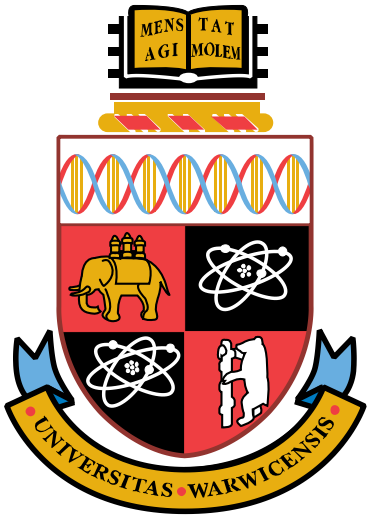



# Analysis and exploitation of AHFCA-dependent signalling systems in *Streptomyces* bacteria

**by**

**Vincent Poon**

A thesis submitted in partial fulfillment of the requirements for the degree of
Doctor of Philosophy in Systems Biology

University of Warwick, Systems Biology Doctoral Training Centre

December 2015

**Table of Contents**

**List of Figures**

## List of Tables

**Acknowledgements**

Firstly, I would like to express my deep gratitude to both my supervisors, Dr. Christophe Corre and Dr. Jonathan Moore, for giving me the opportunity to carry out a PhD project under their patience, guidance, support and supervision. I would like to extend my gratitude to my advisory panel members, Prof. Gregory Challis, Prof. David Hodgson, Dr. Julia Brettschneider, Dr. Leon Danon and Dr. Joseph Christie-Oleza for their helpful discussions and advice throughout my PhD.

I am especially grateful for: Lijiang Song and Ivan Prokes for their help with NMR data analysis, and analysing my samples on the Bruker MaXis/MaXis impact and Bruker Avance; John Sidda for helpful discussions and training with organic extractions and LC-MS analysis; Douglas Roberts for the help in purification of compounds by HPLC; Shanshan Zhou for the help with NMR sample preparation and providing the authentic standards of AHFCA signalling molecules; Daniel Zabala-Alvarez, Chiara Borsetto, Martin Schaefer, Kathryn Styles and Candace Ho for all of their technical advice, support and help throughout my PhD studies.

I would like thank everyone (past and present) in the Corre/ Tosin/ Challis groups from the Chemical Biology Research Facility and Wellington/ Hodgson/ Scanlan/ Christie-Oleza groups from the School of Life Sciences for making my time an enjoyable and memorable experience. Finally, I also would like to thank friends (especially Jenny Chan) and family for their continuous support.

*The PhD was financially supported by Systems Biology Doctoral Training Centre and Biotechnology and Biological Sciences Research Council.*

**Declaration and Inclusion of Material from a Prior Thesis**

This thesis is submitted to the University of Warwick in support of my application for the degree of Doctor of Philosophy. It has been composed by myself and has not been submitted in any previous application for any degree.

The work presented (including data generated and data analysis) was carried out by the author except in the cases outlined below:

Work in Chapter 6 on Compound **2** NMR data analysis was performed by Dr. Lijiang Song and Dr. Christophe Corre

Parts of this thesis have been published by the author:

Sidda, J. D., Song, L., Poon, V., Al-Bassam, M., Lazos, O., Buttner, M. J., Challis, G. L. & Corre, C., Discovery of a family of γ-aminobutyrate ureas *via* rational derepression of a silent bacterial gene cluster. *Chem Sci*. **5**, 86-89. (2014).

**Abstract**

From genome sequencing and analyses, *Streptomyces* bacteria are predicted to produce hundreds of bioactive metabolite compounds; however, the products of some of these cryptic biosynthetic gene clusters are most often silent or produced in very small quantities in laboratory conditions. The AHFCA-dependent signalling system is an example of a regulatory systems that tightly control expression of cryptic biosynthetic gene clusters in *Streptomyces* bacteria. Bioinformatic analyses, using MEME and FIMO, were able to successfully reveal TetR repressor binding sites to gain insights into complex regulatory networks.

The *S. avermitilis* cryptic biosynthetic gene cluster, which is proposed to be under the control of an AHFCA-dependent signalling system, has not been investigated before and genome mining predictions suggest the biosynthetic gene cluster could be involved in the biosynthesis of novel azoxy compounds. The 50 kb predicted azoxy biosynthetic gene cluster was captured using pESAC13A PAC technologies and introduced into *S. coelicolor* M1152 and *S. lividans* TK24 heterologous hosts. Comparative metabolic profiling successfully identified the predicted azoxy compounds in *S. coelicolor* M1152 by LC-MS and identification of the azoxy compounds were supported by UHPLC-ESI-TOF-MS analysis and incorporation of isotope-labelled L-serine.

Consistent with the model of transcriptional regulation, exploitation of the *S. avermitilis* AHFCA-dependent signalling system through the inactivation of *sav*_2268 transcriptional repressor in *S. lividans* TK24 heterologous hosts unlocked the production of azoxy compounds and AHFCA signalling molecules.

**Abbreviations**

| | |
|---|---|
| A | Adenylation (NRPS domain) |
| aaRS | Aminoacyl tRNA synthetase |
| ACP | Acyl carrier protein |
| Act | Actinorhodin |
| AHFCA | 2-Alkyl-4-hydroxymethylfuran-3-carboxylic acids |
| AntiSMASH | Antibiotics & Secondary Metabolite Analysis Shell |
| ARE | Autoregulatory response element |
| AT | Acytransferase (PKS domain) |
| ATP | Adenosine tri-phosphate |
| BAC | Bacterial artificial chromosome |
| BLAST | Basic Local Alignment Search Tool |
| CDA | Calcium dependent antibiotic |
| ChIP-Seq | Chromatin immunoprecipitate - sequencing |
| CoA | Coenzyme A |
| DH | Dehydratase (PKS domain) |
| DNA | Deoxyribonucleic acid |
| DMSO | Dimethyl sulfoxide |
| EDTA | Ethylenediaminetetraacetic acid |
| EMSA | Electromobility shift assay |
| ESAC | *E. coli-Streptomyces* artificial chromosome |
| ESI | Electronspray ionisation |
| FIMO | Finding Individual Motif Occurence |
| GBL | γ-Butyrolactone |
| HCT | High capacity trap |
| HMBC | Heteronuclear multiple-bond correlation |
| HPLC | High performance liquid chromatography |
| HSQC | Heteronuclear single quantum coherence |
| HTH | Helix-turn-helix |
| iChip | Isolation chip |
| IR | Infra-red |
| KR | Ketoreductase (PKS domain) |
| KS | Ketosynthase (PKS domain) |
| LAL | Large ATP binding of LuxR |
| LB | Lysogeny broth |
| LC-MS | Liquid chromatography - mass spectrometry |
| *m/z* | Mass/charge |
| MEME | Motif Em for Motic Elicitation |
| Mm | Methylenomycin |
| MMF | Methylenomycin furan |
| MUSCLE | Multiple Sequence Comparison by Log Expectation |
| NCBI | National Center for Biotechnology Information |
| NMR | Nuclear magnetic resonance |
| NRP | Non-ribosomal peptide |

| | |
|---|---|
| NRPS | Non-ribosomal peptide synthetase |
| OD | Ocular density |
| ORF | Open reading frame |
| PAC | P1-derived artificial chromosome |
| PCR | Polymerase chain reaction |
| PDB | Protein database |
| PK | Polyketide |
| PKS | Polyketide synthase |
| PUFAA | Polyunsaturated fatty acid amine |
| TOF | Time of flight |
| Red | Prodiginine |
| RNA | Ribonucleic acid |
| SARP | *Streptomyces* antibiotic regulatory protein |
| SCB | *Streptomyces coelicolor* butenolide |
| SDS | Sodium dodecyl sulphate |
| SerRS | Seryl tRNA synthetase |
| SFM | Soya flour mannitol |
| SMMS | Supplemented minimal medium solid |
| SMM | Supplemented minimal medium |
| STE | Sodium chloride-tris-EDTA |
| TAE | Tris-acetate-EDTA |
| TE | Thioesterase |
| tRNA | Transfer RNA |
| TSB | Tryptone soya broth |
| UHPLC | Ultra high performance liquid chromatography |
| UV | Ultraviolet |
| VPM | Valanimycin producing media |
| XRE | Xenobiotic response element |

# 1. Introduction

## 1.1. The need for novel and effective antibiotics

Since the discovery of penicillin in 1929, a multitude of novel antibiotics have been discovered to combat pathogenic infections; however, despite the success story of penicillin and the discovery of many antibiotics during 1945-65 (Figure 1-1), many pathogenic microorganisms have now developed resistance to these antibiotics.[1,2,3] Bacterial infections still remain as the second leading cause of death with approximately 17 million mortalities annually.[4] Over the past few decades, antibiotics resistance has been perpetuated by the rapid overuse and misuse of antimicrobial agents, and has led to the increase in multi-drug resistance organisms, such as methicillin and vancomycin-resistant *Staphylococcus aureus*.[5]

Due to the high cost, relatively poor return in investment and difficulty in finding novel and effective compounds, large pharmaceutical companies were losing interest in antimicrobial drug research.[1,6] With the rise and emergence of multidrug-resistant pathogens and long absence in antibiotic research from large pharmaceutical companies, the urgent need for new and novel antibiotics is ever increasing and some of the best sources of antimicrobial agents originate from soil bacterial microorganisms, such as *Streptomyces* bacteria (Figure 1-1).[7,8]

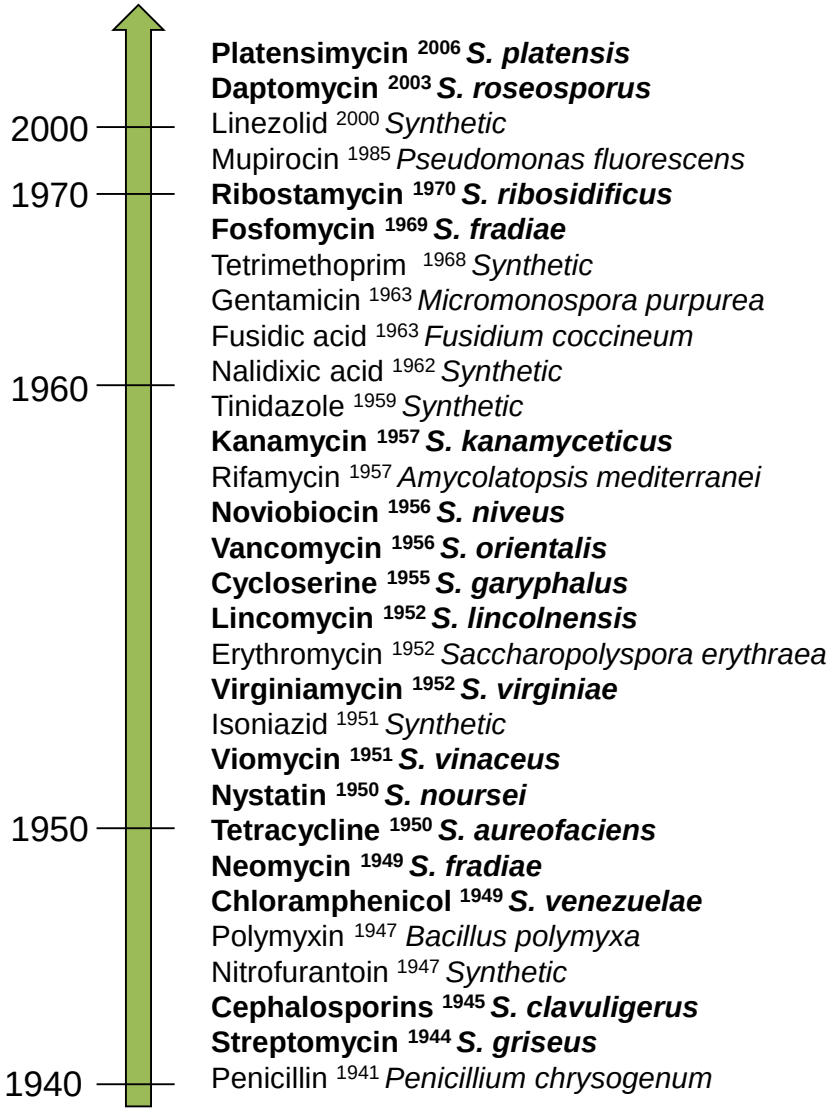



**Figure 1-1** – Many key antibiotics discovered originated from *Streptomyces* bacteria (highlighted in bold). Image adapted from reference 4.

## 1.2. *Streptomyces*, a prolific source of specialised metabolites

*Streptomyces* are Gram-positive filamentous bacteria, belonging to the Actinobacteria class, and have been extensively studied for their unique sporulating life cycle, complex transcriptional regulation, primary and secondary metabolism, and as a host for protein purification (Figure 1-2). [9,10,11] As soil-dwelling bacteria, *Streptomyces* live in an incredibly diverse microbial community, and are continuously competing and co-existing with other microorganisms to thrive in

their natural habitat. In response to specific nutrient conditions or presence of other invading microorganisms, *Streptomyces* are able secrete biologically active specialised metabolites and induce morphological differentiation (Figure 1-2). [12]

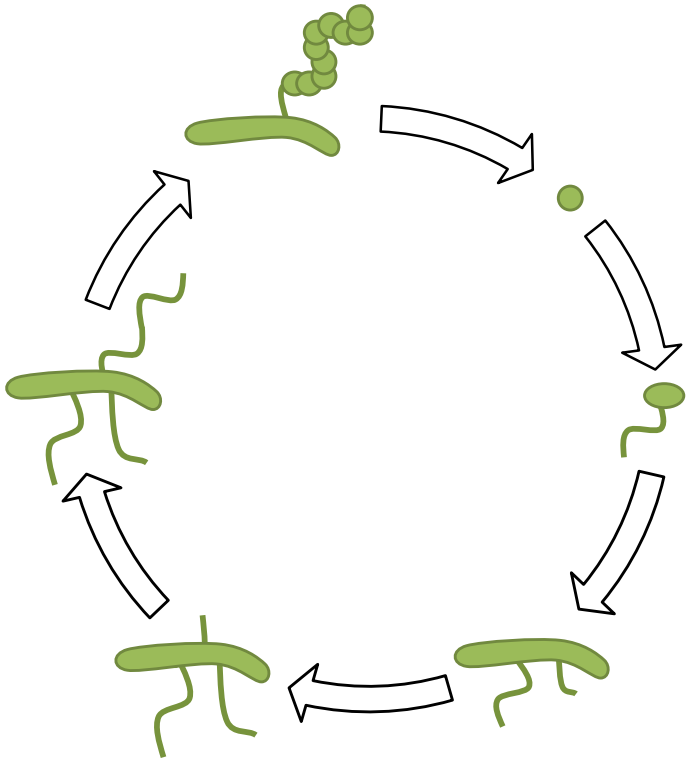


**Figure 1-2** – The life cycle of *Streptomyces* bacteria begins with the germination of a free spore. Hyphae protudes from the bacteria, and grows across and into the substrate media. Under poor environmental conditions, hyphae expand into the air and differentiate into spore chains. Image adapted from reference 13.

The first antibiotic compound ever isolated from a *Streptomyces* bacterium was streptothricin from *S. lavendulae* in 1942. Streptothricin exhibited antimicrobial activity against both Gram-positive and Gram-negative bacteria; however, it was too toxic to be used in humans. [14,15,16] Two years later in 1944, the first antibiotic discovered for human medicine was the bactericidal aminoglycoside streptomycin

from *S. griseus,* and was used to treat the disease, tuberculosis, by specifically targeting and inhibiting protein synthesis in *Mycobacterium tuberculosis*. [17,18,19,20] After the discovery of streptothricin and streptomycin, a number of important antibiotics were discovered and developed during the golden age of antibiotics between the 1940s and 1960s (Figure 1-1); two-thirds of antibiotics approved for clinic originate from the *Streptomyces* genus of bacteria (Figure 1-1). [21] The plethora of specialised metabolites assembled by *Streptomyces* bacteria are of pharmaceutical and medicinal interest as they carry a range of unique biological activity, such as anticancer, anti-inflammatory, antimicrobial, antifungal and immunosuppressive properties. [22] There are currently over 600 *Streptomyces* strains described so far and from bioinformatic analyses, each *Streptomyces* bacteria is predicted to produce dozens of biologically active specialised metabolites. [23,24]

The *Streptomyces* genus of bacteria is one of the most highly sequenced and as of May 2014, there were 19 finished genomic sequences (Table 1-1) and 125 draft assemblies in the Genbank database. [25] The first *Streptomyces* bacteria to have its genome sequence sequenced was *S. coelicolor* A3(2) in 2002 and the linear chromosome of *S. coelicolor* contained 8.7 Mb with 7,825 predicted open reading frames (ORF); the largest number of ORFs reported in a single bacterium at the time of sequencing. [26] Before the genome sequencing of *S. coelicolor* A3(2), five compounds had already been discovered*;* grey spore pigment, actinorhodin (Act), prodiginines (Red), calcium-dependent antibiotic (CDA), and methylenomycin (Mm) compounds (Figure 1-3). The analyses of *S. coelicolor* genome sequence revealed an additional 18 cryptic biosynthetic gene clusters and bioinformatic

analyses has greatly assisted the identification of these specific cryptic compounds.[26]

**Table 1-1** – Genome information of *Streptomyces* bacteria.

| Name | Size (Mbp) | GC % | ORF | # of predicted natural product biosynthetic clusters | Ref. |
|---|---|---|---|---|---|
| ***Streptomyces coelicolor* A3(2)** | | | | | |
| Linear chromosome | 8.67 | 72 | 7,825 | 23 | 26 |
| Linear SCP1plasmid | 0.37 | 69 | 355 | 2 | 26 |
| Circular SCP2 plasmid | 0.03 | 72 | 34 | 0 | 26 |
| ***Streptomyces griseus* IFO 13350** | | | | | |
| Linear chromosome | 8.55 | 72 | 7,138 | 34 | 27 |
| ***Streptomyces venezuelae* ATCC 10712** | | | | | |
| Linear chromosome | 8.23 | 72 | 7,536 | 31 | 28 |
| Circular pSVH1 plasmid | 0.01 | 71 | 11 | 0 | 28 |
| ***Streptomyces avermitilis* MA-4680** | | | | | |
| Linear chromosome | 9.03 | 71 | 7,574 | 38 | 29 |
| Linear SAP1 plasmid | 0.09 | 69 | 96 | 0 | 29 |
| ***Streptomyces hygroscopicus* subsp. Jinggangensis 5008** | | | | | |
| Linear chromosome | 10.15 | 72 | 8,849 | 29 | 30 |
| Linear pSHJG1 plasmid | 0.16 | 69 | 183 | 0 | 30 |
| Circular pSHJG2 plasmid | 0.07 | 71 | 75 | 0 | 30 |

Shortly after the genome sequencing of *S. coelicolor, S. avermitilis* MA-4680, which is responsible for the production of the anthelmintic avermectin compounds, had its genome completely sequenced in 2003 (Table 1-1). [29,31] The discovery of avermectins proved to be significant in the treatment of various parasitic diseases; its semi-synthetic derivative ivermectin is currently used in human and veterinary medicine to treat diseases, such as onchocerciasis and lymphatic filariasis (Figure 1-4). [32] In fact, the 2015 Nobel Prize in Physiology or Medicine was awarded to William Campbell and Satoshi Omura for the discovery and development of

avermectin compounds.[33] Genome analysis of *S. avermitilis* revealed a total of 38 predicted specialised metabolite gene clusters and as of 2014, the products of 16 biosynthetic gene clusters in *S. avermitilis* have been detected (bold lines in Table 1-2); an untapped source of specialised metabolites are waiting to be discovered.[29]

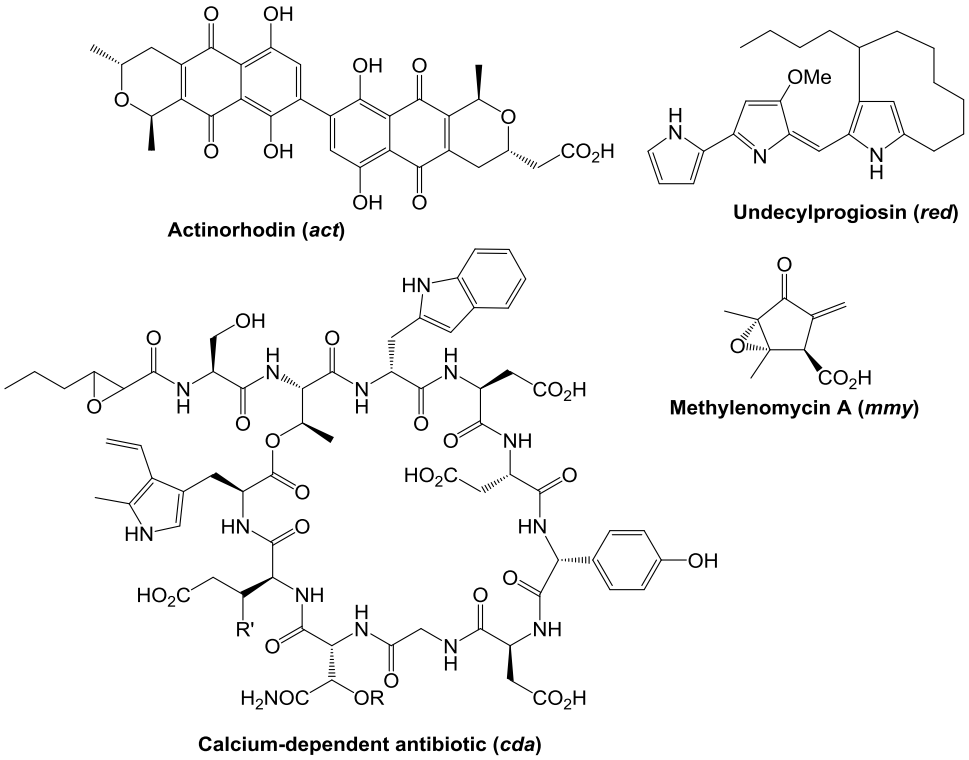
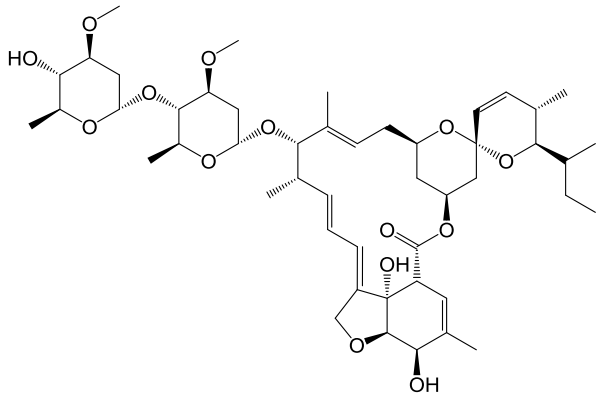



**Figure 1-3** – Specialised metabolites discovered in *S. coelicolor* A3(2).



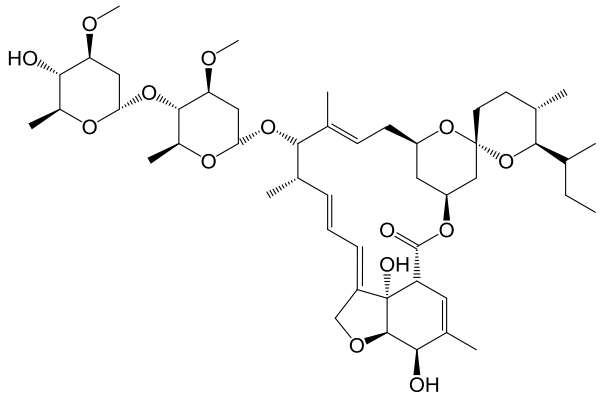




**Figure 1-4** – Structure of avermectin B1a compound discovered in *S. avermitilis* MA-4680. Its semi-synthetic derivative ivermectin contains a single reduced bond.

**Table 1-2** – Specialised metabolites predicted/ detected in *S. avermitilis*.

| # | Predicted/ observed compound | Genes | Approx. size (kb) | Ref. |
|---|---|---|---|---|
| **1** | **Avermitilol, avermitlone** | ***sav_76* (*ams*)** | **1.1** | **34** |
| 2 | PK | *sav_100-101* | 5.2 | |
| 3 | Microcin | *sav_257-sav_259* | 3.2 | |
| **4** | **Filipin** | ***sav_407-419* (pte)** | **81.6** | **35** |
| 5 | NRP | *sav_603-609* | 9.0 | |
| 6 | NRP | *sav_838-869* | 50.8 | |
| **7** | **Avermectin** | ***sav_935-953* (*ave*)** | **80.9** | **31** |
| **8** | **Isorenieratene** | ***sav_1019-1025* (*crt*)** | **8.7** | **35** |
| **9** | **Melanin** | ***sav_1136-1137 (melC)*** | **1.2** | **35** |
| 10 | PK-NRP | *sav_1249-1251* | 4.8 | |
| 11 | PK | *sav_1550-1552* | 19.0 | |
| **12** | **Squalene** | ***sav_1650-1654* (*hop*)** | **6.7** | **36** |
| **13** | **Geosmin, germacradienol** | ***sav_2163-2165* (*geo*)** | **4.4** | **37** |

| | | | | |
|---|---|---|---|---|
| 14 | AHFCA? | *sav_2266-2269* | 3.7 | |
| 15 | PK | *sav_2276-2282* | 11.0 | |
| 16 | PK | *sav_2367-2369* | 15.7 | |
| 17 | Aromatic PK | *sav_2372-2388* | 17.7 | |
| 18 | Siderophore | *sav_2465-2467* | 4.9 | |
| **19** | **Spore pigment** | ***sav_2835-2842 (spp)*** | **7.3** | **35** |
| **20** | **Oligomycin** | ***sav_2890-2903* (*olm*)** | **100.1** | **35** |
| **21** | **Pentalenolactone** | ***sav_2989-sav_3002* (*ptl*)** | **14.1** | **38** |
| **22** | **Albaflavenol, albaflavenone** | ***sav_3031-3032* (*ezs*)** | **2.5** | **39** |
| 23 | NRP | *sav_3155-3164* | 120.0 | |
| 24 | NRP | *sav_3192-3203* | 21.4 | |
| 25 | NRP | *sav_3636-3651* | 32.7 | |
| 26 | Aromatic PK | *sav_3653-3666* | 12.6 | |
| **27** | **Avenolide** | ***sav_3704, sav_3706*** | **3.4** | **40** |
| **28** | **Ochronotic pigment** | ***sav_5149 (hpd)*** | **1.1** | **41** |
| **29** | **Nocardamine, desferrioxamine B** | ***sav_5269-5274 (sid)*** | **7.2** | **42** |
| 30 | Melanin | *sav_5361-5362* | 1.2 | |
| 31 | Microcin | *sav_5686-5689* | 4.3 | |
| **32** | **Ectoine, 5-hydroxyectoine** | ***sav_6395-6398* (*ect*)** | **3.2** | **35** |
| 33 | NRP | *sav_6631-6633* | 5.8 | |
| **34** | **Tetrahydroxynaphthalene** | ***sav_7130-7131* (*rpp*)** | **2.3** | **35** |
| 35 | NRP | *sav_7161-7165* | 7.9 | |
| 36 | PK | *sav_7184-7186* | 8.0 | |
| 37 | Siderophore | *sav_7320-7323* | 6.1 | |
| 38 | PK | *sav_7360-7362* | 12.0 | |

Terpenes, polyketides, peptides and other compounds are highlighted in pale green, purple, blue and red respectively. Bold text indicate compounds that have been detected. Adapted from reference 35 and http://avermitilis.ls.kitasato-u.ac.jp/.

## 1.3. Genome mining methodologies for the identification of novel specialised metabolites

Methodologies and computational tools, such as the widely used Antibiotics & Secondary Metabolite Analysis SHell (antiSMASH), have been developed over the past years to aid in the annotation of genomes and predictions of novel

metabolites. [43] Recent developments of next-generation sequencing technologies and analytical chemistry instruments have resparked interests in antibiotic discovery. The identification and isolation of products from cryptic biosynthetic gene clusters (Figure 1-5) have been facilitated by highly sensitive analytical chemistry techniques, such as liquid chromatography - mass spectrometry (LC-MS). [44,45]

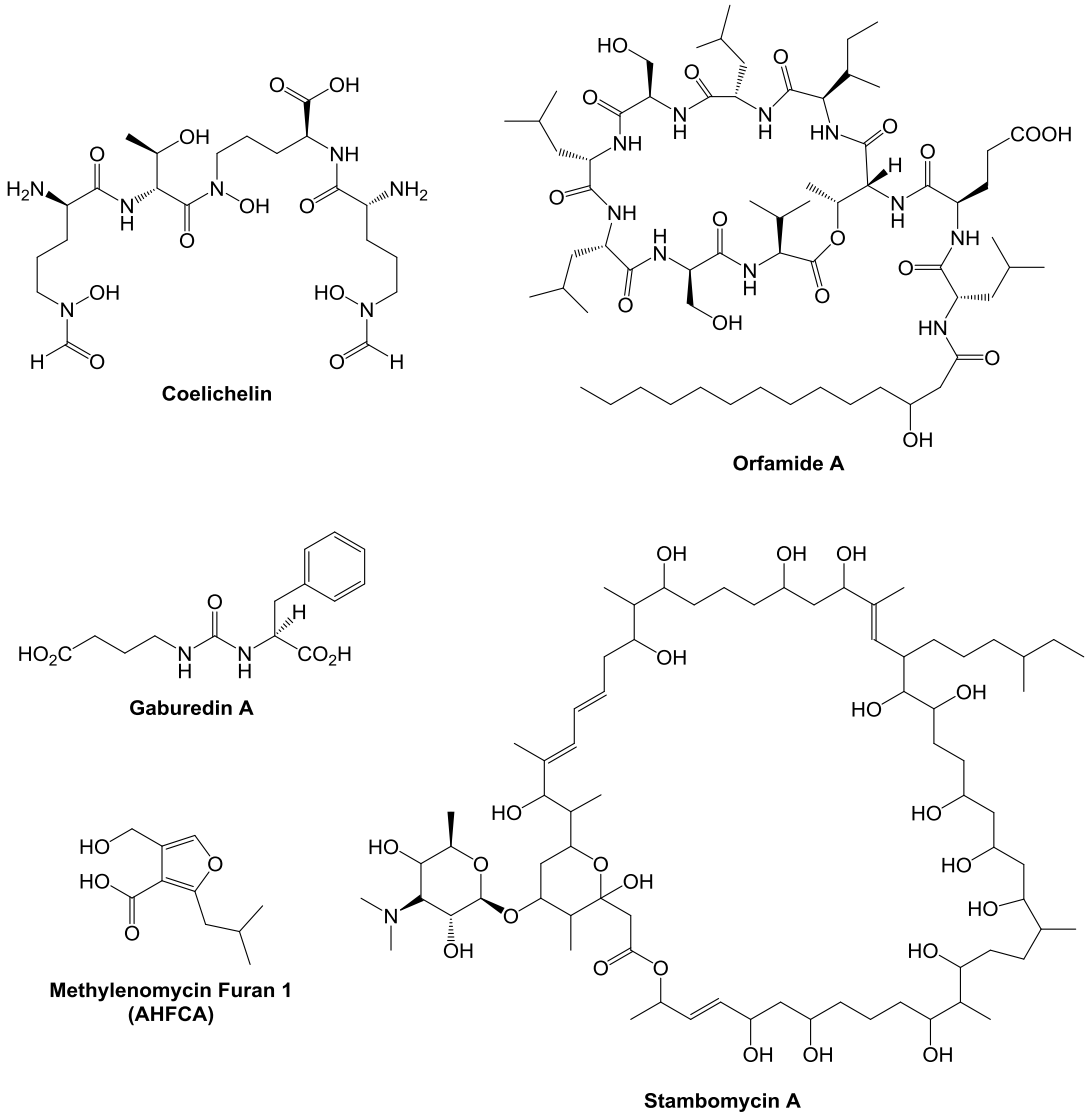



**Figure 1-5** – Natural products isolated from cryptic biosynthetic gene clusters.

Compounds of interest are normally detected by high performance liquid chromatography (HPLC) and can be further identified by ultraviolet (UV)

absorption. Most often the HPLC system is connected to a mass spectrometry and the sample can be directly injected into the mass spectrometry to obtain the mass/charge value (*m/z*) of all the compounds. Molecular formulae for the compounds of interest can be generated from the molecular ion *m/z* given by high resolution MS analysis and compared with similar compounds on the database, such as Reaxys. [46] Novel compounds of interest can be further isolated and purified using semi-preparative HPLC, and analysed by nuclear magnetic resonance (NMR) spectroscopy for structure determination. [47]

### 1.3.1. Prediction of substrate specificity from bioinformatic analysis

Proposed biological functions of biosynthetic genes can provide considerable insights into structure of compounds of interest, as well as its physiochemical properties. Examples of two major classes of specialised metabolites are the polyketides (PK) and non-ribosomal peptides (NRP), which are assembled by large modular biosynthetic machinery enzymes called polyketide synthases (PKS) and non-ribosomal peptide synthetases (NRPS) respectively. [48] Substrate specificities of adenylation and acyltransferase domains, in NRPS and PKS respectively, have been modelled from X-ray crystal structures and the building blocks can be accurately predicted from sequence analysis alone. [49,50,51] Furthermore, sequence analysis can also predict the stereochemistry of ketoreductase tailoring reactions in PK biosynthesis. [52,53] Substrate predictions of NRPS or PKS can be exploited to aid in the identification of products assembled by cryptic biosynthetic gene clusters (Figure 1-6). [44,45]

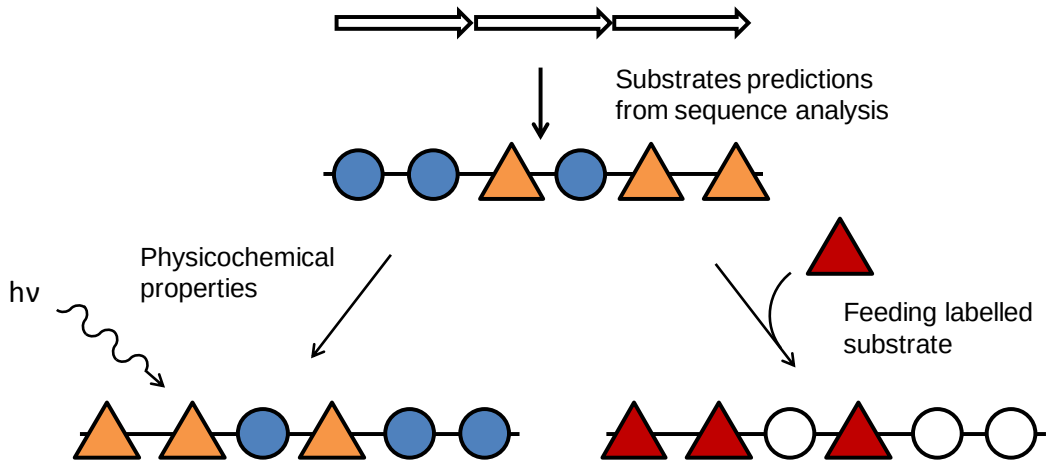



**Figure 1-6** – Prediction of physicochemical properties and genomisotopic approach to identify products of cryptic biosynthetic gene clusters.

The identification of coelichelin (Figure 1-5), from *S. coelicolor*, was greatly assisted by substrate predictions of the tri-modular CchH (SCO0492) NRPS enzyme. From the predicted structure, coelichelin was proposed to function as a siderophore and implied the predicted siderophore biosynthetic gene cluster would be expressed in an iron-deficient medium. Armed with this information, Lautru and co-workers were able to detect coelichelin in iron-deficient medium and consistent with their hypothesis, the production of coelichelin was decreased when ferric iron was added to the culture media. The proposed complex between ferric iron and coelichelin was detectable under UV-vis spectroscopy (Figure 1-6), and helped guide the purification of the ferri-coelichelin complex from *S. coelicolor* supernatant by HPLC and lead to structure determination by NMR spectroscopy. [54]

In another example, a genomisotopic approach was used to discover orfamide (Figure 1-5), a macrocyclic lipopeptide, from *Pseudomonas fluorescens* Pf-5 (Figure 1-6). Leucine amino acids were predicted as substrates for OfaA (PFL2145), OfaB (PFL2146) and OfaC (PFL2147) NRPS. Isotope-labelled L-leucine-$^{15}$N were fed to

the bacteria and the expected incorporation of isotope precursor was detectable by NMR spectroscopy; this helped guide fractionation purification and structure elucidation of orfamide A. [55]

### 1.3.2. Heterologous expression/ comparative metabolic profiling approach to identify products of cryptic biosynthetic gene clusters

Structure predictions in some complicated biosynthetic systems can be particularly challenging and sequence based approach would be unsuitable for identification of these complex metabolites. [44,45]

Putative biosynthetic genes or entire gene cluster can be introduced into heterologous hosts and comparative metabolic profiling approach by LC-MS analysis can detect the production of new compounds (Figure 1-7). Compounds that are only detected in heterologous hosts containing the genes are likely products of the biosynthetic genes, and can be further isolated and characterised. To further confirm the putative activity of biosynthetic genes, deletion of biosynthetic genes is expected to abolish the production of specific natural compounds (Figure 1-7). [44,45]

Heterologous expression approach was demonstrated in the discovery of 2-alkyl-4-hydroxymethylfuran-3-carboxylic acids (AHFCA) (Figure 1-5), which are a novel class of small diffusible signalling molecules that induce the production of specialised metabolites in *Streptomyces* bacteria. [56]

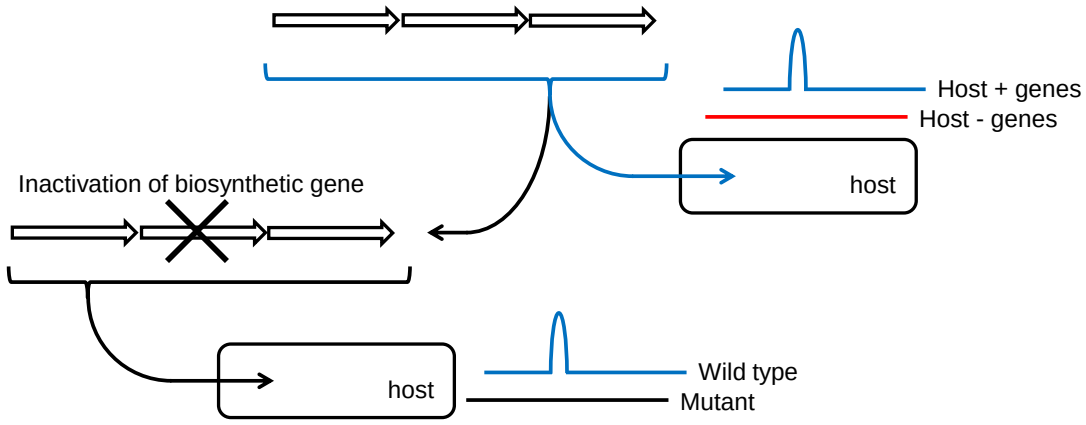



**Figure 1-7** – Heterologous expression/comparative metabolic profiling approach to identify products of cryptic biosynthetic gene cluster.

ScbA (SCO6266) in *S. coelicolor* and AfsA in *S. griseus* are the key biosynthetic enzymes for the assembly of γ-butyrolactones (GBL) signalling molecules (SCBs and A-factor respectively). Bioinformatic analyses identified an analogous protein in *S. coelicolor* called MmfL (SCP1.243) and the three gene operon *mmfL*, *mmfH* (*scp1.244*) and *mmfP* (*scp1.245*) were proposed to be responsible for the biosynthesis of signalling molecules. Comparative metabolic profiling of heterologous hosts with and without the *mmfLHP* operon successfully revealed five compounds named the methylenomycin furans (MMF), which belong to the AHFCA class of signalling molecules. Through a heterologous expression approach, a novel class of signalling molecules was discovered and they were shown to induce the production of methylenomycin antibiotics in *S. coelicolor*.[56] To further confirm the role of *mmfLHP* in the assembly of MMF signalling molecules, deletion of specific genes within the *mmfLHP* operon were shown to abolish the production of MMF and methylenomycins.[57]

### 1.3.3. Genetic modification to unlock and overproduce products of cryptic biosynthetic gene clusters

Identification of products from cryptic biosynthetic gene clusters in heterologous hosts can be challenging if biosynthetic gene clusters are poorly or not expressed in laboratory conditions. Most often, these cryptic biosynthetic gene clusters are tightly regulated at the transcriptional level and are under the control of DNA-binding transcriptional factors. [44,45,58,59] Transcription factors can either activate or repress expression of cryptic biosynthetic gene clusters and genetic manipulation of these specific transcriptional regulators are often necessary to derepress expression and overproduce the compounds of interest. [60]

The awakening of silent and cryptic biosynthetic gene cluster was successfully demonstrated in *S. ambofaciens* ATCC23877 through the overexpression of a cluster-situated, pathway-specific transcriptional activator (Figure 1-8), which led to the discovery of four antitumour macrolide compounds named the stambomycins (Figure 1-5). Stambomycins are biosynthesised by a silent PKS gene cluster, spanning almost 150 kb and containing 25 genes; under laboratory growth conditions, the stambomycin biosynthetic genes were not or very poorly expressed. Within the stambomycin gene cluster, the *samR0484* gene product encodes for a putative large ATP binding of LuxR (LAL) transcriptional activator and the overexpression of *samR0484* was shown to activate the expression of the PKS genes and unlocked the production of the stambomycins (Figure 1-8). [61]

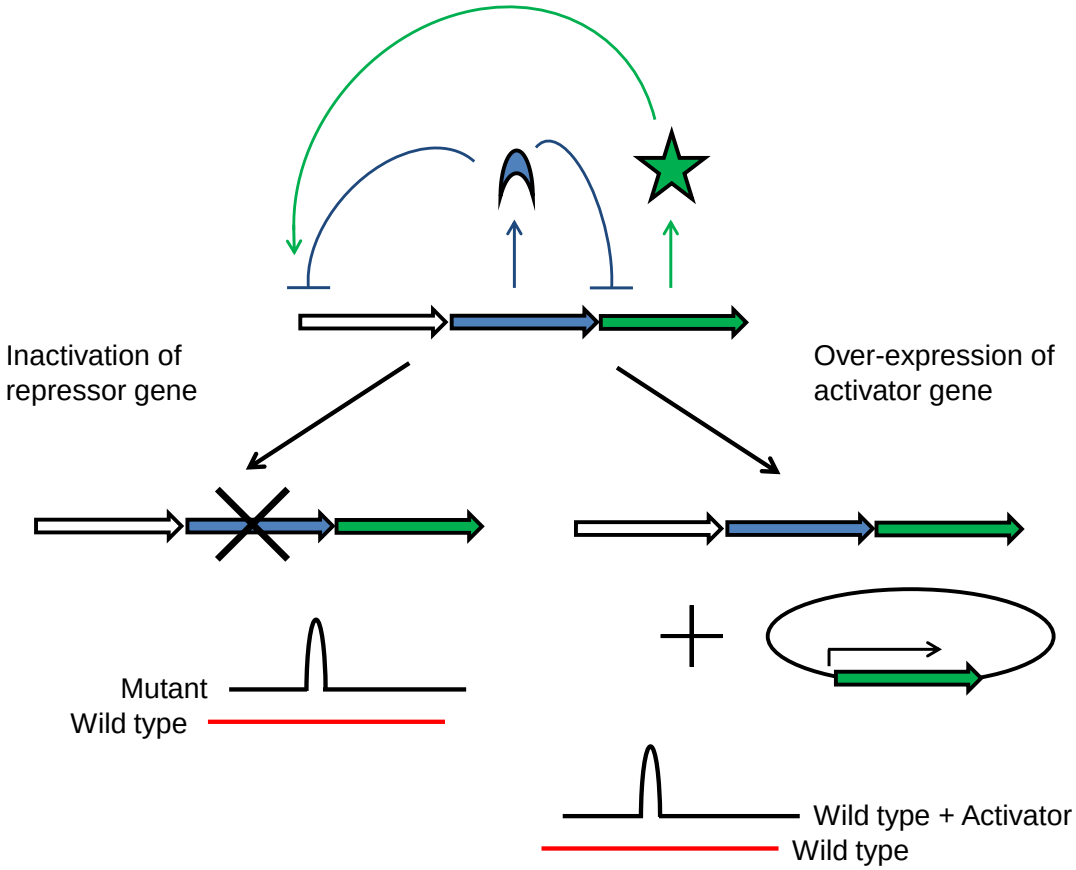



**Figure 1-8** – Inactivation or overexpression of transcriptional regulators/ comparative metabolic profiling approach to unlock and overproduce products of cryptic biosynthetic gene clusters.

In some cryptic biosynthetic gene clusters, pathway-specific transcriptional activators may not be clearly identifiable or are absent. Transcriptional repressors, such as the TetR family of repressors, are essential in the regulation of cryptic biosynthetic gene clusters in response to endogenous or exogenous signals and thus, specific inactivation of transcriptional repressors can sometimes awaken silent biosynthetic gene clusters. [60] The gaburedin biosynthetic gene cluster in *S. venezuelae* does not have any transcriptional activators and is tightly regulated by two cluster-situated TetR repressors. Specific inactivation of *gbnR* repressor (*sven_4187*) in *S. venezuelae* and comparative metabolic profiling led to the discovery of gaburedins (Figure 1-5), a novel family of γ-aminobutyrate-derived

urea compounds (Figure 1-9). Biosynthetic enzyme involved in gaburedin biosynthesis was further confirmed through the inactivation of *gbnB* gene, which encodes for a urea synthetase, and abolished the production of gaburedins (Figure 1-7). [62]

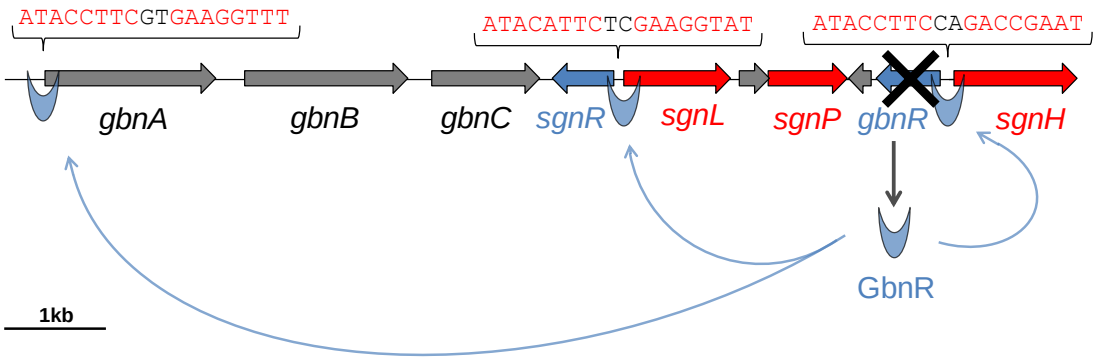



**Figure 1-9** – Inactivation of *gbnR* repressor led to the discovery of gaburedin compounds in *S. venezuelae*. Biosynthetic, repressor and signalling molecule genes are highlighted in grey, blue and red respectively. Image adapted from reference 62.

Continuous and rapid advancement of technologies and methods have accelerated the discoveries of novel specialised metabolite; for instance, the recently developed isolation chip (iChip) has been used to culture the "unculturable" bacteria and has led to the discovery of teixobactin, a new class of antibiotics with undetectable resistance, from *Eleftheria terrae*. [63] In addition, other genetic approaches have been applied to activate silent or cryptic biosynthetic gene clusters, such as the refactoring of biosynthetic genes and genetic modification of pleiotropic transcriptional regulators. [64,65]

## 1.4. The TetR family of transcriptional repressors

TetR repressors are widely distributed in bacterial organisms and most bacterial genomes contain at least one gene encoding for a putative TetR repressor. [66] The

TetR family of transcriptional repressors are primarily involved in regulating vital biological processes, such as antibiotic resistance and production (Table 1-3). [66,67,68]

**Table 1-3** – TetR repressors co-crystallised with cognate ligand or DNA operator.

| TetR name | Function | Crystallised with ligand? | Co-crystallised with DNA operator? | Ref. |
|---|---|---|---|---|
| TetR | Resistance | Tetracycline | TCTATCATTGATAGG | 69, 70 |
| QacR | Resistance | Diamidine | **CTTATAGACCGATC**GATCGGTCTATAAG | 71, 72 |
| CgmR | Resistance | Methylene blue | **CGTAACTGTACCGACC**GGTCGGTACAGTTACG | 73 |
| DesT | Lipid synthesis | Acyl-CoA | AGTGAACGCTTGTTGACT | 74 |
| SimR | Resistance | Simocyclinone | TTCGTACGGCGTACGAA | 75, 76 |
| HrtR | Heme homeostasis | Heme | ATGACACAGTGTCAT | 77 |
| MS6564 | Pleiotropic | No | TCATAAACGAGACGGTACGTCTCGTCTTGTG | 78 |
| SlmA | Cell division | No | **TTACGTGAGT**ACTCACGTAA | 79 |
| CprB | Pleiotropic | No | ATACGGGACGCCCCGTTTAT | 80 |
| ActR | Resistance | Actinorhodin | No | 81 |
| EthR | Antibiotic | Hexadecyl octanoate | No | 82 |
| MphR | Resistance | Erythromycin | No | 83 |

Palindromic sequences are highlighted in red.

To date, the National Center for Biotechnology Information (NCBI) public protein databases contains over 200,000 TetR repressor sequences and over 200 TetR crystal structures. [66,67] The structure of the TetR repressor from *Escherichia coli,* which regulates the resistance mechanism against tetracycline antibiotic, was the first to be elucidated in 1994 by X-ray crystallography. The TetR repressor protein functions as a homodimer and each identical monomer contains 10 α-helices (Figure 1-10). Co-crystallisation studies of the TetR repressor with its cognate ligand and

DNA operator revealed α-helix 1 – 4 at the N-terminus forms the DNA-binding domain and α-helix 4 – 7 are involved in binding to the specific cognate ligand. [69]

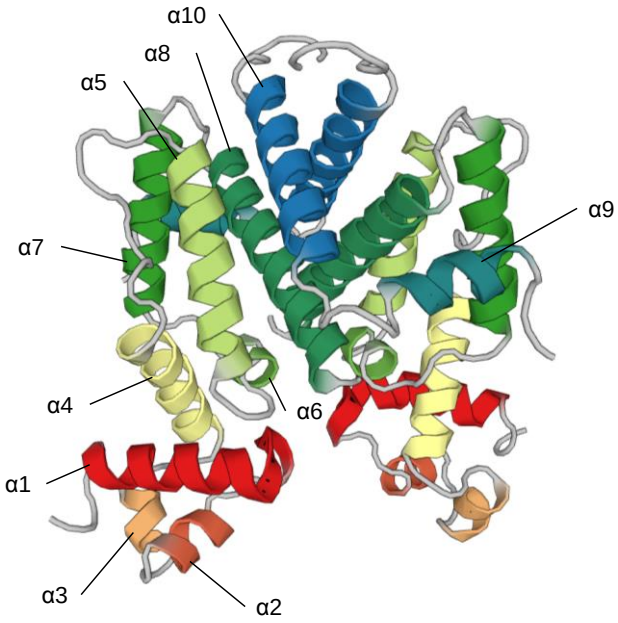



**Figure 1-10** – Structure of TetR repressor (PDB: 2TRT) from *E. coli*.

The TetR family of transcriptional repressors contain a highly conserved helix-turn-helix (HTH) DNA-binding domain, which is primarily composed of α-helix 2 and 3 (Figure 1-10). [70] The HTH DNA-binding domain is one of the most frequently observed DNA-binding domains in prokaryotic transcription factors; in fact, almost 95% of transcriptional factors utilise the HTH DNA-binding domain. [67] Since the TetR repressors are homodimeric, TetR repressors often interact with semi-palindromic or dyad sequences (palindromic sequences are highlighted in red in Figure 1-9 and Table 1-3) and the HTH DNA-binding domain binds to major grooves of DNA. Co-crystallisation and mutational studies have revealed critical amino acid residues required for DNA binding and are mostly located on α-helix 2 and 3 (Figure 1-11). [70] To date, predicting TetR repressor binding sites are still

extremely challenging and modelling of DNA binding sequence from primary amino acid sequences have been attempted.[84]

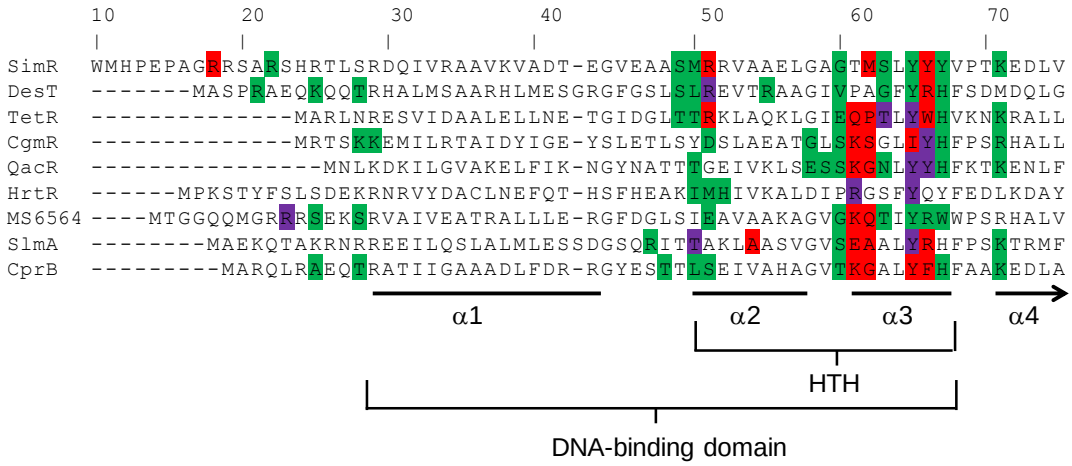



**Figure 1-11** – Clustal Omega alignment of TetR repressor HTH DNA-binding domain. Amino acid residues that interact with DNA bases, phosphate backbone or both are highlighted in red, green and purple respectively. Adapted from reference 79.

The tetracycline/TetR system regulates the resistance mechanism against tetracycline antibiotic in *E. coli.* Tetracycline is a broad spectrum polyketide antibiotic (Figure 1-12) that specifically inhibits protein translation and prevents the binding of aminoacyl-tRNA to the 30S ribosomal subunit. Various resistance mechanisms against tetracycline have evolved and developed in bacteria including ribosome protection and enzymatic inactivation of tetracycline; the most extensively studied resistance mechanism is the tetracycline/TetR efflux system.[85,86]

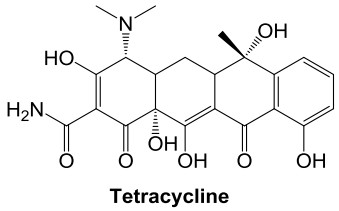



**Figure 1-12** – Structure of tetracycline aromatic polyketide antibiotic.

The *tetA* gene product encodes for a tetracycline efflux pump and is responsible for exporting toxic tetracycline antibiotic out of *E. coli* (Figure 1-13). [87] The *tetR* repressor gene is located divergently adjacent to the *tetA* gene and the TetR repressor tightly regulates expression of *tetA* by binding specifically to the 15 bp *tet* operator (*tetO*) located within the *tetR* - *tetA* intergenic region (Figure 1-13).[70,88,89]

In the presence of tetracycline, this molecule binds to the ligand-binding domain of the TetR repressor, which induces a cascade of conformational changes, and prevents the TetR repressor from binding to the *tetO* sequence (Figure 1-12).[70,90,91] Release of the TetR repressor from the *tetO* sequence allows transcription of *tetA.* The TetA efflux pump transports the toxic tetracycline molecules out of the *E. coli* bacteria (Figure 1-13). [89]

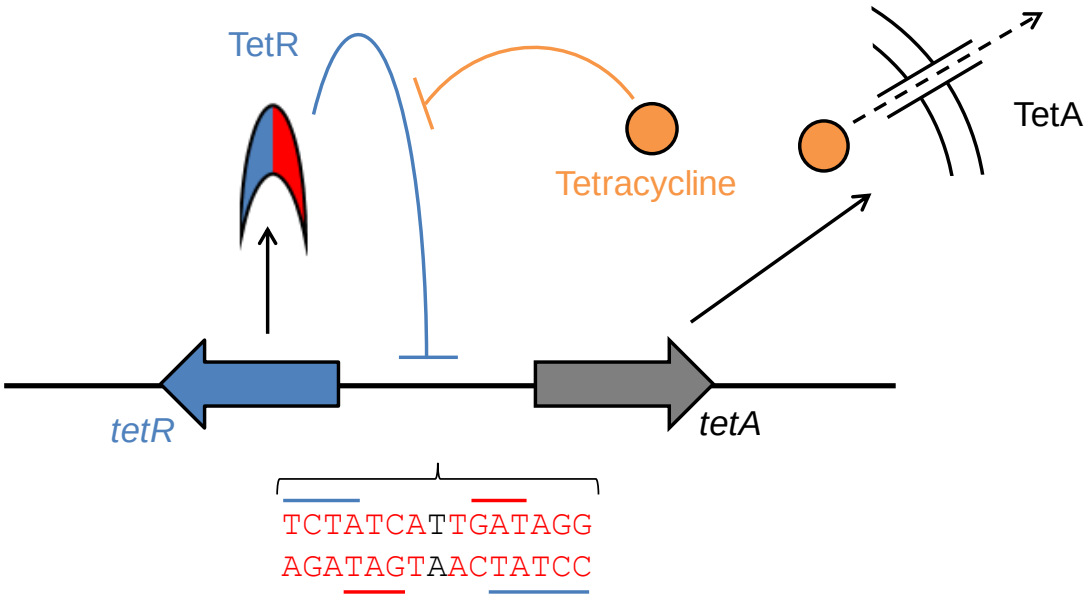



**Figure 1-13** – TetR repressor binds to the promoter of *tetA* and tightly regulates expression of *tetA*. Each identical monomer binds to semi-palindromic sequence (palindromic sequences are highlighted in red - monomer binding is highlighted with blue or red lines). The binding of tetracycline antibiotic (orange circle) to the TetR repressor derepresses the system. The TetA efflux pump exports the tetracycline antibiotic out of *E. coli*.

## 1.5. The ArpA class of transcriptional repressors

The ArpA class of the TetR family is an example of transcriptional repressors involved in regulation of specialised metabolites biosynthesis and are widespread in *Streptomyces* bacteria (Figure 1-14). Similar to the tetracycline/TetR system, presence of specific cognate signalling molecule abolishes the binding of ArpA to DNA operator sequence. However, ArpA repressors are most often involved in complex regulatory networks and controls, not only the expression of biosynthetic genes, but also the expression of itself and genes involved in the assembly of cognate signalling molecules. The exploitation of ArpA-like regulatory systems have led to the discoveries of novel metabolites (Figure 1-14).[92]

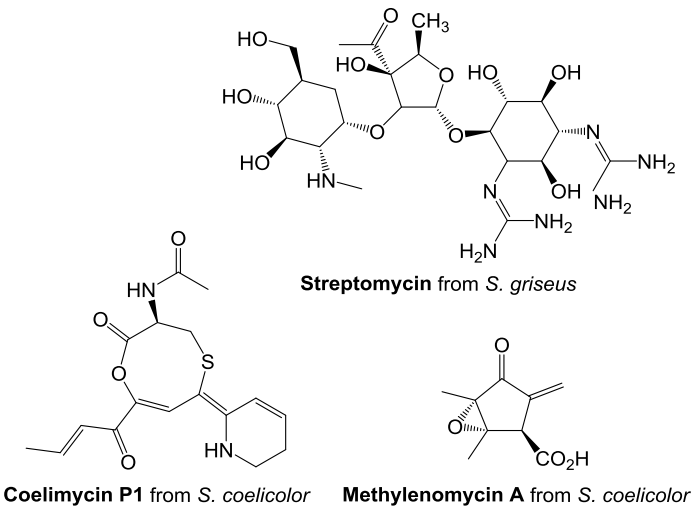



**Figure 1-14** – Specialised metabolites that are regulated by ArpA-like repressors.

The cognate signalling molecules that bind to ArpA-like repressors in *Streptomyces* bacteria, such as the γ-butyrolactones, are derived from intermediates in fatty acid metabolism and glycolysis (Figure 1-15). The A-factor signalling molecule was the first to be discovered and is biosynthesised by AfsA (SGR_RS34400) (Figure 1-15). [93] Analogous AfsA enzymes are widespread in *Streptomyces* bacteria and are proposed to produce a butenolide intermediate, which is subsequently modified to a butanolide or butenolide type signalling molecule (Figure 1-15). [94,95]

Genome mining analyses have identified analogous A-factor/ArpA systems in *Streptomyces* bacteria (Table 1-4) and have been reported to regulate different specialised metabolite biosynthetic gene clusters; for example, SCB/ScbR/ScbR2 and MMF/MmfR/MmyR systems in *S. coelicolor* regulates the biosynthesis of coelimycin and methylenomycins respectively (Figure 1-14). [95]

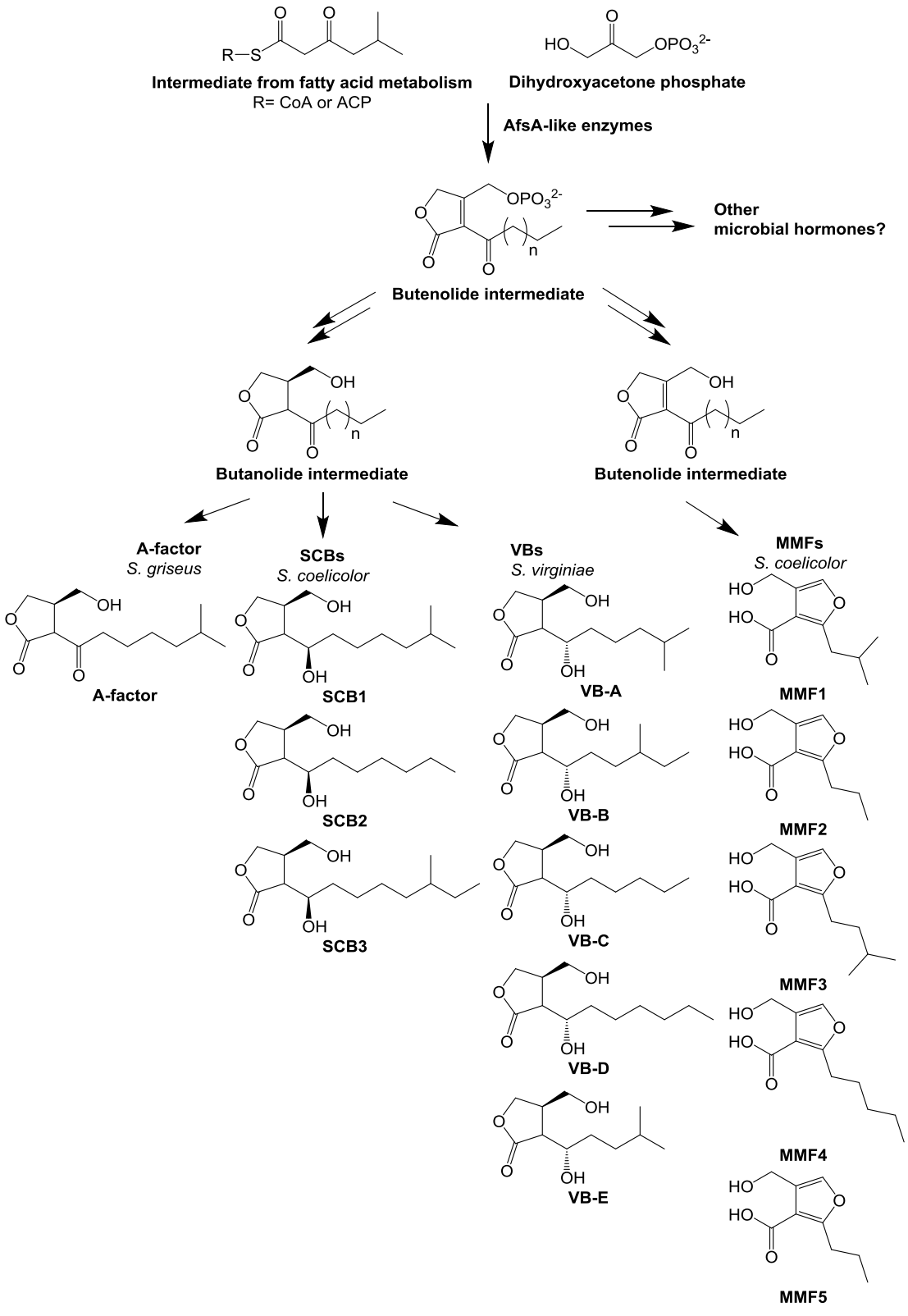



**Figure 1-15** – Diverse butanolide and butenolide signalling molecules are derived from intermediates of fatty acid metabolism and glycolysis. Adapted from reference 94.

**Table 1-4** – Analogous A-factor/ArpA regulatory systems in *Streptomyces* bacteria.

| Organism | Signalling molecule | AfsA-like enzyme | ArpA-like receptor | Antibiotic | Activator | Ref. |
|---|---|---|---|---|---|---|
| *S. acidiscabies* | ? | SabA | SabR, SabS | Polyketide WS5995B | ? | 96 |
| *S. chattanogensis* | CHB | ScgA | ScgR | ? | ? | 97 |
| *S. coelicolor* | SCB | ScbA | ScbR, ScbR2 | Coelimycin | CpkO | 98,99 |
| *S. coelicolor* | AHFCA | MmfL | MmfR, MmyR | Methylenomycin | MmyB | 56,57 |
| *S. griseus* | A-factor | AfsA | ArpA | Streptomycin | AdpA | 27 |
| *S. lavendulae* | IM-2 | FarX | FarA | Indigoidine | FarR1, FarR2 | 100,101 |
| *S. rochei* | SRB1, SRB2 | SrrX | SrrA, SrrB, SrrC | Lankamycin, Lankacydin | SrrY | 102,103, 104 |
| *S. tsukubaensis* | ? | BulS1, BulS2 | BulR1, BulR2 | Tacrolimus | BulY | 105 |
| *S. venezuelae* | AHFCA | SgnL | SgnR, GbnR | Gaburedin | None | 62 |
| *S. venezuelae* | SVB | JadW1 | JadR2, JadR3 | Jadomycin | JadR1 | 106,107 |
| *S. virginiae* | VB | BarX | BarA, BarB | Virginiamycin | ? | 108,109 |

### 1.5.1. Regulation of streptomycin biosynthesis and morphological differentiation by the A-factor/ArpA signalling system in *S. griseus*

The ArpA transcriptional repressor protein (SGR_3731) from *S. griseus* was the first and most extensively studied ArpA repressor in *Streptomyces* bacteria.[110] Deletion of the *arpA* gene in *S. griseus* resulted in overproduction of streptomycin (Figure 1-4) and developed aerial hyphae earlier than the wild type strain.[111] The ArpA repressor binds to the 22 bp autoregulatory response element (ARE) sequence in the promoter of *adpA* (*sgr4742*) and directly regulates expression of *adpA* pleiotrophic transcriptional activator.[112] A-factor, which was discovered from *S. griseus* in 1960s, was demonstrated to be the cognate ligand for the ArpA repressor and in the presence of A-factor, binding of ArpA repressor to the ARE sequence was abolished (Figure 1-16).[112,113] The AdpA pleiotropic transcriptional

activators alters expression of more than a thousand genes and directly regulate expression of over five hundred genes; some of which are involved in morphological differentiation and streptomycin production. [114]

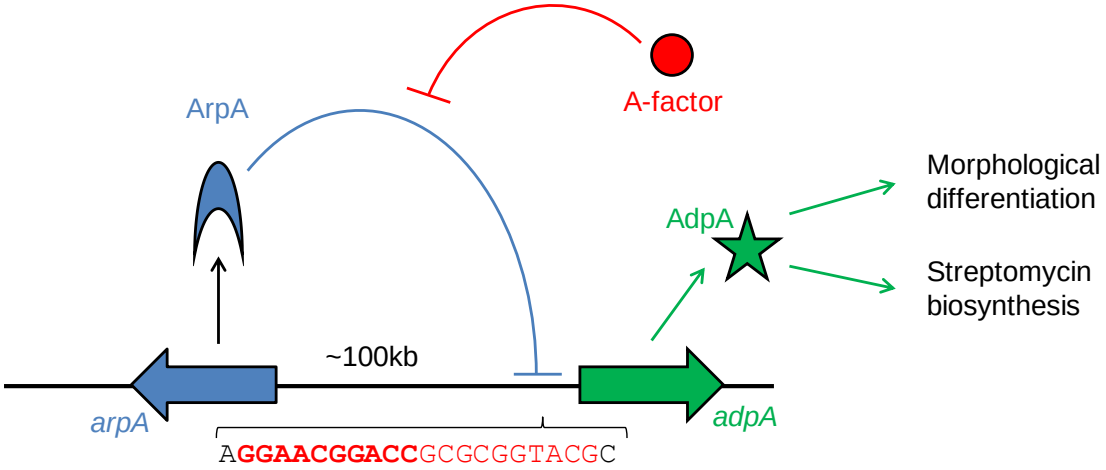



**Figure 1-16** – ArpA repressor binds upstream and regulates the expression of *adpA* pleiotrophic regulator. In the presence of A-factor (red circle), the A-factor ligand binds to ArpA and prevents the repressor from binding to the 22 bp ARE sequence (palindromic sequences are highlighted in red).

