# MINING THE CRYPTIC NONRIBOSOMAL PEPTIDE SYNTHETASE SYSTEMS OF *STREPTOMYCES LAURENTII*

A Thesis
Presented To
The Academic Faculty

By

Tala Mubadda Suidan

In Partial Fulfillment
Of the Requirements for the Degree
B. S. Biochemistry with the Research Option in the
School of Chemistry and Biochemistry College of Sciences

Georgia Institute of Technology

December 2010

# MINING THE CRYPTIC NONRIBOSOMAL PEPTIDE SYNTHETASE SYSTEMS OF *STREPTOMYCES LAURENTII*

Approved by:

Dr. Wendy L. Kelly, R.Ph., Ph.D., Advisor
School of Chemistry & Biochemistry
*Georgia Institute of Technology*

Dr. Raquel L. Lieberman, Ph.D.
School of Chemistry & Biochemistry
*Georgia Institute of Technology*

Dr. Carrie Shepler, Ph.D.
School of Chemistry & Biochemistry
*Georgia Institute of Technology*

Date Approved: 15 December 2010

## ACKNOWLEDGEMENTS

I would like to thank Dr. Wendy L. Kelly and all of the Kelly Research Group members, past and present, in addition to my parents, Dr. Mubadda T. Suidan, Ph.D. and Dr. Aida El-Khazen Suidan, M.D., for all their guidance and support. This work was supported by the Focused Research Project grant from the Georgia Institute of Technology.

TABLE OF CONTENTS

Page

# LIST OF TABLES

# LIST OF FIGURES

# LIST OF FIGURES, CONTINUED

# LIST OF SYMBOLS AND ABBREVIATIONS

| | |
|---|---|
| A | adenylation |
| AMP | adenosine-5'-monophosphate |
| AMT | aminotransferase |
| ATP | adenosine-5'-triphosphate |
| bp | base pairs |
| C | condensation |
| C/E | combination epimerization and condensation |
| Cy | hetrocyclization |
| DMSO | dimethylsulfoxide |
| DNA | deoxyribonucleic acid |
| dNTP | deoxynucleoside triphosphate |
| E | epimerization |
| FMN | flavin mononucleotide |
| FPLC | fast protein liquid chromatography |
| HPLC | high performance liquid chromatography |
| IPTG | isopropyl β-D-1-thiogalactopyranoside |
| kb | kilobases |
| $Mg^{2+}$ | magnesium (II) |
| MonoOx | monooxygenase |
| MS | mass spectrometry |
| MT | methyltransferase |
| MWCO | molecular weight cut off |
| NADPH | nicotinamide adenine dinucleotide phosphate |
| NMR | nuclear magnetic resonance |
| NRPS | nonribosomal peptide synthetase |
| $OD_{600}$ | optical density at 600 nm |
| *orf* | open reading frame |
| Ox | oxidation |
| PCR | polymerase chain reaction |
| PGP | potato glucose peptone |
| PKS | polyketide synthase |
| $PP_i$ | inorganic pyrophosphate |
| PPTase | phosphopantetheinyltransferase |
| RBS | ribosomal binding site |
| Re | C-terminal reductase |
| RNA | ribonucleic acid |
| *S. laurentii* | *Streptomyces laurentii* ATCC 31255 |

## LIST OF SYMBOLS AND ABBREVIATIONS CONTINUED

| | |
|---|---|
| SDS-PAGE | sodium dodecyl sulfate-polyacrylamide gel electrophoresis |
| T | thiolation |
| *Taq* | *Thermus aquaticus* |
| TE | thioesterase |
| TSB | tryptic soy broth |
| UV-Vis | ultraviolet-visible |
| Xgal | 5-bromo-4-chloro-3-indolyl β-D-galactopyranoside |
| ε | molar absorptivity or molar extinction coefficient (L * $mol^{-1}$ * $cm^{-1}$) |
| μg | microgram |
| μM | micromolar |
| μm | micrometer |

# SUMMARY

In bacteria and fungi, nonribosomal peptide synthetase (NRPS) systems produce secondary metabolites that often possess anti-microbial, anti-fungal, anti-cancer, or other biological activities (*1*). Since microorganisms continue to develop resistance against known pharmaceutical treatments, the need for new drugs remains critical. Analyses of bacterial genomes reveal many more biosynthetic gene clusters than the number of known metabolites produced in a given species. Previous investigations of such cryptic clusters resulted in the discovery of novel secondary metabolites (*2-5*). This project evaluated two cryptic NRPS biosynthetic gene clusters encoded in the *Streptomyces laurentii* ATCC 31255 (*S. laurentii*) genome: *nrps2* and *nrps3*. Bioinformatic analyses of the *S. laurentii* genomic data allowed for the identification of these two cryptic NRPS biosynthetic gene clusters and a proposal of the core structures for the associated products. Determination of the loci of the *nrps2* and *nrps3* clusters involved the polymerase chain reaction (PCR), cloning, sequencing, and bioinformatic analysis. The bioinformatic analyses predict that the NRPS3 system produces a novel nonribosomal peptide. On the other hand, NRPS2 is predicted to produce a blue pigment, indigoidine, a metabolite previously isolated from *Streptomyces virginiae* MAFF 6014*, Streptomyces lavendulae* ATCC 11924, and *Erwinia chrysanthemi*. Thus, *S. laurentii* is established as an alternate producer of that compound (*6-8*). For NRPS2, characterizing the protein and identifying the metabolite produced also required PCR and cloning in addition to metal-chelate affinity chromatography of the heterologously-expressed protein, fast protein liquid chromatography (FPLC), *in vitro* assays, and high performance liquid chromatography (HPLC). Inducing

production of the new compound *in vivo* was attempted, and future genetic disruption studies in conjunction with genetic and chemical complementation assays will confirm that the proteins encoded within the genetic cluster are responsible for blue pigment biosynthesis. Study of NRPS2 as an additional indigoidine-producing enzyme illuminates critical residues and motifs that are responsible for the unusual chemistry performed by these enzymes. Overall, the results from both projects aid in the continued battle against microbes, cancer, and other human disease states by adding either a new source for a known metabolite or a new peptide product, which adds to nature's arsenal of biologically active compounds.

# CHAPTER 1: INTRODUCTION

The constant and exponential increase in drug-resistant microorganisms and viruses perpetuates an urgent need for new pharmaceutical agents (*9*). The past few millennia demonstrate the importance of naturally-produced anti-infective agents in medicine: among others, the Egyptians, Chinese, Indians, Greeks, Romans, and Arabs documented the use of naturally-available substances for medicinal purposes (*10*). Between 1981 and 2002, products derived from natural sources comprised 60% of new anticancer pharmaceuticals and 75% of new drugs used to treat infectious diseases (*11*). Bacteria naturally produce many of the pharmaceuticals used in the past and present including some of the β-lactam antibiotics, the nikkomycins, vancomycin, the tetracyclines, and bacitracin (*12-16*). More specifically, the actinomycetes, especially the genus *Streptomyces*, contribute a significant portion of the secondary metabolites useful in medicine. Although bacteria can produce these metabolites under specialized laboratory conditions, such as *S. laurentii* with thiostrepton, certain potentially-useful compounds are generated only under nutritional or environmental stress as a coping strategy or defense mechanism or are produced in response to inducing molecules or other specific ingredients in the medium (*17,18*). Some biosynthetic clusters have no *known* associated product. These clusters are cryptic, and their products are orphan metabolites. Manipulating bacteria to successfully deliver the compounds encoded by cryptic clusters provides additional resources for drug discovery in the fight against cancer and infectious diseases. With the continuous and drastic increase in antimicrobial-agent-resistant strains of harmful viruses, fungi,

and bacteria, finding new natural product drug leads and ways to manipulate their biosyntheses to compensate for resistance or drug-delivery issues becomes paramount.

This study predicts the backbone structure of peptides produced by two NRPS systems. For one system in this study, the enzyme has been purified, and *in vitro* assays have been conducted, but further study and protocol modification is required. Future work on these projects will result in at least one novel metabolite that may contribute to the fight against cancer or infectious diseases. Additionally, improved understanding of the enzyme machinery of these systems provides insight into how they can be applied to satisfy future and current engineering needs.

## Biosynthetic systems

Various biosynthetic systems afford the production of natural products including NRPS and polyketide synthase (PKS) systems. NRPS and PKS systems are found in many antibiotic-producing families of bacteria, including the *Streptomyces* genus of actinomycetes, and these systems yield two large classes of secondary metabolites beneficial to the battle against disease (*1,15*). Secondary metabolites produced by NRPS and PKS systems arise from modular enzymatic assembly lines (*15*). Hybrid PKS-NRPS and NRPS-PKS systems also exist, such as those found in the indanomycin, myxothiazol, or melithiazol biosynthetic clusters in *Streptomyces antibioticus* NRRL 8167, *Stigmatella aurantiaca* DW4/3-1, *Melittangium lichenicola* Me l46 respectively (*19,20*). Polyketides are built from monomers of malonyl-coenzymeA (malonyl-CoA) or one of its derivatives, but a different acyl-CoA thioester may serve as the starting substrate. On the other hand, proteinogenic and nonproteinogenic amino

acids serve as the substrates for NRPSs (*15*). Hybrid systems utilize both malonyl-CoA thioesters and amino acids according to the modules of the individual biosynthetic system. Polyketides and nonribosomal peptides may be modified following megasynthetase assembly or, in certain cases, their monomeric components may be modified pre-assembly (*15*). Typically, all genes encoding the necessary modifying enzymes are clustered next to each other on the bacterial chromosome (*15*). Research on these significant compounds requires understanding their composition and biosynthesis.

## Overview of nonribosomal peptide synthetase systems

This study focuses on NRPS systems due to the large number of cryptic NRPS clusters found in the *S. laurentii* genome (*21*). Figure 1.1 displays the modular and domain organization of a typical NRPS, which includes an initiation, extension, and termination module. In a NRPS system, one module typically exists for each monomer to be incorporated into the metabolite. Three core domains present in almost all NRPS systems are the carrier protein or the thiolation (T) domain, the amino acid adenylation (A) domain, and the condensation (C) domain (*15*). The T domain is posttranslationally modified at a conserved serine residue by a phosphopantetheinyl transferase (PPTase) to bear a phosphopantetheinyl arm that covalently binds the amino acid substrate (*15*). The A domain determines the amino acid monomer to be incorporated and tethered to the T domain by a two-step process. It first catalyzes the formation of an aminoacyl-adenylate intermediate using adenosine-5'-triphosphate (ATP) coordinated with magnesium (II) ($Mg^{2+}$) and releasing inorganic pyrophosphate ($PP_i$), and, in a second step, the A domain then attaches the amino acid substrate to the free thiol on the phosphopantetheinyl arm of the holo-T

domain, releasing adenosine-5'-monophosphate (AMP) (*15*). The C domain is the catalyst that forms the amide bond and extends the polypeptide chain (*15*). The initiation module often lacks a C domain unless another initiating acyl substrate (e.g. a lipid) is to be linked by an amide bond to the N-terminal amino acid (*15*). The termination module usually possesses a C-terminal thioesterase (TE) domain that hydrolyzes the peptidyl thioester bound to the final T domain or, in a few systems such as that of safracin or myxochelin, a reductase (Re) domain appears in place of the TE domain (*22,23*). The Re domain uses the cofactor nicotinamide adenine dinucleotide phosphate (NADPH) to reduce the peptidyl thioester at the final T domain to yield an unstable thioacetal and, ultimately, an aldehyde (*22,23*). Alternatively, a second reduction of the thioacetal can result in a hydroxyl functional group (*22,23*).

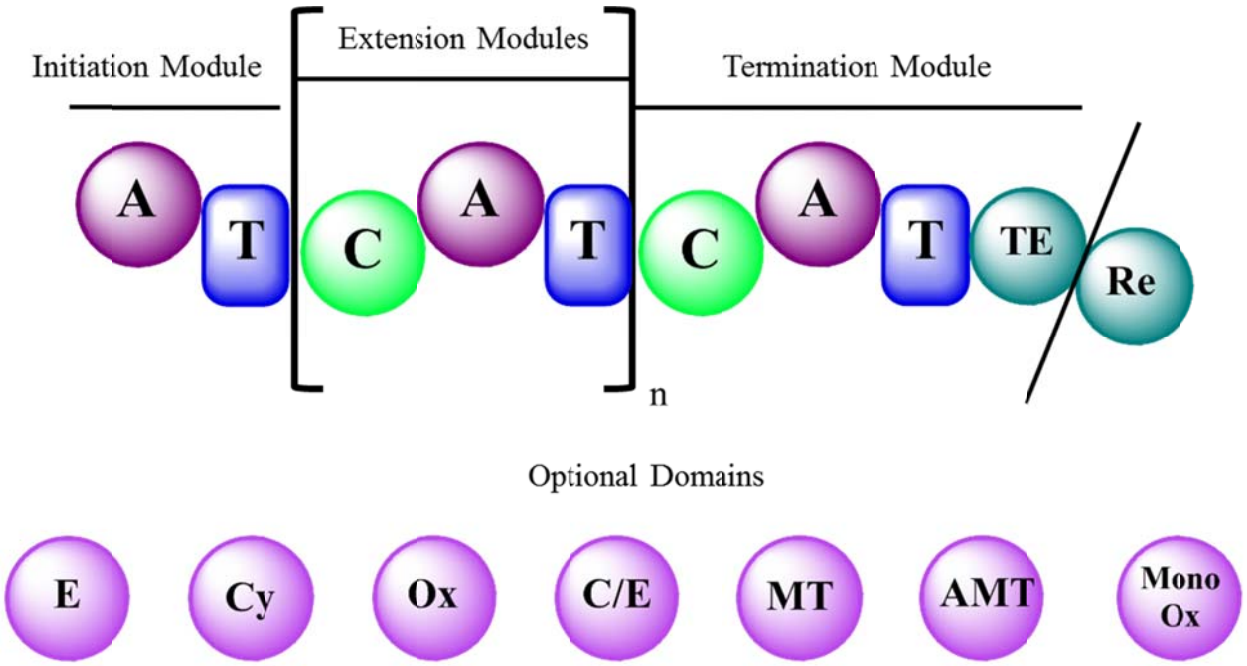



**Figure 1.1** Typical NRPS module organization showing initiation, extension, and termination modules with C, A, T, TE, and Re domains. The "n" indicates that more than one extension module is possible. Optional catalytic domains are shown at the bottom: the epimerization (E), heterocyclization (Cy), oxidase (Ox), methyltransferase (MT), dual-function condensation-epimerization (C/E), monooxygenase (MonoOx), and aminotransferase (AMT) domains (*15,20,22-32*) .

Some domains do not always appear in an NRPS module. These include epimerization (E), heterocyclization (Cy), oxidase (Ox), methyltransferase (MT), dual-function condensation-epimerization (C/E), monooxygenase (MonoOx), and aminotransferase (AMT) domains (Figure 1.1) (*15,20*). E domains convert L amino acids to their D forms, while C/E domains perform both condensation and epimerization functions (*25,32*). Cy domains catalyze amide bond formation using cysteine, serine, or threonine as the acyl acceptor followed by attack of the β-thiol or β-hydroxyl side chain upon the resulting amide carbonyl during cyclodehydration to yield a thiazoline or oxazoline ring (*26,28*). Ox domains often appear in the same module as a Cy domain to catalyze the oxidation of the heterocycle to the fully aromatic species (*29,30*). After amide bond formation, MT domains transfer a methyl group from *S*-adenosylmethionine to the amide nitrogen of the aminoacyl-*S*-T intermediate, (*27,28,33*). In hybrid NRPS-PKS systems, AMT domains transfer the α-amino group from glutamine to a β-ketoacyl group tethered to a T domain resulting in a β-aminoacyl-*S*-T intermediate as observed in the mycosubtilin PKS-NRPS system (*24*). The rarely seen MonoOx domains introduce a hydroxyl group at the α-position of the nascent peptidyl-*S*-T intermediate; this action can result in an intermediate which spontaneously decomposes into a peptide containing a carboxy-terminal amide and glyoxylate or another α-keto acid remaining bound to the T domain (*20,31*). The latter ultimately requires release by the TE domain (*20,31*). These domains occur less frequently than the C, A, and T, and TE domains, however, knowing their function remains necessary to this investigation.

**Analyzing sequences for conserved motifs and substrate specificity in NRPS systems**

Extensive analysis on the deduced amino acid sequences of the NRPSs is based on the work of Marahiel, Stachelhaus, Challis, Rausch, and Conti (*34-38*). Alignment of the amino acid sequences of multiple NRPS domains reveals conserved sequence motifs as reviewed by Marahiel (*36*). These motifs include residues necessary for protein function or structure and are displayed in Table 1.1 (*36*). Based on the crystal structure of the A domain of gramicidin S synthetase A (PheA), the A1 and A2 motifs are conserved for structural reasons but do not participate in the adenylation domain reactions (*35*). The residues of the A3 motif stabilize the pyrophosphate leaving group, while the aspartic acid residue of the A7 motif interacts via hydrogen bonding with the ribose moiety of ATP (*39-41*). The aspartate residue in the A4 motif is thought to stabilize the α-amino group of the amino acid substrate via two hydrogen bonds and proves critical for correctly positioning the substrate within the active site for ATP-dependent activation (*41-44*). In the A5 motif, the side chain carbonyl of the asparagine residue most likely forms a hydrogen bond with the amino group in the adenine of ATP based on mutational studies conducted by Saito, *et al* (*41-44*). The lysine residue in A10 coordinates with the α-carboxy group of the amino acid substrate and the ribose oxygens on the 4' and 5' carbons and is also presumed to fix the position of the amino acid and the ATP substrates (*45*). Mutation studies showed that the A8 sequence is critical for adenylate formation, and another study suggested that this sequence might participate in $Mg^{2+}$ coordination, but the full role remains unclear (*36,46-48*). The roles of the A6 and A9 motifs also remain unclear despite that the motifs still present themselves within the genes of NRPS A domains (*36,46*).

Table 1.1 Highly conserved motifs of the domains of nonribosomal peptide synthetases (*36*).

| Domain | Motif | Consensus Sequence |
|---|---|---|
| Adenylation | A1 | L(TS)YxEL |
| | A2 | LKAGxAYL(VL)P(LI)D |
| | A3 | LAYxxYTSG(ST)TGxPKG |
| | A4 | FDxS |
| | A5 | NxYGPTE |
| | A6 | GELxIxGxG(VL)ARGYL |
| | A7 | Y(RK)TGDL |
| | A8 | GRxDxQVKIRGxRIELGEIE |
| | A9 | LPxYM(IV)P |
| | A10 | NGK(VL)DR |
| Thiolation | T | DxFFxxLGG(HD)S(LI) |
| Condensation | C1 | SxAQxR(LM)(WY)xL |
| | C2 | RHExLRTxF |
| | C3 | MHHxISDG(WV)S |
| | C4 | YxD(FY)AVW |
| | C5 | (IV)GxFVNT(QL)(CA)xR |
| | C6 | (HN)QD(YV)PFE |
| | C7 | RDxSRNPL |
| Thioesterase | TE | G(HY)SxG |
| Epimerization | E1 | PIQxWF |
| | E2 | HHxISDG(WV)S |
| | E3 | DxLLxAxG |
| | E4 | EGHGRE |
| | E5 | RTVGWFTxxYP(YV)PFE |
| | E6 | PxxGxGYG |
| | E7 | FNYLG(QR) |
| *N*-Methylation | M1 | VL(DE)GxGxG |
| | M2 | NELSxYRYxAV |
| | M3 | VExSxARQxGxLD |

*residues within parentheses indicate that either one or the other may be present and "x" indicates that any residue may be present.

The variations in the A domain sequence between the A4 and A5 motifs specify the amino-acid substrate. These variations were analyzed by Stachelhaus, *et al.* and Challis *et al.* who aligned the amino-acid sequences of the region between the A4 and A5 motifs from NRPS subunits with known substrate specificities, including PheA (*34,38*). Based on the PheA crystal structure and these alignments, Stachelhaus, *et al*. identified ten amino acid residues that line the PheA active site and contact the substrate (*35,38*). These residues appear to impart substrate specificity, and Stachelhaus, *et al*. prepared a chart detailing the various sequences that correspond to each potential substrate (*38*). Rausch, *et al.* expanded on the work of Stachelhaus

*et al*. and Challis, *et al*. to create an algorithm which extracts the same ten substrate-specifying residues within 8 Å of the active site in the A domain (*34,35,37,38*). This algorithm then compares these residues with NRPS A domains of known specificity and statistically proposes a substrate for that A domain (*37,38*). Although Rausch's web-based program is a fast method to predict a NRPS module's substrate and includes a more recent database than the previous work of Stachelhaus *et al*., aligning the sequences and manually identifying the specificity-conferring residues still provides beneficial guidance to such studies.

## Genome mining and the genomisotopic approach

The preceding information provides a basic overview of NRPS system components required for this study, while the following tells of how this study came into being. The Kelly lab partially sequenced the *S. laurentii* genome in order to locate the biosynthetic gene cluster for thiostrepton, a ribosomally-produced antibiotic possessing potent activity against Gram-positive bacteria (*49*). During the course of that work, bioinformatic analyses identified NRPS homologs not associated with any known products of *S. laurentii*.

Analyzing genetic data requires prior knowledge of genetics, biochemistry, and streptomycetes. Although notable exceptions exist, DNA is transcribed into RNA and then translated into proteins. Evolution has conserved protein structure and function in related species, genera, or families – and sometimes more broadly as shown by the highly conserved sequences of histone H4 among eukaryotes or cytochrome c in bacteria and mammals (*50,51*). The wealth of genomic data accumulated in publicly accessible databases provides insight into the function of an uncharacterized protein when the sequence of that protein is compared to a database of

proteins with a known role, its homologs. In streptomycetes, an inherent instability in the end regions of linear bacterial chromosomes causes members of this genus to gain and lose DNA, and thus novel functions, through lateral gene transfer in those terminal gene regions (*52*). That gene transfer makes genomes of related species crucial to analyzing a streptomycete's genome and deducing protein function since other species may have been the source of certain genes (*52*). Additionally, knowing the conserved core region for primary metabolism for a genus assists in identifying the potential loci for secondary metabolism. Investigating cryptic clusters starts with bioinformatic analyses and comparing genomic data from the organism of interest to that of related entities, in addition to sequentially-related proteins from a variety of systems.

Both the Challis and Gross research groups pioneered methods to evaluate cryptic biosynthetic clusters. These two methods both use the same bioinformatic analyses in the early stages of evaluating a cryptic cluster (*3,5*). To start, Ishikawa's web-based FramePlot analyzes DNA sequences for open reading frames (*orf*s) in streptomycetes, taking into account the overall base composition within an *orf*, the guanosine and cytosine content of the third codon position (which averages 92% in streptomycetes), and potential ribosomal binding sites (RBSs) (*53,54*). It then deduces the corresponding amino-acid sequence and provides a link to the National Center for Biotechnology Information's Basic Local Alignment Search Tool (BLAST) to check for known similar protein sequences (*53,55*). All of these steps facilitate prediction of protein function.

Investigating any NRPS system requires a thorough bioinformatic analysis such as that described above, and the methods developed by Challis *et al.* and Gross *et al.* do not differ from each other in this respect. Following the *in silico* analysis, however, the two methods diverge. For Challis, genome mining of cryptic NRPS and PKS biosynthetic clusters involves analyzing

genomic data using the tools already described here for all genes encoding proteins within a given locus that might participate in biosynthesis, not just the NRPSs (*3*). Cloning, expression, and purification of key proteins involved in the system can lead to *in vitro* reconstitution of all or part of the system. *In vivo* expression of the cluster in the native bacterium may be achieved by varying culture conditions in different media with and without inclusion of any inducer molecules (*3*). Following *in vivo* or *in vitro* production, nuclear magnetic resonance (NMR) spectrometry, HPLC and other methods aid in purification and structure determination of the metabolite. This genome mining method relies heavily on genetic data and analyses of that data; however, those analyses give a solid foundation for subsequent techniques and assays necessary to produce and detect the orphan metabolite. The Challis group and others applied this method to produce new compounds that include terrequinone, methylenomycin, fuscachelin, and coelichelin (*56-59*).

In another approach, Gross, *et. al.* takes the bioinformatic analyses involved in genome mining, but applies that information differently in the quest to isolate and determine the product of a biosynthetic cluster (*5*). The genomisotopic approach applies the NRPS substrate predictions to propose experimental culture conditions in which the organism of interest may produce the novel compound (*5*). Predicting the function of an NRPS homolog and its anticipated substrate specificity remains necessary in this methodology (*5*). After narrowing down possible growth conditions, an isotopically-labeled substrate predicted for the NRPS may be added to the growth medium (*5*). Following the fate of the labeled isotope via NMR or mass spectrometry (MS) facilitates a fractionation scheme for product purification (*5*). This scheme consists of assays specific to the metabolite, and experimental conditions cannot always be predetermined. Gross, *et al.*, pioneered this method while identifying the biosynthetic machinery that produces

orfamide A in *Pseudomonas fluorescens* Pf-5 (*5*). The genomisotopic method builds on the foundation of the genome mining of the Challis group by following isotope-labeled substrates and basing all subsequent assays on the information obtained from those marked substrates (*3,5*). Recently, Robbel *et al.* adapted this approach to elucidate a new nonribosomally-synthesized siderophore erythrochelin from *Saccharopolyspora erythraea* by adapting this approach (*60*). This alternative method appears equal in value when compared to that of Challis. The genomisotopic approach does not attempt to reconstitute the NRPS system *in vitro* prior to establishing *in vivo* production, eliminating the difficulties of expressing and purifying large NRPS subunits. Researchers mainly use a combination of genome mining and the genomisotopic approach to analyze genomic data and to explore the secondary metabolites that a given bacterium produces.

## Genome mining in *Streptomyces laurentii*

Using a combination of the methods and strategies employed by Challis, *et al*, and Gross, *et al*, this thesis investigates two *S. laurentii* genome fragments encoding NRPS homologs that were identified during the search for the thiostrepton biosynthetic gene cluster: genome fragments 764 and 3305 (*3,5*). The loci surrounding the *S. laurentii* genome fragments 764 and 3305 were annotated as NRPS2 (Figure 1.2) and NRPS3 (Figure 1.3) respectively. Tables 1.2 and 1.3 detail the *orf* and BLAST analyses of genome fragments 764 and 3305, respectively (*53,55*). Genome fragment 764 contains one *orf* encoding an incomplete NRPS homolog and two *orf*s encoding hypothetical proteins (*55*). Genome fragment 3305 includes two partial *orf*s which encode NRPS homologs and one complete *orf* encoding a stand-alone MonoOx domain. The

MonoOx domain encoded by *orf3305b* is comparable to that of the domain observed in the subunits MelG (52% identity and 65% similarity) and MtaG (53% identity and 66% similarity) associated with the hybrid NRPS-PKS systems that produce melithiazol and myxathiazol respectively (*20,55*).

A PCR screen was used to pinpoint the biosynthetic locus for each system from the *S. laurentii* genomic fosmid library with primers specific to the NRPS-encoding genes and/or gene fragments (*49*). Once sequencing results confirmed that the identified fosmids possessed the intended sequences, one fosmid from each set was chosen for shotgun cloning and sequencing. This process was repeated as necessary to ensure that the complete locus was determined for each NRPS system. Extensive bioinformatic analyses were performed to determine all of the *orf*s within the obtained DNA sequence and to predict function based on homology and sequence identity. Once the NRPS homologs were identified, the deduced amino acid sequences for these homologs were further analyzed to determine the module and domain organization in addition to predicting the substrates and a possible core structure for each system. Following the bioinformatic analyses for the NRPS2 system, *nrps2* was cloned into a protein expression vector, and that vector was transformed into the necessary *E. coli* strains. Protein expression and purification followed in addition to *in vitro* assays.

From the results, this study predicts that NRPS3 produces a novel metabolite for which the backbone structure is proposed, but further work is required to determine that product, to activate its production *in vivo*, and to characterize its biosynthesis. Additionally, this study establishes that *S. laurentii* possesses a gene, *nrps2*, which encodes an enzyme that produces a blue pigment. Further work is necessary to determine that pigment's identity and to confirm its

production in *S. laurentii*. Future work on these projects will hopefully result in at least one novel metabolite which will contribute to the fight against cancer or infectious diseases.

