Biosynthetic Lego: Reprogramming RiPP Biosynthesis

Tom Hayon Eyles

John Innes Centre

A thesis submitted September 2018 to the University of East Anglia for the degree of Doctor of Philosophy.

Abstract

Ribosomally synthesised and post translationally modified peptides (RiPPs) are a diverse class of industrially-important and clinically-relevant natural products. Reprogramming the biosynthesis of RiPPs can provide an understanding of their biosynthesis, increases in their yield, and compound derivatives. In this thesis, two RiPP biosynthetic pathways are reprogrammed to achieve these aims.

Bottromycin is a potent antibiotic RiPP, however it is produced in low yields by its native producer and it is rapidly hydrolysed in blood plasma. It was hypothesised that the bottromycin gene cluster could be reprogrammed to increase the production of bottromycin and to derivatise it. Synthetic biology techniques available at the start of the project were deemed inappropriate for use in reprogramming the bottromycin gene cluster as they lacked the ability to conduct refactoring, produce gene insertions/deletions, and make targeted mutations in single steps in the high-GC bottromycin gene cluster. Here, a one-step yeast-based method that enables efficient and flexible modifications to the bottromycin gene cluster is presented. Multiple modifications are showcased, including refactoring, gene deletions and targeted mutations. This facilitated the construction of an inducible, riboswitch-controlled pathway that achieved a 120-fold increase in pathway productivity in a heterologous host. Additionally, an unexpected biosynthetic bottleneck resulted in the production of a suite of new bottromycin-related metabolites.

Thiostreptamide S4 is part of a family of promising antitumor RiPPs, the thioviridamide-like molecules. The gene cluster responsible for thiostreptamide S4 production has been identified, yet the biosynthesis has not been elucidated. It was hypothesised that reprogramming the thiostreptamide S4 gene cluster could provide insights into its biosynthesis. These modified clusters were constructed, and in-depth metabolomics enabled an understanding of the biosynthetic pathway. This biosynthetic understanding could pave the way for future engineering projects and allow key biosynthetic steps to be identified for use in genome mining.

Contents

Figures and Tables

Table of Figures

Table of Tables

Acknowledgements

This PhD has been an amazing experience and there is one person I must thank above everyone else for providing me with this: Andrew Truman. Andy was a supportive supervisor that went above and beyond what was necessary to help me in my PhD. Not only did he provide me with opportunities to grow, but he also tolerated my unending questions about everything from chemistry to careers. Without Andy, this PhD would not have been possible. I am also incredibly grateful for the supervision provided by my secondary supervisors, Barrie Wilkinson and Mervyn Bibb, whose scientific input and career advice were equally invaluable.

Almost everything I have learnt in the lab was taught to me by members of the Truman group, who were always willing to spend time teaching me. I think particularly of Natalia Miguel-Vior, who looked after me when I first arrived and was an excellent teacher. Javier Santos-Aberturas has been a particularly supportive figure; not only teaching me many skills, but also being a great sounding board for ideas. Rodney Lacret was critical in helping me with the chemistry required to complete my projects and Luca Frattaruolo laid the foundations for much of the work I did. Additionally, the rest of the Truman group, including Alicia Russell and David Widdick, have created an atmosphere that has made undertaking this PhD an experience I would happily repeat.

I am particularly privileged to have worked in a department like Molecular Microbiology. The head of the department, Mark Buttner, and the other group leaders have worked hard to foster a collegial and friendly atmosphere that encourages collaboration and sharing of ideas. It will be hard to find somewhere that can match Mol Micro in both the quality of people and the quality of science.

It is important to thank my family, the Doctors Eyles. Mark, Caroline, and Joe have provided more support than I could ever ask for and I am lucky to have them. Finally, I am incredibly grateful to those that have helped keep me sane throughout my PhD: Squadgoalz and Megaladz.

Foreword

Research outputs

The research presented throughout this thesis will lead to published outputs.

Lead Author

Bottromycin

Most of the bottromycin story (Chapter 2) has been published in ACS Synthetic Biology:

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(5), 1211–1218. https://pubs.acs.org/doi/abs/10.1021%2Facssynbio.8b00038

Therefore, any further permissions related to this material should be directed to the ACS. The full paper is also available in the appendices (page 220). It should be noted that in response to reviewers one of the refactored bottromycin gene clusters (pCAPbtm3; section 2.2.3.1) was omitted from the published version of this research. Therefore, the cluster naming system does not match in this thesis beyond pCAPbtm2.

Thiostreptamide S4

The thiostreptamide S4 story (Chapter 3) will be submitted for publication before the end of 2018, pending some minor experimental data. The genome mining arising from this will form the basis for future projects.

Co-Author

Thiovarsolins

Several side projects not mentioned in this thesis will also lead to published outputs. A manuscript is currently submitted to the Journal of the American Chemical Society that describes a new genome

mining technique that has led to the identification of a new family of RiPPs, the thiovarsolins. My role in this was aiding in the TAR cloning of a thiovarsolin gene cluster.

A New RiPP

A research project being carried out by Alicia Russell is currently underway to characterise another new family of RiPPs. My roles in this project were: 1) designing and supervising the assembly and expression of a gene cluster (experimental work carried out by Yan Xunyou), and 2) suggesting an improvement to, and aiding in, TAR cloning of a second gene cluster.

Chapter 1 – Introduction

1.1. Natural Products

1.1.1. A Brief History

"It was found that broth in which the mould had been grown at room temperature for one or two weeks had acquired marked inhibitory, bactericidal and bacteriolytic properties to many of the more common pathogenic bacteria." (Fleming, 1929). This is the sentence that in 1929 started modern natural product research. And whilst at the time it was just a relatively crude extract from mould, *"…to avoid the rather cumbersome phrase "Mould broth filtrate," the name "penicillin" will be used"* (Fleming, 1929), penicillin was then developed to become one of the most famous antibiotics.

However, prior to this there was still interest in natural products. For example, in the nineteenth century Rudolf Emmerich and Oscar Löw wrote *"…pyocyaneus cultures negatively influenced the growth of other microbes e.g. Typhus and Cholera"* [translation] when describing a filtered extract from *Bacillus pyocyaneus* (now called *Pseudomonas aeruginosa*; Emmerich and Löw, 1899). They named the active extract pyocyanase (Emmerich and Löw, 1899). This was in a work named *"Bacteriolytic enzymes as the cause of acquired immunity and through them the healing of infectious diseases"* [translation] and it described the first use in a clinical setting of an antibiotic intentionally extracted from bacteria. The antibiotic was likely a combination of phenazines that included pyocyanin, albeit a variable and often toxic combination; hence its rapid decline from use.

At the same time as this revolution in natural product science was going on, investigations were also happening into ribosomally-synthesized and post-translationally-modified peptides (RiPPs), although they had not been identified as such at the time. In fact, the lanthipeptide nisin was described in 1928 as *"…a soluble and possibly a diffusible substance…"* that inhibited the growth of *Lactobacillus bulgaricus* (Rogers, 1928). Prior to this, a toxin from streptococcus was the source of much research as streptococcal infections were of clinical interest. Understanding the toxin, now thought to be the azoline-containing peptide streptolysin S, was seen as a potentially important part of addressing this

disease. Unfortunately, production of this toxin in the laboratory was low and unreliable and so: *"in order to transform our streptococcus into a microbe capable of giving an abundant supply of toxin, modifications must be made in the nutrient medium."* [translation] (Marmorek, 1902). This is possibly the first published example in contemporary literature of media optimisation towards increasing the yield of a RiPP.

It is, however, Selman Waksman who is known as the father of antibiotics (Kresge, 2004). This is because he developed a technique for selectively enriching samples for organisms that produce antibacterial compounds and then screening them against pathogenic bacteria (Waksman and Woodruff, 1940). The soil turned out to be a particularly rich source of microbes which produce antibiotics, and subsequently other bioactive molecules. This sparked the growth of the field of natural products from then towards the remarkable position researchers are in at the moment, where many estimates put the number of natural products known to be in the low hundreds of thousands, although it should be noted this number varies wildly between sources and it is likely near-impossible to accurately count (Chen et al., 2017).

1.1.2.   Production

1.1.2.1.   Natural Product Producers

Natural products are produced across the entire tree of life, including in higher organisms such as animals and plants. For example, animals and plants produce cyclic peptides crosslinked by disulphide bonds named defensins (Ganz, 2003). These are a vital part of their innate immune system. Plants are also prolific producers of, among other things, terpenes (Singh and Sharma, 2015). These are often produced as defence responses, and are also now developed as flavours, scents, cosmetics, and drugs (Kumari et al., 2014; Schwab et al., 2008). Archaea produce a wide range of natural products, although they lack many of the major classes seen in bacteria and fungi (Charlesworth and Burns, 2015). Fungi and bacteria are responsible for the production of the natural products that were first described in contemporary literature (Emmerich and Löw, 1899; Fleming, 1929; Marmorek, 1902), as well as many of those that have been identified now (Doroghazi et al., 2014). It is generally accepted that many of the natural products produced by bacteria and fungi are being produced as warfare agents that provide selective advantages in the wild (Chater, 2006).

This thesis focusses on natural products from the *Streptomyces* genus of bacteria. These sit within the order Streptomycetales, and the phylum and class are both named Actinobacteria (Barka et al., 2016). *Streptomyces* have a complex lifecycle for bacteria. They live as multicellular mycelial communities primarily in the soil and form environmentally resistant spores for dispersal (Flärdh and Buttner, 2009). The soil environment *Streptomyces* have evolved in contains a rich, complex, and constantly changing microbial community. It has been suggested that these factors are one of the main reasons *Streptomyces* are such prolific producers of natural products (Chater, 2006). Filamentous high-GC Actinobacteria are commonly referred to as the actinomycetes, and it has been predicted that the actinomycetes encode over 17,000 of the major natural product families, with *Streptomyces* spp. containing the most biosynthetic gene clusters (Doroghazi et al., 2014). This is unsurprising, given that the first sequenced genome of a *Streptomyces*, *Streptomyces coelicolor*, revealed that it encodes more than 20 natural product gene clusters (Bentley et al., 2002), and the second sequenced strain, *Streptomyces avermitilis*, encodes at least 25 gene clusters (Ikeda et al., 2003). It is now clear that the incredible natural product capacity is not the exception but the rule for this genus.

1.1.2.2.   Classes of Natural Products

Natural products are grouped in to classes based on their key structural features and/or biosynthetic route. For example, penicillin is a member of the β-lactam antibiotics, named because of their central β-lactam ring (Testero et al., 2010), and of the non-ribosomal peptide natural products, named because their peptide scaffold produced by non-ribosomal peptide synthetases (NRPSs; Felnagle et al., 2008; Walsh, 2016). Other major classes are the polyketides produced by fatty acid synthase-

related polyketide synthases (PKSs; Hertweck, 2009), peptides produced by the ribosome (RiPPs; Arnison et al., 2013), the aminoglycosides (Becker and Cooper, 2013), the terpenes (Oldfield and Lin, 2012), and many more (Figure 1; Katz and Baltz, 2016). As two RiPPs are the focus of this thesis, their features and biosynthesis will be discussed over the next section.
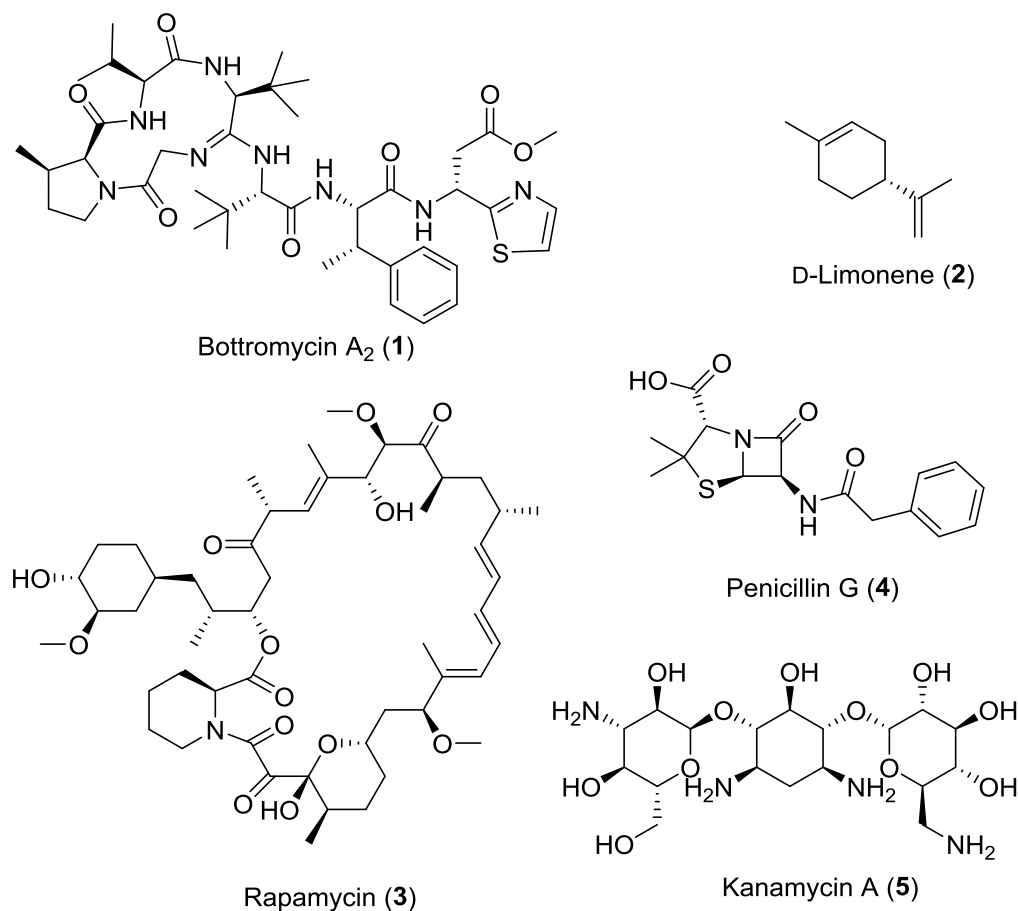
Figure 1. Representative examples of RIPP (bottromycin A$_2$; **1**; Shimamura et al., 2009), terpene (D-limonene; **2**; Hall, 1933), PKS (rapamycin; **3**; Kessler et al., 1993), NRPS (penicillin G; **4**; Bentley, 2004), and aminoglycoside (Kanamycin A; **5**; Cron et al., 1958) natural products.

1.2. RiPPs

1.2.1.   RiPP Classification

RiPPs are unique amongst natural products, as they use the ribosome to produce their scaffold. A general size limit of 10 kDa separates them from modified proteins. They typically consist of a single short precursor peptide that can be separated into two parts: the core peptide and the leader/follower peptide. The core peptide is modified by a series of tailoring enzymes that install complex modifications. It is generally accepted that the leader/follower peptide is used to recruit the tailoring enzymes (Burkhart et al., 2015; Oman and van der Donk, 2010). Finally, the modified core peptide is excised from the leader/follower peptide as a mature or nearly mature product (Figure 2; Arnison et al., 2013).

RiPPs are often classified based on characteristic modifications within their structure. These can be key structural features, such a lanthionine bonds in lanthipeptides (Figure 3; Repka et al., 2017). Alternatively these can be rare combinations of structural features, such as an S-[(Z)-2-aminovinyl](-(3S)-3-methyl)-D-cysteine (Avi(Me)Cys) macrocycle, an N-terminal pyruvyl-like moiety, and
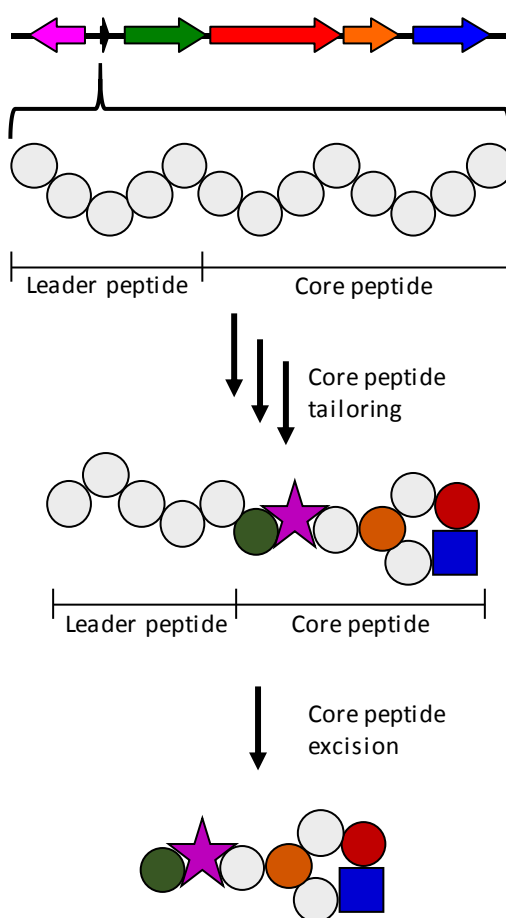
Figure 2. Generalised schematic of RiPP biosynthesis. A ribosomally synthesized core peptide is modified by tailoring enzymes and then excised from the precursor peptide to produce a mature product.
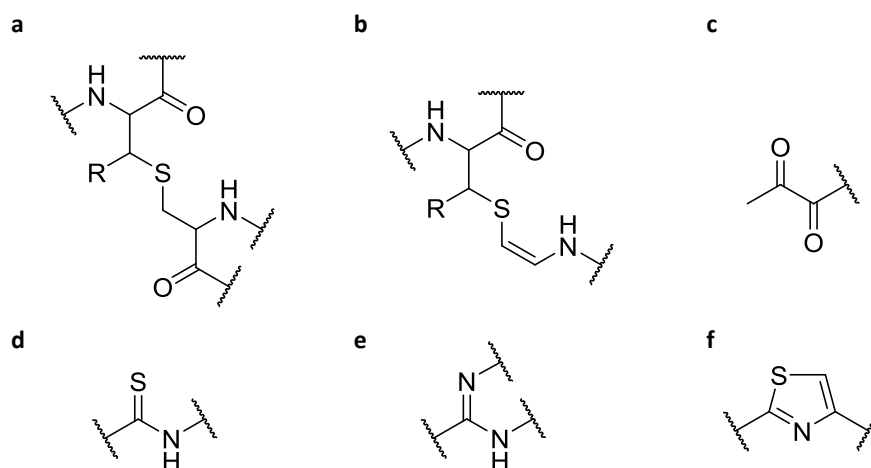
Figure 3. Examples of structural modifications that can be used for RiPP classification. **a.** A (methyl)lanthionine bond, R = H or CH₃. **b.** An Avi(Me)Cys residue, R = H or CH₃. **c.** A pyruvyl moiety. **d.** A thioamide bond. **e.** An amidine residue. **f.** A thiazole ring.

thioamide bonds in the thioviridamide-like molecules (TLMs; Figure 3); none of which are individually unique to the TLMs, but the combination is (Frattaruolo et al., 2017).

Whilst it is useful to be able to split RiPPs into classes with common features, it is imperfect as it can lead to blurred lines between classes. For example, should klebsazolicin (Metelev et al., 2017), a member of the linear azol(in)e-containing peptides (LAP), stay with its current classification? Or, due to its unusual amidine macrocycle, should it become part of a new class of "amidine and azol(in)e containing RiPPs" with the amidine and azoline containing bottromycin-like molecules (Figure 3; Crone et al., 2012)? Ultimately, these classifications do not need to be perfect; they just need to be functional enough to facilitate scientific discourse. Over the next few sections a brief description of some of the main classes of RiPPs will be provided, and the biosynthesis of key features will be discussed in a later section, given that various structural features are common across classes.

1.2.1.1.   Lanthipeptides

Lanthipeptides are RiPPs that contain at least one (methyl)lanthionine bond within their structure (Figure 4). They are a very well characterised and very widespread class of natural products. The (methyl)lanthionine bond is a thioether bond that connects the β-carbon of two residues to form a macrocycle within the RiPP. Lanthionine bonds are structural features within RiPPs that cause them to adopt a configuration that allows them to interact with their target. Lanthipeptides can also contain a diverse array of other modifications, for example disulphide bonds (Xiao et al., 2004), glycosylation (Iorio et al., 2014), N-terminal acylation (Kellner et al., 1989; Mohr et al., 2015; Ortega
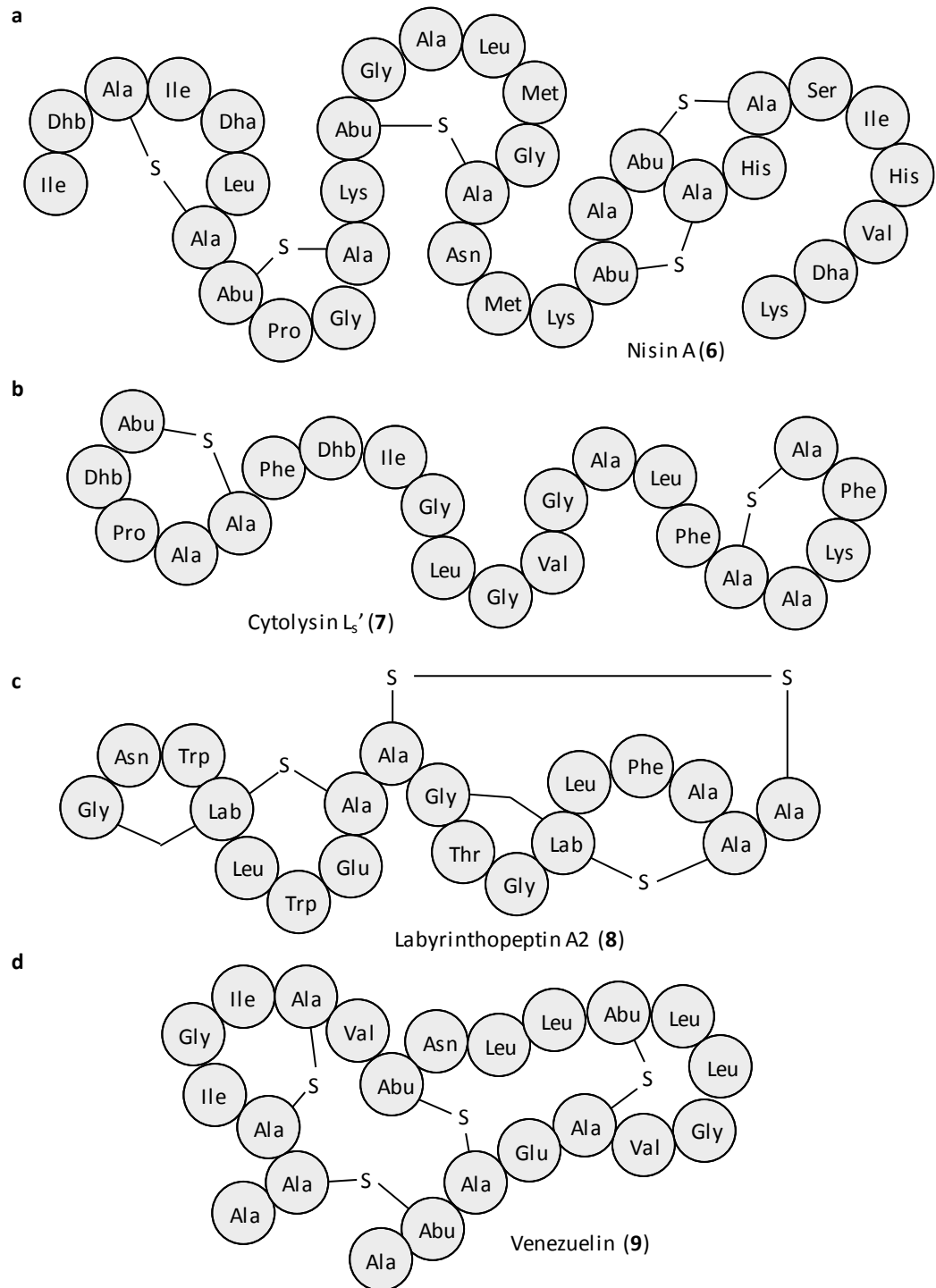
Figure 4. Structures of characteristic lanthipeptides. **a.** Nisin A (**6**), a class I lanthipeptide (Gross and Morell, 1971). **b.** Cytolysin L$_s$' (**7**), a class II lanthipeptide (Van Tyne et al., 2013). **c.** Labyrinthopeptin A2 (**8**), a class III lanthipeptide (Meindl et al., 2010). **d.** Venezuelin (**9**), a class IV lanthipeptide (Zhang et al., 2015).

et al., 2014), dehydration (Gross et al., 1973), halogenation (Castiglione et al., 2008), and many more.

Lanthipeptides are also associated with a diverse array of activities. For example, nisin (**6**) and epidermin are antibacterial, and bind the cell wall precursor lipid II and form pores in bacterial membranes, thereby killing the cells (Breukink et al., 1999; Brötz et al., 1998; Hasper et al., 2004). Two-component lanthipeptides also work in a similar way, for example lacticin 3147: two lanthipeptides are produced by the same cluster, one binds to lipid II and the other is then recruited to form pores in the membrane (Wiedemann et al., 2006). Aside from bactericidal activities, lanthipeptides have been identified that exhibit antiviral (Férir et al., 2013), antifungal (Mohr et al., 2015), and possibly antitumor (Broughton et al., 2016) activities. In *S. coelicolor* a lanthipeptide (SapB) has also been shown to be an important morphogen required for the formation of aerial hyphae (Kodani et al., 2004).

1.2.1.2.   LAPs

Linear azol(in)e containing peptides (LAPs) are characterised by the extensively studied microcin B17 (Li et al., 1996). Microcin B17 is a 43-residue antibacterial peptide containing eight thiazole or oxazole heterocycles (Vizan et al., 1991). Other than the formation of the heterocycles and the cleavage of the leader peptide, it is otherwise unmodified. Other members of the LAPs include the cytotoxic streptolysin S (Molloy et al., 2011), antibacterial plantazolicin (Kalyon et al., 2011), and the morphogenic and antibacterial goadsporin (Igarashi et al., 2001; Onaka et al., 2001).

The activity of microcin B17 has been extensively studied, and it has been shown to act by interacting with a gyrase and DNA complex, stabilizing its cleavage complex, and preventing the re-ligation of DNA (Heddle et al., 2001). Modelling the fold of microcin B17 led to the suggestion that some key aspartate residues interact with the gyrase protein, whilst the heterocycles may bind DNA by base-stacking interactions (Parks et al., 2007). Interestingly, fragments and truncated versions of microcin B17 have been shown to also stabilise the gyrase complex, however if the N-terminus is missing then transport of microcin B17 into the target cells is impaired (Collin et al., 2013; Thompson et al., 2014).

1.2.1.3.   Cyanobactins

Cyanobactins are small head to tail cyclised peptides, similar to LAPs due to their multiple azol(in)e heterocycles (Czekster et al., 2016). They get their name from the phylum of organisms that produce most of them, the cyanobacteria, although similar RiPPs are produced by other non-cyanobacterial microbes. Aside from azol(in)e rings they also contain a variety of other modifications, for example prenylation in trunkamide (Salvatella et al., 2003), disulphide bond in ulithiacyclamide (Ireland and Scheuer, 1980), histidine methylation in microcyclamide (Ishida et al., 2000), and tricyclic prenylated tryptophan in kawaguchipeptin (Ishida et al., 1997; Parajuli et al., 2016).

The cyclic structure of the cyanobactins leads to them being particularly stable and resistant to proteolytic degradation, whilst retaining enough flexibility to maximise binding interactions with targets (Driggers et al., 2008). As such, there is interest in cyanobactins being developed as drugs, especially as many have specific antitumor activities (Lopez et al., 2016; Schmitz et al., 1989). Patellamide and dendroamide are particularly interesting cases; they re-sensitise multidrug-resistant cancer cell lines to anticancer drugs (Ogino et al., 1996; Williams and Jacobs, 1993). This activity prompted the design of some synthetic versions of the cyanobactins (Tao et al., 2011), which have been co-crystallised as inhibitors of P-glycoprotein (Aller et al., 2009). P-glycoprotein is a transporter responsible for exporting a large variety of structurally distinct drugs from cancer cell lines (Sharom, 2008), making drugs that inhibit P-glycoprotein particularly valuable.

1.2.1.4.   Thiopeptides

Thiopeptides are defined by their central macrocycle-forming nitrogen-containing six membered pyridine or piperidine-like ring (Figure 5; Bagley et al., 2005). They are also extensively dehydrated and heterocyclised, and therefore can often contain very few unmodified amino acids. For example the core peptide of geninthiocin has 15 amino acids, whilst the final molecule only contains a single unmodified residue (Sajid et al., 2008). The large thiopeptide diversity can be classified into seven classes (Schwalen et al., 2018): the thiostrepton (**13**)-like (Anderson et al., 1970), thiocillin-like (Aulakh and Ciufolini, 2011), nosiheptide (**11**)-like (Prange et al., 1977), thiomuracin-like (Morris et al., 2009), cyclothiazomycin-like (Aoki et al., 1991), berninamycin-like (Liesch and Rinehart, 1977), and lactazole-like (Hayashi et al., 2014). These are categorised based on structural and chemical features, such as an second indole or quinaldic acid macrocycle (Wei Ding et al., 2017a; Duan et al., 2012; Yu et al., 2009).

The majority of thiopeptides are anti-Gram-positive compounds that target the ribosome (Bagley et al., 2005). Thiostrepton (**13**), nosiheptide (**11**), and micrococcin (**10**) have all been crystallised binding to the bacterial ribosome (Harms et al., 2008). They bind to both ribosomal RNA as well as L11, a ribosomal protein. Thiostrepton (**13**) and nosiheptide (**11**) lock L11 in a conformation that prevents elongation factor G from interacting correctly with the ribosome. Micrococcin (**10**) works in a different manner; it allows elongation factor G binding, yet it also stabilises interactions between L11 and L7. The L7 is then able to constitutively stimulate the GTP hydrolysis activity of elongation factor G (Cameron et al., 2002; Harms et al., 2008).

1.2.1.5.   Bottromycins

Bottromycin A$_2$ (**1**; hereafter called bottromycin) is an extensively modified RiPP, in which only two of the eight constituent residues are unmodified. It contains a rare amidine macrocycle, a unique terminal thiazole, an epimerised aspartate residue, and extensive variable methylation (Figure 6;
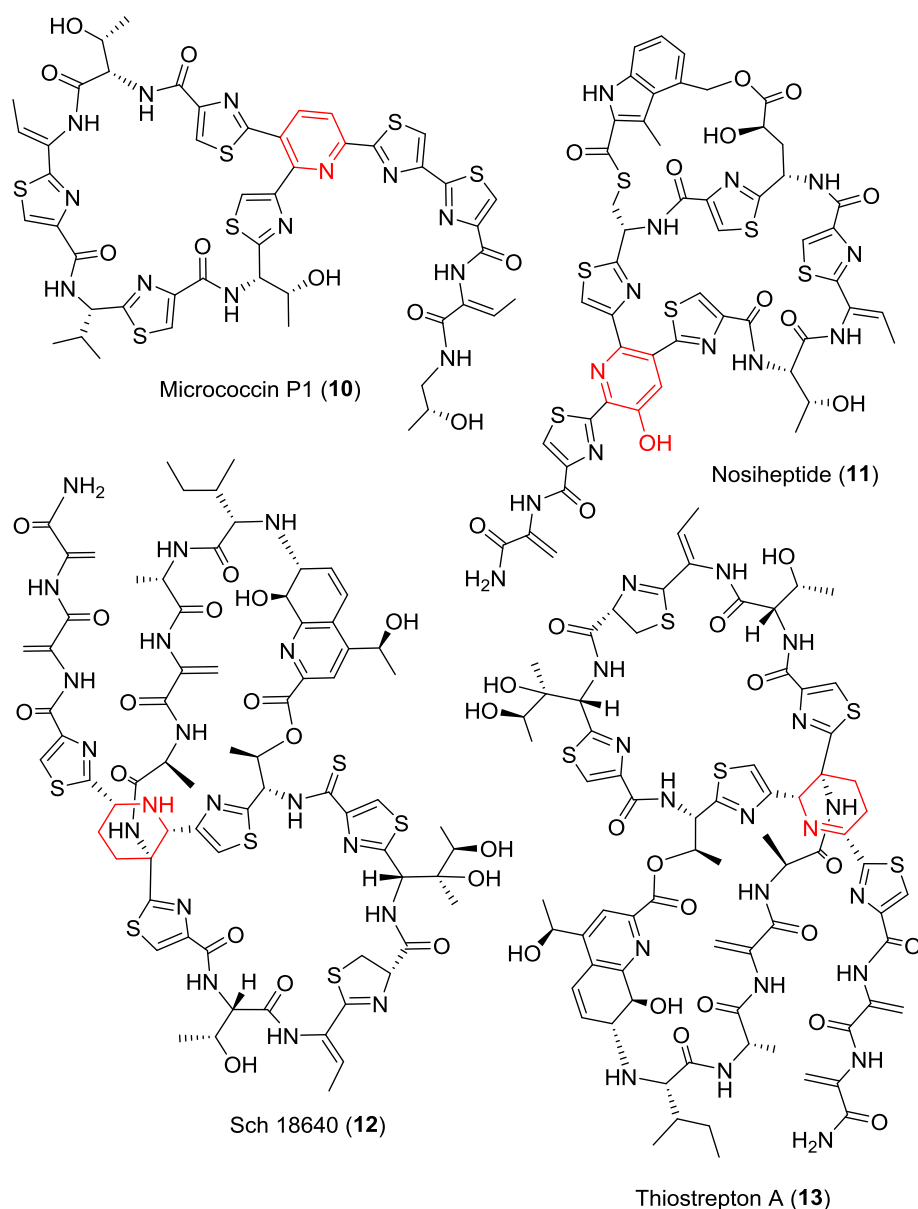
Figure 5. Examples of thiopeptide RiPPs: the pyridine-containing micrococcin P1 (**10**; Hall et al., 1966), the hydroxypyridine-containing nosiheptide (**11**; Prange et al., 1977), the piperidine-containing Sch 18640 (**12**; Puar et al., 1981), and the dehydropiperidine-containing thiostrepton A (**13**; Anderson et al., 1970). The characteristic nitrogen-containing six-membered ring is highlighted in red.
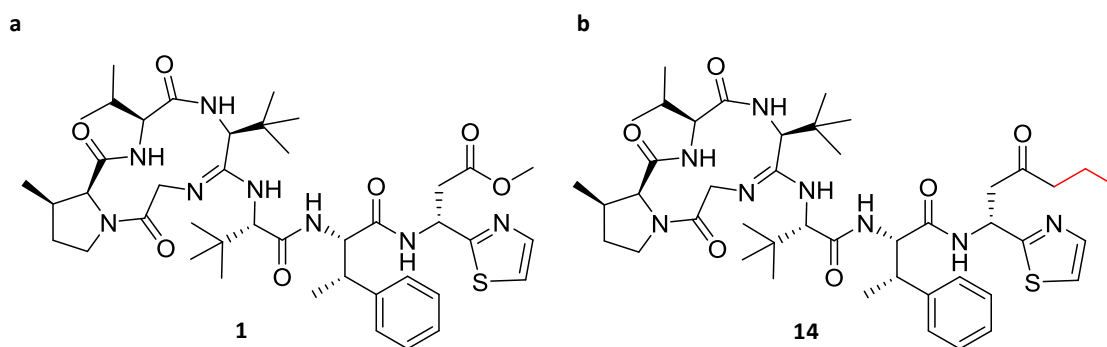
Figure 6. Structure of **a.** bottromycin (**1**) and **b.** a selected semi synthetic derivative (**14**). The modified region is highlighted in red (Kobayashi et al., 2010; Shimamura et al., 2009).

Kaneda, 1992; Shimamura et al., 2009). Bottromycin (**1**) was first extracted from *Streptomyces bottropensis* and has been shown to have potent anti-Gram-positive activity (Shimamura et al., 2009; Waisvisz et al., 1957). Tests of efficacy in treating infected mice failed due to degradation of the compound (Miller et al., 1968), and it has since been shown that this is due to hydrolysis of the aspartate methyl ester in blood plasma (Kobayashi et al., 2010). Since then, many synthetic and semi-synthetic derivatives have been made which retain activity, but lack the methyl ester (Kobayashi et al., 2010; Yamada et al., 2018). For example, if the methyl ester is replaced with a propyl group (**14**; Figure 6), or the aspartate side chain is removed entirely then activity is not lost, yet the methyl ester is no longer available for hydrolysis. These modified versions have yet to be tested *in vivo*.

Whilst bottromycin (**1**) has not yet been crystallised interacting with its target, its activity has been assessed in the past. Bottromycin (**1**) is a translation inhibitor that causes the dissociation of amino acyl (AA)-tRNA from the A-site of the ribosome and prevents further binding of AA-tRNAs to this site (Otaka and Kaji, 1976). This activity is very specific, as the ribosome is still able to translocate and catalyse peptide bond formation in the presence of bottromycin (**1**; Otaka and Kaji, 1983). The importance of the methyl ester for activity of the wild type molecule suggests that this is involved in binding to the ribosome (Kobayashi et al., 2010). Additionally, a 3D structure of bottromycin (**1**) shows that it folds back on itself, and that both the methylated proline and the methylated aspartate are presented on the same surface of the molecule (Gouda et al., 2012), therefore this may also be important in the activity of the molecule. Early tests with bottromycins lacking the proline methylation showed a corresponding four-fold drop in activity (Miller et al., 1968; Yamada et al., 1978).

1.2.1.6.   Thioviridamide-Like Molecules

Thioviridamide (**15**) is a RiPP with a unique structure. It contains an N-terminal 2-hydroxy-2-methyl-4-oxopentanoyl group, a series of five thioamide bonds, an *S*-[(*Z*)-2-aminovinyl]-D-cysteine (AviCys) macrocycle, and a β-hydroxy-N^1,N^3-dimethylhistidinium (Figure 7; Hayakawa et al., 2006b). Heterologous expression of the thioviridamide (**15**) gene cluster in *S. avermitilis* results in the production of a very similar molecule, JBIR-140 (**16**; Figure 7), that only differs based on its N-terminal structure (Izumikawa et al., 2015). Since the discovery of these molecules, this class of RiPPs has been expanded to include five other members (Figure 7), prethioviridamide, thioholgamide (also known as neothioviridamide), thiostreptamide S4 (**17**), thiostreptamide S87, and thioalbamide (**18**; Frattaruolo et al., 2017; Izawa et al., 2017; Kawahara et al., 2018; Kjaerulff et al., 2017). There are also many similar but uncharacterised pathways, likely to also produce thioviridamide-like molecules (Frattaruolo et al., 2017).

Thioviridamide (**15**) and JBIR-140 (**16**) were shown to induce apoptosis selectively in cancer models (Hayakawa et al., 2006a; Izumikawa et al., 2015), and since then both thioalbamide (**18**) and thioholgamide have been shown to also have potent anticancer activities (Frattaruolo et al., 2017; Kjaerulff et al., 2017). Thioalbamide (**18**) was particularly exciting as it was shown to not only
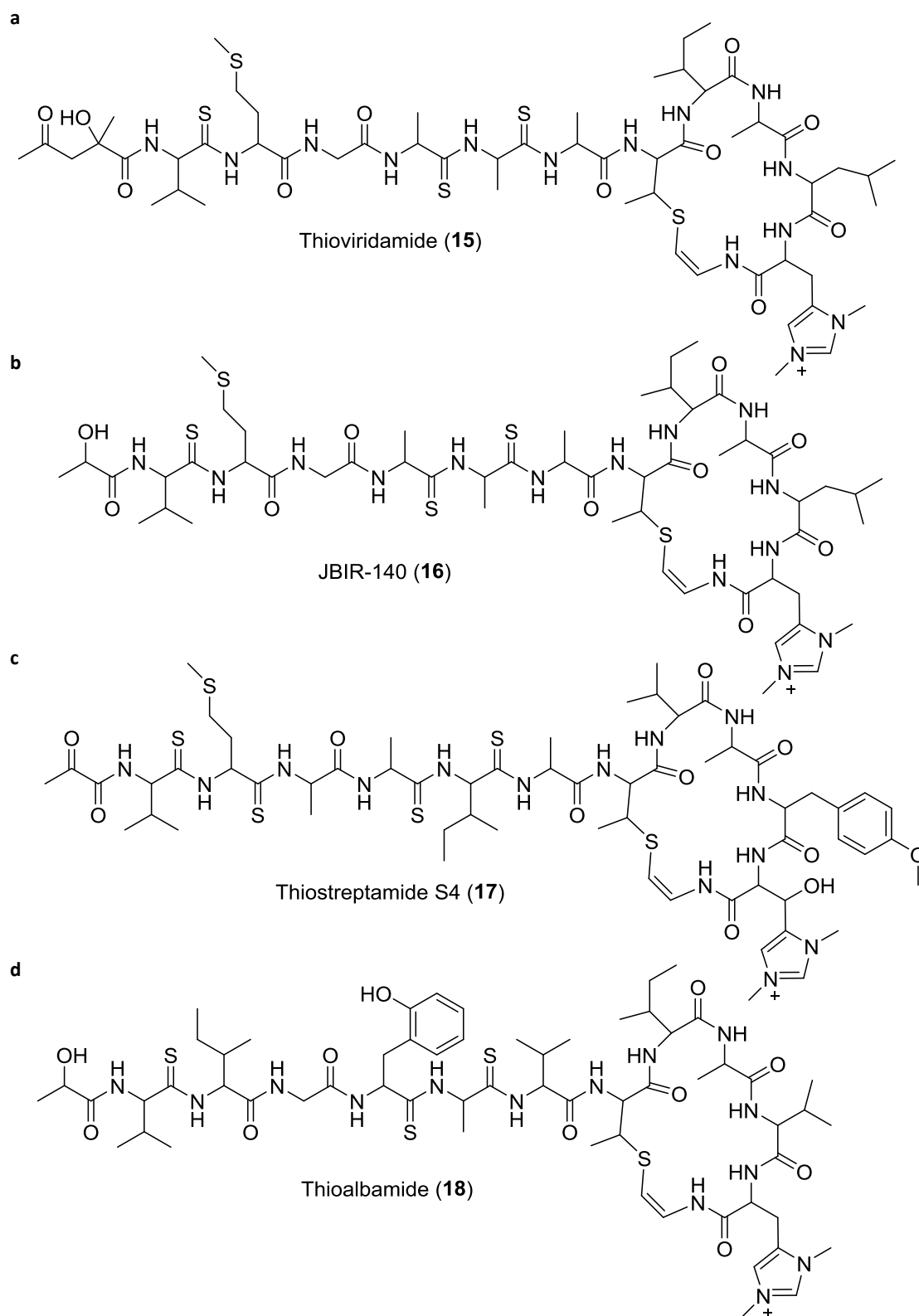
Figure 7. Structures of selected thioviridamide-like RiPPs: **a.** thioviridamide (**15**), **b.** JBIR-140 (**16**), **c.** thiostreptamide S4 (**17**), and **d.** thioalbamide (**18**).

outperform the clinically used doxorubicin when tested against a variety of cancer cell lines, but also to have around 5-fold lower activity against a healthy cell line (Frattaruolo et al., 2017). The mechanism of action of these molecules is not known, yet it will undoubtedly be investigated due to their promise as chemotherapy drugs.

1.2.2. The Biosynthesis of Common RiPP Modifications

1.2.2.1. The Importance of Understanding Biosynthesis

The peptidic scaffold for RiPPs is produced by the ribosome, and therefore is produced in the same way as other peptides and from the same twenty amino acid building blocks. The amazing diversity and properties of RiPPs therefore comes from their small size and their posttranslational modifications (Arnison et al., 2013), which play key roles in the physical structure and the biological activities of RiPPs. The installation of these modifications is often the determining factor as to whether changes to the gene cluster will be tolerated (Deane et al., 2016; Hao et al., 2016). Therefore, to reprogram a RiPP biosynthetic gene cluster, as is the topic of this thesis, it is also vital to understand the biosynthesis of the posttranslational modifications. This section will describe the biosynthesis of some of the most commonly seen posttranslational modifications in RiPP biosynthesis, including all those seen in the subjects of this thesis, bottromycin (**1**) and thiostreptamide S4 (**17**).

1.2.2.2. Lanthionine Bonds

Due to the large diversity of lanthipeptide structures, the lanthipeptides are grouped into one of four classes based on the enzymes that are involved in the formation of the lanthionine bonds. Nisin (**6**; Figure 4) was the founding member of the lanthipeptides (Gross and Morell, 1971; Rogers, 1928), and a mechanism was proposed for the origin of the lanthionine bonds based on the addition of Cys onto dehydroalanine (Dha, resulting from the dehydration of serine) and dehydrobutyrine (Dhb, resulting from the dehydration of threonine; Ingram, 1970). Therefore, the classification is based on the enzymes involved in dehydration and cyclisation.

Nisin (**6**) is a class I lanthipeptide (Rogers, 1928). Of the eleven genes that make up the gene cluster NisA is the precursor peptide and LanB (NisB) and LanC (NisC) are involved in lanthionine bond biosynthesis (Lubelski et al., 2008). LanB proteins, characteristic of class I lanthipeptides, are responsible for the dehydration of serines and threonines in lanthipeptide biosynthesis via glutamation and elimination. The structure of NisB shows a homodimeric enzyme that contains a ~800-residue glutamylation domain and a ~300-residue elimination domain (Ortega et al., 2015). This purified enzyme could dehydrate NisA if it was incubated with ATP, glutamyl-tRNA synthetase, and tRNA[Glu], showing that tRNA[Glu] is the glutamate donor. The FNLD sequence motif in NisA is vital for mediating hydrophobic interactions between NisB and the precursor peptide, a similar motif is found in many class I lanthipeptides, therefore this mechanism of interaction is likely universal to LanB enzymes. The dehydration of the nisin core peptide proceeds in a step-wise manner from the N-terminus to the C-terminus of the molecule (Lubelski et al., 2009). This agrees with a channelling model where NisB, NisC, and NisT (the LanT transporter) localise together at the membrane and
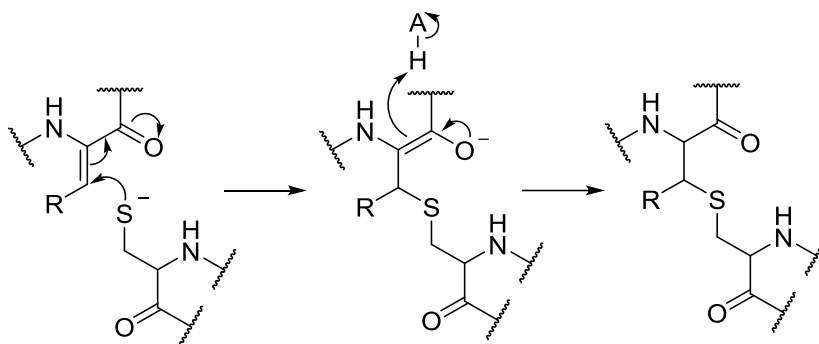
Figure 8. Schematic of lanthionine (R = H) and methyllanthionine (R = CH₃) bond formation from Dha (R = H) or Dhb (R = CH₃). See Figure 9 for a schematic showing the formation of Dha or Dhb.

thread the molecule leader peptide-first via NisB for dehydration, NisC for cyclisation, and NisT for transport, in that order (Lubelski et al., 2009).

In class I lanthipeptide clusters there is a dedicated cyclase responsible for the formation of the lanthionine macrocycles. The crystal structure of NisC shows that it has an α-toroidal core structure with a single zinc ion coordinated by two cysteines and a histidine (Li et al., 2006). Zinc ions coordinated in this way can stabilise cysteine thiolates (Wilker and Lippard, 1995), facilitating the nucleophilic conjugate addition to the β-carbon of Dha or Dhb (Figure 8). Whilst the function of the zinc ion is to stabilise the cysteine thiolate, in basic conditions this is unnecessary and lanthionine rings can occur spontaneously (Okeley et al., 2000). However, without enzymatic control of the regioselectivity of the reaction there is much higher propensity for the lanthionine bond to form with Dha residues over Dhb residues (Zhu et al., 2003); the cyclase is therefore also required to overcome this natural bias.

Class II lanthipeptides use a dedicated LanM enzyme for cyclisation. One of the characteristic members of this family is cytolysin (**7**) produced by *Enterococcus faecalis* (Van Tyne et al., 2013). The crystal structure of CylM (the cytolysin LanM) shows that it contains an N-terminal dehydration domain and a C-terminal cyclase domain (Dong et al., 2015). Whilst the two LanM domains are linked on a single peptide, they are independently catalytic and can therefore act as truncated proteins (Ma et al., 2015; Shimafuji et al., 2015). The cyclase domain of LanM-like enzymes is structurally similar to the LanC enzymes seen in class I lanthipeptide synthesis, and also contains a zinc atom coordinated by two cysteines and a histidine (Dong et al., 2015). The mechanism of cyclisation, with the zinc atom stabilising the thiolate, is therefore likely the same. Investigations into ProcM (a *Prochlorococcus* MIT9313 LanM enzyme) provided additional evidence that the cyclase is also vital for steric control of the regioselectivity of the lanthionine bond formation (Yu et al., 2015).

The LanM dehydration domain is clearly different from the LanB enzymes used for glutamate dependant dehydration in class I clusters. The dehydration domain shows structural similarity to
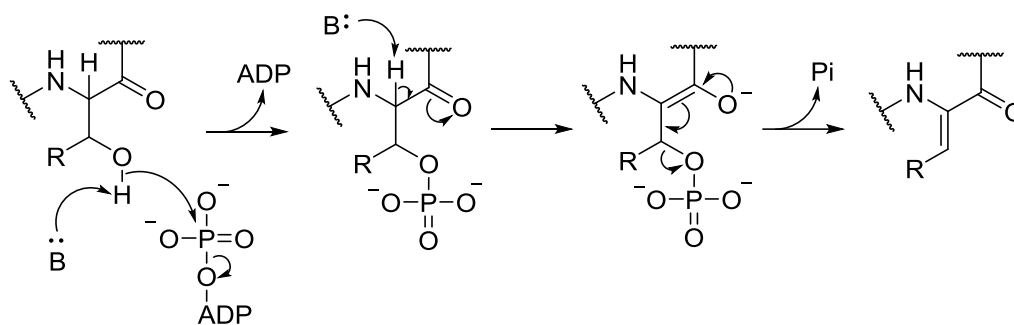
Figure 9. Schematic of phosphorylation and elimination-mediated dehydration of a serine (R = H) or a threonine (R = CH₃).

eukaryotic lipid kinases and lipid kinase-like protein kinases (Dong et al., 2015; Williams et al., 2009). This means the dehydration is dependent on phosphorylation and elimination (Figure 9), which is supported by the ATP dependency of *in vitro* reconstitutions of the dehydration and the observation of phosphorylated intermediates (Chatterjee et al., 2005; Ma et al., 2015; Shimafuji et al., 2015; You and van der Donk, 2007). The residues required for phosphate elimination are situated near the residues for phosphorylation, with one lysine being suggested to continually engage the phosphate throughout phosphorylation and elimination, meaning that in native conditions the entire dehydration reaction happens in the same active site without release of a phosphorylated intermediate (Dong et al., 2015; Ma et al., 2014). The elimination requires ADP to first bind in the active site and it then either helps position the core peptide for elimination or acts as a catalyst to lower the activation energy for elimination (Chatterjee et al., 2005). Therefore, a successful dehydration event requires first ADP binding in the active site, followed by core peptide binding, subsequent ATP-dependant phosphorylation, and finally elimination dependant on the previously bound ADP.

The LctM-catalysed dehydrations proceed in an N- to C-terminal order in the maturation of lacticin 481, an observation that is dependent on the leader peptide still being linked to the core peptide (Thibodeaux et al., 2016). This suggests that the mechanism for ordering the dehydrations is due to the LctM-bound leader peptide holding the core peptide in way that means the more N-terminal serine and threonine residues have favourable spatial positioning towards the dehydration domain, as compared to the C-terminal residues. The timing of cyclisation with respect to this is unclear, although there is evidence that suggests it also proceeds in a largely N- to C- terminal order concurrent with core peptide dehydration (Lee et al., 2009). It is also yet to be seen if this is a universal mechanism for dehydration ordering, or whether in other systems other factors play a part, for example varying dehydration active site affinities for different parts of the core peptide (Thibodeaux et al., 2016).

LanM enzymes are often specific to their precursor peptides, for example in two-component lanthipeptide clusters there is often a LanM enzyme for each precursor peptide (McClerren et al., 2006; Ross et al., 2000; Zhao and van der Donk, 2016). This suggests co-evolution of the LanM enzymes and their precursor peptide substrates. However, there are also very promiscuous LanM enzymes. ProcM is responsible for cyclising a highly diverse array of 29 precursor peptides, and other closely related LanM enzymes (named the CCG group after the CCG motif that replaces the normal CHG motif in other LanM and LanC enzymes) are also associated with many precursor peptides (Li et al., 2010). The zinc atom within these enzymes is coordinated by three cysteines, rather than the two cysteines and a histidine in other LanM and LanC enzymes (Tang and van der Donk, 2012). A thiolate stabilised by a zinc atom coordinated by three cysteines will be more reactive than if the zinc atom is coordinated by two cysteines and a histidine (Chiou et al., 2003); this increased reactivity may account for the promiscuity of CCG group cyclases.

Class III and class IV lanthipeptides use LanKC and LanL proteins for lanthionine formation respectively (Zhang et al., 2012b). The predicted structure of LanKC and LanL proteins is very similar, with both being made up of three independent domains: an N-terminal lyase, a central Ser/Thr kinase, and a C-terminal cyclase. The C-terminal cyclase is homologous to the LanC proteins and the C-terminus of LanM proteins, and yet is also the reason for the distinction between LanKC and LanL enzymes. Whilst LanL proteins contain normal-looking cyclase domains, LanKC proteins lack the zinc-binding residues and have been shown not to bind other metals (Wang and van der Donk, 2012). If this C-terminal domain is deleted from AciKC (the *Catenulispora acidiphila* DSM 44928 LanKC responsible for catenulipeptin maturation) then dehydration occurs normally but cyclisation is severely impaired (Wang and van der Donk, 2012). This indicates that even in the absence of a zinc ligand, this domain is still responsible for proper cycle formation.

Neither LanKC or LanL proteins have been crystallised, yet there are separate crystal structures for orthologues of the lyase (Chen et al., 2008), the Ser/Thr kinase (Young et al., 2003), and the cyclase (Li et al., 2006) domains, allowing a prediction of their activities. Additionally, phosphorylated precursor peptide intermediates can be seen in the production of the class III lanthipeptide, labyrinthopeptin (**8**), showing that the dehydration follows a mechanism of phosphorylation and elimination as in class II lanthipeptides (Müller et al., 2010), albeit with a Ser/Thr kinase instead of the lipid-kinase like domain of LanM enzymes. The LanKC and LanL lyase domains show homology with the OspF protein family (Goto et al., 2010); type III effector proteins from Gram negative bacteria responsible for irreversible phosphate removal from phosphothreonine in plant MAPK proteins (Li et al., 2007). This results in dehydration of the threonine and therefore suggests the role for the OspF-like domains in LanKC and LanL enzymes: elimination and dehydration of serines and threonines. Therefore, whilst LanM enzymes contain bifunctional phosphorylation and elimination
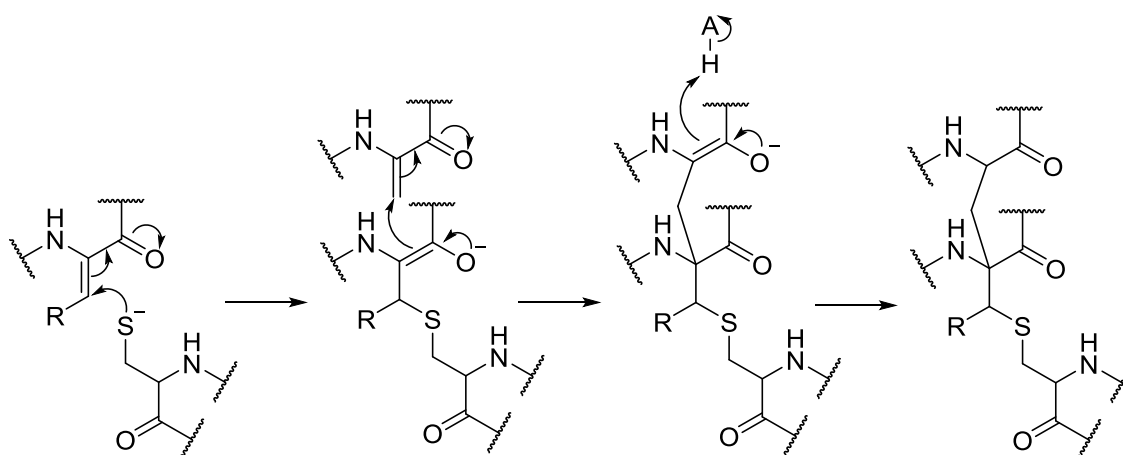
Figure 10. Schematic of labionin (R = H) and methyllabionin (R = CH₃) bond formation.

domains (Dong et al., 2015; H. Ma et al., 2014), LanKC and LanL perform the phosphorylation and elimination with separate domains.

An interesting mechanistic difference between class III (LanKC) and class IV (LanL) cyclases, is the direction in which they process the precursor peptide. SgbL (the globisporin LanL enzyme from *Streptomyces globisporus* subsp. *globisporus* NRRL B2293) was shown to bind the leader peptide with the kinase domain and to process the core peptide in an N- to C-terminal manner, as seen in the maturation of class I and class II lanthipeptides (Hegemann and van der Donk, 2018). In contrast, studies with the LanKC enzymes from the curvopeptin (CurKC from *Thermomonospora curvata*) and the labyrinthopeptin A2 (**8**; LabKC from *Actinomadura namibiensis*) gene clusters showed that they processed the precursor peptide in a broadly C- to N-terminal manner (Jungmann et al., 2014; Krawczyk et al., 2012a). The reasons class III lanthipeptides are processed in a different order from all other characterised lanthipeptide are not known, yet this distinction suggests a significant difference in the ways LanKC enzymes interact with their core peptide substrate in comparison to the cyclisation machinery from other lanthipeptide classes.

Class III lanthipeptide biosynthesis proteins are also uniquely able to catalyse to formation of labionin bonds (Meindl et al., 2010). These are produced when, in the formation of a lanthionine bond, the enolate that is normally protonated instead attacks another Dha residue, which is then protonated to form the labionin (Figure 10). Whilst there are examples of LanKC genes catalysing just the formation of lanthionine bonds, for example curvopeptin (Krawczyk et al., 2012b), just the formation of labionin bonds, for example NAI-112, catenuliptin, and the labyrinthopeptins (Iorio et al., 2014; Meindl et al., 2010; Wang and van der Donk, 2012), and both lanthionine and labionin bonds, for example in the erythreapeptins (Völler et al., 2012), an enzymatic basis for the choice between the two has not been uncovered. The directionality of core peptide processing may be key to why labionin moieties are seen only in class III lanthipeptides, as these are the only core peptides
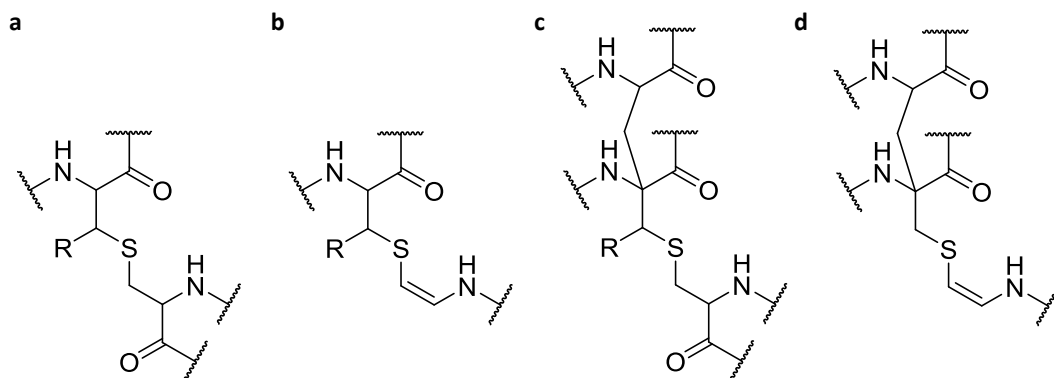
Figure 11. Structures of **a.** lanthionine (R = H) and methyllanthionine (R = CH₃), **b.** AviCys (R = H) and AviMeCys (R = CH₃), **c.** labionin (R = H) and methyllabionin (R = CH₃), and **d.** avionin bonds.

processed in a C- to N-terminal manner (Jungmann et al., 2014; Krawczyk et al., 2012a), and labionin bonds are always seen in the same directionality, with the lanthionine-linked residues towards the C-terminus and the C-C linked residues towards the N-terminus.

RiPP structural features that show remarkable similarity to lanthionine and labionin bonds are AviCys and avionin bonds (Figure 11). The AviCys and avionin bonds are versions of lanthionine and labionin bonds respectively formed with decarboxylated C-terminal cysteines (Sit et al., 2011). The decarboxylation of the terminal cysteine has been shown to be catalysed by homo-oligomeric flavin-containing cysteine decarboxylases (HFCD; often called LanD-enzymes) prior to cyclisation (Kupke et al., 1992, 1995; Majer et al., 2002). AviCys macrocycles are seen in many lanthipeptides, for example class I epidermin, gallidermin, and mutacin 1140 (Götz et al., 2014; Smith et al., 2003), and the class II mersacidin (Chatterjee et al., 1992). The gene clusters responsible for these contain the normal LanB/LanC or LanM proteins responsible for dehydration and cyclisation (Altena et al., 2000; Bierbaum, 1996; Escano et al., 2015). Microvionin, the only RiPP identified so far with an avionin bond, does not contain any conventional lanthionine bonds yet the cluster contains a *lanKC* gene that was shown to be involved in dehydration and cyclisation (Wiebach et al., 2018). LanKC proteins are typical of class III lanthipeptides; considering the similarities between avionin and labionin moieties it is unsurprising that LanKC enzymes are also involved in avionin bond formation.

An interesting observation is that some AviCys containing RiPP gene clusters do not contain lanthionine cyclase enzyme homologues, for example the linaridins and the thioviridamide-like molecules (Claesen and Bibb, 2010; Frattaruolo et al., 2017; Hayakawa et al., 2006a; Kjaerulff et al., 2017). This leads to the interesting possibility that perhaps the thiolate stabilising mechanism of the cyclases is not necessary for AviCys formation. The LanKC involved in microvionin formation does not contain the residues necessary for zinc atom coordination, and the cyclases from epidermin, gallidermin, mutacin 1140, and mersacidin may only need their thiolate stabilising zinc atom for the formation of the other lanthionine bonds in these pathways. The enethiolate that results from HFCD-

catalysed cysteine decarboxylation has a significantly lower pK_a than the thiol side chain of cysteines (Kupke and Götz, 1997), suggesting that the enethiolate in the formation of AviCys bonds does not need to be stabilised in the same way the as the thiolate in the formation of lanthionines.

1.2.2.3.   Pyruvyl-like moieties

The Dha and Dhb that result from serene and threonine dehydration in lanthipeptides are sometimes left unmodified, which can be important for the bioactivity of the molecule (Kuipers et al., 1995; Rollema et al., 1995). However, if these Dha and Dhb residues are at the N-terminus of the core peptide, it has been hypothesised that they can spontaneously form pyruvyl and 2-oxobutyryl moieties (Frattaruolo et al., 2017; Kellner et al., 1989; Mohr et al., 2015; Skaugen et al., 1994; Velásquez et al., 2011). Whilst there is no experimental evidence for this, a plausible schematic can be drawn up (Figure 12). Additionally, the lanthipeptides that contain these modifications have serines and threonines in the correct positions, and in the case of epilancin 15X it has been shown *in vitro* that proteolysis happens in the right position to leave the N-terminal serine that, if dehydrated, would be appropriate for pyruvyl formation (Ortega et al., 2014). A pyruvyl transferase can be seen installing pyruvyl groups during xanthan biosynthesis (Katzen et al., 1998), showing that an enzymatic route to a pyruvyl group is possible. However putative pyruvyl transferases are not seen associated with the pyruvyl-containing lanthipeptide clusters. The evidence suggests that a spontaneous route to pyruvyl groups from Dha is much more likely.

Epilancin K7, epilancin 15X, and thioalbamide (**18**) all contain an N-terminal lactyl group (Ekkelenkamp et al., 2005; Frattaruolo et al., 2017; Van De Kamp et al., 1995). An oxidoreductase (ElxO) from the epilancin 15X cluster was shown to be able to reduce a pyruvyl linked to a core peptide analogue *in vitro* to a lactyl group (Velásquez et al., 2011). This shows an enzymatic basis for the production of the lactyl group from the pyruvyl group and suggests that a similar mechanism is used in the production of the thioalbamide (**18**) lactyl group. Interestingly, ElxO was also shown to be able to reduce 2-oxobutyryl to 2-hydroxybutyryl, a RiPP modification yet to be seen in nature (Figure 12; Ortega et al., 2014).
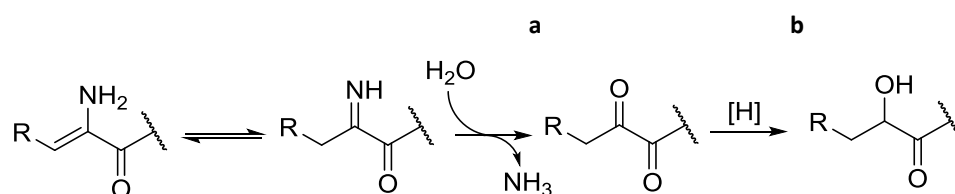
Figure 12. Schematic of **a.** pyruvyl (R = H) and 2-oxobutyryl (R = CH₃), and **b.** lactyl (R = H) and 2-hydroxybutyryl (R = CH₃) formation.
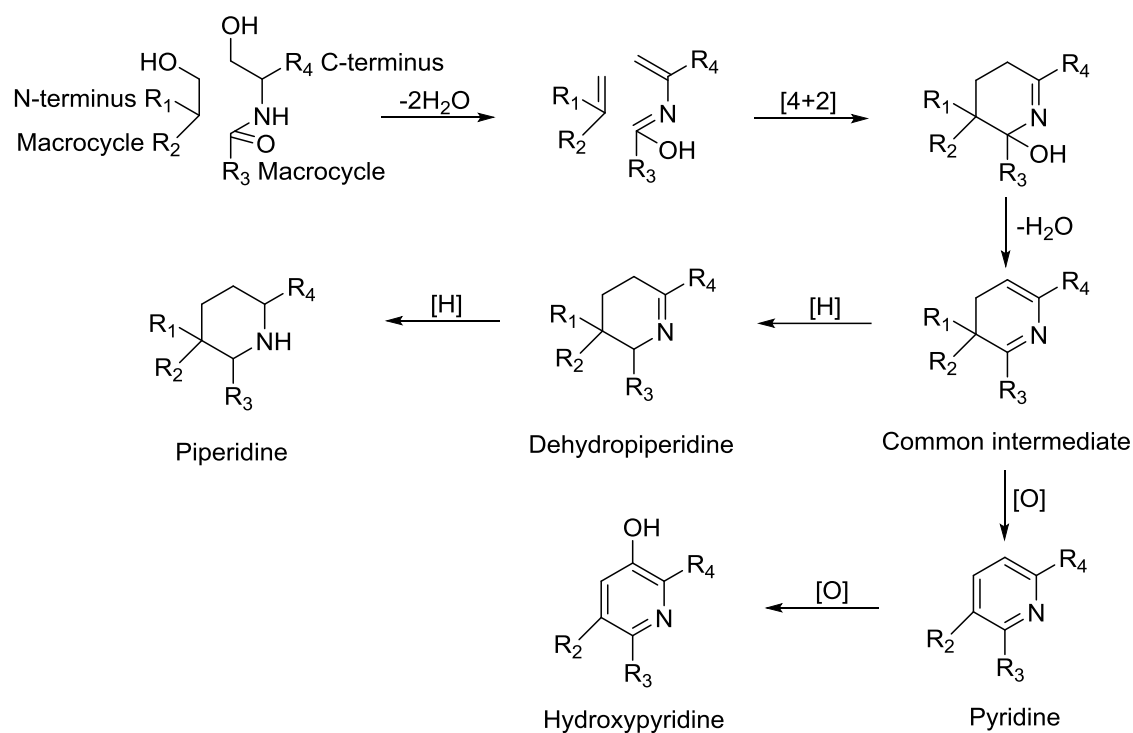
Figure 13. Schematic of pyridine, hydroxypyridine, dehydropiperidine, and piperidine formation (Bagley et al., 2005).

An interesting case is seen in the polytheonamides A and B, which contain an N-terminal 5,5-dimethyl-2-oxohexanoate moiety (Freeman et al., 2012). The biosynthetic proposal follows a similar route as the production of N-terminal 2-oxobutyryl (Freeman et al., 2017). Threonine dehydration by a LanM-like dehydratase produces a Dhb that is subsequently trimethylated by a radical SAM methyltransferase on the γ-carbon, and finally proteolysis allows for the rearrangements seen in the production of pyruvyl and 2-oxobutyryl (Freeman et al., 2017).

1.2.2.4.   Pyridine and Piperidine Rings

One of the defining features of thiopeptides is the six-membered nitrogen-containing macrocycle-forming ring (Bagley et al., 2005). This can take the form of a pyridine (for example micrococcin P1; **10**; Ciufolini and Lefranc, 2010), the hydroxypyridine (for example nosiheptide; **11**; Prange et al., 1977), the piperidines (for example Sch 18640; **12**; Puar et al., 1981), dehydropiperidines (for example thiostrepton; **13**; Anderson et al., 1970), and the dihydroimidazopiperidine (for example Sch 40832; Puar et al., 1998). All of these share a common biosynthetic route via a common intermediate from two serine residues (Figure 13; Bagley et al., 2005). The biosynthesis of dihydroimidazopiperidine has not been elucidated. The two serine residues involved in cyclisation are first dehydrated to Dha. The enzymes responsible for this show similarity to the glutamylation and elimination domains of LanB enzymes (Hudson et al., 2015), and are therefore dependant on tRNA^{Glu}. Following this, a [4+2] cycloaddition between the two Dha residues is catalysed by a Diels-Alderase enzyme that is surprisingly structurally homologous to the elimination domain of LanB

enzymes (Bowers et al., 2010; Cogan et al., 2017). Diels-Alderase enzymes are very rare in natural products, but other examples are found in spirotetronate and spirotetranate biosynthesis (Hashimoto et al., 2015; Tian et al., 2015), as well as in the biosynthesis of the macrolactone spinosyn A (H. J. Kim et al., 2011). The cycloaddition forms the common intermediate that can either be reduced to dehydropiperidine and piperidine rings, or alternatively the N-terminal leader peptide can be cleaved, aromatising the ring to form the pyridines and subsequent oxidation forms the hydroxypyridines (Bagley et al., 2005; Ichikawa et al., 2018; Mocek et al., 1993a, 1993b; Singh et al., 2008).

1.2.2.5.   Azolines and Azoles

Azol(in)e rings are a common feature in many RiPP natural products; the main classes they are seen in are the LAPs (Melby et al., 2011), cyanobactins (Sivonen et al., 2010), thiopeptides (Bagley et al., 2005), and bottromycins (Crone et al., 2016). The most common method of azole formation requires three enzymes catalysing a cyclodehydration to form azolines, and subsequent dehydrogenation to form azoles (Figure 14; Burkhart et al., 2017b). The classic example, microcin B17, used an E1-like protein (McbB), a YcaO enzyme (McbD), and a dehydrogenase (McbC; Li et al., 1996). The dehydrogenase was shown to be a flavin dependant and is responsible for conversion of azolines to azoles (Milne et al., 1999). Therefore, the E1-like protein and the YcaO enzyme are responsible for the formation of thiazolines and oxazolines from cysteine and serine residues respectively. These two enzymes are sometimes fused, for example in the cyanobactin patellamide and trunkamide gene clusters (McIntosh et al., 2010; McIntosh and Schmidt, 2010).

The YcaO protein is the catalytic member of the E1-like and YcaO pair, and it has been shown that the YcaO protein from a three-enzyme complex from *Bacillus* sp. Al Hakam can catalyse the cyclodehydrations on its own (Dunbar et al., 2012; Milne et al., 1998), albeit at a much-reduced rate. YcaO proteins have ATPase activity, yet this was missed for a long time because of its novel active site, which it uses to coordinate ATP between two Mg^{2+} ions (Koehnke et al., 2015). They catalyse the cylodehydrations by phosphorylating the carbonyl oxygen, which turns it into a leaving group and enables the attack onto the carbonyl carbon by the nucleophilic side chain of a nearby cysteine, serine, or threonine residue, to form thiazoline or oxazolines respectively (Dunbar et al., 2012). It is not clear whether the attack of the carbonyl carbon happens prior to or following phosphorylation (Figure 14; Truman, 2016).

The azoline forming YcaO proteins have a proline rich C-terminus with a PxPxP motif necessary for cyclodehydratase activity (Dunbar et al., 2014). This motif sits in the active site and likely plays an important role in substrate coordination. The PxPxP motif is also important for interactions between the YcaO and E1-like proteins, although it does not directly interact with the E1-like proteins (Koehnke et al., 2015). The E1-like protein is a largely structural protein that has a winged helix-turn
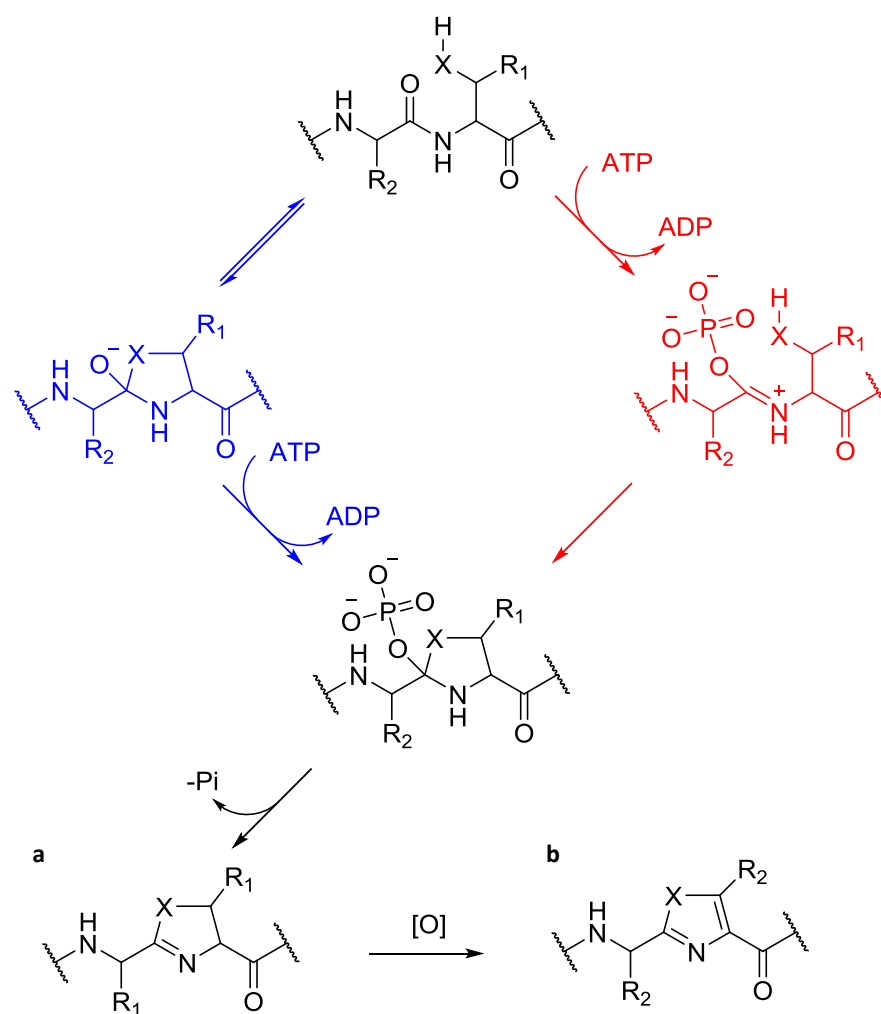
Figure 14. Schematic of **a.** oxazoline (X = O) and thiazoline (X = S), and **b.** azole (X = O) and thiazole (X = S) formation from a serine (R_1 = H, X = O), threonine (R_1 = CH_3, X = O), or cysteine (R_1 = H, X = S) onto a preceding amino acid (R_2 = preceding amino acid side chain). Alternative routes for formation are shown in red and blue (Truman, 2016).

helix domain similar to peptide clamps (Regni et al., 2009). However, in azol(in)e synthesis the E1-like protein does not employ a peptide clamp-like method of peptide binding; it instead uses a three-stranded antiparallel β-sheet that engages the leader peptide as a fourth β-strand (Koehnke et al., 2015), presenting the core peptide for YcaO modification. The E1-like protein additionally plays a role in regulating the ATPase activity of YcaO enzymes, where their rate of ATP hydrolysis is considerably retarded in the absence of the E1-like protein (Dunbar et al., 2014).

In some azol(in)e containing RiPP gene clusters, for example heterocycloanthracin and thiopeptide clusters, the E1-like protein is significantly truncated and fused with the YcaO enzyme and there is an additional protein associated with the heterocylase machinery (Dunbar et al., 2015). This additional protein is named Ocin-ThiF-like, or F protein. The F protein is distantly related to the E1 superfamily and replaces the E1-like protein's precursor peptide binding role in azol(ine) formation. The truncated E1-like protein no longer contains the β-sheet involved in precursor peptide binding,

and the N-terminus of the F protein is predicted to contain a replacement peptide binding motif (Dunbar et al., 2015).

There is not a single rule for YcaO regioselectivity, as there seems to be significant variation in the positions of the residues that are and are not cyclised. The regioselectivity of one LAP gene cluster from *Bacillus* sp. Al Hakam was thoroughly tested by replacing each normally cyclised residue in the core peptide with an alanine or valine, and then individually reverting them to a cysteine (Melby et al., 2012). The heterocyclase was able to process every position in isolation but was much more efficient at processing the position it normally modifies first. Following this it became much more efficient at processing the other positions (Melby et al., 2012). This suggests that the regioselectivity of the enzyme is linked to the order the cycle forming residues in the core peptide are processed. In other cases, the regioselectivity of the heterocyclase has evolved to facilitate later modifications to the precursor peptide. For example, cyanobactins are often fully processed with azol(in)es being formed from every cysteine, serine, and threonine residue (Schmitz et al., 1989), however in ulithiacyclamide two thiazole and two oxazoline rings are formed whilst two cysteines are left unmodified, and this allows for the later formation of a disulphide bond (Ireland and Scheuer, 1980).

Thiopeptides universally require regioselectivity, as during biosynthesis the heterocycles are some of the first modifications made, and two serines must be left unmodified for the formation of the macrocycle (Hudson et al., 2015). Many thiopeptides control the placement of heterocycles biochemically; they show a very strong preference for cyclising cysteines over serines and threonines, possibly due to the lower pKa of the cysteine side chain (Burkhart et al., 2017b). For example, during micrococcin P1 (**10**) maturation six cysteines are formed into thiazol(in)e rings, whilst the four threonines and two serines are not (Ciufolini and Lefranc, 2010). This mechanism is not universal for all thiopeptides, for example methylsulfomycin has more oxazole rings than it does thiazole rings, and yet still has a high level of control over azol(in)e ring formation; it contains two threonine and six serine residues that are not formed into azol(in)e heterocycles (Vijaya Kumar et al., 1999).

Bottromycin (**1**; Figure 15) contains a unique terminal thiazole preceded by an epimerised O-methylated aspartate residue (Shimamura et al., 2009). The gene cluster responsible contains two standalone YcaO domain proteins, one of which catalyses the cyclodehydration (Franz et al., 2017; Schwalen et al., 2017). The other is involved in macroamidine formation (see section 1.2.2.6). The YcaO heterocyclase makes up for the lack of E1- or F-like protein by binding the precursor peptide itself (Franz et al., 2017). Unusually, it does not contain proline rich C-terminus that heterocycle forming YcaO enzymes usually contain, and truncations of this termini do not abolish activity (Schwalen et al., 2017). This is at odds with other azol(in)e forming enzymes, where the proline rich C-terminus plays an important role in the active site (Koehnke et al., 2015). Formation of the terminal
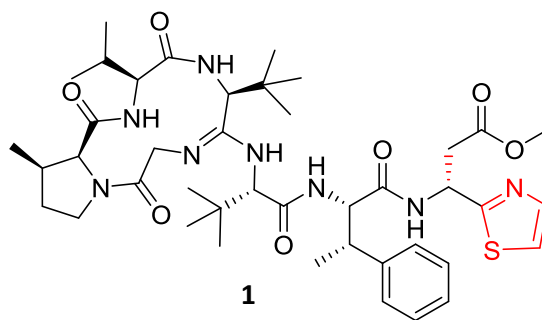
Figure 15. The structure of bottromycin (**1**). The terminal thiazole and epimerised O-methylated aspartate residue are highlighted in red.

thiazole from thiazoline does not rely on the usual flavin dependant dehydrogenase enzyme. Instead, a protease acts to leave the thiazoline with a terminal carboxylic acid, then a P450 monooxygenase catalyses oxidative decarboxylation, which dehydrogenates the ring. This also selects for the epimerised form of the compound (see section 1.2.2.8; Crone et al., 2016). The trifolitoxin gene cluster also has a standalone YcaO enzyme predicted to be responsible for the thiazoline formation, although this has not been experimentally characterised, and is only a bioinformatic prediction (Breil et al., 1993).

Methanobactin (**19**) is a RiPP that contains two oxazolones, oxazoline-like moieties, linked to thioamide bonds (Figure 16; Kim et al., 2004). It is particularly interesting that the gene cluster does not contain YcaO domain proteins, and the oxazolone and thioamide moieties are formed from a cysteine residue (Kenney and Rosenzweig, 2013). This suggests the oxazolone moiety is formed by a different mechanism from other azol(in)e heterocycles. In azol(in)e heterocycles the side chain provides the cyclising nucleophile, meaning a cysteine would form a thiazol(in)e, whereas in methanobactin (**19**) a cysteine has formed an oxazolone. It has been shown the two proteins, MbnB and MbnC, can catalyse the formation of the oxazolone and the thioamide bond *in vitro* (Kenney et al., 2018). Whilst MbnC showed no homology with any characterised proteins and no predicted domains, MbnB shows homology with the DUF692 subfamily that is related to metal dependant enzymes. It was additionally shown that MbnB and MbnC form an iron-containing complex (Kenney et al., 2018).

Based on the observation that one of the genes required for the biosynthesis of methanobactin (**19**) is a metalloprotein a reaction mechanism via an unusual β-lactam-containing intermediate has been suggested (Figure 16; Kenney et al., 2018). Whilst in YcaO-dependant azol(in)e formation the
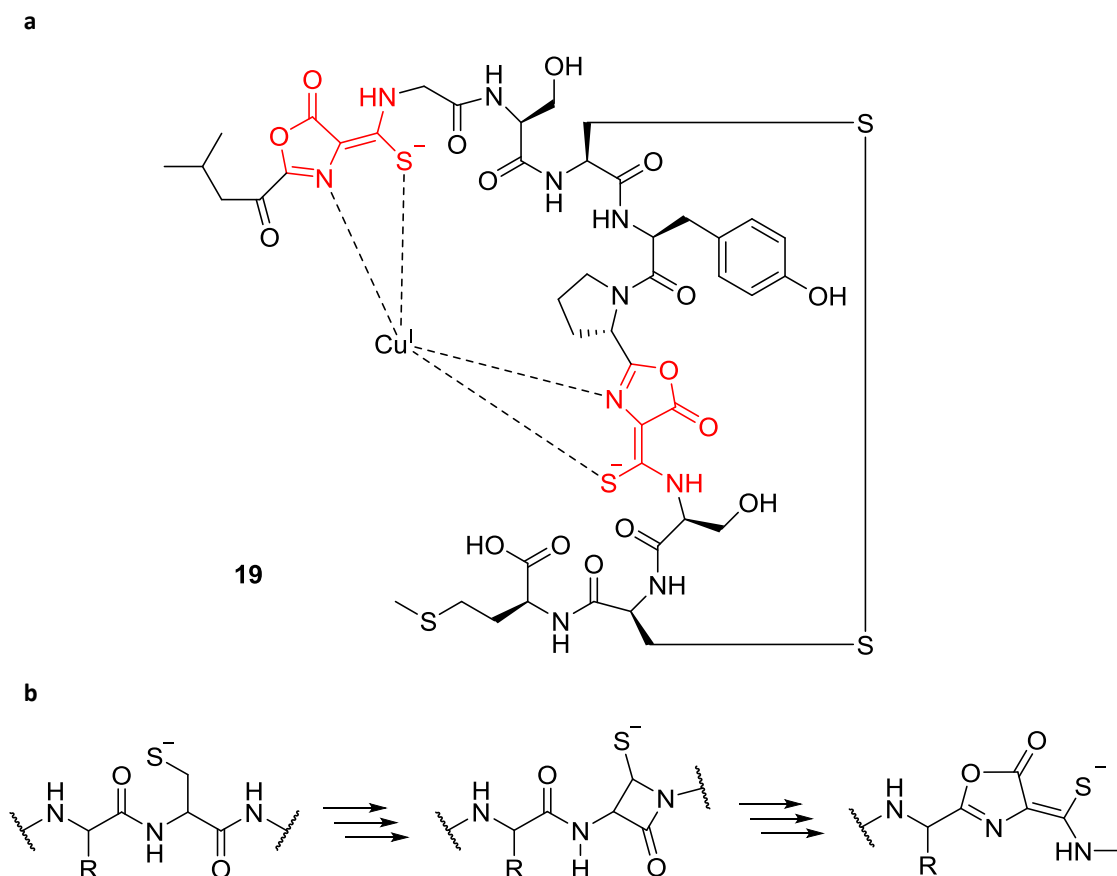
Figure 16. a. The structure of methanobactin (**19**). The oxazolones and thioamide bonds are highlighted in red. **b.** A simplified biosynthetic proposal for the formation of an oxazolone and thioamide bond via a β-lactam-containing intermediate (Kenney et al., 2018).

nucleophile substitutes for the oxygen in the preceding peptide bond, in metalloprotein-dependant oxazolone formation the oxygen from the preceding peptide bond cyclises on to the carbon of the following peptide bond. This explains how an oxazolone forms from a cysteine residue. In the process the nitrogen from the following peptide bond is transferred on to the β carbon of the cysteine side chain, forming the thioamide bond. This proposed mechanism does, however, still require experimental validation.

1.2.2.6.   Amidines

Amidine macrocycles are seen in the RiPPs bottromycin (**1**) and klebsazolicin (Metelev et al., 2017; Shimamura et al., 2009). YcaO domain proteins have been shown to be responsible for the formation of these amidine macrocycles (Crone et al., 2016; Travin et al., 2018). This is proposed to happen with a mechanism very similar to that of azol(in)e formation (Figure 17), where a nucleophilic terminal amine is the attacking nucleophile, rather than a serine, threonine, or cysteine side chain (Franz et al., 2017; Schwalen et al., 2017; Travin et al., 2018). Considering how rare this modification is, there are surprising differences between klebsazolicin and bottromycin (**1**) amidine biosynthesis enzymes.
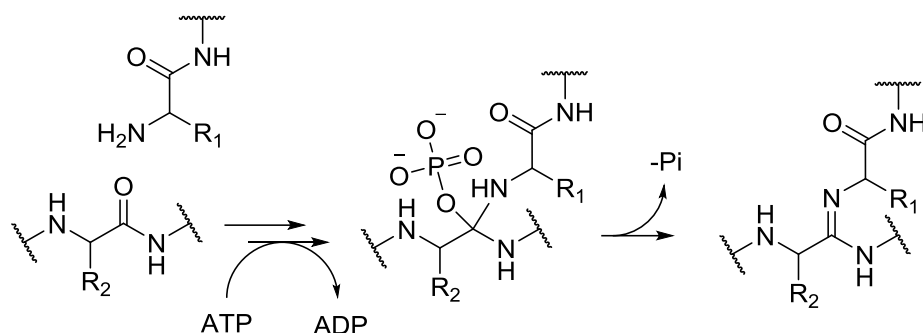
Figure 17. Schematic of amidine formation. The first step proceeds via either the red or the blue route shown in Figure 14. R_1 and R_2 are amino acid side chains.

The klebsazolicin biosynthetic pathway contains a single YcaO domain protein that is associated with an E1-like protein and a dehydrogenase, as is typical of many azol(in)e containing natural product gene clusters, and klebsazolicin contains four azole heterocycles as well as the amidine macrocycle (Travin et al., 2018). The single YcaO domain protein was shown to catalyse first cyclodehydration and then amidine cyclisation. The amidine cyclisation was coupled to leader peptide cleavage; it could not happen prior to leader peptide cleavage, as proteolysis is required to release the terminal amine involved in amidine formation, and just the azole-containing core peptide could not be processed. This is likely because the E1-protein is responsible for leader peptide binding, and therefore substrate recruitment for the YcaO domain protein. Therefore, amidine formation must be very efficient, as the YcaO-E1 complex cannot re-recruit the substrate after the leader peptide is cleaved (Travin et al., 2018). In contrast to this, the bottromycin (**1**) gene cluster has two standalone YcaO domain proteins, one involved in the thiazoline formation and one used for amidine macrocycle formation (Franz et al., 2017; Schwalen et al., 2017). The bottromycin (**1**) precursor peptide has nearly no leader peptide, only a single methionine residue, and instead has a follower peptide (Crone et al., 2012). Therefore, substrate binding follows a different mechanism, and is instead mediated by the YcaO protein itself (Franz et al., 2017), explaining the lack of E1-like proteins. Additionally, the YcaO involved in bottromycin (**1**) amidine macrocyclisation lacks the proline rich C-terminus (Schwalen et al., 2017), suggesting that its exact mechanism may be slightly different from that of klebsazolicin biosynthesis, which contains a proline in the YcaO C-terminus.

1.2.2.7. Thioamide bonds

Thioamide bonds in RiPPs are rare, however there are now three classes confirmed to have them: the copper-chelating methanobactins (Kim et al., 2004), the cytotoxic thioviridamide-like molecules (Hayakawa et al., 2006a), and some of the antibacterial thiopeptides (Hensens and Albers-Schönberg, 1983; Puar et al., 1981). When this PhD was started the biosynthesis of the thioamide bonds had not been elucidated, and therefore investigation into this was part of this project.

However, since the beginning of this project much has been uncovered on thioamide biosynthesis. As previously mentioned (see section 1.2.2.5), the origin of the thioamide in methanobactin (**19**) has been suggested to be the result of an uncharacterised protein and a DUF692 metalloprotein heterodimer catalysing a complex oxidation and rearrangement of a cysteine residue, resulting in a linked oxazolone and thioamide bond (Figure 16; Kenney et al., 2018). Full characterisation of the reaction mechanism requires more work. Homologues of this cluster are widespread, and so this may be a much more common mechanism than the current diversity of the methanobactins represents (Kenney and Rosenzweig, 2013).

In thiopeptide and thioviridamide-like gene clusters there are genes encoding YcaO proteins. It was shown that in the formation of azol(in)e heterocycles YcaO proteins phosphorylate the peptide bond carbon, facilitating the attack of the peptide bond carbon by a nucleophilic residue side chain (Dunbar et al., 2012). It was noted that if the nucleophile was instead an exogenous sulphur then the same mechanism could explain the formation of thioamide bonds (Burkhart et al., 2017b). Evidence was provided for this when it was shown that in methanogenic Archaea the thioamide bond in the protein methyl-coenzyme M reductase is installed by a YcaO and TfuA protein pair (Nayak et al., 2017). Additionally, in methanogenic Archaea the YcaO homologues are not always associated with a TfuA protein. *In vitro* reconstitutions of both a YcaO/TfuA pair and a standalone YcaO-catalysed thioamidation of methyl-coenzyme M reductase were conducted (Mahanta et al., 2018). In the YcaO/TfuA paired system, the YcaO and the TfuA were both necessary for successful thioamidation, whilst in the standalone system the YcaO on its own was sufficient to catalyse thioamidation. The role of the TfuA protein in this is not clear, as sequence and predicted structural homology searches do not provide a clue towards function. The *in vitro* reactions were ATP dependant suggesting that, as predicted, the mechanism mimicked that of azol(in)e formation with an exogenous nucleophilic sulphur (Mahanta et al., 2018).

The thiopeptide RiPPs also use a YcaO and TfuA pair to install a single thioamide bond within their structure, and a large number of uncharacterised homologues have very recently been identified (Schwalen et al., 2018). The thioviridamide-like clusters contain YcaO and TfuA enzymes (Frattaruolo et al., 2017), however in these cases the products contain multiple thioamide bonds. This suggests that, in contrast to the YcaO and TfuAs that thioamidate methyl-coenzyme M reductases and thiopeptides, the YcaO and TfuA proteins in the thioviridamide-like clusters feature a level of programmed promiscuity.

Na_2S was accepted as the sulphur donor in the *in vitro* reconstitution of the methyl-coenzyme M reductase thioamidation (Mahanta et al., 2018). However, these YcaOs are often associated with sulphur metabolism genes such as sulphurtransferases, which may be involved in providing the true

in vivo sulphur donor (Nayak et al., 2017). *In vitro* the sulphur donor could be cysteine and the sulphurtransferase IscS (Mahanta et al., 2018), therefore other sulphur donors are possible.

There are differences in the active site seen when the azol(in)e forming YcaOs are compared to the methanogenic-Archaeal YcaO proteins. For example, only a single Mg^{2+} ion is used in ATP coordination (Mahanta et al., 2018). Additionally, whilst some contain a proline-rich C-terminus, for example YcaO proteins from *Methanopyrus kandleri* and *Methanoculleus sp*. CAG:1088, others do not, for example YcaO proteins from *Methanofollis liminatans* and M*ethanothermus fervidus*. The access of the C-terminus to the active site is blocked by some nearby residues. Therefore, there either must be conformational change on substrate binding to allow C-terminal access, or a different active site morphology is employed (Mahanta et al., 2018).

1.2.2.8.   Epimerisations

One of the structural limitations of RiPPs is that they can only derive from proteinogenic amino acids. This means that every amino acid installed has L- configuration. To achieve the most effective structure for receptor binding L-amino acids are not always optimal, therefore multiple interesting methods have evolved to enable the epimerisation of α-carbons of residues in maturing RiPPs.

Many lanthipeptides have D-amino acids, for example lactosin S (Skaugen et al., 1994), carnolysin (Lohans et al., 2014), bicereucin (Huo and van der Donk, 2016), and lacticin 3147 (Martin et al., 2004). These modifications can be vital for the activity of the molecules (Cotter et al., 2005). The epimerised residues in lanthipeptides are always D-Ala, or D-2-aminobutyrate (D-Abu), and are coded for in the core peptide as serine and threonine respectively. This enables a mechanism to be proposed (Figure 18; Cotter et al., 2005) in which, following dehydration of serines and threonines, stereoselective reduction of the alkene results in a D-amino acid. Whilst there is no reason to believe a specific class of lanthionine dehydratase is necessary, the characterised examples use LanM-like proteins for serine and threonine dehydration (Huo and van der Donk, 2016; Lohans et al., 2014; Martin et al., 2004). There are two types of reductases involved in reduction of the alkenes, named LanJ$_A$ and LanJ$_B$ (Yang and van der Donk, 2015). LanJ$_A$ proteins, for example the one involved in lacticin 3147 biosynthesis, are zinc-dependent dehydrogenases that use an NADPH as a cofactor (Cotter et al.,
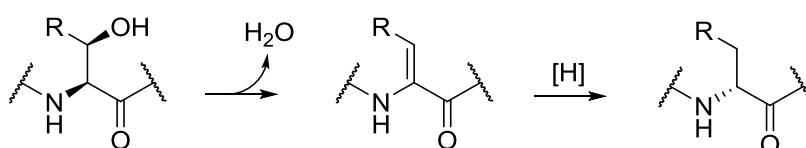
Figure 18. Schematic of epimerisation via dehydration and reduction of L-serine (R = H) and L-threonine (R = CH₃) to D-alanine (R = H) and D-2-aminobutyrate (R = CH₃; Cotter et al., 2005).
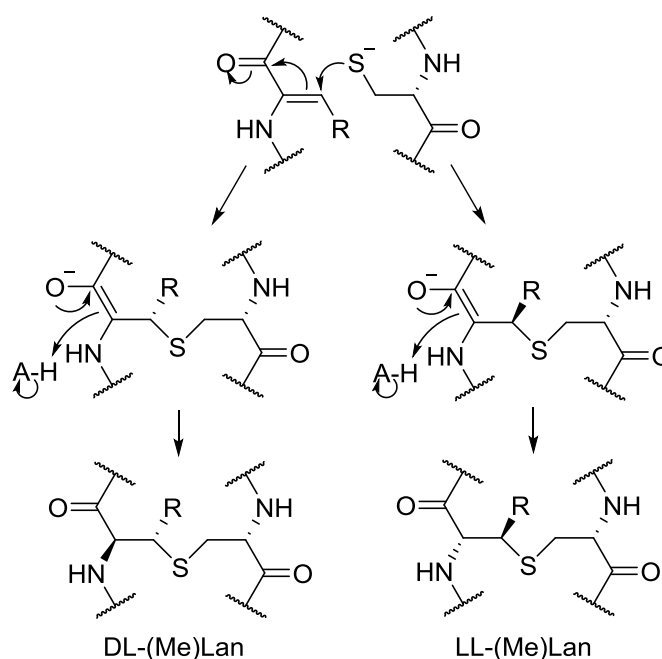
Figure 19. Schematic of the formation of DL- and LL- lanthionine (R = H) and methyllanthionine (R = CH₃) bonds.

2005; Yang and van der Donk, 2015). LanJ₈ proteins, for example in carnolysin biosynthesis, belong to the flavin oxidoreductase family (Lohans et al., 2014; Yang and van der Donk, 2015).

It should also be noted that in lanthionine bond formation DL-rings are usually formed (Figure 19). To form this, the α carbon of the Dha or Dhb residue is protonated after bond formation in a stereoselective manner to result in epimerisation (Repka et al., 2017). In fact, this is so common that lanthionine bond formation in which a residue is not epimerised is rare. In the formation of LL-rings, conjugate attack and protonation still happens on the opposing faces of the alkene, as in DL-formation, however both attack and protonation are swapped to the other faces, to form the LL-isomer. Single LanM enzymes have been characterised catalysing lanthionine residues with both DL- and LL- configurations in a very regiospecific manner, for example in cytolysin (**7**; Dong et al., 2015), showing that single LanM enzymes can be capable of forming both configurations. This is a rare instance of substrate-controlled stereochemistry, where the formation of LL- lanthionine results from a Dhb-Dhb-Xxx-Xxx-Cys motif forcing a non-canonical interaction with the LanM active site (Tang et al., 2015). There are, however, examples where a LL-lanthionine is installed in the absence of a Dhb-Dhb-Xxx-Xxx-Cys motif (Garg et al., 2016), suggesting that more complex factors can be involved in determining stereochemistry.

An interesting property of azoline rings is that they can spontaneously epimerise (Milne et al., 2006). Dynamic kinetic resolution can select for one of the epimers. For example, with the aspartate that precedes the thiazoline ring during bottromycin (**1**) biosynthesis, it is proposed that only the D-form of the aspartate is an appropriate substrate for the P450 decarboxylase (Crone et al., 2016). The
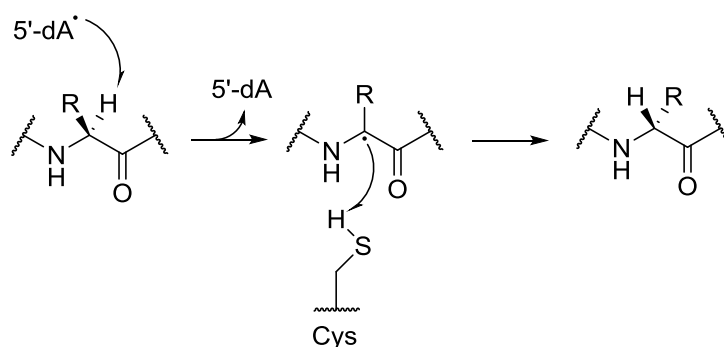
Figure 20. Schematic of radical SAM enzyme-mediated epimerisation of a residue (R is the residue side chain; Benjdia et al., 2017; Morinaka et al., 2014).

P450-catalysed oxidation of the thiazoline to a thiazole locks the aspartate in a D-form, as spontaneous epimerisation cannot happen preceding an azole ring. This is also seen in many cyanobactins, for example the patellamides contain D-amino acids preceding azole rings (Koehnke et al., 2014). It is likely that these are also the result of spontaneous epimerisation and then dynamic kinetic resolution by the dehydrogenase that oxidises the azolines to the azoles (Milne et al., 2006).

Both the epipeptides and the proteusins contain D-amino acids (Benjdia et al., 2017; Freeman et al., 2012). The epipeptide epimerase has been more thoroughly characterised, but the epimerisation is catalysed by radical SAM enzymes in both cases, with the proteusin polytheonamide radical SAM enzyme exhibiting remarkable flexibility and regioselectively by epimerising 18 of the 48 residues in the core peptide (Freeman et al., 2017). The radical SAM enzymes catalyse the epimerisation of a residue by abstracting the hydrogen atom from the α carbon, leaving a radical which is quenched by a hydrogen atom on the opposite side of the residue, resulting in epimerisation (Figure 20; Benjdia et al., 2017; Morinaka et al., 2014). In the case of the epipetide radical SAM enzyme, the hydrogen atom donor that quenches the radical α carbon is a cysteine residue within the enzyme. The radical SAM enzymes from the epipeptides and the proteusins show surprisingly little sequence and predicted structural homology. For example, the radical SAM enzyme from the epipeptides contains a SPASM domain, that is likely involved in regenerating the cysteine residue hydrogen atom donor within the enzyme (Benjdia et al., 2017), whilst the proteusin radical SAM enzyme is not predicted to contain this domain. This suggests that different mechanisms may be employed to regenerate the enzymes and that different hydrogen atom donors may be used.

1.2.2.9.   C-methylations

C-methylation in amino acids is a difficult chemical transformation due to the non-nucleophilic nature of the carbons. Radical S-adenosylmethionine (rSAM) enzymes can use a radical mechanism to manage this unfavourable methylation (Mahanta et al., 2017a). Cleavage of a bond within the SAM cofactor produces a 5'-deoxyadenosyl (5'-dA) radical that can abstract a hydrogen from a substrate. There are four classes (A to D) of rSAM methyltransferases, and two (B and C) have been

identified functioning in RiPP biosynthesis. Class B rSAM enzymes are responsible for the C-methylations in RiPPs such as bottromycin (**1**; Crone et al., 2016), polytheonamide (Parent et al., 2016), thiostrepton (**13**; Benjdia et al., 2015), and siomycin (Liao et al., 2009). Class C rSAM enzymes are responsible for the C-methylations in RiPPs such as Thiomuracin (Mahanta et al., 2017b), nosiheptide (**11**; Wei Ding et al., 2017b), nocathiacin (Zhang et al., 2011), and GE2270A (Mahanta et al., 2017b).

The gentamicin rSAM methyltransferase, whilst not from a RiPP gene cluster, is characteristic of class B rSAM enzymes (Kim et al., 2013). Class B rSAM enzymes are characterised by an N-terminal cobalamin binding domain. In gentamicin biosynthesis one molecule of SAM is cleaved to produce the 5'-dA radical, which abstracts a hydrogen atom from the substrate, and then methyl-cobalamin donates a methyl group to quench the radical (Kim et al., 2013). A methyl group from a second SAM molecule is used to regenerate the methyl-cobalamin (Blaszczyk et al., 2016). The *in vitro* reconstitution of a polytheonamide rSAM methyltransferase suggests that it follows a similar mechanism to the gentamicin rSAM (Parent et al., 2016). The bottromycin (**1**) rSAM methyltransferases have not been as well characterised, yet probably act with a similar mechanism (Crone et al., 2012). Conventionally class B rSAM enzymes contain a CxxxCxxC motif that coordinates the [4Fe-4S] cluster (Zhang et al., 2012a), whilst the bottromycin (**1**) and polytheonamide rSAM methyltransferases contain an unusual CxxxxxxxCxxC motif; the mechanistic or structural implications of this have not been explored (Crone et al., 2012; Parent et al., 2016). The bottromycin (**1**) and polytheonamide rSAM methyltransferases transfer methyl groups to sp^3-hybridised carbon atoms, whilst the thiostrepton (**13**) rSAM methyltransferase transfers a methyl group to sp^2-hybridised carbon atoms (Benjdia et al., 2015). This difference is associated with a difference in mechanism; the thiostrepton (**13**) rSAM methyltransferase has been shown to not catalyse reductive cleavage of SAM to produce a 5'-dA radical (Blaszczyk et al., 2016). Therefore, whilst this enzyme shares predicted structural features with other class B rSAM methyltransferases (Benjdia et al., 2015), it is likely that the methyl transfer does not follow a radical mechanism. The actual mechanism of methyl transfer has yet to be established.

The mechanism of class C rSAM methyltransferases has not been fully characterised, however recent studies on the thiomuracin and the nosiheptide (**11**) methyltransferases has allowed a proposal of how the methyl transfer may take place (Ding et al., 2017c; Zhang et al., 2017). As in most rSAM enzymes, one SAM molecule is cleaved to produce the 5'-dA radical. This is used to abstract a hydrogen atom from a second SAM molecule, forming a methylene radical that can attack the substrate. Re-protonation and heterolytic C-S bond cleavage releases the substrate carrying a methylene radical. The methylene radical is converted to a methyl group when an electron is

transferred from an [4Fe-4S] cluster and subsequent protonation occurs (Ding et al., 2017b; Mahanta et al., 2017b, 2017a).

1.2.2.10.   O/N-Methylations

N-methylation is a very common posttranslational modification of bacterial proteins (Cain et al., 2014). PrmA, conserved across bacteria, tri-methylates the N-terminal α-amino group and ε-amino groups of lysine residues within ribosomal protein L11 (Demirci et al., 2008). N- and O-methylation are features that are also seen in RiPP natural products as well. For example, the cypemycin and other linaridins show α-N-methylations at their N-terminus (Minami et al., 1994). These are performed by SAM-dependant methyltransferases. Interestingly, the linaridin α-N-methyltransferases are more closely related to the SAM-dependant C-methyltransferase UbiE than to PrmA-like methyltransferases (Zhang and van der Donk, 2012). This suggests an unexpected convergent evolution of multiple N-methyltransferases. Analysis of PznL, an α-N-methyltransferase involved in plantazolicin biosynthesis, supports this as it clades separately from UbiE, PrmA, and linaridin-like methyltransferases (Kalyon et al., 2011; Lee et al., 2013; Zhang and van der Donk, 2012).

