Manipulation of *Streptomyces clavuligerus* for the purification of δ-(L-α-aminoadipyl)-L-cysteinyl-D-valine synthetase and the mobilization of plasmid DNA

by ©Marcus A. Moore

A Thesis submitted to the

School of Graduate Studies

in partial fulfillment of the requirements for the degree of

Masters of Science

Department of Biology

Memorial University of Newfoundland

October 2015

St. John's, Newfoundland and Labrador

ABSTRACT

Streptomyces clavuligerus is a filamentous bacterium renowned for its production of medicinally relevant natural products such as cephamycin C and clavulanic acid. Genome sequencing of this organism has revealed a treasure trove of cryptic/silent pathways, including a large linear plasmid with numerous secondary metabolite (SM) gene clusters. This research sought to transfer the *S. clavuligerus* megaplasmid (pSCL4) to a SM overproducing "super" strain, *Streptomyces coelicolor* (M1154), for the purpose of activating putative cryptic gene clusters encoded by it. Furthermore, synthetic constructs were designed for the overexpression and purification of mega-enzymes involved in producing important antibiotics in *Streptomyces* spp. The nonribosomal peptide synthetase, δ-(L-α-aminoadipyl)-L-cysteinyl-D-valine Synthetase (ACVS), is a core enzyme in the biosynthesis of bacterial cephalosporins and penicillins. ACVS was fused to a 8× His-tag and expressed using the constitutive *ermE*p* promoter and was purified using nickel affinity chromatography. Reported here is the first targeted isolation of ACVS and the tools developed in this study can be applied to a variety of enzymes and SM gene clusters.

ACKNOWLEDGEMENTS

The work completed and presented within this thesis was conducted between May 2012 and January 2015, with the majority being done at Memorial University of Newfoundland. I would like to thank Dr. Kapil Tahlan for developing the idea for this research and his continuous support through this story as my supervisor. I would also like to acknowledge the helpful insight and support provided by my supervisory committee, Drs. Dawn Bignell and Andrei Igamberdiev. Work on this thesis was partially conducted as a visiting student to the Laboratorio Nacional de Genomica Para la Biodiversidad in Irapuato México; a special thanks to Dr. Francisco "Paco" Barona-Gomez for hosting me during this time. I would also like to thank Dr. Pablo Cruz-Morales for his helpful comments and conversation, and Hilda Ramos-Aboites for technical support during my time in México. Finally, I would like to acknowledge the Natural Sciences and Engineering Research Council of Canada for gracious financial support through the Canada Graduate Scholarship and Michael Smith Foreign Studies Supplement.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF SYMBOLS, NOMENCLATURE, AND/OR ABBREVIATIONS

This is a list of symbols, nomenclature or abbreviations used throughout this thesis. Acronyms common to life sciences are omitted.

Abbreviation	Definition
aac(3)IV	Apramycin resistance gene
ACP	Acyl carrier protein
ACVS	δ-(L-α-aminoadipyl)-L-cysteinyl-D-valin Synthetase
amp^r	Ampicillin resistance
Apra	Apramycin
cam^r	Chloramphenicol resistance
DCW	Dry cell weight
DTT	Dithiothreitol
EDTA	Ethylenediaminetetraacetic Acid
eGFP	enhanced Green fluorescent protein
FAS	Fatty acid synthases
GI	Genetic Identifiers, Genbank Accession Number
HPLC	High performance liquid chromatography
his	8x His-tag
ISP-4	International *Streptomyces* Project Media 4
LacZ	β-galactosidase gene
LB	Lysogeny broth
MSC	Maximum survivable concentration
MUSCLE	Multiple Sequence Comparison by Log-Expectation
NP	Natural Products
NCBI	National Center for Biotechnology Information
neo	Neomycin resistance gene (kanamycin)
NRPS	Non-ribosomal peptide synthetase
OD	Optical density
ori	Origin of replication
PAGE	Polyacrylamide gel electrophoresis
PBS	Phosphate buffered saline
PCP	Peptidyl carrier protein
PCR	Polymerase chain reaction
PKS	Polyketide synthase
R2YE	Defined *Streptomyces* recovery media
*rpo*B	RNA polymerase β-subunit gene

rpsL	Ribosome protein S12 gene
SARP	*Streptomyces* antibiotic regulatory protein
SDS	Sodium dodecyl sulfate
SFM	Mannitol soya flour medium/Soy Flour Media
SM	Secondary metabolites
TA	Toxin/Antitoxin
TBE	Tris-borate-EDTA
tet	Tetracycline resistance cassette
tev	Tobacco etch virus protease recognition site
tfd	Major transcription terminator phage fd
Tm	PCR annealing temperature
TSB	Trypticase soy broth
TSBS	TSB, with starch
tsr	Thiostrepton resistance cassette
WT	Wild type
YEME	Yeast Extract-malt Extract Medium
YT	Yeast extract-tryptone media

LIST OF APPENDICES

Chapter 1

1.01 Introduction

1.01.01 Secondary metabolites and natural products

Many bioactive chemicals (natural products) with medicinal uses, such as antibiotics (e.g. penicillin and vancomycin), are produced via secondary metabolism by microorganisms. However, their medicinal attributes are not limited to antibiotics: natural products (NP) have been found to act as immune-suppressants, anti-tumor agents, and as tools for developing other drugs (Newman and Cragg 2007). In general, secondary metabolites (SMs) are named so because they are often produced in stationary phase, and are thought to be non-essential for life. Secondary metabolite biosynthesis relies on a number of enzymatic reactions, which evolved from primary metabolism; for example, the biochemical pathway resulting in tetracycline has a large scaffold relying on 72 enzymatic steps (McCormick and Jensen 1965).

Many of the organisms that produce bioactive SMs are soil-dwelling and live in complex ecosystems (Brakhage and Schroeckh 2011). It is hypothesized that they rely on these compounds to act as communication tools and potentially as defensive mechanisms (Yim et al. 2007). Within these ecosystems, competition and communication has resulted in the co-evolution of microbial communities over millions of years. Subsequent investigations using genome sequencing has resulted in the identification of silent or cryptic gene clusters that are not readily expressed under

laboratory conditions and could be potentially involved in secondary metabolism (Yin et al. 2007).

The genes involved in the production of SMs have survived millions of years in their respective hosts, even as they seem to be silent or cryptic at first glance. Therefore, there is great interest in identifying the SMs produced by these clusters and to decipher their biological and ecological functions.

1.01.02 Hidden in the deep: cryptic pathways and antibiotic production potential

As technology has evolved, we have gained the ability to identify the genetic potential of microorganisms to produce SMs. The explosion of genetic information has led to the discovery of many SM gene clusters with no obvious function for the host organisms (Gust et al. 2003). By mining the genomes of such microorganisms, researchers have gained insight into previously untold biosynthetic potentials. Genomic studies on Aspergilli, a group of filamentous fungi known for the production of both therapeutic and toxic SMs, have demonstrated immense diversity in terms of the SM gene clusters that they harbor (Nierman et al. 2005). This diversity and complexity has also raised questions as to why and how both fungi and bacteria have arranged the genes involved in the production of a SM together to form clusters (Chiou et al. 2002). In addition, fungal SM clusters have been shown to reside close to the telomeres, suggesting chromatin-based regulation (Bok and Keller 2004). This led researchers to explore the possibility of using histone modifiers to regulate and activate the production of some of these cryptic SM in fungi (Jenuwein and Allis 2001).

The *Streptomyces* are Gram-positive soil bacteria renowned for their NP production capabilities and for harboring numerous cryptic SM gene clusters. On average, each *Streptomyces* species contains gene clusters for at least 20 predicted secondary metabolic pathways, and this appears to be a ubiquitous phenomenon within the Actinobacteria (Tanaka et al. 2013). Interestingly, it appears that sub-inhibitory concentrations of antibiotics can also induce expression of cryptic SM pathways in certain Actinobacteria (Seyedsayamdost 2014) and there are excellent reviews on the subject (Craney et al. 2013; Moore et al. 2012; Yoon and Nodwell 2014).

Induction of cryptic SM pathways has not been limited to the application of chemicals. Studies have also shown that production can be increased via targeted genetic mutations in ribosome- and RNA polymerase-related genes. Through selection for mutations in the RNA polymerase β-subunit gene (*rpo*B) leading to rifampin resistance, researchers were able to activate pathways involved in SM production in *Streptomyces lividans*, a close relative of the well-studied species *Streptomyces coelicolor* (Hu et al. 2002). This phenomenon was originally discovered when *S. coelicolor* spontaneously developed rifampin resistance leading to the overproduction of actinorhodin, a known blue pigmented antibiotic (Hu and Ochi 2001). It is believed that this mutation mimicked the stringent response, which is involved in the production of the highly phosphorylated nucleotide ppGpp under conditions of amino acid starvation (Hu and Ochi 2001). ppGpp is known to bind to RNA polymerase, and this alters the affinity of the enzyme for certain promoters and changes the corresponding gene expression patterns. The *rpoB* mutation leading to rifampin resistance was thought to alter RNA polymerase in the same way as ppGpp binding (Hu et al. 2002). In addition, previous studies showed that mutations in

the *rpsL* gene encoding the ribosome protein S12 led to activation of cryptic SM biosynthetic genes in *S. lividans* and their overexpression in *S. coelicolor* (Shima et al. 1996). As a consequence, a new technique for the over-expression of both cryptic and non-cryptic gene clusters was developed and called "ribosome engineering", which involves the introduction of mutations that confer drug resistance to certain antibiotics that target the ribosome (Ochi 2007).

These tools have made possible the engineering of host *S. coelicolor* strains for the heterologous expression of SM biosynthetic gene clusters. Due to the complexity of the *Streptomyces* genetic makeup and the large size of many antibiotic biosynthetic gene clusters, previous attempts to utilize bacterial expression strains failed. However, by engineering a strain of *S. coelicolor* through targeted mutations in the *rpoB* and *rpsL* genes, a somewhat relaxed host was created (Gomez-Escribano and Bibb 2011). Under laboratory conditions, *S. coelicolor* is known to produce four SMs (actinorhodin, undecylprodigiosin, CPK, and a calcium-dependent antibiotic), which could affect the levels of available precursors for the biosynthesis of other SMs (Gomez-Escribano and Bibb 2011). By deleting the four known antibiotic biosynthetic pathways, and by introducing the targeted mutations in *rpoB* and *rpsL*, a *S. coelicolor* strain was prepared that could heterologously express the antibiotic gene cluster for congocidine from *Streptomyces ambofaciens* (Gomez-Escribano and Bibb 2011). While the use of different media, chemicals, and genetic manipulation may seem unrelated, they all seem to invoke a stress response, which results in the production of SMs. By combining the three processes/approaches together, it is possible to obtain a viable heterologous host for expressing cryptic SM gene clusters.

1.01.03 The *Streptomyces*

Streptomyces stands as the largest genus in the order Actinomycetales and the family Streptomycetaceae (Goodfellow and Fiedler 2010). The ability of members of this genus to produce economically valuable SMs has led to the identification of many different species (Goodfellow and Fiedler 2010). *Streptomyces* are ubiquitous in terrestrial environments and have the ability to colonize deep soils and into the reaches of the rhizosphere. This is accomplished via the growth/spread of vegetative hyphae and the ability of the *Streptomyces* to differentiate into spores for dispersal (Rueda et al. 2001). The life cycle of the *Streptomyces* allows them to survive long periods of unfavorable conditions (Ensign 1978), with spores being recovered and revived from 70 year old soil samples (Morita 1985).

Streptomyces produce a wide variety of bioactive compounds including upwards of 80% of all clinically used antibiotics of microbial origin (Goodfellow and Fiedler 2010). Antibiotic production is generally growth phase dependent and is also influenced by media and culture conditions. For example, antibiotics and other SMs are produced at the onset of stationary phase in liquid cultures and during morphological differentiation on solid media (Manteca et al. 2008). There are complex global regulatory pathways that control cellular development and secondary metabolism in the *Streptomyces* (Liu et al. 2013). This highlights the complexity of gene regulation within this genus, and stresses the importance of these SM gene clusters in the life cycle of *Streptomyces*.

Many SM biosynthetic capabilities are encoded by large gene clusters, which also contain pathway-specific regulatory genes that are switched on due to a number

of environmental and physiological factors (Liu et al. 2013; Paradkar 2013). Production of SMs is not only stimulated by environmental stresses, but can also be activated in response to metabolic imbalances, physiological stresses and signaling molecules (Eckwall and Schottel 1997; Hood et al. 1992). The exact mechanism for the activation of SM production (including antibiotics) is not well understood; however, it does appear that growth rate and nutritional factors are common denominators in most instances (Frisvad 2012; Kieser et al. 2000).

An immediate concern in antibiotic-producing bacteria is the inherent detrimental impact of the end product on the producing organism. To overcome this, such bacteria have developed a series of solutions to become resistant to the antibiotics they produce (Malla et al. 2010). One such example is the presence of genes within antibiotic biosynthetic gene clusters that encode transmembrane efflux pumps, which pump out the antibiotic from the cell once it is produced (Xu et al. 2012). These resistance genes may also be responsible for cross-resistance to clinically used antibiotics. Therefore, due to their ubiquitous nature and their possible role in passing resistance genes via horizontal gene transfer, *Streptomyces* may play a significant part in the spread of antibiotic resistance (Wiener et al. 1998).

Streptomyces spp. have G+C rich DNA ranging between 61% - 81% (Wright and Bibb 1992). Furthermore, most *Streptomyces* have large genomes and some contain additional linear megaplasmids (Zhou et al. 2012). *S. clavuligerus* ATCC 27064, which is the focus of most of the described work in this thesis, has a 6.76 Mbp chromosome and a 1.796 Mbp megaplasmid (pSCL4), making it one of the largest plasmids to be sequenced to date (Medema et al. 2010). Another distinguishing

feature of many *Streptomyces* is the presence of a linear chromosome (Hopwood 2006), which may be somehow related to the casual amplifications and deletions that their chromosomes undergo at frequencies ranging between 0.1%-1.0% within the population (Cullum et al. 1988; Leblond et al. 1990). Plasmids are found in many species and can be linear or circular with their size covering a broad range (Kinashi 2011). Out of all the medicinally important *Streptomyces,* one described species in particular is hiding many secrets deep within.

1.01.04 *Streptomyces clavuligerus*

Streptomyces clavuligerus is an important producer of commercially important NPs, with the renowned ability to produce the β-lactam antibiotic cephamycin C (Alexander and Jensen 1998) and the β-lactamase inhibitor clavulanic acid (Brown et al. 1976). By itself, clavulanic acid is a weak antibiotic, but it can be used along with certain semisynthetic β-lactams (e.g. amoxicillin) to treat infections that are resistant to β-lactams due to β-lactamase production (Brogden et al. 1981). Although the biosynthetic pathways for producing cephamycin C and clavulanic acid are distinct, the biosynthetic genes for these SMs are clustered together to form a β-lactam super-cluster in *S. clavuligerus* (Liras et al. 2008). Furthermore, several other antibiotics have been discovered that are produced by *S. clavuligerus*, including holomycin (Li and Walsh 2010) and the 5*S* clavams, the latter of which are related to clavulanic acid but differ in their 5*S* stereochemistry as compared to the 5*R* configuration of clavulanic acid (Kwong et al. 2012). In addition, *S. clavuligerus* also produces some SMs with other bioactivities (Li and Walsh 2010).

The genome of *S. clavuligerus* ATCC 27064 was sequenced and analyzed for possible NP biosynthetic gene clusters (Medema et al. 2010). The study showed that *S. clavuligerus* ATCC 27064 only has one plasmid, a 1.8 Mbp linear megaplasmid densely packed with 25 possible NP biosynthetic gene clusters and a 6.8 Mbp core chromosome, which contains 23 putative NP gene clusters (Medema et al. 2010). Interestingly, earlier work exploring the *S. clavuligerus* genome from the strain NRRL3585 found the genomic make up in that organism to include a chromosome along with three linear plasmids: pSCL1 (11.7kbp), pSCL2 (120kbp), and pSCL3 (430kbp) (Wu et al. 2006; Wu and Roy 1993). The discrepancy in the reported presence/numbers of plasmids in *S. clavuligerus* raises intriguing questions about the actual genetic composition of this industrially important bacterium, which warrant further investigation.

1.01.05 The machinery: NRPS and PKS

Secondary metabolite biosynthetic gene clusters encode many different types of proteins, none more important than massive enzymes called nonribosomal peptide synthetases (NRPS) and polyketide synthases (PKS) (Frueh et al. 2008). Certain PKSes (type I only) have similarities to fatty acid synthases (FAS), these complexes are made up of modular units, which act as an assembly line where each conserved domain incorporates a single piece of the growing product (Frueh et al. 2008). PKSes differ from NRPSes in that they condense activated primary metabolites such as acetyl-CoA and malonyl-CoA to form larger polymers. In turn, PKSes differ from FASes as they do not necessarily reduce the final product, resulting in a highly diverse

spectrum of possible polyketide chains (Fujii et al. 2001). Interestingly, many NRPSes and PKSes seem to localize to the inner surface of the cytoplasmic membrane of the organism they are expressed in (Straight et al. 2007).

NRPSes work by condensing both proteinogenic and non-proteinogenic amino acids into peptides. These massive enzymes, or mega-synthetases, are involved in the synthesis of some of the most important bioactive metabolites, ranging from antibiotics like penicillins to immune-suppressants like cyclosporine (Figure 1.01). The diversity of NRPS products is of striking consequence when considering the highly conserved mechanism used for NRP backbone biosynthesis.

In multimodular NRPSes, each module incorporates a new amino acid, where the number of modules matches the number of amino acids contained in the end product (Marahiel 2009). Each of the modules can be broken down into three catalytic domains that act to covalently link substrates to the carrier 4'-phosphopantetheine arm (PPan) (Marahiel 2009). The first domain, which is the adenylation domain (A), acts as a discriminatory gate keeper: it selects the specific amino and generates a aminoacyl AMP-mixed anhydride via ATP hydrolysis (Figure 1.02) (Mootz and Marahiel 1997). Following the initial binding, the intermediate is attached to a thiol moiety on the PPan prosthetic "arm", which is connected to the peptidyl carrier protein (PCP or P) downstream of the A-domain (Stachelhaus et al. 1998). The PCP then acts like a swinging arm: transferring the high-energy thioester substrate, inclusive of all previous elongation intermediates, between catalytic sites (Marahiel 2009). The condensation domain (C) activates the condensation between the two intermediates forming the peptide backbone. A nucleophilic attack is then carried out

by the C-domain and the growing peptide is translocated down the assembly line (Stachelhaus et al. 1998). This allows the upstream PPan-PCP to load the next reaction (Marahiel 2009). The thioesterase domain proceeds to catalyze the release of the end product via hydrolysis from the PPan prosthetic arm. This culminates in four essential domains: C, A, PCP, and TE. The final product of an NRPS can undergo modification during assembly by several tailoring units (Marahiel 2009).

PKSes can be divided into three different types, each producing a different style of compound. Type I comprises one or more large multifunctional proteins, divided into modules, with each module carrying out a single non-iterative cycle of elongation (Cheng et al. 2003). Type II PKSes are multiprotein complexes that can act iteratively (Cheng et al. 2003). Type III enzymes resemble type I PKSes, in that they \ comprise a single polypeptide/protein; however they operate iteratively like the type II complexes (Shen 2003). PKSes are responsible for producing many common antibiotics (erythromycin, epothilone, and rapamycin) (Figure 1.03) (Cheng et al. 2003). Type I synthases (not types II and III) are modular and have multidomain catalytic units that are responsible for a single step of elongation (Thattai et al. 2007). A different module adds each extender unit; therefore, the order of each module in the PKS determines the chemical structure of the product (Figure 1.04). These catalytic domains can tolerate a broad range of substrates, thereby increasing the breadth and diversity of the end product. PKSes have the distinct advantage of still being catalytically active when the modules are rearranged (Thattai et al. 2007). This allows for a potentially limitless diversity of polyketide products; while at the same time also indicating that modules are frequently shuffled during evolution (Thattai et al. 2007).

PKSes generally comprise three core domains per module (type I) or proteins (type II): the acyltransferase (AT), the acyl carrier protein (ACP) and the ketosynthase (KS). In a single step, the AT primes a phosphopanheinyl arm of the ACP domain with a malonyl residue. Following this, the KS domain catalyzes decarboxylative acylation, thereby extending the product (Donadio et al. 2007). The process adds two carbon units per cycle, meaning that a ring containing 14 carbons, such as erythromycin, requires at least 7 cycles to be produced. Several modification domains may also be involved, including ketoreductase, dehyradatase, and enoyl reductase (Donadio et al. 2007). Finally, the thioesterase or cyclase domain catalyzes the release and cyclization of the product (Donadio et al. 2007). The polyketide is cyclized in either an end-to-end manner to give a large ring or internally to form a fused ring product.

NRPSes and PKSes are lynchpins in the production of many medically relevant bioactive compounds. In some cases, hybrids between the two exist where a polyketide is fused to an NRP, leading to even more diversity and complexity in the derived NP (Boettger and Hertweck 2013).

1.01.06 The δ-(L-α-aminoadipyl)-L-cysteinyl-D-valine (ACV) synthetase (ACVS) and cephamycin C production in *S. clavuligerus* as a model system for studying NRPS proteins

The genetic organization of antibiotic gene clusters is complicated. Often times they are organized as large “super-clusters”, sometimes encoding two biosynthetically

different, yet functionally related compounds; such is the case with cephamycin C and clavulanic acid (Ward and Hodgson 1993).

The δ-(L-α-aminoadipyl)-L-cysteinyl-D-valine (ACV) Synthetase (ACVS) is a known NRPS with a well-characterized function in the biosynthesis of the conventional β-lactams, and will act as a model for typical NRPS-based studies. ACVS is one of the initial enzymes involved not only in the cephamycin C biosynthesis pathway, but also in the production of all natural penicillins and cephalosporins and is the link between primary and secondary metabolism (Zhang and Demain 1992). It is classified as a megaenzyme since it is predicted to be 404-426 kDa in size (Hamed et al. 2012). Functionally, ACVS utilizes ATP to condense L-α-aminoadipic acid, L-alanine, and L-cysteine into a tripeptide, δ-(L-α-aminoadipyl)-L-cysteinyl-D-valine (Figure 1.05a/b) (Jensen et al. 1990). The NRP product is then cyclized by ACV cyclase to form isopenicillin N (Liras 1999). Isopenicillin N can then either be passed to ACYL transferase eventually leading to the penicillin G in fungi (Liras 1999), or to the IPN epimerase during cephalosporin biosynthesis (Liras 1999). The IPN epimerase then converts the L-α-aminoadipic acid side chain into the D-configuration (Láiz et al. 1990), resulting in penicillin N that is acted on by penicillin N expandase for expansion of the 5-carbon thiazolidine ring into a 6-carbon dihydrothiazine ring, a characteristic of all cephalosporin antibiotics (Liras 1999). Following this, the DAOC hydroxylase adds a hydroxyl group to the dihydrothiazine ring, and the final steps in cephamycin C production include the carbamoylation of the C-3' and the introduction of a methoxyl group at C-7 positions (Liras 1999).

Given its important role in NP pathways, it is no surprise that ACVS is one of most highly characterized NRPS proteins. However, due to the extremely large size of NRPS proteins, there is very little information available on their native structures. Therefore, novel methods are required for purifying these proteins for future studies.

1.01.07 Megaplasmids from *Streptomyces* have natural product biosynthesis potential

Mini- and giant linear plasmids exists within many *Streptomyces* spp. and are treasure troves of biosynthetic potential (Kinashi 2011). At least two *Streptomyces* spp. have large megaplasmids, *S. clavuligerus* ATCC 27064 (1.8 Mbp megaplasmid, pSCL4) and *Streptomyces cattleya* NRRL 8057 (which harbors a 1.8 Mbp megaplasmid and a 6.3Mbp central chromosome) (Barbe et al. 2011). Both of these megaplasmids are packed with multiple predicted SM biosynthetic gene clusters.

Another pertinent *Streptomyces* harboring plasmids is *S. coelicolor,* which contains SCP1, a 350 kbp plasmid (Kinashi et al. 1987). SCP1 contains the genes for methylenomycin production, the first antibiotic discovered for which the gene cluster was located on a plasmid (Kirby and Hopwood 1977). *Streptomyces rimosus* R6 contains a 387 kbp linear plasmid (pPZG101). A crossover event between the plasmid pPZG101 and the chromosome in *S. rimosus* led to an increase in copy number of the genes involved in oxytetracycline biosynthesis, leading to its overproduction (Pandza et al. 1998). As more genomes are published, it will become apparent how prevalent large plasmids and megaplasmids are throughout the genus. It is becoming clear that *Streptomyces* plasmids act as reservoirs of SM gene clusters.

1.01.08 Focus of described work

Here, we explored two facets of natural product biosynthesis. In Chapter 2, we attempted to introduce the pSCL4 megaplasmid from *S. clavuligerus* into a *S. coelicolor* heterologous host in order to activate the SM biosynthetic gene clusters present on the plasmid. In Chapter 3, we developed a novel tool for the structural study of NRPSes. By using ACVS as a model, we targeted the overexpression and purification of ACVS from *S. clavuligerus* for further structural studies using cryo-electron microscopy and x-ray crystallography.

BIBLIOGRAPHY

Alberts, A.W., Chen, J., Kuron, G., Hunt, V., Huff, J., Hoffman, C., et al. 1980. Mevinolin: a highly potent competitive inhibitor of hydroxymethylglutaryl-coenzyme A reductase and a cholesterol-lowering agent. Proc Natl Acad Sci U S A 77(7): 3957-3961. doi: 10.1073/pnas.77.7.3957.

Alexander, D.C., and Jensen, S.E. 1998. Investigation of the *Streptomyces clavuligerus* cephamycin C gene cluster and its regulation by the CcaR protein. J Bacteriol 180(16): 4068-4079.

Barbe, V., Bouzon, M., Mangenot, S., Badet, B., Poulain, J., Segurens, B., et al. 2011. Complete genome sequence of *Streptomyces cattleya* NRRL 8057, a producer of antibiotics and fluorometabolites. J Bacteriol 193(18): 5055-5056. doi: 10.1128/JB.05583-11

Boettger, D., and Hertweck, C. 2013. Molecular diversity sculpted by fungal PKS-NRPS hybrids. Chembiochem 14(1): 28-42. doi: 10.1002/cbic.201200624.

Bok, J.W., and Keller, N.P. 2004. LaeA, a regulator of secondary metabolism in *Aspergillus* spp. Eukaryot Cell 3(2): 527-535. doi: 10.1128/EC.3.2.527-535.2004.

Brakhage, A.A., and Schroeckh, V. 2011. Fungal secondary metabolites - strategies to activate silent gene clusters. Fungal Genet Biol 48(1): 15-22. doi: 10.1016/j.fgb.2010.04.004.

Brogden, R.N., Carmine, A., Heel, R.C., Morley, P.A., Speight, T.M., and Avery, G.S. 1981. Amoxycillin/clavulanic acid: a review of its antibacterial activity, pharmacokinetics and therapeutic use. Drugs 22(5): 337-362.

Brown, A.G., Butterworth, D., Cole, M., Hanscomb, G., Hood, J.D., Reading, C., et al. 1976. Naturally-occurring beta-lactamase inhibitors with antibacterial activity. J Antibiot (Tokyo) 29(6): 668-669.

Cheng, Y.Q., Tang, G.L., and Shen, B. 2003. Type I polyketide synthase requiring a discrete acyltransferase for polyketide biosynthesis. Proc Natl Acad Sci U S A 100(6): 3149-3154. doi: 10.1073/pnas.0537286100.

Chiou, C.H., Miller, M., Wilson, D.L., Trail, F., and Linz, J.E. 2002. Chromosomal location plays a role in regulation of aflatoxin gene expression in *Aspergillus parasiticus*. Appl Environ Microbiol 68(1): 306-315. doi: 10.1128/AEM.68.1.306-315.2002.

Cullum, J., Flett, F., and Piendl, W. 1988. Genetic instability in *streptomycetes*. Microbiol Sci **5**(8): 233-235.

Craney, A., Ahmed, S., and Nodwell, J. 2013. Towards a new science of secondary metabolism. J Antibiot (Tokyo) 66(7): 387-400. doi: 10.1038/ja.2013.25.

Donadio, S., Monciardini, P., and Sosio, M. 2007. Polyketide synthases and nonribosomal peptide synthetases: the emerging view from bacterial genomics. Natural Product Reports 24(5): 1073-1109. doi: 10.1039/b514050c.

Duggar, B.M. 1948. Aureomycin; a product of the continuing search for new antibiotics. Ann N Y Acad Sci 51(Art. 2): 177-181.

Eckwall, E.C., and Schottel, J.L. 1997. Isolation and characterization of an antibiotic produced by the scab disease-suppressive *Streptomyces diastatochromogenes* strain PonSSII. J Ind Microbiol Biotechnol 19(3): 220-225. doi: 10.1038/sj.jim.2900455.

Ensign, J.C. 1978. Formation, Properties, and Germination of *Actinomycete* Spores. Annu Rev Microbiol **32**(1): 185-219. doi: 10.1146/annurev.mi.32.100178.001153.

Feling, R.H., Buchanan, G.O., Mincer, T.J., Kauffman, C.A., Jensen, P.R., and Fenical, W. 2003. Salinosporamide A: a highly cytotoxic proteasome inhibitor from a novel microbial source, a marine bacterium of the new genus *salinospora*. Angew Chem Int Ed Engl 42(3): 355-357. doi: 10.1002/anie.200390115.

Frisvad, J.C. 2012. Media and growth conditions for induction of secondary metabolite production. Methods Mol Biol 944: 47-58. doi: 10.1007/978-1-62703-122-6_3.

Frueh, D.P., Arthanari, H., Koglin, A., Vosburg, D.A., Bennett, A.E., Walsh, C.T., et al. 2008. Dynamic thiolation-thioesterase structure of a non-ribosomal peptide synthetase. Nature 454(7206): 903-906. doi: 10.1038/nature07162.

Fujii, I., Watanabe, A., Sankawa, U., and Ebizuka, Y. 2001. Identification of Claisen cyclase domain in fungal polyketide synthase WA, a naphthopyrone synthase of *Aspergillus nidulans*. Chem Biol 8(2): 189-197. doi: 10.1016/S1074-5521(00)90068-1.

Goodfellow, M., and Fiedler, H.P. 2010. A guide to successful bioprospecting: informed by actinobacterial systematics. Antonie Van Leeuwenhoek **98**(2): 119-142. doi: 10.1007/s10482-010-9460-2.

Gomez-Escribano, J.P., and Bibb, M.J. 2011. Engineering *Streptomyces coelicolor* for heterologous expression of secondary metabolite gene clusters. Microb Biotechnol 4(2): 207-215. doi: 10.1111/j.1751-7915.2010.00219.x.

Gust, B., Challis, G.L., Fowler, K., Kieser, T., and Chater, K.F. 2003. PCR-targeted *Streptomyces* gene replacement identifies a protein domain needed for biosynthesis of the sesquiterpene soil odor geosmin. Proc Natl Acad Sci U S A 100(4): 1541-1546. doi: 10.1073/pnas.0337542100.

Hamed, R.B., Gomez-Castellanos, J.R., Henry, L., Ducho, C., McDonough, M.A., and Schofield, C.J. 2012. The enzymes of beta-lactam biosynthesis. Nat Prod Rep. doi: 10.1039/c2np20065a.

Higgens, C., and Kastner, R. 1971. *Streptomyces clavuligerus* sp. nov., a β-lactam antibiotic producer. International Journal of Systematic Bacteriology 21(4): 326-331.

Hood, D.W., Heidstra, R., Swoboda, U.K., and Hodgson, D.A. 1992. Molecular genetic analysis of proline and tryptophan biosynthesis in *Streptomyces coelicolor* A3(2): interaction between primary and secondary metabolism--a review. Gene 115(1-2): 5-12. doi: 10.1016/0378-1119(92)90533-U.

Hopwood, D.A. 2006. Soil to genomics: the *Streptomyces* chromosome. Annu Rev Genet **40**: 1-23. doi: 10.1146/annurev.genet.40.110405.090639.

Hu, H., and Ochi, K. 2001. Novel approach for improving the productivity of antibiotic-producing strains by inducing combined resistant mutations. Appl Environ Microbiol 67(4): 1885-1892. doi: 10.1128/AEM.67.4.1885-1892.2001.