### 1.5.2. Regulation of coelimycin biosynthesis by the SCB/ScbR/ScbR2 signalling system in *S. coelicolor*

In *S. coelicolor,* the cryptic polyketide (*cpk*) gene cluster is involved in the biosynthesis of coelimycin, a polyketide alkaloid (Figure 1-14). The two ArpA-like transcriptional repressors, ScbR (*sco6265*) and ScbR2 (*sco6286*), as well as the two genes that encode for *Streptomyces* antibiotic regulatory proteins (SARP) transcriptional activators, *cpkO* (*sco6280*) and *cpkN* (*sco6288*), are all located and clustered together on the 58 kb *cpk* gene cluster. [115]

SARP family of transcriptional activators are commonly involved in increasing the expression of *Streptomyces* specialised metabolite biosynthetic gene clusters.[116] Coelimycin biosynthesis is dependent on the expression of *cpkO* as the deletion of *cpkO* abolishes the production of coelimycin.[117] CpkN has not been as extensively studied; however, overexpression of *cpkN* also results in the increase production of coelimycin.[118]

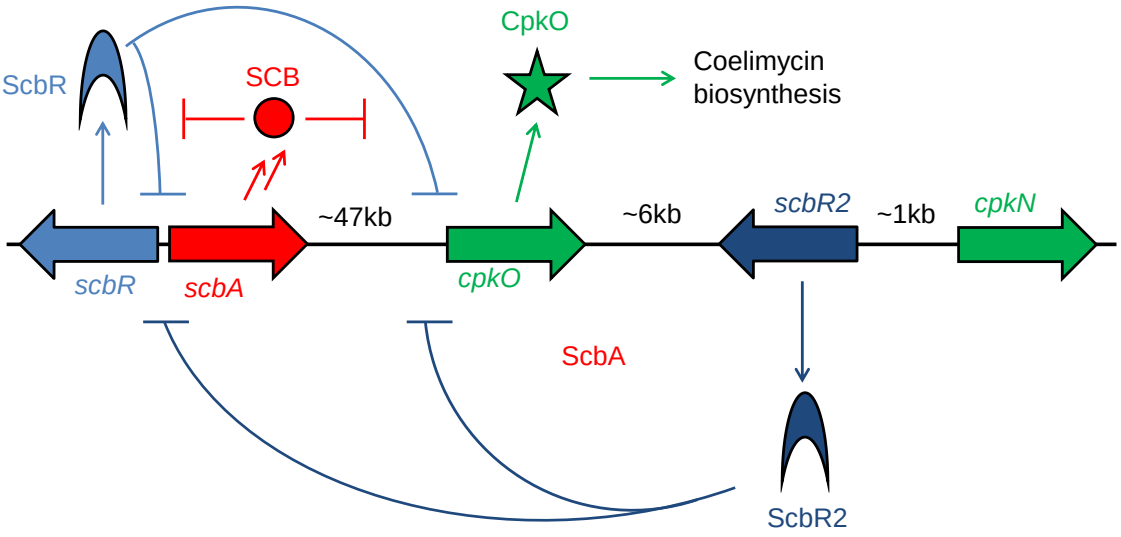
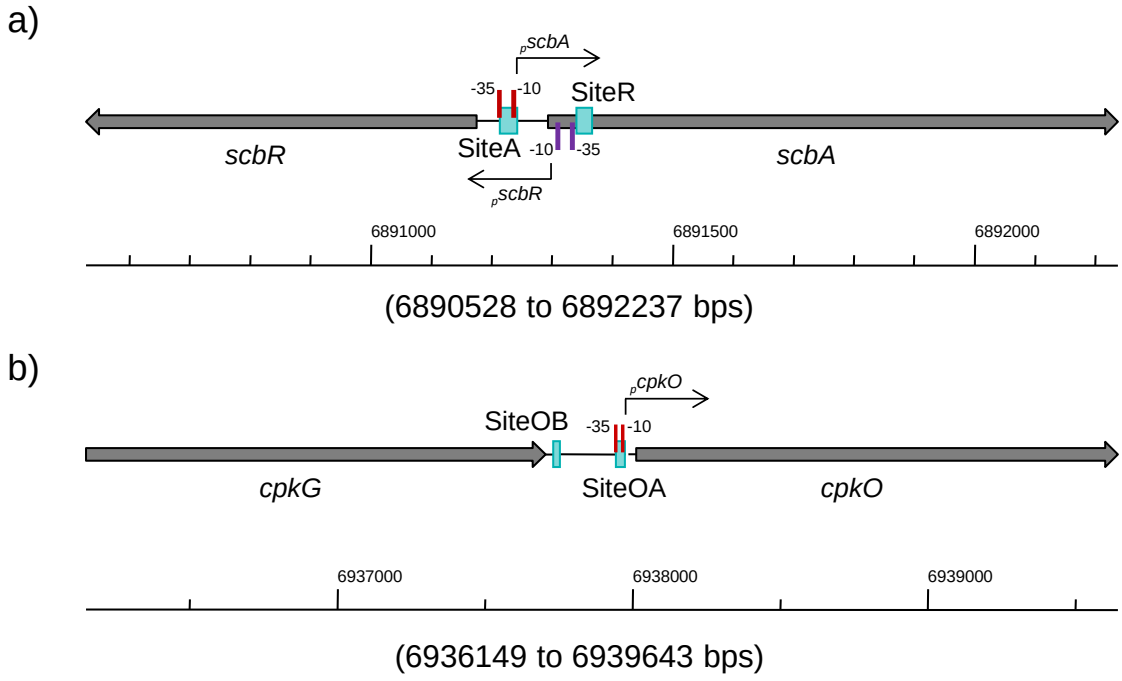



**Figure 1-17** – ScbR and ScbR2 repressor binds to promoters of *scbA* and *cpkO* to regulate the production of SCB signalling molecules and coelimycin respectively. SCB signalling molecules only bind to the ScbR repressors and results in derepression of the system.

In the SCB/ScbR/ScbR2 system, the ScbA enzyme, which shares 65% identity and 74% similarity over 292 amino acids with AfsA, catalyses the analogous reaction in the biosynthesis of SCB signalling molecules and are the cognate ligands for ScbR (Figure 1-15).[119] The *scbA* gene is divergent and adjacent to the *scbR* repressor gene, whilst in the A-factor/ArpA system, the *afsA* gene is located over 100 kb away from *arpA*. From electromobility shift assays (EMSA) and DNase I footprinting, ScbR

repressor was shown to bind to SiteA and SiteR ARE sequences located within the *scbR* - *scbA* intergenic region, as well as to SiteOA and SiteOB in the *cpkO* promoter (Figure 1-17 and Figure 1-18). The binding of ScbR repressor to the ARE sequences were abolished in the presence of SCB signalling molecules (Figure 1-17). [98,117]




c)

```
 SiteA: GAAAAAAAACCGCTCAGTCTGTATCTTAA
SiteOB: CAAAACAGACTGTGTTAGCTGTTT
         **** * ** *       **** *
 SiteR:  GGAACCGGCAATGCGGTTTGTTCGATC
SiteOA: ACAAACCGGTGTGCTGGTTTGTAAAGTCGTGG
          ******       *******    **
```

**Figure 1-18** – ScbR repressor was found to interact with (a) SiteA and SiteR in the *scbR* - *scbA* intergenic region, and (b) SiteOA and SiteOB in the *cpkO*. The ScbR2 repressor could only interacted with SiteA and SiteOB. (c) The ARE sequences are highly conserved. Palindromic sequences are highlighted in red and conserved nucleotides are marked with asterisk.

ScbR2 repressor, which shares 32% identity and 50% similarity over 194 amino acid with ScbR, is located on the other side of the *cpk* gene cluster (Figure 1-17) and

through similar EMSA and DNase I footprinting analyses, ScbR2 repressor was established to only bind to SiteA and SiteOB, which contains one half of the palindromic sequence (Figure 1-18). [120] Furthermore, binding of ScbR2 to the ARE sequences were not abolished in the presence of SCB signalling molecules, which suggests the SCB signalling molecules are not the cognate ligand for ScbR2. Instead, interactions between ScbR2 repressor and ARE sequences were prevented in the presence of endogenous metabolites produced by *S. coelicolor*; actinorhodin and undecylprodigiosin (Figure 1-3). [99,120]

The deletion of the ArpA-like repressors was able to over-produce coelimycin. Inactivation of *scbR*, as well as the addition of SCB signalling molecules did not result in the production of coelimycin, but instead, altered the levels of actinorhodin and undecylprodigiosin. [115,119,121] To further complicate the SCB/ScbR/ScbR2 system, other pleiotropic transcriptional regulators have been reported to directly regulate the *cpk* gene cluster, such as nutrient-sensing DasR regulator (SCO5231), PhoP phosphate response regulator (SCO4230) and AfsQ1/Q2 two component system (SCO4907/SCO4906). [122,123,124,125]

### 1.5.3. Regulation of methylenomycin biosynthesis by the MMF/MmfR/MmyR signalling system in *S. coelicolor*

MmfL, which shares 27% identity and 40% similarity over 274 amino acids with AfsA in *S. griseus,* is required for the assembly of AHFCA signalling molecules (Figure 1-15). Comparative metabolic profiling of heterologous hosts with and without the *mmfLHP* operon led to the discovery of MMF signalling molecules as discussed in Chapter 1.3.2. [56]

The cassette of five genes, *mmfR* (*scp1.242c*), *mmfL*, *mmfH*, *mmfP* and *mmyR* (*scp1.246*), forms the AHFCA-dependent signalling system and is adjacent to a group of genes involved in the biosynthesis of methylenomycins, a colourless cyclopentanoid antibacterial metabolite (Figure 1-3 and Figure 1-19). The 19 kb *Mm* biosynthetic gene cluster contains a total of 21 genes, including biosynthetic, resistance and regulatory genes (Figure 1-19), and are all clustered on the linear SCP1 plasmid. [126] The production of methylenomycin is dependent on the expression of *mmyB*, which encodes for a putative xenobiotic response element (XRE) family of transcriptional activators, and the in-frame deletion of *mmyB* abolishes the production of methylenomycin. Preliminary bioinformatic analyses have identified two direct repeats of 16 bp sequence called the "B-box" sequence and are located in *mmyB* - *mmyY* intergenic region and promoter of *mmyT* (*scp1.241c*); MmyB activator is proposed to bind to the "B-box" sequence to activate expression of methylenomycin biosynthetic genes. [57]

Nutrient sensing transcriptional regulators may also play a role in controlling the production of methylenomycin. Acidic shock and limiting alanine conditions were shown to trigger the production of methylenomycin in *S. coelicolor*. [127] In addition, *mmfL* and *mmyB* genes both contain the rare TTA codon. [57] BldA, which encodes for the rare UUA leucine tRNA, is a global regulator of morphological differentiation and metabolite production, and predominantly accumulates during late exponential and stationary phase. [128,129,130] Therefore, translation of *mmfL* and *mmyB* is dependent on the presence of mature BldA.

Further preliminary bioinformatic analyses of *Mm* gene cluster have identified highly conserved 18 bp ARE sequence in *mmfR* - *mmfL* and *mmyY* - *mmyB* intergenic region and *mmyR* promoter (Figure 1-19).[57] EMSA studies have demonstrated that the MmfR repressor can bind to specific ARE sequences (Dean, R. *et al*. Manuscript in preparation). The MmyR repressor, which is 35% identical and 56% similar to MmfR over 110 amino acids, is also proposed to bind to the same ARE sequences.[57] The exact cognate signalling molecule and DNA binding interactions for MmyR is currently under investigation; however, early results suggest the MmyR repressors do not bind to the same MMF signalling molecules as MmfR (Styles, K. Personal communication).

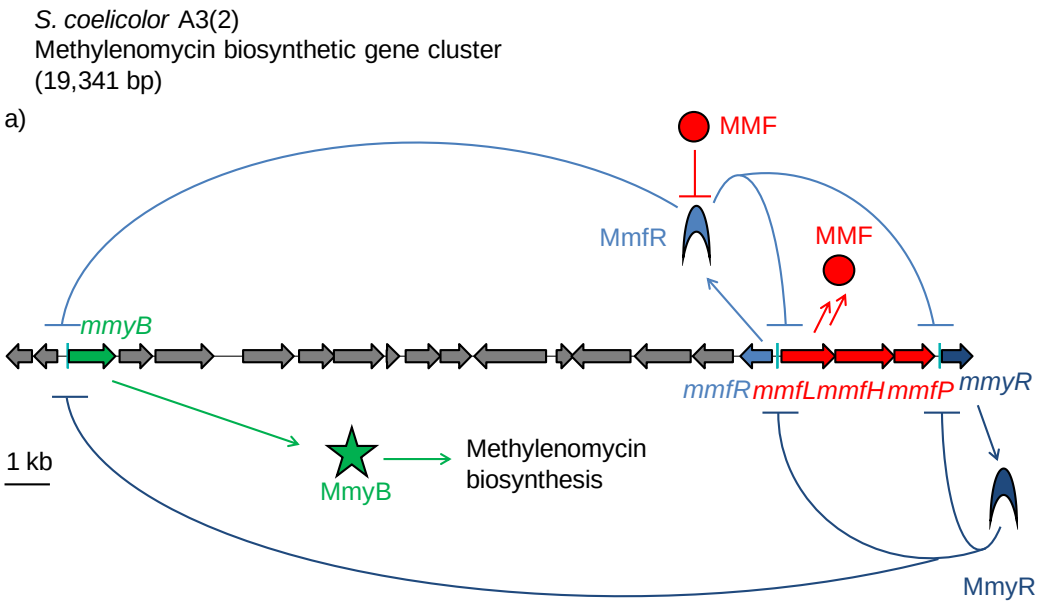



**Figure 1-19** – (a) In absence of MMF signalling molecules (red circle), MmfR repressor binds to *mmfR* - *mmfL* and *mmyY* - *mmyB* intergenic region and *mmyR* promoter to regulate the biosynthesis of methylenomycin and MMFs.
(b) MmfR/MmyR repressors are proposed to bind to highly conserved semi-palindromic sequences (palindromic sequences are highlighted in red).

S1 nuclease protection analyses revealed transcripts of *mmfR* and *mmfL* were transcribed at the beginning of growth, and expression of *mmyB* and *mmyR* was induced 48 hours into growth. The presence of MmfR repressor is proposed to tightly regulate expression of *mmyB, mmfL* and itself by binding to conserved ARE sequences and turning off the production of methylenomycins and MMFs (Figure 1-19). In presence of mature BldA, *mmfL* is translated to a functional protein and assembles the MMF signalling molecules. The signalling molecules bind to the MmfR repressor and begins to derepress the system; however, derepression of *mmfR* would also result in the increase transcription of *mmfR*. Once concentrations of the MMF signalling molecules reaches a critical threshold (through the biosynthesis of *mmfLHP* or from the environment as they are freely diffusible molecules), MmfR repressor is released from the *mmyB* promoter and the biosynthesis of methylenomycins is initiated. MmyR repressor is proposed to act as an off-switch to stop the production of methylenomycin compounds. Hence, specific inactivation of *mmyR* repressor gene unlocks and overproduces methylenomycin and MMF signalling molecules; whilst the deletion of *mmfR* results in no phenotypic difference when compared to the wild type strain. [57]

Genome mining analyses have identified analogous AHFCA-dependent signalling gene clusters in several other *Streptomyces* bacteria (4 to date), such as in *S. venezuelae* (Figure 1-9). Genetic inactivation of the analogous *mmyR* repressor (*gbnR*) in *S. venezuelae* led to the discovery of a novel family of natural products, the gaburedins (discussed in Chapter 1.3.3). [62] Exploitation of AHFCA-dependent signalling systems may lead to the discovery of other novel specialised metabolites in *Streptomyces* bacteria.

## 1.6. Aims of the project

Systems that utilise TetR repressors and butenolide/butanolide signaling molecules are widely distributed in *Streptomyces* bacteria and are often involved in controlling natural product biosynthesis. *In silico* predictions of TetR transcriptional repressor binding sites would provide unprecedented information of transcriptional regulation involved in controlling the expression of cryptic natural product biosynthetic gene clusters. Results from such analyses can be further exploited in experiments aimed at discovering novel compounds. The project aims to investigate a previously identified putative AHFCA-dependent signalling system proposed to tightly regulate the expression of a predicted azoxy natural product biosynthetic gene cluster in *Streptomyces avermitilis*. The following objectives will be pursued:

- Prediction of transcriptional repressor binding sites in putative AHFCA-dependent and GBL-dependent signalling systems in *Streptomyces* bacteria.
- Bioinformatic and genome mining analyses of compounds produced by the cryptic biosynthetic gene cluster proposed to be under the control of a putative AHFCA-dependent signalling system in *S. avermitilis*.
- Identification of predicted cryptic metabolites using a heterologous expression/ comparative metabolic profiling approach.
- Characterisation of predicted metabolites through the incorporation of isotope-labelled precursors and to determine antimicrobial activity of the metabolites. Compounds of interest will also be isolated by semi-preparative HPLC and structure determined by a combination of 1- and 2-D NMR spectroscopy.
- Exploitation of AHFCA-dependent signalling systems to unlock and overproduce the metabolites of interest.

# 2. Materials and Methods

Recipes and protocols were followed and adapted from the manufacturer's instructions and from the "PCR targeting system in *Streptomyces coelicolor* A3(2)" protocol and Practical *Streptomyces* Genetics book. [21,131]

## 2.1. Materials

### 2.1.1. Microbial strains

**Table 2-1** – List of microbial strains used in this work.

| Strain | Uses/ Genotype | Obtained from |
|---|---|---|
| ***Escherichia coli*** | | |
| DH5α | General cloning host.<br><br>*fhuA2 lac(del)U169 phoA glnV44 ϕ80' lacZ(del)M15 gyrA96 recA1 relA1 endA1 thi-1 hsdR17* | New England Biolabs |
| TOP10 | General cloning host.<br><br>*F- mcrA Δ(mrr-hsdRMS-mcrBC) ϕ80dlacZΔM15 ΔlacX74 nupG recA1 araD139 Δ(ara, leu)7697 galE15 galK16 rpsL(Str$^R$) endA1 λ$^-$* | Invitrogen |
| ET12567 | Non-methylating host.<br><br>ET12567: *dam dcm hsdS cat tet* | JIC, Norwich, UK |
| BW25113 | Strain for PCR targeted deletion.<br><br>*F$^-$, DE(araD-araB)567, lacZ4787(del)::rrnB-3, LAM$^-$, rph 1, DE(rhaD-rhaB)568, hsdR514* | JIC, Norwich, UK |
| ET12567/pUZ8002 | ET12567 containing non-transmissible plasmid.<br><br>pUZ8002 plasmid: *tra neo RP4* | JIC, Norwich, UK |
| DH10B/pESAC13A | Host containing pESAC13A constructs.<br><br>DH10B: F$^-$ mcrA *Δ(mrr-hsdRMS-mcrBC) ϕ80dlacZΔM15 ΔlacX74 endA1 recA1 deoR Δ(ara, leu)7697 araD139 galU galK nupG rpsL λ$^-$* | Bio S&T, Canada<br>This study. |

| | | |
|---|---|---|
| | pESAC13A: *aacIV attp-integrase* | |
| DH5α/pUB307 | Strain for mobilising plasmids into bacteria.<br><br>pUB307 plasmid: *RP4 carb* | JIC, Norwich, UK |
| DH5α/pR9604 | Strain for mobilising plasmids into bacteria.<br><br>Derivative of pUB307 with kanamycin resistance replaced with carbenicillin resistance. | JIC, Norwich, UK |
| DH5α/pER2 | DH5α with pER2 plasmid. | Remstam, E. (Corre Group), University of Warwick, Coventry, UK |
| TOP10/pER2Δint | DH5α with pER2Δint plasmid. | This study. |
| BW25113/pIJ790 | Strain for PCR targeted deletions.<br><br>BW25113 with pIJ790 plasmid | JIC, Norwich, UK |
| BW25113/pKD20 | Strain for PCR targeted deletions.<br><br>BW25113 with pKD20 plasmid | JIC, Norwich, UK |
| K12/pMK-RQ_SAV_2301 | K12 with pMK-RQ_SAV_2301 construct. | Life Technologies.<br>This study. |
| ***Streptomyces avermitilis*** | | |
| MA-4680 | Wild type strain. | USDA, USA |
| ***Streptomyces lividans*** | | |
| TK24 | Derivative of *Streptomyces lividans* 66.<br>*str-6*, *SLP2*$^{-}$, *SLP3*$^{-}$ | JIC, Norwich, UK |
| ***Streptomyces coelicolor*** | | |
| M1152 | Derivative of *Streptomyces coelicolor* M145 (SCP1$^{-}$ SCP2$^{-}$).<br><br>*rpoB[C1298T] Δact red cpk cda* | JIC, Norwich, UK |
| ***Bacillus subtilis*** | | |
| 168 | Strain used in antimicrobial assay<br>*ind- tyr+* | Hodgson Group, University of Warwick, Coventry, UK |
| ***Staphylococcus aureus*** | | |
| | Strain used in antimicrobial assay | Wellington Group, University of Warwick, Coventry, UK |

### 2.1.2. Plasmids

**Table 2-2** – List of plasmids used in this work.

| Name | Notes | Obtained from |
|---|---|---|
| pOSV556 | *amp$^R$, hyg$^R$, ermE*, oriT* , *int_pSAM2* | Pernodet Group, University of Paris-sud, France |
| pCC4 | Derivative of pOSV556 with kanamycin resistance replaced with apramycin resistance. | Corre Group, University of Warwick, Coventry, UK |
| pESAC13A-2H | *aacIV,* ~130 kb insert from *S. avermitilis* (containing predicted azoxy gene cluster). | Bio S&T, Canada<br>This study. |
| pESAC13A-19K | *aacIV,* ~130 kb insert from *S. avermitilis* (containing predicted azoxy gene cluster). | Bio S&T, Canada<br>This study. |
| pER2 | Derivative of pCC4.<br>*sav_2267* and *sav_2269* genes cloned into the *Hind*III and *Pst*I sites. | Remstan, E. (Corre Group), University of Warwick, Coventry, UK |
| pER2Δint | Derivative of pER2 with integrase system deleted. | This study. |
| pKD20 | *P araBAD, bet, exo, gam, repA101ts, bla* | JIC, Norwich, UK |
| pIJ790 | Derivative of pKD20 with ampicillin resistance replaced with chloramphenicol resistance. | JIC, Norwich, UK |
| pIJ777 | Kanamycin disruption cassette<br>*amp, neo* | JIC, Norwich, UK |
| pMK-RQ _SAV_2301 | pMK-RQ construct containing<br>*sav_2301* (TTA->GGA)<br>*neo* | Life Technologies.<br>This study. |

### 2.1.3. Primers

Primers were diluted to a final concentration of 100 mM with sterile $dH_2O$ and stored at -20 $^0$C.

**Table 2-3** – List of primers used in this work.

| Primer Name | Primer sequence | Target gene | Expected Size (bp) |
|---|---|---|---|
| for_pESAC_SAV2283a | AGATGCGGGTCGAGAAGGAGTACG | *sav_2283* | 480 |
| rev_pESAC_SAV2283a | AGCCGAAGAAGGACTCGGTGATGG | | |
| for_pESAC13_sav2266a | GTGGCCGACCATGTGCAGCATGGG | *sav_2266* | 425 |
| rev_pESAC13_sav2266a | GCGATGAGGCCAACCGGGAGTACC | | |
| for_pESAC_sav2301a | GCTACCGGCTCACGATCAACGAAA | *sav_2301* | 424 |
| rev_pESAC_sav2301a | GATGCCGAGTTCGTCCGAGAGGTG | | |
| for_integrase_screen | GCGAAGCGGCATGCATTTAC | *int_pSAM2* | 3214 (552) |
| rev_integrase_screen | GATCGGGCCTTGATGTTAC | | |
| for_avaA-B_screen | CGGCGACGTATCGTTCAC | *sav_2267-sav2269* | 831 (132) |
| rev_avaA-B_screen | GCCGACATGACCTGGAAG | | |
| for_sav_2268_deletion_screen | TCGCAGAACACCAAGTAG | *sav_2266-sav_2270* | 3,235 (2,536) |
| rev_sav2268_deletion_screen | CTCATGACGTCCATATCG | | |
| for_Target_SAV_2268 | GTTCACGAACGTATCTTTTATGCTTCGGTCCCCGCCGTGATTCCGGGGATCCGTCGACC | Kanamycin disruption cassette | 1,176 |
| rev_Target_SAV_2268 | CGTGGTGACCCGCGGCTCACCGGGCCGTGGCCGGCCTCATGTAGGCTGGAGCTGCTTC | | |
| for_Target_SAV_2268_2 | GTCCCCGCCGTGGAGCCGGCAAAGGAGGGGAGCGTCATGATTCCGGGGATCCGTCGACC | Kanamycin disruption cassette | 1,176 |
| rev_Target_SAV_2268_2 | TGGCCGGCCTCAGCGGGCCGGGGCGCCGCCGAGCGCCTGTGTAGGCTGGAGCTGCTTC | | |
| for_pKD20_screen | GTACGGCCTTAATCCGTGGAC | *bet* | 534 |
| rev_pKD20_screen | TCGCGGCGAAATATCTGGG | | |
| for_SAV_2301_2 | ACAAAGCTTAGGAGGAACGGCATGCTGAGCTTCTCG | *sav_2301* | 834 |
| rev_SAV_2301_2 | ACACTGCAGGCGGGCTCAGGCCGGCTGG | | |

### 2.1.4. Antibiotic solutions

Antibiotic solutions were filter sterilised before use and stored at 4 $^{0}$C.

**Table 2-4** – List of antibiotics used in this work.

| **Antibiotic** | **Stock Solution (mg/mL)** | ***E. coli* (µg/mL)** | ***Streptomyces* (µg/mL)** | **Solvent** |
|---|---|---|---|---|
| Ampicillin | 100 | 100 | - | $dH_2O$ |
| Apramycin | 50 | 50 | 50 | $dH_2O$ |
| Chloramphenicol | 25 | 25 | - | Ethanol |
| Hygromycin | 50 | - | 50 (100) | $dH_2O$ |
| Kanamycin | 50 | 25 | - | $dH_2O$ |
| Nalidixic Acid | 25 | 25 | 25 | 0.3M NaOH |

### 2.1.5. Solid media

Recipe for solid medium were prepared as follows:

**Lysogeny Broth (LB) Agar**

| | |
|---|---|
| Tryptone | 10 g |
| Yeast extract | 5 g |
| NaCl | 10 g |
| Agar | 15 g |
| $dH_2O$ | to 1 L |

**Soya Flour Mannitol (SFM) Agar**

| | |
|---|---|
| Soya Flour | 20 g |
| Mannitol | 20 g |
| Agar | 20 g |
| Tap water | to 1 L |

**Supplemented Minimal Medium Solid (SMMS) Agar**

| | |
|---|---|
| Casaminoacid | 2 g |
| TES buffer | 8.68 g |
| Agar | 15 g |
| $dH_2O$ | to 1 L |

After autoclave, add the following sterilised solutions.

| | |
|---|---|
| $NaH_2PO_4$ + $K_2HPO_4$ (50 mM each) | 10 mL |
| $MgSO_4$ (1 M) | 5 mL |
| Glucose (50% w/v) | 18 mL |
| Trace element solution | 1 mL |

All solutions were autoclaved before use unless stated otherwise.

### 2.1.6. Liquid media

Recipe for liquid medium were prepared as follows:

**Lysogeny Broth (LB) Medium**

| | |
|---|---|
| Tryptone | 10 g |
| Yeast extract | 5 g |
| NaCl | 10 g |
| $dH_2O$ | to 1 L |

**Supplemented Minimal Medium (SMM)**

| | |
|---|---|
| Casaminoacid | 2 g |
| TES buffer | 8.68 g |
| $dH_2O$ | to 1 L |
| pH | 7.2 |

After autoclave add the following sterilised solutions:

| | |
|---|---|
| $NaH_2PO_4 + K_2HPO_4$ (50 mM each) | 10 mL |
| $MgSO_4$ (1 M) | 5 mL |
| Glucose (50% w/v) | 18 mL |
| Trace element solution | 1 mL |

**2xYT Medium**

| | |
|---|---|
| Tryptone | 16 g |
| Yeast Extract | 10 g |
| NaCl | 5 g |
| $dH_2O$ | to 1L |

**Trace Element Solution**

| | |
|---|---|
| $ZnSO_4.7H_2O$ | 0.1 g |
| $FeSO_4.7H_2O$ | 0.1 g |
| $MnCl_2.4H_2O$ | 0.1 g |
| $CaCl_2.2H_2O$ | 0.1 g |
| NaCl | 0.1 g |
| $dH_2O$ | to 1 L |

Filter sterilised and stored at 4 $^0$C.

**Valanimycin Producing Medium (VPM)**

| | |
|---|---|
| Maltose | 0.2% |
| Peptone | 0.5% |
| Meat Extract | 0.5% |
| Yeast Extract | 0.3% |
| NaCl | 0.3% |
| $MgSO_4.7H_2O$ | 0.1% |
| pH | 7.2 |

**Tryptone Soya Broth (TSB) Medium**

| | |
|---|---|
| Tryptone Soya Broth | 30 g |
| $dH_2O$ | to 1 L |

All solutions were autoclaved before use unless stated otherwise.

## 2.2. Methods

### 2.2.1. Bioinformatics analysis

Genomes of *Streptomyces* bacteria were analysed by AntiSMASH Version 2.0 to determine the number of predicted specialised metabolite gene clusters and help identify putative AHFCA-dependent and GBL-dependent signalling systems. Cryptic

biosynthetic gene clusters were further annotated using Basic Local Alignment Search Tool (BLAST).[132,133]

The TetR repressor binding sites of putative AHFCA-dependent and GBL-dependent signalling systems were discovered using Motif Em for Motif Elicitation (MEME) and Finding Individual Motif Occurances (FIMO) tools from MEME Suite Version 4.9.0 (http://meme-suite.org/).[134] Promoters of the TetR repressors and genes divergently adjacent to the repressor (300 bp upstream of the ORF start codon) were submitted to MEME and the program was allowed to enrich for three motifs. The program was constrained to search for palindromic patterns, and each submitted promoter sequence was searched on the "given strand" and for “zero or one site”. Motifs of interest, which were enriched by MEME, were further submitted to FIMO to search for individual hits to the motif across chromosome or plasmids. Statistical threshold of a *p*-value of less than 1E-4 was used to filter the individual hits.

FIMO results can be viewed in the Electronic Supplementary Material accessible on http://www2.warwick.ac.uk/fac/sci/chemistry/research/corre/corregroup/repositorypoon

### 2.2.2. Growth, storage and manipulation of *E. coli* bacteria

#### 2.2.2.1. Growth of *E. coli* bacteria

*E. coli* was grown overnight in liquid LB media or solid LB agar with appropriate antibiotics at 37 $^{0}$C. For liquid cultures, flasks and/or universal tubes were shaken at 200 rpm.

#### 2.2.2.2. *E. coli* glycerol stocks preparation

*E. coli* glycerol stocks were prepared from overnight *E. coli* cultures and diluted with sterile 50% (w/v) glycerol to a final concentration of 25%. The glycerol stocks were stored at -80 $^{0}$C.

#### 2.2.2.3. Plasmid extraction (GeneJET Plasmid Miniprep Kit)

The Thermo Scientific GeneJET Plasmid Miniprep kit was used for extraction and purification of plasmids from *E. coli*. The protocol was followed per in the manufacturer's instructions with the following changes: [135]

- Plasmids were extracted from 10 mL of overnight *E. coli* culture.
- Prewarmed sterile $dH_2O$ (70 $^{0}$C) was used instead of elution buffer.
- During the elution step, the DNA pellet was allowed to incubate at room temperature for 5 minutes with prewarmed sterile $dH_2O$.

#### 2.2.2.4. Plasmid extraction (phenol/chloroform)

A phenol/chloroform extraction was used for plasmid preparation of pESAC13A constructs. Plasmids were extracted from 10 mL of overnight *E. coli* cultures. The cells were pelleted by centrifugation (2,000 rpm for 10 minutes) and resuspended in 100 µL of solution I (50 mM Tris/HCl and 10 mM EDTA; pH 8). The suspended bacteria was lysed by solution II (200 mM NaOH and 1% SDS) and gently mixed. The proteins were precipitated with 150 µL of solution III (3 M potassium acetate; pH 5.5) and gently mixed. The tube was then centrifuged (13,200 rpm for 5 minutes) to pellet the precipitated proteins. The supernatant was transferred to a fresh tube and 400 µL of pH 8 phenol/chloroform/isoamyl alcohol (25:24:1) was added to extract the DNA. The tubes were mixed gently for 2 minutes and centrifuged (13,200 rpm for 5 minutes) to separate the layers. The upper aqueous

layer was then transferred to a fresh tube and 600 µL of isopropanol was added to precipitate the DNA. The tubes were allowed to incubate on ice for 10 minutes and centrifuged (13,200 rpm for 5 minutes) to pellet the DNA. The pellet was further washed with 200 µL of 70% ethanol and centrifuged (13,200 rpm for 5 minutes). The supernatant was discarded and the DNA pellet was air dryed for 5 minutes at room temperature. The DNA pellet was resuspended in 50 µL of sterile $dH_2O$ and stored at -20 $^0$C.

#### 2.2.2.5. Preparation of electro-competent *E. coli* bacteria

Electro-competent *E. coli* TOP10 or DH5α bacterial cells were prepared by washing mid-exponential phase *E. coli* bacteria ($OD_{600}$ ~0.4 - 0.6) twice with ice cold glycerol (10% w/v). The supernatant was discarded and the pellet was resuspended in residual glycerol. The electro-competent *E. coli* bacteria were aliquotted into cryogenic tubes and stored at -80 $^0$C.

#### 2.2.2.6. Transformation of electro-competent *E. coli* bacteria

Electroporation was performed to introduce plasmids into *E.* coli bacteria and was carried out in 0.2 cm electroporation cuvettes. Electro-competent *E. coli* bacteria (70 µL) and 3 µL of plasmids or ligation mixture were transferred to the ice-cold electroporation cuvette and electroporation was performed at 200 Ω, 25 µF and 2,5 kV. Immediately after electroporation, 1 mL of ice cold LB was added to electroporation cuvette to recover the cells. The transformation mixture was transferred to a fresh tube and incubated at 37 $^0$C for 1 hour at 200 rpm. The mixture was plated onto LB agar with appropriate antibiotics. *E. coli* glycerol stocks were made from positive transformants.

### 2.2.3. Growth, storage and manipulation of *Streptomyces* bacteria

#### 2.2.3.1. Growth of *Streptomyces* bacteria

*Streptomyces* bacteria were grown for 6 days in liquid or solid agar with appropriate antibiotics and incubated at 30 $^{0}$C. For liquid cultures, baffled flasks or flasks containing stainless steel spring were shaken at 220 rpm.

#### 2.2.3.2. *Streptomyces* glycerol spore stock preparation

*Streptomyces* glycerol spore stocks were prepared from a confluent lawn of spores grown on SFM plates with appropriate antibiotics. Sterile $dH_2O$ (3 mL) was added to each plate and the spores were suspended with a sterile loop. The suspension was transferred to a tube and briefly vortexed to liberate the spores from the mycelia. The suspension was filtered through a sterile syringe with non-absorbent wool and pelleted by centrifugation (3,500 rpm for 5 minutes). The spore pellet was resuspended in the residual liquid and diluted with sterile 50% (w/v) glycerol to a final concentration of 25%. The *Streptomyces* glycerol spore stock was stored at -80 $^{0}$C.

#### 2.2.3.3. *Streptomyces* genomic DNA extraction

*Streptomyces* genomic DNA was extracted using the MP Biomedicals FastDNA SPIN Kit for Soil DNA Extraction. The protocol was followed as per the manufacturer's instructions with the following changes: [136]

- Genomic DNA extraction was carried out on *Streptomyces* spores scrapped off a confluent lawn (dimension of 1 x 1 inch square).
- Prewarmed sterile $dH_2O$ (70 $^{0}$C) was used instead of DES solution.
- During the elution step, the DNA pellet was allowed to incubate at room temperature for 5 minutes with prewarmed sterile $dH_2O$.

#### 2.2.3.4. Intergenic conjugation

Bacterial conjugation was performed to transfer plasmids from *E. coli* hosts into *Streptomyces* bacteria. *E. coli* hosts were grown to mid-exponential phase ($OD_{600}$ ~0.4 - 0.6) and centrifuged (2,000 rpm for 10 minutes) to pellet the cells. The cells were washed twice with fresh LB to remove the antibiotics and the pellet was resuspended in 500 µL of LB media. *Streptomyces* spores were germinated by incubating 10 µL of *Streptomyces* spore glycerol stock in 500 µL of 2xYT media at 50 $^0$C for 10 minutes.

*E. coli* and *Streptomyces* bacteria were combined and briefly centrifuged (5,500 rpm for 30 seconds). The supernatant was discarded, and resuspended in 500 µL of sterile $dH_2O$ and serially diluted. Serially diluted conjugation mixture (100 µL) were then plated onto SFM agar + 10 mM $MgCl_2$ and grown overnight at 30 $^0$C. After 16-20 hours, the plates were overlayed with 1 mL of sterile $dH_2O$ with 25 µL of apramycin (50 mg/mL), 20 µL of nalidixic acid (50 mg/mL) and 25 µL of hygromycin (50 mg/mL); for *S. lividans* TK24, 100 µL of hygromycin (50 mg/mL) was used. The plates were further incubated at 30 $^0$C until spores were observed. Positive *Streptomyces* colonies were subcultured onto a fresh SFM agar plate with appropriate antibiotics and *Streptomyces* spore glycerol stocks were made from positive exconjugants.

## 2.3. Molecular biology

### 2.3.1. Polymerase chain reaction

For accurate cloning of genes, New England Biolabs Phusion High-Fidelity DNA polymerase was used. For routine polymerase chain reactions (PCR), Fermentas

Routine PCR taq polymerase (recombinant) was used. The reaction mixture was prepared as follows:

| **Reagent for routine** | **Volume** |
|---|---|
| 10X Taq Buffer with $(NH_4)_2SO_4$ | 5 µL |
| $MgCl_2$ (10 mM) | 3 µL |
| DMSO | 2.5 µL |
| dNTP (10 mM) | 1 µL |
| Primers (100 mM) | 1 µL each |
| DNA template (50 - 200 ng) | 0.5 - 1 µL |
| Fermentas Taq polymerase | 1 unit |
| Sterile $dH_2O$ | to 50 µL |

| **Reagent for high fidelity** | **Volume** |
|---|---|
| 5X Phusion GC buffer | 10 µL |
| DMSO | 2.5 µL |
| dNTP (10 mM) | 1 µL |
| Primers (100 mM) | 1 µL each |
| DNA template (50 - 200 ng) | 0.5 - 1 µL |
| Phusion Taq polymerase | 1 unit |
| Sterile $dH_2O$ | to 50 µL |

For all routine PCR reactions, the following PCR cycle was used:

Step 1: 98 $^0$C for 2 minutes
Step 2: 98 $^0$C for 2 minutes
Step 3: 55 $^0$C for 45 seconds
Step 4: 72 $^0$C for X minutes
Step 5: 72 $^0$C for 15 minutes
Step 6: 4 $^0$C hold

Step 2 to Step 4 were repeated for 30 times and the elongation step (Step 4) was incubated for 1 minute/ 1 kb. For high fidelity PCR reactions, a two step PCR cycle was used (the same PCR cycle as above with Step 3 omitted).

### 2.3.2. Restriction enzyme digestion

Plasmids and/or PCR products were digested using Fermentas FastDigest restriction enzymes. The following reaction mixture was prepared and incubated for 30 - 60 minutes at 37 $^0$C. After incubation, digested products were immediately

analysed by agarose gel electrophoresis and purified using the GeneJet Gel Extraction kit.

| Reagent | Volume |
|---|---|
| 10X FastDigest buffer | 5 µL |
| DNA | 10 - 30 µL |
| FastDigest enzyme | 1.5 µL |
| Sterile $dH_2O$ | to 50 µL |

### 2.3.3. DNA ligation

Plasmids and/or PCR products were ligated together using Invitrogen T4 DNA ligase. An insert to vector molar concentration ratio of 3:1 was used for all DNA ligation reactions. The following reaction mixture was prepared, incubated for 5 minutes at room temperature and transferred to 4 $^0$C overnight before electroporation.

| Reagent | Volume |
|---|---|
| 5X DNA ligase reaction buffer | 4 µL |
| DNA (3:1 ratio) | X µL |
| T4 DNA ligase | 1 unit |
| Sterile $dH_2O$ | to 20 µL |

### 2.3.4. Agarose gel electrophoresis

A 1% agarose gel was prepared by dissolving 1 g of agarose powder in 100 mL of TAE buffer (40 mM Tris, 20 mM acetic acid and 1 mM EDTA) and 5 µL of GelRed was added to the agarose gel. For each sample, 5 µL of DNA was combined with 1 µL of Fermentas 6X DNA gel loading dye and loaded into the well. Fermentas FastRuler Low Range Medium Range, High Range DNA ladder and GeneRuler 1 kb DNA ladder were used as DNA markers (5 µL). Electrophoresis was carried out at 100 V for 40 minutes.

### 2.3.5. Purification of DNA fragments from agarose gel

PCR products and/or digested DNA fragments from agarose gel electrophoresis were excised and purified using the Thermo Scientific GeneJET Gel Extraction kit. The protocol was followed as per the manufacturer's instructions with the following changes: [137]

- Prewarmed sterile $dH_2O$ (70 $^0C$) was used instead of Elution Buffer.
- During the elution step, the DNA pellet was allowed to incubate at room temperature for 5 minutes with prewarmed sterile $dH_2O$.

## 2.4. Analytical chemistry

### 2.4.1. Organic extraction

*Streptomyces* bacteria were grown in liquid (50 mL) or solid agar (25 mL) for 6 days. Culture or media were separated into two tubes for the acidified (pH 3) and neutral (pH 7) organic extraction. An equal volume of ethyl acetate was added to the culture and the pH of ethyl acetate was adjusted with 37% HCl. The upper ethyl acetate layer was evaporated *in vacuo* and residual was dissolved with HPLC grade methanol and water.

For liquid organic extraction, the culture was centrifuged (4,000 rpm for 20 minutes) to remove the bacteria and supernatant was extracted with ethyl acetate. The upper ethyl acetate layer was also treated with excess of magnesium sulphate to remove water and filtered. The ethyl acetate layer was evaporated *in vacuo* as described above.

### 2.4.2. Liquid chromatography – mass spectrometry

LC-MS analysis was performed on the Agilent 1200 HPLC coupled to the Bruker High Capacity Trap (HCT) Ultra mass spectrometer operated in positive mode electrospray ionisation (ESI). Organic extracts were filtered on a spin column (4,500 rpm for 5 minutes) and 20 µL of the filtered extract was injected into the Agilent ZORBAX Eclipse Plus column (C18, 46 x 150 mm, 5 µm). The following solvent and linear gradient profile with a flow rate of 1 mL/minute was used for all LC-MS analysis (Table 2-5).

**Table 2-5** – Gradient profile for LC-MS analysis.

| Time (minutes) | Water + 0.1% formic acid (%) | Methanol + 0.1% formic acid (%) |
|---|---|---|
| 0 | 95 | 5 |
| 5 | 95 | 5 |
| 30 | 0 | 100 |
| 35 | 0 | 100 |
| 40 | 95 | 5 |
| 55 | 95 | 5 |

High resolution LC-MS analysis was performed on the Bruker MaXis or MaXis Impact ESI-Quadrupole-Time of Flight (TOF) mass spectrometry. The Mass spectrometer was connected to the Dionex 3000RS ultra high performance liquid chromatography (UHPLC) system and an Agilent ZORBAX Eclipse Plus column (C18, 2.1 x 100 mm, 5 µm) was used. The following solvent and linear gradient profile with flow rate of 0.2 mL/minute was used for all UHPLC-ESI-TOF-MS analysis (Table 2-6).

**Table 2-6** – Gradient profile for UHPLC-ESI-TOF-MS analysis.

| Time (minutes) | Water + 0.1% formic acid (%) | Acetonitrile + 0.1% formic acid (%) |
|---|---|---|
| 0 | 95 | 5 |
| 5 | 95 | 5 |
| 20 | 0 | 100 |
| 35 | 0 | 100 |

## 2.5. Experimental methods

### 2.5.1. *S. avermitilis* pESAC13A PAC genomic library

*S. avermitilis* pESAC13A P1-derived artificial chromosome (PAC) genomic library was created by Bio S&T (www.biost.com). The following samples were prepared for the *S. avermitilis* pESAC13A PAC genomic library:

- Three tubes containing *S. avermitilis* bacteria.
- Genomic DNA of *S. avermitilis* (minimum of 10 µg).
- Three pairs of primers for screening gene cluster of interest (for_pESAC_SAV2283a and rev_pESAC_SAV2283a; for_pESAC13_sav2266a and rev_pESAC13_sav2266a; for_pESAC_sav2301a and rev_pESAC_sav2301a, see Table 2-3).

#### 2.5.1.1. Preparation of *S. avermitilis* bacterial cells

*S. avermitilis* glycerol spore stock (10 µL) was inoculated in 100 mL of TSB media and cultured for 30 hours. The cells were pelleted by centrifugation (4,000 rpm for 10 minutes at 4 $^{0}$C) and supernatant was discarded. The pellets were then washed twice with 50 mL of buffer (200 mM NaCl, 10 mM Tris-HCl, 100 mM EDTA, pH 8) and flash frozen in liquid nitrogen for 5 minutes and stored at -80 $^{0}$C before shipping.

#### 2.5.1.2. Isolation of *S. avermitilis* genomic DNA

*S. avermitilis* was cultured in TSB media as above and bacterial cells from 25 mL of the culture was pelleted by centrifugation (4,000 rpm for 10 minutes). The cells were washed twice with STE buffer (100 mM NaCl, 10 mM Tris-HCl, 1 mM EDTA, pH 8). The bacterial pellet was resuspended in 5 mL of STE buffer and lysed with lysozyme dissolved in 5 mL of STE buffer (10 mg/mL) for 30 minutes on a gel rocker at room temperature. SDS (10%) was added and further incubated at room temperature for 5 minutes on the gel rocker.

Both SDS and lysozyme were inactivated by incubating the tubes at 70 $^{0}$C for 5 minutes and cooled down on ice for 5 minutes. The sample was then neutralised by addition of 1.25 mL of 5 M potassium acetate and further incubated on ice for 5 minutes. Genomic DNA was purified twice using 5 mL of pH 8 phenol/chloroform/isoamyl alcohol (25:24:1) and mixed gently for 5 minutes. The layers were separated by centrifugation (4,000 rpm for 10 minutes at 4 $^{0}$C). The upper aqueous layer was transferred to a fresh tube and 1 mL of 3 M sodium acetate and 6 mL of isopropanol were added to precipitate the DNA. The precipitated DNA was pelleted by centrifugation (13,200 rpm for 5 minutes) and washed twice with 70% ethanol. The supernatant was discarded and genomic DNA pellet was then air dried for 5 minutes at room temperature. The DNA pellet was resuspended in 500 μL of sterile $dH_2O$ and stored at -80 $^{0}$C before shipping.

### 2.5.2. Heterologous expression of the predicted azoxy gene cluster in *S. coelicolor* M1152 and *S. lividans* TK24

The two pESAC13A-2H and pESAC13A-19K constructs were introduced into *S. coelicolor* M1152 and *S. lividans* TK24 heterologous hosts by tri-parental conjugation and analysed by LC-MS.

#### 2.5.2.1. Tri-parental conjugation in *S. lividans* TK24

Tri-parental conjugation was carried out between *E. coli* DH5α/pUB307, *S. lividans* TK24 and *E. coli* DH10B containing the pESAC13A constructs. *E. coli* strains were grown till mid exponential phase ($OD_{600}$ ~0.4 - 0.6) and washed twice with LB to remove antibiotics. *S. lividans* spores were germinated by inoculating 10 µL of glycerol spore stock in 500 µL of 2xYT media and incubated at 55 $^0$C for 15 minutes. A conjugation mixture was prepared by combining 500 µL of *S. lividans* TK24, *E. coli* DH5α/pUB307 and *E. coli* DH10B containing the pESAC13A constructs and plated on SFM + 10 mM $MgCl_2$. Positive exconjugants were selected with apramycin and nalidixic acid.

#### 2.5.2.2. Tri-parental conjugation in *S. coelicolor* M1152

Tri-parental conjugation was carried out between *E. coli* DH5α/pR9604, *E. coli* ET12567 and *E. coli* DH10B containing the pESAC13A constructs. *E. coli* strains were prepared as above and resuspended in 500 µL of LB media. Tri-parental conjugation was performed by spotting 10 µL of each *E. coli* bacteria on the same location on LB agar. The plates were grown overnight at 37 $^0$C and streaked for single colonies on LB agar with apramycin, ampicillin and chloramphenicol.

Intergenic conjugation was carried out as described above between *S. coelicolor* M1152 and *E. coli* ET12567/pR9604 strains containing the pESAC13A constructs. Positive exconjugants were selected with apramycin and nalidixic acid.

#### 2.5.2.3. Comparative metabolic profiling of *Streptomyces* heterologous hosts

*Streptomyces* glycerol spore stocks (10 µL) were inoculated onto 25 mL of SMMS agar or 50 mL of liquid SMM or VPM and incubated at 30 $^{0}$C for 6 days. Organic extraction was performed with ethyl acetate and extracts were analysed by LC-MS. The pH of the extractions was adjusted with 37% HCl.

### 2.5.3. Genetic inactivation of *sav_2268* repressor

The entire *sav_2268* repressor was deleted in *S. lividans* heterologous host strains by double homologous recombination. Intergenic conjugation was performed between *E. coli* ET12567/pUZ8002/pER2Δint and *S. coelicolor* M1152 or *S. lividans* TK24 heterologous host strains containing the pESAC13A constructs; exconjugants were selected with apramycin, nalidixic acid and hygromycin. Positive exconjugants were passaged on SFM agar containing apramycin and nalidixic acid. After each passage, single colonies were streaked on SFM agar containing apramycin and nalidixic acid and screened for hygromycin sensitivity.

### 2.5.4. Characterisation of predicted metabolites

The predicted metabolites were characterised by the incorporation of isotope precursors, structure determination by NMR spectroscopy and antimicrobial assays.

#### 2.5.4.1. Feeding of isotope labelled amino acids

Isotope-labelled L-serine-$^{13}C_3$-$^{15}N$ and L-threonine-$^{13}C_4$-$^{15}N$ were dissolved in sterile $dH_2O$ to a final concentration of 25 mM, filtered sterilised and stored at 4 $^0$C. *Streptomyces* heterologous host strains (10 µL) were inoculated onto 5 mL SMMS agar and grown for 6 days at 30 $^0$C. After the 1st day of growth, the *Streptomyces* bacteria were overlayed with 200 µL of 25 mM isotope-labelled L-serine or L-threonine for the next 4 days; sterile $dH_2O$ was used as a negative control. After feeding, the metabolites were extracted with ethyl acetate and analysed by LC-MS.

#### 2.5.4.2. Isolation of metabolites for NMR spectroscopy

*S. coelicolor* M1152/pESAC13A-2H1 glycerol spore stocks (10 µL) were inoculated on 30 x 25 mL of SMMS agar plates and incubated for 6 days at 30 $^0$C. Organic extraction was performed on the agar with acidified ethyl acetate and treated with excess of magnesium sulphate to remove water. The ethyl acetate solvent was evaporated *in vacuo* and residual was dissolved in HPLC grade methanol and water.

The scaled up crude extract was filtered on a spin column (4,500 rpm for 5 minutes) and injected into the Agilent ZORBAX Eclipse column (C18, 21.2 x 150 mm, 5 µm) connected to the Agilent 1200 HPLC system. Purification and fraction collection was guided by UV absorbance at 230 nm. The same solvents and gradients were used as in LC-MS analysis with a flow rate of 20 mL/minute (Table 2-5).

Fractions containing the compounds of interest were confirmed by the Agilent 6130B single Quadrupole LC/MS system. The fractions were pooled together, and the methanol solvent was evaporated *in vacuo* and residual water was removed by sublimation. The purified compounds were dissolved in deuterated chloroform and

analysed by a combination of 1- and 2-D NMR spectroscopy ($^{1}H$, $^{13}C$, COSY, HSQC and HMBC) on the Bruker Avance II 700 Mhz.

#### 2.5.4.3. Antimicrobial assays

*E. coli*, *B. subtilis* and *S. aureus* were grown to mid-exponential phase ($OD_{600}$ ~0.4 - 0.6). After growth, 50 µL of the culture was diluted in 2 mL of fresh LB media. The diluted culture was overlayed on 50 mL of acidified (pH 5) LB agar and allowed to fully dry. The crude organic extract (20 µL) was added to 6 mm filter paper discs and allowed to dry. Apramycin (5 µL) and methanol/water (20 µL) were used as positive and negative controls respectively. The filter papers were carefully placed on top of the overlayed LB agar and the plates were incubated overnight at 37 $^{0}C$.

Antimicrobial assays against *Streptomyces* bacteria were performed as above. For the preparation of *Streptomyces* bacteria, *Streptomyces* glycerol spore stock (10 µL) were inoculated in 2 mL of sterile $dH_2O$ and overlayed on 50 mL of SFM agar. The filter paper discs were carefully placed on the dried SFM agar plates and were incubated for 2 days at 30 $^{0}C$.

## 3. AHFCA-dependent Signalling Systems

From genome mining analyses, the cassette of 5 genes that form the AHFCA-dependent signalling system in *S. coelicolor* have been identified in several other *Streptomyces* bacteria, such as in *S. venezuelae* ATCC 10712 and *S. avermitilis* MA-4680 (Figure 3-1). [56,62] The gene organisation of the putative AHFCA-dependent signalling systems appear to be highly conserved across the *Streptomyces* species (Figure 3-1).

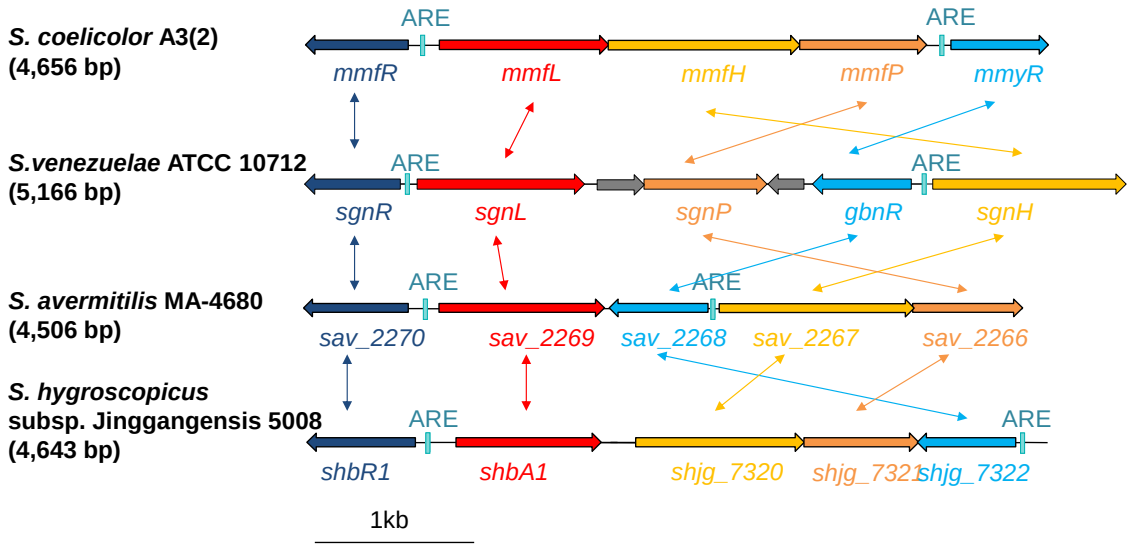



**Figure 3-1** – Analogous AHFCA-dependent signalling systems from *Streptomyces* species. Analogous genes are colour-coded.

In the *S. venezuelae* AHFCA signalling gene cluster, there are two additional genes that are not present in the other systems (Figure 3-1); *sven_4184* and *sven_4186*, which encode for a putative LysR transcriptional regulator and a hypothetical protein respectively. The members within the LysR family of transcriptional regulators are approximately 330 amino acids long and are also known to bind to dyad sequences as a homodimer. [138] The *sven_4184* gene product is only 98 amino

acids long and appears to form a truncated LysR regulator; specifically the N-terminal DNA-binding domain. The *sven_4184* gene product might not encode for a fully functional LysR-like transcriptional regulator, but the DNA-binding domain may still actively interact with DNA sequences. The *sven_4186* gene product encodes for a hypothetical protein and other homologous partners could not be identified.

In *S. coelicolor* and *S. venezuelae*, the AHFCA signalling gene cluster is adjacent to the methylenomycin and gaburedin biosynthetic gene cluster respectively. [57,62] Since the neighbouring biosynthetic gene clusters are all different and organisation of the signalling gene clusters are highly conserved, this implies the biosynthetic and signalling systems have not co-evolved and the signalling gene clusters were most likely transferred horizontally across *Streptomyces* bacteria. AHFCA-dependent signalling systems are predicted to tightly regulate expression of neighbouring cryptic biosynthetic gene clusters and preliminary bioinformatic analyses have identified highly conserved semi-palindromic DNA sequences in the promoters of key regulatory genes (Figure 1-9 and Figure 1-19); TetR repressors are predicted to bind to these sequences and regulate expression of biosynthetic genes. [57,62]

The putative AHFCA-dependent signalling systems discovered in *S. avermitilis* MA-4680 and *S. hygroscopicus* subsp. Jinggangensis 5008 (Figure 3-1) are also proposed to behave analogously and directly regulate neighbouring cryptic biosynthetic gene clusters. To our knowledge, the putative AHFCA-dependent signalling systems, as well as the adjacent cryptic biosynthetic gene clusters in *S. avermitilis* and *S. hygroscopicus* have not been investigated; analyses and exploitation of these complex regulatory systems may result in the discovery of

novel drug-like compounds. The first approach to decipher the complex AHFCA-dependent regulatory mechanisms is to predict TetR repressor binding sites; this would provide strategic targets to genetically manipulate cryptic biosynthetic gene clusters aimed at discovering novel natural compounds.

## 3.1. MEME enrichment analysis of putative AHFCA-dependent signalling systems

The AHFCA-dependent signalling system in *S. coelicolor* contains two TetR repressors named MmfR and MmyR (Figure 3-1). [57] TetR repressors, which are known to bind to semi-palindromic sequences, regulates the expression of itself and divergently adjacent genes. [89] Therefore, semi-palindromic DNA sequences are expected to be present in both of the TetR promoters (*mmfR* and *mmyR*). The MEME suite (http://meme-suite.org/), which contains a series of motif-based sequence analysis tools, was used to analyse TetR repressor binding sites. Firstly, statistically significant and over-represented *de novo* motifs were discovered by the MEME enrichment tool, which uses an expectation maximisation algorithm. [134]

To identify MmyR and MmfR DNA-binding sequences, an initial MEME analysis was performed on the promoters of TetR repressors as well as genes that are divergently adjacent to the repressors (300 bp upstream of start codon of ORFs). The MEME program was constrained to "search given strand only" and to "look for palindromes only". The initial MEME analysis outputs three motifs (Figure 3-2); motifs with E-value of less than 1 or motifs that shared characteristic features of previously identified ARE sequences were further investigated (Figure 1-9, Figure 1-18 and Figure 1-19).

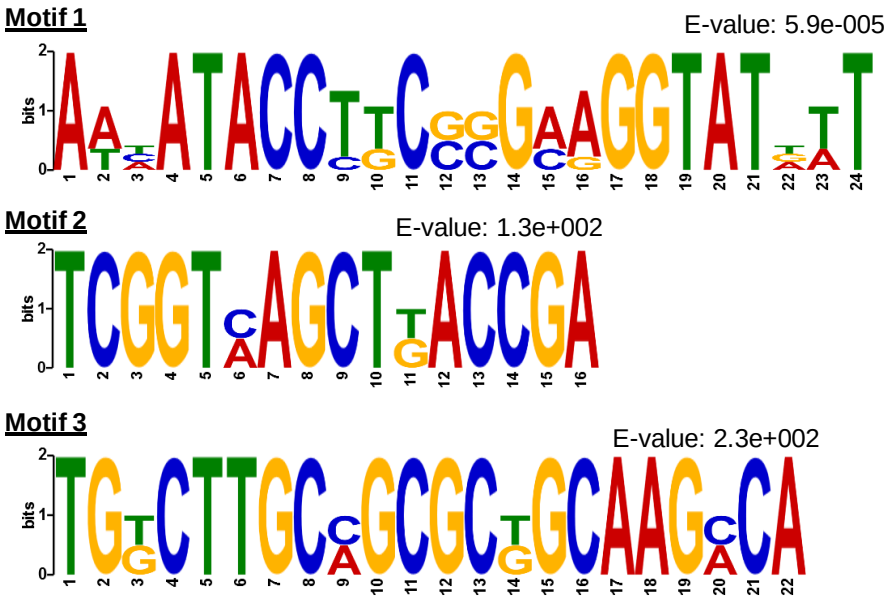



**Figure 3-2** – Output from the initial MEME enrichment analysis of methylenomycin biosynthetic gene cluster. Motif 1 was further investigated as a potential MmyR/MmfR binding sequence.

Initial MEME enrichment analysis of the *S. coelicolor* AHFCA-dependent signalling system identified a highly conserved 24 bp semi-palindromic motif (Figure 3-2) and shares resemblance to ARE sequences previously discovered (Figure 1-19).[57] Initial MEME enrichment analysis was also successful in revealing ARE sequences in the putative AHFCA-dependent signalling systems from *S. venezuelae*, *S. avermitilis* and *S. hygroscopicus* (Figure 3-1, Figure 3-3 and Appendix A).

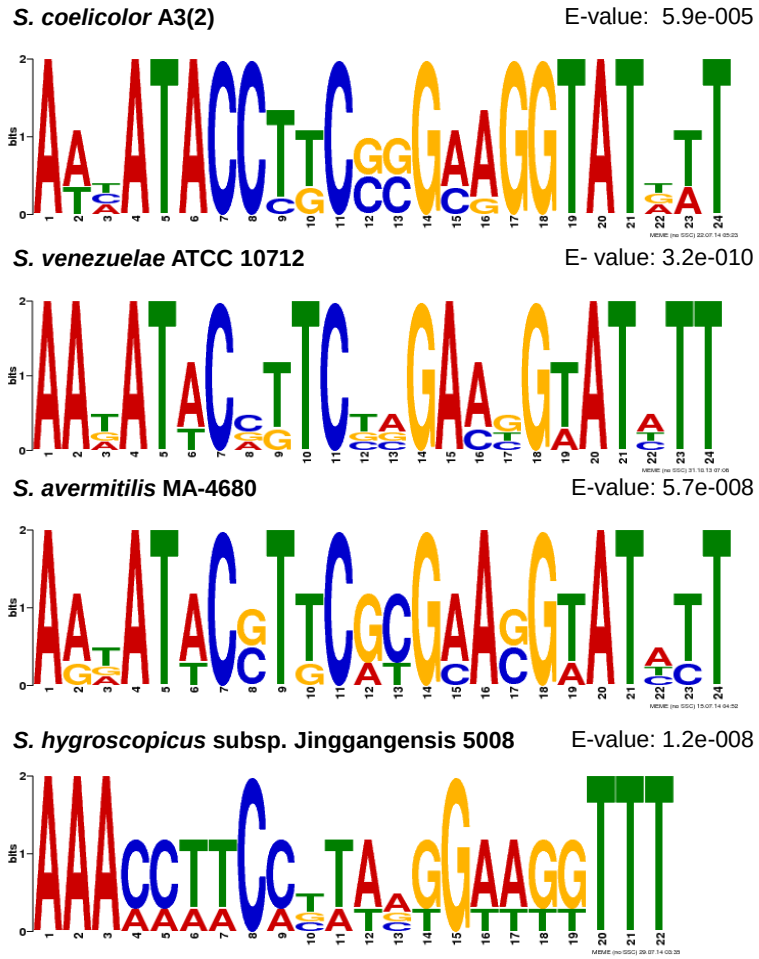



**Figure 3-3** – Initial MEME enrichment analysis identified highly conserved ARE motifs in *S. coelicolor, S. venezuelae*, *S. avermitilis* and *S. hygroscopicus* putative AHFCA-dependent signalling systems. Analogous MmyR and MmfR repressors are proposed to bind to these sequences.

## 3.2. Regulation of neighbouring cryptic biosynthetic gene clusters by AHFCA-dependent signalling systems

Additional TetR binding sites were searched across the whole genome using FIMO. The initial MEME motifs from Figure 3-3 were submitted to FIMO and allowed to scan through *Streptomyces* chromosomes and plasmids. The FIMO program parameters were set as default with a statistical threshold ($p$-value < 1E-4) and were allowed to search on the reverse complement of nucleotide sequences.

The pathway-specific MmyR and MmfR repressors in *S. coelicolor* are proposed to regulate the expression of AHFCA signalling genes and neighbouring *Mm* gene cluster; however, FIMO analyses revealed numerous ARE-like sequences across the whole genome and the TetR repressors may interact with any of these specific sites (Table 3-1 and Electronic Supplementary Material accessible on http://www2.warwick.ac.uk/fac/sci/chemistry/research/corre/corregroup/repositorypoon). [57] In *S. coelicolor* A3(2), there are a total of 151 genes predicted to encode for TetR repressors, which includes 11 MmfR/MmyR paralogs, and would most likely bind to similar semi-palindromic sequences; the FIMO analysis may have identified these "false-positive" sites. [66]

Nonetheless, MmyR and MmfR repressors may bind to these specific ARE-like sequences to regulate expression of other genes outside of the *Mm* gene cluster. Cross-talks between different specialised metabolite gene clusters have been previously reported in *S. coelicolor*. For instance, the ScbR2 repressor in the SCB/ScbR/ScbR2 system could bind to endogenous actinorhodin and undecylprodigiosin metabolites, as well as regulate *act* and *red* biosynthetic genes by directly controlling the expression of *actII-orf4* (*sco5085*), *redD* (*sco5877*) and *redZ* (*sco5881*) and indirect regulation *via adpA* (*sco2792*). [99,120,139,140]

Detailed analysis of FIMO results identified ARE-like sequences in other cryptic biosynthetic gene clusters; for example, individual hits in *S. coelicolor* were located in actinorhodin (Motif 4), coelimycin (Motif 11 and 19), calcium-dependent antibiotic (Motif 15), siderophore (Motif 30) and tetrahydroxynaphthalene (Motif 53) biosynthetic gene clusters (Table 3-2 and Electronic Supplementary

Material). The binding of the TetR repressors from AHFCA-dependent gene clusters were demonstrated to bind across the *Streptomyces* genome by Tan and co-workers. ShbR1 repressor (SHJG_7318), from the *S. hygroscopicus* putative AHFCA-dependent signalling system (Figure 3-1), could directly bind and repress the expression of *adpA* (*shjg_4295*) and was reported to indirectly affected validamycin production. [141,142] The AHFCA-dependent signalling system in *S. coelicolor* may cross-talk with other specialised metabolite biosynthetic gene clusters to coordinate the production of specific natural compounds. The binding of analogous MmyR and MmfR repressors to proposed ARE-like sequences needs to be further confirmed; chromatin immunoprecipitate-sequencing (ChIP-Seq) and transcriptome analysis could be carried out to determine global roles and effects of the TetR repressors.