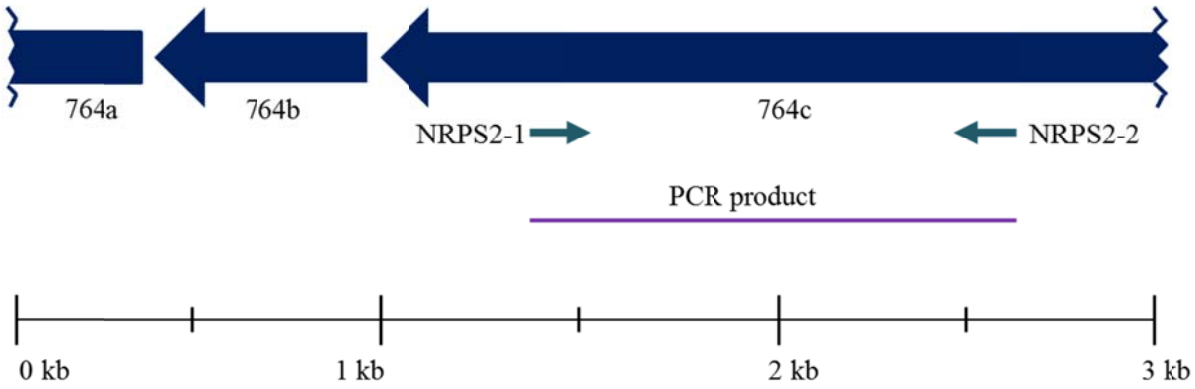



**Figure 1.2** *S. laurentii* genome fragment 764. The *orf*s and *orf* fragments are shown. The PCR primers, NRPS2-1 and NRPS2-2, used to amplify a region of *orf764c* are shown. Blue arrows indicate *orf*s; the jagged breaks indicate an incomplete *orf* that continues beyond the known sequence.

Table 1.2 Analysis of *S. laurentii* genomic DNA fragment 764 (2928 base pairs)

| Gene | Strand | Position | Size (aa) | Homologs (aa) | Species | % ID | % Similarity |
|---|---|---|---|---|---|---|---|
| 764a | - | 1-304 (i) | 101 | Transcriptional regulator, ArsR family (201 aa) | *Geobacillus* sp. Y412MC10 | 39 | 65 |
| | | | | Transcriptional regulator (106 aa) | *Bacillus cereus* BGSC 6E1 | 35 | 51 |
| 764b | - | 347-760 | 137 | MarR family transcriptional regulator (171 aa) | *Streptomyces avermitilis MA-4680* | 31 | 43 |
| 764c | - | 794-2928 (i) | 710 | Blue-pigment synthetase (1282 aa) | *Streptomyces lavendulae* subsp. *lavendulae* | 58 | 71 |
| | | | | Blue-pigment synthetase (1353 aa) | *Streptomyces clavuligerus* ATCC 27064 | 59 | 70 |
| | | | | Indigoidine synthase (1488 aa) | *Erwinia chrysanthemi* | 53 | 68 |

"i" indicates an incomplete *orf*, "aa" refers to the length of the polypeptide in amino acids; for strand, "+" indicates the direct strand, while "-" indicates the complementary strand. % ID represents the percent of residues identical to the homolog while % Sim indicates the number of similar or identical residues in the same position as the homolog.

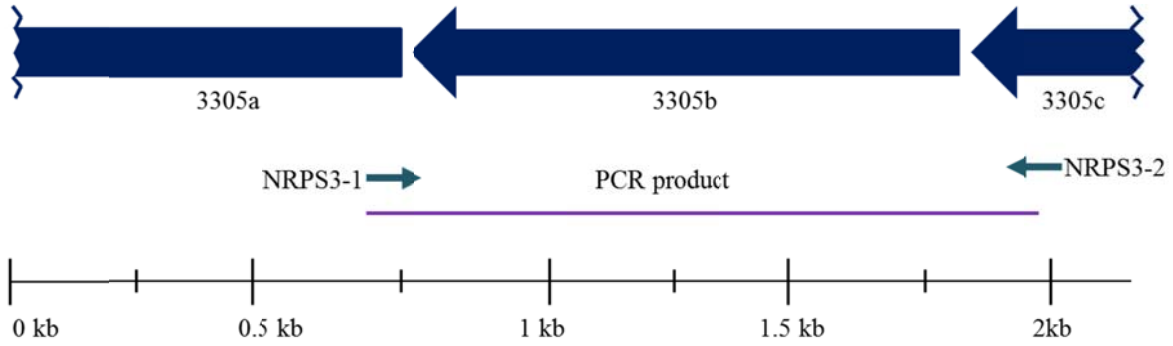



**Figure 1.3** *S. laurentii* genome fragment 3305. The *orf* and *orf* fragments are shown. The PCR primers, NRPS3-2 and NRPS3-2, used to amplify *orf3305b* are shown. Blue arrows indicate *orf*s; the jagged breaks indicate an incomplete *orf* that continues beyond the known sequence.

Table 1.3 Analysis of *S. laurentii* genomic DNA fragment 3305 (2222 base pairs)

| Gene | Strand | Position | Size (aa) | Homologs (aa) | Species | % ID | % Similarity |
|---|---|---|---|---|---|---|---|
| 3305a | - | 1-755 (i) | 250 | Nonribosomal peptide synthetase (8986 aa) | *Streptomyces fungicidicus* | 53 | 63 |
| | | | | Nonribosomal peptide synthetase (6943 aa) | *Streptomyces fungicidicus* | 51 | 60 |
| 3305b | - | 785-1834 | 349 | Non-ribosomal peptide synthetase (3445 aa) | *Sorangium cellulosum* 'So ce 56 | 56 | 66 |
| | | | | MtaG (1750 aa) | *Stigmatella aurantiaca* DW4/3-1 | 53 | 66 |
| | | | | MelG (1747 aa) | *Melittangium lichenicola* | 52 | 65 |
| 3305c | - | 1898-2222 (i) | 107 | Amino acid adenylation domain-containing protein (2386 aa) | *Bacillus cereus* subsp. *cytotoxis* NVH 391-98 | 44 | 62 |
| | | | | DhbF (2385 aa) | *Bacillus licheniformis* ATCC 14580 | 48 | 60 |

"i" indicates an incomplete *orf*, "aa" refers to the length of the polypeptide in amino acids; for strand, "+" indicates the direct strand, while "-" indicates the complementary strand. % ID represents the percent of residues identical to the homolog while % Sim indicates the number of similar or identical residues in the same position as the homolog.

# CHAPTER 2: EXPERIMENTAL METHODS

## General

All chemicals and solvents were reagent grade, purchased from VWR, Fisher, or Sigma-Aldrich, and used without further purification unless otherwise indicated. Other reagents and kits for molecular biology techniques were purchased from NEB, Bio-Rad, Novagen, Qiagen, or Invitrogen. The Vector NTI application suite was purchased from Invitrogen/Life Technologies. All FASTA files and protein sequence descriptions were obtained from National Center for Biotechnology Information (NCBI), and protein-sequence alignments were performed using ClustalW (*61,62*). Individual sequencing reactions were performed by Eurofins MWG Operon (Huntsville, AL), and sequencing of the shotgun fosmid libraries was performed at Functional Biosciences, Inc. (Madison, WI). Protein-coding regions of DNA sequences were predicted using FramePlot 4.0Beta (http://nocardia.nih.go.jp/fp4/) (*53*). Protein sequences were analyzed by BLAST (Basic Local Alignment Search Tool) (*55*).

Polymerase chain reactions (PCRs) were performed using a Robocyler Gradient 96 (Stratagene). Agarose gels were run in a Labnet Gel $X_L$ Ultra V-2. Gels were visualized using an Alpha Innotech Corporation Imager. Beckman-Coulter Microfuge 18, Allegra X-15R, or Microfuge 22R centrifuges were used. *E. coli* was cultivated in a Barnstead Lab-line L-C incubator for solid media or, for liquid media, in a New Brunswick Scientific Innova 4230 refrigerated incubator shaker or in a New Brunswick Innova 44 incubator shaker series. Fast Protein Liquid Chromatography (FPLC) was performed using an Amersham Biosciences ÄKTA

FPLC system with a UV-900, P900, pH/C-900, Box 900, and Fraction 900. A GE Health Sciences Mono Q 5/50 GL column was used for anion exchange chromatography. For ultraviolet-visible (UV-Vis) spectroscopy, a Varian Cary 50 UV-Visible spectrophotometer was used. High-performance liquid chromatography (HPLC) was performed on a Beckman Coulter System Gold equipped with a diode-array detector.

Fosmid and plasmid DNA was purified following the Qiagen's Plasmid Spin Miniprep Kit protocol. When a high concentration fosmid solution was desired, twice the recommended volume of the DNA-containing supernatant was applied to each Qiagen spin-column. The concentrations and purity of the isolated plasmid or fosmid DNA was assessed by UV-visible spectroscopy (*63*).

All DNA primers were purchased from Integrated DNA Technologies. Table 2.1 lists the primers used in this study with the exception of primers used for sequence assembly which are detailed in Table 2.2.

## Bacterial strains, plasmids, and media

The strains and plasmids used in this work are provided in Table 2.3. *S. laurentii* was obtained from American Type Culture Collection (ATCC). All *Escherichia coli* (*E. coli*) strains were grown in Luria-Bertani liquid medium or on solid medium supplemented with the appropriate antibiotic(s). Kanamycin (50 $\mu$g/mL) and chloramphenicol (12.5 $\mu$g/mL) were used for the selective growth of *E. coli*. Tryptic Soy Broth (TSB) and Y6.5 were prepared as described (*8,64*). Y7.2 was prepared as for Y6.5 but at pH 7.2 (*8*). ISP-2, ISP-4, oatmeal, potato glucose

peptone (PGP), and Sabourad dextrose agar media were prepared as previously described (*65-68*). Y6.5 agar was prepared by adding 18 g/L of Bacto agar to Y6.5 medium, and used for the growth and sporulation of *S. laurentii* (*8*).

Table 2.1 Primers used in this study

| Primer Name | Primer Sequence | Purpose |
|---|---|---|
| NRPS2-1 | 5' - CCCCGCCTCCTCCAGGAG - 3' | Primers for screening the genomic library of *S. laurentii* for the NRPS2 biosynthetic gene cluster |
| NRPS2-2 | 5' - CGAGGAGATCCTGCCGCC - 3' | |
| NRPS2-3 | 5'- GGCATATGCCGGGCGCGCGTCGC -3' | Primers for amplification of *nrps2* for heterologous protein expression. The *Nde*I and *Xho*I restriction sites are underlined. |
| NRPS2-4 | 5'- CTCGAGCTACCTCCCCAGCCGGTAGTG-3' | |
| NRPS2-5 | 5'- CTCGAGCCTCCCCAGCCGGTAGTGGAT-3' | |
| NRPS2-V1 | 5' - CGATCAGGAACAGGTTCTCC - 3' | Primers for the confirmation of *nrps2* |
| NRPS2-V2 | 5' - CTCTCCAGAGGCTCCAACTC - 3' | |
| NRPS2-V3 | 5' - TGTAGCCGGAGGACATGAAC - 3' | |
| NRPS2-V4 | 5' - AAGGAGTCCAAGCTCCAGGT - 3' | |
| NRPS2-V5 | 5' - ACCTGGAGCTTGGACTCCTT - 3' | |
| NRPS2-V6 | 5' - TCGCCTACGTCATCTACACG - 3' | |
| NRPS2-V7 | 5' - CCGGACGTGTAGATGACGTA - 3' | |
| NRPS2-V8 | 5' - CTGTTCGTCGCTGATCCTG - 3' | |
| NRPS2-V9 | 5' - CGTCGTAGAAGCGGTAGGTC - 3' | |
| NRPS2-V10 | 5' - GGCGAGGATCTCCCACTG - 3' | |
| NRPS3-1 | 5' - AAAACGACAGAGGGATCACG - 3' | Primers for screening the genomic library of *S. laurentii* for the NRPS3 biosynthetic gene cluster |
| NRPS3-2 | 5' - GGGTCGACGACAACTTCTTC - 3' | |
| NRPS3-3 | 5' - ATGAGAGGGAAAGGCGATCT - 3' | Primers for screening the genomic library of *S. laurentii* to extend the NRPS3 locus upstream of forward end sequence |
| NRPS3-4 | 5' - GACGATGTCGAACTCCTCGT - 3' | |
| NRPS3-5 | 5' - GATTCTCGAACGCTTTCTGC - 3' | Primers for screening the genomic library of *S. laurentii* to extend the NRPS3 locus downstream of reverse end sequence |
| NRPS3-6 | 5' - GATCAGACTGACCCGTTCGT - 3' | |

Table 2.2 Sequencing primers for sequence assembly of fosmids JA18C3, JA4B8, JA9D12, JA5F6, & JA15F8

| Fosmid | Primer | Primer Sequence |
|---|---|---|
| JA18C3 | 18C3-1 | 5'- GGATCACTGGACCCTGTACG - 3' |
| JA18C3 | 18C3-2 | 5' - GCGTTCCAGTTCGTCGAG - 3' |
| JA18C3 | 18C3-3 | 5' - CCGAAGAGTAAGTCCCACCA - 3' |
| JA18C3 | 18C3-4 | 5' - GCCTGATGCGGTATTTTCTC - 3' |
| JA18C3 | 18C3-5 | 5' - CGTCTGGGACCACTGACC - 3' |
| JA18C3 | 18C3-6 | 5' - TGATTTCATCGAGCATGAGG - 3' |
| JA18C3 | 18C3-7 | 5' - CTGGCGTAATAGCGAAGAGG - 3' |
| JA18C3 | 18C3-8 | 5' - CTCCTTGGGCTCCCACTC - 3' |
| JA18C3 | 18C3-9 | 5' - CCGAAGAGTAAGTCCCACCA - 3' |
| JA18C3 | 18C3-10 | 5' - GTGACGTCGTCGCACACC - 3' |
| JA18C3 | 18C3-11 | 5' - TGATTTCATCGAGCATGAGG - 3' |
| JA4B8 | 4B8-1 | 5' - GCTCCTCCAGGTTCTCCTTC - 3' |
| JA4B8 | 4B8-2 | 5' - ATGAAGACCGTCGGTGAACT - 3' |
| JA9D12 | 9D12-1 | 5' - CCCGTACACATTGGTGACCT - 3' |
| JA9D12 | 9D12-2 | 5' - AGCTTCGACGCGTTCTTCT - 3' |
| JA9D12 | 9D12-3 | 5' - CAGTTCGCCCGAAGTGTT -3' |
| JA9D12 | 9D12-4 | 5' - GTACGACGCTGGAGACGTG - 3' |
| JA9D12 | 9D12-5 | 5' - GAGGATCTCGGCGAACAG - 3' |
| JA9D12 | 9D12-6 | 5' - CCACCTGGAACTCCTCCTG - 3' |
| JA9D12 | 9D12-7 | 5' - GATGGCCAGCAGGGAATG - 3' |
| JA9D12 | 9D12-8 | 5' - CACGTCAACATGCCCTTC - 3' |
| JA9D12 | 9D12-9 | 5' - GACTACATGGTTCCCGCTGT - 3' |
| JA9D12 | 9D12-10 | 5' - CGAGCTCCGCGTAACTGAG - 3' |
| JA9D12 | 9D12-11 | 5' - GTACGACGCTGGAGACGTG - 3' |
| JA9D12 | 9D12-12 | 5' - CATGTACCCGGGGAGGTA - 3' |
| JA5F6 | 5F6-1 | 5' - ACATGGACGGCGCTGTAG - 3' |
| JA5F6 | 5F6-2 | 5' - CAGTCGTGGATCTGGAGCTG - 3' |
| JA5F6 | 5F6-3 | 5' - TCGGTTCAGTGGTGTACGAG - 3' |
| JA5F6 | 5F6-4 | 5' - GGTCACCGTCTTCCTCGTC - 3' |
| JA5F6 | 5F6-5 | 5' - CCCCTACAGGCCTACGAGTC - 3' |
| JA5F6 | 5F6-6 | 5' - CGACACCGTGTTCGTGAC - 3' |
| JA5F6 | 5F6-7 | 5' - GAGCAGACACCCGTCGAG - 3' |
| JA5F6 | 5F6-8 | 5' - GTCCGTACGGGAGGAACG - 3' |
| JA5F6 | 5F6-9 | 5' - GACGGAGATCTGTACGAGCA - 3' |
| JA5F6 | 5F6-10 | 5' - GTAGCGGGTTCCGGTCAC - 3' |
| JA15F8 | 15F8-1 | 5' - GTTCGCGTACACCACCTACC - 3' |
| JA15F8 | 15F8-2 | 5' - CCAGGGCGAGTTCGTACC - 3' |
| JA15F8 | 15F8-3 | 5' - ACTACCGTGCACACATGCTC - 3' |
| JA15F8 | 15F8-4 | 5' - GGTGAGGCTGATCTGGAAGA - 3' |
| JA15F8 | 15F8-5 | 5' - CGAACTCGACCACCTGGA - 3' |
| JA15F8 | 15F8-6 | 5' - TGGGTGAAGAACTCCAGGAT - 3' |
| JA15F8 | 15F8-7 | 5' - GATCAAGGACCGAACGTGAC - 3' |
| JA15F8 | 15F8-8 | 5' - CCGAGGTGTACCTGGGACT - 3' |
| JA15F8 | 15F8-9 | 5' - CAGCCGGAGACAGTCCAC - 3' |
| JA15F8 | 15F8-10 | 5' - ACCCAGAACTCCTGGATGG - 3' |
| JA15F8 | 15F8-11 | 5' - AAGGCGGCTTACGAGGAC - 3' |
| JA15F8 | 15F8-12 | 5' - GTGCACGACGGCTTCTTC - 3' |
| JA15F8 | 15F8-13 | 5' - TTCCGTCCTTCATGTGAACC - 3' |
| JA15F8 | 15F8-14 | 5' - AACAGGTCGGTGGCGTACT - 3' |
| JA15F8 | 15F8-15 | 5' - TGAAGAACTCCAGGATGTGG - 3' |
| JA15F8 | 15F8-16 | 5' - CTCCCCACATGGAATGTCTT - 3' |
| JA15F8 | 15F8-17 | 5' - CCGAACGGCAAGATCGAC - 3' |
| JA15F8 | 15F8-18 | 5' - CAGTTCATCGGCGTTTCC - 3' |
| JA15F8 | 15F8-19 | 5' - CCCTTCTGGTGGAGATTGTC - 3' |
| JA15F8 | 15F8-20 | 5' - GGATTCCTGGTCGTCCTGT - 3' |
| JA15F8 | 15F8-21 | 5' - CCCTTCCTGTTGGTACGAGA - 3' |
| JA15F8 | 15F8-22 | 5' - CAGGAAGAGGACCGAGGAC - 3' |
| JA15F8 | 15F8-23 | 5' - ATGCCTCGTTCATCAGTGC - 3' |
| JA15F8 | 15F8-24 | 5' - ACGCTGACCGACAAGGAC - 3' |
| JA15F8 | 15F8-25 | 5' - GCGTTTTCGGAGGGTACTAA - 3' |
| JA15F8 | 15F8-26 | 5' - GACGAGACCGATGAAGAACG - 3' |
| JA15F8 | 15F8-27 | 5' - ACCAGCCCTGTGAACGTC - 3' |
| JA15F8 | 15F8-28 | 5' - CTCGGAGGGTTTGGGAAC - 3' |
| JA15F8 | 15F8-29 | 5' - GTCCTCCTTGGCCTTGTTC - 3' |
| JA15F8 | 15F8-30 | 5' - CTTCTGGGCCTGTTCGAC - 3' |
| JA15F8 | 15F8-31 | 5' - GCGCTAATCGGACAAAACTC - 3' |

Table 2.3 Strains and plasmids used in this study

| Strain/Plasmid | Description | Reference or Source |
|---|---|---|
| *Streptomyces* strain | | |
| *S. laurentii* ATCC 31255 | Wild-type, thiostrepton producer | ATCC |
| *E. coli* strains | | |
| MACH1 | Host for DNA cloning and manipulation | Invitrogen |
| BL21(DE3) | Host for protein expression | Novagen |
| Plasmids | | |
| pSC-B/pSC-A | Plasmid for cloning and sequencing | Agilent (*69,70*) |
| pSC-B-*nrps2*-N4 | Plasmid containing *nrps2* with engineered restriction sites for N-terminal hexahistidine tag | This Study |
| pSC-B-*nrps2*-C7 | Plasmid containing *nrps2* with engineered restriction sites for C-terminal hexahistidine tag | This Study |
| JA4B8 | Fosmid containing partial NRPS2 cluster | (*49*) |
| JA5F6 | Fosmid containing partial NRPS3 cluster | (*49*) |
| JA9D12 | Fosmid containing partial NRPS3 cluster | (*49*) |
| JA15F8 | Fosmid containing partial NRPS3 cluster | (*49*) |
| JA18C3 | Fosmid containing partial NRPS2 cluster | (*49*) |
| pET28b(+) | Protein expression vector for N-terminal hexahistidine fusion protein | Novagen |
| pET29b(+) | Protein expression vector for C-terminal hexahistidine fusion protein | Novagen |
| pET28b(+)-*nrps2* | Protein expression vector for N-terminal hexahistidine fusion protein of NRPS2 | This study |
| pET29b(+)-*nrps2* | Protein expression vector for C-terminal hexahistidine fusion protein of NRPS2 | This study |

## DNA sequence analysis

The DNA sequences obtained from pyrosequencing and shotgun sub-cloning and sequencing were pasted into the FramePlot 4.0 query screen to identify potential open reading frames (*53*). The following parameters were used during analysis: a window size of forty codons, a step size of one codon, a minimum *orf* length of twenty codons, a resolution of one bp/pixel,

showing incomplete *orf*s, a RBS of GACGAACGCTGGCGG, and possible start codons ATG (methionine), GTG (valine), and TTG (leucine). An NCBI BLAST protein search identified homologs of the encoded protein (*55*). Results were tabulated. Deduced protein sequences were aligned with homologs using ClustalW (*61,62*). Further bioinformatic analysis was conducted as described in Chapter 1.

## Primer design & testing

Primer3 v4.0 was used to design the primers for the PCR screen of the *S. laurentii* fosmid library, NRPS2 and NRPS3 gene cluster sequencing, and the sequencing primers required for the fosmid sequence assembly (Tables 2.1 and 2.2) (*49,71*). For PCR, each primer was prepared at a 5 μM working solution in 10 mM Tris-HCl (pH 8.5), and the annealing temperature for each set was optimized using genomic *S. laurentii* DNA as a positive control and pCC1FOS, the empty parent fosmid, as the negative control.

Each 25 μL reaction contained 10 ng template DNA, 20 nmol deoxynucleoside triphosphates (dNTPs), 5 pmol each of a forward primer and a reverse primer, and 1.25 U *Taq* polymerase in 5% dimethylsulfoxide (DMSO) and 10 mM Tris-HCl, 50 mM KCl, 3 mM $MgCl_2$, pH 8.3. These conditions were used for screening the fosmid library and any other PCR where the fidelity of the amplified sequence was less crucial. Elongation times and annealing temperatures varied according to the expected product size and primers used.

## Screening for selected sequences in the *S. laurentii* genomic DNA fosmid library

The *S. laurentii* fosmid library was screened by PCR for the desired sequence. Four rows from each 96-well plate of the fosmid library (48 clones) were combined and the fosmid pool isolated. The four primer pairs to screen the library were NRPS2-1/NRPS2-2, NRPS3-1/NRPS3-2, NRPS3-3/NRPS3-4, and NRPS3-5/NRPS3-6 (Table 2.1). Individual rows (containing a pool of 12 clones) from each plate and individual clones were identified by iterations of PCR. The PCR products were cloned into the pSC-A vector and confirmed by sequence analysis.

For each identified fosmid, a plasmid sublibrary was prepared using Invitrogen's TOPO® Shotgun Subcloning Kit, and the sublibrary was submitted to Functional Biosciences for sequence analysis. The Contig Express application (VectorNTI suite) was used to assess sequence quality from the chromatograms as described by Luckey, *et al.* and to assemble the sequence of the genome fragment cloned into each fosmid (*72*). When necessary, additional sequencing primers were designed to close any gaps in the assembled sequence and generate one continuous sequence. These primers were prepared at 10 μM concentrations and sent to Eurofins MWG Operon with highly-concentrated fosmid DNA preparations.

## Expression and purification of NRPS2

The gene encoding NRPS2 was amplified by PCR using primer pairs NRPS2-3/NRPS2-4 and NRPS2-3/NRPS2-5 (Table 2.1) for the N-terminal and C-terminal hexahistidine fusion

protein constructs respectively. The product was cloned into pSC-B, confirmed by sequence analysis, digested with *Nde*I and *Xho*I, and ligated into appropriately digested pET28b(+) or pET29b(+) to provide pET28b(+)-*nrps2* and pET29b(+)-*nrps2*. The resulting expression vectors were confirmed by sequence analysis.

*Pilot scale expression study.* To assess the optimal temperature and IPTG concentration for NRPS2 production, an expression study was conducted in the following manner (outlined in Table 2.4). From glycerol stocks of *E. coli* BL21(DE3) harboring an expression vector, overnight cultures of pET28b(+), pET29b(+), pET28b(+)-*nrps2*, and pET29b(+)-*nrps2* were grown separately in 3 mL LB broth supplemented with 50 μg/mL kanamycin at 37 °C. From each culture, 100 μL was used to inoculate 13 x 10 mL aliquots of LB broth supplemented with 50 μg/mL kanamycin. Cultures were grown until the optical density at 600 nm ($OD_{600}$) was approximately 0.5 A.U. to 0.6 A.U. At this time, cultures were moved to the appropriate shaking incubator to equilibrate the temperature for 10 min before the appropriate amount of isopropyl-*β*-D-thiogalactopyranoside (IPTG) was added to induce protein expression. After the suitable induction period, cells were pelleted at 1439*xg* for 10 min, the supernatant removed, and the harvested cells were stored at -86 °C until ready for use. Harvested cells were resuspended in 1 mL of lysis buffer [20 mM Tris (pH 8.0), 500 mM NaCl, 2 mM imidazole, 10% glycerol, 1 mg/mL lysozyme, 10 μg/mL deoxyribonuclease I, 2 mM $MgCl_2$] and incubated at 4 °C for 30 min. The cells were disrupted by sonication (8x10-s pulses with a 2 min pause between each pulse). A 100 μL aliquot of the crude lysate was retained, mixed in a 1:1 ratio with 2X SDS-PAGE loading buffer, and stored at -20 °C until needed for SDS-PAGE (*63*). The lysate was clarified via centrifugation at 4 °C; a 100 μL aliquot was retained for SDS-PAGE under the same conditions as the crude lysate. To the cell-free extract, 100 μL of Ni-NTA resin was added, and

the slurry was incubated at 4 °C for 30 min. The resin was pelleted at 640*xg* for 1 min, and the supernatant was discarded. A 500 µL wash with lysis buffer was added, and the resin was pelleted as above, and the supernatant was discarded again. The resin was resuspended in 100 µL of elution buffer [20 mM Tris (pH 8.0), 500 mM NaCl, 300 mM imidazole, 10% glycerol, 2 mM $MgCl_2$], pelleted as above, and the supernatant was retained for SDS-PAGE. Each sample was analyzed by SDS-PAGE.

