Understanding and Engineering the Biosynthesis of Thioamitide

King Wing Louis Fong

John Innes Centre

A thesis submitted March 2023 to the University of East Anglia for the degree of Doctor of Philosophy.

Abstract

Ribosomally synthesised and post-translationally peptides (RiPPs) is a structurally diverse and clinically relevant class of natural products. The thioamitides, a RiPP family, have been drawing attention for their fascinating uncommon chemical features and promising antitumour activities. Moreover, their antitumour activities can vary with a single change of functional group.

With the rapidly developing bioinformatic tools and genome mining efforts, an increasing number of unusual and interesting biosynthetic pathways are being identified in bacteria, which is promising for the discovery of novel thioamitides with diverse molecular structures and consequently different biological activities. Here, thiopotensamide, a novel thioamitide produced by *Nocardiopsis potens* DSM 45234, is presented. This compound is characterised with a dehydration and an unexpected methylation that are not seen in other family members. Confirmation of the molecular structure by NMR was hampered by low productivity and the complexity of secondary metabolite production in the native producer host, as well as unsuccessful genetic manipulation of *N. potens*. However, heterologous expression of the biosynthetic gene cluster in a heterologous host provided promising results that enabled key insights into the biosynthesis of mature thiopotensamide.

In this thesis, the formation of thioamide bonds, one of the key chemical features from thioamitides, is also investigated using the biosynthetic pathway of thiostreptamide S4. This single atom substitution has been reported to play a key role in peptide stability and bioactivities elsewhere. It was hypothesised that the thioamide bond formation was catalysed by a YcaO-TfuA protein pair, which was investigated using *in vitro* biochemical reactions.

Acknowledgements

This PhD has been an amazing experience in my life, and I would like to give my most sincere gratitude to my supervisor Dr Andrew Truman for providing me this life changing opportunity. Andy was one of the most patient and supportive supervisors in the world, who guided me whenever I needed, but never intruded upon me, which allow me to develop my research skills and independence. My PhD would not have been successful without him. Another important person in my PhD life is Dr Natalia Miguel-Vior. Natalia taught me almost every foundation knowledge I need to know for my PhD, in particular the biology and provided answers for my endless questions. Without Natalia, my PhD would have been extremely challenging. Not only was she my teacher, but she was also my friend and like my big sister who looked after me flawlessly. I could not thank her enough for all the supports. Dr Javier Santos-Aberturas also as important in my research career. Not only did he teach me lots of skills, but he also provided endless ideas for my cloning works.

I could not miss the opportunity to thank the rest of Truman lab members and office 218 members who created a friendly and supportive atmosphere to keep me motivated through the down time of my research. I am extremely lucky to have the opportunity to working in Molecular Microbiology where everyone is supportive to each other.

Finally, it is important to express my indebtedness to my family and friends who are always there to support me, in particular the Starkings family who took care of me during my PhD.

Table Of Contents

Table Of Figures

Table Of Tables

Chapter 1 Introduction

1.1 Natural products

1.1.1. Natural products in nature

Compounds occurring in nature are found across all domains of life and typically categorised into three groups (Hanson, 2003). First, primary metabolites are essential for cellular functions, so they exist in all living things. For example, nucleic acids, amino acids, and sugars. The second type of naturally occurring compounds is polymeric materials with high molecular weight, for instance cellulose. Finally, secondary metabolites which are often referred to as natural products, is the most diverse category of the three (Hanson, 2003).

Unlike primary metabolites, natural products are not strictly required for survival, and they differ from organism to organism in both structure and function. Instead, they possess biological effects that will ecologically or physiologically benefit their producers to thrive in particular niches (Hanson, 2003, Gavriilidou et al., 2021). For example, secondary metabolites can regulate growth, development, and defence for the producers (Erb and Kliebenstein, 2020). In plants, volatile organic compounds can attract pollinators and defend against pests or predators (Reinhard et al., 2004, Vivaldo et al., 2017). In microorganisms, as bacteria often grow in dense and multispecies communities, one of their competition mechanisms to achieve ecological and evolutive success is to produce secondary metabolites such as toxins or other antimicrobial molecules as their chemical weapons, small signalling molecules, or scavenging molecules to retrieve nutrients from the environment, among many other functions (Demain and Fang, 2000, Kaper and Sperandio, 2005, Hider and Kong, 2010, Behnsen and Raffatellu, 2016, Granato et al., 2019).

1.1.2. Natural products in human history

The use of natural products in human daily life is potentially as old as human history itself. Natural products have been used for many purposes from food flavouring or odour with herbs and fruits to the use of toxins as poisons. Poisonous materials, for instance alkaloids, were extracted from plants or from animals such as toads, snakes and frogs and applied onto arrows for tribal wars or hunting (Aniszewski, 2007). Most of the recorded historic use of natural products is herb related and it can be tracked back to 5000 years or longer (Aniszewski, 2007, Bent, 2008).

One main use of natural products is as medicines and has been described all over the world and throughout the time in history. Oils of several plants, such as *Cedrus* species (cedar), *Cupressus sempervirens* (cypress), *Clycyrrhiza glabra* (licorice), and *Commiphora* species (myrrh) were found recorded on clay tablets dated from about 2600 B.C. in Mesopotamia (Cragg and Newman, 2005). In China, plants, animals, fungi and minerals were also used for medical purposes since ancient time and they were recorded in the Compendium of Materia Medica (Chinese: 木草綱目), which was written during the 16th century (Li and Luo, 2003). Almost all the uses were the result from centuries of experimenting by trial and error, which in some cases had led to untimely deaths.

Over time, some of these medicinal plants have had their bioactivities confirmed and their bioactive compounds identified. For example, *Coptidis rhizoma*, which is widely known as *Huanglian* in China and is used to treat heat dissipation and detoxification in herbal medicine (Tan et al., 2011). This herb produces berberine that possesses diverse pharmacological activities including cytotoxicity in cancer cells (Tan et al., 2011). Lyciumins, a group of natural products that inhibits the angiotensin-converting enzyme and renin, was

isolated from *Lycium barbarum*, which had been employed for hypertension treatment in Chinese medicine (Yahara et al., 1989). Yet, the majority of natural products from traditional medicinal plants still remain uncharacterised, including the question of whether they are actually bioactive or only bring placebo effect.

While plants were dominating in traditional medicine since ancient times, prokaryote and fungi made their revolution as a source of pharmacologically active substances in a more recent era. Perhaps the most famous example is penicillin, an antibacterial substance produced by *Penicillium rubens* CBS 205.57, formerly *Penicillium rubrum,* first reported in 1929 (Fleming, 1929, Houbraken et al., 2011, Gaynes, 2017). In fact, the application of antibiotics was documented before the discovery of penicillin. Pyocyanase, a filtered extract from *Bacillus pyocyaneus* (later renamed *Pseudomonas aeruginosa*), was clinically utilised as an antibiotic treatment in the 19th century, although the bioactive substance was only identified as pyocyanine decades later (Emmerich and Löw, 1899, Caltrider, 1967). It can be easily rationalised that those antibiotics are produced by the microorganisms to compete with other bacteria for survival. More interestingly, some secondary metabolites actually have antitumour effects, which have helped address one of the other main therapeutic problems, cancer treatment. For example, Coley's toxins, the supernatants of *Streptococcus pyogenes* and *Serratia marcescens* were studied in the late 19th century and utilised for tumour treatments later (Coley, 1910). Other examples include actinomycin D, bleomycin and doxorubicin (Vietti and Valeriote, 1971, Wit et al., 1997, Carvalho et al., 2009). It is interesting that microbial natural products appear to have huge potential as both antimicrobial and antitumour treatments. One potential explanation for antitumour production by bacteria is that the antitumour compounds are cytotoxins that are produced

to target eukaryotic and prokaryotic threats, as many anticancer agents are cytotoxins, including etoposide, mitomycin C, saliniquinones, as well as actinomycin D, bleomycin and doxorubicin that have been mentioned previously (Baldwin and Osheroff, 2005, Sasaki et al., 2006, Murphy et al., 2010).

With the current public health crisis of accelerating development of antimicrobial resistant pathogens and the ongoing challenge of treating cancer, natural products are an excellent source for medicine discovery. A recent study shows that bacteria contain highly biosynthetically diverse phyla and only approximately 3% of the genomic potential for natural product biosynthesis from bacteria has been currently experimentally identified (Gavriilidou et al., 2021). This encouraging result indicates that bacterial natural products have huge potential for novel bioactive compound discovery. Discovery and identification of natural products also provides in-depth knowledge on the chemistry and mode of action the compounds, such as their biosynthetic pathways and interaction mechanisms, which can be used in other areas such as chemical engineering and medicinal chemistry.

1.1.3. Actinobacteria as producers

This thesis focusses on natural products from the actinobacteria phylum of bacteria. Actinobacteria are filamentous high Guanine + Cytosine content Gram-positive bacteria ubiquitous in nature that can be found in diverse habitats such as soil, limestone, freshwater, marine, desert volcanic cave-hot spot, air, insect guts, among many others. Actinobacteria are renowned for producing a remarkable number of very diverse bioactive secondary metabolites, including antibiotics, anti-fungal molecules, anti-inflammatory agents, antiviral agents, and antitumor compounds (Tompsett and McDermott, 1949, Sehgal et al., 1975, Oh et al., 2009, Raveh et al., 2013). For example, actimycin A1a isolated

from a marine actinobacteria *Streptomyces kaviengensis* was found as a potent antiviral agent against a broad spectrum of RNA viruses, and rapamycin produced by *Streptomyces hygroscopicus* was reported to be an multi-functional bioactive agent with immunosuppressive, anti-inflammatory and antifungal activity (Raveh et al., 2013, Liu et al., 2019b). In fact, over 65% of medicinal antibiotics are produced by actinobacteria, in particular by the members of the genus *Streptomyces*, including the example of actinomycin D, chloramphenicol, doxorubicin, and rapamycin (Vietti and Valeriote, 1971, Sehgal et al., 1975, Bartlett, 1982, Bérdy, 2005, Carvalho et al., 2009). Although thousands of natural products have been identified from actinobacteria, bioinformatic study showed a massive potential for novel secondary metabolites due to the presence of cryptic or silent genes clusters in all microbes, including actinobacteria (Gavriilidou et al., 2021).

1.1.4. Classes of natural products

Given their chemical diversity, natural products are classified based on their key structural features and/or biosynthetic pathways (Figure 1.1). For example, one of the most commonly consumed natural products, caffeine, is a member of the alkaloid class, low molecular weight nitrogen-containing heterocyclic compounds, which can have very diverse biosynthetic origins (Rosa et al., 2007, Singh et al., 2017). Another major class of natural products is the terpenes. These compounds are biosynthesised from five-carbon building blocks named isoprenes by enzymes called terpene cyclases (Louie et al., 2020). One of the best-known examples of this class is the multiple bioeffect agent menthol from *Mentha canadensis* (Kamatou et al., 2013).

The pioneering antibiotic penicillin, belonging to the class of β-lactam antibiotics, is a non-ribosomal peptide (NRP) natural product. NRPs are peptide molecules whose scaffolds are

produced by non-ribosomal peptide synthetases (NRPSs) (Felnagle et al., 2008). Another group of natural products classified by their biosynthetic origin is polyketides (PKs), molecules formed via the condensation of acyl-thioester units such as malonyl-CoA and methylmalonyl-CoA, which are synthesised by polyketide synthases (PKSs) (Hertweck, 2009). Antibiotic erythromycin and antitumour agent doxorubicin, among many other clinically important natural products are examples of PKs. Last but not least, another class of peptide natural products is the ribosomally synthesised and post-translationally modified peptides (RiPPs), whose scaffold derives from a genetically encoded peptide that is later modified by the tailoring enzymes in its pathway. Since RiPPs are the main focus of this thesis their biosynthesis will be discussed in more detail in following sections.

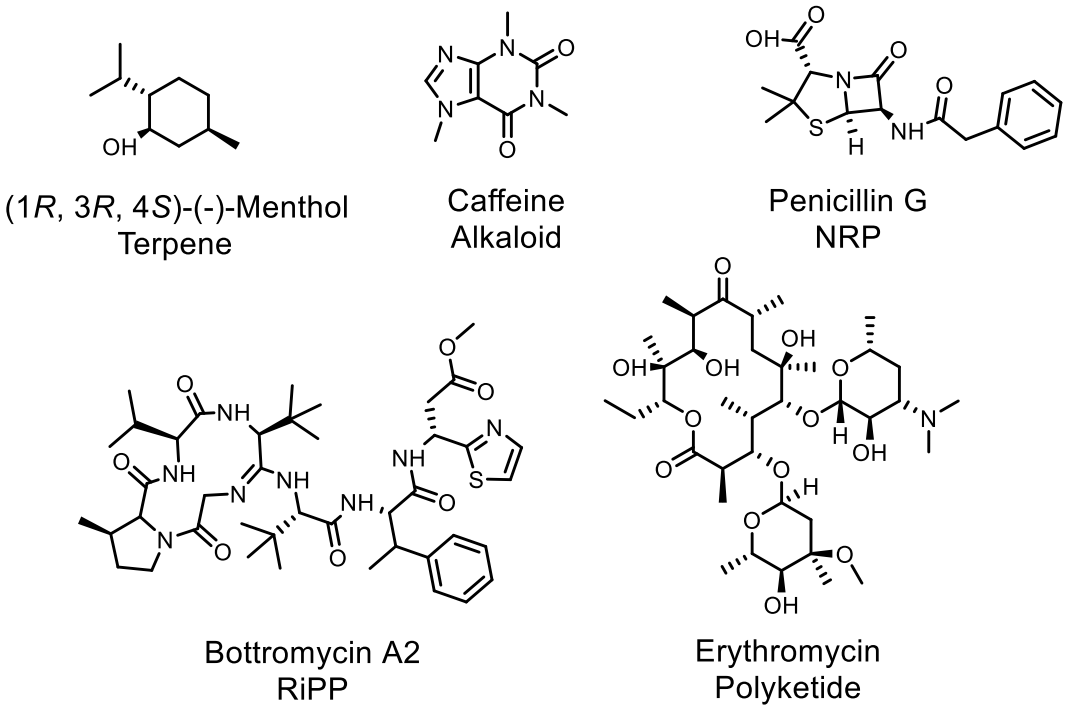


Figure 1.1: Examples of natural products produced by five major biosynthetic routes. Terpene (menthol), alkaloid (caffeine), NRP (penicillin G), RiPP (bottromycin A) and polyketide (erythromycin).

1.1.5. Natural product biosynthesis and biosynthetic gene clusters

The biosynthesis of PKs, NRPs and RiPPs are somehow similar in a way that a backbone scaffold is first formed, and further enzymatic tailoring modifications are added afterward. The backbone formation for PKs and NRPs employ very similar biosynthetic machinery, which involves multidomain modular enzymes (PKSs and NRPSs). PK backbone biosynthesis is dictated by PKSs. Within each module in a PKS, it contains at least 3 essential domains, which are an acyltransferase (AT), acyl carrier protein (ACP) and ketosynthase (KS) domain, and in line with additional tailoring domains to introduce primary structural diversity (Hertweck, 2009). The formation of NRP scaffolds is also through multidomain modular enzymes NRPSs, which contain 3 essential domains (adenylation domain, thiolation domain and condensation domain), an epimerisation domain to convert amino acids between L- and D-isomers and a C-terminal thioesterase domain to promote cyclisation or hydrolysis (Felnagle et al., 2008). Meanwhile, the precursor in RiPPs is encoded in the genome and translated into a short peptide. Once the peptide scaffold is formed, it undergoes a series of subsequent enzymatic tailoring modifications to introduce further diversity in structure.

Enzymes for tailoring modifications, backbone scaffold for PKs and NRPs and the precursor peptide for RiPPs are all genetically coded in the chromosome of the producers. These genes are often found adjacent to another in bacterial derived natural product biosynthetic pathways, as well as the genes for regulatory proteins and transporters, to form biosynthetic gene clusters (BGCs) (Medema and Fischbach, 2015). For example, the actinomycin BGC encoded 28 genes including regulatory and NRPS genes for the biosynthesis of actinomycin, while the BGC for the RiPP bottromycin A_2 consists of 13 genes,

including a precursor peptide gene, genes encoding modifying enzymes, and regulatory and transporter genes (Smith et al., 2006, Franz et al., 2021).

1.1.6. The importance of tailoring modifications

Posttranslational modifications (PTMs) affect the structure and dynamics of proteins and peptides, and the chemical properties of modified sites can be influenced by PTMs on a single type of amino acid or multiple amino acids (Mann and Jensen, 2003, Huang et al., 2018). Therefore, modifications often play an important role in structures and functions as they change a wide range of proteins/peptides behaviours and characteristics, including protein solubility and biomolecular interactions. For example, analogues of rapamycin can possess improved pharmacological properties, derivatives of nystatin A_1 given by various levels of PTMs have different antifungal and hemolytic activities, and a similar effect was also observed in thioamitide analogues when a single modification at the N-termial moiety from thioviridamide to JBIR-140 displayed doubling the bioactivity against Jurkat acute T cell leukemia cells, and the oxidation on the methionine in thioholgamides reduced the bioactivity 10 times against Jurkat cells (Figure 1.2) (Brautaset et al., 2008, Gregory et al., 2013, Izumikawa et al., 2015, Kjaerulff et al., 2017). In some cases, PTMs can also stabilise the peptide or protein, for instance a thioamide bond, a single atom replacement, is capable of stabilising a peptide from proteolysis (Chen et al., 2017).

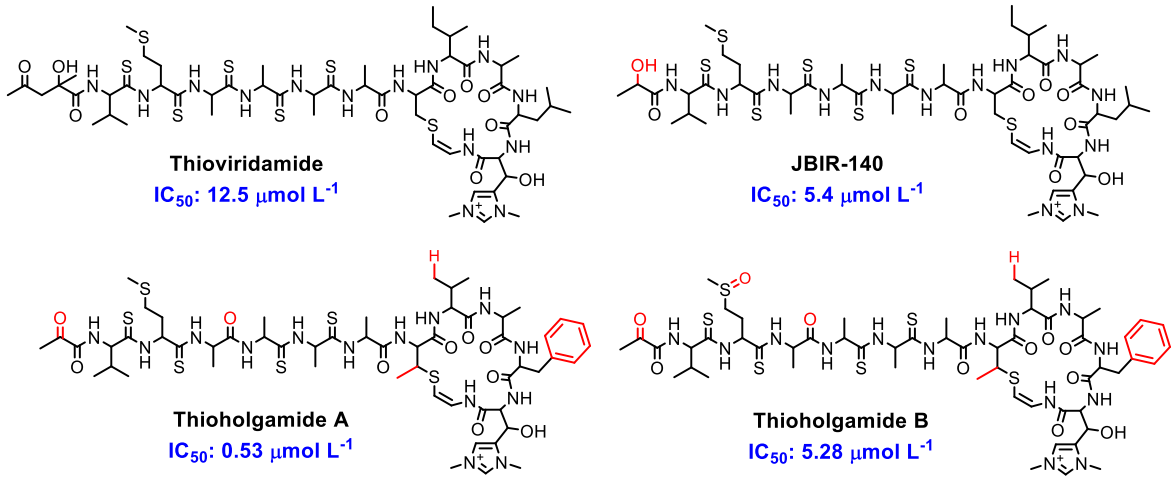


Figure 1.2: Examples of the variation in the biological activity of natural products derived from changes in their structures by diverse tailoring modifications. Structures of thioviridamide and analogues with their anticancer activities. Half maximum inhibitory concentration (IC_{50}) was tested against Jurkat acute T cell leukamia cell line by Izumikawa et al (2015) and Kjaerulff et al (2017). The structural differences with respect to thioviridamide are labelled in red. The stereochemistry of the molecules are not characterised.

1.2 RiPPs

1.2.1. Biosynthetic origin of RiPPs

RiPPs is the most recently discovered class of natural products and comprises gene-encoded peptide natural products as mentioned in section 1.1.5. Before their biosynthetic pathways were characterised and this common nexus between them identified, these compounds were usually classified based on either their producing organisms or biological activities (Klaenhammer, 1993, Duquesne et al., 2007).

As it is stated in its name, biosynthesis of RiPPs initiates with the translation of a precursor peptide that it is longer than the final peptide molecule (Figure 1.3) (Arnison et al., 2013). Often, the short gene encoding this peptide is annotated as gene "A" of its BGC, unless an old designation has been used. The precursor peptide consists of a "recognition" region for binding of biosynthetic proteins, and a core peptide dictating the structure of the final product.

Save for some exceptions, the recognition region can either be located in the N-terminus or C-terminus of the precursor peptide and is termed leader peptide or follower peptide respectively (Arnison et al., 2013). Leader peptides were found to interact with a conserved domain in RiPP modifying enzymes. This conserved domain, known as RiPP recognition element (RRE), was found in the forms of small peptides or fused to a modifying enzyme/tailoring enzyme complex, for instance the adenylating enzyme MccB in microcin C7 BGC, and cyclodehydratase LynD in cyanobactin BGC (Burkhart et al., 2015, Kloosterman et al., 2021). The leader and follower peptide regions are also thought to play a role in self-immunity by reducing the biological activity, stabilising the intermediate peptide from

degradation before the RiPP maturation, and are important for secretion (Oman and van der Donk, 2010).

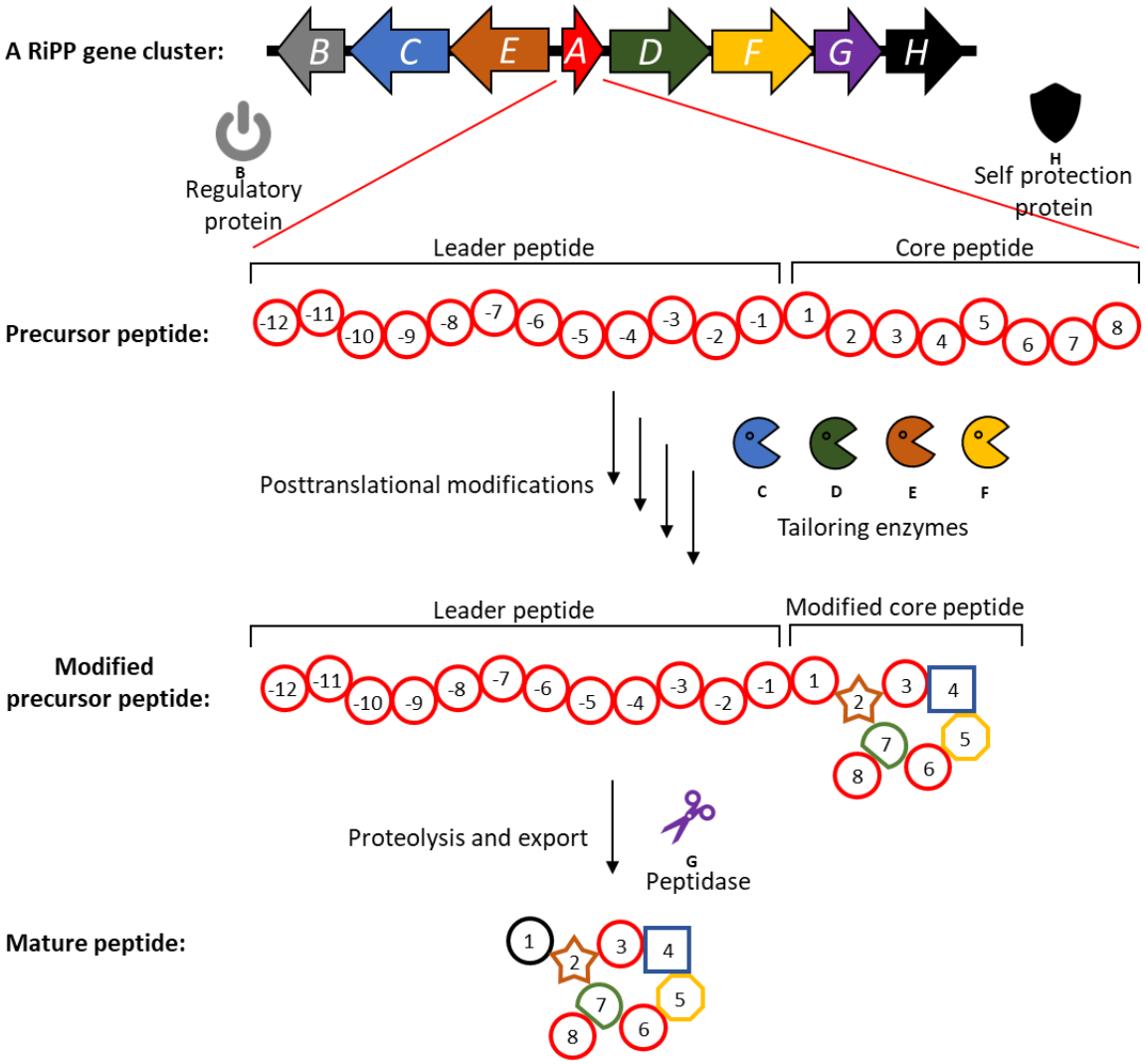


Figure 1.3: Schematic representation of RiPP biosynthesis. A ribosomally synthesised peptide is modified by specific tailoring enzymes (C, D, E and F) and the modified core peptide is then cleaved by a protease (G) to generate the mature peptide that is exported out of the cell. Some RiPP clusters also encode regulatory proteins (B) and self-protection proteins (H). Red circles represent unmodified amino acid residues, and modified amino acids are labelled in other colours and shapes.

The core peptide region constitutes the scaffold of the final molecule. This region is where PTMs are installed and is excised by proteolysis to generate a mature peptide (Oman and van der Donk, 2010). In some RiPP natural products, a single precursor peptide gives multiple mature RiPP products as the precursor peptide includes multiple core peptides.

For example, precursor peptide UstA of ustiloxn B from *Ustilaginoidea virens* contains 16 repeats of YAIG core and KR motifs for recognition, and precursor peptide LbaLycA from *L. barbarum* possesses 12 repeats of core peptide regions for all three lyciumin A, B and D (Umemura et al., 2014, Kersten and Weng, 2018, Luo and Dong, 2019). This was also seen in actinobacteria, for instance thiovarsolin from *Streptomyces varsoviensis*. (Figure 1.4) (Santos-Aberturas et al., 2019). In other cases, a single RiPP BGC can include several precursor peptide genes, each encoding different core sequences, such BGC examples can be found in the cyanobacterium *Prochlorococcus* MIT9313 for the production of multiple lanthionine-containing peptides (lanthipeptides) called the prochlorosins (Li et al., 2010).

As well as a precursor peptide, RiPP BGCs encode multiple tailoring enzymes, responsible for introducing the PTMs in the molecule. As mentioned earlier in this section, one of the purposes for the existence of leader peptides is to interact with RREs which can be part of the modifying enzymes. Hence, tailoring enzymes can be directed by leader peptides to introduce specific PTMs, such as the leader peptide-dependent thioether bond formation by radical *S*-adenosyl-L-methionine (SAM) enzyme in the biosynthesis of subtilosin A (Flühe et al., 2012). However, some modifications are not dependent to the leader peptide, examples include the flavin dependent enzymatic oxidation in the cyanobactin biosynthetic pathway (Gao et al., 2018).

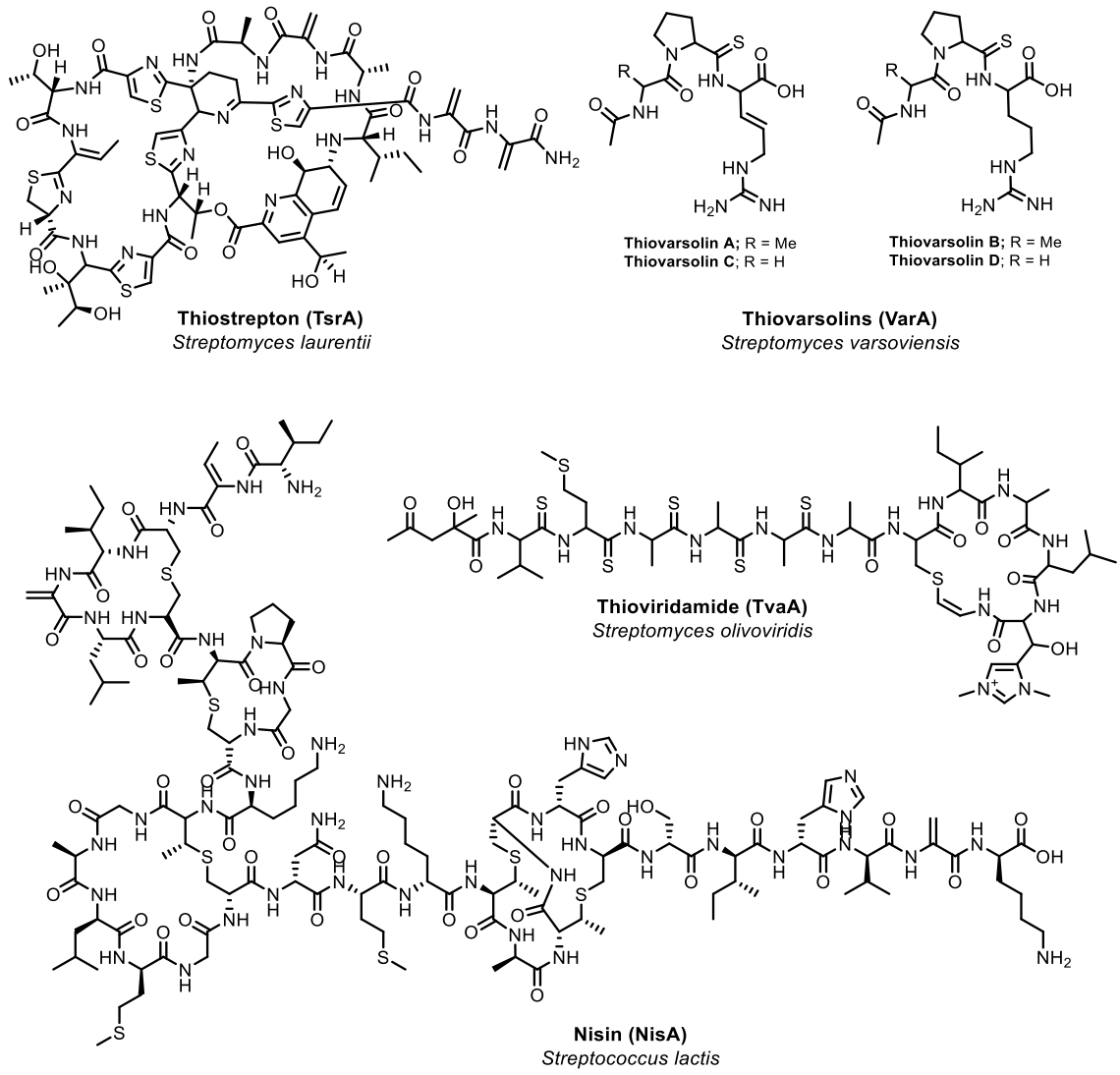
In addition to the core genes for precursor peptide(s) and tailoring enzymes, regulatory proteins, transporters and self-protection proteins are often seen in the BGC. Since secondary metabolites are normally bioactive and with a potential for being lethal to the producer too, self-immunity is essential for the producer. Hence, genes for self-protection proteins are sometimes found in a BGC, for instance the immunity protein McbG in

microcin B17 (MccB17) BGC to prevent MccB17 binding to the target DNA gyrase (Collin and Maxwell, 2019). Another way to achieve self-immunity can be through a physical mechanism, such as exporting the compound from the host. MccB17 producer *Escherichia coli* encodes genes for McbE and McbF, which control the efflux of MccB17 as part of the immunity mechanism (Collin and Maxwell, 2019). While proteolysis and exportation are normally achieved by independent proteins, bifunctional enzymes that proteolyse and export are seen in many class II lanthipeptide biosynthetic pathways. One example is a protein encoded in the BGC of lactococcin DR from *Lactococcus lactis* subsp. *lactis*, which was reported features high level of sequence similarity to ATP-dependent transporters and homology to C39 family of papain-like Cys proteases (Rince et al., 1994, Rawlings et al., 2013). These features supported that leader peptides might play a role in self-immunity.

Since RiPPs are derived from a ribosomal peptide, their peptidic scaffold is originally produced from the twenty amino acid building blocks, as opposed to what happens with NRPs, that often contain non proteinogenic amino acids (Finking and Marahiel, 2004). However, there are multiple ways to introduce chemical diversity in RiPPs. For example, in some cases the core peptide that shapes the backbone of the mature peptide can be highly tolerant to mutation, therefore introducing the first level of chemical diversity (Arnison et al., 2013). In addition, their PTMs can introduce multiple different structural features to the peptide backbone to create a further chemical complexity, which consequently liberate the limitation in structural diversity by canonical proteinogenic amino acids (McIntosh et al., 2009). Finally, changing the stereochemistry of the proteinogenic amino acids can also further diversify the final structure, so the configuration of the amino acid can be converted from L to D. Radical SAM enzyme PoyD in the polytheonamide A biosynthetic pathway

introduces epimerisation of α-carbons of residues in maturing RiPPs (Parent et al., 2018). Epimerisation is also seen in many other natural products, such as bottromycin A_2, carnolysin, bicereucin and lacticin 3147 (Martin et al., 2004, Lohans et al., 2014, Huo and van der Donk, 2016, Franz et al., 2021). There is a very wide spectrum of PTMs described in RiPP biosynthesis and these are usually employed to define different families of RiPPs.

(a)



(b)

```
>TsrA(ACN52291.1)
MSNAALEIGVEGLTGLDVDTLEISDYMDETLLDGEDLTVTMIASASCTTCICTCSCSS
>TvaA(BBI93430.1)
MTEKTQITDVQAFEDLVAKVQEMDGPAQASSTVAALAGLDAAELQNFLEEKSGISPDEEAQGSVMAAAASIALHC
>VarA(WP_030877127.1)
MRKDSDLRLDDLAHVNPEQLQGFLEERSTAAVHGATDAYFSGPRAQTVRGATDAHFSAPRAQTVRGATDAHFSAPRA
QTVRGATDAHFSAPRAQTVRGATDSYFSAPRAQNVR
>NisA(AAA25189.1)
MSTKDFNLDLVSVSKKDSGASPRITSISLCTPGCKTGALMGCNMKTATCHCSIHVSK
```

Figure 1.4: Examples of RiPPs and their corresponding precursor peptides. (a) Structures of thiopeptide (thiostrepton from *S. laurentii)*, derivatives of a thioamidated RiPP (thiovarsolin A-D from *S. varsoviensis*), a thioamitide (thioviridamide from *S. olivoviridis*) and a lanthipeptide (nisin from *S. lactis*). (b) Precursor peptide sequences for the corresponding RiPPs in section a, (accession numbers in brackets), with their core peptide regions highlighted in red. The stereochemistry of thiovarsolines and thioviridamide are not yet determined.

1.2.2. RiPP PTMs and classification

Since the only common biosynthetic feature to all RiPPs is having genetically encoded precursor peptides and both these and their biosynthetic pathways are not conserved across the whole group, RiPPs are often subdivided based on their key structural features and tailoring modifications. For instance, the very widespread class of RiPPs lanthipeptides contain one or more β-thioether crosslinked bis amino acids lanthionine and methyllanthionine. Lanthipeptides are subsequently divided into 5 classes depending on the enzymes involved in the lanthionine formation (Montalbán-López et al., 2021). However, not all RiPPs contain a single key PTM. Instead, some RiPP families possess a combination of multiple rare modifications that make them unique to other RiPP natural products. Thioamitides feature an N-terminal pyruvyl-like moiety, thioamide bonds, an *S*-[(*Z*)-2-aminovinyl](-3*S*)-3-methyl)-D-cysteine (Avi(Me)Cys) macrocycle and a β-hydroxy-N^1,N^3-dimethylhistidinium (hdmHis) (Hayakawa et al., 2006b). For example, the Avi(Me)Cys structure was previously observed in lanthipeptides and linaridins, and thioamide bonds are rare in natural products, but can also be found in thiopeptins and methanobactin (Kenney and Rosenzweig, 2012, Götz et al., 2014, Ichikawa et al., 2018, Ma and Zhang, 2020).

1.2.3. RiPPs and Genome-mining

Thanks to the development of bioinformatic tools, RiPPs are a rapidly expanding group of natural products. The lower costs of sequencing and the use of genome mining techniques have unveiled hundreds of RiPP clusters previously unknown and that would not have been easy to discover due to the lack of unifying features of RiPPs, such as pearlins, spliceotide,

tryglysins, biarylitides and thiovarsolins (Morinaka et al., 2018, Santos-Aberturas et al., 2019, Ting et al., 2019, Rued et al., 2021).

Natural product classes such as NRPs and PKs are produced by multi-modular complexes that provide an obvious feature for genome-mining. On the other hand, the machinery of RiPP biosynthesis only requires a precursor peptide and various PTM enzymes. Hence, genome-mining of RiPPs can be challenging. First of all, precursor peptides are often overlooked by bioinformatic tools and not annotated as they are very short (often less than 50 amino acids) (Russell and Truman, 2020). Also, there is not a single specific tailoring enzyme or domain protein conserved throughout RiPP natural products, but specific enzymes are often conserved within a RiPP family if they catalyse the PTM characteristic of that family. Thus, genome-mining through a specific PTM enzyme could be employed as an alternative strategy to discover hidden biosynthetic pathways. For example, the genome-mining tool RiPP Precursor Peptide Enhanced Recognition (RiPPER) developed in our lab revealed that YcaO and TfuA protein pairs, which catalyse the rare PTM thioamidation, are widespread in actinobacteria, and the study led to a discovery of a novel RiPP family, thiovarsolins (Santos-Aberturas et al., 2019). Hence, it is important to understand RiPP biosynthesis and to identify the key features of RiPPs, as they provided the logic of how to link their key chemical features to their BGCs for discovery of hidden biosynthetic pathways. There are also other genome mining tools for RiPP BGC discovery. Following a similar genome-mining strategy, Rapid ORF Description and Evaluation Online (RODEO) also enables users to define a tailoring enzyme for potential BGC discovery across species (Tietz et al., 2017). Using RODEO, the BGC of a novel thiopeptide saalfelduracin was discovered and characterised from *Amycolatopsis saalfeldensis* (Schwalen et al., 2018). On the other

hand, the most widely used genome-mining tool, antibiotic and Secondary Metabolite Analysis Shell (antiSMASH), searches a query sequence for BGC by comparing encoded gene products with a library of hidden Markov models that describe a range of biosynthetic genes for natural products (Blin et al., 2019).

1.3 Thioamitides

1.3.1. Biological activities

The first reported thioamitide was thioviridamide from *Streptomyces olivoviridis* NA05001 through a screen of cytotoxin producer in 2006 (Hayakawa et al., 2006a). After examining the thioviridamide bioactivity against different 3Y1 rat normal fibroblasts, thioviridamide was suggested to be a potential anticancer agent against oncogene adenovirus E1A gene containing cell lines. Cells were particularly sensitive to the compound regardless the presence of E1B gene which inhibits apoptosis to prolong host cell viability (Debbas and White, 1993, Hayakawa et al., 2006a).

The bioactivities of thioamitides were further confirmed upon the discovery of a novel thioamitide thioalbamide from *Amycolatopsis alba* DSM 44262. A multifactor comparison experiment was carried out for a more comprehensive thioalbamide anticancer activity test in the lab. Comparing the bioactivities of thioalbamide in a non-tumour breast epithelial cell line and a selection of cancer cell lines, it was learned that thioalbamide was highly specific to all tested tumour cell lines at nanomolar concentrations (Frattaruolo et al., 2017). A clinically used doxorubicin was also used to compare their anticancer activities, and the specificity towards cancer cells by doxorubicin was not as remarkable as in the treatments of thioalbamide despite the fact that both compounds showed similar bioactivities against non-tumour epithelial cells (Frattaruolo et al., 2017). In other words, both doxorubicin and thioalbamide showed similar cytotoxicity toward healthy cells, but thioalbamide showed a greater cytotoxic margin between healthy cells and cancer cells. A recent *in vivo* study also indicated that the *in vitro* bioactivities given by thioalbamide were observed in a mouse model which had higher complexity experimental conditions

(Frattaruolo et al., 2023). This encouraging finding suggested a promising pharmaceutical potential for cancer treatment.

It was revealed that the cytotoxic activity from thioalbamide was linked to oxidative stress as the cytotoxicity was abolished when antioxidant vitamin E or *N*-acetyl cysteine was added to thioalbamide treatments, and thus the increase in reactive oxygen species (ROS) production appeared to be caused by dysfunction of energy metabolism (Frattaruolo et al., 2019). It was consequently investigated and reported that the cytotoxicity from thioalbamide could occur through inhibiting the activity of F_0F_1-ATPase complex, which mediated cellular respiration in mitochondria to reduce ROS (Frattaruolo et al., 2023). This finding of F_0F_1-ATPase inhibiting activity was consistent to the finding in the study of prethioviridamide bioactivity mechanism (Takase et al., 2019). This extensive study on thioalbamide showed that not only did thioalbamide induce the apoptotic cancer cell death, but also potentially inhibited the cancer cell invasion and tumour metastasis through suppressing the secretion of matrix metalloproteinase-9 (Frattaruolo et al., 2023).

Not only the mechanism of cytotoxicity showed potential for cancer treatments, but also the relationship between the structure and bioactivity of the compounds (Figure 1.2). For example, thioviridamide, JBIR-140 and thioholgamide A were found to have different half-maximal inhibitory concentration (IC_{50}) in different cancer cell lines (Tang et al., 2018). While thioviridamide and thioaholgamide A are structurally more different, JBIR-140 only has a minor change in structure, an N-terminal lactyl moiety rather than a 2-hydroxy-2-methyl-4-oxopentanoyl group in thioviridamide, indicating that subtle structural modifications can greatly affect the bioactivity of thioamitides.

Since there is substantial evidence indicating the high antitumour potential both *in vitro* and *in vivo* of the thioamitides, as well as the interesting bioactivity changes through trivial structural variations, there is a strong need to explore their diversity and their medicinal potential.

1.3.2. Biosynthetic pathways and genetics

As mentioned in the previous text, thioamitide is a family of RiPP natural products. The BGC of thioviridamide (Figure 1.5) was confirmed by Izawa *et al.* to contain 12 potential genes by heterologously expressing the candidate genes in *Streptomyces lividans* TK23 and *Streptomyces avermitilis* SUKA47 (Izawa et al., 2013, Izawa et al., 2018). Along with the 75-amino acid long precursor peptide (TvaA), the BGC consists of 9 genes for tailoring enzymes including a protease, one gene for membrane protein, and one regulatory gene, based on the analysis on the homologous proteins (Izawa et al., 2013). Genes *H* and *I* were proposed to introduce 5 thioamide bonds into the N-terminal linear chain (Section 1.3.3). Gene *F* shares similarity to phosphoantothenoylcysteine decarboxylase proteins, while genes *C* and *E* exhibit homology to aminoglycoside phosphotransferases. Gene *D* was later identified to be a HopA1-like lyase, since it showed homology to HopA1 effector protein employed to aid plant infection. These four enzymes introduce serine/threonine dehydration and Avi(Me)Cys bond to the molecule (Section 1.3.4) (Izawa et al., 2013, Eyles et al., 2021). Gene *G* (Section 1.3.6) and gene *J* are a putative methyltransferase and a putative oxygenase, respectively, for the formation of β-hydroxy-N^1, N^3-dimethylhistidinium (hdmHis) (Izawa et al., 2013).

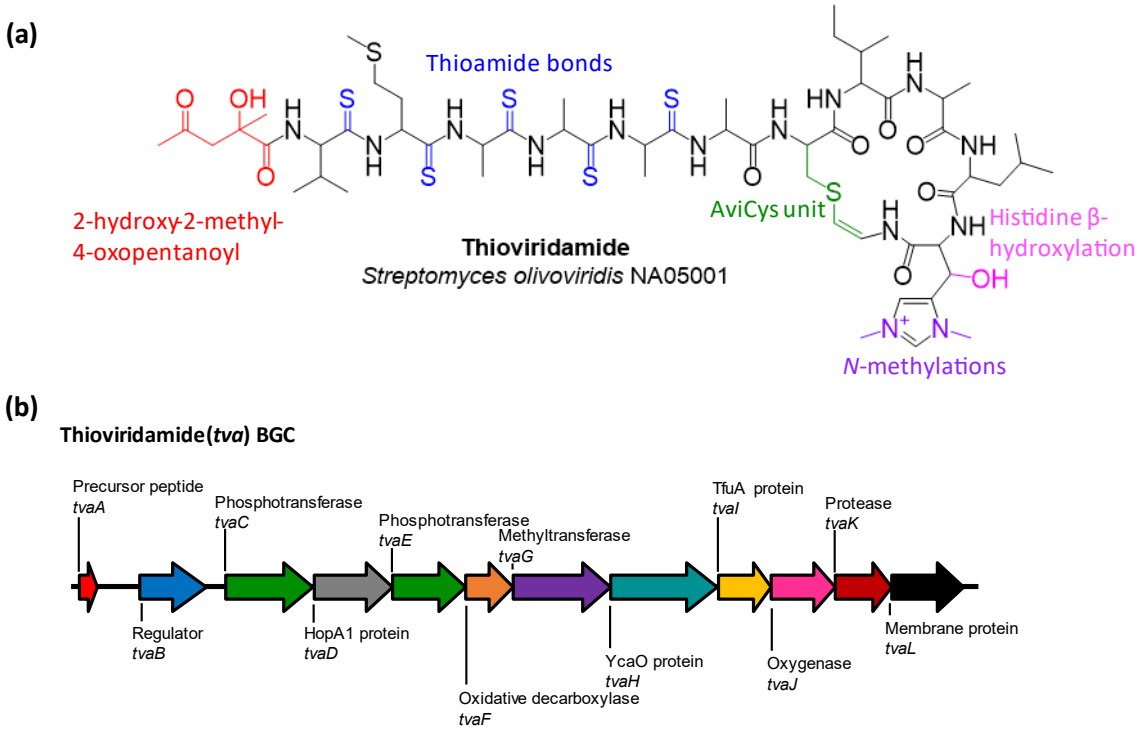


Figure 1.5: Thioviridamide and its BGC. (a) The structure of thioviridamide and the PTMs installed in the molecules are drawn in colours, in which red for 2-hydroxy-2-methyl-4-oxopentanoyl, blue for thioamide bonds, green for AviCys unit, purple for *N*-methylations and magenta for β-hydroxylation in histidine. (b) Thioviridamide BGC.

There was a 7-year gap between the discoveries of thioviridamide and the responsible BGC, and there was no novel thioamitides being reported in between. Either there were difficulties for novel thioamitide production or little interest for thioamitide discovery. However, a few years after identifying the thioviridamide BGC, 14 thioamitide BGCs were discovered from bacterial genomes through a genome-mining study from the Truman lab (Frattaruolo et al., 2017). Upon the discovery of these thioamitide BGCs, three thioamitides were discovered and characterised in the same study, and more thioamitides were discovered shortly after the findings from the genomic-based approach study (Izumikawa et al., 2015, Frattaruolo et al., 2017, Kjaerulff et al., 2017, Izawa et al., 2018, Lu et al., 2021). The discovery of novel thioamitides was therefore facilitated by the genome-mining effort.

Unsurprisingly, the precursor peptides are highly conserved across BGCs, and each thioamitide precursor peptide contains a 13 amino acid core peptide region (Frattaruolo et al., 2017). The identified BGCs are conserved amongst the discovered pathways too, as majority of the genes in a thioamitide pathway could be found in other pathways, but the BGCs also showed a degree of flexibility for encoding additional participating tailoring enzyme genes to introduce further chemical diversity. For example, all identified BGCs enocde a YcaO domain containing protein, but only a few of them encode a methyltransferase (*tsaMT*) homologue that was found essential for the biosynthesis of thiostreptamide S4 (Eyles et al., 2021).

1.3.3. Thioamide bonds and YcaO-domain proteins

One of the key features in thioamitides is the thioamide bonds, which are exceptionally rare in currently known natural products (Izawa et al., 2013). Examples of those rare thioamide containing natural products are mostly reported from bacteria and include apo-methanobactin, closthioamide, saalfelduracin, thiopeptin B, thiovarsolin and thioamitides (Krentz et al., 2010, Lincke et al., 2010, Izawa et al., 2013, Schwalen et al., 2018, Liu et al., 2019a, Santos-Aberturas et al., 2019). The formation of a thioamide bond is an isosteric substitution of the carbonyl oxygen in an amide group to a sulphur (Patani and LaVoie, 1996). This isosteric substitution introduces major changes to the hydrogen bonding interactions between the carbonyl/thiocarbonyl and amide, and consequently improves proteolytic stability, conformational rigidity and can alter bioactivity (Lee et al., 2002, Reiner et al., 2008, Zhang et al., 2010, Chen et al., 2017, Walters et al., 2017).

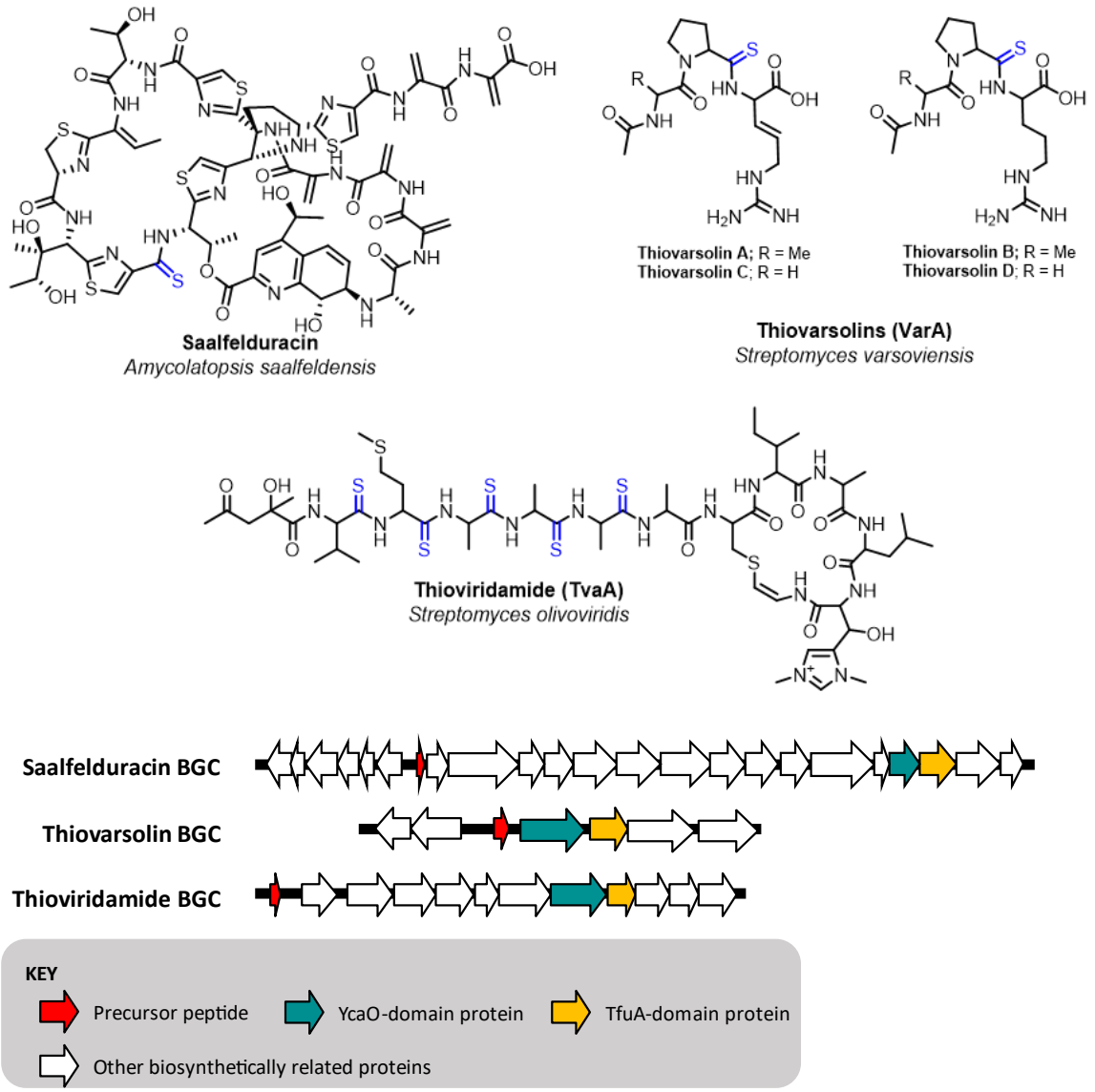


Figure 1.6: Examples of YcaO-TfuA derived thioamide containing compounds from RiPPs. (a) compound structures with the thioamide bonds drawn in blue. (b) corresponding BGC with precursor peptide gene, YcaO-domain protein and TfuA-domain protein highlighted. Size of genes are normalised within the gene cluster.

This intriguing modification is introduced by a YcaO-TfuA protein pair in the biosynthesis of saalfelduracin, thiopeptin B, thiovarsolin and thioamitides (Figure 1.6). The first evidence is that the co-expression of YcaO-TfuA pair from the biosynthetic pathway of saalfelduracin with thiostrepton BGC in *Streptomyces laurentti* generated a thioamidated thiostrepton Sch 18640 (Schwalen et al., 2018). The *in vivo* investigations on biosynthesis on thiovarsolin and a thioamitide (thiostreptamide S4) further supported that the YcaO-TfuA protein pair

in the corresponding biosynthetic pathways are essential to initiate the biosynthesis through thioamide bond installations (Santos-Aberturas et al., 2019, Eyles et al., 2021). However, the current knowledge on YcaO-dependent thioamidation in natural products is still extremely limited.

In fact, the current understanding of YcaO superfamily is that they share a common function in ATP utilisation but the ultimate function of YcaO proteins has not yet being entirely characterised and is being investigated (Dunbar et al., 2012, Dunbar et al., 2014). Besides thioamidation, YcaO proteins can have a role in two other PTMs according to the current knowledge, namely the formation of azolines and cyclic or linear amidines (Li et al., 1996, Kelly et al., 2009, Sivonen et al., 2010, Crone et al., 2016, Metelev et al., 2017, Russell et al., 2021). A YcaO-domain protein often requires a partner protein to complete the function, so the thioamide formation YcaO proteins are also known as TfuA-associated YcaO proteins. Based on evidence from an archaeal YcaO-TfuA protein pair, it was determined that the TfuA protein is involved in the sulphur recruitment by hydrolysing a proteinaceous sulphur donor ThiS and enhanced the affinity of the partner YcaO protein towards the modification substrate, yet very little was known about the YcaO protein itself, nor whether this activity is conserved in bacterial systems (Liu et al., 2021a).

The most extensively studied YcaO proteins are from linear azoline peptide pathways (Figure 1.7a). Most of the azoline forming YcaO protein require an E1-ubiquitin activating (E1-like) domain, either a discrete E1-like protein with a YcaO-domain protein or an E1-like domain fused YcaO protein. For example, an E1-like protein (McbB) and a YcaO protein (McbD) are responsible for the formation of thiazolines and oxazolines in microcin B17 from cysteine and serine residues, respectively; the azol(in)e formation in a thiopeptide

heterocycloanthracin was catalysed by an E1-like domain fused YcaO protein but an additional Ocin-ThiF-like protein was required (Li et al., 1996, Dunbar et al., 2015). In the azol(in)e formation, E1-like protein interacts with the leader peptide and regulates the ATPase activity of the YcaO-domain protein (Dunbar et al., 2014, Koehnke et al., 2015). Meanwhile, the YcaO-domain protein catalyses the core peptide modification by phosphorylating the carbonyl oxygen and therefore activating it into a leaving group (Dunbar et al., 2012). In contrast, the formation of thioazoline formation in bottromycin biosynthesis was dependent on a standalone YcaO protein, but the mechanism is less clear in comparison to the E1-like domain associated YcaO proteins (Crone et al., 2016).

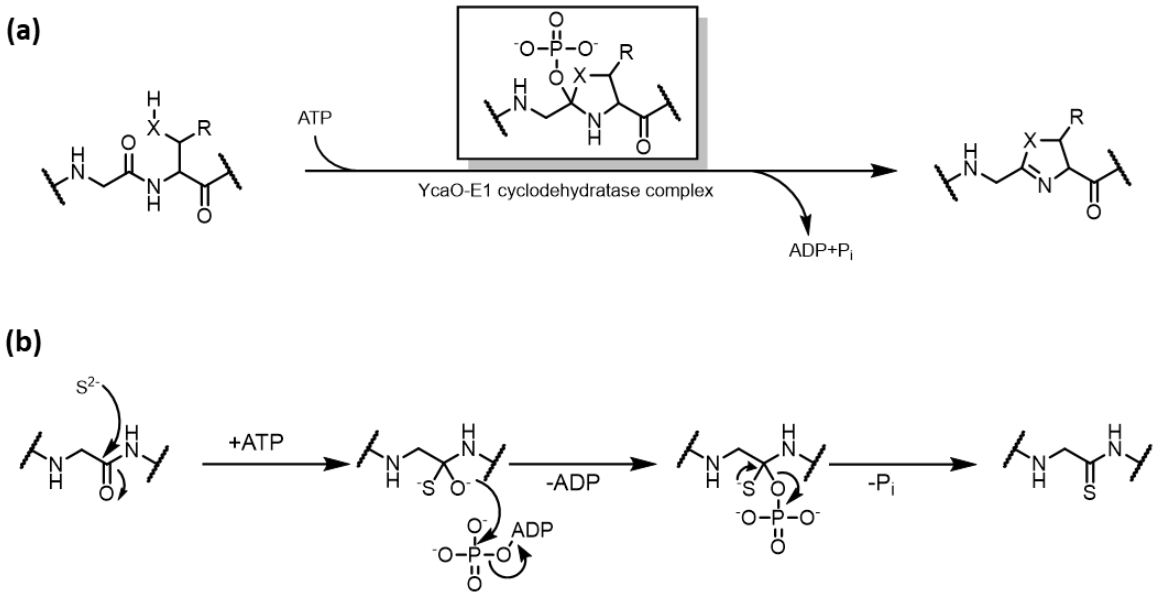


Figure 1.7: Schematic YcaO protein dependent reactions. (a) Dehydration step in Azoline heterocycle formation via cyclodehydratase complex. Where X = S, O; R = H, Me. (b) potential thioamide bond formation mechanism through YcaO-TfuA pair.

Through the current understanding on the azoline forming YcaO-E1 domain protein pair, the formation of thioamide bond through thioamide forming YcaO is proposed (Figure 1.7b).

1.3.4. Aminovinyl-cysteine macrocycle

Another structural characteristic in thioamitides is the presence of a C-terminal macrocycle. While not abundant, this PTM can be found in 4 RiPP families: lanthipeptides, lanthidins, linaridins and thioamitides (Figure 1.8) (Chatterjee et al., 1992, Hayakawa et al., 2006b, Claesen and Bibb, 2010, Götz et al., 2014, Wiebach et al., 2018). Most of these Avi(Me)Cys-containing molecules possess antibacterial activities, with the exception of thioamitides, as was discussed previously (Section 1.3.1). These Avi(Me)Cys-containing natural products exhibit high specificity for their molecular targets and the macrocycle appears to play a critical role towards their mode of action (Ongey et al., 2017). It has been suggested that the enhanced activity and selectivity of these molecules could be due to the peptide rigidity introduced by the unsaturated thioenamide in Avi(Me)Cys moieties (De Leon Rodriguez et al., 2018).

At the beginning of the study, there was very limited knowledge about the formation of the macrocycle in thioamitides. The breakthrough was found by a former lab member (Dr Tom Eyles) that TsaD encoded in the thiostreptamide S4 BGC showed homology to a HopA1 effector family domain, which could potentially act as a C-O lyase to catalyse an elimination of phosphorylated threonine, and the involvement of TsaD in dehydration was then confirmed by *in vivo* examination in *Streptomyces coelicolor* M1146 (Eyles et al., 2021). Whilst my PhD was ongoing, more studies on the dehydration and the formation of Avi(Me)Cys macrocycle were published, supporting the findings in the lab and providing more in-depth understanding of the roles of each protein involved in the biosynthesis.

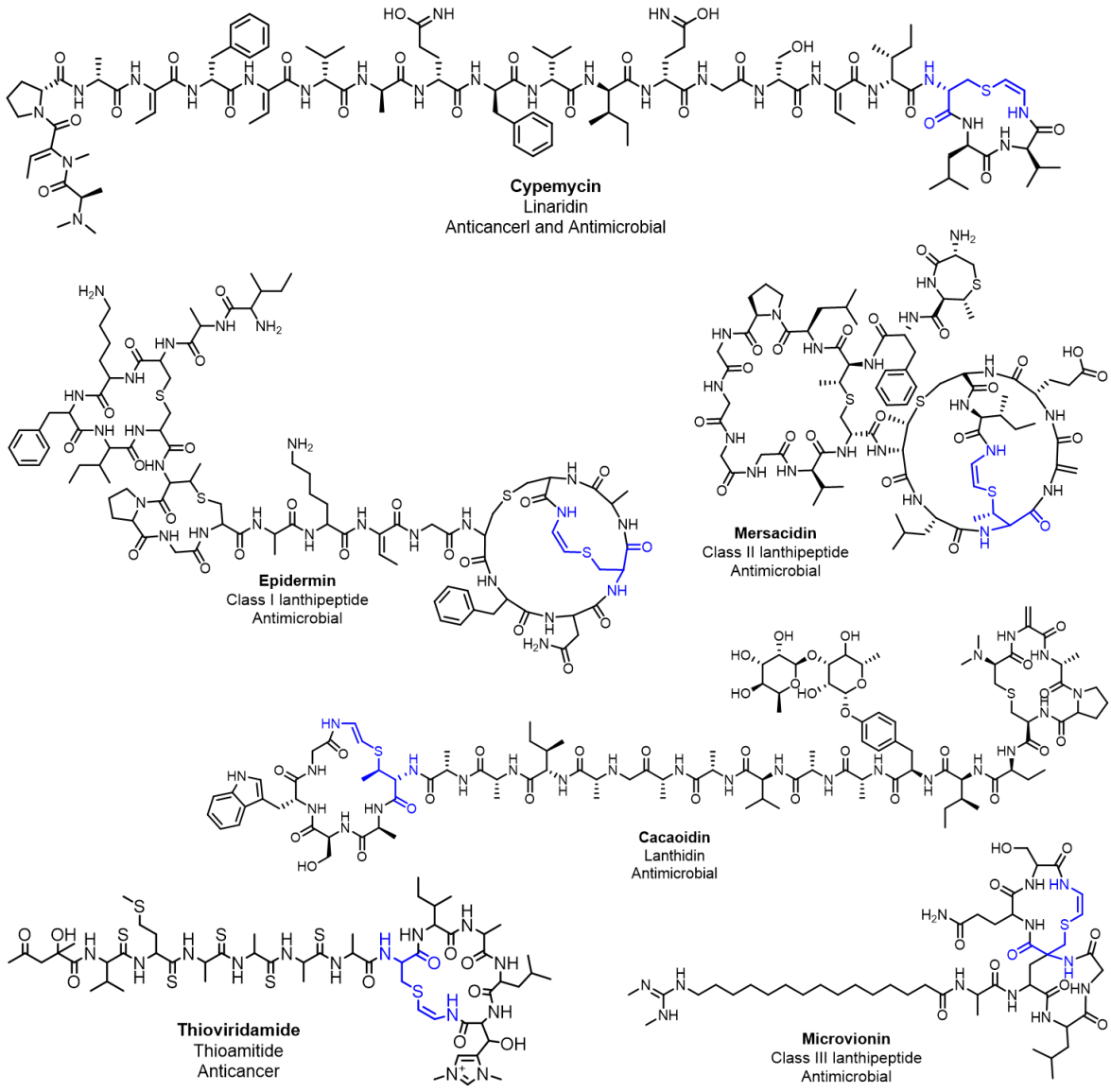


Figure 1.8: Examples of Avi(Me)Cys containing compounds from different RiPP families and their bioactivities, with the Avi(Me)Cys moiety drawn in blue. The stereochemistry of epidermin, thioviridamide and microvionin is not yet confirmed.

According to the study of the AviMeCys unit formation in thiostreptamide S87 biosynthetic pathway that was published in 2021, the formation of AviMeCys unit was driven by 4 of the tailoring enzymes in the biosynthetic pathways: 2 phosphotransferases (TsdC and TsdE), a HopA1-like lyase (TsdD) and a flavin-dependent cysteine decarboxylase (TsdF) (Qiu et al., 2021). The formation of the macrocycle is initiated by two independent enzymatic modifications, a dehydration at a given serine or threonine to form a dehydroalanine (Dha)

or dehydrobutyrine (Dhb) and a flavin-dependent cysteine decarboxylation to give an enethiol at the C-terminus of the peptide (Eyles et al., 2021, Qiu et al., 2021, Sikandar et al., 2022, Xiong et al., 2022). The cyclisation is then achieved by forming a thioether bond, generating either a *S*-[(*Z*)-2-aminovinyl]-D-cysteine (AviCys) or a (2*S*,3*S*)-*S*-[(*Z*)-2-aminovinyl]-3-methyl-D-cysteine (AviMeCys) unit depending on whether the dehydrated hydroxylated amino acid residue is a serine or a threonine, respectively (Figure 1.9).

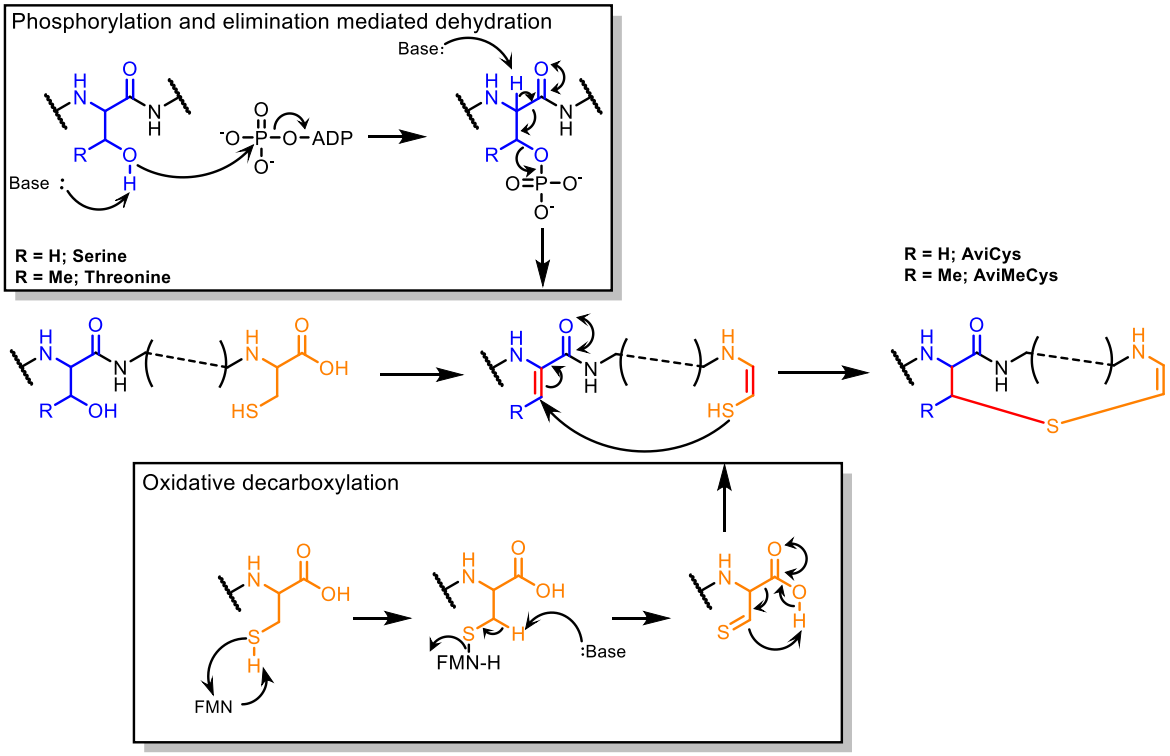


Figure 1.9: Schematic representation of Avi(Me)Cys moiety formation. In blue, dehydration of hydroxylic amino acids serine or threonine to generate dehydroalanine or dehydrobutyrine. In orange, flavin-dependent cysteine decarboxylation. The bond changes from the previous step are drawn in red in Avi(Me)Cys unit formation.

The phosphotransferase TsdC and the HopA1-like lyase TsdD were proposed to form a complex to dehydrate the threonine and serine through phosphorylation of the target residue by phosphotransferase and subsequent removal of the phosphorylated residue by the HopA1-like lyase. This removal would occur via elimination mediated dehydration, potentially an E1cB (elimination, unimolecular, conjugation base) mechanism (Sikandar et

al., 2022). A preliminarily *in vivo* experiment carried out in *E. coli* supported that the dehydration in the biosynthesis of thiostreptamide S87 was phosphorylation dependent as phosphorylated intermediates were observed in an experiment where mutations were introduced in the gene encoding TsdD, inactivating this lyase (Qiu et al., 2021). This result showed that the dehydration in thioamitides potentially would undergo the same mechanism in class II, III and IV lanthipeptides, which had been experimentally confirmed (You and van der Donk, 2007, Liang et al., 2022).

The formation of enethiol from the cysteine residue in the precursor peptide of thiostreptamide S87 is driven by the flavin-dependent oxidoreductase TsdF which catalyses the same oxidative decarboxylation mechanism previously described in lanthipeptides (Figure 1.9) (Qiu et al., 2021). In the biosynthesis of epidermin (a lanthipeptide), the decarboxylation occurs to generate a thioenamide intermediate prior to the cyclisation (Sit et al., 2011). In thiostreptamide S87 biosynthesis, the redox cofactor for TsdF was confirmed to be flavin mononucleotide (FMN), although flavin adenine dinucleotide (FAD) could also function in the biosynthesis of mersacidin (Majer et al., 2002, Qiu et al., 2021). Although a Michael addition could occur spontaneously for the enethiol formation, it was suggested that the other phosphotransferase, TsdE, could bind to the leader peptide and facilitate the correct AviMeCys formation as the absence of TsdE led to poor yield and production of various derivatives (Qiu et al., 2021). This result supported the previous findings from the *in vivo* thiostreptamide S4 biosynthetic pathway reconstitution in *S. coelicolor* M1146 where intermediates with Dha and Dhb were observed without formation of AviMeCys unit, when the TsdE homologue in thiostreptamide S4 biosynthetic pathway was knocked out (Eyles et al., 2021).

1.3.5. Pyruvyl-like moieties and leader peptide removal

Thioamitides have a second serine/threonine residue in their core peptide that undergoes dehydration by the phosphotransferase-HopA1 pair, but is not involved in the formation of Avi(Me)Cys moiety. This other Dha or Dhb is located at the N-terminus, and it has been shown to undergo a spontaneous reaction to form a pyruvyl or 2-oxobutyryl moiety, respectively, after removal of the leader peptide (Eyles et al., 2021, Sikandar et al., 2022, Xiong et al., 2022). This modification can potentially occur through (N-terminal) amide hydrolysis according to the elucidation of proteolytic removal of a lanthipeptide (epilancin 15X) leader peptide (Figure 1.10) (Ortega et al., 2014). The analysis of thioviridamide BGC suggested that TvaK is potentially involved in the leader peptide removal as it has homology to papain family cysteine protease (Izawa et al., 2013). Papain family cysteine proteases are classified as C1A family, which is not common in bacterial genomes (Rawlings and Barrett, 1994, Brömme, 2000). PoyH, also a C1A family protease, is involved in polytheonamide biosynthesis where it catalyses leader peptide cleavage on the precursor peptide, implying that homologues of TvaK in thioamitide biosynthetic pathways can also have the same role (Helf et al., 2019). However, the *in vivo* study of thiostreptamide S4 biosynthesis showed that the TsaK was not essential for the compound maturation (Eyles et al., 2021).

Although the role of the protease has not yet been confirmed, the formation of pyruvyl-like moiety at the N-terminus in thioamitides was spontaneous after leader peptide removal (Sikandar et al., 2022). While most of the currently characterised thioamitides have a pyruvyl N-terminus, the N-terminus in thioalbamide is reduced by an additional tailoring NAD(P)H-dependent reductase encoded in the BGC, generating a lactyl N-

terminus (Hayakawa et al., 2006b, Frattaruolo et al., 2017, Kjaerulff et al., 2017, Lu et al., 2021). In fact, a thioamitide with a lactyl unit was first seen in a derivative of thioviridamide, JBIR-140, from the heterologous expression of the thioviridamide BGC in *S. avermitilis* SUKA17, even though expressing the same BGC in *S. lividans* produced thioviridamide, the same compound identified from the native strain (Izumikawa et al., 2015). This suggested that the lactyl formation in JBIR-140 is also likely to be enzymatic, and this reaction could be linked to a protein that is expressed by *S. avermitilis* SUKA17. A similar modification occurs in epilancin 15X maturation, where the pyruvyl group is subsequentially reduced to a lactyl unit by the short chain dehydrogenase ElxO.

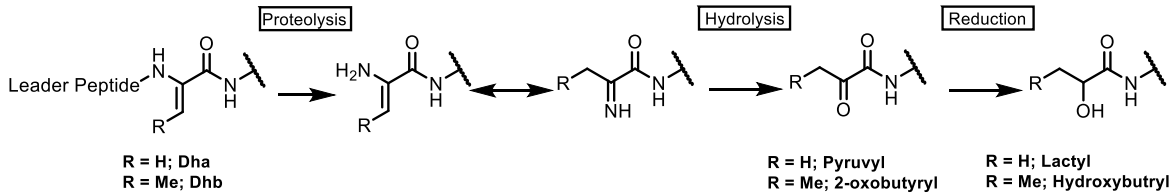


Figure 1.10: Schematic of pyruvyl and 2-oxobutyryl formation and the consequent lactyl and hydroxybutryl formation through reduction.

Although thioviridamide exhibits an N-terminal 2-hydroxy-2-methyl-4-oxopentanoyl group rather than a pyruvyl/lactyl unit, it has been confirmed that this is an experimental artifact given by the extraction solvent during compound purification and that the true product of the pathway, prethioviridamide, carries a pyruvyl moiety (Hayakawa et al., 2006b, Frattaruolo et al., 2017, Izawa et al., 2018). Since both JBIR-140 and thioviridamide were processed with acetone and the lactyl unit in JBIR-140 showed stability when exposed to acetone, it was clear that pyruvyl and lactyl have different stabilities and reactivities (Hayakawa et al., 2006b, Izumikawa et al., 2015). Although it was not clear the role of the pyruvyl-like moiety in thioamitide biological activity, the thioviridamide molecule with

lactyl N-terminus JBIR-140 displayed a higher potency comparing to thioviridamide (Figure 1.2).

1.3.6. Dimethylhistidinium and methylation

The N^1, N^3-dimethylhistidinium (dmHis) group is probably the most unique feature in thioamitides. Interestingly, the formation of this feature is via a common PTM, methylation. Methylations can be observed in nucleic acids, proteins and small molecule metabolites, including natural products (Bennett et al., 2017). Methylations are typically catalysed by *S*-adenosyl-L-methionine (AdoMet or SAM) dependent methyltransferases and can be made on oxygen, nitrogen or sulphur, as is the case in natural products such as *N*-methylation in aeronamide from *Microvirgula aerodenitrificans*, ergothioneine from *Claviceps purpurea* and omphalotin from *Omphalotus olearius*, and *O*-methylation in bottromycin A_2 from *Streptomyces bottropensis* and myxin from *Lysobacter antibioticus* (Sterner et al., 1997, Shimamura et al., 2009, Jiang et al., 2018, Song et al., 2018, Borodina et al., 2020, Cogan et al., 2022). The SAM-dependent methyltransferase normally introduces methylation through binuclear reactions using SAM as a cofactor, for instance, the methylation in the biosynthesis of ergothioneine and myxin (Figure 1.11) (Zubieta et al., 2003, Bennett et al., 2017, Jiang et al., 2018, Borodina et al., 2020). As the methylation is directed by enzymatic reaction, the modifications are also regioselective. For instance, depending on the type of protein arginine methyltransferases, it catalyses mono- and asymmetric dimethylation, or mono- and symmetric dimethylation (Fuhrmann et al., 2015, Maron et al., 2022).

Specifically, the formation of dmHis is catalysed by an *N*-methyltransferase (TvaG in thioviridamide biosynthetic pathway) that installs a methyl group on each of the nitrogen atoms in the imidazole sidechain. While it is not clear the impact of dmHis on thioamitide

bioactivity, it has been proven that the methylation in a molecule changes its lipophilicity and binding affinity and other pharmaceutical properties. While silybin B, isolated from *Silybum marianum*, shows cytotoxic and antiviral activities, synthetically methylating the compound enhanced both cytotoxicity and antiviral activities (Sy-Cordero et al., 2013).

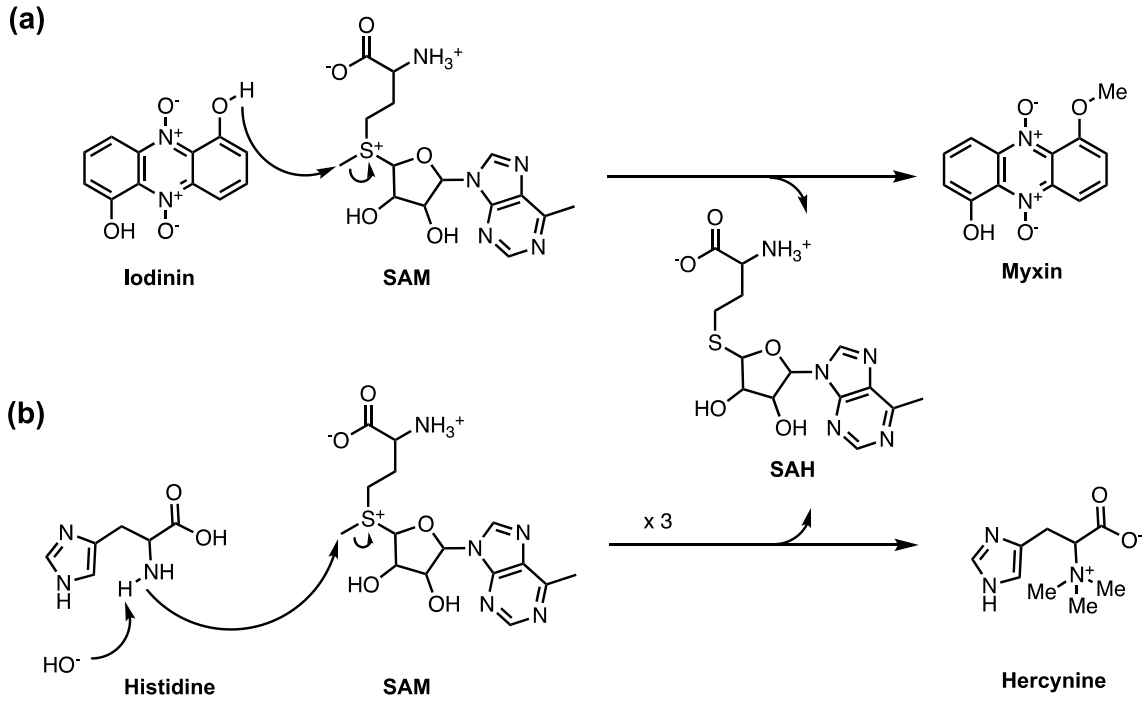


Figure 1.11: Schematic reaction for SAM-dependent reactions. (a) *O*-methylaton on the precursor of myxin, iodinin using SAM as co-factor and producing SAH as byproduct. (b) *N*-methylation on a histidine to produce the precursor of ergothioneine, hercynine.

The methylations in the sidechain also introduce a permanent positive charge to the molecule. Although the mechanism was unclear, positively charged polyethylene glycolated molecules were shown to have better anticancer activity than their neutral and anionic forms due to doubling the cell membrane penetration ability (Wang et al., 2016a). According to the positive inside rule, negatively charged part of transmembrane proteins are often at the outside of the lipid bilayer, so positively charged molecules are more likely attracted to the cell membrane. Also, since the membrane potential is negative, which means the interior concentration of total anionic molecules is higher than the exterior,

positively charged molecules can also passively diffuse into the cell. Thus, the positively charged sidechain could have an important role in thioamitide bioactivity that is associated with cell penetration.

1.3.7. Additional PTMs in thioamitides

While there are core PTMs that can be observed across all currently characterised thioamitides, there are PTMs that are unique to specific thioamitides. Whether or not these additional modifications affect the mode of action is not clear, but they are very likely to have a role of affecting the potency of biological activity. Almost all characterised thioamitides have the hdmHis residue, derived from the histidine residue by hydroxylation as well as the two *N*-methylations, except thioalbamide (Frattaruolo et al., 2017). The hydroxylation is catalysed by a monooxygenase (encoded as gene *J* in thioamitide BGCs), which is homologous to a histidyl-hydroxylase NO66 that hydroxylates 60S ribosomal protein Rpl8 at the β-carbon of the histidine (Ge et al., 2012, Izawa et al., 2013, Eyles et al., 2021, Sikandar et al., 2022). This oxygenase is found encoded in almost all identified BGCs, except in thioalbamide.

Except the reductase responsible for the *N*-lactyl moiety, the thioalbamide BGC also encodes a cytochrome P450 homologue, most of which catalyse oxidation (Frattaruolo et al., 2017). The oxidation by cytochrome P450 involves using heme as cofactor and molecular oxygen O_2 for the reaction, and it can achieve hydroxylation, C-C bond formation or cleavage (Guengerich, 2018). In the biosynthesis of thioalbamide, it is proposed to introduce a hydroxyl group into the phenylalanine residue at the ortho position of the aromatic ring as a structural isomer of tyrosine (Frattaruolo et al., 2017).

Finally, an extra methylation was observed in the tyrosine residue within the macrocycle of thiostreptamide S4 (Frattaruolo et al., 2017). It was proved that the methylation was installed after the leader peptide removal and the extra methyltransferase TsaMT was responsible for the PTM (Eyles et al., 2021). So far, this particular modification, *O*-methylation in macrocycle, is unique to thiostreptamide S4, but several orphan thioamitide BGCs also feature genes with homology to TsaMT (Frattaruolo et al., 2017).

1.4 Synthetic biology and biosynthetic gene cluster manipulation

1.4.1. The purpose of BGC manipulation

The genetic manipulation of a BGC can serve multiple purposes. For example, it has been classically used to investigate the biosynthetic pathway of a natural product through the mutation and complementation of its genetic components. The deletion of a gene can verify whether it is involved in the biosynthesis by monitoring the production of the compound of interest. These approaches help to define the boundaries of the BGC and the functions of its genes, supported in many cases by the accumulation of metabolic intermediates. However, those intermediates can potentially be readily endogenously digested or modified (for example, by acetylation), eventually making their identification challenging, especially if the deleted gene participates in the early stages of the biosynthetic pathway. Fortunately, these intermediates are normally structurally related, so the identification of the intermediates and shunt metabolites can be facilitated by in-depth untargeted metabolomics and metabolite networking (Crone et al., 2016).

Genetic manipulation can also be employed to trigger the production of a cryptic or naturally inactive gene cluster, as well as to increase the yields of a given compound. It is recognised that all microbes possess cryptic and silent biosynthetic pathways that can be found through genome mining, but are not produced at detectable levels under laboratory conditions (Wilkinson and Micklefield, 2007). These cryptic or silent biosynthetic pathways require specific (and mostly unknown) cultivation conditions to trigger the production in the native host through different regulatory mechanisms, which activate or de-repress the expression of the biosynthetic genes. In some cases, the cloning and transfer of the BGC of interest into a suitable heterologous host can activate its expression, arguably because of

the lack of repressors of its expression in the new regulatory context. The discovery of thiovarsolins is an excellent example of this strategy (Santos-Aberturas et al., 2019). The genetic engineering of a biosynthetic pathway can also improve the production, which is crucial for structural and bioactivity characterisation, through the engineering of regulatory components and the deletion of competing pathways to redirect the flux of precursors and other metabolic resources. For example, production of mithramycin was boosted 16 times by the overexpression of a *Streptomyces* antibiotic regulatory protein, while the deletion of certain genes in doxorubicin biosynthetic pathways increased the yield of doxorubicin by reducing the amount of doxorubicin being converted into other undesirable derivatives (Lombó et al., 1999, Lomovskaya et al., 1999).