Not all N-methylations occur on the α-amino group of RiPPs, for example the polytheonamides and the thiopeptide GE2270 contain asparagine N-methylations (Freeman et al., 2012; Tocchetti et al., 2013). The polytheonamide SAM-dependent asparagine N-methyltransferase exhibits remarkable promiscuity as it selectively methylates eight asparagine residues throughout the molecule (Freeman et al., 2017). Microcyclamide contains an uncharacterised histidine methylation (Ziemert et al., 2008), and thioviridamide-like molecules also contain an uncharacterised and unique histidine bis-N-methylation (Frattaruolo et al., 2017). Possibly one of the most exotic N-methylations seen in RiPPs is seen in the fungus *Omphalotus olearius*, which produces the omphalotins (van der Velden et al., 2017). The leader peptide of these small cyclic RiPPs consists of nearly 400 amino acids, half of which form a SAM-dependant N-methyltransferase. This protein autocatalytically methylates 9 of 12 amide nitrogens within the core peptide prior to separation and cyclisation of the core peptide (Song et al., 2018). The mechanism of this involves a water molecule being used to lower the pK_a of the amide nitrogen, which promotes deprotonation and allows attack of the SAM methyl group to produce the methylated amide nitrogen (Song et al., 2018).

SAM-dependant O-methylations are extremely common modifications made to natural products (Liscombe et al., 2012). They follow a similar mechanism to N-methylations; the enzyme lowers the pK_a of the target oxygen, enhancing its nucleophilicity, and this is then followed up by donation of the electrophilic SAM methyl group (Liscombe et al., 2012). Considering the abundance of O-methylation in nature there are surprisingly few RiPPs that are O-methylated. Some examples are O-methyl-glycothiohexide (Northcote et al., 1994), where its hydroxypyridine ring is O-methylated,

bottromycin (**1**) where an aspartic acid is methylated (Crone et al., 2016), and thiostreptamide S4 (**17**) where a tyrosine is O-methylated (Frattaruolo et al., 2017).

1.2.2.11. Hydroxylations

Carbon hydroxylation is a common reaction in metabolite maturation, where it is usually catalysed by cytochrome P450 monooxygenases (Hamdane et al., 2008), α-ketoglutarate-dependant dioxygenases (Wu et al., 2016), and flavin-dependent monooxygenases (Huijbers et al., 2014). Within the RiPP landscape, hydroxylations are primarily seen in the thiopeptides, but are also found in other RiPP classes (Funk and van der Donk, 2017). Hydroxylations in RiPP biosynthesis are catalysed by cytochrome P450 enzymes or α-ketoglutarate-dependant dioxygenases.

The cytochrome P450 monooxygenases and the α-ketoglutarate-dependant dioxygenases work by using iron to bind molecular oxygen. The O-O bond is broken to form a reactive iron-oxo in two different ways. In a cytochrome P450 monooxygenase the molecular oxygen reacts with two protons from the solvent to form water, leaving behind the iron-oxo (He and Montellano, 2004). In an α-ketoglutarate-dependant dioxygenase the molecular oxygen attacks the α-ketoglutarate, triggering the cleavage of the oxygen bond to form the reactive iron-oxo (Wu et al., 2016). An oxygen rebound mechanism then follows to hydroxylate the carbon substrate. The oxygen abstracts a hydrogen from the substrate leaving a radical intermediate. The iron bound hydroxyl then recombines with the radical intermediate to give the final hydroxylation.

The thiopeptides are not uniform in the enzymes used for hydroxylation of residues. In the biosynthetic pathways of thiostrepton (**13**), nosiheptide (**11**), and GE2270, cytochrome P450 monooxygenases are employed to catalyse β- and γ-hydroxylation of residues. Similarly, in nosiheptide (**11**) biosynthesis a cytochrome P450 monooxygenase is used to convert its pyridine into a hydroxypyridine (Kelly et al., 2009; Liu et al., 2013; Tocchetti et al., 2013). In contrast, in thiocillin biosynthesis an α-ketoglutarate-dependant dioxygenase catalyses β-hydroxylation of a valine residue (Wieland Brown et al., 2009).

The promiscuity of these enzymes follows no currently established rule. Whilst the positions of hydroxylations are specific, the number catalysed by different enzymes varies. An α-ketoglutarate-dependant dioxygenase catalyse four hydroxylations in the polytheonamides (Freeman et al., 2017). Whilst a similar enzyme installs only a single hydroxylation in the lanthipeptide cinnamycin (Ökesli et al., 2011). Two families of fungal RiPPs, the amatoxins and phallotoxins, have between one and five hydroxylations, some of which have been shown to be catalysed by the same cytochrome P450 monooxygenases. Therefore, some cytochrome P450 monooxygenases can also hydroxylate more than one location (Walton, 2018).

1.2.2.12.   Core peptide excision

Natural product producers have many ways to digest peptides that could be adopted to excise the core peptide of RiPPs from their precursor peptide, often by removal of a leader peptide. Accordingly, many methods of core peptide excision are seen. As the leader peptide is used for the binding of many tailoring enzymes (Burkhart et al., 2015; Oman and van der Donk, 2010), its removal is often one of the last steps of biosynthesis, and results in the release of the mature product. Some of the best understood methods of core peptide excision are presented here, but this is not an exhaustive list.

The majority of class I lanthipeptides use subtilisin-like serine peptidases to remove the leader peptide; termed LanP enzymes (Repka et al., 2017). The nisin LanP has been characterised as an extracellular protein that removes the leader peptide to produce nisin (**6**) following export of the heterocyclised core peptide bound to the leader peptide (Van der Meer et al., 1993). This method ensures that mature nisin (**6**) is not produced within the cell. Not all subtilisin-like serine peptidases involved in lanthipeptide maturation are extracellular, for example the epilancin 15X LanP is cytoplasmic and excises the core peptide prior to export (Ortega et al., 2014).

Some class I and most class II lanthipeptides use a bifunctional LanT$_P$ enzyme, from the ABC-transporter maturation and secretion (AMS) protein family (Havarstein et al., 1995; Rince et al., 1994), to couple proteolysis and export. These are also seen in many bacteriocin clusters (Franke et al., 1999). These proteins are ABC-transporters with a cytosolic N-terminal papain-like cysteine protease (Ishii et al., 2010), and therefore leader peptide cleavage precedes export. The site of leader peptide cleavage often occurs immediately after a double-Gly (GG) motif (Chen et al., 2001).

Two different mechanisms for leader peptide removal have been identified for class III lanthipeptides. Many clusters are not associated with putative proteases, for example erythreapeptin, avermipeptin, and griseopeptin, and it is likely that endogenous proteases engage in stepwise digestion of the leader peptide releasing products with varying numbers of N-terminal amino acids (Völler et al., 2012). Alternatively, some class III lanthipeptides are associated with prolyl oligopeptidases that selectively cleave the leader peptide to form the mature product (Völler et al., 2013).

LAPs such as microcin B17 and klebsazolicin use a "molecular pencil sharpener" mechanism to remove their leader peptide (Ghilarov et al., 2017; Travin et al., 2018). This protease is formed from two similar proteins that assemble into a spherical heterodimer. A narrow cleft in the centre allows only unstructured peptides, such as the leader peptide of microcin B17, to enter. The leader peptide is then digested by an internal cleavage site provided by one of the proteins. The protease is a zinc

or iron metalloprotein similar to thermolysin, and likely has a similar mechanism of action (Ghilarov et al., 2017).

Thiopeptides that contain a pyridine or hydroxypyridine ring have an interesting method of leader peptide removal, as it is removed by elimination rather than by hydrolysis (Hudson et al., 2015). The [4+2] Diels-Alderase that forms the pyridine ring catalyses the elimination of the leader peptide as a carboxamide, resulting in the dehydrogenation required to form the pyridine ring. This is interesting considering the homology of thiopeptide Diels-Alderases with LanB enzymes that catalyse glutamate elimination (Hudson et al., 2015).

Bottromycin (**1**) is a particularly unusual example, as the precursor peptide contains a follower peptide instead of a leader peptide (Crone et al., 2012). Therefore, excision of the core peptide requires a methionine to be removed from the N-terminus and proteolysis of the C-terminal follower peptide. The enzymes involved are known to be an amidohydrolase that removes the follower peptide, and a separate leucyl aminopeptidase-like methionine aminopeptidase removes the methionine (Crone et al., 2016; Huo et al., 2012). Leucyl aminopeptidases traditionally cleave N-terminal leucines from peptides (Matsui et al., 2006), however it has been shown that some will preferentially cleave methionine (Herrera-Camacho et al., 2007), as seen in bottromycin (**1**) core peptide excision. The methionine aminopeptidase has been crystallised and shown to cleave this methionine *in vitro* (Crone et al., 2016; Mann et al., 2016). The amidohydrolase involved in removing the follower peptide is in the process of being characterised (Andrew Truman, personal communication).

1.3. Synthetic Biology and Cluster Manipulation

1.3.1. The Purpose of Cluster Manipulation

1.3.1.1. Understanding Biosynthesis

There are many reasons to want to manipulate biosynthetic gene clusters. The most common use of biosynthetic gene cluster manipulations is to investigate biosynthesis. Deleting genes can confirm the boundaries of gene clusters, whether genes within the cluster are involved in biosynthesis, and determine the function of biosynthetic genes. Gene clusters can also be built from the bottom up to confirm the function of each gene. Whilst this is a rarer technique because it involves more advanced assemblies, this has been done with the SGR810-815 gene cluster from *Streptomyces griseus* (Luo et al., 2013). Here, different gene cluster lengths were assembled and tested for their production.

The use of gene deletions and gene-by-gene reconstruction of gene clusters in RiPPs is vital to understand their biosynthesis, however it is faced by difficulties. If the deleted biosynthetic gene is one that acts early in a biosynthetic pathway, then the resulting product can often be an unstructured and unmodified peptide. These can be readily digested by endogenous proteases and acetylated endogenously. This can make identification of the intermediates and shunt metabolites challenging and time consuming. In the analysis of bottromycin (**1**) biosynthesis, in-depth untargeted metabolomics and metabolite networking was used to solve this issue (Crone et al., 2016).

1.3.1.2. Increasing Production

Increases in production allow for better detection and enables the isolation of metabolites, which is critical for structural elucidation by NMR and bioactivity assays. Improved detection can aid in the search for poorly produced intermediates or side products. Increases in yield can also allow for the identification of products from a cryptic or naturally inactive gene cluster. In industry, the traditional method to increase the production of natural products is random mutagenesis alongside screening for optimal fermentation conditions. Screening mutants for yield improvements is time consuming and labour intensive, and so is not well suited to academic laboratories. Instead, targeted cluster manipulations can be used to increase production (Luo et al., 2015).

The regulation of RiPP biosynthesis is very diverse (Bartholomae et al., 2017). This means there is no universal targeted method that can be used to increase production. For example, whilst many lanthipeptide gene clusters contain specific activators that can be upregulated to increase production, such as the regulator in the SapB gene cluster (Nguyen et al., 2002), this is not universal amongst lanthipeptides and is much rarer in other classes of RiPPs. This means that successful increases in the production of RiPPs can often require multiple attempts, for example in the

overexpression of the thiopeptide GE2270 eight versions of the gene cluster were constructed (Flinspach et al., 2014). Therefore, efficient methods for gene cluster manipulation are often needed to increase RiPP production.

1.3.1.3.   Product Engineering

In converting lead compounds to drugs there is a success rate of 21.6% (Wong et al., 2018). This means that four out of five potential drugs fail due to properties of the bioactive molecule, such as pharmacokinetics, stability, and *in vivo* activity. As a major application of natural products is drugs, there is a lot of interest in finding ways to increase this poor success rate. One method for this is to rationally alter the molecule's properties. If the reasons a natural product failed to become a drug are known, then this knowledge can be used to design modifications to the natural product to address these issues. A potential way to do this is to modify the gene cluster responsible to make targeted changes to the final product. An alternative is making modifications to a cluster to make libraries of derivatives of the mature natural product, for example "rapalogues" of rapamycin (**3**; Wlodek et al., 2017), with the hope that entering pre-clinical trials with multiple related drug candidates will increase the chances of having a molecule that does not fail the later-stage trials.

It can also be desirable to change the product of a gene cluster to simply make other investigations easier. For example, when doing proof-of-principle engineering experiments it can be beneficial to make an inactive version of the natural product, to avoid self-toxicity. This approach was used when modifying cinnamycin biosynthesis; the gene cluster was altered so the product would be inactive and therefore simplify further experiments (Lopatniuk et al., 2017).

1.3.2.   Currently Available DNA Editing Techniques

1.3.2.1.   *In Situ* Modification of Native Producers

Manipulating biosynthetic gene clusters to understand biosynthesis, increase yield, and modify the product can be done in many ways. Gene clusters can be modified *in situ*, where they are modified in the producing organism, or they can be modified out of the producing organism, where one of many *in vitro* or *in vivo* techniques can be used (see sections 1.3.2.3 and 1.3.2.4). When modifying the gene cluster *in situ* the advantage is that the gene cluster does not need to be cloned on a vector to enable0 *in vitro* or *in vivo* DNA editing techniques. However, the main disadvantage is caused by the wide range of organisms that produce natural products; many organisms are genetically intractable or require bespoke tools to be developed for them. Even in the most prolific and well-studied natural product producers, the actinomycetes, where a wealth of tools are available (Kieser et al., 2000), many strains remain genetically intractable. Therefore, when working with a difficult organism, or one for which bespoke tools have yet to be developed, *in situ* modifications of the gene clusters can be impractical. As many different tools have been developed for many different

organisms, tuned to work with each organism's abilities and limitations, there are too many to summarise fully here. Comprehensive reviews are available for the *in situ* editing tools usable in natural product producers such as *Escherichia coli* (Sathesh-Prabu and Lee, 2018), *Bacillus spp.* (Dong and Zhang, 2014), cyanobacteria (Wolk and Wolk, 2002), *Saccharomyces cerevisiae* (hereafter referred to as yeast; Fraczek et al., 2018), and filamentous fungi (Jiang et al., 2013), to name but a few. Here, a brief summary of some of the tools available and the challenges faced for *in situ* editing in actinomycetes is presented.

The first barrier to *in situ* modification of actinomycete gene clusters is transforming the bacteria. Actinomycetes cannot be transformed by the standard heat shock method of transforming *E. coli*, and whilst electroporation has been reported many times (Fan et al., 2013; Ma et al., 2014; Mazy-Servais et al., 1997; Pigac and Schrempf, 1995), it is not reliable and reproducible amongst many actinomycetes. As such, alternative methods are commonly used. DNA can be transformed into many actinomycetes by protoplast transformation, in which the cell wall is removed to allow DNA uptake, or by conjugation in which *E. coli* is used to transfer the DNA to the actinomycetes (Kieser et al., 2000). Additionally, many actinomycetes have methylation specific restriction systems, such as *S. coelicolor* and *S. griseus* (González-Cerón et al., 2009; Kwak et al., 2006) which often necessitates the passage of DNA through methylation deficient *E. coli*, for example ET12567 (MacNeil et al., 1992), prior to conjugation or transformation. However, some are more tolerant of methylated DNA, for example *Streptomyces lividans* (Liu et al., 2010).

For simple manipulations of biosynthetic gene clusters, where biosynthetic or regulatory genes are to be added, these genes can be expressed *in trans*. For this, phage integrative vectors can be used. These carry phage machinery that integrates the vector into the actinomycete genome in a sequence specific location, for example from the *Streptomyces spp.* phage φC31 (Thorpe and Smith, 1998). However, when regions of DNA need to be removed or replaced during gene cluster modification, alternative methods are needed. Simple suicide vectors can be used, which use large regions of homology to insert an entire plasmid within a target gene by homologous recombination to inactivate it, however this is very disruptive and can lead to off-target effects (Kieser et al., 2000). An alternative is using a double-crossover method, which is similar to the suicide vector method, but the initial vector integration does not delete the target sequence. Instead, the vector includes a pair of DNA sequence "arms" that are homologous to either side of the region to be deleted. A second round of homologous recombination is used to remove the target sequence along with the vector, leading to a clean deletion. The yeast meganuclease I-SceI has also been used to force the second round of homologous recombination, which can speed up the process and minimise screening (Figure 21; Fernández-Martínez and Bibb, 2014). Similar homologous recombination techniques can
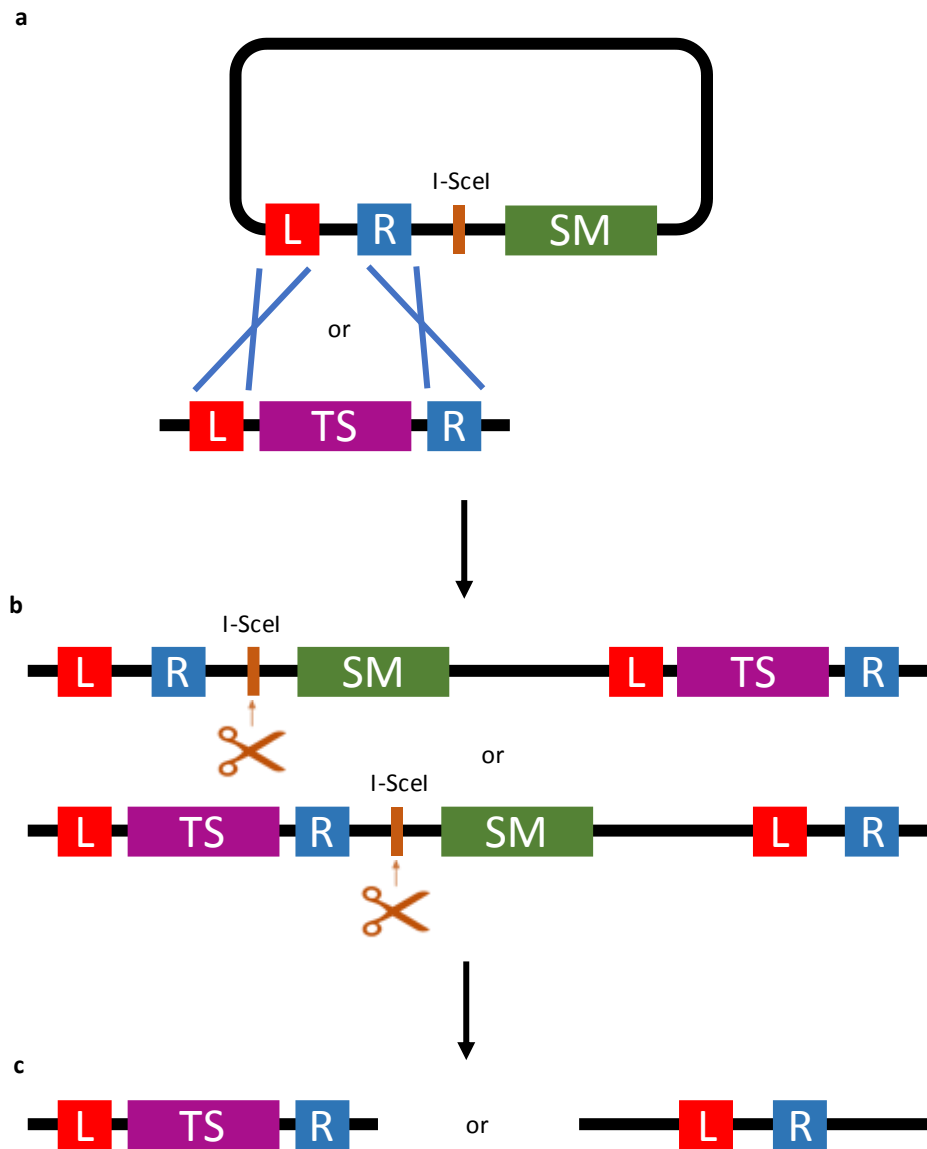
Figure 21. Meganuclease-mediated deletion of a target sequence (TS) in gDNA (Fernández-Martínez and Bibb, 2014). **a.** A plasmid containing a selectable marker (SM), an I-SceI recognition site, and regions of homology (L and R) to either side of the TS is transformed into the target strain. The plasmid integrates at either the L or the R region of homology to produce one of two constructs. **b.** I-SceI is expressed from a separate plasmid and cuts within the newly integrated sequence. This forces a second round of recombination between either the L or the R regions of homology. **c.** 50% of homologous recombination events result in reversion to the original genotype, containing the TS, whilst the other 50% of homologous recombination events result in deletion of the TS.

be used to insert regulatory elements such as promoters upstream of operons to activate them (Olano et al., 2014).

An issue with the homologous recombination-based techniques is that selectable markers must be installed with the modifications. Even if these can be subsequently removed this still slows down the process and leads to off-target effects. An alternative is to cut the bacterial chromosome at a target location and force the organism to repair around the cut with DNA carrying modifications. This requires the expression of a programmed nuclease that will cut the target DNA. The available

programmable nucleases that can be used *in vivo* are zinc-finger nucleases (ZFNs; Miller et al., 2007), transcription activator-like effector nucleases (TALENs; Christian et al., 2010), and clustered regularly interspaced short palindromic repeats-Cas9 (CRISPR-Cas9; Cong et al., 2013; Ran et al., 2013). Neither ZFNs or TALENs have been used to make modifications in actinomycetes, however recently CRISPR-Cas9 has been developed for actinomycetes (Cobb et al., 2015; Jia et al., 2017). This has been showcased by making full gene cluster, single gene, and single nucleotide deletions (Huang et al., 2015), as well as introducing promoters to activate cryptic gene clusters (Zhang et al., 2017).

A particularly interesting technique for *in situ* modification of gene clusters was recently demonstrated in *Streptomyces rapamycinicus* and *Streptomyces fradiae* (Wlodek et al., 2017). The process, named accelerated evolution, involves inserting a plasmid within a polyketide synthase gene cluster. The plasmid's temperature sensitive replicon is activated, causes catastrophic instability within the gene cluster. The repair mechanisms, largely homologous recombination, result in PKS rearrangements which delete and occasionally insert modules, causing a diversification of the products (Wlodek et al., 2017). This technique, whilst powerful, likely has limited applicability outside of modular synthases.

There are many issues associated with *in situ* modification of gene clusters in actinomycetes. For example, it often relies on the actinomycete machinery for homologous recombination, which is not well understood (Hoff et al., 2017; Zhang et al., 2014). Many actinomycetes are also resistant to genetic modification, have slow growing times, and have poorly optimised growth conditions (Kieser et al., 2000). Tools developed for any *in situ* application are not universal. This prevents investigations outside a laboratory's area of expertise, and therefore is not conducive to researching natural products in the broad diversity of organisms they are found in (Cox et al., 2015). A solution to this is manipulating gene clusters outside of their wild type host organism. This allows the development of universal tools that can be applied to all gene clusters. The next three sections will discuss the *in vitro* and *in vivo* methods for cloning, assembling, and modifying DNA, and their limitations.

1.3.2.2.   Gene Cluster Cloning

If a gene cluster is going to be modified outside of the organism it originates from, then it must first be assembled into a vector. DNA synthesis is an exciting possibility for gene cluster assembly and it is likely that many gene clusters will be assembled and modified in this way in the future. However, there are currently many issues with this approach. High GC DNA and repetitive sequences are still troublesome during DNA synthesis (Kosuri and Church, 2014). The actinomycetes contain high GC DNA and many repetitive sequences within their modular gene clusters. DNA synthesis of full gene clusters is also currently prohibitively expensive for most academic researchers to routinely do, especially considering that multiple modified versions of gene clusters are often required. Additionally, a prerequisite for design of a synthetic gene cluster is accurate sequence data, which is

still regularly not available. As DNA synthesis is more a commercial enterprise than an academic one, it will not be reviewed here.

If a gene cluster is small enough, then it can simply be PCR amplified and ligated into a vector. It can also be assembled from multiple PCR fragments, in which case see the next two sections for a summary of assembly methodologies. However, if neither of these techniques are chosen or applicable then a gene cluster must be captured from the genomic DNA. Some of the different methods for capturing gene clusters are described here. A large but incomplete list of many examples of captured gene clusters from *Streptomyces* spp. and the methods used to do them has been collated (Nah et al., 2017). These techniques can be broadly separated into untargeted and targeted techniques.

Untargeted ligation-based techniques are very commonly used (Nah et al., 2017). To do this, genomic DNA is extracted and then ligated in to vectors. This was done for *S. coelicolor,* capturing nearly the entire genome over 637 different cosmids, with an average captured size of 37.5 kb (Redenbach et al., 1996). The cosmid library can then be screened for the presence of a gene cluster of interest. Larger regions of DNA can be captured with the construction of libraries using bacterial artificial chromosomes (BAC) and P1-derived artificial chromosomes (PAC), which can handle up to 300 kb of DNA (Alduina et al., 2003; Alduina and Gallo, 2012). These libraries are a particularly powerful tool for capturing large gene clusters and are especially useful if multiple gene clusters within a single organism are being investigated.

Some targeted methods involve integrating regions of DNA on either side of target gene clusters. For example, a cluster can be cloned from bacteria by inserting conjugal origin of transfer (oriT) sites on either side of the target DNA, along with selection and an *E. coli* origin of replication (Chain et al., 2000; Kvitko et al., 2013). Transfer can be initiated at one of the oriT sites, and the target DNA can be conjugated into a recipient *E. coli* strain. Transfer is terminated by the second oriT, and the transferred DNA is assembled into a circular plasmid by homologous recombination between the two oriTs. A similar technique involves inserting a phage attachment sites (*att*P and *att*B) on either side of the gene cluster, along with selection and an *E. coli* origin of replication (Du et al., 2015). Upon activation of a phage integrase the region between the *att*P and *att*B sites is excised and circularised. This can then be recovered and transformed into *E. coli*. Whilst these techniques are impressively well designed, they suffer from a common issue that severely reduces their broad applicability. These techniques require successful transformation and integration of plasmids within the organism they are capturing the gene cluster from. This relies on methods of transformation being available, which is not always the case, and the organism's ability for homologous recombination, which can be highly variable.

To solve this issue, targeted methods are available which only require the original organism for genomic DNA extraction. Once the genomic DNA is extracted the original organism is not needed. These techniques use capture arms on a vector that are homologous to either end of the target DNA, called a capture vector (Figure 22). The capture vector is then linearised between the two capture arms. This capture vector can then be repaired with the target DNA, either by homologous
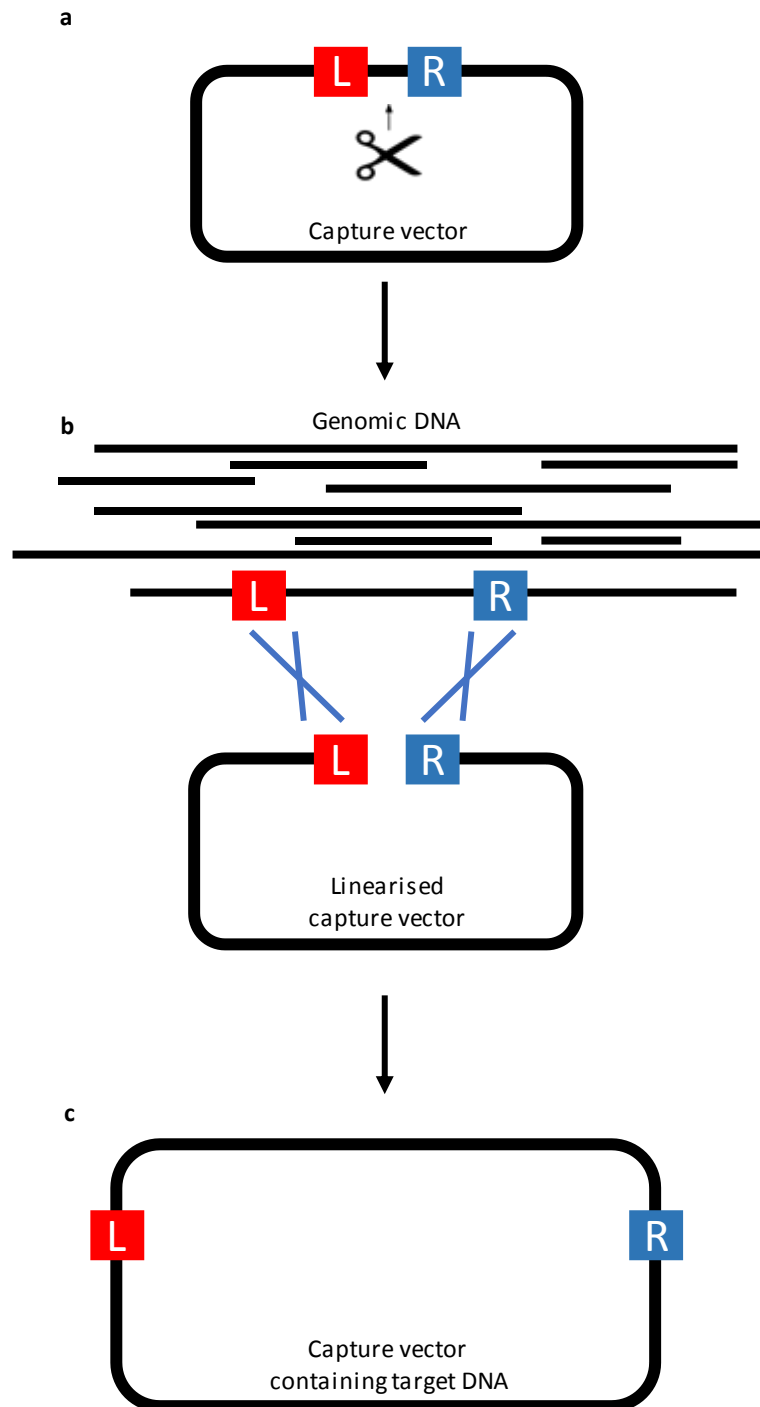
Figure 22. Generalised gene cluster capture from genomic DNA using homologous capture arms. **a.** The capture vector, containing left and right capture arms, is linearised. **b.** A target region of genomic DNA is used to repair the linearised capture vector by homologous recombination with the capture arms. **c.** This produces a circular capture vector containing the target DNA.

recombination *in vivo*, in *E. coli* or yeast (Fu et al., 2012; Kouprina and Larionov, 2016; Lee et al., 2015; Noskov et al., 2003), or alternatively *in vitro*, by Gibson assembly (Jiang et al., 2015). The yeast method, transformation-associated recombination (TAR) cloning, commonly includes counterselection against recirularisation of the capture vector (Noskov et al., 2003). The Gibson assembly method, and a more recently adapted TAR cloning method, uses CRISPR-Cas9 to digest the target DNA out of the genomic DNA (Jiang et al., 2015; Lee et al., 2015). This greatly increases capture efficiency. These techniques are a good compromise for intractable organisms, although they do not allow the capture of entire gene clusters from organisms that do not grow in laboratory conditions. To address this issue, TAR cloning has been shown to be able to capture fragments of gene clusters from environmental DNA, which allows for these fragments to be assembled back together into intact clusters (Kallifidas and Brady, 2012; Kim et al., 2010).

1.3.2.3.   *In Vitro* Techniques for Modifying DNA

Once a gene cluster has been captured in a vector, it can then be modified to help understand biosynthesis, increase yield, and modify the final product. As such, it is important to have robust methodologies for modifying vector-borne DNA. *In vitro* techniques for cloning, assembling, and modifying DNA are particularly attractive methodologies, as it is not limited by the growth speed of an organism or other complicated biological factors. There is a long history of manipulating DNA *in vitro*, starting with digestion and ligation of DNA fragments; a technique that is still a very commonly used method of assembling DNA (Cohen et al., 1973; Dugaiczyk et al., 1975). When paired with PCR it allows the assembly of cloned fragments of DNA with a speed and efficiency that means it has been used for around 30 years (Mullis and Faloona, 1987). Whilst this allows for the assembly of DNA fragments in new arrangements, the greatly diminishing efficiency as more fragments are assembled precludes this from being used in more complex multiple-fragment assemblies (An et al., 2010). Overlap extension PCR can also be used to first assemble fragments together (Horton et al., 1989), however this only slightly increases the usefulness of standard digestion and ligation for larger assemblies, as larger assemblies are difficult to PCR. The overlap extension PCR and recombination (OEPR) method also takes advantage of overlap extension PCR without using digestion and ligation, yet still suffers from the issue of needing to PCR amplify the entire assembly (Liu et al., 2017).

Some of the issues with standard digestion and ligation assembly have been addressed to increase the efficiency of more complex assemblies. For example, it is possible to use restriction enzymes that have different recognition sites, but produce complementary sticky ends (Cost, 2007; Shetty et al., 2011). This means that ligation to the correct target removes the recognition site, whilst unwanted ligation to the original fragment maintains the restriction enzyme recognition site. Therefore, the restriction enzyme continues to digest the incorrect ligations, whilst leaving the correct ligations. In doing this, the efficiency of simultaneous ligation of multiple fragments is greatly increased (Cost,

2007). This can also be used to allow sequential ligation of multiple fragments, exemplified by the BioBrick quick gene assembly (QGA) system (Yamazaki et al., 2017). The initial terminus of the DNA being assembled is first bound to magnetic beads, facilitating washing and replacing of buffers. Sequential iterative digestion and ligation of new fragments of DNA can then be achieved. In this example the two restriction enzymes that have different recognition sites but share sticky ends are SpeI and XbaI. The final assembly is selected for by PCR and size-selective gel purification; the resulting PCR fragment is then ligated into a plasmid and selected for in *E. coli* (Yamazaki et al., 2017).

The efficiency and scope of digestion and ligation-based methods can be improved with the use of type IIS restriction enzymes. There are many published methods that take advantage of this, for example Golden Gate (Engler et al., 2008, 2009), pairwise selection assembly (PSA; Blake et al., 2010), GoldenBraid (Sarrion-Perdigones et al., 2011), modular cloning (MoClo; Weber et al., 2011), methylation-assisted tailorable ends rational ligation (MASTER; Chen et al., 2013), versatile genetic assembly system (VEGAS; Mitchell et al., 2015), and modular idempotent DNA assembly system (MIDAS; Van Dolleweerd et al., 2018). Golden Gate cloning is the most widely-used of these, and so will be discussed further here.

Type IIS restriction enzymes cut outside of their recognition site (Aggarwal, 1995), allowing a single enzyme to leave different sticky ends, depending on where the recognition site is located (Lebedenko et al., 1991). This allows the assembly of multiple fragments in a single pot, as sticky ends can be designed to be non-palindromic (palindromic sticky ends can ligate to themselves) and only complementary to the target fragment's sticky ends. If the assembly proceeds correctly and ligation to the target fragment occurs, then the restriction enzyme recognition site is removed; if it re-ligates to the original fragment then the restriction enzyme recognition site is still present. This allows successive rounds of digestion and ligation at a single temperature in a single reaction mixture to increase the efficiency of the assembly (Figure 23; Engler et al., 2008). A significant issue with methods based on type IIS restriction enzymes is the occurrence of restriction enzyme recognition sites within the DNA being assembled. This can prevent larger assemblies, assemblies with non-standardised parts, or simply assemblies with an unlucky occurrence of recognition sites. This issue is exacerbated by the small number of type IIS restriction enzymes that are commercially available.

Over the past decade many techniques have been published that allow the *in vitro* assembly of DNA fragments with short homologous linkers installed by PCR. This avoids the use of restriction enzymes, removing the sequence limitations imposed by the restriction enzyme recognition sites. Some examples include sequence and ligation–independent cloning (SLIC; Li and Elledge, 2007), uracil-specific excision reagent (USER) DNA Engineering (Bitinaite et al., 2007), Gibson assembly (Gibson et al., 2009), In-Fusion (Sleight et al., 2010; Zhu et al., 2018), DNA assembly with thermostable

Key

■ Fragment to be assembled

■ Restriction enzyme recognition site

■ Restriction enzyme cut site
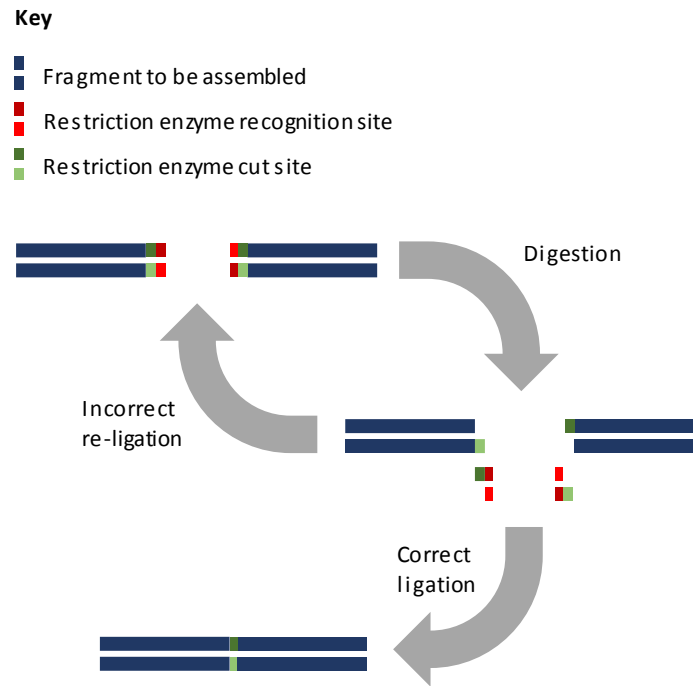
Figure 23. Golden Gate-mediated assembly of DNA (Engler et al., 2008).

exonuclease and ligase (DATEL; Wenwen Ding et al., 2017; Jin et al., 2016), and twin primer assembly (TPA; Liang et al., 2017).

Gibson assembly (isothermal assembly) is characteristic of the homology-based techniques, in which an exonuclease is used to remove one strand from the homologous linkers, leaving a large sticky end (Figure 24; Gibson et al., 2009). After annealing these together, a polymerase can fill in the gaps and a ligase closes the nicks. This is a particularly powerful technique and was used to assemble a synthetic *Mycoplasma genitalium* genome, 583 kb in size (Gibson et al., 2009). SLIC also uses homology and an exonuclease to generate complimentary sticky ends, however instead of a polymerase and ligase, the *E. coli* recombinase RecA is employed to assemble these together (Li and Elledge, 2007). It was also shown that SLIC could be achieved without the exonuclease if the fragment that was going to be assembled was amplified both with and without the regions of homology. These two products can be melted and annealed to each other. One half of the reaction mix would be fragments with single stranded homologous linkers. This was then used with the recombinase to be assembled in to the plasmid (Li and Elledge, 2007). TPA uses this principle of melting and annealing to produce sticky ends in an enzyme free manner, however it was also shown that the repair of the nicked plasmid could be performed in *E. coli*; removing the need for the recombinase that SLIC uses (Figure 24; Liang et al., 2017). Up to ten fragments, and up to 31 kb has been assembled using TPA. Whilst it is clear these techniques are powerful, both Gibson assembly and TPA suffer losses to their efficiency when assembling high GC DNA (Casini et al., 2013; Liang et al., 2017). This is a problem in
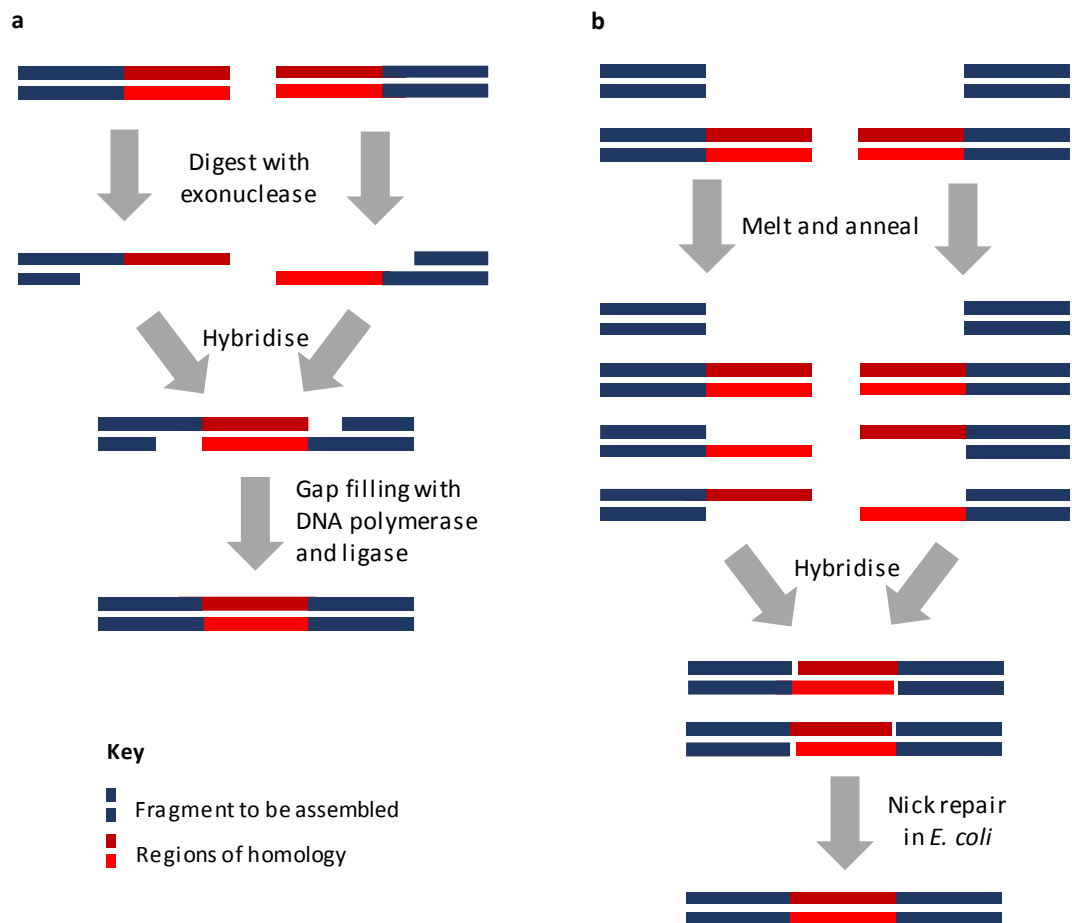
Figure 24. Schematic of **a.** Gibson and **b.** TPA assembly (Gibson et al., 2009; Liang et al., 2017).

editing natural product gene clusters from the actinomycetes, as these have high GC genomes (Bentley et al., 2002).

1.3.2.4.   *In Vivo* **Techniques for Modifying DNA**

Microorganisms have evolved many tools to modify and assemble DNA, for example homologous recombination is vital for DNA repair (Li and Heyer, 2008) and is even used for other key biological purposes, for example sex change in yeast (Haber, 2012; Hanson and Wolfe, 2017; Shore, 1997). These tools can be leveraged to make targeted modifications. The applications of these tools for this can be split into two categories: 1) *in situ* editing of chromosomal DNA in an organism, for example programmable restriction enzymes (see section 1.3.2.1), or 2) editing vector-borne DNA, where it can be used for an application elsewhere. As the application discussed here is modification and heterologous expression of gene clusters, then *in vivo* methods for editing vector-borne DNA are the focus. The key methods for this are summarised here.

It should be noted that *in vivo* and *in vitro* methods do not exclusively fall into one category. In fact, many of the "*in vitro*" methods require a step in *E. coli*, for example nick repair in TPA (Liang et al., 2017) and a single step of homologous recombination in OEPR (Liu et al., 2017). Additionally, all the

in vitro methods suggest selection and screening in *E. coli* following the assembly. Those techniques were still included in the *in vitro* section because the key steps of the assembly happen *in vitro*.

E. coli is not always thought of as a recombinogenic organism, because many laboratory strains of *E. coli* are recombinase deficient mutants; a desirable modification for enhancing plasmid stability (Bryant, 1988). However, *E. coli* can be proficient at recombination, and this has been shown many times to useful in assembling and editing DNA (Fu et al., 2012; Gust et al., 2003; Liu et al., 2003; Rivero-Müller et al., 2007; Zhang et al., 2000, 1998). A particularly well used method is PCR-targeting (Gust et al., 2003). This method is primarily used to delete regions of a plasmid, and the *E. coli* strain used for this is BW25113 carrying an inducible λ RED recombination enhancing plasmid. Homologous recombination is used to replace the target region in a vector with a resistance gene flanked by recombinase recognition sequences. Subsequent expression of the recombinase removes the resistance gene, leaving behind an 81 bp scar (Gust et al., 2003). Whilst these *E. coli*-based techniques are widely used, it is difficult to find examples of them being used to simultaneously make multiple changes, or to make a variety of changes. This is likely because a significant drop in efficiency is seen if the number of fragments being assembled is increased (Jacobus and Gross, 2015).

Bacillus subtilis is less commonly used as a tool than *E. coli*, yet it seems to be able to achieve much more complex assemblies by homologous recombination. It has successfully been used not only to clone the entire mouse mitochondrial chromosome (Yonemura et al., 2007), but also to assemble an entire rice chloroplast chromosome (Itaya et al., 2008). One of the most impressive assemblies presented during the construction of the rice chloroplast chromosome was of 16 fragments to make a 72.9 kb plasmid. This technique has been published for over ten years; and yet very few examples of it being used are reported.

Yeast has been extensively used as a tool for assembling and modifying DNA. Whilst the growth rate of yeast is two to three times slower than *E. coli* and *B. subtilis*, it is preferred for DNA assembly as it is much more capable at homologous recombination than *E. coli* and more widely developed than *B. subtilis*. Since it was first shown to be useful in assembling single genes into digested vectors (Raymond et al., 1999), it has now been shown to be able to assemble and maintain over 100 kb of high GC prokaryote DNA (Noskov et al., 2012), to be able to assemble an entire 592 kb *M. genitalium* genome (Gibson et al., 2008a, 2008b), and to be capable of assembling single-stranded oligonucleotides (Gibson, 2009). Yeast has also been applied to cloning and modifying natural product gene clusters. For example, it has been used to completely rebuild biosynthetic pathways into scaffolds (Pahirulzaman et al., 2012; Shao et al., 2009, 2013, Shao and Zhao, 2012, 2013). These techniques can, however, require multiple assembly steps.

To make modifications to a specific region of a gene cluster, some yeast-based methods employ a yeast-expressed CRISPR-Cas9 to digest the target regions, allowing modifications to be made around that region (Kang et al., 2016). Alternatively, in a method reminiscent of many of the *E. coli*-based recombination techniques, selectable markers can be inserted with modifications; this eliminates the need for digesting the gene cluster at the target location of modification (Montiel et al., 2015; Yamanaka et al., 2014). However, the limitations imposed by having to link modifications to a selectable marker greatly reduce the type of modification that can be made; often being limited to deletions and the insertions of promoters.

1.3.2.5.   Gene Cluster Expression

The previous sections should make it clear that there are many tools available to clone, assemble, and modify DNA. However, once a successfully cloned, assembled, and/or modified biosynthetic gene cluster is in a vector there is still at least one more step: the production from the gene cluster must be tested. It can often be attractive to try and test the expression of the modified gene cluster back in the wild type host of the cluster, as the codon usage will be correct, protein co-factors will be present, and it is capable of producing the metabolic precursors. However, the original, unmodified cluster will still be a present. This may mask the effects of any modifications made. This may not be an issue if the aim was to increase production, although the presence of a repressor in the wild type gene cluster could still cause problems. Alternatively, the wild type gene cluster could be replaced with the modified cluster. For example, the PCR-targeting protocol for making *Streptomyces* mutants suggests that the modified version of the gene cluster, contained on a cosmid, could be integrated back into the chromosome by homologous recombination with either end of the wild type gene cluster (Gust et al., 2003). This would replace the wild type gene cluster with the modified version of the gene cluster in a single step. However, this requires a selectable marker to be associated with the modifications made to the cluster.

An alternative to expressing modified gene clusters in the wild type host is to express them in heterologous hosts. The choice of heterologous host represents a trade-off between closely related organisms and model organisms (Galm and Shen, 2006). Organisms closely related to the wild type strain will often have the correct codon usage, protein co-factors may be present, may be capable of producing the metabolic precursors, and may even have similar regulatory networks. However, they also may not have established transformation, production, or extraction protocols. Model organisms may be more distantly related, and therefore may not have the correct codon usage, protein co-factors, metabolic precursors, or regulatory networks, but they will have established transformation, production, or extraction protocols. Additionally, model organisms will often also have genetic tools, such as promoters, that can be reliably used to modify the gene cluster. This trade-off must be assessed on a case-by-case basis, as no universal model organism has been

developed. This choice means that it is advisable to try multiple hosts to get the balance right. For example, in *Streptomyces spp.*, heterologous expression of biosynthetic gene clusters is performed regularly in a variety of hosts, such as *S. coelicolor*, *S. avermitilis*, *S. lividans*, *Streptomyces albus*, and *Streptomyces venezuelae*, with varying case-specific success (Baltz, 2010; Nah et al., 2017).

Some model organisms have been engineered in attempts to make them more suitable to express a wider variety of gene clusters. This engineering can take many forms. A comprehensive review is not appropriate here, but a couple of typical examples are presented. Some modifications are made to make transformations easier. For example, to allow transformants to be selected for, auxotrophic mutations were introduced into *Aspergillus oryzae* (Gomi et al., 1987). This has enabled it to become a common host for expressing fungal gene clusters (Heneghan et al., 2010; Lazarus et al., 2014; Pahirulzaman et al., 2012). Engineering of model hosts can also be done to provide precursor supply. For example, *E. coli* was engineered with a mevalonate pathway to enable the production of terpenoids (Martin et al., 2003). This allowed it to be used as a host for a gene cluster that produced a potential anti-tumour monoterpenoid (Gupta et al., 2015). *E. coli* has been extensively engineered in other ways for expression of natural product gene clusters, and this is reviewed elsewhere (Fang et al., 2017).

The most relevant engineered model organism to this thesis is *S. coelicolor*. Three engineered versions of this organism were presented as potential hosts for expression of gene clusters: M1146, M1152, and M1154 (Gomez-Escribano and Bibb, 2011). M1146 had four of its strongly expressed gene clusters deleted from its genome. This was done to help relieve stress on the cell during production conditions and to free up precursors for heterologous clusters. M1152 was generated from M1146 with point mutations in its RNA polymerase β-subunit, and M1154 has additional point mutations within ribosomal protein S12. These mutations within each of these genes have been shown to improve antibiotic productivity (Hu et al., 2002; Okamoto-Hosoya et al., 2000; Shima et al., 1996), thereby helping to activate clusters heterologously expressed within these strains. These strains have been used for the expression of a very large number of actinomycete gene clusters (Gomez-Escribano and Bibb, 2014), and M1146 is the primary heterologous host used in this thesis.

1.3.3.   Examples of RiPP Cluster Modifications

1.3.3.1.   Scaffold Engineering

The ribosomal origin of RiPP scaffolds provides a unique opportunity amongst natural products for engineering. Predictable modifications to the RiPP scaffold can be easily made by site-directed mutagenesis of the precursor peptide gene. This is an advantage over other natural product systems, such as NRPSs and PKSs, where modifying the structure of the scaffolds is difficult and requires an understanding of the modular synthases (Barajas et al., 2017; Brown et al., 2018). In RiPPs, the tolerance of tailoring enzymes to these scaffold modifications can limit engineering opportunities, however the ease of precursor peptide modification has still been exploited many times to produce a diverse array of engineered RiPPs.

Modifications to the core peptide have been made very commonly in lanthipeptides and head-to-tail cyclised RiPPs, as they appear to tolerate modifications particularly well (Amagai et al., 2017; Gu et al., 2018; Hetrick et al., 2018; Ruffner et al., 2015; Urban et al., 2017; Yang et al., 2018). However, there are also examples of extensive modifications in other classes, such as thiopeptides, and of a single amino acid change in a bottromycin variant (Hou et al., 2012; Young et al., 2012). Many of these techniques use combinatorial approaches with degenerate oligonucleotides to produce a large diversity of core peptides very easily, however screening these combinatorial libraries can be particularly difficult.

An interesting example worth noting is that of an enrichment process that facilitated efficient screening. A class II lanthipeptide gene cluster and a class I lanthipeptide cluster were assembled using isothermal assembly and were developed for yeast display and phage display respectively (Hetrick et al., 2018). The yeast display system allowed for selection of αvβ3 integrin binding molecules by using fluorescence activated cell sorting. The phage display system allowed for selection for lipid II binding, by using biotinylated lipid II and doing pull down experiments. The core peptides of the lanthipeptides were then randomly mutated using degenerate codons. The cell sorting and the biotin pull down screening methods enriched the displayed lanthipeptides that were not only tolerated by tailoring enzymes, but also capable of binding their targets. The key part of this process, is that enriched lanthipeptides were still associated with the DNA responsible for their production, either in the yeast cell or the phage. Because of this prior enrichment step, deep sequencing could be used to assess the core peptide sequences that led to molecules that had protein binding ability (Hetrick et al., 2018). Screening by sequencing is much higher throughput than metabolomic techniques, as a library can be sequenced in multiplex.

From an engineering perspective, a frustrating aspect of the ribosomal origin of RiPPs is the limitation to proteinogenic amino acids. As such, there has been recent effort towards producing RiPPs with

non-proteinogenic amino acids. This has been done primarily in lanthipeptides (Kuthning et al., 2016; Lopatniuk et al., 2017; Oldach et al., 2012; Zambaldo et al., 2017), however there are also examples in a cyanobactin (Piscotta et al., 2015), a sactipeptide (Himes et al., 2016), and a lasso peptide (Tianero et al., 2012). The current methods for this do not lend themselves towards combinatorial biosynthesis, as each unnatural amino acid incorporated requires the expression of a corresponding tRNA and aminoacyl-tRNA synthetase (Wals and Ovaa, 2014). Whilst this has been an established method in *E. coli*, very recently an example was published using this same technique in which non-proteinogenic amino acids were incorporated into cinnamycin in *S. albus* (Lopatniuk et al., 2017). One of the modifications also increased the antibiotic activity of the cinnamycin derivatives.

1.3.3.2.   Tailoring Engineering

A classic manipulation of RiPP tailoring is the deletion of tailoring genes. This is used as a standard way to understand the biosynthesis of the mature product; gene deletions are often the first experiments that are done when investigating novel biosynthetic pathways. The analysis of the bottromycin (**1**) gene cluster is a particularly well-investigated example of this, where seven tailoring enzymes were deleted using *in situ* homologous recombination (Crone et al., 2016). The modified gene clusters were then used for in-depth untargeted metabolomic analysis to assess the roles of each genes.

There are fewer examples where the removal of genes is done intentionally to make targeted changes to a pathway product. One such example is found during heterologous expression and modification of the cinnamycin gene cluster to include non-proteinogenic amino acids (Lopatniuk et al., 2017). Following TAR cloning of the gene cluster, and expression in *S. albus,* it was noted that a hydroxylation, whilst necessary for activity, was not being efficiently installed, resulting in the production of multiple products. As the activity of the molecule was not necessary for the study, the hydroxylase responsible for the modification was deleted using RedET recombination in *E. coli*. This simplified the cinnamycin-related metabolites being produced and increased production, likely because an active molecule was no longer being produced (Lopatniuk et al., 2017).

Adding new tailoring genes to RiPP biosynthetic pathways is often not as easy as just co-expressing a new tailoring gene. Tailoring genes usually recognise the leader peptides of RiPPs, using a RiPP precursor peptide recognition element (RRE; Burkhart et al., 2015). Therefore, new tailoring genes that are added to biosynthetic pathways have to recognise the leader peptide of the precursor peptide (Oman and van der Donk, 2010). This means that moving tailoring genes between closely related gene clusters is possible. For example, the YcaO and TfuA genes involved in thioamide bond installation in an *Micromonospora arborensis* thiopeptide were expressed in *Streptomyces laurentii* and successfully thioamidated the *S. laurentii*-produced thiostrepton (Schwalen et al., 2018).

The issue of needing the appropriate leader peptide for tailoring enzyme recognition has been addressed by using chimeric leader peptides. For example, a class I lanthipeptide-LAP hybrid molecule was produced by expressing a hybrid core peptide with a leader peptide that contained key elements from the nisin (**6**; lanthipeptide) precursor peptide and the heterocycloanthracin (LAP) precursor peptides (Burkhart et al., 2017a). This was expressed in *E. coli* with the tailoring enzymes responsible for lanthionine bond formation and heterocyclisation. The tailoring enzymes were able to recognise the hybrid leader peptide and process the hybrid core peptide. The same study also presented the production of a sactipeptide-LAP hybrid, a class II lanthipeptide-LAP hybrid, and a class II lanthipeptide-LAP hybrid with an epimerisation catalysed by an enzyme from yet-another gene cluster (Burkhart et al., 2017a). This has the potential to become a generally applicable technique to install new tailoring modifications on RiPPs.

1.3.3.3.   Refactoring

It is often desirable to engineer a gene cluster to increase the yield of the natural product. High yields can have multiple positive effects. They can: facilitate purification for bioactivity data or structural elucidations, prevent deleterious effects on production being a problem when modifications to the gene cluster are made, or remove complex media requirements for production. One brute force approach is to put a strong constitutive promoter in front of each gene, which was carried out with an entire cyanobactin gene cluster. Here, a T7 promoter was placed before every gene on complementary plasmids for expression in *E. coli* (Donia et al., 2006). This was also done during the heterologous expression of the lichenicidin gene cluster in *E. coli* where a remarkable eight steps of ligation and In-Fusion cloning placed each gene responsible for the production of the two-component lanthipeptide under the control of its own T7 promoter, resulting in increased production (Kuthning et al., 2015).

RiPPs in bacteria are often coded on one or very few operons. Therefore, a less brute force approach than complete refactoring is to place a strong constitutive promoter in front of each operon. This has successfully increased production in a LAP (Metelev et al., 2013), a head-to-tail cyclised RiPP (Amagai et al., 2017), a lanthipeptide (Aso et al., 2004), and a thiopeptide (Hayashi et al., 2014). The telomestatin gene cluster is contained on a single operon, and as such a single promoter is only needed to refactor it. However, three different promoters were installed before the cluster by λ-RED recombination and only one gave significant yield improvements (Amagai et al., 2017). This reinforces the fact that refactoring these seemingly simple clusters is still unpredictable.

Overexpression of a single gene can also be used to increase the production of RiPPs. There are obvious targets, for example the precursor peptide. Because the precursor peptide has, at best, a 1:1 ratio of translated protein to the mature RiPP, overexpression should increase the flux passing through the biosynthetic pathway, and therefore increase yield. This is seen with the overexpression

of the lasso peptide microcin J25, where the precursor peptide is placed under the control of an IPTG-inducible T5 promoter whilst the other genes remain under the control of their native promoter (Pan and Link, 2011; Piscotta et al., 2015). This does not, however, always work. In an effort to increase the expression of the thiopeptide GE2270 in the heterologous host, *S. coelicolor* M1146, the highest yield was achieved when a resistance gene was placed under a strong constitutive promoter (Flinspach et al., 2014). The level of production achieved was greater than when a strong promoter was placed in front of the entire gene cluster operon, and when a strong promoter was placed in front of the precursor peptide. This again reinforces how little is known about RiPP pathway regulation, and why the ability to make multiple versions of refactored clusters is important.

1.4. Thesis Objectives

1.4.1.   Natural Product Investigation Pipeline

1.4.1.1.   Pipeline overview

Investigations into biologically active classes of natural products often follow a common pipeline over time (Figure 25). First, they are discovered via bioactivity, metabolomics, or genome guided approaches. Following this, there are investigations into their bioactivity and their biosynthesis. Either simultaneously, or following this, applications of the discovery are often investigated. Perhaps the natural product has interesting bioactivity that gives it potential as a drug lead, interesting biosynthetic genes may provide new genome mining opportunities, or it may lead to new insights into the biology of the host organism. Throughout this process potential issues can be identified. Finally, cluster manipulation and/or medicinal chemistry can be used to address these issues and/or develop new applications. Investigations at all steps of this pipeline are interesting and worthwhile as standalone studies, however following this pipeline allows for a complete story about a class of natural products to be built up.

Not every RiPP class goes through every step of this pipeline, and many RiPP classes are at steps in the middle of this pathway. However, some more established RiPP classes have passed fully through this pipeline, for example the lanthipeptides. Following the discovery of nisin (**6**; Gross and Morell,
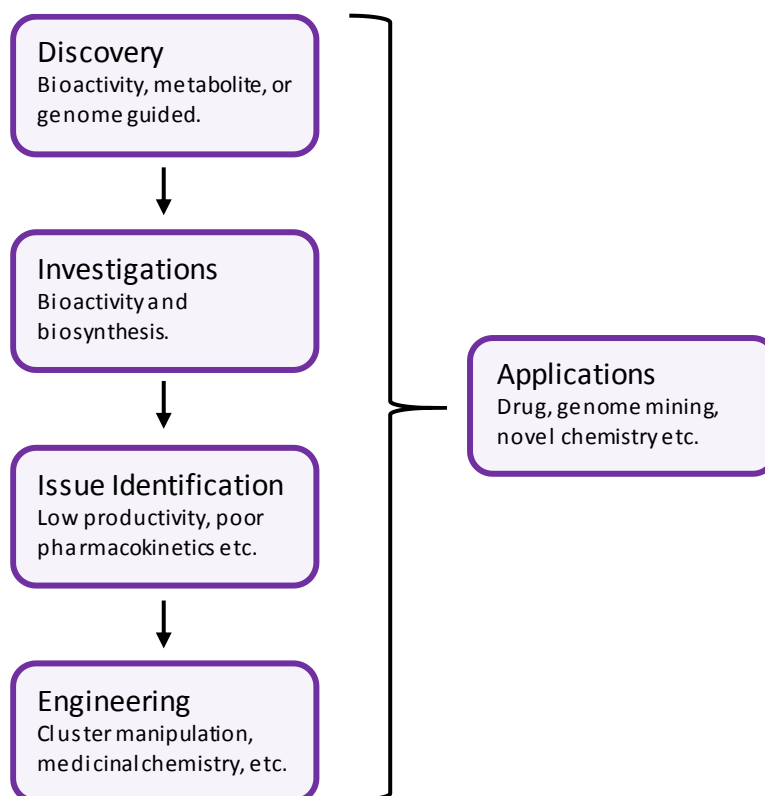
Figure 25. A generalised pipeline for investigations into biologically active natural product classes.

1971; Rogers, 1928) it subsequently found an application as a food preservative (Hansen and Sandine, 1994). The biosynthetic mechanisms and bioactivities of lanthipeptides have since been extensively investigated (Ongey and Neubauer, 2016; Repka et al., 2017; van der Donk and Nair, 2014) and many members of the class now have potential medicinal applications. These have been, and are likely still being, engineered by gene cluster manipulation and medicinal chemistry in an attempt to make them more suitable for those applications (Hetrick et al., 2018; Ongey and Neubauer, 2016; Urban et al., 2017; Yang et al., 2018).

Bottromycin (**1**) and thioviridamide-like molecules are at different steps of this pipeline. They are some of the rarest, most structurally complex, and poorly understood RiPPs, and they both also have very valuable potential applications, bottromycin (**1**) as an antibiotic (Miller et al., 1968) and thioviridamide-like molecules as chemotherapy agents (Hayakawa et al., 2006a). Therefore, they are both worth progressing through this pipeline. This is the aim of this thesis.

1.4.1.2.   Bottromycin in the Pipeline

Bottromycin (**1**) was discovered as a potent antibiotic over sixty years ago (Waisvisz et al., 1957), yet it failed in its application as a drug very early on (Miller et al., 1968). Investigations in to its bioactivity followed (Otaka and Kaji, 1983, 1981, 1976), and with the discovery of the gene cluster investigations in to the biosynthesis proceeded (Crone et al., 2012, 2016; Gomez-Escribano et al., 2012; Hou et al., 2012; Huo et al., 2012). One issue that was preventing bottromycin (**1**) from realising its application as a drug was identified, methyl ester hydrolysis in blood plasma (Kobayashi et al., 2010), and medicinal chemistry approaches have been used to attempt to solve this issue (Kobayashi et al., 2010; Yamada et al., 2018). Therefore, bottromycin (**1**) is currently on the last step of the pipeline: its biosynthesis is well understood, at least one of the issues preventing it becoming a drug have been investigated, and attempts have been made to engineer it to solve this issue. In this thesis, a novel approach is developed to aid with engineering bottromycin (**1**) biosynthesis and increasing pathway productivity. This has the ultimate goal of providing tools to help solve the issues preventing it from becoming a drug.

1.4.1.3.   Thioviridamide-like Molecules in the Pipeline

Thioviridamide-like molecules are at a much earlier stage in the pipeline than bottromycin (**1**). They have been discovered and their bioactivity has been briefly assessed (Frattaruolo et al., 2017; Hayakawa et al., 2006a; Izumikawa et al., 2015; Kawahara et al., 2018; Kjaerulff et al., 2017). Many of the molecules from this class have been shown to have potential applications as a drug, and their use as a drug without engineering has not been ruled out. However, their biosynthesis has not been investigated. Structurally, this class is rare and complex, which is likely to involve unusual biochemistry, so further applications in genome mining will be informed by characterisation of their

biosynthesis. Therefore, in this thesis the biosynthesis of a thioviridamide-like molecule, thiostreptamide S4 (**17**), is investigated.

1.4.2.   Gaps in Current Literature

1.4.2.1.   Technical Development

With the intention of engineering bottromycin (**1**) biosynthesis, it was clear that tools for cluster manipulation would be needed. Addressing the issues with the bottromycin (**1**) gene cluster would likely require multiple complex modifications. The bottromycin (**1**) gene cluster is a large RiPP cluster (over 17 kb), with multiple operons, and with high GC DNA. Many of the tools that have been used with RiPPs so far take advantage of the fact that RiPP clusters often contain very few genes in a single operon (see section 1.3.3). For manipulating larger clusters, current tools either: are not used efficiently, require too many assembly steps, are inappropriate for high GC DNA, or are not suitable for complex modifications (see section 1.3.2). These are the first issues addressed in this thesis. A method for modifying the bottromycin (**1**) gene cluster is developed, inspired by previous yeast-based techniques, that has a good balance between efficiency and complexity of modifications possible.

1.4.2.2.   Bottromycin

Bottromycin (**1**) is produced in low amounts and its methyl ester is hydrolysed in blood plasma (Huo et al., 2012; Kobayashi et al., 2010). Both issues are not conducive to it being used as a drug. In this thesis the first issue to be addressed is the low productivity. It was hypothesised that with the development and use of the aforementioned method for cluster manipulation the yield of bottromycin (**1**) could be increased with targeted modifications to the gene cluster. This would pave the way for further work derivatising bottromycin (**1**) in a biological system and finding variants resistant to hydrolysis in blood plasma.

1.4.2.3.   Thiostreptamide S4

The thioviridamide-like molecules, such as thiostreptamide S4 (**17**), have a unique structure amongst RiPPs (Hayakawa et al., 2006b). The modifications that have been previously seen are often not associated with the biosynthetic enzymes that normally install those modifications (Izumikawa et al., 2015). Yet beyond bioinformatic predictions, the biosynthesis of any thioviridamide-like molecules has not been elucidated. In this thesis, it was hypothesised that the biosynthesis could be investigated in a heterologous system using thiostreptamide S4 (**17**) as a model thioviridamide-like molecule (Frattaruolo et al., 2017). This could lead to applications in genome mining due to the potential biosynthetic novelty and would also lead to potential solutions to any issues encountered when trying to develop the thioviridamide-like molecules for applications as drugs.

Chapter 2 – Bottromycin

2.1. Introduction

2.1.1. Previous Work on Bottromycin

2.1.1.1. Discovery and Activity

Bottromycin (**1**; Figure 26) was discovered in the 1950s as an antibiotic compound that could be extracted from *S. bottropensis* (Waisvisz et al., 1957). Its ability to kill *Staphylococcus aureus* at low concentrations led to a series of papers testing its activity, derivatising it, elucidating its structure, and understanding its mode of action (Miller et al., 1968; Otaka and Kaji, 1983, 1981, 1976; Takahashi and Naganawa, 1976). Poor stability during *in vivo* trials and difficulties in elucidating its structure are likely reasons that caused bottromycin (**1**) to be dropped from active research.

Interest in bottromycin (**1**) was refreshed with a total synthesis study that finally confirmed its structure (Shimamura et al., 2009). Additionally, whilst it was proven that the instability of the molecule derived from hydrolysis of the methyl ester in blood plasma, a series of hydrolysis-resistant derivatives were also presented (Kobayashi et al., 2010). These promising results led to a rush to find the gene cluster responsible for the production of bottromycin (**1**). The bottromycin (**1**) gene cluster was published simultaneously by four groups, identifying the genes responsible in *S. bottropensis*
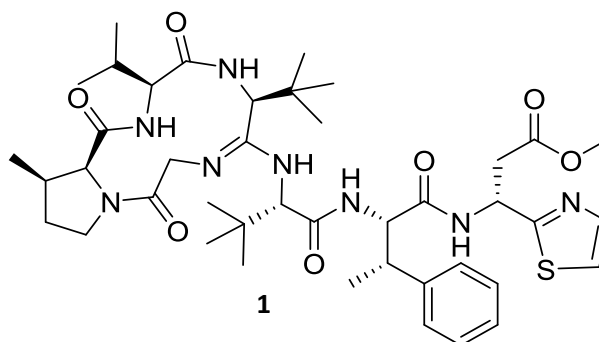
Figure 26. The structure of bottromycin (**1**; Shimamura et al., 2009).

(Gomez-Escribano et al., 2012), *Streptomyces* sp. BC16019 (Huo et al., 2012), *Streptomyces* sp. WMMB272 (Hou et al., 2012), and *Streptomyces scabies* (Crone et al., 2012). The bottromycin (**1**) gene cluster from *S. scabies* is the focus of this chapter.

2.1.1.2. Gene Cluster and Biosynthesis

Bottromycin (**1**) is a RiPP natural product. This means the scaffold (**20**; Figure 27) is encoded as a single gene, *btmD*, and this peptide is then post-translationally modified by proteins encoded by many of the other 12 genes in the gene cluster. The first steps in biosynthesis are a series of β-methylations catalysed by BtmC, BtmG, and BtmK (Crone et al., 2012, 2016; Huo et al., 2012), removal of the N-terminal methionine by BtmM (Crone et al., 2016; Mann et al., 2016), and thiazoline formation by the standalone YcaO, BtmE, to produce **21** (Franz et al., 2017; Schwalen et al., 2017). This is followed by amidine macrocycle formation and follower peptide cleavage, producing **22**. Initial evidence led to the conclusion that BtmF and BtmI would work together to form the macrocycle and BtmH would cleave the leader peptide (Crone et al., 2016), however recent evidence has revealed that BtmF works alone to form the amidine residue and therefore, either BtmH or BtmI cleaves the leader peptide (Franz et al., 2017; Schwalen et al., 2017). Finally, BtmJ catalyses oxidative decarboxylation of the thiazoline to form a thiazole, whilst employing dynamic kinetic resolution to
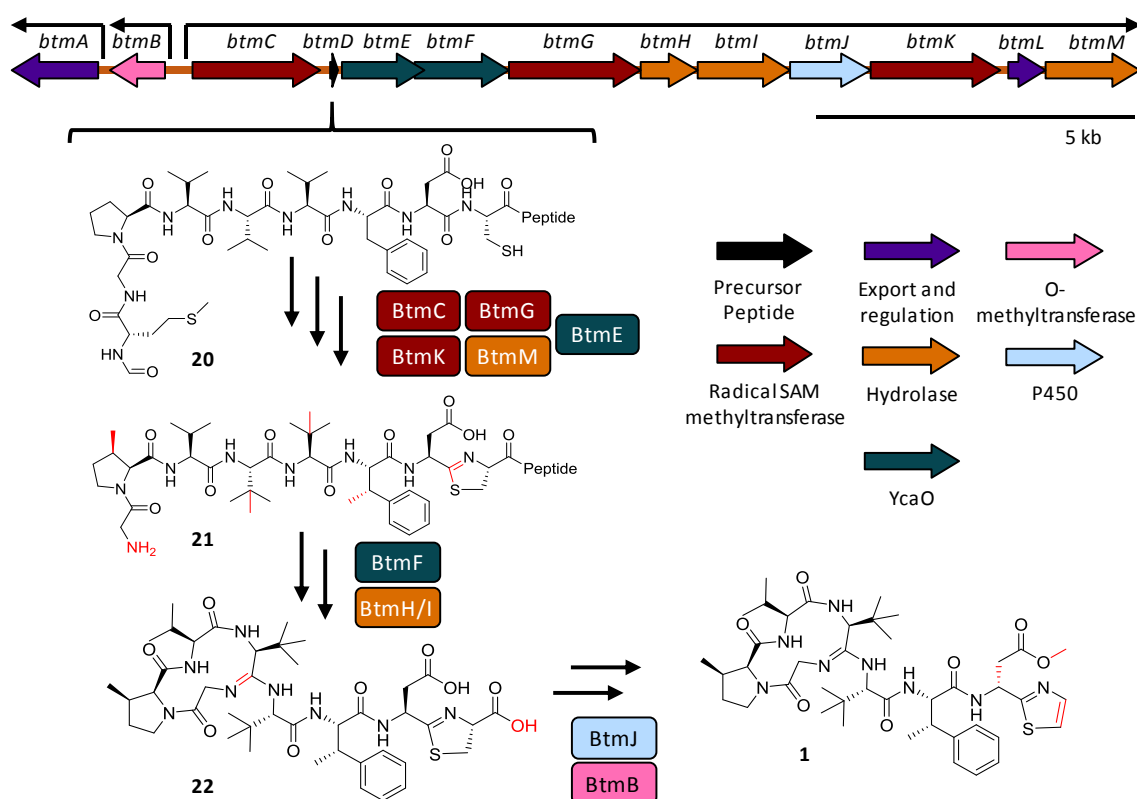
Figure 27. The *S. scabies* bottromycin (**1**) gene cluster, its operon structure, and a revised biosynthetic scheme (Crone et al., 2012, 2016; Franz et al., 2017; Huo et al., 2012; Mann et al., 2016; Schwalen et al., 2017). "Peptide" represents the follower peptide.

select for the spontaneously epimerised aspartic acid, and BtmB forms the aspartic acid methyl ester, leading to mature bottromycin (**1**; Crone et al., 2012, 2016; Huo et al., 2012).

Parallel to this project, other work in the Truman group performed by Natalia Miguel-Vior has focussed on understanding the native regulation of the bottromycin (**1**) gene cluster. This work has revealed that the bottromycin (**1**) gene cluster is organised into three separate operons, two one-gene operons and one 11-gene operon. It was also shown that whilst the pathway contains a transcriptional activator (BtmL), overexpression does not significantly increase production (Miguel-Vior et al., unpublished).

2.1.2.   Aims of this Chapter

2.1.2.1.   Hypothesis

Bottromycin (**1**) is produced in low quantities, but it has previously been shown that upregulating the cluster-specific transporter (BtmA) can increase the production (Huo et al., 2012). Therefore, it was hypothesized that this could be used in conjunction with upregulation of the other operons to increase bottromycin (**1**) production. This increase in production could then facilitate further derivatisation experiments.

2.1.2.2.   Objectives

There were three major objectives of this chapter:

1) Develop a tool to modify the bottromycin (**1**) gene cluster.

2) Use this tool to increase bottromycin (**1**) production.

3) Attempt further derivatisation.

2.2. Results and Discussion

2.2.1. Refactoring the Bottromycin Gene Cluster

2.2.1.1. Cloning the Bottromycin Gene Cluster

Given the in-depth assessment of different *in silico*, *in vitro*, and *in vivo* techniques available to modify gene clusters (see section 1.3.2), it was decided that modifying the bottromycin (**1**) gene cluster with a yeast-based technique would provide the best balance of speed, versatility, and efficiency. Whilst this can be an efficient way to edit large regions of high GC DNA, it requires the gene cluster to be captured into a vector. When this project was started, several gene cluster cloning methods had been reported (for full list see section 1.3.2.2). These include the capture from genomic DNA (gDNA) using transformation-associated recombination (TAR) cloning in yeast (Noskov et al., 2003; Yamanaka et al., 2014), using recombination in *E. coli* (Fu et al., 2012), and building and screening bacterial/P1 artificial chromosome (BAC/PAC) libraries (Alduina et al., 2003; Alduina and Gallo, 2012). There were a few requirements that had to be factored into the decision of how to capture the gene cluster. For example, the vector it was captured onto needed to be a yeast/*E. coli* shuttle vector. This would allow yeast to be used to modify the gene cluster and *E. coli* to be used to propagate the vector, as plasmid yields from yeast are very poor. After modifications had been made, the new gene clusters would need to be conjugated back into *Streptomyces* spp. for testing, so the shuttle vector would also need to contain appropriate features for conjugation and chromosomal integration. The pCAP01 vector used in yeast-based TAR cloning contains the necessary features for this and was constructed specifically for capturing gene clusters (Yamanaka et al., 2014). Therefore, it was decided that TAR cloning would be used for capturing the bottromycin (**1**) gene cluster. pCAP01 contains a tryptophan auxotrophic marker (TRP1) and has the ARSH4/CEN6 region to allow it to act as an artificial chromosome. It contains an pUC *ori* for replication in *E. coli* and a kanamycin resistance gene (aph(3)II) for selection in *E. coli* and *Streptomyces* spp. It also contains an origin of transfer (*oriT*), and ϕC31 integrase and integration site for conjugal transfer and chromosome integration in *Streptomyces* spp.

TAR cloning uses homologous recombination in yeast to repair a linear capture vector with a target region of gDNA (Figure 28). The yeast strain used is the highly recombinogenic VL6-48 (Kouprina et al., 1998). Capture arms on either end of the linearised capture vector specify which section of gDNA is used by yeast to repair the linearised vector, and therefore determine what is captured. Approximately 1 kb capture arms homologous to either side of the bottromycin (**1**) gene cluster from *S. scabies* DSM 41658 were designed. The gene cluster was analysed for putative terminators using the web-based tool WebGeSTer (Mitra et al., 2011). Terminators were predicted spanning up to 90 bp away from the stop codon of *btmA* and 149 bp away from the stop codon of *btmM*. Therefore,

the capture arms were designed to target 120 bp away from the *btmA* stop codon and 158 bp from the *btmM* stop codon. These capture arms were assembled into pCAP01 using recombination in yeast to produce the bottromycin (**1**) specific capture vector (Figure 28).
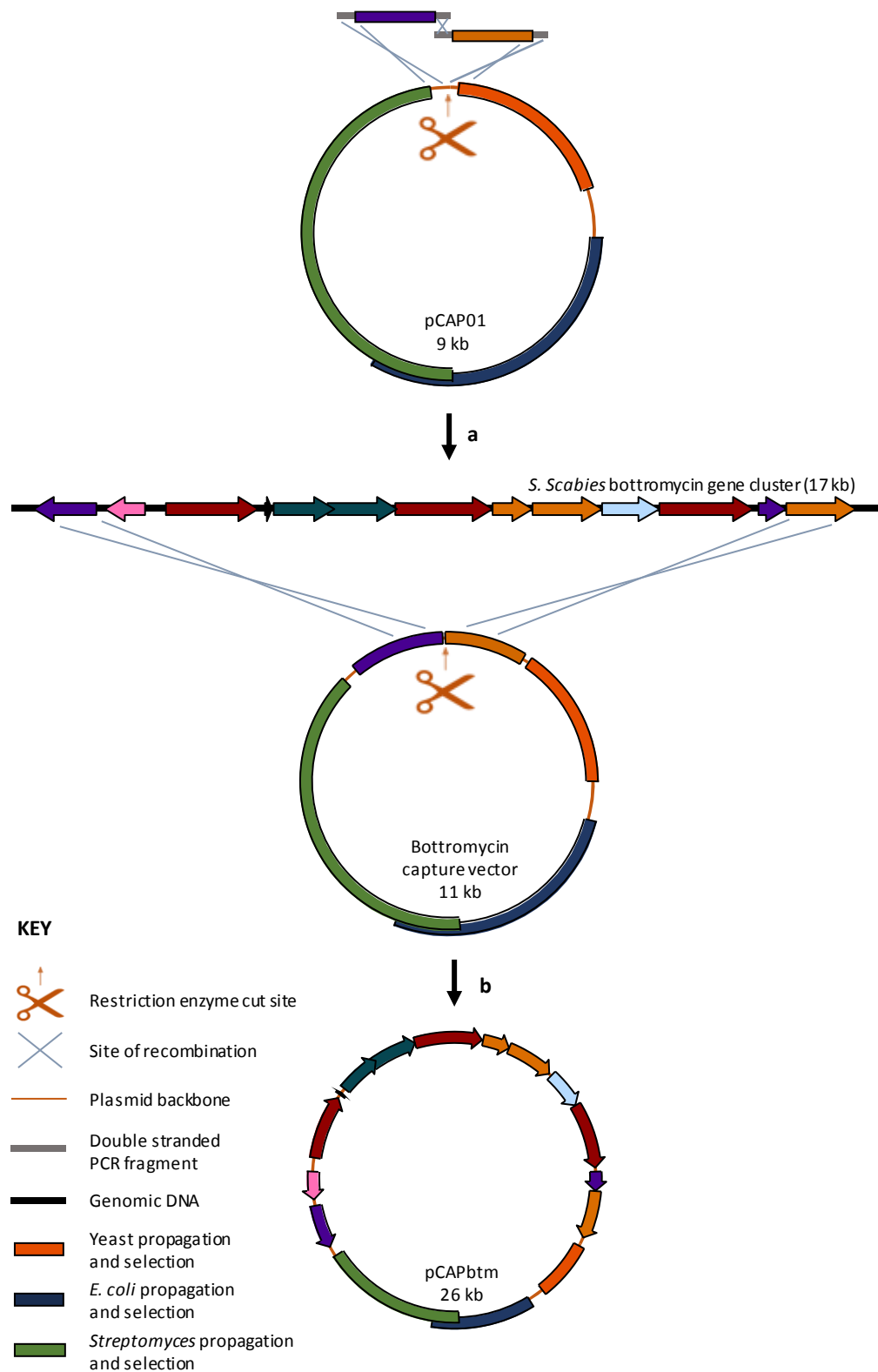
Figure 28. TAR cloning of the bottromycin (**1**) gene cluster. **a.** Capture arms are PCR amplified and assembled into pCAP01 using homologous recombination in yeast to produce a bottromycin-specific capture vector. **b.** The AvrII-digested capture vector is repaired by homologous recombination in yeast with the bottromycin (**1**) gene cluster.

To maximise the likelihood of successful TAR cloning, two different experimental parameters were varied in cloning trials. Firstly, two different methods of transformation into yeast were assessed: (a) spheroplast-mediated, as used in the pCAP01 TAR cloning protocol (Yamanaka et al., 2014), and (b) lithium acetate/polyethylene glycol (LiAc/PEG)-mediated transformation (Gietz and Woods, 2006). Spheroplast transformation is slower and more difficult but is more efficient for transforming larger DNA fragments, whilst LiAc/PEG-mediated transformation is quick and easy but is less efficient for transforming larger DNA fragments (Kawai and Murata, 2015). Secondly, cloning trials were carried out with both digested and undigested gDNA, as TAR cloning efficiency in other systems has been reported to increase if the genomic DNA is digested (Leem et al., 2003). The reduction in DNA size increases the chance the target DNA can enter a yeast cell, and DNA repair by homologous recombination in yeast is more efficient towards the end of DNA strands (Leem et al., 2003). The restriction enzymes BsrDI and PciI were chosen as they could excise the bottromycin (**1**) gene cluster on a 24 kb fragment from *S. scabies* gDNA. Therefore, four different TAR cloning conditions were tested: 1) spheroplast transformation of yeast with the capture vector and digested *S. scabies* gDNA, 2) spheroplast transformation of yeast with the capture vector and undigested *S. scabies* gDNA, 3) LiAc/PEG-mediated transformation of yeast with the capture vector and digested *S. scabies* gDNA, and 4) LiAc/PEG-mediated transformation of yeast with the capture vector and undigested *S. scabies* gDNA.

pCAP01 does not contain features to select against recircularization of the plasmid (Yamanaka et al., 2014), unlike some other TAR cloning systems (Noskov et al., 2003). Therefore, pCAP01-based TAR cloning typically has a high rate of false positives, and as such, extensive screening of yeast colonies is necessary. Four pools of 50 colonies each were tested from each condition (Figure 29) using multiplex PCR to check for the presence of the *btmC-D* intergenic region (~500 bp) and *btmK* (~100 bp). In total, 800 colonies were checked (Figure 29). For the spheroplast transformation with digested gDNA, all four pools contained at least one positive colony, whilst with undigested gDNA only one pool contained at least one positive colony. For the LiAc/PEG-mediated transformation with digested gDNA, three pools contained at least one positive colony, whilst with undigested gDNA no pools contained positive colonies. All 50 colonies from one of the pools from the LiAc/PEG-mediated transformation with digested gDNA condition were checked individually with the same PCR screen (Figure 29). Two of these colonies were positive for the bottromycin (**1**) gene cluster. Whilst not every colony in all the pools was checked, the numbers of positive pools shows that digesting the gDNA prior to TAR cloning greatly increased the TAR cloning efficiency for both methods of transformation.
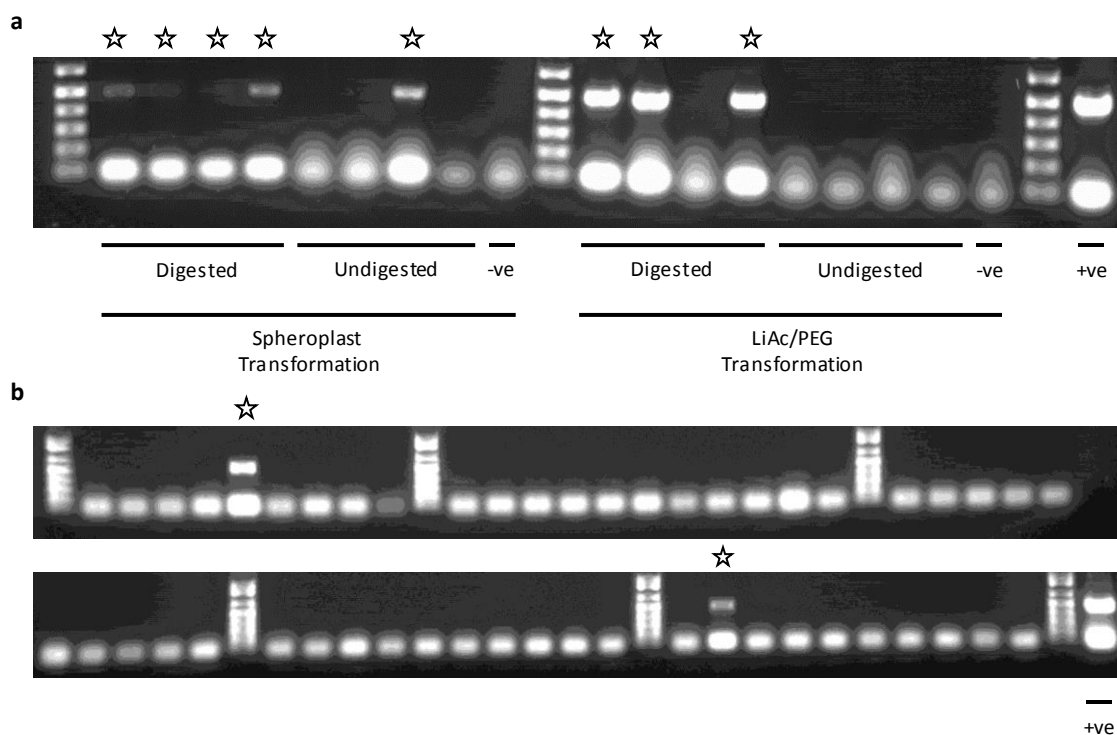
Figure 29. a. Multiplex PCR screening pools of 50 colonies from TAR cloning. Lanes marked with a star contain positive bands. b. Multiplex screening of all colonies from one positive pool.
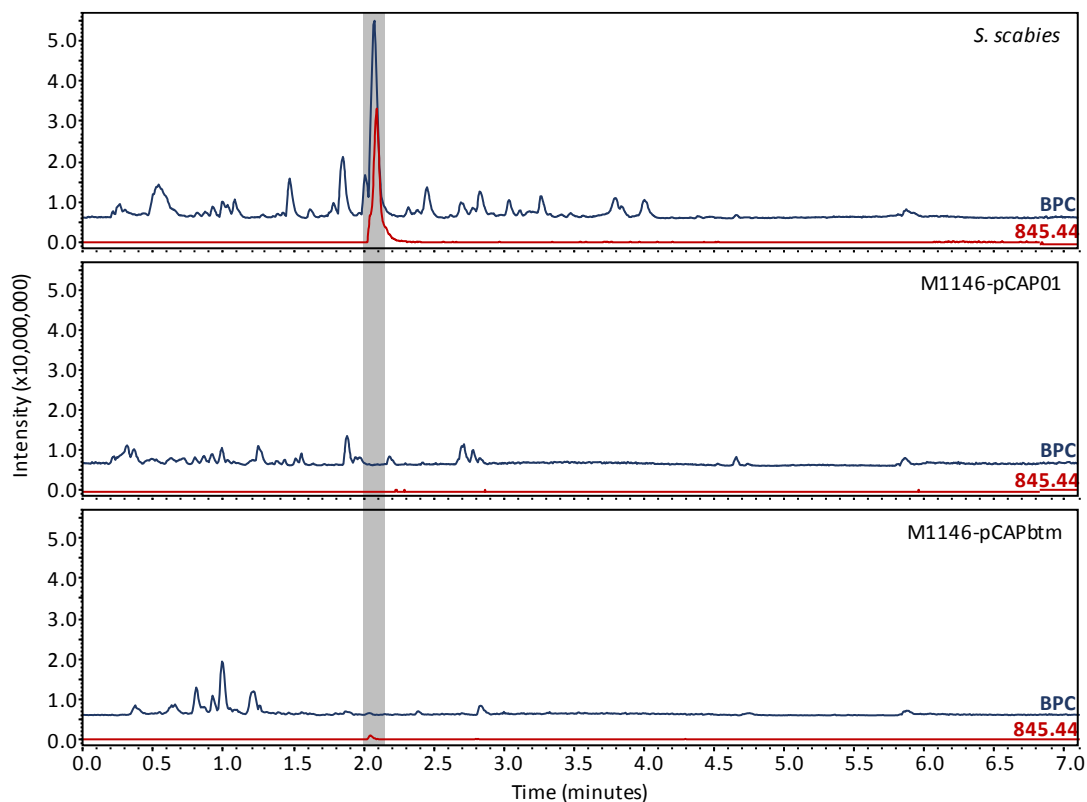
Figure 30. LC-MS analysis of the production of bottromycin (1; [M+Na]$^+$; m/z 845.44) in S. scabies, S. coelicolor M1146-pCAP01, and S. coelicolor M1146-pCAPbtm. The base peak chromatogram (BPC) is in blue, the extracted ion chromatogram (EIC) for bottromycin (1) is in red. The EIC for bottromycin (1) is multiplied by three for clarity.

TAR cloning efficiency was greater when using spheroplast transformation, which appeared to work with undigested gDNA at low efficiency.

The plasmid from one of the positive colonies was purified and passaged through *E. coli* for propagation. Restriction analysis confirmed that it carried the bottromycin (**1**) gene cluster. This vector, pCAPbtm, was conjugated into *S. coelicolor* M1146 using the methylation-deficient *E. coli* ET12567 carrying the helper plasmid pR9604. *S. coelicolor* M1146 was chosen as it has previously been validated as a good host for heterologous expression of RiPP gene clusters (Gomez-Escribano and Bibb, 2014). This produced the strain M1146-pCAPbtm. Bottromycin (**1**) production was tested in bottromycin (**1**) production conditions proven to work in *S. scabies* (Crone et al., 2012) and was assessed by liquid chromatography–mass spectrometry (LC–MS). Mature bottromycins were produced unreliably in barely detectable amounts (Figure 30).

2.2.1.2.   Assessment of Promoter Strength

The low levels of bottromycin (**1**) production in M1146 highlighted the requirement for refactoring of the gene cluster. As upregulation of the putative pathway specific activator gene (*btmL*) had little effect on bottromycin (**1**) production (Natalia Miguel-Vior, personal communication), it was decided that refactoring would be the best way to approach this. Unpublished work in the group (Natalia Miguel-Vior) shows that there are three main operons in the bottromycin (**1**) gene cluster; two one-gene operons, containing *btmA* and *btmB*, and then one eleven-gene operon, containing *btmC-btmM* (Figure 27). Rather than the complete refactoring systems exhibited in some fungal and bacterial systems (Pahirulzaman et al., 2012; Shao et al., 2013), just the promoters in front of each operon would be replaced, to maintain native operon structure. Therefore, three promoters were required. These promoters needed to be different, to avoid potential problems during refactoring, as yeast assembly is based on homology. To help inform the choice of promoters for each assembly, a β-glucuronidase (GUS) assay was undertaken to assess the strength of a series of promoters previously validated in *S. coelicolor*. In a GUS assay the promoter of interest drives expression of *gusA*, the β-glucuronidase gene. GUS breaks down its substrate to produce a colorimetric output that allows the inference of promoter strength. The promoters chosen were PSF14 (Labes et al., 1997; Rudolph et al., 2013), Paac(3)IV (hereinafter referred to as "Paac3"; Sherwood et al., 2013), PkasO* (Wang et al., 2013), PhrdB (Du et al., 2013), and PermE* (Bibb et al., 1994; Table 10). The native promoter responsible for expression of the putative btmC-btmM operon, PbtmC, was also tested as a point of comparison.

To account for any influence of the capture plasmid on gene expression, each promoter-*gusA* fusion was assembled using yeast in pCAP01. The resulting pCAP01-based *gusA* plasmids were tested using a qualitative plate assay in *S. scabies* and *S. coelicolor* M1146 (Figure 31). The standard bottromycin production medium (btmPM) was used so that the results would be directly relatable to production

Media only	No promoter	Media only
	PhrdB	
Paac3	PermE*	PSF14
	PkasO*	
Media only	PbtmC	Media only

24 hours   48 hours   72 hours   120 hours
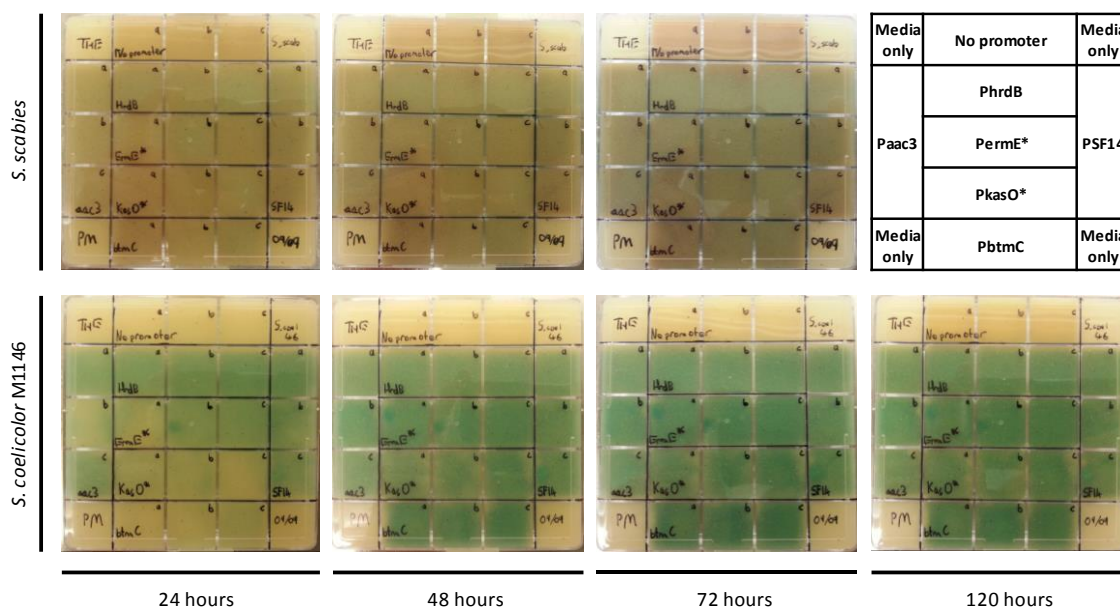
Figure 31. Solid GUS assay on bottromycin (**1**) production medium assessing the strength of different promoters in *S. scabies* and *S. coelicolor* M1146. Results are in triplicate, see key in figure for placement of strains.
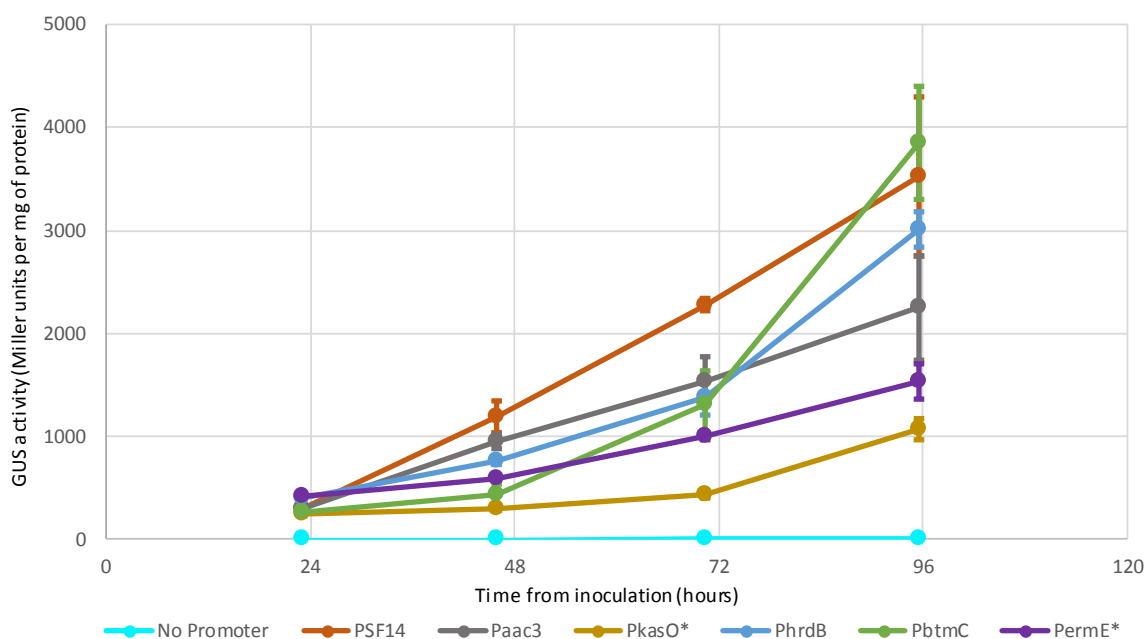
Figure 32. Quantitative GUS assay assessing the strength of different promoters in *S. scabies* and *S. coelicolor* M1146.

experiments (Crone et al., 2012). This indicated that *S. scabies* physiology is not conducive to solid GUS assays, as a brown pigment it produces masks the results. *S. scabies* growth is also significantly inhibited during the assay, leading to very poor sensitivity, and therefore very faint quantities of pigment were produced. The growth of *S. coelicolor* M1146 was inhibited during the GUS assay to a much lesser degree, so analysis of the results was still possible (Figure 31). The *S. coelicolor* M1146 assays showed that PhrdB and SF14 were initially the most active. However, after 48 hours all other

promoters appear equally strong, and relative activities could not be distinguished. The use of PbtmC provided surprisingly high GUS activity, considering the very poor productivity of the bottromycin (**1**) gene cluster in *S. coelicolor* M1146-pCAPbtm (Figure 31).

This experiment was followed up with a quantitative liquid assay to determine any differences between the promoters. To achieve time course data, growths were performed with 14 mL of btmPM in 50 mL Falcon tubes with 2 mL samples taken at periodic 24-hour intervals over four days. These were frozen, so GUS activity could be assessed after all samples had been taken. To assess GUS activity the substrate is added after cells are lysed, however poor data was obtained for *S. scabies* due to unreliable cell lysis. High-quality data were acquired for *S. coelicolor* M1146 (Figure 32). These data showed that, at most time points, the relative promoter strengths were PSF14>PhrdB>Paac3>PermE*>PkasO*. As in the preliminary plate assay, PbtmC is initially a relatively weak promoter but is as strong as PSF14 after 96 h.

2.2.1.3.   Proof of Principle – Btm*

It was hypothesised that unique restriction enzyme cut sites within the bottromycin (**1**) gene cluster could be exploited to make complex modifications to the cluster. This would allow the cluster to be rebuilt around double-strand breaks using yeast-mediated homologous recombination with the intended modifications. As an initial proof of principle for this technique, and to facilitate modification of the region of the bottromycin (**1**) gene cluster containing the promoters *btmA-C*, a BsrGI cut site was installed between *btmA* and *btmB* by homologous recombination in yeast. pCAPbtm was digested with NsiI and SpeI, cutting out *btmA* and *btmB*. This region was then rebuilt with two overlapping PCR products, containing the mutations to install a BsrGI cut site between *btmA* and *btmB*, by homologous recombination in yeast (Figure 33). Plasmids were extracted from four independent yeast colonies, passaged through *E. coli* for amplification, and tested by restriction analysis for a successful assembly reaction. Three of four plasmids checked were correct, whilst the fourth colony contained the original pCAPbtm plasmid. The successful assembly generated pCAPbtm*.

2.2.1.4.   Initial Refactoring – Btm1 and Btm2

pCAPbtm* provided both the proof of concept and the basis for rapid refactoring of the bottromycin (**1**) gene cluster. Its successful assembly showed that restriction enzyme cut sites could be used to modify regions of the gene cluster, and the installation of a BsrGI cut site would make the design of modifications to that region simpler. Rapid modification of the bottromycin (**1**) gene cluster would allow for a series of refactoring trials, with the results from each one informing the next.
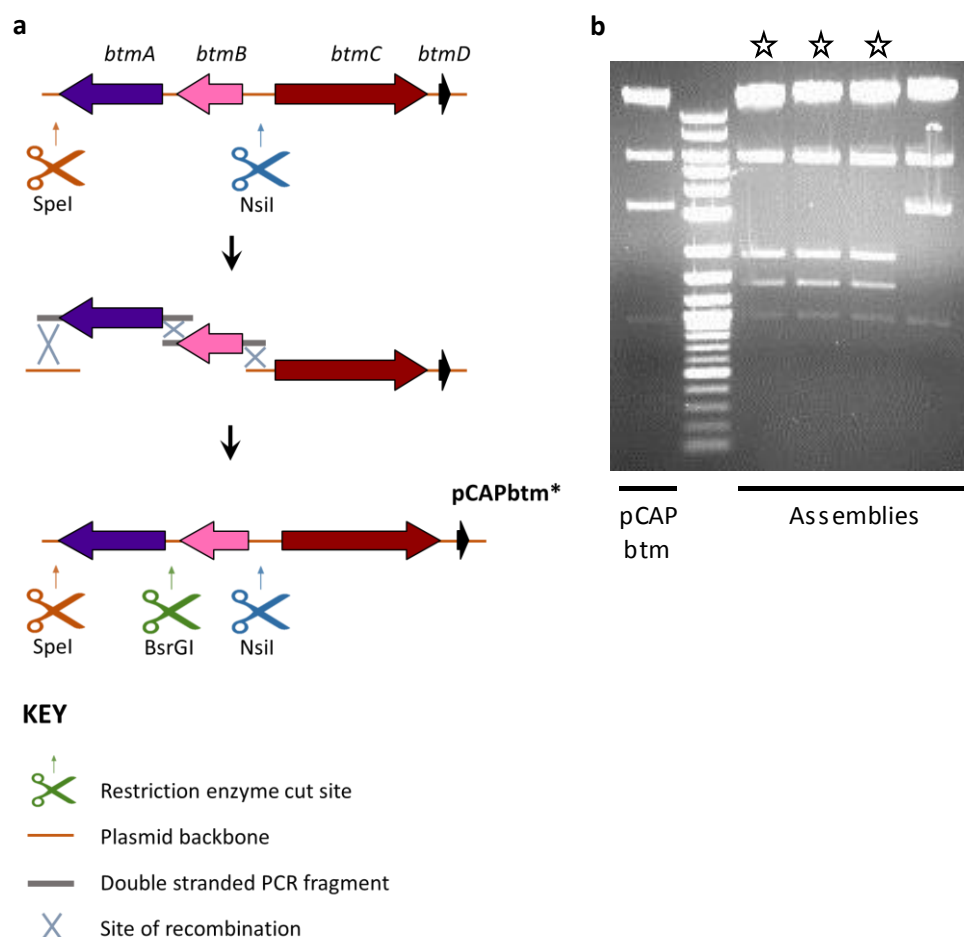
Figure 33. a. Schematic of the assembly of pCAPbtm* from pCAPbtm in yeast; Only the *btmA-btmD* portion of the gene cluster is shown for clarity. **b.** Restriction analysis of assemblies with BsrGI and HindIII. Lanes with digestion patterns indicative of correct assemblies are marked with a star. The ladder is NEB 2-log.

The first attempt at refactoring (Btm1) was to replace the native promoters with strong constitutive promoters. Ideally, this would improve productivity in multiple ways: upregulate the production of the precursor peptide, BtmD, and thereby increasing the metabolic flux through the pathway; upregulate the production of tailoring enzymes, allowing the pathway to deal with the increased flux; and increase expression of the exporter, BtmA, which is believed to protect the host cell from bottromycin (**1**) toxicity. Encouragingly, upregulation of the exporter gene has previously been shown to increase bottromycin (**1**) production on its own (Huo et al., 2012). To achieve this increased productivity, PSF14, PkasO* and PhrdB were chosen to be installed upstream of *btmA*, *btmB* and *btmC* respectively.