Table 2.4 Expression Study to determine optimum protein expression conditions

| C-terminal hexahistidine construct | | | | |
|---|---|---|---|---|
| Culture | [IPTG] (mM) | Temperature after induction (°C) | Growth time after induction (hours) | Vector in *E. coli* BL21(DE3) |
| 1 ($OD_{600}$ control) | N/A | N/A | N/A | pET29b(+) |
| 2 ($OD_{600}$ control) | N/A | N/A | N/A | pET29b(+)-*nrps2* |
| 3 | 0 | 37 | 3 | pET29b(+) |
| 4 | 0.4 | 37 | 3 | pET29b(+) |
| 5 | 0.04 | 37 | 3 | pET29b(+) |
| 6 | 0 | 37 | 3 | pET29b(+)-*nrps2* |
| 7 | 0.4 | 37 | 3 | pET29b(+)-*nrps2* |
| 8 | 0.04 | 37 | 3 | pET29b(+)-*nrps2* |
| 9 | 0 | 28 | 6 | pET29b(+) |
| 10 | 0.4 | 28 | 6 | pET29b(+) |
| 11 | 0.04 | 28 | 6 | pET29b(+) |
| 12 | 0 | 28 | 6 | pET29b(+)-*nrps2* |
| 13 | 0.4 | 28 | 6 | pET29b(+)-*nrps2* |
| 14 | 0.04 | 28 | 6 | pET29b(+)-*nrps2* |
| 15 | 0 | 15 | 24 | pET29b(+) |
| 16 | 0.4 | 15 | 24 | pET29b(+) |
| 17 | 0.04 | 15 | 24 | pET29b(+) |
| 18 | 0 | 15 | 24 | pET29b(+)-*nrps2* |
| 19 | 0.4 | 15 | 24 | pET29b(+)-*nrps2* |
| 20 | 0.04 | 15 | 24 | pET29b(+)-*nrps2* |
| 21 | 0 | 15 | 48 | pET29b(+) |
| 22 | 0.4 | 15 | 48 | pET29b(+) |
| 23 | 0.04 | 15 | 48 | pET29b(+) |
| 24 | 0 | 15 | 48 | pET29b(+)-*nrps2* |
| 25 | 0.4 | 15 | 48 | pET29b(+)-*nrps2* |
| 26 | 0.04 | 15 | 48 | pET29b(+)-*nrps2* |

Table 2.4, continued

| N-terminal hexahistidine construct | | | | |
|---|---|---|---|---|
| Culture | [IPTG] (mM) | Temperature after induction (°C) | Growth time after induction (hours) | Vector in *E. coli* BL21(DE3) |
| 1 ($OD_{600}$ control) | N/A | N/A | N/A | pET28b(+) |
| 2 ($OD_{600}$ control) | N/A | N/A | N/A | pET28b(+)-*nrps2* |
| 3 | 0 | 37 | 3 | pET28b(+) |
| 4 | 0.4 | 37 | 3 | pET28b(+) |
| 5 | 0.04 | 37 | 3 | pET28b(+) |
| 6 | 0 | 37 | 3 | pET28b(+)-*nrps2* |
| 7 | 0.4 | 37 | 3 | pET28b(+)-*nrps2* |
| 8 | 0.04 | 37 | 3 | pET28b(+)-*nrps2* |
| 9 | 0 | 28 | 6 | pET28b(+) |
| 10 | 0.4 | 28 | 6 | pET28b(+) |
| 11 | 0.04 | 28 | 6 | pET28b(+) |
| 12 | 0 | 28 | 6 | pET28b(+)-*nrps2* |
| 13 | 0.4 | 28 | 6 | pET28b(+)-*nrps2* |
| 14 | 0.04 | 28 | 6 | pET28b(+)-*nrps2* |
| 15 | 0 | 15 | 24 | pET28b(+) |
| 16 | 0.4 | 15 | 24 | pET28b(+) |
| 17 | 0.04 | 15 | 24 | pET28b(+) |
| 18 | 0 | 15 | 24 | pET28b(+)-*nrps2* |
| 19 | 0.4 | 15 | 24 | pET28b(+)-*nrps2* |
| 20 | 0.04 | 15 | 24 | pET28b(+)-*nrps2* |
| 21 | 0 | 15 | 48 | pET28b(+) |
| 22 | 0.4 | 15 | 48 | pET28b(+) |
| 23 | 0.04 | 15 | 48 | pET28b(+) |
| 24 | 0 | 15 | 48 | pET28b(+)-*nrps2* |
| 25 | 0.4 | 15 | 48 | pET28b(+)-*nrps2* |
| 26 | 0.04 | 15 | 48 | pET28b(+)-*nrps2* |

The N-terminal hexahistidine fusion protein construct was chosen for large-scale expression. *E. coli* BL21(DE3) containing pET28b(+)-*nrps2* was incubated at 37 °C overnight in 1 × 50 mL Luria-Bertani medium supplemented with 50 $\mu$g/mL kanamycin. The following morning, 4 × 10 mL and 1 × 5 mL of the overnight culture was used to inoculate 4 × 1 L and

1 x 500 mL of Luria-Bertani medium supplemented with 50 $\mu$g/mL kanamycin, and the resulting cultures were grown at 37 °C until $OD_{600}$= 0.4. At this time, the temperature was reduced to 15 °C and protein expression induced with the addition of 0.04 mM isopropyl-$\beta$-D-thiogalactopyranoside (IPTG). Cultures were incubated for an additional 24 h. Harvested cells were resuspended in 30 mL of lysis buffer [20 mM Tris (pH 8.0), 500 mM NaCl, 2 mM imidazole, 10% glycerol, 1 mg/mL lysozyme, 10 µg/mL deoxyribonuclease I, 2 mM $MgCl_2$]. The cells were disrupted by sonication (8x10-s pulses with a 2 min pause between pulses). The lysate was clarified by centrifugation (30000*xg*, 4 °C, 30 min). The cell-free extract was incubated with 4 mL of Ni-NTA resin (Qiagen) for 1.5 h. The slurry was loaded onto a column, and the resin was washed with 20 mL lysis buffer and then with 25 mL wash buffer [20 mM Tris (pH 8.0), 500 mM NaCl, 20 mM imidazole, 10% glycerol, 2 mM $MgCl_2$]. Protein was eluted with a step gradient of 6 mL of elution buffers 1 to 6 [wash buffer containing 40, 60, 80, 100, 150, and 300 mM imidazole], collecting 6 mL fractions. Fractions containing the N-terminal hexahistidine tagged NRPS2 protein were pooled together and dialyzed against 2 × 1 L low salt buffer [50 mM Tris (pH 7.5), 50 mM NaCl, 10% glycerol, 2 mM $MgCl_2$, and 2 mM DTT]. The protein was concentrated with a Millipore-100 kDa molecular weight cut off (MWCO) centrifugal filter, and loaded on to a GE Health Sciences Mono Q 5/50 GL column, equilibrating at low salt buffer conditions. The protein was eluted using a gradient of 50-500 mM NaCl over 60 column volumes. Fractions containing NRPS2 were pooled together and dialyzed against 2 × 1 L storage buffer [50 mM Tris (pH 8.0), 300 mM NaCl, 10% glycerol, 2 mM $MgCl_2$, and 2 mM DTT] and concentrated with a Millipore-100 KDa MWCO centrifugal filter. Protein concentration was determined by the method of Bradford, using bovine serum albumin as a standard (*73*). The protein was flash-frozen in liquid nitrogen and stored at -86 °C. Flavin

mononucleotide (FMN) concentration and fractional loading of NRPS2 by FMN were determined as previously described (*74,75*).

## Blue pigment production in *S. laurentii*

For liquid media, 10 µL of a *S. laurentii* spore suspension was used to inoculate 10 mL aliquots of three different media to which a proposed inducer, γ-nonalactone, was added: TSB, Y6.5, and Y7.2 (*8,49,64*). The inducer was diluted 10-fold with ethanol and to each culture, 3 µL was added for a final concentration of 30 µg/mL γ-nonalactone and 0.027% ethanol. Separate cultures were prepared to which γ-nonalactone was added at 0, 1, 2, 3, 4, 5, 6, 7, 8, 12, 16, 20, 24, 32, 40, 48, 56, 64, 72, 80, 88, and 96 h. Cultures were incubated at 220 rpm at 28 °C for 120 h. The cells were pelleted at 1439*xg* for 10 min, and the supernatant was centrifuged with acetone in a 1:2 supernatant:acetone ratio to precipitate any blue pigment which might have been produced. The pellets were resuspended in sterile deionized water and stored at -20 °C.

For solid media, 10 µL of a *S. laurentii* spore suspension was used to inoculate 10 mL of TSB which was grown for 48 h at 28 °C. For each media type, 3 plates were spread with 100 µL of the liquid culture, and a confluent plaque was allowed to develop. Pictures were taken at 0 and 18 h after which pictures were taken every 24 h. The media used are as follows: ISP2, ISP4, oatmeal, PGP, Saboraud dextrose, and Y6.5 agar.

## NRPS2 *in vitro* blue pigment production

To assess the effect of temperature on the reaction catalyzed by NRPS2, reactions were performed at three temperatures: 15 °C, 25 °C, and 30 °C. To phosphopantetheinylate any apo-NRPS2, 88.0 µL reactions containing 0.5 nmol NRPS2, 0.5 nmol Sfp, 200 nmol $MgCl_2$, and 100 nmol CoA in 50 mM sodium phosphate pH 7.8 were incubated for 30 min at 30 °C or 60 min at the other two temperatures. After this incubation period, 10 µL 30 mM ATP and 2 µL of 50 mM L-Gln were added. Final concentrations were 5 µM NRPS2, 5 µM Sfp, 2 mM $MgCl_2$, 100 µM CoA, 3 mM ATP, and 1 mM L-Gln in 50 mM sodium phosphate (pH 7.8). The UV-Vis absorbance spectrum was monitored between 500 nm and 700 nm with an expected peak around 590 nm at the following time points: 0, 5, 10, 30, 60, 90, and 120 minutes.

# CHAPTER 3: RESULTS

## NRPS2

### Assembly of the NRPS2 genetic locus

The *S. laurentii* genomic fosmid library was screened by polymerase chain reaction (PCR) using primers NRPS2-1 and 2 to identify the fosmids containing the sequence from *S. laurentii* genome fragment 764 which harbors NRPS2 (Figure 3.1). Fosmids JA2E5, JA3B8, JA3C3, JA3C12, JA4B8, JA16B1, JA17B2, JA18C3, and JA21B8 were identified by this process. Sequencing analysis of the PCR products confirmed that all except JA3C3 and JA16B1 did indeed harbor the targeted sequence.

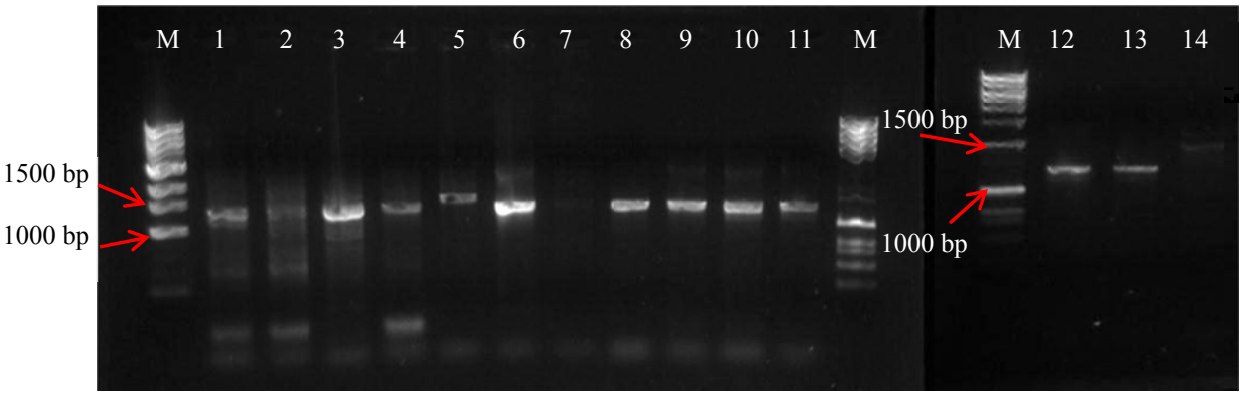



**Figure 3.1** PCR screening of the fosmid library for genomic DNA fragment 764 using primers NRPS2-1 and NRPS2-2 on 1% agarose gels. Expected product size was 1300 bp. Lanes 13 and 14 contain the PCR products from the positive and negative control reactions, genomic *S. laurentii* DNA and pCC1FOS respectively. Lanes 1 through 12 contain the PCR products from the amplification from fosmids JA2E5, JA3B8, JA3C3, JA3C12, JA4B1, JA4B8, JA4B12, JA16B1, JA17B2, JA18C3, JA21B8, and JA23H1.

Fosmids JA18C3 and JA4B8 were subjected to shotgun sequencing. The resulting 64.9 kb sequence was analyzed as shown in Table 3.1 and found to harbor a total of 66 open reading frames (*orf*s). From that analysis, an *orf* map was constructed. Figure 3.2 displays a current *orf* map for the NRPS2 locus spanning a 15 kb region and includes *orf32* to *orf43*, since one enzyme, *orf34*, is expected to compose the complete NRPS system. A putative PPTase, *orf42*, lies downstream of *nrps2* and is also shown since the apo-enzyme requires phosphopantetheinylation.

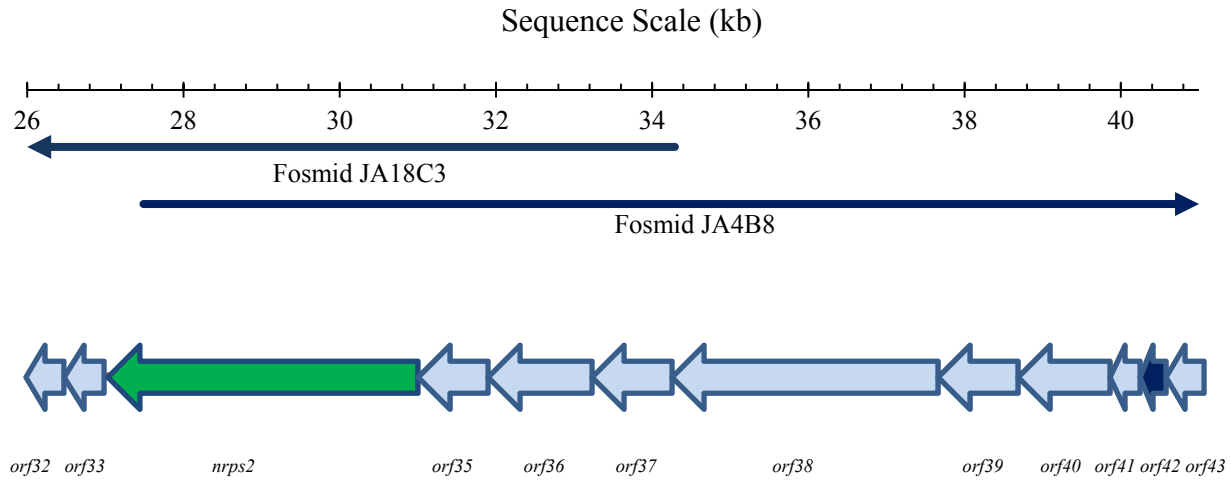



**Figure 3.2** Open reading frame map showing *nrps2*, *orf34*, in green and the putative PPTase-encoding gene, *orf42*, in dark blue. Below the scale, the sequence span of each fosmid is shown. This figure only displays *orf32* to *orf43* from Table 3.1.

Table 3.1 NRPS 2 Assembly (64920 bp) - FramePlot and BLAST analysis

| Gene ID | Strand | Position | aa | Protein Homolog(s) (aa) | Species | %ID/ % Sim | Proposed Function |
|---|---|---|---|---|---|---|---|
| *orf 1* | + | 1-752 (i) | 249 | Hypothetical protein SpriA_01727 (237 aa) | *Streptomyces pristinaespiralis* ATCC 25486 | 50/63 | Unknown |
| *orf 2* | + | 922-1497 | 191 | Hypothetical protein RPC_4006 (183 aa) | *Rhodopseudomonas palustris* BisB18 | 31/50 | Unknown; possible similarity to a subunit of an α-keto acid dehydrogenase complex |
| *orf 3* | + | 1607-2545 | 312 | Hypothetical protein Noca_1578 (332 aa) | *Nocardioides* sp. JS614 | 29/45 | Unknown |
| *orf 4* | + | 2664-2915 | 83 | DNA-binding protein (277 aa) | *Streptomyces sviceus* ATCC 29083 | 65/80 | Unknown |
| *orf 5* | + | 2897-3166 | 89 | Hypothetical protein SAV_5625 (90 aa) | *Streptomyces avermitilis* MA-4680 | 69/81 | Unknown |
| *orf 6* | - | 3198-5075 | 625 | Asparagine synthase, glutamine-hydrolyzing (634 aa) | *Methanococcoides burtonii* DSM 6242 | 28/34 | Amidotransferase |
| *orf 7* | + | 5163-5903 | 248 | LmbE family protein (243 aa) | *Acidothermus cellulolyticus* 11B | 55/69 | Putative deacetylase |
| *orf 8* | + | 6170-7069 | 299 | Hemerythrin HHE cation binding domain-containing protein (279 aa) | *Nocardioides* sp.JS614 | 49/60 | Unknown |
| *orf 9* | + | 7115-7621 | 168 | Hypothetical protein TcurDRAFT_20260 (165 aa) | *Thermomonospora curvata* DSM 43183 | 60/76 | Unknown |
| *orf 10* | + | 8136-8543 | 135 | Endoribonuclease L-PSP superfamily (130 aa) | *Bacillus pseudomycoides* DSM 12442 | 60/70 | Unknown |

"i" indicates an incomplete *orf*, "aa" refers to the length of the polypeptide in amino acids; for strand, "+" indicates the direct strand, while "-" indicates the complementary strand. % ID represents the percent of residues identical to the homolog while % Sim indicates the number of similar or identical residues in the same position as the homolog.

Table 3.1, continued

| Gene ID | Strand | Position | aa | Protein Homolog(s) (aa) | Species | %ID/% Sim | Proposed Function |
|---|---|---|---|---|---|---|---|
| *orf 11* | - | 8588-10195 | 535 | FG-GAP repeat domain protein (516 aa) | *Stigmatella aurantiaca DW4/3-1* | 62/76 | Putative esterase |
| *orf 12* | - | 10366-10746 | 126 | Conserved hypothetical protein (155 aa) | *Streptomyces pristinaespiralis* ATCC 25486 | 62/72 | Putative regulatory protein |
| *orf 13* | - | 10743-10952 | 69 | Predicted protein (82 aa) | *Streptomyces pristinaespiralis* ATCC 25486 | 52/64 | Unknown |
| *orf 14* | + | 11069-11944 | 291 | DNA-binding protein (284 aa) | *Streptomyces pristinaespiralis* ATCC 25486 | 61/73 | Putative esterase |
| *orf 15* | + | 11901-12173 | 90 | Conserved hypothetical protein (96 aa) | *Streptomyces pristinaespiralis* ATCC 25486 | 66/77 | Putative regulatory protein |
| *orf 16* | - | 12175-12987 | 270 | Hypothetical protein SGR_3169 (248 aa) | *Streptomyces griseus* NBRC 13350 | 48/61 | Unknown |
| *orf 17* | + | 13369-14922 | 517 | Unknown (305 aa) | *Streptomyces fungicidus* | 28/40 | Hypothetical nuclease |
| | | | | Endonuclease/exonuclease/phosphatase family protein (304 aa) | *Streptosporangium roseum* DSM 43021 | 35/49 | |
| *orf 18* | - | 15014-15928 | 304 | β-Lactamase (304 aa) | *Streptomyces coelicolor* A3 (2) | 74/85 | Zinc-dependent hydrolase |
| *orf 19* | - | 16346-17059 | 238 | Peptidy-prolyl *cis-trans* isomerase (1236 aa) | *Paracoccidioides brasiliensis* Pb03 | 33/47 | Unknown |
| *orf 20* | - | 17136-17810 | 224 | TetR family transcriptional regulator (219 aa) | *Streptomyces coelicolor* A3 (2) | 81/85 | TetR family transcriptional regulator |

"i" indicates an incomplete *orf*, "aa" refers to the length of the polypeptide in amino acids; for strand, "+" indicates the direct strand, while "-" indicates the complementary strand. % ID represents the percent of residues identical to the homolog while % Sim indicates the number of similar or identical residues in the same position as the homolog.

Table 3.1, continued

| Gene ID | Strand | Position | aa | Protein Homolog(s) (aa) | Species | %ID/ % Sim | Proposed Function |
|---|---|---|---|---|---|---|---|
| *orf 21* | + | 17958-18983 | 341 | Hypothetical protein SCO7720 (368 aa) | *Streptomyces coelicolor* A3 (2) | 87/92 | Unknown |
| *orf 22* | + | 18932-19330 | 132 | Hypothetical protein FRAAL6818 (265 aa) | *Frankia alni* ACN14a | 55/65 | Unknown |
| *orf 23* | - | 19371-19664 | 97 | Translation elongation factor Tu (405 aa) | *Coprothermobacter proteolyticus* DSM 5265 | 30/49 | Unknown |
| *orf 24* | - | 19661-20251 | 196 | Acetyltransferase (188 aa) | *Streptomyces pristinaespiralis* ATCC 25486 | 59/66 | Acetyltransferase |
| *orf 25* | + | 20405-20905 | 166 | Redoxin domain protein / thiol peroxidase (166 aa) | *Pseudomonas putida* KT2440 | 83/91 | Redoxin |
| *orf 26* | - | 21060-21857 | 265 | DNA polymerase beta domain protein region (263 aa) | *Geobacillus* sp. Y412MC10 | 45/60 | Unknown |
| *orf 27* | - | 22026-22652 | 208 | Putative transmembrane protein (604 aa) | *Kordia algicida* OT-1 | 32/51 | Unknown |
| *orf 28* | + | 22817-24214 | 465 | FAD linked oxidase domain protein (458 aa) | *Cyanothece* sp. PCC 7425 | 54/69 | FAD-linked oxidase |
| *orf 29* | - | 24252-24572 | 106 | Conserved hypothetical protein (132 aa) | *Streptomyces pristinaespiralis* ATCC 25486 | 39/53 | Unknown |
| *orf 30* | - | 24646-25278 | 210 | Hypothetical protein (200 aa) | *Streptomyces clavuligerus* | 81/90 | Deaminase-reductase (NADH) |

"i" indicates an incomplete *orf*, "aa" refers to the length of the polypeptide in amino acids; for strand, "+" indicates the direct strand, while "-" indicates the complementary strand. % ID represents the percent of residues identical to the homolog while % Sim indicates the number of similar or identical residues in the same position as the homolog.

Table 3.1, continued

| Gene ID | Strand | Position | aa | Protein Homolog(s) (aa) | Species | %ID/% Sim | Proposed Function |
|---|---|---|---|---|---|---|---|
| *orf 31* | - | 25481-26119 | 212 | Hypothetical protein SAV_7466 (210 aa) | *Streptomyces avermitilis* MA-4680 | 60/69 | Unknown/putative exporter |
| *orf 32* | - | 26116-26517 | 133 | Predicted protein (99 aa) | *Streptomyces pristinaespiralis* ATCC 25486 | 64/72 | Unknown |
| *orf 33* | - | 26551-26967 | 138 | Predicted protein (92 aa) | *Streptomyces pristinaespiralis* ATCC 25486 | 32/55 | Unknown |
| *orf 34* | - | 26998-30915 | 1305 | Putative indigoidine synthase IndC (1283 aa) | *Streptomyces aureofaciens* | 58/71 | NRPS subunit (A-Ox-T-TE) |
| *orf 35* | - | 31158-31910 | 250 | Conserved hypothetical protein (221 aa) | *Streptomyces sviceus* ATCC 29083 | 56/68 | Unknown |
| *orf 36* | - | 31917-33332 | 471 | Putative *S*-adenosyl-L-homocysteine hydrolase (485 aa) | *Streptomyces griseus* NBRC 13350 | 87/91 | SAM hydrolase/NAD+-dependent enzyme |
| *orf 37* | - | 33325-34230 | 301 | 5,10-methylenetetrahydrofolate reductase (286 aa) | *Streptomyces rochei* | 80/87 | 5,10-methylenetetrahydro-folate reductase |
| *orf 38* | - | 34227-37736 | 1169 | Methionine synthase (1162 aa) | *Streptomyces pristinaespiralis* ATCC 25486 | 83/88 | Methionine synthase |
| *orf 39* | - | 37733-38716 | 327 | Putative adenosine kinase (327 aa) | *Streptomyces lavendulae* | 63/74 | Kinase |
| *orf 40* | - | 38731-39939 | 402 | SAM synthetase (402 aa) | *Streptomyces pristinaespiralis* ATCC 25486 | 86/92 | SAM synthetase |

"i" indicates an incomplete *orf*, "aa" refers to the length of the polypeptide in amino acids; for strand, "+" indicates the direct strand, while "-" indicates the complementary strand. % ID represents the percent of residues identical to the homolog while % Sim indicates the number of similar or identical residues in the same position as the homolog.

Table 3.1, continued

| Gene ID | Strand | Position | aa | Protein Homolog(s) (aa) | Species | %ID/% Sims | Proposed Function |
|---|---|---|---|---|---|---|---|
| *orf 41* | - | 39936-40190 | 84 | Hypothetical protein SGR_5184 (83 aa) | *Streptomyces griseus* NBRC 13350 | 41/66 | Unknown / putative carrier protein |
| *orf 42* | - | 40204-40650 | 148 | Holo-acyl carrier protein synthase (125 aa) | *Halothermothrix orenii* H 168 | 38/59 | Phosphopantetheinyl transferase (ACPS family) |
| *orf 43* | - | 40652-41539 | 295 | serS1 (350 aa) | *Streptomyces clavuligerus* ATCC 27064 | 45/58 | Putative seryl-tRNA synthetase |
| *orf 44* | - | 41676-42569 | 297 | Conserved hypothetical protein (291 aa) | *Streptomyces sviceus* ATCC 29083 | 50/68 | Unknown |
| *orf 45* | - | 42634-43677 | 347 | Hypothetical protein bll2645 (374 aa) | *Bradyrhizobium japonicum* USDA 110 | 45/60 | Methyltransferase |
| *orf 46* | - | 43674-44516 | 280 | Hypothetical protein MAP0222c (297 aa) | *Mycobacterium avium subsp. paratuberculosis* K-10 | 42/56 | Methyltransferase |
| *orf 47* | - | 44774-45370 | 198 | MaoC domain-containing protein (146 aa) | *Myxococcus xanthus* DK 1622 | 50/65 | Unknown (dehydratase/hydratase/ epimerase) |
| *orf 48* | - | 45505-46743 | 412 | Methionine γ-lyase (396 aa) | *Saccharopolyspora erythraea* NRRL 2338 | 54/67 | PLP-dependent protein, possible methionine γ-lyase or cystathionine β-lyase |
| *orf 49* | - | 46788-48242 | 484 | Drug resistance transporter EmrB/QacA subfamily protein (481 aa) | *Streptomyces pristinaespiralis* ATCC 25486 | 86/93 | Transporter/resistance |
| *orf 50* | + | 48476-49189 | 237 | Hypothetical protein SGR_3194 (253 aa) | *Streptomyces griseus* NBRC 13350 | 68/77 | Unknown |

"i" indicates an incomplete *orf*, "aa" refers to the length of the polypeptide in amino acids; for strand, "+" indicates the direct strand, while "-" indicates the complementary strand. % ID represents the percent of residues identical to the homolog while % Sim indicates the number of similar or identical residues in the same position as the homolog.

Table 3.1, continued

| Gene ID | Strand | Position | aa | Protein Homolog(s) (aa) | Species | %ID/% Sim | Proposed Function |
|---|---|---|---|---|---|---|---|
| *orf 51* | - | 49260-49505 | 81 | Hypothetical protein SAV_5886 (68 aa) | *Streptomyces avermitilis* MA-4680 | 61/72 | Unknown |
| *orf 52* | + | 49939-51066 | 375 | Hypothetical protein SGR_3852 (417 aa) | *Streptomyces griseus* NBRC 13350 | 40/57 | Unknown / putative regulatory protein |
| *orf 53* | + | 51057-51569 | 170 | Putative acetyltransferase (175 aa) | *Streptomyces griseus* NBRC 13350 | 43/61 | Acetyltransferase |
| *orf 54* | - | 51581-52477 | 298 | Hypothetical protein SCO4676 (290 aa) | *Streptomyces coelicolor* A3 (2) | 39/47 | Unknown |
| *orf 55* | - | 52871-53488 | 205 | Transcriptional regulator (207 aa) | *Streptomyces* sp. Mg1 | 69/81 | Transcriptional regulator |
| *orf 56* | + | 53560-53757 | 65 | Hypothetical protein SSAG_03155 (77 aa) | *Streptomyces* sp. Mg1 | 50/75 | Unknown |
| *orf 57* | + | 53754-53936 | 60 | Hypothetical protein SSAG_03154 (53 aa) | *Streptomyces* sp. Mg1 | 45/69 | Unknown |
| *orf 58* | - | 53933-54637 | 234 | Cytidine/deoxycytidylate deaminase/NUDIX/methyltransferase domain-containing protein (548 aa) | *Deinococcus radiodurans* R1 | 37/50 | Deaminase/hydrolase/ methyltransferase |
| *orf 59* | + | 55057-56835 | 592 | Hypothetical protein SSAG_03543 (618 aa) | *Streptomyces sp. Mg1* | 27/38 | Unknown |
| *orf 60* | - | 56984-58291 | 435 | Metal-dependent hydrolase (430 aa) | *Flavobacterium sp. MED217* | 29/46 | Amidohydrolase / aminoacylase / carboxypeptidase |

"i" indicates an incomplete *orf*, "aa" refers to the length of the polypeptide in amino acids; for strand, "+" indicates the direct strand, while "-" indicates the complementary strand. % ID represents the percent of residues identical to the homolog while % Sim indicates the number of similar or identical residues in the same position as the homolog.

