Doctoral Dissertation

Discovery of cryptic compounds from *Streptomyces lavendulae* FRI-5 using an engineered microbial host

(異種発現系の活用による放線菌休眠化合物の覚醒と同定)

Ivy Grace Umadhay Pait

January 2018

Division of Advanced Science and Biotechnology

Graduate School of Engineering, Osaka University

Contents

Chapter 1
General Introduction

1.1 The role of microbial natural products in drug discovery

Broadly speaking, natural products are defined as "chemical (carbon) compounds isolated from diverse living things" (Bérdy, 2005). These compounds may either be primary metabolites that are common among living organisms (such as polysaccharides, proteins, nucleic and fatty acids), or low molecular weight (Molecular weight < 3000 Da) secondary metabolites that are species-specific and are unnecessary for growth and reproduction (Dias, Urban, & Roessner, 2012). Many microorganisms produce secondary metabolites in order to adapt to environmental changes or defend against threats to survival. Nonetheless, the natural role of these compounds are mostly unclarified (Maplestone, Stone, & Williams, 1992).

Microbial secondary metabolites have a long history of contributions to human medicine, animal health and crop protection. For instance, it was the fortuitous discovery of a simple mold's (*Penicillum rubens*) ability to produce a substance with antibacterial properties (Fleming, 2001) that triggered the age of antibiotics research. Thanks to Penicillin, many survived otherwise deadly bacterial infections during the war-torn era and before long, our dedicated pursuit of other bioactive natural products uncovered treatments for various infectious diseases like tuberculosis, meningitis and pneumonia (Omura, 2011). The term "antibiotics" was originally coined in reference to antibacterials of the early 1990's and was later on modified to include other antimicrobials (antifungals and antiprotozoals), antitumor agents and antiviral compounds (Berdy, 2012). Of the half a million natural compounds reported as of 2012 (Table 1.1), roughly 70,000 are microbial metabolites. Interestingly, nearly half of these reported microbial compounds (47%)

exhibited some form of antibiotic activity or "other" effects (like enzyme inhibition, agricultural effects, etc.), whereas this number is lower in plant (7%) and animal-derived (3%) compounds. Berdy reports that 51% (16,800/33,000) of all known classical "antibiotics" (referring to antimicrobials) actually show other clinically and industrially-relevant biological activities (Figure 1.1). For example, aside from having antitumor or antiviral activities, many antimicrobials can also be enzyme inhibitors and phytotoxic or pharmacologically active agents (Bérdy, 2012).

Table 1.1 Occurring bioactivities and drugs derived from natural products and their origin (Berdy, 2012)

	Natural Product	**%**	**Animal derived**	**%**	**Plant derived**	**%**	**Microbial derived**	**%**
Total	~ 500 000	100	~ 100 000	~20	~ 350 000	~70	~ 70 000	~10
Bioactives	~ 60 000	~12	~ 5000	~3	~ 25 000	~7	~ 33 000	47
Drugs	~ 1300	.03	100-150	0.001	600-1000	0.03	400-500	0.6

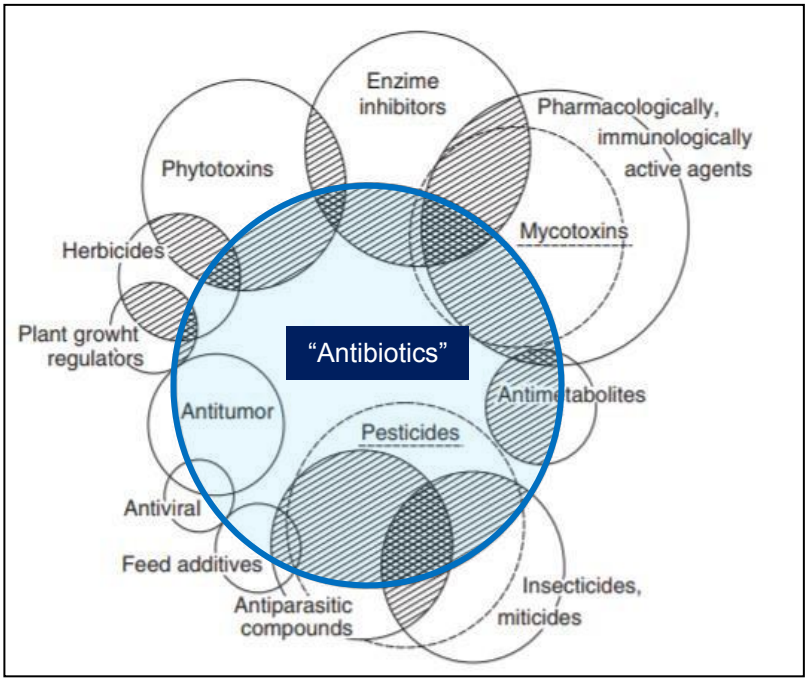


Figure 1.1. Bioactivities of microbial metabolites. 'Antibiotics' refer to its classical definition of antimicrobials (Bérdy, 2012)

This wealth of bioactive compounds from microorganisms yielded many notable therapeutics including the anti-fungal nystatin (Michel, 1977), the anthelmintic drug avermectin (Burg et al., 1979), the cholesterol lowering drug lovastatin (Tobert, 2003), and the immunosuppressant rapamycin (Li, Kim, & Blenis, 2014). It is without doubt that microbial natural products are indispensable resources for the development of therapeutics.

1.2 Bioactivities and biosynthesis of polyketides (PKs) and nonribosomal peptides (NRPs)

Natural products (secondary metabolites) are divided into several structural classes and this work focuses on 2 large groups with exceptional structural diversity and biological activities- the polyketides (PKs) and nonribosomal peptides (NRPs). Polyketides are derived from carboxylic acid building blocks such as simple malonate precursors (Staunton & Weissman, 2001) whereas nonribosomal peptides are small peptides mostly consisting of less than 20 amino acid residues but are highly diverse because of modifications such as acylation, glycosylation, epimerization, cyclization or methylation (Du, Sánchez, & Shen, 2001). PKs and NRPSs constitute a huge portion of commercialized drugs (Fig. 1.2) which includes compounds like the polyketide antibiotics erythromycin and tetracycline, the nonribosomal peptide antibiotics penicillin and vancomycin, as well as hybrid compounds (PK-NRP) rapamycin and the anti-tumor drug bleomycin (Fischbach, Walsh, Fischbach, & Walsh, 2006). The unique structures of thousands of PK, NRP and hybrid compounds arise from sequential actions of large multi-modular enzymes.

1.2.1 The assembly-line enzymology of polyketides

Polyketides are assembled in a way similar to fatty acid biosynthesis which involves simple precursors such as acetyl-coenzyme A and malonyl- units derived from the pool of primary metabolites (Smith & Tsai, 2007). In general, both polyketides and fatty acids are constructed by

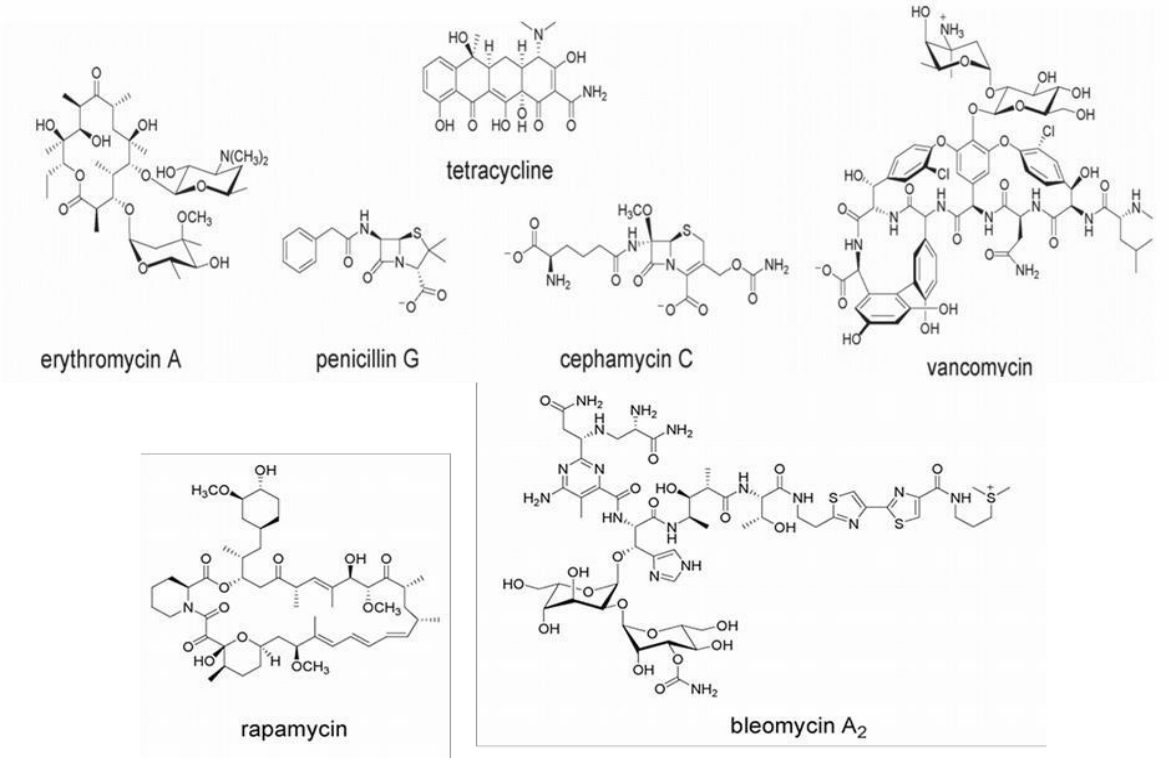


Figure 1.2. Therapeutically relevant PKs, NRPs, and PK-NRP hybrids (Fischbach et al., 2006)

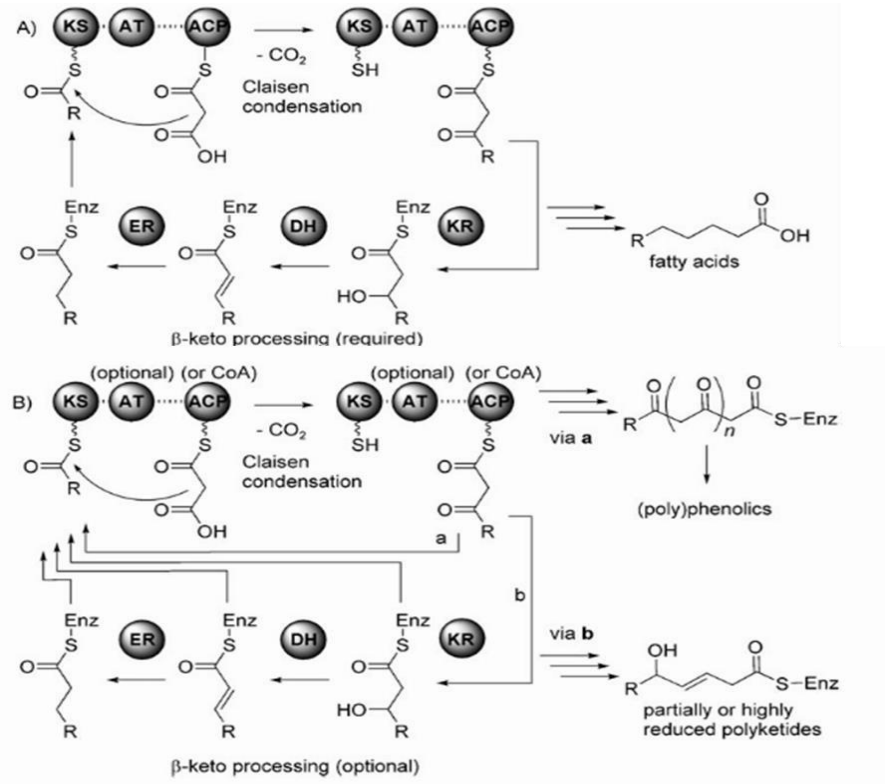


Figure 1.3. Biosynthesis of fatty acids (A) and polyketides (B) (Hertweck, 2009)

repetitive decarboxylative Claisen thioester condensations of an activated acyl starter unit with a malonyl-CoA-derived extender unit (Fig. 1.3) (Hertweck, 2009).

This process usually involves 3 catalytic functions: a ketosynthase (KS) responsible for the Claisen reaction, an optional (malonyl)acyl transferase (MAT/AT) for substrate recognition and activation, and a phosphopantethienylated acyl carrier protein (ACP), which serves as an anchor for the growing chain. For fatty acids, the ketone group undergoes stepwise reduction and dehydration processed by a ketoreductase (KR), dehydratase (DH), and an enoyl reductase (ER), to yield a fully saturated acyl backbone. In contrast, some of these steps can be skipped prior to the next round of polyketide elongation giving polyketides a diversity of functional groups, which can include a mixture of hydroxyls, carbonyls, double bonds and methylene groups. Moreover polyketides can be synthesized using a broader range of biosynthetic precursors (J Staunton & Weissman, 2001). The elongation/reduction cycles are repeated in both pathways until a fixed chain length is obtained, and the thioester-bound substrates are released.

These multiple enzymatic activities are catalyzed by large multifunctional enzymes called polyketide synthases (PKSs). Polyketide synthases can be categorized into three different types (I, II and III) on the basis of their quaternary structure and mechanisms (Hertweck, 2009) as shown in Figure 1.4. Type I PKSs are large enzymes with linearly arranged and covalently fused catalytic domains acting in a step-wise fashion. Type II PKSs involve a complex of several single module proteins with separated enzymatic activities, acting iteratively to produce aromatic polyketides. Lastly, Type III PKSs are characterized by a single active site enzyme that functions as a homodimer and acts repeatedly to form the final product. They do not include an acyl-carrier protein (ACP) domain and are typically associated to plants with a few examples found in microorganisms. Bacteria possess both non-iterative type I PKS (contains individual modules for

every elongation step) and iterative type II PKS. Meanwhile, PKSs in fungi are mainly iterative type I enzymes that have a modular organization, but each enzymatic domain acts repeatedly to elongate the polyketide backbone with varying degrees of reduction (Hertweck, 2009).

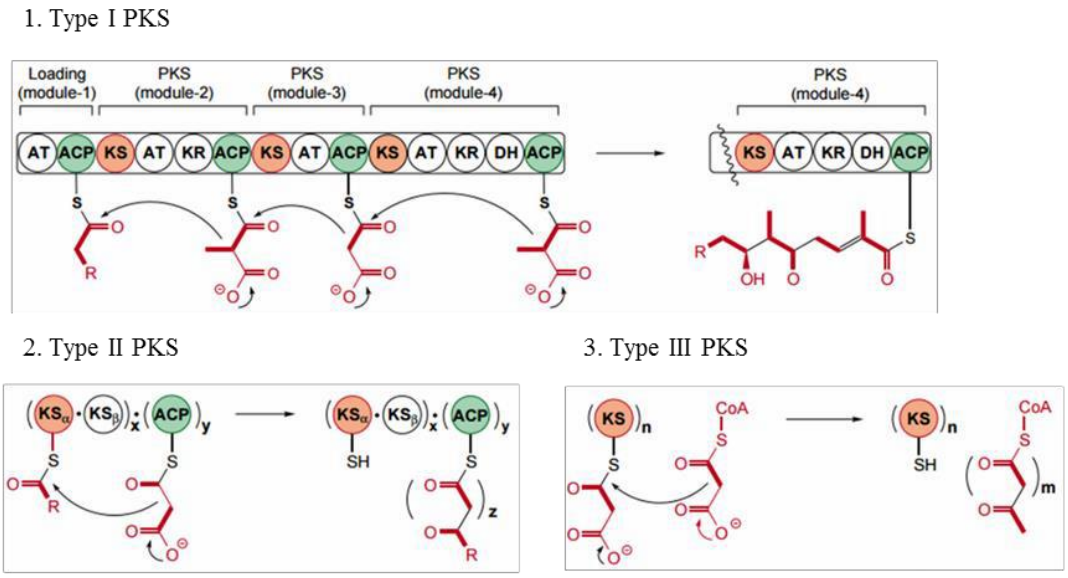


Figure 1.4. Classification of Polyketide Synthases (PKSs) (Shen, 2003)

Once the polyketide backbone is completed, the chain is released from the PKS by hydrolysis or cyclization catalyzed by the thioesterase domain (TE) in modular PKS or sometimes via reductive cleavage (Fischbach et al., 2006). A broad range of tailoring enzymes then decorate the nascent polyketide scaffold resulting into an impressive diversity of structures. These biotransformations may include, but are not limited to, glycosylations, alkylations, hydroxylations, epoxidations and halogenations (Hertweck, 2009). Our comprehensive understanding of PKS assembly has enabled the creation of new derivatives by combinatorial biosynthesis and alterations of the post-PKS tailoring enzymes (J Staunton & Weissman, 2001).

1.2.2 The biosynthetic logic of nonribosomal peptide compounds

Synthesized mainly by bacteria and fungi, nonribosomal peptides (NRPs) are small peptides derived from both proteinogenic and non-proteinogenic amino acids. It is reported that over 500 types of monomers have been incorporated into NRPs (Caradec et al., 2014). These ribosome-independent peptides are structurally diverse, with only 27% being linear and the rest having branched, cyclic or other complex primary structures (Calcott & Ackerley, 2014). Many NRPs have been identified with antibiotic, antiviral, anti-cancer, anti-inflammatory, immunosuppressant and surfactant properties (Sieber & Marahiel, 2005).

Nonribosomal peptides are synthesized by gigantic multi-enzyme complexes known as nonribosomal peptide synthases (NRPS). Just like type I PKSs, NRPS possesses a modular structure and each module is responsible for the addition and optional modification of one amino acid (L. Du et al., 2001). The order and the number of modules in an NRPS protein dictates the sequence and the number of amino acids in the resultant peptide product. Within each module, an adenylation domain (A) recognizes and activates a specific substrate and the activated substrate is subsequently tethered to the thiolation (T) domain sometimes called as the peptidyl carrier protein (PCP) domain (Figure 1.5). The condensation domain (C) then catalyzes the peptide bond formation between the peptidyl and amino acid thioesters to elongate the growing peptide chain (Fischbach et al., 2006; Strieker, Tanović, & Marahiel, 2010). Additionally, modules can also contain optional domains with the ability to modify the amino acid, such as a change in stereochemistry (epimerization domain), a methylation (methylation domain) or heterocyclization of Ser, Thr or Cys residues (cyclization domain). The final peptide is released from the NRPS through the actions of either a thioesterase domain, a variant of the C domain or a reductase domain (Lautru & Challis, 2004) followed by various post-NRPS modifications to generate diverse

structures. Similarly, genetic modifications of the A-domain substrate specificity have been explored to create novel peptide products.

Genes encoding PKSs and NRPSs are usually organized in clusters, perhaps to facilitate coordinated regulation when the need to produce a metabolite arises or to simplify horizontal gene transfer between microbial genomes. These enzymes are also situated next to genes that encode modification enzymes or pathways for unusual monomers, as well as genes encoding resistance mechanisms when the metabolite is toxic. The clustering of these genes into so-called biosynthetic gene clusters (BCGs) have facilitated cloning of complete pathways and exploration of similar machineries in various microorganisms (Fischbach et al., 2006).

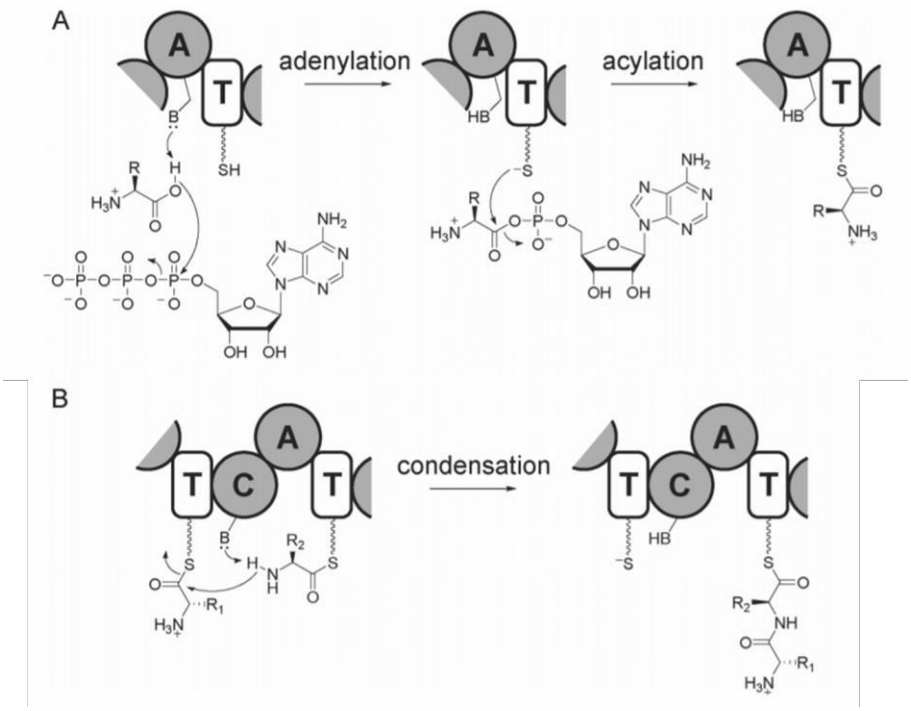


Figure 1.5. The minimum catalytic domains within an NRPS module (Fischbach et al., 2006)

1.3 *Streptomyces*, the proven resource for therapeutics

Actinomycetes are filamentous, high G+C gram positive bacteria that participate in the mineralization processes in nature. They were originally identified as soil-dwelling microorganisms but extensive exploration of other habitats in search of novel isolates and natural products revealed that actinomycetes also inhabit marine environments (Fenical and Jensen 2006; Subramani and Aalbersberg 2012), plants (Golinska et al. 2015), caves (Maciejewska et al. 2016) and even invertebrates (Genilloud et al. 2011). Many members of this class have been explored since the 1940's primarily because of their ability to produce pharmaceutically and agriculturally relevant compounds (Fig. 1.6). Especially in the 50s to the 70s (Golden Age of Antibiotics), 60% of new antibiotics were isolated from actinobacteria, almost exclusively from the genus *Streptomyces* (Bérdy, 2012). In fact, two thirds of all commercially available antibiotics are derived from streptomycetes, including the antihelmintic drug Avermectin whose discovery has been recently recognized with a Nobel Prize (Ikeda, Nonomiya, Usami, Ohta, & Omura, 1999). The ability of streptomycetes to produce anti-infectives, anti-hypertensives, immunosuppressants and other bioactive compounds has made them the most important member of the actinomycetes family.

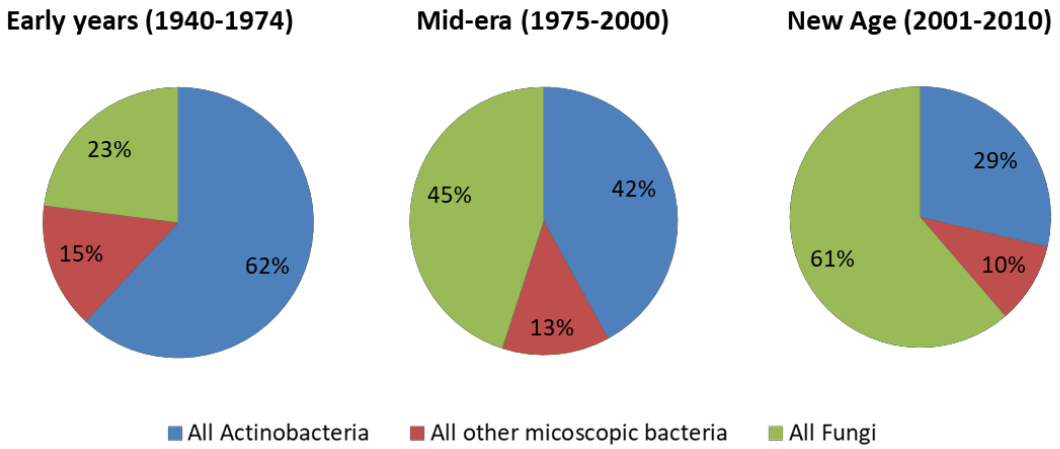


Figure 1.6. Approximate distribution of bioactive secondary metabolites in periods from 1940 to 2010 according to their producers (Berdy 2012).

1.3.1 Life cycle of *Streptomyces*

Within the polymicrobial communities occupying the soil, *Streptomyces* represent the largest genus of the ubiquitous actinomycetes group (Jones et al., 2017). Their lifestyle is strikingly similar to that of filamentous fungi, both sharing three developmental stages as shown in Figure 1.7. Under favorable conditions, unigenomic spore germinates and generates one or two germ tubes that eventually extend at the tips. The continuous branching of these hyphal filaments results in the formation of a substrate mycelium that greatly facilitates soil colonization (Kieser, Bibb, Buttner, Chater, & Hopwood, 2000). As *Streptomyces* continue to grow and local nutrients are exhausted, a complex regulatory cascade then initiates the production of a surfactant that coats some emerging filaments. This hydrophobic coat allows the aerial filaments to break surface tension and grow into the air (Flärdh & Buttner, 2009). Finally, some of these aerial filaments divides synchronously at many sites to form uninucleoid cells that again further develop into mature spores. Sporulation assists streptomycetes in spreading as wells as in their persistence especially in conditions of low nutrient and water availability (Kieser et al., 2000). Interestingly, aerial mycelium and sporulation in *Streptomyces* have been assumed to arise as a result of programmed cell death (PCD) in which the substrate mycelia is sacrificed to release nutrients for aerial mycelium and spore formation (Manteca & Sanchez, 2010). At this stage of morphological development, *Streptomyces* produces many antibiotics and other secondary metabolites as defense mechanism against the scavenging activity of other microorganisms. While most secondary metabolites are generally believed to provide the producer a form of competitive advantage against competitors, the exact role of many soil secondary metabolites is unknown. For the purpose of isolating industrially important compounds, most commercial fermentations are performed in

liquid cultures, where target secondary metabolites are produced by substrate mycelium at the end of the growth phase (Yagüe, Lopez-Garcia, Rioseras, Sanchez, & Manteca, 2012).

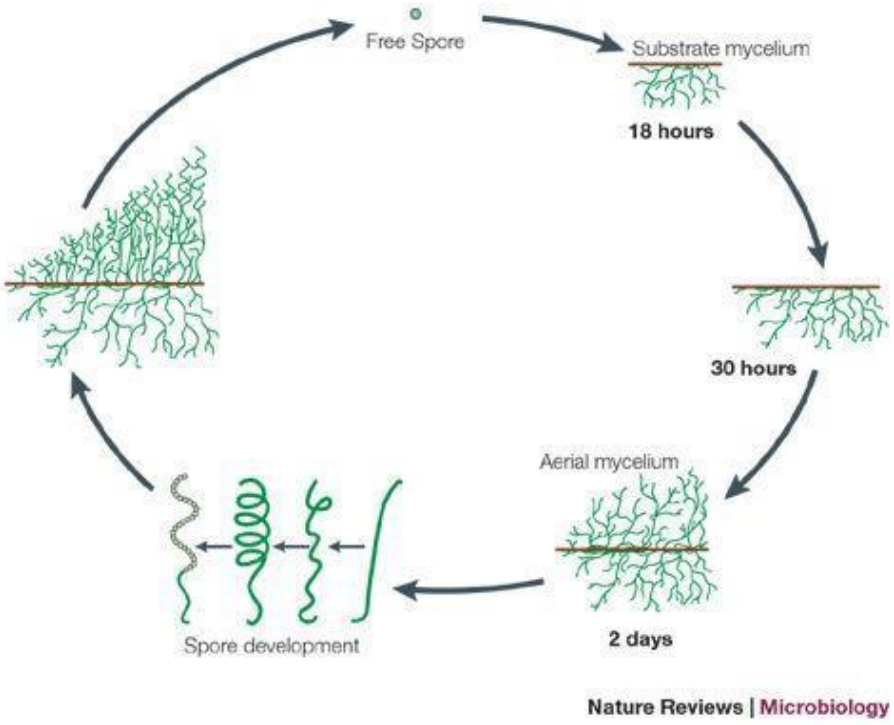


Figure 1.7. Developmental cycle of the model strain *Streptomyces coelicolor* confluent cultures on agar surface (Angert, 2005).

1.3.2 Regulation of secondary metabolite production in *Streptomyces*

Morphological development and secondary metabolism in S*treptomyces* arise from a well-concerted series of regulatory events starting from the detection of environmental stimuli down to the activation of biosynthetic genes encoding bioactive metabolites. Since it is metabolically costly for a bacteria to produce all of its arsenal of metabolites at the same time, the timing and the quantity of secondary metabolites produced is tightly regulated in a hierarchical manner. High-level or global regulators respond to various environmental stress signals (e.g. phosphate or nitrogen starvation) and activate or deactivate the expression of low-level regulators, which are

situated within the biosynthetic gene clusters. This cluster-situated regulators, sometimes known as pathway-specific regulators directly control the transcription and repression of biosynthetic genes for secondary metabolites (van Wezel & McDowall, 2011). Well studied pathway-specific regulators include members of the *Streptomyces* antibiotic regulatory protein (SARP) family characterized by an N-terminal DNA binding domain and a bacterial transcriptional activator domain (BTAD)(Martín & Liras, 2010). SARPs exert positive functions on biosynthetic genes and the deletion of SARP genes resulted in the loss of secondary metabolite production (Arias, Fernández-Moreno, & Malpartida, 1999; Retzlaff & Distler, 1995). In addition, cluster-situated large regulators of the LAL-family, a sub-family of LuxR family of regulators, were also implicated in the activation of polyketide biosynthesis (Kuščer et al., 2007; Wilson, Xue, Reynolds, & Sherman, 2001). In contrast, genes that encode global regulators may either be in the vicinity or located outside of the biosynthetic gene clusters and they exert pleiotropic effects on the production of secondary metabolites and/or morphological differentiation.

In *Streptomyces*, the most extensively studied pyramidal control of secondary metabolism is participated by high-level regulators belonging to the group of γ-butyrolactone autoregulators and their cognate receptor proteins (Takano, 2006). In this system, autoregulators present at nanomolar concentrations elicit antibiotic production by binding to and modulating the DNA-binding activity of its cognate receptor proteins. In the absence of autoregulators, its cognate receptor functions as a transcriptional repressor of target regulatory genes that influence secondary metabolism. Once the production of the autoregulator is initiated and the concentration reaches a threshold level, autoregulators diffuse back into the cell and interact with autoregulator receptor proteins (Bibb, 2005; Takano, 2006). As illustrated in Figure 1.8, the binding of the autoregulator to its specific receptor dissociates the complex from DNA and the protein product, in a series of

steps, activates secondary metabolite production and sometimes morphological differentiation. The γ-butyrolactone autoregulator signalling system have been identified in numerous streptomycetes and other genera of actinomycetes (Choi, Lee, Hwang, Kinosita, & Nihira, 2003; Takano, 2006) with the A-factor system of *Streptomyces griseus* being the most well characterized. A-factor activates streptomycin and grixazone production, simultaneously with sporulation. The binding of A-factor to its receptor ArpA releases the latter from the promoter region of a pleiotropic regulator *adpA*. AdpA is essential for the transcriptional activation of many genes including *strR* (pathway-specific regulator for streptomycin) and some genes for aerial mycelium formation (Bibb, 2005; Ohnishi, Kameyama, Onaka, & Horinouchi, 1999). Although most environment signals that control the onset secondary metabolism remains unclear, studies regarding the complex regulation of secondary metabolites continue every year in an attempt to increase the production of valuable fermentation products or stimulate the production of new compounds.

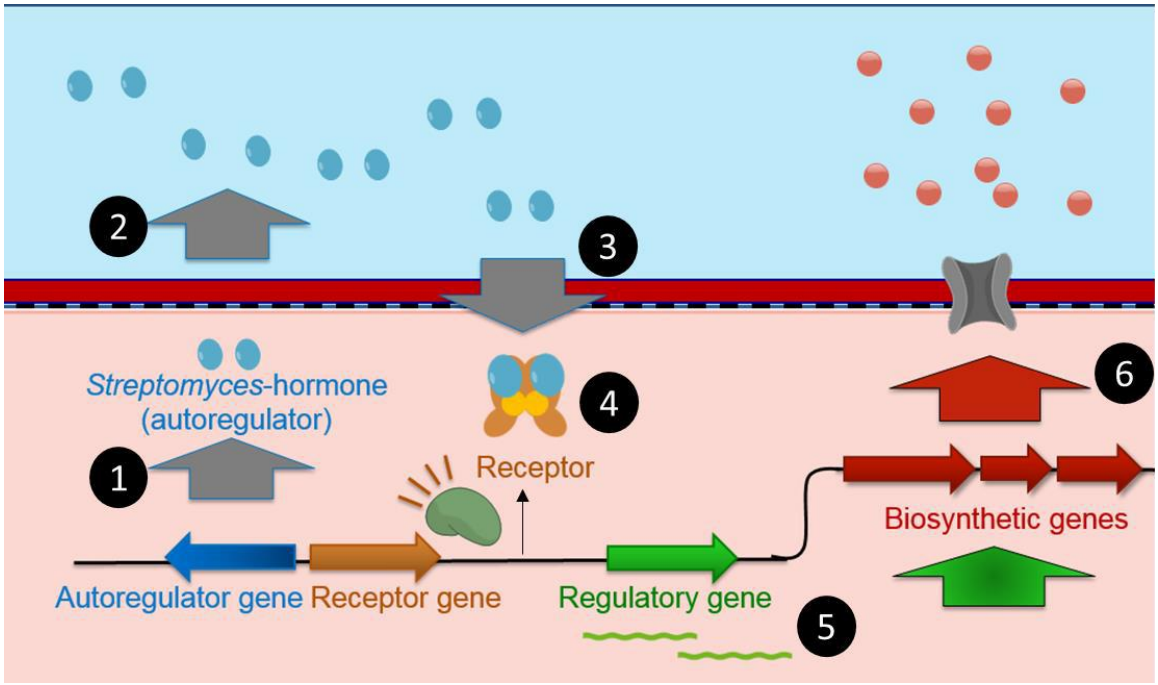


Figure 1.8. The molecular mechanism of an autoregulator signaling cascade

1.4 Genome mining and reviving interest in microbial secondary metabolites

1.4.1 Decline in natural product research

Despite massive contributions to human health and agriculture for over 70 years, research on natural product discovery has declined drastically. In fact, many pharmaceutical companies have abandoned their natural product discovery programs primarily in favor of high-throughput screening of synthetic chemical libraries (J. W.-H. Li & Vederas, 2009). Traditional screening methods to discover microbial natural products usually involve collection and cultivation of strains in various media, compound extraction and bioassay-guided target identification. However, aside from being time-consuming and laborious, only a limited number of samples can be prepared making the entire process incompatible with high-throughput liquid handling-based screening methods that have gained popularity. Chemical libraries on the other hand are already "semi-pure", and are easier to generate and modify making it possible to examine large number of molecules in much shorter timelines.

The roadblock faced by natural product screening is worsened by the high re-discovery rate of already known compounds. During the golden age of antibiotic discovery (1940 to 1980), new bioactive products were discovered at a remarkable rate of roughly 200–300 per year in the late 1970s. However, since the 1980s, it became more difficult to discover new and useful compounds (Demain, 2014). There has been a long "Discovery Void" in the field of antibiotics (Fig. 1.9), wherein no new chemical structures were discovered and majority of new compounds were just variations of existing backbones. Despite this problem, we are in constant need of new classes of antibiotics binding to uncommon targets in order to cope with resistant pathogens, the emergence of new diseases since the 1980s (AIDS, Ebola virus, Hanta virus, Lyme disease, etc.) and the toxicity of some of the current treatments (Bérdy, 2012; Demain, 2014). The reliance of many pharmaceutical companies on combinatorial chemistry alone also failed to generate a variety

of pharmacologically active molecules. To date, only two *de novo* new chemical entities reported in the public domain were approved for drug use, one antitumor drug approved by the FDA in 2005 and a drug for genetic disorder approved in the EU in 2014 (Newman & Cragg, 2016). As a result, the numbers of FDA drug applications, new drug approvals and new chemical entities from the pharmaceutical industry markedly decreased since the late 1990s.

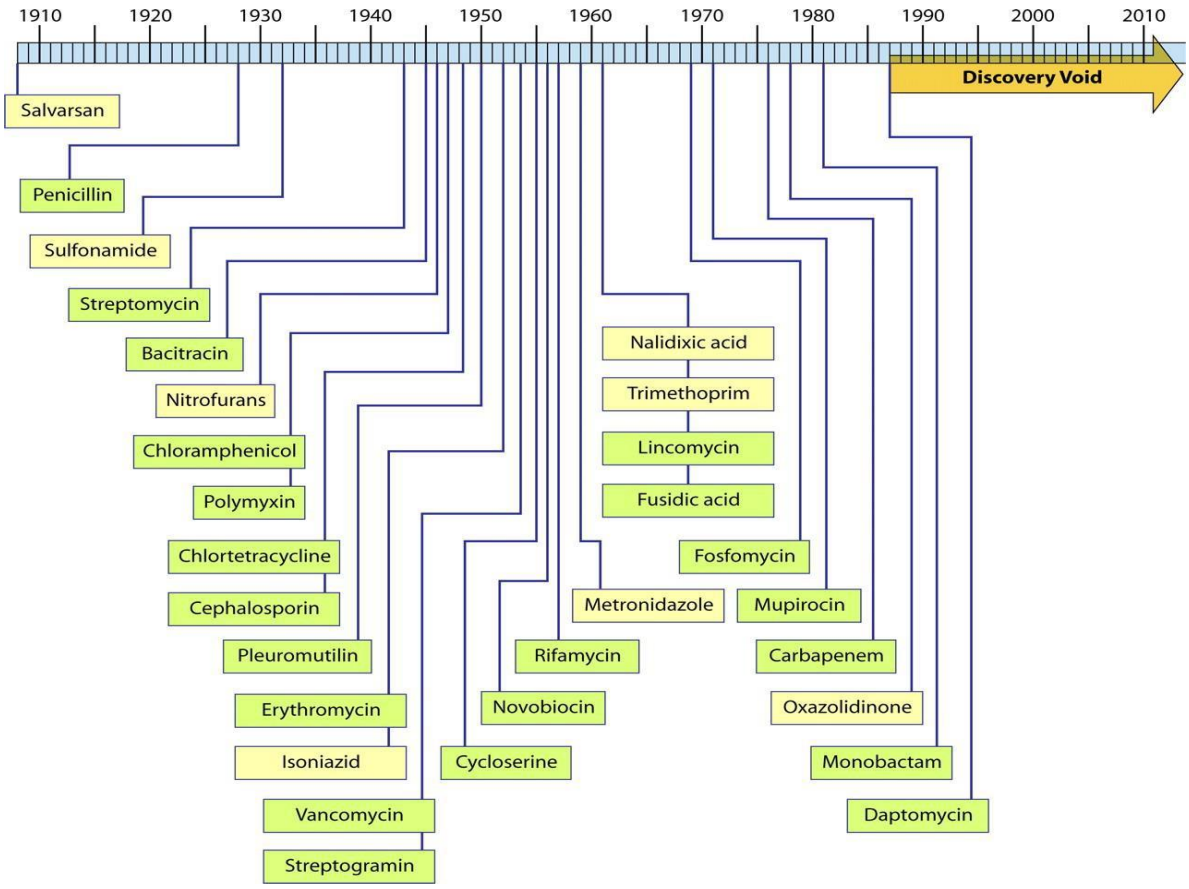


Figure 1.9. Illustration of the "discovery void" wherein no chemical entities have been discovered. Dates indicated are those of reported initial discovery or patent (Silver, 2011)

Undoubtedly, we must not abandon the research on natural products including those from microbial resources because the structural diversity of chemical libraries is far inferior to the novelty, complexity, remarkable diversity of structures and biological activities of natural products. In fact, studies have shown that there is a higher hit rate in high-throughput screening of natural

product collections than of combinatorial libraries of synthetic compounds (Demain, 2014). Instead, natural product discovery must be pursued and made compatible with high-throughput screening and combinatorial chemistry. But most importantly, we must depart from the traditional screening and grinding that have led to frustrating re-discoveries of known compounds in the recent years.

