

**Investigations of Aromatic Metabolism in
Streptomyces Bacteria Using a Multidisciplinary
Approach**

By Rui Zhang

B.S., University of Science and Technology of China, 2010

A Dissertation Submitted in Partial Fulfillment of the Requirements for the Degree of
Doctor of Philosophy in the Department of Chemistry at Brown University

Providence, Rhode Island

May 2016

© Copyright 2016 by Rui Zhang

This dissertation by Rui Zhang is accepted in its present form
by the Department of Chemistry as satisfying the
dissertation requirement for the degree of Doctor of Philosophy.

Date _____

Jason K. Sello, Ph.D., Advisor

Recommended to the Graduate Council

Date _____

Sarah Delaney, Ph.D., Reader

Date _____

Gerwald Jogl, Ph.D., Reader

Approved by the Graduate Council

Date _____

Peter M. Weber, Ph. D.
Dean of the Graduate School

Curriculum Vitae

Rui Zhang

EDUCATION

Brown University, Providence, Rhode Island 2010-2016
Ph.D. in Chemistry Advisor: Dr. Jason K. Sello

University of Science and Technology of China, Hefei, Anhui 2006-2010
Bachelor of Science, Chemistry

RESEARCH EXPERIENCE

Department of Chemistry, Brown University 2011-2016

Graduate Student Researcher

- Studies of *para*-hydroxybenzoic acid catabolism and its regulation in *Streptomyces coelicolor*.
 - Investigated *para*-hydroxybenzoic acid catabolism in *Streptomyces coelicolor* via genetic manipulation
 - Characterized peculiar transcriptional regulator of aromatic catabolism and elucidated molecular mechanism using methods from biochemistry and biophysics
- Resistance and biosynthesis studies of the antibiotic indolmycin in *Streptomyces griseus* ATCC 12648.
 - Conducted research on characterization of indolmycin resistance genes via genetic and biochemical method.
 - Succeeded in the heterologous expression of indolmycin biosynthetic gene cluster.
- Applied new chemical methods (using microwave reactor) to convert plant oil into biodiesel and biolubricants.
- Performed genetic analyses of metabolic flux in the biosynthesis of calcium dependent antibiotic in *Streptomyces coelicolor*.

Department of Chemistry, University of Science and Technology of China 2008-2010

Undergraduate Research Assistant

- Developed research project on theoretical studies of the intramolecular reaction of an aryl radical with the heterocycle. Conducted computational calculation to optimize reaction transition state structures and obtain activation energies.

TEACHING EXPERIENCE

Teaching Assistant

Organic Chemistry Lab in Brown University (CHEM0350) 2011-2013

- Supervised and assisted undergraduate students in organic chemistry techniques during laboratory activities.
- Assessed and graded their quiz and lab reports and provided help with better understanding organic chemistry concepts.

Teaching Assistant

Organic Chemistry Lab in Brown University (CHEM0360) 2010

- Performed GC-MS and NMR analysis on samples submitted in laboratory activities.

AWARDS

Best Poster Award (Student Choice)

- Department of Chemistry, Brown University 2014

Cane-Nambiar Graduate Student Travel Award

- Department of Chemistry, Brown University 2013

Graduate Fellowship

- Department of Chemistry, Brown University 2010-2011

Outstanding Student Scholarship

- University of Science and Technology of China 2006-2008

Outstanding Student Leader in Student Union

- University of Science and Technology of China 2007

ABSTRACTS AND PRESENTATIONS

- Rui Zhang, Dana M. Lord, Rebecca Page, and Jason K. Sello. “Genetic, Biochemical and Structural Studies of a Regulator of Aromatic Catabolism in *Streptomyces* Bacteria”, *XVII. International Symposium on the Biology of Actinomycetes*, Kusadasi, Aydin, Turkey (Oct. 2014)
- Rui Zhang, Dana M. Lord, Rebecca Page, and Jason K. Sello. “Genetic, Biochemical and Structural Studies of a Regulator of Aromatic Catabolism in *Streptomyces* Bacteria”, *Boston Bacterial Meeting*, Boston, MA (June 2014)

PUBLICATIONS

- Rui Zhang, Dana M. Lord, Rebecca Page, and Jason K. Sello, “Catabolism of *para*-hydroxybenzoic acid and its regulation in *Streptomyces coelicolor*” in preparation.
- Rui Zhang, James R. Blum, Tishan Williams, Charles W. Carter Jr., and Jason K. Sello, “Genetic and biochemical analyses of indolymycin resistance and biosynthesis in *Streptomyces gresieus* ATCC 12648” in preparation.
- Rui Zhang, James J. Vacchione, Gabrielle Grandchamp, Ethan Hickey, Jenica Shipley, and Jason K. Sello. “Genetic investigation of three paralogous *trpE* genes in *S. coelicolor* and their contributions to the production of calcium-dependent antibiotic (CDA)” in preparation.

Acknowledgements

To my dear family and friends without whom I would not be able to achieve my goals.

Special thanks to my advisor, Prof. Jason K. Sello. I owe him a debt of gratitude for his guidance, patience and encouragement throughout my entire graduate school study.

Table of Contents

Signature Page.....	iii
Curriculum Vitae.....	iv
Acknowledgements	vi
Table of Contents.....	vii
List of Figures.....	xii
List of Tables	xv
Abstract.....	xvi
Chapter 1. Introduction.....	1
1.1 Introduction to applied microbiology	2
1.2 Research on <i>Streptomyces</i> bacteria in the post-genomic era	5
1.2.1 Bioinformatics centers	6
1.2.2 Bioinformatics tools/databases	7
1.3 Exploiting aromatic metabolism in <i>Streptomyces</i> bacteria.....	13
1.3.1 Catabolism of aromatic compounds in <i>Streptomyces</i> bacteria.....	15
1.3.2 Anabolism of aromatic compounds in <i>Streptomyces</i> bacteria	20
1.3.2.1 Biosynthesis of amino acid tryptophan and its regulation	20
1.3.2.2 Indolmycin	24
1.3.2.3 Calcium Dependent Antibiotic (CDA)	26
1.4 References.....	28
Chapter 2. Catabolism of <i>para</i>-hydroxybenzoic acid and its regulation in <i>Streptomyces coelicolor</i>	37

2.1 Introduction.....	38
2.2 Materials and Methods.....	42
2.2.1 Bacterial strains and culture conditions	42
2.2.2 Plasmids and primers	43
2.2.3 Construction of SCO3084, SCO1308, SCO0266 and SCO3209 null mutants	43
2.2.4 Complementation of <i>pobA</i> (SCO3084) and <i>pobR</i> (SCO3209) null mutants.....	44
2.2.5 Assessment of nutritional capacity.....	45
2.2.6 Transcriptional analyses, RNA isolation and RT-PCR.....	45
2.2.7 Mapping of the <i>pobA</i> (SCO3084) transcription start site.....	46
2.2.8 Molecular cloning and protein purification.....	46
2.2.9 <i>In vitro</i> assay of a <i>para</i> -hydroxybenzoate hydroxylase activity	49
2.2.10 Agarose electrophoretic mobility shift assays.....	49
2.2.11 DNase I footprinting	50
2.2.12 Cloning of full length <i>pobR</i> and truncated mutants for over-expression in <i>S. coelicolor</i>	51
2.2.13 Isothermal Titration Calorimetry	52
2.2.14 Thermofluor Assay	52
2.2.15 Crystallization, data collection, and processing.....	52
2.3 Results and Discussion	56
2.3.1 <i>S. coelicolor</i> has a <i>pobA</i> gene (SCO3084) encoding <i>p</i> -hydroxybenzoate hydroxylase	56
2.3.2 Transcription of the gene <i>pobA</i> is induced by <i>para</i> -hydroxybenzoate (substrate of PobA) and its analogs.....	59
2.3.3 Transcription of the gene <i>pobA</i> was negatively regulated by SCO3209 (<i>pobR</i>).....	60
2.3.4 Bioinformatic analyses revealed that PobR is a peculiar IclR family protein.....	61
2.3.5 PobR binds to a palindromic sequence in the <i>pobA</i> promoter region	63
2.3.6 PHB and its analogs attenuate the binding of PobR to DNA and they bind differently to PobR.....	66
2.3.7 PHB, the physiological relevant ligand of PobR, influenced PobR oligomerization.....	69
2.3.8 Functional analyses of PobR N-terminal domain and C-terminal domains.....	71
2.3.9 Preliminary analyses of the crystal structure of the PobR-PHB complex.....	72

2.4 Conclusions.....	74
2.5 Scientific Contributions	75
2.6 References.....	76
Chapter 3. Genetic and biochemical analyses of indolmycin biosynthesis and resistance in <i>Streptomyces griseus</i> ATCC 12648.....	82
3.1 Introduction.....	83
3.2 Materials and Methods.....	86
3.2.1 Bacterial strains and culture conditions	86
3.2.2 Plasmids and primers	86
3.2.3 Genome sequencing and assembly	87
3.2.4 Growth curve	87
3.2.5 Heterologous overexpression of SGI0214, SGI1617 and SGI6453 in <i>S. coelicolor</i>	87
3.2.6 Indolmycin MIC determination	88
3.2.7 Construction of pETRP1B-His-TrpRS vectors.....	89
3.2.8 Expression of <i>trpRS</i> genes and protein purification.....	89
3.2.9 Assays of tryptophanyl-tRNA synthetase activity	90
3.2.10 Michaelis-Menten kinetics.....	90
3.2.11 Tryptophanyl-tRNA synthetase inhibition assays.....	90
3.2.12 Transcriptional analyses.....	91
3.2.13 Mapping of SGI0214, SGI1617, SGI6453 and SGI6454 transcription start sites	91
3.2.14 Heterologous expression of indolmycin biosynthetic gene cluster in <i>S. coelicolor</i>	93
3.2.15 Detection of indolmycin in liquid culture by LC-MS	94
3.2.16 Bioinformatic Analyses of TrpRSs Sequence.....	95
3.3 Results.....	98
3.3.1 The genome of <i>S. griseus</i> ATCC 12648 and secondary metabolites in <i>S. griseus</i> ATCC 12648.....	98

3.3.2 Genetic analyses of the three tryptophanyl-tRNA synthetases (TrpRSs) from <i>S. griseus</i> ATCC 12648.....	100
3.3.3 Biochemical characterization of the three TrpRSs from <i>S. griseus</i> ATCC 12648.....	103
3.3.4 Identification of the putative indolmycin biosynthetic gene cluster	104
3.3.5 Transcription profiles of the indolmycin biosynthetic gene cluster and the three trpRS genes during the growth of <i>S. griseus</i> ATCC 12648	105
3.3.6 Transcript mapping of the indolmycin biosynthetic gene cluster and tryptophanyl-tRNA synthetase genes.....	107
3.3.7 Heterologous production of indolmycin in <i>S. coelicolor</i>	110
3.4 Discussion.....	111
3.5 Scientific Contributions	116
3.6 References.....	116

Chapter 4. Genetic investigation of three paralogous <i>trpE</i> genes in <i>S. coelicolor</i> and their contributions to the production of calcium-dependent antibiotic (CDA)	123
4.1 Introduction.....	124
4.2 Materials and Methods.....	126
4.2.1 Bacterial strains and culture conditions	126
4.2.2 Plasmids and primers	127
4.2.3 Chemicals.....	127
4.2.4 Construction of <i>trpE1</i> , <i>trpE2</i> and <i>trpE3</i> null mutants.....	127
4.2.5 Genetic complementation of <i>trpE1</i> , <i>trpE2</i> and <i>trpE3</i> null mutants	128
4.2.6 Construction of <i>trpE1</i> , <i>trpE2</i> and <i>trpE3</i> overexpression strains	130
4.2.7 Growth curve	131
4.2.8 CDA bioassay	131
4.2.9 Transcriptional analyses.....	132
4.2.10 Mapping of the <i>trpE1</i> transcription start site	133
4.2.11 Creation of Reporter Constructs	133

4.2.12 Site-Directed Mutagenesis in <i>trpE1</i> leader	134
4.2.13 Assaying Bioluminescence of the Reporter Constructs	134
4.3 Results.....	138
4.3.1 Deletion of <i>trpE</i> genes individually has no effect on the growth of <i>S. coelicolor</i>	138
4.3.2 Deletion of <i>trpE</i> genes differently decreased the CDA production in <i>S. coelicolor</i>	140
4.3.3 The three <i>trpE</i> genes are transcribed differently	143
4.3.4 Identification of the <i>trpE1</i> transcription start site revealed the presence of a putative ribosome-mediated transcriptional attenuator	145
4.3.5 Missense point mutations within the <i>S. coelicolor trpE1</i> leader deregulate transcription	148
4.4 Discussion	150
4.5 Scientific Contributions	154
4.6 References.....	154

List of Figures

Figure 1.1 Life cycle of <i>Streptomyces</i> bacteria, reproduced from reference (14).....	4
Figure 1.2 Commodity chemicals via catabolism of aromatic compounds (48).....	13
Figure 1.3 Tryptophan and its derived aromatic compounds (antibiotic) that are biosynthesized in <i>Streptomyces</i> bacteria.....	14
Figure 1.4 (A) Schematic representation of lignin depolymerization to form dimeric aromatic fragments. (B) Catabolic pathways for the degradation of lignin-derived aromatic compounds by <i>S. paucimobillis</i> SYK-6.	16
Figure 1.5 Aromatic compounds funneled to protocatechuate (PCA), which is further metabolized through the three distinct ring-cleavage pathways.....	17
Figure 1.6 (A-C) Details of three PCA ring-cleavage pathways. (D) <i>pca</i> gene cluster identified in <i>Streptomyces coelicolor</i>	19
Figure 1.7 Tryptophan biosynthetic pathway with chorismate as the starting point.....	21
Figure 1.8 Organization of the <i>trp</i> operon of <i>E. coli</i>	22
Figure 1.9 RNA-based post-transcriptionally regulation utilized in regulating <i>trp</i> operon expression in different bacterial species.	23
Figure 1.10 Overview of CDA gene cluster.....	27
Figure 2.1 Phenotypes of (I) wild-type <i>S. coelicolor</i> M600, (II) <i>S. coelicolor</i> B1653 (Δ SCO3084:: <i>apr</i>), and (III) <i>S. coelicolor</i> B1654 (Δ SCO1308:: <i>apr</i>).	57
Figure 2.2 Phenotypes of (I) wild-type <i>S. coelicolor</i> M600, (II) <i>S. coelicolor</i> B1653 (Δ SCO3084:: <i>apr</i>), and (III) <i>S. coelicolor</i> B1657 (Δ SCO1308:: <i>apr</i> + pJS880).	57
Figure 2.3 Purification and biochemical characterization of PobA.	58
Figure 2.4 Changes in the transcription of <i>pobA</i> in media supplemented with PHB, PCA, and 2,4-DHB.	59
Figure 2.5 Transcription profile of the <i>pobA</i> gene expression in wild-type <i>S. coelicolor</i> and mutant strains lacking either SCO0266 or SCO3209.	61
Figure 2.6 Predicted domains organization of PobR protein.	62

Figure 2.7 Full amino acid sequence alignment of PobR identified in the <i>Streptomyces</i> genus through the protein BLAST bioinformatics analysis.	62
Figure 2.8 PobR recognizes an operator site in the <i>pobA</i> upstream region.....	65
Figure 2.9 (A) Aromatic compounds attenuate the DNA binding of PobR. (B) Thermal denaturation curve of PobR in the absence and presence of ligands using the Thermofluor assay.	67
Figure 2.10 Representative binding isotherms of PobR with aromatic ligands (A) PHB, (B) 2,4-DHB and (C) PCA.	68
Figure 2.11 Size-exclusion chromatography elution of PobR protein in the absence and presence of PHB.	69
Figure 2.12 Transcription profile of <i>pobA</i> gene expression in wild-type <i>S. coelicolor</i> and various mutant strains.....	71
Figure 2.13 The crystal structure of the PobR-PHB complex.....	73
Figure 2.14 Validated model for transcription regulation by PobR.....	74
Figure 3.1 Structure of indolmycin.	84
Figure 3.2 (A) The biosynthetic origins of indolmycin. (B) Proposed biosynthesis of indolmycin.	85
Figure 3.3 Upon the expression of different <i>trpRS</i> genes (SGI0214, SGI1617 and SGI6453), the constructed <i>S. coelicolor</i> strains showed different levels of indolmycin resistance.	101
Figure 3.4 The minimum inhibitory concentration (MIC) assay of indolmycin for <i>S. coelicolor</i> B725 and constructed strains.	102
Figure 3.5 Gene organization of the putative indolmycin biosynthetic gene cluster and enzymes that catalyze the hypothetical reactions in the indolmycin biosynthesis.....	105
Figure 3.6 Expression of indolmycin biosynthetic genes and <i>trpRS</i> genes during the growth of <i>S. griseus</i> ATCC 12648.	106
Figure 3.7 (A) Mapping transcription start site of the SGI0214 locus from <i>S. griseus</i> ATCC 12648. (B) Sequence alignment of the regions upstream of SGI1617 and its orthologs in various <i>Streptomyces</i> species. (C) Ribosome-mediated attenuator identified in <i>S. coelicolor</i> . (D) RNA secondary structure formed by the 195-nucleotide leader of SGI1617 transcript.....	109

Figure 3.8 Heterologous production of indolmycin by <i>S. coelicolor</i> M1154.....	111
Figure 3.9 (A) Sequence alignments of relevant TrpRSs. (B) Phylogenetic tree of TrpRS sequences.	113
Figure 3.10 Reconfiguration of the <i>Bst</i> TrpRS D1 Switch by replacing the sequence FVEL with WARE as is observed in the SG6453 indolmycin-resistant TrpRS.....	115
Figure 4.1 (A) The structure of calcium-dependent antibiotic (CDA). (B) Conversion of chorismate (product of shikimate pathway) to anthranilate under catalysis of TrpE, an anthranilate synthase..	125
Figure 4.2 The growth curves of <i>S. coelicolor</i> wild-type M145 and the <i>trpE</i> null strains.....	139
Figure 4.3 CDA production of wild-type <i>S. coelicolor</i> M145 and <i>trpE</i> null strains.....	141
Figure 4.4 CDA production of <i>S. coelicolor trpE</i> null strains after the introduction of either the wild-type <i>trpE</i> locus or <i>trpE</i> under the control of the constitutive <i>ermE*</i> promoter.	141
Figure 4.5 CDA production of <i>S. coelicolor</i> strains when the growth medium were supplemented with various concentrations of anthranilate or tryptophan.....	142
Figure 4.6 CDA production by wild-type <i>S. coelicolor</i> M145 strain upon the over-expression of the <i>trpE</i> genes.	143
Figure 4.7 Expression of the <i>trpE</i> genes and CDA biosynthetic gene in wild-type <i>S. coelicolor</i> M145.....	144
Figure 4.8 Analyses the region upstream of <i>trpE1</i>	146
Figure 4.9 Schematic representation of the mutually exclusive secondary structures formed by the 163-nucleotide leader of the <i>trpE1</i> transcript.	147
Figure 4.10 Representative <i>trp</i> operon.....	151

List of Tables

Table 1.1 Basic BLAST programs and their functions	9
Table 2.1 Summary of regulatory proteins of aromatic catabolic pathways in bacteria	39
Table 2.2 Strains employed in this study	53
Table 2.3 Plasmids employed in this study	53
Table 2.4 Oligonucleotides employed in this study	55
Table 2.5 PobR homologs in the other <i>Streptomyces</i> genus.	63
Table 2.6 Co-factor effects on PobR thermal stability	66
Table 2.7 The molecular weight (MW) and elution volume of standard proteins	70
Table 3.1 Strains employed in this study	95
Table 3.2 Plasmids employed in this study	95
Table 3.3 Oligonucleotides employed in this study	96
Table 3.4 Synthesis of secondary metabolites in <i>Streptomyces griseus</i> ATCC 12648	99
Table 3.5 Comparison of indolmycin sensitivity of <i>S. coelicolor</i> B725 upon the expression of TrpRS genes from <i>S. griseus</i> ATCC 12648.....	101
Table 3.6 The kinetic parameters of the three TrpRSs from <i>S. griseus</i> ATCC 12648	104
Table 4.1 Strains employed in this study	134
Table 4.2 Plasmids employed in this study	135
Table 4.3 Oligonucleotides employed in this study	136
Table 4.4 Bioluminescence of <i>S. coelicolor</i> M600 WT with various reporters	149

Abstract

Streptomycetes are a large group of Gram-positive, soil-dwelling bacteria that have evolved remarkable metabolic capabilities for survival in a nutrient-poor and a biodiverse ecological niche. *Streptomyces* bacteria are best known for their capacity to produce structurally complex antibiotics for chemical defense from simple intermediates in central metabolism. Indeed, streptomycetes are the source of about half of the 23,000 known antibiotics and two-thirds of the antibiotics that are used in clinical and veterinary medicine. Beyond these anabolic capabilities, streptomycetes are also distinguished by their abilities to consume various forms of carbon in soil. Their ability to catabolize the abundant organic matter in the soil has been ascribed to numerous secreted and cytosolic enzymes. The ecological importance of carbon recycling and the commercial value of antibiotics have motivated much research into the metabolic capabilities of these bacteria.

My thesis work has focused on both how these bacteria consume the lignin component of plant biomass and how they produce antibiotics from the proteinogenic amino acid tryptophan. Aside from their common connection in aromatic metabolism, these projects are linked by origins in in-depth bioinformatic analyses of streptomycete genome sequences. The resulting hypotheses about the structure and regulation of these anabolic and catabolic pathways were tested systematically using a combination of genetic, biochemical and biophysical experiments. My thesis outlines our findings and points that towards the exploitation of the metabolic capabilities of *Streptomyces* bacteria in biotechnology.

Chapter 1. Introduction

1.1 Introduction to applied microbiology

Microorganisms play essential roles in biological technologies that improve and extend human lives (1). For example, the utilization of microorganisms in large scale production of food and industrial products has been done worldwide. In food microbiology, the bacteria are used to produce beverages, yogurts, milk by products such as cheese, sweet chocolates and silage. Actually, preservation of food including the use of fermentation of otherwise perishable raw materials has been used by man since the Neolithic period (around 100,000 years BC) (2). Fermentation of microorganism is also the means by which many medicinal antibiotics are produced. Currently, the broad spectrum antibiotics like penicillin and streptomycin are in use at large throughout the globe. Besides nutritive foods and antibiotics, microorganisms have also been applied in the production of high value chemicals like bioplastics, serum antibodies, and essential hormones (3-5). Apart from production of commodity chemicals, microorganisms play central role in bioremediation to remove the persistent or toxic compounds that are widespread in the environment. It has been reported that oil degrading indigenous microorganisms dramatically reduced the overall environmental impact of both the *Exxon Valdez* and *BP Deepwater Horizon* oil spills happened in 1989 and 2010, respectively (6). Likewise, the mercury-resistant bacteria have a unique way of resisting and subsequently transforming the toxic forms of mercury to non-toxic forms (7). In addition, biological treatment of waste water and industrial exhaust gases is excluded: there is, moreover, an increasing tendency for recovery rather than destruction of valuable compounds of these discharges (8).

Among various microorganisms used in biotechnology, actinobacteria are particularly impactful. In 1943, Waksman and Henrici proposed that the capacity to form branching cells was the hallmark of the actinomycetes and used the degree of branching to define three major groups, two of which they subdivided, giving five genera altogether. These five genera are *Mycobacterium*, *Actinomyces*, *Nocardia*, *Streptomyces* and *Micromonospora* respectively (9). With regard to morphology, many

actinomycetes are filamentous like fungi. These Gram-positive bacteria with high G+C genomes remained rather a Cinderella group until, with the discovery of streptomycin, the world suddenly had to take them seriously. Today, the actinomycetes are best known as producers of antibiotics. With the advent of molecular biology, they have become important to biotechnologists in the search of new antibiotics, vitamins, enzyme inhibitors, etc. They also play an important role in the biodegradation of wastes. Indeed, their wide (natural) distribution in soil, composts, water and elsewhere in the environment makes them important to the agricultural and waste industries (10).

Within the Actinobacteria phylum, members of the *Streptomyces* genus are particularly interesting. Streptomycetes represent the largest genus of Actinobacteria comprising over 900 species (11,12). Members of the genus *Streptomyces* are Gram-positive bacteria distinguished by high G+C genome and a filamentous morphology. The vegetative hyphae (0.5-2.0 μm in diameter) of *Streptomyces* bacteria produce an extensively branched mycelium that rarely fragments. The aerial mycelium at maturity forms chains of at least three non-motile spores. A few species bear short chains of spores on the substrate mycelium. Sclerotia-, pycnidial-, sporangia-, and synnemata-like structures may be formed by some species. The cell wall peptidoglycan contains major amounts of L-diaminopimelic acid (L-DAP). The bacteria are lack of mycolic acids but contain major amounts of saturated, iso-, and anteiso-, fatty acids; possess either hexa- or octahydrogenated menaquinones with nine isoprene units as the predominant isoprenolog; and have complex polar lipid patterns that typically contain diphosphatidylglycerol, phosphatidylethanolamine, phosphatidylinositol, and phosphatidylinositol mannosides. As aerobes, streptomycetes require oxygen during growth (13).

The life cycle of streptomycetes is complex (Fig. 1.1) (14). Under favorable conditions, the growth begins with the germination of a spore which leads to the formation of a germ tube. (The spores are semi-dormant stage in the life cycle which can survive in soil for long periods.) The germ tube elongates and branches to form hyphae that are collectively called a substrate mycelium. As they grow via apical extension, the hyphae secrete hydrolytic enzymes that breakdown protein,

carbohydrates and other macromolecules. The resulting hydrolysis products are absorbed by the colony and are used to support growth. Because streptomycetes are immobile, when essential nutrients are depleted locally, the bacteria become starved. The starvation triggers the initiation of a complex developmental program. Older parts of the mycelium undergo programmed cell death and release nutrients that are used to support the sporulation process and thus complete the life cycle. Prior to sporulation, a subset of the hyphae undergo morphological differentiation into a structure known as aerial hyphae or the aerial mycelium. During programmed cell death, the streptomycetes are under particularly high risk to be scavenged by other competing microorganisms, especially some of which are motile. Thus, the general fact that antibiotic production coincides with the onset of morphological differentiation is believed to be a defensive tactic utilized by streptomycetes to prevent other organisms from scavenging the metabolites and macromolecules released during the autolysis (programmed cell death) (9,15,16).

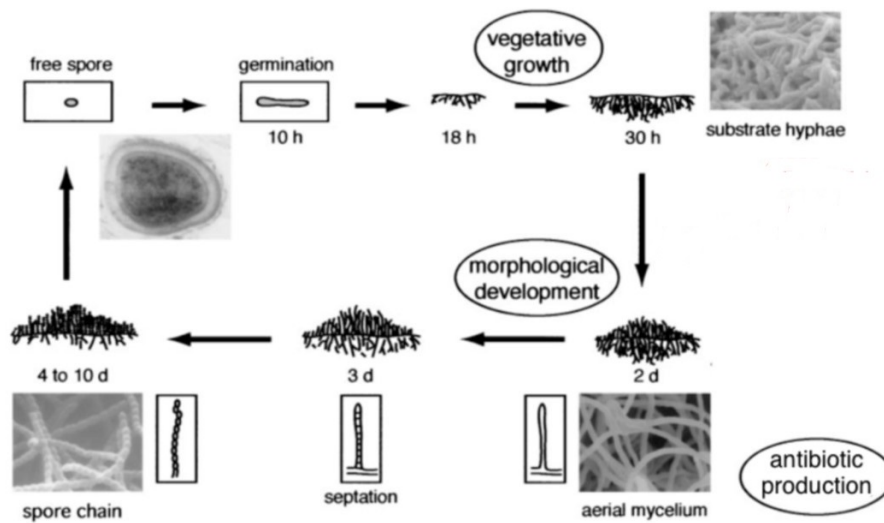


Figure 1.1 Life cycle of *Streptomyces* bacteria, reproduced from reference (14).

Streptomycetes are widely distributed and abundant in soil. In their terrestrial ecological niche, streptomycetes play important roles as general saprobes. Streptomycetes decompose mixtures of insoluble polymers and absorb the soluble breakdown products for their own growth and reproduce. In the context of terrestrial ecology, it is particularly notable that streptomycetes have a unique capacity for the production of a multitude of varied and complex secondary metabolites. These natural products have a wide range of practical bioactivities including: antibacterial, antifungal, antiparasitic antitumor and immunosuppressant (9,16). Interest in streptomycetes bacteria is based largely on the fact that they produce half of the 23,000 known antibiotics and two-thirds of the antibiotics used in clinical and veterinary medicine (17,18). The large scale fermentation of streptomycetes for the production of medicinal antibiotics has been an integral part of the pharmaceutical industry for more than half a century (19-22). The importance of streptomycetes in nature and medicine have made them popular subjects for research and the focus of biotechnology programs.

1.2 Research on *Streptomyces* bacteria in the post-genomic era

Research and exploitation of streptomycetes have been empowered by genomics and bioinformatics. Sequencing the complete genome of the model actinomycete *Streptomyces coelicolor* A3(2) is a milestone in *Streptomyces* genetic studies (9). The Sanger Institute carried out the sequencing project in collaboration led by Prof. David Hopwood of the John Innes Centre. The sequence was generated using a clone-by-clone approach, initially using cosmids generated and mapped by Matthias Redenbach in Germany. Subsequently, a BAC (bacterial artificial chromosome) library was used to fill gaps and confirm the map (17). The *S. coelicolor* sequencing project cost about 2 million dollars (funded by BBSRC) and took almost 4 years to complete the project. The last base pair was completed on July 21, 2001 (9). In the past 15 years, the genomes of more than 250 streptomycetes have been deposited in GenBank. The genome sequence has

increased our understanding of microbial life in the soil. It has also greatly facilitated discovery of new secondary metabolites and generation of new drug candidates by genetic engineering.

The original clone-based Sanger sequencing technique indeed afforded an accurate full genome, but its costs of both money and time were high (23). In the last decade, new sequencing technologies have been developed including next-generation sequencing (NGS) and single molecule real-time (SMRT) sequencing (16,23). The NGS is a high-throughput, imaging-based system that vastly increases the speed of sequencing, as well as the output of data (23). SMRT sequencing is a sequencing-by-synthesis technology based on real-time imaging of fluorescently tagged nucleotides as they are synthesized along each individual DNA template molecule (24). Overall, advances in sequencing technology have decreased the cost of a single genome sequence down from nearly \$1 million in 2007 to close to the \$1,000 mark today (23). In addition, the new technologies allow almost complete genome sequences to be obtained in a matter of weeks or even days (16). Therefore, the rate-limiting (and most costly) step has now moved on to the analysis of genomic data, which is a main emphasis of bioinformatics.

1.2.1 Bioinformatics centers

Bioinformatics is an interdisciplinary field of science that combines computer science, statistics, mathematics, and engineering to analyze and interpret biological data. Currently, there are three major bioinformatics centers around the world: the U.S. National Center for Biotechnology Information (NCBI), the European Bioinformatics Institute (EMBL-EBI) and Swiss Institute for Bioinformatics (SIB) (25).

NCBI (<http://www.ncbi.nlm.nih.gov/>) is a U.S. government-funded national resource for molecular biology information. It is part of the United States National Library of Medicine (NLM), a branch of the National Institutes of Health (NIH). The NCBI collects a series of databases relevant to biotechnology and biomedicine. The major databases include GenBank for DNA sequences and

PubMed, a bibliographic database for the biomedical literature. It also supplies bioinformatics tools and services, among which the most widely-used is BLAST (26). We will talk about this bioinformatics tool in detail in the following part.

EMBL is an intergovernmental organization, consisting of more than 20 members states, associate and prospect members (27). This publicly-funded non-profit institute is housed at five sites in Europe whose expertise covers the whole spectrum of molecular biology. The EMBL-EBI (<http://www.ebi.ac.uk/>) is a hub for bioinformatics research and services. It maintains the world's most comprehensive range of freely available and up-to-date molecular databases and offer a variety of local and online bioinformatics tools including the Clustal series of programs, which will be discussed later.

SIB is an academic non-profit foundation which federates bioinformatics activities throughout Switzerland. ExPASy (<http://www.expasy.org/>) is the SIB Bioinformatics Resource Portal which provides access to scientific databases and software tools (*i.e.*, *resources*) in different areas of life sciences. Historically, ExPASy, known as “Expert Protein Analysis System”, was developed for proteomics analyses (28). Indeed, it has worldwide reputation as one of the main bioinformatics resources for proteomics. But now, ExPASy has evolved and the categories include proteomics, genomics, phylogeny, systems biology, population genetics, transcriptomics etc.

1.2.2 Bioinformatics tools/databases

There are substantial bioinformatics tools/programs available for data analyses. With regard to the functional assessment of a single gene, the three tools (BLAST, the Clustal series of programs, and Gibbs motif sampler) have been widely used. Next, to study a metabolic pathway and related gene cluster in an unknown organism, we can map the genes on to a specific pathway based on gene-by-gene sequence similarities against homologues of genes in KEGG pathways database. Additionally, “antiSMASH” has become a powerful tool for mining secondary metabolism gene clusters in

genome sequences. Finally, StrepDB and the IMG system contain genomes of streptomycetes. They also provide tools and viewers for analyzing and reviewing the annotations of genes and genomes in a comparative context.

All above bioinformatics tools/databases have been applied in my research projects later discussed in chapter 2 to 4. The following information are the detail descriptions of each mentioned bioinformatics tool/database.

a. Basic Local Alignment Search Tool (BLAST)

The BLAST algorithm is one of the most well-known and widely used bioinformatics tools available. It is probably used by more scientists than almost any other bioinformatics tool (29). BLAST is used to compare two sequences and find regions of local similarity between those sequences. Local similarity algorithms seek only relatively conserved subsequences, and a single comparison may yield several distinct subsequence alignment; unconserved regions do not contribute to the measure of similarity (30). Gene and protein sequences are fundamentally important in molecular biology because genes/proteins carrying homologous sequences generally perform similar biological functions (31). The putative function of a newly sequenced gene can be proposed according to the function of its homolog. Thus, the result of a BLAST analysis can be used to infer functional and evolutionary relationships between sequences.

The BLAST program is simple and robust; it can be implemented in a number of ways and it can be used to analyze gene sequences, amino acid sequences, immunoglobulin sequences and gene expression data (29). Table 1.1 listed the basic BLAST programs and their functions. There are also specialized BLAST for specific research purposes. All the programs including basic BLAST program and its many variants are available on NCBI web site (<http://blast.ncbi.nlm.nih.gov/Blast.cgi>).

Table 1.1 Basic BLAST programs and their functions

Program	Function
blastn	Search a nucleotide database using a nucleotide query
blastp	Search protein database using a protein query
blastx	Search protein database using a translated nucleotide query
tblastn	Search translated nucleotide database using a protein query
tblastx	Search translated nucleotide database using a translated nucleotide query

b. The Clustal series of programs

While the BLAST tool enables easy comparisons of two sequences, the Clustal series of programs are widely used in molecular biology for multiple alignments of both nucleic acid and protein sequences and for preparing phylogenetic trees. These programs are used for global multiple sequence alignment, whose algorithms optimize the overall alignment of sequences, which may include large stretches of low similarity (29,32,33).

The programs are seductively easy to use. ClustalW2 (34) is a web server and has been set up at the EBI (<http://www.ebi.ac.uk/Tools/msa/clustalw2/>). This service has been retired recently and substituted by Clustal Omega (35) (<http://www.ebi.ac.uk/Tools/msa/clustalo/>), which is the latest version of Cluster and is suitable for medium-large alignment. Users can enter sequence either by pasting them or by uploading file from local computer. The sequence should be in GCG, FASTA, EMBL, GenBank, PRI, NBRF, PHYLIP or UniProtKB/Swiss-Prot format. The resulting multiple alignments can be displayed as either black and white or color coded text. ClustalX is the local alignment version. It is easy to install, is user-friendly and maintains the portability of the previous generations through the NCBI Vibrant toolkit (<ftp://ncbi.nlm.nih.gov/toolbox/ncbitools/>).

c. Gibbs motif sampler

Gibbs motif sampler is another multiple sequence alignment program which can detect motifs and locally conserved regions in DNA or protein sequences. This new Gibbs strategy (compared to previous Gibbs site sampler) addressed the problem of detecting motifs when little information about the number of occurrences of each motif is available. The Gibbs motif sampler iteratively samples motif models (or possibly no model) for each site and therefore optimally partitions motif-encoding regions into different motifs (36,37). It is particularly useful in the analysis of gene promoter sequence.

This program also provides both a web server (<http://ccmbweb.ccv.brown.edu/gibbs/gibbs.html>) and a local version. The online user interface is more complicated than Clustal Omega. With this software, there are prokaryotic defaults option and eukaryotic defaults option. Or the user can define each parameters (*et. al.* number of different motifs, max sites per seq, motif width and estimate total sites for each motif type). The results are send to user via email in text form. In chapter 2, we performed Gibbs motif sampler to confirm the identification of a regulatory protein binding site in *S. coelicolor*.

d. Kyoto Encyclopedia of Genes and Genomes

KEGG (<http://www.genome.jp/kegg/>) is a knowledge base for systematic analysis of gene function, and for linking it with genomic information to yield higher order functional information. The KEGG project was initiated in May 1995 and KEGG is regarded as a “computer representation” of the biological system according to the developers (38,39). KEGG contains 17 databases that are broadly categorized into systems information, genomic information, chemical information and health information. These categories are distinguished by color coding of web links (40). KEGG database is used most extensively in our lab. It is used in the characterization of genes in genomes underlying metabolism, genetic information processing, environmental information processing,

cellular processes, organismal systems, human diseases and drug development. The best organized part is metabolism, which is represented by about 200 graphical diagrams for the reference metabolic pathway. Each reference pathway can be viewed as a network of enzymes or a network of EC numbers. Combining this database and bioinformatics tool BLAST, researchers can characterize similar metabolism pathway in a newly sequenced organism.

e. antibiotic & Secondary Metabolite Analysis Shell (antiSMASH)

With more and more genomes sequenced in cheaper and faster ways, the needle has moved on the analysis of genomic data (23). Mining genomes for secondary metabolism gene clusters has become a cottage industry. antiSMASH is the first comprehensive pipeline capable of identifying biosynthetic loci covering the whole range of known secondary metabolite classes (polyketide, non-ribosomal peptides, terpenes, aminoglycosides, aminocoumarins, indolocarbazoles, lantibiotics, bacteriocins, nucleosides, β -lactams, butyrolactones, siderophores, melanins and others) (41). antiSMASH has evolved to its version 3.0 (42), a comprehensive resource for the genome mining of biosynthetic gene clusters. Since version 2.0 (43), the program was completely revamped and can be used with highly fragmented draft-quality genome sequences while the previous version only worked with finished genomes, which is tremendously important to the discovery of novel metabolites in the ever-expanding database of draft-quality genome sequences of streptomycetes (44).

antiSMASH is available at <http://antismash.secondarymetabolites.org/>. For nucleotide input, the user can load file files from a local computer. The file should be in GenBank, EMBL format (recommended) or in FASTA format. Results can be retrieved via email or through a Job ID #. The interface displaying results is super user-friendly with all the identified gene clusters listed and color-coded according to the metabolite category of predicted products. After clicking each gene cluster, user will be directed to another page showing the organization of the gene cluster and sometimes the predicted compound structure.

f. StrepDB: the *Streptomyces* Annotation Server

StrepDB (<http://strepdb.streptomyces.org.uk/>) is an online *Streptomyces* annotation server. It holds the genome sequences of widely-used *Streptomyces* species, including *S. coelicolor*, *S. avermitilis* MA-4680, *S. griseus* (producer of streptomycin), *S. lividans*, *S. venezuelae* ATCC 10712 and others. It also provides cosmid library information for *S. coelicolor* and *S. venezuelae* ATCC 10712. Users can extract EMBL files for desired segments of any genome of interest. This web server supplies BLAST analyses within its own database. Each collected genome is also associated with KEGG, which facilitates gene-fishing by KEGG pathways.

g. Integrated Microbial Genomes (IMG)

The IMG database (<https://img.jgi.doe.gov/cgi-bin/m/main.cgi>) was first released in March 2005 and serves as a community resource for comparative analyses of publicly available genomes (45). This system integrates publicly available draft and complete microbial genomes from all three domains of life. It also includes sequence data from a large number of plasmids and viruses. IMG employs NCBI's RefSeq resource (46) as its main source of public genome sequence data and records its primary genome sequence information from RefSeq. It provides gene organization within finished genomes and sequences of scaffolds and/or contigs for draft genomes. In all cases, IMG provides predicted protein-coding sequences (CDSs), protein product names, putative protein functions for all open reading frames and some RNA-coding genes. In essence, IMG provides tools and viewers for analyzing and reviewing the annotation of genes and genomes in a comparative context.

All of the aforementioned bioinformatics tools were essential in the work described in this thesis.

1.3 Exploiting aromatic metabolism in *Streptomyces* bacteria

All of the work outlined in this thesis is focused on the capacity of streptomycetes to metabolize aromatic compounds. Aromatic compounds are cheap and high abundant, but under-utilized in biotechnology. *Streptomyces* and other microorganisms have the capacity to catabolize aromatic compounds into key intermediates in central metabolism. These molecules (central metabolites) can be further converted to commodity chemicals. As shown in Figure 1.2, the aromatic compounds are first degraded to acetyl-CoA, which is the precursor of many high value chemicals. Polyhydroxyalkanoates (PHAs) are the most versatile fully biodegradable polymers with properties similar to conventional plastics. The versatility of PHAs has made them good candidates for the study of their potential in a variety of areas from biomedical/medical fields to food, packaging, textile and household material (47). Triglycerides are the precursors for biodiesel production. Butanol and ethanol are used as clean biofuel for combustion engines.

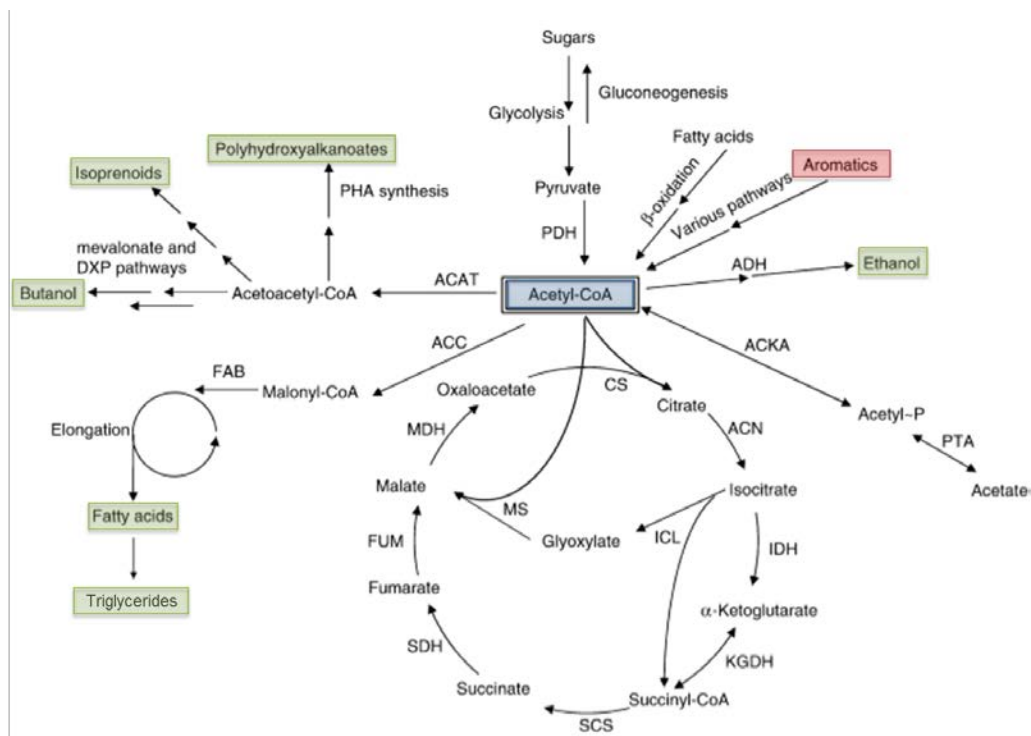


Figure 1.2 Commodity chemicals via catabolism of aromatic compounds (48).

Alternatively, aromatic compounds (especially the proteinogenic aromatic amino acids) can be converted directly to commodity chemicals like antibiotics (Fig 1.3). For example, it has been reported that tryptophan is incorporated into pyrrolnitrin in *Pseudomonas* species (49). Streptomycetes produce many antibiotics from tryptophan. For example, the calcium dependent antibiotic (CDA) produced by *S. coelicolor* contains a D- and L-tryptophan residues (17,50). *S. griseus* ATCC 12648 utilizes tryptophan as direct precursor to synthesize antibiotic indolmycin (51,52). In addition, tryptophan may be a direct or indirect precursor of other streptomycete secondary metabolites such as actinomycin D, produced by *S. parvulus* (53); streptonigrin, produced by *S. flocculus* (54); and candidicin, produced by *S. griseus* (55).

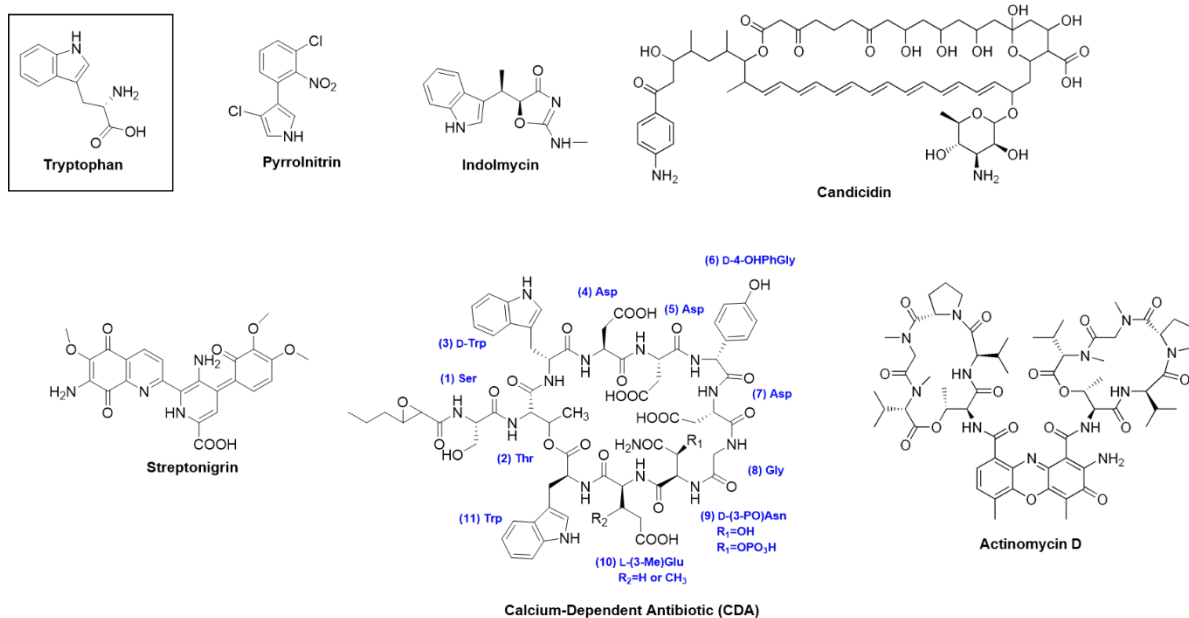


Figure 1.3 Tryptophan and its derived aromatic compounds (antibiotic) that are biosynthesized in *Streptomyces* bacteria.