Table 3.1, continued

| Gene ID | Strand | Position | aa | Protein Homolog(s) (aa) | Species | %ID/% Sim | Proposed Function |
|---|---|---|---|---|---|---|---|
| *orf 61* | + | 58636-59928 | 430 | Hypothetical protein SCO0012 / SCO7835 / SC8E7.32. (136 aa) | *Streptomyces coelicolor A3 (2)* | 78/89 | Unknown[a] |
| | | | | Conserved hypothetical protein (256 aa) | *Streptomyces sviceus* ATCC 29083 | 34/47 | Unknown[b] |
| *orf 62* | + | 60239-61534 | 431 | Chloramphenicol resistance protein (436 aa) | *Streptomyces venezuelae* | 80/87 | Antibiotic resistance |
| *orf 63* | - | 61738-62352 | 204 | Hypothetical protein SGR_427 (204 aa) | *Streptomyces griseus NBRC 13350* | 80/89 | Unknown |
| *orf 64* | - | 62349-62723 | 124 | Hypothetical protein SCO2859 (125 aa) | *Streptomyces coelicolor* A3 (2) | 91/93 | Regulatory protein |
| *orf 65* | - | 62948-63394 | 148 | Conserved hypothetical protein (195 aa) | *Streptomyces sviceus* ATCC 29083 | 46/53 | Unknown |
| *orf 66* | + | 63384-64787 | 467 | Two-component system sensor kinase (469 aa) | *Streptomyces avermitilis* MA-4680 | 57/66 | Regulatory protein |

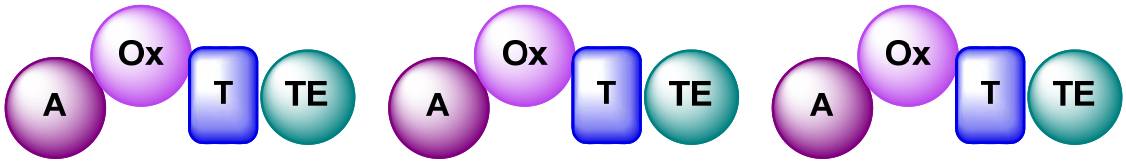
"i" indicates an incomplete *orf*, "aa" refers to the length of the polypeptide in amino acids; for strand, "+" indicates the direct strand, while "-" indicates the complementary strand. % ID represents the percent of residues identical to the homolog while % Sim indicates the number of similar or identical residues in the same position as the homolog

[a] Possible missing stop codon; good similarity for the first 117aa

[b] Similarity for the last 250aa - possible missing stop codon

## Nonribosomal peptide synthetase domain analysis and substrate specificity

The sequence manipulation suite translated the deoxyribonucleic acid (DNA) sequence of *nrps2* into a FASTA-format 1305 amino-acid sequence (*76*). ClustalW aligned the deduced amino acid sequence with provided amino acid sequences from known NRPSs, as detailed in Table 3.2 (*61*). These sequences were downloaded from NCBI and cut appropriately for their specific domains (*62*). The alignments of these NRPS homologs and NRPS2 were then analyzed for the modules, domains, and conserved motifs normally present within each NRPS domain. In addition, the A domain of NRPS2 was inspected for conserved amino acid residues that specify the amino acid monomer activated by the module (*36-38,77,78*). From this analysis, NRPS2 lacks a C domain, but possesses A, Ox, T, and TE domains (Figure 3.3). Figures 3.4 and 3.5 detail the motifs within the adenylation domain, while Figures 3.6, 3.7, and 3.8 display the conserved motifs within the Ox, T, and TE domains respectively. Table 3.3 gives the residues specifying the L-Gln substrate within the adenylation domain. The module and domain organization of NRPS2 mimics those of BpsA (57% identity, 69% similarity) and IndC (52% identity, 67% similarity), both of which produce indigoidine (*6,7,55*). It should be noted that a TE domain was not identified by Reverchon, *et al.*; however, sequence analysis determined its presence (*6*).




NRPS2 (1305 aa)
*Streptomyces laurentii*

BpsA (1282 aa)
*Streptomyces lavendulae*
57% Identity, 69% Similarity

IndC (1488 aa)
*Erwinia chrysanthemi*
52% Identity, 67% Similarity

**Figure 3.3** Module and domain organization of NRPS2, BpsA, and IndC. The source organism and percent identity and similarity to NRPS2 of each protein are shown below each schematic (*6,7,55*).

Table 3.2 Amino acid sequences used for alignments

| Abbreviated Name | Name (aa) | Species | % ID/ % Sim |
|---|---|---|---|
| BpsA | Blue-pigment synthetase (1282 aa) | *S. lavendulae* | 58/71 |
| IndC | Indigoidine synthetase (1488 aa) | *E. chrysanthemi* | 53/68 |
| LicA | Lichenysin synthetase A (3582 aa) | *Bacillus licheniformis* | 36/51 |
| TycIII | Tyrocidine synthetase 3 (6486 aa) | *Brevibacillus brevis* | 40/57 |
| BacA | Bacitracin synthetase 1 (5255 aa) | *Bacillus licheniformis* | 35/54 |
| MycB | Mycosubtilin synthetase B (5369 aa) | *Bacillus subtilis* | 37/53 |
| MtaC | Myxothiazol synthetase C (1290 aa) | *Stigmatella aurantiaca* DW4/3-1 | 33/50 |
| MtaD | Myxothiazol synthetase D (3291 aa) | *Stigmatella aurantiaca* DW4/3-1 | 32/47 |
| BlmIII | Bleomycin synthetase III (935 aa) | *Streptomyces verticillus* | 26/41 |
| EpoB | Epothilone synthetase B (1410 aa) | *Sorangium cellulosum* | 32/47 |
| EntF | Enterobactin synthetase F (1293 aa) | *Escherichia coli* Bl21(DE3) | 36/53 |

* % ID represents the percent of residues identical to NRPS2 while % Sim indicates the number of similar or identical residues to NRPS2.

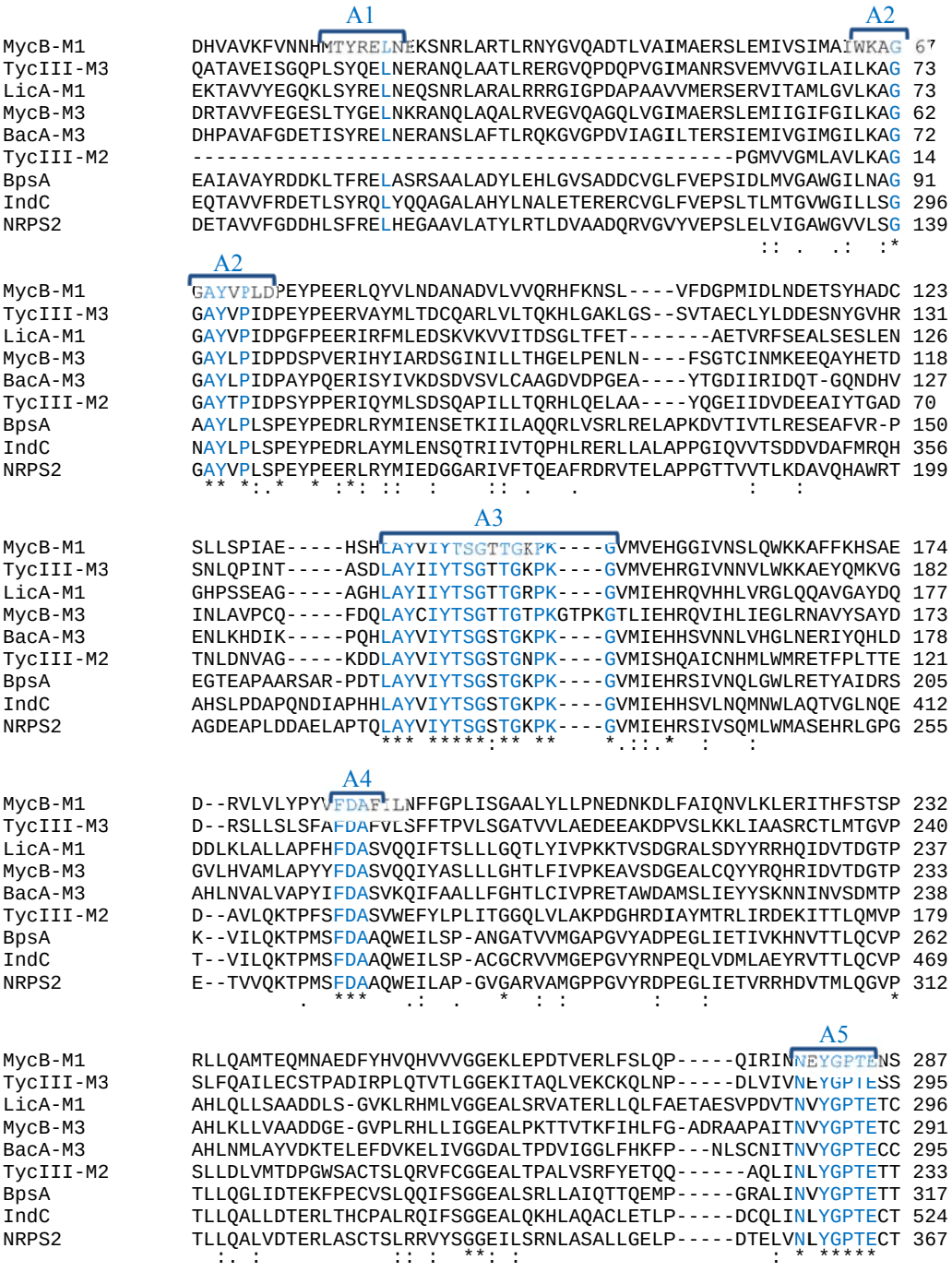



**Figure 3.4** ClustalW alignment and conserved motifs A1 to A5 within the NRPS2 adenylation domain (*36,61*). M# (e.g. M1, M2, etc) refers to the order of the module within a subunit. Brackets demark the conserved motifs and blue font indicates the residues that are strictly conserved within a motif. MycB is a subunit from the mycosubtilin synthetase of *Bacillus subtilis* (*79*). TycIII is a subunit from the tyrocidine synthetase of *Brevibacillus brevis* (*80*). LicA is a subunit from the lichenysin synthetase of *Bacillus licheniformis* (*81*). BacA is a subunit from the bacitracin synthetase of *Bacillus licheniformis* (*82*). BpsA and IndC are the indigoidine synthetases from *Streptomyces lavendulae* and *Erwinia chrysanthemi*, respectively (*6,7*).

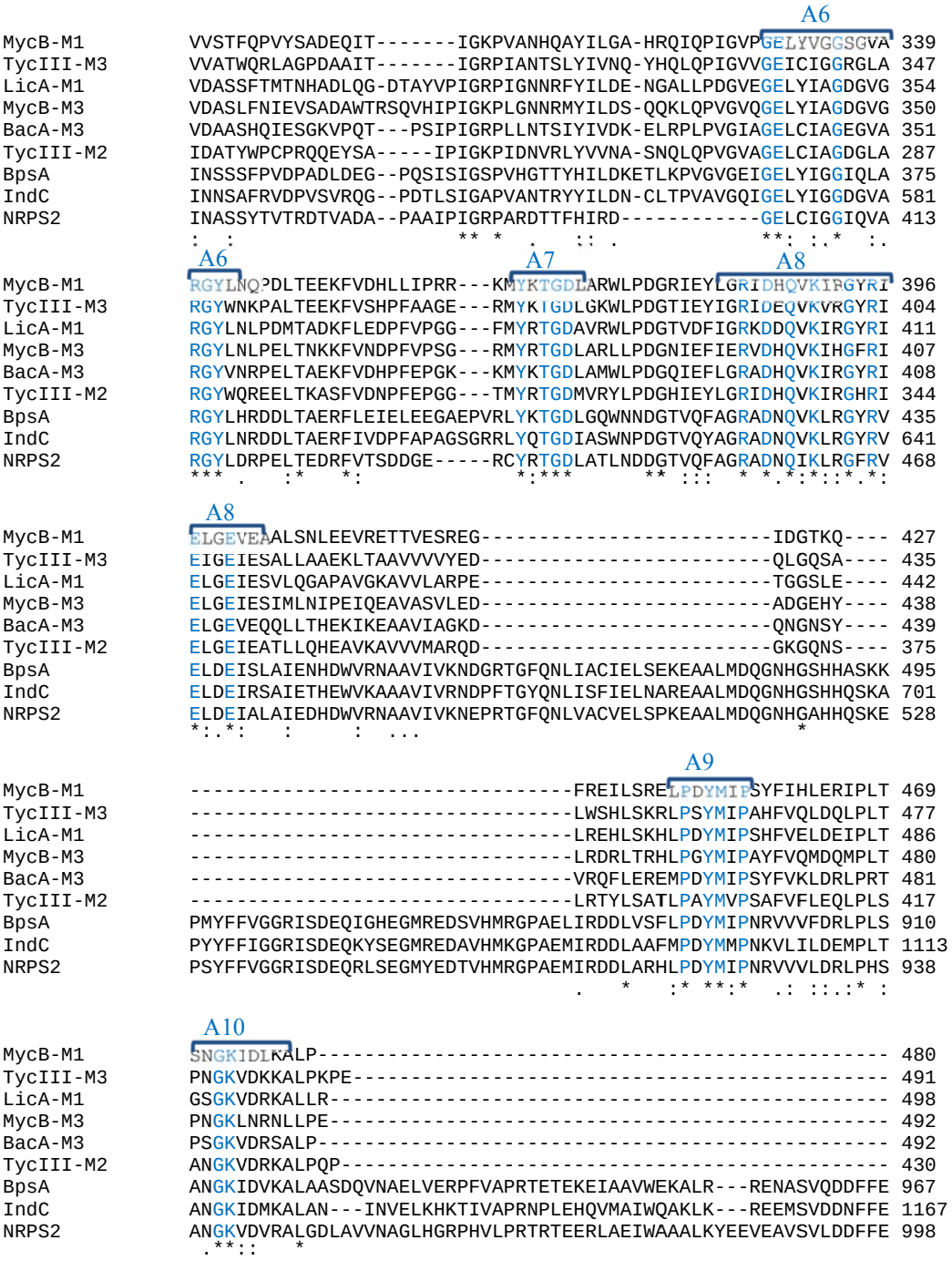



**Figure 3.5** ClustalW alignment and conserved motifs A6 to A10 within the NRPS2 adenylation domain (*36,61*). M# (e.g. M1, M2, etc) refers to the order of the module within a subunit. Brackets demark the conserved motifs and blue font indicates the residues that are strictly conserved within a motif. MycB is a subunit from the mycosubtilin synthetase of *Bacillus subtilis* (*79*). TycIII is a subunit from the tyrocidine synthetase of *Brevibacillus brevis* (*80*). LicA is a subunit from the lichenysin synthetase of *Bacillus licheniformis* (*81*). BacA is a subunit from the bacitracin synthetase of *Bacillus licheniformis* (*82*). BpsA and IndC are the indigoidine synthetases from *Streptomyces lavendulae* and *Erwinia chrysanthemi*, respectively (*6,7*).

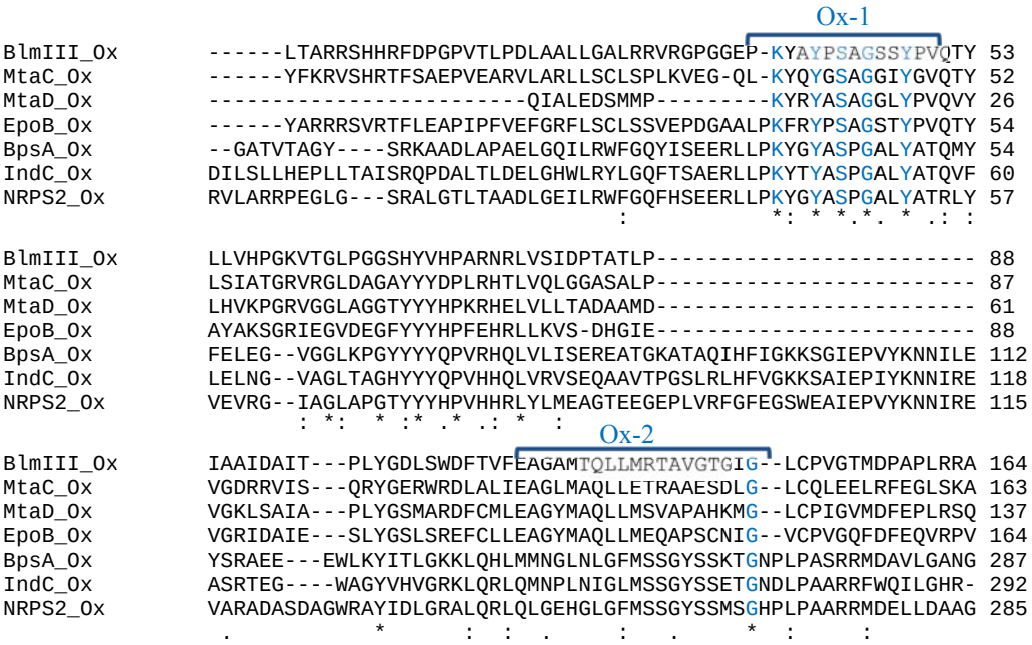



**Figure 3.6** ClustalW alignment and conserved motifs within the NRPS2 oxidation domain (*61,83*). Ox indicates that only the oxidation domain was included in the alignment. Brackets demark the conserved motifs and blue font indicates the residues that are strictly conserved within a motif. BlmIII is a subunit from the bleomycin synthetase of *Streptomyces verticillus* (*84*). MtaC and MtaD are subunits from the myxothiazol synthetase of *Stigmatella aurantiaca* DW4/3-1 (*31*). EpoB is a subunit from the epothilone synthetase of *Sorangium cellulosum* (*85*). BpsA and IndC are the indigoidine synthetases from *Streptomyces lavendulae* and *Erwinia chrysanthemi*, respectively (*6,7*).

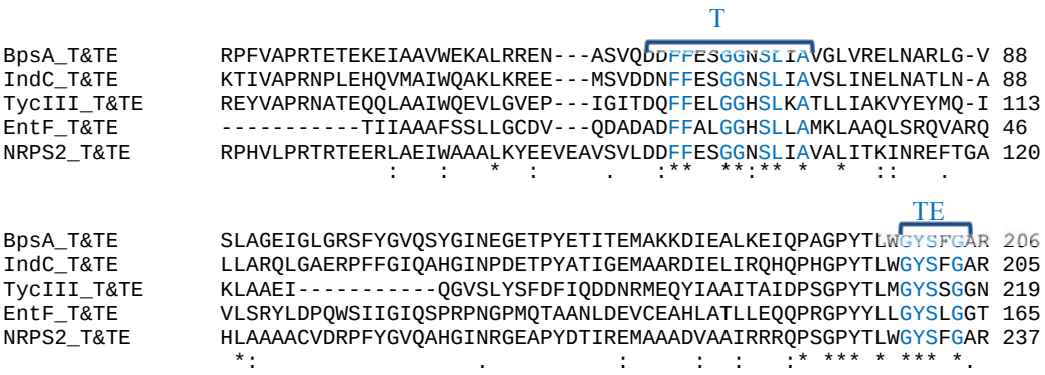



**Figure 3.7** ClustalW alignment and **c**onserved motifs within the NRPS2 T and TE domains (*36,61*). T and TE indicate that only the T and TE domains were included in the alignment. Brackets demark the conserved motifs and blue font indicates the residues that are strictly conserved within a motif. EntF is a subunit from the enterobactin synthetase in *Escherichia coli* BL21(DE3) (*86*). TycIII is a subunit from the tyrocidine synthetase of *Brevibacillus brevis* (*80*). BpsA and IndC are the indigoidine synthetases from *Streptomyces lavendulae* and *Erwinia chrysanthemi*, respectively (*6,7*).

Table 3.3 Substrate specificity alignments for the NRPS2 A domain

| GrsA position | 235 | 236 | 239 | 278 | 299 | 301 | 322 | 330 | 331 | 517 |
|---|---|---|---|---|---|---|---|---|---|---|
| **L-Gln per Stachelhaus, *et al.* (*34,37,38*)** | **D** | **A** | **Q** | **D** | **L** | **G** | **V** | **V** | **D** | **K** |
| BacA M4 (*82*) | D | A | K | D | I | G | V | V | D | K |
| MycB M3 (*79*) | D | A | Q | D | L | G | V | V | D | K |
| LicA M1 (*81*) | D | A | Q | D | L | G | V | V | D | K |
| TycIII M2 (*80*) | D | A | W | Q | F | G | L | V | V | K |
| IndC (*6*) | D | A | W | Q | F | G | L | I | N | K |
| BpsA (*7*) | D | A | W | Q | F | G | V | I | N | K |
| NRPS2 | D | A | W | Q | Y | G | L | I | N | K |
| NRPS2 position | 267 | 268 | 270 | 309 | 334 | 336 | 360 | 368 | 369 | 942 |
| Prediction | L-Gln | | | | | | | | | |

*M# (e.g. M1, M2, etc) refers to the order of the module within a subunit. All A domains used in this alignment with NRPS2 activate L-Gln.

## Expression and purification of NRPS2

The pilot expression study of the N-terminal hexahistidine fusion construct of NRPS2 revealed the production of a blue pigment when the host *E. coli* strain was cultivated at 15 °C for 24 h. This pigment was not observed in the pET28b(+) cultures and was only observed to a lesser degree in the C-terminal hexahistidine fusion protein constructs of NRPS2 as shown in Figure 3.8. The expression study, in conjunction with the blue pigment production for the N-terminal hexahistidine NRPS2 construct, showed that a 24 h induction period with 0.04 mM IPTG at 15 °C produced the most soluble fusion protein (Figure 3.8 and 3.9). The NRPS2 N-terminal hexahistidine fusion protein should be 144.4 kDa. Due to these observations, the cultures for the pilot expression study of the C-terminal NRPS2 hexahistidine construct were never evaluated for NRPS2 expression and solubility.

A. 


B. 


C. 

**Figure 3.8** Photographs of *E. coli* cultures from the pilot expression study of NRPS2 and of 1 L cultures. **A**. N-terminal hexahistidine fusion protein cultures at 15 °C with a 24 h induction period. From left to right, pET28b(+)-*nrps2* with 0.04 mM IPTG, pET28b(+)-*nrps2* with 0.4 mM IPTG, pET28b(+) with 0.04 mM IPTG, and pET28b(+)-*nrps2* with 0 mM IPTG. **B**. C-terminal hexahistidine fusion protein expression study cultures at 15 °C with a 24 h induction period. From left to right, pET29b(+)-*nrps2* with 0.04 mM IPTG, pET29b(+)-*nrps2* with 0.4 mM IPTG, and pET29b(+)-*nrps2* with 0 mM IPTG. **C**. The 1 liter cultures of the NRPS2 N-terminal hexahistidine fusion protein at 15 °C with a 24 h induction period.

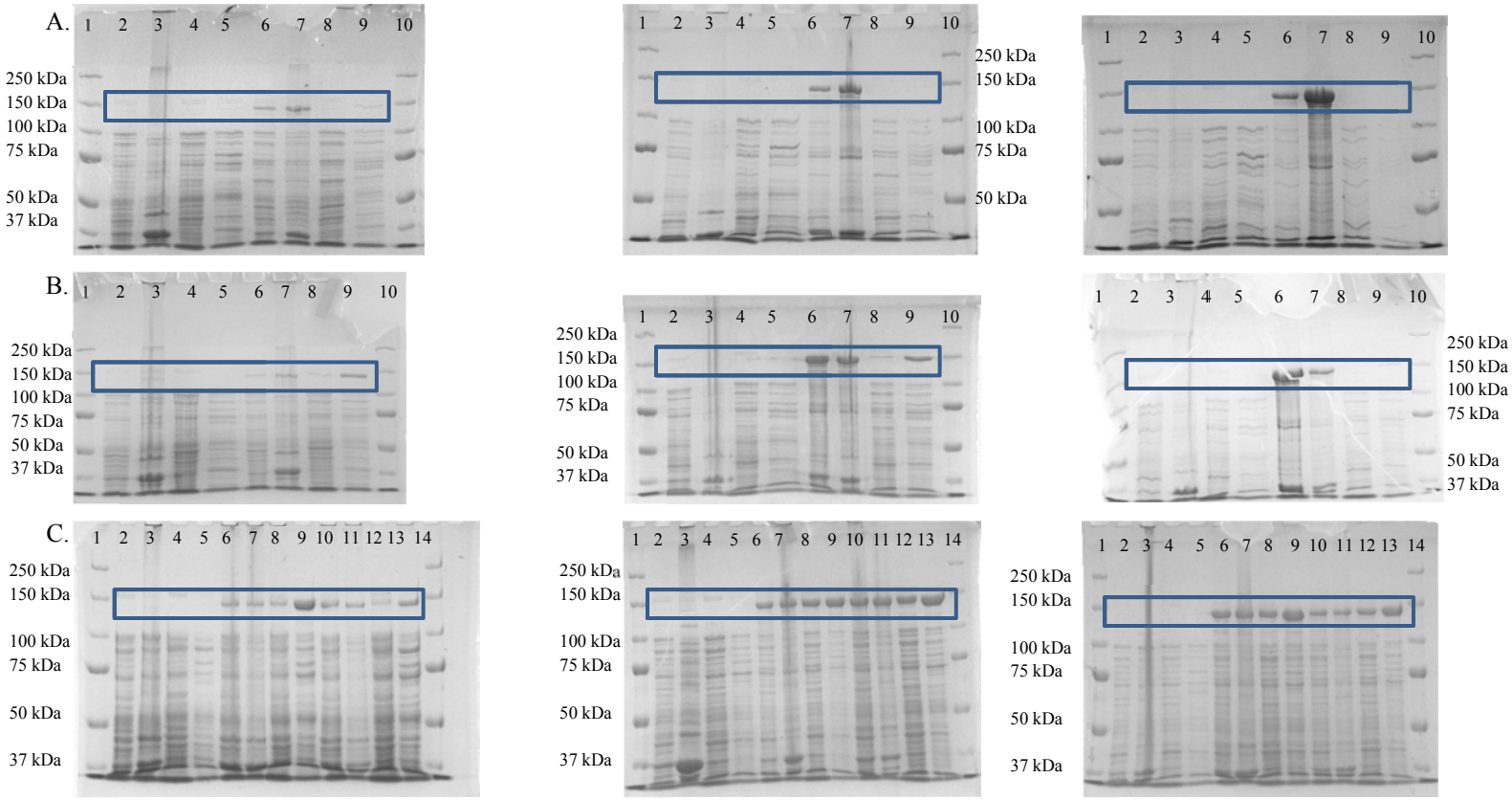



**Figure 3.9** 7% SDS-PAGE analysis from the expression study on the N-terminal hexahistidine fusion construct of NRPS2, 144 kDa outlined in blue. Lanes 1 and 10 in panels A and B are loaded with high molecular weight protein markers, while in panel 3, lanes 1 and 14 are loaded with high molecular weight protein markers. Panel A shows the expression and solubility analysis from the 37 °C 3 hour induction period with IPTG concentration increasing from 0 mM IPTG to 0.04 mM IPTG to 0.4 mM IPTG from left to right. Panel B shows the expression and solubility analysis from the 28 °C 6 hour induction period with IPTG concentration increasing from 0 mM IPTG to 0.04 mM IPTG to 0.4 mM IPTG from left to right. Panel C shows the purifications from the 15 °C 24 and 48 hour incubation with IPTG concentration increasing from 0.0 mM IPTG to 0.04 mM IPTG to 0.4 mM IPTG from left to right. Lanes 2 through 5 in each gel are from pET28b(+) cultures, while lanes 6 through 9 (and 10 through 13 in the 48 hour cultures in panel C) are from the pET28b(+)-*nrps2* cultures. Lanes 2 through 5 and 6 through 9 in addition to lanes 10 through 13 for panel C in order from left to right are the lysate, the insoluble pellet, the cell free extract, and the fraction that eluted from the Ni-NTA resin.

For the large scale production of NRPS2, the induction conditions at 15 °C for 24 h using 0.04 mM IPTG was chosen. Following nickel-chelate chromatography, NRPS2 was subjected to anion-exchange chromatography to remove a majority of contaminants (Figure 3.10).

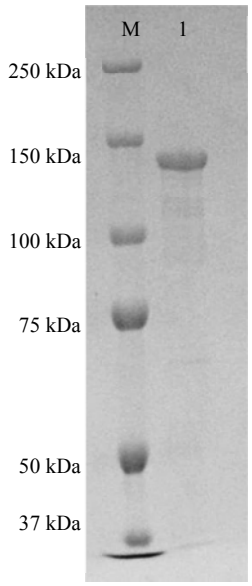



**Figure 3.10** 7% SDS-PAGE analysis from the large scale expression of the 144 KDa N-terminal hexahistidine fusion protein of NRPS2. Lanes M is loaded with a high molecular weight protein marker. Lane 1 is loaded with a fraction containing NRPS2 following anion-exchange chromatography.

## Blue pigment production in *S. laurentii*

In *Streptomyces virginiae* and *Streptomyces lavendulae*, blue pigment production was induced with 30 μg/mL γ-nonalactone (*7,8*). Based on homology to BpsA and locus similarity to *S. lavendulae*, it was hypothesized that γ-nonalactone could activate blue pigment expression in *S. laurentii* as well (*7*). The *S. laurentii* liquid cultures in TSB, Y6.5, and Y7.5 were induced at various time points with γ-nonalactone. Blue pigment production was not observed by visual

inspection, but one culture in Y6.5 medium had a darker tint than the others, and blue pigment production was not described as readily observable in *S. lavendulae* or *S. virginiae* (*7,8*). No studies on the pellets from these cultures have been performed yet.