1.4.2 Bacterial genome mining for new natural products

With time, our knowledge on PKS and NRPS genetics and the biosynthetic machineries for other kinds of compounds has deepened significantly and we have come to know that despite their structural diversity, the core enzymology used to synthesize these bioactive compounds are highly conserved (Fischbach & Walsh, 2006). Using partial genetic information of a PKS or NRPS gene from one strain, it became possible to find microorganisms that are able to produce compounds of the same class. However, as we moved into the 21st century, rapid developments in DNA sequencing technologies have enabled multiple genomes to be sequenced rapidly and inexpensively, revealing that the biosynthetic potential of microbes is surprisingly underexplored.

As explained in the earlier sections, the biosynthetic pathways for secondary metabolite production (for instance NRPS and PKS pathways) often resides within discrete, localized sections of the microbial genome, termed as "biosynthetic gene clusters (BGCs)". Nowadays, thousands of completed bacterial genomes are available in public databases and many bioinformatics tools designed to detect and predict the products of BGCs such as antiSMASH, CLUSTSCAN, and NP Searcher have made genome analysis easier for researchers. It was revealed that many bacterial genomes contain numerous (sometimes 20–25) independent BGCs that may direct the biosynthesis of different types of compounds (Bachmann, Van Lanen, & Baltz, 2014; Scherlach & Hertweck,

2009). For instance, completed and annotated genomes of over 65 *streptomycetes* and hundreds of bacterial species demonstrated that more than half of them harbors genes for polyketide and non-ribosomal peptide biosynthesis (Zarins-Tutt et al., 2016). However despite this immense biosynthetic potential, only a small fraction (<10%) of BCGs are expressed under laboratory fermentation conditions leading to the isolation of only 1 or 2 major secondary metabolites. Although some of these biosynthetic genes may actually be expressed, the production of minor compounds may be overlooked due to low yields, a high metabolic background or unsuitable culture conditions (Scherlach & Hertweck, 2009; Zarins-Tutt et al., 2016). Thus, these "silent" gene loci with unknown products have been commonly referred to as "cryptic" or "orphan" biosynthetic gene clusters and the exact trigger for compound production remains a mystery. It is generally believed that the systematic activation of these cryptic pathways (an emerging field termed as "genome mining") is the long awaited revolutionary approach towards the exploitation of yet unknown secondary metabolites. Genome mining can greatly minimize our reliance on chance and circumvent the historical problem of rediscovery by focusing our resources towards the isolation and identification of novel natural products from uncharacterized gene clusters.

1.4.3 *Streptomyces* genome as the richest bacterial resource for cryptic BGCs

The notion that bacteria has many cryptic gene clusters started when the genome of the model strain *Streptomyces coelicolor* A3(2) was opened by Sanger sequencing in 2002 (Bentley et al., 2002). *S. coelicolor* A3(2) was thought to have been mined to exhaustion with only 4 secondary metabolites. However, the complete sequence and initial annotation of its 8.6 Mbp-genome allowed the identification of another 17 orphan pathways. Structure prediction of the product associated with one tri-modular NRPS initially suggested a tripeptide iron siderophore and the subsequent cultivation in an iron-depleted media yielded the new iron chelator coelichelin (G L Challis & Ravel, 2000). This

became the first successful genome-based attempt towards novel natural product discovery. Soon after, the genome sequences *Streptomyces avermitilis* MA-4680 and *Streptomyces griseus* IFO 13350 were also revealed to have more than 35 BGCs for secondary metabolites (Nett, Ikeda, & Moore, 2009). Prior to genome sequencing, the structures of only 2 polyketides were established from *S. avermitilis*, namely avermectin and oligomycin. From its genome sequence a further 11 PKS gene clusters and 8 NPRS clusters were deduced and so far all the NRPS BGCs were presumed as "cryptic". On the average, *Streptomyces* genomes harbor 20-40 secondary metabolites (Table 1.2) that can potentially be explored for new compounds.

Inexpensive DNA sequencing have made it possible to explore the biosynthetic capacity of other groups of microorganisms. Clusters from metagenomes or unculturable microorganisms were also considered as "unexploited resources" for novel metabolites (Zarins-Tutt et al., 2016). However using the abundance of PKS and NRPS genes as a basis for quantifying biosynthetic potential (given that 60% of important secondary metabolites are related to these classes), clearly large genomes of streptomycetes and other actinomycetes (average length of 8.15 Mb) greatly outnumbered that of non-actinomycete bacteria in terms of the number of BGCs. On the other hand, genomes from unculturable microorganisms are relatively smaller, ranging from 0.60 to 1.57 Mb, with very minimal clusters for secondary metabolites (Baltz, 2016). Just like in the past, genome information points us towards the continued exploration of *Streptomyces* species for the discovery of new chemical entities. Within *Streptomyces* genomes is a pool of biosynthetic resources for new medicines that can potentially rejuvenate the drying drug discovery pipeline and the ongoing challenge lies on how we can stimulate the expression of these genes clusters to obtain our desired compounds.

Table 1.2 Secondary metabolite (SM) encoding capacity of actinomycetes (Baltz, 2016).

Microorganism	Genome Size (Mb)	PKS/NRPS clusters	Total BGCs	% of genome dedicated for SM
Saccharomonospora viridis P101	4.31	4	11	9.7
Salinispora tropica CNB-440	5.18	13	17	10.0
Streptomyces albus J1074	6.84	11	24	14.9
Micromonospora aurantiaca ATCC 7029	7.03	10	17	14.1
Saccharopolyspora erythraea NRRL 338	8.21	16	36	18.6
Streptomyces venezuelae	8.23	9	31	12.2
Streptomyces griseus NBRC 13350	8.55	17	40	20.0
Streptomyces coelicolor A3(2)	8.67	11	27	10.7
Streptomyces avermitilis MA-4680	9.03	17	37	16.7
Streptomyces bingchenggensis BCW-1	11.94	30	53	21.7

1.4.4 Approaches to triggering the production of cryptic metabolites

Several strategies have been employed recently to induce the expression of silent or cryptic gene clusters in bacterial genomes, allowing the discovery of novel compounds. These strategies can be classified into 2 major approaches- (1) Altering of chemical and physical conditions and (2) Genetic/Molecular modification (Zarins-Tutt et al., 2016). Each methodology has its own advantages and limitations, and the choice of a suitable method or combination of methods can be largely dependent on the available genomic information and the amenability of the organism towards genetic manipulation.

1.4.4.1 Altering chemical and physical conditions

(A) OSMAC - It has been known for a long time that even the smallest change in cultivation parameters is critical to the type and the yield of microbial secondary metabolites. Optimization of antibiotic yields and other important fermentation products were popularly carried

out by varying media components and cultivation conditions. To encourage the production of cryptic compounds in one organism, Bode and co-workers developed the OSMAC (one strain/many compounds) approach, which employs the systematic change of readily accessible cultivation parameters such as medium components, pH, aeration and temperature. OSMAC led to the isolation of over 20 different metabolites from streptomycetes (Bode, Bethe, Höfs, & Zeeck, 2002). Similarly in another study, initial cultivation of the terrestrial *Streptomyces* strain C34 in ISP2 medium resulted in the isolation of two novel ansamycin compounds. When the carbon source in ISP2 was replaced with glycerol, two other novel compounds were isolated (Rateb et al., 2011). OSMAC is not a targeted approach because it does not focus on a pre-determined silent BGC. It relies on the global effect on metabolism that results from chemical/physical changes within the culturing system. This makes OSMAC very accessible, inexpensive and relatively simple to implement without complete genome information.

(B) Challenging microorganisms with external cues – Chemicals elicitors were also added to the cultivation media to stimulate secondary metabolism. Antibiotics, for instance, have now been recognized as signals aside from being a line of defense. In a recent study, sub lethal concentrations of the antibiotic trimethoprim were shown to induce the expression of at least 5 biosynthetic pathways in *Burkholderia thailandensis* E264 (Seyedsayamdost, 2014). In addition, other chemical elicitors also activated many silent gene clusters associated with novel compounds such as lunalides A and B, cladochromes F and G and pestalone (Pettit, 2011).

(C) Combined cultures with other microorganisms – Microorganisms live in complex and dynamic communities and they adapt to the presence of competitive species by producing specialized compounds. Moreover, studies showed that interspecies crosstalk between organisms occupying the same environment can affect the activation or suppression of major cellular

metabolites. Thus it was recognized that the combined cultures with other microorganisms to try to mimic natural interactions can also be explored to discover cryptic metabolites. As an example, when *S. coelicolor* was grown with 5 different actinomycetes, a range of compounds with unknown identities were detected, including those that were not detected in pure cultures (Traxler, Watrous, Alexandrov, Dorrestein, & Kolter, 2013). Bacterial co-culturing of *Streptomyces endus* S-522 with the mycolic acid producer, *Tsukamurella pulmonis* also resulted in the discovery of a novel metabolite (Onaka et al., 2015). Microbial co-culture is another successful pleiotropic approach for genome mining although this approach is limited to cultivable microorganisms. Moreover, establishing cultivation conditions where both microorganism can grow well and interact is another important consideration.

1.4.4.1 Genetic modification/Molecular approaches

(A) Engineering the transcription and translation machinery –The successful expression of a gene/set of genes within a cell is also dependent on the cell's transcriptional/translational capacity. Perturbations in these universal systems is expected to have pleiotropic effects on various metabolic pathways. In the ribosome engineering approach, strains are grown in high concentrations of ribosome-targeting antibiotics such as streptomycin or gentamicin. Spontaneous antibiotic-resistant mutants carrying an altered RpsL protein (a critical component of the 30S ribosomal subunit) and other ribosomal proteins are subsequently screened and profiled in various media to detect changes in metabolite profile. This concept emerged when *S. lividans* mutants carrying *rpsL* mutations overproduced certain metabolites including actinorhodin which is normally silent in *S. lividans* (Shima, Hesketh, Okamoto, Kawamoto, & Ochi, 1996). Screening for drug-resistant mutants prior to bioassay among several actinomycetes allowed the discovery of many antibacterial compounds (Hosaka et al., 2009). Similarly, induced

mutations of *rpoB* encoding the RNA polymerase (RNAP) β-subunit after selection with the RNAP-targeting antibiotic rifampicin also unlocked the production of a cytotoxic compound in *Streptomyces somaliensis* (Zarins-Tutt et al., 2016).

(B) Manipulation of global and pathway-specific regulators – Bacterial secondary metabolism is controlled in a pyramidal manner by the combined efforts of global regulators or cluster-situated activators or repressors. Manipulation of global regulators that have the ability to switch on or switch off numerous metabolic pathways have led to the unexpected activation of cryptic compounds (Kalan et al., 2013; Zarins-Tutt et al., 2016). Likewise, the over-expression of activators or the silencing of repressors are straightforward ways to activate a target gene cluster. For example, the constitutive expression of cluster-situated activators of the LAL and SARP families resulted in the discovery of novel compounds in streptomycetes (D. Du et al., 2016; Laureti et al., 2011). In contrast, deletion of cluster-situated repressors of the TetR family freed silent biosynthetic gene clusters of *Streptomyces* sp. PGA64 and *Streptomyces ambofaciens* from their locked state, resulting in the detectible production of corresponding natural products (Bunet et al., 2011; Metsä-Ketelä, Ylihonko, & Mäntsälä, 2004). This approach requires the native strain to be genetically tractable and reasonable regulatory targets to be identified. So far, only a few pleiotropic transcriptional regulators are known in most microorganisms so the scope is limited.

(C) Heterologous expression of pathways- Transferring genes into another host has been a routine method to confirm whether a set of genes is required for the biosynthesis of a certain compound, or to improve yields of target compounds that are unstably produced by the original strain. Moreover, heterologous expression was also employed to obtain unnatural metabolites by combining genes from different pathways (Donsbach & Rück-Braun, 2008). During the early periods of genome mining, the heterologous expression of cryptic genes or pathways in commonly

used *Streptomyces* or *E. coli* wild type hosts sometimes in combination with promoter switching have led to the awakening of a few silent compounds (Corre, Song, O'Rourke, Chater, & Challis, 2008; Lin, Hopson, & Cane, 2006; Palmu, Ishida, Mäntsälä, Hertweck, & Metsä-Ketelä, 2007). The advantage of heterologous expression strategy is that it allows us to identify silent bioactive compounds when genetic manipulation in the original strain is difficult or when plausible pathway-specific regulators are absent within the silent gene cluster. We can also have a certain degree of control over our target metabolite or metabolite class given that the BGC information is available. Moreover it allows us to access the potential of metagenomes and uncultivable microorganisms. Targeting large clusters (>40 kb) may sometimes be difficult, however reconstitution of large pathways into a single vector by cosmid "stitching" is also possible (Zettler et al., 2014).

1.5. Heterologous expression in a genome-minimized *Streptomyces avermitilis* host

Secondary metabolite production is dependent on various cellular processes that correspond to the genetic makeup and biochemistry of the producing microorganism. For this reason, it is generally assumed that the use of a host species closely related to the source of the genes will likely prove optimal for heterologous expression. For instance, expression of exogenous NRPS and PKS gene clusters in *Streptomyces* hosts is favorable because they have the necessary primary metabolism to supply precursors for these compound classes (Baltz, 2010). At the same time, enzyme priming, regulation and resistance mechanisms encoded within the BGCs will likely be more compatible with related microorganisms.

Streptomyces avermitilis is the industrial strain for the anti-parasitic drug avermectin, thus it is a highly robust producer of secondary metabolites. Because this strain is already optimized for the efficient supply of primary metabolic precursors and biochemical energy to support multi-step biosynthesis, our collaborators developed *Streptomyces avermitilis* into a model host for

natural products by controlled genome minimization (Komatsu, Uchiyama, Omura, Cane, & Ikeda, 2010). After aligning the genome of *S. avermitilis* with 2 other model strains *S. coelicolor* A3(2) and *S. griseus*, a conserved core region (~6.28 to 6.50 Mb) containing genes essential for growth was identified. Meanwhile, the subtelomeric regions at the left and right chromosomal ends of *S. avermitilis* were asymmetric in structure (2 Mb and 0.5 Mb in the left and right ends, respectively) whereas that of the 2 strains were symmetric (~1 Mb each side). These subtelomeric regions contain strain-specific genes and several BGCs but no essential genes. Thus, the deletion of the large left subtelomeric region was expected to have no negative impact on either cell growth or primary metabolism-derived precursors for secondary metabolism. A region of more than 1.4 Mb (from its 9.02 Mb-genome) in the left subtelomeric region was removed from the linear chromosome by Cre-*loxP*-mediated site-specific recombination, generating SUKA3 strain (Fig. 1.10). This region includes BGCs for its major secondary metabolites, the avermectins (*ave* cluster) and filipins (*pte* cluster).

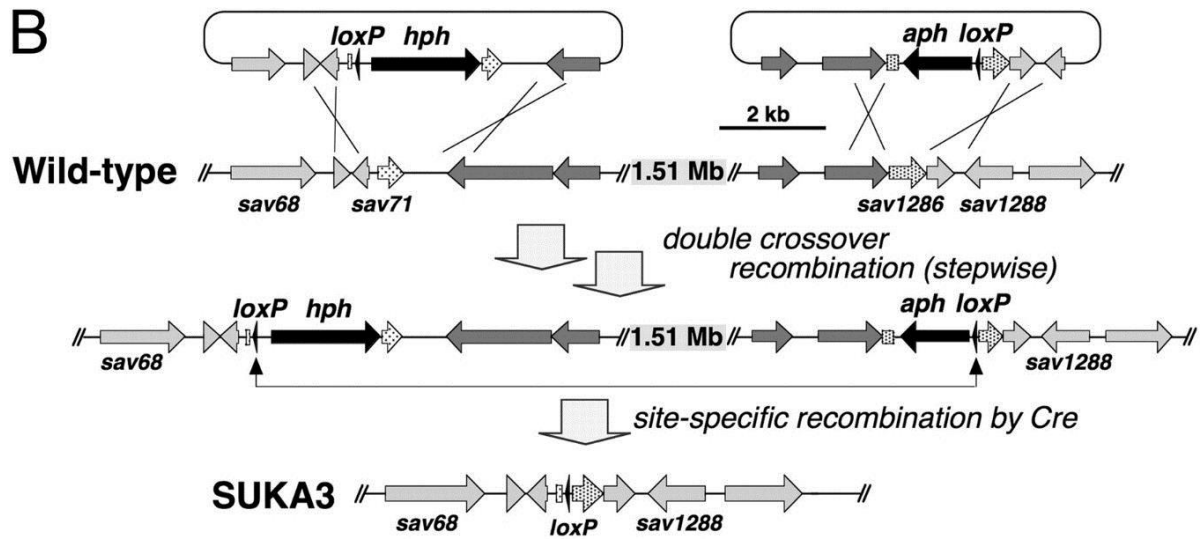


Figure 1.10. The strategy for construction of the large-deletion mutants of *S. avermitilis*. Detailed procedures are described by Komatsu et al., 2010

Continued stepwise deletions by Cre-*loxP* system were employed to remove genes encoding the biosynthesis of oligomycin (*olm* cluster) and 3 endogenous terpene compounds to generate genome-minimized *S. avermitilis* SUKA22 strain that did not produce any of its major endogenous secondary metabolites (Fig. 1.11). This genome-wide deletion provides a pool of precursors that can be channeled to the biosynthesis of exogenous BGCs. Subsequently, a cleaner metabolic background allows easier detection and purification of resulting compounds. Genome-minimized *S. avermitilis* strains, collectively called as SUKA strains, are more genetically stable as a result of deletions of transposons and terminal inverted repeats. Genetic instability is often associated with the loss of secondary metabolite production, which has become an industrial concern (Komatsu et al., 2013, 2010).

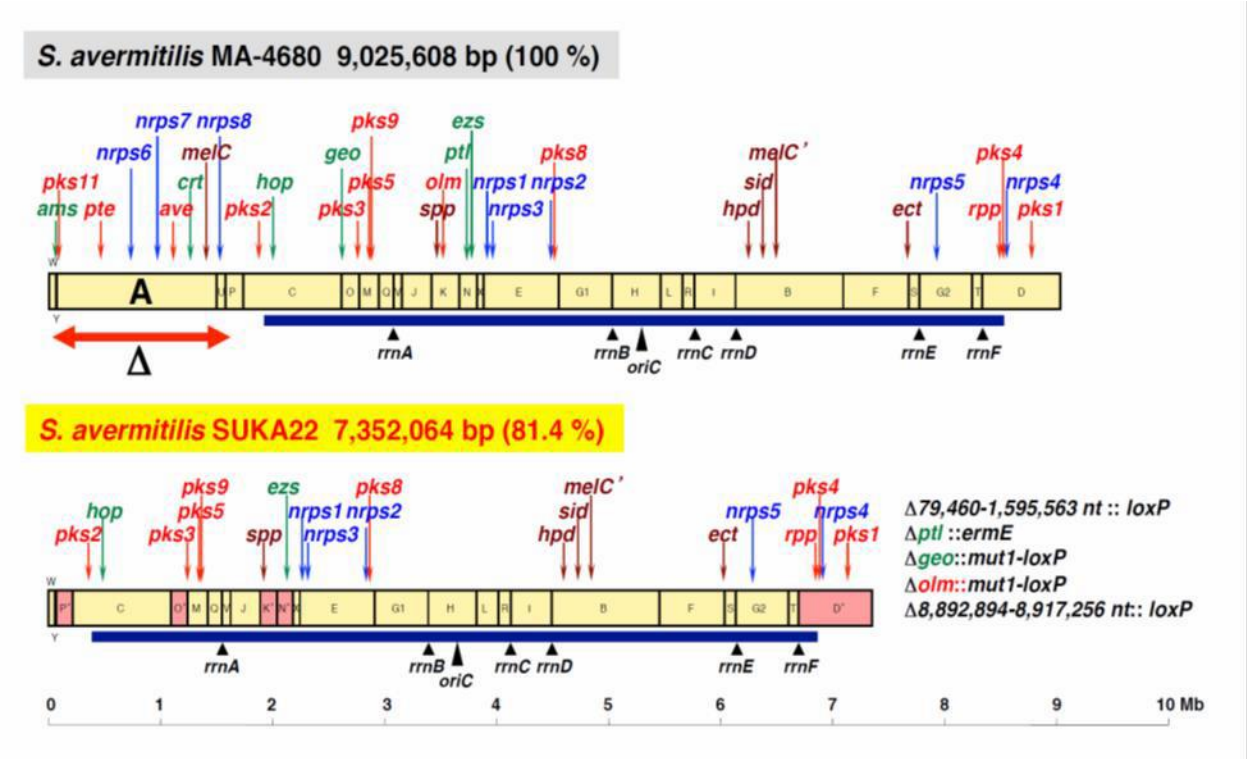


Figure 1.11. Physical map and distribution of gene clusters for BGCs on *S. avermitilis* and its large-deletion mutant SUKA22.

SUKA22 is capable of expressing different pathways including PKS BGCs and NRPS BGCs, with remarkable productivities sometimes surpassing that of the original producer and *S. avermitlis* wild type strain (Fig. 1.12). Moreover, the production of a known compound pholipomycin was activated upon heterologous expression of its corresponding gene cluster in *S. avermitlis* SUKA22 (Komatsu et al., 2013). So far, only active biosynthetic gene clusters or clusters that have been proven to produce a certain metabolite were expressed successfully in SUKA22. However, we believe that the engineered *S. avermitilis* SUKA22 strain might also have the potential to awaken uncharacterized "cryptic" biosynthetic pathways for interesting products and improve compound titers necessary for detection and isolation.

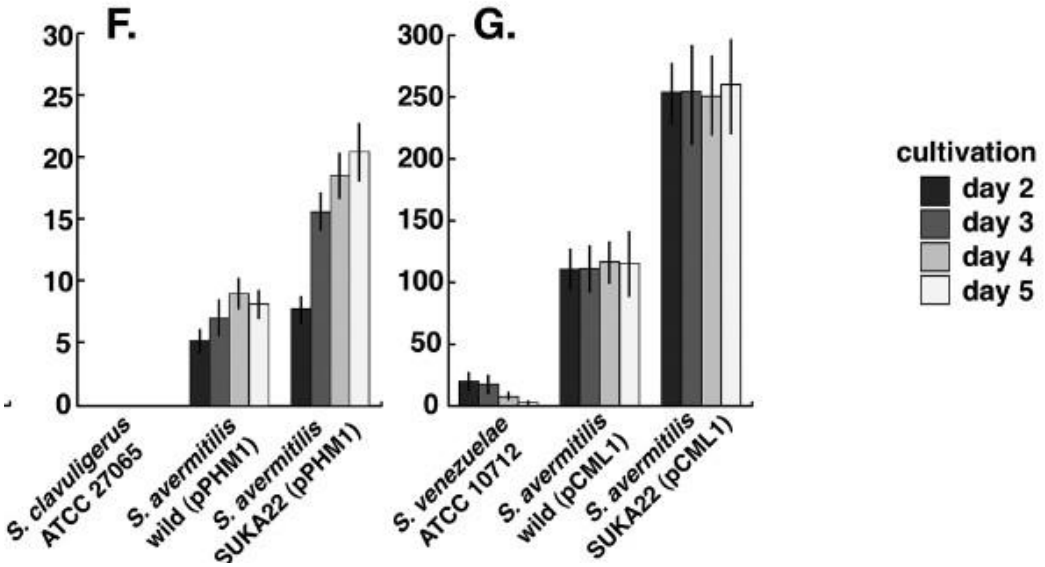


Figure 1.12. Production profiles of (F) pholipomycin and (G) chloramphenicol in original producers, *S. avermitilis* wild-type and SUKA22 carrying the biosynthetic gene cluster (Komatsu et al., 2013)

1.5 Natural products from *Streptomyces lavendulae* FRI-5

Streptomyces lavendulae FRI-5 was isolated from a soil sample in Japan and this strain predominantly produces only 4 secondary metabolites in laboratory fermentation conditions- the blue pigment indigoidine, nucleoside antibiotics (showdomycin and minimycin) and an anti-

tuberculosis drug, D-cycloserine (Fig. 1.13). The production of these compounds is controlled by the γ-butyrolactone autoregulator, IM-2, that initiates a regulatory cascade by binding to its receptor FarA (Hashimoto, Takuya, Sakuda, & Yamada, 1992; Waki, Nihira, & Yamada, 1997). Distinct from other autoregulators that only exert positive effects towards secondary metabolism, IM-2 can switch on the production of nucleoside antibiotics and indigoidine and at the same time block D-cycloserine production. It is this unique feature that prompted years of comprehensive research in our laboratory, and so far, we have already unraveled the involvement of several regulators of the *far* regulatory island in the biosynthesis of IM-2, FarA and indigoidine (Kurniawan et al., 2016; Kurniawan, Kitani, Maeda, & Nihira, 2014; Waki et al., 1997). However, since IM-2 biosynthesis easily responds to slight environmental changes during cultivation, the production of these 4 metabolites is very unstable. Therefore, the biosynthetic gene(s) responsible for these compounds and the mechanism by which the IM-2/FarA regulon controls their expression still needs to be clarified.

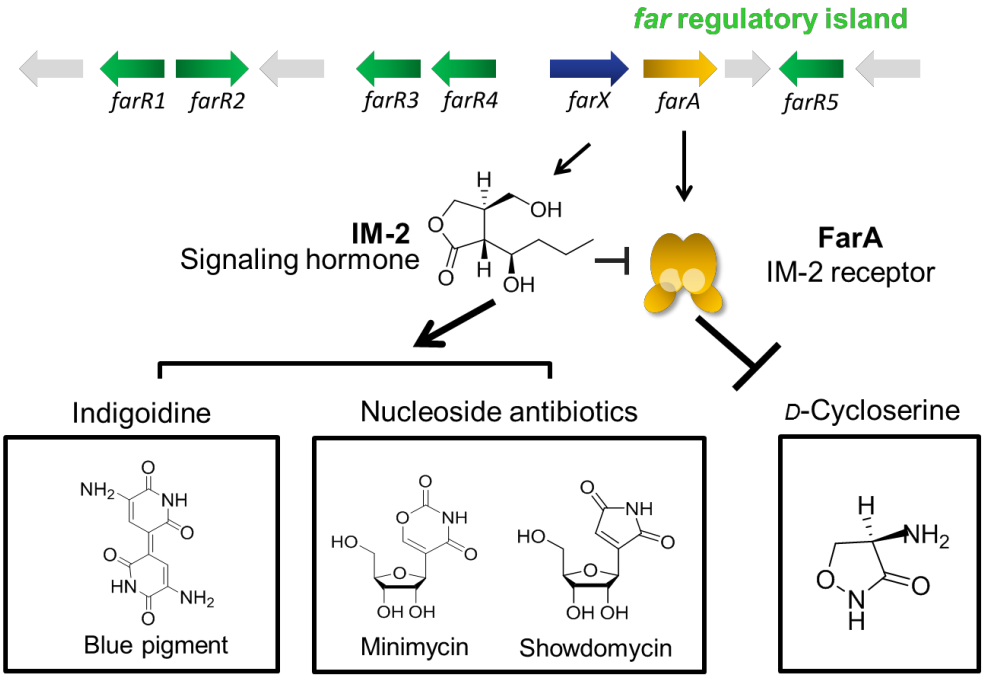


Figure 1.13. Secondary metabolism of *Streptomyces lavendulae* FRI-5

1.6 The aim of the present study

The rapid and inexpensive genome sequencing of many *Streptomyces* strains revealed a wealth of untapped natural product resources within microbial genomes. This led us to hypothesize that *Streptomyces lavendulae* FRI-5 may also have the capacity to biosynthesize additional, potentially valuable secondary metabolites aside from IM-2-controlled secondary metabolites. In this work, I identified several *S. lavendulae* FRI-5 "cryptic" BGCs and focused on clusters with intriguing properties or those with the potential to direct the production of yet undiscovered natural products. I subsequently aimed to isolate the corresponding products of these "cryptic" BGCs using a heterologous expression system involving the genome-minimized host, *S. avermitilis* SUKA22. *S. avermitlis* SUKA22 has successfully expressed known polyketide and nonribosomal peptide BGCs, with yields sometimes surpassing the native producer. Thus, I hypothesized that it might also be used to effectively clarify the products of silent biosynthetic pathways from *S. lavendulae* FRI-5 and to discover new compounds. This work led to the identification of the indigoidine biosynthetic enzyme and other IM-2-responsive cryptic BGCs, as well as the successful isolation of a novel compound with unprecedented biosynthetic properties through the awakening of a silent gene cluster. An overview of this study is as follows:

Chapter 2 describes my attempts to express *S. lavendulae* FRI-5 biosynthetic genes in the *farA*-flanking region. Autoregulator receptor genes are frequently situated with biosynthetic genes under their regulatory control. Thus, I found several uncharacterized biosynthetic genes including a putative indigoidine synthase (*lbpA*) and a cryptic angucycline (*lac*) gene cluster in the vicinity of the *far* regulatory island. I verified the transcriptional control of IM-2 on these newly-identified genes and I initially confirmed the efficiency of using *S. avermitilis* SUKA22 as a host through the functional analysis of LbpA. This work revealed the role of LbpA in the IM-2 signaling cascade.

Next, I also attempted to isolate the product of the *lac* cluster which lacks common angucyline biosynthetic enzymes (ketoreductase and aromatase) but heterologous expression failed to uncover its product, likely because of inherently nonfunctional genes.

In Chapter 3, I focused on a transcriptionally silent type I PKS gene cluster (*lav*) that I assembled from the partial genome sequence of *S. lavendulae* FRI-5. Introduction of the cluster into *S. avermitilis* SUKA22 led to the appearance of several new peaks including the major species lavendiol, a novel polyketide with a terminal diol moiety similar with streptenol compounds. The biosynthetic machinery for streptenols have never been reported whereas here, I proposed the lavendiol biosynthetic pathway which involves a unique PKS initiation mechanism.

Lastly, general conclusions are presented in Chapter 4, together with a summary of other *S. lavendulae* FRI-5 biosynthetic gene clusters that can be targeted for heterologous expression using the genome-minimized host, *S. avermitilis* SUKA22. Here, I also outlined the future directions of this work.

Chapter 2

Identification and heterologous expression of IM-2 controlled biosynthetic genes

2.1 Introduction

The Gram-positive bacteria of the genus *Streptomyces* are well known for their ability to produce numerous secondary metabolites, including industrially important bioactive compounds. In general, these useful secondary metabolites are produced by complex biosynthetic pathways encoded by physically clustered genes (Bibb, 2005; Gregory L. Challis, 2008; Katz & Baltz, 2016). The expression of biosynthetic gene clusters for secondary metabolites is often controlled by pathway-specific transcriptional regulators that are embedded in each cluster and are regarded as lower-level regulators in the regulatory system for *Streptomyces* secondary metabolism (Bibb, 2005; Martín & Liras, 2010). These pathway-specific regulators are hierarchically controlled by higher-level regulators that respond to various physiological and environmental factors and whose genes also occasionally lie within the cognate cluster. Thus, the secondary metabolite production is intricately regulated via several hierarchical layers of regulators. Among the well-known regulatory systems for *Streptomyces* secondary metabolism are the γ-butyrolactone autoregulator signaling cascades (Nishida, Ohnishi, Beppu, & Horinouchi, 2007; Takano, 2006). Autoregulator molecules initiate the production of secondary metabolites by regulating the DNA-binding activity of their own receptors. The binding of γ-butyrolactone autoregulators to the cognate receptor activates the expression of target genes, resulting in the synchronized transcription of biosynthetic genes for secondary metabolites. An autoregulator receptor gene is often clustered with regulatory genes that are under the transcriptional control of both the γ-butyrolactone autoregulator and receptor, and in its vicinity are biosynthetic genes of one or more secondary metabolites. Such

multiple regulatory genes in the same locus are referred to as a regulatory island and are found in many *Streptomyces* species (Arakawa, Mochizuki, Yamada, Noma, & Kinashi, 2007; Bate, Butler, Gandecha, & Cundliffe, 1999; Mingyar, Feckova, Novakova, Bekeova, & Kormanec, 2014; Pulsawat, Kitani, & Nihira, 2007). The regulatory island for virginiamycin production comprises BarA (an autoregulator receptor), BarB (a pseudoreceptor regulator), and other regulators that coordinately regulate the expression of neighboring virginiamycin biosynthetic genes (Matsuno, Yamada, Lee, & Nihira, 2004; Pulsawat et al., 2007). In the absence of a complete genome sequence, regulatory islands could be an important clue to identify potential biosynthetic gene clusters for new secondary metabolites, the production of which is controlled by γ-butyrolactone autoregulators.

Streptomyces lavendulae FRI-5 produces the blue pigment indigoidine as well as other secondary metabolites, such as the antituberculosis agent D-cycloserine and nucleoside antibiotics (Hashimoto, Nihira, Sakuda, & Yamada, 1992). The production of these secondary metabolites is controlled by the *far* regulatory island, which plays a pivotal role in the signaling cascade mediated by IM-2, one of the γ-butyrolactone autoregulators (Shigeru Kitani et al., 2008). The IM-2−specific receptor FarA in the *far* regulatory island exhibits DNA-binding activity toward FarA-response elements (FAREs). The binding of IM-2 to FarA leads to the loss of the DNA-binding activity of FarA and induces the production of indigoidine (S. Kitani, Kinoshita, Nihira, & Yamada, 1999). This indicates that FarA negatively controls the biosynthesis of indigoidine, in addition to positively regulating IM-2 biosynthesis (S. Kitani, Yamada, & Nihira, 2001). We have recently revealed that other regulatory genes in the *far* regulatory island exert their functions in indigoidine production at various levels in the IM-2 signaling cascade. FarR2 (a FarA homologue and pseudoreceptor regulator) is the direct transcriptional target of FarA through the binding of FarA

to FARE3, and it positively controls the onset of indigoidine production in response to the presence of IM-2 (Kurniawan et al., 2016). FarX (an *afsA*-family protein) is an IM-2 biosynthetic enzyme, FarR3 (a *Streptomyces* antibiotic regulatory protein (SARP) family regulator), functions positively in indigoidine production in both a temporal and quantitative manner, and FarR4 (also a regulator of the SARP family) acts as a negative regulator of IM-2 biosynthesis (Shigeru Kitani, Doi, Shimizu, Maeda, & Nihira, 2010; Kurniawan et al., 2014). These regulators in the *far* regulatory island form a hierarchical network for the regulation of indigoidine production. These findings imply that the adjacent regions of the *far* regulatory island may contain biosynthetic genes for indigoidine as well as *D*-cycloserine, nucleoside antibiotics, and other IM-2 cascade-dependent compounds yet to be identified.

In the present study, I found biosynthetic genes for the biosynthesis of indigoidine and some cryptic compounds in the left-flanking region of the *far* regulatory island and sought out to confirm their roles in the IM-2 mediated secondary metabolism. I initially characterized the indigoidine biosynthetic gene *lbpA* through heterologous expression combined with the indigoidine precursor feeding using a *Streptomyces* host suitable for secondary metabolite production. Thereafter, I attempted to activate the nearby cryptic angucycline gene cluster which appears to have intriguing characteristics. The angucycline group of natural products is the largest group of polycyclic aromatic polyketides, rich in chemical scaffolds and various biological activities, predominantly anticancer and anti-bacterial (Kharel et al., 2012). In comparison to well-studied angucycline biosynthetic gene clusters, the *lac* cluster is the only reported cluster lacking two essential genes for the formation of the cyclic angucycline framework, particularly a ketoreductase for the formation of a secondary alcohol, and an aromatase that is essential for the aromatization of the first ring (Kharel et al., 2012; Rohr & Thiericke, 1992). In this chapter, I

aimed to determine whether or not this biosynthetic feature is responsible for the production of a new angucycline or angucyline-related compound upon awakening by heterologous expression.