The BsrGI cut site was used in tandem with the NsiI cut site to modify the promoter regions of the three operons. The first refactored bottromycin (**1**) cluster, pCAPbtm1, was assembled in yeast using BsrGI/NsiI digested pCAPbtm* and four overlapping PCR products (Figure 34). This assembly was screened by PCR; four of the six colonies screened seemed to contain correctly assembled pCAPbtm1. The plasmid from one of the positive colonies was additionally confirmed by sequencing
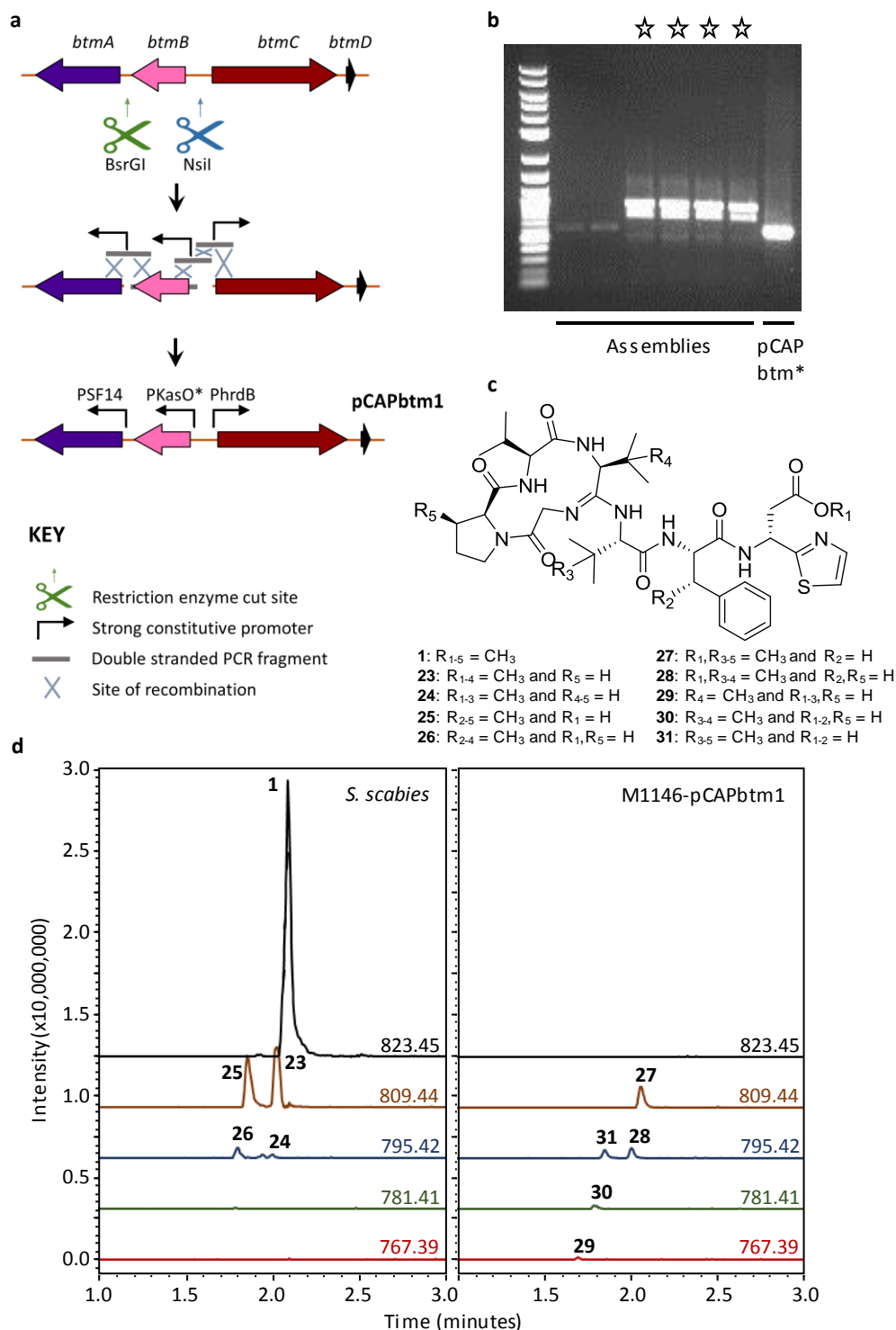
Figure 34. a. Schematic of the assembly of pCAPbtm1 from pCAPbtm* in yeast. Only the *btmA-btmD* portion of the gene cluster is shown for clarity. **b.** PCR screening of assemblies, lanes with band sizes indicative of correct assemblies are marked with a star. The ladder is NEB 2-log. **c.** Mature bottromycins seen. **1, 23-26** were previously identified in *S. scabies* (Crone et al., 2016), **27-31** had not previously been identified in *S. scabies*. **d.** LC-MS analysis of *S. scabies* and M1146-pCAPbtm1 production cultures. EICs of the five masses matching mature bottromycins are shown.

the assembled region (the promoter regions preceding *btmA*, *btmB*, and *BtmC*). This was a particularly encouraging result, as it validated that a single step assembly in yeast could be used to simultaneously make multiple changes to a gene cluster in a highly efficient way.
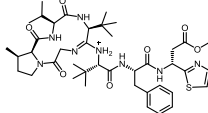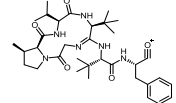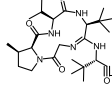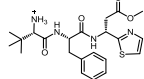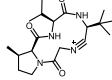
pCAPbtm1 was conjugated into *S. coelicolor* M1146 to produce the strain M1146-pCAPbtm1. Production from this strain was tested in liquid cultures in btmPM. However, LC-MS analysis indicated that no bottromycin (**1**) was produced. Bottromycin-like metabolites are characterised by a highly variable methylation pattern (Crone et al., 2016), and for the purposes of this project, molecules that contain a C-terminal thiazole ring and the amidine macrocycle are referred to as mature bottromycins (Figure 34). As **1** was not seen, a series of other possible mature bottromycins were screened for by LC-MS: *m/z* 823.45, 809.44, 795.42, 781.41, and 767.39 (Figure 34). These reflect the sequential lack of one methyl group starting from bottromycin (**1**; *m/z* 823.45).

Low levels of mature bottromycins were produced by M1146-pCAPbtm1, although all lacked a C-methyl group on phenylalanine (**27-31**; Figure 34), as determined by comparison of tandem mass spectrometry (MS^2) spectra with previously reported spectra (Crone et al., 2016, 2012). The high resolution MS^2 analysis of **27** is shown here (Table 1), whilst the analyses for other compounds are shown in the appendices due to the large number of figures and tables (Figure 97-Figure 99; Table 13-Table 18). The structures presented are based on the MS^2 data and the structure and stereochemical assignment for **1** (Shimamura et al., 2009). Whilst the MS^2 data is extensive, further NMR data would be required for absolute structural elucidation. NMR was not attempted due to the low yield of the compounds. **27-31** are all previously unidentified mature bottromycins. The lack of phenylalanine β-methylation indicated that BtmC, a class B radical SAM methyltransferase, was not functioning in M1146-pCAPbtm1, despite being encoded by the first gene in the *btmC-btmM* operon.

In case this was a result of the transcriptional fusion of the *btmC-btmM* operon with PhrdB, a newly refactored cluster was constructed (Btm2) to assess whether different promoters would result in active BtmC and improve pathway productivity. As part of this refactoring *btmB* was deleted, as BtmB-catalysed methylation of the aspartic acid residue is necessary for bottromycin (**1**) activity (Kobayashi et al., 2010). The resulting gene cluster would have the advantage of producing an inactive bottromycin, and therefore overcome any self-toxicity issues that could arise in a heterologous host. This methyl ester group is rapidly hydrolysed in blood plasma (Kobayashi et al., 2010), so it is unlikely that any clinically relevant derivatives of bottromycin (**1**) will carry this modification. The target refactored plasmid (pCAPbtm2) would therefore feature no *btmB*, as well as an alternative array of promoters: Paac3 upstream of *btmA* and PSF14 upstream of *btmC*. To achieve this in a single step, a modified technique for oligonucleotide assembly in yeast spheroplasts was used (Gibson, 2009). pCAPbtm2 was constructed from BsrGI/NsiI-digested pCAPbtm* and seven single stranded oligonucleotides using LiAc/PEG mediated transformation (Figure 35). Plasmids were extracted from five independent yeast colonies, passaged through *E. coli* for amplification, and tested by restriction analysis for a successful assembly. Four of the five plasmids checked were correctly assembled pCAPbtm2. One of these was additionally confirmed by sequencing of the

Table 1. High-resolution MS^2 fragmentation of molecule **27**.

Formula	Calculated *m/z*	Measured *m/z*	Mass difference (Da)	Proposed structure[a]
$C_{41}H_{61}N_8O_7S^+$	809.4378	809.4380	0.0002	
$C_{34}H_{51}N_6O_5^+$	623.3915	623.3917	0.0002	
$C_{25}H_{42}N_5O_4^+$	476.3231	476.3222	-0.0009	
$C_{22}H_{31}N_4O_4S^+$	447.2061	447.2057	-0.0004	
$C_{19}H_{31}N_4O_3^+$	363.2391	363.2392	0.0001	
$C_{16}H_{20}N_3O_3S^+$	334.1220	334.1218	-0.0002	
$C_{13}H_{22}N_3O_3^+$	268.1656	268.1654	-0.0002	
$C_7H_{11}N_2O_2S^+$	187.0536	187.0535	-0.0001	
$C_8H_{13}N_2O_2^+$	169.0972	169.0973	0.0001	
$C_7H_{13}N_2O^+$	141.1022	141.1027	0.0005	
$C_6H_4NOS^+$	138.0008	138.0012	0.0004	
$C_5H_{12}N^+$	86.0964	86.0965	0.0001	
$C_5H_{10}N^+$	84.0808	84.0812	0.0004	
$C_4H_8N^+$	72.0808	72.0811	0.0003	

[a] The top line of the table provides the MS data and structure for the parent molecule.

assembled region. As with pCAPbtm1, the production of bottromycins was assessed by LC-MS following heterologous expression in *S. coelicolor* M1146. Masses related to mature bottromycins were looked for (*m/z* 823.45, 809.44, 795.42, 781.41, and 767.39; Figure 35). As expected, M1146-

pCAPbtm2 generated mature bottromycins that all lacked the methyl ester group on the aspartic acid (**29-31**; Figure 35), but these compounds were also not β-methylated on phenylalanine by BtmC.
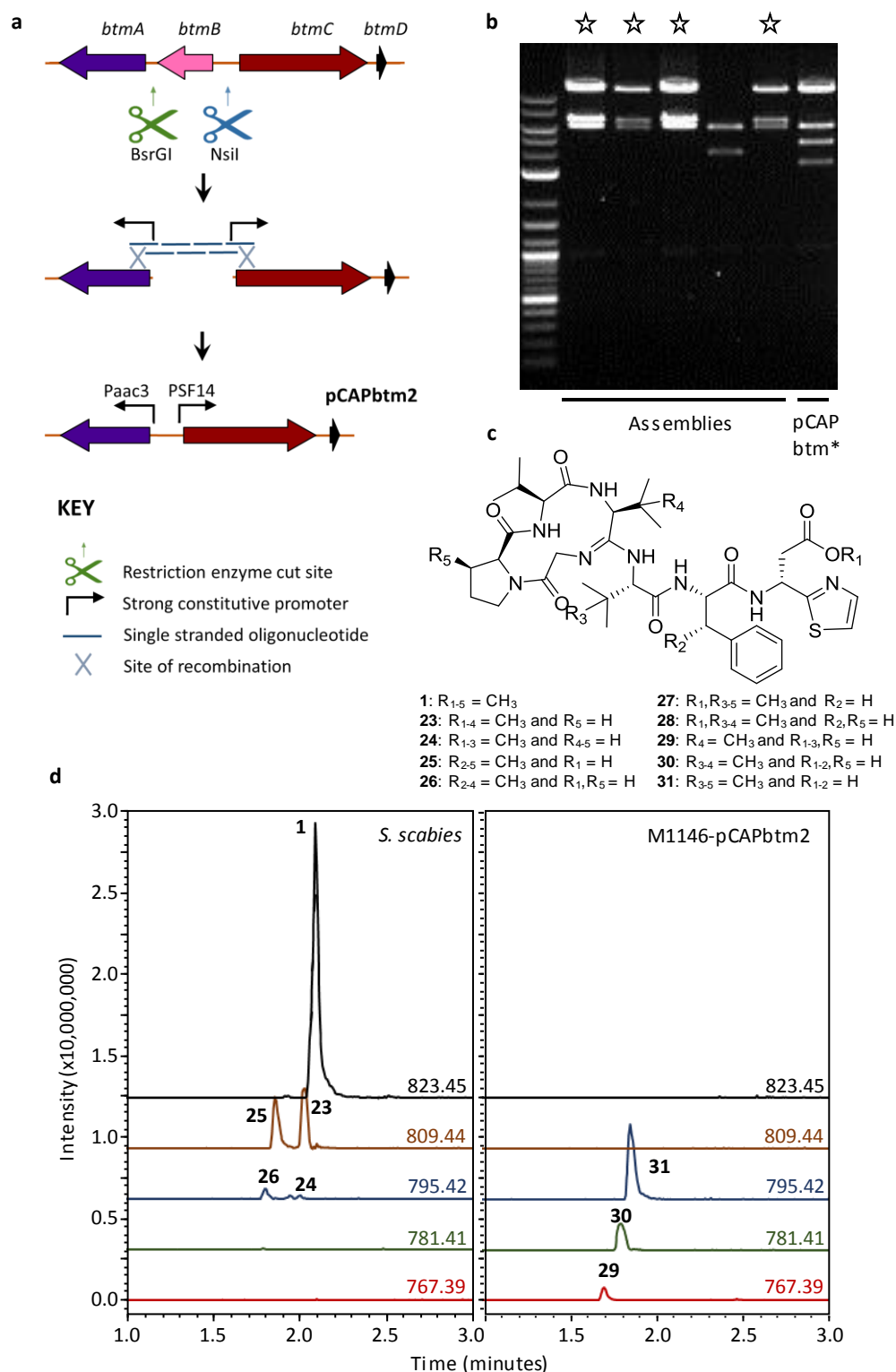
Figure 35. a. Schematic of the assembly of pCAPbtm2 from pCAPbtm* in yeast. Only the *btmA-btmD* portion of the gene cluster is shown for clarity. **b.** Restriction analysis of assemblies with XhoI and HindIII. Lanes with digestion patterns indicative of correct assemblies are marked with a star. The ladder is NEB 2-log. **c.** Mature bottromycins seen. **1**, **23-26** were previously identified in *S. scabies* by Crone et al. (2016), **27-31** had not previously been identified in *S. scabies*. **d.** LC-MS analysis of *S. scabies* and M1146-pCAPbtm2 production cultures. EICs of the five masses matching mature bottromycins are shown.

2.2.2. Understanding the Lack of Active BtmC

2.2.2.1. Metabolomic Analysis

Heterologous expression and gene cluster refactoring studies routinely focus on the final product of a pathway, yet this might not accurately reflect the total productivity of that pathway. When the base peak chromatograms (BPCs) of M1146-pCAP01, M1146-pCAPbtm1, and M1146-pCAPbtm2 were compared it was clear that just looking at the mature bottromycins was providing an incomplete story (Figure 36). Therefore, we employed untargeted metabolomics to identify all metabolomic changes between M1146-pCAPbtm1, M1146-pCAPbtm2, and M1146 containing the empty pCAP01 vector (Appendix Table 12). To identify bottromycin-related metabolites only found in the heterologous host, the metabolomic profile of *S. scabies* was also assessed in parallel. Comparative analysis of LC-MS data enabled the identification of all metabolites only produced by the refactored gene cluster.

To complement this analysis, mass spectral networks were constructed (Nguyen et al., 2013; Wang et al., 2016). The online software used to construct these networks (GNPS; https://gnps.ucsd.edu/) assesses every metabolite fragmented in LC-MS[2] screening and then links these metabolites based on similarities in the fragmentation. This enabled a network of bottromycin-related metabolites to be built, which aided with metabolite identification (Figure 37). These compounds were manually characterized by detailed MS[2] analysis (Figure 97-Figure 106), and any new compounds were subjected to high resolution MS[2] analysis (Table 13-Table 27). Their structures are proposed based on the in-depth MS[2] analysis and the structural relationships indicated by networking. In total, we
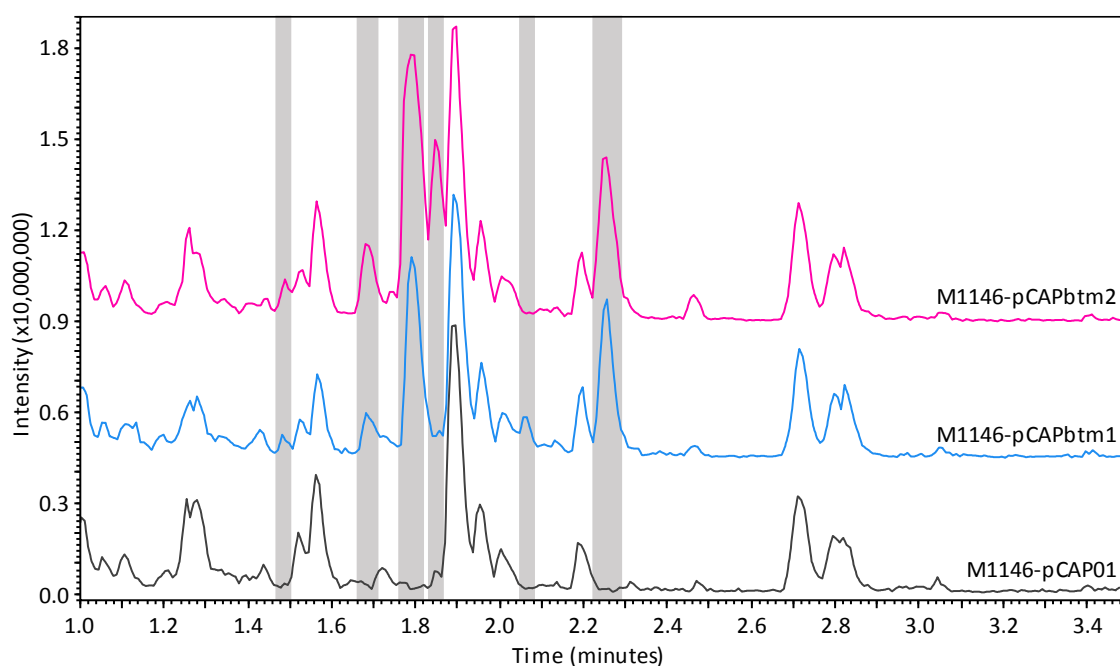
Figure 36. LC-MS analysis of *S. scabies,* M1146-pCAPbtm1, and M1146-pCAPbtm2. Aligned BPCs are shown, zoomed to show the key part of the chromatogram. Grey columns highlight the obvious differences between the strains.
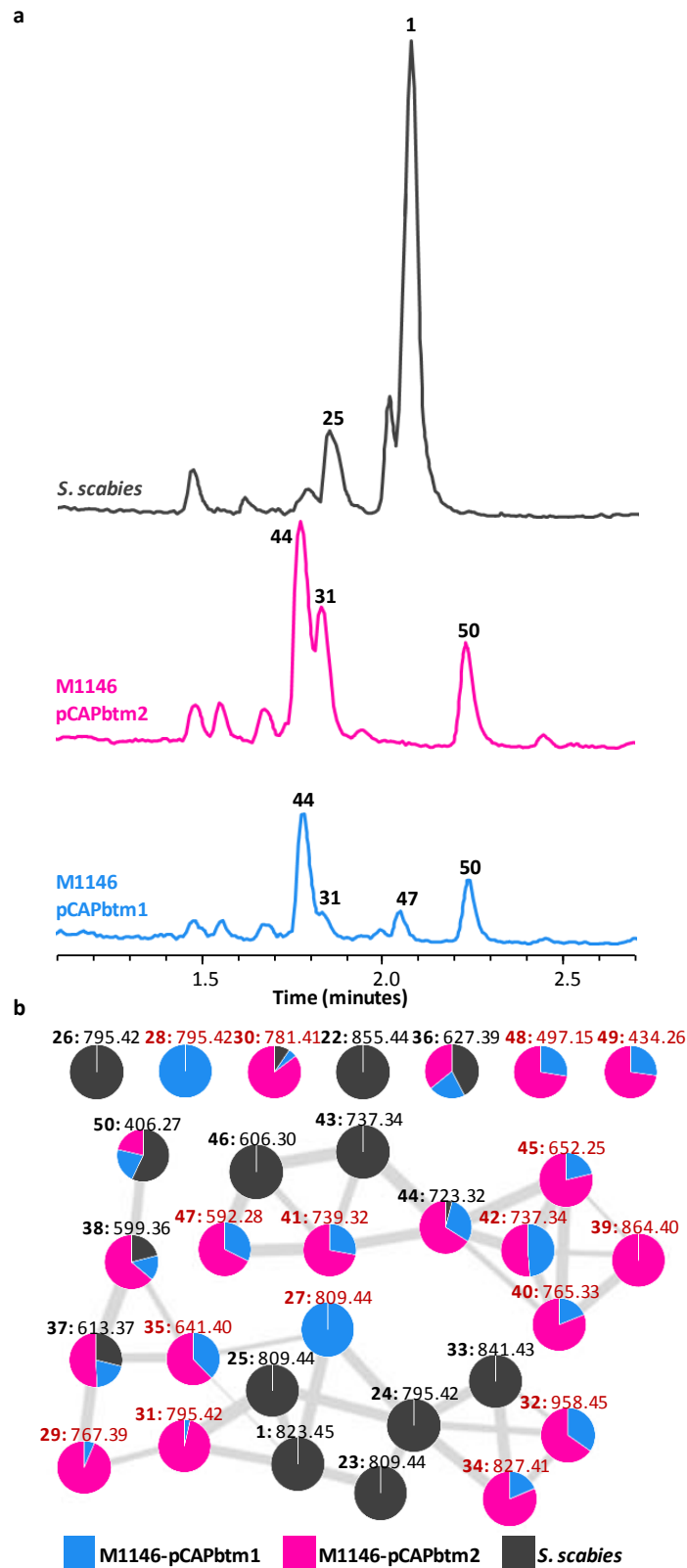
Figure 37. Comparison of bottromycin pathway productivity. **a.** EICs showing all detectable bottromycin-related metabolites from *S. scabies*, *S. coelicolor* M1146-pCAPbtm1 and M1146-pCAPbtm2. The metabolites responsible for the most intense peaks in each spectrum are numbered. **b.** A MS network of the bottromycin-related metabolites produced by M1146-pCAPbtm1, 2, and *S. scabies* in blue, pink, and grey, respectively. Each node represents a single metabolite, with the detected *m/z* listed. Node pie charts indicate the relative abundances of each metabolite between strains. Metabolites identified, but not detected by MS networking, are shown as unconnected nodes. Red node labels indicate molecules identified for the first time in this work.

1: R$_{1-5}$ = CH$_3$

23: R$_{1-4}$ = CH$_3$ and R$_5$ = H

24: R$_{1-3}$ = CH$_3$ and R$_{4-5}$ = H

25: R$_{2-5}$ = CH$_3$ and R$_1$ = H

26: R$_{2-4}$ = CH$_3$ and R$_1$,R$_5$ = H

27: R$_1$,R$_{3-5}$ = CH$_3$ and R$_2$ = H

28: R$_1$,R$_{3-4}$ = CH$_3$ and R$_2$,R$_5$ = H

29: R$_4$ = CH$_3$ and R$_{1-3}$,R$_5$ = H

30: R$_{3-4}$ = CH$_3$ and R$_{1-2}$,R$_5$ = H

31: R$_{3-5}$ = CH$_3$ and R$_{1-2}$ = H

32

22: R$_{1-2}$ = CH$_3$

33: R$_1$ = CH$_3$ and R$_2$ = H

34: R$_{1-2}$ = H

35: R$_{1-3}$ = CH$_3$

36: R$_{1-2}$ = CH$_3$ and R$_3$ = H

37: R$_2$ = CH$_3$ and R$_1$, R$_3$ = H

38: R$_{1-3}$ = H

39

40

41

42: R$_2$ = CH$_3$ and R$_1$ = H

43: R$_1$ = CH$_3$ and R$_2$ = H

44: R$_{1-2}$ = H

45

46: R$_1$ = CH$_3$

47: R$_1$ = H

48

49

50
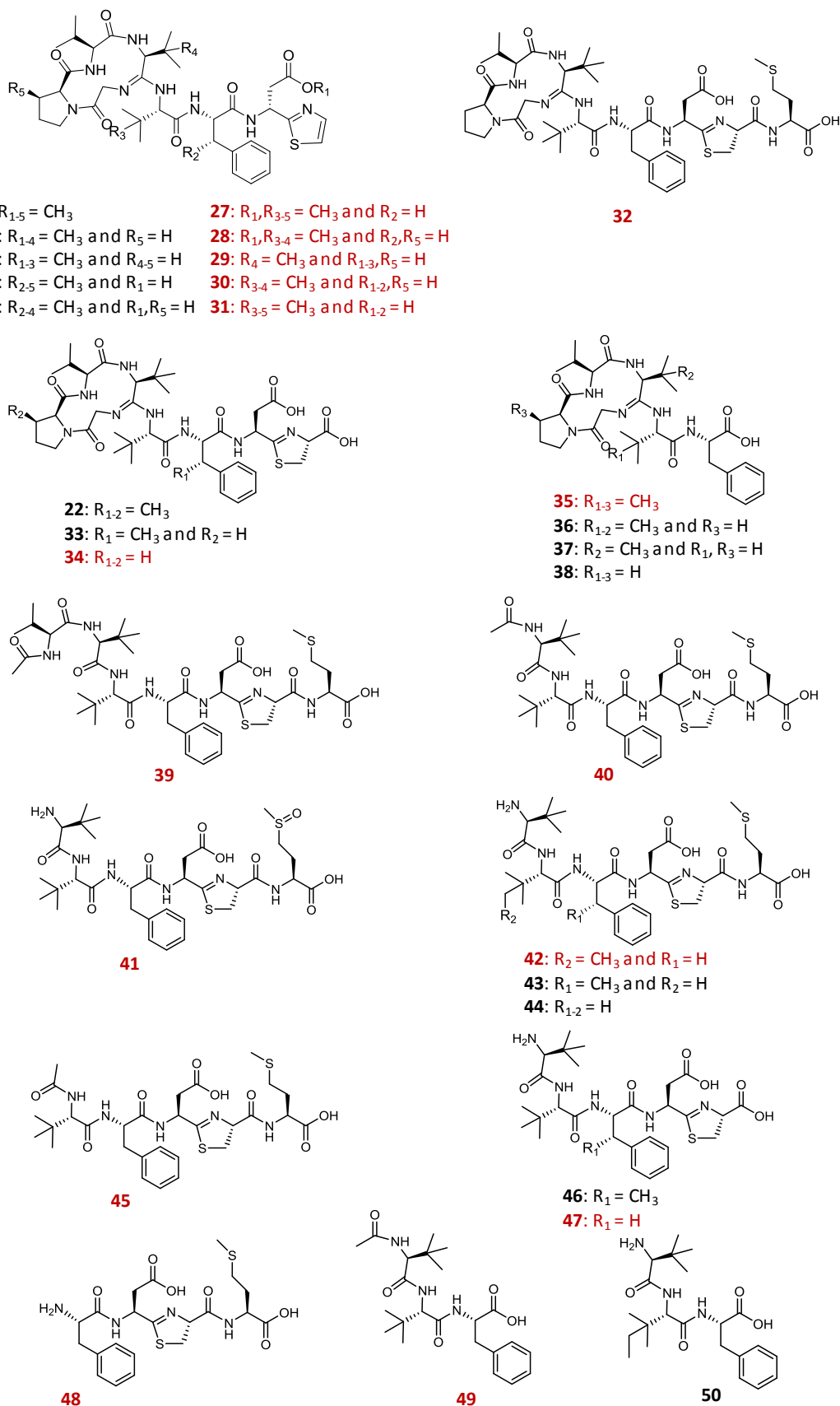
Figure 38. Proposed structures for **1**, **22-50** based on the in-depth MS^2 analysis and the structural relationships indicated by networking. Molecules new to *S. scabies* are labelled in red.

identified 16 new bottromycin-related metabolites by MS analysis (**27**-**31**, **34**-**35**, **39**-**42**, **45**, **47**-**49**), including the aforementioned five new mature bottromycins (compounds **27**-**31**; *m/z* 809.43, *m/z* 795.42, *m/z* 767.39, *m/z* 781.41, and *m/z* 795.42 respectively; Figure 38).

This allowed for a thorough analysis of pathway productivity. The peaks detected in LC-MS for each adduct of each metabolite ([M+H]$^+$, [M+2H]$^{2+}$, and [M+Na]$^+$) were integrated and summed to produce the total MS peak area for all bottromycin-related metabolites. There are flaws associated with quantification based on MS, for example different compounds have different ionisation efficiencies, and are therefore detected with different sensitivities. As all the metabolites compared here are related this is less of an issue, so relative comparisons in yield are still possible. Absolute quantification using this data is currently impossible.

Encouragingly, this analysis showed that a much larger quantity of bottromycin-related metabolites was produced in comparison to M1146-pCAPbtm, resulting in an overall 20-fold increase in M1146-pCAPbtm1 and a 60-fold increase in M1146-pCAPbtm2 (Figure 39) of total MS peak area for these metabolites. The expected C-methyl group on phenylalanine was absent in every detectable bottromycin-related metabolite in M1146-pCAPbtm1 and M1146-pCAPbtm2. Notably, it has been shown that deletion of *btmC* in *S. scabies* abolishes production of mature bottromycins (Crone et al., 2016), which indicates that phenylalanine C-methylation is critical for the efficient activity of downstream enzymes in the native producer. It can be hypothesised that the lack of active BtmC creates a series of bottlenecks that M1146-pCAPbtm1 and M1146-pCAPbtm2 are able to overcome using the strong expression driven by the heterologous promoters, providing an increased metabolic flux through the pathway and thus leading to the suite of new bottromycin-related metabolites.
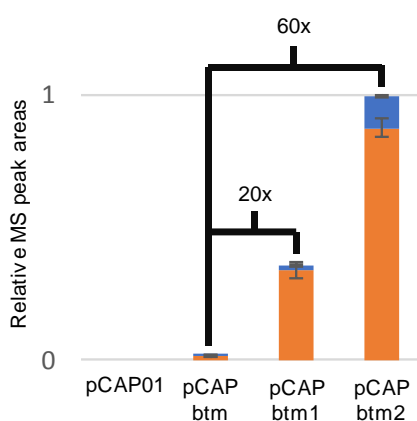
Figure 39. Production of mature bottromycins (blue) and other bottromycin-related metabolites (orange) from refactored bottromycin (**1**) gene clusters expressed in *S. coelicolor* M1146, based on MS peak area.

2.2.2.2.   Understanding the Bottlenecks

As the biosynthesis of bottromycin (**1**) has been very well characterised (Crone et al., 2016), it is possible to look at this large number of metabolites and make hypotheses as to which steps of the pathway are inhibited by the absence of phenylalanine C-methylation in the refactored clusters. Comparisons between the *S. scabies* wild type production and that in M1146-pCAPbtm2 are more appropriate than in other clusters due to the similar total production of bottromycin-related metabolites (Figure 40). Therefore, the following comparisons of intermediates will focus on this refactored cluster.

The biosynthesis of bottromycin (**1**) starts with BtmM-catalysed N-terminal methionine hydrolysis, BtmE-catalysed thiazoline formation, and BtmC, BtmG and BtmK-catalysed β-methylation (Crone et al., 2016). The N-terminal methionine is not seen in any metabolites, suggesting that the N-terminus of the precursor peptide seems to be particularly susceptible to proteolysis by endogenous proteases when the macrocycle is not present. This means that linear peptides without the macrocycle are always lacking most of the macrocycle amino acids (**39-47**, **49**, **50**). Therefore, very little can be predicted about the efficiency of the BtmM-catalysed N-terminal methionine hydrolysis. However, BtmM can function *in vitro* on an unmodified peptide, so it is unlikely that it is affected when the phenylalanine β-methylation is not installed (Crone et al., 2016; Mann et al., 2016). No conclusions can be made about the efficiency of the BtmE-catalysed thiazoline formation. The BtmG and BtmK-catalysed β-methylations appear to happen normally; both M1146 expressing the refactored cluster and *S. scabies* produce a variety of bottromycins with variable methylation patterns on the proline and valines.

Following N-terminal methionine hydrolysis, thiazoline formation, and β-methylation, the next step in bottromycin (**1**) biosynthesis is BtmF-catalysed macrocycle formation (Franz et al., 2017; Schwalen et al., 2017). In M1146-pCAPbtm2 more metabolites in which the macrocycle was not formed are
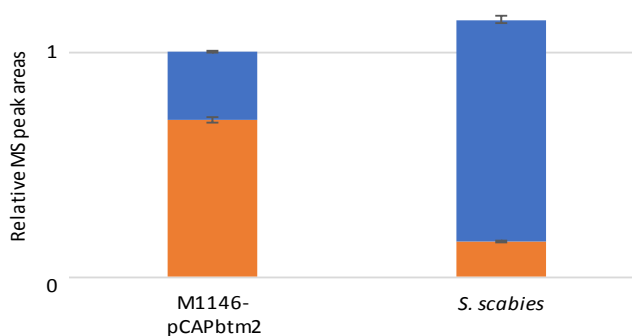
Figure 40. Production of mature bottromycins (blue) and other bottromycin-related metabolites (orange) in *S. scabies* and *S. coelicolor* M1146-pCAPbtm2, based on MS peak areas.
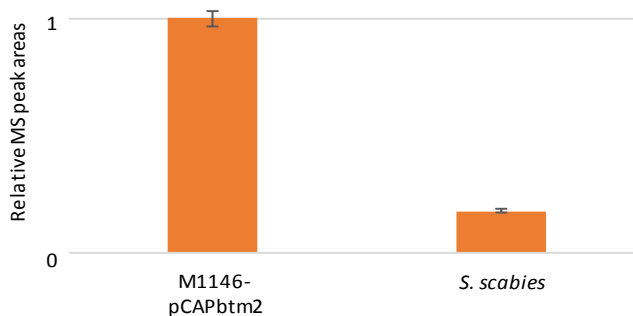
Figure 41. Production of bottromycin-related metabolites in which the macrocycle has failed to form in *S. scabies* and *S. coelicolor* M1146- pCAPbtm2, based on MS peak area.
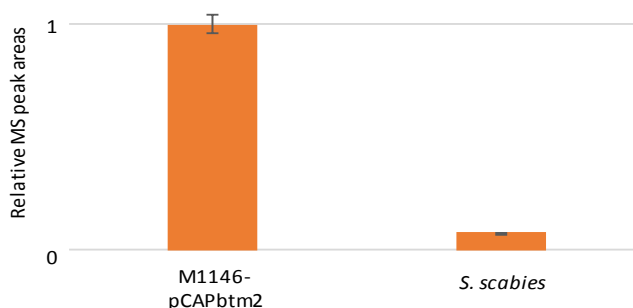
Figure 42. Production of bottromycin-related metabolites in which the C-terminal methionine has failed to be removed in *S. scabies* and *S. coelicolor* M1146- pCAPbtm2, based on MS peak area.

detected (**39**, **40**, **41**, **42**, **44**, **47**, **49**, and **50**), compared to those in *S. scabies* (**43**, **44**, **46**, and **50**; Figure 41). The formation of a macrocycle requires the thiazoline to be present, so this may explain the lack of macrocycle in **49** and **50**, as these are likely to derive from a larger precursor peptide without a heterocycle. The increase in the other non-macrocycle containing metabolites in the refactored cluster is likely due to BtmF functioning poorly, given the presence of a thiazoline in these.

Whilst previous analysis suggested BtmI is involved in macrocycle formation (Crone et al., 2016), this has since been shown to not be the case (Franz et al., 2017; Schwalen et al., 2017). Instead, an unpublished revised biosynthetic proposal suggests that BtmH or BtmI strictly acts after macrocycle formation to catalyse hydrolysis between the thiazoline ring and the first amino acid of the follower peptide, a methionine. Metabolites that contain a methionine at the C-terminus are therefore predicted to result from inefficient BtmH or BtmI-catalysed follower peptide cleavage. In the refactored cluster there seems to be a very large increase in the metabolites with a C-terminal methionine (Figure 42). For example, in pCAPbtm2 significant quantities of the metabolites **32**, **39**, **40**, **41**, **42**, **44**, **45**, and **48** all carrying this methionine can be detected. In *S. scabies* the only metabolites seen with this modification are **43** and **44**. This increase in metabolites that result from inefficient BtmH or BtmI may be because BtmH or BtmI requires the macrocycle to act, as all these metabolites except **32** lack the macrocycle. The presence of **32** in significant quantities and no similar

metabolite in *S. scabies*, however, also suggests that even when the macrocycle is formed BtmH or BtmI efficiency is still negatively affected by the lack of phenylalanine β-methylation.

Once the follower peptide is removed from bottromycin (**1**) intermediates, the P450 BtmJ catalyses oxidative decarboxylation of the thiazoline ring to a thiazole (Crone et al., 2016). Therefore, if the C-terminus of the molecule is a carboxylic acid attached to a thiazoline, it is because BtmJ has not acted. Metabolites where this BtmJ has not acted, such as **22**, **33**, and **46**, are seen in *S. scabies* in small amounts. In the refactored cluster only **34** and **47** are seen, in similarly low quantities. This means there is no discernible increase in the metabolites where the BtmJ has not acted in the M1146-pCAPbtm2. Therefore, BtmJ efficiency does not appear to be negatively affected by the lack of phenylalanine β-methylation. However, if BtmJ efficiency was negatively affected this may not be detected in this analysis. This is because BtmH or BtmI-catalysed proteolysis is required prior to BtmJ-catalysed decarboxylation, and BtmH or BtmI-activity has been shown to be a significant bottleneck in the refactored clusters.

There is a large increase in the production of metabolites that have no C-terminal part of the molecule past the phenylalanine in the refactored cluster (**35**, **36**, **37**, **38**, **49**, and **50**) compared to the production in *S. scabies* (**36**, **37**, **38**, and **50**; Figure 43). The lack of the C-terminal amino acids makes it difficult to guess at which stages of biosynthesis these are generated. The presence of the macrocycle in many of these metabolites suggests the thiazoline ring has been successfully formed. It is possible that inefficient activity of BtmH, BtmI or BtmJ results in intermediates being exposed to endogenous proteases for longer. These endogenous proteases may be cleaving the intermediates between the phenylalanine and the aspartic acid. However, when talking about endogenous proteases, the different cellular background is important: the wildtype cluster is in *S. scabies* and the refactored cluster is in *S. coelicolor* M1146. The different backgrounds may provide a different suite of proteases capably of breaking down the bottromycin-related intermediates between the phenylalanine and aspartic acid residues. Therefore, comparisons of pathway efficiency using these metabolites (**35**, **36**, **37**, **38**, **49**, and **50**) is not possible.
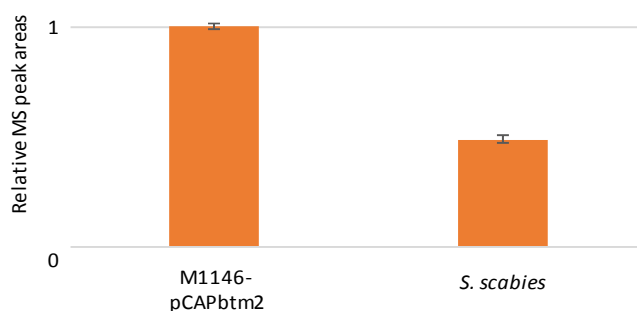
Figure 43. Production of bottromycin-related metabolites in which there is a C-terminal phenylalanine in *S. scabies* and *S. coelicolor* M1146- pCAPbtm2, based on MS peak areas.

The final step in bottromycin (**1**) maturation is the BtmB-catalysed installation of the methyl ester on the aspartate. M1146-pCAPbtm2 lacks this gene, and so cannot be compared to *S. scabies*. However, M1146-pCAPbtm1 contains this gene, and produces small quantities of a mature bottromycin lacking the methyl ester (**31**) compared to those which contain the methyl ester (**27**, **28**). This suggests that the BtmB is not significantly inhibited by the lack of phenylalanine β-methylation.

2.2.2.3.   Testing for Host-Specific Effects

To eliminate the possibility that the lack of active BtmC is a strain-specific effect induced by the heterologous host, *S. coelicolor* M1146, and to assess the role of the host strain on heterologous expression, pCAPbtm2 was also expressed in *S. albus*, *Streptomyces laurentii*, *S. lividans*, and *S. venezuelae* (Figure 44). *S. venezuelae* provided a marginal improvement in production over *S. coelicolor* M1146, while production in *S. laurentii* and *S. lividans* did not differ significantly from *S. coelicolor* M1146. In contrast, the pathway was completely inactive in *S. albus*, indicating the strain-specific nature of heterologous pathway expression. No C-methylation of phenylalanine was observed in any strain, so this particular phenotype does not seem to be a strain-derived effect.
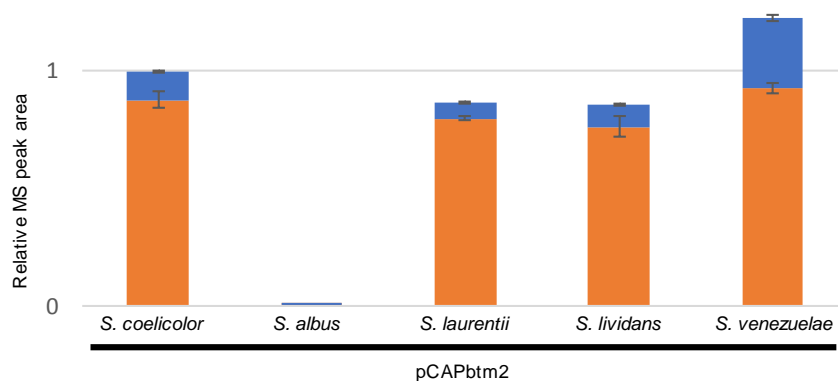
Figure 44. Production of mature bottromycins (blue) and bottromycin-related metabolites (orange) from pCAPbtm2 in five different *Streptomyces* hosts, based on MS peak area.

2.2.2.4.   Sequencing Analysis

One possible cause of BtmC inactivity could have been the introduction of mutations into *btmC* throughout assembly, transformation, or expression. For example, ET12567 cells used in conjugation can be recombinogenic (David Widdick, personal communication), and therefore can mutate plasmids that pass through them. The high expression levels of antibacterial compounds provided by the strong constitutive promoters also put a very high selective pressure towards inactivating the cluster, and inactivating *btmC* could be one method of doing this, as the phenylalanine methylation was recently shown to be important for bottromycin (**1**) activity (Yamada et al., 2018).

Whilst every construct had been validated by sequencing, this assessment was limited to the assembled region and did not cover the whole of *btmC*. Therefore, *btmC* was sequenced in

pCAPbtm1 and pCAPbtm2 using a series of primers spanning the entire gene and its promoter. This showed that the gene had not been mutated during assembly. Alternatively, *btmC* may have been mutationally inactivated during production cultures, as this is when there might be strong selective pressure against a highly active antibiotic-producing cluster. To assess whether the lack of active BtmC was a result of this, DNA was extracted from three replicates of both M1146-pCAPbtm1 and M1146-pCAPbtm2 after production growths, and *btmC* was fully sequenced as described above. Again, no mutations could be seen.

The final mutation-based hypothesis for why no active BtmC was being produced was that there may be mutations elsewhere in the gene cluster having long-range effects on *btmC* expression or function. To assess this M1146-pCAPbtm1 was subjected to whole genome sequencing. The bottromycin (**1**) gene cluster was contained on a single contig and this perfectly aligned against the predicted sequence for the pCAPbtm1 gene cluster. Therefore, the lack of active BtmC is due to a factor other than a mutation.

2.2.2.5.   Transcriptional Analysis

After confirming that the sequence of *btmC* within the refactored clusters was correct, it was important to check that it was being expressed. To carry out transcription analysis, RNA was extracted from M1146-pCAPbtm2 and *S. scabies* production cultures. This was used to produce cDNA. The cDNA was assessed for the presence of transcripts corresponding to *hrdB*, *btmC*, *btmD*, and the region between *btmC* and *btmD* (Figure 45). Transcripts were detected for all of these, showing they are transcribed.

It was at this point that a strong promoter within *btmC* (PbtmD) was identified during an analysis of native bottromycin (**1**) gene cluster regulation in *S. scabies* (Natalia Miguel Vior et al. unpublished). One hypothesis for the lack of active BtmC was that the activity of this internal promoter in a non-native background could be interrupting the expression of *btmC.* Primers were designed that would amplify across the PbtmD transcription start site. These were used in PCR with M1146-pCAPbtm2 and *S. scabies* cDNA and produced a product of the correct size (Figure 45). This showed that the PbtmD transcription start site did not prevent a transcript that contained the whole of *btmC* from being generated.

As the lack of active BtmC is an issue specific to the refactored cluster, it was important to eliminate the possibility that the new promoters were acting incorrectly and initiating transcription from the wrong site, resulting in a transcript that is unable to produce a full length BtmC. Primers were designed to amplify a region starting within the 5'UTR associated with PSF14. These could successfully amplify from M1146-pCAPbtm2 cDNA, showing that transcription is starting from the correct region (Figure 45).
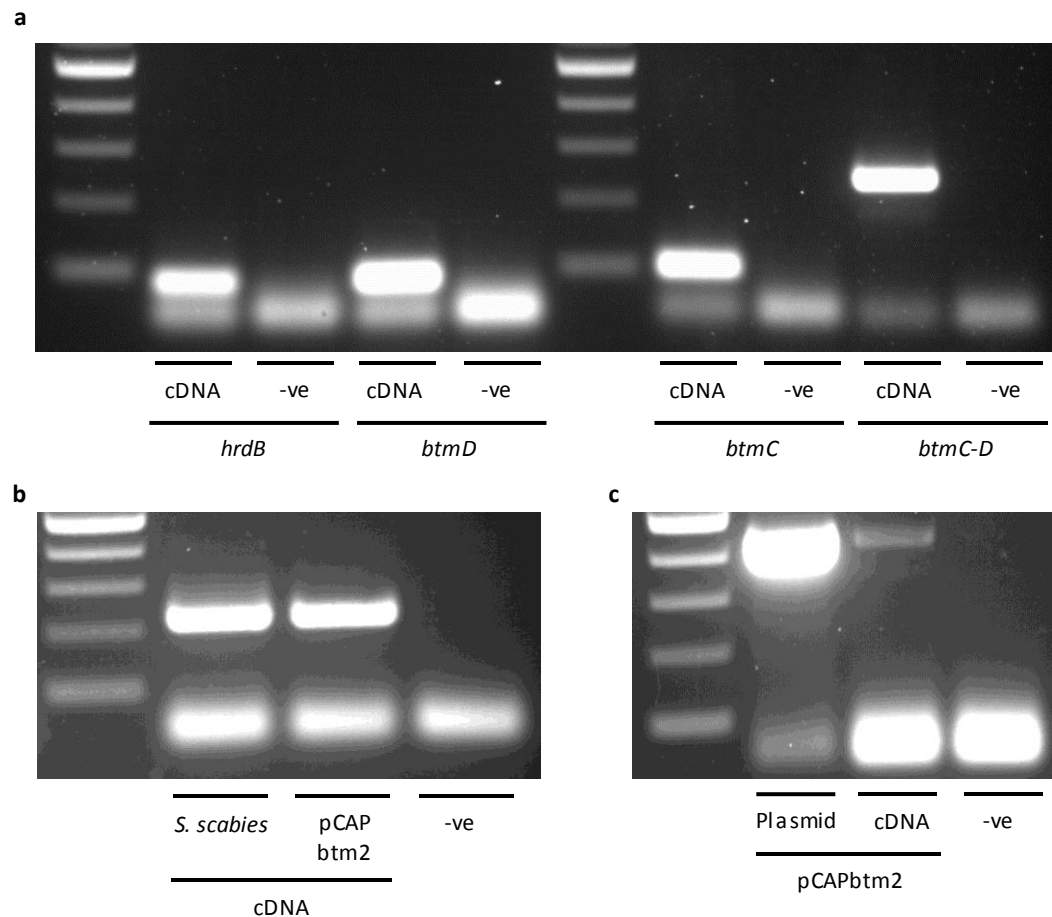
Figure 45. a. RT-PCR of selected fragments from the bottromycin (**1**) gene cluster from M1146-pCAPbtm2 cDNA. The negative control reactions lack reverse transcriptase. **b.** Amplification across PbtmD from *S. scabies* and M1146-pCAPbtm2 cDNA. **c.** Amplification from the 5'UTR associated with PSF14 and *btmC*, from M1146-pCAPbtm2 cDNA and plasmid DNA. The ladder is NEB 2-log.

2.2.2.6.   *In Trans* Expression

A *btmC* deletion in *S. scabies* has previously been shown to be successfully complemented *in trans* by a copy of the gene under the control of PermE* (Crone et al., 2016). This construct could not be used in the *S. coelicolor* M1146 strains containing the refactored cluster, because it uses the same integration site as pCAP01 (φC31). Instead, both the *S. scabies btmC*, and the *S. bottropensis btmC* were cloned in the PermE*-containing pIJ10257 that utilises the φBT1 integration site (Hong et al., 2005). pIJ10257 containing the *S. scabies btmC* was conjugated into M1146-pCAPbtm1 and M1146-pCAPbtm2, whilst the *S. bottropensis btmC* was only conjugated into M1146-pCAPbtm2. Production was tested under normal bottromycin (**1**) production conditions. In all cases, *in trans* expression of *btmC* did not lead to active BtmC.

2.2.2.7.   BtmC Biochemistry

There is the possibility that BtmC inactivity is related to its biochemistry; that of a class B radical SAM methyltransferase. These enzymes require iron-sulphur clusters, SAM, and cobalamin (Bauerle et al.,

2015). One hypothesis is that the lack of active BtmC is due to insufficient supply of one or more of these cofactors, however this is probably not the case as there are two other similar class B radical SAM methyltransferases encoded in the cluster, BtmG and BtmK, which are both functioning normally.

2.2.3.   Probing the Lack of Active BtmC

2.2.3.1.   Investigative Cluster Engineering – Btm3, Btm4, Btm5, and Btm6

The metabolomic analysis (see section 2.2.2.2) revealed that whilst the refactored pathway had significant metabolic flux, the tailoring enzymes were unable to efficiently process intermediates lacking the BtmC-catalysed phenylalanine methylation. Therefore, the inactivity of BtmC was investigated with a series of new versions of the bottromycin (**1**) gene cluster, pCAPbtm3-6. This would take advantage of the efficient yeast-based system that was developed for modifying the gene cluster. These new constructs would all be versions of pCAPbtm2, as this cluster was more productive that pCAPbtm1 when expressed in *S. coelicolor* M1146. These new clusters would therefore have *btmB* deleted, use Paac3 to drive expression of *btmA*, and use PSF14 to drive expression of the *btmC-M* operon.

pCAPbtm3 was generated to assess the importance of the newly identified promoter, PbtmD, by interrupting it. This also allowed a new use for the yeast-based system for modifying the bottromycin (**1**) gene cluster to be tested; instead of refactoring the bottromycin (**1**) gene cluster, yeast-mediated homologous recombination would be used to introduce 33 synonymous mutations within pCAPbtm2. The synonymous mutations were made within the 100 bp region preceding the PbtmD transcription start site. It was hoped that these widespread mutations would disrupt the activity of this promoter. pCAPbtm3 was assembled from BlpI digested pCAPbtm2, two PCR fragments, and three oligonucleotides (Figure 46). Two of six colonies screened by PCR contained the correct plasmid, and one of these plasmids was additionally confirmed by sequencing of the modified region.

As with previous constructs, pCAPbtm3 was expressed in *S. coelicolor* M1146, where it generated similar metabolites as M1146-pCAPbtm2, all also containing unmodified phenylalanine. This indicated that there was still no active BtmC. There was also a slight decrease in production when compared with M1146-pCAPbtm2 (Figure 47). This is consistent with (but does not prove) the predicted role of the intragenic promoter, as altering this region is predicted to reduce precursor supply and decreases metabolic flux through the pathway. This also showed that interrupting this promoter does not rescue expression of *btmC*, suggesting that the lack of active BtmC is not due to the intragenic promoter. It should be noted that further expression analysis such as qRT-PCR would be required to fully confirm that the drop in productivity was in fact due to the interruption of the *btmD*-specific promoter.

pCAPbtm4 and pCAPbtm5 were designed to assess whether the position or presence of *btmC* is important for pathway productivity. In pCAPbtm4 the positions of *btmC* and *btmD* were swapped. In pCAPbtm5 *btmC* was deleted. pCAPbtm4 was assembled from BsrGI and XhoI digested pCAPbtm*, three PCR fragments, and 6 oligonucleotides (Figure 48). All six of the colonies screened by PCR
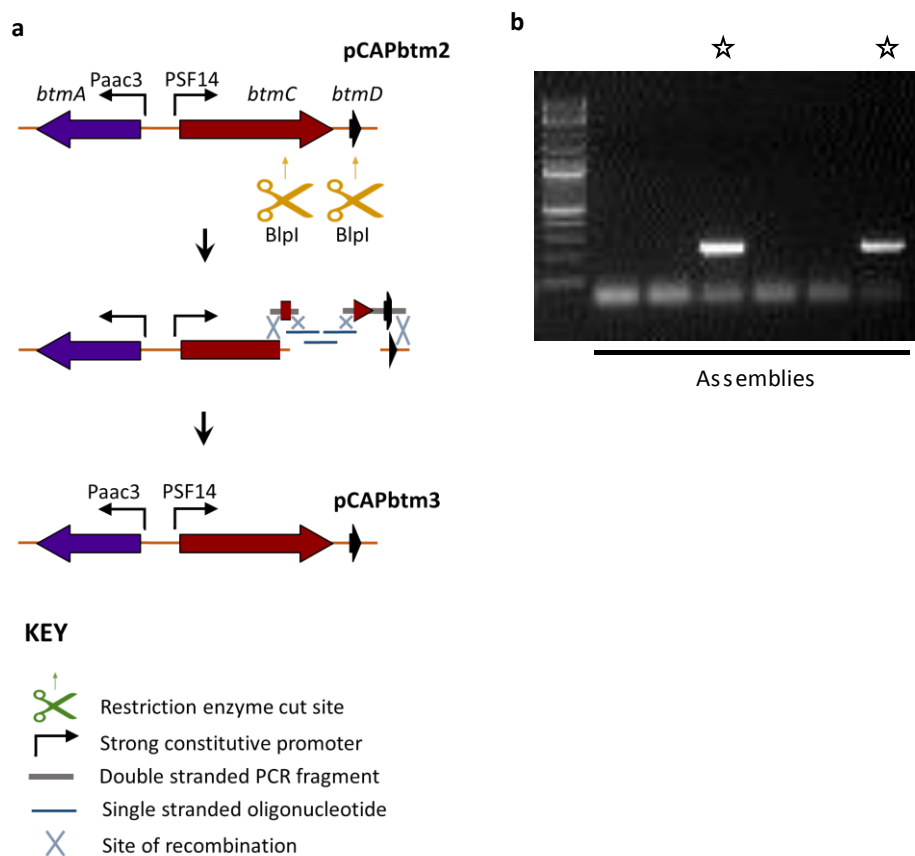
Figure 46. a. Schematic of the assembly of pCAPbtm3 from pCAPbtm2 in yeast. Only the *btmA-btmD* portion of the gene cluster is shown for clarity. **b.** PCR screening of assemblies, lanes with bands indicative of correct assemblies are marked with a star. The ladder is NEB 2-log.
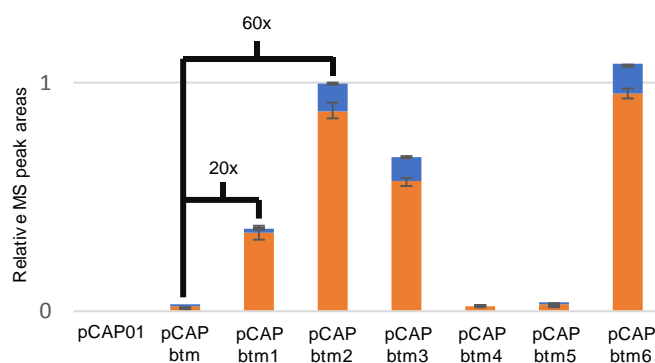
Figure 47. Production of mature bottromycins (blue) and other bottromycin-related metabolites (orange) from refactored bottromycin (**1**) gene clusters expressed in *S. coelicolor* M1146, based on MS peak areas.

contained the correct plasmid from this assembly. pCAPbtm5 was assembled from BsrGI and NsiI digested pCAPbtm* and seven oligonucleotides (Figure 49). All 12 of the colonies checked by PCR contained the correct plasmid. pCAPbtm4 and pCAPbtm5 plasmid were extracted from positive colonies and were confirmed by sequencing the assembled region. The assembly of pCAPbtm4 was the most complicated assembly conducted in this study, with six oligonucleotides and three PCR fragments being assembled in to a plasmid to simultaneously replace promoters, delete a gene, and change the order of genes in a single step. The 100% assembly efficiency of this suggests that this
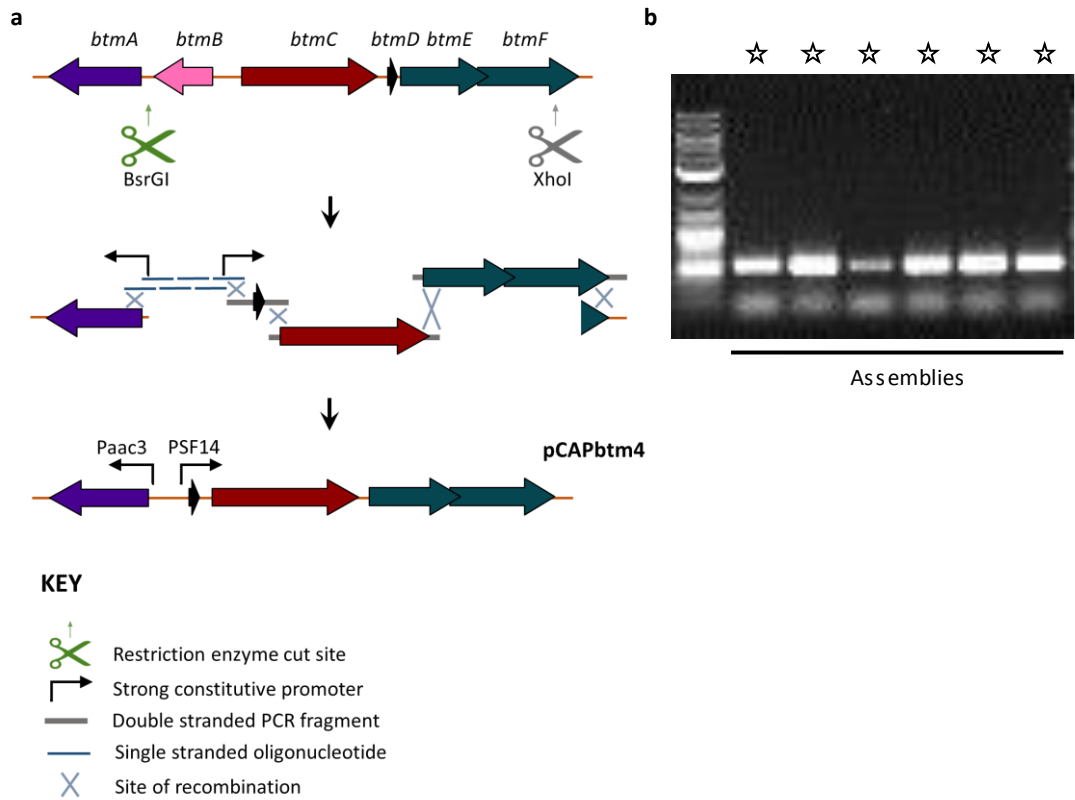
Figure 48. a. Schematic of the assembly of pCAPbtm4 from pCAPbtm* in yeast. Only the *btmA-btmF* portion of the gene cluster is shown for clarity. b. PCR screening of assemblies, lanes with bands indicative of correct assemblies are marked with a star. The ladder is NEB 2-log.
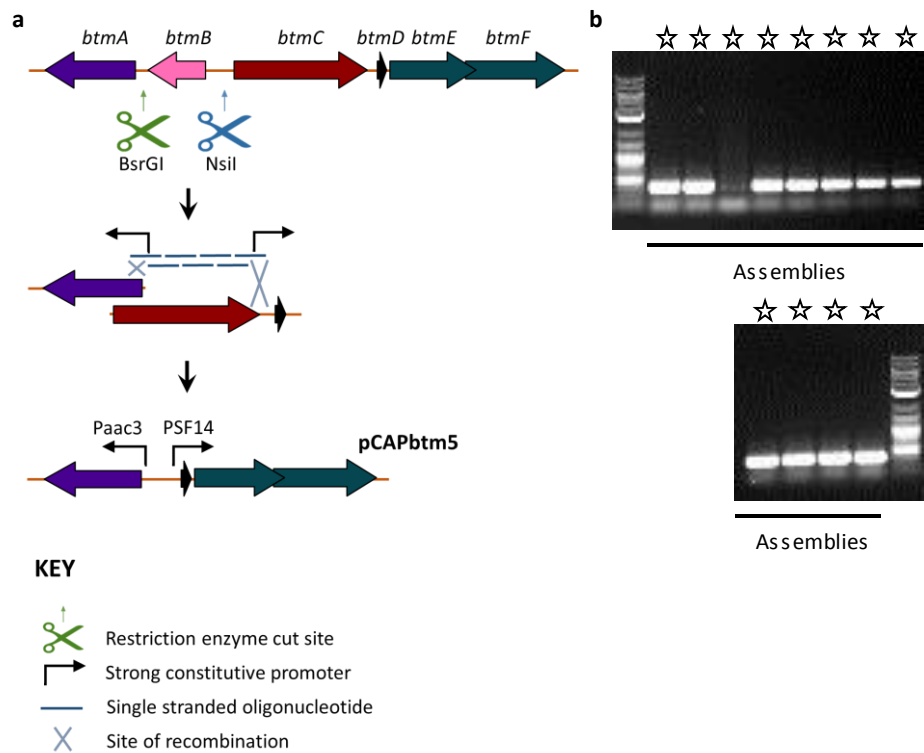
Figure 49. a. Schematic of the assembly of pCAPbtm5 from pCAPbtm* in yeast. Only the *btmA-btmF* portion of the gene cluster is shown for clarity. b. PCR screening of assemblies, lanes with bands indicative of correct assemblies are marked with a star. The ladder is NEB 2-log.

yeast assembly methodology can be exploited further than it was in this study, and that more complicated assemblies making more significant changes to the gene cluster could be attempted in the future. This is particularly appealing considering the one-step nature of this technique makes creating the cluster variants a quick process.

As with other assemblies, the productivity of pCAPbtm4 and pCAPbtm5 were tested in *S. coelicolor* M1146. Almost complete abolition of production was seen when the production from these strains was assessed by LC-MS (Figure 47). This shows that, despite its inactivity, the position and presence of *btmC* within the operon is required for pathway activity. This suggests that the regulation of the bottromycin (**1**) gene cluster is much more complex than previously expected. Whilst the strong promoters can activate the otherwise silent pathway in M1146, they do this at the expense of active BtmC, yet the activated pathways still require *btmC* in its native location for activity.

pCAPbtm6 was designed to replace the 5' UTR associated with PSF14 with the native 5' UTR of *btmC*. This was to establish whether its RNA structure is important for translation initiation or mRNA
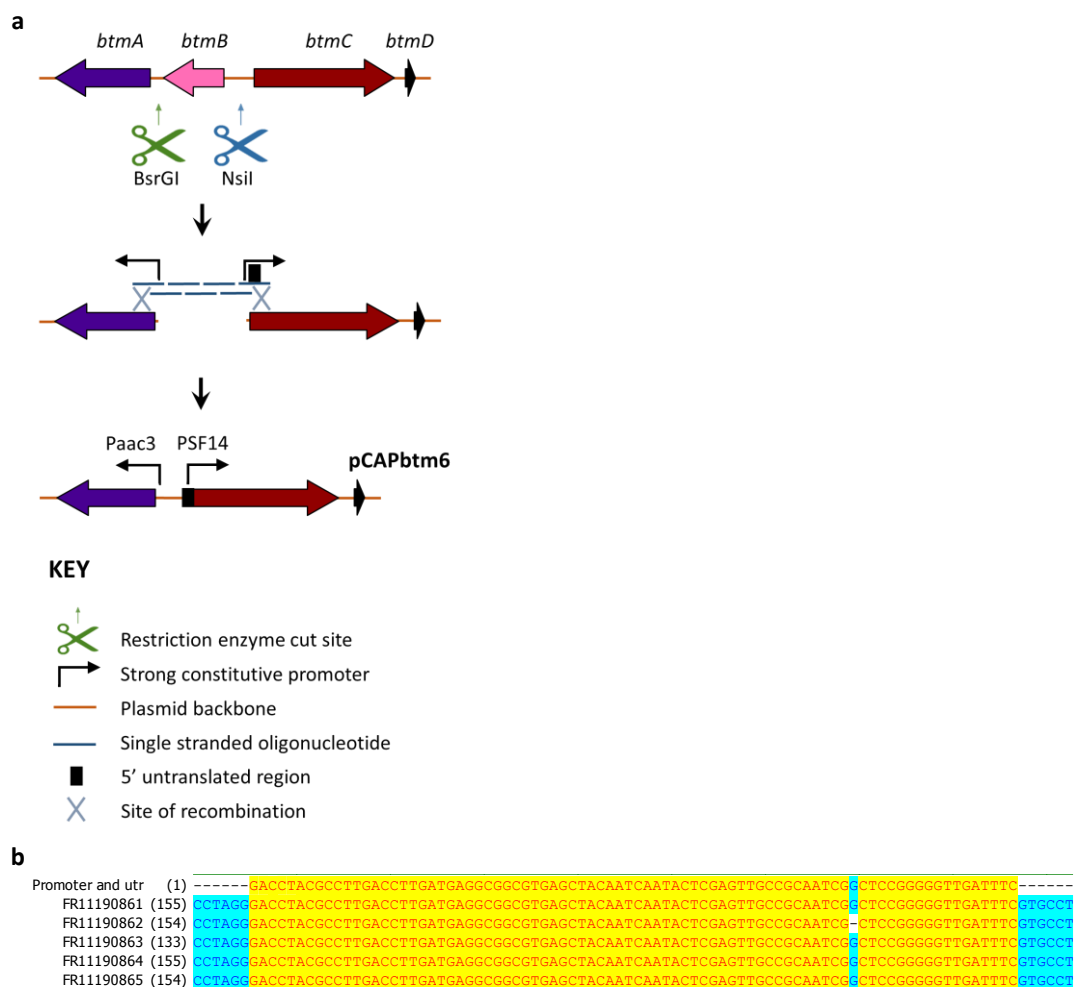
Figure 50. **a.** Schematic of the assembly of pCAPbtm6 from pCAPbtm* in yeast. Only the *btmA-btmF* portion of the gene cluster is shown for clarity. **b.** Screening of assemblies by sequencing, aligned to the promoter and 5'UTR being installed upstream of *btmC*. All but FR11190862 are correct.

stability. This was assembled from BsrGI and NsiI digested pCAPbtm* and seven oligonucleotides. Five assembled plasmids were sequenced and four of the five contained the correct sequence (Figure 50). As with the other constructs, pCAPbtm6 was tested in *S. coelicolor* M1146. The metabolic profile of pCAPbtm6 showed no significant difference from pCAPbtm2 (Figure 47), suggesting that changing the 5' UTR of *btmC* was unlikely to be the reason for the lack of active BtmC.

2.2.3.2.   Attenuating *btmC* expression – Btm7 and Btm8

Class B radical SAM methyltransferases are often insoluble when overexpressed (Bauerle et al., 2015). As *btmC* is being strongly expressed in the refactored clusters, this could explain the lack of active BtmC. There are other class B radical SAM methyltransferases in the clusters that are active and also being overexpressed, but these could be predicted to follow intracistronic attenuating terminators and so it is possible that only *btmC* is overexpressed to a level at which the protein becomes insoluble. Therefore, a possible solution to the lack of active BtmC is attenuating the expression of *btmC*. This could potentially be achieved by using a riboswitch. Riboswitches are regions of mRNA that form secondary structures that respond to the addition or removal of a ligand (Sherwood and Henkin, 2016); this change in secondary structure can have an impact on transcription and/or translation. Installing a riboswitch in this cluster would not only potentially solve the lack of active BtmC issue, but it would also help validate the utility of a riboswitch in refactored *Streptomyces* gene clusters, as there is only one example of a riboswitch-controlled natural product gene cluster (Horbal and Luzhetskyy, 2016). A series of theophylline-dependent riboswitches that enabled inducible gene expression in *S. coelicolor* have been published (Rudolph et al., 2013). Riboswitch E* provided the strongest control of expression in that study, so we constructed versions of pCAPbtm1 and pCAPbtm2 with this riboswitch between PSF14 and btmC to generate pCAPbtm7 and pCAPbtm8, respectively. pCAPbtm7 was assembled from BsrGI and NsiI digested pCAPbtm*, four PCR fragments and two oligonucleotides (Figure 51). pCAPbtm8 was assembled from BsrGI and NsiI digested pCAPbtm*, and eight oligonucleotides (Figure 51). Yeast colonies from the assemblies were checked for the presence of the correct plasmid by PCR. A correct plasmid from each was extracted, passaged through *E. coli*, and additionally confirmed by sequencing of the assembled region.

The production of bottromycin-related metabolites was almost abolished in M1146-pCAPbtm7. The addition or removal of theophylline had almost no effect on production, and in both cases only **38** could be detected (Figure 52). In contrast, bottromycin-related metabolite production was tightly controlled by varying theophylline concentrations in M1146-pCAPbtm8 (Figure 53). Whilst it did not rescue BtmC activity, the addition of theophylline caused a large increase in bottromycin-related metabolite production, and at 8 mM theophylline, M1146-pCAPbtm8 produced 120 times the quantity of bottromycin-related metabolites compared to the wild type cluster in the heterologous
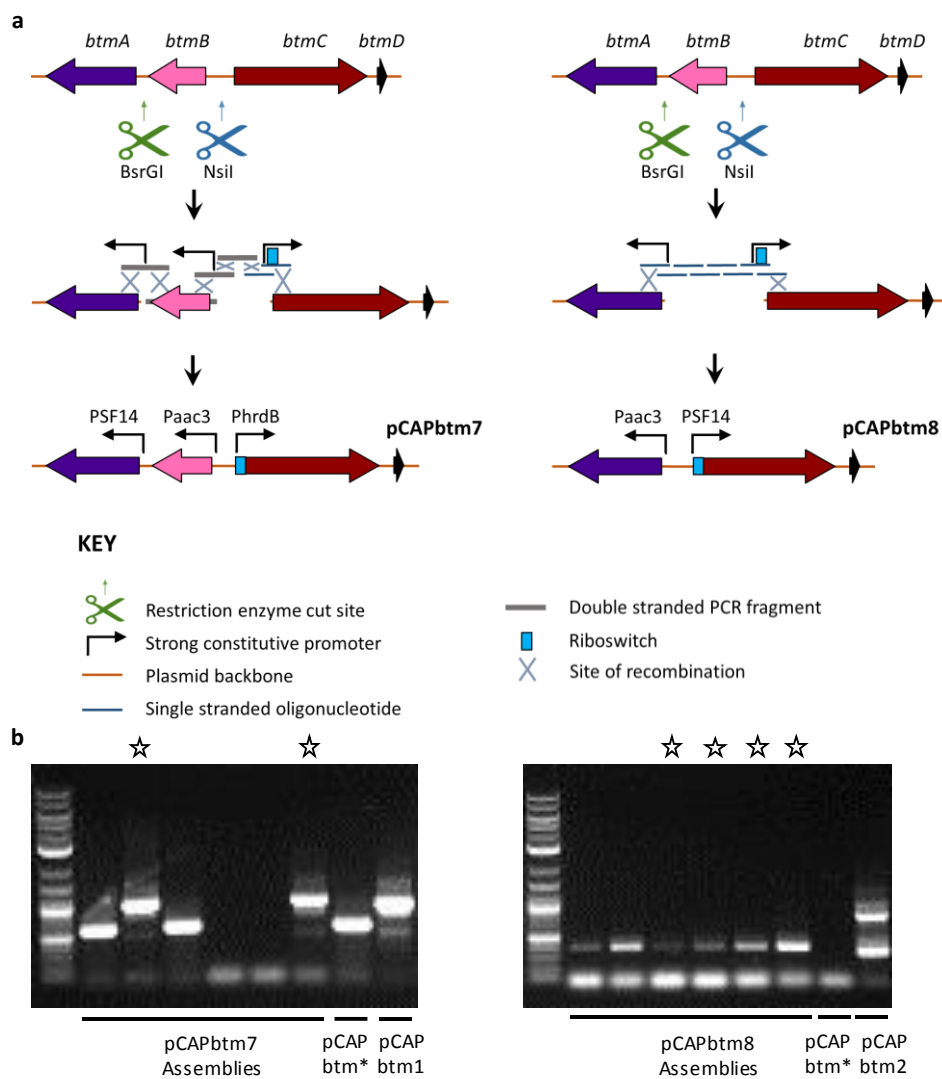
Figure 51. a. Schematic of the assembly of pCAPbtm7 and pCAPbtm8 from pCAPbtm* in yeast. Only the *btmA-btmD* portion of the gene cluster is shown for clarity. **b.** PCR screening of assemblies, lanes with bands indicative of correct assemblies are marked with a star. The ladder is NEB 2-log.
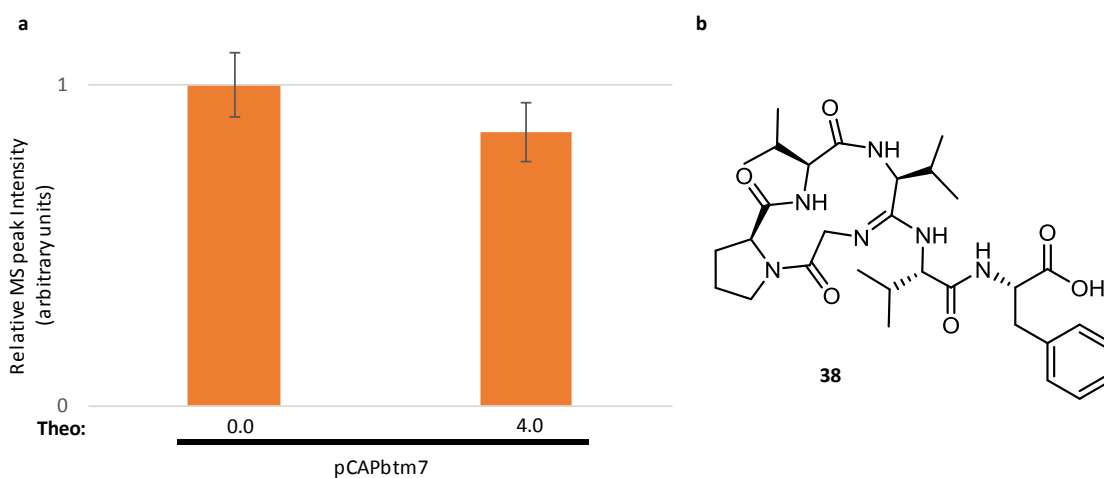
Figure 52. a. Production of bottromycin-related metabolites (orange) from pCAPbtm7 in *S. coelicolor* M1146, based on MS peak area. Theophylline concentrations (mM) are listed. **b.** The structure of **38**, the only metabolite seen produced by *S. coelicolor* M1146 expressing pCAPbtm7
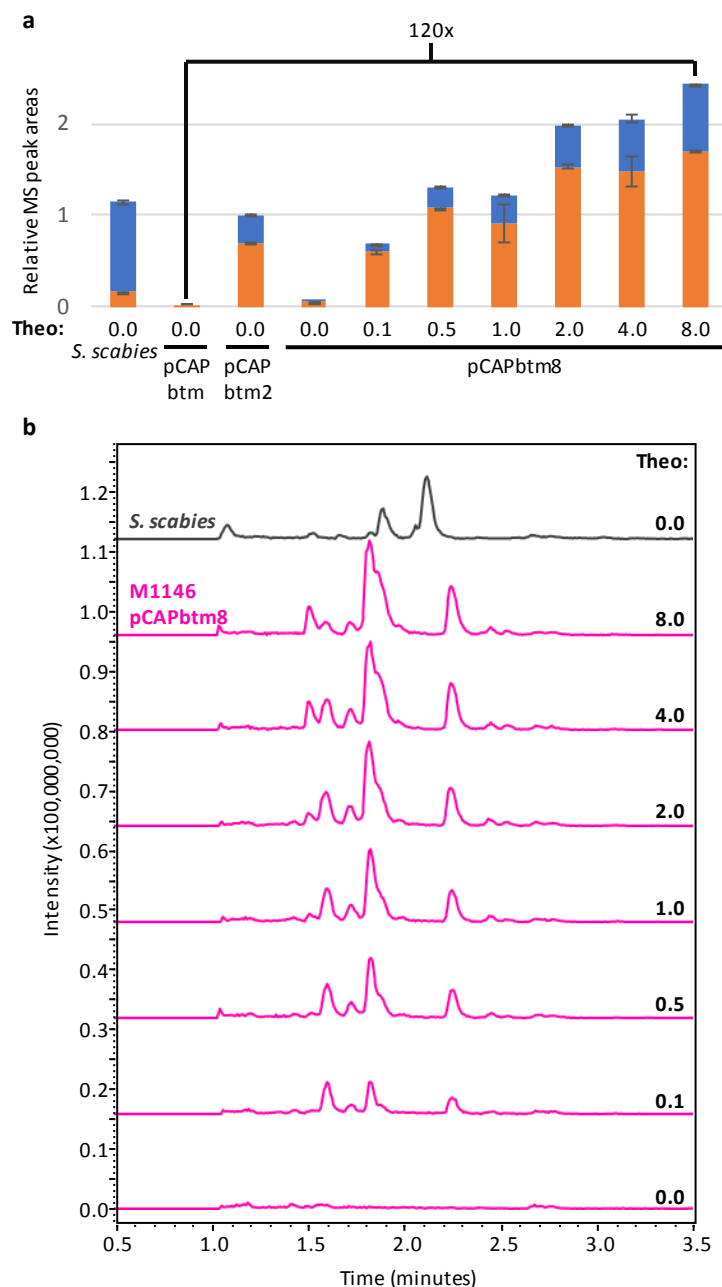
Figure 53. a. Production of mature bottromycins (blue) and other bottromycin-related metabolites (orange) by *S. scabies*, and pCAPbtm, pCAPbtm2 and the riboswitch-containing pCAPbtm8 in *S. coelicolor* M1146, based on MS peak area. Theophylline concentrations (mM) are listed. **b.** EIC of all bottromycin-related metabolites detected in *S. scabies* and M1146-pCAPbtm8 at different theophylline concentrations (mM).

host (M1146-pCAPbtm). The production of mature bottromycins also reached a level comparable to wild type *S. scabies*, and the total yield of bottromycin-related metabolites measured by MS intensity was over twice that of *S. scabies* (Figure 53). The metabolites produced were the same as those produced by M1146-pCAPbtm2 (therefore not C-methylated on phenylalanine) and were detected in comparable ratios. This titratable control of the bottromycin pathway via this riboswitch was a surprising result, as it was predicted to only control translation of *btmC*.

It was unclear how the riboswitch in pCAPbtm8 was controlling bottromycin-related metabolite production, as it influenced the entire pathway's productivity even though it was in front of a gene (*btmC*) that was unable to produce active protein. One hypothesis was that the riboswitch in pCAPbtm8 was controlling transcription instead of translation. To assess whether this was true, cDNA was made from RNA extracted from production cultures of M1146-pCAPbtm2 and M1146-pCAPbtm8 with varying concentrations of added theophylline. This cDNA was used in qPCR experiments to measure the transcript levels of *btmC* and *btmD*, which were then compared to productions levels (Figure 54). This revealed that whilst the riboswitch did influence transcription, this was not entirely correlated with production, and at theophylline concentrations over 0 mM the role of the riboswitch on pathway productivity could not be simply explained by transcript levels. For example, transcript levels of *btmD* with 4 mM theophylline were nine-fold lower than in pCAPbtm2, but this was associated with a significant increase in bottromycin-related metabolite production. Transcription of *btmC* seemed to more closely match production, although there were still some disparities, for example a higher transcript level at 0.1 mM theophylline compared to 0.5 mM theophylline. The fact that the transcript levels of *btmC* and *btmD* react slightly differently to increasing theophylline concentrations is particularly interesting and suggests an important role for
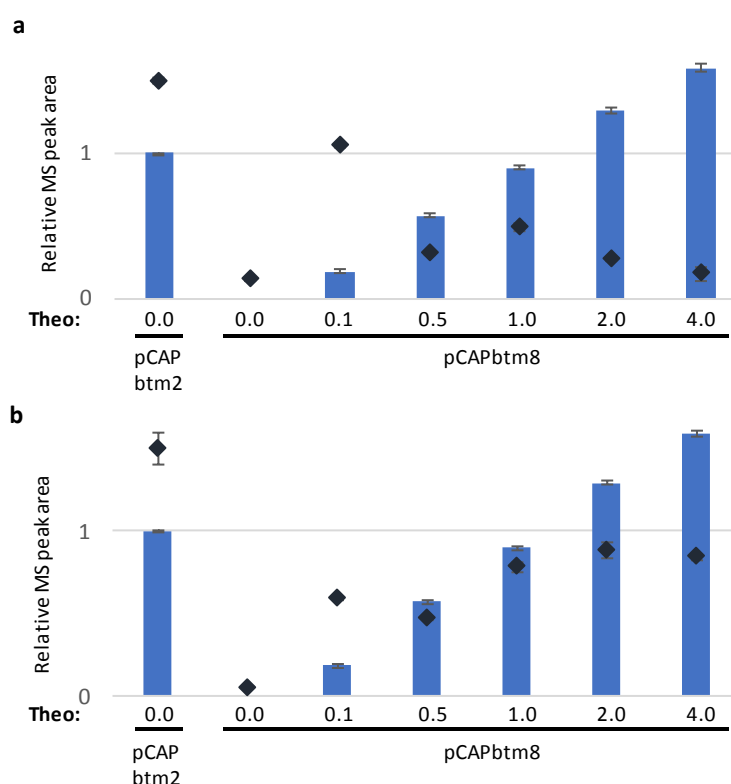
Figure 54. Production of mature bottromycins (blue bars) from pCAPbtm2 and the riboswitch-containing pCAPbtm8 with increasing concentrations of theophylline, based on MS peak area. qRT-PCR measurements of the transcript levels of **a.** *btmD* and **b.** *btmC* normalised to pCAPbtm2 = 1.5 are also shown (dark blue diamonds).

regulation outside of PSF14 and the riboswitch. This result infers a complex interplay between gene transcription and translation that was beyond the scope of this project to fully understand.

2.3. Conclusion

2.3.1.   Summary of Results

2.3.1.1.   Bottromycin Cluster Manipulation

The combination of yeast-mediated TAR cloning and refactoring was an efficient and versatile way to modify the GC-rich bottromycin (**1**) gene cluster (Table 2). Cluster assemblies were achieved in a single step with different combinations of PCR products and single-stranded oligonucleotides. The iterative nature of the refactoring allowed the many parts used in the assemblies to be reused in a modular way (Figure 55), which made successive assemblies quicker to design and cheaper to execute. There was a high average assembly efficiency, where 79% of colonies screened were correctly assembled. The most complex assembly reported here, pCAPbtm4, was constructed from three PCR products, six oligonucleotides and a digested vector, and was correctly assembled in all colonies screened. The method takes advantage of naturally occurring restriction sites and facilitated the introduction of new ones. A potential issue with this technique is the availability of restriction enzymes that do not cut elsewhere in the vector. However, the efficiency of this technique may allow for single oligonucleotides to bridge these unwanted cuts during assemblies. Additionally, the recent

Table 2. Summary of pCAPbtm-based assemblies.

Name	Modification(s)	Parts used	Success rate and screening method	Result
pCAPbtm*	Insertion of BsrGI restriction site.	2 PCR products	3/4 Restriction analysis	Successful proof of principle that simplifies further engineering.
pCAPbtm1	Introduction of strong constitutive promoters.	4 PCR products	4/6 PCR analysis	Improved production and introduced bottleneck.
pCAPbtm2	*btmB* deletion and introduction of strong constitutive promoters.	7 oligos	2/5 Restriction analysis	Improved pathway productivity over pCAPbtm1
pCAPbtm3	btmB deletion, introduction of strong constitutive promoters, and disruption of *btmC* internal promoter.	2 PCR products 3 oligos	2/6 PCR analysis	Slight reduction in production from pCAPbtm2; supports a role for promoter within *btmC*.
pCAPbtm4	*btmB* deletion, introduction of strong constitutive promoters, and swapping of *btmC* and *btmD* positions.	3 PCR products 6 oligos	6/6 PCR analysis	Indicates that gene order is important despite non-functional BtmC.
pCAPbtm5	*btmB* deletion, introduction of strong constitutive promoters, and *btmC* deletion.	7 oligos	12/12 PCR analysis	*btmC* is important for efficient pathway productivity, regardless of whether active protein was produced.
pCAPbtm6	*btmB* deletion, introduction of strong constitutive promoters, and inclusion of *btmC* 5' UTR.	7 oligos	4/5 Sequencing analysis	5' UTR had minimal effect as pathway productivity was similar to pCAPbtm2.
pCAPbtm7	Introduction of strong constitutive promoters, and placement of a riboswitch before *btmC*.	4 PCR products 2 oligos	2/6 PCR analysis	Significantly reduced pathway productivity. The riboswitch provided no control.
pCAPbtm8	*btmB* deletion, introduction of strong constitutive promoters, and placement of a riboswitch before *btmC*..	8 oligos	6/6 PCR analysis	Provided highest production levels. Gave tight control over total cluster productivity.
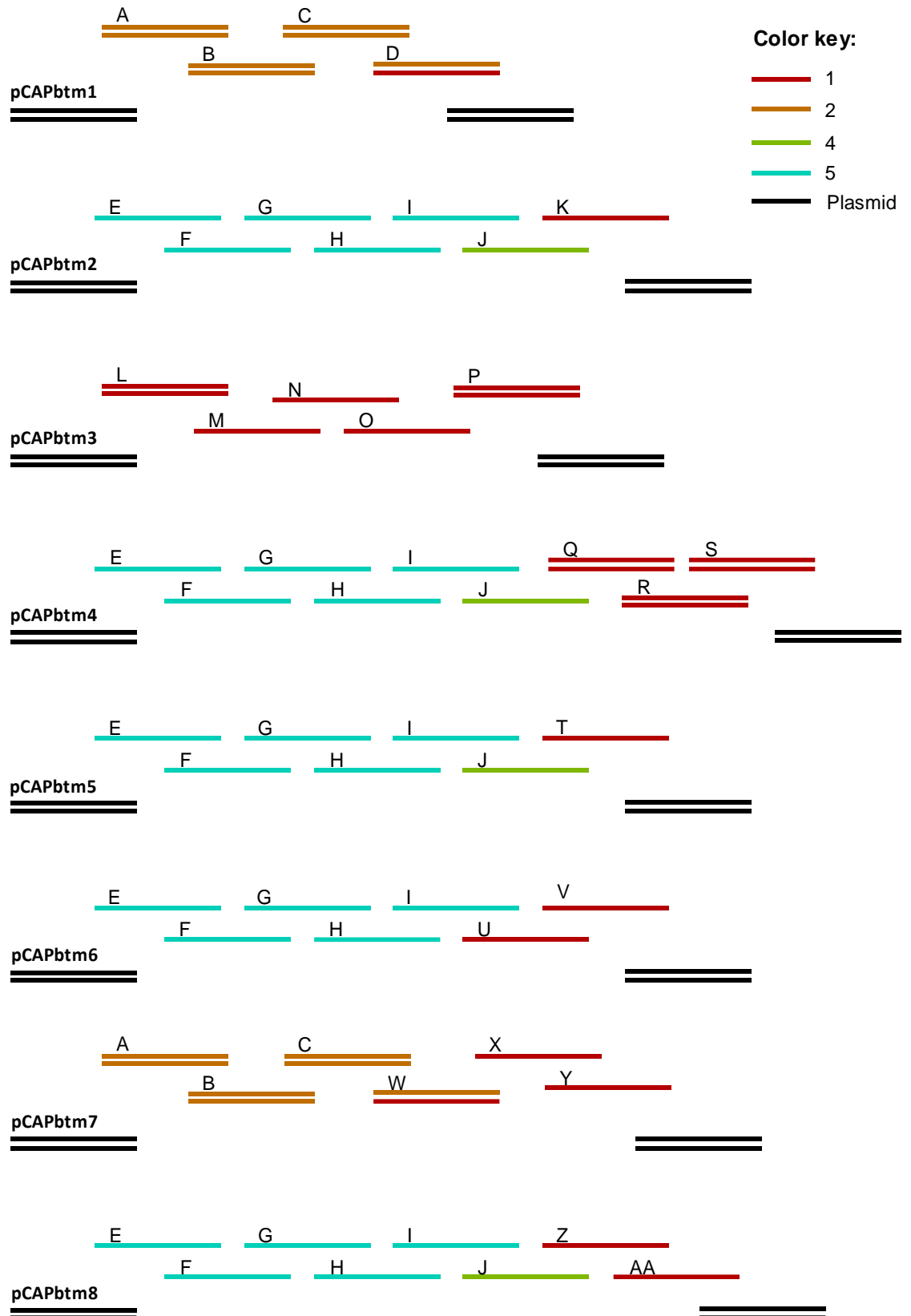
Figure 55. Schematic showing the modularity of cluster assemblies. The colour represents the number of times the parts were used across all assemblies. Double stranded parts have the top strand coloured with regards to the sense primer, and the bottom strand coloured with regards to the anti-sense primer.

development of programmable restriction enzymes potentially overcomes this barrier (Enghiad and Zhao, 2017; Liu et al., 2015).

The yeast-based technique presented here has many advantages over other published yeast-based cluster modification techniques. It does not require a selectable marker to be coupled to the modifications as selection is based on plasmid recircularization, unlike some other techniques (Montiel et al., 2015; Yamanaka et al., 2014). This is significant, as coupling modifications to selectable markers can make more complex modifications difficult, for example swapping gene order as in pCAPbtm4. This process only requires a single assembly step, whilst some techniques for complete refactoring can require multiple assembly steps (Pahirulzaman et al., 2012; Shao et al., 2013). As yeast is a slower growing organism than *E. coli*, it is advantageous to minimise the number of steps yeast is used in. When compared to non-yeast techniques, the methodology presented here also has several advantages. For example, it is not limited by the availability of type IIS restriction enzymes that do not cut within the assembled parts, such as in Golden Gate based techniques (Engler et al., 2009; Weber et al., 2011). There was also no issue encountered with the high GC nature of the cluster, whilst both Gibson assembly and twin primer assembly (TPA) have issues with this in more complex assemblies (Casini et al., 2013; Liang et al., 2017).

The most successful construct (pCAPbtm8) increased production in *S. coelicolor* M1146 120-fold compared to the wild type cluster (pCAPbtm), and the use of a riboswitch led to strict control of production (Figure 53). This is only the second report of riboswitch-dependent inducible expression of an entire Streptomyces gene cluster (Horbal and Luzhetskyy, 2016), and the first example in *S. coelicolor*. The theophylline-controlled riboswitch represents a rapid way to modulate gene expression without the need for a library of gene clusters containing promoters of different strengths.

2.3.1.2.   BtmC-Induced Bottlenecks

Whilst it was ultimately unclear why there was no active BtmC, a significant amount of data was gathered that allowed many hypotheses to be ruled out. Sequencing results showed that there were no mutations in *btmC*. RT-PCR showed that it was transcribed, whilst qRT-PCR showed that transcript level was not sufficient to explain changes in production. pCAPbtm4 and pCAPbtm5 showed that the presence of *btmC* at the start of the cluster was important for pathway activity. The inclusion of the riboswitch showed that attenuating *btmC* expression did not produce an active enzyme. pCAPbtm6, that contained *btmC*'s native 5'UTR, showed that the *btmC* 5'UTR did not contain a regulatory feature vital for expression. pCAPbtm3, in which an attempt to interrupt the *btmD* promoter within *btmC* was made, suggested that PbtmD was not responsible for the lack of active BtmC. Three different promoters, PhrdB in pCAPbtm1, PSF14 in pCAPbtm2, and PermE* in pIJ10257-btmC, were tested in front of *btmC*, therefore the issue was not specific to a promoter.

Untargeted metabolomics and mass spectral networking enabled the identification of the metabolites produced by the refactored cluster and aided in understanding what effect the lack of active BtmC had on production. This can be explained by looking at the pathway and the metabolic flux passing through it. BtmC is one of the first enzymes to act on the precursor peptide, and some tailoring enzymes that catalyse subsequent reactions are inefficient at acting on a core peptide that lacks the BtmC-catalysed phenylalanine β-methylation. The refactored clusters that lack active BtmC have a large amount of metabolic flux being fed into them by the strong promoters, but much of it gets diverted away from the final product (Figure 56), due to the tailoring enzyme inefficiencies. The detected shunt metabolites are results of: enzymes not being able to act (for example BtmJ not functioning, producing **34**), enzymes acting incorrectly (for example valine hypermethylation,
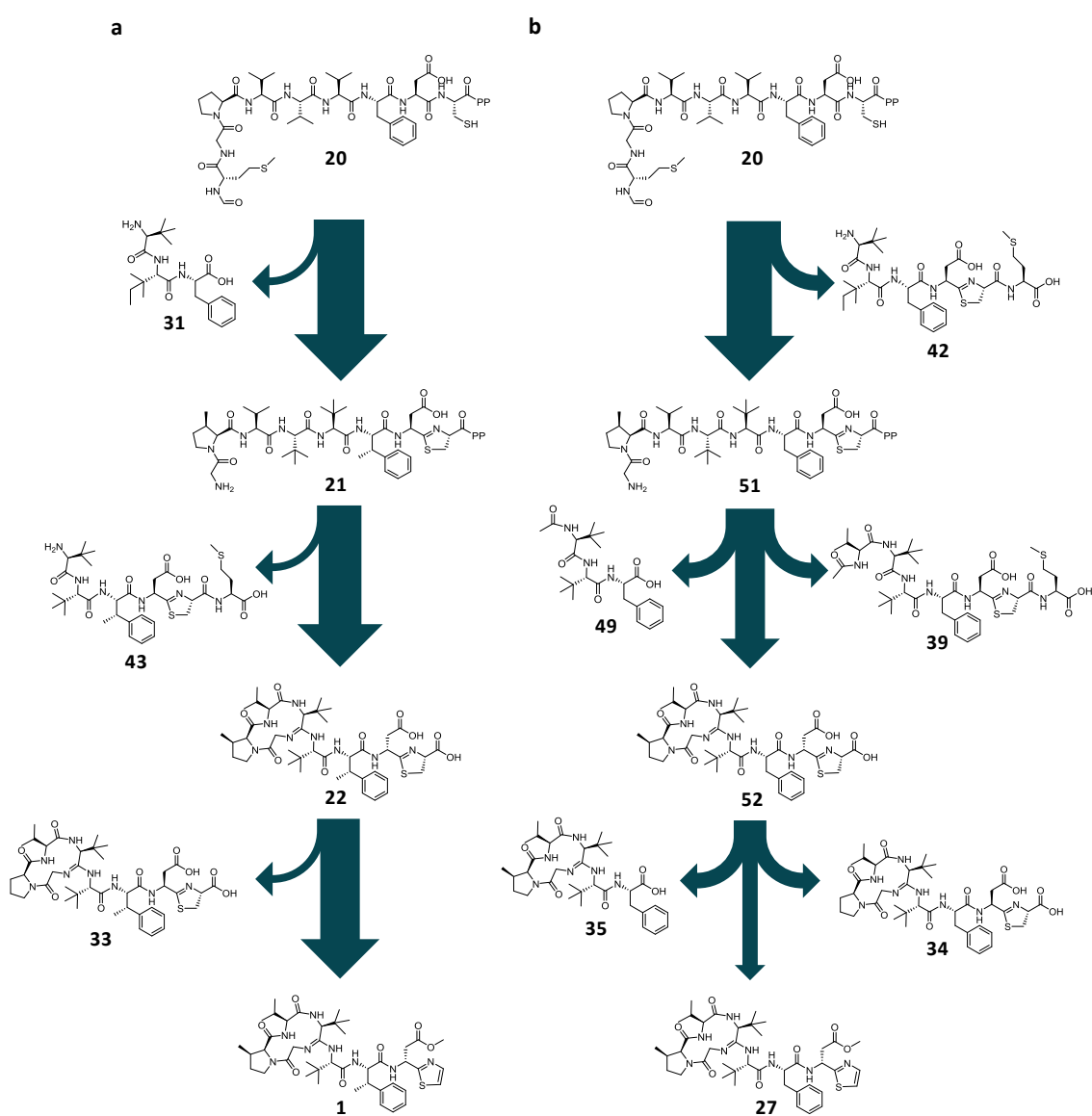
Figure 56. Suggested bottromycin flux model. Selected non-quantitative examples of flux being diverted away from the primary pathway in **a.** *S. scabies*, and **b.** M1146-pCAPbtm1. **20** is the predicted precursor peptide. **21**, **51**, and **52** are hypothetical intermediates. PP labels the C-terminal region of the precursor peptide.

producing **42**), and the slow turnover rate exposing intermediates to proteolysis and acetylation (for example **49**). This is seen in the wild type cluster (Crone et al., 2016), but at a much lower amount.

A more in-depth analysis of the metabolites produced can indicate which steps are impacted in the refactored cluster, likely due to the lack of phenylalanine β-methylation. These results suggest that the main step in bottromycin (**1**) biosynthesis that becomes less efficient is the BtmF-catalysed macrocycle formation. The BtmH or BtmI-catalysed follower peptide hydrolysis is also less efficient, although this is likely caused by the accumulation of intermediates lacking the macrocycle which are poorer substrates for the peptidase.

2.3.2.   Future Implications

2.3.2.1.   Development of Bottromycin Derivatives

This study has revealed a lot about the bottromycin (**1**) pathway. The pathway is particularly resistant to regulatory modification, as all modifications resulted in a lack of active BtmC. It is likely that the regulation of the pathway is not fully understood, and therefore interruptions of this caused the lack of active BtmC. It is possible that a better understanding of the pathway regulation and control would enable this issue to be fixed, however this was beyond the scope of this project. The compound diversification that resulted from the lack of active BtmC also showed that the other tailoring enzymes in this pathway are particularly intolerant to changes in their substrate. This will inhibit attempts to produce a hydrolysis-resistant derivative of bottromycin by modifying the biosynthesis of the pathway. A recently published observation was that the BtmC-catalysed phenylalanine β-methylation is necessary for bottromycin (**1**) activity (Yamada et al., 2018), and so the lack of active BtmC would have to be corrected in refactored systems for their products to be useful. Several active versions of bottromycin have been produced that lack the methyl ester by semisynthesis from mature bottromycin (**1**; Kobayashi et al., 2010) and by total synthesis (Yamada et al., 2018), so this may be a more viable method to produce clinically relevant bottromycin-related metabolites.

One promising observation from the refactored clusters is the large increase in production of metabolites featuring a C-terminal phenylalanine, that therefore lack the aspartic acid and thiazole ring (**35-38**; Figure 57). This is analogous to a synthetic intermediate (**53**) used in a total synthesis and derivatisation of bottromycin study (Yamada et al., 2018), in which four different active bottromycin-related compounds were produced in a single step from this intermediate. It is possible that the refactored clusters could be a quicker route to producing this intermediate, although there are small differences, for example the lack of phenylalanine β-methylation. This methylation may be less important for activity in these highly modified versions of bottromycin. As these intermediates have been predicted to occur at such high levels due to inefficient BtmH, BtmI and/or BtmJ activities,

53: From Yamada et al. (2018)

35: R$_{1-3}$ = CH$_3$
36: R$_{1-2}$ = CH$_3$ and R$_3$ = H
37: R$_2$ = CH$_3$ and R$_1$, R$_3$ = H
38: R$_{1-3}$ = H
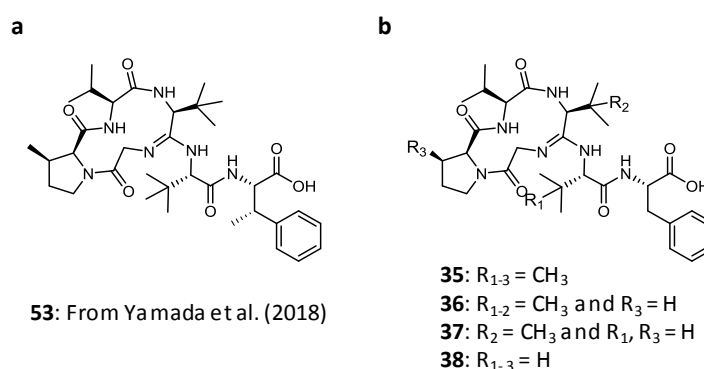
Figure 57. a. Structure of synthetic intermediate used for derivatisation (**53**; Yamada et al., 2018). **b.** Structures of similar compounds **35-38** produced by refactored clusters.

these genes could be deleted in the refactored clusters to try to produce even more of these intermediates.

2.3.2.2.   Modification of Other Biosynthetic Gene Clusters

The most significant result from this investigation is the development of yeast as a tool for reprogramming RiPP gene clusters. The method has proven to be efficient and robust; manipulations could be made within a few weeks with an average efficiency of 79%. Applying this technique to the bottromycin (**1**) gene cluster was hindered by the unusual results, and the analysis thereof. Therefore, it is particularly exciting to envisage this being applied to other gene clusters. This technique was subsequently used to successfully reprogram the thiostreptamide S4 (**17**) gene cluster, which is reported in the next chapter.

Chapter 3 – Thiostreptamide S4

3.1. Introduction

3.1.1.   Previous Work on Thiostreptamide S4

3.1.1.1.   Thioviridamide-Like Molecules

Thioviridamide (**15**) is an apoptosis-inducing compound that was isolated from *Streptomyces olivoviridis* during a screen for cytotoxic compounds (Hayakawa et al., 2006a). The solution of its structure revealed that it is a very complex peptide natural product (Figure 58; Hayakawa et al., 2006b), and heterologous expression of the gene cluster confirmed that it was a RiPP. After the gene cluster was published, heterologous expression and genome mining efforts have revealed other similar thioviridamide-like compounds, including JBIR-140 (**16**; Izumikawa et al., 2015), thioholgamide (also known as neothioviridamide; Kawahara et al., 2018; Kjaerulff et al., 2017), prethioviridamide (Izawa et al., 2017), thioalbamide (**18**), thiostreptamide S87, and thiostreptamide S4 (**17**; Frattaruolo et al., 2017). The focus of this thesis is thiostreptamide S4 (**17**; Figure 58).