Results were more promising on solid media. A purple pigment appeared when *S. laurentii* was cultivated on potato-glucose-peptone (PGP) medium from a 48 hour seed culture inoculated with a *S. laurentii* spore suspension or a 24 hour seed culture inoculated with a *S. laurentii* glycerol stock (Figure 3.11). Other media failed to support any pigment production except for a red pigment that was frequently observed (Figure 3.11). The purple pigment on the PGP agar plates could be from a mixture of blue and red pigments. Pictures were taken of both confluent plaques and streaked plaques. Photographs of the cultures grown on solid media are shown in Figure 3.11.

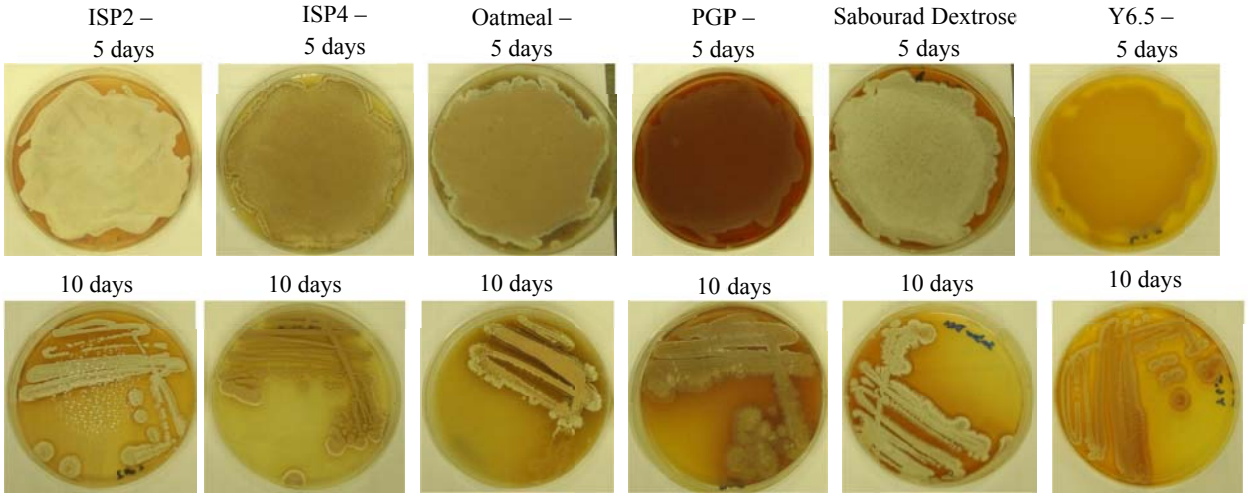



**Figure 3.11** *S. laurentii* cultures grown on solid media. The top row shows the confluent plaques after 5 days of growth. The bottom row displays streaked plaques after 10 days of growth.

### *In vitro* blue pigment production

To assess the capability of NRPS2 to produce a blue pigment *in vitro*, a reaction containing $MgCl_2$, NRPS2, CoenzymeA, and Sfp, a promiscuous PPTase from *Bacillus subtilis*, in a buffered solution were incubated at one of three temperatures to phosphopantetheinylate NRPS2 (*87*). After an appropriate time, ATP and L-Gln were added to the solution, and the reaction was monitored by UV-Vis spectroscopy between 500 and 700 nm where the blue pigment was expected to produce a peak (*7,8*). The first set of results show an increase in absorbance peaking at 603 nm for the three temperatures used: 15 °C, 25 °C, and 30 °C. The greatest increase in absorbance was observed at 25 °C. For all three temperatures, the absorbance at 603 nm maximized at 5 min and decreased at later times.

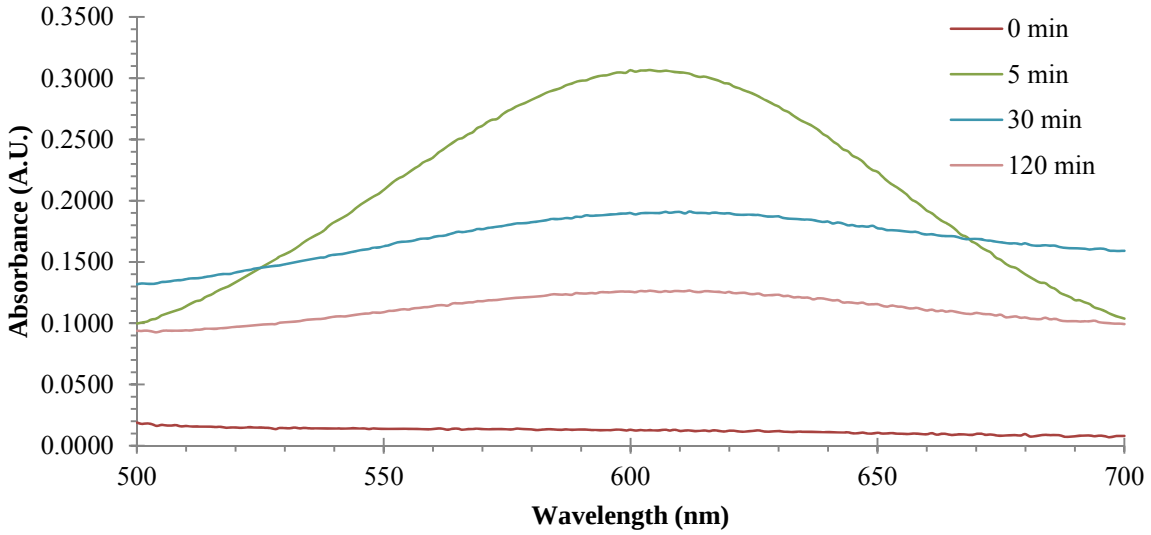



**Figure 3.12** UV-Vis absorption spectrum from the *in vitro* blue pigment production assay at 25 °C. Time points are denoted by the specified line color in the legend. Only the spectra for the following time points are shown: 0 min, 5 min, 30 min, and 120 min.

## NRPS3

### Assembly of the NRPS3 genetic locus

The *S. laurentii* genomic fosmid library was screened by PCR to identify the fosmids containing the sequence from *S. laurentii* genome DNA fragment 3305 which harbors part of the NRPS3 system. Fosmids JA7F9, JA9D12, JA10E1, JA18A5, JA19B8, JA22C11, and JA22C12 were identified by this process as can be seen in Figure 3.13. Sequencing analysis confirmed that all the fosmids did indeed harbor the targeted sequence.

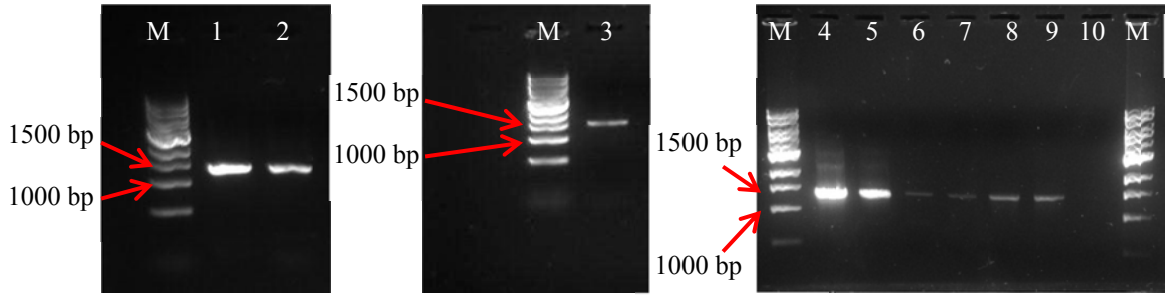



**Figure 3.13** PCR screening of the fosmid library for genomic DNA fragment 3305 using primers NRPS3-1 and NRPS3-2 on 1% agarose gels. Lanes 9 and 10 are loaded with the positive and negative control reactions, genomic *S. laurentii* DNA and pCC1FOS, respectively. Lanes M are loaded with 1 kb marker. Lanes 1 to 8 contain the PCR products from amplification on fosmids JA7F9, JA9D12, JA19B8, JA10E1, JA18A5, JA22C11, and JA22C12. The expected product size was 1300 bp.

Fosmid JA9D12 was subjected to shotgun sequencing. Sequence analysis indicated that extending the sequence was necessary to ensure coverage of the entire locus. PCR screening of the *S. laurentii* genomic fosmid library then identified fosmids that would extend the NRPS3 system locus. Fosmid JA5F6 was identified for extending the sequence upstream of JA9D12, and fosmids JA10E1, JA15F8, JA19B7, JA19B8, and JA21A10 were identified to extend the

sequence downstream of fosmid JA9D12 (Figure 3.14). Fosmids JA5F6 and JA15F8 were chosen for shotgun sequencing. The resulting 103.1 kb sequence was analyzed as shown in in Table 3.4 and found to harbor 74 *orf*s. From that analysis, an *orf* map was constructed; Figure 3.15 shows a partial *orf* map. For NRPS3, four putative NRPS subunits are denoted by *orf30* through *orf33*. *Orf*31 shows identity and similarity only with the MonoOx domain in some hybrid NRPS-PKS systems and stand-alone $F_{420}$-utilizing enzymes; it is only 349 aa which is close to the size of the MonoOx domains of MtaG and MelG (*20,31*). *Orf42* and *orf51* encode a putative PPTase and a MbtH-like protein respectively. At the end of JA15F8 lies another NRPS system which we are denoting as NRPS4. Frameshift mutations appear to exist in the sequencing data covering the incomplete NRPS4 system but these possible mutations must be confirmed by additional sequencing. The region in question is covered only by pyrosequencing data which is prone to frameshift errors (*88*).

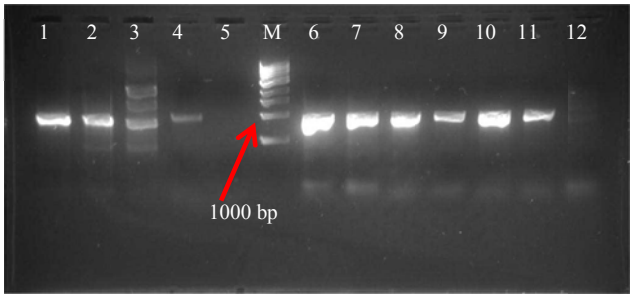



**Figure 3.14** PCR screening of the fosmid library to identify fosmids overlapping with JA9D12. Primers NRPS3-3 and NRPS3-4 were used for lanes 1 through 5 with an expected product size of 952 bp. Lanes 4 and 5 contain the positive and negative control reactions, *S. laurentii* genomic DNA and pCC1FOS, respectively. Lanes 1 to 3 contain the reactions run on fosmids JA5F6, JA7F9, and JA14A2. Primers NRPS3-5 and NRPS3-6 were used for lanes 6 through 12 with an expected product size of 955 bp. Lanes 11 and 12 contain the positive and negative control reactions, *S. laurentii* genomic DNA and pCC1FOS, respectively. Lanes 6 through 10 contain the PCR products from fosmids JA10E1, JA15F8, JA19B7, JA19B8, and JA21A10.

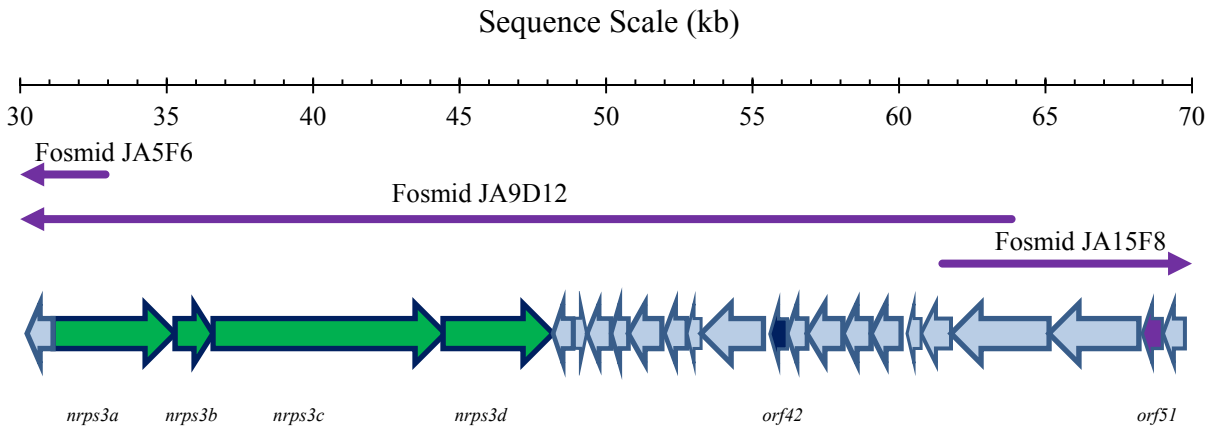



**Figure 3.15** Partial open reading frame map for the NRPS3 locus. *Orf30* to *orf33* display *nrps3a*, *nrps3b*, *nrps3c*, and *nrps3d*, while *orf42* and *orf51* represent the genes encoding the putative PPTase and MbtH-like protein respectively. The fosmids span the sequence as indicated by the purple lines below the scale. This figure only displays *orf29* to *orf52* from Table 3.4

Table 3.4 NRPS3 Assembly (103052 bp) - FramePlot and BLAST analysis

| Gene ID | Strand | Position | aa | Homolog(s) (aa) | Species | %ID/% Sim | Proposed Protein Function |
|---|---|---|---|---|---|---|---|
| *orf1* | + | 1-494 (i) | 163 | Peptide ABC transporter permease (907 aa) | *Streptococcus sanguinis* SK36 | 26/45 | Transporter |
| *orf2* | + | 491-3412 | 973 | Sporulation protein K (931 aa) | *Streptomyces bingchenggensis* BCW-1 | 60/70 | Sporulation protein |
| *orf3* | - | 3430-4485 | 351 | Putative hydrolase (466 aa) | *Streptomyces scabiei* 87.22 | 62/74 | Hydrolase (glycosidic) |
| | | | | Ricin B lectin (496 aa) | *Actinosynnema mirum* DSM 43827 | 67/77 | |
| *orf4* | + | 4561-4743 | 60 | Conserved hypothetical protein (177 aa) | *Streptomyces sviceus* ATCC 29083 | 32/58 | Unknown |
| *orf5* | + | 5104-5526 | 140 | Rare lipoprotein A (140 aa) | *Streptomyces violaceusniger* Tu 4113 | 75/86 | Lipoprotein |
| | | | | Expansin family protein (143 aa) | *Streptomyces bingchenggensis* BCW-1 | 70/82 | |
| *orf6* | + | 5523-6239 | 238 | Hypothetical protein ShygA5_02020 (212 aa) | *Streptomyces hygroscopicus* ATCC 53653 | 65/82 | Unknown |
| | | | | Superoxide dismutase (239 aa) | *Desulfovibrio desulfuricans* subsp. *desulfuricans* str. G20 | 34/50 | |
| | | | | Aminotransferase (504 aa) | *Salinibacter ruber* DSM 13855 | 37/48 | |
| *orf7* | + | 6253-7131 | 292 | Hypothetical protein StrviDRAFT_4987 (300 aa) | *Streptomyces violaceusniger* Tu 4113 | 53/65 | Unknown |
| *orf8* | - | 7246-8772 | 508 | Subtilisin-like protease (511 aa) | *Streptomyces avermitilis* MA-4680 | 67/80 | Protease |
| *orf9* | + | 9040-9507 | 155 | Transcriptional regulator (154 aa) | *Streptomyces coelicolor* A3(2) | 86/92 | Transcriptional regulator |
| | | | | MarR family transcriptional regulator (155 aa) | *Streptomyces bingchenggensis* BCW-1 | 87/92 | |

"i" indicates an incomplete *orf*, "aa" refers to the length of the polypeptide in amino acids; for strand, "+" indicates the direct strand, while "-" indicates the complementary strand. % ID represents the percent of residues identical to the homolog while % Sim indicates the number of similar or identical residues in the same position as the homolog

Table 3.4, continued

| Gene ID | Strand | Position | aa | Homolog(s) (aa) | Species | %ID/% Sim | Proposed Protein Function |
|---|---|---|---|---|---|---|---|
| *orf10* | + | 9721-10929 | 402 | Transmembrane efflux protein (397 aa) | *Streptomyces bingchenggensis* BCW-1 | 87/92 | Membrane channel export protein |
| *orf11* | + | 10955-11323 | 122 | Hypothetical protein SBI_08226 (115 aa) | *Streptomyces bingchenggensis* BCW-1 | 82/91 | Unknown |
| *orf12* | + | 11341-12180 | 279 | Putative oxidoreductase (276 aa) | *Streptomyces bingchenggensis* BCW-1 | 85/90 | Oxidoreductase |
| *orf13* | + | 12606-13691 | 361 | Integral membrane protein (377 aa) | *Streptomyces pristinaespiralis* ATCC 25486 | 57/70 | Putative serine/threonine phosphatase; membrane protein |
| | | | | Protein serine/threonine phosphatase (400 aa) | *Catenulispora acidiphila* DSM 44928 | 45/63 | |
| *orf16* | + | 14007-15200 | 397 | Conserved hypothetical protein (400 aa) | *Streptomyces pristinaespiralis* ATCC 25486 | 62/73 | Unknown |
| *orf17* | - | 15386-15865 | 159 | Predicted protein (440 aa) | *Streptomyces sviceus* ATCC 29083 | 51/62 | Unknown |
| *orf18* | + | 16299-18329 | 676 | Putative ABC transport protein (648 aa) | *Streptomyces* sp. SPB78 | 81/89 | ABC tranporter |
| *orf19* | - | 18525-19340 | 271 | Aldo/keto reductase (345 aa) | *Saccharopolyspora erythraea* NRRL 2338 | 76/86 | Oxidoreductase |
| *orf20* | - | 19782-21623 | 613 | Asparagine synthase (glutamine hydrolyzing) (613 aa) | *Rhodococcus jostii* RHA1 | 75/85 | Asparagine synthase |
| *orf21* | - | 21999-22433 | 144 | Putative IS630 family insertion sequence (363 aa) | *Streptomyces scabiei* 87.22 | 89/91 | Transposase sequence |
| *orf22* | - | 22511-23263 | 250 | Hypothetical protein Noca_2884 (226 aa) | *Nocardioides* sp. JS614 | 67/76 | Reductase [2Fe-2S] |
| | | | | Ferredoxin (217 aa) | *Saccharomonospora viridis* DSM 43017 | 42/53 | |

"i" indicates an incomplete *orf*, "aa" refers to the length of the polypeptide in amino acids; for strand, "+" indicates the direct strand, while "-" indicates the complementary strand. % ID represents the percent of residues identical to the homolog while % Sim indicates the number of similar or identical residues in the same position as the homolog

Table 3.4, continued

| Gene ID | Strand | Position | aa | Homolog(s) (aa) | Species | %ID/% Sim | Proposed Protein Function |
|---|---|---|---|---|---|---|---|
| *orf23* | - | 23260-24030 | 256 | ABC-type cobalamin/$Fe^{3+}$-siderophore transport system, ATPase component (251 aa) | *Saccharomonospora viridis* DSM 43017 | 57/70 | Iron uptake ABC transporter; ATPase component |
| *orf24* | - | 24072-25160 | 362 | ABC-type $Fe^{3+}$-siderophore transport system, permease component (354 aa) | *Saccharomonospora viridis* DSM 43017 | 61/76 | Iron uptake ABC transporter; permease component |
| *orf25* | - | 25157-26209 | 350 | ABC-type $Fe^{3+}$-hydroxamate transport system, periplasmic component (332 aa) | *Saccharomonospora viridis* DSM 43017 | 56/70 | Iron uptake ABC transporter; periplasmic component |
| *orf26* | + | 26629-27684 | 351 | Major facilitator superfamily MFS_1 (446 aa) | *Streptomyces* sp. ACTE | 81/84 | Tranporter protein |
| *orf27* | + | 27681-28118 | 145 | Pyridoxal-dependent decarboxylase (493 aa) | *Streptomyces ghanaensis* ATCC 14672 | 79/82 | PLP-dependent decarboxylase |
| *orf28* | + | 28422-29666 | 414 | Major facilitator superfamily MFS_1 (418 aa) | *Bacillus cereus* 172560W | 34/57 | Transmembrane transporter protein |
| *orf29* | - | 29953-31230 | 425 | Putative export protein (434 aa) | *Streptomyces hygroscopicus* ATCC 53653 | 53/68 | Transmembrane transporter protein |
| *orf30* | + | 31749-35630 | 1293 | Thaxtomin synthetase B (1505 aa) | *Streptomyces turgidiscabies* Car8 | 48/59 | NRPS (A-MT-T) |
| *orf31* | + | 35694-36740 | 349 | Luciferase-like, subgroup (357 aa) | *Streptomyces violaceusniger* Tu 4113 | 57/69 | MonoOx domain |
| | | | | Nonribosomal peptide synthetase (3445 aa) | *Sorangium cellulosum* 'So ce 56' | 56/67 | |
| *orf32* | + | 36773-44635 | 2620 | Amino acid adenylation (4836 aa) | *Streptomyces violaceusniger* Tu 4113 | 46/58 | NRPS (C-A-T-C-A-T-E) |
| *orf33* | + | 44632-48546 | 1304 | Putative NRPS (1323 aa) | *Streptomyces ghanaensis* ATCC 14672 | 44/56 | NRPS (C-A-T-TE) |
| *orf34* | - | 48720-49499 | 259 | Hypothetical protein Caci_4314 (494 aa) | *Catenulispora acidiphila* DSM 44928 | 55/66 | Transport protein |
| | | | | Cation diffusion facilitator family transporter (305 aa) | *Pyrobaculum arsenaticum* DSM 13514 | 28/47 | |

"i" indicates an incomplete *orf*, "aa" refers to the length of the polypeptide in amino acids; for strand, "+" indicates the direct strand, while "-" indicates the complementary strand. % ID represents the percent of residues identical to the homolog while % Sim indicates the number of similar or identical residues in the same position as the homolog

Table 3.4, continued

| Gene ID | Strand | Position | aa | Homolog(s) (aa) | Species | %ID/% Sim | Proposed Protein Function |
|---|---|---|---|---|---|---|---|
| *orf35* | + | 49527-49772 | 82 | Putative IS1647-like transposase (298 aa) | *Streptomyces griseus* subsp. *griseus* NBRC 13350 | 62/81 | Transposase (partial) |
| *orf36* | - | 49805-50443 | 212 | Putative LysE family protein (209 aa) | *Kitasatospora setae* KM-6054 | 60/69 | Amino acid transporter |
| *orf37* | - | 50555-50824 | 89 | Homoserine/threonine efflux protein (208 aa) | *Streptomyces pristinaespiralis* ATCC 25486 | 84/89 | Amino acid exporter protein |
| *orf38* | - | 50916-52250 | 444 | Sulfate adenylyltransferase subunit 1 (444 aa) | *Streptomyces scabiei* 87.22 | 85/91 | Sulfate adenylyl transferase subunit 1 |
| *orf39* | - | 52253-53188 | 311 | Sulfate adenylyltransferase subunit 2 (311 aa) | *Streptomyces avermitilis* MA-4680 | 89/93 | Sulfate adenylyl transferase subunit 2 |
| *orf40* | - | 53185-53793 | 202 | Phosphoadenylyl-sulfate reductase (198 aa) | *Streptomyces* sp. SPB74 | 74/82 | Phosphoadenylyl-sulfatereductase |
| *orf41* | - | 53793-55904 | 703 | NikT protein (421 aa) | *Micromonospora sp.* ATCC 39149 | 64/71 | Aminotransferase |
| | | | | Aminotransferase (348 aa) | *Nocardiopsis dassonvillei subsp. dassonvillei* DSM 43111 | 52/65 | |
| *orf41.5* | - | 54749-54892 | 47 | NikT protein (421 aa) | *Micromonospora sp.* ATCC 39149 | 69/78 | |
| *orf42* | - | 55973-56230 | 85 | Phosphopantetheine-binding (85 aa) | *Frankia* sp. EAN1pec | 45/68 | Acyl carrier protein |
| *orf43* | - | 56261-57298 | 345 | 4-Hydroxy-2-oxovalerate (354 aa) | *Micromonospora* sp. ATCC 39149 | 77/88 | Carboxyl transferase |
| *orf44* | - | 57295-58188 | 297 | NikA protein (298 aa) | *Micromonospora* sp. ATCC 39149 | 67/79 | Acetaldehyde dehydrogenase |
| | | | | Acetaldehyde dehydrogenase (296 aa) | *Streptomyces ansochromogenes* | 60/76 | |

"i" indicates an incomplete *orf*, "aa" refers to the length of the polypeptide in amino acids; for strand, "+" indicates the direct strand, while "-" indicates the complementary strand. % ID represents the percent of residues identical to the homolog while % Sim indicates the number of similar or identical residues in the same position as the homolog

Table 3.4, continued

| Gene ID | Strand | Position | aa | Homolog(s) (aa) | Species | %ID/% Sim | Proposed Protein Function |
|---|---|---|---|---|---|---|---|
| *orf45* | - | 58238-59068 | 276 | 2-Hydroxypenta-2,4-dienoate hydratase (301 aa) | *Micromonospora* sp. ATCC 39149 | 67/74 | Hydrolase |
| *orf46* | - | 59664-60746 | 360 | Streptomycin biosynthesis operon regulator (350 aa) | *Streptomyces griseus* subsp. *griseus* NBRC 13350 | 46/56 | Transcriptional regulator protein |
| *orf47* | - | 60886-61248 | 120 | Transthyretin family protein (108 aa) | *Mycobacterium abscessus* ATCC 19977 | 49/63 | Hydrolase monomer |
| *orf48* | - | 61998-62327 | 109 | Transthyretin (136 aa) | *Pseudomonas fluorescens* Pf0-1 | 30/50 | Hydrolase monomer |
| *orf49* | - | 62478-65546 | 1022 | Transcriptional regulator, SARP family (1022 aa) | *Streptosporangium roseum* DSM 43021 | 46/62 | Transcriptional regulator |
| *orf50* | - | 65734-68499 | 921 | ATPase-like protein (925 aa) | *Streptosporangium roseum* DSM 43021 | 51/65 | Transcriptional regulator |
| | | | | Regulatory protein, LuxR (943 aa) | *Salinispora tropica* CNB-440 | 39/53 | |
| *orf51* | - | 68897-69124 | 75 | Putative MbtH-like protein (75 aa) | *Streptomyces ambofaciens* | 71/81 | MbtH-like protein |
| *orf52* | - | 69264-70556 | 430 | L-Lysine-ε aminotransferase Lat (446 aa) | *Mycobacterium marinum* M | 57/70 | L-Lysine aminotransferase (ε amino group) |
| *orf53* | - | 70558-71202 | 214 | HAD-superfamily hydrolase, subfamily IA, variant (218 aa) | *Meiothermus silvanus* DSM 9946 | 42/54 | Hydrolase |
| *orf54* | - | 71202-72092 | 296 | Spore coat polysaccharide synthesis (282 aa) | *Oceanobacillus iheyensis* HTE831 | 36/53 | dTDP-4-dehydrorhamnose reductase |
| | | | | DTDP-4-dehydrorhamnose reductase (282 aa) | *Bacillus thuringiensis serovar monterrey* BGSC 4AJ1 | 35/53 | |
| *orf55* | - | 72092-72766 | 224 | Dimethylmenaquinone methyltransferase (228 aa) | *Vibrio shilonii* AK1 | 42/62 | Methyltransferase |
| *orf56* | - | 72842-74776 | 644 | Transketolase central region (636 aa) | *Rhizobium leguminosarum* bv. *trifolii* WSM1325 | 50/64 | Transketolase |

"i" indicates an incomplete *orf*, "aa" refers to the length of the polypeptide in amino acids; for strand, "+" indicates the direct strand, while "-" indicates the complementary strand. % ID represents the percent of residues identical to the homolog while % Sim indicates the number of similar or identical residues in the same position as the homolog