Hu, H., Zhang, Q., and Ochi, K. 2002. Activation of antibiotic biosynthesis by specified mutations in the *rpoB* gene (encoding the RNA polymerase beta subunit) of *Streptomyces lividans*. J Bacteriol 184(14): 3984-3991. doi: 10.1128/JB.184.14.3984-3991.2002.

Jensen, S.E., Wong, A., Rollins, M.J., and Westlake, D.W. 1990. Purification and partial characterization of delta-(L-alpha-aminoadipyl)-L-cysteinyl-D-valine synthetase from *Streptomyces clavuligerus.* J Bacteriol 172(12): 7269-7271.

Jenuwein, T., and Allis, C.D. 2001. Translating the histone code. Science 293(5532): 1074-1080. doi: 10.1126/science.1063127.

Kieser, T., Bibb, M.J., Buttner, M.J., Chater, K.F., and Hopwood, D.A. 2000. Practical *Streptomyces* Genetics. 2nd ed. John Innes Foundation, Norwich, U.K.

Kinashi, H. 2011. Giant linear plasmids in *Streptomyces*: a treasure trove of antibiotic biosynthetic clusters. J Antibiot (Tokyo) 64(1): 19-25. doi: 10.1038/ja.2010.146.

Kinashi, H., Shimaji, M., and Sakai, A. 1987. Giant linear plasmids in *Streptomyces* which code for antibiotic biosynthesis genes. Nature 328(6129): 454-456. doi: 10.1038/328454a0.

Kirby, R., and Hopwood, D.A. 1977. Genetic determination of methylenomycin synthesis by the SCP1 plasmid of *Streptomyces coelicolor* A3(2). J Gen Microbiol 98(1): 239-252.

Kwong, T., Zelyas, N.J., Cai, H., Tahlan, K., Wong, A., and Jensen, S.E. 2012. 5S clavam biosynthesis is controlled by an atypical two-component regulatory system in *Streptomyces clavuligerus*. Antimicrob Agents Chemother 56(9): 4845-4855. doi: 10.1128/AAC.01090-12.

Láiz, L., Liras, P., Castro, J.M., and Martín, J.F. 1990. Purification and characterization of the isopenicillin N epimerase from *Nocardia lactamdurans*. J Gen Microbiol 136(4): 663-671.

Leblond, P., Demuyter, P., Simonet, J.M., and Decaris, B. 1990. Genetic instability and hypervariability in *Streptomyces ambofaciens:* towards an understanding of a mechanism of genome plasticity. Mol Microbiol **4**(5): 707-714.

Li, B., and Walsh, C.T. 2010. Identification of the gene cluster for the dithiolopyrrolone antibiotic holomycin in *Streptomyces clavuligerus*. Proc Natl Acad Sci U S A 107(46): 19731-19735. doi: 10.1073/pnas.1014140107.

Liras, P. 1999. Biosynthesis and molecular genetics of cephamycins. Cephamycins produced by actinomycetes. Antonie Van Leeuwenhoek 75(1-2): 109-124.

Liras, P., Gomez-Escribano, J.P., and Santamarta, I. 2008. Regulatory mechanisms controlling antibiotic production in *Streptomyces clavuligerus.* J Ind Microbiol Biotechnol 35(7): 667-676. doi: 10.1007/s10295-008-0351-8.

Liu, G., Chater, K.F., Chandra, G., Niu, G., and Tan, H. 2013. Molecular regulation of antibiotic biosynthesis in *Streptomyces*. Microbiol Mol Biol Rev **77**(1): 112-143. doi: 10.1128/MMBR.00054-12.

Malla, S., Niraula, N.P., Liou, K., and Sohng, J.K. 2010. Self-resistance mechanism in *Streptomyces peucetius*: overexpression of *drrA*, *drrB* and *drrC* for doxorubicin enhancement. Microbiol Res 165(4): 259-267. doi: 10.1016/j.micres.2009.04.002.

Manteca, A., Alvarez, R., Salazar, N., Yague, P., and Sanchez, J. 2008. Mycelium differentiation and antibiotic production in submerged cultures of *Streptomyces coelicolor*. Appl Environ Microbiol **74**(12): 3877-3886. doi: 10.1128/AEM.02715-07.

Marahiel, M.A. 2009. Working outside the protein-synthesis rules: insights into non-ribosomal peptide synthesis. J Pept Sci 15(12): 799-807. doi: 10.1002/psc.1183.

McCormick, J.R., and Jensen, E.R. 1965. Biosynthesis of the Tetracyclines. 8. Characterization of 4-Hydroxy-6-Methylpretetramid. J Am Chem Soc 87: 1794-1795.

McGuire, J.M., Bunch, R.L., Anderson, R.C., Boaz, H.E., Flynn, E.H., Powell, H.M., et al. 1952. Ilotycin, a new antibiotic. Schweiz Med Wochenschr 82(41): 1064-1065.

Medema, M.H., Trefzer, A., Kovalchuk, A., van den Berg, M., Muller, U., Heijne, W., et al. 2010. The sequence of a 1.8-mb bacterial linear plasmid reveals a rich evolutionary reservoir of secondary metabolic pathways. Genome Biol Evol 2: 212-224. doi: 10.1093/gbe/evq013.

Moore, J.M., Bradshaw, E., Seipke, R.F., Hutchings, M.I., and McArthur, M. 2012. Chapter Eighteen - Use and Discovery of Chemical Elicitors That Stimulate Biosynthetic Gene Clusters in *Streptomyces* Bacteria. *In* Methods Enzymol. *Edited by* A.H. David. Academic Press. pp. 367-385.

Mootz, H.D., and Marahiel, M.A. 1997. The tyrocidine biosynthesis operon of *Bacillus brevis*: complete nucleotide sequence and biochemical characterization of functional internal adenylation domains. J Bacteriol 179(21): 6843-6850.

Morita, R. 1985. Starvation and miniaturisation of heterotrophs, with special emphasis on maintenance of the starved viable state. Special Publications of the Society for General Microbiology.

Newman, D.J., and Cragg, G.M. 2007. Natural products as sources of new drugs over the last 25 years. J Nat Prod 70(3): 461-477. doi: 10.1021/np068054v.

Nierman, W.C., Pain, A., Anderson, M.J., Wortman, J.R., Kim, H.S., Arroyo, J., et al. 2005. Genomic sequence of the pathogenic and allergenic filamentous fungus *Aspergillus fumigatus*. Nature 438(7071): 1151-1156. doi: 10.1038/nature04332.

Ochi, K. 2007. From microbial differentiation to ribosome engineering. Biosci Biotechnol Biochem 71(6): 1373-1386. doi: 10.1271/bbb.70007.

Paradkar, A. 2013. Clavulanic acid production by Streptomyces clavuligerus: biogenesis, regulation and strain improvement. J Antibiot (Tokyo). doi: 10.1038/ja.2013.26.

Pandza, S., Biukovic, G., Paravic, A., Dadbin, A., Cullum, J., and Hranueli, D. 1998. Recombination between the linear plasmid pPZG101 and the linear chromosome of *Streptomyces rimosus* can lead to exchange of ends. Mol Microbiol 28(6): 1165-1176. doi: 10.1046/j.1365-2958.1998.00877.x.

Rueda, B., Miguelez, E.M., Hardisson, C., and Manzanal, M.B. 2001. Mycelial differentiation and spore formation by *Streptomyces brasiliensis* in submerged culture. Can J Microbiol **47**(11): 1042-1047.

Seyedsayamdost, M.R. 2014. High-throughput platform for the discovery of elicitors of silent bacterial gene clusters. Proc Natl Acad Sci U S A. doi: 10.1073/pnas.1400019111.

Shen, B. 2003. Polyketide biosynthesis beyond the type I, II and III polyketide synthase paradigms. Curr Opin Chem Biol 7(2): 285-295. doi: 10.1016/S1367-5931(03)00020-6.

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

Stabb, E.V., Jacobson, L.M., and Handelsman, J. 1994. Zwittermicin A-producing strains of *Bacillus cereus* from diverse soils. Appl Environ Microbiol 60(12): 4404-4412.

Stachelhaus, T., Mootz, H.D., Bergendahl, V., and Marahiel, M.A. 1998. Peptide bond formation in nonribosomal peptide biosynthesis. Catalytic role of the condensation domain. J Biol Chem 273(35): 22773-22781. doi: 10.1074/jbc.273.35.22773.

Straight, P.D., Fischbach, M.A., Walsh, C.T., Rudner, D.Z., and Kolter, R. 2007. A singular enzymatic megacomplex from *Bacillus subtilis*. Proc Natl Acad Sci U S A 104(1): 305-310. doi: 10.1073/pnas.0609073103.

Tanaka, Y., Kasahara, K., Hirose, Y., Murakami, K., Kugimiya, R., and Ochi, K. 2013. Activation and products of the cryptic secondary metabolite biosynthetic gene clusters by rifampin resistance (*rpo*B) mutations in actinomycetes. J Bacteriol 195(13): 2959-2970. doi: 10.1128/JB.00147-13.

Thattai, M., Burak, Y., and Shraiman, B.I. 2007. The origins of specificity in polyketide synthase protein interactions. PLoS Comput Biol 3(9): 1827-1835. doi: 10.1371/journal.pcbi.0030186.

Ward, J.M., and Hodgson, J.E. 1993. The biosynthetic genes for clavulanic acid and cephamycin production occur as a 'super-cluster' in three *Streptomyces*. FEMS Microbiol Lett 110(2): 239-242.

Wiener, P., Egan, S., Huddleston, A.S., and Wellington, E.M. 1998. Evidence for transfer of antibiotic-resistance genes in soil populations of streptomycetes. Mol Ecol 7(9): 1205-1216. doi: 10.1046/j.1365-294x.1998.00450.x.

Wright, F., and Bibb, M.J. 1992. Codon usage in the G+C-rich *Streptomyces* genome. Gene **113**(1): 55-65.

Wu, W., Leblanc, S.K., Piktel, J., Jensen, S.E., and Roy, K.L. 2006. Prediction and functional analysis of the replication origin of the linear plasmid pSCL2 in *Streptomyces clavuligerus*. Can J Microbiol 52(4): 293-300. doi: 10.1139/w05-126.

Wu, X., and Roy, K.L. 1993. Complete nucleotide sequence of a linear plasmid from *Streptomyces clavuligerus* and characterization of its RNA transcripts. J Bacteriol 175(1): 37-52.

Xu, Y., Willems, A., Au-Yeung, C., Tahlan, K., and Nodwell, J.R. 2012. A Two-Step Mechanism for the Activation of Actinorhodin Export and Resistance in *Streptomyces coelicolor*. MBio 3(5). doi: 10.1128/mBio.00191-12.

Yim, G., Wang, H.H., and Davies, J. 2007. Antibiotics as signalling molecules. Philos Trans R Soc Lond B Biol Sci 362(1483): 1195-1200. doi: 10.1098/rstb.2007.2044.

Yin, J., Straight, P.D., Hrvatin, S., Dorrestein, P.C., Bumpus, S.B., Jao, C., et al. 2007. Genome-wide high-throughput mining of natural-product biosynthetic gene clusters by phage display. Chem Biol 14(3): 303-312. doi: 10.1016/j.chembiol.2007.01.006.

Yoon, V., and Nodwell, J.R. 2014. Activating secondary metabolism with stress and chemicals. J Ind Microbiol Biotechnol 41(2): 415-424. doi: 10.1007/s10295-013-1387-y.

Zhang, J., and Demain, A.L. 1992. ACV synthetase. Crit Rev Biotechnol 12(3): 245-260. doi: 10.3109/07388559209069194.

Zhou, Z., Gu, J., Li, Y.Q., and Wang, Y. 2012. Genome plasticity and systems evolution in *Streptomyces*. BMC Bioinformatics 13 Suppl 10: S8. doi: 10.1186/1471-2105-13-S10-S8.

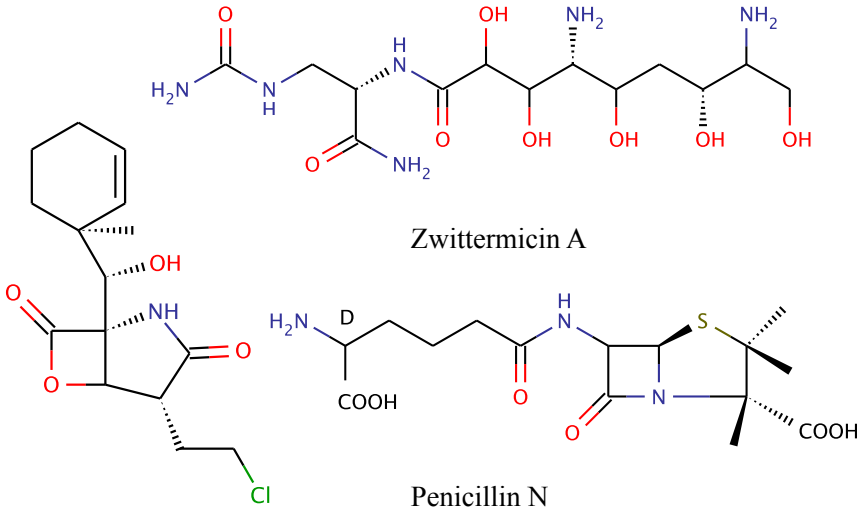


Figure 1.01: Examples of clinically useful nonribosomal peptides. Salinosporimide A was isolated from marine samples and has potent antitumor activity (Feling et al. 2003); Zwittermicin A is an example of a NRPS-PKS hybrid that is used agriculturally to treat plant disease (Stabb et al. 1994); Penicillin N is a precursor to the β-lactam Cephamycin C (Liras 1999).

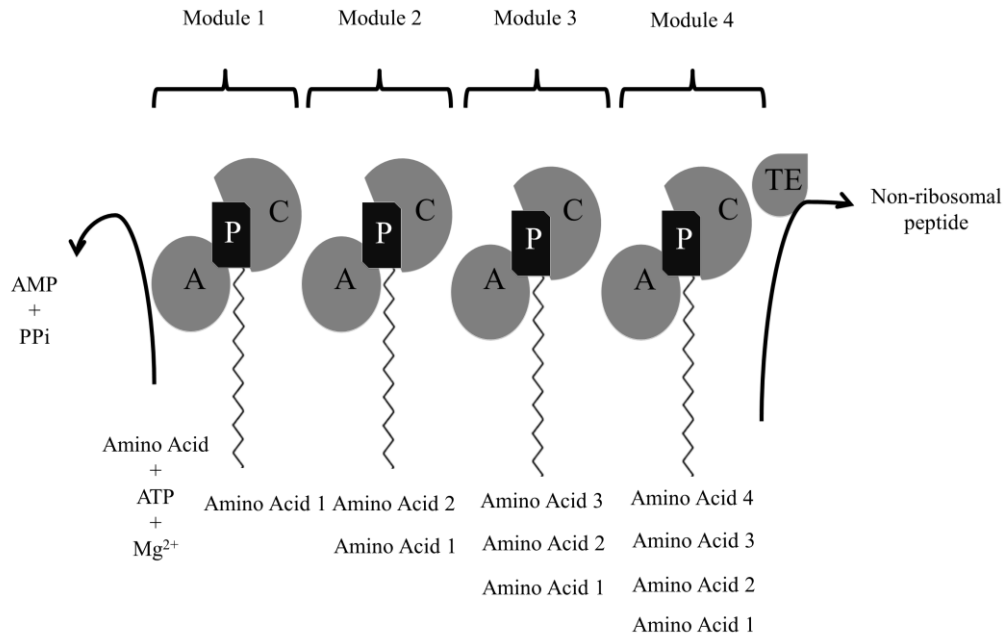


Figure 1.02: Simplified mechanism of the NRPS machinery. A, adenylation domain; P, PCP; C, condensation domain; TE, transesterification domain.

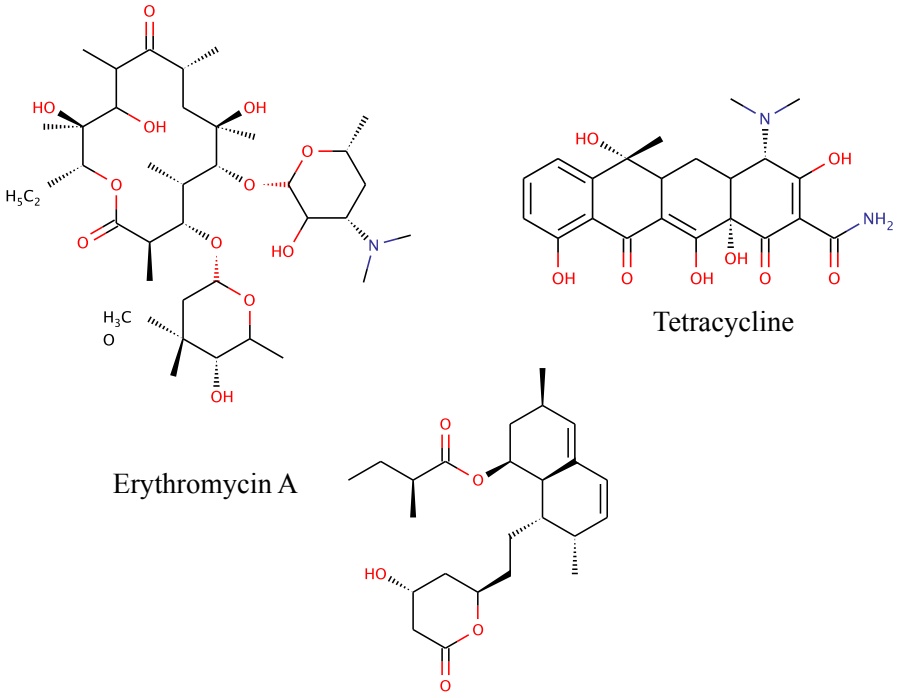


Figure 1.03: Examples of clinically useful polyketides. Erythromycin A is an example of a macrolide antibiotic and was originally isolated from the soil dwelling bacterium *Saccharopolyspora erythreus* (McGuire et al. 1952); tetracycline is characterized by the classic four-ring core structure, which is the hallmark of a subsequent subclass of polyketides (Duggar 1948); Lovastatin is used to treat dyslipidemia and was isolated from the fungus *Aspergillus tereus* (Alberts et al. 1980).

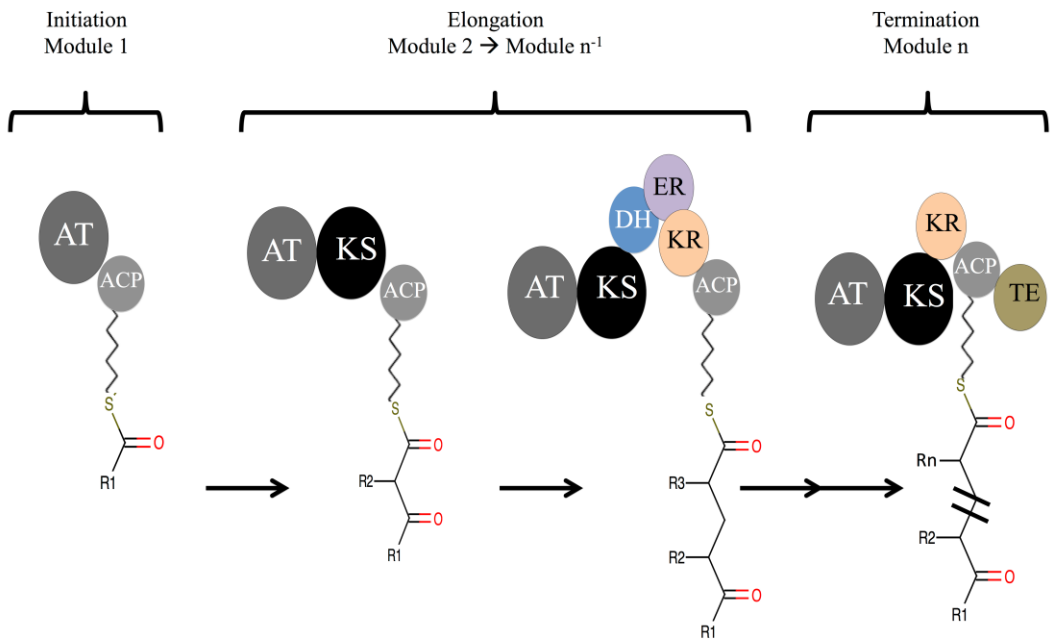


Figure 1.04: Simplified diagram of a type I PKS modular structure. The major components are: AT, acyltransferase; ACP, acyl carrier protein; KS, ketosynthase; DH, dehydratase; ER, enoyl reductase; KR, ketoreductase; TE, thioesterase.

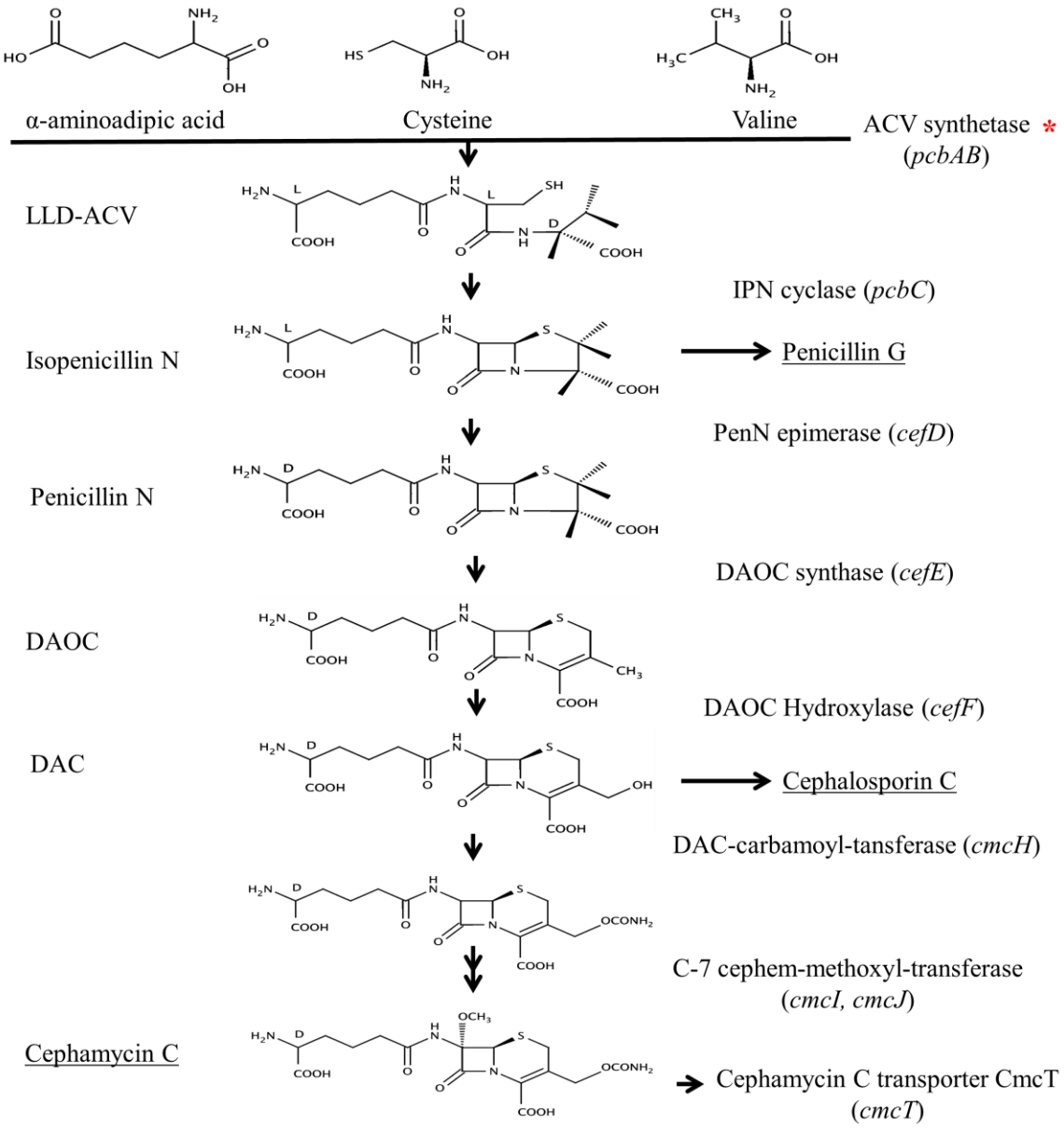


Figure 1.05a: The formation of LLD-ACV by ACVS: L-α-aminoadipic acid, L-cysteine, and L-valine are brought together by ACVS to form LLD-ACV. The initial step in the pathway is the conversion of L-lysine into α-aminoadipic acid by LAT (*lat*) and PCD-DH (*pcd*) (not shown). ACVS (indicated by *) represents an ideal NRPS with typical A, P, and C domain arrangement in each module. Intermediates are listed on the left; enzymes/genes are listed on the right. Potential products are underlined.

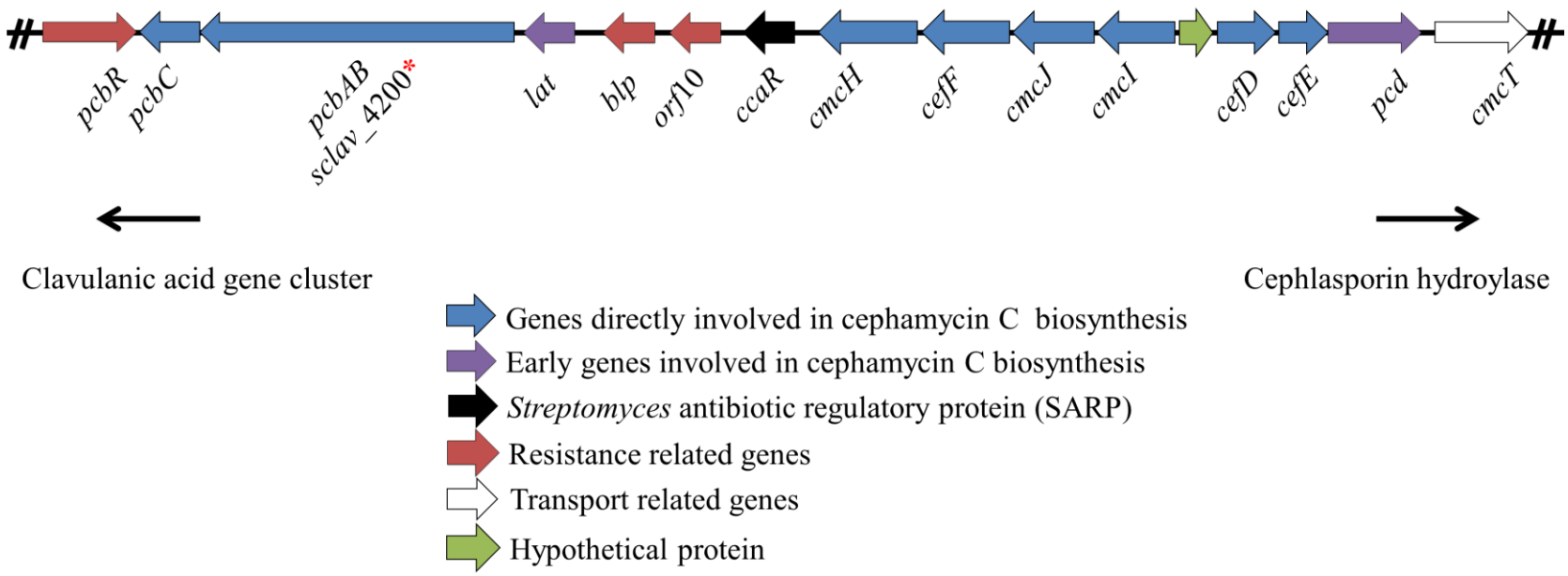


Figure 1.05b: The cephamycin C gene cluster is geographically related to the clavulanic acid gene cluster, along with sharing a regulator, CcaR. Genes that result in NRPS and PKS pathways are often organized as closely packed units; however, gene orientation is not always congruent. *ACVS encoding gene.

Chapter 2

Mobilizing large extra-chromosomal elements from *Streptomyces clavuligerus*

Co-authorship statement

The project was designed and developed by Dr. Kapil Tahlan (Department of Biology, Memorial University of Newfoundland). Experiments were performed by Marcus Moore (Department of Biology, Memorial University of Newfoundland). Reagents/materials/analysis tools were contributed by Drs. Kapil Tahlan and Francisco Barona-Gomez (Langebio Cinvestav-IPN, Irapuato, Mexico). Data analysis was performed by Marcus Moore and by Drs. Kapil Tahlan, Francisco Barona-Gomez and Pablo Cruz-Morales (Langebio Cinvestav-IPN, Irapuato, Mexico). The manuscript was written by Marcus Moore and Dr. Kapil Tahlan.

2.01 - Introduction

The evolution and prevalence of large plasmids within the *Streptomyces* and the analysis of their biological functions is an important area of research (Alvarez-Alvarez et al. 2014; Barbe et al. 2011; Kinashi 2011; Medema et al. 2010; Netolitzky et al. 1995). Megaplasmids (>1Mbp) in *Streptomyces* are untapped sources of SMs as they contain numerous clusters for producing such compounds (Barbe et al. 2011; Medema et al. 2010). Several theories on the origins of these megaplasmids have been suggested and include: (i) a plasmid that has undergone a double crossover with the core chromosome; (ii) a chromosome arm that has fragmented off with an integrated and independent origin of replication; or (iii) multiple consecutive recombination events that have resulted in one linear megaplasmid and one linear core chromosome <7Mbp (Medema et al. 2010). The stochastic nature of *Streptomyces* genomes has made it difficult to study how plasmids arise, are maintained, and how common they are in the genus as a whole. The current crisis of antibiotic resistance has resulted in a tremendous pressure to better understand and exploit the *Streptomyces* for useful natural products (Overbye and Barrett 2005; Wright 2007, 2012).

S. clavuligerus NRRL 3585 harbors four plasmids, two of which are greater than 400kbp (Song et al. 2010). A 1.8Mbp extra-chromosomal element (pSCL4) encodes upwards of 20 potential SM gene clusters (Figure 2.01-A) out of the 48 total clusters identified in this organism (Medema et al. 2010). Many of these so-called "cryptic" clusters (producing no obvious product in the lab) are of interest as they could be involved in the production of novel antibiotic candidates. Targeting each

gene cluster for individual activation, without functional knowledge, is time consuming and may not be successful. Furthermore, functional prediction of gene clusters can only provide insight into the predicted structure of the product and not the bioactive function. Therefore, it is necessary to develop a simple method for large-scale expression and activation of these cryptic genes.

Recently, several *S. coelicolor* "super-strains" (M1146, M1152, and M1154) were developed, which can act as a heterologous hosts for SM gene clusters (Gomez-Escribano and Bibb 2011). These stains have targeted mutations in genes associated with the ribosome (*rpoB* and *rpsL*) that physiologically mimic the stringent response, which is known to increase the production of certain SMs (Gomez-Escribano and Bibb 2011). Furthermore, M1154 has had four known SM pathways knocked-out, resulting in an over-abundance of precursors that can be used by alternate pathways. The introduction of heterologous gene clusters involved in the production of known antibiotics into these strains demonstrated that the super-strains could produce these metabolites at higher levels than the native producer (Gomez-Escribano and Bibb 2011; Jones et al. 2013). Here, we present our attempts to mobilize the *S. clavuligerus* pSCL4 through protoplast fusion into *S. coelicolor* M1154 for the induction of cryptic pathways on the plasmid. It is unknown if pSCL4 can be mobilized between different species or if it can be readily maintained within a heterologous host. Furthermore, if pSCL4 cannot be transferred to *S. coelicolor*, we hope to achieve either the mobilization of one of the smaller plasmids or the production of a genomic chimera between *S. clavuligerus* and *S. coelicolor* M1154, in an attempt to get the production of previously un-described metabolites from either organism.

2.02 Materials and Methods

2.02.01 Culture conditions and general procedures

2.02.01.01 General equipment and procedures

All microbiological manipulations were conducted using aseptic techniques with sterile reagents and glassware. Reagents, antibiotics, and media, unless otherwise stated, were purchased from Sigma Aldrich Canada, Fisher Scientific Canada, or VWR International. Media and solutions were prepared with high performance liquid chromatography (HPLC) grade H_2O and were sterilized by autoclaving or by filter sterilization using 0.2 μm filters (28145-481, VWR International, Edmonton, Alberta) when applicable. Standard equipment used throughout the study include the following: Eppendorf 5424 micro-centrifuge (022620401, Hamburg, Germany); a Thermo Scientific Sorvall ST 16 R centrifuge (75004380, Waltham, Massachusetts, USA) with a TX-400 Swinging Bucket Rotor (75003629, Waltham, Massachusetts, USA); and an Implen P300 NanoPhotometer (Implen GmbH, München, Germany). *Streptomyces* mycelia and *Escherichia coli* cultures were harvested by centrifugation at 3500 rpm for 5 minutes at 4^{o}C, and protoplasts at 1000 rpm for 10 minutes at 4^{o}C, unless otherwise stated. All *Streptomyces* spp. (spore, protoplasts, or mycelial stocks) and *E. coli* strains were stored in 20% v/v glycerol at -80^{o}C, and DNA was stored at -20^{o}C. Additional reagents/chemicals/kits were stored and used as per the manufacturer's directions, unless otherwise specified. All polymerase chain reactions (PCR) were conducted using an Eppendorf VapoProtect Mastercycler Pro (950040015, Hamburg, Germany), and all

images and gels were photographed using a UVP GelDoc-It® TS2 imager (71004-578, Upland, California, USA).