3.1.1.2.   Thiostreptamide S4

The work presented in this thesis directly builds on the work done in the Truman group by Luca Frattaruolo and colleagues. In this work thioviridamide-like clusters were identified across multiple strains and one from *Streptomyces* sp. NRRL S-4 was TAR cloned to produce pCAPtsa (Figure 59; Frattaruolo et al., 2017). Heterologous expression of this cluster in *S. coelicolor* M1146 produced mature thiostreptamide S4 (**17**), providing proof that the entire gene cluster had been captured. Additional work assessing production conditions revealed that solid btmPM provided the best production levels from the heterologous expressed cluster, comparable to those from the wild type strain (Frattaruolo et al., 2017). Studies have also showed that the N-terminus of *S. olivoviridis* metabolite, thioviridamide (**15**), could result from the reaction of acetone with a pyruvyl group, and is therefore an artefact of extraction (Frattaruolo et al., 2017; Izawa et al., 2017). This suggests that a pyruvyl group is the true N-terminus of the molecule produced by *S. olivoviridis*, as in
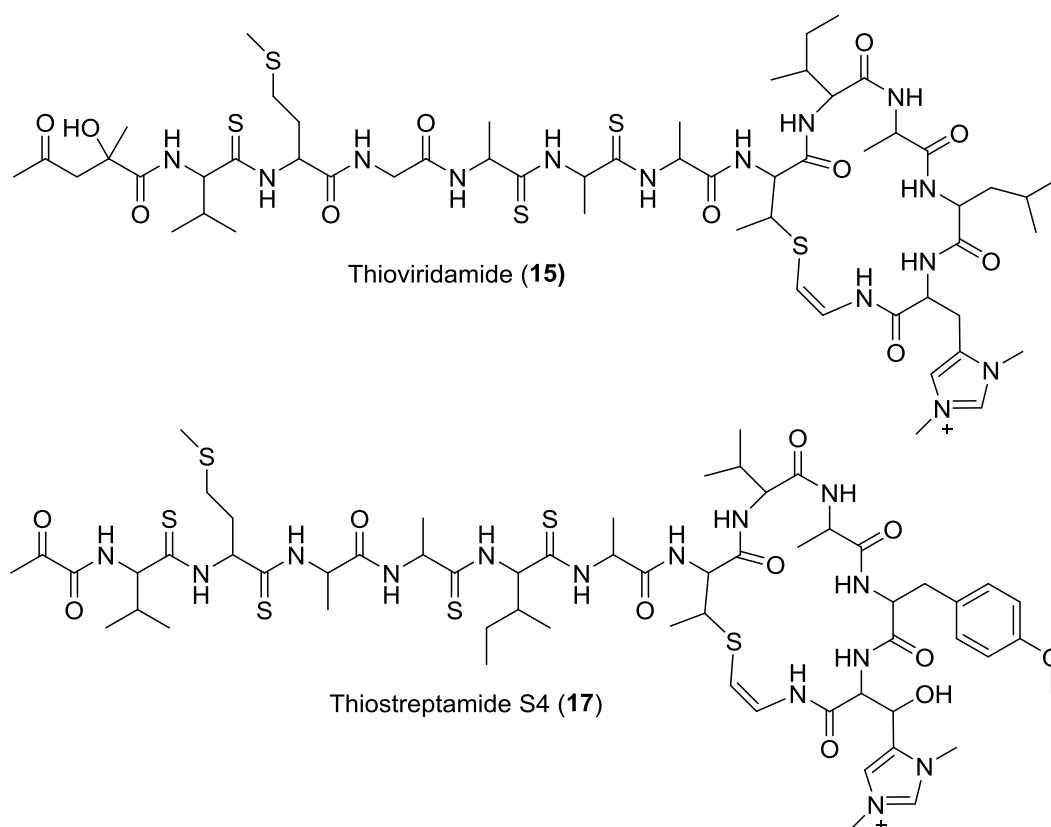
Figure 58. The structure of thioviridamide (**15**) and thiostreptamide S4 (**17**; Frattaruolo et al., 2017; Hayakawa et al., 2006b).
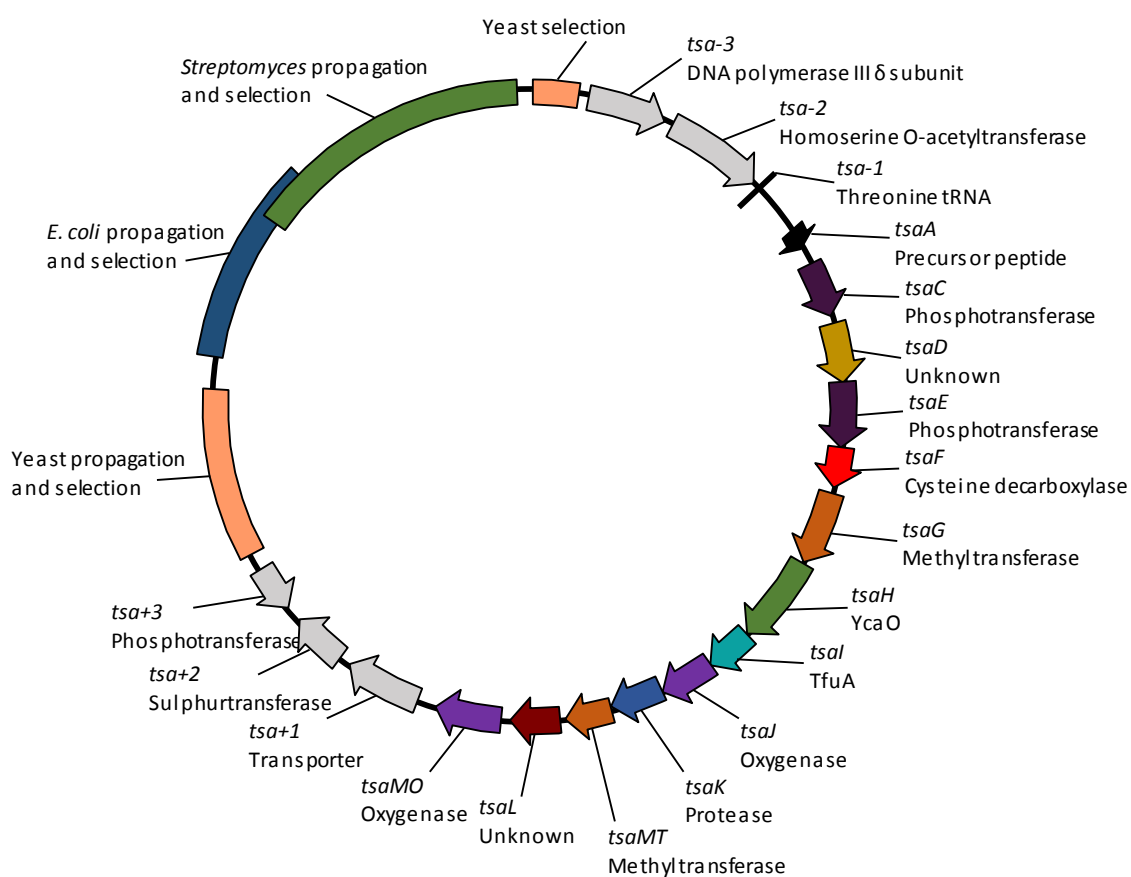
Figure 59. Plasmid map of pCAPtsa (29881 bp). Genes are labelled as published (Frattaruolo et al., 2017).

thiostreptamide S4 (**17**).

3.1.2. Aims of this Chapter

3.1.2.1. Hypothesis

Thiostreptamide S4 (**17**) contains many interesting posttranslational modifications. An N-terminal pyruvyl group, thioamide bonds, an *S*-[(*Z*)-2-aminovinyl](-(3*S*)-3-methyl)-D-cysteine (Avi(Me)Cys) macrocycle, a tyrosine O-methylation, two histidine N-methylations, and a histidine β-hydroxylation are all post-translationally installed. Whilst many of these features are interesting due to their structure and the possible influence they have on bioactivity, there is also the potential for biosynthetic novelty. For example, the histidine methylations are a new modification seen in RiPPs, whilst the installation of thioamide bonds were biosynthetically uncharacterised at the start of this project. Therefore, it was hypothesised that gaining an understanding of the biosynthesis could reveal new biosynthetic mechanisms and machinery involved in RiPP maturation. This could then inform future cluster engineering projects and genome mining for new RiPPs with related biosynthetic machinery.

3.1.2.2. Objectives

There were two major objectives of this chapter:

1) Use the pCAPtsa heterologous expression system to understand thiostreptamide S4 (**17**) biosynthesis

2) Use this information to carry out preliminary engineering and genome mining investigations.

3.2. Results and Discussion

3.2.1.   Understanding the Biosynthesis

3.2.1.1.   Initial Cluster Analysis

To start, an analysis of genes associated with the thiostreptamide S4 (**17**) cluster was carried out using BLASTP (Altschul et al., 1990) and Phyre2 (Kelley et al., 2015). BLASTP looks for proteins similar to the query sequence, and conserved domains within the query sequence. Phyre2 predicts the query's secondary structure, aligns this to known protein structures, and models the structure of the query based on high scoring alignments. The genes submitted to this analysis were based on the previous bioinformatically identified gene cluster boundaries (Figure 60; Frattaruolo et al., 2017). Table 3 and Table 4 show selected BLASTP and Phyre2 results respectively. The selected results (apart from the precursor peptide) ignore identified thioviridamide-like clusters and are high-scoring results that give a good clue towards activity.

It is immediately apparent that many of the predicted proteins produced by the cluster correlate well with the biosynthetic steps that need to be rationalised. For example, both TsaG and TsaMT are
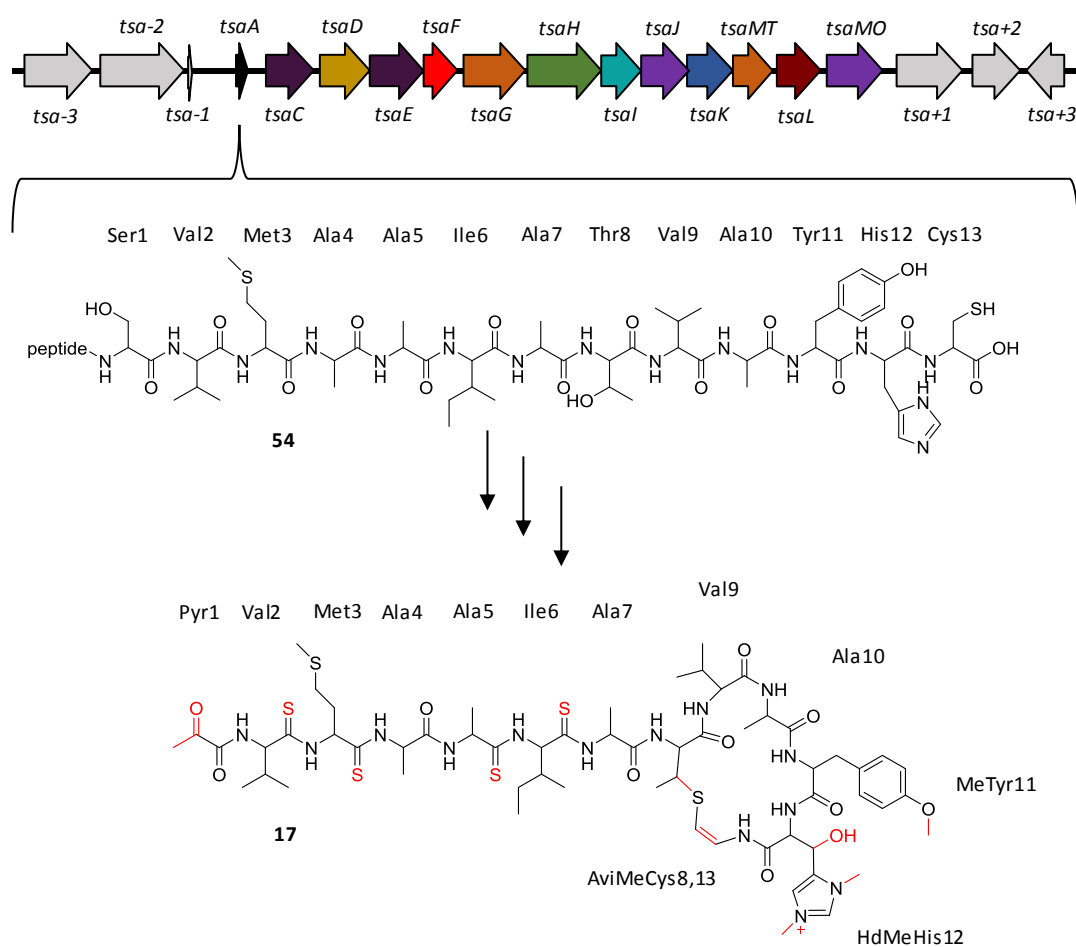
Figure 60. Thiostreptamide S4 (**17**) gene cluster. "Peptide" represents the leader peptide.

Table 3. Selected BLASTP results for each thiostreptamide S4 (**17**) gene. Other than the precursor peptide, the displayed results were chosen from organisms that do not contain an identified thioviridamide-like gene cluster (Frattaruolo et al., 2017).

Query gene	Homologous gene annotation	Organism	Coverage (%)	Identity (%)	Conserved domain
TsaA	BAN83916.1 thioviridamide family RiPP peptide	*Streptomyces olivoviridis*	98	49	None
TsaC	BAZ29304.1 aminoglycoside phosphotransferase	*Cylindrospermum sp. NIES-4074*	96	27	APH ChoK like
TsaD	WP_057178057.1 hypothetical protein	*Cylindrospermopsis sp. CR12*	94	33	None
TsaE	WP_084269265.1 hypothetical protein	*Mycobacterium avium*	12	52	None
TsaF	ADD44321.1 phosphopantothenoylcysteine decarboxylase	*Stackebrandtia nassauensis DSM 44728*	82	39	Flavoprotein
TsaG	ETR69295.1 protein arginine N-methyltransferase	*Candidatus Magnetoglobus multicellularis str. Araruama*	75	29	AdoMet MTase
TsaH	WP_007314232.1 YcaO-related McrA-glycine thioamidation protein	*Methanolinea tarda*	92	38	YcaO
TsaI	WP_044449638.1 TfuA-related McrA-glycine thioamidation protein	*Mastigocladus laminosus*	88	50	TfuA
TsaJ	WP_079103273.1 SDR family NAD(P)-dependent oxidoreductase	*Streptomyces aurantiacus*	94	49	2OG-Fe(II) oxygenase
TsaK	AMY09266.1 Papain family cysteine protease	*Luteitalea pratensis*	87	36	Peptidase C1
TsaMT	WP_084229406.1 class I SAM-dependent methyltransferase	*Mycobacterium sherrisii*	91	38	UbiG
TsaL	WP_093259625.1 hypothetical protein	*Thermostaphylospora chromogena*	73	36	None
TsaMO	WP_067815640.1 LLM class flavin-dependent oxidoreductase	*Actinomadura kijaniata*	97	58	Alkanal monooxygenase

homologues of SAM dependant methyltransferases, and both the tyrosine and the histidine are methylated. Some correlations were less clear, for example TsaJ and TsaMO are both oxidoreductases, and yet only one of these is likely to be responsible for the β-hydroxylation on the histidine. In other cases, there were no obvious candidate enzymes, for example there were no lanthionine bond-forming enzymes, as in epidermin biosynthesis (Bierbaum, 1996), or linaridin-like enzymes, as in cypemycin biosynthesis (Claesen and Bibb, 2010), that could be responsible for forming the AviMeCys residue. Therefore, it was clear that further experiments would be necessary to identify these proteins.

Table 4. Selected Phyre2 results for each thiostreptamide S4 (17) gene.

Query gene	PDB molecule match	Coverage (%)	Confidence (%)[a]	Identity (%)[b]
TsaA	Transcriptional repressor CytR	53	29.3	24
TsaC	Aminoglycoside phosphotransferase	82	99.9	15
TsaD	Type III effector HopA1	59	100.0	18
TsaE	Macrolide 2'-phosphotransferase	36	95.2	25
TsaF	Epidermin decarboxylase	89	100.0	25
TsaG	Protein arginine N-methyltransferase 2	79	100.0	22
TsaH	YcaO	91	100.0	27
TsaI	Plectin	47	99.6	14
TsaJ	CytC3	82	100.0	27
TsaK	Cysteine protease	88	100.0	23
TsaMT	Phosphoethanolamine n-methyltransferase 2	68	99.9	22
TsaL	Copper-containing nitrite reductase	26	41.9	9
TsaMO	3,6-diketocamphane 1,6 monooxygenase	97	100.0	17

[a]Confidence is a measure of how good the secondary structure alignment is
[b]Primary sequence homology

3.2.1.2.   Deletion and Complementation of Biosynthetic Genes

There were several approaches that could have been taken to further analyse the biosynthesis of thiostreptamide S4 (17). A very informative approach is to reconstitute the pathway *in vitro*, as was carried out with the core scaffold of the thiopeptide, thiomuracin (Hudson et al., 2015). The lack of a cellular background makes metabolomic analysis of the intermediates much easier. Endogenous proteases can break down RiPP intermediates, which are often short unstructured peptides; *in vitro* analysis also avoids this issue, enhancing the chance of detecting intermediates. However, this method was not chosen because it also represents a high risk. The difficulty of purifying some proteins in soluble form causes unreliable results. For example, YcaO enzymes (of which there is one in the thiostreptamide S4 (17) pathway) are often difficult to purify in a soluble active form (Franz et al., 2017). Once purified, reaction conditions with correct cofactors must be found, which is not trivial

when working with uncharacterised enzymes. Depending on its biosynthetic timing, a single non-functional enzyme could completely prevent *in vitro* reconstitution of the biosynthetic pathway.

Two other possible approaches were available. Deletions could be made in *Streptomyces* sp. NRRL S-4, the wild type producer of thiostreptamide S4 (**17**), or deletions could be made in pCAPtsa, the plasmid containing the thiostreptamide S4 (**17**) gene cluster, and tested in *S. coelicolor* M1146. Whilst *S.* sp. NRRL S-4 can be manipulated (Frattaruolo et al., 2017), deletions in *Streptomyces* spp. are slow and can be low efficiency. It was decided that it would be easier to delete genes in pCAPtsa by PCR targeting (Gust et al., 2003) in *E. coli* and then test the modified clusters in *S. coelicolor* M1146.

Gene cluster deletions can confirm whether genes are involved in the biosynthesis. However, the exact role of genes can often be harder to ascertain from these results, compared to those from *in vitro* reconstitution. Deletion of genes that act early in RiPP pathways can often result in short unstructured peptides that can be rapidly digested by endogenous proteases and modified by acetylation (Crone et al., 2016). This means that the function of the deleted genes often must be pieced together from multiple possibly-modified breakdown products of intermediates. Several metabolomic tools are available to facilitate this analysis, as used in the bottromycin study (see section 2.2.2.1). Untargeted metabolomics can be conducted using statistical packages to identify all detected differences in the metabolite profile between two cultures. This helps find the breakdown products of intermediates. This removes the need to predict the intermediate expected in advance and allows for the detection of compounds that are produced in very low quantities. Mass spectral networking can also be performed to identify relationships between the compounds detected in cultures (Wang et al., 2016). This does, however, rely on good MS^2 fragmentation data.

The workflow applied to every deletion was as follows. PCR targeting was used for the deletions (Gust et al., 2003). The deletion of each gene within the cluster replaced the gene with an in-frame 81 bp sequence that retained the original start and stop codons. Each deletion (except *tvaA*) that affected production was also complemented, which also provided an opportunity to assess the annotated start codon of each gene. Complementations were carried out by expressing the gene from the strong constitutive promoter PermE* (Bibb et al., 1994) in pIJ10257 (Hong et al., 2005). This plasmid integrates into a φBT1 site in the *S. coelicolor* M1146 genome. Initial complementation attempts used the annotated start codon; other start codons that had to be tested are mentioned in the text. Each deletion construct, and the subsequent complementation, was tested in *S. coelicolor* M1146 in solid bottromycin (**1**) production media plates (btmPM; Figure 61). The extraction process did not distinguish between extracellular and intracellular metabolites. Where the predicted intermediate could not be detected, untargeted metabolomics was carried out to assess what was

a

b
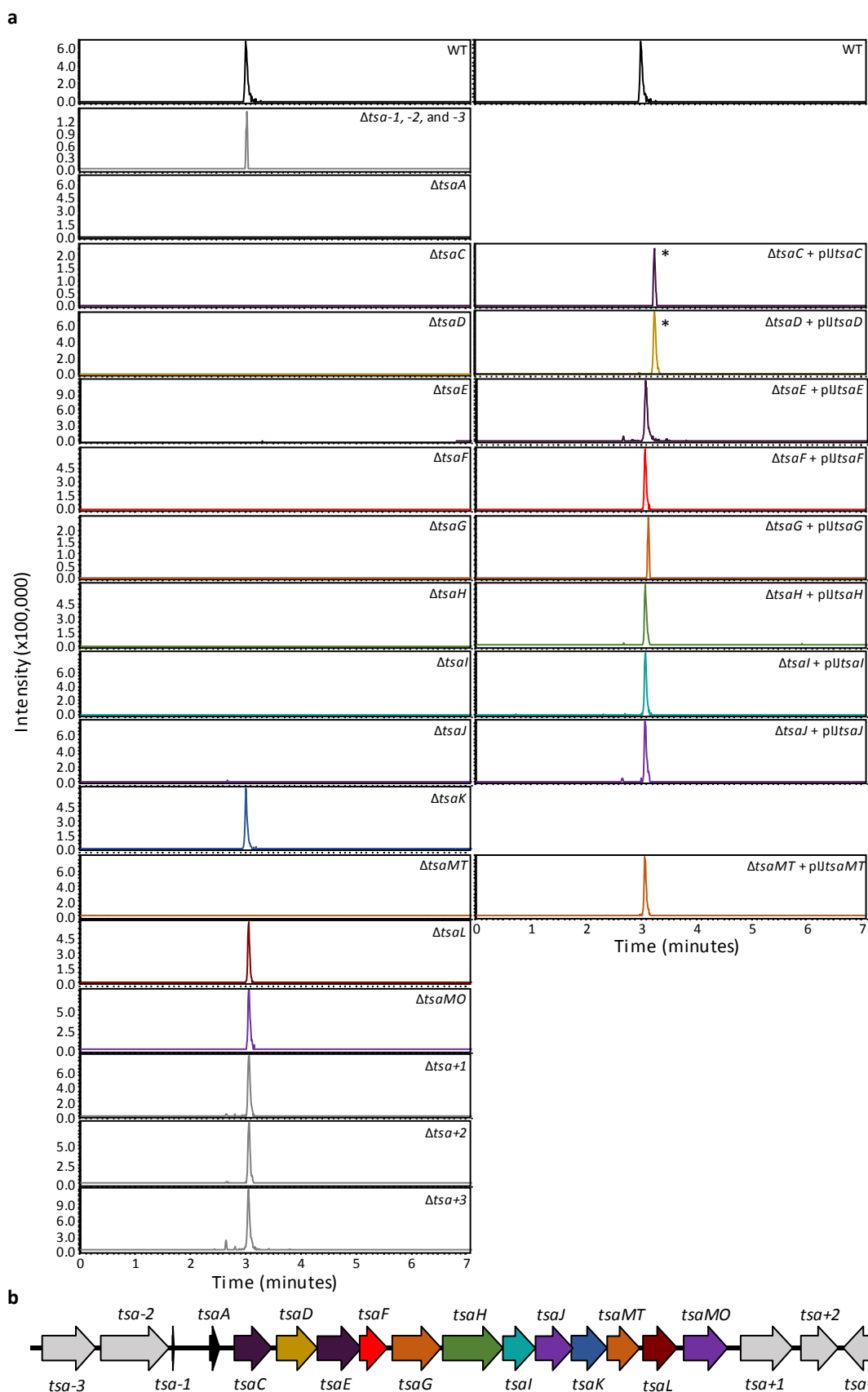
Figure 61. a. LC-MS analysis of each gene deletion cluster and their successful complementations. An EIC of thiostreptamide S4 (**17**; *m/z* 1377.55) is shown. The retention time in the Δ*tsaC* + pIJ*tsaC* and Δ*tsaD* + pIJ*tsaD*, peaks marked with an asterisk, are shifted due to the addition of a guard to the column. Results are not quantitative due to varying MS sensitivity. **b.** The genes captured on pCAPtsa.

being produced. The in-depth analysis of these deletion experiments is explained over the rest of this section.

3.2.1.3.   Establishing the Cluster Boundaries

The region captured whilst TAR cloning the gene cluster covered a larger region than the published gene cluster (*tsaA – tsaMO*; Figure 61; Frattaruolo et al., 2017). Three genes, *tsa-3*, *tsa-2*, and *tsa-1*, were captured upstream from the predicted gene cluster and were predicted to encode a DNA polymerase III δ subunit, a homoserine O-acetyltransferase, and a threonine tRNA respectively. Downstream from the predicted gene cluster three extra ORFs were present, *tsa+1*, *tsa+2*, and *tsa+3*, predicted to encode a transporter, sulphurtransferase, and serine/threonine protein kinase respectively.

The three upstream genes were deleted together in a single step using PCR targeting. The non-coding region between the tRNA and *tsaA* is 755 bp, and 255 bp of this region was also deleted in this experiment. The remaining 500 bp region was predicted to be large enough to feature the *tsaA* promoter. Thiostreptamide S4 (**17**) was still produced by *S. coelicolor* M1146 carrying the cluster lacking these genes, confirming that they do not have a vital biosynthetic role. There was, however, a very significant drop in production (Figure 62). It is unclear what the reason for this could be. It is unlikely either the genes or the tRNA have a regulatory role, however it is possible that there are some long range regulatory recognition elements in the DNA that were removed in this deletion.

Deletion of *tsaMO* and the downstream genes had no effect on production of thiostreptamide S4 (**17**). This includes the deletion of the sulphurtransferase. The gene is not conserved amongst all thioviridamide-like gene clusters, however in archaea it has been noted that sulphur metabolism
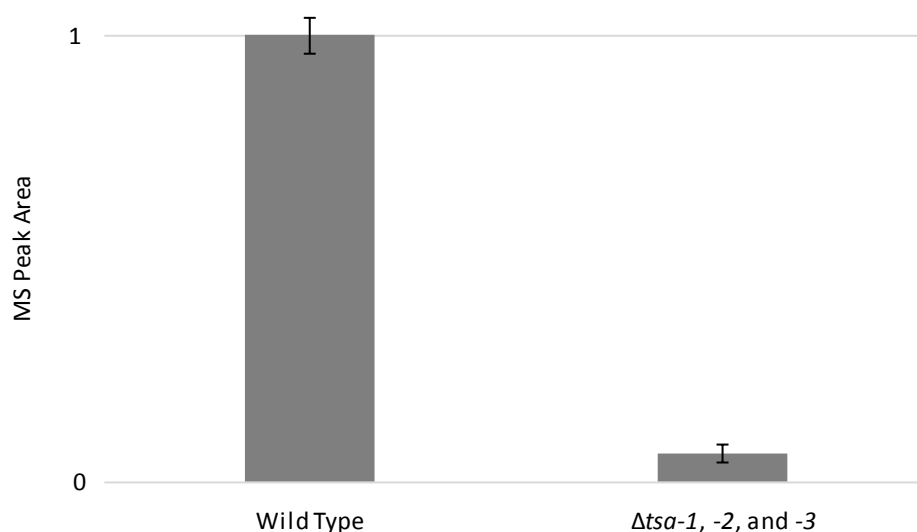
Figure 62. Production of thiostreptamide S4 (**17**; *m/z* 1377.55) by *S. coelicolor* M1146 carrying either the wild type cluster, or the Δ*tsa-1*, *-2*, and *-3* cluster, measured by mass intensity.
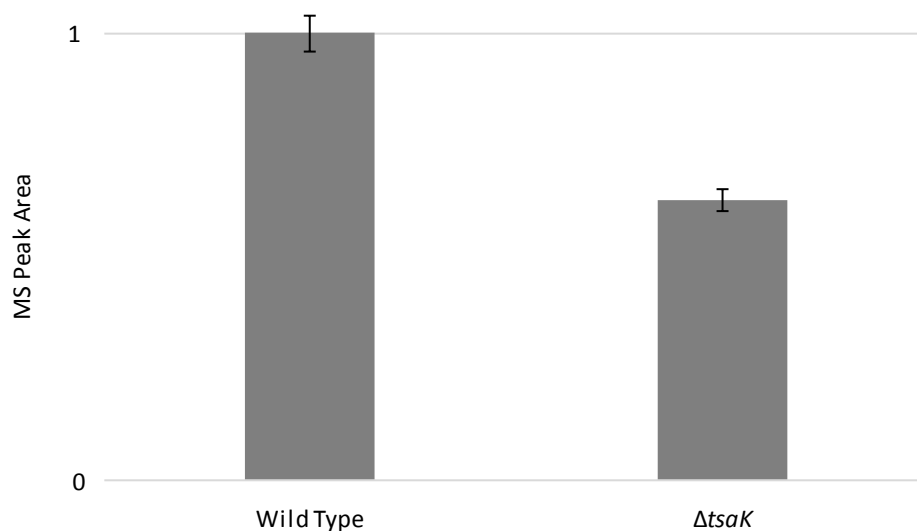
Figure 63. Production of thiostreptamide S4 (**17**; *m/z* 1377.55) by *S. coelicolor* M1146 carrying either the wild type cluster, or the Δ*tsaK* cluster, measured by mass intensity.

genes are often associated with YcaO and TfuA pairs involved in thioamidation (Nayak et al., 2017). A study presenting the *in vitro* reconstitution of thioamide bond installation in an archaeal system showed that either sodium sulphide or cysteine and a sulphurtransferase could be used to provide the sulphur for the thioamidation (Mahanta et al., 2018). Therefore, the sulphurtransferase associated with the thiostreptamide S4 (**17**) pathway may play a role in liberating sulphur. This may only be necessary in certain conditions and is therefore not essential for biosynthesis.

When *tsaK* is deleted, production of thiostreptamide S4 (**17**) is not abolished, although it is slightly reduced (Figure 63). This was surprising given that TsaK homologues are encoded in other thioviridamide-like clusters, which meant that it was predicted to be involved in biosynthesis in some way. TsaK contains a clear cysteine protease domain according to BLASTP domain analysis, and Phyre2 aligns the secondary structure with very high confidence to the cysteine protease, xylellain (Leite et al., 2013). LanT cysteine proteases remove the leader peptides of class I and II lanthipeptides, for example LctT in lacticin 481 biosynthesis (Furgerson Ihnken et al., 2008), and NukT in nukacin ISK-1 biosynthesis (Nishie et al., 2009). Cleavage of the leader peptides in bacteriocins often follows a double glycine (GG) motif (Dirix et al., 2004). The thioviridamide-like precursor peptides contain a QG motif. This glutamine is a significant change from glycine. LctT cysteine protease also requires a glutamic acid residue -8 from the core peptide for recognition (Furgerson Ihnken et al., 2008). Similarly, the thioviridamide-like precursor peptides contain a double acid, DE or EE, motif at -6 and -5 from the core peptide (Figure 64). The conservation of TsaK, the fact that it is predicted to be similar to LanT proteins, and the similar leader peptide motifs all suggest that TsaK is the cysteine protease that is responsible for leader peptide cleavage. As its deletion has little effect on production, it seems likely that peptidases in the heterologous host, *S. coelicolor* M1146, can

```
A. alba         GFSEEELQRFLEEKAGLTAASGVQQSVIGFAVTIAVHC
M. laminosus    GVSPEELQEFLENKAGLSPEEESQGSPMAAAVSIAYHC
N. potens       GLDEAELQTLLESKSGISADEDAQGSVMAAAASVAAHC
S. olivoviridis GLDAAELQNFLEEKSGISPDEEAQGSVMAAAASIALHC
S. NRRL S-87    GVSPEELQAFLEEKAGISPDEEAQGSVMAAAATVAFHC
S. sparsogenes  GVSPEELQAFLEEKAGISPDEEAQGSVMAAAATVAFHC
S. MUSC 125     GVSPEELQAFLEEKAGISPDEEAQGSVMAAAATVAFHC
S. MUSC 14      GVSPEELQAFLEEKAGISPDEEAQGSVMAAAATVAFHC
S. malaysiense  GVSPEELQAFLEEKAGISPDEEAQGSVMAAAATVAFHC
S. NRRL S-4     GVSPEDLQAFLEEKAGISPDEEAQGSVMAAIATVAYHC
S. CNB091       GVSPEDLQAFLEEKAGISPDEEAQGSVMAAIATVAYHC
S. pacifica     GLSPEELQSFLEEKVGVTAEEGVQGTVGGLLVTPATHC
N. fuscirosea   GVSPEDLQQFLEEQVGLSPDEGIQGSFTGIIVTAGVHC
M. eburnea      GVSSEELQKFLEEQVGMNPDEGVQGTFVSVVVTPATHC
```

Figure 64. Aligned precursor peptides encoded in thioviridamide-like gene clusters. Conserved amino acids possibly involved in cysteine protease recognition are shown in red and green. The core peptide is in bold.

complement its deletion. There are two bioinformatically predicted class I lanthipeptide gene clusters in *S. coelicolor* (Marsh et al., 2010), and one of the proteases that processes them may be able process the thiostreptamide S4 (**17**) intermediate. Alternatively, it is common for RiPPs to rely on non-cluster associated endogenous proteases to cleave their leader peptides, for example class III lanthipeptides are usually not associated with peptidases and often rely on stepwise N-terminal digestion to remove the leader peptide (Völler et al., 2012). SapB is an example in *S. coelicolor* of a lanthipeptide where no identified protease is associated with the gene cluster (Kodani et al., 2004). The endogenous protease that SapB biosynthesis relies on may also be able to remove the leader peptide of thiostreptamide S4 (**17**). Therefore, TsaK may only be necessary when the metabolite is produced in the native host or under different regulatory conditions, when different background proteases are present.

Interestingly, deletion of *tsaL* also had no effect on production. *tsaL*-like genes are conserved amongst almost all thioviridamide-like gene clusters (Frattaruolo et al., 2017). BLASTP analysis showed that TsaL has high identity with many proteins, most of which are annotated as "hypothetical protein". The few that are not annotated as "hypothetical protein" have been annotated as membrane proteins, although this appears to be without experimental characterisation. Whilst Phyre2 is unable to find a crystallised structural homologue, it does strongly predict four transmembrane helices. The lack of identified homologues for TsaL makes it hard to predict function. No apparent role for such a well conserved gene could be due to the expression of this gene cluster in a non-native background, or alternatively it could be associated by chance, as some thioviridamide-like gene clusters do not contain a homologue, for example the cluster in *Nocardiopsis potens* (Frattaruolo et al., 2017). Its presence or absence from clusters does not seem to correlate with any features of the predicted structures of the mature molecules. Therefore, if it does play a role it may be non-catalytic, such as regulatory or transport-related.

Both *tsaK* and *tsaL* are well conserved amongst thioviridamide-like gene clusters, and as such cannot be completely discounted from having a role in thiostreptamide S4 (**17**) production. It is not uncommon for proteolysis to be coupled to export in bacteriocins (Havarstein et al., 1995), so it is possible that TsaL is an exporter (based on its predicted transmembrane helices) that couples export to proteolysis with TsaK.

3.2.1.4.   Macrocycle Methylations and Hydroxylation

Based on bioinformatic data, TsaG, TsaJ, and TsaMT were predicted to install the methylations, hydroxylations, and methylations respectively. Predicted masses were looked for during the LC-MS assessment of these deletions. These predictions matched the loss of one, two, or three methyl groups, and one hydroxyl group. Therefore, the *m/z* values that were searched for were 1377.55 (fully mature), 1363.53 (-1 methyl), 1361.55 (-1 hydroxyl), 1347.54 (-1 methyl and -1 hydroxyl), 1349.52 (-2 methyl), 1333.52 (-2 methyl and -1 hydroxyl), 1335.50 (-3 methyl), and 1319.50 (-3 methyl and -1 hydroxyl). In total, 5 of these masses were detected: **17** (1377.55), **55** (1363.53), **56** (1361.55), **57** (1347.54), and **58** (1319.50), and the mass fragmentation data confirmed that the mass differences were on the macrocycle (Figure 65). **17**, **55**, and **56** are seen in the wild type cluster, **58** is seen in the Δ*tsaG* cluster, **56** and **57** are seen in the Δ*tsaJ* cluster, and **55** and **57** are seen in the Δ*tsaMT* cluster (Figure 66).

Deletion of *tsaMT* resulted in the loss of **17**, **56**, and **57** (Figure 66). Instead, **55** was produced, which lacks the tyrosine methylation but is otherwise identical to **17**, therefore confirming that TsaMT is the protein responsible for this modification. This is supported by the bioinformatic analysis, which show *tsaMT* has homology to class I SAM-dependant methyltransferases. Phyre2 predicts that it has high structural homology with the *Haemonchus contortus* phosphoethanolamine N-methyltransferase 2 (Lee and Jez, 2013). This methyltransferase is only found associated with thioviridamide-like clusters in *S.* sp. S4, *S.* sp. S-15, *S. mutomycini*, and *S.* CNB091; clusters which also contain a tyrosine at this location in their core peptides. The BLASTP domain analysis backs this up by linking it to the UbiG family of methyltransferases, which can catalyse the hydroxyl methylations in ubiquinol biosynthesis (Meganathan, 2001). The presence of the histidine hydroxylation and bis-methylation in **55** shows that the tyrosine methylation is not required for the histidine hydroxylase and methyltransferase to function.

S. coelicolor M1146 expressing the Δ*tsaJ* cluster did not produce **17**, whilst it did produce **56** and **57** (Figure 66). Molecules **56** and **57** are both fully mature except for the macrocycle modifications; both lack the histidine hydroxylation, and **57** also lacks the tyrosine methylation. This indicates that TsaJ is responsible for the histidine hydroxylation, and the presence of **56** shows that the histidine
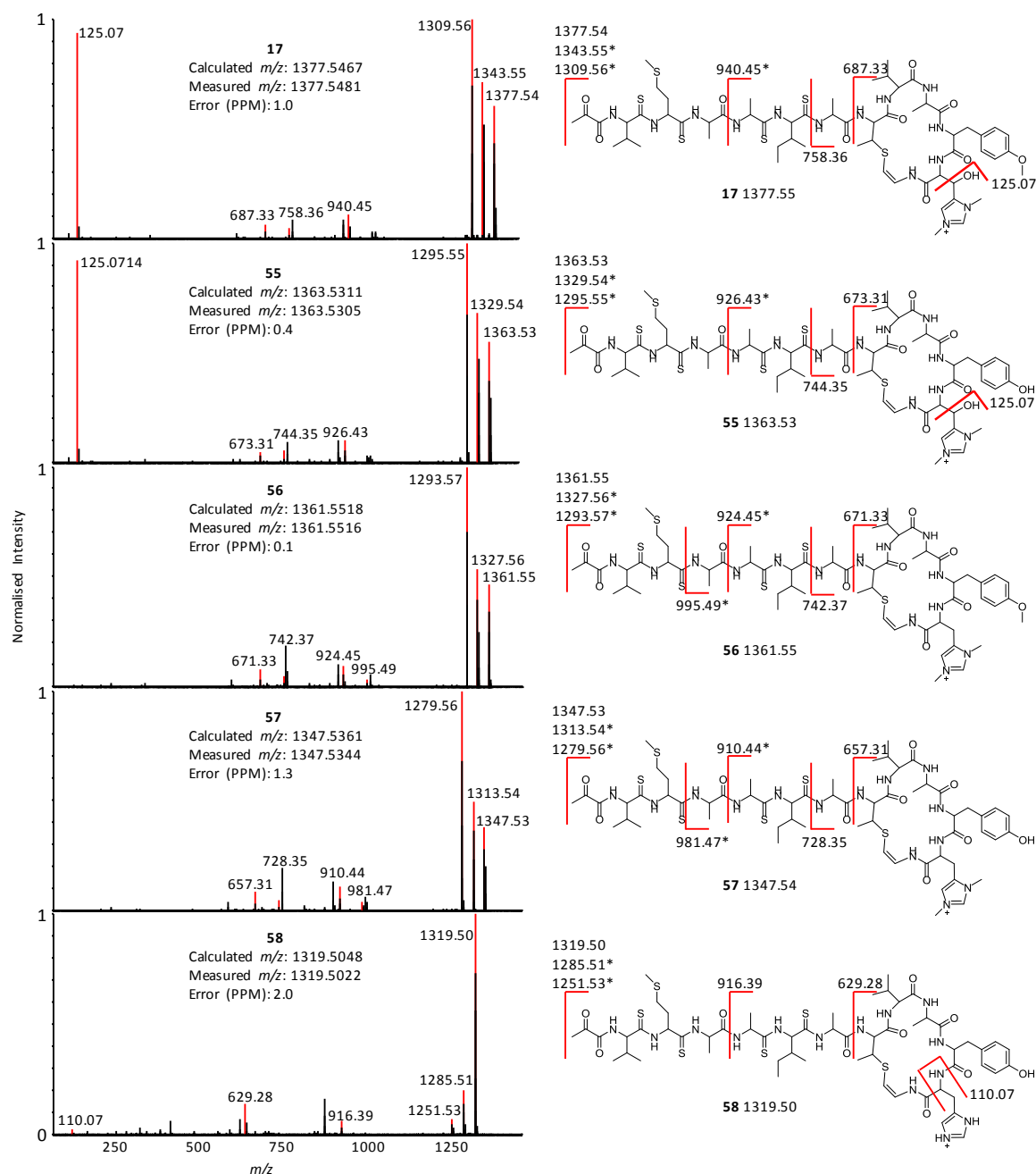
Figure 65. Mass fragmentation data for **17** and **55-58**. Fragments marked with an asterisk show a loss of 33.99, characteristic of the loss of SH_2 from thioamide bonds in fragmentation (see Figure 69). Exact mass measurements are shown within the mass spectra.

hydroxylation is not required for tyrosine methylation. Both BLASTP and Phyre2 analysis suggest that TsaJ has homology to nonheme Fe(II) and α-ketoglutarate dependent enzymes, which are well known to catalyse hydroxylation of unactivated carbon centres (Wu et al., 2016).

Deletion of *tsaG* abolished production of **17**, however the strain also did not produce the expected molecule: a version of **17** just lacking the histidine methylations. Instead it produced molecule **58**, a version of **17** that lacks all modifications to the macrocycle but is otherwise fully mature (Figure 66). This means the TsaJ-catalysed histidine hydroxylation and the TsaMT-catalysed tyrosine methylation
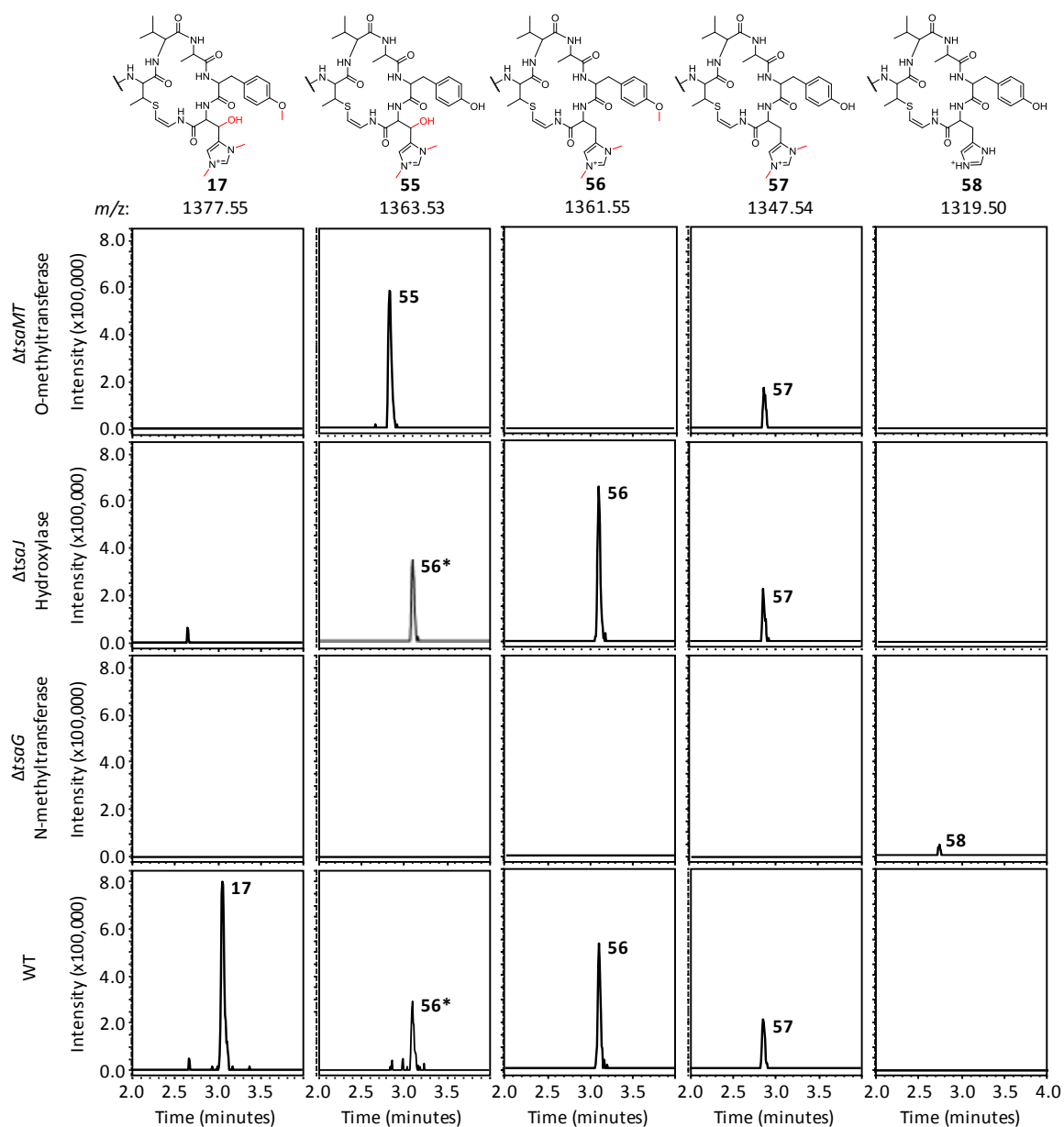
Figure 66. LC-MS analysis of *S. coelicolor* M1146 expressing the WT, Δ*tsaG*, Δ*tsaJ*, and Δ*tsaMT* clusters. EICs showing the varied methylation and hydroxylation patterns are shown. The structure of the macrocycle from each metabolite is shown above the relevant columns; in each case, MS data indicates that the rest of the molecule is identical to fully matured thiostreptamide S4 (**17**). The **56*** label indicates the +2 isotope peaks of **56**.

were both not installed when the histidine methylations were absent. Therefore, the histidine methylations are required for the histidine hydroxylation and the tyrosine methylation. It is not surprising that TsaJ and TsaMT are unable to recognise the molecule without the histidine methylations as the histidine methylations significantly change the macrocycle by introducing a permanent positive charge. TsaG has sequence homology with protein arginine N-methyltransferases. This is the first example of bis-methylated histidine in RiPP natural products, and it has now been shown that the enzyme responsible is very similar to protein arginine N-methyltransferases and likely acts in a similar way.
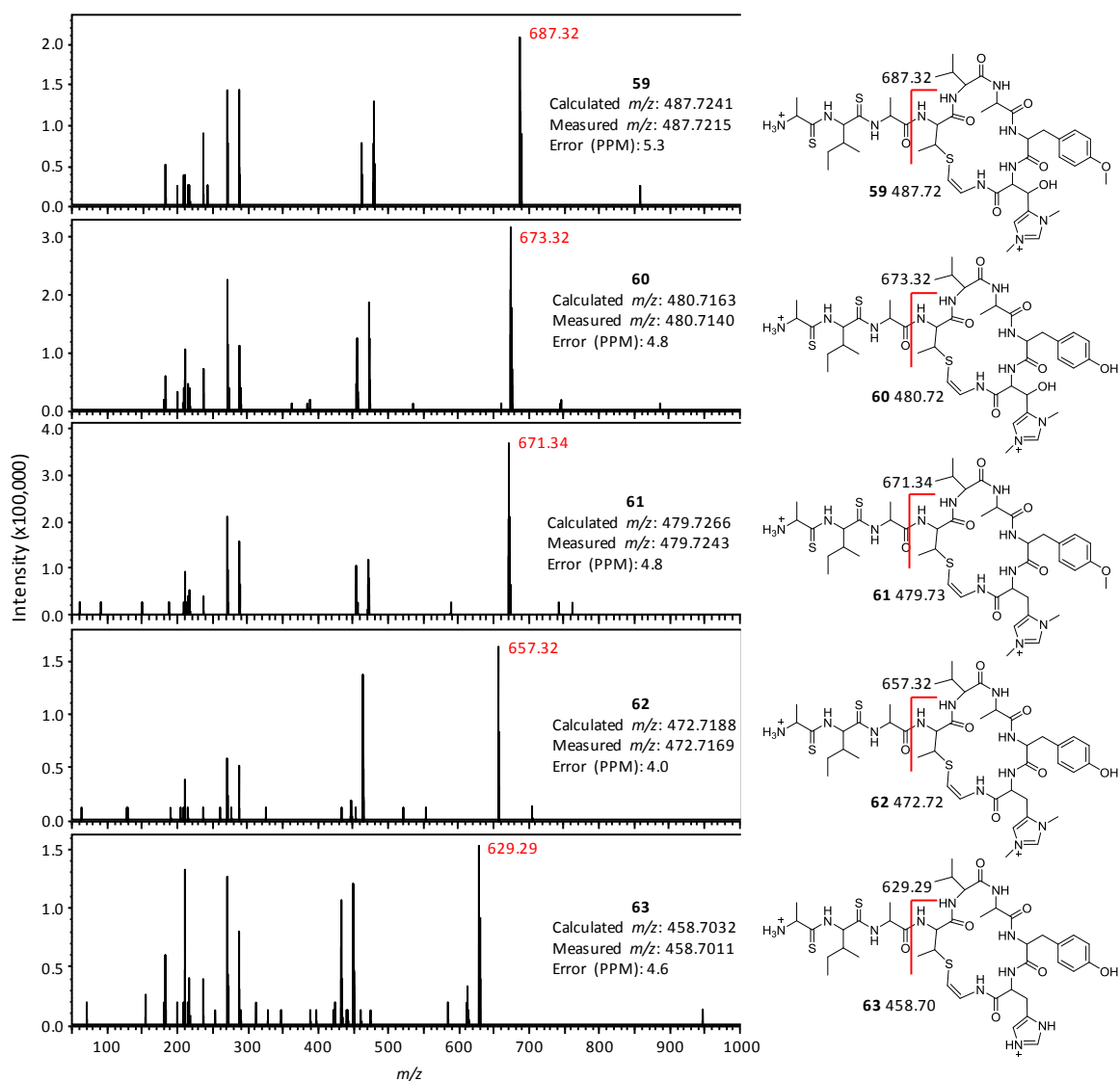
Figure 67. MS^2 data for **59-63**. Exact [M+2H]$^{2+}$ measurements are shown within the mass spectra and the characteristic macrocycle fragments are highlighted.
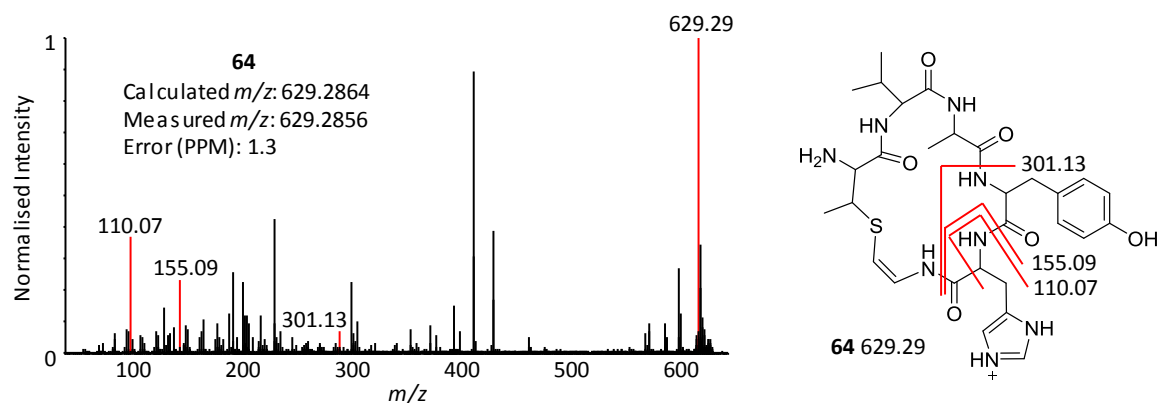
Figure 68. Mass fragmentation data for **64**. Exact mass measurement is shown within the mass spectrum.

These results show that the modifications to the macrocycle; histidine methylation, histidine hydroxylation, and tyrosine methylation, are among the final steps in thiostreptamide S4 (**17**)

biosynthesis and are installed by TsaG, TsaJ, and TsaMT, respectively. TsaG acts as a gatekeeper for TsaJ and TsaMT activity, meaning histidine hydroxylation and tyrosine methylation is only seen in the presence of the histidine methylations. TsaJ and TsaMT are tolerant enough to not need each other's modifications to be active.

To see if other metabolites of interest were seen, the characteristic fragmentation pattern of these metabolites was used to query the LC-MS^2 data from the wild type and Δ*tsaG*, *J*, and *MT* clusters. The macrocycle is often one of the main fragments seen in the deletions, and so the masses of the different macrocycle fragments seen in **17**, **55**, **56**, **57**, and **58** (687.33, 673.31, 671.33, 657.32, and 629.29 respectively) were used to search all fragmentation events in the LC-MS^2 data. This enabled the identification of six new metabolites, **59**-**64**. **59**-**63** were detected as doubly charged species, and whilst their fragmentation was poor, their structures could be proposed to be versions of **17** and **55**-**58** that had been hydrolysed between Ala4 and Ala5 (Figure 67). The bond between Ala4 and Ala5 is one of the only non-thioamidated peptide bond in the linear portion of the molecule, supporting that thioamide bonds protect molecules from proteolysis (Chen et al., 2017). **64** is a version of **58** in which the other non-thioamide bond in the molecule, between Ala7 and AviMeCys8,13, is hydrolysed to leave an unmodified macrocycle (Figure 68). The presence of these metabolites suggest that the macrocycle is highly resistant to proteolysis, a characteristic that is common in peptide macrocycles (White and Craik, 2016).

3.2.1.5.   Macrocycle and Pyruvyl Formation

Genes *tsaC-F* were deleted individually, and each cluster lost the ability to produce thiostreptamide S4 (**17**) in *S. coelicolor* M1146. The role of these genes in the pathway was unclear from bioinformatic data, which made it difficult to predict intermediates for these deletions, and so untargeted metabolomics was employed to detect any relevant metabolites being produced (Table 5). A series of M1146 strains containing the Δ*tsaC-F* clusters were compared to M1146 containing the wildtype cluster, the Δ*tsaA* cluster, and a media only control. After removing all metabolites present in the media and in M1146 expressing the Δ*tsaA* cluster many differences between the strains were identified (Table 5).

It was immediately apparent that several difficulties would need to be overcome in interpreting these data. Identifications by MS^2 fragmentation initially proved difficult due to the small size of the molecules (*m/z* < 600), which resulted in few fragments, and those that were present could not be accounted for by the simple loss of amino acid residue masses, as is common in peptide fragmentation. Many of the metabolites co-eluted, and some of the smaller metabolites showed up in the mass fragmentation of larger metabolites, suggesting source fragmentation. This meant that some of the metabolites were likely to be artefacts of the LC-MS method. Mass spectral networking

Table 5. Untargeted metabolomics with P-values lower than 1E-4 for the wild type cluster and clusters lacking *tsaC*, *tsaD*, *tsaE*, and *tsaF*. The value and intensity of shading in the cells reflects the MS intensity. Ion *m\z* cells are shaded green if their structure was proposed in this project.

Ion *m/z*	Ion RT	PVal	ΔtsaC	ΔtsaD	ΔtsaE	ΔtsaF	WT
1377.55	3.1	5.7E-16	0	0	0	0	696481
465.22	2.0	2.6E-15	0	0	823330	478565	126777
392.11	1.9	3.1E-15	783547	759679	830419	538018	428295
552.23	2.2	9.6E-14	0	0	271762	160966	0
370.12	1.9	1.9E-13	150997	136938	0	0	0
348.14	1.9	2.2E-13	168199	186354	0	0	0
493.73	2.0	2.6E-13	0	0	0	0	140435
487.73	1.6	2.1E-12	0	0	0	0	290551
479.73	1.6	2.3E-12	0	0	0	0	241452
1361.56	3.2	2.6E-12	0	0	0	0	318507
471.18	1.8	6.5E-12	0	94936	0	0	0
262.12	1.3	1.1E-11	89145	0	0	0	0
519.13	1.9	2.1E-11	192625	207592	247175	165182	142112
358.18	1.0	2.9E-11	0	0	78544	0	0
330.13	2.0	5.2E-11	48619	47760	521489	301795	0
503.15	1.9	7.0E-11	534560	588332	635199	406905	312217
453.16	2.0	9.3E-11	0	0	395727	254423	0
481.16	1.9	2.7E-10	325582	289713	324147	167661	183475
472.72	1.4	3.7E-10	0	0	0	0	248165
1347.55	2.9	3.9E-10	0	0	0	0	157561
364.15	1.9	4.0E-10	0	0	83880	0	0
469.16	1.9	1.1E-09	177134	185654	197363	114583	95596
333.07	2.6	2.4E-09	0	0	0	0	95864
531.16	2.3	4.3E-09	113319	106446	107331	0	0
259.09	1.9	4.9E-09	345602	319680	138403	63290	0
372.15	1.3	1.1E-08	0	0	0	64504	240572
330.13	2.2	1.2E-08	0	0	102379	0	0
480.23	1.6	1.3E-08	0	0	0	0	168832
352.12	1.9	3.8E-08	145120	162087	165483	97353	41173
305.08	2.2	1.1E-07	171183	173029	226425	127493	104898
676.25	1.1	1.8E-07	0	0	0	0	104863
564.25	1.3	3.4E-07	0	0	134702	21775	0
447.18	1.8	8.9E-06	152230	131107	145071	69117	0
479.21	1.5	1.8E-05	0	23672	0	0	104027
336.13	1.9	2.3E-05	131275	63697	0	0	0
283.13	1.4	2.4E-05	0	0	101784	0	0
616.18	3.4	3.0E-05	113118	21688	21543	0	100116
585.35	1.4	4.2E-05	81884	113695	32255	0	0
416.11	2.2	4.3E-05	114862	113974	129047	22101	27435
515.18	2.3	5.3E-05	168673	168939	193283	114637	53574
576.26	1.9	8.3E-05	0	0	103787	27383	0
360.23	1.4	8.8E-05	0	45420	128057	0	131936
619.17	3.3	9.1E-05	0	0	0	27907	99799

was employed to try and link related metabolites (Wang et al., 2016), however this was largely unhelpful, as it did not provide an obvious network of thiostreptamide S4-related metabolites.

It has previously been reported that molecules containing thioamide bonds can undergo fragmentation to lose SH$_2$; corresponding to a mass loss of 33.9877 Da (Larsson et al., 1973). This fragmentation does not break the backbone of the molecule. This signature loss can be seen very clearly in the fragmentation of thiostreptamide S4 (**17**; Figure 69). This tends to complicate the fragmentation of metabolites but can also be used as a tool to identify metabolites that contain
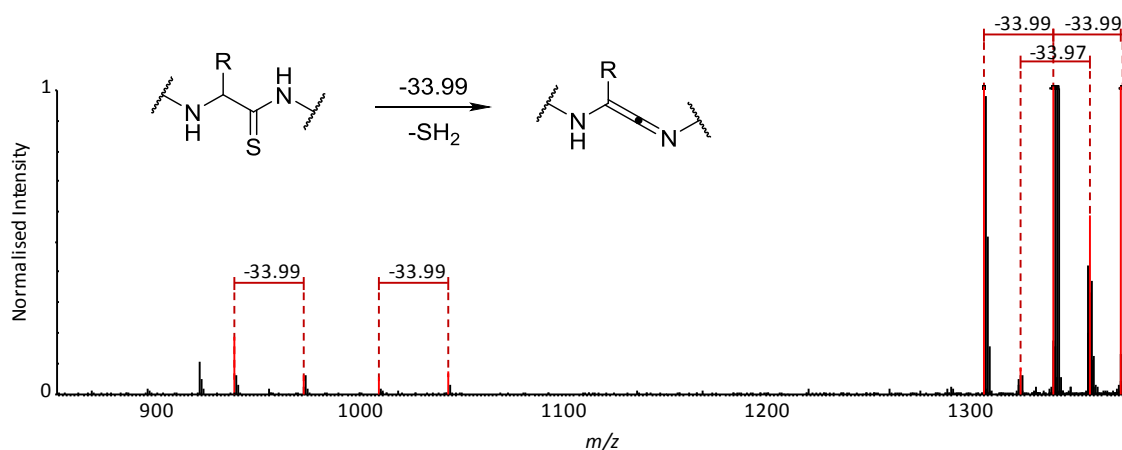
Figure 69. Selected MS fragmentation of thiostreptamide S4 (**17**). Loss of 33.99 Da characteristic of thioamide bond fragmentation is shown, along with the reaction schematic. The mass spectrum is zoomed for clarity.
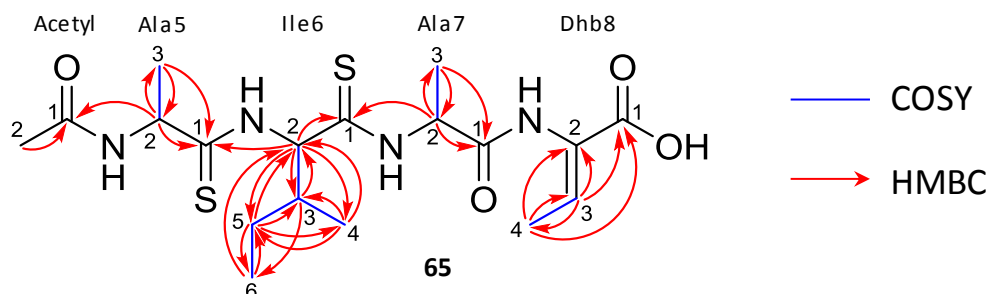
Figure 70. COSY and HMBC correlations for **65** in CD_3OD. Correlations for the NH and OH protons are not visible due to deuterium exchange with the solvent. See Table 6 for NMR assignments table.

thioamide bonds. This indicated that previously unidentified metabolites, with *m/z* of 552.23, 503.15, 465.22, 453.16, 392.11, 370.12, 348.14, 330.13, and 259.09, have thioamide bonds in their structure due to this signature fragmentation. This helps confirm that TsaC-F are not responsible for thioamide bond installation. The only other unexplained tailoring steps are macrocycle formation and N-terminal pyruvyl installation. Therefore, by process of elimination this means it is likely that TsaC-F are responsible for macrocycle formation and N-terminal pyruvyl installation.

This signature loss of 33.98 was used to target metabolites for detailed chemical characterisation. First, **65** (*m/z* 453.16), seen in Δ*tsaE* and Δ*tsaF* clusters, was targeted for purification due to its high production levels (Table 5). **65** was purified from *S. coelicolor* M1146 containing the Δ*tsaE* cluster, yielding 0.7 mg of pure compound. This was characterised by NMR. **65** was found to be the acetylated tetrapeptide AlaIleAlaDhb with thioamide bonds on either side of the isoleucine, with a mass of 430.17 Da (Figure 70 and Table 6). These modifications, and the fact it was detected as a sodium adduct in MS, explain the unusual fragmentation (Figure 71). Based on this structure, the structure of **66**, **67**, and **68** could also be proposed, as they were all acetylated short peptides based around the isoleucine and its thioamide bond (Figure 71).

Table 6. DEPTQ and proton NMR assignments for 65 in CD$_3$OD.

Residue	δC, mult.	δH, mult. (*J*, in Hz)
Acyl		
1	172.84, C	
2	22.63, CH$_3$	1.99 (s)
Ala5		
1	206.01, C	
2	56.72, CH	4.71 (q, 7.2)
3	21.05, CH$_3$	1.39 (d, 7.0)
NH		ND[a]
Ile6		
1	203.66, C	
2	68.85, CH	5.16 (d, 8.9)
3	41.04, CH	2.08 (m)
4	15.55, CH$_3$	0.95 (m)
5	26.21, CH$_2$	1.25, 1.64 (br m)
6	11.34, CH$_3$	ND[b]
NH		ND[a]
Ala7		
1	172.35, C	
2	55.81, CH	5.05 (q, 7.2)
3	17.48, CH$_3$	1.54 (d, 7.1)
NH		ND[a]
Dhb8		
1	ND[c]	
2	ND[c]	
3	136.35, CH	6.83 (q, 7.2)
4	14.39, CH$_3$	1.77 (d, 7.3)
NH		ND[a]
OH		ND[a]

a. The NH and OH protons are not visible due to deuterium exchange with the solvent.
b. The signal for this hydrogen is masked by a contaminant compound, however their COSY and HMBC correlations are visible.
c. Due to low signal these carbons are below the level of noise, however their HMBC correlations are visible (Figure 80).

The identification of 65 revealed that these metabolites could be thioamidated sodium adducts, resulting in hard to resolve fragmentation patterns. Once this was realised it was possible to suggest a structure for 69 (*m/z* 503.14), a metabolite seen in all Δ*tsaC-F* clusters and in the wildtype cluster (Figure 72). Further support for this structure was provided by later precursor peptide modifications that caused targeted changes to this metabolite, producing 70 and 71 (Figure 72; see section 3.2.2.2

a

b
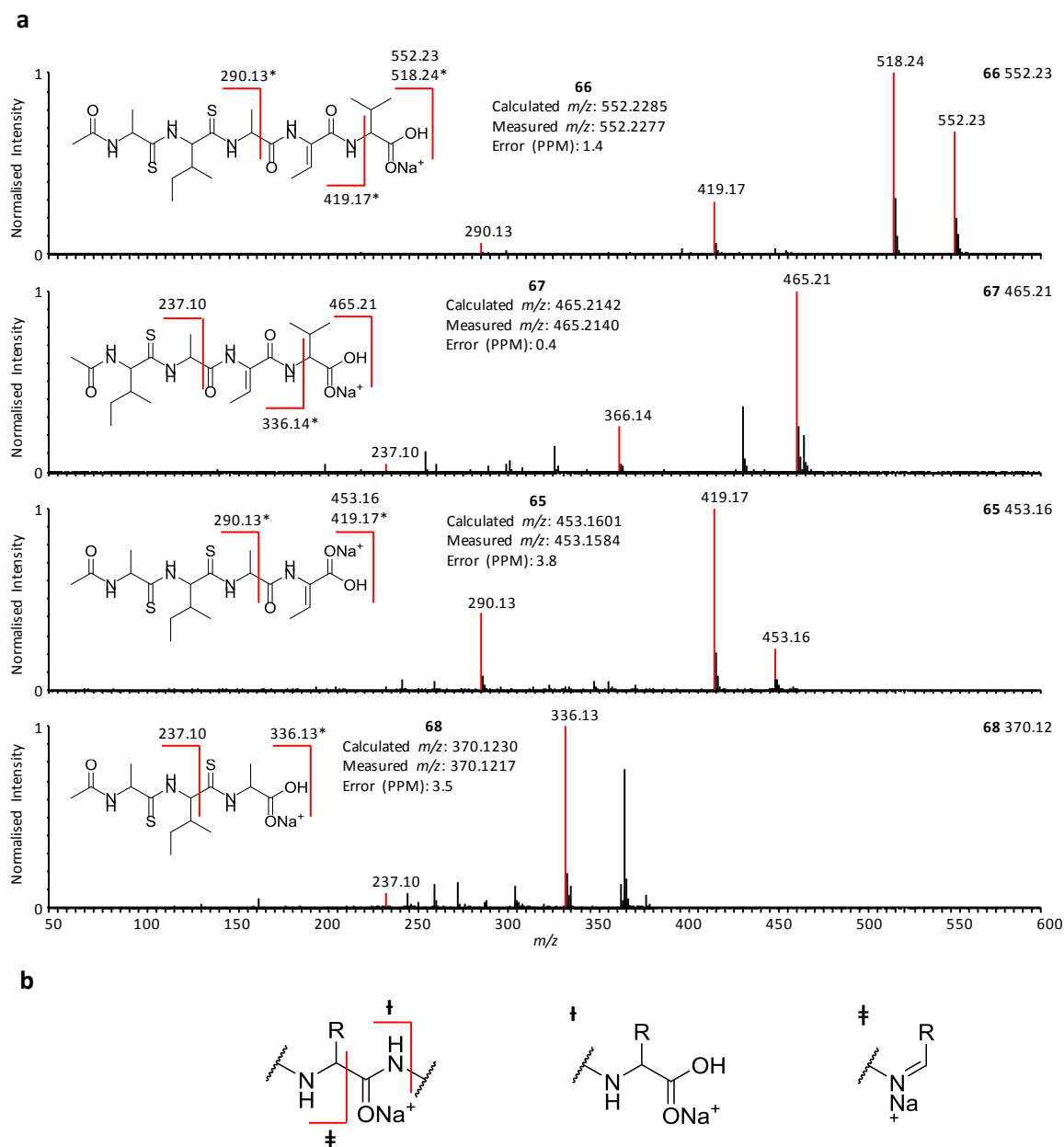
Figure 71. a. MS^2 fragmentation of **65-68** alongside predicted structures. The *m/z* of the parent molecule is shown in the top right of each chromatogram. Fragments marked with an asterisk show a loss of 33.99, characteristic of the loss of SH$_2$ from thioamide bonds in fragmentation. **b.** The unusual position of bond breakage common in the fragmentation of the sodium adduct of peptides (Grese et al., 1989; Newton and McLuckey, 2004).

for descriptions of S1T and M3I). Other metabolites that carried these -33.99 Da signatures in their MS^2 fragmentation were seen, suggesting they also contained a thioamide bond, however these were produced in quantities too low to be readily purified, and the MS^2 data was insufficient for structural characterisation.

As **66**, **69**, **65**, and **68** are all breakdown products, they support the proposal that unmodified peptides are very readily digested in *S. coelicolor* M1146. This is unsurprising as *S. coelicolor* contains a very large complement of proteases (Nagy et al., 2003). Additionally, the structures of **66**, **69**, **65**,
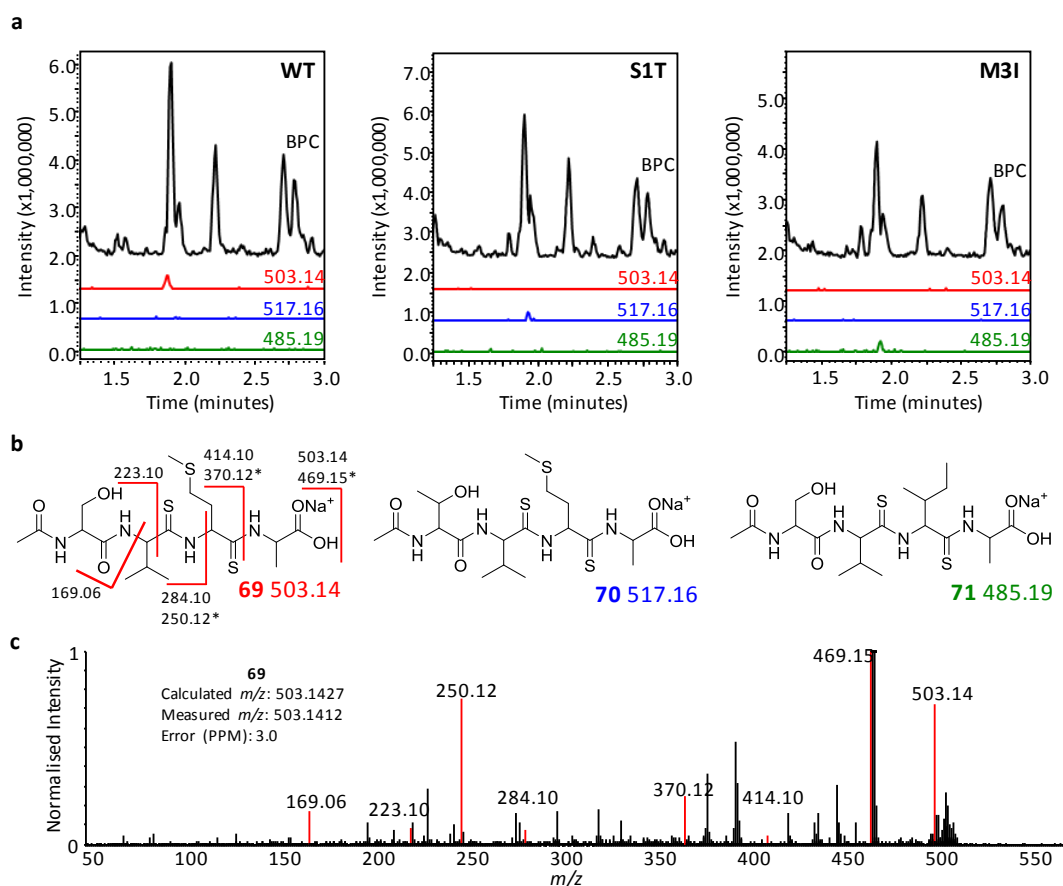
Figure 72. a. LC-MS analysis of *S. coelicolor* M1146 expressing pCAPtsa, pCAPtsa S1T and pCAPtsaM3I. The BPC and EICs of **69-71**, at an *m/z* of 503.14, 517.16, and 485.19 respectively are shown. The EICs are magnified 2x for clarity. **b.** Predicted structures of **69-71** labelled with their *m/z*. Fragments of **69** seen in MS^2 are shown on the molecule. Fragments marked with an asterisk show a loss of 33.99, characteristic of the loss of SH$_2$ from thioamide bonds in fragmentation. **c.** MS^2 fragmentation of **69**. Identified fragments are in red.
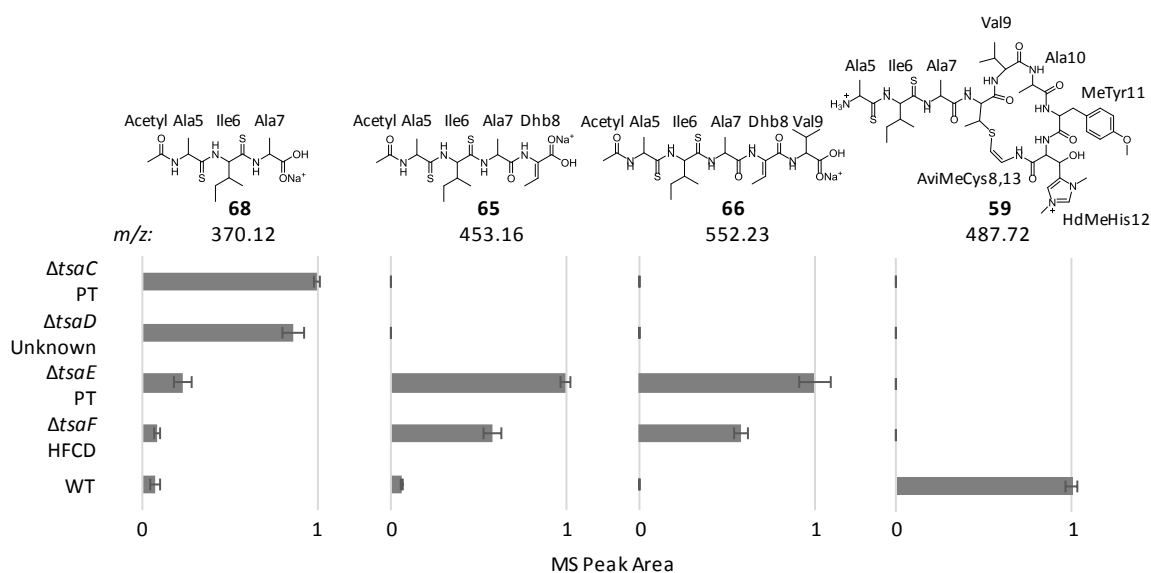
Figure 73. Mass spectral areas of **68** [M+Na$^+$], **65** [M+Na$^+$], **66** [M+Na$^+$], and **59** [M+2H$^+$] in Δ*tsaC-F* and wild type gene clusters. Each bar chart is normalised to the highest mass spectral area. The error bars represent the standard error. PT – phosphotransferase, HFCD – homo-oligomeric flavin-containing cysteine decarboxylase

and **68** give key information towards the roles of TsaC-F. **65** and **66** are breakdown products from an intermediate that lacks the macrocycle yet contains the Dhb8 residue that is required for macrocycle formation. It can be suggested that **68** is a breakdown product from an intermediate that contains an unmodified Thr8 (normally dehydrated to form Dhb8), as modified peptides are more resistant to proteolysis. It can therefore be construed that if **66** and **65** are not seen, the threonine is either not being dehydrated (as seen with **68**), or it is being dehydrated and then further modified, for example by formation of the AviMeCys macrocycle (as seen with **59**). In all strains containing deletions of any of *tsaC-F* no molecules containing the macrocycle are seen (Figure 73). In both the Δ*tsaC* and Δ*tsaD* clusters **68** is seen, whilst **65** and **66** are not seen. Therefore, it can be proposed that TsaC, TsaD, or a combination of the two catalyse dehydration of the Thr8 to Dhb8.

TsaC has low identity with a *Cylindrospermum* sp. NIES-4074 aminoglycoside phosphotransferase (APH), and BLASTP also identifies an APH ChoK-like conserved domain. Phyre2 also aligns 82% of the secondary structure to an APH structure with 99.9% confidence. APHs are resistance enzymes responsible for inactivating aminoglycoside antibiotics by phosphorylation of a sugar hydroxyl group, which interrupts their nucleic acid binding ability, and therefore their mechanism of action (Wright and Thompson, 1999). APHs show very high structural similarity to eukaryotic protein kinases (Hon et al., 1997) and it has been shown that some APH enzymes can also phosphorylate serine residues (Daigle et al., 1999). This suggests that although the BlastP and Phyre2 results indicate TsaC is an aminoglycoside phosphotransferase, it is also possible it could be a serine/threonine kinase. Given that in its absence Thr8 is not dehydrated, it can be predicted that TsaC is responsible for phosphorylation of Thr8, allowing for a subsequent elimination reaction to dehydrate Thr8 (Figure 74).

The role of TsaD in the threonine dehydration is unclear, however the bioinformatic data provides some interesting clues. BLASTP showed that TsaD has many homologues spread across many bacterial phyla, however all were annotated as hypothetical proteins with no identified conserved domains. In contrast, Phyre2 matched the predicted secondary structure of TsaD to a type III effector HopA1 (Hrp-dependent outer protein A1) from *Pseudomonas syringae* (Park et al., 2015). The first
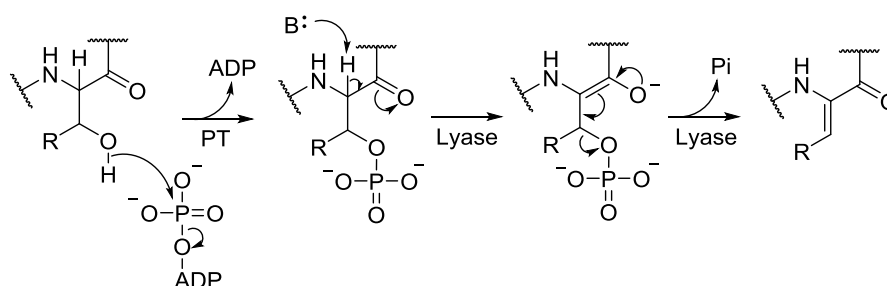
Figure 74. Schematic of phosphorylation and elimination-mediated dehydration of a serine (R = H) or a threonine (R = CH$_3$). PT – phosphotransferase.
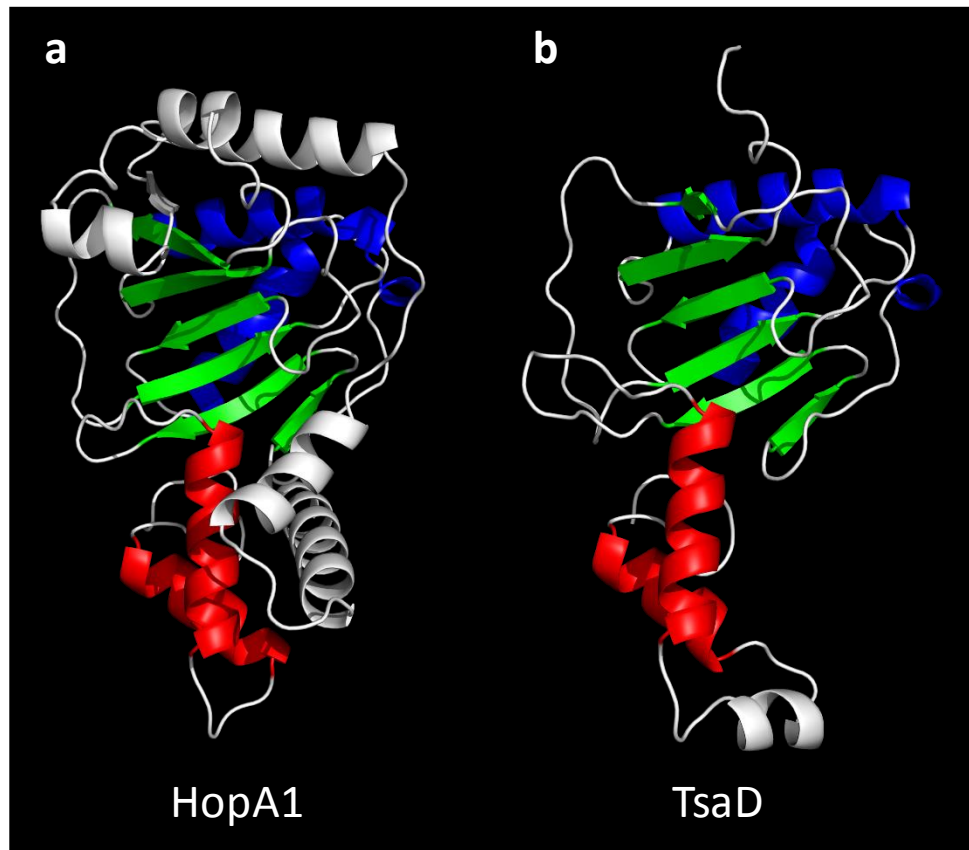
Figure 75 a. Structure of *Pseudomonas syringae* HopA1, amino acids 123-379. **b.** Phyre2 prediction of the structure of TsaD, amino acids 152-331, modelled on HopA1. Regions in which the secondary structure aligns are coloured for clarity.

102 amino acids of HopA1 contain a secretion signal and a ShcA1 chaperone binding domain; a chaperone that targets HopA1 to the type III secretion system (Janjusevic et al., 2013); TsaD lacks homology with these residues but has predicted homology to residues 209-378. Once secreted as an effector protein during plant infection, HopA1 has been shown to interrupt protein complexes formed by the *Arabidopsis* enhanced disease susceptibility1 (EDS1) protein; protein complexes that normally lead to increased expression of defence genes (Bhattacharjee et al., 2011; Heidrich et al., 2011; Kim et al., 2009). However, it is not known if HopA1 has enzymatic activity. Figure 75 shows the predicted secondary structure of TsaD modelled on the tertiary structure of HopA1. These proteins are characterised as an α + β fold with the central β-sheet containing a concave groove (Park et al., 2015).

Despite their shared host and almost identical (and often confused) names, the type III *Pseudomonas* effectors HopA1 (aforementioned) and HopAI1 (an OspF and SpvC like-lyase; Zhu et al., 2007) share no predicted primary or secondary structure similarity based on pairwise alignment and Phyre2 results. However, it has been shown that HopA1 and HopAI1 show similarity in their tertiary structure (Park et al., 2015). HopAI1-like lyases contain a central β-sheet forming a concave groove similar to that in HopA1 (Chen et al., 2008). HopAI1 dephosphorylates plant MAPK proteins involved in

pathogen response (Zhang et al., 2007). Additionally, class III and IV lanthionine bond forming enzymes contain a phosphate elimination lyase domain that is homologous to HopAI1-like lyases (Goto et al., 2010, 2011). The HopAI1 concave groove contains identified active site residues, and whilst similar residues are present in HopA1's groove, they are not in the same positions. The similar tertiary structure between a lyase, HopAI1, and the predicted TsaD homologue, HopA1, combined with the deletion experiment showing that TsaD is required for threonine dehydration allows a function for TsaD to be suggested. This is consistent with a hypothesis that TsaD may act as a lyase to catalyse the elimination of a TsaC-installed phosphate group to dehydrate Thr8. This dehydration of Thr8 to Dhb8 is a key step that needs to occur prior to AviMeCys macrocycle formation. It is rare in RiPPs to see dehydrations that are not catalysed by homologues of the enzymes involved in lanthionine biosynthesis.

In both the ΔtsaE and ΔtsaF clusters, **66**, **65**, and **68** are produced (Figure 73). Therefore, neither enzyme is required for the dehydration of Thr8 to Dhb8. However, the macrocycle is not formed. TsaF shows strong similarity with HFCD family proteins. HFCDs have been shown to decarboxylate cysteines in the formation of AviCys (Blaesse et al., 2000), AviMeCys (Blaesse et al., 2003), and avionin macrocycles (Wiebach et al., 2018). In Avi(Me)Cys containing natural products this decarboxylation forms the thioenolate moiety required for Avi(Me)Cys formation (Figure 76; Blaesse et al., 2003). The lack of macrocycle when *tsaF* is deleted and the bioinformatic results support the role of TsaF as a cysteine decarboxylase that decarboxylates Cys13.

Based on the Phyre2 results, TsaE shows strong predicted structural similarity with macrolide 2'-phosphotransferase (MPH)-like proteins. However, this only covers the second half of the protein. MPH enzymes have very similar structures and functions to APH enzymes (Fong et al., 2017; Pawlowski et al., 2018), but they inactivate macrolide antibiotics such as erythromycin rather than aminoglycosides (O'Hara et al., 1989). According to Phyre2, half of the nucleotide binding site of MPH lies within the region of the protein in which no structural similarity is picked up between TsaE and MPH. The key residues in the other half of the nucleotide binding site and the macrolide binding site are also not well conserved (Fong et al., 2017). If TsaE is a MPH that has evolved to phosphorylate
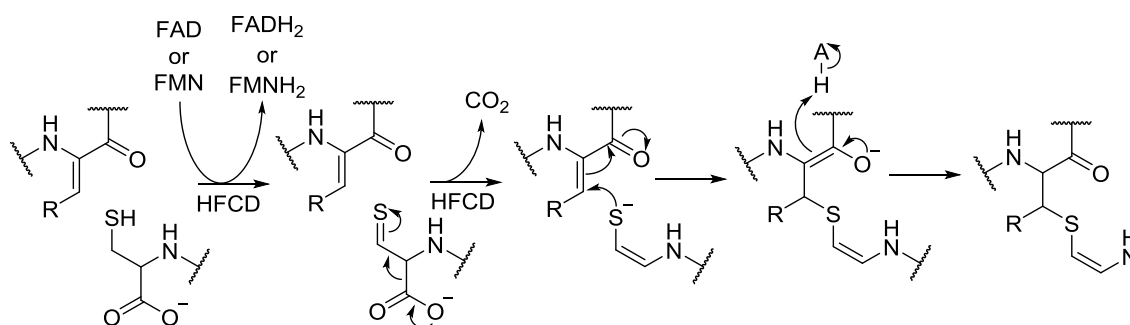
Figure 76. Schematic of HFCD-catalysed cysteine decarboxylation and AviCys (R = H) and AviMeCys (R = CH_3) formation.

the thiostreptamide S4 (**17**) core peptide then it is not a surprise that the macrolide binding site would be different. Given the inconclusive results from Phyre2, TsaE was also submitted to I-TASSER for analysis (Yang et al., 2015). I-TASSER is a more intensive and thorough, yet slower protein structure predictor than Phyre2. Encouragingly, I-TASSER aligned 89% of TsaE with Rv3168, an APH from *Mycobacterium tuberculosis* (S. Kim et al., 2011), with a Z score of 1.17 (a Z score greater than 1 means a good alignment). I-TASSER was also able to find an ATP-binding site analogous to Rv3168's (Kim and Kim, 2011) with a C-score of 0.49 (scores range from 0-1, with higher scores indicating a more reliable prediction) in the correct place in the predicted protein structure. Taken alone, a score of 0.49 is not conclusive, but when taken together with all the predicted structural similarity evidence this strongly suggests TsaE is an APH-like phosphotransferase.

The ΔtsaE cluster has a very similar metabolite profile to the ΔtsaF cluster (Table 5; Figure 73). In the ΔtsaE cluster the macrocycle should be able to form, because Dhb8 can be detected and the HFCD family protein, TsaF, is present. The lack of macrocycle could be explained if TsaE acts as a novel AviMeCys cyclase. However, it is unclear what role a phosphotransferase could play in cyclisation. A different possibility is that TsaE does not play a catalytic role, and instead just brings the thioenolate close to Dhb8 to aid the spontaneous formation of the AviMeCys reside. This seems unlikely as there is no *tsaE* homologue in other Avi(Me)Cys containing clusters such as the linaridins (Claesen and Bibb, 2010). Another alternative hypothesis is that TsaE is a phosphotransferase involved in Ser1 dehydration to Dha1; a dehydration predicted to be necessary for the spontaneous formation of the N-terminal pyruvyl moiety (Figure 77).

It is likely that one of the phosphotransferases, TsaC or TsaE, is responsible for the phosphorylation when dehydrating the Ser1 to Dha1 (with subsequent elimination possibly carried out by TsaD; Figure 74). It has been shown that TsaC is likely responsible for the phosphorylation of Thr8 during dehydration (again with elimination possibly being carried out by TsaD), and so it is catalytically competent enough to be involved in the dehydration of Ser1. Multiple dehydrations catalysed by a single enzyme is commonly seen in conventional lanthipeptide biosynthesis (Repka et al., 2017). However, TsaE mediated phosphorylation during the dehydration of Ser1 would also make sense. **69** gives some interesting clues in this direction (Figure 78). Molecule **69** is predicted to be an acetylated tetrapeptide, SerValMetAla, with thioamide bonds on either side of the methionine, based on MS[2] data. This represents a breakdown product of a pathway intermediate that contains Ser1 to Ala4.
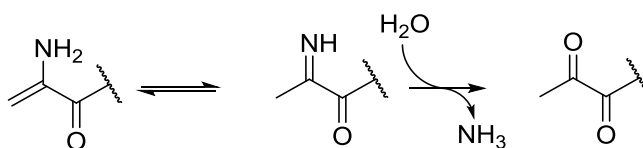
Figure 77. Proposed mechanism for spontaneous formation of a pyruvyl from an N-terminal Dha in water.
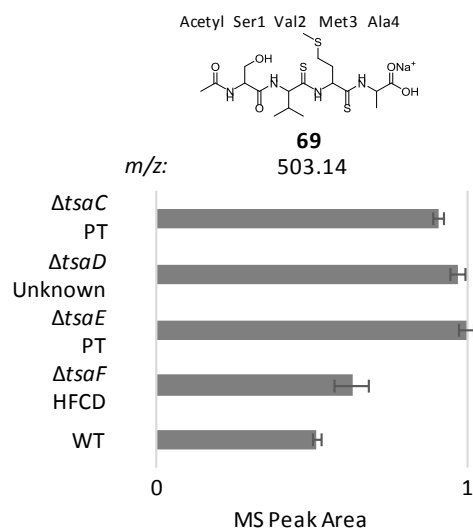
Figure 78. MS peak areas of **69** in Δ*tsaC-F* and wildtype clusters. The bar chart is normalised to the highest mass peak area. The error bars represent the standard error. PT = phosphotransferase, HFCD = homo-oligomeric flavin-containing cysteine decarboxylase.

This is an intermediate from after thioamide bonds have been installed, but before Ser1 dehydration and pyruvyl formation. Unfortunately, a similar molecule with a dehydrated serine or a pyruvyl group cannot be detected. The deletion of either *tsaC, D*, or *E* increases the production of **69** nearly two-fold compared to wild type (Figure 78). It can be suggested that when Ser1 can be dehydrated less of molecule **69** is seen, as in the wild type cluster. Whilst the Δ*tsaF* cluster shows very little increase in the production of molecule **69**, the Δ*tsaC, D*, or *E* clusters show a very significant increase. Therefore, Ser1 is dehydrated less efficiently when any of *tsaC, D*, or *E* are deleted. This result is consistent with the hypothesis that TsaE (and possibly TsaD) is involved in the dehydration of Ser1 following Thr8 dehydration, and that this Ser1 dehydration is required prior to AviMeCys macrocyclisation.

To suggest TsaE catalyses the dehydration of Ser1 to form the pyruvyl, it would first have to be shown that the pyruvyl originates from a dehydrated serine instead of a pyruvyl transferase, as seen in xanthan biosynthesis (Katzen et al., 1998). To confirm this, the precursor peptide was mutated to convert Ser1 to a threonine. As the Δ*tsaA* cluster could not be complemented, this precursor peptide modification was made to the wild type cluster using the same yeast-based method that was used to refactor the bottromycin (**1**) cluster (see section 3.2.2.2 for a full description of precursor peptide modifications; Eyles et al., 2018). The cluster with Ser1 mutated to threonine (pCAPtsaS1T) produced a compound of mass 1391.56, 14 Da higher than thiostreptamide S4 (**17**), and fragmentation data determined that the additional mass was in the N-terminal moiety, the valine, or the methionine (Figure 79). This is consistent with the predicted processing of the first amino acid, and therefore confirms that the N-terminal pyruvyl group comes from a processed serine. This S1T modification to the core peptide is seen naturally in the *Micromonospora eburnea* DSM 44814 and *Salinispora*
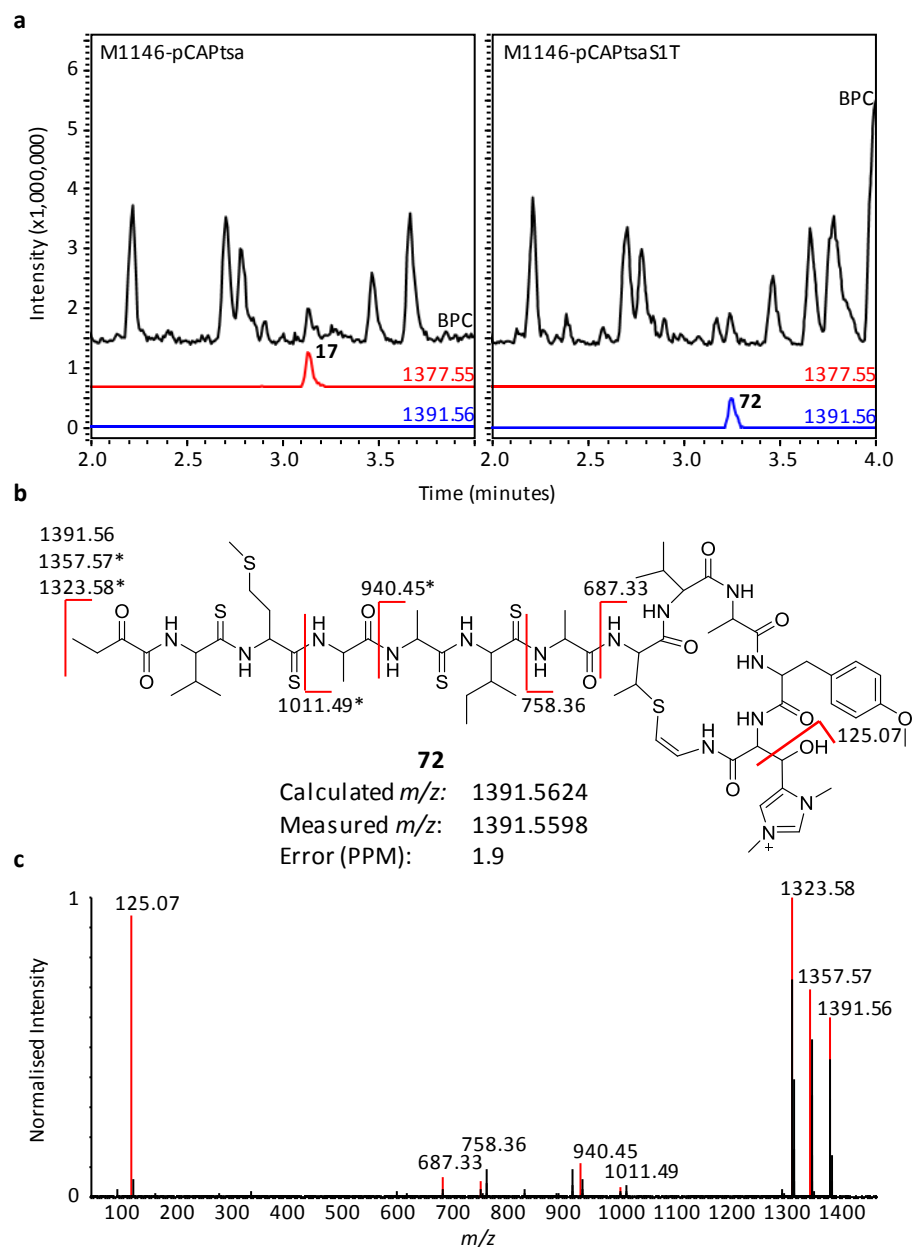
Figure 79. a. LC-MS analysis of M1146-pCAPtsa and M1146-pCAPtsaS1T. EICs of 1377.55 and 1391.56 are shown. EICs of 1377.55 and 1391.56 from M1146-pCAPtsa and M1146-pCAPtsaS1T. **b.** Predicted structure of **72**. Identified fragments are shown on the molecule. Fragments marked with an asterisk show a loss of 33.99, characteristic of the loss of SH_2 from thioamide bonds in fragmentation. **c.** MS^2 Fragmentation of **72**.

pacifica CNT029 predicted core peptides, so it is likely that the N-terminal 2-oxobutyryl produced by pCAPtsaS1T is also produced by these.

This type of N-terminal modification has been known about for a long time. The RiPP Pep5 contains an N-terminal 2-oxobutyryl group and is predicted to derive from a dehydrated threonine (Kellner et al., 1989). Lactocin S contains an N-terminal pyruvyl group that is suggested to originate from a dehydrated serine (Skaugen et al., 1994), and the lactyl group in epilancin 15X has been shown to originate from the reduction of an N-terminal pyruvyl group (Velásquez et al., 2011). Whilst it has

been suggested many times that these moieties originate from dehydrated serines and threonines, converting the pyruvyl into a 2-oxobutyryl with the construction of pCAPtsaS1T is the first experimental evidence of this. This also supports the hypothesis that TsaE is involved in Ser1 dehydration.

In summary, the pathway that most strongly correlates with *tsaC-F* data is as follows:

> 1) Thioamide bond formation precedes their activity, as all identified metabolites contain thioamide bonds (Figure 71; Figure 72).

> 2) Thr8 is dehydrated to Dhb8 by TsaC-mediated phosphorylation and TsaD-mediated elimination. The evidence for this is in the lack of Dhb8 containing metabolites when *tsaC* and *tsaD* are deleted (Figure 73).

> 3) Ser1 is dehydrated to Dha1 by TsaE-mediated phosphorylation and possibly TsaD-mediated elimination. The evidence for this is in the increase in unprocessed serine when *tsaE* and *tsaD* are deleted (Figure 78). The evidence that this follows Thr8 dehydration is that there is also an increase in unprocessed Ser1 when *tsaC* is deleted (Figure 78).

> 4) Cysteine decarboxylation is catalysed by TsaF, forming a reactive thioenolate that can form a cycle with Dhb8. The evidence for this is the lack of macrocycle seen when *tsaF* is deleted and the bioinformatic analysis of TsaF that shows it is homologous to HFCDs involved in Avi(Me)Cys formation.

There are a few issues with this proposed pathway. The largest problem lies in explaining why the macrocycle formation requires Ser1 dehydration. It would be unusual for a HFCD such as TsaF to have substrate requirements so far away from its target, Cys13 (Kupke et al., 1995). Perhaps Cys13 is decarboxylated irrespective of the state of Ser1 dehydration, and it is just that the macrocycle is not formed until after Ser1 dehydration. This hypothesis would, however, require a mechanism to prevent the thioenolate from cyclising onto Dhb8 until after Ser1 dehydration; as the cyclisation is currently suggested to be a non-enzymatic reaction (Figure 76). One or more of TsaC-F may exert conformational control over the core peptide, preventing cyclisation until the Ser1 has been dehydrated. The currently unidentified metabolites seen with these deletions may give more clues (Table 5). It is sensible to predict that a protein exerts some sort of steric control over the core peptide during macrocyclisation as it would likely be necessary to ensure the correct stereochemistry of the resulting AviMeCys residue.

3.2.1.6.   Thioamide Bond Installation

When this project was started the biosynthetic origin of the thioamide bonds was unknown, and it was predicted that YcaO and the TfuA genes could be responsible. Therefore, TsaH and TsaI, YcaO

and TfuA proteins, were predicted to form the thioamide bonds. Upon deletion of each gene the production of every detectable metabolite was abolished. Thioamide bonds can protect peptides from degradation (Chen et al., 2017), and as such it is not a surprise that no intermediate metabolites are detected in the absence of thioamidation. In every other deletion of tailoring enzymes thioamidated compounds could be seen. By process of elimination this shows that TsaH and TsaI are responsible for the thioamidations. This conclusion that the YcaO and TfuA proteins were responsible was supported by recently published work in archaea (Mahanta et al., 2018; Nayak et al., 2017) and in bacterial thiopeptides (Schwalen et al., 2018). It is important to note that macrocycle-containing compounds are not seen, even though the macrocycle also seems to be resistant to degradation. This suggests that the thioamide bond formation precedes and is required for macrocycle formation.

To investigate the tolerance of the thioamide bond formation, a modified version of the cluster was constructed in which the core peptide contained an alanine between the Ser1 and Val2. This was constructed following the same method that was used to produce pCAPtsaS1T, and it was named pCAPtsaSAV. This extended the distance the between the leader peptide and the thioamidated regions of the core peptide by a single amino acid. Interestingly, production of all detectable thiostreptamide S4-related metabolites was abolished. This shows that TsaH and TsaI are unable to thioamidate the core peptide when the amino acids are one residue further from the leader peptide. This suggests that, whilst TsaH and TsaI thioamidate multiple regions on the core peptide, this does not translate to a tolerance of changes to the length of their substrate

3.2.1.7.   Analysis of the Additional Genes in the *A. alba* Gene Cluster

The *A. alba* thioalbamide (**18**) cluster contains genes that code for a predicted P450 (TaaCYP) and reductase (TaaRed) that are not within the thiostreptamide S4 (**17**) gene cluster (Figure 80). The function of these genes could not be tested directly in *A. alba* because previous attempts to genetically manipulate this strain were unsuccessful (Luca Frattaruolo, personal communication). The *A. alba* cluster was previously TAR cloned and conjugation into *Amycolatopsis orientalis* was attempted (Luca Frattaruolo, unpublished). No thioalbamide (**18**) could be detected in the ex-conjugants. Therefore, an alternative method was needed to investigate *taaCYP* and *taaRed*. It was decided that *taaCYP* and *taaRed* would be expressed in M1146 alongside the thiostreptamide S4 (**17**) gene cluster to test if their activity in *A. alba* could be reconstituted on a similar molecule.

Thioalbamide (**18**) has a hydroxylated phenylalanine (Frattaruolo et al., 2017) not seen in other characterised thioviridamide-like compounds, so it was hypothesised that the P450 (TaaCYP) is responsible for this hydroxylation. To test this, the plan was to express TaaCYP alongside two new versions of the thiostreptamide S4 (**17**) cluster with core peptides containing a phenylalanine at amino acid position 5, the position it is found in the thioalbamide (**18**) core peptide (see section 3.2.2.2 for a full description of precursor peptide modifications). The new constructs were
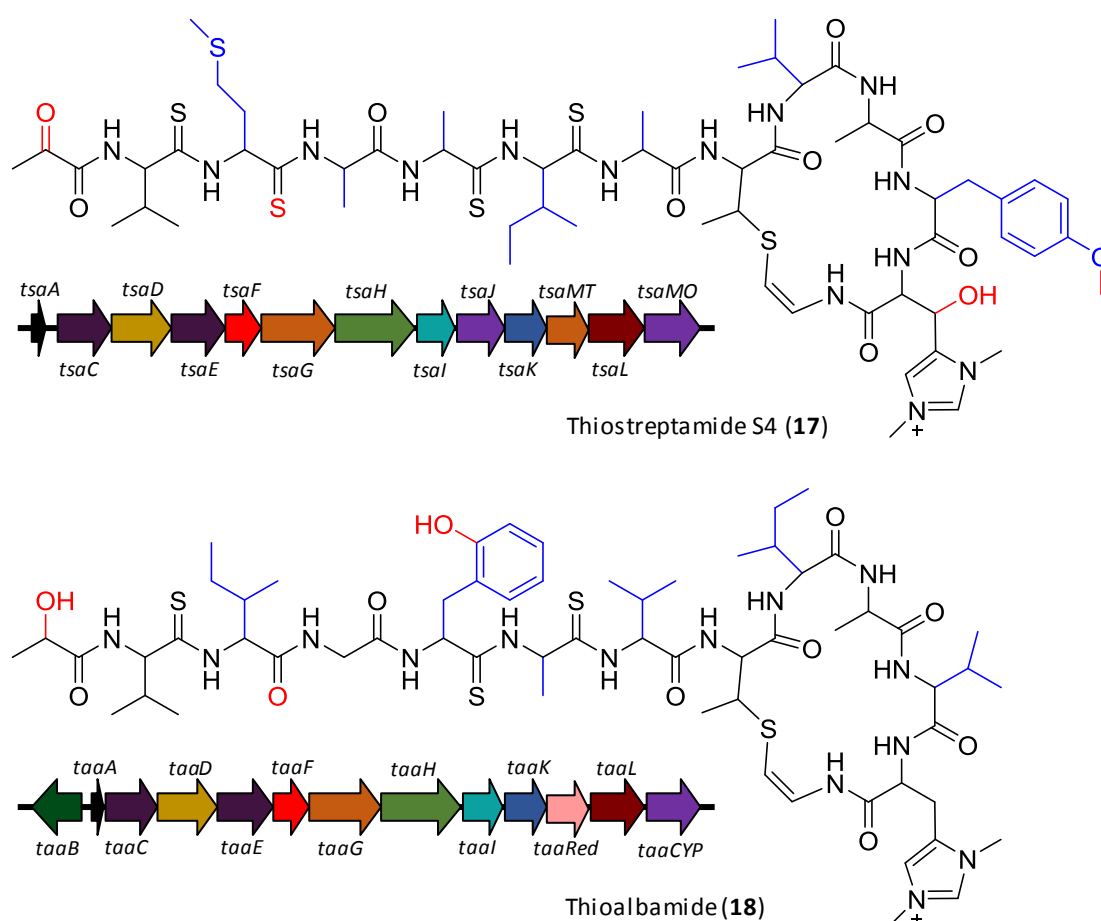
Figure 80. The structures and gene clusters of thiostreptamide S4 (**17**) and thioalbamide (**18**). Differences due to amino acid changes in the core peptide are highlighted in blue, differences due to tailoring enzymes are highlighted in red.

pCAPtsaA5F, with a phenylalanine replacing Ala5, and pCAPtsaTsaCoreTaa, with the entire thioalbamide (**18**) core peptide replacing the thiostreptamide S4 (**17**) core peptide. Unfortunately, when these clusters were expressed in *S. coelicolor* M1146 no related metabolites could be seen, meaning that these modifications were not tolerated by the thiostreptamide S4 (**17**) tailoring enzymes. This meant the P450 could not be tested alongside these modified clusters to see if it could hydroxylate the phenylalanine. The P450 was still expressed in *S. coelicolor* M1146 using PermE* in pIJ10257 alongside the thiostreptamide S4 (**17**) wild type cluster to see if any residual activity could be detected in the absence of the appropriate phenylalanine substrate; unsurprisingly no P450-modified thiostreptamide S4-related metabolites could be detected.