**Table 3-1** – Analogous MmyR/MmfR binding sites across the genome.

| **Name** | **Number of unique sites** |
|---|---|
| *Streptomyces coelicolor* | |
| Chromosome | 65 |
| SCP1 | 5 |
| SCP2 | 0 |
| *Streptomyces venezuelae* | |
| Chromosome | 22 |
| *Streptomyces avermitilis* | |
| Chromosome | 123 |
| SAP1 | 0 |
| *Streptomyces hygroscopicus* | |
| Chromosome | 22 |
| pSHJG1 | 0 |
| pSHJG2 | 0 |

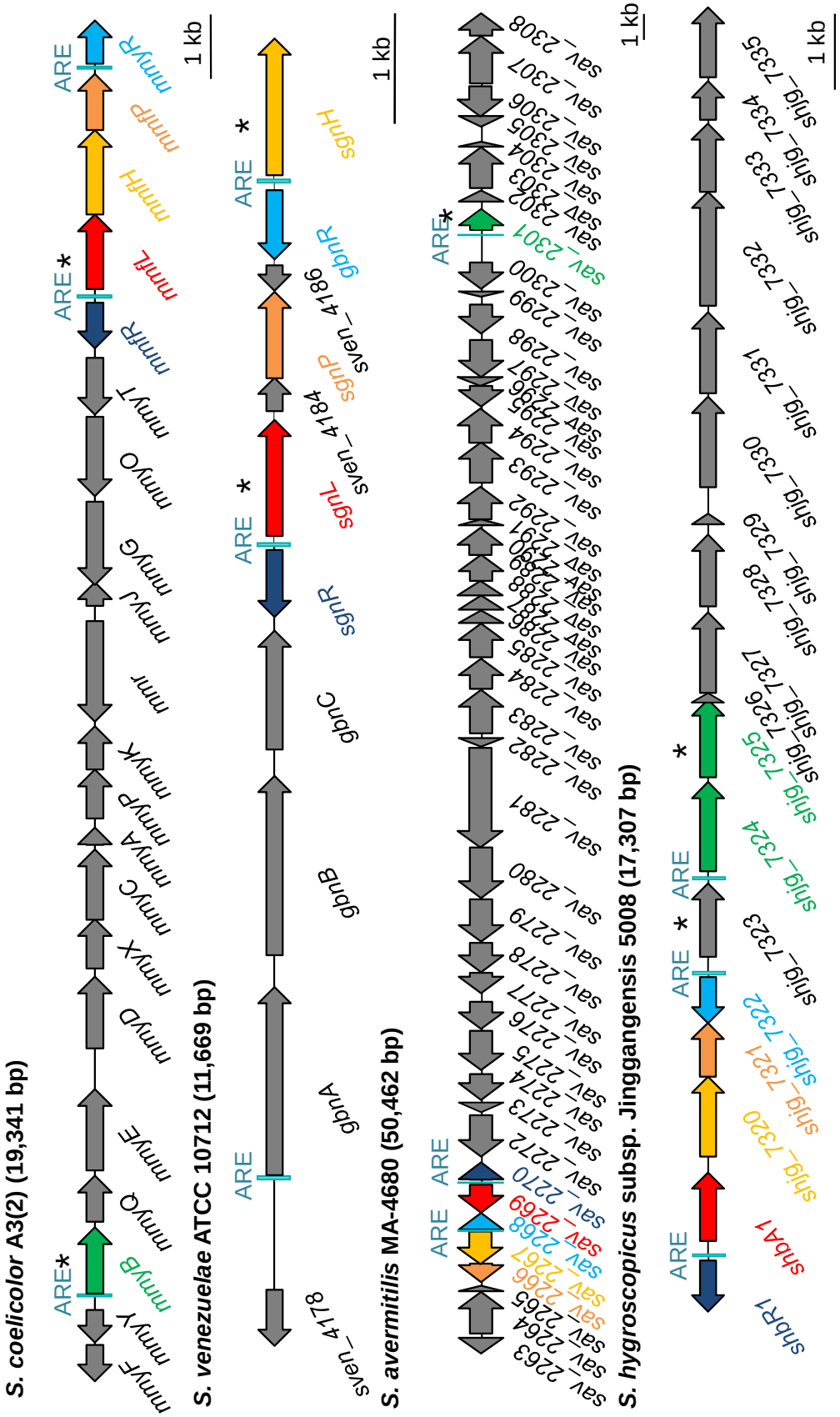
More importantly, FIMO analyses discovered ARE sequences nearby AHFCA signalling gene clusters and were located in the promoters of putative

transcriptional activators and biosynthetic genes (Figure 3-4). The analogous MmyR/MmfR repressors are proposed to regulate putative transcriptional activator to control expression of neighbouring cryptic biosynthetic gene clusters.

**Table 3-2** – MmyR/MmfR binding sites in other specialised metabolite biosynthetic gene clusters in *S. coelicolor.*

| Motif | Start | End | Matched Sequence | Gene downstream of ARE sequence | Proposed Function |
|---|---|---|---|---|---|
| 4 | 5532238 | 5532261 | GATGTCCCTGCGCGAAGGGGTGAT | *sco5089* | Actinorhodin polyketide synthase ACP |
| 11 | 6894127 | 6894150 | TGAAAACCGGCGAGGAGGTTTGTT | *sco6268* | Histidine kinase |
| 15 | 3606273 | 3606296 | GACATGCCTGCCCGCGACTTCTTC | *sco3255* | SpdB protein |
| 19 | 6937942 | 6937965 | TACAAACCAGCACACCGGTTTGTC | *sco6280* | SARP family transcriptional regulator |
| 30 | 6357149 | 6357172 | GACGGTCCCGCTCGCAGGTATGTG | *sco5813* | Hypothetical protein |
| 53 | 1272674 | 1272697 | AAGGACCCGTCCGGGAGGTACGAC | *sco1201* | Reductase |

**Figure 3-4** – (Next page) Specialised metabolite biosynthetic gene clusters are adjacent to AHFCA signalling gene clusters (refer to Appendix D, Appendix E, Appendix F and Appendix G respectively). MEME enrichment analysis reveals ARE sequences are located in the promoters of key regulatory genes and biosynthetic genes. Analogous genes are colour-coded. Asterisk indicate presence of rare TTA codon.


S. coelicolor A3(2) (19,341 bp)
S. venezuelae ATCC 10712 (11,669 bp)
S. avermitilis MA-4680 (50,462 bp)
S. hygroscopicus subsp. Jinggangensis 5008 (17,307 bp)
1 kb
ARE
*
mmyF
mmyY
mmyB
mmyQ
mmyE
mmyD
mmyX
mmyC
mmyA
mmyP
mmyK
mmr
mmyJ
mmyG
mmyO
mmyT
mmyR
mmfL
mmfH
mmfP
mmyR
sven_4178
gbnA
gbnB
gbnC
sgnR
sgnL
sgnP
gbnR
sgnH
sav_2263
sav_2264
sav_2265
sav_2266
sav_2267
sav_2268
sav_2269
sav_2270
sav_2271
sav_2272
sav_2273
sav_2274
sav_2275
sav_2276
sav_2277
sav_2278
sav_2279
sav_2280
sav_2281
sav_2282
sav_2283
sav_2284
sav_2285
sav_2286
sav_2287
sav_2288
sav_2289
sav_2290
sav_2291
sav_2292
sav_2293
sav_2294
sav_2295
sav_2296
sav_2297
sav_2298
sav_2299
sav_2300
sav_2301
sav_2302
sav_2303
sav_2304
sav_2305
sav_2306
sav_2307
sav_2308
shbR1
shbA1
shig_1320
shig_1321
shig_1322
shig_1323
shig_1324
shig_1325
shig_1326
shig_1327
shig_1328
shig_1329
shig_1330
shig_1331
shig_1332
shig_1333
shig_1334
shig_1335

## 3.3. Investigations into the location of ARE sequences

In the four putative AHFCA-dependent signalling systems, organisation of the genes are highly conserved and locations of ARE sequences are consistently found in promoters of the two TetR repressor genes and *mmfL*-like gene (Figure 3-4). The locations of ARE sequences compared to the promoter elements, which were predicted using BPROM, were investigated to determine whether binding of TetR repressor would regulate gene expression. [143]

In the analogous *mmfR* - *mmfL* intergenic region, the ARE sequences were either located downstream or overlapping the -10 to -35 boxes of the *mmfL*-like and *mmfR*-like genes; hence, binding of the TetR repressors to the ARE sequence is predicted to repress expressions of both divergently adjacent genes (Figure 3-5).

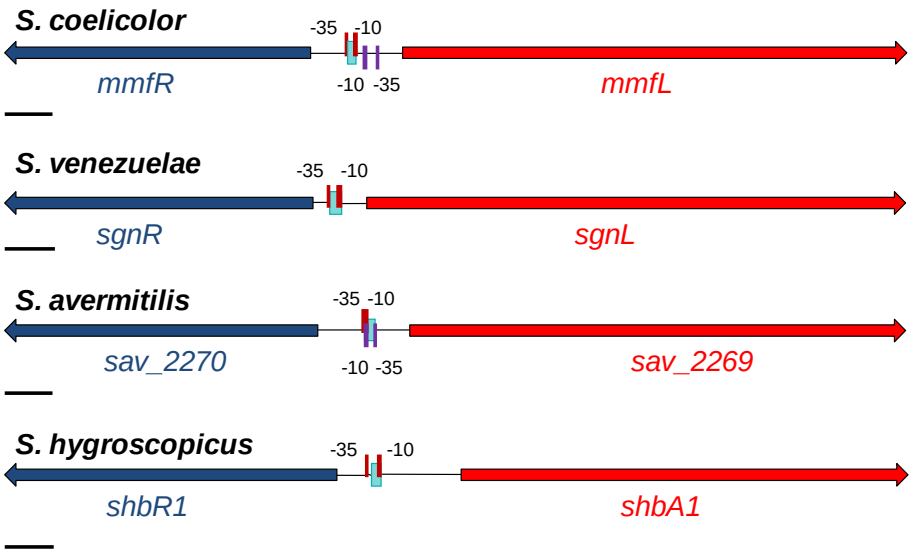



**Figure 3-5** – ARE sequences are downstream/ overlapping promoter elements of analogous *mmyR* and *mmfL* genes. The -10 and -35 boxes were predicted using BPROM (the promoter elements for *sgnR* and *shbR1* could not be predicted). Black bars represent 100 bp in length and ARE sequences are represented as a turquoise box.

Unlike the analogous *mmfR* - *mmfL* intergenic region, genes that are divergently adjacent to the *mmyR*-like repressors are not as well conserved. In *S. venezuelae* and *S. avermitilis*, the *mmyR*-like gene is located divergently adjacent to an *mmfH*-like gene, whilst in *S. coelicolor* there are no genes divergent to the *mmyR* gene and in *S. hygroscopicus*, the *mmyR*-like gene is divergently adjacent to *shjg_7323*, which encodes for a putative monooxygenase (Figure 3-6). The ARE sequences are overlapping the promoter elements of analogous *mmyR* repressor and *mmfH* genes and binding of the repressor is expected to regulate expressions of divergently adjacent genes (Figure 3-6).

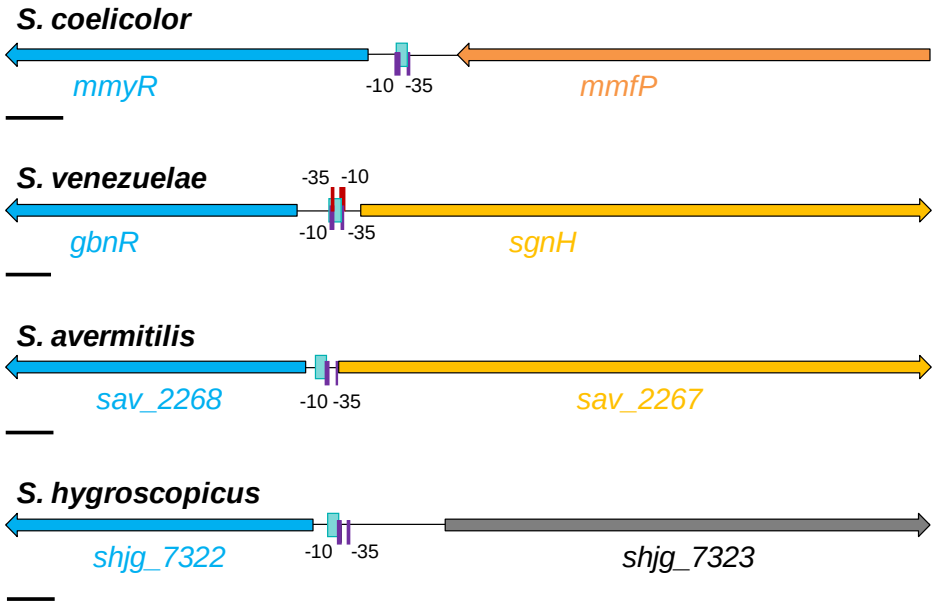



**Figure 3-6** – ARE sequences are downstream/ overlapping promoter elements of analogous *mmfR* gene. The -10 and -35 boxes were predicted using BPROM (promoter elements for *shjg_7323* and *sav_2267* could not be predicted). Black bars represent 100 bp in length and ARE sequences are represented as a turquoise box.

FIMO analyses revealed a third ARE sequence, which was not located in the putative AHFCA signalling gene clusters, but was found in the promoters of genes that encode for proteins with different functions (Figure 3-7 and Table 3-3).

In *S. coelicolor*, *S. avermitilis* and *S. hygroscopicus* systems, the third ARE sequence is located upstream of a putative transcriptional activator (Figure 3-7 and Table 3-3), whilst in *S. venezuelae*, the third ARE sequence is found upstream of *gbnA*, which is proposed to be involved in the biosynthesis of gaburedins.[62]

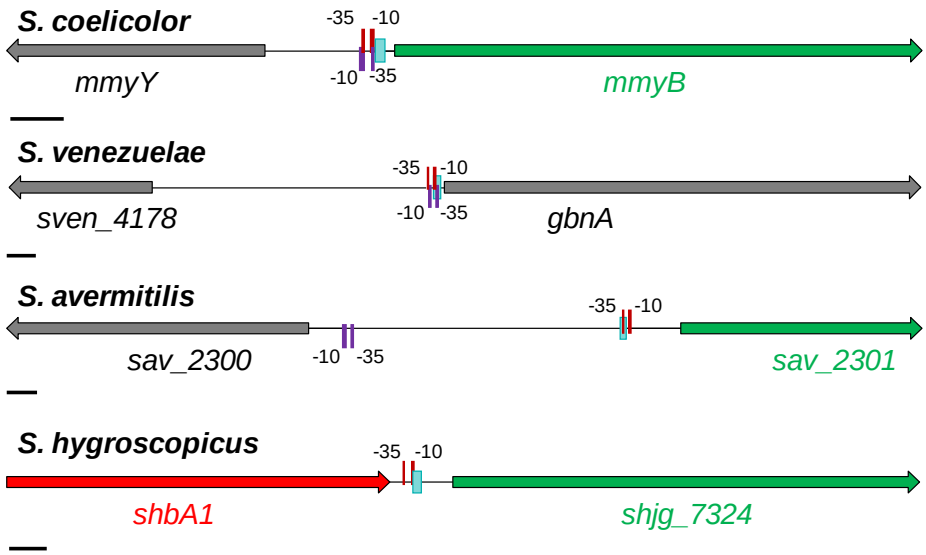



**Figure 3-7** – ARE sequences are downstream/ overlapping promoter elements of putative transcriptional activator/ biosynthetic genes. The -10 and -35 boxes were predicted using BPROM. Black bars represent 100 bp in length and ARE sequences are represented as a turquoise box.

Genes divergently adjacent to putative transcriptional activators also encode for proteins with distinct functions (Figure 3-7 and Table 3-3); the *mmyY*, *sven_4178* and *sav_2300* gene products code for a putative hypothetical protein, acetyltransferase and integral membrane lysyl-tRNA synthetase respectively. The ARE sequences are located downstream or overlap promoter elements of all of the putative transcriptional activators/ biosynthetic genes. In *S. coelicolor* and *S. venezuelae*, the ARE sequences also overlap promoter elements of divergently adjacent genes; binding of the repressor is expected to regulate expressions of the transcriptional activators/ biosynthetic genes, as well as divergently adjacent genes.

However, in *S. venezuelae* the predicted promoter elements for *sven_4178* is located over 1 kb away from the start codon and it is unlikely that the predicted -10 and -35 boxes are the promoter elements for *sven_4178*; this suggests sequencing errors in this specific region may have missed a possible ORF divergently adjacent to *gbnA*. The *sven_4178* - *gbnA* intergenic region needs to be further investigated and resequenced. In *S. avermitilis*, the ARE sequence is located upstream of the *sav_2300* promoter elements; hence, the TetR repressors are proposed to not directly regulate expression of *sav_2300*.

**Table 3-3** – ARE sequences upstream of putative transcriptional activators/ biosynthetic genes in *S. coelicolor*, *S. venezuelae*, *S. avermitilis* and *S. hygroscopicus*.

| Organism/gene downstream of ARE sequence | bp upstream of start codon | Proposed function | ARE sequence |
|---|---|---|---|
| *S. coelicolor mmyB* | 13 | DNA-binding protein | AAAAAACCTTCGGGAAGGTTTGAC |
| *S. venezuelae sven_4179* | 10 | Glutamate decarboxylase | AGTTGATACCTTCGTGAAGGTTTTTATT |
| *S. avermitilis sav_2301* | 175 | SARP family transcriptional activator | AGAAAAATACCTTCTCAAAGGAATTATTCT |
| *S. hygroscopicus shjg_7323* | 211 | Monooxygenase | AGCAAACCATCATTACGGAATTTTTTAT |
| *S. hygroscopicus shjg_7324* | 76 | TylR-like regulatory protein | GGGAAAACTACCTGTGGGAAGGTATTTT |

Putative transcriptional activators are highlighted in green.
Palindromic sequences are highlighted in red.

Predictions of ARE sequences and their respective location to the -10 and -35 boxes suggests that the AHFCA-dependent signalling systems are auto-regulatory. The ARE sequences are downstream or overlap promoter elements of TetR repressors and genes involved in the assembly of the cognate ligand. The TetR repressors are

proposed to not only regulate their own expression, but also regulate the assembly of AHFCA signalling molecules, which abolishes binding of the TetR repressors to ARE sequences (Figure 3-5 and Figure 3-6). In all of the putative AHFCA-dependent signalling systems, with the exception of *S. venezuelae*, the third ARE sequence is located in the promoters of putative pathway-specific transcriptional activator. The presence of transcriptional activators are expected to act as a signal amplifier to overexpress biosynthetic genes and to increase the production of specific natural compounds. In addition, rare TTA codon is also present in *mmyB* and *sav_2301* activators (Figure 3-4) and dependancy on BldA would further fine-tune the regulation of cryptic biosynthetic gene clusters. [57]

### 3.4. Highly conserved ARE motifs across *Streptomyces* bacteria

The ARE sequences, which were found in signalling gene clusters and neighbouring cryptic biosynthetic gene clusters, can be further refined to generate a consensus ARE motif by resubmitting promoter sequences into the MEME program (Figure 3-8 and Appendix B). The refined consensus ARE sequence motifs are highly conserved within and across the *Streptomyces* systems (Figure 3-8). The conserved pattern is more noticeable in the motifs for *S. coelicolor, S. venezuelae* and *S. avermitilis*, where they share the semi-palindromic sequence, AANATACCTTCNNGAAGGTATNTT (palindromic sequence is highlighted in red) (Figure 3-8).

Few differences in the refined consensus ARE sequence motif were observed in *S. hygroscopicus*; for example, the *S. hygroscopicus* consensus motif contains a longer variable spacer of 2 bp and extremities of the motif are represented by a conserved AAA/TTT sequence instead of the ATA/TAT sequence as in *S. coelicolor, S. venezuelae* and *S. avermitilis* consensus motifs (Figure 3-8).

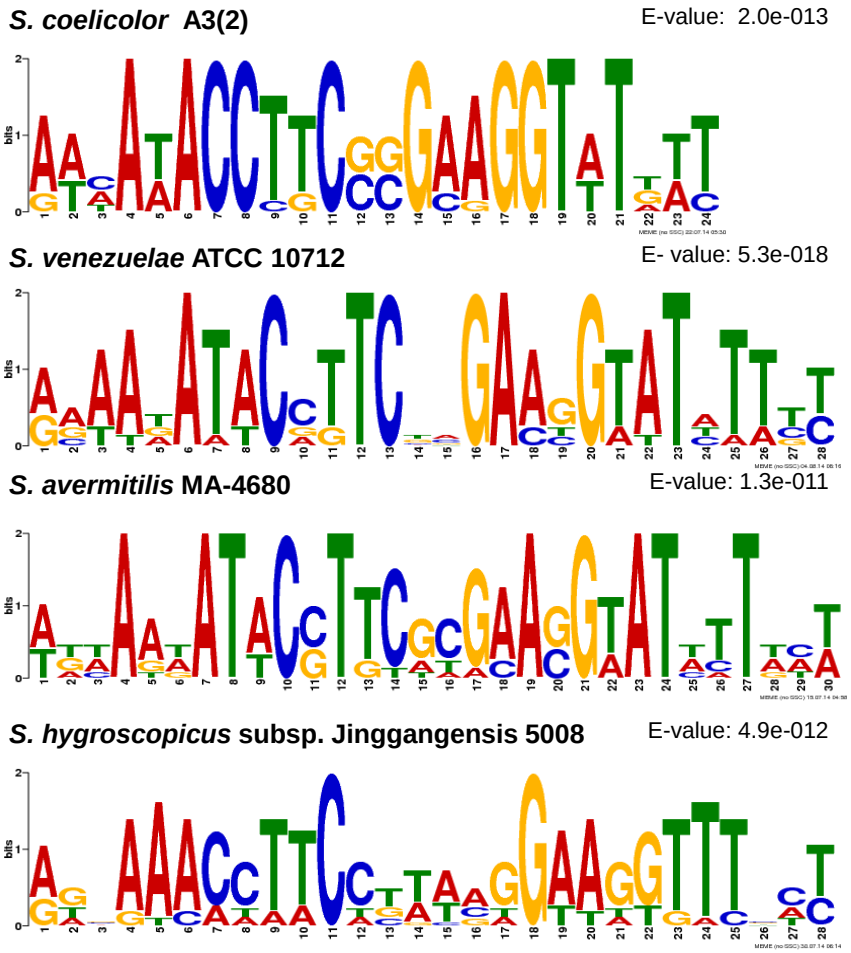



**Figure 3-8** – Refined consensus ARE motifs for *S. coelicolor*, *S. venezuelae*, *S. avermitilis* and *S. hygroscopicus*.

To investigate possible relationships between the consensus ARE motifs and HTH DNA-binding domains, TetR repressors were aligned using Clustal Omega.[144] From alignment results, amino acids proposed to be involved in interactions with DNA bases were different in *S. hygroscopicus* TetR repressors (Figure 3-9).

In the SHJG_7322 repressor, an alanine residue was found in position 48 of α-helix 3, instead of glycine, which is conserved in the other analogous MmyR repressors. Based on the *E. coli* TetR repressor, this specific amino acid residue was shown to be involved in interactions with DNA bases (Figure 3-9). Although alanine and

glycine only differ by a single methyl group, these differences may significantly alter affinity to the ARE sequences. Furthermore, amino acid residues in α-helix 3, also known as the recognition helix, were shown to be primarily involved in the binding to extremities of their respective DNA sequence. [70]

As for the analogous MmfR repressors, amino acid residues in α-helix 3 are all conserved; however, the residue proposed to be involved with DNA contact in α-helix 2 all differ (Figure 3-9). SAV_2268, MmfR and SgnR repressors contain amino acids with non-polar side chains (isoleucine, valine and leucine amino acids respectively), whilst in ShbR1, a methionine amino acid was found, which consists of a straight-chain with a sulphur atom (Figure 3-9). Based on the *E. coli* TetR repressor, this specific amino acid residue was shown to be involved in interactions with beginnings of the variable spacer. The methionine at position 49 of α-helix 2 in ShbR1 may explain differences in the variable spacer region. Although amino acid residues required for DNA binding are based on the *E. coli* TetR repressor, preliminary bioinformatic analysis has provided insights into observed variation in the refined consensus ARE motifs. It should be noted that the amino acid residue differences in SHJG_7322 and ShbR1 are not conserved in the two repressors and implies they may bind to different DNA sequences (Figure 3-9).

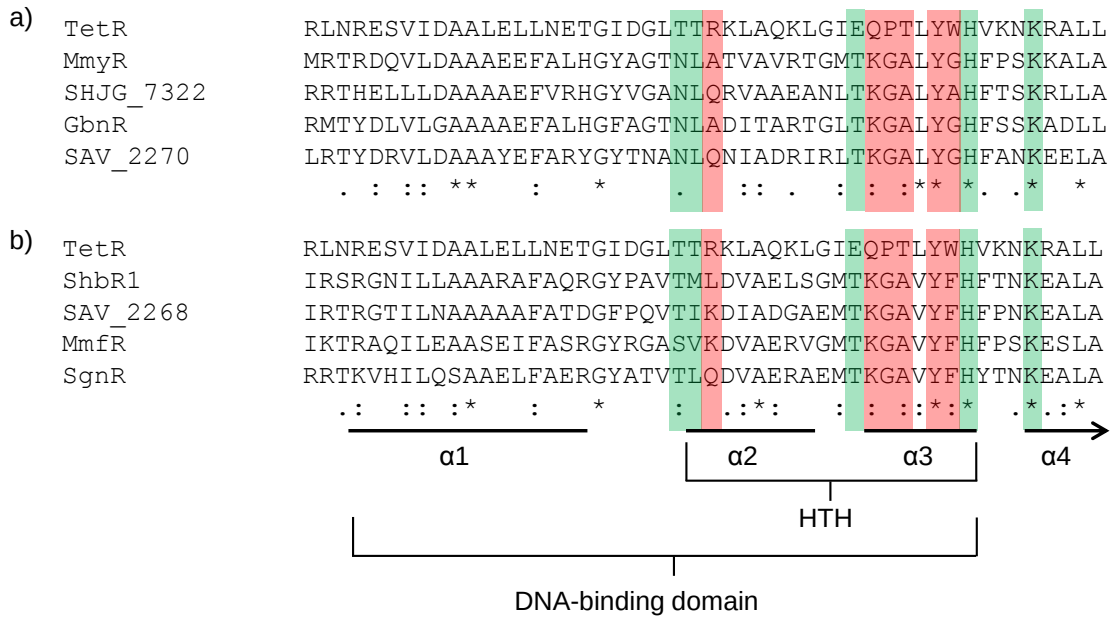



**Figure 3-9 –** Clustal Omega alignment of analogous (a) MmyR and (b) MmfR repressors reveals DNA-binding domains are highly conserved; however, small variations were identified and could contribute to the differences in the consensus ARE motifs. Amino acid residues involved in DNA contact and phosphate backbone are highlighted in red and green respectively (based on *E. coli* TetR repressor).

Early bioinformatic analyses of the *S. coelicolor* methylenomycin biosynthetic gene cluster published by O'Rourke and co-workers, identified highly conserved 18 bp sequences in the promoters of key regulatory genes (Figure 1-19). [57] MEME enrichment analysis successfully identify the same conserved sequences as well as further extend the consensus semi-palindromic sequence to 24 bp (Figure 3-8). MEME enrichment analysis was also successful in discovering ARE sequences in *S. venezuelae*, *S. avermitilis* and *S. hygroscopicus*, which were previously unknown. [62]

The prediction of TetR repressor binding sites upstream of putative transcriptional activators is of particular interest, as it provides specific targets to genetically manipulate in order to overproduce specific compounds from cryptic biosynthetic gene clusters. The *S. hygroscopicus* gene cluster contains two putative transcriptional activator genes (*shjg_7324* and *shjg_7325*, which encode for

TylR-like transcriptional regulator and streptomycin biosynthesis operon regulator respectively) and since the ARE sequence was only discovered in the *shjg_7324* promoter, overexpression of *shjg_7324* is expected to unlock the production of specific natural compounds.

Binding of analogous MmfR repressors to ARE sequences in *S. coelicolor*, *S. venezuelae* and *S. avermitilis* have been demonstrated through a combination of *in vivo* luciferase assays (Styles, K. Personal communication), *in vitro* EMSA (Malet, N., Harrison, P., Zhou, S., and Fullwood, A. Personal communication) and have ultimately led to the crystallisation of MmfR with MMF2 (Dean, R. *et al*. Manuscript in preparation). The MEME enrichment analysis provides an incredibly powerful approach for gaining insights into TetR repressor binding sites and to develop strategies to unlock the production of novel natural compounds.

### 3.5. MEME enrichment analysis in other putative AHFCA-dependent signalling systems

In addition to the four AHFCA-dependent signalling systems, genome mining analysis identified other putative AHFCA signalling gene clusters in *Streptomyces* bacteria (*S. sclerotialus* NRRL ISP-5269, *S.* sp. HGB0020, *S. roseochromogenes* subsp. Oscitans DC 12.976), as well as in *Kitasatospora* (*Kitasatospora cheerianensis* KCTC 2396), a closely related genus (Figure 3-10). [145]

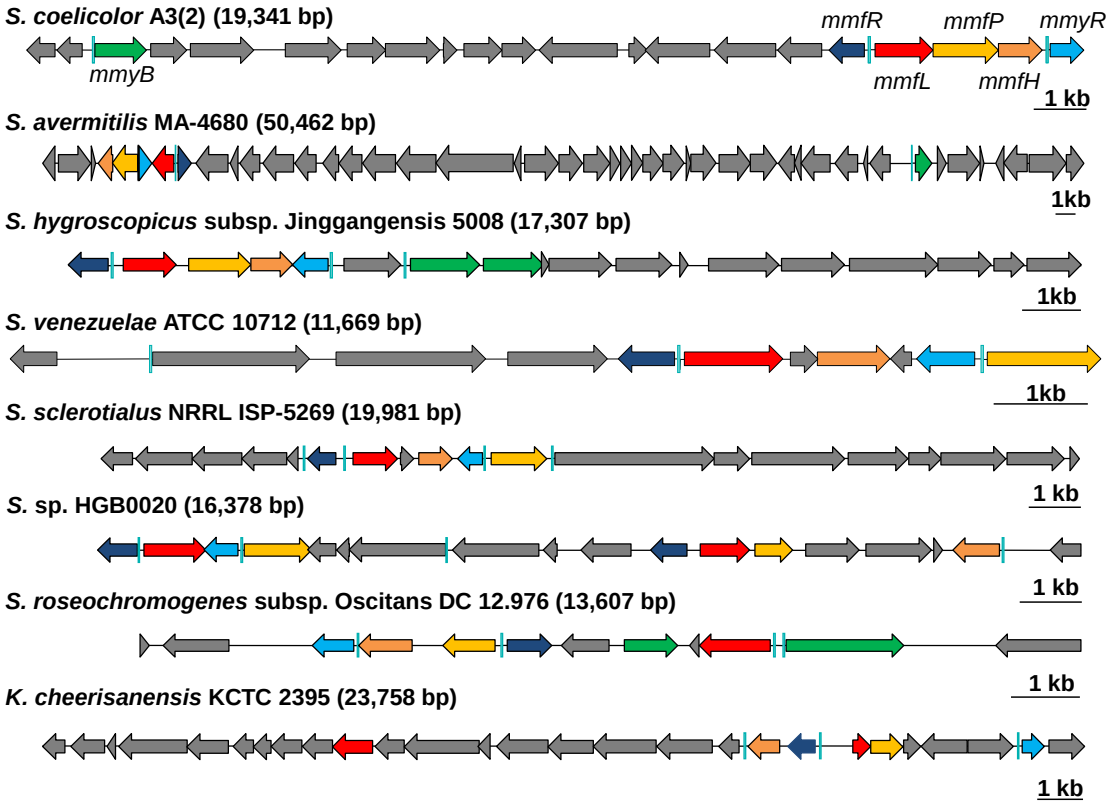



**Figure 3-10** – AHFCA-dependent signalling system are widespread in *Streptomyces* bacteria. ARE sequences are shown as a turquoise bar and genes with proposed analogous activity are colour-coded.

Using the same MEME enrichment analyses, TetR repressor binding sites were successfully identified in other putative AHFCA-dependent signalling systems and as expected, ARE sequences were discovered in the promoters of key regulatory genes (Figure 3-10, Figure 3-11 and Appendix C). Furthermore, newly identified consensus motifs are highly similar to motifs for *S. coelicolor*, *S. venezuelae*, *S. hygroscopicus* and *S. avermitilis* (Figure 3-8 and Figure 3-11).

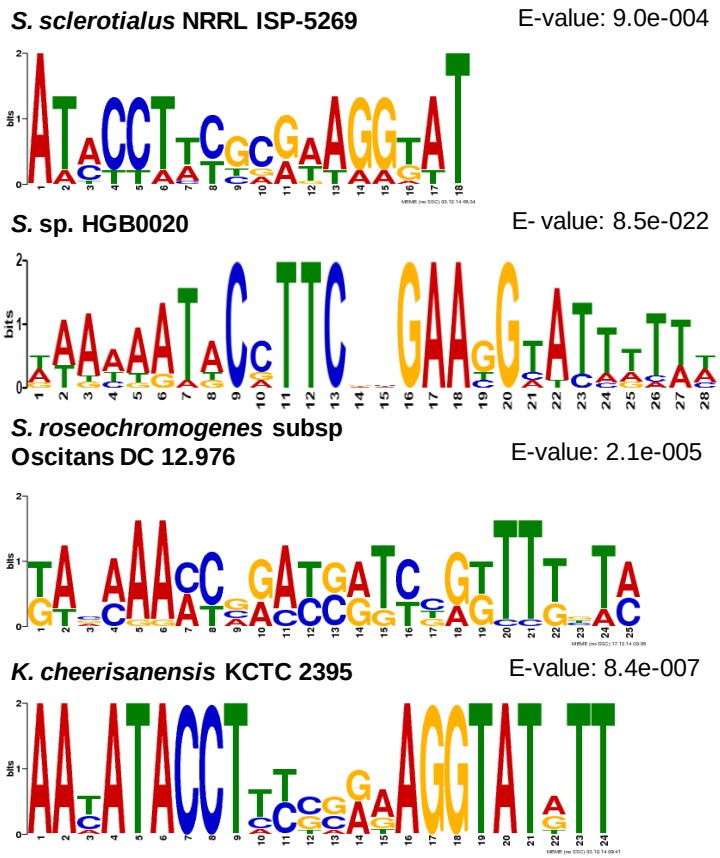



**Figure 3-11** – Refined consensus ARE motifs for *S. sclerotialus*, *S.* sp. HGB0020, *S. roseochromogenes* and *K. cheerisaenensis.*

ARE sequences were found in the promoters of TetR repressor genes, *mmfL*-like genes and putative pathway-specific transcriptional activators or biosynthetic genes (Figure 3-10 and Table 3-4). Surprisingly, the gene organisation of the putative AHFCA signalling gene clusters are not as well conserved in *S. roseochromogenes* and in *S.* sp. HGB0020. In *S. roseochromogenes*, the analogous *mmfL* gene is not divergently adjacent to a TetR repressor, but divergently adjacent to a putative SARP transcriptional activator and in *S.* sp. HGB0020, the analogous *mmfH* gene is over 12 kb away from the *mmfL* analogue (Figure 3-10).

The mechanism of AHFCA-dependent regulation in *S. roseochromogenes* is intriguing as the *mmfL* gene is divergently adjacent to the putative SARP

transcriptional activator and suggests the derepression of the system immediately allows expression of both genes. MEME enrichment analysis was also able to reveal two ARE sequences within the *m878_26325* - *m878_26320* intergenic region and most interestingly, the sequence directly upstream of the *m878_26325* (*mmfL*-like) is not as palindromic as the sequence directly upstream of *m878_26320* (Figure 3-10 and Table 3-4); suggesting that the TetR repressors may have different affinities to the two ARE sequences within this specific intergenic region.

MEME enrichment analysis successfully revealed ARE sequences in promoters of putative pathway-specific transcriptional activators in *S. coelicolor*, *S. avermitilis*, *S. hygroscopicus* and *S. roseochromogenes* (Figure 3-10, Table 3-3 and Table 3-4). The TetR repressors are hypothesised to bind to specific ARE sequence and tightly regulate expression of transcriptional activators; the expression of transcriptional activators amplifies signals for biosynthesis of metabolites. Overexpression of *mmyB*, *sav_2301*, *shjg_7324* and *m878_26320* transcriptional activators in their respective bacterium is expected to drive expression of cryptic biosynthetic gene clusters and overproduce the compounds of interest.

**Table 3-4** – ARE sequence upstream of putative pathway-specific transcriptional activators/ biosynthetic genes in *S. sclerotialus*, *S.* sp. HGB0020, *S. roseochromogenes* and *K. cheerisanensis*.

| Organism/gene downstream of ARE sequence | bp upstream of start codon | Proposed function | ARE sequence |
|---|---|---|---|
| *S. sclerotialus orf1* | 43 | Non-ribosomal peptide synthetase | **ATACCTCCG**CGAAGGAAT |
| *S. sclerotialus ig96_rs0135515* | 52 | QncM-like ACP | ATCCCAATGAGAAAAGAT |
| *S.* sp. HGB0020 *hmpref1211_06619* | 6 | Methylmalonyl-CoA carboxyltransferase | TTAAAAAACCTTCACGAAGGAATTTAAA |
| *S. roseochromogenes m878_26320* | 22<br>156 | SARP family transcriptional activator | AAAAAACCTACGGGACAGAATATT<br>ATAAAACCTTCTCGAAGGTATCTT |
| *K. cheerisaenensis kch_13110* | 117 | 3-oxoacyl-ACP reductase | AAAATACCTTCCCAAAGGTATGTT |

Putative transcriptional activators are highlighted in green.
Palindromic sequences are highlighted in red.

In the remaining 4 neighbouring cryptic biosynthetic gene clusters, MEME enrichment analyses were unable to identify ARE sequences upstream of putative transcriptional activators and this is explained from bioinformatic analyses; which suggests there are no cluster-situated transcriptional activator genes (Figure 3-10). Overexpression of transcriptional activator to discover products of cryptic biosynthetic gene clusters cannot be employed in *S. venezuelae*, *S. sclerotialus*, *S.* sp. HGB0020 and *K. cheerisaenensis*. To overcome this problem, genetic inactivation of analogous *mmyR* repressor in methylenomycin and gaburedin biosynthetic gene cluster unlocked the production of specific compounds; thus, the deletion of *mmyR*-like transcriptional repressor in all of the putative AHFCA-dependent signalling systems is expected to unlock the production of compounds from neighbouring cryptic biosynthetic gene clusters, as well as AHFCA signalling molecules (Figure 3-12). [57,62] Gene annotations for the 8 putative

AHFCA-dependent biosynthetic gene clusters and the identification of MmyR analogue are detailed in Appendix D to Appendix K.

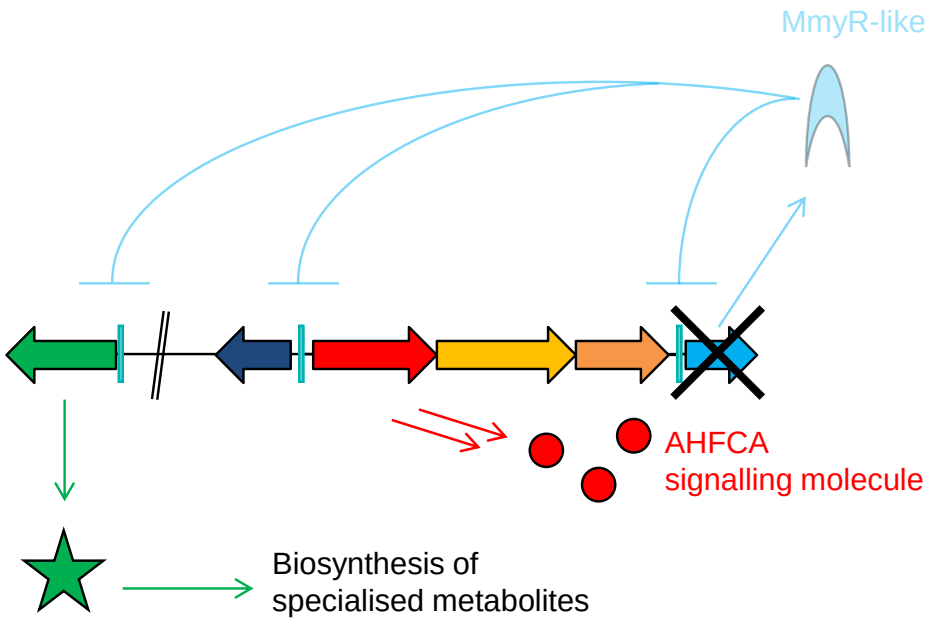



**Figure 3-12** – Deletion of *mmyR*-like repressor (light blue) is expected to unlock the production of specific specialised metabolites and AHFCA signalling molecules (red circles).

## 3.6. Acquisition of AHFCA-dependent signalling systems by horizontal transfer

The ARE sequences, which are positioned in promoters of putative transcriptional activators and biosynthetic genes, vary significantly (Table 3-3 and Table 3-4) and the similarlity of ARE sequences were investigated by phylogenetic analysis (Figure 3-13).

The bootstrap consensus phylogenetic tree reveals that the ARE sequences appear to cluster according to their location within the AHFCA signalling gene cluster; for example, with the exception of *hmpref1211_06616* sequence, all of the *mmfL*-like

ARE sequences are within the same clade and implies the ARE sequences upstream of *mmfL*-like genes are most similar to each other (Figure 3-13).

The same clustering pattern can be observed for the other signalling genes (Figure 3-13). Intriguingly, ARE sequences that are located upstream of putative pathway-specific transcriptional activators do not cluster together and suggests the ARE sequences are not similar to each other (Figure 3-13, Table 3-3 and Table 3-4). To tightly regulate neighbouring cryptic biosynthetic gene cluster, it is essential for TetR repressors to bind strongly to the ARE sequence. ARE sequences upstream of the transcriptional activators/ biosynthetic genes are likely to have evolved individually over time to maximise its binding affinity to the TetR repressors. Clustering of ARE sequences upstream of specific AHFCA signalling genes implies the sequences have also evolved over time to optimise its binding strength to regulate specific genes functions. These observations were supported by preliminary *in vivo* luciferase assay results where MmfR demonstrated the strongest binding to the ARE sequence upstream of *mmyB* and the ARE sequence upstream of the *mmfL* promoter showed the weakest binding affinity (Styles, K. Personal communication).

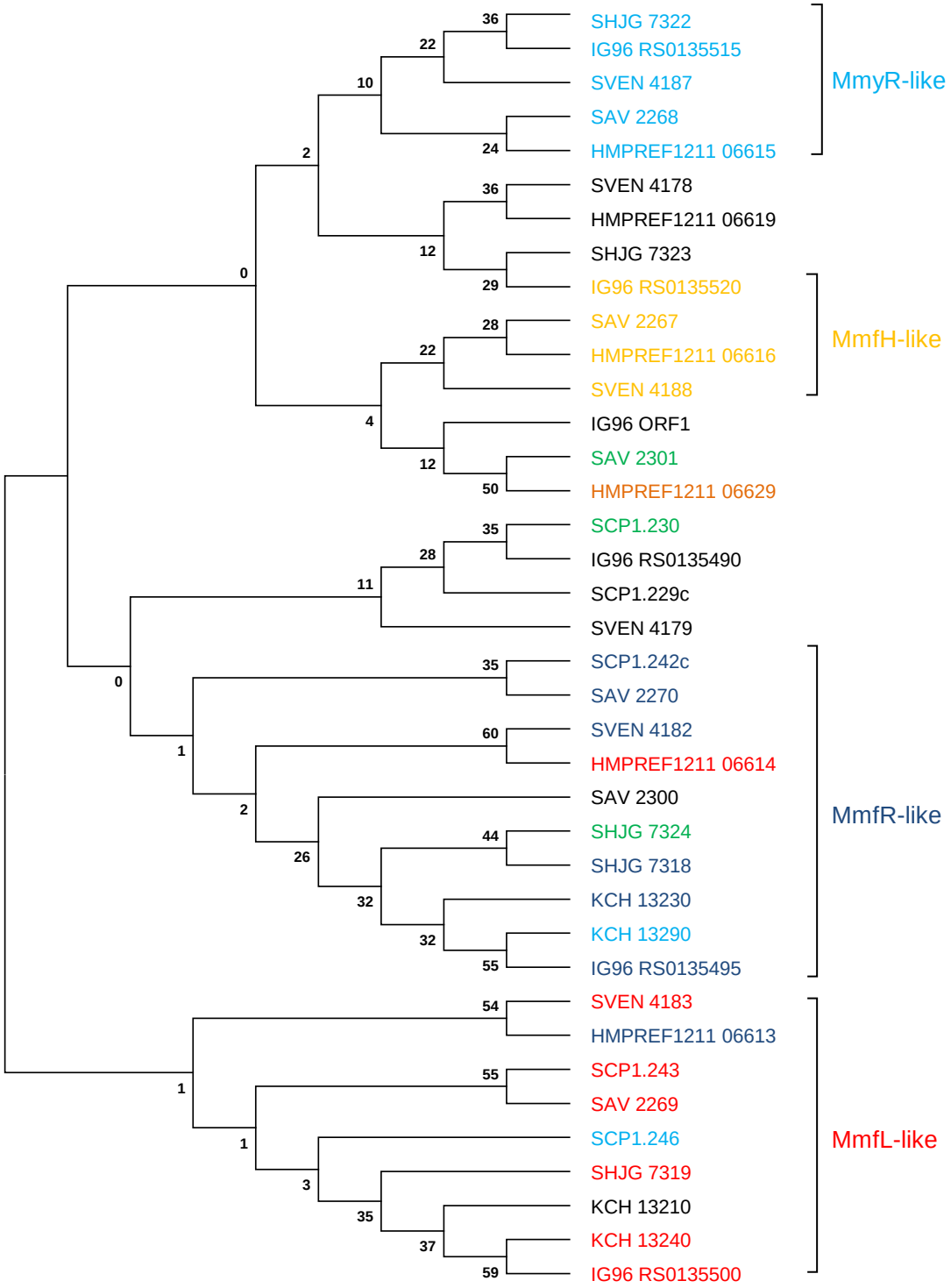



**Figure 3-13** – ARE sequences are clustering according to their respective locations in the AHFCA signalling gene cluster. ARE sequences were aligned using Multiple Sequence Comparison by Log Expectation (MUSCLE) and bootstrap consensus tree was inferred using the maximum likelihood clustering method after 1000 iterations. Genes with analogous functions are colour-coded.

Different cryptic biosynthetic gene clusters were located adjacent to conserved AHFCA signalling gene clusters and observations from phylogenetic analysis, suggests the AHFCA signalling gene clusters were likely acquired by *Streptomyces* bacteria through horizontal transfer. The selective advantage for acquiring AHFCA-dependent signalling system would allow the bacteria to tightly regulate expression of specialised metabolites biosynthetic gene cluster, in consequence more energy and metabolic flux are available for other essential biological functions. [95]

To further investigate differences/ similarities between ARE sequences at the nucleotide level (Figure 3-13), ARE sequences from each clade within the phylogenetic tree were submitted through the MEME program (Figure 3-14). Consensus motifs for AHFCA signalling genes reveal small differences (Figure 3-14). Out of the 4 motifs, the *mmyR*-like motif appears to be the least conserved with only 7 nucleotides fully conserved in the motif, whilst *mmfH*-like, *mmfR*-like and *mmfL*-like motifs have at least 16 nucleotides conserved (Figure 3-14). A possible explanation for the varied ARE sequence upstream of MmyR-like repressor could be due to the predictive function of MmyR-like repressors. The MmyR repressor is speculated to act as an off-switch to turn off methylenomycin production and binding affinity to ARE sequence upstream of *mmyR*-like genes has most likely refined over time within *Streptomyces* bacteria to prevent untimely inhibition. The previously described "variable" region are also surprisingly well conserved in the consensus motifs (Figure 3-14). Selective conservation of the "variable" region would collectively contribute to binding affinity of TetR repressors.

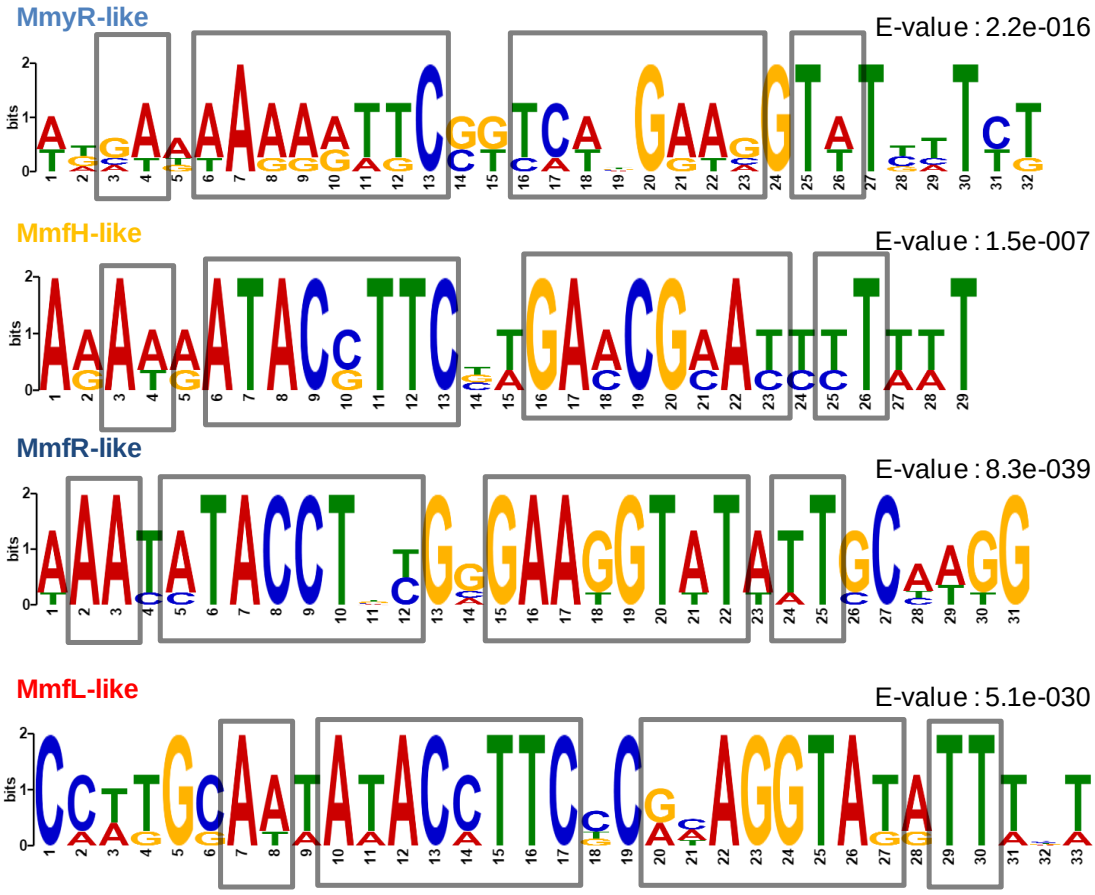



**Figure 3-14** – Consensus ARE sequences upstream of AHFCA signalling genes. The "conserved" regions from Figure 3-8 are highlighted in a grey box.

## 3.7. Conclusions

The production of metabolites from cryptic biosynthetic gene clusters are most often tightly regulated at the transcriptional level, and are silent or produced in very small quantities in laboratory conditions.

The AHFCA-dependent signalling system, which is governed by the two TetR repressors, is an example of complex regulatory networks involved in controlling expression of cryptic biosynthetic gene clusters. From phylogenetic analyses, *Streptomyces* bacteria are speculated to have acquired AHFCA-dependent signalling

system by horizontal transfer to regulate neighbouring biosynthetic gene clusters. Using bioinformatic tools, such as MEME and FIMO, TetR repressor binding sites were successfully identified in the methylenomycin biosynthetic gene cluster and were consistent with published preliminary bioinformatic analysis. [57] Genome mining analyses identified 7 additional putative AHFCA-dependent signalling systems and MEME enrichment analysis successfully revealed TetR binding sites in all of the systems. The TetR binding sites were located in the promoters of key regulatory genes, including putative transcriptional activators and TetR repressors. Specific inactivation of *mmyR*-like repressor, as well as the overexpression of putative transcriptional activators is expected to unlock the production of compounds from neighbouring cryptic biosynthetic gene clusters.

MEME enrichment analysis has proven to be an immensly powerful tool at gaining insights into complex regulatory networks. Exploitations of AHFCA-dependent signalling systems should lead to the discovery of novel specialised metabolites.

## 4. GBL-dependent Signalling Systems

The AHFCA class of compounds is an example of small diffusible signalling molecules involved in switching on the production of bioactive metabolites. Other signalling molecules, such as the γ-butyrolactones (Figure 1-15), derived from intermediates in fatty acid metabolism and glycolysis have been discovered and are also involved in regulating the production of natural compounds from cryptic biosynthetic gene clusters. [92,95]

### 4.1. MEME enrichment analysis of the SCB/ScbR/ScbR2 system

To demonstrate whether the MEME enrichment analysis could predict TetR repressor binding sites in other analogous regulatory systems, the same analysis was performed on the extensively studied SCB/ScbR/ScbR2 system, which controls coelimycin biosynthesis in *S. coelicolor* (discussed in Chapter 1.5.2). [117,121] Previous published results have shown that both ScbR and ScbR2 repressors can bind to highly conserved semi-palindromic ARE sequences within the *scbR* - *scbA* intergenic region and the promoter of *cpkO,* which encodes for a SARP transcriptional activator (Figure 4-1). [98,99,117,120,140]

MEME enrichment analysis was successful at finding both SiteA and SiteR in the *scbR* - *scbA* intergenic region and SiteOB in the promoter of *cpkO*, as well as an ARE sequence upstream of *cpkN* (Figure 4-1, Figure 4-2 and Electronic Supplementary Material). The MEME analysis enriches for highly conserved sequences and despite the palindromic nature of SiteOA, the failure to identify SiteOA suggests it is not as well conserved as the other three ARE sequences (Figure 1-18).

a) *scbR* - *scbA* (*sco6265* – *sco6266*)

b) Upstream of *cpkO* (*sco6280*)


**Figure 4-1** – MEME enrichment analysis was able to discover the ARE sequences (turquoise box) in the (a) *scbR* - *scbA* (*sco6265* - *sco6266*) intergenic region and (b) the promoter of *cpkO* (*sco6280*). The experimentally confirmed ScbR binding sites are underlined in black.

MEME enrichment analysis of AHFCA-dependent signalling systems identified ARE sequences located upstream of both TetR repressors; however, in the SCB/ScbR/ScbR2 system, MEME enrichment analysis could not find any binding sites upstream of *scbR2* and suggests the TetR repressors do not directly regulate expression of *scbR2* repressor. Interestingly, temporal expression analysis have shown that the production of ScbR2 is delayed and is expressed after ScbR; the delay

in expression implies other transcriptional factors are regulating expression of *scbR2*. [140]

Although the MEME enrichment analysis has proven to be an incredibly powerful approach to discover TetR binding sites, the bioinformatic analysis alone cannot deduce which repressor can bind to the ARE sequences. MEME enrichment analysis was only able to identify SiteOB within the *cpkO* promoter and results from EMSA experiments have also shown ScbR2 repressor can bind to SiteOB and not SiteOA. [120,140] The MEME enrichment analysis might be bias towards the prediction of ScbR2 binding sites; this is further supported by the experimentally confirmed binding of ScbR2 to the *cpkN* promoter, which was reported by Li and co-workers, and the successful identification of the ARE sequence located in the *cpkN* promoter. [117,140]

As for the AHFCA-dependent signalling systems, early experimental results have demonstrated the MmfR-like repressors can bind to the ARE sequences within their respective cryptic biosynthetic gene clusters (Harrison, P., Malet, N., Zhou, S., Styles, K. and Fullwood, A. Personal communication) and the analogous MmyR repressor exhibited very weak or no binding to the same ARE sequences (Fullwood, A. and Styles, K. Personal communication); these observations suggests the MEME enrichment analysis might also be bias towards predicting MmfR-like binding sites.

Half palindromic sequences within the *mmr - mmyJ* (*scp1.237c - scp1.238*) intergenic region and upstream of *mmyQ* (*scp1.231*) and *mmyT* (*scp1.241c*) (Table 4-1) were observed within the methylenomycin biosynthetic gene cluster. [57] The

binding of MmfR or MmyR repressors to the half palindromic sites have not been investigated; however, the homodimeric repressors are speculated to bind loosly with the half palindromic sequences. Alternatively, the half-palindromic sequences could be the binding sites for the MmyR repressor.

**Table 4-1** – Half palindromic sequences in the methylenomycin gene cluster.

| Start | End | Half palindrome ARE sequence | bp upstream of start codon | Gene downstream of sequence |
|---|---|---|---|---|
| 3756 | 3764 | ATACCTGCG | 33 | *scp1.231* (*mmyQ*) |
| 12349 | 12357 | GTCGGGTAT | 187 | *scp1.238* (*mmyJ*) |
| 12349 | 12357 | ATACCCGAC | 20 | *scp1.237c* (*mmr*) |
| 16127 | 16135 | TTACGTTCT | 58 | *scp1.241c* (*mmyT*) |

a)


| Name | Bp Upstream | p-value | Sites |
|---|---|---|---|
| SCO6266 | 56 | 2.54E-11 | GAAGGATAGA AAAAAAACCGCTCAGTCTGTAT CTTAACGTTC |
| SCO6265 | 39 | 6.78E-11 | GAACGTTAAG ATACAGACTGAGCGGTTTTTTT TCTATCCTTC |
| SCO6288 | 139 | 2.59E-09 | GCCCACAATG AAACGGACCGCTCGGTCTCTTC TCTCGACTCC |
| SCO6280 | 257 | 3.01E-09 | GGACGGGCCA AAACAGACTGTGTTAGCTGTTT CAGTCTCGAA |
| SCO6265 | 168 | 6.71E-08 | AGCGCCATCG GAACCGGCAATGCGGTTTGTTC GATCGAGTTG |

b)

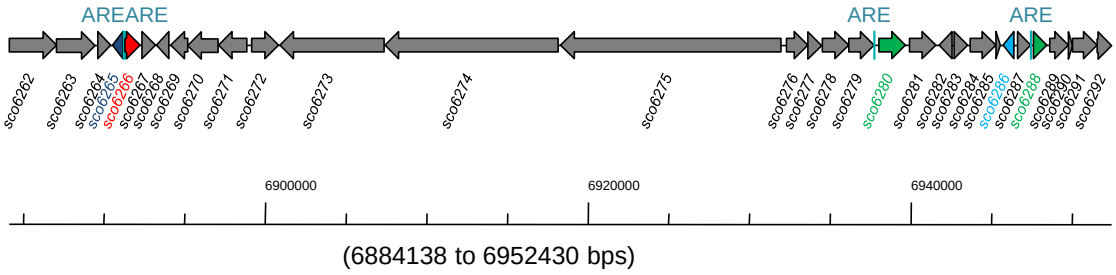



**Figure 4-2** – MEME enrichment analysis of the SCB/ScbR/ScbR2 system identified (a) highly conserved ARE sequences and (b) are location upstream of *sco6265*, *sco6266*, sco6280 and *sco6288* genes, which encode for the TetR repressor (blue), ScbA biosynthetic enzyme (red) and the CpkO and CpkN SARP transcriptional activators (green) respectively.

A combination of ChIP-seq, transcriptome and EMSA analysis carried out by Li and co-workers discovered several additional ScbR and ScbR2 binding sites within the *cpk* gene cluster. [140] Both ScbR and ScbR2 repressors bound upstream of *cpkM* (*sco*6268) and only the ScbR2 repressor was confirmed to bind to the *accA1* - *scF* (*sco6271* - *sco6272*), *cpkA* - *cpkD* (*sco6275* - *sco6276*) and *cpkI* - *cpkJ* (*sco6282* - *sco6283*) intergenic regions; demonstrating that ScbR2 has different binding specificities to ScbR repressor. [140] The SCB/ScbR/ScbR2 system appears to be incredibly complicated and the TetR repressors were discovered to bind to several targets across the genome including other specialised metabolite biosynthetic gene clusters. [140]

FIMO results for the SCB/ScbR/ScbR2 system, which contained a total of 785 individual hits, were analysed in detail to discover potential binding sites across the *S. coelicolor* genome (Electronic Supplementary Material). Binding sites located within the promoters of genes (300 bp from start codon) were investigated (Electronic Supplementary Material); it should be noted that some binding sites would have been missed because individual hits were located more than 300 bp away from the predicted ORF's start codon. From the 785 individual hits, 202 binding sites were located in the promoters of ORFs; however, the MEME enrichment analysis was only able to identify two binding sites, which were consistent with published experimental results (*sco4921* and *sco6312*) (Table 4-2). [140]

The *sco4921* and *sco6312* genes encodes for AccA2 acetyl-Coenzyme A (CoA) carboxylase and CprA TetR transcriptional repressor respectively. The AccA2 enzyme is an essential enzyme in *S. coelicolor* and is involved in the carboxylation of acetyl-CoA to malonyl-CoA, which are precursors for fatty acid biosynthesis and polyketide biosynthesis, such as actinorhodin and coelimycin. [146] CprA is an ArpA-like transcriptional repressor and the exact biological role of CprA is not clear, but CprA is known to regulate metabolite production and morphological differentiation. [147] The direct regulation of *accA2* and *cprA* is proposed to alter the production of other metabolites in *S. coelicolor* and redirect the metabolix flux of primary precursors.

**Table 4-2** – Predicted ScbR/ScbR2 binding sites upstream of *cprA* and *accA2*.

| Start | End | ARE sequence | bp upstream of start codon | Gene downstream of ARE sequence |
|---|---|---|---|---|
| 5357259 | 5357280 | AAAGAAACCGCGTGGCCCGCAG | 203 | *sco4921* (*accA2*) |
| 6972702 | 6972723 | AAAACAGGCACACGGTCTGTTG | 26 | *sco6312* (*cprA*) |

The failure to find the other experimentally confirmed ScbR and ScbR2 binding sites is suspected to be due to the strict nature of the algorithm. The FIMO program searches for individual hits according to the initial motif (positional weight matrix) and consequently does not account for gap sequences. Many of the experimentally confirmed binding sites contained different length of "variable" spacers and in some binding sequences, only one half of the palindrome sequence were present; these sequences would have been significantly different to the initial motif and would not have been picked up by FIMO.[140] Nevertheless, the MEME enrichment analysis was successful at identifying binding sequences upstream of the key regulatory genes within the coelimycin biosynthetic gene cluster.

## 4.2. MEME enrichment analysis of putative GBL-dependent signalling systems in *Streptomyces* bacteria

As of July 2014, the NCBI database contained 20 completely sequenced *Streptomyces* genomes and from these genomes, 679 cryptic specialised metabolites were predicted by AntiSMASH (Table 4-3).[133] Based on the identification of the *afsA*-like gene, 36 of these cryptic biosynthetic gene clusters are proposed to be under the control of regulation systems dependent on butenolide-derived signalling molecules (Table 4-3 and Electronic Supplementary Material).[94] The predicted GBL-dependent signalling systems were further investigated using MEME

enrichment analysis, not only to determine the binding sites for the TetR repressors, but also to examine the effectiveness of the bioinformatic approach at predicting other regulation systems (Electronic Supplementary Material).

**Table 4-3** – Genome information of sequenced *Streptomyces* bacteria.

| ***Streptomyces* Organism** | **Size (Mb)** | **GC %** | **ORF** | **Metabolite clusters** | **Butyrolactone-like clusters** |
|---|---|---|---|---|---|
| *S. albulus* NK660 | 9.37 | 72.3 | 8,238 | 40 | 5 |
| *S. albulus* NK660 (pNKL plasmid) | 0.01 | 71.2 | 22 | 0 | 0 |
| *S. albus* J1074 | 6.84 | 73.3 | 5,937 | 23 | 0 |
| *S. avermitilis* MA-4680 | 9.12 | 70.7 | 7,765 | 23 | 2 |
| *S. avermitilis* MA-4680 (SAP1 plasmid) | 0.09 | 69.2 | 96 | 0 | 0 |
| *S. bingchenggensis* BCW-1 | 11.94 | 70.8 | 10,107 | 49 | 1 |
| *S. cattleya* NRRL 8057 | 8.09 | 73.0 | 7,580 | 27 | 3 |
| *S. cattleya* NRRL 8057 (pSCATT plasmid) | 1.81 | 73.3 | 1,714 | 15 | 1 |
| *S. coelicolor* A3(2) | 9.05 | 72.0 | 8,299 | 23 | 1 |
| *S. coelicolor* A3(2) (SCP1 plasmid) | 0.36 | 69.1 | 355 | 2 | 1 |
| *S. coelicolor* A3(2) (SCP2 plasmid) | 0.03 | 72.1 | 34 | 0 | 0 |
| *S. collinus* Tu 365 | 8.38 | 72.6 | 7,207 | 32 | 0 |
| *S. collinus* Tu 365 (pSCO1 plasmid) | 0.09 | 68.3 | 78 | 0 | 0 |
| *S. collinus* Tu 365 (pSCO2 plasmid) | 0.02 | 69.9 | 30 | 0 | 0 |
| *S. davawensis* JCM 4913 | 9.56 | 70.6 | 8,697 | 35 | 1 |
| *S. davawensis* JCM 4913 (pSDA1 plasmid) | 0.09 | 69.9 | 113 | 0 | 0 |
| *S. fulvissimus* DSM 40593 | 7.91 | 71.5 | 7,027 | 35 | 1 |
| *S. hygroscopicus* subsp. Jinggangensis 5008 | 10.38 | 71.9 | 9,194 | 39 | 3 |

| | | | | | |
|---|---|---|---|---|---|
| *S. hygroscopicus* subsp. Jinggangensis 5008 (pSHJG1 plasmid) | 0.16 | 69.0 | 183 | 0 | 0 |
| *S. hygroscopicus* subsp. Jinggangensis 5008 (pSHJG2 plasmid) | 0.07 | 70.9 | 75 | 0 | 0 |
| *S. niveus* NCIMB 11891 | 8.73 | 70.3 | 7,679 | 28 | 1 |
| *S. niveus* NCIMB 11891 (pSniv1 plasmid) | 0.12 | 68.6 | 101 | 1 | 0 |
| *S. niveus* NCIMB 11891 (pSniv2 plasmid) | 0.05 | 71.4 | 53 | 0 | 0 |
| *S. niveus* NCIMB 11891 (pSniv3 plasmid) | 0.01 | 69.0 | 18 | 0 | 0 |
| *S. niveus* NCIMB 11891 (pSniv4 plasmid) | 0.02 | 67.4 | 17 | 0 | 0 |
| *S. pratensis* ATCC 33331 | 7.66 | 71.0 | 6,857 | 27 | 2 |
| *S. pratensis* ATCC 33331 (pSFLA01 plasmid) | 0.19 | 67.8 | 200 | 0 | 0 |
| *S. pratensis* ATCC 33331 (pSFLA01 plasmid) | 0.13 | 67.2 | 126 | 0 | 0 |
| *S. pristinaespiralis* ATCC 25386 | 8.13 | 69.7 | 7,077 | 20 | 0 |
| *S. rapamycinicus* NRRL 5491 | 12.70 | 70.6 | 10,144 | 48 | 2 |
| *S. roseochromogenes* subsp. Oscitans DS 12.976 | 9.77 | 70.6 | 8,808 | 43 | 2 |
| *S. roseochromogenes* subsp. Oscitans DS 12.976 (pSros1 plasmid) | 0.20 | 69.2 | 134 | 0 | 0 |
| *S. scabiei* 87.22 | 10.15 | 71.5 | 8,901 | 34 | 2 |
| *S. griseus* subsp. Griseus NBRC 13350 | 8.55 | 72.2 | 7,224 | 39 | 1 |
| *S. sviceus* ATCC 29083 | 9.31 | 69.8 | 8,464 | 25 | 2 |
| *S. venezuelae* ATCC 10712 | 8.23 | 72.4 | 7,536 | 31 | 3 |
| *S. violaceusniger* Tu 4113 | 11.14 | 70.9 | 9,610 | 39 | 1 |
| *S. violaceusniger* Tu 4113 (pSTRVI01 plasmid) | 0.29 | 68.2 | 298 | 1 | 1 |
| *S. violaceusniger* Tu 4113 (pSTRVI02 plasmid) | 0.19 | 69.7 | 250 | 0 | 0 |
| | | | **Total** | 679 | 36 |

The 36 putative GBL-dependent signalling systems also include the five previously analysed AHFCA-dependent signallng systems in *S. avermitilis*, *S. coelicolor*,

*S. hygroscopicus*, *S. roseochromogenes* and *S. venezuelae* (Figure 3-8, Figure 3-10 and Figure 3-11) as well as the SCB/ScbR/ScbR2 system (Figure 4-2). For the remaining 30 putative GBL-dependent signalling systems, MEME enrichment analysis successfully discovered conserved semi-palindromic ARE sequences in 12 gene clusters (Figure 4-3 and Figure 4-4). The consensus ARE motifs are listed in Figure 4-3 and correspond to the gene clusters in Figure 4-4 (Electronic Supplementary Material).