2.02.01.02 Strains, media and growth conditions

Cultures of *S. clavuligerus* and *S. coelicolor* were grown at 28°C, and liquid cultures were grown with agitation at 220 rpm in baffled flasks or flasks with springs, unless stated otherwise. *S. clavuligerus* strains were cultured on ISP-4 (M359, HiMedia, Mumbai, India) plates or in Trypticase Soy Broth (TSB) (236950, BD Biosciences, Franklin Lakes, New Jersey, USA) supplemented with 1% w/v starch (TSBS) (S9765, Sigma-Aldrich, St. Louis, Missouri, USA). *S. coelicolor* strains were cultured on Soy Flour Mannitol (SFM) (Kieser et al. 2000) plates or in broth consisting of a 1:1 ratio of Yeast Extract Malt Extract (YEME) (Kieser et al. 2000) and TSB media. When required, the appropriate antibiotics were included as listed in Table 2.01, unless otherwise specified.

E. coli was grown at 37°C and liquid cultures were agitated at 200 rpm. Strains were maintained on LB (Lennox) (BP1426500, Thermo Fisher Scientific, Waltham, Massachusetts, USA) agar plates or in LB broth, supplemented with appropriate antibiotics as listed in Table 2.01, or as otherwise specified. A list of all the strains and plasmids used throughout this study can be found in Table 2.02.

2.02.01.03 Standard DNA manipulation procedures

Standard procedures (agarose gel electrophoresis, etc.) were performed as described previously (Sambrook and Russell 2001). For the isolation of DNA fragments, agarose gels were visualized under ultraviolet (UV) light using a UVP benchtop UV

transluminator (95-0452-01, Upland, California, USA). DNA was extracted from agarose gel slices using either the EZ-10 Spin column DNA gel extraction kit (BS353, Bio Basics Inc, Markham, Ontario) or the Wizard® SV Gel and PCR Clean-Up System (A9281, Promega, Fitchburg, Wisconsin, USA) according to the manufacturer's instructions, with the exception that 33 µl of sterile water warmed to 55°C was used for elution. For plasmid DNA isolation from *E. coli* cultures, 3-5 ml of LB supplemented with appropriate antibiotics was inoculated with either a single colony of *E. coli* or with 5 µl of a glycerol stock, and the broth was incubated with shaking for 16-24 hours. Plasmid DNA was isolated from overnight cultures using the EZ-10 Spin Column Plasmid DNA kit (BS423, Bio Basics Inc, Markham, Ontario), with the exception that DNA was re-suspended in 33 µl of diH_2O. Alternatively, plasmid DNA was extracted using a modified Birnboim and Doly (1979) method as described before (Sambrook and Russell 2001).

2.02.01.04 Preparation and transformation of chemically competent *E. coli* cells

5 ml of 2× YT (1.6% w/v tryptone, 1% w/v yeast extract and 0.5% NaCl w/v) was inoculated with a single *E. coli* colony and grown overnight at 37°C, after which 200 ml of pre-warmed 2× YT was inoculated with 2 ml of the overnight growth. This was agitated at 200 rpm with 5:1 headspace volume at 37°C until the OD_{600} of the culture reached 0.2, following which $MgCl_2$ was added (final concentration 20 mM) and culture was returned to the incubator until the OD_{600} was between 0.45-0.55. The cultures were transferred to 50 ml round bottom centrifuge tubes and were incubated on ice for 2 hours. The cells were harvested by centrifugation at 4°C and were re-suspended in 100 ml of fresh, filter sterilized 1× $Ca^{2+}Mn^{2+}$ solution (0.394 g anhydrous NaOAc, 1.76 g

$CaCl_2$•$2H_2O$, and 1.66 g $MnCl_2$•$4H_2O$, pH5.5 in 60 ml diH_2O) pre-chilled at 4°C and were incubated on ice for an additional 45 minutes. The cells were harvested by centrifugation at 4°C and re-suspended in 10 ml of 1× $Ca^{2+}Mn^{2+}$ solution containing 15% v/v glycerol, and aliquots (200 µl each) were dispensed into pre-chilled 1.5 ml Eppendorf tubes (-20°C for at least 1 hour), and were flash-frozen in liquid nitrogen for storage at -80°C.

To transform the competent cells, aliquots of the chemically competent *E. coli* cells were thawed on ice. Meanwhile, 5 µl of plasmid DNA was combined with diH_2O to bring up the volume to 25 µl in a separate tube, which was then added to 50 µl of the chemically competent *E. coli* and incubated on ice for 30 minutes. Cells were heat shocked for 5 minutes at 37°C, and quickly transferred to 975 µl of ice-cold LB broth in a 1.5 ml Eppendorf tube and incubated for 1 hour at 37°C. The tubes were centrifuged for 2 minutes at 12000 rpm and 900 µl of supernatant was removed and the remaining 100 µl of suspension was plated onto a single LB agar plate containing the appropriate antibiotics.

2.02.02 Genomic manipulation of *Streptomyces* spp.

2.02.02.01 Conjugation with *S. coelicolor*

This procedure was carried out as previously described (Kieser et al. 2000), except that the *S. coelicolor* recipient strain was cultured in 25 ml of 1:1 YEME:TSB for 36 hours, and the *E. coli* donor strain was cultured in 50 ml of LB. The entire conjugation mix was spread onto SFM agar plates and incubated at 28°C, and after 24 hrs the plates were inspected for recovery (sparse growth) and were flooded with the appropriate

antibiotics in water as described previously (Kieser et al. 2000). When isolated colonies were observed after incubation, they were patched onto fresh SFM plates. Following sporulation, spores stocks were prepared for storage by suspending spores in 20% glycerol.

2.02.02.02 Generating *S. clavuligerus* protoplasts

S. clavuligerus spore stocks (10 µl) were used to inoculate 125 ml spring flasks containing 15 ml YEME with 10 ml of TSBS supplemented with 0.5% w/v glycine and appropriate antibiotics. The cultures were incubated at 28^{o}C for 48 hours at 200 rpm, after which the mycelia were harvested by centrifugation and were washed twice with 10 ml of sterile solution of 10.3% w/v sucrose. The mycelia were either frozen at -80^{o}C or were used for protoplast preparation immediately. If frozen mycelia were used for preparing protoplasts, the samples were washed once in 10.3% w/v sucrose prior to use. Washed mycelia were re-suspended in 4 ml of P-Buffer [10.3% w/v sucrose, 0.057 mM K_2SO_4, 10 mM $MgCl_2$, 0.005% w/v KH_2PO_4, 0.368% w/v $CaCl_2$, 0.573% w/v TES (pH 7.2), 0.59 µM $ZnCl_2$, 1.48 µM $FeCl_2$, 0.12 µM $CuCl_2$, 0.10 µM $MnCl_2$, 0.052 µM $Na_2B_4O_7$ and 0.016 µM $(NH_4)_6Mo_7O_{24}$] (Kieser et al. 2000) + 2 mg/ml lysozyme and incubated at 28^{o}C for 1 hour, with gentle manual rotation every 10-15 minutes. An additional 5 ml of P buffer was then added to the sample and the protoplasts were filtered through a sterile non-absorbent cotton wool plug placed in a sterile 5 ml syringe. The protoplasts were then harvested by centrifugation and washed twice with 10 ml of 10.3% w/v sucrose, and were re-suspended in 5 ml of P buffer. Aliquots (1 ml) were dispensed into 1.5 ml Eppendorf tubes and were either used immediately, or were slowly frozen for long-term

storage by placing the tubes on ice, followed by storage at -80^{o}C.

2.02.02.03 Generating *S. coelicolor* protoplasts

S. coelicolor protoplasts were prepared using the same procedure as used for *S. clavuligerus* except that 25 ml of 1:1 YEME: TSB was used as the growth medium.

2.02.02.04 Protoplast fusion

Both parental (*S. clavuligerus* Δ*cvm6p::neo* and Δ*bls1::tsr*) and recipient (*S. coelicolor* M1146/pIJ8660, M1154/pIJ8660, M1154/pIJ8660) protoplast suspensions were adjusted to have approximately the same turbidity, as judged visually, and 500 µl of each were mixed. Protoplasts were harvested by centrifugation and the supernatant was discarded. The protoplast pellet was then re-suspended by gentle tapping in the remaining drop of liquid and was washed with 5 ml of P buffer followed by centrifugation. The supernatant was discarded and the protoplasts were immediately re-suspended in 0.8 ml of 50% w/v PEG 1000 in P buffer by gentle pipetting. The tubes were incubated for 2 minutes at room temperature, after which the protoplasts were plated onto fresh R2YE (Kieser et al. 2000) [(35% sucrose, 0.025% K_2SO_4, 0.1012% $MgCl_2$ $6H_2O$, 0.10% glucose, 0.010% casaminoacids, in di$H2_2O$ + 2.2% biological agar) + (per 100 ml: 5 ml 10% yeast extract, 1 ml 0.5% KH_2PO_4, 8.0 ml 3.86% $CaCl_2$ $2H_2O$, 1.5 ml 20% L-proline, 10 ml 5.73 TES (pH7.2), 0.59 µM $ZnCl_2$, 1.48 µM $FeCl_2$, 0.12 µM $CuCl_2$, 0.10 µM $MnCl_2$, 0.052 µM $Na_2B_4O_7$, 0.016 µM $(NH_4)_6Mo_7O_{24}$, and 0.5 ml 1 N NaOH)] without sucrose (R2S) using a Pasteur pipette with a bent tip in order to ensure that only light pressure was applied to the protoplasts. The plates were incubated at 28^{o}C for 16-20 hours before being flooded with 1ml of sterile diH_2O containing 1250 µg apramycin and

1800 μg thiostrepton or 1250 μg kanamycin. Following 72 hours of incubation, colonies were picked and patched into quadrants on R2S containing 100 μg/ml thiostrepton and were incubated at 28^{o}C for 72 hours. After double selection, the patches were streaked for lawns of growth onto R2YE and were incubated for 7 days at 28^{o}C, following which portions were streaked again onto R2S with 100 μg/ml thiostrepton to confirm resistance. The plates were harvested to make glycerol spore stocks in the case of the M1146 strains, and glycerol mycelial stocks in the case of the M1152 and M1154 strains. The thiostrepton resistant *S. coelicolor* strains potentially harboring the *S. clavuligerus* plasmid were referred to as *Sx* 1146/1152/1154, henceforth.

2.02.03 Chromosomal DNA preparation and confirmation of strain identity

2.02.03.01 Preparation of chromosomal DNA from *Streptomyces* spp.

Chromosomal DNA was prepared from each strain using a modified phenol-chloroform extraction method (Kieser et al. 2000). Mycelia from 10 ml cultures (grown using TSBS and 1:1 TSB:YEME media for *S. clavuligerus* and *S. coelicolor*, respectively) were harvested and washed twice in 10.3% w/v sucrose. The mycelial pellets were either frozen at -80^{o}C for future use or were immediately processed. Mycelia (1ml) were transferred to 15 ml Falcon tubes containing 2 ml of lysozyme buffer (25 mM Tris-HCl pH 8.0, 0.3 M sucrose, 25 mM EDTA, 4 mg/ml lysozyme, and 50 μl/ml RNaseA) and the tubes were incubated at 37^{o}C for 30-45 minutes with gentle manual inversion every 5 to 10 minutes. Next, 500 μl of a sterile 2% w/v SDS solution was added and the tubes were shaken vigorously until the viscosity of the contents decreased. A phenol-chloroform-isoamyl alcohol (25:24:1) solution (500μl) was carefully added to the

tubes, which were then sealed with Parafilm M® and were shaken until the liquid inside developed a "milky" appearance. The tubes were centrifuged at 10000 rpm for 5 minutes, after which the top aqueous phase was transferred to a fresh tube. The DNA was precipitated by the addition of 300 µl of a 3M sodium acetate solution and 1 volume of 100% isopropanol. The tubes were inverted gently and incubated on ice for 10 minutes before vortexing until clumps of DNA became visible. The DNA was carefully spooled and transferred using a pipet tip to Eppendorf tubes containing 700ul of ice-cold 70% ethanol, and the tubes were immediately placed at -20°C for 24 hours. Later, the DNA was harvested by centrifugation and was transferred to new tubes containing 700ul of 70% v/v ethanol, and after incubation for 24 hours the tubes were centrifuged at 10000 rpm for 5 minutes and the supernatant was discarded. The tubes were spun again to compact the pellet further and any residual supernatant was removed using a pipette. The pellets were air-dried for 10 minutes in a biosafety cabinet and an appropriate amount of water was added followed by incubation overnight at 37°C to re-dissolve the DNA. Samples were then frozen at -20°C until further use.

Chromosomal DNA from 10 ml cultures of *S. clavuligerus* and *S. coelicolor* was also isolated using a modified protocol and the DNeasy kit (69506, Qiagen, Venlo, Netherlands). Buffers AL and ATL (200µl) were added to mycelia in a 3 ml screw cap tube that also contained acid-washed beads (BAWZ 400-250-35, OPS Diagnostics, Lebanon, NJ, USA). The samples were homogenized using the SpeedMill PLUS Bead Homogenizer (845-00008-2, Analytik Jena AG, Jena, Germany) with the pre-programmed "Bacteria" protocol (two three-minute on and off cycles). After

centrifugation for 1 minute at 10000 rpm, the supernatants were transferred to fresh 1.5 ml Eppendorf tubes containing 200 µl of 95% v/v ethanol and were vortexed briefly, followed by centrifugation. Next, the supernatants were transferred to spin columns, and the columns were centrifuged at 8000 rpm for 1 minute and the flow-through liquids were discarded. Then, the AW2 buffer (500µl) was added to each spin column, and the columns were centrifuged at 14000 rpm for 3 minutes. Again, the flow-through liquids were discarded, and any residual buffer was removed from each column by an additional centrifugation for 1 minute at 14000 rpm. The spin columns were each placed into a fresh 1.5 ml Eppendorf tube and the DNA was eluted by adding 100 µl of diH_2O to each column and then incubating the column at room temperature for 2-3 minutes followed by centrifugation at 8000 rpm for 1 minute. This was repeated using an additional 100µl of diH_2O in order to maximize the amount of DNA recovered.

2.02.03.02 Confirming the *Sx* strain identity

Confirmation of host strain identity was carried out using PCR amplification and subsequent Sanger sequencing of the products to identify mutations in the *rpoB* and *rpsL* genes (Gomez-Escribano and Bibb 2011; Hu et al. 2002; Shima et al. 1996). Suspected strains isolated by protoplast fusion were subject to a series of PCR amplifications to identify the potential presence of *S. clavuligerus* DNA. To eliminate ambiguity between *S. clavuligerus* and *S. coelicolor*, the *dagA* (not present in *S. clavuligerus*) was amplified to confirm that the isolates were *S. coelicolor.* In order to check for the transfer of the *S. clavuligerus* plasmid, ten regions (~180kbp apart) were selected from the plasmid using the Artemis software (Rutherford et al. 2000). The targeted regions were amplified using

the primers described in Table 2.03 and were Sanger Sequenced at The Centre for Applied Genomics in Toronto, Ontario or at Laboratorio Nacional de Genomica Para la Biodiversidad in Irapuato, México. The PCR products were generated using KAPA Taq (BK1000, KAPABiosystems, Wilington, Massachusetts, USA) as per the following thermocycling conditions: initial denaturation at 95°C for 3 minutes followed by 35 cycles of denaturation at 95°C for 30 seconds, annealing at 55°C for 30 seconds, extension at 72°C for 45 seconds, and final extension at 72°C for 5 minutes.

2.02.04 Bioinformatics

2.02.04.01 Phylogenetic analysis of Actinomycetes

RpoB Actinomycete protein sequences (146 in total) were obtained from the National Center for Biotechnology Information (NCBI). Thirty-seven 16S rDNA sequences (E-Value 0.0, Identity >97%) were obtained from NCBI based on *S. clavuligerus* sequence homology. The RpoB and 16S rDNA sequences from *Catenulispora acidiphila* were included as out-groups, and the *S. cattleya* sequences were also included since this species is also known to have a megaplasmid (Barbe et al. 2011). Sequences were aligned using the MEGA6.0 software (Tamura et al. 2013) with the Multiple Sequence Comparison by Log-Expectation (MUSCLE) algorithm (Edgar 2004). Phylogenetic trees were generated using the MrBayes3.2 software with set parameters (model=4by4 or protein, NST=mixed, covarion=No, # States=4, Rates=Invgamma, Generations=1000000) (Ronquist et al. 2012), and the trees were graphically viewed using FIGTree1.4 (Rambaut 2013). Plasmids were identified by mining NCBI for published sequences indicating the presence of extra chromosomal elements.

Comparative synteny plots were generated using the R2CAT software (Husemann and Stoye 2010).

2.02.04.02 Large plasmid annotation

SM biosynthetic gene clusters on the *S. clavuligerus* (GI: 294810834 and GI: 294816545) (Medema et al. 2010) and *S. cattleya* (GI: 357397620 and GI: 357407371) (Barbe et al. 2011) megaplasmids were annotated using antiSMASH2.0 (Medema et al. 2011), with further annotation being carried out manually using Artemis (Rutherford et al. 2000). Hypothetical or putative proteins and pathways were further annotated using Pfam (Punta et al. 2012) and SBSPKS (Anand et al. 2010). Gene cluster homology was analyzed using MultiGeneBlast (Medema et al. 2013).

2.03 Results and Discussion

2.03.01 Identification and comparison of sequenced megaplasmids in *Streptomyces*

In order to determine the uniqueness of pSCL4, the NCBI genome database was mined for *Streptomyces* that harbor megaplasmids. Only one other species (*Streptomyces cattleya*) was identified with a sequenced megaplasmid (Barbe et al. 2011). Using 16S rDNA and RpoB as a marker, a species tree was generated and analyzed in the context of the evolutionary history of large plasmids (Supplementary Figures 1 and 2). Even though both *S. clavuligerus* and *S. cattleya* harbor megaplasmids, they themselves are not closely related based on the analysis. The *S. cattleya* plasmid (pSCAT) is 1.8 Mbp and is predicted to encode 1713 proteins (Barbe et al. 2011) as compared to pSCL4, which encodes 1581 predicted proteins (Medema et al. 2010). Using AntiSMASH2.0, it was predicted that pSCAT carries 25 SM biosynthetic gene clusters, including NRPSes, β-lactams, siderophores, PKSes, and other putative pathways (Table 2.04). To explore the similarities between the two megaplasmids further, a comparative synteny plot was generated (Supplementary Figure 3). The plot identified 87 similar genes, with no extensive similarities otherwise. This indicates that both plasmids are unique and arose independently, perhaps through a chromosomal arm breaking off or the chromosome undergoing recombination with a plasmid. Both megaplasmids are loaded with SM biosynthetic gene clusters; however, their similarities appear to end there (Supplementary Figures 4 and 5).

2.03.02 Attempted mobilization of pSCL4 based on antibiotic resistance selection marker transfer

The transfer of pSCL4 was attempted by fusing protoplasts of the *S. clavuligerus* donor strain and the *S. coelicolor* host strains. Selection of plasmid transfer from donor to host was based on resistance phenotypes and colony morphology. The pSCL4 donor strain was *S. clavuligerus* Δ*bls1::tsr,* which harbors the thiostrepton resistance gene cassette within the pSCL4 - borne clavulanic acid paralogue gene cluster (Jensen et al. 2000). To select for the *S. coelicolor* host strains, the pIJ8660 integrative plasmid (Sun et al. 1999) was first introduced into each strain to give *S. coelicolor* M1146/pIJ8660, M1152/pIJ8660 and M1154/pIJ8660. The pIJ8660 plasmid integrates into the ϕC31 phage integration site on the *S. coelicolor* chromosome and harbors the apramycin resistance cassette, allowing for selection of the *S. coelicolor* strains in the presence of apramycin. The *S. coelicolor* host strains containing pIJ8660 were tested for apramycin resistance and were shown to be able to grow on R2S with 25 μg/ml apramycin (Figure 2.02-I). As expected, none of the host strains could grow on R2S containing 25 μg/ml apramycin and 60 μg/ml thiostrepton (Figure 2.02-II). The products of protoplast fusion using *S. clavuligerus* Δ*bls1::tsr* and *S. coelicolor* M1154/pIJ8660 showed resistance to both apramycin and thiostrepton (Figure 2.02-III), suggesting that they contain both the M1154/pIJ8660 chromosome harboring the apramycin resistance cassette and the pSCL4 plasmid carrying the thiostrepton resistance gene. Protoplast fusions using M1146 and M1152 did not result in the recovery of any colonies and were therefore not pursued further. Interestingly, each of the three apramycin and thiostrepton resistant colonies

selected for downstream analysis (and designated *Sx* 1, *Sx* 3 and *Sx* 5) showed different phenotypes when grown on R2S. For example, *Sx* 1 (III-E) had distinctly small colonies and grew sparsely. It also grew substantially slower than the other two strains and exhibited delayed aerial mycelium formation and weak sporulation (7 days after *Sx* 3 and *Sx* 5). *Sx* 3 (III-F) was found to grow better than *Sx* 1; however, it lagged behind *Sx* 5 (III-G). Based on these morphological traits, *Sx* 1 was selected as an ideal candidate for downstream genotyping.

2.03.03 Analysis of the *Sx* 1 strain arising from protoplast fusion using PCR "fishing"

Ten screening primer sets were designed based on the DNA sequence of pSCL4, with each set targeting a ~600bp region, ~180kbp apart. Oligonucleotide primer sets specific to *S. clavuligerus* and *S. coelicolor* chromosomal DNA were used to differentiate between the two species using PCR, based on the *pcbAB* and *dagA* genes, respectively. The *tsr* resistance cassette was also targeted for amplification to confirm the transfer of the resistance gene. Positive control reactions were first set up in order to test the primer sets using genomic DNA from the donor strain, *S. clavuligerus* Δ*bls1::tsr*. As shown in Figure 2.03-I, PCR products of the expected size were seen for all 10 of the targeted regions from pSCL4. Bands of the expected size were also seen for *pcbAB, tsr* and *rpsL,* and some non-specific amplification was observed for *dagA* (smearing), which is consistent with the absence of this gene from *S. clavuligerus*. Amplification reactions using genomic DNA from *S. coelicolor* M1154/pIJ8660 confirmed that the fishing primers do not amplify any native chromosomal regions (*S. coelicolor* M1154 is pSCP1$^-$

and pSCP2$^-$) (Gomez-Escribano and Bibb 2011), and no products were observed using the *pcbAB*-specific primers. As expected, PCR products of the correct size were seen in the case of the *dagA* and *rpsL* gene-specific primers (Figure 2.03-III).

The pSCL4 fishing primer sets, the *tsr*-specific primer set and the chromosome-specific primer sets were used in PCR reactions along with genomic DNA from *Sx* 1, *Sx* 3 and *Sx 5* in order to detect the presence of pSCL4 and the chromosomal loci specific to *S. coelicolor* or *S. clavuligerus*. As expected, *Sx* 1 genomic DNA resulted in PCR products with the correct size for both the *dagA* and *rpsL* regions, confirming that the chromosome of *Sx* 1 was from *S. coelicolor* M1154/pIJ8660 (Figure 2.03-II). However, no obvious PCR products for the *tsr* resistance gene could be detected, nor were products observed using any of the pSCL4 fishing primer sets (Figure 2.03-II). Similar results were obtained using genomic DNA from *Sx* 3 and *Sx* 5 (results not shown). This indicates that pSCL4 was not transferred from *S. clavuligerus* to *S. coelicolor* in the protoplast fusion experiment.

Thiostrepton is a macrocyclic thiopeptide that inhibits translation by interacting with prokaryotic ribosomes (Walter et al. 2012). It does this by binding the ribosome within the GTPase-associated centre between the N-terminal domain and the L11/23S rRNA loops (Walter et al. 2012). However, the exact nature of thiostrepton's mechanism of action (MOA) is still debated (Walter et al. 2012). Spontaneous thiostrepton resistance can be traced back to the loss of ribosomal protein L11 (encoded by *rpl*K) in *Bacillus* spp. or a 6 bp deletion in the *rpl*K gene in *S. coelicolor* A3(2) (Bascaran et al. 1991; Kawamoto et al. 1997; Ochi et al. 1997). The mutants in these previous studies were generated using low concentrations of thiostrepton (<5µg/ml); furthermore, spontaneous

mutations are unstable and sensitivity can usually be re-attained through re-selection on replicate media laced with higher concentrations of thiostrepton (Kieser et al. 2000). A drawback of protoplast regeneration is that it can result in a number of genetic changes, resulting in phenotype alterations (Ochi 1990). This includes increased antibiotic resistance, in some cases spontaneous kanamycin resistance as high as 1000 μg/ml (Hotta et al. 1988). This could explain the high level of thiostrepton resistance found in *S. coelicolor* M1154/pIJ8660 after protoplast fusions, without the transfer of pSCL4 carrying the *tsr* resistance gene. The possibility of confirming the mechanisms of resistance observed in these protoplast fusions is discussed in the summary.

Protoplast regeneration is a sensitive process, protoplasts lack a cell well and will lyse with ease. Sucrose has been used with success as an osmotic stabilizer, helping prevent protoplast lysis and increase protoplast regeneration efficiency (Illing et al. 1989). During preliminary experiments, recovering *S. coelicolor* M1154 protoplasts on R2YE (including sucrose) resulted in spontaneous resistance to thiostrepton. Therefore, sucrose was omitted from the recovery media as described by Jones et al. (2013), where thiostrepton was used as a selective marker in *S. coelicolor* M1154. The exclusion of sucrose and increased risk of spontaneous resistance brings into question the use of thiostrepton as a marker during protoplast fusion studies.

2.03.04 Confirmation of the *rpoB* and *rpsL* mutation status in *Sx* 1

PCR analysis was conducted in order to confirm that the *rpoB* and *rpsL* genes in strain *Sx* 1 retained the advantageous point mutations that provide increased SM production in *S. coelicolor* M1154 (Gomez-Escribano and Bibb 2011). Sequencing of the

rpoB and *rpsL* PCR products from the parent *S. coelicolor* M1154/pIJ8660 strain revealed the presence of the expected *rpo*B*[C1298T]* and *rpsL[A262G]* mutations (Supplementary Figure 6), confirming the identity of the strain as *S. coelicolor* M1154. Subsequent sequencing of the *rpoB* and *rpsL* genes from strain *Sx* 1 on the other hand, revealed a different story. In all three strains, both the *rpo*B*[C1298T]* and *rpsL[A262G]* mutations were absent (Supplementary Figure 6). The loss of these mutations could be due to reversions during protoplast preparation/regeneration or due to homologous recombination with *S. clavuligerus* DNA during protoplast fusion. This indicates that low-level selection using rifampin and streptomycin may always be required when working with *S. coelicolor* M1154 to maintain the mutations during protoplast fusion and regeneration.

2.03.05 Repeat of protoplast fusion experiment using kanamycin as the resistance marker on pSCL4

Due to problems with spontaneous thiostrepton resistance in *S. coelicolor* M1154/pIJ8660 as described above, two more attempts at protoplast fusion were done using the *S. clavuligerus* donor strain Δ*cvm6p::neo*, which harbors the neomycin (kanamycin) resistance gene cassette on pSCL4 (Tahlan et al. 2007). This was done to test the hypothesis that spontaneous resistance to kanamycin may be less likely to arise in *S. coelicolor* M1154/pIJ8660 as there are no reports of point mutations leading to kanamycin resistance at significant frequencies. The protoplast fusions were repeated twice using *S. clavuligerus* Δ*cvm6p::neo* and *S. coelicolor* M1154/pIJ8660 and recovered on R2S; however, viable protoplast fusions or spontaneous kanamycin resistance in *S.*

coelicolor M1154/pIJ8660 were not observed (Figure 2.04). Repeating the fusions with R2YE may increase efficiency. The other *S. coelicolor* super-strains (M1146 and M1152) were not tested because they previously did not produce viable fusions; furthermore, M1154 produces greater amounts of heterologous SMs compared to M1152 and M1146 (Gomez-Escribano and Bibb 2011).

2.03.06 *S. coelicolor* M1154 is not the ideal host for protoplast fusion - mediated transfer of pSCL4

The primary goal of this work was to mobilize pSCL4 from *S. clavuligerus* into *S. coelicolor*. Admittedly, this was a lofty goal; however, the benefit of introducing ≥ 25 SM biosynthetic gene clusters into a host that was specifically engineered for SM overproduction far outweighed the difficulties associated with transferring a 1.8Mbp plasmid. Bioinformatics analyses have predicted a wealth of interesting potential pSCL4 - encoded SMs including a meonomycin-like compound, NRPSes with keto-transferase independent biosynthesis, and NRPS/PKS hybrids (Medema et al. 2010; Song et al. 2010).

One of the primary assumptions of this work was that pSCL4 is a plasmid and can transfer between strains and replicate independently. Protoplast fusion is a common method for transferring DNA segments between *Streptomyces* (Baltz and Matsushima 1981). It is efficient and does not depend on external genetic influences. Multiple PCR amplifications on three different protoplast fusions failed to produce any signal for pSCL4 from *S. coelicolor* M1154/pIJ8660. This suggests the possibility there is a mechanism preventing *S. clavuligerus* DNA from entering *S. coelicolor*.

Toxin-antitoxin (TA) systems have been described in *Streptomyces spp.* (Sevillano et al. 2013). These are plasmid stabilization systems where TA proteins are encoded by the chromosome or plasmid (Hayes 2003). Normally the toxin has a longer half-life than the antitoxin and the binding of the two neutralizes the activity of the toxin. In case the replicon encoding the genes is lost, the antitoxin is degraded at a faster rate leaving unbound toxin protein, which is deleterious to the cell (Hayes 2003). These functionally paired proteins are usually only 75 to 100 amino acids long and are highly divergent. TAs are specialized to their particular organism making them difficult to detect using bioinformatics (Hayes 2003). Based on DNA sequence data, pSCL4 encodes at least 4 putative TA - related proteins, which could somehow influence or prevent its entry and establishment in *S. coelicolor*. However, further experimental evidence is required in order to test this hypothesis.

A more plausible explanation as to why pSCL4 was not transferred to *S. coelicolor* could involve the complex restriction endonuclease system found in this organism. Most of the described work with the *S. coelicolor* super hosts involves introducing gene clusters through a non-methylating intermediate host such as *E. coli* ET12567 (Gomez-Escribano and Bibb 2011; Jones et al. 2013). It is well established that *S. coelicolor* does not readily take up DNA from other *Streptomyces* spp. directly (Garcia-Dominguez et al. 1987; Kieser et al. 2000; MacNeil 1988), and historically, alternate methods were used when transforming DNA into *S. coelicolor*. *Streptomyces* spp. are home to over 600 restriction endonucleases and some 100 methylases (MacNeil 1988). These systems usually involve a methylase that methylates host-specific DNA sequences and a corresponding endonuclease, which cleaves foreign DNA that does

contain the specific DNA sequences and that does not match the host methylation pattern (MacNeil 1988). In *E. coli,* DNA is methylated *in vivo* by Dam, Dcm, and Hsd is restricted by *S. coelicolor* (Gonzalez-Ceron et al. 2009). Furthermore, *S. coelicolor* also restricts methylated DNA under *in vitro* conditions (Gonzalez-Ceron et al. 2009). Null-mutants of Dcm and Dam endonucleases in *S. coelicolor* do not alleviate restriction, and novel restriction genes have been detected (Gonzalez-Ceron et al. 2009). These studies noted that the *S. coelicolor* genome has evolved to maintain a strict restriction system and strongly suggests that a non-methylating donor host is required (such as *E. coli* ET12567). A possible solution for introducing pSCL4 into *S. coelicolor* may be the use of an intermediate donor such as *Streptomyces lividans*, which is more closely related to *S. coelicolor* A3(2) than *S. clavuligerus* (Supplementary Figures 1 and 2). *S. lividans* has a unique *in vitro* restriction system that makes it an ideal host for the uptake of DNA from other *Streptomyces* spp. (Zhou et al. 1994). It is known that *S. lividans* 66 is readily transformed by *S. clavuligerus* DNA (Garcia-Dominguez et al. 1987) and it has been shown to act as a non-methylating intermediate for the transfer of DNA into *S. coelicolor* (Zotchev et al. 1995).