Thioalbamide (**18**) contains a lactyl moiety at the N-terminus. In epilancin 15X a pyruvyl moiety is reduced to a lactyl group by ElxO, a SDR family NADP dependant oxidoreductase (Ortega et al., 2014). The thioalbamide (**18**) biosynthetic gene cluster contains *taaRed*, predicted to code for a SDR family reductase not seen in the other thioviridamide-like clusters that could act in an ElxO-like manner to convert the N-terminal pyruvyl to a lactyl. To confirm the function of TaaRed, its gene was cloned into pIJ10257 and integrated into *S. coelicolor* M1146 carrying the wild type thiostreptamide S4 (**17**)

cluster. This strain produced a compound of mass 1379.56 which correlates with the gain of 2 hydrogen atoms. MS^2 analysis showed that this mass change is present in the N-terminal moiety, Val2, or Met3 of the molecule, and is consistent with reduction of the pyruvyl to a lactyl moiety (Figure 81). This confirmed that TaaRed is the reductase responsible for lactyl formation. Further work to determine the implications on bioactivity of this modification is important, but unfortunately time constraints prevented this work from taking place as part of this thesis. Additionally, it would be interesting to see if the 2-oxobutyryl formed by pCAPtsaS1T could also be reduced by TaaRed. JBIR-140 (**16**) contains the lactyl moiety and is produced by heterologous expression of the
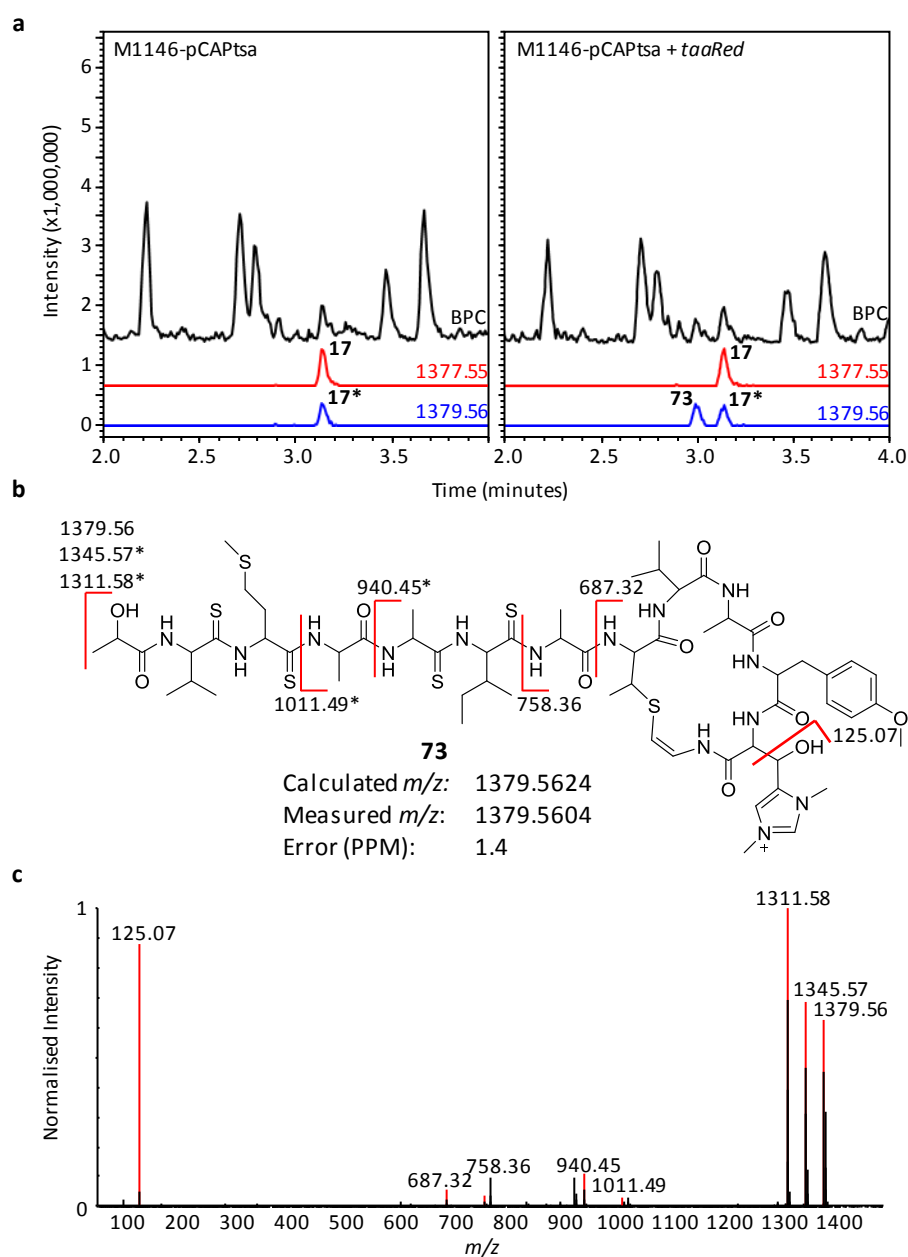
Figure 81. a. LC-MS analysis of M1146-pCAPtsa and M1146-pCAPtsa + *taaRed*. EICs of 1377.55 and 1379.56 are shown. **17*** labels the second isotope peak of **17**. **b.** Predicted structure of **73**. Identified fragments are shown on the molecule. Fragments marked with an asterisk show a loss of 33.99, characteristic of the loss of SH$_2$ from thioamide bonds in fragmentation. **c.** MS^2 Fragmentation of **73**.

thioviridamide (**15**) cluster from *S. olivoviridis* OM13 in *S. avermitilis* SUKA17. Whilst the thioviridamide (**15**) cluster does not contain a reductase, the non-native background may be providing a promiscuous reductase that can catalyse the reduction of the pyruvyl moiety.

3.2.1.8.   Start Codon Assessment

The complementation experiments allowed an assessment of the start codons of each gene and gave information towards the regulatory requirements of genes. The deletions of *tsaA*, *tsaC*, *tsaD* and *tsaG* were not successfully complemented using their annotated start codons, and so the further attempts to complement them are described here.

TsaA was previously identified as the precursor peptide due to the C-terminal SVMAAIATVAYHC motif (Frattaruolo et al., 2017). Unsurprisingly, when it was deleted production of thiostreptamide S4 (**17**) is abolished. Complementation of this deletion was attempted with two versions of the precursor peptide *in trans* under the control of PermE* (Bibb et al., 1994), starting from two potential start codons. Neither complementation worked. This is not unusual; PermE* is often not sufficient for complementation of precursor peptide deletions. For example, the deletion of the GE37468 precursor peptide could not be complemented using PermE* as the promoter, however successful complementation of the GE37468 precursor peptide deletion was achieved if the native promoter was used (Young et al., 2012). Therefore, complementation of the *tsaA* deletion was attempted using *tsaA*'s native promoter. Even with the native *tsaA* promoter driving expression of the precursor peptide, complementation was unsuccessful. It is unclear why this may be, although it is possible that the presence of *tsaA* at the beginning of the operon may be important for proper regulatory control of downstream genes.

The *tsaC* start codon appeared to be annotated correctly, with a suitable RBS. However, the *tsaC* deletion could not be complemented when the complementation construct contained this start codon in front of PermE*. There is a 314 bp region between *tsaA* and *tsaC*, and when this region was included between PermE* and the start codon the complementation worked. This suggests that tsaC has requirements for its efficient expression outside of just a strong promoter, and these are provided by the *tsaA* to *tsaC* intergenic region.

The annotated start codon for *tsaD* did not produce a working complementation construct, so two other potential start codons were also chosen for complementing *tsaD* (Figure 82). Complementation could only be achieved using the start codon that overlapped with the stop codon from *tsaC*. The *tsaG* deletion could not be complemented with the annotated start codon, so complementation was attempted with three different possible start codons (Figure 82). The complementation was successful using the start codon that overlaps with the stop codon of *tsaF* and

a

```
tsaC     ArgAla***
tsaD           MetSerArgTyrAspAlaValPheAlaSerMetAlaGluAspIleGluValLeuAsp
     1   CGCGCATGAGCAGATACGACGCAGTGTTCGCCAGCATGGCGGAGGACATAGAGGTCCTGGAC

tsaD     GlnAlaThrPheArgHisArgGluTrpGlyAspLeuSerProAlaAlaGlyValGluThr
    63   CAGGCAACTTTCCGTCACCGGGAGTGGGGTGACCTGTCCCCTGCTGCTGGGGTGGAGACA
```

b

```
tsaF     GlyGln***
tsaG           ValIleProAspMetThrLysLeuSerProGlnThrValLeuTyrArgAlaProSer
     1   GGACAGTGATCCCAGACATGACCAAGCTCTCACCACAGACGGTTCTCTACCGGGCGCCGTCC

tsaG     PheValAlaGluValAspThrSerAsnGluValLysIleHisPheGluGlyArgValLeuLys
    63   TTTGTCGCCGAGGTCGACACCAGTAACGAGGTGAAGATCCACTTCGAAGGCCGGGTGCTCAAG
```

Figure 82. **a.** Region of thiostreptamide S4 (**17**) gene cluster containing the three potential start codons for *tsaD*. The three potential start codons tested are highlighted in red. The first start codon (bold), coupled to the *tsaC* stop codon, is the only one with which the *tsaD* deletion could be complemented. The third start codon is the annotated start codon. **b.** Region of thiostreptamide S4 (**17**) gene cluster containing the four potential start codons for *tsaG*. The four potential start codons tested are highlighted in red. The first start codon (bold), coupled to the *tsaF* stop codon, and the second start codon (bold) are the start codons with which the *tsaG* deletion could be complemented. The last start codon is the annotated start codon.

with the start codon four codons downstream. The lack of an appropriate RBS for the second start codon suggests that the one linked to the *tsaF* stop codon is the true start codon.

The confirmation of start codons showed that most of the genes in this cluster have start codons that overlap with the preceding stop codon. Two series of genes with overlapping start and stop codons, *tsaC-G* and *tsaH-MT*, are present and the space between *tsaG* and *tsaH* is 28 bp. Overlapping start and stop codons are a good indication of translational coupling (Das and Yanofsky, 1989; Oppenheim and Yanofsky, 1980). Therefore, it is likely many of the genes in this operon are translationally coupled. This would be an important mechanism used to maintain the correct ratio of proteins produced. Translational coupling does not always lead to a 1:1 ratio of proteins produced; the efficiency of coupling can be affected by many things including the sequence of the second gene's ribosome binding site (Das and Yanofsky, 1984) and the translation level of the first gene (Yu et al., 2001). Therefore, the observation of the translational coupling in the thiostreptamide S4 (**17**) gene cluster does not allow a prediction to be made of the relative protein ratios that are produced.

Whilst intragenic promoters are possible, there is not enough space for a terminator between most of the genes. This means it is likely the genes *tsaC-MT* are on a single operon. The gap between *tsaA* and *tsaC* is large enough for a terminator and promoter, therefore there may be two promoters in the gene cluster, one driving expression of *tsaA* and one driving the expression of the *tsaC-MT* operon, but further experiments would be required to determine the true operon structure.

3.2.2. Thiostreptamide S4 Engineering

3.2.2.1. Refactoring

Chapter 2 of this thesis describes the development of a method to refactor gene clusters. This method was therefore used with the thiostreptamide S4 (**17**) gene cluster due to the poor productivity of this pathway. This compound is predicted to have interesting bioactivity, with many thioviridamide-like molecules having selective anti-cancer activities (Frattaruolo et al., 2017; Hayakawa et al., 2006a; Kjaerulff et al., 2017), but the purification necessary to test this with thiostreptamide S4 (**17**) was hampered due to the low production from both *S. coelicolor* M1146-pCAPtsa and *Streptomyces* sp. NRRL S-4. The complex production medium (btmPM) results in significant production of other metabolites; thiostreptamide S4 (**17**) is one of the minor compounds produced (Figure 83). Production from *S. coelicolor* M1146-pCAPtsa is also better on solid media, which makes scaling up growths much less efficient than if production was good in liquid media. Refactoring was conducted with the aim of achieving one or more of the following goals: a general increase in production, an increase in production in liquid medium, and/or an increase in production in simpler medium.

The region spanning from before *tsaA* to before *tsaC* was predicted to be the regulatory region, after analysis of the start codons from the cluster showed that the only intergenic gaps large enough for regulatory elements were before *tsaA* and *tsaC*. Therefore, this area was targeted for modification. When deciding how to refactor this region there were a few things to consider. Attenuating terminators (rather than complete terminators) often follow precursor peptide genes at the start of operons. This is to ensure an excess of precursor peptides compared to the tailoring enzymes, for example in microcin C (Zukher et al., 2014). This may not be the case with the thiostreptamide S4 (**17**) cluster, as there is a large (314 bp) gap between *tsaA* and *tsaC*, which allows room for both a terminator and a promotor. As it was unclear how the regulation of this region worked, it was
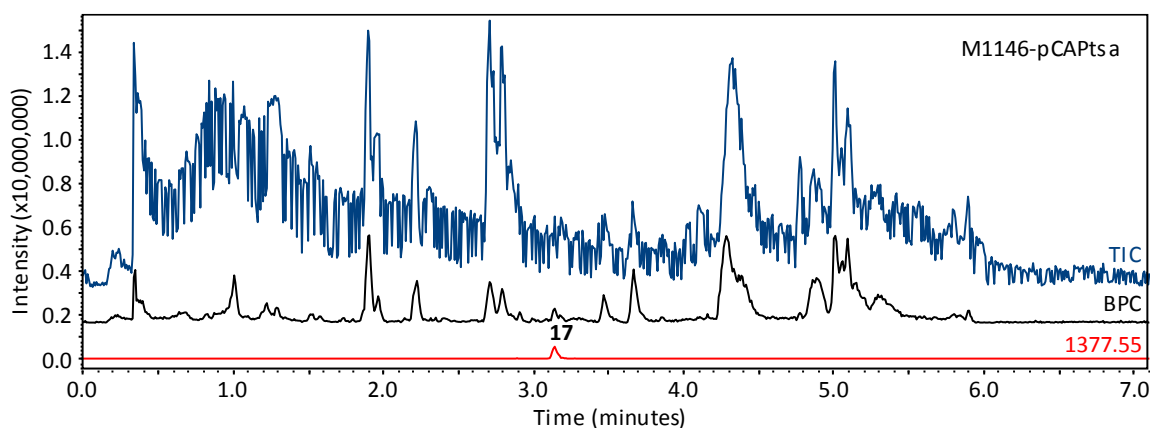
Figure 83. LC-MS analysis of *S. coelicolor* M1146-pCAPtsa. TIC, BPC, and EIC of 1377.55 (**17**) are shown.

decided that two promoters would be used, one to drive expression of *tsaA*, and one to drive expression of the *tsaC-MT* operon. PSF14 and Paac3 were chosen due to their robust activity, explored in the previous chapter (section 2.2.1.2). Four versions of the cluster were made: pCAPtsa1 with PSF14 driving expression of *tsaA* and another PSF14 driving expression of the *tsaC-MT* operon, pCAPtsa2 with PSF14 driving expression of *tsaA* and Paac3 driving expression of the *tsaC-MT* operon, pCAPtsa3 with Paac3 driving expression of *tsaA* and PSF14 driving expression of the *tsaC-MT* operon, and pCAPtsa4 with Paac3 driving expression of *tsaA* and Paac3 driving expression of the *tsaC-MT* operon (Figure 84). This would test the effect of two different strength promoters in all combinations in the two positions available.

As with the refactoring of the bottromycin (**1**) gene cluster, naturally occurring unique restriction enzyme recognition sites were chosen to base the modifications around. These were AflII, situated before *tsaA*, and SrfI, situated after *tsaC* (Figure 84). Using these recognition sites, *tsaA* and *tsaC* could be digested out of pCAPtsa and this region could be rebuilt with a series of PCR products. Each cluster, pCAPtsa1-4, was built by assembling the AflII and SrfI digested plasmid backbone with four
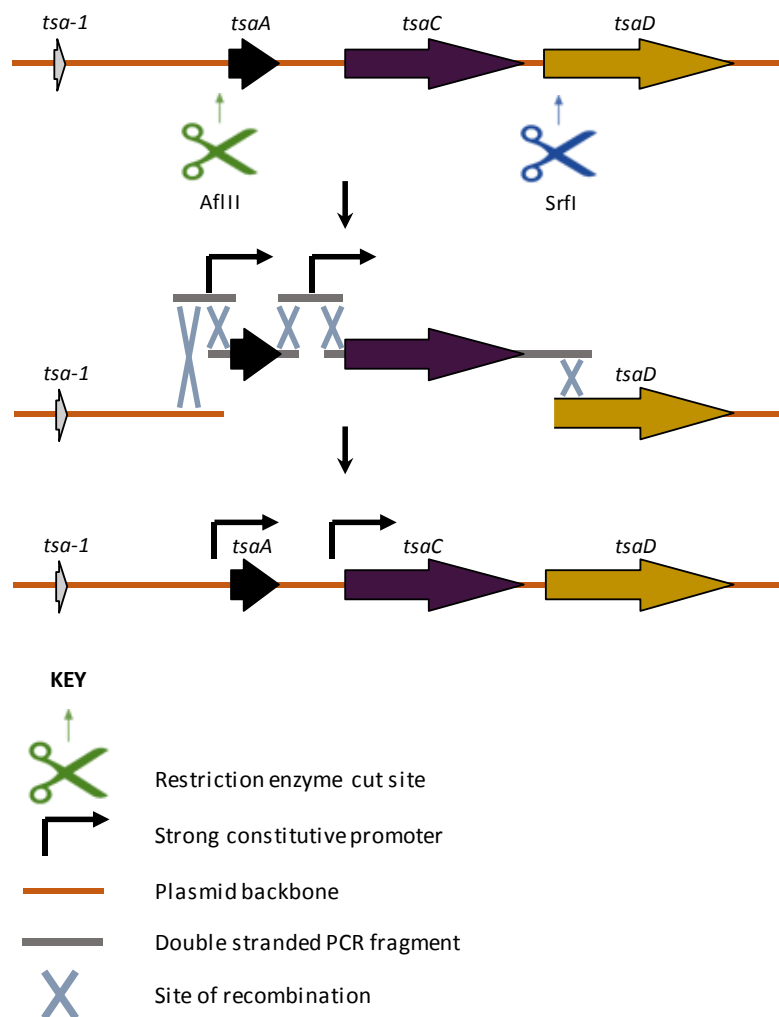
Figure 84. Refactoring pCAPtsa by assembly in yeast. The strong constitutive promoters are PSF14 or Paac3, organised in all four possible combinations to produce pCAPtsa1-4.

PCR fragments in yeast (Figure 84). Each PCR fragment either contained *tsaA*, *tsaB*, PSF14, or Paac3, and had the appropriate linker to allow it to assemble with the digested plasmid backbone or neighbouring PCR fragments.

These clusters were conjugated into *S. coelicolor* M1146, but no thiostreptamide S4-related metabolites could be detected by LC-MS following fermentation in btmPM, a media shown to support the activity from the promoters PSF14 and Paac3. There are many reasons this could have failed. As shown by the complementation experiments, *tsaA* under the control of PermE* *in trans* could not rescue production of thiostreptamide S4 (**17**). It is possible Paac3 and PSF14 are unable to drive sufficient expression of *tsaA* to support thiostreptamide S4 (**17**) production, or it is possible that the position of *tsaA* at the start of the operon may be important for pathway productivity. Whilst *tsaA* is still in the same position after refactoring, the promoter placed between *tsaA* and *tsaC* may interrupt the role *tsaA* plays in the expression of downstream genes. A new refactoring experiment was therefore designed, to see if an alternative construct could avoid these potential issues.

A riboswitch was successfully used in controlling the expression of the bottromycin (**1**) gene cluster (see section 2.2.3.2). When placed between PSF14 and an operon, this riboswitch controls the expression of the operon in response to the addition of theophylline in a dose dependant manor (Rudolph et al., 2013). Given that the most successful bottromycin (**1**) cluster variant contained this riboswitch (pCAPbtm8), it was decided to use the riboswitch in concert with PSF14 to drive expression of *tsaA*, and then use Paac3 to drive expression of the *tsaC-MT* operon. As PSF14 is the stronger promoter, this would allow an excess of TsaA over the tailoring enzymes. The riboswitch would allow the avoidance of any toxic effects resulting from overexpression until production, although the similar metabolite thioalbamide (**18**) has poor activity against bacteria (Frattaruolo et al., 2017). This new cluster was assembled from the AflII and SrfI digested plasmid backbone, two PCR fragments, and three oligonucleotides (Figure 85). The two PCR fragments contained *tsaA* and *tsaB* with the appropriate regions of homology with the backbone or oligonucleotides. Two of the oligonucleotides contained PSF14 and the riboswitch, whilst the other had Paac3 on it. The resulting construct was named pCAPtsa5.

When pCAPtsa5 was tested no thiostreptamide S4 (**17**) related metabolites could be detected by LC-MS. It is unclear why all attempts at refactoring this cluster failed. Whilst it would be interesting to explore why this cluster appeared to be resistant to refactoring, that was not within the scope of this project. Two promotor regions were replaced in these refactoring experiments, and it is possible that either region could be the issue although the evidence available from these experiments does not

KEY

  Restriction enzyme cut site

  Strong constitutive promoter

  Plasmid backbone

  Double stranded PCR fragment

  Single stranded oligonucleotide

  Riboswitch

  Site of recombination
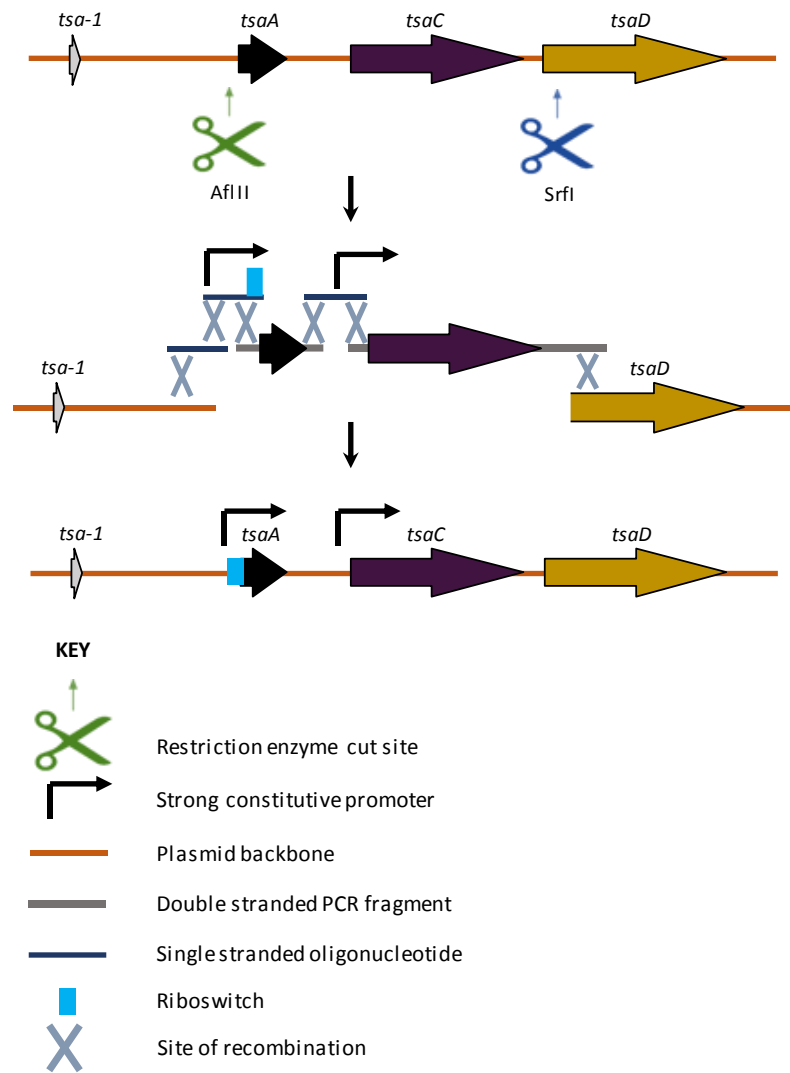
Figure 85. Refactoring pCAPtsa by assembly in yeast to produce pCAPtsa5

give enough information to unpick the reasons for this. The fact that the Δ*tsaA* cluster could not be complemented from another construct suggests that the position of *tsaA* at the beginning of the cluster could be important. Hence, new promotors inserted either side of *tsaA* would disrupt this native regulatory system, which may severely limit pathway production. The fact that the *tsaC* deletion could only be complemented by a construct containing its native promoter suggests that the region of DNA between *tsaA* and *tsaC* is important for *tsaC* expression. Due to the predicted extensive translational coupling of the *tsaC-MT* operon it is possible that most of the operon's expression is also reliant on *tsaC* expression. Therefore, the expression of the entire *tsaC-MT* operon is likely dependant on the native region of DNA between *tsaA* and *tsaC.* If the refactoring simply led to *tsaC* inactivity but other genes were properly expressed, then a series of distinctive metabolites would be detected, such as **69** and **68** (Figure 73). These metabolites were not detected. To unpick whether the disruption of the *tsaA* promoter region, the disruption of the *tsaC* promotor region, or

the disruption of both are the blame for the lack of production, future experiments should look at refactoring P*tsaA* or P*tsaC* independently.

3.2.2.2.   Precursor Peptide Modification

Natural product peptide backbones that originate from NRPSs are very difficult to engineer; making targeted changes to the NRPS that will result in a specific change in the peptide backbone is time consuming and very often unsuccessful (Winn et al., 2016). In contrast to this, the peptide backbone of RiPPs is encoded as a short peptide in the gene cluster. This means targeted modifications to the amino acids in this backbone are very easy to design; codons encoding one amino acid can be swapped to others. These changes can be made for many reasons, for example they can be used to probe biosynthesis, to test the substrate specificity of enzymes, and to install specific properties in to a compound. It was decided that this could be done with the thiostreptamide S4 (**17**) gene cluster. A series of modifications that were made to probe biosynthesis and test the substrate tolerance of enzymes have been previously mentioned: pCAPtsaS1T, SAV, A5F, and TsaCoreTaa (sections 3.2.1.5, 3.2.1.6, and 3.2.1.7). Five other modifications were also made to install specific properties and test substrate tolerance: pCAPtsaM3I, T8S, Y11V, H12A, and H12W (Figure 86).

The modifications to the precursor peptide were all made using the same system as previously described, rebuilding the region between the AflII and SrfI cut sites with two PCR fragments and a
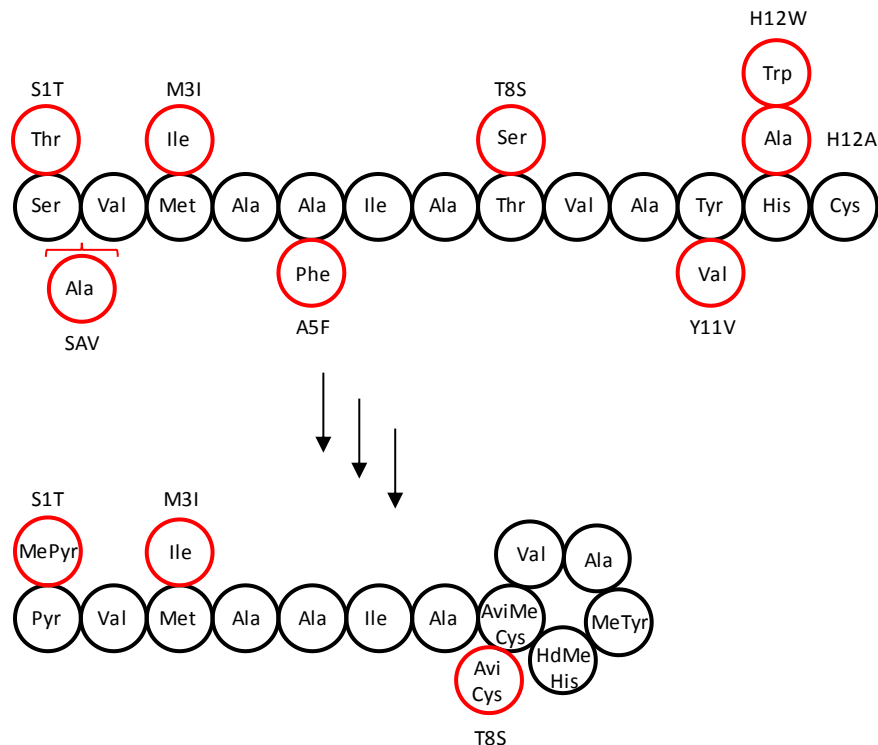
Figure 86. The attempted modifications made to the core peptide. The thiostreptamide S4 (**17**) core peptide and mature molecule is shown in black. At the top of the equation the attempted modifications are shown in red, and at the bottom of the equation the tolerated modifications are shown in red. The swap of the entire core peptide with the thioalbamide (**18**) core peptide (pCAPtsaTsaCoreTaa) is not shown.

single oligonucleotide (Figure 87). These assemblies were assessed by sequencing, showing that an 82% efficiency was achieved. The core peptide was contained entirely on the oligonucleotide, so by using the same reaction mix and just swapping in an oligonucleotide for each version of the core cluster different mutants were made in a very quick and easy manner. It has not gone unnoticed that this method is appropriate for upscaling to more high-throughput tests of the pathway's substrate tolerance, but time constraints prevented this from being fully explored within this PhD. Amino acid
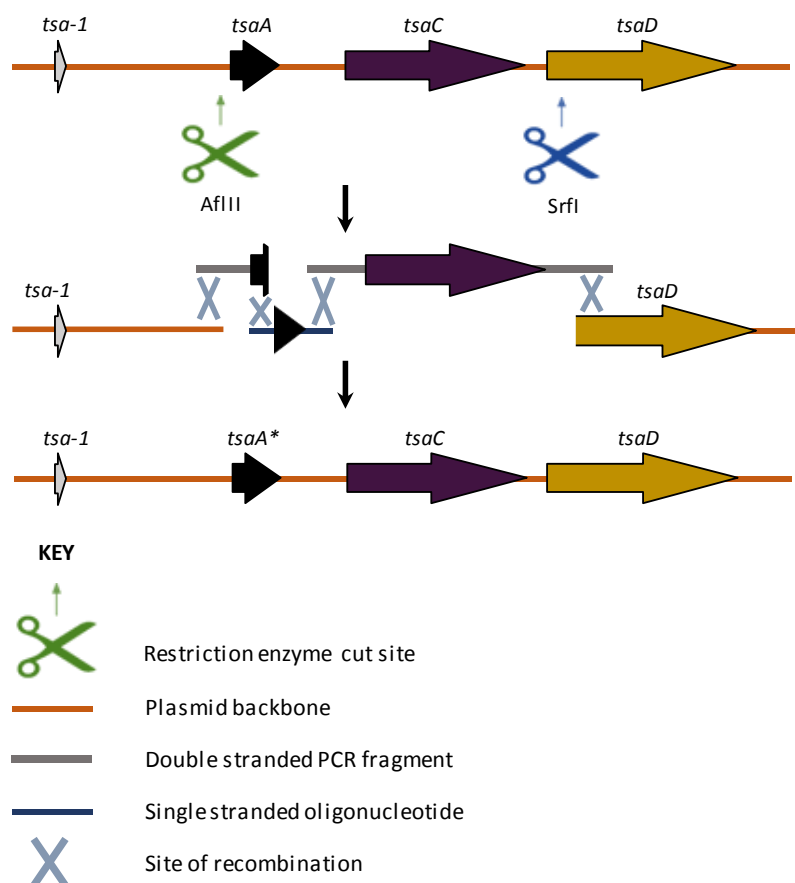
Figure 87. Yeast assembly-based modification of the core peptide on pCAPtsa.

```
A. alba         SVIGFAVTIAVHC
M. laminosus    SPMAAAVSIAYHC
N. potens       SVMAAAASVAAHC
S. olivoviridis SVMAAAASIALHC
S. NRRL S-87    SVMAAAATVAFHC
S. sparsogenes  SVMAAAATVAFHC
S. MUSC 125     SVMAAAATVAFHC
S. MUSC 14      SVMAAAATVAFHC
S. malaysiense  SVMAAAATVAFHC
S. NRRL S-4     SVMAAIATVAYHC
S. CNB091       SVMAAIATVAYHC
S. pacifica     TVGGLLVTPATHC
N. fuscirosea   SFTGIIVTAGVHC
M. eburnea      TFVSVVVTPATHC
```

Figure 88. Aligned core peptides from thioviridamide-like clusters.

substitutions, often inspired by other thioviridamide-like precursor peptides (Figure 88), were designed to test the substrate tolerance of the pathway, complement other investigations into the biosynthesis, and to have specific effects on the activity and stability of the molecule.

pCAPtsaT8S was constructed to assess whether the macrocycle could form with a dehydrated serine instead of a threonine. A serine residue is found in this position in the *S. olivoviridis*, *N. potens* DSM 45234 and *Mastigocladus laminosus* precursor peptides. pCAPtsaT8S was tested as usual in *S. coelicolor* M1146 on solid btmPM. This modification was tolerated by the thiostreptamide S4 (**17**) cluster; a version of the molecule with a mass (*m/z* 1363.53) and fragmentation that matches the
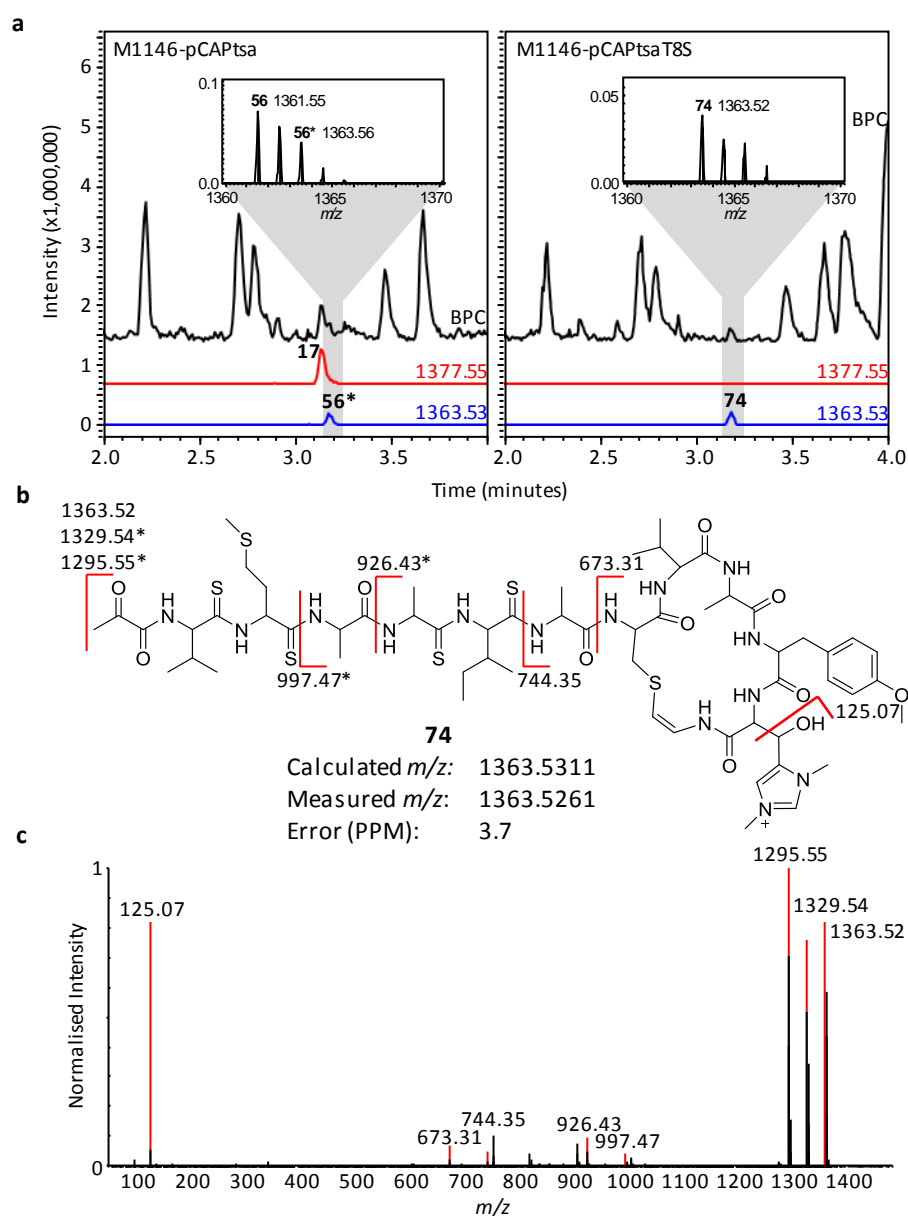
Figure 89. LC-MS analysis of M1146-pCAPtsa and M1146-pCAPtsaT8S. EICs of 1377.55 and 1363.53 are shown. **56*** labels the second isotope peak of **56**. Mass spectrums are shown to enable the distinguishing between **56** and **74**, as their retention times are the same. **b.** Predicted structure of **74**. Identified fragments are shown on the molecule. Fragments marked with an asterisk show a loss of 33.99, characteristic of the loss of SH$_2$ from thioamide bonds in fragmentation. **c.** MS^2 Fragmentation of **74**.

AviMeCys moiety being replaced with the expected AviCys moiety was detected (Figure 89). It would be interesting to see if there are any effects on bioactivity of the molecule, as both are seen in nature.

Position 11 of the core peptide seems to be highly variable across the TLM family (Figure 88), with aromatic phenylalanine and tyrosine residues found in this position, as well as smaller hydrophobic residues (alanine, valine and isoleucine), and threonine. To test whether the methylated tyrosine was important for the activity of thiostreptamide S4 (**17**), it was mutated to a valine, which is present in the equivalent position in the *A. alba* precursor peptide, to make pCAPtsaY11V. This was not tolerated by the pathway and no mature product was seen. However, an increase in the production of **65** was seen (Figure 90). Therefore, the threonine involved in macrocyclisation is dehydrated but a complete macrocycle is not formed. This is indicative of either inactive TsaE or TsaF, as shown by the deletion experiments (Figure 73). The proximity of the Y11V mutation to Cys13 suggests that the increase in yield of **65** may be due to poorly functioning TsaF, which is the HFCD predicted to be responsible for the cysteine decarboxylation required for macrocycle formation. This strongly suggests that replacing Tyr11 with a valine causes the HFCD to be less able to recognise this region of the molecule. This is comparable to results reported for the HFCDs responsible for the AviCys moiety in many lanthipeptides. These rely on the amino acids preceding the cysteine for recognition of their substrate (Ortega et al., 2017). Therefore, if Tyr11 is an important part of the TsaF recognition motif, it is possible that a HFCD from a cluster that does not have an aromatic residue at position 11 could recover activity to pCAPtsaY11V. The *A. alba* HFCD usually recognises a valine in position 11, so future experiments could look at expressing the *A. alba* HFCD alongside pCAPtsaY11V.

Position 12 of the core peptide is a histidine in every version of the precursor peptide identified. This position has a permanent positive charge installed on it by TsaG-mediated bis-methylation.
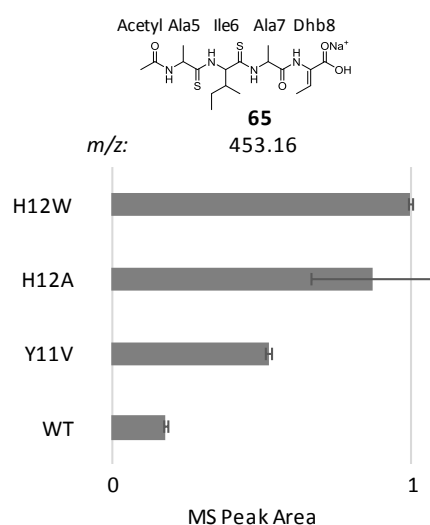
Figure 90. MS peak areas of **65** in H12W, H12A, Y11V, and wildtype clusters. The bar chart is normalised to the highest mass spectral area. The error bars represent the standard error.

Mitochondria within tumour cells show a more negative membrane potential, therefore favouring being targeted by positively charged drugs (Kalyanaraman et al., 2018). It is possible that thioviridamide-like molecules act as apoptosis inducers by targeting mitochondria with their permanent positive charge. To test the tolerance of the pathway to modifications here, and to test the residue's importance for activity, mutations were made at this location. Two modifications were made for this purpose: the histidine was swapped to an alanine, to determine if a neutral modification was tolerated and if the loss of the positively charged histidine affects activity, and to a
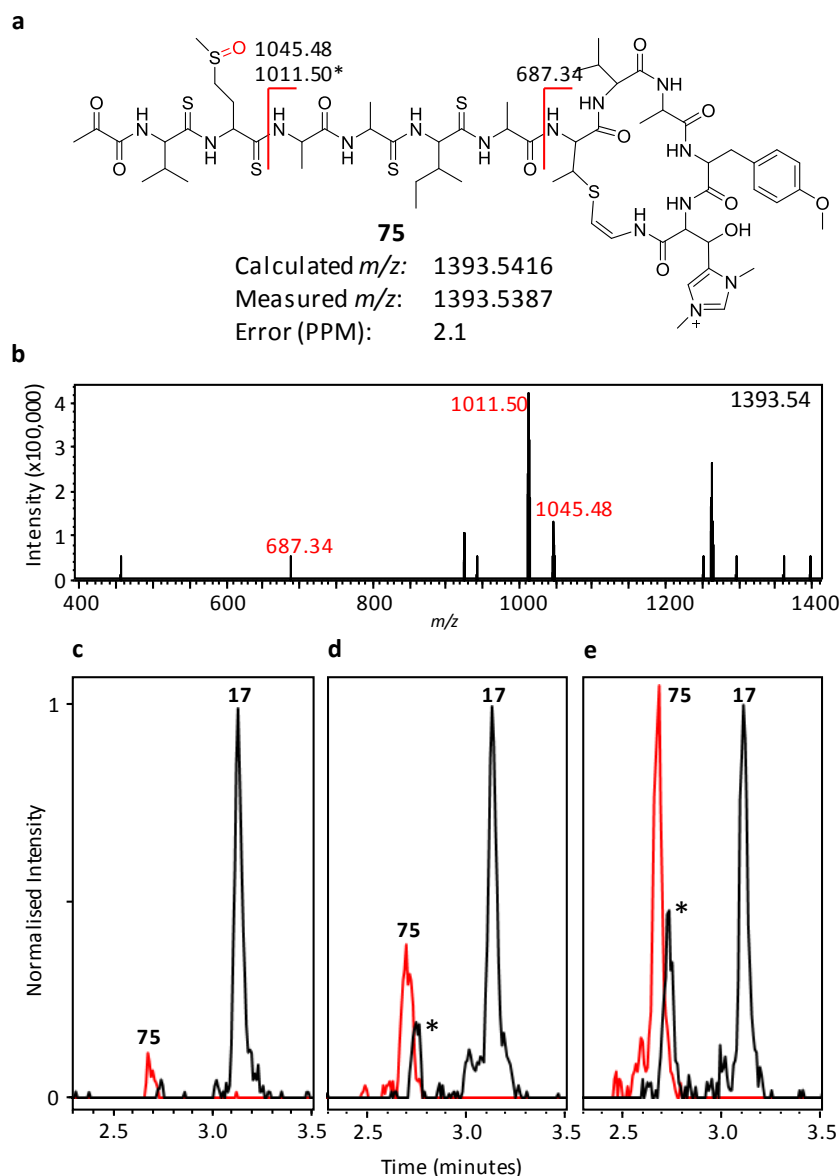
Figure 91. a. Structure of **75**, the methionine sulphoxide version of **17**. The oxidation is highlighted in red. The fragments seen during fragmentation are marked on the molecule. Fragments marked with an asterisk show a loss of 33.99, characteristic of the loss of SH_2 from thioamide bonds in fragmentation. **b.** Mass fragmentation of **75**. **c, d,** and **e.** EICs of *m/z* 1377.55 (black) and 1393.54 (red). Molecules **17** and **75** are labelled, the peaks labelled with an ***** are predicted to be a methionine sulphoxide version of **56**. Each chromatogram is normalised to the intensity of **17**, as quantitative comparison between these samples was not possible. **c.** EICs following step one of **75** purification; the crude methanol extract from M1146-pCAPtsa. **d.** EICs following step two of **17** purification; a liquid-liquid extraction using EtOAc. **e.** EICs following step three of **17** purification; Sephadex separation.

tryptophan, to see if a different nitrogen-containing aromatic amino acid was tolerated. These constructs were named pCAPtsaH12A and pCAPtsaH12W respectively. As with pCAPtsaY11V, produced an increased quantity of **65** whilst failing to produce any macrocycle containing products. This again suggests that TsaF is intolerant to modifications to this region of the molecule.

A common metabolite detected throughout growth and extraction of thiostreptamide S4 (**17**) is the methionine sulphoxide derivative (**75**; Figure 91). Met3 is particularly susceptible to oxidation, which is a problem for working with the molecule. During an attempted purification of **17**, the ratio of compounds **17** to **75** was 13:1 after the first step of purification, 3:1 after the second step of
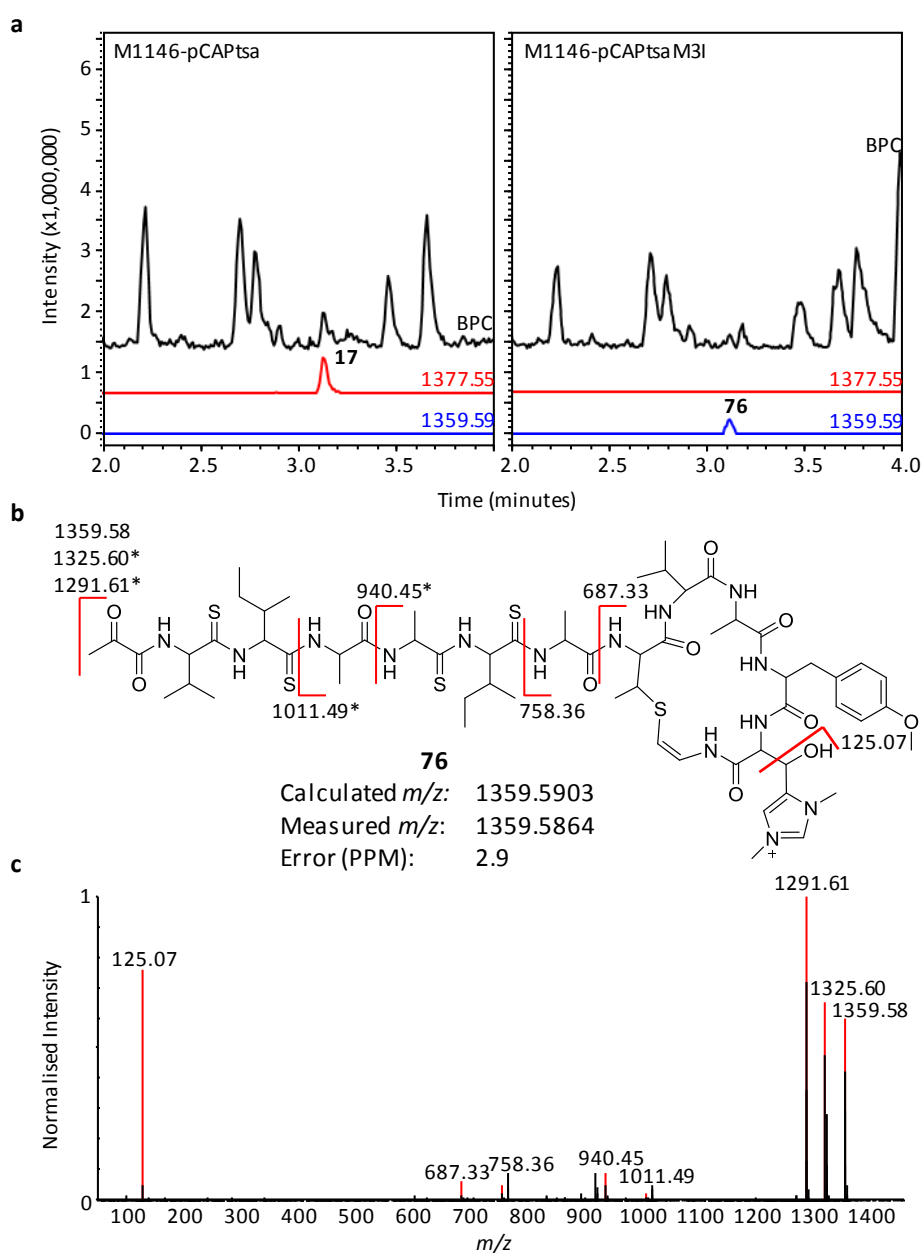
Figure 92. LC-MS analysis of M1146-pCAPtsa and M1146-pCAPtsaM3I. EICs of 1377.55 and 1359.59 are shown. **b.** Predicted structure of **76**. Identified fragments are shown on the molecule. Fragments marked with an asterisk show a loss of 33.99, characteristic of the loss of SH_2 from thioamide bonds in fragmentation. **c.** MS^2 Fragmentation of **76**.

purification, and 1:1 after the third step (Figure 91). This represents significant conversion of **17** to **75** throughout purification. This would likely be a problem if this molecule was used in a clinical setting, as the methionine sulphoxide version of a similar molecule, thioholgamide, is around ten times less active than un-oxidised thioholgamide (Kjaerulff et al., 2017).

Whilst most thioviridamide-like core peptides contain a methionine as their third residue, there is variation. For example, core peptides in the pathways in *S. pacifica*, *M. eburnea*, and *A. alba* have a glycine, a valine, and an isoleucine at this position respectively. To engineer thiostreptamide S4 (**17**) into a more stable molecule, a version was made with Met3 swapped for an isoleucine, as in the *A. alba* precursor peptide (Figure 88). Thioalbamide (**18**) has strong cytotoxic activity (Frattaruolo et al., 2017) and so it was predicted that the methionine to isoleucine mutation would not lose activity, whilst becoming more resistant to oxidation. This plasmid was named pCAPtsaM3I, and production was tested following conjugation into *S. coelicolor* M1146. A thiostreptamide S4-like molecule of *m/z* 1359.59 was detected, and the fragmentation allowed us to pin down the change in mass the N-terminal moiety, the first amino acid, or second amino acid (Figure 92). This is consistent with the methionine being swapped to an isoleucine. Whilst there was slightly lower production than the wild type thiostreptamide S4 (**17**), it was still produced in levels that will be sufficient for future purification for activity assays.

3.2.3. Discovery of New Clusters

3.2.3.1. TsaD Homologues

The role of TsaD in the macrocycle formation was unclear, however it was partnered with a phosphotransferase and the metabolomic data indicated that it was required for threonine dehydration, and so could be acting as a lyase. This is therefore a previously unidentified and uncharacterised enzyme involved in RiPP biosynthesis. It was hypothesized that TsaD could be leveraged to direct the discovery of novel RiPPs. To do this, relatives of TsaD were analysed.

BLASTP was used to identifiy homologues of TsaD. Using the default settings 568 homologues were found. 15 of these homologues were from thioviridamide-like clusters, with identities varying from 99% down to 37%. Those that were not from thioviridamide-like clusters had low identities varying from 35% to 24%. When a diverse selection of these were submitted to Phyre2, they were predicted to have the same structural similarity to HopA1 as TsaD. This reassured that, whilst they had low identities, they still were genuine homologues of TsaD. Interestingly, the only three proteins from this analysis that are annotated as anything other than "hypothetical protein" are annotated as LanM or LanC-like cyclases. These turned up in the lower half of the 568 results, indicating poor similarity and explaining why these had not been noticed before. A representative set of 171 proteins across the 568 homologues were aligned using MUSCLE (Edgar, 2004) and used to construct a phylogenetic tree with RAxML-HPC2 on XSEDE via the CIPRES Science Gateway (Figure 93; Miller et al., 2010; Stamatakis, 2014). The genomic context of the TsaD homologues in this tree was assessed. By combining this analysis with the position of the TsaD homologues on the tree, it could be identified that the TsaD homologues grouped on the tree based on the gene clusters they were a part of. Five main groups of TsaD homologues could be identified: those part of lanthipeptide-like clusters, those part of thioviridamide-like clusters, and those part of three other similar clusters that were labelled type 1, type 2, and type 3 (Figure 93). In all clusters the *tsaD*-like gene was preceded by a phosphotransferase gene.

3.2.3.2. Lanthipeptide-Like Clusters

The lanthipeptide-like clusters are identified as such because they contain a LanC-like cyclase and the putative precursor peptides show a cysteine and serine rich C-terminal region typical of lanthipeptide core peptides. The example shown is the *Mycobacterium tuberculosis* cluster (Figure 93). TsaD showed good identity with the N-terminal half of what is annotated as a LanM in this cluster. However, if this *M. tuberculosis* LanM is compared using BLASTP against the enterococcal cytolysin (7) LanM (Dong et al., 2015), only the C-terminus shows a small amount identity (18%). This
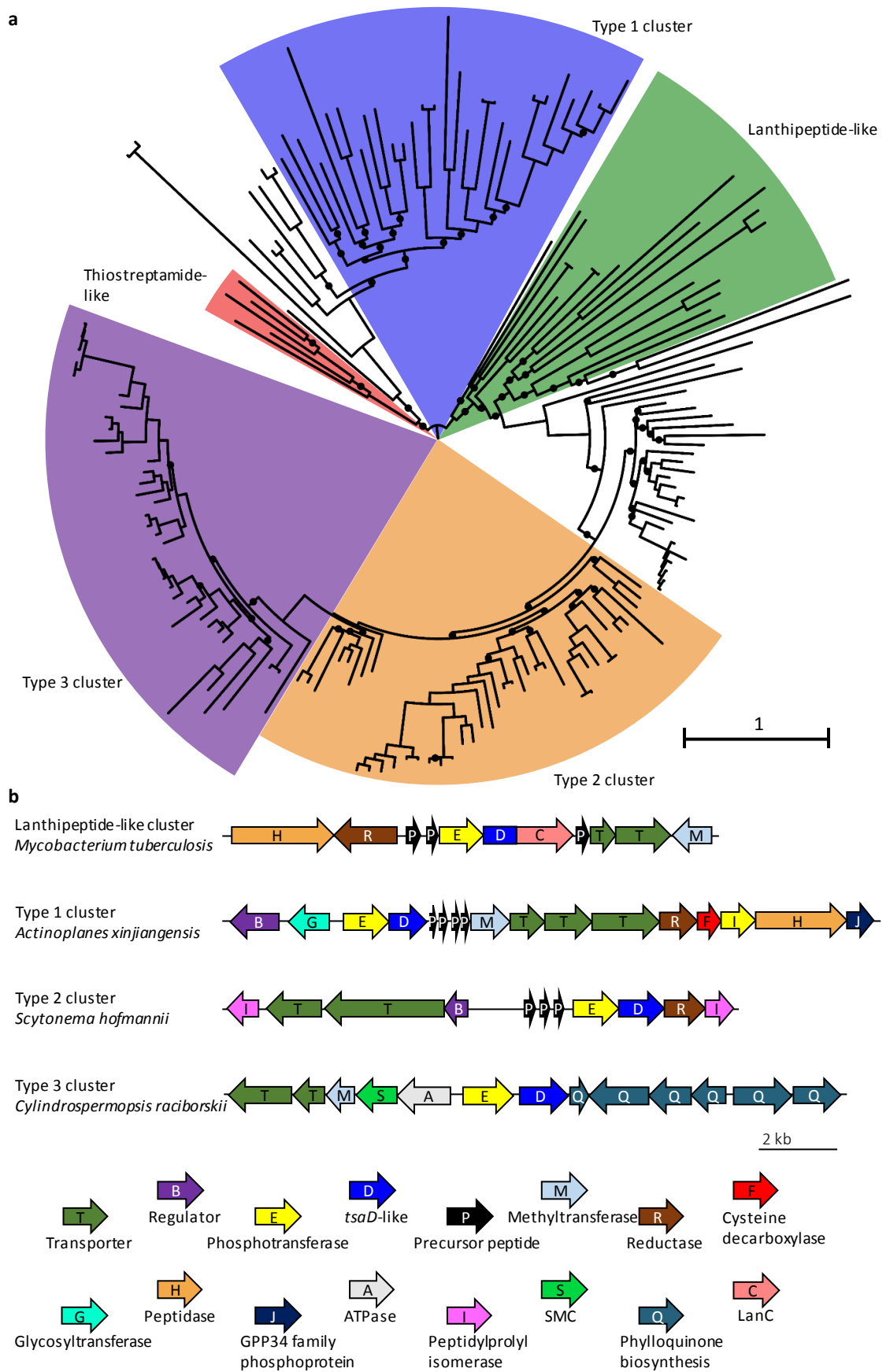
Figure 93 a. Phylogenetic tree of TsaD-related proteins. Coloured wedges group the proteins into similar genetic contexts. Dots indicate bootstrap values <50 %. **b.** Representative example of lanthipeptide-like and type 1, 2, and 3 clusters.

suggests the gene was annotated as *lanM* due to its C-terminus, however the N-terminus has been replaced by a TsaD-like enzyme. The N-terminus of LanM proteins is responsible for phosphorylation mediated dehydration of the serine/threonine residues required for lanthionine bond formation (van der Donk and Nair, 2014), whilst the C-terminus is responsible for cyclisation, using a zinc atom to activate the cysteine sulphur atom. It can therefore be suggested that the dehydratase domain of the *M. tuberculosis* LanM has been replaced by a TsaD-like enzyme (forming a protein fusion) and its associated preceding phosphotransferase; with this pair of enzymes now catalysing the dehydration. The cyclase like C-terminus of the *M. tuberculosis* LanM is still present with the residues needed to coordinate a zinc atom to activate the cysteine sulphur atom (Li et al., 2006). This is all backed up by Phyre2 results for the *M. tuberculosis* LanM, which show the C-terminus having a 100% confidence alignment with the cyclase region of enterococcal cytolysin (**7**) LanM, whilst the N-terminus has a 100% confidence alignment with the *P. syringae* type III effector HopA1 (the alignment that TsaD also shares). It is possible that by investigating TsaD homologues, a new class of lanthionine bond forming enzymes has been uncovered.

3.2.3.3. Type 1 Clusters

It is possible that the type 1 clusters will make Avi(Me)Cys or avionin containing RiPPs, as they are also always associated with HFCDs that are similar to the ones encoded in thioviridamide-like clusters. A representative gene cluster example (Figure 93) from *Actinoplanes xinjiangensis* is shown. This has a sequence of four putative precursor peptides all containing a SxxTxxxxxC motif at their C termini. This not only provides the potential for AviMeCys residue, but also the potential for an avionin moiety as seen in the lipolanthines (Wiebach et al., 2018). These clusters also contain a putative glycosyltransferase, suggesting this is possibly a glycosylated RiPP. There are a few examples of glycosylated RiPPs, for example the glycocins (Norris and Patchett, 2016) and NAI-112 (Iorio et al., 2014). NAI-112 is a tryptophan-glycosylated lanthipeptide that contains a C-terminal labionin moiety arising from a SxxTxxxC motif. The *Actinoplanes xinjiangensis* precursor peptides also contain tryptophan residues in the precursor peptides, which could be similar targets for glycosylation. The lanthionine bonds in NAI-112 are synthesized using class III enzymes, whereas the type 1 clusters lack these enzymes, suggesting that the type 1 clusters are distinct from the NAI-112-like clusters.

One unanswered question with thiostreptamide S4 (**17**) macrocycle formation is the precise role for TsaE, a phosphotransferase whose gene precedes that of the HFCD, in biosynthesis. The type 1 clusters also contain an additional phosphotransferase associated with their HFCD gene. The fact that that both thioviridamide-like clusters and type 1 clusters contain homologues of *tsaC-F* supports the conclusion from the deletion experiments that the phosphotransferases (TsaC and TsaF) in the thioviridamide-like clusters perform distinct reactions.

3.2.3.4.   Type 2 Clusters

The type 2 clusters are again associated with putative precursor peptides, for example the cyanobacterial *Sytonema hofmannii* cluster encodes three short peptides (Figure 93). The putative precursor peptides in these clusters are identified as being Nif11-related; Nif11-like proteins have been shown to act as precursor peptides in cyanobacterial genomes (Haft et al., 2010). These previously observed clusters contain multiple precursor peptides with highly conserved leader peptides, a classic GG cleavage motif, and highly variable core peptides, and these features are all seen in the type 2 clusters. The Nif11-like precursor peptides have been reported to be associated with thiazole, oxazole, and lanthipeptide biosynthetic machinery (Haft et al., 2010), however the type 2 clusters differ in that they are associated with peptidylprolyl isomerases. Intriguingly, the precursor peptides do not contain residues capable of forming lanthionine, AviCys, or AviMeCys macrocycles; the *S. hofmannii* precursor peptides do not even contain a single cysteine. This suggests that TsaD may not be limited to working in dehydrations involved in lanthionine or AviCys bond formation. Identification of the products of type 2 clusters, and then gene deletion experiments, would be required to demonstrate the role the TsaD-like protein and its associated phosphotransferase play here. The transporter associated with these clusters is of the ABC-transporter maturation and secretion (AMS) type (Havarstein et al., 1995). These bifunctional enzymes are commonly associated with class I and II lanthipeptides and couple proteolysis of the leader peptide with export of the final product (Franke et al., 1999). This not only suggests that the product of type 2 clusters is cleaved off the leader peptide by this transporter/protease fusion protein as it is exported from the cell, but it also provides support that a genuine RiPP gene cluster has been discovered.

3.2.3.5.   Type 3 Clusters

The type 3 clusters are distinct from type 1 and 2 clusters, as putative precursor peptides could not be found associated with them, and so may represent non-RiPP clusters. They are reliably associated with phylloquinone biosynthesis genes (Widhalm and Rhodes, 2016). They are also regularly associated with a structural maintenance of chromosomes (SMC) gene. The role the TsaD-like protein and its associated phosphotransferase plays here is very unclear and identifying the true product of the type 3 clusters would be the first step in uncovering this. It is also possible that this is a part of primary metabolism, and that the TsaD-like homologue and its associated phosphotransferase are involved in the phosphate handling that is required in phylloquinone biosynthesis (Nowicka and Kruk, 2010).

3.2.3.6.   Hypothesis of TsaD Function

This phylogenetic analysis has revealed that TsaD-like proteins are likely involved in a variety of RiPP pathways. In the lanthipeptide-like clusters and the thioviridamide-like clusters it is associated with

cyclase machinery. The type 1 clusters are predicted to form Avi(Me)Cys or avionin cycles, so the TsaD-like protein is again associated with cyclase machinery. The type 2 and type 3 clusters are slightly stranger, as the precursor peptides of the type 2 clusters do not appear to be capable of forming lanthionine, labionin, Avi(Me)Cys, or avionin macrocycles, and the type 3 clusters do not appear to be RiPPs. As it is always associated with a preceding phosphotransferase, it can be suggested that TsaD works as a partner enzyme with the phosphotransferase. As a TsaD-like protein is fused to a LanC in the lanthipeptide-like clusters, and LanC contains the precursor peptide-binding region (Dong et al., 2015), it is unlikely TsaD-like proteins are solely responsible for substrate recognition/binding. Therefore, this provides support to the hypothesis that TsaD-like proteins act as lyases to remove phosphate groups after the partner phosphotransferase has installed it.

3.3. Conclusion

3.3.1.   Summary of Results

3.3.1.1.   Final Pathway

The analysis of the metabolites produced by expressing the thiostreptamide S4 (**17**) wildtype cluster and the deletion clusters in *S. coelicolor* M1146 resulted in the identification of **17** and **55-69** (Figure 94; Table 7), mainly by detailed LC-MS characterisation. This information allows a biosynthetic scheme to be proposed (Figure 95). The first step is the thioamidation of the core peptide, **54**, to produce **77**. The biosynthetic origin of this modification was previously unknown. By process of elimination the deletion experiments show that TsaH and TsaI, the YcaO and TfuA, are responsible for this. This is supported by recent publications describing YcaO and TfuA mediated thioamidation in archaea (Mahanta et al., 2018; Nayak et al., 2017) and in bacterial thiopeptides (Schwalen et al., 2018).

The next step is dehydration of Thr8 to Dhb8 to produce **78**. This is catalysed by TsaC and TsaD. TsaC is a phosphotransferase, and so dehydration likely follows the mechanism typical of class II, III, and IV lanthipeptides; phosphorylation and elimination. The data presented here suggests that TsaD is the lyase responsible for elimination. Following this step, Ser1 is dehydrated to produce **79**. This likely follows a similar mechanism but with TsaE catalysing the phosphorylation. It also cannot be ruled out that TsaC catalyses this phosphorylation. It is unusual for dehydrations in RiPPs to be performed by non-lanthipeptide proteins (Repka et al., 2017); this therefore represents a novel set of enzymes catalysing this.

Following dehydrations, the macrocycle is formed. This requires Cys13 to be decarboxylated, producing **80**. Bioinformatic predictions of function give strong support to TsaF being a HFCD responsible for cysteine decarboxylation. The deletion experiments also support this by showing that a macrocycle cannot form in the absence of *tsaF*. The formation of the AviMeCys macrocycle that produces **81** is predicted to be non-enzymatic, however the fact that it seems to follow Ser1 dehydration suggests some control over the timing of this macrocycle formation. As it would be unusual for a HFCD to have substrate requirements as far away as Ser1, there is possibly a mechanism that delays Cys13 decarboxylation or macrocycle formation; this could be a steric effect caused by any of the tailoring enzymes that bind the precursor peptide prior to macrocycle formation.

The next step is methylation of His12, producing **82**. Deletion experiments show that TsaG is responsible for this step. Histidine bis-methylation has not been seen in natural products before, and it provides a positive charge that may be key in making these compounds active against cancer cell

17 1377.55: R$_1$ = **a**, R$_2$ = **c**
55 1363.53: R$_1$ = **b**, R$_2$ = **c**
56 1361.55: R$_1$ = **a**, R$_2$ = **d**
57 1347.54: R$_1$ = **b**, R$_2$ = **d**
58 1318.50: R$_1$ = **b**, R$_2$ = **e**

59 974.44: R$_1$ = **a**, R$_2$ = **c**
60 960.43: R$_1$ = **b**, R$_2$ = **c**
61 958.45: R$_1$ = **a**, R$_2$ = **d**
62 944.43: R$_1$ = **b**, R$_2$ = **d**
63 915.39: R$_1$ = **b**, R$_2$ = **e**

a   **b**   **c**   **d**   **e**

64 628.28: R$_1$ = **b**, R$_2$ = **e**

66 529.24

69 480.15

67 442.23

65 430.17

68 347.13

Figure 94. Structures of compounds **17** and **65**, and predicted structures of compounds **55-64** and **66-68**.

Table 7. Identified Metabolites produced by the clusters containing deletions of each gene. Dark shading indicates the detection of a metabolite by LC-MS. MT – methyltransferase, PT – phosphotransferase, HFCD – homo-oligomeric flavin-containing cysteine decarboxylase, PP – precursor peptide.

Molecule	17	55	56	57	58	59	60	61	62	63	64	66	69	67	65	68
Pyruvyl	Yes	Yes	Yes	Yes	Yes	No	No	No	No	No	No	No	No	No	No	No
Macrocycle	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No	No	No	No	No
Thioamide	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes
WT	■		■	■		■	■	■					■	■	■	■
Δ*tsaK* Protease	■		■			■	■	■					■	■	■	■
Δ*tsaMT* O-MT		■		■		■		■	■	■			■	■	■	■
Δ*tsaJ* Hydroxylase			■	■				■	■	■			■	■	■	
Δ*tsaG* N-MT					■					■	■		■	■	■	
Δ*tsaE* PT												■	■	■	■	■
Δ*tsaF* HFCD												■	■	■	■	■
Δ*tsaC* PT													■	■	■	■
Δ*tsaD* Unknown													■			■
Δ*tsaH* YcaO-like																
Δ*tsaI* TfuA-like																
Δ*tsaA* PP																

Figure 95. Proposed thiostreptamide S4 (**17**) biosynthetic pathway. Peptide labels the leader peptide. **54** is the predicted precursor peptide. **77**-**82** are predicted intermediates.

lines (Kalyanaraman et al., 2018). Whilst TsaG is not a structurally new enzyme, and is homologous to protein arginine N-methyltransferases, this is an example of it catalysing a new, and possibly clinically relevant, modification. Gene deletion experiments show that this methylation acts as a gatekeeper for subsequent modifications: His12 β-hydroxylation and Tyr11 O-methylation, installed by TsaJ and TsaMT respectively. It is likely that these proteins have evolved to only act on substrates containing a bis-methylated histidine. Whilst mature thiostreptamide S4-like molecules are rarely seen lacking the histidine N-methylations, it is not uncommon to see significant quantities of mature thiostreptamide S4-like molecules lacking the histidine β-hydroxylation and tyrosine O-methylation. Therefore, these modifications are not a prerequisite for the leader peptide cleavage and associated pyruvyl formation.

There is no good experimental evidence that allows the assignment of an enzyme to the cleavage of the leader peptide. However, bioinformatic analysis shows that that TsaK is closely related to cysteine proteases, and these are seen cleaving the leader peptides of RiPPs with recognition motifs similar to those seen in the thioviridamide-like precursor peptides (Furgerson Ihnken et al., 2008). The small change in production observed when *tsaK* is deleted is suggested to be because endogenous proteases can also remove the leader peptide, as is seen in many class III lanthipeptides (Völler et al., 2012). Following proteolysis, the pyruvyl group is likely to be formed spontaneously from Dha1. The production of a 2-oxobutyryl group following the Ser1 to Thr1 substitution provided the first experimental evidence that these moieties arise from amino acids and are not installed by pyruvyl transferases (Katzen et al., 1998).

Considering its strong association with thioviridamide-like clusters, it was strange that deletion of *tsaL* had no effect. TsaL is predicted to be a membrane associated protein. A couple of possibilities are that it could play a regulatory role related to signal transduction or it could play a role in thiostreptamide S4 (**17**) transport. If it is regulatory, it could be that the non-native background of *S. coelicolor* M1146 makes its role unimportant. If it plays a role in transport, then our experiments may not pick that up, as the extraction procedure used did not distinguish between extracellular and intracellular metabolites.

3.3.1.2.   **Precursor Peptide Modifications**

Initial experiments to modify the precursor peptide were conducted. These showed that the precursor peptide was not very tolerant to changes (Figure 86). Minor changes, such as serine to threonine (S1T), methionine to isoleucine (M3I), and threonine to serine (T8S), produced the expected variants on thiostreptamide S4 (**17**). This is a complex biosynthetic pathway with many tailoring enzymes that would all have to tolerate these changes, and as such it is still an encouraging result that these changes were tolerated. Major changes, like an alanine to phenylalanine (A5F)

mutation and a full core peptide swap with the thioalbamide (**18**) core (TsaCoreTaa), were not tolerated.

The results obtained with the modifications close to the C-terminal cysteine (Y11V, H12A, and H12W) were of interest because, whilst they were not fully tolerated, the metabolomic analysis indicated why they were not tolerated. Mutational analysis of the biosynthetic gene cluster identified metabolites that were seen with each different gene deletion. Production from *S. coelicolor* M1146 expressing Y11V, H12A, and H12W modified clusters could be matched up to with the increase in yield of a metabolite (**65**) from the *tsaF* deletion cluster, suggesting that the strains expressing these clusters were unable to produce mature thiostreptamide S4-like molecules because of the intolerance of TsaF, the cysteine decarboxylase. This gives a target for future cluster engineering projects when trying to improve the substrate tolerance of the cluster.

Possibly the most exciting modification that was tolerated was the methionine to isoleucine (M3I) mutation. This is because it represents a targeted modification to install a specific beneficial property into thiostreptamide S4 (**17**): an issue was identified (the propensity to oxidise), a solution was designed (swapping the problem amino acid), and the engineered modification was tolerated.

3.3.1.3.   Gene Cluster Discovery

Understanding the biosynthesis of the pathway enables the identification of interesting biosynthetic steps. These new interesting steps can then be used to search for new natural product diversity. For example, the macrocycle formation is clearly a very interesting step. The formation of Avi(Me)Cys macrocycles in the absence of lanthipeptide-like genes is not well understood. There are two enzymatic steps needed for macrocycle formation in thiostreptamide S4 (**17**): Thr8 dehydration and Cys13 decarboxylation. It is unusual to see the dehydration step catalysed by non-lanthionine-type enzymes. The proteins proposed to be a new type of lyase involved in this step, TsaD, inspired the building of a phylogenetic tree of its relatives (Figure 93).

By analysing relatives of TsaD two new classes of RiPP and a new class of lanthipeptides may have been discovered. These classes have likely been missed before due to the way RiPPs are typically identified during genome mining. Many methods are guided by commonly RiPP-associated biosynthetic genes, for example YcaO-like proteins (Cox et al., 2015), because new precursor peptides are difficult to identify (Tietz et al., 2017). TsaD was only revealed as a RiPP-associated enzyme with biosynthetic importance because of the work in this thesis, and so has not been used to guide genome mining efforts before. The predicted lanthipeptide-like clusters that were revealed as a part of this analysis could have been identified in LanC-guided genome mining, but because the TsaD-like-LanC-like fusion genes were misannotated as classical lanthionine bond forming enzymes they would be easy to miss. The type 1 and type 2 RiPP clusters would also be easy to miss in classical

genome mining efforts due to their unique architecture and gene complement, for example the novel close association of type 2 clusters with peptidylprolyl isomerases.

3.3.2.   Future Work

3.3.2.1.   Bioactivity Assessments

Whilst the biosynthesis of thiostreptamide S4 (**17**) is particularly novel, and the interest in this was the driving force behind this project, there is also reason to believe it may have clinical potential. Thioviridamide-like molecules are often very promising anticancer compounds (Frattaruolo et al., 2017; Hayakawa et al., 2006a; Kjaerulff et al., 2017). It is therefore important that thiostreptamide S4 (**17**) is also tested for its activity.   The histidine methylations are predicted to be vital for thioviridamide-like molecules' activities. A version of thiostreptamide S4 (**17**) lacking these histidine methylations (**58**) was produced when *tsaG* was deleted, and so testing this molecule for activity would be key in starting to understand the structure-activity relationship. The activity of **76**, produced when Met3 was swapped to an Ile (pCAPtsaM3I), would also be very interesting to know, as this compound represents an engineered more stable version of thiostreptamide S4 (**17**). If **76** retains activity it could be a version of thiostreptamide S4 (**17**) more likely to be clinically relevant. These molecules are in the process of being purified in the group, and a collaboration is in place to test their bioactivity (Frattaruolo et al., 2017).

3.3.2.2.   *In Vitro* Characterisation

The only *in vitro* characterisation of the installation of a thioamide bonds is in an archaeal system (Mahanta et al., 2018). As this is a modification made to the methyl-coenzyme M reductase enzyme in archaea, it may have significant differences to the bacterial system that works on thioviridamide-like molecules. An additional difference is that whilst in the archaeal system only a single thioamide bond is installed, the thioviridamide-like compounds are the only known examples of multiple thioamide bonds being introduced by a single YcaO and TfuA pair. Therefore, it would be interesting to attempt to reconstitute this step of thiostreptamide S4 (**17**) biosynthesis *in vitro*. This would allow investigation into the selectivity of the enzymes and help explain why some bonds are thioamidated, but not others. This may also help reveal the role of the TfuA protein, as a standalone YcaO protein is catalytically sufficient to activate the peptide backbone in some archaea (Mahanta et al., 2018). Additionally, the sulphurtransferase associated with the thiostreptamide S4 (**17**) gene cluster could then be tested to see if it can liberate the sulphur for the thioamide bonds from cysteine or another substrate. There would likely be some challenges when attempting this, as YcaO enzymes are often difficult to purify in a soluble form (Franz et al., 2017). This is, however, a good system to attempt this in as the deletion experiments show that the YcaO and TfuA proteins, TsaH and TsaI, are likely to act first in biosynthesis. Therefore, the substrate would be an easy-to-obtain unmodified peptide.

3.3.2.3.   Biosynthesis-Informed Cluster Engineering

The good understanding of thiostreptamide S4 (**17**) biosynthesis gained in this work is important for future cluster engineering projects. Being able to assign roles to the genes in the cluster makes it easier to plan the modifications needed to make targeted changes to the final molecule. For example, the deletion of *tsaG* allowed the production of a demethylated histidine version of thiostreptamide S4 (**17**) that, if purified, could be tested to uncover whether these modifications are key for activity. An understanding of the biosynthesis also aids investigations into what goes wrong when modifications are made, for example in the precursor peptide modifications. An example is given here of one insight these deletion experiments gave, and how this could be used to fix modifications that were not tolerated.

Metabolomic data from the Y11V, H12A, and H12W mutant clusters allowed the suggestion that TsaF, the HFCD, could not tolerate the precursor peptide modifications. In looking to address the intolerance of TsaF, it is useful to assess the natural diversity of HFCD enzymes. The thioviridamide-like compounds are not the only natural products that use HFCD enzymes. There are also linaridins, lanthipeptides, and lipolanthines that use HFCD enzymes to decarboxylate their C-terminal cysteines (Goto et al., 2010; Kupke et al., 1995; Wiebach et al., 2018). In addition, the HFCD involved in the synthesis of the Avi(Me)Cys-containing lanthipeptides has been reported to be similar to those involved in the biosynthesis of the dithiolopyrrolones (Li and Walsh, 2010; Ortega et al., 2017). When diverse examples from these classes are built into a phylogenetic tree, the HFCDs appear to form three clades based on their class (Figure 96). The linaridins and the lipolanthines clade close together, the thioviridamide-like HFCDs sit quite distinct from other any of the others. The lanthipeptide clade
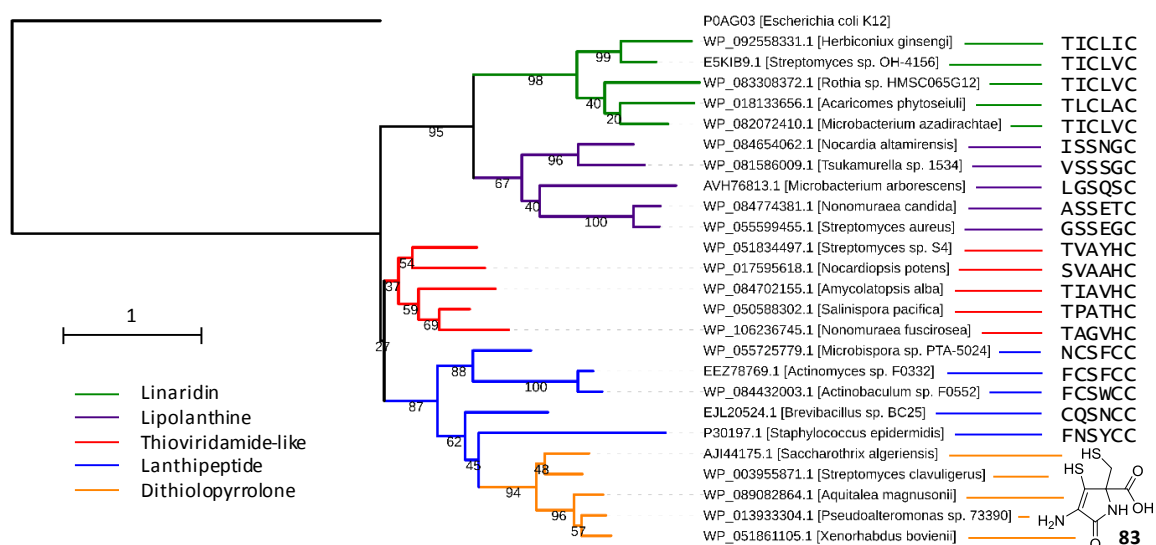
Figure 96. Phylogenetic tree of HFCD genes built using 5 examples from each natural product class that uses a HFCD enzyme in biosynthesis. The HFCD substrate is shown alongside each branch, for RiPPs this is the terminal 6 amino acids. UbiX (P0AG03), a similar non-HFCD protein, is used as the outgroup (Gulmezian et al., 2007; Rangarajan et al., 2004).

contains the dithiolopyrrolones HFCDs. This is particularly interesting because the lanthipeptide examples all have a C-terminal cysteine-cysteine dipeptide and the HFCD substrate (**83**) in dithiolopyrrolone biosynthesis is predicted to derive from a cysteine-cysteine dipeptide (Li and Walsh, 2010). Therefore, it is possible that the dithiolopyrrolone HFCDs have evolved from the lanthipeptide HFCDs, with just a small change in substrate tolerance required.

The precursor peptides associated with each HFCD seem to clade based on the amino acids preceding the terminal cysteine, and it is likely that these features are important for HFCD recognition (Ortega et al., 2017). Therefore, this phylogenetic tree could aid in picking HFCDs that could complement modifications to the macrocycle amino acids that are not tolerated by the native HFCD. The linaridins and lanthipeptides seem to have very little variation in the 5 amino acids preceding the cysteine. Whilst many of the lanthipeptide HFCDs seem to be able to tolerate a large aromatic residue in the same position as the tyrosine in the thioviridamide-like molecules, they seem to almost universally have a cysteine in the same position as the histidine in the thioviridamide-like molecules. They therefore would likely have trouble rescuing any modification to the thiostreptamide S4 (**17**) histidine other than a change to cysteine. Instead, the information here could inform experiments in which all the thiostreptamide S4 (**17**) macrocycle amino acids are replaced with a sequence seen on the tree, and then this change could be complemented with the matching HFCD.

An alternative method for rescuing changes to the macrocycle amino acids would be seeing which HFCD has the highest substrate tolerance. The HFCD EpiD is very tolerant to changes in the peptide leading up to the C-terminal cysteine (Kupke et al., 1995) and can tolerate many amino acids including tyrosine at -2 from the cysteine, and a variety of amino acids at -1 from the cysteine. It is therefore possible that this could also act on the thiostreptamide S4 (**17**) cysteine, whilst being tolerant of changes to the amino acids preceding this cysteine. It would be interesting to see if EpiD could complement the *tsaF* deletion, and whether it could restore productivity to the pCAPtsaY11V, H12A, and H12W clusters. This would then allow many changes to be made to the core peptide without having to associate each change with a specific HFCD enzyme.

The ability to modify the production of the gene cluster would be particularly useful when paired with bioactivity data and information on mechanism of action. This information could feed back into modifications of the gene cluster to tailor thiostreptamide S4 (**17**) towards its desired therapeutic use in oncology. Ultimately this could also lead to a suite of thiostreptamide S4-like molecules that have different properties, for example molecule **76** from M1146-pCAPtsaM3I that was engineered for enhanced stability. Drugs for oncology have some of the lowest success rates whilst going through clinical trials, at 3.4% (Wong et al., 2018). Therefore, being able to approach trials with a

suite of related thiostreptamide S4-like molecules will increase the chance of finding a viable and clinically useful drug.

Chapter 4 – Materials and Methods

4.1. Methods

4.1.1. Chemicals and Media Components

Unless otherwise noted, all chemicals and media components were from Sigma Aldrich. All restriction enzymes were purchased from New England Biolabs. The final concentrations of antibiotics were: 50 µg mL^{-1} kanamycin, 50 µg mL^{-1} hygromycin, 50 µg mL^{-1} apramycin, 50 µg mL^{-1} carbenicillin, 25 µg mL^{-1} chloramphenicol, and 25 µg mL^{-1} nalidixic acid. All primers and oligonucleotides were ordered at a High Purity Salt Free (HPSF) grade from Eurofins Genomics. Ultrapure water was obtained using a Milli-Q purification system (Merck) and all media and solutions were autoclaved prior to use, unless otherwise stated.

4.1.2. Strains

Streptomyces scabies DSM 41658, *Streptomyces coelicolor* M1146, *Streptomyces lividans* TK24, and *Streptomyces albus* J1074 were cultured on SFM (2% soy flour (Holland & Barrett), 2% mannitol, and 2% agar (Formedium)). *Streptomyces venezuelae* NRRL B-65442 was cultured on MYM-TAP (0.4% maltose, 0.4% yeast extract (Merck), 1% malt extract, 40 µg mL^{-1} ZnCl$_2$, 200 µg mL^{-1} FeCl$_2$•6H$_2$O, 10 µg mL^{-1} CuCl$_2$•2H$_2$O, 10 µg mL^{-1} MnCl$_2$•4H$_2$O, 10 µg mL^{-1} Na$_2$B$_4$O$_7$•10H$_2$O, and 10 µg mL^{-1} (NH$_4$)$_6$Mo$_7$O$_{24}$•4H$_2$O). Spore stocks were stored in 20% glycerol at -20°C. *Escherichia coli* DH5α (Invitrogen) was used for plasmid propagation and *E. coli* ET12567 (MacNeil et al., 1992) was used for conjugations; these were grown in lysogeny broth (LB) and stored in 20% glycerol at -20°C. Electrocompetent *E. coli* was stored in 10% glycerol at -80°C. *Saccharomyces cerevisiae* VL6-48 (Kouprina et al., 1998) was grown on YPAD (1% yeast extract, 2% peptone (BD Biosciences), 2% glucose (Fisher), 20 µg mL^{-1} adenine, and 2% agar), and was stored in 20% glycerol at -80°C, or as single colonies on YPAD plates at 7°C.

4.1.3.   Transforming *E. coli*

4.1.3.1. Making Electrocompetents

E. coli were inoculated into 10 mL SOB-Mg (2% tryptone, 0.5% yeast extract, 0.58% NaCl, and 0.186% KCl) and grown overnight at 250 rpm, 28°C. A 1% inoculum of this starter culture was added to ten 250 mL conical flasks, each containing 50 mL SOB-Mg. These were incubated for 3 to 5 hours and at 250 rpm, 28°C until an OD_{600} between 0.2 and 0.4 was reached. Cells were harvested by centrifugation at 800 x *g* for 20 minutes and resuspended in a total of 250 mL 10 % glycerol. This was repeated three times, resuspending in totals of 50 mL, 50 mL, and 1 mL successively. This produced electrocompetent cells, which were separated into 50 µL aliquots in microcentrifuge tubes and flash frozen for storage at -80°C.

4.1.3.2. Transforming Electrocompetents

Electrocompetent cells were thawed and electroporation cuvettes (2 mm) were cooled on ice. 1 µL of DNA solution was added to each aliquot of cells. This was mixed gently by pipetting and transferred to the electroporation cuvette. The outside of the cuvette was dried, and it was inserted into a Gene Pulser (Bio-Rad) with the pulse generator set to 25 µFD, 2.5 kV, and 200 Ω. The pulse was delivered, and the cuvette was immediately placed on ice. 200 µL of LB was added. The cells were transferred back to a microcentrifuge tube and were incubated for an hour at 250 rpm, 37°C. The entire transformation mix was then plated on LB with the appropriate selection. If hygromycin selection was used cells were plated on DNA (2.3% Difco Nutrient Agar and 2% agar).

4.1.4.   DNA Extraction from *Streptomyces*

Genomic DNA was extracted from *Streptomyces* spp. using a salting out procedure (Kieser et al., 2000). Spores were inoculated into 50 mL TSB (3% tryptone soya broth; Oxoid) and incubated for three days at 250 rpm, 28°C. Half of this culture was used for a genomic DNA extraction. Mycelia were harvested and resuspended in 5 mL SET buffer (75 mM NaCl, 25 mM EDTA pH 8, and 20 mM Tris-HCl pH 7.5). 100 µL of lysozyme solution (50 mg mL^{-1} in water) was added, and this was incubated at 37°C for 30 to 60 minutes. 140 µL of proteinase K solution (20 mg mL^{-1} in water) and 600 µL of 10% SDS were added. This was mixed gently and incubated at 55°C for 2 hours. 2 mL of NaCl (5 M) were added, it was mixed gently, and 5 mL of chloroform was added. This was mixed by inversion at 20°C for 30 minutes. The mixture was separated by centrifugation at 4500 g at 4°C for 15 minutes. The upper supernatant was transferred to a new 50 mL tube, and 0.6 volumes isopropanol was added and gently mixed by inversion. The DNA that precipitated as long strands was spooled onto sealed glass pipette, transferred to a microcentrifuge tube, washed with 70 % ethanol, and left to dry. This was then resuspended in water and dissolved at 55°C.

4.1.5.   General Yeast Methods

4.1.5.1.   Transformations

Yeast transformations all followed a modified version of the Gietz and Woods (Gietz and Woods, 2006) lithium acetate/polyethylene glycol (LiAc/PEG) protocol. Single colonies of *S. cerevisiae* VL6-48 were inoculated into 10 mL of liquid YPAD in a 50 mL centrifuge tube and grown overnight at 250 rpm, 30°C. This starter culture was added to 40 mL liquid YPAD in a 250 mL conical flask and incubated for 4 hours at 250 rpm, 30°C. These cells were harvested by centrifuging for 5 minutes at 1,789 x *g* and washed twice in equal volumes of sterile water. Cells were resuspended in 1 mL 0.1 M LiAc, transferred to a microcentrifuge tube and pelleted at 3,000 x *g* for 15 seconds. These cells were resuspended in 400 µL 0.1 M LiAc to a final volume of 500 µL and then transferred to microcentrifuge tubes as aliquots of 50 µL for single transformations. Prior to a transformation, an aliquot was briefly centrifuged, and the supernatant was removed. The following solutions were then added to the cells, in this order: 240 µL PEG solution (50% PEG 3350), 36 µL 1 M LiAc, 50 µL salmon sperm DNA (Invitrogen; 2 mg mL^{-1}; boiled for ten minutes and cooled in ice water), and 34 µL DNA to be assembled. Cells were resuspended by pipetting and incubated at 250 rpm, 30°C for 30 minutes, which was followed by heat shock at 42°C for 30 minutes. Cells were pelleted by centrifugation at 3,000 x *g* for 30 seconds, the supernatant was removed, and cells were resuspended in 200 µL sterile water. This final volume of 220 µL was separated into 200 µL and 20 µL aliquots that were plated on selective media SD+CSM-Trp (0.17% YNB-AA-$(NH_4)_2SO_4$ (Formedium), 0.5% $(NH_4)_2SO_4$, 2% glucose, 2% agar, 20 µg mL^{-1} adenine, and 740 µg mL^{-1} CSM-Trp (Formedium)). Plates were then incubated for 3 days at 30°C.

4.1.5.2.   Colony Screening

To screen yeast colonies by PCR, colonies were picked using a pipette tip, and cells were resuspended in 50 µL 1 M sorbitol (Fisher). 2 µL of zymolyase (5 U µL^{-1}; Zymo Research) was added to each cell suspension and incubated at 30°C for 1 hour. Cell suspensions were then boiled for 10 minutes, centrifuged for 15 seconds at 1,000 x *g* and 1 µL of the supernatant was used as a template for PCR.

4.1.5.3.   Plasmid Extraction

To shuttle the plasmids from yeast into *E. coli*, colonies of yeast were grown in 10 mL of liquid SD+CSM-Trp overnight at 250 rpm, 30°C. Cells were harvested by centrifuging for 5 minutes at 1,789 x *g* and resuspended in 200 µL 1 M sorbitol plus 2 µL of zymolyase (5 U µL^{-1}). Cell suspensions were incubated at 30°C for 1 hour to produce spheroplasts. Spheroplasts were pelleted at 600 x *g* for 10 minutes, and the supernatant was aspirated. Plasmid DNA was extracted from the spheroplasts using a standard Wizard miniprep protocol (Promega). 1 µL plasmid DNA was then transformed into *E. coli*

by electroporation. Transformations were plated on LB-agar and transformants were selected for with kanamycin.

4.1.6.   PCR and Sequencing

PCRs of fragments for assemblies were carried out with Herculase II Fusion DNA Polymerase (Agilent), and analytical PCRs were carried out Go Taq G2 Flexi DNA Polymerase (Promega). These followed the manufacturers' protocols and annealing temperatures were based on predicted primer melting temperatures. All sequencing, unless otherwise noted, was carried out on plasmid miniprep templates using a Mix2Seq Kit (Eurofins Genomics).

4.1.7.   TAR Cloning

TAR cloning the bottromycin (**1**) gene cluster in yeast followed the Yamanaka et al. protocol, (Yamanaka et al., 2014) with a few modifications as described below. Capture arms specific to the bottromycin (**1**) gene cluster were amplified from *S. scabies* genomic DNA using the primer pairs: RSPbtmarm and RASPbtmarm, and LSPbtmarm and LASPbtmarm. These primers also introduced 30 bp regions of homology with pCAP01, 30 bp regions of homology with the corresponding capture arm, and an AvrII recognition site between the two arms. pCAP01 was linearised using XhoI and was purified alongside the capture arms using a Wizard PCR clean-up kit (Promega). The capture arms were assembled into the linearized pCAP01 by transformation into yeast as described above. This produced the bottromycin-specific capture vector pCAPbtmarm, which was linearised using AvrII prior to gene cluster capture. *S. scabies* genomic DNA was extracted using the salting out procedure (Kieser et al., 2000) and digested with PciI and BsrDI to cut the bottromycin (**1**) gene cluster out of the gDNA. Both the digested gDNA and the linearized pCAPbtmarm were precipitated using a sodium acetate (NaOAc) method: 0.1 volume of 3 M NaOAc was added to the DNA digestion mixes, followed by 4 volumes of ethanol, and was then incubated at -20°C for four hours. The DNA was pelleted out of the solution by centrifugation at 20,238 x g for 30 minutes, before aspirating the supernatant. DNA pellets were washed twice in 500 µL ice cold ethanol and air dried. These were then resuspended to final concentrations of: 1.25 µg µL^{-1} for the gDNA and 1 µg µL^{-1} for the plasmid. To capture the bottromycin (**1**) gene cluster yeast was transformed using the LiAc/PEG mediated transformation procedure with 3 µg of the digested gDNA and 1 µg of the linearized capture vector. This transformation was replicated in parallel four times. Fifty colonies from each of the four transformations were patched onto master plates. These four pools were screened by multiplex PCR using the primer pairs: probe_CD_fw and probe_CD_rv, and qRT-Kfw and qRTKrv. These primers amplify a 500 bp fragment across the gap between *btmC* and *btmD*, and a 100 bp fragment within *btmK*, respectively. Three of the four pools contained positive colonies, and one pool was picked to

screen each colony individually by PCR, following the same protocol as before. Two out of these fifty colonies were positive, hence confirming the presence of the plasmid containing the captured bottromycin (**1**) gene cluster, pCAPbtm.

4.1.8.   Promoters

All promoter sequences are listed in (Table 10). PSF14 and Paac3 were synthesized as single-stranded oligonucleotides. PhrdB was amplified from *S. coelicolor* gDNA using the primers AvrIISPPHrdB and ASPPHrdB, PkasO* was amplified from pSS76 (provided by Susan Schlimpert, John Innes Centre, unpublished) using the primers AvrIISPPKasO* and ASPPKasO*, and PermE* was amplified from pIJ10257 (Hong et al., 2005) using the primers AvrIISPPErmE* and ASPPErmE.

4.1.9.   Assembly of pCAP-Based Plasmids

4.1.9.1.   General Assembly and Screening

pCAP-based plasmid assemblies were all carried out using LiAc/PEG mediated transformation into yeast as described previously. The efficiency and flexibility of the system meant no strict adherence to concentration or molar ratios of DNA was necessary. As a guide: 40 ng of linearized plasmid, 80 ng of each PCR product, and 500 pmol of each oligonucleotide in an assembly was sufficient. PCR products and linearized plasmids were column purified. Unless otherwise noted, the PCR products used in assemblies were produced with approximately 60 bp regions of overlap with other assembly parts. The oligonucleotides used in assemblies overlapped by between 30 and 60 bp, and if they were assembled to another oligonucleotide they were also complementary. Plasmid screening was accomplished using restriction analysis, PCR, or sequencing, in a case dependent manner. All primers and oligonucleotides are listed in Table 11.

4.1.9.2.   *gusA* Plasmids

gusA was amplified from pMS82-gus1 (Feeney et al., 2017) using SPpCAP01gusv2 and ASPpCAP01gus. These primers added 30 bp extensions with homology to pCAP01 on either side of the KpnI recognition site. The PCR product was assembled into KpnI-digested pCAP01 to produce pCAP01gus. The promoters for the GUS assay were amplified using the primer pairs SPpCAPSF14::gus and ASPpCAPSF14::gus, SPpCAPHrdB::gus and ASPpCAPHrdB::gus, SPpCAPKasO::gus and ASPpCAPKasO::gus, SPpCAPaac3::gus and ASPpCAPaac3::gus, SPpCAPErmE::gus and ASPpCAPErmE::gus, and SPpCAPbtmC::gus and ASPpCAPbtmC::gus, from the templates PSF14, PhrdB, PkasO*, Paac3, PermE*, and pCAPbtm respectively. These primers added 30 bp extensions homologous to pCAP01gus on either side of the SpeI recognition site. These were all assembled into SpeI-digested pCAP01gus to form pCAP01PSF14gus, pCAP01PhrdBgus, pCAP01PkasO*gus,

pCAP01Paac3gus, pCAP01PermE*gus, and pCAP01PbtmCgus. These constructs were confirmed by sequencing using pCAP01insertFv2 and pCAP01insertR.

4.1.9.3.   pCAPbtm*

Two overlapping regions between the NsiI and SpeI recognition sites in pCAPbtm were obtained using the primer pairs: SPbtmA*v2 and ASPtbtmA*v1, and SPbtmB*v2 and ASPbtmB*v2. These overlapped with each other by 30 bp in the *btmA* promoter region and included the necessary mutations for introducing a BsrGI recognition site. These PCR products also included 30 bp on either side of the NsiI and SpeI recognition sites to facilitate homologous recombination. pCAPbtm was digested with NsiI and SpeI and was assembled with the PCR products. This was confirmed by BsrGI and HindIII restriction analysis (Figure 33).

Table 8. Assemblies, and their constituent parts.

Construct	Backbone	Enzymes used	Parts
pCAPbtm1	pCAPbtm*	BsrGI and NsiI	A,B,C,D
pCAPbtm2	pCAPbtm*	BsrGI and NsiI	E,F,G,H,I,J,K
pCAPbtm3	pCAPbtm2	BlpI	L,M,N,O,P
pCAPbtm4	pCAPbtm*	BsrGI and XhoI	E,F,G,H,I,J,Q,R,S
pCAPbtm5	pCAPbtm*	BsrGI and NsiI	E,F,G,H,I,J,T
pCAPbtm6	pCAPbtm*	BsrGI and NsiI	E,F,G,H,I,U,V
pCAPbtm7	pCAPbtm*	BsrGI and NsiI	A,B,C,W,X,Y
pCAPbtm8	pCAPbtm*	BsrGI and NsiI	E,F,G,H,I,J,Z,AA
pCAPtsa1	pCAPtsa	AflII and SrfI	AB,AD,AH,AJ
pCAPtsa2	pCAPtsa	AflII and SrfI	AB,AE,AI,AK
pCAPtsa3	pCAPtsa	AflII and SrfI	AC,AF,AH,AJ
pCAPtsa4	pCAPtsa	AflII and SrfI	AC,AG,AI,AK
pCAPtsa5	pCAPtsa	AflII and SrfI	AL,AM,AN,AO,AP
pCAPtsaTsaCoreTaa	pCAPtsa	AflII and SrfI	AQ,AR,BA
pCAPtsaS1T	pCAPtsa	AflII and SrfI	AR,AS,BA
pCAPtsaSAV	pCAPtsa	AflII and SrfI	AR,AT,BA
pCAPtsaM3I	pCAPtsa	AflII and SrfI	AR,AU,BA
pCAPtsaA5F	pCAPtsa	AflII and SrfI	AR,AV,BA
pCAPtsaT8S	pCAPtsa	AflII and SrfI	AR,AW,BA
pCAPtsaY11V	pCAPtsa	AflII and SrfI	AR,AX,BA
pCAPtsaH12A	pCAPtsa	AflII and SrfI	AR,AY,BA
pCAPtsaH12W	pCAPtsa	AflII and SrfI	AR,AZ,BA

4.1.9.4.   pCAPbtm1-8, pCAPtsa1-5, and pCAPtsa Precursor Peptide Modifications

The parts used in each assembly are shown in Table 8, and Table 9 has a description of each part. A schematic of each assembly and the assessment of efficiency for pCAPbtm1-8 is also shown in the main text (Figure 33; Figure 34; Figure 35; Figure 46; Figure 48; Figure 49; Figure 50; Figure 51; Figure 84; Figure 85; Figure 87). The assembly of pCAPtsa plasmids were assessed by sequencing using the primers SPPTvaAseq, ASPPTvaBseq, and ASPSrfITvaBseq.

Table 9. The parts used in cluster assemblies.

Part	Primer pairs/Oligonucleotides	Size of PCR products (base pairs) or oligonucleotides (bases)	Contains	Left linker (towards *btmA* or *tsaA*)	Right linker (towards *btmM* or *tsaL*)
A	SPSF14btmAv2 ASPSF14btmAv2	215 bp	PSF14	*btmB*	*btmA*
B	SPbtmbB* ASPbtmB*	1072 bp	*btmB*	No linker	No linker
C	SPKasO*btmBv2 ASPKasO*btmBv2	375 bp	PKasO*	PHrdB	*btmB*
D	SPHrdBbtmCv2 ASPHrdBbtmCv2	548 bp	PHrdB	PKasO*	*btmC*
E	pCAP01btmAASC1	90 b	Part of Paac3	*btmA*	Middle of Paac3
F	pCAP01btmAASC2	60 b	Part of Paac3	Middle of Paac3	Middle of Paac3
G	pCAP01btmAASC3	60 b	Part of Paac3	Middle of Paac3	Middle of Paac3
H	pCAP01btmAASC4	60 b	Part of Paac3	Middle of Paac3	Paac3
I	pCAP01btmAASC5	60 b	Part of PSF14	Paac3	Middle of PSF14
J	pCAP01btmAASC6	60 b	Part of PSF14	Middle of PSF14	Middle of PSF14
K	pCAP01btmAASC7	105 b	Part of PSF14	Middle of PSF14	*btmC*
L	SPbtmC*DPbtmD1 ASPbtmC*DPbtmD1	372 bp	Part of *btmC*	Middle of *btmC*	Middle of *btmC*
M	OligoDPbtmD1	120 b	Part of mutated PbtmD	*btmC*	Middle of PbtmD
N	OligoDPbtmD2	119 b	Part of mutated PbtmD	Middle of PbtmD	Middle of PbtmD
O	OligoDPbtmD3	120 b	Part of mutated PbtmD	Middle of PbtmD	*btmC*
P	SPbtmC*DPbtmD2 ASPbtmC*DPbtmD2	479 bp	Part of *btmC* and *btmD*	Middle of *btmC*	Middle of *btmD*
Q	SPSF14btmDAASDC ASPCbtmDAASDC	273 bp	*btmD*	PSF14	*btmC*
R	SPDbtmCAASDC ASPEbtmCAASDC	2130 bp	*btmC*	*btmD*	*btmE*
S	SPCbtmEAASDC ASPFbtmFAASDC	2292 bp	*btmE* and most of *btmF*	*btmC*	End of *btmF*
T	pCAP01btmAASD7	120 b	Part of PSF14	Middle of PSF14	*btmD*
U	pCAP01btmAASutrC6	55 b	Part of PSF14	Middle of PSF14	Middle of PSF14
V	pCAP01btmAASutrC7	98 b	Part of PSF14 and *btmC* 5'UTR	Middle of PSF14	*btmC*
W	SPHrdBbtmCv2 ASPHrdBrs	536 bp	PHrdB	PKasO*	E* riboswitch
X	rsbtmC1	120 b	Part of E* riboswitch	PHrdB	Middle of E* riboswitch
Y	rsbtmC2	120 b	Part of E* riboswitch	Middle of E* riboswitch	*btmC*
Z	pCAP01btmAASrsC7	69 b	Part of PSF14	Middle of PSF14	E* riboswitch

AA	pCAP01btmAASrsC8	108 b	E* riboswitch	PSF14	*btmC*
AB	SPTvaPSF14 ASPTvaAPSF14	165 bp	PSF14	Plasmid backbone	*tsaA*
AC	SPTvaPaac3 ASPTvaAPaac3	176 bp	Paac3	Plasmid backbone	*tsaA*
AD	SPPSF14S4TvaA ASPPSF14S4TvaA	566 bp	*tsaA*	PSF14	PSF14
AE	SPPSF14S4TvaA ASPPaac3S4TvaA	566 bp	*tsaA*	PSF14	Paac3
AF	SPPaac3S4TvaA ASPPSF14S4TvaA	566 bp	*tsaA*	Paac3	PSF14
AG	SPPaac3S4TvaA ASPPaac3S4TvaA	566 bp	*tsaA*	Paac3	Paac3
AH	SPTvaAPSF14 ASPTvaBPSF14	147 bp	PSF14	*tsaA*	*tsaC*
AI	SPTvaAPaac3 ASPTvaBPaac3	158 bp	Paac3	*tsaA*	*tsaC*
AJ	SPPSF14S4TvaB ASPS4TvaTvaB	1156 bp	*tsaC*	PSF14	*tsaD*
AK	SPPaac3S4TvaB ASPS4TvaTvaB	1156 bp	*tsaC*	Paac3	*tsaD*
AL	TvaPSF14rsOligo1	120 b	PSF14	Plasmid backbone	E* riboswitch
AM	PSF14rsOligo2	120 b	E* riboswitch	PSF14	No linker
AN	SPPSF14rsTvaAv2 ASPPaac3terTvaA	290 bp	*tsaA*	E* riboswitch	Paac3
AO	Paac3	103 b	Paac3	No linker	No linker
AP	SPPaac3S4TvaB ASPTvaCTvaB	1167 bp	*tsaC*	Paac3	*tsaD*
AQ	SPTvaA ASPTvaA	473 bp	*tsaA*	Plasmid backbone	No linker
AR	TsaCoreAlba	119 b	Mutated *tsaA* core	*tsaA*	*tsaC*
AS	S1T	119 b	Mutated *tsaA* core	*tsaA*	*tsaC*
AT	SAV	120 b	Mutated *tsaA* core	*tsaA*	*tsaC*
AU	M3I	119 b	Mutated *tsaA* core	*tsaA*	*tsaC*
AV	A5F	119 b	Mutated *tsaA* core	*tsaA*	*tsaC*
AW	T8S	119 b	Mutated *tsaA* core	*tsaA*	*tsaC*
AX	Y11V	119 b	Mutated *tsaA* core	*tsaA*	*tsaC*
AY	H12A	119 b	Mutated *tsaA* core	*tsaA*	*tsaC*
AZ	H12W	119 b	Mutated *tsaA* core	*tsaA*	*tsaC*
BA	SPTvaB ASPTvaB	1081 bp	*tsaC*	No linker	*tsaD*

4.1.10. pCAPtsa Gene Deletions

4.1.10.1. Gene Disruption

To obtain gene deletions in pCAPtsa, PCR targeting was used following the published protocol (Gust et al., 2003). The disruption cassette (apramycin resistance cassette bracketed by FLP-recombinase recognition sites) was amplified from pIJ773-oriT, a version of pIJ773 modified to have the oriT removed, with the following primers: the *tsaA* cassette used SPPCRtTvaA and ASPPCRtTvaA; the *tsaC* cassette used SPPCRtTvaB and ASPPCRtTvaBv3; the *tsaD* cassette used SPPCRtTvaC and ASPPCRtTvaCv4; the *tsaE* cassette used SPPCRtTvaD and ASPPCRtTvaDv4; the *tsaF* cassette used

SPPCRt11840 and ASPPCRt11840v3; the *tsaG* cassette used SPPCRtTvaE and ASPPCRtTvaE; the *tsaH* cassette used SPPCRtTvaF and ASPPCRtTvaFv2; the *tsaI* cassette used SPPCRtTvaG and ASPPCRtTvaGv2; the *tsaJ* cassette used SPPCRt11820 and ASPPCRt11820; the *tsaK* cassette used SPPCRtTvaH and ASPPCRtTvaHv2; the *tsaMT* cassette used SPPCRt11775 and ASPPCRt11775; the *tsaL* cassette used SPPCRtTvaI and ASPPCRtTvaI; the *tsaMO* cassette used SPPCRtTvaJ and ASPPCRtTvaJ; the *tsa+1* cassette used SPPCRtTvaK and ASPPCRtTvaK; the *tsa+2* cassette used SPPCRtTvaL and ASPPCRtTvaL; the *tsa+3* cassette used SPPCRtTvaRgenes and ASPPCRtTvaRgenes; and the *tsa-1, -2,* and *-3* cassette used SPPCRtTvaLgenes and ASPPCRtTvaLgenes.

pCAPtsa was introduced to *E. coli* BW25113-pIJ790 by electroporation, and transformants were selected on LB-agar containing chloramphenicol and kanamycin at 30°C. A single colony was used to make electrocompetent cells using the standard protocol, with one modification; the second round of growth was performed with 10 mM arabinose added to the medium. These cells were then transformed with the gene specific disruption cassettes by electroporation. Transformants were selected for on LB-agar containing kanamycin and apramycin at 37°C. Plasmids were extracted from *E. coli* colonies, and gene disruptions were first confirmed by PCR using the following primer pairs: Δ*tsaA* was assessed with SPS4TvaA and ASPS4TvaA; Δ*tsaC* was assessed with SPS4TvaB and ASPS4TvaB; Δ*tsaD* was assessed with SPS4TvaC and ASPS4TvaCv2; Δ*tsaE* was assessed with SPS4TvaD and ASPS4TvaDv2; Δ*tsaF* was assessed with SPS4Tva40v3 and ASPS4Tva40v3; Δ*tsaG* was assessed with SPS4TvaE and ASPS4TvaE; Δ*tsaH* was assessed with SPS4TvaF and ASPS4TvaF; Δ*tsaI* was assessed with SPS4TvaG and ASPS4TvaG; Δ*tsaJ* was assessed with SPS411820 and ASPS411820; Δ*tsaK* was assessed with SPS4TvaHv3 and ASPS4TvaHv3; Δ*tsaMT* was assessed with SPS411775 and ASPS411775; Δ*tsaL* was assessed with SPS4TvaI and ASPS4TvaI; Δ*tsaMO* was assessed with SPS4TvaJ and ASPS4TvaJ; Δ*tsa+1* was assessed with SPS4TvaK and ASPS4TvaK; Δ*tsa+2* was assessed with SPS4TvaLv3 and ASPS4TvaLv3; Δ*tsa+3* was assessed with SPS4Rgenes and ASPS4Rgenes; and Δ*tsa-1,-2,* and *-3* was assessed with SPLgPCRtcheckv4 and ASPLgPCRtcheckv4. Additionally, the primers SPaac(3)IVseq and ASPaac(3)IVseq were used to sequence outwards from the inserted selectable marker.

