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

Heul, H.U. van der

General Introduction

The biology of *Streptomyces*

Most of the antibiotics used in the clinic are based on microbial natural products (Demain, 2014, Hutchings *et al.*, 2019, Wright, 2014). Especially *Streptomyces* and other genera of the Actinobacteria produce a plethora of bioactive natural products that are used in the pharmaceutical industry and agricultural sector; not only as antibiotics, but also as anti-cancer, anti-fungal, anti-viral agents or immunosuppressants (Barka *et al.*, 2016, Berdy, 2005). Streptomycetes are Gram-positive bacteria with a G+C rich genome that undergo a complex morphological development. Antibiotic production by streptomycetes is generally linked to their life cycle; this starts with the germination of a spore, initiating the growth of vegetative hyphae (Flärdh & Buttner, 2009). Via tip extension and branching, these grow out to form an intricate vegetative mycelium. Under stress conditions, such as nutrient depletion, new hyphae are formed on top of the vegetative mycelium that erect into the air, and are therefore called aerial hyphae. During this transition, a part of the vegetative mycelium undergoes programmed cell death, whereby nutrients are released from the lysed cell and used for the synthesis of the aerial hyphae (Chater & Chandra, 2006, Barka *et al.*, 2016, Flärdh & Buttner, 2009). Antibiotics are mostly produced during this stage of development, probably to protect the host against other microbes that seek to scavenge the released nutrients (Bibb, 2005, Tenconi *et al.*, 2020). The aerial hyphae further develop into spores that can then be spread in the environment and start a new lifecycle (Flärdh & Buttner, 2009).

Antibiotic discovery

In the Golden Age of antibiotic discovery (1940s-1960s), many clinically used antibiotics have been discovered from *Streptomyces* (Hutchings *et al.*, 2019). However, by the 1970s many strains and compounds were readily found again, a process called replication. Eventually, the return of investment of high-throughput screening in combination with targeted based approaches became too low (Kolter & van Wezel, 2016, Silver, 2011). Therefore, large industrial companies stopped investing in the discovery of novel antibiotics (Payne *et al.*, 2007, Brown & Wright, 2016). While drug discovery is failing, antimicrobial resistance (AMR) is on the rise; thus, pathogenic bacteria that once seemed like a thing from the past, are once more posing a huge threat to human health. These multi-drug resistant (MDR) pathogens include the so-called ESKAPE pathogens (*Enteroccocis faecium*, *Staphyloccocus aureus*, *Klebsiella pneumoniae*, *Acinetobacter baumanni*,

Pseudomonas aeruginosa, and *Enterobacter* species) and *Mycobacterium tuberculosis* (MDR-TB) (Rice, 2008, Ventola, 2015).

A possible solution came via next-generation sequencing, which revived the interest in antibiotic discovery in the early 21st century (Katz & Baltz, 2016, Bentley *et al.*, 2002). The production of antibiotics and other natural products is mediated by a cluster of genes that code for enzymes required for the biosynthesis of the compound(s). Such a cluster is called a biosynthetic gene cluster (BGC) (Medema *et al.*, 2015). Sequencing of *Streptomyces* genomes revealed that scientists had perhaps only scratched the surface in terms of the number of BGCs, whereby each *Streptomyces* species contains many more BGCs than initially thought. Even the model organism *Streptomyces coelicolor*, which had been worked on by thousands of scientists worldwide, had many BGCs that had never been seen (Bentley *et al.*, 2002). Until the genome sequence was published, no fewer than four antibiotics had been identified, namely undecylprodigiosin (Red), actinorhodin (Act), calcium dependent antibiotic (Cda) and the plasmid-encoded methylenomycin (Mmy). However, the genome sequence revealed in total 22 BGCs (Bentley *et al.*, 2002, Hoskisson & Seipke, 2020). One of these, the cryptic *cpk* cluster, is repressed by the global regulator DasR (Rigali *et al.*, 2008). It was later shown that this BGC specifies a novel antibiotic, known as coelimycin (Gomez-Escribano *et al.*, 2012). This is a well-know example of the principle of cryptic or silent BGCs, which started the postgenomic era of antibiotic discovery. Still, the major question we need to address is, are cryptic BGCs the answer to the AMR crisis? Can we really expect new chemistry, and in the end new drugs from them?