A more sophisticated genetic engineering approach, known as “refactoring”, proceeds by the systematic substitution of the promoters present in the original BGC by constitutive or inducible promoters not subjected to the original regulatory mechanisms governing the expression of the biosynthetic pathway (Temme et al., 2012). Although the refactoring of a biosynthetic pathway can be technically complex, one of its advantages is that it does not require previous knowledge about the regulation of the original pathway (Eyles et al., 2018, Horbal et al., 2018). One of its disadvantages, however, is that the stoichiometry of the enzymes involved in the pathway can substantially differ from the original system. In some cases, this altered stoichiometry can affect the biosynthetic performance of the engineered pathway (Eyles et al., 2018).

Genetic engineering can also be employed to produce non-natural derivatives or completely new compounds, as illustrated by the engineering of modular assembly lines for PKs and NRPs, or the engineering of precursor peptides in the case of RiPP biosynthetic

pathways (Bozhüyük et al., 2019a, Bozhüyük et al., 2019b, Franz and Koehnke, 2021). Exchanging modular domains in rapamycin synthase successfully altered the final product and produced 6-deoxyerythronolid analogues (McDaniel et al., 1999). Additional structural diversity for the final product of a biosynthetic pathway can be introduced through combinatorial biosynthesis efforts, where genetic elements belonging to different pathways are combined. For example, combinatorial biosynthesis approaches led to productions of vatiamide analogues, the production of new polyene macrolides or the production of thioamidated thiostrepton (Santos-Aberturas et al., 2015, Schwalen et al., 2018, Moss et al., 2019).

1.4.2. Examples of genetic engineering techniques

1.4.2.1. *in situ* modification

As stated above, genetic modification allows for an understanding of biosynthesis, can increase yields and diversify the final products. Many different tools have been developed for the genetic manipulation of bacteria. In particular, the manipulation of BGCs can be made *in situ* or *ex situ*. The main advantage of the modification of a BGC in its native host is the conservation of the original biochemical, biological, and regulatory context, and this ensures that certain unknown requirements for the production of the compound will remain unaltered. The greatest challenge in editing genes *in situ* is caused by the wide range of natural product producers, so bespoke tools and methods need to be developed for the producers. Unfortunately, despite actinobacteria being extensively studied and although a reasonable range of genetic modifying tools are available, the extraordinary diversity of this group of bacteria means that many strains remain genetically intractable by standard methods (Kieser et al., 2000).

Several methods have been developed for the introduction of exogenous DNA into several groups of actinobacteria. In some cases, protoplasts can be transformed with plasmids by electroporation or using polyethylene glycol as chemical transformation inducer (Musiol-Kroll et al., 2019). However, the conditions required for the generation of protoplasts strongly vary even across species belonging to the same genus, and method optimization can be tricky and time consuming. In some cases, it is possible to perform direct electroporation-mediated transformation of washed vegetative cells, but the right choice of the culture density, culture conditions and growth phase can be critical in this case (Kieser et al., 2000). Transfection using actinophage vectors is also possible, although is laborious and not widely employed since the development of intergeneric conjugation protocols employing *Escherichia coli* strains as DNA donors. This method is simple and allows the introduction into the host of a variety of origin of transfer (*oriT*) containing vectors. Depending on their behaviour once inside of the host, plasmids can be classified as suicidal (only allowing homologous recombination into the host genome), integrative (thus depending on the presence of an integration attB site compatible with an attP site within the plasmid, for instance pIJ10257 integrates into ϕBT1 phage integration site) or replicative (Kieser et al., 2000, Hong et al., 2005). As many actinobacteria (such as *S. coelicolor* A3(2) and *S. avermitilis*) have strong methyl-specific restriction systems which prevent the uptake of methylated DNA, the use of methylation-deficient *E. coli* strains for the propagation and conjugation of plasmids is a general requirement to ensure transformation and conjugation efficiency (MacNeil, 1988).

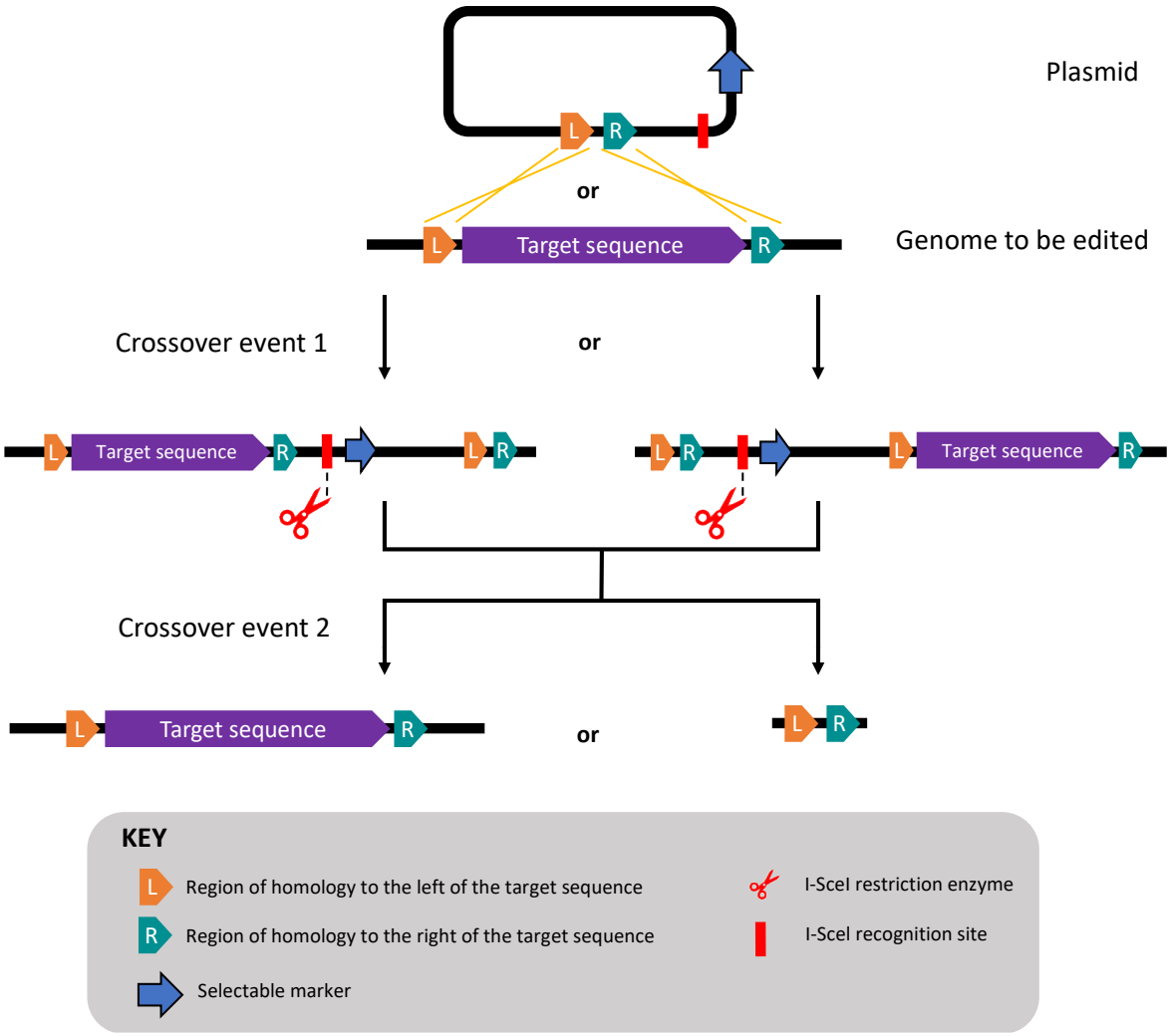


Figure 1.12: Meganuclease mediated deletion or replacement of a target sequence in gDNA (Fernández-Martínez and Bibb, 2014) through double crossing over. I-SceI can be expressed from a separate plasmid to introduce DSB and the recombination events result in 50% of deletion of the target sequence or reversion to the original genotype.

When a gene or genes in a BGC need to be removed or replaced, although insertional inactivation via a single crossover can employed, this method is disruptive to the BGC as the plasmid integrates into the genome and can cause polar effects affecting the expression of genes situated downstream in the transcriptional organization of the BGC. Instead, insertional inactivation via double cross over or using an in-frame deletion is preferred. This insertional method for gene disruption can be achieved by a suicide vector containing a region or regions for homologous recombination (Kieser et al., 2000). To promote the in-

frame deletion, I-SceI endonuclease has been employed (Lu et al., 2010). Firstly, a plasmid containing a rare sequence for I-SceI recognition as well as the regions for homologous recombination is integrated into the genome by single crossover. In a second step, I-SceI endonuclease introduces double-strand breaks (DSB) at this rare sequence to promote second single-crossover recombination events as the only option for the survival of the host (Figure 1.12). This method was used in *S. coelicolor* and approximately 50% of the clones contained the desired genotype (Lu et al., 2010, Fernández-Martínez and Bibb, 2014).

Precise DNA deletion in actinobacteria can also be achieved by cooperating clustered regularly short palindromic repeats (CRISPR) arrays and CRISPR associated (Cas) genes. The working principle is to introduce DSBs by a complex containing Cas9 nuclease and a synthetic guide RNA (sgRNA), which is a fusion of a CRISPR RNA (crRNA) and a trans-activating crRNA (tracrRNA) and directs the Cas9 to cleave at the defined locations. The use of pCRISPomyces efficiently removed targeted DNA in the length of between 20 and 30,000 bp from 3 different streptomycetes chromosomes (Cobb et al., 2015). Many different CRISPR/Cas9 related tools and techniques have been developed (Mitousis et al., 2020). Apart for DNA deletion, CRISPR/Cas9 can also be used to introduce DNA to the genome. Hence, the strategy can also activate silent gene clusters by introducing promoters or regulatory genes, or by the direct activation of transcription by the use of dCas9 (dead Cas9)-transcriptional activator fusion proteins which are able to bind the region targeted by the sgRNA but not to produce a DBS, thus allowing the fused transcriptional activator to recruit the RNA polymerase complex to the targeted promoter (Zhang et al., 2017, Ho et al., 2020).

1.4.2.2. BGC cloning

If genetic manipulation is performed in a native host, the researcher is limited by whether the genome editing tools are compatible with the host strain. Despite the impressive potential of actinobacteria for the production of diverse secondary metabolite production, the main bottleneck for *in situ* genetic modification is that many actinobacterial strains remain genetically intractable. As an alternative approach, the BGC of interest can be isolated by capturing it from the native strain genomic DNA or assembled using polymerase chain reaction (PCR) products and then transferred, expressed, and characterised in a model heterologous host, such as *S. coelicolor* M1146 and *S. lividans* TK24 (Gomez-Escribano and Bibb, 2011, Ahmed et al., 2020). This strategy has been successfully used for many BGCs, including the SGR810-815 BGC, thioviridamide BGC in *S. lividans*, plus the bottromycin BGC, the thiovarsolin BGC and the thiostreptamide S4 BGC in *S. coelicolor* (Izawa et al., 2013, Luo et al., 2013, Eyles et al., 2018, Santos-Aberturas et al., 2019, Eyles et al., 2021). To manipulate a BGC outside of the native host, the BGC needs to be assembled into a vector. This can be achieved by targeted or untargeted ligation, golden gate cloning, Gibson assembly, transformation-associated recombination (TAR), to name but a few (Nah et al., 2017).

Gene clusters can simply be amplified by PCR and ligated into a vector. Untargeted ligation method utilises the restriction enzyme recognition sites naturally existing in genomic DNA (gDNA) and ligates the restriction fragments into vectors. The resulting library of plasmids/cosmids can be screened for BGCs of interest. This method has captured almost the entire chromosomal DNA of *S. coelicolor* into 637 constructs (Redenbach et al., 1996). The method of untargeted ligation cosmid libraries has been extremely successful and widely used during the last few decades in bacterial natural product research. However,

novel targeted methods bypassing the library screening required for cosmid libraries have gained importance in recent years.

BGC assembly can also be achieved by homologous recombination *in vivo* or *in vitro* using gDNA from the original organism, or synthetic DNA if feasible. TAR cloning is a robust method for large chromosomal fragment isolation using the yeast *Saccharomyces cerevisiae*, which can perform homologous recombination with very high efficiency (Orr-Weaver et al., 1981, Larionov et al., 1997). To perform TAR cloning, a capture vector is normally generated containing a pair of homologous capture arms taken from both ends of the target DNA (Figure 1.13). The vector is normally trifunctional, containing a yeast origin of replication, an *E. coli* origin of replication and an attB site for integration into the genome of the final expression host, as well as a counterselection marker against the empty capture vector and an antibiotic selection marker for downstream selection (Noskov et al., 2003, Zhang et al., 2019).

Gibson assembly assembles multiple overlapping DNA fragments in a single-step isothermal reaction *in vitro* (Gibson et al., 2009). DNA fragments are assembled through 5'-chewback by T5 exonuclease to generate sticky 3'-ends of fragments, followed by annealing and DNA repair by ligase and DNA polymerase (Figure 1.14). Gibson assembly is also a robust method, and it has a shorter process time as it does not require preparation of yeast spheroplasts and the longer times required for the growth and manipulation of yeast.

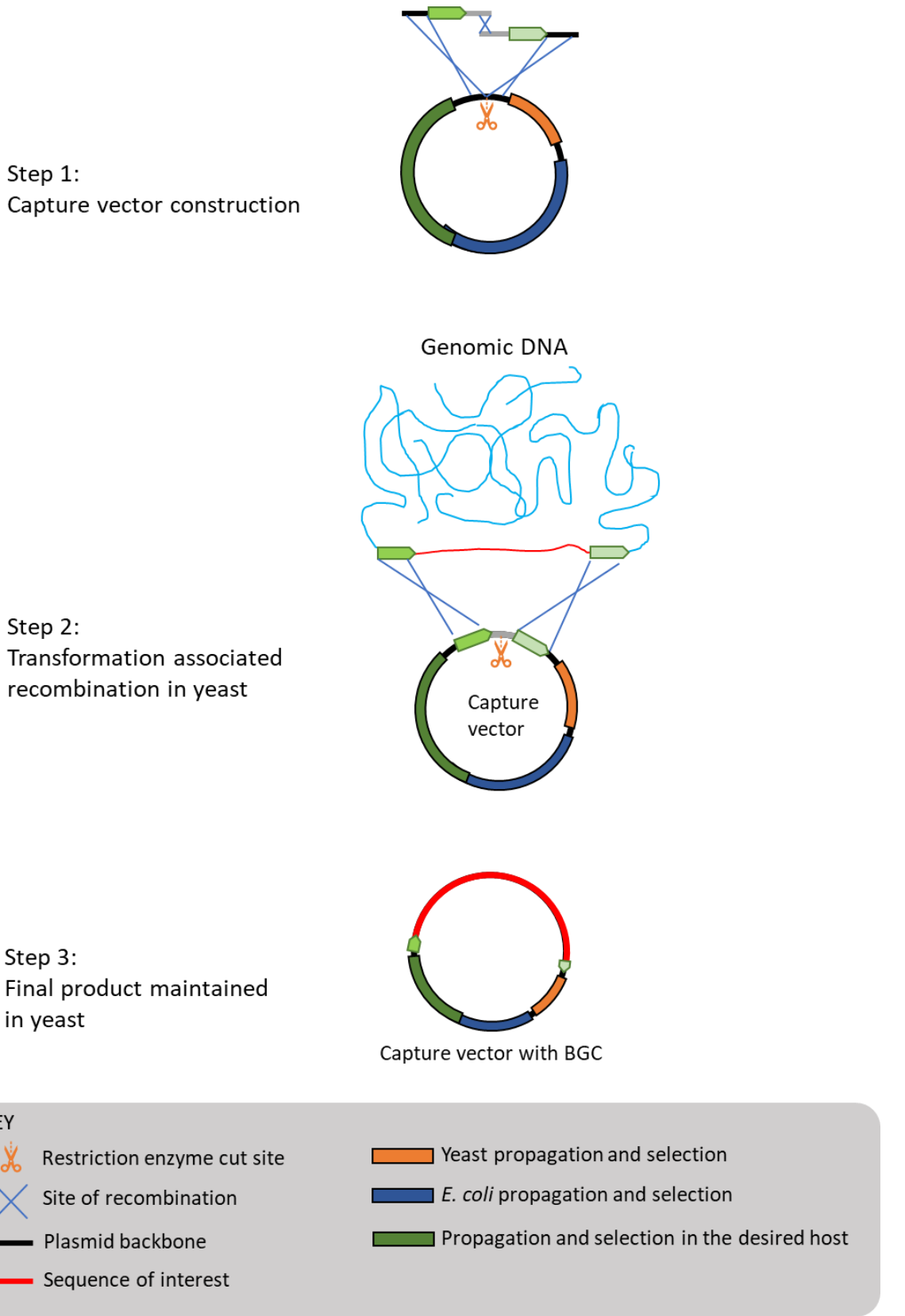


Figure 1.13: General transformation associated recombination in yeast for gene cluster capture from gDNA using homologous capture arms.

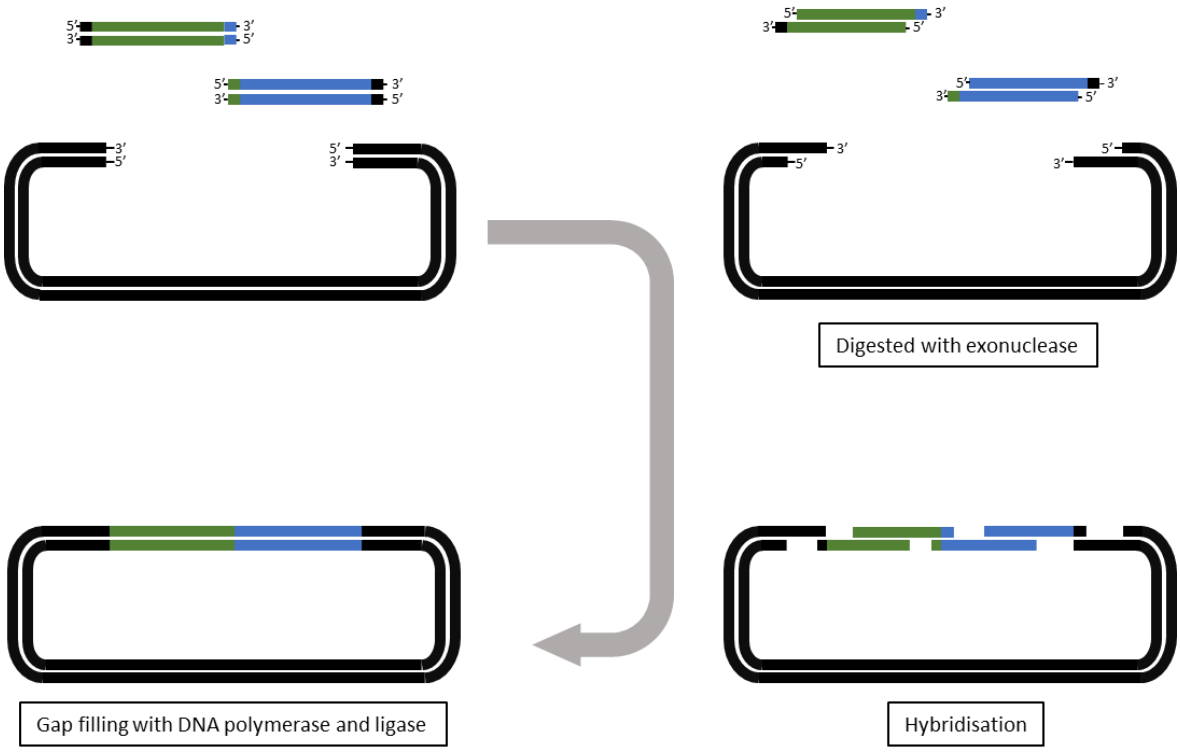


Figure 1.14: Schematic of a single-step isothermal Gibson assembly showing the 3 events during assembly.

1.4.2.3. *in vitro* techniques for BGC modification

Once the BGC is isolated and introduced into a well-characterised host cell, it can be modified using the engineering techniques described in section 1.4.2.1. *in vivo* genetic manipulation can be time consuming as it is limited by the life cycle (growth speed) of an organism or other complicated biological factors, like it was briefly mentioned in TAR cloning and Gibson assembly.

The modification of a BGC can be simultaneously done during the BGC assembly process. For instance, when a homologous recombination method is used, the homologous regions for recombination can be introduced by extended primers with overhangs to stitch two PCR amplicons amplified from discontinuous genomes or from different sources. By doing this, gene deletion, insertion or rearrangement can be achieved. Using Gibson assembly, Kalva *et al* generated a scarless deletion in a single *in vitro* assembly step (Kalva et al., 2018).

Using this kind of approach, it is also possible to perform the complete refactoring of a BGC in a single step by combining multiple overlapping PCR products corresponding to multiple promoters and coding sequences. TAR cloning can also be used to simultaneously modify a BGC alongside BGC assembly (Eyles et al., 2018).

The assembly of PCR-generated fragments by different methods can be employed for multiple purposes, as gene deletion, refactoring, or addition. However, PCR-based methods are limited by the efficiency of DNA polymerase, which drastically decreases when amplifying long amplicons or difficult templates. Although this can be overcome with assembly by splitting the region into two or more, assembly limits itself with low efficiency when there are too many fragments. Also, in the case of Gibson assembly, small DNA fragments are not efficiently digested and can be problematic (Roth et al., 2014).

Golden gate cloning is an alternative *in vitro* technique that utilises Type IIS restriction enzymes to stitch multiple fragments through ligation for BGC construction and modifications (Figure 1.15) (Bird et al., 2022). Type IIS restriction enzymes cleave DNA at a fixed distance to the recognition site, which means that a single restriction enzyme can generate numbers of fragments with unique sticky ends, so the number of assembly building blocks is not limited by the available enzymes. For instance, a high-yield assembly of 24 fragments was achieved in a single rection to assemble a *lac* cassette (Potapov et al., 2018).

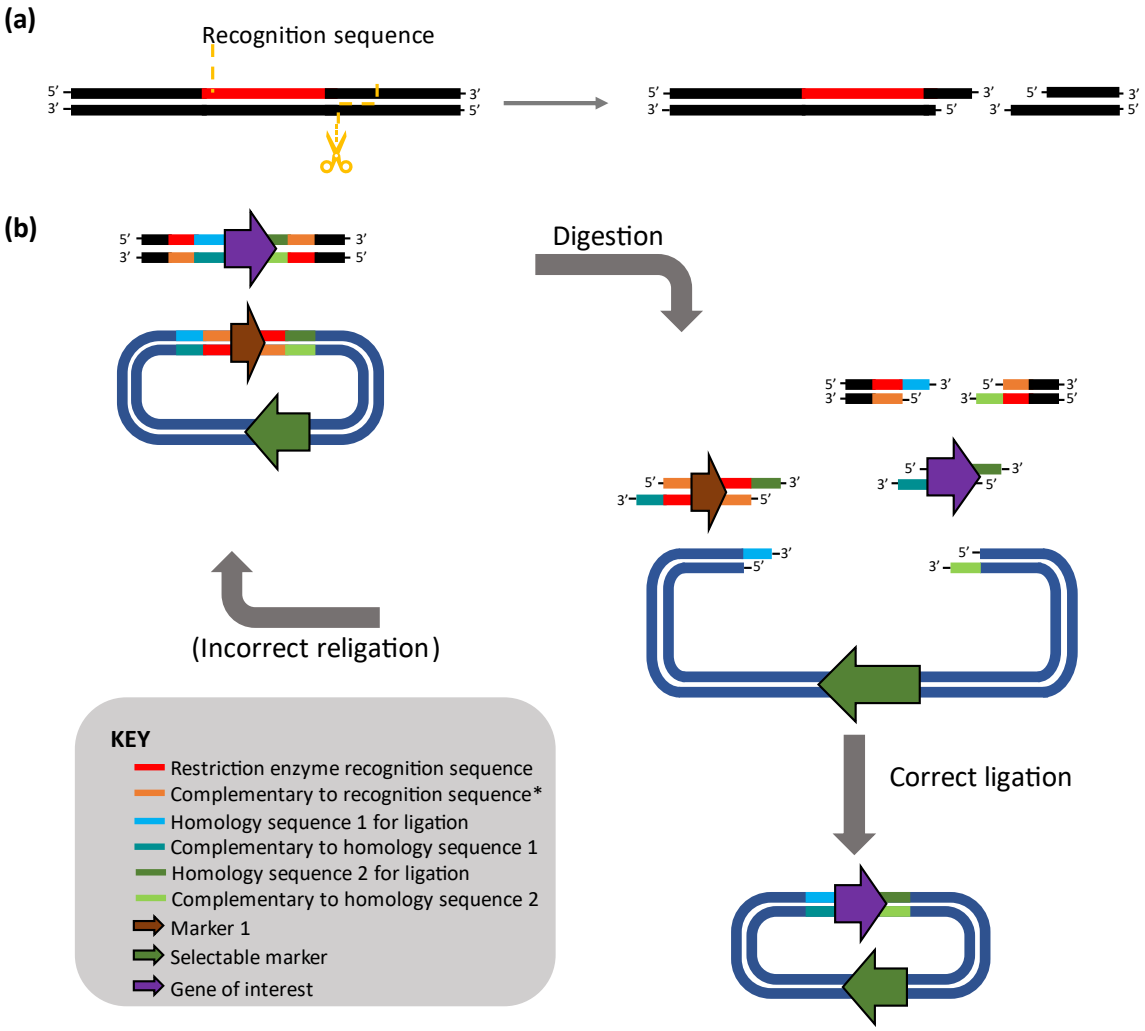


Figure 1.15: Schematic reaction for Golden gate assembly. (a) General digestion behaviour of type IIS restriction enzymes, which recognise a particular sequence in one strand of the DNA and cut outside the enzyme recognition site. (b) The scheme represented one of constructs in the reaction mix. The combination of homology sequence 1 and 2 are specific to this particular construct, and one of the homology sequences can be used in another construct to dictate the position of the gene in the next level of construction. Marker 1 is optional and helps in identifying colonies containing plasmids from religations.

1.4.2.4. *in vivo* techniques for BGC modifications

Along with many *in vitro* modification techniques, a BGC captured in a plasmid/cosmid can also be edited *in vivo*. Many tools to modify BGCs *in vivo* have been mentioned in section 1.4.2.1, where the tools for *in situ* modifications were discussed. PCR-targeted mutagenesis is especially efficient and recommended once the BGC of interest is cloned into a plasmid, allowing replacement to delete a region of plasmid (Gust et al., 2003). The deletion is

prompted by a phage protein λ-Red by recombination in *E. coli* BW25113 between the DNA in the plasmid and a linear PCR amplified antibiotic resistance cassette surrounded by recombination arms. The antibiotic selection marker that replaces the gene from the plasmid can subsequentially be removed by expression of a recombinase to leave an 81 bp scar, thus allowing the use of the same resistance marker in consecutive mutagenesis steps (Gust et al., 2003).

There are still many different tools for *in vivo* DNA modification that have not been mentioned here, and it is worth emphasising that some tools are not strictly being used for a single purpose (Liu et al., 2021b).

1.5 Aims of the thesis

Thioamitides are a group of structurally and biologically fascinating compounds, but the available knowledge about them is rather limited, both in terms of biosynthetic machinery and structure/bioactivity relationship. There are currently six structurally distinct natural thioamitides (Hayakawa et al., 2006b, Izumikawa et al., 2015, Frattaruolo et al., 2017, Izawa et al., 2018). All thioamitides tested for their biological activity showed potent cytotoxicity against cancer cells (Hayakawa et al., 2006a, Izumikawa et al., 2015, Kjaerulff et al., 2017). Given that thioamitides with different structures showed changes in their bioactivity levels, the additional structural diversity of unknown thioamitides pointed out by genome-mining predictions was suggestive of a wider bioactivity potential within this group of compounds. Thus, it was crucial to trigger the production and to identify the thioamitides being encoded in these silent BGCs. The discovery of a novel thioamitide is the focus of Chapter 2.

One potential issue that creates a bottleneck for compound characterisations is the low productivity (Lombó et al., 1999). This could be addressed by genetic manipulation to increase the yield by refactoring as discussed before. In order to achieve that, BGCs may need to be cloned into a heterologous expression host. By making the BGC accessible to genetic manipulation, this also unlocks the gate to gain insights into its biosynthesis. This work is discussed in Chapter 2.

At the time when the project started, the biochemistry underlying the origin of the multiple thioamide bonds featured by thioamitides had not been characterised. It was hypothesised that the thioamide bonds in thioamitides could be installed by the YcaO-TfuA pairs according to the study of thioamidation in the α-subunit of methyl-coenzyme M reductase

in archaea (Mahanta et al., 2018). Hence, one of the objectives in this thesis was to investigate the thioamidation through *in vitro* reconstitution. This work is described in Chapter 3.

Chapter 2 Discovery and study of the biosynthesis of thiopotensamides

2.1 Introduction

2.1.1. Thioamitides, a group of structurally fascinating molecules

Thioviridamide, the founding member of thioamitides, was first isolated from *Streptomyces olivoviridis* through a screen for cytotoxic compounds (Hayakawa et al., 2006a). This molecule is a very complex peptide natural product possessing a 2-hydroxy-2-methyl-4-oxopentanoyl group, five thioamide bonds, an *S*-[(*Z*)-2-aminovinyl](-(3S)-3-methyl)-D-cysteine (Avi(Me)Cys) macrocycle, two histidine *N*-methylations and a histidine β-hydroxylation (Figure 2.1) (Hayakawa et al., 2006b). The compound was confirmed as a ribosomally synthesised and post-translationally modified peptide (RiPP) via heterologous expression of its biosynthetic gene cluster (BGC) in *Streptomyces lividans* (Izawa et al., 2013).

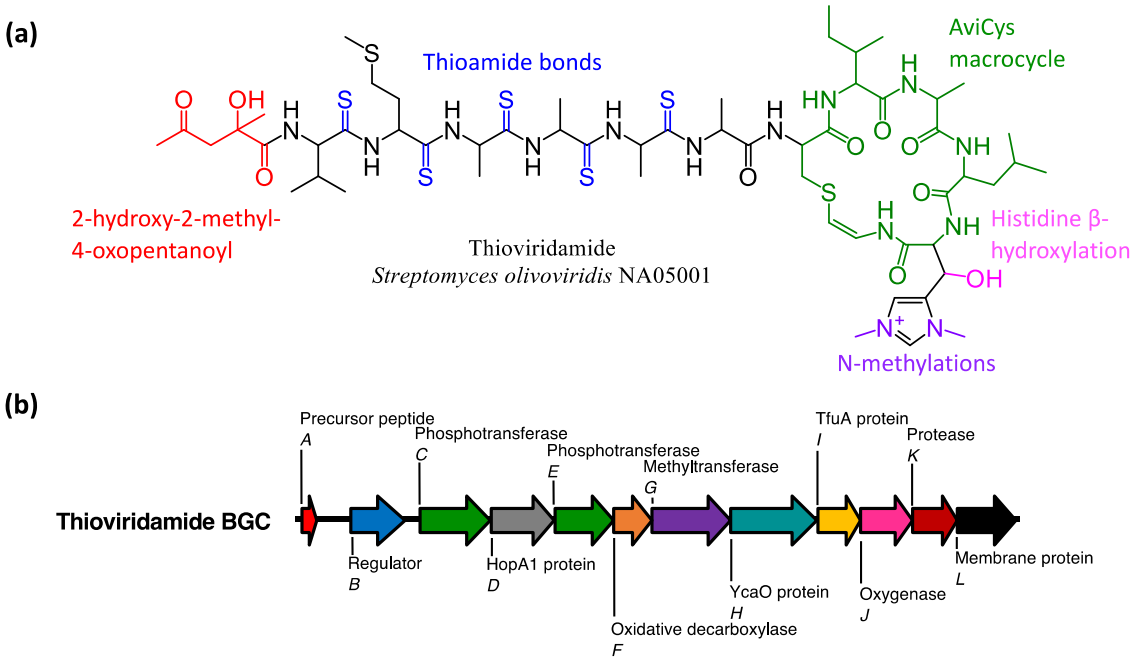


Figure 2.1: Thioviridamide structural features (a) and its BGC (b).

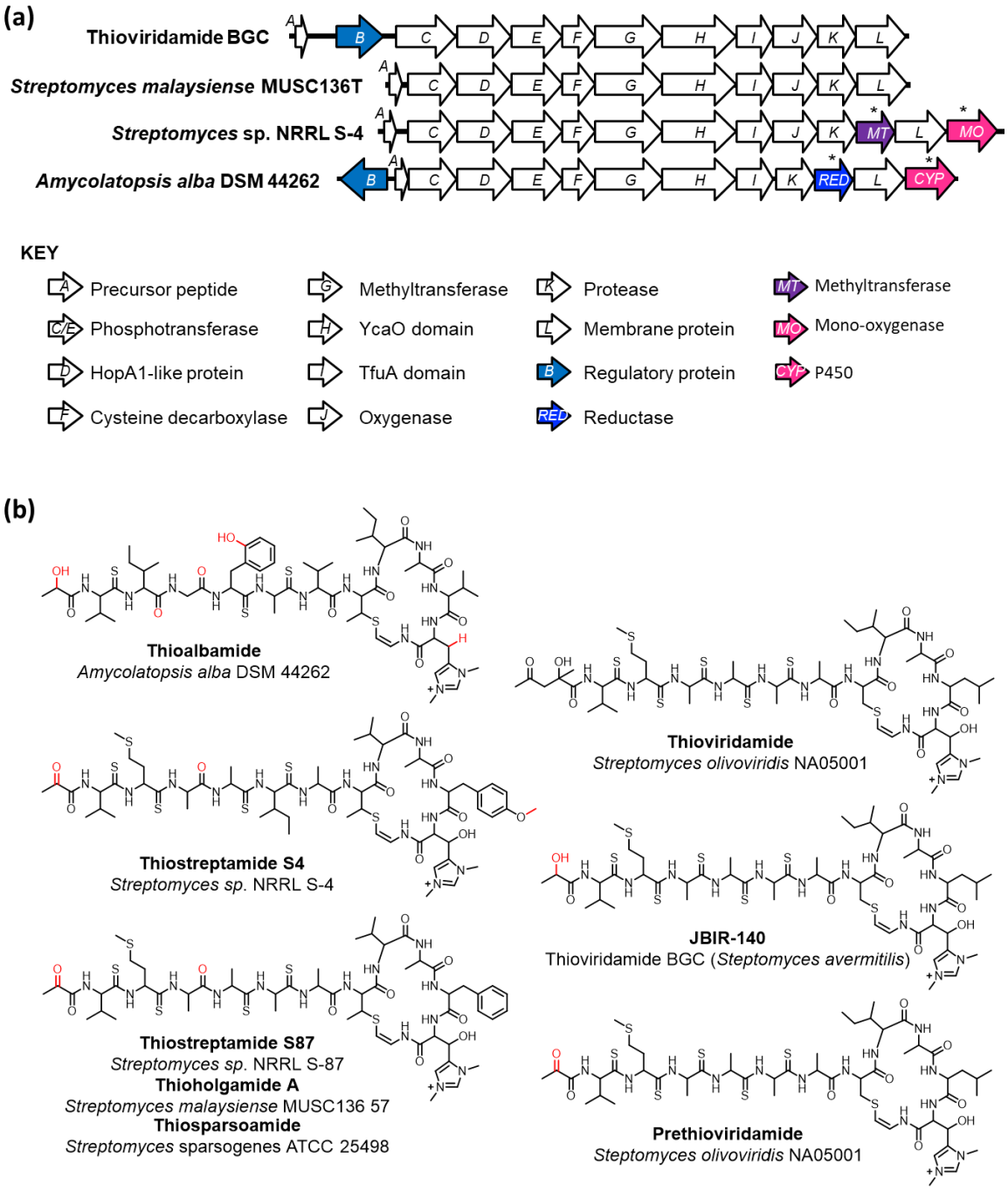


Figure 2.2: Thioamitides and biosynthetic pathways. (a) Comparison of characterised thioamitide BGCs. Unconventional genes involved in the biosynthesis are showed in colour. (b) Structures of currently known thioamitides. The PTMs that are different to thioviridamide are labelled in red. The stereochemistry of the listed thioamitides is not confirmed.

The characterisation of the thioviridamide cluster supported the discovery of several novel thioamitides through genome mining, such as JBIR-140, thioholgamides, thioalbamide, thiostreptamide S4 and thiostreptamide S87 (Izumikawa et al., 2015, Frattaruolo et al.,

2017, Kjaerulff et al., 2017, Izawa et al., 2018). As more thioamitides were discovered, the true N-terminal functional group was found to be a pyruvyl group (or lactyl for JBIR-140 and thioalbamide) instead of a 2-hydroxy-2-methyl-4-oxopentanoyl group, which was an artifact generated during extraction of the molecule using acetone as organic solvent (Frattaruolo et al., 2017). This was further confirmed by the characterization of prethioviridamide in *S. olivoviridis*, which also contains an N-terminal pyruvyl group (Izawa et al., 2018).

Besides amino acid changes in their core peptide sequences, all these molecules share almost all the posttranslational modifications (PTMs) featured in thioviridamide, with the exception of thioalbamide, where the histidine β-hydroxylation is not installed as the responsible gene is absent in the gene cluster. There are additional modifications present in specific thioamitides, such as a tyrosine *O*-methylation in thiostreptamide S4, or a reduction on the N-terminal pyruvyl group for the lactyl formation in thioalbamide and JBIR-140 (Izumikawa et al., 2015, Frattaruolo et al., 2017). This suggests the biosynthesis of thioamitides is highly conserved but additional modifications can be introduced in a fully mature thioamitide. It is also a promising indication that further structurally diverse thioamitides could still be found, with potential implications in their bioactivity profiles.

Preliminary results suggested thioviridamide was a promising candidate for cancer treatment as the compound appeared to specifically target tumour cells that express E1A-like oncogene (Hayakawa et al., 2006a). This sparked the interest in the study of thioamitides as potential antitumour compounds (detailed discussion in section 1.3.1). After the discovery of thioalbamide, its antitumour activity was compared against the currently approved chemotherapeutic medicine doxorubicin. Unlike doxorubicin,

thioalbamide was found to selectively target tumour cells, which means thioalbamide could be potentially safer than doxorubicin. This exciting finding suggested thioamitides show a huge potential as candidates for cancer therapy. Interestingly, different thioamitides show different anticancer activities. For example, thioviridamide, JBIR-140 and thioholgamide A were found to have different half-maximal inhibitory concentration (IC_{50}) in different cancer cell lines (Izumikawa et al., 2015, Kjaerulff et al., 2017). While thioviridamide and thioholgamide A are structurally more different, JBIR-140 only has a minor change in structure, an N-terminal lactyl moiety rather than a 2-hydroxy-2-methyl-4-oxopentanoyl group in thioviridamide, indicating that subtle structural modifications can greatly affect the bioactivity of thioamitides (Figure 1.2). Hence, discovering structurally diverse thioamitides and building a thioamitide library can provide an important resource for candidate substances for chemotherapy in the future.

2.1.2. The bottleneck for novel thioamitide discovery

Through a genome-mining analysis of the National Center for Biotechnology Information (NCBI) records carried out in our lab, 14 thioamitide BGC-containing bacteria had already been identified at the start of this project (Frattaruolo et al., 2017). Most of these 14 discovered BGCs were from Actinobacteria, and many of them did not yet have their products identified. Some of the producing organisms are rare actinomycetes, whose biology is still not well known. In addition, it is a common occurrence that BGCs are not active in normal laboratory conditions, as the triggers for production found in the original environment might not be present (Walsh and Fischbach, 2010, Rutledge and Challis, 2015). A way to address this bottleneck for compound discovery and characterisation is the cloning and heterologous expression of these BGCs in a different host, usually a well-

characterised, genetically tractable strain that can be easily manipulated. For instance, bottromycin, thiovarsolin and thiostreptamide S4 were successfully expressed in a *Streptomyces* model *S. coelicolor* M1146, as well as thioviridamide in *Streptomyces lividans* (Izawa et al., 2013, Eyles et al., 2018, Santos-Aberturas et al., 2019, Eyles et al., 2021). The study of natural products can often suffer from low specialised metabolite production levels, which created another bottleneck for elucidating the chemical structure for such large and complex molecules particularly when using nuclear magnetic resonance (NMR) spectroscopy, but this problem can be overcome by refactoring the biosynthetic pathway to improve its expression especially when expressed heterologously and by upscaling the production volume, among other strategies, such as increasing the substrate supply and refactoring the biosynthetic pathway by altering the promoter or supplementing a well-characterised promotor/regulatory protein (Chen et al., 2013, Eyles et al., 2018, Mingyar et al., 2021).

2.2 Aims of the chapter

Since the currently characterised thioamitides show great potential as anticancer molecules, and even minor changes in their chemical structure can affect their anticancer activity, identifying novel thioamitides could lead to discovery of potent anticancer compounds with even better bioactivity profiles.

The main objective of this part of my PhD was therefore the discovery and structural characterisation of novel thioamitides. In order to achieve this, there were 4 specific goals:

1) Expand and update the genome mining search for novel thioamitide BGCs;
2) Fermentation screening of thioamitide production by growing candidate strains on a selection of production media;
3) Purification and characterisation of any novel thioamitides identified;
4) Genetic manipulation of the producing BGC to increase production and to understand the biosynthetic pathway of the novel compound(s).

2.3 Results and discussion

2.3.1. Genome mining of thioamitide BGCs and selection of screening candidates

2.3.1.1. Thioamitide BGC identification and analysis

This project began with a genome mining analysis of the NCBI databases looking for any novel thioamitide BGCs. For that purpose, a BLASTX search was carried out using the gene *tvaH* (BAN83923.1) as query. The genomic loci of the top 50 hits to potential homologues were investigated to determine whether the potential homologues were encoded in thioamitide-like BGCs. Using this approach, four new thioamitide BGCs were identified in the databases, increasing the total known thioamitide BGCs from 14 to 18 (Appendix 1-3). These new clusters were all identified in Actinobacterial genomes corresponding to *Nonomuraea fuscirosea* DSM 45880, *Streptomyces caelestis* DSM 40084, *Streptomyces griseocarneus* JCM 4580 and *Streptomyces sparsogenes* ATTC 25498. The genetic organisation of these novel gene clusters is very similar to the previously known thioamitide BGCs (Figure 2.3). Besides their precursor peptide and the *tvaH* homologue that used for the candidate cluster identification, all of the gene clusters contain a cysteine decarboxylase, two phosphotransferase genes and at least one HopA1 gene, suggesting their products should all contain the Avi(Me)Cys macrocycle (section 1.3.4). Similarly, all of them encode the TfuA protein (TvaI homologues) required to pair with a YcaO protein (TvaI homologues) for thioamidation of the precursor peptide (section 1.3.3). Finally, a methyltransferase (TvaG homologue) and a monooxygenase (TvaJ homologue) in the gene clusters indicated that they should all have a β-hydroxy-N^1, N^3-dimethylhistidium (hdmHis) residue in the macrocycle (section 1.3.6 and 1.3.7). In addition to this core of tailoring genes, each of the clusters contains extra specific genes involved in further modifications. For

example, a radical *S*-adenosyl-L-methionine (SAM) protein is encoded in the gene cluster from *S. griseocarneus*. With respect to the newly identified clusters, the BGC from *S. griseocarneus* is very similar to one from *Micromonospora eburnea*, except an endonuclease gene is found in between the precursor peptide (gene *A*) and phosphotransferase (gene *C*) in the *M. eburnea* BGC. On the other hand, the *S. caelestis* and *S. sparsogenes* have the exact same genetic organisation as *Streptomyces* sp. NRRL S-87. Among all the clusters, the smallest one belongs to the cyanobacterium *Mastigocladus laminosus.* Although it was identified previously, the poor-quality sequencing data did not allow the complete annotation of this BGC (Frattaruolo et al., 2017). This was updated during this bioinformatic analysis, where a class I SAM-dependent methyltransferase (homologue of *tsaMT* from thiostreptamide S4 biosynthetic pathway) is encoded after the oxygenase (gene *J*), which appears to be the last gene in the thioamitide BGC from *M. laminosus*, the second thioamitide BGC in the list without the membrane protein gene (*L*). However, this gene has been reported to be non-essential for the biosynthesis in thiostreptamide S4 (Eyles et al., 2021). More importantly, the exportation mechanism of thioamitide in cyanobacteria might be different, so a different membrane protein/transporter might be involved and could be encoded elsewhere in the genome.

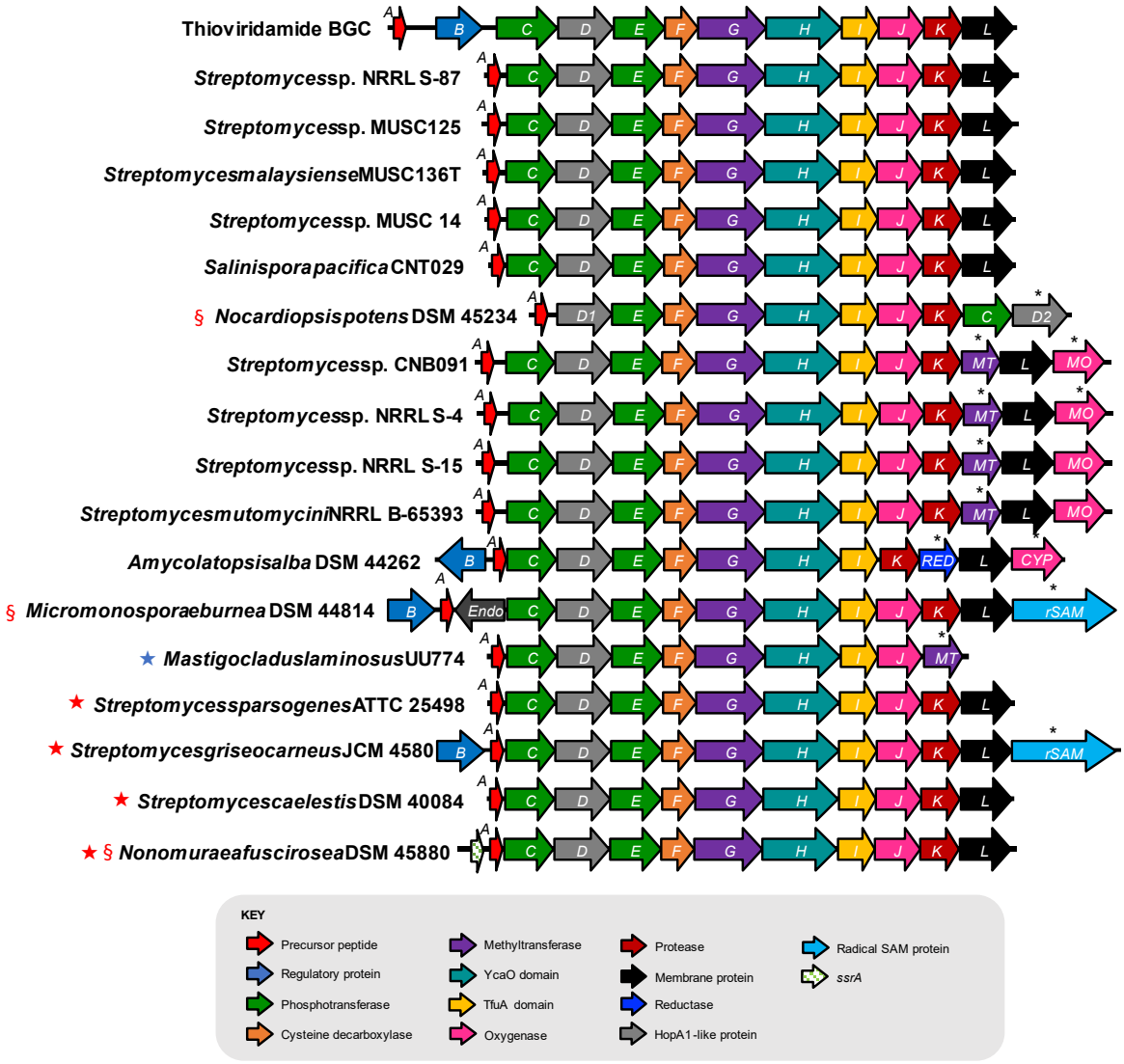


Figure 2.3: Comparison of thioamitide BGCs studied in this project. The thioviridamide BGC at the top is used as reference for comparison. Core tailoring genes are presented in the key below the clusters, while putative additional tailoring enzymes are highlighted with an asterisk above. (CYP: P450; Endo: endonuclease MT: methyltransferase; MO: flavin-dependent oxygenase; RED: reductase). Newly identified BGCs are labelled with a red star, and updated pathways are labelled with a blue star. Strains selected for the study were indicated with a symbol §.

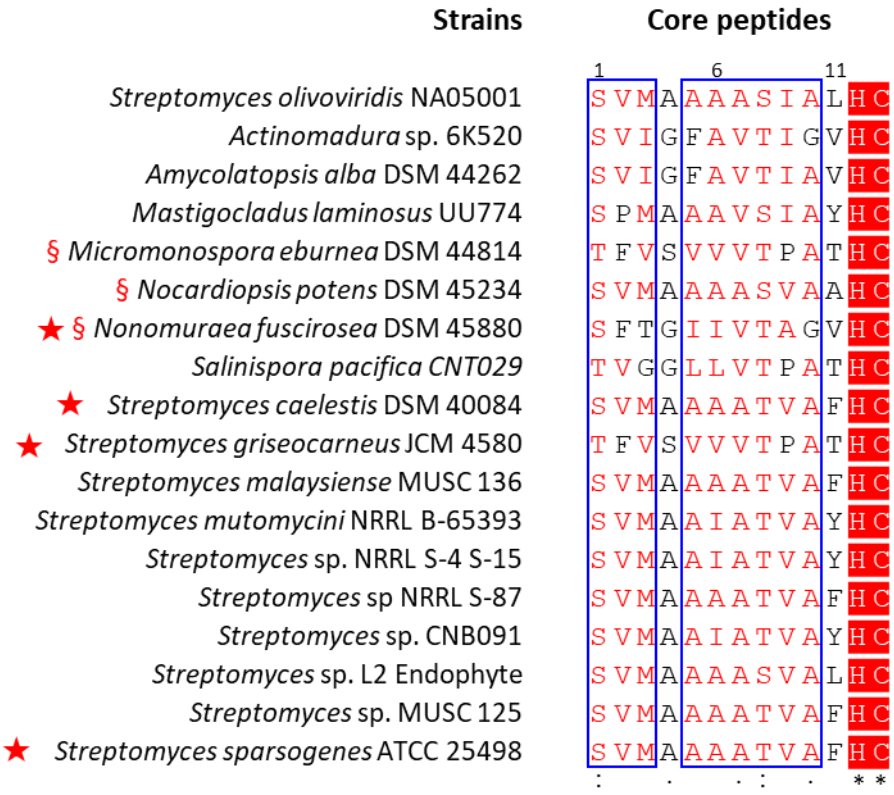


Figure 2.4: Multiple alignment of the 18 thioamitide core peptides from the identified BGCs. MUSCLE was used for alignment, which is visualised using ESPript 3.0 (Robert and Gouet, 2014, Madeira et al., 2022). Identical residues are shown in white with a red background, while similar residues (Risler matrix set with global score of 0.7) are coloured red and are boxed. The core peptides from the newly identified BGCs are labelled with a red star. Strains selected for the study were indicated with a symbol §.

The precursor peptides of these BGCs showed a highly conserved C-terminal region across all precursor peptides, in particular the core peptide region (Figure 2.4, Appendix 4). Generally speaking, all precursor peptides shared the same C-terminal residues, histidine-cystine. A histidine residue is essential for one of the key thioamitide features, dimethylhistidium, and the C-terminal cystine is required for Avi(Me)Cys macrocycle formation. Both positions 1 and 8 of the core peptide are either occupied by hydroxylated amino acids (serine or threonine). Only in rare occasions that the position 1 is a threonine according to these BGC, and there are not any currently characterised thioamitide molecules with a threonine-derived N-terminus. On the other hand, a threonine at position

8 is more common. A serine or threonine at position 8 determines whether an AviCys or an AviMeCys bond is formed respectively. For example, Ser8 (serine in position 8 of the core peptide) from prethioviridamide gives AviCys bond for the macrocycle, while Thr8 from thiostreptamide S4 gives AviMeCys bond for the macrocycle. The rest of the core peptide region is less conserved, but there are some similar amino acid residues in most cases, such as the regular occurrence of alanine residues between positions 4 and 7.

2.3.1.2. Strain selection for screening experiments

Since the aim of the project was to discover structurally diverse thioamitides, the identified clusters were filtered to shortlist interesting candidate clusters to take forward for screening experiments. Therefore, producers of previously characterised molecules such as *Streptomyces* sp. NRRL S-87 and *Streptomyces* sp. NRRL S-4 as well as strains containing almost identical clusters like *Streptomyces malaysiense* and *Streptomyces mutomycini* (Figure 2.3) were not selected for the study. This left the clusters containing the most diverse precursor peptide sequences (Figure 2.4) as well as tailoring genes not previously described in thioamitide pathways. Among these and based on availability of the producer strains in culture collections, three Actinobacteria, *Micromonospora eburnea* DSM 44814, *Nonomuraea fuscirosea* DSM 45880 and *Nocardiopsis potens* DSM 45234, (Figure 2.3) were selected for screening experiments as potential producers of structurally interesting thioamitides.

M. eburnea has a radical SAM protein encoded after a highly conserved membrane protein gene in thioamitide BGCs (Figure 2.3). The radical SAM gene overlaps slightly with the previous gene, indicating they are likely part of the same operon and translationally coupled. Radical SAM proteins can catalyse a variety of post-translational modifications

(PTMs), including methylations, ring formation and anaerobic oxidation (Sofia et al., 2001). However, a BLASTX search on this radical SAM protein showed it contains a B12-binding domain, a typical feature of class B radical SAM methyltransferases, which use cobalamin as a cofactor to add methyl groups to their substrates (Fujimori, 2013, Benjdia et al., 2017, Lu et al., 2020). From perspective of the core peptide, not only does it feature a threonine at position 1, which is rare, the core peptide region is one of the most dissimilar (positions 2 to 9) to the others (Figure 2.4).

With respect to *N. fuscirosea*, its thioamitide BGC was originally annotated with a long non-coding region after the oxygenase gene, but a BLASTX search on this region revealed the presence of a protease homologous to the ones present in other thioamitide BGCs and thought to be responsible for the cleavage of the core peptide of the molecules (Izawa et al., 2013, Sikandar et al., 2022). On the other side of the BGC, this strain has a *ssrA* gene, which encodes a small stable RNA also known as a transfer-messenger RNA, before the precursor peptide gene. The general function of a SsrA RNA is to tag and release stalled ribosomes by providing a recognition site for protease binding (Karzai et al., 2000, Keiler and Ramadoss, 2011). Putting the information together, it was speculated that the expression of protease required assistance from SsrA RNA. Hence, the BGC was originally found interesting as the *ssrA* gene could potentially take part in thioamitide biosynthesis.

Finally, the BGC from *N. potens* was also interesting because of its unusual gene arrangement. In the other thioamitide BGCs, gene *D* (encoding a HopA1-like domain containing protein) is normally clustered in between both phosphotransferase genes, *C* and *E*. However, in *N. potens* one of the phosphotransferases (*C*) is relocated at the end of the BGC, and clustered with an extra HopA1-like gene (*D2*) after the protease gene. Previous

work in our lab showed that the HopA1-phosphotransferase pair (by gene *C* and *D*) functions together to catalyse an elimination mediated dehydration on serine and threonine (Eyles et al., 2021). This was later confirmed by *in vitro* reconstitution of the dehydration of serine and threonine in the thioholgamide and thiostreptamide S87 core peptides (Qiu et al., 2021, Sikandar et al., 2022). Given the rearrangement of gene *C and D* with an extra copy of gene *D* in this BGC, it is not clear whether they would dehydrate the substrate same way as in the other biosynthetic pathways. While the role of the extra HopA1 gene was not clear, making this BGC one of the most intriguing in the whole dataset.

2.3.2. Screening for thioamitide production conditions

Many BGCs are silent or cryptic under standard laboratory conditions (Walsh and Fischbach, 2010, Rutledge and Challis, 2015, Mao et al., 2018). The "one strain-many compound" (OSMAC) approach, growing the bacteria in different culture conditions is one of the simplest and most traditional ways to approach this problem (Bode et al., 2002, Pinedo-Rivilla et al., 2022). As the production of secondary metabolites is often linked to the carbon and nitrogen source of the environment, 11 culture media (section 4.1.4) with variety of carbon and nitrogen sources were selected for thioamitide production screening (van der Meij et al., 2017). Alongside the selected strains, the three actinobacteria, *Amycolatopsis alba*, *Streptomyces* sp. NRRL S-4 and *Streptomyces* sp. NRRL S-87, that had their thioamitides characterised in the lab were also used as positive controls. The cultures were carried out at 30 ^{o}C in Petri dishes over 14 days after which agar plugs were extracted with methanol and analysed by LCMS. Fermentations were also carried out in liquid, where the strains were grown in 10 mL of liquid medium at 30 ^{o}C with 200 rpm shaking in 50 mL Falcon tubes topped with sponge bung to allow air exchange. LCMS samples were prepared by

mixing 500 μL of culture with equal volume of methanol after 14 days of fermentation. However, novel thioamitide production was not consistent in liquid culture, so solid production was preferred in this project.

Both *M. eburnea* and *N. fuscirosea* grew slowly, and only very limited growth was seen on the culture media when they were fermented to 14 days, so cultures were extended up to 30 days for these strains, where both of them appeared to have much better growth and production samples were also taken at this time point.

The discovery of natural products is often complicated by the low production levels, the complexity of the samples for screening and the uncertainties predicting the products of biosynthetic pathways. In the case of RiPPs, the core section of the precursor peptide provides the backbone structure of the final product, and the PTMs on the backbone can usually be predicted based on the tailoring enzymes encoded in the BGC, facilitating product identification.

However, thioamitide product prediction is less precise for 2 reasons. First of all, thioamidation can occur on multiple residues, and the number of thioamide bonds in the molecule varies from species to species (Hayakawa et al., 2006b, Frattaruolo et al., 2017, Kjaerulff et al., 2017). The control of multiple thioamide bond installation has not yet been characterised so it is impossible to determine exactly how many thioamide bonds will be present in a molecule. As per the currently characterised compounds, thioamidation occurs only on the linear part of the molecule, so it was expected that there should be no more than 7 thioamide bonds in a thioamitide. Second, although all 3 selected pathways have the conventional tailoring enzymes, the role of the additional enzymes such as the radical SAM protein at the end of *M. eburnea* or the extra HopA1 in *N. potens* was still uncertain.

Thus, the potential novel thioamitides from those three strains were expected to have masses anywhere between 1100 and 1500 Da (Table 2.1). Potential masses for the putative products of these pathways were calculated using the mass of the expected core peptides in each case plus the conventional modifications catalysed by the core tailoring enzymes: dehydrations (phosphotransferase, HopA1 protein), decarboxylation (cysteine decarboxylase), methylations (methyltransferase), and hydroxylation (oxygenase). To account for the potential different degrees of thioamidation, minimum masses corresponding to molecules carrying just 1 thioamide bond as well as maximum masses for molecules with all 7 possible thioamidations in the linear N-terminal chain were calculated. In the case of *M. eburnea*, 14 Da was also added to the maximum mass to account for a potential methylation by the radical SAM protein in the cluster.

Table 2.1: Summary of predicted molecular masses of putative thioamitides from the strains selected for screening. Minimum masses were predicted using the core peptide with 1 thioamidation and the conventional modifications; maximum masses were predicted the same manner but with 7 thioamidations. *The thioamitide from *M. eburnea* also had a methylation added to predict the maximum mass.

Strains	Core peptide sequence	Mass (Da)		
		Unmodified	Minimum	Maximum
M. eburnea	TFVSVVVTPATHC	1359.68	1339.67	1449.67*
N. fuscirosea	SFTGIIVTAGVHC	1303.66	1283.64	1379.64
N. potens	SVMAAAASVAAHC	1187.54	1167.53	1263.53

In addition to the predicted mass range for potential thioamitide molecules, previous reports indicated that MS/MS spectra of thioamitides often show a signature neutral loss pattern, a sequential loss of 33.99, corresponding to hydrogen sulphide (H_2S) resulting from the fragmentation of thioamide bonds (Larsson, 1973, Frattaruolo et al., 2017, Santos-

Aberturas et al., 2019). This characteristic mass loss should help in the identification of any novel molecules.

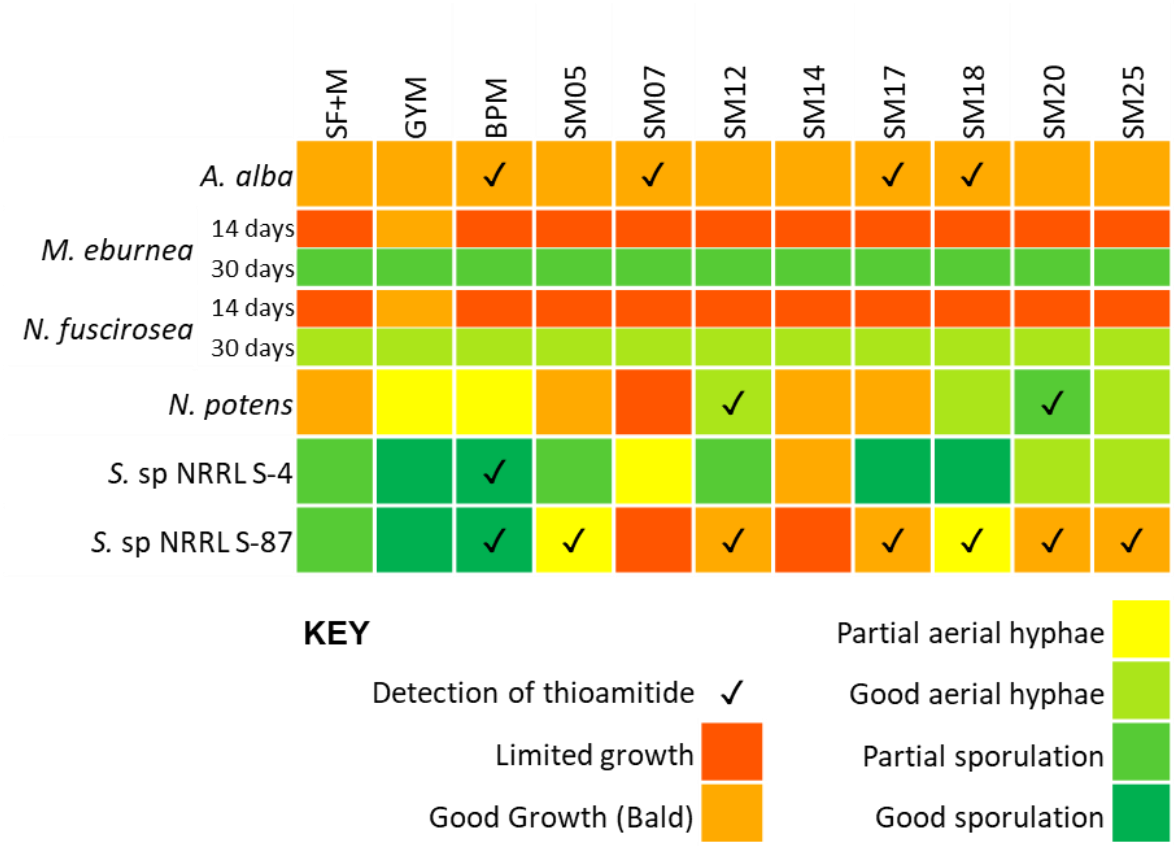


Figure 2.5: Summary of growth and thioamitide production in the screening media panel (section 4.1.4). The observations on the growth of strains and detection of thioamitide production were taken on day 14 unless stated otherwise.

Based on the parameters established for the LCMS search of thioamitides, no production was detected for *M. eburnea* or *N. fuscirosea* either at 14 or 30 days. On the other hand, *N. potens* seemed to produce a potential novel thioamitide (Figure 2.6). This compound, now named thiopotensamide A, had a *m/z* of 1243.421 and sequential neutral losses of 33.99 were observed in its MS/MS spectrum. Therefore, further characterisation of this compound was carried out.

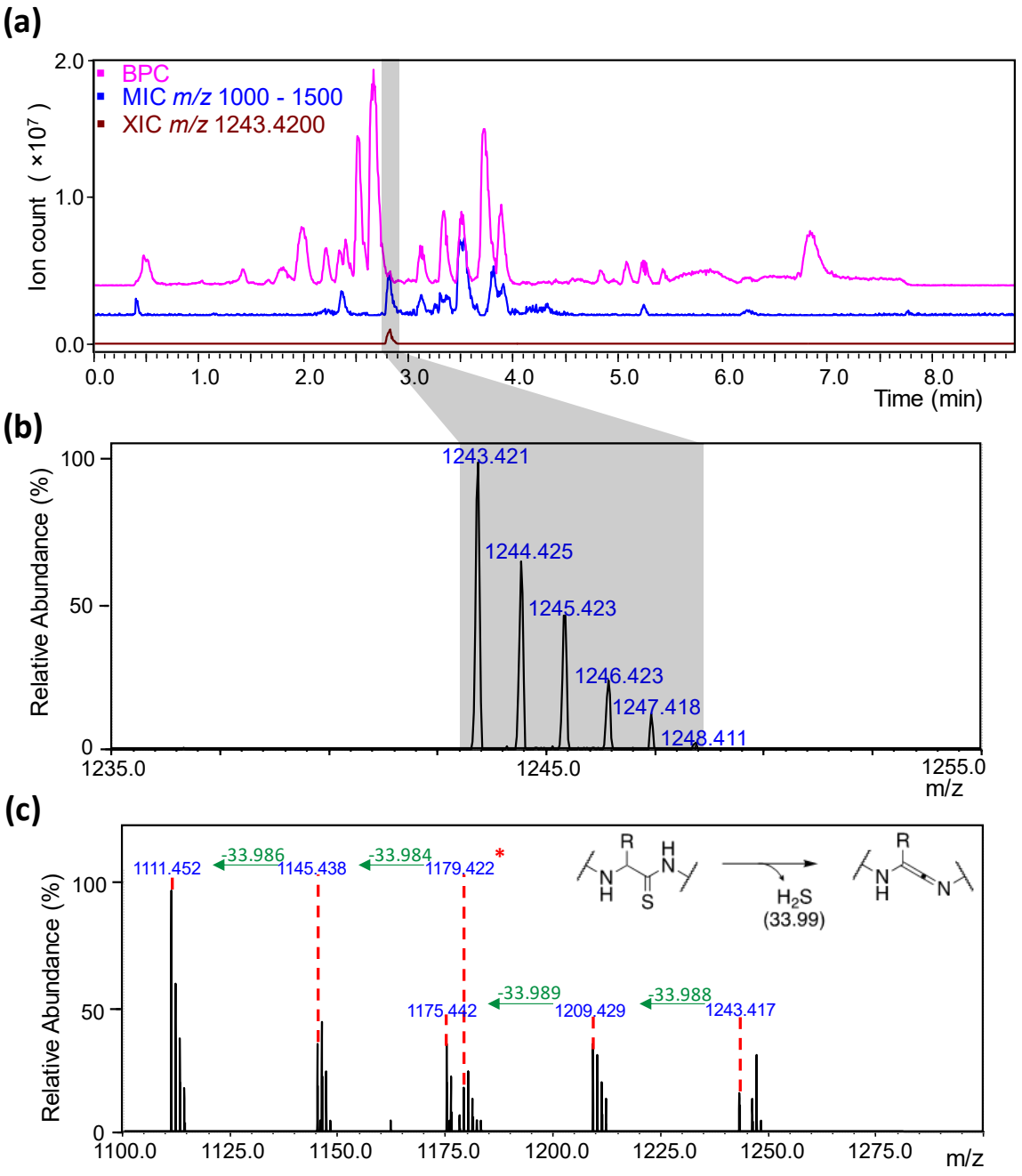


Figure 2.6: LCMS analysis of thiopotensamide A in a culture sample from *N. potens* on SM20 after fermentation for 14 days. (a) LCMS chromatograms of the sample: base peak chromatogram (BPC); multiple ion chromatogram (MIC) covering the mass range between 1000 and 1500 Da used for screening; extracted ion chromatogram (XIC) for *m/z* 1243.42 corresponding to thiopotensamide A. (b) MS spectrum of thiopotensamide A. (c) MS/MS spectrum showing the signature neutral loss of hydrogen sulphide (H_2S) from thioamide bonds, along with the schematic representation of the fragmentation process.

2.3.3. LCMS based structural characterisation of thiopotensamide A from *Nocardiopsis potens*

As mentioned previously, the MS/MS spectrum of thiopotensamide A was a strong indication that this molecule was in fact a thioamitide, since it contained four 33.99 Da signature mass losses associated with the presence of thioamide bonds. Interestingly, only two of these neutral losses could be directly associated with the precursor ion with *m/z* 1243.42, while the other two seemed derived from fragment with *m/z* 1179.42, which is 63.99 Da smaller than the precursor ion (Figure 2.6C). This neutral mass loss has been reported to correspond to the loss of methane sulfenic acid and is a signature of the presence of an oxidized methionine in the molecule (Figure 2.7) (Guan et al., 2003, Miles et al., 2014). A similar neutral loss of 63.99 from the molecular ion had previously been observed during the characterisation of two analogues of thiostreptamide S4 and S87, where it was hypothesised that the sidechain sulphur in the methionine residue from the core peptide had been oxidised to generate methionine sulphoxide, probably by reactive oxygen species (ROS) (Liang et al., 2012, Frattaruolo et al., 2017). Therefore, this observation was consistent with thiopotensamide A being a thioamitide featuring a sulfoxide group on its methionine residue.

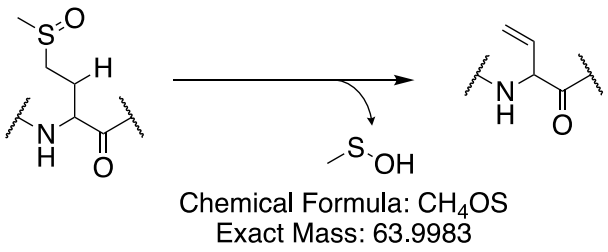


Figure 2.7: Schematic MS/MS fragmentation of methionine sulphoxide (MetO).

Based on the genes present in the *N. potens* BGC, and previous *in vivo* studies in thioviridamide and thiostreptamide S4 and the *in vitro* study in thioholgamide, a series of tailoring modifications potentially present in thiopotensamide A could be predicted (Izawa et al., 2013, Eyles et al., 2021, Sikandar et al., 2022). These modifications included dehydrations on Ser1 and Ser8 to form two 2,3-dehydroalanine (Dha) residues (putatively catalysed by the products of *tpaC*, *tpaD1* and/or *tpaD2*), hydroxylation and double methylations (carried out by the products of *tpaJ* and *tpaG* respectively) to form β-hydroxy-N^1, N^3-dimethylhistidinium (hdmHis12), cysteine decarboxylation-led ring formation (by the enzymes encoded by *tpaE*, *tpaF*) on Dha8, and pyruvyl formation through leader peptide removal next to Dha1 (Eyles et al., 2021, Qiu et al., 2021, Sikandar et al., 2022). With respect to the number of thioamidations, although the MS/MS spectrum of thiopotensamide A indicated the presence of four thioamide bonds, the presence of further thioamidations could not be discarded, so two structures (predicted structure I and II) were predicted *in silico*, each containing four and five thioamidations respectively (Figure 2.8).

Interestingly, the observed mass for thiopotensamide A did not match either predicted structure I nor II. The calculated mass for predicted structure I, carrying 4 thioamide bonds, had a difference of -12 Da, while predicted structure II, with 5 thioamide bonds, had a mass difference of +4 Da. These mass differences implied at least 1 or more extra PTMs were installed in thiopotensamide A but were missing in the original predictions.

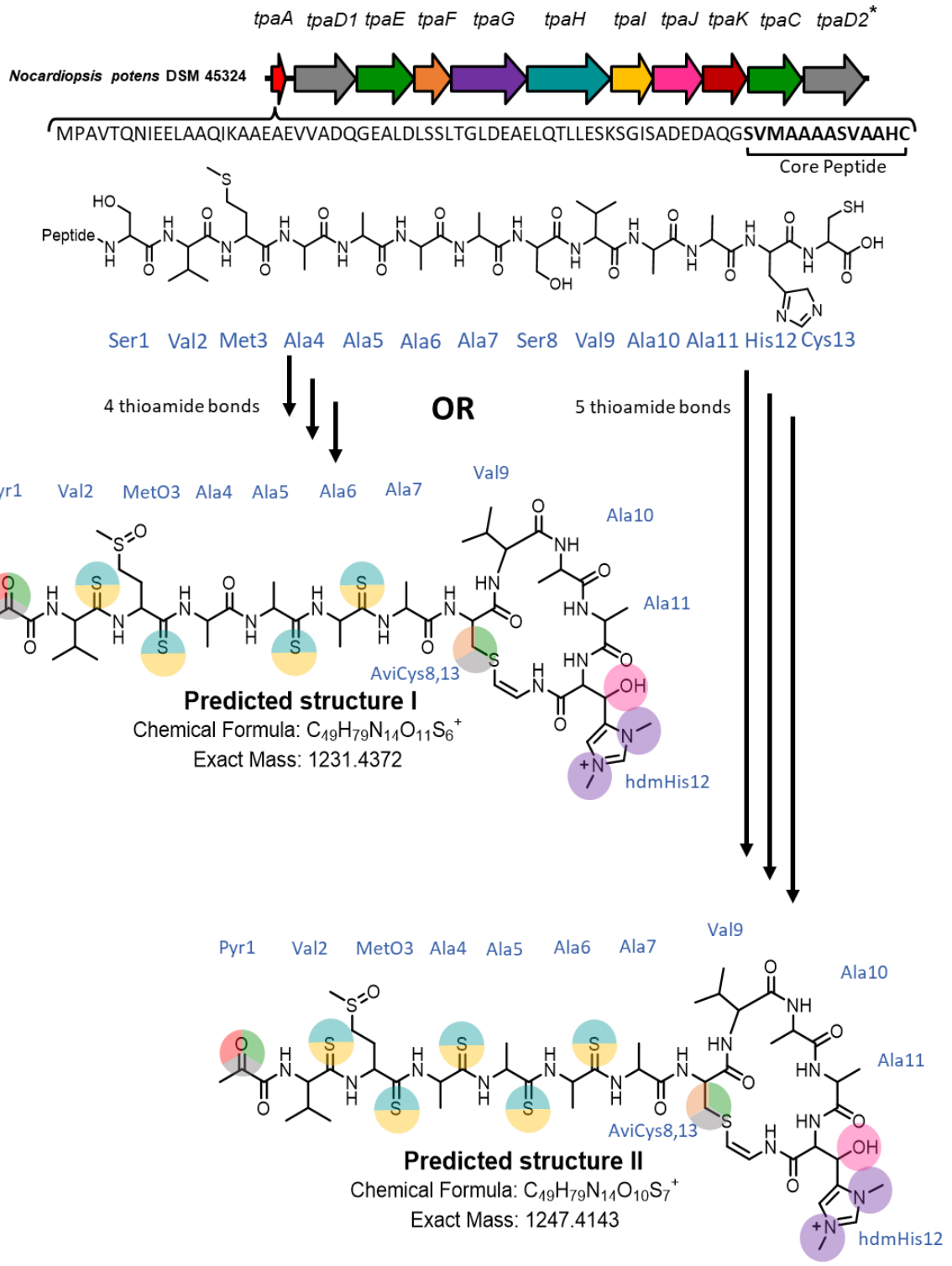


Figure 2.8: *in silico* generated potential structures (predicted structure I and II) of the thioamitide product from *N. potens* BGC. "Peptide" represents the leader peptide.

There was not a single PTM that would modify the molecule by +12 or -4 Da in a single step, suggesting at least 2 extra modifications on the molecule were required to give a mass of

1243.42 Da as final product. In order to identify the extra modifications, culture extracts containing thiopotensamide A were analysed on a Synapt G2-Si mass spectrometer applying targeted fragmentation on the molecule. The structural information acquired from thioamitides previously characterised in the lab was used as reference for comparison. In particular, the MS/MS fragments corresponding to the AviCys macrocycle from each of the molecules were particularly useful for locating where the extra modifications were installed.

The expected mass for the macrocycle fragment resulting from predicted structures I or II could not be observed in MS/MS fragmentation of thiopotensamide A, suggesting that one or more of the missing modifications in these predictions were located in the macrocycle region. The *m/z* range between 50 and 200 features fragments that are derived from the histidine in the AviCys cycle. The spectra from thiostreptamide S4 and thiostreptamide S87 share a fragment with *m/z* 125.07 consistent with the retro-aldo fragmentation from hydroxylated histidine, whereas neither thioalbamide nor thiopotensamide A have this fragment (Figure 2.9). In the case of thioalbamide, it is known this hydroxylation does not occur in the histidine residue due to the absence of a monooxygenase gene (*J*) in the BGC.

Unlike the thioalbamide biosynthetic pathway, this retro-aldo fragment was expected from MS/MS of thiopotensamide A as the monooxygenase gene *tpaJ* is in the BGC, so its absence was surprising and indicated that the histidine was not hydroxylated in thiopotensamide A. Given the presence of *tpaJ*, it is possible that hydroxylation occurs, but the residue is further modified. One possibility was that the hydroxyl group was removed through elimination. There are also fragments between thiopotensamide A and thioalbamide MS/MS spectra showing -2 unit differences, for example (95.06 vs 97.07 and 136.09 vs

138.10). This mass difference indicates the potential presence of unsaturated hydrocarbon in thiopotensamide A and saturated hydrocarbon in thioalbamide. Although this kind of modification was not previously described in any thioamitide, it was hypothesised it could occur in an analogous way to the dehydration of the serine residues in the molecule (Figure 2.10a).

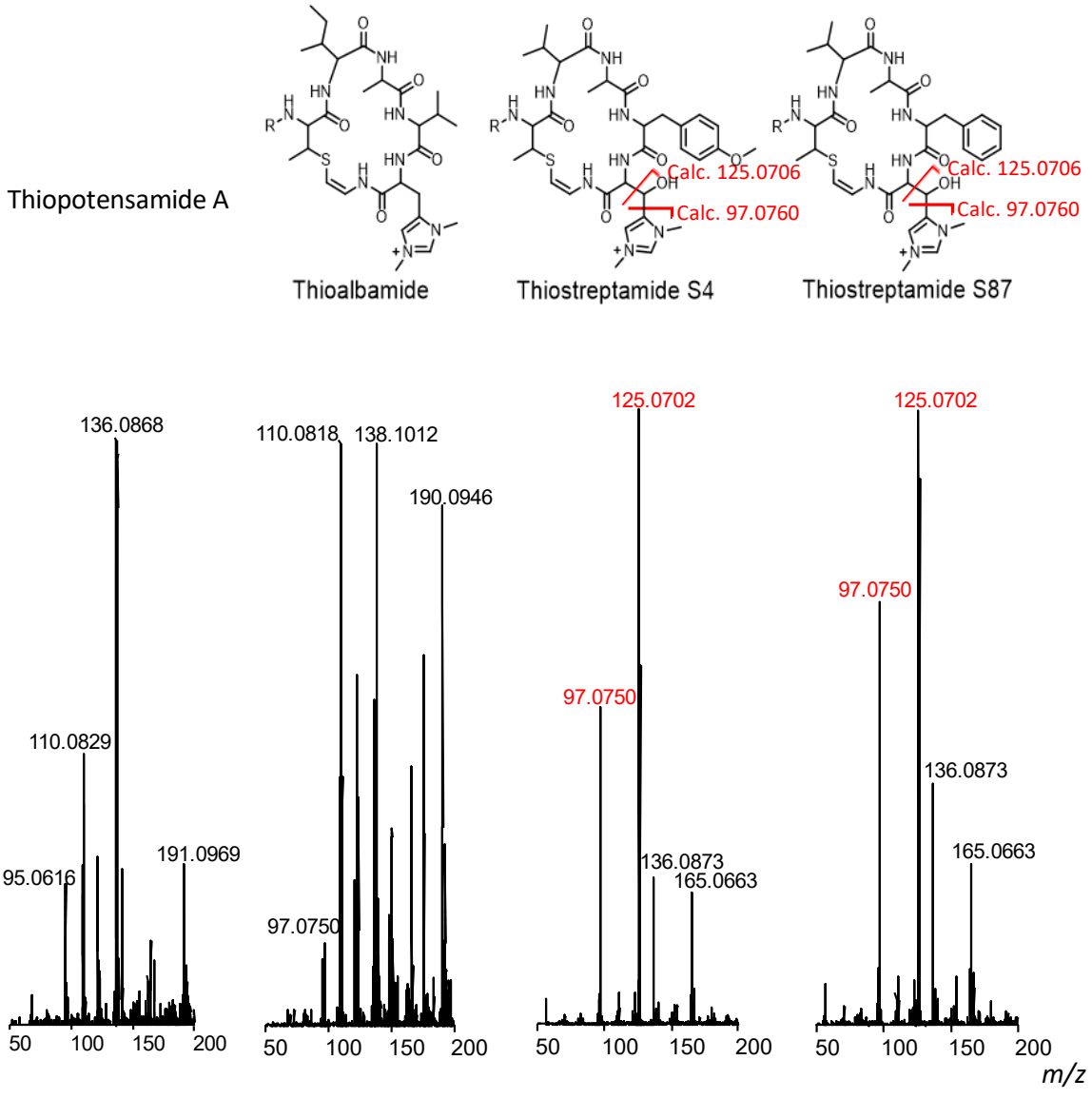


Figure 2.9: low *m/z* range of MS/MS spectra from thiopotensamide A, thioalbamide, thiostreptamide S4 and thiostreptamide S87. Fragmentation patterns for the dimethylhistidinium side chain in these molecules is depicted above the spectra and the corresponding observed fragments are highlighted in red.

Adding the potential dehydration to predicted structure II reduced its mass by 18 Da and generated an *in silico* structure with a calculated exact mass of 1229.4037 Da (predicted structure III, Figure 2.10b). The predicted fragmentation pattern for this structure was entirely consistent with the fragments observed in the mid-range (*m/z* 500 – 900) MS/MS spectrum from thiopotensamide A (Figure 2.11), providing strong support that this modification is present in the molecule.

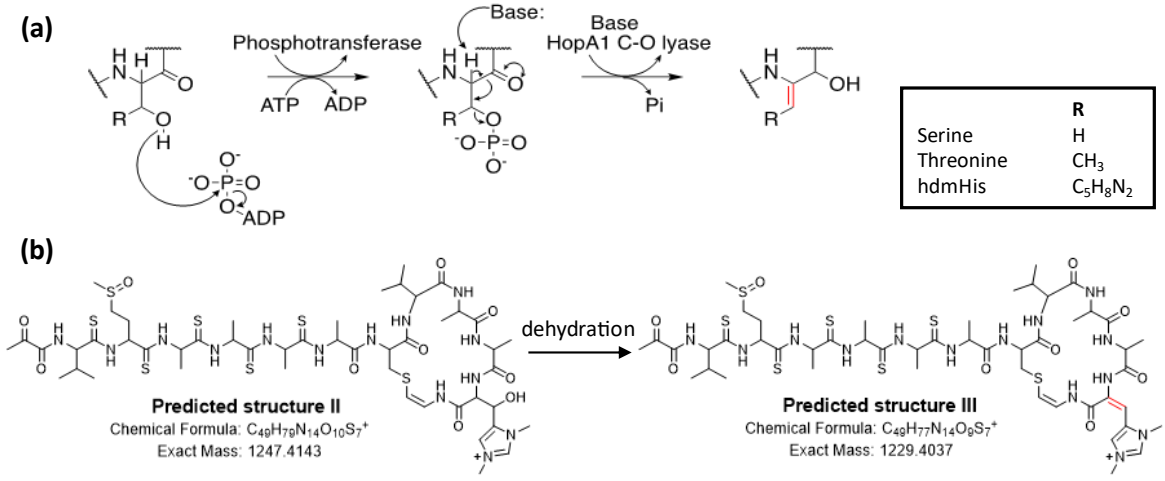


Figure 2.10: Predicted dehydration of dimethylhistidinium (hdmHis) in thiopotensamide A. (a) Schematic representation of phosphotransferase/HopA1 mediated dehydration hydroxylated side chains of serine, threonine or hydroxylated hdmHis. (b) Elimination mediated dehydration on predicted structure II to generate predicted structure III.