The *S. coelicolor* super-strains have been effective at over-expressing heterologous antibiotic gene clusters cloned into cosmid vectors, which are introduced through an intermediate non-methylating *E. coli* host. The transfer of megaplasmids directly into *S. coelicolor* M1154 is probably not a viable method for introducing biosynthetic pathways and activating cryptic gene clusters. The advantageous mutations within *S. coelicolor* M1154 also need to be maintained with basal levels of rifampicin and streptomycin, during protoplast regeneration. However, using the lessons learned during

the current study, we hope to circumnavigate the *S. coelicolor* restriction system by repeating the above-described experiment with *S. lividans* 66 as a host for pSCL4. In addition, *S. lividans* 66 can then be bioengineered to perhaps act as a SMs producing super strain (through mutations of *rpoB* and *rpsL* genes). However, *S. lividans* 66 does not produce SMs at the same level compared to its close relative, *S. coelicolor* A3(2) (Shima et al. 1996). Alternatively, *S. lividans* 66 could act as an intermediate host for pSCL4 between *S. clavuligerus* and *S. coelicolor* M1154. If this were successful, it would be interesting to begin transferring other megaplasmids through *S. lividans* 66 into *S. coelicolor* M1154, such as the *S. cattleya* megaplasmid to investigate and activate the production of biologically relevant SMs for future applications.

BIBLIOGRAPHY

Alvarez-Alvarez, R., Rodriguez-Garcia, A., Martinez-Burgo, Y., Robles-Reglero, V., Santamarta, I., Perez-Redondo, R., et al. 2014. A 1.8-Mb-reduced *Streptomyces clavuligerus* genome: relevance for secondary metabolism and differentiation. Appl Microbiol Biotechnol 98(5): 2183-2195. doi: 10.1007/s00253-013-5382-z.

Anand, S., Prasad, M.V., Yadav, G., Kumar, N., Shehara, J., Ansari, M.Z., et al. 2010. SBSPKS: structure based sequence analysis of polyketide synthases. Nucleic Acids Res 38(Web Server issue): W487-496. doi: 10.1093/nar/gkq340.

Baltz, R.H., and Matsushima, P. 1981. Protoplast fusion in *Streptomyces*: conditions for efficient genetic recombination and cell regeneration. J Gen Microbiol 127(1): 137-146. doi: 10.1099/00221287-127-1-137.

Barbe, V., Bouzon, M., Mangenot, S., Badet, B., Poulain, J., Segurens, B., et al. 2011. Complete genome sequence of *Streptomyces cattleya* NRRL 8057, a producer of antibiotics and fluorometabolites. J Bacteriol 193(18): 5055-5056. doi: 10.1128/JB.05583-11.

Bascaran, V., Sanchez, L., Hardisson, C., and Brana, A.F. 1991. Stringent response and initiation of secondary metabolism in *Streptomyces clavuligerus*. J Gen Microbiol 137(7): 1625-1634. doi: 10.1099/00221287-137-7-1625.

Benson, D.A., Clark, K., Karsch-Mizrachi, I., Lipman, D.J., Ostell, J., and Sayers, E.W. 2014. GenBank. Nucleic Acids Res 42(Database issue): D32-37. doi: 10.1093/nar/gkt1030.

Bethesda Research Laboratories. 1986. BRL pUC host: *E. coli* DH5α competent cells. Focus 8(2): 9.

Birnboim, H.C., and Doly, J. 1979. A rapid alkaline extraction procedure for screening recombinant plasmid DNA. Nucleic Acids Res 7(6): 1513-1523.

Blin, K., Medema, M.H., Kazempour, D., Fischbach, M.A., Breitling, R., Takano, E., et al. 2013. antiSMASH 2.0--a versatile platform for genome mining of secondary metabolite producers. Nucleic Acids Res 41(Web Server issue): W204-212. doi: 10.1093/nar/gkt449.

Edgar, R.C. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res 32(5): 1792-1797. doi: 10.1093/nar/gkh340.

Garcia-Dominguez, M., Martin, J.F., Mahro, B., Demain, A.L., and Liras, P. 1987. Efficient plasmid transformation of the beta-lactam producer *Streptomyces clavuligerus*. Appl Environ Microbiol 53(6): 1376-1381.

Gomez-Escribano, J.P., and Bibb, M.J. 2011. Engineering *Streptomyces coelicolor* for heterologous expression of secondary metabolite gene clusters. Microb Biotechnol 4(2): 207-215. doi: 10.1111/j.1751-7915.2010.00219.x.

Gonzalez-Ceron, G., Miranda-Olivares, O.J., and Servin-Gonzalez, L. 2009. Characterization of the methyl-specific restriction system of *Streptomyces coelicolor* A3(2) and of the role played by laterally acquired nucleases. FEMS Microbiol Lett 301(1): 35-43. doi: 10.1111/j.1574-6968.2009.01790.x.

Hayes, F. 2003. Toxins-Antitoxins: Plasmid Maintenance, Programmed Cell Death, and Cell Cycle Arrest. Science 301(5639): 1496-1499. doi: 10.1126/science.1088157.

Hotta, K., Ishikawa, J., Ichihara, M., Naganawa, H., and Mizuno, S. 1988. Mechanism of increased kanamycin-resistance generated by protoplast regeneration of *Streptomyces griseus*. I. Cloning of a gene segment directing a high level of an aminoglycoside 3-N-acetyltransferase activity. J Antibiot (Tokyo) 41(1): 94-103. doi: 10.7164/antibiotics.41.94.

Hu, H., Zhang, Q., and Ochi, K. 2002. Activation of antibiotic biosynthesis by specified mutations in the *rpoB* gene (encoding the RNA polymerase beta subunit) of *Streptomyces lividans*. J Bacteriol 184(14): 3984-3991. doi: 10.1128/JB.184.14.3984-3991.2002.

Husemann, P., and Stoye, J. 2010. r2cat: synteny plots and comparative assembly. Bioinformatics 26(4): 570-571. doi: 10.1093/bioinformatics/btp690.

Illing, G.T., Normansell, I.D., and Peberdy, J.F. 1989. Protoplast isolation and regeneration in *Streptomyces clavuligerus*. J Gen Microbiol **135**(8): 2289-2297.

Jensen, S.E., Elder, K.J., Aidoo, K.A., and Paradkar, A.S. 2000. Enzymes catalyzing the early steps of clavulanic acid biosynthesis are encoded by two sets of paralogous genes in

Streptomyces clavuligerus. Antimicrob Agents Chemother 44(3): 720-726. doi: 10.1128/AAC.44.3.720-726.2000.

Jones, A.C., Gust, B., Kulik, A., Heide, L., Buttner, M.J., and Bibb, M.J. 2013. Phage p1-derived artificial chromosomes facilitate heterologous expression of the FK506 gene cluster. PLoS One 8(7): e69319. doi: 10.1371/journal.pone.0069319.

Kawamoto, S., Zhang, D., and Ochi, K. 1997. Molecular analysis of the ribosomal L11 protein gene (*rpl*K = *rel*C) of *Streptomyces griseus* and identification of a deletion allele. Mol Gen Genet 255(6): 549-560. doi: 10.1007/s004380050528.

Kieser, T., Bibb, M.J., Buttner, M.J., Chater, K.F., and Hopwood, D.A. 2000. Practical *Streptomyces* Genetics. 2nd ed. John Innes Foundation, Norwich, U.K.

Kinashi, H. 2011. Giant linear plasmids in *Streptomyces*: a treasure trove of antibiotic biosynthetic clusters. J Antibiot (Tokyo) 64(1): 19-25. doi: 10.1038/ja.2010.146.

MacNeil, D.J. 1988. Characterization of a unique methyl-specific restriction system in *Streptomyces avermitilis*. J Bacteriol 170(12): 5607-5612.

MacNeil, D.J., Gewain, K.M., Ruby, C.L., Dezeny, G., Gibbons, P.H., and MacNeil, T. 1992. Analysis of *Streptomyces avermitilis* genes required for avermectin biosynthesis utilizing a novel integration vector. Gene 111(1): 61-68. doi: 10.1016/0378-1119(92)90603-M.

Medema, M.H., Blin, K., Cimermancic, P., de Jager, V., Zakrzewski, P., Fischbach, M.A., et al. 2011. antiSMASH: rapid identification, annotation and analysis of secondary metabolite biosynthesis gene clusters in bacterial and fungal genome sequences. Nucleic Acids Res 39(Web Server issue): W339-346. doi: 10.1093/nar/gkr466.

Medema, M.H., Takano, E., and Breitling, R. 2013. Detecting sequence homology at the gene cluster level with MultiGeneBlast. Mol Biol Evol 30(5): 1218-1223. doi: 10.1093/molbev/mst025.

Medema, M.H., Trefzer, A., Kovalchuk, A., van den Berg, M., Muller, U., Heijne, W., et al. 2010. The sequence of a 1.8-mb bacterial linear plasmid reveals a rich evolutionary reservoir of secondary metabolic pathways. Genome Biol Evol 2: 212-224. doi: 10.1093/gbe/evq013.

Netolitzky, D.J., Wu, X., Jensen, S.E., and Roy, K.L. 1995. Giant linear plasmids of beta-lactam antibiotic producing *Streptomyces*. FEMS Microbiol Lett 131(1): 27-34. doi: 10.1016/0378-1097(95)00230-3.

Ochi, K. 1990. Streptomyces *relC* mutants with an altered ribosomal protein ST-L11 and genetic analysis of a *Streptomyces griseus relC* mutant. J Bacteriol 172(7): 4008-4016.

Ochi, K., Zhang, D., Kawamoto, S., and Hesketh, A. 1997. Molecular and functional analysis of the ribosomal L11 and S12 protein genes (*rpl*K and *rpsL*) of *Streptomyces coelicolor* A3(2). Mol Gen Genet 256(5): 488-498. doi: 10.1007/s004380050593.

Overbye, K.M., and Barrett, J.F. 2005. Antibiotics: where did we go wrong? Drug Discov Today 10(1): 45-52. doi: 10.1016/S1359-6446(04)03285-4.

Paget, M.S., Chamberlin, L., Atrih, A., Foster, S.J., and Buttner, M.J. 1999. Evidence that the extracytoplasmic function sigma factor sigmaE is required for normal cell wall structure in *Streptomyces coelicolor* A3(2). J Bacteriol 181(1): 204-211.

Punta, M., Coggill, P.C., Eberhardt, R.Y., Mistry, J., Tate, J., Boursnell, C., et al. 2012. The Pfam protein families database. Nucleic Acids Res 40: D290-301. doi: 10.1093/nar/gkr1065.

Rambaut, A. 2013. FigTree v1.4. Available from http://tree.bio.ed.ac.uk/software/figtree/ [accessed 28 April 2014 2014].

Ronquist, F., Teslenko, M., van der Mark, P., Ayres, D.L., Darling, A., Hohna, S., et al. 2012. MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. Syst Biol 61(3): 539-542. doi: 10.1093/sysbio/sys029.

Rutherford, K., Parkhill, J., Crook, J., Horsnell, T., Rice, P., Rajandream, M.A., et al. 2000. Artemis: sequence visualization and annotation. Bioinformatics 16(10): 944-945. doi: 10.1093/bioinformatics/16.10.944.

Sambrook, J., and Russell, D.W. 2001. Molecular Cloning: A Laboratory Manual. Cold Spring Harbor Laboratory Press.

Sevillano, L., Diaz, M., and Santamaria, R. 2013. Stable expression plasmids for *Streptomyces* based on a toxin-antitoxin system. Microb Cell Fact 12(1): 39. doi: 10.1186/1475-2859-12-39.

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

Song, J.Y., Jeong, H., Yu, D.S., Fischbach, M.A., Park, H.S., Kim, J.J., et al. 2010. Draft genome sequence of *Streptomyces clavuligerus* NRRL 3585, a producer of diverse secondary metabolites. J Bacteriol 192(23): 6317-6318. doi: 10.1128/JB.00859-10.

Sun, J., Kelemen, G.H., Fernandez-Abalos, J.M., and Bibb, M.J. 1999. Green fluorescent protein as a reporter for spatial and temporal gene expression in *Streptomyces coelicolor* A3(2). Microbiology 145 (Pt 9): 2221-2227.

Tahlan, K., Anders, C., Wong, A., Mosher, R.H., Beatty, P.H., Brumlik, M.J., et al. 2007. 5S clavam biosynthetic genes are located in both the clavam and paralog gene clusters in *Streptomyces clavuligerus.* Chem Biol 14(2): 131-142. doi: 10.1016/j.chembiol.2006.11.012.

Tamura, K., Stecher, G., Peterson, D., Filipski, A., and Kumar, S. 2013. MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. Mol Biol Evol 30(12): 2725-2729. doi: 10.1093/molbev/mst197.

Walter, J.D., Hunter, M., Cobb, M., Traeger, G., and Spiegel, P.C. 2012. Thiostrepton inhibits stable 70S ribosome binding and ribosome-dependent GTPase activation of elongation factor G and elongation factor 4. Nucleic Acids Res 40(1): 360-370. doi: 10.1093/nar/gkr623.

Wright, G.D. 2007. The antibiotic resistome: the nexus of chemical and genetic diversity. Nat Rev Microbiol 5(3): 175-186. doi: 10.1038/nrmicro1614.

Wright, G.D. 2012. Antibiotics: a new hope. Chem Biol 19(1): 3-10. doi: 10.1016/j.chembiol.2011.10.019.

Zhou, X., Deng, Z., Hopwood, D.A., and Kieser, T. 1994. *Streptomyces lividans* 66 contains a gene for phage resistance which is similar to the phage lambda ea59 endonuclease gene. Mol Microbiol 12(5): 789-797. doi: 10.1111/j.1365-2958.1994.tb01065.x.

Zotchev, S.B., Schrempf, H., and Hutchinson, C.R. 1995. Identification of a methyl-specific restriction system mediated by a conjugative element from *Streptomyces bambergiensis.* J Bacteriol 177(16): 4809-4812.

A

S. clavuligerus

Predicted Product	locus
Terpene	*sclav_p0067*
NRPS-Terpene	*sclav_p0290*
Terpene	*sclav_p0497*
Terpene	*sclav_p0564*
Putative	*sclav_p0610*
Terpene-lantipeptide	*sclav_p0640*
Putative	*sclav_p0676*
Terpene	*sclav_p0766*
Butyrolactone	*sclav_p0808*
Putative	*sclav_p0935*
NRPS	*sclav_p1005*
β-lactam	*sclav_p1075*
Bacteriocin	*sclav_p1116*
Terpene	*sclav_p1177*
PKS	*sclav_p1233*
t1PKS	*sclav_p1272*
t1PKS-NRPS	*sclav_p1369*
NRPS	*sclav_p1427*
Putative	*sclav_p1474*
Putative	*sclav_p1586*

B

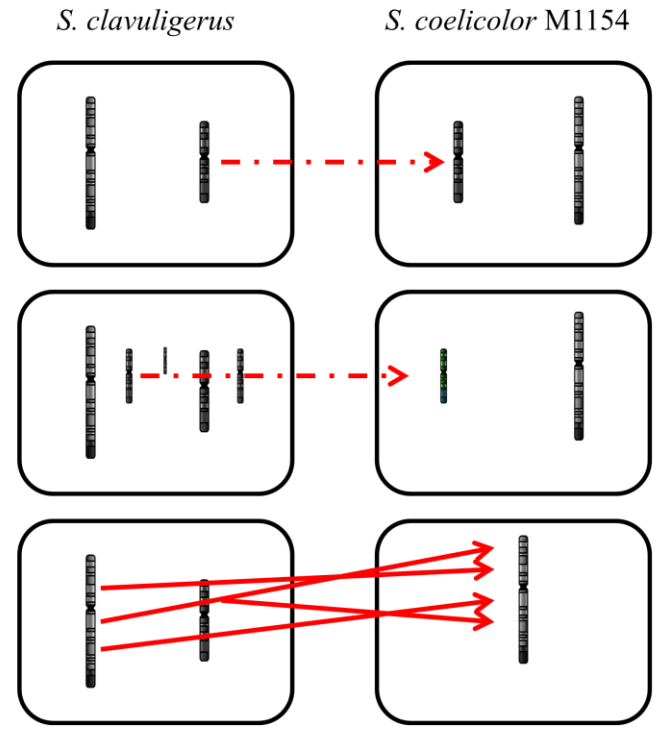


Figure 2.01: (A) pSCL4 from *S. clavuligerus* harbors upwards of 20 secondary metabolite gene clusters with incredible diversity. Product prediction and loci identification was performed using AntiSMASH2.0 (Blin et al. 2013). **(B)** During protoplast fusion there are three possible outcomes. **(i)** pSCL4 is mobile and can move between different species; **(ii)** pSCL4 is not mobile, one or more of the other previously described plasmids from *S. clavuligerus* are mobile; **(iii)** The plasmids are not mobile and recombination occurs between the *S. clavuligerus* and the *S. coelicolor* M1154 chromosomes leading to chimeras.

Table 2.01: Antibiotics used throughout this study with stock and experimental concentrations.

Antibiotic	Stock Concentration (mg/ml)	Experimental concentration (µg/ml)		
		S. clavuligerus	*S. coelicolor*	*E. coli*
Apramycin	50	50	50	50
Chloramphenicol	34	25	25	25
Kanamycin	50	50	200	50
Nalidixic Acid	50	25	25	25
Thiostrepton	50	10 (liquid) and 60 (solid media)		NA[1]
Rifampin	50	NA	10	NA
Streptomycin	100	NA	50	NA
Ampicillin	100	NA	NA	100

[1] NA = Not applicable, antibiotic was not used with this organism

Table 2.02: Plasmids and strains used during this study.

Name	Relative genotype and comments	Reference
pIJ8660	Integrative plasmid containing *egfp,* and *aac(3)IV*	(Sun et al. 1999)
pUZ8002	Helper plasmid for conjugation, *neo*	(Paget et al. 1999)
E. coli		
ET12567	Non-methylating *E. coli* strain for conjugation, cam^r , tet^r	(MacNeil et al. 1992)
DH5α	General lab strain for the introduction of plasmid DNA	(Bethesda Research Laboratories 1986)
S. clavuligerus		
NRRL3585	Wild Type (WT), also referred to as ATCC27064	Northern Regional Research Laboratory, Peoria, Ill.
cvm6p	Δ*cvm6p::neo*	(Tahlan et al. 2007)
bls1	Δ*bls1::tsr*	(Jensen et al. 2000)
S. coelicolor		
M1146	Δ*act*, Δ*red*, Δ*cpk*, Δ*cda*	(Gomez-Escribano and Bibb 2011)
M1146/pIJ8660	Δ*act*, Δ*red*, Δ*cpk*, Δ*cda,* pIJ8660	This study
M1152	Δ*act*, Δ*red*, Δ*cpk*, Δ*cda, rpo*B*[C1298T]*	(Gomez-Escribano and Bibb 2011)
M1152/pIJ8660	Δ*act*, Δ*red*, Δ*cpk*, Δ*cda, rpo*B*[C1298T],* pIJ8660	This study
M1154	Δ*act*, Δ*red*, Δ*cpk*, Δ*cda, rpo*B*[C1298T]*, *rpsL[A262G]*	(Gomez-Escribano and Bibb 2011)
M1154/pIJ8660	Δ*act*, Δ*red*, Δ*cpk*, Δ*cda, rpo*B*[C1298T]*, *rpsL[A262G],* pIJ8660	This study

Table 2.03: Oligonucleotides used during this study.

Primer Name	Sequence (5'-3')	Purpose	Tm(^{o}C)
L-TSRF	ATGACTGAGTTGGACACCAT	PCR amplify *tsr* resistance gene	59
L-TSRR	TTATCGGTTGGCCGCGAGAT	PCR amplify *tsr* resistance gene	59
L-SCP1F	TGGGGCAACTGTCTCTTGTA	Confirmation of *S. clavuligerus* pSCL4	59
L-SCP1R	AAGACCCAGACTTCACAGCG	Confirmation of *S. clavuligerus* pSCL4	59
L-SCP2F	GAATACACATCCCGGGTCCT	Confirmation of *S. clavuligerus* pSCL4	59
L-SCP2R	TATACGGCGAGCATCACGAT	Confirmation of *S. clavuligerus* pSCL4	59
L-SCP3F	GTAGACGACGACTGAGCAGA	Confirmation of *S. clavuligerus* pSCL4	59
L-SCP3R	AAGTACGAGTATCCCGACCC	Confirmation of *S. clavuligerus* pSCL4	59
L-SCP4F	TGCGCCTGTAAAATCAAGGG	Confirmation of *S. clavuligerus* pSCL4	59
L-SCP4R	TTGTTGGTGTAGTCGACGGA	Confirmation of *S. clavuligerus* pSCL4	59
L-SCP5F	GGAGCGGAGAAGAAGGAAGA	Confirmation of *S. clavuligerus* pSCL4	59
L-SCP5R	AACAGGAGTGGGACTTCGAC	Confirmation of *S. clavuligerus* pSCL4	59
L-SCP6F	CCGAGGGATGAAGTGTGACT	Confirmation of *S. clavuligerus* pSCL4	59
L-SCP6R	GGTCTTTCCGATCATGCTGG	Confirmation of *S. clavuligerus* pSCL4	59
L-SCP7F	GTGTCCTTCTCCTCCTCGAC	Confirmation of *S. clavuligerus* pSCL4	59
L-SCP7R	CACTGTGACCGGGGAGTATG	Confirmation of *S. clavuligerus* pSCL4	59
L-SCP8F	CTGGGACTCGTGATGATCCA	Confirmation of *S. clavuligerus* pSCL4	59
L-SCP8R	CTTGAGGATGCTGGAGTGGT	Confirmation of *S. clavuligerus* pSCL4	59
L-SCP9F	GAGTTCATTCCGCCATGTCC	Confirmation of *S. clavuligerus* pSCL4	59
L-SCP9R	CCCACGTATATCCCGGTGTT	Confirmation of *S. clavuligerus* pSCL4	59
L-SCP10F	CTGGATCACGGGCTTCTTG	Confirmation of *S. clavuligerus* pSCL4	59
L-SCP10R	TCTCCGGTGTTCTCGTACTG	Confirmation of *S. clavuligerus* pSCL4	59
P1 *rpoB S. coelicolor* F	CCGAGTTCACCAACAACGAGACC	PCR amplify mutations of the *rpoB* gene from *S. coelicolor*	59

P2 *rpoB S. coelicolor R*	CGATGACGAAGCGGTCCTCC	PCR amplify mutations of the *rpoB* gene from *S. coelicolor*	58
P3 *rpoB S. coelicolor F (SEQ)*	GGCCGCTACAAGGTGAACAAGAAG	PCR amplify mutations of the *rpoB* gene from *S. coelicolor* for sequencing	59
rpsL S. coelicolor F	ATTCGGCACACAGAAAC	PCR amplify mutations of the *rpsL* gene from *S. coelicolor*	45
rpsL S. coelicolor R	AGAGGAGAACCGTAGAC	PCR amplify mutations of the *rpsL* gene from *S. coelicolor*	47
L-SCOAGF	CAGGCGGGAATCGAAGA	*S. coelicolor dagA* gene	61
L-SCOAGR	CGTTCCGTGAGGTGCTG	*S. coelicolor dagA* gene	61
L-SCACVF	GACGTGACCATGCGATGTAC	*S. clavuligerus pcbAB* gene	59
L-SCACVR	TACGGGTCGATGAGGAACAG	*S. clavuligerus pcbAB* gene	59

Table 2.04: The SM gene clusters located on pSCAT are as diverse as those found on pSCL4.

S. cattleya		*S. clavuligerus*	
Predicted Product	**Locus[1]**	**Predicted Product**	**Locus[1]**
NRPS	*scat_p0038*	Terpene	*sclav_p0067*
Terpene	*scat_p0077*	NRPS-Terpene	*sclav_p0290*
NRPS	*scat_p0179*	Terpene	*sclav_p0497*
Terpene	*scat_p0222*	Terpene	*sclav_p0564*
t1PKS[2]	*scat_p0312*	Putative	*sclav_p0610*
Putative	*scat_p0420*	Terpene-lantipeptide	*sclav_p0640*
Terpene-NRPS	*scat_p0615*	Putative	*sclav_p0676*
Putative	*scat_p0755*	Terpene	*sclav_p0766*
β-lactam	*scat_p0820*	Butyrolactone	*sclav_p0808*
Terpene	*scat_p0891*	Putative	*sclav_p0935*
Putative	*scat_p0927*	NRPS	*sclav_p1005*
Butyrolactone	*scat_p0948*	β-lactam	*sclav_p1075*
Putative	*scat_p0976*	Bacteriocin	*sclav_p1116*
Putative	*scat_p1107*	Terpene	*sclav_p1177*
NRPS	*scat_p1181*	PKS	*sclav_p1233*
Siderophore	*scat_p1214*	t1PKS	*sclav_p1272*
Putative	*scat_p1258*	t1PKS-NRPS	*sclav_p1369*
Putative	*scat_p1307*	NRPS	*sclav_p1427*
Putative	*scat_p1367*	Putative	*sclav_p1474*
Putative	*scat_p1428*	Putative	*sclav_p1586*
t1PKS	*scat_p1475*		
Putative	*scat_p1523*		
Putative	*scat_p1589*		
Putative	*scat_p1629*		
t1PKS	*scat_p1670*		

[1] Approximate locus number, relative to complete gene cluster
[2] t1PKS = type 1 polyketide synthase

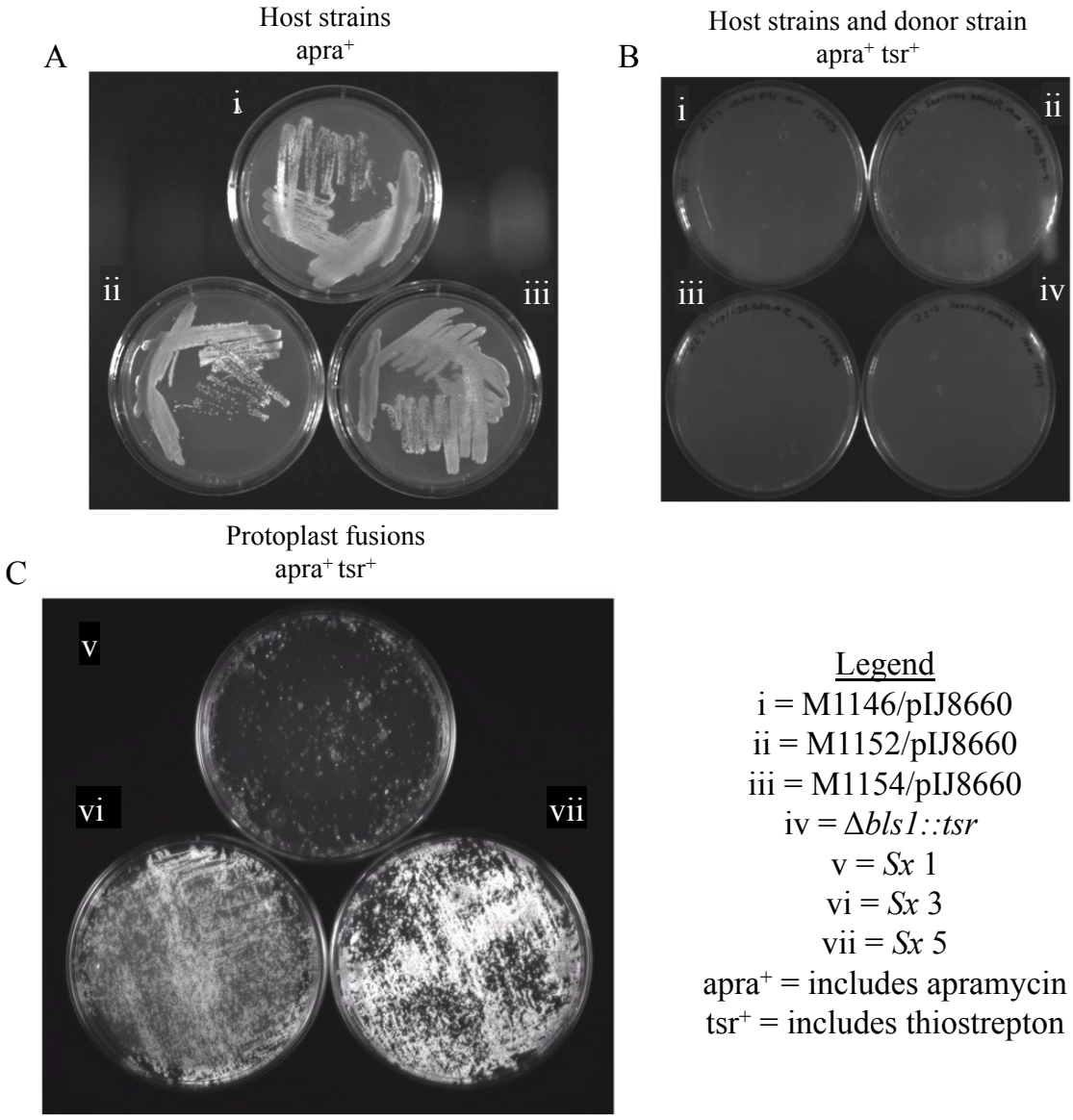


Figure 2.02: Thiostrepton and apramycin resistance profiles of host and donor protoplast regenerations and protoplast fusions. **(A)** Host strains are resistant to apramycin. Protoplasts of *S. coelicolor* M1146/pIJ8660 **(i)**, M1152/pIJ8660 **(ii)**, and M1154/pIJ8660 **(iii)**, grown on R2S with apramycin 25 μg/ml for 7 days at 28°C. **(B)** Host and donor strains are not resistant to apramycin and thiostrepton. Protoplasts of *S. coelicolor* recovered on R2S with 60 μg/ml thiostrepton and 25 μg/ml apramycin for 7 days at 28°C. **(i)** M1146/pIJ8660; **(ii)**, M1152/pIJ8660; **(iii)**, M1154/pIJ8660; **(iv)**, and *S. clavuligerus* Δ*bls1::tsr*. **(C)** *Sx* strains obtained following protoplast fusion are resistant to apramycin 25 μg/ml and thiostrepton 60 μg/ml when grown on R2S for 7 days at 28°C. Ten colonies were recovered after protoplast fusion, however, only three were able to survive reselection: **(v)** *Sx* 1; **(vi)** *Sx* 3; **(vii)** *Sx* 5.