Activation of silent biosynthetic gene clusters

The reason why many BGCs are silent in the laboratory is most likely that the signals needed to activate them are not yet understood, and are to be found in the natural habitat, often the soil or a eukaryotic host. Indeed, streptomycetes are found everywhere, in the soil, sea, plants, animals and humans, and are part of a community with their own signals that can trigger antibiotic production in the bacterium (van der Meij *et al.*, 2017, van Bergeijk *et al.*, 2020). Regulatory networks play a key role in this phenomenon as it links environmental signals, such as the availability of sugars, nitrogen and phosphate supply to the activation of a BGC (van der Heul *et al.*, 2018, Liu *et al.*, 2013, van Bergeijk *et al.*, 2020).

All BGCs are regulated by one or more pleiotropic and cluster-situated regulators (Bibb, 2005, van der Heul *et al.*, 2018). These may either activate the transcription of the BGCs, or conversely, block the transcription of the BGC and thus act as repressors. External molecules can bind to these regulators and function as allosteric effectors, and thereby inhibit or activate the DNA binding of the activators or repressors (Liu *et al.*, 2013). One well-known example is the binding of glucosamine-6-phosphate to the pleiotropic regulator DasR, which releases DasR from the DNA, upon which the expression of the actinorhodin (Act), undecylprodigiosin (Red), coelimycin and calcium-dependent antibiotic (Cda) BGC is activated in *S. coelicolor* (Fillenberg *et al.*, 2016, Swiatek *et al.*, 2012, Rigali *et al.*, 2008).

One of the methods to activate antibiotic production is to find the specific elicitor (Zhu *et al.*, 2014b, van Bergeijk *et al.*, 2020, van Bergeijk *et al.*, 2022). A classical method is OSMAC (One Strain Many Compounds), whereby streptomycetes are grown at various conditions to activate multiple BGCs (Machushynets *et al.*, 2019). A recently developed method to activate silent BGCs in a more targeted manner is the high throughput elicitor screening method (HiTES). In this method a reporter gene is inserted in a BGC of interest and screened against a library of elicitors (Seyedsayamdost, 2014).

Genome mining

Genome mining using specific algorithms to predict BGCs has become a routine method (Medema *et al.*, 2011). These include antiSMASH, BAGEL, ClustScan, CLUSEAN, NP searcher, and PRISM (Medema *et al.*, 2011, de Jong *et al.*, 2006, Starcevic *et al.*, 2008, Weber *et al.*, 2009, Medema *et al.*, 2014, Li *et al.*, 2009, Skinnider *et al.*, 2015, Lee *et al.*, 2020). These detection methods are based on finding core genes of the known BGC classes, such as the non-ribosomal peptides (NRPS), Polyketide synthases type I, -II, -III (PKS I, PKS II, PKS IIII) and ribosomally synthesised and post-translationally modified peptides (RiPPs) (Medema *et al.*, 2021, Skinnider *et al.*, 2017, Helfrich *et al.*, 2019, Chevrette *et al.*, 2017). Based on the presence of accessory genes, such as transporters and modification genes, a border is predicted (Medema *et al.*, 2011). Prediction of BGCs is also based on comparison with known orthologous BGCs. However, these tools overlook BGCs of unknown compound classes. Therefore, it is important to develop new genomic mining algorithms. An example is the improvement on the identification

of novel subclasses of RiPPs using the machine-learning algorithm decRiPPter (Data-driven Exploratory Class-independent RiPP TrackER), which was developed in our laboratory and implemented in the novel version of antiSMASH, version 6.0 (Blin *et al.*, 2021, Kloosterman *et al.*, 2020a, Kloosterman *et al.*, 2020b).