More specifically, multiple fragment ions from thiopotensamide A have the expected *m/z* given from fragments (y_1, y_2, y_3, y_4 and y_5) from predicted structure III, with y_1 corresponding to the macrocycle fragment itself (Figure 2.11). Not only could the fragment ions directly derived from the precursor ion be identified, but extra mass losses from some of these fragments could also be detected in the MS/MS spectrum. For example, all product ions from y_1 to y_4 showed alternative fragments with masses consistent with ammonia losses (NH_3; 17.0265 Da), from their N-terminal amines (Figure 2.11). In addition, y_3 to y_5 showed alternative fragmentation consistent with H_2S loss (33.99 Da) corresponding to thioamitide bonds. y_1 and y_2 did not show 33.99-minus fragments, indicating that Ala4, Ala5

and Ala6 had thioamide bonds, whilst Ala7 was not likely to have this modification. In addition to this, the analysis of the high mass range (above *m/z* 1000) MS/MS fragments of 1 corresponding to fragments containing increasing lengths of the N-terminal linear chain revealed the two-remaining predicted thioamidations are likely to be installed in Val2 and MetO3.

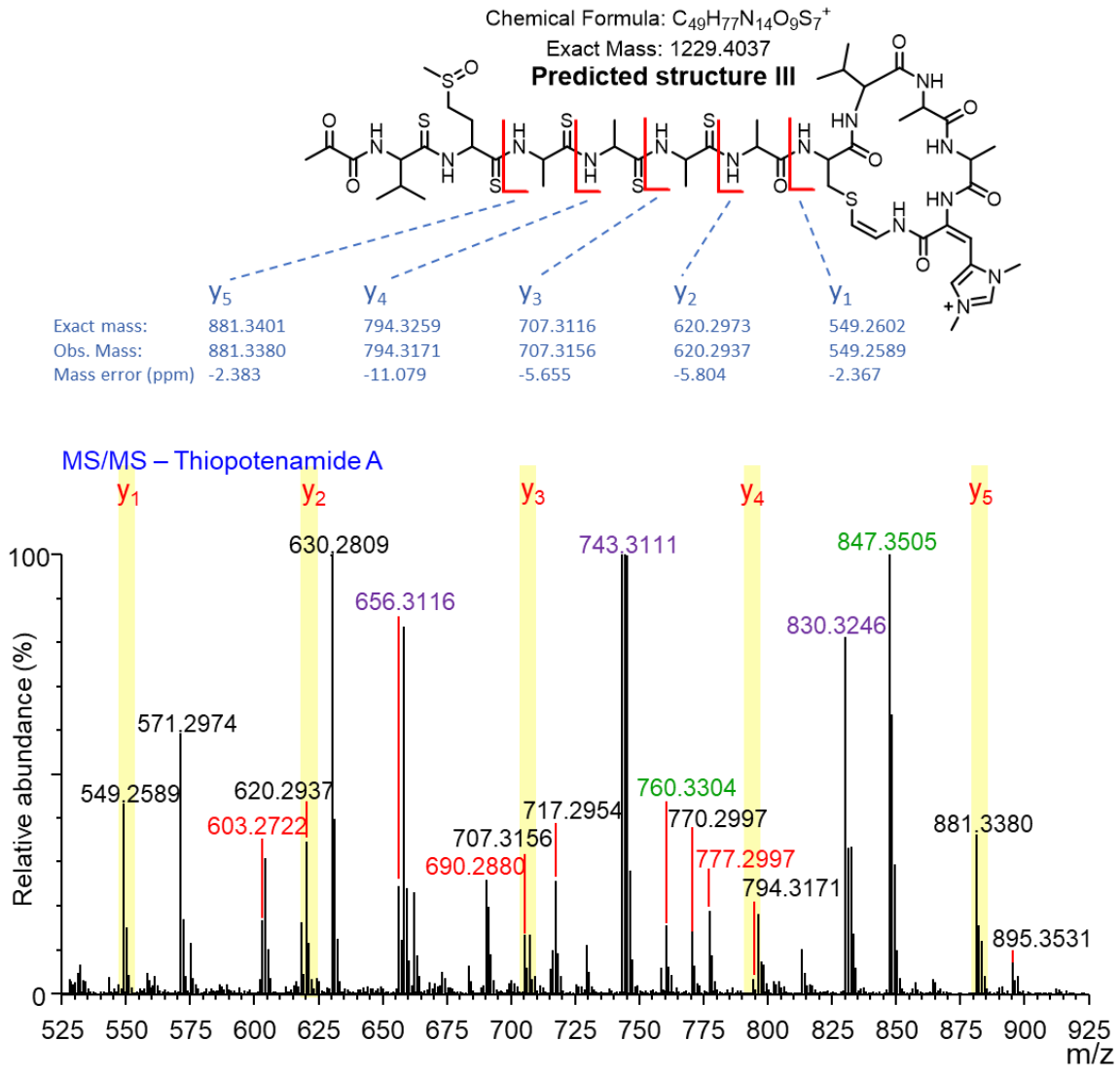


Figure 2.11: Mid-range of MS/MS spectrum from thiopotensamide A obtained with a Synapt mass spectrometer showing fragments consistent with predicted structure III. Main peptide fragments are highlighted in yellow and their predicted and observed masses are listed in the figure with mass errors. Fragments with ammonia and hydrogen sulphide losses are labelled in red and green respectively, while loss of both is labelled in purple.

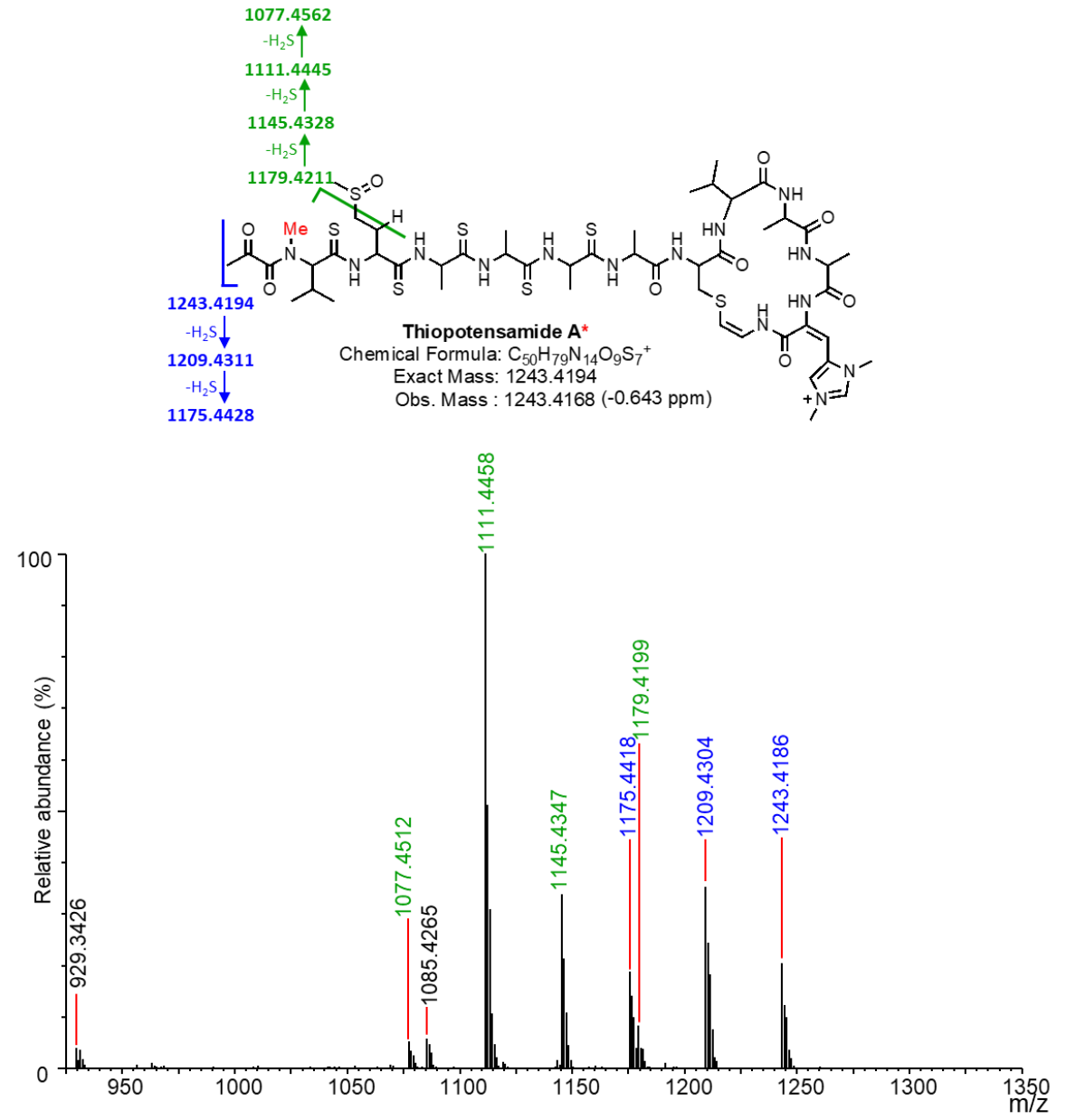


Figure 2.12: Tentative thiopotensamide A structure and the high *m/z* range of MS/MS spectrum from thiopotensamide A obtained with a Synapt mass spectrometer showing fragments consistent with the proposed structure. Masses in the spectra are colour coded to match the predictions; predicted masses are labelled in bold. * The methyl group (in red) was drawn on one of the possible modified positions.

Finally, the remaining mass difference between thiopotensamide A and predicted structure III was 14 Da, indicating the other missing modification was a methylation. Fragmentation data indicated that this extra methylation had to be installed in the N-terminus of the molecule, likely on Pyr1, Val2 or MetO3. However, the exact methylation location was not

clear as no diagnostic fragments for this part of the molecule could be detected. Figure 2.12 shows a tentative structure with the methyl group on the N-terminal amide nitrogen, but the methyl group could be elsewhere.

In summary, the accurate mass and fragmentation pattern of thiopotensamide A are consistent with a thioamitide containing an AviCys macrocycle with a dehydrated dimethyl histidinium, 5 thioamidations, a sulphoxide methionine residue, an N-terminal pyruvyl group and a methylation located in an undefined position in the N-terminus of the molecule. This molecule was therefore named thiopotensamide A as the main product of the thioamitide BGC from *N. potens*. Since some of the proposed modifications in thiopotensamide A are novel and in particular the methylation position could not be determined, NMR spectroscopy is required in order to confirm the structure of the molecule (Figure 2.12).

From all the proposed PTMs in thiopotensamide A, the most intriguing ones are the histidine dehydration and the extra methylation. With regards to the dehydration, it was speculated that the additional HopA1 domain containing gene (*tpaD2*) in the BGC could be involved in it, since it is unique to the *tpa* BGC. The HopA1 (TvaD homologue) and phosphotransferase (TvaC homologue) protein pair were originally proposed to dehydrate serine and threonine in thioamitide core peptides (Eyles et al., 2021). This was later validated genetically and *in vitro* showing the protein pair catalyse an elimination mediated dehydration (Figure 2.10a) (Qiu et al., 2021, Sikandar et al., 2022). Since the side chains of serine/threonine and the hydroxylated histidine both feature a β-hydroxyl group, having a phosphotransferase-HopA1 pair catalyse this dehydration of hydroxylated histidine was theoretically possible. As this dehydration had not been reported in any of the

characterised thioamitides to date, this dehydration was not likely to be introduced by the regular phosphotransferase-HopA1 pair. However, the additional HopA1 gene (either *tpaD1* or *tpaD2*) in the cluster might have a role in the additional dehydration to the molecule.

Similarly, there was no clear evidence of which enzyme was responsible for the methylation at the N-terminus of thiopotensamide A. The *tpa* BGC was thought to span from *tpaA* to *tpaD2*, in a similar way to the other characterised thioamitide BGCs, which have their essential tailoring enzyme genes clustered in one operon. This cluster does have one methyl transferase gene (*tpaG*), which is predicted to be responsible for the double methylation of the histidine residue. While there was a possibility the extra methylation was introduced by the product of this gene, this was unlikely since such modification had not been detected in any other thioamitides. Alternatively, the possibility that the methylation was introduced by other enzymes outside the described BGC originally could not be ruled out.

2.3.4. Purification of thiopotensamide A

2.3.4.1. Preliminary separation

While MS can provide a lot of information regarding the structure of a molecule, NMR is considered the gold standard to characterise unknown compounds, especially in the absence of a standard that could be used for LCMS comparisons. MS only requires small amount of sample, but NMR requires a larger amount and higher purity of sample. In order to obtain enough material to perform NMR spectroscopy on thiopotensamide A**,** production cultures of *N. potens* had to be upscaled and the compound extracted from the culture and purified.

Initial purification efforts started from batches of 2 L of solid SM12 medium cultures plated on 400x400 mm square Petri dishes and incubated at 30 ^{o}C for two weeks. The agar was then chopped and extracted with an equal volume of methanol. This methanol crude extract was then filtered to remove cell debris, dried and resuspended in 500 mL of water.

Liquid-liquid extraction of the crude aqueous suspension with ethyl acetate and a normal phase flash chromatography were used to remove the bulk of salts and sugars from the extract. The resulting sample was subjected to preparative high performance liquid chromatography (HPLC) on a C18 column applying the method used to purify previously characterised thioamitides (Hayakawa et al., 2006b, Izumikawa et al., 2015, Frattaruolo et al., 2017). When the preparative HPLC results were analysed, thiopotensamide A appeared to co-elute with a molecule (compound 528) with *m/z* 528.35, which was found to have very high ion intensity, which was usually over two orders of magnitude more than the intensity for thiopotensamide A regardless of HPLC conditions.

2.3.4.2. Size exclusion separation test

The observed mass for compound 528 corresponded to a singly charged adduct, likely a $[M+H]^+$ indicating that this compound could have a molecular mass of 527.35 Da. Since compound 528 only has half of the molecular weight of thiopotensamide A, separation of compound 528 from thiopotensamide A was attempted on a Sephadex LH20 size exclusion column (SEC). Frustratingly, the SEC fractions containing thiopotensamide A also had substantial amounts of compound 528 (Figure 2.13). It was then briefly considered that compound 528 could be a product from the degradation of thiopotensamide A. However, a MS/MS analysis showed that thiopotensamide A and compound 528 had completely different fragmentation profile, so this hypothesis was soon abandoned. Other possible

explanations for this surprising result could be related to the polarities of thiopotensamide A (with native positive charge) and compound 528 (neutral molecule), and/or the differences in molecular shape and volume. Calderon and Baumann also found that octadecenyl ethanediol ether (molecular weight 312.5) eluted earlier than cholesterol (molecular weight 386.7) from a LH20 column despite the fact that they have very little polarity differences in the system, and tests on separation of isomers (*cis*- and *trans*-9-octadecenyl ethanediol ether) showed the early elution of ethanediol monoether ether was due to their molecular dimensions (Calderon and Baumann, 1970).

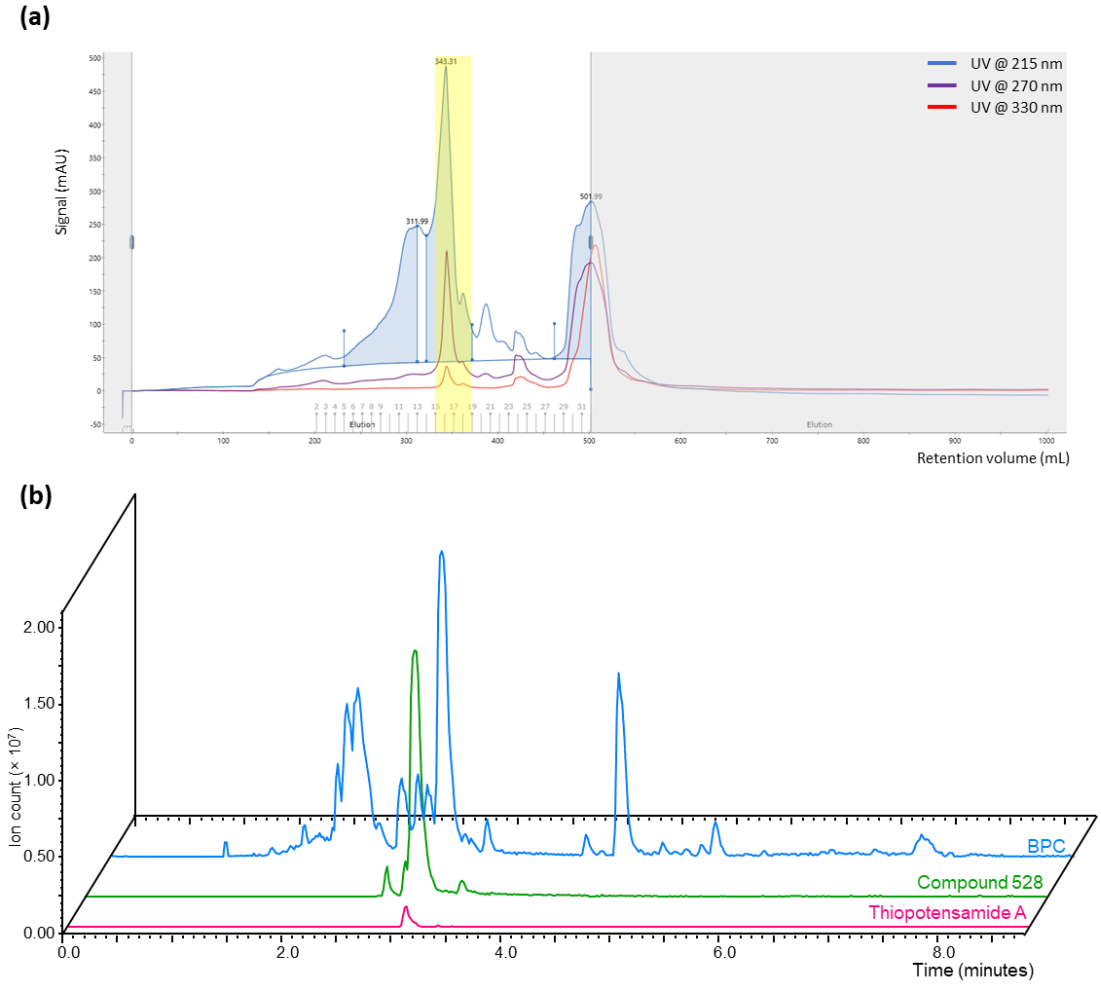


Figure 2.13: (a) Size exclusion purification with fractions containing thiopotensamide A highlighted in yellow and (b) Chromatograms of a representative SEC fraction sample showing a substantial amount of the key impurity, compound 528.

2.3.4.3. Preparative HPLC separation

In parallel, this semi-purified material was tested with a variety of analytical columns with different stationary and mobile phase combinations in an attempt to improve the separation between thiopotensamide A and compound 528. Kinetex C18, Luna Omega polar C18 and Luna PFP columns were tested in combinations with two different solvent systems: H_2O (0.1% formic acid, FA) against acetonitrile (ACN with 0.1% FA) and H_2O (0.1% FA) against methanol (0.1% FA) (Figure 2.14) as well as with different solvent gradients. None of these tests made a difference in either Kinetex C18 nor Luna Omega polar C18 columns, but separation of thiopotensamide A and compound 528 was achieved in a 50 mm analytical PFP column with a gradient of 40 to 70% methanol over 6 minutes (Figure 2.14, section 4.2.11). The separation obtained in the PFP column was considered as significant as the elution gap between thiopotensamide A and compound 528 was more than 2 column volumes (CVs). Hence, purification was attempted using a PFP semipreparative column with H_2O against methanol with 0.1% FA as additive to each.

The separation of thiopotensamide A and compound 528 in the semipreparative PFP column was attempted using 3 different gradients (Figure 2.15). The first attempt was a 45 to 98% methanol gradient in 42 minutes. Fractions containing thiopotensamide A were collected at retention time (R_t) 38-39 minutes (78.0% methanol) (Figure 2.15a). Unfortunately, these fractions still contained a mixture of thiopotensamide A and compound 528, possibly due to the high abundance of compound 528 leading to peak broadening and tailing in the elution. The fractions from the previous separation round were pooled and re-injected. This sample should contain reduced amount of compound 528, and the elution time was expected to be shifted. This sample with reduced amount of compound 528 did not change the R_t of the compound and once again the fractions

containing thiopotensamide A were a mixture of thiopotensamide A and compound 528 as shown from LCMS analyses.

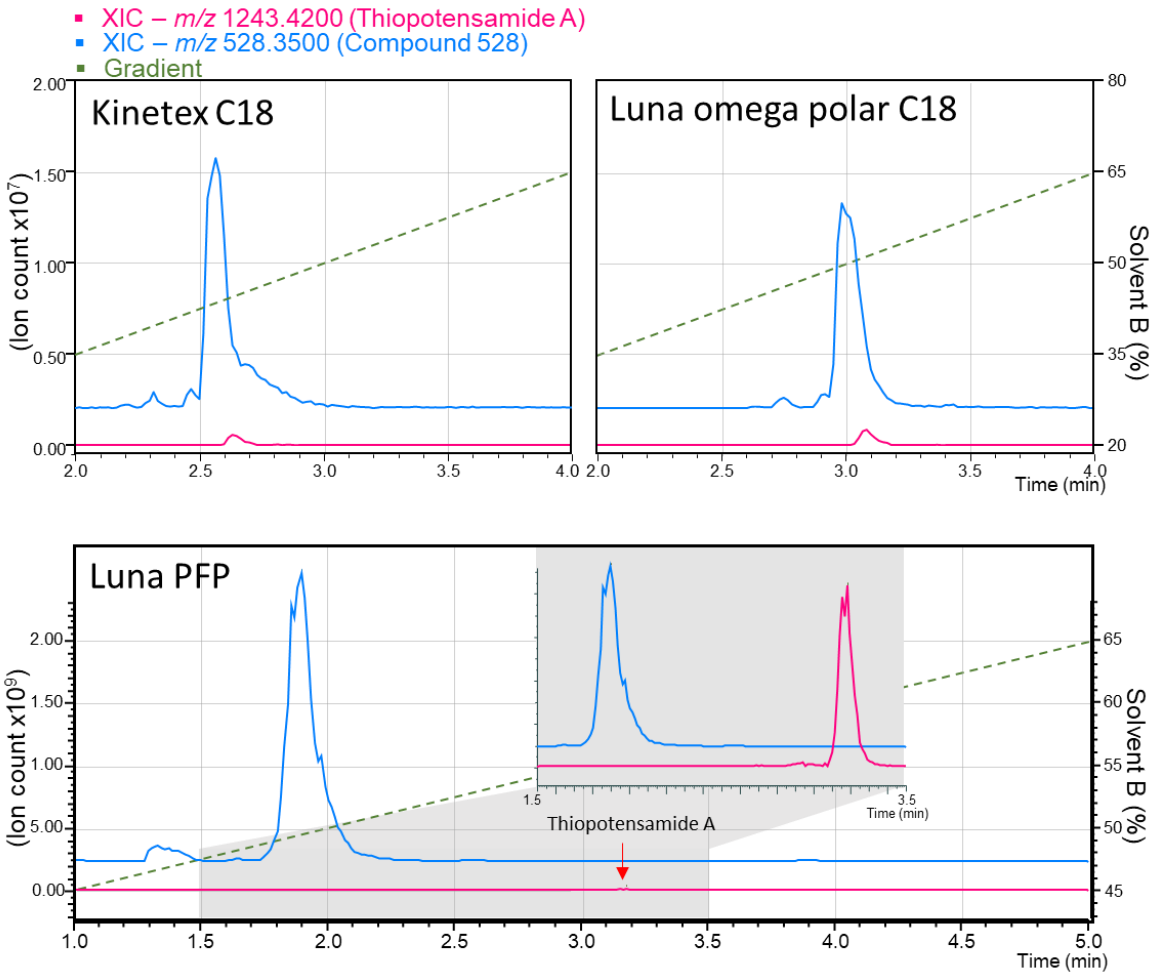


Figure 2.14: Extracted ion chromatograms of thiopotensamide A (*m/z* 1243.420, pink) and compound 528 (*m/z* 528.354 blue) from culture samples separated in analytical columns with different stationary phases. Chromatograms were normalised to the ion peak intensity of compound 528 in each chromatogram. For visualisation purposes, a detail of the chromatogram from Luna PFP is shown in grey where the XIC are not normalised to the same scale to show the presence of thiopotensamide A.

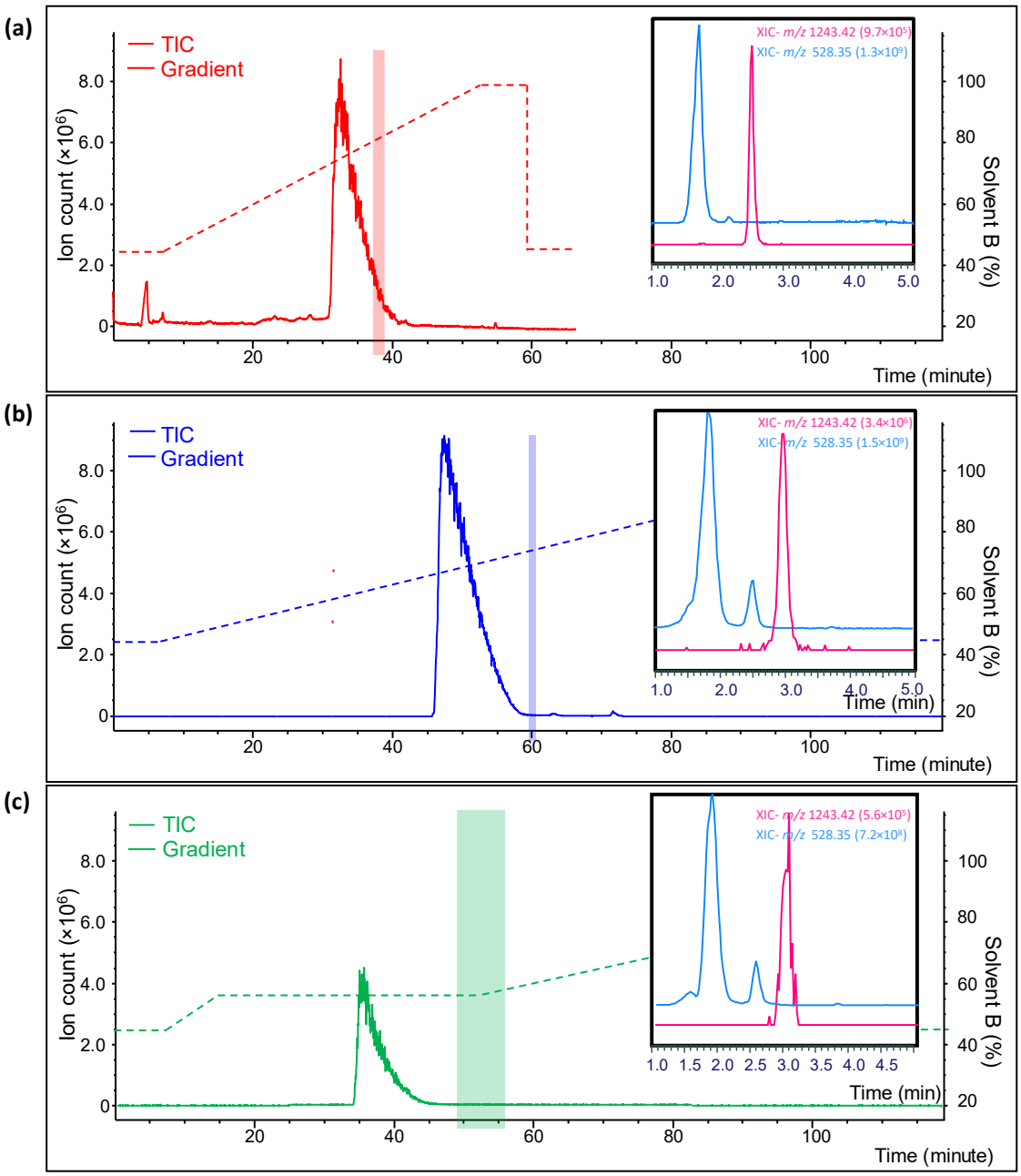


Figure 2.15: Results of the different gradients tested on the 250 mm semipreparative Luna PFP column. (a) HPLC separation using a 42-minute gradient from 45 to 90% solvent B (methanol). (b) HPLC separation using an 84-minute gradient from 45 to 90% solvent B. (c) HPLC separation using a 3-step gradient. The solid line in each chromatogram represents the TIC of the run, while the dotted line represents the gradient employed. In each chromatogram, the shaded R_t region indicates where thiopotensamide A was eluting in each case and the XICs of the representative fractions from each experiment are shown in the box on the right. For visualisation purposes, the XICs are not normalised to the same scale and the intensity of the peak in each case is shown in brackets after the *m/z*.

It was also considered that the gradient was potentially too steep in terms of CVs, so the extent of thiopotensamide A and compound 528 separation was not as great as in the analytical trials. An extended run was also tested with the same gradient but over 84 minutes instead of 49 minutes, to allow the elution of 7 CVs from the 250 mm semipreparative column. In this case, fractions containing thiopotensamide A were collected at R_t 59.5-60.5 minutes (73.1% of methanol) (Figure 2.15b). However, the separation of thiopotensamide A and compound 528 was again unsuccessful as the fractions were also the mixture of two. In order to gain a better separation and to avoid overloading the column, an isocratic step was introduced to the gradient and half amount of the material was injected for the following run (Figure 2.15c). The elution time of thiopotensamide A widened to 7 minutes between R_t 49 and 56 minutes, but again HPLC-HR-MS of these fractions revealed compound 528 was still present in them.

A potential explanation for the poorer separation observed in the semipreparative runs versus the analytical ones could be that the analytical PFP column had slightly smaller particle size (semipreparative: 5 µm; analytical: 3 µm), which could affect the separation resolution. Hence, as a last resort, a low scale purification was attempted using the analytical PFP column. To build up enough sample for analysis, ten separations were carried out using the analytical gradient and flowrate (40 to 70% methanol over 6 minutes, 0.6 mL min^{-1}) and the resulting fractions were pooled, concentrated to 1/10 volume, and analysed by LCMS. However, in this case, only compound 528 was detected in some of the fractions, while thiopotensamide A was not observed in any of the fractions, indicating that the compound had been diluted beyond detection.

2.3.4.4. Ion-exchange solid phase extraction

Another strategy was to employ cation exchange chromatography, since compound 528 is likely to be a neutral compound, whereas thiopotensamide A is natively carrying a positive charge at the dimethylhistidinium residue. In general, ion exchange separation is mainly based on the electrostatic interaction of ions between the mobile phase and a stationary phase with integrated ionic moieties (Anand et al., 2001). The ionic species in the solute bind to the stationary phase. This binding is reversible and can be controlled by several factors, such as the pH and the nature of the mobile phase, as well as the ionic strength, molecular weight and charge intensity of the analyte (Anand et al., 2001). In short, thiopotensamide A was expected to bind to cation exchange resin whereas compound 528 should not interact with the resin, and thiopotensamide A would be eluted with acidic buffer after washing compound 528 off with non-ionic buffer.

AmberChrom® 50WX4 resin was used in this ion exchange experiment. Semi-purified material collected after liquid-liquid extraction and flash chromatography was used in this trial. After binding the sample to the resin, it was washed with 7 aliquots of 4 mL methanol (2 times v:w to the resin), followed by 3 aliquots of 5 mL 1.2 M HCl in methanol for elution. However, neither thiopotensamide A nor compound 528 was found in any of the 7 wash fractions, and all 3 eluates had compound 528 detected but not thiopotensamide A.

The first question from the trial was whether compound 528 had low binding ability towards resins so that the non-ionic washes was not strong enough to elute compound 528. A stepwise increased of HCl concentration was added to the next trial as washes. The second mystery was why thiopotensamide A was undetectable in the fractions of the acidic eluent. The potential reasons could either be that thiopotensamide A had very strong

binding towards to resin so the HCl was not strong enough to elute the compound or the compound had been degraded.

A second attempt of cation exchange purification was also performed with increased volume of wash buffers to ensure the resins were sufficiently washed and conditioned. The final fractions were concentrated to 500 µL (4 times more concentrated than the original material) for LCMS analysis. Frustratingly, compound 528 was detected in all collected fractions from low acidic to high acidic eluents, yet thiopotensamide A was not detected in any of the fractions (Table 2.2). Hence, purification using ion exchange resin was considered to be unsuccessful at this stage.

Table 2.2: Summary of ion exchange purification attempt of thiopotensamide A. After applying to the resin, the sample was allowed to flow through and then washed with increasing concentrations of HCl in methanol. HPLC samples were prepared by pooling all fractions from each elution step and concentrated to an equal volume (500 µL). *the abundance of compound 528 was much lower (at least 50 times lower).

Steps	Wash Volume (mL)	Detection of compound528 (*m/z* 528.3540)	Detection of thiopotensamideA (*m/z* 1243.4200)
Flow through	80	✓*	✗
MeOH only	40	✓*	✗
10 mM HCl	40	✓	✗
30 mM HCl	20	✓	✗
60 mM HCl	20	✓	✗
120 mM HCl	20	✓	✗
300 mM HCl	20	✓	✗
600 mM HCl	20	✓	✗
1200 mM HCl	10	✓	✗

To check whether thiopotensamide A was potentially degraded during purification, semi-purified material was also treated with a series of increasing concentrations of HCl in

methanol, which was used as elution buffer in the ion exchange experiment. The material was tested in parallel with the ion exchange experiment (in room temperature for approximately 3 hours). The results (Figure 2.16) showed that thiopotensamide A was not stable in acidic environment, as the amount of detected thiopotensamide A declined when the concentration of acid increased, with the compound degrading almost completely at concentration of HCl over 120 mM. In contrast, compound 528 was still detectable even at the highest HCl concentration tested, 1.2M (Figure 2.16).

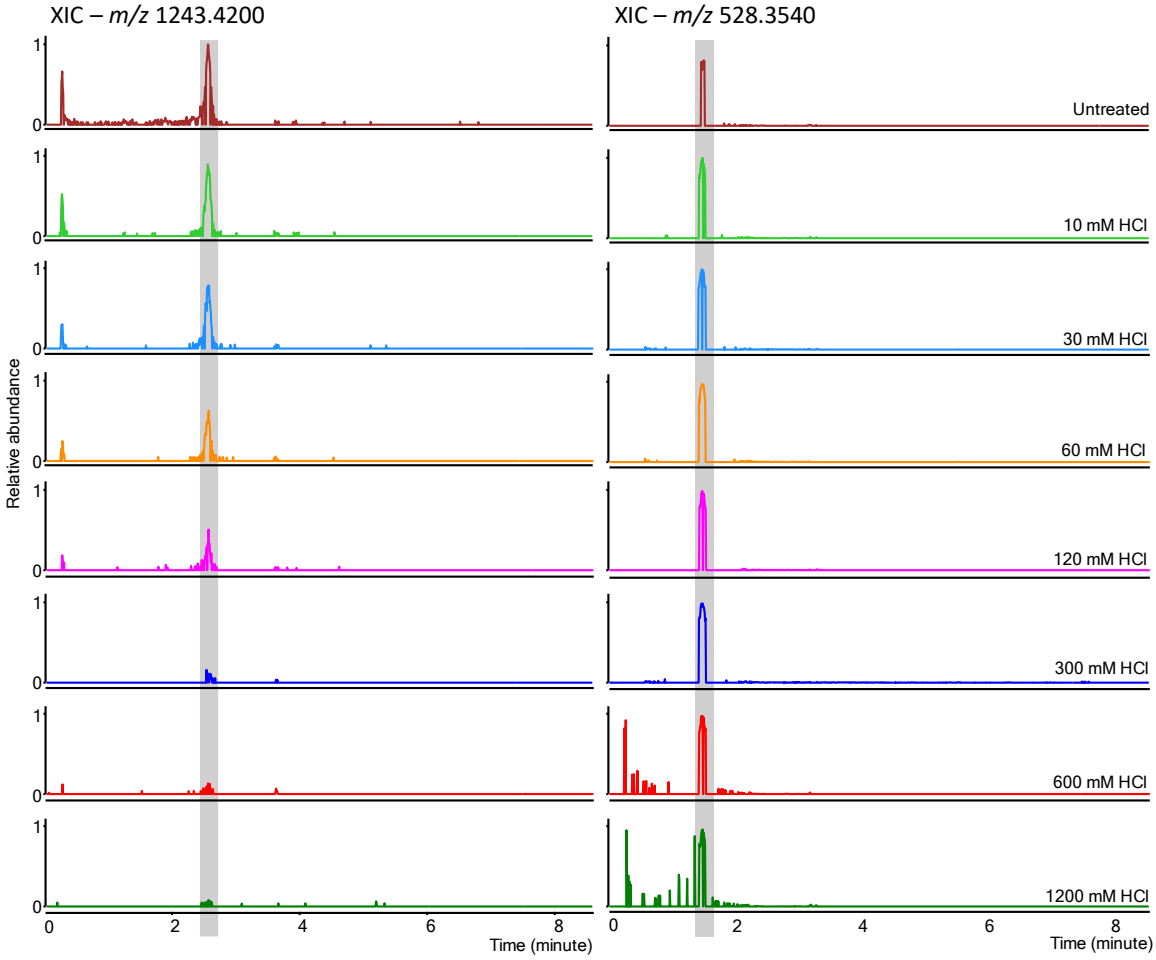


Figure 2.16: Stability tests for thiopotensamide A and compound 528 after treatment with increasing HCl concentrations. Extracted ion chromatograms of thiopotensamide A (*m/z* 1243.4200, left), and compound 528 (*m/z* 528.3540, right) were normalised with respect to the sample with maximum intensity in each case. Untreated sample was incubated with methanol only.

The results of the preliminary ion exchange experiment and the stability test indicated that it would be difficult to optimize the protocol to allow successful purification of thiopotensamide A. In summary, despite the use of several orthogonal separation techniques and multiple optimisation attempts, purification of thiopotensamide A proved extremely challenging and it was not possible to obtain material appropriate for NMR analysis from wild type *N. potens* cultures.

2.3.4.5. Identification of impurity with *m/z* 528

The lack of success separating thiopotensamide A and compound 528 with all the different techniques employed (flash chromatography, HPLC, SEC) triggered an interest in discovering the identity of compound 528. Identification of this compound could then help develop a targeted approach to remove compound 528 from the sample. The first step was assessing whether compound 528 was a metabolite from *N. potens* or a medium component and whether its production was medium dependent. Hence, MS data from *N. potens* cultures on different production media were reviewed. Compound 528 was not detected in any of the media controls (uninoculated media extracts) but was present in all *N. potens* production culture samples, regardless of whether thiopotensamide A was produced or not (Figure 2.17). This finding confirms that compound 528 is a metabolite produced by *N. potens* but unfortunately there was no single condition where selective production of thiopotensamide A was achieved.

Once it was established that compound 528 was a true metabolite of *N. potens*, the next step was determining whether it was a known molecule. The compound had an *m/z* of 528.35 likely corresponding to a protonated adduct [M+H$^+$], but the possibility of it being a

natively charged molecule [M^+] like thioamitides are could not be discarded, so the molecular mass of compound 528 could either be [M] = 527.35 Da or [M^+] = 528.35 Da.

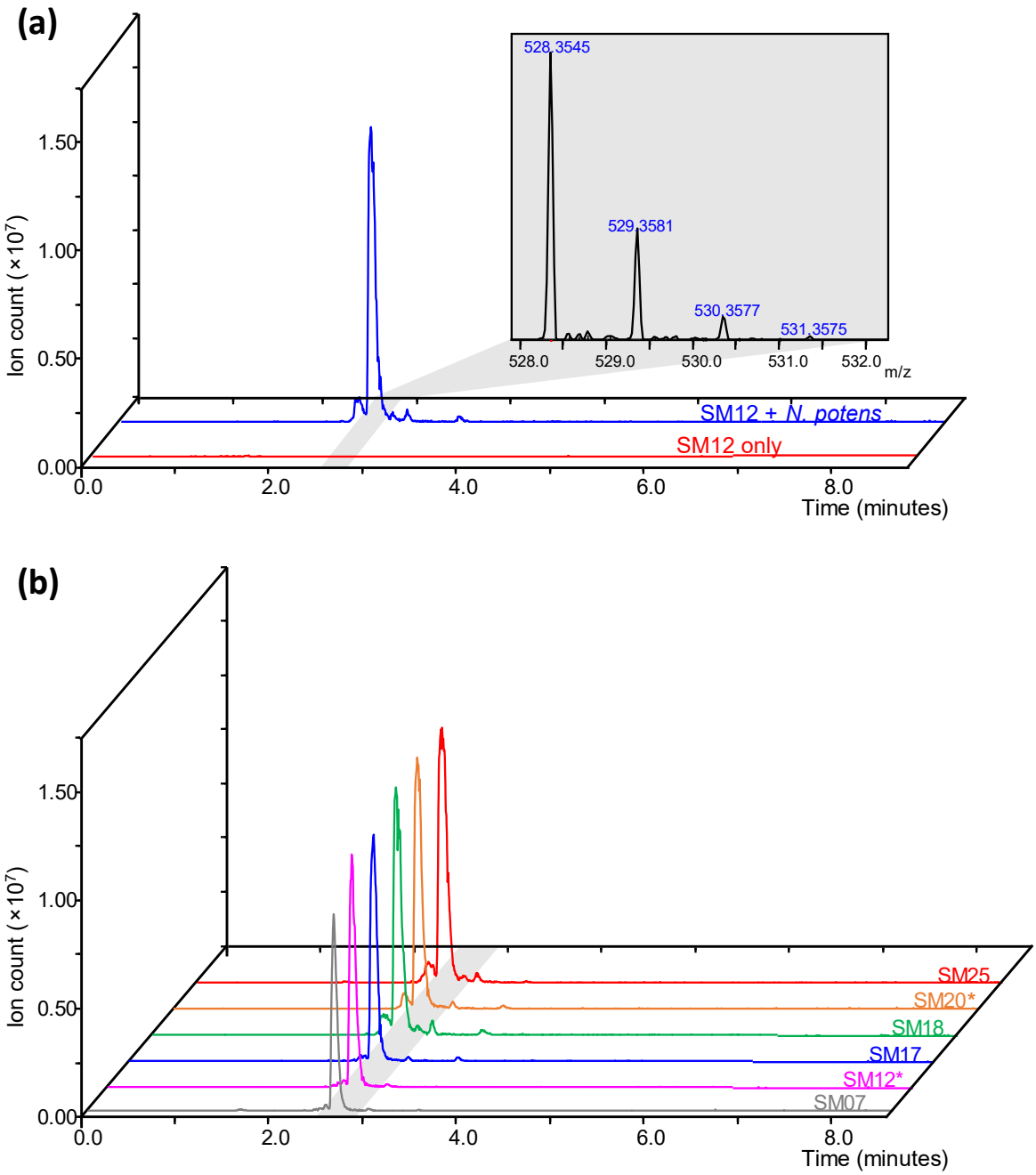


Figure 2.17: Screening of compound 528 (*m/z* 528.354) production in various conditions. (a) XICs for *m/z* 528.354 in uninoculated SM12 (red) and SM12 incoculated with *N. potens* (blue) (b) XICs for *m/z* 528.354 in culture samples of *N. potens* grown in various production media. *indicates the medium used for production of thiopotensamide A.

A mass-based search was carried out on the Dictionary of Natural Products, using a mass range between 527.3 and 527.4 (CHEMnetBase, 2022). This search returned 15 hits (Table 2.3), three of which were isomers and shared the same molecular mass with compound 528, with mass errors below 2 ppm. All 3 compounds (kayamycin, picromycin and proactinomycin) are macrolides, and more interestingly, kayamycin was reported being produced by *Nocardiopsis* (Chain, 1942, Rengaraju, 1985, Betlach et al., 1998). Digging further in the literature since information about kayamycin was extremely limited, an analogue of kayamycin had also been reported as a product of the same *N. potens* strain used in this work. This molecule, potensimicin, was discovered through an automated platform for the discovery of glycosylated molecules based on genomic data (Johnston et al., 2015). Interestingly, reviewing the genome after this new information came to light, it turned out that the potensimicin BGC is located immediately upstream of the thiopotensamide cluster. Indeed, potensimicin was found in the production culture (*m/z* 526.34) too, but it did not draw the first attention as it appeared to be the precursor of kayamycin and had different retention time in HPLC (Appendix 5). The structural difference between potensimicin and kayamycin is the unsaturated hydrocarbons in the narbonolide, so the conversion from potensimicin to kayamycin only requires a reduction (Figure 2.18). To definitively verify the identity of compound 528, an NMR analysis was performed, confirming this molecule shares the skeletal formula of kayamycin, and NOESY NMR spectroscopy of the molecule demonstrates they shares the relative stereochemistry (Appendix 6-16).

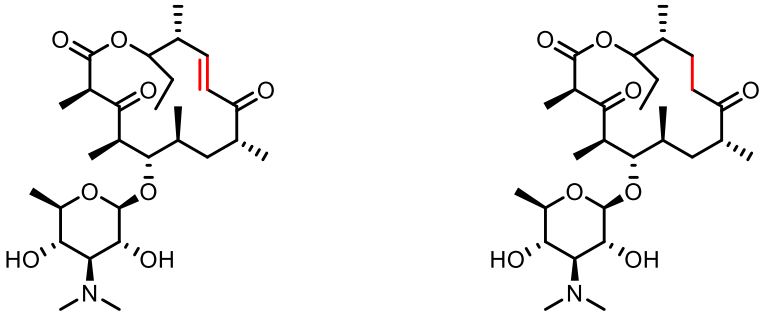


Figure 2.18: Structure of potensimicin and kayamycin

Table 2.3: Chemical search results from the Dictionary of Natural products for the mass range between 527.3 and 527.4 Da. Mass errors in ppm between the theoretic proton adduct mass ($[M+H]^+$) of each hit and the observed mass for compound 528 (528.3543 Da) are shown to the right. Molecules with masses within 5 ppm of compound 528 are highlighted.

Chemical Name	Molecular Formula	Theoretic mass [M] (Da)	Theoretic mass $[M+H]^+$ (Da)	Mass Error (ppm)
DaphnezomineD	$C_{32}H_{49}NO_5$	527.3611	528.3689	-27.63
Daphniphylline	$C_{32}H_{49}NO_5$	527.3611	528.3689	-27.63
DindygulerinoneA; N-[2-(3,4-Dihydroxyphenyl)ethyl]	$C_{32}H_{49}NO_5$	527.3611	528.3689	-27.63
Embellicine A	$C_{34}H_{41}NO_4$	527.3036	528.3114	81.20
Kayamycin	$C_{28}H_{49}NO_8$	527.3458	528.3536	1.32
MDN 0104	$C_{29}H_{53}NO_7$	527.3822	528.3900	-67.56
MelantherasideC; 3-Ac	$C_{32}H_{49}NO_5$	527.3611	528.3689	-27.63
PF 1171D	$C_{28}H_{41}N_5O_5$	527.3108	528.3185	67.76
PF 1171E	$C_{28}H_{41}N_5O_5$	527.3108	528.3185	67.76
PhomapyrrolidoneA	$C_{34}H_{41}NO_4$	527.3036	528.3114	81.20
PhomapyrrolidoneA; 15-Epimer	$C_{34}H_{41}NO_4$	527.3036	528.3114	81.20
Picromycin; 10,11-Dihydro	$C_{28}H_{49}NO_8$	527.3458	528.3536	1.32
Proactinomycin; Proactinomycin B	$C_{28}H_{49}NO_8$	527.3458	528.3536	1.32
Pyrrocidine C	$C_{34}H_{41}NO_4$	527.3036	528.3114	81.20
Secodaphniphylline, 22S-Acetoxy	$C_{32}H_{49}NO_5$	527.3611	528.3689	-27.63

2.3.5. Heterologous expression of the thiopotensamide BGC

To advance the work on the thiopotensamide pathway, genetic manipulation of *N. potens* was attempted. The idea was to probe the essential genes for hdmHis dehydration, as well as to disrupt the biosynthesis of kayamycin to facilitate purification of thiopotensamide A. The genetic tractability of *N. potens* was tested via attempts to introduce several different plasmids via conjugation: replicative expression vector pIJ86, integrative expression vector pIB139 and pCRISPOmyces, a pSG5 based plasmid used for gene editing in actinomycetes (Kieser et al., 2000, Wilkinson et al., 2002, Cobb et al., 2015). All three vectors had apramycin resistance selection markers. First of all, *N. potens* was grown on solid soy flour medium (SFM) and SFM with apramycin. *N. potens* showed no resistance to apramycin as it grew on the non-selective plate but not when apramycin was added. After testing the selectable marker, pIJ86 was attempted to be introduced through conjugation via *E. coli* ET12567 with pUZ8002 as a helper plasmid. Apramycin resistant *N. potens* exconjugant were taken to obtain patches on a fresh SFM plate prepared with apramycin. However, the results were inconsistent. The first conjugation showed apramycin resistant colonies, but the patches from the exconjugants did not grow on the SFM plates prepared with apramycin. When the experiment was repeated, there was no signs of growth even in an extended period (to 14 days) indicating the plasmid might potentially be lost from the host.

After conjugation, electroporation of mycelium was employed to individually introduce all 3 different types of plasmids. The integrative expression vector pIB139 requires ϕC31 integration, but no similar sequences were found when BLASTN against *N. potens* genome even though the programme selection was optimised for “somewhat similar sequences”. However, the genome sequence on NCBI database was incomplete at the time, so the

possibility that ϕC31 was in the chromosome could not be ruled out. The transformants from all three types of plasmids showed apramycin resistance again, but the isolated colonies did not continue to grow after transferring to a fresh plate with selection antibiotic. Hence, the experiment was considered unsuccessful as none of the plasmids seemed to be maintained in *N. potens*.

2.3.5.1. PCR-based TAR cloning

Given the challenges in the purification of thiopotensamide A and the genetic manipulation of *N. potens*, it became essential to clone the thiopotensamide BGC in a suitable host for heterologous expression. Following the success cloning the BGC for thiostreptamide S4 via transformation-associated recombination (TAR) using *Saccharomyces cerevisiae* VL6-48N and expressing it in *S. coelicolor*, a similar method from Zhang *et al* was applied to the *tpa* BGC (Zhang et al., 2019, Eyles et al., 2021). It was found by Lee *et al* that the efficiency of the original TAR cloning method could be improved by treating the genomic DNA (gDNA) with CRISPR/Cas9 to isolate the region of interest, increasing the yield of region-positive clones up to 35-fold (Lee et al., 2015). This could be due to increased penetration of the smaller DNA fragments into the yeast cells as well as increased recombination rates when the regions involved in the capture are at the ends of the fragments. Another way to increase efficiency is using purified DNA, to ensure high quality and quantity on the genetic material used in the transformation. Hence, a PCR-based TAR cloning method was applied to provide suitable size and high quality of DNA for improved efficiency TAR cloning (Figure 2.19) (Eyles et al., 2018). One advantage of this approach is that it allows the assembly of long DNA sequences split in multiple fragments small enough to be amplified by PCR. For example, 25 DNA fragments were successfully assembled in yeast to generate a complete

synthetic *Mycoplasma genitalium* genome, indicating that yeast is capable to assemble numerous DNA fragments (Gibson et al., 2008). Using PCR-based TAR cloning also allows to simultaneously refactor the biosynthetic pathway to potentially improve production in the heterologous expression host. This can be achieved by substituting the native regulatory sequences in the cluster for strong promoters and ribosome binding sites validated to work in the destination host. These sequences can be PCR amplified separately and designed to carry sequences for homologous recombination and insertion in the correct locations in the cluster (Eyles et al., 2018).

The trifunctional vector pCAP03 was employed for this TAR cloning work (Zhang et al., 2019). This vector is suitable for yeast and *E. coli* propagation and selection, as well as for chromosomal integration in the ϕC31 site of *Streptomyces.* Selection in yeast, where pCAP03 is maintained as single copy, is achieved using a tryptophan auxotrophy marker. The plasmid also possesses a pUC origin of replication for multicopy plasmid propagation in *E. coli* to obtain better plasmid yields. Kanamycin is the selection marker for pCAP03 both in *E. coli* and *Streptomyces*.

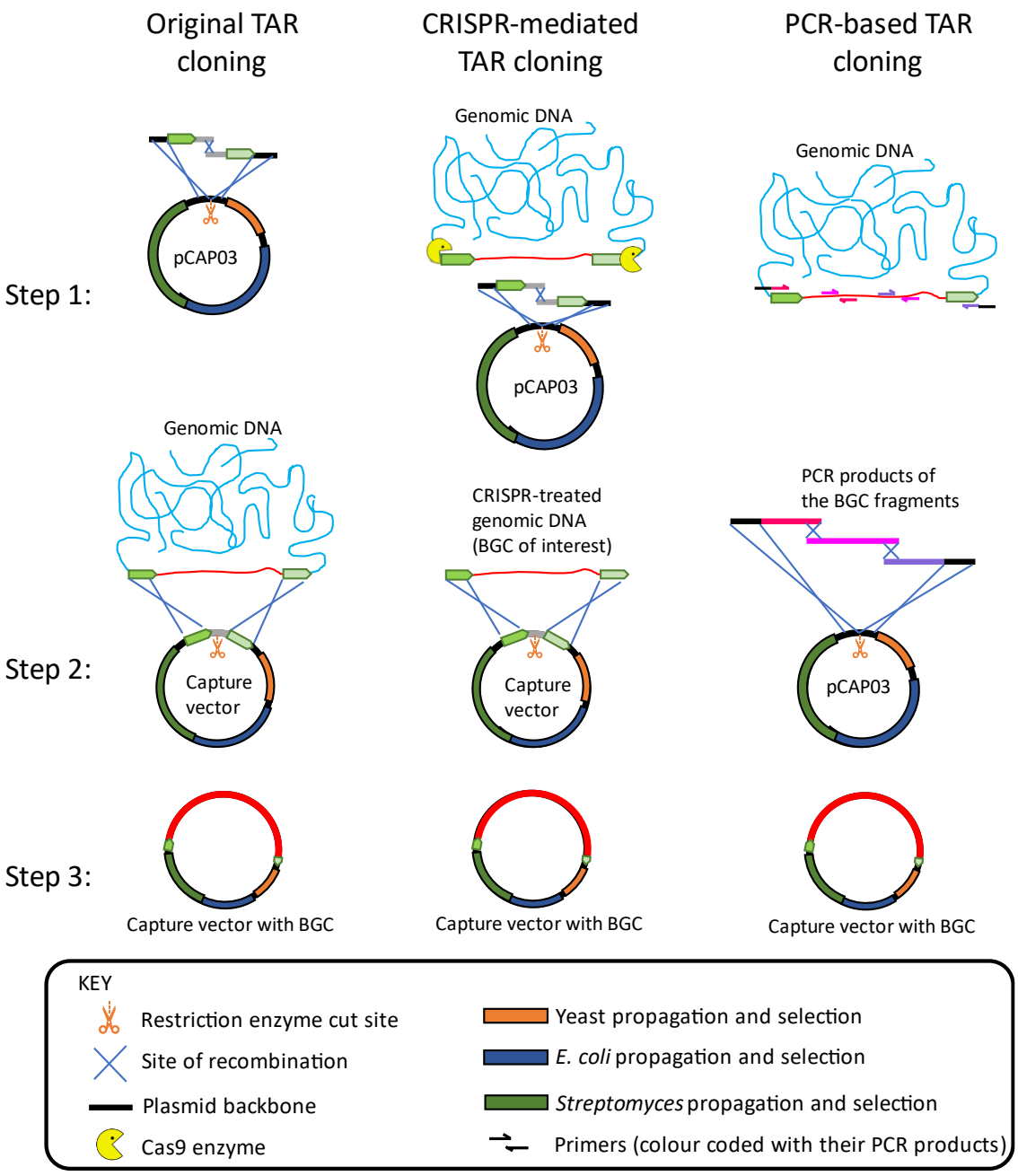


Figure 2.19: Schematic representation of TAR cloning techniques discussed. In step 1, original and CRISPR-mediated TAR cloning require construction of a capture vector containing DNA hooks that are specific to the boundaries of the genes of interest. CRISPR-mediated method also requires construction of specific Cas9-guide RNA complex for digestion at the boundaries of the region to clone followed by gDNA clean up. On the other hand, the PCR-based method only requires PCR amplification of products with homologous regions for recombination between them as well as with the flanking regions of the plasmid where they will be cloned, without need for a specific capture vector. Step 2 is to transform the prepared materials (linearised plasmid and insert DNA: gDNA for original method, CRISPR-treated gDNA for CRISPR mediated method and PCR products for PCR-based method). The final step in all three methods is the selection of transformant colonies in yeast and PCR screening for the correct assembly.

Two constructs were made for the experiment (Figure 2.21). In one of the constructs, the cluster was amplified and assembled with its native organisation, from the beginning of the intergenic region before *tpaA* (to capture the potential *tpaA* promoter) up to *tpaD2*. In the second version, the BGC was refactored to optimise its expression in *S. coelicolor* M1146 (Figure 2.21). First, it was important to use a RiPP strong promoter known to work in M1146, as it has been reported that RiPP precursor peptides require very high expression levels for compound production since the precursor peptide itself is the scaffold of the molecule. It had been reported that the native promoter for the RiPP thiovarsolin precursor peptide was found to successfully drive the expression of the thiovarsolin BGC in *S. coelicolor* M1146 and lead to good levels of production in the heterologous host (Santos-Aberturas et al., 2019). Therefore, the promoter region of the thiovarsolin BGC was utilised to refactor the thiopotensamide BGC before *tpaA*. On the other hand, it is well known that RiPP BGCs often have attenuator sequences after their precursor peptide genes so that the downstream tailoring genes are transcribed at lower levels than the precursor peptide (Foulston and Bibb, 2010, O'Rourke et al., 2017, Vior et al., 2020). Analysis of the intergenic region after *tpaA* for the formation of secondary structures with the ViennaRNA Package 2.0, revealed a high probability hairpin the sequence. (Figure 2.20) (Lorenz et al., 2011). The potential formation of a RNA hairpin in the intergenic region before the tailoring enzyme gene cluster was likely to negatively impact the expression of the downstream the tailoring enzymes, plus any native promoters located after it might not be functional in M1146, so the constitutive promoter P_{hrdB}, which was previously proven to be suitable for refactoring in *S. coelicolor* was inserted before *tpaD1* to ensure the expression of tailoring enzymes (Eyles et al., 2018).

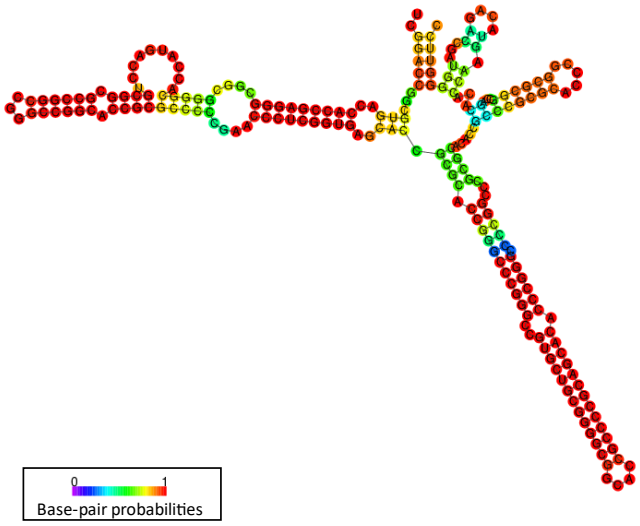


Figure 2.20: Results of the RNA fold analysis for minimum free energy prediction for the intergenic region between *tpaA* and *tpaD1*.

All the fragments for the TAR cloning were designed and amplified with minimum of 60 bp overlapping with either their neighbouring fragments or with the pCAP03 vector in the cases of fragments 1 (1a from the refactored cluster) and 4, to generate homologous regions for recombination in yeast. The overlapping regions are either introduced by the overhangs of primers (fragments 1, 1a, 1b, 1c, 1d and 4) or extended PCR region (fragments 2, 3 and 4). For the wildtype cluster assembly, the native BGC was amplified in 4 fragments from *N. potens* gDNA and assembled into pCAP03 to generate construct pLF015 (Figure 2.21). For the refactored cluster, the thiovarsolin promoter region (fragment 1a), and the P_{hrdB} promoter (fragment 1c) were amplified from *Streptomyces varsoviensis* gDNA and *S. coelicolor* gDNA respectively, with overhangs to be inserted either side of the *tpaA* gene (fragment 1b). A fourth fragment (1d) was designed to contain the rest of the genes originally contained in the wildtype fragment 1 so fragments 2, 3 and 4 could be reused in the assembly of the refactored cluster. In this case all 7 fragments were assembled into pCAP03 to generate the refactored construct pLF016 (Figure 2.21)

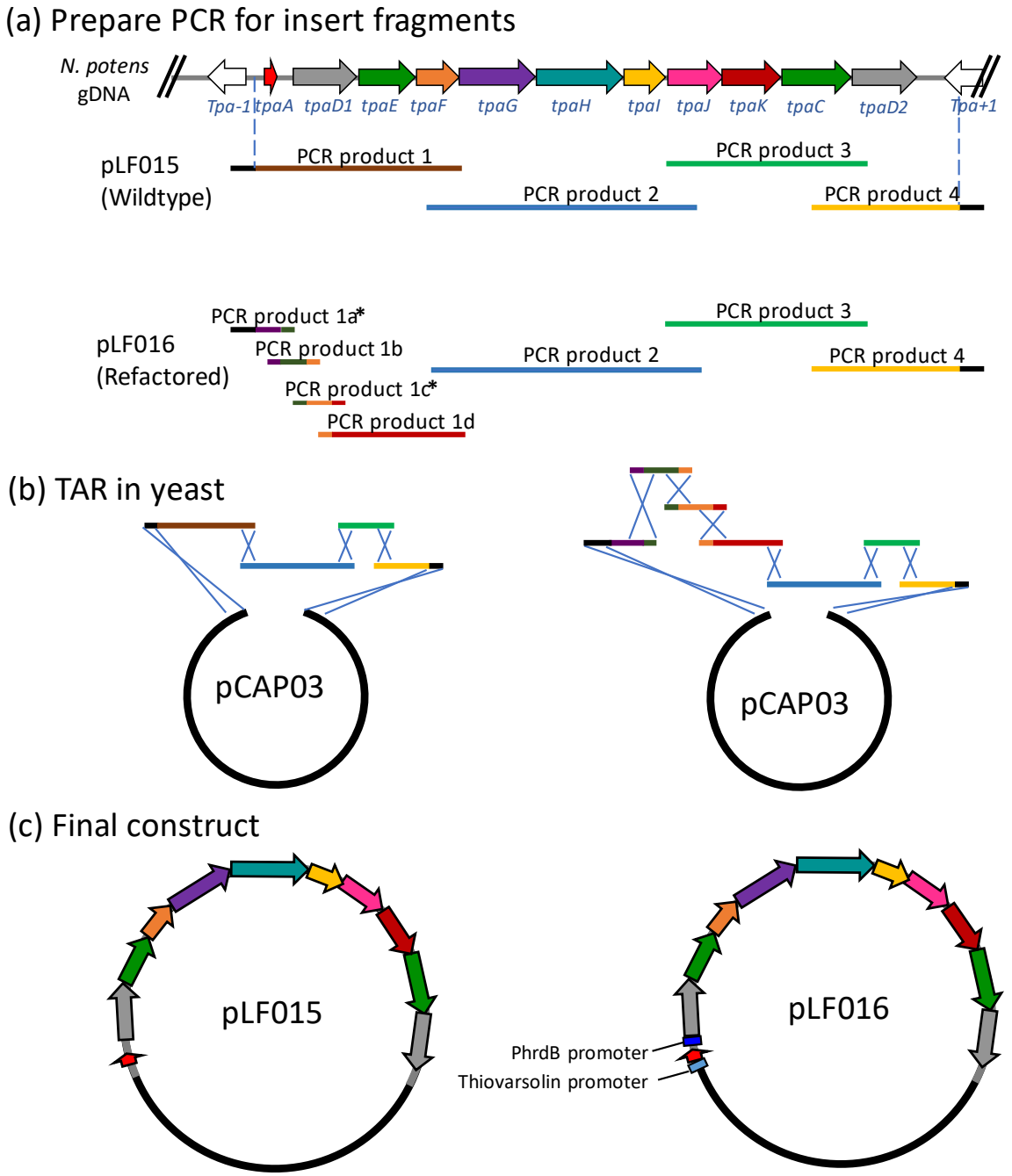


Figure 2.21: PCR-based TAR cloning scheme for the main stages of construction for pLF015 (wild type BGC) and pLF016 (refactored BGC). (a) Design and PCR amplification of assembly fragments. (b) PCR products and linearised pCAP03 were transformed into yeast for recombination and assembly. (c) The final constructs: pLF015 containing the wild type thiopotensamide BGC and pLF016 containing a refactored thiopotensamide BGC where intergenic region before *tpaA* was replaced with the promoter region from thioavarsolin and PhrdB promoter was inserted between *tpaD1* and the intergenic region after *tpaA*. DNA fragments are colour coded to represent the overlapping sequences required for recombination. Refactoring fragments amplified from *S. varsoviensis* and *S. coelicolor* are highlighted with an asterisk.

pCAP03 derived from pCAP01 contains the strong promoter P_{ADH1} to express the *URA3* gene to select against recircularization of the plasmid, so a low false positive rate was expected (Zhang et al., 2019). Four pools of 10 colonies were tested from each construct using PCR to check for the presence of overlapping region between PCR products 3 and 4 (553 bp) (Figure 2.21) that was used for recombination. In total, 80 colonies were checked between both constructs and all pools showed at least one positive colony per pool (Figure 2.22a). All 10 colonies from pools 2 (for pLF015) and 1 (for pLF016) were checked individually with the same PCR screen. Although the DNA gel bands were very weak when the yeast colonies were individually tested, five colonies tested positive in each pool 2 and pool 1 (Figure 2.22b). Four positive colonies from each pool were selected for plasmid transfer to *E. coli* and purification. The plasmids were checked by PCR using primers amplifying a region covering from the pCAP03 backbone to downstream of the start codon of *tpaA* gene. The products were expected to be 720 bp for pLF015 and 1060 bp for pLF016, as the promoter region from *S. varsoviensis* is approximately 350 bp bigger than the native region upstream of *tpaA*. The PCR results confirmed that the PCR-based TAR cloning was successful, with two positive clones for each construct (Figure 2.22d).

After being confirmed with whole plasmid sequencing, pLF015 and pLF016 were conjugated into *S. coelicolor* M1146 via methylation-deficient *E. coli* ET12567 carrying helper plasmid pR9604. *S. coelicolor* M1146 was employed as it had been validated as a host for enhanced level of secondary metabolite production and reduced level of background metabolite production (Gomez-Escribano and Bibb, 2011). This resulted in the generation of strains M1146/pLF015 with the native cluster and M1146/pLF016 with the refactored one. Several clones of each were tested for thiopotensamide production on

SM20 agar as it was one of the production media for thiopotensamide A in the native strain *N. potens*. Production was evaluated by HPLC-HR-MS/MS.

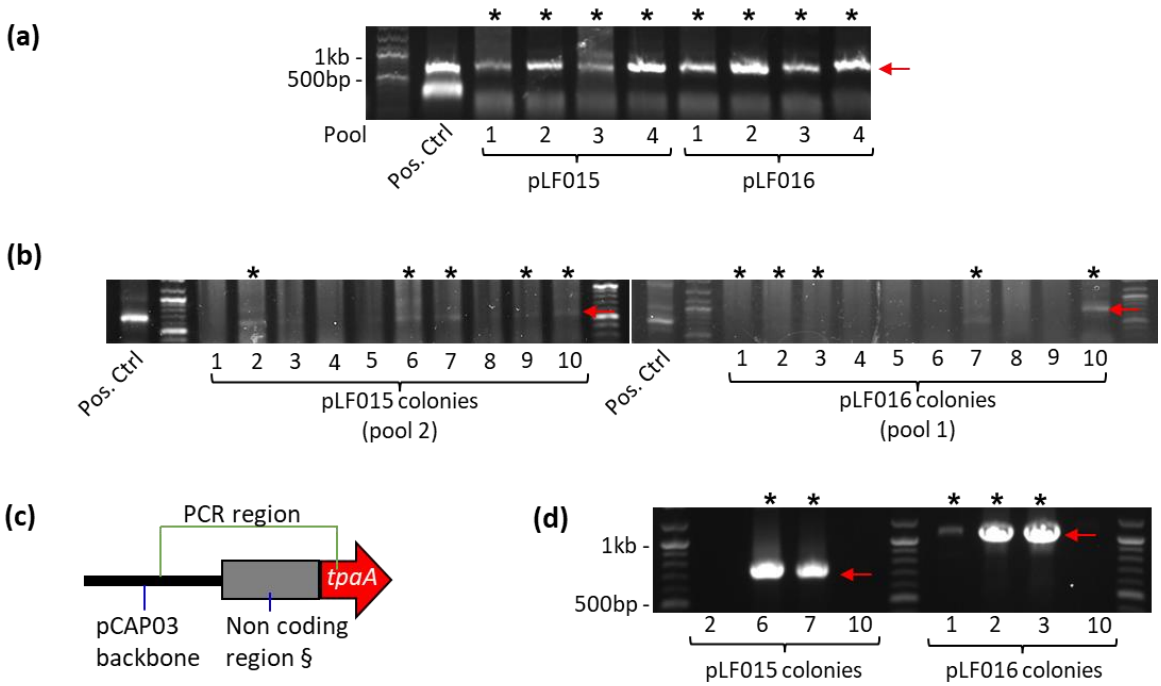


Figure 2.22: PCR screening results for TAR cloning. Red arrows indicate the location of positive band in the lane and asterisks mark the lanes containing a positive band. (a) Colony PCR screening on pools of 10 colonies from TAR cloning. Positive control (Pos. Ctrl) used gDNA from N. potens as template for amplification. (b) Individual PCR screening for yeast colonies from pool 2 (pLF015) and pool 1 (pLF016). Positive control (Pos. Ctrl) was a repeat of the corresponding pooled PCR. (c) PCR region using the same primer pairs to obtain distinct DNA bands. § Non coding region in pLF015 (native intergenic region upstream *tpaA* from *N. potens*) is approx. 350 bp smaller than in pLF016 (thiovarsolin promoter region from *S. varsoviensis*). (d) PCR using purified plasmid from 4 selected positive colonies from each construct.

2.3.5.2. Thiopotensamide B, an unexpected BGC product

Interestingly, neither of the constructs in *S. coelicolor* M1146 produced thiopotensamide A. Instead, an analogue of thiopotensamide A (*m/z* 1231.4176) was detected from the heterologous expression of the cloned BGC (Figure 2.23). This novel compound was assigned as thiopotensamide B. This compound was clearly related to the cloned BGC as it

was not detected in the cultures of *S. coelicolor* M1146 with the empty vector. This compound has a *m/z* 1231.4176 on a MS spectrum and possesses consecutive neutral loss of 33.99 on a MS/MS spectrum. This mass was very similar to the mass from predicted structure I (exact mass: 1231.4371, Figure 2.8), but the mass error (-15.835 ppm) between the observed mass and the exact mass from predicted structure I was significant, which suggested predicted structure I was not the correct stucture.

It became clear that predicted structure II was very close to thiopotensamide B. The MS/MS spectrum (Figure 2.24) showed that the methionine was not oxidised as the neutral loss of 63.99 from the molecular ion was absent. The 16 mass units of the oxidised methionine on the theoretical structure II could be removed to generate a structure with exact mass of 1231.4194 (mass error: -1.462 ppm). Except the oxidation on the methionine, another difference between thiopotensamide A and B was a hydroxyl group at the hdmHis residue according to the MS/MS spectrum. The hydroxyl group in thiopotensamide A was hypothesised to be removed after installation, and the removal of hydroxyl group led to the absence of MS/MS fragment *m/z* 125.0709 (Figure 2.9; Section 2.3.1). Whereas the hydroxyl group was seemingly preserved in thiopotensamide B, as the MS/MS fragment was present in the spectrum. This suggested that the histidine residue has a difference of hydroxyl group between thiopotensamide A and B.

Combining the fragmentation information from MS/MS spectra of thiopotensamide A and B, it is supportive of the hypothesis that the absence of the hydroxyl group is due to elimination mediated dehydration. In the middle range of the MS/MS spectra, thiopotensamide A and B showed a very similar neutral loss pattern (Figure 2.25). Here, the MS/MS fragments of thiopotensamide B had a *m/z* +18 shift in comparison to

thiopotensamide A. This shift of MS/MS pattern suggested that both compounds were related, and one had an extra modification than the other. If thiopotensamide A never had the hydroxyl group installed, the shift would have been a +16 in thiopotensamide B instead of +18, as the removal of hydroxyl group lead to a formation of double bond between α and β position, which consequently led to a loss of 18 in molecular mass.

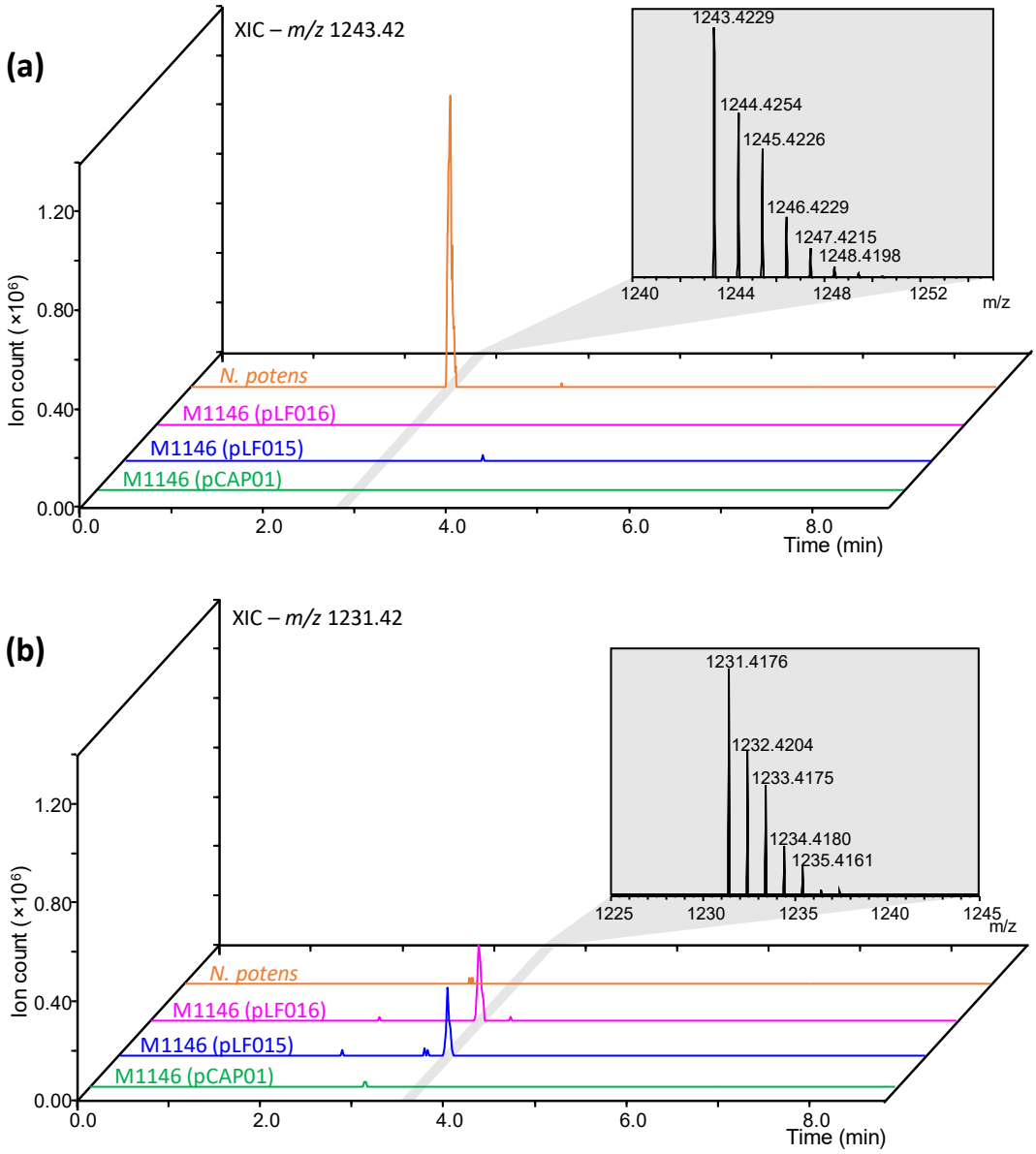


Figure 2.23: Extracted ion chromatograms (XICs) of thiopotensamide A and thiopotensamide B from different expression strains. (a) XICs of thiopotensamide A (*m/z* 1243.4200) showed only the native strain produced thiopotensamide A. (b) XIC of thihopotensamide B (*m/z* 1231.4200) showed both M1146/pLF015 and M1146/pLF016

produced thiopotensamide B. M1146/pCAP01 was employed as negative control and showed neither of the compounds were produced.

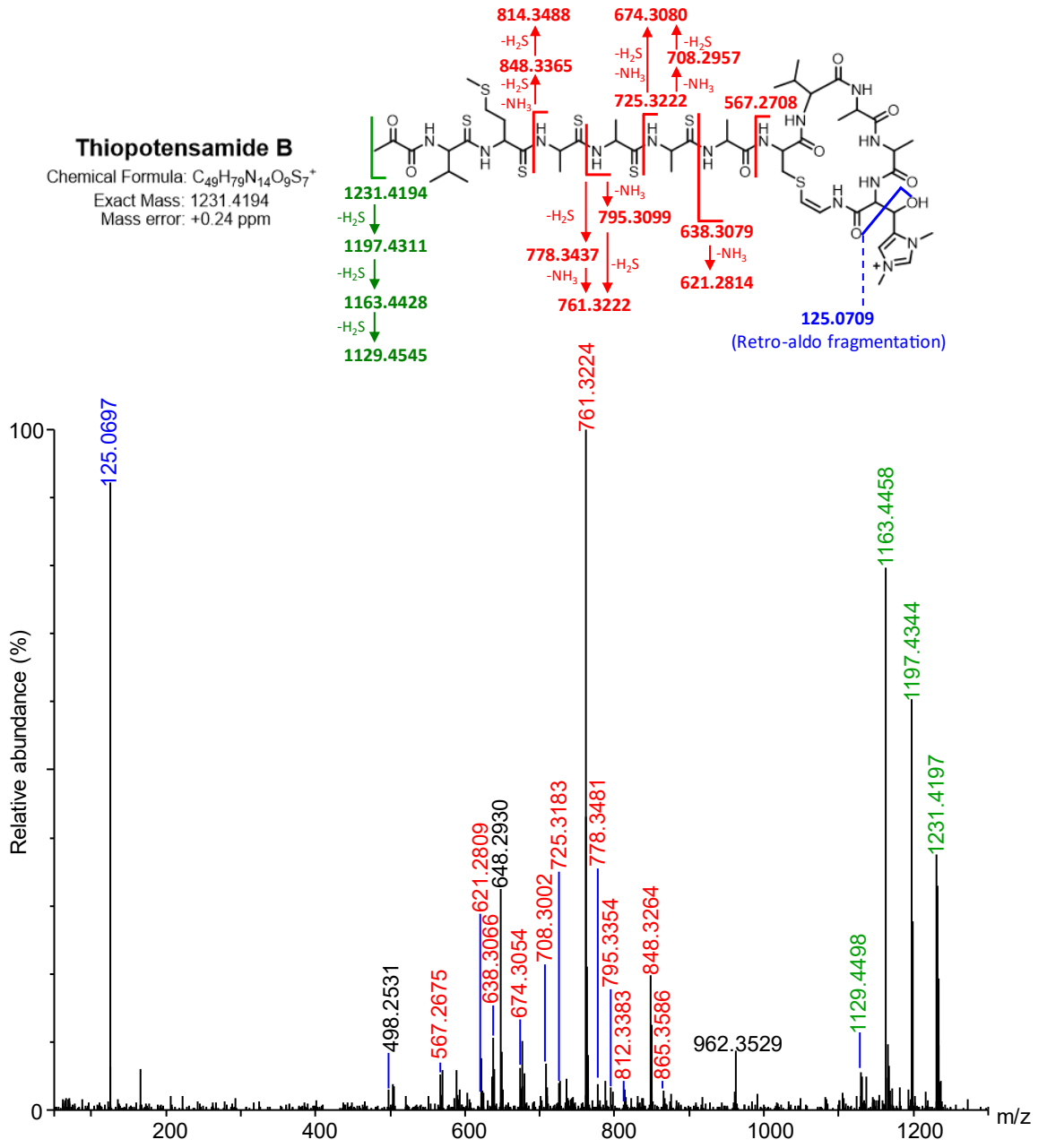


Figure 2.24: MS/MS spectrum of thiopotensamide B from Synapt HR-MS. Expected fragmentation and the calculated mass in bold. Peaks are labelled in colour to match the prediction masses. The spectrum showed 3 consecutive neutral losses of 33.99, which was given by thioamide bond fragmentation. The fragment ion at *m/z* 125.0697 is consistent fragmentation on hdmHis that was not seen in the MS/MS spectrum of thiopotensamide A.

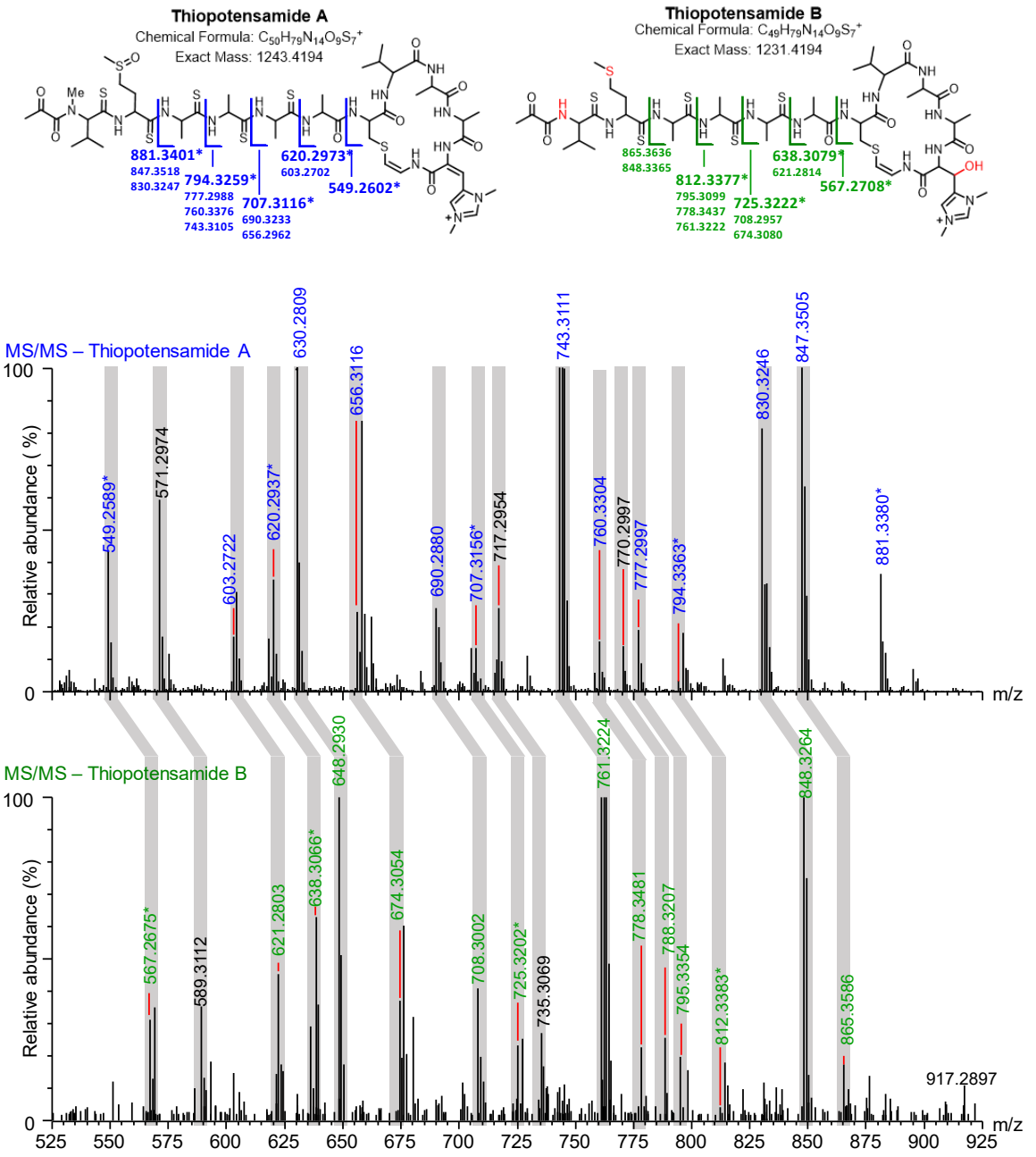


Figure 2.25: Middle range (*m/z* 525 – 925) of MS/MS spectra from thiopotensamide A and thiopotensamide B and the expected fragments (mass in bold). Peaks are labelled in colour to match the prediction masses and peaks related to identified product ions were highlighted in grey.