Table 3.4, continued

| Gene ID | Strand | Position | aa | Homolog(s) (aa) | Species | %ID/% Sim | Proposed Protein Function |
|---|---|---|---|---|---|---|---|
| *orf57* | - | 74803-75564 | 253 | Oxidoreductase, short chain dehydrogenase (272 aa) | *Rhodospirillum centenum* SW | 47/60 | Oxidoreductase |
| *orf58* | - | 75596-76663 | 355 | *N*-acetylneuraminic acid synthase (365 aa) | *Rhodospirillum centenum* SW | 55/69 | Neuramic acid synthase |
| *orf59* | - | 76669-77997 | 442 | Putative transcriptional regulator, GntR family (427 aa) | *Kribbella flavida* DSM 17836 | 63/74 | Transcriptional regulator or aminotransferase |
| | | | | Aminotransferase class I and II (441 aa) | *Micromonospora sp.* ATCC 39149 | 63/74 | |
| *orf60* | - | 77994-79103 | 369 | FMN-dependent alpha-hydroxy acid dehydrogenase (376 aa) | *Actinosynnema mirum* DSM 43827 | 59/71 | FMN-dependent dehydrogenase |
| *orf61* | - | 79100-80137 | 345 | 4-hydroxyphenylpyruvate dioxygenase (354 aa) | *Streptomyces roseosporus* NRRL 15998 | 50/60 | 4-Hydroxyphenylpyruvate dioxygenase |
| *orf62* | + | 80356-81858 | 500 | AMP-dependent synthetase and ligase (519 aa) | *Catenulispora acidiphila* DSM 44928 | 45/57 | NRPS type II A domain |
| *orf63* | + | 81855-82979 | 374 | *O*-methyltransferase (340 aa) | *Streptomyces tubercidicus* | 37/53 | *O*-methyltransferase |
| *orf64* | + | 82986-84251 | 421 | Cytochrome P450 monooxygenase (405 aa) | *Saccharopolyspora erythraea* NRRL 2338 | 45/60 | Cytochrome P450 |
| | | | | NikQ protein (396 aa) | *Streptomyces tendae* | 36/53 | |
| *orf65* | + | 84275-84511 | 78 | Nonribosomal peptide synthetase modules and related protein-like protein (2720 aa) | *Streptosporangium roseum* DSM 43021 | 64/69 | NRPS type II T domain |
| *orf66* | - | 85070-85570 | 166 | Secreted protein (169 aa) | *Streptomyces lividans* TK24 | 39/53 | Secreted protein |
| *orf67* | + | 85881-89828 | 1315 | WD-40 repeat-containing protein (1303 aa) | *Streptomyces roseosporus* NRRL 11379 | 49/60 | Transcriptional regulator |

"i" indicates an incomplete *orf*, "aa" refers to the length of the polypeptide in amino acids; for strand, "+" indicates the direct strand, while "-" indicates the complementary strand. % ID represents the percent of residues identical to the homolog while % Sim indicates the number of similar or identical residues in the same position as the homolog

Table 3.4, continued

| Gene ID | Strand | Position | aa | Homolog(s) (aa) | Species | %ID/% Sim | Proposed Protein Function |
|---|---|---|---|---|---|---|---|
| *orf68* | - | 90004-91317 | 437 | ATP/GTP-binding protein (444 aa) | *Streptomyces clavuligerus* ATCC 27064 | 40/55 | Membrane protein / peptidase |
| *orf69* | - | 91489-91740 | 83 | Conserved hypothetical protein (80 aa) | *Streptomyces* sp. SPB74 | 45/56 | Unknown |
| *orf70* | - | 91834-93063 | 409 | PE-PGRS family protein (410 aa) | *Streptomyces* sp. e14 | 70/80 | Transcriptional regulator |
| | | | | Transcriptional regulator (420 aa) | *Streptomyces clavuligerus* ATCC 27064 | 48/64 | |
| *orf71* | - | 93130-94932 | 600 | Amino acid adenylation (1345 aa) | *Anabaena variabilis* ATCC 29413 | 42/59 | NRPS subunit |
| *orf72* | - | 93068-95395 | 775 | Hypothetical protein bcere0007_52480 (2787 aa) | *Bacillus cereus* AH621 | 60/75 | Seems to be frameshift linking orf71 and orf73 |
| *orf73* | - | 93068-100411 | 1801 | Amino acid adenylation domain protein (4836 aa) | *Streptomyces violaceusniger* Tu 4113 | 47/60 | NRPS subunit |
| *orf74* | + | 100435-103052 (i) | 871 | Peptide synthetase (4898 aa) | *Streptomyces lavendulae* | 49/60 | NRPS subunit |

"i" indicates an incomplete *orf*, "aa" refers to the length of the polypeptide in amino acids; for strand, "+" indicates the direct strand, while "-" indicates the complementary strand. % ID represents the percent of residues identical to the homolog while % Sim indicates the number of similar or identical residues in the same position as the homolog

**Nonribosomal peptide synthetase domain analysis and substrate specificity**

Module and domain analysis was performed identically to that of NRPS2. This analysis indicates that NRPS3 is composed of four subunits. NRPS3A is predicted to have A-MT-T domains and activate L-phenylalanine whose α-amino group would be methylated by this subunit. NRPS3B shows significant similarity to MonoOx domains, but it is a discrete subunit. A MonoOx domain could hydroxylate a saturated carbon as it does in the myxathiazol and melithiazol systems (*20,31*). If the MonoOx domain acts on the glycine residue, it would facilitate a glyoxal leaving group as it does in the aforementioned systems, but it could also hydroxylate an aliphatic carbon on valine. NRPS3C is a dimodular subunit of C-A-T-C-A-T-E domain organization and is predicted to activate L-Val and Gly. The third module which activates Gly has an E domain; it is unclear which substrate the E domain epimerizes since Gly is achiral. NRPS3D is a single module subunit which is predicted to activate D-lysergic acid of domain organization C-A-T-TE. Table 3.5 shows the proteins used for sequence alignments. Figure 3.16 shows the predicted partial peptide backbone for the NRPS3 system. Figures 3.17 through 3.25 show the various alignments for the subunits. Table 3.6 gives the alignment of residues thought to dictate substrate specificity within the A domain.

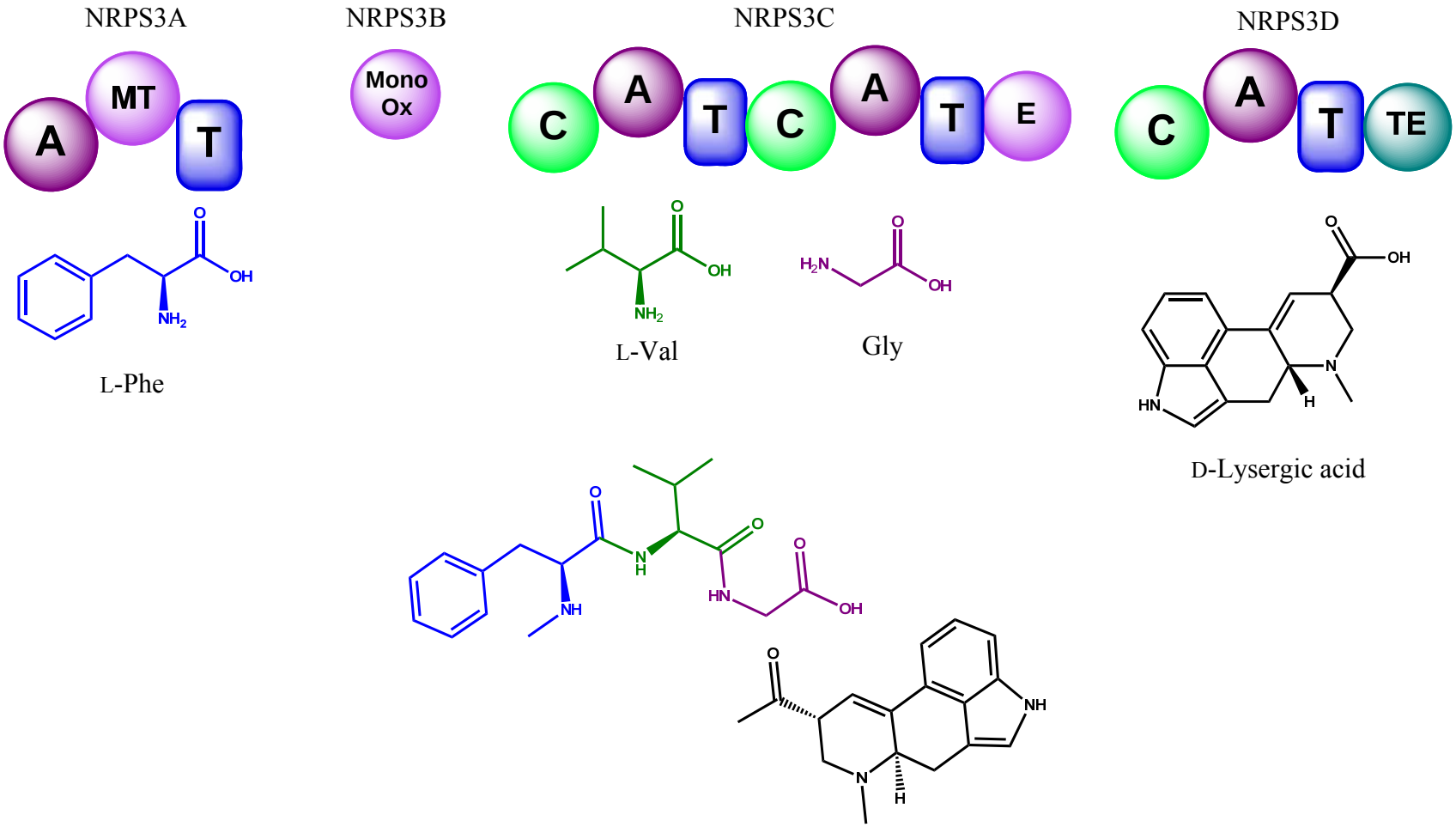



**Figure 3.16** NRPS3 domains, predicted substrates, and a proposed peptide backbone.

Table 3.5 Proteins used for sequence alignments with NRPS3 (*62*)

| Abbreviated Name | Name | Species |
|---|---|---|
| ACMS III | Actinomycin synthetase III | *Streptomyces anulatus* |
| BacA | Bacitracin synthetase 1 | *Bacillus licheniformis* |
| CDAI | Calcium-dependent antibiotic peptide synthetase I | *Streptomyces coelicolor* A3(2) |
| CDAII | Calcium-dependent antibiotic peptide synthetase II | *Streptomyces coelicolor* A3(2) |
| CDAIII | Calcium-dependent antibiotic peptide synthetase III | *Streptomyces coelicolor* A3(2) |
| CtaG | Cystothiazole A synthetase G | *Cystobacter fuscus* |
| EntF | Enterobactin synthetase F | *Escherichia coli* BL21(DE3) |
| GrsA | Gramicidin S synthetase 1 | *Aneurinibacillus migulanus* |
| GrsB | Gramicidin S synthetase 2 | *Aneurinibacillus migulanus* |
| LPS2 | Ergopeptine synthetase 2 | *Claviceps purpurea* |
| LtxA | Lyngbyatoxin synthetase A | *Lyngbya majuscula* |
| MelG | Melithiazol synthetase G | *Melittangium lichenicola* |
| MtaG | Myxothiazole synthetase G | *Stigmatella aurantiaca DW4/3-1* |
| MycB | Mycosubtilin synthetase B | *Bacillus subtilis* |
| PrisI-3,4 | Pristinamycin synthetase 3,4 | *Streptomyces pristinaespiralis* |
| PseurotinA | Pseurotin A synthetase | *Aspergillus fumigatus* |
| PstC | Friulimicin synthetase C | *Actinoplanes friuliensis* |
| TxtA | Thaxtomin A synthetase A | *Streptomyces acidiscabies* |
| TxtB | Thaxtomin A synthetase B | *Streptomyces acidiscabies* |
| VirS | Virginiamycin S synthetase | *Streptomyces virginiae* |


**Figure 3.17** ClustalW alignments and conserved motifs A1 to A5 within the A domains of NRPS3A, NRPS3C, and NRPS3D (*36,61*). M# (e.g. M1, M2, etc) refers to the order of the module within a subunit. Brackets demark the conserved motifs and blue font indicates the residues that are strictly conserved. BacA is a subunit from the bacitracin synthetase of *Bacillus licheniformis* (*82*). CDAII is a subunit from the calcium dependent antibiotic synthetase of *Streptomyces coelicolor* A3(2)(*89*). CpaA is a subunit from the cyclopiazonic acid synthetase of *Aspergillus flavus* (*90*). PseurotinA is a subunit from the pseurotin A synthetase of *Aspergillus fumigatus* (*91*). EntF is a subunit from the enterobactin synthetase of *Escherichia coli* (*86*). GrsA and GrsB are subunits from the gramicidin synthetase of *Aneurinibacillus migulanus* (*92,93*).

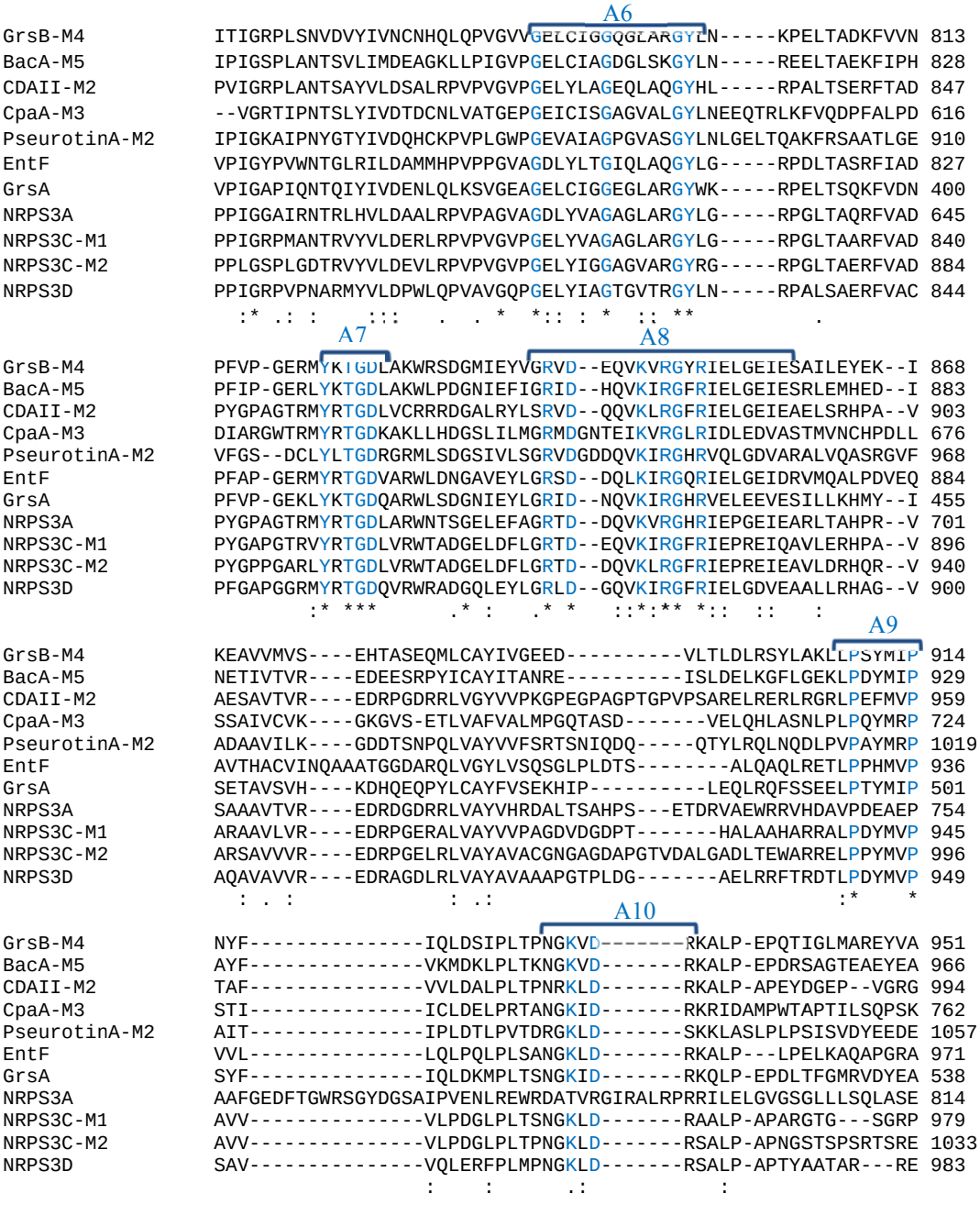



**Figure 3.18** ClustalW alignments and conserved motifs A6 to A8 within the A domains of NRPS3A, NRPS3C, and NRPS3D, and A9 and A10 for NRPS3C and NRPS3D (*36,61*). NRPS3A has an MT domain that sequentially interrupts the A domain between motifs A8 and A9. M# (e.g. M1, M2, etc) refers to the order of the module within a subunit. Brackets demark the conserved motifs and blue font indicates the residues that are strictly conserved. BacA is a subunit from the bacitracin synthetase of *Bacillus licheniformis* (*82*). CDAII is a subunit from the calcium dependent antibiotic synthetase of *Streptomyces coelicolor* A3(2)(*89*). CpaA is a subunit from the cyclopiazonic acid synthetase of *Aspergillus flavus* (*90*). PseurotinA is a subunit from the pseurotin A synthetase of *Aspergillus fumigatus* (*91*). EntF is a subunit from the enterobactin synthetase of *Escherichia coli* (*86*). GrsA and GrsB are subunits from the gramicidin synthetase of *Aneurinibacillus migulanus* (*92,93*).

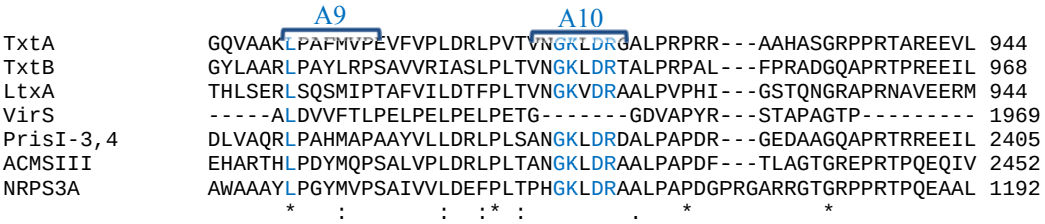



**Figure 3.19** ClustalW alignment and conserved motifs A9 and A10 for NRPS3A (*36,61*). NRPS3A has a MT domain which sequentially interrupts the A domain between motifs A8 and A9. Brackets demark the conserved motifs and blue font indicates the residues that are strictly conserved within a motif. TxtA and TxtB are subunits from the thaxtomin A synthetase of *Streptomyces acidiscabies* (*94*). LtxA is a subunit from the lyngbyatoxin synthetase of *Lynghya masjuscula* (*95*). VirS is a subunit from the virginiamycin S synthetase of *Streptomyces virginiae* (*96*). PrisI-3,4 is a subunit from the pristinamycin synthetase of *Streptomyces pristinaespiralis* (*97*). ACMSIII is a subunit from the actinomycin synthetase of *Streptomyces anulatus* (*98*).

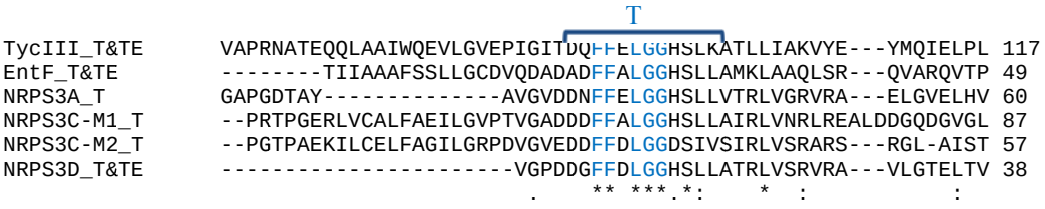



**Figure 3.20** ClustalW alignment and conserved motif for the T domain of NRPS3A, NRPS3C, and NRPS3D (*36,61*). T and TE indicate that only the T and/or TE domains were included in the alignment. Brackets demark the conserved motifs and blue font indicates the residues that are strictly conserved within a motif. EntF is a subunit from the enterobactin synthetase of *Escherichia coli* BL21(DE3) (*86*). TycIII is a subunit from the tyrocidine synthetase of *Brevibacillus brevis* (*80*).

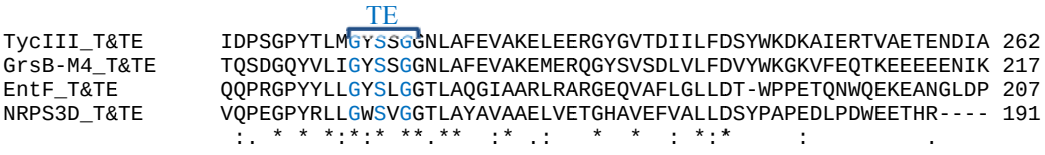



**Figure 3.21** ClustalW alignment and conserved motif for the TE domain of NRPS3D (*36,61*). T and TE indicate that only the T and TE domains were included in the alignment. Brackets demark the conserved motifs and blue font indicates the residues that are strictly conserved within a motif. EntF is a subunit from the enterobactin synthetase of *Escherichia coli* (*86*). GrsB is a subunit from the gramicidin synthetase of *Aneurinibacillus migulanus* (*92,93*). TycIII is a subunit from the tyrocidine synthetase of *Brevibacillus brevis* (*80*).

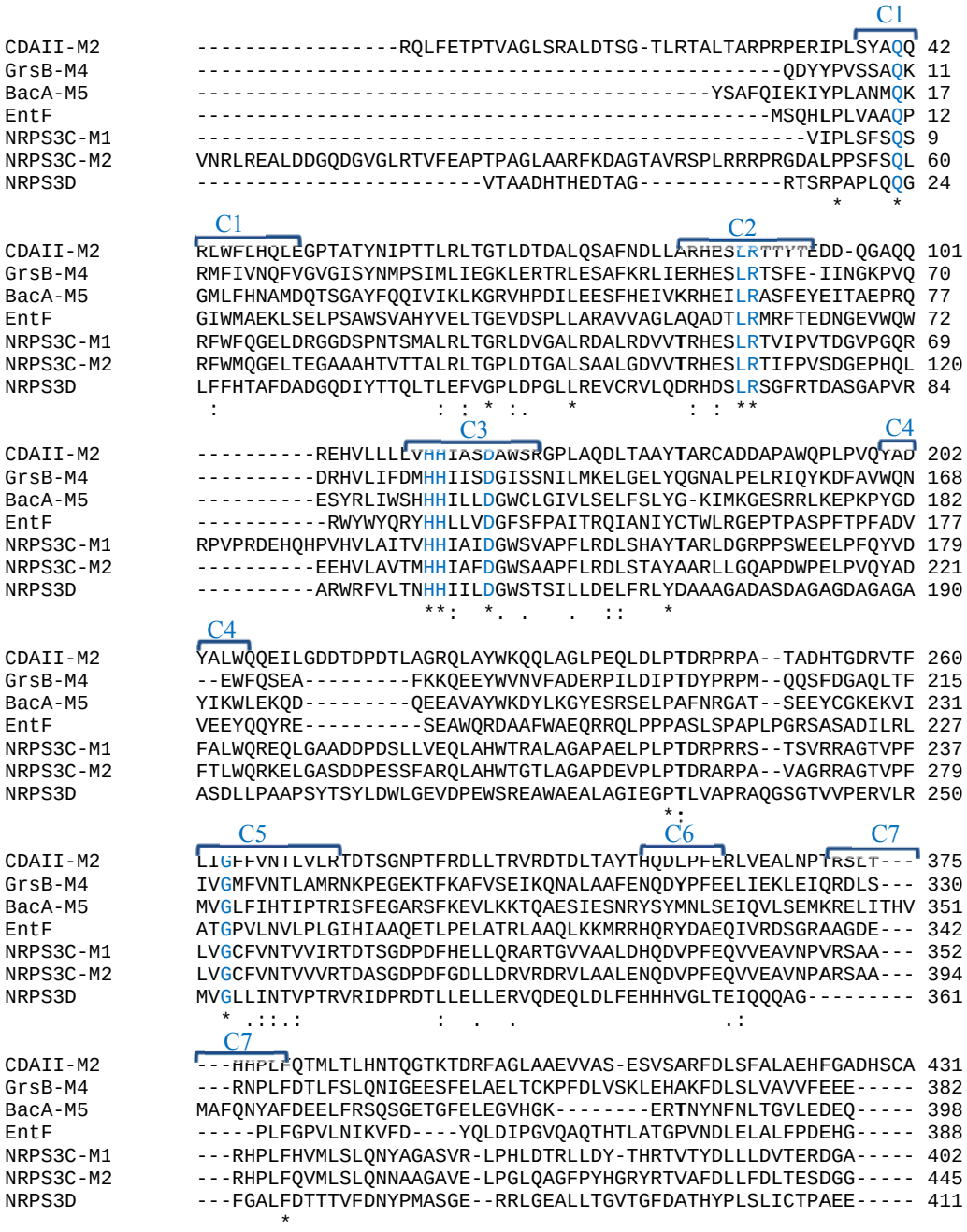



**Figure 3.22** ClustalW alignment and conserved motifs for the C domains in NRPS3C and NRPS3D (*36*). M# (e.g. M1, M2, etc) refers to the order of the module within a subunit. Brackets demark the conserved motifs and blue font indicates the residues that are strictly conserved. BacA is a subunit from the bacitracin synthetase of *Bacillus licheniformis* (*82*). CDAII is a subunit from the calcium dependent antibiotic synthetase of *Streptomyces coelicolor* A3(2)(*89*). EntF is a subunit from the enterobactin synthetase of *Escherichia coli* BL21(DE3) (*86*). GrsB is a subunit from the gramicidin synthetase of *Aneurinibacillus migulanus* (*92,93*).


**Figure 3.23** ClustalW alignment and conserved motifs for the MT domain of NRPS3A (*36,61*). Brackets demark the conserved motifs and blue font indicates the residues that are strictly conserved within a motif. TxtA and TxtB are subunits from the thaxtomin A synthetase of *Streptomyces acidiscabies* (*94*). LtxA is a subunit from the lyngbyatoxin synthetase of *Lynghya masjuscula* (*95*). VirS is a subunit from the virginiamycin S synthetase of *Streptomyces virginiae* (*96*). PrisI-3,4 is a subunit from the pristinamycin synthetase of *Streptomyces pristinaespiralis* (*97*). ACMSIII is a subunit from the actinomycin synthetase of *Streptomyces anulatus* (*98*).

```
CtaG            ISFDISVLEILYSLTRGFQVVLRGEQTTHPPVRAGASKPLEFSLFYFASDER-ERPGDKY 778
MelG            MSFDISVLEIFYSLTRGFQVVLREERTTQQPVRAGASKPLEFSLFYFASDER-ERSGDKY 779
MtaG            MSFDISILEILFSLTRGFQVVIRGEQGAGLPVSAGHRKAPQFSLFYFASDER-ESTHGKY 771
NRPS3B          ------------------------------------MDFSLFYFADDSTPDGDAGRY 21
                                                     :*******.*.  :   .:*

CtaG            RLLLEGARFADEHGFTAVWTPERHFHSFGGVYPNPSVISAAIAATTRNLRIRAGSVVLPL 838
MelG            RLLMEGARFADEHGFTAVWTPERHFHSFGGIYPNPSVVSAAIAATTRNLRIRAGSVVLPL 839
MtaG            QLLLEGARFADQHGFTAVWTPERHFHPFGGIFPNPSVVSAALAATTRNIRIRAGSVVLPL 831
NRPS3B          ELLLEGARFADRHGFRAVWTPERHFHPFGGLYPNPSVVGAALAMVTNTVAIRAGSVVAPL 81
                .**:*******.*** **********.***::*****:.**:* .*..: ******* **

CtaG            HSPIRVAEEWSVVDNLSGGRVDLSIASGWHPNDFVLAPERYADARSQLFSQLDSVRKLWR 898
MelG            HSPIRVAEEWSIVDNLSGGRVDLSFASGWHPNDFVLAPERYAGARGQLMSQIEAFQKLWR 899
MtaG            HSPIRVAEEWSIVDNLSNGRVDLSFASGWHPNDFVLAPERFADARSGLAGQIQTFKKLWR 891
NRPS3B          HHPLRIAEEWSVVDNLSGGRAGVSLASGWHPVDFALGRVPYEDRKRALADGIEQLRGLWR 141
                * *:*:*****:*****.**..:*:****** **.*.   : . :  * . :: .: ***

CtaG            GEAVVFPNGLGQDVEVRTLPRPIQPDVAIWLTAAGNPETFRAAGERGLNVLTHLLGQNPA 958
MelG            GEAVVFPNGLGQDVEVRTLPRPIQPDVAIWLTAAGNPETFRAAGERGLNVLTHLLGQNLA 959
MtaG            GEKVNFRNGVGTDVAVQALPRPIQPDVAVWLTAAGNPETFRLAGELGTNILTHLLGQNLT 951
NRPS3B          GDGHEVVDGNGVPATVRIFPPPVQRELPLWVTSAGDAGTFRTAAATRAGVLTHLLHQDVD 201
                *:   . :* *  . *: :* *:* ::.:*:*:**:. *** *.    .:***** *:

CtaG            ELAKKIQIYRDAWKAAGHGPGEGHVTLMLHTFLGEDPEAVRQKVQGPLRQYLKNSVGLLR 1018
MelG            ELAKKIQIYRDAWKAAGHGPGAGHVTLMLHTFLGEDRGAVRQKIQGPLREYLKSSVGLLR 1019
MtaG            ELEKKIQIYRDAWKAAGHGPGEGHVTLMLHTFVGDDAAEVRQKVQGPLRQYLKSSVGLLK 1011
NRPS3B          ELAAKIAAYRRTARETHDG-WEGHVVLMLHTFLGTDRDEVRATVDGPLRAYLRSSVHLLA 260
                **  **  ** : : : .*   ***.******:* *   ** .::**** **:.** **

CtaG            SVIGPLQEGTEFESLGEADIELLLSRAIERYLDQMGLFGTVESCLPRVHQLRELGVDEIA 1078
MelG            SVIGPLPHGAEFESLSEADIDVLLSKAIERYFEQMGLFGTVESCLPMVHQLRELGVDEIA 1079
MtaG            TVIGPLAHGAEFESLSEADIDVLLSRAIERYQDQMGLFGTPESCLPMVAKLRDLGVDEIA 1071
NRPS3B          RSFGALDPDFDIDALEDEDLDFLVDQSFDTYFGQRGLFGTVEDAAPVVERLRGAGVDEIA 320
                  :*.*  . ::::* : *::.*:.:::: *  * ***** *.. * * :**  ******

CtaG            SLIDFGVDLESTLEGLRHLNALRERATSREEQEDLPTLVARHGVTHFQCTPSMLRMLLLE 1138
MelG            SLIDFGVDMESTLAGLQHLNALRERATSREEKEEIPELVARHGVTHFQCTPSMLRMLLME 1139
MtaG            CLIDFGVDRESTLAGLHHLNELRERSTQHGEPEDIPALVARHGVTHFQCTPSMLRMLLLE 1131
NRPS3B          CLIDFGIGTKTVLDGLHHLNTLREASAAH------------------------------ 349
                .*****:. ::.* **:*** *** :: :
```

**Figure 3.24** ClustalW alignment for the MonoOx domain of NRPS3B (*61*). No defined motifs for a MonoOx domain have yet been identified, but conserved residues appear throughout the domain sequence. CtaG is a subunit from the cystothiazole A synthetase of *Cystobacter fuscus* (*99*). MelG is a subunit from the melithiazol synthetase of *Melittangium lichenicola* (*20*). MtaG is a subunit from the myxothiazol synthetase of *Stigmatella aurantiaca* DW4/3-1 (*31*).