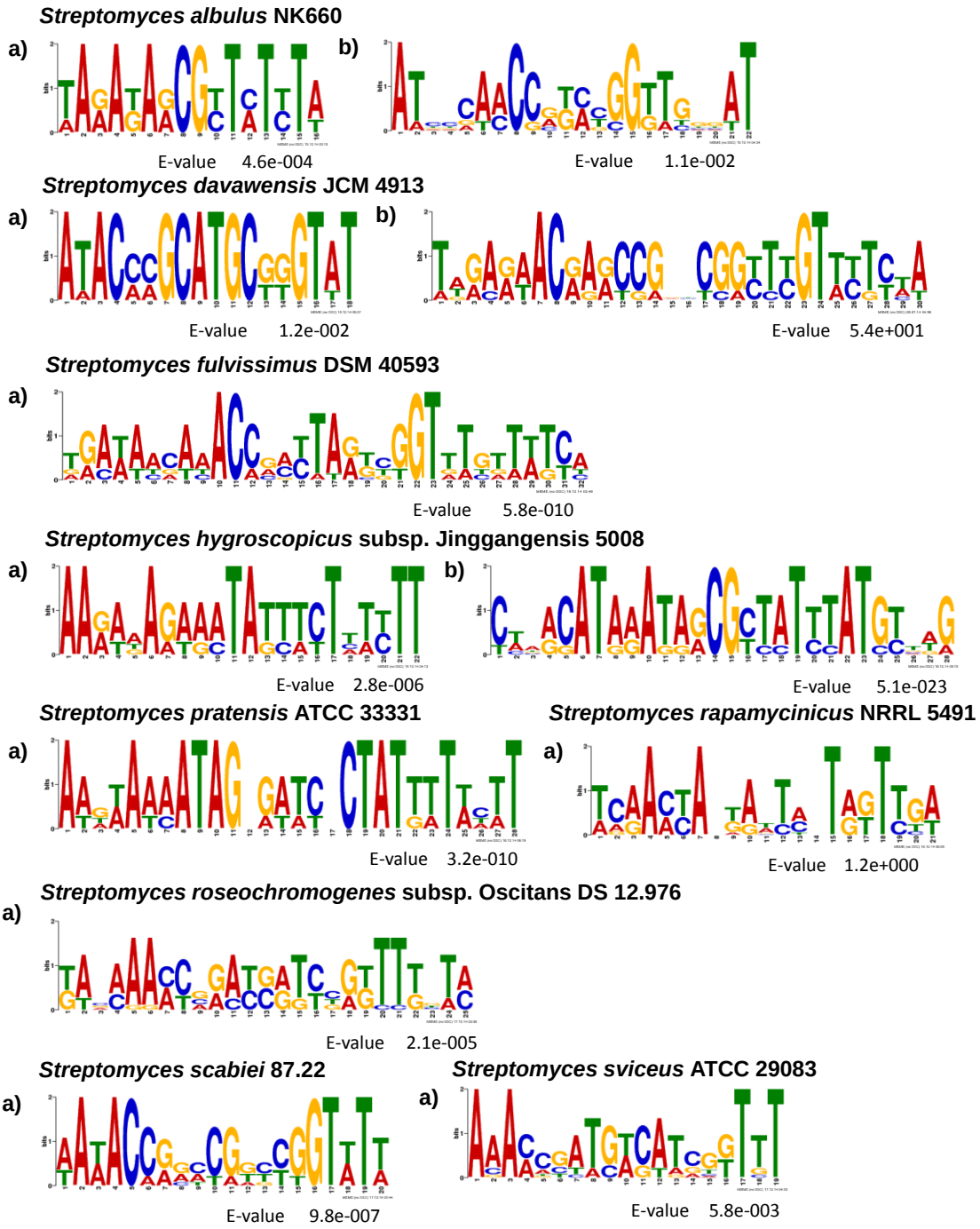



**Figure 4-3** – Motifs for putative GBL-dependent signalling systems; the TetR repressors are proposed to bind to ARE sequences to regulate biosynthetic genes. Consensus motifs correspond to cryptic biosynthetic gene clusters listed in Figure 4-4.

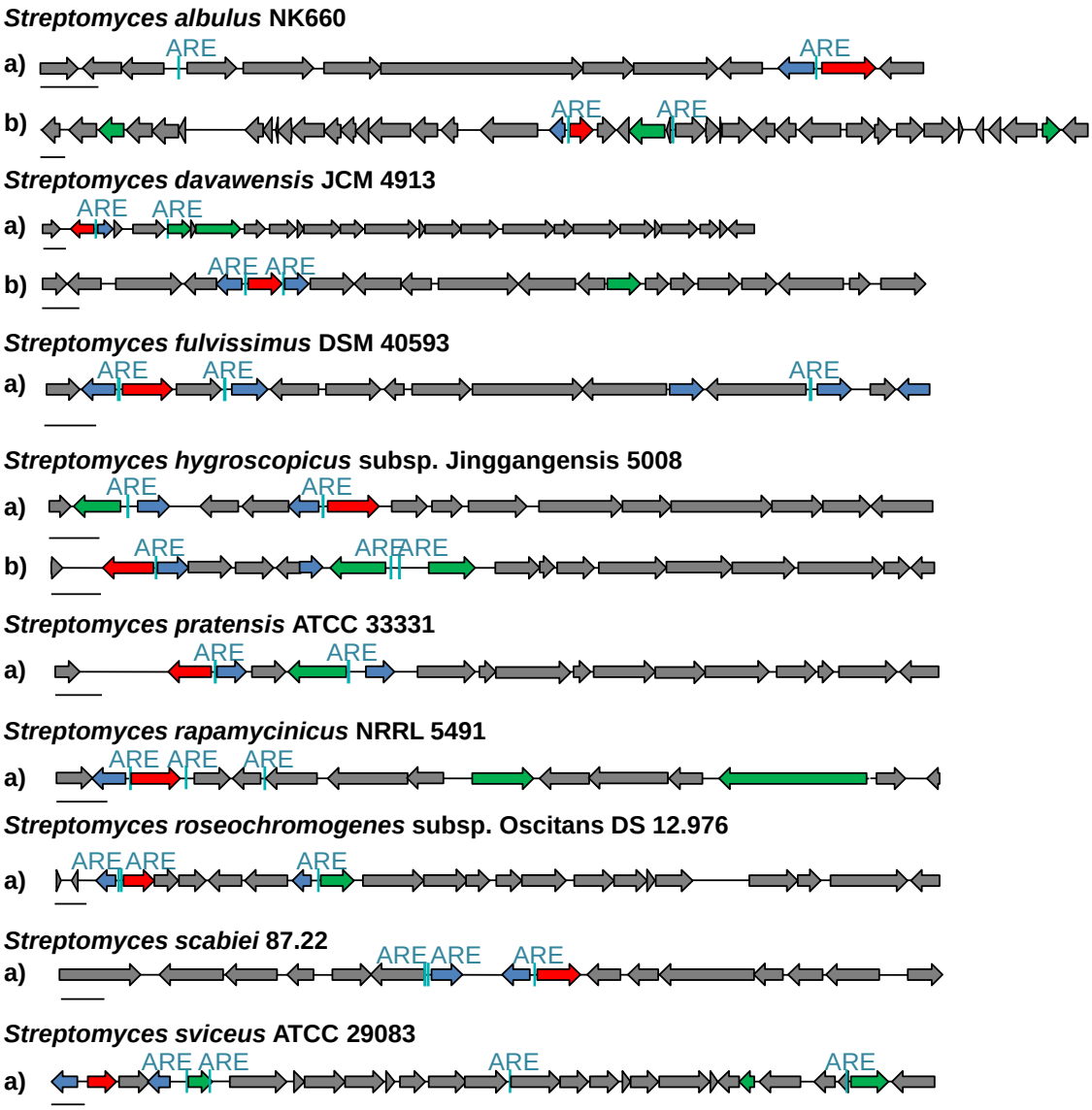



**Figure 4-4** – Putative GBL signalling gene clusters are adjacent to cryptic biosynthetic gene clusters. Analogous regulatory genes are colour-coded; TetR repressors (blue); A-factor biosynthesis enzyme (red); transcriptional activator (green). Predicted TetR binding sites (ARE) are highlighted as a turquoise box and black lines represent 1 kb in length.

The GBL consensus motifs (Figure 4-3) share characteristic patterns with the AHFCA consensus motifs (Figure 3-8 and Figure 3-11); the highly conserved extremities of the consensus motifs primarily consists of As and Ts, whilst the "variable" spacer region consists of Cs and Gs. In addition, the ARE sequences are always found in the intergenic region between the TetR repressor and *afsA*-like gene (Figure 4-4 and Table 4-4); however, in *Streptomyces sviceus* ATCC 29083, the ARE

sequence could not be identified upstream of the *afsA*-like gene, but was found in the promoters of putative transcriptional activators and biosynthetic genes (Figure 4-4). The failure to identify ARE sequence may be due to sequencing errors of the genome or alternatively, the putative GBL-dependent signalling system in *S. sviceus* may have a novel transcriptional regulation mechanism to control the neighbouring cryptic biosynthetic gene cluster.

Unlike the GBL consensus motifs, the ARE sequences in AHFCA-dependent signalling systems are conserved across the *Streptomyces* bacteria (Figure 4-3), and implies the GBL-dependent signalling systems have likely evolved for a longer period of time and novel regulations systems may have branched off and developed. Furthermore, the organisation of putative GBL signalling gene clusters (Figure 4-4) are not as well conserved and clustered as the AHFCA signalling genes (Figure 3-10); however, in both regulation systems, the TetR repressor gene is always divergent and adjacent to the *afsA*-like gene (Figure 3-10 and Figure 4-4).

**Table 4-4** – ARE sequences identified in GBL-dependent signalling systems.

| Organism (cluster) Gene downstream of ARE sequence | ARE sequence | bp upstream of start codon | Proposed function |
|---|---|---|---|
| *S. albulus* (a) *dc74_7607* | ATAAATAACGTTCTCTAT | 85 | A-factor biosynthesis enzyme |
| *S. albulus* (a) *dc74_7608* | ATAGAGAACGTTATTTAT | 27 | TetR family transcriptional repressor |
| *S. albulus* (a) *dc74_7615* | AAAGATAGCGCTCTTTAT | 138 | Short-chain dehydrogenase/ reductase |
| *S. albulus* (a) *dc74_7616* | ATAAAGAGCGCTATCTTT | 249 | Nitrate/nitrite sensor protein |
| *S. albulus* (b) *dc74_7317* | CGTGATCTCACCCGGATGGTTGGGATCCCG | 97 | Hypothetical protein |
| *S. albulus* (b) *dc74_7317* | CGGGATCCCAACCATCCGGGTGAGATCACG | 100 | Hypothetical protein |
| *S. albulus* (b) *dc74_7322* | TAAGATACGTACCATCCGGTTTTCTTTTCA | 136 | TetR family transcriptional repressor |
| *S. albulus* (b) *dc74_7321* | TGAAAAGAAAACCGGATGGTACGTATCTTA | 43 | A-factor biosynthesis enzyme |
| *S. davawensis* (a) *bn159_4439* | AAGATACAAGCATGCGGGTATGTT | 76 | TetR family transcriptional repressor |
| *S. davawensis* (a) *bn159_4440* | AACATACCCGCATGCTTGTATCTT | 59 | A-factor biosynthesis enzyme |
| *S. davawensis* (a) *bn159_4436* | AGAAAACCCGCATGCGGGTATGTT | -5 | AraC family transcriptional regulator |
| *S. davawensis* (b) *bn159_2279* | TTGAAAACAAACCGGTCGGCTCGTACTCTA | 60 | A-factor biosynthesis enzyme |
| *S. davawensis* (b) *bn159_2280* | TAGAGTACGAGCCGACCGGTTTGTTTTCAA | 86 | TetR family transcriptional repressor |
| *S. davawensis* (b) *bn159_2278* | TGAAAAACAAACGGCATGACCCGTTCGTTT | 19 | TetR family transcriptional repressor |
| *S. fulvissimus* (a) *sful_260* | AAACCACACCCACTAAGTGGTTAGTAAG | 116 | TetR family transcriptional repressor |

| | | | |
|---|---|---|---|
| *S. fulvissimus* (a) *sful_261* | CTTACTAACCACTTAGTGGGTGTGGTTT | 71 | Glucan endo-1,6-beta-glucosidase |
| *S. fulvissimus* (a) *sful_269* | AAATCAAACCGGTAAGTTGGTTTGATAT | 121 | TetR family transcriptional repressor |
| *S. fulvissimus* (a) *sful_271* | AAAAAAAACCGCTTAGTCTGTATCTTAT | 56 | A-factor biosynthesis enzyme |
| *S. fulvissimus* (a) *sful_272* | ATAAGATACAGACTAAGCGGTTTTTTTT | 61 | TetR family transcriptional repressor |
| *S. hygroscopicus* (a) *shjg_4003* | AAGATAGAGATAGTTCTTTTTT | 75 | TetR family transcriptional repressor |
| *S. hygroscopicus* (a) *shjg_4004* | AAAAAAGAACTATCTCTATCTT | 79 | A-factor biosynthesis enzyme |
| *S. hygroscopicus* (a) *shjg_4000* | AAGAGAGTAATATTTTTTATTT | 182 | TetR family transcriptional repressor |
| *S. hygroscopicus* (a) *shjg_3999* | AAATAAAAAATATTACTCTCTT | 132 | DNA-binding protein |
| *S. hygroscopicus* (b) *shjg_8650* | ACATAGATAGCGTCATTTATGT | 49 | A-factor biosynthesis enzyme |
| *S. hygroscopicus* (b) *shjg_8651* | ACATAAATGACGCTATCTATGT | 20 | TetR family transcriptional repressor |
| *S. hygroscopicus* (b) *shjg_8657* | GCATAAATAGCGTTATCTATCT | 747 | StrR-like regulatory protein |
| *S. hygroscopicus* (b) *shjg_8656* | AGATAGATAACGCTATTTATGC | 98 | StrR-like regulatory protein |
| *S. hygroscopicus* (b) *shjg_8656* | GCATAAAGAGCGCTATTCATGT | 269 | StrR-like regulatory protein |
| *S. hygroscopicus* (b) *shjg_8657* | ACATGAATAGCGCTCTTTATGC | 576 | StrR-like regulatory protein |
| *S. pratensis* (a) *sfla_6322* | AAGTAACATAGGGATTACTATTTTAATT | 22 | TetR family transcriptional repressor |
| *S. pratensis* (a) *sfla_6323* | AATTAAAATAGTAATCCCTATGTTACTT | 71 | A-factor biosynthesis enzyme |
| *S. pratensis* (a) *sfla_6320* | ATAAAAAATAGCGAACTCTATTATTCTT | 33 | TylR-like regulatory protein |

| | | | |
|---|---|---|---|
| *S. pratensis* (a) *sfla_6319* | AAGAATAATAGAGTTCGCTATTTTTTAT | 363 | TetR family transcriptional repressor |
| *S. rapamycinicus* (a) *m271_07485* | TCAAACATTATTAGTATTTTA | 2 | A-factor biosynthesis enzyme |
| *S. rapamycinicus* (a) *m271_07490* | TAAAATACTAATAATGTTTGA | 82 | TetR family transcriptional repressor |
| *S. rapamycinicus* (a) *m271_07475* | ACAACCACTACTTTTGGTCCA | 66 | Hypothetical protein |
| *S. rapamycinicus* (a) *m271_07480* | TCAAACTAACGTAGTGATCGT | 155 | Hydrolase |
| *S. roseochromogenes* (a) *m878_11420* | AAAACCGATCAGTCAGTTTT | 58 | A-factor biosynthesis enzyme |
| *S. roseochromogenes* (a) *m878_11425* | AAAACTGACTGATCGGTTTT | 175 | A-factor biosynthesis enzyme |
| *S. roseochromogenes* (a) *m878_11395* | CAAACCAGACGATCGGTTTT | 223 | TetR family transcriptional repressor |
| *S. roseochromogenes* (a) *m878_11390* | AAAACCGATCGTCTGGTTTG | 57 | StrR-like regulatory protein |
| *S. roseochromogenes* (a) *m878_11425* | CAAACAAACCATCCAGTTGT | 95 | A-factor biosynthesis enzyme |
| *S. roseochromogenes* (a) *m878_11420* | ACAACTGGATGGTTTGTTTG | 138 | A-factor biosynthesis enzyme |
| *S. scabiei* (a) *scab_12101* | AATAAACCAACCGTGCGGTATATT | 68 | TetR family transcriptional repressor |
| *S. scabiei* (a) *scab_12111* | AATATACCGCACGGTTGGTTTATT | 74 | Transposase |
| *S. scabiei* (a) *scab_12081* | ACAATACAGGTCAGCCGGTTTTTT | 53 | A-factor biosynthesis enzyme |
| *S. scabiei* (a) *scab_12091* | AAAAAACCGGCTGACCTGTATTGT | 100 | TetR family transcriptional repressor |
| *S. scabiei* (a) *scab_12101* | AGAAAACCAACCGTCCGGTATTGT | 142 | TetR family transcriptional repressor |
| *S. scabiei* (a) *scab_12111* | ACAATACCGGACGGTTGGTTTTCT | 0 | Transposase |

| | | | |
|---|---|---|---|
| *S. sviceus* (a) *sseg_05943* | AAAAAGATGACATAGGTTT | 122 | DNA-binding protein |
| *S. sviceus* (a) *sseg_05944* | AAACCTATGTCATCTTTTT | -56 | Hypothetical protein |
| *S. sviceus* (a) *sseg_11168* | AAACGGATGTGATCGTTGT | 22 | Monooxygenase, FAD-binding |
| *S. sviceus* (a) *sseg_02730* | AAACCGATGAGTAGGTTTT | 37 | Phosphate regulon regulatory protein |
| *S. sviceus* (a) *sseg_02731* | AAAACCTACTCATCGGTTT | 521 | TetR family transcriptional repressor |
| *S. sviceus* (a) *sseg_02728* | ACACCCATGTCAGCAGTCT | 612 | Acetyl-CoA carboxylase, biotin carboxylase |
| *S. sviceus* (a) *sseg_02730* | ACACCCATGTCAGCAGTCT | 677 | Phosphate regulon regulatory protein |

Putative transcriptional activators are highlighted in green.

### 4.3. Exploitation of putative signalling systems

To our current knowledge, the putative GBL-dependent signalling systems have not been investigated and the exploitation of regulation systems is expected to unlock the production of novel natural compounds.

In 7 of the GBL-dependent gene clusters, the MEME enrichment analysis successfully identified ARE sequences upstream of putative transcriptional activators (highlighted in green in Table 4-4). The specific overexpression of pathway-specific transcriptional activator is expected to increase expression of neighbouring cryptic biosynthetic gene clusters. Furthermore, in the partially characterised GBL-dependent regulatory systems of SCB/ScbR/ScbR2 in *S. coelicolor* and SVB/JadR3/JadR2 in *S. venezuelae*, the deletion of specific transcriptional repressors have resulted in the production of compounds from cryptic biosynthetic

gene clusters; for instance, the deletion of *scbR2* and *jadR2* "pseudo" GBL receptors in *S. coelicolor* and *S. venezuelae* unlocked the production of coelimycin and jadomycin respectively. [121,148] To identify specific TetR repressor to inactivate, a phylogenetic analysis of all of the TetR repressors in the AHFCA/GBL dependent signalling systems were carried out (Figure 4-5).

The phylogenetic tree reveals 4 main clades of receptors (Figure 4-5). Other "smaller" clades were also observed in the phylogenetic tree (Figure 4-5) and were primary found divergently adjacent to biosynthetic genes; the exact biological function or the cognate ligand for these TetR repressors are not known.

The "cognate AHFCA binding receptor" clade consists of the analogous MmfR repressors and have been experimentally confirmed to bind to AHFCA signalling molecules (Malet, N. Zhou, S., Styles, K. and Fullwood, A. Personal communication). In the ""pseudo" AHFCA binding receptor" clade, which includes the paralogous MmyR repressors, *in silico* analysis and early investigations have shown the MmyR repressors do not bind to the same AHFCA signalling molecules (Styles, K. Personal communication).

The deletion of "psuedo" AHFCA binding receptors in *S. coelicolor* and *S. venezuelae* resulted in the overproduction of methylenomycin and gaburedins respectively and hence, the deletion of MmyR orthologues is also expected to result in a similar phenotype in other cryptic biosynthetic gene clusters (Figure 4-5). [57,62]

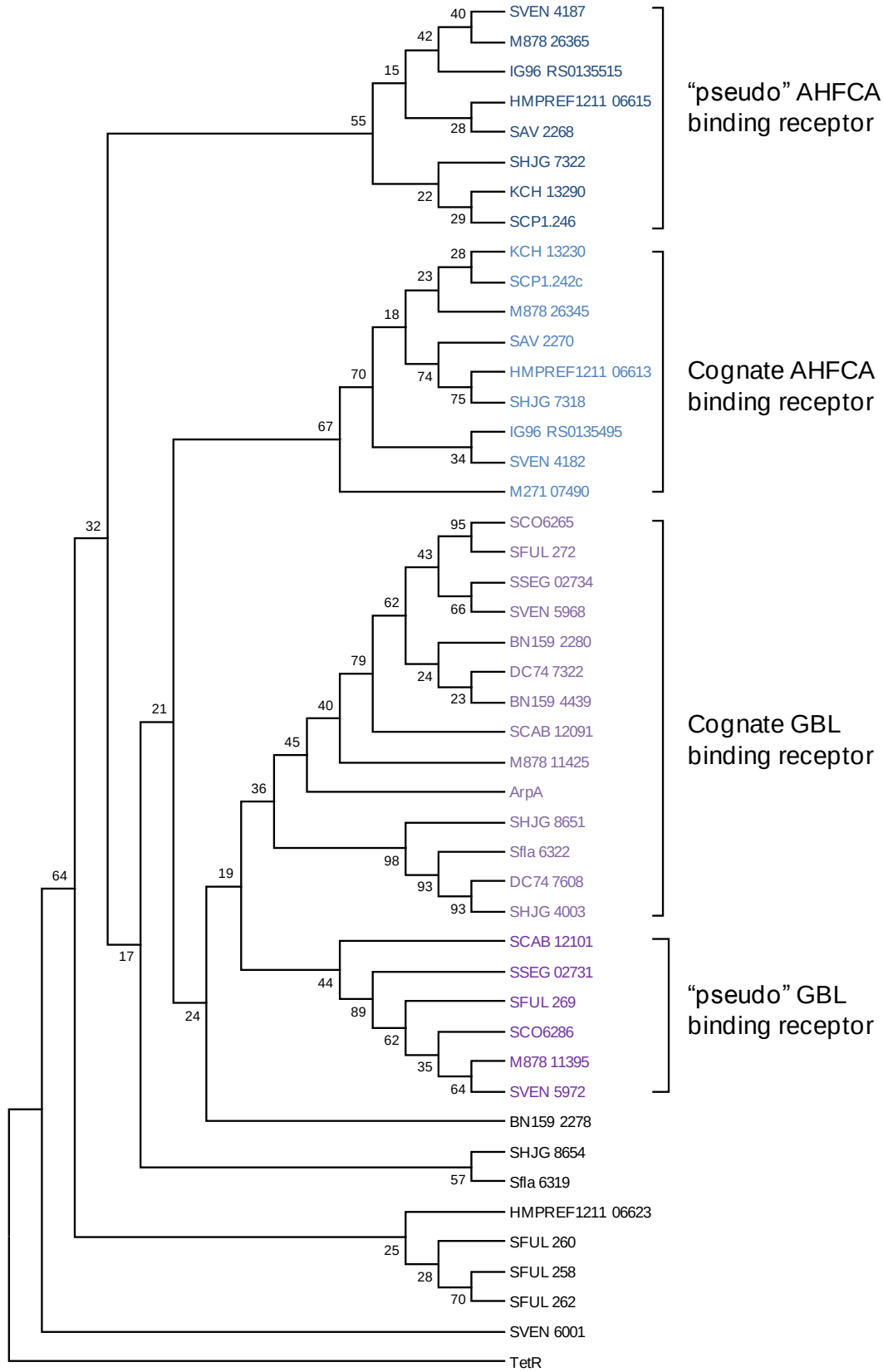



**Figure 4-5** – Phylogenetic analysis reveals the TetR repressors are clustering into four clades. TetR repressor protein sequences were aligned using MUSCLE and phylogenetic tree was generated using maximum likelihood method. Bootstrap consensus tree was inferred from 1000 iterations.

Intriguingly, the M271_07490 repressor from *S. rapamycinicus* also falls into the "cognate AHFCA binding receptor" clade (Figure 4-5); this implies that the M271_07490 repressor share similar properties with MmfR and may interact with AHFCA-like signalling molecules. To determine whether M271_07490 could bind to AHFCA-like signalling molecules, the key amino acid residues required for binding with the cognate ligand were compared with analogous MmfR repressors. Co-crystallisation of MmfR with MMF2 has revealed Y85 and Y144 are directly involved in hydrogen bonding with the cognate ligand (Figure 4-6) and these amino acid residues are not present in the analogous MmyR repressors (Dean., R. *et al*. Manuscript in preparation).

All of the MmfR analogues contain the conserved Y85 and Y144 residue, except for SHJG_7318, which has a phenylalanine residue instead of tyrosine at position 144. The M271_07490 repressor contains phenylalanine in both of the key amino acid positions and suggests M271_07490 may also bind to AHFCA-like signalling molecules. Of particular interest is that the cryptic biosynthetic gene cluster in *S. rapamycinicus* does not contain analogous *mmfH* or *mmfP* genes; therefore, the signalling system in *S. rapamycinicus* is proposed to be dependent on a novel class of signalling molecule. Alternatively, the cryptic biosynthetic gene cluster could cross-talk with other *Streptomyces* bacteria that produce AHFCA signalling molecules to coordinate the biosynthesis of specialised metabolites.

```
M271_07490            VVEEHFATWPPLIERF-20-AFRDDIVVRAGARLWAERTLIEAPMPPPFVGWID
KCH_13230             VVEALYATWPRTLEAV-20-AFLNDPIMQAGTRLQNERDVIDVELPLPYVDWTT
SCP1.242c             VVEEHYARWPAAMEEI-20-AFRDDPVMQAGARLQSERAFIDAELPLPYVDWTH
M878_26345            VVEAHYSRWPELLEGI-20-AFATDIVVQAGARLQLERSLIDAQLPQPYVGWTQ
SAV_2270              VLEEFYRRMQEAVNGA-21-AFHEDVFIHAGARLQIERPYIKAELPVPYVGTLK
SHJG_7318             VTTEFYKRLPAISDTV-20-ALRDEPMMKAGARLQIERGLIDADLPIPFQDYTE
HMPREF1211_06613      VAEEFYHVLGVIGQEV-20-AFRDDTMVQAGARLQIERSLIGAEMPMPYLGYTE
IG96_RS0135495        VIMEHYARWEPLVSEV-20-AFRDDVMVQAGARLQIERSLIKADLPVPYVGWQE
SVEN_4182             VVQEHYARWPEILKGA-20-AFARDIVVQAGARLQIERALIDAELPEPYVGWED
                      *    :       .      *:  : .::**:**  **  * . :* *:
```

**Figure 4-6** – Clustal Omega alignment of the cognate AHFCA binding receptor clade. Residues proposed to be involved in hydrogen bond with the cognate AHFCA ligand are highlighted in pale red.

There are also two separate clades for the GBL binding receptors (Figure 4-5). Based on the TetR repressors in jadomycin and coelimycin biosynthetic gene clusters, the ""real" GBL binding receptor" were reported to bind to GBL signalling molecules, whilst the ""pseudo" GBL binding receptor" do not bind to the same GBL molecules, but to structurally diverse metabolites, such as actinorhodin, undecylprodigiosin and jadomycin.[98,99,117,139] In addition, the deletion of *sco6286*, which encodes for the ScbR2 "pseudo" binding receptor, resulted in the production of coelimycin in *S. coelicolor*.[115,121] Therefore, the deletion of these specific "pseudo" GBL receptors in putative GBL-dependent regulatory systems is expected to unlock the production of compounds from cryptic biosynthetic gene clusters.

Some of the putative GBL-dependent signalling systems contain a varied number of TetR repressor genes; for example, the putative GBL-dependent signalling systems in *S. albulus*, *S. davawensis* and *S. rapamycinicus* appears to be under the control of a single TetR repressor whilst *S. fulvissimus* contains five TetR repressors. From the predictions of ARE sequences, the transcriptional regulation of these cryptic biosynthetic gene clusters could function differently and novel mechanisms could have evolved.

## 4.4. Conclusions

The GBL-dependent signalling systems are widely distributed in *Streptomyces* bacteria; in fact, the genome analysis of 20 sequenced *Streptomyces* bacteria identified 36 cryptic biosynthetic gene clusters proposed to be under the control of putative GBL-dependent signalling systems. To test the effectiveness of the MEME enrichment analysis at discovering TetR binding sites, the bioinformatic analysis was performed on 36 gene clusters.

From the 36 gene clusters, the MEME enrichment analysis was successful at discovering the ARE sequences for 18 biosynthetic gene clusters. The ARE sequences were identified in the promoters of putative transcriptional activators and TetR repressors; the deletion of specific "pseudo" ArpA-like transcriptional repressors or overexpression of pathway-specific transcriptional activators is expected to unlock the production of novel compounds from cryptic biosynthetic gene clusters.

The SCB/ScbR/ScbR2 system, which is involved in regulating coelimycin biosynthesis, has been extensively studied by global transcriptomic analysis and ChIP-seq. MEME enrichment analysis was unable to identify all of the experimentally confirmed binding sites and is speculated to be due to the strict nature of the searching algorithm.

## 5. Prediction and Identification of Azoxy Compounds from *S. avermitilis*

From genome mining analyses, a number of cryptic biosynthetic gene clusters that appear to be under the control of putative AHFCA-dependent regulatory systems were identified (Figure 3-10, Appendix D to Appendix K). To our knowledge, the neighbouring 50 kb cryptic biosynthetic gene cluster in *S. avermitilis* has not been investigated (cluster #15 in Table 1-2) and exploitation of the AHFCA-dependent signalling system is expected to result in the discovery of novel natural compounds. The 50 kb cryptic biosynthetic gene cluster contains approximately 45 genes in total including 6 genes involved in regulation; the cassette of 5 genes (*sav_2266* to *sav_2270*) that forms an AHFCA signalling gene cluster and *sav_2301*, which encodes for a SARP transcriptional activator (Appendix F).

Structural prediction of cryptic metabolites would assist their identification and characterisation. Detailed gene annotations of the 50 kb cryptic biosynthetic gene cluster in *S. avermitilis* reveals several biosynthetic genes including polyketide synthases, which code for large family of multi-domain enzymes involved in producing polyketide natural products (Figure 6-1). [149]

### 5.1. Seryl-tRNA synthetase involved in natural product biosynthesis

The *sav_2294* gene product encodes for a putative seryl-tRNA synthetase (serRS). Aminoacyl-tRNA synthetases (aaRS) are enzymes that catalyse esterification of a specific amino acid to their cognate tRNA to produce an aminoacyl-tRNA and are essential substrates for protein translation. Aminoacyl-tRNAs are also substrates for the bridge formation in peptidoglycan biosynthesis of bacterial cell walls. [150] The

presence of a serRS in biosynthetic gene clusters is not common, however a few biosynthetic gene clusters with serRS orthologues have recently been identified, such as albomycin and valanimycin from *S.* sp. ATCC 700974 and *S. viridifaciens* MG456-hF10 respectively (Figure 5-1). [151,152]

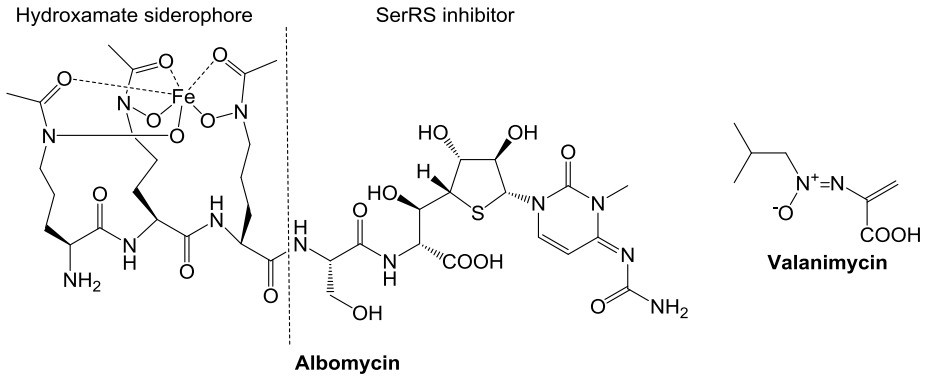



**Figure 5-1** – Structure of albomycin and valanimycin. The albomycin antibiotic consists of a serRS inhibitor linked to hydroxamate siderophore through a serine residue. VlmL serRS is actively involved in the biosynthesis of valanimycin natural compound and provides charged serine.

The genome of *S. avermitilis*, *S.* sp. ATCC 700974 and *S. viridifaciens* contains two serRS paralogs. [151,152] One of the serRS paralogs in *S.* sp. ATCC 700974 and *S. viridifaciens* encodes for a housekeeping serRS, and has high similarity and kinetic parameters with other *Streptomyces* housekeeping aaRS. [151,153] In *S. avermitilis*, both *sav_2294* and *sav_4244* genes encode for serRS and from phylogenetic analysis, the *sav_4244* is most likely the housekeeping serRS (Figure 5-2).

Paralogous serRS in *S.* sp. ATCC 700974 and *S. viridifaciens* are suggested to have different biological functions. In *S.* sp. ATCC 700974, the *alb10* gene product, which encodes for a paralogous serRS, is located within the albomycin biosynthetic gene cluster. [151] The albomycin antibiotic has a thioribosyl nucleoside linked to a siderophore through a serine residue and the thioribosyl nucleoside is proposed to

mimic a serine substrate to inhibit serRS (Figure 5-1).[154] Biochemical investigations confirmed albomycin inhibited the activity of housekeeping serRS in *S.* sp. ATCC 700974, but not of Alb10 serRS, which is proposed to confer albomycin resistance in the natural producer.[151]

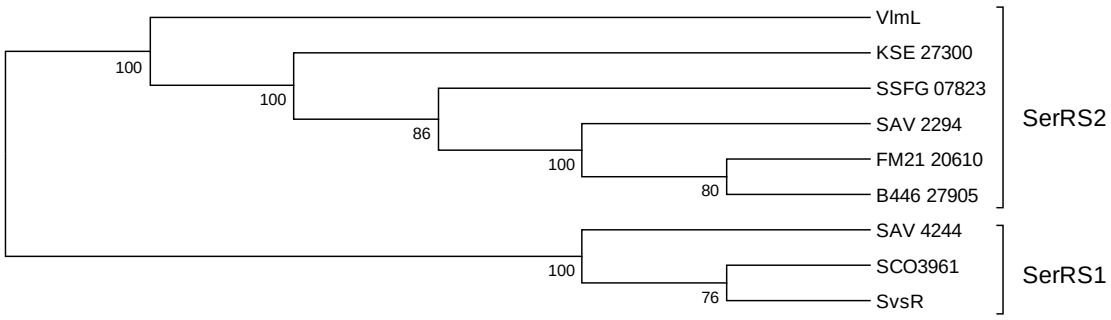



**Figure 5-2** – Phylogenetic analysis of seryl-tRNA synthetases in *Streptomyces* bacteria reveals housekeeping serRS (SerRS1) clusters separately to paralogous serRS (SerRS2). SerRS were aligned with MUSCLE and a consensus bootstrap phylogenetic tree was generated using the maximum likelihood method from 1000 iterations.

Valanimycin, which has been extensively investigated by Garg and co-workers, is derived from L-valine and L-serine *via* the intermediacy of isobutylamine and isobutylhydroxylamine (Figure 5-3).[155,156] The biological functions of 6 genes involved in valanimycin biosynthesis have been established (Figure 5-3). VlmD catalyses the decarboxylation of L-valine into isobutylamine while the flavin-dependent VlmH/VlmR two-component system reduces isobutylamine to isobutylhydroxylamine (Figure 5-3).[157,158,159] The unusual involvement of VlmL serRS was found to charge L-serine to the cognate tRNA and the seryl residue is then transferred to the hydroxyl group of isobutylhydroxylamine by VlmA (Figure 5-3).[152] Currently, the mechanism for the formation of the azoxy moiety remains unclear and possible mechanisms are discussed in reference 152. The final

biosynthetic step involves the dehydration of valanimycin hydrate to yield valanimycin. Two enzymes are responsible for catalysing the phosphorylation (VlmJ), which then favours the elimination reaction (VlmK) (Figure 5-3). [160]

VlmL serRS is therefore directly involved in the biosynthesis of a specialised metabolite (Figure 5-3). Interestingly, the Alb10 serRS, which has demonstrated to confer resistance to albomycin, is also speculated to be involved in peptide formation between the serRS inhibitor moiety and hydroxamate siderophore (Figure 5-1). [151,152,153] The SAV_2294 serRS in *S. avermitilis*, which shares 59% identity and 70% similarity over 427 amino acids with VlmL, is also predicted to participate in the biosynthesis of cryptic metabolite(s).

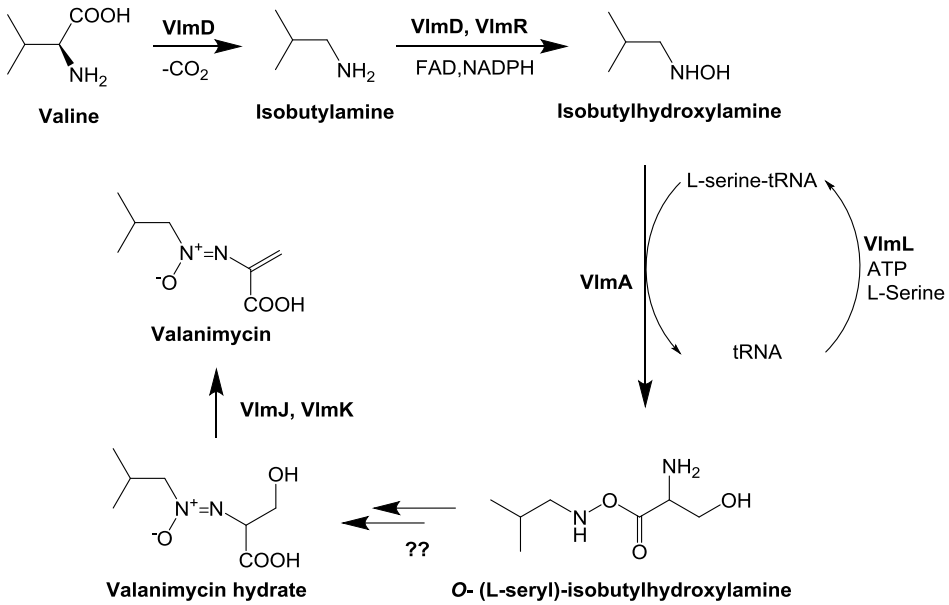



**Figure 5-3** – Proposed valanimycin biosynthetic pathway. Valine is converted into valanimycin through the intermediacy of isobutylamine and isobutylhydroxamine intermediates. VlmL provides charged serine substrate for valanimycin biosynthesis. Adapted from reference 161.

## 5.2. Prediction of novel azoxy natural compound produced by the *S. avermitilis* cryptic biosynthetic gene cluster

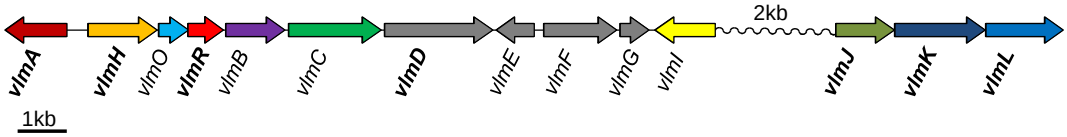
The valanimycin biosynthetic gene cluster contains a total of 14 genes including resistance and regulatory genes (Figure 5-4).[162] The *vlmE* gene encodes for a TetR transcriptional repressor and *vlmI* encodes for a SARP transcriptional activator.[162] The *vlmE* gene is divergently adjacent to the *vlmF* valanimycin resistance transporter (Figure 5-4) and similar to the tetracycline resistance system in *E. coli*, VlmE, most likely, regulates the expression of *vlmF*. From genetic and complementation studies, VlmI is essential for valanimycin biosynthesis and expression of valanimycin biosynthetic genes are under the positive control of VlmI.[161]

The biological functions of 7 biosynthetic genes in the valanimycin gene cluster have been experimentally validated (Figure 5-3 and Figure 5-4). All of these genes except *vlmD* were also identified in the *S. avermitilis* 50 kb cryptic biosynthetic gene cluster (Figure 5-3 and Figure 5-4), suggesting this cluster directs the biosynthesis of azoxy-containing compound derivatives. No orthologues of VlmD, which provides the isobutylamine precursor in valanimycin biosynthesis, is present in *S. avermitilis* (Figure 5-3). A different precursor is therefore proposed to be incorporated into the *S. avermitilis* azoxy compounds. In addition, the *S. avermitilis* biosynthetic gene cluster is significantly larger than that of valanimycin and contains a number of genes involved in polyketide biosynthesis (Figure 5-4). A polyketide precursor (or possibly another amino acid) could therefore be incorporated into the azoxy metabolites in *S. avermitilis* (Figure 5-5). Homologues of all of the key enzymes involved in formation of the dehydroalanine moiety in valanimycin are present in

*S. avermitilis* and is expected to be intact in the final predicted azoxy compounds (Figure 5-5).

a)

***Streptomyces viridifaciens* MG456-hF10**
**Valanimycin biosynthetic gene cluster (17,496 bp)**




| Gene in *S. viridifaciens* | Analogous gene in *S. avermitilis* | Proposed Function |
|---|---|---|
| ***vlmA*** | ***sav_2300*** | **O-seryl acyltransferase** |
| ***vlmH*** | ***sav_2284*** | **Isobutylamine N-hydroxylase** |
| *vlmO* | *sav_2287* | Putative integral membrane protein |
| ***vlmR*** | ***sav_2288*** | **NADPH-flavin oxidoreductase** |
| *vlmB* | *sav_2289* | Hypothetical protein |
| *vlmC* | *sav_2293* | Amino acid permease |
| ***vlmD*** | | **Valine decarboxylase** |
| *vlmE* | | TetR family transcriptional repressor |
| *vlmF* | | Valanimycin resistance |
| *vlmG* | | Hypothetical protein |
| *vlmI* | *sav_2301* | SARP family transcriptional regulator |
| ***vlmJ*** | ***sav_2272*** | **Diacylglycerol kinase** |
| ***vlmK*** | ***sav_2272*** | **MmgE/PrpD family protein** |
| ***vlmL*** | ***sav_2294*** | **Seryl-tRNA synthetase** |

b)

***Streptomyces avermitilis* MA-4680**
**(50, 462 bp)**

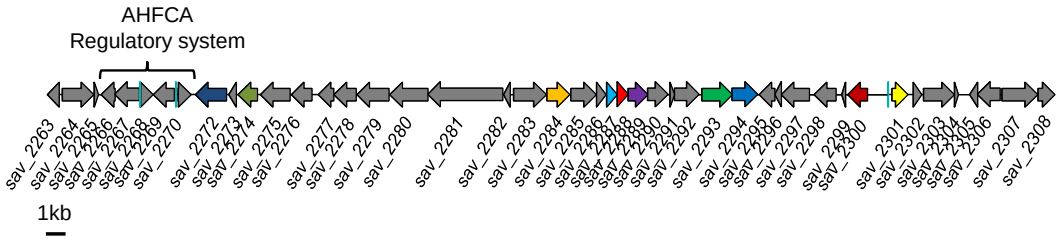
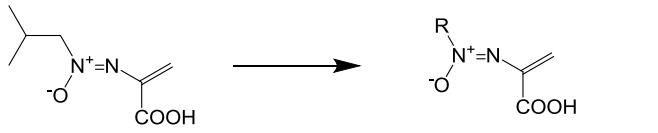



**Figure 5-4** – The valanimycin biosynthetic gene cluster contains 14 genes involved in biosynthesis, regulation and resistance. (a) The functions of valanimycin biosynthetic genes have been established through biochemical studies are highlighted in bold. (b) Nearly all of the valanimycin biosynthetic genes are present in the *S. avermitilis* biosynthetic gene cluster. Analogous genes are colour-coded.




Valanimycin Predicted azoxy core structure

**Figure 5-5** – Proposed core azoxy structure produced by the predicted azoxy biosynthetic gene cluster in *S. avermitilis*. Due to absence of valine decarboxylase enzyme in *S. avermitilis*, azoxy natural compounds in *S. avermitilis* are speculated to incorporate a different precursor, possibly of polyketide origin.

## 5.3. Capture of the entire azoxy biosynthetic gene cluster into PAC vectors

Compared to the model organism, *S. coelicolor* A3(2), genetic manipulations in *S. avermitilis* are relatively more difficult. For instance, transformation is primarily carried out *via* protoplast rather than intergenic conjugation. [163] Cloning entire biosynthetic gene clusters into well characterised heterologous hosts or superhost, such as *S. lividans* TK24 or *S. coelicolor* M1152 respectively, does not only make the genetic work easier but also improves yield and facilitate identification/ purification processes. [164,165,166] Genomic DNA libraries, which are a collection of DNA fragments captured onto a specific vector, have been routinely used to acquire biosynthetic gene clusters for heterologous expression. [167] Several cloning vector systems are available depending on the size of DNA fragment required and cloning purpose (Table 5-1).

**Table 5-1** – Cloning vectors for genomic DNA library.

| **Vector** | **Host** | **Typical insert size (kb)** |
|---|---|---|
| Plasmids | *E. coli* | Up to 10 |
| Cosmids/ Fosmids | *E. coli* | Up to 45 |
| Bacterial artificial chromosome | *E. coli* | 100 to 300 |
| P1-derived artificial chromosome | *E. coli* | 100 to 300 |
| Yeast artificial chromosome | *Saccharomyces cerevisiae* | 100 to 2,000 |

It is difficult to determine the boundaries of the predicted azoxy biosynthetic gene cluster in *S. avermitilis*. With the identification of ARE sequences on the opposite ends of the biosynthetic gene cluster, the azoxy biosynthetic gene cluster is proposed to be at least 50 kb (Appendix F). Therefore, plasmid or modified λ phage vectors, such as cosmid or fosmids, are not appropriate to capture the 50 kb gene cluster and artifical chromosomes are more suitable. Examples of artificial chromosomes include the bacterial artificial chromosomes (BAC) and P1-derived artificial chromosomes (PAC), which are based on the replicon of the F factor fertility plasmid and bacteriophage P1 respectively. [168,169] Sosio and co-workers were able to produce an *E. coli* - *Streptomyces* artificial chromosomes (ESAC) shuttle vector by introducing the ΦC31 *attP-int* system into pCYPAC2 to generate a non-conjugative pPAC-S1 vector. [170] The pPAC-S1 vector could replicate autonomously in *E. coli*, as well as allow direct site-specific integration into *Streptomyces* bacterial genome. The pPAC-S1 vector was further developed by introducing *oriT*, from RK2 replicon, to create a conjugative pESAC13 vector (Figure 5-6), which is currently licensed to Bio S&T, and is able to produce pESAC13A PAC genomic libraries with insert sizes of approximately 130 kb. [171]

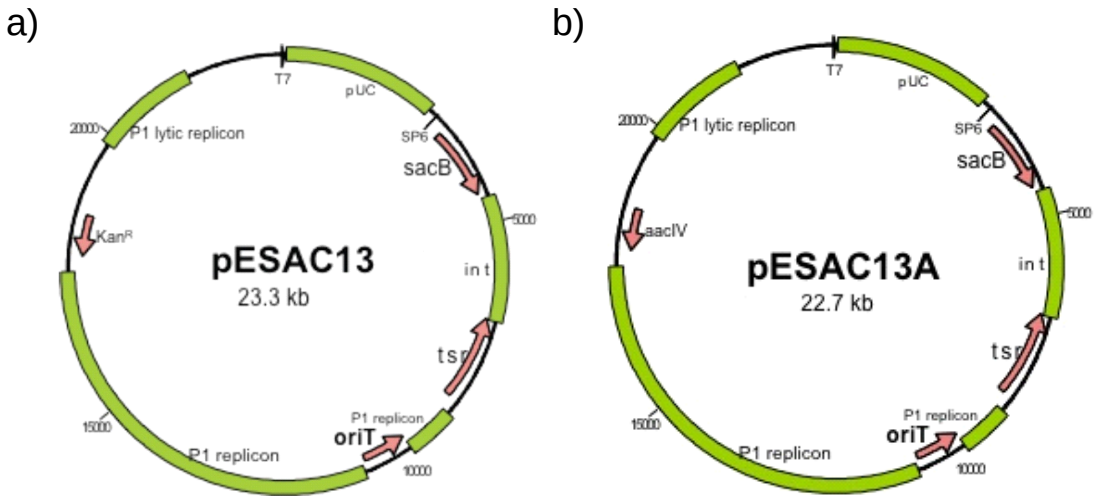



**Figure 5-6** – The pESAC13 and pESAC13A PAC vector. Kanamycin resistance gene in (a) pESAC13 was replaced with an apramycin resistance gene to generate the (b) pESAC13A PAC vector. Image obtained from Bio S&T (www.biost.com).

The *S. avermitilis* pESAC13A PAC genomic DNA library was used to capture the putative azoxy gene cluster. An apramycin resistant pESAC13A PAC genomic library of *S. avermitilis* MA-4680 was constructed for the project by Bio S&T and a total of 1,920 individual clones of *E. coli* DH10B each containing a single pESAC13A were obtained (Figure 5-6). Bio S&T also screened the pESAC13A genomic library for the biosynthetic gene cluster of interest (Figure 5-7).

Three pairs of primers were designed to bind to the left, center and right of the predicted azoxy biosynthetic gene cluster (for_pESAC_SAV2283a and rev_pESAC_SAV2283a; for_pESAC13_sav2266a and rev_pESAC13_sav2266a; for_pESAC_sav2301a and rev_pESAC_sav2301a, see Table 2-3). The primers were first tested on genomic DNA and then the pESAC13A constructs for the predicted azoxy biosynthetic gene cluster. Two individual constructs, named pESAC13A-2H and pESAC13A-19K, were positive for all three pairs of primers (Figure 5-7). The predicted azoxy biosynthetic gene cluster is present in pESAC13A-2H and

pESAC13A-19K constructs and contains approximately 130 kb and 135 kb insert sizes respectively as determined by restriction digests (Figure 5-7); however, the extremities of the cloned DNA fragment were not determined and sequencing of pESAC13A constructs would be required to define the fragment ends.

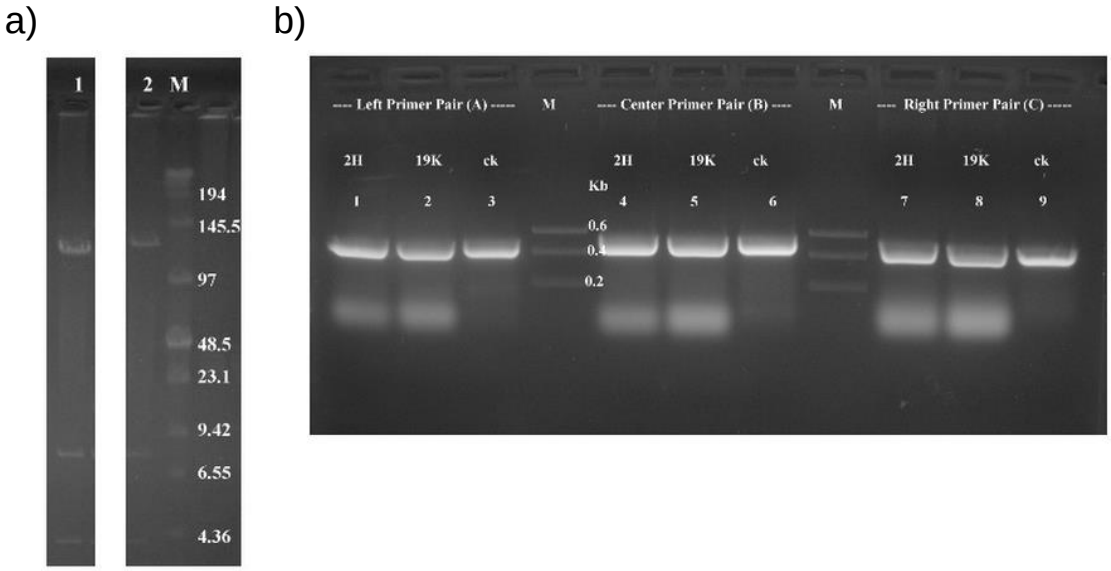



**Figure 5-7** – PCR identified two pESAC13A constructs containing the entire predicted azoxy biosynthetic gene cluster. (a) The pESAC13A-2H (Lane 1) and pESAC13A-19K (Lane 2) contain an approximate insert size of 130 kb and 135 kb respectively. (b) The pESAC13A-2H (Lanes 1, 4 and 7) and pESAC13A-19K (Lanes 2, 5, 8) constructs were positive for left (*sav_2283*), center (*sav_2266*) and right (*sav_2301*) primers. *S. avermitilis* genomic DNA (Lane 3, 6 and 9) was used as a positive control. Isolation and screening of colonies containing the pESAC13A construct of interest was performed by Bio S&T.

From bioinformatic and genome analysis, the genome of *S. avermitilis* is predicted to produce 38 specialised metabolites and as of 2014, 16 compounds have been detected (Table 1-2). [35] Putative biosynthetic gene clusters in *S. avermitilis* are all under 130 kb (Table 1-2) and the pESAC13A PAC genomic library could be further exploited to find other cryptic biosynthetic gene clusters in *S. avermitilis*; identification of products from these cryptic biosynthetic gene clusters could be

carried out using a heterologous expression/comparative metabolic profiling approach.

## 5.4. Heterologous expression of the predicted azoxy biosynthetic gene cluster from *S. avermitilis*

The positive pESAC13A constructs were introduced into two different *Streptomyces* heterologous hosts; *S. coelicolor* M1152 and *S. lividans* TK24. *S. coelicolor* M1152 (Δ*act*, Δ*red*, Δ*cpk*, Δ*cda*, *rpoB*[C1298T]) is an engineered strain designed for heterologous expression of metabolite gene clusters. Most importantly, *S. coelicolor* M1152 is based on *S. coelicolor* M145 (SCP1$^-$, SCP2$^-$) and the actinorhodin, prodiginines, coelimycin and CDA biosynthetic pathways have been deleted; in consequence primary precursors and more energy are available for heterologous metabolic pathways. [166] *S. coelicolor* M1152 also contains a single amino acid mutation in RpoB, which encodes for a RNA polymerase β-subunit, and mutations in RpoB were shown to enhance the level of metabolite production. [166,172] *S. lividans* TK24 (SLP2$^-$, SLP3$^-$) is a derivative of *S. lividans* 66, which had its genome sequenced in 2013, and is a closely related species to *S. coelicolor*. [173]

### 5.4.1. Introduction of pESAC13A-2H and pESAC13A-19K into *S. lividans* TK24 by tri-parental conjugation

The advantage of using *S. lividans* as a heterologous hosts is that it does not contain a methyl-sensing restriction system; therefore, methylated DNA from *E. coli* can be directly introduced into *S. lividans* TK24 without the need to passage the vector through a non-methylating *E. coli* host (Figure 5-8). [21]

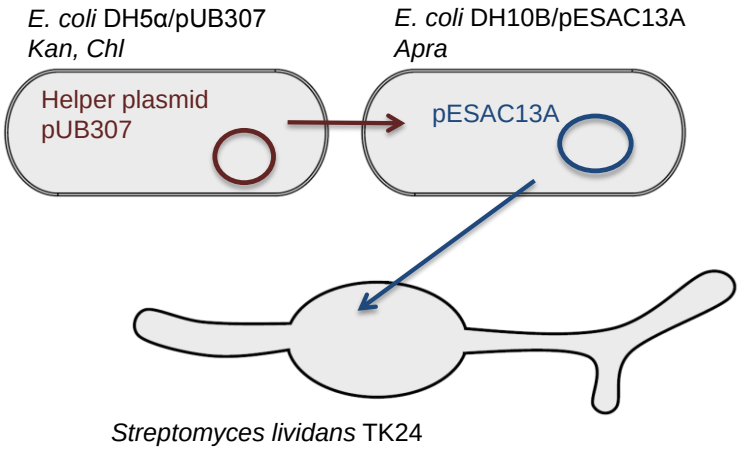



**Figure 5-8** – Schematic of tri-parental conjugation in *S. lividans* TK24. The pUB307 helper plasmid from *E. coli* DH5α is mobilised into *E. coli* DH10B containing pESAC13A and transfers the pESAC13A constructs into *S. lividans* TK24.

Tri-parental conjugation was carried out between *E. coli* DH5α/pUB307, *E. coli* DH10B containing pESAC13A constructs and *S. lividans* TK24 to mobilise pESAC13A-2H and pESAC13A-19K into *S. lividans* heterologous hosts; two colonies of *S. lividans* TK24/pESAC13A-2H and *S. lividans* TK24/pESAC13A-19K were obtained (Figure 5-9). The successful integration of pESAC13A-2H and pESAC13A-19K in *S. lividans* heterologous host strains were confirmed by PCR using the same primer pairs that were used to identify pESAC13A-2H and pESAC13A-19K (Figure 5-10).

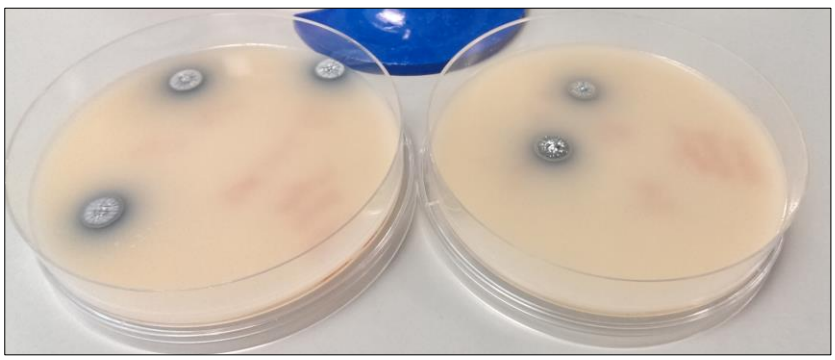


**Figure 5-9** – Successful integration of apramycin resistant pESAC13A-2H (left) and pESAC13A-19K (right) constructs in *S. lividans* TK24.

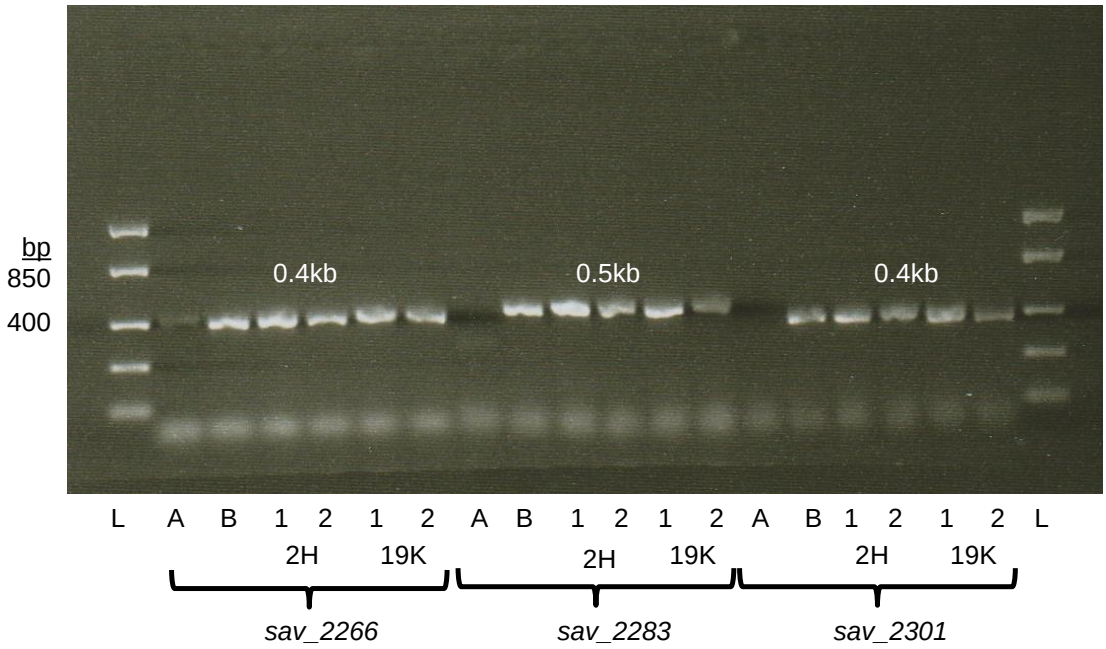



**Figure 5-10** – PCR confirmed the integration of pESAC13A-2H and pESAC13A-19K constructs in *S. lividans* heterologous host strains. The primers pairs (*sav_2266*, *sav_2283* and *sav_2301*) were used and a band corresponding to the expected product of 0.4 - 0.5 kb were obtained in all colonies. *S. lividans* TK24 genomic DNA was used as a negative control (Lane A) and *S. avermitilis* genomic DNA was used as a positive control (Lane B).

### 5.4.2. Introduction of pESAC13A-2H and pESAC13A-19K into *S. coelicolor* M1152 by tri-parental conjugation

To introduce DNA into *S. coelicolor*, DNA needs to be passaged into a non-methylating DNA host, such as *E. coli* ET12567. [21] Firstly, tri-parental conjugation was carried out between *E. coli* DH5α/pR9604, *E. coli* ET12567 and the *E. coli* DH10B containing the pESAC13A constructs to introduce the self transmissible pR9604 plasmid and pESAC13A constructs into non-methylating *E. coli* ET12567 hosts (Figure 5-11). *E. coli* ET12567/pR9604/pESAC13A-2H and *E. coli* ET12567/pR9604/pESAC13A-19K exconjugants were screened by PCR using the same three primer pairs as in the previous section (Figure 5-12).

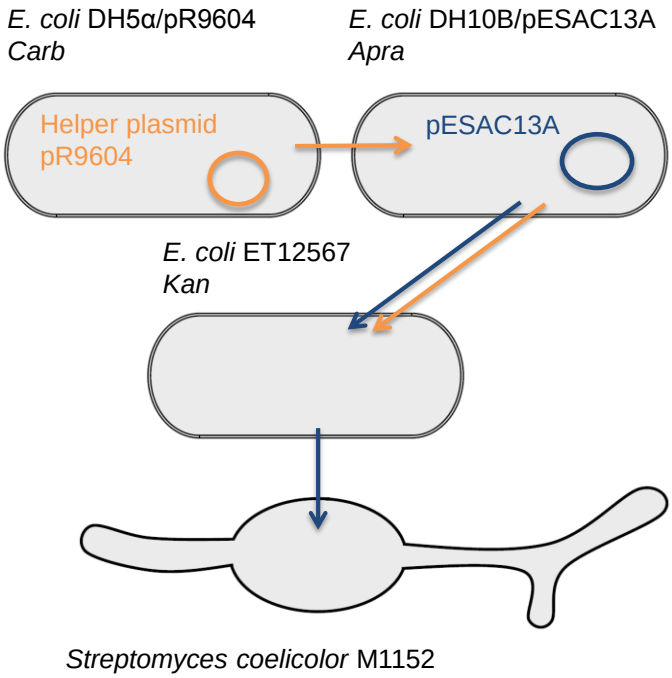



**Figure 5-11** – Schematic of tri-parental conjugation in *S. coelicolor* M1152. The pR9604 helper plasmid is transferred into *E. coli* DH10B containing the pESAC13A constructs. The pR9604 plasmid mobilises the pESAC13A constructs into *E. coli* ET12567 resulting in a kanamycin, carbenicillin and apramycin resistant strain. Standard *E. coli*/ *Streptomyces* conjugation is carried out to transfer the pESAC13A constructs into *S. coelicolor* M1152.

After transfer of pESAC13A construct and pR9604 plasmid into non-methylating *E. coli* ET12567, intergenic conjugation was successfully carried out between *S. coelicolor* M1152 and *E. coli* ET12567/pR9604 transformants containing the pESAC13A constructs; two colonies of *S. coelicolor* M1152/pESAC13A-2H and *S. coelicolor* M1152/pESAC13A-19K were obtained. Successful integration of pESAC13A-2H and pESAC13A-19K constructs into *S. coelicolor* heterologous hosts were confirmed by PCR screening (Figure 5-13). No noticeable phenotype change was observed in *S. coelicolor* heterologous host strains when compared to the wild type *S. coelicolor* M1152.

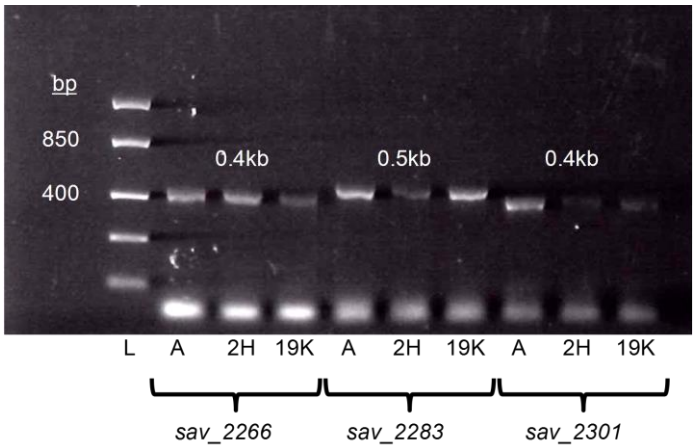



**Figure 5-12** – PCR confirmed the transfer of pESAC13A-2H and pESAC13A-19K constructs into *E. coli* ET12567. The 3 primer pairs (*sav_2266*, *sav_2283* and *sav_2301*) were used and the band corresponding to the expected product of 0.4 - 0.5 kb were obtained in all of the colonies. *S. avermitilis* genomic DNA was used as a positive control (Lane A).

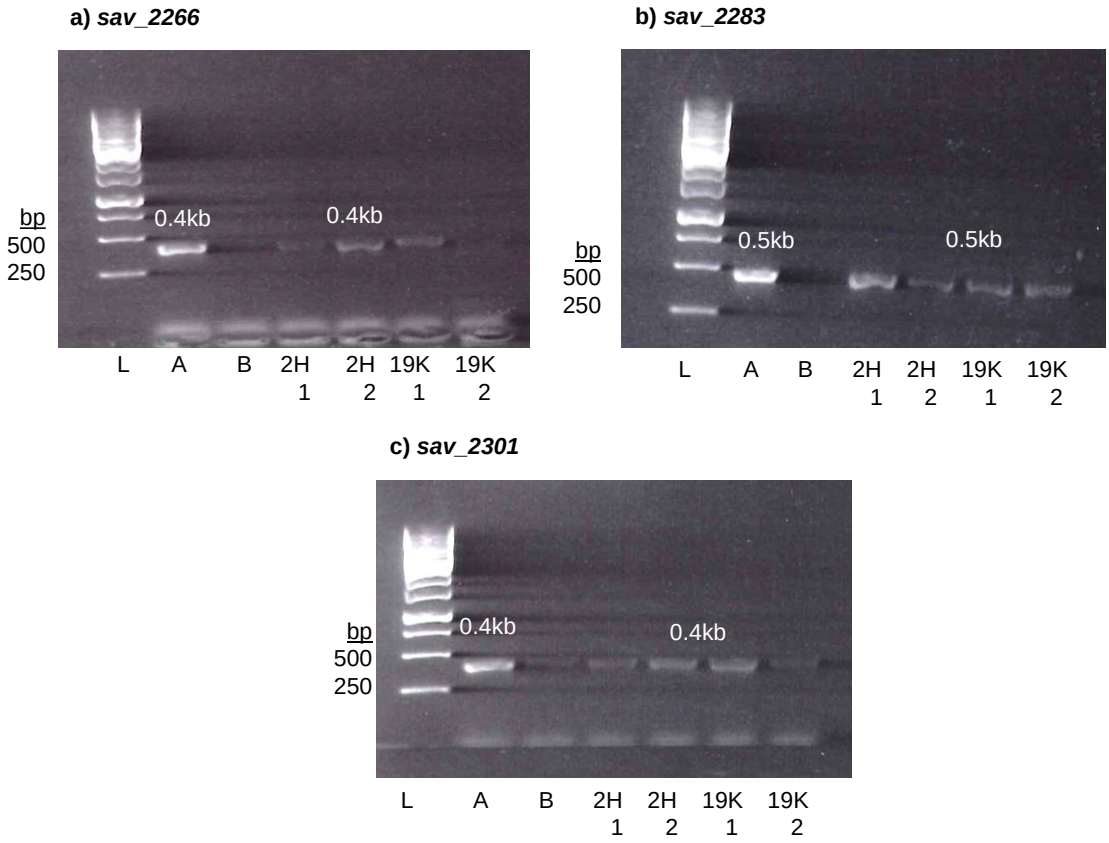



**Figure 5-13** – PCR confirmed the integration of pESAC13A-2H and pESAC13A-19K constructs in *S. coelicolor* heterologous host strains. The 3 primers pairs (*sav_2266*, *sav_2283* and *sav_2301*) were used and the band corresponding to the expected product of 0.4 - 0.5 kb were obtained in all of the colonies. *S. avermitilis* genomic DNA was used as a positive control (Lane A) and *S. coelicolor* M1152 genomic DNA was used as a negative control (Lane B).

## 5.5. Comparitive metabolic profiling of *Streptomyces* heterologous hosts grown on SMMS agar

Comparative metabolic profiling of *S. coelicolor* and *S. lividans* heterologous host strains by LC-MS analysis was expected to reveal the presence of novel metabolites resulting from the integration of pESAC13A-2H or pESAC13A-19K into the chromosome. [166]

### 5.5.1. Comparative metabolic profiling of *S. lividans* heterologous hosts

Metabolites from *S. lividans* heterologous host strains containing the pESAC13A constructs were compared against metabolites produced by the wild type *S. lividans* TK24 by LC-MS. LC-MS analysis of both the neutral and acidified organic extraction did not reveal any compounds produced due to the introduction of the pESAC13A constructs. Absence of compounds of interest may be due to poor precursor supply in *S. lividans* TK24 or active transcriptional regulation of biosynthetic genes in *S. lividans* heterologous hosts.