Although the heterologous expression host did not produce thiopotensamide A, the production of thiopotensamide B provided valuable information to support the hypothesis that the hdmHis residue underwent an elimination to form thiopotensamide A. Besides the

dehydration, the methylation that was hypothesised to be present on Pyr1, Val2 or MetO3 was not installed in thiopotensamide B, based on MS data (Figure 2.24).

It was intriguing that neither of the unconventional modifications were instated in thiopotensamide B, suggesting that at least one or more responsible genes for one of or both modifications were inactive or from outside the cloned BGC. There was also a possibility that one of these modifications could only be installed after the other, but the chance was relatively low considering those modifications were on two different ends of the molecule.

It was also interesting to see that the methionine was not oxidised in the heterologously expressed product. The oxidation on the methionine was thought to be introduced by ROS, which means that the oxidation should be non-enzymatic, so the reasons for the absence of the oxidation were unclear. One of the factors affecting the oxidations on a methionine is the accessibility of the residue. If a methionine is embedded inside a protein, the oxidation reaction is less likely to take place in comparison to a surface residue (Kulczyk and Leustek, 2022, Walker et al., 2022). Although the susceptibility of the methionine residue could affect oxidative reaction, the structural difference between thiopotensamide A and B is rather trivial to change the secondary structure of the peptide. This methionine oxidation could relate to expression host. First of all, if *N. potens* possessed a higher level of ROS than *S. coelicolor*, the methionine would have a higher chance to be oxidised. Alternatively, methionine sulphoxide (MetO) could be more rapidly reduced back to methionine in *S. coelicolor* than *N. potens*. Since oxidation on methionine can lead to loss of protein function, methionine sulphoxide reductases (MSRA for *S*-MetO and MSRB for *R*-MetO) exist in bacteria and higher organisms to restore normal protein function (Brot and

Weissbach, 2000). Although at least one pair of *MSRA* and *MSRB* genes can be found in most of the bacteria, many bacteria have multiple copies of the genes (Ezraty et al., 2005, Sasindran et al., 2007). Hence, it is possible that *S. coelicolor* is more likely to maintain non-oxidised thiopotensamide. It was possible that the oxidation was not solely introduced by ROS, it could also be introduced by enzymatic activities. It is possible then that the oxidation could also be introduced by enzymes encoded outside the cloned BGC. Although it is rare in bacteria, members of the MICAL (microtubule associated monooxygenase, calponin and LIM) family proteins, monooxygenase domain containing proteins from eukaryotes, can produce ROS to oxidise certain methionine residues (Ortegón Salas et al., 2020).

2.3.5.3. Refactored BGC improved thiopotensamide B production

Despite both *S. coelicolor* strains M1146/pLF015 and M1146/pLF016 producing only intermediate versions of thiopotensamide, the question of whether refactoring of the cluster lead to increased production still needed an answer. Solid SM20 cultures were fermented at 30 ^{o}C for 15 days. Two biological replicates and five analytical replicates were taken from each strain and cultures were checked at 5 time-points: 5, 8, 10, 13 and 15 days. The production curves show production of thiopotensamide B peaks between 8 and 10 days, and the production is significantly increased in the refactored BGC, pLF016 (Figure 2.26). Since both strains gave the same product, and the refactored BGC containing strain *S. coelicolor* M1146/pLF016 was found to be more productive, this strain was selected to move forward for production and genetic manipulation (Section 2.3.3).

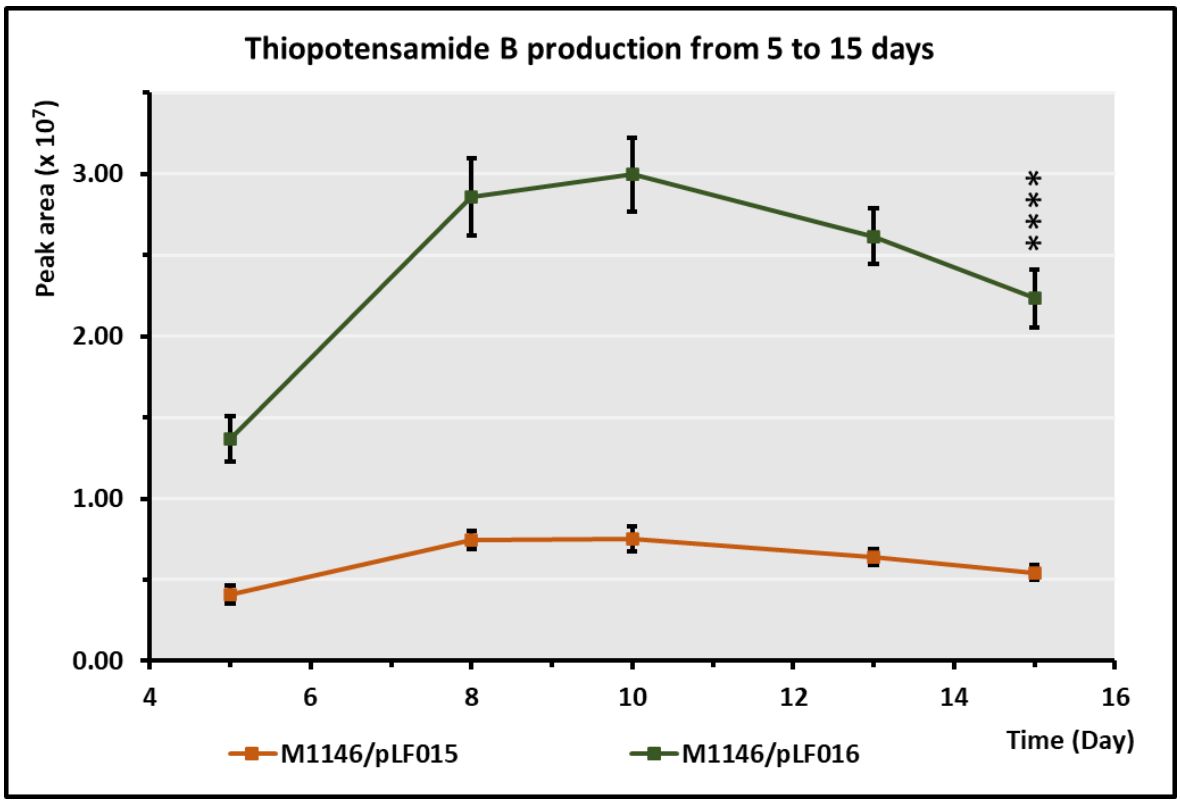


Figure 2.26: Production curves of thiopotensamide B over a 15-day period by the wildtype *tpa* BGC in *S. coelicolor* M1146/pLF015 and the refactored BGC in *S. coelicolor* M1146/pLF016. The asterisk indicates the production of thiopotensamide B is statistically significant at the end of the production period (p-value = 3.89×10^{-8}), and the data were collected from 2 biological replicates and 5 analytical replicates and error bars in both figures correspond to standard mean error between replicas.

2.3.6. Genetic manipulation of thiopotensamide BGC containing *S. coelicolor*

2.3.6.1. Potential genetic basis for the extra dehydration

Given the unique organisation of the phosphotransferase and HopA1 protein encoding genes in the *N. potens* BGC and their proposed role in the unique dehydration on hdmHis12, these proteins were compared with their homologues from other thioamitide BGCs. The proteins responsible for hdmHis dehydration were anticipated to be more dissimilar to the ones involved in serine/threonine dehydration, given their proposed substrate, hdmHis12 residue, is more sterically hindered than serine or threonine and is located in different position in the core peptide sequence. In particular, if the protein was specific to hdmHis12

only, the binding pocket would be expected to be bigger than the homologues that only make modifications on serine and threonine. It could also be the case that either or both the phosphotransferase-HopA1 pairs have a more flexible binding site to enable the dehydration on the hdmHis12 residue. In either case, the conserved binding site of the protein was anticipated to have some degree of modification.

First of all, a phylogenetic analysis of the homologues for both phosphotransferases in the thioamitide clusters was performed as the dehydration was initiated by phosphorlyation (Figure 2.27). Although both TpaC and TpaE are phosphotransferases, the *in vitro* reconstitution of AviMeCys bond formation from thiostreptamide S87, showed that TsdC (the homologue of TpaC) initiated the dehydration through phosphorylation and TsdE (the homologue of TpaE) was involved in the AviMeCys formation (Qiu et al., 2021). The TpaC homologues and TpaE homologues clustered neatly, showing the C and E homologues were phylogenetically distinct from each other as anticipated. Both TpaC and TpaE clustered with their respective homologues from the biosynthetic pathways of multiple characterised thioamitides, even though these compounds had a hydroxyl group installed and no elimination. Curiously, proteins TvaC and TvaE from the thioviridamide pathway were found to be more distantly related to these proteins than TpaC and TpaE, respectively. In other words, the phylogenetic analysis of phosphotransfereases TpaC and TpaE was inconclusive as the proteins were not significantly different enough from their counterparts to justify a distinct enzymatic activity.

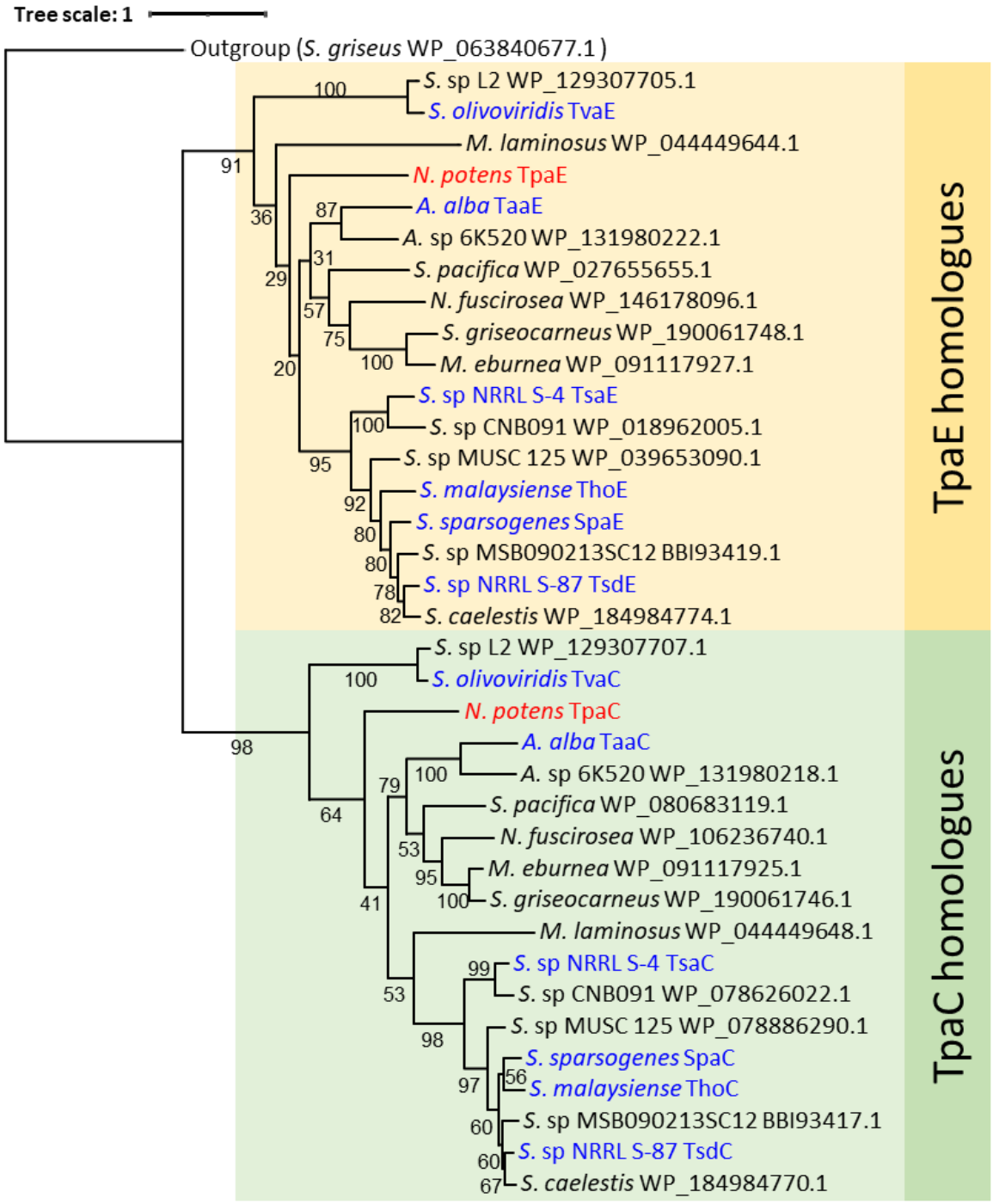


Figure 2.27: Maximum likelihood phylogenetic tree of the phosphotransferase (TpaC and TpaE) protein homologues from thioamitide BGCs. Aminoglycoside *O*-phosphotransferase from *Streptomyces griseus* was not from a thioamitide biosynthetic pathway and was used as the outgroup for the tree. Homologues of TpaC are highlighted in green, and homologues of TpaE are highlighted in yellow. Phosphotransferases from the biosynthetic pathways of characterised thioamitides are labelled in blue. Phosphotransferases from the thiopotensamide A BGC are labelled in red. Numbers at the nodes are the bootstrap values.

Following the phosphotransferases analysis, the HopA1 proteins were assessed next. As mentioned before, a unique feature of the *tpa* cluster is that there are two HopA1 domain containing genes *tpaD1* and *tpaD2*, versus the single copy (D) present in the rest of thioamitide BGCs. Hence, one of the two HopA1 proteins was assumed to be essential for the conventional modifications, namely the dehydrations on both serine residues, while the other one could be devoted to the unique dehydration observed in thiopotensamide A.

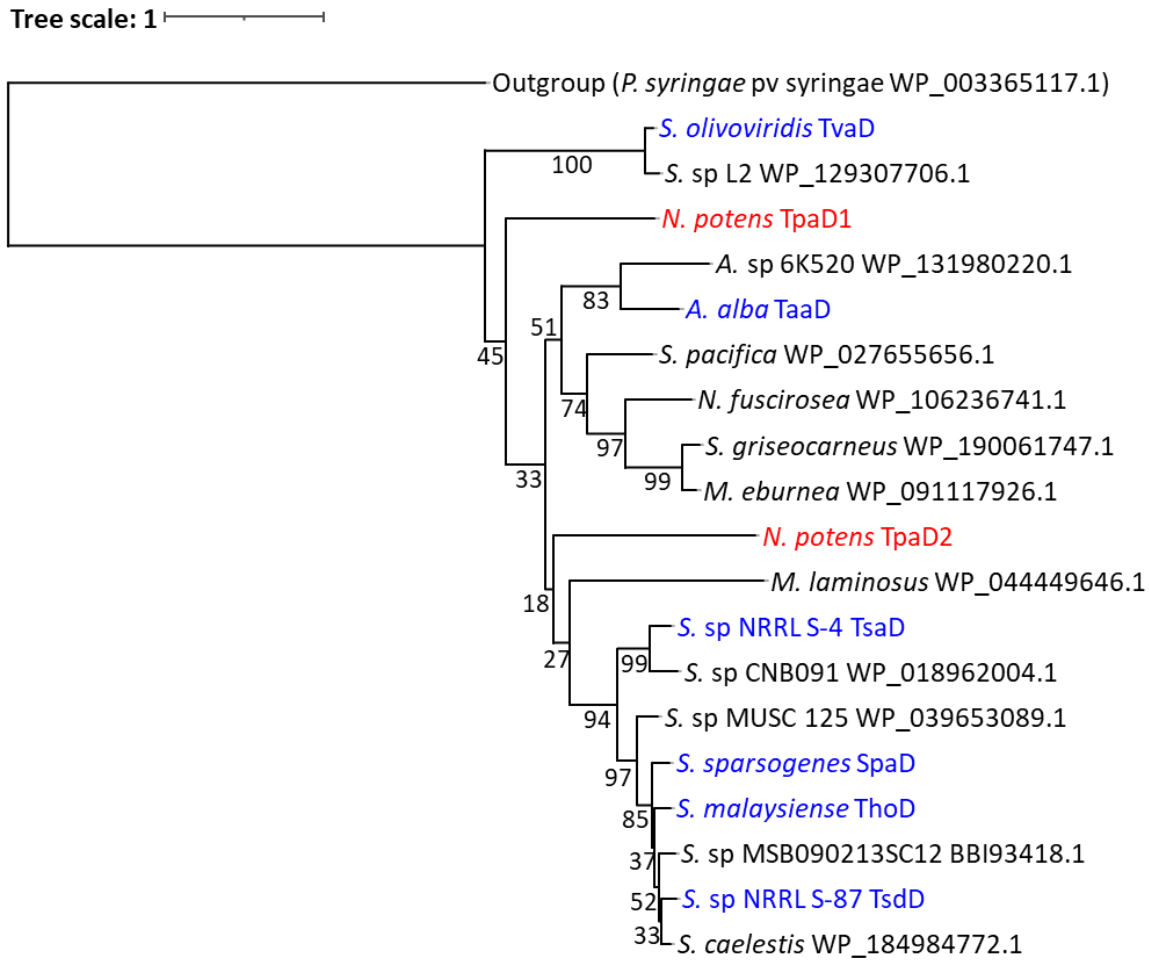


Figure 2.28: Maximum likelihood phylogenetic tree of HopA1-like proteins from thioamitide BGCs. HopA1 protein from *Psuedomonas syringae pv syringae*, not belonging to a thioamitide BGC was employed as the outgroup for the tree. HopA1-like proteins from the biosynthetic pathways of characterised thioamitides are labelled in blue, and from the biosynthetic pathway of thiopotensamide A was labelled in red. Numbers at the nodes are the bootstrap values.

Once again, the phylogenetic analysis of the HopA1 proteins was inconclusive (Figure 2.28). While TpaD2 seemed to cluster closer to most of the HopA1 proteins from known thioamitides, TpaD1 is by no means an outlier, and in fact it is more closely related to the rest of the "D" homologues than TvaD from the thioviridamide pathway, as was the case with the phosphotransferase homologues. Interestingly, TpaD1 and TpaD2 branch out independently in the phylogenetic tree instead of as a pair, indicating that even if one of them originated as a duplication of the other in an ancestral cluster, they have had time to evolve and become quite distinct proteins.

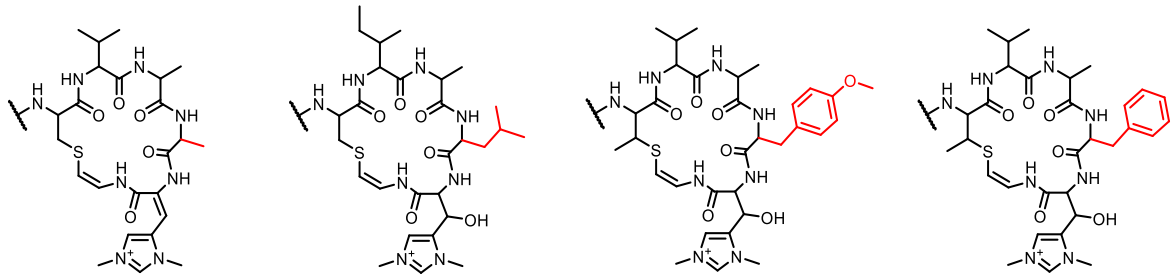


Figure 2.29: Avi(Me)Cys macrocycles from currently available thioamitides. The sidechain of the amino acid neighbouring the His12 residue was highlighted in red. *the structure of thiopotensamide A is tentative.

Another possible explanation for the presence of the dehydration in thiopotensamide A, assuming no additional specific enzymatic activities were needed, could be substrate related, with dehydration happening simply because this histidine residue is more accessible for modification than its equivalent in other thioamitides. If this was true, the reaction would be less likely to take place with a bulkier sidechain on the adjacent amino acid(s). Hence, the adjacent amino acids to His12 were compared in several thioamitides (Figure 2.29). The core peptide of thiopotensamide A has the least sterically hindered group neighbouring the hdmHis12 compared to the other core peptides, as the -1 amino acid of

His12 in thiopotensamide A is alanine, where in the rest of characterised hydroxylated thioamitides this position is occupied by leucine in thioviridamide and derivatives, tyrosine in thiostreptamide S4, and phenylalanine in thiostreptamide S87, thioholgamide and thiosparsoamide. Whether this hypothesis is correct could be proven by introducing different amino acids in the position preceding histidine in the core peptide.

Following these analyses, and given the struggle to obtain pure thiopotensamide A, it became clear that genetic manipulation of the cluster was required to further validate the structure of the molecule as well as to determine the genetic origin of the additional tailoring modifications in the molecule.

2.3.6.2. *tpaC* and *tpaD2* are essential to the biosynthesis

Since the final product from the heterologous expression lacked the dehydration in the hdmHis residue, it meant either the responsible proteins were either not being expressed properly or not encoded within the cloned BGC. If the genes were cloned but not expressed in *S. coelicolor* M1146, between the two phosphotransferase-HopA1 pairs the more likely to malfunction would be *tpaC* and *tpaD2*. Since all the tailoring enzymes appeared to be in the same operon and the other phosphotransferase-HopA1 pair (*tpaD1* and *tpaE*) were encoded at the beginning of it, other modifications such as thioamidation, histidine methylations, hydroxylation and macrocyclization were expected to be absent if the first two genes in the operon failed to be transcribed. Since hydroxylation was observed in thiopotensamide B, the responsible oxygenase encoded by *tpaJ* was extremely likely to be correctly expressed as well as the YcaO and TfuA proteins (encoded by *tpaH* and *tpaI*) responsible for the thioamidations. *tpaK*, the gene after *tpaJ* in the BGC, encodes the protease supposedly involved in cleaving the leader peptide, but previous work on *tsaK*,

the *tpaK* homologue, from thiostreptamide S4 BGC showed it was non-essential for biosynthesis, so it was uncertain whether this gene or *tpaC* and *tpaD2* after it were being expressed properly. Furthermore, the fact that these two genes had clearly been involved in a genetic rearrangement, could have led to problems with their expression. If one or both of those two genes were not correctly expressed, this could lead to the lack of hdmHis dehydration.

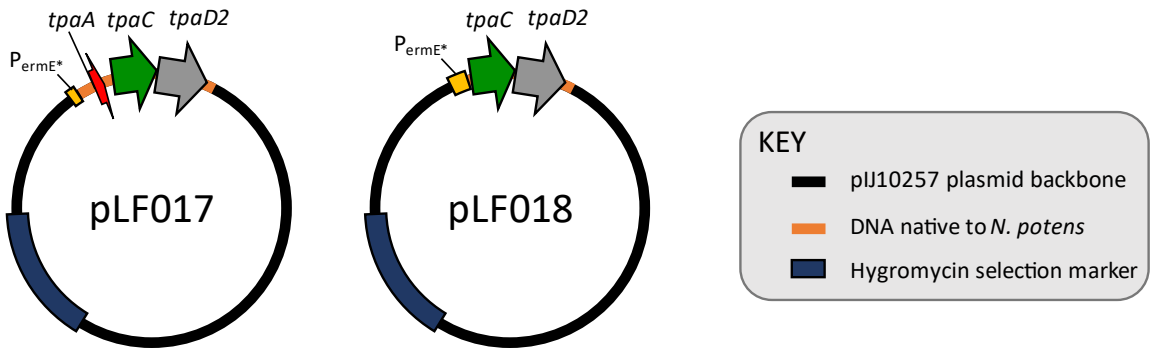


Figure 2.30: Constructs for *in trans* overexpression of *tpaC* and *tpaD2*.

To test this hypothesis, both *tpaC* and *tpaD2* were cloned into pIJ10257, which contains strong promoter P_{ermE*} and utilises the ϕBT1 integration site (Bibb et al., 1994, Hong et al., 2005). Two constructs were made for the *in trans* overexpression of the genes (Figure 2.30). Both constructs contained *tpaC and tpaD2*, but one of those constructs also contained a copy of *tpaA* including the intergenic regions before and after this gene. The construct including *tpaA* was made to test in parallel if providing more precursor peptide would enhance the productivity (Section 2.3.6.5). Neither of the constructs complemented the dehydration at the hdmHis residue, so it was not possible to confidently determine from this experiment whether these genes had a role in hdmHis dehydration or even were involved in thiopotensamide biosynthesis at all.

Gene deletions could provide a better understanding of the roles of these genes in the biosynthesis. If deleting either *tpaC* or *tpaD2* from the BGC interrupted the biosynthesis, that would indicate the gene(s) have an essential role in the biosynthesis of thiopotensamide B. This could be further confirmed by performing an *in trans* complementation with the constructs described previously. On the other hand, if the biosynthesis of thiopotensamide B was not affected when one of or both genes was removed from the BGC, the enzymes could be considered as non-essential for the biosynthesis.

tpaC and *tpaD2* in pLF016 were individually deleted using PCR targeting in *E. coli*, and then the modified clusters were integrated to *S. coelicolor* M1146 to test the production of thiopotensamide. The deletion of *tpaC* generated construct pLF016_ΔC, and the deletion of *tpaD2* generated a construct pLF016_ΔD2. Neither of the constructs produced thiopotensamide A nor B, suggesting both genes were required for the biosynthesis of thiopotensamides (Figure 2.31a) The knockout mutants were complemented with the plasmid constructs (pLF017 and pLF018) that were made for over expressing *tpaC* and *tpaD2*. Production of thiopotensamide B was restored by pLF018 in both mutant strains, but pLF017 only complemented the production in pLF016_ΔD2 and did so at much lower levels than pLF018. These results confirm that *tpaC* and *tpaD2* are essential for the biosynthesis of thiopotensamides, likely for the dehydration of its serine residues. The results from the complementation also suggest that neither *tpaC* nor *tpaD2* are responsible for the removal of hydroxyl group on hdmHis residue. Instead, this role could be carried out by *tpaD1* and *tpaE*, or by genes encoded outside the cloned BGC. Although *tpaD1* and *tpaE* were hypothesised to be expressed in the BGC, it was still possible that one of or both

genes were not expressed, *tpaD1* particularly. Heterologous expression of bottromycin (*btm*) BGC in *S. coelicolor* M1146 led to production of immature bottromycin metabolites due to the first gene *btmC* in one of the operons appearing to be inactive despite no mutations to the gene (Eyles et al., 2018). If *tpaD1* is involved in the dehydration in the hdmHis residue and was inactive, production of thiopotensamide without dehydration at hdmHis would be possible.

It was surprising that the knockout mutants behaved differently when complemented with pLF017 and with pLF018 (Figure 2.31). These results indicated the presence of *tpaA* appeared to have a negative impact to the expression of *tpaC* and *tpaD2*. This could be due to the fact that the intergenic region between *tpaA* and *tpaC* in the plasmid did not possess a suitable ribosome binding site (RBS) for *tpaC* expression, since this intergenic region was the one originally following *tpaA* and not the one before *tpaC* in the original BGC. Analysis of the sequence preceding *tpaC* in the plasmid did show an AG rich region that could act as an RBS, but whether this would be suitable for expression of *tpaC* is not clear. More likely, this result is the effect of an attenuator sequence located after the coding sequence of *tpaA* (Figure 2.20), as it was discussed previously when describing the refactoring of the cluster (Foulston and Bibb, 2010, O'Rourke et al., 2017, Vior et al., 2020).

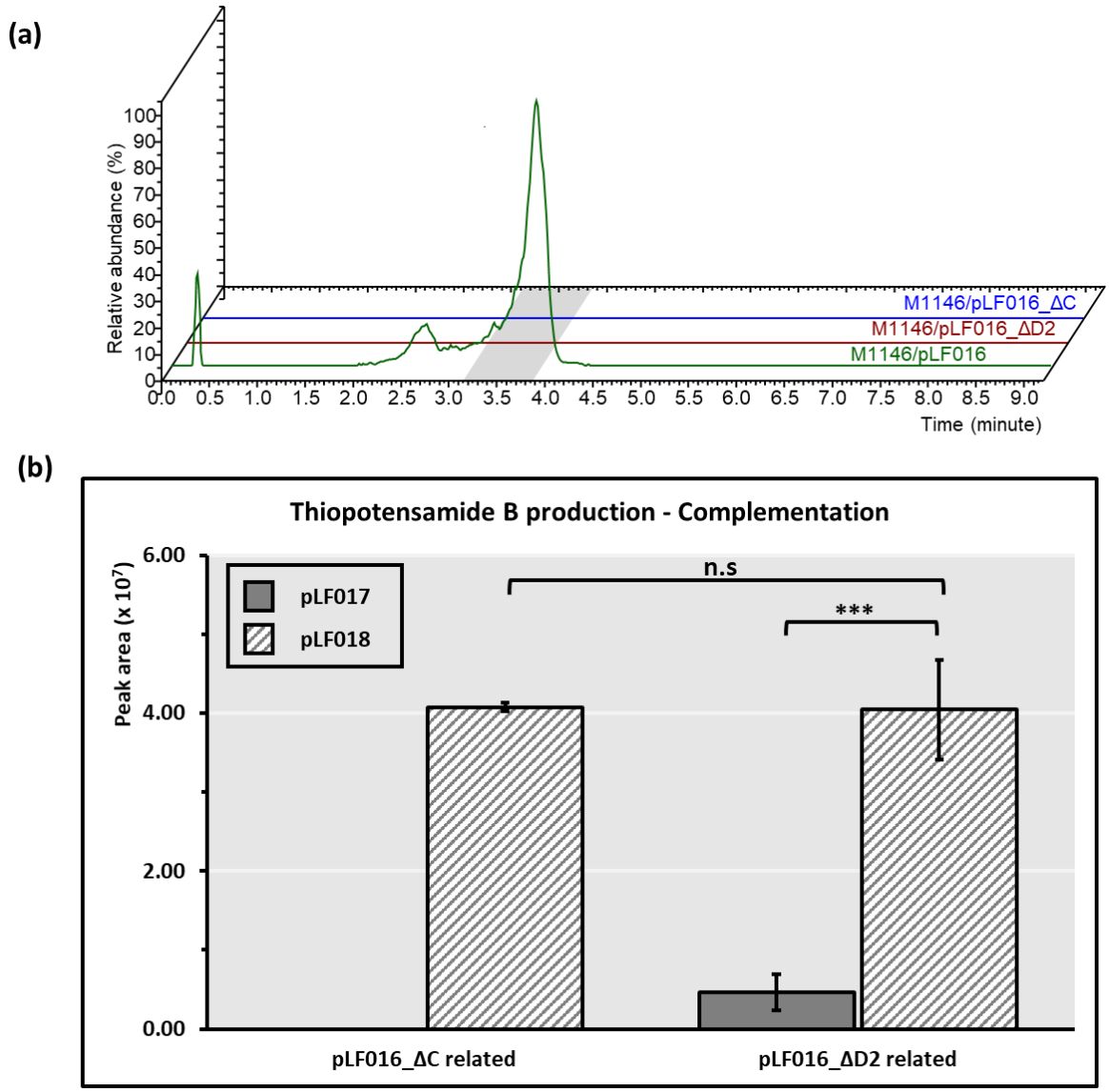


Figure 2.31: Production of thiopotensamide B in *S. coelicolor* M1146 mutants related to genes *tpaC* and *tpaD2*. XIC of thiopotensamide B in (a) the knockout mutant M1146/pLF016_ΔC and M1146/pLF016_ΔD2 in comparison to the unmodified refactored cluster, (b)Summary of production from mutants M1146/pLF016_ΔC and M1146/pLF016_ΔD2 when the mutant was complemented with pLF017 and pLF018. Production of thiopotensamide B was not significantly different between the two mutants when complemented with pLF018 (p-value = 0.960). Whereas, when M1146/pLF016_ΔC and M1146/pLF016_ΔD2 were complemented with pLF017, there was no production or significantly lower production (p-value = 0.0005) of thiopotensamide B. Data were collected from 5 biological replicates and error bars were plotted by standard mean error.

As an alternative responsible for the histidine dehydration in thiopotensamide A, a lanthipeptide synthetase gene from *N. potens* was tested. Previous work in the thiosparsoamide cluster showed a class III lanthipeptide synthetase homologue (SpaKC) in

S. sparsogenes was capable of dehydrating the serine and threonine in the non-thioamidated core peptide region from thiosparsoamide precursor peptide (Lu et al., 2021) . In the *N. potens* genome there is also a SpaKC homologue (PotKC, accession number: WP_017595403.1) encoded outside the thiopotensamide BGC. If PotKC was a dehydratase with more relaxed specificity than its SpaKC homologue, it could potentially carry out the dehydration reaction at the hdmHis residue in thiopotensamide B. Consequently, *potKC* was cloned into pIJ10257 to yield pLF021 for the expression of PotKC *in trans* with the BGC of thiopotensamide.

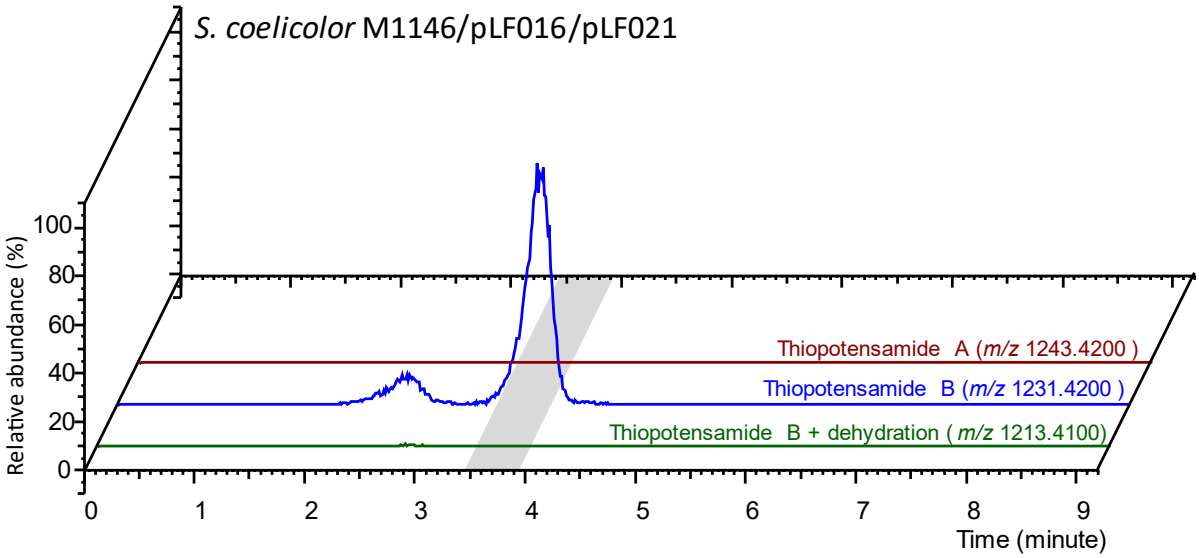


Figure 2.32: Production of thiopotensamides in *S. coelicolor* M1146/pLF016/pLF021. XICs for thiopotensamide A in red, thiopotensamide B in blue and a potentially dehydrated version of thiopotensamide B in green.

The LCMS results showed that PotKC did not dehydrate the hdmHis residue and the final product remained as thiopotensamide B, which indicated that the PotKC could not modify thiopotensamide B on its own (Figure 2.32). This result was not entirely surprising. If this was a generalised modification mechanism, some of the other hdmHis containing thioamitides should also undergo dehydration, as their producing strains also possess a SpaKC homologues. For example, *S.* sp. NRRL S-4 (SnrKC, accession number:

WP_053928640.1) and *S. malaysiense* (MalKC, accession number: WP_071381943.1), but respectively, neither thiostreptamide S-4 nor thioholgamide are dehydrated at the histidine.

Although the responsible dehydratase protein(s) was unclear, the results ruled out the hypotheses that the dehydration was solely made by the TpaC-TpaD2 pair or the external lanthipeptide synthetase PotKC. It is possible that the modification could potentially be made by both the other phosphotransferase-HopA1 pair (TpaE-TpaD1) or at least one of them whilst this responsible protein in the BGC is inactive, or the modification required another enzyme or enzyme complex that was encoded outside of the cloned BGC.

2.3.6.3. Restoring the methylation expands the BGC boundary

Since there were multiple modifications missing in thiopotensamide B with respect to the original thiopotensamide, which are methylation near the N-terminal, oxidation at the methionine and the dehydration at hdmHis residue, the BGC boundaries established initially needed to be re-evaluated as two of these modifications are unprecedent in other characterised thioamitides. The aim was to express any genes potentially belonging to the real cluster in the refactored cluster (pLF016) background to restore the missing modifications.

BLASTP was employed to analyse the proteins encoded by the genes within 15 kb either side of *tpaH* gene (Figure 2.33, Table 2.4). Two methyltransferases (*tpa-7* and *tpa+3*) were found outside but near the original BGC. Both methyltransferases as well as *tpaG*, the histidine methyltransferase from the cluster are SAM-dependent methyltransferases. This class of proteins can install methylations in different target atoms including oxygen, nitrogen and sulphur. In order to identify whether either of the external methyltransferases

was involved in the biosynthesis of thiopotensamide A, *tpa-7* and *tpa+3* were individually cloned into pIJ10257 to create pLF024 and pLF025 respectively for *in trans* methyltransferase overexpression.

Table 2.4: BLASTP results for the proteins encoded surrounding of the thiopotensamide BGC. The displayed results correspond to the top hit after *N. potens*.

Query gene	Assession number	Conserved domains	Homologous gene annotation	Organism	Coverage (%)	Identity (%)
Tpa+6	WP_017595605.1	COG4249/DNA2	Caspase family protein	*Streptomonospora halotolerans*	99	61.14
Tpa+5	WP_017595606.1	None	Hypothetical protein	*Nocardiopsis* sp. RSe5-2	98	59.15
Tpa+4	WP_040710961.1	Penicil_amidase	Penicillin acylase family protein	*Streptomonospora litoralis*	99	73.98
Tpa+3	WP_017595608.1	AdoMet_Mtases	Methyltransferase	*Mesorhizobium* sp.	98	42.65
Tpa+2	WP_086003916.1	PRK06184	FAD-dependent monooxygenase	*Candidatus Eremiobacteraeota bacterium*	89	36.75
Tpa+1	WP_017595610.1	HTH_MerR-like	MerR family transcriptional regulator	*Nocardiopsis* sp. SBT366	93	58.50
			***Tpa* BGC**			
Tpa-1	WP_026120726.1	GTB-type	Activator-dependent family glycosyltransferase	*Nocardiopsis* sp. JB362	98	46.86
Tpa-2	WP_017595623.1	None	Hypothetical protein	*Nocardiopsis* sp. JB363	80	33.75
Tpa-3	WP_017595624.1	None	Possible integral membrane protein	*Nonmuraea gerenzanensis*	99	56.59
Tpa-4	WP_017595625.1	None	LmbU family transcriptional regulator	*Nocardiopsis* sp. MG754419	99	57.50
Tpa-5	WP_017595626.1	Hexose_dehydrat ase	NDP-hexose 2,3-dehydratase family protein	*Nocardiopsis* sp. L17-MgMaSL7	99	75.89
Tpa-6	WP_017595627.1	AAT_I	DegT/DnrJ/EryC1/StrS family aminotransferase	*Streptomyces* sp. BR123	99	81.77
Tpa-7	WP_017595628.1	KetoArg_3Met	Class I SAM-dependent methyltransferase	*Nocardiopsis eucommiae*	95	72.15
Tpa-8	WP_026120727.1	Cupin_Rm1C-like	dTDP-4-dehydrorhamnose 3,5-epimerase family protein	*Nocardiopsis* sp. Huas11	100	78.50
Tpa-9	WP_017595630.1	Epimerase	NAD-dependent epimerase/dehydratase family protein	*Streptomyces* sp. I6	94	61.61

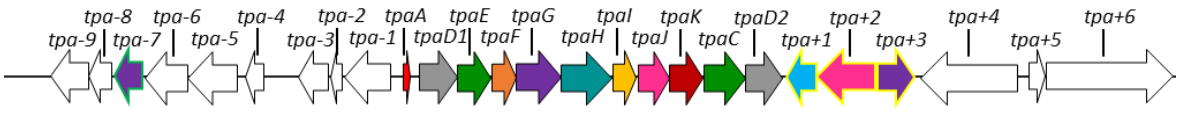


Figure 2.33: Thiopotensamide BGC and flanking genes. Gene labels correspond to query gene in Table 2.4.

Over expression of methyltransferase (*tpa-7*) from upstream of the BGC in the heterologous host did not result in methylation of thiopotensamide B, but this modification was successfully restored when the pathway was complemented with methyltransferase (*tpa+3*) from downstream (Figure 2.34). In the production screening for M1146/pLF016/pLF024, thiopotensamide B was not detected. Instead, a peak with *m/z* 1245.4396 was seen. This mass was 14 Da greater than thiopotensamide B. This novel molecule, corresponding to a methylated thiopotensamide B (Exact mass: 1245.4350) with a mass error of +3.69 ppm, was named thiopotensamide C.

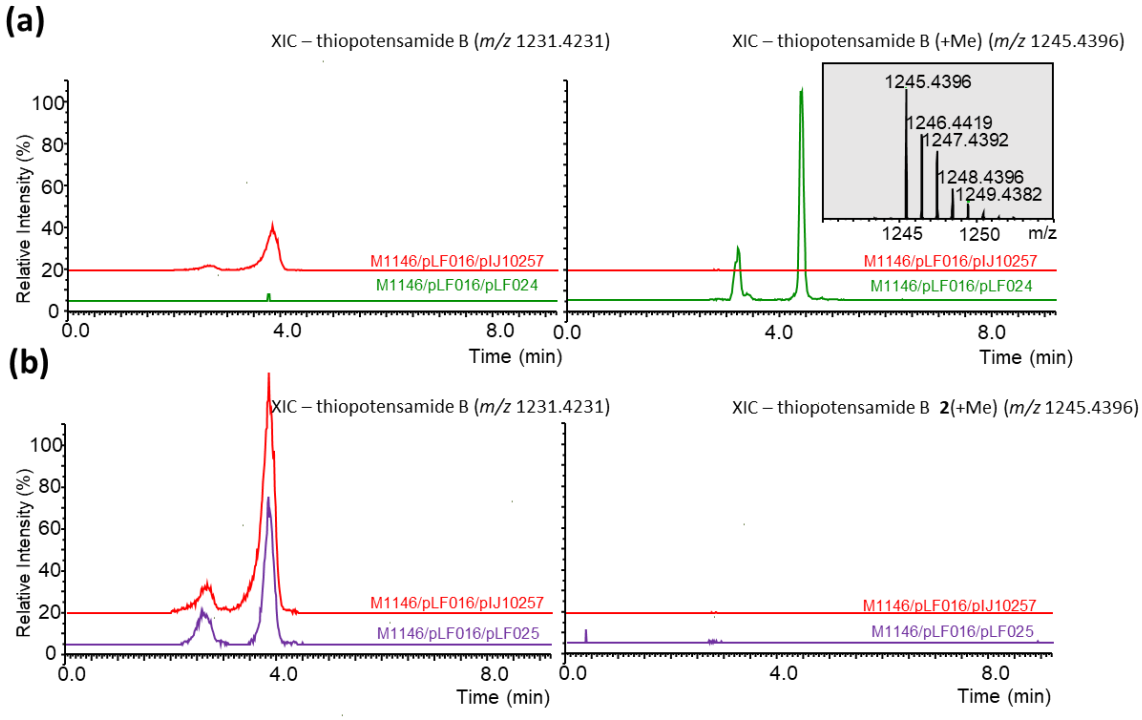


Figure 2.34: LC-MS analysis of thiopotensamide production after expression potential additional tailoring genes for methylation of thiopotensamide B. (a) XIC for thiopotensamide B (left) and potentially methylated product (right) in strain expressing of methyltransferase *tpa+3* (green) versus empty vector control (red) (b) Expression results for strain expressing methyltransferase *tpa-7* (purple trace). The chromatograms were globally scaled to the highest XIC within the sample.

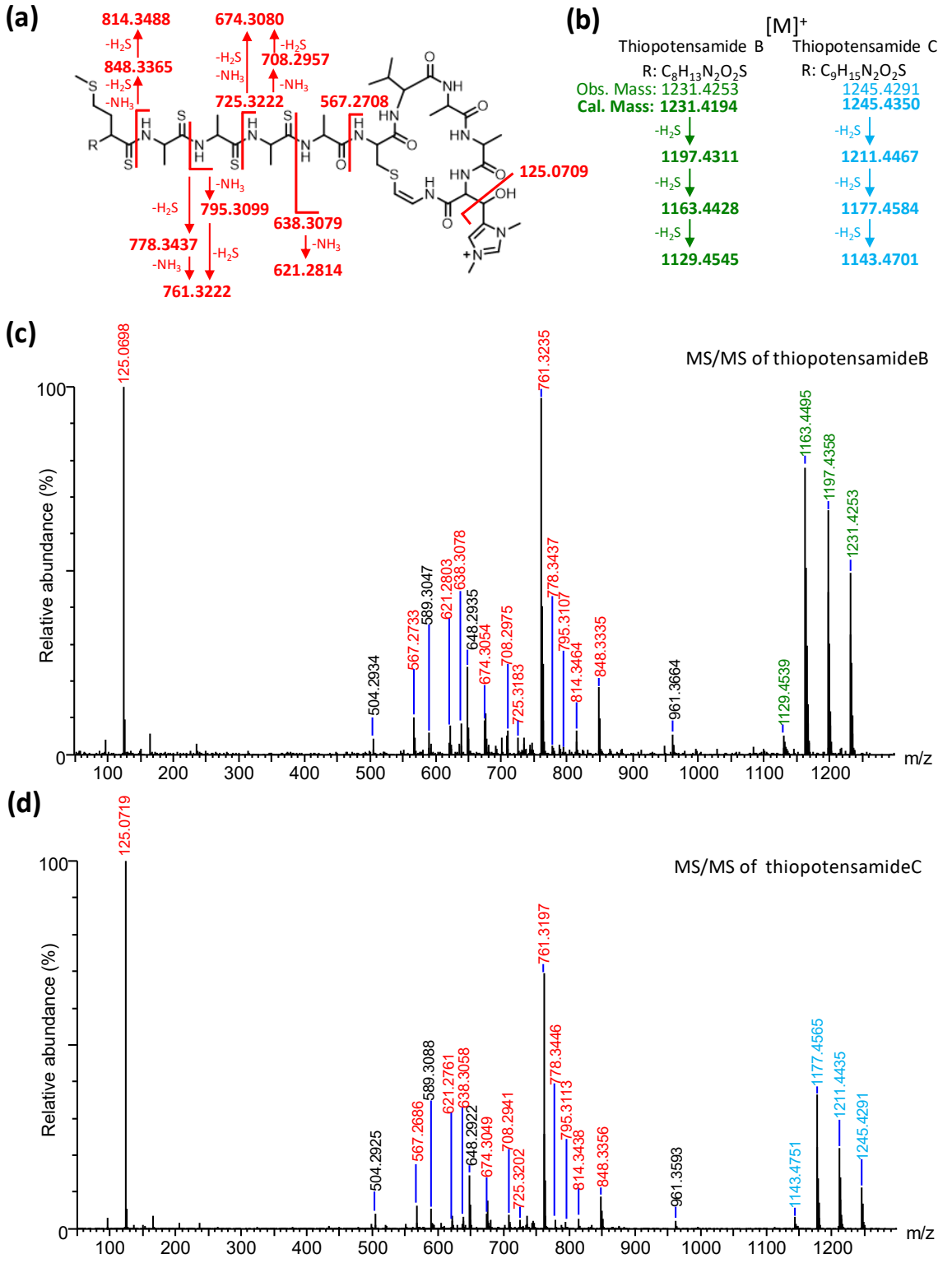


Figure 2.35: Fragmentation patterns of thiopotensamide B and C and their MS/MS spectra. (a) Expected fragments for the part of the molecule identical in thiopotensamide B and C. (b) Theoretical and observed masses of the molecular ions for thiopotensamide B and C and the calculated masses resulting from neutral losses of H_2S. (c) MS/MS spectrum of thiopotensamide B. (d) MS/MS spectrum of thiopotensamide C. Masses in the spectra are colour coded to match the predictions in panel a and b; predicted masses are labelled in bold.

The MS/MS fragmentation from thiopotensamide C (Figure 2.35) showed that this molecule has the same neutral losses of *m/z* 33.99 and the hdmHis fragment with *m/z* 125.07 is observed in thiopotensamide B. In fact, the entire middle mass range MS/MS spectra (*m/z* 500-1000) of both thiopotensamide B and C is identical, and a mass difference of 14 Da can only be observed in the biggest fragments (above *m/z* 1000). Therefore, thiopotensamide C was confirmed to be a methylated derivative of thiopotensamide B. What was more encouraging was that the methylation also appeared to be at Pyr1, Val2 or Met3 position, consistent with the hypothesis for thiopotensamide A. This result confirms that the product of *tpa+3* is involved in the biosynthesis of thiopotensamides and is part of its BGC. Therefore, gene *tpa+3* is reannotated as *tpaMT*. Despite these results, there was still no clear evidence to show where exactly TpaMT introduces the methylation.

As for *tpa-7*, this methyltransferase is in fact part of the potensimicin BGC responsible for the production of kayamycin, which is located upstream of the thiopotensamide cluster. This methyltransferase is required for mycaminose sugar formation, so it is not surprising it did not methylate thiopotenamide B.

The secondary structure of TpaMT was predicted by Phyre2, a protein fold recognition server to gain more information about the protein (Kelley et al., 2015). Phyre2 modelled the predicted secondary structure of TpaMT after an *O*-methyltransferase, LaPhzM, for the formation of Phenazine in *Lysobacter antibioticus* with 100% confidence and 97% coverage (Figure 2.36) (Jiang et al., 2018). The significant matching result from Phrye2, suggests that TapMT could potentially be an *O*-methyltransferase and the methylation could be installed on the Pyr1 residue (Figure 2.37). Since this protein is a SAM dependent domain

methyltransferase, the catalytic reaction relies on SAM as a methyl donor and produces *S*-adenosyl-L-homocysteine (SAH) as by-product.

Although bioinformatic analysis suggests that TpaMT has high similarity with an *O*-methyltransferase, and methylation catalysed by TpaMT was consequently hypothesised on the pyruvyl moiety on thiopotensamide B, the active site residue that was found in LaPhzM could not be identified in TpaMT. Hence, it is not clear whether TpaMT is an *O*-methyltransferase or an *N*-methyltransferase, which means experimental data such as NMR analysis is required to support the proposed position of the extra methylation in thiopotensamide A and C.

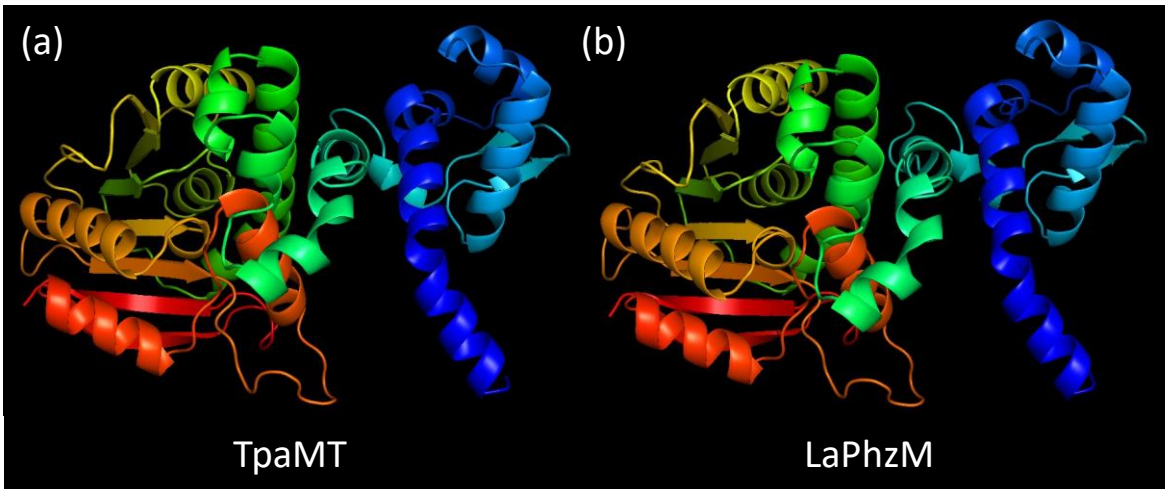


Figure 2.36: Prediction of TpaMT secondary structure from Phyre 2 (a) and the secondary structure of LaPhzM from *Lysobacter antibioticus* OH13 (b).

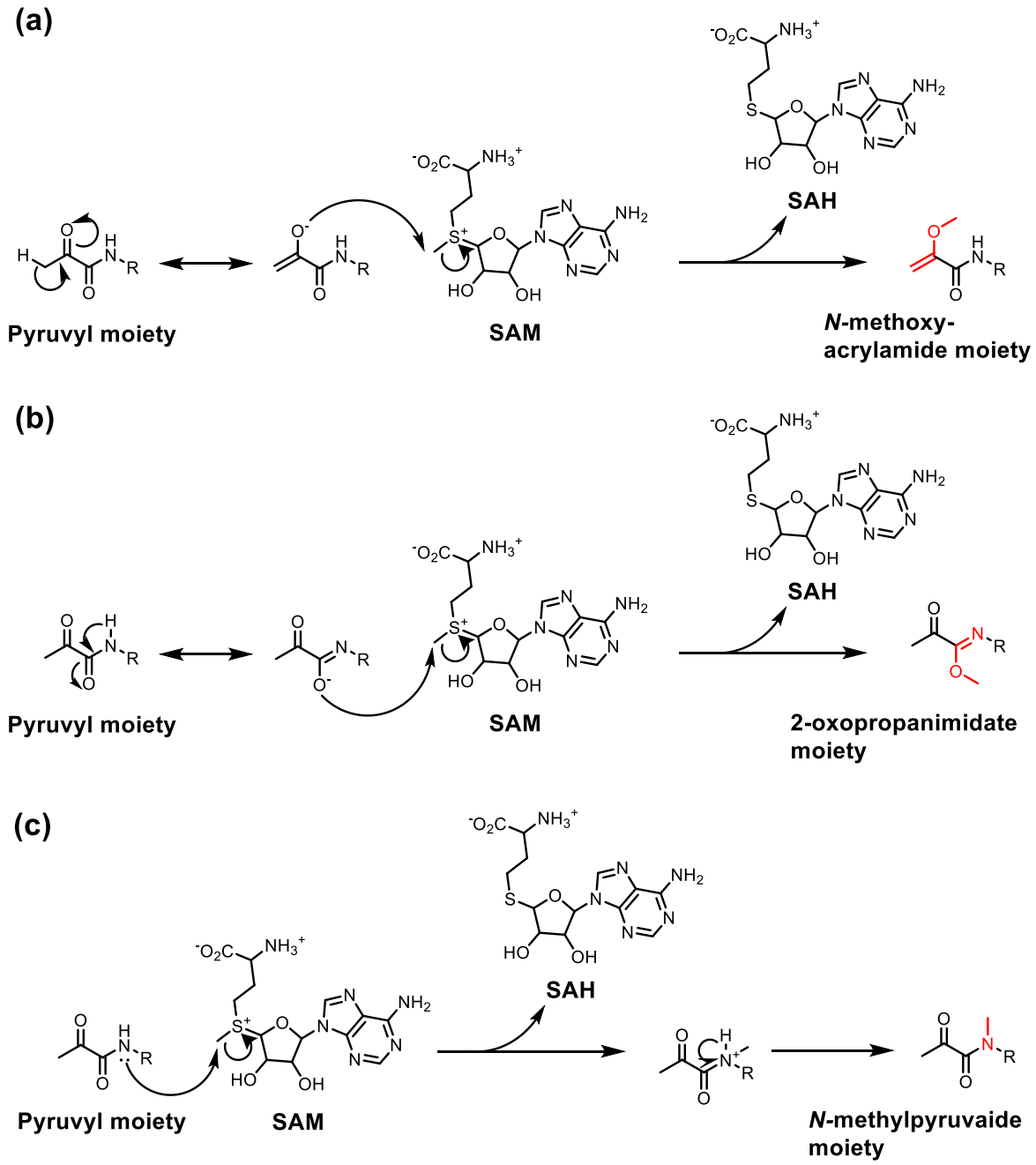


Figure 2.37: The 3 potential methylation positions in pyruvyl moiety.

2.3.6.4. Attempt to restore the MetO by overexpressing the monooxygenase *tpa+2*

Since methyltransferase *tpaMT* was proved to be part of the thiopotenamide BGC, it seemed feasible that the genes before could also be part of it. A MerR transcriptional regulator (*tpa+1*) and a FAD-dependent monooxygenase (*tpa+2*) were located between *tpaD2* and *tpaMT*. In the native producer strain *N. potens*, the thiopotensamide is always

oxidised at the methionine residue. Oxidation was thought to be introduced by non-enzymatic reactions, but the thiopotensamide produced through heterologous expression does not possess an oxidation on the methionine. One of the possibilities for this phenomenon could be that the oxidation on methionine could be enzymatic. A monooxygenase domain containing protein was previously reported to oxidise defined methionine residues in actins (Ortegón Salas et al., 2020). Hence, *tpa+2* was also individually cloned into pIJ10257 to generate pLF023.

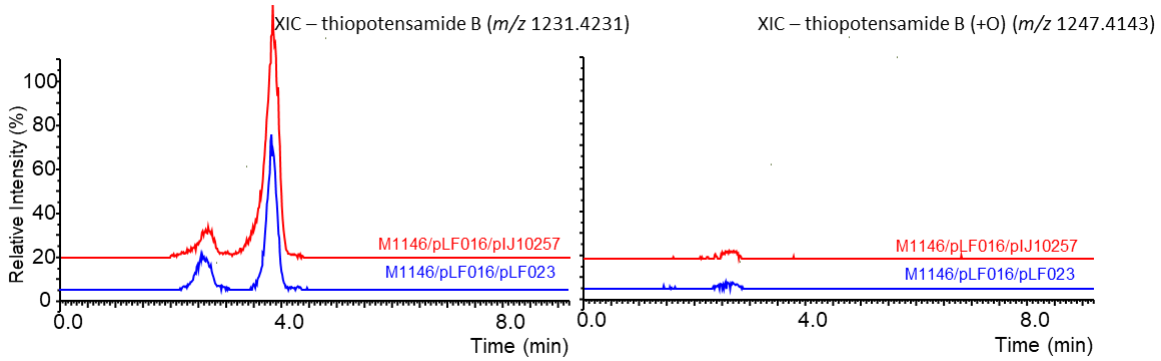


Figure 2.38: LC-MS analysis of thiopotensamide production after expression potential additional tailoring genes for oxidation of thiopotensamide B. XIC for thiopotensamide B (left) and potentially oxidised product (right) in strain expressing monooxygenase *tpa+2* (blue trace) versus empty vector control (red).

If thiopotensamide B was oxidised, a mass shift of +16 was anticipated to yield a compound with a theoretical mass of 1247.4143. When pLF023, containing monooxygenase *tpa+2*, was co-expressed *in trans* with the refactored thiopotensamide BGC in M1146/pLF016/pLF023, there was no detection of *m/z* 1247.4143, while thiopotensamide B was still detected (*m/z* 1231.4123, Figure 2.38). This result indicates that either the monooxygenase Tpa+2 cannot oxidise thiopotensamide B, or at least it requires either a partner protein that is not in the constructs or a modification that is missing in thiopotensamide B.

2.3.6.5. Regulatory genes and other production improvement attempts

Not all thioamitide BGCs have a regulatory gene in their BGC, although there are 2 characterised thioamitides that do have them near or within their BGCs: *tvaB* for thioviridamaide and *taaB* for thioalbamide. A MerR transcriptional regulator gene (*tpa+1*) is in the region between *tpaD2* and *tpaMT*, so we were curious about whether this MerR transcriptional regulator had a role in the biosynthesis of thiopotensamide. MerR family transcriptional regulators are ubiquitous across bacteria, and they in general act as activators (Brown et al., 2003). Therefore, the *tpa+1* was cloned in the integrative vector pIJ10257, generating plasmid pLF022, and expressed *in trans* with the refactored BGC.

In addition, during the characterisation of thiostreptamide S4, thiostreptamide S87 and thioalbamide BGCs in the lab, thioalbamide had the highest purified yield amongst all three, and it was the only one out of three that has a *Streptomyces* antibiotic regulatory protein gene (*taaB*) in its pathway, although very little was known about its role in thioalbamide biosynthesis. It would be interesting to see whether expressing the gene in parallel would change the productivity. For this reason, a *taaB* containing construct (pLF026) was made using pIJ10275 and introduced in the heterologous host. According to the production test results, there were no significant changes in production in either of the strains (Figure 2.39).

One main concern of the experiment was that the refactored construct (pLF016) substituted the intergenic region before *tpaA* from the original genome, which may contain the biniding region for activating the transcription of the operon. In the study of microbisporicin BGC, the regulatory protein boosted the production of β-glucuronidase (GusA), a production marker for the experiment, only when the correct promoter region

was provided upstream the *gusA* gene (Fernández-Martínez et al., 2015). Hence, the regulatory genes were also tested with the wildtype construct (pLF015).

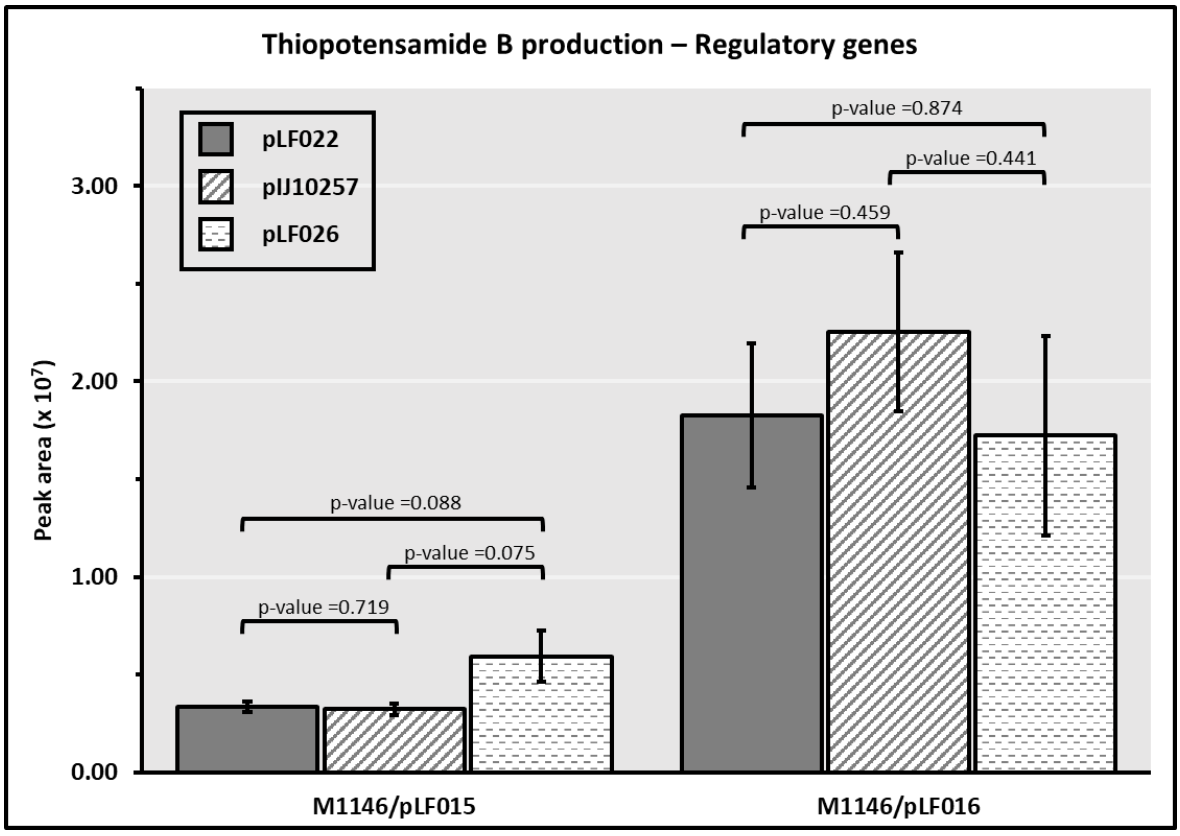


Figure 2.39: Thiopotensamide B production in different *S. coelicolor* M1146/pLF015 and M1146/pLF016 overexpressing potential thioamitide production regulators *in trans*. M1146/pLF015 contains wildtype thiopotensamide BGC and M1146/pLF016 contains refactored thiopotensamide BGC. pLF022 contains MerR regulator *tpa+1*. pLF026 contains *taaB* from the thioalbamide BGC. Data was collected from 5 biological replicates and the error bars represent the standard mean error between replicas.

It was clear that overexpressing the regulatory genes *in trans* to the refactor BGC did not statistically change the level of thiopotensamide B production. Meanwhile, interesting results were seen in the wildtype construct (pLF015). When the strain was provided MerR gene *tpa+1*, it showed no changes in production level of thiopotensamide B in comparison to its negative control M1146/pLF015/pIJ10257. However, its role in thiopotensamide production cannot be completely discarded because the pathway was expressed in *S. coelicolor* as it would be in its native host, a different genus, and the regulatory activity

could be affected by this. Intriguingly, when the construct was co-expressed with the regulatory gene *taaB* from thioalbamide BGC, the production appeared to be enhanced, despite the fact that none of the BGC (from *Nocaridiopsis*), regulatory protein TaaB (form *Amycolatopsis*) and the expression host (*Streptomyces*) were native to each other. Although the increase of production by TaaB was not statistically significant (p-value = 0.075), it indicated a potential increase for thioamitide production.

Another strategy to try to increase thiopotensamide production was to provide additional precursor peptide for the biosynthesis. As it was discussed previously, in RiPP biosynthesis the precursor peptide is the scaffold for the final molecule, so it was hypothesised that the presence of an extra copy of *tpaA* under a strong promoter could lead to increased production. In order to test this hypothesis, the strains carrying the plasmids pLF017 and pLF018 described in section 2.3.6.2 were compared as the only difference between those two strains is the presence of an extra copy the *tpaA* gene driven by a strong promoter P_{ermE*} in pLF017. However, the production of thiopotensamide did not show any statistically significant difference between M1146/pLF015/pLF017 and M1146/pLF015/pLF018 (Figure 2.40).

There could be multiple reasons that the production in M1146/pLF015/pLF017 was not significantly increased, but it could be linked to degradation of the compound. After all, thiopotensamide B is in fact an intermediate molecule of the biosynthetic pathway to thiopotensamide A. This leaves the question of whether the expression host recognises thiopotensamide B as an immature molecule and degrades it, as it has been reported to happen in other RiPPs (Eyles et al., 2018). As an alternative, one of the missing modifications could be the limiting factor instead of the numbers of precursor peptide. MS

networking and comparative metabolomics would the appropriate tools to identify the shunt molecules (Johnston et al., 2015, Eyles et al., 2018).

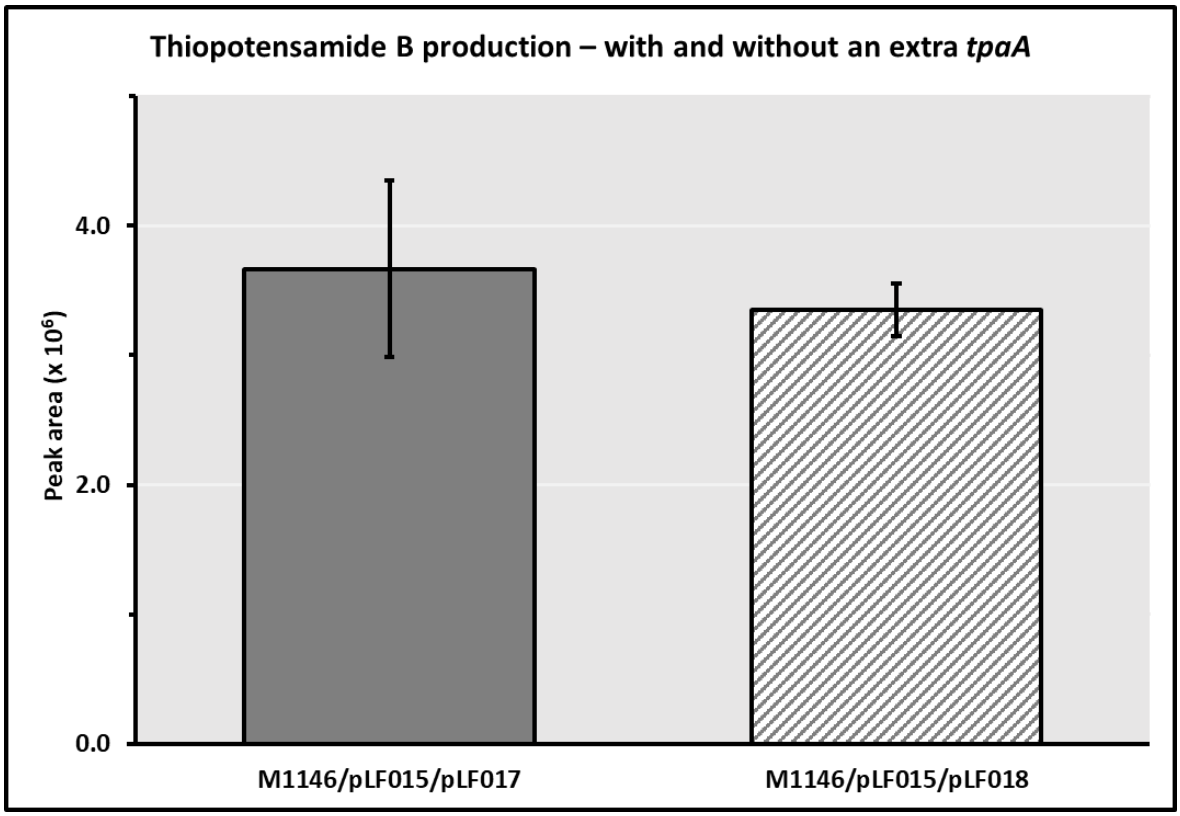


Figure 2.40: Thiopotensamide B production in different *S. coelicolor* M1146/pLF015 strains. M1146/pLF015/pLF017 contains an extra copy of *tpaA*, *tpaC* and *tpaD2*. M1146/pLF015/pLF018 contains only an extra *tpaC* and *tpaD2*. Data was collected from 5 biological replicates and the error bars represent the standard mean error. The difference on thiopotensamide B production was not statistically significant (p-value=0.67).

2.4 Conclusions

Genome mining for novel thioamitide clusters led to the discovery of four additional thioamitide BGCs in addition to the 14 previously known clusters (Frattaruolo et al., 2017). Three actinobacterial clusters from *M. eburnea*, *N. fuscirosea* and *N. potens* were chosen for follow up screening experiments. After fermenting the candidate strains on a selection of screening media, a novel thioamitide, thiopotensamide A, was discovered being produced by *N. potens* (Figure 2.6). Similar to other thioamitides, thiopotensamide A can be structurally divided into 2 parts, an AviCys macrocycle and a thioamide containing linear chain. Thiopotensamide A consists of 5 thioamide bonds. A neutral loss (-63.99) found in the MS/MS spectrum of thiopotensamide A indicated that the methionine residue from thiopotensamide A is oxidised (Guan et al., 2003, Miles et al., 2014, Frattaruolo et al., 2017). The lack of the hdmHis fragment (*m/z* 125.0709) present in other characterised thioamitides suggests an absence of hydroxyl group in the histidine residue of thiopotensamide A, and this absence of hydroxyl group is likely due to an E1cB elimination mediated dehydration (Ke et al., 2011, Frattaruolo et al., 2017, Sikandar et al., 2022). One of the two pairs of phosphotransferase-HopA1 genes present in the cluster were speculated to be responsible for the dehydration, but phylogenetic analyses did not provide any clear direction on which pair was more likely to catalyse the dehydration. Finally, a methylation on either Pyr1, Val2 or MetO3 position was also proposed based on the molecular ion mass and according to the MS/MS fragmentation pattern.

Purification of thiopotensamide A was attempted for NMR analysis to confirm the proposed additional modifications and to locate the methylation. Multiple orthogonal separation techniques were employed, but another metabolite co-eluted with

thiopotensamide A, which hindered the purification. This metabolite was later identified as the macrolide kayamycin, previously described as a product of *N. potens* (Rengaraju, 1985). Despite trying techniques specifically targeting the separation of thiopotensamide A and kayamycin, the purification was unsuccessful at the time of writing this thesis.

Genetic manipulation of *N. potens* was attempted, but the strain proved to be difficult to manipulate. Consequently, a PCR-based TAR cloning approach was employed to heterologously express the BGC for the production and purification of thiopotensamide A and to investigate its biosynthesis through genetic manipulation of the cluster. The genes from *tpaA* to *tpaD2* were selected for yeast-based cloning and two constructs were made for heterologous expression in *S. coelicolor* M1146: a wildtype version of the cluster and a refactored one. The wildtype construct included the intergenic regions preceding and following the precursor peptide gene with their putative native promoters, whereas the refactored one had intergenic region before *tpaA* replaced by a sequence UTR containing thiovarsolin promoter and RBS, as well as the strong constitutive promoter P_{hrdB} before the *tpaD1* gene.

Curiously, the heterologous expression of the BGC did not produce a fully mature thiopotensamide A, instead it produced thiopotensamide B, which did not seem to possess methylation near the N-terminus, dehydration on hdmHis residue and oxidation on the methionine residue, according to the analysis of its MS/MS spectra (Figure 2.24). It was interesting to find that the methionine was not oxidised, as was expected to be introduced by non-enzymatic activities (Liang et al., 2012, Frattaruolo et al., 2017). Production of thiopotensamide B from *S. coelicolor* M1146 peaked between day 8 and day 10 for both construct pLF015 (wildtype BGC) and pLF016 (refactored BGC). The refactored BGC was

found to have significant greater productivity in heterologous host comparing to the wildtype BGC. Hence, the refactored construct was used to investigate the missing modifications in thiopotensamide B.

The role of genes *tpaC* and *tpaD2* in thiopotensamide biosynthesis was investigated. Both genes were individually deleted by PCR targeting. Knockouts of either *tpaC* or *tpaD2* abolished production of thiopotensamide B, indicating that the products of both of these genes are essential for the biosynthesis of thiopotensamide B. This was further supported by the production of thiopotensamide B being restored by *in trans* complementation experiments (Figure 2.31). This result, alongside their overexpression in the wildtype background also ruled out their role in hdmHis dehydration. The potential role of a lantipeptide synthetase, PotKC was also investigated by overexpressing the gene under the control of P_{ermE*}, but this gene was also not responsible for the dehydration (Figure 2.32).

Experiments investigating the flanking genes of the cloned BGC to redefine its boundaries were performed. A monooxygenase gene near the BGC was speculated to oxidise the methionine residue but expressing the gene *in trans* in the heterologous expression strain did not show any oxidation being installed on thiopotensamide B. Although the enzyme did not appear to modify thiopotensamide B, it is still not clear whether this monooxygenase is just not involved in the pathway, or it requires partner proteins/pre-installed modifications for the oxidation.

Two methyltransferases located on either side of the BGC were also investigated to address the N-terminal methylation that was missing in thiopotensamide B. Expression of the downstream methyltransferase, now annotated as *tpaMT*, introduced a mass shift of +14 to thiopotensamide B, consistent with a methylation. The MS/MS fragmentation pattern

suggests this methyl group is also at the N-terminus as it was the case for the methyl group in thiopotensamide A. However, there is not sufficient data to precisely locate the methylation position in either of the two molecules.

To increase the production of thiopotensamide in the heterologous host, several strategies were applied in addition to the wildtype cluster and/or the refactoring of the cluster. First, two potential regulatory genes, *tpa+1* from *N. potens* and *TaaB* from the thioalbamide pathway were co-expressed with both clusters in *S. coelicolor*. The regulatory gene from thioalbamide pathway (*taaB*) appeared to improve the production from the wildtype cluster (Figure 2.39). Although the production in refactored cluster did not appear to be affected by either of the regulatory genes (*tpa+1* and *taaB*), the production was still significantly greater than the improved production in wildtype (wildtype cluster with overexpression of *taaB*). Meanwhile, an extra copy of the precursor peptide *tpaA* was introduced in wildtype of the overexpression construct. Unfortunately, neither of over expressing an extra copy of *tpaA* with the wildtype nor the refactored clusters led to increased production of thiopotensamide (Figure 2.40). However, there are a number of factors that might have affected the results, so future work should revisit these strategies.

2.5 Future work

2.5.1. Thiopotensamide characterisation

The project has come very close to the characterisation of thiopotensamide A but has been extremely challenged by the presence of kayamycin in the cultures, due to a combination of high production level of kayamycin, low production level of thiopotensamide A and similar chromatographic behaviours (Figure 2.15). Since kayamycin and its BGC are identified, more bespoke separation techniques that target kayamycin could be tested. One of the strategies was to break down kayamycin at the glycosidic bond between the narbonolide aglycon and the mycaminose sugar. It was previously reported that mycaminose could be released by acidic hydrolysis on magnamycin, a mycamino sugar containing antibiotic (Wagner et al., 1953). If the mycaminose could be removed from kayamycin, the narbonolide aglycon could potentially have a drastic change of behaviour in the employed purification systems. However, the stability tests in the project showed that treating the sample with acid also had negative impacts on thiopotensamide A. Hence, the experiment conditions might require adjusting. There is also further optimisation that could be attempted on some of the techniques already attempted, like the HPLC purification using a PFP column or other orthogonal stationary phases.

Genetic manipulation of *N. potens* was unsuccessful using common tools in the lab for *Streptomyces* sp., therefore to increase thiopotensamide production or disrupt kayamycin production in the native host, novel or compatible genetic editing tools need to be developed for *N. potens*. While obtaining purified thiopotensamide A from the native host *N. potens* might be challenging, heterologous expression has showed promising results. Despite the fact that heterologous expression of the BGC in *S. coelicolor* does not lead to

the production of thiopotensamide A, the cluster is partially functional and is capable of producing the intermediate thiopotensamide B. Moreover, co-expression of the cluster with the newly identified methyltransferase TpaMT leads to the production of thiopotensamide C. All these molecules have been thoroughly characterized by LCMS and MS/MS but NMR characterisation is still required to definitely confirm their structure. If purification of thiopotensamide A is not accomplished, the structures of thiopotensamide B and C can also provide valuable information, in particular characterisation of the structure of thiopotensamide C is an alternative option to locate exact position of the methylation at the N-terminus of thiopotensamide.

In addition, *S. coelicolor* cultures produce a much cleaner extract for purification, so heterologously expressing the BGC in *S. coelicolor* is a promising approach to obtain sufficient thiopotensamide for purification. Purification of thiopotensamide B is currently under way and will hopefully be a starting point for follow up structural work.

The putative hdmHis dehydration is still one important missing piece of the puzzle in this study. This novel modification makes thiopotensamide A unique to other characterised thioamitides. While experiments have been carried out to test potential candidates to catalyse it, the responsible tailoring enzymes remain uncertain. After confirming the role of *tpaC* and *tpaD2* in biosynthesis, the phosphotransferase-HopA1 pair of *tpaD1* and *tpaE* are now the focus of interest. It was expected that these genes were being expressed since genes downstream of them were clearly active (section 2.3.6.1), however based on the results it is possible that either *tpaD1* or *tpaE* are inactive. This scenario was observed in heterologous expression of the bottromycin cluster in *S. coelicolor*, where it was observed that the methylation that BtmC normally introduces was not installed in the pathway

products, even though its gene was at the beginning of the BGC operon (Eyles et al., 2018). Hence, knockout mutants of *tpaD1* and *tpaE* as well as overexpression of these genes could be carried out to monitor their role in production. The production of thiopotensamide B should remain unaffected in the knockouts and overexpression strains if the genes are inactive or not involved in the biosynthesis of thiopotensamide B. Given the difficulties in the purification of thiopotensamide A from *N. potens*, it would be important for the project to recover the dehydration in the heterologous expression host. While the mass shift in the fragmentation pattern between the MS/MS spectra from thiopotensamide A and B is consistent with a dehydrated hdmHis, NMR verification is the ultimate goal. Thus, if the hdmHis dehydration was recovered in the heterologous host, the resulting molecule could be purified to characterise the structure.

Once the dehydration and the responsible tailoring enzymes are identified, it is crucial to understand the modification mechanism. The presence of a hydroxylation in thiopotensamide B supports the proposed dehydration mediated elimination, but this ideally would be verified *in vitro* (Ke et al., 2011, Sikandar et al., 2022). One question to answer is whether the sterics of neighbouring amino acid sidechains influence the modification. For example, the position 11 of the thiopotensamide core peptide right before the histidine is occupied by an alanine. If the alanine were replaced with a tyrosine, which is bulkier, it would be interesting to observe if the hdmHis would still undergo E1cB dehydration. Prior to this experiment, the dehydration reaction needs to be restored in the heterologous host first. Although the MS/MS spectra from thiopotensamide A and B provided valuable data in support of the hdmHis dehydration hypothesis, the dehydrated

histidine can potentially be labelled with an *N*-acetylcysteamine (Guo et al., 2008). The labelled product would give a defined mass shift of fragments in MS/MS.

2.5.2. Generation of structural diversity in thioamitides

Besides structure elucidation, purification of the compounds is also essential for bioactivity assays. It has been proven that thioamitides possess anticancer activity and the potency varies between analogues (Izumikawa et al., 2015, Kjaerulff et al., 2017). Evaluation of thiopotensamide derivatives should provide excellent information about the importance of their PTMs for bioactivity. Three thiopotensamide derivatives were discovered in the study, and each of them had different levels of modification. By comparing their bioactivities, it should provide an overview on how the unconventional dehydration on the macrocycle and the methylation on the linear chain affect the bioactivity in terms of both potency and the cytotoxicity margin between cancer and healthy cells.

Another potential line of work is how to generate more thiopotensamide derivatives. In fact, this is applicable to all thioamitides, as the ultimate goal is to obtain optimised anticancer substances. One method that was briefly mentioned previously is to replace amino acid residues in the core peptide region, as the amino acid sidechains define the general structure of RiPPs, including thioamitides. This has already been done in the case of other RiPPs such as nosiheptide, thiocillin, thiostrepton and thiostreptamide S4 (Acker et al., 2009, Li et al., 2012, Wang et al., 2016b, Eyles et al., 2021). This could also potentially unlock a site for a particular modification, for example the tyrosine for *O*-methylation in thiostreptamide S4.

