, Cao *et al.*, 2021, Kibret *et al.*, 2018). The genes *lugA–OII*, covering the genes *lugA–C* which encode the so-called minimal PKS enzymes required for the production of the angucycline backbone (KSα, KSβ and ACP), were deleted and the metabolomics profile was compared with that of the wild-type strain (Wu *et al.*, 2019). This *lug-pks* mutant failed to produce any rearranged, non-rearranged angucyclines and lugdunomycin, which confirmed that this gene cluster is indeed required for the production of these molecules (**Chapter 3**).

RNA-seq for transcriptome profiling was conducted on RNA isolated from *Streptomyces* sp. QL37 grown for different incubation times (24, 36, 48 and 60 h) on lugdunomycin production medium. This revealed that the genes *lugRI–lugOV*

showed a similar expression pattern, suggesting that these genes are controlled by the same regulatory mechanism and are thus part of one gene cluster. Thus, according to these data, the *lug* gene cluster comprises 27 genes, including amongst other genes encoding for the minimal PKS, five oxygenase genes and five regulatory ones.

To further verify the extent of the *lug* gene cluster and to find other possible lugdunomycin producers, a collection of over 1000 genomes of *Streptomyces* and *Kitasatospora* strains was searched for the presence of angucycline BGCs. Interestingly, some 25% of these strains contained the minimal PKS, that is characteristic for the angucycline BGCs, suggesting that angucycline production is widespread. We identified 36 strains (incl. *Streptomyces* sp. QL37) that had angucycline BGCs, and these shared at least 18 genes with the *lug* gene cluster. Comparison of the homologous BGCs with the *lug* gene cluster, revealed that it comprises 28 genes covering *lugM–lugOV,* which is close to what was predicted using RNA-seq data. Thus, based on the bioinformatics data in combination with the RNA-seq data we conclude that the *lug* gene cluster comprises 28 genes. Possibly the expression of *lugM* is controlled differently compared to *lugRI–lugOV* resulting in different expression levels.

Among the 35 angucycline BGCs, only one contained an orthologue of the *lugRI* regulator which was in the genome of *Streptomyces* sp. 94 (Hulcr *et al.*, 2011). On the other hand, all of them lacked homologues of *lugX* and *lugK*. Not a single strain had a homologue of *lugX* even outside its identified angucycline BGC, and this gene was not found anywhere in the NCBI database; therefore, it is unclear whether this encodes a functional protein and if so, what its function is. Transcripts of *lugX* were detected, strongly suggesting that it is not a pseudogene. Homologues of *lugK* were found elsewhere in the genomes of 28 out of the 35 strains. *lugK* encodes a phosphopantetheinyl transferase (PPTase), which can be used in multiple pathways (Bunet *et al.*, 2014)

A phylogenetic tree was generated based on housekeeping genes from 1020 *Streptomyces* and *Kitasatospora* strains. This revealed that strains containing a *lug*-type gene cluster were spread all over the phylogenetic tree and thus the *lug* gene cluster is not correlated to phylogeny. Strains that were closely related to

Streptomyces sp. QL37, such as *Streptomyces* sp. LamerLS-316 did not contain the *lug* gene cluster.

Comparison of the *lug* gene cluster with 27 well-characterised angucycline BGCs that direct the production of either non-rearranged, A-ring or B-ring cleaved angucyclines revealed that in total eight genes were specific to the *lug*(-type) gene clusters or those that direct the production of C-ring cleaved molecules. These included the regulatory genes *lugRI–lugRIII* and *lugRV*, the oxygenase genes *lugOIII* and *lugOIV* and the transporter gene *lugTI* (**Chapter 3**). The genes *lugOIII* and *lugOV* are most likely required for C-ring cleavage (see below). Although genes could be assigned that differentiate *lug*-type gene clusters from non-rearranged and B-ring cleaved angucycline BGCs, the question remained why the *lug* gene cluster is so special that it can direct the production of such as complex molecule. Heterologous expression of the *lug* gene cluster and flanking regions in the engineered host *Streptomyces coelicolor* M1152 resulted in the production of rearranged and non-rearranged angucyclines, but not lugdunomycin (**Chapter 4**). This further indicated that the *lug* gene cluster is truly an angucycline gene cluster directing the production of non-rearranged and C-ring cleaved angucyclines. However, which other genes are then required for lugdunomycin biosynthesis?