### 5.5.2. Identification of predicted azoxy compounds in *S. coelicolor* heterologous hosts

Similarly, metabolites between the *S. coelicolor* heterologous host strains and wild type *S. coelicolor* M1152 were compared by LC-MS. Three new compounds (**1** - **3**) were identified in the acidified *S. coelicolor* organic extracts (Figure 5-14). The *m/z* values for compounds **1** - **3** were 219.0, 201.1 and 201.1 respectively and were only present in *S. coelicolor* strains containing pESAC13A-2H or pESAC13A-19K and were absent in *S. coelicolor* M1152 control.

Due to the nitrogen rule, molecular ion masses of compounds **1** - **3** indicates they contain an even number of nitrogen atoms, which was confirmed from the molecular formulae generated by UHPLC-ESI-TOF-MS data (Table 5-2); compound **1** has a molecular formula of $C_9H_{19}N_2O_4$ and both compounds **2** and **3** have a molecular formula of $C_9H_{17}N_2O_3$.

The molecular formulae therefore imply that the observed compounds **1** - **3** are related and similar to the valanimycin compound, which has a molecular formula of $C_7H_{12}N_2O_3$ and an *m/z* value of 173.1 (Figure 5-15).[155,161] Compared to valanimycin, compound **1** has an additional $C_2H_6O$ whilst compounds **2** and **3** have an extra $C_2H_4$. Of particular interest are compounds **2** and **3**, which have the exact same molecular formulae and are likely to be isomers (Table 5-2). A mass difference of 18 Da, which corresponds to a loss of water, was also observed from **1** to **2** and **3**.

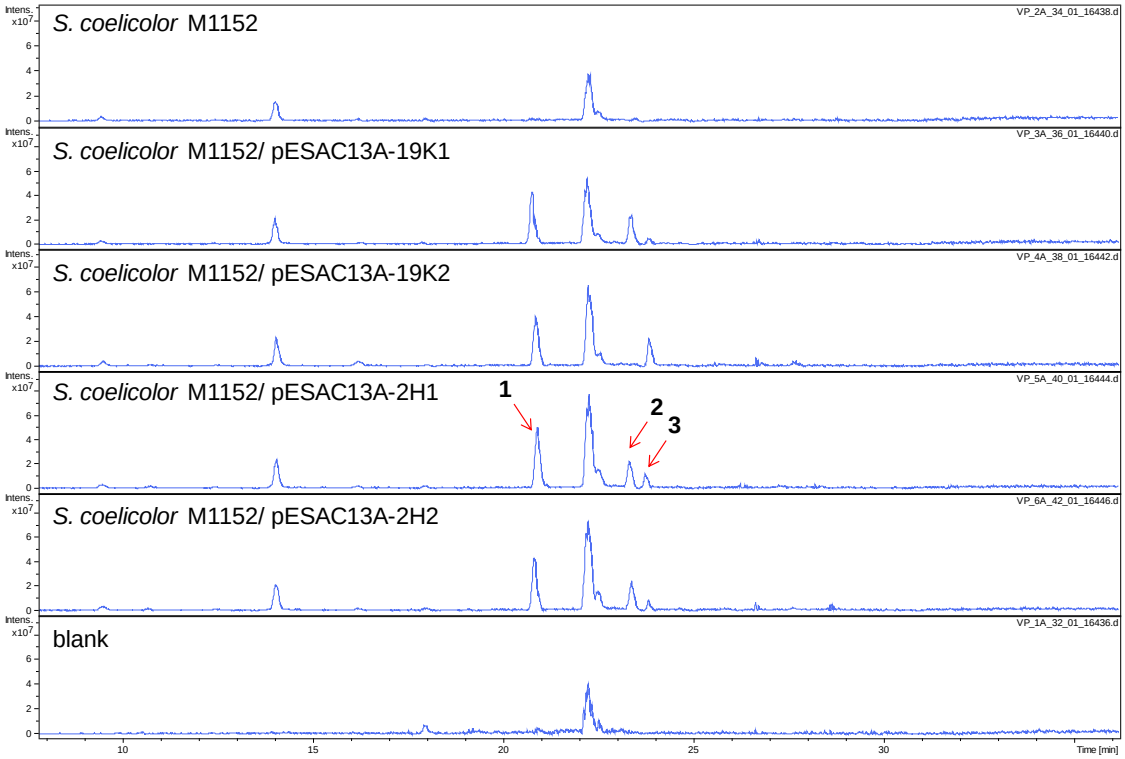



**Figure 5-14** – Predicted azoxy compounds (**1** - **3**) were identified in the acidified *S. coelicolor* organic extracts by LC-MS. Extracted ion chromatograms (*m/z* = 219.0 and 201.0) detected compounds **1** - **3** only in *S. coelicolor* M1152 containing pESAC13A-2H or pESAC13A-19K and not in *S. coelicolor* M1152 control or blank.

**Table 5-2** – Compounds detected in the acidified organic extraction of *S. coelicolor* heterologous host strains grown on SMMS agar.

| # | Retention Time (min) | $[M+H]^+$ | | | $[M+Na]^+$ | | |
|---|---|---|---|---|---|---|---|
| | | Formula | Observed | Calculated | Formula | Observed | Calculated |
| 1 | 20.0 | $C_9H_{19}N_2O$ | 219.1336 | 219.1345 | $C_9H_{18}N_2O_4Na$ | 241.1154 | 241.1164 |
| 2 | 23.3 | $C_9H_{17}N_2O$ | 201.1226 | 201.1239 | $C_9H_{16}N_2O_3Na$ | 223.1042 | 223.1059 |
| 3 | 23.7 | $C_9H_{17}N_2O$ | 201.1231 | 201.1239 | $C_9H_{16}N_2O_3Na$ | 223.1045 | 223.1059 |

With the hypothesis that the dehydroalanine moiety is still intact in the predicted azoxy compounds, the branched isobutyl moiety in valanimycin is expected to be extended by two $CH_2$ units and the proposed structures of possible azoxy compounds (**2** and **3**) are shown in Figure 5-15.

Due to the absence of valine decarboxylase in *S. avermitilis*, a different precursor was predicted to be incorporated in the azoxy compounds. Another amino acid could participate in the biosynthesis of the azoxy compounds; however, there are no amino acids that are larger than valine by two units of $CH_2$. Therefore, an amino acid is unlikely to be incorporated in the predicted azoxy compounds, but a precursor of polyketide origin as proposed earlier.

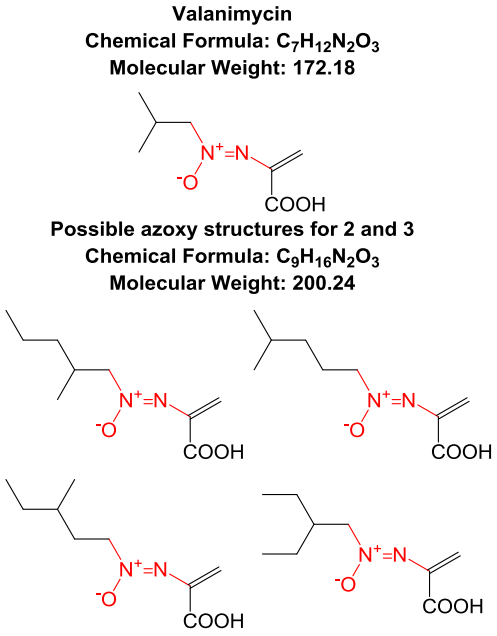



**Figure 5-15** – Structure of valanimycin and proposed structures for the predicted azoxy compounds (**2** and **3**), which contain two extra units of $CH_2$.

### 5.5.3. Identification of predicted aromatic amine compounds in *S. coelicolor* heterologous hosts

In addition to the acidified organic extraction, metabolites from *S. coelicolor* strains were extracted with ethyl acetate at neutral pH and analysed by LC-MS. Three new compounds (**4** - **6**) were detected in the neutral *S. coelicolor* organic extracts, and were only observed in *S. coelicolor* strains containing pESAC13A-2H or

pESAC13A-19K (Figure 5-16). The aromatic amine compounds (**4** - **6**), and not the predicted azoxy compounds (**1** - **3**), were found in the neutral extracts, and *vice versa* for the acidified extracts.

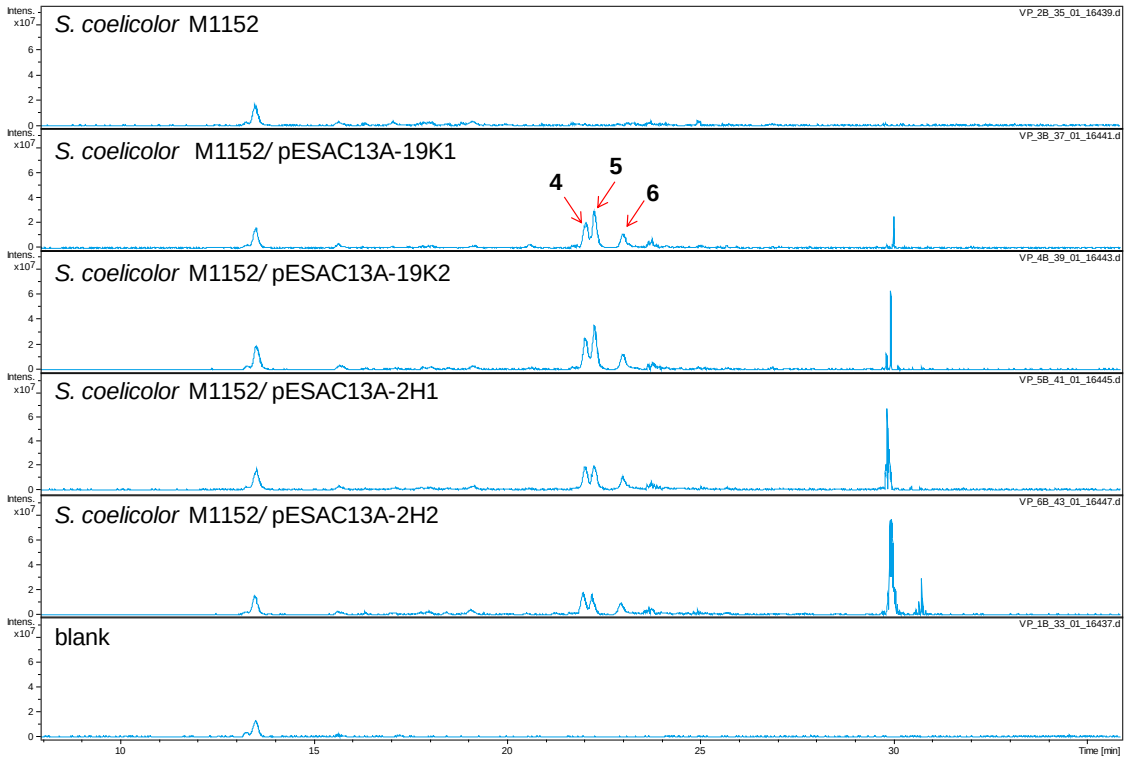



**Figure 5-16** – Aromatic amine compounds (**4** - **6**) were identified in the neutral *S. coelicolor* organic extracts by LC-MS. Extracted ion chromatograms (*m/z* = 262.0, 278.0 and 276.0) detected the compounds **4** - **6** only in *S. coelicolor* M1152 containing pESAC13A-2H or pESAC13A-19K and not in *S. coelicolor* M1152 control or blank.

The molecular ion masses for these three new compounds (where *m/z* for **4** is 262.06, for **5** is 278.06 and for **6** is 276.06) indicate they contain an odd number of nitrogens and are most likely not related to the predicted azoxy compounds (**1** - **3**). This was confirmed by molecular formulae generated from UHPLC-ESI-TOF-MS data. The molecular formulae for compounds **4** - **6** are speculated to be aromatic amine compounds containing at least 3 aromatic rings; this is implied from the ratio of carbons and hydrogens (Table 5-3).

**Table 5-3** – Compounds detected in the neutral organic extraction of *S. coelicolor* heterologous host strains grown on SMMS agar.

| # | Retention Time (min) | $[M+H]^+$ | | | $[M+Na]^+$ | | |
|---|---|---|---|---|---|---|---|
| | | Formula | Observed | Calculated | Formula | Observed | Calculated |
| 4 | 22.0 | $C_{19}H_{20}N$ | 262.1591 | 262.1596 | $C_{19}H_{19}NNa$ | 284.1410 | 284.1415 |
| 5 | 22.2 | $C_{19}H_{20}NO$ | 278.1540 | 278.1545 | $C_{19}H_{19}NONa$ | 300.1359 | 300.1364 |
| 6 | 23.0 | $C_{19}H_{22}NO$ | 276.1745 | 276.1752 | $C_{20}H_{21}NNa$ | 298.1566 | 298.1572 |

The predicted aromatic amine compounds (**4** - **6**) were only observed when pESAC13A-2H or pESAC13A-19K constructs were introduced in *S. coelicolor* heterologous hosts and are likely products of the metabolic pathway introduced *via* the pESAC13A. Both of the pESAC13A constructs contain an insert size of approximately 130 - 135 kb (Figure 5-7); therefore, including the 50 kb predicted azoxy biosynthetic gene cluster (*sav_2266* to *sav_2301*), there are an additional 80 - 85 kb of DNA in the inserts of these pESAC13A constructs. The theoretical DNA insert could span between *sav_2202* to *sav_2378,* and from the predicted metabolite gene clusters published by Ikeda and co-workers, the polyketide (*sav_2367* - *sav_2369*; cluster #16 in Table 1-2) and half of the aromatic polyketide (*sav_2372* - *sav_2388*; cluster #17 in Table 1-2) biosynthetic gene clusters could have been captured along with the predicted azoxy biosynthetic gene cluster. [35] The aromatic amine **4** - **6** may be products of the polyketide or aromatic polyketide biosynthetic gene cluster.

Alternatively, introduction of pESAC13A constructs in *S. coelicolor* M1152 may have activated expression of cryptic biosynthetic gene clusters or modified endogenous metabolites produced by *S. coelicolor* M1152. [174] More importantly, the observed compounds (**1** - **6**) were detected in strains containing either pESAC13A-2H and

pESAC13A-19K, and implies that the two pESAC13A constructs are similar. Sequencing of pESAC13A-2H and pESAC13A-19K constructs would allow the insert ends to be determined and these interpretations to be validated.

## 5.6. Incorporation of isotope labelled L-serine in the predicted azoxy compounds

Incorporation experiments using isotopically-labelled precursors, carried out by Ronald and co-workers, established that valanimycin is derived from L-valine and L-serine (Figure 5-3).[155,156] The seryl residue, from seryl-tRNA, is transferred onto the hydroxyl group of isobutylhydroxylamine to form an intermediate, which is subsequently converted into valanimycin hydrate (Figure 5-3).[152] The seryl residue in valanimycin hydrate is dehydrated to form valanimycin (Figure 5-3).[160] Since serine and threonine amino acids only differ by a single methyl group, L-threonine may be incorporated into the predicted azoxy compounds, instead of L-serine. Thus, dehydration of the threonyl residue would result in the expected cis/trans isomers (**2** and **3**) (Figure 5-17); however, if serine was the precursor no configurational isomers would result from the dehydration. The feeding of isotope-labelled L-serine-$^{13}C_3$-$^{15}N$ and L-threonine-$^{13}C_4$-$^{15}N$ was carried out to determine whether they are incorporated into the predicted azoxy compounds (**1** - **3**).

The mass spectra of azoxy compounds (**1** - **3**) resulting from isotope-labelled feeding experiments revealed incorporation of L-serine-$^{13}C_3$-$^{15}N$ and not L-threonine-$^{13}C_4$-$^{15}N$ (Figure 5-18). In the feeding of L-serine-$^{13}C_3$-$^{15}N$, the expected increase of 4 Da for the $[M+H]^+$ and $[M+Na]^+$ molecular ion mass were observed for all of the azoxy compounds (**1** - **3**) (Figure 5-18). The 4 Da increase implies intact

incorporation of L-serine into the azoxy compounds (**1** - **3**), which is consistent with the proposed valanimycin biosynthesis pathway (Figure 5-3). [160] Incorporation of isotope-labelled serine in all of the predicted azoxy compounds further supports the predictions that compounds **1** - **3** are structurally related and are the novel azoxy compounds of interest.

For the feeding of L-threonine-$^{13}C_4$-$^{15}N$, the molecular mass increase of 5 Da was not observed for the predicted azoxy compounds (**1** - **3**) (Figure 5-18), which indicates that L-threonine is not incorporated into the azoxy compounds (**1** - **3**). Dehydration of the seryl residue of compounds **1** is therefore not responsible for generating the cis/trans isomers, compound **2** and **3** are different type of isomers. The molecular formulae for the *S. avermitilis* azoxy compounds are highly similar to that of valanimycin and indicate they contain an additional two $CH_2$ groups extending the isobutyl moiety in valanimycin. Depending on the chain branching, different structural isomers may be envisaged for **1** - **3** (Figure 5-15).

For the aromatic amine compounds (**4** - **6**), isotope-labelled feeding experiments revealed these compounds do not incorporate either L-serine-$^{13}C_3$-$^{15}N$ or L-threonine-$^{13}C_4$-$^{15}N$ and further supports their unrelatedness with the identified azoxy compounds (**1** - **3**).

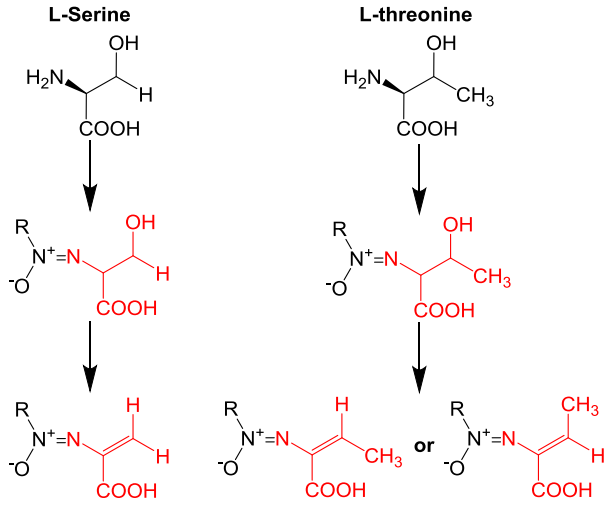
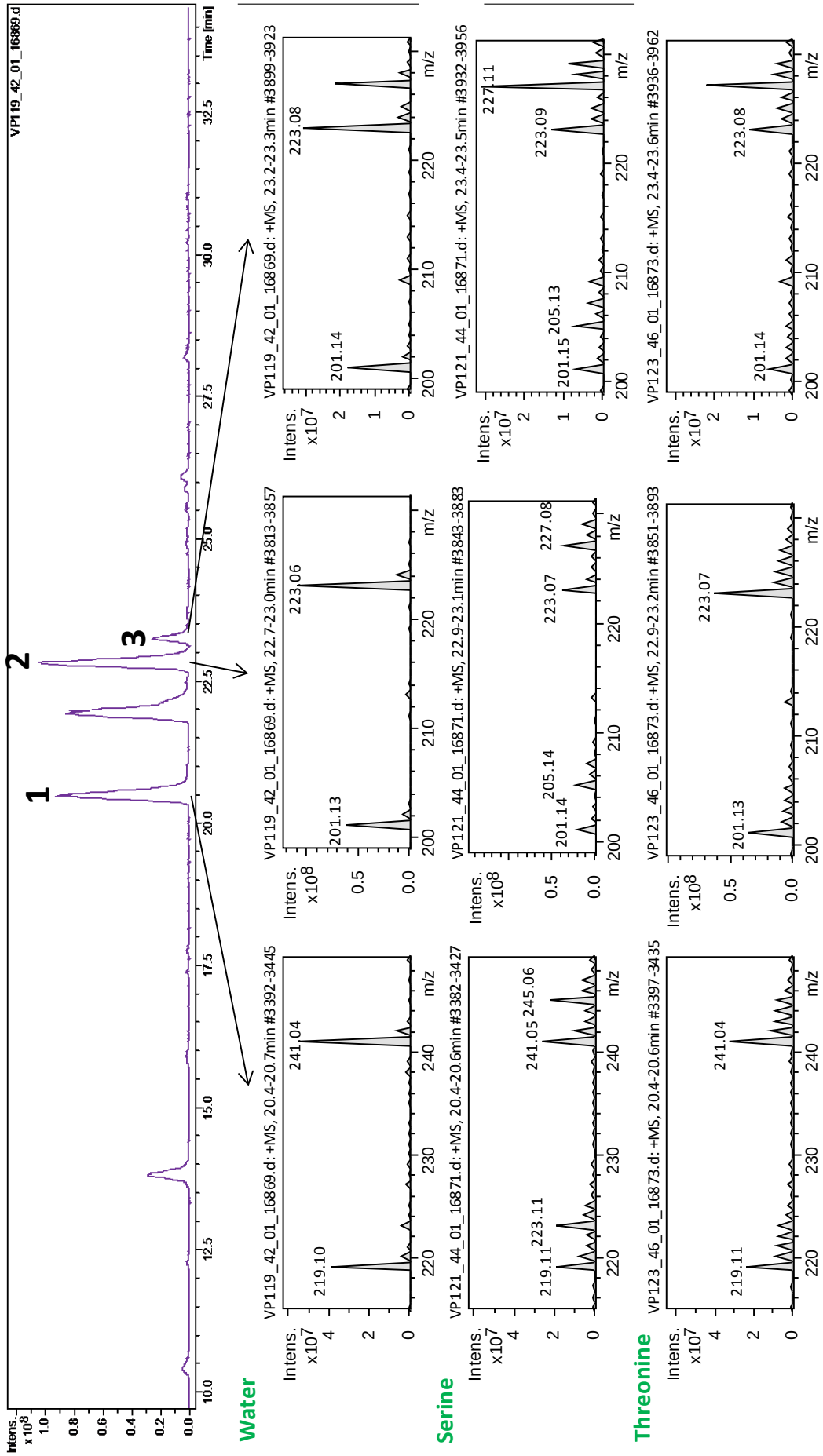



**Figure 5-17** – Proposed incorporation of L-serine or L-threonine into the predicted azoxy compound and subsequent dehydration of the seryl and threonyl residues. In the case of threonyl dehydration cis/trans isomers could be expected.

**Figure 5-18** – (Next Page) The predicted azoxy compounds (**1** - **3**) are incorporating L-serine-$^{13}C_3$-$^{15}N$ and not L-threonine-$^{13}C_4$-$^{15}N$; water was used as a negative control. Extracted ion chromatograms (*m/z* = 201.0 and 219.0) detected the predicted azoxy compounds (**1** - **3**) and their respective mass spectra are shown; molecular ion mass of $[M+H]^+$ and $[M+Na]^+$ are highlighted.


VP119_42_01_16869.d
Time [min]
1
2
3
Water
Serine
Threonine
VP119_42_01_16869.d: +MS, 20.4-20.7min #3392-3445
VP119_42_01_16869.d: +MS, 22.7-23.0min #3813-3857
VP119_42_01_16869.d: +MS, 23.2-23.3min #3899-3923
VP121_44_01_16871.d: +MS, 20.4-20.6min #3382-3427
VP121_44_01_16871.d: +MS, 22.9-23.1min #3843-3883
VP121_44_01_16871.d: +MS, 23.4-23.5min #3932-3956
VP123_46_01_16873.d: +MS, 20.4-20.6min #3397-3435
VP123_46_01_16873.d: +MS, 22.9-23.2min #3851-3893
VP123_46_01_16873.d: +MS, 23.4-23.6min #3936-3962
Intens.
m/z

## 5.7. Purification of azoxy compounds for structure determination by NMR spectroscopy

To obtain enough product for the structure elucidation of compounds **1** - **3**, the growth of *S. coelicolor* M1152/pESAC13A-2H was scaled up to 750 mL on SMMS agar (30 x 25 mL plates). Organic extraction was performed with ethyl acetate and solvent was evaporated *in vacuo*. The residual crude extract was dissolved in methanol and water before LC-MS analysis. Once the presence of compounds of interest was confirmed, purification using semi-preparative HPLC was initiated. The predicted azoxy compounds (**1** - **3**) were previously shown to strongly absorb wavelengths of 220 nm and 230 nm (Figure 5-19), which is also consistent with previously isolated azoxy metabolites. [155,175] For semi-preparative HPLC, the UV chromatogram at 230 nm was used to guide their purification.

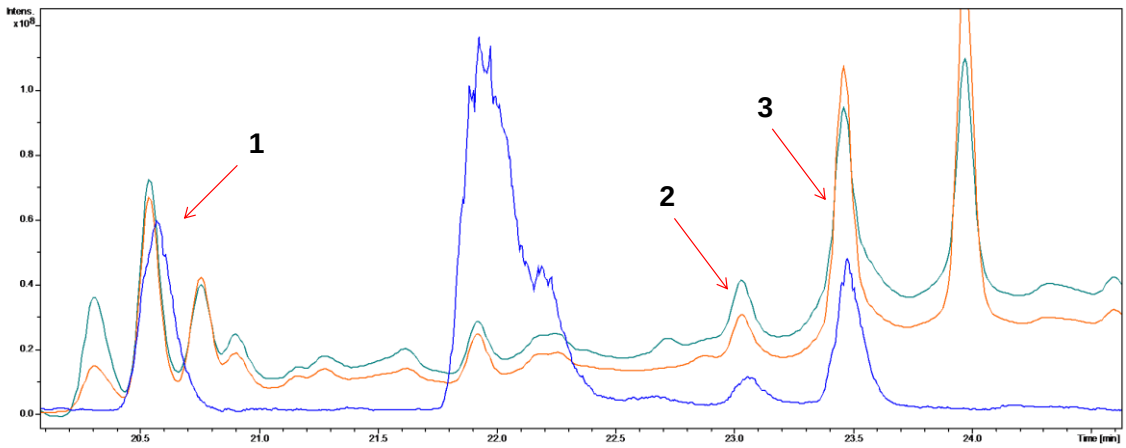



**Figure 5-19** – UV chromatogram of the predicted azoxy compounds (**1** - **3**). Extracted ion chromatograms (blue trace - $m/z$ = 201.0 and $m/z$ = 219.0) detected the azoxy compounds (**1** - **3**) that strongly absorb wavelengths of 220 nm (green trace) and 230 nm (orange trace).

Fractions collected by semi-preparative HPLC were further analysed by mass spectroscopy to confirm the presence of the compounds of interest and pooled together. The samples were concentrated by evaporating the methanol *in vacuo* and

remaining water was removed by freeze drying. The resulting dried residue sample was dissolved in deuterated chloroform and analysed by NMR spectroscopy. Compound **1** (7.3mg) and **3** (17.7mg) were analysed by $^{1}$H-NMR (700 Mhz), and compound **2** (6.6mg) was analysed by a combination of 1- and 2-D NMR spectroscopy ($^{1}$H, $^{13}$C, COSY, HSQC and HMBC) (Figure 5-20).

The signals on the $^{1}$H-NMR spectra obtained for **1** and **3** were not intense enough for determining their structure. For compound **2**, analysis of the NMR spectra (performed by Corre, C. and Song, L.) was discovered to be consistent with NMR spectroscopy data of the SCB1 signalling molecule, which was previously characterised by Takano and co-workers in 2008.[119] The SCB1 signalling molecule has a characteristic mass spectrum pattern with molecular ion peaks for $[M+Na]^+$ ($m/z$ = 267.07), $[M+H]^+$ ($m/z$ = 245.11), $[M+H-H_2O]^+$ ($m/z$ = 227.14) and $[M+H-2H_2O]^+$ ($m/z$ = 209.17) and detailed analysis of mass spectra obtained for the purified compounds, undeniably, revealed SCB1 signalling molecules were co-purified with compounds **2** and **3** (Figure 5-21). Purification of predicted azoxy compounds would need to be further optimised to remove SCB1 signalling molecules. Interestingly, extracted ion chromatogram of the scaled up extract revealed that SCB1 was overproduced in *S. coelicolor* M1152 co-eluate with compound **2** and **3** (Figure 5-21). Optimisation of the purification of **2** and **3** to remove SCB1 signalling molecules would be difficult and an alternative solution would be to genetically inactivate the *scbA* gene, which encodes for the key enzyme involved in the assembly of SCBs.[176] Inactivation of *scbA* in *S. coelicolor* M1152 is expected to abolish the production of SCB signalling molecules and would generate a "cleaner" *S. coelicolor* superhost. For compound **1**, only small amounts of SCBs

were co-eluating (Figure 5-21). Therefore, purification of the proposed hydrated azoxy compounds (**1**) could be further optimised for establishing its structure. The amount of purified material from 750 mL of SMMS agar was too low to obtain signals intense enough by NMR spectroscopy, and would need to be further scaled up to obtain more product. Production of predicted azoxy compounds (**1** - **3**) were still detectable in 50 mL of liquid SMM at 30 $^{0}$C, shaking at 220 rpm for 6 days and compounds **1** - **3** could be further scaled up in liquid culture. Furthermore, previous attempts to scale up the production of the predicted azoxy compounds (**1** - **3**) were carried out in *S. coelicolor* M1152/pESAC13A-2H1 strain and genetic manipulations, such as overexpression of the *sav_2301* activator or inactivation of the *sav_2268* repressor, is expected to over produce the compounds of interest. In addition, a significant decrease in valanimycin was observed in *S. viridifaciens* after 40 hours of incubation; growth conditions of *S. coelicolor* heterologous hosts could be further optimised to obtain the maximum yield of azoxy compounds. [155]

With greater production of the compounds of interest, the samples could be further analysed by infra-red (IR) spectroscopy to confirm the azoxy moiety, which has a characteristic IR spectrum at ~1540 cm$^{-1}$. [177]

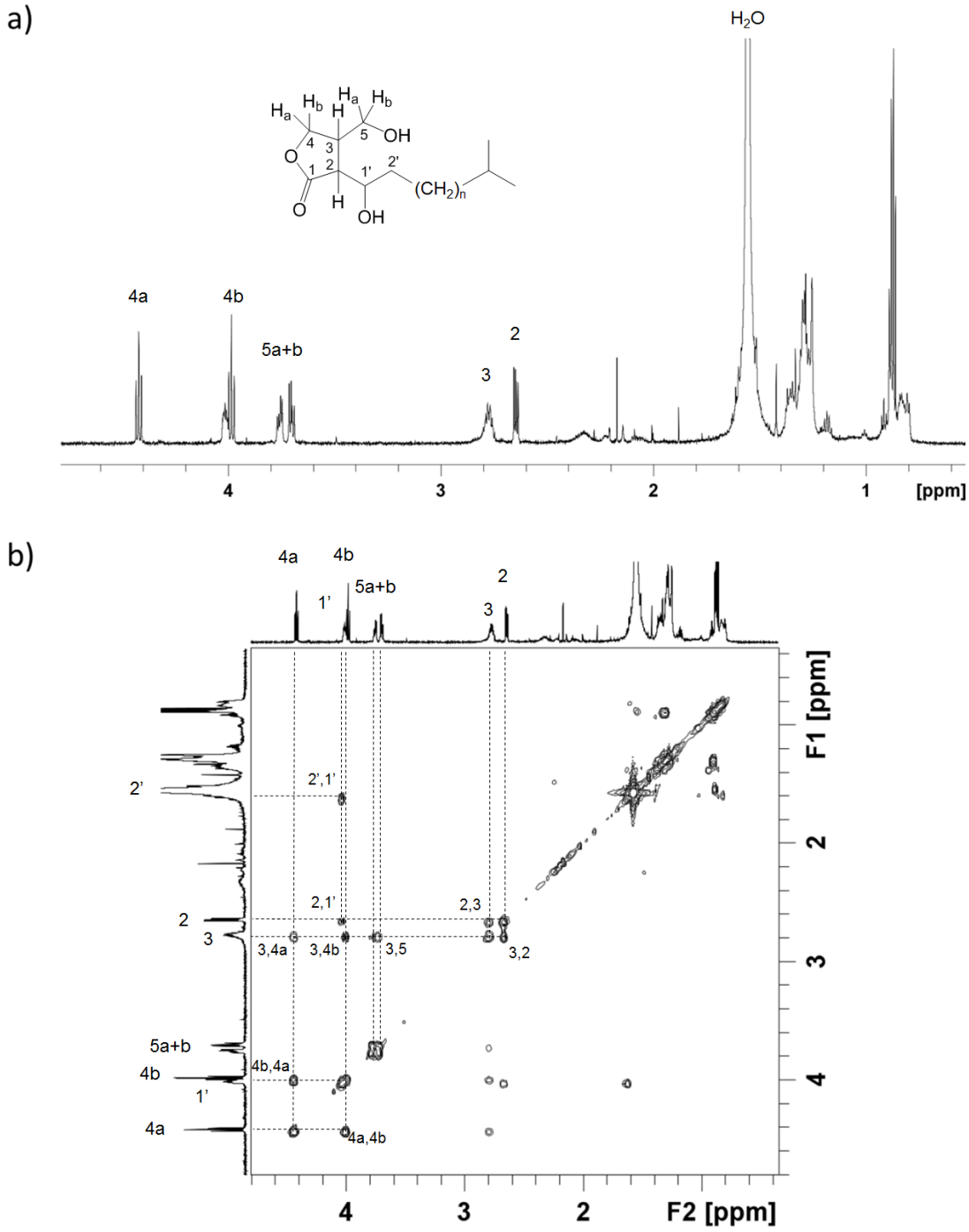



**Figure 5-20** – (a) Proton NMR (700 Mhz) and (b) 2-D COSY spectrum of compound **2** dissolved in deuterated chloroform. The NMR data for compound **2** matched the structure of the SCB1 signalling molecule published by Takano and co-workers (See reference 117).

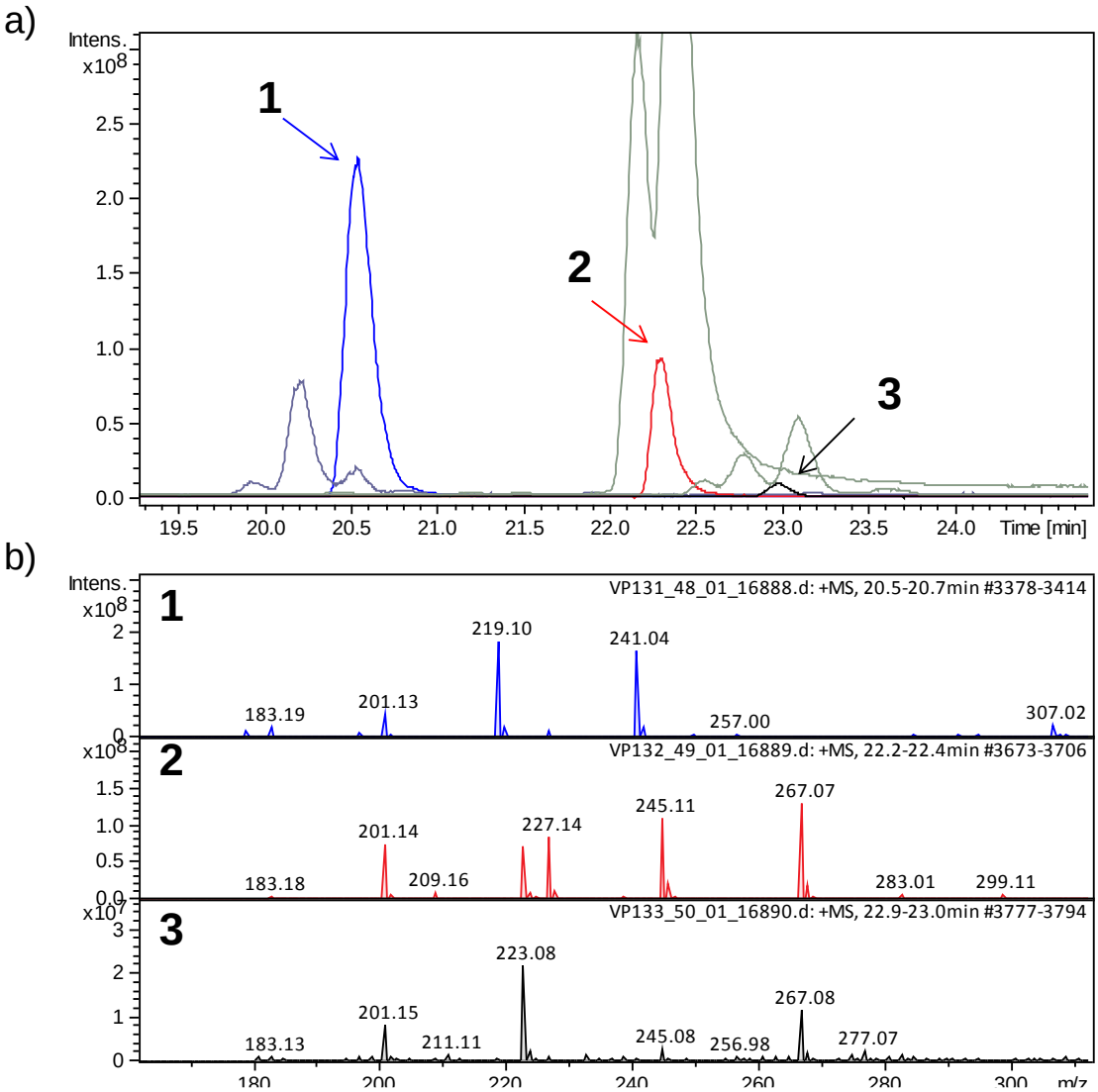



**Figure 5-21** – (a) Extracted ion chromatogram (*m/z* = 219.0, 201.0 and 201.0 traces are coloured blue, red and black respectively) and (b) mass spectra of compounds **1** - **3** reveal the predicted azoxy compounds are co-purified with SCB signalling molecules, which are overproduced in *S. coelicolor* M1152 (*m/z* = 209.17, 227.14, 245.11 and 267.07, grey trace).

## 5.8. Conclusions

To our knowledge, the cryptic biosynthetic gene cluster in *S. avermitilis*, which is proposed to be under the control of the neighbouring putative AHFCA-dependent signalling system, has not been investigated. Genome mining analyses have revealed a set of genes that are shared between the valanimycin biosynthetic gene cluster from *S. viridifaciens* and the cryptic biosynthetic gene cluster from *S. avermitilis*; therefore, the cryptic biosynthetic gene cluster is predicted to be involved in the biosynthesis of novel azoxy natural compounds.

A *S. avermitilis* pESAC13A PAC genomic library was constructed for the project by Bio S&T. The pESAC13A-2H and pESAC13A-19K constructs were confirmed to have captured the predicted azoxy biosynthetic gene cluster. The pESAC13A constructs were successfully introduced into *S. coelicolor* M1152 and *S. lividans* TK24 by tri-parental conjugation and comparative metabolic profiling identified 6 new compounds (**1** - **6**). The molecular formulae, generated through the analysis of UHPLC-ESI-TOF-MS data, confirmed compounds **1** - **3** are the predicted azoxy compounds of interest and compounds **4** - **6** are aromatic amines.

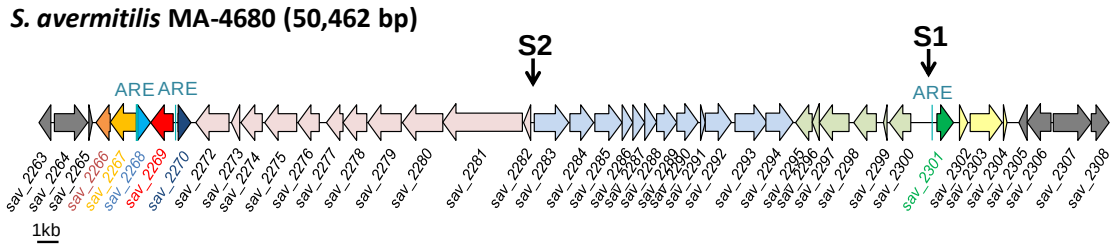
To further support the identification of the predicted azoxy compounds, compounds **1** - **3** were demonstrated to incorporate isotope-labelled L-serine-$^{13}C_3$-$^{15}N$. Production of compound **1** - **3** were further scaled up and purified by semi-preparative HPLC in order to determine the structure of compounds by a combination of 1- and 2-D NMR spectroscopy; however, the structure of compounds **1** - **3** could not be solved and purification needs to be optimised by further scale up and removal of SCB signalling molecules.

# 6. Exploitation of AHFCA-dependent Signalling Systems

The MEME enrichment analysis has proven to be an efficient tool for gaining insights into complex regulatory networks and was successful at revealing TetR binding sites for all of the putative AHFCA-dependent signalling systems. In the *S. avermitilis* AHFCA-dependent signalling system, TetR binding sites were identified in promoters of genes involved in the assembly of AHFCA signalling molecules, TetR repressors and upstream of *sav_2301* SARP transcriptional activator (Figure 6-1); hence, the signalling system is proposed to regulate expression of *sav_2301* and control the biosynthesis of the predicted azoxy compounds. The SARP family of transcriptional activators are commonly found in *Streptomyces* biosynthetic gene clusters and are involved in pathway-specific regulation of biosynthetic gene clusters by binding to direct heptameric DNA repeats located in promoter regions. [116]

## 6.1. Predictions of SAV_2301 SARP transcriptional activator binding sites

Organisation of the predicted azoxy biosynthetic gene cluster suggests there are four possible operons; *sav_2282* - *sav_2272* operon that is divergently adjacent to *sav_2283* - *sav_2294* operon and *sav_2300* - *sav_2295* operon is divergently adjacent to *sav_2301* - *sav_2304* operon (Coloured in pale red, blue, green and yellow respectively in Figure 6-1). Bioinformatic analysis identified SARP-like transcriptional activator binding sites in the intergenic regions between these divergently adjacent putative operons (Figure 6-1 and Figure 6-2).




| Name | Proposed Function |
|---|---|
| SAV_2263 | TetR family transcriptional regulator |
| SAV_2264 | Transmembrane efflux protein |
| SAV_2265 | Hypothetical protein |
| SAV_2266 | MmfP homologue |
| SAV_2267 | MmfH homologue |
| SAV_2268 | MmyR homologue |
| SAV_2269 | MmfL homologue |
| SAV_2270 | MmfR homologue |
| SAV_2272 | MmgE/PrpD family protein |
| SAV_2273 | Isomerase |
| SAV_2274 | Diacylglycerol kinase |
| SAV_2275 | Transmembrane efflux protein |
| SAV_2276 | 3-oxoacyl-ACP synthase |
| SAV_2277 | Thioesterase |
| SAV_2278 | F420-dependent dehydrogenase |
| SAV_2279 | Long-chain-fatty-acid-CoA ligase |
| SAV_2280 | Modular polyketide synthase |
| SAV_2281 | Modular polyketide synthase |
| SAV_2282 | ACP |
| SAV_2283 | Aldehyde dehydrogenase |
| SAV_2284 | Isobutylamine N-hydroxylase |
| SAV_2285 | Ornithine aminotransferase |
| SAV_2286 | Cyclase |

| Name | Proposed Function |
|---|---|
| SAV_2287 | Integral membrane protein |
| SAV_2288 | NADPH-flavin oxidoreductase |
| SAV_2289 | Hypothetical protein |
| SAV_2290 | 3-oxoacyl-ACP synthase |
| SAV_2291 | ACP |
| SAV_2292 | 3-oxoacyl-ACP synthase |
| SAV_2293 | Amino acid permease |
| SAV_2294 | Seryl-tRNA synthetase |
| SAV_2295 | Enoyl-ACP reductase |
| SAV_2296 | Ferredoxin |
| SAV_2297 | MFS transporter |
| SAV_2298 | Fatty acid desaturase |
| SAV_2299 | Hypothetical protein |
| SAV_2300 | Integral membrane lysyl-tRNA synthetase |
| SAV_2301* | SARP family transcriptional regulator * |
| SAV_2302 | Membrane protein |
| SAV_2303 | Methylmalonyl-CoA carboxyltransferase |
| SAV_2304 | Hypothetical protein |
| SAV_2305 | Enamine deaminase |
| SAV_2306 | Acyl-CoA dehydrogenase |
| SAV_2307 | Acyl-CoA synthetase |
| SAV_2308 | PaaX family transcriptional regulator |

**Figure 6-1** – The 50 kb cryptic biosynthetic gene cluster appears to be under the control of an AHFCA-dependent signalling system. MEME enrichment analysis identified ARE sequences upstream of *sav_2301* (highlighted in green). The SAV_2301 transcriptional activator is proposed to bind to SARP binding sites (S1 and S2) to regulate expression of the four operons (highlighted in pale red, blue, green and yellow). *contains TTA codon.

The predicted SARP-like binding sites are similar to other experimentally confirmed SARP binding sites; SAV_2301 is proposed to bind to these 2 specific sequences to recruit RNA polymerase and activate the 4 divergently adjacent operons (Figure 6-1 and Figure 6-2).[161]

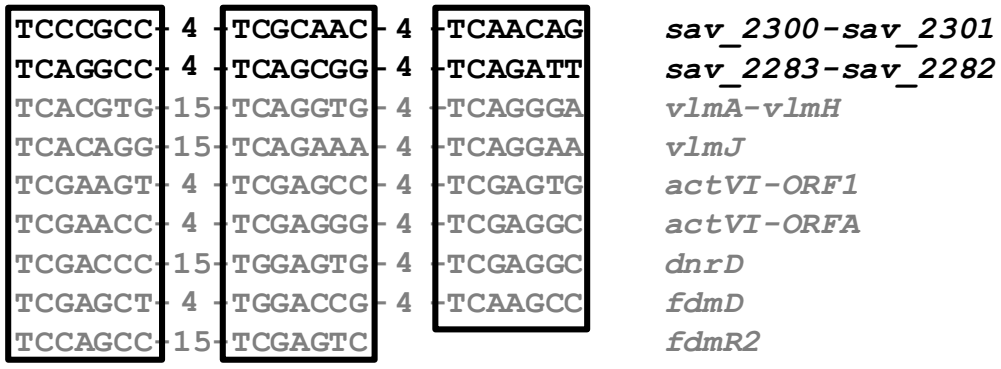



**Figure 6-2** – Conserved direct heptameric repeats were identified in the promoters of divergently adjacent putative operons. SARP-like binding sequences are similar to other experimentally confirmed SARP binding sites identified in *Streptomyces* bacteria. SAV_2301 activator is proposed to bind to the sequences to activate expression of downstream genes. Image adapted from reference 161.

From earlier MEME enrichment analysis, one of the TetR binding sites was found in the *sav_2301* promoter. The ARE sequence was discovered to overlap the -10 and -35 boxes; this implies the binding of TetR repressors to the specific ARE sequence is expected to prevent expression of *sav_2301* (Figure 6-1 and Figure 6-3). The predicted SARP binding sequence is located upstream of the promoter elements of *sav_2301* and binding of SAV_2301 activator to the SARP binding sequence is expected to drive expression of itself and further amplify signals for the azoxy biosynthetic genes (Figure 6-3). Intriguingly, the ARE sequence is located downstream of the predicted SARP binding sequence and there is a 2 bp nucleotide overlap between the two sequences; binding of TetR repressors is speculated to block binding of SARP activator and/or RNA polymerase (Figure 6-3). Therefore, overexpression of *sav_2301* transcriptional activator alone may not result in the activation and expression of the silent biosynthetic gene cluster, genetic inactivation of TetR repressors or addition of a specific amount of AHFCA signalling molecules is proposed to be required for the overproduction of compounds of interest. Another

conserved SARP-like binding sequence was also identified within the *sav_2282* - *sav_2283* intergenic region and SAV_2301 activator is expected to drive expression of the two divergently adjacent operons. Additionally, MEME enrichment analysis did not reveal any ARE sequence within the *sav_2282* - *sav_2283* intergenic region; therefore, overexpression of *sav_2301* is expected to increase expression of divergently adjacent *sav_2282* - *sav_2272* and *sav_2283* - *sav_2294* operons (Figure 6-1).

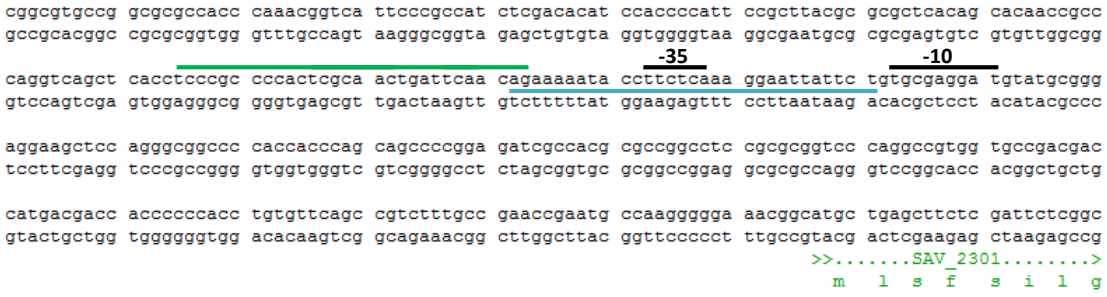



**Figure 6-3** – Binding of TetR repressors to the ARE sequence (turquoise line), which overlaps the -35 and -10 boxes for *sav_2301* (black line), is expected to prevent expression of *sav_2301*. SAV_2301 activator is predicted to bind to heptameric repeats (green line) located upstream of the promoter elements to drive expression of itself. The -10 and -35 boxes were predicted using BPROM.

## 6.2. Cross-talk with other specialised metabolite biosynthetic gene clusters in *S. avermitilis*

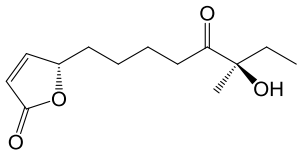
FIMO analysis of the *S. avermitilis* putative AHFCA-dependent signalling system reveals a total of 246 TetR binding sites across the *S. avermitilis* genome. Of these 246 predicted TetR binding sites, 81 binding sites were located in promoter of ORFs (within 300 bp of an ORF start codon) and 8 of these binding sites were located within specialised metabolite biosynthetic gene clusters (Table 6-1 and Electronic Supplementary Material). The *S. avermitilis* AHFCA-dependent signalling system

may cross-talk and regulate (activating or repressing) other biosynthetic gene clusters, such as filipin or oligomycin clusters (Table 6-1). Furthermore, detailed analysis of FIMO results revealed a binding site upstream of *aco* gene (*sav_3706*), which encodes for a key enzyme involved in the biosynthesis of avenolides (Table 6-1). [40]

**Table 6-1** – Proposed SAV_2268/SAV_2270 binding sites in *S. avermitilis* specialised metabolite biosynthetic gene clusters.

| Start | End | ARE Sequence | Gene downstream of ARE sequence | Gene Cluster |
|---|---|---|---|---|
| 527141 | 527164 | AAGCCGCCTTCGCGCACGTACGTG | *sav_416* | Filipin |
| 2766276 | 2766299 | AGAATTCGTTCACGAACGTATCTT | *sav_2267* | AHFCA |
| 2766276 | 2766299 | AAGATACGTTCGTGAACGAATTCT | *sav_2268* | AHFCA |
| 2768078 | 2768101 | AATATACCTGCGCGAAGGTATATT | *sav_2269* | AHFCA |
| 2803762 | 2803762 | AAAATACCTTCTCAAAGGAATTAT | *sav_2301* | Azoxy |
| 2910600 | 2910623 | AGAACTCGTCCATGATGGACTTCT | *sav_2385* | Aromatic |
| 3536473 | 3536496 | AAACTTCGGCCGAGCAGGAATTCA | *sav_2892* | Oligomycin |
| 3536473 | 3536496 | TGAATTCCTGCTCGGCCGAAGTTT | *sav_2891* | Oligomycin |
| 4585568 | 4585591 | AAGATACGTACTAGACGGTTTTGT | *sav_3706* | Avenolide |
| 8790017 | 8790040 | ACTCATCCTGCGCGTACGTGTCCT | *sav_7362* | PK |

The avenolide signalling molecules represents a structually distinct family of signalling molecules compared to the AHFCAs (Figure 6-4); it is a small diffusible butenolide molecule produced in *S. avermitilis* and their mechanism of action is proposed to be analogous to that of GBL and AHFCA signalling molecules. [40] The avenolides have been shown to be essential for avermectin production and deletion of the *aco* gene results in the abolished production of both avermectins and avenolides. [40]




Avenolide

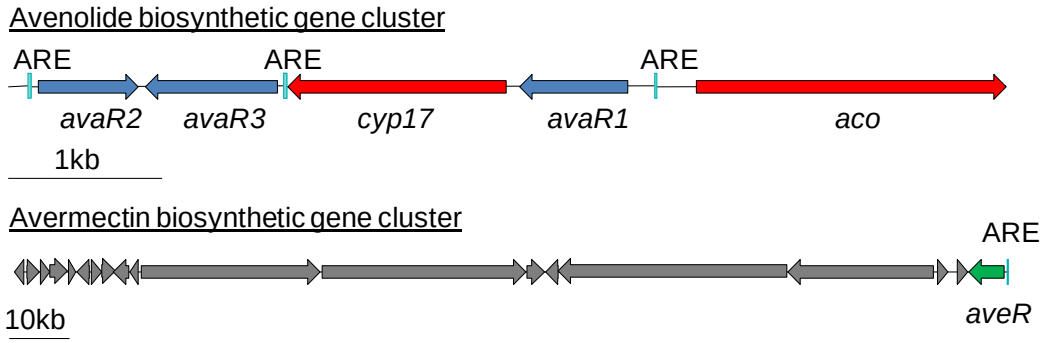
**Figure 6-4** – Avenolide signalling molecules in *S. avermitilis a*re essential for avermectin biosynthesis.

The avenolide signalling gene cluster contains a total of 5 genes (Figure 6-5), which consists of 3 TetR repressor genes, *avaR1* (*sav_3705*), *avaR2* (*sav_3703*) and *avaR3* (*sav_3702*), and 2 biosynthetic genes, *aco* and *cyp17* (*sav_3704*); *aco* and *cyp17* gene products encodes for a putative acyl-CoA oxidase and cytochrome P450 respectively. [40] AvaR1 and AvaR3 are negative regulators of the avermectin biosynthetic gene cluster as deletion of *avaR1* and *avaR3* resulted in increased production of avermectins; the biological role and activity of AvaR2 has yet to be investigated. [178,179]

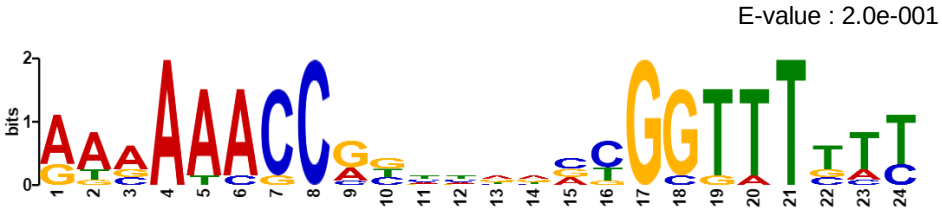
MEME enrichment analysis was performed on the avenolide-dependent signalling system to determine AvaR1, AvaR2 and AvaR3 binding sites across the *S. avermitilis* genome. Highly conserved semi-palindromic sequences in promoters of TetR repressor genes and upstream of *aco* gene were successfully identified (Figure 6-5). Previously published EMSA and DNase I footprinting analyses established that AvaR1 repressor can bind upstream of the *aco* promoter region (126 bp upstream of *aco* start codon) and protected a 30 bp sequence. [40,178] However, MEME enrichment analysis could not identify this specific DNA sequence, but a different ARE sequence was found in the *aco* promoter (Figure 6-5). The experimentally confirmed AvaR1 binding sequence (AAGACAAAACCGTCTAGTACGTATCTTTGA; palindromic sequences are highlighted in red) contains a longer variable spacer

compared to other ARE sequences and is most likely the reason why the MEME enrichment analysis failed to find this specific sequence (Figure 6-5); these observations suggests AvaR1, AvaR2 and AvaR3 may have different binding specificities and the MEME enrichment analysis may be bias towards the prediction of AvaR2 and/or AvaR3 repressors assuming that AvaR1 recognises a different sequences.

a)




b)




| Name | bp upstream of start codon | *p*-value | Sites |
|---|---|---|---|
| SAV_3706 | 250 | 1.23e-11 | GACCGGGACG ATAAAACCAGCTGAACGGTTTTTC AATGCTCCCC |
| SAV_3705 | 170 | 1.23e-11 | GGGGAGCATT GAAAAACCGTTCAGCTGGTTTTAT CGTCCCGGTC |
| SAV_3703 | 36 | 1.29e-10 | CTGATGCGCG AGAAAACCGCACAAGCGGTTTGTT AAAGTTGCTG |
| SAV_3702 | 41 | 1.42e-10 | GGGCGCCACA AACAAACCGCTATGCTGGTTTCTT GCAGGGCCCT |
| SAV_935 | 24 | 3.08e-7 | TCAAGCAATC GAAATCCCCGAGACACGCTTTCTT GGAAAAAGGA |

**Figure 6-5** – The avenolide signalling system (a) is proposed to regulate avermectin biosynthetic gene cluster by controlling expression of *aveR*. MEME enrichment analysis (b) revealed conserved semi-palindromic ARE sequences in the avenolide-dependent signalling system and promoter of *aveR*.

The predicted ARE sequences for the AHFCA-dependent and avenolide-dependent signalling system in *S. avermitilis* were found to overlap *avaR1* and *aco* promoter

elements respectively (Figure 6-6), and binding of TetR repressors are expected to repress the expression of *avaR1* and *aco*. Most significantly, the ARE sequence, which was predicted for the AHFCA-dependent signalling system, is consistent with the experimentally confirmed AvaR1 binding site (Figure 6-6); SAV_2268/SAV_2270 repressors may have similar binding specificities as AvaR1. [40,178]

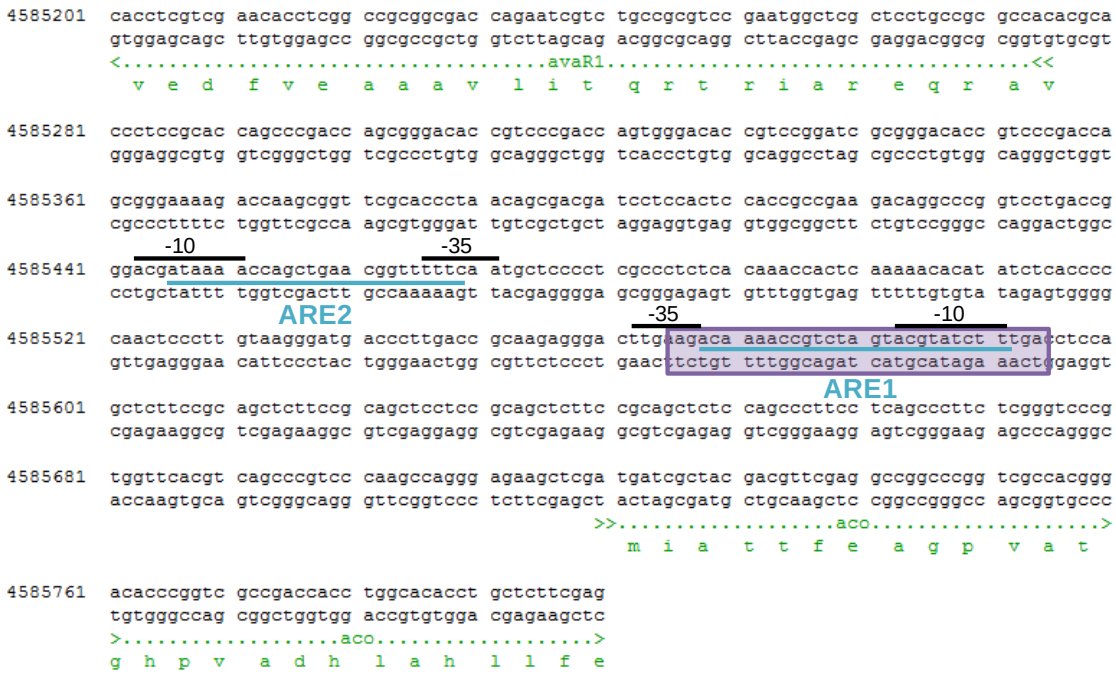



**Figure 6-6** – MEME enrichment analysis predicted a TetR binding sites (underlined in turquoise) for AHFCA-dependent (ARE1) and avenolide-dependent (ARE2) signalling system within the *avaR1* - *aco* intergenic region; both ARE sequences overlap the -10 and -35 boxes of *avaR1* and *aco* respectively. The experimentally confirmed *avaR1* binding is highlighted in a purple box.

Detailed analysis of the avenolide-dependent signalling system reveals a total of 253 individual binding sites across the *S. avermitilis* genome and 58 binding sites were found in promoters of ORFs (within 300 bp of the start codon) (Electronic Supplementary Material). One binding site of interest was located upstream of the *aveR* promoter, also known as *sav_935* (Figure 6-5 and Table 6-2). AveR encodes for

a transcriptional activator that belongs to the LAL family of transcriptional regulator and has been found to be essential for the production of avermectins and also appears to negatively regulate oligomycin biosynthesis. [180,181] EMSA and *in vivo* assays have also confirmed the binding of AvaR1 repressor to the promoter of *aveR* (Wang, W. Personal communication).

**Table 6-2** – Proposed AvaR1/AvaR2/AvaR3 binding sites in *S. avermitilis* specialised metabolite biosynthetic gene clusters.

| Start | End | ARE Sequence | Gene downstream of ARE sequence | Biosynthetic gene cluster |
|---|---|---|---|---|
| 1131715 | 1131735 | AATACCGGGTGACCGGTATGT | *sav_935* (*aveR*) | Avermectin |
| 3930109 | 3930129 | CACAACCGTCCGGCGGGTGTT | *sav_3155* | NRP |
| 4581364 | 4581384 | ACAAACCGCTATGCTGGTTTC | *sav_3702* (*avaR2*) | Avenolide |
| 4583028 | 4583048 | AAAACCGCACAAGCGGTTTGT | *sav_3703* (*avaR3*) | Avenolide |
| 4585448 | 4585468 | AAAAACCGTTCAGCTGGTTTT | *sav_3705* (*avaR1*) | Avenolide |
| 4585448 | 4585468 | AAAAACCGTTCAGCTGGTTTT | *sav_3706* (*aco*) | Avenolide |

Bioinformatic analysis of TetR repressor binding sites in the *S. avermitilis* avenolide-dependent signalling system has therefore provided unprecedented information of the complex regulatory network. The avenolide-dependent signalling system appears to be very complicated and does not specifically regulate any neighbouring biosynthetic gene cluster, but controls expression of several biosynthetic gene clusters across the genome. Binding of AvaR1/AvaR2/AvaR3 and SAV_2268/SAV_2270 to the predicted ARE sequence needs to be further investigated in order determine their exact biological roles and functions. From genome mining analysis, analogous avenolide signalling gene clusters have been

identified in *S. fradiae, S. ghanaensis* and *S. griseoauranticus* and investigations into these clusters may result in the discovery of avenolide-like signalling molecules. [40]

Inactivation of *sav_2268* or *sav_2270* repressors in the AHFCA-dependent signalling system may indirectly affect regulation of several other gene clusters, such as those for avenolide, avermectin, oligomycin and filipin biosynthesis. Furthermore, global transcriptional regulators may regulate the azoxy biosynthetic gene cluster and deletion of *sav_2270* may not result in the unlocked production of the predicted azoxy compounds. Using heterologous hosts would prevent the cross-talk between diverse biosynthetic gene clusters from *S. avermitilis*; although cross-communication with another set of gene clusters may still occur in heterologous hosts. The azoxy-containing metabolites (**1** - **3**) were not detected in *S. lividans* heterologous hosts containing the pESAC13A constructs, which could be explained by the fact that this cluster may still be tightly regulated.

## 6.3. Inactivation of *sav_2268* to unlock the production of azoxy compounds and AHFCA signalling molecules

In our previous studies of AHFCA-dependent signalling systems in *S. coelicolor* and *S. venezuelae*, inactivation of *mmyR* and *gbnR* repressor led to the overproduction of specialised metabolites (methylenomycins and gaburedins respectively), as well as overproduction of AHFCAs (Sidda, J. D. Manuscript in preparation). [57,62] Inactivation of *sav_2268* (*mmyR* orthologue) in the azoxy gene cluster was therefore expected to result in the constitutive production of azoxy compounds (**1** - **3**) and AHFCAs (Figure 3-12).

Various approaches and methodologies have been attempted to exploit regulatory systems that controls production of metabolites, including the deletion of transcriptional repressors and overexpression of transcriptional activators. [131,182]

The pER2 plasmid, which was previously constructed by separately cloning *sav_2267* and *sav_2269* into the pCC4 vector, was used to delete the entire *sav_2268* repressor gene in *S. lividans* heterologous host strains by double homologous recombination (Figure 6-7 and Figure 6-8). The pER2 plasmid, which is based on the pCC4 vector, contains the integrative element of pSAM2 that is needed to be deleted to prevent prevent site-specific recombination (Figure 6-8). [183] The pER2 plasmid was digested with *Acl*I restriction enzyme, and the 8.3 kb DNA fragment was recovered by gel extraction and re-ligated to generate a pER2Δint construct (Figure 6-9).

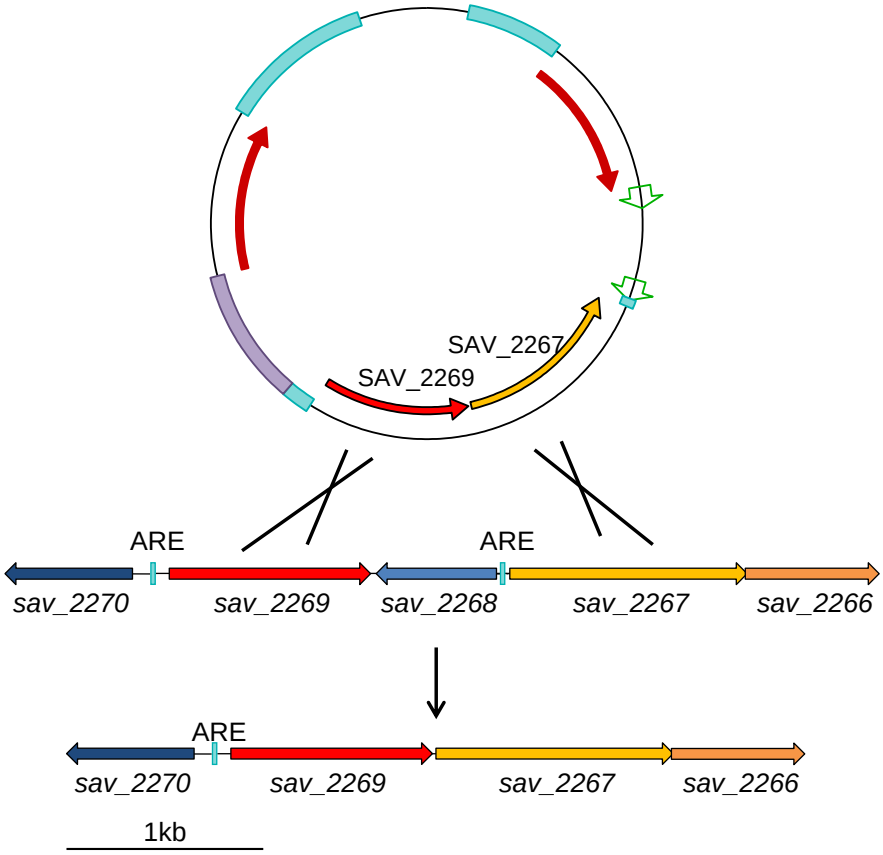



**Figure 6-7** – Homologous recombination strategy to delete the *sav_2268* repressor gene in pESAC13A-2H and pESAC13A-19K.

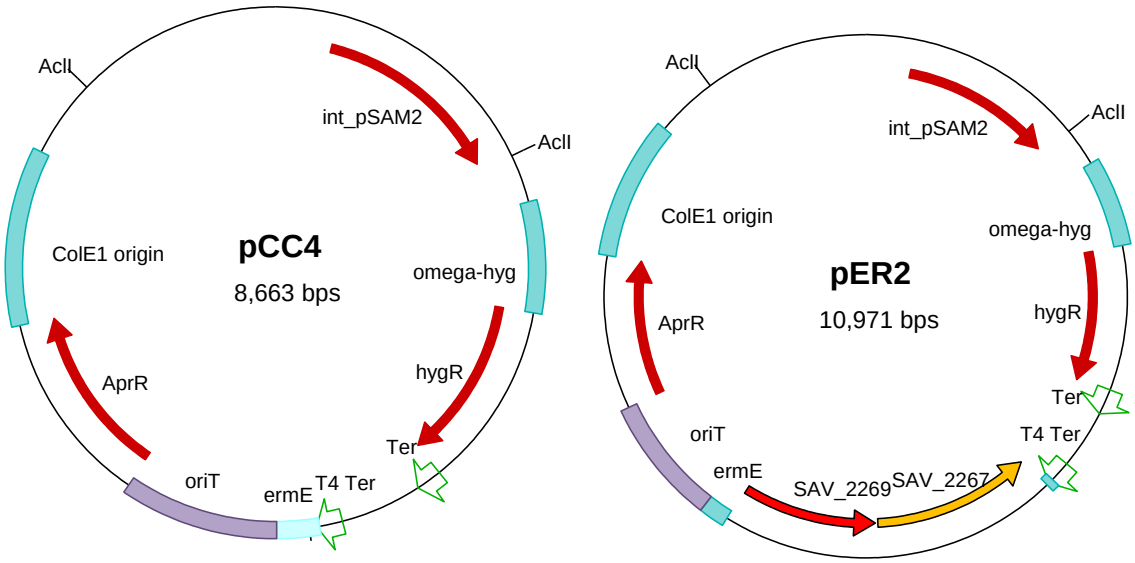



**Figure 6-8** – The *sav_2269* and *sav_2267* genes were separately cloned into the integrative pCC4 vector to produce the pER2 plasmid.