Another method to generate diversity involves genetically engineering the biosynthetic pathway by introducing genes from other pathways and/or removing existing genes. This

approach is called combinatorial biosynthesis and has been used in the case of thiostrepton A and nisin for example (Li et al., 2018, Schwalen et al., 2018). This can be challenging for complex RiPPs, as the interactions between the tailoring enzymes and the leader peptide need to be considered when designing experiments. Thus, understanding the interactions between tailoring enzymes and precursor peptide is key for thioamitide engineering.

Last but not least, the regulatory protein TaaB for thioalbamide biosynthesis preliminarily showed an increase production of thiopotensamide B in the *in trans* co-expression experiment. Although the increase was not statistically significant, increasing the number of biological replicas might increase the resolution of the experiment. Also, the regulatory protein is not native to the construct as they are from two different genera, *Amycolaoptsis* and *Nocardiopsis*. It would be intriguing to observe the production, when the upper region from *tpaA* region is replaced by the upstream intergenic regions both sides of *taaB*.

Chapter 3 *in vitro* study of YcaO-TfuA catalysed thioamidation

3.1. Introduction

One of the key structural features of thioamitides, responsible for the name of this RiPP class, is the presence of thioamide bonds in their N-terminal linear chain. Interestingly, each thioamitide contains multiple thioamides, and the numbers of thioamide bonds varies between molecules. For example, thioalbamide has 3 thioamide bonds, thioholgamide (also known as neothioviridamide), thiosparsoamide, thiostreptamide S4 and thiostrepatmide S87 have 4 and JBIR-140, thioviridamide and prethioviridamide have 5 each (Figure 3.1) (Hayakawa et al., 2006b, Izumikawa et al., 2015, Frattaruolo et al., 2017, Kjaerulff et al., 2017, Izawa et al., 2018, Kawahara et al., 2018, Lu et al., 2021). These are the only peptides in nature that feature multiple thioamides.

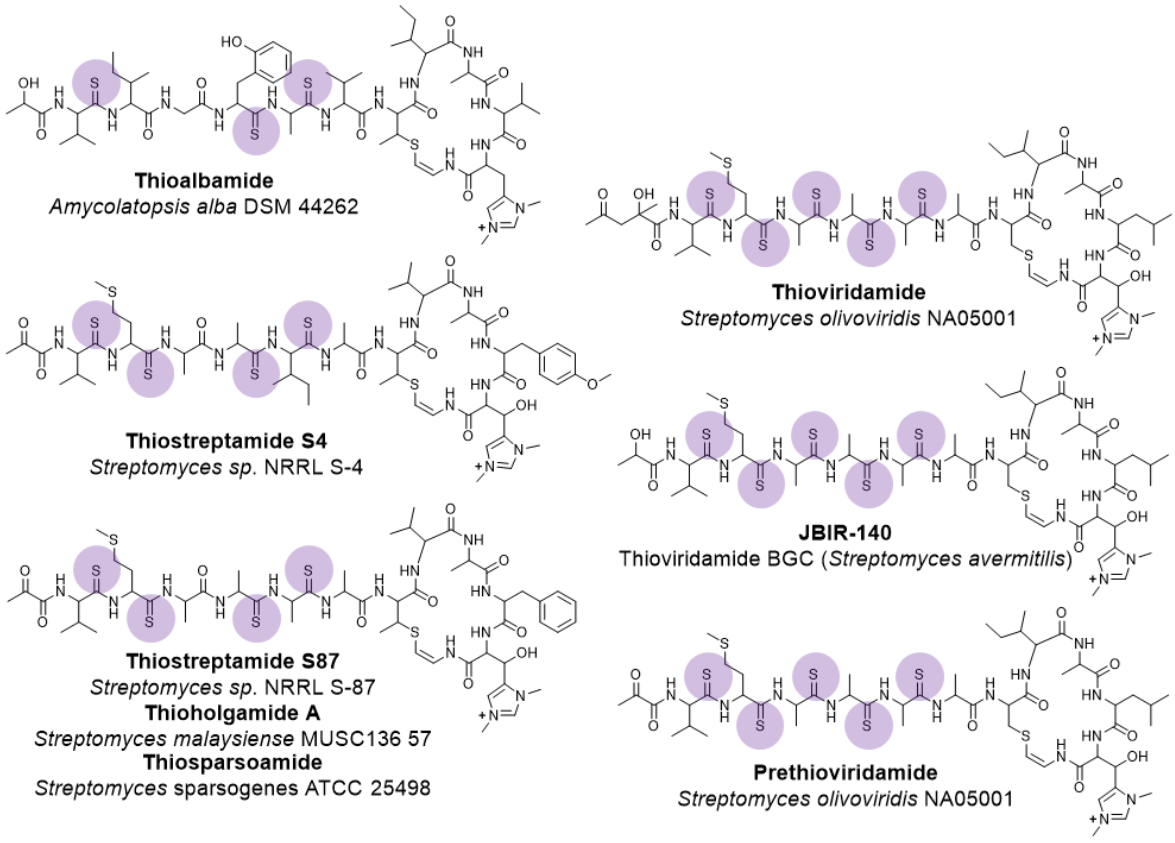


Figure 3.1: Characterised thioamitides known to-date with their thioamide bonds highlighted.

The elucidation of thioamide bond formation only began in the very recent years. Multiple thioamide bond formations in RiPPs were reported to be YcaO dependent. YcaO domain proteins are responsible for several other post-translational modifications (PTMs). For instance, azole heterocycle formation for thiazole/oxazole-modified microcin (TOMM) (such as microcin B17), and amidine formation for bottromycin and streptamidine (Li et al., 1996, Crone et al., 2016, Russell et al., 2021). In spite of the fact that variety modifications were catalysed by YcaO superfamily, the common function across YcaO domain proteins were claimed to utilise adenosine triphosphate (ATP) and potentially hydrolyse ATP (details in section 1.3.3) (Dunbar et al., 2012, Dunbar et al., 2014). For instance, the azole heterocycles in a TOMM from *Bacillus* sp. A1 Hakam was initiated by cyclodeydratase complex that was formed by YcaO protein and E1-like protein (Figure 3.2) (Dunbar et al., 2012).

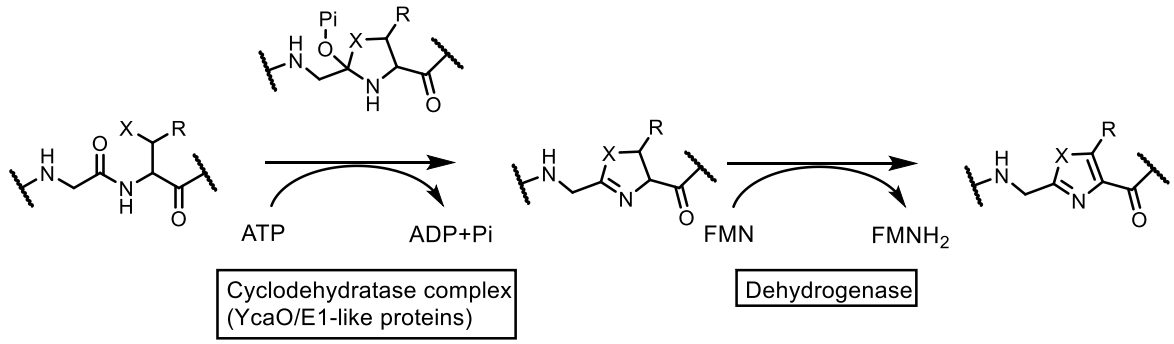


Figure 3.2: Role of YcaO protein in azole heterocycle formation in a TOMM biosynthetic pathway from *Bacillus* sp.. X=S, O; R= H, CH_3.

Following the discovery of thioviridamide, the YcaO and TfuA domain containing proteins in its biosynthetic gene cluster were speculated to be responsible for the thioamidation (Izawa et al., 2013). More recently, a YcaO and TfuA pair were confirmed to be involved in thioamidation of methyl-coenzyme M reductase, an enzyme from the methanogenic

archaeon *Methanosarcina acetivorans* through both genetic manipulation and *in vitro* assays (Nayak et al., 2017, Mahanta et al., 2018). This was further supported by the formation of a thioamidated thiostrepton when the YcaO-TfuA pair from saalfelduracin BGC from *Amycolatopsis saalfeldensis* was introduced into the thiostrepton expression host *Streptomyces laurentii* (Schwalen et al., 2018). Furthermore, YcaO and TfuA pair was highly conserved in all currently discovered thioamitide BGCs through genome-mining (Figure 3.3a), and genetic manipulation of a thioamitide BGC in a heterologous expression model in our lab right before this project was started also showed similar results (Eyles et al., 2021).

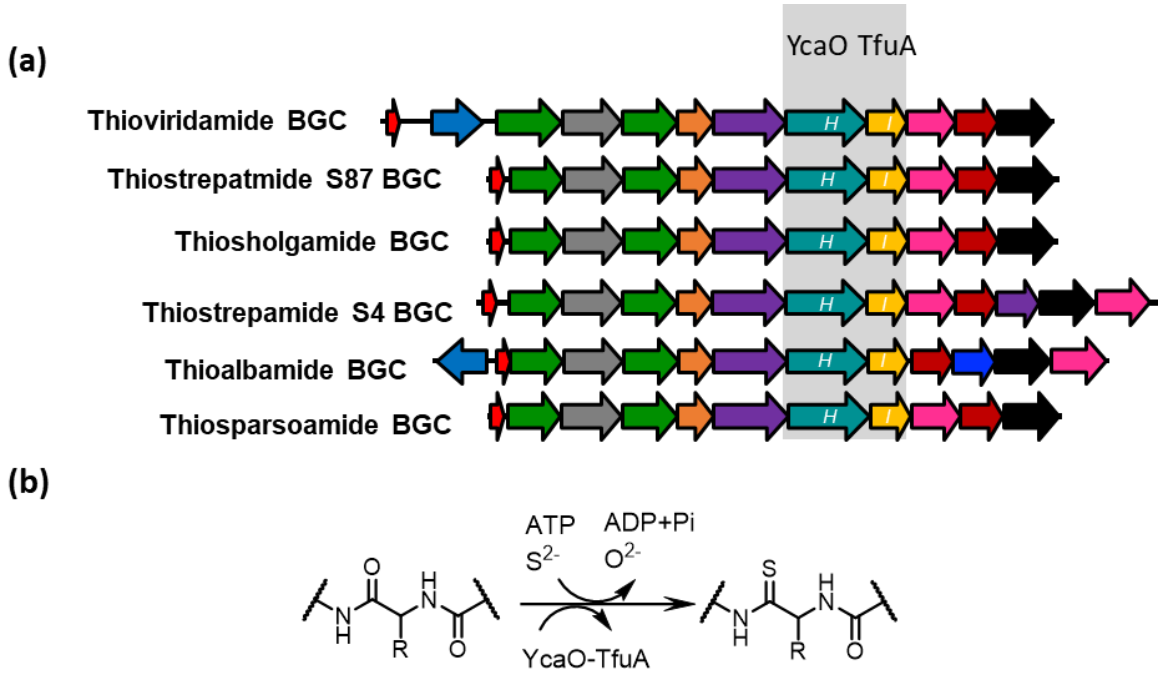


Figure 3.3: (a) An alignment of the characterised thioamitide BGCs with YcaO-TfuA pair highlighted (b) a schematic installation of thioamide bond by YcaO and TfuA proteins.

Confirming the role through genetic manipulation was a key discovery for understanding the biosynthesis of thioamitides, but it also generated curiosity about how this pair of genes leads to multiple modifications in multiple but defined amino acid residues and what is the exact role of these proteins in RiPP biosynthesis. Since the genetic manipulation study

was the only proof that those proteins had an essential role for the thioamidation, we aimed to reconstitute the reaction *in vitro* to provide a true understanding of the reaction, such as the co-factors involved and the role of the TfuA protein. At the time the project started, information about the sulphur source and the role of TfuA protein was extremely limited. Knocking out the corresponding gene in an *in vivo* system can provide clear evidence for the essential of the TfuA, but *in vitro* reconstitution offers opportunities to confirm whether and investigate how the biochemical, such as sulphur source and ATP, play their roles in YcaO-TfuA driven thioamidation. Similarly, although YcaO-TfuA pair were responsible for the thioamidation in methyl-coenzyme M reductase in *M. acetivorans* and a third protein in the reaction might not be likely, the possibility of other protein involvement is not certain. Hence, testing what exactly is required for the thioamide bond formation would be the most fundamental question before any protein-protein or protein-peptide interaction studies can be carried out.

3.2. Aims of the chapter

At the beginning of this project there was limited knowledge about thioamitide formation. Thioamidation was proposed to be the first modification in the biosynthesis of thioamitides based on genetic data, so it was crucial to elucidate the reaction *in vitro* (Franz et al., 2021). Although it was confirmed that YcaO-domain and TfuA-domain proteins were essential in the biosynthesis by engineering the biosynthetic pathway *in vivo*, the components involved in the reaction could only be confirmed *in vitro* so any essential co-factors, such as ATP and source of sulphur, could also be identified.

The main objectives of this chapter were to test the hypothesis that thioamidation on a thioamitide precursor peptide requires YcaO and TfuA proteins from the biosynthetic pathway by observing thioamide bonds being installed on the substrate peptide, as well as confirm that the required cofactors were just ATP and a sulphur source. In order to achieve this, there were 3 specific objectives of this chapter:

1) Clone each of the candidate genes required for thioamidation from a thioamitide BGC into a suitable protein expression vector;
2) Heterologous expression of the genes and optimisation of expression conditions to obtain soluble YcaO and TfuA proteins and precursor peptide;
3) Perform and analyse *in vitro* thioamidation reactions with the purified proteins and precursor peptide.

3.3. Results and discussion

3.3.1. Selecting candidate proteins for expression

In order to perform *in vitro* assays, the participating proteins or peptides of interest need to be expressed in a soluble form that can be purified for enzymatic assays. Unfortunately, YcaO-domain proteins had previously been reported to be sparingly soluble, so this could potentially be a challenge in this study (Franz et al., 2017). One strategy to avoid this was to test proteins from different species with the expectation that amino acid sequence variation between them would contribute differences in solubility. While a previous genome-mining approach from the lab had identified 14 strains (to-date, 18 strains) possessing a thioamitide BGC, a subset of them were selected for expression based on sequence diversity and access to the source strain (Frattaruolo et al., 2017). Hence, six strains were selected for the study, *Amycolatopsis alba*, *Micromonospora eburnea*, *Nonomuraea fuscirosea*, *Nocardiopsis potens*, *Streptomyces* sp. NRRL S-4 and *Streptomyces* sp. NRRL S-87 (Figure 3.4, Appendix 17).

3.3.2. Tackling the solubility problem of YcaO-domain proteins

3.3.2.1. Expression with a polyhistidine tag

The *ycaO* genes were amplified from gDNA and cloned into pET28a (Table 3.1), so a polyhistidine tag (His_6-tag) could be introduced to the N-termini of the proteins to support their isolation. Despite the fact that cloning into pET28a allows the introduction of the His_6-tag to either C-terminus or N-terminus of the protein, it was previously reported that the C-terminus of YcaO-domain protein, BalhD, played an important role in protein activity in *Bacillus sp* for cyclodehydration in TOMM (Dunbar et al., 2014). Dunbar *et al* revealed that when introducing mutations to the C-terminus of BalhD, the protein decreased or lost its

catalytic activity *in vitro*. Introduction the His_6-tag to the N-terminus of the protein was less likely to affect the activity of the enzyme. pET28a also provides a thrombin cleavage site was between the protein and the tag, so the His_6-tag could be removed by thrombin cleavage.

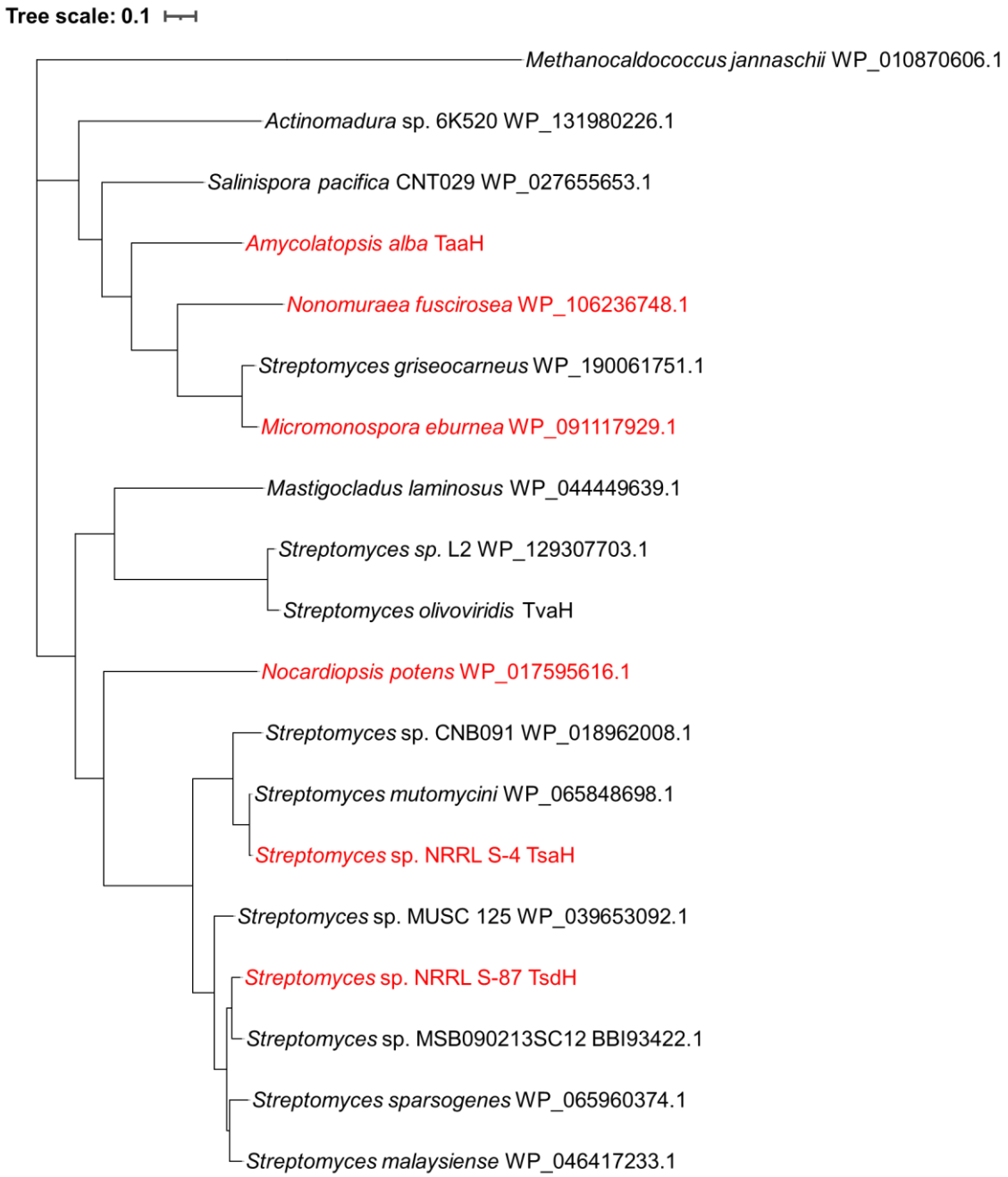


Figure 3.4: Maximum likelihood phylogenetic tree of YcaO proteins from thioamitide BGCs. YcaO protein from *Methanocaldococcus jannaschii*, not belonging to a thioamitide BGC was used as the outgroup for the tree, YcaO proteins selected for the project are labelled in red.

Table 3.1: Details of His_6-tagged YcaO proteins selected for overexpression in BL21(DE3). Abbreviations: Da (Dalton), pI (isoelectric point). *Names of YcaO protein were taken from published data (TpaH was from the previous chapter). **Molecular mass was calculated with His tag, *** Mass of His_6 tagged peptide when no gene is in pET28a.

YcaO protein source	Expression vector name	Accession Number/Protein name*	Molecular mass** (Da)	Theoretical pI
A. alba	pET_YcaO_AA	WP_020631615.1/TaaH	51282	6.02
M. eburnea	pET_YcaO_ME	WP_091117929.1	51125	6.41
N. fuscirosea	pET_YcaO_NF	WP_106236748.1	51215	6.65
N. potens	pET_YcaO_NP	WP_017595616.1/TpaH	51563	5.94
S. sp. NRRL S-4	pET28a_S4_YcaO	WP_031090932.1/TsaH	51781	6.10
S. sp. NRRL S-87	pET_YcaO_S87	WP_030193566.1/TsdH	50975	6.75
N/A	pET28a	N/A	5877***	11.60

Along with an empty pET28a vector as a negative control, the pET28a-based constructs were introduced to *E. coli* BL21(DE3), which enables overexpression the proteins of interest after expression is induced by the addition of IPTG. Comparison of the overexpression constructs with the negative control showed that the His_6-YcaO homologues were successfully expressed, and the expected bands of approximately 51 kDa given by the proteins could be immediately identified on an SDS-PAGE gel (Figure 3.5). Small scale solubility tests were carried out lysing the cells and separating the soluble and insoluble fractions by centrifugation. This was a quick test to check if the selected YcaO proteins from thioamitide biosynthetic pathways had better solubility than the YcaO protein in bottromycin biosynthetic pathway, which was reported to be challenging to obtain a functional and soluble protein (Franz et al., 2017). However, all six His_6-YcaO homologues appeared in the pellet fraction, which meant they were not soluble in the conditions tested (Figure 3.5). In case a small fraction of the His_6-YcaO homologues were soluble but could not be observed due to background from *E. coli* BL21(DE3), a larger scale protein expression was carried out, but the SDS-PAGE results did not show any potential soluble His_6-YcaO proteins.

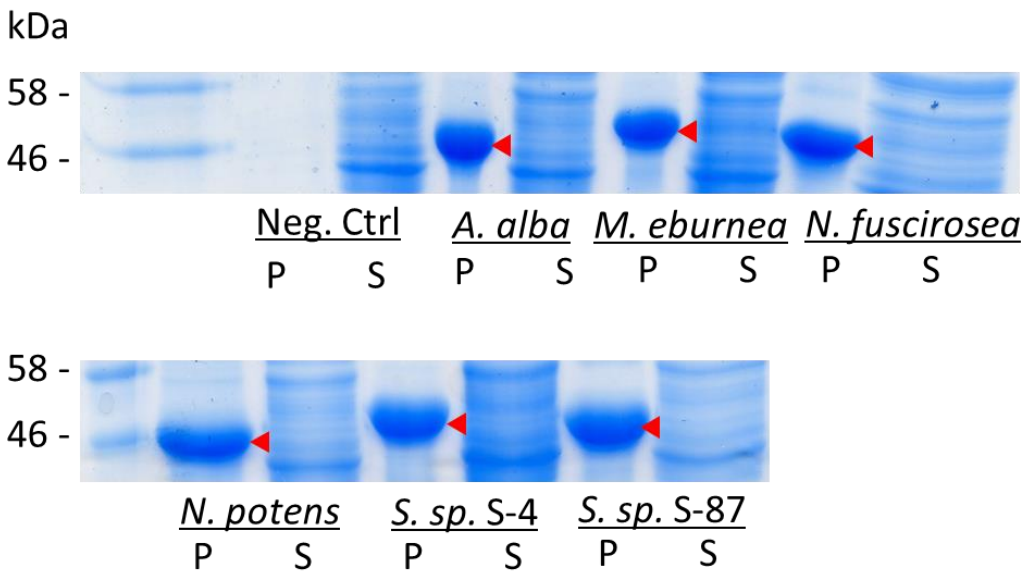


Figure 3.5: SDS-PAGE analysis of the His_6-YcaO proteins solubility. The His_6-YcaO proteins are indicated by the red triangles. Each construct had 2 samples taken for the analysis, a sample from the pellet (P) and a sample from supernatant (S). The constructs were pET28a (Neg. Ctrl), pET_YcaO_AA (*A. alba*), pET_YcaO_ME (*M. eburnea*), pET_YcaO_NF (*N. fuscirosea*), pET_YcaO_NP (*N. potens*), pET28a_S4_YcaO (*S.* sp. S-4) and pET_YcaO_S87 (*S.* sp. S-87).

3.3.2.2. Expression with a small ubiquitin-like modifier tag

In addition to assisting with the isolation of target proteins, fusion tags can also help solubilise proteins. For example, fusion with a maltose binding protein (MBP) tag was able to solubilise BalhD from *Bacillus* sp. (Dunbar et al., 2014). However, there were also reports suggesting that small ubiquitin-like modifier (SUMO) tag had better performance than MBP tag. For example, two difficult-to-express proteins, matrix metalloprotease-13 and myostatin, had greater expression level when they were fused to SUMO than to MBP (Marblestone et al., 2006). Also, a mammalian antibody fused with MBP tag was reported to have reduced activity in comparison to that the same antibody was fused to a SUMO tag (Liu et al., 2018). Meanwhile, SUMO tag was reported to successfully solubilise insoluble proteins with improved expression (Marblestone et al., 2006, Guerrero et al., 2015).

pTB146 is a plasmid for introducing a His_6-SUMO tag to the N-terminus of the protein (Guerrero et al., 2015). Hence, pTB146 used in this study.

Table 3.2: Plasmid details of His_6-SUMO-YcaO proteins selected for overexpression. Abbreviations: Da (Dalton), pI (isoelectric point). *Names of YcaO protein were taken from published data (TpaH was from the previous chapter). **Molecular masses were calculated with His_6-SUMO tag, *** Mass of His_6-SUMO tagged peptide when no gene is cloned into pTB146.

YcaO protein source	Constructs in BL21 (DE3)	Accession Number/Protein name*	Molecular mass** (Da)	Theoretical pI
A. alba	pLF001	WP_020631615.1/TaaH	62760	5.98
M. eburnea	pLF002	WP_091117929.1	62038	6.01
N. fuscirosea	pLF003	WP_106236748.1	62128	6.24
N. potens	pLF004	WP_017595616.1/TpaH	62475	5.67
S. sp. NRRL S-4	pLF005	WP_031090932.1/TsaH	62694	5.76
N/A	pTB146	N/A	15459***	8.64

The selected genes were cloned into pTB146 (Table 3.2), and along with the empty vector, pTB146, the expression test on *E. coli* BL21(DE3) showed that the tagged proteins had high level of expression. The solubility tests also showed pLF005 might have a small fraction of soluble protein as a gel band appeared at the appropriate level from the supernatant sample, but none of the other showed the same results. Hence, pLF005 was taken forward to perform a purification test. Figure 3.6 shows the comparison of the purified fractions from pLF005 expression and the negative control (pTB146). An extra gel band aligned with 72k Da marker on all fractions from pLF005, but not on the negative control. The gel band appeared about 10k Da greater than the calculated mass (62k Da). This unusual scenario could have been due insufficient unfolding of the protein, insufficient binding of SDS to the protein, and/or a combination of the two (Tiwari et al., 2019). Since this band did not appear on the negative control, it was proposed to be His_6-SUMO-YcaO protein.

Although qualitative proteomic analysis was required to confirm the identity of protein and the majority of the expressed protein still appeared to be in the pellet fraction, the protein was purified using nickel-affinity chromatography. The SDS-PAGE result showed that a small portion of the His_6-SUMO-TsaH from construct pLF005 was soluble and purified. This result was very promising for the downstream *in vitro* experiments. However, Figure 3.6 also showed His_6-SUMO-TsaH was not the only protein being isolated, there were still some background proteins being extracted by the nickel (II) (Ni(II)) resin. The reason that background proteins were present in the eluate was that the low solubility of tagged protein leaved unoccupied resin available for those naturally histidine-rich proteins from *E. coli* (Andersen et al., 2013). Improving the protein solubility could be a solution to the issues of quality and quantity.

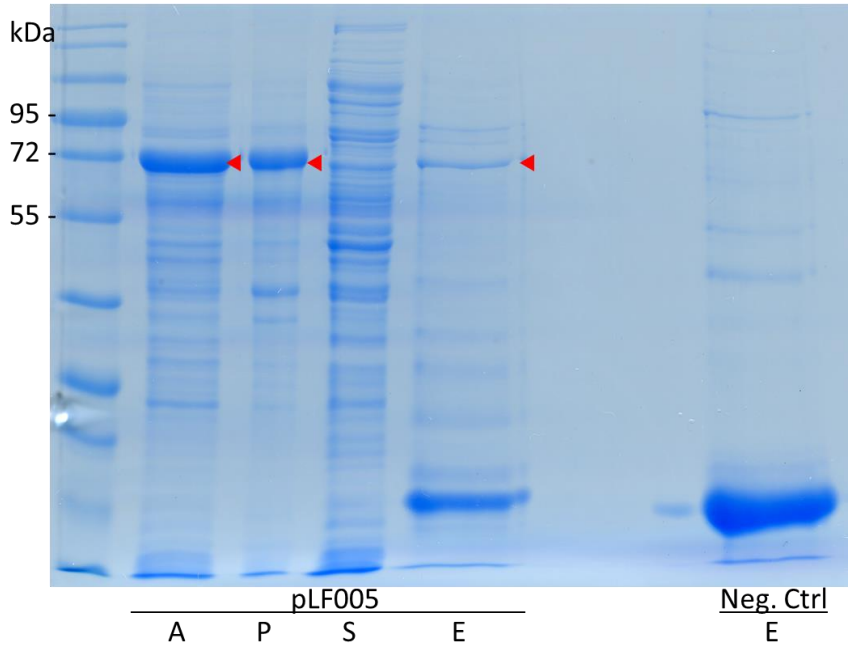


Figure 3.6: SDS-PAGE analysis of His_6-SUMO-TsaH purification. His_6-SUMO-TsaH (pLF005) was expressed in *E. coli* BL21 and purified with Ni(II) resin. The protein gel bands of interest are indicated by red triangles. Samples on the gel are (from left to right) total lysate (A), pellet fraction (P), soluble fraction (S), purified eluate (1 M imidazole) after Ni(II) resin purification (E) and the purified eluate of pTB146/BL21 (E).

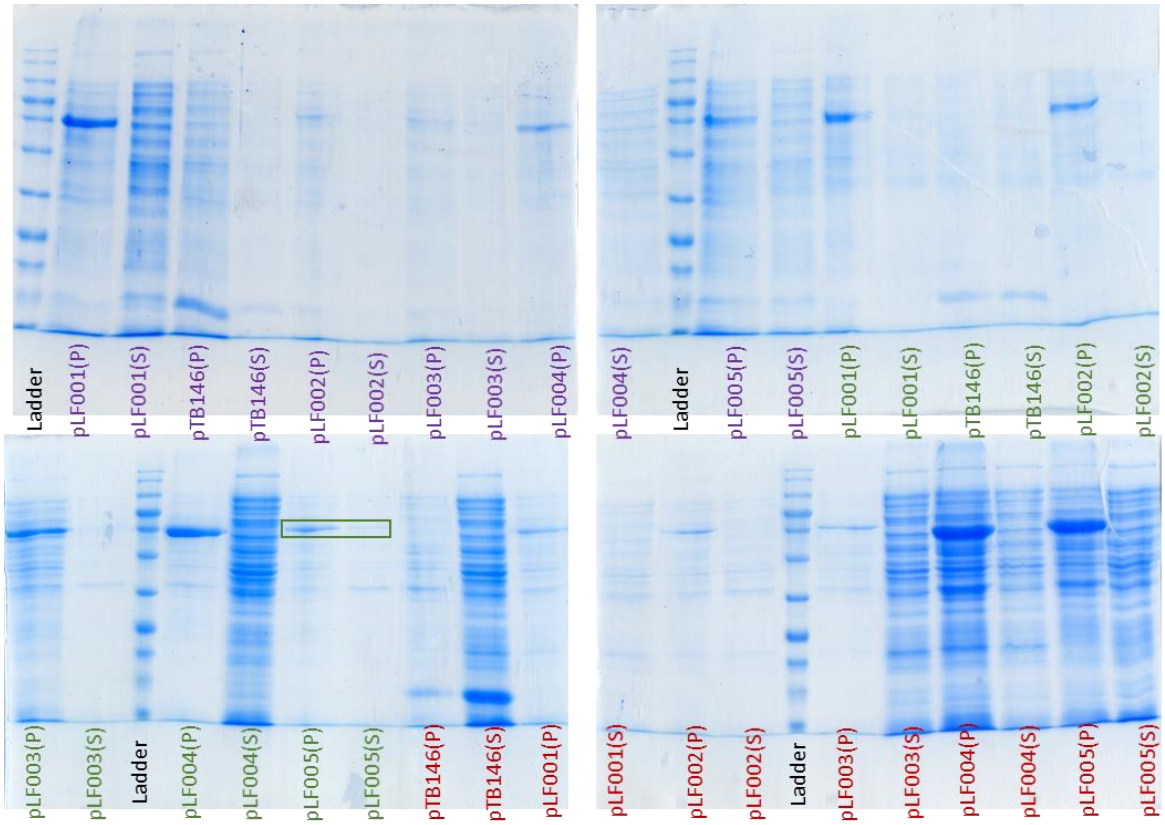


Figure 3.7: SDS-PAGE gel images from solubility tests showing the insoluble fraction (P) and soluble fraction (S) of a lysate. The SDS-PAGE images were obtained from expression of 5 YcaO homologues and a negative control (pTB146). Proteins were expressed by SoluBL21™ (purple labels), BL21 (green labels) and Rosetta™ (red labels). A very small portion of the YcaO protein from *S.* sp NRRL S-4 was seen in the soluble fraction (highlighted in green).

Since these preliminary results showed that the YcaO-domain protein from pLF005 (origin: *S.* sp. NRRL S-4) had a potential for purification, follow up work focused on this protein (TsaH) rather than other YcaO proteins. The first idea to enhance the protein solubility was to use different expression strains optimised for this purpose. A potential issue was the codon usage in the host cell. *E. coli*, as well as other species, has a codon bias which leads to an imbalanced population of mRNA molecules and cognate tRNA (Kane, 1995). However, all 6 genes were directly taken from actinomycetes, which have high GC content and different codon bias from *E. coli*. Therefore, *E. coli* might struggle to properly translate the high GC content genes due to codon bias, and this might prevent proper protein folding

(Rosano and Ceccarelli, 2009). *E. coli* Rosetta^{TM}(DE3) has improved expression of rare codons in *E. coli* thanks to a plasmid carrying the corresponding tRNAs (Tegel et al., 2010). Another alternative expression strain is *E. coli* SoluBL21^{TM}(DE3), a mutant strain of *E. coli* BL21(DE3). This strain has been reported to successfully produce soluble proteins that were insoluble in regular *E. coli* BL21(DE3) (Deatherage et al., 2012). Both *E. coli* SoluBL21^{TM}(DE3) and Rosetta^{TM}(DE3) were transformed with the pTB146-based constructs, and expression and solubility tests were carried out. However, these tests (Figure 3.7) showed that the new hosts did not improve solubility and *E. coli* BL21(DE3) was still the best strain to use for His_6-SUMO-TsaH expression from pLF005.

Another strategy assessed for improved protein production was codon-optimising the gene for *E. coli* expression. Changing the nucleotide sequence to generate synonymous codons has been found to alter both the expression level of the protein recombinants, as well as their solubility (Cortazzo et al., 2002). Substituting the wildtype with synonymous codons for specific amino acids in EgFABP1 (*Echinococcus granulosus* fatty acid binding protein1) altered the protein solubility (Cortazzo et al., 2002). The *ycaO* gene *tsaH* from *S.* sp. NRRL S-4 was optimised for *E. coli* expression and synthesised by Twist Bioscience. The optimised gene was cloned into pTB146 to form construct pLF006 and expressed in *E. coli* BL21(DE3), as this was the best expression strain for soluble His_6-SUMO-TsaH from pLF005. Both constructs pLF005 and pLF006 expressed the same His_6-SUMO-TsaH, and the protein from construct pLF005 was previously shown to be soluble, so a purification experiment was conducted for the proteins from both constructs in parallel. The results from purifying the proteins from both constructs showed that soluble His_6-SUMO-TsaH from pLF006 was

obtained too, but there was no visible improvement in the amount of soluble protein (Figure 3.8).

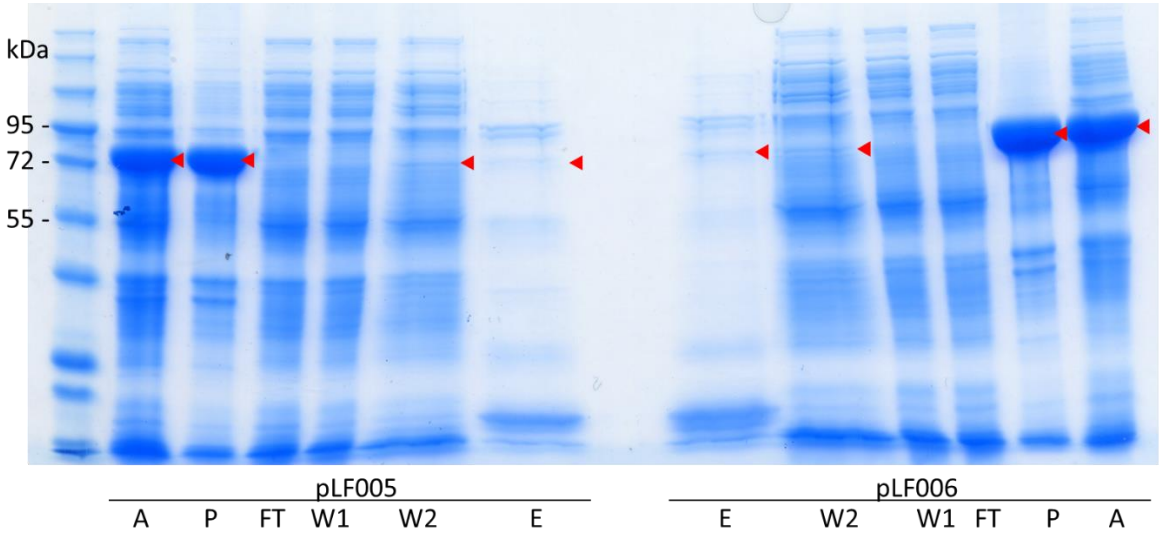


Figure 3.8: SDS-PAGE for purification fractions from expression of pLF005 and pLF006. The red triangles indicate the His_6-SUMO-TsaH in each gel sample. Abbreviations: total lysate content (A), pellet fraction (P), Flowthrough of the soluble fraction of the lysate (FT), Wash fraction with 5 mL 5 mM imidazole (W1), further wash fraction with 5 mL 5 mM imidazole (W2), eluate fraction with 5 mL 1 M imidazole (E).

3.3.2.3. Co-expression with partner protein

Another strategy to help solubilise TsaH is to co-express the protein with its interacting partners. Overexpression of recombinant proteins without an appropriate binding partner could suffer from misfolding leading to inclusion body formation and low production yield (Sørensen and Mortensen, 2005, Sadeghian-Rizi et al., 2019). For example, co-expression of bacteriophage T4 gene *gp31* with *gp23* increased the soluble expression product of *gp23* through assisting gp23 protein folding correctly (Kurochkina and Mesyanzhinov, 1999). Since TsaH and TsaI proteins were playing an essential role in thioamidation, it was reasonable to assess whether the YcaO protein becomes more soluble when it is co-expressed with the TfuA protein from the same pathway. If the YcaO protein became more soluble, it would indicate that the partner protein also played a role in regulating the folding

of the YcaO protein. Parallel work on the TfuA-domain protein (TsaI) from *S.* sp. NRRL S-4 showed that it was soluble when it was expressed on its own (see section 3.3.3). For the co-purification experiment, *tsaH* and *tsaI* were expressed in two separate inducible plasmids, pTB146 and pET28a backbone respectively, via *E. coli* BL21(DE3). The purification of the expressed proteins could be isolated by using Ni(II) resin, as both expressed proteins will have a His_6 tag attached to their N-termini. In this case, a larger volume of Ni(II) resin (4 mL slurry instead of 1 mL) was applied to ensure unoccupied Ni(II) resin was available for both proteins in the experiment. Although previously mentioned that excessive resin introduced undesired proteins to the purification mix, the nature of the trial was to ensure the resins would not be occupied by His_6-TsaI, which was found much more soluble than His_6-SUMO-TsaH. Both His_6-TsaI and His_6-SUMO-TsaH proteins were expressed and present the total lysate indicating that both proteins were expressed (Figure 3.9). However, TsaI did not increase the solubility of TsaH. Instead, TsaI protein that was soluble when expressed by itself (section 3.3.3) was not soluble in the co-expression experiment.

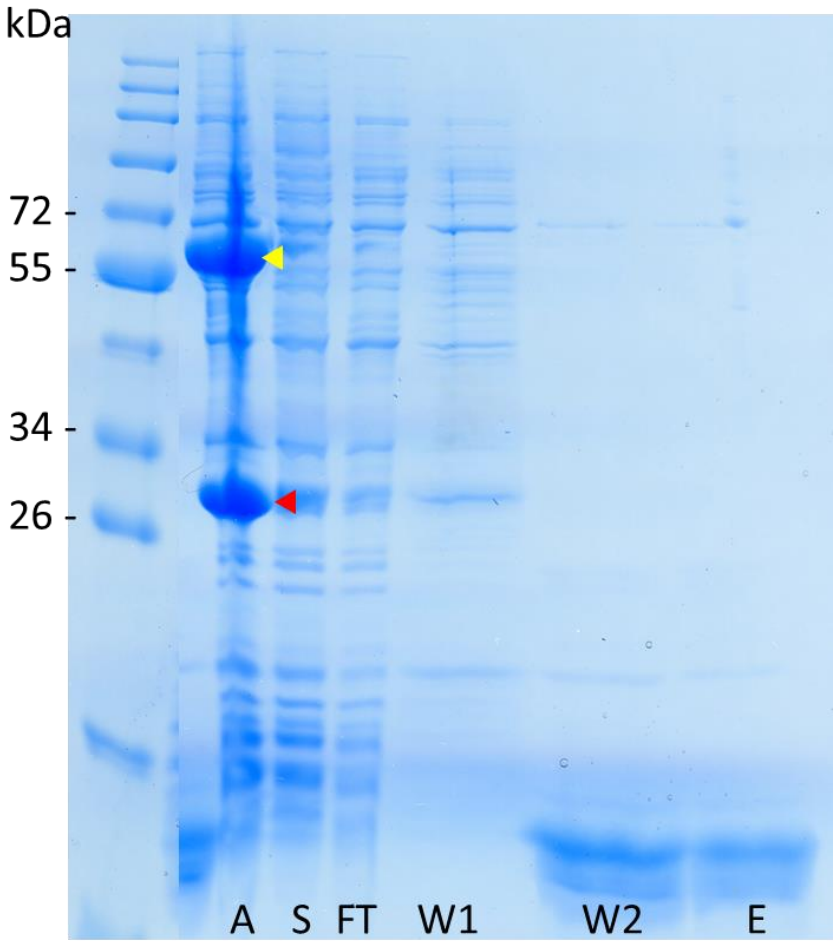


Figure 3.9: SDS-PAGE analysis of TsaI and TsaH co-expression and purification. The yellow triangle indicates the His_6-SUMO-TsaH and red triangle indicates the His_6-TsaI in each gel sample. Abbreviations: total lysate (A), Soluble fraction (S), Flowthrough of the soluble fraction of the lysate (FT), Wash fraction with 20 mL 5 mM imidazole (W1), further wash fraction with 20 mL 5 mM imidazole (W2), eluate fraction with 15 mL 1 M imidazole (E).

3.3.3. Purification of thioamidation proteins and substrate peptide

After the solubility tests, *E. coli* BL21(DE3)/pLF005 was used for upscale expression of TsaH. Purification was first carried out from 6 L of culture for purification by Ni(II) resins. A small sample of each purification step was taken for SDS-PAGE analysis along with the whole cell lysate. The most intense gel band from the eluate sample was cut for proteomics confirmation of the identity of the protein. Identification of protein are normally achieved by either bottom-up or top-down proteomics. In bottom-up proteomic analysis, the protein

would be firstly digested into peptides that are often readily ionised, so fragments are more predictable for protein identification (Dupree et al., 2020). On the other hand, top-down proteomics is performed without the digestion step, and the peptide fragments of the protein are obtained by fragmentation in the mass analyser, so the connectivity information is preserved (Zhu et al., 2017).

Bottom-up proteomic analysis was used for the identification and the results showed that the protein present in the gel band provided the highest score for His_6-SUMO-TsaH, against the *E. coli* protein database and His_6-SUMO-TsaH. Although only about 48% of the sequence was identified, there were between 23 and 24 unique peptides seen in the spectra, which was significant enough to confirm the presence of the protein (Figure 3.10b). The His_6-SUMO-TsaH obtained from a reduced volume of Ni(II) resins (6 litre culture with 1 mL Ni(II) resin slurry) was concentrated and further purified on a LH200 size exclusion column to obtain a purer protein sample.

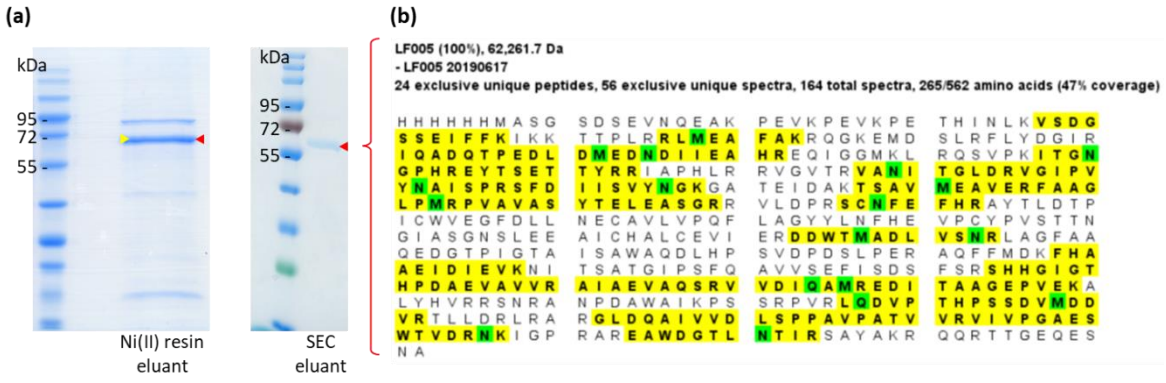


Figure 3.10: (a) SDS-PAGE analysis demonstrated that purity of His_6-SUMO-TsaH from only Ni(II) resin (left) and further purified with size exclusion column (right). The protein gel bands indicated in red correspond to the target protein, and indicated in yellow is the sample analysed by bottom-up proteomics. (b) Bottom-up proteomics result of His_6-SUMO-TsaH visualised with the software Scaffold. The peptides identified are highlighted in yellow, and the amino acids with PTMs are highlighted in green

Table 3.3: Plasmid details of constructs expressed in BL21(DE3) used for protein extraction. Molecular masses in the table included the tag. Abbreviations: Da (Dalton)

Constructs in BL21 (DE3)	Protein/ peptide	Accession Number	Molecular mass (Da)	N-terminal protein tag
pET28a_S4_TsaA	TsaA	WP_107418454.1	11405	His_6 tag
pLF005	TsaH	WP_031090932.1	62720	His_6-SUMO tag
pET28a_S4_TsaI	TsaI	WP_051834498.1	28294	His_6 tag

The precursor peptide gene (*tsaA*) and TfuA-domain gene (*tsaI*) from the thiostreptamide S4 BGC had previously been cloned into pET28a to express each protein with a His_6-tag at their N-termini. Solubility tests showed that majority of the expressed protein and peptide in *E. coli* BL21(DE3) were soluble. Hence, large-scale expression and purification was performed for both proteins. The purified fractions were much cleaner than His_6-SUMO-TsaH. The SDS-PAGE gel bands were excised and analysed by bottom-up proteomics. The results confirmed that the gel bands corresponded to TsaA and TsaI.

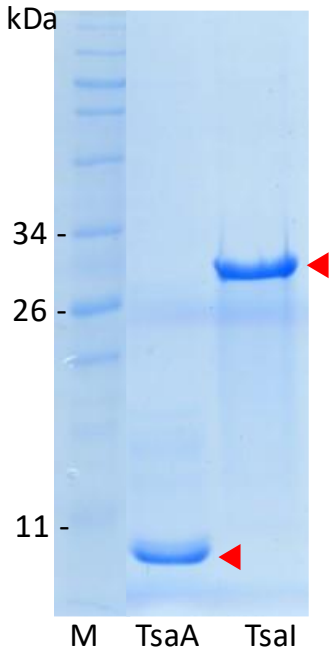


Figure 3.11: SDS-PAGE analysis demonstrated that purity of His_6-TsaA and His_6-TsaI protein from only Ni(II) resin. M is the protein ladder. The protein gel bands indicated in red correspond to the target protein and the gel bands were sent for bottom-up proteomic analysis.

In summary, both the precursor peptide TsaA and the TsaI could be easily expressed and purified using Ni(II) resin with His_6-tag and were confirmed by proteomic analysis. On the other hand, only a very small fraction of TsaH with His_6-SUMO tag was soluble, and a second step of purification was required to remove background proteins. A few approaches were taken to increase the solubility of TsaH, including testing various expression strains, codon optimisation and co-expression, but none of those seemed to improve the solubility of the protein.

3.3.4. Top-down MS analysis on biological materials

3.3.4.1. Top-down proteomic MS pattern from His_6-TsaA

In order to monitor the *in vitro* thioamidation assays and determine if or where the modification had taken place, top-down MS was used in the following experiments. "Top-down" mass spectrometry is an emerging and powerful analytical tool for identifying PTMs on a peptide/protein of interest. The key difference between bottom-up and top-down mass spectrometry is that the latter one does not require trypsin digestion prior to the analysis and instead uses intact protein. The main benefit of this approach is that intact protein mass preserves all the information from primary structure and the modifications of the protein, whereas the bottom-up analysis may lose fragments of proteins and these fragments might contain the PTMs of interest (Chait, 2006, Zhang et al., 2011).

A sample of purified His_6-TsaA was subjected to intact mass analysis. The MS spectrum showed 4 deconvoluted molecular masses, which were 10343, 10521, 11273 and 11451 Da (Figure 3.12). According to the predicted protein mass in Table 3.3, the tagged TsaA should have 108 amino acid residues and a molecular weight of 11405 Da. None of the masses on the spectrum matched this predicted mass. However, it has been reported that the N-

terminal methionine that initiates protein synthesis is often co-translationally removed by enzyme methionine aminopeptidase (Wingfield, 2017). When the peptide mass was calculated without the methionine residue, the molecular mass was then expected to be 11273 Da. In this case, one of the observed masses (11272.8 Da) and the calculated mass (11273.4 Da) had a mass difference of 0.6 Da, which was considered to be an insignificant mass difference (Kou et al., 2014).

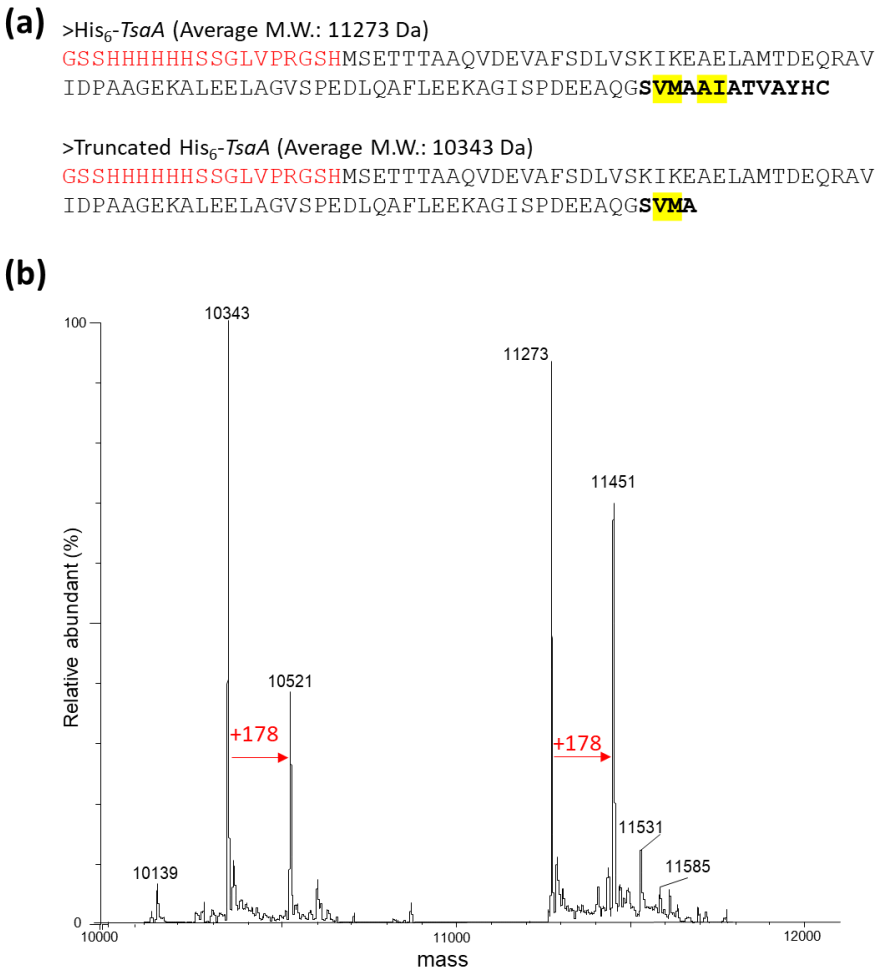


Figure 3.12: (a) Amino acid sequence of the detected full TsaA peptide (top) and its truncated version (bottom) with the calculated peptide mass. His_6 tag region is displayed in red, the core peptide region is displayed in bold, and amino acids where thioamide bonds are expected to be installed are highlighted in yellow. (b) Deconvoluted spectrum of His_6-TsaA. Four main molecular ions were detected 10343, 10521, 11273 and 11451, corresponding to truncated His_6-TsaA, truncated His_6-TsaA with gluconolyation, intact His_6-TsaA and intact His_6-TsaA with gluconoylation, respectively. Mass difference of 178 Da indicated in red corresponds to gluconolyation.

The deconvoluted peaks of 11451 and 11273 Da had a mass difference of 178 Da between them, which was also observed between 10521 and 10343. This addition of 178 Da was proposed to be gluconoylation (Figure 3.12). It has been reported that protein expression strain *E. coli* BL21(DE3) is likely to accumulate 6-phosphogluconolactone, which subsequently promotes gluconoylation of recombinant proteins, due to the lack of 6-phosphogluconolactonase (Meier et al., 2012, Schweida et al., 2019). Another study also claimed that the His_6-tag, which has an N-terminal sequence of (M)GSSHHHH, is more reactive to be gluconoylated at the N-terminus (Geoghegan et al., 1999). This suggested the two peaks with smaller mass were given by TsaA, but with part of the C-terminal core peptide missing, as the presence of a gluconoylated version indicated the N-terminal His_6-tag was still attached. Indeed, after excluding 9 C-terminal residues (AIATVAYHC), the predicted mass was 10343 Da, which was again consistent with one of the observed masses (Figure 3.12). Although the truncated His_6-TsaA had part of the core peptide missing, there were still two amino acid residues that could be thioamidated according to the final product. Therefore, the truncated peptide might still be able to function as a substrate for thioamide bond installations.

Summarising all the information, the deconvoluted peak at 11273 corresponded to intact His_6-TsaA, and the other 3 peaks were modifications of His_6-TsaA, corresponding to gluconoylated intact peptide, truncated His_6-TsaA and its gluconoylated version.

3.3.4.2. TsaA stability investigations

The presence of truncated TsaA after purification indicated one potential challenge of the *in vitro* assay was the precursor peptide stability. Whether the truncated peptides behaved the same way as the full-length precursor peptide was not clear. If the sites for

thioamidation were too close to the C-terminus for efficient modification, not only did the degradation reduce the amounts of available substrates for thioamidation, but it also introduced undesired materials to the spectra. It has been reported that the signal-to-noise ratio (S/N) decreases exponentially with an increase of molecular mass, so when a peptide appears in different forms, it can weaken the signal from the analyte (Gregorich and Ge, 2014). If the signal given by the thioamidated peptide was low, it might not appear in the deconvoluted spectrum. As these peaks appeared together in our analysis, it was vital to tackle the degradation of the protein.

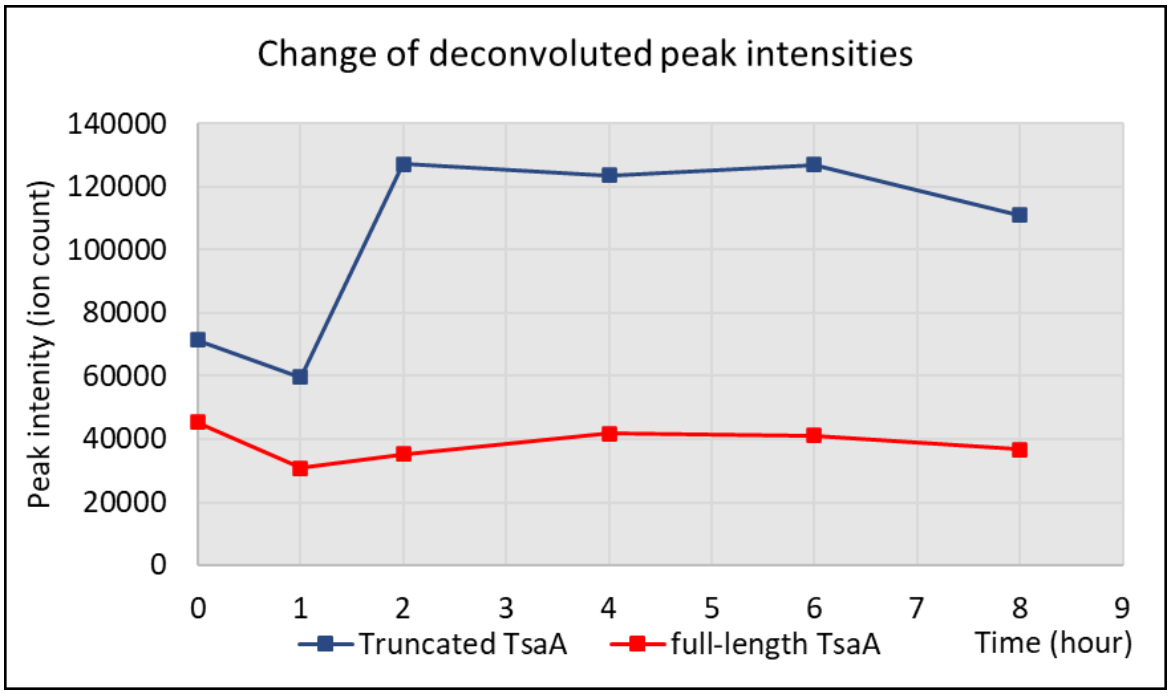


Figure 3.13: Change of deconvoluted peak intensities over time. For the truncated version of the precursor peptide (blue line) and the intact precursor peptide (orange).

A stability test was carried out to check when the peptide was degraded over time in the reaction conditions. Since the *in vitro* assay was carried out throughout a relatively long time (up to 8 hours) at 30 ^{o}C, it would require that the peptide would remain stable throughout the period. For the stability test, six aliquots of freshly prepared His_6-TsaA were incubated for up to 8 hours under the *in vitro* assay conditions without any tailoring

enzymes. One aliquot of the test sample was taken to -20 ^{o}C at 0, 1, 2, 4, 6 and 8 hours. All samples were desalted and stored at -20 ^{o}C until intact mass analysis. The peak intensities of the full-length peptide mass 11273 Da were very similar. Figure 3.13 shows the peak intensity changes from sample to sample. The trivial fluctuation amongst samples revealed that the peptide was stable in the reaction buffer for at least 8 hours, so the purified His_6-TsaA was suitable for *in vitro* assay.

Peptides are often readily enzymatically and/or chemically degraded *in vivo* compared to other biologics (Evans et al., 2020). This could be even more rapid if the peptide was expressed in a heterologous host as the peptide was not native to the expression host. The fact that truncated peptide could be observed in the Ni(II) resin eluate already indicated that the peptide was degraded either during the purification process, or more likely have happened *in vivo* even before purification. It was possible that the degradation of the peptide began immediately after translation. His_6-TsaA had been expressed in the same manner as His_6-TsaI and His_6-SUMO-TsaH, which were expressed in lower temperature (18 ^{o}C) and longer expression time (between 12 and 16 hours). Reducing *in vivo* incubation by expressing the peptide in a higher temperature environment (30 ^{o}C) with a shorter expression time (3 to 4 hours) was attempted to resolve the degradation problem but did not lead to greater amounts of complete peptide. If the peptide was degraded within the cells, a greater improvement was anticipated as the peptide spent less time in the cells and the cells were lysed with protease inhibitors. Hence, the truncated peptide was a degradation product, it would be more likely to be produced during sonication and purification. Alternatively, the truncated peptide is a product of intact peptide fragmentation during ionisation in the MS analysis.

3.3.4.3. Tailoring enzyme quality assurance

Although both the tailoring enzymes had their identity confirmed by bottom-up proteomics, and the functionality of the tagged TsaH was confirmed by *in trans* complementation (Section 3.3.6), concerns regarding the protein qualities remained as *in vitro* reconstitution assays failed to obtain any expected results (Section 3.3.5). Therefore, top-down proteomics analyses were carried out to check whether intact protein could be seen. This was to verify the tailoring enzymes were not degraded during purification.

The intact TsaI protein mass was identified from deconvoluted spectra (Figure 3.14). The intact mass showed that the His_6-TsaI protein also had the N-terminal methionine removed, and a relatively small amount of gluconoylated version was detected too. This result verified the primary structure of His_6-TsaI protein was correct, but this top-down analysis could obviously not confirm whether the secondary and higher structures of the protein were correct too.

When the experiment was repeated for His_6-SUMO-TsaH, neither the masses with the N-terminal methionine (62719 ±1) nor without the methionine (62588 ±1) were observed in the deconvoluted spectra. It was also concerning that not all tryptic peptides fragments identified from the bottom-up proteomic analysis could be found. Figure 3.10b, shows that neither of C- nor N-termini of TsaH as well as some fragments in the middle of the protein were identified by the analysis. The missing fragments could be due to the detection issues in particular the peptide fragments in the middle of the protein, because the sample was excised from an SDS-PAGE gel band corresponding to an appropriate size protein.

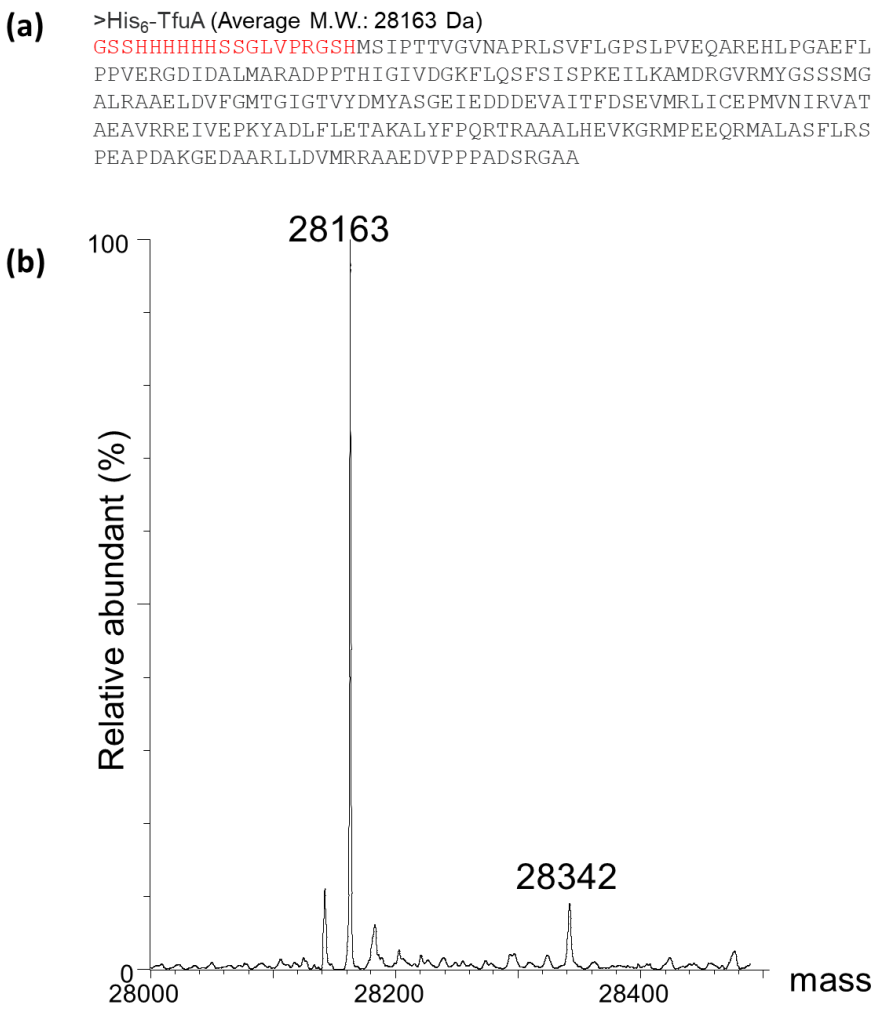


Figure 3.14: (a) Amino acid sequence of the His_6-TsaI with its calculated peptide mass. His_6 tag region is displayed in red. (b) Deconvoluted spectrum of His_6-TsaI. The main molecular mass of 28163, corresponds to intact His_6-TsaI protein, whereas a minor peak at 28342 matches the mass of intact His_6-TsaI with gluconolyation.

However, the undetectable peptides could also mean some parts of the protein were simply not present, in particular the C-terminus as the N-terminal His_6 tag was essential for purification. If a small part of the protein was removed via degradation, the protein would still appear as similar size on the SDS-PAGE. If the protein was incomplete, especially the C-terminus, the purified protein may not be functional, and the reaction would not be likely to take place. To check this, a more extensive examination on the protein mass was taken. There were up to 39 amino acid residues missing from the N-terminus and up to 18 amino acid residues from the C- terminus, so a list with the combination of 760 potential masses

ranging from 56000 to 63000 Da were generated. If the protein lost a small part from either or both termini, a mass or masses from the list was expected to appear in the deconvoluted spectra. Although some deconvoluted masses were detected, none of those protein masses matched any in the list. These detected masses could be related to TsaH (intact or shortened) with PTMs, but they were more likely to be background proteins in the sample. Although the top-down proteomic data could not provide a conclusion on whether the protein was truncated at the C-terminus, combining both of top-down and bottom-up proteomic results suggested that there was a possibility that the purified TsaH in the sample was damaged and non-functional. Another explanation for this result could be that TsaH in the sample was not concentrated enough to provide reliable MS data.

3.3.5. *in vitro* thioamidation reconstitution

As mentioned in the introduction (Section 3.1), precursor peptide thioamidation by YcaO and TfuA proteins was proposed to be completed in the presence of ATP and a sulphur donor (Figure 3.3). Since the natural source of sulphur remained uncertain, sodium sulphide (Na_2S) was incorporated in the reaction to ensure a supply of this element. Therefore, the reaction mix should contain His_6-TsaA, His_6-TsaI, His_6-SUMO-TsaH, ATP and Na_2S to ensure the basic requirements for the reaction were fulfilled. For each thioamidation on the precursor peptide, every replacement of oxygen (16 Da) with sulphur (32 Da) would lead to an increase of 16 Da, up to 64 Da corresponding the four thioamidations expected to be observed in the fully modified peptide (thiostreptamide S4 in Figure 3.1).

The *in vitro* assays were carried out with 80 µM His_6-TsaA, 2 µM His_6-SUMO-TsaH, 2 µM His_6-TsaI by putting all the components in a range of reaction buffer (50 mM Tris-HCl, 125

mM NaCl, 20 mM $MgCl_2$, 5 mM DTT, 5 mM ATP and 2 mM N_2S) in the range of pH 6 to 9, and incubating the reaction for 8 hours at 25 ^oC and 30 ^oC. This reaction condition was acquired from the successful *in vitro* reconstitution of YcaO-TfuA thioamidation in methyl-coenzyme M reductase from *M. acetivorans* where buffer pH at 7.5 and incubation temperature at 30 ^oC were used (Nayak et al., 2017). Negative control contained the same compositions but replacing His_6-SUMO-TsaH with reaction buffer.

If TsaA was thioamidated, a mass shift of +16 was anticipated for each thioamide bond being installed to the peptide and to yield an expected mass of 11337 from 11273 (Full-length TsaA), or partial thioamidated with 1, 2 or 3 thioamide bonds to give expected masses of 11289, 11305 or 11321, respectively. However, no modifications were detected from any of the reactions. With the respect to TsaA, the deconvoluted spectra between a reaction sample and a negative control showed the peak for the full-length TsaA. In fact, all TsaA related peaks, including the gluconoylated, and both gluconoylated and non gluconoylated truncated TsaA remained unchanged between reaction samples and negative controls. Although the experiment did not show expected results for thioamidation, the samples in different buffer pH for the *in vitro* assays showed similar full-length and truncated TsaA suggesting that within the Tris-HCl buffering range, the peptide behaved the same.

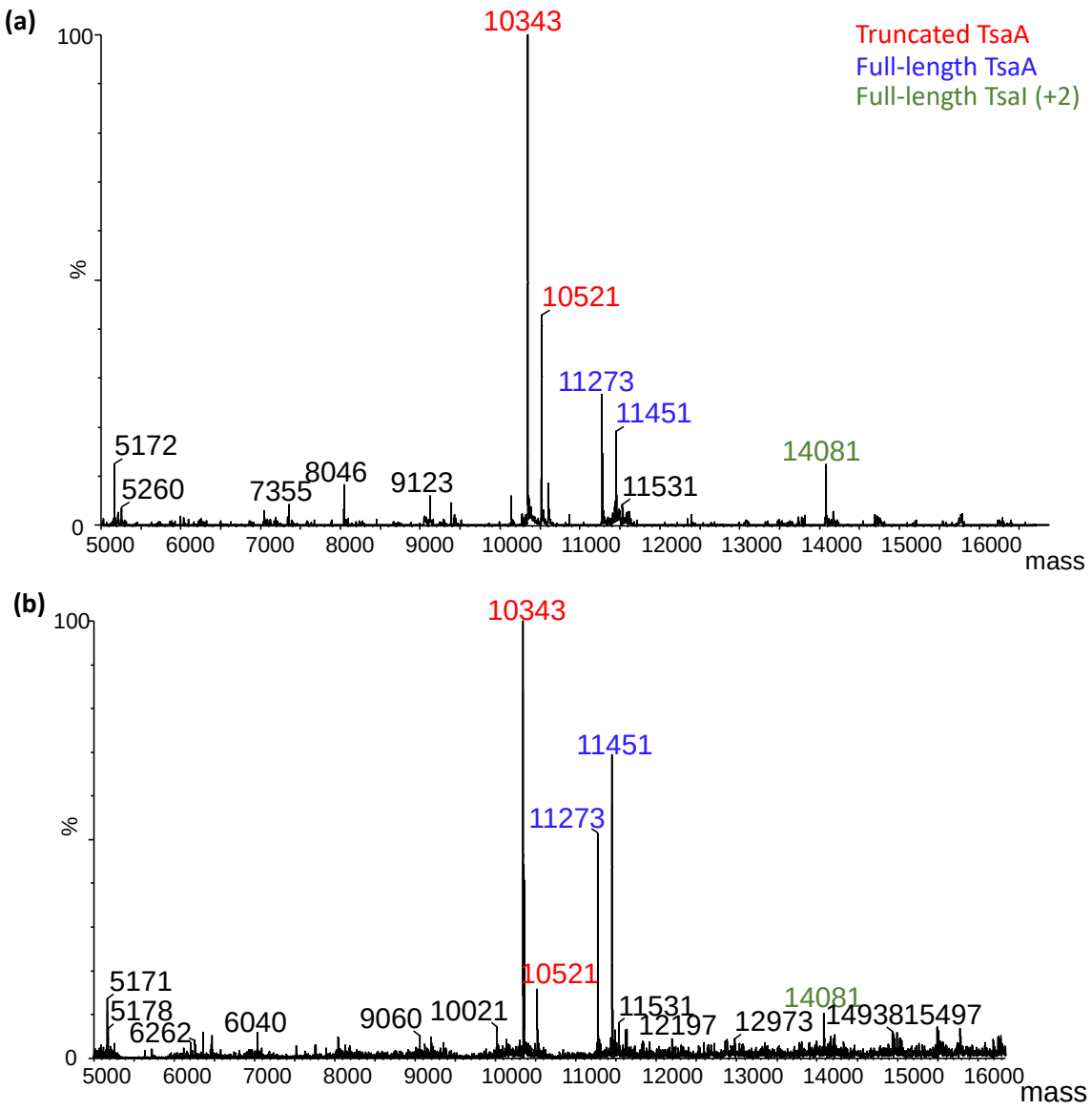


Figure 3.15: Deconvoluted spectra of *in vitro* assay samples. (a) From a negative *in vitro* assay control by replacing His_6-SUMO-TsaH with reaction buffer; (b) a representative *in vitro* assay sample where all reaction components were added.

One potential problem was that both tailoring enzymes and precursor peptide still had protein tags attached. Hence, a new set of *in vitro* assays was carried out adding proteases to cleave these tags. Thrombin and His_6-Ulp1 (His_6-tag ubiquitin-like-specific protease 1) was purified and confirmed by proteomics for the experiment. A summary of the different reactions carried out with and without proteases can be found in Table 3.4.

Reactions 1 to 4 correspond to YcaO-minus negative controls for reactions 5 to 8, respectively. For example, it was expected to see mass 11273 Da corresponding to unmodified full-length peptide in reaction 1. On the other hand, in reaction 5 (containing all the reaction components but not the proteases), if the substrate peptide had been modified, a peak corresponding to fully modified substrate (4 thioamide bonds: 11337 Da) or peaks with various level of modifications (1 thioamidation: 11289 Da, 2 thioamidation: 113053 Da, 3 thioamidation: 11321 Da and 4 thioamidation 11337 Da) would have been observed from deconvoluted spectra. Reactions 2 and 6 were set up as same as reactions 1 and 5 with the addition of His_6-Ulp1 to remove the SUMO tag on TsaH in case it affected its enzymatic activity.

After thorough MS analysis of the *in vitro* reactions, none of the predicted thioamidated peptide masses (9539, 9555, 9571 and 9587 Da in the case of untagged full length TsaA) were detected. As mentioned previously the truncated peptide still had 2 remaining thioamidation sites and its leader peptide was undamaged, so potential masses of thioamidated truncated TsaA (10359, 10375, 8608 and 8624 Da) were also investigated. In both tagged and untagged versions, an increase of 16 Da was observed. However, this increase of 16 was also detected in their corresponding negative controls lacking YcaO protein, indicating the peak resulted from a non YcaO protein related PTM, such as an oxidation. In summary, these results indicated that no thioamidation was achieved in the *in vitro* assays.

The His_6-tag on precursor peptide and TsaI protein was also taken into consideration, so reactions 3, 4, 7 and 8 were made as replicates for reaction 1, 2, 5 and 6 respectively with thrombin to cleave off the tag. After the removal of His_6-tag the full length TsaA peptide

mass without modification was 9523 Da instead of 11273 Da. The MS analysis of the *in vitro* assays was performed by comparing the deconvoluted peaks in the mass range between 10000 and 12000 Da on the spectra with the negative control. The analysis was focused on the non-gluconoylated versions of the peptides as it was not clear whether gluconoylation would have an impact on the thioamidation.

Table 3.4: Summary of the *in vitro* thioamidation assays carried out and their results. Mass 11273.3 corresponds to the non-modified full length His_6-TsaA, and the following masses correspond to 4 different levels of thioamidation (11289 to 11337 Da). Full length His6-TsaA digested by thrombin would give TsaA (9523 Da), and the following masses correspond to the 4 potential thioamidations (9539 to 9587 Da). Truncated His_6-TsaA had part of the core peptide truncated (unmodified truncated peptide: 10343 Da). 10359 and 10375 Da were the potential modified products as it only has 2 sites for thioamidation. When the unmodified truncated peptide was digested by thrombin, a mass of 8592 Da was expected. Abbreviation: TA (thioamidation)

	Reaction	1	2	3	4	5	6	7	8
Reaction components	His_6-TsaA	✓	✓	✓	✓	✓	✓	✓	✓
	His_6-TfuA	✓	✓	✓	✓	✓	✓	✓	✓
	His_6-SUMO-YcaO	×	×	×	×	✓	✓	✓	✓
	Thrombin	×	×	✓	✓	×	×	✓	✓
	His_6-Ulp1	×	✓	×	✓	×	✓	×	✓
Full-length His_6-TsaA	(unmodified) 11273								
	(+1 TA) 11289								
	(+2 TA) 11305								
	(+3TA) 11321								
	(+4 TA) 11337								
Full-length TsaA	(unmodified) 9522								
	(+1 TA) 9539								
	(+2 TA) 9555								
	(+3 TA) 9571								
	(+4 TA) 9587								
Truncated His_6-TsaA	(unmodified)10343								
	(+1 TA)10359								
	(+2 TA)10375								
Truncated TsaA	(unmodified)8592								
	(+1 TA)8608								
	(+2 TA)8624								

KEY

- ✓ A presence of the given component
- × Absence of the given component
- Detection of a relevant mass
- No detection of a relevant mass
- Detection of a relevant mass that was not expected in the reaction
- No detection of a relevant mass that was not expected in the reaction

3.3.6. Functionality of the tagged YcaO protein by *in trans* complementation

Although the purification tags were designed to have minimal impacts on the protein they were fused to, whether and how His_6-SUMO tag would affect the folding and activity of TsaH was not certain. An effective way to assess whether the tagged protein was still functional was to use it to complement a biosynthetic pathway knockout *in trans*. In a previous study, the BGC for thiostreptamide S4 (TsaS4BGC) had been cloned into *Streptomyces coelicolor* M1146, as well as a BGC with a deletion of *tsaH* gene (TsaS4BGC_ΔH) (Eyles et al., 2021). In the study, thiostreptamide S4 was not produced from *S. coelicolor* M1146 with TsaS4BGC_ΔH, but only when TsaH complemented the biosynthetic pathway (Eyles et al., 2021).

This complementation experiment was mimicked with the tagged TsaH. If the tagged protein complemented the knockout mutant, the protein would be considered to be unaffected by the tag. To do so, the His_6-SUMO-TsaH fusion was amplified from pLF005 and cloned into to pIJ10257 to generate pLF008, so that the tagged protein could be integrated into the ϕBT1 attB site of *S. coelicolor* M1146, a different chromosomal locus to where TsaS4BGC_ΔH was integrated. Production of thiostreptamide S4 was recovered when the M1146 TsaS4BGC_ΔH mutant was complemented with the gene for His_6-SUMO-TsaH. This showed that the His_6-SUMO tag did not destroy the TsaH protein function. However, the efficiency of the protein appeared to be affected by the tag. Under the same production and extraction conditions, the wildtype pathway (TsaS4BGC) and the knockout mutant complemented with wildtype *tsaH* gene (TsaS4BGC_ΔH + TsaH) had a very similar level of thiostreptamide S4 production, but when the knockout mutant was complemented with the tagged TsaH, the production was much lower in comparison of the other two (Figure

3.16). If the modification efficiency of TsaH was affected by the tag, either the reaction time might need to be extended or the amount of TsaH need to be added to the reaction to obtain sufficient modified peptide for detection.

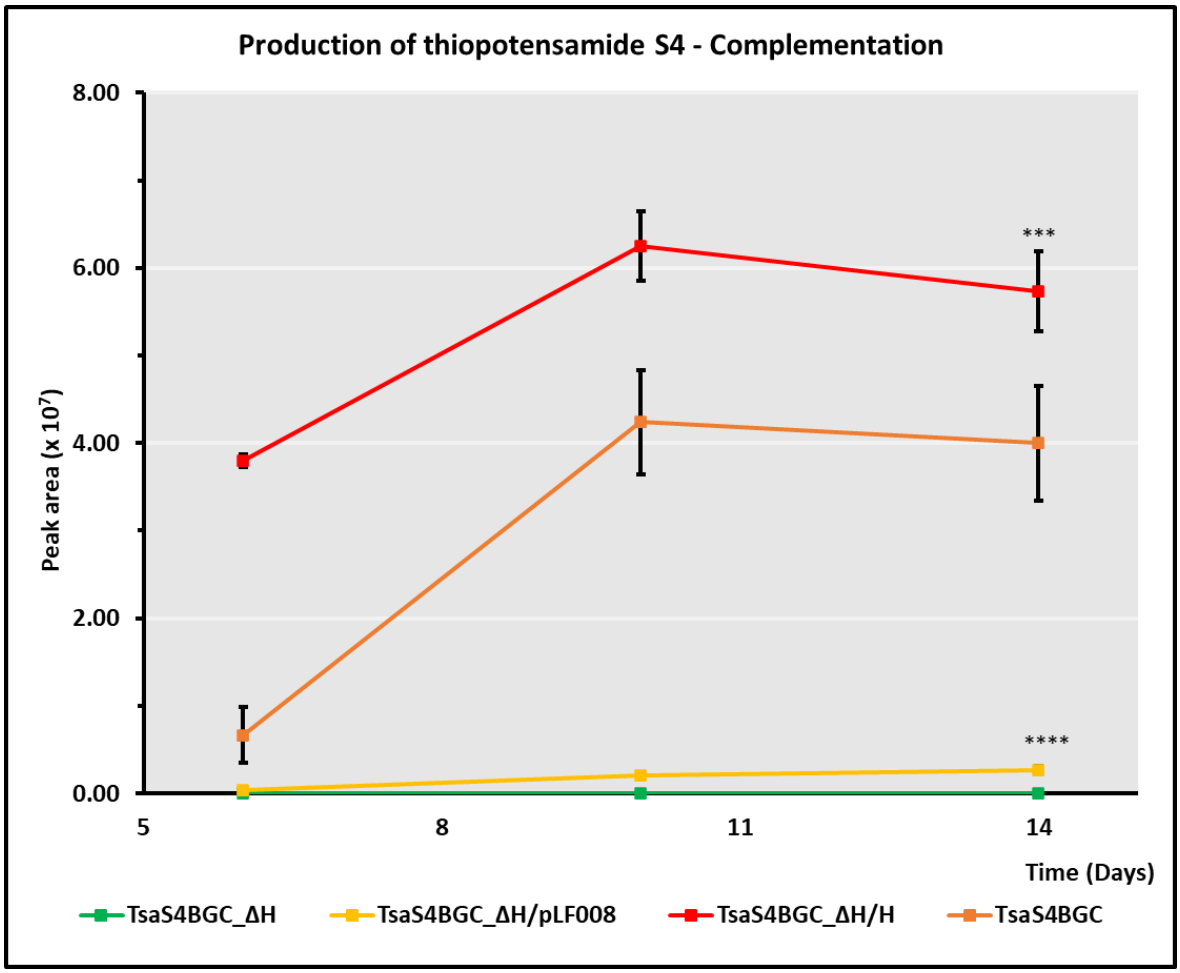


Figure 3.16: Production of thiostreptamide S4 in complementation experiment. Production was carried out for 14 days and 3 time-points were taken on day 6, 10 and 14.Wild type BGC (TsaS4BGC) was used as production positive control; TsaBGCS4_ΔH was a knockout mutant that had the ycaO gene (*tsaH*) deleted as a negative control; TsaBGCS4_ΔH + TsaH was the knockout mutant complemented with the wildtype *tsaH*; TsaBGCS4_ΔH + pLF008 was the knockout mutant complemented in trans with the His_6-SUMO-YcaO gene fusion. The production was significantly lower when the mutant was complemented with pLF008 comparing to the wild type BGC (p-value = 8.05×10^{-7}). Data were collected from 5 analytical replicas and error bars were plotted by standard mean error.

If the catalytic reaction was limited by the reaction efficiency only, positive results would have appeared in the *in vitro* reactions with proteases added, as SUMO tag should not

affect the protein function permanently (Georgiou and Valax, 1996). However, the results showed otherwise (Table 3.4). When Ulp1 and/or thrombin was added to the experiment, no thioamidated TsaA was observed suggesting either one or more of the catalytic proteins was absent or malfunctioning.