This thesis describes the use of bioinformatics to study how streptomycetes make tryptophan and convert it to antibiotics.

1.3.1 Catabolism of aromatic compounds in *Streptomyces* bacteria

Chapter 2 concerns pathways for the catabolism of lignin-derived aromatic compounds. Lignin is a heterogeneous polymer of phenylpropanoid units derived from three major phenolic components, *p*-coumaryl alcohol (H), coniferyl alcohol (G) and sinapyl alcohol (S). In nature, lignin makes up about 25-30% of the mass of dry wood, which makes it the most abundant aromatic material on the Earth. The phenylpropane units are linked through C-C and C-O-C bonds and are recalcitrant to degradation by microorganisms (56).

Up to present, white-rot fungi are the best known organisms that can completely break down the lignin to carbon dioxide and water (57,58). Heme-dependent lignin peroxidases (Lip), manganese peroxidases (MnP), and versatile peroxidases and copper-dependent laccases are all extracellular lignolytic enzymes produced by white-rot fungi used in oxidative depolymerisation of lignin (Fig. 1.4A) (59). These enzymes catalyze depolymerization via radical reactions (60). Besides fungi, there are also several bacteria that are able to break down lignin as well as lignin-derived low-molecular-weight compounds. There are reports of lignin-related biphenyl degraded by *Pseudomonas putida* and *Pseudomonas fluorescens*. *Sphingomonas paucimobilis* SYK-6 is capable of breaking down dimeric lignin fragments, including biaryls, β -aryl ethers, biphenyl, phenylcoumarane and diarylpropane to produce monomeric aromatic compounds (Fig. 1.4B) (60). Interestingly, the first bacterial enzyme capable of depolymerizing lignin was isolated from *Streptomyces viridosporus* (61). This lignin-consuming strain was one of many isolated by D. L. Crawford and his group at University of Idaho in the 1970s (62). His team published the first conclusive evidence showing that *Streptomyces* can decompose lignin in 1978 (62). The ensuing studies demonstrated that both *S. viridosporus* and *Amcolatopsis setonii* (formerly *Streptomyces*

setonii) converted the lignin in corn stover into *p*-coumarate, ferulate, vanillate, *p*-hydroxybenzoate, and protocatechuic acid via formation of “acid-precipitable, polymeric lignin” intermediates (63-65).

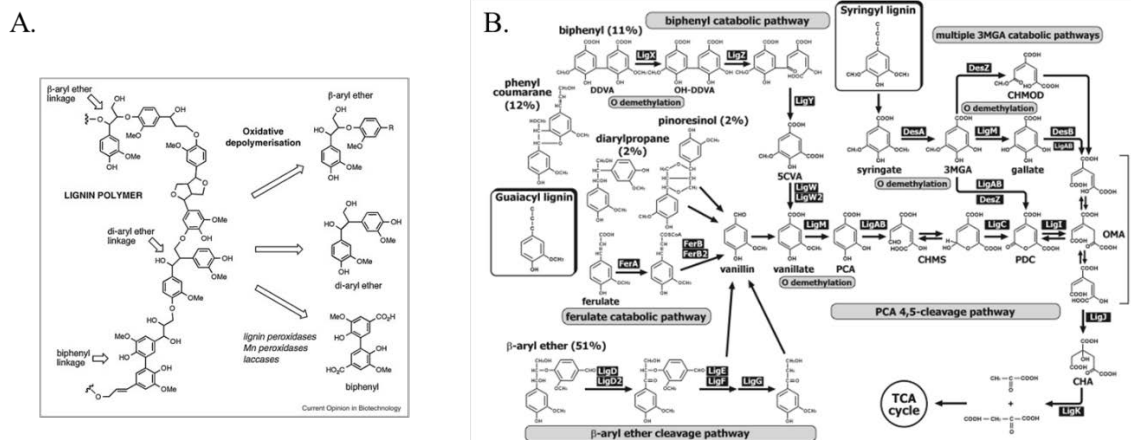


Figure 1.4 (A) Schematic representation of lignin depolymerization to form dimeric lignin aromatic fragments. Reproduced from reference (59) (B) Catabolic pathways for the degradation of lignin-derived aromatic compounds by *S. paucimobillis* SYK-6. Reproduced from reference (60).

As shown in Figure 1.4B, protocatechuic acid (PCA) is one of the key intermediate metabolites in the bacterial degradation of lignin and its derived aromatic compounds. Three distinct catabolic pathways of PCA degradation have been identified in microorganisms: *meta*-fission between C2 and C3; *ortho*-fission between C3 and C4; and *meta*-fission between C4 and C5 (Fig. 1.5). All the three pathways funnel PCA to the tricarboxylic acid (TCA) cycle, while the final products of each pathway are different.

The cleavage of PCA between C2 and C3 was first reported by R. L. Crawford in 1975 when they investigated the ability of members of the genus *Bacillus* to utilize various aromatic compounds (66). In *Bacillus macerans* (later reclassified as *Paenibacillus*) strain JJ-1B, the aromatic ring of PCA is cleaved by PCA 2,3-dioxygenase to yield 5-carboxy-2-hydroxymuconic semialdehyde, which is further degraded into pyruvate and acetaldehyde (67). The genes responsible for this

pathway were identified and characterized later in 2009 (68). Up to now, this is still the only example of PCA 2,3-cleavage pathway.

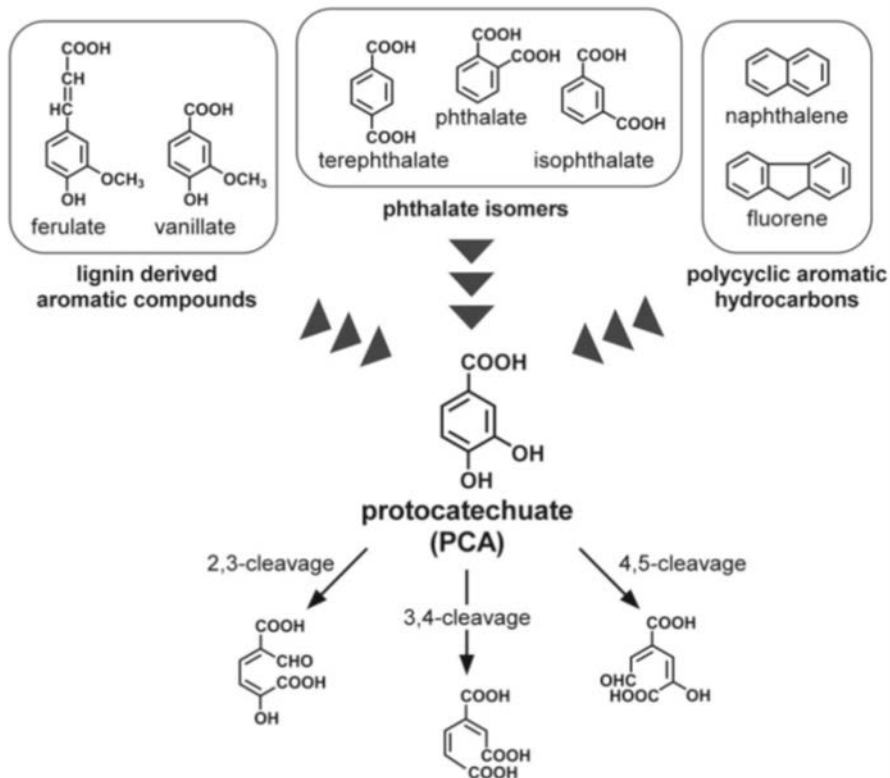


Figure 1.5 Aromatic compounds funneled to protocatechuate (PCA), which is further metabolized through the three distinct ring-cleavage pathways. Reproduced from reference (69).

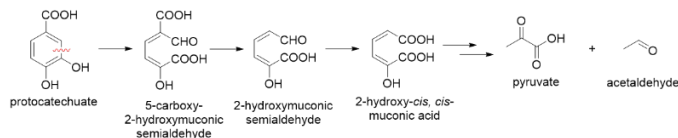
The PCA 4,5-cleavage pathway was reported in 1960 (70) and has been extensively studied since then (71,72). In this pathway, PCA is initially converted to 4-carboxy-2-hydroxymuconate-6-semialdehyde (CHMS) by PCA 4,5-dioxygenase. The subsequent product is finally degraded to pyruvate and oxaloacetate. Genes coding enzyme involved in this pathway were first isolated from *Sphingobium* sp. strain (formerly *Sphingomonas paucimobilis*) SYK-6 (Fig 1.4B). This pathway is confined to a few organisms and is not the focus of our studies.

The PCA 3,4-cleavage pathway is one branch of β -keto adipate pathway, which is a multistep, convergent metabolic route used by many microorganisms. In this pathway, PCA is transformed to β -carboxymuconate firstly by the reaction catalyzed by PCA 3,4-dioxygenase. Then, in a sequence of five enzymatic reactions, β -carboxymuconate is reductively to β -keto adipate, which is further enzymatically transformed into acetyl coenzyme A and succinyl coenzyme A (73,74). The β -keto adipate pathway is widely distributed in microorganisms and the amino acid sequences of the enzymes involved in this pathway are highly conserved in many bacteria including *Acinetobacter baylyi* (formerly *Acinetobacter calcoaceticus*) ADP1, *Pseudomonas putida* PRS2000, *Agrobacterium tumefaciens* A348, *Rhodococcus erythropolis*, and many others (73). In 2000, Davies and co-workers reported the identification of a seven-gene operon encoding homologs of β -keto adipate pathway enzymes in *Streptomyces* sp. strain 2065 and they predicted that enzymes involved in this pathway are widely distributed in the *Streptomyces* genus (75). This prediction was later confirmed by the observation of homologous *pca* gene clusters in the genome of *S. avermitilis* (76) and *S. coelicolor* (17). The following genetic and biochemical analyses conducted by previous lab member in the Sello group first characterized the regulation of *pca* gene cluster in *Streptomyces* bacteria (77). It was found that the expression of *pca* genes is induced by protocatechuate and other aromatic compounds. Gene (*pcaV*) clustered with *pca* locus and transcribed divergently was identified encoding a regulatory factor (PcaV) which can repress the transcription of adjacent *pca* gene. In brief, the repressor PcaV binds to the *pcaV-pcaI* intergenic region and blocks the access of RNA polymerase in the absence of aromatic compounds. In the presence of corresponding aromatic compounds, PcaV is released from the binding site and thus turn on the transcription of *pca* genes.

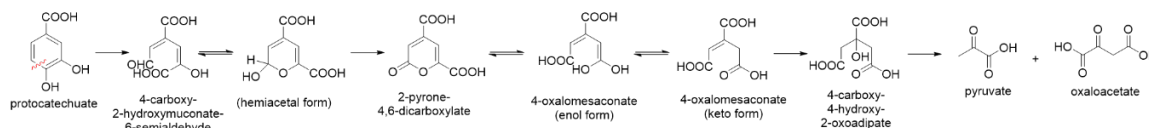
Para-hydroxybenzate (PHB) is one of compounds observed by Crawford in his studies of lignin catabolism by streptomycetes. Under catalysis of the enzyme *para*-hydroxybenzoate hydroxylase (encoded by gene *pobA*), PHB can be converted to PCA and fluxed into the β -keto adipate pathway.

The lower cost of PHB (\$0.62/gram) relative to PCA (\$5.20/gram) makes this compound a more commercially viable carbon source for commodity chemicals.

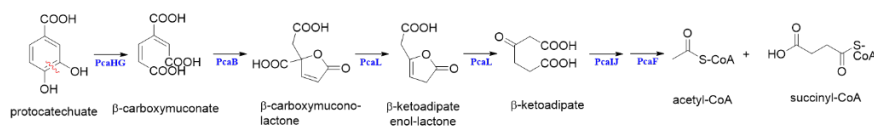
A. PCA 2,3-cleavage pathway



B. PCA 4,5-cleavage pathway



C. PCA 3,4-cleavage pathway and reaction enzymes



D. Organization of the *pca* gene cluster in *Streptomyces coelicolor*

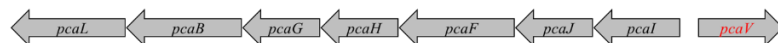


Figure 1.6 (A-C). Details of three PCA ring-cleavage pathways. **(D).** *pca* gene cluster identified in *Streptomyces coelicolor*. The functions of the gene products are as follows: *pcaL* encodes putative γ -carboxymuconolactone decarboxylase/ β -ketoadipate enol-lactone, *pcaB* encodes putative β -carboxymuconate cycloisomerase, *pcaGH* encode procatechuate 3,4-dioxygenase, α - and β -subunits, *pcaF* encodes β -ketoadipyl CoA thiolase, *pcaIJ* encode β -ketoadipate succinyl-CoA transferase α - and β -subunits, and *pcaV* encodes a MarR family transcriptional regulator. The enzymes were labeled in blue and shown below the arrows in the PCA 3,4-cleavage pathway.

The catabolism of PHB is very well-studied in *Acinetobacter* and *Pseudomonas* (78-82). In fact, the properties of enzyme PobaA have been well characterized in a series of *Pseudomonas* species back to 1960s (83-86). Regulation of the *poba* gene has also been studied in both *Pseudomonas* and *Acinetobacter* (78,79,81). Generally, in these two organisms, there is a gene adjacent to *poba* gene called *pobR*. This gene encodes a transcriptional factor PobR. It has been determined that

when bound to *para*-hydroxybenzoate, PobR activates the transcription of *pobA* gene by recruiting RNA polymerase to the promoter of the *pobA* open reading frame. Despite of substantial information about conversion from PHB to PCA, this biological process has never been studied in *Streptomyces* bacteria.

In chapter 2, via a series of BLAST analyses performed on StrepDB, we were able to identify gene candidates involved in the catabolism of PHB and its regulation in *S. coelicolor*. In particular, these analyses revealed a novel regulator of PHB catabolism called PobR, a peculiar IclR-family transcription factor. We also explained the regulatory mechanism about how this transcriptional factor regulates the degradation of PHB in *S. coelicolor*.

1.3.2 Anabolism of aromatic compounds in *Streptomyces* bacteria

Chapters 3 and 4 concern anabolism of tryptophan into high value compounds, namely indolmycin and calcium dependent antibiotic (CDA).

1.3.2.1 Biosynthesis of amino acid tryptophan and its regulation

Tryptophan is the one of the least abundant and the most costly α - amino acid building blocks of proteins. It is derived from the shikimate pathway wherein shikimic acid is converted to chorismate. After the formation of chorismate, a series of reactions involving seven genes and their corresponding enzymes complete the biosynthesis of tryptophan (Fig. 1.7). The details of chorismate biosynthesis will not be discussed here, but the conversion of chorismate to tryptophan will be discussed. The first step is the conversion of chorismate into anthranilate as catalyzed by enzymes encoded by *trpG* and *trpE* in the presence of glutamine. Then enzyme encoded by *trpD* converts anthranilate to phosphoribosyl-anthranilate, which is transformed into indole-3-glycerol phosphate by the sequential activity of *trpF* and *trpC* gene products. In the last step, the indole-3-glycerol phosphate is condensed with L-serine to form tryptophan under the catalysis of a synthetase constituted by *trpA* and *trpB* gene products (50,87). The enzymatic reactions involved

in tryptophan biosynthesis are conserved in all tryptophan producing microorganisms (15,88). The model organism *S. coelicolor* possesses multiple paralogs of the *trp* genes which are spread across the chromosome and organized either in polycistronic units (*trpCIXBA* and *trpC2D2GE2*) or as single open reading frame (*trpE3*, *trpE1*, *priA* [homolog to *trpF*] and *trpD1*) (17,50).

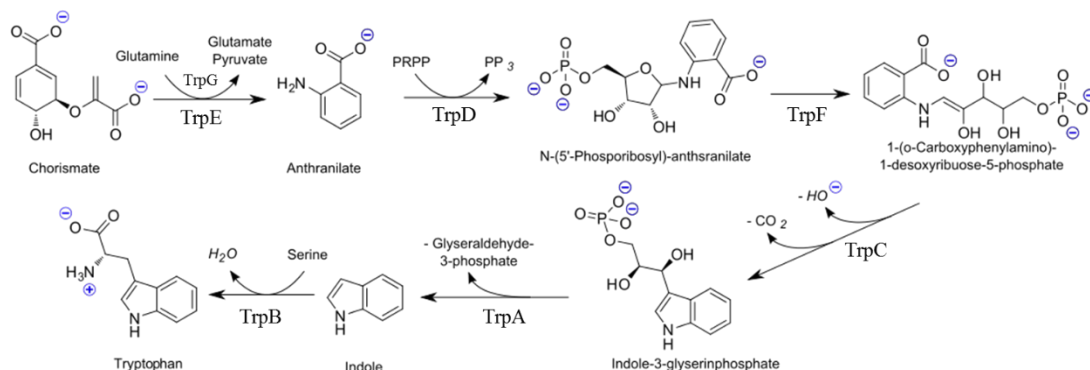


Figure 1.7 Tryptophan biosynthetic pathway with chorismate as the starting point. The enzymes that catalyze the seven reactions in tryptophan formation are shown. Only one of the reactions is reversible. Two of the tryptophan pathway enzymes often function as polypeptide complexes: anthranilate synthetase, consisting of the TrpG and TrpE polypeptides, and tryptophan synthase, consisting of the TrpB and TrpA polypeptides. PRPP: phosphoribosylpyrophosphate.

The biosynthesis of tryptophan is a metabolically expensive and complicated process. As shown in Figure 1.7, the products of four metabolic pathways contribute carbon and/or nitrogen atoms to tryptophan formation (*i.e.*, chorismate, glutamine, phosphoribosylpyrophosphate, and L-serine). The “value” of tryptophan is also high because the chorismate is also used as a precursor of tyrosine and phenylalanine as well as the precursor of *p*-aminobenzoic acid and several other metabolites. Given the complexity and various carbon/nitrogen resources for tryptophan biosynthesis, tryptophan producing microorganisms have regulatory strategies to ensure the production of sufficient levels of chorismate, glutamine, phosphoribosylpyrophosphate, and L-serine and to

balance the incorporation of tryptophan into between polypeptide chains and other related secondary metabolites (87).

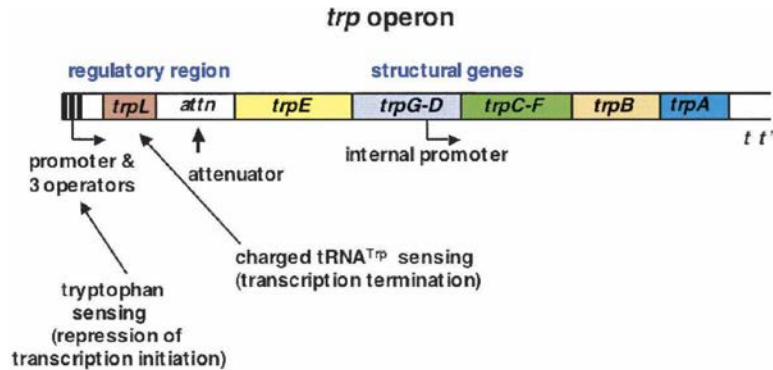


Figure 1.8 Organization of the *trp* operon of *E. coli*. The genes of *E. coli* required for tryptophan biosynthesis from chorismate are organized in a single operon, or transcriptional unit. The relative order of the seven genetic segments, *trpEGDFCBA*, corresponds roughly to the relative order of the respective enzymatic reactions. (t=terminator) Reproduced from reference (87).

Several regulatory mechanisms have been identified in controlling transcription of the *trp* operon in different bacterial species. In the first type, transcription is controlled at the stage of RNA synthesis by a DNA binding transcription repressor. This repressive mode is best studied in *E. coli*. In this bacterium, the genes that are responsible for tryptophan biosynthesis are organized as a single transcription unit (Fig. 1.8) (89). There are three operator sites located in the *trp* operon's promoter region. The repressor (tryptophan-bound form) binds to one or more operators and prevents access by the RNA polymerase, thus inhibiting the transcription initiation of the *trp* operon (90). The second type of regulation of tryptophan biosynthesis is by *cis*-acting RNA regulatory structures that are influenced by ribosome-, protein-, and tRNA-mediated attenuators (Fig. 1.9A) (87). For ribosome-mediated (in *E. coli*) and protein-mediated (in *B. subtilis*) attenuation (87,91), a leader sequence in the mRNA adopts a transcriptional terminator structure in the presence of adequate tryptophan, whereas it adopts an anti-terminator in the absence of charged tryptophan

(Fig. 1.8, 1.9B, 1.9C). The uncharged tRNA-mediated attenuation (T box) has been identified in many Gram-positive bacteria (Fig. 1.9D). The T box is a riboswitch that has high affinity for uncharged tRNA^{trp}. In the absence of tryptophan, the uncharged tRNA^{trp} binds the riboswitch and triggers a conformational change that permits transcription of the *trp* operon.

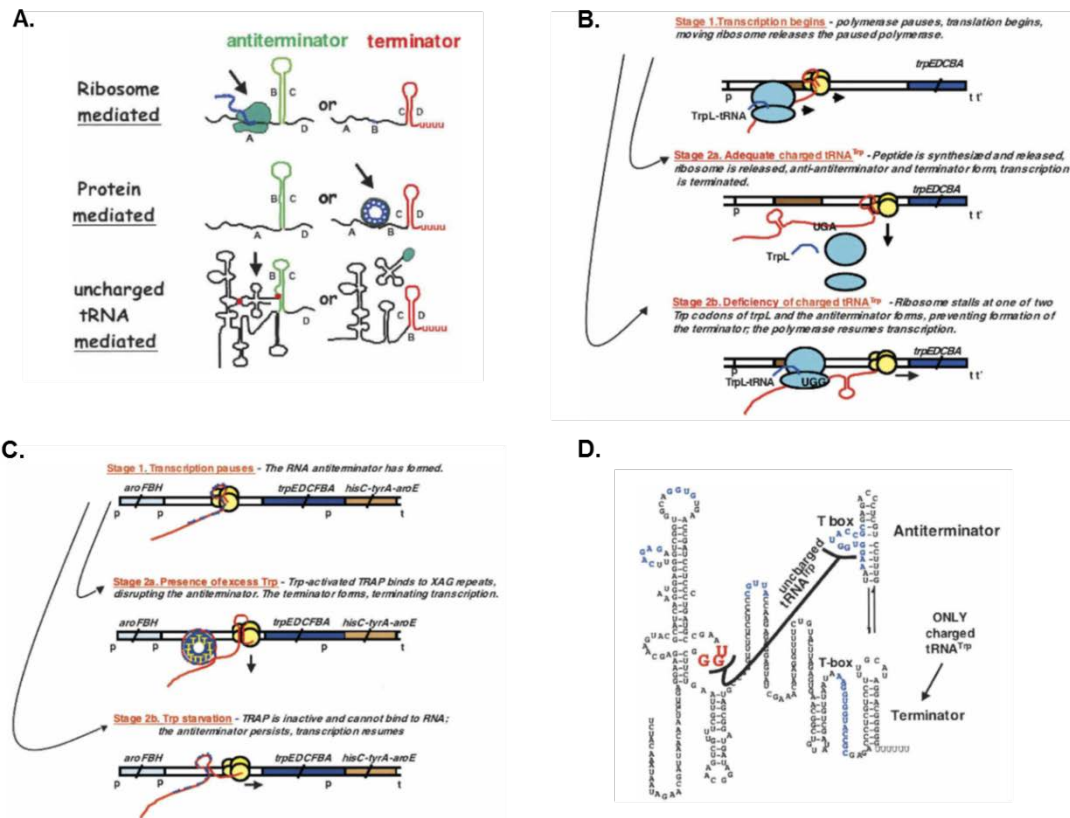


Figure 1.9 RNA-based post-transcriptionally regulation utilized in regulating *trp* operon expression in different bacterial species. (A) Overall three types of transcription attenuation regulatory mechanisms including ribosome-, protein- and uncharged tRNA-mediated attenuation. Alternative RNA hairpin structures can be formed within the leader transcript segments and one of the formed structure is an intrinsic transcription terminator. (B) Detail process of ribosome-mediated attenuation in *E. coli*. (C) Detail process of protein (TRAP)-mediated attenuation in *B. subtilis*. (D) Sequence and presumed alternative structures formed in the T box leader RNA segment of the transcript of the *at* operon of *B. subtilis*. T box binds uncharged tRNA^{trp}. All the figures are reproduced from reference (87).

1.3.2.2 Indolmycin

Indolmycin is a tryptophan-derived bioactive secondary metabolite produced by *Streptomyces griseus* ATCC 12648 (formerly named *Strptomycetes albus*) (92,93). The antibiotic was first isolated in 1960 and was patented by Pfizer in 1965. It had good activity against *Staphylococci* strains that were resistant to several clinically used antibacterial drugs (94,95). By virtue of its structural similarity to L-tryptophan, indolmycin acts as a potent inhibitor of bacterial tryptophanyl-tRNA synthetase (TrpRS, 50% inhibitory concentration [IC₅₀] = 9.25 nM for *E. coli* TrpRS) (96). Importantly, it displays minimal activity against several the eukaryotic counterpart (IC₅₀ = 4.04 mM for bovine liver TrpRS) (97). Unfortunately, the compound showed relatively low bioactivity against several common bacterial pathogens, including *Streptococci*, *Enterococci*, and members of the family *Enterobacteriaceae* (96). The spectrum of antibacterial activity is restricted mainly to Gram-positive bacteria as indolmycin is transported into Gram-positive bacteria whose system for uptake of tryptophan can also act on indolmycin. The hydrophobic characteristic of indolmycin prevents its penetration through the hydrophilic barrier of the Gram-negative outer membrane (98,99). Pfizer stopped development of indolmycin after the discovery that it exhibited liver toxicity in rats (100).

Indolmycin analogs with enhanced hydrophilic properties, 5-hydroxyindolmycin and 5-methoxyindolmycin did show a moderate increase in antimicrobial activity compared with indolmycin (99,101). However, these analogs were not suitable for further development. The interest was renewed when Takeda Pharmaceuticals (Japan) discovered that indolmycin showed potent activity against clinical isolates of *Helicobacter pylori* (MICs = 0.008 to 0.031 µg/ml) through another effect other than inhibiting TrpRS (96). Later, it was further reported that indolmycin demonstrated good activity against MRSA (methicillin-resistant *S. aureus*) and strains that are resistant to mupirocin or fusidic acid (102). Thus, indolmycin was proposed to be a candidate for development as a topical anti-bacterial agent.

Aside from its biological activity, indolmycin has attracted the attention of both chemists and biochemists because of its oxazolinone core structure. It is the only known secondary metabolite with this unusual structure. Chemists have reported several total synthesis routes to indolmycin and its analogs (103-106). Biochemists have been working on unveiling the biosynthesis mystery of indolmycin since 1970s. Floss and co-workers first established that tryptophan, arginine, and methionine are indolmycin precursors and later that β -methylindolepyruvate and indolmycenic acid are intermediates through feeding the producer bacteria with isotope-labeled precursors (51,107,108). Based on the above observations, they proposed the biosynthesis pathway of indolmycin, which later remained as an unsolved problem for almost half century until 2015 (52).

In this thesis, we aimed to discover the indolmycin biosynthetic gene locus in *S. griseus* ATCC 12648. Our strategy for identification of biosynthetic gene cluster was predicated on the idea that antibiotic resistance and biosynthetic genes are clustered. Thus, we used BLAST to look for known indolmycin-resistance genes. From the literature, auxiliary TrpRS genes in bacteria were reported to be associated with the resistance to indolmycin. In 2002, Dieter Söll and co-workers discovered two TrpRS genes in *S. coelicolor*, *trpRS1* and *trpRS2*, that share 47% amino acid similarity (109). Biochemical analyses revealed that *trpRS2* is indolmycin-sensitive, while *trpRS1* is indolmycin-resistant (K_i for TrpRS1 as 150 times greater than the K_i for TrpRS2). The kinetic results indicated that TrpRS2 is the housekeeping TrpRS because it catalyzed tryptophanyl-tRNA formation more efficiently than TrpRS1 *in vitro*. The transcription of *trpRS2* gene is constitutively on, while the transcription of *trpRS1* gene is induced only when the bacteria are exposed to indolmycin, which was further elucidated through the regulation of attenuator located in the 5' leader of *trpRS1* gene (18). Later, the SAV4725 in *S. avermitilis* ATCC 31267 and SGR3809 in *S. griseus* NBRC 13350 were identified as the auxiliary TrpRS genes that confer high-level resistance to indolmycin with the difference that SAV4725 was transcribed in the response to indolmycin, while SGR3809 was constitutively transcribed. *S. griseus* ATCC 12648 is the producer of indolmycin (110).

Theoretically, this bacterium is expected to be highly resistant to indolmycin. In chapter 3, the indolmycin-resistant mechanism of *S. griseus* ATCC 12648 was systematically studied and used to discover the indolmycin biosynthetic genes.

1.3.2.3 Calcium Dependent Antibiotic (CDA)

In 1983, Hopwood and co-workers isolated a substance from *Streptomyces coelicolor* that inhibits the growth of Gram-positive bacteria in the presence of Ca^{2+} ions. The compound was thus named the calcium-dependent antibiotic (CDA) (Fig. 1.3) (111). CDA is a non-ribosomal, cyclic lipopeptide consisting of 11 amino acid residues (including both proteinogenic and non-proteinogenic amino acids) with 6-carbon hydrocarbon chain at its N-terminus (112,113). It is similar in structure and mechanism with other acidic lipopeptide antibiotics, including friulimicin, amphomycin, and daptomycin (114). In daptomycin's mechanism of action, coordination to Ca^{2+} enable it to aggregate and penetrate the membrane of Gram-positive bacteria. The subsequent disruption of the membrane results in the loss of potassium ions, depolarization of membrane potential, and ultimately cell death (115). This potent and distinctive mechanism of action makes the daptomycin-related compounds, including CDA, important weapons against severe antibiotic resistant pathogens, such as methicillin-resistant *S. aureus* strains (MRSA) and vancomycin-resistant *Enterococci* (VRE) (114). In 2003, daptomycin successfully completed phase III clinical trials and became the first acidic lipopetide antibiotic of its class to be approved for clinical use under the trade name Cubicin (116).

The biosynthesis of CDA has long been a subject of interest. The location of the CDA gene cluster on chromosome was first determined with degenerate probes targeting known conserved motifs of nonribosomal peptide synthetases (NRPSs) (117). Later, the 82 kb CDA gene cluster, which comprises approximately 1% of the genome, was sequenced in entirety as part of the *S. coelicolor* genome project (17,117). The CDA gene cluster is located in within the “core” region of the linear chromosome and consists of at least 40 ORFs. Genes that are at the boundaries of the cluster are

SCO3210 (encoding DAHP synthetase homolog) and SCO3249 (encoding acyl carrier protein [ACP] homolog) (Fig. 1.10) (113).

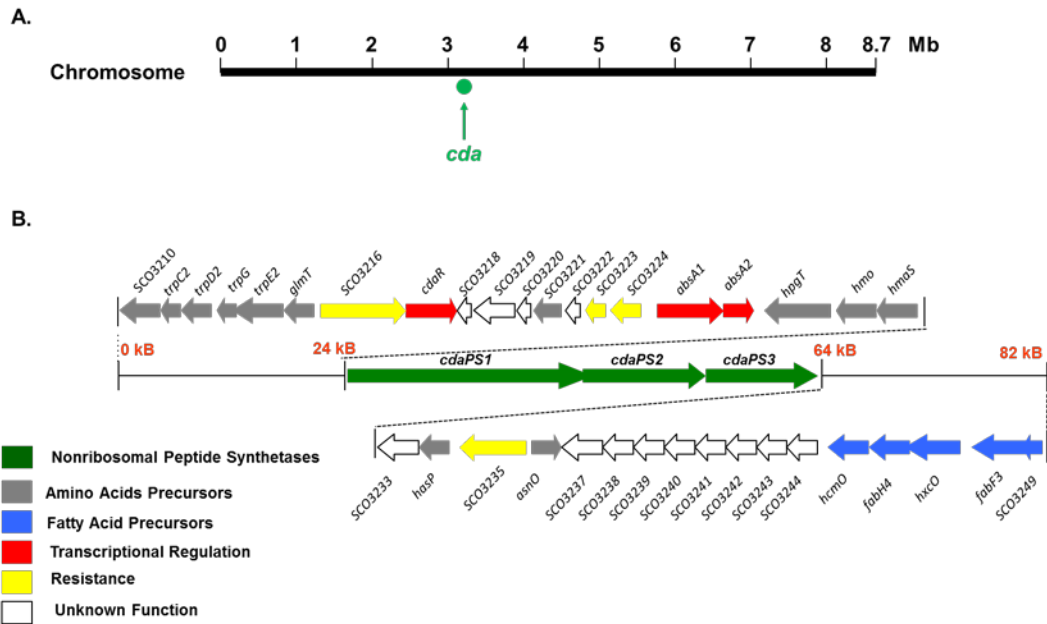


Figure 1.10 Overview of CDA gene cluster. (A) The location of CDA gene cluster on chromosome of *S. coelicolor*. (B) Organization of the CDA biosynthetic gene cluster. Genes are color-coded according to their functions and contributions to the CDA production. Figure B is reproduced from reference (113).

Noticeably, there is a tryptophan biosynthesis locus (SCO3211-3214, *trpCDGE*) within the CDA gene cluster. As the peptide core of CDA contains D- and L-tryptophan residues, it was proposed that the local cluster may ensure adequate tryptophan for CDA biosynthesis at the appropriate stage in the life cycle, independent of the needs of protein synthesis (17).

The presence of the tryptophan biosynthesis genes in the CDA biosynthetic gene cluster raised questions about metabolic flux in antibiotic biosynthesis. There has been much interest in the flux of amino acids into the CDA biosynthetic pathway. A. M. Puglia and co-workers reported that deletion of gene (*kynU*) involved in tryptophan catabolism via kynurenine formation leads to increased amount of CDA, which is likely due to the accumulation of direct precursor, tryptophan

(118). Recently, an integrated study based on quantitative reverse transcription-PCR (qRT-PCR) revealed that tryptophan specifically induces the transcription of tryptophan biosynthetic genes present in the CDA gene cluster. In addition, tryptophan supplementation increased the production of CDA. It was explained that the stimulatory effect on CDA production may not be due to the direct precursor feeding, but due to the possibility that tryptophan induces the transcription of the CDA gene cluster via the transcriptional regulator *cdaR* (50). Chapter 4 in this thesis is focused on tryptophan biosynthetic clusters within and outside of the CDA biosynthetic gene cluster and their relevance in CDA production.

1.4 References

1. Perlman, D. (1980) Some problems on the new horizons of applied microbiology. *Dev. Indust. Microbiol.*, **21**, 15-23.
2. Bourdichon, F., Casaregola, S., Farrokh, C., Frisvad, J.C., Gerds, M.L., Hammes, W.P., Harnett, J., Huys, G., Laulund, S., Ouwehand, A. *et al.* (2012) Food fermentations: microorganisms with technological beneficial use. *Int J Food Microbiol*, **154**, 87-97.
3. Mosbach, K., Birnbaum, S., Hardy, K., Davies, J. and Bülow, L. (1983) Formation of proinsulin by immobilized *Bacillus subtilis*. *Nature*, **302**, 543-545.
4. Luengo, J.M., García, B., Sandoval, A., Naharro, G. and Olivera, E.a.R. (2003) Bioplastics from microorganisms. *Current Opinion in Microbiology*, **6**, 251-260.
5. Spadiut, O., Capone, S., Krainer, F., Glieder, A. and Herwig, C. (2014) Microbials for the production of monoclonal antibodies and antibody fragments. *Trends Biotechnol*, **32**, 54-60.
6. Atlas, R.M. and Hazen, T.C. (2011) Oil biodegradation and bioremediation: a tale of the two worst spills in U.S. history. *Environ Sci Technol*, **45**, 6709-6715.
7. Dash, H.R. and Das, S. (2012) Bioremediation of mercury and the importance of bacterial mer genes. *International Biodeterioration & Biodegradation*, **75**, 207-213.
8. Allard, A.-S. and Neilson, A.H. (1997) Bioremediation of organic waste sites: A critical review of microbiological aspects. *International Biodeterioration & Biodegradation*, **39**, 253-285.
9. Hopwood, D.A. (2007) *Streptomyces in nature and medicine: the antibiotic makers* (Oxford University Press, Inc. New York).

10. Goodfellow, M., Williams, S.T. and Mordarski, M. (1988) *Actinomycetes in Biotechnology* (Academic Press, Inc., San Diego).
11. Nett, M., Ikeda, H. and Moore, B.S. (2009) Genomic basis for natural product biosynthetic diversity in the actinomycetes. *Nat Prod Rep*, **26**, 1362-1384.
12. Romero-Rodriguez, A., Robledo-Casados, I. and Sanchez, S. (2015) An overview on transcriptional regulators in *Streptomyces*. *Biochim Biophys Acta*, **1849**, 1017-1039.
13. Holt, J.G., Krieg, N.R., Sneath, P.H.A., Staley, J.T. and Williams, S.T. (2000) *Bergey's Manual of Determinative Bacteriology* (Ninth Edition, LIPPINCOTT WILLIAMS & WILKINS, Philadelphia).
14. Horinouchi, S. (2002) A microbial hormone, A-factor, as a master switch for morphological differentiation and secondary metabolism in *streptomyces griseus*. *Frontiers in Bioscience*, **7**, d2045.
15. Hodgson, D.A. (2000) Primary metabolism and its control in streptomycetes: A most unusual group of bacteria. **42**, 47-238.
16. van Wezel, G.P. and McDowall, K.J. (2011) The regulation of the secondary metabolism of *Streptomyces*: new links and experimental advances. *Nat Prod Rep*, **28**, 1311-1333.
17. Bentley, S.D., Chater, K.F., Cerdeno-Tarraga, A.M., Challis, G.L., Thomson, N.R., James, K.D., Harris, D.E., Quail, M.A., Kieser, H., Harper, D. *et al.* (2002) Complete genome sequence of the model actinomycete *Streptomyces coelicolor* A3(2). *Nature*, **417**, 141-147.
18. Vecchione, J.J. and Sello, J.K. (2010) Regulation of an auxiliary, antibiotic-resistant tryptophanyl-tRNA synthetase gene via ribosome-mediated transcriptional attenuation. *J Bacteriol*, **192**, 3565-3573.
19. Shirato, S. and Nagatsu, C. (1965) Fermentation studies with *Streptomyces griseus*. I. Carbohydrate sources for the production of protease and streptomycin. *Appl Microbiol*, **13**, 669-672.
20. Gupte, M.D. and Kulkarni, P.R. (2002) A study of antifungal antibiotic production by *Streptomyces chattanoogensis* MTCC 3423 using full factorial design. *Letters in Applied Microbiology*, **35**, 22-26.
21. Sanchez, J., Yague, P. and Manteca, A. (2012) New Insights in Fermentations. *Ferment Technol*, **1**.
22. Bhasin, S. and Modi, H.A. (2012) Optimization of Fermentation Medium for the Production of Glucose Isomerase Using *Streptomyces* sp. SB-P1. *Biotechnol Res Int*, **2012**, 874152.
23. Mullin, R. (2011) The Next Generation In Genome Sequencing, Advances in technology create new challenges in data analysis. *Chemical & Engineering News*, **89**, 17-21.

24. Roberts, R.J., Carneiro, M.O. and Schatz, M.C. (2013) The advantages of SMRT sequencing. *Genome Biol*, **14**.
25. Rigden, D.J., Fernandez-Suarez, X.M. and Galperin, M.Y. (2016) The 2016 database issue of Nucleic Acids Research and an updated molecular biology database collection. *Nucleic Acids Res*, **44**, D1-6.
26. Mizrachi, I. (2002) The NCBI Handbook.
27. (2015) EMBL member states (European Molecular Biology Laboratory).
28. Artimo, P., Jonnalagedda, M., Arnold, K., Baratin, D., Csardi, G., de Castro, E., Duvaud, S., Flegel, V., Fortier, A., Gasteiger, E. *et al.* (2012) ExPASy: SIB bioinformatics resource portal. *Nucleic Acids Res*, **40**, W597-603.
29. Casey, R.M. (2005) BLAST Sequences Aid in Genomics and Proteomics. *Business Intelligence Network*.
30. Altschul, S.F., Gish, W., Miller, W., Myers, E.W. and Lipman, D.J. (1990) Basic local alignment search tool. *Journal of Molecular Biology*, **215**, 403-410.
31. Koonin, E.V. (2005) Orthologs, paralogs, and evolutionary genomics. *Annu Rev Genet*, **39**, 309-338.
32. Higgins, D.G. and Sharp, P.M. (1988) CLUSTAL: a package for performing multiple sequence alignment on a microcomputer. *Gene*, **73**, 237-244.
33. Chenna, R. (2003) Multiple sequence alignment with the Clustal series of programs. *Nucleic Acids Research*, **31**, 3497-3500.
34. Larkin, M.A., Blackshields, G., Brown, N.P., Chenna, R., McGettigan, P.A., McWilliam, H., Valentin, F., Wallace, I.M., Wilm, A., Lopez, R. *et al.* (2007) Clustal W and Clustal X version 2.0. *Bioinformatics*, **23**, 2947-2948.
35. Sievers, F., Wilm, A., Dineen, D., Gibson, T.J., Karplus, K., Li, W., Lopez, R., McWilliam, H., Remmert, M., Soding, J. *et al.* (2011) Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol Syst Biol*, **7**, 539.
36. Neuwald, A.F., Liu, J.S. and Lawrence, C.E. (1995) Gibbs motif sampling: detection of bacterial outer membrane protein repeats. *Protein Sci*, **4**, 1618-1632.
37. Thompson, W. (2003) Gibbs Recursive Sampler: finding transcription factor binding sites. *Nucleic Acids Research*, **31**, 3580-3585.
38. Kanehisa, M. (2000) KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Research*, **28**, 27-30.

39. Kanehisa, M., Goto, S., Hattori, M., Aoki-Kinoshita, K.F., Itoh, M., Kawashima, S., Katayama, T., Araki, M. and Hirakawa, M. (2006) From genomics to chemical genomics: new developments in KEGG. *Nucleic Acids Res*, **34**, D354-357.
40. Kanehisa, M., Goto, S., Sato, Y., Kawashima, M., Furumichi, M. and Tanabe, M. (2014) Data, information, knowledge and principle: back to metabolism in KEGG. *Nucleic Acids Res*, **42**, D199-205.
41. Medema, M.H., Blin, K., Cimermancic, P., de Jager, V., Zakrzewski, P., Fischbach, M.A., Weber, T., Takano, E. and Breitling, R. (2011) antiSMASH: rapid identification, annotation and analysis of secondary metabolite biosynthesis gene clusters in bacterial and fungal genome sequences. *Nucleic Acids Res*, **39**, W339-346.
42. Weber, T., Blin, K., Duddela, S., Krug, D., Kim, H.U., Brucoleri, R., Lee, S.Y., Fischbach, M.A., Muller, R., Wohlleben, W. *et al.* (2015) antiSMASH 3.0-a comprehensive resource for the genome mining of biosynthetic gene clusters. *Nucleic Acids Res*, **43**, W237-243.
43. Blin, K., Medema, M.H., Kazempour, D., Fischbach, M.A., Breitling, R., Takano, E. and Weber, T. (2013) antiSMASH 2.0--a versatile platform for genome mining of secondary metabolite producers. *Nucleic Acids Res*, **41**, W204-212.
44. Harrison, J. and Studholme, D.J. (2014) Recently published *Streptomyces* genome sequences. *Microb Biotechnol*, **7**, 373-380.
45. Markowitz, V.M., Chen, I.M., Palaniappan, K., Chu, K., Szeto, E., Grechkin, Y., Ratner, A., Jacob, B., Huang, J., Williams, P. *et al.* (2012) IMG: the Integrated Microbial Genomes database and comparative analysis system. *Nucleic Acids Res*, **40**, D115-122.
46. Pruitt, K.D., Tatusova, T., Brown, G.R. and Maglott, D.R. (2012) NCBI Reference Sequences (RefSeq): current status, new features and genome annotation policy. *Nucleic Acids Res*, **40**, D130-135.
47. Keshavarz, T. and Roy, I. (2010) Polyhydroxyalkanoates: bioplastics with a green agenda. *Curr Opin Microbiol*, **13**, 321-326.
48. Timmis, K.N. (2010) *Handbook of Hydrocarbon and Lipid Microbiology* (Springer-Verlag, Berlin Heidelberg).
49. Chang, C.J., Floss, H.G., Hook, D.J., Mabe, J.A., Manni, P.E., Martin, L.L., Schroder, K. and Shieh, T.L. (1981) The biosynthesis of the antibiotic pyrrolnitrin by *Pseudomonas aureofaciens*. *J Antibiot* (Tokyo), **34**, 555-566.
50. Palazzotto, E., Renzone, G., Fontana, P., Botta, L., Scaloni, A., Puglia, A.M. and Gallo, G. (2015) Tryptophan promotes morphological and physiological differentiation in *Streptomyces coelicolor*. *Appl Microbiol Biotechnol*, **99**, 10177-10189.
51. Hornemann, U., Hurley, L.H., Speedie, M.K. and Floss, H.G. (1971) Biosynthesis of indolmycin. *Journal of the American Chemical Society*, **93**, 3028-3035.