Deletion of the integrative system in pER2 plasmid was screened by *Acl*I restriction enzyme digestion, PCR, and DNA sequencing. For the restriction enzyme digestion screening, a single product of 8,309 bp was expected and a band consistent with this size was observed for the digested pER2Δint; confirming the deletion of the integrase system (Figure 6-9 and Figure 6-10). To further support the deletion of the integration system, primers (for_integrase_screen and rev_integrase_screen, see Table 2-3) were designed to anneal upstream and downstream of the *Acl*I restriction enzyme site and a band consistent with the expected PCR product of 3,214 bp for pER2 and 552 bp for pER2Δint were obtained by PCR (Figure 6-10). The 552 bp PCR product for pER2Δint was sent for DNA sequencing and results further confirmed the successful deletion of the integrase system.

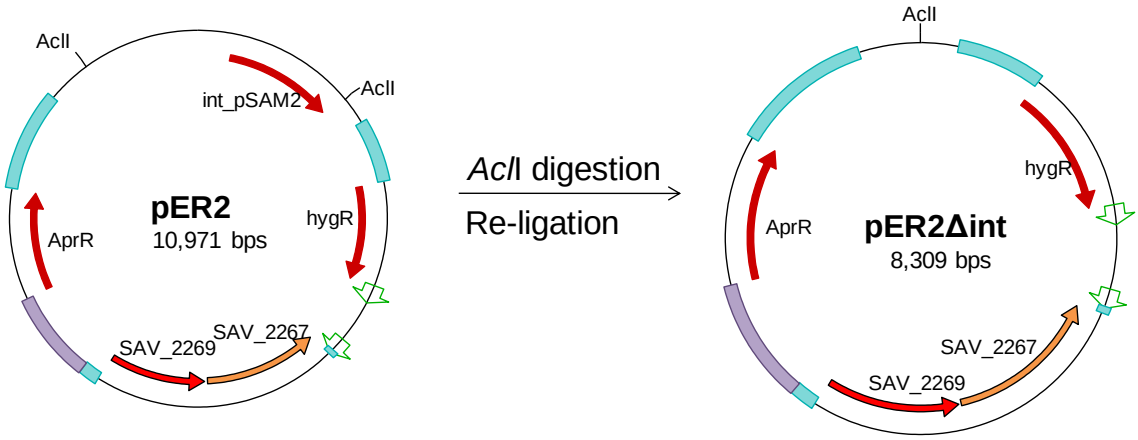



**Figure 6-9** – The pER2 plasmid was digested with *Acl*I restriction enzyme to delete the integrase system. The resulting 8.3 kb digested fragment was re-ligated to generate the pER2Δint construct.

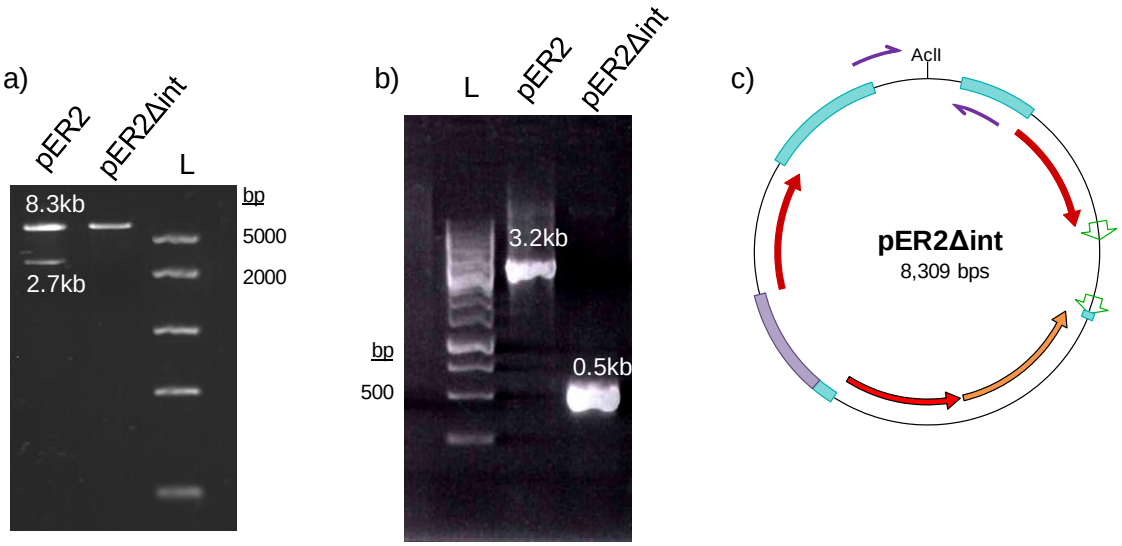



**Figure 6-10** – The integrase element was confirmed to be deleted in the pER2Δint construct by (a) *Acl*I restriction enzyme digestion and (b) PCR. For the *Acl*I restriction enzyme digestion, bands corresponding to the expected product of 8,303 bp was obtained in the pER2Δint plasmid (c) Primers (purple arrows) were designed to flank the *Acl*I restriction enzyme site and (b) a band consistent the expected PCR product of 552 bp was obtained in the pER2Δint plasmid.

The pER2Δint construct was successfully introduced into conjugative *E. coli* ET12567/pUZ8002 host by electroporation and intergenic conjugation was successfully carried out between *E. coli* ET12567/pUZ8002/pER2Δint and *S. lividans* heterologous host strains containing pESAC13A-2H and pESAC13A-19K; exconjugants were selected with apramycin and hygromycin. Since the integrative system of pER2 were deleted in pER2Δint construct, the first recombination event was expected to occur in the *sav_2269* or *sav_2267* arms with homologous DNA sequence to those in the azoxy gene cluster integrated into the chromosome (Figure 6-7 and Figure 6-11). The first recombination event would integrate the entire pER2Δint construct, which contains the hygromycin resistance gene, into *S. lividans* TK24/pESAC13A-2H and *S. lividans* TK24/pESAC13A-19K to generate hygromycin resistant *S. lividans* colonies (Figure 6-7 and Figure 6-11). The second recombination event was obtained after passaging *S. lividans* colonies on SFM agar

with apramycin and nalidixic acid but no hygromycin. After each passage, *S. lividans* colonies were streaked for single colonies and selected for hygromycin sensitivity (Figure 6-11). The second recombination event where the pER2Δint backbone was lost resulting in hygromycin sensitive *S. lividans* mutant strains (Figure 6-7 and Figure 6-11) was desired and importantly deletion of the *sav_2268* repressor gene was expected in some of these new mutants. A total of 4 hygromycin sensitive *S. lividans* TK24/pESAC13A-2H Δ*sav_2268* mutant strains were obtained and deletion of *sav_2268* repressor gene was assessed by a series of PCR and DNA sequencing.

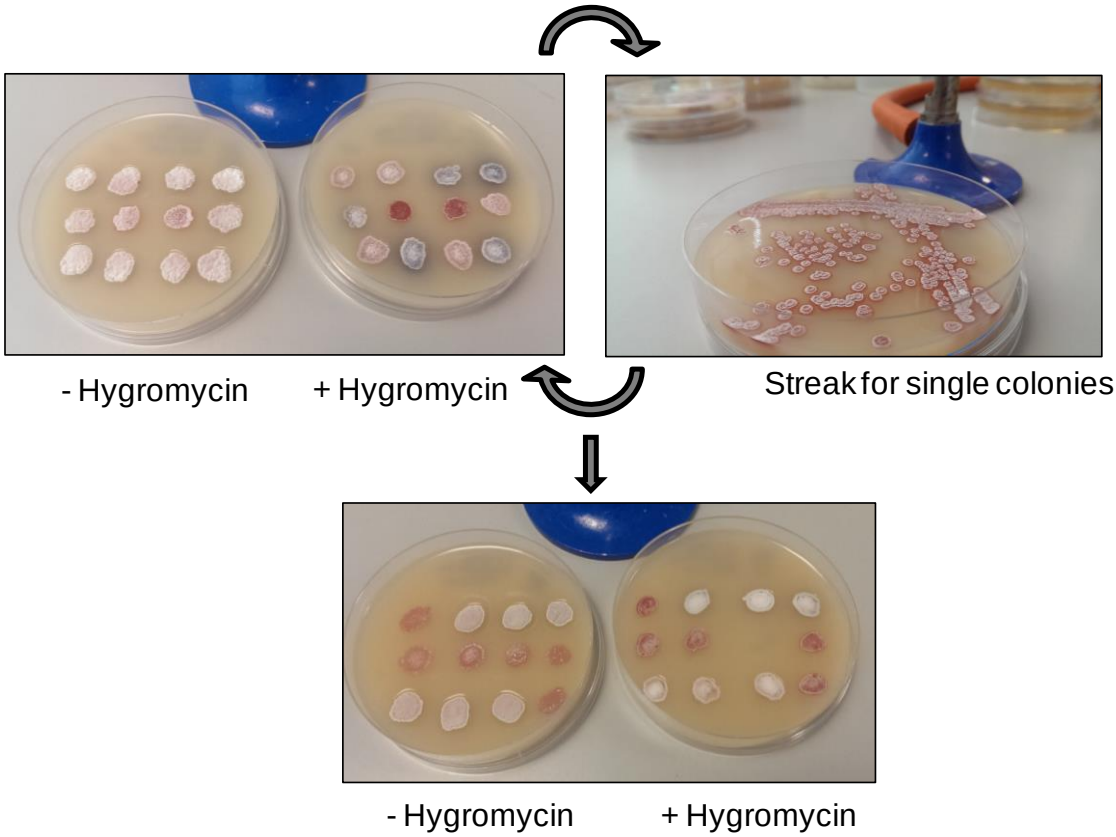



**Figure 6-11** – *S. lividans* TK24/pESAC13A-2H/pER2Δint were passaged on SFM agar to obtain double cross-over mutations. Hygromycin sensitive *S. lividan* TK24/pESAC13A-2H Δ*sav_2268* mutant colonies were obtained after 4 passages.

Primers (for_integrase_screen and rev_integrase_screen, see Table 2-3), which annealed to the pER2Δint backbone, were used to confirm the absence of the pER2Δint construct and hence that double-cross over had occured in *S. lividans* mutant strains. A band corresponding to the expected PCR product of 552 bp was only observed in positive controls of pER2Δint and *S. lividans*/pESAC13A-2H/pER2Δint (Figure 6-12) and were absent in negative controls of pESAC13A-2H and *S. lividans* TK24/pESAC13A-2H. Most significantly, the expected PCR product of 552 bp was absent in all of the four *S. lividans* mutant strains, which confirms the deletion of the pER2Δint backbone and hence, double homologous recombination has occurred in the 4 hygromycin sensitive mutant strains.

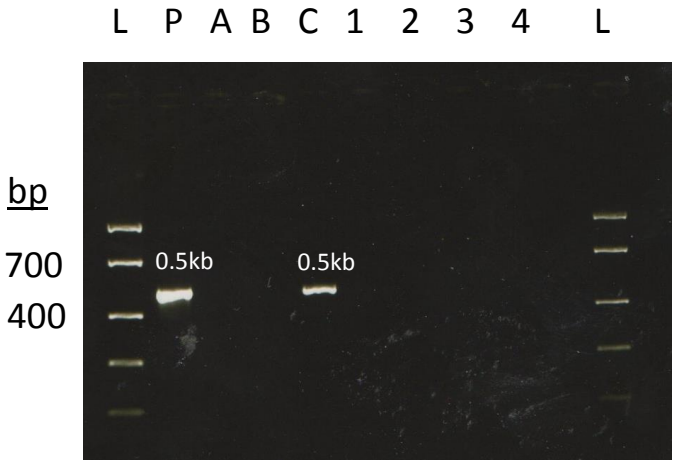



**Figure 6-12** – PCR confirmed the removal of pER2Δint backbone in *S. lividans* mutant strains. A band corresponding to the expected PCR product of 552 bp was obtained in the positive control pER2Δint (Lane P) and hygromycin resistant *S. lividans* TK24/pESAC13A-2H/pER2Δint (Lane C), but is absent in the 4 double cross-over *S. lividans* mutant strains (Lane 1 - 4); thus confirming the successful removal of the pER2Δint backbone. DNA of pESAC13A-2H (Lane A) and *S. lividans* TK24/pESAC13A-2H (Lane B), which do not have the pER2Δint construct, were used as negative controls.

Primers (for_sav_2268_deletion_screen and rev_sav2268_deletion_screen, see Table 2-3) were designed to confirm the deletion of *sav_2268* repressor by annealing to *sav_2266* and *sav_2270* gene (Figure 6-13). A band corresponding to the expected PCR product of 2,536 bp was observed in the 4 *S. lividans* mutant strains; thus, further confirming the successful deletion of *sav_2268* gene (Figure 6-13). The negative controls, pESAC13A-2H and *S. lividans* TK24/pESAC13A-2H do not have the *sav_2268* repressor deleted and a band corresponding to the expected product of 3,235 bp was obtained by PCR (Figure 6-13).

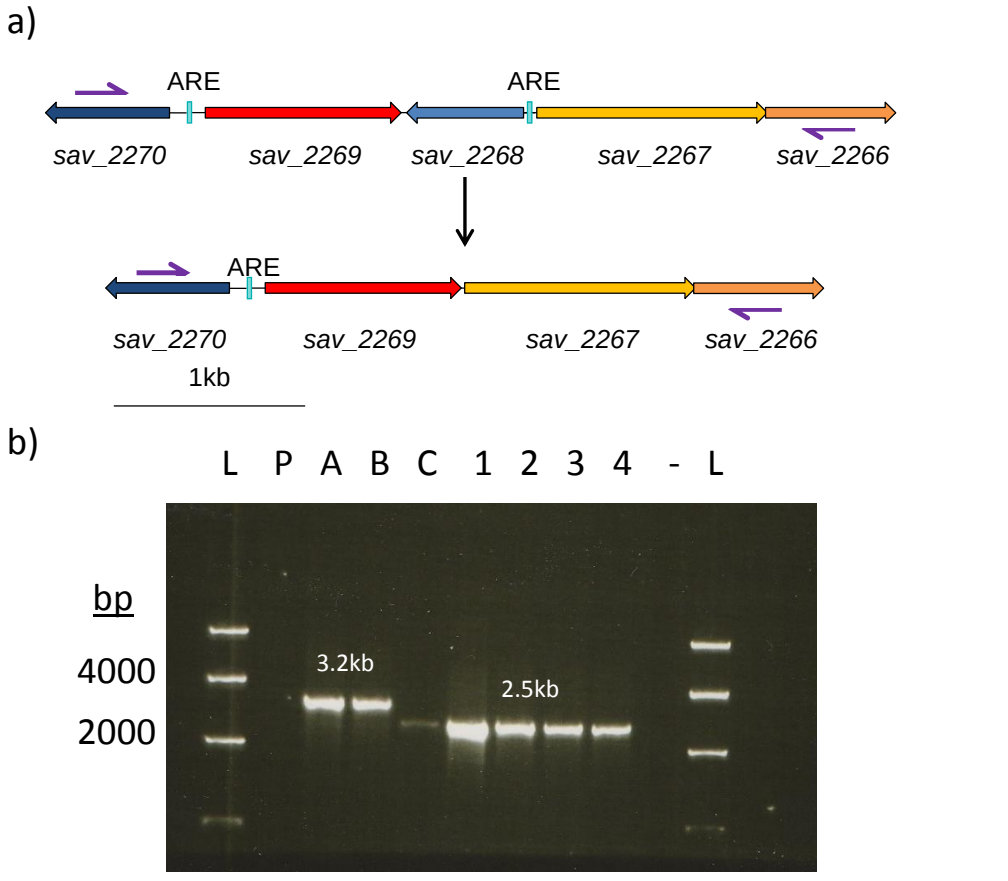



**Figure 6-13** – PCR confirmed the deletion of *sav_2268* gene in the 4 *S. lividans* mutants strains. (A) Primers were designed to anneal to *sav_2266* and *sav_2270* gene. (B) A band consistent with the expected product of 2,536 bp was observed in the 4 *S. lividans* mutant strains (Lane 1 – 4). As expected, no PCR products was observed in the pER2Δint (Lane P) and the expected band corresponding to 3,235 bp was obtained in pESAC13A-2H and *S. lividans* TK24/pESAC13A-2H negative controls.

Finally, the DNA region between *sav_2267* and *sav_2269* was sequenced to confirm the correct deletion of *sav_2268*. Primers (for_avaA-B_screen and rev_avaA-B_screen, see Table 2-3), which flanked *sav_2268*, were used to amplify the *sav_2267* - *sav_2269* intergenic region. PCR product was analysed by DNA sequencing and results confirmed the successful deletion of *sav_2268* in all 4 of the mutant strains (Figure 6-14); errors were observed near the ends of sequencing reads, but were corrected by sequencing of the reverse strand (Figure 6-14).


**Figure 6-14** – DNA sequencing confirmed the successful deletion of *sav_2268* gene in the 4 repressor mutant strains. Sequencing reads of for_avaA-B_screen (54JG71, 54JG73, 54JG75 and 54JG77) and rev_avaA-B_screen (54JG72, 54JG74, 54JG76 and 54JG78) were aligned against the Δ*sav_2268* mutant template. Nucleotide positions of *sav_2267* and *sav_2269* reading frames are highlighted as black arrows and nucleotide matches are highlighted in green.

### 6.3.1. Additional attempts to inactivate the *sav_2268* repressor gene

#### 6.3.1.1. Double homologous recombination in *S. avermitilis*

The same double homologous recombination strategy was attempted in the native organism, *S. avermitilis* (Figure 6-7); however, intergenic conjugations between *E. coli* ET12567/pER2Δint and *S. avermitilis* were unsuccessful in our hands. The deletion of *sav_2268* could therefore not be carried out using the same approach as in the *S. lividans* heterologous host strains. An alternative solution would be to

introduce the pER2Δint construct into *S. avermitilis* through polyethylene glycol assisted protoplast transformation, which involves removing the *Streptomyces* bacteria's cell wall and introducing the DNA material through the cell membrane directly by electroporation. [184,185,186]

#### 6.3.1.2. Double homologous recombination in *S. coelicolor* heterologous host strains

While the predicted azoxy compounds **1** - **3** were produced in *S. coelicolor* host containing the azoxy pathway, deletion of *sav_2268* was expected to result in the overproduction of these metabolites. As with *S. lividans* hosts, the same double homologous recombination strategy (Figure 6-7) was attempted with *S. coelicolor* hosts containing the azoxy pathway. Conjugation between *E. coli*/ET12567/pER2Δint and *S. coelicolor* was successful, and after the 8th passage, one *S. coelicolor* M1152/pESAC13A-2H and two *S. coelicolor* M1152/pESAC13A-19K hygromycin sensitive colonies were obtained. Deletion of *sav_2268* gene in hygromycin sensitive *S. coelicolor* mutant strains remains to be confirmed by PCR and DNA sequencing.

#### 6.3.1.3. PCR targeting approach

Due to the slow growth of *Streptomyces* bacteria, deletion of genes *via* double homologous recombination can take a considerable amount of time to accomplish. Therefore, deletion of genes within *E. coli* hosts can provide a quicker approach to producing mutants. The PCR targeting system, which takes advantage of the λ-RED recombination system, has been successful in generating thousands of mutants in *E. coli* K12, as well as creating knock-outs in gene clusters. [56,131,187] The "PCR targeting system in *Streptomyces coelicolor* A3(2)" protocol was followed in an

attempt to replace the *sav_2268* repressor gene with a kanamycin disruption cassette in *E. coli* (Figure 6-15). [131,182]

Due to the temperature sensitive pKD20 plasmid, introduction of low copy pESAC13A constructs in *E. coli* BW25113/pKD20 by electroporation needed to be carried out at 30 $^{0}$C; however, electroporation of the pESAC13A constructs into an *E. coli* BW25113/pKD20 host were unsuccessful and is most likely due to the inefficiency of introducing large fragile pESAC13A-2H and pESAC13A-19K DNA into the slow growing *E. coli* bacteria. To overcome this problem, pESAC13A-2H and pESAC13A-19K were successfully introduced in empty *E. coli* BW25113 by electroporation performed at 37 $^{0}$C. The pKD20 plasmid, which contains the λ-RED recombination system, was later reintroduced into *E. coli* BW25113/pESAC13A-2H and *E. coli* BW25113/pESAC13A-19K by electroporation and confirmed by PCR (for_PKD20_screen and rev_pKD20_screen, see Table 2-3).

The kanamycin disruption cassette, which was cloned from pIJ777 (for_Target_SAV_2268 and rev_Target_SAV_2268, see Table 2-3), was introduced into arabinose induced *E. coli* BW25113/pKD20/pESAC13A-2H and *E. coli* BW25113/pKD20/pESAC13A-19K by electroporation; however, no kanamycin resistant colonies were obtained and deletion of the *sav_2268* gene was unsuccessful. Alternative kanamycin disruption cassettes were amplified with different primers (for_Target_SAV_2268_2 and rev_Target_SAV_2268_2, see Table 2-3) and had also failed to produce kanamycin resistant colonies. Unsuccessful attempts to delete *sav_2268 via* PCR targeting may be due to the resulting

overexpression of the azoxy biosynthetic genes, which could cause a toxic environment and hence may kill the *E. coli* bacteria.

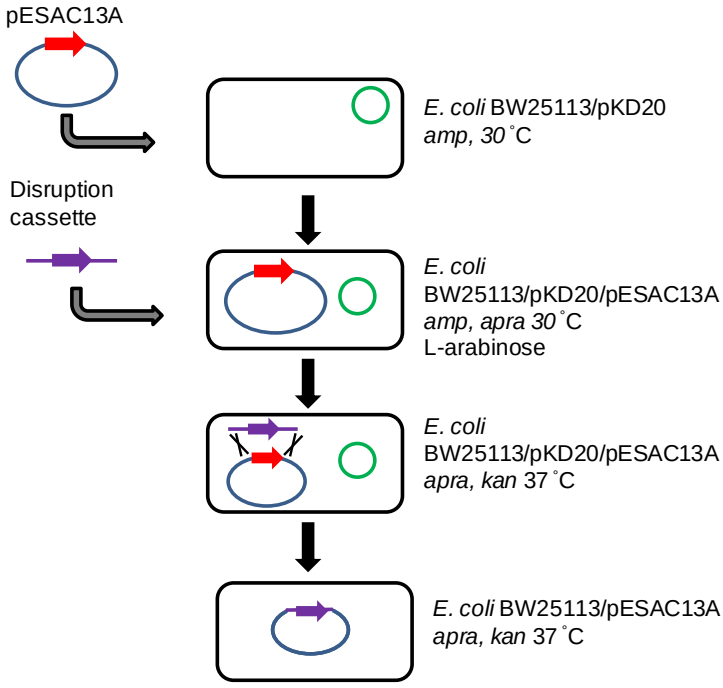



**Figure 6-15** – PCR targeting approach to replacing *sav_2268* gene with a kanamycin disruption cassette. *E. coli* BW25113/pKD20 contains a temperature sensitive and arabinose inducible λ-RED recombination system, which increases recombination probability between extremities of the kanamycin resistance gene and *sav_2268* gene.

### 6.3.2. Attempts to overexpress the *sav_2301* transcriptional activator

Overexpression of *sav_2301* cluster-situated pathway-specific transcriptional activators in the predicted azoxy biosynthetic gene cluster provides an alternative strategy to overproduce the azoxy compounds. Cloning of *sav_2301* gene (for_SAV_2301_2 and rev_SAV_2301_2, see Table 2-3), with appropriate ribosome binding site and flanking restriction sites, was successful; however, the cloned *sav_2301* DNA fragment could not be introduced downstream of *ermE** promoter of pCC4 and pOSV556 vectors *via Hind*III and *Pst*I sites (Figure 6-16).

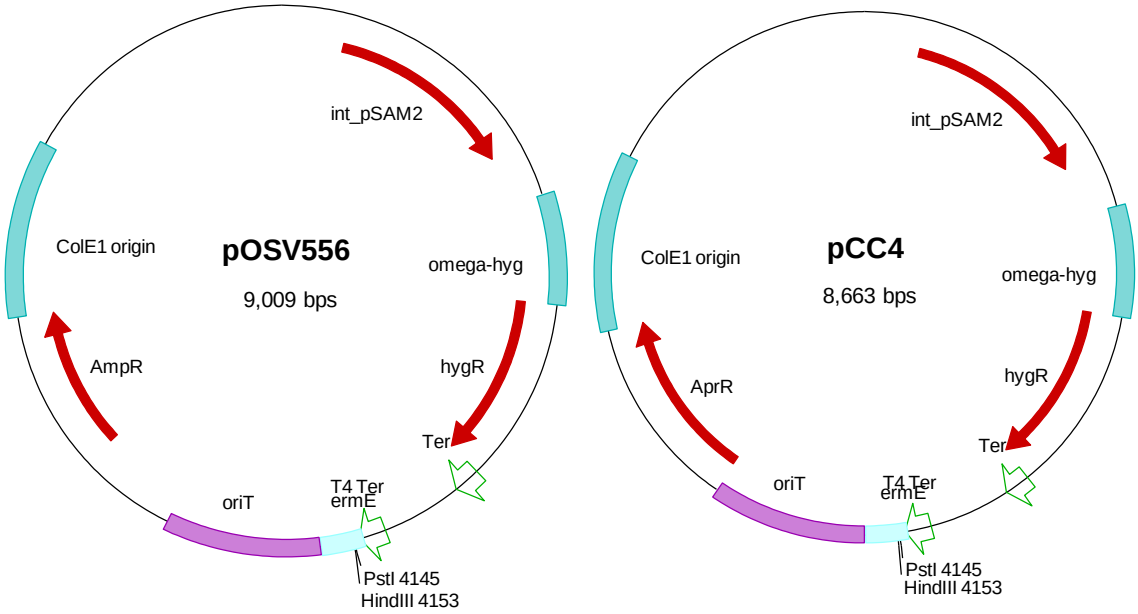



**Figure 6-16** – The pCC4 plasmid is a derivative of the pOSV556 plasmid where the ampicillin resistance gene was replaced with an apramycin resistance gene. *Hind*III and *Pst*I sites are located downstream of the *ermE** constitutive promoter.

A synthetic *sav_2301* gene was also ordered and subcloning into pOSV556 was requested (Life Technologies). The *sav_2301* transcriptional activator gene contains TTA codon between position 46 to 48. BldA encodes for the rare UUA leucine tRNA and regulates many biological processes in *Streptomyces* bacteria including morphological differentiation and antibiotic production. This leucine tRNA is known to be functional only in specific growth phases; [130] hence, the TTA codon was changed to GAG to prevent the BldA-dependent regulation of *sav_2301*. [57]

The altered *sav_2301* transcriptional activator gene was successfully synthesised into pMK-RQ vector and introduced into *E. coli* K12 by Life Technologies (Figure 6-17). However, Life Technologies were also unsuccessful in subcloning the synthetic *sav_2301* gene into pOSV556 vector and suggests there might be toxicity issues with overexpression of the SARP transcriptional activator in *E. coli* hosts; this

may explain why the deletion of *sav_2268* repressor, which is expected to result in *sav_2301* expression, in *E. coli* BW25113 hosts was unsuccessful. To overcome toxicity issues, inducible vectors may provide an alternative approach in controlling toxic effects of the transcriptional activator.

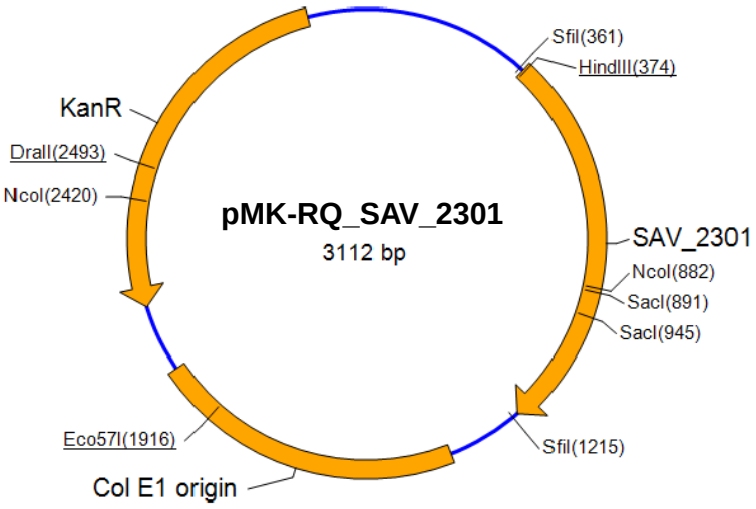



**Figure 6-17** – The altered *sav_2301* gene was synthesised and introduced into the kanamycin resistant pMK-RQ vector by Life Technologies.

## 6.4. Comparitive metabolic profiling of *S. lividans* mutant strains grown in SMMS agar

Comparative metabolic profiling of *S. lividans* mutant strains were carried out to determine whether the production of the predicted azoxy compounds (**1** - **3**) were overproduced.

Different phenotypes were observed when the *S. lividans* mutant strains were grown on SMMS agar (Figure 6-18); the repressor mutants were a significantly darker red in colour when compared with wild type *S. lividans* TK24.

*S. lividans* TK24 also contains the prodiginine gene cluster and the repressor mutant is speculated to have altered production of the prodigines. However, base peak chromatogram (*m/z* = 392.0 - 396.0) was unable to detect the red-pigmented prodiginine compounds. [188]

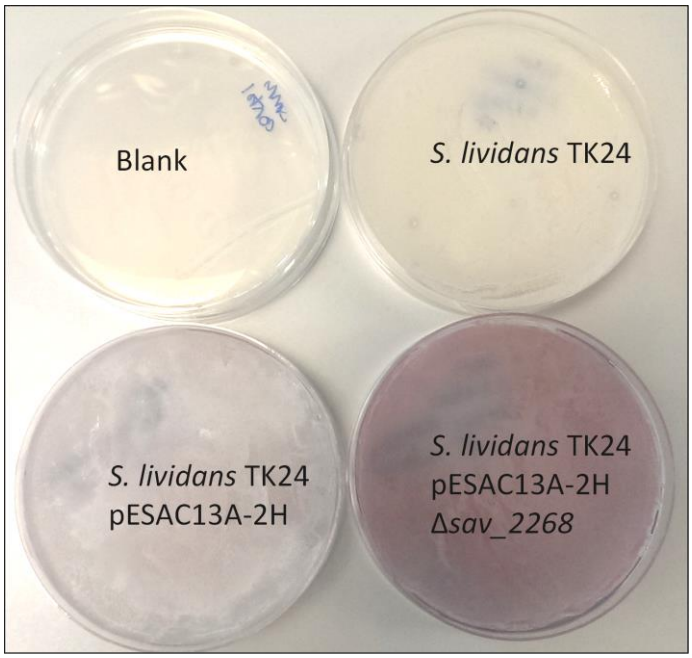



**Figure 6-18** – Different phenotypes were observed in the *S. lividans* heterologous host strains when grown on SMMS agar for 6 days.

Productions of the predicted azoxy compounds (**1** - **3**) was expected to be unlocked in the repressor mutant. Extracted ion chromatograms could not detect either the predicted azoxy compounds (**1** - **3**) (*m/z* = 201.0 and 219.0) or aromatic amine compounds (**4** - **6**) (*m/z* = 268.0, 278.0 and 272.0) in *S. lividans* mutant strains. This suggests the predicted azoxy biosynthetic gene cluster is still tightly regulated at the transcriptional level or specific precursor supply is required for production of azoxy compounds.

Whilst the previously observed compounds (**1** - **6**) were not detected in *S. lividans* mutant strains, LC-MS analysis of the acidified and neutral extracts revealed a series of compounds (**7** - **16**) that were unlocked and overproduced in mutant strains (Figure 6-19). UHPLC-ESI-TOF-MS analyses of compounds **7** - **14** (*m/z* = 270.2) and compounds **15** and **16** (*m/z* = 254.2) suggest that compounds **7** - **16** are related polyunsaturated fatty acid amines (PUFAA) (Table 6-3) and confirms that compounds **7** - **14** and compounds **15** and **16** are isomers.

The exact biological relevance of the observed PUFAA compounds remains unclear; however, it is interesting to note that the production of the predicted PUFAA compounds (**7** - **14**) were only observed in the deletion of *sav_2268* repressor gene in *S. lividans* TK24/pESAC13A-2H heterologous hosts and hence, the predicted PUFAA compounds (**7** - **14**) are speculated to be intermediates or products of the predicted azoxy biosynthetic gene cluster. However one cannot exclude at this stage that deletion of the repressor gene may have unlocked the production of novel cryptic compounds, alternatively, the azoxy biosynthetic genes could be modifying endogenous metabolites produced by *S. lividans* TK24.

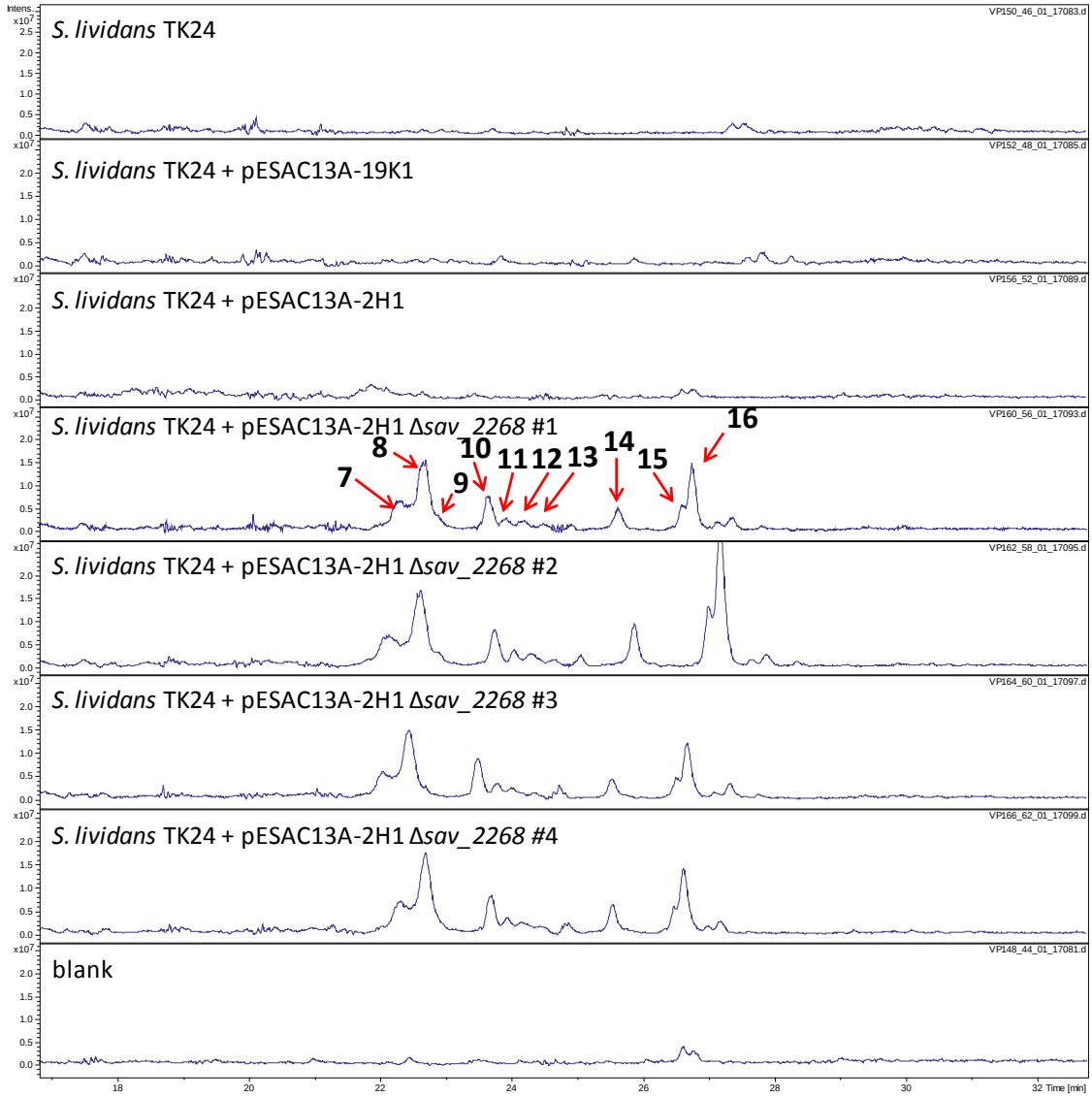



**Figure 6-19** – Comparative metabolic profiling of *S. lividans* mutant strains identified a series of PUFAA compounds (**7** - **16**). Extracted ion chromatograms (*m/z* = 254.0 and 270.0) detected compounds (**7** - **16**) only in the mutant strains grown on SMMS agar.

Detailed analysis of UHPLC-ESI-TOF-MS traces revealed the predicted PUFAA compounds (**7** - **16**), which were overproduced in mutant strains, were almost undetectable in *S. lividans* heterologous host strains (Figure 6-20) and deletion of the *sav_2268* gene in *S. lividans* heterologous host strains has been essential to identify these metabolites (**7** - **16**).

**Table 6-3** – Compounds detected in acidified and neutral organic extraction of *S. lividans* mutant strains grown on SMMS agar.

| # | Retent. Time (min) | $[M+H]^+$ | | | $[M+Na]^+$ | | |
|---|---|---|---|---|---|---|---|
| | | Formula | Observ. | Calc. | Formula | Observ. | Calc. |
| 7 | 22.3 | $C_{15}H_{28}NO_3$ | 270.2070 | 270.2069 | $C_{15}H_{27}NO_3Na$ | 292.1886 | 292.1889 |
| 8 | 22.6 | $C_{15}H_{28}NO_3$ | 270.2071 | 270.2069 | $C_{15}H_{27}NO_3N_a$ | 292.1887 | 292.1889 |
| 9 | 22.8 | $C_{15}H_{28}NO_3$ | 270.2067 | 270.2069 | $C_{15}H_{27}NO_3Na$ | 292.1884 | 292.1889 |
| 10 | 23.9 | $C_{15}H_{28}NO_3$ | 270.2073 | 270.2069 | $C_{15}H_{27}NO_3Na$ | 292.1886 | 292.1889 |
| 11 | 24.1 | $C_{15}H_{28}NO_3$ | 270.2073 | 270.2069 | $C_{15}H_{27}NO_3Na$ | 292.1885 | 292.1889 |
| 12 | 24.5 | $C_{15}H_{28}NO_3$ | 270.2062 | 270.2069 | $C_{15}H_{27}NO_3Na$ | 292.1884 | 292.1889 |
| 13 | 24.8 | $C_{15}H_{28}NO_3$ | 270.2068 | 270.2069 | $C_{15}H_{27}NO_3Na$ | 292.1895 | 292.1889 |
| 14 | 25.6 | $C_{15}H_{28}NO_3$ | 270.2067 | 270.2069 | $C_{15}H_{27}NO_3Na$ | 292.1885 | 292.1889 |
| 15 | 26.5 | $C_{15}H_{28}NO_2$ | 254.2124 | 254.2120 | $C_{15}H_{27}NO_2Na$ | 276.1941 | 276.1939 |
| 16 | 25.7 | $C_{15}H_{28}NO_2$ | 254.2125 | 254.2120 | $C_{15}H_{27}NO_2Na$ | 276.1941 | 276.1939 |

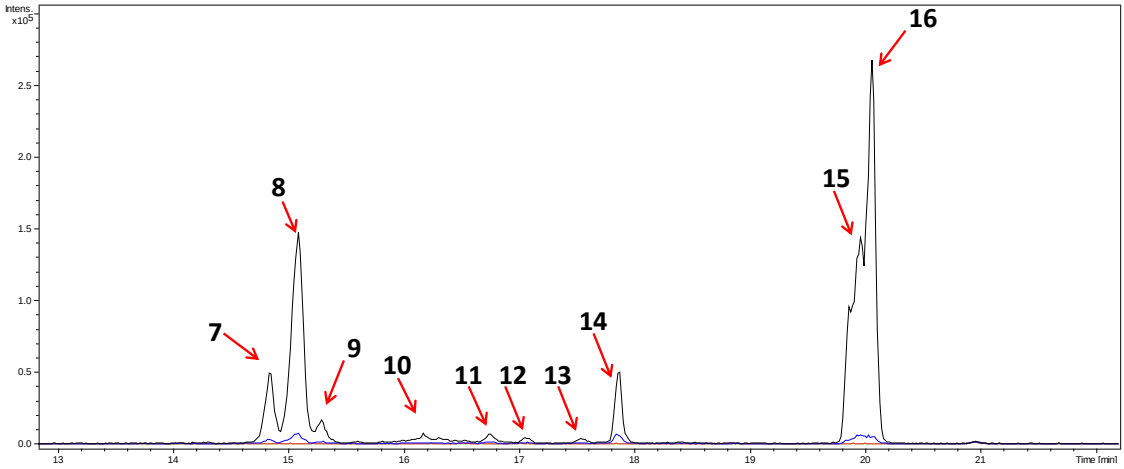



**Figure 6-20** – Extracted ion chromatogram (*m/z* = 270.2067 and 254.2124) of UHPLC-ESI-TOF-MS traces revealed the predicted PUFAA compounds (**7** - **16**) are produced in minute quantities in *S. lividans* TK24/pESAC13A-2H (blue trace). Compounds **7** - **16** were overproduced in *S. lividans* mutant strains (black trace) and were absent in the wild type *S. lividans* TK24 (red trace).

## 6.5. Comparitive metabolic profiling of *S. lividans* mutant strains in VPM

Previous work on molecular characterisation and analysis of the valanimycin biosynthetic gene cluster by Garg and co-workers had similarly captured the valanimycin biosynthetic gene cluster from *S. viridifaciens* on a cosmid and subsequently introduced the cosmid construct into *S. lividans* TK24 as a heterologous hosts. To determine whether low precursor supply was responsible for the absence of azoxy compounds in SMMS, comparative metabolic profiling of *S. lividans* mutant strains grown in VPM was performed.[162]

The predicted azoxy compounds (**1** - **3**) were finally detected in *S. lividans* mutant strains grown in VPM by LC-MS (Figure 6-21). In addition to the predicted azoxy compounds (**1** - **3**), base peak chromatogram (*m/z* = 200 - 300) also detected AHFCA signalling molecules (**17** and **18**); the identification of the predicted azoxy compounds (**1** - **3**) and AHFCA signalling molecules (**17** and **18**) were confirmed by UHPLC-ESI-TOF-MS analysis (Table 6-4).

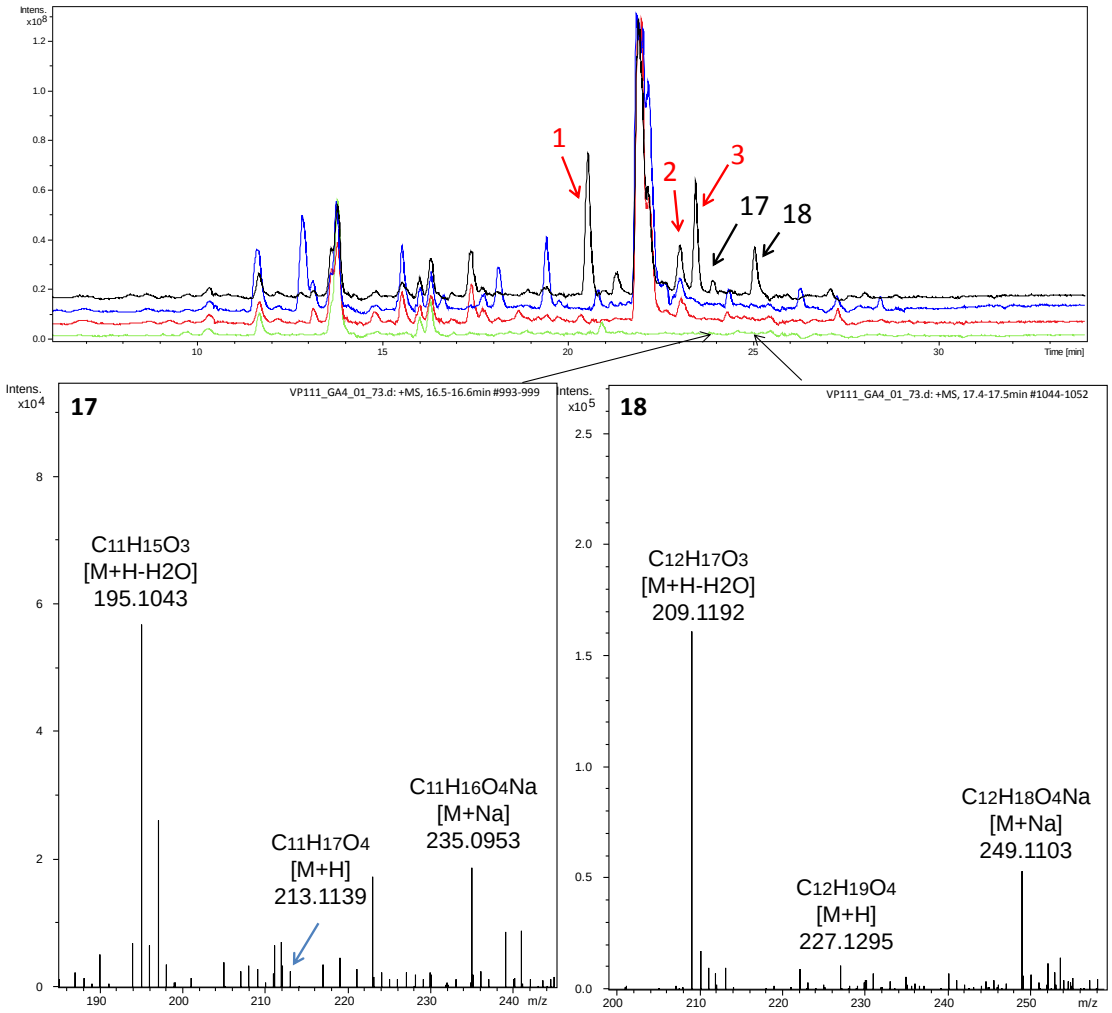



**Figure 6-21** – Production of the predicted azoxy compounds (**1** - **3**) and AHFCAs (**17** - **18**) were unlocked in *S. lividans* mutant strains (black trace), and are absent in wild type *S. lividans* TK24 (red trace), *S. lividans* TK24/pESAC13A-2H1 (blue trace) and blank (green trace). Base peak chromatogram (*m/z* = 200 - 300) detected AHFCA signalling molecules (**17** and **18**); the high resolution mass spectra data are shown for these AHFCA compounds.

**Table 6-4** – AHFCA compounds detected in acidified organic extracts of *S. lividans* mutant strains grown in VPM.

| # | Retention Time (min) | $[M+H]^+$ | | | $[M+Na]^+$ | | |
|---|---|---|---|---|---|---|---|
| | | Formula | Observed | Calculated | Formula | Observed | Calculated |
| 17 | 24.0 | $C_{11}H_{17}O_4$ | 213.1138 | 213.1127 | $C_{11}H_{16}O_4Na$ | 235.0953 | 235.0946 |
| 18 | 25.0 | $C_{12}H_{19}O_4$ | 227.1294 | 227.1283 | $C_{12}H_{17}O_4Na$ | 249.1103 | 248.1025 |

### 6.5.1. Characterisation of *S. avermitilis* AHFCA signalling molecules

The presence of AHFCAs were further supported by their mass spectra. Indeed these molecules present a characteristic mass spectrum pattern. Intense molecular ion peaks for $[M+H-H_2O]^+$ and $[M+Na]^+$, but very weak peak for $[M+H]^+$ patterns were obtained for compounds **17** and **18** (Figure 6-21 and Table 6-4). [56]

Two main structural isomers could be envisaged for each of the predicted AHFCAs (**17** and **18**), which have *m/z* value for $[M+H]^+$ of 213.1 and 227.1 respectively (Figure 6-22). To determine their precise structure, authentic standard of AHFCAs consistent with the *m/z* values observed, i.e. AHFCA3, AHFCA5, AHFCA6 and AHFCA7 (obtained from Zhou, S., Challis Group - Warwick) were analysed by LC-MS in parallel with organic extracts.

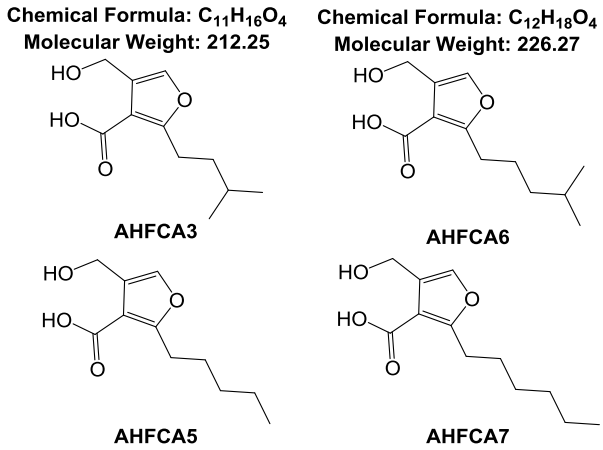



**Figure 6-22** – Structural isomers of AHFCA3, AHFCA5, AHFCA6 and AHFCA7 signalling molecules.

Extracted ion chromatogram for AHFCA signalling molecules (*m/z* = 195.0, 209.0, 235.0 and 249.0) in diluted organic extracts revealed an unexpected small shoulder peak for compounds **17** and **18** and line up with the authentic standards (Figure

6-23); this implies compounds **17-1**, **17-2**, **18-1** and **18-2** are AHFCA3, AHFCA5, AHFCA6 and AHFCA7 respectively.

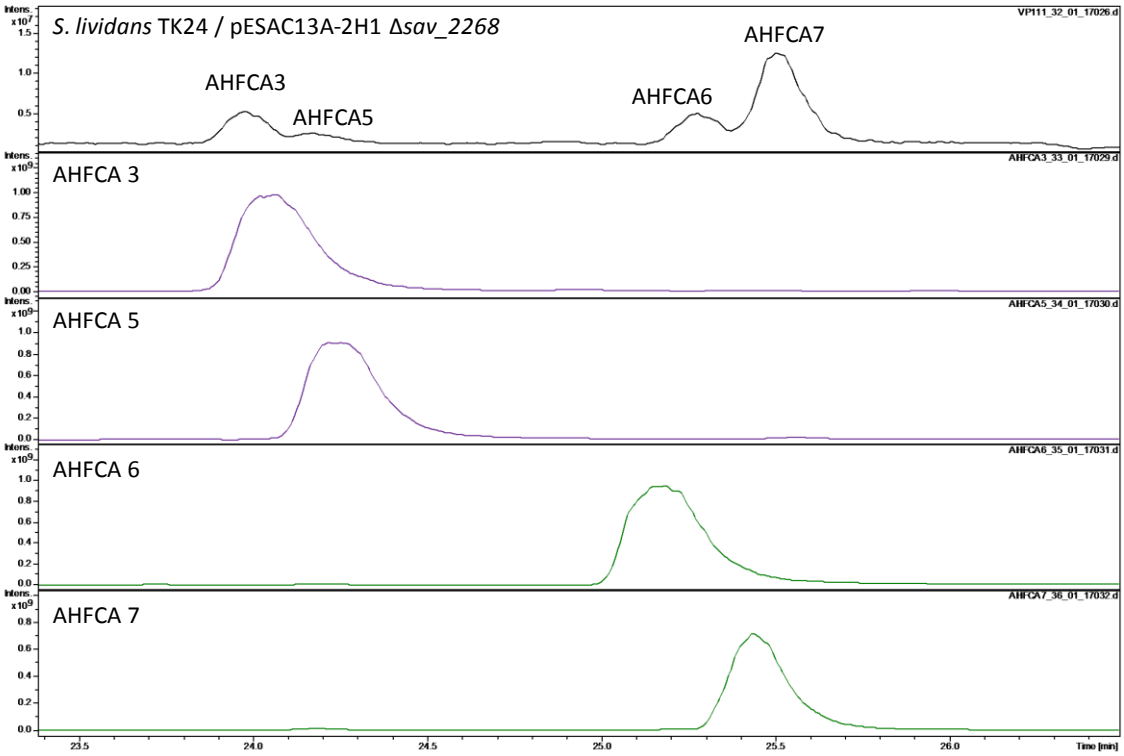



**Figure 6-23** – *S. lividans* organic extracts were compared with AHFCA3, AHFCA5, AHFCA6 and AHFCA7 authentic standards. Extracted ion chromatogram (*m/z* = 195.0, 209.0, 235.0 and 249.0) showing AHFCA3, AHFCA5, AHFCA6 and AHFCA7 were overproduced in *S. lividans* mutant strains.

Due to high concentration of authentic standards and the use of low resolution LC-MS, the retention time of the AHFCA standards did not line up exactly with that observed for *S. avermitilis* AHFCAs (**17-1**, **17-2**, **18-1** and **18-2**) (Figure 6-23). Co-injection of authentic standards with organic extracts would confirm the nature of *S. avermitilis* AHFCA compounds. In addition UHPLC-MS instead of LC-MS analysis would provide a greater retention time resolution. Finally, comparison of tandem MS fragmentation patterns would also further confirm the identity of the observed compounds (**17-1**, **17-2**, **18-1** and **18-2**).

### 6.5.2. Identification of predicted polyunsaturated fatty acid amine compounds

Different phenotypes between *S. lividans* strains were also observed when grown in liquid VPM for 6 days (Figure 6-24). The culture of wild type *S. lividans* TK24 and *S. lividans* TK24/pESAC13A-2H1 were deep blue and red in colour (Figure 6-24) and implies actinorhodin and undecylprogiosin are overproduced respectively. The base peak chromatogram was able to detect the overproduction of prodigiosins (*m/z* = 392 - 396) in *S. lividans* TK24/pESAC13A-2H1; however, the base peak chromatogram for actinorhodin (*m/z* = 632 - 636) could not identify the specific compounds. The mutant strains were neither blue or red and implies production of actinorhodin and prodigiosins were abolished; the abolished production of prodigosin (*m/z* = 392 for streptorubin B and 394 for undecylprodigiosin) was confirmed by UHPLC-ESI-TOF-MS analysis.

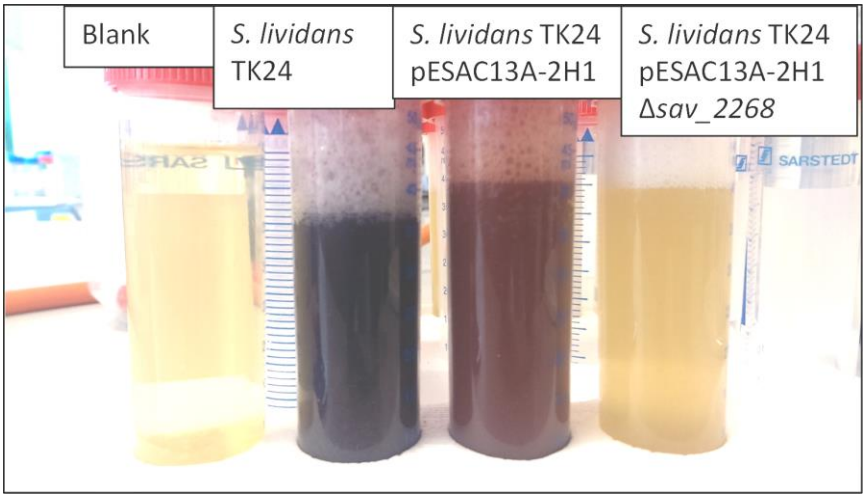



**Figure 6-24** – Different phenotypes were observed when *S. lividans* strains were grown in VPM for 6 days.

Comparative metabolic profiling of acidified organic extracts revealed a series of compounds (**19 - 45**) that were only produced in *S. lividans* heterologous host and

mutant strains (Figure 6-25 and Figure 6-26) using VPM culture medium but not SMM. The molecular formulae generated by UHPLC-ESI-TOF-MS analysis for compounds **19** - **45** suggests they are related PUFAA compounds (Table 6-5). For each *m/z* value corresponding to PUFAAs, a number of isomers were detected by UHPLC-ESI-TOF-MS analysis; for instance, compound **22** has a molecular formula of $C_{16}H_{27}NO_2$ and the extracted ion chromatogram (*m/z* = 266.2113 and 288.1939) highlighted another 10 compounds (**23** - **33**) with the same molecular formula eluating at different retention times (Figure 6-25 and Table 6-5).

Surprisingly, the production of the PUFAA compounds (**19** - **45**) were altered in *S. lividans* mutant strains; compounds **19** - **38** were abolished and compounds **39** - **45** were significantly reduced in mutant strains (Figure 6-25 and Figure 6-26). It is difficult to rationalise why the production of some of the PUFAA compounds (**39** - **45**) are abolished in the repressor mutants.

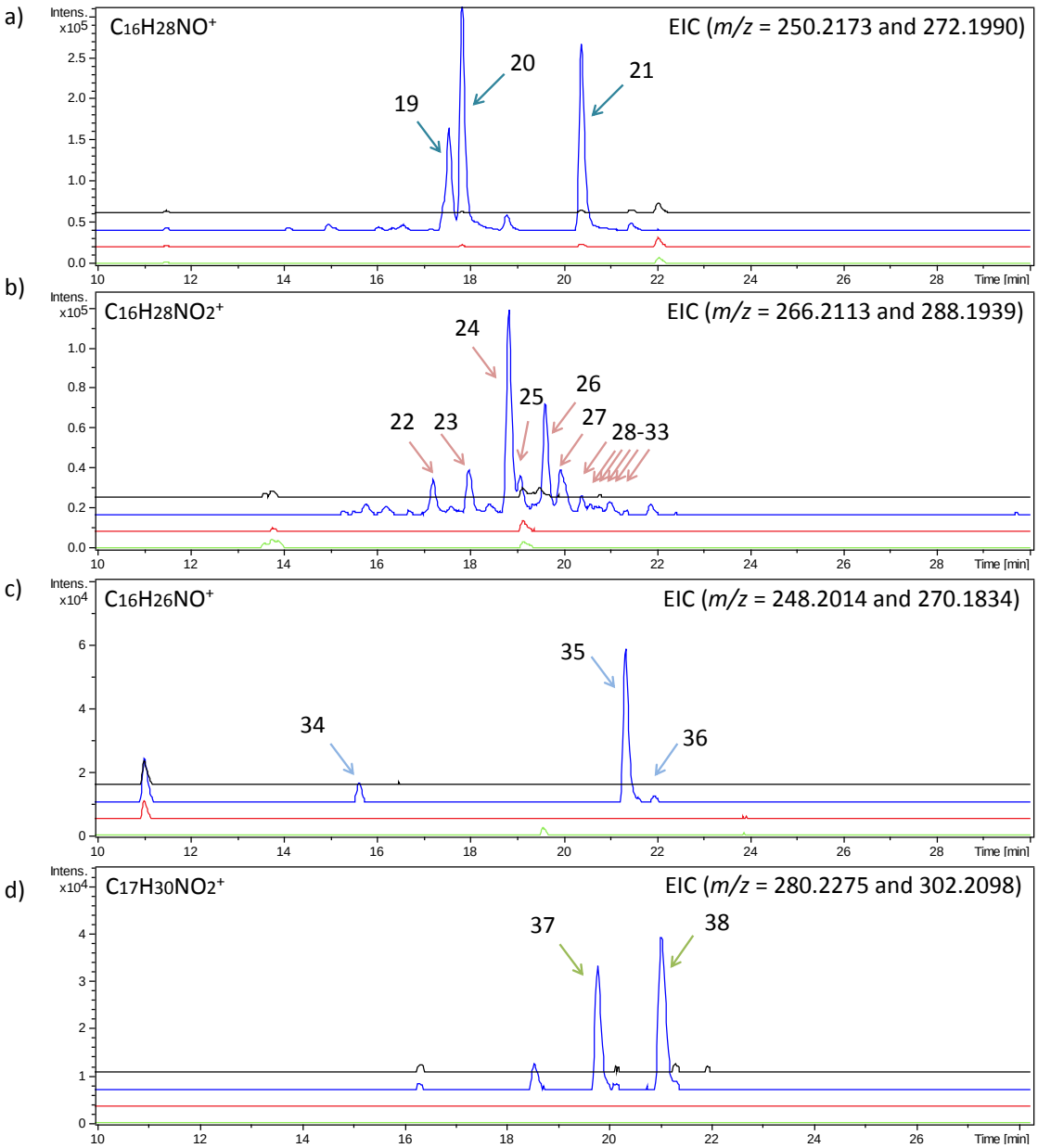



**Figure 6-25** – From UHPLC-ESI-TOF-MS traces, extracted ion chromatogram detected compounds **19** - **38** from *S. lividans* TK24/pESAC13A-2H1 (blue trace). Compounds **19** - **38** were absent in wild type *S. lividans* TK24 (red trace), blank (green trace) and mutant strains (black trace). Arrows are colour-coded according to Table 6-5.

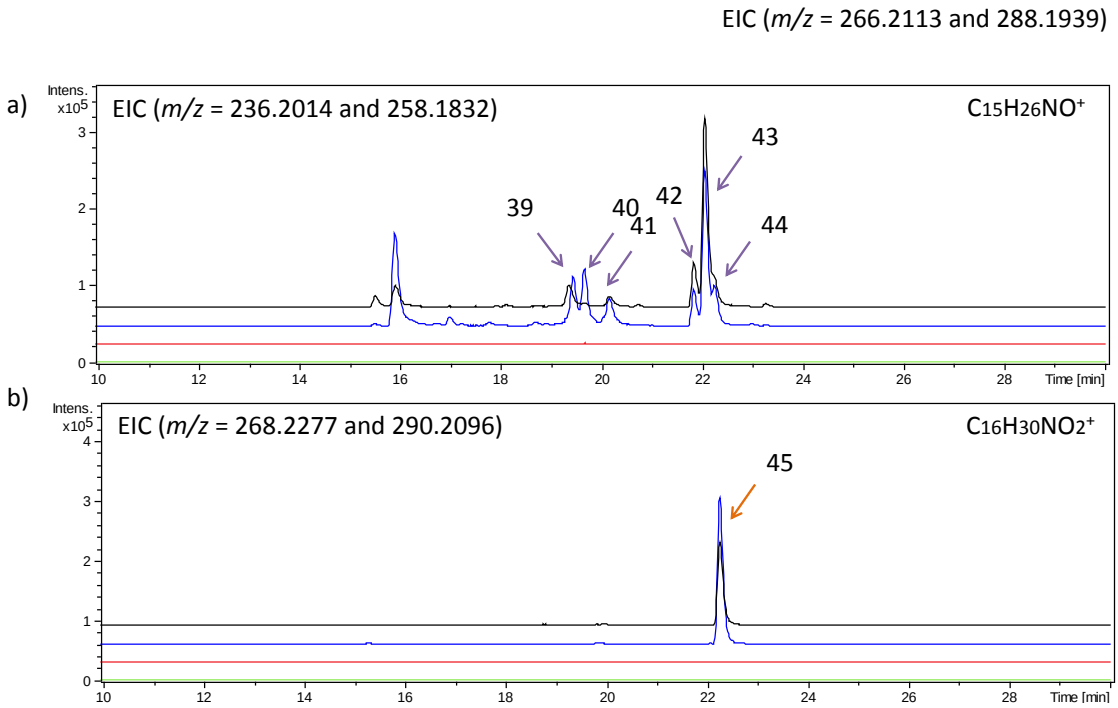



**Figure 6-26** – From UHPLC-ESI-TOF-MS traces, extracted ion chromatogram detected the compounds (**39** - **45**) in *S. lividans* TK24/pESAC13A 2H1 (blue trace) and were significantly reduced in *S. lividans* mutant strains (black trace). Compounds (**19** - **38**) were absent in wild type *S. lividans* TK24 (red trace) and blank (green trace). Arrows are colour-coded according to Table 6-5.

**Table 6-5** – Compounds detected in acidified organic extraction of *S. lividans* heterologous host strains grown in VPM for 6 days.

| # | Retent. Time (min) | $[M+H]^+$ | | | $[M+Na]^+$ | | |
|---|---|---|---|---|---|---|---|
| | | Formula | Observed | Calculated | Formula | Observed | Calculated |
| 19 | 24.9 | $C_{16}H_{28}NO$ | 250.2168 | 250.2171 | $C_{16}H_{27}NONa$ | 272.1954 | 272.1990 |
| 20 | 25.1 | $C_{16}H_{28}NO$ | 250.2168 | 250.2171 | $C_{16}H_{27}NONa$ | 272.1981 | 272.1990 |
| 21 | 28.2 | $C_{16}H_{28}NO$ | 250.2174 | 250.2171 | $C_{16}H_{27}NONa$ | 272.1990 | 272.1990 |
| 22 | 24.6 | $C_{16}H_{28}NO_2$ | 266.2113 | 266.2120 | $C_{16}H_{27}NO_2Na$ | 288.1934 | 288.1939 |
| 23 | 25.1 | $C_{16}H_{28}NO_2$ | 266.2114 | 266.2120 | $C_{16}H_{27}NO_2Na$ | 288.1935 | 288.1939 |
| 24 | 25.8 | $C_{16}H_{28}NO_2$ | 266.2116 | 266.2120 | $C_{16}H_{27}NO_2Na$ | 288.1933 | 288.1939 |
| 25 | 26.1 | $C_{16}H_{28}NO_2$ | 266.2113 | 266.2120 | $C_{16}H_{27}NO_2Na$ | 288.1935 | 288.1939 |
| 26 | 26.4 | $C_{16}H_{28}NO_2$ | 266.2117 | 266.2120 | $C_{16}H_{27}NO_2Na$ | 288.1935 | 288.1939 |
| 27 | 26.7 | $C_{16}H_{28}NO_2$ | 266.2118 | 266.2120 | $C_{16}H_{27}NO_2Na$ | 288.1935 | 288.1939 |
| 28 | 27.4 | $C_{16}H_{28}NO_2$ | 266.2119 | 266.2120 | $C_{16}H_{27}NO_2Na$ | 288.1936 | 288.1939 |
| 29 | 27.6 | $C_{16}H_{28}NO_2$ | 266.2118 | 266.2120 | $C_{16}H_{27}NO_2Na$ | 288.1944 | 288.1939 |
| 30 | 27.9 | $C_{16}H_{28}NO_2$ | 266.2120 | 266.2120 | $C_{16}H_{27}NO_2Na$ | 288.1948 | 288.1939 |
| 31 | 28.6 | $C_{16}H_{28}NO_2$ | 266.2118 | 266.2120 | $C_{16}H_{27}NO_2Na$ | 288.1948 | 288.1939 |
| 32 | 28.9 | $C_{16}H_{28}NO_2$ | 266.2188 | 266.2120 | $C_{16}H_{27}NO_2Na$ | 288.1938 | 288.1939 |
| 33 | 29.1 | $C_{16}H_{28}NO_2$ | 266.2116 | 266.2120 | $C_{16}H_{27}NO_2Na$ | 288.1939 | 288.1939 |
| 34 | 22.9 | $C_{16}H_{26}NO$ | 248.2021 | 248.2014 | $C_{16}H_{25}NONa$ | 270.1826 | 270.1834 |
| 35 | 28.7 | $C_{16}H_{26}NO$ | 248.2019 | 248.2014 | $C_{16}H_{25}NONa$ | 270.1836 | 270.1834 |
| 36 | 29.0 | $C_{16}H_{26}NO$ | 248.2015 | 248.2014 | $C_{16}H_{25}NONa$ | 270.1828 | 270.1834 |
| 37 | 27.2 | $C_{17}H_{30}NO_2$ | 280.2274 | 280.2277 | $C_{17}H_{29}NO_2Na$ | 302.2100 | 302.2096 |
| 38 | 28.8 | $C_{17}H_{30}NO_2$ | 280.2275 | 280.2277 | $C_{17}H_{29}NO_2Na$ | 302.2098 | 302.2096 |
| 39 | 26.7 | $C_{15}H_{26}NO$ | 236.2014 | 236.2014 | $C_{15}H_{25}NONa$ | 258.1832 | 258.1834 |
| 40 | 27.1 | $C_{15}H_{26}NO$ | 236.2014 | 236.2014 | $C_{15}H_{25}NONa$ | 258.1830 | 258.1834 |
| 41 | 27.6 | $C_{15}H_{26}NO$ | 236.2015 | 236.2014 | $C_{15}H_{25}NONa$ | 258.1836 | 260.1990 |
| 42 | 28.3 | $C_{15}H_{26}NO$ | 236.2017 | 236.2014 | $C_{15}H_{25}NONa$ | 258.1833 | 258.1834 |
| 43 | 29.1 | $C_{15}H_{26}NO$ | 236.2018 | 236.2014 | $C_{15}H_{25}NONa$ | 258.1834 | 258.1834 |
| 44 | 29.6 | $C_{15}H_{26}NO$ | 236.2018 | 236.2014 | $C_{15}H_{25}NONa$ | 258.1835 | 258.1834 |
| 45 | 29.1 | $C_{16}H_{30}NO_2$ | 268.2279 | 268.2277 | $C_{16}H_{29}NO_2Na$ | 290.2096 | 290.2096 |

## 6.6. Prediction of precursors for the azoxy compounds

From genome mining analyses discussed in Chapter 5.2, the azoxy compounds produced in *S. avermitilis* were proposed to incorporate precursor of polyketide origin (Figure 5-5). Detailed annotation of the azoxy biosynthetic genes in *S. avermitilis* show *sav_2282* to *sav_2280* encodes for polyketide synthase enzymes and appears to only contain a single extension module with ketoreductase and dehydratase domains (Figure 6-27).