2.2. Materials and Methods

2.2.1 Bacterial strains, plasmids, and growth conditions

Streptomyces lavendulae FRI-5 wild-type strain (MAFF10-06015; National Food Research Institute, Tsukuba, Japan) was grown on ISP medium 2 (Becton, Dickinson and Company, Franklin Lakes, NJ, USA) for spore formation, and *Streptomyces avermitilis* SUKA22 (Komatsu et al., 2010) was used as the host for heterologous expression of *lbpA* and the *lac* genes. *Escherichia coli* DH5α was used for general DNA manipulation *E. coli F- dcm* Δ*(srl-recA)306::Tn10 carrying pUB307-aph::Tn7 was used for E.coli/Streptomyces* conjugation, *Escherichia coli* BW25141 containing pKD119 (Komatsu, Tsuda, Omura, Oikawa, & Ikeda, 2008) was used for the cloning of *lbpA* and *lac* genes by λ Red-mediated recombination (Sharan, Thomason, Kuznetsov, & Court, 2009), and *Escherichia coli* GM2929 *hsdS*::Tn*10* was used to prepare unmethylated plasmid DNAs for transformation into *S. avermitilis* SUKA22. pBluescript II SK was used for general cloning, a small pRED vector (Komatsu et al., 2010) for *in vivo* cloning of the biosynthetic genes, pSET152 (Bierman et al., 1992) for heterologous expression of *lbpA*, pKU492 (Komatsu et al., 2010) for heterologous expression of *lac* genes and pLT129 (Daduang et al., 2015) for adding the missing ketoreductase and aromatase genes. The phenotypes of *S. lavendulae* FRI-5 strains and *S. avermitilis* SUKA22 strains were analyzed after growth either in liquid or solid medium B (containing [in grams per liter] yeast extract, 7.5; glycerol, 7.5; NaCl, 1.25 [pH 6.5]) (Hashimoto, Nihira, et al., 1992) or in the *S. avermitlis* synthetic medium (Cane, He, Kobayashi, Omura, & Ikeda, 2006). The media conditions and general *E. coli* and

Table 2.1 Oligonucleotides used in this study

Primer	Sequence (5' - 3')
For PCR-amplification of a gene encoding a homologue of an indigoidine biosynthetic gene	
dlbpA-Fw	CCGACAAYCAGGTCAAGYTSCG
dlbpA-Re	GCCRAARGAGTAGCCCCACAG
For PCR-amplification of a region from *lbpA* to *farD*	
farD-2	ACGTACGGGCCAAGATGGTCTCGGTACT
Lbp-4	ATGCCGCTCTTCTTGCCGATGAAAGT
For RT-PCR analysis	
or14-Fw	TGGTTCCGAAGACCTGCTGCGGTA
orf14-Rv	TTACCGCGTACCGGACATCTGCCA
orf13-Fw	TAGCCGGAGCTCTTGGCGTCTTTGA
orf13-Rv	AGTTCCAGCCCTGGATCCTCAAA
orf12-Fw	ACCCGTGCGTGAGGCGATCTTGTCCT
orf12-Rv	TACCGTGCGTTCTCCTGAGGGGAGAT
orf11-Fw	TTCCGCTCGTCCTCTAGAACGGGAA
orf11-Rv	ATCGGGACGGCAGAGCAGTTCGCGAA
orf10-Fw	TACCTGCGTCCGGACGGGAAGACCCA
orf10-Rv	ATGCCCCCGTACGTGTCGATGATGAT
orfR-Fw	ACGTGCGGATCTTCGAGCGGCTCGTCA
orfR-Rv	ATGAGCCGGCTGTGGAACCACTCGTT
lbpA-Fw	ACATCGAGGCCCTCAAGGAGATCCA
lbpA-Rv	TTGAGCTCGCTGATGAACTCGGCGAA
orf5-Fw	TGATGAAGACCAGCCGTACGCTT
orf5-Rv	ATCGGCTTGATCTTGTCCTTCAGCA
orf4-Fw	AGTCGGTCGTGCCGAAGTAGCGGT
orf4-Rv	CCAAGCTCCCGCACCTGTACTACAA
orf3-Fw	TTGACGAAGCGGATCGCGTCGGCGA
orf3-Rv	CGTCCAGCTGCACGGTGACGTCA
orf2-Fw	AGTAGGCGTTCCTCCAGGAAGGCCA
orf2-Rv	TGCGCTTCTACCCGCCGGAGGCCTT
lacH-Fw	CGAACAGCGAGTAGACGGACACGG
lacH-Rv	TCACCCGGGTCCAGTGGATCGTCA

Table 2.1 continued. Oligonucleotides used in this study

Primer	Sequence (5' - 3')
lacG-Fw	CTCGACCTCGGTCTGCCCG
lacG-Rv	TGCCGGGGAAGTCGAAGCCGCCGA
lacD-Fw	CCGCCCGAACCACTCCAC
lacD-Rv	GGGCACGAAGTGGTCGTACAGGT
lacB-Fw	GGCTTCTGTTTCGTATGTCT
lacB-Rv	TGGCCTGCCACTTGTCGACGGCCT
lacA-Fw	ATGCCCCACCTGGAACTGACC
lacA-Rv	TCCCGAACCAGCGGTCGAGCGCCAT
For heterologous expression of *lbpA*	
pRED-lbpA-Fw	CTTCGGACGGACGGGCCCCGTCGGGCCCGTCCGTCCGGGCGAGGATCCTGCCAGGAAGATACTTAACAG
pRED-lbpA-Re	CCGGCCATCAGAGGGCGTCGGGTGGCTCAATTCTGACGCAGAGGATCCCCATTCATCCGCTTATTATC
For heterologous expression of *lac* cluster	
pRED-lac-Fw	CAGGACCACTACGCCCGCAAGCACGGAAACCCGCCGCGGTGAGCTAGCTGCCAGGAAGATACTTAACAG
pRED-lac-Re	CGAGTGCGGACCGCTGTCGGGGGTCGTGTGGACGGGACCGTCAAGCTTCCATTCATCCGCTTATTATC
pRED-OElac-Fw	CAGGACCACTACGCCCGCAAGCACGGAAACCCGCCGCGGTGAAAGCTTTGCCAGGAAGATACTTAACAG
pRED-OElac-Re	ACCCGCGCCGACGACGATCACAGGGGCGTCCATCACGTCTCCACTAGTCATTCATCCGCTTATTATC
OErpsJ-Fw	CCGGAATTCGGCTTATGTCCAATTCCACG
OErpsJ-Re	TGCTCTAGATGTACTCAGTAGTCCTTCGTC
jadDE-Fw	CGTGCCGGTTGGTAGGAAGGAGCGACACCGCATGTCCCAG
jadDE-Re	CTTTAGATTCTAGAGGTCAGCGCTTGCCCTCGGCG
pRED-OEmin-Fw	CGCGCCACGATCAGGTTGCTGTGCATTGTTTCTACTCTCTCCACTAGTCATTCATCCGCTTATTATC
pRED-OEmin-Re	CCACCTCCGCCCCCACCCAGCCCACAGAAAGGGAGTACGCCCACTAGTTGCCAGGAAGATACTTAACAG

Restriction sites are underlined

Streptomyces manipulations were as described previously (Kieser et al., 2000). The primers used in this work are listed in Table 2.1.

2.2.2 Cloning of *lbpA* and *lbpA*-flanking regions, and sequence analysis

The degenerate primers dlbpA-Fw and dlbpA-Re were designed and used to amplify an internal segment of a *bpsA* homologue from *S. lavendulae* FRI-5. The PCR product was clonedinto pBluescript II SK, analyzed by DNA sequencing, and used as a probe for further screening The 22.7-kb DNA fragment covering the region from *lbpA* to *farD* was amplified using the primer pair farD-2 and Lbp-4. The PCR product was digested with *Pst*I, and each DNA fragment was also employed as a probe for screening. Genomic DNA of *S. lavendulae* FRI-5 was analyzed by Southern hybridization with these DNA probes. Partial genomic libraries were constructed with size-fractionated *Bam*HI fragments, *Eco*RI fragments, *Mlu*I fragments, or *Sph*I fragments and pBluescript II SK. These libraries were screened by colony PCR with primers designed based on sequences of probe DNAs, and the DNA sequences of the inserts on the positive plasmids were determined by primer walking. Open reading frames (ORFs) and gene functions were annotated manually using the FramePlot 4.0beta program (http://nocardia.nih.go.jp/fp4/), the BLAST algorithm, and the web-based PKS/NRPS analysis program (http://nrps.igs.umaryland.edu/nrps). The nucleotide sequence data reported in this paper have been deposited in the DDBJ under accession number LC209815.

2.2.3 Analysis of indigoidine production

One milliliter of each culture supernatant was collected at the indicated intervals and filtered through a 0.2-μm-pore-size filter, and absorbance at 590 nm was measured for the production of indigoidine.

2.2.4 Gene expression analysis of the identified genes

Total RNAs were extracted from mycelia harvested at the indicated cultivation times by an RNeasy Mini Kit (QIAGEN Sciences, Germantown, MD, USA) and treated with DNase I (Takara Bio, Shiga, Japan). The complementary DNA (cDNA) was synthesized using GoScript Reverse Transcriptase (Promega, Madison, WI, USA) and random primers (Invitrogen, Carlsbad, CA, USA). For semiquantitative RT-PCR, PCR amplification was carried out under the following conditions: a single round of denaturation at 95°C for 2 min and 27 cycles of 98°C for 30 s, 57°C for 30 s, and 72°C for 1 min, followed by a single extension at 72°C for 1 min. The absence of DNA contamination was confirmed by RT-PCR without reverse transcriptase (Fig S2). The *hrd*B gene (Kurniawan et al., 2014) was used as an internal control.

2.2.5 Expression of *lbpA* in *S. avermitilis* SUKA22

The cloning vector pRED was amplified by PCR with the primer pair pRED-lbpA-Fw and pRED-lbpA-Re. The PCR product was treated with *Dpn*I to remove template DNA. A 1.7-kb fragment of pRED harboring upstream and downstream sequences of *lbpA* was cotransformed into *L*-arabinose−induced *E. coli* BW25141 carrying pKD119 with the *S. lavendulae* FRI-5 cosmid cE93, which contains the 38.8-kb region covering 20 genes from *lacH* to *orf14*. pLT135 harboring the entire *lbpA* gene with its 100-bp upstream and downstream regions was generated after λ Red-mediated recombination and a 4.1-kb *Bam*HI fragment, recovered from pLT135, was inserted into the *Bam*HI site of pSET152 to yield pLT136. After demethylation by *E. coli* GM2929 *hsdS*::Tn*10*, pLT136 was introduced into *S. avermitilis* SUKA22 by protoplast transformation, followed by apramycin selection and PCR analysis. Spores (1.0 x10^8 CFU) of *S. avermitilis* strains were inoculated into 70 mL of liquid medium B in a 500-mL baffled flask at 28°C.

2.2.6 Isolation and MS analysis of the blue pigment

The blue pigment, produced by *S. avermitilis* SUKA22 expressing the *lbpA* gene, was isolated as described by Kurniawan *et al*. (Kurniawan et al., 2014) with slight modifications. The culture supernatant supplemented with 100 mM L-glutamine was collected after 2 days of cultivation and centrifuged at 25,000 x *g* for 45 minutes. The precipitated blue pigment was washed twice with water and methanol, and dried *in vacuo*. The molecular mass of the sample dissolved in DMSO was analyzed by electron impact−mass spectrometry (EI−MS) with a JMS-700 mass spectrometer (JEOL, Tokyo, Japan).

2.2.7 Expression of the *lac* cluster in *S. avermitilis* SUKA22

To clone the *lac* cluster and its native promoter region, the cloning vector pRED was amplified by PCR with the primer pair pRED-lac-Fw and pRED-lac-Re. The PCR product was treated ith *Dpn*I to remove template DNA. A 12.2-kb fragment of pRED harboring *lacA* upstream and *lacK* downstream sequences was cotransformed into L-arabinose−induced *E. coli* BW25141 carrying pKD119 with the *S. lavendulae* FRI-5 cosmid cWW51, which contains the region covering *orf6* to the *far* cluster. pLTlac harboring the entire *lac* cluster with the *farD* intergenic region and 100-bp *lacK* downstream region was generated after λ Red-mediated recombination. A 12.2-kb *Hind*III/*Nhe*I fragment, recovered from pLTlac, was inserted into the same site of pKU492 to yield pKUlac. To generate the *lac* cluster over-expression strain, the pRED plasmid was amplified using the primer pair pRED-OElac-Fw and pRED-OElac-Re. The PCR product was used to clone the *lac* operon from cosmid cWW51 without its native promoter by the λ Red-mediated recombination method previously described to generate plasmid pLTOElac. The promoter region of *rpsj* gene encoding the ribosomal protein S10 was prepared by PCR using the plasmid pKU1021 (Komatsu et al., 2013) and the primer pair OErpsJ-Fw and OErpsJ-Rv. The

PCR amplicon digested with *EcoR*I and *Xba*I was ligated into the *EcoR*I/*Xba*I long fragment of pKU492 to generate pKU492/*rpsJp*. The 11.8-kb *Spe*I/*Hind*III fragment, recovered from pLTOElac, was inserted into the same site of pKU492/*rpsJp* to yield pKUOElac. After demethylation by *E. coli* GM2929 *hsdS*::Tn*10*, vectors pKUlac or pKUOElac was introduced into *S. avermitilis* SUKA22 by protoplast transformation, followed by apramycin selection and PCR analysis.

In a similar way, the minimum set of angucycline genes (*lacC to lacF*) was cloned from cosmid cWW51 by λ Red-mediated recombination method using the primer pair pRED-OEmin-Fw and pRED-OEmin-Re to generate pLTOEmin. The 3.4-kb *Spe*I/*Hind*III fragment, recovered from pLTOEmin, was inserted into the same site of pKU492/*rpsJp* to yield pKUOEmin. The resultant plasmid was introduced into *S. avermitilis* SUKA22 as previously described.

2.2.8 Detection of secondary metabolites in the *lac* cluster-expression strain

Spores (1.0 x10^8 CFU) of *S. avermitilis* strains were inoculated into 70 mL of synthetic production medium and incubated on a rotary shaker (160 rpm) for 7 days at 28ºC. The culture broth was extracted with equal volume of ethyl acetate, and the ethyl acetate layer was evaporated and dissolved in methanol. The extract was analyzed by reversed-phase HPLC on a Cadenza C_{18} column (3 µm; 4.6 i.d. x 75 mm; Imtakt, USA) developed with a gradient system of CH_3CN (15% for 0-3 min; 15-85% for 3-25 min; 85% for 25-29 min, 85-15% for 29-32 min) containing 0.1% formic acid (flow rate, 1.2 mL/min; UV detection, 254 nm).

2.2.9 Adding the missing KR and ARO genes

Genes *jadE* and *jadD* encoding an angucycline ketoreductase and aromatase, respectively, was PCR-amplified from the genomic DNA of *Streptomyces venezuelae* ATCC 10712 (Han, Yang,

Ramalingam, Mosher, & Vining, 1994) using the primer pair *jadDE*-Fw and *jadDE*-Rv and cloned into the *Bam*HI site of pLT129 (Daduang et al., 2015) using a GeneArt Seamless Cloning and Assembly Kit (Life Technologies, CA, USA). The plasmid designated as pKU/KRARO was introduced into the *lac* overexpression strains by intergeneric conjugation and the integration of the plasmid into the ΦC31 attachment site was confirmed by hygromycin resistance and PCR analysis.

2.3. Results

2.3.1 Cloning and sequence analysis of biosynthetic genes flanking the IM-2 receptor gene in *S. lavendulae* FRI -5

We previously found that *S. lavendulae* FRI-5 produces 4 secondary metabolites including the blue pigment indigoidine and that indigoidine biosynthesis is under the control of the IM-2/FarA system (Hashimoto, Takuya, et al., 1992; S. Kitani et al., 2001). Indigoidine has been identified also in other streptomycetes (Novakova, Odnogova, Kutas, Feckova, & Kormanec, 2010; Takahashi et al., 2007) and is synthesized by a nonribosomal peptide synthetases (NRPSs) such as BpsA of *Streptomyces lavendulae* ATCC11924 (Takahashi et al., 2007). Thus, we assumed that indigoidine from *S. lavendulae* FRI-5 is synthesized by a similar NRPS biosynthetic pathway. We designed two degenerate primers (dlbpA-Fw and dlbpA-Re) based on the amino acid sequence of the oxidation domain of BpsA, and successfully PCR-amplified a target DNA region that encodes a part of a protein highly similar to BpsA (Fig. S1). To reveal the entire amino acid sequence encoded by the *bpsA* homologue gene, this PCR product was used as a probe to screen a partial genomic library, and a 6.9-kb BamHI fragment was cloned. The nucleotide sequence of the DNA fragment suggests the presence of two complete genes, named *lbpA* (S. lavendulae FRI-5

blue pigment synthetase A) and *orfR*, and two incomplete genes (*orf9* and *orf10*). As described below, *lbpA* encodes a plausible NRPS for indigoidine biosynthesis, and *orfR* encodes a homologue of a cluster-situated regulatory gene for the production of secondary metabolites. In streptomycetes, multiple regulatory genes controlling the biosynthesis of one or more secondary metabolites are frequently localized in the vicinity of the biosynthetic gene cluster (Bibb, 2005; Martín & Liras, 2010). Thus, we assumed that *lbpA* is likely to be in the vicinity of the receptor gene *farA*, together with other biosynthetic genes for secondary metabolites. We attempted to PCR-amplify the region from *lbpA* to *farD*, and obtained a 22.7-kb DNA fragment (data not shown). This revealed that *lbpA* is located in the left-flanking region of the *far* regulatory island. Partial genomic libraries of *S. lavendulae* FRI-5 were constructed and screened with a chromosome-walking procedure to clone the downstream region of *farD*. The DNA fragments and the PCR-amplified gaps between the fragments were sequenced and assembled. Annotation analysis of the sequence and comparison with genes in a public database showed 27 newly identified ORFs. The deduced genetic organization is shown in Figure 2.1 and summarized in Table 2.1. The left-flanking regions of *farD* was revealed to comprise several uncharacterized biosynthetic genes including those similar to type II polyketide synthase (PKS) gene clusters from several *Streptomyces* species, as well as genes involved in nucleoside metabolism and secondary metabolism, in addition to *lbpA* and *orfR*. Proteins encoded by 11 genes, designated as *lacA*-*lacK* (lavendulae angucycline cluster), exhibited high similarity to genes involved in the biosynthesis of the angucycline-group antibiotic auricin in *Streptomyces aureofaciens* CCM 3239 (Kormanec, Novakova, Mingyar, & Feckova, 2014). The deduced gene products of *lacD* and *lacE* have similarities with β-ketoacyl synthase (KS) units KSα and KSβ, respectively, and the *lacF* product encodes an acyl-carrier protein (ACP) domain containing the conserved 4'-phosphopantetheine

attachment site serine residue. These three gene products may form a minimal PKS that generates a decaketide backbone in the typical type II PKS biosynthetic pathway (Kharel et al., 2012). The lac cluster also contains *lacC*, which encodes putative cyclase for the closure of the last aromatic ring in angucycline biosynthesis. However, this cluster has no gene encoding a ketoreductase (KR) responsible for the reduction of the C-9 keto group into a hydroxyl group and an aromatase (ARO) that is essential for the closure and aromatization of the first ring. The lack of these two important enzymatic genes might have led to the absence of any

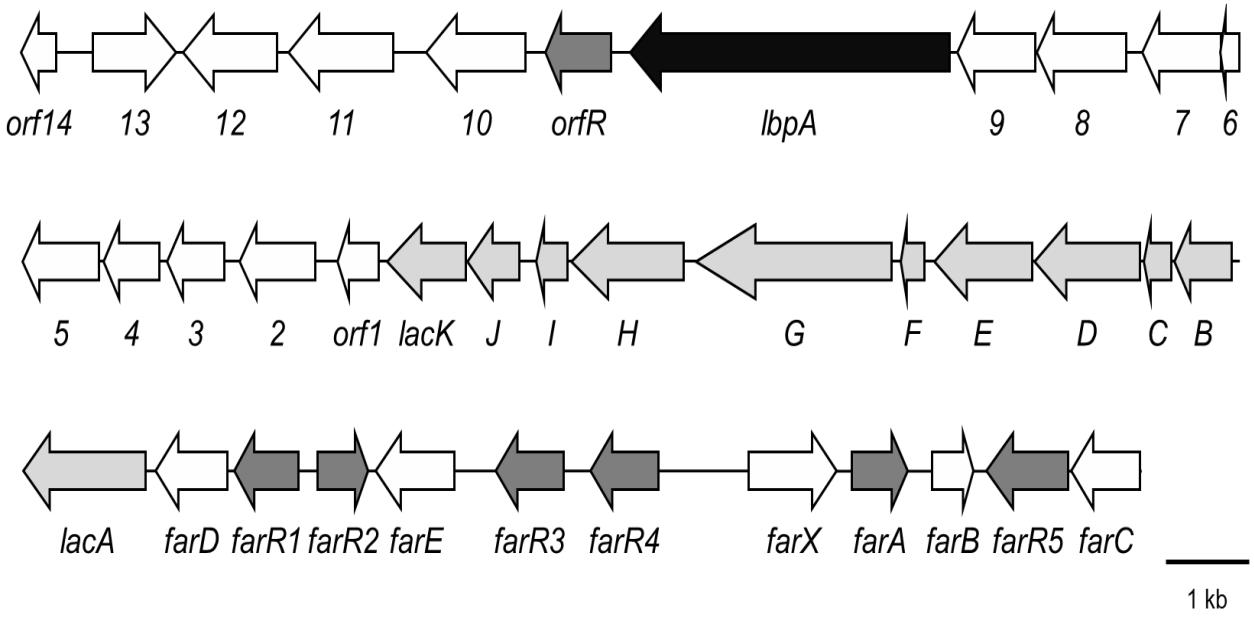


Figure 2.1. Gene organization of a 43.3-kb region containing the *far* regulatory island. Each arrow indicates the direction of transcription and relative gene size. The genes shaded in dark gray correspond to regulatory genes, and those shaded in light gray to a type II polyketide synthase gene cluster.

angucycline-group antibiotics in the culture of *S. lavendulae* FRI-5 or may result into the formation of another aromatic polyketide compound. Despite the missing genes KR and ARO genes, this gene cluster still harbors other putative biosynthetic genes, such as *lacA* and *lacG*, that are involved in the tailoring steps of the intermediate, as well as those, such as *lacH*, that provide self-resistance against the final product.

Table 2.1. Deduced functions of gene products identified in the left-flanking region of the IM-2 receptor gene *farA*

Gene	Size[a]	Putative function	Homolog and origin	Identity /similarity (%)	Accession number
orf14	160	Hypothetical protein	SAMR1079, *S. ambofaciens*	70/79	CAJ88788
orf13	352	Transglycosylase	SAMR1080, *S. ambofaciens*	69/80	CAJ88789
orf12	379	Antibiotic production activating factor	SAMR1083, *S. ambofaciens*	60/75	CAJ88792
orf11	422	Hypothetical protein	Hypothetical protein, *Streptomyces* sp. RSD-27	77/86	WP_042819335
orf10	402	*S*-Adenosylmethionine synthase	Methionine adenosyltransferase, *S. pristinaespiralis*	92/95	WP_005319909
orfR	262	SARP family transcriptional regulator	Aur1PR4, *S. aureofaciens*	82/90	ACK77758
lbpA	1282	Indigoidine synthase	BpsA, *S. lavendulae*	98/98	BAE93896
orf9	313	Ribose-phosphate pyrophosphokinase	Sa7, *S. aureofaciens*	97/98	ACK77756
orf8	359	Oxidoreductase	Sa6, *S. aureofaciens*	92/95	ACK77755
orf7	314	Pseudouridine-5'-phosphate glycosidase	Sa5, *S. aureofaciens*	97/99	ACK77754
orf6	77	Hypothetical protein	Sa4, *S. aureofaciens*	82/93	ACK77753
orf5	307	ABC transporter permease	Sa3, *S. aureofaciens*	89/89	ACK77752
orf4	221	Uracil phosphoribosyltransferase	Sa2, *S. aureofaciens*	83/92	ACK77751
orf3	229	Anthranilate isomerase	Sa1, *S. aureofaciens*	94/96	ACK77750
orf2	304	Acetyltransferase	Aur1U, *S. aureofaciens*	75/82	ACK77749
orf1	165	Hypothetical protein	Aur1X, *S. aureofaciens*	72/81	ACK77747
lacK	316	Malonyl-CoA-ACP transcylase	Aur1M, *S. aureofaciens*	84/87	AAX57200
lacJ	206	Phosphopantetheinyl transferase	Aur1L, *S. aureofaciens*	86/88	AAX57199
lacI	125	Hypothetical protein	Aur1K, *S. aureofaciens*	90/91	AAX57198
lacH	450	Transporter protein	Aur1J, *S. aureofaciens*	69/70	AAX57197

Gene	Size	Putative function	Homolog and origin	Identity Similarity	Acession No.
lacG	786	Oxygenase	Aur1I, *S. aureofaciens*	78/83	AAX57196
lacF	94	Acyl carrier protein	Aur1F, *S. aureofaciens*	80/87	AAX57193
lacE	404	Ketoacyl synthase β	Aur1E, *S. aureofaciens*	82/84	AAX57192
lacD	422	Ketoacyl synthase α	Aur1D, *S. aureofaciens*	85/90	AAX57191
lacC	109	Cyclase	Aur1C, *S. aureofaciens*	81/90	AAX57190
lacB	231	Transcriptional regulator	Aur1B, *S. aureofaciens*	77/82	AAX57189
lacA	489	Oxygenase	Aur1A, *S. aureofaciens*	90/93	AAX57188

[a] Numbers refer to amino acid residues.

The *lac* cluster was followed by nine genes (*orf1−orf9*) encoding proteins assigned with no specific functions in secondary metabolism. The gene product of *orf3* might share similar activity with phosphoribosylanthranilate isomerases in tryptophan biosynthesis (Priestle et al., 1987). The protein encoded by *orf4* is homologous to the members of uracil phosphoribosyltransferases that are key enzymes for the salvage pathway in the synthesis of uridine 5'-monophosphate (Hughes, Beck, & O'Donovan, 2005; Sinha & Smith, 2001). The deduced gene product of *orf9* shows significant similarity to ribose-phosphate pyrophosphokinases that produce phosphoribosyl pyrophosphate in the synthesis of nucleotides (Kilstrup, Hammer, Jensen, & Martinussen, 2005). The homologues of Orf4 and Orf9 are found in the genomes of *Streptomyces coelicolor* A3(2), *Streptomyces avermitilis* MA-4680, and *Streptomyces griseus* IFO 13350, but none of the homologues is clustered in the genome. Furthermore, the homologues of Orf3 are not frequently found in the streptomycete genomes. Thus, the *orf1-orf9* genes are assumed to also be involved in the biosynthesis of cryptic secondary metabolites. In addition, the region

downstream of *orfR* contains several genes related to secondary metabolism, such as *orf10, orf12*, and *orf13.* However, their involvement or function in the secondary metabolism of *S. lavendulae* FRI-5 remains to be elucidated.

2.3.2 Transcriptional analysis of genes identified in the *farA*-flanking regions

External addition of IM-2 to culture at the early exponential stage induces earlier production of large amounts of indigoidine with growth retardation (Fig. 2.2, (S. Kitani et al., 2001; Shigeru Kitani et al., 2010)). In addition, regulatory genes under the control of the IM-2/FarA system are tightly regulated by IM-2 at the transcriptional level (Shigeru Kitani et al., 2010; Kurniawan et al., 2016, 2014). To know whether the newly identified genes are involved in the IM-2−mediated secondary metabolism, the effect of IM-2 on their transcription was monitored by semiquantitative RT-PCR analysis. The transcriptional levels of genes from *orfR* to *lacA* showed remarkable increases in the presence of external IM-2, whereas genes from *orf14* to *orf10* did not significantly respond to it (Fig. 2.3). These results suggest that the putative angucycline cluster (*lacA* to *lacK*), putative biosynthetic genes *(orf1* to *orf9*) for unidentified compounds, *lbpA,* and *orfR* are new members of the IM-2 regulon in addition to the far regulatory genes, and they are likely involved in IM-2−mediated secondary metabolism.

2.3.3 Transcriptional control by the IM-2/FarA system of *lbpA*

Among the newly identified members of the IM-2 regulon, the product of *lbpA* was the initial target of interest to confirm the efficacy of the genome-minimized host *S. avermitlis* SUKA22 in expressing biosynthetic genes from *S. lavendulae* FRI-5. LbpA showed high similarity and identity with BpsA, an indigoidine synthetase from *S. lavendulae* ATCC11924 (Table 1). LbpA possesses a single NRPS module, including an adenylation domain (A) for substrate

recognition and activation, an oxidation (Ox) domain that is integrated into the A-domain and is implicated in the

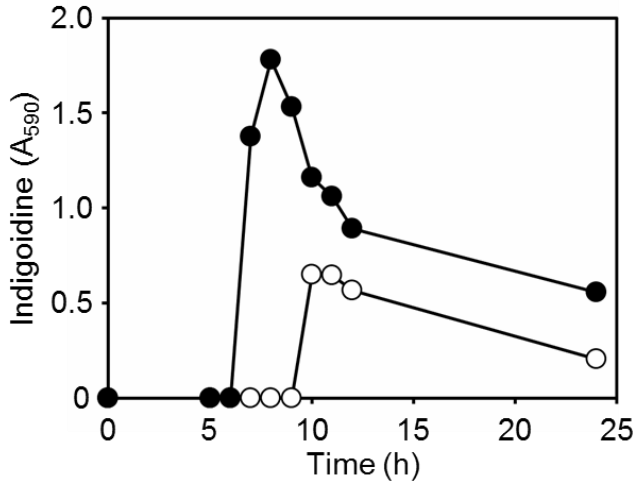


Figure 2.2 Effect of IM-2 on indigoidine production. Indigoidine production in liquid culture: open and solid circles indicate indigoidine production without and with the addition of IM-2 (100 nM) at 5 h of cultivation.

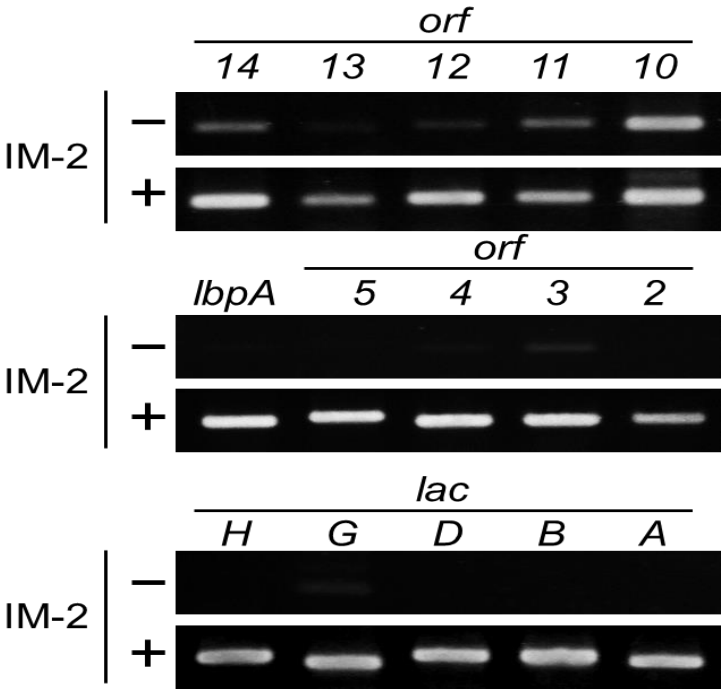


Figure 2.3 Effect of IM-2 on the transcription of the identified genes. Transcriptional analysis by semi-quantitative RT-PCR. Total RNAs were extracted from mycelia collected at 6.5 h of cultivation without (-) or with (+) the exogenous IM-2 addition.

binding of flavin mononucleotide (FMN), a thiolation (T) domain, and a thioesterase (TE) domain (Fig. 2.4A). Bioinformatic analysis revealed that the A-domain has the substrate recognition sequence (DAWQFGVI) for the incorporation of L-glutamine, the direct precursor of indigoidine (34) (Fig. S1). These findings strongly support the idea that LbpA is an enzyme for the bioconversion of *L*-glutamine into indigoidine.

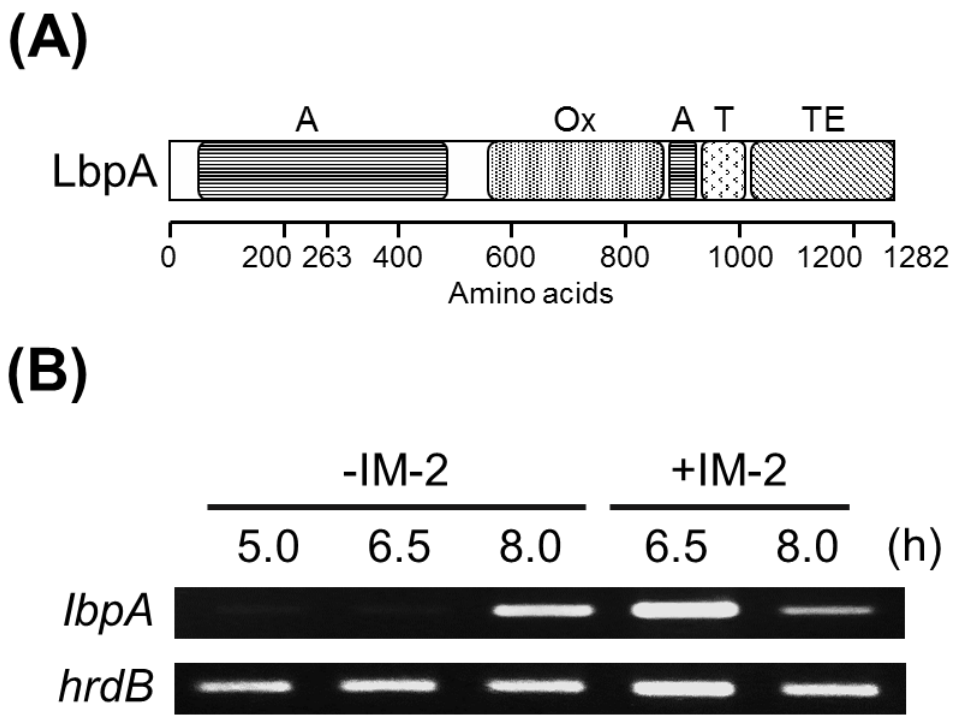


Figure 2.4 Transcriptional regulation of *lbpA* by IM-2 (A) Structural organization of LbpA. The functional domains are shown: A, adenylation; Ox, oxidation; T, thiolation; TE, thioesterase. (B) Semi-quantitative RT-PCR analysis of the *lbpA*. Total RNAs were extracted from mycelia collected at the indicated times without (-) or with (+) exogenous IM-2 addition (final concentration, 100 nM). The hrdB-like gene was used as a control, because this gene is expressed fairly constantly throughout growth.

To learn more about transcriptional regulation of *lbpA*, its transcription was analyzed in a time-dependent manner in the wild-type strain (Fig.2.4B). Under the normal growth conditions of the wild-type strain with no IM-2 addition, the transcripts of *lbpA* were clearly detected at 8 h of

cultivation, before the onset of indigoidine production (Fig. 2.2, Fig. 2.4B). An excess of IM-2 at 5 h of cultivation significantly reduced both transcriptions at 8 h cultivation and induced earlier transcription of *lbpA* at 6.5 h of cultivation (Fig. 2.4B), coinciding with the temporal manner of indigoidine production as shown in Figure 2.2. Gene disruption of *farA* led to the earlier production of indigoidine in a much-reduced amount compared to the wild-type strain (15). Consistent with the observed decrease in indigoidine production, the mRNA levels of *orfR* and *lbpA* decreased significantly in the *farA* disruptant in contrast to the wild-type strain (data not shown). This indicates that the *lbpA* transcription is tightly and temporally controlled by the IM-2/FarA regulatory system.

2.3.4 Heterologous expression of *lbpA* in *S. avermitilis* SUKA22

To confirm the in vivo function of LbpA, the plasmid pLT136 containing the entire *lbpA* gene with its promoter region was constructed and introduced into *S. avermitilis* SUKA22, an engineered host suitable for heterologous expression. The engineered *S. avermitilis* SUKA22 strain carrying pLT136 (pSET152/*lbpA*) produced small amounts of unidentified blue pigment on both ISP medium 2 and solid medium B, while no such production was observed during the cultivation of S. avermitilis SUKA22 carrying pSET152 alone (Fig. 2.5A). In addition, production of the blue pigment in *S. avermitilis* SUKA22 carrying pLT136 was observed not only in solid culture but also in liquid culture (Fig. 2.5B). It was clear from these results that the heterologous expression of LbpA confers the ability to biosynthesize the blue pigment in *S. avermitilis* SUKA22.

2.3.5 The engineered strain expressing *lbpA* produces indigoidine

The low yield of blue pigment produced by the engineered strain expressing *lbpA* hindered further characterization. In light of the similarity of LbpA to indigoidine synthetases, it is plausible

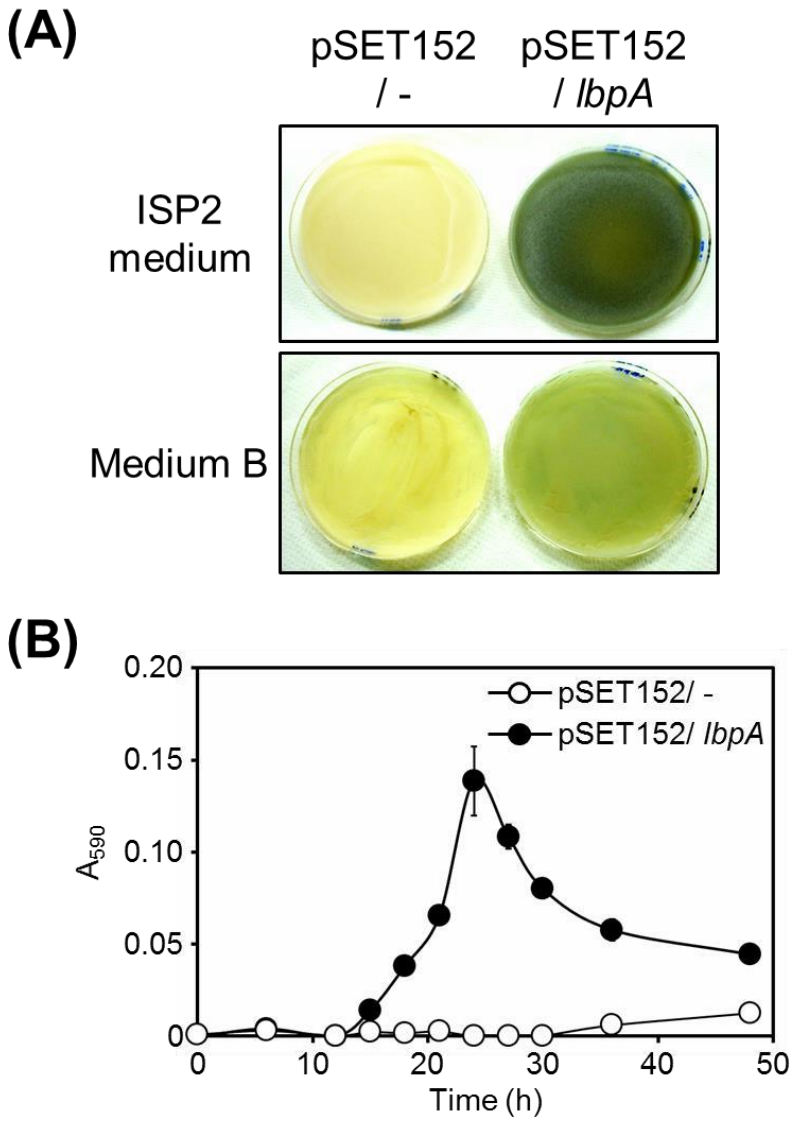


Figure 2.5 Production of blue pigment in the heterologous host harboring the *lbpA* gene. pSET152/-, *S. avermitilis* SUKA22 carrying pSET152 (vector control); pSET152/*lbpA*, *S. avermitilis* SUKA22 carrying pLT136 harboring the *lbpA* gene. (A) Each strain was cultivated on ISP medium 2 and solid medium B, and incubated for 2 days at 28°C. (B) Each strain was inoculated into liquid medium B and cultivated at 28°C. Error bars show standard deviations from triplicate experiments.

that LbpA utilizes two molecules of L-glutamine as a direct precursor. Thus, we conjectured that the addition of L-glutamine to the medium of the engineered strain would enhance blue pigment production. Indeed, we found that exogenous addition of L-glutamine led to a remarkable increase

in the production of blue pigment in a dose-dependent manner (Fig. 2.6A and 2.6B) along with a slight enhancement of cell growth (Fig. 2.7).