52. Du, Y.L., Alkhalaf, L.M. and Ryan, K.S. (2015) In vitro reconstitution of indolmycin biosynthesis reveals the molecular basis of oxazolinone assembly. *Proc Natl Acad Sci U S A*, **112**, 2717-2722.
53. Hitchcock, M.J. and Katz, E. (1988) Purification and characterization of tryptophan dioxygenase from *Streptomyces parvulus*. *Arch Biochem Biophys*, **261**, 148-160.
54. Hartley, D.L. and Speedie, M.K. (1984) A tryptophan C-methyltransferase involved in streptonigrin biosynthesis in *Streptomyces flocculus*. *Biochem J*, **220**, 309-313.
55. Gil, J.A., Liras, P., Naharro, G., Villanueva, J.R. and Martin, J.F. (1980) Regulation by aromatic amino acids of the biosynthesis of candicidin by *Streptomyces griseus*. *J Gen Microbiol*, **118**, 189-195.
56. Rubin, E.M. (2008) Genomics of cellulosic biofuels. *Nature*, **454**, 841-845.
57. Martinez, D., Larrondo, L.F., Putnam, N., Gelpke, M.D., Huang, K., Chapman, J., Helfenbein, K.G., Ramaiya, P., Detter, J.C., Larimer, F. *et al.* (2004) Genome sequence of the lignocellulose degrading fungus *Phanerochaete chrysosporium* strain RP78. *Nat Biotechnol*, **22**, 695-700.
58. Bugg, T.D., Ahmad, M., Hardiman, E.M. and Rahmanpour, R. (2011) Pathways for degradation of lignin in bacteria and fungi. *Nat Prod Rep*, **28**, 1883-1896.
59. Bugg, T.D., Ahmad, M., Hardiman, E.M. and Singh, R. (2011) The emerging role for bacteria in lignin degradation and bio-product formation. *Curr Opin Biotechnol*, **22**, 394-400.
60. Masai, E., Katayama, Y. and Fukuda, M. (2007) Genetic and biochemical investigations on bacterial catabolic pathways for lignin-derived aromatic compounds. *Biosci Biotechnol Biochem*, **71**, 1-15.
61. Magnuson, T.S., Roberts, M.A., Crawford, D.L. and Hertel, G. (1991) Immunologic relatedness of extracellular ligninases from the actinomycetes *Streptomyces viridosporus* T7A and *Streptomyces badius* 252. *Appl Biochem Biotechnol*, **28-29**, 433-443.
62. Crawford, D.L. (1978) Lignocellulose decomposition by selected *streptomyces* strains. *Appl Environ Microbiol*, **35**, 1041-1045.
63. Antai, S.P. and Crawford, D.L. (1981) Degradation of softwood, hardwood, and grass lignocelluloses by two *streptomyces* strains. *Appl Environ Microbiol*, **42**, 378-380.
64. Crawford, D.L. (1981) Microbial conversion of lignin to useful chemicals using a lignin-degradation *Streptomyces*. *Biotechnol. Bioeng. Symp*, **11**, 275-291.
65. Pometto, A.L. and Crawford, D.L. (1986) Catabolic Fate of *Streptomyces Viridosporus* T7a-Produced, Acid-Precipitable Polymeric Lignin Upon Incubation with Ligninolytic *Streptomyces* Species and *Phanerochaete-Chrysosporium*. *Appl Environ Microb*, **51**, 171-179.

66. Crawford, R.L. (1975) Novel pathway for degradation of protocatechuic acid in *Bacillus* species. *J Bacteriol*, **121**, 531-536.
67. Crawford, R.L., Bromley, J.W. and Perkins-Olson, P.E. (1979) Catabolism of protocatechuate by *Bacillus macerans*. *Appl Environ Microbiol*, **37**, 614-618.
68. Kasai, D., Fujinami, T., Abe, T., Mase, K., Katayama, Y., Fukuda, M. and Masai, E. (2009) Uncovering the protocatechuate 2,3-cleavage pathway genes. *J Bacteriol*, **191**, 6758-6768.
69. Naofumi Kaminura, E.M. (2014) The Protocatechuate 4,5-Cleavage Pathway: Overview and New Findings. *Biodegradative Bacteria: How Bacteria Degrade, Survive, Adapt, and Evolve* (Springer, Japan), 207-226.
70. Dagley, S., Evans, W.C. and Ribbons, D.W. (1960) New pathways in the oxidative metabolism of aromatic compounds by microorganisms. *Nature*, **188**, 560-566.
71. Arciero DM, L.J., Huynh BH, Kent TA, Münck E. (1983) EPR and Mössbauer studies of protocatechuate 4,5-dioxygenase. Characterization of a new Fe 2+ environment. *J Biol Chem*, **258**, 14981–14991.
72. Arciero DM, O.A., Lipscomb JD. (1985) [17 O]Water and nitric oxide binding by protocatechuate 4,5-dioxygenase and catechol 2,3-dioxygenase. Evidence for binding of exogenous ligands to the active site Fe 2+ of extradiol dioxygenases. *J Biol Chem*, **260**, 14035–14044.
73. Harwood, C.S. and Parales, R.E. (1996) The beta-ketoadipate pathway and the biology of self-identity. *Annu Rev Microbiol*, **50**, 553-590.
74. Seo, J.S., Keum, Y.S. and Li, Q.X. (2009) Bacterial degradation of aromatic compounds. *Int J Environ Res Public Health*, **6**, 278-309.
75. Iwagami, S.G., Yang, K. and Davies, J. (2000) Characterization of the Protocatechuic Acid Catabolic Gene Cluster from *Streptomyces* sp. Strain 2065. *Applied and Environmental Microbiology*, **66**, 1499-1508.
76. Omura, S., Ikeda, H., Ishikawa, J., Hanamoto, A., Takahashi, C., Shinose, M., Takahashi, Y., Horikawa, H., Nakazawa, H., Osonoe, T. *et al.* (2001) Genome sequence of an industrial microorganism *Streptomyces avermitilis*: deducing the ability of producing secondary metabolites. *Proc Natl Acad Sci U S A*, **98**, 12215-12220.
77. Davis, J.R. and Sello, J.K. (2010) Regulation of genes in *Streptomyces* bacteria required for catabolism of lignin-derived aromatic compounds. *Appl Microbiol Biotechnol*, **86**, 921-929.
78. DiMarco, A.A., Averhoff, B. and Ornston, L.N. (1993) Identification of the transcriptional activator pobR and characterization of its role in the expression of pobA, the structural gene for *p*-hydroxybenzoate hydroxylase in *Acinetobacter calcoaceticus*. *J Bacteriol*, **175**, 4499-4506.

79. DiMarco, A.A. and Ornston, L.N. (1994) Regulation of *p*-hydroxybenzoate hydroxylase synthesis by PobR bound to an operator in *Acinetobacter calcoaceticus*. *J Bacteriol*, **176**, 4277-4284.
80. Ornston, L.N. and Parke, D. (1977) The evolution of induction mechanisms in bacteria: insights derived from the study of the beta-ketoadipate pathway. *Curr Top Cell Regul*, **12**, 209-262.
81. Bertani, I., Kojic, M. and Venturi, V. (2001) Regulation of the *p*-hydroxybenzoic acid hydroxylase gene (*pobA*) in plant-growth-promoting *Pseudomonas putida* WCS358. *Microbiology*, **147**, 1611-1620.
82. Paul, D., Chauhan, A., Pandey, G. and Jain, R.K. (2004) Degradation of *p*-hydroxybenzoate via protocatechuate in *Arthrobacter protophormiae* RKJ100 and *Burkholderia cepacia* RKJ200. *Curr Sci India*, **87**, 1263-1268.
83. Hosokawa, K. and Stanier, R.Y. (1966) Crystallization and properties of *p*-hydroxybenzoate hydroxylase from *Pseudomonas putida*. *J Biol Chem*, **241**, 2453-2460.
84. Howell, L.G., Spector, T. and Massey, V. (1972) Purification and properties of *p*-hydroxybenzoate hydroxylase from *Pseudomonas fluorescens*. *J Biol Chem*, **247**, 4340-4350.
85. Entsch, B., Ballou, D.P. and Massey, V. (1976) Flavin-oxygen derivatives involved in hydroxylation by *p*-hydroxybenzoate hydroxylase. *J Biol Chem*, **251**, 2550-2563.
86. Entsch, B. and van Berkel, W.J. (1995) Structure and mechanism of *para*-hydroxybenzoate hydroxylase. *FASEB J*, **9**, 476-483.
87. Yanofsky, C. (2007) RNA-based regulation of genes of tryptophan synthesis and degradation, in bacteria. *RNA*, **13**, 1141-1154.
88. Yanofsky, C. (2000) Transcription Attenuation: Once Viewed as a Novel Regulatory Strategy. *Journal of Bacteriology*, **182**, 1-8.
89. Yanofsky, C., Crawford, I.P. (1987) The tryptophan operon. In *Escherichia coli* and *Salmonella typhimurium*. *Cellular and molecular biology* (eds. F.C. Neidhardt et al., American Society for Microbiology, Washington D.C.), 1452-1472.
90. Lawson, C.L., Benoff, B., Berger, T., Berman, H.M. and Carey, J. (2004) *E. coli* trp repressor forms a domain-swapped array in aqueous alcohol. *Structure*, **12**, 1099-1108.
91. Yanofsky, C. (2004) The different roles of tryptophan transfer RNA in regulating *trp* operon expression in *E. coli* versus *B. subtilis*. *Trends Genet*, **20**, 367-374.
92. Marsh, W.S., Garretson, A.L. and Wesel, E.M. (1960) PA 155 A, B, and X antibiotics produced by a strain of *Streptomyces albus*. *Antibiot Chemother*, **10**, 316-320.

93. Routien, J.B. (1966) Identity of streptomycete producing antibiotic PA155A. *J Bacteriol*, **91**, 1663.
94. Floss, H.G. (1981) Biosynthesis of some aromatic antibiotics. *Antibiotic biosynthesis* (Springer-Verlag, Berlin Heidelberg), **4**, 236-261.
95. Beaulieu, D. and Ohemeng, K.A. (1999) Patents on bacterial tRNA synthetase inhibitors: January 1996 to March 1999. *Expert Opinion on Therapeutic Patents*, **9**, 1021-1028.
96. Kanamaru, T., Nakano, Y., Toyoda, Y., Miyagawa, K.I., Tada, M., Kaisho, T. and Nakao, M. (2001) *In Vitro* and *In Vivo* Antibacterial Activities of TAK-083, an Agent for Treatment of *Helicobacter pylori* Infection. *Antimicrobial Agents and Chemotherapy*, **45**, 2455-2459.
97. Werner, R.G., Thorpe, L.F., Reuter, W. and Nierhaus, K.H. (1976) Indolmycin Inhibits Prokaryotic Tryptophanyl-tRNA Ligase. *European Journal of Biochemistry*, **68**, 1-3.
98. Werner, R.G. (1980) Uptake of indolmycin in gram-positive bacteria. *Antimicrob Agents Chemother*, **18**, 858-862.
99. Werner, R.G. and Demain, A.L. (1981) Directed biosynthesis of new indolmycins. *J Antibiot* (Tokyo), **34**, 551-554.
100. Werner, R.G. and Reuter, W. (1979) Interaction of indolmycin in the metabolism of tryptophan in rat liver. *Arzneimittelforschung*, **29**, 59-63.
101. Hurdle, J.G., O'Neill, A.J. and Chopra, I. (2005) Prospects for aminoacyl-tRNA synthetase inhibitors as new antimicrobial agents. *Antimicrob Agents Chemother*, **49**, 4821-4833.
102. Hurdle, J.G., O'Neill, A.J. and Chopra, I. (2004) Anti-staphylococcal activity of indolmycin, a potential topical agent for control of staphylococcal infections. *J Antimicrob Chemother*, **54**, 549-552.
103. Preobrazhenskaya, M.N., Balashova, E.G., Turchin, K.F., Padeiskaya, E.N., Uvarova, N.V., Pershin, G.N. and Suvorov, N.N. (1968) Total synthesis of antibiotic indolmycin and its stereoisomers. *Tetrahedron*, **24**, 6131-6143.
104. Takeda, T. and Mukaiyama, T. (1980) Asymmetric total synthesis of indolmycin. *Chemistry Letters*, 163-166.
105. Shue, Y.-K. (1996) Total synthesis of (\pm) indolmycin. *Tetrahedron Letters*, **37**, 6447-6448.
106. Witty, D.R., Walker, G., Bateson, J.H., O'Hanlon, P.J., Eggleston, D.S. and Haltiwanger, R.C. (1996) Synthesis of conformationally restricted analogues of the tryptophanyl tRNA synthetase inhibitor indolmycin. *Tetrahedron Letters*, **37**, 3067-3070.
107. Hornemann, U., Hurley, L.H., Speedie, M.K., Guenther, H.F. and Floss, H.G. (1969) Biosynthesis of the antibiotic indolmycin by *Streptomyces griseus*. C-Methylation at the β -carbon atom of the tryptophan side-chain. *J. Chem. Soc. D*, **0**, 245-246.

108. Zee L, H.U., Gloss HG. (1975) Further studies on the biosynthesis of the antibiotic indolmycin in *Streptomyces griseus*. *Biochem Physiol Pflanzen*, **168**, 19-25.
109. Kitabatake, M., Ali, K., Demain, A., Sakamoto, K., Yokoyama, S. and Soll, D. (2002) Indolmycin resistance of *Streptomyces coelicolor* A3(2) by induced expression of one of its two tryptophanyl-tRNA synthetases. *J Biol Chem*, **277**, 23882-23887.
110. Vecchione, J.J. and Sello, J.K. (2009) A novel tryptophanyl-tRNA synthetase gene confers high-level resistance to indolmycin. *Antimicrob Agents Chemother*, **53**, 3972-3980.
111. Kempster, C., Kaiser, D., Haag, S., Nicholson, G., Gnau, V., Walk, T., Gierling, K.H., Decker, H., Zähner, H., Jung, G. *et al.* (1997) CDA: Calcium-Dependent Peptide Antibiotics from *Streptomyces coelicolor* A3(2) Containing Unusual Residues. *Angewandte Chemie International Edition in English*, **36**, 498-501.
112. Kieser T, B.M., Buttner MJ, Chater KF, Hopwood DA. (2000) *Practical Streptomyces Genetics* (John Innes Foundation, Norwich, England).
113. Hojati, Z., Milne, C., Harvey, B., Gordon, L., Borg, M., Flett, F., Wilkinson, B., Sidebottom, P.J., Rudd, B.A.M., Hayes, M.A. *et al.* (2002) Structure, Biosynthetic Origin, and Engineered Biosynthesis of Calcium-Dependent Antibiotics from *Streptomyces coelicolor*. *Chemistry & Biology*, **9**, 1175-1187.
114. Kim, H.B., Smith, C.P., Micklefield, J. and Mavituna, F. (2004) Metabolic flux analysis for calcium dependent antibiotic (CDA) production in *Streptomyces coelicolor*. *Metab Eng*, **6**, 313-325.
115. Pogliano, J., Pogliano, N. and Silverman, J.A. (2012) Daptomycin-mediated reorganization of membrane architecture causes mislocalization of essential cell division proteins. *J Bacteriol*, **194**, 4494-4504.
116. Kirkpatrick, P., Raja, A., LaBonte, J. and Lebbos, J. (2003) Fresh from the pipeline: Daptomycin. *Nat Rev Drug Discov*, **2**, 943-944.
117. Chong, P.P., Podmore, S.M., Kieser, H.M., Redenbach, M., Turgay, K., Marahiel, M., Hopwood, D.A. and Smith, C.P. (1998) Physical identification of a chromosomal locus encoding biosynthetic genes for the lipopeptide calcium-dependent antibiotic (CDA) of *Streptomyces coelicolor* A3(2). *Microbiology*, **144** (Pt 1), 193-199.
118. Zummo, F.P., Marineo, S., Pace, A., Civiletti, F., Giardina, A. and Puglia, A.M. (2012) Tryptophan catabolism via kynurenine production in *Streptomyces coelicolor*: identification of three genes coding for the enzymes of tryptophan to anthranilate pathway. *Appl Microbiol Biotechnol*, **94**, 719-728.

Chapter 2. Catabolism of *para*-hydroxybenzoic acid and its regulation in *Streptomyces coelicolor*

2.1 Introduction

Microbial degradation of aromatic compounds has long attracted the interest of the scientific community. Decades of research have yielded important genetic, biochemical and physiological knowledge about aromatic compound utilization by microbes. Microorganisms have evolved remarkable capabilities to consume these highly stable compounds because they are so prevalent in nature. Indeed, there are abundant aromatic compounds derived from lignin, a constituent of plant biomass (1-3). This aromatic polymer and its building blocks make up as much as one-quarter of the biomass of land plants (4). Utilization of this huge reservoir of renewable carbon to produce biofuels and commodity chemicals is the current scientific focus given the increasing worldwide energy demand and the finite supply of fossil fuels. In addition to lignin, there are also aromatic compounds produced by humans (*e.g.*, benzene, toluene, ethylbenzene and xylene). These prominent environment pollutants are derived from petroleum derivatives such as gasoline (5,6). Pesticides and industrial wastes also contain aromatic and halogenated aromatic constituents, many of which are toxic (7). Studies of aromatic compounds degradation by microbes are often motivated by the need for strategies to remove the toxic and xenobiotic compounds from the environment.

Though aromatic compounds tend to be highly stable, they are consumed by microorganisms. Pathways for utilization of aromatic compounds are widely distributed in the microbial world (8,9). In bacteria, diverse aromatic compounds are enzymatically transformed to common aromatic intermediates, namely catechol and protocatechuate (PCA) (9,10). Remarkably, there are three distinct catabolic pathways through which these aromatic compounds are catabolized to intermediates of the Krebs cycle (10). The predominant pathway for catabolism of catechol and PCA involves the 3,4-cleavage of the diol and the subsequent conversion to β -keto adipate. This eponymous β -keto adipate pathway ultimately yields succinate and acetyl-coenzyme A, intermediates of the Krebs cycle. The other catabolic pathways occur via 2,3-cleavage and 4,5-cleavage of PCA. They have been observed in *Paenibacillus* sp. Strain JJ-1B (11-13) and

Sphingobium sp. strain SYK-6 (14), respectively. The former pathway converts PCA to pyruvate and acetyl-CoA, while the later one generates pyruvate and oxaloacetate (10). The physiological significance of these pathways is that they enable the carbon of aromatic compounds to enter intermediary metabolic pathways that underly cell growth and metabolism (8). Since PCA is a key intermediate in aromatic catabolism and its catabolism yields building blocks of commodity chemicals (*e.g.*, antibiotics, biofuels), reactions that lead to the formation of PCA have attracted the interest of many groups. In this context, the conversion of *para*-hydroxybenzoate (PHB) to PCA has been the subject of numerous investigations. PHB, a well-known product of the microbial degradation of lignin, is highly abundant and cheap (15). It is envisioned that the conversion of PHB to PCA by *para*-hydroxybenzoate hydroxylase can be exploited in microbial production of commodity chemicals (15-20).

Table 2.1 Summary of regulatory proteins of aromatic catabolic pathways in bacteria (21-23)

Family	Numbers of identified protein	Function	Substrates of the pathways
NtrC	19	Transcriptional activators	Toluene, Phenol, Trimethylbenzene, Phenanthrene/Naphthalene, Biphenyl, 2-OH-biphenyl, Aryl esters
LysR	19	Transcriptional activators	Phenol, Catechol, Protocatechuate, Benzoate, 4-OH-benzoate, et al.
AraC	7	Transcriptional activators	Toluene (<i>meta</i> pathway), 4-OH-phenylacetate, Tyramine, Isopropylbenzene, 4-OH-benzoate,
IclR	8	Transcriptional activators	4-OH-benzoate, Protocatechuate, Catechol, 3-OH-phenylpropionate, <i>o</i> -Chlorobenzoate
TCST	9	Transcriptional activators	Toluene, Biphenyl/PCB, Styrene
FNR	2	Transcriptional activators	4-OH-benzoate, Benzoate
MarR	4	One activators/ three repressors	Benzoate, Protocatechuate, Homoprotocatechuate
TetR	1	Transcriptional repressors	<i>p</i> -Cymene
GntR	6	Transcriptional repressors	Phenylacetate, Phenol, Vanillate, Biphenyl/4-Chlorobiphenyl

Because aromatic compound degradation requires many highly reactive enzymes and is secondary to carbohydrate catabolism, microorganisms tend to exert tight regulatory control over aromatic catabolism (21,24). An understanding of the regulatory mechanisms are critical for the exploitation of microorganisms in biotechnology platforms. The regulation is often mediated at the level of transcription. There are various regulatory proteins belonging to different families of prokaryotic transcriptional regulators that control the expression of aromatic catabolic pathways. Table 2.1 is the summary of regulatory proteins of aromatic catabolic pathways in bacteria (21-23). With respect to aromatic degradation in bacteria, the most prevalent pathway is the β -keto adipate pathway for catabolism of protocatechuate and catechol. This pathway is regulated by members of the LysR, IclR and MarR families. Interestingly, members of the LysR and IclR families are all transcriptional activators, whereas the MarR family contains both transcriptional activators and repressors. Studies of these proteins have provided important insights into the regulation of gene expression in bacteria (21).

It is well-known that *Streptomyces* bacteria are well-equipped with pathways for aromatic catabolism and have unusual systems for regulation of these pathways (25-28). For example, the Sello group reported that the β -keto adipate pathway in *S. coelicolor* and other streptomycetes was regulated by a MarR family member (rather than a NtrC, LysR or IclR family member as is the case for other bacteria) (22,23). Recently, we conducted a series of bioinformatic analyses on the genomes of multiple streptomycetes and discovered a large gene encoding a peculiar protein, which is an apparent translational fusion of two IclR type proteins with homology to PobR, a regulator of genes encoding *para*-hydroxybenzoate hydroxylase. The IclR family is named after the transcriptional regulator isocitrate lyase regulator (IclR) of *E. coli* (29,30), which is the most comprehensively characterized member of this group at the genetic and biochemical level (31). Roughly 1000 IclR-type transcriptional factors have been identified (32,33) in both Gram-positive and Gram-negative bacteria as well as Archaea (34-37). Members of the IclR family are involved

in various metabolic processes such as glyoxylate shunt, multidrug resistance, degradation of aromatics, inactivation of quorum-sensing signals, determination of plant pathogenicity, and sporulation (37). The sizes of IclR-type proteins are generally 238 to 280 amino acid residues (25 to 30 kDa). There is a helix-turn-helix (HTH) DNA binding motif at the N-terminus of IclR-type proteins, while the C-terminus is involved in protein oligomerization and effector binding (38). This family comprises regulators acting as repressors, activators and proteins with a dual role, namely activators for some genes and repressors for others (37). Interestingly, all of the characterized IclR regulators of aromatic catabolism are described as activators (21). The most studied members of the family are from *Acinetobacter calcoaceticus* and fluorescent *Pseudomonads*, which are known to grow rapidly on aromatic substrates (8).

The aforementioned translation fusion of two IclR family members in *S. coelicolor* attracted our attention for two reasons. First, a protein of this organization is unique to biology. Second, homologs of its two domains are known to regulate the transcription of genes encoding the *para*-hydroxybenzoate hydroxylase that converts PHB into the key intermediate of aromatic catabolism PCA. These observations led us to carry out a series of experiments to characterize the regulation of the *pobA* genes and the structure and function of this unusual transcription factor. Unlike all the other IclR-family members that regulate aromatic catabolism, the PobR protein in *S. coelicolor* acts as a transcriptional repressor rather than as an activator. Bioinformatic analyses, *in vivo* transcriptional assays, electrophoretic mobility shift assays (EMSAs), DNase I footprinting, and isothermal calorimetry (ITC) were used to elucidate the regulatory mechanism of PobR. We found that PobR loses its high affinity for DNA (*i.e.*, the *pobA* operator) in the presence of PHB, the inducer of *pobA* transcription. Size-exclusion chromatography and small-angle X-ray scattering were performed to study the effect of binding ligand on the oligomerization state of PobR. In addition, the crystal structure of PobR bound to its ligand yielded new insights into the molecular

mechanisms of IclR family members. Collectively, our studies have expanded existing views of gene expression in streptomycetes.

2.2 Materials and Methods

2.2.1 Bacterial strains and culture conditions

A complete list of bacterial strains employed in this study is provided in Table 2.2. *Escherichia coli* strains DH5 α , ET12567/pUZ8002 and BL21-Gold (DE3) were grown on Luria-Bertani medium at 37°C for routine sub-cloning (39). While *E. coli* strain BW25113/pIJ790 was grown at 30°C to maintain selection of pIJ790. *S. coelicolor* strains were grown at 30°C on mannitol soya flour medium (SFM), yeast extract-malt extract medium (YEME), minimal liquid medium (NMMP) (40) or minimal medium (MM) [3 g Gelzan, 1 g (NH₄)₂SO₄, 5 g K₂HPO₄, 2 g MgSO₄·7H₂O, 0.1 g FeSO₄·7H₂O per liter]. For minimal medium, carbon sources were sterilized separately and added to media at the following the final concentrations: glucose, 5 g per liter; *para*-hydroxybenzoic acid (PHB), 20 mM; protocatechuic acid (PCA), 20 mM; and 1% (vol/vol) EtOH as negative control. Genomic DNA was isolated from *S. coelicolor* grown in YEME and total mRNA was isolated from *S. coelicolor* from NMMP. MM was used for the different carbon sources growth experiments. For selection of *E. coli*, antibiotics were used at the following concentration: ampicillin (100 μ g/ml), apramycin (50 μ g/ml), chloramphenicol (25 μ g/ml), hygromycin (75 μ g/ml), kanamycin (50 μ g/ml), spectinomycin (50 μ g/ml) and streptomycin (50 μ g/ml). For selection of *S. coelicolor*, apramycin (50 μ g/ml), hygromycin (75 μ g/ml), spectinomycin (200 μ g/ml) and streptomycin (10 μ g/ml) were used. In conjugations between *E. coli* and *S. coelicolor*, nalidixic acid (20 μ g/ml) was used to counterselect *E. coli*.

2.2.2 Plasmids and primers

All of the plasmids employed in this study are listed in Table 2.3. DNA sequencing was performed by Davis Sequencing (Davis, CA). Table 2.4 shows the primers used in this work, all of which were synthesized by Invitrogen. PCR was performed with *Taq* (Invitrogen), *Pfu* (Stratagene, Agilent Technologies), Crimson *Taq* DNA polymerase (New England Biolabs) and Expand High Fidelity PCR system (Roche). All PCR reactions with *Taq*, *Pfu*, Crimson *Taq* DNA polymerase and Phusion High-Fidelity DNA polymerase were performed with 5% (vol/vol) DMSO (40).

2.2.3 Construction of SCO3084, SCO1308, SCO0266 and SCO3209 null mutants

PCR-targeted mutagenesis was used to replace the *S. coelicolor* SCO3084, SCO0266 and SCO3209 genes with an apramycin resistance cassette, *apr* (41). The requisite PCR products were amplified from the apramycin-resistance gene insert of pIJ773 using primers SCO3084 KO For/SCO3084 KO Rev, and SCO0266 KO For/SCO0266 KO Rev and SCO3209 KO For/SCO3209 KO Rev. The PCR products were individually introduced into *E. coli* BW25113/pIJ790 containing cosmid StE25 (SCO3084) or cosmid 11B10 (SCO0266) or cosmid StE8 (SCO3209) and λ RED recombinase. The resultant recombinant cosmids, StE25 Δ SCO3084::*apr*, 11B10 Δ SCO0266::*apr* and StE8 Δ SCO3209::*apr*, were introduced individually via non-methylating *E. coli* strain ET12567/pUZ8002 into *S. coelicolor* M600 by conjugation as previously described (42). M600 is a plasmid free-derivative of the wild-type strain. Exconjugants lacking the SCO3084, SCO0266 and SCO3209 genes were selected by apramycin resistance and kanamycin sensitivity. Gene replacements were confirmed via PCR on isolated genomic DNA from the null mutants using primers SCO3084 KO Det For/SCO3084 KO Det Rev for *S. coelicolor* B1653 Δ SCO3084::*apr*, and primers SCO0266 KO Det For/SCO0266 KO Det Rev for *S. coelicolor* B1655 Δ SCO0266::*apr* and primers SCO3209 KO Det For/SCO3209 KO Det Rev for *S. coelicolor* B1656 Δ SCO3209::*apr*. The same strategy was also performed to construct strains B1654 Δ SCO1308::*spec*. SCO1308 gene was replaced with spectinomycin resistance cassette, *spec*. The requisite PCR products were

amplified from spectinomycin-resistance cassette from pIJ770 using primers SCO1308 KO For/SCO1308 KO Rev. Recombinant cosmid 5F05 Δ SCO1308::spec was introduced into *S. coelicolor* M600, yielding *S. coelicolor* B1654. Exconjugants were selected by spectinomycin resistance and kanamycin sensitivity. Gene replacements were confirmed via PCR on isolated genomic DNA from the null mutants using primers SCO1308 KO Det For/SCO1308 KO Det Rev.

2.2.4 Complementation of *pobA* (SCO3084) and *pobR* (SCO3209) null mutants

A 1,731-bp fragment was excised from cosmid StE25 containing *pobA* (SCO3084) ORF and 407-bp upstream of its translational start site by restriction digest with BstXI and NcoI. The fragment was treated with DNA polymerase I, Large (Klenow) Fragment (New England Biolabs) according to the manufacturer's protocol. The blunt-ended fragment then was ligated into the SmaI site of pBluescript KS+ to yield pJS878. The fragment was then excised from pJS878 with SpeI and EcoRV and ligated into pMS81 pre-treated with the same enzymes to yield pJS880. pMS81 is a site-specific integrating vector that inserts into the Φ BT1 *attB* site of *S. coelicolor* (SCO4848) (43). pJS880 was transformed into non-methylating *E. coli* strain ET12567/pUZ8002 and introduced into *S. coelicolor* via conjugation (42), yielding *S. coelicolor* B1657 Δ SCO3084::*apr* pMS81-SCO3084. Hygromycin-resistant exconjugants were purified by single colony isolation. Phenotypic analyses of *S. coelicolor* B1657 were carried out by growing the strain on media having different carbon sources.

A 1,981-bp fragment was excised from cosmid StE8 containing *pobR* (SCO3209) ORF and 367 upstream base pairs by restriction digest with BlnI and BspEI. The fragment was treated with DNA polymerase I, Large (Klenow) Fragment (New England Biolabs) according to the manufacturer's protocol. The blunt-ended fragment then was ligated into the SmaI site of pBluescript KS+ to yield pJS879. The fragment was then excised from pJS879 and ligated into pMS81 as previously described to yield pJS881, which was introduced into the *pobR* null mutant strain *S. coelicolor*

B1658 Δ SCO3209::*apr* as previously described. Phenotypic analysis of *S. coelicolor* B1658 was accomplished via RT-PCR analysis of *pobA* (SCO3084).

2.2.5 Assessment of nutritional capacity

Strains were cultivated on minimal medium containing different carbon sources as described above. For the tested five strains (*S. coelicolor* M600, *S. coelicolor* B1653, *S. coelicolor* B1654, *S. coelicolor* B1657 and *S. coelicolor* B1658), 2×10^9 spores were streaked on MM plate with Q-tips. Growth was assessed after incubation at 30°C for 5 days.

2.2.6 Transcriptional analyses, RNA isolation and RT-PCR

Wild-type *S. coelicolor* M600, Δ SCO0266::*apr*, Δ SCO3209::*apr* and Δ SCO3209::*apr* pMS81-SCO3209 shaken liquid cultures were grown in NMMP for 14 h (mid-exponential phase) after which either *para*-hydroxybenzoic acid (PHB), protocatechuic acid (PCA), or ethanol as a negative control were added. Both PHB and PCA used in this experiments were dissolved in ethanol. The final inducing concentration of aromatic compounds was 2 mM. After a 1 h induction period, an aliquot of culture sufficient to attain a cell pellet volume of 100 μ l was removed for total mRNA isolation. The cells were washed once with 10.3% (w/v) sucrose solution followed by the addition of 100 μ L of 10 mg/ml lysozyme solution (50 mM Tris – HCl, 1mM EDTA, pH 8.0). The cells were incubated at 37°C for 15 min. Total RNA was isolated using the RNeasy Mini Kit (Qiagen) following the manufacturer's protocol. The concentration of each purified, DNA-free total RNA isolated was measured with a NanoDrop ND-1000 spectrophotometer. An equal amount of total RNA (1200 ng) was employed in all RT-PCR. RT-PCR was performed with the OneStep RT-PCR kit (Qiagen) according to the manufacturer's protocol. A 221-bp complementary DNA (cDNA) corresponding to the *pobA* (SCO3084) transcript were detected with SCO3084 RT For and SCO3084 RT Rev primers. A 486-bp cDNA corresponding to the *hrdB* (SCO5820) transcript, was detected with *hrdB* RT For/Rev primers as a positive control. All primer sequences are listed in

Table 2.4. The PCR program used for detection of both transcripts in *S. coelicolor* was 50°C for 30 min, 95°C for 15 min, 30 cycles of 94°C for 30 s, 56°C for 30 s, and 72°C for 1 min, followed by a final elongation at 72°C for 10 min. Detection of possible contaminating DNA in RNA samples was accomplished by PCR with *Pfu* polymerase under the same conditions. No bands were observed in the controls, indicating that all RT-PCR products correspond to amplification of RNA transcripts.

2.2.7 Mapping of the *pobA* (SCO3084) transcription start site

The transcriptional start sites of the *pobA* (SCO3084) were identified with the 5'-RACE System - Rapid Amplification of cDNA Ends, Version 2.0 (Invitrogen). For mapping the *pobA* open reading frame, total RNA was isolated, as previously described, from an NMMP culture of wild-type *S. coelicolor* M600 treated with 2 mM *para*-hydroxybenzoic acid to elevate levels of the *pobA* transcript. The first-strand cDNA synthesis was performed using 20 pmol of primer SCO3084 GSP1 and 2000 ng of RNA template according to the manufacturer's protocol for high GC-content transcripts. After cDNA synthesis, the reaction mixture was treated with an RNase Mix to remove template RNA and purified using the S.N.A.P. column procedure. An oligo-dC tail was added to the purified cDNA using the TdT-tailing reaction. Amplification of dC-tailed cDNA was accomplished using a 2.5 μ l aliquot of the preceding reaction as template, the primer SCO3084 GSP2 and abridged Anchor primer (AAP, supplied with kit) and *Taq* DNA polymerase. The purified PCR product was then ligated into the pGEM-T easy vector and transformed into the *E. coli* strain DH5 α to yield pJS882. DNA sequencing of cloned inserts was performed by Davis Sequencing (Davis, CA).

2.2.8 Molecular cloning and protein purification

For PobA production, the *pobA* gene (SCO3084) was PCR-amplified from cosmid StE25 and ligated into pBluescript KS+ to yield pJS891. The fragment containing the *pobA* gene was excised

and ligated into the vector pET-28a(+) (Novagen) using the restriction sites NdeI and HindIII to yield pJS892. The plasmid was transformed into *E. coli* BL21-Gold (DE3) Competent Cells (Agilent) and subsequently inoculated into 50 ml cultures of LB containing 50 mg/L kanamycin. The cells were grown at 37°C shaking at 250 rpm to an OD₆₀₀ between 0.5-0.9, at which point the cells were transferred to 4°C for 1 hour. The cultures were then induced with 1mM IPTG and grown at 18°C for 18 h.

Purification of PobA was carried out using Ni-NTA Spin Kit (Qiagen) following the manufacturer's protocol. The pellet derived from 50 ml cell culture was resuspended in 2 ml lysis buffer (50 mM Tris pH 7.4, 300 mM NaCl, 10 mM imidazole) containing 150 units Benzonase Nuclease (Sigma). The cells were lysed by sonication (1 s on, 2 s off for 5 min) using a 450 sonifier digital cell disruptor (Branson) and the cell debris was removed by centrifugation (12,000 x g/30 min/4°C. The cleared lysate containing the 6xHis-tagged PobA was loaded onto the Ni-NTA spin column (from the kit) and His-tagged PobA was eluted with elution buffer (50 mM Tris pH 7.4, 300 mM NaCl, 500 mM imidazole). Protein was further exchanged to storage buffer (50 mM Tris pH 7.4, 300 mM NaCl, 10% glycerol) using Amicon Ultra 30K cutoff centrifugal filters (Merck Millipore). The purification of PobA was analyzed by SDS-PAGE electrophoresis.

For PobR-FL, the *pobR* gene (SCO3084) was PCR-amplified from *S. coelicolor* cosmid StE8 and ligated into pBluescript KS+ to yield pJS883. The fragment containing the *pobR* gene was excised and sub-cloned into the RPIB bacterial expression vector (44), which contains an N-terminal His₆-tag and Tobacco Etch Virus (TEV) cleavage site. The plasmid was transformed into *E. coli* BL21-Gold (DE3) Competent Cells (Agilent) and subsequently inoculated into 1L cultures of LB containing 50 mg/L kanamycin. The cells were grown at 37°C with vigorous shaking to an OD₆₀₀ between 0.5-0.9, at which point the cells were transferred to 4°C for 1 hour. The cultures were then induced with 0.5 mM IPTG and grown at 18°C for 18 hours.

The PobR-truncated construct consists of amino acid residues 12-511. A 1515-bp DNA fragment encoding this protein construct was PCR-amplified from *S. coelicolor* cosmid StE8 and ligated into pBluescript KS+ to yield pJS884. The fragment was then excised and sub-cloned into the RP1B bacterial expression vector. For PobR N-terminal domain and C-terminal domain, a 783-bp DNA fragment encoding PobR N-terminus (12-267 amino acid) and a 672-bp DNA fragment encoding PobR C-terminus (293-511 amino acid) were PCR-amplified from plasmid pJS883, which carries the *pobR* gene ORF. The resulting PCR products were ligated into pBluescript KS+ to yield pJS885 (N-terminus) and pJS886 (C-terminus). Both fragments were then excised and sub-cloned into the RP1B bacterial expression vector to yield pJS889 and pJS890. The expression conditions were the same as described above for the expression of PobR-FL protein, except that the induced culture of PobR C-terminal domain was grown at 37°C for 4 h.

For the purification, the pellet was resuspended in lysis buffer (50 mM Tris pH 8.0, 500 mM NaCl, 0.1% Triton X-100, 5 mM imidazole, Complete tabs-EDTA free (Roche)). The cells were lysed by high-pressure homogenization using a C3 Emulsiflex high-pressure cell homogenizer (Avestin), and the cell debris was removed by centrifugation (45,000 x *g*/45 min/4°C). The supernatant containing soluble proteins was filtered through a 0.22- μ m membrane (Millipore) and loaded onto a HisTrap HP column (GE Healthcare). His-tagged proteins were eluted with a 5-500 mM imidazole gradient. The fractions containing PobR were identified by SDS-PAGE gel electrophoresis and pooled. The His₆-tag was cleaved through overnight incubation with TEV protease (dialysis buffer: 20 mM Tris pH 8.0, 500 mM NaCl; 4°C). PobR was further purified using a Ni-NTA column to separate cleaved protein from the protease, and cleavage was verified by SDS-PAGE electrophoresis. The final purified protein was run over a size exclusion chromatography (SEC) column (Superdex 75 26/60, GE Healthcare; buffer: 20 mM Tris pH 7.8, 300 mM NaCl, 0.5 mM TCEP, 10% glycerol, 4°C). For PobR bound to PHB (PobR:PHB), the complex was mixed in a 1:5 and 1:10 ratio, incubated on ice for 1 hour, and run over the same SEC

column with identical buffer as apo PobR.

2.2.9 *In vitro* assay of a *para*-hydroxybenzoate hydroxylase activity

The reaction mixture (3 ml) contained 5 μ M protein, 0.5 mM *para*-hydroxybenzoic acid, 0.4 mM NADPH and 5 μ M FAD in 50 mM Tris·Cl (pH 7.4) and was incubated at 30°C. At 0 h, 0.5 h, 1 h and 1.5 h four time points, 500 μ l reaction was taken out and heated at 100°C for 5 min to stop the reaction. Precipitated protein was removed by using Amicon Ultra 10K cutoff centrifugal filters (Merck Millipore). A rapid colorimetric assay described previously (45) was performed to detect protocatechic acid (PCA) formed in the reaction. 1.5 mM FeCl₃ and 50 μ g/ml *p*-toluidine was added to the reaction supernatant followed by 5 min incubation at 30°C. Appearance of a purple color indicated the formation of PCA.

2.2.10 Agarose electrophoretic mobility shift assays

The 118-bp DNA probe was amplified by PCR using *Pfu* polymerase and the *pobA* promoter For and *pobA* promoter Rev primers. Cosmid StE25 was used as template. The PCR program used for amplification was 94°C for 3 min, 33 cycles of 94°C for 30 s, 57°C for 30 s and 72°C for 50 s, following by a final elongation time of 10 min at 72°C. PCR products were purified using Qiagen's PCR Purification Kit following the manufacturer's protocol. DNA concentration was determined using a Nanodrop ND-1000 spectrophotometer. Binding of PobR to DNA fragment was performed in 20 μ l reaction containing 50 mM Tris-HCl pH 6.0, 50 mM KCl, 5 mM MgCl₂, 1 mM EDTA, 10% (vol/vol) glycerol, 0.1 μ g poly(dI-dC), 1 mM dithiothreitol (DTT). Each reaction contained 41 nM DNA and between 10 and 500 nM PobR. The effect of aromatic compounds on the binding of PobR to DNA probe was determined by adding the compounds to the pre-formed PobR-DNA reaction mixture. All compounds were dissolved in dimethyl sulfoxide (DMSO). Reactions were incubated at 30°C for 15 min, loaded onto 2% agarose gel and run in tris-borate EDTA (TBE) buffer at 60 V for 2 h. In experiments where the capacity of aromatic compounds to destabilize the PobR-

DNA was assessed, reactions were incubated at 30°C for further 30 min after adding compounds. Gels were subsequently stained in ethidium bromide for visualization of DNA.

2.2.11 DNase I footprinting

For DNase I footprinting, the 118-bp DNA probe was generated as described above in agarose EMSAs part except that 5' end of primer *pobA* promoter For was labeled with [γ -³²P] ATP using T4 polynucleotide kinase (New England Biolabs) following the manufacturer's protocol. Binding of PobR to DNA fragment was first carried out in 40 μ l reaction with 50 mM Tris-HCl pH 6.0, 50 mM KCl, 5 mM MgCl₂, 1 mM EDTA and 1 mM dithiothreitol (DTT). The reaction contained 584 nM DNA probe and 4.96 μ M PobR. After 20 min incubation at 30°C, 0.2 unit of DNase I (New England Biolabs), 5 μ l of DNase I buffer [100 mM Tris pH 7.6, 25 mM MgCl₂, and 5 mM CaCl₂] and extra 2.5 μ l 100 mM CaCl₂ were added to the reaction (EDTA in binding experiment can inhibit the activity of DNase I, more CaCl₂ was added to competitively chelate the EDTA). The total volume was adjusted to 50 μ l with milli-Q water. The reaction was incubated at 37°C for 5 min and then quenched by the addition of 12 μ l 0.1 M EDTA. When the reaction was complete, DNA sample was then extracted twice with 95 μ l of phenol/chloroform/isoamyl alcohol mixture and precipitated with 400 μ l 96% ethanol, and the resulting pellet was dried under vacuum. Reaction was also performed with the free duplex. As control, the same volume of PobR protein storage buffer was used to replace PobR in protein-DNA binding experiment. The enzyme concentration of DNase I optimized for DNase I footprinting resulted in over-cleavage of the free duplex, so 0.08 unit of enzyme DNase I was used and the incubation time was shortened to 1.5 min.

Standard Maxam-Gilbert A/G and C/T reactions performed on coding strand of the DNA probe was used to create marker. The single stand DNA probe was generated by PCR program as described above using 5'-³²P end-labeled primer *pobA* promoter For and 118-bp unlabeled double-strand DNA probe as template.

After the DNA pellets were dried, the number of counts were determined in a Perkin-Elmer scintillation counter. After dilution with loading buffer [80% formamide, 10 mM EDTA, 1 mg/ml bromophenol blue and 1 mg/ml xylene cyanol] to 5000 CMP/ μ l. 2.1 μ l (for DNase I footprinting) and 2.4 μ l (for A/G and C/T marker) samples were applied to a 10% denaturing polyacrylamide gel (0.4 mm thick, 59 cm long) and electrophoresed at 80 W. (Electrophoresis was performed until the xylene cyanol dye migrated 7 inches.) The gel was dried and used to expose a phosphor-plate for 5 days.

Phosphor-plate was imaged using a Bio-Rad Pharos scanner and the image processing was done using Bio-Rad's Image Lab. First, the gel image was aligned and the lanes was identified. The lane width and position were adjusted, and the position of each band was manually determined. Using the accompanying Maxam-Gilbert sequencing reactions, each band was assigned by nucleotide position and the sequence of PobR binding region was determined.

2.2.12 Cloning of full length *pobR* and truncated mutants for over-expression in *S. coelicolor*

The *pobR* ORF was excised from plasmid pJS883 with NdeI and HindIII and ligated into vector pIJ10257 pre-treated with the same enzyme to yield pJS917. pIJ10257 (46) is a site-specific integrating vector that insert into the Φ BT1 *attB* site of *S. coelicolor* (SCO4848) (43). Gene that is cloned into NdeI-HindIII site is under control of strong constitutive promoter *ermEp** (47). pJS917 was then transformed into ET12567/pUZ8002 and introduced into *S. coelicolor* *pobR* null strain B1656 (yielding strain B1665) via conjugation, as previously described (42). Exconjugants were selected by hygromycin resistance. The same strategy was used to construct *S. coelicolor* strains B1666, B1667, and B1668. Confirmation of gene over-expression was accomplished through RT-PCR analysis of *pobA* (SCO3084). Vector pIJ10257 was also introduced into *S. coelicolor* B1656 separately to yield B1669 strain as the negative control.

2.2.13 Isothermal Titration Calorimetry

Isothermal titration calorimetry (ITC) experiments were performed using a VP-ITC (GE Healthcare) at 25°C. Ligands were dissolved in ITC assay buffer (20 mM Tris pH 7.8, 300 mM NaCl, 0.5 mM TCEP, 10% glycerol), and both protein and ligand were degassed under vacuum. Ligands (*para*-hydroxybenzoic acid: 600 μ M, and 2,4-dihydroxybenzoic acid: 1.2 mM) were titrated (10 μ L injections every 300 s, 35 times) into the cell containing 60 μ M monomeric PobR under constant stirring (307 rpm). Titration of the ligand into buffer resulted in a negligible heat of dilution for each ligand. Data were analyzed using NITPIC (48) and SEDPHAT (49).

2.2.14 Thermofluor Assay

A series of protein thermal denaturation assays were performed with a final concentration of 3 μ M dimeric protein, 300 μ M ligand, and 5x SYPRO Orange (Invitrogen). As control, buffer was added instead of ligand. Samples were aliquoted in a 96-well PCR plate (Applied Biosystems) and sealed with optical adhesive film (Applied Biosystems) to prevent evaporation. The temperature was gradually increased from 25°C to 95°C using a 7900HT Fast Real-Time PCR System (Agilent). A charge-coupled device detector monitored changes in the intensity of the SYPRO orange fluorescence. Data were analyzed and T_m values computed using the DSF analysis calculation software (50,51).

2.2.15 Crystallization, data collection, and processing

Apo PobR was concentrated to 6.5869 mg/ml, mixed with PHB in 1:10 ratio, and immediately used to set-up crystallization trials. Crystals were obtained in 7 days at 4°C in 12% PEG MME 2000, 0.06 M MES, 0.04 M MES Na pH 5.8, 0.0175 M Sulfate NH_4 using sitting-drop vapor diffusion. Crystals were cryoprotected in precipitant containing 20% glycerol and frozen in liquid nitrogen for data collection. Data for PobR:PHB was collected on a Rigaku FR-E+ SuperBright generator using a Saturn 944G CCD detector. Diffraction data were processed to 2.4 Å with HKL3000 (52).