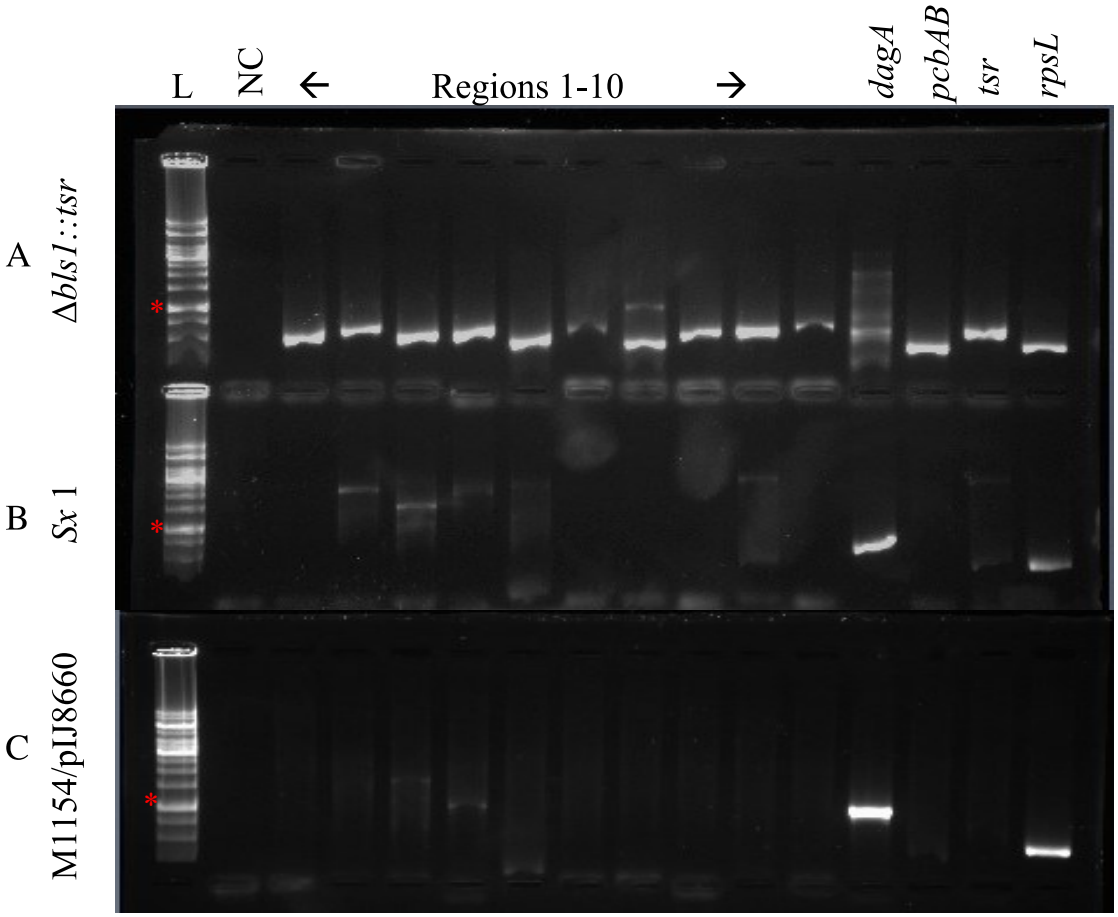


Figure 2.03: PCR screening for pSCL4 from *S. clavuligerus* NRRL 3585 indicated that the megaplasmid did not mobilize into *S. coelicolor* M1154/pIJ8660. L, GeneRuler 1kb ladder; NC, negative control; Regions 1-10 represent targeted regions based off *S. clavuligerus* NRRL 3585 pSCL4; *dagA,* agarase gene; *pcbAB,* ACVS gene; *tsr,* thiostrepton resistance cassette; *rpsL,* 30S ribosomal protein S12 gene. 1kbp marker band identified by *. Chromosomal DNA isolated from **(A)** *S. clavuligerus Δbls1::tsr,* **(B)** *Sx* 1 and **(C)** *S. coelicolor* M1154/pIJ8660 was used as template DNA for PCR analysis, respectively.

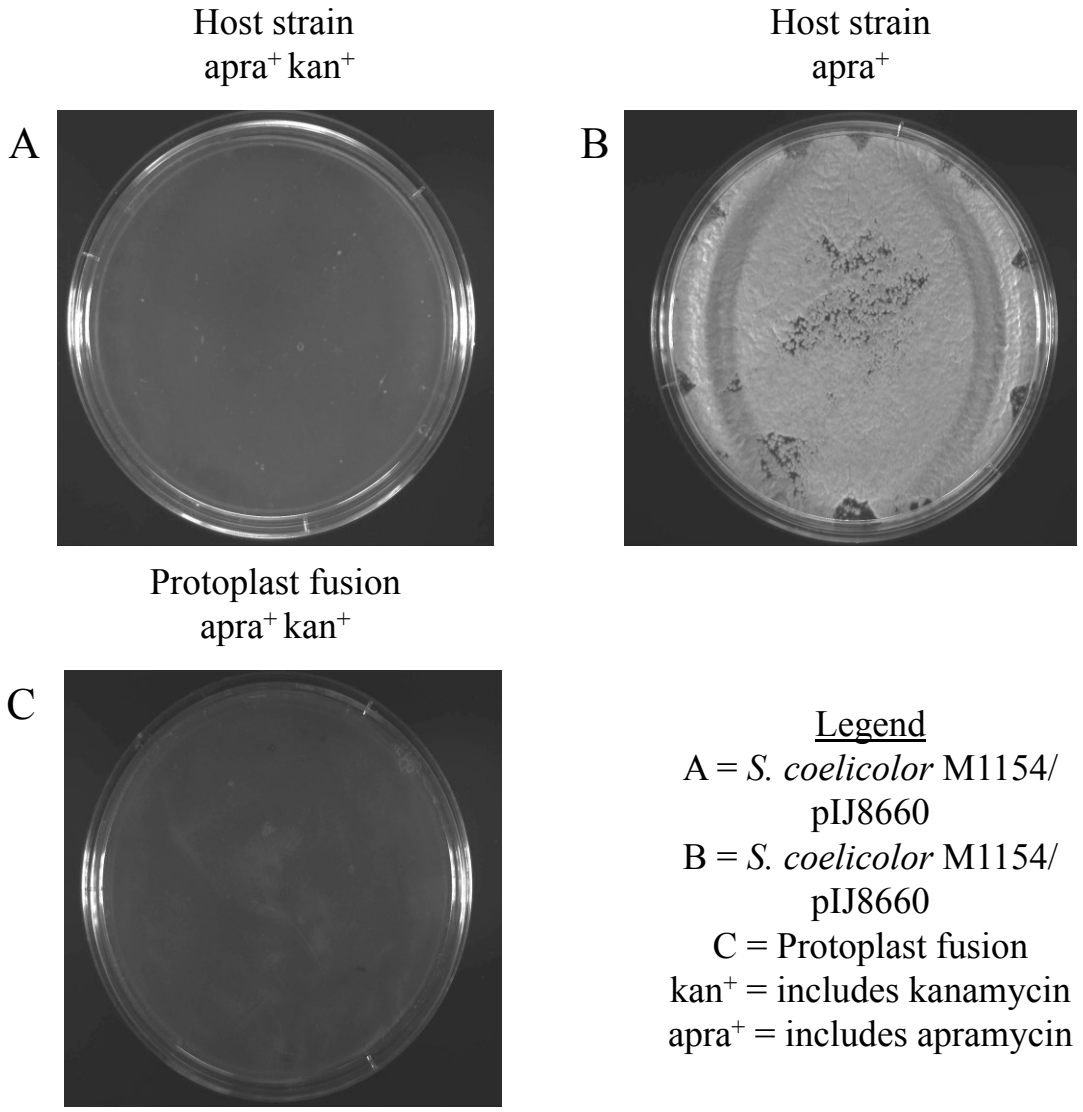


Figure 2.04: Kanamycin and apramycin resistance profiles of host and donor protoplast regenerations and protoplast fusions. **(A)** Host strain (*S. coelicolor* M1154/pIJ8660) is not resistant to apramycin and kanamycin. **(B)** Host strain (*S. coelicolor* M1154/pIJ8660) is resistant to apramycin. **(C)** Protoplast fusions are not resistant to apramycin and kanamycin. All protoplasts were recovered on R2S for 7 days at 28^{o}C.

Chapter 3

The large-scale purification of the δ-(L-α-aminoadipyl)-L-cysteinyl-D-valine synthetase from *Streptomyces clavuligerus*

Co-authorship statement

Project was designed and developed by Dr. Kapil Tahlan (Department of Biology, Memorial University of Newfoundland). Experiments were performed by Marcus Moore (Department of Biology, Memorial University of Newfoundland) and Dr. Kapil Tahlan. Reagents/materials/analysis tools were contributed by Dr. Kapil Tahlan. Analysis was done by Marcus Moore, and Dr. Kapil Tahlan. Manuscript was written by Marcus Moore and Dr. Kapil Tahlan.

3.01 Introduction

Mega-enzymes, such as nonribosomal peptide synthetases (Strieker et al. 2010) (NRPS) and the modular polyketide synthases (Shen 2003) (PKS) or fusions of the two (NRPS/PKS) (Weissman and Muller 2008), are mega-enzyme assembly lines responsible for numerous and diverse bioactive natural products, many of which have applications in medicine and agriculture (Whicher et al. 2014). These enzymes are often scaffolds that offer numerous opportunities for synthetic biology and bioengineering (Kittendorf and Sherman 2006; Ongley et al. 2013; Walsh 2002). However, the size and complexity of NRPSes and PKSes make it difficult to express and purify functional enzymes from their native or from heterologous hosts in large quantities for detailed studies (Ongley et al. 2013).

The δ-(L-α-aminoadipyl)-L-cysteinyl-D-valine (ACV) synthetase (ACVS) is an important NRPS involved in the formation of the tripeptide backbone of all penicillin and cephalosporin family antibiotics (Jensen et al. 1988). The enzyme is found in all fungi and bacteria that produce the respective metabolites including *S. clavuligerus,* which produces cephamycin C. ACVS condenses L-aminoadipic acid, L-cysteine, and L-valine using ATP hydrolysis, and the tripeptide is then processed by other enzymes from the pathways leading to the respective antibiotics (Hollander et al. 1984; Jensen et al. 1990; Schwecke et al. 1992). Like other NRPSes, isolation of large quantities of ACVS from a native host has proved difficult.

Described here is a novel tool for the expression of ACVS, and potentially other NRPSes and PKSes, in the natural host and from the native genomic locus

(Figure 3.01). The functional basis for this system is the directional knockin of an inducible promoter (*tipA*p) (Murakami et al. 1989) or a constitutive promoter (*ermE*p*) (Bibb et al. 1994; Schmitt-John and Engels 1992) upstream of the ACVS gene (*sclav_4200* or *pcbAB*) and a NRPS/PKS hybrid (*sclav_4466*) in *S. clavuligerus*. SCLAV_4466 is a cryptic megasynthetase with a predicted molecular weight of ~523 kDa. Similar to ACVS, SCLAV_4466 is encoded within the chromosome in a large gene cluster. However, while little is known about the terminal product, SCLAV_4466 is predicted to combine a malonyl with the amino acids, serine and proline.

Expression from *tipA*p is induced by the exogenous addition of thiostrepton (an antibiotic), whereas *ermE*p* has been shown to be a strong, constitutive promoter that works well to drive the expression of genes in *Streptomyces* spp. Furthermore, this system has been optimized to include sequences for purification and isolation of intact enzyme using nickel affinity purification employing a cleavable 8× histidine tag (His-tag). The long-term goal of this work is to develop tools for large-scale purification of ACVS and other mega-proteins for collaborative structural studies using cryo-electron microscopy and x-ray crystallography. In addition, the tagged protein can also be used for protein localization and interaction studies.

3.02 Materials and Methods

3.02.01 Culture conditions and general procedures

All *E. coli* and *Streptomyces* culturing, manipulations and procedures were performed as previously described in chapter 2. In the current chapter, conjugations were carried out using *S. clavuligerus* spores instead of mycelia as described in chapter 2. A list of all the plasmids and strains used throughout this study can be found in table 3.01. All oligonucleotide primers used in the current study are listed in Table 3.02. All restriction endonucleases were purchased from New England BioLabs or Thermo Fisher Scientific. All mycelia were harvested for growth determination and fluorescence studies by centrifugation at 12000 rpm for 2 minutes using a microcentrifuge.

3.02.02 Design of synthetic promoter constructs

Two plasmids were constructed based on the thiostrepton inducible promoter *tipA*p (Murakami et al. 1989) and the strong, constitutive promoter *ermE*p* (Bibb et al. 1994; Schmitt-John and Engels 1992) (Figure 3.02). Geneious software (Biomatters Ltd., Auckland, New Zealand) was used to design the nucleotide sequence with the incorporation of the required properties and enzyme restriction sites. Synthetic promoter constructs were synthesized by Integrated DNA Technologies (Coralville, Iowa, USA) to our specifications in proprietary plasmids (pIDTSmart-KAN-*tipA*p and pIDTSmart-KAN-*ermE*p*).

3.02.03 Cloning the synthetic promoter constructs

The synthetic promoters were cloned into pIJ8668 and pIJ8660 (Sun et al. 1999) using EcoRV and NdeI restriction sites engineered in the constructs. pIJ8668, pIJ8660, pIDTSmart-KAN-*tipA*p and pIDTSmart-KAN-*ermE*p* were digested for 18 hours with EcoRV-HF® (R3195S), followed by heat inactivation (65^{o}C for 20 minutes). Buffers were exchanged using the EZ-10 Spin Column PCR DNA kit (BS363, Bio Basics Inc, Markham, Ontario) and plasmids were digested with NdeI. Phosphate groups were removed from pIJ8660 and pIJ8668 using FastAP thermosensitive alkaline phosphatase (EF0651, Thermo Fisher Scientific, Waltham, MA, USA). Digested pIDTSmart-KAN-*tipA*p, and pIDTSmart-KAN-*ermE*p* were separated on 1.5% agarose gels and the DNA fragments containing the promoter constructs were extracted using the Wizard® SV and PCR clean-up system (A9281, Promega, Madison, WI, USA). The synthetic promoters were ligated into pIJ8660 and pIJ8668 using the Fermentas Rapid DNA Ligation Kit (K1422, Thermo Scientific, Waltham, MA, USA) resulting in, pIJ8660-*tipA*p, pIJ8660-*ermE*p* and pIJ8668-*ermE*p*. *tipAp* was cloned between EcoRV and NotI in pIJ8668 resulting in, pIJ8668-*tipA*p. Plasmids were chemically transformed into DH5α and confirmed by restriction digest with EcoRV/NotI.

A 1.5kbp DNA fragment from the start of each gene of interest (*pcbAB*or *sclav_4200* GI:294327632 and *sclav_4466* GI:294327894) was ligated downstream of the synthetic promoter constructs using NdeI and NotI restriction sites engineered into the PCR primers. Fragments were PCR amplified using oligonucleotide primers and *S. clavuligerus* chromosomal DNA as template (Alexander et al. 2007) with Phusion High-Fidelity DNA Polymerase (F-5302, Fisher Scientific-Canada, Ottawa, Ontario) using the

following thermocycling conditions: initial denaturation 98°C for 30 seconds, 10× (denaturation 98°C 10 seconds, annealing 55°C 30 seconds, extension 72°C 1 minute), 30× (denaturation 98°C 10 seconds, annealing 63°C 30 seconds, extension 72°C 1 minute), final extension 72°C 5 minutes. PCR amplicons were gel-purified from 0.8% agarose gels. "A" overhangs were added to the end of the PCR products with Taq DNA polymerase (EP0401, Thermo Fisher Scientific, Waltham, MA, USA) using the following thermocycling conditions: 72.0°C for 10 minutes. Fragments were ligated into pGEM-T® easy vector (A1360, Promega, Fitchburg, WI, USA) resulting in two new plasmids (pGEM-4200 and pGEM-4466), and transformed into DH5α. Successful transformations were selected based on blue/white selection on LB agar plates containing ampicillin overlaid with 100 µl of 100 mM IPTG and 20 µl of 50 mg/ml X-Gal. The pGEM-4200 and pGEM-4466 insert ligations were confirmed by Sanger sequencing at The Centre for Applied Genomics (Toronto, Ontario) using M13 primers. The respective DNA fragments were then cloned into pIJ8668-*tipA*p and pIJ8668-*ermE*p* using the engineered NdeI and NotI restriction sites, resulting in pIJ8668-*tipA*p-4466, pIJ8668-*tipA*p-4200, pIJ8668-*ermE*p*-4200, pIJ8668-*ermE*p*-4466, pIJ8660-*ermE*p*, and pIJ8660-*tipA*p (Figure 3.03). Each plasmid was purified and transformed into chemically competent ET12567 (MacNeil et al. 1992) containing pUZ8002 (Paget et al. 1999) and conjugated into *S. clavuligerus* Δ*ccaR::tsr, S. clavuligerus* NRRL3585*, S. clavuligerus* Δ*bls1::tsr,* and *S. clavuligerus* Δ*cmcT::neo*, as previously described. The integration of the plasmids was confirmed by genomic DNA PCR and the products were sequenced.

3.02.04 Determination of protein concentrations using the Bradford assay

To verify that the synthetic promoters were active, *S. clavuligerus* Δ*ccaR*::*tsr*/pIJ8660-*ermE*p* and *S. clavuligerus* Δ*ccaR*::*tsr*/pIJ8660 were grown in 100 ml TSBS for 144 hours and 1 ml samples were drawn at 24-hour intervals after the first sample, which was taken at 12 hours. Cells were pelleted by centrifugation and frozen at -80^{o}C for future analysis.

Bradford assays were carried out to quantify protein concentrations using a kit according to the manufacturer's recommendations (EMD Milliore, Merck KGaA, Darmstadt, Germany). For measuring relative growth, mycelial pellets from above were suspended in 100 µl of TE and heated to 100^{o}C for 10 minutes. Twenty microliters of each sample were mixed with 80 µl 0.15M NaCl to which 900 µl of Bradford reagent was added and incubated at room temperature for 5 minutes. The absorbance of each sample was measured at 595 nm and the concentration of protein was determined using a bovine serum albumin (BSA) standard curve (Table 3.03)**.**

3.02.05 eGFP fluorescence assay

S. clavuligerus Δ*ccaR*::*tsr*/pIJ8660 and *S. clavuligerus* Δ*ccaR*::*tsr*/pIJ8660-*ermE*p* (or *tipA*p based constructs) were inoculated into 10 ml of TSBS and were grown for 48 hours. Growth of the strains was quantified as described previously (Brana et al. 1985) by mixing 0.5 ml of culture with 0.5 ml of 2.5N HCl and diluting with 3.0 ml of diH_2O. The suspension was homogenized with ultrasonic treatment for 30 seconds and the absorbance was measured at 600nm. Samples were diluted until the OD_{600} was lower than 0.6; under these conditions, an OD_{600} of 2.15 is equivalent to 1.0 mg/ml dry cell

weight (DCW). TSBS seed culture equivalent to 0.59 mg of DCW was then transferred into 25 ml of fresh TSBS, and 1.5 ml samples were drawn at 12, 24, 48, 72 and 96 hours. Mycelia were washed twice with 200 µl phosphate buffered saline (PBS) (20 mM sodium phosphate and 300 mM sodium chloride, pH 7.4) and re-suspended in 200 µl PBS, which was pipetted into black flat bottom polystyrene 96 well plates (137101, Nalgene NUNC, Rochester, NY, USA). Fluorescence due to the expression of eGFP was measured using a Synergy H1MG plate reader (BTH1MG, Biotek, Winooski, VT, USA) with excitation at 488 nm, emission at 510 nm, gain (manual) at 50 and probe height at 7.5 mm. For high-resolution fluorescent microscopic images, 10 µl of mycelia were placed onto a microscope slide followed by a cover slip that was sealed around the edges with clear nail polish. Fluorescence was viewed using a Zeiss Axio Imager A1 and images were taken with an AxioCam HRc camera.

3.02.06 Protein extraction by sonication

Mycelia were washed in 10.3% w/v sucrose and suspended immediately in ½ pellet volume of equilibration buffer (0.1M MOPS pH7.5, 0.05M KCl, 1 mM DTT, 20% v/v glycerol, and 10 mM Imidazole). Mycelia were sonicated using a Sonicator Q125 (Sonicator Q125, Qsonica, Newtown, CT, USA) with a 3 mm probe (part number 4422) at 50% power for 3 minutes, 15 seconds on and 15 seconds off cycles. Crude soluble extract was clarified by centrifugation for 10 minutes at 4500 rpm at 4^{o}C and the protein concentration in the supernatant was determined using the Bradford protein assay as described above.

3.02.07 Dot-blot detection of ACVS protein

Soluble protein (6µl) was dotted onto a methanol - soaked Millipore Immobilon®-P PVDF Membrane (IPFL00010, Millipore, Darmstadt, Germany). The membrane was allowed to dry, after which it was blocked in TBS-T (20 mM Tris-HCl, 150 mM NaCl, pH7.5 and 0.05% v/v Tween20) containing 5% w/v nonfat milk (Instant Skim Milk, Carnation) overnight at 4^{o}C. ACVS antibodies (Tahlan, K. unpublished) were diluted 1:5000 in TBS-T containing 5% w/v nonfat milk and were incubated with the membrane for one hour with gentle rocking, after which the membrane was washed three times in TBS-T for 10 minutes with gentle rocking. Secondary antibody (SA1-200, Thermo Fisher Scientific, Waltham, MA, USA) was diluted 1:1000 in TBS-T and was incubated for one hour with gentle rocking, after which the membrane was washed three times in TBS-T for 5 minutes each. The membrane was developed with ECL Western Blot Substrate (W1001, Promega, Fitchburg, WI, USA) and imaged on a GE ImageQuant LAS 4000 Digital Imaging System (28-9558-10, GE Healthcare, Baie d'Urfe, Quebec).

3.02.08 Protein purification using nickel affinity resin

Proteins of interest were purified using Ni-NTA resin (88221, Thermo Fisher Scientific, Waltham, MA, USA) and a batch centrifugation protocol according to the manufacturer's recommendations with some changes. Mycelia (23.6 mg DCW) from a 25 ml starter culture were inoculated into 1L of TSBS, and the culture was grown for 48 hours. The mycelia were harvested in 50 ml falcon tubes and were resuspended in one pellet volume of MDKG equilibration buffer (0.1M MOPS pH7.5, 0.05M KCl, 1 mM DTT, 20% v/v glycerol, and 10 mM Imidazole). Ni-NTA resin (500µl) was prepared by

pelleting the resin by centrifugation at 700 × *g* for 2 minutes, and then washing the resin in two resin-bed volumes of equilibration buffer. The soluble protein fraction obtained after sonication and centrifugation was then added to the resin and incubated at 4^{o}C for 30 minutes with manual inversion every 5 to 10 minutes. The resin was washed twice in 2 resin volumes of wash buffer (0.1M MOPS pH7.5, 0.05M KCl, 1 mM DTT, 20% v/v glycerol, and 35 mM Imidazole) and the supernatant was collected after each wash. His-tag proteins were eluted twice with 1 resin volume of MDKG elution buffer (0.1M MOPS pH7.5, 0.05M KCl, 1 mM DTT, 20% v/v glycerol, and 250 mM Imidazole). For figure 3.08, protein was isolated in PBS with 10 mM imidazole (pH 7.4), washed in PBS with 25 mM imidazole, and eluted in PBS with 250 mM imidazole. Later on, MDKG was used in other protein preparations to stabilize ACVS, as previously described (Jensen et al. 1990). The protein preparation was concentrated using Microcon® centrifugal filters (MRCPRTO10, EMD Millipore, Etoicoke, Ontario) and buffer exchanged into MDKG buffer for storage.

3.02.09 Protein analysis by sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE)

Protein extracts were visualized using SDS-PAGE with 10% or 5% w/v polyacrylamide gels as described previously (Sambrook and Russell 2001). Proteins from insoluble pellets were extracted by vigorously vortexing with one volume 6M urea. Samples were mixed with 6× protein loading dye (375 mM Tris-HCl pH6.8, 9% v/v SDS, 50% v/v glycerol, and 0.03% v/v bromophenol blue) and 2.5% w/v dithiothreitol (DTT) and boiled at 100^{o}C for 5 minutes prior to analysis along with PageRuler Plus protein

ladder (Thermo Scientific, Rockford, IL, USA). Gels were run at 57V at 4^{o}C and were stained in Coomassie brilliant blue or silver stained (PI24612, Thermo Scientific, Rockford, IL, USA) for imaging as previously described (Sambrook and Russell 2001).

3.03 Results and Discussion

3.03.01 Establishing an eGFP reporter system for *tipA*p and *ermE*p*

The purpose of the described work was to design a tool for the expression and purification of ACVS and other mega proteins from *S. clavuligerus*. This system was designed using two expression promoters, *tipA*p and *ermE*p*, which were cloned into the plasmid pIJ8668. A reporter system was also constructed to determine if the synthetic *ermE*p* and *tipA*p promoters used in the current study were active in *S. clavuligerus*. This was done by introducing the respective promoter constructs upstream of a promoterless *egfp* gene in the plasmid pIJ8660, which is a well - established integrative reporter system for use in *Streptomyces* (Sun et al. 1999).

*tipA*p is a thiostrepton inducible promoter (Schmitt-John and Engels 1992) and thiostrepton being an antibiotic is known to inhibit the growth of *S. clavuligerus*. Therefore, a *S. clavuligerus* Δ*ccaR*::*tsr* mutant strain (Alexander and Jensen 1998), in which the *ccaR* gene has been inactivated by the insertion of a thiostrepton resistance gene (*tsr*), was selected as a host for this system. Furthermore, *S. clavuligerus* Δ *ccaR*::*tsr* is an ideal host for downstream expression studies of ACVS as CcaR regulates the expression of the gene encoding ACVS and the Δ*ccaR*::*tsr* knockout does not produce ACVS (Alexander and Jensen 1998). CcaR is a SARP (*Streptomyces* antibiotic regulatory protein), which regulates the expression of the tricistronic operon that contains *pcbAB* (ACVS) (Alexander and Jensen 1998). Therefore, strains with insertions or deletions in *ccaR* are unable to transcribe the genes for all three proteins, and subsequently cannot produce cephamycin C. The use of this strain would avoid any potential complications

that could arise from the expression of the operon encoding ACVS, during the regulated expression of the tagged version of the protein in the knockin strain.

Previous studies showed that induction of *tipA*p in *S. coelicolor* and *S. lividans* involved thiostrepton concentrations upwards of 30 μg/ml (Ali et al. 2002; McKenzie and Nodwell 2009). Before proceeding with *tipA*p induction in *S. clavuligerus*, the maximum survivable concentration (MSC) of thiostrepton for the *S. clavuligerus* Δ*ccaR*::*tsr* strain was determined. pIJ8660-*tipA*p and pIJ8660 were introduced into *S. clavuligerus* Δ*ccaR*::*tsr,* and the resulting strains tested for resistance to thiostrepton and apramycin, as the *tsr* gene is located in *ccaR* and the *apra* gene on the pIJ8660 backbone. Results of the testing suggested that much higher quantities of apramycin and thiostrepton can be used as compared to previously published concentrations for *S. clavuligerus* (Table 3.04). As a result, 30 μg/ml was used for induction of *tipA*p in *S. clavuligerus* containing the *tsr* gene.

Next, to determine the optimal induction time point, thiostrepton was added at 0 hours (inoculation of the culture) or after 30 hours of growth in TSBS, as previously described in *S. lividans* using a different growth medium (Ali et al. 2002). Adding thiostrepton at 0 hours represents induction during early stages when the organism is retooling its physiology to begin growth under the given conditions, whereas induction at 30 hours is during exponential growth phase, when the organism is growing and reproducing rapidly. pIJ8660 was used as a negative control to subtract non eGFP related auto-fluorescence in *S. clavuligerus,* which can result in false positive results. Relative fluorescence levels were measured for both the *S. clavuligerus* Δ*ccaR*::*tsr*/pIJ8660 and Δ*ccaR*::*tsr*/pIJ8660-*tipA*p strains over the course of 96 hours to determine optimal activity of *tipA*p (Figure 3.04). Strong induction of eGFP was observed when induction was

carried out using thiostrepton at 0 hours. However, expression from *tipA*p quickly dropped by 96 hours of growth. ACVS is a large protein and may require consistent strong expression to get large amounts of protein for purification; as a consequence, *ermE*p* was next tested for consistent expression.

*ermE*p* is a strong and constitutive promoter, resulting in the continuous expression of the genes under its control. The *ermE*p* promoter fragment was cloned into pIJ8660 and was introduced into *S. clavuligerus* Δ*ccaR::tsr* as described for *tipA*p above. Relative fluorescence was measured for *S. clavuligerus* Δ*ccaR::tsr/*pIJ8660 and Δ*ccaR::tsr/*pIJ8660-*ermE*p*, as described above for *tipA*p (Figure 3.05-A). pIJ8660-*ermE*p* produced strong and consistent signal from eGFP up to 96 hours. A growth curve was constructed based on protein concentration to determine that both the *S. clavuligerus* Δ*ccaR::tsr/*pIJ8660 and Δ*ccaR::tsr/*pIJ8660-*ermE*p* cultures were at similar stages of growth (Figure 3.05-B). To further confirm promoter activity, high-resolution fluorescence microscopy images showed increased pIJ8660-*ermE*p* activity when compared to the pIJ8660 control (Figure 3.05-C and D).

Unpredictable expression from *tipA*p over 96 hours resulted in the use of *ermE*p* instead to avoid any downstream problems. It is known that *ermE*p* results in stronger protein expression as compared to *tipA*p (Schmitt-John and Engels 1992; Temuujin et al. 2011), and due to the sizes of the mega proteins that were to be expressed (~414 kDa for ACVS and ~526 kDa for SCLAV_4466), *tipA*p may not be suitable for their expression and purification from *S. clavuligerus*.

3.03.02 Confirming construct crossover at *pcbAB* and *sclav_4466*

The plasmid pIJ8668 can freely replicate in *E. coli*, but cannot do so in *Streptomyces* due to the lack of an appropriate origin of replication. In addition, pIJ8668 is non-integrative (does not contain phage integration sites or other mechanism) and has to undergo homologous recombination for insertion into the chromosome using sequences cloned into it. Along with the *ermE*p* promoter, 1.5kbp of the 5′ regions of each respective gene (*pcbAB* and *sclav_4466*) was also cloned into pIJ8668, yielding an integrative plasmid that would then recombine into the host genome at the specified gene locus. This would result in a tagged copy of each gene regulated by *ermE*p* and a second disrupted copy. In order to confirm the correct arrangement of the integrated plasmids, oligonucleotide primers were designed to amplify a ~1.8kbp region starting from the *ermE*p* promoter to ~100bp downstream of the 1.5kbp cloned fragments from the *S. clavuligerus* chromosome. As shown in Figure 3.06, the expected ~1.8kbp PCR products were obtained following amplification of genomic DNA from both the *pcbAB* and *sclav_4466* knockin strains. In addition, the DNA sequences of the PCR products were determined to confirm the insertions.

3.03.03 Preliminary purification of ACVS

Large proteins can sometimes be difficult to resolve by SDS-PAGE for staining and detection. In order to quickly detect increased ACVS expression in the *S. clavuligerus* knockin strain, a simple dot-blot assay was used. Dot-blots are similar to western blots, with the exception that the protein sample is directly applied to a polyvinylidene fluoride membrane. Dot-blots fail to differentiate between individual

proteins in a crude extract; however, they can indicate the presence of a protein based on antibody binding. Soluble proteins from *S. clavuligerus* Δ*ccaR::tsr/*pIJ8668-*ermE*p*-4200 and *S. clavuligerus* WT were blotted and then probed with ACVS antibodies. As shown in Figure 3.07, a stronger ACVS signal was detected in the *ermE*p* knockin strain relative to WT *S. clavuligerus* after 24 to 72 hours of growth, confirming that ACVS production is higher in the *ermE*p* knockin strain than in the WT strain.

Alexander et al. (2000) were able to resolve low levels of ACVS from a crude *S. clavuligerus* protein extract using SDS-PAGE with a 10% w/v polyacrylamide gel, and this was used as a starting point for preliminary analysis of ACVS expression under control of *ermE*p* in the *S. clavuligerus* knockin strain. In order to determine the presence of ACVS in *S. clavuligerus* Δ*ccaR::tsr/*pIJ8668-*ermE*p*-4200, nickel resin-based protein purification using extracts from the strain was performed to bind the His-tag that should be present on the N-terminus of ACVS. Samples were compared to extracts from the ACVS-deficient *S. clavuligerus* Δ*ccaR::tsr* strain as described by Alexander et al. (2000). As expected, soluble and insoluble protein extracts from *S. clavuligerus* Δ*ccaR::tsr* did not indicate the presence of ACVS (Figure 3.08). Protein extracts from *S. clavuligerus* Δ*ccaR::tsr*/pIJ8668-*ermE*p*-4200 and eluted fractions after nickel resin binding showed the presence of a potential protein band that was >175 kDa in size, which could potentially be ACVS (Figure 3.08). Two other smaller proteins (~62 kDa and ~29 kDa) appear to co-elute with the large protein, the latter of which has also been reported by Jensen et al. (1990). These are most likely proteins with some affinity to Ni-NTA or proteins that interact with ACVS or stable degradation products of ACVS itself.

A large amount of ACVS was not purified during the preliminary experiments described above, which could be due to the instability ACVS under the conditions used. Numerous reports indicate that both fungal and *Streptomyces* ACVS will degrade quickly once purified and that the preparations require glycerol and dithiothretol (DTT) to stabilize the protein (Jensen et al. 1990; Theilgaard et al. 1997; van der Lende et al. 2002). The initial purification was done using buffers and conditions recommended by the supplier of the Ni-NTA resin, which does not contain glycerol or DTT.