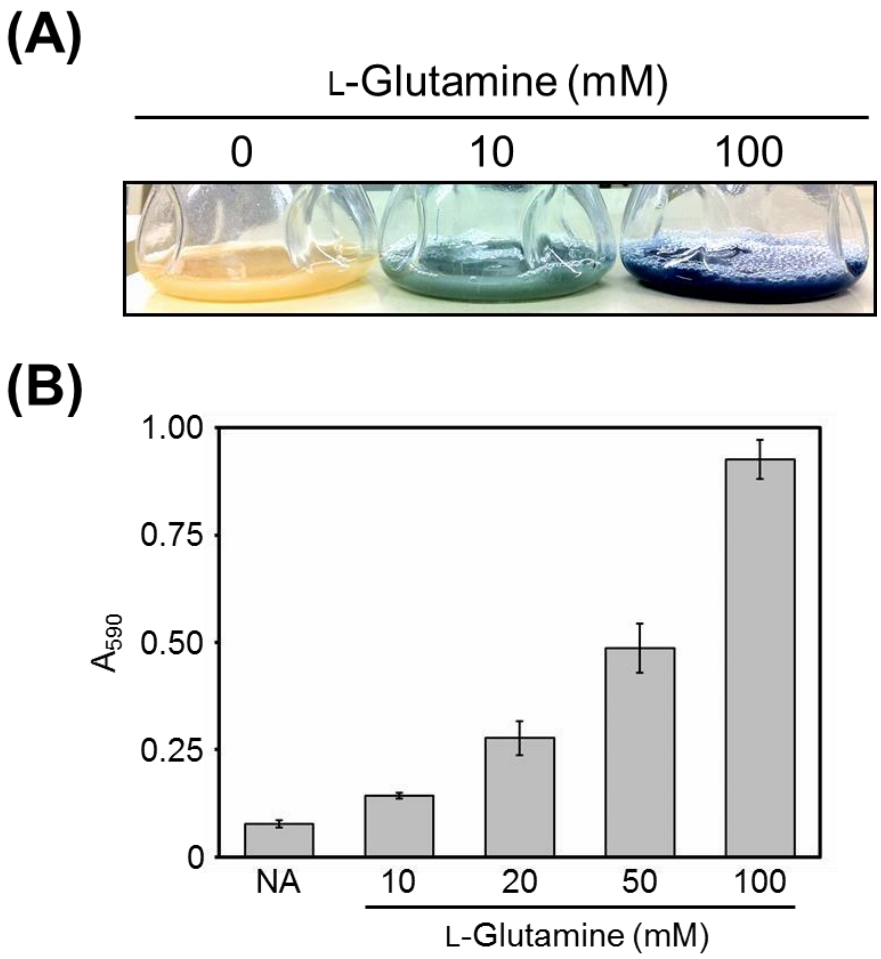


Figure 2.6 Production of blue pigment in the heterologous host harboring the *lbpA* gene. pSET152/-, *S. avermitilis* SUKA22 carrying pSET152 (vector control); pSET152/*lbpA*, *S. avermitilis* SUKA22 carrying pLT136 harboring the *lbpA* gene. (A) Each strain was cultivated on ISP medium 2 and solid medium B, and incubated for 2 days at 28ºC. (B) Each strain was inoculated into liquid medium B and cultivated at 28ºC. Error bars show standard deviations from triplicate experiments.

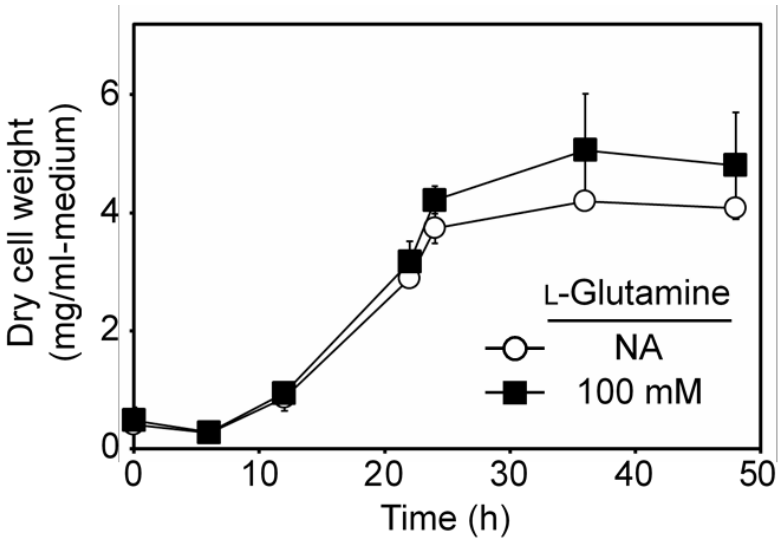


Figure 2.7 Growth curve of *S. avermitilis* harboring *lbpA* with external addition of L-glutamine. Growth curves were measured by dry cell weight. NA, no external addition of L-glutamine after 12 h of cultivation. Error bars show standard deviations from triplicate experiments.

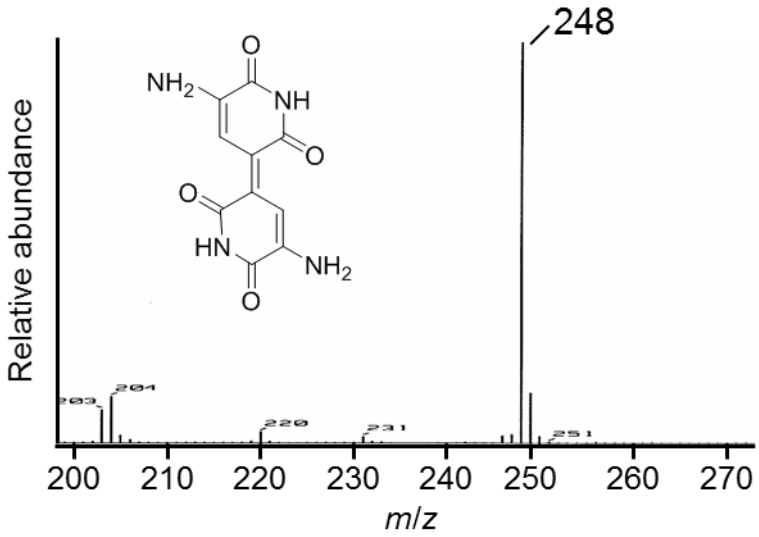


Figure 2.8 EI−MS analysis of the extracted blue pigment. The peak m/z = 248 was detected and was consistent with the molecular weight of indigoidine ($C_{10}H_8N_4O_4$).

To characterize the blue pigment, it was extracted from the culture supernatant for further spectral analyses. The purified compound in DMSO had a maximum UV absorption at 607 nm, which was identical to that of the indigoidine produced by *S. lavendulae* FRI-5 identified in our previous work. Moreover, the EI−MS spectrometry of the compound yielded a peak at 248 m/z (Fig. 2.8), corresponding to the molecular formula of indigoidine ($C_{10}H_8N_4O_4$)(Takahashi et al., 2007). These results clearly demonstrated that the engineered strain expressing *lbpA* can produce stable amounts the blue pigment indigoidine. We thus concluded that LbpA functions as an indigoidine synthetase in *S. lavendulae* FRI-5. Moreover, my results have confirmed that the genome minimized host *S. avermitilis* SUKA22 can effectively express biosynthetic genes from *S. lavendulae* FRI-5.

2.3.6 Attempts to activate the IM-2 controlled cryptic angucyline cluster

Even though the transcription of the *lac* genes is induced by the exogenous addition of IM-2 (Fig. 2.3), no angucycline compounds or aromatic polyketides were detected in the fermentation broth of *S. lavendulae* FRI-5. In order to isolate the product of the cryptic *lac* genes, I constructed a genome integrative vector carrying the entire *lac* cluster (*lacA* to *lacK*) with its putative promoter region. The resulting plasmid pKUlac was introduced by protoplast transformation into *S. avermitilis* SUKA22 to obtain strains that harbor a single copy of the *lac* cluster in its genome. The *S. avermitilis* SUKA22 strains containing plasmid pKUlac (*S. avermitilis* SUKA22/pKUlac) did not produce any new metabolite in comparison with the original host strain (*S. avermitilis* SUKA22) (Fig. 2.9), or *S. avermitilis* SUKA22 with the empty vector (data not shown). Many gene clusters for secondary metabolites contain one or more regulators that activate the expression of proximal biosynthetic genes; however no putative positive regulators were found within the *lac* cluster. Since all of the genes formed a polycistron, a constitutive promoter, *rpsJ*p, from the gene

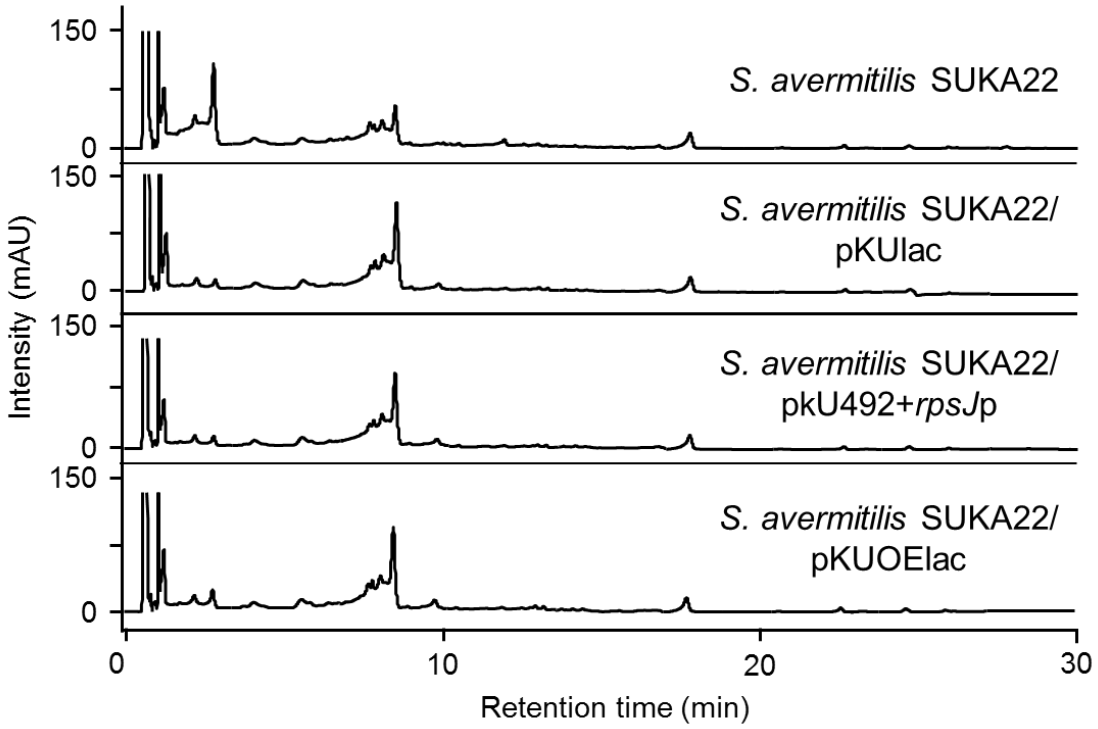


Figure 2.9 HPLC analysis of metabolite profiles from *S. avermitilis* SUKA22 carrying different vectors for *lac* cluster expression. *S. avermitilis* SUKA22, S. avermitilis SUKA22 original strain; *S. avermitilis* SUKA22/pKUlac, *S. avermitilis* SUKA22 carrying vector pKUlac, which includes the lac cluster (lacA to *lacK*) and its plausible promoter region; *S. avermitilis* SUKA22/pKU492+*rpsJ*p, *S. avermitilis* SUKA22 carrying pkU492 with the *rpsJ* promoter; *S. avermitilis* SUKA22/pKUOElac, *S. avermitilis* SUKA22 carrying vector pKUOElac, with the *lac* operon is expressed under the consitutive *rpsJ*p promoter; mAU, milliabsorbance units at 254 nm

encoding the ribosomal protein S10 (Komatsu et al., 2010) was introduced at the beginning of *lacA* to generate the *lac* overexpression vector pKUOElac. The vector pKUOElac was similarly introduced in the genome minimized host, *S. avermitilis* SUKA22. My current results showed that the generated *lac* cluster overexpression strains (*S. avermitilis* SUKA22/pKUOElac) also did not produce any detectible cluster-specific metabolite (Fig. 2.9) in several solid and liquid media.

2.3.7 Supplementing genes for angucyline ketoreductase and aromatase

In comparison to well-studied angucycline biosynthetic gene clusters such as the jadomycin cluster of *Streptomyces venezuelae* ATCC10712, the *lac* cluster lacks two essential

genes for the formation of the cyclic angucycline framework, particularly a ketoreductase for the formation of a secondary alcohol, and an aromatase that is essential for the aromatization of the first ring (Fig. 2.10A). From the extensive studies of jadomycin biosynthesis, it was established that a minimum of 6 genes (Fig. 210B) are important for the formation of the benz[a]anthracene moiety, UMW6 (Rohr & Thiericke, 1992) that is a common angucycline intermediate.

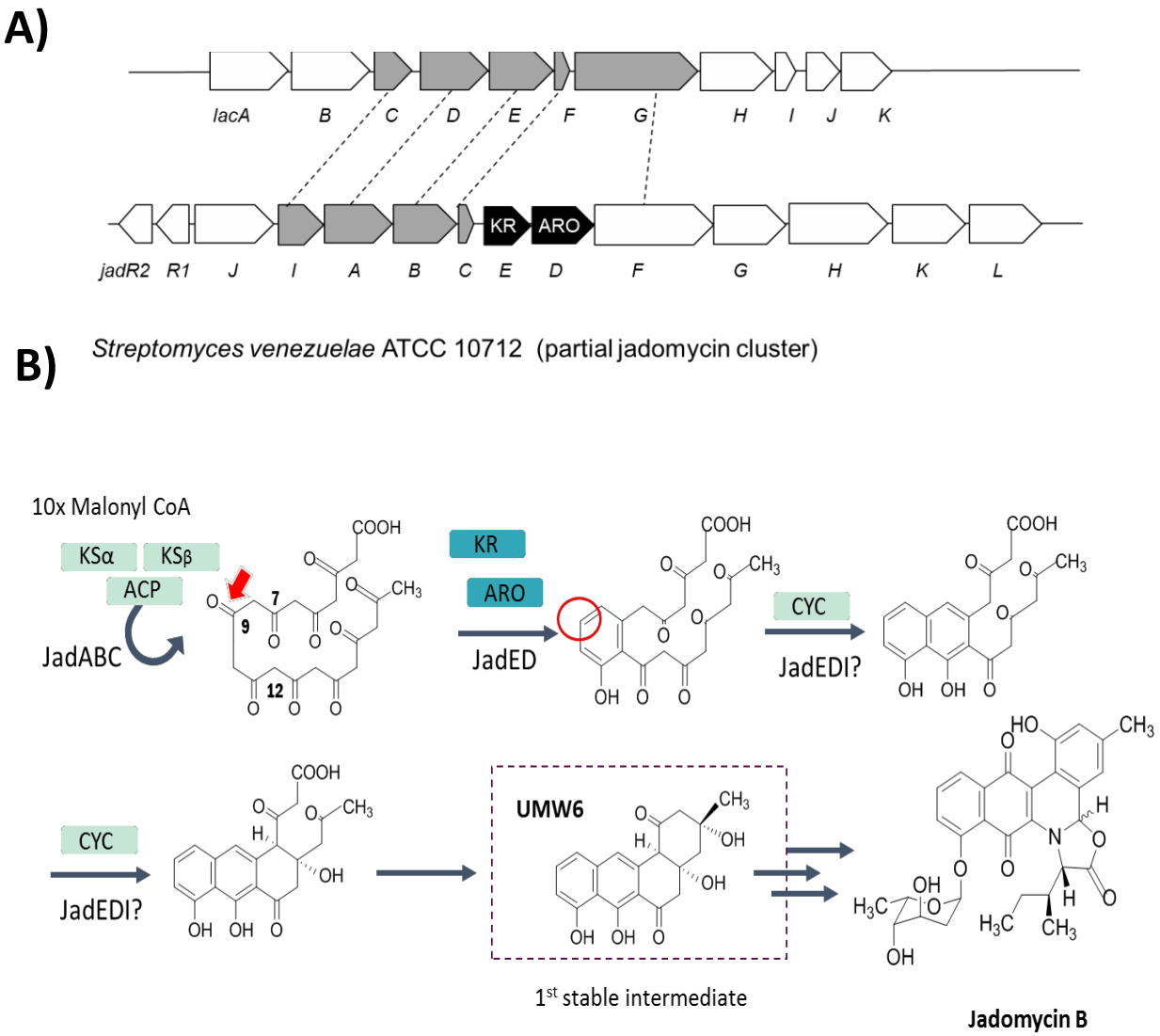


Figure 2.10 The biosynthetic gene cluster (A) and proposed biosynthetic pathway for jadomycin. (A) Comparison of the *lac* gene cluster from *Streptomyces lavendulae* FRI-5 with a portion of the jadomycin cluster from *Streptomyces venezuelae* ATCC10712. Genes that responsible for the biosynthesis UMW6 are highlighted. Missing ketoreductase (KR) and aromatase (ARO) genes are shaded in black. (B) The proposed role of the polyketide synthase genes in jadomycin biosynthesis.

To confirm whether the failure to detect any angucycline compound from the culture broth of the *lac* cluster-expression strains was due to the missing ketoreductase (KR) and aromatase (ARO) enzymes, *jadE* and *jadD* genes of the jadomycin cluster encoding a KR and ARO, respectively, were cloned and introduced into *lac* cluster overexpression strain *S. avermitilis* SUKA22/pKUOElac. The vector harboring the *jadED* operon under a constitutive promoter was integrated into a separate site of the SUKA22 genome to yield a *lac* cluster overexpression strain with missing KR and ARO genes (*S. avermitilis* SUKA22/pKUOElac+KRARO). Additionally, *S. avermitilis* SUKA strain overexpressing the minimum angucycline genes (*lacC to lacF)* was also constructed and the resulting transformant (*S. avermitilis* SUKA/pKUOEmin) was similarly supplemented with *jadED* to generate *S. avermitilis* SUKA/pKUOEmin+KRARO.

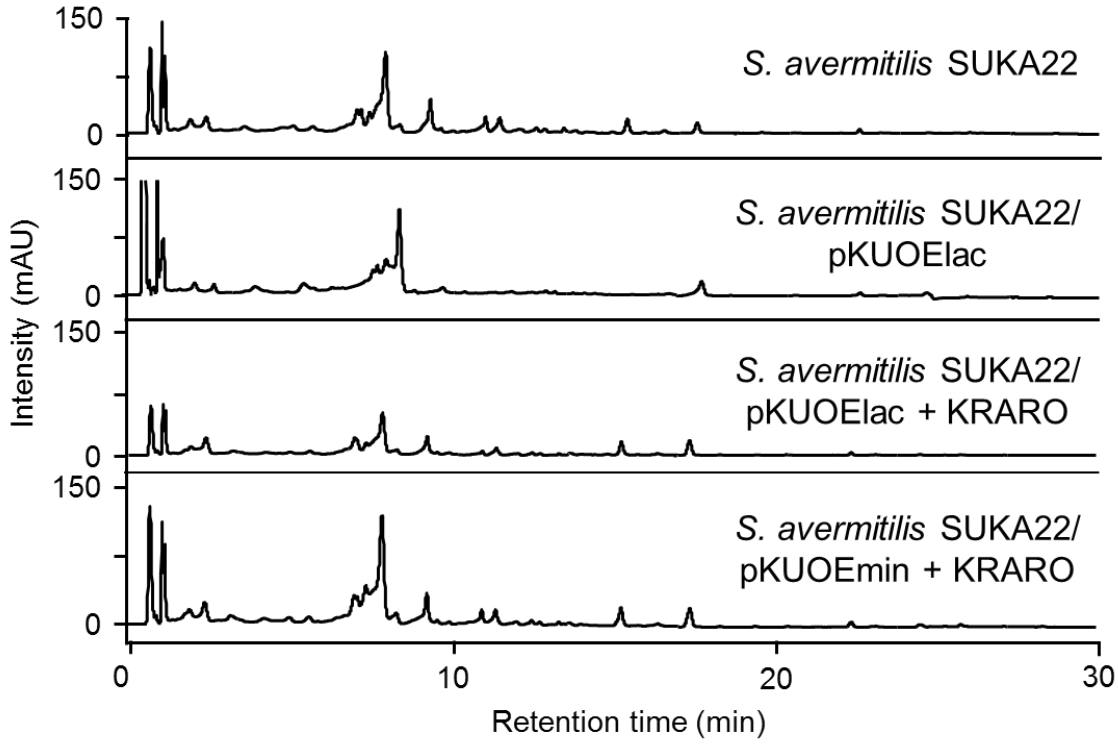


Figure 2.12 HPLC analysis of metabolite profiles from *S. avermitilis* SUKA22 *lac*-overexpression strains supplemented with missing KR and ARO genes. *S. avermitilis* SUKA22, *S. avermitilis* SUKA22 original strain; *S. avermitilis* SUKA22/pKUOElac, *S. avermitilis* SUKA22 carrying vector pKUOElac; *S. avermitilis* SUKA22/pKUOElac+KRARO, *S. avermitilis* SUKA22 carrying

vector pKUOElac with *jadED*; *S. avermitilis* SUKA22/pKUOEmin+KRARO, *S. avermitilis* SUKA22 carrying vector pKUOEmin with *jadED;* mAU, milliabsorbance units at 254 nm.

The *S. avermitilis* SUKA22 strains overexpressing either the complete *lac* cluster or the minimum *lac* genes did not produce any apparent metabolite even after the introduction of additional KR and ARO genes from the jadomycin cluster (Fig 2.12). I also tried to cultivate the *lac* cluster-expression strains in various liquid and solid media but our current laboratory conditions failed to reveal the product of the cryptic *lac* cluster.

2.4. Discussion

The streptomycetes genome has secondary metabolite biosynthetic gene clusters that are frequently accompanied by regulatory genes for controlling the biosynthesis of these secondary metabolites (Bibb, 2005; Martín & Liras, 2010). We previously showed that the *far* regulatory genes play complicated roles in the production of the blue pigment indigoidine, nucleoside antibiotics, and D-cycloserine as secondary metabolites of *S. lavendulae* FRI-5 (S. Kitani et al., 2001; Shigeru Kitani et al., 2010; Kurniawan et al., 2016, 2014). Therefore, we hypothesized that the adjacent regions of the *far* regulatory island might contain biosynthetic genes of these secondary metabolites. In the present study, we have demonstrated that several putative biosynthetic genes under the transcriptional regulation of IM-2 are localized in the left-hand region of the *far* regulatory island. Of these, *lbpA* encodes an indigoidine biosynthetic NRPS enzyme utilizing L-glutamine as a precursor. With respect to indigoidine production, three types of transcriptional regulators in the *far* regulatory island are involved at several regulation layers in the IM-2/FarA system. The IM-2 receptor FarA is a negative regulator for indigoidine biosynthesis before IM-2 production and controls the amount of indigoidine (S. Kitani et al., 2001). In the second layer of regulation, FarR2, a FarA homologue regarded as a pseudoreceptor, serves as a

positive regulator for the onset of indigoidine biosynthesis in response to IM-2 (Kurniawan et al., 2016), while FarR3, a SARP-family regulator, positively controls both the onset of indigoidine biosynthesis and the amount of indigoidine (Kurniawan et al., 2014). However, epistatic analyses indicated that FarR2 and FarR3 independently control the onset of indigoidine biosynthesis. The upstream region of *lbpA* has neither a plausible binding site of FarR2, which recognizes ARE-like sequences, nor a typical sequence for the binding of the SARP family regulators, which are heptameric sequences separated by 4 nt or 15 nt. This suggests that both types of regulators (FarR2 and FarR3) operate via another regulator that controls the biosynthesis of indigoidine positively. The identification of factors that bind to the *lbpA* upstream region and control indigoidine biosynthesis would help to decipher the complicated regulatory mechanism underlying indigoidine biosynthesis governed by the IM-2/FarA system.

Indigoidine has been found in several different bacteria, such as *Erwinia chrysanthemi* and *S. lavendulae* ATCC11924 (Reverchon, Rouanet, Expert, & Nasser, 2002; Takahashi et al., 2007). It naturally functions as a powerful radical scavenger that enables phytopathogens to tolerate oxidative stress during plant defense response, and thus is a strong antioxidant. Recent efforts have also extended the utility of indigoidine to various fields. For instance, indigoidine exerts antimicrobial activity against the pathogenic marine bacterium *Vibrio fischeri*, and it was recently developed into a versatile reporter of gene expression for bacteria and mammalian cells, as well as a chromogenic reporter for colony screening in *Streptomyces* (Cude et al., 2012; P. Li et al., 2015; Müller, Ausländer, Ausländer, Kemmer, & Fussenegger, 2012). Further, indigoidine is seen as a viable natural substitute for industrial blue dyes. Microbial pigments are a safer alternative to chemically synthesized dyes, and could be more economical than natural pigments from plants and animals. Thus, expanding the potential application of indigoidine requires an efficient and

stable method to produce it on a large scale. We recently verified that the blue pigment produced by *S. lavendulae* FRI-5 is indigoidine (Kurniawan et al., 2014). However, its production is unstable under normal conditions because endogenous IM-2 biosynthesis easily responds to slight environmental changes during cultivation. While the external and excess addition of chemically synthesized IM-2 into medium leads to more stable and higher indigoidine titers (S. Kitani et al., 2001; Shigeru Kitani et al., 2010), the complexity of the regulatory network for indigoidine production, as detailed in our previous works, has hampered genetic manipulations to improve the indigoidine yield of the original producer. In order to achieve high production levels of indigoidine independent of cultivation conditions, we constructed an engineered strain expressing *lbpA* by using *S. avermitilis* SUKA22, which is a suitable host for heterologous production of exogenous secondary metabolites. An NRPS is initially formed as an inactive apoform and must be posttranslationally modified for activity by phosphopantetheinyl transferases (PPTases). These enzymes transfer the cofactor 4'-phosphopantetheine from cognate CoA to a highly conserved serine residue in the T-domain to make it functional (L. Du et al., 2001). Introducing only *lbpA* successfully led to indigoidine production suggesting that the 4-phosphopantetheinylation of LbpA from *S. lavendulae* FRI-5 was achieved by endogenous PPTase activity of *S. avermitilis*. The feeding of indigoidine precursor to the engineered strain resulted in enhanced levels of indigoidine compared to those of the original producer with no addition of IM-2. This demonstrated that the combination of both a heterologous expression system using *S. avermitilis* SUKA22 and precursor feeding in the engineered strains is a simple and effective approach to achieve high-level and stable production of useful actinomycete natural compounds. Heterologous expression in SUKA22 is especially useful in improving yields of compounds with associated biosynthetic genes, and whose production in the native producer is tightly regulated and highly dependent on unidentified

environmental signals. On the contrary, without prior knowledge of the biosynthetic gene or complete biosynthetic gene cluster for a target compound or if the substrate specificity of the target biosynthetic enzyme cannot be predicted only by *in silico* analysis, this approach may not necessarily be suitable for increasing the yield of actinomycete secondary metabolites.

The region between *lbpA* and the *far* regulatory genes contains potential biosynthetic genes for secondary metabolites—namely, the *lac* cluster for potential biosynthesis of a putative aromatic polyketide, and putative biosynthetic genes (*orf1*−*orf9*) for potential biosynthesis of unidentified compounds. The *lac* cluster includes the minimal PKS genes (*lacD*, *lacE*, and *lacF*) for the generation of linear poly-β-keto thioester linked poly-β-ketones, and it shows high similarity to a part of the *aur1* cluster for the angucycline antibiotic auricin in *S. aureofaciens* CCM 3239. However, unlike the typical type II PKS gene cluster, the *lac* cluster lacks genes encoding a ketoreductase for position C-9 of the polyketide intermediate and an aromatase for the aromatization of the first ring in the aromatic polyketides. These biosynthetic enzymes are important for the maturation of the intermediate. Therefore, it would appear to be difficult to predict the final natural product biosynthesized by the *lac* cluster using *in silico* analysis alone. We attempted to characterize the product of the cryptic *lac* cluster by heterologous expression in a genome minimized host. Unfortunately, the constitutive expression of the biosynthetic genes alone did not result into compound detection and isolation under our cultivation conditions. One regulatory gene *lacB* was found to have sequence homology with *aur1B*, which contains a DNA-binding motif of the TetR family of transcriptional repressors. The exact role of *aur1B* in auricin biosynthesis has not yet been clarified but the disruption of its homolog, *pgaY*, resulted in the activation of 2 cryptic angucycline metabolites (Metsä-Ketelä et al., 2004). Taken together, I considered that the lack of enzymes for ring formation and tight repression might explain why no

angucycline compound was detected in the laboratory cultures of *S. lavendulae* FRI-5. However, separately cloning and overexpressing the minimum *lac* genes (*lacC* to *lacF*) and introducing ketoreductase and aromatase functions in the heterologous host did not yield any stable angucycline intermediate. Our current results suggests that the *lac* cluster's inactivity is due to reasons other than the absence of essential angucycline genes or transcriptional repression. Based on sequence alignment, there was no clear evidence that the PKS genes were nonfunctional. Our failure to detect the product of the *lac* cluster even after heterologous expression in another *Streptomyces* host (data not shown) specific for angucylines (Palmu et al., 2007), strongly indicates that the *lac* genes may be genetically inactive or aberrant. This result shows that despite the abundance of uncharacterized BGCs in the genome of Streptomyces, not all may be activated by heterologous expression or not all may be functional.

2.5. Summary

In the absence of a complete genome sequence, we suggested that the autoregulator receptor gene can be used as a probe to fish for biosynthetic gene clusters. Here, I have successfully found several gene clusters in the vicinity of the IM-2 receptor *farA*, including the indigoidine biosynthetic gene *lbpA* and the cryptic angucycline cluster (*lac* cluster). I also showed that these biosynthetic genes are newly identified members of the IM-2 regulon, and confirmed that LbpA is the indigoidine synthase of the IM-2 signalling cascade. These findings can be used to further clarify the regulatory mechanism governed by the IM-2/FarA system. Furthermore, heterologous expression of *lbpA* in the versatile model host *Streptomyces avermitilis* SUKA22 led to indigoidine production, which was enhanced dramatically by feeding of the indigoidine precursor L-glutamine. This demonstrated that the heterologous expression system using *S. avermitilis* SUKA22 and

precursor feeding in the engineered strains can achieve high-level and stable production, which is essential for the detection and isolation of cryptic compounds.

Unfortunately, the product of the *lac* cluster was not detected even after overexpression, addition of essential genes (*jadE* and *jadD*) from the jadomycin pathway, and use of an alternative heterologous host. This inability to isolate the corresponding *lac* cluster metabolite is likely caused by genetically aberrant genes and it illustrates the impending difficulty in activating cryptic BGCs. Nevertheless, as occurred in the successful heterologous production of indigoidine in the present study, the heterologous expression of other cryptic genes in the versatile model host *S. avermitilis* SUKA22, sometimes with a precursor supply (like the L-glutamine addition in the indigoidine production) and genetic supplementation (for example, co-expression of exogenous ketoreductase and/or aromatase genes) may allow the discovery of novel bioactive compounds potentially hidden within the genome of *S. lavendulae* FRI-5.

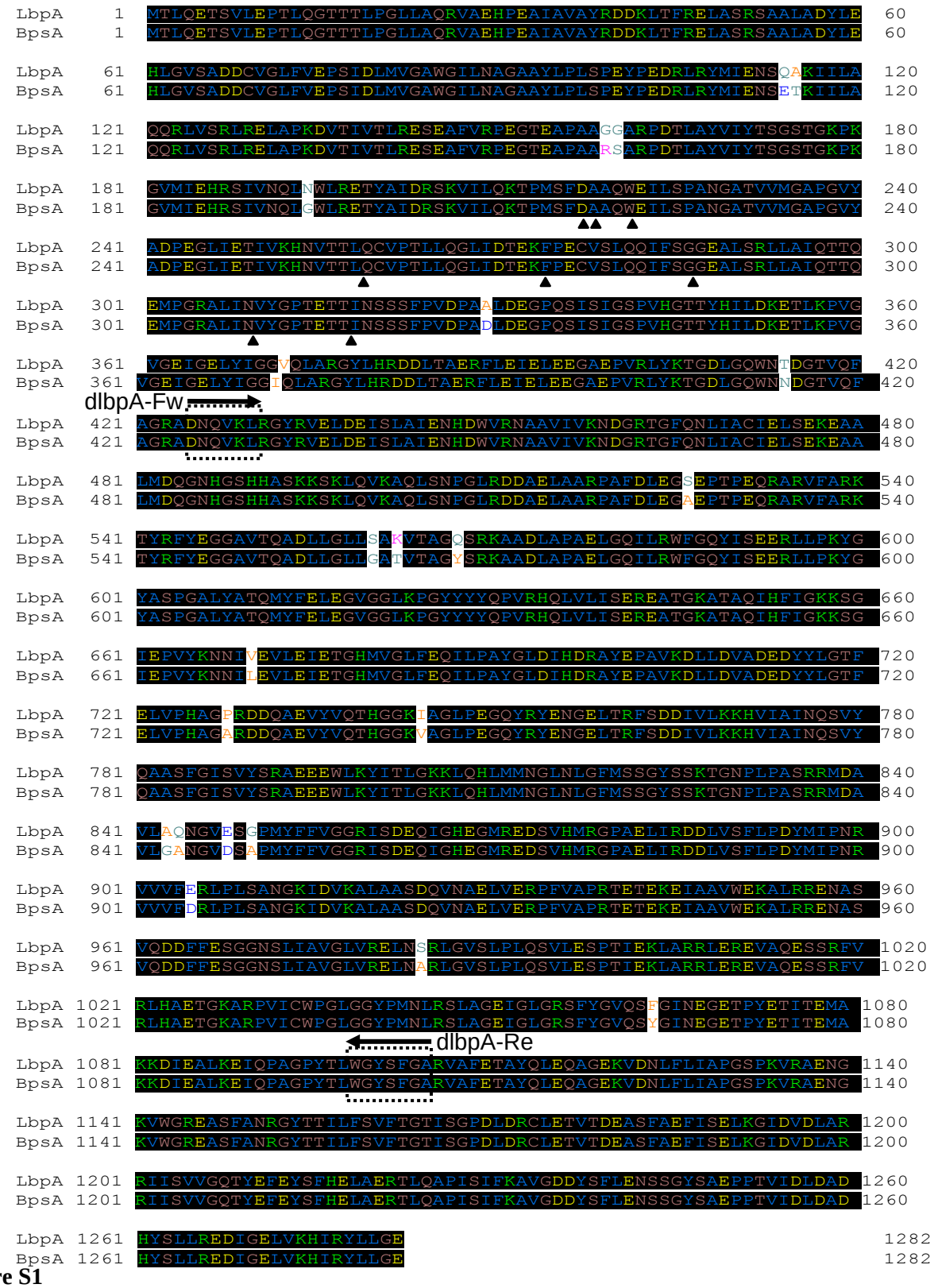


Figure S1

Sequence alignment of *S. lavendulae* FRI-5 LbpA with *S. lavendulae* ATCC11924 BpsA. Identical amino acid residues are highlighted in black. The signature residues for recognition of L-glutamine are denoted by triangles. Each arrow shows the region of designed degenerate primers (dlbpA-Fw and dlbpA-Re), which encode the amino acids indicated by dashed boxes.

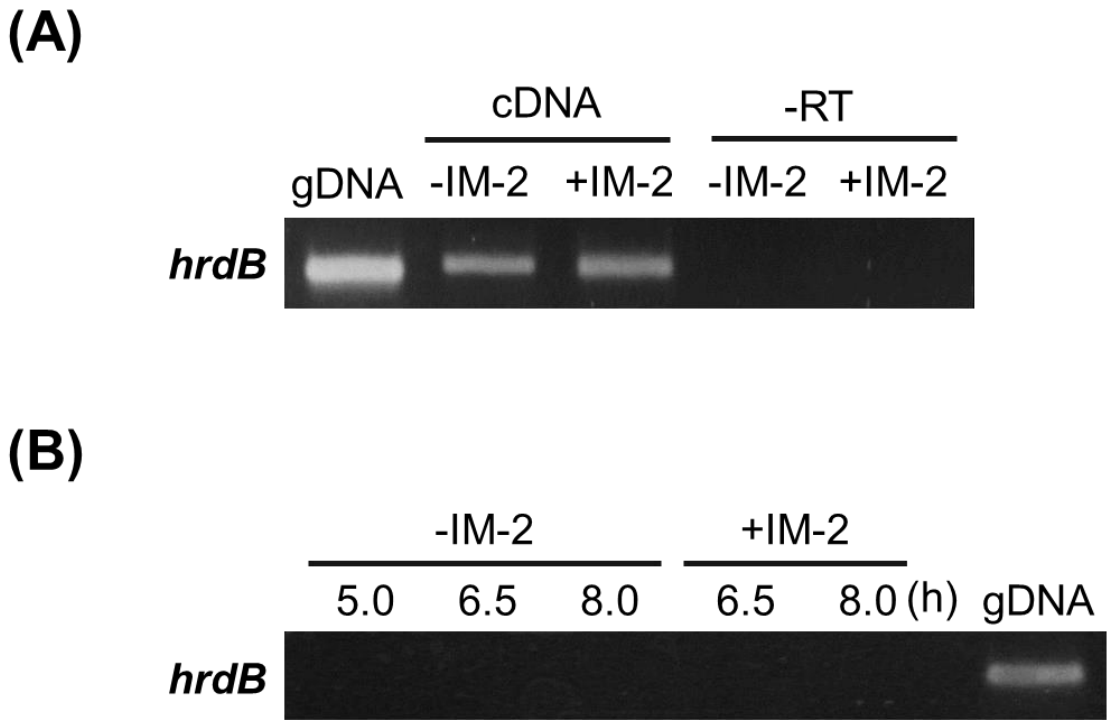


Figure S2
RT-PCR detection of *hrdB* transcripts to confirm the absence of DNA contamination in RNA samples for Fig. 2.3 and Fig. 2.4b, respectively. (A) Lane 1 shows the PCR amplification of genomic DNA as positive control; Lanes 2 and 3 are transcripts detected at growth conditions without and with IM-2 addition; Lanes 4 and 5 shows negative control without RT (B) Lanes 1-5 shows negative control without RT; Lane 6 shows the PCR amplification of genomic DNA as positive control.

Chapter 3

Discovery of the novel compound lavendiol through the awakening of a silent, uncharacterized PKS gene cluster

3.1 Introduction

Natural products from microorganisms are structurally complex and have diverse biological activities, making them indispensable building blocks for the biosynthesis of therapeutic compounds, pesticides, and anti-parasitic agents (Koehn & Carter, 2005; Newman & Cragg, 2016; Omura, 2011). Traditional bioactivity-guided screenings have been plagued by the frequent rediscovery of already known compounds. Nonetheless, current rapid developments in microbial genome analysis have unveiled a plethora of untapped biosynthetic gene clusters for promising bioactive compounds (Bachmann et al., 2014; Gregory L. Challis, 2008). However, a considerable portion of these gene clusters are expressed either way below detectable levels under standard laboratory cultivation conditions or not at all, indicating that many novel and potentially useful compounds encoded by these "silent" biosynthetic pathways still await discovery (Ochi & Hosaka, 2013; Scherlach & Hertweck, 2009; Zarins-Tutt et al., 2016).