An additional BGC is required for the final steps of lugdunomycin biosynthesis

Lugdunomycin production requires two substrates: a C-ring cleaved rearranged angucycline and *iso*-maleimycin. We previously hypothesised that *iso*-maleimycin was derived from the limamycins (Wu *et al.*, 2019). However, we found that *iso*-maleimycin was not only observed in the wild-type, but also in the extracts of the *lug-pks* mutant (Uiterweerd, 2020). This suggested that *iso*-maleimycin is not derived from the limamycins, as previously hypothesised (Wu *et al.*, 2019). The temporal RNA-seq data revealed that the *lug* gene cluster was co-expressed with amongst others BGC 23. This BGC contains both β-lactone related genes and amino group carrier protein related genes (BGC 23a), together with γ-butyrolactone related genes (BGC23b) (Wolf *et al.*, 2017a, Robinson *et al.*, 2019, Matsuda *et al.*, 2017). Suggestively, an amino group carrier protein was proposed to be involved in maleimycin biosynthesis, a molecule produced by *S. showdoensis* ATCC 15227 (Makato, 2012-2017, Prima *et al.*, 2017, Matsuda *et al.*, 2017, Elstner *et al.*, 1973). Indeed, BGC23 and the maleimycin BGC are highly similar, and we therefore

propose that this cluster is involved in the production of *iso*-maleimycin, one of the substrates for the final reaction that generate lugdunomycin. The lack of a BGC similar to BGC23 in *S. coelicolor* explains why the heterologous host *S. coelicolor* M1152 carrying the *lug* gene cluster did not produce lugdunomycin (**Chapter 4**).

The role of the oxygenase genes of the *lug* gene cluster in lugdunomycin biosynthesis

One of the key reactions in the production of lugdunomycin is the cleavage of the C-ring in the angucycline, which is required to generate rearranged angucyclines as well as lugdunomycin, and is likely catalysed by a Baeyer–Villiger oxidation (Wu *et al.*, 2019). The *lug* gene cluster encodes five putative oxygenases, namely *lugOI–lugOV*, and their functional role in lugdunomycin production was studied by deletion experiments and metabolomic studies (**Chapter 5**).

The presence of a range of oxygenase genes with different functionalities in angucyclines BGCs is important for the diversification of the angucyclines that are biosynthesised (Fan & Zhang, 2018). Orthologues of LugOI and LugOII, such as PgaE and PgaM are important for the diversification of non-rearranged angucyclines (Patrikainen *et al.*, 2012, Fan & Zhang, 2018). Deletion of *lugOI* or *lugOII* resulted in mutants that had lost the ability to produce most of the unrearranged angucyclines and were incapable of producing the rearranged angucyclines and lugdunomycin. Thus, these genes are likely involved in early post-PKS modifications. These results are well in line with the known role of the homologues of LugOI (PgaE, UrdE, LanE) and LugOII (PgaM, UrdM, LanV), which are involved in the biosynthesis of gaudimycin, urdamycin and landomycin, respectively (Kharel & Rohr, 2012, Mayer *et al.*, 2005). Functional analysis showed that LugOII catalyses C6 ketoreduction, similar to other LugOII-like enzymes. Interestingly, previous *in vitro* experiments also revealed that LugOII can perform C1 ketoreduction, a reaction that has not been reported before for LugOII homologues (Xiao *et al.*, 2020). Deletion of *lugOIV* did not significantly affect angucycline biosynthesis, and its role in lugdunomycin biosynthesis, if any, so far remains elusive.

LugOIII and LugOV showed low similarity to antibiotic monooxygenases. The *lugOIII* and the *lugOV* mutants produced canonical angucyclines, but failed to produce the rearranged ones. These results suggest that these enzymes are candidates for the cleavage of the C-ring. Interestingly, analysis of the molecular

network of the *lugO* mutants revealed a likely hierarchical role between LugOIII and LugOV, whereby LugOIII catalyses the production of an angucycline epoxide that is used by LugOV for the subsequent Baeyer–Villiger oxidation reaction, resulting in the C-ring cleavage of angucyclines. These results are in line with the observation that the production of C-ring cleaved angucyclines governed by the *tac* gene cluster of *Streptomyces* sp. CB00072 requires TacS and TacT, which are orthologues of LugOIII and LugOV, respectively (Cao *et al.*, 2021). Altogether this study lays groundwork for the generation of novel C-ring cleaved rearranged angucyclines using combinatorial biosynthesis. *In vitro* assays using the purified oxygenases should further give insight into the functional role of the oxygenases encoded by the *lug* gene cluster.