The lack of intact detection from top-down proteomics and the absent of C-terminal peptide fragments from bottom-up proteomics by TsaH indicates that TsaH was likely to be degraded and responsible for challenges in the *in vitro* reconstitutions. However, the stability of TsaH in the reaction mix was not certain, and the low solubility of TsaH was main limiting factor. Although it is not likely, the possibility of His_6 tag influencing the TsaI function cannot be ruled out at this stage.

3.4. Conclusion

After examining the solubility of YcaO proteins from 6 different thioamitide pathways, the enzyme from the biosynthetic pathway of thiostreptamide S4 was chosen because the YcaO protein (TsaH) was slightly soluble when a SUMO tag was fused to its N-terminus. Although the tagged TsaH was partially soluble when produced in *E. coli*, low solubility was still a bottleneck for this study. Improvement of the solubility was attempted by changing the expression strains and co-expressing the protein with the cognate TfuA protein (TsaI), but none of the attempts showed an increase in TsaH solubility.

Both the precursor peptide (TsaA) and TsaI were readily soluble when produced in *E. coli*, and were isolated using immobilised nickel-affinity chromatography, whereas TsaH required a further purification step on a size exclusion column. Production of the peptide and proteins were initially confirmed by SDS-PAGE and were then assessed by both bottom-up and top-down proteomics for their identities. Although bottom-up proteomics showed positive for all 3 materials, tagged TsaH could not be detected using top-down proteomics, and it was not clear whether it was caused by the low concentration, degradation or the combination of both. Peptide substrate stability was also performed using the *in vitro* assay conditions, and the substrate appeared to be stable over 8 hours.

Although *in vitro* reconstitution reactions were performed with this material, no thioamidation was observed in the experiments. TsaH functionality investigation showed that the tagged TsaH is capable of complementing the pathway *in vivo*, even though the production was significantly reduced. Hence, *in vitro* reconstitution was carried out with

proteases that cleaved the tags. However, no thioamidation was observed in the experiments.

Overall, it appeared that TsaH was the bottleneck for the success of this experiment. This protein was hard to purify, and very little amount of soluble protein was obtained for the experiments. It is not entirely clear whether the purified TsaH was intact and even if it was the complementation experiment indicated the tagged version is significantly less efficient than the wild type one.

Although the experiment was inconclusive for this particular pathway, a study by Sikandar et al published recently showed the *in vitro* reconstitution of the thioamidation on thioholgamide biosynthetic pathway (Sikandar et al., 2022). Sikandar *et al* confirmed that the precursor peptide was modified by YcaO and TfuA proteins when ATP and a sulphur source were supplied (Sikandar et al., 2022). Their study also revealed the mystery in the experiment that a precursor peptide with truncated C-terminus would not undergo thioamidation even if all thioamidation components were supplied (Sikandar et al., 2022).

3.5. Future implication

The *in vitro* study of the thioamidation through TsaH-TsaI pair was challenging due to the low solubility of TsaH and the potential malfunction of enzyme(s). There are a few potential follow-up experiments can be performed. First of all, the solubility of TsaH could potentially be improved using MBP tag, given the success on the YcaO protein purification using MBP tag for thioholgamide and thiostreptamide S87 *in vitro* biosynthesis (Sikandar et al., 2022, Xiong et al., 2022). In the thioholgamide experiment, using additive, high concentration of urea, in the lyse buffer solubilised the YcaO protein (ThoH), so simply adding additives to the lysis buffer can potentially improve the yield of soluble TsaH in the future. Given a bulkier tag MBP was used in ThoI and TsdI (TsaI homologues from thioholgamide and thiostreptamide S87 biosynthetic pathways, respectively), it was not very likely that a His_6 tag will lead to an unfunctional TsaI. However, this can be tested with an *in trans* complementation.

The recent success of *in vitro* reconstitution of thioholgamide was an encouraging breakthrough to get more insights into the biosynthesis of thioamitides. It enables many downstream experiments for more in-depth understanding of these pathways. While it has been confirmed that YcaO and TfuA proteins are essential for thioamidation, and TfuA protein was known as a partner protein for YcaO protein the role of this partner protein is less clear (Mahanta et al., 2018, Eyles et al., 2021, Liu et al., 2021a, Sikandar et al., 2022, Xiong et al., 2022). The TfuA domain in *M. acetivorans* was reported to have a role in support of sulphur sourcing to its YcaO partner by hydrolysing thiocarboxylated ThiS, as well as enhancing the YcaO-substrate peptide interaction (Liu et al., 2021a). Interestingly, the enzymatic activities between YcaO and TfuA proteins appeared to be bidirectional, as

YcaO-substrate peptide complex facilitated the thiocarboxylated ThiS hydrolysation from TfuA protein. This archaea study provided an insight into the role of TfuA proteins, but further investigation to identify the responsibilities of TfuA protein in thioamitide biosynthesis is required. Besides the function of TfuA proteins in the biosynthesis, there are a few questions that are more specific to thioamitide formation.

Although it has been shown that the length of the core peptide was essential rather than its sequence for successful thioamidation, it is also known that the leader peptide in precursor peptide serves the purpose for enzyme recognition and binding (Oman and van der Donk, 2010, Sikandar et al., 2022). Therefore, it would be necessary to identify whether there are any protein-peptide interactions. This could also reveal the mechanism that controls how a specific number of thioamide bonds are installed in each specific biosynthetic pathway, especially as different thioamitides feature different numbers of thioamides (Figure 3.1). It had also been discussed previously how installing thioamide bond could improve the bioactivities. If there was a correlation between the anticancer activity of thioamitides and their number of thioamide bonds, revealing the secret of regiochemical control would be crucial for accurately and effectively engineering the biosynthetic pathway for novel thioamitides with improved bioactivity, as well as engineering the production of new thioamide-containing peptides.

The success of *in vitro* reconstitution also enables the optimisation of the catalytic efficiency from YcaO and TfuA proteins. For example, the change of catalytic constant can be observed by changing the amino acid sequence of the proteins. In section 3.3.6, results of *in trans* complementation with tagged YcaO protein, showed the efficiency dropped dramatically in comparison to when the mutant was complemented with the wildtype YcaO

protein. It would be interesting to obtain the catalytic constant from *in vitro* assays with TsaH and TsaI carrying a relatively bulky tag. This will confirm whether the tag affects the interaction between the TsaI protein or the substrate directly or indirectly. Finally, as with many other of natural products, thioamitides are also produced in very low amounts. Given their potential as antitumour compounds, there is a chance there would be an interest to develop thioamitides for clinical use. Sikandar et al have successfully synthesised thioholgamide completely *in vitro*, so techniques like immobilised enzymes could be employed for higher scale synthesis of these molecules (Sikandar et al., 2022).

Chapter 4 Materials and methods

4.1. Materials

4.1.1. Materials and equipment

Unless otherwise stated, all chemicals and media components were from Merck (formerly Sigma-Alrich). NEBuilder HiFi DNA assembly master mix, Q5 High-Fidelity polymerase and all restriction enzymes were purchased from New England BioLabs (NEB). GoTaq® DNA polymerase and Wizard® Plus MiniPrep DNA purification system were purchased from Promega. MP Biomedicals™ FastDNA spin kit for soil was used to extract the genomic DNA from actinomycetes. Unless otherwise specified, water used in the project was filtered by MILLPORE MilliQ advantage A10 system. Primers for PCR and Sanger sequencing were ordered and manufactured from Integrated DNA technologies and the sequences are listed in section 4.3.3.

The final concentrations of antibiotics were at 50 μg mL^{-1} for apramycin, 50 μg mL^{-1} for carbenicillin, 25 μg mL^{-1} for chloramphenicol, 50 μg mL^{-1} for erythromycin, 50 μg mL^{-1} for hygromycin, 50 μg mL^{-1} for kanamycin and 25 μg mL^{-1} for nalidixic acid.

A NanoDrop 2000 spectrophotometer (Thermo Scientific) was used in the project to determine the concentration and purity of gDNA, plasmids, and gel purified DNA materials such as digested vector backbones and PCR products.

4.1.2. Strains

Amycolatopsis alba DSM 44262, *Nocardiopsis pontes* DSM 45324, *Streptomyces* sp. NRRL S-4 and *Streptomyces* sp. NRRL S-87 were previously obtained by the Truman group (Frattaruolo et al., 2017). *Nonomuraea fuscirosea* DSM 45880 and *Micromonospora eburnea* DSM 44814 were purchased from the DSMZ culture collection (Braunschweig,

Germany). *Streptomyces coelicolor* M1146 was used as the heterologous expression host for actinomycete gene clusters (Gomez-Escribano and Bibb, 2011). *Escherichia coli* DH5α (Invitrogen) was used for routine plasmid propagation, *E. coli* DH10β (New England BioLab) was used for large plasmid (≥20 kb) propagation and *E. coli* ET12567 was used for conjugations (MacNeil et al., 1992). *E. coli* BL21 (DE3) from Invitrogen, Rosetta™ (DE3) from Novagen and SoluBL21™ from Genlatins were used for protein expression. *E. coli* BW11253 was used for gene deletions (Datsenko and Wanner, 2000). A highly transformable yeast, *Saccharomyces cerevisiae* VL6-48N, was used for transformation associated recombination (TAR) cloning (Zhang et al., 2019).

4.1.3. Plasmids

Table 4.1: Backbone plasmid used in the thesis

Plasmid	Features	Selectable Markers	Application
pET28a (Merck)	T7 RNA polymerase under *lac* control. N- or C- His tagging.	Kanamycin	Heterologous expression in *E coli*
pTB145 (Bendezú et al., 2009)	*lacI*^{Q} P_{T7}:: *His*$_6$*-ulp1*(403- 621) *bla*	Carbenicillin	Ubiqoutin
pTB146 (Bendezú et al., 2009)	*lacI*^{Q} P_{T7}:: *His*$_6$*-SUMO bla* carries a unique sapI cloning site for seamless fusion to SUMO tag	Carbenicillin	Heterologous expression in *E coli*
pCAP01	*ARSH4/CEN6-Trp1*, pUC ori, ϕC31 int-attP-oriT-aph, *ADH1*	Kanamycin	Negative control to pCAP backbone constrcuts
pCAP03 (Zhang et al., 2019)	*ARSH4/CEN6-Trp1*, pUC ori, ϕC31 int-attP-oriT-aph, *URA3*, *ADH1*	Kanamycin	TAR cloning
pIB139 (Wilkinson et al., 2002)	ϕC31integration site, Apramycin resistance, Perm* promoter	Apramycin	Genetic editing
pIJ10257 (Hong et al., 2005)	*ermE** promoter2 ϕBT1 phage integration site Origin of transfer oriT from RK2	Hygromycin	Over expression or complementation
pIJ86 (Kieser et al., 2000)	pIJ101 origin of replication, bifunctional replication origin	Apramycin	Genetic editing
pCRISPOmyces	pSG5 origin of replication, spCas9 bifunctional replication origin	Apramycin	Genetic ediiting
pIJ773ϕOriT (JIC collection)	oriT, non-conjugative, flippase recognition target (FRT) sites	Apramycin	Gene deletions via PCR targeting
pUZ8002(Kieser et al., 2000)	Helper strain in triparental mating	Hygromycin	Conjugation helper plasmid
pR9604 (Jones et al., 2013)	Helper strain in triparental mating	Carbenicillin	Conjugation helper plasmid

4.1.4. Media components

E. coli were grown in lysogeny broth (LB) or DN Agar (2.3% Difco Nutrient Agar (BD)) depending on the antibiotic used for selection. Other media that were used for *E. coli* were Lennox broth (L), super optimal broth without magnesium (SOC-Mg: 2% tryptone, 0.5% yeast extract, 10 mM NaCl, 2.5 mM KCl) and super optimal broth with catabolite repression (SOC: 2% Tryptone, 0.5% Yeast extract, 10 mM NaCl, 2.5 mM KCl, 10 mM $MgSO_4$, 10 mM $MgCl_2$, 20 mM glucose).

Unless specified, *A. alba, S. coelicolor, S.* sp. NRRL S-4 and *S.* sp. NRRL S-87 were cultured on SFM (2% soy flour (Holland & Barrett), 2% mannitol and 2% agar (Formedium)). *M. fuscirosea*, *M. eburnea* and *N. potens* were grown on GYM (0.4% D-glucose anhydrous (Fisher), 0.4% yeast extract (BD), 1% malt extract (Oxoid), 0.2% $CaCO_3$ (BD) and 2% agar, pH 7.2). All actinomycete strains in the project could be grown on tryptone soy broth (TSB) to gain biomass. CR medium (1% glucose, 10.3% sucrose, 1.5% tryptic soy broth, 0.5% yeast extract, 50 mM $MgCl_2$) was employed to prepare actively growing *N. potens* for electrocompetent mycelium.

Media used for thioamitide production screening were:

SM05 (2% polypeptone (BD), 0.8% beef extract powder, 1.5% D-glucose anhydrous, 1% glycerol, 0.04% $CaCO_3$, pH 7.2), SM07 (2.09% MOPS, 1.5% L-proline, 2% sucrose (Fisher), 0.15% L-glutamic acid, 0.05% NaCl, 0.2% K_2HPO_4, 2 mM $MgSO_4$ (Fisher), 0.2 mM $CaCl_2$, 0.5% R5 trace element stock, pH 6.5)

SM12 (1% soya flour, 5% D-glucose anhydrous, 0.4% polypeptone, 0.4% beef extract powder, 0.1% Yeast extract, 0.25% NaCl, 0.5% $CaCO_3$, pH7.6), SM14 (1% D-glucose

anhydrous, 2% soy peptone, 0.5% beef extract powder, 0.5% NaCl, 1.7mM $ZnSO_4 \cdot 7H_2O$, 0.2 mM $CaCl_2 \cdot 2H_2O$, 0.5% R5 trace element stock, pH 6.5)

SM17 (0.2% D-glucose anhydrous, 4% glycerol, 0.2% soluble starch (BD), 0.5% soya flour, 0.5% polypeptone, 0.5% yeast extract, 0.5% NaCl, 0.2% $CaCO_3$, pH6.4), SM18 (1.5% D-glucose anhydrous, 4% soluble starch, 2.5% pharmedia, 1.6% cane molasses, 0.8% $CaCO_3$, pH7.2), SM20 (2% maltose, 0.5% polypeptone, 0.5% beef extract powder, 0.3% yeast extract, 0.3% NaCl, 0.1% $MgSO_4 \cdot 7H_2O$, pH 7.2)

SM25 (1% polypeptone, 2.1% malt extract, 4% glycerol, pH 6.3)

Bottromycin production medium (BPM, 1% glucose, 1.5% soluble starch, 0.5% yeast extract, 1% soya flour, 0.5% NaCl, 0.3% $CaCO_3$, pH 7.6).

The screening media were made with milliQ water with exception of that SM17, SM18 and SM20 were made with tap water from John Innes Centre, and the pH were adjusted with 5 M KOH at 95% of the desired volume before making up to the correct volume. R5 trace element stock prepared by the laboratory support team in John Innes Centre and contained 0.004% $ZnCl_2$, 0.02% $FeCl_3 \cdot 6H_2O$, 0.001% $CuCl_2 \cdot 2H_2O$, 0.001% $MnCl_2 \cdot 4H_2O$, 0.001% $Na_2B_4O_7 \cdot 10H_2O$, 0.001% $(NH_4)_6Mo_7O_{24}$.

S. cerevisiae VL6-48N was grown in yeast extract peptone dextrose medium (YPD medium: 2% D-glucose anhydrous, 1% yeast extract, 2% peptone). For selection of transformants, top selective agar (1 M sorbitol, 2.2% dextrose, 0.17% yeast nitrogen base without amino acids and ammonium sulphate, 0.19% yeast synthetic drop-out medium supplements without tryptophan, 0.5% ammonium sulphate, 0.01% adenine, 3% agar) and bottom selective agar (1 M sorbitol, 2.2% dextrose, 0.17% yeast nitrogen base without amino acids

and ammonium sulphate, 0.19% yeast synthetic drop-out medium supplements without tryptophan, 0.5% ammonium sulphate, 0.01% adenine, 0.0001% 5-FOA, 2% agar).

4.2. Methods

4.2.1. Bioinformatic analysis

4.2.1.1. Mining thioamitide biosynthetic pathways

The gene sequence of the YcaO domain protein in the thioviridamide pathway from *Streptomyces olivoviridis* NA05001 was used to obtain its homologues. The database from National Center for Biotechnology Information (NCBI) was searched by using BLASTX with the genetic code set to Bacteria and Archaea. The genes of interest were selected from the BLAST results, and GenBank file of the potential gene clusters were downloaded from -15,000 bp to +15000 bp of the gene of interest, so a total size of approximate 31,500 bp of a genomic DNA section was downloaded for each candidate strain. The functions of the genes in the DNA sections were manually annotated by its putative conserved domain according to their BLASTX result (Lu et al., 2020).

4.2.1.2. Phylogenetic analysis TvaCDE homologues in thioamitide pathways

The HopA1 proteins (TvaD homologues) of interest including the outgroup firstly had their protein sequences aligned by MUSCLE from the EMBL-EBI search and sequence analysis tools with Pearson/FASTA as the output format (Chojnacki et al., 2017). The alignments were uploaded to CIPRES science gateway for phylogenetic analysis by RAxML-HPC2 on XSEDE (Miller et al., 2011). The parameters were set to "raxmlHPC-HYBRID -T 4 -N autoMRE -n HopA1_filtered_RiPR -s infile.txt -p 12345 -m PROTGAMMABLOSUM62 -k -f a -x 12345 -o AAF71481.2 HopA1 [Pseudomonas syringae pv. syringae] --asc-corr lewis", AAF71481.2 HopA1 [Pseudomonas syringae pv. syringae] was the outgroup name, and the outgroup was a HopA1 protein (accession number: AAF71481.2) from *Pseudomonas syringae pv. syringae*. The phylogenetic trees were then visualised by iTOL (Letunic and Bork, 2016).

For phosphotransferases (TvaCE homologues), the same setup was used, but the outgroup was an aminoglycoside *O*-phosphotransferase (accession number: WP_063840677.1) from *Streptomyces griseus*.

4.2.2. Transforming *E. coli*

4.2.2.1. Making Electrocompetent cells

E. coli were inoculated into 10 mL LB and grown overnight at 250 rpm, 37 ^{o}C as a starter culture. A 500 μL (1%) inoculum of the starter culture was added to each of the two 50 mL LB in 250 mL conical flasks. They were incubated for 3-5 hours at 250 rpm, 37 ^{o}C until the reading of the OD_{600} reached between 0.3 and 0.5. Cells were harvested by centrifugation at 800 × *g* for 10 minutes, 4 ^{o}C and resuspended in a total of 100 mL of ice-cold 10% glycerol. This was repeated further 4 times, with resuspending in total of 100 mL, 100 mL, 50 mL and 1 mL of ice-cold 10% glycerol respectively. The final 1 mL electrocompetent cells were separated into aliquots of 50 μL in 1.5 mL Eppendorf tubes to be used or flash frozen in liquid nitrogen for storage at -80 ^{o}C.

4.2.2.2. Transforming electrocompetent cells

Electrocompetent cells were thawed, and electroporation cuvettes (2 mm) were cooled on ice. Each aliquot of cells had 1 μL of DNA solution added and mixed by stirring with the pipette tip. The DNA-cell mix were gently transferred to the electroporation cuvette and further incubated on ice for 30 minutes. The outside of the cuvette was dried before the cuvette being inserted into a Gene Pulser (Bio-Rad) with the pulse generator set to 25 μFD, 2.5 kV, and 200 Ω. Once the pulse was delivered, 950 μL of SOC was added. The cells were returned to a microcentrifuge tube and were incubated to recover for 1 hour at 250 rpm,

37 °C. The recovered cells were plated on LB (DN Agar if hygromycin was used as selection marker) with the appropriate antibiotic as selection marker. The plates were incubated at 37 °C overnight to obtain colonies.

4.2.3. General yeast methods

4.2.3.1. Transformations

Yeast transformations followed the TAR cloning protocol from Zhang et al (2019). A single colony of *S. cerevisiae* VL6-48N from a YPD plate were inoculated in 10 mL of liquid YPD containing adenine and grown overnight at 250 rpm, 30 °C as a seed culture. A 1 mL (2%) inoculum of the seed culture was added to 50 mL liquid YPD media containing adenine in a 250 mL conical flask and grown for 8 hours at 30 °C, 250 rpm, until the OD_{600} reached 0.9. The cells were then harvested by a 3-minute centrifugation at 1800 ×*g*, 4 °C. The cell pellet was resuspended in 50 mL ice-cold sterile water and harvested by centrifuge at 1800 ×*g*, 4 °C for 3 minutes. The cells were resuspended in 50 mL of 1 M sterile ice-cold sorbitol and left on ice overnight.

The cells were harvested by centrifuge at 1800 ×*g* at 4 °C. After removing the supernatant completely, the cells were resuspended in 20 mL SPE solution (10 mM pH 7.5 HEPES, 0.2% EDTA, 1M sorbitol). The cells in SPE had 40 µL of 2-mercaptoethanol added followed by 80 µL zymolyase (Zymo Research, 5 U μL^{-1}). The cells were incubated in 30 °C for 40 minutes to obtain spheroplasts. The spheroplast mix was topped with 1 M ice-cold sorbitol to 50 mL, and the cells were then harvested by centrifuging for 10 minutes at 600 ×*g*, 4 °C. After removing the supernatant completely, the spheroplasts were gently resuspended in 50 mL of 1 M ice-cold sorbitol. The spheroplasts were harvested by centrifuging for 10 minutes at 600 ×*g*, 4 °C and had the supernatant removed. The spheroplasts were gently resuspended

in 2 mL STC solution (10 mM pH 7.5 Tris, 10 mM $CaCl_2$, 1 M sorbitol). The spheroplasts were left at room temperature for 10 minutes. 200 μL of the spheroplasts in STC were transferred to a microcentrifuge tube that already contained transforming DNA solution. The spheroplasts/DNA mixtures were incubated at room temperature for 10 minutes. After adding 800 μL of 20% PEG solution (10 mM Tris, 10 mM $CaCl_2$, 20% PEG8000, pH 7.5), the cells were further incubated at room temperature for 20 minutes. The cells were pelleted by centrifuge at 700 ×*g,* 4 °C for 10 minutes. The supernatant was removed, and the cells were gently resuspended by 800 μL SOS solution (6.5 mM $CaCl_2$, 0.25% w/v yeast extract, 1 M sorbitol, 1% w/v peptone). The cells were recovered at 30 °C with occasional inversion. The cells were then mixed with 8 mL 60 °C melted top agar, and immediately poured onto a selective bottom agar plate containing 0.0001% 5-FOA. The cells on plates were then incubated at 30 °C for 5 days to obtain colonies.

4.2.3.2. Colony screening

Colonies were picked to make patches onto a new selective agar plate without 5-FOA for each construct, and the plates were incubated at 30 °C for 2 days. A quarter of a patch was taken by a toothpick and resuspended in 50 μL 1M sorbitol with 5 U zymolyase. The cells were digested by zymolyase by incubating in 37 °C for 2 hours before being boiled for 10 minutes. The cells were centrifuged for 10 minutes at 600 ×*g*, and the 1 μL of the supernatant was used as DNA template for colony PCRs.

4.2.4. DNA extraction

4.2.4.1. Genomic DNA extraction from Actinomycetes

To collect genomic DNA from *A. alba* DSM 44262, *M. eburnea* DSM 44814, *N. fuscirosea* DSM45880, *N. potens* DSM45324, *S.* sp. NRRL S-4 and *Streptomyces* sp. NRRL S-87, the

stock cells were streaked on solid SF+M and grown in 30 °C for 5 days to obtain single isolated colonies. A colony was then collected from each plate and individually inoculated in 50 mL liquid DNB in a 250 mL conical flask with a sprig at 30 °C with 250 rpm shaking. The liquid cultures were then centrifuged at 1789 ×*g* for 10 minutes. Approximate 0.5 g of a pellet was collected for genomic DNA extraction by FastDNA SPIN kit for Soil. The extraction process was carried out following the manual provided. The DNA concentrations were measured using a NanoDrop.

4.2.4.2. Plasmid DNA extraction from *E. coli*

E. coli containing the target plasmid DNA were inoculated in 10 mL LB (L, if hygromycin was used as selection marker) with appropriate antibiotic for selection at 250 rpm, 37 °C overnight. The cells harvested by centrifugation at 1789 ×*g* for 10 minutes. The plasmids were harvested from the cells by using Wizard® miniprep kit. The plasmids were eluted with 100 μL of 55 °C prewarmed nuclease-free water. DNA concentrations were measured using a NanoDrop.

4.2.4.3. Plasmid DNA extraction from yeast

Colonies of yeast that were confirmed positive by colony PCRs were grown in 10 mL of liquid selective liquid medium without Trp at 30 °C, 250 rpm for 2 days. The cells were pelleted by by centrifuging for 10 minutes at 1789 ×*g* and then resuspended in 200 μL 1 M sorbitol with 2 μL of zymolyase (5 U μL^{-1}). The digested cells were then pelleted by centrifuging at 600 ×*g* for 10 minutes. The plasmids were then harvested using Wizard® miniprep kit.

4.2.5. Electrophoresis

4.2.5.1. Agarose gel electrophoresis

0.8% agarose in 1× TBE (Tris/Borate/EDTA) buffer with 3 μL ethidium bromide per 100 mL was used for electrophoresis of DNA. DNA samples were mixed with 6 × gel loading dye (NEB) before loading into the agarose gel wells. 5 μL of 1 kb plus DNA ladder (NEB) was used as a marker. The electrophoresis was run 1 × TBE running buffer using a BioRad PowerPac™ Basic (90V/400mA/75W) for 30 to 45 minutes depending on the length of the gel. Gel bands were visualised and imaged under UV.

4.2.5.2. SDS-PAGE analysis and preoteomic analysis

Sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE) analysis was carried out using a 12% resolving gel with 4% stacking gel, unless otherwise stated. The samples were mixed with equal volume of protein loading dye and heated at 100 °C for 10 minutes before loading onto the gel. The gel was run for 60 minutes with settings of 160 mV, 72 mA and 40 W and occasional checks to avoid overrunning. The gels were stained by InstantBlue® and washed by water for at least 3 hours prior to photography.

The separated protein bands from SDS-PAGE were excised to confirm the protein sequence by proteomics. In order to perform proteomic analysis, the gel bands were firstly destained with 1 mL 30% ethanol for 30 minutes at 60 °C, and the destaining step was repeated until the gel band became clear. The gel band was then washed for 20 minutes by vortex with 1 mL 50 mM TEAB in 50% ACN. The gel band was incubated with 1 mL 10 mM dithiothreitol in 50 mM TEAB for 30 minutes at 55 °C. After removing dithiothreitol, 1 mL 30 mM iodoacetamide in 50 mM TEAB was added to the gel band and incubated in dark and room temperature. The gel bands were sliced into small pieces and washed for 20 minutes with

1 mL 50 mM TEAB in 50% ACN. Finally, the gel pieces were washed twice with 1 mL ACN before sending to proteomics for trypsin digestion and LC-MS/MS.

Both digestion and proteomic analysis were performed by Dr Carlo de Oliveira Martins and Dr Gerhard Saalbach (John Innes Centre). The tryptic digested proteins were analysed by LC-MS/MS using an Acquity UPLC (Waters) coupled with a Synapt G2-Si mass spectrometer. Peptide sequences were analysed using Scaffold (Proteome Software) and compared against the target peptide sequences and the *E. coli* protein database to determine the identity of the extracted protein band.

4.2.6. Molecular cloning

4.2.6.1. YcaO domain protein codon optimisation

YcaO domain protein from *S.* sp. S-4 had its codon optimised and synthesised from Twist Bioscience for *E. coli*. The wildtype YcaO domain gene (locus tag: ADK82_RS1180, accession number: WP_031090932.1) with 131 bp after the stop codon with a total length of 1499 bp was uploaded to Twist Bioscience. Once the gene had been uploaded, it was edited for flanking sites and optimisation. Sequence (GGCTCACAGAGAACAGATTGGTGGT) taken from pTB146 was added to the 5' end as a flanking region. The ORF was selected from 26 to 1393 and *E. coli* was chosen as the host strain. Finally, options for "restriction enzyme sites to avoid" and "regions to preserve" were left as "unselected". The synthetic DNA product was used as insert for Gibson assembly.

4.2.6.2. Preparing DNA materials for cloning

The genes of interest were amplified by Q5 high fidelity polymerase using primers that contain approximately 30-nucleotide overhangs taken from corresponding backbone vectors (Table 4.2). The PCR mix followed the manufacturer's instruction with high-GC

enhancer. The temperatures for annealing stage were depending on the melting temperature of the primers, and the extension time was set to 40 second per 1000 bp. The qualities of the PCR products were checked by 0.8% agarose electrophoresis to determine if gel purification was required. The appropriate clean up steps were used for the PCR products by Illustra GFX PCR DNA and gel band purification kits.

Table 4.2: Primers and DNA templates used in PCR to amplify DNA for Gibson assembly

PCR products	**Primer 1**	**Primer 2**	DNA template
YcaO_AA_ins	AA-H-Start	AA-H-End	gDNA from *A. alba*
YcaO_ME_ins	ME-H-Start	ME-H-End	gDNA from *M. eburnea*
YcaO_NP_ins	NP-H-Start	NP-H-End	gDNA from *N. potens*
YcaO_NF_ins	NF-H-Start	NF-H-End	gDNA from *N. fuscirosea*
YcaO_S87_ins	S87-H-Start	S87-H-End	gDNA from *S. sp NRRL S-87*
001002	KW001	KW002	gDNA from *A. alba*
019020	KW019	KW020	gDNA from *M. eburnea*
023023	KW022	KW023	gDNA from *N. fuscirosea*
024025	KW024	KW025	gDNA from *N. potens*
009010	KW009	KW010	gDNA from *S. sp* NRRL S-4
037038	KW037	KW038	pLF005
031032	KW031	KW032	gDNA from *S. sp* NRRL S-4
100101	KW100	KW101	gDNA from *N. potens*
102103	KW102	KW103	gDNA from *N. potens*
104105	KW104	KW105	gDNA from *N. potens*
102106	KW102	KW106	gDNA from *N. potens*
104106	KW104	KW106	gDNA from *N. potens*
114115	KW114	KW115	gDNA from *N. potens*
116117	KW116	KW117	gDNA from *N. potens*
118119	KW118	KW119	gDNA from *N. potens*
120121	KW120	KW121	gDNA from *N. potens*
122123	KW122	KW123	gDNA from *N. potens*
124125	KW124	KW125	gDNA from *N. potens*

Plasmids used in the project were listed in table with the restriction enzymes that were used to linearise the backbone vectors. Backbone plasmids were digested overnight at 37

°C, and all digested backbone vectors underwent a gel separation in a 0.8% agarose prior to Illustra GFX PCR DNA and gel band purification kits.

Table 4.3: Restriction enzymes used to linearise plasmid backbone for Gibson assembly

Plasmid backbone	Enzyme 1	Enzyme 2
pTB146	SapI	PstI
pET28a	NdeI	SacI
pIJ10257	NdeI	HindIII
pCAP03	NdeI	XhoI
pMAL-28a	NdeI	EcoRI

4.2.6.3. Gibson assembly for plasmid constructions

Constructs for overexpressing proteins in *E. coli*, and *in trans* expression or complementation genes in *S. coelicolor* M1146 were assembled using NEBuilder HiFi DNA assembly master mix. For each assembly, the molar ratio of PCR product for insert and linearised backbone vector was approximately 3:1. Two constructs had 2 inserts, where the molar ratio of the two PCR products and the backbone was approximately 3:3:1. Regardless of the numbers of DNA fragments in the assembly mix, the total amount of DNA was between 200 and 300 ng in each reaction. Once an assembly reaction mix was prepared, the reaction took place at 50 °C for 30 minutes. After the 30-minute incubation period, the reaction mix was used to transform *E. coli* DH5α.

Table 4.4: DNA materials required for Gibson assembly and their selection markers. Backbone vector was prepared using the restriction enzymes stated in Table 4.3, and inserts were prepared following Table 4.2

Constructs	Backbone vector	Insert(s)	Selection marker
pET_YcaO_AA	pET28a	YcaO_AA_ins	Kanamycin
pET_YcaO_ME	pET28a	YcaO_ME_ins	Kanamycin
pET_YcaO_NP	pET28a	YcaO_NP_ins	Kanamycin
pET_YcaO_NF	pET28a	YcaO_NF_ins	Kanamycin
pET_YcaO_S87	pET28a	YcaO_S87_ins	Kanamycin
pLF001	pTB146	001002	Carbenicillin
pLF002	pTB146	019020	Carbenicillin
pLF003	pTB146	022023	Carbenicillin
pLF004	pTB146	024025	Carbenicillin
pLF005	pTB146	009010	Carbenicillin
pLF006	pTB146	Synthetic ycaO gene	Carbenicillin
pLF008	pIJ10257	037038	Hygromycin
pLF013	pMAL-28a	031032	Carbenicillin
pLF017	pIJ10257	100101 & 102103	Hygromycin
pLF018	pIJ10257	104105	Hygromycin
pLF019	pSS88	100101 & 102106	Hygromycin
pLF020	pSS88	104106	Hygromycin
pLF021	pIJ10257	114115	Hygromycin
pLF022	pIJ10257	116117	Hygromycin
pLF023	pIJ10257	118119	Hygromycin
pLF024	pIJ10257	120121	Hygromycin
pLF025	pIJ10257	122123	Hygromycin
pLF026	pIJ10257	124125	Hygromycin

4.2.6.4. Verifying the cloned plasmids

Up to eight DH5α colonies from each assembly were first patched on a fresh LB plate (L plate if hygromycin was used). To reduce the work and increase the efficiency, these colonies were also preliminary confirmed by colony PCRs with an external upstream primer, an internal downstream primer and GoTaq flexi polymerase. The annealing temperature was set at 55 ^{o}C and the extension time in the thermocycler was set to 1 minute per 1 kb. The PCR products were checked with 0.8% agarose electrophoresis.

Up to two of the positive clones from colony PCRs had their plasmids extracted, and unless otherwise specified, plasmids were sequenced using a Mix2Seq Kit (Eurofins Genomics).

4.2.6.5. PCR based TAR cloning

The protocol for TAR cloning the thiopotensamide gene cluster in yeast was modified from Zhang et al (2019). For the wildtype construct (pLF015), the biosynthetic gene cluster (BGC) was amplified from *N. potens* genomic DNA into 4 PCR products of the BGC fragments, and each BGC fragment had at least 60 bp overlapping with the neighbouring fragments. The beginning of the first fragment and the end of the last fragment also had 60 bp regions of homology with pCAP03 introduced by primers. The BGC was also refactored by replacing the UTR before the precursor peptide gene with the promotor containing region before thiovarsolin precursor peptide from *Streptomyces varsoviensis* and adding the P_{hrdB} promotor before the first tailoring enzyme in the gene cluster (Santos-Aberturas et al., 2019). Hence, there were 6 PCR products in total for the refactored construct (pLF016).

Table 4.5: Primers and DNA templates used in PCR to amplify DNA for TAR cloning

PCR products	Primer 1	Primer 2	DNA template
064065	KW064	KW065	gDNA from N. potens
066067	KW066	KW067	gDNA from N. potens
068083	KW068	KW083	gDNA from N. potens
086069	KW086	KW069	gDNA from N. potens
070071	KW070	KW071	gDNA from S. varsoviensis
072073	KW072	KW073	gDNA from N. potens
074075	KW074	KW075	gDNA from S. coelicolor
076067	KW076	KW067	gDNA from N. potens

The linearised pCAP03 and the PCR products of BGC fragments were purified by 0.8% agarose gel, followed by Illustra GFX PCR DNA and gel band purification kits. The purified

PCR products and linearised pCAP03 were transformed into yeast as described in Zhang's protocol (Zhang et al., 2019).

Table 4.6: DNA materials used for TAR cloning. Backbone vector was prepared using the restriction enzymes stated in Table 4.3, and inserts were prepared following Table 4.5

Constructs	Backbone vector	Inserts	Selection marker
pLF015	pCAP03	064065, 066067, 068083, 086069	Kanamycin
pLF016	pCAP03	070071, 072073, 074075, 076067, 068083, 086069	Kanamycin

40 yeast colonies were taken for each construct to make patches and perform 4 pooled colony PCRs using primer pair, KW089 and KW090. All pooled colony PCRs showed there was at least one positive colony, so individual colony PCRs were performed from a pool to identify the positive colony. Two positive clones were taken from each construct and had their plasmid extracted by using phenol-chloroform steps from the protocol, and the plasmid were transformed into *E. coli* DH10β for plasmid propagation. The plasmid constructs were sent to Plasmidsaurus for sequencing.

4.2.6.6. PCR targeting mutagenesis

To obtain gene deletions from pLF015 and pLF016, PCR targeting was used following the published protocol from Gust et al (2003). The disruption cassette containing apramycin resistance gene bracketed by FLP-recombinase recognition sites was amplified from pIJ773-oriT, a version of pIJ773 with the *oriT* removed in John Innes collection, using the following primers: the *tpaC* cassette used KW130 and KW131; the *tpaD2* cassette used KW129 and KW132.

pLF015 and pLF016 were individually introduced to *E. coli* BW25113/pIJ790 by electroporation, and the transformed cells were grown on LB-agar with chloramphenicol

and kanamycin at 30 °C. A single colony from each transformation was taken to grow overnight on 5 mL SOB-Mg with chloramphenicol and kanamycin at 30 °C, 250 rpm for electrocompetent cells. A 1% inoculum of this overnight culture was added to 10 mL SOB-Mg containing chloramphenicol, kanamycin and 10 mM L-arabinose. The cultures were then grown at 30 °C, 250 rpm for 4 hours. The cells were then harvested by centrifuge at 1789 ×*g* for 5 mins. Each cell pellet was washed twice with 10 mL ice-cold sterile water, and twice with 10 mL of ice-cold 10% glycerol. Each of the cell pellets was resuspended in 200 μL of 10% glycerol and separated into 50 μL aliquots in microcentrifuge tubes as electrocompetent cells. These cells were then transformed with the gene specific disruption cassettes by electroporation. The transformants were selected on LB-agar containing apramycin and kanamycin. Colonies containing the disrupted genes were checked by colony PCRs using the following primer pairs: Δ*TpaC* by KW086 and KW131; Δ*TpaD2* by KW132 and KW099. The *E. coli* BW25113/pIJ790 colonies with non-disrupted constructs were used as negative controls. Plasmids were extracted from *E. coli* BW25113 and then transformed into *E. coli* ET12567/pR9604 for conjugations, and the transformants were selected by LB-agar containing apramycin, carbenicillin and chloramphenicol.

4.2.6.7. *Streptomyces* conjugations

E. coli ET12567 carrying a helper plasmid was used for intergenic conjugations into *S. coelicolor* M1146, and the protocol was adapted from Kieser et al (2000). There were 2 helper plasmids used in the project. When a pCAP-based plasmid was used, pR9604 was used as helper plasmid. When a pIJ10257-based plasmid was used, pUZ8002 was used as helper plasmid. pCAP-based plasmids were transformed into *E. coli* ET12567/pR9604 by electroporation. The transformed cells were then selected by growing on LB-agar

containing carbenicillin, chloramphenicol and kanamycin. A single colony was used to inoculate 10 mL LB containing the same antibiotics at 37 °C, 250 rpm overnight.

A fresh growth of *E. coli* was initiated by adding 200 μL of the inoculum to 10 mL LB containing the same antibiotics, and the cells were grown in the same conditions for approximately 4 hours when OD_{600} reached between 0.4 and 0.6. The cells were then washed twice in 10 mL 10% glycerol and once in 1 mL LB. Finally, the cells were resuspended in 200 μL of LB.

5 μL of *S. coelicolor* M1146 spores were mixed with 500 μL of 10% glycerol, and heat-shocked at 50 °C for 10 minutes. The heat-shocked spores were left in room temperature for 10 minutes to cool before being pipetted to the washed *E. coli* ET12567 cells. This mixture was then plated on 30 mL SFM with 10 mM $MgCl_2$ and incubated at 30 °C overnight. The plate was overlaid with nalidixic acid and kanamycin to select for *S. coelicolor* M1146 pCAP-based plasmid exconjugants. Once isolated colonies of *S. coelicolor* appeared, the clones were patched on SFM containing nalidixic acid and kanamycin.

For pIJ10257-based plasmids, the transformed *E. coli* were grown on L instead of LB, and the selection antibiotics were chloramphenicol, hygromycin and kanamycin. *S. coelicolor* was selected by nalidixic acid and hygromycin.

4.2.6.8. Electroporation of *Nocardiopsis* mycelium

N. potens was inoculated in 100 mL CR medium (1% glucose, 10.3% sucrose, 1.5% tryptic soy broth, 0.5% yeast extract, 50 mM $MgCl_2$) from 100 μL mycelium stock at 30 °C for 24 hours. The mycelium was centrifuged at 4 °C 1789 ×*g* for 10 minutes. The harvested mycelium was washed in 100 mL ice-cold 10% sucrose and recentrifuged at using the same

programme (4 °C 1789 ×*g* for 10 minutes) to wash twice in 50 mL ice-cold 15% glycerol. The mycelium was then resuspended in 10 mL 15% glycerol and had lysozyme (final concentration of 100 μg mL^{-1}) added. The mycelium was incubated at 37 °C. The digested mycelium was then washed twice with 10 mL ice-cold 15% glycerol, and resuspended in 1 mL of 3% PEG 1000, 10% glycerol, 6.5% sucrose.

An aliquot of 50 μL of the mycelium suspension was mixed with 1 μL of plasmid (pIJ86, pCRISPOmyces and pIB139) and transferred to an ice-cold 2 mm-gapped electrocuvette. The cuvette containing mycelium was incubated on ice for 30 minutes and the outside of the cuvette was dried before the cuvette being inserted into a Gene Pulser (Bio-Rad) with the pulse generator set to 25 μFD, 2 kV, and 400 Ω. Once the pulse was delivered, 750 μL of ice-cold CR medium was added to the mycelium and incubate with shaking for 3 hours at 30 °C. The transformants then plated on 30 mL SFM incubated at 30 °C overnight. The plate was overlaid with apramycin to select for *N. potens* containing pIJ86, pCRISPomyces or pIB139 plasmid. Once isolated colonies of *N. potens* appeared, the clones were patched on SFM containing apramycin.

4.2.7. Protein expression and purification

4.2.7.1. Protein expression tests

Precursor peptide (*TsaA*), YcaO domain protein (*TsaH*) and TufA domain protein (*TsaI*) genes had previously cloned into pET28a plasmid by Dr Tom Eyles in the group (pET28a_S4_TsaA, pET28a_S4_TsaH and pET28a_S4_TsaI, respectively). pET28a, pTB146, pET_YcaO_AA, pET_YcaO_ME, pET_YcaO_NP, pET_YcaO_NF, pET_YcaO_S87, pLF001, pLF002, pLF003, pLF004, pLF005, pLF006, pLF013, pET28a_S4_TsaA, pET28a_S4_TsaH, pET28a_S4_TsaI were individually transformed into *E. coli* BL21 (DE3). The transformants

were grown overnight on LB agar plates containing appropriate antibiotics at 30 °C to obtain colonies. A single colony was used to inoculate in 10 mL LB with antibiotics overnight at 30 °C, 250 rpm. Colonies grown of LB plates were kept at 4 °C for up to 1 week. The expression of proteins was then induced with IPTG at final concentration of 1 mM, and the culture was further incubated at 30 °C with shaking for 4 hours. The cells were pelleted by centrifugation at 1789 ×*g* for 10 minutes, followed by lysis in 2 mL of 2% SDS for SDS-PAGE.

4.2.7.2. Protein solubility tests

An inoculum was grown from a colony in 10 mL LB containing appropriate antibiotic overnight at 30 °C, 250 rpm. The expression cultures were prepared by adding 1 mL of the overnight inoculum into 100 mL LB with antibiotic, and the expression cultures were grown at 30 °C, 250 rpm for approximately 3 hours. When the OD_{600} reached 0.6, the expression cultures were induced with 0.1 mM IPTG. After adding IPTG, the cultures were moved to be grown at 18 °C overnight, 200 rpm. The expression cultures were washed with equal volume of 20% glycerol by centrifugation at 3472 ×*g*, 4 °C for 30 minutes and resuspension. The cells were resuspended in 10 mL of lyse-buffer (500 mM NaCl, 50 mM Tris-HCl, 10% glycerol at pH 8.0) for sonication. The sonication was done on ice, and the sonicator (Sonics Vibra Cell VC 750) was connected to a 220-B probe and set to 40% amplitude, 8 minutes with a 3-second interval every second. The sonicated lysates were then centrifuged for 40 minutes at 51428 ×*g*, 4 °C to separate the soluble and insoluble fractions. The insoluble pellet was resuspended in 10 mL 2% SDS for SDS-PAGE.

4.2.7.3. Bulk production of proteins for purification

pET28a_S4_TsaA and pET28a_S4_TsaI contained thiostreptamide S-4 precursor peptide (*TsaA*) and TfuA domain protein (*TsaI*) from *S.* sp. NRRL S-4 in pET28a, which was kanamycin

resistant. pTB145 contained Ubl-specific protease 1 (*Ulp1*) with polyhistidine tag and pLF005 contained YcaO domain protein (*TsaH*) from *S.* sp. S-4 in pTB146, which contained a carbenicillin selection marker. The target proteins were expressed in *E. coli* BL21 (DE3). An inoculum was grown from a colony in 10 mL LB containing appropriate antibiotic overnight at 250 rpm, 37 ^{o}C as seed culture.

The expression cultures were prepared by adding 10 mL of the seed culture to each 500 mL LB with appropriate antibiotic. The cultures were initiated at 200 rpm, 30 ^{o}C for approximately 3 hours. Once the OD_{600} reached 0.6, the expression was induced with IPTG at the final concentration of 0.1 mM, and the cultures were then moved to 18 ^{o}C for overnight shaking at 200 rpm.

The cells were harvested by centrifugation at 5422 ×*g*, 4 ^{o}C for 40 minutes. Cell pellet from each 500 mL culture were then resuspended in 25 mL of His-binding buffer (500 mM NaCl, 20 mM Tris-HCl, 5 mM imidazole, 10% glycerol, pH 7.9). The cell resuspension was then lysed by sonication (65% amplitude) for 8 minutes with a 3-second interval every second on ice. The cell lysate was then centrifuged at 51428 ×*g*, 4^{o}C for 60 minutes. Every 25 mL of supernatant was collected and loaded onto 0.5 mL Nickel(II)-agarose slurry, which had been pre-conditioned with 10 mL His-binding buffer, without disrupting the resin. The resin was then washed twice with 4 mL wash buffer (500 mM NaCl, 20 mM Tris-HCl, 50 mM imidazole, 10% glycerol, pH 7.9). The protein of interest was then eluted by 10 mL elute buffer (500 mM NaCl, 20 mM Tris-HCl, 1 M imidazole, pH 7.9). The final eluate fraction was then desalted by buffer exchanged to storage buffer (50 mM Tris-HCl, 100 mM NaCl, 10% glycerol, pH8.0) and concentrated by 30k Amicon® ultra-15 for TsaH protein or 10k Amico® ultra-15 for Ulp1, TsaA and TsaI proteins.

4.2.7.4. Size exclusion chromatography for protein purification

TsaH and Ulp1 were concentrated to 2 mL and injected to a pre-conditioned AKTA pure 25M chromatography system connected to a 16/600 Superdex 200pg column (bed volume: 120.637 mL), a UV detector (U9-M). The flow rate was set to 1 mL min^{-1} and the separation was done by passing through 120 mL of the storage buffer described in section 4.2.7.3. Fractions were collected every 1 mL from 10 to 80 mL retention volume by a fraction collector based on retention volume.

4.2.8. *in vitro* assays for thioamidation

in vitro assays were carried out in a buffer that had a pH of 7.5 and contained 50 mM Tris-HCl, 125 mM NaCl, and 20 mM $MgCl_2$. The compositions of the reactions are specified in the results section. The reactions were incubated in 30 ^oC overnight. At the end of the reaction, materials were desalted by C18 ZipTips (EMD Millipore) and eluted with 10 μL ACN for proteomics. The desalted samples were analysed by LC-MS/MS using Acquity UPLC (Waters) coupled with Synapt G2-Si mass spectrometer.

4.2.9. Screening for *thioamitide* production

The actinomycete strains was firstly inoculated in 10 mL TSB in a 50-mL Falcon® tube with a sponge stopper at 30 ^oC, 250 rpm for 5 days to gain biomass. 1 mL of the inoculum was plated on a 90-mm petri dish containing 25 mL screening medium with 2% agar. The plates were then grown at 30 ^oC for 14 days. During the 14-day period. Three time points of samples were taken at day 5, 9 and 14. Sampling were achieved by using the wide end of P1000 pipette tips to extract a plug from the production culture. Triplicate samples were taken for each strain at each time point for LC-MS sample preparation. Each plug was mixed

with 1 mL methanol and shaken vigorously for 1 hour. After a 20-minute centrifuge at 20000 ×*g*, the 500 µL of the supernatant was taken for LC-MS analysis.

4.2.10. Purification of thiopotensamide A

4.2.10.1. Fermentation of *N. potens*

100 µL of *N. potens* mycelium stock was inoculated in two 50 mL TSB at 30 °C, 250 rpm for 3 days. The mycelium culture was concentrated by removing 40 mL of the supernatant after centrifuging at 3472 ×*g* for 10 minutes. Each of the 1 mL concentrated mycelium was then transferred to a 245 mm × 245 mm × 25 mm plate containing 200 mL solid SM12. The plates were then grown in 30 °C for 14 days.

4.2.10.2. Extraction of thiopotensamide A for purification

The production culture was extracted twice with equal volume of methanol. The methanol extract was filtered through a filter paper and had the solvent removed on a rotary evaporator at 30 °C. Every batch of dried materials from a 2-liter culture was resuspended in 500 mL of water and the container were rinsed with 500 mL ethyl acetate. Both aqueous and ethyl acetate fraction was loaded onto a separation funnel to perform a liquid-liquid extraction. The aqueous layer was returned to the funnel and further extracted with equal volume of ethyl acetate twice. The collected ethyl acetate fractions were then dried in rotary evaporator at 30 °C to produce a brown solid.

4.2.10.3. Flash chromatography purification

The brown solid was redissolved in minimal volume of methanol. The materials were mixed and dried with twice weight of silica gel (Davisil Grade 635, pore size 60 Å, 60-100 mesh). The crude extract and silica gel mixture was then loaded onto a BioTage® SNAP cartridge KP-Sil 50g for flash chromatography. The flash chromatography was performed by a

BioTage Isolera system using methanol (solvent B) against chloroform (solvent A) with a flow rate of 100 mL/min. The separation began with isocratically flushing 1 column volume (CV: 90 mL) of 10% B, then two gradients from 10 to 50% B over 10 CVs followed by 50 to 90% B over 2 CVs. Finally, the separation finished with a 90% B isocratic hold for 2 CVs. Cartridge was flush with 2 CVs of 100% B, followed by 2 CVs air flushing to dry. Elution was monitored at 254 nm and 270 nm, and each fraction was 22 mL. The flush solvent and solvent residue from air flushing were also collected. All collected fractions were sampled and analysed by LC-MS/MS. Fractions containing thiopotensamide A were combined and dried by rotary evaporation.

4.2.10.4. Cation exchange for purification

1.0 g of cation exchange resin (AmberChrom® 50WX4 hydrogen form, 200-400 mesh) was washed with 15 mL methanol. 1.5 mg of the brown solid from liquid-liquid extraction was dissolved in 40 mL methanol. The materials in methanol were carefully loaded onto the column containing the conditioned resin without disturbing the resin. The materials flowed through the resin via gravity. The resin was then washed 8 times with 5 mL methanol, and the solvent pass through the column by centrifuge at 500 ×*g* for 1 minute. The wash step was repeated with 8 × 5 mL 10 mM $HCl_{(MeOH)}$, 4 × 5 mL 30 mM $HCl_{(MeOH)}$, 4 × 5 mL 60 mM $HCl_{(MeOH)}$, 4 × 5 mL 120 mM $HCl_{(MeOH)}$, 4 × 5 mL 300 mM $HCl_{(MeOH)}$ and 4 × 5 mL 600 mM $HCl_{(MeOH)}$. Finally, the resin was stripped by passing 10 mL 1.2 M $HCl_{(MeOH)}$ via gravity. Each of the fraction had a 10 times concentrated sample prepared for HR-MS/MS.

4.2.10.5. Size exclusion chromatography

25 mg of materials from flash chromatography was dissolved in 2 mL methanol and filtered using a 0.45 μm PTFE filter (Whatman). The filtered material was then injected in to a pre-

conditioned AKTA pure 25M chromatography system connected to a Sephadex LH-20 column (XK 26 mm × 100 cm, bed volume: 477.6 mL), an external refractive index (RI) detector (Optilab, Wyatt Technology) and a UV detector (U9-M). The flow rate was set to 1 mL min^{-1} and the separation was done by passing through 1 L of methanol. Fractions were collected from 200 to 500 mL retention volume by a fraction collector (F9-R) based on retention volume. Fractions were sampled and analysed by HR-MS/MS. The results of separations were viewed in software Unicorn 7.6 from Cytiva Fractions containing thiopotensamide A were pooled and dried by GeneVac.

4.2.10.6. Preparative-scale HPLC

Flash chromatography fractions were redissolved in 80% methanol to obtain a sample with a concentration of approximate 2 mg mL^{-1}. The sample was filtered by a 0.45 μm PTFE filter. 2 mL of the filtered sample were injected to an Agilent 1260 HPLC system, which consisted of a preparative binary preparative pump (Agilent 1260 Prep Pump G7161A), a high-performance preparative auto-sampler (Agilent 1290 Prep HiP ALS G7169B) to be separated by a 250 × 21.2 mm Kinetex® 5 μm XB-C18 (pore size 100 Å) column from Phenomenex in a preparative column compartment (Agilent 1290 Prep Column Comp G7163B). The solvent system used ACN (+ 0.1% FA) as solvent B and H_2O (+0.1% FA) as solvent A. The separation was achieved by running a 12-minute gradient from 20 to 70% solvent B and the flow rate was 20 mL min^{-1}. The separation was monitored by a quadruple MS (LC/MSD G6125B – G1948B). MS data was obtained in a positive ionisation mode, and the detection range was set to *m/z* range 100 -1500 with a target *m/z* 1243. Ten 7 mL Fractions were collected from 10 minutes to 13.5 minutes by a preparative fraction collector (Agilent 1290 Prep HiP FC G7159B). Data were analysed with the software

ControlPanel from Agilent. Fractions were sampled and analysed by HR-MS/MS. Fractions containing thiopotensamide A were pooled and dried by GeneVac.

4.2.10.7. Analytical-scale preparative HPLC

Materials from preparative HPLC fractions were dissolved in 80% methanol with a concentration between 1.0 and 2.0 mg mL^{-1} to be polished using an Agilent 1290 HPLC system. The system consisted of a dual-needle multisampler (Agilent 1290 multisampler G7167B), analytical scale fraction collector (Agilent 1260 FC-AS G1364F). Materials were separated by a 250 × 10 mm Luna® 5 μm PFP(2) column (pore size 100 Å) from Phenomenex, which was kept in a column oven (Agilent 1290 MCT G7116B) that set to 40 °C. Solvent A (0.1% FA in water) and solvent B (0.1% FA in MeOH) were delivered to the column by a binary pump (Agilent 1290 High Speed Pump G7120A), and the flow rate was set to 3 mL min^{-1}. The fraction collection was monitored by UV at 270 nm using a diode-array detector (Agilent 1290 DAD FS G7117A), and the separation was monitored by a quadrupole MS (Agilent LC/MSD G6125B – G1948B). MS data was obtained in a positive ionisation mode, and the detection range was set to *m/z* range 100 – 1500 with a target *m/z* 1243. Fractions were sampled and analysed by HR-MS/MS. Fractions containing thiopotensamide A were pooled and dried by GeneVac.

4.2.11. HPLC-HR-MS/MS systems

In this project, three HPLC-HR-MS systems were used, depending on the availability. Unless otherwise stated, 10 μL of the sample was injected for LC analysis, and the flow rate was set to 600 μL min^{-1}. There were two analytical columns used in the project, so two mobile phase gradients were used, unless specifically indicated.

When a 50×21 mm Kinetex® 2.6 µm XB-C18 column (pore size 100 Å) from Phenomenex was applied, the chromatography was done by a 6-minute gradient of combinations of 0.1% FA in water (solvent A) and 0.1% FA in ACN (solvent B) from 5% to 95% B. At the end of the gradient, the column was washed by passing 1.2 mL 95% B and reconditioned by passing 1.2 mL 5% B.

When a 50 × 2 mm Luna® 3 µm PFP(2) column (pore size 100 Å) from Phenomenex was used, combinations of 0.1% FA in water (solvent A) and 0.1% FA in methanol (solvent B) was used as mobile phase. The gradient started from 40% B isocratically for 0.5 minutes, followed by a 6-minute gradient going from 40% to 70% B. After the gradient separation, the column was washed with 95% B for 1.5 minutes, then reconditioned with 40% B for 1.5 minutes for the next sample.

4.2.11.1. Shimadzu IT-ToF system

A Shimadzu Nexera X2 ultrahigh performance liquid chromatography (UHPLC) system, which consisted of a solvent delivery module (LC-30AD), auto-sampler (SIL-30AC), was employed. A 50×21 mm Kinetex® 2.6 µm XB-C18 column (pore size 100 Å) from Phenomenex was used in the HPLC and kept in the column oven (CTO20A/CTO-20AC) that was set to 40 °C. The LC analysis was monitored by a Shimadzu ion-trap-time-of-flight (IT-ToF) tandem mass spectrometer. MS data was obtained in positive ionisation mode and the detection range was set to *m/z* 500 – 2000, unless otherwise stated. Data were analysed with the software packages, LabSolution and ProfilingSolution, from Shimadzu.

4.2.11.2. Thermo Q Extractive system

A Thermo scientific Vanquish UHPLC system, which consisted of a binary pump (VF-P10), split samplers (VF-A10), column compartments (VH-C10) and a diode-array detector (VF-

D11) was used. Columns were kept in the column compartment to retain at 40 °C. The LC analysis was also monitored by a Themo Scientific Q Exactive orbitrap MS, and the detection range was set to *m/z* 150 – 2000 under positive ionisation mode, unless otherwise stated. Data were analysed with software FreeStyle from Thermo Fisher.

4.2.11.3. Agilent Q-ToF system

Agilent quadrupole-time-of-flight (Q-ToF) tandem mass spectrometer (G6546) was used to analyse the chromatographic separation from an Agilent 1290 Infinity II HPLC system, which consisted of a multisampler (G7167B), column compartment (G7167B), diode-array detector (G7117B) and a binary pump (G7120A). The column compartment was set to 40 °C to obtain a consistent separation temperature. The detection for Q-ToF was set to positive ion polarity and auto MS/MS segmentation, and the *m/z* range was set to 100 – 1700 in both MS and MS/MS. Data were analysed with software Qualitative Analysis from MassHunter package, Agilent.

4.2.11.4. HR-MS/MS for fragmentation on Synapt

High-resolution mass spectra for targeted compound fragmentations were acquired from Acquity UPLC (Waters) – Synapt G2-Si mass spectrometer. The samples were sent to Dr Carlo de Oliveira Martins (John Innes Centre) to obtain LC-MS/MS data. The MS data were collected in positive mode with the following parameters: capillary voltage = 2.5 kV; cone voltage = 40 V; source temperature = 125 °C; desolvation temperature = 300 °C. Leu-enkephalin peptide was used to generate a lock-mass calibration with *m/z* = 556.2766 measured every 20 s during the run. The compounds of interest had their masses put on an inclusion list and fragmented by data directed analysis (DDA) method with the following

parameters: top 3 precursor ions (inclusion list only); intensity threshold = 10000; scan time = 1 s; isolation window = 0.5 Da; collision energy was ramped between 30 – 40 eV at low mass (*m/z* 50) to 70 – 120 eV at high mass (*m/z* 1600).

4.3. DNA sequences

4.3.1. Codon optimised synthetic DNA

>Codon optimised YcaO domain protein from S. sp. S-4
ATTGAGGCTCACAGAGAACAGATTGGTGGTGTAAAACTTCGTCAATCGGTGCCAAAGATT
ACTGGTAATGGGCCTCATCGCGAGTATACCAGTGAAACAACATATCGTCGTATTGCACCC
CATTTACGCCGCGTAGGAGTGACGCGCGTTGCAAATATTACGGGACTTGATCGTGTGGGT
ATTCCGGTGTATAATGCCATTAGCCCGCGTAGTTTTGATATTATTTCTGTGTATAATGGT
AAAGGGGCGACGGAAATTGATGCGAAAACGAGCGCAGTAATGGAAGCGGTAGAGCGCTTT
GCGGCGGGCTTGCCAATGCGCCCAGTTGCAGTGGCTAGTTATACAGAATTAGAAGCAAGT
GGTCGTCGTGTCTTGGATCCTCGTTCGTGTAATTTTGAATTCCACCGTGCTTATACCCTT
GATACTCCCATTTGTTGGGTTGAAGGTTTTGATCTTTTAAATGAATGTGCGGTTCTGGTT
CCGCAATTCTTAGCTGGTTATTATTTAAATTTCCACGAAGTTCCATGTTATCCGGTAAGC
ACGACGAACGGTATTGCCTCTGGTAATAGCCTTGAAGAAGCAATTTGCCATGCGTTATGT
GAAGTTATTGAACGTGATGATTGGACAATGGCGGATTTGGTTAGCAATCGTCTCGCAGGA
TTTGCGGCTCAAGAAGATGGTACTCCGATTGGAACCGCGATTTCGGCGTGGGCGCAAGAT
TTACATCCGTCGGTGGATCCTGATTCGTTACCAGAACGTGCGCAATTCTTTATGGATAAA
TTTCACGCTGCAGAAATTGATATTGAAGTCAAGAATATTACTTCAGCGACTGGTATTCCG
TCATTCCAAGCCGTAGTATCAGAATTTATCTCTGATAGTTTTAGTCGTTCGCATCATGGT
ATTGGCACACACCCTGACGCTGAAGTGGCTGTAGTACGTGCTATTGCTGAAGTGGCACAA
AGTCGCGTTGTAGACATTCAAGCGATGCGCGAAGATATTACGGCAGCTGGTGAACCCGTC
GAAAAGGCCTTATATCATGTTCGCCGTTCCAATCGCGCTAACCCAGACGCTTGGGCTATT
AAACCTAGCAGTCGCCCTGTACGCCTTCAAGATGTCCCAACTCATCCTAGTAGCGATGTG
ATGGATGACGTTCGTACACTTTTGGATCGTCTGCGTGCGCGTGGATTAGATCAAGCTATT
GTAGTAGATCTTTCACCCCCAGCAGTGCCGGCAACGGTAGTCCGTGTAATTGTTCCTGGA
GCCGAAAGCTGGACGGTGGATCGTAATAAAATTGGGCCTCGCGCCCGTGAGGCCTGGGAT
GGGACTCTCAATACTATTCGTTCTGCGTACGCCAAACGCCAACAACGTACTACTGGAGAA
CAAGAAAGTAATGCTTAAGCATCCCCACCACCGTCGGTGTGAACGCACCACGCCTATCCG
TCTTCCTCGGCCCCAGCCTGCCTGTCGAGCAGGCACGGGAGCACCTGCCCGGCGCCGAGT
TCCTGCCGCCCGTCGAGCGCTTTGTTAGCAGCCGGATCCCCTTCCTGCA

4.3.2. Promoters for refactoring

>Upstream UTR of thiovarsolin precursor peptide
CGATGGGCTCTCCTCTGGGGCTTCGAAAGGACCGGGGCAGCAGCATCGGGGACGGGACGG
GGGGTGAGGGCCGGGCGAATGATGGGCCGATGACGCGCCGAAGAATGGCGGATTTTTCTT
TATACAGAACGCGTGGGACGTTCGGCAATGCCCGATCCGAAGATCACAACTCCCCGGTAC
GCCCCCCCTGTCATCCCGCCGCCTCCTTTTGATCTTCGACCGTCGGCCGTTCGGATTCCT
CCGCCAAAGGACGGCACACGGCGGAAACCGCCGACCGTGCAACATGGCTGAAATATTGCG
CCCCGGGATCGACACCCCTTGGAGGGTCCTTGCGTGACCCCTTAGGCGACCCTTGAGACG
ACACCCCGGGCGCCAACTACACCATGCTGCCCGACTGTTGACTAGTCTTGCCCGCTGTGC
CATCGTGGCGAAATGGCCTCCTGTCGGGGGGAATGACCGGAGGGAATATCTCCCCTTCCG
TCATCCGGGGCGACCGAGTAATTGCGCGTTCCGTTCGGTCTCCATGTGCGCCCGGCCGAC
AATAGGACAAACAGCGGCGTGAGAACTCGCGCACAACTCGGCAGAGGCGGAGTCGGAACG
ACGTTCCTGTGCGGGACAGTGCGTATTACCGGTCACTGGTTACCGGTTACTGGTTACCGG
TTGCCGACGCTGTCGACCGTGCCCGGCCTCCAAACCCGTAAGGAAAGGAAAAGGCCCTC

```
>PhrdB promotor
CCGCCTTCCGCCGGAACGGCGGGGTCCGGGCACGCCAAACCCCTCCTGTGGCTGTGGCCG
GCCACCGCCGTCACCTTCGGACCCCGTGGAGCCGCTCCCGGTTCCACGGGGTCCGAAGGT
GTGATGAGCAGGCTGCGCCTTCCTCGCGCGGCCGCAAGGTACGAGTTGATGACCTTGTTT
ATCCGCATCTGACCAATTTTGATCGCTTACGGGGTGTGACTCGGGCCACGCGGATTGGGC
GTAACGCTCTTGGGAACAACACGATGACCTAAGAGGTGACAGCCGCGGAGGGAATACGGA
CGCCGTTCACGGCGCTGTGCATCTCCCCGGCCCGCCCGCACCGTCGGCCCATTCCCAAGC
CGGTGGTCGGCCCCTGTCCGCCGTGGACGGGGCCGGAAGCCGTTTTTCAACGTTCCGAGA
GGTTGTTCATG
```

4.3.3. Primers

Table 4.7: Primers and oligonucleotides used in the project

Primers /Oligonucleotide	**Sequence (5' - 3')**
NF-H-Start	GGCCTGGTGCCGCGCG GCAGCCATATGAAGCTG CGCACCCACG
NF-H-End	GTGCGGCCGCAAGCTTG TCGACGGAGCTCCTGGA TCGGCGGCAGGTAC
AA-H-Start	GGCCTGGTGCCGCGCG GCAGCCATATGCATATC CGCGAAAACGC
AA-H-End	GTGCGGCCGCAAGCTTG TCGACGGAGCTCTCGAC GATGCCGATCACC
S87-H-Start	GGCCTGGTGCCGCGCG GCAGCCATATGAAGCTG CGTTCCAGCGTC
S87-H-End	GTGCGGCCGCAAGCTTG TCGACGGAGCTCGCTTG TGCCATCAGGGCGTC
ME-H-Start	GGCCTGGTGCCGCGCG GCAGCCATATGAAACTA CGCGCCAGCG
ME-H-End	GTGCGGCCGCAAGCTTG TCGACGGAGCTCTCGAC AGGAACTTGCCGTCG
NP-H-Start	GGCCTGGTGCCGCGCG GCAGCCATATGAAGCTC CGGAACCGGATC
NP-H-End	GTGCGGCCGCAAGCTTG TCGACGGAGCTCGTGCA GGAACTGGCCGTCG
Seq_NF001	GCACAGTTGATCTACGCCTTCCAC
Seq_NP001	CTACTACTCCCGCTACCACGAG
Seq_NF002	TACCGCATCACCACCACCAAC
Seq_NF003	CCTCAGCGTGGACCTCGAAC
Seq_NF004	GTCTTCCTTCGGCTGATCGC
Seq_ME001	CAACTCTGTCAGGTCCTGGAAAAG
KW001	GGCTCACAGAGAACAGATTGGTGGT ATGCATATCCGCGAAAACGC
KW002	CAGCCGGATCCCCTTCCTGCA TCGACGATGCCGATCACCG
KW003	GGCTCACAGAGAACAGATTGGTGGT ATGAAACTACGCGCCAGCG
KW004	CAGCCGGATCCCCTTCCTGCA TCGACAGGAACTTGCCGTCG
KW005	GGCTCACAGAGAACAGATTGGTGGT ATGAAGCTGCGCACCCACG
KW006	CAGCCGGATCCCCTTCCTGCA CTGGATCGGCGGCAGGTAC
KW007	GGCTCACAGAGAACAGATTGGTGGT ATGAAGCTCCGGAACCGGATC
KW008	CAGCCGGATCCCCTTCCTGCA GTGCAGGAACTGGCCGTCG
KW009	GGCTCACAGAGAACAGATTGGTGGT ATGAAACTGCGCCAGAGCGTC
KW010	CAGCCGGATCCCCTTCCTGCA GATGTCACCACGCTCGACGG
KW011	GGCTCACAGAGAACAGATTGGTGGT ATGAAGCTGCGTTCCAGCGTC

KW012	CAGCCGGATCCCCTTCCTGCA GCTTGTGCCATCAGGGCGTC
KW013	ACCACCAATCTGTTCTCTGTGAGCC
KW014	TGCAGGAAGGGGATCCGGCTG
KW015	GGCTCACAGAGAACAGATTGGTGGT ATGCATATCCGCGAAAACGC
KW016	CAGCCGGATCCCCTTCCTGCA TCGACGATGCCGATCACCG
KW017	CCGCAAGCTTGTCGACGGAGTCGACGATGCCGATCACC
KW018	GATTCTTGTACGACGGTATTAGAATTCAAG
KW019	ATTGAGGCTCACAGAGAACAGATTGGTGGTATGAAACTACGCGCCAGCG
KW020	CTTTGTTAGCAGCCGGATCCCCTTCCTGCATCGACAGGAACTTGCCGTCG
KW021	CTTTGTTAGCAGCCGGATCCCCTTCCTGCAGTACTCCGCGTCGAGTTCCTC
KW022	ATTGAGGCTCACAGAGAACAGATTGGTGGTATGAAGCTGCGCACCCACG
KW023	CTTTGTTAGCAGCCGGATCCCCTTCCTGCACTGGATCGGCGGCAGGTAC
KW024	ATTGAGGCTCACAGAGAACAGATTGGTGGTATGAAGCTCCGGAACCGGATC
KW025	CTTTGTTAGCAGCCGGATCCCCTTCCTGCAGTGCAGGAACTGGCCGTCG
KW026	CTTTGTTAGCAGCCGGATCCCCTTCCTGCATCCGGGCGGTTCATCAGCTC
KW027	ATTGAGGCTCACAGAGAACAGATTGGTGGTATGAAGCTGCGTTCCAGCGTC
KW028	CTTTGTTAGCAGCCGGATCCCCTTCCTGCACGTTTCCGGTGATCTTGGGG
KW029	GGCCTGGTGCCGCGCGGCAGCCATATGAGCATCCCCACCACCGTC
KW030	GTGCGGCCGCAAGCTTGTCGACGGAGCTCGTAAGTAGGTGCGGCAGTCGGTC
KW031	GGCCTGGTGCCGCGCGGCAGCCATATGTCTGAAACCACGACAGCAGC
KW032	GTGCGGCCGCAAGCTTGTCGACGGAGCTCCAATGATTCGCCCGTGATCC
KW033	AGGATCGTCTAGAACAGGAGGCCCCATATGCACCACCACCACCACCATATG
KW034	GCTCATGAGAACCCTAGGGGATCCAAGCTTGCAGCCAACTCAGCTTCCTTTC
KW035	AATTTCACACAGGAAACAGAATTCGAGCTCCTCTAGAAATAATTTTGTTTAACTTTAAG
KW036	CTCACTGATCCGCTAGTCCGAGGCCTCGATGTCACCACGCTCGACG
KW037	AGGATCGTCTAGAACAGGAGGCCCCATATGCTCTAGAAATAATTTTGTTTAACTTTAAG
KW038	GCTCATGAGAACCCTAGGGGATCCAAGCTTGATGTCACCACGCTCGACG
KW039	TCTTGACGGCTGGCGAGAGGTGC
KW040	GCTCACAGAGAACAGATTGGTGG
KW041	CCAGTTTCTCGCCGGGTACTAC
KW042	GAACTTGTCCATGAAGAACTG
KW043	ATGGGGCCTCCTGTTCTAGACGATCCT
KW044	AAGCTTGGATCCCCTAGGGTTCTCATGAGC
KW045	ATTTTGTTTAACTTTAATAAGGAGATATACCTCTAGAAATAATTTTGTTTAACTTTAAG
KW046	TTATGCGGCCGCAAGCTTGTCGACCTGCAGGATGTCACCACGCTCGACG
KW047	CAACCTCGGGATCGAGGGAAGGATTTCACAT ATGTCTGAAACCACGACAGCAGC
KW048	TTTATTTGAAGCTTATTTAATTACCTGCAG CAATGATTCGCCCGTGATCC
KW049	GGCTCACAGAGAACAGATTGGTGGT ATGTCTGAAACCACGACAGCAGC
KW050	CAGCCGGATCCCCTTCCTGCA CAATGATTCGCCCGTGATCC
KW051	CAACCTCGGGATCGAGGGAAGGATTTCACAT ATGAAACTGCGCCAGAGCGTC
KW052	TTTATTTGAAGCTTATTTAATTACCTGCAG GATGTCACCACGCTCGACGG
KW053	AGCAGCGGCCTGGTGCCGCGCGGCAGCCAC ATGAAAATCGAAGAAGGTAAACTG

KW054	ATCTCAGTGGTGGTGGTGGTGCTCGAG CAAAAAGGCCATCCGTCAG
KW055	CTCTTTGTTCGGACTGGCGGCG
KW056	CAACCTCGGGATCGAGGGAAGGATTTCACAC ATGGGCAGCAGCCATCATCATC
KW057	CGACTGAGCCTTTCGTTTTATTTGAAGCTT CCGGATATAGTTCCTCCTTTCAGC
KW058	CCGCCTGGCTAATGTTCTG
KW059	CCCCATCGGTCTGTACTTG
KW060	CGATGAAAGCACCACCAAAG
KW061	ATTTTGTTTAACTTTAATAAGGAGATATAC ATGAGCATCCCCACCACCGTC
KW062	TTATGCGGCCGCAAGCTTGTCGACCTGCA GTAAGTAGGTGCGGCAGTCGGTC
KW064	AGGTGCGATAATCCTTTCGCAAAAACTCGGTTTGACGCCTCCCATGGTATAAATAGTGGC GGCGGCATCTCCAGACGTCTAGAAG
KW065	CCGCCAGGACGACCGAGGTGAGCAG
KW066	GCACATCCTGGCCTGGCTCCCGCTC
KW067	CTCCTCGGCGTCGCGGTGCTTCTGG
KW068	GATGTACTTCCCGGACCGCAGCTAC
KW069	GCTTGGCAGCAACAGGACTAGGATGAGTAGCAGCACGTTCCTTATATGTAGCTTTCGACA GGCTCCTCGACTCCTTCCCTCCCGC
KW070	AGGTGCGATAATCCTTTCGCAAAAACTCGGTTTGACGCCTCCCATGGTATAAATAGTGGC CGATGGGCTCTCCTCTGGGGCTTCGAAAGG
KW071	GGCCTCGGCGGCCTTGATCTGCGCGGCGAGCTCTTCGATGTTCTGCGTGACGGCAGGCA TGAGGGCCTTTTCCTTTCCTTACGGGTTTGG
KW072	GTTGCCGACGCTGTCGACCGTGCCCGGCCTCCAAACCCGTAAGGAAAGGAAAAGGCCCT CATGCCTGCCGTCACGCAGAACATCG
KW073	CGGCCACAGCCACAGGAGGGGTTTGGCGTGCCCGGACCCCGCCGTTCCGGCGGAAGGC GGGGAACCCCATCGGTCTGTACTTGGTG
KW074	GCGCACACCGCCCGCGCACCCGGCGCGGCCAGCACACCAAGTACAGACCGATGGGGTT CCCCGCCTTCCGCCGGAACGGCGGGGTC
KW075	GTCGGGGGAGACCACCTTCACCTGTTCGGCGATGCTTCTGAAGACTCTCTCGTACCGGTC CATGAACAACCTCTCGGAACGTTG
KW076	CCCTGTCCGCCGTGGACGGGGCCGGAAGCCGTTTTTCAACGTTCCGAGAGGTTGTTCATG GACCGGTACGAGAGAGTCTTCAGAAG
KW077	CGCCTGTTCGCCTTCCTCACCG
KW078	GCTTGGCAGCAACAGGACTAGGATGAGTAGCAGCACGTTCCTTATATGTAGCTTTCGACA CCCCGTGTTGATCTTGGACTTGTCG
KW079	GCTTGGCAGCAACAGGACTAGGATGAGTAGCAGCACGTTCCTTATATGTAGCTTTCGACA CGACAAGTCCAAGATCAACACAGAGG
KW080	GATCAGCTACGAGTCGGACACCCTG
KW081	CGGCCAGTTCCTGCACGCGCTCTC
KW082	CGAGAGGGTGTGTGAATGACAGCCG
KW083	GGGTCACTCTCTTTCGCCTCCACTC
KW084	GTCGCTGTTCAGGTCGGCGTCGTAG
KW085	GATCTCGTCCTCGGCGGAGCTCACC
KW086	CATCCTGCTGTCGTTCGGCACCG
KW087	CCTGCTGTCCGGCTACGACGACACC
KW088	CCCACTACGGCAGGCTGGACCGCTG
KW089	CTGAAGATTCCCGAAATGTGTCCACTTTG

KW090	CAGGGTCTGCAGCTCGGCCTCGTC
KW091	GTGAGAACTCGCGCACAACTCGGC
KW092	CTCCACGGGGTCCGAAGGTGACG
KW093	CTGCTGATCGGACCGGCGTGACC
KW094	CGAGGGAGGAACAACCTGATGACCG
KW095	CTCCTTCGCTGGGCTTCCGTCGTC
KW096	GATCGTCCCCGGCATCGAGTCG
KW097	GAAGTGGAAGCCGTCCCGCCGC
KW098	CGGTCGACAAGTCCAAGATCAACACGG
KW099	GGATAATGCCTTTAGCGGCTTAACTGTGC
KW100	GGTAGGATCGTCTAGAACAGGAGGCCCCAT GGCGGCATCTCCAGACGTCTAGAAG
KW101	CTCGTCCTCGGCCTTCAGCGACGGGGTCAC GGAACCCCATCGGTCTGTACTTGGTG
KW102	CAGCACACCAAGTACAGACCGATGGGGTTCC GTGACCCCGTCGCTGAAGGC
KW103	GCTCATGAGAACCCTAGGGGATCCAAGCTT GGCTCCTCGACTCCTTCCCTCC
KW104	GGTAGGATCGTCTAGAACAGGAGGCCCCATA TGACCCCGTCGCTGAAGGC
KW105	GCTCATGAGAACCCTAGGGGATCCAAGCTT GGCTCCTCGACTCCTTCCCTCC
KW106	GTCCTCCTCGCCCTTGGAGACCATCTCGAGCCTCCT CAAGATCAACACAGAGGCGGGGC
KW107	GCACCGCCCCGCAGCACACC
KW108	CATCGGTCTGTACTTGGTGTGCTGGC
KW109	CGCTCGTTGTCGCTGTTCAGGTC
KW110	GTCCACTTCGACCTCAGCCCG
KW111	CTCTACCTCCACTACCACGCCCG
KW112	CGCCCTCGCCCTCGATCTCG
KW113	GAAAGACAATCCCCGATCCGC
KW114	GGTAGGATCGTCTAGAACAGGAGGCCCCA ATGACCGACAACGACTCCG
KW115	GCTCATGAGAACCCTAGGGGATCCAAGCTT TCAAGCTCCGTCGTCGTCTC
KW116	GGTAGGATCGTCTAGAACAGGAGGCCCCATA TGCTCACCATTGGAGAGGCCG
KW117	GCTCATGAGAACCCTAGGGGATCCAAGCTT TCAGGGCGCGGGGCGC
KW118	GGTAGGATCGTCTAGAACAGGAGGCCCCATATG AGACCCGGAGCGGGAAATCAAG
KW119	GCTCATGAGAACCCTAGGGGATCCAAGCTT CTAGGCGGCGGCCCGCTG
KW120	GGTAGGATCGTCTAGAACAGGAGGCCCCAT ATGGCCGACACTGAATTCGATATC
KW121	GCTCATGAGAACCCTAGGGGATCCAAGCTT TCAGACCGGGGTCCCGTG
KW122	GGTAGGATCGTCTAGAACAGGAGGCCCCATA TGTACGGAGCGGAATTCACC
KW123	GCTCATGAGAACCCTAGGGGATCCAAGCTT TCAGTCTTCGGTGGACTCGG
KW124	GGTAGGATCGTCTAGAACAGGAGGCCCCATA TGAGATACGAGGTTCTTGGTTCGCTG
KW125	GCTCATGAGAACCCTAGGGGATCCAAGCTT TCAGGAGGCCAGGCGGTCC
KW126	GAAGGCGTCGGTGCACGCCGCGAGTGGAGGCGAAAGAGA ATTCCGGGGATCCGTCGACC
KW127	TCTCCTCCGCGATCCGCTCCAGGGCCCGGAGGGGCGCGC TGTAGGCTGGAGCTGCTTC
KW128	ACACGCTCTTCGGGCCCGGACTGCTCCG ATTCCGGGGATCCGTCGACC
KW129	CGCAGGTGAGCACTCCTTGACCGGTCGACAAGTCCAAGA TGTAGGCTGGAGCTGCTTC
KW130	CTGAAGGCCGAGGACGAGAGCTCCCTGCGGCGCTGGATC ATTCCGGGGATCCGTCGACC

KW131	CGGGCCCGAAGAGCGTGTCCGCCATGGCCTCGGGGTCGC TGTAGGCTGGAGCTGCTTC
KW132	GGGCCCTGGAGCGGATCGCGGAGGAGACCCGGGTCCTC ATTCCGGGGATCCGTCGACC
KW134	GGAGACCAGCGACGACCG
KW135	CACGTTGAAGGTGCTGGC
KW136	GCGGCTCCCACCTGCTCG
KW137	CCCGAGCAGGCCCCCGAG
KW138	GCACCAGGGCGTAGGGG
KW139	CAGGATGTGCGGGACGGC
KW140	GACGACCTCACCGCCCCG
KW141	CGGTGCGCGGTGCTCAC

Abbreviations

Abbreviation	Definition
5-FOA	5-fluoroorotic acid
ACN	Acetonitrile
ACP	Acyl carrier protein
ADP	Adenosine diphosphat
antiSMASH	antibiotic and Secondary Metabolite Analysis Shell
AT	Acyltransferase
ATP	Adenosine triphosphat
Avi(Me)Cys	*S*-[(*Z*)-2-aminovinyl](-(3S)-3-methyl)-d-cysteine
BPC	Base peak chromatogram
Cas	CRISPR associated
CIPRES	Cyberinfrastructure for Phylogenetic Research
CRISPR	Clustered regularly short palindromic repeats
crRNA	CRISPR RNA
CV	Column volume
Da	Dalton
Dha	2,3-dehydralanine
DN agar	Difco nutrient agar
DN broth	Difco nutrient broth
DNA	Deoxyribonucleic acid
DSB	Double-strand break
DTT	Dithiothreitol
E1cB	Elimination, unimolecular, conjugate base
EDTA	Ethylenediaminetetraacetic acid
EMBL-EBI	European Molecular Biology Laboratory - The European Bioinformatics Institute
FA	Formic acid
FAD	Flavin adenine dinucleotide
FMN	Flavin mononucleotide

gDNA	Genomic DNA
GYM	Glucose yeast extract malt extract medium
hdmHis	β-hydroxy-N^1, N^3-dimethylhistidinium
HEPES	2-[4-(2-hydroxyethyl)piperazin-1-yl]ethanesulfonic acid
His_6-tag	Polyhistidine tag
HPLC	High preformance liquid chromatograph
HR-MS	High resolusion mass spectrometer
IPTG	Isopropyl β-D-1- thiogalactopyranoside
iTOL	Interactive Tree Of Life
IT-ToF	Ion-trap-time-of-flight
KS	Ketosynthase
L	Lennox broth
LB	Lysogeny broth
LC	Liquid chromatography
MBP	Maltose binding protein
MOPS	3-morpholinopropane-1-sulfonic acid
MS	Mass spectrometery
MS/MS	Tandem mass spectrometry
MUSCLE	MUltiple Sequence Comparison by Log-Expectation
NCBI	National center for Biotechnology information
NEB	New England BioLab
NMR	Nuclear magnetic resonance
NRP	Non-ribosomal peptide
OD600	Optical density at 600 nm
oriT	Origin of transfer
PCR	Polymerase chain reaction
PEG	Polyethylene glycol
pI	Isoelectric point
PTM	Post-translational modification
Q-ToF	Quadrupole-time-of-flight
RBS	Ribosom binding site
RiPP	Ribosomally synthesized and post-translationally modified peptide
RiPPER	RiPP Precursor Peptide Enhanced Recognition
RODEO	Rapid ORF Description and Evaluation Online
RREs	RiPP recognition elements
S/N	Signal to noise ratio
SAM	*S*-adenosyl-L-methionine
SDS	Sodium dodecyl sulphate
SDS-PAGE	Sodium dodecyl sulphate polyacrylamide gel electrophoresis
SFM	Soya flour mannitol
sgRNA	Synthetic guide RNA
SOC	Super optimal broth with catabolite repression
SUMO	Small ubiquitin-like modifier
TAR	Transformation-associated recombination

TEAB	Triethylammonium bicarbonate
TIC	Total ion chromatogram
tracrRNA	Trans-activating crRNA
TSB	Tryptic soy broth
UTR	Untranslated region
XIC	Extracted ion chromatogram
YPD	Yeast extract peptone dextrose medium

References

ACKER, M. G., BOWERS, A. A. & WALSH, C. T. 2009. Generation Of Thiocillin Variants By Prepeptide Gene Replacement And *in vivo* Processing By *Bacillus cereus*. *Journal of the American Chemical Society,* 131**,** 17563-17565.