Members of the genus *Streptomyces* are the most prolific microbial producers of bioactive compounds as secondary metabolites that are used in human/veterinary medicine and agriculture (Bérdy, 2012). A comparison of sequenced bacterial genomes has reinforced this notion, revealing that large genomes of *Streptomyces* are richest in terms of biosynthetic capacity, with around 20-40 biosynthetic gene clusters of secondary metabolites per strain (Baltz, 2016). Several strategies have been employed recently to induce the expression of these silent gene clusters, allowing the discovery of novel compounds. For example, in the OSMAC (one strain/many compounds) approach, the systematic change of readily accessible cultivation parameters leads to the identification of diverse bioactive compounds (Bode et al., 2002; Zarins-Tutt et al., 2016; Zhang

et al., 2014). Combined cultures with other microorganisms to try to mimic natural interactions and modification of the transcriptional/translational machinery have also been explored to discover new natural products (Onaka et al., 2015; Tanaka et al., 2013; Tanaka, Tokuyama, & Ochi, 2009; Traxler et al., 2013) . In addition, more targeted, genome-guided manipulation of pathway-specific regulators in a silent gene cluster has enabled the production of new biologically active compounds (D. Du et al., 2016; Laureti et al., 2011). Another effective strategy is to use a versatile host for heterologous expression (Yamada, Arima, et al., 2015). The advantage of this strategy is that it allows us to identify silent bioactive compounds when genetic manipulation in the original strain is difficult or when plausible pathway-specific regulators are absent within the silent gene cluster.

Streptomyces lavendulae FRI-5 predominantly produces four secondary metabolites: blue pigment indigoidine, anti-tuberculosis drug D-cycloserine, and nucleoside antibiotics [showdomycin and minimycin] (Fig. 1.13) (Hashimoto, Nihira, et al., 1992). As explained in the previous chapter, we characterized the indigoidine biosynthetic gene *lbpA* using heterologous expression method with the genome-minimized *Streptomyces avermitilis* SUKA22 (Pait et al., 2017). *S. avermitilis* SUKA22 is an engineered host that is suitable for heterologous expression because it is incapable of producing most of its endogenous secondary metabolites and because it provides an excellent clean background for expression of biosynthetic genes for exogenous secondary metabolites (Komatsu et al., 2010; Yamada, Kuzuyama, et al., 2015). Similar to the genomes of other *Streptomyces* strains, that of *S. lavendulae* FRI-5 is likely to include numerous silent gene clusters with the potential to direct the production of yet-undiscovered natural products. Thus, heterologous expression of silent, uncharacterized gene clusters from *S. lavendulae* FRI-5 into the "clean" host *S. avermitilis* SUKA22 may reveal products associated with intriguing biosynthetic pathways identified by bioinformatics.

In this work, I focused on a gene cluster encoding a type I modular polyketide synthase (PKS) of *S. lavendulae* FRI-5. Polyketide compounds constitute an important group of structurally diverse secondary metabolites that have great pharmacological value (J Staunton & Weissman, 2001). In type I PKS systems, polyketide compounds are assembled from acyl-CoA building blocks in a stepwise fashion by large, multifunctional enzymes with various catalytic domains, and the actions of tailoring enzymes during or after polyketide chain elongation diversify the resulting aglycon (Fischbach & Walsh, 2006). Here I report the isolation and structural elucidation of lavendiol, a new diol-containing polyketide resulting from the activation of an uncharacterized type I PKS gene cluster by heterologous expression. I also proposed the lavendiol assembly line, which has some unprecedented biosynthetic features.

3.2. Materials and Methods

3.2.1 Bacterial strains, plasmids, and growth conditions

Streptomyces lavendulae FRI-5 (MAFF10-06015; National Food Research Institute, Tsukuba, Japan) was grown as previously described for spore formation (S. Kitani et al., 2001). *Streptomyces avermitilis* SUKA22 (Yamada, Kuzuyama, et al., 2015) was used as the host for heterologous expression of the *lav* cluster, and was grown on YMS-MC medium for spore formation. *Escherichia coli* DH5α was used for general DNA manipulation, *E. coli* BW25141 containing pKD119 was used for λ Red-mediated recombination, *E. coli* GM2929 *hsdS*::Tn*10* was used to prepare unmethylated cosmid DNAs for transformation into *S. avermitilis* SUKA22, and *E. coli* F^- *dcm* Δ(*srl-recA*)*306*::Tn*10* carrying pUB307-*aph*::Tn*7* was used for *E.coli/Streptomyces* conjugation (Komatsu et al., 2008). pBluescript II KS was used for general cloning, pKU250 and pKU475 (Komatsu et al., 2010) for gene disruption, and a small pRED (Komatsu et al., 2010) vector for *in vivo* cloning of 9 genes (*orf4* to *orf12*). Spores (1.0×10^8 CFU) of the *S. avermitilis*

strains were inoculated into 70 mL of APM (Kitani et al. 2009) medium in a 500-ml baffled flask, and mycelia were harvested after 40 h of cultivation at 28ºC. The mycelia were washed, resuspended in fresh APM medium and stored at -80ºC until use as a seed culture. The media conditions and general *E. coli* and *Streptomyces* manipulations were as described previously (Kieser et al., 2000). The primers used in this work are listed in Table 3.1.

3.2.2 Genome sequencing and bioinformatics analyses

Genomic DNA sequencing was performed using Illumina Genome Analyzer GAII. The sequencing data were assembled using Velvet assembly software to give 5,190 contigs with a mean length of 1,525 bp. The assembled contig sequences and the Genome Matcher software (Ohtsubo, Ikeda-Ohtsubo, Nagata, & Tsuda, 2008) were used to screen for potential type I PKS gene clusters in the genome. Open reading frames (ORFs) and gene functions were annotated manually by using the FramePlot 4.0beta program (http://nocardia.nih.go.jp/fp4/), 2ndFind program (http://biosyn.nih.go.jp/2ndfind/), the BLAST algorithm, and the PKS/NRPS analysis online tool (http://nrps.igs.umaryland.edu/nrps/). The nucleotide sequence data reported in this paper have been deposited in the DDBJ data bank under accession number LC330869.

3.2.3 Construction and screening of cosmid library of genomic DNA

Genomic DNA of *S. lavendulae* FRI-5 was partially digested with *Sau*3AI, and 30-40 kb DNA fragments were ligated into *Xba*I/*Bam*HI-digested and dephosphorylated SuperCos 1. *E. coli* XL-1 Blue MR and Gigapack III plus packaging extract (both from Agilent Technologies) were used for library construction according to the manufacturer's instructions. The cosmid library

Table 3.1 Oligonucleotides used in this study.

Primer	Sequence (5' - 3')
For transcriptional analysis of the *lav* cluster	
orf21-Fw	GGAGGTGACGGTGATGATCG
orf21-Rv	GACGGGATACTCCACCGAAC
orf22-Fw	TCGCTGTGCTCCTGGTACTC
orf22-Rv	GAACAGCCGGTGACAGAAGC
lavA-Fw	GACACGCGATATTCGGCCAG
lavA-Rv	GTTCACTGCTGCTGGTCAGC
For heterologous expression of the *lav* cluster	
pRED-orf12-Fw	GCTCCGGTTTCCCGTCCCTGCTCCCGGGCGATCCGGAAA AGGTCTAGATGCCAGGAAGATACTTAACAG
pRED-orf4-Re	CCCTAGGCCGTTTCTTTCAGATCATCTGATCGTTGGTCT GGTCTGCAGCATTCATCCGCTTATTATC
hph-Fw	CGGGGTACCCAAGCGGTCAAACATGAGAATTCGCGGCCG CATAATACGACTCCCGTTGGATACACCAAGGAAAGT
hph-Re	CCCAAGCTTCTTCATGAGCTCAGCCAATCGACTGGCGAG
For construction of *lav* disruptants	
dislavA-Fw	CCGGAATTCGGCGTGAACCTCAACCTCCTC
dislavA-Re	CCGGAATTCCCTCCACGGCGATCATGGCT
dislavB-Fw	CCGGAATTCGCGTGTACGACTCCGGTCAG
dislavB-Re	CCGGAATTCCTGAGGCTCGGGTCGTTCAA
tsr-Fw	ATGACTGAGTTGGACACCATCGCA
tsr-Re	TTATCGGTTGGCCGCGAGATTCCT
pKU250-bla	GCTGCCAGTGGCGATAAGTC
lavA-dw	GAAGGCGTGGCTGACGGTCA
lavB-up	GTACGGCGACCAACCAGGAC
lavB-dw	CTCGATGTCGAAGGCCGTGG

*Restriction sites are underlined.

(5,000 colonies) was screened by PCR using primers designed from the assembled contig sequence data, yielding two overlapping cosmid clones (DD60 and AA94). The gaps between neighboring contigs were closed by cosmid-based PCR amplification and sequencing.

3.2.4 Gene expression analysis of the newly identified *lav* genes

Total RNAs were extracted from mycelia harvested at the indicated cultivation times by an RNeasy Mini Kit (QIAGEN Sciences, Germantown, MD, USA) and treated with DNase I (Takara Bio, Shiga, Japan). The complementary DNA (cDNA) was synthesized using SuperScript III Rnase H^- Reverse Transcriptase and random primers (Invitrogen, Carlsbad, CA, USA). For semiquantitative RT-PCR, PCR amplification was carried out under the following conditions: a single round of denaturation at 95°C for 2 min and 30 cycles of 98°C for 30 s, 57°C for 30 s, and 72°C for 1 min, followed by a single extension at 72°C for 5 min. The absence of DNA contamination was confirmed by RT-PCR without reverse transcriptase.

3.2.5 Heterologous expression of the *lav* cluster in *S. avermitilis* SUKA22

The construction of the refactored cosmid cDD60α is depicted in Figure 3.2. The *neo* gene in the Supercos 1 backbone of cosmid DD60 was replaced with a 5.2 kb *Ssp*I fragment from pIJ10702 (Yanai, Murakami, & Bibb, 2006), which contains an apramycin resistance gene (*aac3(IV)*) and a ΦC31 integration cassette by PCR targeting, yielding the integrative cosmid cDD60. For *in vivo* cloning of the 9 genes (*orf4* to *orf12*), the pRED vector was PCR-amplified by the primer pair (pRED-orf12-Fw/pRED-orf4-Re). The PCR product was treated with *Dpn*I to remove template DNA and cotransformed into L-arabinose−induced *E. coli* BW25141 carrying pKD119 with the cosmid AA94, to generate pLT161. An 8.2 kb *Xba*I/*Pst*I fragment recovered from pLT161 was inserted into the *Xba*I and *Pst*I sites of pBluescript II KS to yield pLT162. pLT162 was digested with *Kpn*I and *Hin*dIII, and was ligated together with a fragment containing a hygromycin resistance gene (*hph*) with a 42 bp sequence identical to the SuperCos 1 sequence (PCR-amplified with pKU475 and the hph-Fw and hph-Re primer pair), resulting in pLT163. An 10.1 kb *Kpn*I/*Xba*I fragment, recovered from pLT163, was introduced into the integrative cosmid cDD60 by λ Red-

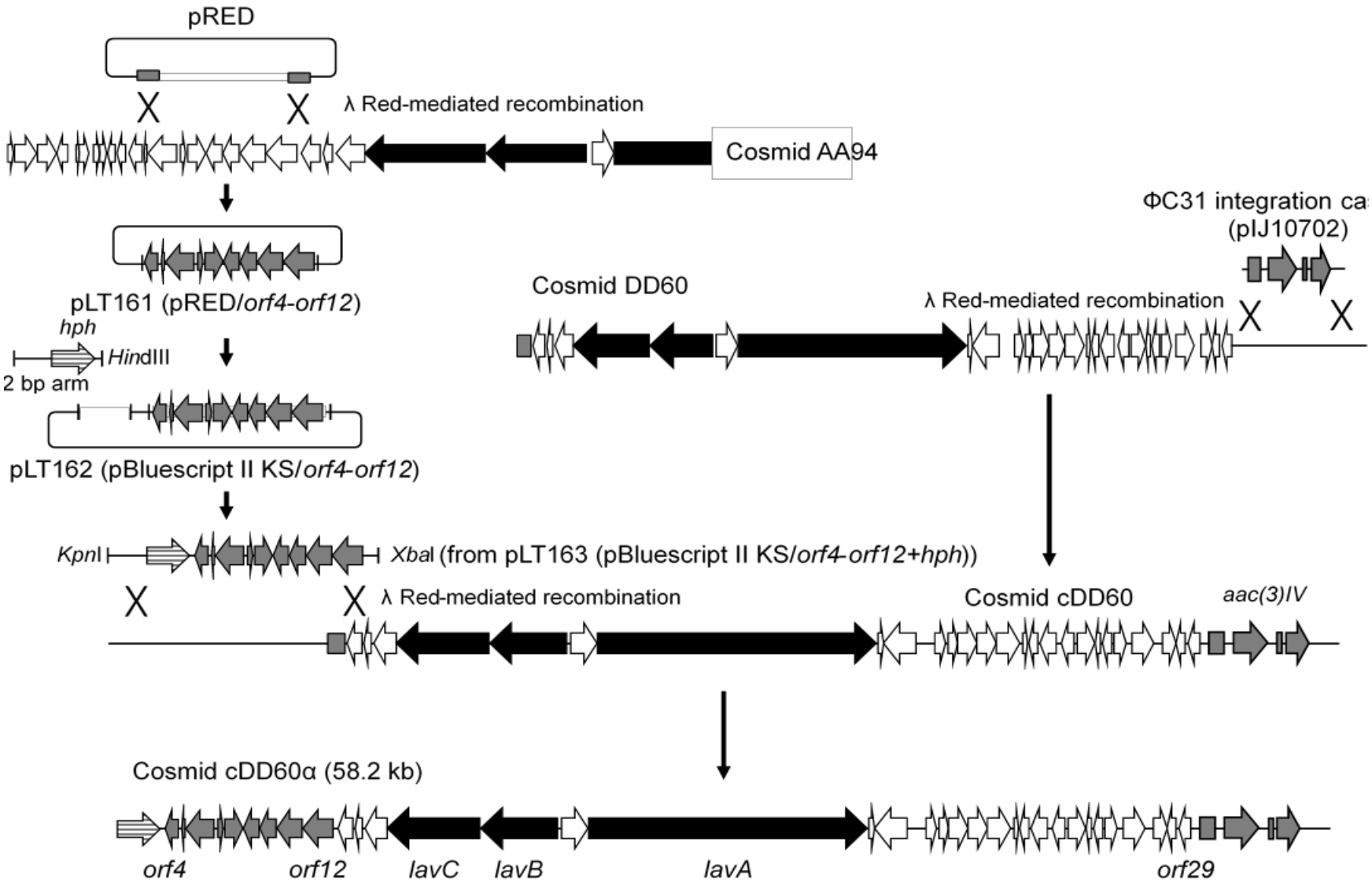


Figure 3.1 Strategy for the construction of cosmid cDD60α.

mediated recombination, and the resultant cosmid was designated as cosmid cDD60α. After demethylation by *E. coli* GM2929 *hsdS*::Tn*10*, cosmid cDD60 or cDD60α was introduced into *S. avermitilis* SUKA22 by protoplast transformation, followed by apramycin or hygromycin selection and PCR analysis.

3.2.6 Detection of secondary metabolites in the heterologous host

The seed cultures of the *S. avermitilis* strains were inoculated into 70 mL of synthetic production medium (Cane et al., 2006) supplemented with 25 mM L-glutamic acid in a 500-ml baffled flask, and incubated on a rotary shaker (160 rpm) for 4 days at 28ºC. The culture broth was extracted with a half volume of *n*-butanol, and the *n*-butanol layer was evaporated and dissolved in dimethyl sulfoxide. The *n*-butanol extract was analyzed by reversed-phase HPLC on a Capcell-Pak C_{18} column (UG80; 5 μm; 4.6 i.d. x 250 mm; Shiseido, Tokyo) developed with a gradient system of CH_3CN (10% for 0-3 min; 10-100% for 3-32 min; 100% for 32-35 min) containing 0.1% formic acid (flow rate, 1.2 mL/min; UV detection, 290 nm). The culture broth of *S. lavendulae* FRI-5 was prepared and analyzed in the same way as that of the *S. avermitili*s strains.

3.2.7 Isolation and structural elucidation of compound 1

Culture broth (5 L) of *S. avermitilis* SUKA22 carrying cosmid cDD60α was extracted with 2 volumes of ethyl acetate, after which the ethyl acetate layer was evaporated to dryness. The crude extract (2.2 g) was fractionated by silica gel column chromatography eluted with a stepwise gradient of hexane/ethyl acetate (3:1, 1:3, 1:4, and 0:1 v/v). Fraction 2 was concentrated to give 900 mg of yellowish oil, which was further purified by preparative reversed-phase HPLC using a Capcell-Pak C_{18} column (UG80; 5 μm; 10 i.d. x 250 mm; Shiseido) developed with 25% CH_3CN containing 0.1% formic acid at 3 ml/min followed by detection at 290 nm, yielding 15.7 mg of

purified compound **1**. The UV spectrum was recorded on a Hitachi U-3210 spectrophotometer, and the IR spectrum was recorded on an FTIR-8400 S (Shimadzu, Kyoto, Japan). Optical rotation was measured on a P-2200 polarimeter (JASCO, Tokyo). HR-ESI-TOF-MS was measured on a Waters Q-TOF Ultima API mass spectrometer. NMR (^{1}H, 600 MHz; ^{13}C, 150 MHz) spectra were recorded on a Bruker UltraShield 600 Plus spectrometer, and the ^{1}H and ^{13}C chemical shifts were referenced to the solvent signal ([MeOH]-d_4: δ_C 49.1, δ_H 3.31).

3.2.8 Physicochemical properties of lavendiol

Yellow amorphous solid; UV (MeOH) λ_{max} (logε) 217 (3.79), 318 (4.49) nm; IR (film) ν_{max} 3392 (br), 3387 (br), 3373 (br), 3367 (br), 1674, 1635, and 1584 cm^{-1}; $[\alpha]^{20}_{D}$+7.8 (*c* 0.10, $CHCl_3$); HRESITOFMS *m/z* 247.1308 $[M + Na]^+$ (calculated for $C_{13}H_{20}O_3Na$, 247.1310); ^{1}H and ^{13}C NMR data, see Table 3.2.

3.2.9 Disruption of the *lav* genes in the heterologous host

A 3.1 kb internal fragment of *lavA* was amplified by PCR with the primer pair dislavA-Fw/dislavA-Re. Similarly, a 2.5 kb internal fragment of *lavB* was amplified with the primer pair dislavB-Fw/dislavB-Re. The resultant fragments were digested with *Eco*RI and inserted into the *Eco*RI site of pKU250 to yield pLT164 for *lavA* disruption and pLT165 for *lavB* disruption. These plasmids were introduced into *S. avermitilis* SUKA22 carrying the cosmid cDD60α via intergeneric conjugation, and each *lav* gene was inactivated by plasmid insertion. The genotype of each candidate strain for gene disruption was confirmed by PCR analysis. *S. avermitilis* SUKA22 carrying the *lavA*-disrupted cosmid cDD60α and that carrying the *lavB*-disrupted cosmid cDD60α were abbreviated as Δ*lavA* and Δ*lavB*, respectively. After three rounds of growth in liquid TSB (tryptic soy broth; Oxoid, Basingstoke, UK)+YEME (1:1 v/v) medium in the absence of

thiostrepton, revertant strains were selected based on thiostrepton sensitivity. The genotype was confirmed by PCR analysis for the absence of a thiostrepton-resistant gene (*tsr*).

3.3. Results

3.3.1 Genome mining of a silent type I PKS gene cluster

To find potential biosynthetic gene clusters for secondary metabolites, the genome of *S. lavendulae* FRI-5 was sequenced using the Illumina sequencing technology. Analysis of the draft sequence with AntiSMASH (Blin et al., 2013) predicted the presence of biosynthetic genes for polyketides, peptides, and terpenes, including the indigoidine biosynthetic gene (*lbpA*) and the *lac* cluster for a putative aromatic polyketide (Pait et al., 2017). A 9.9-kb contig (H27_Node344) possesses six open reading frames (ORFs) (*orf20* to *orf25*) and an incomplete ORF (*lavA*) that partially encodes a functional module of a typical type I PKS. The six ORFs putatively encode transcriptional regulators, a membrane protein, and hypothetical proteins. Transcriptional analysis demonstrated that putative regulatory genes *orf21* and *orf22* were transcribed in a time-dependent manner, but the transcript of *lavA* was not detected (Fig. 3.2), suggesting that the newly identified type I PKS gene cluster comprising the *lavA* gene seems to be silent in *S. lavendulae* FRI-5.

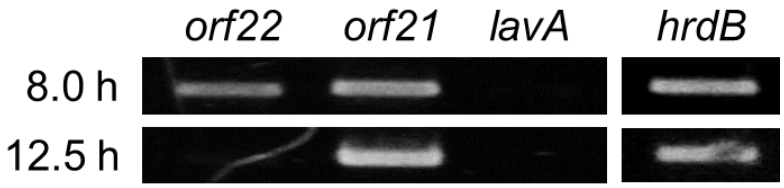


Figure 3.2 Gene expression analysis of the *lavA* gene and putative regulatory genes. Total RNAs were extracted from mycelia collected at the indicated cultivation times. The *hrdB* gene was used as an internal control, because this gene is expressed constantly throughout growth

To isolate the complete type I PKS gene cluster, a cosmid library of the *S. lavendulae* FRI-5 genome was constructed and screened by PCR, resulting in two overlapping cosmids designated AA94 and DD60. The entire *lav* cluster was identified after cosmid-based genomic walking and sequencing of PCR-amplified gaps between neighboring contigs. We opened a total of 55.1 kb of contiguous DNA, and annotation analysis of this region and comparison with genes in public databases showed 31 predicted ORFs. The genetic organization is shown in Fig. 3.3, and the deduced gene functions are summarized in Table 3.2.

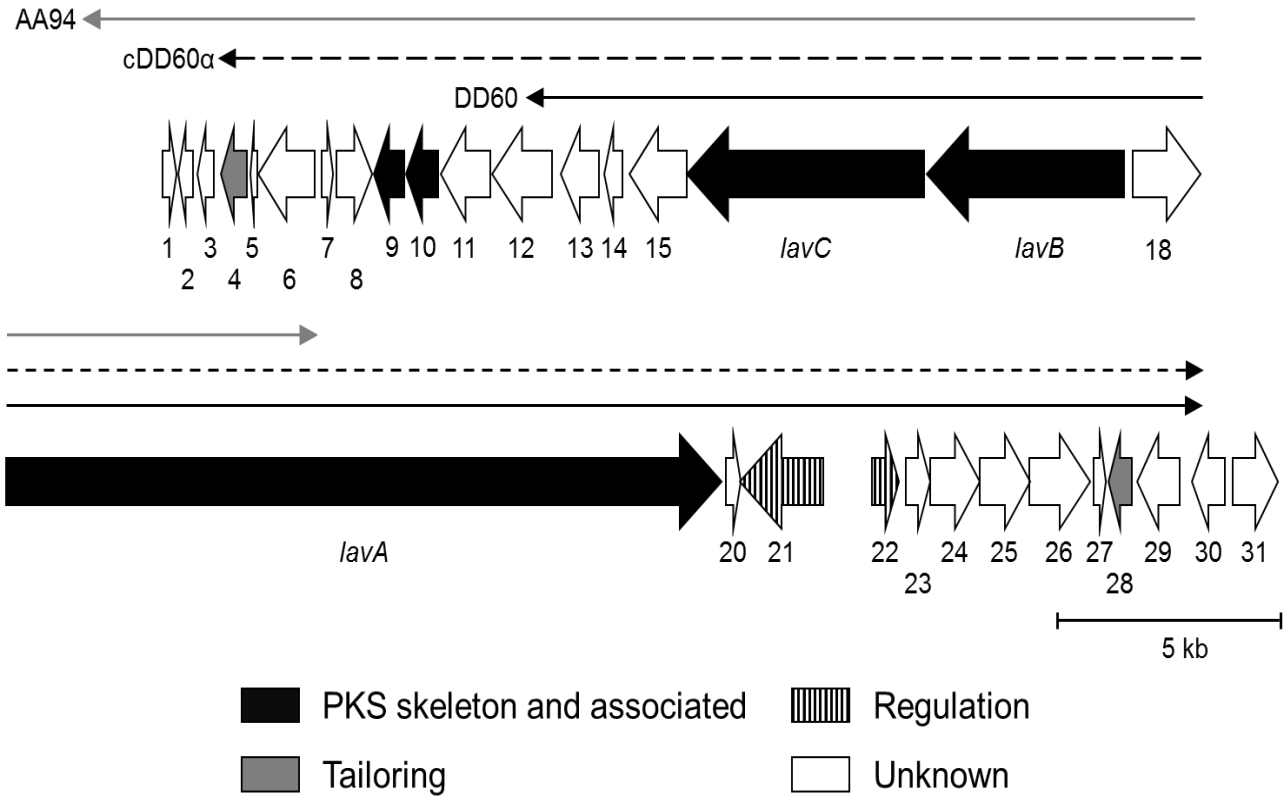


Figure 3.3 Genetic organization of a 55.1-kb region containing the lavendiol biosynthetic gene cluster. The solid arrow, the dashed arrow, and the gray arrow indicate the boundaries of DNA fragments from cosmids cDD60, cDD60α, and AA94, respectively. Each arrow indicates the direction of transcription and relative gene size. ORFs predicted to participate in lavendiol biosynthesis are shaded. The proposed functions of individual ORFs are indicated here and summarized in Table 1.

Table 3.2. Deduced functions of ORFs in the lavendiol biosynthetic gene cluster

Gene	Size[a]	Proposed function	Homolog[b] and origin	Identity /similarity (%)
orf1	109	Hypothetical protein	Hypothetical protein (WP_033225719), *S. virginiae*	55/70
orf2	117	Transposase	Putative transposase (ALO13484), *S. venezuelae* ATCC 15439	91/94
orf3	130	Transposase	IS5 family transposase (OLZ65495), *S. amritsarensis* MTCC 11845	93/95
orf4	209	SAM-dependent methyltransferase	Orf7 (BAI70387), *S. lavendulae* ATCC 11924	92/95
orf5	83	Hypothetical protein	Hypothetical protein (WP_030620314), *S. sclerotialus*	58/70
orf6	446	Hypothetical protein	Hypothetical protein (WP_033524496), *S. galbus*	56/67
orf7	85	Hypothetical protein	Orf9 (BAI70388), *S. lavendulae* ATCC 11924	94/97
orf8	283	Lipoprotein	Orf10 (BAI70389), *S. lavendulae* ATCC 11924	89/93
orf9	243	Phosphopantetheinyl transferase	EntD (AJF67787), *S. vietnamensis* GIMV4	68/76
orf10	255	Type II thioesterase	Type II thioesterase (CAQ52621), *S. violaceoruber* 3844-33C	57/67
orf11	395	Cytochrome P450	BN6_47180 (CCH31997), *Saccharothrix espanaensis* DSM 44229	81/90
orf12	479	Amine oxidase	BN6_47190 (CCH31998), *Saccharothrix espanaensis* DSM 44229	78/86
orf13	297	NAD-dependent oxidoreductase	BN6_47200 (CCH31999), *Saccharothrix espanaensis* DSM 44229	81/89
orf14	141	DUF4440 domain-containing protein	BN6_47210 (CCH32000), *Saccharothrix espanaensis* DSM 44229	80/88
orf15	454	FAD-binding monooxygenase	BN6_47220 (CCH32001), *Saccharothrix espanaensis* DSM 44229	71/78
lavC	1890	Type I polyketide synthase	BN6_47230 (CCH32002), *Saccharothrix espanaensis* DSM 44229	75/81
lavB	1562	Type I polyketide synthase	BN6_47240 (CCH32003), *Saccharothrix espanaensis* DSM 44229	74/82
orf18	536	Aminotransferase	BN6_47250 (CCH32004), *Saccharothrix espanaensis* DSM 44229	71/83
lavA	5660	Type I polyketide synthase	BN6_47260 (CCH32005), *Saccharothrix espanaensis* DSM 44229	74/81
orf20	118	Nuclear transport factor 2-family protein	BN6_47280 (CCH32006), *Saccharothrix espanaensis* DSM 44229	68/75
orf21	751	SARP-family transcriptional regulator	BN6_47290 (CCH32007), *Saccharothrix espanaensis* DSM 44229	70/78
orf22	216	XRE-family transcriptional regulator	Orf11 (BAI70390), *S. lavendulae* ATCC 11924	96/97
orf23	197	Hypothetical protein	Orf12 (BAI70391), *S. lavendulae* ATCC 11924	95/98

Table 3.2 cont.

Gene	Size[a]	Proposed function	Homolog[b] and origin	Identity /similarity (%)
orf24	391	Membrane protein	VR44_04565 (KJY37951), *S. katrae* NRRL ISP-5550	91/94
orf25	395	Hypothetical protein	Hypothetical protein (WP_079432212), *S. katrae* NRRL ISP 5550	85/89
orf26	476	Pyridine nucleotide-disulfide oxidoreductase	VR44_04575 (KJY37968), *S. katrae* NRRL ISP-5550	89/91
orf27	93	Hypothetical protein	VR44_04580 (KJY37952), *S. katrae* NRRL ISP-5550	83/86
orf28	181	GCN5 family acetyltransferase	VR44_04585 (KJY37953), *S. katrae* NRRL ISP-5550	87/93
orf29	333	UDP-glucose 4-epimerase	UDP-glucose 4-epimerase GalE (KJS58022), *S. rubellomurinus* ATCC 31215	70/79
orf30	260	Dehydrogenase	Short-chain dehydrogenase (KUF18065), *S. silvensis* ATCC 53525	80/87
orf31	359	ATPase	ATPase (ALO12333), *S. venezuelae* ATCC 15439	87/90

[a] Numbers refer to amino acid residues.

[b] Parenthetical codes are National Center for Biotechnology accession number

Three PKS genes (*lavA*, *lavB*, and *lavC*) were identified in the cluster, and the *lavA* gene, together with a putative aminotransferase gene (*orf18*) and a hypothetical gene (*orf20*), is divergently located from the *lavB* and *lavC* genes. Based on the predicted function, seven genes (*orf9* to *orf15*) in the region downstream of the *lavB* and *lavC* genes are probably involved in the type I PKS machinery. Proteins encoded by these *lav* genes and the adjacent eight genes (*orf11* to *orf21*) showed high similarity to those in an uncharacterized type I PKS gene cluster of *Saccharothrix espanaensis* DSM 44229, a non-*Streptomyces* actinomycete, and the gene arrangement at this locus was found to be highly conserved (Fig. 3.4) (Strobel et al., 2012). The region downstream of *orf21* contains several enzymatic genes (Orf26 as a disulfide oxidoreductase and Orf30 as a dehydrogenase), but their functions are still unknown in the secondary metabolism of *S. lavendulae* FRI-5.

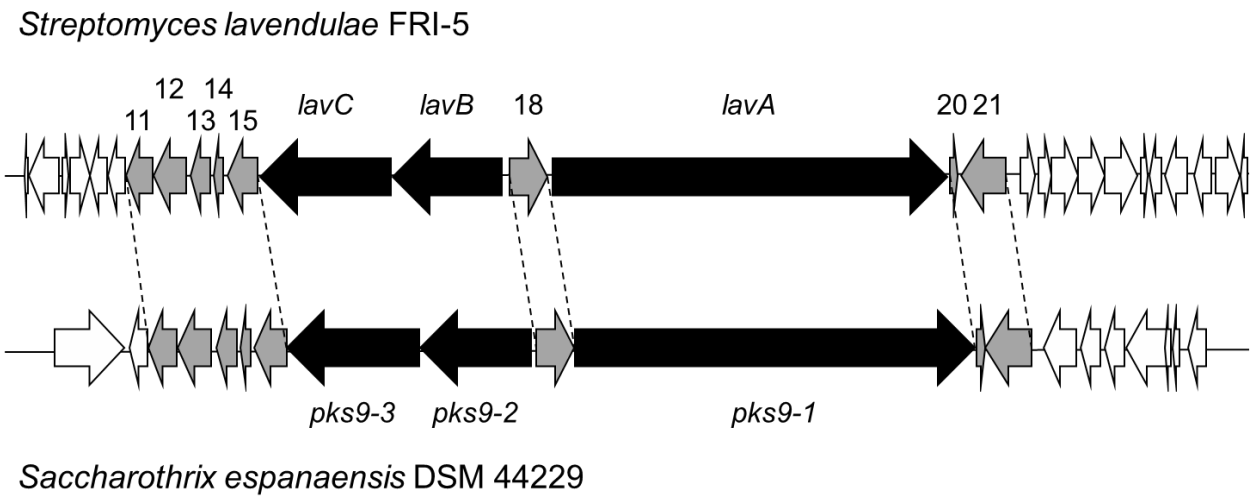


Figure 3.4 Comparison of the lav gene cluster from *Streptomyces lavendulae* FRI-5 with the uncharacterized type I PKS gene cluster from *Saccharothrix espanaensis* DSM 44229.

3.3.2 Activation of the silent type I PKS gene cluster by heterologous expression

In order to activate the silent *lav* genes, we constructed the genome-integrative cosmid cDD60 by replacing the kanamycin resistance gene (*neo*) in cosmid DD60 with a fragment containing both an apramycin resistance gene and a ΦC31 integration cassette. The resulting cosmid cDD60 was introduced by protoplast transformation into *S. avermitilis* SUKA22, an engineered host suitable for heterologous expression, and was integrated into the chromosome. The *S. avermitilis* SUKA22 carrying cosmid cDD60 (*S. avermitilis* SUKA22/cDD60) produced no apparent metabolites, similar to *S. avermitilis* carrying an empty vector alone (Fig. 3.5). The downstream region of *orf13*, which was not included in cosmid cDD60, possesses important genes for the biosynthesis of polyketide compounds: Orf4, a putative *S*-adenosylmethionine (SAM)-dependent methyltransferase; Orf9, a putative phosphopantetheinyl transferase; and Orf10, a putative type II thioesterase. To incorporate the functions of these enzymatic genes in cosmid cDD60, we constructed cosmid cDD60α, which covers 26 genes from *orf4* to *orf29*, by λ Red-mediated recombination (Fig. 3.1). Subsequently, the transformants of *S. avermitilis* SUKA22

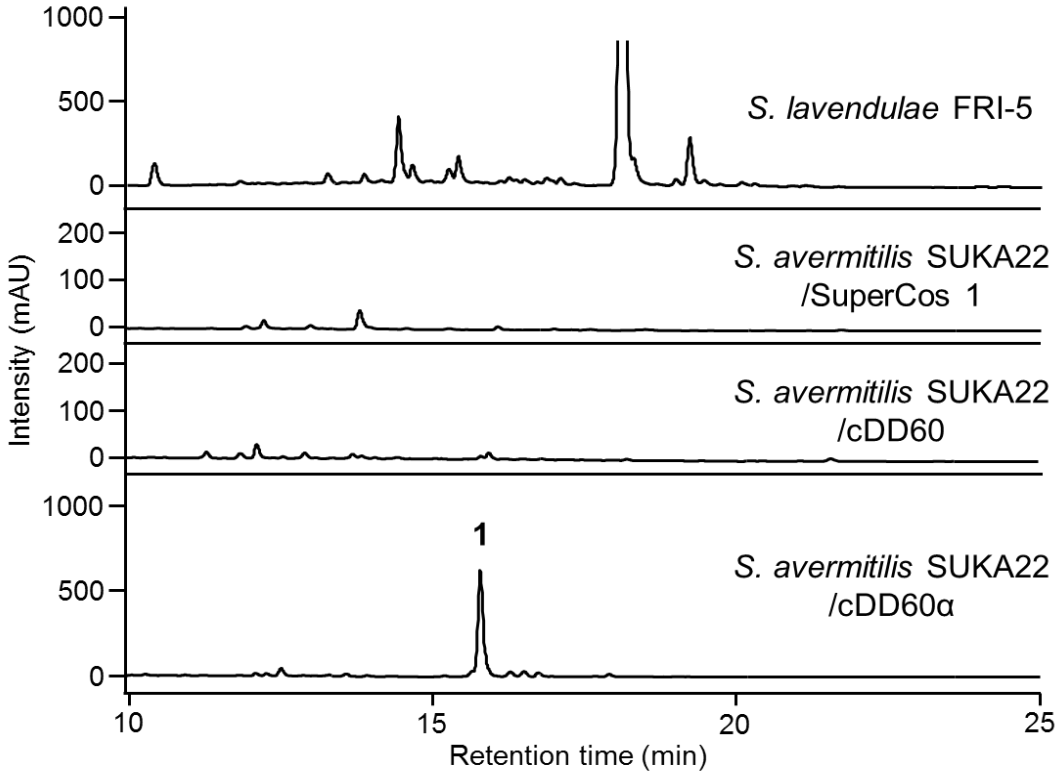


Figure 3.5 HPLC analysis of metabolite profiles from *S. lavendulae* FRI-5 and *S. avermitilis* SUKA22 carrying different cosmids. *S. avermitilis* SUKA22/SuperCos 1, *S. avermitilis* SUKA22 carrying SuperCos 1 (vector control); *S. avermitilis* SUKA22/cDD60, *S. avermitilis* SUKA22 carrying cosmid cDD60, where nine genes (*orf4* to *orf12*) are not included; *S. avermitilis* SUKA22/cDD60α, *S. avermitilis* SUKA22 carrying cosmid cDD60α, where the nine genes are present together with the *lav* genes. mAU, milliabsorbance units at 290 nm. A peak for lavendiol (**1**) is indicated by a compound number.

were grown in several liquid media, and *n*-BuOH extracts of the culture broth were analyzed by reversed-phase HPLC. Unlike the case of *S. avermitilis* SUKA22 carrying cosmid cDD60, *S. avermitilis* SUKA22 carrying cosmid cDD60α (*S. avermitilis* SUKA22/cDD60α) was found to produce one remarkably large peak (compound 1, 15.8 min) (Fig. 3.5). This peak was not observed in the HPLC chromatogram of *S. lavendulae* FRI-5. These results clearly indicated that cosmid cDD60α confers the ability to synthesize compound **1**, a cryptic metabolite of *S. lavendulae* FRI-5, to the engineered host *S. avermitilis* SUKA22.

3.3.3 Isolation and structural elucidation of new diol-containing polyketide

In order to isolate compound **1,** *S. avermitilis* SUKA22/cDD60α was grown in 5 L of production medium containing 25 mM L-glutamic acid for 4 days. Compound **1** was successively purified from the culture broth by silica gel column chromatography and preparative reversed-phase HPLC, and about 15 mg of **1** was obtained as a yellowish powder. The molecular formula of compound **1** was deduced to be $C_{13}H_{20}O_3$ by HRESITOFMS (*m/z* 247.1308 $[M + Na]^+$, calculated for 247.1310), indicating the presence of four sites of unsaturation in the structure. The structure of compound **1** was established by single-dimensional NMR (^1H and ^{13}C NMR) and two-dimensional NMR analysis (HMBC, HSQC, ^1H-^1H COSY, and NOESY) (Table 3.3, Fig. 3.6).