The *Rhodococcus jostii* RHA1 RHA06195 (PDB: 2IA2) was identified by the Fold and Function Assignment Server (-71.2 score, 46% sequence identity) as a suitable initial molecular replacement (MR) model. Chainsaw was used to create a poly-alanine model of domain A of 2IA2 which was subsequently used for MR (McCoy et al., 2007). The iterative model building and refinement is still in process by our collaborator.

Table 2.2 Strains employed in this study

Strain	Genotype	Reference/Source
<i>S. coelicolor</i> A3(2)		
M600	Prototroph SCP1-, SCP2-	(40)
B1653	M600 Δ SCO3084:: <i>apr</i>	This Study
B1654	M600 Δ SCO1308:: <i>spec</i>	This Study
B1655	M600 Δ SCO0266:: <i>apr</i>	This Study
B1656	M600 Δ SCO3209:: <i>apr</i>	This Study
B1657	M600 Δ SCO3084:: <i>apr</i> - pJS880	This Study
B1658	M600 Δ SCO3209:: <i>apr</i> - pJS881	This Study
B1665	M600 Δ SCO3084:: <i>apr</i> - pJS917	This Study
B1666	M600 Δ SCO3084:: <i>apr</i> - pJS918	This Study
B1667	M600 Δ SCO3084:: <i>apr</i> - pJS919	This Study
B1668	M600 Δ SCO3084:: <i>apr</i> - pJS920	This Study
B1669	M600 Δ SCO3084:: <i>apr</i> - pIJ10257	This Study
<i>E. coli</i>		
DH5 α	F- ϕ lacZ Δ M15 Δ (lacZYA-argF)U169 <i>recA1 endA1 hsdR17</i> (r_k^- , m_k^-) <i>phoA supE44 thi-1 gyrA96 relA1 λ^-</i>	Invitrogen
BW25113	Δ (<i>araD-araB</i>)567 Δ lacZ4787(::rrnB-4) <i>lacI_p</i> -4000(lacI ^Q) λ^- <i>rpoS369</i> (Am) <i>rph-1</i> Δ (<i>rhaD-rhaB</i>)568 <i>hsdR514</i>	(53)
ET12567	<i>dam dcm hsdS cat tet</i>	(41)
BL21-Gold (DE3)	<i>E. coli</i> B F ⁻ <i>ompT hsdS</i> (r_B^- m_B^-) <i>dcm</i> ⁺ Tet ^r <i>gal λ</i> (DE3) <i>endA Hte</i>	Agilent

Table 2.3 Plasmids employed in this study

Plasmid	Description	Vector Backbone	Reference/Source
pIJ790	[<i>oriR101</i>], [<i>repA101</i> (<i>ts</i>)], <i>araBp-gam-bet-exo</i> , Chl ^r		(41)
pUZ8002	RP4 Derivative OriT ⁻ , Kan ^r		(41)
pBluescript KS+	pUC <i>ori</i> , MCS, Amp ^r		Agilent Technologies (Stratagene)
pGem-T Easy	pUC-derived, <i>lacZ</i> , Amp ^r		Promega
pMS81	<i>oriT</i> , Φ BT1 <i>attB-int</i> , Hyg ^r	pSET152	(43)
pETRP1B	pBR322 <i>ori</i> , fl origin, <i>lacI</i> , MCS, Kan ^r		(44)

pET-28a(+)	pBR322 <i>ori</i> , <i>f1</i> origin, <i>lacI</i> , MCS, Kan ^r		Novagen
pIJ10257	<i>oriT</i> , ΦBT1 <i>attB-int</i> , Hyg ^r , <i>ermEp*</i>	pMS81	(46)
pJS878	1731bp BstXI and NcoI fragment from cosmid StE25 containing <i>pobA</i> (SCO3084) with 407 upstream bp, Amp ^r	pBluescript KS+	This Study
pJS879	1981bp BlnI and BspEI fragment from cosmid StE8 containing <i>pobR</i> (SCO3209) with 367 upstream bp, Amp ^r	pBluescript KS+	This Study
pJS880	1731bp BstXI and NcoI fragment from cosmid StE25 containing <i>pobA</i> (SCO3084) with 407 upstream bp, Hyg ^r	pMS81	This Study
pJS881	1981bp BlnI and BspEI fragment from cosmid StE8 containing <i>pobR</i> (SCO3209) with 367 upstream bp, Hyg ^r	pMS81	This Study
pJS882	<i>pobA</i> 5'-RACE product of Primer AAP(kit) and SCO3084 GSP2	pGem-T Easy	This Study
pJS883	<i>pobR</i> (SCO3209) ORF with engineered NdeI and HindIII sites at each end, Amp ^r	pBluescript KS+	This Study
pJS884	1515bp fragment encoding truncated PobR with engineered NdeI and HindIII site at each end, Amp ^r	pBluescript KS+	This Study
pJS885	783bp fragment encoding PobR N-terminus with engineered NdeI and HindIII site at each end, Amp ^r	pBluescript KS+	This Study
pJS886	672bp fragment encoding PobR C-terminus with engineered NdeI and HindIII site at each end, Amp ^r	pBluescript KS+	This Study
pJS887	<i>pobR</i> (SCO3209) ORF with engineered NdeI and HindIII sites at each end, Kan ^r	pETRP1B	This Study
pJS888	1515bp fragment encoding truncated PobR with engineered NdeI and HindIII site at each end, Kan ^r	pETRP1B	This Study
pJS889	783bp fragment encoding PobR N-terminus with engineered NdeI and HindIII site at each end, Kan ^r	pETRP1B	This Study
pJS890	672bp fragment encoding PobR C-terminus with engineered NdeI and HindIII site at each end, Kan ^r	pETRP1B	This Study
pJS917	<i>ermEp*</i> - <i>pobR</i> (SCO3209) ORF, Hyg ^r	pIJ10257	This Study
pJS918	<i>ermEp*</i> - 1515bp fragment encoding truncated PobR 12-511, Hyg ^r	pIJ10257	This Study
pJS919	<i>ermEp*</i> - 783bp fragment encoding PobR N-terminus, Hyg ^r	pIJ10257	This Study
pJS920	<i>ermEp*</i> - 672bp fragment encoding PobR C-terminus, Hyg ^r	pIJ10257	This Study
pJS891	<i>pobA</i> (SCO3084) ORF with engineered NdeI and HindIII sites at each end, Amp ^r	pBluescript KS+	This Study
pJS892	<i>pobR</i> (SCO3209) ORF with engineered NdeI and HindIII sites at each end, Kan ^r	pET-28a(+)	This Study

Table 2.4 Oligonucleotides employed in this study

Primer Name	Application/Function	Sequence
SCO3084 KO For	PCR-Targeting/Disruption of SCO3084	5'-CTCACGGCCGGGGTCGCACGGGCCGCCGG GCGGCGCTCAACTAGTTGTAGGCTGGAGCTG CTTC-3'
SCO3084 KO Rev	PCR-Targeting/Disruption of SCO3084	5'-GAGCAAGGCAACGACGCCGATCGAAGGG AACGACCGATGACTAGTATTCCGGGGATCCG TCGACC-3'
SCO1308 KO For	PCR-Targeting/Disruption of SCO1308	5'-TTCCGGTGGCCCGTGCGCCGCCGACATAA TGTCCGCGTACTAGTATTCCGGGGATCCGT CGACC-3'
SCO1308 KO Rev	PCR-Targeting/Disruption of SCO1308	5'-CGACTGAGCGACCTGCCGTCCCGGCTGCC CGGGCGTCAACTAGTTGTAGGCTGGAGCTG CTTC-3'
SCO0266 KO For	PCR-Targeting/Disruption of SCO0266	5'-GGCACGCAGCGCGCCGACTTCCACGTCCT TGCCGTCGTCATTCCGGGGATCCGTCGACC- 3'
SCO0266 KO Rev	PCR-Targeting/Disruption of SCO0266	5'-GGTCGGAACCCCGGGGCCGGTGCGCCGCCG CCTGCCACGCTGTAGGCTGGAGCTGCTTC-3'
SCO3209 KO For	PCR-Targeting/Disruption of SCO3209	5'-ATGCCCGCCACACCACTCACGACGCCGC CCGGACGAGATTCCGGGGATCCGTCGACC- 3'
SCO3209 KO Rev	PCR-Targeting/Disruption of SCO3209	5'-CGTCGCCCCCGGGCCGCTGTCCGCCCCCG CGGTGGGTATGTAGGCTGGAGCTGCTTC-3'
SCO3084 KO Det For	Verification of SCO3084 Disruption	5'-CTCGAGCAAGGCAACGAC-3'
SCO3084 KO Det Rev	Verification of SCO3084 Disruption	5'-TACGCCGATCAGTGATTC-3'
SCO1308 KO Det For	Verification of SCO1308 Disruption	5'-GCGCCGCCGACATAATG-3'
SCO1308 KO Det Rev	Verification of SCO1308 Disruption	5'-GTCGACGAGGCGCTGAAC-3'
SCO0266 KO Det For	Verification of SCO0266 Disruption	5'-TGGCTCTTCTCGTTCGATG-3'
SCO0266 KO Det Rev	Verification of SCO0266 Disruption	5'-CCTCGGGATGACCACCTG-3'
SCO3209 KO Det For	Verification of SCO3209 Disruption	5'-AAGGGCGTCACGATCAAG-3'
SCO3209 KO Det Rev	Verification of SCO3209 Disruption	5'-GGACTCCGTACCCTCGTG-3'
SCO3084 RT For	Detection of SCO3084 transcript	5'-ACTCGGTCGTCCTGGAGAG-3'
SCO3084 RT Rev	Detection of SCO3084 transcript	5'-GTAGACCATGACGGATCTG-3'
<i>hrdB</i> RT For	Detection of the <i>hrdB</i> (SCO5820) transcript	5'-CTCGAGGAAGAGGGTGTGAC-3'
<i>hrdB</i> RT Rev	Detection of the <i>hrdB</i> (SCO5820) transcript	5'-TGCCGATCTGCTTGAGGTAG-3'
SCO3084 For	5' RACE, <i>pobA</i> (SCO3084)	5'-CTCGAGCAAGGCAACGAC -3'
SCO3084 GSP1	5' RACE, <i>pobA</i> (SCO3084)	5'-TCTGGGCGTAGACCATGAC -3'
SCO3084 GSP2	5' RACE, <i>pobA</i> (SCO3084)	5'-GTGCCCTGCTCCAGGATC -3'
<i>pobA</i> promoter For	Amplification of probe containing PobR binding sites in <i>S. coelicolor</i>	5'-GCGCAGTTGATGGTGAAG-3'
<i>pobA</i> promoter Rev	Amplification of probe containing PobR binding sites in <i>S. coelicolor</i>	5'-GCATCGGTGCTTCCCTTC-3'
PobR-FL Clone For	Cloning of the <i>pobR</i> ORF (SCO3209)	5'- <u>CATATG</u> CCCCGCCACACCAC-3' ^a

PobR-FL Clone Rev	Cloning of the <i>pobR</i> ORF (SCO3209)	5'- <u>AAGCTT</u> CGAGTCCGGACCCGTC-3' ^b
Truncated PobR Clone For	Cloning of sequence corresponding to truncated PobR protein	5'-CATATGGACGAGGCGGTGCGGCCGCTG-3'
Truncated PobR Clone Rev	Cloning of sequence corresponding to truncated PobR protein	5'- <u>AAGCTT</u> <u>TC</u> ACTCGGTCTCGACCAGGTC-3' ^{b,c}
PobR-N Clone For	Cloning of sequence corresponding to PobR N-terminus	5'-CATATGGACGAGGCGGTGCGGC-3' ^a
PobR-N Clone Rev	Cloning of sequence corresponding to PobR N-terminus	5'- <u>AAGCTT</u> <u>TC</u> AGTCGTCCTCCATCGCCG-3' ^{b,c}
PobR-C Clone For	Cloning of sequence corresponding to PobR C-terminus	5'-CATATGGAAGTGGGCCGGGAGTTC-3' ^a
PobR-C Clone Rev	Cloning of sequence corresponding to PobR C-terminus	5'- <u>AAGCTT</u> <u>TC</u> ACTCGGTCTCGACCAGGTC-3' ^{b,c}
PobA Clone For	Cloning of the <i>pobA</i> ORF (SCO3084)	5'-CATATGCGCACCCACCGTCG-3' ^a
PobA Clone Rev	Cloning of the <i>pobA</i> ORF (SCO3084)	5'- <u>AAGCTT</u> ACGCTACGCCGATCAGTG-3' ^b

- The engineered NdeI site is underlined.
- The engineered HindIII site is underlined.
- The engineered stop codon is double underlined.

2.3 Results and Discussion

2.3.1 *S. coelicolor* has a *pobA* gene (SCO3084) encoding *p*-hydroxybenzoate hydroxylase

Bioinformatic analyses of the *S. coelicolor* genome sequence revealed two *pobA* genes (SCO1308 and SCO3084) encoding enzymes that are homologous to *p*-hydroxybenzoate hydroxylase from *Pseudomonas fluorescens* (54,55) (<http://streptomyces.org.uk>). To test the predicted functions of these two genes, PCR-targeting procedure was used to construct a SCO3084 null mutant (*S. coelicolor* B1653 Δ SCO3084::*apr*) and a SCO1308 null mutant (*S. coelicolor* B1654 Δ SCO1308::*spec*). The genes of interest in these strains were replaced by markers for resistance to apramycin and spectinomycin, respectively. The functions of the genes of interest were inferred from the capacities of the null strains to grow on media having PHB as the sole carbon source. A strain lacking a functional *para*-hydroxybenzoate hydroxylase would not be able to grow on media having PHB as the sole carbon source. We found that SCO3084 null mutant lost its ability growing on Gelzan plates having *p*-hydroxybenzoate (PHB) as the sole carbon source; whereas it can still

grow on plates having protocatechuate (PCA), product of PHB hydroxylation, as the sole carbon source. In contrast, there was no significant difference of growth between wild-type *S. coelicolor* and SCO1308 null strain on media supplemented with either PHB or PCA (Fig. 2.1). These observations indicate that only SCO3084 encodes a functional *para*-hydroxybenzoate hydroxylase. To confirm that SCO3084 gene was necessary and sufficient for metabolism of PHB, a complementation strain was constructed. Provision of SCO3084 under the control of its native promoter to the null strain restored PHB consumption to the SCO3084 null strain (Fig. 2.2).

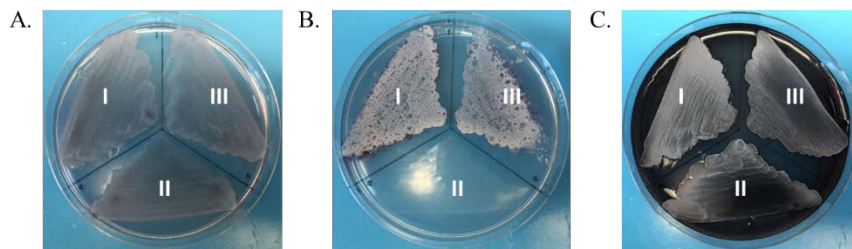


Figure 2.1 Phenotypes of (I) wild-type *S. coelicolor* M600, (II) *S. coelicolor* B1653 (Δ SCO3084::*apr*), and (III) *S. coelicolor* B1654 (Δ SCO1308::*apr*). The strains were grown for 10 days at 30°C on MM media supplemented with different carbon sources. From left to right, (A) 0.5% (w/v) glucose, (B) 20 mM PHB, and (C) 20 mM PCA. Deletion of SCO3084 in *S. coelicolor* abolished its ability of growing on plates using PHB as the sole carbon source while disruption of SCO1308 did not affect growth on PHB.

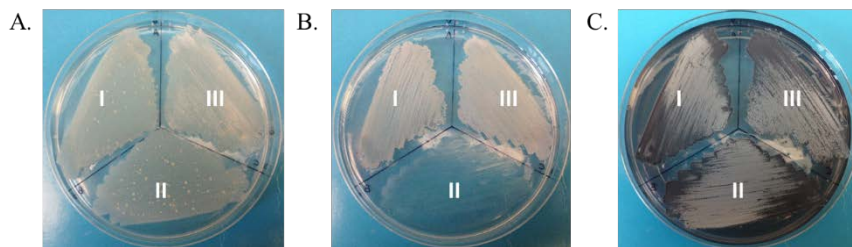


Figure 2.2 Phenotypes of (I) wild-type *S. coelicolor* M600, (II) *S. coelicolor* B1653 (Δ SCO3084::*apr*), and (III) *S. coelicolor* B1657 (Δ SCO1308::*apr* + pJS880). The strains were grown for 10 days at 30°C on MM media supplemented with different carbon sources. From left to right, a. 0.5% (w/v) glucose, b. 20 mM PHB, and c. 20 mM PCA.

In addition to the aforementioned reverse genetic analyses, a colorimetric biochemical assay of the SCO3084 gene product was performed. For this experiment, we cloned the *pobA* gene (SCO3084) with N-terminal His₆ tag, over-expressed it in *E. coli*, and purified the gene product by metal-affinity chromatography (Fig. 2.3A). PHB was incubated with PobA, NADPH and FAD in Tris·Cl buffer. Formation of PCA in this reaction was successfully detected colorimetrically as described previously (15) (Fig. 2.3B). Furthermore, the production of PCA was also detected by ESI-MS (data not shown). Both genetic and biochemical results indicate that the SCO3084 gene encodes *p*-hydroxybenzoate hydroxylase.

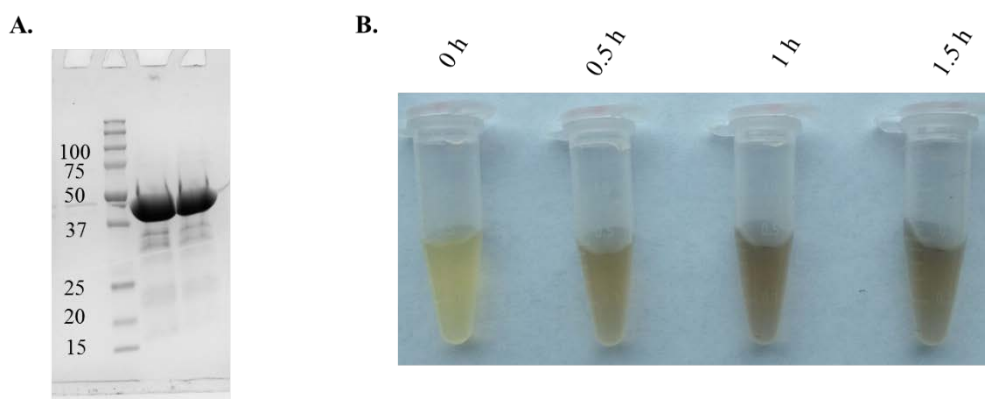


Figure 2.3 Purification and biochemical characterization of PobA. (A) SDA-PAGE analysis of PobA purification. The theoretical molecular weight (MW) of PobA with His-tag is 45,406.9 Da. (B) *In vitro* assay of a *para*-hydroxybenzoate hydroxylase activity. For this assay, 5 μ M purified PobA was incubated with PHB, NADPH and FAD in Tris·Cl (pH 7.4) buffer at 30°C. At 0 h, 0.5 h, 1 h and 1.5 h four time points, the reaction was quenched by heating at 100°C for 5 min. 1.5 mM FeCl₃ and 50 μ g/ml *p*-toluidine were then added to the reaction supernatant followed by 5 min incubation at 30°C. Appearance of a purple color indicated the formation of PCA.

2.3.2 Transcription of the gene *pobA* is induced by *para*-hydroxybenzoate (substrate of PobA) and its analogs

It was reported that *para*-hydroxybenzoate induced the transcription of *pobA* in *Acinetobacter calcoaceticus* (18,19). By analogy, we hypothesized that the transcription of *pobA* in *S. coelicolor* might be induced by *para*-hydroxybenzoate. Semi-quantitative reverse transcription polymerase chain reaction (RT-PCR) was utilized to determine if transcription of the gene *pobA* was induced by PHB. Indeed, there was robust transcription of *pobA* when *S. coelicolor* was grown in liquid media supplemented with *para*-hydroxybenzoate, whereas there was only a basal level transcription of *pobA* when the bacterium was grown in media lacking PHB (Fig. 2.4). Interestingly, we found that *para*-hydroxybenzoate analogs, including protocatechuate (PCA, the product of PobA catalyzed hydroxylation) and 2,4-dihydroxybenzoate (2,4-DHB) could also induce the transcription of *pobA*. These observations differ from studies in *A. calcoaceticus* wherein only PHB induces the transcription of *pobA* (18).

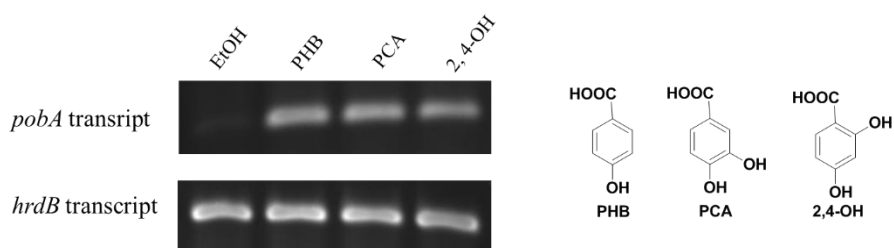


Figure 2.4 Changes in the transcription of *pobA* in media supplemented with PHB, PCA, and 2,4-DHB. These molecules are the substrate of PobA, product of the PobA-catalyzed reaction, and an analog of PHB, respectively. All the compounds were dissolved in EtOH. RT-PCR was performed to detect the transcription of *pobA*. The *hrdB* control transcript was detected in all of the RNA samples. The *hrdB* gene, which encodes a vegetative sigma factor, is constitutively transcribed (40).

2.3.3 Transcription of the gene *pobA* was negatively regulated by SCO3209 (*pobR*)

Based on the regulated transcription of *pobA*, we sought to identify the cognate transcriptional regulator. In *Acinetobacter calcoaceticus*, expression of *pobA* was regulated by a transcriptional activator PobR, encoded by the gene *pobR*, which was divergently transcribed (18,19). Accordingly, we sought *pobR* homologs in the *S. coelicolor* genome. Interestingly, we found no homologs of *pobR* in close proximity to SCO3084. Thus, BLASTp (56) was performed to identify gene candidates encoding regulators that are homologous to PobR. Analyses revealed that the best candidate is a putative transcription regulator encoded by gene SCO3209 and it is 35% identical in amino acid sequence to PobR in *A. calcoaceticus*. Additionally, analysis of annotations of the *S. coelicolor* genome also showed that SCO0266 encodes a putative regulator, which is 30.3% identical to PobR from *Pseudomonas sp.* Strain HR199. To determine which one regulates the transcription of SCO3084 and how it functions to regulate *pobA* transcription, PCR-targeted gene replacement was applied to construct a SCO3209 null strain (*S. coelicolor* B1656 Δ SCO3209::*apr*) and a SCO0266 null strain (*S. coelicolor* B1655 Δ SCO0266::*apr*). We found that the gene *pobA* was constitutively transcribed in the SCO3209 null mutant in the absence and presence of *para*-hydroxybenzoate. In contrast, deletion of gene SCO0266 did not affect the transcription of *pobA* (Fig. 2.5A). The phenotype of the SCO3209 null mutant was suppressed by complementation. The complementation strain *S. coelicolor* B1658 exhibited PHB-dependent transcription of *pobA* (Fig. 2.5B). These observations indicate that gene SCO3209 encodes a repressor of *pobA* gene transcription. Based on the apparent function of the SCO3209 gene, we decided to name its product PobR.

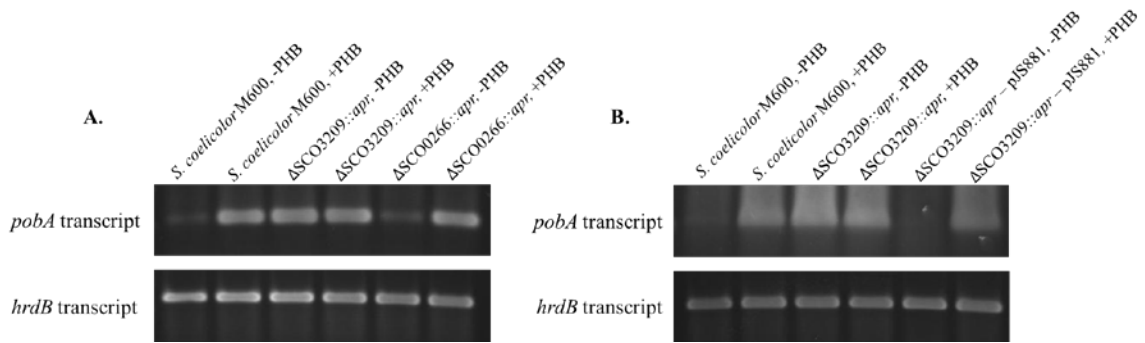


Figure 2.5 Transcription profile of the *pobA* gene expression in wild-type *S. coelicolor* and mutant strains lacking either SCO0266 or SCO3209. RT-PCR was used to detect the transcription of *pobA*. The *hrdB* transcript (positive control) was detected in all of the RNA samples. The *hrdB* gene, which encodes a vegetative sigma factor, is constitutively transcribed. The *pobA* gene was transcribed in wild-type *S. coelicolor* M600 culture grown in NMMP media treated with PHB. (A) *pobA* was constitutively transcribed in the SCO3209 null mutant in the absence and presence of PHB, while deletion of gene SCO0266 didn't affect the transcription of *pobA*. (B) The transcription defect in SCO3209 null strain was suppressed in the complementation strain, *S. coelicolor* B1658 Δ SCO3209::apr + pJS881.

2.3.4 Bioinformatic analyses revealed that PobR is a peculiar IclR family protein

According to the gene annotation in StrepDB, SCO3209 encodes a putative IclR family transcriptional regulator. However, the gene product was almost the twice the size of typical IclR-type regulator (*i.e.*, 55 kDa). In fact, SCO3209 appears to encode a fusion of two IclR family members. HMMER (57) and protein BLAST (56) predicted that both N- and C-terminal domains of the PobR protein are like an intact IclR protein (Fig. 2.6). The two domains of PobR were predicted to be structurally similar by FFAS (58) and connected by disordered linker based on IUPred (59) prediction. However, in comparison with other known IclR family regulators, protein sequence alignment indicated that the PobR N-terminal domain contains an extra insert, while the PobR C-terminal domain is missing this insert plus another long stretch of sequence. In many respects, PobR is a peculiar IclR-type regulator. Furthermore, protein BLAST (Table 2.5) against

streptomyete genomes in the IMG (Integrated Microbial Genomes) database showed that this peculiar PobR is highly conserved (Fig. 2.7).

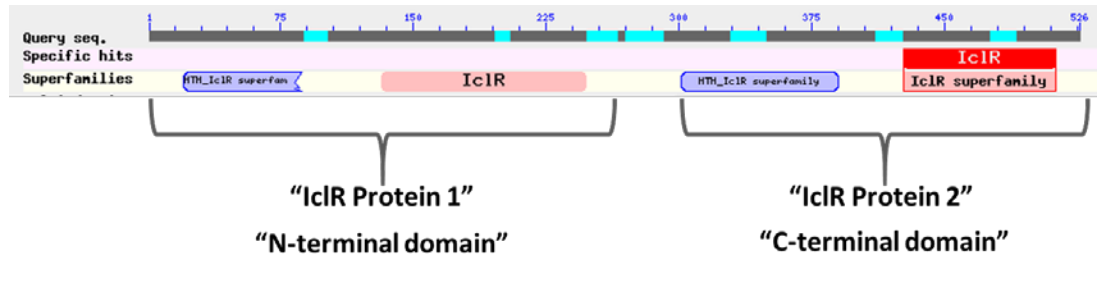


Figure 2.6 Predicted domains organization of PobR protein. Both N-terminal domain and C-terminal domain of PobR are homolog to an intact IcIR family protein. The monomeric PobR protein is proposed to share similar structure with currently known dimeric IcIR family protein.

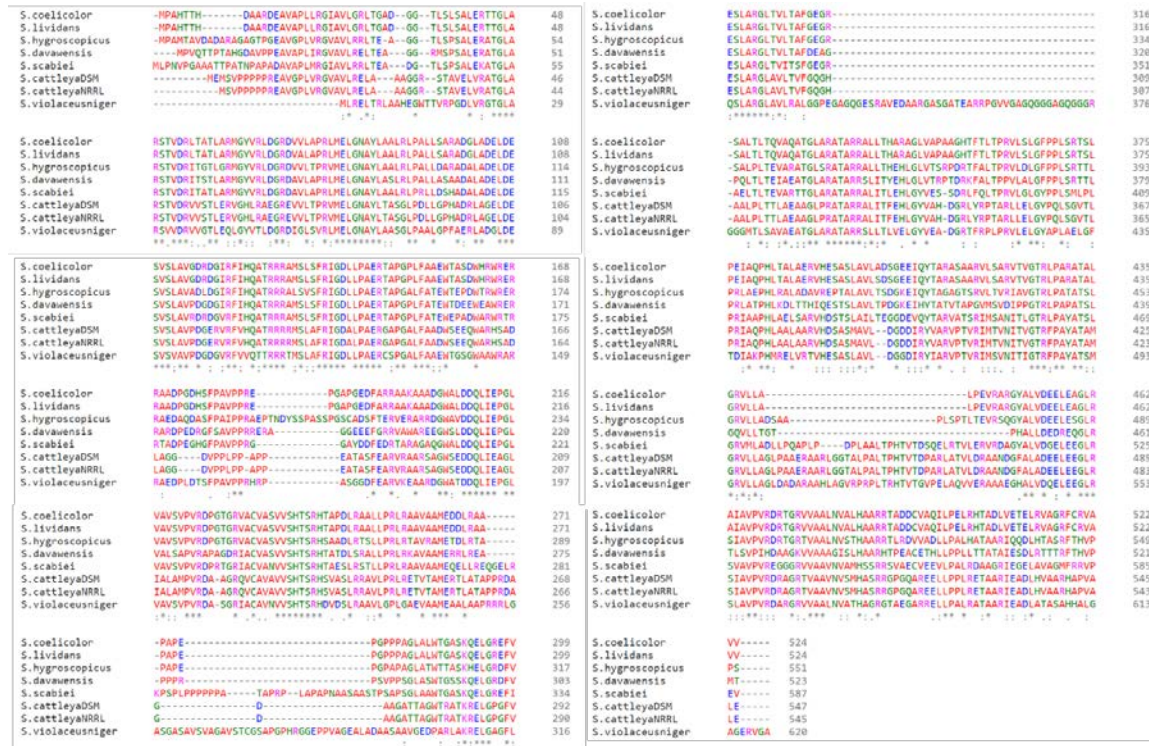


Figure 2.7 Full amino acid sequence alignment of PobR identified in the *Streptomyces* genus through the protein BLAST bioinformatics analysis. Conserved amino acids are highlighted by color and indicated with star.

Table 2.5 PobR homologs in the other *Streptomyces* genus.

No.	<i>Streptomyces</i> genus	Predicted identity	Protein AA length	E-value	Identity (%)	Positives (%)	Gaps (%)
1	<i>S. lividans</i> TK24	transcriptional regulator, IclR family	524	0	99	99	0
2	<i>S. hygroscopicus jinggangensis</i> TL01	transcriptional regulator, IclR family	551	7e-180	72	78	5
3	<i>S. avermitilis</i> MA-4680	transcriptional regulator, IclR family	572	3e-172	68	75	10
4	<i>S. davawensis</i> JCM 4913	transcriptional regulator, IclR family	523	4e-154	66	75	2
5	<i>S. scabiei</i> 87.22	transcriptional regulator, IclR family	587	3e-108	82	87	1
6	<i>S. cattleya</i> DSM 46488	transcriptional regulator, IclR family	547	6e-107	53	64	8
7	<i>S. cattleya</i> NRRL 8057	transcriptional regulator, IclR family	545	9e-107	53	64	8
8	<i>S. violaceusniger</i> Tu 4113	transcriptional regulator, IclR family	620	9e-75	70	77	0

2.3.5 PobR binds to a palindromic sequence in the *pobA* promoter region

Our genetic analyses indicated that SCO3209 encodes a transcriptional repressor that is functionally affected by PHB. The typical mode of repression mediated by IclR regulator is that it binds to a specific DNA sequence (usually a palindrome) overlapping the RNA polymerase binding site, such that occupancy of the site blocks the binding of the RNA polymerase. By analogy to this mode, we hypothesized that the molecular mechanism of transcriptional regulation involved binding of PobR to a palindromic sequence close to transcription start site of the *pobA* gene. To test our hypothesis, we first used 5' rapid amplification of cDNA ends (5'-RACE) experiments to map the transcription start site of the *pobA* gene. The transcription of *pobA* was mapped to two

sites, 109 bp and 110 bp upstream of the *pobA* translation start site (-109 and -110) (Fig. 2.8A). The putative promoter sequences (-10 and -35) have also been assigned on the basis of conformity to a consensus sequence among streptomycete promoters (60). Next, electrophoretic mobility shift assays (EMSAs) were used to assess the binding of PobR to *pobA* promoter region. A 118-bp PCR-amplified fragment spanning the entire region between the predicted promoter and the *pobA* open reading frame initiation codon (ATG) was used as DNA probe in EMSAs. As shown in Figure 2.8B, titration with increasing quantities of purified PobR resulted in a shift in the migration of the DNA probe. In order to identify the precise DNA sequence of the binding region, we performed DNase I footprinting analysis. The same 118-bp DNA segment labeled at 5' end with [γ -³²P] ATP was treated with DNase I in the presence and absence of purified PobR protein. As shown in Figure 2.8C, a single region adjacent to the putative transcriptional promoter of *pobA* was protected from nucleolytic digestion. This 30 bp protected site contains perfect inverted repeat sequences which may be involved in sequence-specific recognition by PobR. As there was only one shifted protein-DNA complex in the EMSAs and only one single inverted sequence was protected by PobR, we concluded that there is only one operator existing for protein binding in the *pobA* upstream region. Further sequence alignments (ClustalX2) (61) of the *pobA* upstream regions (up to -150 bp nucleotide) and Gibbs motif sampler analysis (62) revealed that PobR binding site sequences are conserved in multiple *Streptomyces* species (Fig. 2.8D). Interestingly, besides the inverted repeat sequences, the protected sequence at 3' end (CGAAC) is also highly conserved.

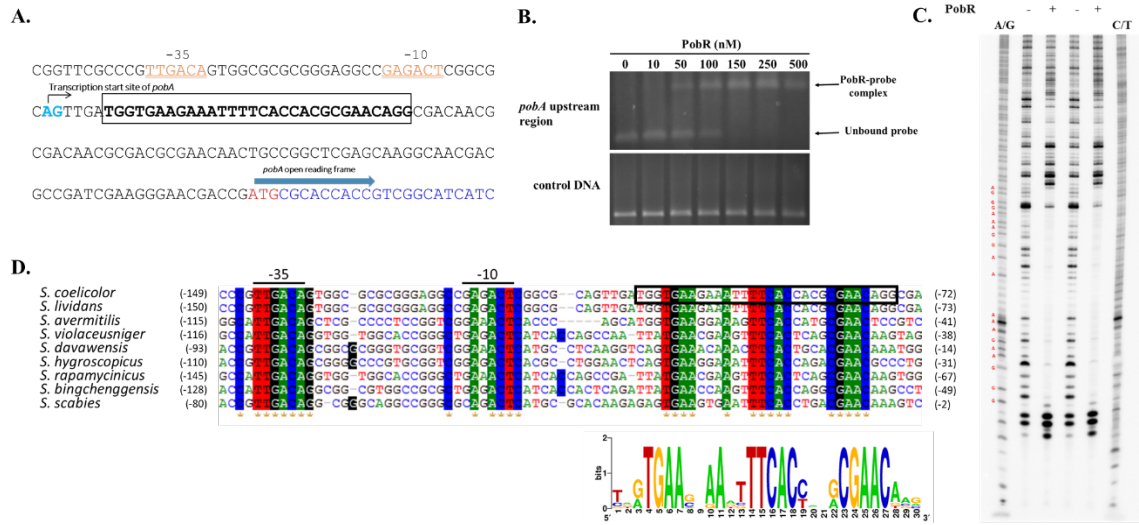


Figure 2.8 PobR recognizes an operator site in the *pobA* upstream region. **(A)** The overall view of *pobA* upstream region coding strand sequence, which shows the transcriptional start site of *pobA* (highlighted in turquoise), putative -35 and -10 *pobA* promoter elements (underlined and highlighted in orange) and PobR regulator binding site sequence (in bold and boxed). **(B)** Agarose EMSA experiments using a PCR-amplified DNA fragment spanning the entire region between predicted promoter and initiation codon (ATG) and increasing concentration of PobR. In comparison with the unbound DNA probe in the electrophoresis, retarded migration of the DNA probe was observed on addition of PobR. Both PobR-probe complex and unbound probe were indicated with a single arrow. The control DNA is 486-bp fragment amplified from *hrdB* ORF and does not contain PobR binding operator. **(C)** DNase I footprinting analysis of PobR binding site in the *pobA* upstream region. The 5'-³²P end-labeled 118-bp DNA fragment was digested with DNase I in the absence and presence of purified PobR protein. The reactions were performed in duplicate. In comparison with the digest in the absence of PobR, a clear area which indicated the protected region was observed when DNA was incubated with PobR before nucleolytic digestion. Lanes C/T and A/G represent as ladders and were generated from Maxam-Gilbert reaction. **(D)** Full sequence alignment of *pobA* upstream region. Putative -35 and -10 *pobA* promoter elements and PobR binding site sequence (boxed) identified in DNase I footprinting reaction are conserved in several members of the *Streptomyces* genus. The conserved PobR binding site was also identified using the Gibbs Motif Sampler (62). Sequence logo images (63) depict the sequence conservation of PobR binding site.

2.3.6 PHB and its analogs attenuate the binding of PobR to DNA and they bind differently to PobR

IcIR family repressors typically regulate transcription through ligand-mediated attenuation of DNA binding. As discussed above, transcription of the gene *pobA* was induced by PHB and its analogs (PCA and 2, 4-DHB). We predicted that these aromatic compounds would reduce the affinity of PobR for the operator of *pobA*. We found that PobR dissociated from DNA having the *pobA* promoter region in EMSAs. Specifically, pre-formed PobR-DNA complexes were destabilized upon treatment with increasing quantities of aromatic compounds (Fig. 2.9A). Thus, these small molecules are all considered ligands for PobR.

Table 2.6 Co-factor effects on PobR thermal stability*

Cofactor	T _m (°C)	ΔT _m (°C)
PHB	50.27 ± 0.43	2.01
PCA	47.66 ± 0.58	-0.60
2,4-DHB	49.30 ± 0.57	1.04
EtOH	48.26 ± 0.30	--
Water	48.26 ± 0.18	--

*T_m shift data: Values represent means ± SD of four independent experiments

Next, a thermofluor assay was used to profile the thermal stability of PobR in the presence of various aromatic ligands (PHB, PCA, and 2,4-DHB). All the compounds were dissolved in EtOH first and then diluted in protein buffer. It was confirmed that incorporation of EtOH in the system did not affect the melting pointing of PobR. The compound/ligand that induced a significant change in the melting temperature (2.01°C difference) was PHB, which has a T_m of 50.27 ± 0.43°C compared to 48.26 ± 0.30°C for the control buffer (EtOH). 2,4-DHB resulted in a relatively moderate shift with a ΔT_m = 1.04°C, which is indicative of weaker binding. Curiously, treatment

with PCA caused a decrease in the T_m (*i.e.*, $\Delta T_m = -0.6^\circ\text{C}$), which suggested that it destabilizes PobR (Fig. 2.9B, Table 2.6).

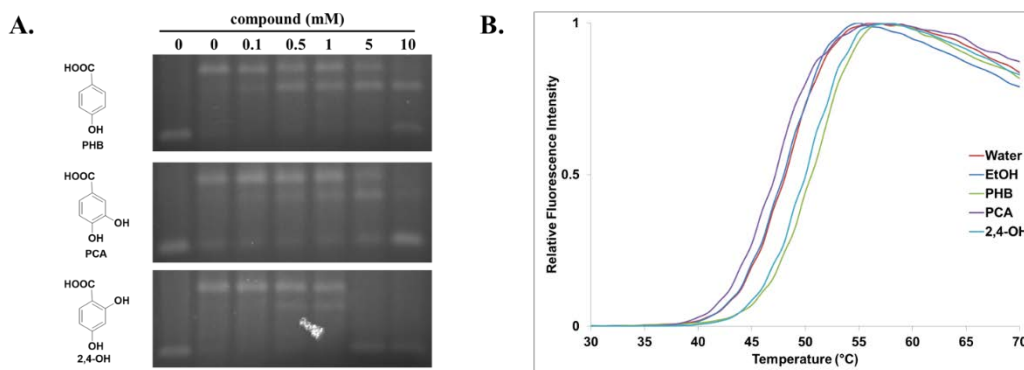


Figure 2.9 (A) Aromatic compounds attenuate the DNA binding of PobR. PobR-probe complex and increasing concentrations of aromatic compounds (PHB, PCA and 2,4-DHB) were used in agarose EMSA experiments. The dissociation of PobR from the DNA probe is observed on addition of increasing amounts of the aromatic compounds. Unbound DNA was used as control in the first lane and the structure of each aromatic compound is shown on the left. (B) Thermal denaturation curve of PobR in the absence and presence of ligands using the Thermofluor assay. At the beginning of the assay, the protein is folded and the dye is occluded from the hydrophobic residues located in the core. Upon heating, the protein unfolds in a cooperative manner allowing the dye access to the internal hydrophobic residues. Upon binding, fluorescence increases as indicated by the Y-axis. When the protein is completely unfolded, maximum fluorescence is achieved. However, as the protein aggregates, these interaction replace the dye-protein interaction and fluorescence decreases. The melting temperature T_m is reflected by the point where 50% protein is unfolded. All the tested aromatic compounds were first dissolved in EtOH and then diluted in protein buffer (20 mM Tris pH 7.5, 300 mM NaCl, 0.5 mM TCEP, 10% glycerol, RT). The T_m increased when PobR binds to PHB and 2,4-DHB, while the T_m decreased when PobR binds to PCA.

Isothermal titration calorimetry (ITC) was further used to determine the PobR-PHB stoichiometry and quantify the binding affinity of PobR for PHB (Fig. 2.10). The stoichiometry of the complex is 1:1 (monomer:ligand), and the calculated dissociation constant K_d is $5.75 \pm 0.49 \mu\text{M}$, which is

comparable with other observed K_d for IclR family protein with corresponding ligand (31). We also performed ITC to determine the affinities of 2,4-DHB and PCA for Pobr. The binding of 2,4-DHB to Pobr was found to be weaker compared to PHB with a K_d of $12.21 \pm 0.70 \mu\text{M}$. The binding of PCA to Pobr was very weak and the signal did not reach saturation in our ITC experiments. The preference of Pobr for PHB is consistent with the fact that the ligand is catabolized by the product of gene (*i.e.*, *pobA*) whose expression is regulated by Pobr.

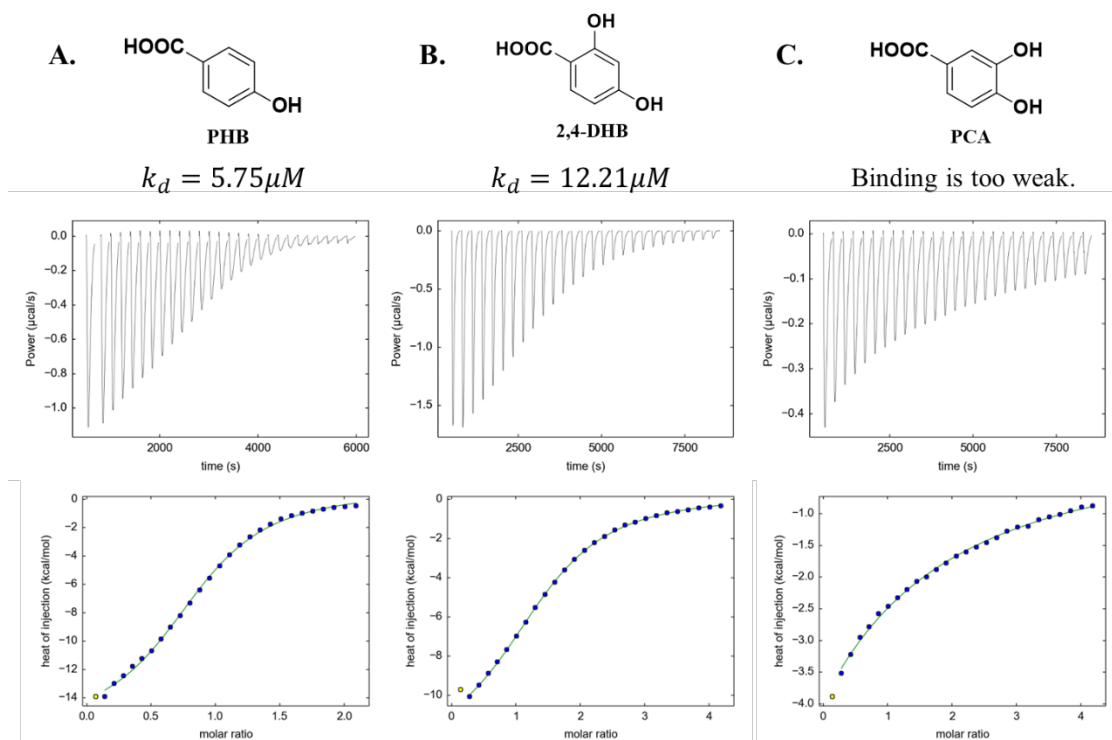


Figure 2.10 Representative binding isotherms of Pobr with aromatic ligands (A) PHB, (B) 2,4-DHB and (C) PCA. Average dissociation constant k_d for each ligand are provided except that the binding of PCA to Pobr is too weak to determine k_d . The upper panel displays raw data and the lower panel depicts the derived binding isotherm plotted against molar ratio of titrant.

2.3.7 PHB, the physiological relevant ligand of PobR, influenced PobR oligomerization

The *E. coli* IclR protein, which was the founding member of this family, was found to exist as tetramer, dimer, and monomer or in an oligomeric equilibrium of these three states (64,65). Mass spectroscopy studies indicated that IclR was predominantly a tetramer in solution (66). A number of IclR family regulators have been reported as a tetramer in complex with its target DNA (*i.e.*, TtgV and BlcR) (67-69).

