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

Nuñez Santiago, I.

06

CHAPTER 6

General discussion

Introduction

The increasing prevalence of resistance mechanisms, such as antibiotic-resistant pathogenic bacteria and drug-resistant tumor cells, necessitates the search for novel drug candidates. *Streptomyces* species, known for their remarkable chemical diversity, are a prolific source of bioactive natural products (Alam *et al.*, 2022). These filamentous soil bacteria are responsible for the production of nearly half of all antibiotics currently utilized in clinical settings, as well as various natural products employed in medical, biotechnological, and agricultural applications. Advances in genome mining strategies have revealed that *Streptomyces* genomes harbor a significantly higher number of biosynthetic gene clusters (BGCs) for natural products than previously estimated, many of which remain silent under standard screening conditions (Nett *et al.*, 2009). This untapped genetic potential presents an opportunity to explore a vast chemical space that may lead to the discovery of new bioactive compounds.

Several approaches are currently employed to explore this untapped chemical space. However, these strategies are not without their limitations as discussed in **Chapter 1**. In recent years, the advent of big data and high-throughput screening of extensive natural product collections has often led to the frequent rediscovery of known molecules, revealing only a small fraction of the available chemical diversity. This phenomenon highlights the challenge of going beyond the 'tip of the iceberg' to uncover novel compounds with unique chemical scaffolds.

In this thesis, *Streptomyces* species are explored as a primary source for the discovery of novel natural products, aiming to explore the chemical space through a combination of advanced methodologies. This work integrates a variety of analytical chemistry techniques with bioactivity assays, as well as cultivation approaches like elicitors and OSMAC (One Strain Many Compounds), to trigger the production of diverse natural products. Advanced techniques for BGC screening are employed, combining tools like antiSMASH with state-of-the-art predictors such as PARAS. In a multi-faceted approach, genomic data is integrated with metabolomic analysis, utilizing methods like feature-based molecular networking (FBMN) via GNPS and MS-mining using MassQL. Additionally, the study explores the complex lugdunomycin biosynthetic pathway as a case study, challenging the traditional 'one BGC-one molecule' paradigm. By adopting these diverse techniques, this thesis seeks to push the boundaries of natural product research, uncovering previously unexplored molecular diversity from *Streptomyces.*

Targeted prioritization in early-stage drug discovery by reducing the candidate complexity.

A variety of bioactive molecules are sought for specific targets, ranging from anticancer treatments aimed at particular cell lines to antibiotics targeting significant public health pathogens. When the objectives of bioactivity are clearly defined, implementing mechanisms for targeted prioritization during the early stages of drug discovery becomes essential. The nanoRAPIDS analytical platform (Nunez Santiago *et al.*, 2024) facilitates the rapid discovery, efficient identification, and dereplication of bioactive metabolites through nanofractionation (as discussed in **Chapter 2**). This platform is particularly effective in identifying low-abundance molecules in complex mixtures, which may otherwise be obscured by more abundant compounds, possibly less relevant too.

In addition to identifying bioactive compounds, challenges often arise when attempting to activate silent BGCs in the search for novel bioactive molecules. While variations in bioactivity may occur, the complex composition of the extract complicates the interpretation of results, hindering the ability to ascertain the specific contributions of newly activated BGCs among the numerous compounds present. Addressing these complexities is crucial for accurately assessing the potential of elicitor-induced changes in bioactivity.

We utilized nanoRAPIDS to identify the bioactive congeners that correlated to increased antimicrobial activity of *Streptomyces* sp. MBT84 after challenge with catechol, as compared to the significantly lower bioactivity observed in the control medium (van Bergeijk *et al.*, 2022). This approach allowed pinpointing the specific molecules within the complex extract responsible for the enhanced bioactivity. Most features identified were dereplicated as angucyclines, a well-known family of bioactive compounds. Notably, the application of nanoRAPIDS also led to the identification of a novel bioactive feature, resulting in the discovery of saquayamycin N, an angucycline derivative containing an *N,N*-acetylcysteine moiety.