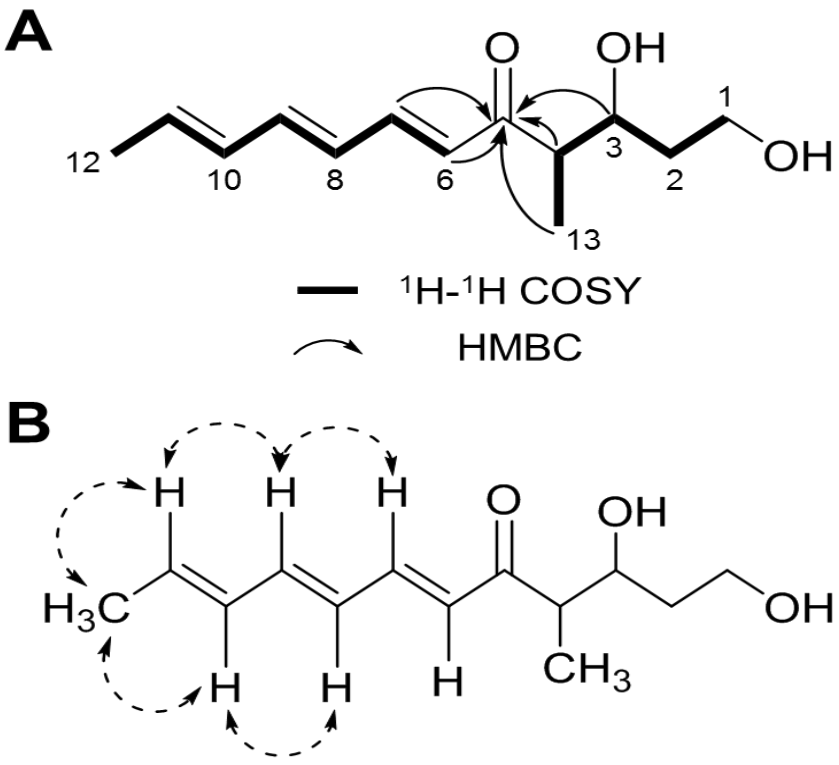


Figure 3.6 ^1H-^1H COSY and key HMBC correlations of compound **1** (A), and key NOESY correlations of **5** (B).

The ^{1}H-^{1}H COSY correlations established two sequences from olefinic methine protons H-6 (δ_H 6.25) to methyl proton H_3-12 (δ_H 1.79), demonstrating the presence of the triene moiety, and from methylene protons H_2-1 (δ_H 3.67) to methyl protons H_3-13 (δ_H 1.04). HMBC correlations of **1** from olefinic methine protons (H-6 and H-7), methyl protons (H_3-13), and a methine proton (δ_H 3.91, H-3) to a carbonyl carbon C-5 (δ_C 204.2) confirmed a connection between two partial fragments: C-1 to C-13 and C-6 to C-12. The IR absorption bands of compound **1** suggested the presence of OH or NH groups (broad absorption bands: ν_{max} 3392, 3387, 3373, and 3367 cm^{-1}). These data, taken together with the predicted molecular formula based on HRMS analysis, suggested that hydroxy groups were connected to C-1 (δ_C 58.6) and C-3 (δ_C 70.3). The geometries of the double bond between C-6 and C-8 were assigned as 6*E*, 8*E*, on the basis of the vicinal coupling constant ($J_{6,7}$ = 15.4 Hz in average and $J_{8,9}$ = 14.8 Hz) (Table 2), and as 10*E* by NOESY correlations for H-8/H10, H-9/H-11, and H-10/H_3-12 (Fig. 3.6b). Thus, the structure of **1** was elucidated as (6*E*,8*E*,10*E*)-1,3-dihydroxy-4-methyldodeca-6,8,10-trien-5-one, which has not been isolated from any natural sources. We named compound **1** "lavendiol" (for *S.* *lavendulae* FRI-5 diol-containing polyketide).

Table 3.3 NMR data for lavendiol (**1**) in methanol-d_4.

Position	δ_C	δ_H mult (*J* in Hz)	HMBC
1	58.6	3.67, m	2, 3
2a	36.1	1.53, m	1, 3, 4
2b		1.72, m	1, 3, 4
3	70.3	3.91, ddd (9.8, 7.4, 2.4)	1, 2, 4, 5, 13
4	50.0	2.94, dq (7.1,7.1)	2, 3, 5, 13
5	204.2		
6	127.7	6.25, d (15.5)	5, 7, 8
7	143.8	7.25, dd (15.2, 11.2)	5, 8, 9
8	128.0	6.27, dd (14.8, 11.2)	6, 7, 10
9	142.6	6.65, dd (14.8, 10.8)	7, 10, 11
10	131.3	6.20, m	8, 9, 12
11	135.3	5.98, m	9, 12
12	17.2	1.79, d (6.8)	10, 11
13	12.2	1.04, d (7.0)	3, 4, 5

3.3.4 Disruption of *lavA* and *lavB* genes

To determine the cluster responsible for the biosynthesis of lavendiol, we disrupted the *lavA* and *lavB* genes in the *S. avermitilis* SUKA22/cDD60α strain by insertional inactivation via a single crossover. The schematic representation of PKS gene deletions and the genotype of the

disruptant and revertant strains are shown in Figure 3.7. The disruption of either gene completely abolished the production of lavendiol (Fig. 3.8). In addition, revertant strains (Δ*lavA*/revertant and Δ*lavB*/revertant) were obtained from these disruptants by a second crossover, and the production of lavendiol was recovered. These results clearly indicated that the enzymatic function of LavA and that of LavB are indispensable for lavendiol production and that the cluster in cosmid cDD60α is involved in lavendiol biosynthesis.

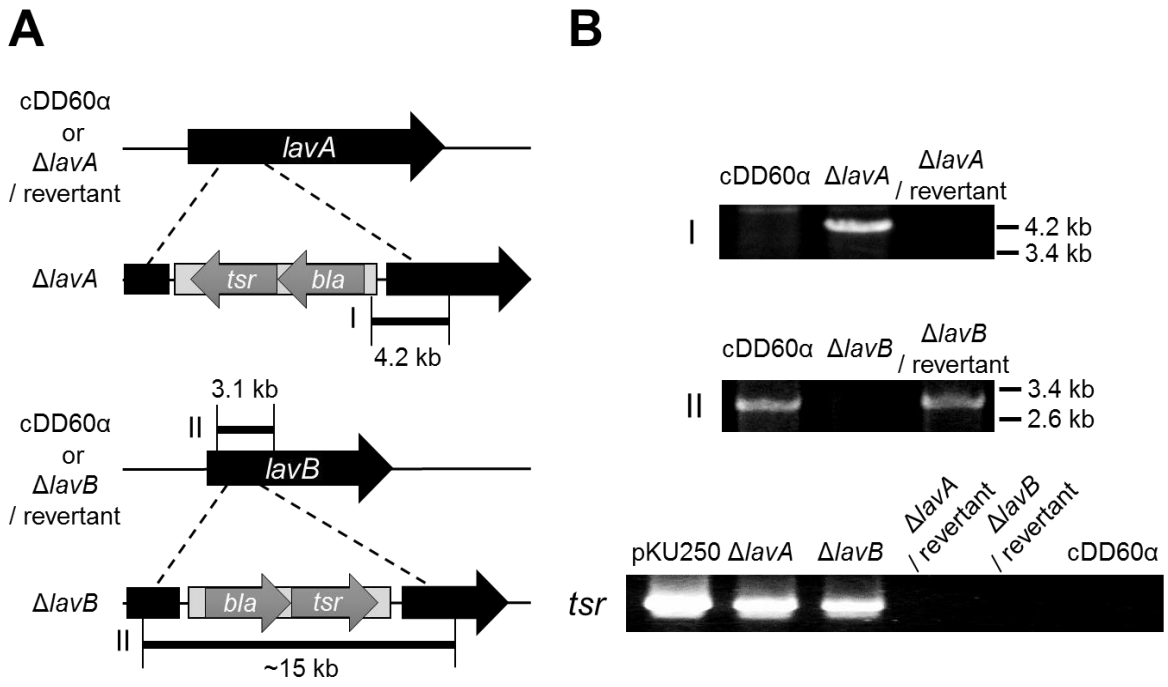


Figure 3.7 Disruption of the *lavA* and *lavB* genes in *S. avermitilis* SUKA22 carrying cosmid cDD60α. cDD60α, *S. avermitilis* SUKA22 carrying cosmid cDD60α; Δ*lavA*, *S. avermitilis* SUKA22 carrying the *lavA*-disrupted cosmid cDD60α; Δ*lavB*, *S. avermitilis* SUKA22 carrying the *lavB*-disrupted cosmid cDD60α; Δ*lavA*/revertant, a revertant strain derived from the Δ*lavA* strain; Δ*lavB*/revertant, a revertant strain derived from the Δ*lavB* strain. (A) Schematic representation of the strategy for insertional inactivation of *lavA* (upper panel) and *lavB* (lower panel). *tsr* and *bla* genes confer resistance for thiostrepton and ampicillin, respectively. (B) PCR analysis to confirm genotype of Δ*lavA* and Δ*lavB* strains and the revertants. A 4.2 kb fragment (I) containing a part of *lavA* and pKU250 was detected by using the primer pair pKU250-bla/lavA-dw. A 3.1 kb internal fragment (II) of the *lavB* gene was detected by using the primer pair lavB-up/lavB-dw. The *tsr* gene (810 bp) was amplified using the primer pair tsr-Fw/tsr-Re.

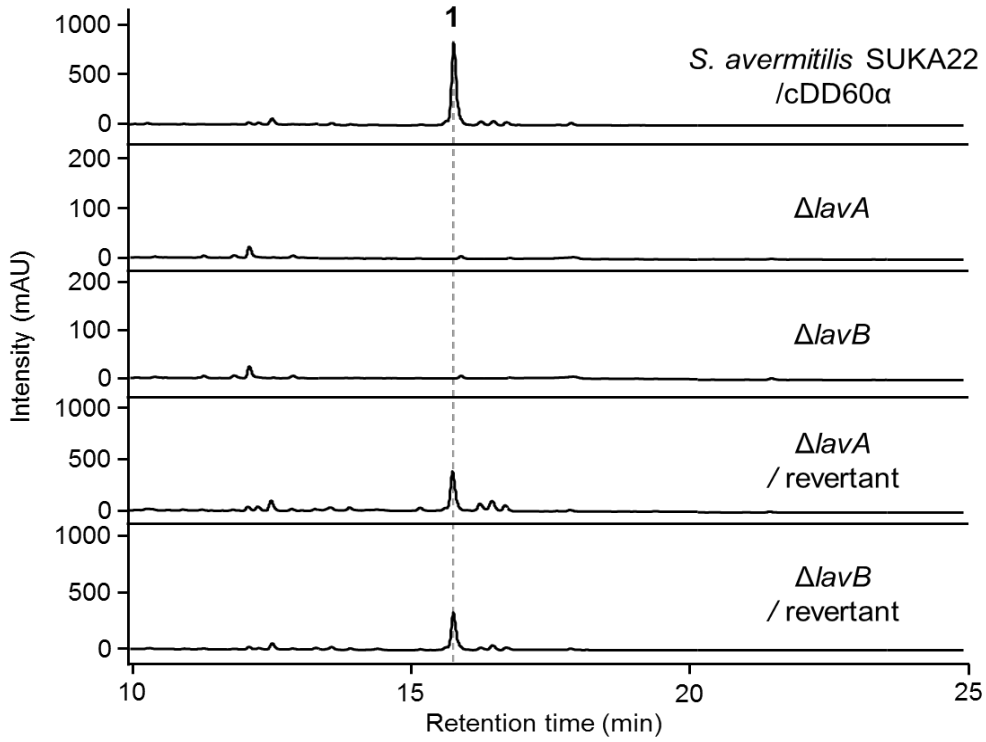


Figure 3.8 Lavendiol (**1**) production analyzed after gene disruption of *lavA* and *lavB* in the heterologous hosts. *S. avermitilis* SUKA22/cDD60α, *S. avermitilis* SUKA22 carrying the cosmid cDD60α; Δ*lavA*, *S. avermitilis* SUKA22 carrying the *lavA*-disrupted cosmid cDD60α; Δ*lavB*, *S. avermitilis* SUKA22 carrying the *lavB*-disrupted cosmid cDD60α; Δ*lavA*/revertant, a revertant strain derived from the Δ*lavA* strain; Δ*lavB*/revertant, a revertant strain derived from the Δ*lavB* strain. mAU, milliabsorbance units at 290 nm. A peak for lavendiol (**1**) is indicated by a compound number.*lavA* and pKU250 was detected by using the primer pair pKU250-bla/lavA-dw. A 3.1 kb internal fragment (II) of the *lavB* gene was detected by using the primer pair lavB-up/lavB-dw. The *tsr* gene (810 bp) was amplified using the primer pair tsr-Fw/tsr-Re.

3.3.5 *In silico* analysis of the gene cluster for lavendiol biosynthesis

The inserted region of cosmid cDD60α harbors 26 genes, including 3 PKS genes (*lavA*, *lavB*, and *lavC*) and several genes putatively involved in polyketide assembly. To predict the polyketide assembly process for lavendiol biosynthesis, these PKSs were analyzed *in silico*, demonstrating that a total of 26 enzymatic domains are organized into six modules (Fig. 3.9). The PKS module in the canonical type I PKS machinery employs a minimal domain set that comprises acyltransferase (AT), ketosynthase (KS), and acyl carrier protein (ACP) domains (James Staunton

& Weissman, 2001). The first module of LavA contains a mutant KS^Q domain (Fig. 3.10) and lacks the cognate AT domain that loads a coenzyme A-activated short-chain acyl group onto the ACP domain. KS^Q domains decarboxylate the ACP-bound malonate unit to form an acetyl group at the polyketide chain assembly (Fischbach & Walsh, 2006). The downstream of the KS^Q domain in the first module of LavA has a short segment (184 amino acids) as remnant AT residues, and this segment is expected to serve as an AT-docking domain in the *trans*-AT type I PKS system (Cheng, Tang, & Shen, 2003; Piel, 2010). Thus, in the initiation step of lavendiol biosynthesis, we predict that a malonyl-CoA unit is incorporated into the ACP domain of the loading module by an acyltransferase acting in *trans*, and the resulting malonyl-*S*-ACP is decarboxylated by the KS^Q domain to afford an acetyl starter unit. The gene product of *orf28* is the only discrete acetyltransferase in the vicinity of the *lav* genes, which is classified into the GCN5-related *N*-acetyltransferase (GNAT) family. In curacin A biosynthesis, the cognate GNAT domain in the loading module catalyzes decarboxylation of malonyl-CoA to acetyl-CoA and *S*-acetyl transfer from acetyl-CoA to load an ACP domain (Gu et al., 2007). However, the Orf28 protein has no important residues (neither His nor Thr) for malonyl-CoA decarboxylation, suggesting that Orf28 could play a role only in the transfer of malonyl-CoA into the ACP domain. Another possibility is that a malonyl-CoA transferase in the fatty acid biosynthesis (Wesener, Potharla, & Cheng, 2011)or a free-standing AT enzyme of *S. avermitilis* SUKA22 provides a malonyl-CoA unit into the loading module of LavA. Thus, in the initiation step of lavendiol biosynthesis, a malonyl-CoA unit may be incorporated into the ACP domain of the loading module, and the resulting malonyl-*S*-ACP is decarboxylated by the KS^Q domain to afford an acetyl starter unit. All the remaining KS domains and ACP domains were

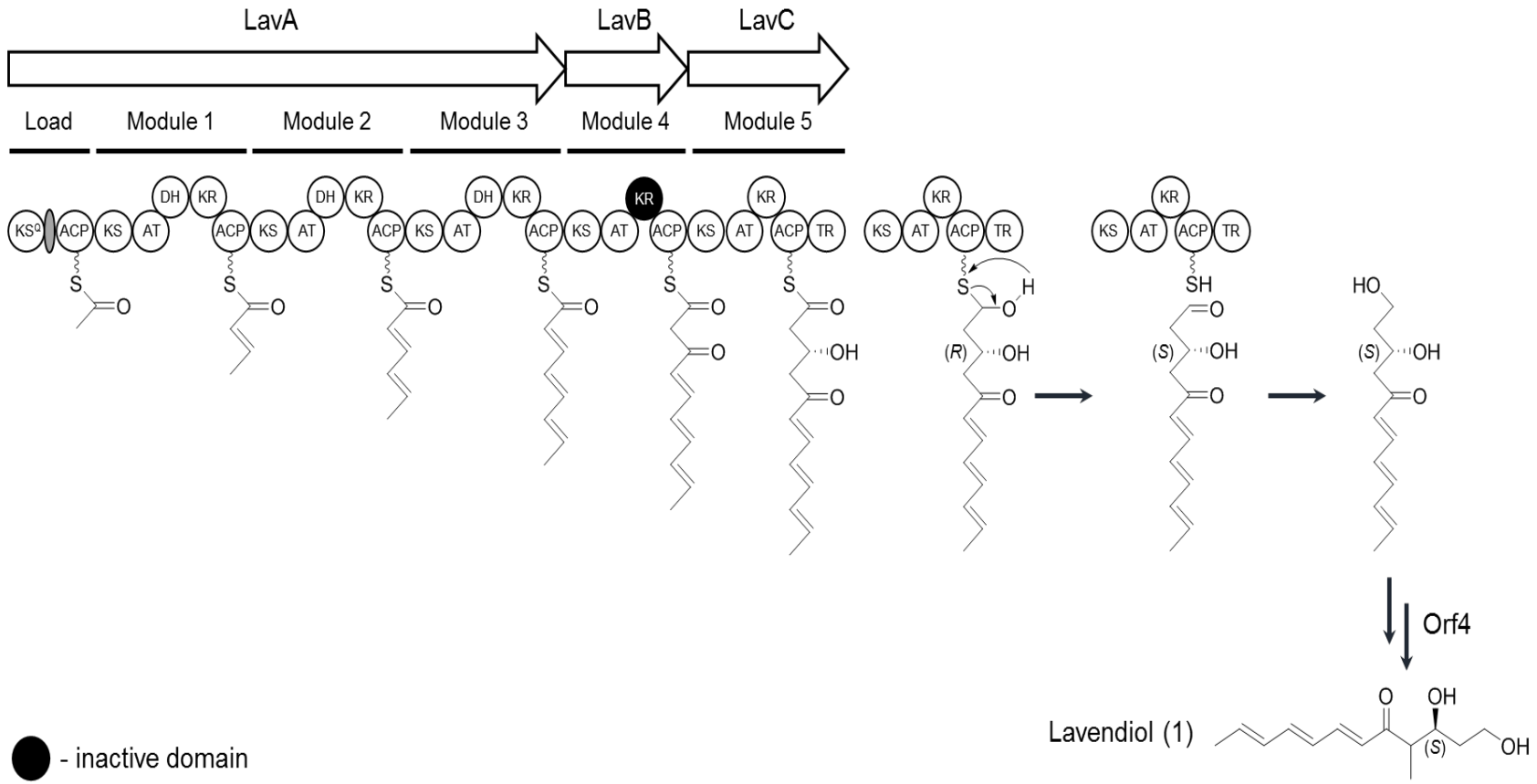


Figure 3.9 Predicted biosynthetic model of lavendiol (**1**). The modules and enzymatic domains (circles) in the PKS polypeptides are shown. KS, ketosynthase; KS^Q, KS in which the active site Cys is replaced by Gln; AT, acyltransferase; ACP, acyl carrier protein; DH, dehydratase; KR, ketoreductase. The presumed inactive KR domain of module 4 is shaded in black. The configuration of a hydroxy group at C-3 was deduced based on sequence analysis of the KR domain of module 5.

resumed to be functional based on the presence of conserved active sites (Fig. 3.10) (Min, Varoglu, & Sherman, 2000). On the other hand, the substrate specificities of the AT domains were predicted based on the phylogenetic tree analysis, which suggested that all the AT domains are specific for malonyl-CoA (Fig. 3.11). These predicted substrates are in agreement with the carbon skeleton of lavendiol, except for the methyl group at C-13. Because a methyltransferase (MT) domain is not found in the Lav PKSs, the methyl group is probably generated by a discrete MT enzyme, such as Orf4, which is a putative SAM-dependent methyltransferase (Skiba et al., 2016; Yu et al., 2013). With respect to the KR domains, analysis of the conserved active site (catalytic triad of Ser-Tyr-Asn) demonstrated that all the KR domains should be active in the polyketide assembly line (Fig. 3.10) (Reid et al., 2003). However, the biosynthetic pathway of lavendiol does not require the KR activity of module 4, suggesting that KR4 is probably inactive in the assembly line. A similar contradiction is found in the biosynthetic pathways of other polyketides, such as avermectin and nemadectin (Ikeda et al., 1999). Furthermore, sequence analysis revealed that all KR domains contain the stereochemistry signature "LDD" motif for the B1-type KR domains to yield a hydroxy group of *R* configuration (Keatinge-Clay, 2007), suggesting that lavendiol might have an R configuration at C-3 (Fig. 3.9). Regarding the dehydratase (DH) domains, each of the three DH domains possesses the conserved His residue in the conserved HXXXGXXXXP motif (Wu, Zaleski, Valenzano, Khosla, & Cane, 2005), indicating that the DH domains are active and appear to be responsible for the formation of a triene moiety. Module 5 encoded by *lavC* has a thioester reductase (TR) domain, which is homologous (61% identity and 72% similarity) to that of CpkC in the biosynthesis of coelimycin P1(Gomez-Escribano et al., 2012a). Thus, we assumed that the TR domain of LavC catalyzes the NAD(P)H-dependent reductive cleavage to yield an aldehyde intermediate. Based on the predicted functions of Lav PKSs, we suggest that lavendiol biosynthesis

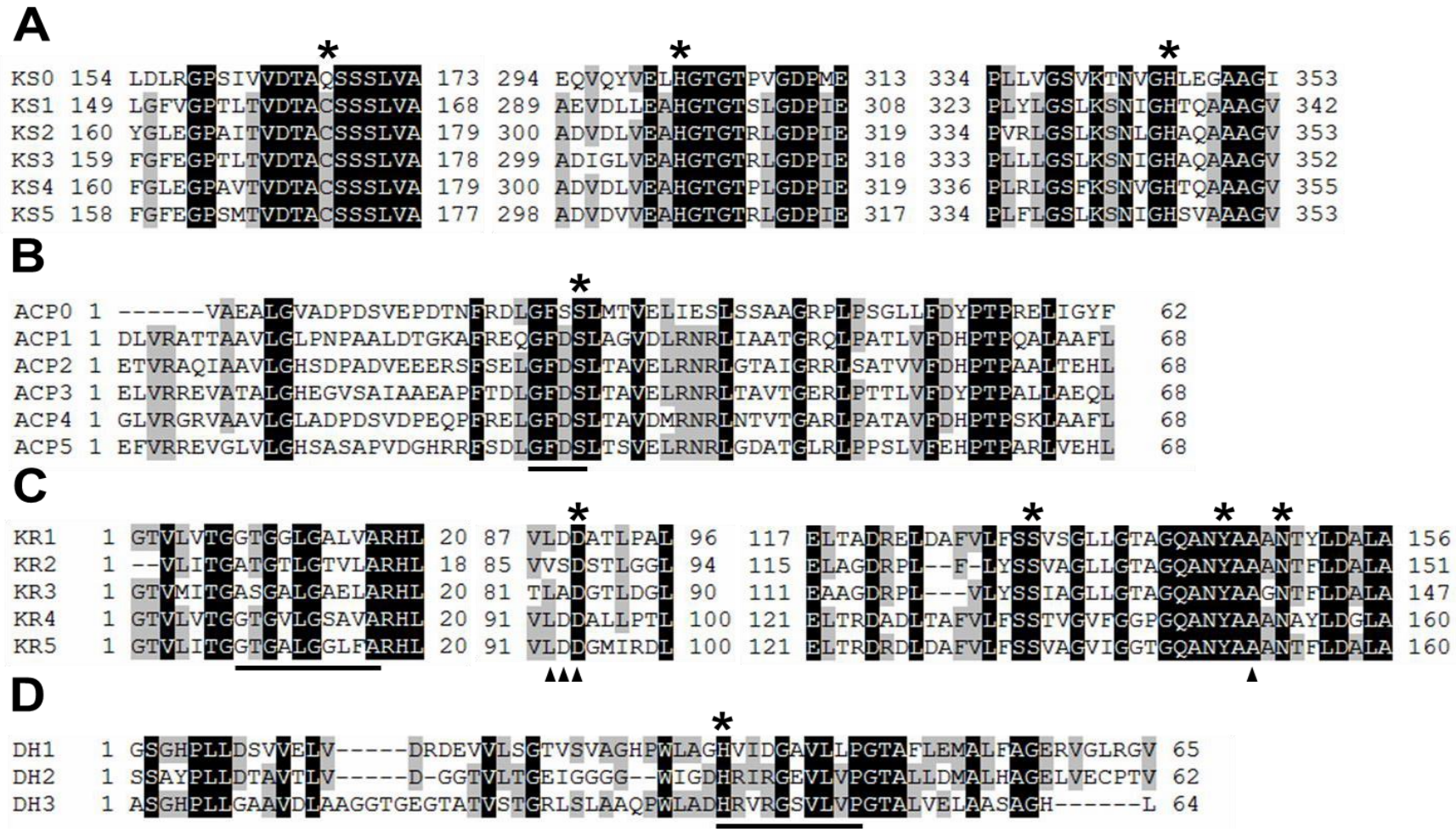


Figure 3.10 Sequence motifs in domains for KS (A), ACP (B), KR (C), and DH (D) from Lav PKSs. The numbers indicate amino acid positions within each domain. Black and grey boxes in the alignment indicate positions at which the same amino acid is found and at least 70% of the amino acids match, respectively. (A) Asterisks indicate the conserved catalytic triad of C-H-H. (B) The consensus sequence for the phosphopantetheine attachment site is underlined. The active site Ser residue is indicated by an asterisk. (C) The consensus sequence for the NADPH-binding site is underlined. The conserved Asp residue and the catalytic triad of S-Y-N are indicated by asterisks. Solid triangles denote the consensus motifs for B1-type KR (LDD, no P). (D) The conserved HXXXGXXXXP motif for DH domains is underlined. Asterisks denote the active site His residue.

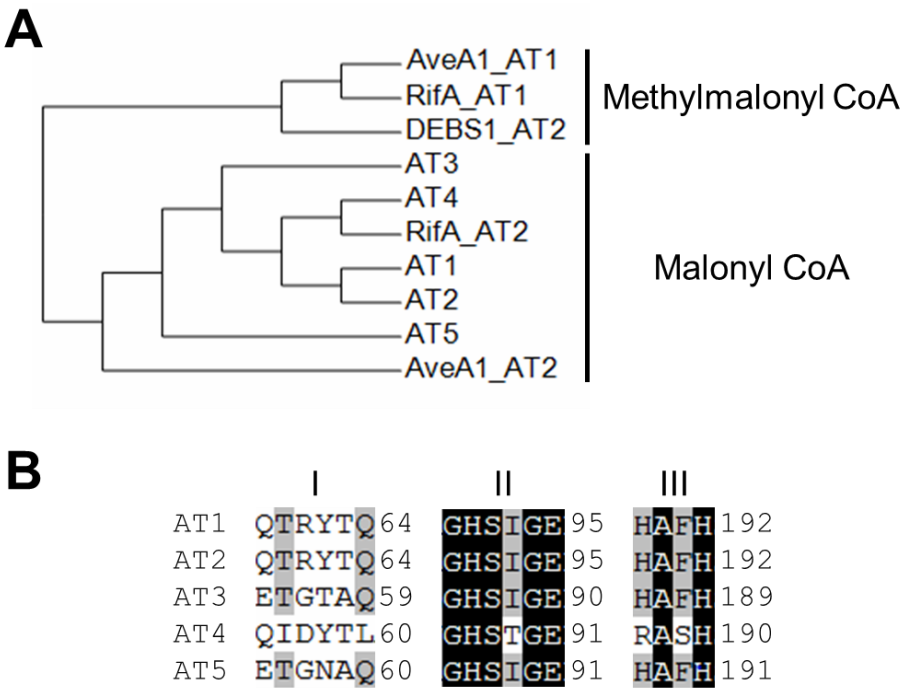


Figure 3.11 Analysis of the AT domains from Lav PKSs by phylogenetic tree (A) and amino acid sequence alignment (B). (A) Multiple sequence alignment was conducted with the CLUSTALW program (http://www.genome.jp/tools/clustalw/). Phylogenetic tree was constructed by the unweighted-pair group method with arithmetic mean. The names of the AT domains and corresponding sequence accession numbers are as follows: AveA1 (BAA84474) for avermectin in *Streptomyces avermitilis*; RifA (AAC01710) for rifamycin in *Amycolatopsis mediterranei*; DEBS1 (Q03131) for erythromycin in *Saccharopolyspora erythraea*. (B) Sequence alignment of the conserved motifs in AT domains from Lav PKS modules. Three dominant motifs, composed of six, six, and four amino acid residues, are assigned in each AT domain. Black and grey boxes in the alignment indicate positions at which the same amino acid is found and at least 70% of the amino acids match, respectively.

starts with the incorporation of malonyl-CoA to the ACP domain of the loading module (Fig. 3.9). Following the generation of an ACP-bound acetyl unit by the KS^{Q} domain, three modules (modules 1 to 3) produce the triene moiety. After additional modules (4 and 5) elongate the polyketide chain, the TR domain of module 5 yields an aldehyde intermediate. This compound is further reduced to a putative primary alcohol intermediate. The C-4 position of the diol compound is methylated by Orf4 to complete lavendiol biosynthesis.

3.4 Discussion

The initiation process of the lavendiol assembly line has several unprecedented features. A type I PKS module minimally contains AT, KS, and ACP domains, while *trans*-AT type I PKS systems lack an AT domain in the PKS modules and have one or a few free-standing ATs (called discrete AT enzymes) that act *in trans* to supply extender units for each elongation step (Piel, 2010). Interestingly, in the *lav* cluster, only the loading module of LavA is AT-deficient, whereas all of other PKS modules (modules 1 to 5) have their own AT domains, implying that lavendiol biosynthesis has two different AT systems. In typical *trans*-AT type I PKS systems, essential AT activities are provided by enzymes encoded within the same gene cluster (Piel, 2010). A putative discrete AT enzyme (Orf28) was found in the *lavA* downstream region, and is a member of GNAT-family acetyltransferases. The loading module of the curacin A PKS has a GNAT domain that bears bifunctional decarboxylase/*S*-acetyltransferase activity to catalyze the loading of an acetyl group (Gu et al., 2007). In contrast, Orf28 appears to be active only for the malonyl-group transfer into the ACP domain, because the important amino acids for malonyl-CoA decarboxylation are missing. Regarding the supply of the malonyl group, the recruitment of malonyl-CoA:ACP acyl-transferase from the primary metabolism of the *Streptomyces* host is also another possible route for polyketide chain initiation. Functional crosstalk between fatty acid biosynthesis and modular polyketide biosynthesis was first demonstrated in the PKS-NRPS hybrid gene cluster for the anti-cancer drug, FK228 (Wesener et al., 2011). Moreover, since endogenous host enzymes have also been shown to interact with the introduced heterologous pathway, putative discrete acyltransferases (SAV_4733 or SAV_5444) present in the *S. avermitlis* SUKA22 host genome might function to successfully express the *lav* cluster. The supplied malonyl group would be modified by the KS^{Q} domain of the loading module to form an acetyl group during the polyketide

chain assembly. This two-step process for the generation of an acetyl starter unit is unusual and intriguing in the PKS system, although gene disruption and *in vitro* experiments will be necessary to clarify the detailed reaction mechanism. Future investigations of the loading module of lavendiol PKS will provide more insights on polyketide chain initiation and the role of host enzymes during heterologous expression of silent gene clusters.

Reductive polyketide chain release from a type I PKS has been sometimes identified in the biosynthetic pathways of alkaloid compounds, such as coelimycin, cyclizidine, and argimycin P(Gomez-Escribano et al., 2012b; Huang, Kim, Liu, & Zhang, 2015; Ye et al., 2017). The last module of the polyketide assembly lines for these alkaloid compounds has a TR domain that yields a putative aldehyde intermediate, which acts as a substrate for an aminotransferase utilizing the amino donor glutamate (Gottelt, Kol, Gomez-Escribano, Bibb, & Takano, 2010). Thereafter, the resultant aminated compound could spontaneously cyclize or could be tailored by some modification enzymes prior to the formation of a nitrogen-containing ring. The upstream region of *lavA* and *lavB* also has a putative class III aminotransferase gene (*orf18*). The gene product of *orf18* shows moderate similarity to class III aminotransferases such as CpkG, CycI, and ArpN for the biosynthesis of coelimycin, cyclizidine, and argimycin P, respectively. Furthermore, several minor *lav* PKS gene-related peaks were detected in the fermentation broth of *S. avermitilis* SUKA22 carrying the cosmid cDD60α (Fig. 3.8), and the production of some of these compounds increased when the cosmid cDD60α was introduced into another heterologous host (data not shown). These findings suggest that the assembly line composed of Lav PKSs, Orf18, and other putative modification enzymes is capable of biosynthesizing an aldehyde intermediate, alkaloid compounds, and other lavendiol-related compounds. The successful activation of the *lav* cluster sets the basis for generating new derivatives through genetic engineering and combinatorial

biosynthesis.

The chemical structure of lavendiol resembles that of streptenol-family compounds (Fig. 3.12), which show a richness of biological activities, such as cholesterol biosynthesis inhibitory activity, immunoregulatory activity, and anti-tumor activity (Mizutani et al., 1989; Tarazona et al., 2017). We evaluated the growth inhibitory activity of lavendiol against several microorganisms (*Bacillus subtilis*, *Escherichia coli*, *Saccharomyces cerevisiae*, *Ralstonia solanacearum*, and *Phytophthora sojae*), but there was no bioactivity against these biological targets. Biosynthetic gene clusters of streptenol-family compounds have never been reported until now. In this work, we successfully identified the lavendiol biosynthetic genes and clarified portions of its biosynthetic pathway. Genetic information of the *lav* cluster may provide useful insights on the formation of the diol moiety that is common to both compounds.

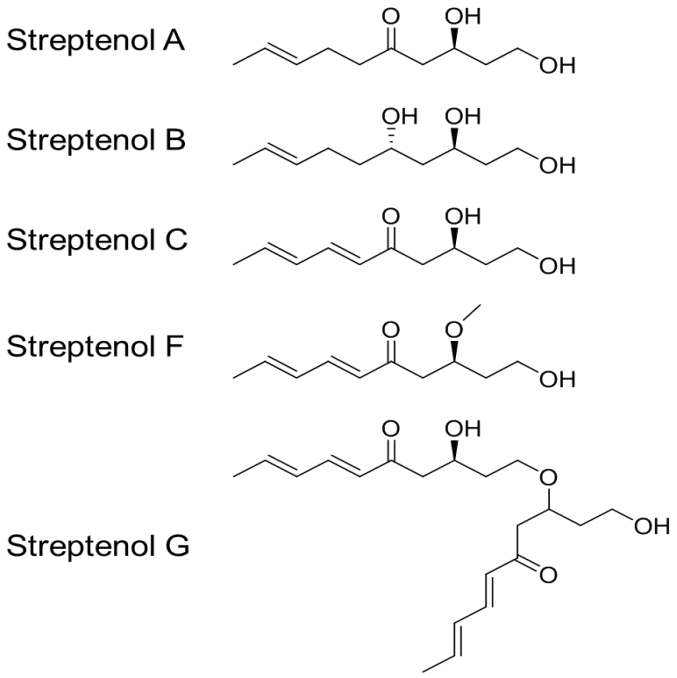


Figure 3.12 Structures of streptenols A, B, C, F, and G. Streptenol A, streptenol B, and streptenol C are isolated from *Streptomyces luteogriseus* strain FH-S 1307; streptenols F and G are isolated from *Streptomyces misionensis* BAT-10-03-023.

The efficacy of using of *S. avermitilis* SUKA22 as a genome mining tool was shown through the successful expression of the *lav* cluster, leading to the isolation and identification of a novel polyketide. Because SUKA22 was derived from an industrial polyketide (Avermectin) producer, its primary metabolism is already optimized to ensure the efficient supply of precursors for the production of other polyketide compounds. In addition, *S. avermitilis* SUKA22 can also support the production of nonribosomal peptides, terpenes and compounds derived from the shikimate pathway as proven by the successful expression of several characterized BGCs (Komatsu et al., 2013). Thus, this genome-minimized host can most likely be used to target a huge diversity of silent biosynthetic gene clusters. Moreover, the production of lavendiol in the *lav* cluster expression strains was quite low (data not shown) prior to media optimization and it would have easily been masked if there were other major metabolites. However, since the metabolic background of SUKA22 is cleaner owing to the deletion of its BGCs for endogenous secondary metabolites, even cryptic compounds produced at small amounts are less likely to be overlooked.

Based on the proposed lavendiol biosynthetic pathway, many of the introduced biosynthetic genes have unclarified roles thus we assumed that lavendiol may not be the only product. Although some other PKS-dependent peaks were detected in the culture broths of SUKA22 *lav*-expression strains, none of them accumulated to levels high enough for further clarification. The toxicity of the introduced BGC's product to the host may prevent it from accumulating, especially if a resistance gene is not in place. When the *lav* cluster was introduced to an alternative heterologous host, *S. coelicolor* M1152, a set of compounds were produced more abundantly and these were confirmed to inhibit the growth of *S. avermitilis* SUKA22 (data not shown). These results showed that different host strains would have different degrees of tolerance against the heterologous compound(s). As in this case, SUKA22 may be sensitive to some

compounds and by introducing cryptic BGCs into not just one but several hosts, we could possibly activate an array of cryptic compounds.

3.5. Summary

In this chapter, I identified a partial type I PKS gene cluster (*lav* cluster) from the draft genome sequence of *S. lavendulae* FRI-5 and assembled its complete sequence by cosmid-based primer walking and sequencing. The *lav* PKS genes were transcriptionally silent and the cluster is therefore cryptic in *S. lavendulae* FRI-5. The *lav* cluster is non-homologous with any characterized type I PKS biosynthetic gene cluster suggesting that it may direct the production of an atypical compound. The original cosmid DD60 harboring all the *lav* PKS genes was insufficient for compound production. However, supplementation of an additional genetic fragment with genes encoding PKS-priming and tailoring enzymes led to successful expression in SUKA22 and to the subsequent detection of multiple compounds that were absent in the fermentation broths of *Streptomyces lavendulae* FRI-5. The major product was purified and characterized as lavendiol, a new polyketide with a terminal diol moiety similar to streptenols and it has not been reported from any natural sources. Gene disruption of *lavA* and *lavB* confirmed the involvement of this cluster in lavendiol biosynthesis. Finally, I also proposed the biosynthetic pathway of lavendiol based on *in silico* analysis of gene functions, and reported a unique PKS loading module-specific *trans* AT system possibly participated by Orf28, an acyltransferase of the GNAT family. We evaluated the growth inhibitory activity of lavendiol against several microorganisms (*Bacillus subtilis*, *Escherichia coli*, *Saccharomyces cerevisiae*, *Ralstonia solanacearum*, and *Phytophthora sojae*), but there was no bioactivity against these biological targets

Chapter 4
General conclusion and future perspectives

Microbial secondary metabolites are exceptionally diverse chemical entities and their derivatives have a long history of contributions to human medicine, animal health and crop protection. In particular, natural products from the *Streptomyces* genus make up two thirds of all commercially available antibiotics, in addition to anti-hypertensives, immunosuppressants and other clinically-relevant compounds. Traditional screening methods to discover microbial natural products usually involve the collection and cultivation of strains in various media, followed by compound extraction and bioassay-guided target identification. However because of extensive research since the 1940's, traditional bioactivity-guided screenings have recently been plagued by the frequent rediscovery of already known compounds. This problem has led many pharmaceutical companies to abandon their natural product research programs completely, despite our constant need of new classes of therapeutics to cope with antibiotic resistance and the emergence of new diseases.