BGCs with high similarity to known gene clusters are often discarded as they are likely to lead to the production of known molecules (Martinet *et al.*, 2020). However, while such BGCs may encode similar enzymes that catalyse known enzymatic reactions, they may still use different substrates and thus lead to new chemical diversity (Martinet *et al.*, 2020, Yan *et al.*, 2016). In addition, similar BGCs may be differentially expressed amongst different strains, due to the absence or presence of certain regulatory elements (Amos *et al.*, 2017). Thus, when an interesting, but silent BGC is found in one strain, it is possible to look for a similar BGC with different regulatory elements in another strain, where the cluster may be active. This offers more possibilities to identify the natural product that is related to the BGC.

Expanding the chemical space

A means to find novel antibiotics or other bioactive compounds is the screening of thousands of extracts derived from natural sources and study the present chemistry, which is the field of metabolomics (Zdouc *et al.*, 2021, Wu *et al.*, 2015b). However, the analysis of such extracts is challenging due to their complexity; they contain primary and secondary metabolites and residues from the growth medium (Wolfender *et al.*, 2019). A valuable development in the detection of natural products is high resolution mass spectrometry (MS). This method allows the efficient detection of molecules that would have been missed using the classical methods, such as thin layer chromatography (TLC) and liquid chromatography (LC) combined with UV detection, thus broadening the chemical space. The LC-MS data of extracts derived from different growth conditions can be used to prioritise metabolites for further study by applying uni-and multivariate statistical analyses (Beniddir *et al.*, 2021, Machushynets *et al.*, 2019, Wu *et al.*, 2016a). However, LC-MS data only comprise a list of masses and obtaining more information on structural features of the molecule of interest still requires isolation of the molecule combined with structural elucidation using nuclear magnetic resonance spectroscopy (NMR)(Rogers *et al.*, 2019). In addition, it remains difficult to dereplicate known compounds and identify yet unknown ones. For this reason,

the Global Natural Products Social Molecular Networking (GNPS) platform was developed, which is based on the data obtained from tandem mass spectrometry (MS/MS) fragmentation spectra. This tool made a huge contribution to the field of metabolomics, clustering molecules with similar MS/MS fragmentation patterns and therefore molecules with similar structures (Nothias *et al.*, 2020, Nguyen *et al.*, 2013, Schmid, 2020). An important implementation is the comparison of the MS/MS data with spectral libraries (Wang *et al.*, 2016). This allows rapid dereplication of known natural products and - more importantly - identification of yet unknown analogues of already known compounds [30, 50].

Linking Natural products with their cognate BGCs

Fermentation under different growth conditions in combination with GNPS networking and transcriptomics or proteomics can readily facilitate the linkage of natural products with their cognate BGCs (Du & van Wezel, 2018, Gubbens *et al.*, 2014, Wu *et al.*, 2015a, Amos *et al.*, 2017). The identified BGC may in turn be genetically modified and/ or expressed heterologous in an optimised *Streptomyces* host, so as to increase the production of the natural product of interest for upscaling, structural elucidation and bioactivity assays (Li *et al.*, 2019, Jones *et al.*, 2013, Gomez-Escribano & Bibb, 2011, Zheng *et al.*, 2007, Hutchinson & Colombo, 1999). The characterisation of a BGC may help in the discovery of the biosynthetic pathway leading to the synthesis of natural product and thereby the discovery of novel modification enzymes that may be used to synthesise new natural product analogues which are more bioactive or can overcome antibiotic resistance mechanisms (Xiao *et al.*, 2020, Kim *et al.*, 2015a, Baltz, 2018, Hulst *et al.*, 2021).