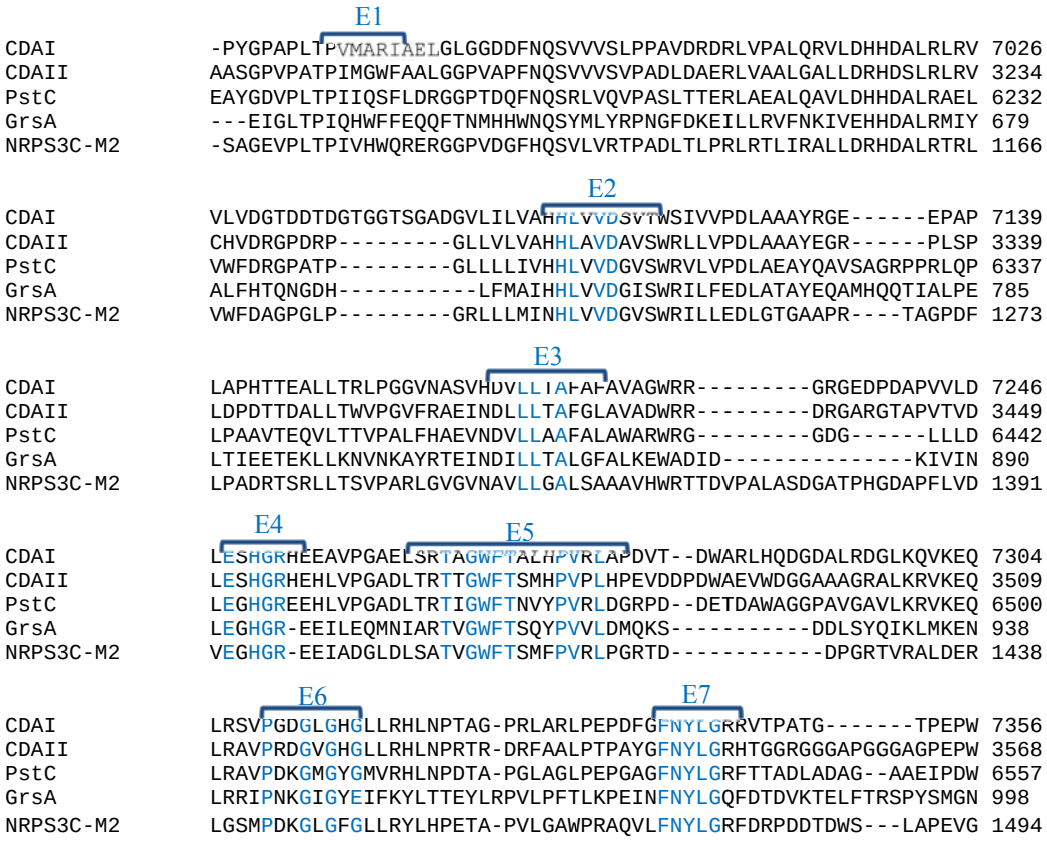



**Figure 3.25** ClustalW alignment and conserved motifs for the epimerization domain in NRPS3C (*36,61*). M# (e.g. M1, M2, etc) refers to the order of the module within a subunit. Brackets demark the conserved motifs and blue font indicates the residues that are strictly conserved within a motif. CDAI and CDAII are subunits from the calcium dependent antibiotic synthetase of *Streptomyces coelicolor* A3(2)(*89*). GrsA is a subunit from the gramicidin synthetase of *Aneurinibacillus migulanus* (*92,93*). PstC is a subunit from the fiulimicin synthetase of *Actinoplanes friuliensis* (*100*).

Table 3.6 Substrate specificity alignments for NRPS3 subunits A, C, and D

NRPS3A

| GrsA position | 235 | 236 | 239 | 278 | 299 | 301 | 322 | 330 | 331 | 517 |
|---|---|---|---|---|---|---|---|---|---|---|
| GrsA (L-Phe) (*92,93*) | D | A | W | T | I | A | A | I | C | K |
| TycA (L-Phe) (*80*) | D | A | W | T | I | A | A | I | C | K |
| TycB M2 (L-Phe) (*80*) | D | A | W | T | I | A | G | V | C | K |
| NRPS3A | D | A | W | T | V | A | A | V | C | K |
| NRPS3A position | 481 | 482 | 485 | 524 | 545 | 547 | 568 | 576 | 577 | 1161 |
| Prediction | L-Phe | | | | | | | | | |

NRPS3C module 1

| GrsA position | 235 | 236 | 239 | 278 | 299 | 301 | 322 | 330 | 331 | 517 |
|---|---|---|---|---|---|---|---|---|---|---|
| GrsB - M2 (L-Val) (*92,93*) | D | A | F | L | G | A | G | T | F | K |
| TycC - M4 (L-Val) (*80*) | D | A | F | L | G | A | G | T | F | K |
| LicB - M1 (L-Val) (*81*) | D | A | F | L | G | A | G | T | F | K |
| NRPS3C- M1 | D | A | Y | L | G | Q | V | T | F | K |
| NRPS3C-M1 position | 668 | 669 | 672 | 711 | 737 | 739 | 760 | 768 | 769 | 961 |
| Prediction | L-Val | | | | | | | | | |

NRPS3C module 2

| GrsA position | 235 | 236 | 239 | 278 | 299 | 301 | 322 | 330 | 331 | 517 |
|---|---|---|---|---|---|---|---|---|---|---|
| CDAII - M2 (Gly) (*89*) | D | I | L | Q | V | G | L | I | W | K |
| DhbF - M1 (Gly) (*101*) | D | I | L | Q | V | G | L | I | W | K |
| MelG (Gly) (*20*) | D | I | L | Q | V | G | M | I | W | K |
| NosC - M2 (Gly) (*102*) | D | I | L | Q | C | G | L | I | W | K |
| NRPS3C-M2 | D | I | L | Q | V | G | V | I | W | K |
| NRPS3C-M2 position | 1737 | 1738 | 1741 | 1780 | 1804 | 1806 | 1827 | 1835 | 1836 | 2032 |
| Prediction | Gly | | | | | | | | | |

NRPS3D

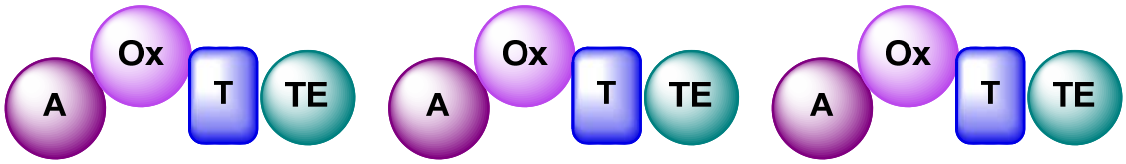
| GrsA position | 235 | 236 | 239 | 278 | 299 | 301 | 322 | 330 | 331 | 517 |
|---|---|---|---|---|---|---|---|---|---|---|
| CpaA (L-Trp) (*90*) | D | M | A | L | C | G | S | A | C | K |
| pseurotin A (L-Trp) (*91*) | D | A | Y | T | M | A | A | I | C | K |
| LPS2 (D-lysergic acid) (*103*) | D | V | F | S | V | G | L | V | M | K |
| NRPS3D | D | V | F | L | V | L | I | T | Y | K |
| NRPS3D position | 680 | 681 | 684 | 715 | 740 | 742 | 762 | 774 | 775 | 965 |
| LPS2 position | 454 | 455 | 458 | 495 | 516 | 518 | 539 | 546 | 547 | 768 |
| Prediction | D-lysergic acid | | | | | | | | | |

The amino acid in parentheses indicates which amino acid is activated by the specific A domain.
§Predictions based on Stachelhaus, Challis and Rausch (*34,37,38*)
.

# CHAPTER 4: DISCUSSION

## NRPS2

The study of the NRPS2 system required identifying the genetic locus, extensive bioinformatic analysis, gene cloning, protein expression, protein purification, and *in vitro* assays. The NRPS2 system consists of a single NRPS subunit of one module consisting of A, Ox, T, and TE domains (Figure 4.1). It produces an unidentified blue pigment, as evidenced by the protein expression cultures in *Escherichia coli*. The blue pigment is most likely indigoidine, but attempts to purify it have not yet met with success. Confirmation of the blue pigment's identity therefore remains a subject for future efforts.




NRPS2 (1305 aa)
*Streptomyces laurentii*

BpsA (1282 aa)
*Streptomyces lavendulae*
57% Identity, 69% Similarity

IndC (1488 aa)
*Erwinia chrysanthemi*
52% Identity, 67% Similarity

**Figure 4.1** Module and domain organization of NRPS2, BpsA, and IndC. The source organism and percent identity and similarity of NRPS2 to each protein are shown below each schematic (*6,7,55*).

A polymerase chain reaction (PCR)-based screen of the genomic fosmid library led to the identification of ten fosmids possessing *S. laurentii* genomic DNA fragment 764. From these fosmids, JA18C3 and JA4B8 were sequenced, analyzed and assembled to a 64920 bp fragment of the *S. laurentii* chromosome. Analysis of the resulting sequence revealed 66 *orf*s including

*orf34* (*nrps2*) and *orf42* (a putative PPTase). NRPS2 shows homology to two blue-pigment producing single-module NRPS enzymes that activate L-Gln: BpsA and IndC (*6,7*). The *nrps2* locus lacks a MbtH like protein. MbtH-like proteins play an ambiguous role in NRPS systems, but they are commonly found within NRPS clusters, and a recent study indicates they play an integral role in the NRPS systems with which they are associated by stabilizing A domain activity (*104,105*). Not all NRPS clusters encode an MbtH homolog. Indeed, the two closest homologs of NRPS2, IndC and BpsA, lack an MbtH-like protein within their loci (*6,7*).

More extensive bioinformatic analysis determined the domain organization in Figure 4.1, which matches that of BpsA and IndC, two indigoidine synthetases (*6,7*). Additionally, the core motifs highlighted in Figures 3.4 through 3.7 of each domain revealed conservative variations to some residues, but similar differences were observed in other amino acid sequences in characterized NRPS subunits. Significant variances exist between the Ox-2 motifs of BlmIII, MtaC, MtaD, and EpoB versus those found in IndC, BpsA, and NRPS2. The altered motifs, however, are consistent among IndC, BpsA, and NRPS2, and their presence could be due to a subtle difference in function. The analysis of the substrate specificity-conferring residues of NRPS2 does not match exactly with a set defined by Stachelhaus, *et al*., but it did match the residues from BpsA and IndC, which activate L-Gln. When analyzed by the algorithm provided by Rausch, *et al*, L-Gln is also predicted to be the substrate for NRPS2. Therefore, the prediction that L-Gln is the amino acid substrate for NRPS2 seems to be reliable (*6,7,37,38*).

Based on a predicted common substrate and the strong identity to BpsA and IndC, NRPS2 is predicted to produce indigoidine (*6,7*). Figure 4.2 shows a proposed reaction sequence (*6,7*). This reaction sequence requires the usual NRPS substrate activation and attachment to the holo-T domain, but also requires several oxidations by the Ox domain. The first two oxidations

proposed here are dehydrogenations to form a double bond between the α- and β-carbons on each L-Gln substrate. The next step is the carbon-carbon bond formation between the γ-carbon on each L-Gln, and a final dehydrogenation is required to introduce another double bond. While oxidation of activated single bonds to double bonds is common for Ox domains, oxidative coupling to form a new carbon-carbon bond is not. Thus there is unusual oxidative chemistry in these blue pigment synthetases (*30,83,85*). The oxidizing cofactor is assumed to be FMN, and the domain is likely an oxidase, meaning air-reoxidation can regenerate the cofactor (*30*). Confirmation of the cofactor's identity remains as future work, and further support for this reaction sequence could be provided by detailed UV-Vis spectroscopy of the FMN cofactor during catalysis, mutagenesis studies of the Ox domain, and crystal structure determination of this unusual Ox domain.

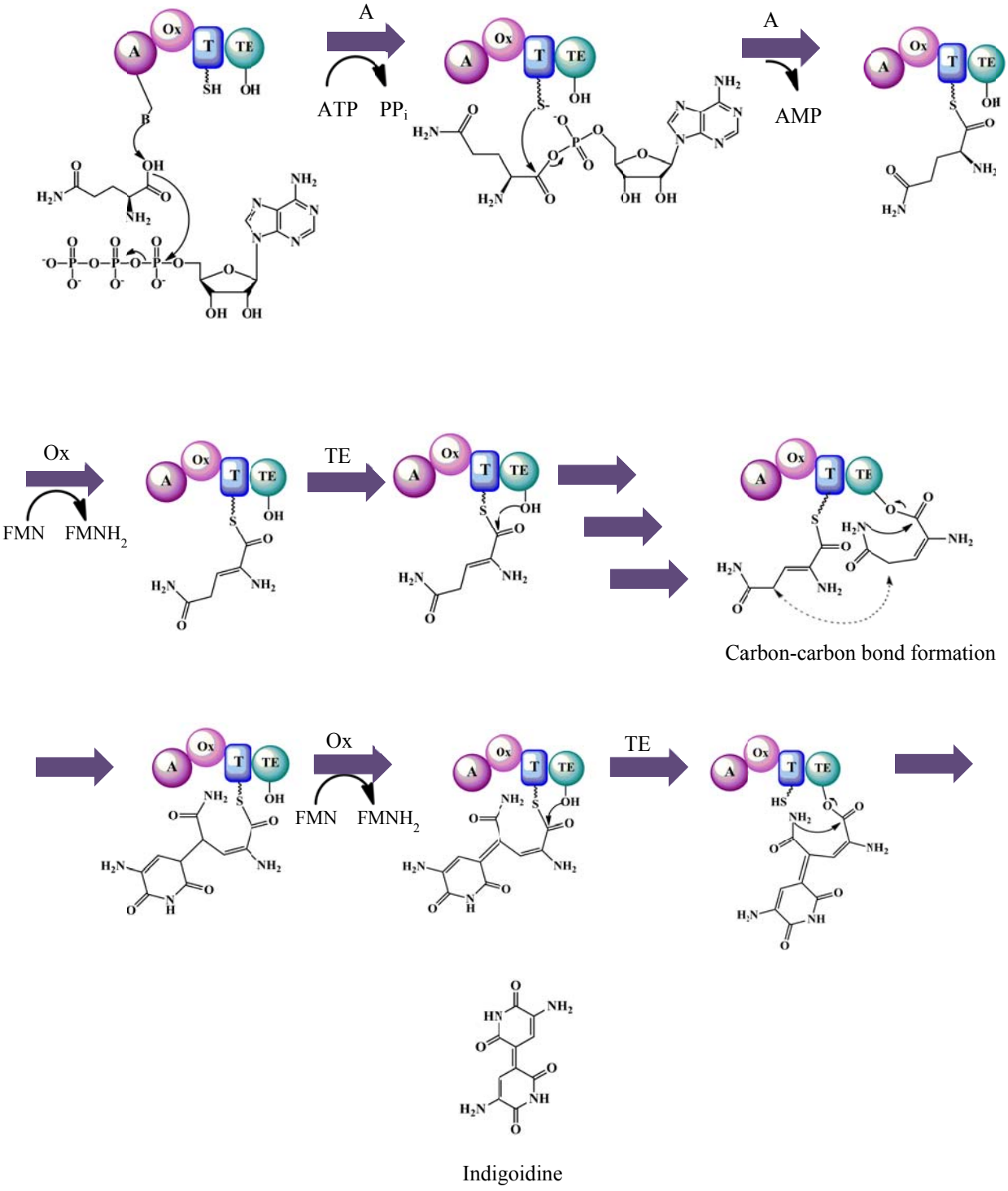



**Figure 4.2** A proposed reaction sequence for indigoidine production by NRPS2 (*6,7*).

Heterologous protein expression studies revealed that the N-terminal hexahistidine fusion construct of NRPS2 induced by 0.04 mM IPTG at 15 °C for 24 hours produced the greatest amount of soluble protein. An observation in keeping with the functional predictions of NRPS2 is the intense yellow color of the purified protein, consistent with a FMN cofactor. Another indicator of the function of NRPS2 was the intense blue color in the culture following protein expression. Since NRPS2 is similar to indigoidine synthetases and is predicted to require L-Gln, it is assumed that this blue pigment is indigoidine. A blue color was observed in the cultures expressing NRPS2 fused at the C-terminal to hexahistidine, but to a lesser degree. The blue pigment production during NRPS2 expression indicates that not only is the protein functional, but also that the *E. coli* PPTase can prime the T domain with the phosphopantetheinyl prosthetic group, an atypical activity when the substrate is an actinomycete NRPS T domain (*106,107*); this assumption will have to be confirmed by mass spectrometry (MS). The reduced production of blue pigment by cultures expressing the C-terminal hexahistidine fusion to NRPS2 could mean that the *E. coli* PPTase has not primed the T domain or that most of the protein produced was insoluble. No firm conclusions can be drawn from pigment production by the N-terminal or C-terminal hexahistidine fusion constructs of NRPS2 without completing the expression and solubility study. The bioinformatic analyses and the blue pigment produced by the protein strongly suggest that NRPS2 produces indigoidine from two L-Gln units. Without purified pigment, nuclear magnetic resonance (NMR) and mass spectrometry (MS) studies cannot be completed and the pigment's identity cannot be confirmed. However, using the purified protein, NRPS2 may be partially characterized by *in vitro* assays.

Initial results from the *in vitro* assays revealed an increase in absorbance at 600 nm that appears after five minutes and decreases at later time points. One reason causing this event could

be the production of hydrogen peroxide during the reoxidation of the FMN cofactor (*30*). Hydrogen peroxide could not only oxidize the cofactor, but can also oxidize both the product and the enzyme. These latter two oxidations would not only continuously degrade the blue pigment formed, leading to a loss of absorbance at 603 nm, but could also inactivate the NRPS2 and prevent further product formation. After testing to assess the levels of superoxide and hydrogen peroxide in the reaction, a solution to this problem may lie in the use of catalase in the *in vitro* assays to degrade the hydrogen peroxide produced during cofactor regeneration and tests (*30,108*). Future work will determine whether or not the inclusion of catalase will be beneficial for blue pigment formation by NRPS2.

Another question arises concerning blue pigment production in *S. laurentii*: how is production induced? Reverchon, *et al*, and others indicate that the PecS transcriptional regulator, a MarR type regulator, is involved in indigoidine biosynthesis in *Erwinia chrysanthemi* and *Streptomyces aureofaciens* CCM 3329 (*6,109*). Studying the transcriptional regulators within the *nrps2* locus for those similar to PecS could suggest ways in which to activate production in *S. laurentii.* By adding γ-nonalactone to cultures, Yanagimoto, *et al*, was able to produce a blue pigment in *Streptomyces virginiae* with identical properties to the blue pigment synthesized by NRPS2 (*8*). The DNA locus for the gene of *Streptomyces virginiae*'s blue pigment-producing enzyme is unavailable. Using the same medium and inducer molecule, however, Takahashi, *et al.* induced blue pigment production in *Streptomyces lavendulae* (*7*); *S. lavendulae*'s BpsA shares strong identity with NRPS2. Including γ-nonalactone in *S. laurentii* cultivated in Y6.5, as described by Yanagimoto, *et al*., resulted in a culture with a green pigment. It is noted that aged samples of the semi-purified blue pigment from *E. coli* also adopt a green hue, perhaps indicative of oxidation (data not included). Neither substance was ever analyzed, however, so no direct

conclusions can be made. When *S. laurentii* was cultivated on PGP solid medium, a purple substance was excreted into the agar from a confluent plaque (Figure 3.11). *S. laurentii* produces a red pigment on oatmeal agar, and the possibility exists that a red pigment and a blue pigment are simultaneously produced when the bacterium is cultivated on PGP medium. To support this postulate, the cultures would need to be repeated, and the pigments extracted and purified. Additionally, it could be that once an optimum temperature for enzyme activity is identified, this information could be applied to induce blue pigment production by *S. laurentii*.

The utility of the blue pigment remains to be determined. It is possible that, like other pigments, this pigment could serve as an anti-oxidant (*110*). Reverchon, *et al*, also suggests that indigoidine scavenges oxygen radicals, a capability contributes to the pathogenicity of *E. chrysanthemi* (*6*). *S. laurentii* produces thiostrepton, but possesses biosynthetic clusters which could produce other metabolites (*18,21*). Although a role for the blue pigment in *S. laurentii* is not currently known, the blue pigment could serve as an anti-oxidant for the bacterium.

In conclusion, NRPS2 produces a blue pigment thought to be indigoidine. To confirm the structure of the blue pigment, a purification protocol must be established to facilitate characterization by NMR and MS. To optimize the *in vitro* blue pigment production, adding catalase might be necessary to remove any hydrogen peroxide and prevent degradation of the product or the enzyme. A and Ox domain activity assays must be performed, and mutagenesis studies can be employed to confirm the function of the T and TE domains. Additionally, once the temperature for optimum enzyme activity is determined, blue pigment production in *S. laurentii* must be reattempted at that temperature, even if it means longer incubation periods due to slowed growth rates. Additionally, a crystal structure of this unusual Ox domain could facilitate understanding the chemistry it performs.

## NRPS3

Currently, the *nrps3* locus has been identified, and, from bioinformatic analyses, a backbone structure for the NRPS3 product has been predicted. The assembly of fosmids JA5F6, JA9D12, and JA15F8 resulted in a 103052 bp sequence. Analyzing that sequence revealed 74 *orf*s including *orf30, orf31, orf32*, and *orf33* which encode the NRPS subunits, NRPS3A, NRPS3B, NRPS3C, and NRPS3D. Additionally, *orf42* encodes a PPTase, while *orf51* encodes an MbtH-like protein often associated with NRPS systems (*104*). It was observed that a NRPS subunit, denoted as NRPS4A, lies at the end of the 103052 bp sequence assembled around the *nrps3* locus. It is unclear whether the MbtH-like protein and other proteins encoded within the DNA sequence are part of the NRPS3 biosynthetic gene cluster or that of NRPS4. Delineating the product resulting from the NRPS3 biosynthetic gene cluster is complicated by not knowing the exact boundaries of the locus.

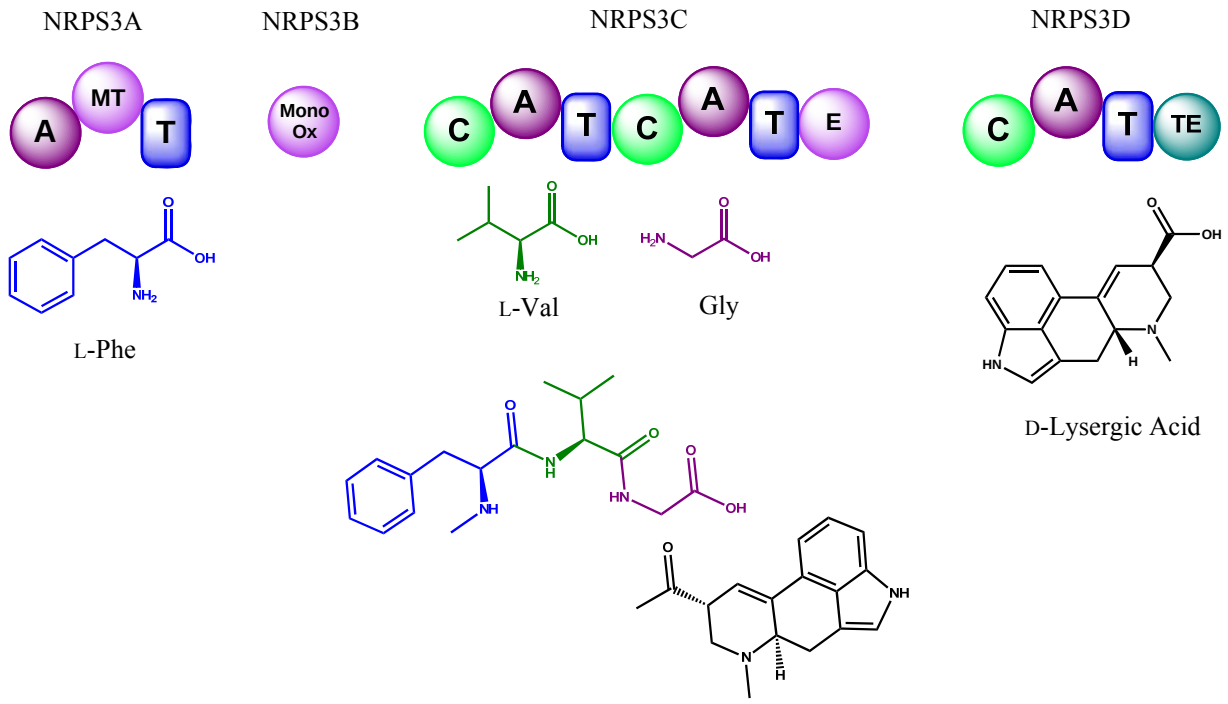



**Figure 4.3** NRPS3 domains and predicted substrates with a proposed backbone for subsequent modifications.

Bioinformatic analyses determined the module and domain organization of each NRPS homolog in addition to the conserved motifs and substrate specificity of each module. NRPS3A, NRPS3C, and NRPS3D encode subunits with $A_1$-$MT_1$-$T_1$, $C_2$-$A_2$-$T_2$-$C_3$-$A_3$-$T_3$-$E_3$, and $C_4$-$A_4$-$T_4$-TE domain organization respectively. NRPS3B is a discrete MonoOx domain and shows several residues conserved with characterized MonoOx domains. Variances exist in the conserved motifs of the NRPS3 domains, but those differences are conservative and exist in other characterized NRPS subunits. For the MT domain, the M3 motif is missing, but M1 and M2 are present and several other residues are highly conserved. The domain analysis appears to be firm, but the substrate specificity of each module's A domain requires deeper examination.

The three subunits, NRPS3A, NRPS3C, and NRPS3D, are predicted to activate L-Phe, L-Val, Gly, and D-lysergic acid respectively (Table 3.6). The substrate predictions show some variances from what is expected but a few reasons could explain those disparities. The substrate for the MonoOx domain is not known. It could hydroxylate a saturated carbon on a free valine, and the NRPS3C-M1 A domain could activate the resulting hydrolyated valine. If so, it could better explain the variances from the ten residues defined by Stachelahus, *et al*. in the alignment for the NRPS3C-M1 A domain. Additionally, the only known NRPS subunit to activate lysergic acid is from a fungal species (*103*). Specific enzymes are required to produce lysergic acid, but those enzymes are not encoded in the *nrps3* locus, as expected (*103,111*). Genes encoding lysergic acid biosynthetic enzymes could lie elsewhere in the *S. laurentii* genome, as do the amino acid biosynthetic genes, but this physical separation of biosynthetic genes is rare (*103,111*). Additionally, lysergic acid lacks the free amine required for peptide bond formation with the preceding substrate and lysergic acid. After considering these facts, the possibility exists that rather than activating lysergic acid, NRPS3D activates a yet-to-be-determined substrate.