Genes encoding the biosynthetic pathway and regulators for the production of a specialized metabolite are usually physically clustered in the genome and are thus referred to as biosynthetic gene clusters or BGCs. With the massive developments in sequencing technology combined with bioinformatics, a wealth of untapped natural product resources has been discovered from microbial genomes. In 2002, sequencing of the *Streptomyces* genome revealed for the first time that microbial genomes harbor biosynthetic gene clusters for more than 20 secondary metabolites. A vast number of these pathways are tightly controlled by the constraints of precursors, biochemical energy, and environmental cues, and appear to be expressed poorly, if at all. Thus, they have been

referred to as silent/cryptic biosynthetic pathways. The activation of these cryptic BGCs, known as "genome mining", is a radical approach that can stimulate the drying pipeline of natural product discovery. For instance, genome mining makes it possible to focus on the isolation and identification of novel natural products from unique gene clusters and circumvent the historical problem of rediscovery.

In this dissertation, I focused on the genome mining of *Streptomyces lavendulae* FRI-5 which predominantly produces only four secondary metabolites, all of which are controlled by the signaling hormone IM-2 and its receptor FarA. Similar to other streptomycetes, we hypothesized that the genome of this strain may also have the potential to biosynthesize more useful secondary metabolites.

The basic knowledge related to my study including the approaches to activate silent BGCs are given in Chapter 1. The advantage of heterologous expression strategy is that it allows us to identify silent bioactive compounds from a specific target BGC when genetic manipulation in the original strain is difficult or when plausible pathway-specific regulators cannot be identified. Here, I used a derivative of the avermectin industrial producer, *S. avermitilis*, in which DNA segments for majority of its endogenous gene clusters were deleted such that its primary metabolism-derived precursors and biochemical energy can be effectively exploited for the production of the heterologous compounds. *S. avermitlis* SUKA22 has successfully expressed known polyketide and nonribosomal peptide BGCs, with yields sometimes surpassing that of the native producer. In this study, I hypothesized that the utility of *S. avermitlis* SUKA22 might also be extended towards isolating the products of cryptic biosynthetic pathways especially those encoding unpredictable compounds.

In Chapter 2, I have successfully found several gene clusters in the vicinity of the IM-2 receptor *farA*, including the indigoidine biosynthetic gene *lbpA* and a cryptic angucycline cluster (*lac* cluster). This is in agreement with earlier observations that biosynthetic gene clusters usually flank regulatory genes or regulatory islands controlling their expression. These findings suggest that the autoregulator receptor gene can indeed be used as a probe to fish cryptic biosynthetic gene clusters, even in the absence of complete genome sequence. I also showed that the transcription of *lbpA* and the *lac* genes were induced by IM-2 addition, showing that these biosynthetic genes are members of the IM-2 regulon. In our previews works, we have already unraveled the involvement of several regulators of the *far* regulatory island in the biosynthesis of IM-2, FarA and indigoidine; however until now, the biosynthetic gene(s) responsible for the IM-2-responsive compounds and the mechanism by which the IM-2/FarA regulon controls their expression is still unknown. Heterologous expression of *lbpA* in *S. avermitilis* SUKA22 confirmed that LbpA is the indigoidine synthase of the IM-2 signalling cascade. The identification of factors binding to *lbpA* upstream region can be used to further clarify the regulatory network for secondary metabolism in *Streptomyces.*

The cryptic *lac* cluster identified next to the *far* regulatory island was intriguing because of the lack of ketoreductase and aromatase enzymes that are necessary for the formation of the angucycline intermediate. The *lac* cluster is the only reported putative angucycline cluster without these 2 essential genes. This unique feature led me to initially hypothesize that the *lac* genes might encode an atypical compound. I tried to express the *lac* cluster both under its native promoter and under a constitutive promoter, and afterwards added essential genes (*jadE* and *jadD*) from the jadomycin pathway. I also introduced the *lac* genes into an alternative host that can successfully express angucycline gene clusters. However, none of these efforts resulted in the production of

any cluster-related compound. This inability to isolate the corresponding *lac* cluster metabolite is likely caused by genetically aberrant genes whose detection is currently beyond available PKS structural information and bioinformatics analysis. Failure to express the *lac* cluster illustrates the possible presence of non-functional BGCs in the genome of streptomycetes that may hamper the activation of silent BGCs. Nonetheless, the successful expression of *lbpA* and enhanced production of indigoidine demonstrated that the heterologous expression system using *S. avermitilis* SUKA22 in combination with precursor feeding in the engineered strains is a simple and effective approach to achieve high-level and stable production of useful actinomycete natural products or cryptic compounds.

In Chapter 3, I identified a partial type I PKS gene and I assembled the sequence of the corresponding biosynthetic gene cluster by cosmid-based primer walking and sequencing. This cluster was predicted to encode a novel compound and was successfully activated by heterologous expression in *S. avermitilis* SUKA22. The introduction of the uncharacterized *lav* cluster resulted in the appearance of multiple cryptic compounds that were absent in the fermentation broths of *S. lavendulae* FRI-5. Moreover, genetic disruptions of the PKS genes *lavA* and *lavB* confirmed that these new compounds were direct products of the introduced gene cluster. The SUKA22 strain with the completed *lav* cluster produced the major product lavendiol, a new diol-containing polyketide. My current findings exemplify the efficacy of using the genome-minimized *S. avermitilis* SUKA22 host in awakening cryptic biosynthetic gene clusters. There are many uncharacterized gene clusters from actinomycetes genome and as in the case of the *lav* cluster, heterologous expression of other intriguing “cryptic” pathways may also lead to the discovery of other novel chemical entitites.

Based on its proposed biosynthetic pathway, the initiation process of the lavendiol assembly line has several unprecedented features. A type I PKS module minimally contains AT, KS, and ACP domains, while *trans*-AT type I PKS systems lack an AT domain in the PKS modules and have one or a few free-standing ATs (called discrete AT enzymes) that act *in trans* to supply extender units for each elongation step. Interestingly, only the loading module of LavA is AT-deficient, implying that lavendiol biosynthesis has two different AT systems. This is the first report of a polyketide synthesized by a type I PKS that has a loading module-specific *trans* AT system. The only putative acyltransferase in the cluster is encoded by *orf28*, whose gene product is a member of GNAT-family acetyltransferases. GNAT acyltransferase implicated in the loading module of polyketides appear as cognate domains of the PKS enzyme, not as discrete enzymes, and bears bifunctional decarboxylase/*S*-acetyltransferase activity to catalyze the loading of an acetyl group. In contrast, from sequence analysis Orf28 appears to be active only for the malonyl-group transfer into the ACP domain. With regards to the supply of the malonyl group, a malonyl-CoA transferase in the fatty acid biosynthesis or a discrete AT enzyme of the *S. avermitilis* SUKA22 host may also participate in lavendiol biosynthesis. Future investigations of the loading module of lavendiol PKS will provide more insights on polyketide chain initiation and the role of host enzymes during heterologous expression of silent gene clusters.

The *lav* cluster has many tailoring enzymes and hypothetical proteins with no predicted roles in lavendiol biosynthesis, suggesting that it may direct the production compounds other than lavendiol. This was supported by the appearance of multiple *lav*-related minor peaks in SUKA22 and other new peaks when the cluster was expressed in an alternative host (data not shown). My current results show that through the heterologous expression of an uncharacterized BGCs in more than one heterologous host, we can possible activate a series of cryptic metabolites. Moreover,

further deciphering the lavendiol biosynthetic pathway and its products could open the door to the rational development of novel lavendiol derivatives and streptenol-family compounds with interesting biological activities.

My analysis of the draft genome of *S. lavendulae* FRI-5 also identified other short contigs harboring uncharacterized PKS or NRPS genes. I assembled 2 additional clusters by cosmid based primer walking and their genetic organizations are shown in Figure 4.1. Cluster A is a type III PKS biosynthetic gene cluster which is moderately homologous to the *srs* operon of *S. griseus* (Funabashi, Funa, & Horinouchi, 2008). Cluster B includes three single module NRPS genes whose substrates cannot be predicted *in silico.* The sequence of cluster B is still incomplete although it roughly has little homology with any characterized NRPS cluster, suggesting that it might direct the production of a yet undiscovered compounds. Both of these clusters are transcriptionally silent in *S. lavendulae* FRI-5 and I likewise set out to activate them in *S. avermitlis* SUKA22 in preliminary experiments.

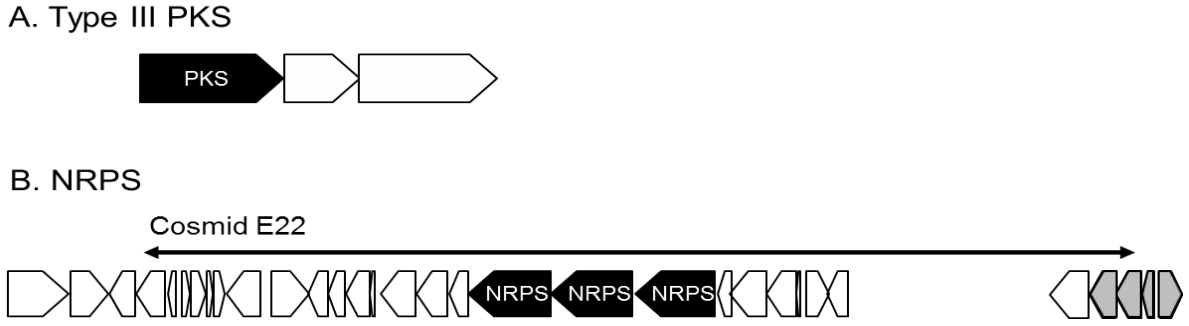


Figure 4.1 Genetic organization of additional cryptic biosynthetic gene clusters identified from the genome of *S. lavendulae* FRI-5. (A) Type III PKS gene cluster homologous to the *srs* operon of *Streptomyces griseus* (B) NRPS gene cluster containing 3 single module NRPS genes shaded in black. The solid arrow indicates the region of the gene cluster covered by cosmid E22 that was used for sequence assembly. Genes shaded in gray encode enzymes for the biosynthesis of non-proteinogenic amino acids. The large gap in between the sequence represents the area that cannot be sequenced.

Cluster A was cloned under the control of the constitutive *ermE** promoter and introduced into the genome of *S. avermitilis* SUKA22, however no product was detected under our culture conditions. A recent study has proven that the presence of endogenous type III PKS clusters in the heterologous host can interfere with the biosynthesis of heterologous type III PKS gene clusters (Thanapipatsiri, Claesen, Gomez-Escribano, Bibb, & Thamchaipenet, 2015). The genome of SUKA22 still harbors a putative type III PKS gene (*sav_7131*) whose gene product is usually responsible for the production 1,3,6,8-tetrahydroxynaphthalene (THN), a precursor for bacterial melanin production (Funa et al., 1999). Removing the endogenous type III PKS gene of *S. avermitilis* SUKA22 may prove useful in improving precursor pools for the future analysis of type III PKS genes not just from *S. lavendulae* FRI-5 but also from a wide range of actinomycetes.

On the other hand, cosmid E22 containing all of the NRPS genes in cluster B was modified to integrate into the genome of *S. avermitilis* in a similar fashion described in Chapter 3. Introduction of cosmid E22 alone also did not result in any NRPS-related product. NRPS enzymes may incorporate non-proteinogenic or highly modified amino acids whose biosynthesis are also encoded by genes within the cluster. Some genes related to amino acid metabolism but are outside of the coverage of cosmid E22 (Fig. 4.1) may be crucial for metabolite production, and thus needs to be "stitched" into cosmid E22, as in the *lav* cluster, prior to heterologous expression.

References

Angert, E. R. (2005). Alternatives to binary fission in bacteria. Nature Reviews Microbiology 3(3): 214–224.

Arakawa, K., Mochizuki, S., Yamada, K., Noma, T., & Kinashi, H. (2007). γ-Butyrolactone autoregulator-receptor system involved in lankacidin and lankamycin production and morphological differentiation in *Streptomyces rochei*. Microbiology 153(6): 1817–1827.

Arias, P., Fernández-Moreno, M. A., & Malpartida, F. (1999). Characterization of the pathway-specific positive transcriptional regulator for actinorhodin biosynthesis in *Streptomyces coelicolor* A3(2) as a DNA-binding protein. Journal of Bacteriology 181(22): 6958–6968.

Bachmann, B. O., Van Lanen, S. G., & Baltz, R. H. (2014). Microbial genome mining for accelerated natural products discovery: Is a renaissance in the making? Journal of Industrial Microbiology and Biotechnology 41: 175–184.

Baltz, R. H. (2010). *Streptomyces* and *Saccharopolyspora* hosts for heterologous expression of secondary metabolite gene clusters. Journal of Industrial Microbiology and Biotechnology 37(8): 759–772.

Baltz, R. H. (2016). Gifted microbes for genome mining and natural product discovery. Journal of Industrial Microbiology and Biotechnology 44: 573–588

Bate, N., Butler, A. R., Gandecha, A. R., & Cundliffe, E. (1999). Multiple regulatory genes in the tylosin biosynthetic cluster of *Streptomyces fradiae*. Chemistry and Biology 6(9): 617–624.

Bentley, S., Chater, K., Cerdeño-Tárraga, A.-M., Challis, G. L., Thomson, N. R., James, K. D., Harris, D. E., et al. (2002). Complete genome sequence of the model actinomycete *Streptomyces coelicolor* A3(2). Nature 417(6885):141–147.

Bérdy, J. (2005). Bioactive Microbial Metabolites. The Journal of Antibiotics 58(1): 1–26.

Bérdy, J. (2012). Thoughts and facts about antibiotics: Where we are now and where we are heading. The Journal of Antibiotics 65(8): 441–441.

Bibb, M. J. (2005). Regulation of secondary metabolism in streptomycetes. Current Opinion in Microbiology 8(2): 208–215.

Bierman, M., Logan, R., O'Brien, K., Seno, E. T., Rao, R. N., & Schoner, B. E. (1992). Plasmid cloning vectors for the conjugal transfer of DNA from *Escherichia coli* to *Streptomyces* spp. Gene 116(1): 43–49.

Blin, K., Medema, M. H., Kazempour, D., Fischbach, M. a., Breitling, R., Takano, E., & Weber, T. (2013). antiSMASH 2.0--a versatile platform for genome mining of secondary metabolite producers. Nucleic acids research 41(Web Server issue): 204–212. Retrieved from 10.1093/nar/gkt449

Bode, H. B., Bethe, B., Höfs, R., & Zeeck, A. (2002). Big effects from small changes: Possible ways to explore nature's chemical diversity. ChemBioChem 3(7): 619-627.

Bunet, R., Song, L., Mendes, M. V., Corre, C., Hotel, L., Rouhier, N., Framboisier, X., et al. (2011). Characterization and manipulation of the pathway-specific late regulator AlpW reveals *Streptomyces ambofaciens* as a new producer of kinamycins. Journal of Bacteriology 193(5): 1142–1153.

Burg, R. W., Miller, B. M., Baker, E. E., Birnbaum, J., Currie, S. a, Hartman, R., Monaghan, R. L., et al. (1979). Avermectins , New Family of Potent Anthelmintic Agents : Producing Organism and Avermectins , New Family of Potent Anthelmintic Agents : Producing Organism and Fermentation. Antimicrobial Agents and Chemotherapy 15(3): 361–367.

Calcott, M. J., & Ackerley, D. F. (2014). Genetic manipulation of non-ribosomal peptide synthetases to generate novel bioactive peptide products. Biotechnology Letters 36(12): 2407-2416.

Cane, D. E., He, X., Kobayashi, S., Omura, S., & Ikeda, H. (2006). Geosmin biosynthesis in *Streptomyces avermitilis*. Molecular cloning, expression, and mechanistic study of the germacradienol/geosmin synthase. The Journal of Antibiotics 59(8): 471–479.

Caradec, T., Pupin, M., Vanvlassenbroeck, A., Devignes, M. D., Smaïl-Tabbone, M., Jacques, P., & Leclère, V. (2014). Prediction of monomer isomery in florine: A workflow dedicated to nonribosomal peptide discovery. PLoS ONE 9(1): e85667. Retrieved from 10.1371/journal.pone.0085667

Challis, G. L. (2008). Mining microbial genomes for new natural products and biosynthetic pathways. Microbiology 154(6): 1555–1569.

Challis, G. L., & Ravel, J. (2000). Coelichelin, a new peptide siderophore encoded by the *Streptomyces coelicolor* genome: structure prediction from the sequence of its non-ribosomal peptide synthetase. FEMS Microbiol Lett 187(2): 111–114.

Cheng, Y.-Q., Tang, G.-L., & Shen, B. (2003). Type I polyketide synthase requiring a discrete acyltransferase for polyketide biosynthesis. Proceedings of the National Academy of Sciences 100(6): 3149–3154.

Choi, S. U., Lee, C. K., Hwang, Y. Il, Kinosita, H., & Nihira, T. (2003). γ-Butyrolactone autoregulators and receptor proteins in non-*Streptomyces* actinomycetes producing commercially important secondary metabolites. Archives of Microbiology 180(4): 303–307.

Corre, C., Song, L., O'Rourke, S., Chater, K. F., & Challis, G. L. (2008). 2-Alkyl-4-hydroxymethylfuran-3-carboxylic acids, antibiotic production inducers discovered by Streptomyces coelicolor genome mining. Proceedings of the National Academy of Sciences 105(45): 17510–17515.

Cude, W. N., Mooney, J., Tavanaei, A. A., Hadden, M. K., Frank, A. M., Gulvik, C. A., May, A. L., et al. (2012). Production of the antimicrobial secondary metabolite indigoidine contributes to competitive surface colonization by the marine roseobacter *Phaeobacter* sp. strain Y4I. Applied and Environmental Microbiology 78(14): 4771–4780.

Daduang, R., Kitani, S., Sudoh, Y., Grace Umadhay Pait, I., Thamchaipenet, A., Ikeda, H., Igarashi, Y., et al. (2015). 29-Deoxymaklamicin, a new maklamicin analogue produced by a genetically engineered strain of *Micromonospora* sp. NBRC 110955. Journal of Bioscience

and Bioengineering 120(6): 608–613.

Demain, A. L. (2014). Importance of microbial natural products and the need to revitalize their discovery. Journal of Industrial Microbiology and Biotechnology 41(2): 185-201.

Dias, D. A., Urban, S., & Roessner, U. (2012). A historical overview of natural products in drug discovery. Metabolites 2(4): 303–336.

Donsbach, K., & Rück-Braun, K. (2008). Combinatorial biosynthesis of polyketides. Organic Synthesis Highlights IV: 343–349.

Du, D., Katsuyama, Y., Onaka, H., Fujie, M., Satoh, N., Shin-ya, K., & Ohnishi, Y. (2016). Production of a novel amide-containing polyene by activating a cryptic biosynthetic gene cluster in *Streptomyces* sp. MSC090213JE08. ChemBioChem, 1464–1471.

Du, L., Sánchez, C., & Shen, B. (2001). Hybrid peptide-polyketide natural products: biosynthesis and prospects toward engineering novel molecules. Metabolic engineering, 3(1), 78–95.

Fischbach, M. A., & Walsh, C. T. (2006). Assembly-line enzymology for polyketide and nonribosomal peptide antibiotics: Logic machinery, and mechanisms. Chemical Reviews 106: 3468–3496.

Flärdh, K., & Buttner, M. J. (2009). Streptomyces morphogenetics: dissecting differentiation in a filamentous bacterium. Nature reviews. Microbiology 7(1): 36–49.

Fleming, A. (2001). On the antibacterial action of cultures of a penicillium, with special reference to their use in the isolation of B. influenzae. 1929. Bulletin of the World Health Organization 79(8): 780–790.

Funa, N., Ohnishi, Y., Fujii, I., Shibuya, M., Ebizuka, Y., & Horinouchi, S. (1999). A new pathway for polyketide synthesis in microorganisms. Nature 400(6747): 897–899.

Funabashi, M., Funa, N., & Horinouchi, S. (2008). Phenolic lipids synthesized by type III polyketide synthase confer penicillin resistance on *Streptomyces griseus*. Journal of Biological Chemistry 283(20): 13983–13991.

Gomez-Escribano, J. P., Song, L., Fox, D. J., Yeo, V., Bibb, M. J., & Challis, G. L. (2012b). Structure and biosynthesis of the unusual polyketide alkaloid coelimycin P1, a metabolic product of the *cpk* gene cluster of *Streptomyces coelicolor* M145. Chemical Science 3(9): 2716.

Gottelt, M., Kol, S., Gomez-Escribano, J. P., Bibb, M., & Takano, E. (2010). Deletion of a regulatory gene within the cpk gene cluster reveals novel antibacterial activity in *Streptomyces coelicolor* A3(2). Microbiology 156(8): 2343–2353.

Gu, L., Geders, T. W., Wang, B., Gerwick, W. H., Hakansson, K., Smith, J. L., & Sherman, D. H. (2007). GNAT-like strategy for polyketide chain initiation. Science 318(5852): 970–974.

Han, L., Yang, K., Ramalingam, E., Mosher, R. H., & Vining, L. C. (1994). Cloning and characterization of polyketide synthase genes for jadomycin B biosynthesis in *Streptomyces venezuelae* ISP5230. Microbiology 140(12): 3379–3389.

Hashimoto, K., Nihira, T., Sakuda, S., & Yamada, Y. (1992). IM-2, a butyrolactone autoregulator,

induces production of several nucleoside antibiotics in *Streptomyces* sp. FRI-5. Journal of Fermentation and Bioengineering 73(6): 449–455.

Hertweck, C. (2009). The biosynthetic logic of polyketide diversity. Angewandte Chemie - International Edition 48(26): 4688–4716.

Hosaka, T., Ohnishi-Kameyama, M., Muramatsu, H., Murakami, K., Tsurumi, Y., Kodani, S., Yoshida, M., et al. (2009). Antibacterial discovery in actinomycetes strains with mutations in RNA polymerase or ribosomal protein S12. Nature biotechnology 27(5): 462–464.

Huang, W., Kim, S. J., Liu, J., & Zhang, W. (2015). Identification of the polyketide biosynthetic machinery for the indolizidine alkaloid cyclizidine. Organic Letters 17(21): 5344–5347.

Hughes, L. E., Beck, D. A., & O'Donovan, G. A. (2005). Pathways of pyrimidine salvage in *Streptomyces*. Current Microbiology 50(1): 8–10.

Ikeda, H., Nonomiya, T., Usami, M., Ohta, T., & Omura, S. (1999). Organization of the biosynthetic gene cluster for the polyketide anthelmintic macrolide avermectin in *Streptomyces avermitilis*. Proceedings of the National Academy of Sciences of the United States of America, 96(17), 9509–9514.

Jones, S. E., Ho, L., Rees, C. A., Hill, J. E., Nodwell, J. R., & Elliot, M. A. (2017). Streptomyces exploration is triggered by fungal interactions and volatile signals. eLife, 6, e21738. Retrieved from 10.7554/eLife.21738

Kalan, L., Gessner, A., Thaker, M. N., Waglechner, N., Zhu, X., Szawiola, A., Bechthold, A., et al. (2013). A cryptic polyene biosynthetic gene cluster in *Streptomyces calvus* is expressed upon complementation with a functional *bldA* gene. Chemistry and Biology 20(10): 1214–1224.

Katz, L., & Baltz, R. H. (2016). Natural product discovery: past, present, and future. Journal of Industrial Microbiology & Biotechnology 43: 155–176.

Keatinge-Clay, A. T. (2007). A tylosin ketoreductase reveals how chirality is determined in polyketides. Chemistry and Biology 14(8): 898–908.

Kharel, M. K., Pahari, P., Shepherd, M. D., Tibrewal, N., Nybo, S. E., Shaaban, K.A. & Rohr, J. (2012). Angucyclines: Biosynthesis, mode-of-action, new natural products, and synthesis. Natural Product Reports 29(2): 264.

Kieser, T., Bibb, M. J., Buttner, M. J., Chater, K. F., & Hopwood, D. A. (2000). Practical *Streptomyces* genetics. Norwich: John Innes Foundation.

Kilstrup, M., Hammer, K., Jensen, P. R., & Martinussen, J. (2005). Nucleotide metabolism and its control in lactic acid bacteria. FEMS Microbiology Reviews.

Kitani, S., Doi, M., Shimizu, T., Maeda, A., & Nihira, T. (2010). Control of secondary metabolism by *farX*, which is involved in the γ-butyrolactone biosynthesis of Streptomyces lavendulae FRI-5. Archives of Microbiology 192(3): 211–220.

Kitani, S., Iida, A., Izumi, T. aki, Maeda, A., Yamada, Y., & Nihira, T. (2008). Identification of genes involved in the butyrolactone autoregulator cascade that modulates secondary

metabolism in *Streptomyces lavendulae* FRI-5. Gene 425: 9–16.

Kitani, S., Ikeda, H., Sakamoto, T., Noguchi, S., & Nihira, T. (2009). Characterization of a regulatory gene, *aveR*, for the biosynthesis of avermectin in *Streptomyces avermitilis*. Applied Microbiology and Biotechnology 82(6): 1089–1096.

Kitani, S., Kinoshita, H., Nihira, T., & Yamada, Y. (1999). In vitro analysis of the butyrolactone autoregulator receptor protein (FarA) of *Streptomyces lavendulae* FRI-5 reveals that FarA acts as a DNA-binding transcriptional regulator that controls its own synthesis. J. Bacteriol. 181(16): 5081–4.

Kitani, S., Yamada, Y., & Nihira, T. (2001). Gene replacement analysis of the butyrolactone autoregulator receptor (FarA) reveals that FarA acts as a novel regulator in secondary metabolism of *Streptomyces lavendulae* FRI-5. Journal of Bacteriology, 183(14), 4357–4363.

Koehn, F. E., & Carter, G. T. (2005). The evolving role of natural products in drug discovery. Nature reviews. Drug discovery 43:206–220.

Komatsu, M., Komatsu, K., Koiwai, H., Yamada, Y., Kozone, I., Izumikawa, M., Hashimoto, J., et al. (2013). Engineered *Streptomyces avermitilis* host for heterologous expression of biosynthetic gene cluster for secondary metabolites. ACS Synthetic Biology 2(7): 384–396.

Komatsu, M., Tsuda, M., Omura, S., Oikawa, H., & Ikeda, H. (2008). Identification and functional analysis of genes controlling biosynthesis of 2-methylisoborneol. Proceedings of the National Academy of Sciences of the United States of America 105(21): 7422–7427.

Komatsu, M., Uchiyama, T., Omura, S., Cane, D. E., & Ikeda, H. (2010). Genome-minimized *Streptomyces* host for the heterologous expression of secondary metabolism. Proceedings of the National Academy of Sciences of the United States of America 107(6): 2646–51.

Kormanec, J., Novakova, R., Mingyar, E., & Feckova, L. (2014). Intriguing properties of the angucycline antibiotic auricin and complex regulation of its biosynthesis. Applied Microbiology and Biotechnology 98(1): 45–60.

Kurniawan, Y. N., Kitani, S., Iida, A., Maeda, A., Lycklama a Nijeholt, J., Lee, Y. J., & Nihira, T. (2016). Regulation of production of the blue pigment indigoidine by the pseudo γ-butyrolactone receptor FarR2 in *Streptomyces lavendulae* FRI-5. Journal of Bioscience and Bioengineering 121(4): 372–379.

Kurniawan, Y. N., Kitani, S., Maeda, A., & Nihira, T. (2014). Differential contributions of two SARP family regulatory genes to indigoidine biosynthesis in *Streptomyces lavendulae* FRI-5. Applied Microbiology and Biotechnology 98(23): 9713–9721.

Kuščer, E., Coates, N., Challis, I., Gregory, M., Wilkinson, B., Sheridan, R., & Petković, H. (2007). Roles of *rapH* and *rapG* in positive regulation of rapamycin biosynthesis in Streptomyces hygroscopicus. Journal of Bacteriology 189(13): 4756–4763.

Laureti, L., Song, L., Huang, S., Corre, C., Leblond, P., Challis, G. L., & Aigle, B. (2011). Identification of a bioactive 51-membered macrolide complex by activation of a silent polyketide synthase in *Streptomyces ambofaciens*. Proceedings of the National Academy of Sciences of the United States of America 108(15): 6258–6263.

Lautru, S., & Challis, G. L. (2004). Substrate recognition by nonribosomal peptide synthetase multi-enzymes. Microbiology 150(6): 1629–1636.

Li, J., Kim, S. G., & Blenis, J. (2014). Rapamycin: One drug, many effects. Cell Metabolism 19(3): 373-379.

Li, J. W.-H., & Vederas, J. C. (2009). Drug Discovery and Natural Products: End of an era or an endless frontier? Science 325: 161–165.

Li, P., Li, J., Guo, Z., Tang, W., Han, J., Meng, X., Hao, T., et al. (2015). An efficient blue-white screening based gene inactivation system for *Streptomyces*. Applied Microbiology and Biotechnology 99(4): 1923–1933.

Lin, X., Hopson, R., & Cane, D. E. (2006). Genome mining in *Streptomyces coelicolor:* molecular cloning and characterization of a new sesquiterpene synthase. Journal of the American Chemical Society 128(18): 6022–6023.

Manteca, A., & Sanchez, J. (2010). Streptomyces developmental cycle and secondary metabolite production. Applied and environmental microbiology 75(9): 560–566.

Maplestone, R. A., Stone, M. J., & Williams, D. H. (1992). The evolutionary role of secondary metabolites--a review. Gene 115: 151–7.

Martín, J. F., & Liras, P. (2010). Engineering of regulatory cascades and networks controlling antibiotic biosynthesis in *Streptomyces*. Current Opinion in Microbiology 13(3): 263-273.

Matsuno, K., Yamada, Y., Lee, C. K., & Nihira, T. (2004). Identification by gene deletion analysis of barB as a negative regulator controlling an early process of virginiamycin biosynthesis in *Streptomyces virginiae*. Archives of Microbiology 181(1): 52–59.

Metsä-Ketelä, M., Ylihonko, K., & Mäntsälä, P. (2004). Partial activation of a silent angucycline-type gene cluster from a rubromycin beta producing *Streptomyces* sp. PGA64. The Journal of Antibiotics 57(8): 502–510.

Michel, G. W. (1977). Nystatin. Analytical Profiles of Drug Substances, 6, 341–421.

Min, H., Varoglu, M., & Sherman, D. H. (2000). Structural modeling and site-directed mutagenesis of the actinorhodin β-ketoacyl-acyl carrier protein synthase. Journal of Bacteriology 182(9): 2619–2623.

Mingyar, E., Feckova, L., Novakova, R., Bekeova, C., & Kormanec, J. (2014). A γ-butyrolactone autoregulator-receptor system involved in the regulation of auricin production in *Streptomyces aureofaciens* CCM 3239. Applied Microbiology and Biotechnology 99(1): 309–325.

Mizutani, S., Odai, H., Masuda, T., Iijima, M., Osono, M., Hamada, M., Naganawa, H., et al. (1989). Biological activities of IC201 ((3S,8E)-1,3-dihydroxy-8-decen-5-one), a low molecular weight immunomodulator produced by *Streptomyces*. The Journal of Antibiotics 42(6): 952–959.

Müller, M., Ausländer, S., Ausländer, D., Kemmer, C., & Fussenegger, M. (2012). A novel reporter system for bacterial and mammalian cells based on the non-ribosomal peptide

indigoidine. Metabolic Engineering 14(4): 325–335.

Nett, M., Ikeda, H., & Moore, B. S. (2009). Genomic basis for natural product biosynthetic diversity in the actinomycetes. Natural Product Reports, 26(11), 1362-1184.

Newman, D. J., & Cragg, G. M. (2016). Natural Products as Sources of New Drugs from 1981 to 2014. Journal of Natural Products 79(3): 629-661.

Nishida, H., Ohnishi, Y., Beppu, T., & Horinouchi, S. (2007). Evolution of γ-butyrolactone synthases and receptors in *Streptomyces*. Environmental Microbiology 9(8): 1986–1994.

Novakova, R., Odnogova, Z., Kutas, P., Feckova, L., & Kormanec, J. (2010). Identification and characterization of an indigoidine-like gene for a blue pigment biosynthesis in *Streptomyces aureofaciens* CCM 3239. Folia Microbiologica 55(2): 119–125.

Ochi, K., & Hosaka, T. (2013). New strategies for drug discovery: Activation of silent or weakly expressed microbial gene clusters. Applied Microbiology and Biotechnology 97(1): 87–98.

Ohnishi, Y., Kameyama, S., Onaka, H., & Horinouchi, S. (1999). The A-factor regulatory cascade leading to streptomycin biosynthesis in *Streptomyces griseus*: Identification of a target gene of the A-factor receptor. Molecular Microbiology 34(1): 102–111.

Ohtsubo, Y., Ikeda-Ohtsubo, W., Nagata, Y., & Tsuda, M. (2008). GenomeMatcher: A graphical user interface for DNA sequence comparison. BMC Bioinformatics 9(1): 376.

Omura, S. (2011). Microbial metabolites: 45 years of wandering, wondering and discovering. Tetrahedron 67(35): 6420–6459.

Onaka, H., Ozaki, T., Mori, Y., Izawa, M., Hayashi, S., & Asamizu, S. (2015). Mycolic acid-containing bacteria activate heterologous secondary metabolite expression in *Streptomyces lividans*. The Journal of Antibiotics 68(2): 1–4.

Pait, I. G. U., Kitani, S., Kurniawan, Y. N., Asa, M., Iwai, T., Ikeda, H., & Nihira, T. (2017). Identification and characterization of *lbpA* , an indigoidine biosynthetic gene in the γ-butyrolactone signaling system of *Streptomyces lavendulae* FRI-5. Journal of Bioscience and Bioengineering 124(4): 369-375

Palmu, K., Ishida, K., Mäntsälä, P., Hertweck, C., & Metsä-Ketelä, M. (2007). Artificial reconstruction of two cryptic angucycline antibiotic biosynthetic pathways. ChemBioChem 8(13): 1577–1584.

Pettit, R. K. (2011). Small-molecule elicitation of microbial secondary metabolites. Microbial Biotechnology 4(4): 471-478.

Piel, J. (2010). Biosynthesis of polyketides by *trans*-AT polyketide synthases. Natural product reports 27(7): 996–1047.

Priestle, J. P., Grütter, M. G., White, J. L., Vincent, M. G., Kania, M., Wilson, E., Jardetzky, T. S., et al. (1987). Three-dimensional structure of the bifunctional enzyme N-(5'-phosphoribosyl)anthranilate isomerase-indole-3-glycerol-phosphate synthase from *Escherichia coli*. Proceedings of the National Academy of Sciences of the United States of America 84(16): 5690–5694.

Pulsawat, N., Kitani, S., & Nihira, T. (2007). Characterization of biosynthetic gene cluster for the production of virginiamycin M, a streptogramin type A antibiotic, in *Streptomyces virginiae*. Gene 393: 31–42.

Rateb, M. E., Houssen, W. E., Harrison, W. T. A., Deng, H., Okoro, C. K., Asenjo, J. A., Andrews, B. A., et al. (2011). Diverse metabolic profiles of a *Streptomyces* strain isolated from a hyper-arid environment. Journal of Natural Products 74(9): 1965–1971.

Reid, R., Piagentini, M., Rodriguez, E., Ashley, G., Viswanathan, N., Carney, J., Santi, D. V., et al. (2003). A model of structure and catalysis for ketoreductase domains in modular polyketide synthases. Biochemistry 42(1): 72–79.

Reiss, T., & Breit, B. (2009). A Unified Strategy for the Stereospecific Construction of Propionates and Acetate-Propionates Relying on a Directed Allylic Substitution. Chemistry - A European Journal 15(26): 6345–6348.

Retzlaff, L., & Distler, J. (1995). The regulator of streptomycin gene expression, StrR, of Streptomyces griseus is a DNA binding activator protein with multiple recognition sites. Molecular Microbiology 18(1): 151–162.

Reverchon, S., Rouanet, C., Expert, D., & Nasser, W. (2002). Characterization of indigoidine biosynthetic genes in *Erwinia chrysanthemi* and role of this blue pigment in pathogenicity. Journal of Bacteriology, 184(3), 654–665.

Rohr, J., & Thiericke, R. (1992). Angucycline group antibiotics. Natural product reports 9(2): 103–137.

Scherlach, K., & Hertweck, C. (2009). Triggering cryptic natural product biosynthesis in microorganisms. Organic & Biomolecular Chemistry 7(9): 1753.

Seyedsayamdost, M. R. (2014). High-throughput platform for the discovery of elicitors of silent bacterial gene clusters. Proceedings of the National Academy of Sciences of the United States of America 111(20): 7266–7271.

Sharan, S. K., Thomason, L. C., Kuznetsov, S. G., & Court, D. L. (2009). Recombineering: a homologous recombination-based method of genetic engineering. Nature protocols 4(2): 206–223.

Shima, J., Hesketh, A., Okamoto, S., Kawamoto, S., & Ochi, K. (1996). Induction of actinorhodin production by *rpsL* (encoding ribosomal protein S12) mutations that confer streptomycin resistance in *Streptomyces lividans* and *Streptomyces coelicolor* A3(2). Journal of Bacteriology 178(24): 7276–7284.

Sieber, S. A., & Marahiel, M. A. (2005). Molecular mechanisms underlying nonribosomal peptide synthesis: Approaches to new antibiotics. Chemical Reviews 105(2): 715-738.

Sinha, S. C., & Smith, J. L. (2001). The PRT protein family. Current Opinion in Structural Biology 11(6): 733-739.

Skiba, M. A., Sikkema, A. P., Fiers, W. D., Gerwick, W. H., Sherman, D. H., Aldrich, C. C., & Smith, J. L. (2016). Domain Organization and Active Site Architecture of a Polyketide Synthase C -methyltransferase. ACS Chemical Biology 11(12): 3319–3327.

Smith, S., & Tsai, S.-C. (2007). The type I fatty acid and polyketide synthases: a tale of two megasynthases. Nat. Prod. Rep., 24(5): 1041–1072.

Staunton, J., & Weissman, K. J. (2001). Polyketide biosynthesis: a millennium review. Natural product reports 18(4): 380–416.

Staunton, J., & Weissman, K. J. (2001). Polyketide biosynthesis: a millennium review. Natural Product Reports 18(4): 380–416.