4.1.10.2.   Removal of Selectable Marker

Gene-disrupted pCAPtsa plasmids were transformed in to *E. coli* DH5α-BT340 by electroporation. Transformants were selected for on LB-agar containing apramycin, chloramphenicol, and kanamycin at 30°C. Individual colonies were picked and spread to single colonies on LB-agar without antibiotics and grown overnight at 42°C. Isolated colonies were spread as patches on LB-agar containing kanamycin and then LB-agar containing kanamycin and apramycin. Patches that only grew on the LB-agar plates containing kanamycin were assumed to have lost the disruption cassette, leaving behind an 81 bp scar. This was additionally confirmed by sequencing using the following primers: Δ*tsaA* was

assessed with SPS4TvaA and ASPS4TvaA; Δ*tsaC* was assessed with SPS4TvaB and ASPS4TvaB; Δ*tsaD* was assessed with SPS4TvaC and ASPS4TvaCv2; Δ*tsaE* was assessed with SPS4TvaD and ASPS4TvaDv2; Δ*tsaF* was assessed with SPS4Tva40v3 and ASPS4Tva40v3; Δ*tsaG* was assessed with SPseqS4tvaE and ASPS4TvaE; Δ*tsaH* was assessed with SPseqS4tvaF and ASPS4TvaF; Δ*tsaI* was assessed with SPseqS4tvaG and ASPS4TvaG; Δ*tsaJ* was assessed with SPS411820 and ASPS411820; Δ*tsaK* was assessed with SPseqS4tvaH and ASPS4TvaHv3; Δ*tsaMT* was assessed with SPseqS4tva75 and ASPS411775; Δ*tsaL* was assessed with SPseqS4tvaI and ASPS4TvaI; Δ*tsaMO* was assessed with SPS4TvaJ and ASPS4TvaJ; Δ*tsa+1* was assessed with SPS4TvaK and ASPS4TvaK; Δ*tsa+2* was assessed with SPseqS4tvaL and ASPS4TvaLv3; Δ*tsa+3* was assessed with SPS4Rgenes and ASPS4Rgenes; and Δ*tsa-1,-2,* and *-3* was assessed with SPLgPCRtcheckv4 and ASPLgPCRtcheckv4. Additionally, the primers SPaac(3)IVseq and ASPaac(3)IVseq were used to sequence outwards from the inserted selectable marker.

4.1.11.   Construction of Single Gene Expression Plasmids

Constructs for *in trans* expression of *btmC*, thioalbamide (**18**) genes, and thiostreptamide S4 (**17**) complementation genes were assembled using standard digestion and ligation of PCR fragments in to pIJ10257 (Hong et al., 2005). One version of *btmC* was amplified from *S. scabies*, using the primers SPNdeIbtmC and ASPPacIbtmC, and three versions of *btmC* (with different start sites) were amplified from *S. bottropensis* using the primers SPpIJbotPbtmC and ASPpIJbotPbtmC; SPpIJbotmetbtmC and ASPpIJbotPbtmC; and SPpIJbotvalbtmC and ASPpIJbotPbtmC. TaaRed and TaaCYP were amplified from *A. alba* DNA using the primer pairs: SPNdeITaaRedu and ASPNdeITaaRedu, and SPNdeITaaP450 and ASPPacITaaP450 respectively. Thiostreptamide S4 (**17**) genes *tsaA, tsaC-tsaJ* were amplified from pCAPtsa using the following primers: for *tsaA* SPNdeIS4tvaA, SPNdeIS4TvaAv2, SPNdeIS4TvaAv3 and ASPPacIS4TvaA were used; for *tsaC* SPNdeIS4TvaB, SPNdeIS4TvaBv2, and ASPPacIS4TvaB were used; for *tsaD* SPNdeIS4TvaC, SPNdeIS4TvaCv2, SPNdeIS4TvaCv3, and ASPPacIS4TvaCv2 were used; for *tsaE* SPNdeIS4TvaD and ASPPacIS4TvaD were used; for *tsaF* SPNdeIS411840 and ASPPacIS411840 were used; for *tsaG* SPNdeIS4TvaE, SPNdeIS4TvaEv2, SPNdeIS4TvaEv3, SPNdeIS4TvaEv4, and ASPPacIS4TvaE were used; for *tsaH* SPNdeIS4TvaF and ASPPacIS4TvaF were used; for *tsaI* SPNdeIS4TvaG and ASPPacIS4TvaG were used; for *tsaJ* SPNdeIS411820 and ASPPacIS411820 were used; and for *tsaMT* SPNdeIS411775 and ASPPacIS411775 were used.

PCR fragments and pIJ10257 were digested with NdeI and PacI. The PCR fragments were ligated in to pIJ10257 using T4 ligase (Invitrogen). Ligation reactions were transformed into *E. coli* DH5α by electroporation and transformants were selected for on DN-agar containing hygromycin. Colonies were screened for correct ligations by PCR using primers SPpIJ10257ins and ASPpIJ10257ins.

Plasmids extracted from these colonies were additionally checked by sequencing using the same primers.

4.1.12.   *Streptomyces* Conjugations

E. coli ET12567 carrying the helper plasmid pR9604 was used for intergenic conjugations into *Streptomyces* strains using a protocol described in Kieser et al. (2000) with a few modifications. Plasmids were transformed into *E. coli* ET12567-pR9604 by electroporation. Transformants were selected on LB-agar containing carbenicillin, chloramphenicol, and kanamycin or hygromycin for pCAP-based or pIJ10257-based plasmids respectively. A single colony was used to inoculate 10 mL liquid LB containing the same antibiotics, and grown overnight at 37°C, 250 rpm. A 1% inoculum was used to start a fresh growth in 10 mL of LB containing the same antibiotics. This was grown in the same conditions to an OD_{600} of 0.4. Cells were washed twice in 2 mL LB and resuspended in 1 mL LB. 5 µL of *Streptomyces* spores were mixed with 0.5 mL 2xYT (1.6% tryptone (BD Biosciences), 1% yeast extract, and 0.5% NaCl; adjusted to pH 7.4 before autoclaving) and heat-shocked at 50°C for 10 minutes. 0.5 mL of washed *E. coli* ET12567-pR9604 were added to the heat-shocked spores. This mixture was pelleted by centrifugation at 3000 x *g* for 5 minutes and resuspended in 0.2 mL water. Cells were plated on SFM with $MgCl_2$ (Fisher) added to a final concentration of 10 mM and incubated at 30°C overnight. These plates were overlaid with nalidixic acid and kanamycin or hygromycin to select for *Streptomyces* pCAP-based or pIJ10257-based plasmid exconjugants.

4.1.13.   Production Cultures

Unless otherwise noted, the same conditions were used for all bottromycin (**1**) production cultures. All cultures were conducted in triplicate in 50 mL centrifuge tubes with their lids replaced with foam bungs and were incubated at 28°C at 250 rpm. For seed cultures, 10 µL spores were added to 10 mL GYM (0.4% glucose, 0.4% yeast extract, and 1% malt extract; adjusted to pH 7.2 before autoclaving) and grown for 3 days. For production cultures, a 5% inoculum from the seed culture was added to 10 mL of btmPM (1% glucose, 1.5% starch (BD Biosciences), 0.5% yeast extract, 1% soy flour, 0.5% NaCl, and 0.3% $CaCO_3$ (AnalaR); adjusted to pH 7.0 before autoclaving; $CoCl_2$ was added after autoclaving to a final concentration of 20 µg mL^{-1}), and grown for 5 days at 28°C. For thiostreptamide S4 (**17**) production cultures, 10 µL of spores were spread directly on btmPM-agar and this was grown for 5 days at 28°C.

4.1.14.   GUS Assay

pCAP01gus,   pCAP01PSF14gus,   pCAP01PhrdBgus,   pCAP01PkasO*gus,   pCAP01Paac3gus, pCAP01PermE*gus, and pCAP01PbtmCgus were each conjugated into *S. coelicolor* M1146 and *S.*

scabies. Spores of each mutant strain were added to 10 mL GYM and incubated overnight at 28°C at 250 rpm. A 5% inoculum of each seed culture was added in triplicate to 12 mL btmPM and incubated at 28°C at 250 rpm. 1 mL samples were taken at 23, 46, 70.5, and 95.5 hours of growth. These mycelial samples were washed twice in equal volumes of water, and frozen at -80°C.

To assess GUS activity, defrosted mycelia were first resuspended in 1 mL lysis buffer (60 mM $Na_2HPO_4\bullet7H_2O$, 40 mM $NaH_2PO_4\bullet H_2O$ (AnalaR), 10 mM KCl, 1 mM $MgSO_4\bullet7H_2O$, 50 mM β-mercaptoethanol, 0.1% Triton X-100 (BDH Chemicals), and 1 mg mL^{-1} lysozyme) and incubated for 30 minutes at 37°C. Cellular debris was removed by centrifugation at 13,000 x *g* for 4 minutes. Protein content in 2 μL lysate was quantified using a Bradford assay. 0.1 mL lysate was added to 0.9 mL Z buffer (60 mM $Na_2HPO_4\bullet7H_2O$, 40 mM $NaH_2PO_4\bullet H_2O$, 10 mM KCl, 1 mM $MgSO_4\bullet7H_2O$, and 50 mM β-mercaptoethanol); blank controls contained 1 mL Z buffer. Reactions were conducted in a heat block at 30°C and were started with the addition of 0.2 mL 4-Nitrophenyl-β-D-glucopyranosiduronic acid (PNPG) solution (4 mg mL^{-1} in Z buffer). Once a pale-yellow color had developed the reactions were quenched by the addition of 1 M Na_2CO_3 (0.5 mL) and the reaction time was recorded. Reactions that had no GUS activity were quenched after two to three hours. 0.2 mL of each quenched reaction mixture was added to a flat-bottomed 96-well plate and the optical density at 420 and 550 nm was measured three times, with the average being taken. Activity in Miller units per mg of protein were calculated using the equation:

$$A = \frac{1000.(OD_{420} - 1.75.OD_{550})}{t.v.p}$$

Where A = GUS activity (Miller units per mg of protein), t = reaction time (minutes), v = lysate volume (mL), and p = protein concentration (mg mL^{-1}).

4.1.15. Metabolite Analysis

4.1.15.1. Standard LC-MS^2 Analysis

Samples were taken from the supernatant of production cultures and mixed with one volume methanol. LC-MS^2 data were acquired using a Shimadzu Nexera X2 UHPLC connected to a Shimadzu ion-trap time-of-flight (IT-TOF) mass spectrometer and analysed using LabSolutions software (Shimadzu). 10 μL samples were injected onto a Phenomenex Kinetex 2.6 μm XB-C18 column (50 mm x 2.1 mm, 100 Å). The samples were eluted over 5 minutes using a 5 to 95% gradient of acetonitrile in water + 0.1% formic acid. After the first minute of each run, positive mode MS data were collected between *m/z* 200 and 2000, with an ion accumulation window of 20 ms and automatic sensitivity control of 70% of the base peak. The curved desolvation line (CDL) temperature was 250°C and the heat block temperature was 300°C. MS^2 data were collected between *m/z* 50 and 2000 in a data-dependent manner for parent ions between *m/z* 200 and 2000, using collision-

induced dissociation energy of 50% and a precursor ion width of 3 Da. The instrument was calibrated using sodium trifluoroacetate cluster ions prior to every run.

4.1.15.2.   High Resolution LC-MS^2 Analysis

High resolution LC-MS^2 data were acquired by Gerhard Saalbach (JIC, UK). A Waters Synapt G2-Si mass spectrometer was operated in positive mode with a scan time of 0.5 s in the mass range of m/z 50 to 1200. 7µL samples were injected onto a Luna Omega 1.6 µm Polar C18 column (50 mm x 2.1 mm, 100 Å, Phenomenex) and eluted with a linear gradient of 1 to 50 % acetonitrile in water + 0.1% formic acid over 20 minutes. Synapt G2-Si MS data were collected with the following parameters: capillary voltage = 3kV; cone voltage = 40 V; source temperature = 120°C; desolvation temperature = 350°C. Leu-enkephalin peptide was used to generate a dual lock-mass calibration with m/z = 278.1135 and m/z = 556.2766 measured every 30 s during the run.

4.1.15.3.   Untargeted Metabolomic Analysis

LC-MS data acquired on the Shimadzu IT-TOF was assessed using the statistical package Profiling Solution 1.1 (Shimadzu; Table 5 and Table 12). The following Profiling Solution parameters were used: ion m/z tolerance = 0.1 Da, ion intensity threshold = 50,000, LabSolutions compatible ion m/z tolerances = ON, De-Isotope matrix = ON. Metabolites detected in the negative controls were filtered out during post-processing in Microsoft Office 365 ProPlus Excel.

4.1.15.4.   Metabolite Networking

Global Natural Products Social Molecular Networking (GNPS; http://gnps.ucsd.edu) was used to construct the metabolite network from data acquired on the Shimadzu IT-TOF (Wang et al., 2016). LC-MS data from pCAPbtm1, pCAPbtm2 and *S. scabies* production experiments was used for GNPS analysis using the following GNPS settings: Parent Mass Tolerance = 1 Da, Min Pairs Cos = 0.6, Min Matched Peaks = 3, Network TopK = 15, MSCluster = ON, Minimum Peak Intensity = 25, Filter Precursor Window = OFF, Filter Library = OFF, and Filter peaks in 50 Da Window = OFF. The resulting networks were then formatted in Cytoscape 2.8.0 and Microsoft Office 365 ProPlus PowerPoint. Duplicated nodes (same mass, retention time and fragmentation patterns) were removed, and molecules identified by Profiling Solution analysis were manually added as unlinked nodes.

4.1.15.5.   Identification and Quantification

Following identification by metabolomics, MS^2 data for previously unreported metabolites were compared to MS^2 data of known metabolites (Crone et al., 2016; Frattaruolo et al., 2017) to aid with metabolite structural elucidation. For each quantification experiment, all relevant metabolites (stated with each figure) were used to estimate pathway productivity by integration of LC–MS peak areas, obtained on the Shimadzu IT-TOF, using Browser software (Shimadzu). All data reported from

riboswitch experiments are the average of duplicate cultures, while all other MS quantification data are the average of triplicate cultures. Error bars on all graphs represent the standard error.

4.1.15.6.   Purification of Compound 65

65 was purified from *S. coelicolor* expressing the Δ*tsaE* gene cluster. Both standard LC-MS and high-resolution LC-MS analysis were used to assess the progress of purification and track **65**. 100 μL of spores were spread on 1.6 L of btmPM-agar and were incubated at 28°C for five days. This was extracted with 3.2 L of ethyl acetate, which resulted in an organic fraction containing a small proportion of **65** and the solid leftover containing the rest of **65**. The solid leftover was subsequently extracted with 3.2 L of methanol and reduced on the rotary evaporator at 30°C to 800 mL. This 800 mL extract contained undetermined combination of water and methanol that was immiscible in ethyl acetate. A liquid-liquid extract with 3.2 L of ethyl acetate was performed and the remainder of **65** partitioned into the organic phase. This organic phase was combined with the original ethyl acetate extract and was evaporated on a rotary evaporator at 30°C to produce a brown solid. This was resuspended in $CHCl_3$ and dried on to silica gel. This was packed in to a silica column and separated into 12 fractions using $CHCl_3$ and an increasing concentration of CH_3OH. An initial 400 mL elution of 100% $CHCl_3$ was followed by 10 elutions of 200 mL, ranging from 5% to 25% CH_3OH in $CHCl_3$. The final elution was with 100% CH_3OH. Fractions eluted with 10% CH_3OH and upwards contained **65**, so were combined and evaporated to a brown solid using a rotary evaporator at 30°C. This was resuspended in CH_3OH, dried onto a 1 g DSC-18 column (Discovery), and separated into 17 fractions using water and acetonitrile. 14 elutions of 10 mL, ranging from 0% to 60% acetonitrile were collected as the first 14 fractions. The final 3 fractions were collected using 10 mL of 100% acetonitrile. All fractions with acetonitrile concentrations equal to or less than 20% contained **65**, and so were combined and evaporated using a rotary evaporator at 30°C to give 493 mg of brown solid. This was resuspended in CH_3OH and separated by HPLC Gemini-NX 2.6 μ C18 column (150 mm x 21.2 mm, 100 Å) using a gradient from 10% to 60% acetonitrile over X minutes. Compound collection was guided by absorbance at 272 nm, found to be characteristic of thioamide bonds. **65** eluted at 21.5 minutes and was collected and evaporated using a Genevac EZ-2 Elite. This provided 0.7 mg of pure **65** as a white powder.

4.1.15.7.   Purification of Compounds 31 and 50

Thiostreptamide S4 (**17**) and its oxidised version (**75**) were purified from M1146-pCAPtsa. Both standard LC-MS and high-resolution LC-MS analysis were used to assess the progress of purification and track **17**. 10 μL of spores were inoculated in to three 250 mL conical flasks containing 50 mL GYM. These seed cultures were incubated at 28°C at 250 rpm for three days. 5% inoculums from the seed culture were added to six 2 L conical flasks containing 400 mL btmPM each. These production cultures were incubated at 28°C at 250 rpm for five days. The pellet was separated from the

supernatant by centrifugation at 4000 x *g* for 20 minutes. **17** was associated with the pellet so the pellet was extracted with 2.4 liters of ethyl acetate. **17** was still associated with the pellet so the pellet was subsequently extracted with 2.4 liters of methanol, which liberated **17** from the pellet. The methanol extract was reduced on a rotary evaporator at 30°C to 400 mL. This 400 mL extract contained undetermined combination of water and methanol that was immiscible in ethyl acetate. The extract was subjected to a liquid-liquid extract with 2.4 L ethyl acetate. **17** partitioned into the organic phase, which was evaporated on the rotary evaporator at 30°C to 0.91 g of a brown powder. This powder was resuspended in 2:1 CH$_3$OH:CHCl$_3$, a white pellet was removed by brief centrifugation, and the supernatant was separated on a Sephadex column into 13 ~7 mL fractions. The positive fractions were combined and evaporated on the rotary evaporator to yield 0.58 g of brown powder. At this point the significant conversion of **17** to **75** was noted (Figure 91).

4.1.15.8.   NMR

Raw NMR data were acquired by Rodney Lacret (JIC, UK). NMR spectra were recorded on a Bruker Avance 600 MHz NMR spectrometer. Chemical shifts were reported in ppm using the signals of the residual solvents as internal references (δ_H 3.31 and δ_C 49.0 for CD$_3$OD; δ_H 2.50 and δ_C 39.52 for DMSO-d_6). Data were analysed using MestReNova 6.0.2. See Figure 107, Figure 108, Figure 109, Figure 110, and Figure 111 for the raw data, and Table 6 and Figure 70 in section 3.2.1.5 for the analysis of this data.

4.1.16.   Transcript Analysis

4.1.16.1.   RNA Extraction and cDNA Production

M1146-pCAPbtm2 and M1146-pCAPbtm8 were fermented for 3 days in duplicate with varying theophylline concentrations. Mycelia were washed and stored at -80°C in 1 volume RNA*later*® (ThermoFisher Scientific). Cells were resuspended in 1 mL RLT buffer from an RNeasy kit (Qiagen) and subsequently mechanically lysed in lysing matrix B tubes (MP Biomedicals) using a FastPrep (Bio101) cell disrupter. The cells were subjected to three rounds of 30 seconds of disruption at 6 m s^{-1} with intervals of one minute on ice. Cell debris and the lysing matrix were pelleted by centrifugation at 16,000 x *g* for 10 minutes. RNA was purified from the supernatant using an RNeasy kit following the manufacturer's protocol. Contaminating gDNA was removed using on-column digestion with DNAseI (Qiagen), and after elution using the TURBO DNA-*free* Kit (Ambion, Invitrogen). 490 ng RNA was subjected to another round of DNA removal and then used for cDNA synthesis using QuantiTect Reverse Transcription Kit (Qiagen), following the manufacturer's protocol. cDNA was stored at -20°C.

4.1.16.2.   PCR Analysis of cDNA

The presence of key regions of cDNA was assessed using PCR. Primers were provided by Natalia Miguel-Vior. The presence of *hrdB* cDNA was confirmed with the primers hrdBfw and hrdBrv. The presence of *btmD* cDNA was confirmed with the primers btmDfw and btmDrv. The presence of *btmC* cDNA was confirmed with the primers btmCfw and btmCrv. The presence of the region between *btmC* and *btmD* as cDNA was confirmed with the primers btmC-Dfw and btmC-Drv. The presence of the PbtmD region as cDNA was confirmed using the primers probe_CD_fw and probe_CD_rv. The presence of the 5'UTR of SF14 in the cDNA was confirmed using the primers SPSF145'UTR and btmCp_rv.

4.1.16.3.   qPCR

qPCR reactions used the SensiFAST SYBR No-ROX Kit (Bioline) following the manufacturer's protocol and were conducted and tracked using a C1000 Touch Thermal Cycler with a CFX96 Real-Time System reaction module (Bio-Rad). Data collection and analysis was controlled using Bio-Rad CFX Manager 3.1. Each biological replicate was analyzed in duplicate, so four data points were obtained for each condition. The quantity of the *hrdB* transcript was used as an internal control and was assessed using the primer pair qhrdBsfw and qhrdBsrv. *btmC* transcript quantity was determined using the primer pair qRT-Cfw and qRT-Crv. *btmD* transcript quantity was determined using the primer pair: qRT-Dfw and qRT-Drv.

4.1.17.   Sequencing *btmC*

The sequence of *btmC* in M1146-pCAPbtm1 was assessed after production growths. gDNA was extracted from production cultures using FastDNA SPIN Kit for Soil (MP Biomedicals) and used as a template for PCR using the primer pair seqbtmC1 and seqPbtmB. This PCR product was sequenced using primers seqbtmC1-3.

To sequence the entire bottromycin gene cluster in M1146-pCAPbtm1, DNA was extracted from mycelium using the salting out procedure and subjected to whole genome sequencing. The sample was processed to generate a TruSeq PCR-free library and then sequenced using Illumina MiSeq (600-cycle, 2x300 bp) at the DNA Sequencing Facility, Department of Biochemistry, University of Cambridge (UK). The raw data output comprised 11,835,896 paired-end 301 bp Illumina MiSeq reads. The assembly was carried out by Genome assembly and annotation were carried out by Martin Trick (JIC, UK) using SPAdes v3.6.2 (Bankevich et al., 2012) with the k-mer flag set to -k 21,33,55,77,99,127. All assembly tasks were conducted using 16 CPUs on a 256 GB compute node within the Norwich Bioscience Institutes (NBI) High Performance Computing (HPC) cluster. The initial assembly was trimmed to remove any contigs with cov <2 and/or with a length less than 300 bp. Genome statistics: Total size = 8,463,237 bp; number of contigs = 69; largest contig = 937,253 bp;

N50 = 407,385 bp; L50 = 8. pCAPbtm1 was fully sequenced as part of a larger contig, which showed that the plasmid had integrated into the expected phiC31 site within SCO3798.

4.1.18. Phylogenetic Analysis

Homologues of TsaD and HFCDs were identified by a BLASTP analysis. A pseudo-random selection of 171 homologues spread over the 568 homologues of TsaD identified was generated by using systematic sampling approach whilst avoiding multiple homologues from the same species. 5 HFCDs from each of 5 classes of natural products were selected. Each dataset was aligned using MUSCLE (Edgar, 2004) set to default parameters in the EMBL-EBI online tool suite (https://www.ebi.ac.uk/services). Alignments were trimmed to remove unnecessary sequence. These alignments were used to generate a maximum likelihood phylogenetic tree with RAxML through the CIPRES science gateway (https://www.phylo.org/; Miller et al., 2010; Stamatakis, 2014). The RAxML analysis was set to use a protein GAMMA model with the BLOSUM62 protein substitution matrix, and to conduct rapid bootstrapping stopped automatically with an autoMRE majority rule criterion. The HFCD RAxML analysis additionally had the protein UbiX (P0AG03) from *E. coli* set as the outgroup (Gulmezian et al., 2007; Rangarajan et al., 2004). The phylogenetic trees were rooted, visualized, and cosmetically altered with iTOL (Letunic and Bork, 2016). The TsaD-homologue phylogenetic tree was midpoint rooted and the HFCD phylogenetic tree was outgroup rooted.

4.2. DNA Sequences

4.2.1.   Promoters

Table 10. Promoter sequences used in this study.

Promoter	Reference	Sequence (5'-3'). Transcription start site (if known), RBS, and start codon are capitalised.
PSF14	Labes et al. (1997) and Rudolph et al. (2013)	gacctacgccttgaccttgatgaggcggcgtgagctacaatcaatactCaatacgactcactataggttcctgctAAGGAGGcaacaagATG
PhrdB	Du et al. (2013)	ccgccttccgccggaacggcggggtccgggcacgccaaaccctcctgtggctgtggccggccaccgccgtcaccttcggaccccgtggagccgctcccggttccacggggtccgaaggtgtgatgagcaggctgcgccttcctcgcgcggccgcaaggtacgagttgatgaccttgtttatccgcatctgaccaattttgatcgcttacggggtgtgactcgggccacgcggattgggcgtaacgctcttgggaacaacacgatgacctaagaggtgacagccgcggagggaatacggacgccgttcacggcgctgtgcatctccccggcccgcccgcaccgtcggcccattcccaagccggtggtcggcccctgtccgccgtggacggggccggaagccgttttttcaacgttccGAGAGGttgttcATG
PkasO*	Wang et al. (2013)	gtcgactctagagctgagttggctgctgccaccgctgagcaataactagcataaccccttggggcctctaaacgggtcttgaggggtttttttgctgaaaggaggaactatatccgcgggatcctgttcacattcgaacggtctctgctttgacaacatgctgtgcggtgttgtaaagtcgtggccaggagaatacgacagcgtgcaggactgggggagttactagtatctgagttGAAGAGGtgaccatATG
Paac3	Sherwood et al. (2013)	caccaccgactatttgcaacagtgccgttgatcgtgctatgatcgactgAtgtcatcagcggtggagtgcaatgtcgtgcaatacgAAGGAGGactacacATG
PermE*	Bibb et al. (1994)	gcggtcgatcttgacggctggcgagaggtgcggggaggatctgaccgacgcggtccacacgtggcaccgcgatgctgttgtgggcacaatcgtgccggttggtaggatcgtctagaacAGGAGGccccatATG

4.2.2. Primers

Table 11. The primers and oligonucleotides used in this thesis.

Primer/Oligonucleotide	Sequence (5'-3')
seqPbtmA	cagccagggcagggcgatgg
seqPbtmB	atgtggtcaccgagcagagg
seqPbtmC	cagaataccgacctgaatcg
seqbtmC1	ggcgctcttctcggttcc
seqbtmC2	gatgtgcggcggcgtcag
seqbtmC3	ggccacgccggtggtgagc
SPrtbtmDtssv2	gcgcccatcacgagtcc
ASPrtbtmDtssv3	gtagacccaccgcttcctg
AvrIISPPHrdB	cctaggccgccttccgccggaac
ASPPHrdB	catgaacaacctctcggaacgt tga aa aac ggcttcc
AvrIISPPKasO*	cctagggtcgactctagagctg
ASPPKasO*	catatggtcacctcttcaactcag atac
AvrIISPPErmE*	cctagggcggtcgatcttgacg
ASPPErmE*	catatggggcctcctgttctagac
SPpCAP01gusv2	acaaagatcgactagta acctcg ag acttg atgc tgcg gcccgtcg aa ac
ASPpCAP01gus	aacccctatttgtttatttttctaa at actcact gcttcccgcccc tgctg
SPpCAPSF14::gus	atagctgcgccgatggtttctac aa aga tcccta gggacctac gccttg
ASPpCAPSF14::gus	gcgggtcggggtttcgacgggccgc agc atctt gttgcctcc tta gcag
SPpCAPHrdB::gus	atagctgcgccgatggtttctac aa aga tccta ggccgcct tccgccg ga
ASPpCAPHrdB::gus	gcgggtcggggtttcgacgggccgc agc atg aac aacctcc tgg aac g
SPpCAPKasO::gus	atagctgcgccgatggtttctac aa aga tcccta gggtc gactc tag agc
ASPpCAPKasO::gus	gcgggtcggggtttcgacgggccgc agc ata tggtc acctcctc aactc
SPpCAPaac3::gus	tagctgcgccgatggtttctaca aagatcct aggc accacc gact attt g
ASPpCAPaac3::gus	gcgggtcggggtttcgacgggccgc agc atgt gta gtcctccttc g
SPpCAPErmE::gus	atagctgcgccgatggtttctac aa aga tcccta gggc ggtcg atctt g
ASPpCAPErmE::gus	gcgggtcggggtttcgacgggccgc agc ata tgg ggcctcct gttc
SPpCAPbtmC::gus	atagctgcgccgatggtttctac aa aga tccta gggc acat gg acg
ASPpCAPbtmC::gus	gcgggtcggggtttcgacgggccgc agc atg aa atca acctcct gagcc
pCAP01insertR	tatagcacgtgatgaaa agg
pCAP01insertFv2	gcagggcgaagaatctcg
RSPbtmarm	cattaaattgcacctaggta aa ttg aag acctg acg gagc tgg gtctcac
RASPbtmarm	tatttgtttattttttctaaat acag gtacc acttcctcc atcg ggt ggtg
LSPbtmarm	cgccgatggtttctacaaaga tcgac tag ttggcc actctccct gtcttc
LASPbtmarm	gtcttcaatttacctaggtgca attt aat gcctcacc gtcgtct ggttc g
probe_CD_fw	ctcaccgacgaacgacccg
probe_CD_rv	tggggtcgtcgttgaggaa gtc
qRT-Kfw	ggcgtccgtccatccaggtg
qRT-Krv	gccaggccgtagaaggag agg ta
SPbtmA*v2	gggcgtggatatgtcctgcggg ta aat agc
ASPbtmA*v2	gaccagaatgtacagat gctgccgc tctac
SPbtmB*v2	gtagagcggcagcatctgtac attct ggtc
ASPbtmB*v2	ggtgcccgccggcctgttcttgatcaactc

SPSF14btmAv2	ggaggccacggtgttgcccagcagggacaggccgtgggagatcaggtagagcggcagcatcttgttgcctccttagcagg
ASPSF14btmAv2	cggaacccccgcccgggcccgcaccacccgacctcccaccgccgggacacccccggccgcccctaggggacctacgccttga
SPKasO*btmBv2	cgcgcgggcacagagcggcgcgagcagtgtctccgccgcgccggtcagggaaatcttcacatggtcacctcttcaactca
ASPKasO*btmBv2	tcttgcggaagacttaacacgataggaagcccgacacgagccccggctagcgtgcttcctcctagggtcgactctagagc
SPHrdBbtmCv2	aggaagcacgctagccgggctcgtgtcgggcttcctatcgtgttaagtcttccgcaagaccgccttccgccggaacgg
ASPHrdBbtmCv2	actgttccagggcatggaaacaaggagaacgttcttggcgataggtgtggcggtaggcacgaacaacctctcggaacgtt
SPbtmbB*	ctggggcccgtcggatggtg
ASPbtmB*	cgctcgtccatgtgccgtagaag
pCAP01btmAASC1	ggaggccacggtgttgcccagcagggacaggccgtgggagatcaggtagagcggcagcatgtgtagtcctccttcgtattgcacgacatt
pCAP01btmAASC2	gatcgactgatgtcatcagcggtggagtgcaatgtcgtgcaatacgaaggaggactacac
pCAP01btmAASC3	gcactccaccgctgatgacatcagtcgatcatagcacgatcaacggcactgttgcaaata
pCAP01btmAASC4	tcaaggcgtaggtccctaggcaccaccgactatttgcaacagtgccgttgatcgtgctat
pCAP01btmAASC5	gtcggtggtgcctagggacctacgccttgaccttgatgaggcggcgtgagctacaatcaa
pCAP01btmAASC6	gcaggaacctatagtgagtcgtattgagtattgattgtagctcacgccgcctcatcaagg
pCAP01btmAASC7	tactcaatacgactcactataggttcctgctaaggaggcaacaaggtgcctaccgccacacctatcgccaagaacgtttcccttgtttccatgcccctggaacagt
SPbtmC*DPbtmD1	tgcgggccgactggaac
ASPbtmC*DPbtmD1	cgaagccgagccggtag
SPbtmC*DPbtmD2	aggagtggcagggggaac
ASPbtmC*DPbtmD2	gccgacgctcatgaggtg
OligoDPbtmD1	gcgtccggctgtggcgcgcccatcacgagtccgcgcgcttcggctaccggctcggcttcgactcgctgacgcgacgacgagcggccgggggctcccgccccatacgacgacgctccgcg
OligoDPbtmD2	ccctgcagatcccggaagcgcgtcccaccccaccaccgcccggaagaggcggcgcctcttccccgcggagcgtcgtcgtatgggcggggagccccggccgctcgtccgtcagcgtcagcga
OligoDPbtmD3	aagaggcccgcctcttccgggcggtggtggtggtgggacgcgcttccgggatctgcaggggcaggagtggcaggggagacgctggggaccaggccctggacacgctgcacgcctggcacagga
SPCbtmEAASDC	acgctcaccgctcgaccgtaaggaggccgccccgcacgcggcacacgggcccccgcgtcccagtcggcgcgcgacggactacc
ASPFbtmFAASDC	ggcctcggtgtccgcgtccaccggcccgttctcgaacatcgtccgcacctggaggtactcgagcagcgcccgctgcaccgcgtacg
SPDbtmCAASDC	tggaggagctggagtcctgggcgcctgggacggtgaagccacctcatgagcgtcggccggagttgccgcaatcggctcc
ASPEbtmCAASDC	cgcactccgtcgccgtcgcttcgcgcatgctccctctcccggtagtccgtcgcgccgactgggacgcggggggcccgtgtc
SPSF14btmDAASDC	gaggcggcgtgagctacaatcaatactcaatacgactcactataggttcctgctaaggaggcaacaagatgggacccgtagtcgtattcg
ASPCbtmDAASDC	cttggcgataggtgtggcggtaggcacgaaatcaaccccccggagccgatgcggcaactccggccgacgctcatgaggtg
pCAP01btmAASD7	tactcaatacgactcactataggttcctgctaaggaggcaacaagatgggaccccgtagtcgtattcgactgcatgaccgcggacttcctcaacgacgaccccaacaacgcggagctgagc
pCAP01btmAASutrC6	ggagccgattgcggcaactcgagtattgattgtagctcacgccgcctcatcaagg
pCAP01btmAASutrC7	tactcgagttgccgcaatcggctccggggttgatttcgtgcctaccgccacacctatcgccaagaacgtttcccttgtttccatgccctggaacagt
ASPHrdBrs	ggtgctgccaagggcatcaagacgatgctggtatcacccgtagctatactgactggtaatggaacgttgaaaaacggcttc
rsbtmC1	attaccagtcagtatagctacgggtgatacagcatcgtcttgatgcccttggcagcaccctgctaaggaggcaacaaggtgcctaccgccacacctatcgccaagaacgtttcccttgt
rsbtmC2	actgttccagggcatggaaacaaggagaacgttcttggcgataggtgtggcggtaggcaccttgttgcctccttagcagggtgctgccaagggcatcaagacgatgctggtatcacccgt
pCAP01btmAASrsC7	tactcaatacgactcactataggttcctgcggtgatacagcatcgtcttgatgcccttggcagcaccc
pCAP01btmAASrsC8	actgttccagggcatggaaacaaggagaacgttcttggcgataggtgtggcggtaggcaccttgttgcctccttagcagggtgctgccaagggcatcaagacgatgct
hrdBfw	accatagcgtcacaccctc
hrdBrv	ctgaccagattccggccact
btmDfw	gtattcgactgcatgaccgc
btmDrv	ccaggactccagctcctcca
btmCfw	gcaactgcgagggttccatg

btmCrv	gtggaggagggtgacgaagg
btmC-Dfw	tcgtcccccccgtacctctca
btmC-Drv	gctcagctccgcgttgttgg
SPPSF145'UTR	aatacgactcactataggttcc tgct
btmCp_rv	cggtgtcctccacatgacggt
SPTvaPSF14	tccgctgtgggccgacaaggcgttt ttcgg acgc gacgc aa ag ata agt gg aatct ta agg acctac gccttg accttt ga
ASPTvaAPSF14	tcccgctaatcgaatcaatcgtg ag acccg agca gga acct ata gtg agt
SPTvaPaac3	tccgctgtgggccgacaaggcgttt ttcgg acgc gacgc aa ag ata agt gg aatct ta agcacc accg acta tt
ASPTvaAPaac3	tcccgctaatcgaatcaatcgtg ag acccgcgt att gcacg aca ttgc ac
SPPSF14S4TvaA	actcaatacgactcactatag gttcct gctcgg gtctc acga ttg attcg
ASPPSF14S4TvaA	cgccgcctcatcaaggtcaaggcgt ag gtcccta ggtcct gggc gtca atg attc g
ASPPaac3S4TvaA	caacggcactgttgcaaata gtcg gtgg tgcct aggtcc tgg gcgtc aat gattc g
SPPaac3S4TvaA	cagcggtggagtgcaatgtc gtgc aat acgcg ggtctc acg attg attc g
SPTvaAPSF14	ggatcacgggcgaatcattg acgccc agg acct agg gacct acgcctt gacct tga
ASPTvaBPSF14	ccgaaatgccgcatcaatcgca acccctgc agc agg aacct at agtg agt
SPTvaAPaac3	ggatcacgggcgaatcattg acgccc agg acct aggc accacc gact att
ASPTvaBPaac3	ccgaaatgccgcatcaatcgca acccctgcc gtat tgcac gac attgc ac
SPPSF14S4TvaB	actcaatacgactcactatag gttcct gctgc agg gggt gcga ttg atgc
ASPS4TvaTvaB	acgggggccccgtcgatcagccagc gggc ggcccg ggca tcgccg gcgt agt agctc
SPPaac3S4TvaB	cagcggtggagtgcaatgtc gtgc aat acggc ag gggt tgcg attg atgc
TvaPSF14rsOligo1	tggcaatccgctgtgggccgaca aggc gtttt tcgg acgcg acgc aa ag ata agt gga atg acct acgcctt gacctt gat ga ggcg gcgtg agct aca atcaatactcggtaccaatac
PSF14rsOligo2	cctccttagcagggtgctgccaagg gcatc aa gac gat gctgg tatc accgg aacc tat agt gag tcgt attg gtacc gag tatt gat tgt agctc acgc cgcctcatcaaggtcaaggc
SPPSF14rsTvaAv2	tccggtgataccagcatcgtcttg atgccctt ggc agcaccc tgct aag ga ggca aca ag atg gacg ag gtcgcg ttca
ASPPaac3terTvaA	gctgatgacatcagtcgatca tagc acg atca acg gcact gttgc aa ata gtcg gtgg tgttcct gggc gtca atg attc
Paac3	cctaggcaccaccgactatttgca acag tgccgt tga tcgtgc tat gatc gact gat gtcatc agc ggtg ga gtgc aat gtcgt gca atac ga agg agg actacacatg
SPPaac3S4TvaB	cgactgatgtcatcagcggtgg agt gca atgtc gtgc aat acg aa gga agg actac acat gttgc gtctg gtgcct g
ASPTvaCTvaB	cctcccgccgggtggtgccgaggaccac ggg ggccccgtc gatc agcc agcg ggcg gcccgg gcatc gccggc gta gta g
SPTvaA	ccgcacccgagagagtttcag
ASPTvaA	gccttgcgcctcctcgtccg
TsaCoreAlba	ggagaaggccggcatatcgcc gg acga gg aggc gca aggc agcg tcatc ggcttc gcggtc acca tcgccgt gcact gctg agtt ag aca atg gca a gcagcgcttccgagcagaaggc
S1T	ggagaaggccggcatatcgcc gg acga gg aggc gca aggc accgtc atg gccgcca tcgccacc gtgg gcttacc actgc tgagtt agac aa tggc aa gcagcgcttccgagcagaaggc
SAV	gagagaggccggcatatcgccg gac gag ga ggcgc aa ggca gcgccg tcat ggccgcc atcgcc accgt ggctt accact gctg agtt ag aca atg gc aagcagcgcttccgagcaga agg
M3I	ggagaaggccggcatatcgcc gg acga gg aggc gca aggc agcg tcatc gccgcca tcgccacc gtgg gcttacc actgc tga gtt agac aa tggc aa gcagcgcttccgagcagaaggc
A5F	ggagaaggccggcatatcgcc gg acga gg aggc gca aggc agcg tcat ggccttc atcgcc accgt ggctt accact gctg agtt ag aca atg gca a gcagcgcttccgagcagaaggc
T8S	ggagaaggccggcatatcgcc gg acga gg aggc gca aggc agcg tcat ggccgcc atcgcctcc gtg gcttacc actgc tga gtt agac aa tggc aa gcagcgcttccgagcagaaggc
Y11V	ggagaaggccggcatatcgcc gg acga gg aggc gca aggc agcg tcat ggccgcc atcgcc accgt ggctg tccact gctg agtt ag aca atggc aa gcagcgcttccgagcagaaggc
H12A	ggagaaggccggcatatcgcc gg acga gg aggc gca aggc agcg tcat ggccgcc atcgcc accgt ggctt acgcct gctg agtt ag aca atggc aa gcagcgcttccgagcagaaggc
H12W	ggagaaggccggcatatcgcc gg acga gg aggc gca aggc agcg tcat ggccgcc atcgcc accgt ggctt actg gtgct gag tta gac aat ggca a gcagcgcttccgagcagaaggc

SPTvaB	tgagttagacaatggcaagc
ASPTvaB	ccagccgggctcggggtatc
SPPCRtTvaA	ggttgaagcatgtctgaaaccacgacagcagctcaggtgattccggggatccgtcgacc
ASPPCRtTvaA	ccttctgctcggaagcgctgcttgccattgtctaactcatgtaggctggagctgcttc
SPPCRtTvaB	gttgcgattgatgcggcatttcggaatgaggaggcgatgattccggggatccgtcgacc
ASPPCRtTvaBv3	ctgcgtcgtatctgctcatgcgcggctgctccgccccttttgtaggctggagctgcttc
SPPCRtTvaC	caccgggagtggggtgacctgtccccctgctgctgggggtgattccggggatccgtcgacc
ASPPCRtTvaCv4	cggtctcctgggtgctcatatggaggtcgaggggcgggctgtaggctggagctgcttc
SPPCRtTvaD	gcgcccctatctcaacgcccgcccctcgacctccatatgattccggggatccgtcgacc
ASPPCRtTvaDv4	cctcagcggtctctttcatgcagtggtcacccttcggtatgtaggctggagctgcttc
SPPCRtTvaE	accagtaacgaggtgaagatccacttcgaaggccgggtgattccgggggatccgtcgacc
ASPPCRtTvaE	agtttcactggtgctccttgcgtggtggtatcggaatcatgtaggctggagctgcttc
SPPCRtTvaF	cggcctgattccgataccaccacgcaaggagcaccagtgattccggggatccgtcgacc
ASPPCRtTvaFv2	cggtggtggggatgctcatgcgttgctctcctgctcgcctgtaggctggagctgcttc
SPPCRtTvaG	gcagcagcgcaccacgggcgagcaggagagcaacgcatgattccggggatccgtcgacc
ASPPCRtTvaGv2	tgttcaccgccgtcctcatgccgctcctcgactgtcagctgtaggctggagctgcttc
SPPCRtTvaH	ccggatcgaccacaccgtgcggcaccggagcgacgacatgattccggggatccgtcgacc
ASPPCRtTvaHv2	atgcagcttctgccgtcatgccgtctgctcccagtcctctgtaggctggagctgcttc
SPPCRtTvaI	gtgattctcttacaggttgtgctggtgtggtccgaggtgattccggggatccgtcgacc
ASPPCRtTvaI	cagtttccggcaacggtgcgcccggaacactgcggctatgtaggctggagctgcttc
SPPCRtTvaJ	cttgagaaaccgtcaggacgcttctaaaattctgccatgattccggggatccgtcgacc
ASPPCRtTvaJ	tcgcagggatgggatcgggcggtcggccgggccggatcatgtaggctggagctgcttc
SPPCRtTvaK	acggacgagcggaagggtacggggcggttggccgccgtgattccggggatccgtcgacc
ASPPCRtTvaK	cgtggttcaagggcacgcggtggaccggcacggcggctatgtaggctggagctgcttc
SPPCRtTvaL	ggggagatgtctttcatgaccgtcgagcacgaggggcatgattccggggatccgtcgacc
ASPPCRtTvaL	cgtcgggcaggaggggctggcaggaggggctgccggtcatgtaggctggagctgcttc
SPPCRtLgenes	tcggtttgacgcctcccatggtataaatagtggctcgagattccggggatccgtcgacc
ASPPCRtLgenes	ctgtggcgccgactcgttccggacgaagtgtggggggcggtgtaggctggagctgcttc
SPPCRtRgenes	acgcgtccgccgggccggctcatcgcaccggcgcgacgcattccggggatccgtcgacc
ASPPCRtRgenes	agtagcagcacgttccttatatgtagctttcgacatatgtgtaggctggagctgcttc
SPPCRt11840	cctcgccgccgtcgggtaccgaagggtgaccactgcatgattccggggatccgtcgacc
ASPPCRt11840v3	tggtcatgtctgggatcactgtccttctgctcgcgcggctgtaggctggagctgcttc
SPPCRt11820	ggatgtgccacctcccgctgacagtcgaggagcggcatgattccggggatccgtcgacc
ASPPCRt11820	cggccgtgctcgtgctcatgtcgtgctccggtgccgcactgtaggctggagctgcttc
SPPCRt11775	ctggatagtgatcgaagaggactgggagcagacggcatgatttccggggatccgtcgacc
ASPPCRt11775	agaatcactgccggggcggcccaacgtgccgccgccttatgtaggctggagctgcttc
SPPTvaAseq	gcacccgagaggtttcagc
ASPPTvaBseq	ccatgacgagggttccactc
ASPSrfITvaBseq	gccgggctcggggtatc
SPS4TvaA	ggttgaagcatgtctgaaac
ASPS4TvaA	ccttctgctcggaagcgctg
SPS4TvaB	gttgcgattgatgcggcatt
ASPS4TvaB	cctccgccatgctggcgaac
SPS4TvaC	caccgggagtggggtgacct

ASPS4TvaCv2	gcggtcgctcccggaggttc
SPS4TvaD	gcgcccctatctcaacgccc
ASPS4TvaDv2	gacagcgcctgctgtatctc
SPS4Tva40v3	tgcgtgcatgggcggcatc
ASPS4Tva40v3	gcttgagcacccggccttc
SPS4TvaE	accagtaacgaggtgaagat
ASPS4TvaE	agtttcactggtgctccttg
SPS4TvaF	cggcctgattccgataccac
ASPS4TvaF	cggtggtggggatgctcatg
SPS4TvaG	gcagcagcgcaccacgggcg
ASPS4TvaG	tgttcaccgccgtcctcatg
SPS411820	ggatgtgccacctcccgctg
ASPS411820	cggccgtgctcgtgctcatg
SPS4TvaHv3	aggcccgtcccctctc
ASPS4TvaHv3	atctcgccactgatacgtcg
SPS411775	ctggatagtgatcgaagagg
ASPS411775	agaatcactgccggggcggc
SPS4TvaI	gtgattctcttacaggttgt
ASPS4TvaI	cagtttccggcaacggtgcg
SPS4TvaJ	cttgagaaaccgtcaggacg
ASPS4TvaJ	tcgcagggatgggatcgggc
SPS4TvaK	acggacgagcggaagggtac
ASPS4TvaK	cgtggttcaagggcacgcgg
SPS4TvaLv3	gtcgcgtcggccgggtac
ASPS4TvaLv3	cacggcgaggacaagacggt
SPLgPCRtcheckv4	ctctctcgctttcctcattc
ASPLgPCRtcheckv4	gagaaatgcgtgggctctc
SPS4Rgenes	acgcgtccgccgggccggct
ASPS4Rgenes	agtagcagcacgttccttat
SPseqS4tvaE	accaagctctcaccacag
SPseqS4tvaF	ggcagcaagtcaaccttg
SPseqS4tvaG	gccgggcgcggagtcctg
SPseqS4tvaH	cgaggcccgtcccctctc
SPseqS4tvaI	agaccgtgaagaaatacg
SPseqS4tvaL	tgttccctgtctatgctg
SPseqS4tva75	gccctacgcctttctgac
SPaac(3)IVseq	cgctacggaaggagctgtgg
ASPaac(3)IVseq	cttcctgcatccgccagagg
SPNdeIS4TvaA	ggcccccatatggacgaggtcgcgttcagcga
ASPPacIS4TvaA	ctctagttaattaagctgcttgccattgtctaac
SPNdeIS4TvaB	ggcccccatatgttgcgtctggtgcctgtcgc
ASPPacIS4TvaB	ctctagttaattaagaacactgcgtcgtatctgc
SPNdeIS4TvaC	ggcccccatatggagacagagccgggcactct
ASPPacIS4TvaCv2	ctctagttaattaaccccccggtctcctgggtgctcagatg

SPNdeIS4TvaD	ggcccccatatgagcacccaggagaccggggg
ASPPacIS4TvaD	ctctagttaattaacggtacctcagcggtctctt
SPNdeIS4TvaE	ggcccccatatgctcaagctgggccagtccgc
ASPPacIS4TvaE	ctctagttaattaacttgcgtggtggtatcggaa
SPNdeIS4TvaF	ggcccccatatgaaactgcgccagagcgtccc
ASPPacIS4TvaF	ctctagttaattaaaaccgacggtggtggggatgc
SPNdeIS4TvaG	ggcccccatatgagcatccccaccaccgtcgg
ASPPacIS4TvaG	ctctagttaattaacaggttgttcaccgccgtcc
SPNdeIS4TvaH	ggcccccatatgagcacgagcacggccgatcc
ASPPacIS4TvaH	ctctagttaattaatgaagatgcagcttctgccg
SPNdeIS4TvaI	ggcccccatatgctgtccgtcggccaggccgt
ASPPacIS4TvaI	ctctagttaattaatgcgcccggaacactgcggg
SPNdeIS4TvaJ	ggcccccatatgaatattcccttctccgtgct
ASPPacIS4TvaJ	ctctagttaattaatccggcccggccgaccgccc
SPNdeIS4TvaK	ggcccccatatgtgggtgttcatgaccggtgg
ASPPacIS4TvaK	ctctagttaattaagcggtggaccggcacggcgg
SPNdeIS4TvaL	ggcccccatatggctgtggacgccggggcggg
ASPPacIS4TvaL	ctctagttaattaactggcaggaggggctgccgg
SPNdeIS411820	ggcccccatatgaggacggcggtgaacaacct
ASPPacIS411820	ctctagttaattaagggatcggccgtgctcgtgc
SPNdeIS411840	ggcccccatatgaaagagaccgctgaggtacc
ASPPacIS411840	ctctagttaattaagagcttggtcatgtctggga
SPNdeIS411775	ggcccccatatgacggcagaagctgcatcttc
ASPPacIS411775	ctctagttaattaacggcccaacgtgccgccgcc
SPNdeIS4TvaAv2	ggcccccatatgtctgaaaccacgacagcagc
SPNdeIS4TvaAv3	ggcccccatatggaaacccggcgacaacaacg
SPNdeIS4TvaBv2	ggcccccatatggttagacaatggcaagcagc
SPNdeIS4TvaCv2	ggcccccatatgagcagatacgacgcagtgtt
SPNdeIS4TvaCv3	ggcccccatatgttcgccagcatggcggagga
SPNdeIS4TvaEv2	ggcccccatatgatcccagacatgaccaagct
SPNdeIS4TvaEv3	ggcccccatatgaccaagctctcaccacagac
SPNdeIS4TvaEv4	ggcccccatatgaagatccacttcgaaggccg
SPNdeITaaP450	ggcccccatatgatgacggaaaccgtgaatc
ASPPacITaaP450	ctctagttaattaatcagcgtggcgtgaagag
SPNdeITaaRedu	ggcccccatatgatgaccggcgagaaccag
ASPNdeITaaRedu	ctctagttaattaatcaccggctcgtgctgctg
SPNdeIbtmC	ggttgatcatatgcctaccgccacacctatcg
ASPPacIbtmC	atggcgttaattaatcttctcggttccggcgctc
SPpIJbotPbtmC	tggtaggatcgtctagaacaggaggccccacggcgggaaatattcttc
ASPpIJbotbtmC	ccaagcttatgcaggactctagttaattaacggggttgctgtgcgatgc
SPpIJbotmetbtmC	aggatcgtctagaacaggaggccccatatgccatggaacagtgtcatc
SPpIJbotvalbtmC	aggatcgtctagaacaggaggccccatatgcctaccgccaaaaacgtttc
SPpIJ10257ins	cacgcggtcgatcttg
ASPpIJ10257ins	cgagctgaagaaagacaatc

References

Aggarwal, A.K., 1995. Structure and function of restriction endonucleases. Curr. Opin. Struct. Biol. 5, 11–19. https://doi.org/10.1016/0959-440X(95)80004-K

Alduina, R., Gallo, G., 2012. Artificial Chromosomes to Explore and to Exploit Biosynthetic Capabilities of Actinomycetes. J. Biomed. Biotechnol. 2012, 1–10. https://doi.org/10.1155/2012/462049

Alduina, R., Grazia, S., Dolce, L., Salerno, P., Sosio, M., Donadio, S., Puglia, A.M., 2003. Artificial chromosome libraries of Streptomyces coelicolor A3(2) and Planobispora rosea. Fed. Eur. Microbiol. Soc. Microbiol. Lett. 218, 181–186. https://doi.org/10.1111/j.1574-6968.2003.tb11516.x

Aller, S.G., Yu, J., Ward, A., Weng, Y., Chittaboina, S., Zhuo, R., Harrell, P.M., Trinh, Y.T., Zhang, Q., Urbatsch, I.L., Chang, G., 2009. Structure of P-glycoprotein reveals a molecular basis for poly-specific drug binding. Science (80-.). 323, 1718–1722. https://doi.org/10.1126/science.1168750

Altena, K., Guder, A., Cramer, C., Bierbaum, G., 2000. Biosynthesis of the lantibiotic mersacidin: organization of a type B lantibiotic gene cluster. Appl. Environ. Microbiol. 66, 2565–71. https://doi.org/10.1128/AEM.66.6.2565-2571.2000

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., Lipman, D.J., 1990. Basic local alignment search tool. J. Mol. Biol. 215, 403–410. https://doi.org/10.1016/S0022-2836(05)80360-2

Amagai, K., Ikeda, H., Hashimoto, J., Kozone, I., Izumikawa, M., Kudo, F., Eguchi, T., Nakamura, T., Osada, H., Takahashi, S., Shin-Ya, K., 2017. Identification of a gene cluster for telomestatin biosynthesis and heterologous expression using a specific promoter in a clean host. Sci. Rep. 7, 3382. https://doi.org/10.1038/s41598-017-03308-5

An, Y., Wu, W., Lv, A., 2010. A PCR-after-ligation method for cloning of multiple DNA inserts. Anal. Biochem. 402, 203–205. https://doi.org/10.1016/J.AB.2010.03.040

Anderson, B., Hodgkin, D.C., Viswamitra, M.A., 1970. The structure of thiostrepton. Nature 225, 233–235. https://doi.org/10.1038/225233a0

Aoki, M., Ohtsuka, T., Yamada, M., Ohba, Y., Watanabe, J., Yokose, K., Roche, N., 1991. Cyclothiazomycin, a novel polythiazole-containing peptide with renin inhibitory activity. Taxonomy, fermentation, isolation and physico-chemical characterization. J. Antibiot. (Tokyo). 44, 582–588. https://doi.org/10.7164/antibiotics.44.582

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. Nat. Prod. Rep. 30, 108–160. https://doi.org/10.1039/c2np20085f

Aso, Y., Nagao, J.-I., Koga, H., Okuda, K.-I., Kanemasa, Y., Sashihara, T., Nakayama, J., Sonomoto, K., 2004. Heterologous expression and functional analysis of the gene cluster for the biosynthesis of and immunity to the lantibiotic, nukacin ISK-1. J. Biosci. Bioeng. 98, 429–436. https://doi.org/10.1016/S1389-1723(05)00308-7

Aulakh, V.S., Ciufolini, M.A., 2011. Total Synthesis and Complete Structural Assignment of Thiocillin I. J. Am. Chem. Soc. 133, 5900–5904. https://doi.org/10.1021/ja110166x

Bagley, M.C., Dale, J.W., Merritt, E.A., Xiong, X., 2005. Thiopeptide Antibiotics. Chem. Rev. 105, 685–714. https://doi.org/10.1021/cr0300441

Baltz, R.H., 2010. Streptomyces and Saccharopolyspora hosts for heterologous expression of secondary metabolite gene clusters. J. Ind. Microbiol. Biotechnol. 37, 759–772. https://doi.org/10.1007/s10295-010-0730-9

Bankevich, A., Nurk, S., Antipov, D., Gurevich, A.A., Dvorkin, M., Kulikov, A.S., Lesin, V.M., Nikolenko, S.I., Pham, S., Prjibelski, A.D., Pyshkin, A. V., Sirotkin, A. V., Vyahhi, N., Tesler, G., Alekseyev, M.A., Pevzner, P.A., 2012. SPAdes: A New Genome Assembly Algorithm and Its Applications to Single-Cell Sequencing. J. Comput. Biol. 19, 455–477. https://doi.org/10.1089/cmb.2012.0021

Barajas, J.F., Blake-Hedges, J.M., Bailey, C.B., Curran, S., Keasling, J.D., 2017. Engineered polyketides: Synergy between protein and host level engineering. Synth. Syst. Biotechnol. 2, 147–166. https://doi.org/10.1016/J.SYNBIO.2017.08.005

Barka, E.A., Vatsa, P., Sanchez, L., Gaveau-Vaillant, N., Jacquard, C., Klenk, H.-P., Clément, C., Ouhdouch, Y., van Wezel, G.P., 2016. Taxonomy, Physiology, and Natural Products of Actinobacteria. Microbiol. Mol. Biol. Rev. 80, 1–43. https://doi.org/10.1128/MMBR.00019-15

Bartholomae, M., Buivydas, A., Viel, J.H., Montalbán-López, M., Kuipers, O.P., 2017. Major gene-regulatory mechanisms operating in ribosomally synthesized and post-translationally modified peptide (RiPP) biosynthesis. Mol. Microbiol. 106, 186–206. https://doi.org/10.1111/mmi.13764

Bauerle, M.R., Schwalm, E.L., Booker, S.J., 2015. Mechanistic Diversity of Radical *S*-Adenosylmethionine (SAM)-dependent Methylation. J. Biol. Chem. 290, 3995–4002. https://doi.org/10.1074/jbc.R114.607044

Becker, B., Cooper, M.A., 2013. Aminoglycoside Antibiotics in the 21st Century. Am. Chem. Soc. Chem. Biol. 8, 105–115. https://doi.org/10.1021/cb3005116

Benjdia, A., Guillot, A., Ruffié, P., Leprince, J., Berteau, O., 2017. Post-translational modification of ribosomally synthesized peptides by a radical SAM epimerase in Bacillus subtilis. Nat. Chem. 9, 698–707. https://doi.org/10.1038/nchem.2714

Benjdia, A., Pierre, S., Gherasim, C., Guillot, A., Carmona, M., Amara, P., Banerjee, R., Berteau, O., 2015. The thiostrepton A tryptophan methyltransferase TsrM catalyses a cob(II)alamin-dependent methyl transfer reaction. Nat. Commun. 6, 8377. https://doi.org/10.1038/ncomms9377

Bentley, R., 2004. The Molecular Structure of Penicillin. J. Chem. Educ. 81, 1462. https://doi.org/10.1021/ed081p1462

Bentley, S.D., Chater, K.F., Cerdeño-Tárraga, A.-M., Challis, G.L., Thomson, N.R., James, K.D., Harris, D.E., Quail, M.A., Kieser, H., Harper, D., Bateman, A., Brown, S., Chandra, G., Chen, C.W., Collins, M., Cronin, A., Fraser, A., Goble, A., Hidalgo, J., Hornsby, T., Howarth, S., Huang, C.-H., Kieser, T., Larke, L., Murphy, L., Oliver, K., O'Neil, S., Rabbinowitsch, E., Rajandream, M.-A., Rutherford, K., Rutter, S., Seeger, K., Saunders, D., Sharp, S., Squares, R., Squares, S., Taylor, K., Warren, T., Wietzorrek, A., Woodward, J., Barrell, B.G., Parkhill, J., Hopwood, D.A., 2002. Complete genome sequence of the model actinomycete Streptomyces coelicolor A3(2). Nature 417, 141–147. https://doi.org/10.1038/417141a

Bhattacharjee, S., Halane, M.K., Kim, S.H., Gassmann, W., 2011. Pathogen Effectors Target Arabidopsis EDS1 and Alter Its Interactions with Immune Regulators. Science (80-.). 334, 1405–1408. https://doi.org/10.1126/science.1211592

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, 533–545. https://doi.org/10.1111/j.1365-2958.1994.tb02187.x

Bierbaum, G., 1996. The biosynthesis of the lantibiotics epidermin, gallidermin, Pep5 and epilancin K7. Antonie van Leeuwenhoek, Int. J. Gen. Mol. Microbiol. 69, 119–127. https://doi.org/10.1007/BF00399417

Bitinaite, J., Rubino, M., Varma, K.H., Schildkraut, I., Vaisvila, R., Vaiskunaite, R., 2007. USER™ friendly DNA engineering and cloning method by uracil excision. Nucleic Acids Res. 35, 1992–2002. https://doi.org/10.1093/nar/gkm041

Blaesse, M., Kupke, T., Huber, R., Steinbacher, S., 2003. Structure of MrsD, an FAD-binding protein of the HFCD family. Acta Crystallogr. - Sect. D Biol. Crystallogr. 59, 1414–1421. https://doi.org/10.1107/S0907444903011831

Blaesse, M., Kupke, T., Huber, R., Steinbacher, S., 2000. Crystal structure of the peptidyl-cysteine decarboxylase EpiD complexed with a pentapeptide substrate. Eur. Mol. Biol. Organ. J. 19, 6299–6310. https://doi.org/10.1093/emboj/19.23.6299

Blake, W.J., Chapman, B.A., Zindal, A., Lee, M.E., Lippow, S.M., Baynes, B.M., 2010. Pairwise selection assembly for sequence-independent construction of long-length DNA. Nucleic Acids Res. 38, 2594–2602. https://doi.org/10.1093/nar/gkq123

Blaszczyk, A.J., Silakov, A., Zhang, B., Maiocco, S.J., Lanz, N.D., Kelly, W.L., Elliott, S.J., Krebs, C., Booker, S.J., 2016. Spectroscopic and Electrochemical Characterization of the Iron–Sulfur and Cobalamin Cofactors of TsrM, an Unusual Radical *S* -Adenosylmethionine Methylase. J. Am. Chem. Soc. 138, 3416–3426. https://doi.org/10.1021/jacs.5b12592

Bowers, A.A., Walsh, C.T., Acker, M.G., 2010. Genetic Interception and Structural Characterization of Thiopeptide Cyclization Precursors from *Bacillus cereus*. J. Am. Chem. Soc. 132, 12182–12184. https://doi.org/10.1021/ja104524q

Breil, B.T., Ludden, P.W., Triplett, E.W., 1993. DNA sequence and mutational analysis of genes involved in the production and resistance of the antibiotic peptide trifolitoxin. J. Bacteriol. 175, 3693–3702. https://doi.org/10.1128/jb.175.12.3693-3702.1993

Breukink, E., Wiedemann, I., Van Kraaij, C., Kuipers, O.P., Sahl, H.G., De Kruijff, B., 1999. Use of the cell wail precursor lipid II by a pore-forming peptide antibiotic. Science (80-.). 286, 2361–2364. https://doi.org/10.1126/science.286.5448.2361

Brötz, H., Josten, M., Wiedemann, I., Schneider, U., Götz, F., Bierbaum, G., Sahl, H.-G., 1998. Role of

lipid-bound peptidoglycan precursors in the formation of pores by nisin, epidermin and other lantibiotics. Mol. Microbiol. 30, 317–327. https://doi.org/10.1046/j.1365-2958.1998.01065.x

Broughton, L.J., Giuntini, F., Savoie, H., Bryden, F., Boyle, R.W., Maraveyas, A., Madden, L.A., 2016. Duramycin-porphyrin conjugates for targeting of tumour cells using photodynamic therapy. J. Photochem. Photobiol. B Biol. 163, 374–384. https://doi.org/10.1016/J.JPHOTOBIOL.2016.09.001

Brown, A.S., Calcott, M.J., Owen, J.G., Ackerley, D.F., 2018. Structural, functional and evolutionary perspectives on effective re-engineering of non-ribosomal peptide synthetase assembly lines. Nat. Prod. Rep. https://doi.org/10.1039/C8NP00036K

Bryant, F.R., 1988. Construction of a recombinase-deficient mutant recA protein that retains single-stranded DNA-dependent ATPase activity. J. Biol. Chem. 263, 8716–8723.

Burkhart, B.J., Hudson, G.A., Dunbar, K.L., Mitchell, D.A., 2015. A prevalent peptide-binding domain guides ribosomal natural product biosynthesis. Nat. Chem. Biol. 11, 564–570. https://doi.org/10.1038/nchembio.1856

Burkhart, B.J., Kakkar, N., Hudson, G.A., Van Der Donk, W.A., Mitchell, D.A., 2017a. Chimeric Leader Peptides for the Generation of Non-Natural Hybrid RiPP Products. Am. Chem. Soc. Cent. Sci. 3, 629–638. https://doi.org/10.1021/acscentsci.7b00141

Burkhart, B.J., Schwalen, C.J., Mann, G., Naismith, J.H., Mitchell, D.A., 2017b. YcaO-Dependent Posttranslational Amide Activation: Biosynthesis, Structure, and Function. Chem. Rev. 117, 5389–5456. https://doi.org/10.1021/acs.chemrev.6b00623

Cain, J.A., Solis, N., Cordwell, S.J., 2014. Beyond gene expression: The impact of protein post-translational modifications in bacteria. J. Proteomics 97, 265–286. https://doi.org/10.1016/J.JPROT.2013.08.012

Cameron, D.M., Thompson, J., March, P.E., Dahlberg, A.E., 2002. Initiation Factor IF2, Thiostrepton and Micrococcin Prevent the Binding of Elongation Factor G to the Escherichia coli Ribosome. J. Mol. Biol. 319, 27–35. https://doi.org/10.1016/S0022-2836(02)00235-8

Casini, A., MacDonald, J.T., Jonghe, J. De, Christodoulou, G., Freemont, P.S., Baldwin, G.S., Ellis, T., 2013. One-pot DNA construction for synthetic biology: The Modular Overlap-Directed Assembly with Linkers (MODAL) strategy. Nucleic Acids Res. 42, e7–e7. https://doi.org/10.1093/nar/gkt915

Castiglione, F., Lazzarini, A., Carrano, L., Corti, E., Ciciliato, I., Gastaldo, L., Candiani, P., Losi, D., Marinelli, F., Selva, E., Parenti, F., 2008. Determining the Structure and Mode of Action of Microbisporicin, a Potent Lantibiotic Active Against Multiresistant Pathogens. Chem. Biol. 15, 22–31. https://doi.org/10.1016/J.CHEMBIOL.2007.11.009

Chain, P.S., Hernandez-Lucas, I., Golding, B., Finan, T.M., 2000. oriT-directed cloning of defined large regions from bacterial genomes: identification of the Sinorhizobium meliloti pExo megaplasmid replicator region. J. Bacteriol. 182, 5486–5494.

Charlesworth, J.C., Burns, B.P., 2015. Untapped Resources: Biotechnological Potential of Peptides and Secondary Metabolites in Archaea. Archaea 2015, 1–7. https://doi.org/10.1155/2015/282035

Chater, K.F., 2006. Streptomyces inside-out: A new perspective on the bacteria that provide us with antibiotics. Philos. Trans. R. Soc. B Biol. Sci. 361, 761–768. https://doi.org/10.1098/rstb.2005.1758

Chatterjee, C., Miller, L.M., Leung, Y.L., Xie, L., Yi, M., Kelleher, N.L., van der Donk, W.A., 2005. Lacticin 481 Synthetase Phosphorylates its Substrate during Lantibiotic Production. J. Am.

Chem. Soc. 127, 15332–15333. https://doi.org/10.1021/JA0543043

Chatterjee, S., Chatterjee, S., Lad, S.J., Phansalkar, M.R., Rupp, R.H., Ganguli, B.N., 1992. Mersacidin, a New Antibiotic from Bacillus - Fermentation, Isolation, Purification and Chemical Characterization. J. Antibiot. (Tokyo). 45, 832–838. https://doi.org/10.7164/antibiotics.45.832

Chen, L., Wang, H., Zhang, J., Gu, L., Huang, N., Zhou, J.-M., Chai, J., 2008. Structural basis for the catalytic mechanism of phosphothreonine lyase. Nat. Struct. Mol. Biol. 15, 101–102. https://doi.org/10.1038/nsmb1329

Chen, P., Qi, F.X., Novak, J., Krull, R.E., Caufield, P.W., 2001. Effect of amino acid substitutions in conserved residues in the leader peptide on biosynthesis of the lantibiotic mutacin II. Fed. Eur. Microbiol. Soc. Microbiol. Lett. 195, 139–144. https://doi.org/10.1016/S0378-1097(00)00565-6

Chen, W.-H., Qin, Z.-J., Wang, J., Zhao, G.-P., 2013. The MASTER (methylation-assisted tailorable ends rational) ligation method for seamless DNA assembly. Nucleic Acids Res. 41, e93–e93. https://doi.org/10.1093/nar/gkt122

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. https://doi.org/10.1021/jacs.7b08417

Chen, Y., de Bruyn Kops, C., Kirchmair, J., 2017. Data Resources for the Computer-Guided Discovery of Bioactive Natural Products. J. Chem. Inf. Model. 57, 2099–2111. https://doi.org/10.1021/acs.jcim.7b00341

Chiou, S.-J., Riordan, C.G., Rheingold, A.L., 2003. Synthetic modeling of zinc thiolates: Quantitative assessment of hydrogen bonding in modulating sulfur alkylation rates. Proc. Natl. Acad. Sci. 100, 3695–3700. https://doi.org/10.1073/pnas.0637221100

Christian, M., Cermak, T., Doyle, E.L., Schmidt, C., Zhang, F., Hummel, A., Bogdanove, A.J., Voytas, D.F., 2010. Targeting DNA double-strand breaks with TAL effector nucleases. Genetics 186, 756–761. https://doi.org/10.1534/genetics.110.120717

Ciufolini, M.A., Lefranc, D., 2010. Microccocin P1: Structure, biology and synthesis. Nat. Prod. Rep. 27, 330–342. https://doi.org/10.1039/b919071f

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. 107, 16297–16302. https://doi.org/10.1073/pnas.1008608107

Cobb, R.E., Wang, Y., Zhao, H., 2015. High-efficiency multiplex genome editing of Streptomyces species using an engineered CRISPR/Cas system. Am. Chem. Soc. Synth. Biol. 4, 723–8. https://doi.org/10.1021/sb500351f

Cogan, D.P., Hudson, G.A., Zhang, Z., Pogorelov, T. V, van der Donk, W.A., Mitchell, D.A., Nair, S.K., 2017. Structural insights into enzymatic [4+2] aza-cycloaddition in thiopeptide antibiotic biosynthesis. Proc. Natl. Acad. Sci. U. S. A. 114, 12928–12933. https://doi.org/10.1073/pnas.1716035114

Cohen, S.N., Chang, A.C., Boyer, H.W., Helling, R.B., 1973. Construction of biologically functional bacterial plasmids in vitro. Proc. Natl. Acad. Sci. 70, 3240–3244. https://doi.org/10.1073/PNAS.70.11.3240

Collin, F., Thompson, R.E., Jolliffe, K.A., Payne, R.J., Maxwell, A., 2013. Fragments of the Bacterial Toxin Microcin B17 as Gyrase Poisons. Public Libr. Sci. One 8, e61459.

https://doi.org/10.1371/journal.pone.0061459

Cong, L., Ran, F.A., Cox, D., Lin, S., Barretto, R., Habib, N., Hsu, P.D., Wu, X., Jiang, W., Marraffini, L.A., Zhang, F., 2013. Multiplex Genome Engineering Using CRISPR/Cas Systems. Science (80-.). 339, 819–823. https://doi.org/10.1126/science.1231143

Cost, G.J., 2007. Enzymatic ligation assisted by nucleases: simultaneous ligation and digestion promote the ordered assembly of DNA. Nat. Protoc. 2, 2198–2202. https://doi.org/10.1038/nprot.2007.325

Cotter, P.D., O'Connor, P.M., Draper, L.A., Lawton, E.M., Deegan, L.H., Hill, C., Ross, R.P., 2005. Posttranslational conversion of L-serines to D-alanines is vital for optimal production and activity of the lantibiotic lacticin 3147. Proc. Natl. Acad. Sci. 102, 18584–18589. https://doi.org/10.1073/pnas.0509371102

Cox, C.L., Doroghazi, J.R., Mitchell, D.A., 2015. The genomic landscape of ribosomal peptides containing thiazole and oxazole heterocycles. BioMed Cent. genomics 16, 778. https://doi.org/10.1186/s12864-015-2008-0

Cron, M.J., Fardig, O.B., Johnson, D.L., Whitehead, D.F., Hooper, I.R., Lemieux, R.U., 1958. Kanamycin. IV. The Structure of Kanamycin. J. Am. Chem. Soc. 80, 4115–4115. https://doi.org/10.1021/ja01548a077

Crone, W.J.K., Leeper, F.J., Truman, A.W., 2012. Identification and characterisation of the gene cluster for the anti-MRSA antibiotic bottromycin: expanding the biosynthetic diversity of ribosomal peptides. Chem. Sci. 3, 3516–3521. https://doi.org/10.1039/c2sc21190d

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. Angew. Chemie Int. Ed. 55, 9639–9643. https://doi.org/10.1002/anie.201604304

Czekster, C.M., Ge, Y., Naismith, J.H., 2016. Mechanisms of cyanobactin biosynthesis. Curr. Opin. Chem. Biol. 35, 80–88. https://doi.org/10.1016/j.cbpa.2016.08.029

Daigle, D.M., McKay, G.A., Thompson, P.R., Wright, G.D., 1999. Aminoglycoside antibiotic phosphotransferases are also serine protein kinases. Chem. Biol. 6, 11–18. https://doi.org/10.1016/S1074-5521(99)80016-7

Das, A., Yanofsky, C., 1989. Restoration of a translational stop-start overlap reinstates translational coupling in a mutant trpb'-trpa gene pair of the Escherichia coli tryptophan operon. Nucleic Acids Res. 17, 9333–9340. https://doi.org/10.1093/nar/17.22.9333

Das, A., Yanofsky, C., 1984. A ribosome binding site sequence is necessary for efficient expression of the distal gene of a translationally-coupled gene pair. Nucleic Acids Res. 12, 4757–4768. https://doi.org/10.1093/nar/12.11.4757

Deane, C.D., Burkhart, B.J., Blair, P.M., Tietz, J.I., Lin, A., Mitchell, D.A., 2016. In Vitro Biosynthesis and Substrate Tolerance of the Plantazolicin Family of Natural Products. Am. Chem. Soc. Chem. Biol. 11, 2232–2243. https://doi.org/10.1021/acschembio.6b00369

Demirci, H., Gregory, S.T., Dahlberg, A.E., Jogl, G., 2008. Multiple-Site Trimethylation of Ribosomal Protein L11 by the PrmA Methyltransferase. Structure 16, 1059–1066. https://doi.org/10.1016/j.str.2008.03.016

Ding, W., Ji, W., Wu, Y., Wu, R., Liu, W.-Q., Mo, T., Zhao, J., Ma, X., Zhang, W., Xu, P., Deng, Z., Tang, B., Yu, Y., Zhang, Q., 2017a. Biosynthesis of the nosiheptide indole side ring centers on a cryptic carrier protein NosJ. Nat. Commun. 8, 437. https://doi.org/10.1038/s41467-017-00439-1

Ding, W., Li, Y., Zhao, J., Ji, X., Mo, T., Qianzhu, H., Tu, T., Deng, Z., Yu, Y., Chen, F., Zhang, Q., 2017b.

The Catalytic Mechanism of the Class C Radical S-Adenosylmethionine Methyltransferase NosN. Angew. Chemie - Int. Ed. 56, 3857–3861. https://doi.org/10.1002/anie.201609948

Ding, W., Weng, H., Jin, P., Du, G., Chen, J., Kang, Z., 2017. Scarless assembly of unphosphorylated DNA fragments with a simplified DATEL method. Bioengineered 8, 296–301. https://doi.org/10.1080/21655979.2017.1308986

Ding, W., Wu, Y., Ji, X., Qianzhu, H., Chen, F., Deng, Z., Yu, Y., Zhang, Q., 2017c. Nucleoside-linked shunt products in the reaction catalyzed by the class C radical S-adenosylmethionine methyltransferase NosN. Chem. Commun. 53, 5235–5238. https://doi.org/10.1039/C7CC02162C

Dirix, G., Monsieurs, P., Dombrecht, B., Daniels, R., Marchal, K., Vanderleyden, J., Michiels, J., 2004. Peptide signal molecules and bacteriocins in Gram-negative bacteria: a genome-wide in silico screening for peptides containing a double-glycine leader sequence and their cognate transporters. Peptides 25, 1425–1440. https://doi.org/10.1016/J.PEPTIDES.2003.10.028

Dong, H., Zhang, D., 2014. Current development in genetic engineering strategies of Bacillus species. Microb. Cell Fact. 13, 63. https://doi.org/10.1186/1475-2859-13-63

Dong, S.H., Tang, W., Lukk, T., Yu, Y., Nair, S.K., van der donk, W.A., 2015. The enterococcal cytolysin synthetase has an unanticipated lipid kinase fold. Elife 4. https://doi.org/10.7554/eLife.07607

Donia, M.S., Hathaway, B.J., Sudek, S., Haygood, M.G., Rosovitz, M.J., Ravel, J., Schmidt, E.W., 2006. Natural combinatorial peptide libraries in cyanobacterial symbionts of marine ascidians. Nat. Chem. Biol. 2, 729–735. https://doi.org/10.1038/nchembio829

Doroghazi, J.R., Albright, J.C., Goering, A.W., Ju, K.-S., Haines, R.R., Tchalukov, K.A., Labeda, D.P., Kelleher, N.L., Metcalf, W.W., 2014. A roadmap for natural product discovery based on large-scale genomics and metabolomics. Nat. Chem. Biol. 10, 963–968. https://doi.org/10.1038/nchembio.1659

Driggers, E.M., Hale, S.P., Lee, J., Terrett, N.K., 2008. The exploration of macrocycles for drug discovery — an underexploited structural class. Nat. Rev. Drug Discov. 7, 608–624. https://doi.org/10.1038/nrd2590

Du, D., Wang, L., Tian, Y., Liu, H., Tan, H., Niu, G., 2015. Genome engineering and direct cloning of antibiotic gene clusters via phage φBT1 integrase-mediated site-specific recombination in Streptomyces. Sci. Rep. 5, 8740. https://doi.org/10.1038/srep08740

Du, D., Zhu, Y., Wei, J., Tian, Y., Niu, G., Tan, H., 2013. Improvement of gougerotin and nikkomycin production by engineering their biosynthetic gene clusters. Appl. Microbiol. Biotechnol. 97, 6383–6396. https://doi.org/10.1007/s00253-013-4836-7

Duan, L., Wang, S., Liao, R., Liu, W., 2012. Insights into Quinaldic Acid Moiety Formation in Thiostrepton Biosynthesis Facilitating Fluorinated Thiopeptide Generation. Chem. Biol. 19, 443–448. https://doi.org/10.1016/j.chembiol.2012.02.008

Dugaiczyk, A., Boyer, H.W., Goodman, H.M., 1975. Ligation of EcoRI endonuclease-generated DNA fragments into linear and circular structures. J. Mol. Biol. 96, 171–184. https://doi.org/10.1016/0022-2836(75)90189-8

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. Nat. Chem. Biol. 10, 823–829. https://doi.org/10.1038/nchembio.1608

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, 569–575.

https://doi.org/10.1038/nchembio.944

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, 7672–7677. https://doi.org/10.1021/jacs.5b04682

Edgar, R.C., 2004. MUSCLE: Multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 32, 1792–1797. https://doi.org/10.1093/nar/gkh340

Ekkelenkamp, M.B., Hanssen, M., Danny Hsu, S.-T., de Jong, A., Milatovic, D., Verhoef, J., van Nuland, N.A.J., 2005. Isolation and structural characterization of epilancin 15X, a novel lantibiotic from a clinical strain of Staphylococcus epidermidis. Fed. Eur. Biochem. Soc. Lett. 579, 1917–1922. https://doi.org/10.1016/j.febslet.2005.01.083

Emmerich, R., Löw, O., 1899. Bakteriolytische Enzyme als Ursache der erworbenen Immunität und die Heilung von Infectionskrankheiten durch dieselben. Zeitschrift für Hyg. und Infect. 31, 1–65. https://doi.org/10.1007/BF02206499

Enghiad, B., Zhao, H., 2017. Programmable DNA-Guided Artificial Restriction Enzymes. Am. Chem. Soc. Synth. Biol. 6, 752–757. https://doi.org/10.1021/acssynbio.6b00324

Engler, C., Gruetzner, R., Kandzia, R., Marillonnet, S., 2009. Golden Gate Shuffling: A One-Pot DNA Shuffling Method Based on Type IIs Restriction Enzymes. Public Libr. Sci. One 4, e5553. https://doi.org/10.1371/journal.pone.0005553

Engler, C., Kandzia, R., Marillonnet, S., 2008. A One Pot, One Step, Precision Cloning Method with High Throughput Capability. Public Libr. Sci. One 3, e3647. https://doi.org/10.1371/journal.pone.0003647

Escano, J., Stauffer, B., Brennan, J., Bullock, M., Smith, L., 2015. Biosynthesis and transport of the lantibiotic mutacin 1140 produced by Streptococcus mutans. J. Bacteriol. 197, 1173–1184. https://doi.org/10.1128/JB.02531-14

Eyles, T.H., Vior, N.M., Truman, A.W., 2018. Rapid and Robust Yeast-Mediated Pathway Refactoring Generates Multiple New Bottromycin-Related Metabolites. Am. Chem. Soc. Synth. Biol. 7, 1211–1218. https://doi.org/10.1021/acssynbio.8b00038

Fan, Y.-Q., Liu, H.-J., Yan, L., Luan, Y.-S., Zhou, H.-M., Yang, J.-M., Yin, S.-J., Wang, Y.-L., 2013. Optimized transformation of Streptomyces sp. ATCC 39366 producing leptomycin by electroporation. J. Microbiol. 51, 318–322. https://doi.org/10.1007/s12275-013-2428-y

Fang, L., Zhang, G., Pfeifer, B.A., 2017. Engineering of E. coli for Heterologous Expression of Secondary Metabolite Biosynthesis Pathways Recovered from Metagenomics Libraries, in: Functional Metagenomics: Tools and Applications. Springer International Publishing, Cham, pp. 45–63. https://doi.org/10.1007/978-3-319-61510-3_3

Feeney, M.A., Chandra, G., Findlay, K.C., Paget, M.S.B., Buttner, M.J., 2017. Translational Control of the SigR-Directed Oxidative Stress Response in Streptomyces via IF3-Mediated Repression of a Noncanonical GTC Start Codon. MBio 8, e00815-17. https://doi.org/10.1128/mBio.00815-17

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. Mol. Pharm. 5, 191–211. https://doi.org/10.1021/mp700137g

Férir, G., Petrova, M.I., Andrei, G., Huskens, D., Hoorelbeke, B., Snoeck, R., Vanderleyden, J., Balzarini, J., Bartoschek, S., Brönstrup, M., Süssmuth, R.D., Schols, D., 2013. The Lantibiotic Peptide Labyrinthopeptin A1 Demonstrates Broad Anti-HIV and Anti-HSV Activity with Potential for Microbicidal Applications. Public Libr. Sci. One 8, e64010. https://doi.org/10.1371/journal.pone.0064010

Fernández-Martínez, L.T., Bibb, M.J., 2014. Use of the Meganuclease I-SceI of Saccharomyces cerevisiae to select for gene deletions in actinomycetes. Sci. Rep. 4, 7100. https://doi.org/10.1038/srep07100

Flärdh, K., Buttner, M.J., 2009. Streptomyces morphogenetics: dissecting differentiation in a filamentous bacterium. Nat. Rev. Microbiol. 7, 36–49. https://doi.org/10.1038/nrmicro1968

Fleming, A., 1929. On the antibacterial action of cultures of a penicillium, with special reference to their use in the isolation of *B. influenzae* . Br. J. Exp. Pathol. 10, 226–236.

Flinspach, K., Kapitzke, C., Tocchetti, A., Sosio, M., Apel, A.K., 2014. Heterologous expression of the thiopeptide antibiotic GE2270 from Planobispora rosea ATCC 53733 in Streptomyces coelicolor requires deletion of ribosomal genes from the expression construct. Public Libr. Sci. One 9, e90499. https://doi.org/10.1371/journal.pone.0090499

Fong, D.H., Burk, D.L., Blanchet, J., Yan, A.Y., Berghuis, A.M., 2017. Structural Basis for Kinase-Mediated Macrolide Antibiotic Resistance. Structure 25, 750–761.e5. https://doi.org/10.1016/j.str.2017.03.007

Fraczek, M.G., Naseeb, S., Delneri, D., 2018. History of genome editing in yeast. Yeast 35, 361–368. https://doi.org/10.1002/yea.3308

Franke, C.M., Tiemersma, J., Venema, G., Kok, J., 1999. Membrane topology of the lactococcal bacteriocin ATP-binding cassette transporter protein LcnC. Involvement of LcnC in lactococcin a maturation. J. Biol. Chem. 274, 8484–8490. https://doi.org/10.1074/JBC.274.13.8484

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. https://doi.org/10.1021/jacs.7b09898

Frattaruolo, L., Lacret, R., Cappello, A.R., Truman, A.W., 2017. A Genomics-Based Approach Identifies a Thioviridamide-Like Compound with Selective Anticancer Activity. Am. Chem. Soc. Chem. Biol. 12, 2815–2822. https://doi.org/10.1021/acschembio.7b00677

Freeman, M.F., Gurgui, C., Helf, M.J., Morinaka, B.I., Uria, A.R., Oldham, N.J., Sahl, H.G., Matsunaga, S., Piel, J., 2012. Metagenome mining reveals polytheonamides as posttranslationally modified ribosomal peptides. Science (80-.). 338, 387–390. https://doi.org/10.1126/science.1226121

Freeman, M.F., Helf, M.J., Bhushan, A., Morinaka, B.I., Piel, J., 2017. Seven enzymes create extraordinary molecular complexity in an uncultivated bacterium. Nat. Chem. 9, 387–395. https://doi.org/10.1038/nchem.2666

Fu, J., Bian, X., Hu, S., Wang, H., Huang, F., Seibert, P.M., Plaza, A., Xia, L., Müller, R., Stewart, A.F., Zhang, Y., 2012. Full-length RecE enhances linear-linear homologous recombination and facilitates direct cloning for bioprospecting. Nat. Biotechnol. 30, 440–446. https://doi.org/10.1038/nbt.2183

Funk, M.A., van der Donk, W.A., 2017. Ribosomal Natural Products, Tailored To Fit. Acc. Chem. Res. 50, 1577–1586. https://doi.org/10.1021/acs.accounts.7b00175

Furgerson Ihnken, L.A., Chatterjee, C., van der Donk, W.A., 2008. In vitro reconstitution and substrate specificity of a lantibiotic protease. Biochemistry 47, 7352–63. https://doi.org/10.1021/bi800278n

Galm, U., Shen, B., 2006. Expression of biosynthetic gene clusters in heterologous hosts for natural product production and combinatorial biosynthesis. Expert Opin. Drug Discov. 1, 409–437. https://doi.org/10.1517/17460441.1.5.409

Ganz, T., 2003. Defensins: antimicrobial peptides of innate immunity. Nat. Rev. Immunol. 3, 710–

720. https://doi.org/10.1038/nri1180

Garg, N., Goto, Y., Chen, T., van der Donk, W.A., 2016. Characterization of the stereochemical configuration of lanthionines formed by the lanthipeptide synthetase GeoM. Biopolymers 106, 834–842. https://doi.org/10.1002/bip.22876

Ghilarov, D., Serebryakova, M., Stevenson, C.E.M., Hearnshaw, S.J., Volkov, D.S., Maxwell, A., Lawson, D.M., Severinov, K., 2017. The Origins of Specificity in the Microcin-Processing Protease TldD/E. Structure 25, 1549–1561.e5. https://doi.org/10.1016/j.str.2017.08.006

Gibson, D.G., 2009. Synthesis of DNA fragments in yeast by one-step assembly of overlapping oligonucleotides. Nucleic Acids Res. 37, 6984–6990. https://doi.org/10.1093/nar/gkp687

Gibson, D.G., Benders, G.A., Andrews-Pfannkoch, C., Denisova, E.A., Baden-Tillson, H., Zaveri, J., Stockwell, T.B., Brownley, A., Thomas, D.W., Algire, M.A., Merryman, C., Young, L., Noskov, V.N., Glass, J.I., Venter, J.C., Hutchison, C.A., Smith, H.O., 2008a. Complete chemical synthesis, assembly, and cloning of a Mycoplasma genitalium genome. Science (80-.). 319, 1215–1220. https://doi.org/10.1126/science.1151721

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., 2008b. One-step assembly in yeast of 25 overlapping DNA fragments to form a complete synthetic Mycoplasma genitalium genome. Proc. Natl. Acad. Sci. 105, 20404–20409. https://doi.org/10.1073/pnas.0811011106

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. Nat. Methods 6, 343–345. https://doi.org/10.1038/nmeth.1318

Gietz, R.D., Woods, R.A., 2006. Yeast Transformation by the LiAc/SS Carrier DNA/PEG Method, in: Yeast Protocols. Humana Press, New Jersey, pp. 107–120. https://doi.org/10.1385/1-59259-958-3:107

Gomez-Escribano, J.P., Bibb, M.J., 2014. Heterologous expression of natural product biosynthetic gene clusters in Streptomyces coelicolor: from genome mining to manipulation of biosynthetic pathways. J. Ind. Microbiol. Biotechnol. 41, 425–431. https://doi.org/10.1007/s10295-013-1348-5

Gomez-Escribano, J.P., Bibb, M.J., 2011. Engineering Streptomyces coelicolor for heterologous expression of secondary metabolite gene clusters. Microb. Biotechnol. 4, 207–215. https://doi.org/10.1111/j.1751-7915.2010.00219.x

Gomez-Escribano, J.P., Song, L., Bibb, M.J., Challis, G.L., 2012. Posttranslational β-methylation and macrolactamidination in the biosynthesis of the bottromycin complex of ribosomal peptide antibiotics. Chem. Sci. 3, 3522–3525. https://doi.org/10.1039/c2sc21183a

Gomi, K., Iimura, Y., Hara, S., 1987. Integrative Transformation of *Aspergillus oryzae* with a Plasmid Containing the *Aspergillus nidulans argB* Gene. Agric. Biol. Chem. 51, 2549–2555. https://doi.org/10.1080/00021369.1987.10868429

González-Cerón, G., Miranda-Olivares, O.J., Servín-González, L., 2009. Characterization of the methyl-specific restriction system of Streptomyces coelicolor A3(2) and of the role played by laterally acquired nucleases. Fed. Eur. Microbiol. Soc. Microbiol. Lett. 301, 35–43. https://doi.org/10.1111/j.1574-6968.2009.01790.x

Goto, Y., Li, B., Claesen, J., Shi, Y., Bibb, M.J., van der Donk, W.A., 2010. Discovery of Unique Lanthionine Synthetases Reveals New Mechanistic and Evolutionary Insights. Public Libr. Sci. Biol. 8, e1000339. https://doi.org/10.1371/journal.pbio.1000339

Goto, Y., Ökesli, A., van der Donk, W.A., 2011. Mechanistic Studies of Ser/Thr Dehydration

Catalyzed by a Member of the LanL Lanthionine Synthetase Family. Biochemistry 50, 891–898. https://doi.org/10.1021/bi101750r

Götz, F., Perconti, S., Popella, P., Werner, R., Schlag, M., 2014. Epidermin and gallidermin: Staphylococcal lantibiotics. Int. J. Med. Microbiol. 304, 63–71. https://doi.org/10.1016/J.IJMM.2013.08.012

Gouda, H., Kobayashi, Y., Yamada, T., Ideguchi, T., Sugawara, A., Hirose, T., Ōmura, S., Sunazuka, T., Hirono, S., 2012. Three-Dimensional Solution Structure of Bottromycin A2: A Potent Antibiotic Active against Methicillin-Resistant Staphylococcus aureus and Vancomycin-Resistant Enterococci. Chem. Pharm. Bull. (Tokyo). 60, 169–171. https://doi.org/10.1248/cpb.60.169

Grese, R.P., Cerny, R.L., Gross, M.L., 1989. Metal ion-peptide interactions in the gas phase: a tandem mass spectrometry study of alkali metal cationized peptides. J. Am. Chem. Soc. 111, 2835–2842. https://doi.org/10.1021/ja00190a015

Gross, E., Kiltz, H.H., Nebelin, E., 1973. Subtilin, VI: structure of subtilin. Hoppe-Seyler's Z. Physiol. Chem. 354, 810–812.

Gross, E., Morell, J.L., 1971. Structure of nisin. J. Am. Chem. Soc. 93, 4634–4635. https://doi.org/10.1021/ja00747a073

Gu, W., Dong, S.-H., Sarkar, S., Nair, S.K., Schmidt, E.W., 2018. The Biochemistry and Structural Biology of Cyanobactin Pathways: Enabling Combinatorial Biosynthesis. Methods Enzymol. 604, 113–163. https://doi.org/10.1016/BS.MIE.2018.03.002

Gulmezian, M., Hyman, K.R., Marbois, B.N., Clarke, C.F., Javor, G.T., 2007. The role of UbiX in Escherichia coli coenzyme Q biosynthesis. Arch. Biochem. Biophys. 467, 144–153. https://doi.org/10.1016/j.abb.2007.08.009

Gupta, S., Marko, M.G., Miller, V.A., Schaefer, F.T., Anthony, J.R., Porter, J.R., 2015. Novel Production of Terpenoids in Escherichia coli and Activities Against Breast Cancer Cell Lines. Appl. Biochem. Biotechnol. 175, 2319–2331. https://doi.org/10.1007/s12010-014-1382-4

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. Proc. Natl. Acad. Sci. U. S. A. 100, 1541–6. https://doi.org/10.1073/pnas.0337542100

Haber, J.E., 2012. Mating-type genes and MAT switching in Saccharomyces cerevisiae. Genetics 191, 33–64. https://doi.org/10.1534/genetics.111.134577

Haft, D.H., Basu, M.K., Mitchell, D.A., 2010. Expansion of ribosomally produced natural products: a nitrile hydratase- and Nif11-related precursor family. BioMed Cent. Biol. 8, 70. https://doi.org/10.1186/1741-7007-8-70

Hall, G.E., Sheppard, N., Walker, J., 1966. Chemistry of micrococcin P. X. Proton magnetic resonance spectrum of dimethyl micrococcinate, and the probable mode of biosynthesis of micrococcinic acid. J. Chem. Soc. Perkin 1 16, 1371–1373.

Hall, J.A., 1933. Biogenetics in the Terpene Series. Chem. Rev. 13, 479–499. https://doi.org/10.1021/cr60046a002

Hamdane, D., Zhang, H., Hollenberg, P., 2008. Oxygen activation by cytochrome P450 monooxygenase. Photosynth. Res. 98, 657–666. https://doi.org/10.1007/s11120-008-9322-1

Hansen, J.N., Sandine, W.E., 1994. Nisin as a model food preservative. Crit. Rev. Food Sci. Nutr. 34, 69–93. https://doi.org/10.1080/10408399409527650

Hanson, S.J., Wolfe, K.H., 2017. An Evolutionary Perspective on Yeast Mating-Type Switching.