AHMED, Y., REBETS, Y., ESTÉVEZ, M. R., ZAPP, J., MYRONOVSKYI, M. & LUZHETSKYY, A. 2020. Engineering Of *Streptomyces lividans* For Heterologous Expression Of Secondary Metabolite Gene Clusters. *Microbial Cell Factories,* 19**,** 5.

ANAND, V., KANDARAPU, R. & GARG, S. 2001. Ion-Exchange Resins: Carrying Drug Delivery Forward. *Drug Discovery Today,* 6**,** 905-914.

ANDERSEN, K. R., LEKSA, N. C. & SCHWARTZ, T. U. 2013. Optimized *E. coli* Expression Strain LOBSTR Eliminates Common Contaminants From His-Tag Purification. *Proteins: Structure, Function, and Bioinformatics,* 81**,** 1857-1861.

ANISZEWSKI, T. 2007. Alkaloids - Secrets Of Life. *Alkaloids - Secrets of Life*.

ARNISON, P. G., BIBB, M. J., BIERBAUM, G., BOWERS, A. A., BUGNI, T. S., BULAJ, G., CAMARERO, J. A., CAMPOPIANO, D. J., CHALLIS, G. L., CLARDY, J., COTTER, P. D., CRAIK, D. J., DAWSON, M., DITTMANN, E., DONADIO, S., DORRESTEIN, P. C., ENTIAN, K.-D., FISCHBACH, M. A., GARAVELLI, J. S., GÖRANSSON, U., GRUBER, C. W., HAFT, D. H., HEMSCHEIDT, T. K., HERTWECK, C., HILL, C., HORSWILL, A. R., JASPARS, M., KELLY, W. L., KLINMAN, J. P., KUIPERS, O. P., LINK, A. J., LIU, W., MARAHIEL, M. A., MITCHELL, D. A., MOLL, G. N., MOORE, B. S., MÜLLER, R., NAIR, S. K., NES, I. F., NORRIS, G. E., OLIVERA, B. M., ONAKA, H., PATCHETT, M. L., PIEL, J., REANEY, M. J. T., REBUFFAT, S., ROSS, R. P., SAHL, H.-G., SCHMIDT, E. W., SELSTED, M. E., SEVERINOV, K., SHEN, B., SIVONEN, K., SMITH, L., STEIN, T., SÜSSMUTH, R. D., TAGG, J. R., TANG, G.-L., TRUMAN, A. W., VEDERAS, J. C., WALSH, C. T., WALTON, J. D., WENZEL, S. C., WILLEY, J. M. & VAN DER DONK, W. A. 2013. Ribosomally Synthesized And Post-Translationally Modified Peptide Natural Products: Overview And Recommendations For A Universal Nomenclature. *Natural product reports,* 30**,** 108-160.

BALDWIN, L. E. & OSHEROFF, N. 2005. Etoposide, Topoisomerase II And Cancer. *Current Medicinal Chemistry - Anti-Cancer Agents,* 5**,** 363-372.

BARTLETT, J. G. 1982. Chloramphenicol. *Med Clin North Am,* 66**,** 91-102.

BEHNSEN, J. & RAFFATELLU, M. 2016. Siderophores: More Than Stealing Iron. *mBio,* 7**,** e01906-16.

BENDEZÚ, F. O., HALE, C. A., BERNHARDT, T. G. & DE BOER, P. A. J. 2009. RodZ (YfgA) Is Required For Proper Assembly Of The Mreb Actin Cytoskeleton And Cell Shape In *E. coli*. *The EMBO Journal,* 28**,** 193-204.

BENJDIA, A., BALTY, C. & BERTEAU, O. 2017. Radical SAM Enzymes In The Biosynthesis Of Ribosomally Synthesized And Post-Translationally Modified Peptides (Ripps). *Frontiers in Chemistry,* 5.

BENNETT, M. R., SHEPHERD, S. A., CRONIN, V. A. & MICKLEFIELD, J. 2017. Recent Advances In Methyltransferase Biocatalysis. *Current Opinion in Chemical Biology,* 37**,** 97-106.

BENT, S. 2008. Herbal Medicine In The United States: Review Of Efficacy, Safety, And Regulation: Grand Rounds At University Of California, San Francisco Medical Center. *J Gen Intern Med,* 23**,** 854-9.

BÉRDY, J. 2005. Bioactive Microbial Metabolites. *The Journal of Antibiotics,* 58**,** 1-26.

BETLACH, M. C., KEALEY, J. T., BETLACH, M. C., ASHLEY, G. W. & MCDANIEL, R. 1998. Characterization Of The Macrolide P-450 Hydroxylase From *Streptomyces venezuelae* Which Converts Narbomycin To Picromycin. *Biochemistry,* 37**,** 14937-14942.

BIBB, M. J., WHITE, J., WARD, J. M. & JANSSEN, G. R. 1994. The mRNA For The 23s rRNA Methylase Encoded By The ermE Gene Of *Saccharopolyspora erythraea* Is Translated In The Absence Of A Conventional Ribosome-Binding Site. *Mol Microbiol,* 14**,** 533-45.

BIRD, J. E., MARLES-WRIGHT, J. & GIACHINO, A. 2022. A User's Guide To Golden Gate Cloning Methods And Standards. *ACS Synthetic Biology,* 11**,** 3551-3563.

BLIN, K., SHAW, S., STEINKE, K., VILLEBRO, R., ZIEMERT, N., LEE, S. Y., MEDEMA, M. H. & WEBER, T. 2019. antiSMASH 5.0: Updates To The Secondary Metabolite Genome Mining Pipeline. *Nucleic Acids Research,* 47**,** W81-W87.

BODE, H. B., BETHE, B., HÖFS, R. & ZEECK, A. 2002. Big Effects From Small Changes: Possible Ways To Explore Nature's Chemical Diversity. *Chembiochem,* 3**,** 619-27.

BORODINA, I., KENNY, L. C., MCCARTHY, C. M., PARAMASIVAN, K., PRETORIUS, E., ROBERTS, T. J., VAN DER HOEK, S. A. & KELL, D. B. 2020. The Biology Of Ergothioneine, An Antioxidant Nutraceutical. *Nutrition Research Reviews,* 33**,** 190-217.

BOZHÜYÜK, K. A. J., LINCK, A., TIETZE, A., KRANZ, J., WESCHE, F., NOWAK, S., FLEISCHHACKER, F., SHI, Y.-N., GRÜN, P. & BODE, H. B. 2019a. Modification And De Novo Design Of Non-Ribosomal Peptide Synthetases Using Specific Assembly Points Within Condensation Domains. *Nature Chemistry,* 11**,** 653-661.

BOZHÜYÜK, K. A. J., MICKLEFIELD, J. & WILKINSON, B. 2019b. Engineering Enzymatic Assembly Lines To Produce New Antibiotics. *Current Opinion in Microbiology,* 51**,** 88-96.

BRAUTASET, T., SLETTA, H., NEDAL, A., BORGOS, S. E. F., DEGNES, K. F., BAKKE, I., VOLOKHAN, O., SEKUROVA, O. N., TRESHALIN, I. D., MIRCHINK, E. P., DIKIY, A., ELLINGSEN, T. E. & ZOTCHEV, S. B. 2008. Improved Antifungal Polyene Macrolides Via Engineering Of The Nystatin Biosynthetic Genes In *Streptomyces noursei*. *Chemistry & Biology,* 15**,** 1198-1206.

BRÖMME, D. 2000. Papain-like Cysteine Proteases. *Current Protocols in Protein Science,* 21**,** 21.2.1-21.2.14.

BROT, N. & WEISSBACH, H. 2000. Peptide Methionine Sulfoxide Reductase: Biochemistry And Physiological Role. *Peptide Science,* 55**,** 288-296.

BROWN, N. L., STOYANOV, J. V., KIDD, S. P. & HOBMAN, J. L. 2003. The MerR Family Of Transcriptional Regulators. *FEMS Microbiology Reviews,* 27**,** 145-163.

BURKHART, B. J., HUDSON, G. A., DUNBAR, K. L. & MITCHELL, D. A. 2015. A Prevalent Peptide-Binding Domain Guides Ribosomal Natural Product Biosynthesis. *Nature Chemical Biology,* 11**,** 564-570.

CALDERON, M. & BAUMANN, W. J. 1970. Gel Permeation Chromatography Of Neutral Hydroxy Lipids On Sephadex LH-20. *J Lipid Res,* 11**,** 167-9.

CALTRIDER, P. G. 1967. Pyocyanine. *In:* GOTTLIEB, D. & SHAW, P. D. (eds.) *Antibiotics: Volume I Mechanism of Action.* Berlin, Heidelberg: Springer Berlin Heidelberg.

CARVALHO, C., SANTOS, X. R., CARDOSO, S., CORREIA, S., OLIVEIRA, J. P., SANTOS, S. M. & MOREIRA, I. P. 2009. Doxorubicin: The Good, The Bad And The Ugly Effect. *Current Medicinal Chemistry,* 16**,** 3267-3285.

CHAIN, A. D. G. E. 1942. Proactinomycin: A “Bacteriostatic” Produced By A Species Of *Proactinomyces*. *The British Journal of Experiental Pathology,* 23**,** 123-127.

CHAIT, B. T. 2006. Mass Spectrometry: Bottom-Up or Top-Down? *Science,* 314**,** 65-66.

CHATTERJEE, S., CHATTERJEE, S., LAD, S. J., PHANSALKAR, M. S., RUPP, R. H., GANGULI, B. N., FEHLHABER, H. W. & KOGLER, H. 1992. Mersacidin, A New Antibiotic From *Bacillus* Fermentation, Isolation, Purification And Chemical Characterization. *J Antibiot (Tokyo),* 45**,** 832-8.

CHEMNETBASE. 2022. *Dictionary Of Natural Product* [Online]. Available: https://dnp.chemnetbase.com/chemical/ChemicalSearch.xhtml?dswid=1492 [Accessed].

CHEN, X., MIETLICKI-BAASE, E. G., BARRETT, T. M., MCGRATH, L. E., KOCH-LASKOWSKI, K., FERRIE, J. J., HAYES, M. R. & PETERSSON, E. J. 2017. Thioamide Substitution Selectively Modulates Proteolysis And Receptor Activity Of Therapeutic Peptide Hormones. *J Am Chem Soc,* 139**,** 16688-16695.

CHEN, Y., DAVIET, L., SCHALK, M., SIEWERS, V. & NIELSEN, J. 2013. Establishing A Platform Cell Factory Through Engineering Of Yeast Acetyl-CoA Metabolism. *Metab Eng,* 15**,** 48-54.

CHOJNACKI, S., COWLEY, A., LEE, J., FOIX, A. & LOPEZ, R. 2017. Programmatic Access To Bioinformatics Tools From EMBL-EBI Update: 2017. *Nucleic acids research,* 45**,** W550-W553.

CLAESEN, J. & BIBB, M. 2010. Genome Mining And Genetic Analysis Of Cypemycin Biosynthesis Reveal An Unusual Class Of Posttranslationally Modified Peptides. *Proc Natl Acad Sci U S A,* 107**,** 16297-302.

COBB, R. E., WANG, Y. & ZHAO, H. 2015. High-Efficiency Multiplex Genome Editing of *Streptomyces* Species Using an Engineered CRISPR/Cas System. *ACS Synthetic Biology,* 4**,** 723-728.

COGAN, D. P., BHUSHAN, A., REYES, R., ZHU, L., PIEL, J. & NAIR, S. K. 2022. Structure And Mechanism For Iterative Amide N-Methylation In The Biosynthesis Of Channel-Forming Peptide Cytotoxins. *Proceedings of the National Academy of Sciences,* 119**,** e2116578119.

COLEY, W. B. 1910. The Treatment Of Inoperable Sarcoma By Bacterial Toxins (The Mixed Toxins Of The *Streptococcus eysipelas* And The *Bacillus prodigiosus*). *Proc R Soc Med,* 3**,** 1-48.

COLLIN, F. & MAXWELL, A. 2019. The Microbial Toxin Microcin B17: Prospects For The Development Of New Antibacterial Agents. *Journal of Molecular Biology,* 431**,** 3400-3426.

CORTAZZO, P., CERVEÑANSKY, C., MARíN, M., REISS, C., EHRLICH, R. & DEANA, A. 2002. Silent Mutations Affect *in vivo* Protein Folding In *Escherichia coli*. *Biochemical and Biophysical Research Communications,* 293**,** 537-541.

CRAGG, G. M. & NEWMAN, D. J. 2005. Biodiversity: A Continuing Source Of Novel Drug Leads. *Pure and Applied Chemistry,* 77**,** 7-24.

CRONE, W. J. K., VIOR, N. M., SANTOS-ABERTURAS, J., SCHMITZ, L. G., LEEPER, F. J. & TRUMAN, A. W. 2016. Dissecting Bottromycin Biosynthesis Using Comparative Untargeted Metabolomics. *Angewandte Chemie International Edition,* 55**,** 9639-9643.

DATSENKO, K. A. & WANNER, B. L. 2000. One-Step Inactivation Of Chromosomal Genes In *Escherichia coli* K-12 Using PCR Products. *Proceedings of the National Academy of Sciences,* 97**,** 6640-6645.

DE LEON RODRIGUEZ, L. M., WILLIAMS, E. T. & BRIMBLE, M. A. 2018. Chemical Synthesis Of Bioactive Naturally Derived Cyclic Peptides Containing Ene-Like Rigidifying Motifs. *Chemistry – A European Journal,* 24**,** 17869-17880.

DEATHERAGE, C. L., HADZISELIMOVIC, A. & SANDERS, C. R. 2012. Purification And Characterization Of The Human γ-Secretase Activating Protein. *Biochemistry,* 51**,** 5153-5159.

DEBBAS, M. & WHITE, E. 1993. Wild-Type P53 Mediates Apoptosis By E1A, Which Is Inhibited By E1B. *Genes Dev,* 7**,** 546-54.

DEMAIN, A. L. & FANG, A. 2000. The Natural Functions Of Secondary Metabolites. *In:* FIECHTER, A. (ed.) *History of Modern Biotechnology I.* Berlin, Heidelberg: Springer Berlin Heidelberg.

DUNBAR, K. L., CHEKAN, J. R., COX, C. L., BURKHART, B. J., NAIR, S. K. & MITCHELL, D. A. 2014. Discovery Of A New Atp-Binding Motif Involved In Peptidic Azoline Biosynthesis. *Nature Chemical Biology,* 10**,** 823.

DUNBAR, K. L., MELBY, J. O. & MITCHELL, D. A. 2012. YcaO Domains Use ATP To Activate Amide Backbones During Peptide Cyclodehydrations. *Nat Chem Biol,* 8**,** 569-75.

DUNBAR, K. L., TIETZ, J. I., COX, C. L., BURKHART, B. J. & MITCHELL, D. A. 2015. Identification Of An Auxiliary Leader Peptide-Binding Protein Required For Azoline Formation In Ribosomal Natural Products. *J Am Chem Soc,* 137**,** 7672-7.

DUPREE, E. J., JAYATHIRTHA, M., YORKEY, H., MIHASAN, M., PETRE, B. A. & DARIE, C. C. 2020. A Critical Review Of Bottom-Up Proteomics: The Good, The Bad, And The Future Of This Field. *Proteomes,* 8**,** 14.

DUQUESNE, S., DESTOUMIEUX-GARZÓN, D., PEDUZZI, J. & REBUFFAT, S. 2007. Microcins, Gene-Encoded Antibacterial Peptides From Enterobacteria. *Natural Product Reports,* 24**,** 708-734.

EMMERICH, R. & LÖW, O. 1899. Bakteriolytische Enzyme Als Ursache Der Erworbenen Immunität Und Die Heilung Von Infectionskrankheiten Durch Dieselben. *Zeitschrift für Hygiene und Infektionskrankheiten,* 31**,** 1-65.

ERB, M. & KLIEBENSTEIN, D. J. 2020. Plant Secondary Metabolites As Defenses, Regulators, And Primary Metabolites: The Blurred Functional Trichotomy. *Plant Physiol,* 184**,** 39-52.

EVANS, B. J., KING, A. T., KATSIFIS, A., MATESIC, L. & JAMIE, J. F. 2020. Methods To Enhance The Metabolic Stability Of Peptide-Based Pet Radiopharmaceuticals. *Molecules,* 25**,** 2314.

EYLES, T. H., VIOR, N. M., LACRET, R. & TRUMAN, A. W. 2021. Understanding Thioamitide Biosynthesis Using Pathway Engineering And Untargeted Metabolomics. *Chemical Science,* 12**,** 7138-7150.

EYLES, T. H., VIOR, N. M. & TRUMAN, A. W. 2018. Rapid And Robust Yeast-Mediated Pathway Refactoring Generates Multiple New Bottromycin-Related Metabolites. *ACS Synthetic Biology,* 7**,** 1211-1218.

EZRATY, B., AUSSEL, L. & BARRAS, F. 2005. Methionine Sulfoxide Reductases In Prokaryotes. *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics,* 1703**,** 221-229.

FELNAGLE, E. A., JACKSON, E. E., CHAN, Y. A., PODEVELS, A. M., BERTI, A. D., MCMAHON, M. D. & THOMAS, M. G. 2008. Nonribosomal Peptide Synthetases Involved In The

Production Of Medically Relevant Natural Products. *Molecular Pharmaceutics,* 5**,** 191-211.

FERNÁNDEZ-MARTÍNEZ, L. T. & BIBB, M. J. 2014. Use Of The Meganuclease I-SceI Of *Saccharomyces cerevisiae* To Select For Gene Deletions In *Actinomycetes*. *Scientific Reports,* 4**,** 7100.

FERNÁNDEZ-MARTÍNEZ, L. T., GOMEZ-ESCRIBANO, J. P. & BIBB, M. J. 2015. A Rela-Dependent Regulatory Cascade For Auto-Induction Of Microbisporicin Production In *Microbispora corallina*. *Molecular Microbiology,* 97**,** 502-514.

FINKING, R. & MARAHIEL, M. A. 2004. Biosynthesis Of Nonribosomal Peptides. *Annual Review of Microbiology,* 58**,** 453-488.

FLEMING, A. 1929. On The Antibacterial Action Of Cultures Of A *Penicillium*, With Special Reference To Their Use In The Isolation Of *B. influenzæ*. *British journal of experimental pathology,* 10**,** 226-236.

FLÜHE, L., KNAPPE, T. A., GATTNER, M. J., SCHÄFER, A., BURGHAUS, O., LINNE, U. & MARAHIEL, M. A. 2012. The Radical SAM Enzyme AlbA Catalyzes Thioether Bond Formation In Subtilosin A. *Nat Chem Biol,* 8**,** 350-7.

FOULSTON, L. C. & BIBB, M. J. 2010. Microbisporicin Gene Cluster Reveals Unusual Features Of Lantibiotic Biosynthesis In *Actinomycetes*. *Proceedings of the National Academy of Sciences,* 107**,** 13461-13466.

FRANZ, L., ADAM, S., SANTOS-ABERTURAS, J., TRUMAN, A. W. & KOEHNKE, J. 2017. Macroamidine Formation In Bottromycins Is Catalyzed By A Divergent YcaO Enzyme. *J Am Chem Soc,* 139**,** 18158-18161.

FRANZ, L., KAZMAIER, U., TRUMAN, A. W. & KOEHNKE, J. 2021. Bottromycins-Biosynthesis, Synthesis And Activity. *Natural Product Reports,* 38**,** 1659-1683.

FRANZ, L. & KOEHNKE, J. 2021. Leader Peptide Exchange To Produce Hybrid, New-To-Nature Ribosomal Natural Products. *Chemical Communications,* 57**,** 6372-6375.

FRATTARUOLO, L., FIORILLO, M., BRINDISI, M., CURCIO, R., DOLCE, V., LACRET, R., TRUMAN, A. W., SOTGIA, F., LISANTI, M. P. & CAPPELLO, A. R. 2019. Thioalbamide, A Thioamidated Peptide From *Amycolatopsis alba*, Affects Tumor Growth And Stemness By Inducing Metabolic Dysfunction And Oxidative Stress. *Cells,* 8**,** 1408.

FRATTARUOLO, L., LACRET, R., CAPPELLO, A. R. & TRUMAN, A. W. 2017. A Genomics-Based Approach Identifies A Thioviridamide-Like Compound With Selective Anticancer Activity. *ACS Chemical Biology,* 12**,** 2815-2822.

FRATTARUOLO, L., MALIVINDI, R., BRINDISI, M., RAGO, V., CURCIO, R., LAURIA, G., FIORILLO, M., DOLCE, V., TRUMAN, A. W. & CAPPELLO, A. R. 2023. Thioalbamide Inhibits FoF1-ATPase In Breast Cancer Cells And Reduces Tumor Proliferation And Invasiveness In Breast Cancer *in vivo* Models. *Molecular Metabolism,* 68**,** 101674.

FUHRMANN, J., CLANCY, K. W. & THOMPSON, P. R. 2015. Chemical Biology of Protein Arginine Modifications in Epigenetic Regulation. *Chemical Reviews,* 115**,** 5413-5461.

FUJIMORI, D. G. 2013. Radical SAM-Mediated Methylation Reactions. *Curr Opin Chem Biol,* 17**,** 597-604.

GAO, S., GE, Y., BENT, A. F., SCHWARZ-LINEK, U. & NAISMITH, J. H. 2018. Oxidation Of The Cyanobactin Precursor Peptide Is Independent Of The Leader Peptide And Operates In A Defined Order. *Biochemistry,* 57**,** 5996-6002.

GAVRIILIDOU, A., KAUTSAR, S. A., ZABURANNYI, N., KRUG, D., MÜLLER, R., MEDEMA, M. H. & ZIEMERT, N. 2021. A Global Survey Of Specialized Metabolic Diversity Encoded in Bacterial Genomes. *bioRxiv***,** 2021.08.11.455920.

GAYNES, R. 2017. The Discovery Of Penicillin—New Insights After More Than 75 Years of Clinical Use. *Emerging Infectious Disease journal,* 23**,** 849.

GE, W., WOLF, A., FENG, T., HO, C.-H., SEKIRNIK, R., ZAYER, A., GRANATINO, N., COCKMAN, M. E., LOENARZ, C., LOIK, N. D., HARDY, A. P., CLARIDGE, T. D. W., HAMED, R. B., CHOWDHURY, R., GONG, L., ROBINSON, C. V., TRUDGIAN, D. C., JIANG, M., MACKEEN, M. M., MCCULLAGH, J. S., GORDIYENKO, Y., THALHAMMER, A., YAMAMOTO, A., YANG, M., LIU-YI, P., ZHANG, Z., SCHMIDT-ZACHMANN, M., KESSLER, B. M., RATCLIFFE, P. J., PRESTON, G. M., COLEMAN, M. L. & SCHOFIELD, C. J. 2012. Oxygenase-Catalyzed Ribosome Hydroxylation Occurs In Prokaryotes And Humans. *Nature Chemical Biology,* 8**,** 960-962.

GEOGHEGAN, K. F., DIXON, H. B., ROSNER, P. J., HOTH, L. R., LANZETTI, A. J., BORZILLERI, K. A., MARR, E. S., PEZZULLO, L. H., MARTIN, L. B., LEMOTTE, P. K., MCCOLL, A. S., KAMATH, A. V. & STROH, J. G. 1999. Spontaneous Alpha-N-6-Phosphogluconoylation Of A "His Tag" In *Escherichia coli*: The Cause Of Extra Mass Of 258 Or 178 Da In Fusion Proteins. *Anal Biochem,* 267**,** 169-84.

GEORGIOU, G. & VALAX, P. 1996. Expression Of Correctly Folded Proteins In *Escherichia coli*. *Curr Opin Biotechnol,* 7**,** 190-7.

GIBSON, D. G., BENDERS, G. A., AXELROD, K. C., ZAVERI, J., ALGIRE, M. A., MOODIE, M., MONTAGUE, M. G., VENTER, J. C., SMITH, H. O. & HUTCHISON, C. A. 2008. One-Step Assembly In Yeast Of 25 Overlapping DNA Fragments To Form A Complete Synthetic Mycoplasma Genitalium Genome. *Proceedings of the National Academy of Sciences,* 105**,** 20404-20409.

GIBSON, D. G., YOUNG, L., CHUANG, R.-Y., VENTER, J. C., HUTCHISON, C. A. & SMITH, H. O. 2009. Enzymatic Assembly Of DNA Molecules Up To Several Hundred Kilobases. *Nature Methods,* 6**,** 343-345.

GOMEZ-ESCRIBANO, J. P. & BIBB, M. J. 2011. Engineering *Streptomyces coelicolor* For Heterologous Expression Of Secondary Metabolite Gene Clusters. *Microbial Biotechnology,* 4**,** 207-215.

GÖTZ, F., PERCONTI, S., POPELLA, P., WERNER, R. & SCHLAG, M. 2014. Epidermin And Gallidermin: Staphylococcal Lantibiotics. *International Journal of Medical Microbiology,* 304**,** 63-71.

GRANATO, E. T., MEILLER-LEGRAND, T. A. & FOSTER, K. R. 2019. The Evolution And Ecology Of Bacterial Warfare. *Curr Biol,* 29**,** R521-r537.

GREGORICH, Z. R. & GE, Y. 2014. Top-Down Proteomics In Health And Disease: Challenges And Opportunities. *PROTEOMICS,* 14**,** 1195-1210.

GREGORY, M. A., KAJA, A. L., KENDREW, S. G., COATES, N. J., WARNECK, T., NUR-E-ALAM, M., LILL, R. E., SHEEHAN, L. S., CHUDLEY, L., MOSS, S. J., SHERIDAN, R. M., QUIMPERE, M., ZHANG, M.-Q., MARTIN, C. J. & WILKINSON, B. 2013. Structure Guided Design Of Improved Anti-Proliferative Rapalogs Through Biosynthetic Medicinal Chemistry. *Chemical Science,* 4**,** 1046-1052.

GUAN, Z., YATES, N. A. & BAKHTIAR, R. 2003. Detection And Characterization Of Methionine Oxidation In Peptides By Collision-Induced Dissociation And Electron Capture Dissociation. *Journal of the American Society for Mass Spectrometry,* 14**,** 605-613.

GUENGERICH, F. P. 2018. Mechanisms Of Cytochrome P450-Catalyzed Oxidations. *ACS Catalysis,* 8**,** 10964-10976.

GUERRERO, F., CIRAGAN, A. & IWAÏ, H. 2015. Tandem SUMO Fusion Vectors For Improving Soluble Protein Expression And Purification. *Protein Expression and Purification,* 116**,** 42-49.

GUO, J., WANG, J., LEE, J. S. & SCHULTZ, P. G. 2008. Site-Specific Incorporation Of Methyl- And Acetyl-Lysine Analogues Into Recombinant Proteins. *Angewandte Chemie International Edition,* 47**,** 6399-6401.

GUST, B., CHALLIS, G. L., FOWLER, K., KIESER, T. & CHATER, K. F. 2003. PCR-Targeted *Streptomyces* Gene Replacement Identifies A Protein Domain Needed For Biosynthesis Of The Sesquiterpene Soil Odor Geosmin. *Proceedings of the National Academy of Sciences,* 100**,** 1541-1546.

HANSON, J. R. 2003. *The Classes Of Natural Product And Their Isolation*, Royal Society of Chemistry.

HAYAKAWA, Y., SASAKI, K., ADACHI, H., FURIHATA, K., NAGAI, K. & SHIN-YA, K. 2006a. Thioviridamide, A Novel Apoptosis Inducer In Transformed Cells From *Streptomyces olivoviridis*. *The Journal Of Antibiotics,* 59**,** 1.

HAYAKAWA, Y., SASAKI, K., NAGAI, K., SHIN-YA, K. & FURIHATA, K. 2006b. Structure Of Thioviridamide, A Novel Apoptosis Inducer From *Streptomyces olivoviridis*. *J Antibiot (Tokyo),* 59**,** 6-10.

HELF, M. J., FREEMAN, M. F. & PIEL, J. 2019. Investigations Into PoyH, A Promiscuous Protease From Polytheonamide Biosynthesis. *Journal of Industrial Microbiology & Biotechnology,* 46**,** 551-563.

HERTWECK, C. 2009. The Biosynthetic Logic Of Polyketide Diversity. *Angew Chem Int Ed Engl,* 48**,** 4688-716.

HIDER, R. C. & KONG, X. 2010. Chemistry And Biology Of Siderophores. *Natural Product Reports,* 27**,** 637-657.

HO, H.-I., FANG, J. R., CHEUNG, J. & WANG, H. H. 2020. Programmable CRISPR-Cas Transcriptional Activation In Bacteria. *Molecular Systems Biology,* 16**,** e9427.

HONG, H.-J., HUTCHINGS, M. I., HILL, L. M. & BUTTNER, M. J. 2005. The Role Of The Novel Fem Protein Vank In Vancomycin Resistance In *Streptomyces coelicolor*. *Journal of Biological Chemistry,* 280**,** 13055-13061.

HORBAL, L., MARQUES, F., NADMID, S., MENDES, M. V. & LUZHETSKYY, A. 2018. Secondary Metabolites Overproduction Through Transcriptional Gene Cluster Refactoring. *Metab Eng,* 49**,** 299-315.

HOUBRAKEN, J., FRISVAD, J. C. & SAMSON, R. A. 2011. Fleming's Penicillin Producing Strain Is Not *Penicillium chrysogenum* but *P. rubens*. *IMA Fungus,* 2**,** 87-95.

HUANG, K.-Y., LEE, T.-Y., KAO, H.-J., MA, C.-T., LEE, C.-C., LIN, T.-H., CHANG, W.-C. & HUANG, H.-D. 2018. DbPTM in 2019: Exploring Disease Association And Cross-Talk Of Post-Translational Modifications. *Nucleic Acids Research,* 47**,** D298-D308.

HUO, L. & VAN DER DONK, W. A. 2016. Discovery And Characterization Of Bicereucin, An Unusual D-Amino Acid-Containing Mixed Two-Component Lantibiotic. *Journal of the American Chemical Society,* 138**,** 5254-5257.

ICHIKAWA, H., BASHIRI, G. & KELLY, W. L. 2018. Biosynthesis Of The Thiopeptins And Identification Of An F420H2-Dependent Dehydropiperidine Reductase. *Journal of the American Chemical Society,* 140**,** 10749-10756.

IZAWA, M., KAWASAKI, T. & HAYAKAWA, Y. 2013. Cloning And Heterologous Expression Of The Thioviridamide Biosynthesis Gene Cluster From *Streptomyces olivoviridis*. *Applied and environmental microbiology,* 79**,** 7110-7113.

IZAWA, M., NAGAMINE, S., AOKI, H. & HAYAKAWA, Y. 2018. Identification Of Essential Biosynthetic Genes And A True Biosynthetic Product For Thioviridamide. *J Gen Appl Microbiol,* 64**,** 50-53.

IZUMIKAWA, M., KOZONE, I., HASHIMOTO, J., KAGAYA, N., TAKAGI, M., KOIWAI, H., KOMATSU, M., FUJIE, M., SATOH, N., IKEDA, H. & SHIN-YA, K. 2015. Novel Thioviridamide Derivative—JBIR-140: Heterologous Expression Of The Gene Cluster For Thioviridamide Biosynthesis. *The Journal of Antibiotics,* 68**,** 533-536.

JIANG, J., GUIZA BELTRAN, D., SCHACHT, A., WRIGHT, S., ZHANG, L. & DU, L. 2018. Functional and Structural Analysis of Phenazine O-Methyltransferase LaPhzM from *Lysobacter antibioticus* OH13 and One-Pot Enzymatic Synthesis of the Antibiotic Myxin. *ACS Chemical Biology,* 13**,** 1003-1012.

JOHNSTON, C. W., SKINNIDER, M. A., WYATT, M. A., LI, X., RANIERI, M. R. M., YANG, L., ZECHEL, D. L., MA, B. & MAGARVEY, N. A. 2015. An Automated Genomes-To-Natural Products Platform (GNP) For The Discovery Of Modular Natural Products. *Nature Communications,* 6**,** 8421.

JONES, A. C., GUST, B., KULIK, A., HEIDE, L., BUTTNER, M. J. & BIBB, M. J. 2013. Phage P1-Derived Artificial Chromosomes Facilitate Heterologous Expression of the FK506 Gene Cluster. *PLOS ONE,* 8**,** e69319.

KALVA, S., BOEKE, J. D. & MITA, P. 2018. Gibson Deletion: A Novel Application Of Isothermal *in vitro* Recombination. *Biol Proced Online,* 20**,** 2.

KAMATOU, G. P. P., VERMAAK, I., VILJOEN, A. M. & LAWRENCE, B. M. 2013. Menthol: A Simple Monoterpene With Remarkable Biological Properties. *Phytochemistry,* 96**,** 15-25.

KANE, J. F. 1995. Effects Of Rare Codon Clusters On High-Level Expression Of Heterologous Proteins In *Escherichia coli*. *Current Opinion in Biotechnology,* 6**,** 494-500.

KAPER, J. B. & SPERANDIO, V. 2005. Bacterial Cell-To-Cell Signaling In The Gastrointestinal Tract. *Infection and Immunity,* 73**,** 3197-3209.

KARZAI, A. W., ROCHE, E. D. & SAUER, R. T. 2000. The SsrA–SmpB System For Protein Tagging, Directed Degradation And Ribosome Rescue. *Nature Structural Biology,* 7**,** 449-455.

KAWAHARA, T., IZUMIKAWA, M., KOZONE, I., HASHIMOTO, J., KAGAYA, N., KOIWAI, H., KOMATSU, M., FUJIE, M., SATO, N., IKEDA, H. & SHIN-YA, K. 2018. Neothioviridamide, a Polythioamide Compound Produced by Heterologous Expression of a *Streptomyces* sp. Cryptic RiPP Biosynthetic Gene Cluster. *Journal of Natural Products,* 81**,** 264-269.

KE, Z., SMITH, G. K., ZHANG, Y. & GUO, H. 2011. Molecular Mechanism For Eliminylation, A Newly Discovered Post-Translational Modification. *Journal of the American Chemical Society,* 133**,** 11103-11105.

KEILER, K. C. & RAMADOSS, N. S. 2011. Bifunctional Transfer-Messenger RNA. *Biochimie,* 93**,** 1993-7.

KELLEY, L. A., MEZULIS, S., YATES, C. M., WASS, M. N. & STERNBERG, M. J. E. 2015. The Phyre2 Web Portal For Protein Modeling, Prediction And Analysis. *Nature Protocols,* 10**,** 845.

KELLY, W. L., PAN, L. & LI, C. 2009. Thiostrepton Biosynthesis: Prototype For A New Family Of Bacteriocins. *Journal of the American Chemical Society,* 131**,** 4327-4334.

KENNEY, G. E. & ROSENZWEIG, A. C. 2012. Chemistry And Biology Of The Copper Chelator Methanobactin. *ACS Chemical Biology,* 7**,** 260-268.

KERSTEN, R. D. & WENG, J.-K. 2018. Gene-Guided Discovery And Engineering Of Branched Cyclic Peptides In Plants. *Proceedings of the National Academy of Sciences,* 115**,** E10961-E10969.

KIESER, T., BIBB, M. J., BUTTNER, M. J., CHATER, K. F. & DAVID A, H. 2000. *Practical Streptomyces Genetics,* Norwich, England, The John Innes Foundation.

KJAERULFF, L., SIKANDAR, A., ZABURANNYI, N., ADAM, S., HERRMANN, J., KOEHNKE, J. & MÜLLER, R. 2017. Thioholgamides: Thioamide-Containing Cytotoxic RiPP Natural Products. *ACS Chemical Biology,* 12**,** 2837-2841.

KLAENHAMMER, T. R. 1993. Genetics Of Bacteriocins Produced By Lactic Acid Bacteria. *FEMS Microbiol Rev,* 12**,** 39-85.

KLOOSTERMAN, A. M., MEDEMA, M. H. & VAN WEZEL, G. P. 2021. Omics-Based Strategies To Discover Novel Classes Of RiPP Natural Products. *Current Opinion in Biotechnology,* 69**,** 60-67.

KOEHNKE, J., MANN, G., BENT, A. F., LUDEWIG, H., SHIRRAN, S., BOTTING, C., LEBL, T., HOUSSEN, W., JASPARS, M. & NAISMITH, J. H. 2015. Structural Analysis Of Leader Peptide Binding Enables Leader-Free Cyanobactin Processing. *Nat Chem Biol,* 11**,** 558-563.

KOU, Q., WU, S. & LIU, X. 2014. A New Scoring Function For Top-Down Spectral Deconvolution. *BMC Genomics,* 15**,** 1140.

KRENTZ, B. D., MULHERON, H. J., SEMRAU, J. D., DISPIRITO, A. A., BANDOW, N. L., HAFT, D. H., VUILLEUMIER, S., MURRELL, J. C., MCELLISTREM, M. T., HARTSEL, S. C. & GALLAGHER, W. H. 2010. A Comparison of Methanobactins from *Methylosinus trichosporium* OB3b and *Methylocystis* Strain SB2 Predicts Methanobactins Are Synthesized from Diverse Peptide Precursors Modified To Create a Common Core for Binding and Reducing Copper Ions. *Biochemistry,* 49**,** 10117-10130.

KULCZYK, A. W. & LEUSTEK, T. 2022. You Cannot Oxidize What You Cannot Reach: Oxidative Susceptibility Of Buried Methionine Residues. *Journal of Biological Chemistry,* 298**,** 101973.

KUROCHKINA, L. P. & MESYANZHINOV, V. V. 1999. Co-Expression Of Gene 31 And 23 Products Of Bacteriophage T4. *Biochemistry (Mosc),* 64**,** 379-83.

LARIONOV, V., KOUPRINA, N., SOLOMON, G., BARRETT, J. C. & RESNICK, M. A. 1997. Direct Isolation Of Human *BRCA2* Gene By Transformation-Associated Recombination In Yeast. *Proceedings of the National Academy of Sciences,* 94**,** 7384-7387.

LARSSON, F. C. V. L., S.-O.; MØLLER, JØRGEN; SCHROLL, GUSTAV 1973. Mass Spectra Of Thioamide. *Acta Chemica Scandinavica,* 27**,** 747-755.

LEE, H.-J., CHOI, Y.-S., LEE, K.-B., PARK, J. & YOON, C.-J. 2002. Hydrogen Bonding Abilities of Thioamide. *The Journal of Physical Chemistry A,* 106**,** 7010-7017.

LEE, N. C. O., LARIONOV, V. & KOUPRINA, N. 2015. Highly Efficient CRISPR/Cas9-Mediated Tar Cloning Of Genes And Chromosomal Loci From Complex Genomes In Yeast. *Nucleic Acids Research,* 43**,** e55-e55.

LETUNIC, I. & BORK, P. 2016. Interactive Tree Of Life (iTOL) V3: An Online Tool For The Display And Annotation Of Phylogenetic And Other Trees. *Nucleic Acids Research,* 44**,** W242-W245.

LI, B., SHER, D., KELLY, L., SHI, Y., HUANG, K., KNERR, P. J., JOEWONO, I., RUSCH, D., CHISHOLM, S. W. & VAN DER DONK, W. A. 2010. Catalytic Promiscuity In The Biosynthesis Of Cyclic Peptide Secondary Metabolites In Planktonic Marine Cyanobacteria. *Proceedings of the National Academy of Sciences,* 107**,** 10430-10435.

LI, C., ZHANG, F. & KELLY, W. L. 2012. Mutagenesis Of The Thiostrepton Precursor Peptide At Thr7 Impacts Both Biosynthesis And Function. *Chemical Communications,* 48**,** 558-560.

LI, Q., MONTALBAN-LOPEZ, M. & KUIPERS, O. P. 2018. Increasing The Antimicrobial Activity Of Nisin-Based Lantibiotics Against Gram-Negative Pathogens. *Applied and Environmental Microbiology,* 84**,** e00052-18.

LI, S. & LUO, X. 2003. *Compendium Of Materia Medica : Bencao Gangmu / Compiled By Li Shizhen ; translated and annotated by Luo Xiwen,* Beijing, China, Foreign Languages Press.

LI, Y.-M., MILNE, J. C., MADISON, L. L., KOLTER, R. & WALSH, C. T. 1996. From Peptide Precursors to Oxazole and Thiazole-Containing Peptide Antibiotics: Microcin B17 Synthase. *Science,* 274**,** 1188-1193.

LIANG, H., LOPEZ, I. J., SÁNCHEZ-HIDALGO, M., GENILLOUD, O. & VAN DER DONK, W. A. 2022. Mechanistic Studies On Dehydration In Class V Lanthipeptides. *ACS Chemical Biology,* 17**,** 2519-2527.

LIANG, X., KAYA, A., ZHANG, Y., LE, D. T., HUA, D. & GLADYSHEV, V. N. 2012. Characterization Of Methionine Oxidation And Methionine Sulfoxide Reduction Using Methionine-Rich Cysteine-Free Proteins. *BMC Biochemistry,* 13**,** 21.

LINCKE, T., BEHNKEN, S., ISHIDA, K., ROTH, M. & HERTWECK, C. 2010. Closthioamide: An Unprecedented Polythioamide Antibiotic From The Strictly Anaerobic Bacterium *Clostridium cellulolyticum*. *Angewandte Chemie International Edition,* 49**,** 2011-2013.

LIU, A., SI, Y., DONG, S.-H., MAHANTA, N., PENKALA, H. N., NAIR, S. K. & MITCHELL, D. A. 2021a. Functional Elucidation Of TfuA In Peptide Backbone Thioamidation. *Nature Chemical Biology,* 17**,** 585-592.

LIU, J., LIN, Z., LI, Y., ZHENG, Q., CHEN, D. & LIU, W. 2019a. Insights Into The Thioamidation Of Thiopeptins To Enhance The Understanding Of The Biosynthetic Logic Of Thioamide-Containing Thiopeptides. *Organic & Biomolecular Chemistry,* 17**,** 3727-3731.

LIU, Y., YANG, F., ZOU, S. & QU, L. 2019b. Rapamycin: A Bacteria-Derived Immunosuppressant That Has Anti-atherosclerotic Effects and Its Clinical Application. *Frontiers in Pharmacology,* 9.

LIU, Z.-H., HUANG, D., FU, X.-J., CHENG, P. & DU, E.-Q. 2018. Comparison Of Three Commonly Used Fusion Tags For The Expression Of Nanobodies In The Cytoplasm Of *Escherichia coli*. *Biotechnology & Biotechnological Equipment,* 32**,** 462-469.

LIU, Z., ZHAO, Y., HUANG, C. & LUO, Y. 2021b. Recent Advances in Silent Gene Cluster Activation in *Streptomyces*. *Frontiers in Bioengineering and Biotechnology,* 9.

LOHANS, C. T., LI, J. L. & VEDERAS, J. C. 2014. Structure And Biosynthesis Of Carnolysin, A Homologue Of Enterococcal Cytolysin With D-Amino Acids. *Journal of the American Chemical Society,* 136**,** 13150-13153.

LOMBÓ, F., BRAÑA, A. F., MÉNDEZ, C. & SALAS, J. A. 1999. The Mithramycin Gene Cluster Of *Streptomyces argillaceus* Contains A Positive Regulatory Gene And Two Repeated DNA Sequences That Are Located At Both Ends Of The Cluster. *J Bacteriol,* 181**,** 642-7.

LOMOVSKAYA, N., OTTEN, S. L., DOI-KATAYAMA, Y., FONSTEIN, L., LIU, X.-C., TAKATSU, T., INVENTI-SOLARI, A., FILIPPINI, S., TORTI, F., COLOMBO, A. L. & HUTCHINSON, C. R. 1999. Doxorubicin Overproduction In *Streptomyces peucetius*: Cloning And

Characterization Of The *dnrU* Ketoreductase Anddnrv Genes And The *doxA* Cytochrome P-450 Hydroxylase Gene. *Journal of Bacteriology,* 181**,** 305-318.

LORENZ, R., BERNHART, S. H., HÖNER ZU SIEDERDISSEN, C., TAFER, H., FLAMM, C., STADLER, P. F. & HOFACKER, I. L. 2011. ViennaRNA Package 2.0. *Algorithms Mol Biol,* 6**,** 26.

LOUIE, K. B., KOSINA, S. M., HU, Y., OTANI, H., DE RAAD, M., KUFTIN, A. N., MOUNCEY, N. J., BOWEN, B. P. & NORTHEN, T. R. 2020. Mass Spectrometry For Natural Product Discovery. *In:* LIU, H.-W. & BEGLEY, T. P. (eds.) *Comprehensive Natural Products III.* Oxford: Elsevier.

LU, J., WU, Y., LI, Y. & WANG, H. 2021. The Utilization Of Lanthipeptide Synthetases Is A General Strategy For The Biosynthesis Of 2-Aminovinyl-Cysteine Motifs In Thioamitides. *Angew Chem Int Ed Engl,* 60**,** 1951-1958.

LU, S., WANG, J., CHITSAZ, F., DERBYSHIRE, M. K., GEER, R. C., GONZALES, N. R., GWADZ, M., HURWITZ, D. I., MARCHLER, G. H., SONG, J. S., THANKI, N., YAMASHITA, R. A., YANG, M., ZHANG, D., ZHENG, C., LANCZYCKI, C. J. & MARCHLER-BAUER, A. 2020. CDD/SPARCLE: The Conserved Domain Database In 2020. *Nucleic Acids Res,* 48**,** D265-d268.

LU, Z., XIE, P. & QIN, Z. 2010. Promotion Of Markerless Deletion Of The Actinorhodin Biosynthetic Gene Cluster In *Streptomyces coelicolor*. *Acta Biochimica et Biophysica Sinica,* 42**,** 717-721.

LUO, S. & DONG, S.-H. 2019. Recent Advances In The Discovery And Biosynthetic Study Of Eukaryotic RiPP Natural Products. *Molecules,* 24**,** 1541.

LUO, Y., HUANG, H., LIANG, J., WANG, M., LU, L., SHAO, Z., COBB, R. E. & ZHAO, H. 2013. Activation And Characterization Of A Cryptic Polycyclic Tetramate Macrolactam Biosynthetic Gene Cluster. *Nature Communications,* 4**,** 2894.

MA, S. & ZHANG, Q. 2020. Linaridin Natural Products. *Natural Product Reports,* 37**,** 1152-1163.

MACNEIL, D. J. 1988. Characterization Of A Unique Methyl-Specific Restriction System In *Streptomyces Avermitilis*. *Journal of Bacteriology,* 170**,** 5607-5612.

MACNEIL, D. J., GEWAIN, K. M., RUBY, C. L., DEZENY, G., GIBBONS, P. H. & MACNEIL, T. 1992. Analysis Of *Streptomyces Avermitilis* Genes Required For Avermectin Biosynthesis Utilizing A Novel Integration Vector. *Gene,* 111**,** 61-8.

MADEIRA, F., PEARCE, M., TIVEY, A. R. N., BASUTKAR, P., LEE, J., EDBALI, O., MADHUSOODANAN, N., KOLESNIKOV, A. & LOPEZ, R. 2022. Search And Sequence Analysis Tools Services From EMBL-EBI In 2022. *Nucleic acids research***,** gkac240.

MAHANTA, N., LIU, A., DONG, S., NAIR, S. K. & MITCHELL, D. A. 2018. Enzymatic Reconstitution Of Ribosomal Peptide Backbone Thioamidation. *Proc Natl Acad Sci U S A,* 115**,** 3030-3035.

MAJER, F., SCHMID, D. G., ALTENA, K., BIERBAUM, G. & KUPKE, T. 2002. The Flavoprotein MrsD Catalyzes The Oxidative Decarboxylation Reaction Involved In Formation Of The Peptidoglycan Biosynthesis Inhibitor Mersacidin. *Journal of Bacteriology,* 184**,** 1234-1243.

MANN, M. & JENSEN, O. N. 2003. Proteomic Analysis Of Post-Translational Modifications. *Nature Biotechnology,* 21**,** 255-261.

MAO, D., OKADA, B. K., WU, Y., XU, F. & SEYEDSAYAMDOST, M. R. 2018. Recent Advances In Activating Silent Biosynthetic Gene Clusters In Bacteria. *Current Opinion in Microbiology,* 45**,** 156-163.

MARBLESTONE, J. G., EDAVETTAL, S. C., LIM, Y., LIM, P., ZUO, X. & BUTT, T. R. 2006. Comparison Of SUMO Fusion Technology With Traditional Gene Fusion Systems: Enhanced Expression And Solubility With SUMO. *Protein Science,* 15**,** 182-189.

MARON, M. I., CASILL, A. D., GUPTA, V., ROTH, J. S., SIDOLI, S., QUERY, C. C., GAMBLE, M. J. & SHECHTER, D. 2022. Type I And II PRMTs Inversely Regulate Post-Transcriptional Intron Detention Through Sm And CHTOP Methylation. *eLife,* 11**,** e72867.

MARTIN, N. I., SPRULES, T., CARPENTER, M. R., COTTER, P. D., HILL, C., ROSS, R. P. & VEDERAS, J. C. 2004. Structural Characterization Of Lacticin 3147, A Two-Peptide Lantibiotic With Synergistic Activity. *Biochemistry,* 43**,** 3049-3056.

MCDANIEL, R., THAMCHAIPENET, A., GUSTAFSSON, C., FU, H., BETLACH, M., BETLACH, M. & ASHLEY, G. 1999. Multiple Genetic Modifications Of The Erythromycin Polyketide Synthase To Produce A Library Of Novel "Unnatural" Natural Products. *Proceedings of the National Academy of Sciences,* 96**,** 1846-1851.

MCINTOSH, J. A., DONIA, M. S. & SCHMIDT, E. W. 2009. Ribosomal Peptide Natural Products: Bridging The Ribosomal And Nonribosomal Worlds. *Nat Prod Rep,* 26**,** 537-59.

MEDEMA, M. H. & FISCHBACH, M. A. 2015. Computational Approaches To Natural Product Discovery. *Nature Chemical Biology,* 11**,** 639-648.

MEIER, S., JENSEN, P. R. & DUUS, J. Ø. 2012. Direct Observation Of Metabolic Differences In Living *Escherichia coli* Strains K-12 And Bl21. *ChemBioChem,* 13**,** 308-310.

METELEV, M., OSTERMAN, I. A., GHILAROV, D., KHABIBULLINA, N. F., YAKIMOV, A., SHABALIN, K., UTKINA, I., TRAVIN, D. Y., KOMAROVA, E. S., SEREBRYAKOVA, M., ARTAMONOVA, T., KHODORKOVSKII, M., KONEVEGA, A. L., SERGIEV, P. V., SEVERINOV, K. & POLIKANOV, Y. S. 2017. Klebsazolicin Inhibits 70s Ribosome By Obstructing The Peptide Exit Tunnel. *Nature Chemical Biology,* 13**,** 1129-1136.

MILES, C. O., MELANSON, J. E. & BALLOT, A. 2014. Sulfide Oxidations For LC-MS Analysis Of Methionine-Containing Microcystins In *Dolichospermum flos-aquae* NIVA-CYA 656. *Environmental Science & Technology,* 48**,** 13307-13315.

MILLER, M. A., PFEIFFER, W. & SCHWARTZ, T. 2011. The CIPRES Science Gateway: A Community Resource For Phylogenetic Analyses. *Proceedings of the 2011 TeraGrid Conference: Extreme Digital Discovery.* Salt Lake City, Utah: Association for Computing Machinery.

MINGYAR, E., MÜHLING, L., KULIK, A., WINKLER, A., WIBBERG, D., KALINOWSKI, J., BLIN, K., WEBER, T., WOHLLEBEN, W. & STEGMANN, E. 2021. A Regulator Based "Semi-Targeted" Approach To Activate Silent Biosynthetic Gene Clusters. *International Journal of Molecular Sciences,* 22**,** 7567.

MITOUSIS, L., THOMA, Y. & MUSIOL-KROLL, E. M. 2020. An Update On Molecular Tools For Genetic Engineering Of *Actinomycetes*—The Source Of Important Antibiotics And Other Valuable Compounds. *Antibiotics,* 9**,** 494.

MONTALBÁN-LÓPEZ, M., SCOTT, T. A., RAMESH, S., RAHMAN, I. R., VAN HEEL, A. J., VIEL, J. H., BANDARIAN, V., DITTMANN, E., GENILLOUD, O., GOTO, Y., GRANDE BURGOS, M. J., HILL, C., KIM, S., KOEHNKE, J., LATHAM, J. A., LINK, A. J., MARTÍNEZ, B., NAIR, S. K., NICOLET, Y., REBUFFAT, S., SAHL, H.-G., SAREEN, D., SCHMIDT, E. W., SCHMITT, L., SEVERINOV, K., SÜSSMUTH, R. D., TRUMAN, A. W., WANG, H., WENG, J.-K., VAN WEZEL, G. P., ZHANG, Q., ZHONG, J., PIEL, J., MITCHELL, D. A., KUIPERS, O. P. & VAN DER DONK, W. A. 2021. New Developments In RiPP Discovery, Enzymology And Engineering. *Natural Product Reports,* 38**,** 130-239.

MORINAKA, B. I., LAKIS, E., VEREST, M., HELF, M. J., SCALVENZI, T., VAGSTAD, A. L., SIMS, J., SUNAGAWA, S., GUGGER, M. & PIEL, J. 2018. Natural Noncanonical Protein Splicing Yields Products With Diverse β-Amino Acid Residues. *Science,* 359**,** 779-782.

MOSS, N. A., SEILER, G., LEÃO, T. F., CASTRO-FALCÓN, G., GERWICK, L., HUGHES, C. C. & GERWICK, W. H. 2019. Nature's Combinatorial Biosynthesis Produces Vatiamides A–F. *Angewandte Chemie International Edition,* 58**,** 9027-9031.

MURPHY, B. T., NARENDER, T., KAUFFMAN, C. A., WOOLERY, M., JENSEN, P. R. & FENICAL, W. 2010. Saliniquinones A-F, New Members Of The Highly Cytotoxic Anthraquinone-γ-Pyrones From The Marine Actinomycete *Salinispora arenicola*. *Australian Journal of Chemistry,* 63**,** 929-934.

MUSIOL-KROLL, E. M., TOCCHETTI, A., SOSIO, M. & STEGMANN, E. 2019. Challenges And Advances In Genetic Manipulation Of Filamentous *Actinomycetes* – The Remarkable Producers Of Specialized Metabolites. *Natural Product Reports,* 36**,** 1351-1369.

NAH, H.-J., PYEON, H.-R., KANG, S.-H., CHOI, S.-S. & KIM, E.-S. 2017. Cloning And Heterologous Expression Of A Large-Sized Natural Product Biosynthetic Gene Cluster In *Streptomyces* Species. *Frontiers in Microbiology,* 8.

NAYAK, D. D., MAHANTA, N., MITCHELL, D. A. & METCALF, W. W. 2017. Post-Translational Thioamidation Of Methyl-Coenzyme M Reductase, A Key Enzyme In Methanogenic And Methanotrophic Archaea. *eLife,* 6**,** e29218.

NOSKOV, V. N., KOUPRINA, N., LEEM, S. H., OUSPENSKI, I., BARRETT, J. C. & LARIONOV, V. 2003. A General Cloning System To Selectively Isolate Any Eukaryotic Or Prokaryotic Genomic Region In Yeast. *BMC Genomics,* 4**,** 16.

O'ROURKE, S., WIDDICK, D. & BIBB, M. 2017. A Novel Mechanism Of Immunity Controls The Onset Of Cinnamycin Biosynthesis In *Streptomyces cinnamoneus* DSM 40646. *Journal of Industrial Microbiology and Biotechnology,* 44**,** 563-572.

OH, D.-C., POULSEN, M., CURRIE, C. R. & CLARDY, J. 2009. Dentigerumycin: A Bacterial Mediator Of An Anti-Fungus Symbiosis. *Nature Chemical Biology,* 5**,** 391-393.

OMAN, T. J. & VAN DER DONK, W. A. 2010. Follow The Leader: The Use Of Leader Peptides To Guide Natural Product Biosynthesis. *Nat Chem Biol,* 6**,** 9-18.

ONGEY, E. L., YASSI, H., PFLUGMACHER, S. & NEUBAUER, P. 2017. Pharmacological And Pharmacokinetic Properties Of Lanthipeptides Undergoing Clinical Studies. *Biotechnology Letters,* 39**,** 473-482.

ORR-WEAVER, T. L., SZOSTAK, J. W. & ROTHSTEIN, R. J. 1981. Yeast Transformation: A Model System For The Study Of Recombination. *Proceedings of the National Academy of Sciences,* 78**,** 6354-6358.

ORTEGA, M. A., VELÁSQUEZ, J. E., GARG, N., ZHANG, Q., JOYCE, R. E., NAIR, S. K. & VAN DER DONK, W. A. 2014. Substrate Specificity Of The Lanthipeptide Peptidase ElxP And The Oxidoreductase ElxO. *ACS Chemical Biology,* 9**,** 1718-1725.

ORTEGÓN SALAS, C., SCHNEIDER, K., LILLIG, C. H. & GELLERT, M. 2020. Signal-Regulated Oxidation Of Proteins Via MICAL. *Biochemical Society Transactions,* 48**,** 613-620.

PARENT, A., BENJDIA, A., GUILLOT, A., KUBIAK, X., BALTY, C., LEFRANC, B., LEPRINCE, J. & BERTEAU, O. 2018. Mechanistic Investigations Of PoyD, A Radical S-Adenosyl-L-Methionine Enzyme Catalyzing Iterative And Directional Epimerizations In Polytheonamide A Biosynthesis. *Journal of the American Chemical Society,* 140**,** 2469-2477.

PATANI, G. A. & LAVOIE, E. J. 1996. Bioisosterism: A Rational Approach in Drug Design. *Chem Rev,* 96**,** 3147-3176.

PINEDO-RIVILLA, C., ALEU, J. & DURÁN-PATRÓN, R. 2022. Cryptic Metabolites From Marine-Derived Microorganisms Using OSMAC And Epigenetic Approaches. *Marine Drugs,* 20**,** 84.

POTAPOV, V., ONG, J. L., KUCERA, R. B., LANGHORST, B. W., BILOTTI, K., PRYOR, J. M., CANTOR, E. J., CANTON, B., KNIGHT, T. F., EVANS, T. C., JR. & LOHMAN, G. J. S. 2018. Comprehensive Profiling Of Four Base Overhang Ligation Fidelity By T4 DNA Ligase And Application To DNA Assembly. *ACS Synthetic Biology,* 7**,** 2665-2674.

QIU, Y., LIU, J., LI, Y., XUE, Y. & LIU, W. 2021. Formation Of An Aminovinyl-Cysteine Residue In Thioviridamides Occurs Through A Path Independent Of Known Lanthionine Synthetase Activity. *Cell Chemical Biology,* 28**,** 675-685.e5.

RAVEH, A., DELEKTA, P. C., DOBRY, C. J., PENG, W., SCHULTZ, P. J., BLAKELY, P. K., TAI, A. W., MATAINAHO, T., IRANI, D. N., SHERMAN, D. H. & MILLER, D. J. 2013. Discovery Of Potent Broad Spectrum Antivirals Derived From Marine Actinobacteria. *PLoS One,* 8**,** e82318.

RAWLINGS, N. D. & BARRETT, A. J. 1994. Families Of Cysteine Peptidases. *Methods Enzymol,* 244**,** 461-86.

RAWLINGS, N. D., WALLER, M., BARRETT, A. J. & BATEMAN, A. 2013. MEROPS: The Database Of Proteolytic Enzymes, Their Substrates And Inhibitors. *Nucleic Acids Research,* 42**,** D503-D509.

REDENBACH, M., KIESER, H. M., DENAPAITE, D., EICHNER, A., CULLUM, J., KINASHI, H. & HOPWOOD, D. A. 1996. A Set Of Ordered Cosmids And A Detailed Genetic And Physical Map For The 8 *Mb Streptomyces coelicolor* A3(2) Chromosome. *Molecular Microbiology,* 21**,** 77-96.

REINER, A., WILDEMANN, D., FISCHER, G. & KIEFHABER, T. 2008. Effect Of Thioxopeptide Bonds On α-Helix Structure And Stability. *Journal of the American Chemical Society,* 130**,** 8079-8084.

REINHARD, J., SRINIVASAN, M. V. & ZHANG, S. 2004. Scent-triggered navigation in honeybees. *Nature,* 427**,** 411.

RENGARAJU, S. N., S; GANJU, P L; AMIN, M A; IYENGAR, M R S; SHOMURA, T; SEZAKI, M 1985. A New Macrolide Antibiotic Kayamycin 10,11-Dihydro-5-O-Mycaminosyl Narbonolide Produced By *Nocardiopsis*. *J-GLOBAL***,** 52-54.

RINCE, A., DUFOUR, A., POGAM, S. L., THUAULT, D., BOURGEOIS, C. M. & PENNEC, J. P. L. 1994. Cloning, Expression, And Nucleotide Sequence Of Genes Involved In Production Of Lactococcin DR, A Bacteriocin From *Lactococcus lactis* Subsp. *lactis*. *Applied and Environmental Microbiology,* 60**,** 1652-1657.

ROBERT, X. & GOUET, P. 2014. Deciphering Key Features In Protein Structures With The New Endscript Server. *Nucleic Acids Research,* 42**,** W320-W324.

ROSA, E. A. S., BENNETT, R. N. & AIRES, A. 2007. 15 Levels And Potential Health Impacts Of Nutritionally Relevant Phytochemicals In Organic And Conventional Food Production Systems. *In:* COOPER, J., NIGGLI, U. & LEIFERT, C. (eds.) *Handbook of Organic Food Safety and Quality.* Woodhead Publishing.

ROSANO, G. L. & CECCARELLI, E. A. 2009. Rare Codon Content Affects The Solubility Of Recombinant Proteins In A Codon Bias-Adjusted *Escherichia coli* Strain. *Microbial Cell Factories,* 8**,** 41.

ROTH, T. L., MILENKOVIC, L. & SCOTT, M. P. 2014. A Rapid And Simple Method For DNA Engineering Using Cycled Ligation Assembly. *PLOS ONE,* 9**,** e107329.

RUED, B. E., COVINGTON, B. C., BUSHIN, L. B., SZEWCZYK, G., LACZKOVICH, I., SEYEDSAYAMDOST, M. R. & FEDERLE, M. J. 2021. Quorum Sensing In *Streptococcus* Mutans Regulates Production Of Tryglysin, A Novel RaS-RiPP Antimicrobial Compound. *mBio,* 12.

RUSSELL, A. H. & TRUMAN, A. W. 2020. Genome Mining Strategies For Ribosomally Synthesised And Post-Translationally Modified Peptides. *Computational and Structural Biotechnology Journal,* 18**,** 1838-1851.

RUSSELL, A. H., VIOR, N. M., HEMS, E. S., LACRET, R. & TRUMAN, A. W. 2021. Discovery And Characterisation Of An Amidine-Containing Ribosomally-Synthesised Peptide That Is Widely Distributed In Nature. *Chemical Science,* 12**,** 11769-11778.

RUTLEDGE, P. J. & CHALLIS, G. L. 2015. Discovery Of Microbial Natural Products By Activation Of Silent Biosynthetic Gene Clusters. *Nature Reviews Microbiology,* 13**,** 509-523.

SADEGHIAN-RIZI, T., EBRAHIMI, A., MOAZZEN, F., YOUSEFIAN, H. & JAHANIAN-NAJAFABADI, A. 2019. Improvement Of Solubility And Yield Of Recombinant Protein Expression In *E. coli* Using A Two-Step System. *Research in Pharmaceutical Sciences,* 14**,** 400-407.

SANTOS-ABERTURAS, J., CHANDRA, G., FRATTARUOLO, L., LACRET, R., PHAM, T., MIGUEL-VIOR, N., EYLES, T. & TRUMAN, A. 2019. Uncovering The Unexplored Diversity Of Thioamidated Ribosomal Peptides In Actinobacteria Using The RiPPER Genome Mining Tool. *Nucleic Acids Research,* 47.

SANTOS-ABERTURAS, J., ENGEL, J., DICKERHOFF, J., DÖRR, M., RUDROFF, F., WEISZ, K. & BORNSCHEUER, U. T. 2015. Exploration Of The Substrate Promiscuity Of Biosynthetic Tailoring Enzymes As A New Source Of Structural Diversity For Polyene Macrolide Antifungals. *ChemCatChem,* 7**,** 490-500.

SASAKI, M., OKAMURA, M., IDEO, A., SHIMADA, J., SUZUKI, F., ISHIHARA, M., KIKUCHI, H., KANDA, Y., KUNII, S. & SAKAGAMI, H. 2006. Re-Evaluation Of Tumor-Specific Cytotoxicity Of Mitomycin C, Bleomycin and Peplomycin. *Anticancer Res,* 26**,** 3373-80.

SASINDRAN, S. J., SAIKOLAPPAN, S. & DHANDAYUTHAPANI, S. 2007. Methionine Sulfoxide Reductases And Virulence Of Bacterial Pathogens. *Future Microbiology,* 2**,** 619-630.

SCHWALEN, C. J., HUDSON, G. A., KILLE, B. & MITCHELL, D. A. 2018. Bioinformatic Expansion And Discovery Of Thiopeptide Antibiotics. *Journal of the American Chemical Society,* 140**,** 9494-9501.

SCHWEIDA, D., BARRAUD, P., REGL, C., LOUGHLIN, F. E., HUBER, C. G., CABRELE, C. & SCHUBERT, M. 2019. The NMR Signature Of Gluconoylation: A Frequent N-Terminal Modification Of Isotope-Labeled Proteins. *Journal of Biomolecular NMR,* 73**,** 71-79.

SEHGAL, S. H., BAKER, H. & C., V. 1975. Rapamycin (AY-22, 989), A New Antifungal Antibiotic. *The Journal of Antibiotics,* 28**,** 727-732.

SHIMAMURA, H., GOUDA, H., NAGAI, K., HIROSE, T., ICHIOKA, M., FURUYA, Y., KOBAYASHI, Y., HIRONO, S., SUNAZUKA, T. & OMURA, S. 2009. Structure Determination And Total Synthesis Of Bottromycin A_2: A Potent Antibiotic Against MRSA And VRE. *Angew Chem Int Ed Engl,* 48**,** 914-7.

SIKANDAR, A., LOPATNIUK, M., LUZHETSKYY, A., MÜLLER, R. & KOEHNKE, J. 2022. Total *in vitro* Biosynthesis Of The Thioamitide Thioholgamide And Investigation Of The Pathway. *Journal of the American Chemical Society,* 144**,** 5136-5144.

SINGH, S. P., KAUR, S. & SINGH, D. 2017. Toxicological Profile Of Indian Foods—Ensuring Food Safety In India. *In:* GUPTA, R. K., DUDEJA & SINGH, M. (eds.) *Food Safety in the 21st Century.* San Diego: Academic Press.

SIT, C. S., YOGANATHAN, S. & VEDERAS, J. C. 2011. Biosynthesis of Aminovinyl-Cysteine-Containing Peptides and Its Application in the Production of Potential Drug Candidates. *Accounts of Chemical Research,* 44**,** 261-268.

SIVONEN, K., LEIKOSKI, N., FEWER, D. P. & JOKELA, J. 2010. Cyanobactins-Ribosomal Cyclic Peptides Produced By Cyanobacteria. *Appl Microbiol Biotechnol,* 86**,** 1213-25.

SMITH, A. W., CAMARA-ARTIGAS, A., WANG, M., ALLEN, J. P. & FRANCISCO, W. A. 2006. Structure Of Phenoxazinone Synthase From *Streptomyces antibioticus* Reveals A New Type 2 Copper Center. *Biochemistry,* 45**,** 4378-87.

SOFIA, H. J., CHEN, G., HETZLER, B. G., REYES-SPINDOLA, J. F. & MILLER, N. E. 2001. Radical SAM, A Novel Protein Superfamily Linking Unresolved Steps In Familiar Biosynthetic Pathways With Radical Mechanisms: Functional Characterization Using New Analysis And Information Visualization Methods. *Nucleic acids research,* 29**,** 1097-1106.

SONG, H., VAN DER VELDEN, N. S., SHIRAN, S. L., BLEIZIFFER, P., ZACH, C., SIEBER, R., IMANI, A. S., KRAUSBECK, F., AEBI, M., FREEMAN, M. F., RINIKER, S., KÜNZLER, M. & NAISMITH, J. H. 2018. A Molecular Mechanism For The Enzymatic Methylation Of Nitrogen Atoms Within Peptide Bonds. *Science Advances,* 4**,** eaat2720.

SØRENSEN, H. P. & MORTENSEN, K. K. 2005. Soluble Expression Of Recombinant Proteins In The Cytoplasm Of *Escherichia coli*. *Microbial Cell Factories,* 4**,** 1.

STERNER, O., ETZEL, W., MAYER, A. & ANKE, H. 1997. Omphalotin, A New Cyclic Peptide With Potent Nematicidal Activity From *Omphalotus olearius* II. Isolation And Structure Determination. *Natural Product Letters,* 10**,** 33-38.

SY-CORDERO, A. A., GRAF, T. N., RUNYON, S. P., WANI, M. C., KROLL, D. J., AGARWAL, R., BRANTLEY, S. J., PAINE, M. F., POLYAK, S. J. & OBERLIES, N. H. 2013. Enhanced Bioactivity Of Silybin B Methylation Products. *Bioorganic & Medicinal Chemistry,* 21**,** 742-747.

TAKASE, S., KUROKAWA, R., KONDOH, Y., HONDA, K., SUZUKI, T., KAWAHARA, T., IKEDA, H., DOHMAE, N., OSADA, H., SHIN-YA, K., KUSHIRO, T., YOSHIDA, M. & MATSUMOTO, K. 2019. Mechanism Of Action Of Prethioviridamide, An Anticancer Ribosomally Synthesized And Post-Translationally Modified Peptide With A Polythioamide Structure. *ACS Chemical Biology,* 14**,** 1819-1828.

TAN, W., LU, J., HUANG, M., LI, Y., CHEN, M., WU, G., GONG, J., ZHONG, Z., XU, Z., DANG, Y., GUO, J., CHEN, X. & WANG, Y. 2011. Anti-Cancer Natural Products Isolated From Chinese Medicinal Herbs. *Chinese Medicine,* 6**,** 27.

TANG, J., LU, J., LUO, Q. & WANG, H. 2018. Discovery And Biosynthesis Of Thioviridamide-Like Compounds. *Chinese Chemical Letters,* 29**,** 1022-1028.

TEGEL, H., TOURLE, S., OTTOSSON, J. & PERSSON, A. 2010. Increased Levels Of Recombinant Human Proteins With The *Escherichia coli* Strain Rosetta(DE3). *Protein Expression and Purification,* 69**,** 159-167.

TEMME, K., ZHAO, D. & VOIGT, C. A. 2012. Refactoring The Nitrogen Fixation Gene Cluster From *Klebsiella oxytoca*. *Proceedings of the National Academy of Sciences,* 109**,** 7085-7090.

TIETZ, J. I., SCHWALEN, C. J., PATEL, P. S., MAXSON, T., BLAIR, P. M., TAI, H.-C., ZAKAI, U. I. & MITCHELL, D. A. 2017. A New Genome-Mining Tool Redefines The Lasso Peptide Biosynthetic Landscape. *Nature Chemical Biology,* 13**,** 470-478.

TING, C. P., FUNK, M. A., HALABY, S. L., ZHANG, Z., GONEN, T. & VAN DER DONK, W. A. 2019. Use Of A Scaffold Peptide In The Biosynthesis Of Amino Acid-Derived Natural Products. *Science,* 365**,** 280-284.

TIWARI, P., KAILA, P. & GUPTASARMA, P. 2019. Understanding Anomalous Mobility Of Proteins On SDS-PAGE With Special Reference To The Highly Acidic Extracellular Domains Of Human E- And N-Cadherins. *ELECTROPHORESIS,* 40**,** 1273-1281.

TOMPSETT, R. & MCDERMOTT, W. 1949. Recent Advances In Streptomycin Therapy. *The American Journal of Medicine,* 7**,** 371-381.

UMEMURA, M., NAGANO, N., KOIKE, H., KAWANO, J., ISHII, T., MIYAMURA, Y., KIKUCHI, M., TAMANO, K., YU, J., SHIN-YA, K. & MACHIDA, M. 2014. Characterization Of The Biosynthetic Gene Cluster For The Ribosomally Synthesized Cyclic Peptide Bstiloxin B In *Aspergillus flavus*. *Fungal Genetics and Biology,* 68**,** 23-30.

VAN DER MEIJ, A., WORSLEY, S. F., HUTCHINGS, M. I. & VAN WEZEL, G. P. 2017. Chemical Ecology Of Antibiotic Production By *Actinomycetes*. *FEMS Microbiology Reviews,* 41**,** 392-416.

VIETTI, T. & VALERIOTE, F. 1971. Actinomycin D: Kinetics Of Its Lethal Action On A Transplantable Leukemia. *J Natl Cancer Inst,* 46**,** 1177-81.

VIOR, N. M., CEA-TORRESCASSANA, E., EYLES, T. H., CHANDRA, G. & TRUMAN, A. W. 2020. Regulation Of Bottromycin Biosynthesis Involves An Internal Transcriptional Start Site And A Cluster-Situated Modulator. *Frontiers in Microbiology,* 11.

VIVALDO, G., MASI, E., TAITI, C., CALDARELLI, G. & MANCUSO, S. 2017. The Network Of Plants Volatile Organic Compounds. *Scientific Reports,* 7**,** 11050.

WAGNER, R. L., HOCHSTEIN, F. A., MURAI, K., MESSINA, N. & REGNA, P. P. 1953. Magnamycin. A New Antibiotic1,2. *Journal of the American Chemical Society,* 75**,** 4684-4687.

WALKER, E. J., BETTINGER, J. Q., WELLE, K. A., HRYHORENKO, J. R., MOLINA VARGAS, A. M., O'CONNELL, M. R. & GHAEMMAGHAMI, S. 2022. Protein Folding Stabilities Are A Major Determinant Of Oxidation Rates For Buried Methionine Residues. *Journal of Biological Chemistry,* 298.

WALSH, C. T. & FISCHBACH, M. A. 2010. Natural Products Version 2.0: Connecting Genes To Molecules. *Journal of the American Chemical Society,* 132**,** 2469-2493.

WALTERS, C. R., SZANTAI-KIS, D. M., ZHANG, Y., REINERT, Z. E., HORNE, W. S., CHENOWETH, D. M. & PETERSSON, E. J. 2017. The Effects Of Thioamide Backbone Substitution On Protein Stability: A Study In α-Helical, β-Sheet, And Polyproline II Helical Contexts. *Chemical Science,* 8**,** 2868-2877.

WANG, H.-X., ZUO, Z.-Q., DU, J.-Z., WANG, Y.-C., SUN, R., CAO, Z.-T., YE, X.-D., WANG, J.-L., LEONG, K. W. & WANG, J. 2016a. Surface Charge Critically Affects Tumor Penetration And Therapeutic Efficacy Of Cancer Nanomedicines. *Nano Today,* 11**,** 133-144.

WANG, S., ZHENG, X., PAN, Q. & CHEN, Y. 2016b. Mutagenesis Of Precursor Peptide For The Generation Of Nosiheptide Analogues. *RSC Advances,* 6**,** 94643-94650.

WIEBACH, V., MAINZ, A., SIEGERT, M. J., JUNGMANN, N. A., LESQUAME, G., TIRAT, S., DREUX-ZIGHA, A., ASZODI, J., LE BELLER, D. & SÜSSMUTH, R. D. 2018. The Anti-

Staphylococcal Lipolanthines Are Ribosomally Synthesized Lipopeptides. *Nat Chem Biol,* 14**,** 652-654.

WILKINSON, B. & MICKLEFIELD, J. 2007. Mining And Engineering Natural-Product Biosynthetic Pathways. *Nature Chemical Biology,* 3**,** 379-386.

WILKINSON, C. J., HUGHES-THOMAS, Z. A., MARTIN, C. J., BÖHM, I., MIRONENKO, T., DEACON, M., WHEATCROFT, M., WIRTZ, G., STAUNTON, J. & LEADLAY, P. F. 2002. Increasing The Efficiency Of Heterologous Promoters In *Actinomycetes*. *J Mol Microbiol Biotechnol,* 4**,** 417-26.

WINGFIELD, P. T. 2017. N-Terminal Methionine Processing. *Current Protocols in Protein Science,* 88**,** 6.14.1-6.14.3.

WIT, R. D., STOTER, G., KAYE, S. B., SLEIJFER, D. T., JONES, W. G., HUININK, W. W. T. B., REA, L. A., COLLETTE, L. & SYLVESTER, R. 1997. Importance Of Bleomycin In Combination Chemotherapy For Good-Prognosis Testicular Nonseminoma: A Randomized Study Of The European Organization For Research And Treatment Of Cancer Genitourinary Tract Cancer Cooperative Group. *Journal of Clinical Oncology,* 15**,** 1837-1843.

XIONG, J., LUO, S., QIN, C.-X., CUI, J.-J., MA, Y.-X., GUO, M.-X., ZHANG, S.-S., LI, Y., GAO, K. & DONG, S.-H. 2022. Biochemical Reconstitution Reveals The Biosynthetic Timing And Substrate Specificity For Thioamitides. *Organic Letters,* 24**,** 1518-1523.

YAHARA, S., SHIGEYAMA, C., NOHARA, T., OKUDA, H., WAKAMATSU, K. & YASUHARA, T. 1989. Structures Of Anti-Ace And -Renin Peptides From *Lycii radicis cortex*. *Tetrahedron Letters,* 30**,** 6041-6042.

YOU, Y. O. & VAN DER DONK, W. A. 2007. Mechanistic Investigations Of The Dehydration Reaction Of Lacticin 481 Synthetase Using Site-Directed Mutagenesis. *Biochemistry,* 46**,** 5991-6000.

ZHANG, J., GUY, M. J., NORMAN, H. S., CHEN, Y.-C., XU, Q., DONG, X., GUNER, H., WANG, S., KOHMOTO, T., YOUNG, K. H., MOSS, R. L. & GE, Y. 2011. Top-Down Quantitative Proteomics Identified Phosphorylation Of Cardiac Troponin I As A Candidate Biomarker For Chronic Heart Failure. *Journal of Proteome Research,* 10**,** 4054-4065.

ZHANG, J. J., YAMANAKA, K., TANG, X. & MOORE, B. S. 2019. Direct Cloning And Heterologous Expression Of Natural Product Biosynthetic Gene Clusters By Transformation-Associated Recombination. *In:* SHUKLA, A. K. (ed.) *Methods in Enzymology.* Academic Press.

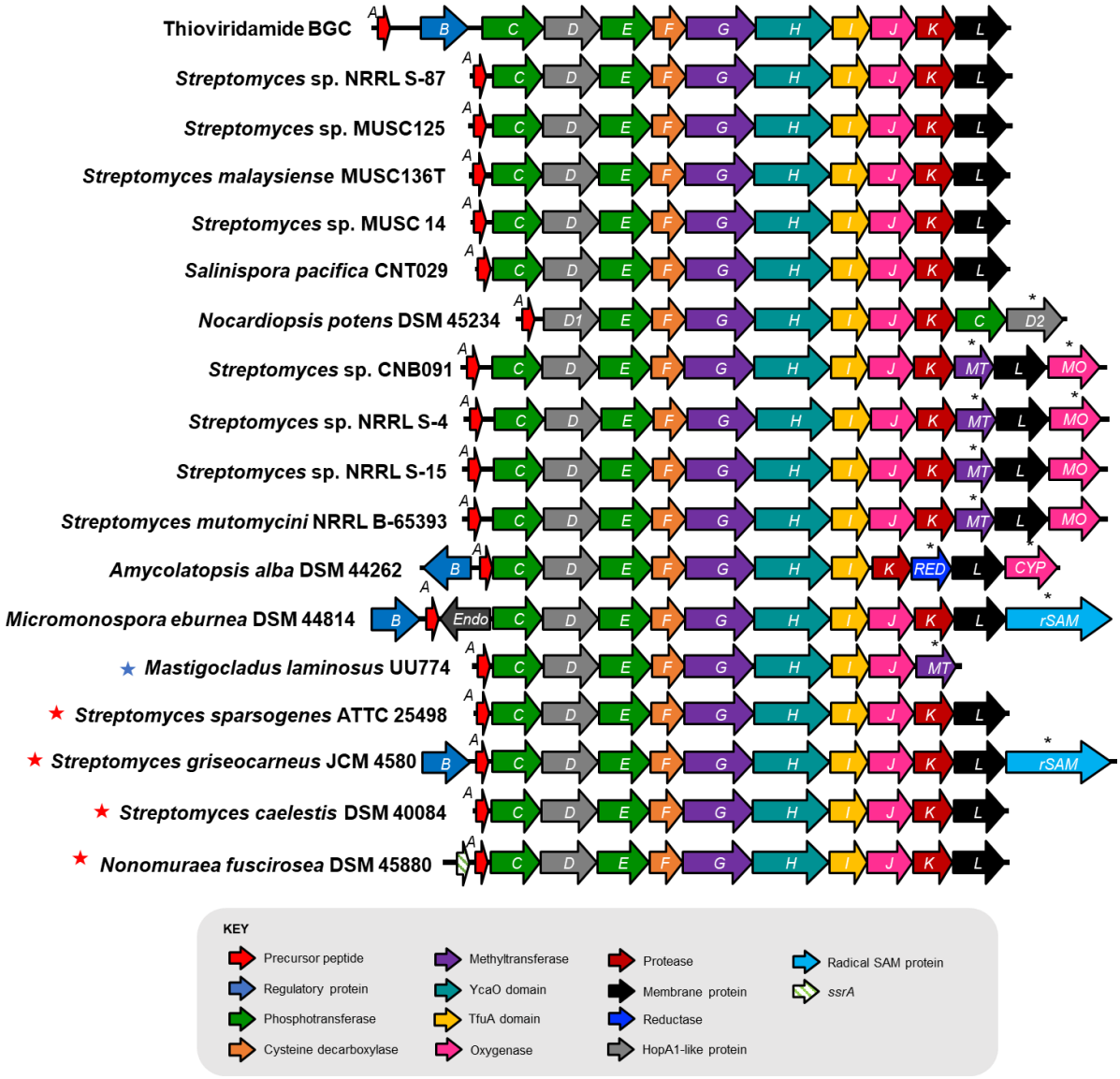
ZHANG, M. M., WONG, F. T., WANG, Y., LUO, S., LIM, Y. H., HENG, E., YEO, W. L., COBB, R. E., ENGHIAD, B., ANG, E. L. & ZHAO, H. 2017. CRISPR–Cas9 Strategy For Activation Of Silent *Streptomyces* Biosynthetic Gene Clusters. *Nature Chemical Biology,* 13**,** 607-609.