3.03.04 Step up and purification of ACVS from large cultures and modified protocol

In order to purify larger amounts of ACVS, crude extracts were harvested from a 1 L *S. clavuligerus* Δ*ccaR::tsr*/pIJ8668-*ermE*p*-4200 culture using a MDKG buffer-based system along with Ni-NTA. The imidazole present in the eluted ACVS protein was removed immediately by concentrating the protein into fresh MDKG buffer. The eluted fractions were then resolved by SDS-PAGE using a 5% w/v polyacrylamide gel and under similar conditions described by Jensen et al. (1990), leading to the detection of a protein >250 kDa in size as determined by silver staining (Figure 3.09). The predicted molecular weight of ACVS is ~414 kDa (based on the predicted amino acid sequence), which is in agreement with experimental studies using both *Streptomyces* and certain fungal ACVS (van der Lende et al. 2002; Zhang and Demain 1992b).

However, similar attempts at purifying and resolving SCLAV_4466 were not successful. SCLAV_4466 (predicted to be ~523 kDa) is substantially larger than ACVS, and may not have even migrated into the resolving gel during SDS-PAGE. Furthermore, SCLAV_4466 is a NRPS/PKS, and in *Streptomyces* PKSes have been shown to be

produced during a relatively short window (approximately after 24 hours of growth) and the 48 hour culturing time used in the current study may not favor SCLAV_4466 overexpression (Pfeifer and Khosla 2001). Cephamycin C is known to be produced in TSBS (Alexander and Jensen 1998); however, the media requirements for the SCLAV_4466 dependent pathway have not been determined, and different media may be required for the expression of other proteins/factors required for the translation and stabilization of the protein. The production of the mega protein in the producer organism probably requires changes to the ribosomes and other cellular machinery to translate and fold such large structures. Therefore, it is possible that more optimization will need to be performed before exact conditions for expressing SCLAV_4466 are identified.

A number of studies have purified both fungal and bacterial ACVS (Jensen et al. 1990; Kallow et al. 1998; Theilgaard et al. 1997; van der Lende et al. 2002; Zhang and Demain 1992a; Zhang and Demain 1992b). Purification in these studies was based on native promoters and painstaking protein purification procedures, resulting in low yields that did not allow for further study or manipulation. Presented here is a tool that can be used for the purification of ACVS for detailed downstream studies. This system may also be optimized for the expression of other NRPSes and PKSes. Recent NRPS/PKS structural studies have focused on PKS subunits (Whicher et al. 2014), and to date, no structural studies have been conducted on an intact NRPS, NRPS/PKS hybrid, or PKS purified from a native host. This work provides an important tool required to successfully elucidate the structure of ACVS, while also lending itself to numerous other applications, which are discussed in the summary.

BIBLIOGRAPHY

Alexander, D.C., Anders, C.L., Lee, L., and Jensen, S.E. 2007. *pcd* mutants of *Streptomyces clavuligerus* still produce cephamycin C. J Bacteriol 189(16): 5867-5874. doi: 10.1128/JB.00712-07.

Alexander, D.C., Brumlik, M.J., Lee, L., and Jensen, S.E. 2000. Early cephamycin biosynthetic genes are expressed from a polycistronic transcript in *Streptomyces clavuligerus*. J Bacteriol 182(2): 348-356. doi: 10.1128/JB.182.2.348-356.2000.

Alexander, D.C., and Jensen, S.E. 1998. Investigation of the *Streptomyces clavuligerus* cephamycin C gene cluster and its regulation by the CcaR protein. J Bacteriol 180(16): 4068-4079.

Ali, N., Herron, P.R., Evans, M.C., and Dyson, P.J. 2002. Osmotic regulation of the *Streptomyces lividans* thiostrepton-inducible promoter, *ptipA*. Microbiology 148(Pt 2): 381-390.

Bethesda Research Laboratories. 1986. BRL pUC host: *E. coli* DH5 α competent cells. Focus 8(2): 9.

Bibb, M.J., White, J., Ward, J.M., and Janssen, G.R. 1994. The mRNA for the 23S rRNA methylase encoded by the *ermE* gene of *Saccharopolyspora erythraea* is translated in the absence of a conventional ribosome-binding site. Mol Microbiol 14(3): 533-545. doi: 10.1111/j.1365-2958.1994.tb02187.x.

Brana, A.F., Wolfe, S., and Demain, A.L. 1985. Ammonium repression of cephalosporin production by *Streptomyces clavuligerus*. Can J Microbiol 31(8): 736-743.

Flett, F., Mersinias, V., and Smith, C.P. 1997. High efficiency intergeneric conjugal transfer of plasmid DNA from *Escherichia coli* to methyl DNA-restricting streptomycetes. FEMS Microbiol Lett 155(2): 223-229. doi: 10.1016/S0378-1097(97)00392-3.

Hollander, I.J., Shen, Y.Q., Heim, J., Demain, A.L., and Wolfe, S. 1984. A pure enzyme catalyzing penicillin biosynthesis. Science 224(4649): 610-612. doi: 10.1126/science.6546810.

Jensen, S.E., Elder, K.J., Aidoo, K.A., and Paradkar, A.S. 2000. Enzymes catalyzing the early steps of clavulanic acid biosynthesis are encoded by two sets of paralogous genes in *Streptomyces clavuligerus*. Antimicrob Agents Chemother 44(3): 720-726. doi: 10.1128/AAC.44.3.720-726.2000.

Jensen, S.E., Westlake, D.W.S., and Wolfe, S. 1988. Production of the penicillin precursor δ-(L-α-aminoadipyl)-L-cysteinyl-D-valine (ACV) by cell-free extracts from *Streptomyces clavuligerus*. FEMS Microbiol Lett 49(2): 213-218. doi: 10.1111/j.1574-6968.1988.tb02718.x.

Jensen, S.E., Wong, A., Rollins, M.J., and Westlake, D.W. 1990. Purification and partial characterization of delta-(L-alpha-aminoadipyl)-L-cysteinyl-D-valine synthetase from *Streptomyces clavuligerus*. J Bacteriol 172(12): 7269-7271.

Kallow, W., von Dohren, H., and Kleinkauf, H. 1998. Penicillin biosynthesis: energy requirement for tripeptide precursor formation by delta-(L-alpha-aminoadipyl)-L-cysteinyl-D-valine synthetase from *Acremonium chrysogenum*. Biochemistry 37(17): 5947-5952. doi: 10.1021/bi971741o.

Kittendorf, J.D., and Sherman, D.H. 2006. Developing tools for engineering hybrid polyketide synthetic pathways. Curr Opin Biotechnol 17(6): 597-605. doi: 10.1016/j.copbio.2006.09.005.

MacNeil, D.J., Gewain, K.M., Ruby, C.L., Dezeny, G., Gibbons, P.H., and MacNeil, T. 1992. Analysis of *Streptomyces avermitilis* genes required for avermectin biosynthesis utilizing a novel integration vector. Gene 111(1): 61-68. doi: 10.1016/0378-1119(92)90603-M.

McKenzie, N.L., and Nodwell, J.R. 2009. Transmembrane topology of the AbsA1 sensor kinase of *Streptomyces coelicolor*. Microbiology 155(Pt 6): 1812-1818. doi: 10.1099/mic.0.028431-0.

Murakami, T., Holt, T.G., and Thompson, C.J. 1989. Thiostrepton-induced gene expression in *Streptomyces lividans*. J Bacteriol 171(3): 1459-1466.

Ongley, S.E., Bian, X., Neilan, B.A., and Muller, R. 2013. Recent advances in the heterologous expression of microbial natural product biosynthetic pathways. Nat Prod Rep 30(8): 1121-1138. doi: 10.1039/c3np70034h.

Paget, M.S., Chamberlin, L., Atrih, A., Foster, S.J., and Buttner, M.J. 1999. Evidence that the extracytoplasmic function sigma factor sigmaE is required for normal cell wall structure in *Streptomyces coelicolor* A3(2). J Bacteriol 181(1): 204-211.

Pfeifer, B.A., and Khosla, C. 2001. Biosynthesis of polyketides in heterologous hosts. Microbiol Mol Biol Rev 65(1): 106-118. doi: 10.1128/MMBR.65.1.106-118.2001.

Sambrook, J., and Russell, D.W. 2001. Molecular Cloning: A Laboratory Manual. Cold Spring Harbor Laboratory Press.

Schmitt-John, T., and Engels, J.W. 1992. Promoter constructions for efficient secretion expression in *Streptomyces lividans*. Appl Microbiol Biotechnol 36(4): 493-498. doi: 10.1007/BF00170190.

Schwecke, T., Aharonowitz, Y., Palissa, H., von Dohren, H., Kleinkauf, H., and van Liempt, H. 1992. Enzymatic characterisation of the multifunctional enzyme delta-(L-alpha-aminoadipyl)-L-cysteinyl-D-valine synthetase from *Streptomyces clavuligerus*. Eur J Biochem 205(2): 687-694. doi: 10.1111/j.1432-1033.1992.tb16830.x.

Shen, B. 2003. Polyketide biosynthesis beyond the type I, II and III polyketide synthase paradigms. Curr Opin Chem Biol 7(2): 285-295. doi: 10.1016/S1367-5931(03)00020-6.

Strieker, M., Tanovic, A., and Marahiel, M.A. 2010. Nonribosomal peptide synthetases: structures and dynamics. Curr Opin Struct Biol 20(2): 234-240. doi: 10.1016/j.sbi.2010.01.009.

Sun, J., Kelemen, G.H., Fernandez-Abalos, J.M., and Bibb, M.J. 1999. Green fluorescent protein as a reporter for spatial and temporal gene expression in *Streptomyces coelicolor* A3(2). Microbiology 145 (Pt 9): 2221-2227.

Temuujin, U., Chi, W.J., Lee, S.Y., Chang, Y.K., and Hong, S.K. 2011. Overexpression and biochemical characterization of DagA from *Streptomyces coelicolor* A3(2): an endo-type beta-agarase producing neoagarotetraose and neoagarohexaose. Appl Microbiol Biotechnol 92(4): 749-759. doi: 10.1007/s00253-011-3347-7.

Theilgaard, H.B., Kristiansen, K.N., Henriksen, C.M., and Nielsen, J. 1997. Purification and characterization of delta-(L-alpha-aminoadipyl)-L-cysteinyl-D-valine synthetase from *Penicillium chrysogenum*. Biochem J 327 (Pt 1): 185-191.

van der Lende, T.R., van de Kamp, M., Berg, M., Sjollema, K., Bovenberg, R.A., Veenhuis, M., et al. 2002. delta-(L-alpha-Aminoadipyl)-L-cysteinyl-D-valine synthetase, that mediates the first committed step in penicillin biosynthesis, is a cytosolic enzyme. Fungal Genet Biol 37(1): 49-55. doi: 10.1016/S1087-1845(02)00036-1.

Walsh, C.T. 2002. Combinatorial biosynthesis of antibiotics: challenges and opportunities. Chembiochem 3(2-3): 125-134. doi: 10.1002/1439-7633(20020301)3:2/3%3C124::AID-CBIC124%3E3.0.CO;2-J.

Weissman, K.J., and Muller, R. 2008. Protein-protein interactions in multienzyme megasynthetases. Chembiochem 9(6): 826-848. doi: 10.1002/cbic.200700751.

Whicher, J.R., Dutta, S., Hansen, D.A., Hale, W.A., Chemler, J.A., Dosey, A.M., et al. 2014. Structural rearrangements of a polyketide synthase module during its catalytic cycle. Nature 510(7506): 560-564. doi: 10.1038/nature13409.

Zhang, J., and Demain, A. 1992a. In vitro stabilization of acv synthetase activity from *Streptomyces clavuligerus*. Appl Biochem Biotechnol 37(2): 97-110. doi: 10.1007/BF02921663.

Zhang, J., and Demain, A.L. 1992b. ACV synthetase. Crit Rev Biotechnol 12(3): 245-260. doi: 10.3109/07388559209069194.

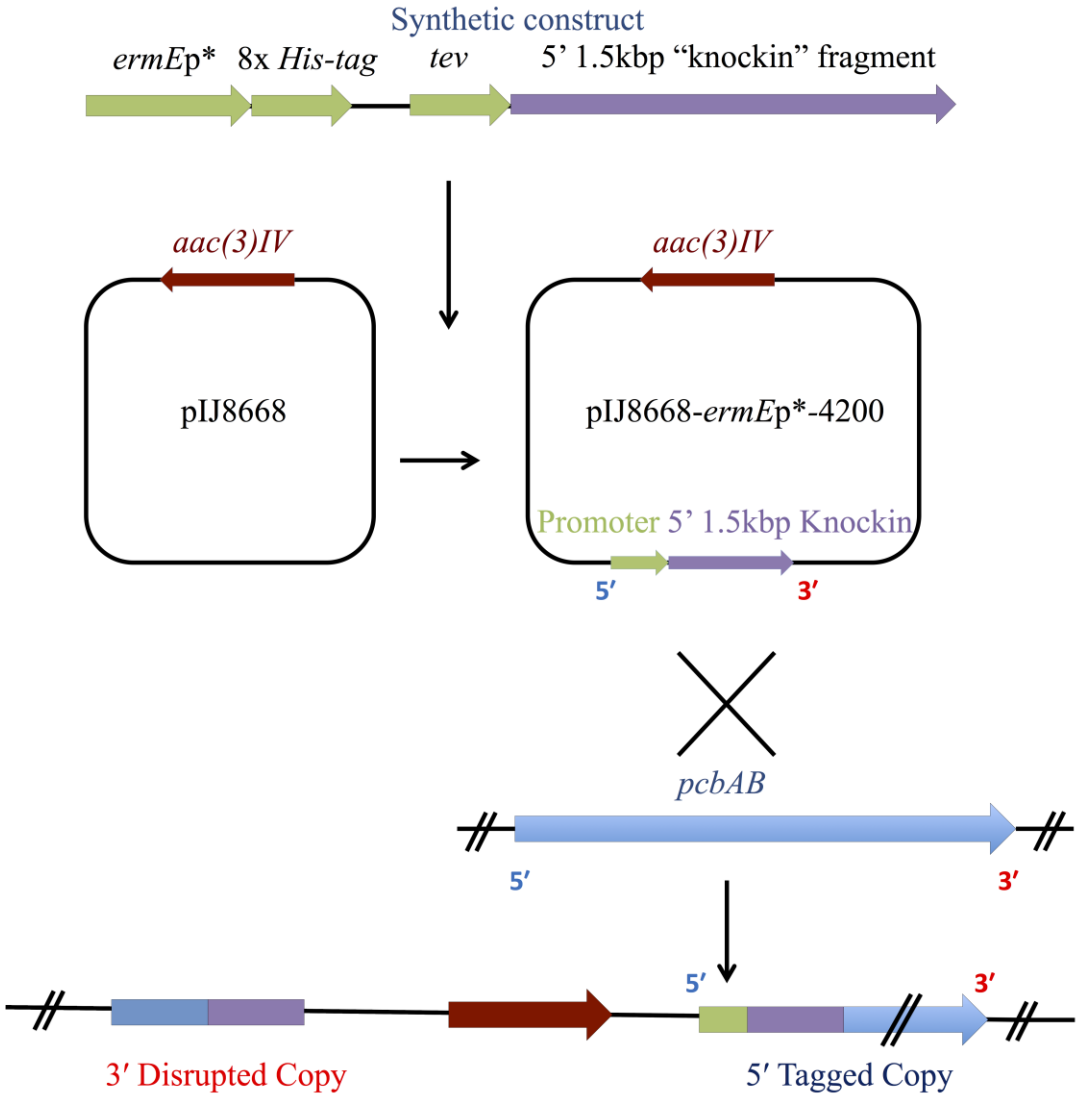


Figure 3.01 ACVS expression and purification tool is based on the recombination of a cleavable His-tagged promoter directly upstream of a gene of interest (*pcbAB*). Construct containing heterologous promoter and purification tools will be cloned into integration vector. The vector will then be conjugated into *Streptomyces* strain of interests where it will integrate upstream of gene of interest. This will result in the over-production of a tagged protein that can be easily purified.

Table 3.01: Plasmids and strains used during this study.

Name	Relative genotype and comments	Reference
pIDTSmart-KAN-*tipA*p	Proprietary plasmid in which *tipA*p *(his, tev*) was delivered, *neo*	This study
pIDTSmart-KAN-*ermE*p*	Proprietary plasmid in which *ermE*p* *(his, tev*) was delivered, *neo*	This study
pIJ8668	Non-integrative plasmid containing *egfp,* and *aac(3)IV*	(Sun et al. 1999)
pIJ8660	Integrative plasmid containing *egfp,* and *aac(3)IV*	(Sun et al. 1999)
pGEM-T® Easy	Proprietary plasmid used for cloning, *lacZ,* amp^r	Promega
pIJ8660-*tipA*p	pIJ8660 with *tipA*p *(his, tev*)	This study
pIJ8660-*ermE*p*	pIJ8660 with *ermE*p* *(his, tev*)	This study
pGEM-4200	pGEM-T® Easy with 5' 1.5Kb of *sclav_4200*	This study
pGEM-4466	pGEM-T® Easy with 5' 1.5Kb of s*clav_4466*	This study
pIJ8668-*ermE*p*	pIJ8668 with *ermE*p* *(his, tev*)	This study
pIJ8668-*tipA*p	pIJ8668 with *tipA*p *(his, tev*)	This study
pIJ8668-*ermE*p*-4200	pIJ8668-*ermE*p* with 5' 1.5kbp of *sclav_4200*	This study
pIJ8668-*ermE*p*-4466	pIJ8668-*ermE*p* with 5' 1.5kbp of sclav_4466	This study
pIJ8668-*tipA*p-4200	pIJ8668-*tipA*p with 5' 1.5kbp of *sclav_4200*	This study
pIJ8668-*tipA*p-4466	pIJ8668-*tipA*p with 5' 1.5kbp of sclav_4466	This study

pUZ8002	Helping plasmid for conjugation, *neo*	(Paget et al. 1999)
E. coli		
ET12567	Non-methylating *E. coli* strain for conjugation, cam^r, tet^r	(MacNeil et al. 1992)
DH5α	General lab strain for the introduction of plasmid DNA	(Bethesda Research Laboratories 1986)
S. clavuligerus		
NRRL3585	Wild Type (WT)	Northern Regional Research Laboratory, Peoria, Ill.
Δ*lat*	Genomic DNA template for 5' knockin PCR, Δ*lat::acc(3)IV*,	(Alexander et al. 2000)
WT/pIJ8668-*ermE*p*-4200	*acc(3)IV, ermE*p* (*his, tev*) integrated upstream of *sclav_4200*	This study
WT/pIJ8668-*ermE*p*-4466	*acc(3)IV, ermE*p* (*his, tev*) integrated upstream of *sclav_4466*	This study
WT/pIJ8660	*acc(3)IV, ermE*p* *(his, tev*), *egfp*	This study
WT/pIJ8668	*acc(3)IV*	This study
Δ*ccaR::tsr*	Δ*ccaR::tsr*[10]	(Alexander and Jensen 1998)
Δ*ccaR::tsr/*pIJ8668-*ermE*p*-4200	Δ*ccaR::tsr*, *acc(3)IV, ermE*p* *(his, tev*) integrated upstream of *sclav_4200*	This study
Δ*ccaR::tsr/*pIJ8668-*ermE*p*-4466	Δ*ccaR::tsr*, *acc(3)IV, ermE*p* *(his, tev*) integrated upstream of *sclav_4466*	This study

Δ*ccaR::tsr/*pIJ8660-*ermE*p*	Δ*ccaR::tsr*, *acc(3)IV, ermE*p* *(his, tev*), *egfp*	This study
Δ*ccaR::tsr/*pIJ8660-*tipA*p	Δ*ccaR::tsr*, *acc(3)IV, tipA*p *(his, tev*), *egfp*	This study
Δ*ccaR::tsr/*pIJ8660	Δ*ccaR::tsr*, *acc(3)IV, egfp*	This study
Δ*bls1::tsr*	Δ*bls1::tsr*	(Jensen et al. 2000)
Δ*bls1::tsr/*pIJ8668-*ermE*p*-4200	Δ*bls1::tsr*, *acc(3)IV, ermE*p* *(his, tev*) integrated upstream of *sclav_4200*	This study
Δ*bls1::tsr/*pIJ8668-*ermE*p*	Δ*bls1::tsr*, *acc(3)IV, ermE*p* *(his, tev*) integrated upstream of *sclav_4466*	This study
Δ*bls1::tsr*/pIJ8660-*ermE*p*	Δ*bls1::tsr*, *acc(3)IV, ermE*p* *(his, tev*), *egfp*	This study
Δ*bls1::tsr*/pIJ8660	Δ*bls1::tsr*, *acc(3)IV*, *egfp*	This study

Table 3.02: Oligonucleotides used during this study.

Primer Name	Sequence (5' – 3')	Purpose	Tm(^{o}C)
M13-R	GGTCATAGCTGTTTCC	Sequencing using MI3 loci	45
M13-F	TTTTCCCAGTCACGAC	Sequencing using MI3 loci	45
4200 - F	CGATACATATGATGTCAGCACGGT ACCCGA	Inserts NdeI restriction site and amplifies 1.5kb 5' sequence from *sclav_4200*	65
4200 - R	GCGGCCGCCCGTCGCGTTGAGGTA	Inserts NotI restriction site and amplifies 1.5kb 5' sequence from *sclav_4200*	65
4466 - F	GTACATATGATGTCGACACGTCGT TCGT	Inserts NdeI restriction site and amplifies 1.5kb 5' sequence from *sclav_4466*	60
4466 - R	GCGGCCGCCCCCATCAGCCCCAG	Inserts NotI restriction site and 1.5kb 5' sequence from *sclav_4466*	60
4200 Kn	TACCACTTCACATTCCGCAC	Reverse primer for knockin mutants of *sclav_4200*	52
4466 Kn	AGACACGGACAGCACACCA	Revers primer for knockin mutants of *sclav_4466*	52
4200 - KnSeq	AGCAACAGCCGTTTC	Sequencing the PCR product of Knockin with *sclav_4200*	52
4466 - KnSeq	ATCAGAAACGACCGG	Sequencing the PCR product of Knockin with *sclav_4466*	52
*ermE*p* Seq	GATATCGGTACCAGCCC	Sequencing the PCR product of *ermE*p* Knockin	52

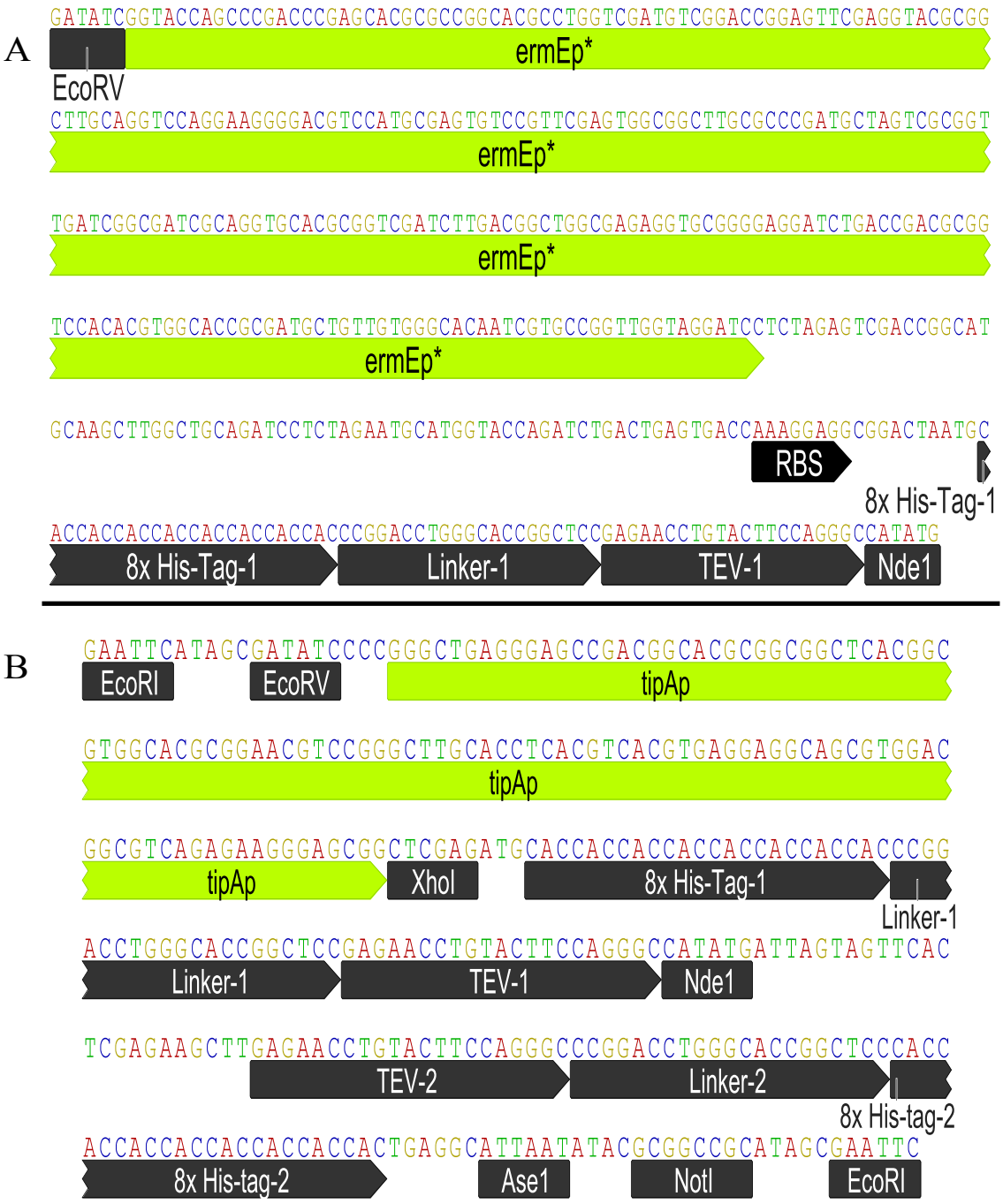


Figure 3.02: Restriction maps of the synthetic promoter constructs. **(A)** *ermE*p*, constitutive *Streptomyces* promoter; 8× His-Tag, polyhistidine tag; Linker-1/Linker-2, linker sequence between His-tag and TEV site; TEV, tobacco etch virus protease recognition site. **(B)** *tipA*p, thiostrepton inducible promoter; 8× His-tag, polyhistidine tag; Linker-1/Linker-2, linker sequence between His-tag and TEV site; TEV, tobacco etch virus protease recognition site.

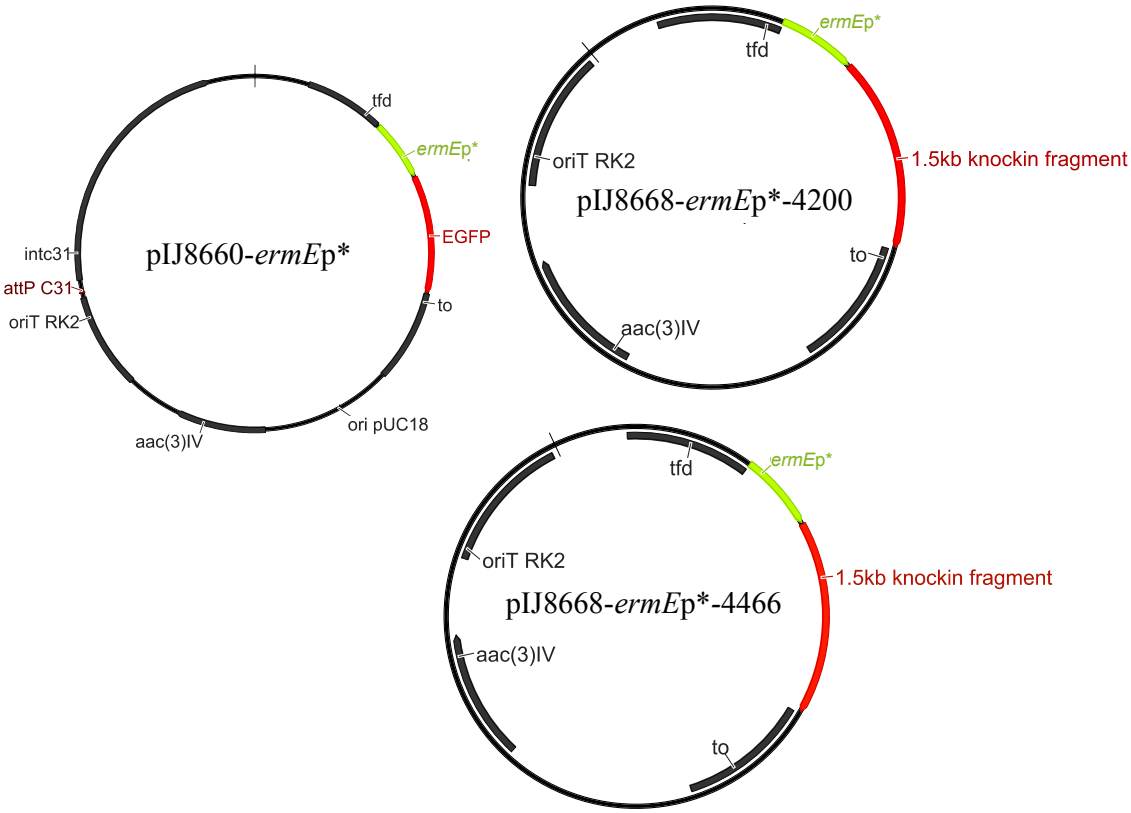


Figure 3.03: Plasmid maps of pIJ8660-*ermE*p*, pIJ8668-*ermE*p*-4200 and pIJ8668-*ermE*p*-4466. *tfd*, major transcription terminator of phage fd; *to*, transcription terminator of phage λ; *ori*T RK2, origin of transfer from RK2; *aac*(3)*IV*, apramycin resistance cassette; *ori* pUC18 is located between *to* and *aac*(3)IV in pIJ8668. pIJ8668-*ermE*p*-4466 is identical to pIJ8668-*ermE*p*-4200, with the exception of the knockin fragment.

Table 3.03: The BSA standards used to construct a standard curve based on final BSA concentration and absorbance at OD_{595} to extrapolate protein concentrations from unknown samples

Final BSA concentration µg/µl	**µl of BSA mixed to a final volume of 100µl 0.15 M NaCl**	**Absorbance at OD_{595}**
0.025	5	0.148
0.050	10	0.209
0.075	15	0.285
0.100	20	0.407
0.200	40	0.674
0.400	80	0.938
0.500	100	0.990

Table 3.04: Maximum survivable concentration of thiostrepton and apramycin in *tipA*p - based *S. clavuligerus* expression strains

	MSC (µg/ml)	
Strain	**Apramycin**	**Thiostrepton**
Wild Type	0	0
Δ*ccaR*	0	160
Δ*ccaR*/pIJ8668-*tipA*p-4200	400	160
Δ*ccaR*/pIJ8668-*tipA*p-4466	400	160

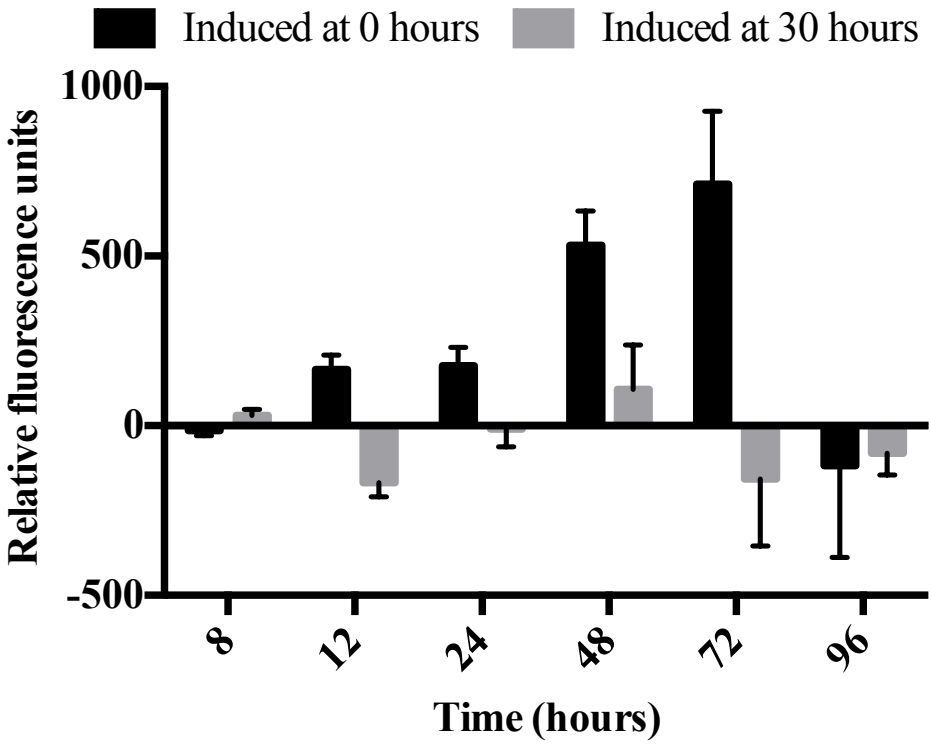


Figure 3.04: *S. clavuligerus* Δ*ccaR::tsr*/pIJ8660-*tipA*p produces signal from eGFP for 72 hours. *S. clavuligerus* Δ*ccaR::tsr*/pIJ8660-*tipA*p and pIJ8660 relative fluorescence units (RFU) were measured in triplicate over the course of 96 hours. Control strain (pIJ8660) was subtracted from *tipA*p constructs (pIJ8660-*tipA*p). Bars represent standard deviation.