Scope of this thesis

In the attempt to find novel antibiotics, a strain isolated from the Qinling mountains in China was screened using the OSMAC method. This led to the discovery of lugdunomycin (Wu *et al.*, 2019), which is produced by *Streptomyces* sp. QL37 (Zhu *et al.*, 2014b) . The molecule is characterised by its heptacyclic ring, a spiroatom, benzaza[4,3,3]propellant moiety, and two all carbon stereocenters (Wu *et al.*, 2019). Lugdunomycin is derived from the angucyclines, molecules with a benz[a] anthracene backbone that form one of the largest classes of natural products and are produced by polyketide type II synthases (Kharel *et al.*, 2012). Angucyclines are characterised by their benz[a]anthracene core with an additional carbohydrate

residue. Lugdunomycin is derived from the angucyclinones; angucyclines without the sugar residue (Mikhaylov *et al.*, 2021, Wu *et al.*, 2019). In this thesis we further refer to these natural products as angucyclines, which are well-known for their antibacterial and anti-cancer activities. However, their toxicity makes them less attractive for clinical application, which may be solved by creating chemical analogues (Mikhaylov *et al.*, 2021). A biosynthetic pathway was proposed for the production of lugdunomycin (Figure 1). This starts with one molecule of acetyl-CoA and nine molecules of malonyl-CoA. The minimal polyketide machinery, consisting of the heterodimer ketosynthase alpha (KSα), ketosynthase beta (KSβ) and acyl carrier protein (ACP) use these primary metabolites for the synthesis of a decaketide. Ketoreductases and cyclases then catalyse this carbon chain to generate the basic angucycline framework. The first stable intermediate is the angucycline UWM6 or its dehydrated form prejadomycin. Post-PKS tailoring enzymes, including a methyltransferase, reductases and oxygenases subsequently modify the angucyclines to a final product (Kharel *et al.*, 2012). The hypothesised biosynthesis pathway of lugdunomycin itself requires two key steps, namely (I) the rearrangement of an angucycline by opening of the C-ring of an angucycline backbone through a Baeyer-Villiger reaction; and (II) a final reaction wherein the rearranged angucycline reacts with a dienophile in an intermolecular Diels-Alder reaction (Wu *et al.*, 2019). Though angucyclines are produced in high quantities under most conditions, lugdunomycin itself is produced in very small amounts. More insights into the mode of action and biosynthesis of lugdunomycin is important to further augment our understanding of angucyclines in general, and of lugdunomycin in particular.