Considering the ambiguity that arises from the bioinformatic analyses, triggering the expression of the NRPS3 system and determining the product's structure might prove easier than trying to predict a final product from the biosynthetic gene cluster alone. Future work for NRPS3 includes analyzing the transcriptional regulators encoded within the *nrps3* locus for clues on how to trigger the cluster's expression, structure determination of the cluster's metabolite, cloning of the genes for the individual subunits, and expression, purification, and characterization of these proteins.

## Conclusion

These two genome mining projects of *S. laurentii* remain incomplete; however the direction of the remaining work is known. The NRPS3 system is predicted to produce a novel peptide product. Once that product is produced and structurally characterized, the biosynthetic cluster can be delineated and fully explored, and biological assays can be used to determine the product's potency against a variety of diseases. On the other hand, since the blue pigment produced by NRPS2 is likely indigoidine, the NRPS2 system does not add a new metabolite to nature's arsenal, but further study on indigoidine could reveal medicinally useful characteristics of this NRPS product. Also, NRPS2 is another enzyme that produces a known metabolite. For both systems, comparison of the enzymes that can produce a particular compound or structure can illuminate critical structures and specific residues that are necessary for the chemistry that the enzymes or enzymatic domains perform. Once these structures and residues are identified, biosynthetic engineering may be employed to exploit that chemistry and to produce additional compounds that can aid in the fight against infectious disease.

# References


1. Caboche, S., Pupin, M., Leclere, V., Fontaine, A., Jacques, P., and Kucherov, G. (2008) NORINE: a database of nonribosomal peptides, *Nucleic Acids Res. 36*, D326-D331.

2. Bode, H. B., and Müller, R. (2005) The impact of bacterial genomics on natural product research, *Angew. Chem. Int. Ed. 44*, 6828-6846.

3. Challis, G. L. (2008) Genome mining for novel natural product discovery, *J. Med. Chem. 51*, 2618-2628.

4. Corre, C., and Challis, G. L. (2009) New natural product biosynthetic chemistry discovered by genome mining, *Nat. Prod. Rep. 26*, 977-986.

5. Gross, H., Stockwell, V. O., Henkels, M. D., Nowak-Thompson, B., Loper, J. E., and Gerwick, W. H. (2007) The genomisotopic approach: a systematic method to isolate products of orphan biosynthetic gene clusters, *Chem. Biol. 14*, 53-63.

6. Reverchon, S., Rouanet, C., Expert, D., and Nasser, W. (2002) Characterization of indigoidine biosynthetic genes in *Erwinia chrysanthemi* and role of this blue pigment in pathogenicity, *J. Bacteriol. 184*, 654-665.

7. Takahashi, H., Kumagai, T., Kitani, K., Mori, M., Matoba, Y., and Sugiyama, M. (2007) Cloning and characterization of a *Streptomyces* single module type non-ribosomal peptide synthetase catalyzing a blue pigment synthesis, *J. Biol. Chem. 282*, 9073-9081.

8. Yanagimoto, M., and Enatsu, T. (1983) Regulation of a blue pigment production by γ-nonalactone in *Streptomyces* sp., *J. Ferment. Technol. 61*, 545-550.

9. Spellberg, B., Powers, John H., Brass, Eric P., Miller, Loren G., and Edwards, J., John E. (2004) Trends in antimicrobial drug development: implications for the future, *Clin. Infect. Dis. 38*, 1279-1286.

10. Newman, D. J., Cragg, G. M., and Snader, K. M. (2000) The influence of natural products upon drug discovery, *Nat. Prod. Rep. 17*, 215-234.

11. Newman, D. J., Cragg, G. M., and Snader, K. M. (2003) Natural products as sources of new drugs over the period 1981-2002, *J. Nat. Prod. 66*, 1022-1037.

12. Butler, M. S. (2004) The role of natural product chemistry in drug discovery, *J. Nat. Prod. 67*, 2141-2153.

13. Butler, M. S. (2008) Natural products to drugs: natural product-derived compounds in clinical trials, *Nat. Prod. Rep. 25*, 475-516.

14. Dahn, U., Hagenmaier, H., Hohne, H., Konig, W. A., Wolf, G., and Zahner, H. (1976) Stoffwechselprodukte von mikroorganismen. 154. Mitteilung. Nikkomycin, ein neuer hemmstoff der chitinsynthese bei pilzen, *Arch. Microbiol. 107*, 143-160.

15. Fischbach, M. A., and Walsh, C. T. (2006) Assembly-line enzymology for polyketide and nonribosomal peptide antibiotics: logic, machinery, and mechanisms, *Chem. Rev. 106*, 3468-3496.

16. Johnson, B. A., Anker, H., and Meleney, F. L. (1945) Bacitracin: a new antibiotic produced by a member of the *B. subtilis* group, *Science 102*, 376-377.

17. Scherlach, K., and Hertweck, C. (2009) Triggering cryptic natural product biosynthesis in microorganisms, *Org. Biomol. Chem. 7*, 1753-1760.

18. Trejo, W. H., Dean, L. D., Pluscec, J., Meyers, E., and Brown, W. E. (1977) *Streptomyces laurentii*, a new species producing thiostrepton, *J. Antibiot. 30*, 639-643.

19. Li, C., Roege, K. E., and Kelly, W. L. (2009) Analysis of the indanomycin biosynthetic gene cluster from *Streptomyces antibioticus* NRRL 8167, *ChemBioChem 10*, 1064-1072.

20. Weinig, S., Hecht, H.-J., Mahmud, T., and Müller, R. (2003) Melithiazol biosynthesis: further insights into myxobacterial PKS/NRPS systems and evidence for a new subclass of methyl transferases, *Chem. Biol. 10*, 939-952.

21. Kelly, W. L., Pan, L. T., Li, C., Suidan, T. M., Roege, K. E., and Aragon, J. P. (2008) *Streptomyces laurentii* ATCC 31255 genomic DNA analysis, Georgia Institute of Technology, Atlanta, Georgia.

22. Gaitatzis, N., Kunze, B., and Müller, R. (2001) *In vitro* reconstitution of the myxochelin biosynthetic machinery of *Stigmatella aurantiaca* Sg a15: Biochemical characterization of a reductive release mechanism from nonribosomal peptide synthetases, *Proc. Natl. Acad. Sci. U.S.A. 98*, 11136-11141.

23. Velasco, A., Acebo, P., Gomez, A., Schleissner, C., Rodriguez, P., Aparicio, T., Conde, S., Munoz, R., de la Calle, F., Garcia, J. L., and Sanchez-Puelles, J. M. (2005) Molecular characterization of the safracin biosynthetic pathway from *Pseudomonas fluorescens* A2-2: designing new cytotoxic compounds, *Mol. Microbiol. 56*, 144-154.

24. Aron, Z. D., Dorrestein, P. C., Blackhall, J. R., Kelleher, N. L., and Walsh, C. T. (2005) Characterization of a new tailoring domain in polyketide biogenesis: the amine transferase domain of MycA in the mycosubtilin gene cluster, *J. Am. Chem. Soc. 127*, 14986-14987.

25. Balibar, C. J., Vaillancourt, F. H., and Walsh, C. T. (2005) Generation of D amino acid residues in assembly of arthrofactin by dual condensation/epimerization domains, *Chem. Biol. 12*, 1189-1200.

26. Gehring, A. M., Mori, I., Perry, R. D., and Walsh, C. T. (1998) The nonribosomal peptide synthetase HMWP2 forms a thiazoline ring during biogenesis of yersiniabactin, an iron-chelating virulence factor of *Yersinia pestis*, *Biochemistry 37*, 11637-11650.

27. Hacker, C., Glinski, M., Hornbogen, T., Doller, A., and Zocher, R. (2000) Mutational analysis of the *N*-methyltransferase domain of the multifunctional enzyme enniatin synthetase, *J. Biol. Chem. 275*, 30826-30832.

28. Miller, D. A., Walsh, C. T., and Luo, L. (2001) *C*-methyltransferase and cyclization domain activity at the intraprotein PK/NRP switch point of yersiniabactin synthetase, *J. Am. Chem. Soc. 123*, 8434-8435.

29. Sakai, T. T., Riordan, J. M., and Glickson, J. D. (1982) Models of bleomycin interactions with poly(deoxyadenylylthymidylic acid). Fluorescence and proton nuclear magnetic resonance studies of cationic thiazole amides related to bleomycin A2, *Biochemistry 21*, 805-816.

30. Schneider, T. L., Shen, B., and Walsh, C. T. (2003) Oxidase domains in epothilone and bleomycin biosynthesis: thiazoline to thiazole oxidation during chain elongation, *Biochemistry 42*, 9722-9730.

31. Silakowski, B., Schairer, H. U., Ehret, H., Kunze, B., Weinig, S., Nordsiek, G., Brandt, P., Blocker, H., Hofle, G., Beyer, S., and Muller, R. (1999) New lessons for combinatorial biosynthesis from myxobacteria: the myxothiazol biosynthetic gene cluster of *Stigmatella aurantiaca* DW4/3-1, *J. Biol. Chem. 274*, 37391-37399.

32. Stachelhaus, T., and Walsh, C. T. (2000) Mutational analysis of the epimerization domain in the initiation module PheATE of gramicidin S synthetase, *Biochemistry 39*, 5775-5787.

33. Kleinkauf, H., and von Dohren, H. (1990) Nonribosomal biosynthesis of peptide antibiotics, *Eur. J. Biochem. 192*, 1-15.

34. Challis, G. L., Ravel, J., and Townsend, C. A. (2000) Predictive, structure-based model of amino acid recognition by nonribosomal peptide synthetase adenylation domains, *Chem. Biol. 7*, 211-224.

35. 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*, 4174-4183.

36. Marahiel, M. A., Stachelhaus, T., and Mootz, H. D. (1997) Modular peptide synthetases involved in nonribosomal peptide synthesis, *Chem. Rev. 97*, 2651-2674.

37. Rausch, C., Weber, T., Kohlbacher, O., Wohlleben, W., and Huson, D. H. (2005) Specificity prediction of adenylation domains in nonribosomal peptide synthetases (NRPS) using transductive support vector machines (TSVMs), *Nucleic Acids Res. 33*, 5799-5808.

38. Stachelhaus, T., Mootz, H. D., and Marahiel, M. A. (1999) The specificity-conferring code of adenylation domains in nonribosomal peptide synthetases, *Chem. Biol. 6*, 493-505.

39. Gocht, M., and Marahiel, M. A. (1994) Analysis of core sequences in the D-Phe activating domain of the multifunctional peptide synthetase TycA by site-directed mutagenesis, *J. Bacteriol. 176*, 2654-2662.

40. Hamoen, L. W., Eshuis, H., Jongbloed, J., Venema, G., and van Sinderen, D. (1995) A small gene, designated *comS*, located within the coding region of the fourth amino acid-activation domain of *srfA*, is required for competence development in *Bacillus subtilis*, *Mol. Microbiol. 15*, 55-63.

41. Saito, M., Hori, K., Kurotsu, T., Kanda, M., and Saito, Y. (1995) Three conserved glycine residues in valine activation of gramicidin S synthetase 2 from *Bacillus brevis*, *J. Biochem. 117*, 276-282.

42. Kleinkauf, H., and Von Dohren, H. (1996) A nonribosomal system of peptide biosynthesis, *Eur. J. Biochem. 236*, 335-351.

43. Marahiel, M. A. (1992) Multidomain enzymes involved in peptide synthesis, *FEBS Lett. 307*, 40-43.

44. Turgay, K., Krause, M., and Marahiel, M. A. (1992) Four homologous domains in the primary structure of GrsB are related to domains in a superfamily of adenylate-forming enzymes, *Mol. Microbiol. 6*, 529-546.

45. Pavela-Vrancic, M., Pfeifer, E., Schroder, W., von Dohren, H., and Kleinkauf, H. (1994) Identification of the ATP binding site in tyrocidine synthetase 1 by selective modification with fluorescein 5'-isothiocyanate, *J. Biol. Chem. 269*, 14962-14966.

46. Pavela-Vrancic, M., Pfeifer, E., van Liempt, H., Schafer, H. J., von Dohren, H., and Kleinkauf, H. (1994) ATP binding in peptide synthetases: determination of contact sites of the adenine moiety by photoaffinity labeling of tyrocidine synthetase 1 with 2-azidoadenosine triphosphate, *Biochemistry 33*, 6276-6283.

47. Pavela-Vrancic, M., Van Liempt, H., Pfeifer, E., Freist, W., and Von Dohren, H. (1994) Nucleotide binding by multienzyme peptide synthetases, *Eur. J. Biochem. 220*, 535-542.

48. Tokita, K., Hori, K., Kurotsu, T., Kanda, M., and Saito, Y. (1993) Effect of single base substitutions at glycine-870 codon of gramicidin S synthetase 2 gene on proline activation, *J. Biochem. 114*, 522-527.

49. Kelly, W. L., Pan, L., and Li, C. (2009) Thiostrepton biosynthesis: prototype for a new family of bacteriocins, *J. Am. Chem. Soc. 131*, 4327-4334.

50. Kisters-Woike, B., Vangierdegom, C., and Müller-Hill, B. (2000) On the conservation of protein sequences in evolution, *Trends Biochem. Sci. 25*, 419-421.

51. Margoliash, E. (1963) Primary structure and evolution of cytochrome c, *Proc. Natl. Acad. Sci. U.S.A. 50*, 672-679.

52. Choulet, F., Aigle, B., Gallois, A., Mangenot, S., Gerbaud, C., Truong, C., Francou, F.-X., Fourrier, C., Guerineau, M., Decaris, B., Barbe, V., Pernodet, J.-L., and Leblond, P. (2006) Evolution of the terminal regions of the *Streptomyces* linear chromosome, *Mol. Biol. Evol. 23*, 2361-2369.

53. Ishikawa, J., and Hotta, K. (1999) FramePlot: a new implementation of the Frame analysis for predicting protein-coding regions in bacterial DNA with a high G+C content, *FEMS Microbiol. Lett. 174*, 3.

54. Nakamura, Y., Gojobori, T., and Ikemura, T. (1997) Codon usage tabulated from the international DNA sequence databases, *Nucleic Acids Res. 25*, 244-245.

55. Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. (1990) Basic local alignment search tool, *J. Mol. Biol. 215*, 403-410.

56. Balibar, C. J., Howard-Jones, A. R., and Walsh, C. T. (2007) Terrequinone A biosynthesis through L-tryptophan oxidation, dimerization and bisprenylation, *Nat. Chem. Biol. 3*, 584-592.

57. Corre, C., Song, L., O'Rourke, S., Chater, K. F., and Challis, G. L. (2008) 2-Alkyl-4-hydroxymethylfuran-3-carboxylic acids, antibiotic production inducers discovered by *Streptomyces coelicolor* genome mining, *Proc. Natl. Acad. Sci. U.S.A. 105*, 17510-17515.

58. Dimise, E. J., Widboom, P. F., and Bruner, S. D. (2008) Structure elucidation and biosynthesis of fuscachelins, peptide siderophores from the moderate thermophile *Thermobifida fusca*, *Proc. Natl. Acad. Sci. U.S.A. 105*, 15311-15316.

59. Lautru, S., Oves-Costales, D., Pernodet, J. L., and Challis, G. L. (2007) MbtH-like protein-mediated cross-talk between non-ribosomal peptide antibiotic and siderophore biosynthetic pathways in *Streptomyces coelicolor* M145, *Microbiology 153*, 1405-1412.

60. Robbel, L., Knappe, T. A., Linne, U., Xie, X., and Marahiel, M. A. (2010) Erythrochelin - a hydroxamate-type siderophore predicted from the genome of *Saccharopolyspora erythraea*, *FEBS J. 277*, 663-676.

61. Larkin, M. A., Blackshields, G., Brown, N. P., Chenna, R., McGettigan, P. A., McWilliam, H., Valentin, F., Wallace, I. M., Wilm, A., Lopez, R., Thompson, J. D., Gibson, T. J., and Higgins, D. G. (2007) Clustal W and Clustal X version 2.0, *Bioinformatics 23*, 2947-2948.

62. Pruitt, K. D., Tatusova, T., and Maglott, D. R. (2007) NCBI reference sequences (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins, *Nucleic Acids Res. 35*, D61-65.

63. Sambrook, J., and Russell, D. W. (2001) *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

64. Fildes, P. (1920) A new medium for the growth of *B. influenzae*, *Br. J. Exp. Pathol. 1*, 129-130.

65. Kuhn, R., Starr, M. P., Kuhn, D. A., Bauer, H., and Knackmuss, H.-J. (1965) Indigoidine and other bacterial pigments related to 3,3′-bipyridyl, *Arch. of Microbiol. 51*, 71-84.

66. Lehmann, P. (1999) P.R. Murray, E.J. Baron, M.A. Pfaller, F.C. Tenover and R.H. Yolken, eds. Manual of Clinical Microbiology, 7th ed, *Mycopathologia 146*, 107-108.

67. Shirling, E. B., and Gottlieb, D. (1966) Methods for characterization of *Streptomyces* species, *Int. J. Syst. Bacteriol. 16*, 313-340.

68. Sabouraud, R. (1892) Contribution a l'etude de la trichophytie humaine. Etude clinique, microscopique et bacterioloqique sur la pluralite des trichophytons de l'homme, *Ann. Dermatol. syphil. 3*, 1061-1087.

69. Abremski, K., Hoess, R., and Sternberg, N. (1983) Studies on the properties of P1 site-specific recombination: evidence for topologically unlinked products following recombination, *Cell 32*, 1301-1311.

70. Shuman, S. (1994) Novel approach to molecular cloning and polynucleotide synthesis using vaccinia DNA topoisomerase, *J. Biol. Chem. 269*, 32678-32684.

71. Rozen, S., and Skaletsky, H. J. (2000) Primer3 on the www for general users and for biologist programmers, in *Bioinformatics Methods and Protocols: Methods in Molecular Biology* (Krawetz, S., and Misener, S., Eds.), pp 365-386, Humana Press, Totowa, NJ.

72. Luckey, J. A., Drossman, H., Kostichka, A. J., Mead, D. A., D'Cunha, J., Norris, T. B., and Smith, L. M. (1990) High speed DNA sequencing by capillary electrophoresis, *Nucleic Acids Res. 18*, 4417-4421.

73. Bradford, M. M. (1976) A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding, *Anal. Biochem. 72*, 248-254.

74. Cerletti, P. (1959) Properties of riboflavin phosphates, *Anal. Chim. Acta 29*, 243-250.

75. Schneider, T. L., and Walsh, C. T. (2004) Portability of oxidase domains in nonribosomal peptide synthetase modules, *Biochemistry 43*, 15946-15955.

76. Stothard, P. (2000) The sequence manipulation suite: JavaScript programs for analyzing and formatting protein and DNA sequences, *BioTechniques 28*, 1102-1104.

77. Anand, S., Prasad, M. V., Yadav, G., Kumar, N., Shehara, J., Ansari, M. Z., and Mohanty, D. (2010) SBSPKS: structure based sequence analysis of polyketide synthases, *Nucleic Acids Res. 38 Suppl*, W487-496.

78. Ansari, M. Z., Yadav, G., Gokhale, R. S., and Mohanty, D. (2004) NRPS-PKS: a knowledge-based resource for analysis of NRPS/PKS megasynthases, *Nucleic Acids Res. 32*, W405-413.

79. Duitman, E. H., Hamoen, L. W., Rembold, M., Venema, G., Seitz, H., Saenger, W., Bernhard, F., Reinhardt, R., Schmidt, M., Ullrich, C., Stein, T., Leenders, F., and Vater, J. (1999) The mycosubtilin synthetase of *Bacillus subtilis* ATCC6633: A multifunctional hybrid between a peptide synthetase, an amino transferase, and a fatty acid synthase, *Proc. Natl. Acad. Sci. U.S.A. 96*, 13294-13299.

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

81. Konz, D., Doekel, S., and Marahiel, M. A. (1999) Molecular and biochemical characterization of the protein template controlling biosynthesis of the lipopeptide lichenysin, *J. Bacteriol. 181*, 133-140.

82. Konz, D., Klens, A., Schorgendorfer, K., and Marahiel, M. A. (1997) The bacitracin biosynthesis operon of *Bacillus licheniformis* ATCC 10716: molecular characterization of three multi-modular peptide synthetases, *Chem. Biol. 4*, 927-937.

83. Du, L., Chen, M., Sánchez, C., and Shen, B. (2000) An oxidation domain in the BlmIII non-ribosomal peptide synthetase probably catalyzing thiazole formation in the biosynthesis of the anti-tumor drug bleomycin in *Streptomyces verticillus* ATCC15003, *FEMS Microbiol. Lett. 189*, 171-175.

84. Du, L., Sanchez, C., Chen, M., Edwards, D. J., and Shen, B. (2000) The biosynthetic gene cluster for the antitumor drug bleomycin from *Streptomyces verticillus* ATCC15003 supporting functional interactions between nonribosomal peptide synthetases and a polyketide synthase, *Chem. Biol. 7*, 623-642.

85. Chen, H., O'Connor, S., Cane, D. E., and Walsh, C. T. (2001) Epothilone biosynthesis: assembly of the methylthiazolylcarboxy starter unit on the EpoB subunit, *Chem. Biol. 8*, 899-912.

86. Rusnak, F., Sakaitani, M., Drueckhammer, D., Reichert, J., and Walsh, C. T. (2002) Biosynthesis of the *Escherichia coli* siderophore enterobactin: sequence of the entF gene, expression and purification of EntF, and analysis of covalent phosphopantetheine, *Biochemistry 30*, 2916-2927.

87. Quadri, L. E., Weinreb, P. H., Lei, M., Nakano, M. M., Zuber, P., and Walsh, C. T. (1998) Characterization of Sfp, a *Bacillus subtilis* phosphopantetheinyl transferase for peptidyl carrier protein domains in peptide synthetases, *Biochemistry 37*, 1585-1595.

88. Ahmadian, A., Ehn, M., and Hober, S. (2006) Pyrosequencing: history, biochemistry and future, *Clin. Chim. Acta 363*, 83-94.

89. Redenbach, M., Kieser, H. M., Denapaite, D., Eichner, A., Cullum, J., Kinashi, H., and Hopwood, D. A. (1996) A set of ordered cosmids and a detailed genetic and physical map for the 8 Mb *Streptomyces coelicolor* A3(2) chromosome, *Mol. Microbiol. 21*, 77-96.

90. Seshime, Y., Juvvadi, P. R., Tokuoka, M., Koyama, Y., Kitamoto, K., Ebizuka, Y., and Fujii, I. (2009) Functional expression of the *Aspergillus flavus* PKS-NRPS hybrid CpaA involved in the biosynthesis of cyclopiazonic acid, Isao Fujii Iwate Medical University, School of Pharmacy, Iwate, Japan.

91. Maiya, S., Grundmann, A., Li, X., Li, S. M., and Turner, G. (2007) Identification of a hybrid PKS/NRPS required for pseurotin A biosynthesis in the human pathogen *Aspergillus fumigatus*, *ChemBioChem 8*, 1736-1743.

92. Hori, K., Yamamoto, Y., Minetoki, T., Kurotsu, T., Kanda, M., Miura, S., Okamura, K., Furuyama, J., and Saito, Y. (1989) Molecular cloning and nucleotide sequence of the gramicidin S synthetase 1 gene, *J. Biochem. 106*, 639-645.

93. Kratzschmar, J., Krause, M., and Marahiel, M. A. (1989) Gramicidin S biosynthesis operon containing the structural genes *grsA* and *grsB* has an open reading frame encoding a protein homologous to fatty acid thioesterases, *J. Bacteriol. 171*, 5422-5429.

94. Healy, F. G., Wach, M., Krasnoff, S. B., Gibson, D. M., and Loria, R. (2000) The *txtAB* genes of the plant pathogen *Streptomyces acidiscabies* encode a peptide synthetase required for phytotoxin thaxtomin A production and pathogenicity, *Mol. Microbiol. 38*, 794-804.

95. Edwards, D. J., and Gerwick, W. H. (2004) Lyngbyatoxin biosynthesis: sequence of biosynthetic gene cluster and identification of a novel aromatic prenyltransferase, *J. Am. Chem. Soc. 126*, 11432-11433.

96. de Crecy-Lagard, V. A., Saurin, W., Thibaut, D., Gil, P., Naudin, L., Crouzet, J., and Blanc, V. (2005) Streptogramin B biosynthesis in *Streptomyces pristinaespiralis* and *Streptomyces virginiae*: molecular characterization of the last structural peptide synthase gene, Institute Pasteur, Unit de Biochimie Microbienne, Paris, France.

97. de Crecy-Lagard, V., Blanc, V., Gil, P., Naudin, L., Lorenzon, S., Famechon, A., Bamas-Jacques, N., Crouzet, J., and Thibaut, D. (1997) Pristinamycin I biosynthesis in *Streptomyces pristinaespiralis:* molecular characterization of the first two structural peptide synthetase genes, *J. Bacteriol. 179*, 705-713.

98. Schauwecker, F., Pfennig, F., Grammel, N., and Keller, U. (2000) Construction and *in vitro* analysis of a new bi-modular polypeptide synthetase for synthesis of *N*-methylated acyl peptides, *Chem. Biol. 7*, 287-297.

99. Feng, Z., Qi, J., Tsuge, T., Oba, Y., Kobayashi, T., Suzuki, Y., Sakagami, Y., and Ojika, M. (2005) Construction of a bacterial artificial chromosome library for a myxobacterium of the genus *Cystobacter* and characterization of an antibiotic biosynthetic gene cluster, *Biosci. Biotechnol. Biochem. 69*, 1372-1380.

100. Muller, C., Nolden, S., Gebhardt, P., Heinzelmann, E., Lange, C., Puk, O., Welzel, K., Wohlleben, W., and Schwartz, D. (2007) Sequencing and analysis of the biosynthetic gene cluster of the lipopeptide antibiotic friulimicin in *Actinoplanes friuliensis*, *Antimicrob. Agents Chemother. 51*, 1028-1037.

101. May, J. J., Wendrich, T. M., and Marahiel, M. A. (2001) The *dhb* operon of *Bacillus subtilis* encodes the biosynthetic template for the catecholic siderophore 2,3-dihydroxybenzoate-glycine-threonine trimeric ester bacillibactin, *J. Biol. Chem. 276*, 7209-7217.

102. Hoffmann, D., Hevel, J. M., Moore, R. E., and Moore, B. S. (2003) Sequence analysis and biochemical characterization of the nostopeptolide A biosynthetic gene cluster from *Nostoc* sp. GSV224, *Gene 311*, 171-180.

103. Correia, T., Grammel, N., Ortel, I., Keller, U., and Tudzynski, P. (2003) Molecular cloning and analysis of the ergopeptine assembly system in the ergot fungus *Claviceps purpurea*, *Chem. Biol. 10*, 1281-1292.

104. Felnagle, E. A., Barkei, J. J., Park, H., Podevels, A. M., McMahon, M. D., Drott, D. W., and Thomas, M. G. (2010) MbtH-like proteins as integral components of bacterial nonribosomal peptide synthetases, *Biochemistry 49*, 8815-8817.

105. Zhang, W., Heemstra, J. R., Jr., Walsh, C. T., and Imker, H. J. (2010) Activation of the pacidamycin PacL adenylation domain by MbtH-like proteins, *Biochemistry 49*, 9946-9947.

106. Gehring, A. M., Lambalot, R. H., Vogel, K. W., Drueckhammer, D. G., and Walsh, C. T. (1997) Ability of *Streptomyces* spp. acyl carrier proteins and coenzyme A analogs to serve as substrates in vitro for *E. coli* holo-ACP synthase, *Chem Biol 4*, 17-24.

107. Lambalot, R. H., Gehring, A. M., Flugel, R. S., Zuber, P., LaCelle, M., Marahiel, M. A., Reid, R., Khosla, C., and Walsh, C. T. (1996) A new enzyme superfamily - the phosphopantetheinyl transferases, *Chem Biol 3*, 923-936.

108. Allen, J. F. (1978) Induction of a Mehler reaction in chloroplast preparations by flavin mononucleotide: effects on photosynthesis by intact chloroplasts, *Plant Sci. Lett. 12*, 151-159.

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

110. Beutner, S., Bloedorn, B., Frixel, S., Blanco, I. H., Hoffmann, T., Martin, H.-D., Mayer, B., Noack, P., Ruck, C., Schmidt, M., Schülke, I., Sell, S., Ernst, H., Haremza, S., Seybold, G., Sies, H., Stahl, W., and Walsh, R. (2001) Quantitative assessment of antioxidant properties of natural colorants and phytochemicals: carotenoids, flavonoids, phenols and indigoids. The role of beta-carotene in antioxidant functions, *J. Sci. Food Agric. 81*, 559-568.

111. Taylor, E. H., and Ramstad, E. (1960) Biogenesis of lysergic acid in ergot, *Nature 188*, 494-495.