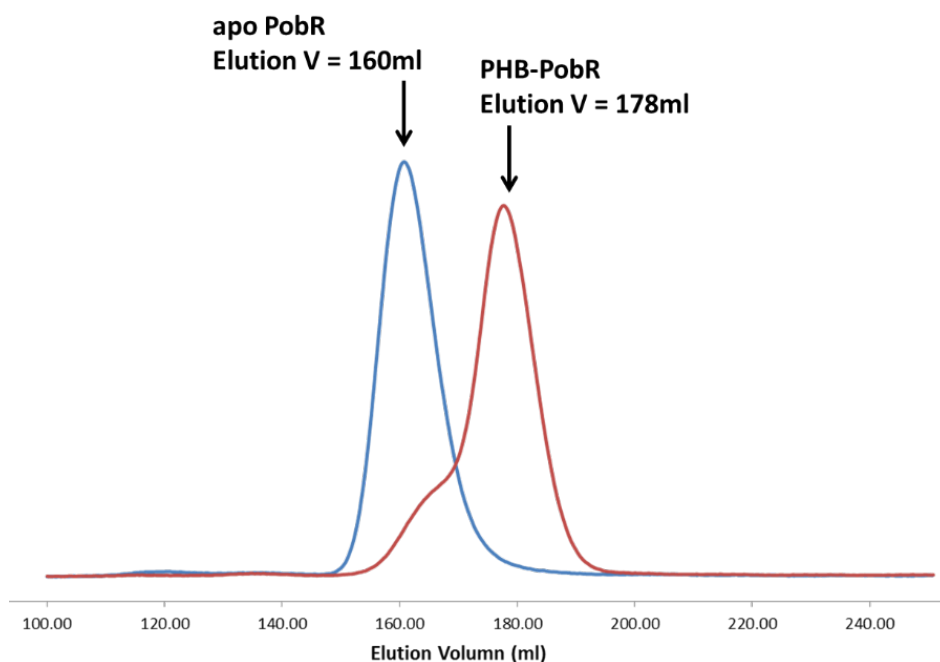


Figure 2.11 Size-exclusion chromatography elution of PobR protein in the absence and presence of PHB, which is the physiological relevant ligand. The protein was incubated with PHB in a molar ratio of 1:10 (monomeric PobR/PHB) in ice for 1 hour before the sample was loaded on Superdex 75 Hiload 26/60 column. The mobile phases used to elute both protein samples were the same (20 mM Tris pH 7.8, 300 mM NaCl, 0.5 mM TCEP, 10% glycerol, 4°C). Thyroglobulin (670 kDa), Gamma-globulin (158 kDa), Ovalbumin (44 kDa), Myoglobin (17 kDa) and Vitamin B12 (1.35 kDa) were used as standards to plot the calibration curve. The blue line represents the apo PobR elution trace and the red line represents the PHB-PobR elution trace. The elution volume of PobR sample was labeled next to the peak.

Given the unusual sequence and domain organization of PobR in *S. coelicolor*, we predicted that PobR would bind DNA as a dimer and exist as a dimer in the absence of ligand. Initially, we found that PobR was eluted as dimer in sizing exclusion chromatography purification (Fig. 2.11).

Table 2.7 The molecular weight (MW) and elution volume of standard proteins

Standard protein	MW (kDa)	Volume (ml)
Thyroglobulin	670	129
Gamma-globulin	158	146
Ovalbumin	44	179
Myoglobin	17	217
Vitamin B12	1.35	284

For the detailed calculations, Thyroglobulin, Gamma-globulin, Ovalbumin, Myoglobin, and Vitamin B12 were used as standards to plot the calibration curve. The molecular weight (MW) and elution volume of each standard is listed in Table 2.7. The calibration curve was plotted based on equation (1), wherein V_0 represents void volume of the column, V_C represents the geometric bed volume of the column and V_R is the elution volume of the protein. The column we used to purify PobR is Superdex 75 26/60. The V_0 is 121 ml and the V_C is 318 ml. Based on the elution volume of PobR, we were able to determine the oligomerization state of the protein.

$$K_{av} = \frac{V_R - V_0}{V_C - V_0} \quad (1)$$

To test whether PHB affected oligomerization of PobR, apo PobR was incubated with different concentrations of PHB (with molar ratio of monomeric PobR/PHB=1:5 and 1:10) on ice for 1 hour and then run over the same SEC column with identical buffer as apo PobR. PobR-PHB eluted predominantly as monomer in solution (Fig. 2.11). This observation indicated that PHB induces

conformational changes in PobR that favor a monomeric species. Importantly, this monomeric species is analogous to the known dimeric state of canonical IclR proteins.

2.3.8 Functional analyses of PobR N-terminal domain and C-terminal domains

The biochemical studies of the binding of PobR to DNA and the assessment of its binding to PHB and other ligands raised interesting questions about the structure of PobR. Accordingly, we set out to use protein crystallography to assess the structures of full length PobR and/or its N- and C-terminal domains. To construct synthetic genes that would yield proteins amenable to crystallization, we used bioinformatic tools. First, we found that the first 11 and the last 13 amino acids of PobR were predicted to be disordered. Accordingly, we constructed PobR 12-511 via PCR amplification of a truncated *pobR* gene. This PCR product was cloned into pETRP1B for expression in *E. coli*. Second, we predicted that the truncated PobR and the individual domains would be more amenable to crystallization. We used bioinformatics to define the boundaries between the N- and C-terminal domains. N-terminal domain (12-267) and C-terminal domain (293-511), open reading frames for these individual domains were amplified by PCR and cloned into pETRP1B for heterologous expression in *E. coli*.



Figure 2.12 Transcription profile of *pobA* gene expression in wild-type *S. coelicolor* and various mutant strains. RT-PCR was used to detect the transcription of *pobA*. Only the over-expression of the full length *pobR* ORF is capable of suppressing the phenotype of *S. coelicolor* *pobR* null mutant. The *pobA* gene was transcribed constitutively upon the over-expression of DNA fragment encoding PobR N-terminal domain, DNA fragment encoding PobR C-terminal domain and DNA fragment encoding truncated PobR.

For PobR 12-511, it was found that the truncated *pobR* gene expressed well at 18°C under the induction of IPTG. However, the protein construct turned out to be insoluble during purification. The same problem was encountered while purifying the N-terminal domain of PobR. The gene encoding PobR C-terminal domain only showed moderate level of expression and the corresponding protein was also insoluble when we performed protein purification.

In order to elucidate the function of each domain and understand how each domain contributes to the function of the overall protein, we switched to *in vivo* genetic analyses. All the gene fragments including the *pobR* ORF, DNA fragment encoding PobR N-terminal domain, DNA fragment encoding PobR C-terminal domain and DNA fragment encoding truncated PobR were heterologously expressed in *S. coelicolor* B1656 individually. For these experiments, each gene fragment was cloned into integrative plasmid pIJ10257 for expression under the control of the constitutive *ermE** promoter and then introduced into the *S. coelicolor pobR* null mutant through conjugation. The transcription of *pobA* gene was examined via RT-PCR analysis. However, it was found that the *pobA* gene was constitutively transcribed in all the newly constructed strains as it behaved in the *pobR* null strain, which means that over-expression of partial *pobR* gene was not capable of suppressing the phenotype of *pobR* null mutant (Fig. 2.12). A possible reason for this observation is that the expressed gene products were not active and insoluble in the *S. coelicolor pobR* null strain. Overall, both biochemical as well as genetic experiments did not generate any promising results concerning the functions of PobR N-terminal and C-terminal domains.

2.3.9 Preliminary analyses of the crystal structure of the PobR-PHB complex

In collaboration with Prof. Rebecca Page in the department of molecular biology, cell biology, and biochemistry at Brown, we sought to grow crystals of PobR in the presence and absence of PHB. We were able to crystallize both apo PobR and the PobR-PHB complex. Unfortunately, only the crystals of the PobR-PHB complex yielded an interpretable diffraction pattern. We determined the PobR-PHB complex structure to 2.4Å (Fig. 2.13A). The in-depth analyses of the crystal structure

are in progress and not ready for disclosure. In addition, Small Angle X-ray Scattering (SAXS) experiment was also performed for both apo PobR as well as PobR-PHB and these analyses are also in progress.

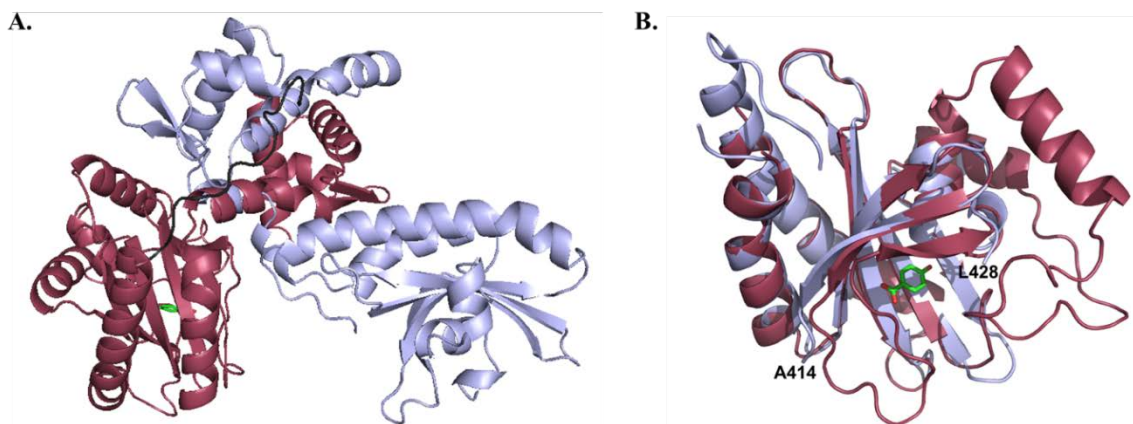


Figure 2.13 The crystal structure of the PobR-PHB complex. (A) The PobR monomer is shown as a cartoon representation with the N-terminus in raspberry and the C-terminus in light blue. The PHB molecule is shown as sticks and colored by elements. (B) Structures alignment between PobR N-terminal domain and C-terminal domain. It is shown that the loop closing on PHB in N-terminal domain is the loop disordered in C-terminal domain.

Through preliminary analyses of the crystal structure of PobR-PHB complex, it was found that only one PHB molecule binds to the N-terminal domain of PobR, which is consistent with the stoichiometry results deduced from the ITC experiments. As expected, the structures of N-terminal domain and C-terminal domain are distinct. The loop closing on PHB in N-terminal domain is the loop disordered in C-terminal domain, which also accounts for the single PHB molecule binding. (Fig. 2.13B)

2.4 Conclusions

In this chapter, we characterized the catabolism of the lignin-derived aromatic compound PHB and its regulation in *S. coelicolor*. The PHB hydroxylase, encoded by gene *pobA* (SCO3084), is responsible for fluxing PHB through PCA into the β -ketoadipate pathway and ultimately the Krebs cycle. This hydroxylation was found under the regulation of an IclR family protein PobR, encoded by the *pobR* gene (SCO3209). The proposed model is that PobR binds to the promoter region of *pobA* gene and blocks the access of RNA polymerase, thus represses the transcription of *pobA* gene in the absence of PHB. While upon addition of PHB, the aromatic ligand binds to the regulator, which dissociates PobR from *pobA* promoter so that the structural gene can be transcribed (Fig. 2.14).

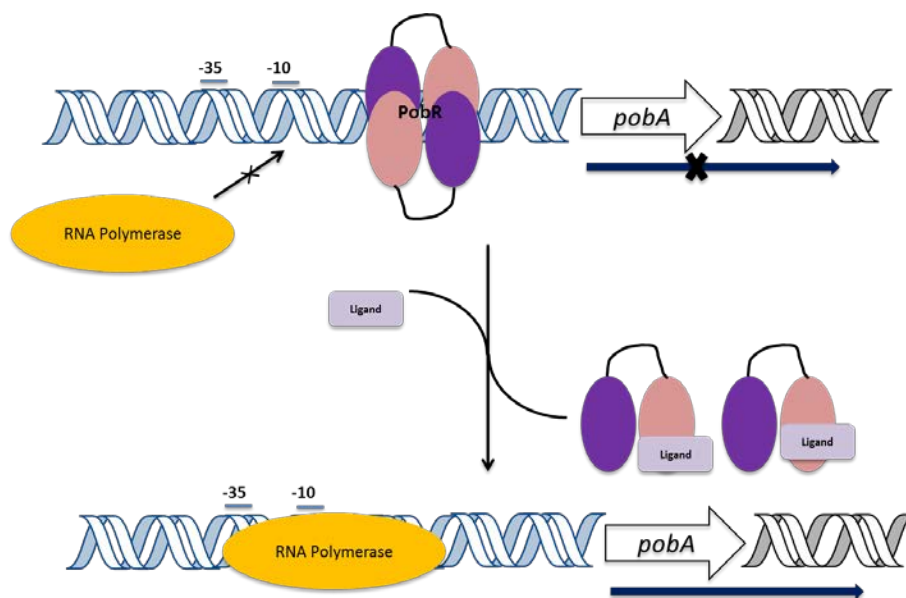


Figure 2.14 Validated model for transcription regulation by PobR.

Biologically, PobR is intriguing because it is the first identified IclR family regulator that functions as a repressor in the degradation of aromatic compounds (21). PobR is peculiar owing to its sequence and structure. Each domain (N-terminal domain and C-terminal domain) is structurally

similar to an intact IclR family protein. Curiously, sequence alignment indicated that there is no nucleotide sequence similarity between the two DNA fragments encoding N-terminal domain and C-terminal domain. The amino acid sequence similarity of the two domains is also low with an identity of only 44%. These alignment results rule out the probability that gene duplication accounts for the generation of *pobR* gene. We believe that gene excision and re-assembly likely occurred during evolution to yield the peculiar organization of the *pobR* gene. These genetic events were likely recent because PobR is only observed and conserved in members of the *Streptomyces* genus as shown in Table 2.5. Currently, our preliminary biophysical results demonstrated that aromatic compound PHB interacts with PobR, with the K_d of 5.75 μM . Moreover, they indicated that PHB trigger a conformational change in dimeric PobR that favors a monomeric structure. We believe that our on-going analyses on PHB-PobR crystal structure will generate more interesting conclusions. Overall, this work provides novel insights into the molecular mechanisms of IclR family transcriptional repressors in aromatic catabolism and the structures of IclR family members. These studies of aromatic catabolism open the door to the exploitation of microorganisms for the conversion of aromatic compounds to commodity chemicals and biofuels.

2.5 Scientific Contributions

In this chapter, Dr. Jennifer Davis (MPPB¹, Brown University) constructed *S. coelicolor* strain B1653 (M600 $\Delta\text{SCO3084}::\text{apr}$). Dr. Dana Lord (MPPB¹, Brown University) and Prof. Rebecca Page (MCB², Brown University) solved the crystal structure of PobR-PHB complex. Dr. Dana Lord also processed data analyses in ITC and thermafluor experiments. All the other experiments and data analyses were performed by Rui Zhang.

1. MPPB: Department of Molecular Pharmacology, Physiology and Biotechnology
2. MCB: Department of Molecular Biology, Cell Biology and Biochemistry

2.6 References

1. Adler, E. (1977) Lignin Chemistry - Past, Present and Future. *Wood Sci Technol*, **11**, 169-218.
2. Rubin, E.M. (2008) Genomics of cellulosic biofuels. *Nature*, **454**, 841-845.
3. Weng, J.K., Li, X., Bonawitz, N.D. and Chapple, C. (2008) Emerging strategies of lignin engineering and degradation for cellulosic biofuel production. *Curr Opin Biotechnol*, **19**, 166-172.
4. Chundawat, S.P., Beckham, G.T., Himmel, M.E. and Dale, B.E. (2011) Deconstruction of lignocellulosic biomass to fuels and chemicals. *Annu Rev Chem Biomol Eng*, **2**, 121-145.
5. Dean, B.J. (1985) Recent findings on the genetic toxicology of benzene, toluene, xylenes and phenols. *Mutation Research/Reviews in Genetic Toxicology*, **154**, 153-181.
6. Fuchs, G., Boll, M. and Heider, J. (2011) Microbial degradation of aromatic compounds - from one strategy to four. *Nat Rev Microbiol*, **9**, 803-816.
7. Max M. Haggblom, P.W.M. (2000) Anaerobic Biosdegradation of Halogenated Pesticides: Influence of Alternate Electron Acceptors. *Soil Biochemistry (Marcel Dekker, Inc. New York)*, **10**, 1-22.
8. Ornston, L.N. and Parke, D. (1977) The evolution of induction mechanisms in bacteria: insights derived from the study of the beta-ketoadipate pathway. *Curr Top Cell Regul*, **12**, 209-262.
9. Harwood, C.S. and Parales, R.E. (1996) The beta-ketoadipate pathway and the biology of self-identity. *Annu Rev Microbiol*, **50**, 553-590.
10. Naofumi Kaminura, E.M. (2014) The Protocatechuate 4,5-Cleavage Pathway: Overview and New Findings. *Biodegradative Bacteria: How Bacteria Degrade, Survive, Adapt, and Evolve* (Springer, Japan), 207-226.
11. Crawford, R.L. (1975) Novel pathway for degradation of protocatechuic acid in *Bacillus species*. *J Bacteriol*, **121**, 531-536.
12. Crawford, R.L., Bromley, J.W. and Perkins-Olson, P.E. (1979) Catabolism of protocatechuate by *Bacillus macerans*. *Appl Environ Microbiol*, **37**, 614-618.
13. Kasai, D., Fujinami, T., Abe, T., Mase, K., Katayama, Y., Fukuda, M. and Masai, E. (2009) Uncovering the protocatechuate 2,3-cleavage pathway genes. *J Bacteriol*, **191**, 6758-6768.
14. Masai, E., Katayama, Y. and Fukuda, M. (2007) Genetic and biochemical investigations on bacterial catabolic pathways for lignin-derived aromatic compounds. *Biosci Biotechnol Biochem*, **71**, 1-15.

15. Paul, D., Chauhan, A., Pandey, G. and Jain, R.K. (2004) Degradation of *p*-hydroxybenzoate via protocatechuate in *Arthrobacter protophormiae* RKJ100 and *Burkholderia cepacia* RKJ200. *Curr Sci India*, **87**, 1263-1268.
16. Hosokawa, K. and Stanier, R.Y. (1966) Crystallization and properties of *p*-hydroxybenzoate hydroxylase from *Pseudomonas putida*. *J Biol Chem*, **241**, 2453-2460.
17. Howell, L.G., Spector, T. and Massey, V. (1972) Purification and properties of *p*-hydroxybenzoate hydroxylase from *Pseudomonas fluorescens*. *J Biol Chem*, **247**, 4340-4350.
18. DiMarco, A.A., Averhoff, B. and Ornston, L.N. (1993) Identification of the transcriptional activator *pobR* and characterization of its role in the expression of *pobA*, the structural gene for *p*-hydroxybenzoate hydroxylase in *Acinetobacter calcoaceticus*. *J Bacteriol*, **175**, 4499-4506.
19. DiMarco, A.A. and Ornston, L.N. (1994) Regulation of *p*-hydroxybenzoate hydroxylase synthesis by *PobR* bound to an operator in *Acinetobacter calcoaceticus*. *J Bacteriol*, **176**, 4277-4284.
20. Bertani, I., Kojic, M. and Venturi, V. (2001) Regulation of the *p*-hydroxybenzoic acid hydroxylase gene (*pobA*) in plant-growth-promoting *Pseudomonas putida* WCS358. *Microbiology*, **147**, 1611-1620.
21. Díaz, E. and Prieto, M.a.A. (2000) Bacterial promoters triggering biodegradation of aromatic pollutants. *Current Opinion in Biotechnology*, **11**, 467-475.
22. Davis, J.R. and Sello, J.K. (2010) Regulation of genes in *Streptomyces* bacteria required for catabolism of lignin-derived aromatic compounds. *Appl Microbiol Biotechnol*, **86**, 921-929.
23. Davis, J.R., Brown, B.L., Page, R. and Sello, J.K. (2013) Study of *PcaV* from *Streptomyces coelicolor* yields new insights into ligand-responsive MarR family transcription factors. *Nucleic Acids Res*, **41**, 3888-3900.
24. Lorenzo, V. and Pérez-Martín, J. (1996) Regulatory noise in prokaryotic promoters: how bacteria learn to respond to novel environmental signals. *Molecular Microbiology*, **19**, 1177-1184.
25. Pometto, A.L. and Crawford, D.L. (1986) Catabolic Fate of *Streptomyces Viridosporus* T7a-Produced, Acid-Precipitable Polymeric Lignin Upon Incubation with Ligninolytic *Streptomyces* Species and *Phanerochaete-Chrysosporium*. *Appl Environ Microb*, **51**, 171-179.
26. Grund, E., Knorr, C. and Eichenlaub, R. (1990) Catabolism of benzoate and monohydroxylated benzoates by *Amycolatopsis* and *Streptomyces spp.* *Appl Environ Microbiol*, **56**, 1459-1464.

27. Iwagami, S.G., Yang, K. and Davies, J. (2000) Characterization of the protocatechuic acid catabolic gene cluster from *Streptomyces* sp. strain 2065. *Appl Environ Microbiol*, **66**, 1499-1508.
28. Romero-Rodriguez, A., Robledo-Casados, I. and Sanchez, S. (2015) An overview on transcriptional regulators in *Streptomyces*. *Biochim Biophys Acta*, **1849**, 1017-1039.
29. Kornberg, H.L. (1967) The regulation of anaplerotic enzymes in *E. coli*. *Bull Soc Chim Biol (Paris)*, **49**, 1479-1490.
30. Maloy, S.R. and Nunn, W.D. (1982) Genetic regulation of the glyoxylate shunt in *Escherichia coli* K-12. *J Bacteriol*, **149**, 173-180.
31. Lorca, G.L., Ezersky, A., Lunin, V.V., Walker, J.R., Altamentova, S., Evdokimova, E., Vedadi, M., Bochkarev, A. and Savchenko, A. (2007) Glyoxylate and pyruvate are antagonistic effectors of the *Escherichia coli* IclR transcriptional regulator. *J Biol Chem*, **282**, 16476-16491.
32. Fillet, S., Krell, T., Morel, B., Lu, D., Zhang, X. and Ramos, J.L. (2011) Intramolecular signal transmission in a tetrameric repressor of the IclR family. *Proc Natl Acad Sci U S A*, **108**, 15372-15377.
33. Krell, T., Molina-Henares, A.J. and Ramos, J.L. (2006) The IclR family of transcriptional activators and repressors can be defined by a single profile. *Protein Sci*, **15**, 1207-1213.
34. Hugouvieux-Cotte-Pattat, N., Dominguez, H. and Robert-Baudouy, J. (1992) Environmental conditions affect transcription of the pectinase genes of *Erwinia chrysanthemi* 3937. *J Bacteriol*, **174**, 7807-7818.
35. Nasser, W., Reverchon, S. and Robert-Baudouy, J. (1992) Purification and functional characterization of the KdgR protein, a major repressor of pectinolysis genes of *Erwinia chrysanthemi*. *Molecular Microbiology*, **6**, 257-265.
36. Nasser, W., Reverchon, S., Condemine, G. and Robert-Baudouy, J. (1994) Specific interactions of *Erwinia chrysanthemi* KdgR repressor with different operators of genes involved in pectinolysis. *J Mol Biol*, **236**, 427-440.
37. Molina-Henares, A.J., Krell, T., Eugenia Guazzaroni, M., Segura, A. and Ramos, J.L. (2006) Members of the IclR family of bacterial transcriptional regulators function as activators and/or repressors. *FEMS Microbiol Rev*, **30**, 157-186.
38. Tropel, D. and van der Meer, J.R. (2004) Bacterial transcriptional regulators for degradation pathways of aromatic compounds. *Microbiol Mol Biol Rev*, **68**, 474-500, table of contents.
39. Sambrook J., F.E.F., T. Maniatis. (1989) *Molecular cloning: a laboratory manula* (2nd ed, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N. Y.).

40. Kieser T, B.M., Buttner MJ, Chater KF, Hopwood DA. (2000) *Practical Streptomyces Genetics* (John Innes Foundation, Norwich, England).
41. Gust, B., Challis, G.L., Fowler, K., Kieser, T. and Chater, K.F. (2003) PCR-targeted *Streptomyces* gene replacement identifies a protein domain needed for biosynthesis of the sesquiterpene soil odor geosmin. *Proc Natl Acad Sci U S A*, **100**, 1541-1546.
42. Jones, G.H., Paget, M.S.B., Chamberlin, L. and Buttner, M.J. (1997) Sigma-E is required for the production of the antibiotic actinomycin in *Streptomyces antibioticus*. *Molecular Microbiology*, **23**, 169-178.
43. Gregory, M.A., Till, R. and Smith, M.C.M. (2003) Integration Site for *Streptomyces* Phage BT1 and Development of Site-Specific Integrating Vectors. *J Bacteriol*, **185**, 5320-5323.
44. Peti, W. and Page, R. (2007) Strategies to maximize heterologous protein expression in *Escherichia coli* with minimal cost. *Protein Expr Purif*, **51**, 1-10.
45. Parke, D. (1992) Application of *p*-Toluidine in Chromogenic Detection of Catechol and Protocatechuate, Diphenolic Intermediates in Catabolism of Aromatic Compounds. *Appl Environ Microbiol*, **58**, 2694-2697.
46. Hong, H.J., Hutchings, M.I., Hill, L.M. and Buttner, M.J. (2005) The role of the novel Fem protein VanK in vancomycin resistance in *Streptomyces coelicolor*. *J Biol Chem*, **280**, 13055-13061.
47. Bibb, M.J., White, J., Ward, J.M. and Janssen, G.R. (1994) The mRNA for the 23S rRNA methylase encoded by the *ermE* gene of *Saccharopolyspora erythraea* is translated in the absence of a conventional ribosome-binding site. *Molecular Microbiology*, **14**, 533-545.
48. Keller, S., Vargas, C., Zhao, H., Piszczek, G., Brautigam, C.A. and Schuck, P. (2012) High-precision isothermal titration calorimetry with automated peak-shape analysis. *Anal Chem*, **84**, 5066-5073.
49. Houtman, J.C., Brown, P.H., Bowden, B., Yamaguchi, H., Appella, E., Samelson, L.E. and Schuck, P. (2007) Studying multisite binary and ternary protein interactions by global analysis of isothermal titration calorimetry data in SEDPHAT: application to adaptor protein complexes in cell signaling. *Protein Sci*, **16**, 30-42.
50. Vedadi, M., Niesen, F.H., Allali-Hassani, A., Fedorov, O.Y., Finerty, P.J., Jr., Wasney, G.A., Yeung, R., Arrowsmith, C., Ball, L.J., Berglund, H. *et al.* (2006) Chemical screening methods to identify ligands that promote protein stability, protein crystallization, and structure determination. *Proc Natl Acad Sci U S A*, **103**, 15835-15840.
51. Niesen, F.H., Berglund, H. and Vedadi, M. (2007) The use of differential scanning fluorimetry to detect ligand interactions that promote protein stability. *Nat Protoc*, **2**, 2212-2221.

52. Otwinowski, Z. and Minor, W. (1997) [20] Processing of X-ray diffraction data collected in oscillation mode. **276**, 307-326.
53. Datsenko, K.A. and Wanner, B.L. (2000) One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. *Proc Natl Acad Sci U S A*, **97**, 6640-6645.
54. Muller, F., Voordouw, G., Berkel, W.J.H., Steennis, P.J., Visser, S. and Rooijen, P.J. (1979) A Study of *p*-Hydroxybenzoate Hydroxylase from *Pseudomonas fluorescens*. Improved Purification, Relative Molecular Mass, and Amino Acid Composition. *European Journal of Biochemistry*, **101**, 235-244.
55. Bentley, S.D., Chater, K.F., Cerdeno-Tarraga, A.M., Challis, G.L., Thomson, N.R., James, K.D., Harris, D.E., Quail, M.A., Kieser, H., Harper, D. *et al.* (2002) Complete genome sequence of the model actinomycete *Streptomyces coelicolor* A3(2). *Nature*, **417**, 141-147.
56. Altschul, S.F., Gish, W., Miller, W., Myers, E.W. and Lipman, D.J. (1990) Basic local alignment search tool. *Journal of Molecular Biology*, **215**, 403-410.
57. Finn, R.D., Clements, J. and Eddy, S.R. (2011) HMMER web server: interactive sequence similarity searching. *Nucleic Acids Res*, **39**, W29-37.
58. Jaroszewski, L., Li, Z., Cai, X.H., Weber, C. and Godzik, A. (2011) FFAS server: novel features and applications. *Nucleic Acids Res*, **39**, W38-44.
59. Dosztanyi, Z., Csizmok, V., Tompa, P. and Simon, I. (2005) IUPred: web server for the prediction of intrinsically unstructured regions of proteins based on estimated energy content. *Bioinformatics*, **21**, 3433-3434.
60. Strohl, W.R. (1992) Compilation and analysis of DNA sequences associated with apparent streptomycete promoters. *Nucleic Acids Res*, **20**, 961-974.
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. *et al.* (2007) Clustal W and Clustal X version 2.0. *Bioinformatics*, **23**, 2947-2948.
62. Thompson, W. (2003) Gibbs Recursive Sampler: finding transcription factor binding sites. *Nucleic Acids Research*, **31**, 3580-3585.
63. Crooks, G.E., Hon, G., Chandonia, J.M. and Brenner, S.E. (2004) WebLogo: a sequence logo generator. *Genome Res*, **14**, 1188-1190.
64. Nègre, D., Cortay, J.-C., Galinier, A., Sauve, P. and Cozzone, A.J. (1992) Specific interactions between the IclR repressor of the acetate operon of *Escherichia coli* and its operator. *Journal of Molecular Biology*, **228**, 23-29.
65. Donald, L.J., Chernushevich, I.V., Zhou, J., Verentchikov, A., Poppe-Schriemer, N., Hosfield, D.J., Westmore, J.B., Ens, W., Duckworth, H.W. and Standing, K.G. (1996) Preparation and properties of pure, full-length IclR protein of *Escherichia coli*. Use of time-

of-flight mass spectrometry to investigate the problems encountered. *Protein Sci*, **5**, 1613-1624.

66. Donald, L.J., Hosfield, D.J., Cuvelier, S.L., Ens, W., Standing, K.G. and Duckworth, H.W. (2001) Mass spectrometric study of the *Escherichia coli* repressor proteins, Ic1R and Gc1R, and their complexes with DNA. *Protein Sci*, **10**, 1370-1380.
67. Guazzaroni, M.E., Krell, T., Gutierrez del Arroyo, P., Velez, M., Jimenez, M., Rivas, G. and Ramos, J.L. (2007) The transcriptional repressor TtgV recognizes a complex operator as a tetramer and induces convex DNA bending. *J Mol Biol*, **369**, 927-939.
68. Pan, Y., Fiscus, V., Meng, W., Zheng, Z., Zhang, L.H., Fuqua, C. and Chen, L. (2011) The *Agrobacterium tumefaciens* transcription factor BlcR is regulated via oligomerization. *J Biol Chem*, **286**, 20431-20440.
69. Pan, Y., Wang, Y., Fuqua, C. and Chen, L. (2013) In vivo analysis of DNA binding and ligand interaction of BlcR, an IclR-type repressor from *Agrobacterium tumefaciens*. *Microbiology*, **159**, 814-822.

**Chapter 3. Genetic and biochemical analyses of
indolmycin biosynthesis and resistance in
Streptomyces griseus ATCC 12648**

3.1 Introduction

Indolmycin is an antibiotic that was first isolated from *Streptomyces griseus* ATCC 12648. It has potent activity against Gram-positive and Gram-negative bacteria (1-4). Given its structural similarity to tryptophan, indolmycin competitively inhibits bacterial tryptophanyl-tRNA synthetases (TrpRSs) (5). Pfizer, who patented indolmycin as an antibacterial drug lead (6), suspended drug development activities after the discovery that indolmycin perturbed the metabolism of tryptophan in rat liver by inhibiting enzymes involved in the catabolism of this amino acid (7). Interest in the pharmacological potential of indolmycin has been renewed again because the compound demonstrated good bioactivity against important human pathogens like *Helicobacter pylori* and methicillin-resistant *Staphylococcus aureus* (MRSA) (8,9). Because the compound's inherent toxicity suggests that it cannot be administered internally for the treatment of infections, it has been proposed that indolmycin may be useful as a topical anti-infective (9).

The emergence of drug resistance must be anticipated when evaluating the clinical potential of any antibacterial drug lead. For this reason, there has been some interest in bacterial resistance to indolmycin. It was reported that *Staphylococci* strains developed spontaneous indolmycin resistance by acquiring missense point mutations within the *trpRS* gene (9). In *E. coli*, indolmycin resistance is ascribed to up-regulated expression of *trpRS* gene (10). Besides these two cases of resistance, studies by Söll and Sello have shown that resistance to indolmycin is associated with auxiliary tryptophanyl-tRNA synthetases that are not inhibited by the antibiotic. Interestingly, they have found that these auxiliary genes in *Streptomyces* bacteria are transcriptionally up-regulated when the bacteria are exposed to indolmycin (11-14).

In addition to its antibacterial activity, indolmycin has attracted attention due to its unique structure. Indolmycin is the only secondary metabolite having an oxazolinone ring (Fig 3.1). There has been interest in the biosynthesis origins of this molecule since late 1960s. The first experiments in biosynthesis studies involved feeding *S. griseus* ATCC 12648 isotopically labelled precursors and

mapping the fates of the labeled atoms within indolmycin (15-17). It was found that L-tryptophan, arginine and S-adenosyl methionine (SAM) are the three precursors in indolmycin biosynthetic pathway (Fig 3.2A). Those studies revealed that the backbone of indolmycin comes from tryptophan, whereas the oxazolinone ring is derived from the guanidinium moiety of arginine. The two methyl groups were proposed to be derived from SAM. Following the labeling experiments, the Floss group proposed a tentative biosynthetic pathway (Fig 3.2B) (17). The biosynthesis of indolmycin begins with the transamination of L-tryptophan to yield indolepyruvic acid. The indolpyruvate is methylated at position 3 to yield 3-methylindolepyruvate. This methylated indole pyruvate is then reduced to yield indolmycenic acid. The amidino group of arginine is transferred to indolmycenic acid enabling cyclization. N-methylation is thought to be the final step in indolmycin biosynthesis (18).

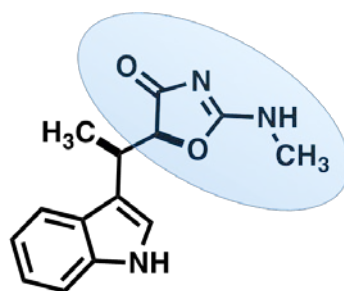
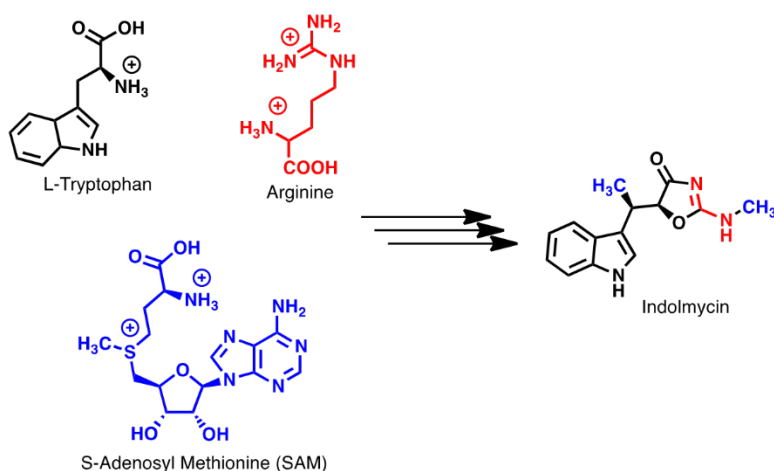


Figure 3.1 Structure of indolmycin. The oxazolinone ring is highlighted in blue.

Since the early 1970s, no work has been conducted on indolmycin biosynthesis. By sequencing the genome of *S. griseus* ATCC 12648, we decided to revisit this long-time unanswered mystery of how indolmycin is biosynthesized. The challenge was to identify the indolmycin biosynthetic genes among all of the genes in this bacterium. Our strategy for identifying the biosynthetic gene was based on the idea that antibiotic resistance genes are typically clustered with antibiotic biosynthetic genes (19). Based on precedents of auxiliary tryptophanyl-tRNA synthetases, we examined the genome of *S. griseus* ATCC 12648. We discovered three genes encoding tryptophanyl-tRNA

synthetase and used genetic and biochemical approaches to assess the susceptibility of the cognate enzymes to inhibition by indolmycin. We found that the most highly indolmycin resistant synthetase was clustered with genes whose functions were predicted to be capable of catalyzing transformation in the indolmycin biosynthesis. Cloning of these genes and their heterologous expression in *S. coelicolor* enabled the bacterium to produce indolmycin. Our genetic studies are complemented by recent biochemical studies by Ryan and coworkers (20).

A. The biosynthetic origins of Indolmycin



B. Proposed biosynthesis of indolmycin

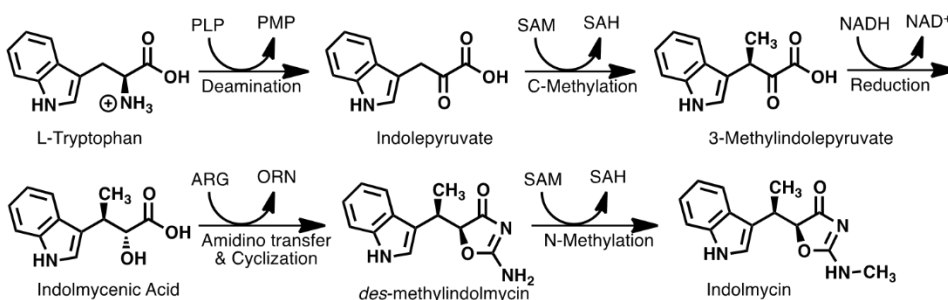


Figure 3.2 (A) Labeling studies have revealed that indolmycin is biosynthesized from L-tryptophan, arginine, and SAM (15-17). (B) The proposed biosynthesis of indolmycin with intermediates and necessary cofactors (17,18).

3.2 Materials and Methods

3.2.1 Bacterial strains and culture conditions

A complete list of bacterial strains employed in this study is provided in Table 3.1. *Escherichia coli* strains DH5 α , 10-beta, ET12567/pUZ8002 were grown on Luria-Bertani medium at 37°C for routine subcloning (21). *E. coli* BL21/pGro7 was grown in ZYM-505 media (22) for heterologous expression of genes encoding the tryptophanyl-tRNA synthetases. *S. griseus* ATCC 12648 wild-type strain and *S. coelicolor* strains were grown at 30°C on mannitol soya flour medium (SFM), Difco Nutrient Agar medium (DNA), minimal liquid medium (NMMP) (21) and R3 medium (23). SFM was used for spore generation of *S. griseus* and *S. coelicolor*, and for *S. coelicolor*/*E. coli* conjugations. DNA medium was used for MIC (minimum inhibitory concentration) assay. NMMP was used to grow *S. griseus* ATCC 12648 wild-type strain for growth curve plotting, total mRNA isolation and indolmycin production. The optimal medium for indolmycin production is R3 medium. For selection of *E. coli*, antibiotics were used at the following concentration: ampicillin (100 μ g/ml), apramycin (50 μ g/ml), chloramphenicol (25 μ g/ml and 34 μ g/ml), hygromycin (75 μ g/ml), kanamycin (50 μ g/ml and 100 μ g/ml). For selection of *S. coelicolor*, apramycin (50 μ g/ml) and hygromycin (75 μ g/ml) were used. In conjugation with *S. coelicolor*, nalidixic acid (20 μ g/ml) was used to counterselect *E. coli*.

3.2.2 Plasmids and primers

All of the plasmids employed in this study are listed in Table 3.2. DNA sequencing was performed by GENEWIZ (South Plainfield, NJ). Table 3.3 shows the primers used in this work, all of which were synthesized by Invitrogen. PCR was performed with *Taq* DNA Polymerase (Invitrogen), *Pfu* (Stratagene, Agilent Technologies), Crimson *Taq* DNA Polymerase (New England Biolabs) and Q5 High-Fidelity DNA Polymerase (New England Biolabs). All PCR reactions with *Taq*, *Pfu* and

Crimson *Taq* DNA polymerase were performed with 5% (vol/vol) DMSO. 5× Q5 High GC Enhancer was added into PCR reactions with Q5 High-Fidelity DNA Polymerase (24).

3.2.3 Genome sequencing and assembly

The draft genome sequence of *Streptomyces griseus* ATCC 12648 was generated by BGI Tech (a subsidiary of Beijing Genomic Institute). High-throughput Illumina sequencing technology (25) was used to construct library and generate original reads. The subsequent assembly of the short reads into genome sequence was performed using SOAP *de novo*, (<http://soap.genomics.org.cn/soapdenovo.html>; version: 1.05). Briefly, the short reads were first fragmented into kmer as nodes. Two nodes would be connected on a line if they have relationship. De Bruijn graph was then constructed by these linkages. The second step of assembly is construction of contigs using located sequences. Thirdly, scaffold was assembled by connecting contigs through paired-end reads support. Finally, all the reads were mapped to assembled scaffolds and paired-end reads relationship was used to fill gaps.

3.2.4 Growth curve

A approximately 1×10^9 spores of wild-type *S. griseus* ATCC 12648 were germinated in 2 ×YT (24) media, inoculated into flasks containing 30 ml NMMP, and then grown as shaken liquid cultures for 54 h at 30°C. At each indicated time, 1.5 ml sample was collected from flask and dry cell weights were measured.

3.2.5 Heterologous overexpression of SGI0214, SGI1617 and SGI6453 in *S. coelicolor*

The SGI0214 ORF was amplified from *S. griseus* ATCC 12648 genomic DNA using primers SGI0214 Clone For and SGI0214 Clone Rev. The PCR product, with an engineered NdeI site at 5' end and an engineered HindIII site at its 3' end, was ligated into pBluescript KS+ to yield pJS893. The fragment containing SGI0214 ORF was then excised from pJS893 with NdeI and HindIII and ligated into vector pIJ10257 (26) pretreated with the same enzymes to yield pJS896. pIJ10257 is a

site-specific integrating vector that insert into the Φ BT1 *attB* site of *S. coelicolor* (SCO4848) (27). Gene that is cloned into NdeI-HindIII site is under control of strong constitutive promoter *ermEp** (28). pJS896 was then transformed into ET12567/pUZ8002 and introduced into *S. coelicolor* B725 strain (yielding strain B1659) via conjugation, as previously described (29). Exconjugants were selected by hygromycin resistance. Vector pIJ10257 was also introduced into *S. coelicolor* B725 separately to yield B1662 strain as the negative control.

The SGI1617 ORF was amplified from *S. griseus* ATCC 12648 genomic DNA using primers SGI1617 Clone For and SGI1617 Clone Rev. The PCR product, with an engineered NdeI site at 5' end and an engineered HindIII site at 3' end, was cloned into pBluescript KS+ and subsequently subcloned into pIJ10257 to yield pJS897, which was then introduced into *S. coelicolor* B725 strain (yielding strain B1660) through conjugation.

The SGI6453 ORF was amplified from *S. griseus* ATCC 12648 genomic DNA using primers SGI6453 Clone For and SGI6453 Clone Rev. The PCR product, with an engineered NdeI site at 5' end and an engineered HindIII site at 3' end, was cloned into pBluescript KS+ and subsequently subcloned into pIJ10257 to yield pJS898, which was then introduced into *S. coelicolor* B725 strain (yielding strain B1661) through conjugation.

3.2.6 Indolmycin MIC determination

The MIC assay was performed on Difco Nutrient Agar medium supplemented with the indicated concentrations of indolmycin. Briefly, indolmycin was resuspended in DMSO and tested in a range from 0.125 to 200 μ g/ml. For all the tested strains, after addition of 2×10^8 spores into each well, the 96-well plates were incubated at 30°C for 48 h. The lowest drug concentration that inhibited $\geq 90\%$ growth was considered to be the MIC.

3.2.7 Construction of pETRP1B-His-TrpRS vectors

All the three TrpRS gene ORFs were amplified via PCR and cloned into a plasmid for expression in *E. coli*. To construct the TrpRS gene expression vector, the fragment containing SGI0214 ORF was excised from pJS893 with NdeI and HindIII and ligated into vector pETRP1B (30) pretreated with the same enzymes to yield the expression vector pJS899. The DNA fragment carrying SGI1617 ORF was excised from pJS894 and ligated into pETRP1B to yield pJS900 in the same way. Finally, the fragment with SGI6453 ORF was excised from pJS895 and ligated into pETRP1B to yield pJS901.

3.2.8 Expression of *trpRS* genes and protein purification

Each of the three *trpRS* genes was co-expressed with GroEL/ES chaperones by auto-induction (22). Briefly, BL21 competent cells bearing the pGro7 plasmid (Takara) were transformed with a pETRP1B expression vector for one of the three TrpRS variants. Small, 5 mL cultures were grown overnight in ZYM-505 media supplemented with chloramphenicol (34 µg/ml) and kanamycin (100 µg/ml) to select for the pGro7 and pETRP1B plasmids, respectively. Arabinose (0.05%) was included to induce expression of the GroEL/ES chaperones. One milliliter of the overnight culture was added to 1 L of fresh ZYM-5052 media and grown at 37°C until the cultures reached an OD₆₀₀ of 1. At this point, the temperature was lowered to 25°C and the cells continued to grow for an additional 16 hours. The cells were harvested at 4500 rpm for 30 minutes, resuspended in 80 mL of buffer (25 mM Hepes, 0.3 M NaCl, 1 M urea, 10 mM BME, 30 mM imidazole, pH 8.0) per 1 L of culture and frozen at -20°C. Upon thawing, cells were sonicated and centrifuged (16000 rpm, 4°C, and 1 hour). His-TrpRS was captured from the lysate on Ni-NTA resin and eluted with 0.3 M imidazole. Purified SGI_0214 and SGI_1617 were incubated with TEV protease to remove the N-terminal his-tag. The cleaved protein mixture was passed back over a Ni-NTA column to capture both uncleaved protein and his-TEV protease. SGI_6453 was not efficiently cleaved by TEV protease.

3.2.9 Assays of tryptophanyl-tRNA synthetase activity

Active sites were titrated by following the loss of ATP to determine the fraction of molecules competent for catalysis as described by Fersht (1975) (31,32). Final reaction conditions were as follows: 50 mM Hepes pH 7.5, 5000 cpm/ μ l (γ^{32} P)-ATP, 10 μ M ATP, 0.5 mM tryptophan, 5 mM $MgCl_2$, 0.05 U/ μ l pyrophosphatase, and 3 μ M total TrpRS active sites. The reaction, performed at 37°C, was initiated with enzyme. At discrete time points between 15 seconds and 30 minutes 3 μ l of the reaction were added to 6 μ l of quench buffer (sodium acetate pH 5.3, 1% SDS) and placed on ice. Three microliters of each quenched reaction were spotted onto a cellulose-PEI TLC plate and run in 0.75 M KH_2PO_4 pH 3.5 with 4 M urea to separate $^{32}P_i$ and (γ^{32} P)-ATP. Plates were developed using a Typhoon Imager and analyzed with ImageJ (33) and JMP (34).