Characterisation of the regulatory genes of the *lug* gene cluster

Streptomycetes harbour up to 1000 different transcription factors, forming a hugely complex regulatory circuitry that allows them to respond to the cacophony of signals they receive from the environment, so as to formulate appropriate responses, including decisions when to initiate development and specialised metabolism. Insights into the regulation of the *lug* gene cluster could provide insight into the role of lugdunomycin in the environment of *Streptomyces* sp. QL37 and why it is hardly produced in the laboratory (van Bergeijk *et al.*, 2020). Regulation of BGCs is controlled by both pleiotropic and cluster-situated regulators (Liu *et al.*, 2013). In this study the cluster-situated regulators (CSRs) encoded by the *lug* gene cluster were studied.

A mutational study of *lugRII–lugRV* showed that LugRII (LuxR-type), LugRIV (atypical response regulator) and LugRV (SARP-regulator) are required for angucyclines production in *Streptomyces* sp. QL37 and hence also for lugdunomycin production, while LugRIII (TetR-regulator) only plays a minor role in the regulation of the *lug* gene cluster. Overexpression of *lugRIV* and *lugRV* led to enhanced angucycline production compared to the control strain in early growth stages and - importantly - overexpression of *lugRV* led to improved lugdunomycin production. Based on these results, the role of LugRIV and LugRV was investigated further. The role of LugRIV and LugRV as transcriptional activators was further validated by quantitative proteomics, which showed that the proteins of the lugdunomycin BGC were significantly upregulated. Notably at later stages of growth, overexpression of *lugRV* increased the production of the angucyclines,

but not lugdunomycin (**Chapter 6**). As mentioned above, the likely final reaction required for lugdunomycin includes two substrates; a rearranged angucycline and *iso*-maleimycin (Wu *et al.*, 2019, Uiterweerd, 2020). Therefore, BGC23 also needs to be upregulated to achieve lugdunomycin biosynthesis, assuming that no other enzymes encoded by genes outside the two BGCs are required for the biosynthesis of lugdunomycin

The CSRs of a regulatory network act in a hierarchical process and the DNA activity of the regulators belonging to the class of LuxR, TetR and atypical response regulators can be modulated by the intermediates and the final product of the biosynthesis pathway (van der Heul *et al.*, 2018, Wang *et al.*, 2009). These two aspects lead to (multiple) feedback and/or feedforward loops, which does not yet take pleiotropic regulators into account. Examples of these complex regulatory networks are those of the auricin- and jadomycin BGC of *S. aureofaciens* CCM 3239 and *S. venezuelae* ISP5230, respectively (Kormanec *et al.*, 2014, Zou *et al.*, 2014). Cross-complementation of the generated mutants suggested that the response regulator LugRIV activates the expression of the SARP regulator LugRV that in turn controls the expression of the structural *lug* genes (**Chapter 6**). Further molecular biological studies, such as DNA-binding experiments and extensive transcriptional analysis, are required to obtain more insights into precisely how the LugR proteins control the expression of the *lug* gene cluster.

Chemical diversity of the secondary metabolome of *Streptomyces* sp. QL37.
With 35 BGCs predicted in its genome, *Streptomyces* sp. QL37 has the capacity to produce a plethora of secondary metabolites other than angucyclines and *iso*-maleimycin. Indeed, under the culture conditions tested in this study, the wild-type strain was found to also produce γ-butyrolactones (produced from BGC 16, 23 or 33), tetramate macrolactams (produced from BGC 6) and sceliphrolactam (produced from BGC 9) (Yang *et al.*, 2005, Moree *et al.*, 2014, Low *et al.*, 2018) (**Chapter 4**). Furthermore, some molecular families that were unrelated to angucyclines were specifically observed in some of the *lugO* null mutants. On MM agar the accumulation of molecules related to *N*-acyl glutamine and aminoalcohols were observed, in particular in *lugOI*, *lugOII* and *lugOIII* null mutants (Battista *et al.*, 2019, Won *et al.*, 2014, Harrison *et al.*, 2018). On R5 agar the accumulation of a novel lipopeptide was observed in extracts derived from the *lugOI* null mutant (possibly produced from BGC 21) (**Chapter 5**). A more extensive