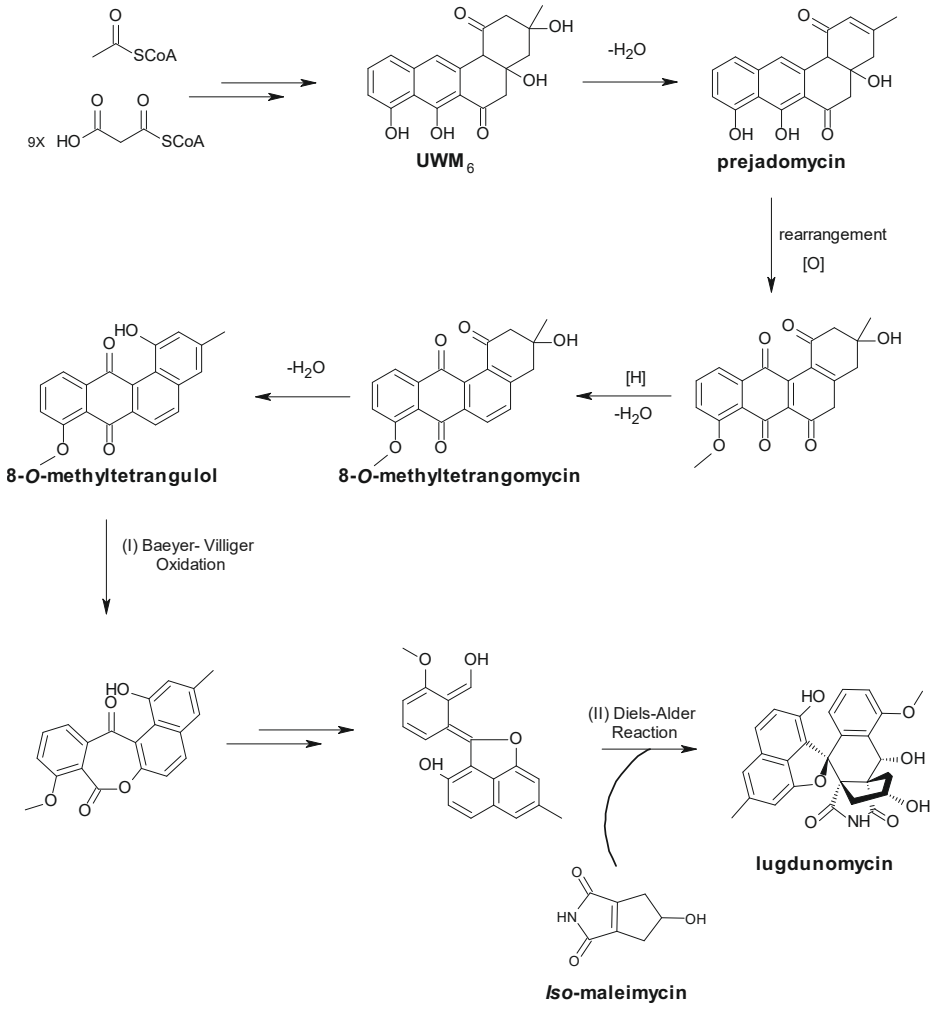

Figure 1 Postulated simplified lugdunomycin biosynthesis pathway
The exceptional transformation involves a C-ring cleavage between bond C6a–C7 in the angucycline framework and a Diels-Alder reaction between an angucycline derived diene and the dienophile *iso*-maleimycin (Wu *et al.*, 2019).

The aim of this thesis is to obtain novel insights into the transcriptional control of the lugdunomycin BGC and into its biosynthesis pathway, and to apply this knowledge to increase the production of lugdunomycin.

Chapter 2 reviews the regulatory networks that control antibiotic production in Actinobacteria. The includes specific and global control of BGCs, the impact of

primary metabolism (such as the availability of nitrogen, phosphate and carbon sources) on antibiotic production, the link between the complex life cycle and antibiotic production and the involvement of specialised metabolites in the regulation of BGCs (van der Heul *et al.*, 2018).

In **Chapter 3** the BGC governing lugdunomycin biosynthesis is characterised. The *lug* gene cluster is a polyketide type II BGC comprising 28 genes. The importance of this gene cluster in lugdunomycin biosynthesis was verified using gene deletion experiments and the extent of the cluster was assessed using temporal RNA-seq. Furthermore, bioinformatic comparisons were used to find other *lug*-type gene clusters in other *Streptomyces* and *Kitasatospora* strains and we evaluated whether the presence of a *lug*-type gene cluster in a strain is correlated to phylogeny.

The functional role of the *lug* gene cluster in lugdunomycin biosynthesis was further examined by heterologous expression in **Chapter 4.** A paired-omics approach, including genomics, metabolomics and transcriptomics, was applied to find a missing link in lugdunomycin biosynthesis.

The opening of the C-ring of the angucyclines is an important step in lugdunomycin biosynthesis. It was hypothesised that a C-ring rearranged angucycline is used as a substrate in the final reactions of lugdunomycin biosynthesis, presumably via a Baeyer-Villiger oxidation. In **Chapter 5** the five oxygenases encoded by the *lug* gene cluster have been studied in detail, using gene deletion strategies combined with metabolomics and GNPS networking tools. A primary focus was to discover the oxygenases required for cleavage of the angucycline C-ring.

The *lug* gene cluster contains five regulatory genes, which are studied in **Chapter 6**. Gene deletion experiments in combination with metabolomics and proteomics were applied to identify the role of the five *lugR* genes in the transcriptional control of the *lug* cluster, and identify the key transcriptional activators. Over-expression of the regulators was used as an approach to increase the production of angucyclines and in particular of lugdunomycin.

Chapter 7 presents a summarising discussion of the main findings of this thesis, including conceptual ideas on how to proceed with studies of the biosynthesis of lugdunomycin, aimed at obtaining a full picture of the mode of action of this exciting molecule.