With the detection of polyunsaturated fatty acid amine and aromatic amine compounds (**4** - **16** and **19** - **45**) in *S. lividans* heterologous host and mutant strains, the 3 PKS enzymes are proposed to act in an iterative manner to reuse domains and assemble a long fatty acid-like polyketide chain. Presence of a nitrogen atom on the predicted PUFAA compounds (**7** – **16** and **19** - **45**) and aromatic amine compounds (**4** - **6**) could be explained by *sav_2279* and *sav_2307* gene products, which are proposed to encode for a long chain fatty acid-CoA ligase and acyl-CoA synthetase respectively; both contain adenylation domains (Figure 6-27). Adenylation domains often activates amino acids using ATP, the activated amino acid could be condensed to the polyketide derived precursor. The sequence analysis of SAV_2279 and SAV_2307 adenylation domains were unable to predict the substrate specificity. [51] After the assembly of the polyketide chain, SAV_2277 thioesterase is proposed to cleave the PUFAA compounds from the ACP.

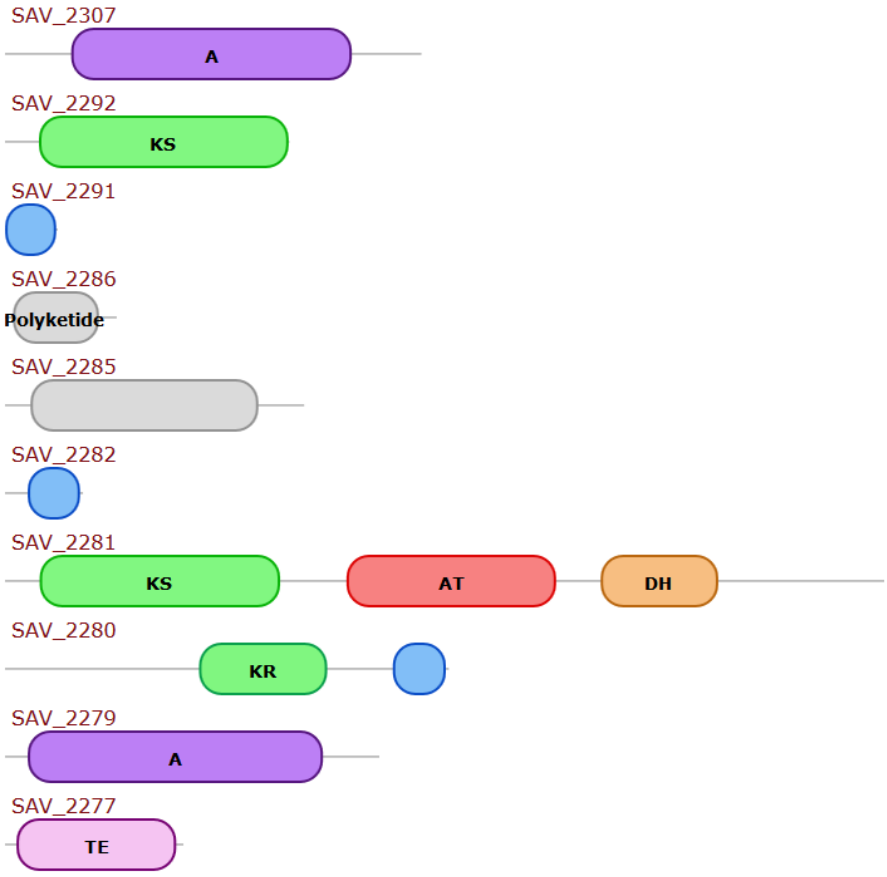



**Figure 6-27** – Detailed antiSMASH annotation of biosynthetic genes from the predicted azoxy biosynthetic gene cluster in *S. avermitilis*. Abbreviations are as follows: A – adenylation; KS – ketosynthase; AT – acyltransferase; DH – dehydrogenase; KR – ketoreductase; TE – thioesterase. Blue circles are acyl carrier proteins.

Furthermore, elaiomycins, which are azoxy compounds produced by *S.* sp. strain HKI0708, were shown to incorporate isotope-labelled *N*-octylamine.[175] In addition, azoxyalkenes, produced by *S. rochei* 7434AN4, were also reported to incorporate *N*-hexylamine.[189] With the idea that the predicted PUFAA compounds might be involved in the biosynthesis of the predicted azoxy compounds, the expected molecular formula and molecular weight from the incorporation of the observed compounds (**4** - **16** and **19** - **45**) in the biosynthesis of the final azoxy compounds are detailed in Table 6-6; however, azoxy compounds consistent with these molecular formulae could not be identified by LC-MS.

**Table 6-6** – Expected azoxy compounds generated from incorporation of the observed compounds as precursors

| Compound | R group | Expected Formula | Expected MW |
|---|---|---|---|
| 1 | $C_6H_{15}O$ | $C_9H_{18}N_2O_4$ | 218 |
| 2 - 3 | $C_6H_{13}$ | $C_9H_{16}N_2O_3$ | 200 |
| 4 - 5 | $C_{19}H_{16}$ | $C_{21}H_{19}N_2O_3$ | 347 |
| 6 | $C_{19}H_{18}$ | $C_{21}H_{21}N_2O_3$ | 349 |
| 7 - 14 | $C_{15}H_{24}O_2$ | $C_{18}H_{27}N_2O_5$ | 351 |
| 15 - 16 | $C_{15}H_{24}O$ | $C_{18}H_{27}N_2O_4$ | 335 |
| 19 - 21 | $C_{16}H_{24}$ | $C_{19}H_{27}N_2O_3$ | 331 |
| 22 - 33 | $C_{16}H_{24}O$ | $C_{19}H_{27}N_2O_4$ | 347 |
| 34 - 36 | $C_{16}H_{22}$ | $C_{19}H_{25}N_2O_3$ | 329 |
| 37 - 38 | $C_{17}H_{26}O$ | $C_{20}H_{29}N_2O_4$ | 361 |
| 39 - 44 | $C_{15}H_{22}$ | $C_{18}H_{25}N_2O_3$ | 317 |
| 45 | $C_{16}H_{26}O$ | $C_{19}H_{29}N_2O_3$ | 333 |

The valanimycin compounds were reported to be extremely unstable and structural determination of valanimycin was impossible without the formation of stable derivatives. A valanimycin ammonia adduct was found to be the most stable and MS analysis and structure elucidation by NMR spectroscopy were subsequently performed on this stable derivative.[155] The azoxy compounds in *S. avermitilis*, which were proposed to incorporate the PUFAA compounds (**4 - 16** and **19 - 45**), may be unstable and undetectable by LC-MS (Table 6-6). Organic extracts from *S. coelicolor, S. lividans* heterologous host or mutant strains could be treated similarly to the valanimycin isolation procedures by Yamato and co-workers to obtain stable derivatives and they could then be characterised by MS and NMR spectrum analysis. [155] The predicted azoxy compounds are known to incorporate L-serine and a genomisotopic guided approach could be used to detect our azoxy compounds of interest.[55]

The predicted azoxy biosynthetic gene cluster in *S. avermitilis* also contains other enzymes including an intriguing ornithine aminotransferase, which is encoded by

*sav_2285* (Figure 6-1). The exact biological role of the ornithine aminotransferase enzyme in *S. avermitilis* is unclear in the biosynthesis of the novel azoxy compounds. Ornithine aminotransferase is involved in creating ornithine from amino acid precursor (Figure 6-28); hence, feeding of isotope-labelled ornithine would provide insights into the biosynthesis of the azoxy compounds.

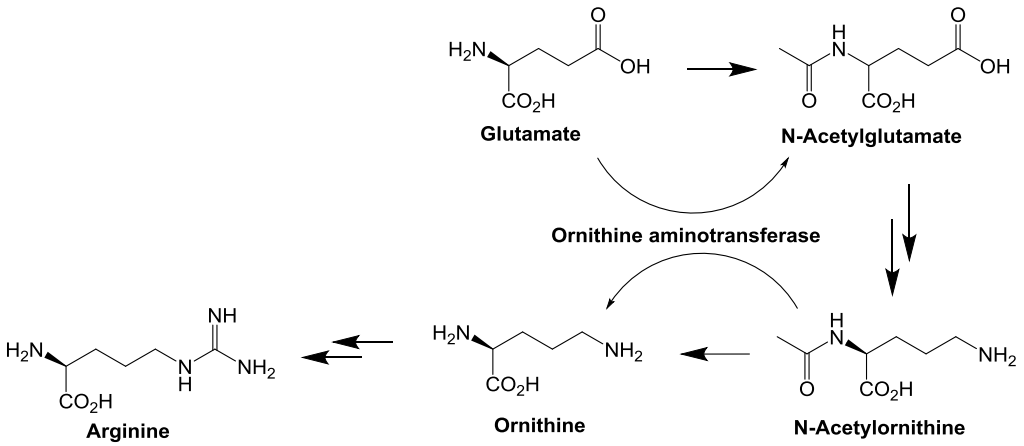



**Figure 6-28** – Orf6 ornithine aminotransferase enzyme from *S. clavuligerus* is involved in one of the key steps in the biosynthesis of arginine by catalysing the conversion of glutamate to ornithine via *N*-acetylglutamate semialdehyde. Image adapted from reference 190.

The azoxy compounds (**1** - **3**), which were detected in *S. coelicolor* heterologous host and *S. lividans* mutant strains, may not be the true metabolic product of the *S. avermitilis* azoxy biosynthetic gene cluster; this is supported by the size of the gene cluster as well as the nature of some of the genes present in the silent and cryptic azoxy biosynthetic gene cluster.

## 6.7. Antimicrobial assays to determine the biological activity of the predicted azoxy compounds

The family of azoxy compounds have demonstrated a broad range of biological activities including antifungal, antibacterial and antitumour

properties. [155,175,191,192,193] The azoxy compounds are highly cytotoxic and carcinogenic, and are proposed to act as alkylating agents by specifically reacting with DNA and resulting in strand breaks. Valanimycin was reported to be active against both Gram-positive and Gram-negative bacteria, but was also shown to be more effective against DNA-repair deficient bacteria, which further supports the hypothesis that the azoxy compounds primarily targets DNA. [194] To determine the biological activity of the observed azoxy compounds (**1** - **3**), crude organic extracts were tested against *E. coli*, *B. subtilis*, *S. aureus*, *S. coelicolor* and *S. lividans*.

Crude extract of *S. coelicolor* heterologous host strains grown on SMM agar or VPM did not exhibit any differences in antimicrobial activity when compared to the wild type *S. coelicolor* M1152 crude extracts. This implies that the predicted azoxy compounds (**1** - **3**) and aromatic amine compounds (**4** - **6**) are not active against the Gram-positive and Gram-negative bacteria tested at the concentration of extract used. Scaled up production and purification of specific compounds, rather than crude extracts, could be used to investigate the biological activities of the compounds of interest.

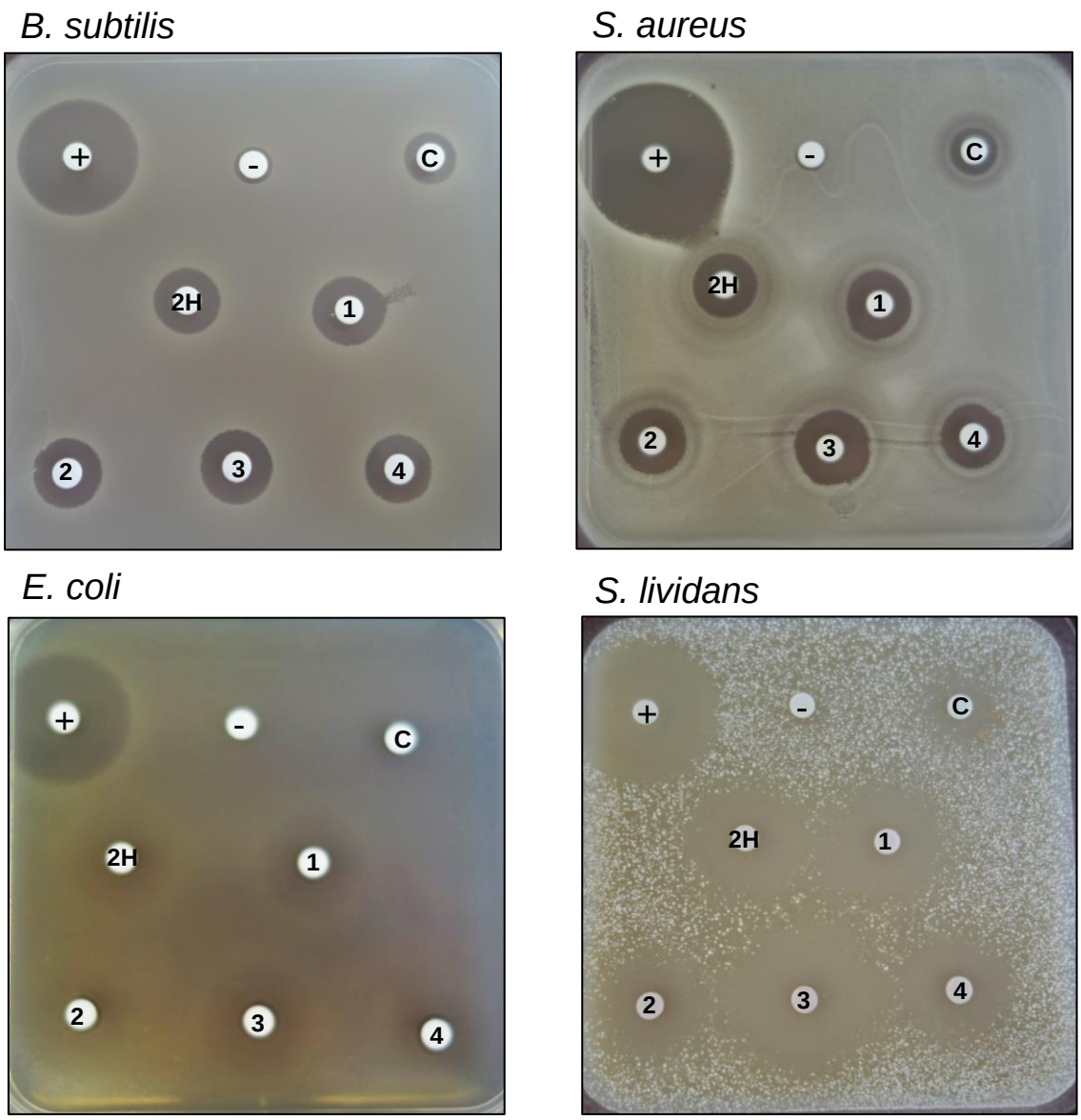



**Figure 6-29** – Zone of inhibition assays to determine antimicrobial activity of crude extract of *S. lividans* mutant strains grown in VPM. Crude extract of *S. lividans* TK24/pESAC13A-2H1 (2H) and mutant strains (1 - 4) were active against *B. subtilis*, *S. aureus*, and *S. lividans* and weakly active against *E. coli*. Apramycin (+), methanol (-), wild type *S. lividans* TK24 crude extract (c) were used as controls.

Of significant interest, crude organic extracts of *S. lividans* TK24/pESAC13A-2H grown in VPM showed significant antimicrobial activity against both Gram-positive and Gram-negative bacteria when compared to extracts from *S. lividans* TK24 wild type (Figure 6-29). The same area of inhibition were observed in *S. lividans* heterologous host and mutant strains, suggesting compounds with antimicrobial

activity are produced when pESAC13A-2H is introduced into *S. lividans* TK24. From LC-MS analysis, the production levels of specific compounds are not under the control of the SAV_2268 repressor. Comparative metabolic profiling did not reveal any compound absent from *S. lividans* TK24 and produced in similar amounts between *S. lividans* containing the pESAC13A-2H or mutated version of pESAC13A-2H.

Interestingly, crude extract from *S. lividans* heterologous host and mutant strains were active against wild type *S. lividans* TK24 implying pESAC13A-2H also contains the resistance genes against these antimicrobial compounds. It should be noted that the wild type *S. lividans* TK24 crude extract were also active against itself and the resistance genes may not be expressed. The antimicrobial assays needs to be repeated to confirm the antimicrobial activity of the crude organic extracts.

### 6.8. Potential cross-talk in *S. coelicolor* and *S. lividans* heterologous hosts

The production of the azoxy compounds (**1** - **3**) were observed in *S. coelicolor* heterologous hosts grown on SMMS agar; however in the case of *S. lividans*, the azoxy compounds (**1** - **3**) were still not produced in rich media and deletion of the *sav_2268* repressor was required to unlock the production of **1** - **3**. These observations suggest there are two components to the regulation of the predicted azoxy biosynthetic gene cluster; firstly, the *sav_2268* repressor needs to be deleted to allow expression of biosynthetic genes and secondly, there needs to be appropriate precursors.

The azoxy compounds (**1** - **3**) were detected in *S. coelicolor* heterologous host strains grown on SMMS agar and appears to be greatly affected by the availability of

nutrients. Indeed the production of **1** - **3** were abolished when *S. coelicolor* heterologous host strains were grown in VPM; nutrient-sensing pleiotropic transcription factors may play a role in regulating the azoxy compounds. The predicted azoxy compounds (**1** - **3**) were not detected when *S. lividans* heterologous host strains were grown on SMMS agar and implies either nutrient supply is insufficient or other nutrient-sensing transcriptional repressors may still be regulating the predicted azoxy biosynthetic gene cluster.

*S. coelicolor* M1152 and *S. lividans* TK24 are highly similar species; however, many endogenous biosynthetic gene clusters were deleted in *S. coelicolor*, including the ScbR2 "pseudo" GBL-binding receptor (the role of ScbR2 repressor in *S. coelicolor* is discussed in Section 1.5.2). [166] SLIV_06715 repressor in *S. lividans* TK24, which shares 99% identity and 99% similarity with ScbR2 repressor, may regulate expression of the predicted azoxy biosynthetic gene cluster; this is supported by detection of **1** - **3** in *S. coelicolor,* but not in *S. lividans*, when grown on SMMS. Using MEME enrichment analysis, only one ScbR binding site (GGACGGCTGGCGCTGGCTGTTC) was discovered within *sav_2297* (582 bp downstream of *sav_2297* start codon), which encodes for a major facilitator superfamily transporter; the binding of the SLIV_06715 repressor to the specific ARE sequence may regulate expression of the operon (Figure 6-1).

Using MEME enrichment analysis, SAV_2268/SAV_2270 repressor binding sites were also revealed across the *S. lividans* TK24 genome. From FIMO analysis, 14 unique individual hits were discovered and 10 hits were found in the promoters of ORFs (within 300 bp of start codon) (Table 6-7 and Electronic Supplementary

Material). Surprisingly, some of the SAV_2268/SAV_2270 binding sites were discovered in promoters of transcriptional regulators involved in nutrient-response, such as the PhoU family transcriptional regulator (*sliv_rs17285*). The PhoU family transcriptional regulator is essential in *E. coli* for the repression of the *pho* regulon in high phosphate concentrations and in *S. coelicolor*, PhoU regulator has been implicated to be involved in regulating the *pho* regulon. [92,195] MEME enrichment analysis was successful at revealing ARE sequences in *S. lividans* TK24 hosts and suggests potential cross-activity between the heterologous host and predicted azoxy biosynthetic gene cluster. The predicted ARE sites needs to be further confirmed to determine cross-talk between pathways of heterologous hosts and that contained in the integrated pESAC13A. Introduction of pESAC13A-2H and pESAC13A-19K constructs in other heterologous hosts, such as *S. albus* J1074, which does not contain any AHFCA-dependent or GBL-dependent signalling systems, might provide a better solution to prevent cross-talk between ArpA-like repressors.

In *S. coelicolor* heterologous hosts, production of the AHFCA signalling molecules (**17-1**, **17-2**, **18-1**, **18-2**) could not be detected. The AHFCA-dependent signalling system in *S. coelicolor* heterologous host strains could have been rearranged and this may explain why the predicted azoxy compounds (**1** - **3**) were detected in the *S. coelicolor* heterologous host strains when grown on SMMS agar. Although, the assembly of AHFCA signalling molecules could still be tightly regulated by SAV_2268, the deletion of *sav_2268* gene in *S. coelicolor* heterologous host strains is expected to unlock production of AHFCA signalling molecules. Sequencing of *S. coelicolor* and *S. lividans* heterologous host and mutant strains would confirm if

any rearrangements of the azoxy biosynthetic gene cluster have occured and may provide insights into the compounds observed by LC-MS.

**Table 6-7** – Predicted SAV_2268/SAV_2270 binding sites in *S. lividans* TK24.

| Matched Sequence | bp upstream of gene | Gene downstream of sequence | Proposed function of gene |
|---|---|---|---|
| CGAAGGACTTCATCGAGAAGGACTACTACA | -37 | SLIV_RS19960 | DnaJ Molecular Chaperone |
| TGATAGATGACTCTGGAAAGGAATCATTCA | 24 | SLIV_RS36610 | Oxidoreductase |
| TGAATGATTCCTTTCCAGAGTCATCTATCA | 16 | SLIV_RS36605 | TetR family transcriptional regulator |
| CGACGCGTACGTCCGTGACCGTATCCTGCT | -95 | SLIV_RS20910 | N-acetylglutamate synthase |
| CGAAACGGACGTACGAGAGGGAATCCTGAT | -2 | SLIV_RS17285 | PhoU family transcriptional Regulator |
| ATCAGGATTCCCTCTCGTACGTCCGTTTCG | 186 | SLIV_RS17280 | Two-component sensor histidine kinase |
| TGCAGAAGGGCTTCGAGGAGGTCGACTTCT | -70 | SLIV_RS21110 | Hydrolase |
| AGAAGTCGACCTCCTCGAAGCCCTTCTGCA | 114 | SLIV_RS21115 | AraC family transcriptional regulator |
| TGTACAAGACCTTCTTCAAGCTCGTCTTCT | -31 | SLIV_RS30350 | Dihydroorotate dehydrogenase |
| ACTTGAACACCTTCGCGAAGGTGCCGTTGA | -73 | SLIV_RS19075 | Hypothetical protein |

## 6.9. Conclusions

Deletion of the *sav_2268* repressor, as well as overexpression of the *sav_2301* activator was expected to unlock the production of predicted azoxy compounds (**1** - **3**) and AHFCA signalling molecules. The *sav_2268* repressor was inactivated in *S. lividans* heterologous host strains by double homologous recombination and LC-MS analysis of mutant strains grown in VPM revealed the production of azoxy compounds (**1** - **3**) and AHFCA signalling molecules (**17-1**, **17-2**, **18-1** and **18-2**); identification of these specific compounds were confirmed by UHPLC-ESI-TOF-MS.

In addition to the observed compounds, a series of PUFAA compounds (**4** - **16** and **19** – **45**) were also detected in *S. lividans* heterologous host strains and were significantly reduced or abolished in mutant strains. From genome mining analyses, the azoxy compounds were proposed to derive from a precursor of polyketide origin and the observed PUFAA compounds were speculated to be involved in the biosynthesis of the azoxy compounds; however, such compounds could not be detected by LC-MS due to possible instability of these specific azoxy compounds and their isolation could be further optimised by converting them into stable ammonia adduct derivatives. The azoxy biosynthetic gene cluster appears to be regulated differently between *S. lividans* TK24 and *S. coelicolor* M1152; further investigations are required to determine whether there are possible cross-talks between different specialised metabolite gene clusters or rearrangements of the azoxy biosynthetic gene cluster.

# 7. Conclusions, Summary and Future Work

## 7.1. MEME enrichment analysis

Due to the rise in multi-drug resistance microorganism, there is an urgent need for novel antimicrobial agents. Two-thirds of antibiotics approved for use in the clinic today originate from the *Streptomyces* genus of bacteria and each *Streptomyces* bacteria is predicted to contain dozens of gene clusters involved in the direct biosynthesis of specialised metabolites. [21] Production of these specific bioactive molecules are sometimes not produced or produced in very small quantities in laboratory conditions and most often, these cryptic biosynthetic gene clusters are tightly regulated at the transcriptional level and are under the control of DNA-binding transcription factors. [116]

AHFCA-dependent and GBL-dependent signalling systems are widespread in *Streptomyces* bacteria and represent examples of complex regulatory networks involved in controlling cryptic biosynthetic gene clusters. Genome mining analysis has identified other putative AHFCA-dependent signalling systems and TetR repressor binding sites could be predicted using MEME enrichment analysis. This information provides unprecedented insights into the complex regulatory networks and genetic manipulation targets to unlock the production of novel compounds from silent and cryptic biosynthetic gene clusters. To our knowledge, the specific natural product assembled by cryptic biosynthetic gene clusters are not known and exploitation of these signalling systems, through the deletion of specific transcriptional repressors and overexpression of transcriptional activators, are

expected to lead to the discovery of novel natural compounds. In our previous studies, exploitation of AHFCA-dependent signalling system in *S. coelicolor* and *S. venezuelae* (published in reference 62) unlocked the production of methylenomycins and gaburedins respectively. [57,62] In this project, the exploitation of the AHFCA-dependent signalling system in *S. avermitilis* resulted in the discovery of azoxy compounds and AHFCAs.

## 7.2. Azoxy natural compounds

The first azoxy compound that was discovered was the antifungal compound, macrozamin, from the plant cycad, *Macrozamia spiralis* (Figure 7-1). [191,196] Since the discovery of macrozamin in 1940, which was also the first naturally occurring compound isolated to contain an N-N bond, a series of other azoxy compounds have been isolated, including cycasin from *Cycas revoluta* Thunb, azoxybacillins from *bacillus cereus* and more recently, azoxyalkenes from *S. rochei* 7434AN4 (Figure 7-1). [189,192,197]

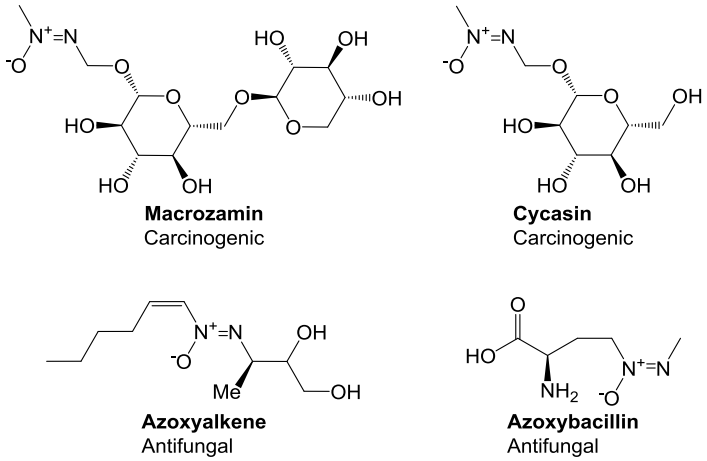



**Figure 7-1** – Azoxy compounds isolated from other organisms.

Azoxy compounds exhibit a broad range of biological activities, including antimicrobial, antifungal and anticancer properties (Figure 7-1 and Figure 7-2). Most interestingly, modifications of the aliphatic chain or seryl residue of elaiomycin compounds significantly altered the biological properties; for instance, amidation in elaiomycin H displayed activity against *Aspergillus* (Figure 7-2). [175,193,198]

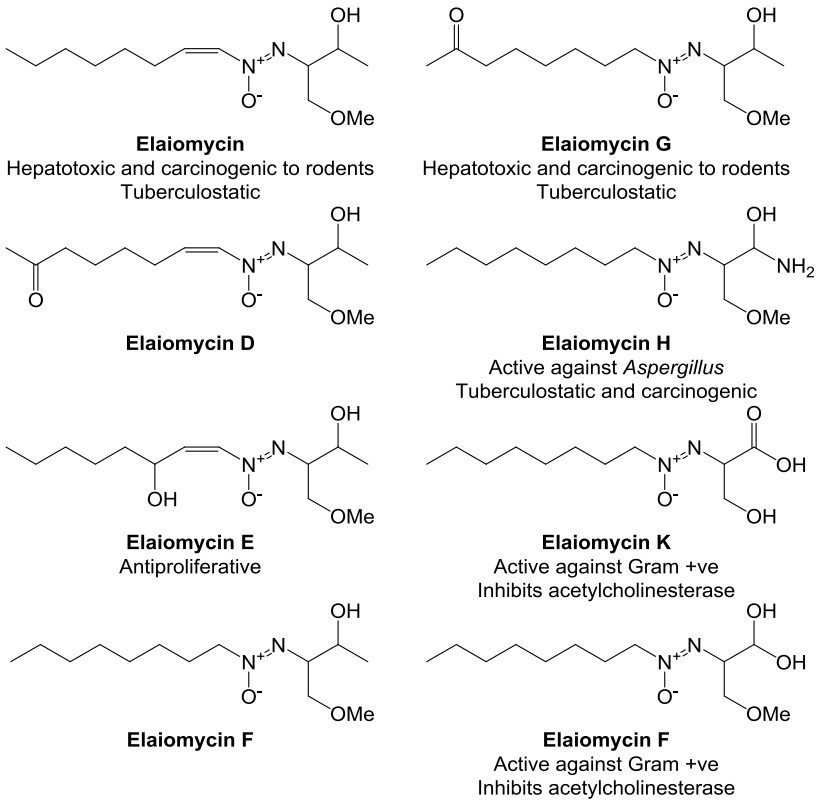



**Figure 7-2** – Elaiomycin compounds isolated from *Streptomyces* bacteria.

To date, the only sequenced azoxy biosynthetic gene cluster is the valanimycin gene cluster from *S. viridifaciens.* Sequencing of the gene cluster has proved to be greatly beneficial in the investigation of valanimycin biosynthesis, which has led to the establishment of the unusual involvement of seryl-tRNA synthetase in valanimycin biosynthesis. Despite extensive investigations, formation of the key azoxy moiety still remains unclear. [152,162] Biosynthesis of other specialised metabolites that utilise

aminoacyl-tRNA synthetase are widespread and have been found to participate in the biosynthesis of pacidamycins and cyclodipeptides. [199,200]

Most significantly, sequencing of the valanimycin biosynthetic gene cluster has assisted with genome mining analysis of the *S. avermitilis* cryptic biosynthetic gene cluster, which is proposed to be under the control of an AHFCA-dependent signalling system. Since nearly all of the key valanimycin biosynthetic genes were also identified in the *S. avermitilis* cryptic biosynthetic gene cluster, the cryptic biosynthetic gene cluster is predicted to direct biosynthesis of novel azoxy natural compounds.

### 7.3. Identification of the predicted azoxy compounds

Due to the relatively poor tractability of *S. avermitilis*, a heterologous expression/ comparative metabolic profiling approach was used to identify the compounds of interest. The predicted azoxy biosynthetic gene cluster was captured using pESAC13A PAC technologies (pESAC13A-2H and pESAC13A-19K) by Bio S&T. In the first instance, pESAC13A-2H and pESAC13A-19K, which contains the predicted azoxy biosynthetic gene cluster, were successfully introduced into *S. coelicolor* M1152 and *S. lividans* TK24 heterologous hosts *via* tri-parental conjugation. Comparative metabolic profiling of *S. coelicolor* heterologous host strains grown on SMMS agar successfully detected the predicted azoxy compounds (**1** – **3**). Identification of the azoxy compounds were further supported by UHPLC-ESI-TOF-MS data and incorporation of isotope-labelled L-serine-$^{13}C_3$-$^{15}N$.

As of May 2014, the products of 16 cryptic biosynthetic gene clusters in *S. avermitilis* are known and the pESAC13A PAC genomic library can be further exploited to

screen for the remaining 22 cryptic biosynthetic gene clusters. [35] Similar heterologous expression/ comparative metabolic profiling can be applied to identify these specific natural compounds.

## 7.4. Genetic manipulations of the *S. avermitilis* AHFCA-dependent signalling system

The predicted azoxy compounds (**1** - **3**) were not detected in *S. lividans* heterologous host strains by LC-MS, which implies that the biosynthetic gene cluster is still tightly regulated. MEME enrichment analysis successfully revealed TetR binding sites in the promoter of *sav_2301* transcriptional activator. The neighbouring AHFCA-dependent signalling system is speculated to tightly regulate expression of *sav_2301*, which is predicted to control expression of the azoxy biosynthetic genes. Therefore, deletion of TetR repressor, *sav_2268*, is expected to unlock the production of azoxy compounds (**1** - **3**) and AHFCAs.

The *sav_2268* repressor was successfully deleted in *S. lividans* heterologous host strains by double homologous recombination. Comparative metabolic profiling of *S. lividans* mutant strains by LC-MS revealed the predicted production of azoxy compounds (**1** - **3**) and AHFCA 3, 5, 6 and 7 signalling molecules (**17-1**, **17-2**, **18-1** and **18-2**); identification of AHFCA signalling molecules were established using AHFCA synthetic standards and UHPLC-ESI-TOF-MS analysis.

The presence of AHFCA signalling molecules needs to be further confirmed by tandem MS and co-injection of *S. avermitilis* AHFCAs and authentic standards on LC-MS. In addition, the activity of these AHFCAs will need to be tested to determine

whether the specific signalling molecules can abolish binding of the TetR repressors to ARE sequences.

### 7.5. Identification of other metabolites

Detailed analysis of LC-MS results revealed a series of compounds that were, not only produced in heterologous host strains containing the pESAC13A constructs, but were also significantly reduced or abolished in repressor mutant strains. Molecular formulae, generated by UHPLC-ESI-TOF-MS data, implies these compounds are aromatic amines (**4** - **6**) and polyunsaturated fatty acid amines (**7** - **16** and **19** - **45**). The exact biological roles and functions of these compounds remains unclear; however, the predicted azoxy compounds are proposed to incorporate a precursor of polyketide origin. The *S. avermitilis* azoxy biosynthetic gene cluster contains a few genes that encode for PKS enzymes and the biosynthetic machinaries are speculated to be involved in the assembly of PUFAA compounds (**7** - **16** and **19** - **45**). The PUFAA compounds (**7** - **16** and **19** - **45**) are hypothesised to be incorporated in the final azoxy compounds; however, LC-MS was unable to detect these specific azoxy compounds, which could be due to possible instability of the azoxy compounds.

Isolation of the predicted azoxy compounds needs to be further optimised to produce stable derivatives that could be characterised by MS and NMR spectroscopy analyses. Since the azoxy compounds should incorporate L-serine, a genomisotopic approach can be used to identify these specific azoxy compounds and also aid in isolation for structure determination by NMR spectroscopy. [55]

Furthermore, DNA sequencing of heterologous host strains and pESAC13A constructs to determine insert ends and possible rearrangements of the azoxy

biosynthetic gene cluster would provide insights and explanations to observations from comparative metabolic profiling.

### 7.6. Structure determination of the predicted azoxy compounds

To determine the structure of compounds **1** - **3**, production of compounds were scaled up, purified by semi-preparative HPLC and analysed by a combination of 1- and 2-D NMR spectroscopy. Analyses of the NMR spectra revealed the main compound was the signalling molecule, SCB1, and not the azoxy metabolite. Furthermore, detailed analysis of mass spectra data, undoubtedly, showed compounds **1** - **3** were co-purified with SCB signalling molecules.

Isolation of **1** - **3** will therefore require further optimisation by determining optimal growth conditions for the maximum yield of product and through genetic manipulations of the AHFCA-dependent signalling system to overproduce the azoxy compounds. The SCB1 signalling molecules, which are overproduced in *S. coelicolor* M1152, needs to be removed for structure determination of compounds **1** - **3** and deletion of the *scbA* gene, which is involved in assembly of SCB signalling molecules, would abolish the production of SCB signalling molecules in *S. coelicolor* M1152. Alternatively, introduction of the pESAC13A constructs in other "cleaner" heterologous hosts could help with the purification of the specific azoxy compounds. An increased availability of purified compound material would facilitate structure determination by NMR spectroscopy and permit the use of other analytical chemistry techniques, such as IR spectroscopy.

## 7.7. Cross-talks between *Streptomyces* heterologous host and the predicted azoxy biosynthetic gene cluster

The conditions required to detect the predicted azoxy compounds (**1** - **3**) implies that the predicted azoxy biosynthetic gene cluster is regulated differently between *S. coelicolor* and *S. lividans* heterologous hosts. MEME enrichment analysis was used to determine cross-talk between different specialised metabolite biosynthetic gene clusters and revealed one ScbR/ScbR2 repressor binding site located within the predicted azoxy biosynthetic gene cluster. The analogous ScbR2 repressor in *S. lividans* may affect expression of the azoxy biosynthetic gene cluster. In addition, several SAV_2268/SAV_2270 repressor binding sites were discovered across *S. lividans* TK24 including nutrient-sensing transcription factors, such as the PhoU transcriptional regulator. [92,195]

Nutrient-sensing transcriptional repressors and analogous ArpA-like repressors are speculated to tightly regulate the azoxy biosynthetic gene cluster and possible cross-activity needs to be confirmed through binding assays. Alternatively, introduction of the pESAC13A constructs into different heterologous hosts, such as *S. albus*, which does not produce GBLs, is expected to abolish cross-activity with other ArpA-like transcription factors.

## 7.8. Genome mining analysis of other putative azoxy biosynthetic gene clusters

Azoxy compounds are a rare class of N-N bond-containing compounds; in fact, to date, there are only 18 compounds, containing the azoxy moiety, within the Dictionary of Natural Products. [201] Genome mining analyses have identified putative biosynthetic gene clusters that are proposed to direct the biosynthesis of distinct

novel azoxy natural compounds, all sharing key valanimycin biosynthetic genes (Figure 7-3). Intriguingly, no AHFCA-dependent signalling systems are present in these gene clusters and all putative azoxy biosynthetic gene clusters are suggested to be unique. Experimental investigations of these systems are expected to result in the discovery of further azoxy compounds (Figure 7-3).

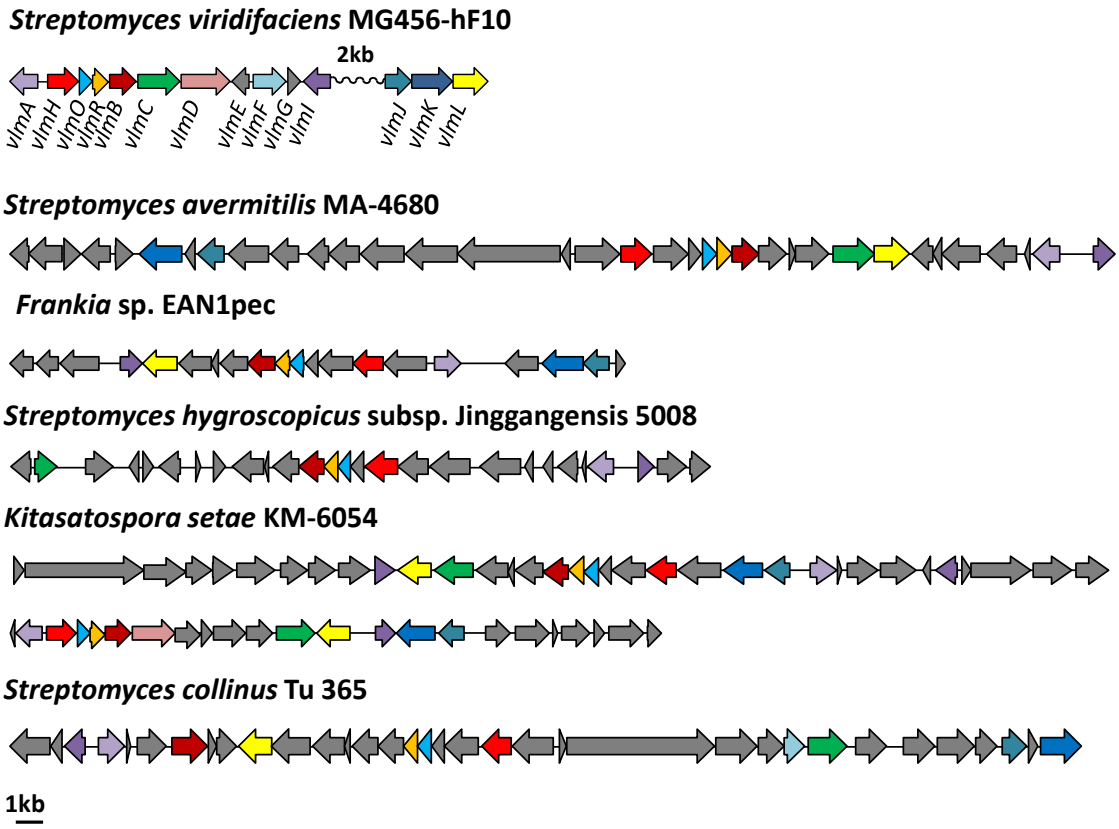



**Figure 7-3** – Other specialised metabolite biosynthetic gene clusters proposed to direct the biosynthesis of novel azoxy compounds. Analogous genes are colour-coded.

## 8. References


1. Davies, J. & Davies, D., Origins and evolution of antibiotic resistance. *Microbiol Mol Biol Rev.* **74** (3), 417-33. (2010).
2. Fleming, A., The discovery of penicillin. *Br Med J.* **2** (1), 4-5. (1944).
3. Fleming, A., 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** (3), 226-236. (1929).
4. Procópio, R. E., Silva, I. R., Martins, M. K., Azevedo, J. L. & Araújo, J. M., Antibiotics produced by *Streptomyces*. *Braz J Infect Dis.* **16** (5), 466-71. (2012).
5. Sakoulas, G. & Moellering, R. C. J., Increasing antibiotic resistance among methicillin-resistant *Staphylococcus aureus* strains. *Clin Infect Dis.* **46** (Suppl 5), S360-7. (2008).
6. Silver, L. L. & Bostian, K. A., Discovery and development of new antibiotics: the problem of antibiotic resistance. *Antimicrob Agents Chemother.* **37** (3), 377-83. (1993).
7. Shlaes, D. M., Projan, S. J. & Edwards, J. E., Antibiotic discovery: state of the state. *ASM News.* **70** (6), 275-81. (2004).
8. Watve, M. G., Tickoo, R., Jog, M. M. & Bhole, B. D., How many antibiotics are produced by the genus *Streptomyces*? *Arch Microbiol.* **176** (5), 386-90. (2001).
9. Hodgson, D. A., Primary metabolism and its control in streptomycetes: a most unusual group of bacteria. *Adv Microb Physiol.* **42**, 47-238. (2000).
10. Zhou, Z., Gu, J., Du, Y. L., Li, Y. Q. & Wang, Y., The -omics era- toward a systems-level understanding of *Streptomyces*. *Curr Genomics.* **12** (6), 404-16. (2011).
11. Anné, J., Maldonado, B., Van Impe, J., Van Mellaert, L. & Bernaerts, K., Recombinant protein production and streptomycetes. *J Biotechnol.* **158** (4), 159-67. (2012).
12. Hibbing, M. E., Fuqua, C., Parsek, M. R. & Peterson, S. B., Bacterial competition: surviving and thriving in the microbial jungle. *Nat Rev Microbiol.* **8** (1), 15-25. (2010).
13. Flärdh, K. & Buttner, M. J., *Streptomyces* morphogenetics: dissecting differentiation in a filamentous bacterium. *Nat Rev Microbiol.* **7** (1), 36-49. (2009).
14. Waksman, S. A., Streptothricin, a new selective bacteriostatic and bactericidal agent, particularly active against Gram-negative bacteria. *Exp Biol Med (Maywood).* **49** (2), 207-10. (1942).
15. Waksman, S. A., Production and activity of streptothricin. *J Bacteriol.* **46** (3), 299-310. (1943).

16. Robinson, H. J., Studies on the toxicity and activity of streptothricin. *Science.* **99** (2583), 540-2. (1944).
17. Waksman, S. A., Reilly, H. C. & Johnstone, D. B., Isolation of streptomycin-producing strains of *Streptomyces griseus*. *J Bacteriol.* **52** (3), 393–7. (1946).
18. Singh, B. & Mitchison, D. A., Bactericidal activity of streptomycin and isoniazid against tubercle bacilli. *Br Med J.* **1** (4854), 130-2. (1954).
19. Kincade, G. F., Saxton, G. D., Morse, P. W. & Mathisen, A. K., Streptomycin in the treatment of tuberculosis (A Report of Its Use in a Series of 100 Cases). *Can Med Assoc J.* **59** (2), 105-12. (1948).
20. Finken, M., Kirschner, P., Meier, A., Wrede, A. & Böttger, E. C., Molecular basis of streptomycin resistance in *Mycobacterium tuberculosis*: alterations of the ribosomal protein S12 gene and point mutations within a functional 16S ribosomal RNA pseudoknot. *Mol Microbiol.* **9** (6), 1239-46. (1993).
21. Kieser, T., Bibb, M. J., Buttner, M. J., Chater, K. F. & Hopwood, D. A., *Practical Streptomyces Genetics.* (The John Innes Foundation, Norwich, 2000).
22. Chaudhary, A. K., Dhakal, D. & Sohng, J. K., An insight into the "-omics" based engineering of streptomycetes for secondary metabolite overproduction. *Biomed Res Int.*, doi:10.1155/2013/968518. (2013).
23. LPSN, Genus *Streptomyces*, Available at http://www.bacterio.net/s/streptomycesa.html. (Accessed January 2015).
24. Hopwood, D. A., *Streptomyces in Nature and Medicine: The Antibiotic Makers* (Oxford University Press, New York, 2007).
25. Harrison, J. & Studholme, D. J., Recently published *Streptomyces* genome sequences. *Microb Biotechnol.* **7** (5), 373-80. (2014).
26. Bentley, S. D. *et al.*, Complete genome sequence of the model actinomycete *Streptomyces coelicolor* A3(2). *Nature.* **417** (6885), 141-7. (2002).
27. Ohnishi, Y., Kameyama, S., Onaka, H. & Horinouchi, S., The A-factor regulatory cascade leading to streptomycin biosynthesis in *Streptomyces griseus* : identification of a target gene of the A-factor receptor. *Mol Microbiol.* **34** (1), 102-11. (1999).
28. Pullan, S. T., Chandra, G., Bibb, M. J. & Merrick, M., Genome-wide analysis of the role of GlnR in *Streptomyces venezuelae* provides new insights into global nitrogen regulation in actinomycetes. *BMC Genomics.* **12** (175), doi:10.1186/1471-2164-12-175. (2011).
29. Ikeda, H. *et al.*, Complete genome sequence and comparative analysis of the industrial microorganism *Streptomyces avermitilis*. *Nat Biotechnol.* **21** (5), 526-31. (2003).
30. Wu, H. *et al.*, Genomic and transcriptomic insights into the thermo-regulated biosynthesis of validamycin in *Streptomyces hygroscopicus* 5008. *Nat Biotechnol.* **21** (5), 526-31. (2003).
31. Burg, R. W. *et al.*, Avermectins, new family of potent anthelmintic agents: producing organism and fermentation. *Antimicrob Agents Chemother.* **15** (3), 361–7. (1979).
32. Crump, A. & Omura, S., Ivermectin, 'Wonder drug' from Japan: the human use perspective. *Proc Jpn Acad Ser B Phys Biol Sci.* **87** (2), 13-28. (2011).

33. Nobel Media AB, The Nobel Prize in Physiology or Medicine 2015, Available at http://www.nobelprize.org/nobel_prizes/medicine/laureates/2015/. (Accessed December 2015).
34. Chou, W. K. W. *et al.*, Genome mining in *Streptomyces avermitilis*: cloning and characterization of SAV_76, the synthase for a new sesquiterpene, avermitilol. *J Am Chem Soc.* **132** (26), 8850-1. (2010).
35. Ikeda, H., Kazuo, S. Y. & Omura, S., Genome mining of the *Streptomyces avermitilis* genome and development of genome-minimized hosts for heterologous expression of biosynthetic gene clusters. *J Ind Microbiol Biotechnol.* **41** (2), 233-50. (2014).
36. Metz, P. A. *et al.*, Characterization of a lipid-rich fraction synthesized by *Streptomyces avermitilis*. *J Chromatogr.* **441** (1), 31-44. (1988).
37. Cane, D. E., He, X., Kobayashi, S., Omura, S. & Ikeda, H., Geosmin biosynthesis in *Streptomyces avermitilis*. Molecular cloning, expression, and mechanistic study of the germacradienol/geosmin synthase. *J Antibiot (Tokyo).* **59** (8), 471-9. (2006).
38. Tetzlaff, C. N. *et al.*, A gene cluster for biosynthesis of the sesquiterpenoid antibiotic pentalenolactone in *Streptomyces avermitilis*. *Biochemistry.* **45** (19), 6179-86. (2006).
39. Takamatsu, S. *et al.*, Characterization of a silent sesquiterpenoid biosynthetic pathway in *Streptomyces avermitilis* controlling epi-isozizaene albaflavenone biosynthesis and isolation of a new oxidized epi-isozizaene metabolite. *Microb Biotechnol.* **4** (2), 184-91. (2011).
40. Kitani, S. *et al.*, Avenolide, a *Streptomyces* hormone controlling antibiotic production in *Streptomyces avermitilis*. *Proc Natl Acad Sci USA.* **108** (39), 16410-5. (2011).
41. Denoya, C. D., Skinner, D. D. & Morgenstern, M. R., A *Streptomyces avermitilis* gene encoding a 4-hydroxyphenylpyruvic acid dioxygenase-like protein that directs the production of homogentisic acid and an ochronotic pigment in Escherichia coli. *J Bacteriol.* **176** (17), 5312-9 (1994).
42. Ueki, M. *et al.*, Nocardamin production by *Streptomyces avermitilis*. *Actinomycetologica.* **23** (2), 34-9. (2009).
43. Weber, T. *et al.*, antiSMASH 3.0 — a comprehensive resource for the genome mining of biosynthetic gene clusters. *Nucl Acids Res.* **43** (W1), W237-43. (2015).
44. Challis, G. L., Mining microbial genomes for new natural products and biosynthetic pathways. *Microbiology.* **154** (Pt 6), 1555-69. (2008).
45. Corre, C. & Challis, G. L., Heavy tools for genome mining. *Chem Biol.* **14** (1), 7-9. (2007).
46. Reller, T., New 'Reaxys' solution improves workflow quality and efficiency for chemists (2009), Available at https://www.elsevier.com/about/press-releases/science-and-technology/new-reaxys-solution-improves-workflow-quality-and-efficiency-for-chemists. (Accessed November 2015).
47. Rutledge, P. J. & Challis, G. L., Discovery of microbial natural products by activation of silent biosynthetic gene clusters. *Nat Rev Microbiol.* **13** (8), 509-23. (2015).

48. Fischbach, M. A. & Walsh, C. T., Assembly-line enzymology for polyketide and nonribosomal peptide antibiotics: logic, machinery, and mechanisms. *Chem Rev.* **106** (8), 3469-96. (2006).
49. Haydock, S. F. *et al.*, Divergent sequence motifs correlated with the substrate specificity of (methyl)malonyl-CoA:acyl carrier protein transacylase domains in modular polyketide synthases. *FEBS Lett.* **374** (2), 246-8. (1995).
50. Stachelhaus, T., Mootz, H. D. & Marahiel, M. A., The specificity-conferring code of adenylation domains in nonribosomal peptide synthetases. *Chem Biol.* **6** (8), 493-505. (1999).
51. Challis, G. L., Ravel, J. & Townsend, C. A., Predictive, structure-based model of amino acid recognition by nonribosomal peptide synthetase adenylation domains. *Chem Biol.* **7** (3), 211-24. (2000).
52. Caffrey, P., Conserved amino acid residues correlating with ketoreductase stereospecificity in modular polyketide synthases. *Chembiochem.* **4** (7), 654-7. (2003).
53. Reid, R. *et al.*, A model of structure and catalysis for ketoreductase domains in modular polyketide synthases. *Biochemistry.* **42** (1), 72-9. (2003).
54. Lautru, S., Deeth, R. J., Bailey, L. M. & Challis, G. L., Discovery of a new peptide natural product by *Streptomyces coelicolor* genome mining. *Nat Chem Biol.* **1** (5), 265-9. (2005).
55. Gross, H. *et al.*, The genomisotopic approach: a systematic method to isolate products of orphan biosynthetic gene clusters. *Chem Biol.* **14** (1), 53-63. (2007).
56. Corre, C., Song, L., O'Rourke, S., Chater, K. F. & Challis, G. L., 2-Alkyl-4-hydroxymethylfuran-3-carboxylic acids, antibiotic production inducers discovered by *Streptomyces coelicolor* genome mining. *Proc Natl Acad Sci USA.* **105** (45), 17510-5. (2008).
57. O'Rourke, S. *et al.*, Extracellular signalling, translational control, two repressors and an activator all contribute to the regulation of methylenomycin production in *Streptomyces coelicolor*. *Mol Microbiol.* **71** (3), 763-78. (2009).
58. Chiang, Y. M., Chang, S. L., Oakley, B. R. & Wang, C. C., Recent advances in awakening silent biosynthetic gene clusters and linking orphan clusters to natural products in microorganisms. *Curr Opin Chem Biol.* **15** (1), 137-43. (2011).
59. Van Wezel, G. P. & McDowall, K. J., The regulation of the secondary metabolism of *Streptomyces*: new links and experimental advances. *Nat Prod Rep.* **28** (7), 1311-33. (2011).
60. Aigle, B. & Corre, C., Waking up *Streptomyces* secondary metabolism by constitutive expression of activators or genetic disruption of repressors. *Methods Enzymol.* **517**, 343-66. (2012).
61. Laureti, L. *et al.*, Identification of a bioactive 51-membered macrolide complex by activation of a silent polyketide synthase in *Streptomyces ambofaciens*. *Proc Natl Acad Sci USA.* **108** (15), 6258-63. (2011).
62. Sidda, J. D. *et al.*, Discovery of a family of γ-aminobutyrate ureas *via* rational derepression of a silent bacterial gene cluster. *Chem Sci.* **5**, 86-9. (2014).

63. Ling, L. L. *et al.*, A new antibiotic kills pathogens without detectable resistance. *Nature.* **517** (7535), 455-9. (2015).
64. Olano, C. *et al.*, Activation and identification of five clusters for secondary metabolites in *Streptomyces albus* J1074. *Microb Biotechnol.* **7** (3), 242-56. (2014).
65. Mckenzie, N. L. *et al.*, Induction of antimicrobial activities in heterologous streptomycetes using alleles of the *Streptomyces coelicolor* gene *absA1*. *J Antibiot (Tokyo).* **63** (4), 177-82. (2010).
66. Cuthbertson, L. & Nodwell, J. R., The TetR family of regulators. *Microbiol Mol Biol Rev.* **77** (3), 440-75. (2013).
67. Ramos, J. L. *et al.*, The TetR family of transcriptional repressors. *Microbiol Mol Biol Rev.* **69** (2), 326-56. (2005).
68. Yu, Z., Reichheld, S. E., Savchenko, A., Parkinson, J. & Davidson, A. R., A comprehensive analysis of structural and sequence conservation in the TetR family transcriptional regulators. *J Mol Biol.* **400** (4), 847-64. (2010).
69. Hinrichs, W. *et al.*, Structure of the Tet repressor-tetracycline complex and regulation of antibiotic resistance. *Science.* **264** (5157), 418-20. (1994).
70. Orth, P., Schnappinger, D., Hillen, W., Saenger, W. & Hinsrichs, W., Structural basis of gene regulation by the tetracycline inducible Tet repressor–operator system. *Natl Struct Biol.* **7** (3), 215-9. (2000).
71. Murray, D. S., Schumacher, M. A. & Brennan, R. G., Crystal structures of QacR-diamidine complexes reveal additional multidrug-binding modes and a novel mechanism of drug charge neutralization. *J Biol Chem.* **279** (14), 14365-71. (2004).
72. Schumacher, M. A. *et al.*, Structural basis for cooperative DNA binding by two dimers of the multidrug-binding protein QacR. *EMBO J.* **21** (5), 1210-8. (2002).
73. Itou, H., Watanabe, N., Yao, M., Shirakihara, Y. & Tanaka, I., Crystal structures of the multidrug binding repressor *Corynebacterium glutamicum* Cgmr in complex with inducers and with an operator. *J Mol Biol.* **403** (2), 174-84. (2010).
74. Miller, D. J., Zhang, Y. M., Subramanian, C., Rock, C. O. & White, S. W., Structural basis for the transcriptional regulation of membrane lipid homeostasis. *Nat Struct Mol Biol.* **17** (8), 971-5. (2010).
75. Le, T. B., Schumacher, M. A., Lawson, D. M., Brennan, R. G. & Buttner, M. J., The crystal structure of the TetR family transcriptional repressor SimR bound to DNA and the role of a flexible N-terminal extension in minor groove binding. *Nucleic Acids Res.* **39** (21), 9433-47. (2011).
76. Le, T. B. *et al.*, Structures of the TetR-like simocyclinone efflux pump repressor, SimR, and the mechanism of ligand-mediated derepression. *J Mol Biol.* **408** (1), 40-56. (2011).
77. Sawai, H., Yamanaka, M., Sugimoto, H., Shiro, Y. & Aono, S., Structural basis for the transcriptional regulation of heme homeostasis in *Lactococcus lactis*. *J Biol Chem.* **287** (36), 30755-68. (2012).

78. Yang, S. *et al.*, Structural basis for interaction between *Mycobacterium smegmatis* Ms6564, a TetR family master regulator, and its target DNA. *J Biol Chem.* **288** (33), 23687-95. (2013).
79. Tonthat, N. K. *et al.*, SlmA forms a higher-order structure on DNA that inhibits cytokinetic Z-ring formation over the nucleoid. *Proc Natl Acad Sci USA.* **110** (26), 10586-91. (2013).
80. Bhukya, H., Bhujbalroa, R., Bitra, A. & Anand, R., Structural and functional basis of transcriptional regulation by TetR family protein CprB from *S. coelicolor* A3(2). *Nucleic Acids Res.* **42** (15), 10122-33. (2014).
81. Willems, A. R. *et al.*, Crystal structures of the *Streptomyces coelicolor* TetR-like protein ActR alone and in complex with actinorhodin or the actinorhodin biosynthetic precursor (S)-DNPA. *J Mol Biol.* **376** (5), 1377-87. (2008).
82. Frénois, F., Engohang-Ndong, J., Locht, C., Baulard, A. R. & Villeret, V., Structure of EthR in a ligand bound conformation reveals therapeutic perspectives against tuberculosis. *Mol Cell.* **16** (2), 301-7. (2004).
83. Zheng, J. *et al.*, Structure and function of the macrolide biosensor protein, MphR(A), with and without erythromycin. *J Mol Biol.* **387** (5), 1250-60. (2009).
84. Sahota, G. & Stormo, G. D., Novel sequence-based method for identifying transcription factor. *Bioinformatics.* **26** (21), 2672-7. (2010).
85. Speer, B. S. & Salyers, A. A., Novel aerobic tetracycline resistance gene that chemically modifies tetracycline. *J Bacteriol.* **171** (1), 148-53. (1989).
86. McMurry, L., Petrucci, J. R. E. & Levy, S. B., Active efflux of tetracycline encoded by four genetically different tetracycline resistance determinants in *Escherichia coli*. *Proc Natl Acad Sci USA.* **77** (7), 3974-7. (1980).
87. Kaneko, M., Yamaguchi, A. & Sawai, T., Energetics of tetracycline efflux system encoded by Tn10 in *Escherichia coli*. *FEBS Lett.* **193** (2), 194-8. (1985).
88. Hillen, W. & Berens, C., Mechanisms underlying expression of Tn10 encoded tetracycline resistance. *Annu Rev Microbiol.* **48**, 345-69. (1994).
89. Meier, I., Wray, L. V. & Hillen, W., Differential regulation of the Tn10-encoded tetracycline resistance genes *tetA* and *tetR* by the tandem *tet* operators O1 and O2. *EMBO J.* **7** (2), 567-72. (1988).
90. Orth, P., Saenger, W. & Hinrichs, W., Tetracycline-chelated Mg2+ ion initiates helix unwinding in Tet repressor induction. *Biochemistry.* **38** (1), 191-8. (1999).
91. Orth, P. *et al.*, Conformational changes of the Tet repressor induced by tetracycline trapping. *J Mol Biol.* **279** (2), 439-47. (1998).
92. Liu, G., Chater, K. F., Chandra, G., Niu, G. & Tan, H., Molecular regulation of antibiotic biosynthesis in *Streptomyces*. *Microbiol Mol Biol Rev.* **77** (1), 112-43. (2013).
93. Ando, N., Matsumori, N., Sakuda, S., Beppu, T. & Horinouchi, S., Involvement of *afsA* in A-factor biosynthesis as a key enzyme. *J Antibiot (Tokyo).* **50** (10), 847-52. (1997).

94. Sidda, J. D. & Corre, C., Gamma-butyrolactone and furan signaling systems in *Streptomyces*. *Methods Enzymol.* **517**, 71-87. (2012).
95. Nishida, H., Ohnishi, Y., Beppu, T. & Horinouchi, S., Evolution of gamma-butyrolactone synthases and receptors in *Streptomyces*. *Environ Microbiol.* **9** (8), 1986-94. (2007).
96. Healy, F. G. *et al.*, Characterization of gamma-butyrolactone autoregulatory signaling gene homologs in the angucyclinone polyketide WS5995B producer *Streptomyces acidiscabies*. *J Bacteriol.* **191** (15), 4786-97. (2009).
97. Du, Y. L., Shen, X. L., Yu, P., Bai, L. Q. & Li, Y. Q., Gamma-butyrolactone regulatory system of *Streptomyces chattanoogensis* links nutrient utilization, metabolism, and development. *Appl Environ Microbiol.* **77** (23), 8415-26. (2011).
98. Takano, E., Chakraburtty, R., Nihira, T., Yamada, Y. & Bibb, M. J., A complex role for the gamma-butyrolactone SCB1 in regulating antibiotic production in *Streptomyces coelicolor* A3(2). *Mol Microbiol.* **41** (5), 1015-28. (2001).
99. Xu, G. *et al.*, "Pseudo" gamma-butyrolactone receptors respond to antibiotic signals to coordinate antibiotic biosynthesis. *J Biol Chem.* **285** (35), 27440-8. (2010).
100. Kitani, S. *et al.*, Identification of genes involved in the butyrolactone autoregulator cascade that modulates secondary metabolism in *Streptomyces lavendulae* FRI-5. *Gene.* **425** (1-2), 9-16. (2008).
101. Kitani, S., Doi, M., Shimizu, T., Maeda, A. & Nihira, T., Control of secondary metabolism by *farX*, which is involved in the γ-butyrolactone biosynthesis of *Streptomyces lavendulae* FRI-5. *Arch Microbiol.* **192** (3), 211-20. (2010).
102. Arakawa, K., Mochizuki, S., Yamada, K., Noma, T. & Kinashi, H., Gamma-butyrolactone autoregulator-receptor system involved in lankacidin and lankamycin production and morphological differentiation in *Streptomyces rochei*. *Microbiology.* **153** (Pt 6), 1817-27. (2007).
103. Yamamoto, S., He, Y., Arakawa, Y. & Kinashi, H., γ-butyrolactone-dependent expression of the *Streptomyces* antibiotic regulatory protein gene *srrY* plays a central role in the regulatory cascade leading to lankacidin and lankamycin production in *Streptomyces rochei*. *J Bacteriol.* **190** (4), 1308-16. (2008).
104. Arakawa, K., Tsuda, N., Taniguchi, A. & Kinashi, H., The butenolide signaling molecules SRB1 and SRB2 induce lankacidin and lankamycin production in *Streptomyces rochei*. *Chembiochem.* **13** (10), 1447-57. (2012).
105. Salehi-Najafabadi, Z. *et al.*, The gamma-butyrolactone receptors BulR1 and BulR2 of *Streptomyces tsukubaensis*: tacrolimus (FK506) and butyrolactone synthetases production control. *Appl Microbiol Biotechnol.* **98** (11), 4919-36. (2014).
106. Zhang, Y., Zou, Z., Niu, G. & Tan, H., JadR and JadR2 act synergistically to repress jadomycin biosynthesis. *Sci China Life Sci.* **56** (7), 584-90. (2013).
107. Zou, Z. *et al.*, A γ-butyrolactone-sensing activator/repressor, JadR3, controls a regulatory mini-network for jadomycin biosynthesis. *Mol Microbiol.* **94** (3), 490-505. (2014).
108. Kawachi, R. *et al.*, Identification of an AfsA homologue (BarX) from *Streptomyces virginiae* as a pleiotropic regulator controlling autoregulator

biosynthesis, virginiamycin biosynthesis and virginiamycin M1 resistance. *Mol Microbiol.* **36** (2), 302-13. (2000).

109. Kinoshita, H. *et al.*, Butyrolactone autoregulator receptor protein (BarA) as a transcriptional regulator in *Streptomyces virginiae*. *J Bacteriol.* **179** (22), 6986-93. (1997).
110. Onaka, H. *et al.*, Cloning and characterization of the A-factor receptor gene from *Streptomyces griseus*. *J Bacteriol.* **177** (21), 6083-92. (1995).
111. Kato, J. Y., Miyahisa, J., Mashiko, M., Ohnishi, Y. & Horinouchi, S., A single target is sufficient to account for the biological effects of the A-factor receptor protein of *Streptomyces griseus*. *J Bacteriol.* **186** (7), 2206-11. (2004).
112. Onaka, H. & Horinouchi, S., DNA-binding activity of the A-factor receptor protein and its recognition DNA sequences. *Mol Microbiol.* **24** (5), 991-1000. (1997).
113. Khokhlov, A. S. *et al.*, The A-factor, responsible for streptomycin biosynthesis by mutant strains of *Actinomyces streptomycini*. *Dokl Akad Nauk SSSR.* **177** (1), 232-5. (1967).
114. Higo, A., Hara, H., Horinouchi, S. & Ohnishi, Y., Genome-wide distribution of AdpA, a global regulator for secondary metabolism and morphological differentiation in Streptomyces, revealed the extent and complexity of the AdpA regulatory network. *DNA Res.* **19** (3), 259-74. (2012).
115. Gomez-Escribano, J. P. *et al.*, Structure and biosynthesis of the unusual polyketide alkaloid coelimycin P1, a metabolic product of the *cpk* gene cluster of *Streptomyces coelicolor* M145. *Chem Sci.* **3** (9), 2716-20. (2012).
116. Bibb, M. J., Regulation of secondary metabolism in streptomycetes. *Curr Opin Microbiol.* **8** (2), 208-15. (2005).
117. Takano, E. *et al.*, A bacterial hormone (the SCB1) directly controls the expression of a pathway-specific regulatory gene in the cryptic type I polyketide biosynthetic gene cluster of *Streptomyces coelicolor*. *Mol Microbiol.* **56** (2), 465-79. (2005).
118. Dyson, P., *Streptomyces: Molecular Biology and Biotechnology.* (Caister Academic Press, 2011).
119. Takano, E. *et al.*, Purification and structural determination of SCB1, a γ-butyrolactone that elicits antibiotic production in *Streptomyces coelicolor* A3(2). *J Biol Chem.* **275** (15), 11010-6. (2000).
120. Wang, J. *et al.*, A novel role of 'pseudo' γ-butyrolactone receptors in controlling γ-butyrolactone biosynthesis in *Streptomyces*. *Mol Microbiol.* **82** (1), 236-50. (2011).
121. Gottelt, M., Kol, S., Gomez-Escribano, J. P., Bibb, M. & Takano, E., Deletion of a regulatory gene within the cpk gene cluster reveals novel antibacterial activity in *Streptomyces coelicolor* A3(2). *Microbiology.* **156** (8), 2343-53. (2010).
122. Świątek-Połatyńska, M. A. *et al.*, Genome-wide analysis of *in vivo* binding of the master regulator DasR in *Streptomyces coelicolor* identifies novel non-canonical targets. *PLoS One.* **10** (4), doi: 10.1371/journal.pone.0122479. (2015).