Genetics 206, 9–32. https://doi.org/10.1534/genetics.117.202036

Hao, Y., Pierce, E., Roe, D., Morita, M., McIntosh, J.A., Agarwal, V., Cheatham, T.E., Schmidt, E.W., Nair, S.K., 2016. Molecular basis for the broad substrate selectivity of a peptide prenyltransferase. Proc. Natl. Acad. Sci. U. S. A. 113, 14037–14042. https://doi.org/10.1073/pnas.1609869113

Harms, J.M., Wilson, D.N., Schluenzen, F., Connell, S.R., Stachelhaus, T., Zaborowska, Z., Spahn, C.M.T., Fucini, P., 2008. Translational Regulation via L11: Molecular Switches on the Ribosome Turned On and Off by Thiostrepton and Micrococcin. Mol. Cell 30, 26–38. https://doi.org/10.1016/j.molcel.2008.01.009

Hashimoto, T., Hashimoto, J., Teruya, K., Hirano, T., Shin-ya, K., Ikeda, H., Liu, H., Nishiyama, M., Kuzuyama, T., 2015. Biosynthesis of versipelostatin: identification of an enzyme-catalyzed [4+2]-cycloaddition required for macrocyclization of spirotetronate-containing polyketides. J. Am. Chem. Soc. 137, 572–5. https://doi.org/10.1021/ja510711x

Hasper, H.E., de Kruijff, B., Breukink, E., 2004. Assembly and Stability of Nisin–Lipid II Pores. Biochemistry 43, 11567–11575. https://doi.org/10.1021/BI049476B

Havarstein, L.S., Diep, D.B., Nes, I.F., 1995. A family of bacteriocin ABC transporters carry out proteolytic processing of their substrates concomitant with export. Mol. Microbiol. 16, 229–240. https://doi.org/10.1111/j.1365-2958.1995.tb02295.x

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. J. Antibiot. (Tokyo). 59, 1–5. https://doi.org/10.1038/ja.2006.1

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. https://doi.org/10.1038/ja.2006.2

Hayashi, S., Ozaki, T., Asamizu, S., Ikeda, H., Ōmura, S., Oku, N., Igarashi, Y., Tomoda, H., Onaka, H., 2014. Genome Mining Reveals a Minimum Gene Set for the Biosynthesis of 32-Membered Macrocyclic Thiopeptides Lactazoles. Chem. Biol. 21, 679–688. https://doi.org/10.1016/J.CHEMBIOL.2014.03.008

He, X., Ortiz De Montellano, P.R., 2004. Radical rebound mechanism in cytochrome P-450-catalyzed hydroxylation of the multifaceted radical clocks ??- and ??-thujone. J. Biol. Chem. 279, 39479–39484. https://doi.org/10.1074/jbc.M406838200

Heddle, J.G., Blance, S.J., Zamble, D.B., Hollfelder, F., Miller, D.A., Wentzell, L.M., Walsh, C.T., Maxwell, A., 2001. The antibiotic microcin B17 is a DNA gyrase poison: characterisation of the mode of inhibition. J. Mol. Biol. 307, 1223–1234. https://doi.org/10.1006/JMBI.2001.4562

Hegemann, J.D., van der Donk, W.A., 2018. Investigation of Substrate Recognition and Biosynthesis in Class IV Lanthipeptide Systems. J. Am. Chem. Soc. 140, 5743–5754. https://doi.org/10.1021/jacs.8b01323

Heidrich, K., Wirthmueller, L., Tasset, C., Pouzet, C., Deslandes, L., Parker, J.E., 2011. Arabidopsis EDS1 Connects Pathogen Effector Recognition to Cell Compartment-Specific Immune Responses. Science (80-.). 334, 1401–1404. https://doi.org/10.1126/science.1211641

Heneghan, M.N., Yakasai, A.A., Halo, L.M., Song, Z., Bailey, A.M., Simpson, T.J., Cox, R.J., Lazarus, C.M., 2010. First Heterologous Reconstruction of a Complete Functional Fungal Biosynthetic Multigene Cluster. ChemBioChem 11, 1508–1512. https://doi.org/10.1002/cbic.201000259

Hensens, O.D., Albers-Schönberg, G., 1983. Total structure of the highly modified peptide antibiotic components of thiopeptin. J. Antibiot. (Tokyo). 36, 814–831.

https://doi.org/10.7164/antibiotics.36.814

Herrera-Camacho, I., Rosas-Murrieta, N.H., Rojo-Domínguez, A., Millán, L., Reyes-Leyva, J., Santos-López, G., Suárez-Rendueles, P., 2007. Biochemical characterization and structural prediction of a novel cytosolic leucyl aminopeptidase of the M17 family from Schizosaccharomyces pombe. Fed. Eur. Biochem. Soc. J. 274, 6228–6240. https://doi.org/10.1111/j.1742-4658.2007.06142.x

Hertweck, C., 2009. The Biosynthetic Logic of Polyketide Diversity. Angew. Chemie Int. Ed. 48, 4688–4716. https://doi.org/10.1002/anie.200806121

Hetrick, K., J., Walker, M.C., Donk, W.A. van der, 2018. Development and Application of Yeast and Phage Display of Diverse Lanthipeptides. Am. Chem. Soc. Cent. Sci. 4, 458–467. https://doi.org/10.1021/ACSCENTSCI.7B00581

Himes, P.M., Allen, S.E., Hwang, S., Bowers, A.A., 2016. Production of Sactipeptides in Escherichia coli: Probing the Substrate Promiscuity of Subtilosin A Biosynthesis. Am. Chem. Soc. Chem. Biol. 11, 1737–1744. https://doi.org/10.1021/acschembio.6b00042

Hoff, G., Bertrand, C., Piotrowski, E., Thibessard, A., Leblond, P., 2017. Implication of RuvABC and RecG in homologous recombination in Streptomyces ambofaciens. Res. Microbiol. 168, 26–35. https://doi.org/10.1016/j.resmic.2016.07.003

Hon, W.-C., McKay, G.A., Thompson, P.R., Sweet, R.M., Yang, D.S.C., Wright, G.D., Berghuis, A.M., 1997. Structure of an Enzyme Required for Aminoglycoside Antibiotic Resistance Reveals Homology to Eukaryotic Protein Kinases. Cell 89, 887–895. https://doi.org/10.1016/S0092-8674(00)80274-3

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. J. Biol. Chem. 280, 13055–13061. https://doi.org/10.1074/jbc.M413801200

Horbal, L., Luzhetskyy, A., 2016. Dual control system – A novel scaffolding architecture of an inducible regulatory device for the precise regulation of gene expression. Metab. Eng. 37, 11–23. https://doi.org/10.1016/j.ymben.2016.03.008

Horton, R.M., Hunt, H.D., Ho, S.N., Pullen, J.K., Pease, L.R., 1989. Engineering hybrid genes without the use of restriction enzymes: gene splicing by overlap extension. Gene 77, 61–68. https://doi.org/10.1016/0378-1119(89)90359-4

Hou, Y., Tianero, M.D.B., Kwan, J.C., Wyche, T.P., Michel, C.R., Ellis, G.A., Vazquez-Rivera, E., Braun, D.R., Rose, W.E., Schmidt, E.W., Bugni, T.S., 2012. Structure and Biosynthesis of the Antibiotic Bottromycin D. Org. Lett. 14, 5050–5053. https://doi.org/10.1021/ol3022758

Hu, H., Zhang, Q., Ochi, K., 2002. Activation of antibiotic biosynthesis by specified mutations in the rpoB gene (encoding the RNA polymerase beta subunit) of Streptomyces lividans. J. Bacteriol. 184, 3984–3991. https://doi.org/10.1128/JB.184.14.3984-3991.2002

Huang, H., Zheng, G., Jiang, W., Hu, H., Lu, Y., 2015. One-step high-efficiency CRISPR/Cas9-mediated genome editing in Streptomyces. Acta Biochim. Biophys. Sin. (Shanghai). 47, 231–243. https://doi.org/10.1093/abbs/gmv007

Hudson, G.A., Zhang, Z., Tietz, J.I., Mitchell, D.A., van der Donk, W.A., 2015. In Vitro Biosynthesis of the Core Scaffold of the Thiopeptide Thiomuracin. J. Am. Chem. Soc. 137, 16012–16015. https://doi.org/10.1021/jacs.5b10194

Huijbers, M.M.E., Montersino, S., Westphal, A.H., Tischler, D., van Berkel, W.J.H., 2014. Flavin dependent monooxygenases. Arch. Biochem. Biophys. 544, 2–17. https://doi.org/10.1016/j.abb.2013.12.005

Huo, L., Rachid, S., Stadler, M., Wenzel, S.C., Müller, R., 2012. Synthetic Biotechnology to Study and Engineer Ribosomal Bottromycin Biosynthesis. Chem. Biol. 19, 1278–1287. https://doi.org/10.1016/j.chembiol.2012.08.013

Huo, L., van der Donk, W.A., 2016. Discovery and Characterization of Bicereucin, an Unusual d - Amino Acid-Containing Mixed Two-Component Lantibiotic. J. Am. Chem. Soc. 138, 5254– 5257. https://doi.org/10.1021/jacs.6b02513

Ichikawa, H., Bashiri, G., Kelly, W.L., 2018. Biosynthesis of the Thiopeptins and Identification of an F420H2-Dependent Dehydropiperidine Reductase. J. Am. Chem. Soc. 140, jacs.8b04238. https://doi.org/10.1021/jacs.8b04238

Igarashi, Y., Kan, Y., Fujii, K., Fujita, T., Harada, K.-I., Naoki, H., Tabata, H., Onaka, H., Furumai, T., 2001. Goadsporin, a Chemical Substance which Promotes Secondary Metabolism and Morphogenesis in Streptomycetes. II. Structure Determination. J. Antibiot. (Tokyo). 54, 1045– 1053. https://doi.org/10.7164/antibiotics.54.1045

Ikeda, H., Ishikawa, J., Hanamoto, A., Shinose, M., Kikuchi, H., Shiba, T., Sakaki, Y., Hattori, M., Ōmura, S., 2003. Complete genome sequence and comparative analysis of the industrial microorganism Streptomyces avermitilis. Nat. Biotechnol. 21, 526–531. https://doi.org/10.1038/nbt820

Ingram, L., 1970. A ribosomal mechanism for synthesis of peptides related to nisin. Biochim. Biophys. Acta - Nucleic Acids Protein Synth. 224, 263–265. https://doi.org/10.1016/0005-2787(70)90642-8

Iorio, M., Sasso, O., Maffioli, S.I., Bertorelli, R., Monciardini, P., Sosio, M., Bonezzi, F., Summa, M., Brunati, C., Bordoni, R., Corti, G., Tarozzo, G., Piomelli, D., Reggiani, A., Donadio, S., 2014. A Glycosylated, Labionin-Containing Lanthipeptide with Marked Antinociceptive Activity. Am. Chem. Soc. Chem. Biol. 9, 398–404. https://doi.org/10.1021/cb400692w

Ireland, C., Scheuer, P.J., 1980. Ulicyclamide and ulithiacyclamide, two new small peptides from a marine tunicate. J. Am. Chem. Soc. 102, 5688–5691. https://doi.org/10.1021/ja00537a053

Ishida, K., Matsuda, H., Murakami, M., Yamaguchi, K., 1997. Kawaguchipeptin B, an Antibacterial Cyclic Undecapeptide from the Cyanobacterium Microcystis aeruginosa. J. Nat. Prod. 60, 724– 726. https://doi.org/10.1021/NP970146K

Ishida, K., Nakagawa, H., Murakami, M., 2000. Microcyclamide, a cytotoxic cyclic hexapeptide from the cyanobacterium Microcystis aeruginosa. J. Nat. Prod. 63, 1315–1317. https://doi.org/10.1021/np000159p

Ishii, S., Yano, T., Ebihara, A., Okamoto, A., Manzoku, M., Hayashi, H., 2010. Crystal structure of the peptidase domain of Streptococcus ComA, a bifunctional ATP-binding cassette transporter involved in the quorum-sensing pathway. J. Biol. Chem. 285, 10777–10785. https://doi.org/10.1074/jbc.M109.093781

Itaya, M., Fujita, K., Kuroki, A., Tsuge, K., 2008. Bottom-up genome assembly using the Bacillus subtilis genome vector. Nat. Methods 5, 41–43. https://doi.org/10.1038/nmeth1143

Izawa, M., Nagamine, S., Aoki, H., Hayakawa, Y., 2017. Identification of essential biosynthetic genes and a true biosynthetic product for thioviridamide. J. Gen. Appl. Microbiol. 64, 50–53. https://doi.org/10.2323/jgam.2017.05.002

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. J. Antibiot. (Tokyo). 68, 533–536. https://doi.org/10.1038/ja.2015.20

Jacobus, A.P., Gross, J., 2015. Optimal Cloning of PCR Fragments by Homologous Recombination in Escherichia coli. PLoS One 10, e0119221. https://doi.org/10.1371/journal.pone.0119221

Janjusevic, R., Quezada, C.M., Small, J., Erec Stebbins, C., 2013. Structure of the HopA1(21-102)-ShcA chaperone-effector complex of Pseudomonas syringae reveals conservation of a virulence factor binding motif from animal to plant pathogens. J. Bacteriol. 195, 658–664. https://doi.org/10.1128/JB.01621-12

Jia, H., Zhang, L., Wang, T., Han, J., Tang, H., Zhang, L., 2017. Development of a CRISPR/Cas9-mediated gene-editing tool in Streptomyces rimosus. Microbiology 163, 1148–1155. https://doi.org/10.1099/mic.0.000501

Jiang, D., Zhu, W., Wang, Y., Sun, C., Zhang, K.-Q., Yang, J., 2013. Molecular tools for functional genomics in filamentous fungi: Recent advances and new strategies. Biotechnol. Adv. 31, 1562–1574. https://doi.org/10.1016/J.BIOTECHADV.2013.08.005

Jiang, W., Zhao, X., Gabrieli, T., Lou, C., Ebenstein, Y., Zhu, T.F., 2015. Cas9-Assisted Targeting of CHromosome segments CATCH enables one-step targeted cloning of large gene clusters. Nat. Commun. 6, 8101. https://doi.org/10.1038/ncomms9101

Jin, P., Ding, W., Du, G., Chen, J., Kang, Z., 2016. DATEL: A Scarless and Sequence-Independent DNA Assembly Method Using Thermostable Exonucleases and Ligase. Am. Chem. Soc. Synth. Biol. 5, 1028–1032. https://doi.org/10.1021/acssynbio.6b00078

Jungmann, N.A., Krawczyk, B., Tietzmann, M., Ensle, P., Süssmuth, R.D., 2014. Dissecting Reactions of Nonlinear Precursor Peptide Processing of the Class III Lanthipeptide Curvopeptin. J. Am. Chem. Soc. 136, 15222–15228. https://doi.org/10.1021/ja5062054

Kallifidas, D., Brady, S.F., 2012. Reassembly of functionally intact environmental DNA-derived biosynthetic gene clusters. Methods Enzymol. 517, 225–239. https://doi.org/10.1016/B978-0-12-404634-4.00011-5

Kalyanaraman, B., Cheng, G., Hardy, M., Ouari, O., Lopez, M., Joseph, J., Zielonka, J., Dwinell, M.B., 2018. A review of the basics of mitochondrial bioenergetics, metabolism, and related signaling pathways in cancer cells: Therapeutic targeting of tumor mitochondria with lipophilic cationic compounds. Redox Biol. 14, 316–327. https://doi.org/10.1016/j.redox.2017.09.020

Kalyon, B., Helaly, S.E., Scholz, R., Nachtigall, J., Vater, J., Borriss, R., Süssmuth, R.D., 2011. Plantazolicin A and B: Structure Elucidation of Ribosomally Synthesized Thiazole/Oxazole Peptides from Bacillus amyloliquefaciens FZB42. Org. Lett. 13, 2996–2999. https://doi.org/10.1021/ol200809m

Kaneda, M., 1992. Studies on bottromycins. I. 1H and 13C NMR assignments of bottromycin A2, the main component of the complex. J. Antibiot. (Tokyo). 45, 792–796. https://doi.org/10.7164/antibiotics.45.792

Kang, H.-S., Charlop-Powers, Z., Brady, S.F., 2016. Multiplexed CRISPR/Cas9- and TAR-Mediated Promoter Engineering of Natural Product Biosynthetic Gene Clusters in Yeast. ACS Synth. Biol. 5, 1002–1010. https://doi.org/10.1021/acssynbio.6b00080

Katz, L., Baltz, R.H., 2016. Natural product discovery: past, present, and future. J. Ind. Microbiol. Biotechnol. 43, 155–176. https://doi.org/10.1007/s10295-015-1723-5

Katzen, F., Ferreiro, D.U., Oddo, C.G., Ielmini, M.V., Becker, A., Pühler, A., Ielpi, L., 1998. Xanthomonas campestris pv. campestris gum mutants: Effects on xanthan biosynthesis and plant virulence. J. Bacteriol. 180, 1607–1617.

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. J. Nat. Prod. 81, 264–269. https://doi.org/10.1021/acs.jnatprod.7b00607

Kawai, S., Murata, K., 2015. Transformation of Saccharomyces cerevisiae: Spheroplast Method, in: Genetic Transformation Systems in Fungi, Volume 1. Springer, Cham, pp. 61–63. https://doi.org/10.1007/978-3-319-10142-2

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. Nat. Protoc. 10, 845–58. https://doi.org/10.1038/nprot.2015.053

Kellner, R., Jung, G., Josten, M., Kaletta, C., Entian, K.-D., Sahl, H.-G., 1989. Pep5: Structure Elucidation of a Large Lantibiotic. Angew. Chemie Int. Ed. English 28, 616–619. https://doi.org/10.1002/anie.198906161

Kelly, W.L., Pan, L., Li, C., 2009. Thiostrepton Biosynthesis: Prototype for a New Family of Bacteriocins. J. Am. Chem. Soc. 131, 4327–4334. https://doi.org/10.1021/ja807890a

Kenney, G.E., Dassama, L.M.K., Pandelia, M.E., Gizzi, A.S., Martinie, R.J., Gao, P., DeHart, C.J., Schachner, L.F., Skinner, O.S., Ro, S.Y., Zhu, X., Sadek, M., Thomas, P.M., Almo, S.C., Bollinger, J.M., Krebs, C., Kelleher, N.L., Rosenzweig, A.C., 2018. The biosynthesis of methanobactin. Science (80-.). 359, 1411–1416. https://doi.org/10.1126/science.aap9437

Kenney, G.E., Rosenzweig, A.C., 2013. Genome mining for methanobactins. BMC Biol. 11, 17. https://doi.org/10.1186/1741-7007-11-17

Kessler, H., Haessner, R., Schüler, W., 1993. Structure of Rapamycin: An NMR and Molecular-Dynamics Investigation. Helv. Chim. Acta 76, 117–130. https://doi.org/10.1002/hlca.19930760106

Kieser, T., Bibb, M.J., Buttner, M.J., Chater, K.F., Hopwood, D.A., 2000. Practical Streptomyces Genetics, John Innes Centre Ltd. https://doi.org/10.4016/28481.01

Kim, H.J., Graham, D.W., DiSpirito, A.A., Alterman, M.A., Galeva, N., Larive, C.K., Asunskis, D., Sherwood, P.M.A., 2004. Methanobactin, a Copper-Acquisition Compound from Methane-Oxidizing Bacteria. Science (80-.). 305, 1612–1615. https://doi.org/10.1126/science.1098322

Kim, H.J., McCarty, R.M., Ogasawara, Y., Liu, Y., Mansoorabadi, S.O., LeVieux, J., Liu, H., 2013. GenK-catalyzed C-6' methylation in the biosynthesis of gentamicin: isolation and characterization of a cobalamin-dependent radical SAM enzyme. J. Am. Chem. Soc. 135, 8093–6. https://doi.org/10.1021/ja312641f

Kim, H.J., Ruszczycky, M.W., Choi, S., Liu, Y., Liu, H., 2011. Enzyme-catalysed [4+2] cycloaddition is a key step in the biosynthesis of spinosyn A. Nature 473, 109–12. https://doi.org/10.1038/nature09981

Kim, J.H., Feng, Z., Bauer, J.D., Kallifidas, D., Calle, P.Y., Brady, S.F., 2010. Cloning large natural product gene clusters from the environment: piecing environmental DNA gene clusters back together with TAR. Biopolymers 93, 833–844. https://doi.org/10.1002/bip.21450

Kim, S., Kim, K.J., 2011. Cloning, expression, purification, crystallization and X-ray crystallographic analysis of Rv2606c from Mycobacterium tuberculosis H37Rv. Acta Crystallogr. Sect. F. Struct. Biol. Cryst. Commun. 69, 578–80. https://doi.org/10.1107/S1744309113010683

Kim, S., Nguyen, C.M.T., Kim, E.-J., Kim, K.-J., 2011. Crystal structure of Mycobacterium tuberculosis Rv3168: A putative aminoglycoside antibiotics resistance enzyme. Proteins Struct. Funct. Bioinforma. 79, 2983–2987. https://doi.org/10.1002/prot.23119

Kim, S.H., Kwon, S. Il, Saha, D., Anyanwu, N.C., Gassmann, W., 2009. Resistance to the

Pseudomonas syringae effector HopA1 is governed by the TIR-NBS-LRR protein RPS6 and is enhanced by mutations in SRFR1. Plant Physiol. 150, 1723–32. https://doi.org/10.1104/pp.109.139238

Kjaerulff, L., Sikandar, A., Zaburannyi, N., Adam, S., Herrmann, J., Koehnke, J., Müller, R., 2017. Thioholgamides: Thioamide-Containing Cytotoxic RiPP Natural Products. Am. Chem. Soc. Chem. Biol. 12, 2837–2841. https://doi.org/10.1021/acschembio.7b00676

Kobayashi, Y., Ichioka, M., Hirose, T., Nagai, K., Matsumoto, A., Matsui, H., Hanaki, H., Masuma, R., Takahashi, Y., Ōmura, S., Sunazuka, T., 2010. Bottromycin derivatives: Efficient chemical modifications of the ester moiety and evaluation of anti-MRSA and anti-VRE activities. Bioorg. Med. Chem. Lett. 20, 6116–6120. https://doi.org/10.1016/j.bmcl.2010.08.037

Kodani, S., Hudson, M.E., Durrant, M.C., Buttner, M.J., Nodwell, J.R., Willey, J.M., 2004. From The Cover: The SapB morphogen is a lantibiotic-like peptide derived from the product of the developmental gene ramS in Streptomyces coelicolor. Proc. Natl. Acad. Sci. 101, 11448–11453. https://doi.org/10.1073/pnas.0404220101

Koehnke, J., Bent, A.F., Houssen, W.E., Mann, G., Jaspars, M., Naismith, J.H., 2014. The structural biology of patellamide biosynthesis. Curr. Opin. Struct. Biol. 29, 112–121. https://doi.org/10.1016/J.SBI.2014.10.006

Koehnke, J., Mann, G., Bent, A.F., Ludewig, H., Shirran, S., Botting, C., Lebl, T., Houssen, W.E., Jaspars, M., Naismith, J.H., 2015. Structural analysis of leader peptide binding enables leader-free cyanobactin processing. Nat. Chem. Biol. 11, 558–563. https://doi.org/10.1038/nchembio.1841

Kosuri, S., Church, G.M., 2014. Large-scale de novo DNA synthesis: technologies and applications. Nat. Methods 11, 499–507. https://doi.org/10.1038/nmeth.2918

Kouprina, N., Annab, L., Graves, J., Afshari, C., Barrett, J.C., Resnick, M.A., Larionov, V., 1998. Functional copies of a human gene can be directly isolated by transformation-associated recombination cloning with a small 3′ end target sequence. Proc. Natl. Acad. Sci. 95, 4469–4474. https://doi.org/10.1073/pnas.95.8.4469

Kouprina, N., Larionov, V., 2016. Transformation-associated recombination (TAR) cloning for genomics studies and synthetic biology. Chromosoma 125, 621–32. https://doi.org/10.1007/s00412-016-0588-3

Krawczyk, B., Ensle, P., Müller, W.M., Süssmuth, R.D., 2012a. Deuterium Labeled Peptides Give Insights into the Directionality of Class III Lantibiotic Synthetase LabKC. J. Am. Chem. Soc. 134, 9922–9925. https://doi.org/10.1021/ja3040224

Krawczyk, B., Völler, G.H., Völler, J., Ensle, P., Süssmuth, R.D., 2012b. Curvopeptin: A New Lanthionine-Containing Class III Lantibiotic and its Co-substrate Promiscuous Synthetase. ChemBioChem 13, 2065–2071. https://doi.org/10.1002/cbic.201200417

Kuipers, O.P., Rollema, H.S., Beerthuyzen, M.M., Siezen, R.J., de Vos, W.M., 1995. Protein engineering and biosynthesis of nisin and regulation of transcription of the structural nisA gene. Int. Dairy J. 5, 785–795. https://doi.org/10.1016/0958-6946(95)00032-1

Kumari, S., Pundhir, S., Priya, P., Jeena, G., Punetha, A., Chawla, K., Firdos Jafaree, Z., Mondal, S., Yadav, G., 2014. EssOilDB: a database of essential oils reflecting terpene composition and variability in the plant kingdom. Database (Oxford). 2014, bau120. https://doi.org/10.1093/database/bau120

Kupke, T., Götz, F., 1997. The enethiolate anion reaction products of EpiD: pK(α) value of the enethiol side chain is lower than that of the thiol side chain of peptides. J. Biol. Chem. 272, 4759–4762. https://doi.org/10.1074/jbc.272.8.4759

Kupke, T., Kempter, C., Jung, G., Götz, F., 1995. Oxidative decarboxylation of peptides catalyzed by flavoprotein EpiD. Determination of substrate specificity using peptide libraries and neutral loss mass spectrometry. J. Biol. Chem. 270, 11282–11289. https://doi.org/10.1074/JBC.270.19.11282

Kupke, T., Stevanović, S., Sahl, H.G., Götz, F., 1992. Purification and characterization of EpiD, a flavoprotein involved in the biosynthesis of the lantibiotic epidermin. J. Bacteriol. 174, 5354–61. https://doi.org/10.1128/JB.174.16.5354-5361.1992

Kuthning, A., Durkin, P., Oehm, S., Hoesl, M.G., Budisa, N., Süssmuth, R.D., 2016. Towards Biocontained Cell Factories: An Evolutionarily Adapted Escherichia coli Strain Produces a New-to-nature Bioactive Lantibiotic Containing Thienopyrrole-Alanine. Sci. Rep. 6, 33447. https://doi.org/10.1038/srep33447

Kuthning, A., Mösker, E., Süssmuth, R.D., 2015. Engineering the heterologous expression of lanthipeptides in Escherichia coli by multigene assembly. Appl. Microbiol. Biotechnol. 99, 6351–6361. https://doi.org/10.1007/s00253-015-6557-6

Kvitko, B.H., McMillan, I.A., Schweizer, H.P., 2013. An improved method for oriT-directed cloning and functionalization of large bacterial genomic regions. Appl. Environ. Microbiol. 79, 4869–4878. https://doi.org/10.1128/AEM.00994-13

Kwak, J., Jiang, H., Kendrick, K.E., 2006. Transformation using in vivo and in vitro methylation in Streptomyces griseus. Fed. Eur. Microbiol. Soc. Microbiol. Lett. 209, 243–248. https://doi.org/10.1111/j.1574-6968.2002.tb11138.x

Labes, G., Bibb, M., Wohlleben, W., 1997. Isolation and characterization of a strong promoter element from the Streptomyces ghanaensis phage 119 using the gentamicin resistance gene (aacC1) of Tn1696 as reporter. Microbiology 143, 1503–1512. https://doi.org/10.1099/00221287-143-5-1503

Larsson, F.C. V., Lawesson, S.O., Møller, J., Schroll, G., 1973. Mass Spectra of Thioamides. Acta Chem. Scand. 27, 747–755. https://doi.org/10.3891/acta.chem.scand.27-0747

Lazarus, C.M., Williams, K., Bailey, A.M., 2014. Reconstructing fungal natural product biosynthetic pathways. Nat. Prod. Rep. 31, 1339–1347. https://doi.org/10.1039/C4NP00084F

Lebedenko, E.N., Birikh, K.R., Plutalov, O.V., Berlin, Y.A., 1991. Method of artificial DNA splicing by directed ligation (SDL). Nucleic Acids Res. 19, 6757–6761. https://doi.org/10.1093/nar/19.24.6757

Lee, J., Hao, Y., Blair, P.M., Melby, J.O., Agarwal, V., Burkhart, B.J., Nair, S.K., Mitchell, D.A., 2013. Structural and functional insight into an unexpectedly selective N-methyltransferase involved in plantazolicin biosynthesis. Proc. Natl. Acad. Sci. 110, 12954–12959. https://doi.org/10.1073/pnas.1306101110

Lee, M.V., Ihnken, L.A.F., You, Y.O., McClerren, A.L., van der Donk, W.A., Kelleher, N.L., 2009. Distributive and directional behavior of lantibiotic synthetases revealed by high-resolution tandem mass spectrometry. J. Am. Chem. Soc. 131, 12258–12264. https://doi.org/10.1021/ja9033507

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 Res. 43, e55. https://doi.org/10.1093/nar/gkv112

Lee, S.G., Jez, J.M., 2013. Evolution of structure and mechanistic divergence in di-domain methyltransferases from nematode phosphocholine biosynthesis. Structure 21, 1778–87. https://doi.org/10.1016/j.str.2013.07.023

Leem, S.H., Noskov, V.N., Park, J.-E., Kim, S. Il, Larionov, V., Kouprina, N., 2003. Optimum conditions for selective isolation of genes from complex genomes by transformation-associated recombination cloning. Nucleic Acids Res. 31, 29e–29. https://doi.org/10.1093/nar/gng029

Leite, N.R., Faro, A.R., Dotta, M.A.O., Faim, L.M., Gianotti, A., Silva, F.H., Oliva, G., Thiemann, O.H., 2013. The crystal structure of the cysteine protease Xylellain from Xylella fastidiosa reveals an intriguing activation mechanism. Fed. Eur. Biochem. Soc. Lett. 587, 339–344. https://doi.org/10.1016/j.febslet.2013.01.009

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 Res. 44, W242–W245. https://doi.org/10.1093/nar/gkw290

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. Proc. Natl. Acad. Sci. U. S. A. 107, 10430–5. https://doi.org/10.1073/pnas.0913677107

Li, B., Walsh, C.T., 2010. Identification of the gene cluster for the dithiolopyrrolone antibiotic holomycin in Streptomyces clavuligerus. Proc. Natl. Acad. Sci. 107, 19731–19735. https://doi.org/10.1073/pnas.1014140107

Li, B., Yu, J.P.J., Brunzelle, J.S., Moll, G.N., Van Der Donk, W.A., Nair, S.K., 2006. Structure and mechanism of the lantibiotic cyclase involved in nisin biosynthesis. Science (80-.). 311, 1464–1467. https://doi.org/10.1126/science.1121422

Li, H., Xu, H., Zhou, Y., Zhang, J., Long, C., Li, S., Chen, S., Zhou, J.-M., Shao, F., 2007. The Phosphothreonine Lyase Activity of a Bacterial Type III Effector Family. Science (80-.). 315, 1000–1003. https://doi.org/10.1126/science.1138960

Li, M.Z., Elledge, S.J., 2007. Harnessing homologous recombination in vitro to generate recombinant DNA via SLIC. Nat. Methods 4, 251–256. https://doi.org/10.1038/nmeth1010

Li, X., Heyer, W.-D., 2008. Homologous recombination in DNA repair and DNA damage tolerance. Cell Res. 18, 99–113. https://doi.org/10.1038/cr.2008.1

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 (80-.). 274, 1188–1193. https://doi.org/10.1126/SCIENCE.274.5290.1188

Liang, J., Liu, Z., Low, X.Z., Ang, E.L., Zhao, H., 2017. Twin-primer non-enzymatic DNA assembly: an efficient and accurate multi-part DNA assembly method. Nucleic Acids Res. 45, e94–e94. https://doi.org/10.1093/nar/gkx132

Liao, R., Duan, L., Lei, C., Pan, H., Ding, Y., Zhang, Q., Chen, D., Shen, B., Yu, Y., Liu, W., 2009. Thiopeptide Biosynthesis Featuring Ribosomally Synthesized Precursor Peptides and Conserved Posttranslational Modifications. Chem. Biol. 16, 141–147. https://doi.org/10.1016/j.chembiol.2009.01.007

Liesch, J.M., Rinehart, K.L., 1977. Berninamycin. 3. Total structure of berninamycin A. J. Am. Chem. Soc. 99, 1645–1646. https://doi.org/10.1021/ja00447a061

Liscombe, D.K., Louie, G. V., Noel, J.P., 2012. Architectures, mechanisms and molecular evolution of natural product methyltransferases. Nat. Prod. Rep. 29, 1238. https://doi.org/10.1039/c2np20029e

Liu, C.J., Jiang, H., Wu, L., Zhu, L.Y., Meng, E., Zhang, D.Y., 2017. OEPR cloning: An efficient and seamless cloning strategy for large- and multi-fragments. Sci. Rep. 7, 44648. https://doi.org/10.1038/srep44648

Liu, G., Ou, H.-Y., Wang, T., Li, L., Tan, H., Zhou, X., Rajakumar, K., Deng, Z., He, X., 2010. Cleavage of phosphorothioated DNA and methylated DNA by the type IV restriction endonuclease ScoMcrA. Public Libr. Sci. One Genet. 6, e1001253. https://doi.org/10.1371/journal.pgen.1001253

Liu, P., Jenkins, N.A., Copeland, N.G., 2003. A highly efficient recombineering-based method for generating conditional knockout mutations. Genome Res. 13, 476–84. https://doi.org/10.1101/gr.749203

Liu, W., Xue, Y., Ma, M., Wang, S., Liu, N., Chen, Y., 2013. Multiple Oxidative Routes towards the Maturation of Nosiheptide. ChemBioChem 14, 1544–1547. https://doi.org/10.1002/cbic.201300427

Liu, Y., Tao, W., Wen, S., Li, Z., Yang, A., Deng, Z., Sun, Y., 2015. In Vitro CRISPR/Cas9 System for Efficient Targeted DNA Editing. MBio 6, e01714-15. https://doi.org/10.1128/mBio.01714-15

Lohans, C.T., Li, J.L., Vederas, J.C., 2014. Structure and Biosynthesis of Carnolysin, a Homologue of Enterococcal Cytolysin with d -Amino Acids. J. Am. Chem. Soc. 136, 13150–13153. https://doi.org/10.1021/ja5070813

Lopatniuk, M., Myronovskyi, M., Luzhetskyy, A., 2017. Streptomyces albus : A New Cell Factory for Non-Canonical Amino Acids Incorporation into Ribosomally Synthesized Natural Products. Am. Chem. Soc. Chem. Biol. 12, 2362–2370. https://doi.org/10.1021/acschembio.7b00359

Lopez, J.A. V., Al-Lihaibi, S.S., Alarif, W.M., Abdel-Lateff, A., Nogata, Y., Washio, K., Morikawa, M., Okino, T., 2016. Wewakazole B, a Cytotoxic Cyanobactin from the Cyanobacterium *Moorea producens* Collected in the Red Sea. J. Nat. Prod. 79, 1213–1218. https://doi.org/10.1021/acs.jnatprod.6b00051

Lubelski, J., Khusainov, R., Kuipers, O.P., 2009. Directionality and coordination of dehydration and ring formation during biosynthesis of the lantibiotic nisin. J. Biol. Chem. 284, 25962–25972. https://doi.org/10.1074/jbc.M109.026690

Lubelski, J., Rink, R., Khusainov, R., Moll, G.N., Kuipers, O.P., 2008. Biosynthesis, immunity, regulation, mode of action and engineering of the model lantibiotic nisin. Cell. Mol. Life Sci. 65, 455–476. https://doi.org/10.1007/s00018-007-7171-2

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. Nat. Commun. 4, 2894. https://doi.org/10.1038/ncomms3894

Luo, Y., Li, B.-Z., Liu, D., Zhang, L., Chen, Y., Jia, B., Zeng, B.-X., Zhao, H., Yuan, Y.-J., 2015. Engineered biosynthesis of natural products in heterologous hosts. Chem. Soc. Rev. 44, 5265–5290. https://doi.org/10.1039/C5CS00025D

Ma, H., Gao, Y., Zhao, F., Wang, J., Teng, K., Zhang, J., Zhong, J., 2014. Dissecting the catalytic and substrate binding activity of a class II lanthipeptide synthetase BovM. Biochem. Biophys. Res. Commun. 450, 1126–1132. https://doi.org/10.1016/j.bbrc.2014.06.129

Ma, H., Gao, Y., Zhao, F., Zhong, J., 2015. Individual catalytic activity of two functional domains of bovicin HJ50 synthase BovM. Wei Sheng Wu Xue Bao 55, 50–58.

Ma, Z., Liu, J., Shentu, X., Bian, Y., Yu, X., 2014. Optimization of electroporation conditions for toyocamycin producer *Streptomyces diastatochromogenes* 1628. J. Basic Microbiol. 54, 278–284. https://doi.org/10.1002/jobm.201200489

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, 61–68. https://doi.org/10.1016/0378-1119(92)90603-M

Mahanta, N., Hudson, G.A., Mitchell, D.A., 2017a. Radical *S* -Adenosylmethionine Enzymes Involved
    in RiPP Biosynthesis. Biochemistry 56, 5229–5244.
    https://doi.org/10.1021/acs.biochem.7b00771

Mahanta, N., Liu, A., Dong, S., Nair, S.K., Mitchell, D.A., 2018. Enzymatic reconstitution of ribosomal
    peptide backbone thioamidation. Proc. Natl. Acad. Sci. 115, 3030–3035.
    https://doi.org/10.1073/pnas.1722324115

Mahanta, N., Zhang, Z., Hudson, G.A., van der Donk, W.A., Mitchell, D.A., 2017b. Reconstitution
    and Substrate Specificity of the Radical *S* -Adenosyl-methionine Thiazole *C* -Methyltransferase
    in Thiomuracin Biosynthesis. J. Am. Chem. Soc. 139, 4310–4313.
    https://doi.org/10.1021/jacs.7b00693

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. J. Bacteriol. 184, 1234–43.
    https://doi.org/10.1128/JB.184.5.1234-1243.2002

Mann, G., Huo, L., Adam, S., Nardone, B., Vendome, J., Westwood, N.J., Müller, R., Koehnke, J.,
    2016. Structure and Substrate Recognition of the Bottromycin Maturation Enzyme BotP.
    ChemBioChem 17, 2286–2292. https://doi.org/10.1002/cbic.201600406

Marmorek, A., 1902. La Toxine Streptococcique. Ann. Inst. Pasteur (Paris). 16, 169–178.

Marsh, A.J., O'Sullivan, O., Ross, R.P., Cotter, P.D., Hill, C., 2010. In silico analysis highlights the
    frequency and diversity of type 1 lantibiotic gene clusters in genome sequenced bacteria.
    BioMed Cent. Genomics 11, 679. https://doi.org/10.1186/1471-2164-11-679

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. https://doi.org/10.1021/bi0362065

Martin, V.J.J., Pitera, D.J., Withers, S.T., Newman, J.D., Keasling, J.D., 2003. Engineering a
    mevalonate pathway in Escherichia coli for production of terpenoids. Nat. Biotechnol. 21,
    796–802. https://doi.org/10.1038/nbt833

Matsui, M., Fowler, J.H., Walling, L.L., 2006. Leucine aminopeptidases: diversity in structure and
    function. Biol. Chem. 387, 1535–1544. https://doi.org/10.1515/BC.2006.191

Mazy-Servais, C., Baczkowski, D., Dusart, J., 1997. Electroporation of intact cells of Streptomyces
    parvulus and Streptomyces vinaceus. Fed. Eur. Microbiol. Soc. Microbiol. Lett. 151, 135–138.
    https://doi.org/10.1111/j.1574-6968.1997.tb12561.x

McClerren, A.L., Cooper, L.E., Quan, C., Thomas, P.M., Kelleher, N.L., van der Donk, W.A., 2006.
    Discovery and in vitro biosynthesis of haloduracin, a two-component lantibiotic. Proc. Natl.
    Acad. Sci. U. S. A. 103, 17243–17248. https://doi.org/10.1073/pnas.0606088103

McIntosh, J.A., Donia, M.S., Schmidt, E.W., 2010. Insights into Heterocyclization from Two Highly
    Similar Enzymes. J. Am. Chem. Soc. 132, 4089–4091. https://doi.org/10.1021/ja9107116

McIntosh, J.A., Schmidt, E.W., 2010. Marine Molecular Machines: Heterocyclization in Cyanobactin
    Biosynthesis. ChemBioChem 11, 1413–1421. https://doi.org/10.1002/cbic.201000196

Meganathan, R., 2001. Ubiquinone biosynthesis in microorganisms. Fed. Eur. Microbiol. Soc.
    Microbiol. Lett. 203, 131–139. https://doi.org/10.1016/S0378-1097(01)00330-5

Meindl, K., Schmiederer, T., Schneider, K., Reicke, A., Butz, D., Keller, S., Gühring, H., Vértesy, L.,
    Wink, J., Hoffmann, H., Brönstrup, M., Sheldrick, G.M., Süssmuth, R.D., 2010.
    Labyrinthopeptins: A New Class of Carbacyclic Lantibiotics. Angew. Chemie Int. Ed. 49, 1151–
    1154. https://doi.org/10.1002/anie.200905773

Melby, J.O., Dunbar, K.L., Trinh, N.Q., Mitchell, D.A., 2012. Selectivity, Directionality, and Promiscuity in Peptide Processing from a *Bacillus* sp. Al Hakam Cyclodehydratase. J. Am. Chem. Soc. 134, 5309–5316. https://doi.org/10.1021/ja211675n

Melby, J.O., Nard, N.J., Mitchell, D.A., 2011. Thiazole/oxazole-modified microcins: Complex natural products from ribosomal templates. Curr. Opin. Chem. Biol. 15, 369–378. https://doi.org/10.1016/j.cbpa.2011.02.027

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. Nat. Chem. Biol. 13, 1129–1136. https://doi.org/10.1038/nchembio.2462

Metelev, M., Serebryakov, M., Ghilarov, D., Zhao, Y., Severinov, K., 2013. Structure of microcin B-like compounds produced by pseudomonas syringae and species specificity of their antibacterial action. J. Bacteriol. 195, 4129–4137. https://doi.org/10.1128/JB.00665-13

Miller, J.C., Holmes, M.C., Wang, J., Guschin, D.Y., Lee, Y.-L., Rupniewski, I., Beausejour, C.M., Waite, A.J., Wang, N.S., Kim, K.A., Gregory, P.D., Pabo, C.O., Rebar, E.J., 2007. An improved zinc-finger nuclease architecture for highly specific genome editing. Nat. Biotechnol. 25, 778–785. https://doi.org/10.1038/nbt1319

Miller, M.A., Pfeiffer, W., Schwartz, T., 2010. Creating the CIPRES Science Gateway for inference of large phylogenetic trees, in: 2010 Gateway Computing Environments Workshop (GCE). IEEE, pp. 1–8. https://doi.org/10.1109/GCE.2010.5676129

Miller, W., Chaiet, L., Rasmussen, G., Christensen, B., Hannah, J., Miller, A.K., Wolfe, F., 1968. Bottromycin. Separation of Biologically Active Compounds and Preparation and Testing of Amide Derivatives. J. Med. Chem. 11, 746–749. https://doi.org/10.1021/jm00310a603

Milne, B.F., Long, P.F., Starcevic, A., Hranueli, D., Jaspars, M., 2006. Spontaneity in the patellamide biosynthetic pathway. Org. Biomol. Chem. 4, 631–638. https://doi.org/10.1039/b515938e

Milne, J.C., Eliot, A.C., Kelleher, N.L., Walsh, C.T., 1998. ATP/GTP Hydrolysis Is Required for Oxazole and Thiazole Biosynthesis in the Peptide Antibiotic Microcin B17. Biochemistry 22, 13250–13261. https://doi.org/10.1021/BI980996E

Milne, J.C., Roy, R.S., Eliot, A.C., Kelleher, N.L., Wokhlu, A., Nickels, B., Walsh, C.T., 1999. Cofactor Requirements and Reconstitution Of Microcin B17 Synthetase: A Multienzyme Complex that Catalyzes the Formation of Oxazoles and Thiazoles in the Antibiotic Microcin B17. Biochemistry 13, 4768–4781. https://doi.org/10.1021/BI982975Q

Minami, Y., Yoshida, K., Azuma, R., Urakawa, A., Kawauchi, T., Otani, T., Komiyama, K., Ōmura, S., 1994. Structure of cypemycin, a new peptide antibiotic. Tetrahedron Lett. 35, 8001–8004. https://doi.org/10.1016/0040-4039(94)80033-2

Mitchell, L.A., Chuang, J., Agmon, N., Khunsriraksakul, C., Phillips, N.A., Cai, Y., Truong, D.M., Veerakumar, A., Wang, Y., Mayorga, M., Blomquist, P., Sadda, P., Trueheart, J., Boeke, J.D., 2015. Versatile genetic assembly system (VEGAS) to assemble pathways for expression in S. cerevisiae. Nucleic Acids Res. 43, 6620–6630. https://doi.org/10.1093/nar/gkv466

Mitra, A., Kesarwani, A.K., Pal, D., Nagaraja, V., 2011. WebGeSTer DB-A transcription terminator database. Nucleic Acids Res. 39, D129–D135. https://doi.org/10.1093/nar/gkq971

Mocek, U., Knaggs, A.R., Tsuchiya, R., Nguyen, T., Beale, J.M., Floss, H.G., 1993a. Biosynthesis of the modified peptide antibiotic nosiheptide in Streptomyces actuosus. J. Am. Chem. Soc. 115, 7557–7568. https://doi.org/10.1021/ja00070a001

Mocek, U., Zeng, Z., O'Hagan, D., Zhou, P., Fan, L.D.G., Beale, J.M., Floss, H.G., 1993b. Biosynthesis of the modified peptide antibiotic thiostrepton in Streptomyces azureus and Streptomyces laurentii. J. Am. Chem. Soc. 115, 7992–8001. https://doi.org/10.1021/ja00071a009

Mohr, K.I., Volz, C., Jansen, R., Wray, V., Hoffmann, J., Bernecker, S., Wink, J., Gerth, K., Stadler, M., Müller, R., 2015. Pinensins: The First Antifungal Lantibiotics. Angew. Chemie Int. Ed. 54, 11254–11258. https://doi.org/10.1002/anie.201500927

Molloy, E.M., Cotter, P.D., Hill, C., Mitchell, D.A., Ross, R.P., 2011. Streptolysin S-like virulence factors: the continuing sagA. Nat. Rev. Microbiol. 9, 670–681. https://doi.org/10.1038/nrmicro2624

Montiel, D., Kang, H.-S., Chang, F.-Y., Charlop-Powers, Z., Brady, S.F., 2015. Yeast homologous recombination-based promoter engineering for the activation of silent natural product biosynthetic gene clusters. Proc. Natl. Acad. Sci. 112, 8953–8958. https://doi.org/10.1073/pnas.1507606112

Morinaka, B.I., Vagstad, A.L., Helf, M.J., Gugger, M., Kegler, C., Freeman, M.F., Bode, H.B., Piel, J., 2014. Radical S -Adenosyl Methionine Epimerases: Regioselective Introduction of Diverse D - Amino Acid Patterns into Peptide Natural Products. Angew. Chemie Int. Ed. 53, 8503–8507. https://doi.org/10.1002/anie.201400478

Morris, R.P., Leeds, J.A., Naegeli, H.U., Oberer, L., Memmert, K., Weber, E., LaMarche, M.J., Parker, C.N., Burrer, N., Esterow, S., Hein, A.E., Schmitt, E.K., Krastel, P., 2009. Ribosomally Synthesized Thiopeptide Antibiotics Targeting Elongation Factor Tu. J. Am. Chem. Soc. 131, 5946–5955. https://doi.org/10.1021/ja900488a

Müller, W.M., Schmiederer, T., Ensle, P., Süssmuth, R.D., 2010. In Vitro Biosynthesis of the Prepeptide of Type-III Lantibiotic Labyrinthopeptin A2 Including Formation of a C⬚C Bond as a Post-Translational Modification. Angew. Chemie Int. Ed. 49, 2436–2440. https://doi.org/10.1002/anie.200905909

Mullis, K.B., Faloona, F.A., 1987. Specific Synthesis of DNA in Vitro via a Polymerase-Catalyzed Chain Reaction. Methods Enzymol. 155, 335–350. https://doi.org/10.1016/0076-6879(87)55023-6

Nagy, I., Banerjee, T., Tamura, T., Schoofs, G., Gils, A., Proost, P., Tamura, N., Baumeister, W., De Mot, R., 2003. Characterization of a novel intracellular endopeptidase of the alpha/beta hydrolase family from Streptomyces coelicolor A3(2). J. Bacteriol. 185, 496–503. https://doi.org/10.1128/JB.185.2.496-503.2003

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. Front. Microbiol. 8, 394. https://doi.org/10.3389/fmicb.2017.00394

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. https://doi.org/10.7554/eLife.29218

Newton, K.A., McLuckey, S.A., 2004. Generation and manipulation of sodium cationized peptides in the gas phase. J. Am. Soc. Mass Spectrom. 15, 607–615. https://doi.org/10.1016/J.JASMS.2003.12.014

Nguyen, D.D., Wu, C.-H., Moree, W.J., Lamsa, A., Medema, M.H., Zhao, X., Gavilan, R.G., Aparicio, M., Atencio, L., Jackson, C., Ballesteros, J., Sanchez, J., Watrous, J.D., Phelan, V. V., van de Wiel, C., Kersten, R.D., Mehnaz, S., De Mot, R., Shank, E.A., Charusanti, P., Nagarajan, H., Duggan, B.M., Moore, B.S., Bandeira, N., Palsson, B.O., Pogliano, K., Gutierrez, M., Dorrestein, P.C., 2013. MS/MS networking guided analysis of molecule and gene cluster families. Proc. Natl. Acad. Sci. 110, E2611–E2620. https://doi.org/10.1073/pnas.1303471110

Nguyen, K.T., Willey, J.M., Nguyen, L.D., Nguyen, L.T., Viollier, P.H., Thompson, C.J., 2002. A central regulator of morphological differentiation in the multicellular bacterium Streptomyces coelicolor. Mol. Microbiol. 46, 1223–1238. https://doi.org/10.1046/j.1365-2958.2002.03255.x

Nicole Kresge, R.D.S. and R.L.H., 2004. Selman Waksman: the Father of Antibiotics. J. Biol. Chem. 279, 101–102. https://doi.org/10.1074/jbc.X400012200

Nishie, M., Shioya, K., Nagao, J., Jikuya, H., Sonomoto, K., 2009. ATP-dependent leader peptide cleavage by NukT, a bifunctional ABC transporter, during lantibiotic biosynthesis. J. Biosci. Bioeng. 108, 460–464. https://doi.org/10.1016/J.JBIOSC.2009.06.002

Norris, G.E., Patchett, M.L., 2016. The glycocins: in a class of their own. Curr. Opin. Struct. Biol. 40, 112–119. https://doi.org/10.1016/J.SBI.2016.09.003

Northcote, P.T., Siegel, M., Borders, D.B., Lee, M.D., 1994. Glycothiohexide alpha, a novel antibiotic produced by "Sebekia" sp., LL-14E605. III. Structural elucidation. J. Antibiot. (Tokyo). 47, 901–8.

Noskov, V.N., Karas, B.J., Young, L., Chuang, R.-Y., Gibson, D.G., Lin, Y.-C., Stam, J., Yonemoto, I.T., Suzuki, Y., Andrews-Pfannkoch, C., Glass, J.I., Smith, H.O., Hutchison, C.A., Venter, J.C., Weyman, P.D., 2012. Assembly of Large, High G+C Bacterial DNA Fragments in Yeast. Am. Chem. Soc. Synth. Biol. 1, 267–273. https://doi.org/10.1021/sb3000194

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. BioMed Cent. genomics 4, 16. https://doi.org/10.1186/1471-2164-4-16

Nowicka, B., Kruk, J., 2010. Occurrence, biosynthesis and function of isoprenoid quinones. Biochim. Biophys. Acta - Bioenerg. 1797, 1587–1605. https://doi.org/10.1016/J.BBABIO.2010.06.007

O'Hara, K., Kanda, T., Ohmiya, K., Ebisu, T., Kono, M., 1989. Purification and characterization of macrolide 2'-phosphotransferase from a strain of Escherichia coli that is highly resistant to erythromycin. Antimicrob. Agents Chemother. 33, 1354–1357. https://doi.org/10.1128/AAC.33.8.1354

Ogino, J., Moore, R.E.M., Patterson, G.M.L.P., Smith, C.D., 1996. Dendroamides, New Cyclic Hexapeptides from a Blue-Green Alga. Multidrug-Resistance Reversing Activity of Dendroamide A. J. Nat. Prod. 59, 581–586. https://doi.org/10.1021/NP960178S

Okamoto-Hosoya, Y., Sato, T.-A., Ochi, K., 2000. Resistance to Paromomycin is conferred by rpsL mutations, accompanied by an enhanced antibiotic production in *Streptomyces coelicolor* A3(2). J. Antibiot. (Tokyo). 53, 1424–1427. https://doi.org/10.7164/antibiotics.53.1424

Okeley, N.M., Zhu, Y., van der Donk, W.A., 2000. Facile Chemoselective Synthesis of Dehydroalanine-Containing Peptides. Org. Lett. 2, 3603–3606. https://doi.org/10.1021/OL006485D

Ökesli, A., Cooper, L.E., Fogle, E.J., Van Der Donk, W.A., 2011. Nine post-translational modifications during the biosynthesis of cinnamycin. J. Am. Chem. Soc. 133, 13753–13760. https://doi.org/10.1021/ja205783f

Olano, C., García, I., González, A., Rodriguez, M., Rozas, D., Rubio, J., Sánchez-Hidalgo, M., Braña, A.F., Méndez, C., Salas, J.A., 2014. Activation and identification of five clusters for secondary metabolites in Streptomyces albus J1074. Microb. Biotechnol. 7, 242–256. https://doi.org/10.1111/1751-7915.12116

Oldach, F., Al Toma, R., Kuthning, A., Caetano, T., Mendo, S., Budisa, N., Süssmuth, R.D., 2012. Congeneric Lantibiotics from Ribosomal In Vivo Peptide Synthesis with Noncanonical Amino Acids. Angew. Chemie Int. Ed. 51, 415–418. https://doi.org/10.1002/anie.201106154

Oldfield, E., Lin, F.-Y., 2012. Terpene Biosynthesis: Modularity Rules. Angew. Chemie Int. Ed. 51, 1124–1137. https://doi.org/10.1002/anie.201103110

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. https://doi.org/10.1038/nchembio.286

Onaka, H., Tabata, H., Igarashi, Y., Sato, Y., Furumai, T., 2001. Goadsporin, a Chemical Substance which Promotes Secondary Metabolism and Morphogenesis in Streptomycetes. I. Purification and Characterization. J. Antibiot. (Tokyo). 54, 1036–1044. https://doi.org/10.7164/antibiotics.54.1036

Ongey, E.L., Neubauer, P., 2016. Lanthipeptides: chemical synthesis versus in vivo biosynthesis as tools for pharmaceutical production. Microb. Cell Fact. 15, 97. https://doi.org/10.1186/s12934-016-0502-y

Oppenheim, D.S., Yanofsky, C., 1980. Translational coupling during expression of the tryptophan operon of Escherichia coli. Genetics 95, 785–795.

Ortega, M.A., Cogan, D.P., Mukherjee, S., Garg, N., Li, B., Thibodeaux, G.N., Maffioli, S.I., Donadio, S., Sosio, M., Escano, J., Smith, L., Nair, S.K., van der Donk, W.A., 2017. Two Flavoenzymes Catalyze the Post-Translational Generation of 5-Chlorotryptophan and 2-Aminovinyl-Cysteine during NAI-107 Biosynthesis. Am. Chem. Soc. Chem. Biol. 12, 548–557. https://doi.org/10.1021/acschembio.6b01031

Ortega, M.A., Hao, Y., Zhang, Q., Walker, M.C., Van Der Donk, W.A., Nair, S.K., 2015. Structure and mechanism of the tRNA-dependent lantibiotic dehydratase NisB. Nature 517, 509–512. https://doi.org/10.1038/nature13888

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. Am. Chem. Soc. Chem. Biol. 9, 1718–1725. https://doi.org/10.1021/cb5002526

Otaka, T., Kaji, A., 1983. Mode of action of bottromycin A2: Effect of bottromycin A2on polysomes. Fed. Eur. Biochem. Soc. Lett. 153, 53–59. https://doi.org/10.1016/0014-5793(83)80118-5

Otaka, T., Kaji, A., 1981. Mode of action of bottromycin A 2 : effect on peptide bond formation. Fed. Eur. Biochem. Soc. Lett. 123, 173–176. https://doi.org/10.1016/0014-5793(81)80280-3

Otaka, T., Kaji, A., 1976. Mode of action of bottromycin A2. Release of aminoacyl or peptidyl tRNA from ribosomes. J. Biol. Chem. 251, 2299–2306.

Pahirulzaman, K., Williams, K., Lazarus, C.M., 2012. A Toolkit for Heterologous Expression of Metabolic Pathways in Aspergillus oryzae, in: Methods in Enzymology. pp. 241–260. https://doi.org/10.1016/B978-0-12-404634-4.00012-7

Pan, S.J., Link, A.J., 2011. Sequence Diversity in the Lasso Peptide Framework: Discovery of Functional Microcin J25 Variants with Multiple Amino Acid Substitutions. J. Am. Chem. Soc. 133, 5016–5023. https://doi.org/10.1021/ja1109634

Parajuli, A., Kwak, D.H., Dalponte, L., Leikoski, N., Galica, T., Umeobika, U., Trembleau, L., Bent, A., Sivonen, K., Wahlsten, M., Wang, H., Rizzi, E., De Bellis, G., Naismith, J., Jaspars, M., Liu, X., Houssen, W., Fewer, D.P., 2016. A Unique Tryptophan C-Prenyltransferase from the Kawaguchipeptin Biosynthetic Pathway. Angew. Chem. Int. Ed. Engl. 55, 3596–3599. https://doi.org/10.1002/anie.201509920

Parent, A., Guillot, A., Benjdia, A., Chartier, G., Leprince, J., Berteau, O., 2016. The B $_{12}$ -Radical SAM Enzyme PoyC Catalyzes Valine C $_\beta$ -Methylation during Polytheonamide Biosynthesis. J. Am. Chem. Soc. 138, 15515–15518. https://doi.org/10.1021/jacs.6b06697

Park, Y., Shin, I., Rhee, S., 2015. Crystal structure of the effector protein HopA1 from Pseudomonas syringae. J. Struct. Biol. 189, 276–280. https://doi.org/10.1016/J.JSB.2015.02.002

Parks, W.M., Bottrill, A.R., Pierrat, O.A., Durrant, M.C., Maxwell, A., 2007. The action of the bacterial toxin, microcin B17, on DNA gyrase. Biochimie 89, 500–507. https://doi.org/10.1016/J.BIOCHI.2006.12.005

Pawlowski, A.C., Stogios, P.J., Koteva, K., Skarina, T., Evdokimova, E., Savchenko, A., Wright, G.D., 2018. The evolution of substrate discrimination in macrolide antibiotic resistance enzymes. Nat. Commun. 9, 112. https://doi.org/10.1038/s41467-017-02680-0

Pigac, J., Schrempf, H., 1995. A simple and rapid method of transformation of Streptomyces rimosus R6 and other streptomycetes by electroporation. Appl. Environ. Microbiol. 61, 352–356.

Piscotta, F.J., Tharp, J.M., Liu, W.R., Link, A.J., 2015. Expanding the chemical diversity of lasso peptide MccJ25 with genetically encoded noncanonical amino acids. Chem. Commun. 51, 409–412. https://doi.org/10.1039/c4cc07778d

Prange, T., Ducruix, A., Pascard, C., Lunel, J., 1977. Structure of nosiheptide, a polythiazole-containing antibiotic [40]. Nature 265, 189–190. https://doi.org/10.1038/265189a0

Puar, M.S., Chan, T.M., Hegde, V., Patel, M., Bartner, P., Ng, K.J., Pramanik, B.N., MacFarlane, R.D., 1998. Sch 40832: a novel thiostrepton from Micromonospora carbonacea. J. Antibiot. (Tokyo). 51, 221–224. https://doi.org/10.7164/antibiotics.51.221

Puar, M.S., Ganguly, A.K., Afonso, A., Brambilla, R., Mangiaracina, P., Sarre, O., MacFarlane, R.D., 1981. Sch 18640. A new thiostrepton-type antibiotic. J. Am. Chem. Soc. 103, 5231–5233. https://doi.org/10.1021/ja00407a047

Ran, F.A., Hsu, P.D., Wright, J., Agarwala, V., Scott, D.A., Zhang, F., 2013. Genome engineering using the CRISPR-Cas9 system. Nat. Protoc. 8, 2281–2308. https://doi.org/10.1038/nprot.2013.143

Rangarajan, E.S., Li, Y., Iannuzzi, P., Tocilj, A., Hung, L.-W., Matte, A., Cygler, M., 2004. Crystal structure of a dodecameric FMN-dependent UbiX-like decarboxylase (Pad1) from Escherichia coli O157: H7. Protein Sci. 13, 3006–3016. https://doi.org/10.1110/ps.04953004

Raymond, C.K., Pownder, T.A., Sexson, S.L., 1999. General method for plasmid construction using homologous recombination. Biotechniques 26, 134–8, 140–1.

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. Mol. Microbiol. 21, 77–96. https://doi.org/10.1046/j.1365-2958.1996.6191336.x

Regni, C.A., Roush, R.F., Miller, D.J., Nourse, A., Walsh, C.T., Schulman, B.A., 2009. How the MccB bacterial ancestor of ubiquitin E1 initiates biosynthesis of the microcin C7 antibiotic. EMBO J. 28, 1953–64. https://doi.org/10.1038/emboj.2009.146

Repka, L.M., Chekan, J.R., Nair, S.K., van der Donk, W.A., 2017. Mechanistic Understanding of Lanthipeptide Biosynthetic Enzymes. Chem. Rev. 117, 5457–5520. https://doi.org/10.1021/acs.chemrev.6b00591

Rince, A., Dufour, A., Le Pogam, S., Thuault, D., Bourgeois, C.M., Le Pennec, J.P., 1994. Cloning, expression, and nucleotide sequence of genes involved in production of lactococcin DR, a bacteriocin from lactococcus lactis subsp. lactis. Appl. Environ. Microbiol. 60, 1652–1657.

Rivero-Müller, A., Lajić, S., Huhtaniemi, I., 2007. Assisted large fragment insertion by Red/ET-recombination (ALFIRE)--an alternative and enhanced method for large fragment recombineering. Nucleic Acids Res. 35, e78. https://doi.org/10.1093/nar/gkm250

Rogers, L.A., 1928. The inhibiting effect of Streptococcus lactis on Lactobacillus bulgaricus. J. Bacteriol. 16, 321–325.

Rollema, H.S., Kuipers, O.P., Both, P., De Vos, W.M., Siezen, R.J., 1995. Improvement of solubility and stability of the antimicrobial peptide nisin by protein engineering. Appl. Environ. Microbiol. 61, 2873–2878.

Ross, R.P., McAuliffe, O., Hill, C., 2000. Each peptide of the two-component lantibiotic lacticin 3147 requires a separate modification enzyme for activity. Microbiology 146, 2147–2154. https://doi.org/10.1099/00221287-146-9-2147

Rudolph, M.M., Vockenhuber, M.-P., Suess, B., 2013. Synthetic riboswitches for the conditional control of gene expression in Streptomyces coelicolor. Microbiology 159, 1416–1422. https://doi.org/10.1099/mic.0.067322-0

Ruffner, D.E., Schmidt, E.W., Heemstra, J.R., 2015. Assessing the combinatorial potential of the RiPP cyanobactin tru pathway. Am. Chem. Soc. Synth. Biol. 4, 482–92. https://doi.org/10.1021/sb500267d

Sajid, I., Shaaban, K.A., Frauendorf, H., Hasnain, S., 2008. Val-Geninthiocin : Structure Elucidation and MS n Fragmentation of Thiopeptide Antibiotics Produced by Streptomyces sp . RSF18. Zeitschrift für Naturforsch. 63, 1223–1230. https://doi.org/10.1515/znb-2008-1014

Salvatella, X., Caba, J.M., Albericio, F., Giralt, E., 2003. Solution structure of the antitumor candidate trunkamide A by 2D NMR and restrained simulated annealing methods. J. Org. Chem. 68, 211–215. https://doi.org/10.1021/jo026464s

Sarrion-Perdigones, A., Falconi, E.E., Zandalinas, S.I., Juárez, P., Fernández-del-Carmen, A., Granell, A., Orzaez, D., 2011. GoldenBraid: an iterative cloning system for standardized assembly of reusable genetic modules. Public Libr. Sci. One 6, e21622. https://doi.org/10.1371/journal.pone.0021622

Sathesh-Prabu, C., Lee, S.K., 2018. Genome Editing Tools for *Escherichia coli* and Their Application in Metabolic Engineering and Synthetic Biology, in: Emerging Areas in Bioengineering. Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim, Germany, pp. 307–319. https://doi.org/10.1002/9783527803293.ch17

Schmitz, F.J., Ksebati, M.B., Chang, J.S., Wang, J.L., Hossain, M.B., Van der Helm, D., Engel, M.H., Serban, A., Silfer, J.A., 1989. Cyclic peptides from the ascidian Lissoclinum patella: conformational analysis of patellamide D by x-ray analysis and molecular modeling. J. Org. Chem. 54, 3463–3472. https://doi.org/10.1021/jo00275a036

Schwab, W., Davidovich-Rikanati, R., Lewinsohn, E., 2008. Biosynthesis of plant-derived flavor compounds. Plant J. 54, 712–732. https://doi.org/10.1111/j.1365-313X.2008.03446.x

Schwalen, C.J., Hudson, G.A., Kille, B., Mitchell, D.A., 2018. Bioinformatic Expansion and Discovery of Thiopeptide Antibiotics. J. Am. Chem. Soc. 140, 9494–9501. https://doi.org/10.1021/jacs.8b03896

Schwalen, C.J., Hudson, G.A., Kosol, S., Mahanta, N., Challis, G.L., Mitchell, D.A., 2017. In Vitro Biosynthetic Studies of Bottromycin Expand the Enzymatic Capabilities of the YcaO Superfamily. J. Am. Chem. Soc. 139, 18154–18157. https://doi.org/10.1021/jacs.7b09899

Shao, Z., Rao, G., Li, C., Abil, Z., Luo, Y., Zhao, H., 2013. Refactoring the silent spectinabilin gene cluster using a plug-and-play scaffold. Am. Chem. Soc. Synth. Biol. 2, 662–669. https://doi.org/10.1021/sb400058n

Shao, Z., Zhao, H., 2013. Construction and Engineering of Large Biochemical Pathways via DNA Assembler, in: Methods in Molecular Biology. pp. 85–106. https://doi.org/10.1007/978-1-

62703-625-2_9

Shao, Z., Zhao, H., 2012. DNA assembler: a synthetic biology tool for characterizing and engineering natural product gene clusters. Methods Enzymol. 517, 203–24. https://doi.org/10.1016/B978-0-12-404634-4.00010-3

Shao, Z., Zhao, H., Zhao, H., 2009. DNA assembler, an in vivo genetic method for rapid construction of biochemical pathways. Nucleic Acids Res. 37, e16. https://doi.org/10.1093/nar/gkn991

Sharom, F.J., 2008. ABC multidrug transporters: structure, function and role in chemoresistance. Pharmacogenomics 9, 105–127. https://doi.org/10.2217/14622416.9.1.105

Sherwood, E.J., Hesketh, A.R., Bibb, M.J., 2013. Cloning and Analysis of the Planosporicin Lantibiotic Biosynthetic Gene Cluster of Planomonospora alba. J. Bacteriol. 195, 2309–2321. https://doi.org/10.1128/JB.02291-12

Sherwood, A. V., Henkin, T.M., 2016. Riboswitch-Mediated Gene Regulation: Novel RNA Architectures Dictate Gene Expression Responses. Annu. Rev. Microbiol. 70, 361–374. https://doi.org/10.1146/annurev-micro-091014-104306

Shetty, R., Lizarazo, M., Rettberg, R., Knight, T.F., 2011. Assembly of BioBrick Standard Biological Parts Using Three Antibiotic Assembly. Methods Enzymol. 498, 311–326. https://doi.org/10.1016/B978-0-12-385120-8.00013-9

Shima, J., Hesketh, A., Okamoto, S., Kawamoto, S., Ochi, K., 1996. Induction of actinorhodin production by rpsL (encoding ribosomal protein S12) mutations that confer streptomycin resistance in Streptomyces lividans and Streptomyces coelicolor A3(2). J. Bacteriol. 178, 7276–84.

Shimafuji, C., Noguchi, M., Nishie, M., Nagao, J., Shioya, K., Zendo, T., Nakayama, J., Sonomoto, K., 2015. In vitro catalytic activity of N-terminal and C-terminal domains in NukM, the post-translational modification enzyme of nukacin ISK-1. J. Biosci. Bioeng. 120, 624–629. https://doi.org/10.1016/J.JBIOSC.2015.03.020

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 A2: A potent antibiotic against MRSA and VRE. Angew. Chemie - Int. Ed. 48, 914–917. https://doi.org/10.1002/anie.200804138

Shore, D., 1997. Genetic recombination: Sex-change operations in yeast. Curr. Biol. 7, R24–R27. https://doi.org/10.1016/S0960-9822(06)00012-1

Singh, B., Sharma, R.A., 2015. Plant terpenes: defense responses, phylogenetic analysis, regulation and clinical applications. 3 Biotech 5, 129–151. https://doi.org/10.1007/s13205-014-0220-2

Singh, S.B., Herath, K., Yu, N.X., Walker, A.A., Connors, N., 2008. Biosynthetic studies of Nocathiacin-I. Tetrahedron Lett. 49, 6265–6268. https://doi.org/10.1016/J.TETLET.2008.08.031

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. Acc. Chem. Res. 44, 261–268. https://doi.org/10.1021/ar1001395

Sivonen, K., Leikoski, N., Fewer, D.P., Jokela, J., 2010. Cyanobactins-ribosomal cyclic peptides produced by cyanobacteria. Appl. Microbiol. Biotechnol. 86, 1213–1225. https://doi.org/10.1007/s00253-010-2482-x

Skaugen, M., Nissen-Meyer, J., Jung, G., Stevanovic, S., Sletten, K., Abildgaard, C.I.M., Nes, I.F., 1994. In vivo conversion of L-serine to D-alanine in a ribosomally synthesized polypeptide. J. Biol. Chem. 269, 27183–27185.

Sleight, S.C., Bartley, B.A., Lieviant, J.A., Sauro, H.M., 2010. In-Fusion BioBrick assembly and re-engineering. Nucleic Acids Res. 38, 2624–2636. https://doi.org/10.1093/nar/gkq179

Smith, L., Zachariah, C., Thirumoorthy, R., Rocca, J., Novák, J., Hillman, J.D., Edison, A.S., 2003. Structure and dynamics of the lantibiotic mutacin 1140. Biochemistry 42, 10372–10384. https://doi.org/10.1021/bi034490u

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. Sci. Adv. 4, eaat2720. https://doi.org/10.1126/sciadv.aat2720

Stamatakis, A., 2014. RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 30, 1312–1313. https://doi.org/10.1093/bioinformatics/btu033

Takahashi, Y., Naganawa, H., 1976. The revised structure of bottromycin A2. J. … 2, 1120–1123. https://doi.org/10.7164/antibiotics.29.1120

Tang, W., Jiménez-Osés, G., Houk, K.N., van der Donk, W.A., 2015. Substrate control in stereoselective lanthionine biosynthesis. Nat. Chem. 7, 57–64. https://doi.org/10.1038/nchem.2113

Tang, W., van der Donk, W.A., 2012. Structural Characterization of Four Prochlorosins: A Novel Class of Lantipeptides Produced by Planktonic Marine Cyanobacteria. Biochemistry 51, 4271–4279. https://doi.org/10.1021/bi300255s

Tao, H., Weng, Y., Zhuo, R., Chang, G., Urbatsch, I.L., Zhang, Q., 2011. Design and synthesis of Selenazole-containing peptides for cocrystallization with P-glycoprotein. Chembiochem 12, 868–873. https://doi.org/10.1002/cbic.201100048

Testero, S.A., Fisher, J.F., Mobashery, S., 2010. β-Lactam Antibiotics, in: Burger's Medicinal Chemistry and Drug Discovery. John Wiley & Sons, Inc., Hoboken, NJ, USA, pp. 257–402. https://doi.org/10.1002/0471266949.bmc226

Thibodeaux, C.J., Wagoner, J., Yu, Y., van der Donk, W.A., 2016. Leader Peptide Establishes Dehydration Order, Promotes Efficiency, and Ensures Fidelity During Lacticin 481 Biosynthesis. J. Am. Chem. Soc. 138, 6436–6444. https://doi.org/10.1021/jacs.6b00163

Thompson, R.E., Collin, F., Maxwell, A., Jolliffe, K.A., Payne, R.J., 2014. Synthesis of full length and truncated microcin B17 analogues as DNA gyrase poisons. Org. Biomol. Chem. 12, 1570–1578. https://doi.org/10.1039/c3ob42516a

Thorpe, H.M., Smith, M.C.M., 1998. In vitro site-specific integration of bacteriophage DNA catalyzed by a recombinase of the resolvase/invertase family. Proc. Natl. Acad. Sci. 95, 5505–5510. https://doi.org/10.1073/pnas.95.10.5505

Tian, Z., Sun, P., Yan, Y., Wu, Z., Zheng, Q., Zhou, S., Zhang, H., Yu, F., Jia, X., Chen, D., Mándi, A., Kurtán, T., Liu, W., 2015. An enzymatic [4+2] cyclization cascade creates the pentacyclic core of pyrroindomycins. Nat. Chem. Biol. 11, 259–265. https://doi.org/10.1038/nchembio.1769

Tianero, M.D.B., Donia, M.S., Young, T.S., Schultz, P.G., Schmidt, E.W., 2012. Ribosomal route to small-molecule diversity. J. Am. Chem. Soc. 134, 418–25. https://doi.org/10.1021/ja208278k

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. Nat. Chem. Biol. 13, 470–478. https://doi.org/10.1038/nchembio.2319

Tocchetti, A., Maffioli, S., Iorio, M., Alt, S., Mazzei, E., Brunati, C., Sosio, M., Donadio, S., 2013. Capturing Linear Intermediates and C-Terminal Variants during Maturation of the Thiopeptide GE2270. Chem. Biol. 20, 1067–1077. https://doi.org/10.1016/J.CHEMBIOL.2013.07.005

Travin, D.Y., Metelev, M., Serebryakova, M., Komarova, E.S., Osterman, I.A., Ghilarov, D., Severinov, K., 2018. Biosynthesis of Translation Inhibitor Klebsazolicin Proceeds through Heterocyclization and N-Terminal Amidine Formation Catalyzed by a Single YcaO Enzyme. J. Am. Chem. Soc. 140, 5625–5633. https://doi.org/10.1021/jacs.8b02277

Truman, A.W., 2016. Cyclisation mechanisms in the biosynthesis of ribosomally synthesised and post-translationally modified peptides. Beilstein J. Org. Chem. 12, 1250–68. https://doi.org/10.3762/bjoc.12.120

Urban, J.H., Moosmeier, M.A., Aumüller, T., Thein, M., Bosma, T., Rink, R., Groth, K., Zulley, M., Siegers, K., Tissot, K., Moll, G.N., Prassler, J., 2017. Phage display and selection of lanthipeptides on the carboxy-terminus of the gene-3 minor coat protein. Nat. Commun. 8, 1500. https://doi.org/10.1038/s41467-017-01413-7

Van De Kamp, M., Van Den Hooven, H.W., Konings, R.N.H., Bierbaum, G., Sahl, H. -G, Kuipers, O.P., Siezen, R.J., De Vos, W.M., Hilbers, C.W., Van De Ven, F.J.M., 1995. Elucidation of the Primary Structure of the Lantibiotic Epilancin K7 from Staphylococcus epidermidis K7: Cloning and Characterisation of the Epilancin-K7–Encoding Gene and NMR Analysis of Mature Epilancin K7. Eur. J. Biochem. 230, 587–600. https://doi.org/10.1111/j.1432-1033.1995.0587h.x

van der Donk, W.A., Nair, S.K., 2014. Structure and mechanism of lanthipeptide biosynthetic enzymes. Curr. Opin. Struct. Biol. 29, 58–66. https://doi.org/10.1016/j.sbi.2014.09.006

Van der Meer, J.R., Polman, J., Beerthuyzen, M.M., Siezen, R.J., Kuipers, O.P., De Vos, W.M., 1993. Characterization of the Lactococcus lactis nisin A operon genes nisP, encoding a subtilisin-like serine protease involved in precursor processing, and nisR, encoding a regulatory protein involved in nisin biosynthesis. J. Bacteriol. 175, 2578–2588. https://doi.org/10.1128/jb.175.9.2578-2588.1993

van der Velden, N.S., Kälin, N., Helf, M.J., Piel, J., Freeman, M.F., Künzler, M., 2017. Autocatalytic backbone N-methylation in a family of ribosomal peptide natural products. Nat. Chem. Biol. 13, 833–835. https://doi.org/10.1038/nchembio.2393

Van Dolleweerd, C.J., Kessans, S.A., Van De Bittner, K.C., Bustamante, L.Y., Bundela, R., Scott, B., Nicholson, M.J., Parker, E.J., 2018. MIDAS: A Modular DNA Assembly System for Synthetic Biology. Am. Chem. Soc. Synth. Biol. 7, 1018–1029. https://doi.org/10.1021/acssynbio.7b00363

Van Tyne, D., Martin, M.J., Gilmore, M.S., 2013. Structure, function, and biology of the Enterococcus faecalis cytolysin. Toxins (Basel). 5, 895–911. https://doi.org/10.3390/toxins5050895

Velásquez, J.E., Zhang, X., van der Donk, W.A., 2011. Biosynthesis of the antimicrobial peptide epilancin 15X and its N-terminal lactate. Chem. Biol. 18, 857–867. https://doi.org/10.1016/j.chembiol.2011.05.007

Vijaya Kumar, E.K.S., Kenia, J., Mukhopadhyay, T., Nadkarni, S.R., 1999. Methylsulfomycin I, a new cyclic peptide antibiotic from a streptomyces sp. HIL Y-9420704. J. Nat. Prod. 62, 1562–1564. https://doi.org/10.1021/np990088y

Vizan, J.L., Hernandez-Chico, C., Del Castillo, I., Moreno1, F., 1991. The peptide antibiotic microcin Bi 7 induces double-strand cleavage of DNA mediated by E.coli DNA gyrase. EMBO J. 10, 467–476. https://doi.org/10.1002/j.1460-2075.1991.tb07969.x

Völler, G.H., Krawczyk, B., Ensle, P., Süssmuth, R.D., 2013. Involvement and Unusual Substrate Specificity of a Prolyl Oligopeptidase in Class III Lanthipeptide Maturation. J. Am. Chem. Soc. 135, 7426–7429. https://doi.org/10.1021/ja402296m

Völler, G.H., Krawczyk, J.M., Pesic, A., Krawczyk, B., Nachtigall, J., Süssmuth, R.D., 2012.

Characterization of New Class III Lantibiotics-Erythreapeptin, Avermipeptin and Griseopeptin from Saccharopolyspora erythraea, Streptomyces avermitilis and Streptomyces griseus Demonstrates Stepwise N-Terminal Leader Processing. ChemBioChem 13, 1174–1183. https://doi.org/10.1002/cbic.201200118

Waisvisz, J.M., van der Hoeven, M.G., van Peppen, J., Zwennis, W.C.M., 1957. Bottromycin. I. A New Sulfur-containing Antibiotic. J. Am. Chem. Soc. 79, 4520–4521. https://doi.org/10.1021/ja01573a072

Waksman, S.A., Woodruff, H.B., 1940. The Soil as a Source of Microorganisms Antagonistic to Disease-Producing Bacteria*1. J. Bacteriol. 40, 581–600.

Wals, K., Ovaa, H., 2014. Unnatural amino acid incorporation in E. coli: current and future applications in the design of therapeutic proteins. Front. Chem. 2, 15. https://doi.org/10.3389/fchem.2014.00015

Walsh, C.T., 2016. Insights into the chemical logic and enzymatic machinery of NRPS assembly lines. Nat. Prod. Rep. 33, 127–135. https://doi.org/10.1039/C5NP00035A

Walton, J., 2018. The cyclic peptide toxins of Amanita and other poisonous mushrooms. Springer International Publishing.

Wang, H., van der Donk, W.A., 2012. Biosynthesis of the Class III Lantipeptide Catenulipeptin. Am. Chem. Soc. Chem. Biol. 7, 1529–1535. https://doi.org/10.1021/cb3002446

Wang, M., Carver, J.J., Phelan, V. V, Sanchez, L.M., Garg, N., Peng, Y., Nguyen, D.D.D.-T., Watrous, J., Kapono, C.A., Luzzatto-Knaan, T., Porto, C., Bouslimani, A., Melnik, A. V, Meehan, M.J., Liu, W.-T., Crüsemann, M., Boudreau, P.D., Esquenazi, E., Sandoval-Calderón, M., Kersten, R.D., Pace, L.A., Quinn, R.A., Duncan, K.R., Hsu, C.-C., Floros, D.J., Gavilan, R.G., Kleigrewe, K., Northen, T., Dutton, R.J., Parrot, D., Carlson, E.E., Aigle, B., Michelsen, C.F., Jelsbak, L., Sohlenkamp, C., Pevzner, P., Edlund, A., McLean, J., Piel, J., Murphy, B.T., Gerwick, L., Liaw, C.-C., Yang, Y.-L., Humpf, H.-U., Maansson, M., Keyzers, R.A., Sims, A.C., Johnson, A.R., Sidebottom, A.M., Sedio, B.E., Klitgaard, A., Larson, C.B., Boya P, C.A., Torres-Mendoza, D., Gonzalez, D.J., Silva, D.B., Marques, L.M., Demarque, D.P., Pociute, E., O'Neill, E.C., Briand, E., Helfrich, E.J.N., Granatosky, E.A., Glukhov, E., Ryffel, F., Houson, H., Mohimani, H., Kharbush, J.J., Zeng, Y., Vorholt, J.A., Kurita, K.L., Charusanti, P., McPhail, K.L., Nielsen, K.F., Vuong, L., Elfeki, M., Traxler, M.F., Engene, N., Koyama, N., Vining, O.B., Baric, R., Silva, R.R., Mascuch, S.J., Tomasi, S., Jenkins, S., Macherla, V., Hoffman, T., Agarwal, V., Williams, P.G., Dai, J., Neupane, R., Gurr, J., Rodríguez, A.M.C., Lamsa, A., Zhang, C., Dorrestein, K., Duggan, B.M., Almaliti, J., Allard, P.-M., Phapale, P., Nothias, L.-F., Alexandrov, T., Litaudon, M., Wolfender, J.-L., Kyle, J.E., Metz, T.O., Peryea, T., Nguyen, D.D.D.-T., VanLeer, D., Shinn, P., Jadhav, A., Müller, R., Waters, K.M., Shi, W., Liu, X., Zhang, L., Knight, R., Jensen, P.R., Palsson, B.Ø., Pogliano, K., Linington, R.G., Gutiérrez, M., Lopes, N.P., Gerwick, W.H., Moore, B.S., Dorrestein, P.C., Bandeira, N., 2016. Sharing and community curation of mass spectrometry data with Global Natural Products Social Molecular Networking. Nat. Biotechnol. 34, 828–837. https://doi.org/10.1038/nbt.3597

Wang, W., Li, X., Wang, J., Xiang, S., Feng, X., Yang, K., 2013. An Engineered Strong Promoter for Streptomycetes. Appl. Environ. Microbiol. 79, 4484–4492. https://doi.org/10.1128/AEM.00985-13

Weber, E., Engler, C., Gruetzner, R., Werner, S., Marillonnet, S., 2011. A Modular Cloning System for Standardized Assembly of Multigene Constructs. Public Libr. Sci. One 6, e16765. https://doi.org/10.1371/journal.pone.0016765

White, A.M., Craik, D.J., 2016. Discovery and optimization of peptide macrocycles. Expert Opin. Drug Discov. 11, 1151–1163. https://doi.org/10.1080/17460441.2016.1245720

Widhalm, J.R., Rhodes, D., 2016. Biosynthesis and molecular actions of specialized 1,4-naphthoquinone natural products produced by horticultural plants. Hortic. Res. 3, 16046. https://doi.org/10.1038/hortres.2016.46

Wiebach, V., Mainz, A., Siegert, M.-A.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. https://doi.org/10.1038/s41589-018-0068-6

Wiedemann, I., Bottiger, T., Bonelli, R.R., Wiese, A., Hagge, S.O., Gutsmann, T., Seydel, U., Deegan, L., Hill, C., Ross, P., Sahl, H.-G., 2006. The mode of action of the lantibiotic lacticin 3147 - a complex mechanism involving specific interaction of two peptides and the cell wall precursor lipid II. Mol. Microbiol. 61, 285–296. https://doi.org/10.1111/j.1365-2958.2006.05223.x

Wieland Brown, L.C., Acker, M.G., Clardy, J., Walsh, C.T., Fischbach, M.A., 2009. Thirteen posttranslational modifications convert a 14-residue peptide into the antibiotic thiocillin. Proc. Natl. Acad. Sci. 106, 2549–2553. https://doi.org/10.1073/pnas.0900008106

Wilker, J.J., Lippard, S.J., 1995. Modeling the DNA Methylphosphotriester Repair Site in Escherichia coli Ada. Why Zinc and Four Cysteines? J. Am. Chem. Soc. 117, 8682–8683. https://doi.org/10.1021/ja00138a031

Williams, A.B., Jacobs, R.S., 1993. A marine natural product, patellamide D, reverses multidrug resistance in a human leukemic cell line. Cancer Lett. 71, 97–102. https://doi.org/10.1016/0304-3835(93)90103-G

Williams, R., Berndt, A., Miller, S., Hon, W.-C., Zhang, X., 2009. Form and flexibility in phosphoinositide 3-kinases. Biochem. Soc. Trans. 37, 615–626. https://doi.org/10.1042/BST0370615

Winn, M., Fyans, J.K., Zhuo, Y., Micklefield, J., 2016. Recent advances in engineering nonribosomal peptide assembly lines. Nat. Prod. Rep. 33, 317–347. https://doi.org/10.1039/C5NP00099H

Wlodek, A., Kendrew, S.G., Coates, N.J., Hold, A., Pogwizd, J., Rudder, S., Sheehan, L.S., Higginbotham, S.J., Stanley-Smith, A.E., Warneck, T., Nur-E-Alam, M., Radzom, M., Martin, C.J., Overvoorde, L., Samborskyy, M., Alt, S., Heine, D., Carter, G.T., Graziani, E.I., Koehn, F.E., McDonald, L., Alanine, A., Rodríguez Sarmiento, R.M., Chao, S.K., Ratni, H., Steward, L., Norville, I.H., Sarkar-Tyson, M., Moss, S.J., Leadlay, P.F., Wilkinson, B., Gregory, M.A., 2017. Diversity oriented biosynthesis via accelerated evolution of modular gene clusters. Nat. Commun. 8, 1206. https://doi.org/10.1038/s41467-017-01344-3

Wolk, O.K.C., Wolk, C., 2002. Genetic tools for cyanobacteria. Appl. Microbiol. Biotechnol. 58, 123–137. https://doi.org/10.1007/s00253-001-0864-9

Wong, C.H., Siah, K.W., Lo, A.W., 2018. Estimation of clinical trial success rates and related parameters. Biostatistics 00, 1–14. https://doi.org/10.1093/biostatistics/kxx069

Wright, G.D., Thompson, P.R., 1999. Aminoglycoside phosphotransferases: proteins, structure, and mechanism. Front. Biosci. 4, d9. https://doi.org/10.2741/Wright

Wu, L.-F., Meng, S., Tang, G.-L., 2016. Ferrous iron and α-ketoglutarate-dependent dioxygenases in the biosynthesis of microbial natural products. Biochim. Biophys. Acta - Proteins Proteomics 1864, 453–470. https://doi.org/10.1016/j.bbapap.2016.01.012

Xiao, H., Chen, X., Chen, M., Tang, S., Zhao, X., Huan, L., 2004. Bovicin HJ50, a novel lantibiotic produced by Streptococcus bovis HJ50. Microbiology 150, 103–108. https://doi.org/10.1099/mic.0.26437-0

Yamada, T., Takashima, K., Miyazawa, T., 1978. Studies of unusual amino acids and their peptides.

IX. The synthetic study of bottromycins B1and B2. Bull. Chem. Soc. Jpn. 51, 878–883. https://doi.org/10.1246/bcsj.51.878

Yamada, T., Yagita, M., Kobayashi, Y., Sennari, G., Shimamura, H., Matsui, H., Horimatsu, Y., Hanaki, H., Hirose, T., Ōmura, S., Sunazuka, T., 2018. Synthesis and Evaluation of Antibacterial Activity of Bottromycins. J. Org. Chem. 83, 7135–7149. https://doi.org/10.1021/acs.joc.8b00045

Yamanaka, K., Reynolds, K.A., Kersten, R.D., Ryan, K.S., Gonzalez, D.J., Nizet, V., Dorrestein, P.C., Moore, B.S., 2014. Direct cloning and refactoring of a silent lipopeptide biosynthetic gene cluster yields the antibiotic taromycin A. Proc. Natl. Acad. Sci. 111, 1957–1962. https://doi.org/10.1073/pnas.1319584111

Yamazaki, K., de Mora, K., Saitoh, K., 2017. BioBrick-based 'Quick Gene Assembly' in vitro. Synth. Biol. 2. https://doi.org/10.1093/synbio/ysx003

Yang, J., Yan, R., Roy, A., Xu, D., Poisson, J., Zhang, Y., 2015. The I-TASSER Suite: protein structure and function prediction. Nat. Methods 12, 7–8. https://doi.org/10.1038/nmeth.3213

Yang, X., Lennard, K.R., He, C., Walker, M.C., Ball, A.T., Doigneaux, C., Tavassoli, A., van der Donk, W.A., 2018. A lanthipeptide library used to identify a protein–protein interaction inhibitor. Nat. Chem. Biol. 14, 375–380. https://doi.org/10.1038/s41589-018-0008-5

Yang, X., van der Donk, W.A., 2015. Post-translational Introduction of d -Alanine into Ribosomally Synthesized Peptides by the Dehydroalanine Reductase NpnJ. J. Am. Chem. Soc. 137, 12426–12429. https://doi.org/10.1021/jacs.5b05207

Yonemura, I., Nakada, K., Sato, A., Hayashi, J.-I., Fujita, K., Kaneko, S., Itaya, M., 2007. Direct cloning of full-length mouse mitochondrial DNA using a Bacillus subtilis genome vector. Gene 391, 171–177. https://doi.org/10.1016/J.GENE.2006.12.029

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. https://doi.org/10.1021/bi602663x

Young, T.A., Delagoutte, B., Endrizzi, J.A., Falick, A.M., Alber, T., 2003. Structure of Mycobacterium tuberculosis PknB supports a universal activation mechanism for Ser/Thr protein kinases. Nat. Struct. Biol. 10, 168–174. https://doi.org/10.1038/nsb897

Young, T.S., Dorrestein, P.C., Walsh, C.T., 2012. Codon randomization for rapid exploration of chemical space in thiopeptide antibiotic variants. Chem. Biol. 19, 1600–1610. https://doi.org/10.1016/j.chembiol.2012.10.013

Yu, J.-S., Madison-Antenucci, S., Steege, D.A., 2001. Translation at higher than an optimal level interferes with coupling at an intercistronic junction. Mol. Microbiol. 42, 821–834. https://doi.org/10.1046/j.1365-2958.2001.02681.x

Yu, Y., Duan, L., Zhang, Q., Liao, R., Ding, Y., Pan, H., Wendt-Pienkowski, E., Tang, G., Shen, B., Liu, W., 2009. Nosiheptide biosynthesis featuring a unique indole side ring formation on the characteristic thiopeptide framework. Am. Chem. Soc. Chem. Biol. 4, 855–864. https://doi.org/10.1021/cb900133x

Yu, Y., Mukherjee, S., van der Donk, W.A., 2015. Product Formation by the Promiscuous Lanthipeptide Synthetase ProcM is under Kinetic Control. J. Am. Chem. Soc. 137, 5140–5148. https://doi.org/10.1021/jacs.5b01409

Zambaldo, C., Luo, X., Mehta, A.P., Schultz, P.G., 2017. Recombinant Macrocyclic Lanthipeptides Incorporating Non-Canonical Amino Acids. J. Am. Chem. Soc. 139, 11646–11649. https://doi.org/10.1021/jacs.7b04159

Zhang, J., Shao, F., Li, Y., Cui, H., Chen, L., Li, H., Zou, Y., Long, C., Lan, L., Chai, J., Chen, S., Tang, X.,

Zhou, J.-M., 2007. A Pseudomonas syringae Effector Inactivates MAPKs to Suppress PAMP-Induced Immunity in Plants. Cell Host Microbe 1, 175–185. https://doi.org/10.1016/J.CHOM.2007.03.006

Zhang, L., Nguyen, H.C., Chipot, L., Piotrowski, E., Bertrand, C., Thibessard, A., Leblond, P., 2014. The adnAB Locus, Encoding a Putative Helicase-Nuclease Activity, Is Essential in Streptomyces. J. Bacteriol. 196, 2701–2708. https://doi.org/10.1128/JB.01513-14

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. Nat. Chem. Biol. 13, 607–609. https://doi.org/10.1038/nchembio.2341

Zhang, Q., Chen, D., Lin, J., Liao, R., Tong, W., Xu, Z., Liu, W., 2011. Characterization of NocL involved in thiopeptide nocathiacin I biosynthesis: a [4Fe-4S] cluster and the catalysis of a radical S-adenosylmethionine enzyme. J. Biol. Chem. 286, 21287–31294. https://doi.org/10.1074/jbc.M111.224832

Zhang, Q., Doroghazi, J.R., Zhao, X., Walker, M.C., van der Donk, W.A., 2015. Expanded Natural Product Diversity Revealed by Analysis of Lanthipeptide-Like Gene Clusters in Actinobacteria. Appl. Environ. Microbiol. 81, 4339–4350. https://doi.org/10.1128/AEM.00635-15

Zhang, Q., van der Donk, W.A., 2012. Catalytic promiscuity of a bacterial α-N-methyltransferase. Fed. Eur. Biochem. Soc. Lett. 586, 3391–3397. https://doi.org/10.1016/j.febslet.2012.07.050

Zhang, Q., van der Donk, W.A., Liu, W., 2012a. Radical-Mediated Enzymatic Methylation: A Tale of Two SAMS. Acc. Chem. Res. 45, 555–564. https://doi.org/10.1021/ar200202c

Zhang, Q., Yu, Y., Vélasquez, J.E., van der Donk, W.A., 2012b. Evolution of lanthipeptide synthetases. Proc. Natl. Acad. Sci. 109, 18361–18366. https://doi.org/10.1073/pnas.1210393109

Zhang, Y., Buchholz, F., Muyrers, J.P.P., Stewart, A.F., 1998. A new logic for DNA engineering using recombination in Escherichia coli. Nat. Genet. 20, 123–128. https://doi.org/10.1038/2417

Zhang, Y., Muyrers, J.P.P., Testa, G., Stewart, A.F., 2000. DNA cloning by homologous recombination in Escherichia coli. Nat. Biotechnol. 18, 1314–1317. https://doi.org/10.1038/82449

Zhang, Z., Mahanta, N., Hudson, G.A., Mitchell, D.A., van der Donk, W.A., 2017. Mechanism of a Class C Radical S -Adenosyl- l -methionine Thiazole Methyl Transferase. J. Am. Chem. Soc. 139, 18623–18631. https://doi.org/10.1021/jacs.7b10203

Zhao, X., van der Donk, W.A., 2016. Structural Characterization and Bioactivity Analysis of the Two-Component Lantibiotic Flv System from a Ruminant Bacterium. Cell Chem. Biol. 23, 246–256. https://doi.org/10.1016/j.chembiol.2015.11.014

Zhu, B., Cai, G., Hall, E.O., Freeman, G.J., 2018. In-Fusion™ assembly: seamless engineering of multidomain fusion proteins, modular vectors, and mutations. Biotechniques 43, 354–359. https://doi.org/10.2144/000112536

Zhu, Y., Gieselman, M.D., Zhou, H., Averin, O., van der Donk, W.A., 2003. Biomimetic studies on the mechanism of stereoselective lanthionine formation. Org. Biomol. Chem. 1, 3304. https://doi.org/10.1039/b304945k

Zhu, Y., Li, H., Long, C., Hu, L., Xu, H., Liu, L., Chen, S., Wang, D.-C., Shao, F., 2007. Structural Insights into the Enzymatic Mechanism of the Pathogenic MAPK Phosphothreonine Lyase. Mol. Cell 28, 899–913. https://doi.org/10.1016/J.MOLCEL.2007.11.011

Ziemert, N., Ishida, K., Quillardet, P., Bouchier, C., Hertweck, C., De Marsac, N.T., Dittmann, E.,

2008. Microcyclamide biosynthesis in two strains of Microcystis aeruginosa: From structure to genes and vice versa. Appl. Environ. Microbiol. 74, 1791–1797. https://doi.org/10.1128/AEM.02392-07

Zukher, I., Novikova, M., Tikhonov, A., Nesterchuk, M. V., Osterman, I.A., Djordjevic, M., Sergiev, P. V., Sharma, C.M., Severinov, K., 2014. Ribosome-controlled transcription termination is essential for the production of antibiotic microcin C. Nucleic Acids Res. 42, 11891–11902. https://doi.org/10.1093/nar/gku880

Appendices
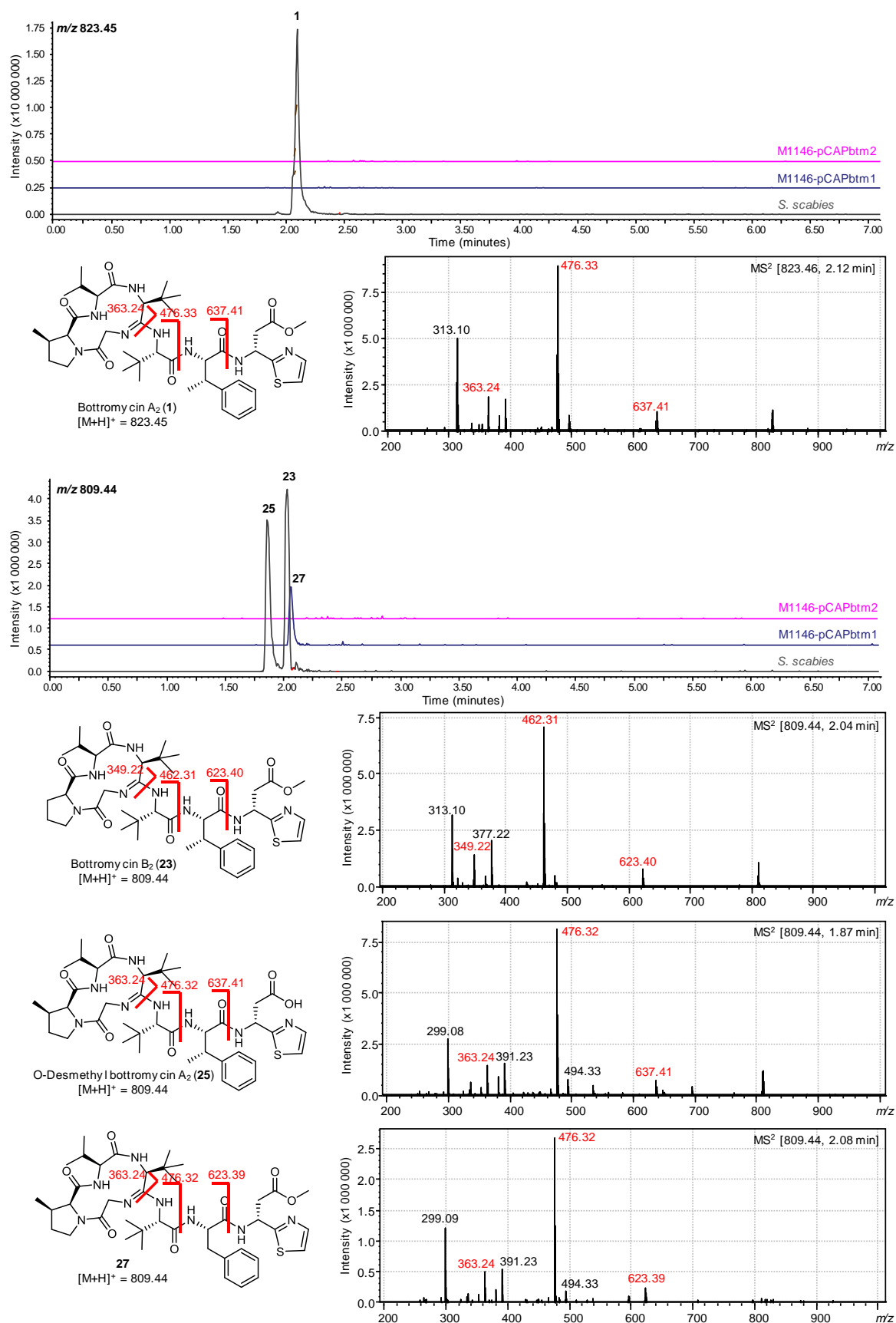
Figure 97. LC-MS^2 data for mature bottromycins **1**, **23**, **25**, and **27**.
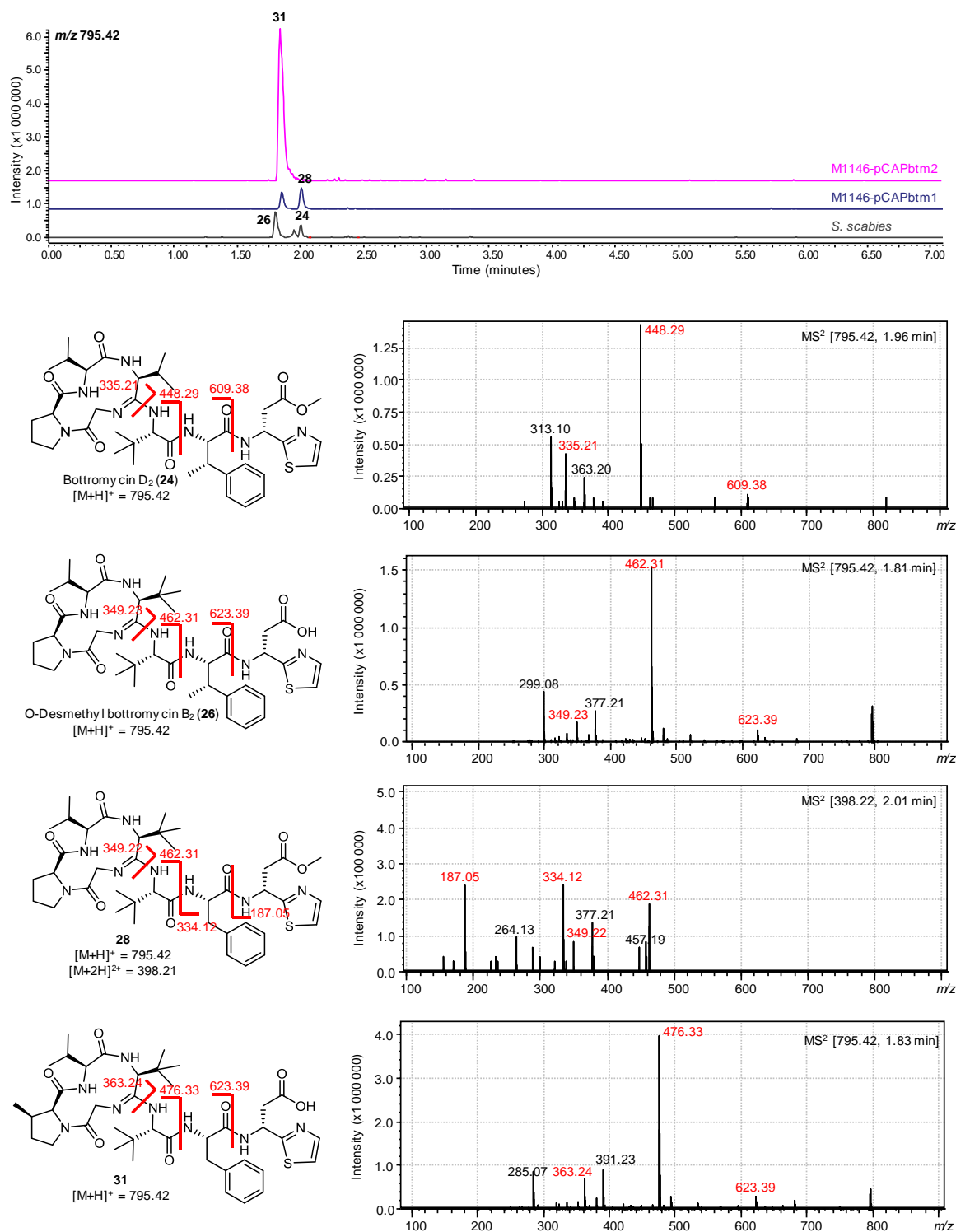
Figure 98. LC-MS^2 data for mature bottromycins **24**, **26**, **28**, and **31**.
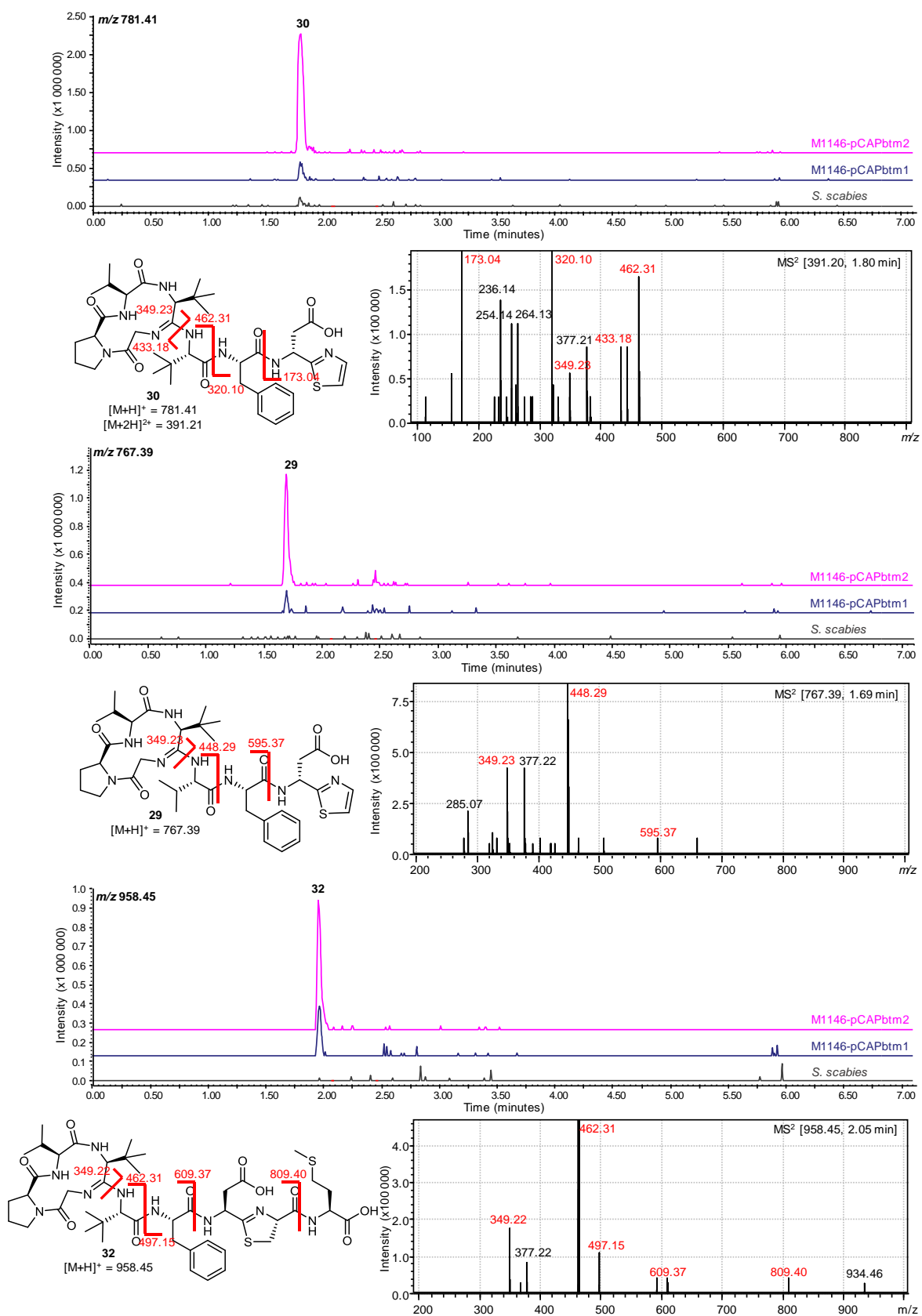
Figure 99. LC-MS² data for mature bottromycins **29** and **30**, and bottromycin-related metabolite **32**.
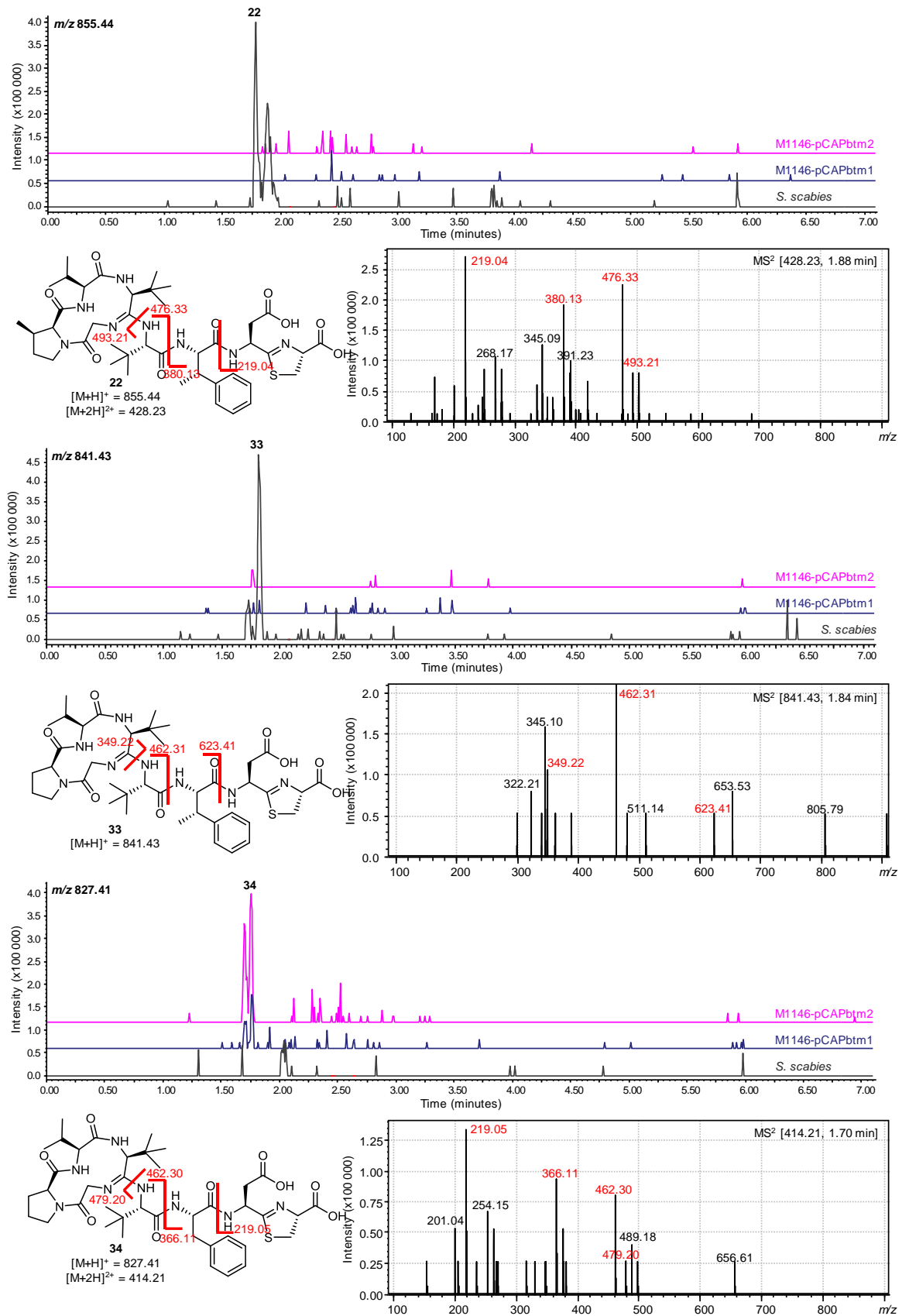
Figure 100. LC-MS^2 data for bottromycin-related metabolites **22**, **33**, and **34**.
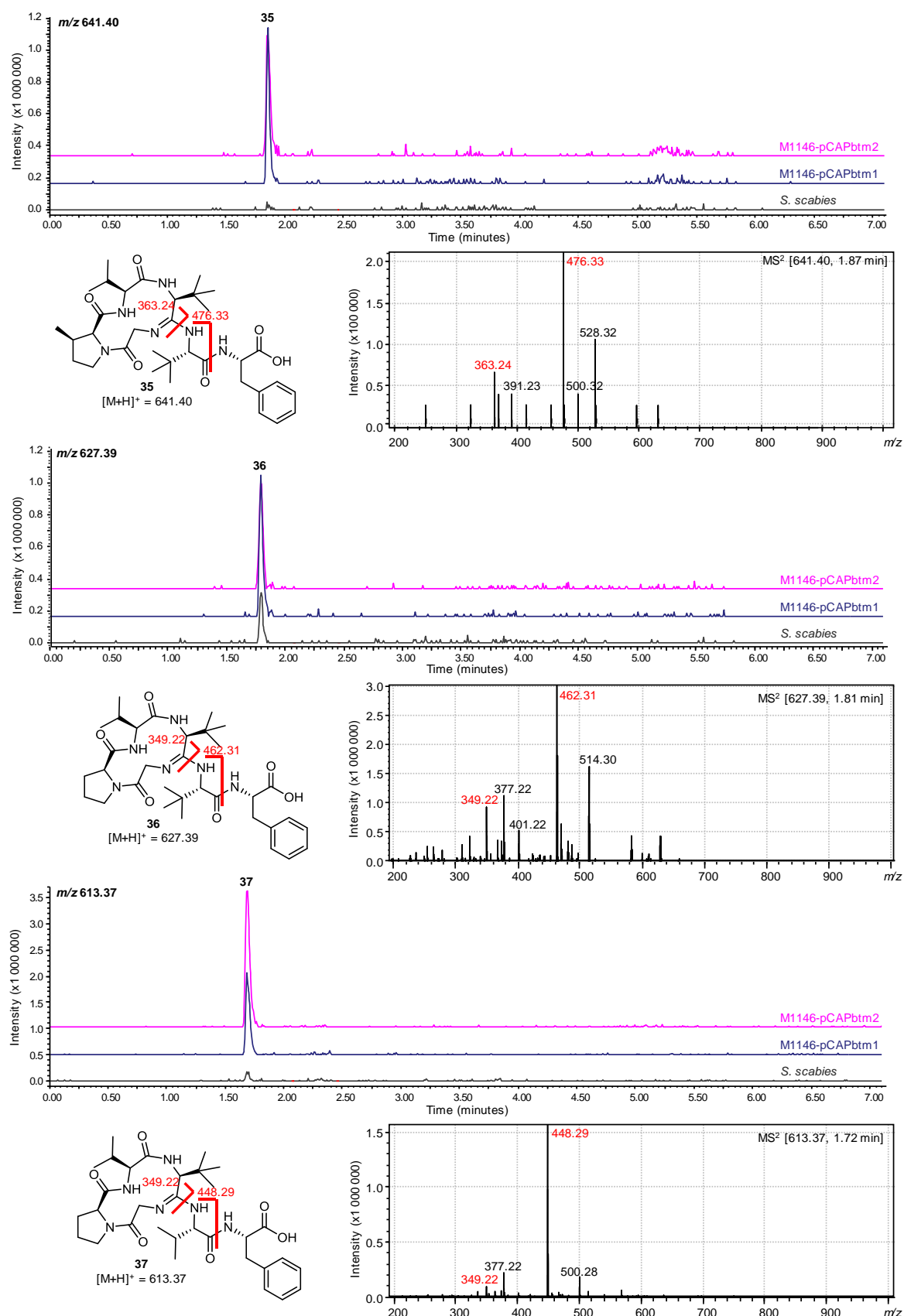
Figure 101. LC-MS^2 data for bottromycin-related metabolites **35**, **36**, and **37**.