OSMAC approach, combined with genetic manipulation of the strain, should shed more light on the extent of the specialised metabolome of *Streptomyces* sp. QL37, and discover the chemical space of the natural products it can produce. Different studies already reported that affecting the production levels of natural products that are abundantly produced by the host, can lead to significant alterations in the production of other unrelated secondary metabolites (Iorio *et al.*, 2021, Culp *et al.*, 2019). This was demonstrated by the fact that inactivation of the BGCs for the highly produced antibiotics streptomycin and streptothricin led to the production of previously hidden rare unknown natural products (Culp *et al.*, 2019). In addition, blocking several steps in the pseudouridimycin (PUM) pathway in *Streptomyces* sp. ID38640 led to the production of previously unseen and unrelated metabolites. Thus, interfering with one BGC can unveil a new potential of the producer strain and possibly be used to discover new secondary metabolites and antibiotics (Iorio *et al.*, 2021).

Morphological changes associated with angucycline production.

Production of angucyclines disturbs growth and blocks development, most likely due to cytotoxicity as a result of their DNA-degrading properties (Kharel *et al.*, 2012). In line with that, we observed major differences in morphological differentiation between the wild-type strain on the one hand, and the mutants that failed to produce angucyclines (Δ*lug-pks*, Δ*lugRII*, Δ*lugRIV* and Δ*lugRV*) on the other, particularly on R5 agar (**Chapter 6**). Angucycline non-producers developed well, whereas the strains producing angucyclines, including the wild-type strain, were inhibited in their development, resulting in a nonsporulating phenotype. Previously it was observed that the cytotoxic DNA-damaging prodiginines produced by *S. coelicolor* postpone sporulation of the strain on rich media. The *redD* mutant of *S. coelicolor*, which is unable to produce these molecules, showed precocious morphological differentiation on R2YE agar plates as compared to the parent strain (Tenconi *et al.*, 2020). The authors proposed that prodiginines mediate the onset of programmed cell death of the vegetative mycelium, postponing the developmental program until sufficient nutrients are produced to support extensive aerial growth. The accelerated sporulation and smaller colonies of *redD* mutants are then explained by the lack of cell death and hence accelerated life cycle. Angucyclines may play a similar role in the control of cell death and development due to their DNA-damaging properties (Tenconi & Rigali, 2018).

Concluding remarks and future perspectives

This study provides new insights into the biosynthesis and regulation of angucyclines and the highly rearranged angucycline derivative lugdunomycin. These new insights may be applied to increase lugdunomycin production by *Streptomyces* sp. QL37 with the aim to isolate significant quantities of lugdunomycin and determine its mode-of-action and that of related molecules.

The *lug* gene cluster consists of 28 genes, including the minimal PKS, five oxygenase genes (*lugOI–lugOV*) and five regulatory genes (*lugRI–lugRV*). LugOIII and LugOV likely mediate C-ring cleavage during angucycline biosynthesis, which is one of the key reactions in lugdunomycin biosynthesis. In addition, we hypothesise that an angucycline epoxide is possibly the required substrate for C-ring opening. Homologous BGCs were found in other *Streptomyces* strains. BGCs that contain *lugOIII* and/or *lugOV* homologues likely specify a diversity of C-ring cleaved angucyclines and perhaps also lugdunomycin. However, the final reaction in the lugdunomycin pathway likely requires two BGCs, one producing the diene (angucycline) and the other producing the dienophile (*iso*-maleimycin) of a Diels-Alder reaction, which is proposed to be spontaneous, based on the fact that a racemic mixture is produced (Wu *et al.*, 2019).

An interesting question that arises is whether lugdunomycin biosynthesis offers an evolutionary advantage to the producer. In a general sense, the concept that molecules may be produced by multiple BGCs may be much more common than we currently anticipate. We already observed this phenomenon for the production of endophenasides by *Kitasatospora* sp. MBT 66 and actinomycin L from *Streptomyces* sp. MBT 27 (Wu *et al.*, 2016b, Machushynets *et al.*, 2022). Indeed, most of the drug discovery research emanates from the idea that one molecule is derived from one BGC, and this is also the basis for synthetic biology approaches. The example of lugdunomycin shows that in this way interesting chemical space may be missed and we should still rely on the chemistry isolated from the producer strain.