3.2.10 Michaelis-Menten kinetics

The incorporation of (32 P)-PPi into ATP was tracked by thin-layer chromatography (32). Reactions contained 5000 cpm/ μ l (32 P)-PPi, 0.1 M Tris pH 8.0, 70 mM BME, 5 mM $MgCl_2$, 10 mM KF, 2 mM PPi, 2 mM ATP and eight tryptophan concentrations from 10 – 600 μ M for SGI_0214 and 65 – 4000 μ M for SGI_1617 and SGI_6453. Reactions were initiated with 100 nM SGI_0214, 600 nM SGI_1617 or 100 nM SGI_6453 at 37°C.

3.2.11 Tryptophanyl-tRNA synthetase inhibition assays

Inhibition assays were performed under the same conditions as the Michaelis-Menten experiments for tryptophan described above at various indolmycin concentrations. Seven indolmycin concentrations from 0.02 – 0.50 μ M for SGI_0214, 0.18 – 60 μ M for SGI_1617, and 0.03 – 900 μ M for SGI_6453 were tested. Results fitted to a competitive inhibition model (1) using non-linear regression to eq. (1) allowed for determination of K_M for tryptophan and K_i for indolmycin using JMP (34).

$$Rate = \frac{[Tryptophan] * k_{cat}}{\left(1 + \frac{[Indolmycin]}{K_i} + [Tryptophan]\right)} \quad (1)$$

3.2.12 Transcriptional analyses

To assess the transcriptional profiles of the three tryptophanyl-tRNA synthetase genes and the indolmycin biosynthesis genes, *S. griseus* ATCC 12648 wild-type shaken liquid culture (30 ml) was grown in NMMP for 64 h at 30°C. At each indicated time (13 h, 18 h, 24 h, 30 h, 36 h and 47 h), an aliquot of culture sufficient to attain a cell pellet volume of 100 μ l was removed for total mRNA isolation. The cells were washed once with 10.3% (w/v) sucrose solution followed by the addition of 100 μ L of 10 mg/ml lysozyme solution (50 mM Tris – HCl, 1 mM EDTA, pH 8.0). The cells were incubated at 37°C for 15 min. Total RNA was isolated using the RNeasy Mini Kit (Qiagen) following the manufacturer’s protocol. The concentration of each purified, DNA-free total RNA isolated was measured with a NanoDrop ND-1000 spectrophotometer. An equal amount of total RNA (1800 ng) was employed in all RT-PCR. RT-PCR was performed with the OneStep RT-PCR kit (Qiagen) according to the manufacturer’s protocol for the transcripts with high GC content (which requires the use of Q-solution). Primers used for RT-PCR analyses are listed in Table 3.3 (The function is labeled as detection of gene transcript for all the primers). The RT-PCR program used for the detection of all transcripts was 50°C for 30 min, 95°C for 15 min, 28 cycles of 94°C for 30 s, 56°C for 30 s, and 72°C for 1 min, followed by a final elongation at 72°C for 10 min. Detection of possible contaminating DNA in RNA samples was accomplished by PCR with *Taq* DNA Polymerase (Invitrogen) using primers SGI6453 RT-PCR For/Rev. No bands were observed in the controls, indicating that all RT-PCR products correspond to amplification of RNA transcripts.

3.2.13 Mapping of SGI0214, SGI1617, SGI6453 and SGI6454 transcription start sites

The transcriptional start sites of SGI0214, SGI1617, SGI6453 and SGI6454 were identified with the 5’-RACE System for Rapid Amplification of cDNA Ends, Version 2.0 (Invitrogen). For

mapping the transcription start site of SGI0214, total mRNA isolated at 30 h as described above was used to synthesize cDNA according to RT-PCR result. The first-strand cDNA synthesis was performed using 20 pmol of primer SGI0214 GSP B and 2000 ng of RNA template according to the manufacturer's protocol for high GC-content transcripts. Two reactions were set up in order to get sufficient cDNA for the following steps. After cDNA synthesis, the reaction mixtures were treated with an RNase Mix to remove template RNA, pooled together and then purified by running through the S.N.A.P column. An oligo-dC tail was added to the purified cDNA using the TdT-tailing reaction. Amplification of dC-tailed cDNA was accomplished using a 2.5 μ l aliquot of the preceding reaction as template, the primer SGI0214 GSP C and abridge anchor primer (AAP, supplied with kit) and *Taq* DNA polymerase. The resulting polymerase chain reaction (PCR) product was used in a second PCR using the nested primer SGI0214 GSP D and universal amplification primer (UAP, supplied with kit). The first and second PCR products were purified and then ligated into the pGem-T easy vector and transformed into the *E. coli* strain DH5 α to yield pJS909 and pJS910. DNA sequencing of cloned inserts was performed by GENEWIZ (South Plainfield, NJ). The location of the transcription start site was corroborated through independent trials.

The details of mapping the transcriptional start sites of SGI1617, SGI6453 and SGI6454 are almost the same as described for SGI0214 with a few modifications. For SGI1617, because transcription level of the gene was low along the whole growth curve, 47 h mRNA sample was used as template and 6 reactions were set up together for first-strand cDNA synthesis. For both SGI6453 and SGI6454, 30 h mRNA sample was used as template. One reaction of cDNA synthesis was set up for SGI6453, while two reactions of cDNA synthesis were set up for SGI6454. All the corresponding primers and plasmids are listed in Table 3.3 and Table 3.2.

3.2.14 Heterologous expression of indolmycin biosynthetic gene cluster in *S. coelicolor*

An 11261-bp DNA fragment (seq0307), including the whole indolmycin biosynthetic gene cluster and the indolmycin high-resistant gene SGI6453, was cloned. Specifically, the whole cluster was assembled from three shorter fragments that were amplified from the *S. griseus* ATCC 12648 genomic DNA with Q5 High-Fidelity DNA Polymerase using primers Seq0307 A Clone For/Rev, Seq0307 Clone B For/Rev and Seq0307 Clone C For/Rev. The sizes of each smaller fragment are 5040-bp (A fragment), 6343-bp (B fragment) and 1901-bp (C fragment). Each PCR product, with engineered XbaI sites at both 5' end and 3' end, was ligated into pBluescript KS+ to yield pJS902 (containing A fragment), pJS903 (containing B fragment) and pJS904 (containing C fragment). A 6230-bp fragment was excised from pJS903 with NsiI and HindIII and ligated into vector pJS902 pretreated with the same enzyme to yield pJS905. Assembled A and B fragment was sub-cloned from pJS905 using XbaI and ligated into vector pSET152 (35) pretreated with XbaI to yield pJS906. C fragment was also sub-cloned from pJS904 using XbaI and ligated into vector pSET152 pretreated with XbaI to yield pJS907. Finally, the 9293-bp DNA fragment was sub-cloned from pJS906 with EcoRV and then ligated into vector pJS907 pretreated with then same enzyme EcoRV and then Shrimp Alkaline Phosphatase (rSAP) to yield pJS908. The desired plasmid harboring the locus of interest (pJS908) was selected by performing *E. coli* colony PCR using primers Seq0307 Colony PCR For/Rev and was further confirmed with restriction digests (BamHI, KpnI and XbaI respectively).

pJS906 was then transformed into ET12567/pUZ8002 and introduced into *S. coelicolor* M1154 (36,37) strain (yielding strain B1663) via conjugation, as previously described. Exconjugants were selected by apramycin resistance. Phenotypic analyses of *S. coelicolor* B1663 was accomplished via detection of indolmycin produced in NMMP and R3 liquid culture as described below. Vector pSET152 was also introduced into *S. coelicolor* M1154 separately to yield B1664 strain as a negative control.

3.2.15 Detection of indolmycin in liquid culture by LC-MS

Cultivation of *S. griseus* ATCC 12648 and *S. coelicolor* strains for the production of indolmycin were carried out as described previously (37). For the production of indolmycin in liquid culture, the preliminary experiments were performed in NMMP without PEG. 2×10^8 spores of *S. griseus* ATCC 12648 or the *S. coelicolor* strain were germinated in 500 μ l of 2 \times YT at 50°C for 15 min and then transferred into 250-ml flasks containing 50 ml of NMMP without PEG. The liquid culture were incubated with shaking (200 rpm) at 30°C for 48 h before the extraction of indolmycin. The optimized experiments with better yield of indolmycin were carried out in R3 medium. 2×10^8 spores of *S. griseus* ATCC 12648 or the *S. coelicolor* strains were germinated in 10 ml of 2 \times YT at 30°C for 8 h with 300 rpm shaking speed. Germlings were harvested by centrifugation and resuspended in 1 ml fresh R3 medium by vortexing. The germinated spores were inoculated in 50 ml R3 medium in a 250-ml flask and the resulting culture were incubated at 250 rpm and 30°C for 68 h.

A published procedure was adapted for the isolation and detection of indolmycin (13,38). For both condition described above, the entire culture (50 ml) was filtered to remove the mycelium. Culture filtrates then were adjusted to pH 3 with 4N HCl and extracted twice with 40 ml ethyl acetate. The organic phases were combined and extracted with 8 ml of a 10% (wt/vol) Na_2CO_3 solution and then dried over Na_2SO_4 . Solvent was removed under reduced pressure to leave residues, which was further dissolved in 100 μ l methanol for LC-MS analysis.

LC-MS analysis was performed with an Agilent 6530 accurate-mass Q-TOF LC/MS system operated in positive ion electrospray modes. LC analysis was performed with an Agilent 1260 Infinity HPLC system and Agilent ZORBAX Extend-C18 column (2.1 \times 50 mm). Elution was carried out at the flow rate of 0.3 ml/min with a gradient of mobile-phase A (water, 0.1% (v/v)

formic acid) and mobile-phase B (acetonitrile) (95:5 to 5:95, 0 to 9 min). An indolmycin standard compound acquired from C. W. Carter, Jr. was used as the control.

3.2.16 Bioinformatic Analyses of TrpRSs Sequence

TrpRS sequences were aligned using MUSCLE (39); the alignments were then analyzed for maximum probability trees using BEAST (40) and assembled and annotated using Treeannotator (40).

Table 3.1 Strains employed in this study

Strain	Genotype	Reference / Source
<i>S. griseus</i> ATCC 12648	Wild-type indolmycin producer	ATCC
<i>S. coelicolor</i>		
B725	M600 <i>trpRS1</i> (SCO3334):: <i>apr</i>	(12)
B728	M600 <i>trpRS2</i> (SCO4839):: <i>apr</i>	(12)
B1659	B725 - pJS896	This Study
B1660	B725 - pJS897	This Study
B1661	B725 - pJS898	This Study
B1662	B725 - pIJ10257	This Study
M1154	M145 Δact (SCO5071-5092), Δred (SCO5877-5898), Δcda (SCO3210-3249), Δcpk (SCO6269-6288), <i>rpoB</i> (C1298T, S433L), <i>rpsL</i> (A262G, K88E)	
B1663	M1154 - pJS907	This Study
B1664	M1154 - pSET152	This Study
<i>E. coli</i>		
DH5 α	F- $\phi lacZ \Delta M15 \Delta (lacZYA-argF)U169 recA1 endA1 hsdR17(r_k^-, m_k^-) phoA supE44 thi-1 gyrA96 relA1 \lambda^-$	Invitrogen
10-beta competent cell	$\Delta (ara-leu) 7697 araD139 fhuA \Delta lacX74 galK16 galE15 e14- \phi 80dlacZ \Delta M15 recA1 relA1 endA1 nupG rpsL$ (StrR) <i>rph spoT1</i> $\Delta (mrr-hsdRMS-mcrBC)$	New England BioLabs
ET12567	<i>dam dcm hsdS cat tet</i>	
BL21	F ⁻ <i>ompT hsdS_B(r_B^m_B) dcm gal</i>	Takara

Table 3.2 Plasmids employed in this study

Plasmid	Description	Vector Backbone	Reference/Source
pUZ8002	RP4 Derivative OriT, Kan ^r		
pBluescript KS+	pUC <i>ori</i> , MCS, Amp ^r		Agilent Technologies (Stratagene)
pIJ10257	<i>oriT</i> , $\Phi BT1 attB-int$, Hyg ^r , <i>ermEp</i> *	pMS81	(26)
pGem-T Easy	pUC-derived, <i>lacZ</i> , Amp ^r		Promega
pSET152	<i>oriT</i> , PUC18 <i>ori</i> , $\Phi C31 attB-int$, <i>lacZα</i> , Apra ^r		(35)
pETRP1B	pBR322 <i>ori</i> , f1 origin, <i>lacI</i> , MCS, Kan ^r		(30)
pGro7	pACYC <i>ori</i> , <i>groES/EL</i> , <i>araC</i> , <i>araB</i> , Cm ^r		Takara
pJS893	SGI0214 ORF with engineered NdeI and HindIII sites at each end, Amp ^r	pBluescript KS+	This Study

pJS894	SGI1617 ORF with engineered NdeI and HindIII sites at each end, Amp ^r	pBluescript KS+	This Study
pJS895	SGI6453 ORF with engineered NdeI and HindIII sites at each end, Amp ^r	pBluescript KS+	This Study
pJS896	<i>ermEp</i> *- SGI0214 ORF, Hyg ^r	pIJ10257	This Study
pJS897	<i>ermEp</i> *- SGI1617 ORF, Hyg ^r	pIJ10257	This Study
pJS898	<i>ermEp</i> *- SGI6453 ORF, Hyg ^r	pIJ10257	This Study
pJS899	SGI0214 ORF with engineered NdeI and HindIII sites at each end, Kan ^r	pETRP1B	This Study
pJS900	SGI1617 ORF with engineered NdeI and HindIII sites at each end, Kan ^r	pETRP1B	This Study
pJS901	SGI6453 ORF with engineered NdeI and HindIII sites at each end, Kan ^r	pETRP1B	This Study
pJS902	5040bp A fragment of indolmycin biosynthetic gene cluster (seq0307) with engineered XbaI sites at each end, Amp ^r	pBluescript KS+	This Study
pJS903	6343bp B fragment of indolmycin biosynthetic gene cluster (seq0307) with engineered XbaI sites at each end, Amp ^r	pBluescript KS+	This Study
pJS904	1901bp C fragment of indolmycin biosynthetic gene cluster (seq0307) with engineered XbaI sites at each end, Amp ^r	pBluescript KS+	This Study
pJS905	Assembled A and B fragment of seq0307, Amp ^r	pBluescript KS+	This Study
pJS906	Assembled A and B fragment of seq0307, Apra ^r	pSET152	This Study
pJS907	C fragment of seq0307, Apra ^r	pSET152	This Study
pJS908	Assembled A, B and C fragment, the whole sequence of seq0307, Apra ^r	pSET152	This Study
pJS909	SGI0214 5'-RACE product of Primer AAP (kit) and SGI0214 GSP C	pGem-T Easy	This Study
pJS910	SGI0214 5'-RACE product of Primer UAP (kit) and SGI0214 GSP D	pGem-T Easy	This Study
pJS911	SGI1617 5'-RACE product of Primer AAP (kit) and SGI1617 GSP C	pGem-T Easy	This Study
pJS912	SGI1617 5'-RACE product of Primer UAP (kit) and SGI1617 GSP E	pGem-T Easy	This Study
pJS913	SGI6453 5'-RACE product of Primer AAP (kit) and SGI6453 GSP C	pGem-T Easy	This Study
pJS914	SGI6453 5'-RACE product of Primer UAP (kit) and SGI6453 GSP D	pGem-T Easy	This Study
pJS915	SGI6454 5'-RACE product of Primer AAP (kit) and SGI6454 GSP D	pGem-T Easy	This Study
pJS916	SGI6454 5'-RACE product of Primer UAP (kit) and SGI6454 GSP E	pGem-T Easy	This Study

Table 3.3 Oligonucleotides employed in this study

Primer Name	Application/Function	Sequence
SGI0214 Clone For	Cloning of SGI0214 ORF	5' - <u>CATATGGCCTCCGACCTCCC</u> - 3' ^a
SGI0214 Clone Rev	Cloning of SGI0214 ORF	5' - <u>AAGCTTGCCGTGATCTCGCCAG</u> - 3' ^b
SGI1617 Clone For	Cloning of SGI1617 ORF	5' - <u>CATATGACCACCACCGCCCCGGCCGCCAG</u> - 3' ^a
SGI1617 Clone Rev	Cloning of SGI1617 ORF	5' - <u>AAGCTTGCCGCCGGTACCGATG</u> - 3' ^b
SGI6453 Clone For	Cloning of SGI6453ORF	5' - <u>CATATGATCAAGCTGTCTCGGGAATC</u> - 3' ^a
SGI6453 Clone Rev	Cloning of SGI6453 ORF	5' - <u>AAGCTTGGAGCTCGACCCGCTC</u> - 3' ^b
SGI0214 RT-PCR For	Detection of SGI0214 transcript	5' - CGCCTTCTACATGGTCGTG - 3'
SGI0214 RT-PCR Rev	Detection of SGI0214 transcript	5' - TGGAGAGGATCGTGAGCAG - 3'

SGI1617 RT-PCR For	Detection of SGI1617 transcript	5' - CATCCCCGACTACCAGGTC - 3'
SGI1617 RT-PCR Rev	Detection of SGI1617 transcript	5' - GGAGAGCGGAATGGAGTTG - 3'
SGI6453 RT-PCR For	Detection of SGI6453 transcript	5' - CGCATGGTGCAGTTCAAG - 3'
SGI6453 RT-PCR Rev	Detection of SGI6453 transcript	5' - CGCCAACCTCCTTGACGAC - 3'
SGI6454 RT-PCR For	Detection of SGI6454 transcript	5' - ATCGGGGCTCGCTATCTC - 3'
SGI6454 RT-PCR Rev	Detection of SGI6454 transcript	5' - AGCCGCGCAGTACTCATC - 3'
SGI6457 RT-PCR For	Detection of SGI6457 transcript	5' - AACAACTCTGACGCGCTACG - 3'
SGI6457 RT-PCR Rev	Detection of SGI6457 transcript	5' - AGACCCACCGTCAGGTTTC - 3'
SGI6460 RT-PCR For	Detection of SGI6460 transcript	5' - CGGTCCCAGCACTCTCAC - 3'
SGI6460 RT-PCR Rev	Detection of SGI6460 transcript	5' - GTGGTGAGGACGGTGGTC - 3'
SGI6455-6456 intergenic For	Detection of SGI6455 and SGI6456 cotranscription	5' - GTTCGCGCTCGACGAC - 3'
SGI6455-6456 intergenic Rev	Detection of SGI6455 and SGI6456 cotranscription	5' - TCGGAGCGGTCCAGATAC - 3'
SGI6456-6457 intergenic For	Detection of SGI6456 and SGI6457 cotranscription	5' - GTCCAGGTGGAGTTTGTGG - 3'
SGI6456-6457 intergenic Rev	Detection of SGI6456 and SGI6457 cotranscription	5' - CCATCGGCGAGATTGATG - 3'
SGI6459-6460 intergenic For	Detection of SGI6459 and SGI6460 cotranscription	5' - GCGGAGAGGAGTGGAAACC - 3'
SGI6459-6460 intergenic Rev	Detection of SGI6459 and SGI6460 cotranscription	5' - TCCGCATCCGTAGTCCAG - 3'
SGI6460-6461 intergenic For	Detection of SGI6460 and SGI6461 cotranscription	5' - GAGGTCACGGTCGAGCAC - 3'
SGI6460-6461 intergenic Rev	Detection of SGI6460 and SGI6461 cotranscription	5' - CAGGGAGAGCAGGAGCTG - 3'
Seq0307 A Clone For	Cloning of A fragment of indolmycin biosynthetic gene cluster	5' - <u>TCTAGAGATGCCGTT</u> CATAGACGGTC - 3' ^c
Seq0307 A Clone Rev	Cloning of A fragment of indolmycin biosynthetic gene cluster	5' - <u>TCTAGACCTGTCTCGCACCCTT</u> CAC - 3' ^c
Seq0307 B Clone For	Cloning of B fragment of indolmycin biosynthetic gene cluster	5' - <u>TCTAGAAAATATTCGCCCCAGGATTC</u> - 3' ^c
Seq0307 B Clone Rev	Cloning of B fragment of indolmycin biosynthetic gene cluster	5' - <u>TCTAGAACATGCCGATAGGTGACGAG</u> - 3' ^c
Seq0307 C Clone For	Cloning of C fragment of indolmycin biosynthetic gene cluster	5' - <u>TCTAGAGACTGGGAGACATCGGAGTC</u> - 3' ^c
Seq0307 C Clone Rev	Cloning of C fragment of indolmycin biosynthetic gene cluster	5' - TCTAGACACCCAACACTATCCCGTG - 3' ^c
Seq0307 Colony PCR For	Selection of correct plasmid pJS908	5' - CACTGCCCCGGACTACTCG - 3'
Seq0307 Colony PCR Rev	Selection of correct plasmid pJS908	5' - CAGGGAGAGCAGGAGCTG - 3'
SGI0214 For	5' RACE, SGI0214	5' - CCCGCACCTGAGACAATG - 3'
SGI0214 GSP B	5' RACE, SGI0214	5' - ACGTGGCTCTGGACGAAG - 3'
SGI0214 GSP C	5' RACE, SGI0214	5' - GTGTTGGCCCCGAGCTC - 3'

SGI0214 GSP D	5' RACE, SGI0214	5' - GTGATCGCGTGCAGGTC - 3'
SGI1617 For	5' RACE, SGI1617	5' - GTCCCTCACGCACCACTC - 3'
SGI1617 GSP A	5' RACE, SGI1617	5' - ACCGAGGGCCAGGTAGTC - 3'
SGI1617 GSP C	5' RACE, SGI1617	5' - GTCCACGCCGAGGTTCTG - 3'
SGI1617 GSP E	5' RACE, SGI1617	5' - CGGTGAGGACCCGGAAC - 3'
SGI6453 For	5' RACE, SGI6453	5' - TCAAGCTGTCGGGAATCAC - 3'
SGI6453 GSP B	5' RACE, SGI6453	5' - CGTCGGACTTCTCCTTGAAC - 3'
SGI6453 GSP C	5' RACE, SGI6453	5' - AGGTCCGACTGCACGAAC - 3'
SGI6453 GSP D	5' RACE, SGI6453	5' - AGCTGGTGCTCGGTCAAC - 3'
SGI6454 For	5' RACE, SGI6454	5' - CCGGTCAAATCTCCATC - 3'
SGI6454 GSP A	5' RACE, SGI6454	5' - GGGGTGGTCCTGAGTGAGAG - 3'
SGI6454 GSP D	5' RACE, SGI6454	5' - ACGCGCAGGAGTCTGG - 3'
SGI6454 GSP E	5' RACE, SGI6454	5' - GCGACTCCAGCAGATTGAC - 3'

- The engineered NdeI site is underlined.
- The engineered HindIII site is underlined.
- The engineered XbaI site is underlined.

3.3 Results

3.3.1 The genome of *S. griseus* ATCC 12648 and secondary metabolites in *S. griseus* ATCC 12648

To enable identification of the indolmycin biosynthetic gene cluster, we contracted sequencing of the genome of the indolmycin producer *S. griseus* ATCC 12648. For sequencing, a 500-bp PCR-free library (1,495 Mb raw data) was constructed, followed by low quality data elimination to generate total 1,007 Mb of clean data, which provide an average 28.44× coverage of the genome. The elimination included removal of 0.754% adapter and 0.084% duplication in raw data. For assembly, key parameter K was set at 57 according to the optimal assembly strategy. 339 contigs and 63 scaffolds were constructed after the final assembly. The total genome size is 7,147,681 bp with a G+C content of 72.62%. This Whole Genome Shotgun project has been deposited at DDBJ/ENA/GenBank under the accession LVHX00000000. The version described in this paper is version LVHX01000000. Genome annotation by NCBI Prokaryotic Genome Annotation Pipeline (PGAP) is still in process. The raw reads have also been submitted to the Sequence Read Archive (SRA) database under the accession SRP072367.

The genome of *S. griseus* ATCC 12648 was analyzed with regard to secondary metabolism using the software tool antiSMASH (41). This bioinformatic tool is used to identify multimodular enzymes that catalyze the biosynthesis of polyketides and non-ribosomal peptides. A total of 31 gene clusters for putative biosynthetic secondary metabolites were predicted in *S. griseus* ATCC 12648. Information about all the gene clusters, how similar they are to the current known cluster as well as possible types of the secondary metabolites are summarized in Table 3.4. The production of these hypothetical secondary metabolites was not investigated in this thesis work. Among the predicted gene clusters, some of them are highly conserved in the *Streptomyces* genus (e.g., *amf* cluster, PTMs biosynthetic gene cluster and nonactin biosynthetic gene cluster). Curiously, many of them appear to be unique to *S. griseus* ATCC 12648, as no similar known cluster was identified via antiSMASH analysis. We note that the indolmycin biosynthetic gene cluster was not predicted by antiSMASH because it is not a non-ribosomal peptide or polyketide.

Table 3.4 Synthesis of secondary metabolites in *Streptomyces griseus* ATCC 12648

Type	Gene cluster	Most similar known cluster (similarity)
Melanin	SGI0234-SGI0236	
Melanin	SGI5235-SGI5250	Melanin biosynthetic gene cluster (100%)
Lantipeptide	SGI0489-SGI0501	AmfS ^d _biosynthetic gene cluster (100%)
Lantipeptide	SGI2129-SGI2145	
Lantipeptide	SGI4950-SGI4984	
Thiopeptide-Lantipeptide	SGI5664-SGI5685	
Ectoine	SGI1228-SGI1240	Ectoine biosynthesis gene cluster (50%)
Butyrolactone-Ectoine	SGI1733-SGI1752	Skyllamycin biosynthetic gene cluster (8%)
Butyrolactone	SGI4904-SGI4925	A40926 ^e biosynthetic gene cluster (3%)
Terpene	SGI2513-SGI2523	Steffimycin biosynthetic gene cluster (13%)
Terpene	SGI2524-SGI2547	Isorenieratene biosynthetic gene cluster (100%)
Terpene	SGI2856-SGI2878	A-500359s ^f biosynthetic gene cluster (5%)
Terpene	SGI3210-SGI3233	Hopene biosynthetic gene cluster (61%)
Terpene	SGI6165-SGI6183	Arylomycin biosynthetic gene cluster (33%)
Terpene	SGI4845-SGI4864	
NRPS ^a	SGI3278-SGI3284	
NRPS	SGI4791-SGI4827	Griseobactin biosynthetic gene cluster (94%)
NRPS	SGI5500-SGI5530	Coelichelin biosynthetic gene cluster (81%)
NRPS	SGI5731-SGI5735	
NRPS	SGI5823-SGI5863	C-1027 ^g biosynthetic gene cluster (57%)
NRPS-PKS ^b (type 1) mix	SGI3326-SGI3359	SGR_PTMs ^h biosynthetic gene cluster (100%)

NRPS-PKS (type 1) mix	SGI0401-SGI0412	Leinamycin biosynthetic gene cluster (4%)
PKS (type 2)	SGI2975-SGI2991	Nonactin biosynthetic gene cluster (85%)
PKS (type 3)	SGI5429-SGI5471	Herboxidiene biosynthetic gene cluster (6%)
PKS (type 3)	SGI5540-SGI5574	Alkylresorcinol biosynthetic gene cluster (100%)
Bacteriocin	SGI3809-SGI3832	
Bacteriocin	SGI5373-SGI5383	Tetronasin biosynthetic gene cluster (3%)
Siderophore	SGI5793-SGI5801	
Otherks ^c	SGI3364-SGI3375	
Otherks-Butyrolactone	SGI3376-SGI3389	SF2575 ⁱ biosynthetic gene cluster (4%)
Other	SGI6395-SGI6409	

- a. NRPS, nonribosomal peptide synthetases.
- b. PKS, polyketide synthetase.
- c. Otherks, other types of PKS cluster.
- d. Amf, aerial-mycelium formation.
- e. A40926, a new glycopeptide antibiotic.
- f. A-500359s, bacterial phosphor-N-acetylmuramylpentapeptide-translocase inhibitor.
- g. C-1027, an enediyne antitumor antibiotic.
- h. PTMs, polycyclic tetramate macrolactams.
- i. SF2575, an anticancer tetracycline polyketide.

3.3.2 Genetic analyses of the three tryptophanyl-tRNA synthetases (TrpRSs) from *S. griseus*

ATCC 12648

Bioinformatic analyses of the *S. griseus* ATCC 12648 genome did not reveal the identities of the indolmycin biosynthetic genes. Resistance and biosynthetic genes for a specific secondary metabolite are usually clustered in bacteria (24,42). We proposed that the indolmycin biosynthetic genes could be identified by virtue of their proximity to an indolmycin resistance gene. Resistance to indolmycin in *Streptomyces* bacteria is linked to auxiliary isoforms of tryptophanyl-tRNA synthetase (11-13). There are two well characterized indolmycin resistant TrpRS enzymes – gene product of SCO3334 in *S. coelicolor* and gene product of SGR3809 in *S. griseus* NBRC 13350. In our study, bioinformatic analysis of the resulting draft genome data revealed that *Streptomyces griseus* ATCC 12648 carries three tryptophanyl-tRNA synthetase (TrpRS) genes (SGI0214, SGI1617 and SGI6453); the corresponding gene products will be recognized as SGI_0214, SGI_1617 and SGI_6453, respectively. The TrpRS encoded by SGI0214 is 97% identical in amino

acid sequence to the indolmycin-sensitive enzyme encoded by SGR2702 in *S. griseus* NBRC 13350, while TrpRS encoded by SGI1617 is 88% identical to indolmycin-resistant enzyme encoded by SGR3809. This bioinformatics analysis yielded the hypothesis that SGI0214 encodes an indolmycin-sensitive enzyme and SGI1617 encodes an indolmycin-resistant enzyme. Curiously, SGI6453 encoded an enzyme that shows low similarity to both of the known TrpRSs (43% identical AA sequence to SGR2702 and 32% identical AA sequence to SGR3809). We hypothesized SGI6453 encoded an indolmycin-resistant tryptophanyl-tRNA synthetase that could be clustered with the biosynthetic genes.

Table 3.5 Comparison of indolmycin sensitivity of *S. coelicolor* B725 upon the expression of TrpRS genes from *S. griseus* ATCC 12648

Entry	ORF ^a expressed in B725 (indolmycin-sensitive <i>S.coelicolor</i>)	MIC ^b ($\mu\text{g/ml}$)	Fold increase in MIC ^c
1	<i>S. coelicolor</i> B725 (<i>trpRS1</i> null strain)	0.25	1
2	SGI0214	0.25	1
3	SGI1617	15	60
4	SGI6453	>250	>1000

- All ORFs were introduced into strain B725 under the *ermE** promoter for equalized expression *in vivo* (see Materials and Methods).
- The MIC is the concentration of indolmycin sufficient to inhibit the growth of each strain after 48 h on Difco nutrient agar (see Materials and Methods)
- Fold increase in MIC relative to that of indolmycin-sensitive *S. coelicolor* strain B725 (MIC = 0.25 $\mu\text{g/ml}$).

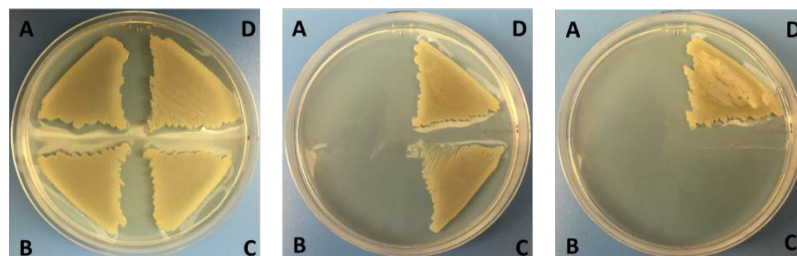


Figure 3.3 Upon the expression of different *trpRS* genes (SGI0214, SGI1617 and SGI6453), the constructed *S. coelicolor* strains showed different levels of indolmycin resistance relative to the parental, indolmycin-sensitive strain *S. coelicolor* B725 ($\Delta\text{trpRS1}::\text{apr}$). All the tested strains: A) *S. coelicolor* B725, B) *S.*

coelicolor B725 with SGI0214, C) *S. coelicolor* B725 with SGI1617, D) *S. coelicolor* B725 with SGI6453. The bacteria were grown for 2 days at 30°C on Difco nutrient agar medium containing various concentrations of indolmycin. From left to right, A) 0 µg/ml, B) 7.5 µg/ml C) 100 µg/ml. SGI0214 encodes a sensitive TrpRS, SGI1617 confers mild resistance to indolmycin and SGI6453 confers strong resistance to indolmycin.

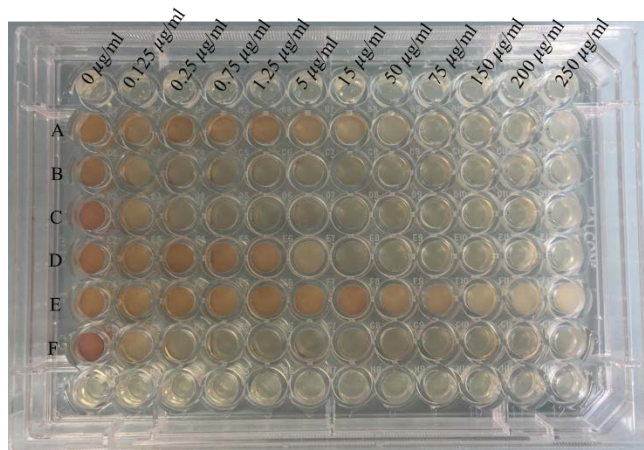


Figure 3.4 The minimum inhibitory concentration (MIC) assay of indolmycin for *S. coelicolor* B725 and constructed strains. The bacteria were grown for 2 days at 30°C on Difco nutrient agar medium containing various concentrations of indolmycin. All the tested strains: (A) *S. coelicolor* M600 WT, (B) *S. coelicolor* B725 (*trpRS1* null strain), (C) *S. coelicolor* B725 with SGI0214, (D) *S. coelicolor* B725 with SGI1617, (E) *S. coelicolor* B725 with SGI6453 (F) *S. coelicolor* with vector pIJ10257. The concentration of indolmycin increased from left to right. They are 0, 0.125, 0.25, 0.75, 1.25, 5, 15, 50, 75, 150, 200 and 250 µg/ml respectively and indicated on figure.

To test our hypothesis that SGI1617 and SGI6453 are antibiotic resistance genes, each gene was heterologously expressed in *Streptomyces coelicolor* B725, the indolmycin-sensitive *trpRS1* null strain. As a control, we also expressed SGI0214, which was homologous to indolmycin-sensitive *trpRS* genes. In detail, the genes were cloned into pIJ10257 for expression under the control of the constitutive *ermE** promoter and then individually introduced into the *S. coelicolor trpRS1* null strain. Then, we examined the *in vivo* indolmycin susceptibility of the constructed strains. The

constructed strains showed different levels of indolmycin resistance relative to parental strain *S. coelicolor* B725 (Fig. 3.3). We found that there is no distinct difference of the indolmycin MIC for *S. coelicolor* B725 upon the expression of SGI0214. In contrast, upon expression of SGI1617, the indolmycin MIC increased from 0.25 to about 15 $\mu\text{g/ml}$, which is consistent with the observation of MIC elevation upon the expression of SGR3809 in the same genetic background. Intriguingly, upon the expression of SGI6453, *S. coelicolor* B725 exhibited robust growth and resistance at concentrations as high as 250 $\mu\text{g/ml}$ indolmycin (Table 3.5, Fig. 3.4), which is more than 1,000-fold higher than the MIC of the parent strain. These findings were consistent with predictions about the indolmycin susceptibility of SGI_0214 and SGI_1617. In addition, they established that SGI_6453 was a novel and highly indolmycin-resistant tryptophanyl-tRNA synthetase.

3.3.3 Biochemical characterization of the three trpRSs from *S. griseus* ATCC 12648

The exceedingly high degree of indolmycin resistance conferred by heterologous expression of SGI6453 raised many questions about the activity and structure of its product. Accordingly, we initiated a series of biochemical studies of the product of SGI6453. In parallel investigations, we also studied the products of SGI0214 and SGI1617. This work was carried out in collaboration with Prof. Charlie W. Carter, Jr. at UNC. Chapel Hill. The ORF of each gene (SGI0214, SGI1617 and SGI6453) was cloned into the expression vector pETRP1B (30) and the three TrpRS variants were co-expressed in *E. coli* with GroEL/ES chaperones by auto-induction. The protein were purified to homogeneity via nickel affinity chromatography. Michaelis-Menten experiments were then conducted at increasing tryptophan concentrations in the presence of various indolmycin concentrations. After fitting results to a competitive inhibition model using non-linear regression, we were able to determine kinetic parameters of the three TrpRS variants (Table 3.6).

Table 3.6 The kinetic parameters^a of the three TrpRSs from *S. griseus* ATCC 12648

	SGI_0214	SGI_1617	SGI_6453
k_{cat} (s^{-1})	9.11 ± 0.49	3.13 ± 0.06	8.59 ± 0.40
K_M (μM)	63.17 ± 11.67	133.14 ± 12.76	570.77 ± 83.05
K_i (μM)	0.08 ± 0.02	2.70 ± 0.31	144.88 ± 28.98
K_M/K_i	789.63	49.31	3.94
k_{cat}/K_M ($\text{s}^{-1} \cdot \text{M}^{-1}$)	1.44×10^5	2.35×10^4	1.50×10^4

a. The catalytic efficiency, tryptophan binding affinity, and indolmycin inhibition of TrpRS variants were determined by steady-state kinetics.

Consistent with the *in vivo* observation that gene SGI0214 encodes an indolmycin sensitive TrpRS, the enzyme SGI_0214 has a relative small $K_{i, \text{indolmycin}}$, $0.08 \mu\text{M}$, and binds indolmycin ~ 780 times tighter than tryptophan. The K_i of SGI_1617 ($2.7 \mu\text{M}$) was greater than SGI_0214 ($0.08 \mu\text{M}$) by about 34-fold and the K_i of SGI_6453 ($144.7 \mu\text{M}$) was greater than SGI_0214 dramatically by about 1800-fold, suggesting that gene SGI1617 and SGI6453 confer different levels of indolmycin resistance. According to rate constants, both SGI_0214 and SGI_6453 were equally active, but the K_M of SGI_6453 ($570.8 \mu\text{M}$) is one order of magnitude higher than SGI_0214 ($63.2 \mu\text{M}$). The results indicate that SGI_0214 is a more efficient TrpRS catalyzing tryptophanyl-tRNA formation *in vitro*. Compared to SGI_0214 and SGI_6453, the enzyme SGI_1617 is less active and has intermediated level of tryptophan binding affinity and catalytic efficiency.

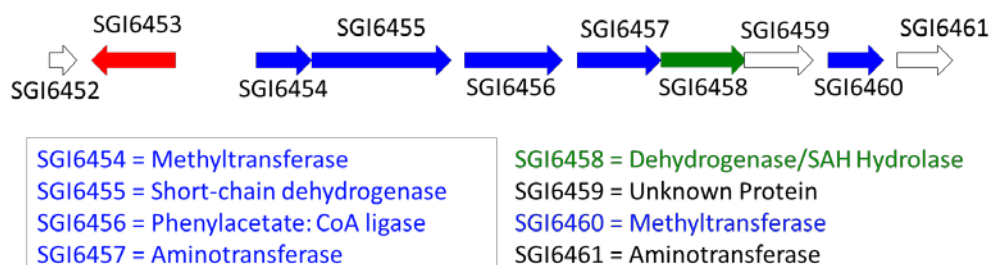
3.3.4 Identification of the putative indolmycin biosynthetic gene cluster

Our discovery of SGI6453 as an indolmycin resistance determinant was made more interesting by our observation that it was physically clustered with genes encoding enzymes that could catalyze the hypothetical reactions in the indolmycin biosynthesis (Fig 3.5).

When considering the regulation of indolmycin biosynthesis, it has been reported that many of the pathway-specific regulatory proteins that control secondary metabolism in streptomycetes belong to the SARP family (43,44). They have been found associated with secondary metabolic gene clusters that encode aromatic polyketides (45-47), ribosomally and non-ribosomally synthesized peptides (48,49), undecylprodiginines (50), Type I polyketides (51-53), β -lactams (54) and azoxy

compounds (55). Interestingly, we did not find any transcription factors in the cluster. Given the identification of the putative indolmycin biosynthetic gene cluster, we set about a series of experiments to examine the expression and function of the hypothetical biosynthesis genes.

A. Organization of the putative indolmycin biosynthetic gene cluster



B. Reaction enzymes of the indolmycin biosynthesis

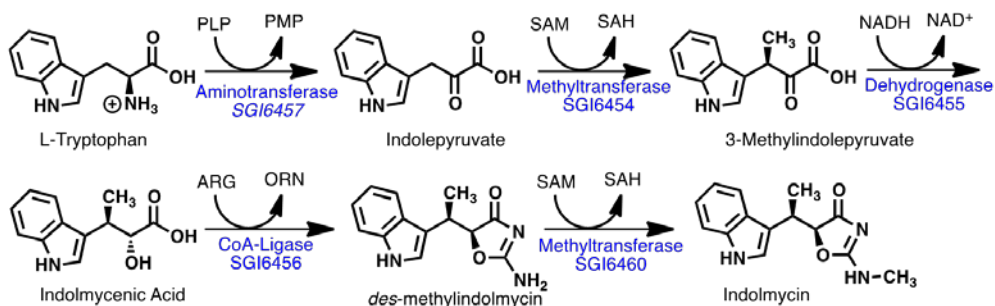


Figure 3.5 Gene organization of the putative indolmycin biosynthetic gene cluster and enzymes that catalyze the hypothetical reactions in the indolmycin biosynthesis.

3.3.5 Transcription profiles of the indolmycin biosynthetic gene cluster and the three *trpRS* genes during the growth of *S. griseus* ATCC 12648

Initially, reverse transcription (RT) – PCR was utilized to access the transcription profiles of the indolmycin biosynthetic and resistance genes. Growth curve of *S. griseus* ATCC 12648 was obtained by inoculating pre-germinated spores into NMMP medium. Similar to what has been reported for *S. coelicolor* (56), the growth curve could be separated into four components: rapid

growth 1 (RG1), transition (T), rapid growth 2 (RG2) and stationary (S) (Fig. 3.6A). Along the growth curve, 13, 18, 24, 30, 36 and 47 h six time points that cover all the four distinct stages were chosen for total mRNA isolation. Three genes, SGI6454, SGI6457 and SGI6460 were selected for RT-PCR analysis as they are physically located in different positions in the gene cluster. We found that transcription of the three genes was influenced by growth phase and the transcripts increased in abundance in the middle of RG1 phase (18 h) (Fig. 3.6B). The observations of indolmycin biosynthetic genes transcription are consistent with the fact that antibiotic production in streptomycetes is generally growth phase-dependent (56).

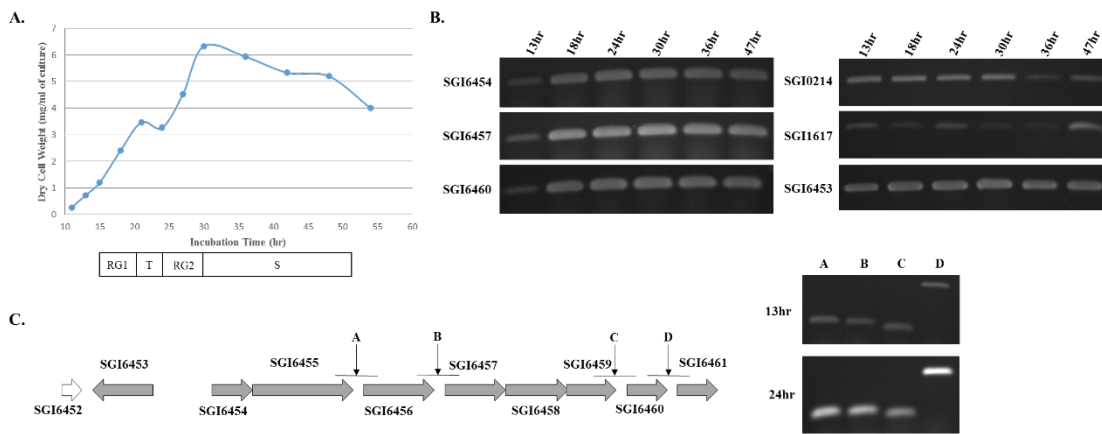


Figure 3.6 Expression of indolmycin biosynthetic genes and *trpRS* genes during the growth of *S. griseus* ATCC 12648. **(A)** The growth curve of *S. griseus* ATCC 12648 wild-type. The bacteria was grown as shaken liquid culture in NMMP medium at 30°C. An initial period of rapid growth (RG1) beginning around 11 h after inoculation of liquid culture was followed by a brief growth slowdown starting around 21 h after inoculation (transition phase). After 3 hours, culture resumed another rapid growth phase (RG2) before entering stationary (S) phase. **(B)** Transcription profile of indolmycin biosynthetic genes and TrpRS genes during growth of *S. griseus* ATCC 12648. RT-PCR was used to detect the transcription of 3 indolmycin biosynthetic genes (SGI6454, SGI6457 and SGI6460) and all the TrpRS genes (SGI0214, SGI1617 and SGI6453). Six timepoints were chosen according to the growth curve and they were indicated above the figure. Expression of indolmycin biosynthetic genes were growth-phase dependent. For *trpRS* genes, transcription of SGI0214 and SGI1617 were influenced by growth phase while SGI6453 was constitutively

transcribed. (C) Four intergenic regions within the indolmycin biosynthetic gene cluster. RT-PCR using 13 h and 24 h RNA sample were performed to detect the transcription of DNA fragments spanning the whole intergenic regions. The presence of transcript of all the intergenic regions indicated that the biosynthetic gene cluster is transcribed as a single unit.

Though all the 8 functionally related genes within the cluster are arranged contiguously in one orientation on the chromosomes of *S. griseus* ATCC 12648, it was unclear whether this cluster is one polycistronic operon as there are four modestly sized intergenic regions within the cluster. To test whether these intergenic regions contained transcription terminators or promoters, RT-PCR with 13 h and 24 h RNA sample was performed using primers whose PCR product spanned the whole intergenic region. The presence of RT-PCR product bands for all the intergenic regions indicated that all the indolmycin biosynthetic genes are co-transcribed as one polycistronic message (Fig. 3.6C).

In parallel with transcription profiling of the presumptive biosynthetic genes, we studied expression of the tryptophanyl-tRNA synthetase genes. We found that the transcription of SGI0214 and SGI1617 were influenced by growth phase while SGI6453 was constitutively transcribed robustly (Fig. 3.6B).