Expanding our understanding of biosynthetic complexity and the hidden potential in nature's chemistry

The discovery of lugdunomycin, a molecule with highly complex chemical architecture that is produced by *Streptomyces sp.* QL37, has provided key insights into the complexity of natural product biosynthesis. For years, production of the compound remained a mystery. Biosynthesis of the molecule required an angucycline biosynthetic gene cluster (BGC12 or *lug*) (Wu *et al.*,

2019b). However, the structure of lugdunomycin revealed a distinct moiety that could not be explained by this BGC alone. Work in this thesis and in close collaboration with the University of Groningen showed that lugdunomycin is the product of not one, but two independent BGCs (Uiterweerd *et al.*, 2024). The *lug* gene cluster thereby gives rise to the rearranged-angucyclinone portion of the molecule, which consists of elmonin, while a second BGC specifies the biosynthesis of *iso*-maleimycin, which is the second portion that gets incorporated to form lugdunomycin. The production of *iso*-maleimycin by *Streptomyces* sp. QL37 was previously demonstrated (Uiterweerd *et al.*, 2024). The requirement of two BGCs to synthesize one compound challenges the classical notion of "one molecule, one BGC" and highlights the intricate interplay between different biosynthetic pathways within a single organism. For more details and examples on this concept the reader is referred to Chapter 1.

The final lugdunomycin structure is synthesized *in vivo* by an intermolecular Diels-Alder reaction, which uses elmonin as diene and *iso*-maleimycin as dienophile, thus linking two molecules from different BGCs. The requirement of two BGCs also explains why heterologous expression of the *lug* BGC in *S. coelicolor* M1152 results in the production of angucyclines and derivatives, but not of lugdunomycin. This underscores the importance of in-depth investigation into the molecular mechanisms underlying natural product biosynthesis, and also issues a warning to scientists who routinely employ heterologous expression to identify the product of unknown BGCs. After all, when two BGCs are needed, the final product will never be seen, unless the second BGC is also present in the host. Although comprehending these complex processes may seem time-consuming and challenging, it is essential for advancing our fundamental understanding of how nature constructs secondary metabolites. This effort is not merely an academic exercise, but rather a way to expand our appreciation of the hidden chemical space that remains to be explored. By doing so, we unlock new opportunities for innovation, particularly in the discovery of novel antibiotics or modification of old ones, where nature has repeatedly inspired breakthroughs.

The case of lugdunomycin also highlights the limitations of current bioinformatics-driven approaches, which often focus on individual BGCs in isolation. This approach assumes that each cluster is responsible for a distinct molecular product. However, as demonstrated by lugdunomycin, the reality is often more complex. Multiple BGCs may be responsible for the biosynthesis of

chemical entities that would be missed if BGCs are expressed individually. This emphasizes the need to rethink bioinformatic strategies to better account for inter-BGC interactions and their potential to yield new chemistry.

Angucyclines are widespread secondary metabolites commonly produced by *Streptomyces* species (Kharel *et al.*, 2012). This large class of molecules is built upon a core four-ring polyketide synthase (PKS) structure known as the angucyclinone. However, what makes angucyclines so chemically diverse is the extensive array of post-PKS modifications facilitated by a variety of tailoring enzymes (Fan & Zhang, 2018), leading to heavily rearranged structures and expanding their bioactivity profiles (**Chapter 4**).For example, enzymes capable of ring opening are known to be key players in the structural diversity of angucyclines. In *Streptomyces* sp. QL37, which produces lugdunomycin, the presence of enzymes responsible for C-ring cleavage is one such modification (Elsayed *et al.*, 2023, Xiao *et al.*, 2020), showcasing how rearrangements can lead to novel molecular frameworks. Similarly, *Streptomyces albus* produces atypical angucyclinones with a 6-5-6-6 ring structure and highly oxidized A-rings (Jin *et al.*, 2018), further highlighting the role of enzymatic modifications in generating unique chemical entities. Another notable example is *Streptomyces* sp. CS057, whose angucycline BGC contains a glycosyltransferase gene, suggesting the possibility of glycosylation, a modification known to enhance the pharmacological properties of natural products.

Building on this knowledge, the idea of creating an “enzyme toolbox” for angucycline biosynthesis becomes an exciting prospect. This toolbox could consist of various tailoring enzymes, each capable of modifying the same PKS-derived angucyclinone core in different ways. While such an approach presents significant challenges, it also offers the potential to generate an almost limitless number of novel compounds. These modifications could be strategically applied to create new molecules with enhanced or entirely novel biological activities. Moreover, as discussed in the context of lugdunomycin, the potential synergy between different classes of molecules opens even more possibilities. By combining the biosynthetic capabilities of multiple BGCs within a single organism, researchers could generate hybrid molecules that may exhibit unexpected and valuable bioactivities.