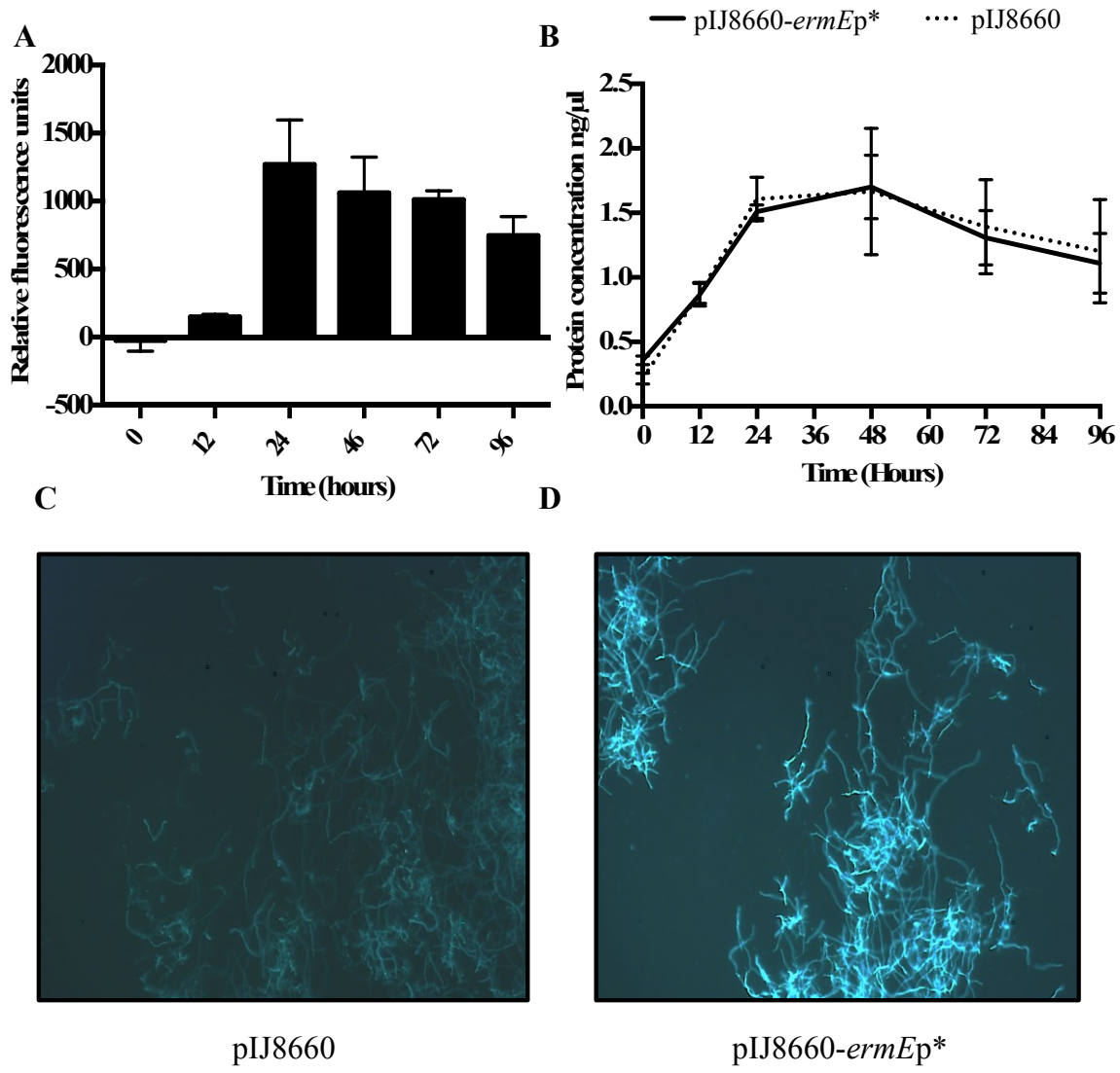


Figure 3.05: *S. clavuligerus* Δ*ccaR::tsr*/pIJ8660-*ermE*p* produces strong signal from eGFP. **(A)** *S. clavuligerus* Δ*ccaR::tsr*/pIJ8660-*ermE*p* and pIJ8660 RFU was measured in triplicate over the course of 96 hours, pIJ8660 RFU was subtracted from pIJ8660-*ermE*p*. Bars represent standard deviation. **(B)** Measurement of growth of each culture was estimated through protein concentrations (ng/µl) over 120 hours. Measurements done in triplicate and bars represent standard deviation. **(C)** Fluorescence microscopy of *S. clavuligerus* Δ*ccaR::tsr*/pIJ8660 after 72 hours. **(D)** Fluorescence microscopy of *S. clavuligerus* Δ*ccaR::tsr/*pIJ8660-*ermE*p* after 72 hours.

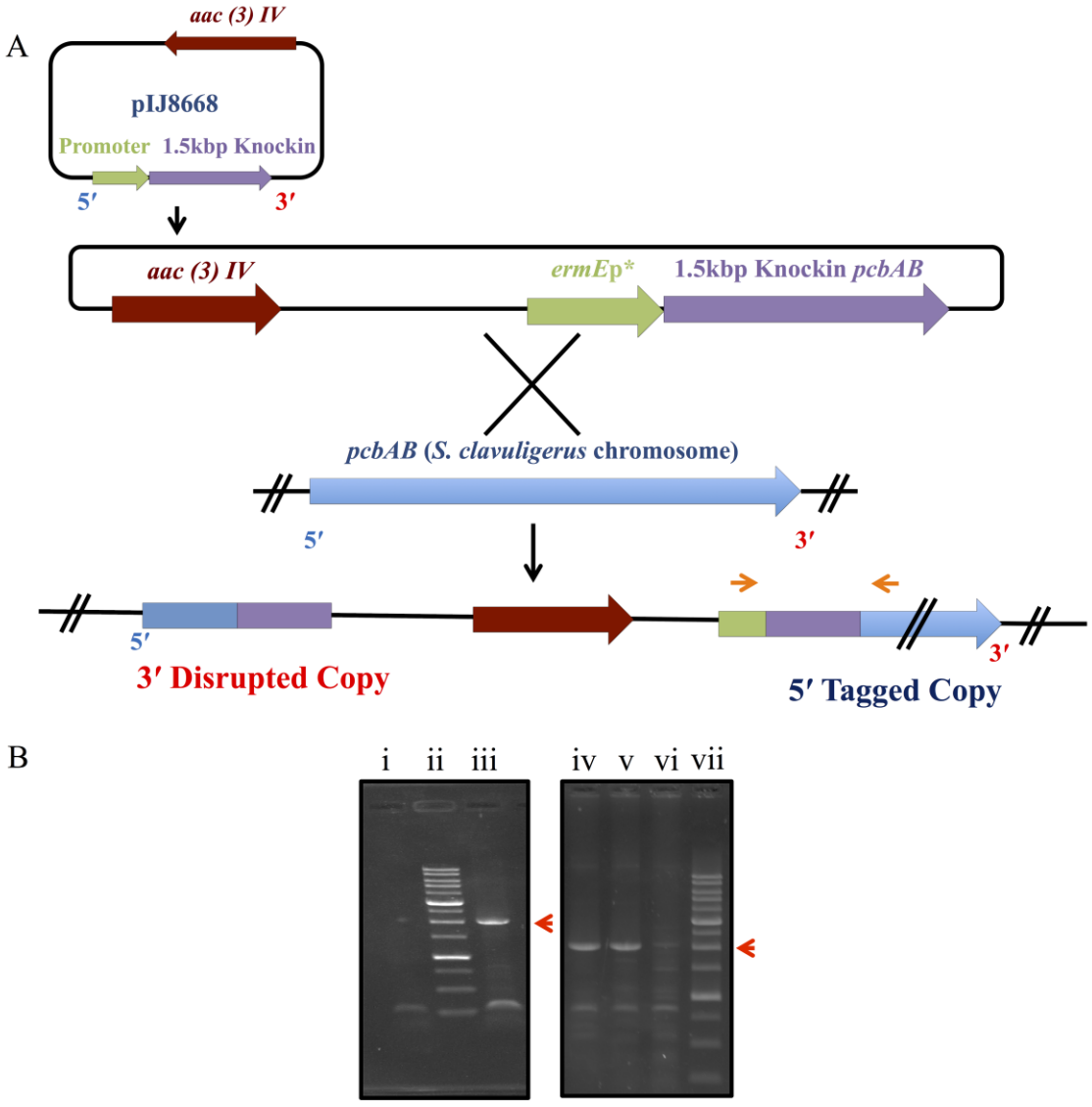


Figure 3.06: PCR amplifications confirming crossover at *pcbAB* and *sclav_4466* to give the corresponding *S. clavuligerus* expression strains. **(A)** Schematic representation of the process leading to the homologous recombination and promoter knockin for *pcbAB*. Primers run from promoter to just beyond knockin fragment (**orange arrows**). **(B)** Results from PCR assays confirming the correct knockin arrangement of the promoters in the respective *S. clavuligerus* strains. Expected fragment size of ~1.8kb is indicated just below the 2kbp marker (**red arrows**). **(i)**, negative control; **(ii)**, 1kb ladder; **(iii)**, *pcbAB* 1.8kb confirmation fragment; **(iv)**, *sclav_4466* 1.8kbp confirmation fragment; **(v)**, *sclav_4466* 1.8kbp confirmation fragment; **(vi)**, *sclav_4466* 1.8kbp confirmation fragment; **(vii)**, 1kb ladder.

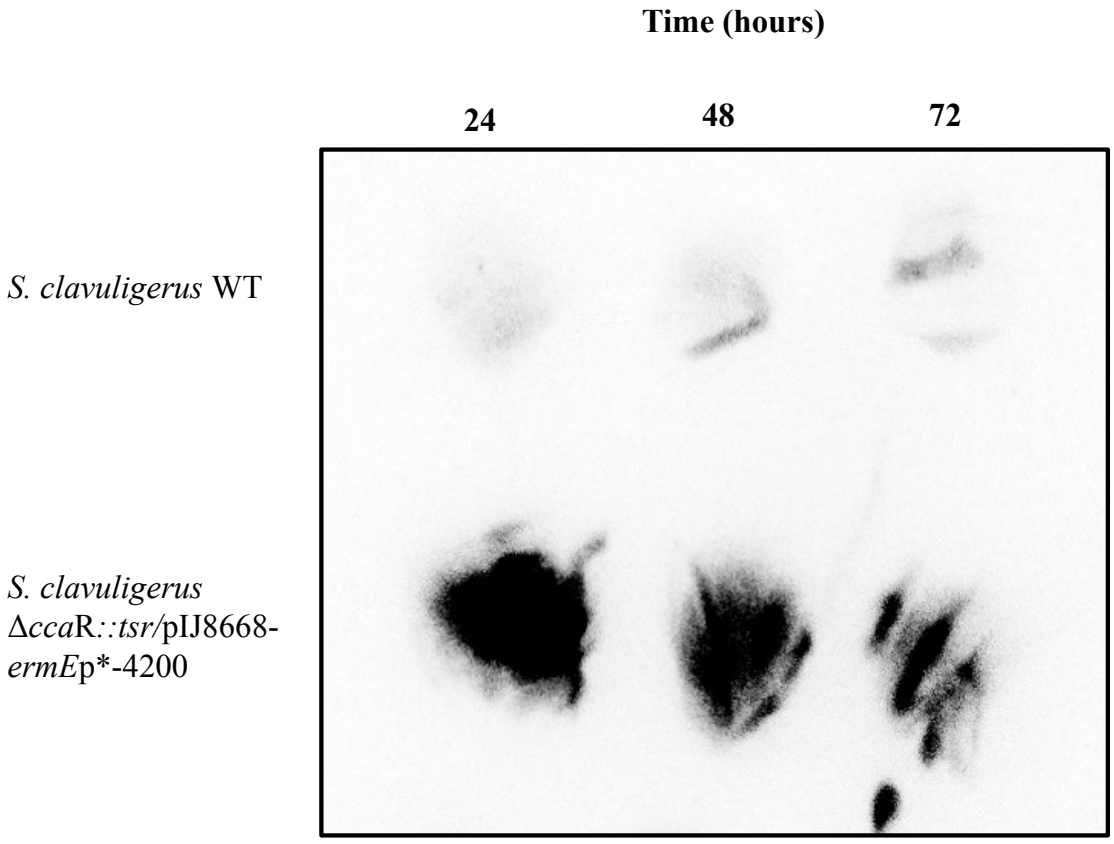


Figure 3.07: Preliminary dot-blot indicates that *S. clavuligerus* Δ*ccaR::tsr/*pIJ8668-*ermE*p*-4200 produces better ACVS signal than *S. clavuligerus* WT. Samples of crude soluble protein from *S. clavuligerus* WT and *S. clavuligerus* Δ*ccaR::tsr/*pIJ8668-*ermE*p*-4200 were drawn off at 24, 48, and 72 hours and tested against ACVS antibodies.

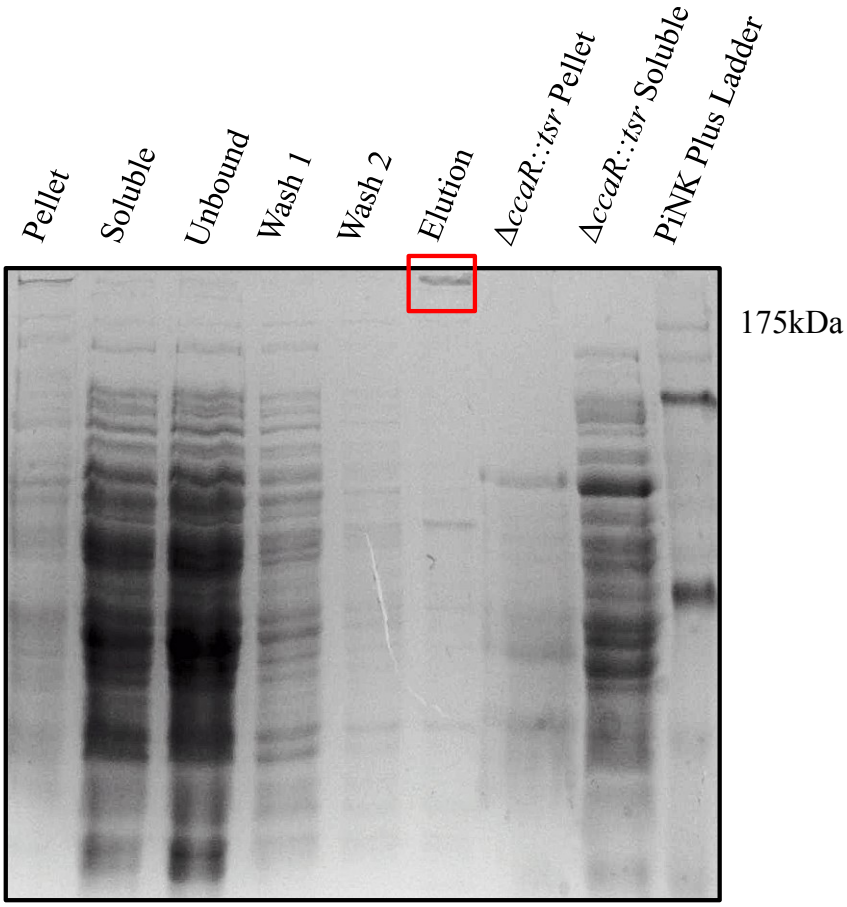


Figure 3.08: Preliminary analysis of purified ACVS by SDS-PAGE using a 10% w/v gels. Extracts from the *S. clavuligerus* Δ*ccaR::tsr/*pIJ8668-*ermE*p*-4200 strain was collected at various stages of the purification process as indicated above each lane. The red box indicates the putative ACVS protein recovered using Ni-NTA resin. Smaller bands observed below ACVS may be co-expressed or related proteins and were previously observed by Jensen et al. (1990). *S. clavuligerus* Δ*ccaR::tsr* produces no ACVS (Alexander et al. 2000). The gel was run at 57V at 4^{o}C until bromophenol blue ran off gel. Proteins were then visualized using Coomassie brilliant blue staining. Pellet, crude non-soluble fraction obtained after centrifugation of the lysate; Soluble, crude soluble proteins present in the supernatant after centrifugation of the lysate (prior to Ni-NTA resin); Unbound, crude soluble proteins present in supernatant after Ni-NTA resin treatment; Wash 1, first wash of Ni-NTA resin with wash buffer; Wash 2, second wash of Ni-NTA resin with wash buffer; Elution, eluted products from Ni-NTA resin; Δ*ccaR::tsr* Pellet, crude non-soluble proteins from Δ*ccaR::tsr;* Δ*ccaR::tsr* Soluble, crude soluble proteins from Δ*ccaR::tsr;* PiNK Plus Ladder, protein standards ladder.

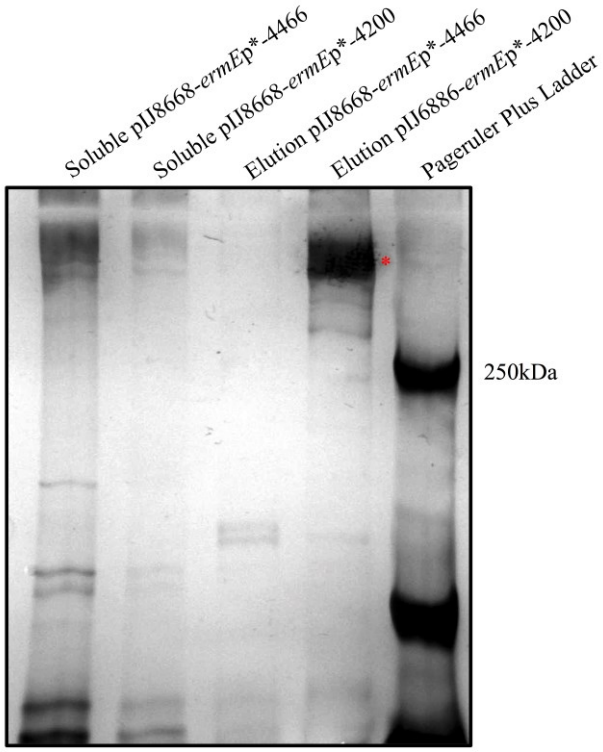


Figure 3.09: Large-scale purification of ACVS and SDS-PAGE analysis. Potential ACVS is indicated by *. 15 µg of *S. clavuligerus* Δ*ccaR::tsr* protein was loaded into each well of a 5% w/v polyacrylamide gel, which was run for 8 hours at 4°C and 57V. Visualization of the proteins in the gel was performed using silver staining. Soluble pIJ8668-*ermE*p*-4466, 15 µg of crude soluble extract; soluble pIJ8668-*ermE*p*-4200, 15 µg of crude soluble extract; elution pIJ8668-*ermE*p*-4466, 15 µg of 4466 protein eluted from nickel resin; elution pIJ8668-*ermE*p*-4200, 15 µg of ACVS protein eluted from nickel resin.

Chapter 4

4.01 Summary

4.01 Summary of completed work and future directions

Two systems were developed for the study of natural product biosynthesis: one, an approach for the activation and expression of silent/cryptic SM pathways; and two, a tool for studying the precise biochemical machinery involved in antibiotic production.

Protoplast fusion between *S. clavuligerus* and *S. coelicolor* M1154 did not produce favorable results. However, given the lessons learned from this study, it is possible that protoplast fusions can be produced using *S. lividans* and *S. clavuligerus*. In addition, it would be interesting to map the thiostrepton resistance observed after protoplast fusions in the current study to determine if it is a result of a previously described mutation or due to a novel mechanism. Heterologous expression of cryptic clusters from pSCL4 will most likely be possible by using *S. lividans* as an intermediate for passing pSCL4 from *S. clavuligerus* to *S. coelicolor* M1154.

The comparison between the *S. cattleya* and *S. clavuligerus* megaplasmids revealed limited similarity between the two elements, suggesting they are not closely related. The analysis of the 16S species trees (Supplementary Figures 1 and 2) showed a close relationship between *S. clavuligerus* and two clavulanic acid producers, *S. katsurahamenas* and *S. jumonjinensis*. It will be interesting to see if these other closely related species contain any megaplasmids. As a result, sequencing and annotation of these two strains was completed and analysis is ongoing. These data will help determine the

minimum genome for clavulanic acid and if megaplasmids are essential to its production. If megaplasmids are detected, then what sorts of secondary metabolite gene clusters are present on them and how do they compare to pSCL4 from *S. clavuligerus*?

Two synthetic promoter constructs were designed based on *ermE*p* and *tipA*p and were used to prepare multiple plasmids and strains. Although the overexpression and Ni-affinity purification of His-tagged ACVS was achieved using *ermE*p*, *tipA*p may be useful for the regulated expression of smaller proteins. The number of strains developed should allow for numerous downstream applications of this technology for the expression of proteins in their native *Streptomyces* hosts. The technology could be used for comparative studies on the production of cephamycin C in an ACVS overexpression strain since ACVS is the rate limiting enzyme in the pathway (Zhang and Demain 1992). Such strategies could also be used to overexpress other mega proteins to see if product formation is enhanced or activated in the case of cryptic gene clusters.

This system further lends itself to localization studies using the constitutive or regulated expression of epitope-tagged mega proteins. ACVS in fungal β-lactam producers is a cytosolic enzyme (van der Lende et al. 2002), whereas some other proteins involved in the pathway are found in membrane-bound compartments (van der Lende et al. 2002). It has long been postulated that ACVS in *S. clavuligerus* might somehow interact with the cephamycin C membrane transport protein, CmcT (K. Tahlan, personal communication). In order to study the localization of ACVS, and other proteins in *S. clavuligerus*, the knockin constructs could also be introduced into *S. clavuligerus*. The cellular localization of ACVS and any other protein of interest could be detected via immunostaining for the fused His-tag, followed by transmission electron microscopy.

The most promising aspect of this work is the future collaborative application of these constructs for structural studies of ACVS. X-ray studies have been done on C domains (Bloudoff and Schmeing 2013; Keating et al. 2002; Samel et al. 2007) and A domains (Conti et al. 1997; May et al. 2002) of NRPS proteins, but not on a fully intact protein. In NRPSes, substrate specificity is dictated by the A domain; however, the C domain does have some role in substrate selection, the nature and mechanism of which is unclear. The C domain also has a highly conserved histidine motif involved in NRP catalysis (Marahiel et al. 1997), though there is no agreement on what role this plays in condensation (Bloudoff and Schmeing 2013). Solving a complete and intact NRPS structure will help clarify the mechanism and substrate specificity of C domains, as well as provide broader insight into the workings of an NRPS as a whole. This has numerous future potential applications for bioengineering or developing designer peptides.

BIBLIOGRAPHY

Bloudoff, K., and Schmeing, T.M. 2013. Crystallization and preliminary crystallographic analysis of the first condensation domain of viomycin synthetase. Acta Crystallogr Sect F Struct Biol Cryst Commun 69(Pt 4): 412-415. doi: 10.1107/S1744309113004004.

Conti, E., Stachelhaus, T., Marahiel, M.A., and Brick, P. 1997. Structural basis for the activation of phenylalanine in the non-ribosomal biosynthesis of gramicidin S. EMBO J 16(14): 4174-4183.

Keating, T.A., Marshall, C.G., Walsh, C.T., and Keating, A.E. 2002. The structure of VibH represents nonribosomal peptide synthetase condensation, cyclization and epimerization domains. Nat Struct Biol 9(7): 522-526. doi: 10.1038/nsb810.

Liras, P. 1999. Biosynthesis and molecular genetics of cephamycins. Cephamycins produced by actinomycetes. Antonie Van Leeuwenhoek 75(1-2): 109-124.

Marahiel, M.A., Stachelhaus, T., and Mootz, H.D. 1997. Modular Peptide Synthetases Involved in Nonribosomal Peptide Synthesis. Chem Rev 97(7): 2651-2674.

May, J.J., Kessler, N., Marahiel, M.A., and Stubbs, M.T. 2002. Crystal structure of DhbE, an archetype for aryl acid activating domains of modular nonribosomal peptide synthetases. Proc Natl Acad Sci U S A 99(19): 12120-12125. doi: 10.1073/pnas.182156699.

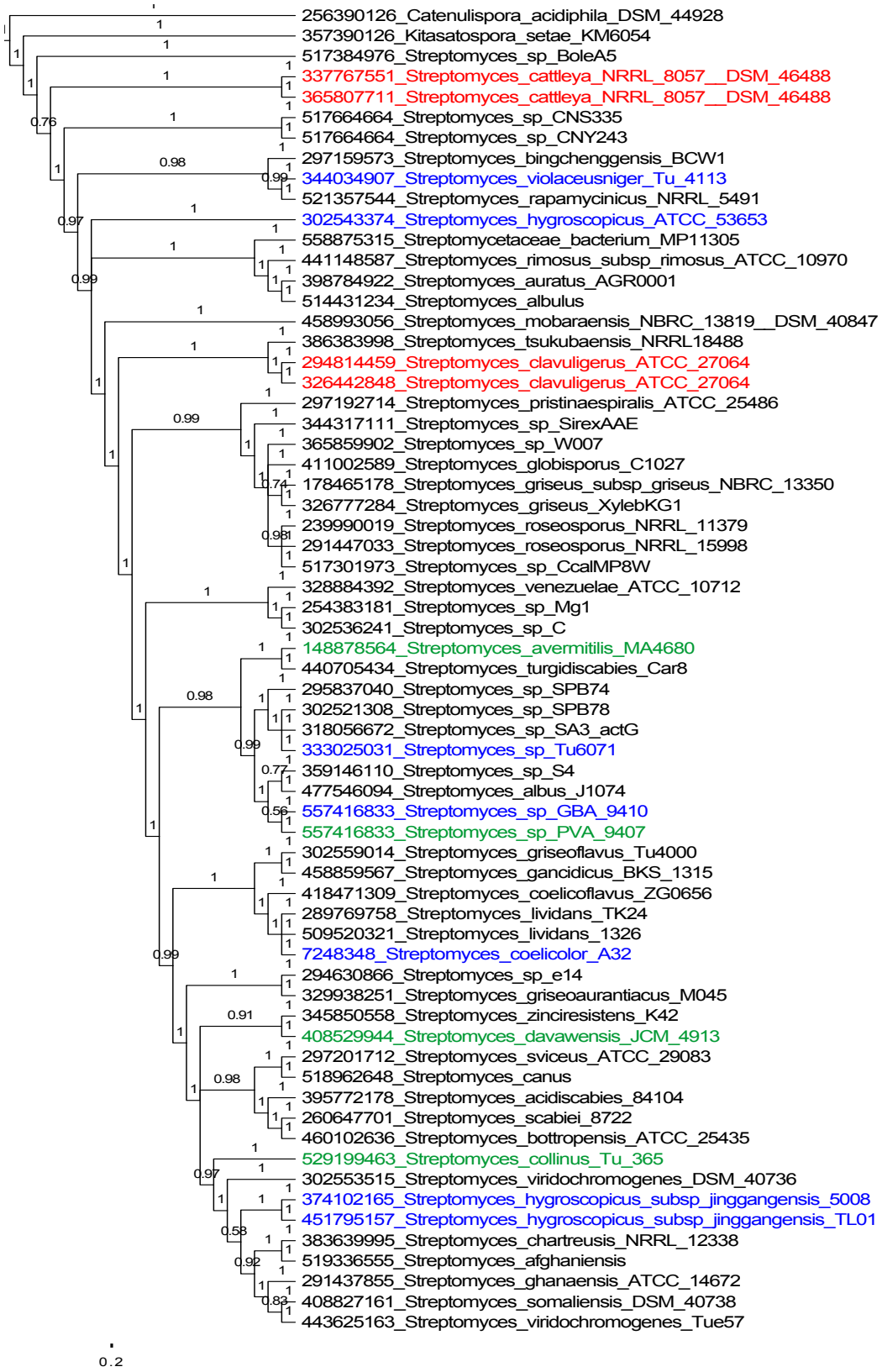
Samel, S.A., Schoenafinger, G., Knappe, T.A., Marahiel, M.A., and Essen, L.O. 2007. Structural and functional insights into a peptide bond-forming bidomain from a nonribosomal peptide synthetase. Structure 15(7): 781-792. doi: 10.1016/j.str.2007.05.008.

van der Lende, T.R., van de Kamp, M., Berg, M., Sjollema, K., Bovenberg, R.A., Veenhuis, M., et al. 2002. delta-(L-alpha-Aminoadipyl)-L-cysteinyl-D-valine synthetase, that mediates the first committed step in penicillin biosynthesis, is a cytosolic enzyme. Fungal Genet Biol 37(1): 49-55. doi: 10.1016/S1087-1845(02)00036-1.

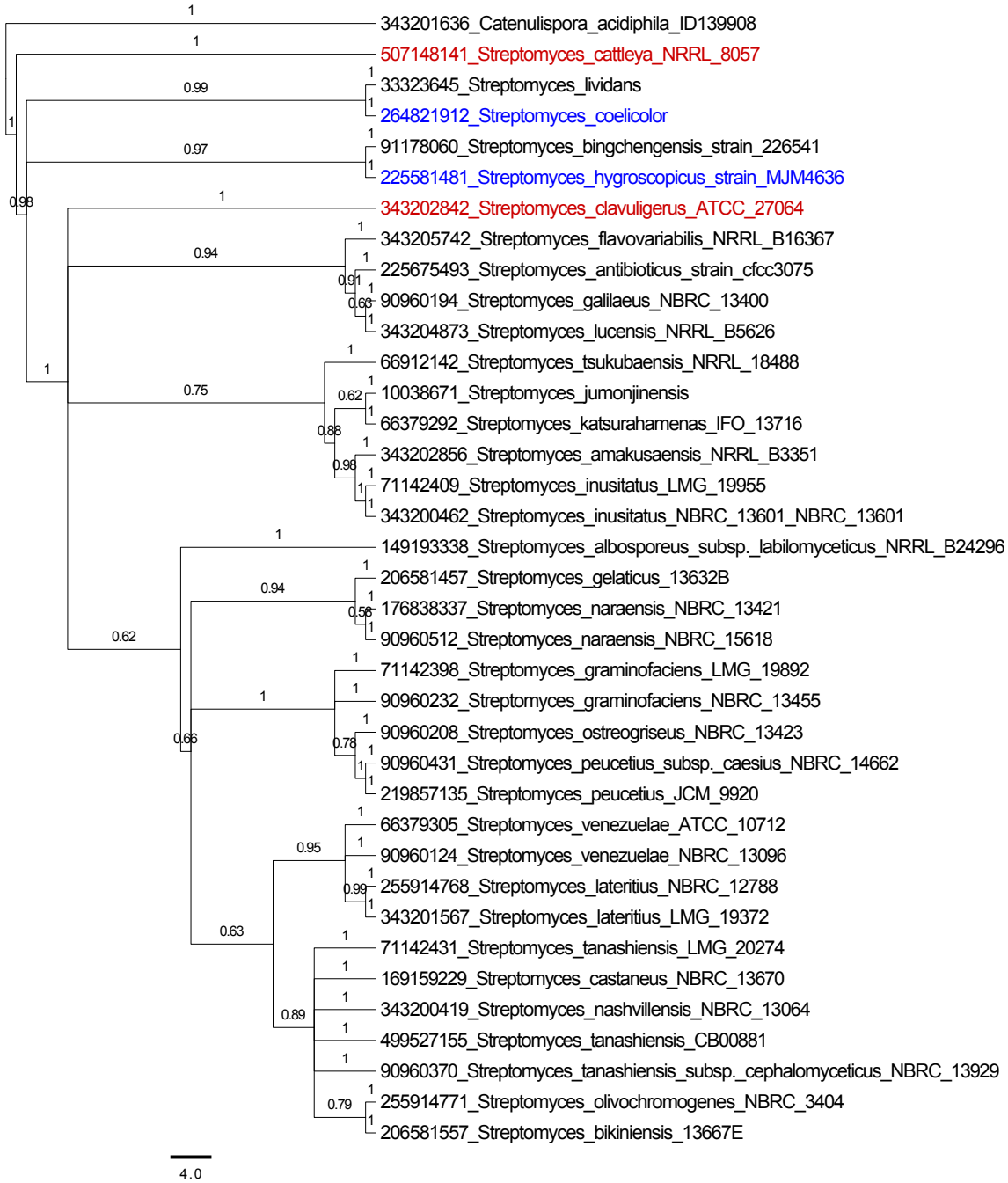
Zhang, J., and Demain, A.L. 1992. ACV synthetase. Crit Rev Biotechnol 12(3): 245-260. doi: 10.3109/07388559209069194.