3.3.6 Transcript mapping of the indolmycin biosynthetic gene cluster and tryptophanyl-tRNA synthetase genes

To further investigate the transcription of the indolmycin biosynthetic and resistance genes, we used 5' rapid amplification of cDNA ends (5'-RACE) experiments to map the transcription start sites of SGI6453, SGI1617, SGI0214, and SGI6454 (the first gene in the presumptive biosynthetic operon). For the indolmycin biosynthetic gene cluster, curiously, the transcription start site (TSS) was identified at -425 bp relative to its predicted translational initiation site (TIS). A mRNA transcript with such a long 5' - untranslated region (5' - UTR) is rare in streptomycetes. Generally,

long leader regions (>150 nt) of transcripts may be a target for posttranscriptional regulation or carry cis encoded regulatory RNA elements such as riboswitches. To test the presence of any regulatory possibility, an RNA secondary structure prediction tool, MFOLD (57) was firstly applied to the 425-bp non-coding sequence in front of the indolmycin biosynthetic gene cluster. The prediction did reveal 14 mutually exclusive secondary structure, however, no conserved sequence for ribosome-mediated attenuator or T-box antiterminator were detected in the region. The sequence was then analyzed by MATCHTM (58) and PROMO (59,60), both of which are the tools for searching transcription factor binding sites in DNA sequences. The analyses revealed several transcription factor binding sites, none of which was further proved to be a promising transcriptional binding site for streptomycetes.

We found that the transcription start site of SGI0214 coincided with the ATG translation initiation site of the open reading frame (Fig. 3.7A). This observation is consistent with our group's previous report that its homolog *trpRS2* in *S. coelicolor* is transcribed as a leaderless message (12,61). Leaderless transcripts lack Shine-Dalgarno sequence for translation initiation, but transcriptional signals such as Pribnow box (Bacteria) or TATA box (Archaea) have been found upstream of the TIS (62-66). Recently, Huaiqiu Zhu *et al.* developed an algorithm, which is statistically validated, to classify the initiation regulatory signals upstream translation start sites into SD-like, TA-like (which resemble the Pribnow or TATA box) and atypical signals (66,67). They found that 10 out of 13 leaderless genes documented for *S. coelicolor* contain TA-like signal upstream to their TISs (67). Based on this statistical study, we searched the upstream sequence of SGI0214 TIS and did find the TA-like signal (Fig. 3.7A).

The transcription of SGI1617 was mapped to two sites 194-bp and 195-bp upstream of its predicted TIS, which raised the possibility that the gene is transcriptionally regulated via a *cis*-acting RNA regulatory structure. We previously reported that genes encoding auxiliary indolmycin-resistant tryptophanyl-tRNA synthetases in other species of *Streptomyces* bacteria are transcribed differently

(13). Specifically, the transcription of SGR3809 in *S. griseus* NBRC 13350 and SGI1617 was not influenced by indolmycin, whereas the transcription of *trpRS1* in *S. coelicolor* and SAV4725 in *S. avermitilis* was induced by indolmycin via the regulatory element: ribosome-mediated transcriptional attenuator located within the 5'-UTR (13). Sequence alignments (ClustalX2) (68) revealed a pseudo *trp*-RNA-binding attenuator with similar structure organization (Fig. 3.7B).

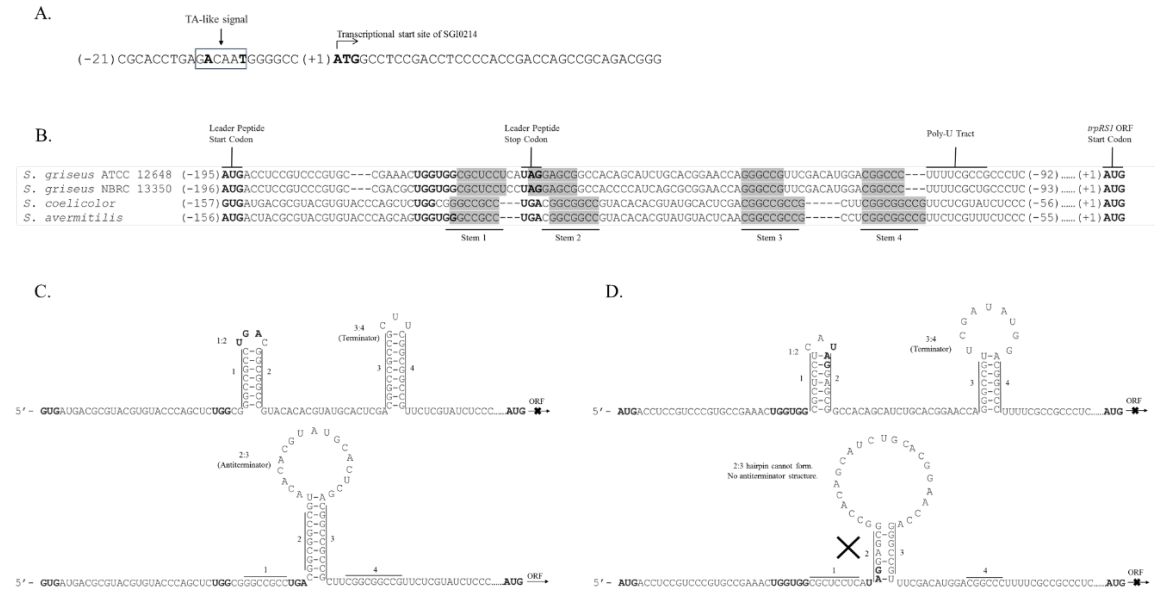


Figure 3.7 (A) Mapping transcription start site of the SGI0214 locus from *S. griseus* ATCC 12648. The transcription start site was mapped at 0 bp (indicated by arrow on figure), which is coincide with the translation initiation site (bold). TA-like signal was circle in box with nucleotide A and T in bold. (B) Sequence alignment of the regions upstream of SGI1617 and its orthologs in various *Streptomyces* species. The upstream region of SGI1617 is highly homologous to the upstream region of SGR3809 (*S. griseus* NBRC 13350), while its similarities to the upstream region of SCO3334 (*S. coelicolor*) and SAV4725 (*S. avermitilis*) are relative low. The identified sequences of the dyads of symmetry are shaded grey. The start codon (AUG/GUG), tryptophan codon (UGG) and the stop codon (UGA) of the leader peptide ORFs are in bold. The translation initiation sites (AUG) are also in bold. (C) Ribosome-mediated attenuator identified in *S. coelicolor*. The 157-nucleotide leader of SCO3334 transcript forms two mutually exclusive secondary structures. The top one is proposed terminator structure and the bottom one is proposed antiterminator structure. (D) RNA secondary structure formed by the 195-nucleotide leader of SGI1617 transcript. The top

one is similar to proposed terminator structure in *S. coelicolor*. In SGI1617, the leader region probably cannot stably form a hairpin because fewer nucleotides constitute the 2/3 stem (relative to the *S. coelicolor* hairpin).

Within the leader sequence of SGI1617 and SGR3809, there are ribosome-mediated attenuator elements including two distinct dyads of symmetry (each of them forms a hairpin), a small ORF containing two *trp* codons positioned adjacent to sequence 1 of the upstream dyad and a poly-U tract. As the 3:4 hairpin is G+C rich and immediately upstream of the poly-U tract, it resembles a *rho*-independent transcription termination (Fig. 3.7C). But different from the cases in *S. coelicolor* and *S. avermitilis*, stem2 and stem3 within the leader region of SGI1617 and SGR3809 cannot hybridize to form 2:3 hairpin and convert the structure to antiterminator from terminator. The absence of a regulatory structure was corroborated by our finding that the transcription of SGI1617 was not influenced by indolmycin (Fig. 3.7D).

Finally, the transcription of SGI6453 was located at two sites as well, -147 bp and -148 bp relative to its predicted TIS. The same bioinformatics analyses was performed to the 148 bp 5' UTR of SGI6453, but no conserved *cis*-encoded regulatory RNA element or regulator protein binding site was identified within the region. This prediction result was consistent with the observation that SGI6453 was constitutive transcribed robustly during the growth of *S. griseus* ATCC 12648 in liquid culture.

3.3.7 Heterologous production of indolmycin in *S. coelicolor*

Bioinformatic analyses of the genes physically clustered with the indolmycin resistance gene SGI6453 revealed multiple enzymes whose homologs catalyze reactions analogous to those in the hypothetical indolmycin biosynthetic pathway (Fig. 3.5). Accordingly, we predicted that genes SGI6454-6461 constituted the biosynthetic locus. To test this prediction, we amplified, assembled, and cloned the entire biosynthetic cluster and the resistance gene (an 11261-bp DNA fragment) into the integrative plasmid pSET152. Via conjugation, the constructed plasmid was transferred to

the heterologous expression host *S. coelicolor* M1154, whose genome was engineered specifically for secondary metabolite gene cluster heterologous expression. The constructed strain was named as B1663. The exconjugants were found to produce indolmycin. The antibiotic was present in the supernatant of shaken liquid cultures of *S. coelicolor* B1663 in both NMMP and R3 media 68 hours of growth. R3 was the optimal media with an indolmycin yield of 2 mg/L of culture. The production of indolmycin was not detected in the control strain *S. coelicolor* M1154 with vector pSET152 (Fig. 3.8). This observation confirmed our hypothesis and we genetically proved the biosynthesis of indolmycin.

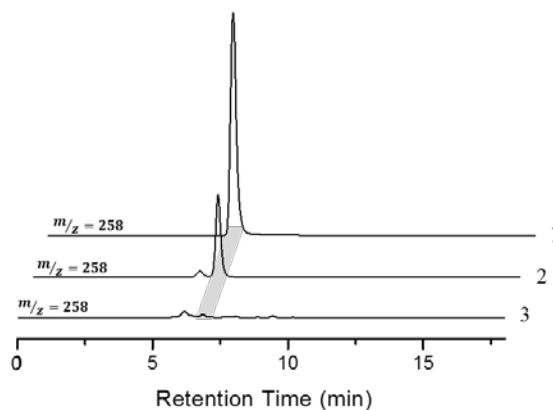


Figure 3.8 Heterologous production of indolmycin by *S. coelicolor* M1154. LC-MS was used to detect indolmycin. Trace 1 is indolmycin standard. Trace 2 is extracted ion chromatogram (EIC) from LC-MS analysis for isolation from liquid culture supernatant of *S. coelicolor* M1154 with pSET152 carrying indolmycin biosynthetic gene cluster. Trace 3 is EIC for isolation from liquid culture supernatant of *S. coelicolor* M1154 with only vector pSET152.

3.4 Discussion

The bioactivity and peculiarity of the indolmycin molecule make it an appealing antibiotic to study. However, the biosynthesis of indolmycin has been a mystery for decades. In this work, we have

used a combination of genetic and biochemical approaches to characterize indolmycin biosynthesis and resistance in *S. griseus* ATCC 12648. Benefitting from the ultra-fast and low-cost DNA sequencing technology in the post genomic era, we contracted sequencing the genome of indolmycin producer *S. griseus* ATCC 12648. However, even with the genome in hand, the fact that indolmycin is not a non-ribosomal peptide or polyketide rendered powerful bioinformatics tools for analyses of secondary metabolism futile in the search for the biosynthetic genes. Our strategy for identifying the biosynthetic genes was based on the common scenario wherein the biosynthetic genes are clustered together in a specific locus and this locus is typically adjacent to the gene conferring resistance to the cognate antibiotic (19).

As a structural analog of tryptophan, indolmycin displays inhibition for TrpRS. Generally, auxiliary aminoacyl-tRNA synthetase genes are highly correlated with antibiotic resistance (11-14,69). Through genome mining, three TrpRS gene (SGI0214, SGI1617 and SGI6453) were identified in indolmycin producer strain, *S. griseus* ATCC 12648. Our genetic and biochemical analyses indicated that one gene (SGI0214) encodes an indolmycin-sensitive TrpRS (which is also the primary TrpRS catalyzing tryptophanyl-tRNA formation), while the other two (SGI1617 and SGI6453) encode indolmycin-resistant TrpRSs.

A comparative analysis of the available streptomycete genome sequence revealed that several species possess TrpRSs that are homologous to SGI_0214 and SGI_1617 (13), while homologs of SGI_6453 with more than 80% similarity were only identified in two isolated *Streptomyces* species (*Streptomyces* sp. AVP053U2, 99% and *Streptomyces* NRRL. F-5135, 96%) and *Kutzemia albida* (83%) via bioinformatics analysis. As these two unnamed isolates have not been characterized using traditional methods, the reason why these two species harbor a high-level indolmycin-resistant gene is still not clear. However, it is widely known that *Streptomyces* species often harbor genes conferring resistance to antibiotics that they do not produce (70). Additionally, there is

considerable evidence indicating transfer of resistance elements between streptomycetes or from streptomycetes to other bacteria genera (70-72).

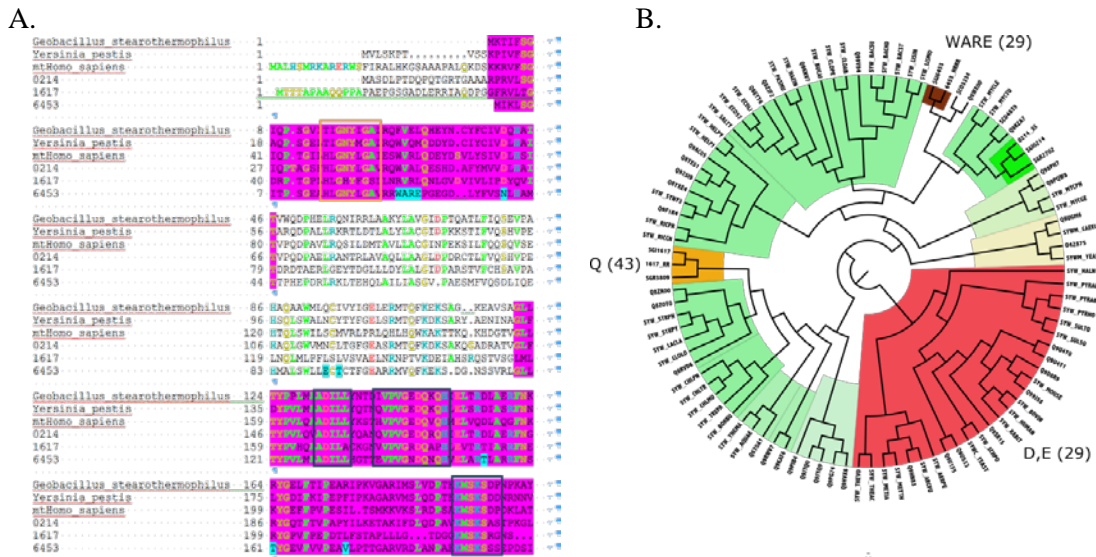


Figure 3.9 (A) Sequence alignments of relevant TrpRSs. Magenta segments are included in the TrpRS Urzyme (73). Other segments are in what is conventionally called “connecting peptide 1” (CP1), and the anticodon-binding domain is not shown. Relevant sequence differences are highlighted in turquoise. (B) Phylogenetic tree of TrpRS sequences, annotated with known indolmycin resistance (bold red) or sensitivity (bold green). The sensitivity of most bacterial sequences have not been measured (pastel green). Conspicuous sequence signatures of the different classes of resistant TrpRSs are included outside the circle.

When taken in context with other types of information, sequences of the indolmycin-sensitive and resistant TrpRSs from *S. griseus* ATCC 12648 furnish significant clues to the molecular bases of indolmycin resistance. We found that the two indolmycin resistant enzymes, SGI_1617 ($K_i = 2.7 \mu\text{M}$) and SGI_6453 ($K_i = 145 \mu\text{M}$) have distinct features in their primary sequences (Fig. 3.9A), as confirmed by their presence on distinct clades in the cladogram in Figure 3.9B. SGI_1617 has a glutamine in the position homologous to H43 in the *Bst* TrpRS. H43 in *Bst* TrpRS is known to play a critical role in binding to indolmycin as a participant in a hydrogen-binding network (74). However, glutamine cannot hydrogen bond in a fashion that will complement the network

stabilized by indolmycin. Thus, the modest resistance of that strain is likely a consequence of the H=>Q substitution.

Interestingly, the highly resistant enzyme, SGI_6453 has a consensus histidine in position 43. However, it has an unusual WARE sequence in the first β - α - β crossover of the Rossmann fold. Two residues of this sequence (*i.e.*, W and E), occupy positions homologous to what is known as the D1 switch. The D1 Switch is a master-switch that mediates interdomain shear during the conformational change undergone by *Bst* TrpRS in its catalytic cycle. Previous studies have shown that these residues, which form a broadly conserved tertiary packing motif, are coupled by ~5 kcal/mole to both the active-site metal ion (75) and the decision whether or not to activate the amino acid occupying the amino acid pocket (75). To understand how this unusual WARE sequence affect the activity of SGI_6453, reconfiguration of the *Bst* TrpRS D1 Switch by replacing the sequence FVEL with WARE was performed (Fig 3.10). The tryptophan is homologous to a phenylalanine that interchanges positions relative to F37, L29, and I4. That motion must be altered in some way by the substitution of phenylalanine with tryptophan at position 26 and, more importantly by the substitution of leucine with glutamic acid at position 29. The glutamate can assume a rotamer in which it is strongly attracted by electrostatic forces to arginine 175. Owing to the strong coupling of the conformational change to tryptophan binding affinity, this change is also consistent with the ~10-fold increase in K_M for substrate tryptophan in this enzyme, and hence for its comparable decrease in catalytic proficiency, k_{cat}/K_M .

Finally, using heterologous expression in *S. coelicolor*, we were able to confirm that the hypothetical gene cluster was sufficient for indolmycin biosynthesis. However, our current yield of indolmycin in *S. coelicolor* is much lower than its production in original producer strain *S. griseus* ATCC 12648. Our genetic analyses were recently corroborated by studies of Prof. K.S. Ryan and co-workers at the University of British Columbia (20). They reported reconstitution of indolmycin *in vitro* using enzymes. They also reported that deletion of either of the PLP-dependent

aminotransferases (SGI6457 and SGI6461) did not abolish production of indolmycenic acid, which is a precursor of indolmycin. They also found that production of indolepyruvate by SGI6454 was barely detectable, and the production by SGI6461 was also low. The low catalytic efficiency of aminotransferases within the cloned indolmycin biosynthetic gene cluster might be responsible for the low yield of indolmycin in *S. coelicolor*.

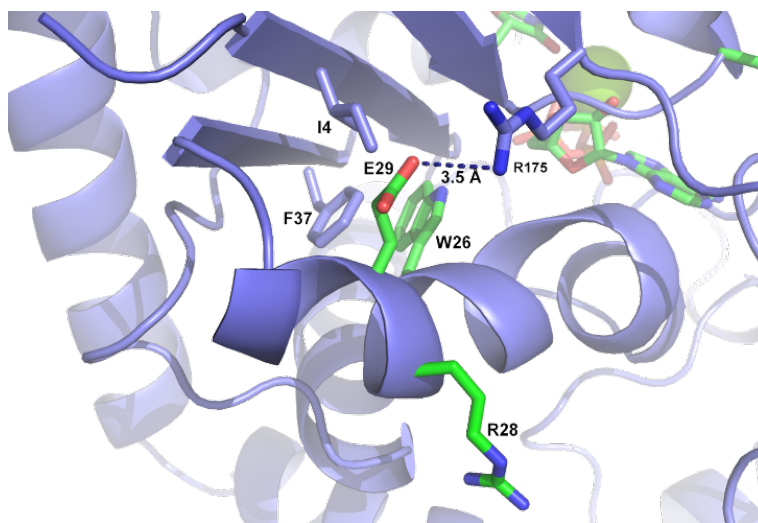


Figure 3.10 Reconfiguration of the *Bst* TrpRS D1 Switch by replacing the sequence FVEL with WARE as is observed in the SG6453 indolmycin-resistant TrpRS. Other residues in the D1 switch are shown as indicated, as is arginine 175, which can form a strong salt bridge interaction with the mutant glutamic acid in position 29.

In conclusion, we characterized three TrpRSs in indolmycin producer strain *S. griseus* ATCC 12648, one of which is a novel, very high indolmycin resistant. We also provided transcription profile of both indolmycin biosynthetic genes and *trpRS* genes along the growth of the bacteria. Finally, regardless of the low yield, we proved that the 11kb fragment cloned in this study is genetically sufficient for the production of indolmycin.

3.5 Scientific Contributions

In this chapter, Dr. Jennifer Davis (MPPB¹, Brown University) carried out the initial bioinformatic analyses of the *Streptomyces griseus* ATCC 12648 genome sequence and identified the three tryptophanyl-tRNA synthetase genes. James Blum 14' (Brown University) constructed *S. coelicolor* strains B1659, B1660, B1661 and B1662. He also cloned the three tryptophanyl-tRNA synthetase genes from *S. griseus* ATCC 12648 and performed the initial RT-PCR experiments for studies of the transcription of the indolmycin biosynthesis and resistance genes. Dr. Tishan Williams (UNC. Chapel Hill) purified three TrpRS enzymes from *S. griseus* ATCC 12648 and performed the biochemistry characterization of TrpRS enzymes. Prof. Charlie W. Carter Jr. (UNC. Chapel Hill) provided the discussion about molecular bases of TrpRS resistance to indolmycin. All the other experiments and data analyses were performed by Rui Zhang.

1. MPPB: Department of Molecular Pharmacology, Physiology and Biotechnology

3.6 References

1. Marsh, W.S., Garretson, A.L. and Wesel, E.M. (1960) PA 155 A, B, and X antibiotics produced by a strain of *Streptomyces albus*. *Antibiot Chemother*, **10**, 316-320.
2. Routien, J.B. (1966) Identity of streptomycete producing antibiotic PA155A. *J Bacteriol*, **91**, 1663.
3. Floss, H.G. (1981) Biosynthesis of some aromatic antibiotics. *Antibiotic biosynthesis* (Springer-Verlag, Berlin Heidelberg), **4**, 236-261.
4. Hurdle, J.G., O'Neill, A.J. and Chopra, I. (2005) Prospects for aminoacyl-tRNA synthetase inhibitors as new antimicrobial agents. *Antimicrob Agents Chemother*, **49**, 4821-4833.
5. Werner, R.G., Thorpe, L.F., Reuter, W. and Nierhaus, K.H. (1976) Indolmycin Inhibits Prokaryotic Tryptophanyl-tRNA Ligase. *European Journal of Biochemistry*, **68**, 1-3.
6. Beaulieu, D. and Ohemeng, K.A. (1999) Patents on bacterial tRNA synthetase inhibitors: January 1996 to March 1999. *Expert Opinion on Therapeutic Patents*, **9**, 1021-1028.

7. Werner, R.G. and Reuter, W. (1979) Interaction of indolmycin in the metabolism of tryptophan in rat liver. *Arzneimittelforschung*, **29**, 59-63.
8. Kanamaru, T., Nakano, Y., Toyoda, Y., Miyagawa, K.I., Tada, M., Kaisho, T. and Nakao, M. (2001) *In Vitro* and *In Vivo* Antibacterial Activities of TAK-083, an Agent for Treatment of *Helicobacter pylori* Infection. *Antimicrobial Agents and Chemotherapy*, **45**, 2455-2459.
9. Hurdle, J.G., O'Neill, A.J. and Chopra, I. (2004) Anti-staphylococcal activity of indolmycin, a potential topical agent for control of staphylococcal infections. *J Antimicrob Chemother*, **54**, 549-552.
10. Bogosian, G., Haydock, P.V. and Somerville, R.L. (1983) Indolmycin-Mediated Inhibition and Stimulation of Transcription at the Trp Promoter of *Escherichia-Coli*. *Journal of Bacteriology*, **153**, 1120-1123.
11. Kitabatake, M., Ali, K., Demain, A., Sakamoto, K., Yokoyama, S. and Soll, D. (2002) Indolmycin resistance of *Streptomyces coelicolor* A3(2) by induced expression of one of its two tryptophanyl-tRNA synthetases. *J Biol Chem*, **277**, 23882-23887.
12. Vecchione, J.J. and Sello, J.K. (2008) Characterization of an inducible, antibiotic-resistant aminoacyl-tRNA synthetase gene in *Streptomyces coelicolor*. *J Bacteriol*, **190**, 6253-6257.
13. Vecchione, J.J. and Sello, J.K. (2009) A novel tryptophanyl-tRNA synthetase gene confers high-level resistance to indolmycin. *Antimicrob Agents Chemother*, **53**, 3972-3980.
14. Vecchione, J.J. and Sello, J.K. (2010) Regulation of an auxiliary, antibiotic-resistant tryptophanyl-tRNA synthetase gene via ribosome-mediated transcriptional attenuation. *J Bacteriol*, **192**, 3565-3573.
15. Hornemann, U., Hurley, L.H., Speedie, M.K., Guenther, H.F. and Floss, H.G. (1969) Biosynthesis of the antibiotic indolmycin by *Streptomyces griseus*. C-Methylation at the β -carbon atom of the tryptophan side-chain. *J. Chem. Soc. D*, **0**, 245-246.
16. Hornemann, U., Hurley, L.H., Speedie, M.K. and Floss, H.G. (1970) Isolation and absolute configuration of indolmycenic acid, an intermediate in the biosynthesis of indolmycin by *Streptomyces griseus*. *Tetrahedron Lett*, 2255-2258.
17. Hornemann, U., Hurley, L.H., Speedie, M.K. and Floss, H.G. (1971) Biosynthesis of indolmycin. *Journal of the American Chemical Society*, **93**, 3028-3035.
18. Speedie, M.K., Hornemann, U. and Floss, H.G. (1975) Isolation and characterization of tryptophan transaminase and indolepyruvate C-methyltransferase. Enzymes involved in indolmycin biosynthesis in *Streptomyces griseus*. *J Biol Chem*, **250**, 7819-7825.
19. Cundliffe, E. (1989) How Antibiotic-Producing Organisms Avoid Suicide. *Annual Review of Microbiology*, **43**, 207-233.

20. Du, Y.L., Alkhalaf, L.M. and Ryan, K.S. (2015) In vitro reconstitution of indolmycin biosynthesis reveals the molecular basis of oxazolinone assembly. *Proc Natl Acad Sci U S A*, **112**, 2717-2722.
21. Sambrook J., F.E.F., T. Maniatis. (1989) *Molecular cloning: a laboratory manula* (2nd ed, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N. Y.).
22. Studier, F.W. (2005) Protein production by auto-induction in high density shaking cultures. *Protein Expr Purif*, **41**, 207-234.
23. Shima, J., Hesketh, A., Okamoto, S., Kawamoto, S. and Ochi, K. (1996) Induction of actinorhodin production by *rpsL* (encoding ribosomal protein S12) mutations that confer streptomycin resistance in *Streptomyces lividans* and *Streptomyces coelicolor* A3(2). *J Bacteriol*, **178**, 7276-7284.
24. Kieser T, B.M., Buttner MJ, Chater KF, Hopwood DA. (2000) *Practical Streptomyces Genetics* (John Innes Foundation, Norwich, England).
25. Bennett, S. (2004) Solexa Ltd. *Pharmacogenomics*, **5**, 433-438.
26. Hong, H.J., Hutchings, M.I., Hill, L.M. and Buttner, M.J. (2005) The role of the novel Fem protein VanK in vancomycin resistance in *Streptomyces coelicolor*. *J Biol Chem*, **280**, 13055-13061.
27. Gregory, M.A., Till, R. and Smith, M.C.M. (2003) Integration Site for *Streptomyces* Phage BT1 and Development of Site-Specific Integrating Vectors. *J Bacteriol*, **185**, 5320-5323.
28. Bibb, M.J., White, J., Ward, J.M. and Janssen, G.R. (1994) The mRNA for the 23S rRNA methylase encoded by the *ermE* gene of *Saccharopolyspora erythraea* is translated in the absence of a conventional ribosome-binding site. *Molecular Microbiology*, **14**, 533-545.
29. Jones, G.H., Paget, M.S.B., Chamberlin, L. and Buttner, M.J. (1997) Sigma-E is required for the production of the antibiotic actinomycin in *Streptomyces antibioticus*. *Molecular Microbiology*, **23**, 169-178.
30. Peti, W. and Page, R. (2007) Strategies to maximize heterologous protein expression in *Escherichia coli* with minimal cost. *Protein Expr Purif*, **51**, 1-10.
31. Fersht, A.R., Ashford, J.S., Bruton, C.J., Jakes, R., Koch, G.L. and Hartley, B.S. (1975) Active site titration and aminoacyl adenylate binding stoichiometry of aminoacyl-tRNA synthetases. *Biochemistry*, **14**, 1-4.
32. Francklyn, C.S., First, E.A., Perona, J.J. and Hou, Y.M. (2008) Methods for kinetic and thermodynamic analysis of aminoacyl-tRNA synthetases. *Methods* (San Diego, Calif.), **44**, 100-118.
33. Schneider, C.A., Rasband, W.S. and Eliceiri, K.W. (2012) NIH Image to ImageJ: 25 years of image analysis. *Nat Methods*, **9**, 671-675.

34. SAS. (2013). 9 ed. SAS Institute, Cary, NC.
35. Bierman, M., Logan, R., O'Brien, K., Seno, E.T., Nagaraja Rao, R. and Schoner, B.E. (1992) Plasmid cloning vectors for the conjugal transfer of DNA from *Escherichia coli* to *Streptomyces spp.* *Gene*, **116**, 43-49.
36. Gomez-Escribano, J.P. and Bibb, M.J. (2011) Engineering *Streptomyces coelicolor* for heterologous expression of secondary metabolite gene clusters. *Microb Biotechnol*, **4**, 207-215.
37. Gomez-Escribano, J.P. and Bibb, M.J. (2012) *Streptomyces coelicolor* as an Expression Host for Heterologous Gene Clusters. *Method Enzymol*, **517**, 279-300.
38. Werner, R.G. and Demain, A.L. (1981) Directed biosynthesis of new indolmycins. *J Antibiot* (Tokyo), **34**, 551-554.
39. Edgar, R.C. (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res*, **32**, 1792-1797.
40. Drummond, A.J. and Rambaut, A. (2007) BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evol Biol*, **7**, 214.
41. Blin, K., Medema, M.H., Kazempour, D., Fischbach, M.A., Breitling, R., Takano, E. and Weber, T. (2013) antiSMASH 2.0--a versatile platform for genome mining of secondary metabolite producers. *Nucleic Acids Res*, **41**, W204-212.
42. Hopwood, D.A. (2007) *Streptomyces in nature and medicine: the antibiotic makers* (Oxford University Press, Inc. New York).
43. Bibb, M.J. (2005) Regulation of secondary metabolism in streptomycetes. *Curr Opin Microbiol*, **8**, 208-215.
44. Wietzorrek, A. and Bibb, M. (1997) A novel family of proteins that regulates antibiotic production in streptomycetes appears to contain an OmpR-like DNA-binding fold. *Molecular Microbiology*, **25**, 1181-1184.
45. Pang, X., Aigle, B., Girardet, J.M., Mangenot, S., Pernodet, J.L., Decaris, B. and Leblond, P. (2004) Functional Angucycline-Like Antibiotic Gene Cluster in the Terminal Inverted Repeats of the *Streptomyces ambofaciens* Linear Chromosome. *Antimicrobial Agents and Chemotherapy*, **48**, 575-588.
46. Sheldon, P.J., Busarow, S.B. and Hutchinson, C.R. (2002) Mapping the DNA-binding domain and target sequences of the *Streptomyces peucetius* daunorubicin biosynthesis regulatory protein, DnrI. *Molecular Microbiology*, **44**, 449-460.
47. Ichinose, K., Ozawa, M., Itou, K., Kunieda, K. and Ebizuka, Y. (2003) Cloning, sequencing and heterologous expression of the medermycin biosynthetic gene cluster of *Streptomyces* sp. AM-7161: towards comparative analysis of the benzoisochromanquinone gene clusters. *Microbiology*, **149**, 1633-1645.

48. Widdick, D.A., Dodd, H.M., Barraille, P., White, J., Stein, T.H., Chater, K.F., Gasson, M.J. and Bibb, M.J. (2003) Cloning and engineering of the cinnamycin biosynthetic gene cluster from *Streptomyces cinnamoneus cinnamoneus* DSM 40005. *Proc Natl Acad Sci U S A*, **100**, 4316-4321.
49. Ryding, N.J., Anderson, T.B. and Champness, W.C. (2002) Regulation of the *Streptomyces coelicolor* calcium-dependent antibiotic by *absA*, encoding a cluster-linked two-component system. *J Bacteriol*, **184**, 794-805.
50. Cerdeño, A.M., Bibb, M.J. and Challis, G.L. (2001) Analysis of the prodiginine biosynthesis gene cluster of *Streptomyces coelicolor* A3(2): new mechanisms for chain initiation and termination in modular multienzymes. *Chemistry & Biology*, **8**, 817-829.
51. Takano, E., Kinoshita, H., Mersinias, V., Bucca, G., Hotchkiss, G., Nihira, T., Smith, C.P., Bibb, M., Wohlleben, W. and Chater, K. (2005) A bacterial hormone (the SCB1) directly controls the expression of a pathway-specific regulatory gene in the cryptic type I polyketide biosynthetic gene cluster of *Streptomyces coelicolor*. *Mol Microbiol*, **56**, 465-479.
52. Sun, Y., Zhou, X., Dong, H., Tu, G., Wang, M., Wang, B. and Deng, Z. (2003) A Complete Gene Cluster from *Streptomyces nanchangensis* NS3226 Encoding Biosynthesis of the Polyether Ionophore Nanchangmycin. *Chemistry & Biology*, **10**, 431-441.
53. Oliynyk, M., Stark, C.B.W., Bhatt, A., Jones, M.A., Hughes-Thomas, Z.A., Wilkinson, C., Oliynyk, Z., Demydchuk, Y., Staunton, J. and Leadlay, P.F. (2003) Analysis of the biosynthetic gene cluster for the polyether antibiotic monensin in *Streptomyces cinnamomensis* and evidence for the role of *monB* and *monC* genes in oxidative cyclization. *Molecular Microbiology*, **49**, 1179-1190.
54. Núñez, L.E., Méndez, C., Braña, A.F., Blanco, G. and Salas, J.A. (2003) The Biosynthetic Gene Cluster for the β -Lactam Carbapenem Thienamycin in *Streptomyces cattleya*. *Chemistry & Biology*, **10**, 301-311.
55. Garg, R.P., Ma, Y., Hoyt, J.C. and Parry, R.J. (2002) Molecular characterization and analysis of the biosynthetic gene cluster for the azoxy antibiotic valanimycin. *Molecular Microbiology*, **46**, 505-517.
56. Huang, J., Lih, C.J., Pan, K.H. and Cohen, S.N. (2001) Global analysis of growth phase responsive gene expression and regulation of antibiotic biosynthetic pathways in *Streptomyces coelicolor* using DNA microarrays. *Genes Dev*, **15**, 3183-3192.
57. Walter, A.E., Turner, D.H., Kim, J., Lyttle, M.H., Muller, P., Mathews, D.H. and Zuker, M. (1994) Coaxial stacking of helices enhances binding of oligoribonucleotides and improves predictions of RNA folding. *Proc Natl Acad Sci U S A*, **91**, 9218-9222.
58. Kel, A.E. (2003) MATCHTM: a tool for searching transcription factor binding sites in DNA sequences. *Nucleic Acids Research*, **31**, 3576-3579.

59. Messeguer, X., Escudero, R., Farre, D., Nunez, O., Martinez, J. and Alba, M.M. (2002) PROMO: detection of known transcription regulatory elements using species-tailored searches. *Bioinformatics*, **18**, 333-334.
60. Farre, D., Roset, R., Huerta, M., Adsuara, J.E., Rosello, L., Alba, M.M. and Messeguer, X. (2003) Identification of patterns in biological sequences at the ALGGEN server: PROMO and MALGEN. *Nucleic Acids Research*, **31**, 3651-3653.
61. Janssen, G.R. (1993) Eubacterial, archebacterial, and eukaryotic genes that encode leaderless promoters. *Industrial microorganisms: basic and applied molecular genetics* (American Society for Microbiology, Washington, D.C.), 59-67.
62. Moll, I., Grill, S., Gualerzi, C.O. and Blasi, U. (2002) Leaderless mRNAs in bacteria: surprises in ribosomal recruitment and translational control. *Mol Microbiol*, **43**, 239-246.
63. Torarinsson, E., Klenk, H.P. and Garrett, R.A. (2005) Divergent transcriptional and translational signals in Archaea. *Environ Microbiol*, **7**, 47-54.
64. Londei, P. (2005) Evolution of translational initiation: new insights from the archaea. *FEMS Microbiol Rev*, **29**, 185-200.
65. Zhu, H., Hu, G.Q., Yang, Y.F., Wang, J. and She, Z.S. (2007) MED: a new non-supervised gene prediction algorithm for bacterial and archaeal genomes. *BMC Bioinformatics*, **8**, 97.
66. Hu, G.Q., Zheng, X., Yang, Y.F., Ortet, P., She, Z.S. and Zhu, H. (2008) ProTISA: a comprehensive resource for translation initiation site annotation in prokaryotic genomes. *Nucleic Acids Res*, **36**, D114-119.
67. Zheng, X., Hu, G.Q., She, Z.S. and Zhu, H. (2011) Leaderless genes in bacteria: clue to the evolution of translation initiation mechanisms in prokaryotes. *BMC Genomics*, **12**, 361.
68. Larkin, M.A., Blackshields, G., Brown, N.P., Chenna, R., McGettigan, P.A., McWilliam, H., Valentin, F., Wallace, I.M., Wilm, A., Lopez, R. *et al.* (2007) Clustal W and Clustal X version 2.0. *Bioinformatics*, **23**, 2947-2948.
69. Zeng, Y., Roy, H., Patil, P.B., Ibba, M. and Chen, S. (2009) Characterization of two seryl-tRNA synthetases in albomycin-producing *Streptomyces* sp. strain ATCC 700974. *Antimicrob Agents Chemother*, **53**, 4619-4627.
70. D'Costa, V.M., McGrann, K.M., Hughes, D.W. and Wright, G.D. (2006) Sampling the antibiotic resistome. *Science*, **311**, 374-377.
71. Benveniste R., J.D. (1973) Aminoglycoside antibiotic-inactivating enzymes in actinomycetes similar to those present in clinical isolates of antibiotic-resistant bacteria. *Proc Natl Acad Sci U S A*, **70**, 2276-2280.
72. Wiener, P., Egan, S. and Wellington, E.M.H. (1998) Evidence for transfer of antibiotic-resistance genes in soil populations of streptomycetes. *Molecular Ecology*, **7**, 1205-1216.

73. Pham, Y., Kuhlman, B., Butterfoss, G.L., Hu, H., Weinreb, V. and Carter, C.W., Jr. (2010) Tryptophanyl-tRNA synthetase Urzyme: a model to recapitulate molecular evolution and investigate intramolecular complementation. *J Biol Chem*, **285**, 38590-38601.
74. Williams, T.L., Yin, Y.W. and Carter, C.W., Jr. (2016) Selective Inhibition of Bacterial Tryptophanyl-tRNA Synthetases by Indolmycin Is Mechanism-based. *J Biol Chem*, **291**, 255-265.
75. Weinreb, V., Li, L. and Carter, C.W., Jr. (2012) A master switch couples Mg(2)(+)-assisted catalysis to domain motion in *B. stearothermophilus* tryptophanyl-tRNA Synthetase. *Structure*, **20**, 128-138.

Chapter 4. Genetic investigation of three paralogous *trpE* genes in *S. coelicolor* and their contributions to the production of calcium-dependent antibiotic (CDA)

4.1 Introduction

Genetic redundancy often underlies the common phenomenon wherein deleting or mutating a gene from a genome has minimal or no impact on the phenotype or fitness of the organism because of functional compensation conferred by one or more other genes (1). Genetic redundancy is potentially of great relevance to organismal evolution for two reasons. First, it may protect organisms from potentially harmful mutations. Second, maintaining pools of functionally similar, yet diverse gene products, enables the evolution of new biochemical functions (2). At the individual gene level, genetic redundancy is caused either by ancient paralogy via gene duplication or horizontal gene transfer (HGT) (1-3).

Bioinformatic analyses of the genomes of streptomycetes have revealed that a great deal of genetic redundancy. The presence of the paralogous genes has raised a number of compelling questions for research (4). In one of the earliest bioinformatics analyses of paralogous genes, it was revealed that about 140 out of the 7,825 predicted genes in *S. coelicolor* has at least one paralog with more than 70% similarity. Several studies of paralogous genes in streptomycetes have revealed connections to antibiotic resistance, wherein the genes encode isoforms of enzymes that are not susceptible to antibiotics (5-11). For example, paralogous genes have been shown to underly the resistance of streptomycetes to indolmycin (9,10) and pentalenolactone (11). In other cases, paralogous genes are known to encode enzyme isoforms that are produced and active at different stages in the *Streptomyces* developmental cycle (4). One of the examples is the differential expression of gene clusters encoding genes for glycogen synthesis in vegetative and aerial mycelia (12).

Among the ~140 paralogous genes in *S. coelicolor*, we were particularly intrigued by the redundancy in genes for the biosynthesis of the proteinogenic amino acid, tryptophan. Indeed, there were four different genetic loci encoding genes for the biosynthesis of tryptophan (4). Tryptophan biosynthesis is the most biochemically expensive of the amino acid pathways, requiring incorporation of products generated from four other pathways (13-15). Though it is not obvious

why *S. coelicolor* would have so much capacity for producing this metabolically expensive amino acid, we thought that the redundancy might be important for ensuring that the bacterium has sufficient tryptophan for primary metabolism and for the production of the calcium-dependent antibiotic (CDA), a peptide antibiotic that contains two units of tryptophan (Fig. 4.1A) (16). This perspective was partially justified by the fact that one tryptophan biosynthesis locus (*trpCDGE*) is located inside the genetic locus for CDA biosynthesis (4,16).

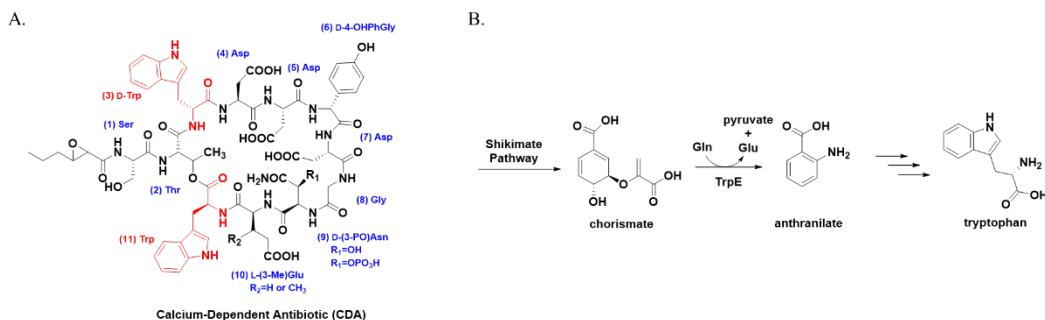


Figure 4.1 (A) The structure of calcium-dependent antibiotic (CDA). CDA is a cyclic lipopeptide consisting of 11 amino acid residues (including both proteinogenic and non-proteinogenic amino acids). The decapeptide lactone core of CDA contains two tryptophan units (L-tryptophan and D-tryptophan), both of which are colored in red. (B) Conversion of chorismate (product of shikimate pathway) to anthranilate under catalysis of TrpE, an anthranilate synthase. This is the first reaction in the branch of the aromatic biosynthesis pathway leading specifically to tryptophan.

We noted that there were three copies of the gene encoding anthranilate synthase in *S. coelicolor* (*trpE1*: SCO2117, *trpE2*: SCO3214, and *trpE3*: SCO2043) (4). The anthranilate synthase (AS)-catalyzed conversion from chorismate to anthranilate is the first committed step in the biosynthetic pathway leading to tryptophan (Fig. 4.1B) (17). TrpE, which catalyzes formation of the key C-N bond in anthranilate from ammonia and anthranilate, is metabolically sufficient for tryptophan biosynthesis when ammonia is abundant. To avoid the toxicity associated with high levels of

ammonia, TrpE functions often functions in collaboration with TrpG, an enzyme that provides ammonia on demand via catalytic hydrolysis of glutamine's side chain (3,17).

We propose that the physiological significance of the three anthranilate synthases in *S. coelicolor* is to ensure that the bacterium has sufficient tryptophan for primary metabolism and secondary metabolism (*i.e.*, CDA biosynthesis). This chapter describes how we used molecular genetic methods to study the contribution of each *trpE* gene to the growth of *S. coelicolor* and the production of CDA. Because paralogous genes often exhibit growth-phase dependent expression, we also investigated the regulation of each *trpE* gene. These studies led to the identification of a ribosome-dependent transcriptional attenuator in the 5' leader of the *trpE1* transcript.

4.2 Materials and Methods

4.2.1 Bacterial strains and culture conditions

A complete list of bacterial strains employed in this study is provided in Table 4.1. *Escherichia coli* strains DH5 α and ET12567/pUZ8002 were grown on Luria-Bertani medium at 37°C for routine sub-cloning (18). While *E. coli* strain BW25113/pIJ790 was grown at 30°C to maintain selection of pIJ790. *S. coelicolor* strains were grown at 30°C on mannitol soya flour medium (SFM), Difco Nutrient Agar medium (DNA), OXOID nutrient agar, yeast extract-malt extract medium (YEME) (19), or minimal liquid medium (NMMP). SFM was used for *S. coelicolor/E. coli* conjugations and for spore formation. Genomic DNA was isolated from *S. coelicolor* grown in YEME and total mRNA was isolated from *S. coelicolor* from NMMP. OXOID nutrient agar was used for the Calcium-Dependent Antibiotic (CDA) bioassay (19). For selection of *E. coli*, antibiotics were used at the following concentration: ampicillin (100 μ g/ml), apramycin (50 μ g/ml), chloramphenicol (25 μ g/ml), hygromycin (75 μ g/ml), kanamycin (50 μ g/ml), spectinomycin (50 μ g/ml) and streptomycin (50 μ g/ml). For selection of *S. coelicolor*, apramycin (50 μ g/ml), hygromycin

(75 µg/ml), spectinomycin (200 µg/ml) and streptomycin (10 µg/ml) were used. In conjugation with *S. coelicolor*, nalidixic acid (20 µg/ml) was used to counterselect *E. coli*. *Bacillus mycoides* was grown at 30°C on soft nutrient agar supplemented with calcium nitrate [Ca(NO₃)₂].