123. Allenby, N. E. E., Laing, E., Bucca, G., Kierzek, A. M. & Smith, C. P., Diverse control of metabolism and other cellular processes in *Streptomyces coelicolor* by the PhoP transcription factor: genome-wide identification of *in vivo* targets. *Nucl Acids Res.* **40** (19), 9543-56. (2012).
124. Wang, R. *et al.*, Identification of two-component system AfsQ1/Q2 regulon and its cross-regulation with GlnR in *Streptomyces coelicolor*. *Mol Microbiol.* **87** (1), 30-48. (2013).
125. Rodríguez, H., Rico, S., Díaz, M. & Santamaría, R. I., Two-component systems in *Streptomyces*: key regulators of antibiotic complex pathways. *Microb Cell Fact.* **12** (127), doi: 10.1186/1475-2859-12-127. (2013).
126. Chater, K. F. & Bruton, C. J., Resistance, regulatory and production genes for the antibiotic methylenomycin are clustered. *EMBO J.* **4** (7), 1893-7. (1985).
127. Hayes, A., Hobbs, G., Smith, C. P., Oliver, S. G. & Butler, P. R., Environmental signals triggering methylenomycin production by *Streptomyces coelicolor* A3(2). *J Bacteriol.* **179** (17), 5511-5. (1997).
128. Takano, E. *et al.*, A rare leucine codon in *adpA* is implicated in the morphological defect of *bldA* mutants of *Streptomyces coelicolor*. *Mol Microbiol.* **50** (2), 475-86. (2003).
129. Leskiw, B. K., Lawlor, E. J., Fernadez-Abalos, J. M. & Chater, K. F., TTA codons in some genes prevent their expression in a class of developmental, antibiotic-negative, *Streptomyces* mutants. *Proc Natl Acad Sci USA.* **88** (6), 2461-5. (1991).
130. Leskiw, B. K., Mah, R., Lawlor, E. J. & Chater, K. F., Accumulation of *bldA*-specified tRNA is temporally regulated in *Streptomyces coelicolor* A3(2). *J Bacteriol.* **175** (7), 1995-2005. (1993).
131. Gust, B., Kieser, T. & Chater, K., PCR targeting system in *Streptomyces coelicolor* A3(2), Available at http://streptomyces.org.uk/redirect/protocol_V1_4.pdf. (Accessed October 2015).
132. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J., Basic local alignment search tool. *J Mol Biol.* **215** (3), 403-10. (1990).
133. Blin, K. *et al.*, antiSMASH 2.0 — a versatile platform for genome mining of secondary metabolite producers. *Nucleic Acids Res.* **41**, W204-12. (2013).
134. Bailey, T. L. *et al.*, MEME SUITE: tools for motif discovery and searching. *Nucleic Acids Res.* **37**, W202-8. (2009).
135. Thermo Scientific, GeneJET Plasmid Miniprep Kit #K0502, #K0503, Available at https://tools.thermofisher.com/content/sfs/manuals/MAN0012655_GeneJET_Plasmid_Miniprep_UG.pdf. (Accessed October 2015).
136. MP Biomedicals, FastDNA SPIN Kit for Soil , Available at http://www.mpbio.com/includes/technical/soil.pdf. (Accessed October 2015).
137. Thermo Scientific, GeneJET Gel Extraction Kit, Available at https://tools.thermofisher.com/content/sfs/manuals/MAN0012661_GeneJET_Gel_Extraction_UG.pdf. (Accessed October 2015).

138. Muraoka, S. *et al.*, Crystal structure of a full-length LysR-type transcriptional regulator, CbnR: unusual combination of two subunit forms and molecular bases for causing and changing DNA bend. *J Mol Biol.* **328** (3), 555-66. (2003).
139. Wang, W. *et al.*, Angucyclines as signals modulate the behaviors of *Streptomyces coelicolor*. *Proc Natl Acad Sci USA.* **111** (15), 5688-93 (2014).
140. Li, X. *et al.*, ScbR- and ScbR2-mediated signal transduction networks coordinate complex physiological responses in *Streptomyces coelicolor*. *Sci Rep.* **5** (14831), doi: 10.1038/srep14831. (2015).
141. Tan, G., Bai, L. & Zhong, J., Exogenous 1,4-butyrolactone stimulates A-factor-like cascade and validamycin biosynthesis in *Streptomyces hygroscopicus* 5008. *Biotechnol Bioeng.* **110** (11), 2984-93. (2013).
142. Tan, G., Peng, Y., Lu, C., Bai, L. & Zhong, J., Engineering validamycin production by tandem deletion of γ-butyrolactone receptor genes in *Streptomyces hygroscopicus* 5008. *Metab Eng.* **28**, 74-81. (2015).
143. Softberry, BPROM, Available at http://linux1.softberry.com/berry.phtml?topic=bprom&group=programs&subgroup=gfindb. (Accessed November 2015).
144. Sievers, F. *et al.*, Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol Syst Biol.* **7** (539), doi: 10.1038/msb.2011.75. (2011).
145. Hsiao, N. H. & Kirby, R., Comparative genomics of *Streptomyces avermitilis*, *Streptomyces cattleya*, *Streptomyces maritimus* and *Kitasatospora aureofaciens* using a *Streptomyces coelicolor* microarray system. *Antonie Van Leeuwenhoek.* **93** (1-2), 1-25. (2008).
146. Rodríguez, E. & Gramajo, H., Genetic and biochemical characterization of the alpha and beta components of a propionyl-CoA carboxylase complex of *Streptomyces coelicolor* A3(2). *Microbiology.* **145** (Pt 11), 3109-19. (1999).
147. Onaka, H., Nakagawa, T. & Horinouchi, S., Involvement of two A-factor receptor homologues in *Streptomyces coelicolor* A3(2) in the regulation of secondary metabolism and morphogenesis. *Mol Microbiol.* **28** (4), 743-53. (1998).
148. Yang, K., Han, L. & Vining, L. C., Regulation of jadomycin B production in *Streptomyces venezuelae* ISP5230: involvement of a repressor gene, *jadR2*. *J Bacteriol.* **177** (21), 6111-7. (1995).
149. Shen, B., Polyketide biosynthesis beyond the type I, II and III polyketide synthase paradigms. *Curr Opin Chem Biol.* **7** (2), 285-95. (2003).
150. Shepherd, J. & Ibba, M., Direction of aminoacylated transfer RNAs into antibiotic synthesis and peptidoglycan-mediated antibiotic resistance. *FEBS Lett.* **587** (18), 2895-904. (2013).
151. Zeng, Y., Roy, H., Patil, P. B., Ibba, M. & Chen, S., Characterization of two seryl-tRNA synthetases in albomycin-producing *Streptomyces* sp. strain ATCC 700974. *Antimicrob Agents Chemother.* **53** (11), 4619-27. (2009).
152. Garg, R. P., Qian, X. L., Alemany, L. B., Moran, S. & Parry, R. J., Investigations of valanimycin biosynthesis: elucidation of the role of seryl-tRNA. *Proc Natl Acad Sci USA.* **105** (18), 6543-7. (2008).

153. Garg, R. P., Gonzalez, J. M. & Parry, R. J., Biochemical characterization of VlmL, a seryl-tRNA synthetase encoded by the valanimycin biosynthetic gene cluster. *J Biol Chem.* **281** (37), 26785-91. (2006).
154. Stefanska, A. L., Fulston, M., Houge-Frydrych, C. S., Jones, J. J. & Warr, S. R., A potent seryl tRNA synthetase inhibitor SB-217452 isolated from a *Streptomyces* species. *J Antibiot (Tokyo).* **53** (12), 1346-53. (2000).
155. Yamato, M. *et al.*, Isolation and properties of valanimycin, a new azoxy antibiotic. *J Antibiot (Tokyo).* **39** (2), 184-91. (1986).
156. Parry, R. J., Li, Y. & Lii, F. L., Biosynthesis of azoxy compounds. Investigations of valanimycin biosynthesis. *J Am Chem Soc.* **114** (25), 10062-4. (1992).
157. Parry, R. J. & Li, W., Purification and characterization of isobutylamine N-hydroxylase from the valanimycin producer *Streptomyces viridifaciens* MG456-hF10. *Arch Biochem Biophys.* **339** (1), 47-54. (1997).
158. Parry, R. J. & Li, W., An NADPH:FAD oxidoreductase from the valanimycin producer, *Streptomyces viridifaciens*. Cloning, analysis, and overexpression. *J Biol Chem.* **272** (37), 23303-11. (1997).
159. Parry, R. J., Li, W. & Cooper, H. N., Cloning, analysis and overexpression of the gene encoding isobutylamine N-hydroxylase from the valanimycin producer, *Streptomyces viridifaciens*. *J Bacteriol.* **179** (2), 409-16. (1997).
160. Garg, R. P., Alemany, L. B., Moran, S. & Parry, R. J., Identification, characterization, and bioconversion of a new intermediate in valanimycin biosynthesis. *J Am Chem Soc.* **131** (28), 9608-9. (2009).
161. Garg, R. P. & Parry, R. J., Regulation of valanimycin biosynthesis in *Streptomyces viridifaciens*: characterization of VlmI as a *Streptomyces* antibiotic regulatory protein (SARP). *Microbiology.* **156** (Pt 2), 472-83. (2010).
162. Garg, R. P., Ma, Y., Hoyt, J. C. & Parry, R. J., Molecular characterization and analysis of the biosynthetic gene cluster for the azoxy antibiotic valanimycin. *Mol Microbiol.* **46** (2), 505-17. (2002).
163. Zhao, J. L., Wen, Y., Chen, Z., Song, Y. & Li, J. L., An *adpA* homologue in *Streptomyces avermitilis* is involved in regulation of morphogenesis and melanogensis. *Chin Sci Bull.* **52** (5), 623-30. (2007).
164. Hong, S. T., Carney, J. R. & Gould, S. J., Cloning and heterologous expression of the entire gene clusters for PD 116740 from *Streptomyces* strain WP 4669 and tetrangulol and tetrangomycin from *Streptomyces rimosus* NRRL 3016. *J Bacteriol.* **179** (2), 470-6. (1997).
165. Eustáquio, A. S. *et al.*, Heterologous expression of novobiocin and clorobiocin biosynthetic gene clusters. *Appl Environ Microbiol.* **71** (5), 2452-9. (2005).
166. Gomez-Escribano, J. P. & Bibb, M. J., Engineering *Streptomyces coelicolor* for heterologous expression of secondary metabolite gene clusters. *Microb Biotechnol.* **4** (2), 207-15. (2011).
167. Jones, A. C. *et al.*, Phage P1-derived artificial chromosomes facilitate heterologous expression of the FK506 gene cluster. *PLoS One.* **8** (7), doi: 10.1371/journal.pone.0069319. (2013).

168. Shizuya, H. *et al.*, Cloning and stable maintenance of 300-kilobase-pair fragments of human DNA in *Escherichia coli* using an F-factor-based vector. *Proc Natl Acad Sci USA.* **89** (18), 8794-7. (1992).
169. Inoannou, P. A. *et al.*, A new bacteriophage P1-derived vector for the propagation of large human DNA fragments. *Nat Genet.* **6** (1), 84-9. (1994).
170. Sosio, M. *et al.*, Artificial chromosomes for antibiotic-producing actinomycetes. *Nat Biotechnol.* **18** (3), 343-5. (2000).
171. Bio S&T Inc, PAC Libraries - pESAC13 vector, Available at http://www.biost.com/page/PAC.aspx. (Accessed September 2014).
172. Hu, H., Zhang, Q. & Ochi, K., Activation of antibiotic biosynthesis by specified mutations in the *rpoB* gene (encoding the RNA polymerase beta subunit) of *Streptomyces lividans*. *J Bacteriol.* **184** (14), 3984-91. (2002).
173. Rückert, C. *et al.*, Complete genome sequence of *Streptomyces lividans* TK24. *J Biotechnol.* **199**, 21-2. (2015).
174. Braña, A. F. *et al.*, Activation and silencing of secondary metabolites in *Streptomyces albus* and *Streptomyces lividans* after transformation with cosmids containing the thienamycin gene cluster from *Streptomyces cattleya*. *Arch Microbiol.* **196** (5), 345-55. (2014).
175. Ding, L., Ndejouong, B. S., Maier, A., Fiebig, H. H. & Hertweck, C., Elaiomycins D-F, antimicrobial and cytotoxic azoxides from *Streptomyces* sp. strain HKI0708. *J Nat Prod.* **75** (10), 1729-34. (2012).
176. Hsiao, N. H. *et al.*, ScbA from *Streptomyces coelicolor* A3(2) has homology to fatty acid synthases and is able to synthesize gamma-butyrolactones. *Microbiology.* **153** (Pt 5), 1394-404. (2007).
177. Blair, L. M. & Sperry, J., Natural products containing a nitrogen-nitrogen bond. *J Nat Prod.* **76** (4), 794-812. (2013).
178. Wang, J. B. *et al.*, Characterization of AvaR1, an autoregulator receptor that negatively controls avermectins production in a high avermectin-producing strain. *Biotechnol Lett.* **36** (4), 813-9. (2014).
179. Nihira, T., Ikeda, H., Komatsu, M., Kitani, S. & Miyamoto, K. T., The autoregulator receptor homologue AvaR3 plays a regulatory role in antibiotic production, mycelial aggregation and colony development of *Streptomyces avermitilis*. *Microbiology.* **157** (Pt 8), 2266-75. (2011).
180. Kitani, S., Ikeda, H., Sakamoto, T., Noguchi, S. & Nihira, T., Characterization of a regulatory gene, *aveR*, for the biosynthesis of avermectin in *Streptomyces avermitilis*. *Appl Microbiol Biotechnol.* **82** (6), 1089-96. (2009).
181. Guo, J. *et al.*, The pathway-specific regulator AveR from *Streptomyces avermitilis* positively regulates avermectin production while it negatively effects oligomycin biosynthesis. *Mol Genet Genomics.* **283** (2), 123-33. (2010).
182. Gust, B., Challis, G. L., Fowler, K., Kieser, T. & Chater, K. F., PCR-targeted *Streptomyces* gene replacement identifies a protein domain needed for biosynthesis of the sesquiterpene soil odor geosmin. *Proc Natl Acad Sci USA.* **100** (4), 1541-6. (2003).

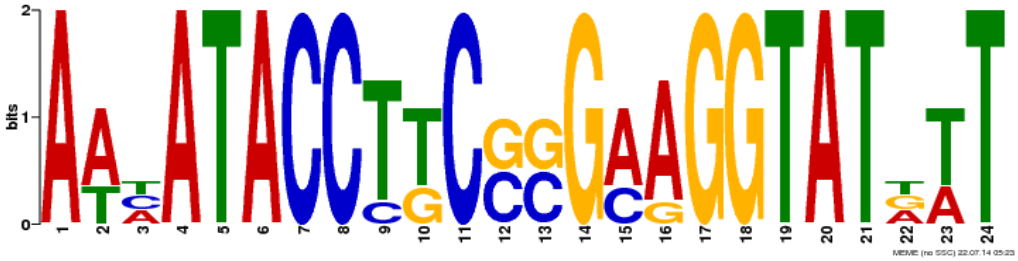
183. Raynal, A., Friedmann, A., Tuphile, K., Guerineau, M. & Pernodet, J. L., Characterization of the *attP* site of the integrative element pSAM2 from *Streptomyces ambofaciens*. *Microbiology.* **148** (Pt 1), 71-7. (2002).
184. Bibb, M. J., Ward, J. M. & Hopwood, D. A., Transformation of plasmid DNA into *Streptomyces* at high frequency. *Nature.* **274** (5669), 398-400. (1978).
185. MacNeil, D. J. & Klapko, L. M., Transformation of *Streptomyces avermitilis* by plasmid DNA. *J Ind Microbiol.* **2** (4), 209-18. (1987).
186. MacNeil, D. J., Characterization of a unique methyl-specific restriction system in *Streptomyces avermitilis*. *J Bacteriol.* **170** (12), 5607-12. (1988).
187. Baba, T. *et al.*, Construction of *Escherichia coli* K-12 in-frame, single-gene knockout mutants: the Keio collection. *Mol Syst Biol.* **2** (2006.0008), doi: 10.1038/msb4100050. (2006).
188. Mo, S. J., Kim, B. S. & Reynolds, K. A., Production of branched-chain alkylprodiginines in *S. coelicolor* by replacement of the 3-ketoacyl ACP synthase III initiation enzyme, RedP. *Chem Biol.* **12** (2), 191-200. (2005).
189. Kunitake, H., Hiramatsu, T., Kinashi, H. & Arakawa, K., Isolation and biosynthesis of an azoxyalkene compound produced by a multiple gene disruptant of *Streptomyces rochei*. *Chembiochem.* **16** (15), 2237-43. (2015).
190. Elkins, J. M., Kershaw, N. J. & Schofield, C. J., X-ray crystal structure of ornithine acetyltransferase from the clavulanic acid biosynthesis gene cluster. *Biochem J.* **385** (Pt 2), 565-73. (2005).
191. Cooper, J. M., Isolation of a toxic principle from the seeds of *Macrozamia spiralis*. *J Proc R Soc NSW.* **74**, 450-4. (1941).
192. Nishida, K., Kobayashi, A. & Nagahama, T., Studies on cycasin, a new toxic glycoside, of *Cycas revoluta* Thunb. *B Agr Chem Soc Japan.* **19** (1), 77-84. (1955).
193. Manderscheid, N. *et al.*, Elaiomycins K and L, new azoxy antibiotics from *Streptomyces* sp. Tü 6399*. *J Antibiot (Tokyo).* **66** (2), 85-8. (2013).
194. Yamato, M. & Umezawa, H., Valanimycin acts on DNA in bacterial cells. *J Antibiot (Tokyo).* **40** (4), 558-60. (1987).
195. Li, Y. & Zhang, Y., PhoU is a persistence switch involved in persister formation and tolerance to multiple antibiotics and stresses in *Escherichia coli*. *Antimicrob Agents Chemother.* **51** (6), 2092-9. (2007).
196. Riggs, N. V., The occurence of Macrozamin in the seeds of cycads. *Aust J Chem.* **7** (1), 123-3. (1954).
197. Fujiu, M. *et al.*, Azoxybacilin, a novel antifungal agent produced by *Bacillus cereus* NR2991. Production, isolation and structure elucidation. *J Antibiot (Tokyo).* **47** (7), 833-5. (1994).
198. Haskell, T. H., Ryder, A. & Bartz, Q. R., Elaiomycin, a new tuberculostatic antibiotic; isolation and chemical characterization. *Antibiot Chemother (Northfield).* **4** (2), 141-4. (1954).
199. Zhang, W., Ntai, I., Kelleher, N. L. & Walsh, C. T., tRNA-dependent peptide bond formation by the transferase PacB in biosynthesis of the pacidamycin group of pentapeptidyl nucleoside antibiotics. *Proc Natl Acad Sci USA.* **108** (30), 12249-53. (2011).

200. Sauguet, L. *et al.*, Cyclodipeptide synthases, a family of class-I aminoacyl-tRNA synthetase-like enzymes involved in non-ribosomal peptide synthesis. *Nucleic Acids Res.* **39** (10), 4475-89. (2011).
201. CRC Press, CHEMnetBASE, Available at http://dnp.chemnetbase.com/. (Accessed December 2015).

# 9. Appendices

## 9.1. Appendix A – Initial consensus ARE motif for *S. coelicolor*, *S. venezuelae*, *S. avermitilis* and *S. hygroscopicus* AHFCA-dependent signalling systems

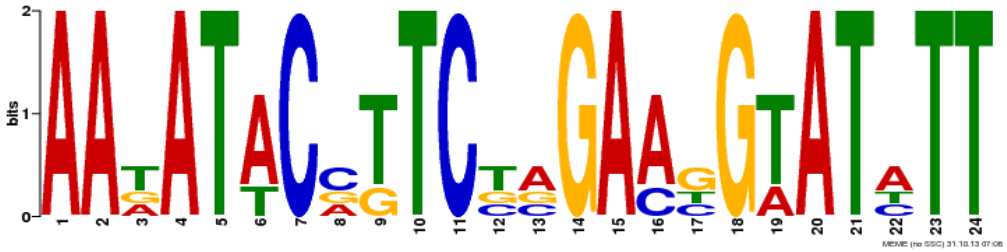
### *Streptomyces coelicolor* A3(2)



E-value: 5.9e-005

| Name | Bp Upstream of start codon | *p*-value | Sites |
|---|---|---|---|
| SCP1.243 | 94 | 2.10e-14 | AACCCATTGC ATAATACCTTCCCGCAGGTATATT TCTCTCGGTC |
| SCP1.242c | 76 | 2.44e-14 | GACCGAGAGA AATATACCTGCGGGAAGGTATTAT GCAATGGGTT |
| SCP1.246 | 44 | 1.27e-13 | CGCCCCCACT AACATACCTTCCCGAGGGTATGTT TTCCGGGCCC |

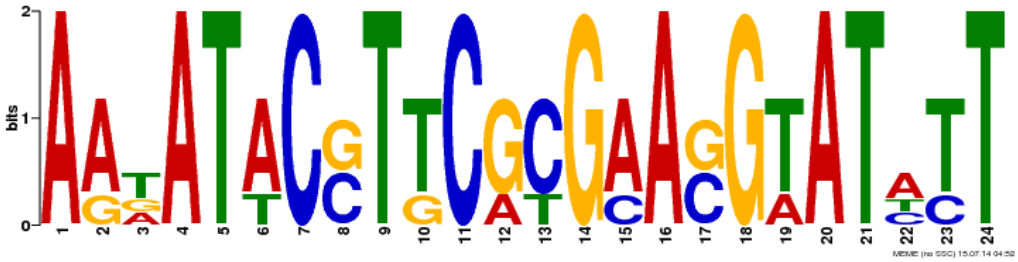
### *Streptomyces venezuelae* ATCC 10712



E- value: 3.2e-010

| Name | Bp Upstream of start codon | *p*-value | Sites |
|---|---|---|---|
| SVEN_4183 | 65 | 2.15e-15 | TATGCCTTGC AATATACATTCTCGAAGGTATATT TCTTGCTGCC |
| SVEN_4182 | 32 | 2.15e-15 | GGCAGCAAGA AATATACCTTCGAGAATGTATATT GCAAGGCATA |
| SVEN_4188 | 41 | 8.82e-13 | GGCCCCACAG AAAATACCTTCCAGACCGAATCTT TTTCTACTCG |
| SVEN_4187 | 69 | 1.25e-12 | CGAGTAGAAA AAGATTCGGTCTGGAAGGTATTTT CTGTGGGGCC |

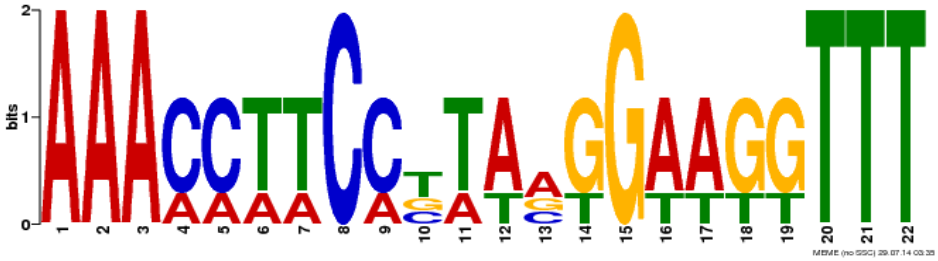
## ***Streptomyces avermitilis* MA-4680**



E-value: 5.7e-008

| Name | Bp Upstream of start codon | *p*-value | Sites |
|---|---|---|---|
| SAV_2270 | 95 | 1.41e-14 | AGGCATCATT AATATACCTGCGCGAAGGTATATT GCAAGGCTCG |
| SAV_2269 | 72 | 1.41e-14 | CGAGCCTTGC AATATACCTTCGCGCAGGTATATT AATGATGCCT |
| SAV_2268 | 19 | 8.44e-13 | CTCGATGATT AGAATTCGTTCACGAACGTATCTT TTATGCTTCG |
| SAV_2267 | 25 | 8.44e-13 | CGAAGCATAA AAGATACGTTCGTGAACGAATTCT AATCATCGAG |

## ***Streptomyces hygroscopicus* subsp. Jinggangensis 5008**

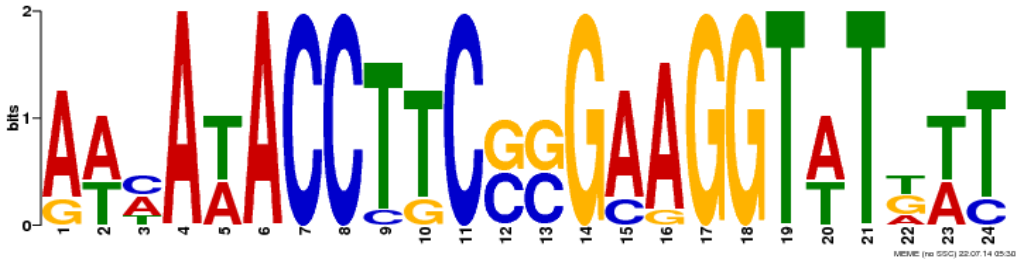


E-value: 1.2e-008

| Name | Bp Upstream of start codon | *p*-value | Sites |
|---|---|---|---|
| SHJG_7319 | 163 | 1.28e-12 | GCCAGGGAAT AAACCTTCCCAAAGGTAGGTTT GACGTGGGGG |
| SHJG_7322 | 30 | 1.45e-12 | TCCGCTGATA AAAAATTCCGTAATGATGGTTT GCTCTTCGTT |
| SHJG_7318 | 68 | 1.45e-12 | CCCCCACGTC AAACCTACCTTTGGGAAGGTTT ATTCCCTGGC |
| SHJG_7323 | 214 | 1.62e-12 | AACGAAGAGC AAACCATCATTACGGAATTTTT TATCAGCGGA |

## 9.2. Appendix B – Refined consensus ARE motif for *S. coelicolor*, *S. venezuelae*, *S. avermitilis* and *S. hygroscopicus* AHFCA-dependent signalling systems

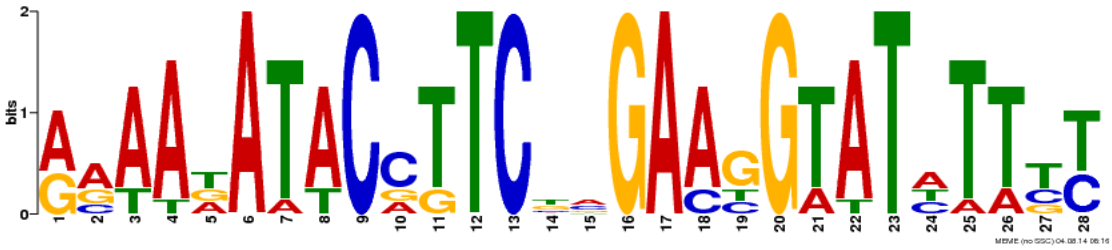
### *Streptomyces coelicolor* A3(2)



E-value: 2.0e-013

| Name | Bp Upstream of start codon | *p*-value | Sites |
|---|---|---|---|
| SCP1.243 | 94 | 2.40e-13 | AACCCATTGC ATAATACCTTCCCGCAGGTATATT TCTCTCGGTC |
| SCP1.242c | 76 | 2.66e-13 | GACCGAGAGA AATATACCTGCGGGAAGGTATTAT GCAATGGGTT |
| SCP1.230 | 13 | 5.55e-13 | TACGAGTCAT AAAAAACCTTCGGGAAGGTTTGAC ACTGTGAGGC |
| SCP1.229c | 193 | 6.62e-13 | GCCTCACAGT GTCAAACCTTCCCGAAGGTTTTTT ATGACTCGTA |
| SCP1.246 | 44 | 7.82e-13 | CGCCCCCACT AACATACCTTCCCGAGGGTATGTT TTCCGGGCCC |

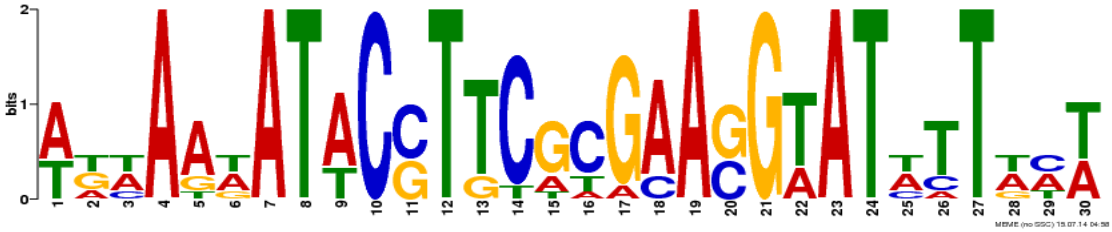
### *Streptomyces venezuelae* ATCC 10712



E-value: 5.3e-018

| Name | Bp Upstream of start codon | *p*-value | Sites |
|---|---|---|---|
| SVEN_4182 | 30 | 3.31e-16 | CGGCAGCAA GAAATATACCTTCGAGAATGTATATTGC AAGGCATAC |
| SVEN_4183 | 63 | 4.67e-16 | GTATGCCTT GCAATATACATTCTCGAAGGTATATTTC TTGCTGCCG |
| SVEN_4188 | 39 | 1.88e-14 | CGGCCCCAC AGAAAATACCTTCCAGACCGAATCTTTT TCTACTCGC |
| SVEN_4187 | 67 | 2.21e-14 | GCGAGTAGA AAAAGATTCGGTCTGGAAGGTATTTTCT GTGGGGCCG |
| SVEN_4179 | 10 | 5.01e-13 | CCTGATTAC AGTTGATACCTTCGTGAAGGTTTTTATT TCATGGGTC |

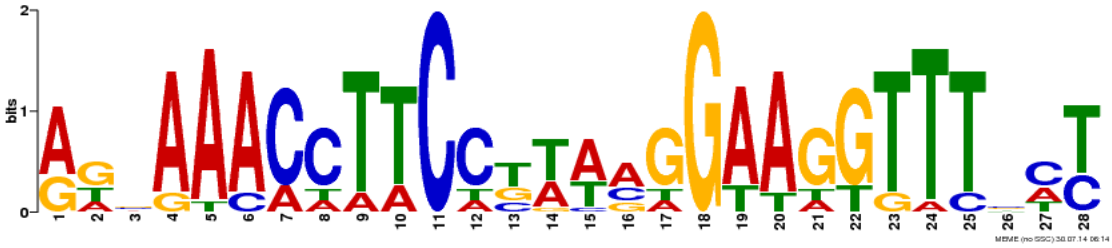
## ***Streptomyces avermitilis* MA-4680**



E-value: 1.3e-011

| Name | Bp Upstream of start codon | *p*-value | Sites |
|---|---|---|---|
| SAV_2270 | 92 | 1.55e-15 | TAGGCATC ATTAATATACCTGCGCGAAGGTATATTGCA AGGCTCGAC |
| SAV_2269 | 69 | 1.72e-15 | TCGAGCCT TGCAATATACCTTCGCGCAGGTATATTAAT GATGCCTAG |
| SAV_2267 | 22 | 1.98e-14 | CCGAAGCA TAAAAGATACGTTCGTGAACGAATTCTAAT CATCGAGTG |
| SAV_2268 | 14 | 2.90e-14 | ACTCGATG ATTAGAATTCGTTCACGAACGTATCTTTTA TGCTTCGGT |
| SAV_2301 | 175 | 3.36e-14 | GATTCAAC AGAAAAATACCTTCTCAAAGGAATTATTCT GTGCGAGGA |

## ***Streptomyces hygroscopicus* subsp. Jinggangensis 5008**

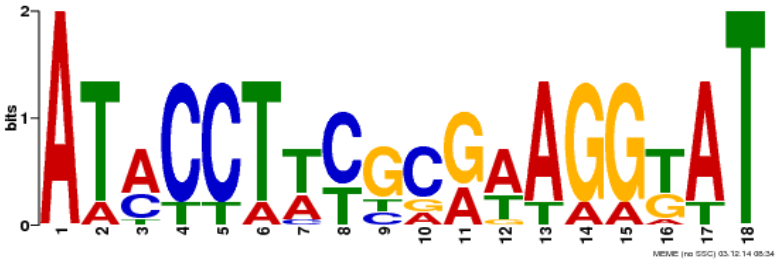


E-value: 4.9e-012

| Name | Bp Upstream of start codon | *p*-value | Sites |
|---|---|---|---|
| SHJG_7318 | 65 | 2.20e-13 | ACCCCCCAC GTCAAACCTACCTTTGGGAAGGTTTATT CCCTGGCAA |
| SHJG_7319 | 159 | 3.04e-13 | TTGCCAGGG AATAAACCTTCCCAAAGGTAGGTTTGAC GTGGGGGGT |
| SHJG_7323 | 211 | 3.38e-13 | GAAACGAAG AGCAAACCATCATTACGGAATTTTTTAT CAGCGGAGC |
| SHJG_7322 | 27 | 5.12e-13 | GCTCCGCTG ATAAAAAATTCCGTAATGATGGTTTGCT CTTCGTTTC |
| SHJG_7504 | 3 | 3.39e-11 | CACCCGGTA AGGAAACTTTCCTAACAGAAGGGTCCCC CA |
| SHJG_7505 | 176 | 3.63e-11 | ACGCATGTG GGGGACCCTTCTGTTAGGAAAGTTTCCT TACCGGGTG |
| SHJG_7324 | 76 | 5.70e-11 | CCTACATCC GGGAAAACTACCTGTGGGAAGGTATTTT TGGCTGCGG |

## 9.3. Appendix C – Refined consensus ARE motifs for *S. scerotialus*, *S.* sp. HGB0020, *S. roseochromogenes* and *K. cheerisanensis* AHFCA-dependent signalling systems

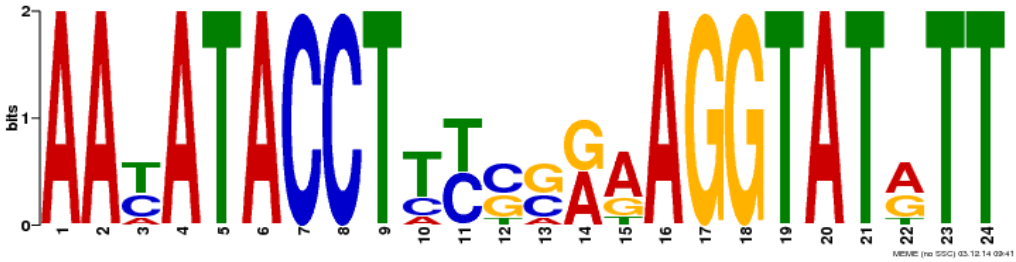
### *Streptomyces sclerotialus* NRRL ISP-5269



E-value: 9.0e-004

| Name | Bp Upstream of start codon | *p*-value | Sites |
|---|---|---|---|
| IG96_RS0135495 | 171 | 5.43e-10 | GCGAAAAAAT ATACCTATGGGAAGGTAT ATTGCATGGG |
| IG96_RS0135500 | 162 | 6.96e-10 | CCCATGCAAT ATACCTTCCCATAGGTAT ATTTTTTCGC |
| IG96_RS0135490 | 114 | 3.19e-9 | ATCTCAATTG AAACCTTCGCGAAGGTTT GATTCTGTGG |
| ORF1 | 43 | 1.25e-8 | TGACATTTGA ATACCTCCGCGAAGGAAT GATTTCGTGA |
| IG96_RS0135515 | 52 | 1.59e-7 | TGTGTTGGGC ATCCCAATGAGAAAAGAT CCGCCTTGGA |
| IG96_RS0135520 | 104 | 1.75e-7 | TCCAAGGCGG ATCTTTTCTCATTGGGAT GCCCAACACA |

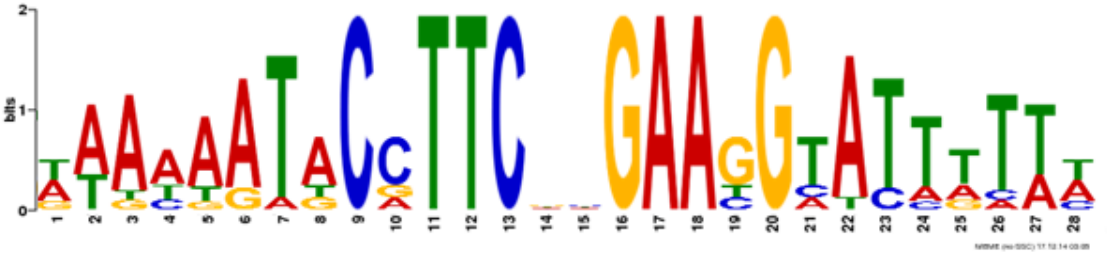
### *Kitasatospora cheerisanensis* KCTC 2395



E-value: 8.4-007

| Name | Bp Upstream of start codon | *p*-value | Sites |
|---|---|---|---|
| KCH_13210 | 117 | 1.47e-15 | GGCTCCGAAG AAAATACCTTCCCAAAGGTATGTT GCCTGTGGCA |
| KCH_13290 | 78 | 3.51e-15 | TTGCGCCAAT AACATACCTATTGGAAGGTATGTT ATCTGTCACC |
| KCH_13230 | 104 | 3.83e-15 | ACGATGAGAA AATATACCTCTGGGAAGGTATATT GCTAGGGGGA |

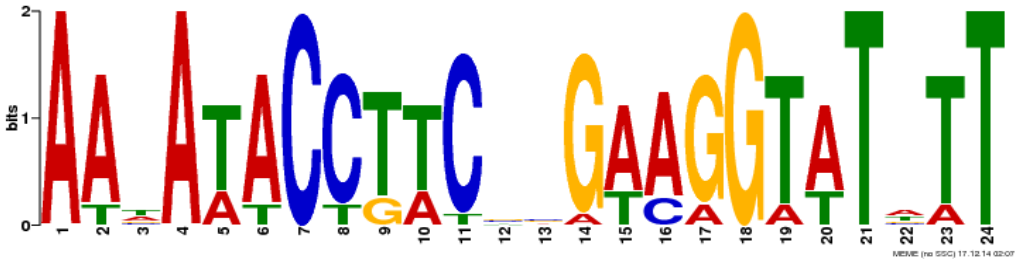
## ***Streptomyces* sp. HGB0020**



E-value: 8.5e-022

| Name | Bp Upstream of start codon | *p*-value | Sites |
|---|---|---|---|
| HMPREF1211_06616 | 29 | 1.18e-14 | GGATCTCC AAATAATACCTTCTTGAACGCACTTTTT TGCTACC |
| HMPREF1211_06615 | 48 | 1.56e-14 | GGGTAGCA AAAAAGTGCGTTCAAGAAGGTATTATTT GGAGATC |
| HMPREF1211_06619 | 6 | 2.03e-14 | AAGCCCGG TTAAAAAACCTTCACGAAGGAATTTAAA TGC |
| HMPREF1211_06613 | 12 | 3.42e-14 | GCCGCAAC TTACAATACATTCCCGAAGGTATATTTC AAGGAAC |
| HMPREF1211_06614 | 70 | 5.62e-14 | TGTTCCTT GAAATATACCTTCGGGAATGTATTGTAA GTTGCGG |
| HMPREF1211_06629 | 45 | 1.32e-13 | CCTCCTCA TAGTGATACCTTCTCGAAGGAATTTTTT CGTGAGC |

## ***Streptomyces roseochromogenes* subsp Oscitans DC 12.976**

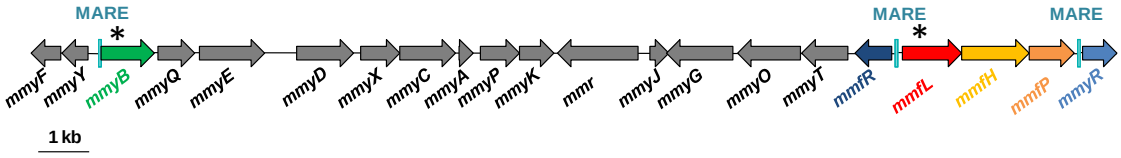


E-value: 9.4-e025

| Name | Bp Upstream of start codon | *p*-value | Sites |
|---|---|---|---|
| M878_26350 | 85 | 1.12e-14 | ACCACTTTT AATATACCTTCCTGTAGGTATATT TGCCACGGGC |
| M878_26345 | 65 | 1.50e-14 | CCCGTGGCA AATATACCTACAGGAAGGTATATT AAAAGTGGTA |
| M878_26320 | 22 | 1.44e-12 | TGAGCAGTT ATAAAACCTTCTCGAAGGTATCTT ACGGAGACTG |
| M878_26325 | 183 | 1.91e-12 | AGTCTCCGT AAGATACCTTCGAGAAGGTTTTAT AACTGCTCAC |
| M878_26325 | 156 | 1.53e-11 | TGAACGGTG AATATTCTGTCCCGTAGGTTTTTT TCGCGAGCCA |
| M878_26365 | 46 | 1.62e-11 | TCCGCAAAT AACATACCGTTGAGAAGGTATGTT GTTCCTGCCG |
| M878_26320 | 156 | 1.84e-11 | GGCTCGCGA AAAAAACCTACGGGACAGAATATT CACCGTTCAA |

## 9.4. Appendix D – *S. coelicolor* A3(2) methylenomycin gene cluster

***S. coelicolor* A3(2)**
**(19,341 bp)**

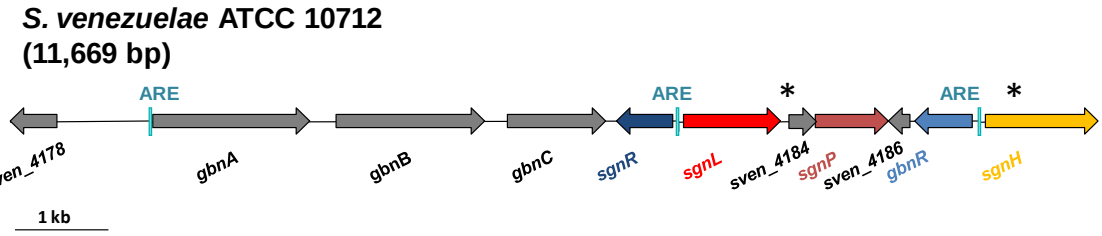



*Asterisks indicate presence of TTA codon

| Protein | Homologue (% identity / % similarity) | Proposed function |
|---|---|---|
| MmyF (SCP1.228c) | Monooxygenase *S. mirabilis* (62 / 75) | Oxidoreductase |
| MmyY (SCP1.229c) | Hypothetical protein *S. natalensis* (58 / 67) | Hypothetical protein |
| MmyB (SCP1.230) | DNA-binding protein *S. sclerotialus* (46 / 62) | DNA-binding protein |
| MmyQ (SCP1.231) | Monoxygenase *S. mirabilis* (76 / 87) | Oxidoreductase |
| MmyE (SCP1.232) | 12-oxophytodienoate reductase *Paenibacillus alvei* (58 / 70) | Enoyl reductase |
| MmyD (SCP1.233) | Hypothetical protein *S. rochei* (48 / 60) | Hypothetical protein |
| MmyX (SCP1.233A) | ATP/GTP-binding protein *S. mirabilis* (79 / 89) | ATP/GTP-binding protein |
| MmyC (SCP1.233B) | 3-oxoacyl-ACP synthase *S. roseus* (74 / 84) | ACP synthase |
| MmyA (SCP1.234) | Acyl carrier protein *S. albus* (37 / 60) | ACP |
| MmyP (SCP1.235) | Haloacid dehalogenase family hydrolase *S. regensis* (46 / 58) | Phosphatase |
| MmyK (SCP1.236) | ATP/GTP-binding protein *S. hygroscopicus* (60 / 76) | ATP/GTP-binding protein |
| MmR (SCP1.237c) | MFS transporter *S. mirabilis* (75 / 84) | Efflux membrane protein |
| MmyJ (SCP1.238) | ArsR family transcriptional regulator *S. bingchenggensis* (63 / 82) | ArsR family transcriptional regulator |
| MmyG (SCP1.239c) | Oxidoreductase *S. acidscabies* (54 / 64) | Oxidoreductase |
| MmyO (SCP1.240c) | Monoxygenase *S. albus* (55 / 68) | Monooxygenase |
| MmyT (SCP1.241c) | Thioesterase *S. roseus* (38 / 49) | Thioesterase |
| MmfR (SCP1.242c) | TetR family transcriptional regulator *S. olindensis* (66 / 77) | TetR family transcriptional regulator |
| MmfL (SCP1.243) | A-factor biosynthesis protein *Rhodococcus imtechensis* (34 / 48) | AfsA homologue |

| | | |
|---|---|---|
| MmfH (SCP1.244) | Oxidoreductase *S. aureofaciens* (58 / 68) | Oxidoreductase |
| MmfP (SCP1.245) | Haloacid dehalogenase family hydrolase *Streptomyces himastatinicus* (42 / 51) | Phosphatase |
| MmyR (SCP1.246) | TetR family transcriptional regulator *Streptomyces lavenduligriseus* (36/ 50) | TetR family transcriptional regulator |

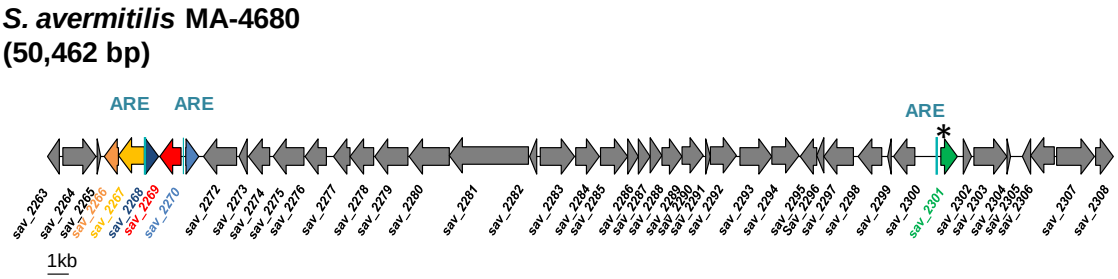
## 9.5. Appendix E – *S. venezuelae* ATCC 10712 gaburedin gene cluster




*Asterisks indicate presence of TTA codon

| Protein | Homologue (% identity / % similarity) | Proposed function |
|---|---|---|
| SVEN_4178 | Acetyltransferase<br>*Nocardio gilva* (42 / 64) | Acyltransferase |
| GbnA<br>(SVEN_4179) | CadA<br>*E. coli* (25 / 41) | Glutamate decarboxylase |
| GbnB<br>(SVEN_4180) | EntE<br>*E. coli* (25 / 40) | Urea synthase |
| GbnC<br>(SVEN_4181) | EamA family transporter<br>*Salmonella enterica* (55 /71) | Transmembrane efflux |
| SgnR<br>(SVEN_4182) | MmfR<br>*S. coelicolor* (54 / 71) | MmfR homologue |
| SgnL<br>(SVEN_4183) | MmfL<br>*S. coelicolor* (33 / 44) | MmfL homologue |
| SVEN_4184 | LysR family transcriptional regulator<br>*S. vietnamensis* (98 / 100) | LysR family transcriptional regulator |
| SgnP<br>(SVEN_4185) | MmfP<br>*S. coelicolor* (42 / 56) | MmfP homologue |
| SVEN_4186 | Hypothetical protein | Hypothetical protein |
| GbnR<br>(SVEN_4187) | MmyR<br>*S. coelicolor* (47 / 54) | MmyR homologue |
| SgnH<br>(SVEN_4188) | MmfH<br>*S. coelicolor* (45 / 57 | MmfH homologue |

## 9.6. Appendix F – *S. avermitilis* MA-4680 azoxy biosynthetic gene cluster



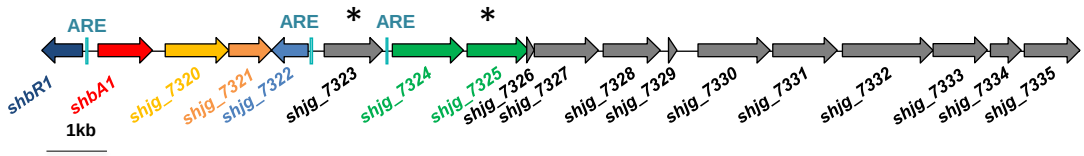

*Asterisks indicate presence of TTA codon

| Protein | Homologue (% identity / % similarity) | ProposedFunction |
|---|---|---|
| SAV_2263 | TetR family transcriptional regulator<br>*S. rimosus* (66 / 77) | TetR family transcriptional regulator |
| SAV_2264 | Transporter<br>*S. mirabilis* (67 / 77) | Transmembrane efflux protein |
| SAV_2265 | Hypothetical protein | Hypothetical protein |
| SAV_2266 | MmfP<br>*S. coelicolor* (37 / 51) | MmfP homologue |
| SAV_2267 | MmfH<br>*S. coelicolor* (45 / 58) | MmfH homologue |
| SAV_2268 | MmyR<br>*S. coelicolor* (39 / 54) | MmyR homologue |
| SAV_2269 | MmfL<br>*S. coelicolor* (34 / 48) | MmfL homologue |
| SAV_2270 | MmfR<br>*S. coelicolor* (47 / 64) | MmfR homologue |
| SAV_2272 | VlmK<br>*S. viridifaciens* (55 / 70) | MmgE/PrpD family protein |
| SAV_2273 | Isomerase<br>*S. collius* (80 / 89) | Isomerase |
| SAV_2274 | VlmJ<br>*S. viridifaciens* (44 / 59) | Diacylglycerol kinase |
| SAV_2275 | Major facilitator superamily transporter<br>*S. collinus* (51 / 65) | Transmembrane efflux protein |
| SAV_2276 | 3-oxoacyl-ACP synthase<br>*S. griseofuscus* (73 / 84) | 3-oxoacyl-ACP synthase |
| SAV_2277 | Thioesterase<br>*S. mutabilis* (86 / 93) | Thioesterase |
| SAV_2278 | F420-dependent dehydrogenase<br>*S. mutabilis* (87 / 90) | F420-dependent dehydrogenase |
| SAV_2279 | Long-chain-fatty-acid-CoA ligase<br>*S. mutabilis* (84 / 90) | Long-chain-fatty-acid-CoA ligase |
| SAV_2280 | Modular polyketide synthase<br>*S. niger* (53 / 62) | Modular polyketide synthase |
| SAV_2281 | Modular polyketide synthase<br>*Longispora albida* (48 / 61) | Modular polyketide synthase |
| SAV_2282 | Acyl carrier protein<br>*S. mutabilis* (77 / 90) | ACP |
| SAV_2283 | Aldehyde dehydrogenase<br>*S. collinus* (95 / 97) | Aldehyde dehydrogenase |

| | | |
|---|---|---|
| SAV_2284 | VlmH<br>*S. viridifaciens* (56 / 74) | Isobutylamine N-hydroxylase |
| SAV_2285 | Ornithine aminotransferase<br>*S. collinus* (93 / 96) | Ornithine aminotransferase |
| SAV_2286 | Cyclase<br>*S. mutabilis* (86 / 91) | Cyclase |
| SAV_2287 | VlmO<br>*S. viridifaciens* (63 / 71) | Integral membrane protein |
| SAV_2288 | VlmR<br>*S. viridifaciens* (38 / 54) | NADPH-flavin oxidoreductase |
| SAV_2289 | VlmB<br>*S. viridifaciens* (61 / 77) | Hypothetical protein |
| SAV_2290 | 3-oxoacyl-ACP synthase<br>*S. collinus* (85 / 92) | 3-oxoacyl-ACP synthase |
| SAV_2291 | Acyl carrier protein<br>*S. bingchenggensis* (70 / 81) | ACP |
| SAV_2292 | 3-oxoacyl-ACP synthase<br>*S. mutabilis* (93 / 96) | 3-oxoacyl-ACP synthase |
| SAV_2293 | VlmC<br>*S. viridifaciens* (54 / 71) | Amino acid permease |
| SAV_2294 | VlmL<br>*S. viridifaciens* (59 / 70) | Seryl-tRNA synthetase |
| SAV_2295 | Enoyl-ACP reductase<br>*S. collinus* (87 / 91) | Enoyl-ACP reductase |
| SAV_2296 | Ferredoxin<br>*S. mutabilis* (81 / 87) | Ferredoxin |
| SAV_2297 | Major facilitator superfamily transporter<br>*S. collinus* (83 / 89) | MFS transporter |
| SAV_2298 | Fatty acid desaturase<br>*S. collinus* (91 / 95) | Fatty acid desaturase |
| SAV_2299 | Hypothetical protein<br>*S. aureocirculatus* (52 / 65) | Hypothetical protein |
| SAV_2300 | VlmA<br>*S. viridifaciens* (57 / 68) | Integral membrane lysyl-tRNA synthetase |
| SAV_2301 | VlmI<br>*S. viridifaciens* (47 / 61) | SARP family transcriptional regulator |
| SAV_2302 | Membrane protein<br>*S. mutabilis* (81 / 82) | Membrane protein |
| SAV_2303 | Methylmalonyl-CoA carboxyltransferase<br>*S. griseofuscus* (89 / 93) | Methylmalonyl-CoA carboxyltransferase |
| SAV_2304 | Hypothetical protein<br>*S. griseofuscus* (68 / 77) | Hypothetical protein |
| SAV_2305 | Enamine deaminase<br>*S. sviceus* (84 /91) | Enamine deaminase |
| SAV_2306 | Acyl-CoA dehydrogenase<br>*S. fulvoviolaceus* (89 / 92) | Acyl-CoA dehydrogenase |
| SAV_2307 | Fatty acid-CoA synthetase<br>*S. collinus* (89 / 91) | Acyl-CoA synthetase |
| SAV_2308 | PaaX family transcriptional regulator<br>*S. fulvoviolaceus* (91 / 92) | PaaX family transcriptional regulator |

## 9.7. Appendix G – *S. hygroscopicus* subsp. Jinggangensis 5008 gene cluster

**_S. hygroscopicus_ subsp. Jinggangensis 5008 (17,307 bp)**

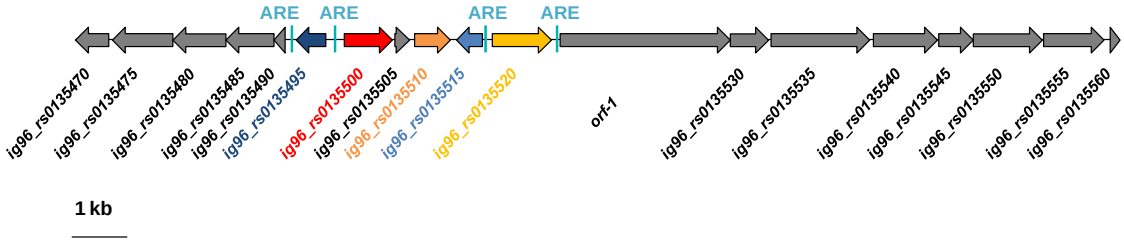



*Asterisks indicate presence of TTA codon

| Protein | Homologue (% identity / % similarity) | Proposed function |
|---|---|---|
| SHJG_7318 (ShbR1) | MmfR *S. coelicolor* (51 / 67) | MmfR homologue |
| SHJG_7319 (ShbA1) | MmfL *S. coelicolor* (36 / 46) | MmfL homologue |
| SHJG_7320 | MmfH *S. coelicolor* (48 / 57) | MmfH homologue |
| SHJG_7321 | MmfP *S. coelicolor* (42 / 53) | MmfP homologue |
| SHJG_7322 | MmyR *S. coelicolor* (37 / 53) | MmyR homologue |
| SHJG_7323 | 5,10-methylene tetrahydromethanopterin reductase *S. scabiei* (46 / 61) | Monooxygenase |
| SHJG_7324 | TylR *S. fradiae* (32 / 45) | TylR-like regulatory protein |
| SHJG_7325 | StrR *S. griseus* (48 / 63) | Streptomycin biosynthesis operon regulator |
| SHJG_7326 | Hypothetical protein | Hypothetical protein |
| SHJG_7327 | Hypothetical protein *S. pluripotens* (46 / 56) | Hypothetical protein |
| SHJG_7328 | Hypothetical protein *S. lavendulae* (42 / 51) | Hypothetical protein |
| SHJG_7329 | Hypothetical protein *S. rapamycinicus* (44 / 57) | Hypothetical protein |
| SHJG_7330 | Acyl-CoA dehydrogenase *S. achromogenes* (75 / 85) | Acyl-CoA dehydrogenase |
| SHJG_7331 | Alcohol dehydrogenase *S. achromogenes* (86 / 93) | Alcohol dehydrogenase |
| SHJG_7332 | Trehalose-phosphate synthase *S. achromogenes* (71 / 78) | Trehalose-phosphate synthase |
| SHJG_7333 | UTP-glucose-1-phosphate uridylyltransferase *S. vietnamensis* (80 / 87) | UTP-glucose-1-phosphate uridylyltransferase |
| SHJG_7334 | Flavin reductase *S. achromogenes* (77 / 82) | Flavin reductase |
| SHJG_7335 | Oxidoreductase *S. achromogenes* (48 / 54) | Oxidoreductase |

## 9.8. Appendix H – *S. sclerotialus* NRRL ISP-5269 gene cluster

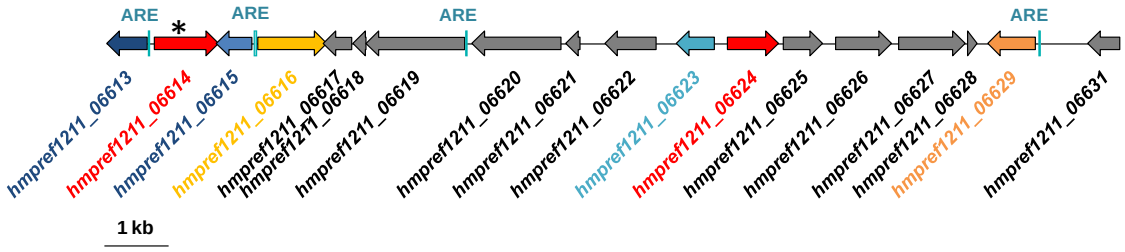
***S. sclerotialus* NRRL ISP-5269**
**(19,981 bp)**




| Protein | Homologue (% identity / % similarity) | Proposed function |
|---|---|---|
| IG96_RS0135470 | 4-phosphopantetheinyl transferase *S. pristinaespiralis* (99 / 54) | 4-phosphopantetheinyl transferase |
| IG96_RS0135475 | QncL *S. melanovinaceus* (37 / 49) | QncL-like acyltransferase |
| IG96_RS0135480 | QncL *S. melanovinaceus* (97 / 51) | QncL-like dehydrogenase |
| IG96_RS0135485 | QncN *S. melanovinaceus* (96 / 61) | QncN-like dehydrogenase |
| IG96_RS0135490 | QncM *S. melanovinaceus* (97 / 32) | QncM-like ACP |
| IG96_RS0135495 | MmfR *S. coelicolor* (56 / 73) | MmfR-like transcriptional regulator |
| IG96_RS0135500 | MmfL *S. coelicolor* (35 / 47) | MmfL homologue |
| IG96_RS0135505 | LysR family transcriptional regulator *S. venezuelae* (92 / 61) | LysR family transcriptional regulator |
| IG96_RS0135510 | MmfP *S. coelicolor* (46/ 56) | MmfP homologue |
| IG96_RS0135515 | MmyR *S. coelicolor* (43 / 61) | MmyR homologue |
| IG96_RS0135520 | MmfH *S. coelicolor* (49 / 59) | MmfH homologue |
| ORF-1 | Non-ribosomal peptide synthetase *S.* sp. DSM 40835 (96 / 34) | Non-ribosomal peptide synthetase |
| IG96_RS0135530 | Thioesterase *S. sviceus* (45 / 56) | Thioesterase |
| IG96_RS0135535 | Anthranilate synthase *S. griseus* (57 / 70) | Anthranilate syntase |
| IG96_RS0135540 | Phospho-2-dehydro-3-deoxyheptonate aldolase *S. scabiei* (60 / 71) | Phospho-2-dehydro-3-deoxyheptonate aldolase |
| IG96_RS0135545 | Isochorismatase *S. scabiei* ( 59 / 70) | Isochorismatase |
| IG96_RS0135550 | Hypothetical protein *S. yeochonensis* (50 / 65) | Hypothetical protein |
| IG96_RS0135555 | Major Facilitator Superfamily Permease *S. tsukubaensis* (30 / 49) | MFS transporter |
| IG96_RS0135560 | Chorismate synthase *S. ochraceiscleroticus* (62 / 63) | Chorismate synthase |

## 9.9. Appendix I – *S.* sp. HGB0020 gene cluster

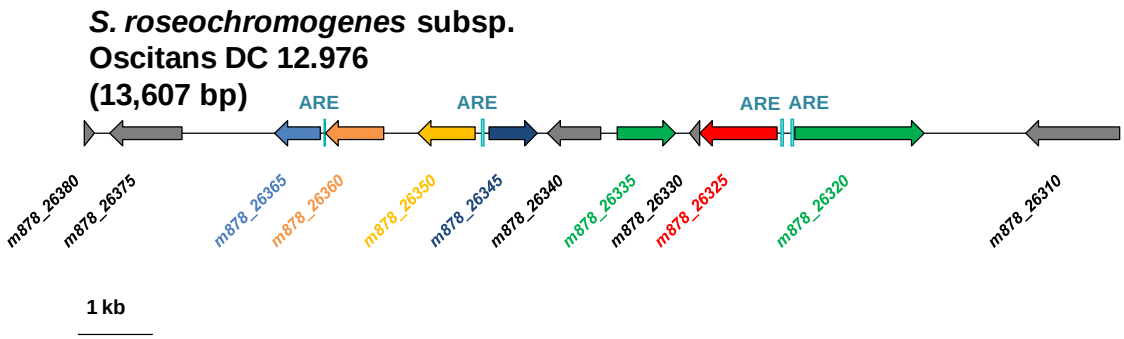
***Streptomyces* sp. HGB0020**
**(16,378 bp)**




*Asterisks indicate presence of TTA codon

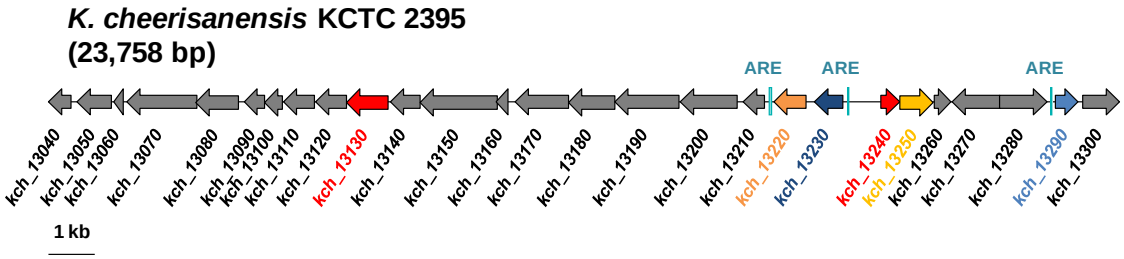
| Protein | Homologue (% identity / % similarity) | Proposed function |
|---|---|---|
| HMPREF1211_06613 | MmfR *S. coelicolor* (52 / 70) | MmfR homologue |
| HMPREF1211_06614 | BarX *S. virginiae* (36 / 48) | AfsA homologue |
| HMPREF1211_06615 | MmyR *S. coelicolor* (48 / 58) | MmyR homologue |
| HMPREF1211_06616 | MmfH *Streptomyces coelicolor* (45 / 56) | MmfH homologue |
| HMPREF1211_06617 | MerR family transcriptional regulator *S. griseus* (42 / 51) | MerR Family transcriptional regulator |
| HMPREF1211_06618 | Hypothetical protein *S. pristinaespiralis* (45 / 53) | Hypothetical protein |
| HMPREF1211_06619 | Methylmalonyl-CoA carboxyltransferase *S. vinaceus* (82 / 90) | Methylmalonyl-CoA carboxyltransferase |
| HMPREF1211_06620 | Major Facilitator Superfamily Protein *S. natalensis* (77 / 86) | MFS transporter |
| HMPREF1211_06621 | Tautomerase *S. xiamenensis* (86 /76) | Tautomerase |
| HMPREF1211_06622 | 3- oxoacyl-ACP reductase *S. avermitilis* (39 / 50) | 3-oxoacyl-ACP reductase |
| HMPREF1211_06623 | TetR family transcriptional regulator *S. incarnatus* (52 / 68) | TetR family transcriptional regulator |
| HMPREF1211_06624 | A-factor biosynthesis protein *S. clavuligerus* (35 / 48) | AfsA homologue |
| HMPREF1211_06625 | Hypothetical protein *S. roseus* (57 / 72) | Hypothetical protein |
| HMPREF1211_06626 | Hypothetical protein *S. roseus* (94 / 66) | Hypothetical protein |
| HMPREF1211_06627 | 3-oxoacyl-ACP synthase *S. virginiae* (69 / 77) | 3-oxoacyl-[acyl-carrier-protein] synthase III |
| HMPREF1211_06628 | Hypothetical protein *S. griseorubens* (67 / 66) | Hypothetical protein |
| HMPREF1211_06629 | MmfP *S. coelicolor* (40 / 53) | Phosphatase |
| HMPREF1211_06631 | Hypothetical protein *S. cyaneogriseus* (72 / 33) | Hypothetical protein |

## 9.10. Appendix J – *S. roseochromogenes* subsp. Oscitans DC 12.976 cluster




| Protein | Homologue (% identity / % similarity) | Proposed function |
|---|---|---|
| M878_26380 | Hypothetical protein<br>*S. avermitilis* (64 / 70) | Hypothetical protein |
| M878_26375 | Methylglyoxal synthase<br>*Leptospira santarosai* (59 / 76) | Methylglyoxal synthae |
| M878_26365 | MmyR<br>*S. coelicolor* (45 / 65) | MmyR homologue |
| M878_26360 | MmfP<br>*S. coelicolor* (47 / 56) | MmfP homologue |
| M878_26350 | MmfH<br>*S. coelicolor* (74 /59) | MmfH homologue |
| M878_26345 | MmfR<br>*S. coelicolor* (61 / 78) | MmfR homologue |
| M878_26340 | Hypothetical protein<br>*S. lydicus* (79 / 86) | Hypothetical protein |
| M878_26335 | MarR family transcriptional regulator<br>*S. rimosus* (85 / 89) | MarR family transcriptional regulator |
| M878_26330 | Hypothetical protein<br>*Streptacidiphilus rugosus* (51 / 55) | Hypothetical protein |
| M878_26325 | BarX<br>*S. virginiae* (43 / 54) | AfsA homologue |
| M878_26320 | SARP family transcriptional regulator<br>*S. albus* (45 / 59) | SARP family transcriptional regulator |
| M878_26310 | Carboxylate-amine ligase<br>*S. roseus* (48 / 55) | Carboxylate-amine ligase |

## 9.11. Appendix K – *K. cheerisanensis* KCTC 2395 gene cluster




| Protein | Homologue (% identity / % similarity) | Proposed Function |
|---|---|---|
| KCH_13040 | Polyketide cyclase<br>*S. aureofaciens* (46 / 59) | Polyketide cyclase |
| KCH_13050 | 4'-phosphopantetheinyl transferase<br>*S. actuosus* (62 / 68) | 4'-phosphopantetheinyl transferase |
| KCH_13060 | Hypothetical protein<br>*S. nodosus* (50 / 61) | Hypothetical protein |
| KCH_13070 | Methylmalonyl-CoA carboxyltransferase<br>*S. iranensis* (88 / 93) | Methylmalonyl-CoA carboxyltransferase |
| KCH_13080 | Malonyl CoA-ACP transacylase<br>*Nocardia vinacea* (63 / 74) | Malonyl CoA-ACP transacylase |
| KCH_13090 | 3-hydroxyacyl-ACP dehydratase<br>*S. lavendulae* (37 / 53) | 3-hydroxyacyl-ACP dehydratase |
| KCH_13100 | 3-hydroxyacyl-ACP dehydratase<br>*Stappia stellulata* (42 / 55) | 3-hydroxyacyl-ACP dehydratase |
| KCH_13110 | 3-oxoacyl-ACP reductase<br>*Streptacidiphilus melanogenes* (91/ 95) | 3-oxoacyl-ACP reductase |
| KCH_13120 | Haloacid dehalogenase<br>*Rothia dentocariosa* (33 / 53) | Haloacid dehalogenase |
| KCH_13130 | A-factor biosynthesis protein<br>*S. himastatinicus* (37 / 48) | AfsA homologue |
| KCH_13140 | 3-oxoacyl-ACP synthase<br>*S. katrae* (41 / 49) | 3-oxoacyl-ACP synthase |
| KCH_13150 | 3-oxoacyl-ACP synthase<br>*S. lavenduligriseus* (57 / 76) | 3-oxoacyl-ACP synthase |
| KCH_13160 | Polyketide-8 synthase ACP<br>*Catelliglobosispora koreensis* (34 / 60) | Polyketide-8 synthase ACP |
| KCH_13170 | Monoxygenase<br>*S. scabiei* (38 / 47) | Monooxygenase |
| KCH_13180 | 3-oxoacyl-ACP synthase<br>*S. toyocaensis* (36 / 53) | 3-oxoacyl-ACP synthase |
| KCH_13190 | MFS transporter<br>*S. virginiae* (46 / 61) | MFS transporter |
| KCH_13200 | Serine hydrolase<br>*Nocardia aobensis* (49 / 61) | Serine hydrolase |
| KCH_13210 | Thioesterase<br>*S. avermitilis* (44 / 60) | Thioesterase |
| KCH_13220 | SgnP<br>*S. venezuelae* (47/ 58) | MmfP homologue |
| KCH_13230 | MmfR<br>*S. coelicolor* (59 / 76) | MmfR homologue |

| | | |
|---|---|---|
| KCH_13240 | MmfL<br>*S. coelicolor* (44 / 57) | MmfL homologue |
| KCH_13250 | MmfH<br>*S. coelicolor* (53 / 61) | MmfH homologue |
| KCH_13260 | Oxidoreductase<br>*S. aureofaciens* (54 / 63) | Oxidoreductase |
| KCH_13270 | AraC family transcriptional regulator<br>*S. azureus* (78 / 85) | AraC family transcriptional regulator |
| KCH_13280 | Sacchraopine dehydrogenase<br>*S. venezuelae* (73 / 82) | Sacchraopine dehydrogenase |
| KCH_13290 | MmyR<br>*S. coelicolor* (39 / 47) | MmyR homologue |
| KCH_13300 | Patatin<br>*S. vietnamensis* (44 / 53) | Patatin |