Appendix A

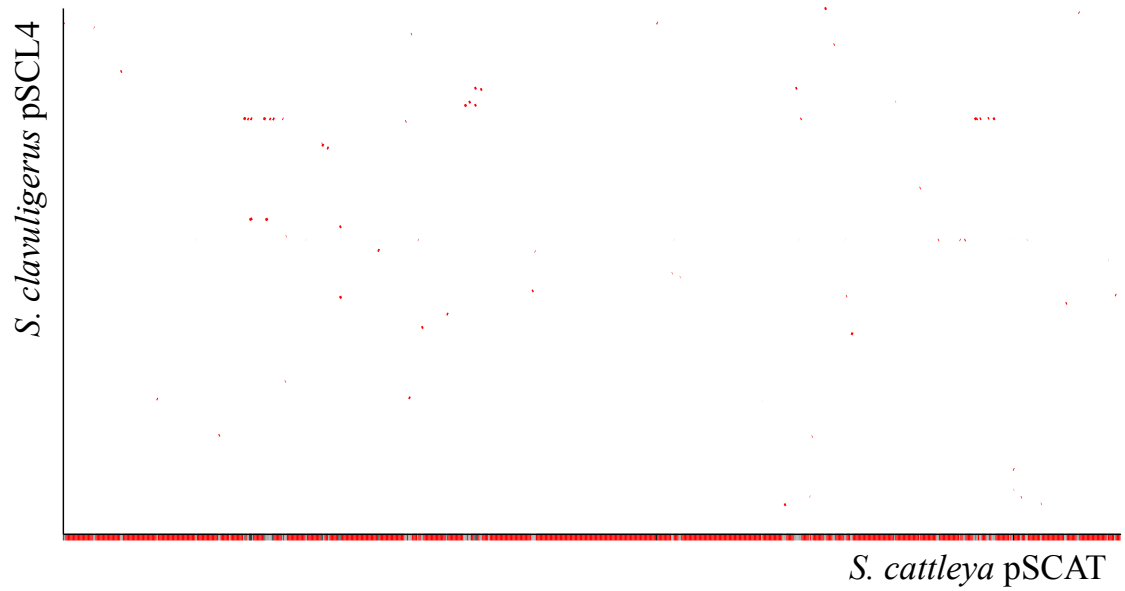
Supplementary data supporting chapter 1


256390126_Catenulispora_acidiphila_DSM_44928
357390126_Kitasatospora_setae_KM6054
517384976_Streptomyces_sp_BoleA5
337767551_Streptomyces_cattleya_NRRL_8057__DSM_46488
365807711_Streptomyces_cattleya_NRRL_8057__DSM_46488
517664664_Streptomyces_sp_CNS335
517664664_Streptomyces_sp_CNY243
297159573_Streptomyces_bingchenggensis_BCW1
344034907_Streptomyces_violaceusniger_Tu_4113
521357544_Streptomyces_rapamycinicus_NRRL_5491
302543374_Streptomyces_hygroscopicus_ATCC_53653
558875315_Streptomycetaceae_bacterium_MP11305
441148587_Streptomyces_rimosus_subsp_rimosus_ATCC_10970
398784922_Streptomyces_auratus_AGR0001
514431234_Streptomyces_albulus
458993056_Streptomyces_mobaraensis_NBRC_13819__DSM_40847
386383998_Streptomyces_tsukubaensis_NRRL18488
294814459_Streptomyces_clavuligerus_ATCC_27064
326442848_Streptomyces_clavuligerus_ATCC_27064
297192714_Streptomyces_pristinaespiralis_ATCC_25486
344317111_Streptomyces_sp_SirexAAE
365859902_Streptomyces_sp_W007
411002589_Streptomyces_globisporus_C1027
178465178_Streptomyces_griseus_subsp_griseus_NBRC_13350
326777284_Streptomyces_griseus_XylebKG1
239990019_Streptomyces_roseosporus_NRRL_11379
291447033_Streptomyces_roseosporus_NRRL_15998
517301973_Streptomyces_sp_CcalMP8W
328884392_Streptomyces_venezuelae_ATCC_10712
254383181_Streptomyces_sp_Mg1
302536241_Streptomyces_sp_C
148878564_Streptomyces_avermitilis_MA4680
440705434_Streptomyces_turgidiscabies_Car8
295837040_Streptomyces_sp_SPB74
302521308_Streptomyces_sp_SPB78
318056672_Streptomyces_sp_SA3_actG
333025031_Streptomyces_sp_Tu6071
359146110_Streptomyces_sp_S4
477546094_Streptomyces_albus_J1074
557416833_Streptomyces_sp_GBA_9410
557416833_Streptomyces_sp_PVA_9407
302559014_Streptomyces_griseoflavus_Tu4000
458859567_Streptomyces_gancidicus_BKS_1315
418471309_Streptomyces_coelicoflavus_ZG0656
289769758_Streptomyces_lividans_TK24
509520321_Streptomyces_lividans_1326
7248348_Streptomyces_coelicolor_A32
294630866_Streptomyces_sp_e14
329938251_Streptomyces_griseoaurantiacus_M045
345850558_Streptomyces_zinciresistens_K42
408529944_Streptomyces_davawensis_JCM_4913
297201712_Streptomyces_sviceus_ATCC_29083
518962648_Streptomyces_canus
395772178_Streptomyces_acidiscabies_84104
260647701_Streptomyces_scabiei_8722
460102636_Streptomyces_bottropensis_ATCC_25435
529199463_Streptomyces_collinus_Tu_365
302553515_Streptomyces_viridochromogenes_DSM_40736
374102165_Streptomyces_hygroscopicus_subsp_jinggangensis_5008
451795157_Streptomyces_hygroscopicus_subsp_jinggangensis_TL01
383639995_Streptomyces_chartreusis_NRRL_12338
519336555_Streptomyces_afghaniensis
291437855_Streptomyces_ghanaensis_ATCC_14672
408827161_Streptomyces_somaliensis_DSM_40738
443625163_Streptomyces_viridochromogenes_Tue57
0.2

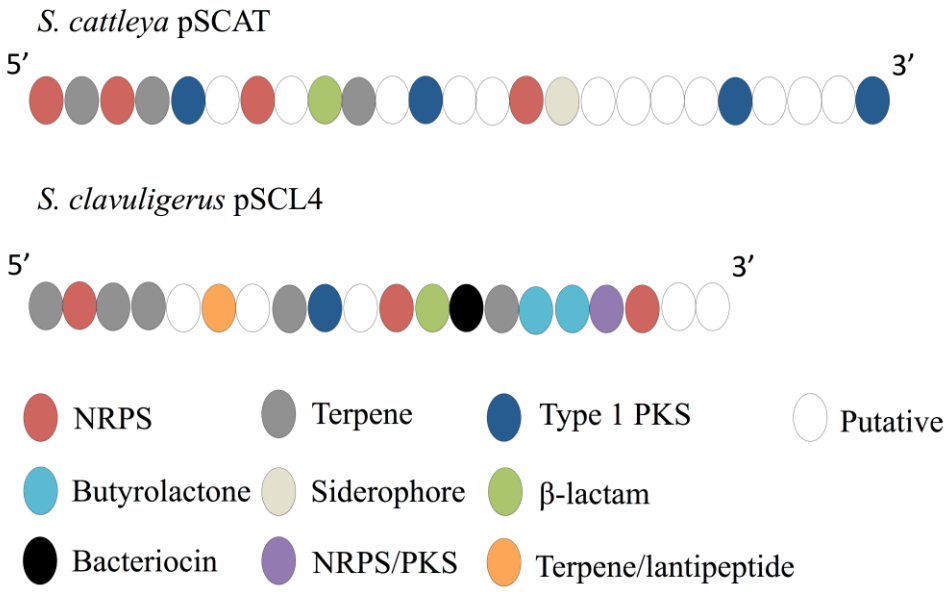
Supplementary Figure 1: Phylogenetic reconstruction based on RpoB protein sequences of *Streptomyces*. Large extra-chromosomal elements appear to be evolutionarily diverse throughout the genus *Streptomyces.* NCBI was canvased for complete genomic sequences of *Streptomyces*. Of particular interest was the presence of large plasmids. The gene identifiers (GI) of each sequence are identified before its name. Species labeled with colour highlight the varying sizes of these elements within *Streptomyces*: red, >400kb; blue, 100kb to 400kb; green, 50kb to 100kb. Multiple sequencing projects of *S. clavuligerus* and *S. cattleya* result in duplications within the tree.



Supplementary Figure 2: 16S rDNA phylogenetic reconstruction of closely related *Streptomyces* closely related to *S. clavuligerus* and clavulanic acid producers (*S. clavuligerus, S. katsurahamenas,* and *S. jumonjinensis)*. The genetic identifiers (GI) of each sequence are identified before its name. Colours correspond with Supplementary Figure 1.

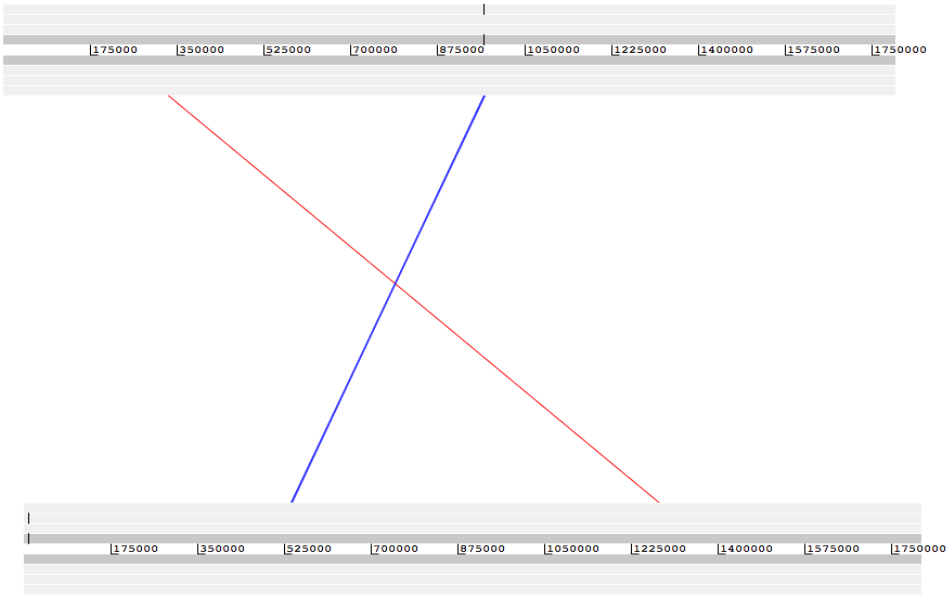


Supplementary Figure 3: Synteny plot between pSCL4 and pSCAT shows little structural relationship between the two megaplasmids. pSCL4 and pSCAT were aligned using R2CAT to determine any structural relationship between the two megaplasmids. R2CAT found limited synteny between the two megaplasmids, indicating limited structural relationship.



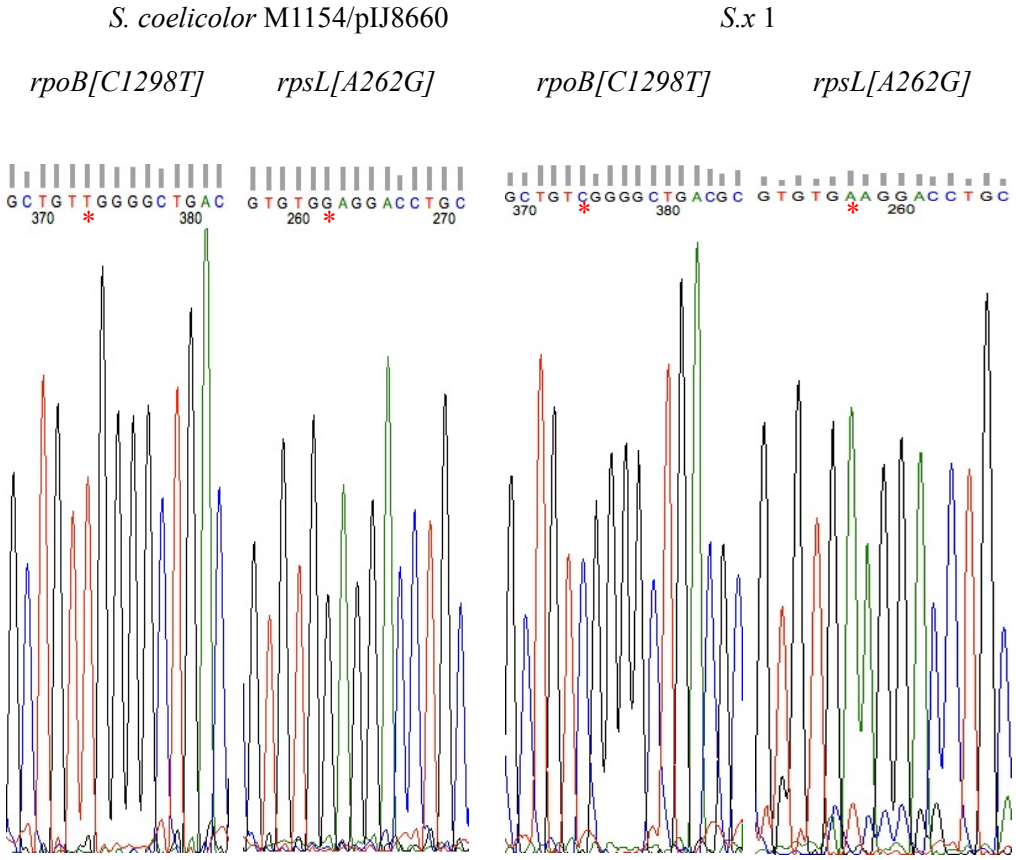
Supplementary Figure 4: Visual comparison of the secondary metabolite gene clusters encoded by the linear pSCAT and pSCL4 showing that both megaplasmids are packed with a variety of potential biosynthetic pathways. Clusters were identified using the AntiSMASH2.0 software and are shown in a linear manner, starting from 5' to 3'.

S. clavuligerus pSCL4



S. cattleya pSCAT

Supplementary Figure 5: Plasmid alignment and comparison shows little similarity between pSCL4 and pSCAT. The alignment was generated using the Artemis Comparison Tool (ACT) with the correlation score set to 700.



Supplementary Figure 6: Advantageous mutations in *rpoB* and *rpsL* gene sequences are present in *S. coelicolor* M1154/pIJ8660 but not present in isolates obtained after protoplast fusions. * Represents mutated nucleotide.

Supplementary Table 1: Complete prediction of interesting secondary metabolite gene clusters located on pSCL4, including predicted terminal product, protein description, and organisms harboring similar pathways. Homologous gene clusters contain >40% of genes with sequence identity >45%.

Product	Locus Tag	Description	Organisms with homologous Gene clusters
NRP[1]	*sclav_p0286*	Phosphoribosylglycinamide synthetase	*Saccharopolyspora erythraea* NRRL 2338
	sclav_p0287	Aspartate decarboxylase	*Pseudomonas entomophila* L48
	sclav_p0288	Ornithine carbamoyltransferase	*Streptomyces flavogriseus* ATCC 33331
	sclav_p0289	Methyltransferase	*Amycolatopsis mediterranei*
	sclav_p0290	NRPS	*Saccharomonospora glauca* K62
	sclav_p0291	AMP-dependent synthetase	*Micromonospora sp.* ATCC 39149
	sclav_p0292	Cysteine synthase	
	sclav_p0293	Ornithine cyclodeaminase	
	sclav_p0294	FAD dependent oxidoreductase family	
	sclav_p0295	AMP-dependent synthetase	
	sclav_p0296	Putative acyl carrier protein	
	sclav_p0297	Acyl-CoA dehydrogenase	
	sclav_p0298	Aminotransferase	
	sclav_p0299	Ornithine cyclodeaminase	
	sclav_p0300	Cysteine synthase	
	sclav_p0301	Condensation domain containing protein	
	sclav_p0302	Argininosuccinate lyase	
	sclav_p0303	NRPS	
	sclav_p0304	NRPS	
	sclav_p0305	AMP-dependent synthetase and ligase – Adenylation and PCP domains	
	sclav_p0306/ sclav_p0307	ABC transporter related protein	
	sclav_p0308	AMP-dependent synthetase and ligase	
PK[2]	*sclav_p0327*	Cytochrome 450	No Hit
	sclav_p0328	Terpene Synthase	
	sclav_p0329	Methyltransferase	
	sclav_p0330	Flavin-containing amine	

		oxidoreductase	
	sclav_p0331/ sclav_p0332	Beta-ketoacyl synthase	
	sclav_p0331	Beta-ketoacyl synthase	
NRP - like	*sclav_p0563*	Histidine Kinase	*Streptomyces bingchenggensis* BCW-1
	sclav_p0564	Adenylation domain, thioesterase	*Streptomyces scabiei* 87.22
	sclav_p0565	Acyl Carrier protein	*Streptomyces hygroscopicus* ATCC 53653
	sclav_p0570	Acyltransferase	*Streptomyces sp. Sirex*AA-E
	sclav_p0571	Terpene Synthase	*Streptomyces cf. griseus*
	sclav_pt001	tRNA-Ala	
	sclav_p0573	Cytochrome P450	
	sclav_p0573	Pentalenene synthase	
NRP	*sclav_p1007*	Metallo-beta-lactamase family protein	*Nocardia brasiliensis* ATCC 700358
	sclav_p1005	NRPS with PKS like domains	*Streptomyces roseosporus* NRRL 15998
	sclav_p1004	Sulfotransferase	*Streptomyces lividans* TK24
	sclav_p1003	Decarboxylases	*Burkholderia thailandensis* E264
	sclav_p1002	Short-chain dehydrogenases	*Rhodococcus opacus* B4
			Burkholderia multivorans CGD2M
β-Lactam	*sclav_p1069*	Cysteine synthase	*Streptomyces antibioticus* strain Tu 1718
	sclav_p1070	Aminotransferase	*Streptomyces flavogriseus* ATCC 33331
	sclav_p1072	Aminotransferase	Streptomyces clavuligerus ATCC 27064 - chromosome
	sclav_p1074	Pyruvate oxidase	*Saccharomonospora viridis* DSM 43017
	sclav_p1075	Asparagine Synthase	*Anoxybacillus flavithermus* WK1
	sclav_p1076	Arginase	*Nocardia brasiliensis* ATCC 700358
	sclav_p1077	ArgJ protein family	
	sclav_p1078	Aminotransferase	
	sclav_p1079	SARP	
	sclav_p1080	Histidine Kinase	

PK	*sclav_p1225*	Histidine decarboxylase	*Streptomyces sp. W007*
	sclav_p1227	Haloalkane dehalogenase	*Streptomyces ghanaensis* ATCC 14672
	sclav_p1233	PKS	*Streptosporangium roseum* DSM 43021
	sclav_p1234	Thioesterase	*Streptomyces carzinostaticus subsp. neocarzinostaticus*
			Verrucosispora maris AB-18-032
			Salinispora tropica CNB-440
NRP/PK[3]	*sclav_p1272*	AMP-dependent synthetase and ligase	No Hit
	sclav_p1273	Thioesterase	
	sclav_p1269	Acyltransferase	
	sclav_p1268	Acyl carrier protein	
	sclav_p1267	Transposase	
	sclav_p1266	Acyl-ACP Thioesterase	
	sclav_p1265	Methyltransferase	
	sclav_p1264	AMP-dependent synthetase	
	sclav_p1263	Dehydrogenase	
	sclav_p1278	NAD-dependent epimerase	
	sclav_p1279	Asparagine synthase	
	sclav_p1280	Glycosyltransferase, MGT family	
	sclav_p1281	Asparagine synthase	
NRP	*sclav_p1301*	AMP-dependent synthetase and ligase - Adenylation domain	
	sclav_p1302	Condensation domain containing protein	
	sclav_p1303	NRPS	
	sclav_p1304	NRPS	
	sclav_p1305	Isomerase	
	sclav_p1307	AMP-dependent synthetase and ligase - Adenylation PCP	
	sclav_p1308	Cytochrome P450	
	sclav_p1309	Thioesterase	
	sclav_p1311	β-Lactamase	
NRP/PK	*sclav_p1325*	ACP	No Hit

	sclav_p1327	Acyl-CoA Synthetase - AMP-dependent synthetase and ligase	
	sclav_p1328	O-Methyltransferase	
	sclav_p1335	Acyl-CoA Synthetase - AMP-dependent synthetase and ligase	
	sclav_p1339	Peptide synthetase - AMP-dependent synthetase and ligase	
	sclav_p1340	NRPS	
	sclav_p1341	NRPS	
	sclav_p1350	PKS	
	sclav_p1351	Thioesterase	
	sclav_p1352	Reductase	
	sclav_p1353	Cytochrome P450	
	sclav_p1368	SARP	
NRP	*sclav_p1420*	Short-chain dehydrogenase	*Stackebrandtia nassauensis* DSM 44728
	sclav_p1422	Serine protease	*Streptomyces bingchenggensis* BCW-1
	sclav_p1423	N-acetylgalactosamine 6-sulfatase	*Streptomyces hygroscopicus* ATCC 53653
	sclav_p1426	Methyltransferase	*Streptomyces sp. Tu6071*
	sclav_p1427	NRPS	*Actinomadura sp.* ATCC 39727
	sclav_p1428	Prenyltransferases	*Streptomyces albus* J1074
	sclav_p1429	Terpene synthase family	*Amycolatopsis mediterranei* S699
	sclav_p1439	Histidine Kinase	
NRP	*sclav_p1474*	NRPS	*Streptomyces roseosporus* NRRL 15998
	sclav_p1475	Transporter	*Streptomyces bingchenggensis* BCW-1
	sclav_p1476	Chloramphenicol acetyltransferase	*Agrobacterium radiobacter* K84 plasmid pAtK84c
	sclav_p1477	Cysteine synthase	*Amycolatopsis mediterranei* S699
	sclav_p1484	Oxidoreductase	*Actinosynnema mirum* DSM 43827
			Streptomyces griseus subsp. griseus NBRC 13350

[1] NRP = nonribosomal peptide [3] NRP/PK = hybrid [2] PK = polyketide

Supplementary Table 2: Complete prediction of interesting secondary metabolite gene clusters located on pSCAT, including predicted terminal product, protein description, and organisms harboring similar pathways. Homologous gene clusters contain >40% of genes with sequence identity >45%.

Product	Locus Tag	Description	Organisms with homologous gene clusters
NRP[1]	*scat_p0034*	Histidine kinase	*Streptomyces bingchenggensis* BCW-1
	scat_p0037	Oxidoreductase	*Streptomyces hygroscopicus subsp. jinggangensis* 5008
	scat_p0038	Condensation domain containing protein	*Streptomyces albus* J1074
	scat_p0039	NRPS	*Nocardia cyriacigeorgica* GUH-2
	scat_p0040	Thioesterase	*Nocardia brasiliensis* ATCC 700358
	scat_p0041	Condensation domain containing protein	*Stackebrandtia nassauensis* DSM 44728
	scat_p0042	NRPS	*Streptomyces bingchenggensis* BCW-1
	scat_p0072	Methyltransferase	*Verrucosispora maris* AB-18-032
	scat_p0074	PKS Enoyl reductase	
	scat_p0092	Condensation domain containing protein	
NRP	*scat_p0157*	Methyltransferase	*Stackebrandtia nassauensis* DSM 44728
	scat_p0168	PK - Crotonyl-CoA Reductase	*Salinispora tropica* CNB-440
	scat_p0172	Keto reductase	*Streptomyces bingchenggensis* BCW-1
	scat_p0176	Thioesterase	*Streptomyces albus* J1074
	scat_p0179 /p0180	NRPS	*Actinosynnema mirum* DSM 43827
	scat_p0192	PKS fragment, Enoyl reductase	*Streptomyces griseus subsp. griseus* NBRC 13350

			Ralstonia solanacearum GMI1000 plasmid pGMI1000MP
			Nocardiopsis dassonvillei subsp. dassonvillei DSM 4311
PK[2]	*scat_p0291*	Keto Reductase	*Streptomyces rochei* 7434AN4 plasmid pSLA2-L
	scat_p0297	PKS	*Streptomyces tsukubaensis* NRRL18488
	scat_p0298	Acetyl-CoA Carboxylase	*Amycolatopsis mediterranei* U32
	scat_p0303 /p0304	Drug Resistance ABC transporter	*Nocardia brasiliensis* ATCC 700358
	scat_p0305	Transmembrane efflux pump	*Catenulispora acidiphila* DSM 44928
	scat_p0306	Thioesterase	*Streptomyces sp. KCTC* 11604BP FK506
	scat_p0308	Acyl-CoA dehydrogenase	*Streptomyces bingchenggensis* BCW-1
	scat_p0309	Putative Acyl Carrier protein	
	scat_p0311	Beta-Ketoacyl synthase	
	scat_p0312	PKS	
	scat_p0313	PKS	
	scat_p0314	PKS	
	scat_p0315	PKS	
	scat_p0317	Acetyltransferase	
	scat_p0327	Polyketide cyclase	
NRP	*scat_p0609*	AMP-dependant synthetase	*Stackebrandtia nassauensis* DSM 44728

	scat_p0615	NRPS	*Streptomyces sp. W007*
	scat_p0616	NRPS	*Amycolatopsis mediterranei* S699
	scat_p0617	NRPS	*Amycolatopsis mediterranei* U32
	scat_p0618	AMP-dependant synthetase	*Salinispora arenicola* CNS-205
	scat_p0620	Thioesterase	*Streptomyces clavuligerus* ATCC 27064 plasmid
	scat_p0624	Antibiotic transporter	*Ralstonia eutropha* H16
	scat_p0629	NRPS	*Streptomyces venezuelae* ATCC 10712
	scat_p0634	Thioesterase	*Salinispora tropica* CNB-440
	scat_p0635	NRPS	
	scat_p0649	AMP-dependant synthetase	
β-lactam	*scat_p0835*	Methyltransferase	*Streptomyces flavogriseus* ATCC 33331
	scat_p0834	Oxidoreductase	*Saccharomonospora viridis* DSM 43017
	scat_p0833	AMP-dependant synthetase	*Streptomyces flavogriseus* ATCC 33331
	scat_p0832	β-Lactam synthetase	*Streptomyces clavuligerus* ATCC 27064
	scat_p0831	Methyltransferase	*Streptomyces griseus subsp. griseus* NBRC 13350
	scat_p0830	Methyltransferase	*Streptomyces sp. SA3* actG
	scat_p0829	Drug resistance transporter	*Streptomyces griseus subsp. griseus* NBRC 13350
	scat_p0828	LysR Family regulator	*Kitasatospora setae* KM-6054
	scat_p0827	Hydrolase	*Streptomyces cf. griseus* XylebKG-1

	scat_p0824	Enoyl-CoA hydratase	
	scat_p0823	Crotonyl-CoA reductase	
NRP/PK[3]	*scat_p1161*	SARP Family regulator	*Streptomyces hygroscopicus* ATCC 53653
	scat_p1180	NRPS	*Catenulispora acidiphila* DSM 44928
	scat_p1181	AMP-dependant synthetase	*Streptomyces bingchenggensis* BCW-1
	scat_p1184	Taurine catabolism dioxygenase	*Micromonospora lupini str. Lupac*
	scat_p1186	Phosphoesterase family	*Streptomyces griseoaurantiacus* M045
	scat_p1187	Glycoside-hydrolase family	*Thermobifida fusca* YX
	scat_p1189	Alcohol dehydrogenase	*Streptomyces violaceusniger* Tu 4113
	scat_p1190	Carbohydrate kinase	*Streptomyces griseus subsp. griseus* NBRC 13350
			Streptomyces bingchenggensis BCW-1
Siderophore	*scat_p1210*	Short-chain dehydrogenase	*Streptomyces hygroscopicus* ATCC 53653
	scat_p1211	AMP-dependant synthetase	*Catenulispora acidiphila* DSM 44928
	scat_p1212	3-oxoacyl-ACP synthase	*Streptomyces bingchenggensis* BCW-1
	scat_p1213	Malonyl CoA-acyl carrier protein	*Streptomyces violaceusniger* Tu 4113
	scat_p1214	Acyl-CoA oxidase	*Micromonospora lupini str. Lupac*
	scat_p1215	Oxidoreductase	*Streptomyces pristinaespiralis* ATCC 25486
	scat_p1216	L,D-transpeptidase catalytic domain	*Frateuria aurantia* DSM 6220

	scat_p1218 - *scat_p1221*	Iron related transport proteins	*Yersinia aldovae* ATCC 35236
	scat_p1222	Siderophore-interacting and FAD-binding domain	*Streptomyces chartreusis* NRRL 12338
	scat_p1223	Decarboxylase, pyridoxal-dependent	*Streptomyces avermitilis* MA-4680
	scat_p1224	L-lysine 6-monooxygenase	
	scat_p1225	Putative siderophore biosynthesis protein	
	scat_p1226	IucA/IucC family domains	
	scat_p1227	Drug resistance transporter, EmrB/QacA - efflux pump	
	scat_p1234	Fe^{2+} Transporter	
Putative	*scat_p1428*	Molybdopterin oxidoreductase	*Kribbella flavida* DSM 17836
	scat_p1429	Iron-sulfur binding protein	*Burkholderia pseudomallei* 668
	scat_p1430	Nitrate reductase delta subunit	*Burkholderia thailandensis* E264
	scat_p1431	Nitrate reductase gamma subunit	*Halomonas elongata* DSM 2581
	scat_p1433	Enoyl-(Acyl carrier protein) reductase	*Streptoalloteichus hindustanus* ATCC 31158
	scat_p1434	Epsilon-poly-L-lysine synthase	*Gordonia bronchialis* DSM 43247
	scat_p1435	Glycoside hydrolase	*Gordonia alkanivorans* NBRC 16433
PK	*scat_p1466*	NAD-dependent epimerase	*Streptomyces longisporoflavus* NCIMB 11426
	scat_p1465	Phytanoyl-CoA dioxygenase	*Streptomyces hygroscopicus* ATCC 53653
	scat_p1467	GDP-Mannose Dehydrogenase	*Frankia sp. EAN1pec*

	scat_p1468	Drug resistance transporter	*Streptomyces bingchenggensis* BCW-1
	scat_p1472	Alcohol dehydrogenase GroES-like domain	*Saccharomonospora xinjiangensis* XJ-54 \
	scat_p1475	3-ketoacyl-ACP synthase	*Verrucosispora maris* AB-18-032
	scat_p1476	3-Oxoacyl-[acyl-carrier-protein (ACP)] synthase III	*Streptomyces pristinaespiralis* ATCC 25486
	scat_p1478	Transketolase	
	scat_p1479	1-deoxy-D-xylulose-5-phosphate synthase	
	scat_p1480	2-oxoacid dehydrogenases acyltransferase	
	scat_p1481	Aldehyde dehydrogenase family	
	scat_p1482	Methyltransferase	
	scat_p1483	Beta-ketoacyl synthase	
	scat_p1484	Beta-ketoacyl synthase	
PK	*scat_p1656*	Drug resistance transporter	*Streptomyces hygroscopicus* ATCC 53653
	scat_p1657	Alpha/beta hydrolase family	*Saccharopolyspora erythraea* NRRL 2338
	scat_p1669	Methyltransferase	*Streptomyces griseoaurantiacus* M045
	scat_p1670	PKS	*Streptomyces bingchenggensis* BCW-1
	scat_p1671	Phosphopantetheine attachment site	*Streptomyces violaceusniger* Tu 4113
	scat_p1672	Thioesterase domain	*Streptomyces cinnamonensis*
	scat_p1676	Phosphoenolpyruvate mutase	
	scat_p1677	PrpF protein	

	scat_p1678	MmgE/PrpD family
	scat_p1679 /scat_p1680	Aconitate hydratase

[1] NRP = nonribosomal peptide
[2] PK = polyketide
[3] NRP/PK = hybrid NRP and PK