ZHANG, W., LI, J., LIU, L.-W., WANG, K.-R., SONG, J.-J., YAN, J.-X., LI, Z.-Y., ZHANG, B.-Z. & WANG, R. 2010. A Novel Analog Of Antimicrobial Peptide Polybia-MPI, With Thioamide Bond Substitution, Exhibits Increased Therapeutic Efficacy Against Cancer And Diminished Toxicity In Mice. *Peptides,* 31**,** 1832-1838.

ZHU, R., ZACHARIAS, L., WOODING, K. M., PENG, W. & MECHREF, Y. 2017. Chapter Twenty-One - Glycoprotein Enrichment Analytical Techniques: Advantages and Disadvantages. *In:* SHUKLA, A. K. (ed.) *Methods in Enzymology.* Academic Press.

ZUBIETA, C., ROSS, J. R., KOSCHESKI, P., YANG, Y., PICHERSKY, E. & NOEL, J. P. 2003. Structural Basis For Substrate Recognition In The Salicylic Acid Carboxyl Methyltransferase Family. *The Plant Cell,* 15**,** 1704-1716.

Appendices



Appendix 1: Comparison of thioamitide BGCs studied in this project. The thioviridamide BGC at the top is used as reference for comparison.

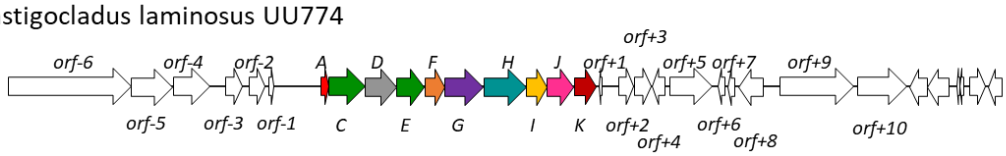
Appendix 2: BLASTP analysis of the proteins summarised in relation to thioviridamide biosynthetic pathway. *Proteins showed homology to TsaMT in thiostreptamide S4 biosynthesis

Protein annotation from Tva pathway	TvaB	TvaC	TvaD	TvaE	TvaF	TvaG
Predicted function	putative SARP family regulator	aminoglycoside phosphotransferase family protein	HopA1	phosphotransferase	Putative flavoprotein decarboxylase	methyltransferase
Actinomadura sp	-	WP_131980218	WP_131980220	WP_131980222	WP_131980224	WP_131980306
Amycolatopsis alba	WP_020634203	WP_020634201	WP_020634200	WP_020634199	WP_020634198	WP_020634197
Micromonospora eburnea	WP_091117923	WP_091117925	WP_091117926	WP_091117927	WP_091127036	WP_091117928
Nocardiopsis potens	-	WP_017595612	WP_017595620 WP_017595611	WP_017595619	WP_017595618	WP_017595617
Nonomuraea fuscirosea	WP_106236753	WP_106236740	WP_106236741	WP_146178096	WP_106236745	WP_106236747
Salinispora pacifica	-	WP_080683119	WP_027655656	WP_027655655	WP_050588302	WP_027655654
Steptomyces sp MSB090213SC12	-	BBI93417	BBI93418	BBI93419	BBI93420	BBI93421
Streptomyces malaysiense MUSC	-	WP_079175246	WP_046417231	WP_046417232	WP_053055417	WP_046417320
Streptomyces olivoviridis	BAN83917	BAN83918	BAN83919	BAN83920	BAN83921	BAN83922
Streptomyces sp CNB091	-	WP_078626022	WP_018962004	WP_018962005	WP_018962006	WP_063738229
Streptomyces sp L12	WP_129307708	WP_129307707	WP_129307706	WP_129307705	WP_129307704	WP_129311875
Streptomyces sp MUSC125	-	WP_078886290	WP_039653089	WP_039653090	WP_052269935	WP_052269936
Streptomyces sp S-4 S-15	-	WP_078850504	WP_053929453	WP_053929452	WP_051834497	WP_031090931
Streptomyces sp S-87	-	WP_078909880	WP_030193573	WP_107048384	WP_051794611	WP_030193568
Streptomyces sparsogenes	-	WP_079150874	WP_065960366	WP_065960368	WP_065960370	WP_079150890
Streptomyces griseocarneus JCM 4580	WP_190061744	WP_190061746	WP_190061747	WP_190061748	WP_190061749	WP_190061750
Streptomyces caelestis DSM 40084	-	WP_184984770	WP_184984772	WP_184984774	WP_184984776	WP_184993455
Mastigocladus laminosus UU774	-	WP_044449648	WP_044449646	WP_044449644	WP_052508307	WP_135107295

Protein annotation from Tva pathway	TvaH	TvaI	TvaJ	TvaK	TvaL	TsaMT*
Predicted function	YcaO-related McrA-glycine thioamidation protein	TfuA-like core domain-containing protein	putative phytanoyl-CoA dioxygenase family protein	papain family cycsteine protease	membrane protein	methyltransferase domain-containing protein
Actinomadura sp	WP_131980226	WP_131980228	WP_131980230	WP_131980232	WP_131980214	-
Amycolatopsis alba	WP_020634196	WP_020634195	-	WP_051137548	WP_039794118	-
Micromonospora eburnea	WP_091117929	WP_091117930	WP_091117931	WP_091117932	-	-
Nocardiopsis potens	WP_017595616	WP_017595615	WP_017595614	WP_017595613	WP_017595624	-
Nonomuraea fuscirosea	WP_106236748	WP_106236750	WP_106236751	WP_146178097	-	-
Salinispora pacifica	WP_027655653	WP_027655652	WP_080683118	WP_080683117	WP_027655649	-
Steptomyces sp MSB090213SC12	BBI93422	BBI93423	BBI93424	BBI93425	BBI93426	-
Streptomyces malaysiense MUSC	WP_046417320	WP_046417234	WP_053055418	WP_046417235	WP_046417323	-
Streptomyces olivoviridis	BAN83923	BAN83924	BAN83925	BAN83926	BAN83927	
Streptomyces sp CNB091	WP_018962008	WP_051114702	WP_018962010	WP_018962011	WP_106434510	WP_078626023
Streptomyces sp L12	WP_129307703	WP_129307702	WP_129307701	WP_129307700	WP_129307699	-
Streptomyces sp MUSC125	WP_039653092	WP_039653094	WP_052269937	WP_039653095	WP_039653248	-
Streptomyces sp S-4 S-15	WP_031090932	WP_051834498	WP_051834499	WP_031090940	WP_042836196	WP_078850506
Streptomyces sp S-87	WP_030193566	WP_030193564	WP_051794609	WP_078909879	WP_051794607	WP_051794607
Streptomyces sparsogenes	WP_065960374	WP_065960379	WP_065960381	WP_065960382	WP_065960597	WP_065960597
Streptomyces griseocarneus JCM 4580	WP_190061751	WP_190061752	WP_190061753	WP_190061754	WP_190061755	-
Streptomyces caelestis DSM 40084	WP_184984777	WP_184984778	WP_184984779	WP_184984780	WP_184984781	-
Mastigocladus laminosus UU774	WP_044449639	WP_044449638	WP_044449636	-	-	WP_135107297.1

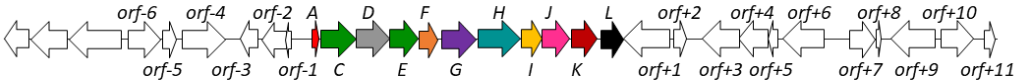
Appendix 3: Surrounding proteins near the novel and updated thioamitide BGCs

Mastigocladus laminosus UU774



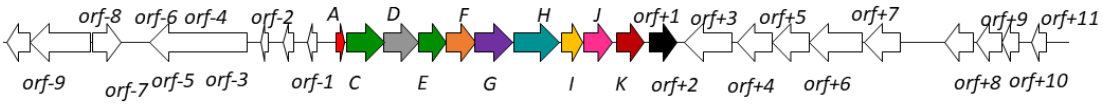
Query gene	Assession number	Conserved domains	Homologous gene annotation	Organism	Coverage (%)	Identity (%)
orf-6	WP_135107293.1	PEP_TPR_lipo	tetratricopeptide repeat protein	Nostoc sp. GBBB01	100	83.14
orf-5	WP_044449656.1	TadD	tetratricopeptide repeat protein	Nostoc sp. LEGE 12447	100	82.23
orf-4	WP_044449654.1	TPR_12	tetratricopeptide repeat protein	Nostoc sp. LEGE 12447	98	85.71
orf-3	WP_044449652.1	InsG superfamily	ISH3 family transposase	Nostoc punctiforme	97	92.35
orf-2	WP_044449731.1	InsG superfamily	ISH3 family transposase	Nostoc sp. KVJ3	100	92.35
orf-1	-	No putative conserved domains	IS4 transposase	Nostoc flagelliforme CCNUN1	100	91.67
			Thioamitide BGC			
orf+1	-	No putative conserved domains	IS1 family transposase	Fischerella muscicola	96.00	96.77
orf+2	WP_135105769.1	Transpos_IS630	helix-turn-helix domain-containing protein	Hapalosiphonaceae cyanobacterium JJU2	98.00	87.80
orf+3	-	Transpos_IS630	IS630 family transposase	Rivularia sp. T60_A2020_040	99	86.17
orf+4	-	COG5659	IS701 family transposase	Nostoc sp. C052	100.00	90.23
orf+5	-	No putative conserved domains	Transposase	Fischerella muscicola	97.00	86.13
orf+6	-	Pep_deformylase	Peptide deformylase	Fischerella sp. PCC 9339	100.00	97.01
orf+7	WP_017309361.1	No putative conserved domains	Hypothetical protein	Fischerella major NIES-592	100.00	91.67
orf+8	WP_044451813.1	No putative conserved domains	Hypothetical protein	Fischerella sp. PCC 9339	100.00	99.24
orf+9	-	P-type_ATPase_cation	HAD-IC family P-type ATPase	Westiellopsis prolifica	99.00	99.59
orf+10	WP_044451481.1	NhaP superfamily	Sodium proton antiporter	Hapalosiphonaceae	100.00	99.42

Streptomyces sparsogenes strain ATCC 25498



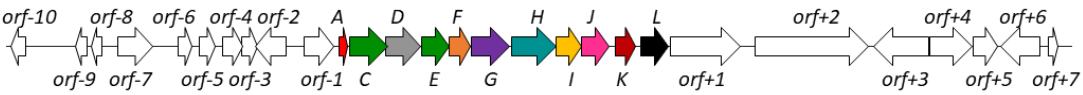
Query gene	Assession number	Conserved domains	Homologous gene annotation	Organism	Coverage (%)	Identity (%)
orf-7	WP_104531259.1	Zinc peptidase	Allantoate amidohydrolase	Streptomyces sp. NBS 14/10	99.00	94.67
orf-6	WP_076972185.1	Urocanate hydratase	Urocanate hydratase	Streptomyces sp. NBS 14/10	99.00	97.29
orf-5	WP_065960356.1	Helix-Turn-Helix DNA binding domain	MerR family transcriptional regulator	Streptomyces sp. NBS 14/10	100.00	84.87
orf-4	WP_065960357.1	Type III Pyridoxal 5-phosphate (PLP)-Dependent	Diaminopimelate decarboxylase	Streptomyces bingchenggensis BCW-1	95.00	89.26
orf-3	WP_245737961.1	No putative conserved domains	Hypothetical protein	Streptomyces hygroscopicus	99.00	80.50
orf-2	WP_245737962.1	RpiR superfamily	MurR/RpiR family transcriptional regulator	Streptomyces sp. NBS 14/10	100.00	94.87
orf-1	-	No putative conserved domains	Esterase-like activity of phytase family protein	Prauserella halophila	62.00	59.38
			Thioamitide BGC			
orf+1	WP_065960597.1	No putative conserved domains	Hypothetical protein	Streptomyces caelestis	99.00	85.66
orf+2	WP_065960386.1	PRK07787	acyl-CoA synthetase	Streptomyces bingchenggensis BCW-1	100.00	90.74
orf+3	WP_065960391.1	No putative conserved domains	Membrane protein	Streptomyces bingchenggensis BCW-1	97.00	84.67
orf+4	WP_141712972.1	Methyltr_grsp	Protein-L-isoaspartate O-methyltransferase	Streptomyces zagrosensis	93.00	57.22
orf+5	WP_065960393.1	GRASP_SAV_5884	RimK domain-containing protein	Streptomyces sparsogenes DSM 40356	100.00	100.00
orf+6	WP_107500864.1	No putative conserved domains	Putative ATP-grasp-modified RiPP	Streptomyces sp. IMTB 2501	88.00	92.47
orf+7	WP_079150877.1	Nudix_hydrolase	NUDIX hydrolase	Streptomyces sp. AJS327	90.00	53.99
orf+8	WP_065960399.1	HTH_XRE/DUF5753 superfamily	Helix-turn-helix transcriptional regulator	Streptomyces yokosukanensis	100.00	84.17
orf+9	WP_065960401.1	DUF397	DUF397 domain-containing protein	Streptomyces monashensis	100.00	74.24
orf+10	WP_065960402.1	Cysta_beta	Cystathionine beta-synthase	Streptomyces sp. NBS 14/10	99.00	96.31
orf+11	WP_079150878.1	FeeA_FeeB_like/SGNH_hydrolase	SGNH/GDSL hydrolase family protein	Streptomyces sp. NBS 14/10	90.00	92.63
orf+12	WP_065960406.1	PRK07851 superfamily	Acetyl-CoA C-acetyltransferase	Streptomyces sp. NBS 14/10	100.00	98.28

Streptomyces caelestis strain DSM 40084



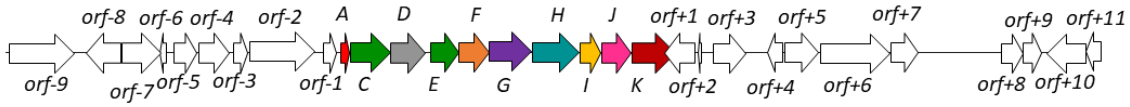
Query gene	Assession number	Conserved domains	Homologous gene annotation	Organism	Coverage (%)	Identity (%)
orf-7	WP_184984756.1	MdlB superfamily	ABC transporter ATP-binding protein	Streptomyces spinoverrucosus	100.00	97.72
orf-6	WP_184984758.1	MdlB superfamily	ABC transporter ATP-binding protein	Streptomyces flaveus	100.00	96.81
orf-5	MBB5795398.1	HTH_18	AraC family transcriptional regulator	Streptomyces albicerus	100.00	96.06
orf-4	WP_184984760.1	DAP2	Hypothetical protein	Streptomyces coeruleorubidus	100.00	96.48
orf-3	WP_184984762.1	No putative domains	Hypothetical protein	Streptomyces caelestis	100.00	100.00
orf-2	WP_184984764.1	Ribosomal L7 L12	Ribosomal protein L7/L12	Streptomyces coeruleorubidus	100.00	96.59
orf-1	WP_184984766.1	No putative domains	Hypothetical protein	Streptomyces sp. tea 10	58.00	71.74
			Thioamitide BGC			
orf+1	WP_184984783.1	GH43-Bt3655-like/AbfB	Glycoside hydrolase family 43 protein	Streptomyces azureus	100.00	98.11
orf+2	WP_184984784.1	PRK11618	Sugar ABC transporter permease YjfF	Streptomyces lincolnensis	100.00	98.54
orf+3	WP_184984785.1	AraH superfamily	ABC transporter permease	Streptomyces qaidamensis	100.00	99.45
orf+4	WP_184984786.1	AglA superfamily	Sugar ABC transporter ATP-binding protein	Streptomyces variegatus	100.00	98.44
orf+5	WP_184984787.1	PBP1_galactofuranose_YtfQ-like	ABC transporter substrate-binding protein	Streptomyces variegatus	100.00	99.40
orf+6	WP_184984788.1	Periplasmic_Binding_Protein Type II	Transporter substrate-binding domain-containing protein	Streptomyces sp. SA15	96.00	89.96
orf+7	WP_184984789.1	No putative domais	Hypotentical protein	Streptomyces sp. SA15	99.00	86.96
orf+8	WP_184984791.1	No putative domais	Hypotentical protein	Streptomyces sp. SA15	100.00	77.65
orf+9	WP_184993457.1	No putative domais	LPXTG cell wall anchor domain-containing protein	Streptomyces phyllanthi	93.00	79.82
orf+10	WP_184984792.1	Tryp_SPc/P-loop_NTPase superfamily	Serine protease	Streptomyces phyllanthi	99.00	89.36

Streptomyces griseocarneus strain JCM 4580



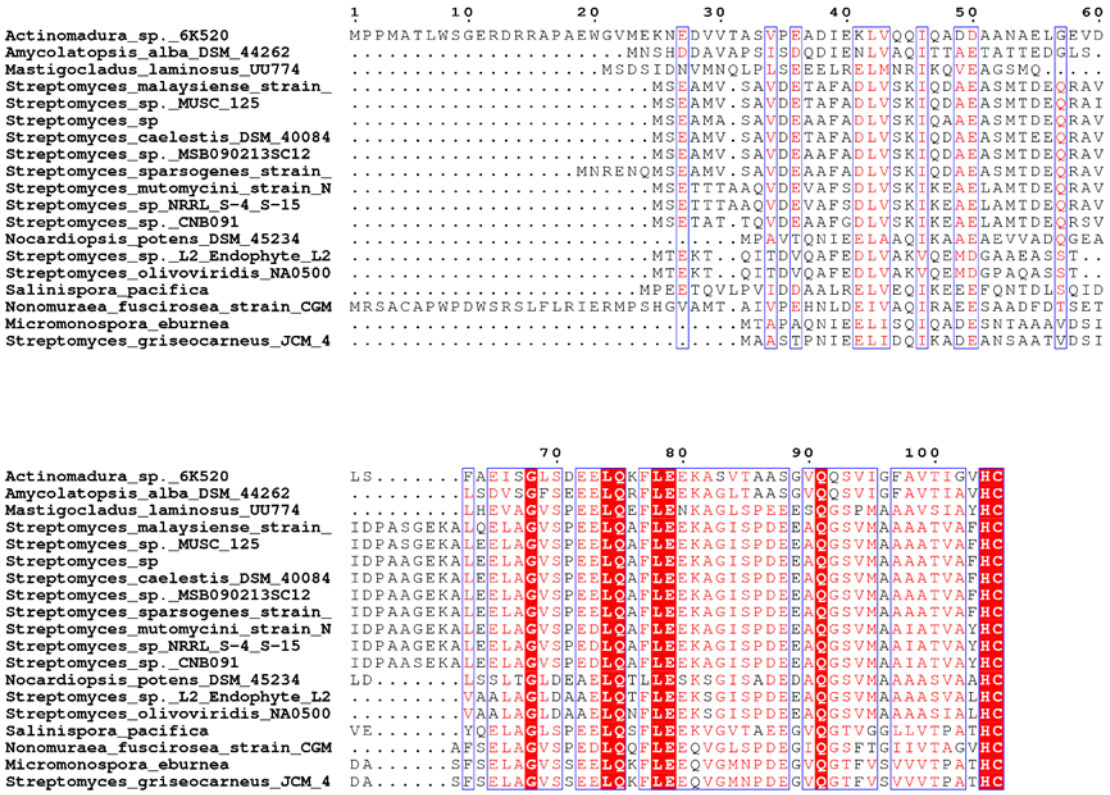
Query gene	Assession number	Conserved domains	Homologous gene annotation	Organism	Coverage (%)	Identity (%)
orf-10	WP_190061737.1	No putative conserved domains	Hypotentical protein	Streptomyces sp. CB03234	100.00	69.64
orf-9	WP_190061738.1	PRCH superfamily	PRC-barrel domain-containing protein	Streptomyces sp. ISL-100	100.00	89.83
orf-8	WP_190061739.1	PUF2267 superfamily	Hypothetical protein GCM10012286_80460	Streptomyces lasiicapitis	100.00	98.47
orf-7	-	IbpA superfmily	Hsp20/alpha crystallin family protein	Streptomyces scabiei	99.00	98.25
orf-6	WP_190061740.1	PDDEXK_nucleaase-like	Restriction endonuclease	Streptomyces sp. UNOC14_S4	99.00	93.95
orf-5	WP_190061741.1	STAS superfamily	STAS domain-containing protein	Streptomyces cinnamoneus	100.00	78.67
orf-4	WP_190062374.1	No putative conserved domains	Hypothetical protein BLA24_18970	Streptomyces cinnamoneus	100.00	89.24
orf-3	WP_190061742.1	No putative conserved domains	Hypotentical protein	Streptomyces sp. ISL-11	93.00	87.72
orf-2	WP_190061743.1	No putative conserved domains	Hypotentical protein	Streptomyces sp. NRRL B-1677	100.00	87.57
orf-1	-	No putative conserved domains	Hypotentical protein	Streptomyces avermitilis	29.00	34.88
			Thioamitide BGC			
orf+1	WP_190061756.1	YgiQ superfamily	radical SAM protein	Streptomyces sp. UNOB3_S3	97.00	96.13
orf+2	WP_190062375.1	PRK07003/BRR2	DEAD/DEAH box helicase	Streptomyces noursei	96.00	87.91
orf+3	WP_190061757.1	Abhydrolase superfamily	Alpha/beta hydrolase	Streptomyces sp. HU2014	88.00	78.81
orf+4	WP_190061758.1	COG4585	Histidine kinase	Streptomyces inhibens	100.00	84.67
orf+5	WP_190062376.1	CitB superfamily	Response regulator transcription factor	Streptomyces kasugaensis	96.00	96.28
orf+6	WP_190061759.1	DUF6242	Hypothetical protein	Streptomyces roseifaciens	93.00	68.02
orf+7	WP_190062437.1	No putative conserved domains	Hypothetical protein	Wenjunlia vitaminophila	64.00	89.06

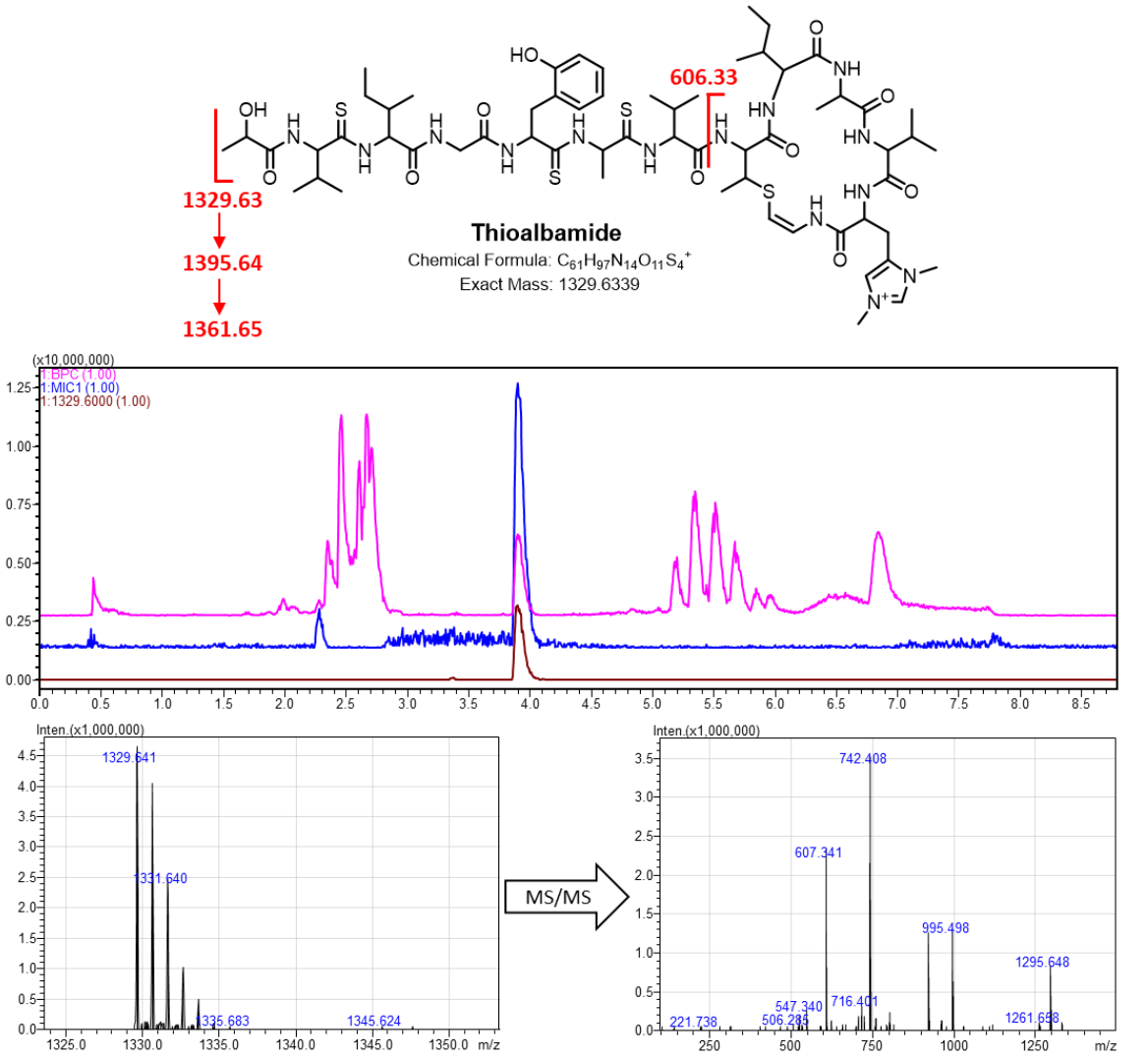
Nonomuraea fuscirosea strain CGMCC 4.7104 Ga0171605



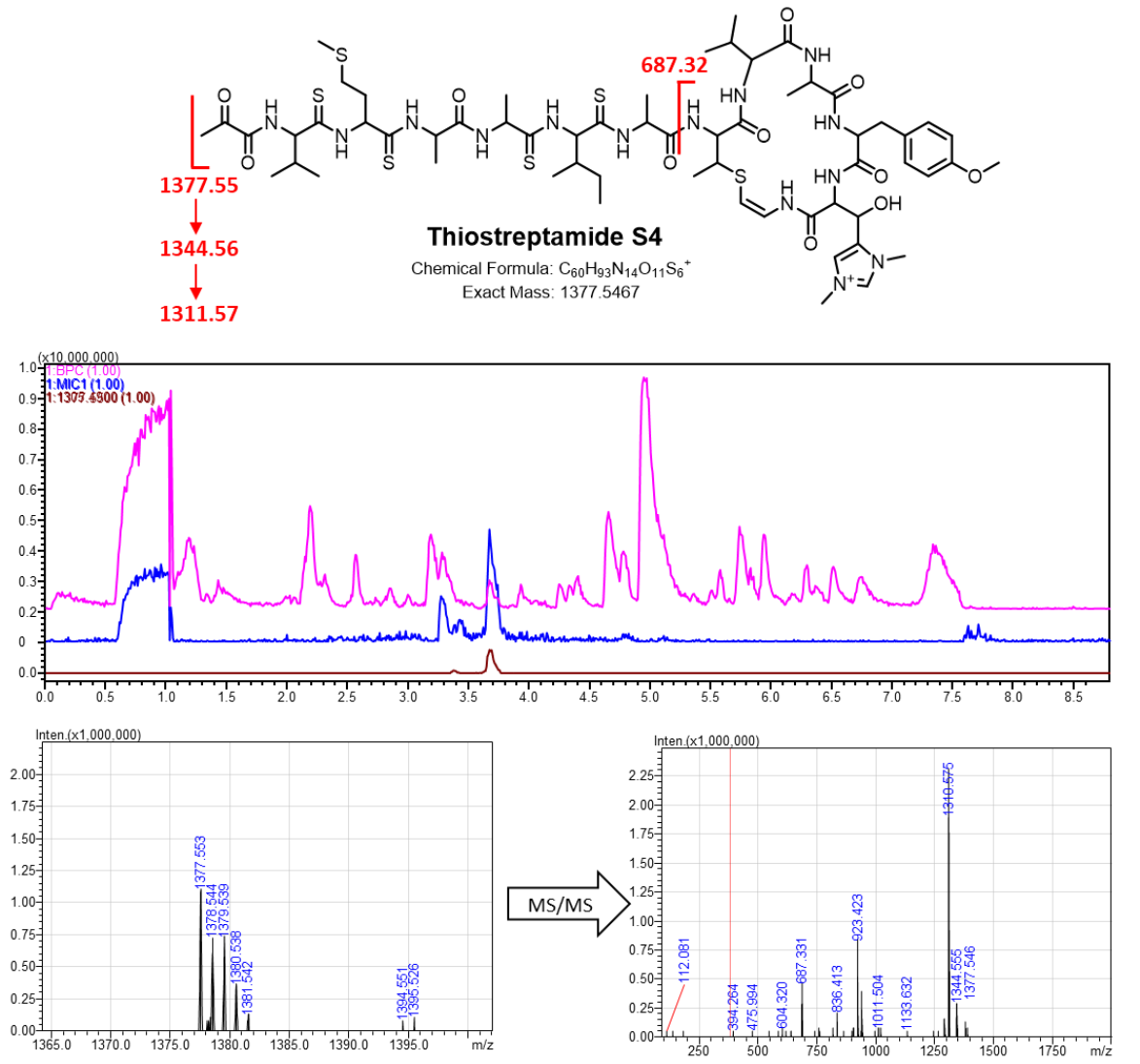
Query gene	Assession number	Conserved domains	Homologous gene annotation	Organism	Coverage (%)	Identity (%)
orf-8	WP_106236727.1	Bac_GDH	NAD-glutamate dehydrogenase	Nonomuraea thailandensis	100.00	93.23
orf-7	WP_104531259.1	Zinc peptidase	Allantoate amidohydrolase	Streptomyces sp. NBS 14/10	99.00	94.67
orf-6	WP_076972185.1	Urocanate hydratase	Urocanate hydratase	Streptomyces sp. NBS 14/10	99.00	97.29
orf-5	WP_065960356.1	Helix-Turn-Helix DNA binding domain	MerR family transcriptional regulator	Streptomyces sp. NBS 14/10	100.00	84.87
orf-4	WP_065960357.1	Type III Pyridoxal 5-phosphate (PLP)-Dependent	Diaminopimelate decarboxylase	Streptomyces bingchenggensis BCW-1	95.00	89.26
orf-3	WP_245737961.1	No putative conserved domains	Hypothetical protein	Streptomyces hygroscopicus	99.00	80.50
orf-2	WP_245737962.1	RpiR superfamily	MurR/RpiR family transcriptional regulator	Streptomyces sp. NBS 14/10	100.00	94.87
orf-1	-	No putative conserved domains	Esterase-like activity of phytase family protein	Prauserella halophila	62.00	59.38
			Thioamitide BGC			
orf+1	WP_106236753.1	BTAD	AfsR/SARP family transcriptional regulator	Nonomuraea aurantiaca	89.00	84.59
orf+2	WP_146178098.1	No putative conserved domains	Hypothetical protein	Nonomuraea thailandensis	99.00	83.66
orf+3	WP_106236757.1	No putative conserved domains	Hypothetical protein	Nonomuraea solani	95.00	84.02
orf+4	WP_106236758.1	No putative conserved domains	Hypothetical protein	Nonomuraea thailandensis	98.00	85.93
orf+5	WP_106236759.1	GlxA superfamily	GlxA family transcriptional regulator	Nonomuraea zeae	100.00	89.75
orf+6	WP_106236761.1	PKc-like superfamily	Serine/threonine-protein kinase	Nonomuraea candida	84.00	71.00
orf+7	WP_245955697.1	HTH_BTAD superfamily	AfsR/SARP family transcriptional regulator	Nonomuraea terrae	89.00	83.26
orf+8	WP_146178113.1	NB-ARC/PEP_TPR_lipo	AfsR/SARP family transcriptional regulator	Nonomuraea polychroma	99.00	77.03
orf+9	WP_106236762.1	No putative conserved domains	Hypothetical protein	Nonomuraea gerenzanensis	100.00	77.61
orf+10	WP_106236763.1	OmpA_C like superfamily	OmpA family protein	Nonomuraea basaltis	100.00	80.00
orf+11	WP_106236765.1	degP-htrA_Do	Trypsin-like peptidase domain-containing protein	Actinomadura parvosata subsp. Kistnae	100.00	86.54
orf+12	WP_106236766.1	PRK08276	AMP-binding protein	Nonomuraea thailandensis	100.00	91.60

Appendix 4: Multiple alignment of the 18 thioamitide precursor peptides from the identified BGCs. MUSCLE used for alignment, which is visualised using ESPript 3.0 (Robert and Gouet, 2014, Madeira et al., 2022). Identical residues are showed in white with a red background, while similar residues (Risler matrix set with global score of 0.7) are coloured red and are boxed.

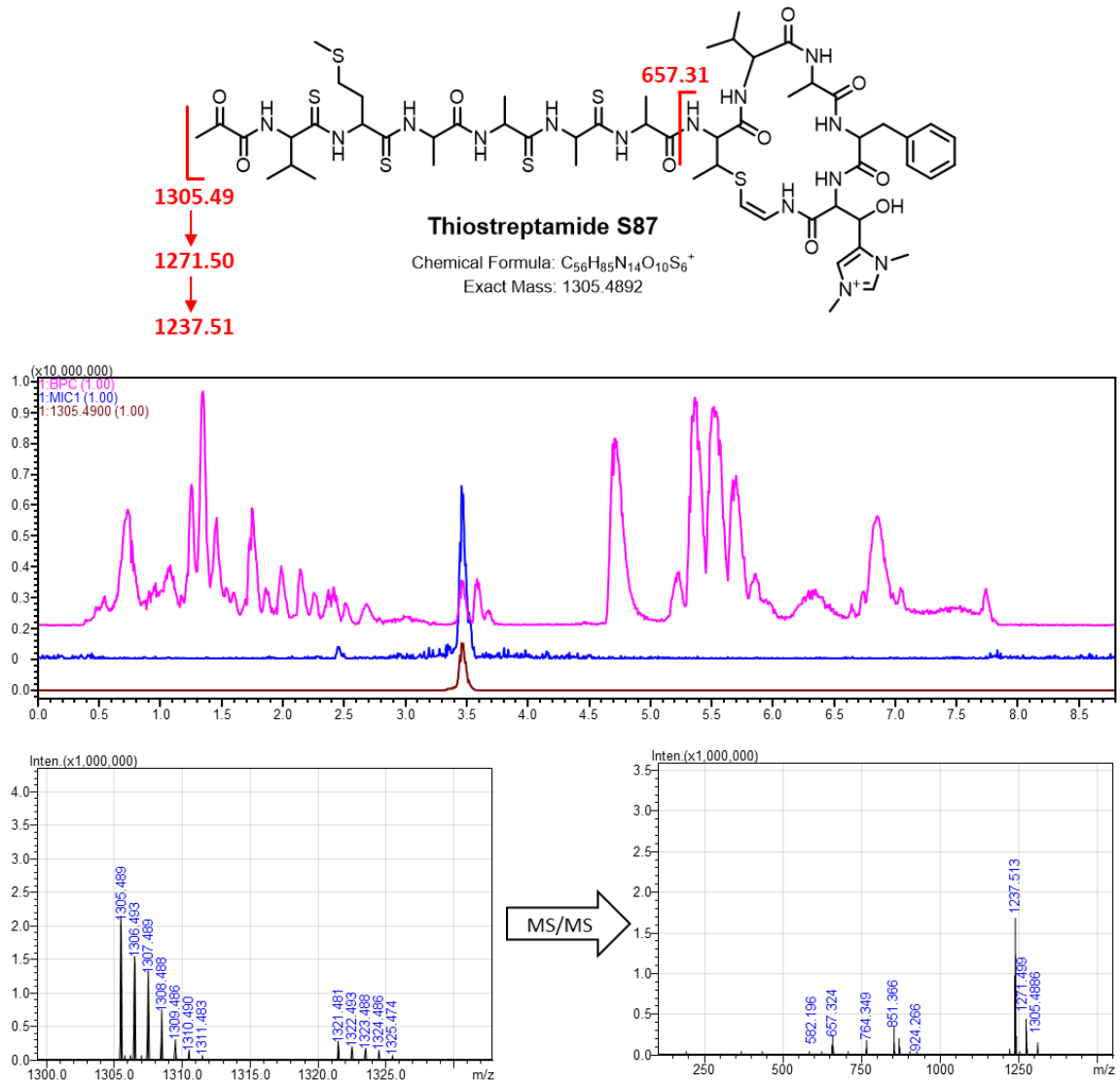




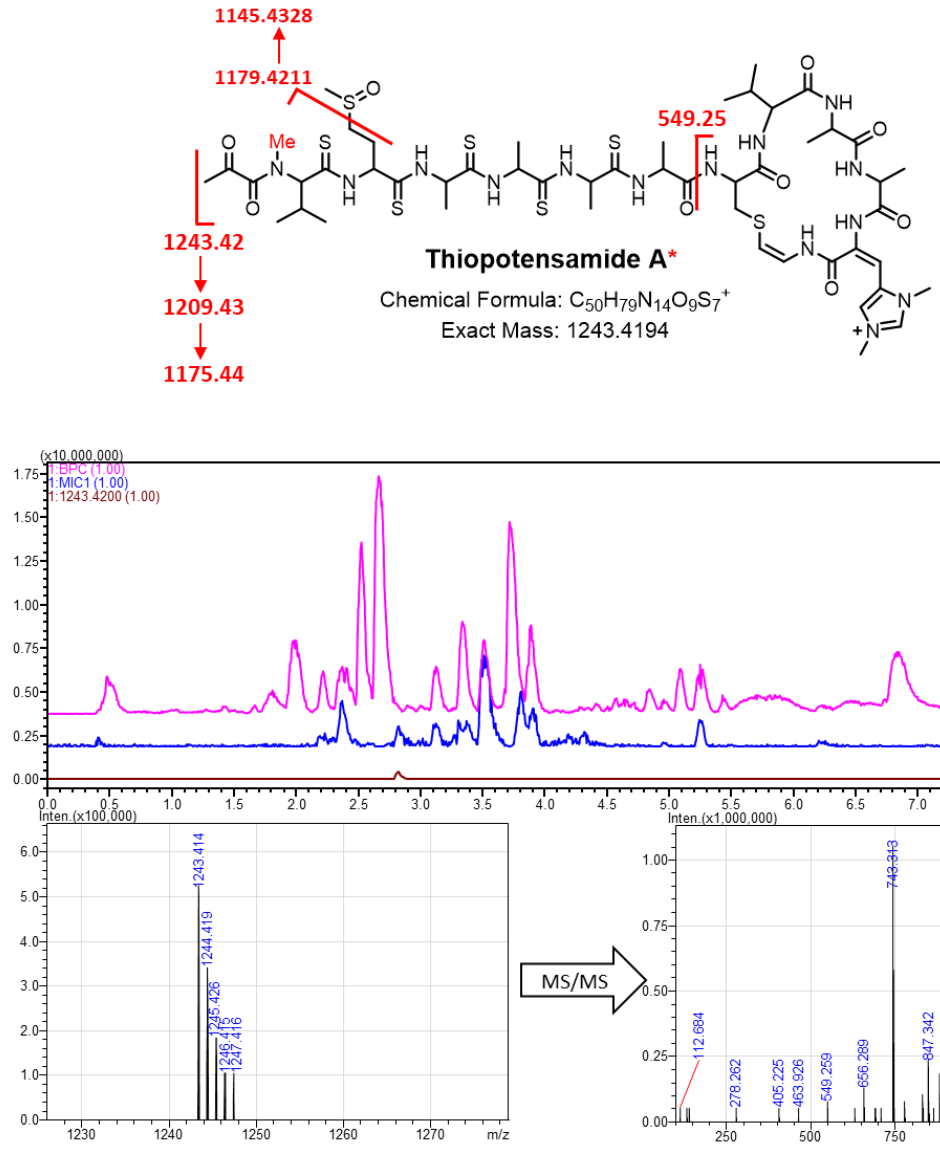
Appendix 5: Production of thioalbamide by *Amycolatopsis alba* DSM 44262 and MS/MS analysis of thioalbamide. Precursor ion was given by M^+ *m/z* 1329.63. Expected product ions were shown on the molecule from neutral loses of H_2S ($\Delta m/z$ = 33.99) and macrocycle.



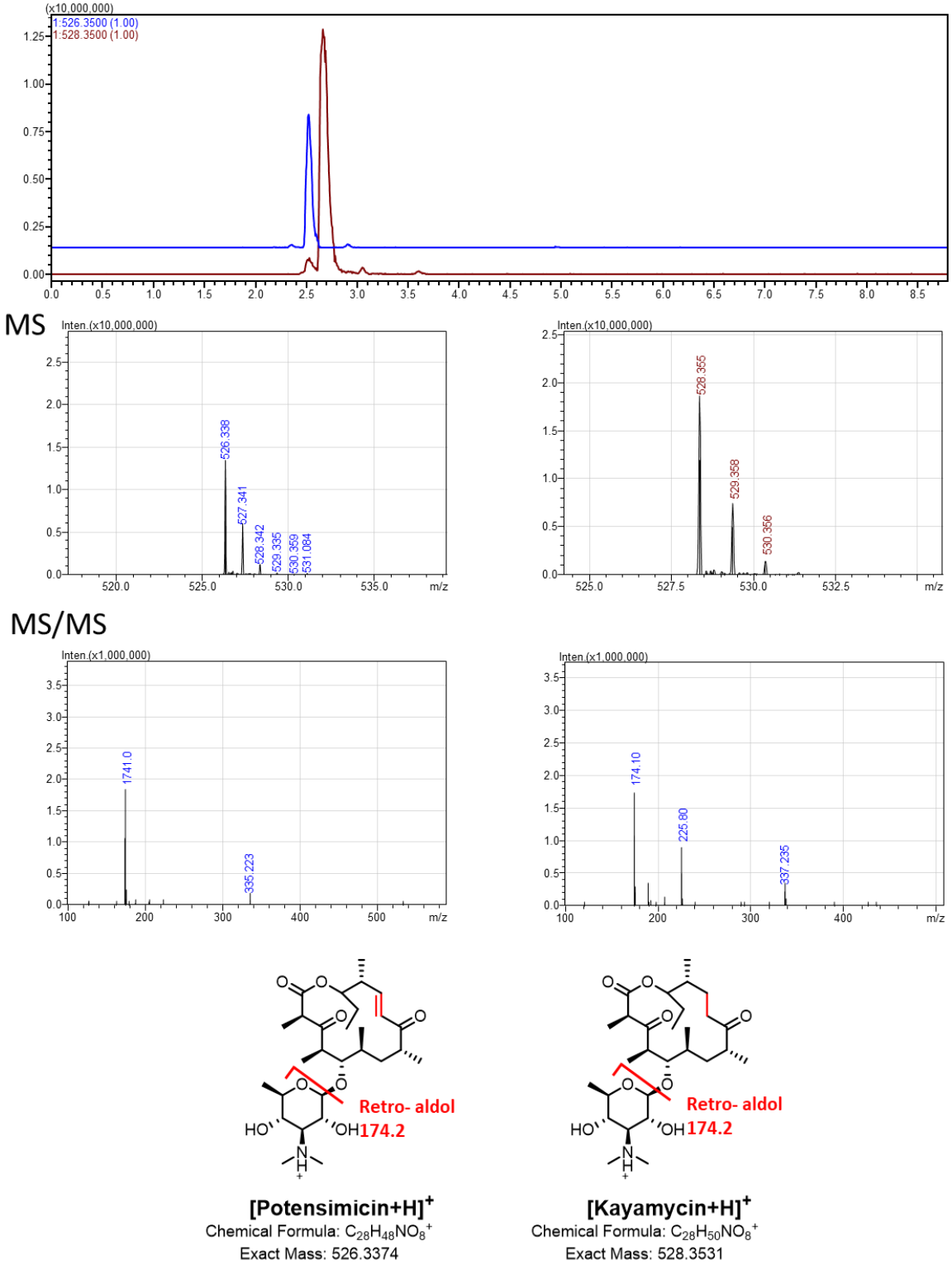
Appendix 6: Production of thiostreptamide S4 by *Streptomyces* sp. NRRL S-4 and MS/MS analysis of thiostreptamide S4. Precursor ion was given by M^+ *m/z* 1377.55. Expected product ions were shown on the molecule from neutral loses of H_2S ($\Delta m/z$ = 33.99) and macrocycle.



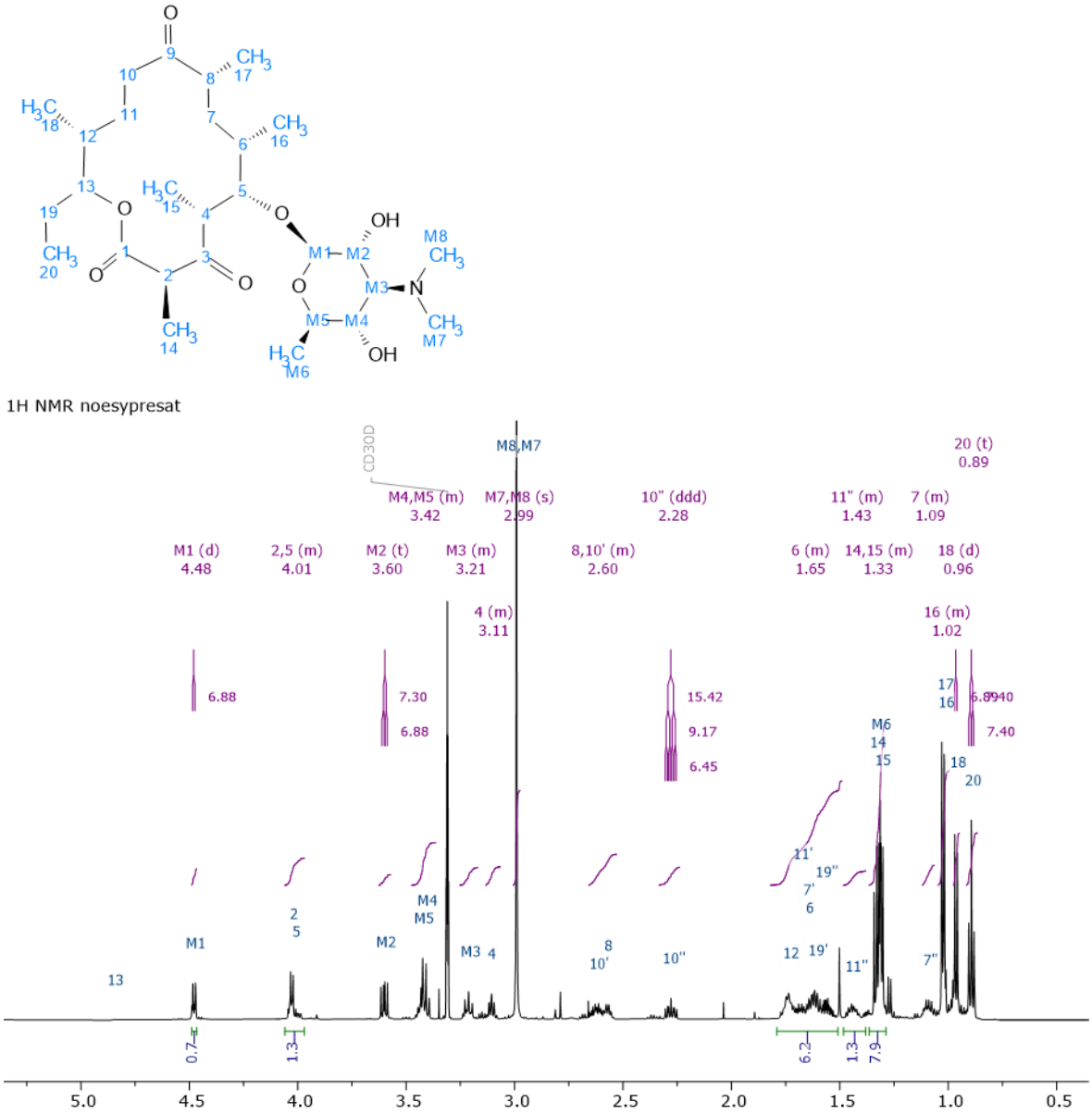
Appendix 7: Production of thiostreptamide S87 by *Streptomyces* sp. NRRL S-87 and MS/MS analysis of thiostreptamide S87. Precursor ion was given by M^+ *m/z* 1305.49. Expected product ions were shown on the molecule from neutral loses of H_2S ($\Delta m/z$ = 33.99) and macrocycle.



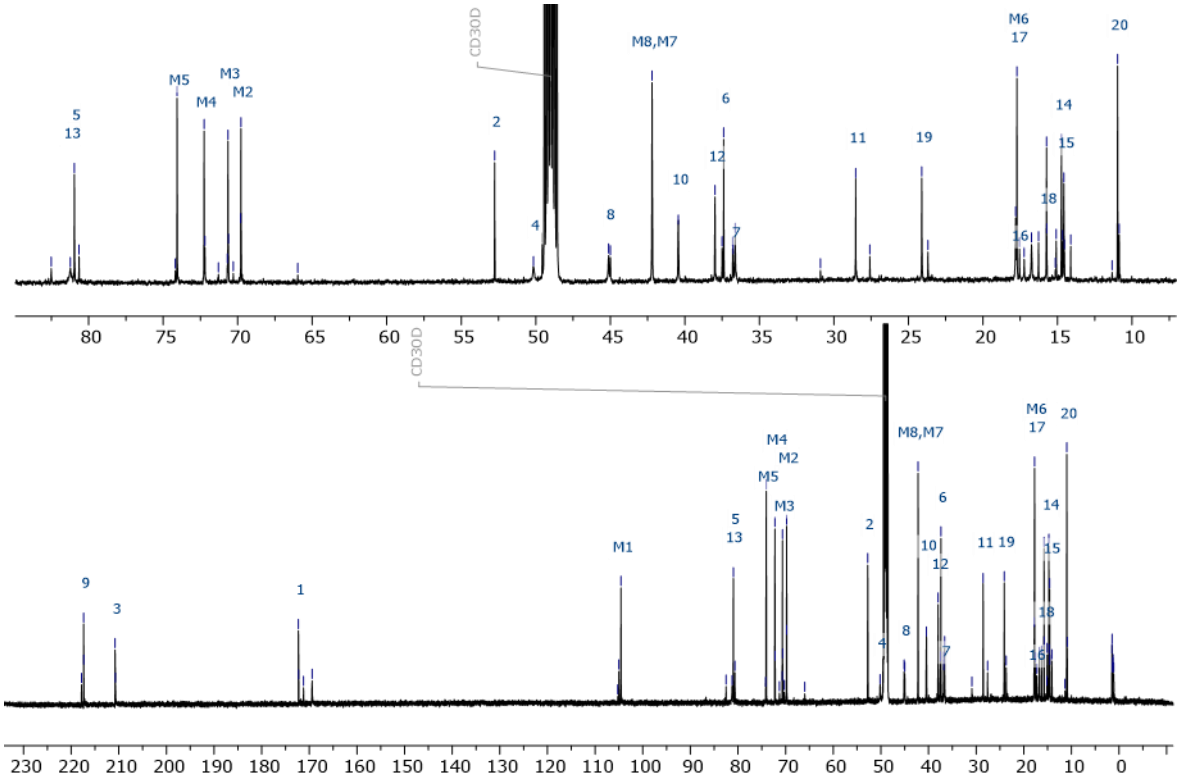
Appendix 8: Appendix 9: Production of thiopotensamide A by *Nocardiopsis potens* DSM45234 and MS/MS analysis of thiopotensamide A. Precursor ion was given by M^+ *m/z* 1243.42. Expected product ions were shown on the molecule from neutral loses of H_2S ($\Delta m/z$ = 33.99) and macrocycle.



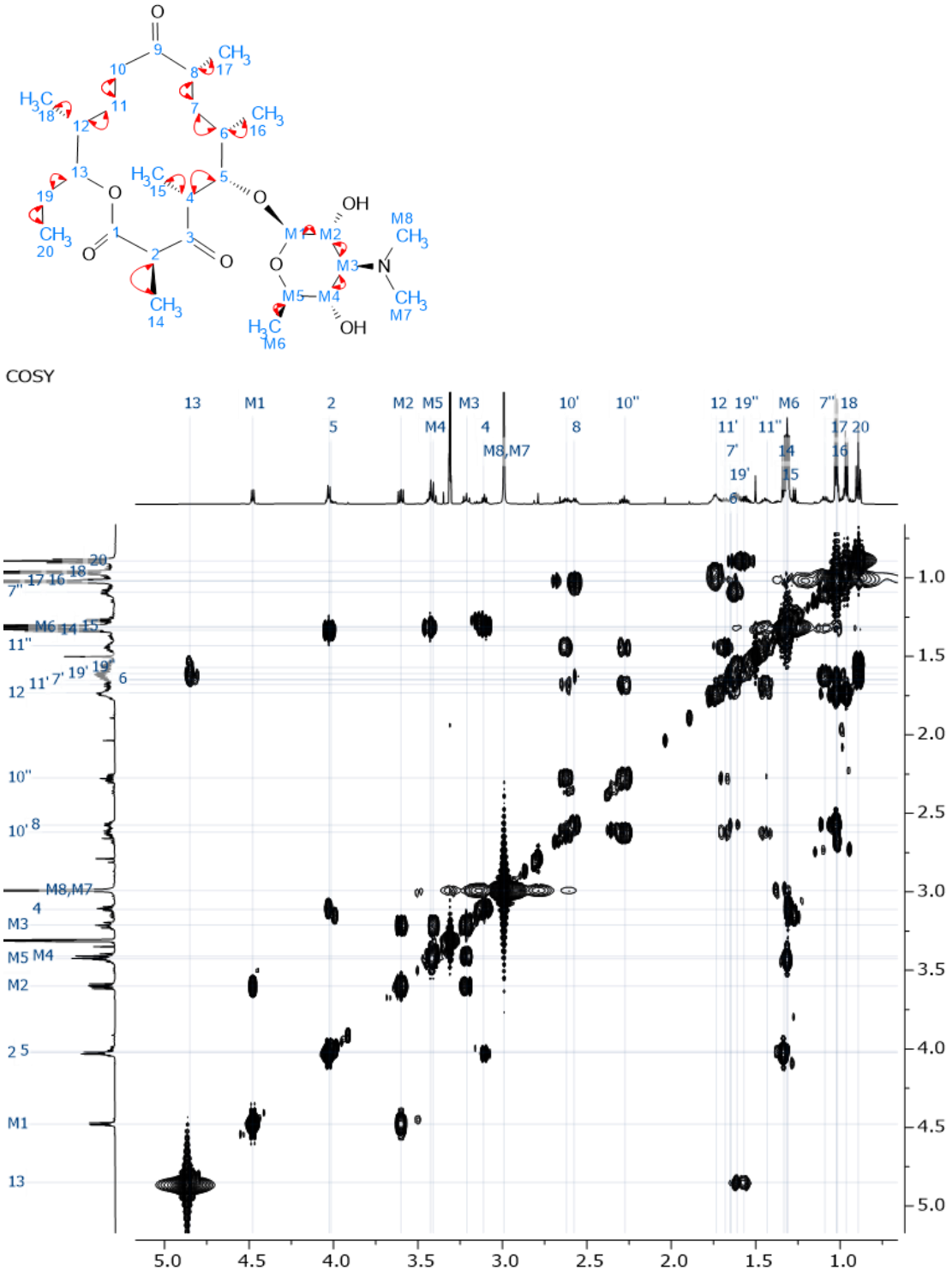
Appendix 10: Production of potensimicin and kayamycin from *N. potens*. MS/MS fragmentation showed both compounds had mycaminose sugar.



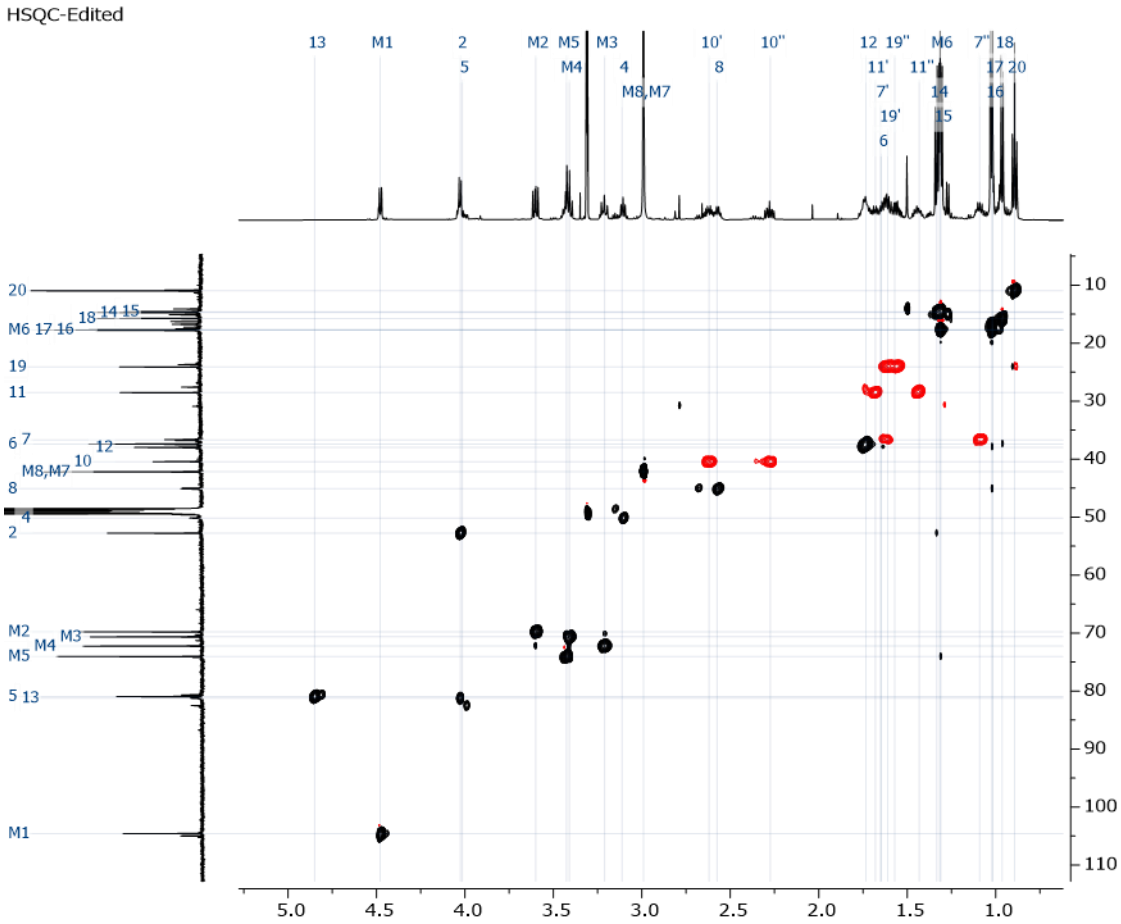
Appendix 11: ^{1}H NMR (600 MHz, CD3OD, 298K) of kayamycin recorded with water suppression (noesypresat experiment). Note that peak of H-13 is not visible because its position is coincided with suppressed water peak.



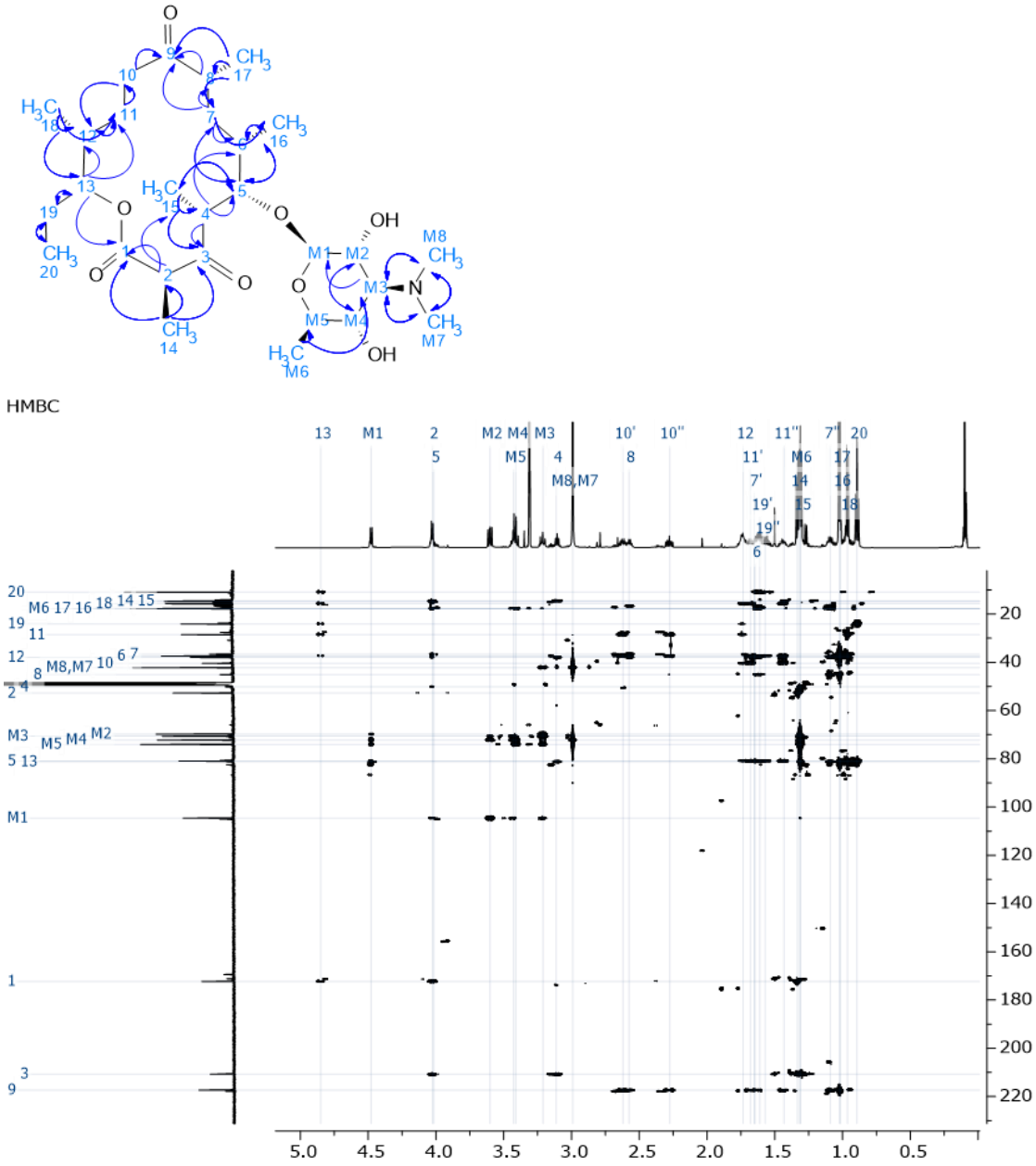
Appendix 12: ^{13}C NMR (150 MHz, CD_3OD, 298K) of kayamycin



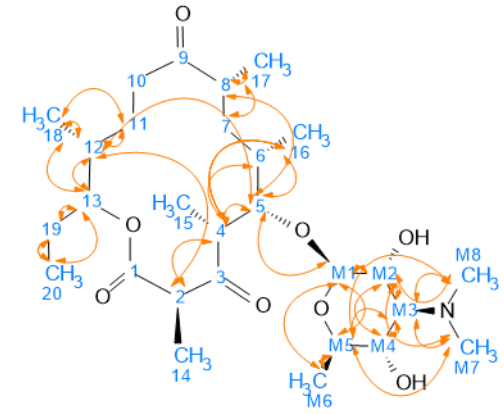
Appendix 13: COSY NMR of kayamycin and the interactions observed in COSY NMR.



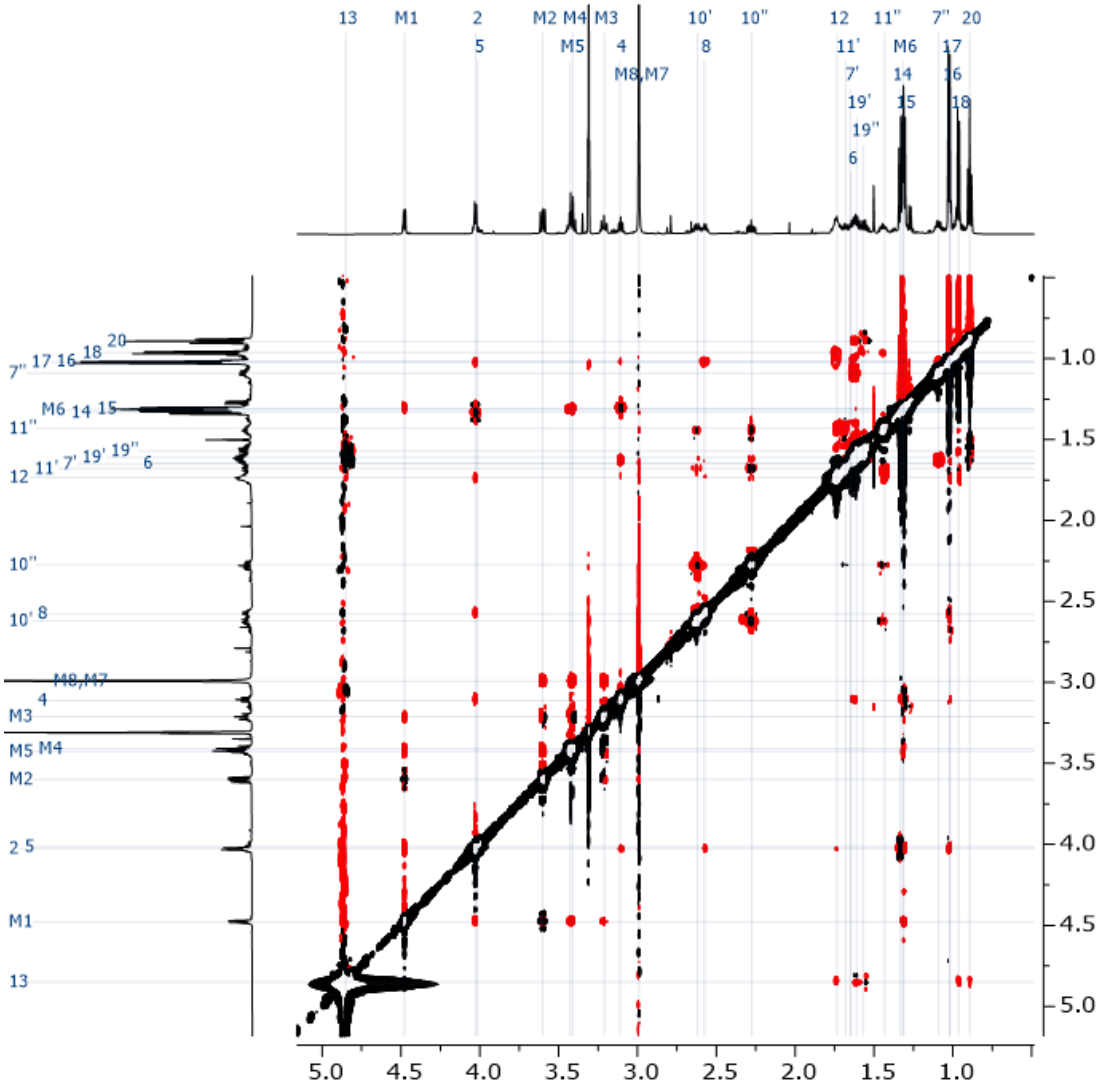
Appendix 14: (^{1}H, ^{13}C)-HSQC edited NMR of kayamycin



Appendix 15: Correlation in 2D HMBC spectra of kayamycin



NOESY

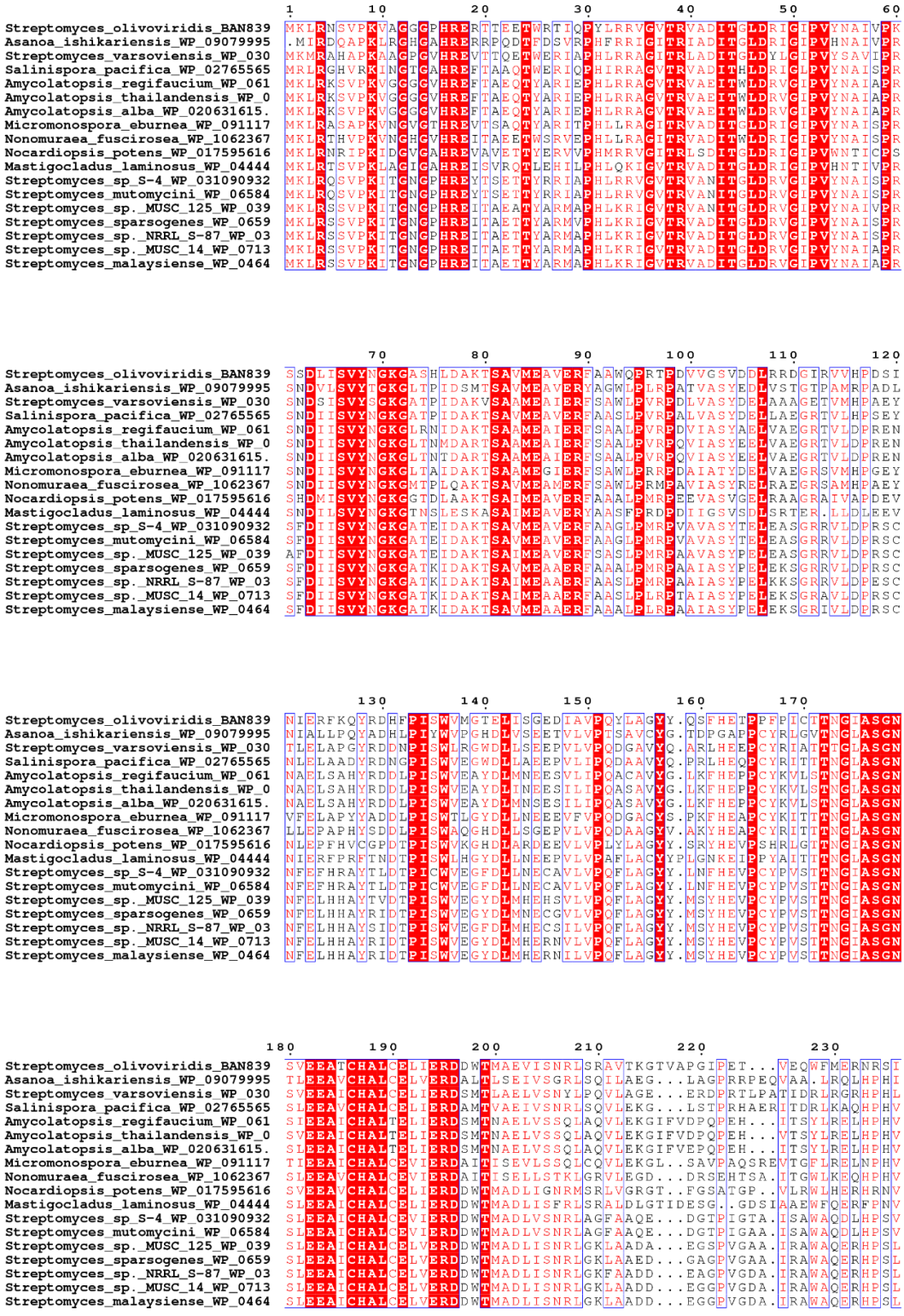


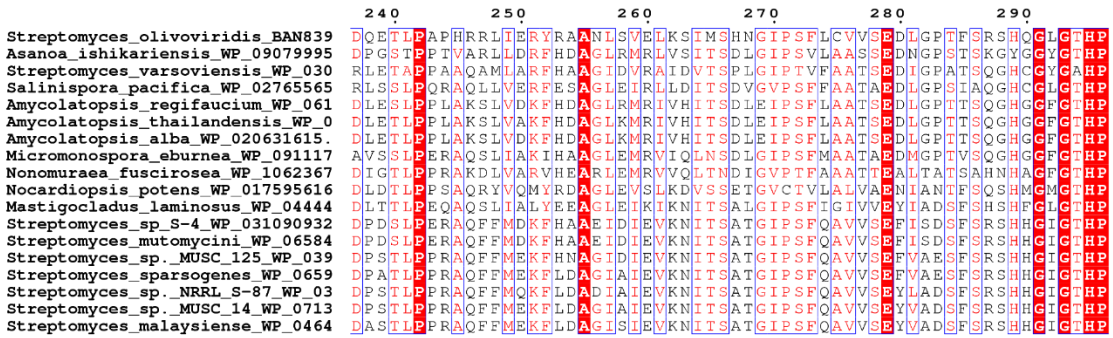
Appendix 16: Correlation in 2D NOESY spectra of kayamycin

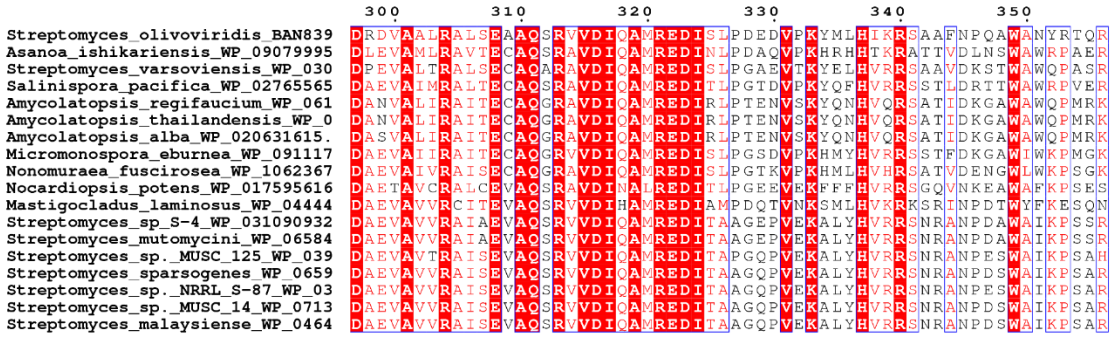
Appendix 17: Tables of NMR J values and interactions

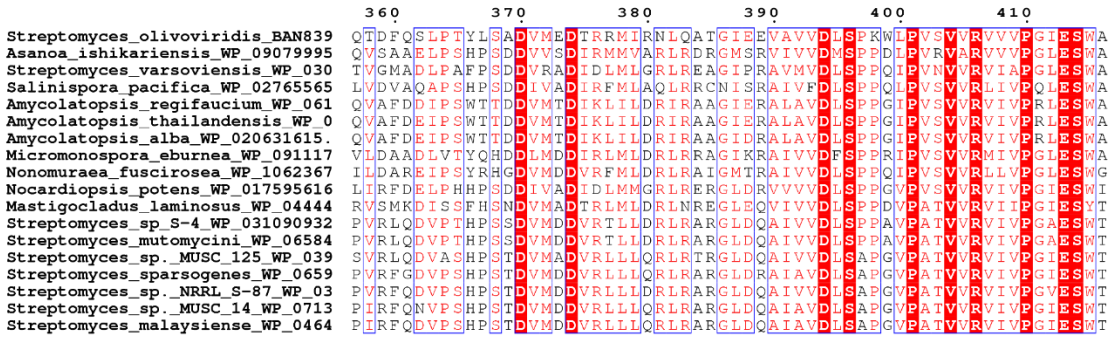
No	δ_H	J (Hz)	δ_C	HMBC	COSY	NOESY
1	-		172.3	2,13, 14	-	-
2	4.03 (m)		52.8	1, 14, 15	14	4, 12
3	-		210.8	4, 14, 15	-	-
4	3.11 (m)		50.1	3,5, 6, 15	5, 15	2,5, 6, 16
5	4.01 (m)		81.0	M1, 7, 15, 16	4	M1, 4, 8, 12, 16
6	1.65 (m)		37.4	-	7'', 7',16	4, 7''
8	2.58 (m)		45.1	9	7'', 7', 17	5, 17
9	-		217.4	7'', 8, 10'', 10', 17	-	-
12	1.73 (m)		38.0	11, 18	11', 18	2,5, 11'', 13, 18
13	4.85 (m)		81.2	1,11, 12, 19, 20	19'', 19'	12, 18, 19', 20
14	1.33 (m)		14.7	1, 2,3	2	-
15	1.31 (m)		14.6	3, 4,5	4	-
16	1.02 (m)		17.6	5, 6, 7	6	4,5
17	1.02 (m)		17.7	7, 8, 9	8	7'', 8
18	0.96 (d)	6.90	15.7	11, 12, 13	12	11'', 12, 13, 19''
20	0.89 (dd)	7.40	11.0	13, 19	19'', 19'	13, 19'', 19'
10'	2.62 (m)		40.4	9, 11, 12	10'', 11'', 11'	10'', 11''
10''	2.27 (ddd)	6.5, 9.2, 15.4	40.4	9, 11, 12	10', 11'', 11'	10', 11''
11'	1.68 (m)		28.5	10, 12	10'', 10', 11'', 12	11''
11''	1.43 (m)		28.5	-	10'', 10', 11'	10'', 10', 11', 12, 18
19'	1.61 (m)		24.1	-	13, 19'', 20	13, 20
19''	1.57 (m)		24.1	13, 20	13, 19', 20	18, 20
7'	1.65 (m)		36.7	-	6, 7'', 8	-
7''	1.09 (m)		36.7	5, 6, 8, 9, 16, 17	6, 7', 8	6, 17
M1	4.48 (d)	6.90	104.6	M5, M4, 5	M2	M5, M2, M6, 5
M2	3.60 (t)	6.9, 7.3	69.8	M4, M1	M3, M1	M4, M3, M1, M8, M7
M3	3.21 (m)		70.6	M1, M8, M7	M4, M2	M5, M4, M2, M8, M7
M4	3.41 (m)		72.2	M2	M3	M3, M2
M5	3.43 (m)		74.1	M4, M3, M1, M6	M6	M3, M1, M6, M8, M7
M6	1.32 (m)		17.7	M5, M3	M5	M5, M1
M7	2.99 (s)		42.2	M3, M8	-	M5, M3, M2
M8	2.99 (s)		42.2	M3, M7	-	M5, M3, M2

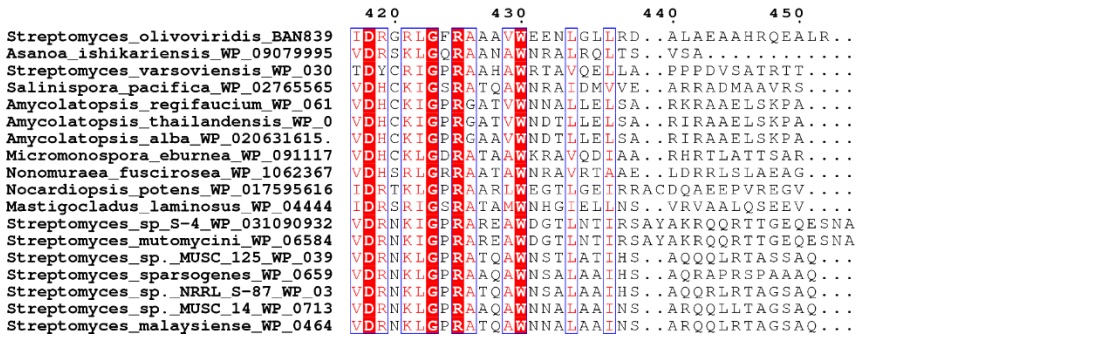
Appendix 18: Multiple alignment of the 18 thioamitide precursor peptides from the identified BGCs. MUSCLE used for alignment, which is visualised using ESPript 3.0 (Robert and Gouet, 2014, Madeira et al., 2022). Identical residues are showed in white with a red background, while similar residues (Risler matrix set with global score of 0.7) are coloured red and are boxed.




240 250 260 270 280 290
Streptomyces_olivoviridis_BAN839
Asanoa_ishikariensis_WP_09079995
Streptomyces_varsoviensis_WP_030
Salinispora_pacifica_WP_02765565
Amycolatopsis_regifaucium_WP_061
Amycolatopsis_thailandensis_WP_0
Amycolatopsis_alba_WP_020631615.
Micromonospora_eburnea_WP_091117
Nonomuraea_fuscirosea_WP_1062367
Nocardiopsis_potens_WP_017595616
Mastigocladus_laminosus_WP_04444
Streptomyces_sp_S-4_WP_031090932
Streptomyces_mutomycini_WP_06584
Streptomyces_sp._MUSC_125_WP_039
Streptomyces_sparsogenes_WP_0659
Streptomyces_sp._NRRL_S-87_WP_03
Streptomyces_sp._MUSC_14_WP_0713
Streptomyces_malaysiense_WP_0464


300 310 320 330 340 350
Streptomyces_olivoviridis_BAN839
Asanoa_ishikariensis_WP_09079995
Streptomyces_varsoviensis_WP_030
Salinispora_pacifica_WP_02765565
Amycolatopsis_regifaucium_WP_061
Amycolatopsis_thailandensis_WP_0
Amycolatopsis_alba_WP_020631615.
Micromonospora_eburnea_WP_091117
Nonomuraea_fuscirosea_WP_1062367
Nocardiopsis_potens_WP_017595616
Mastigocladus_laminosus_WP_04444
Streptomyces_sp_S-4_WP_031090932
Streptomyces_mutomycini_WP_06584
Streptomyces_sp._MUSC_125_WP_039
Streptomyces_sparsogenes_WP_0659
Streptomyces_sp._NRRL_S-87_WP_03
Streptomyces_sp._MUSC_14_WP_0713
Streptomyces_malaysiense_WP_0464


360 370 380 390 400 410
Streptomyces_olivoviridis_BAN839
Asanoa_ishikariensis_WP_09079995
Streptomyces_varsoviensis_WP_030
Salinispora_pacifica_WP_02765565
Amycolatopsis_regifaucium_WP_061
Amycolatopsis_thailandensis_WP_0
Amycolatopsis_alba_WP_020631615.
Micromonospora_eburnea_WP_091117
Nonomuraea_fuscirosea_WP_1062367
Nocardiopsis_potens_WP_017595616
Mastigocladus_laminosus_WP_04444
Streptomyces_sp_S-4_WP_031090932
Streptomyces_mutomycini_WP_06584
Streptomyces_sp._MUSC_125_WP_039
Streptomyces_sparsogenes_WP_0659
Streptomyces_sp._NRRL_S-87_WP_03
Streptomyces_sp._MUSC_14_WP_0713
Streptomyces_malaysiense_WP_0464


420 430 440 450
Streptomyces_olivoviridis_BAN839
Asanoa_ishikariensis_WP_09079995
Streptomyces_varsoviensis_WP_030
Salinispora_pacifica_WP_02765565
Amycolatopsis_regifaucium_WP_061
Amycolatopsis_thailandensis_WP_0
Amycolatopsis_alba_WP_020631615.
Micromonospora_eburnea_WP_091117
Nonomuraea_fuscirosea_WP_1062367
Nocardiopsis_potens_WP_017595616
Mastigocladus_laminosus_WP_04444
Streptomyces_sp_S-4_WP_031090932
Streptomyces_mutomycini_WP_06584
Streptomyces_sp._MUSC_125_WP_039
Streptomyces_sparsogenes_WP_0659
Streptomyces_sp._NRRL_S-87_WP_03
Streptomyces_sp._MUSC_14_WP_0713
Streptomyces_malaysiense_WP_0464