Figure 102. LC-MS² data for bottromycin-related metabolites **38**, **39**, and **40**.
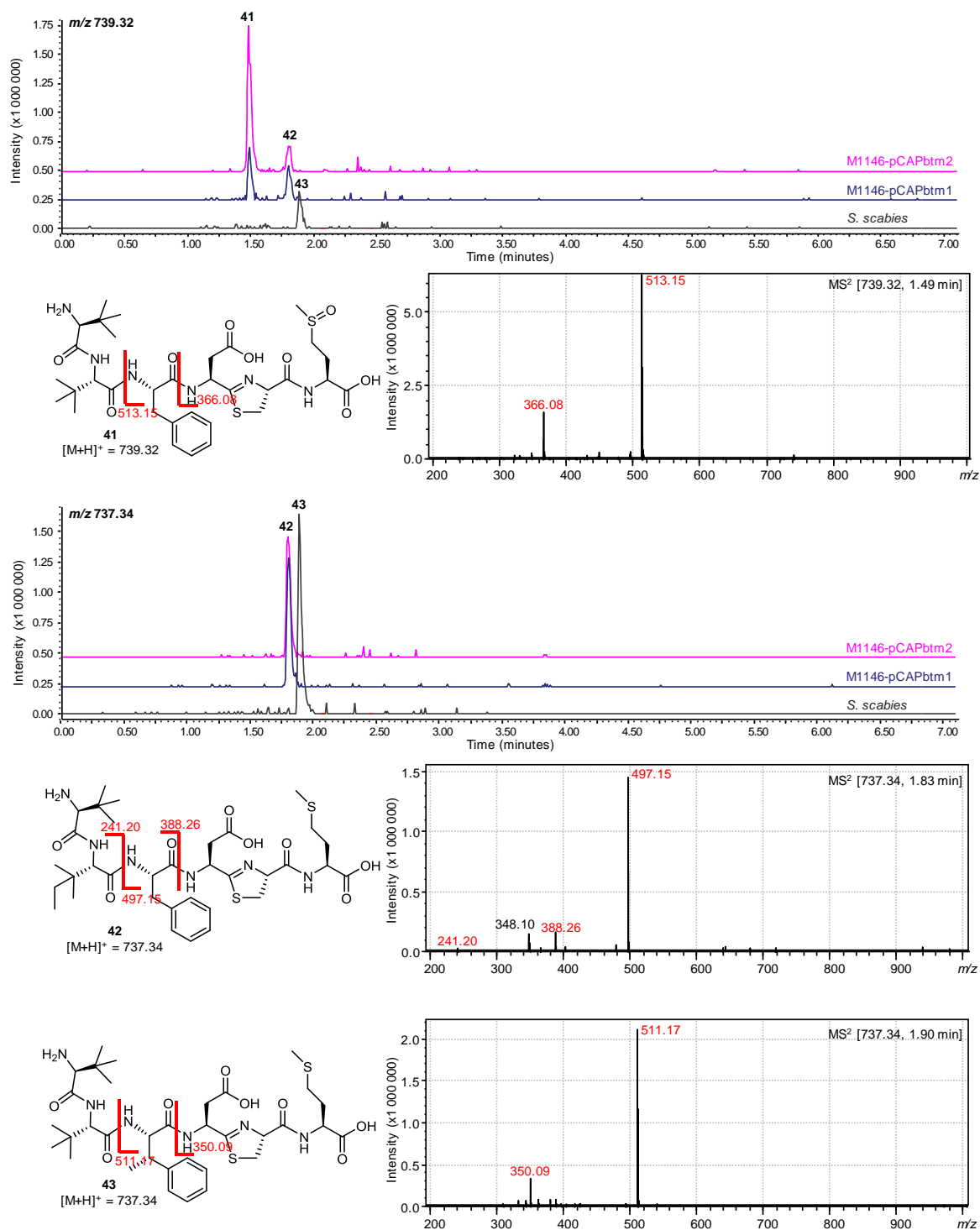
Figure 103. LC-MS² data for bottromycin-related metabolites **41**, **42**, and **43**. The peaks for **42** and **43** in the top chromatogram are from the +2 isotopes of these compounds. The precise site of additional methylation in **42** cannot be determined by this analysis, so the structure represents one possible proposal. This unusual additional methylation is predicted to also exist in compound **50**.
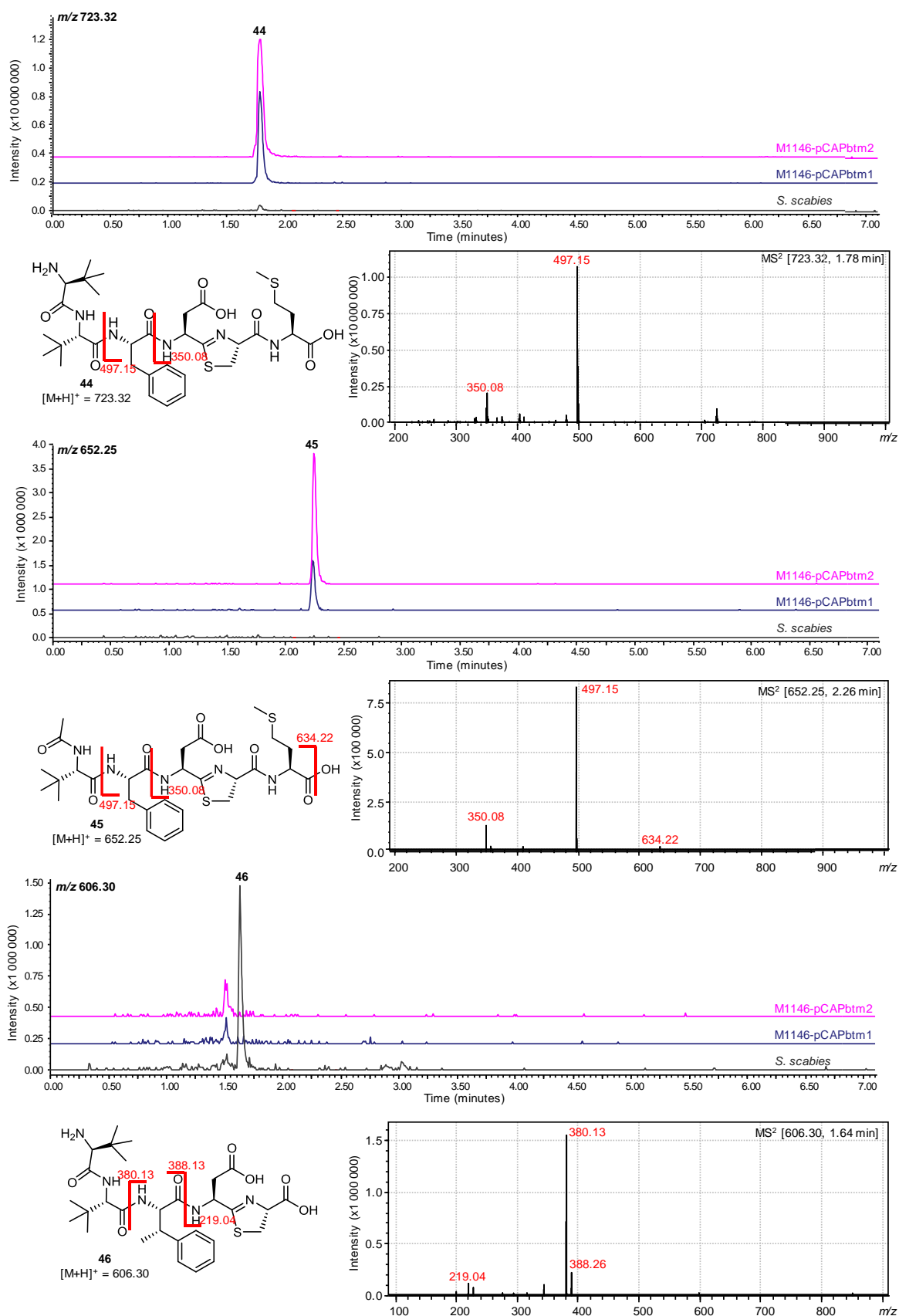
Figure 104. LC-MS^2 data for bottromycin-related metabolites **44**, **45**, and **46**.
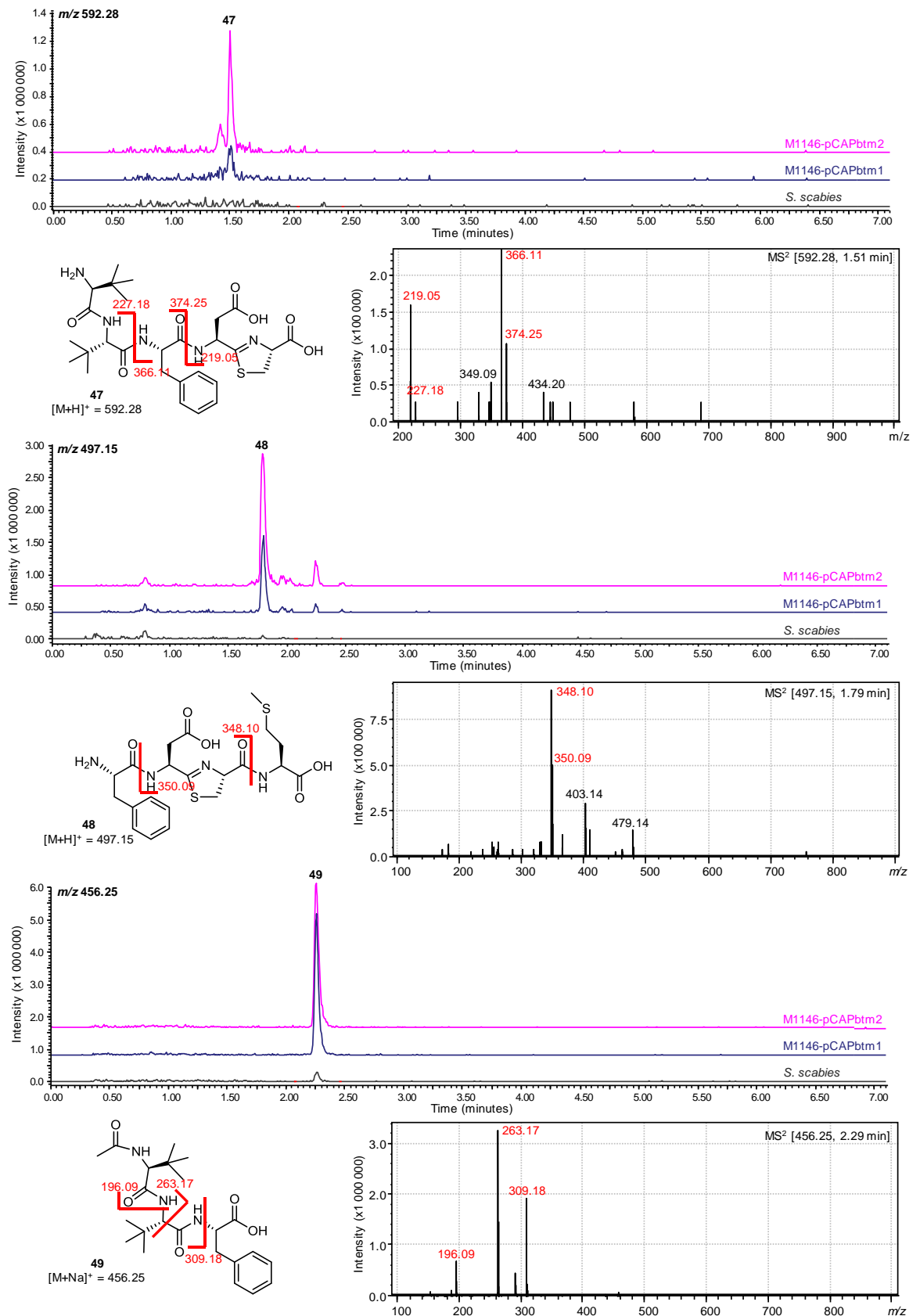
Figure 105. LC-MS^2 data for bottromycin-related metabolites **47**, **48**, and **49**. The unusual fragmentation for **49** is consistent with previously reported fragmentation patterns for [M+Na]$^+$ peptides, where abundant [b+M+OH]$^+$ fragments are observed (Grese et al., 1989; Newton and McLuckey, 2004).
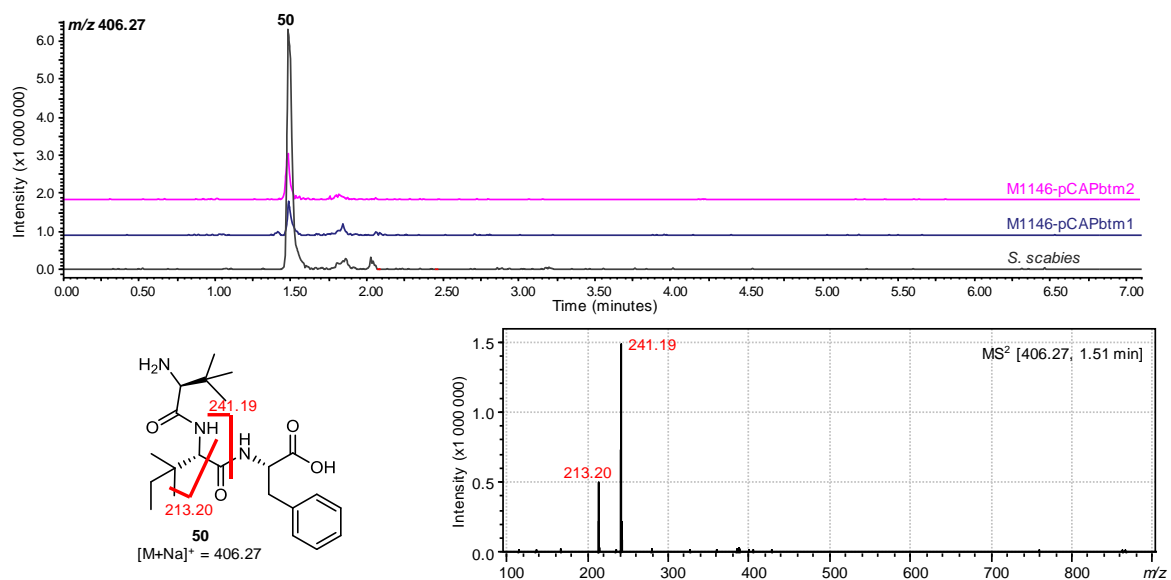
Figure 106. LC-MS^2 data for bottromycin-related metabolite **50**.
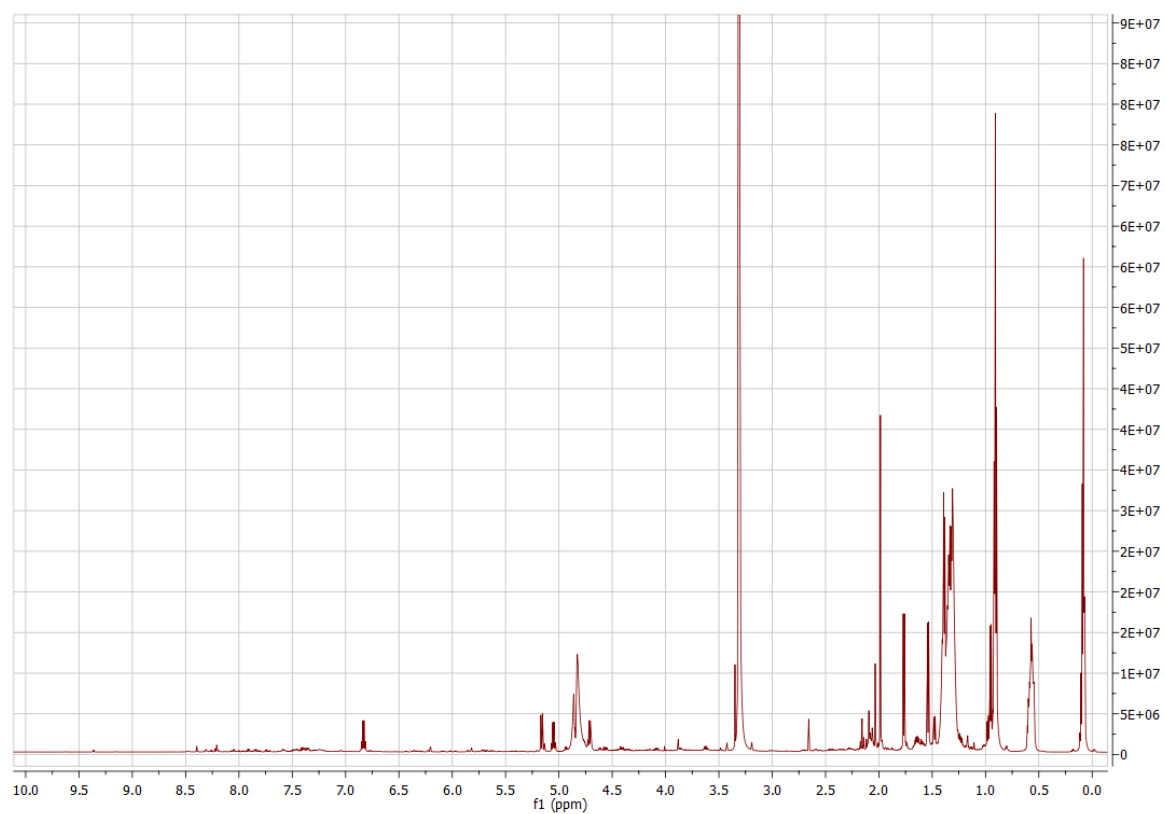
Figure 107. Proton NMR data for molecule **65**.
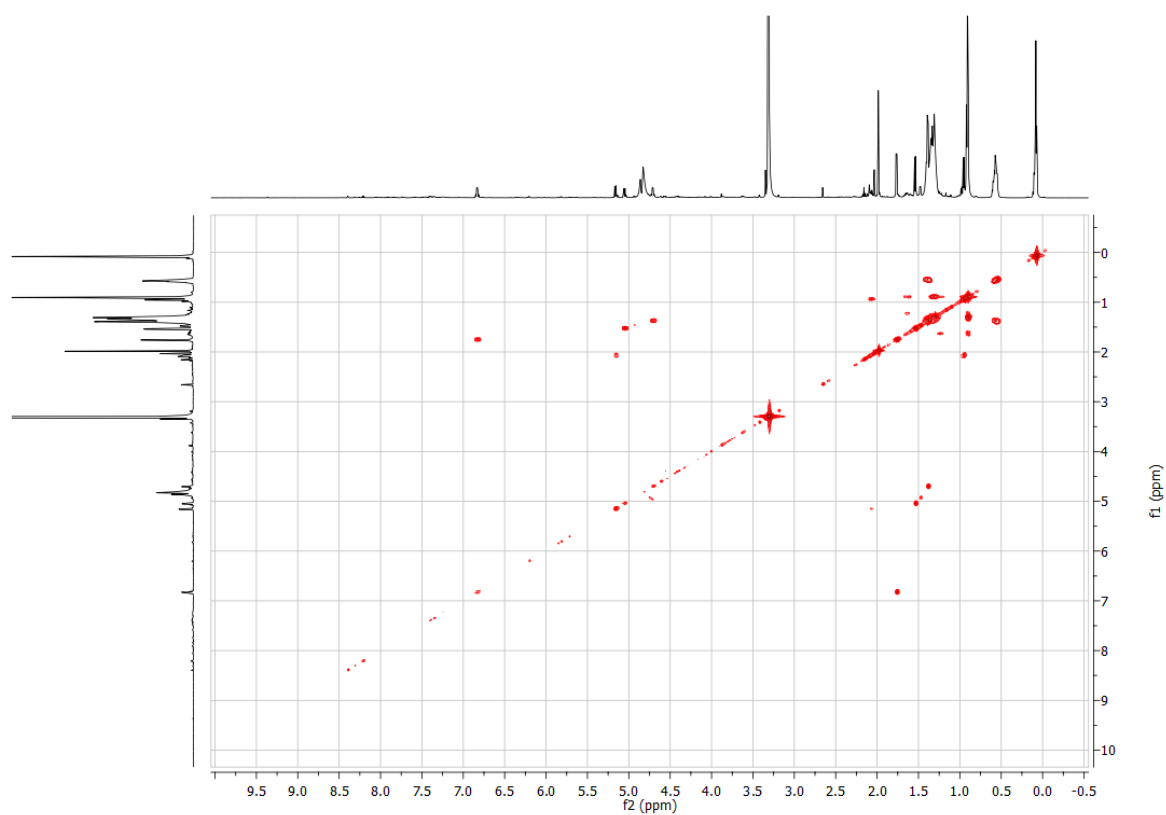
Figure 108. COSY NMR data for molecule **65**.
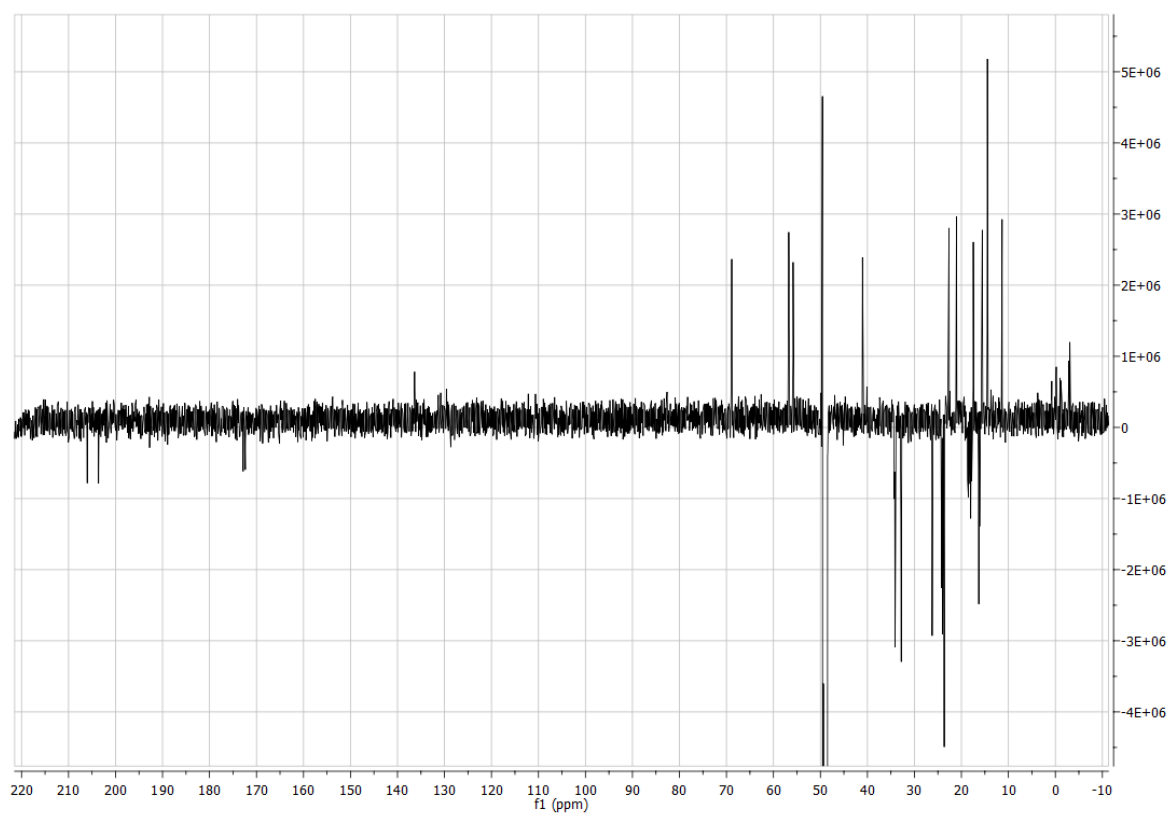
Figure 109. DEPTQ NMR data for molecule **65**.
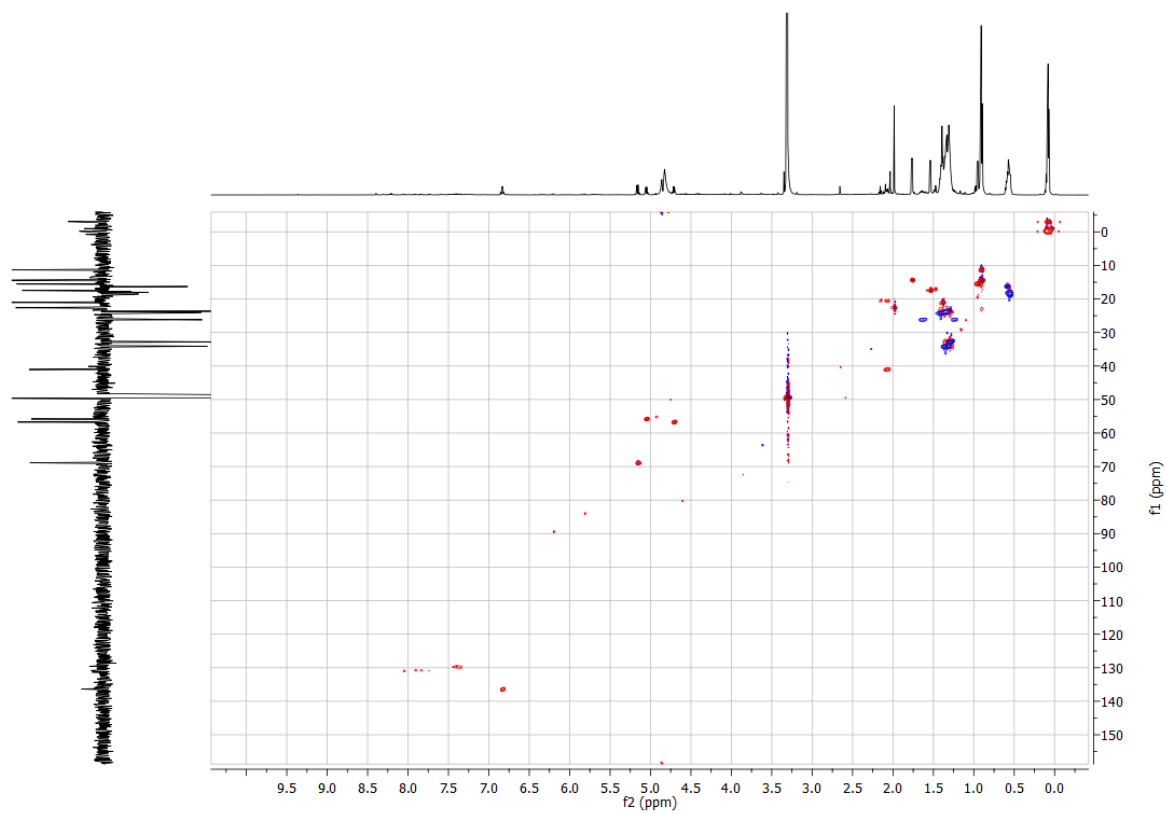
Figure 110. HSQC NMR data for molecule **65**.
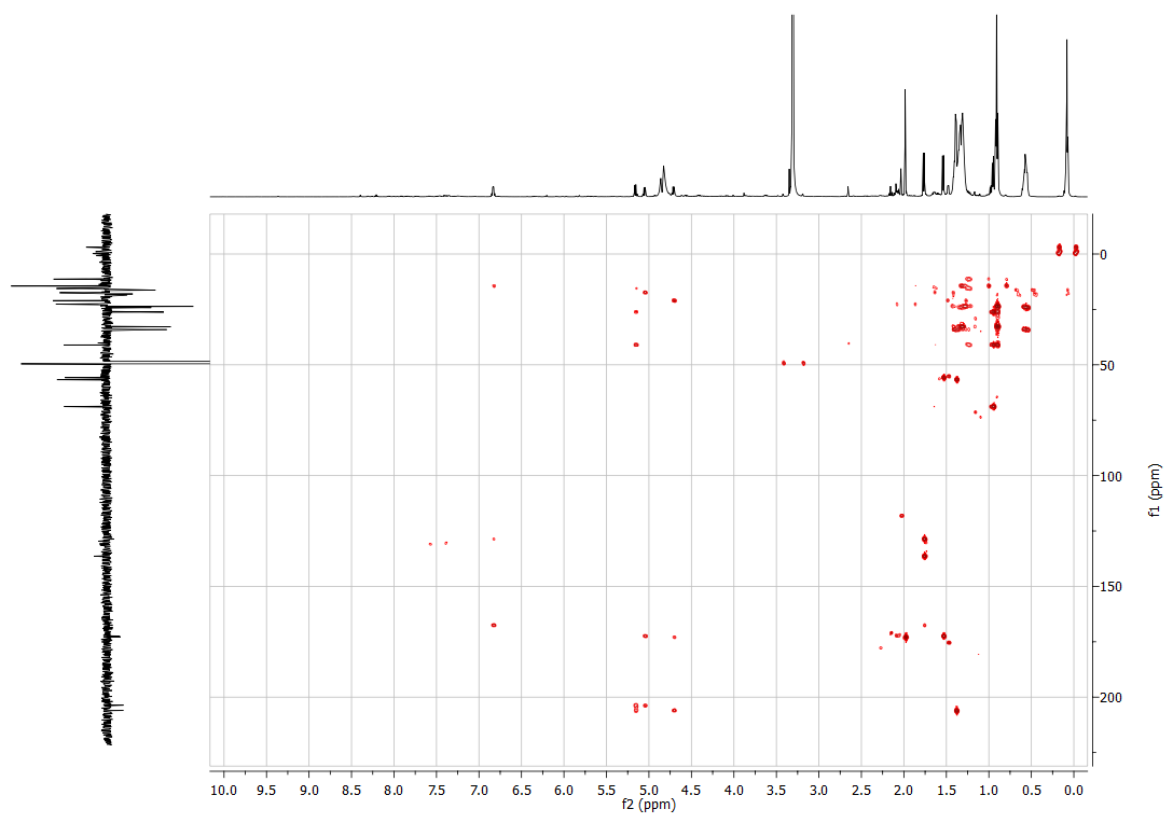
Figure 111. HMBC NMR data for molecule **65**.

Table 12. Selected triplicate untargeted metabolomic data sorted by p-value. MS data are coloured by intensity. All metabolites present in M1146-pCAP01 were removed.
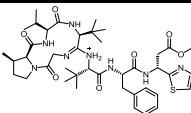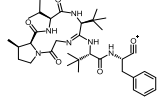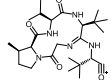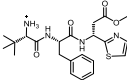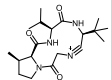
Ion m/z	Ion RT	PVal	M1146-pCAPbtm2			M1146-pCAPbtm1			M1146-pCAP01		
798.43	1.84	7.20E-10	242049	243260	230332	0	0	0	0	0	0
1009.50	2.22	7.80E-10	0	0	0	101437	100266	95743	0	0	0
902.75	5.68	1.40E-09	0	0	0	165611	165550	156128	0	0	0
466.24	1.46	4.60E-08	97021	86253	90116	0	0	0	0	0	0
709.30	1.74	9.70E-08	105609	115433	101179	0	0	0	0	0	0
981.44	1.97	1.20E-07	137249	142303	123991	0	0	0	0	0	0
795.42	1.85	1.50E-07	5233945	5200113	5731382	576290	182717	130894	0	0	0
781.41	1.79	1.80E-07	1814056	1874300	2058233	236477	175845	127563	0	0	0
366.11	1.50	3.30E-07	197279	195779	226375	0	0	0	0	0	0
391.71	1.79	3.40E-07	778593	690576	664891	0	0	0	0	0	0
990.48	2.03	3.60E-07	284502	240103	264432	0	0	0	0	0	0
404.23	1.57	5.30E-07	70272	84233	77722	0	0	0	0	0	0
745.31	1.79	6.50E-07	810322	909693	779220	1102958	1194006	1241712	0	0	0
877.75	5.66	8.70E-07	0	0	0	125424	129750	107384	0	0	0
599.36	1.57	1.00E-06	3450445	3286746	3741146	1811263	2150722	1792404	0	0	0
315.66	1.79	1.10E-06	145280	122432	148480	0	0	0	0	0	0
388.16	1.62	1.20E-06	0	0	0	175564	144474	151870	0	0	0
697.28	1.70	1.40E-06	187494	151536	170590	0	0	0	0	0	0
958.45	1.96	1.50E-06	826668	795632	785303	319643	186270	197217	0	0	0
472.22	2.25	1.50E-06	395822	445818	477186	471043	426315	432321	0	0	0
683.59	5.68	1.60E-06	68736	83222	82687	117351	107983	104202	0	0	0
408.25	1.95	1.60E-06	375540	452096	410558	253302	243088	260105	0	0	0
817.41	1.86	1.70E-06	544165	506715	437039	0	0	0	0	0	0
521.21	1.92	2.10E-06	176177	141200	152589	0	0	0	0	0	0
429.22	1.28	2.70E-06	0	0	0	95109	120325	113343	0	0	0
331.17	0.70	2.70E-06	132390	113531	110955	124291	138842	122464	0	0	0
613.37	1.68	3.90E-06	2604665	2354922	2587080	1539939	1152800	1105936	0	0	0
496.24	2.02	4.00E-06	200096	258344	230828	0	0	0	0	0	0
640.40	2.23	5.20E-06	86144	82944	92800	115840	93120	99564	0	0	0
803.38	1.79	5.50E-06	183321	151006	143080	0	0	0	0	0	0
726.32	1.79	5.50E-06	492773	571620	535312	305948	339620	245218	0	0	0
723.32	1.79	6.70E-06	8819825	9722209	10025299	6679022	5420708	4783407	0	0	0
836.40	1.96	7.00E-06	131252	102656	133146	0	0	0	0	0	0
339.66	1.79	7.10E-06	86080	109248	112320	0	0	0	0	0	0
877.72	5.28	7.20E-06	0	0	0	76808	101228	95577	0	0	0
655.25	2.24	7.60E-06	167405	126457	142740	0	0	0	0	0	0
406.27	1.48	8.00E-06	1171607	996673	1136745	834884	663948	650990	0	0	0
343.16	1.91	9.50E-06	328156	368788	294144	205987	167699	179613	0	0	0
391.21	1.79	1.00E-05	1697341	1398358	1329710	159588	0	0	0	0	0
462.31	1.80	1.20E-05	152589	138115	185329	0	0	0	0	0	0
767.40	1.70	1.30E-05	884176	784002	733477	146039	160497	0	0	0	0
434.26	2.25	1.30E-05	276319	251945	278682	186853	122432	140091	0	0	0
398.72	1.85	1.30E-05	2873866	2354555	3175334	178048	0	0	0	0	0
350.09	1.79	1.30E-05	533830	547395	557317	288531	148889	179699	0	0	0
711.27	1.67	1.70E-05	100567	129664	138918	0	0	0	0	0	0
320.11	1.85	1.80E-05	329950	245685	333757	0	0	0	0	0	0
348.10	1.79	2.10E-05	102464	72960	92736	0	0	0	0	0	0
275.19	2.67	2.20E-05	0	0	0	66112	59840	82784	0	0	0
433.19	1.71	2.40E-05	0	0	0	123312	90396	95281	0	0	0
344.18	0.59	2.60E-05	0	0	0	150145	146283	197818	0	0	0
403.14	1.79	2.80E-05	191552	181815	135269	0	0	0	0	0	0
624.25	1.59	3.20E-05	120395	95157	137655	0	0	0	0	0	0
899.71	4.71	3.40E-05	0	0	0	110660	96064	76287	0	0	0
432.29	4.42	3.60E-05	0	0	0	74858	66496	94844	0	0	0
362.67	1.79	4.10E-05	3098209	2868715	3010576	1250454	395328	372189	0	0	0
437.19	1.04	4.70E-05	0	0	0	146739	212119	164557	0	0	0
652.25	2.24	5.10E-05	2869978	3138472	3814067	1085101	530867	508712	0	0	0
497.15	1.79	5.20E-05	2006897	2237395	2253225	1181125	711988	502674	0	0	0
476.32	1.90	5.40E-05	607719	503617	437575	0	0	87570	0	0	0
787.32	2.47	5.60E-05	924204	890300	1056125	429547	175748	176031	0	0	0
294.16	1.19	5.60E-05	89777	70887	106415	0	0	0	0	0	0
739.32	1.49	6.00E-05	1433242	1060529	1357804	500038	322287	269986	0	0	0
422.21	1.21	6.30E-05	156897	134239	159191	159650	112436	157003	0	0	0
213.13	0.28	6.40E-05	76352	79424	108938	0	0	0	0	0	0
543.19	1.92	6.50E-05	181568	199831	265505	0	0	0	0	0	0
957.68	5.66	7.30E-05	0	0	0	96024	64626	94800	0	0	0
402.22	1.83	7.90E-05	0	0	0	101051	71488	71104	0	0	0
489.31	3.50	8.00E-05	0	0	0	85255	130791	119572	0	0	0
899.71	5.77	9.60E-05	415589	550702	1050544	1708136	1792930	1834028	0	0	0

Table 13. High resolution mass analysis of the new bottromycin-related metabolites.

Compound	Formula	Adduct	Predicted m/z	Observed m/z	Error (ppm)
6	$C_{41}H_{60}N_8O_7S$	H^+	809.4378	809.4380	0.2
7	$C_{40}H_{58}N_8O_7S$	H^+	795.4222	795.4211	-1.4
8	$C_{38}H_{54}N_8O_7S$	H^+	767.3909	767.3901	-1.0
9	$C_{39}H_{56}N_8O_7S$	$2H^{2+}$	391.2069	391.2072	0.8
10	$C_{40}H_{58}N_8O_7S$	$2H^{2+}$	398.2147	398.2150	0.8
11	$C_{45}H_{67}N_9O_{10}S_2$	H^+	958.4525	958.4518	-0.7
14	$C_{40}H_{58}N_8O_9S$	H^+	827.4120	ND[a]	
15	$C_{34}H_{52}N_6O_6$	H^+	641.4021	641.4014	-1.1
19	$C_{40}H_{61}N_7O_{10}S_2$	H^+	864.3994	864.3979	-1.7
20	$C_{35}H_{52}N_6O_9S_2$	H^+	765.3310	765.3298	-1.6
21	$C_{33}H_{50}N_6O_9S_2$	H^+	739.3153	739.3152	-0.1
22	$C_{34}H_{52}N_6O_8S_2$	H^+	737.3361	ND[a]	
25	$C_{29}H_{41}N_5O_8S_2$	H^+	652.2469	652.2463	-0.9
27	$C_{28}H_{41}N_5O_7S$	H^+	592.2799	592.2792	-1.2
28	$C_{21}H_{28}N_4O_6S_2$	H^+	497.1523	497.1532	1.8
29	$C_{23}H_{35}N_3O_5$	Na^+	456.2469	456.2469	0.0

[a] No data (ND) indicates the molecule was undetectable on the high resolution Synapt G2-Si mass spectrometer.

Table 14. High-resolution MS^2 fragmentation of molecule **27**.

| Formula	Calculated *m/z*	Measured *m/z*	Mass difference (Da)	Proposed structure[a]
$C_{41}H_{61}N_8O_7S^+$	809.4378	809.4380	0.0002	
$C_{34}H_{51}N_6O_5^+$	623.3915	623.3917	0.0002	
$C_{25}H_{42}N_5O_4^+$	476.3231	476.3222	-0.0009	
$C_{22}H_{31}N_4O_4S^+$	447.2061	447.2057	-0.0004	
$C_{19}H_{31}N_4O_3^+$	363.2391	363.2392	0.0001	
$C_{16}H_{20}N_3O_3S^+$	334.1220	334.1218	-0.0002	
$C_{13}H_{22}N_3O_3^+$	268.1656	268.1654	-0.0002	
$C_7H_{11}N_2O_2S^+$	187.0536	187.0535	-0.0001	
$C_8H_{13}N_2O_2^+$	169.0972	169.0973	0.0001	
$C_7H_{13}N_2O^+$	141.1022	141.1027	0.0005	
$C_6H_4NOS^+$	138.0008	138.0012	0.0004	
$C_5H_{12}N^+$	86.0964	86.0965	0.0001	
$C_5H_{10}N^+$	84.0808	84.0812	0.0004	
$C_4H_8N^+$	72.0808	72.0811	0.0003	

[a] The top line of the table provides the MS data and structure for the parent molecule.

Table 15. High-resolution MS^2 fragmentation of molecule 28.

Formula	Calculated m/z	Measured m/z	Mass difference (Da)	Proposed structure[a]
$C_{40}H_{59}N_8O_7S^+$	795.4222	795.4211	-0.0011	
$C_{33}H_{49}N_6O_5^+$	609.3759	609.3751	-0.0008	
$C_{24}H_{40}N_5O_4^+$	462.3075	462.3062	-0.0013	
$C_{22}H_{31}N_4O_4S^+$	447.2061	447.2043	-0.0018	
$C_{18}H_{29}N_4O_3^+$	349.2234	349.2227	-0.0007	
$C_{16}H_{20}N_3O_3S^+$	334.122	334.1225	0.0005	
$C_{12}H_{20}N_3O_3^+$	254.1499	254.1491	-0.0008	
$C_7H_{11}N_2O_2S^+$	187.0536	187.0546	0.0010	
$C_7H_8NO_2S^+$	170.0270	170.0267	-0.0003	
$C_7H_{11}N_2O_2^+$	155.0815	155.0815	0.0000	
$C_6H_4NOS^+$	138.0008	138.0013	0.0005	
$C_6H_{11}N_2O^+$	127.0866	127.0871	0.0005	
$C_8H_{10}N^+$	120.0808	120.0813	0.0005	
$C_5H_{12}N^+$	86.0964	86.0962	-0.0002	
$C_4H_{10}N^+$	72.0808	72.0804	-0.0004	
$C_4H_8N^+$	70.0651	70.0647	-0.0004	

[a] The top line of the table provides the MS data and structure for the parent molecule.

Table 16. High-resolution MS^2 fragmentation of molecule **29**.

| Formula	Calculated *m/z*	Measured *m/z*	Mass difference (Da)	Proposed structure[a]
$C_{38}H_{55}N_8O_7S^+$	767.3909	767.3901	-0.0008	
$C_{20}H_{27}N_4O_4S^+$	595.3602	595.3603	0.0001	
$C_{23}H_{38}N_5O_4^+$	448.2918	448.2921	0.0003	
$C_{18}H_{29}N_4O_3^+$	349.2234	349.2235	0.0001	
$C_{12}H_{20}N_3O_3^+$	254.1499	254.1493	-0.0006	
$C_7H_{11}N_2O_2^+$	155.0815	155.0813	-0.0002	
$C_6H_{11}N_2O^+$	127.0866	127.0876	0.0010	
$C_8H_{10}N^+$	120.0808	120.0817	0.0009	
$C_5H_{12}N^+$	86.0964	86.0965	0.0001	
$C_4H_{10}N^+$	72.0808	72.0797	-0.0011	
$C_4H_8N^+$	70.0651	70.0651	0.0000	

[a] The top line of the table provides the MS data and structure for the parent molecule.

Table 17. High-resolution MS^2 fragmentation of molecule **30**.

Formula	Calculated *m/z*	Measured *m/z*	Mass difference (Da)	Proposed structure[a]
$C_{39}H_{57}N_8O_7S^+$	781.4065	781.4077	0.0012	
$C_{33}H_{49}N_6O_5^+$	609.3759	609.3746	-0.0013	
$C_{24}H_{40}N_5O_4^+$	462.3075	462.3075	0.0000	
$C_{18}H_{29}N_4O_3^+$	349.2234	349.2238	0.0004	
$C_{15}H_{18}N_3O_3S^+$	320.1063	320.1062	-0.0001	
$C_{12}H_{20}N_3O_3^+$	254.1499	254.1500	0.0001	
$C_6H_9N_2O_2S^+$	173.0379	173.0377	-0.0002	
$C_6H_6NO_2S^+$	156.0114	156.0117	0.0003	
$C_7H_{11}N_2O_2^+$	155.0815	155.0819	0.0004	
$C_6H_{11}N_2O^+$	127.0866	127.0870	0.0004	
$C_8H_{10}N^+$	120.0808	120.0808	0.0000	
$C_5H_{12}N^+$	86.0964	86.0967	0.0003	
$C_3H_4NS^+$	86.0059	86.0071	0.0012	
$C_4H_8N^+$	70.0651	70.0654	0.0003	

[a] The top line of the table provides the MS data and structure for the parent molecule.

Table 18. High-resolution MS^2 fragmentation of molecule 31.

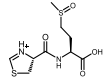| Formula	Calculated *m/z*	Measured *m/z*	Mass difference (Da)	Proposed structure[a]
$C_{40}H_{59}N_8O_7S^+$	795.4222	795.4233	0.0011	
$C_{34}H_{51}N_6O_5^+$	623.3915	623.3911	-0.0004	
$C_{25}H_{42}N_5O_4^+$	476.3231	476.3232	0.0001	
$C_{21}H_{29}N_4O_4S^+$	433.1904	433.1884	-0.0020	
$C_{19}H_{31}N_4O_3^+$	363.2391	363.2391	0.0000	
$C_{15}H_{18}N_3O_3S^+$	320.1063	320.1060	-0.0003	
$C_{15}H_{15}N_2O_3S^+$	303.0798	303.0788	-0.0010	
$C_{13}H_{22}N_3O_3^+$	268.1656	268.1651	-0.0005	
$C_6H_9N_2O_2S^+$	173.0379	173.0377	-0.0002	
$C_8H_{13}N_2O_2^+$	169.0972	169.0975	0.0003	
$C_7H_{13}N_2O^+$	141.1022	141.1028	0.0006	
$C_8H_{10}N^+$	120.0808	120.0809	0.0001	
$C_5H_{12}N^+$	86.0964	86.0968	0.0004	
$C_5H_{10}N^+$	84.0808	84.0808	0.0000	
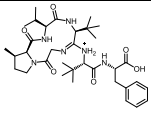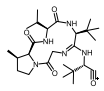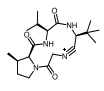
[a] The top line of the table provides the MS data and structure for the parent molecule.

Table 19. High-resolution MS^2 fragmentation of molecule **32**.

Formula	Calculated *m/z*	Measured *m/z*	Mass difference (Da)	Proposed structure[a]
$C_{45}H_{68}N_9O_{10}S_2^+$	958.4525	958.4518	-0.0007	
$C_{21}H_{29}N_4O_6S_2^+$	497.1523	497.1508	-0.0015	
$C_{24}H_{40}N_5O_4^+$	462.3075	462.3082	0.0007	
$C_{12}H_{20}N_3O_5S_2^+$	350.0839	350.0840	0.0001	
$C_{18}H_{29}N_4O_3^+$	349.2234	349.2218	-0.0016	
$C_{12}H_{20}N_3O_3^+$	254.1499	254.1499	0.0000	
$C_7H_{11}N_2O_2^+$	155.0815	155.0813	-0.0002	
$C_6H_{11}N_2O^+$	127.0866	127.0868	0.0002	
$C_8H_{10}N^+$	120.0812	120.0808	-0.0004	
$C_5H_{12}N^+$	86.0964	86.0974	0.0010	
$C_4H_8N^+$	70.0651	70.0650	-0.0001	
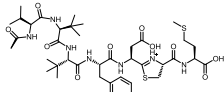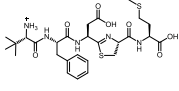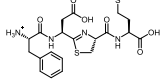
[a] The top line of the table provides the MS data and structure for the parent molecule

Table 20. High-resolution MS^2 fragmentation of molecule 35.

| Formula	Calculated *m/z*	Measured *m/z*	Mass difference (Da)	Proposed structure[a]
$C_{34}H_{53}N_6O_6^+$	641.4021	641.4014	-0.0007	
$C_{25}H_{42}N_5O_4^+$	476.3231	476.3249	0.0018	
$C_{19}H_{31}N_4O_3^+$	363.2391	363.2384	-0.0007	
$C_{13}H_{22}N_3O_3^+$	268.1656	268.1646	-0.0010	
$C_8H_{13}N_2O_2^+$	169.0972	169.0974	0.0002	
$C_7H_{13}N_2O^+$	141.1022	141.1024	0.0002	
$C_8H_{10}N^+$	120.0808	120.0781	-0.0027	
$C_5H_{12}N^+$	86.0964	86.0950	-0.0014	
$C_5H_{10}N^+$	84.0808	84.0811	0.0003	
$C_4H_8N^+$	72.0808	72.0800	-0.0008	
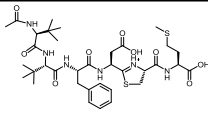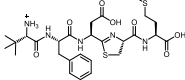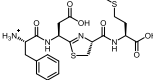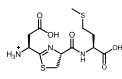
[a] The top line of the table provides the MS data and structure for the parent molecule.

Table 21. High-resolution MS^2 fragmentation of molecule **39**.

| Formula	Calculated *m/z*	Measured *m/z*	Mass difference (Da)	Proposed structure[a]
$C_{40}H_{62}N_7O_{10}S_2^+$	864.3994	864.3979	-0.0015	
$C_{27}H_{40}N_5O_7S_2^+$	610.2364	610.2314	-0.0050	
$C_{21}H_{29}N_4O_6S_2^+$	497.1523	497.1534	0.0011	
$C_{19}H_{34}N_3O_4^+$	368.2544	368.2527	-0.0017	
$C_9H_{15}N_2O_3S_2^+$	263.0519	263.0504	-0.0015	
$C_{13}H_{23}N_2O_3^+$	255.1703	255.1706	0.0003	
$C_7H_{12}NO_2^+$	142.0863	142.0844	-0.0019	
$C_8H_{10}N^+$	120.0808	120.0811	0.0003	
$C_5H_{12}N^+$	86.0964	86.0968	0.0004	
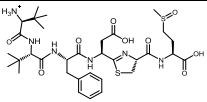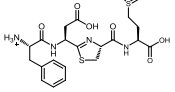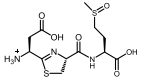
[a] The top line of the table provides the MS data and structure for the parent molecule.

Table 22. High-resolution MS^2 fragmentation of molecule **40**.

Formula	Calculated *m/z*	Measured *m/z*	Mass difference (Da)	Proposed structure[a]
$C_{35}H_{53}N_6O_9S_2^+$	765.3310	765.3298	-0.0012	
$C_{27}H_{40}N_5O_7S_2^+$	610.2332	610.2364	0.0032	
$C_{21}H_{29}N_4O_6S_2^+$	497.1523	497.1514	-0.0009	
$C_{12}H_{20}N_3O_5S_2$	350.0839	350.0839	0.0000	
$C_{14}H_{25}N_2O_3^+$	269.1860	269.1867	0.0007	
$C_9H_{15}N_2O_3S_2^+$	263.0519	263.0517	-0.0002	
$C_7H_{14}NO^+$	128.1070	128.1050	-0.0020	
$C_8H_{10}N^+$	120.0808	120.0805	-0.0003	
$C_3H_6NO_2^+$	88.0393	88.0395	0.0002	
$C_5H_{12}N^+$	86.0964	86.0964	0.0000	
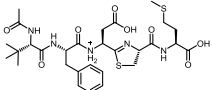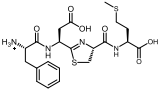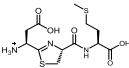
[a] The top line of the table provides the MS data and structure for the parent molecule.

Table 23. High-resolution MS^2 fragmentation of molecule 41.

| Formula	Calculated m/z	Measured m/z	Mass difference (Da)	Proposed structure[a]
$C_{33}H_{51}N_6O_9S_2^+$	739.3153	739.3152	-0.0001	
$C_{21}H_{29}N_4O_7S_2^+$	513.1472	513.1479	0.0007	
$C_{12}H_{20}N_3O_6S_2^+$	366.0788	366.0784	-0.0004	
$C_9H_{15}N_2O_4S_2^+$	279.0468	279.0448	-0.0020	
$C_{11}H_{23}N_2O^+$	199.1805	199.1817	0.0012	
$C_8H_{10}N^+$	120.0808	120.0806	-0.0002	
$C_5H_{12}N^+$	86.0964	86.0969	0.0005	
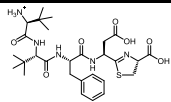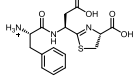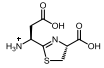
[a] The top line of the table provides the MS data and structure for the parent molecule.

Table 24. High-resolution MS^2 fragmentation of molecule **45**.

Formula	Calculated *m/z*	Measured *m/z*	Mass difference (Da)	Proposed structure[a]
$C_{29}H_{42}N_5O_8S_2^+$	652.2469	652.2463	-0.0006	
$C_{21}H_{29}N_4O_6S_2^+$	497.1523	497.1512	-0.0011	
$C_{12}H_{20}N_3O_5S_2^+$	350.0839	350.0829	-0.0010	
$C_9H_{15}N_2O_3S_2^+$	263.0519	263.0510	-0.0009	
$C_7H_{14}NO^+$	128.1070	128.1085	0.0015	
$C_8H_{10}N^+$	120.0808	120.0812	0.0004	
$C_5H_{12}N^+$	86.0964	86.0967	0.0003	

[a] The top line of the table provides the MS data and structure for the parent molecule.

Table 25. High-resolution MS^2 fragmentation of molecule 47.

Formula	Calculated *m/z*	Measured *m/z*	Mass difference (Da)	Proposed structure[a]
$C_{28}H_{42}N_5O_7S^+$	592.2799	592.2792	-0.0007	
$C_{16}H_{20}N_3O_5S^+$	366.1118	366.1125	0.0007	
$C_7H_{11}N_2O_4S^+$	219.0434	219.0433	-0.0001	
$C_{11}H_{23}N_2O^+$	199.1805	199.1811	0.0006	
$C_4H_6NO_2S^+$	132.0114	132.0117	0.0003	
$C_8H_{10}N^+$	120.0808	120.0809	0.0001	
$C_3H_6NO_2^+$	88.0393	88.0394	0.0001	
$C_5H_{12}N^+$	86.0964	86.0967	0.0003	

[a] The top line of the table provides the MS data and structure for the parent molecule.
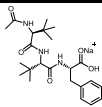
Table 26. High-resolution MS^2 fragmentation of molecule **48**.

Formula	Calculated *m/z*	Measured *m/z*	Mass difference (Da)	Proposed structure[a]
$C_{21}H_{29}N_4O_6S_2{}^+$	497.1523	497.1532	0.0009	
$C_{12}H_{20}N_3O_5S_2{}^+$	350.0839	350.0822	-0.0017	
$C_{16}H_{18}N_3O_4S^+$	348.1013	348.1008	-0.0005	
$C_9H_{15}N_2O_3S_2{}^+$	263.0519	263.0539	0.0020	
$C_5H_{12}NO_2S^+$	150.0583	150.0574	-0.0009	
$C_8H_{10}N^+$	120.0808	120.0808	0.0000	
$C_3H_6NO_2{}^+$	88.0393	88.0407	0.0014	

[a] The top line of the table provides the MS data and structure for the parent molecule.

Table 27. High-resolution MS^2 fragmentation of molecule 49.

Formula	Calculated *m/z*	Measured *m/z*	Mass difference (Da)	Proposed structure[a]
$C_{23}H_{35}N_3NaO_5^+$	456.2469	456.2469	0.0000	
$C_{14}H_{26}N_2NaO_4^+$	309.1785	309.1787	0.0002	
$C_{13}H_{24}N_2NaO_2^+$	263.1726	263.1730	0.0004	
$C_8H_{15}NNaO_3^+$	196.0944	196.0945	0.0001	
$C_6H_{13}NNaO_2^+$	154.0851	154.0838	-0.0013	

[a] The top line of the table provides the MS data and structure for the parent molecule.

ACS SyntheticBiology

Rapid and Robust Yeast-Mediated Pathway Refactoring Generates Multiple New Bottromycin-Related Metabolites

Tom H. Eyles, Natalia M. Vior, and Andrew W. Truman*

Department of Molecular Microbiology, John Innes Centre, Norwich, NR4 7UH, U.K.

S *Supporting Information*

ABSTRACT: Heterologous expression of biosynthetic gene clusters (BGCs) represents an attractive route to the production of new natural products, but is often hampered by poor yields. It is therefore important to develop tools that enable rapid refactoring, gene insertion/deletion, and targeted mutations in BGCs. Ideally, these tools should be highly efficient, affordable, accessible, marker free, and flexible for use with a wide range of BGCs. Here, we present a one-step yeast-based method that enables efficient, cheap, and flexible modifications to BGCs. Using the BGC for the antibiotic bottromycin, we showcase multiple modifications including refactoring, gene deletions and targeted mutations. This facilitated the construction of an inducible, riboswitch-controlled pathway that achieved a 120-fold increase in pathway productivity in a heterologous streptomycete host. Additionally, an unexpected biosynthetic bottleneck resulted in the production of a suite of new bottromycin-related metabolites.

KEYWORDS: *bottromycin, refactor, biosynthesis, streptomyces, riboswitch, natural product*
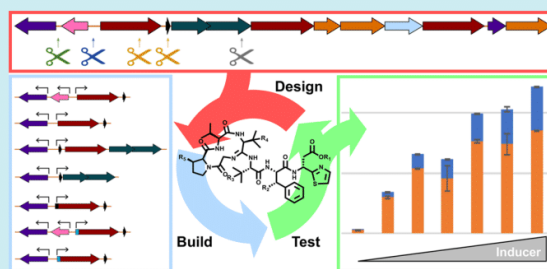
Bacteria produce an incredible variety of bioactive secondary metabolites, including the majority of clinically used antibiotic classes and a myriad of compounds that are used across medicine.[1] The study of natural product biosynthetic pathways is often hindered by low yields or the genetic intractability of the native producing organism. Therefore, the heterologous expression of biosynthetic gene clusters (BGCs) can provide multiple benefits. It facilitates genetic modification of the gene cluster, aids with the identification of products from uncharacterized BGCs,[2] and allows for heterologous expression in a defined background, such as *Streptomyces coelicolor* M1146,[3] where multiple BGCs have been removed. However, there are significant issues associated with heterologous expression. Regulatory pathways are not fully conserved between species, so moving a BGC between organisms can drastically change the productivity of a pathway. It is therefore often necessary to refactor the regulation of a BGC using well-characterized promoters to enable expression in a heterologous host.[4] However, there are limited methods to efficiently refactor pathways in a scarless and sequence-independent manner.

There are a variety of *in vitro* and *in vivo* methods for cloning and engineering biosynthetic gene clusters. *In vitro* methods have proven to be very rapid and versatile,[5−9] but can be limited by various factors, such as the availability of type II restriction enzymes that do not cut within the BGC when using Golden Gate cloning,[6−8] and drops in efficiency for Gibson assembly of regulatory regions and GC-rich DNA.[9,10] *Saccharomyces cerevisiae* (yeast) is highly recombinogenic,[11] which makes it a valuable tool for *in vivo* cloning and

engineering of large regions of DNA.[12] This is arguably more versatile than *in vitro* methods, and yeast can assemble and maintain over 100 kbp of high GC prokaryote DNA.[13] BGC modification systems need to handle complex changes to long stretches of DNA. Some methods do this by completely rebuilding biosynthetic pathways into scaffolds,[14,15] although this can require multiple assembly steps. Alternatively, a recently reported approach uses a yeast-expressed CRISPR-Cas9 system to cut and modify target regions.[16] BGC regulatory elements can also be coupled to selectable markers,[17,18] although this limits the types of modifications that can be made.

Double-strand breaks in DNA increase the recombination efficiency of yeast,[19] and site-specific repair of these breaks enables the addition, deletion, and modification of DNA. Here, we integrate transformation-associated recombination (TAR) cloning in yeast,[12,17] yeast-mediated engineering,[20,21] and recently developed regulatory tools[22,23] into a BGC modification process that we applied to the bottromycin gene cluster.[24−27] We show how commercially available restriction enzymes, PCR products, and single stranded oligonucleotides can be used to simultaneously refactor, delete genes, make targeted mutations, and reorganize gene order in the bottromycin cluster. This approach is viable for high GC gene fragments, allows for a variety of scarless changes to be made to a BGC in a single step, and uses cheap parts. The
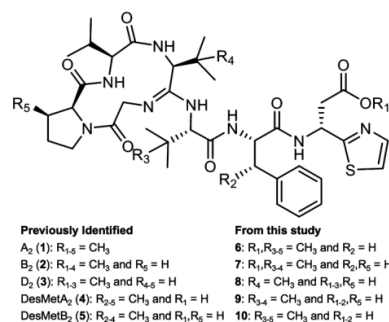
refactoring and remodeling of this gene cluster led to a BGC that was 120-fold more productive in a heterologous host and yielded a variety of new bottromycins.

■ RESULTS AND DISCUSSION

Bottromycin is a heavily modified RiPP antibiotic characterized by a macrocyclic amidine and a terminal thiazole (Figure 1). It


Figure 1. Mature bottromycins seen in this study. **1–5** were previously identified in *S. scabies*,[34] **6–10** had not previously been identified in *S. scabies*.

has a unique architecture for an antibiotic, but biological studies have been hampered by low yields in native producing strains.[26] The bottromycin biosynthetic cluster in *Streptomyces scabies* consists of genes *btmA* to *btmM* (Figure 2), for which *btmD* encodes the precursor peptide, and the cluster has an overall GC content of 73.7%. To facilitate both heterologous

expression and yeast-mediated refactoring, we used TAR cloning in yeast to capture the bottromycin cluster from *S. scabies* DSM 41658 in the yeast/*E. coli* shuttle vector pCAP01.[17] To increase TAR cloning efficiency, a 24 kbp fragment containing the full gene cluster was excised out of *S. scabies* genomic DNA by digestion.[28] This facilitated lithium acetate/polyethylene glycol mediated transformation of yeast,[29] which is a quicker and easier technique than the spheroplast transformation used previously for TAR cloning larger DNA fragments.[17] The resulting vector carrying the bottromycin cluster, pCAPbtm, was conjugated into *S. coelicolor* M1146, but mature bottromycins (defined as compounds containing both the macrocyclic amidine and terminal thiazole) were produced in very low amounts in the resulting strain, M1146-pCAPbtm. To improve pathway productivity, we hypothesized that restriction sites naturally found in the bottromycin gene cluster could be exploited to efficiently introduce marker-free heterologous promoters, gene deletions, and gene rearrangements *via* the repair of double-strand DNA breaks by homologous recombination in yeast.

Genes *btmC* to *btmM* are tightly clustered and unidirectional, so we predicted that pathway expression could be modified by refactoring the region between *btmA* and *btmD* (Figure 2). We initially focused on testing a series of constitutive promoters that had previously been validated in *S. coelicolor*: PSF14,[22,30] Paac(3)IV (hereinafter referred to as "Paac3"),[31] PkasO*,[23] PhrdB,[32] and PermE*[33] (Table S1). A time-course β-glucuronidase (GUS) assay was undertaken to confirm the strength of each of these promoters in *S. coelicolor* M1146 alongside PbtmC, the native promoter responsible for expression of the putative *btmC-btmM* operon (Figure S1).
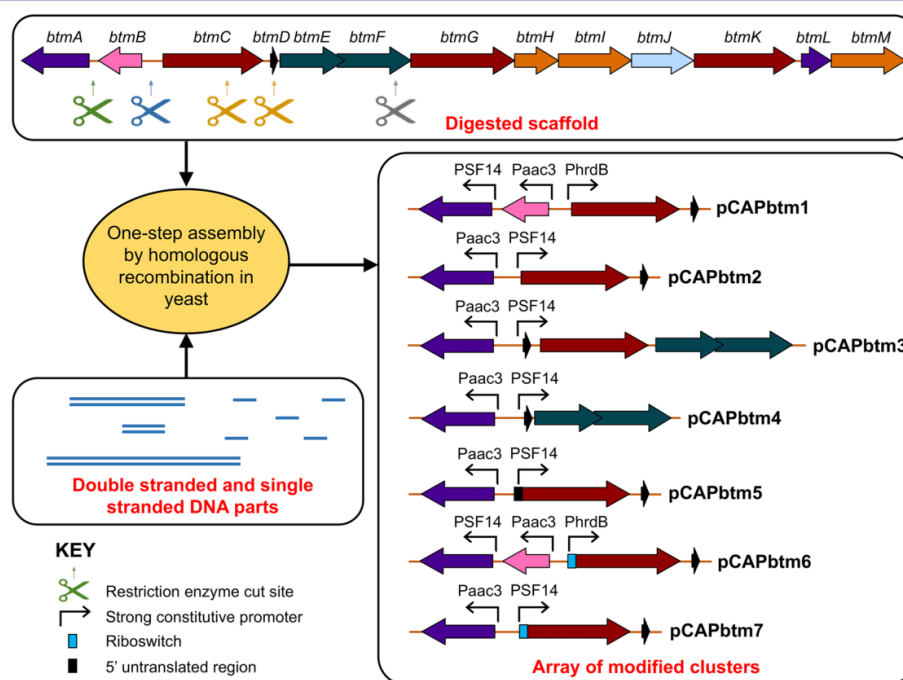
Figure 2. Schematic of yeast-mediated engineering of the bottromycin biosynthetic gene cluster. A detailed assembly schematic is shown in Figure S2.

Table 1. Summary of Plasmid Assemblies

name	modification(s)	parts used[a]	success rate and screening method	result
pCAPbtm*	insertion of BsrGI restriction site.	2 PCR products	3/4 restriction analysis	successful proof of principle that simplifies further engineering
pCAPbtm1	introduction of strong constitutive promoters.	4 PCR products	4/6 PCR analysis	improved production and introduced bottleneck
pCAPbtm2	btmB deletion; strong constitutive promoters.	7 oligos	2/5 restriction analysis	improved pathway productivity over pCAPbtm1
pCAPbtm3	btmB deletion; strong constitutive promoters; swapping btmC and btmD	3 PCR products 6 oligos	6/6 PCR analysis	indicates that gene order is important despite nonfunctional BtmC
pCAPbtm4	btmB deletion; strong constitutive promoters; btmC deletion	7 oligos	12/12 PCR analysis	btmC is important for efficient pathway productivity
pCAPbtm5	btmB deletion; strong constitutive promoters; inclusion of btmC 5′ UTR	7 oligos	4/5 sequencing analysis	5′ UTR had minimal effect on pathway productivity
pCAPbtm6	strong constitutive promoters; riboswitch before btmC	4 PCR products 2 oligos	2/6 PCR analysis	significantly reduced pathway productivity; the riboswitch provided no control
pCAPbtm7	btmB deletion; strong constitutive promoters; riboswitch before btmC	8 oligos	6/6 PCR analysis	provided highest production levels and inducible control over cluster productivity

[a]Details of each part are listed in Tables S20−21.

The relative promoter strengths were PSF14 > PhrdB > Paac3 > PermE* > PkasO*, while PbtmC is initially a relatively weak promoter but is as strong as PSF14 after 96 h, which is surprising considering the lack of bottromycins seen when the wild type cluster was expressed in S. coelicolor M1146.

To modify the btmA-D region, we exploited a unique NsiI recognition site in the btmC promoter region. Homologous recombination in S. cerevisiae VL6−48 was used to repair digested pCAPbtm with two overlapping PCR products, introducing a BsrGI recognition site into the intergenic region between btmA and btmB and generating pCAPbtm* (Figure 2 and Table 1; see Supplementary Methods for a detailed description). This made the cluster easier to manipulate and served as a proof of principle for further in vivo modifications of the cluster. The first refactored bottromycin cluster, pCAPbtm1, was made by introducing PSF14, PkasO* and PhrdB upstream of btmA, btmB, and btmC, respectively (Figure 2 and Table 1), and was assembled in yeast using BsrGI-NsiI digested pCAPbtm* and four overlapping PCR products (Figure 2, Figure S2, and Table 1). Low levels of mature bottromycins were produced by S. coelicolor M1146-pCAPbtm1, although all lacked a C-methyl group on phenylalanine (Figure 1, R$_2$), as determined by comparison of tandem mass spectrometry (MS^2) spectra with previously reported spectra[26,34] (Figures S3−12). This indicated that BtmC, a class B radical SAM methyltransferase, was seemingly not functioning in M1146-pCAPbtm1, despite being encoded by the first gene in the btmC-btmM operon.

Heterologous expression and BGC refactoring studies routinely focus on the final product of a pathway, yet this might not accurately reflect the total productivity of that pathway. Therefore, we employed untargeted metabolomics to identify all metabolomic changes between M1146-pCAPbtm1 and M1146 containing the empty pCAP01 vector (Table S2). To identify bottromycin-related metabolites (BRMs) only found in the heterologous host, the metabolomic profile of S. scabies was also assessed in parallel. Comparative analysis of liquid chromatography−mass spectrometry (LC−MS) data enabled the identification of all metabolites only produced by the refactored gene cluster. These compounds were then manually characterized by detailed MS^2 analysis (Figures S3−12), and any new compounds were subjected to high resolution MS^2 analysis (Tables S3−17). To complement this analysis, we constructed mass spectral networks,[35,36] which aided with

metabolite identification (Figure 3d). In total, we identified 16 new BRMs by MS analysis (6−11, 14−15, 19−22, 25, 27−29; Figure 3; Figures S3−12; Tables S3−17), including five new mature bottromycins (compounds 6−10; m/z 809.43, m/z 795.42, 767.39, m/z 781.41, and m/z 795.42 respectively; Figure 1).

Encouragingly, this analysis showed that a much larger quantity of BRMs was produced in comparison to M1146-pCAPbtm1, resulting in an overall 20-fold increase (Figure 4) of total MS intensity for these metabolites. The expected C-methyl group on phenylalanine was absent in every detectable BRM in M1146-pCAPbtm1. Notably, it has been shown that deletion of btmC in S. scabies abolishes production of mature bottromycins, which indicates that phenylalanine C-methylation is critical for the efficient activity of downstream enzymes in the native producer.[24] We propose that inactive BtmC creates a bottleneck that M1146-pCAPbtm1 is able to overcome by strong expression from the heterologous promoters, providing an increased metabolic flux through the pathway and thus leading to the suite of new BRMs.

Despite its apparent inactivity, there were no mutations to btmC in M1146-pCAPbtm1, and whole genome sequencing of this strain showed that there were no mutations elsewhere in the gene cluster. We therefore generated newly refactored BGC assemblies to assess whether different promoters would result in active BtmC and improve pathway productivity. As part of this refactoring we deleted btmB, as BtmB-catalyzed methylation of the aspartic acid (Figure 1, R$_1$) is necessary for bottromycin activity.[37] The resulting BGC would have the advantage of producing an inactive bottromycin, and therefore overcome any self-toxicity issues that could arise in a heterologous host. This methyl ester group is rapidly hydrolyzed in blood plasma,[37] so it is unlikely that any clinically relevant derivatives of bottromycin will carry this modification.

The target refactored plasmid (pCAPbtm2) would therefore feature no btmB, as well as an alternative array of promoters: Paac3 upstream of btmA and PSF14 upstream of btmC. To achieve this in a single step, we modified a technique for oligonucleotide assembly in yeast spheroplasts[21] to construct pCAPbtm2 from pCAPbtm* and seven single stranded oligonucleotides using lithium acetate/polyethylene glycol mediated transformation (Figure 2 and Table 1). As expected, M1146-pCAPbtm2 generated mature bottromycins that all
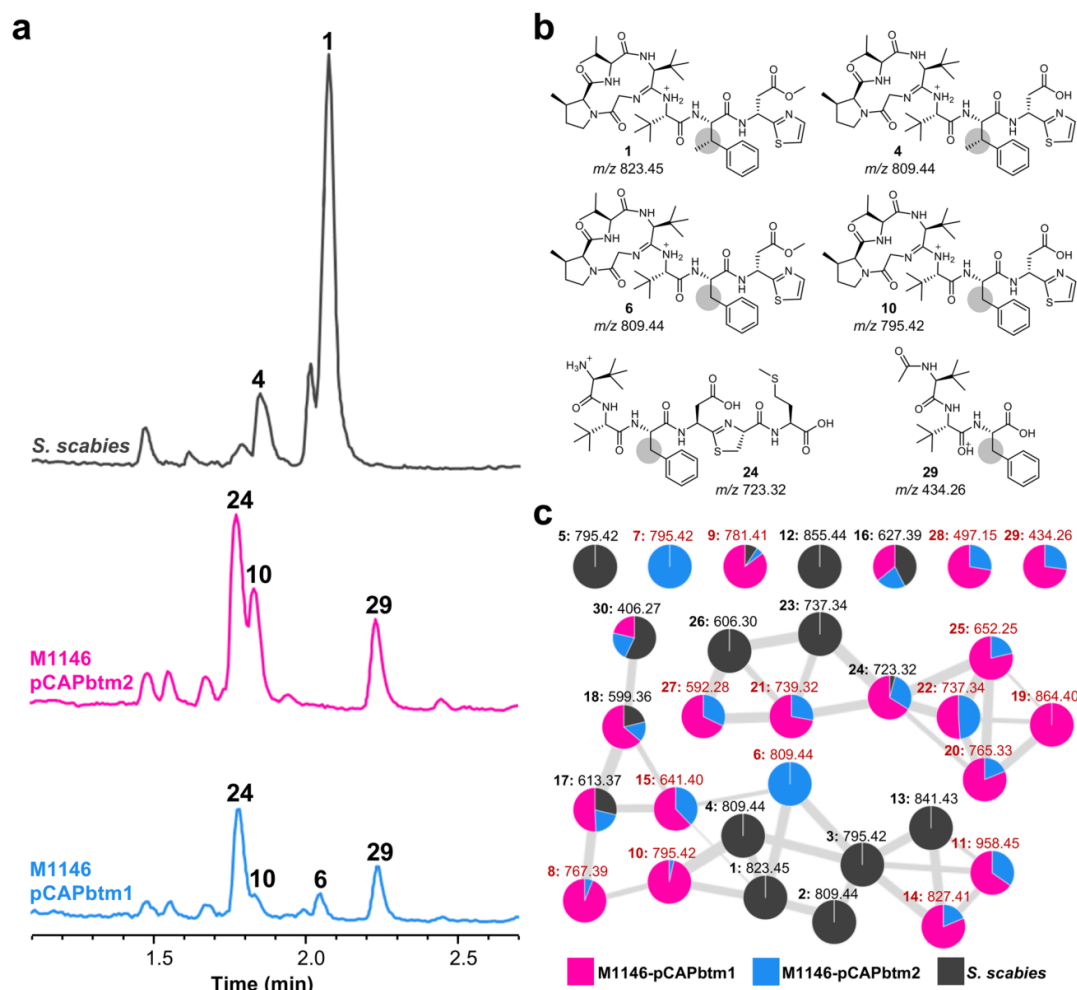
Figure 3. Comparison of bottromycin pathway productivity. (a) Extracted ion chromatograms showing all detectable bottromycin related metabolites from *S. scabies*, *S. coelicolor* M1146-pCAPbtm1 and M1146-pCAPbtm2. The metabolites responsible for the most intense peaks in each spectrum are numbered. (b) Structures of the metabolites numbered in panel a. The gray shaded circles highlight the absence or presence of phenylalanine C-methylation. (c) A mass spectral network of the bottromycin-related metabolites produced by M1146-pCAPbtm1, 2, and *S. scabies* in blue, pink, and gray, respectively. Each node represents a single metabolite, with the detected *m/z* listed (Figures S3−12). Node pie charts indicate the relative abundances of each metabolite between strains. Metabolites identified, but not detected by MS networking, are shown as unconnected nodes. Black node labels indicate molecules previously identified in *S. scabies*, while red node labels indicate molecules identified for the first time in this work.

lacked the methyl ester group on the aspartic acid (**8**−**10**; Figures S4,S5), but these compounds were also not C-methylated on phenylalanine. As before, comparative metabolomics was used to map the full array of pathway products (Figure 3 and Table S2). This highlighted a significant increase in total production over M1146-pCAPbtm1 for both mature bottromycins and other BRMs. On the basis of the LC−MS peak areas, M1146-pCAPbtm2 produced 60 times more BRMs than M1146-pCAPbtm (Figure 4). To assess the role of the host strain on heterologous expression, pCAPbtm2 was also expressed in *Streptomyces albus*, *Streptomyces laurentii*, *Streptomyces lividans*, and *Streptomyces venezuelae* (Figure S13). *S. venezuelae* provided a marginal improvement in production over M1146, while production in *S. laurentii* and *S. lividans* did not differ significantly from M1146. In contrast, the pathway

was completely inactive in *S. albus*, indicating the strain-specific nature of the heterologous pathway expression. No C-methylation of phenylalanine was observed in any strain.

Yeast-mediated cluster engineering enabled us to construct a variety of gene clusters to investigate BtmC inactivity, as well as the importance of the gene for pathway productivity, while showcasing the versatility of the system with more complicated assemblies. Three questions were asked: (a) What would happen by swapping the positions of *btmC* and *btmD*? (b) Is inactive *btmC* important for pathway productivity? (c) Is a native 5′ untranslated region (5′ UTR) required for *btmC* mRNA stability? Three independent variants of pCAPbtm2 were therefore generated as described below (pCAPbtm3−5; Figure 2 and Table 1). Single step assemblies were designed using an array of PCR products and oligonucleotides that were
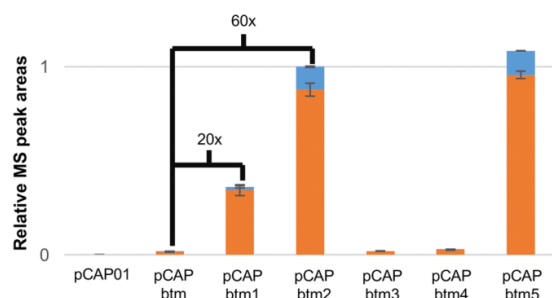
Figure 4. Production of mature bottromycins (blue) and other bottromycin-related metabolites (orange) from refactored bottromycin gene clusters expressed in *S. coelicolor* M1146.

varied depending on the desired construct. The assembly reactions proceeded with high fidelity (Table 1 and Table S18).

To assess whether the position of *btmC* is important for its expression and activity, *btmC* and *btmD* were swapped in pCAPbtm3. To determine whether the presence of the *btmC* gene is important for efficient operon expression, *btmC* was deleted to generate pCAPbtm4. Finally, the 5′ UTR associated with PSF14 was replaced with the native 5′ UTR of *btmC* to yield pCAPbtm5, to establish whether its RNA structure is important for translation initiation or mRNA stability (Figure 2 and Table 1). Production profiles of each construct in M1146 indicated that BtmC remained inactive in all cases (data not

shown). Production of BRMs was almost abolished in M1146-pCAPbtm3 and M1146-pCAPbtm4 (Figure 4). This indicates that, despite its inactivity, the presence *btmC* in its original location within the operon is important for pathway activity. The inclusion of the native 5′ UTR of *btmC* in pCAPbtm5 provided a slight drop in the production of mature bottromycins compared to pCAPbtm2, accompanied by slight boost to the production of other BRMs (Figure 4).

BtmC is a class B radical SAM methyltransferase, which are often insoluble when overexpressed in a heterologous host.[38] Therefore, a possible solution to BtmC inactivity is attenuating the expression of *btmC*, which could potentially be achieved by using a riboswitch. Rudolph *et al.*[22] identified a series of theophylline-dependent riboswitches that enabled inducible gene expression in *S. coelicolor*. Riboswitch E* provided the strongest control of expression in that study, so we constructed versions of pCAPbtm1 and pCAPbtm2 with this riboswitch between PhrdB/PSF14 and *btmC* to generate pCAPbtm6 and pCAPbtm7, respectively (Figure 2 and Table 1). This would also help validate the utility of a riboswitch in refactored *Streptomyces* BGCs, as there are very few examples of inducible riboswitch systems.[39]

Bottromycin production was almost abolished in M1146-pCAPbtm6, both with and without theophylline (Figure S14), and only two BRMs could be detected. In contrast, bottromycin production was tightly controlled by varying theophylline concentrations in M1146-pCAPbtm7 (Figure 5a). While it did not rescue BtmC activity, the addition of theophylline caused a large increase in bottromycin production,
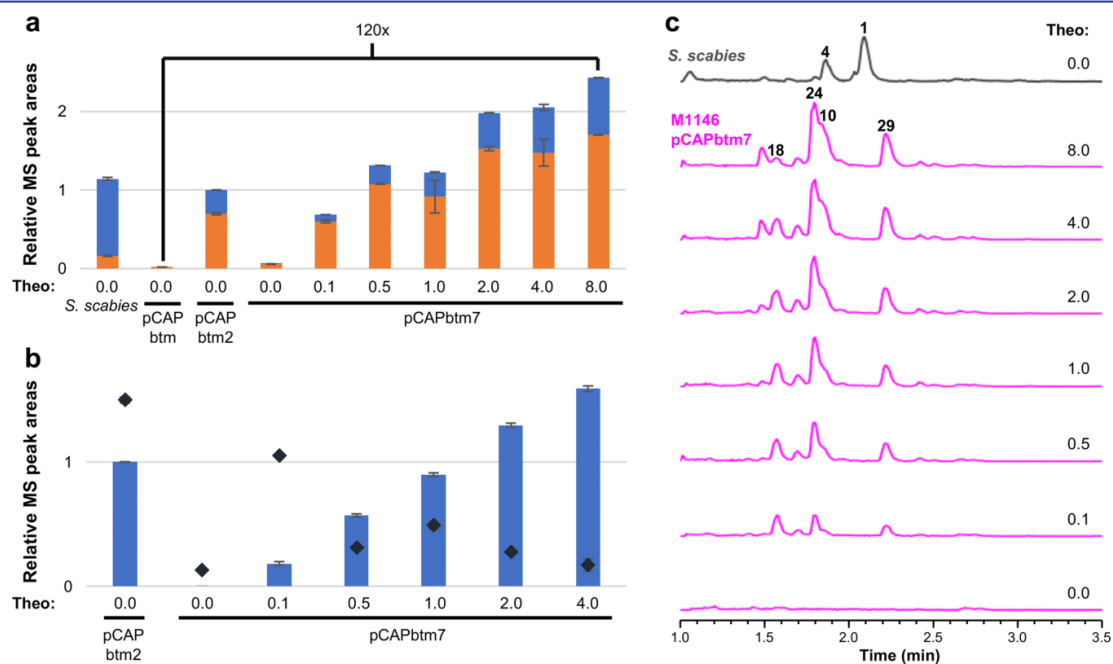
Figure 5. Assessment of inducible bottromycin production by the riboswitch-containing pCAPbtm7. (a) Production of mature bottromycins (blue) and other bottromycin-related metabolites (orange) from *S. scabies*, pCAPbtm, pCAPbtm2, and pCAPbtm7 in *S. coelicolor* M1146. Theophylline concentrations (mM) are listed. (b) Production of mature bottromycins (blue bars) from pCAPbtm2 and pCAPbtm7 with increasing concentrations of theophylline. qRT-PCR measurements of the transcript levels of precursor peptide *btmD* normalized to pCAPbtm2 = 1.5 are also shown (diamonds). (c) Extracted ion chromatograms of all bottromycin-related metabolites detected in *S. scabies* and M1146-pCAPbtm7 at different theophylline concentrations (mM).

and at 8 mM theophylline, M1146-pCAPbtm7 produced 120 times the quantity of BRMs compared to the nonrefactored cluster (M1146-pCAPbtm) (Figure 5a). The production of mature bottromycins also reached a level comparable to wild type *S. scabies*, and the total yield of BRMs was over twice that of *S. scabies*. The metabolites produced were the same as those produced by M1146-pCAPbtm2 (therefore not C-methylated on phenylalanine) and were detected in comparable ratios. This titratable control of the bottromycin pathway *via* this riboswitch was a surprising result, as it was predicted to only control translation of inactive *btmC*. Therefore, qRT-PCR was used to measure transcript levels of *btmC* (Figure S15) and *btmD* (Figure 5b). This revealed that while the riboswitch did influence transcription, this was not entirely correlated with production, and at theophylline concentrations over 0.1 mM the role of the riboswitch on pathway productivity could not be simply explained by transcript levels. For example, transcript levels of *btmD* in pCAPbtm7 with 4 mM theophylline were 9-fold lower than in pCAPbtm2, but this was associated with a significant increase in bottromycin production in pCAPbtm7. This infers a complex interplay between gene transcription and translation, as well as the relative activity of pathway enzymes.

In summary, the combination of TAR cloning and yeast-mediated refactoring based on homologous recombination is an efficient and versatile way to introduce scarless modifications into GC-rich natural product BGCs. Cluster assemblies were achieved in a single step with different combinations of reusable PCR products and single-stranded oligonucleotides (Figure S2; see Table 1 for a summary of all assemblies and results). There was a high average assembly efficiency, where 79% of colonies screened were correctly assembled. The most complex assembly reported here, pCAPbtm3, was constructed from three PCR products, six oligonucleotides, and a digested vector, and was correctly assembled in all colonies screened. This method was not hindered by the high GC content of the bottromycin gene cluster or the assembly of multiple regulatory regions, which can hinder *in vitro* methods.[9,10] The method takes advantage of naturally occurring restriction sites and facilitated the introduction of new ones. While this potentially limits the regions that can be modified, the recent development of programmable restriction enzymes potentially overcomes this barrier.[40,41]

Our inability to activate BtmC indicates that while BGC refactoring and heterologous expression are powerful techniques, they can lead to unexpected results. Untargeted metabolomics and mass spectral networking enabled us to assess the total productivity of the bottromycin gene cluster and identify novel pathway products resulting from a nonfunctioning BtmC. The most successful construct (pCAPbtm7) increased production in *S. coelicolor* M1146 120-fold compared to the wild type cluster (pCAPbtm), and the use of a riboswitch led to strict control of production (Figure 5). This is only the second report of riboswitch-dependent inducible expression of a *Streptomyces* BGC,[39] and the first example in *S. coelicolor*. The theophylline-controlled riboswitch represents a rapid way to modulate gene expression without the need for a library of gene clusters containing promoters of different strengths.

■ METHODS

Assembly of pCAP01-Based Plasmids. pCAP01-Based plasmid assemblies were all carried out using LiAc/PEG mediated transformation into yeast (see Supplementary Methods). The efficiency and flexibility of the process meant that no strict adherence to concentration or molar ratios of DNA was necessary. As a guide, the following ratio of DNA in an assembly was sufficient: 40 ng of linearized plasmid, 80 ng of each PCR product, and 500 pmol of each oligonucleotide. PCR products and linearized plasmids were column purified. Unless otherwise noted, the PCR products used in assemblies were produced with approximately 60 bp regions of overlap with other assembly parts. The oligonucleotides used in assemblies overlapped by between 30 and 60 bp, and if they were assembled to another oligonucleotide they were also complementary. Plasmid screening methods are described in Table S18, and all primers and oligonucleotides are listed in Table S19.

To assemble pCAPbtm*, two overlapping regions between the NsiI and SpeI recognition sites in pCAPbtm were obtained using the primer pairs: SPbtmA*v2 and ASPtbtmA*v1, and SPbtmB*v2 and ASPbtmB*v2. These overlapped with each other by 30 bp in the btmA promoter region and included the necessary mutations for introducing a BsrGI recognition site. These PCR products also included 30 bp on either side of the NsiI and SpeI recognition sites to facilitate homologous recombination. pCAPbtm was digested with NsiI and SpeI, and pCAPbtm* was assembled with the PCR products and confirmed by restriction analysis (Table S18). The parts used in the assembly of pCAPbtm1–7 are shown in Table S20, and Table S21, which provide a description of each part. A schematic of each assembly is also shown in Figure S2. The assessment of assembly efficiency is described in Table S18, and each assembly was additionally confirmed by sequencing. Details of *gusA* plasmid construction are provided in the Supplementary Methods.

Liquid Chromatography–Mass Spectrometry (LC–MS). Samples were taken from the supernatant of production cultures and mixed with one volume methanol. LC–MS data were acquired using a Shimadzu Nexera X2 UHPLC connected to a Shimadzu ion-trap time-of-flight (IT-TOF) mass spectrometer and analyzed using LabSolutions software (Shimadzu). Metabolites produced by strains containing pCAPbtm-based plasmids were identified by a combination of the statistical package Profiling Solution 1.1 (Shimadzu; Table S2) and mass spectral networking using Global Natural Products Social Molecular Networking[36] (GNPS; http://gnps.ucsd.edu). LC–MS data from pCAPbtm1, pCAPbtm2, and *S. scabies* production experiments were used for GNPS analysis (Figure 3).

Following identification by metabolomics, MS² data for previously unreported BRMs were compared to MS² data of known BRMs[26,34] to aid with metabolite identification (Figures S3–S12). For each sample, all BRMs (compounds **1–30**) were used to estimate pathway productivity by integration of LC–MS peak areas using Browser software (Shimadzu). All data reported from riboswitch experiments are the average of duplicate cultures, while all other MS quantification data are the average of triplicate cultures. Error bars represent the standard error. High resolution mass and MS² data of the new BRMs were acquired on a Waters Synapt G2-Si mass spectrometer. Results are reported in Tables S3–S17.

■ ASSOCIATED CONTENT

ⓢ Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acssynbio.8b00038.

ACS Synthetic Biology   Letter

Further experimental details, including general micro-
biological and chemical methods, qRT-PCR experiments,
and genome sequencing; additional figures, tables, and
their references (PDF)

AUTHOR INFORMATION

Corresponding Author
*Tel: +44(0)1603 450750. E-mail: andrew.truman@jic.ac.uk.

ORCID
Andrew W. Truman: 0000-0001-5453-7485

Author Contributions
T.H.E. performed TAR cloning, heterologous pathway
expression, gene cluster refactoring and LC–MS analysis.
N.M.V. and T.H.E. carried out gene expression analysis. A.W.T.
conceived and supervised the study.

Notes
The authors declare no competing financial interest.

REFERENCES

(1) Newman, D. J., and Cragg, G. M. (2016) Natural Products as
Sources of New Drugs from 1981 to 2014. *J. Nat. Prod. 79*, 629−661.
(2) Ongley, S. E., Bian, X., Neilan, B. A., and Müller, R. (2013)
Recent advances in the heterologous expression of microbial natural
product biosynthetic pathways. *Nat. Prod. Rep. 30*, 1121−1138.
(3) Gomez-Escribano, J. P., and Bibb, M. J. (2011) Engineering
Streptomyces coelicolor for heterologous expression of secondary
metabolite gene clusters. *Microb. Biotechnol. 4*, 207−215.
(4) Smanski, M. J., Zhou, H., Claesen, J., Shen, B., Fischbach, M. A.,
and Voigt, C. A. (2016) Synthetic biology to access and expand
nature's chemical diversity. *Nat. Rev. Microbiol. 14*, 135−149.
(5) Gibson, D. G., Young, L., Chuang, R.-Y., Venter, J. C., Hutchison,
C. A., III, and Smith, H. O. (2009) Enzymatic assembly of DNA
molecules up to several hundred kilobases. *Nat. Methods 6*, 343−345.
(6) Engler, C., Gruetzner, R., Kandzia, R., and Marillonnet, S. (2009)
Golden gate shuffling: a one-pot DNA shuffling method based on type
IIs restriction enzymes. *PLoS One 4*, e5553.
(7) Weber, E., Engler, C., Gruetzner, R., Werner, S., and Marillonnet,
S. (2011) A modular cloning system for standardized assembly of
multigene constructs. *PLoS One 6*, e16765.
(8) Smanski, M. J., Bhatia, S., Zhao, D., Park, Y., Woodruff, L. B. A.,
Giannoukos, G., Ciulla, D., Busby, M., Calderon, J., Nicol, R., Gordon,
D. B., Densmore, D., and Voigt, C. A. (2014) Functional optimization
of gene clusters by combinatorial design and assembly. *Nat. Biotechnol.
32*, 1241−1249.

(9) Liang, J., Liu, Z., Low, X. Z., Ang, E. L., and Zhao, H. (2017)
Twin-primer non-enzymatic DNA assembly: an efficient and accurate
multi-part DNA assembly method. *Nucleic Acids Res. 45*, e94.
(10) Casini, A., MacDonald, J. T., De Jonghe, J., Christodoulou, G.,
Freemont, P. S., Baldwin, G. S., and Ellis, T. (2014) One-pot DNA
construction for synthetic biology: the Modular Overlap-Directed
Assembly with Linkers (MODAL) strategy. *Nucleic Acids Res. 42*, e7.
(11) Orr-Weaver, T. L., Szostak, J. W., and Rothstein, R. J. (1981)
Yeast transformation: a model system for the study of recombination.
Proc. Natl. Acad. Sci. U. S. A. 78, 6354−6358.
(12) Kouprina, N., and Larionov, V. (2008) Selective isolation of
genomic loci from complex genomes by transformation-associated
recombination cloning in the yeast *Saccharomyces cerevisiae*. *Nat.
Protoc. 3*, 371−377.
(13) Noskov, V. N., Karas, B. J., Young, L., Chuang, R.-Y., Gibson, D.
G., Lin, Y.-C., Stam, J., Yonemoto, I. T., Suzuki, Y., Andrews-
Pfannkoch, C., Glass, J. I., Smith, H. O., Hutchison, C. A., III, Venter,
J. C., and Weyman, P. D. (2012) Assembly of large, high G+C
bacterial DNA fragments in yeast. *ACS Synth. Biol. 1*, 267−273.
(14) Pahirulzaman, K. A. K., Williams, K., and Lazarus, C. M. (2012)
A toolkit for heterologous expression of metabolic pathways in
Aspergillus oryzae. *Methods Enzymol. 517*, 241−260.
(15) Shao, Z., Rao, G., Li, C., Abil, Z., Luo, Y., and Zhao, H. (2013)
Refactoring the Silent Spectinabilin Gene Cluster Using a Plug-and-
Play Scaffold. *ACS Synth. Biol. 2*, 662−669.
(16) Kang, H.-S., Charlop-Powers, Z., and Brady, S. F. (2016)
Multiplexed CRISPR/Cas9- and TAR-Mediated Promoter Engineer-
ing of Natural Product Biosynthetic Gene Clusters in Yeast. *ACS
Synth. Biol. 5*, 1002−1010.
(17) Yamanaka, K., Reynolds, K. A., Kersten, R. D., Ryan, K. S.,
Gonzalez, D. J., Nizet, V., Dorrestein, P. C., and Moore, B. S. (2014)
Direct cloning and refactoring of a silent lipopeptide biosynthetic gene
cluster yields the antibiotic taromycin A. *Proc. Natl. Acad. Sci. U. S. A.
111*, 1957−1962.
(18) Montiel, D., Kang, H.-S., Chang, F.-Y., Charlop-Powers, Z., and
Brady, S. F. (2015) Yeast homologous recombination-based promoter
engineering for the activation of silent natural product biosynthetic
gene clusters. *Proc. Natl. Acad. Sci. U. S. A. 112*, 8953−8958.
(19) Storici, F., Durham, C. L., Gordenin, D. A., and Resnick, M. A.
(2003) Chromosomal site-specific double-strand breaks are efficiently
targeted for repair by oligonucleotides in yeast. *Proc. Natl. Acad. Sci. U.
S. A. 100*, 14994−14999.
(20) Raymond, C. K., Pownder, T. A., and Sexson, S. L. (1999)
General method for plasmid construction using homologous
recombination. *BioTechniques 26*, 134−141.
(21) Gibson, D. G. (2009) Synthesis of DNA fragments in yeast by
one-step assembly of overlapping oligonucleotides. *Nucleic Acids Res.
37*, 6984−6990.
(22) Rudolph, M. M., Vockenhuber, M.-P., and Suess, B. (2013)
Synthetic riboswitches for the conditional control of gene expression
in *Streptomyces coelicolor*. *Microbiology 159*, 1416−1422.
(23) Wang, W., Li, X., Wang, J., Xiang, S., Feng, X., and Yang, K.
(2013) An engineered strong promoter for streptomycetes. *Appl.
Environ. Microbiol. 79*, 4484−4492.
(24) Crone, W., Leeper, F. J., and Truman, A. W. (2012)
Identification and characterisation of the gene cluster for the anti-
MRSA antibiotic bottromycin: expanding the biosynthetic diversity of
ribosomal peptides. *Chem. Sci. 3*, 3516−3521.
(25) Gomez-Escribano, J. P., Song, L., Bibb, M. J., and Challis, G. L.
(2012) Posttranslational β-methylation and macrolactamidination in
the biosynthesis of the bottromycin complex of ribosomal peptide
antibiotics. *Chem. Sci. 3*, 3522−3525.
(26) Huo, L., Rachid, S., Stadler, M., Wenzel, S. C., and Müller, R.
(2012) Synthetic biotechnology to study and engineer ribosomal
bottromycin biosynthesis. *Chem. Biol. 19*, 1278−1287.
(27) Hou, Y., Tianero, M. D. B., Kwan, J. C., Wyche, T. P., Michel, C.
R., Ellis, G. A., Vazquez-Rivera, E., Braun, D. R., Rose, W. E., Schmidt,
E. W., and Bugni, T. S. (2012) Structure and Biosynthesis of the
Antibiotic Bottromycin D. *Org. Lett. 14*, 5050−5053.

(28) Leem, S.-H., Noskov, V. N., Park, J.-E., Kim, S. I., Larionov, V., and Kouprina, N. (2003) Optimum conditions for selective isolation of genes from complex genomes by transformation-associated recombination cloning. *Nucleic Acids Res. 31*, e29.

(29) Gietz, R. D., and Woods, R. A. (2006) Yeast transformation by the LiAc/SS Carrier DNA/PEG method. *Methods Mol. Biol. 313*, 107–120.

(30) Labes, G., Bibb, M., and Wohlleben, W. (1997) Isolation and characterization of a strong promoter element from the *Streptomyces ghanaensis* phage I19 using the gentamicin resistance gene (aacC1) of Tn 1696 as reporter. *Microbiology 143*, 1503–1512.

(31) Sherwood, E. J., Hesketh, A. R., and Bibb, M. J. (2013) Cloning and analysis of the planosporicin lantibiotic biosynthetic gene cluster of *Planomonospora alba. J. Bacteriol. 195*, 2309–2321.

(32) Du, D., Zhu, Y., Wei, J., Tian, Y., Niu, G., and Tan, H. (2013) Improvement of gougerotin and nikkomycin production by engineering their biosynthetic gene clusters. *Appl. Microbiol. Biotechnol. 97*, 6383–6396.

(33) Bibb, M. J., White, J., Ward, J. M., and 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*, 533–545.

(34) Crone, W. J. K., Vior, N. M., Santos-Aberturas, J., Schmitz, L. G., Leeper, F. J., and Truman, A. W. (2016) Dissecting Bottromycin Biosynthesis Using Comparative Untargeted Metabolomics. *Angew. Chem., Int. Ed. 55*, 9639–9643.

(35) Nguyen, D. D., et al. (2013) MS/MS networking guided analysis of molecule and gene cluster families. *Proc. Natl. Acad. Sci. U. S. A. 110*, E2611–E2620.

(36) Wang, M., et al. (2016) Sharing and community curation of mass spectrometry data with Global Natural Products Social Molecular Networking. *Nat. Biotechnol. 34*, 828–837.

(37) Kobayashi, Y., Ichioka, M., Hirose, T., Nagai, K., Matsumoto, A., Matsui, H., Hanaki, H., Masuma, R., Takahashi, Y., Ōmura, S., and Sunazuka, T. (2010) Bottromycin derivatives: Efficient chemical modifications of the ester moiety and evaluation of anti-MRSA and anti-VRE activities. *Bioorg. Med. Chem. Lett. 20*, 6116–6120.

(38) Bauerle, M. R., Schwalm, E. L., and Booker, S. J. (2015) Mechanistic diversity of radical S-adenosylmethionine (SAM)-dependent methylation. *J. Biol. Chem. 290*, 3995–4002.

(39) Horbal, L., and Luzhetskyy, A. (2016) Dual control system - A novel scaffolding architecture of an inducible regulatory device for the precise regulation of gene expression. *Metab. Eng. 37*, 11–23.

(40) Liu, Y., Tao, W., Wen, S., Li, Z., Yang, A., Deng, Z., and Sun, Y. (2015) *In Vitro* CRISPR/Cas9 System for Efficient Targeted DNA Editing. *mBio 6*, e01714–15.

(41) Enghiad, B., and Zhao, H. (2017) Programmable DNA-Guided Artificial Restriction Enzymes. *ACS Synth. Biol. 6*, 752–757.

Abbreviations

ABC	ATP-binding cassette
ADP	adenosine diphosphate
AMS	ATP transporter maturation and secretion
APH	aminoglycoside phosphotransferase
ATP	adenosine triphosphate
AviCys	*S*-[(*Z*)-2-aminovinyl]-D-cysteine
AviMeCys	*S*-[(*Z*)-2-aminovinyl](-(3*S*)-3-methyl)-D-cysteine
BAC	bacterial artificial chromosome
BLAST	basic local alignment search tool
BLOSUM	blocks substitution matrix
BPC	base peak chromatogram
Btm	bottromycin
btmPM	bottromycin production medium
cDNA	complimentary DNA
CIPRES	cyberinfrastructure for phylogenetic research
COSY	correlation spectroscopy
CRISPR	clustered regularly interspaced short palindromic repeats
DA	deoxyadenosyl
DATEL	DNA assembly with thermostable exonuclease and ligase
DEPTQ	distortionless enhancement by polarization transfer including the detection of quaternary nuclei
Dha	dehydroalanine
Dhb	dehydrobutyrine
DNA	deoxyribonucleic acid
DUF	domain of unknown function
EIC	extracted ion chromatogram
EMBL-EBI	European molecular biology laboratory - European bioinformatics institute
FAD	flavin adenine dinucleotide
FMN	flavin mononucleotide
GNPS	global natural product social molecular networking
GTP	guanosine triphosphate
HFCD	homo-oligomeric flavin-containing cysteine decarboxylase
HMBC	heteronuclear multiple bond correlation
HPSF	high purity salt free
HSQC	heteronuclear single quantum correlation
I-TASSER	iterative threading assembly refinement
iTOL	interactive tree of life
IT-TOF	ion trap - time of flight
K_a	acid disassociation constant
LAP	linear azol(in)e-containing peptide

LB	lysogeny broth
LC	liquid chromatography
LiAc	lithium acetate
MIDAS	modular idempotent DNA assembly system
MoClo	modular cloning
MPH	macrolide phosphotransferase
MS	mass spectrometry
MUSCLE	multiple sequence comparison by log-expectation
NADPH	nicotinamide adenine dinucleotide phosphate
NMR	nuclear magnetic resonance
NRPS	non-ribosomal peptide synthetase
OEPR	overlap extension PCR and recombination
PAC	phage artificial chromosome
PCR	polymerase chain reaction
PEG	polyethylene glycol
PKS	polyketide synthase
PSA	pairwise selection assembly
QGA	quick gene assembly
qRT	quantitative reverse transcription
RAxML	randomized axelerated maximum likelihood
RiPP	ribosomally synthetised and post translationally modified peptide
RNA	ribonucleic acid
RRE	RiPP precursor peptide recognition element
rSAM	radical S-adenosyl-L-methionine
RT	reverse transcription
SAM	S-adenosyl-L-methionine
SLIC	sequence and ligation - independent cloning
TALEN	transcription activator-like effector nucleases
TAR	transformation associated recombination
TIC	total ion chromatogram
TPA	twin primer assembly
tRNA	transfer ribonucleic acid
Tsa	thiostreptamide S4
USER	uracil-specific excision reagent
UTR	untranslated region
VEGAS	versatile genetic assembly system
ZFN	zinc-finger nucleases