Strieker, M., Tanović, A., & Marahiel, M. A. (2010). Nonribosomal peptide synthetases: Structures and dynamics. Current Opinion in Structural Biology 20(2): 234-240.

Strobel, T., Al-Dilaimi, A., Blom, J., Gessner, A., Kalinowski, J., Luzhetska, M., Pühler, A., et al. (2012). Complete genome sequence of *Saccharothrix espanaensis* DSM 44229T and comparison to the other completely sequenced Pseudonocardiaceae. BMC Genomics 13(1): 465.

Takahashi, H., Kumagai, T., Kitani, K., Mori, M., Matoba, Y., & Sugiyama, M. (2007). Cloning and characterization of a Streptomyces single module type non-ribosomal peptide synthetase catalyzing a blue pigment synthesis. Journal of Biological Chemistry 282(12): 9073–9081.

Takano, E. (2006). γ-Butyrolactones: Streptomyces signalling molecules regulating antibiotic production and differentiation. Current Opinion in Microbiology 9(3): 287–294.

Tanaka, Y., Kasahara, K., Hirose, Y., Murakami, K., Kugimiya, R., & Ochi, K. (2013). Activation and products of the cryptic secondary metabolite biosynthetic gene clusters by rifampin resistance (rpoB) mutations in actinomycetes. Journal of Bacteriology 195(13): 2959–2970.

Tanaka, Y., Tokuyama, S., & Ochi, K. (2009). Activation of secondary metabolite-biosynthetic gene clusters by generating rsmG mutations in Streptomyces griseus. The Journal of Antibiotics 62(12): 669–673.

Tarazona, G., Schleissner, C., Rodriguez, P., Perez, M., Cañedo, L., & Cuevas, C. (2017). Streptenols F–I Isolated from the Marine-Derived *Streptomyces misionensis* BAT-10-03-023. Journal of Natural Products 80(4): 1034–1038.

Thanapipatsiri, A., Claesen, J., Gomez-Escribano, J.-P., Bibb, M., & Thamchaipenet, A. (2015). A *Streptomyces coelicolor* host for the heterologous expression of Type III polyketide synthase genes. Microbial Cell Factories, 14(1), 145.

Tobert, J. a. (2003). Lovastatin and beyond: the history of the HMG-CoA reductase inhibitors. Nature reviews. Drug discovery, 2(7), 517–26.

Traxler, M. F., Watrous, J. D., Alexandrov, T., Dorrestein, P. C., & Kolter, R. (2013). Interspecies interactions stimulate diversification of the *Streptomyces coelicolor* secreted metabolome. mBio, 4(4):e00459-13. Retrieved from doi:10.1128/mBio.00459-13.

Waki, M., Nihira, T., & Yamada, Y. (1997). Cloning and characterization of the gene (*farA*) encoding the receptor for an extracellular regulatory factor (IM-2) from *Streptomyces* sp. strain FRI-5. Journal of Bacteriology 179(16): 5131–5137.

Wesener, S. R., Potharla, V. Y., & Cheng, Y. Q. (2011). Reconstitution of the FK228 biosynthetic

pathway reveals cross talk between modular polyketide synthases and fatty acid synthase. Applied and Environmental Microbiology 77(4): 1501–1507.

van Wezel, G. P., & McDowall, K. J. (2011). The regulation of the secondary metabolism of Streptomyces: new links and experimental advances. Natural Product Reports 28(7): 1311.

Wilson, D. J., Xue, Y., Reynolds, K. A., & Sherman, D. H. (2001). Characterization and analysis of the *pikD* regulatory factor in the pikromycin biosynthetic pathway of *Streptomyces venezuelae*. Journal of Bacteriology 183(11): 3468–3475.

Wu, J., Zaleski, T. J., Valenzano, C., Khosla, C., & Cane, D. E. (2005). Polyketide double bond biosynthesis. Mechanistic analysis of the dehydratase-containing module 2 of the picromycin/methymycin polyketide synthase. Journal of the American Chemical Society 127(49): 17393–17404.

Yagüe, P., Lopez-Garcia, M. T., Rioseras, B., Sanchez, J., & Manteca, A. (2012). New insights on the development of *Streptomyces* and their relationships with secondary metabolite production. Current Trends in Microbiology 8: 65–73.

Yamada, Y., Arima, S., Nagamitsu, T., Johmoto, K., Uekusa, H., Eguchi, T., Shin-Ya, K., et al. (2015). Novel terpenes generated by heterologous expression of bacterial terpene synthase genes in an engineered *Streptomyces* host. The Journal of antibiotics 68: 1–10.

Yamada, Y., Kuzuyama, T., Komatsu, M., Shin-ya, K., Omura, S., Cane, D. E., & Ikeda, H. (2015). Terpene synthases are widely distributed in bacteria. Proceedings of the National Academy of Sciences 112(3): 857–862.

Yanai, K., Murakami, T., & Bibb, M. (2006). Amplification of the entire kanamycin biosynthetic gene cluster during empirical strain improvement of *Streptomyces kanamyceticus*. Proceedings of the National Academy of Sciences of the United States of America 103(25): 9661–9666.

Ye, S., Molloy, B., Braña, A. F., Zabala, D., Olano, C., Cortes, J., Moris, F., et al. (2017). Identification by genome mining of a type I polyketide gene cluster from *Streptomyces argillaceus* involved in the biosynthesis of pyridine and piperidine alkaloids argimycins P. Frontiers in Microbiology 8(FEB): 1–18.

Yu, D., Xu, F., Zhang, S., Shao, L., Wang, S., & Zhan, J. (2013). Characterization of a methyltransferase involved in herboxidiene biosynthesis. Bioorganic and Medicinal Chemistry Letters 23(20): 5667–5670.

Zarins-Tutt, J. S., Barberi, T. T., Gao, H., Mearns-Spragg, A., Zhang, L., Newman, D. J., & Goss, R. J. M. (2016). Prospecting for new bacterial metabolites: a glossary of approaches for inducing, activating and upregulating the biosynthesis of bacterial cryptic or silent natural products. Nat. Prod. Rep. 33(1): 54–72.

Zettler, J., Xia, H., Burkard, N., Kulik, A., Grond, S., Heide, L., & Apel, A. K. (2014). New Aminocoumarins from the rare actinomycete *Catenulispora acidiphila* DSM 44928: Identification, structure elucidation, and heterologous production. ChemBioChem 15(4): 612–621.

Zhang, W., Li, S., Zhu, Y., Chen, Y., Chen, Y., Zhang, H., Zhang, G., et al. (2014). Heronamides

D-F, polyketide macrolactams from the deep-sea-derived *Streptomyces* sp. SCSIO 03032. Journal of Natural Products 77(2): 388–391.

List of Publications

Pait, I. G. U., Kitani, S., Kurniawan, Y. N., Asa, M., Iwai, T., Ikeda, H., & Nihira, T. (2017). Identification and characterization of *lbpA* , an indigoidine biosynthetic gene in the γ-butyrolactone signaling system of *Streptomyces lavendulae* FRI-5. Journal of Bioscience and Bioengineering 124(4): 369-375

Pait, I. G. U., Kitani, S., Wahidah, F. R., Ulanova, D., Arai, M., Ikeda, H., & Nihira, T. (2017). Discovery of a new diol-containing polyketide by heterologous expression of a silent biosynthetic gene cluster from *Streptomyces lavendulae* FRI-5. Advance online publication. doi: https://doi.org/10.1007/s10295-017-1997-x

Acknowledgements

Pursuing this PhD degree has been a period of intense learning and growth for me, not only academically, but also on a personal level. The path was never straight, in fact, the first few years were mostly dead ends. So as I write this dissertation, I would like to reflect on the people whose guidance, support and companionship have helped me bounce off every roadblock, and find my direction amidst all unexpected twists and spirals of this PhD journey.

First of all, I would like to express my sincerest appreciation to my advisor Prof. Takuya Nihira for encouraging my research in his laboratory and for the chance to come to Japan, which opened up so many other opportunities for me. His continuous support and advice for my research and personal life have been priceless. I would also like to thank my co-advisor Dr. Shigeru Kitani for his unwavering patience and efforts in guiding me with my project and teaching me how to be a good researcher. His invaluable inputs during countless discussions have helped my research move forward, and also helped me improve my presentation and technical writing skills. I would also like to acknowledge the helpful suggestions and encouragement from Dr. Hiroshi Kinoshita especially during the difficult periods of this work.

I also extend my thanks to my committee members Prof. Toshiya Muranaka and Prof. Hajime Watanabe whose comments and questions during my midterm defense improved the story of this PhD dissertation. I also would like to acknowledge our collaborators Prof. Haruo Ikeda of Kitasato University who kindly provided us with the SUKA22 strain and Prof. Masayoshi Arai of the Graduate School of Pharmaceutical Sciences, Osaka University for their significant contributions to the completion of this work. I also would like to thank Inoue Sensei of the Analytical Instrumentation Facility for her guidance during the structural elucidation of my

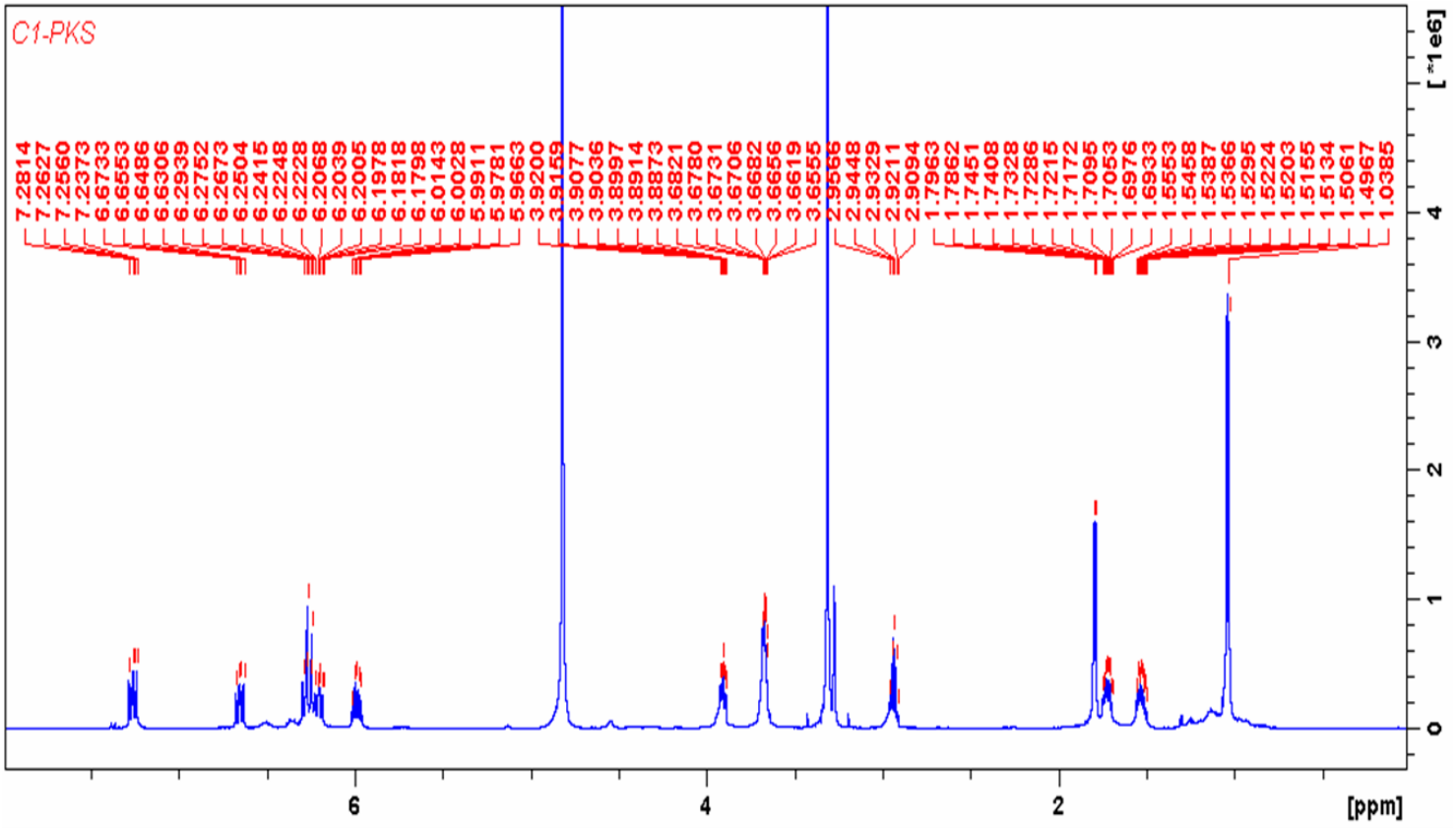
compound. And of course, without the financial assistance of Japan Ministry of Education, Culture, Sports, Science and Technology (MEXT), this experience would also be impossible.

I am very grateful for my labmates and other members of the International Center for Biotechnology for promoting a very supportive and stimulating working environment in this 6 years. Special thanks to all my senpai (Ai, Pang, Hansky, Dian, Fitri, Andi) for their patience in teaching me and to my fellow international students especially Thao, Havy, Tanty and Nindy for these many years of "fun" in the lab. I will never forget all the lessons, troubleshooting guidance, discussions over lunch, laughter and especially the friendship that supported me especially during moments of trouble. For all my Japanese labmates who were always friendly and ready to help me not just with lab matters but also with learning Japanese culture, the language and job-hunting, thank you from the bottom of my heart.

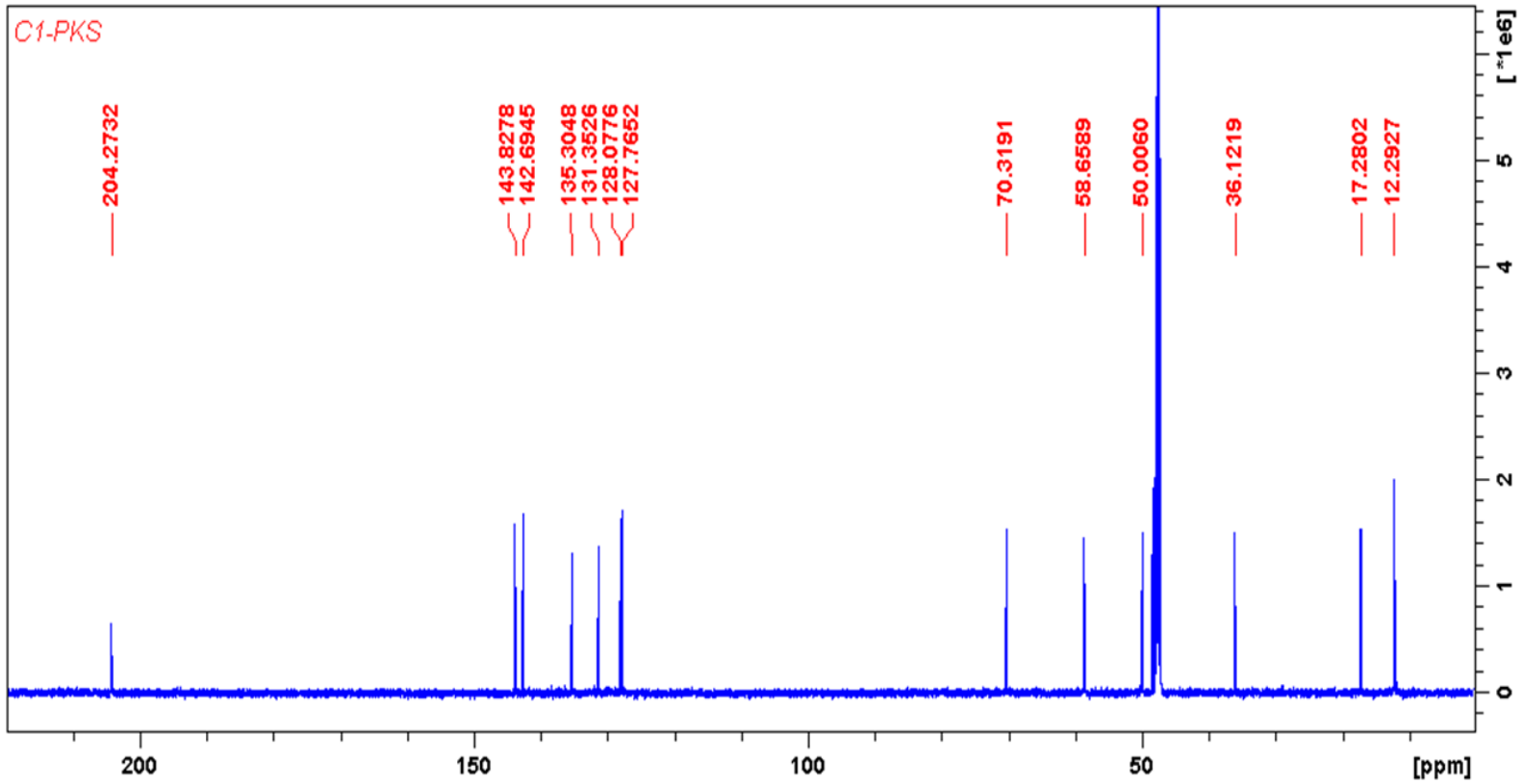
To the sweet kids of EPION, my friends, to the Filipino Students Society of Osaka (FILSSO), my family and loved ones who were always cheering me on, pouring prayers and never-ending support, I would not have made it this far alone and will be forever indebted.

Finally, I am thankful for God's grace and tiny miracles as He works in the background of every day.

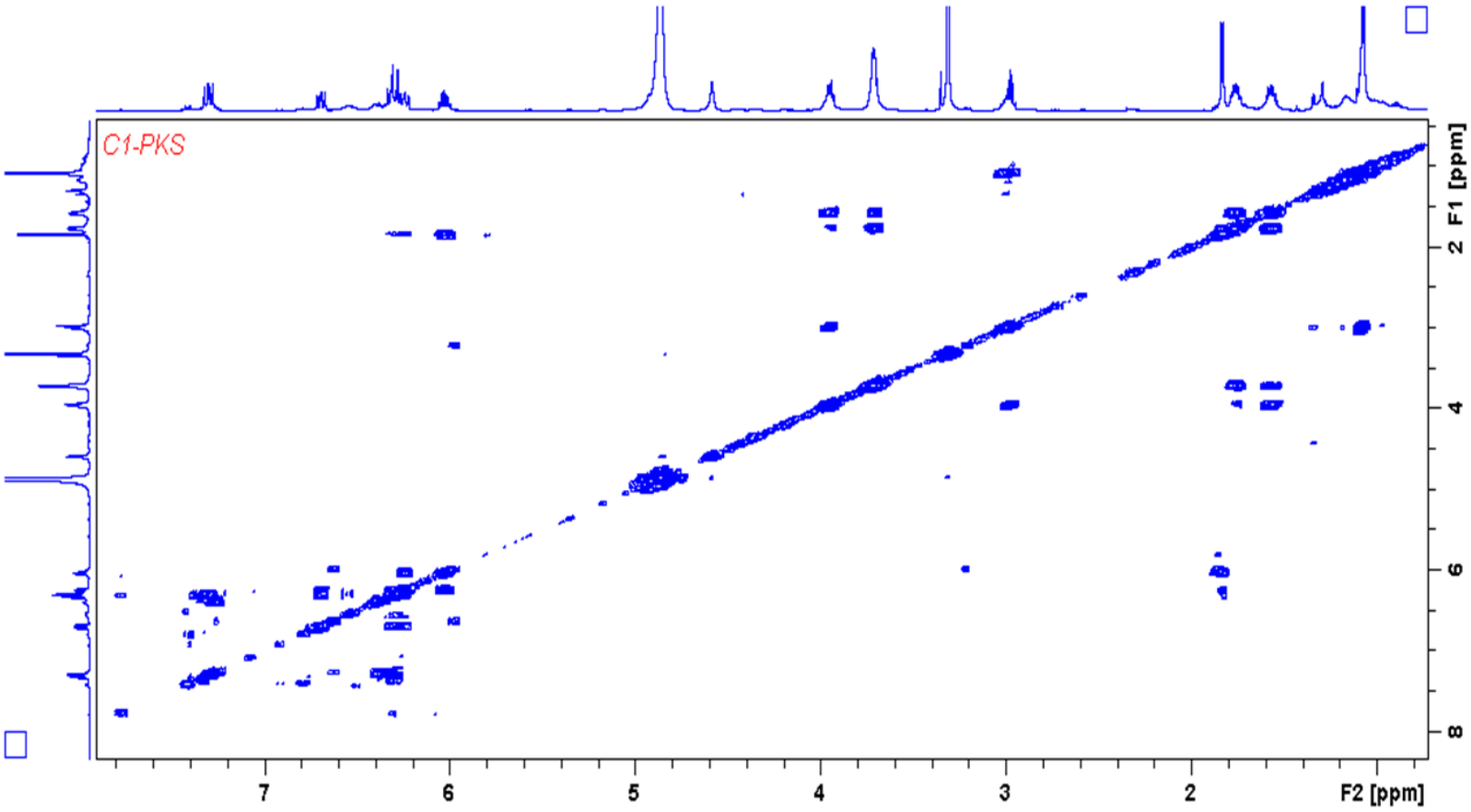
Appendix 1. ^{1}H NMR spectrum (600 MHz) of lavendiol in CD_3OD



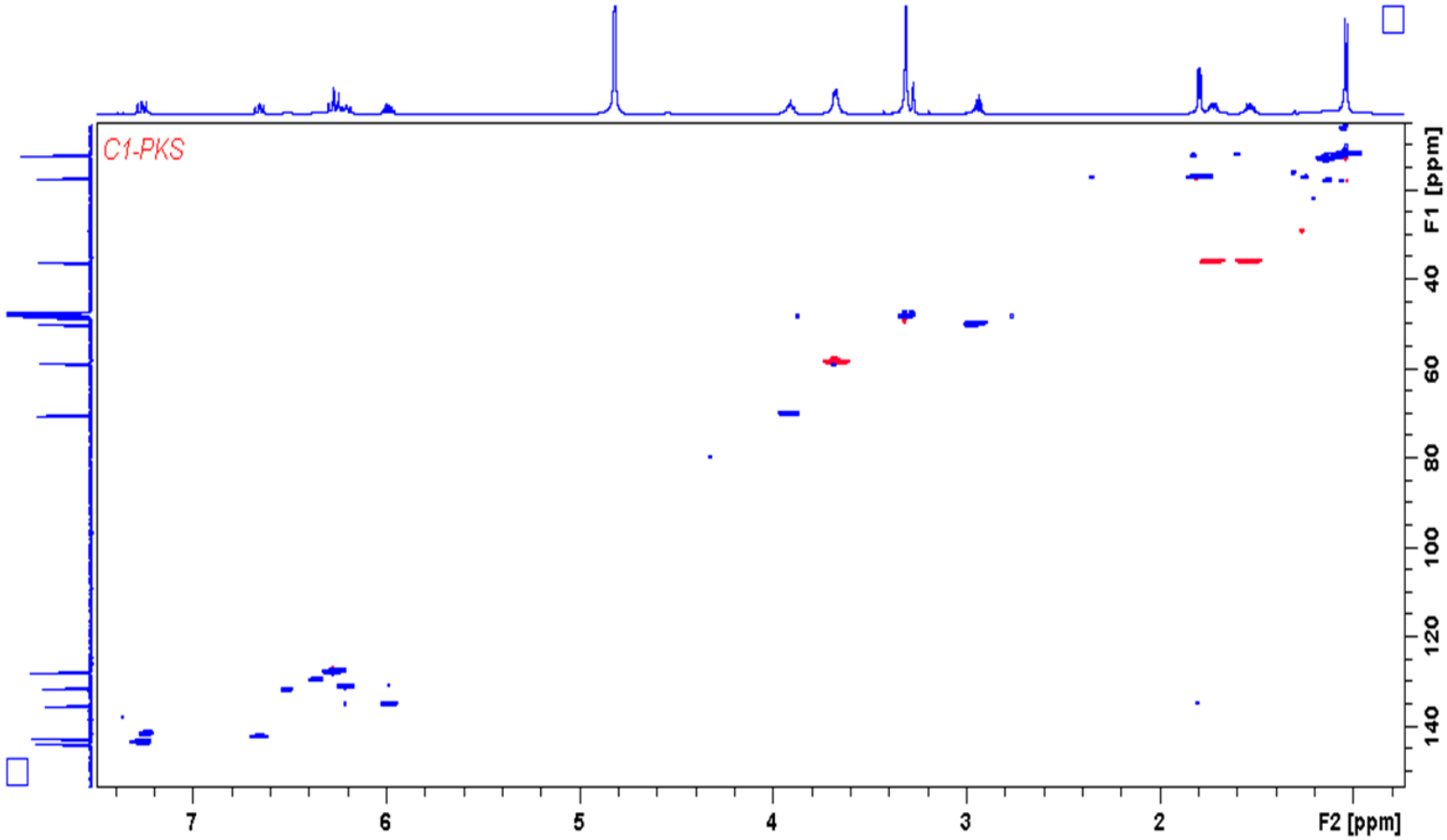
Appendix 2. ^{13}C NMR spectrum (150 MHz) of lavendiol in CD_3OD



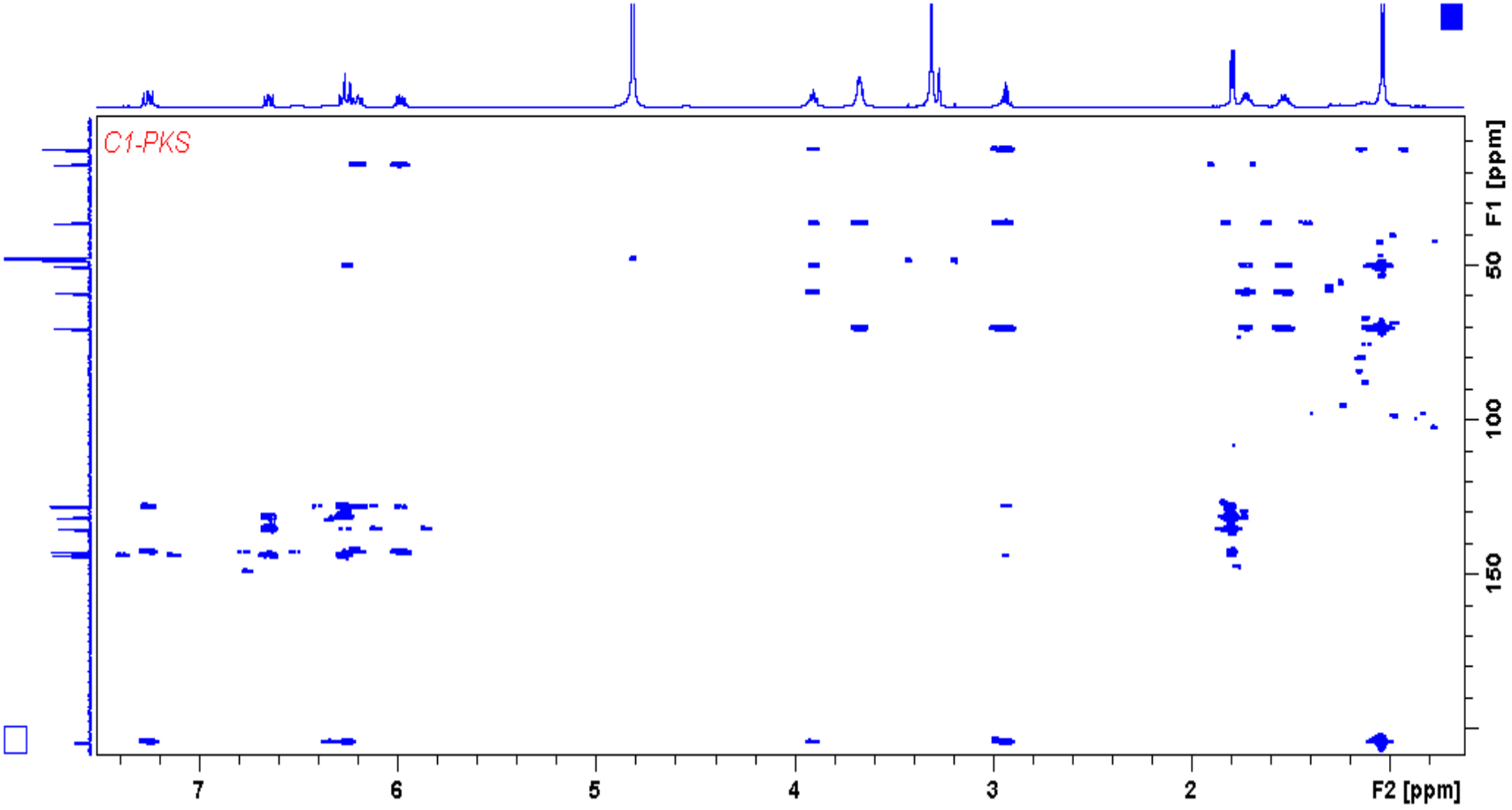
Appendix 3. COSY spectrum of lavendiol in CD_3OD



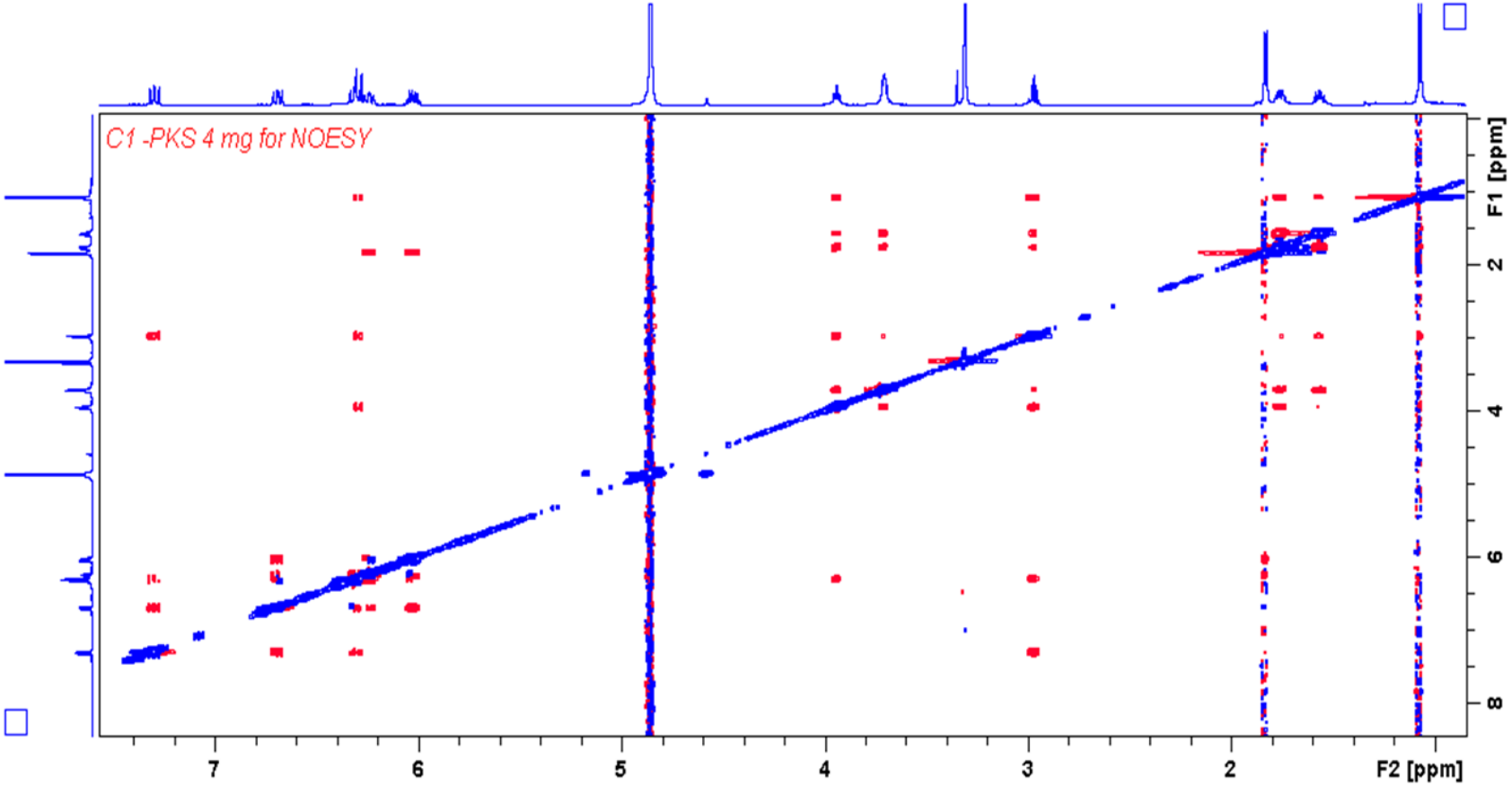
Appendix 4. HSQC spectrum of lavendiol in CD_3OD



Appendix 5. HMBC spectrum of lavendiol in CD_3OD

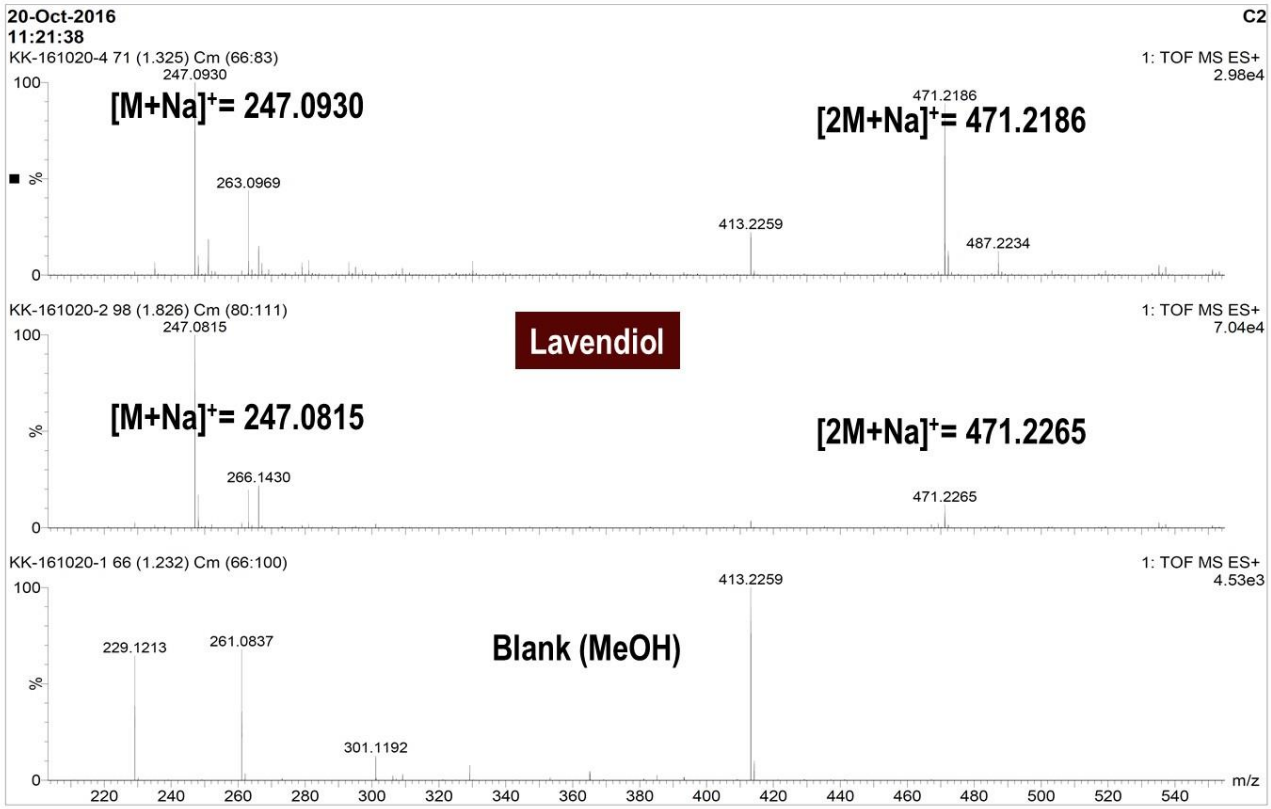


Appendix 6. NOESY spectrum of lavendiol in CD_3OD



Appendix 7. Low resolution ESIMS analysis of lavendiol

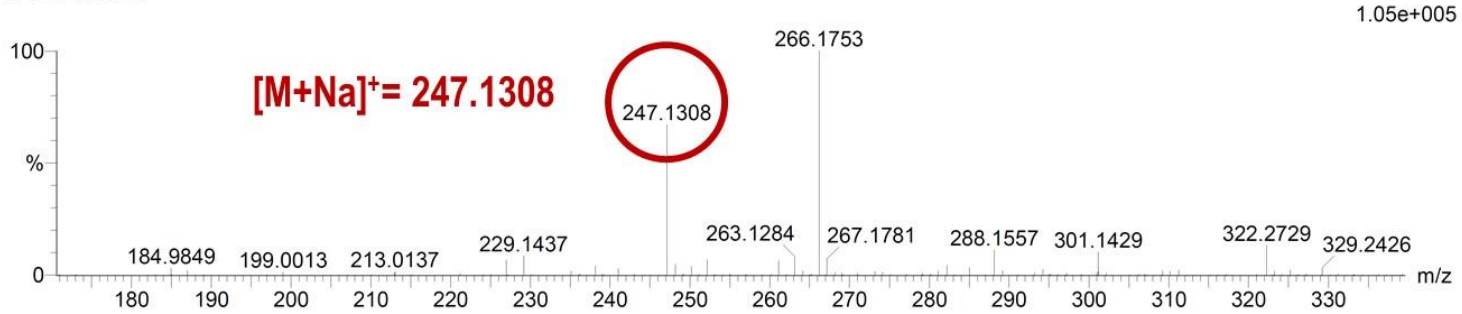
(Waters Q-TOF Ultima ESI (Positive Ion Mode)



Appendix 7. High resolution ESIMS analysis of lavendiol

(Waters Q-TOF Ultima ESI (Positive Ion Mode)

Monoisotopic Mass, Even Electron Ions
45 formula(e) evaluated with 6 results within limits (up to 50 closest results for each mass)
Elements Used:
C: 10-20 H: 15-30 N: 0-2 O: 0-5 Na: 0-1
20-Oct-2016
12:00:36
1: TOF MS ES+
KK-161020-10 60 (1.492) AM (Cen,4, 80.00, Ar,5000.0,556.28,0.18,LS 5); Sm (SG, 2x3.00); Cm (37:65)



Minimum:				-1.5						
Maximum:		20.0	50.0	23.0						
Mass	Calc. Mass	mDa	PPM	DBE	i-FIT	Formula				
247.1308	247.1310	-0.2	-0.8	3.5	4674.2	C13	H20	O3	Na	
	247.1294	1.4	5.7	2.5	4082.0	C10	H19	N2	O5	
	247.1334	-2.6	-10.5	6.5	6681.3	C15	H19	O3		
	247.1235	7.3	29.5	11.5	7806.8	C17	H15	N2		
	247.1211	9.7	39.3	8.5	5804.5	C15	H16	N2	Na	
	247.1422	-11.4	-46.1	3.5	4164.0	C12	H20	N2	O2	Na

Appendix 8. IR Spectrum of lavendiol (in methylene chloride)