The bioactivity of angucyclines, particularly their effectiveness against cancerous cell lines and Gram-positive pathogens such as methicillin-resistant *Staphylococcus aureus* (MRSA), has already been well-documented. As

antibiotic resistance continues to rise globally, the discovery of new antibiotics remains a critical priority. Given the structural diversity and proven antimicrobial activity of angucyclines, further exploration of their biosynthesis, particularly through the lens of combinatorial biosynthesis, could unlock new antibiotics with potent efficacy against resistant strains.

Beyond the low-hanging fruit: uncovering new molecules from intensively studied strains

The term "talented strain" is often used in natural product discovery to describe strains capable of producing a variety of secondary metabolites. However, this is particularly true for most *Streptomyces* species, which typically harbour between 25 to 70 distinct BGCs (Belknap *et al.*, 2020). While many of these BGCs remain silent under laboratory conditions, their latent biosynthetic potential is undeniable. The challenge lies in exploring and investigating as many of these BGCs as possible, not only to expand our understanding of microbial metabolism but also to uncover novel compounds that may hold therapeutic potential.

Streptomyces sp. QL37 serves as a prime example of this untapped potential and had more surprises in store for us. Genomic analysis revealed 30 BGCs, each with the potential to encode structurally diverse natural products. One NRPS BGC stood out because the specialized bioinformatic tool PARAS (Predictive Algorithm for Resolving A-domain Specificity) predicted the production of a peptide containing several non-proteinogenic amino acids (**Chapter 5**). These included three piperazic acid residues, one 3-amino-2-methylpropanoic acid (AMPA), and one *allo*-isoleucine, highlighting the potential for a structurally novel peptide.

Through the integration of advanced culturing techniques, metabolomics tools such as GNPS, and data analysis platforms like MassQL, the molecules marushamycin A, B and C were identified. To the best of our knowledge, marushamycin represents a new family of piperazic acid-containing NRPS products. The 2D structure of marushamycin A was elucidated using a combination of NMR experiments, MS/MS analysis, and genomic predictions. Genomic analysis combined with *de novo* sequencing and NMR, allowed us to propose a biosynthetic pathway for marushamycin.

Given the vast collections of *Streptomyces* strains that are available in the private and academic setting, there is an enormous potential for further in-depth analysis of individual strains to explore their biosynthetic capacities. Combining various advanced technologies—such as bioinformatics, metabolomics, genomics, chemical synthesis, and structural biology— is important for the successful exploration of the chemical space for new natural products that has been overlooked in the past. These interdisciplinary approaches, which are also employed in this thesis, not only allow the discovery of novel compounds but also reveal biosynthetic pathways that may otherwise remain hidden.

This exploration of new chemical space is essential, as only about 3% of the chemical space is currently known (Gavriilidou *et al.*, 2021). This vast unexplored territory underscores the need for innovative approaches to broaden our chemical repertoire. AI-driven methodologies are emerging as powerful tools for predicting novel compounds and uncovering biosynthetic pathways that have not been considered in traditional discovery efforts (Mullowney *et al.*, 2023). Additionally, new screening strategies, such as the use of elicitors informed by a deeper understanding of microbial ecology, offer promising avenues for discovering previously unknown natural products that could be triggered under specific environmental conditions (Okada & Seyedsayamdost, 2017).

Furthermore, heterologous expression offers significant potential for revealing new chemistry. Achieving heterologous expression of natural products at medium or high throughput could greatly enhance our ability to uncover novel compounds, accelerating the discovery of valuable molecules (Kolter & van Wezel, 2016). This is particularly critical in the context of antimicrobial resistance (AMR), where the urgent need for new antibiotics demands more efficient and effective methods for exploring chemical diversity.

As the exploration of chemical space continues to evolve, the use of new tools and techniques to interrogate strains will hopefully reveal new biosynthetic potential. Only through a combination of innovative approaches and interdisciplinary collaboration can we further unravel the complex pathways for the biosynthesis of microbial natural products and fully realize their untapped potential. This should allow scientists to find the new drugs we so badly need in our continuing fight against multi-drug-resistant pathogens and hopefully turn the tide in the AMR crisis.