4.2.2 Plasmids and primers

All of the plasmids employed in this study are listed in Table 4.2. DNA sequencing was performed by Davis Sequencing (Davis, California). Table 4.3 shows the primers used in this work, all of which were synthesized by Invitrogen. PCR was performed with *Tap* DNA Polymerase (Invitrogen), *Pfu* (Stratagene, Agilent Technologies), Crimson *Taq* DNA polymerase (New England Biolabs), Phusion High-Fidelity DNA polymerase (New England Biolabs) and Expand High Fidelity PCR system (Roche). All PCR reactions with *Taq*, *Pfu*, Crimson *Taq* DNA polymerase and Phusion High-Fidelity DNA polymerase were performed with 5% (vol/vol) DMSO (19).

4.2.3 Chemicals

(±)-Indolmycin was chemically synthesized in the Sello laboratory by James J. Vecchione according to the procedure developed by Kamiyama *et al* (20). The ¹H and ¹³C NMR spectra, as well as mass spectra, of the synthetic material were identical to those reported for authentic indolmycin. The contaminating enantiomer of indolmycin in the synthetic racemate (1:1 mixture of enantiomers) does not have biological activity. All amounts of (±)-synthetic indolmycin described herein are corrected to reflect only the active enantiomer. L-(-)-tryptophan was purchased from Acros Organics and anthranilic acid was bought from Sigma-Aldrich.

4.2.4 Construction of *trpE1*, *trpE2* and *trpE3* null mutants

PCR-targeted mutagenesis was used to replace the *S. coelicolor trpE1*, *trpE2* and *trpE3* genes with an apramycin resistance cassette, *apr* (21). The requisite PCR products were amplified from the apramycin-resistance gene insert of pIJ773 using primers TrpE1 KO For/TrpE1 KO Rev, and

TrpE2 KO For/TrpE2 KO Rev and TrpE3 KO For/TrpE3 KO Rev. The PCR products were individually introduced into *E. coli* BW25113/pIJ790 containing cosmid St6E10 (*trpE1*), or cosmid StE8 (*trpE2*), or cosmid St4G6 (*trpE3*) and λ RED recombinase. The resultant recombinant cosmids, St6E10 Δ SCO2117::*apr* (*trpE1* null), StE8 Δ SCO3210-3215::*apr* (*trpE2* null) and St4G6 Δ SCO2043::*apr* (*trpE3* null), were introduced individually via non-methylating *E. coli* strain ET12567/pUZ8002 into *S. coelicolor* M145 by conjugation as previously described (22). M145 is a plasmid free-derivative of the wild-type strain (19). Exconjugants lacking the *trpE1*, *trpE2* cluster and *trpE3* genes were selected by apramycin resistance and kanamycin sensitivity. Gene replacements were confirmed via PCR on isolated genomic DNA from the null mutants using primers TrpE1 KO Det For/TrpE1 KO Det Rev for *S. coelicolor* B16570 Δ SCO2117::*apr*, and primers TrpE2 KO Det For/TrpE2 KO Det Rev for *S. coelicolor* B1674 Δ SCO3210-3215::*apr* and primers TrpE3 KO Det For/TrpE3 KO Det Rev for *S. coelicolor* B1678 Δ SCO2043::*apr*. The same strategy was also performed to construct *trpE1* and *trpE3* double knockout mutant strain *S. coelicolor* B1682 Δ SCO2043::*spec* Δ SCO2117::*apr*. As *trpE1* was replaced by apramycin resistance cassette, spectinomycin resistance cassette, *spec*, was used to replace *trpE3* gene. The requisite PCR products was amplified from spectinomycin-resistance gene insert of pIJ770 using the same primers TrpE3 KO For/TrpE3 KO Rev. Recombinant cosmid St4G6 Δ SCO2043::*spec* was introduced into *S. coelicolor* B1670 Δ SCO2117::*apr*, yielding *S. coelicolor* B1682. Exconjugants were selected by spectinomycin resistance and kanamycin sensitivity.

4.2.5 Genetic complementation of *trpE1*, *trpE2* and *trpE3* null mutants

To enable complementation of the *trpE1* null strain with *trpE1* gene under the control of its native promoter, a fragment for complementation harboring a 2,480-bp fragment was excised from cosmid St6E10 containing *trpE1* (SCO2117) ORF and 179 bp upstream of its translational start site by restriction digest with SphI. The fragment was treated with DNA polymerase I, Large (Klenow) Fragment (New England Biolabs) according to the manufacturer's protocol. The blunt-ended

fragment then was ligated into the SmaI site of pBluescript KS+ to yield pJS921. The fragment was then excised from pJS921 with SpeI and EcoRV and ligated into pMS81 pre-treated with the same enzymes to yield pJS924. pMS81 is a site-specific integrating vector that inserts into the Φ BT1 *attB* site of *S. coelicolor* (SCO4848) (23). pJS924 was transformed into non-methylating *E. coli* strain ET12567/pUZ8002 and introduced into *S. coelicolor* via conjugation (22), yielding *S. coelicolor* B1671 Δ SCO2117::*apr* pMS81-SCO2117. Hygromycin-resistant exconjugants were purified by single colony isolation. Phenotype analysis of *S. coelicolor* B1671 was accomplished through the CDA bioassay.

For complementation of the *trpE2* null strain, a 7,562-bp fragment was excised from cosmid StE8 containing the whole *trpE2* gene cluster (SCO3210-15) and 225 upstream base pairs by restriction digests with SfcI and XcmI. The fragment was treated with DNA polymerase I, Large (Klenow) Fragment (New England Biolabs) according to the manufacturer's protocol. The blunt-ended fragment then was ligated into the SmaI site of pBluescript KS+ to yield pJS922. The fragment was then excised from pJS922 and ligated into pMS81 as previously described to yield pJS925, which was introduced into *S. coelicolor* B1674 Δ SCO3210-3215::*apr* as previously described.

To enable complementation of the *trpE3* null strain with *trpE3* gene under the control of its native promoter, a 2,074-bp fragment was excised from cosmid St4G6 containing *trpE3* (SCO2043) ORF and 175 upstream base pairs by restriction digest with ScaI and BlnI. The fragment was treated with DNA polymerase I, Large (Klenow) Fragment (New England Biolabs) according to the manufacturer's protocol. The blunt-ended fragment then was ligated into the SmaI site of pBluescript KS+ to yield pJS923. The fragment was then excised from pJS923 and ligated into pMS81 as previously described to yield pJS926, which was introduced into *trpE3* null mutant strain *S. coelicolor* B1678 Δ SCO2304::*apr* as previously described.

4.2.6 Construction of *trpE1*, *trpE2* and *trpE3* overexpression strains

The *trpE1* ORF was amplified from cosmid St6E10 using primers TrpE1 Clone For and TrpE1 Clone Rev. The PCR product, with an engineered NdeI site at its 5' end and an engineered HindIII site at its 3' end, was ligated into pBluescript KS+ to yield pJS927. The fragment containing *trpE1* gene ORF was then excised from pJS927 with NdeI and HindIII and ligated into vector pIJ10257 pre-treated with the same enzymes to yield pJS930. pIJ10257 is a site-specific integrating vector that inserts into the Φ BT1 *attB* site of *S. coelicolor* (SCO4848) (23,24). Gene that is cloned into NdeI-HindIII site is under control of strong constitutive promoter *ermE** (25). pJS930 was then transformed into ET12567/pUZ8002 and introduced into *S. coelicolor* M145 wild-type strain (yielding strain B1673) and B1670 strain (yielding strain B1672 as an additional complementation of *trpE1* null mutant) via conjugation, as previously described (22). Exconjugants were selected by hygromycin resistance. Phenotypic analyses of *S. coelicolor* B1673 and B1672 strains were accomplished through CDA bioassay using the CDA-susceptible bacterium *Bacillus mycoides* (19).

The *trpE2* ORF was amplified from cosmid StE8 using primers TrpE2 Clone For and TrpE2 Clone Rev. The PCR product, with an engineered NdeI site at its 5' end and an engineered HindIII site at its 3' end, was cloned into pBluescript KS+ and subsequently sub-cloned into pIJ10257 to yield pJS931, which was introduced into *S. coelicolor* M145 wild-type strain (yielding strain B1677) and B1674 strain (yielding strain B1676 as an additional complementation of *trpE2* null mutant) through conjugation.

The *trpE3* ORF was amplified from cosmid St4G6 using primers TrpE3 Clone For and TrpE3 Clone Rev. The PCR product, with an engineered NdeI site at its 5' end and an engineered HindIII site at its 3' end, was cloned into pBluescript KS+ and subsequently sub-cloned into pIJ10257 to yield pJS929, which was introduced into *S. coelicolor* M145 wild-type strain (yielding strain B1681) and B1678 strain (yielding strain B1680 as an additional complementation of *trpE3* null mutant) through conjugation.

4.2.7 Growth curve

For all the 4 experimental strains, wild-type *S. coelicolor* M145, *S. coelicolor* B1670 (Δ SCO2117::*apr*), *S. coelicolor* B1674 (Δ SCO3210-3215::*apr*) and *S. coelicolor* B1678 (Δ SCO2304::*apr*), approximately 2×10^9 spores were germinated in 2 \times YT media (19), inoculated into flasks containing 30 ml NMMP, and then grown as shaken liquid cultures for 84 h at 30°C. At each indicated time, 1.5 ml samples were collected from each flask and dry cell weights were measured.

4.2.8 CDA bioassay

CDA production assays were performed essentially as described previously (19). For *S. coelicolor* M145 wild-type as well as all the constructed strains, approximately 2×10^8 spores were suspended in 5 μ l of water and spotted onto 25 ml OXOID nutrient agar. The plates were incubated at 30°C for 48 h, after which they were overlaid with 10 ml soft nutrient agar containing CDA-sensitive bacterium, *Bacillus mycoides* and $\text{Ca}(\text{NO}_3)_2$ at a concentration to give 12 mM through the plate. (For the CDA bioassay, *B. mycoides* was grown at 30°C in Difco nutrient broth to an OD600 of 0.7, and 0.5 ml of the overnight culture was added into 10 ml of Soft nutrient agar.) The plates with both *S. coelicolor* and *B. mycoides* were then incubated at 30°C for another 12 h, then the zones of inhibition, which are absent when $\text{Ca}(\text{NO}_3)_2$ is omitted, were measured.

For chemical complementation of *trpE* null mutants, 25 ml OXOID nutrient agar was supplemented with various concentrations of L(-)-tryptophan (0.5 mM, 5 mM, 25 mM and 37.5 mM) or anthranilic acid (0.5 mM, 1 mM, 2 mM and 3 mM). Wild-type *S. coelicolor* M145, *S. coelicolor* B1653 (Δ SCO2117::*apr*), *S. coelicolor* B1657 (Δ SCO3210-3215::*apr*) and *S. coelicolor* B1661 (Δ SCO2304::*apr*) were grown at 30°C only for 42 h before the plates were overlaid with soft nutrient agar.

4.2.9 Transcriptional analyses

Wild-type *S. coelicolor* M145 was grown as a shaken liquid culture (30 ml) in NMMP for 48 h. At each indicated time (18 h, 24 h and 48 h), an aliquot of culture sufficient to attain a cell pellet volume of 100 μ l was removed for total mRNA isolation. The cells were washed once with 10.3% (w/v) sucrose solution followed by the addition of 100 μ l of 10 mg/ml lysozyme solution (50 mM Tris-HCl, 1 mM EDTA, pH 8.0). The cells were incubated at 37°C for 15 min. Total RNA was isolated using the RNeasy Mini Kit (Qiagen) following the manufacturer's protocol. The concentration of each purified, DNA-free total RNA isolated was measured with a NanoDrop ND-1000 spectrophotometer. An equal amount of total RNA (1800 ng) was employed in all RT-PCR. RT-PCR was performed with the OneStep RT-PCR kit (Qiagen) according to the manufacturer's protocol. A 361-bp complementary DNA (cDNA) corresponding to the *trpE1* (SCO2117) transcript, a 214-bp cDNA corresponding to the *trpE2* (SCO3214) transcript, a 377-bp cDNA corresponding to the *trpE3* (SCO2043) transcript and a 417-bp cDNA corresponding to the *cdaPSI* (SCO3230, encoding CDA peptide synthetase I) transcript were detected with TrpE1 RT For/Rev primers, Trp2 RT For/Rev primers, TrpE3 RT For/Rev primers, and *cdaPSI* RT For/Rev primers, respectively. A 486-bp cDNA corresponding to the *hrdB* (SCO5820) transcript, was detected with *hrdB* RT For/Rev primers as a positive control. All primer sequences are listed in Table 4.3. The PCR program used for detection of all the transcripts was 50°C for 30 min, 95°C for 15 min, 30 cycles of 94°C for 30 s, 56°C for 30 s, and 72°C for 1 min, followed by a final elongation at 72°C for 10 min. Detection of possible contaminating DNA in RNA samples was accomplished by PCR with *Pfu* polymerase under the same conditions. No bands were observed in the controls, indicating that all RT-PCR products correspond to amplification of RNA transcripts.

For indolmycin induction experiments, wild-type *S. coelicolor* M145 shaken liquid culture (30 ml) was grown in NMMP for 21 h, followed by treatment with 100 μ M indolmycin, or DMSO as a negative control for a three-hour period. Total RNA isolation and RT-PCR were performed the

same as described above except that 1500 ng RNA was employed for all the “induced” samples in RT-PCR analyses.

4.2.10 Mapping of the *trpE1* transcription start site

The *trpE1* transcription start site was identified using the 5' RACE System for Rapid Amplification of cDNA Ends, Version 2.0 (Invitrogen). For the experiments, total RNA was isolated as previously described from an NMMP culture of wild-type *S. coelicolor* M600 induced with 100 μ M indolmycin to elevate levels of the *trpE1* transcript. The first-strand cDNA synthesis was performed using 30 pmol of primer SCO2117 GSP E and 1800 ng of RNA template according to the manufacturer's protocol for high GC-content transcripts. After cDNA synthesis, the reaction mixture was treated with an RNase Mix to remove template RNA and purified using the S.N.A.P. column procedure. An oligo-dC tail was added to the purified cDNA using the TdT-tailing reaction. Amplification of dC-tailed cDNA was accomplished using a 5 μ l aliquot of the preceding reaction as template, the primer SCO2117 GSP D and abridged Anchor primer (AAP, supplied with kit) and *Taq* DNA polymerase. The resulting polymerase chain reaction (PCR) product was used in a second PCR using the nested primer SCO2117 GSP C and universal amplification primer (UAP, supplied with kit). The first and second PCR products were purified and then ligated into the pGEM-easy vector and transformed into the *E. coli* strain DH5 α to yield pJS933 and pJS934. DNA sequencing of cloned inserts was performed by Davis Sequencing (Davis, CA). The location of the transcription start site was corroborated by independent trials.

4.2.11 Creation of Reporter Constructs

Using *S. coelicolor* cosmid St2C1A as template, a 374-base pair region spanning the *trpE1* leader and promoter was amplified by PCR using primers *trpE1* L-374 F and *trpE1* L-374R. The *trpE1* leader was then ligated into pMU1 (26) by engineered SpeI and NdeI restriction sites as were the various mutations of the leader.

4.2.12 Site-Directed Mutagenesis in *trpE1* leader

Specific point mutations in the *S. coelicolor trpE1* leader were made using the Stratagene QuikChange site-directed mutagenesis procedure. The primer pairs used to generate all point mutations in the *trpRS1* leader are provided in Table 4.3. Plasmid pJS361 was used as template in these reactions. All mutant *trpRE1* leaders were introduced into pMU1, yielding plasmids pJS362 – pJS368 (Table 4.2). Plasmids pMU1 and pJS361-pJS375 were independently introduced into wild-type *S. coelicolor* strain M600 (19) via conjugation, yielding strains B798 and B783- B790, respectively (Table 4.1).

4.2.13 Assaying Bioluminescence of the Reporter Constructs

Strains harboring the reporter constructs based on pMU1 were plated on Verdor 96 well plates at 10^6 cfu (colony forming unit) per well and grown for 3 days at 30°C. Each well contained 300 μ l of Difco Nutrient Agar. Luciferase activity was measured using the Envision Multilabel Plate Reader following standard protocol at 450 nm wavelength. At least six experimental replicates were made for each strain. The average value of the replicates was then obtained for each strain and reported here. Obvious statistical outliers were excluded from mean calculation.

Table 4.1 Strains employed in this study

Strain	Genotype	Reference / Source
<i>S. coelicolor</i>		
M145	Prototroph SCP1-, SCP2-	(19)
M600	Prototroph SCP1-, SCP2-	(19)
B783	M600 – pJS361	This Study
B784	M600 – pJS362	This Study
B785	M600 – pJS363	This Study
B786	M600 – pJS364	This Study
B787	M600 – pJS365	This Study
B788	M600 – pJS366	This Study
B789	M600 – pJS367	This Study
B790	M600 – pJS368	This Study
B798	M600 – pMU1	This Study
B1670	M145 Δ SCO2117:: <i>apr</i>	This Study
B1671	M145 Δ SCO2117:: <i>apr</i> – pJS924	This Study

B1672	M145 ΔSCO2117::apr – pJS930	This Study
B1673	M145 wild-type – pJS930	This Study
B1674	M145 ΔSCO3210-3215::apr	This Study
B1675	M145 ΔSCO3210-3215::apr – pJS925	This Study
B1676	M145 ΔSCO3210-3215::apr – pJS931	This Study
B1677	M145 wild-type – pJS931	This Study
B1678	M145 ΔSCO2304::apr	This Study
B1679	M145 ΔSCO2304::apr – pJS926	This Study
B1680	M145 ΔSCO2304::apr – pJS932	This Study
B1681	M145 wild-type – pJS932	This Study
B1682	M145 ΔSCO2117::apr ΔSCO2304::spec	This Study
<i>E. coli</i>		
DH5α	F-ϕlacZΔM15 Δ(lacZYA-argF)U169 recA1 endA1 hsdR17(r _k ⁻ , m _k ⁻) phoA supE44 thi-1 gyrA96 relA1 λ ⁻	Invitrogen
BW25113	Δ(araD-araB)567 ΔlacZ4787(::rrnB-4) lacI ^p -4000(lacI ^Q) λ- rpoS369(Am) rph-1 Δ(rhaD-rhaB)568 hsdR514	(27)
ET12567	dam dcm hsdS cat tet	(21)
<i>Bacillus mycoides</i>		

Table 4.2 Plasmids employed in this study

Plasmid	Description	Vector Backbone	Reference/Source
pIJ790	[<i>oriR101</i>], [<i>repA101(ts)</i>], <i>araBp-gam-bet-exo</i> , <i>Chl</i> ^r		(21)
pUZ8002	RP4 Derivative <i>OriT</i> ^r , <i>Kan</i> ^r		(21)
pGem-T Easy	pUC-derived, <i>lacZ</i> , <i>Amp</i> ^r		Promega
pBluescript KS+	pUC <i>ori</i> , MCS, <i>Amp</i> ^r		Agilent Technologies (Stratagene)
pMS81	<i>oriT</i> , ΦBT1 <i>attB-int</i> , <i>Hyg</i> ^r	pSET152	(23)
pIJ10257	<i>oriT</i> , ΦBT1 <i>attB-int</i> , <i>Hyg</i> ^r , <i>ermEp</i> [*]	pMS81	(24)
pMU1	<i>oriT</i> , <i>luxCDABE</i> , ΦBT1 <i>attB-int</i> , <i>Apr</i> ^r		(26)
pJS361	<i>trpE1</i> wild type leader	pMU1	This Study
pJS362	<i>trpE1</i> leader A(-163)G	pMU1	This Study
pJS363	<i>trpE1</i> leader G(-132)C	pMU1	This Study
pJS364	<i>trpE1</i> leader G(-135)C	pMU1	This Study
pJS365	<i>trpE1</i> leader G(-138)C	pMU1	This Study
pJS366	<i>trpE1</i> leader G(-132-135)C	pMU1	This Study
pJS367	<i>trpE1</i> leader G(-135-138)C	pMU1	This Study
pJS368	<i>trpE1</i> leader G(-132-135-138)C	pMU1	This Study
pJS370	<i>trpE1</i> wild type leader	pBluescript KS+	This Study
pJS371	<i>trpE1</i> leader A(-163)G	pBluescript KS+	This Study
pJS372	<i>trpE1</i> leader G(-132)C	pBluescript KS+	This Study
pJS373	<i>trpE1</i> leader G(-135)C	pBluescript KS+	This Study
pJS374	<i>trpE1</i> leader G(-138)C	pBluescript KS+	This Study
pJS375	<i>trpE1</i> leader G(-132-135)C	pBluescript KS+	This Study
pJS376	<i>trpE1</i> leader G(-135-138)C	pBluescript KS+	This Study
pJS377	<i>trpE1</i> leader G(-132-135-138)C	pBluescript KS+	This Study

pJS921	2480bp SphI fragment from cosmid St6E10 containing <i>trpE1</i> (SCO2117) with 179 upstream bp, Amp ^r	pBluescript KS+	This Study
pJS922	7562bp SfcI and XcmI fragment from cosmid StE8 containing <i>trpE2</i> cluster (SCO3210-15) with 225 upstream bp, Amp ^r	pBluescript KS+	This Study
pJS923	2074bp ScaI and BlnI fragment from cosmid St4G6 containing <i>trpE3</i> (SCO2043) with 175 upstream bp, Amp ^r	pBluescript KS+	This Study
pJS924	2480bp SphI fragment from cosmid St6E10 containing <i>trpE1</i> (SCO2117) with 179 upstream bp, Hyg ^r	pMS81	This Study
pJS925	7562bp SfcI and XcmI fragment from cosmid StE8 containing <i>trpE2</i> cluster (SCO3210-15) with 225b upstream bp, Hyg ^r	pMS81	This Study
pJS926	2074bp ScaI and BlnI fragment from cosmid St4G6 containing <i>trpE3</i> (SCO2043) with 175 upstream bp, Hyg ^r	pMS81	This Study
pJS927	<i>trpE1</i> (SCO2117) ORF with engineered NdeI and HindIII sites at each end, Amp ^r	pBluescript KS+	This Study
pJS928	<i>trpE2</i> (SCO3214) ORF with engineered NdeI and HindIII sites at each end, Amp ^r	pBluescript KS+	This Study
pJS929	<i>trpE3</i> (SCO2043) ORF with engineered NdeI and HindIII sites at each end, Amp ^r	pBluescript KS+	This Study
pJS930	<i>ermEp</i> *- <i>trpE1</i> (SCO2117) ORF, Hyg ^r	pIJ10257	This Study
pJS931	<i>ermEp</i> *- <i>trpE2</i> (SCO3214) ORF, Hyg ^r	pIJ10257	This Study
pJS932	<i>ermEp</i> *- <i>trpE3</i> (SCO2043) ORF, Hyg ^r	pIJ10257	This Study
pJS933	<i>trpE1</i> 5'-RACE product of Primer AAP (kit) and SCO2117 GSP2D	pGem-T Easy	This Study
pJS934	<i>trpE1</i> 5'-RACE product of Primer AUAP (kit) and SCO2117 GSP2C	pGem-T Easy	This Study

Table 4.3 Oligonucleotides employed in this study

Primer Name	Application/Function	Sequence
TrpE1 KO For	PCR-Targeting/Disruption of <i>trpE1</i> (SCO2117)	5'-CCCGCATCGCGGGATGTGCCGCGGCCCGCGC GCGCATGACTAGTATTCCGGGGATCCGTCGACC-3' ^a
TrpE1 KO Rev	PCR-Targeting/Disruption of <i>trpE1</i> (SCO2117)	5'-GGTCAGCACCGACTCGGGGTGGAAGTGGACGC CCGCGAAACTAGTTGTAGGCTGGAGCTGCTTC-3' ^a
TrpE2 KO For	PCR-Targeting/Disruption of <i>trpE2</i> cluster (SCO3210-3215)	5'-TCCCGACGCATGCGCGCACAGACGAGCGGACC CACCGTGACTAGTATTCCGGGGATCCGTCGACC-3' ^a
TrpE2 KO Rev	PCR-Targeting/Disruption of <i>trpE2</i> cluster (SCO3210-3215)	5'-CTCCGCCCCACCCTCCCGCCCCGACGCACCCG CCCTCAACTAGTTGTAGGCTGGAGCTGCTTC-3' ^a
TrpE3 KO For	PCR-Targeting/Disruption of <i>trpE3</i> (SCO2043)	5'-ATCCCGGTGAGCCGCAAGCTCCTCGCCGACGGC GACACC <u>ACTAGT</u> ATTCCGGGGATCCGTCGACC-3' ^a
TrpE3 KO Rev	PCR-Targeting/Disruption of <i>trpE3</i> (SCO2043)	5'-CTCGGTGTCTCGGCGACCGGGTCGGAGTCGGC GACGAC <u>ACTAGT</u> TGTAGGCTGGAGCTGCTTC-3' ^a
TrpE1 KO Det For	Verification of <i>trpE1</i> (SCO2117) Disruption	5'-CGGCCTGATCATCTGAGG-3'
TrpE1 KO Det Rev	Verification of <i>trpE1</i> (SCO2117) Disruption	5'-TCAGGGTCAGCACCGACTC-3'
TrpE2 KO Det For	Verification of <i>trpE2</i> cluster (SCO3210-3215) Disruption	5'-CGACAGGGACGGATCTTG-3'
TrpE2 KO Det Rev	Verification of <i>trpE2</i> cluster (SCO3210-3215) Disruption	5'-GCACGAGGGTACGGAGTC-3'

TrpE3 KO Det For	Verification of <i>trpE3</i> (SCO2043) Disruption	5'-GCTCCTCGCGGATCAGTA-3'
TrpE3 KO Det Rev	Verification of <i>trpE3</i> (SCO2043) Disruption	5'-GGTACGGGGGTGAGGAAC-3'
TrpE1 RT For	Detection of the <i>trpE1</i> (SCO2117) transcript	5'-GACGGCAAGGAGATCGAG-3'
TrpE1 RT Rev	Detection of the <i>trpE1</i> (SCO2117) transcript	5'-GGATCAGGATGGGGGAG-3'
TrpE2 RT For	Detection of the <i>trpE2</i> (SCO3214) transcript	5'-CGTGATGCACCTGTCGTC-3'
TrpE2 RT Rev	Detection of the <i>trpE2</i> (SCO3214) transcript	5'-GATGGCGAAGTCGGTCAG-3'
TrpE3 RT For	Detection of the <i>trpE3</i> (SCO2043) transcript	5'-GCCCTGCAGATCATCGAC-3'
TrpE3 RT Rev	Detection of the <i>trpE3</i> (SCO2043) transcript	5'-GGTACGGGGGTGAGGAAC-3'
<i>cdaPSI</i> RT For	Detection of the <i>cdaPSI</i> (SCO3230) transcript	5'-GGATCCTGCCTGGAGATC-3'
<i>cdaPSI</i> RT Rev	Detection of the <i>cdaPSI</i> (SCO3230) transcript	5'-CAGCCGCTCGTAGAACAG-3'
<i>trpB</i> RT For	Detection of the <i>trpB</i> (SCO2037) transcript	5'-AAGTCCGACCCGGAGTTC-3'
<i>trpB</i> RT Rev	Detection of the <i>trpB</i> (SCO2037) transcript	5'-CTTGTGTGAGCCGGTGTG-3'
<i>trpC1</i> RT For	Detection of the <i>trpC1</i> (SCO2039) transcript	5'-GCTGCGCAAGGACTTCAT-3'
<i>trpC1</i> RT Rev	Detection of the <i>trpC1</i> (SCO2039) transcript	5'-TCGACCTCGAGCGTCTTC-3'
<i>hrdB</i> RT For	Detection of the <i>hrdB</i> (SCO5820) transcript	5'-CTCGAGGAAGAGGGTGTGAC-3'
<i>hrdB</i> RT Rev	Detection of the <i>hrdB</i> (SCO5820) transcript	5'-TGCCGATCTGCTTGAGGTAG-3'
TrpE1 Clone For	Cloning of the <i>trpE1</i> ORF (SCO2117)	5'-CATATGACACCGCTCGACAGC-3' ^b
TrpE1 Clone Rev	Cloning of the <i>trpE1</i> ORF (SCO2117)	5'-AAGCTTCTTCCTCGACAAGTCCCTG-3' ^c
TrpE2 Clone For	Cloning of the <i>trpE2</i> ORF (SCO3214)	5'-GAAGGCGGCCATATGACCAC-3' ^b
TrpE2 Clone Rev	Cloning of the <i>trpE2</i> ORF (SCO3214)	5'-AAGCTTTTGTGATCACCGCGAC-3' ^c
TrpE3 Clone For	Cloning of the <i>trpE3</i> ORF (SCO2043)	5'-CATATGGACGTCACGCATGAC-3' ^b
TrpE3 Clone Rev	Cloning of the <i>trpE3</i> ORF (SCO2043)	5'-AAGCTTTTCCCATTGAGGGCGTG-3' ^c
SCO2117 GSP C	5' RACE, <i>trpE1</i>	5'-GAGTCTGCCGTGCTC-3'
SCO2117 GSP D	5' RACE, <i>trpE1</i>	5'-CGGATCTGCCGGAAG-3'
SCO2117 GSP E	5' RACE, <i>trpE1</i>	5'-GATGTCGTGGCGCTC-3'
<i>trpE1</i> L-374 F	Cloning of the <i>trpE1</i> leader	5'-ACTAGTGTGCTCTATTCTACAAGGGCAC-3' ^a
<i>trpE1</i> L-374 R	Cloning of the <i>trpE1</i> leader	5'-CATATGGGGTCCGTGTCCTTCTC-3' ^b
E1 A(-163)G F	Site-directed mutagenesis of the <i>trpE1</i> leader, creation of A(-163)G mutation	5'-GGTCGGCGGCCTGTTCCGCGCAC-3'
E1 A(-163)G R	Site-directed mutagenesis of the <i>trpE1</i> leader, creation of A(-163)G mutation	5'-GTGCGCGAACAGGCCCGACC-3'
E1 G(-138)C F	Site-directed mutagenesis of the <i>trpE1</i> leader, creation of G(-138)C mutation	5'-GACCCGGAACCTCGTGGTGGAC-3'
E1 G(-138)C R	Site-directed mutagenesis of the <i>trpE1</i> leader, creation of G(-138)C mutation	5'-GTCCACCACGAGTTCGGGTC-3'

E1 G(-135)C F	Site-directed mutagenesis of the <i>trpE1</i> leader, creation of G(-135)C mutation	5'- <u>CCC</u> GGA <u>ACT</u> GGT <u>CGT</u> GGACCGCTC-3'
E1 G(-135)C R	Site-directed mutagenesis of the <i>trpE1</i> leader, creation of G(-135)C mutation	5'-GTGAGCGGTCCACGACCAGTTC-3'
E1 G(-132)C F	Site-directed mutagenesis of the <i>trpE1</i> leader, creation of G(-132)C mutation	5'-GGA <u>ACT</u> GGTGGT <u>CGA</u> CCGCTC-3'
E1 G(-132)C R	Site-directed mutagenesis of the <i>trpE1</i> leader, creation of G(-132)C mutation	5'-GTGAGCGGT <u>CGA</u> CCAG-3'
E1 G(-135 -138)C F	Site-directed mutagenesis of the <i>trpE1</i> leader, creation of G(-135-138)C mutation	5'- <u>CCG</u> GAACTCGT <u>CGT</u> GGACCGCTC-3'
E1 G(-135 -138)C R	Site-directed mutagenesis of the <i>trpE1</i> leader, creation of G(-135-138)C mutation	5'- <u>CCG</u> GAACTCGT <u>CGT</u> GGACCGCTC-3'
E1 G(-132 -135)C F	Site-directed mutagenesis of the <i>trpE1</i> leader, creation of G(-132-135)C mutation	5'-GAACTGGT <u>CGT</u> CGACCGCTC-3'
E1 G(-132 -135)C R	Site-directed mutagenesis of the <i>trpE1</i> leader, creation of G(-132-135)C mutation	5'-GAGCGGT <u>CGA</u> CCAGTTC-3'
E1 G(-132 -135-138)C F	Site-directed mutagenesis of the <i>trpE1</i> leader, creation of G(-132-135-138)C mutation	5'-GACCCGGA <u>ACT</u> CGT <u>CGT</u> CGAC-3'
E1 G(-132 -135-138)C R	Site-directed mutagenesis of the <i>trpE1</i> leader, creation of G(-132-135-138)C mutation	5'-GTGAGCGGT <u>CGA</u> CCAGTTC-3'

- The engineered SpeI site is underlined.
- The engineered NdeI site is underlined.
- The engineered HindIII site is underlined.

4.3 Results

4.3.1 Deletion of *trpE* genes individually has no effect on the growth of *S. coelicolor*

The *trpE* genes (SCO2117, SCO3214 and SCO2043) in *S. coelicolor* encode anthranilate synthases, which are the essential enzymes in tryptophan biosynthesis. To investigate the significance of each TrpE isoform in the physiology of *S. coelicolor*, gene replacement by PCR-targeting procedure was performed to construct four mutant strains, namely *S. coelicolor* B1670 (Δ SCO2117::*apr*), *S. coelicolor* B1674 (Δ SCO3210-3215::*apr*), *S. coelicolor* B1678 (Δ SCO2043::*apr*) and *S. coelicolor* B1682 (Δ SCO2117::*apr* Δ SCO2043::*spec*). Because *trpE2* (SCO3214) is part of a polycistronic unit, the entire operon was replaced by *apr* cassette to avoid polar effects that would result if only

trpE2 was replaced. We found that all *trpE* null mutants were visually indistinguishable from wild-type *S. coelicolor* with respect to colony size and sporulation on three different solid media (SFM, DNA and OXOID nutrient agar). To complement growth analyses on solid media, we carried out phenotypic analyses of the null mutants in the liquid medium NMMP. We inoculated wild-type *S. coelicolor*, the *trpE1* null strain, the *trpE2* null strain and the *trpE3* null strain in NMMP medium separately and let them grow as shaken liquid cultures for 80 hours. The overall growth were very similar according to the plotted growth curve (Fig. 4.2).

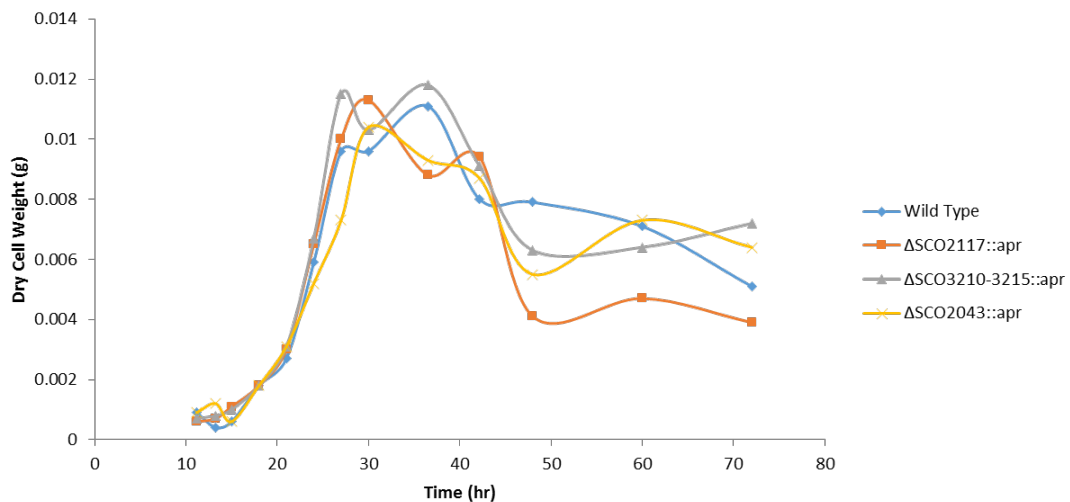


Figure 4.2 The growth curves of *S. coelicolor* wild-type M145 and the *trpE* null strains. The bacteria was grown as shaken liquid culture in NMMP medium at 30°C for 84 hours. At each indicated time, 1.5 ml samples were collected from each flask and dry cell weights were measured. Disruption of each *trpE* gene individually did not have a marked effect on the growth of *S. coelicolor*.

We also investigated the nutritional requirements and growth rates of strains lacking two of the three *trpE* genes. Interestingly, we found that the *trpE1/trpE3* double mutant (Δ SCO2117::apr Δ SCO2043::spec) was a tryptophan auxotroph. (The auxotrophy was evident in the fact that all steps in the construction of strain B1682 required bacterial cultivation on media supplement with tryptophan). Those observations indicated that at least one of the two genes (*trpE1* and *trpE3*) is

required to ensure sufficient levels of tryptophan to support the growth of *S. coelicolor*. Moreover, they indicated that *trpE2* gene does not encode the housekeeping anthrilate synthase and suggested that may have a special role in tryptophan biosynthesis.

4.3.2 Deletion of *trpE* genes differently decreased the CDA production in *S. coelicolor*.

Given that the decapeptide lactone core of CDA contains two tryptophan units (Fig. 4.1), we proposed that less CDA would be produced in the *trpE* null strains compared to the wild-type strain. To test our hypothesis, a biological CDA bioassay was performed to compare the CDA production in wild-type *S. coelicolor* and the *trpE* null strains. The assay was carried out by growing the strains on Oxoid Nutrient Agar and then overlaying with soft Nutrient Agar containing *Bacillus mycoides* and $\text{Ca}(\text{NO}_3)_2$ at a concentration to give 12 mM throughout the plate (19). A zone of inhibition is diagnostic for CDA. The inhibition zone can be attributed solely to CDA because it is absent when $\text{Ca}(\text{NO}_3)_2$ is omitted. As evidenced by differences in the area of the zones of inhibition, all the three *trpE* null strains produced various degrees of less CDA than the wild-type strain (Fig. 4.3). Interestingly, the strain that exhibited the most severe defect in CDA production was the *trpE3* null strain. Despite the fact that the *trpE2* gene is proximal to the CDA biosynthetic, its disruption had the least impact on CDA biosynthesis. It was also found that provision of neither the wild-type *trpE* locus nor *trpE* under the control of the constitutive *ermE** promoter was able to completely suppress the CDA-defective phenotypes of the *trpE* null strains (Fig. 4.4).

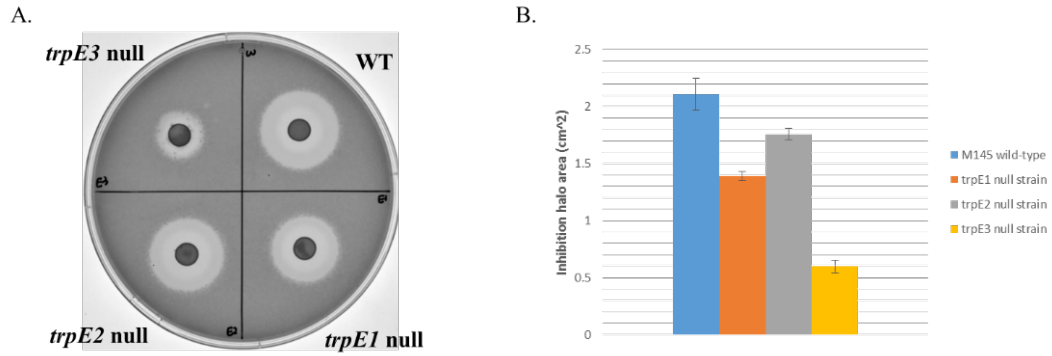


Figure 4.3 CDA production of wild-type *S. coelicolor* M145 and *trpE* null strains. CDA biosynthesis was assessed and compared using a standard bioassay in which soft-agar containing CDA-susceptible bacterium, *Bacillus mycoides* in soft agar inoculated over *S. coelicolor*. (A) Representative bioassay. (B) Measured zones of inhibition in experiments performed in triplicate. Each column stands for one tested strain and colored in a different color as indicated. The standard deviations are reported as error bars. Deletion of *trpE* genes all decreased the production of CDA, among which *trpE3* is the most with 70% reduction.

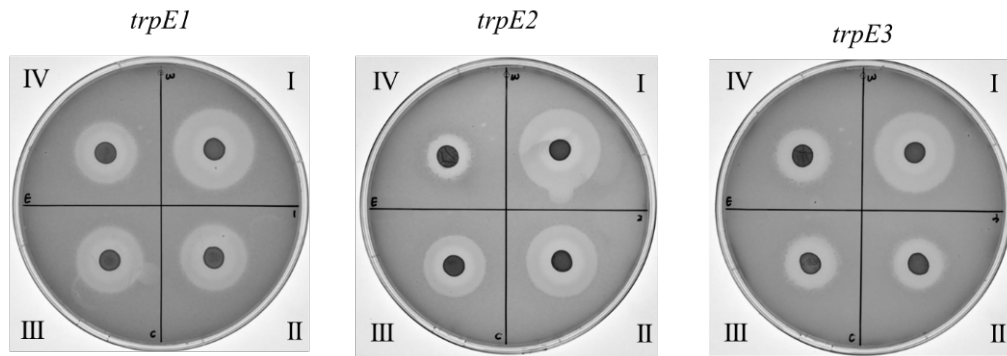


Figure 4.4 CDA production of *S. coelicolor* *trpE* null strains after the introduction of either the wild-type *trpE* locus or *trpE* under the control of the constitutive *ermE** promoter. In the clockwise direction of each plate shown above: (I) wild-type *S. coelicolor* M145 (II) *trpE* null strains. (III) *trpE* null strains harboring plasmids with the cognate native *trpE* loci. (IV) *trpE* null strains with the cognate *trpE* ORFs under the *ermE** promoter. The phenotypes of all the *trpE* null strains were not completely suppressed via the genetic complementation.

As disruption of the *trpE* genes in *S. coelicolor* could not be complemented genetically, we added the direct precursor, either anthranilate (the direct product of TrpE enzyme) or tryptophan into the growth medium. These experiments were regarded as “chemical complementations”. Because high anthranilate exhibited toxicity to *S. coelicolor* above 4 mM (data not shown), we used it in the chemical complementation experimentations at concentrations ranging from 0.5 to 3 mM. We found that, for both wild-type *S. coelicolor* and the *trpE* null strains, the addition of anthranilate increased the production of CDA. In particular, compared to wild-type *S. coelicolor*, higher concentration (3 mM) of anthranilate promoted more CDA production in the *trpE* null strains especially the *trpE3* null by more than 100% (Fig. 4.5A). However, this increment of CDA production was still not able to make up the CDA defect caused by disruption of *trpE* genes. A similar trend was observed when tryptophan was added into the growth medium in concentrations ranging from 0.5 – 37.5 mM (Fig. 4.5B).

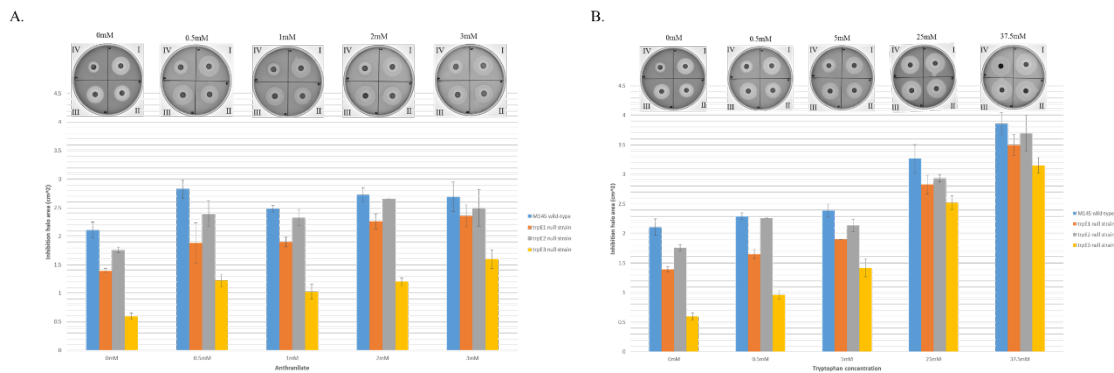


Figure 4.5 CDA production of *S. coelicolor* strains when the growth medium were supplemented with various concentrations of anthranilate or tryptophan. **(A)** Anthranilate was added into growth medium to give various final concentrations (0.5 mM, 1 mM, 2 mM and 3 mM). Addition of anthranilate increased the production of CDA for both wild-type *S. coelicolor* M145 as well as the *trpE* null strains. However, provision of anthranilate did not completely suppress the phenotype of the *trpE* null strains. **(B)** The tested final concentrations of tryptophan are 0.5 mM, 5 mM, 25 mM and 37.5 mM. Lower concentrations of tryptophan (0.5 mM and 5 mM) promoted more production of CDA in the *trpE* null strain than the wild-type strain.

Higher concentrations of tryptophan (25mM and 37.5mM) induced more production of CDA for both wild-type *S. coelicolor* and the *trpE* null strains in an obvious manner.

To further assess the contribution of the *trpE* genes to CDA production, each *trpE* gene was cloned into pIJ10257 for expression under the control of the constitutive *ermE** promoter and subsequently introduced into the *S. coelicolor* wild-type M145 strain. Interestingly, we found that *S. coelicolor* produced more CDA only upon the over-expression of *trpE3* gene, which is consistent our observation that disruption of *trpE3* gene had the most injurious effect on CDA production (Fig. 4.6).

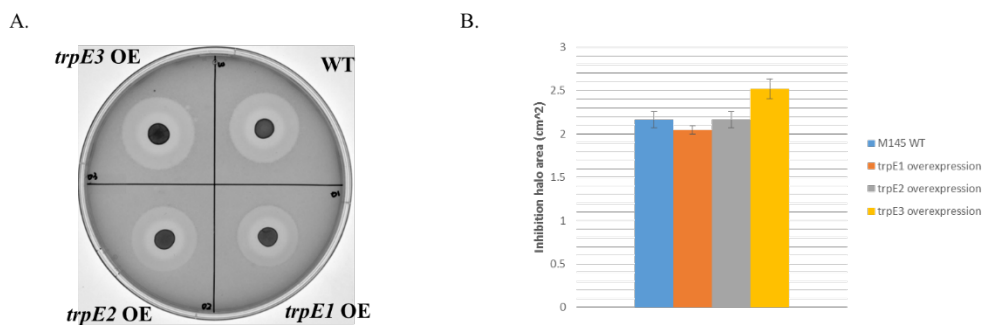


Figure 4.6 CDA production by wild-type *S. coelicolor* M145 strain upon the over-expression of the *trpE* genes. All *trpE* ORFs were introduced into *S. coelicolor* M145 wild-type under the *ermE** promoter for equalized expression *in vivo*. Antibiotic measurements were assessed and compared using a standard bioassay as described previously. (A) Representative bioassay. (B) Inhibition halo areas of triplicated experiments. Each column stands for one tested strain and colored in a different color as indicated. Standard deviation is reported for each bar. Only the overexpression of *trpE3* gene increased the production of CDA in *S. coelicolor* moderately.

4.3.3 The three *trpE* genes are transcribed differently

To explore the possibility that the *trpE* genes are differentially expressed, we investigated the *trpE* genes transcription profiles along with the growth curve and compared the transcription profiles

with that of a gene encoding a CDA biosynthetic enzyme using RT-PCR. Specifically, we analyzed the transcription of *trpE1*, *trpE2*, *trpE3* and *cdaPSI* (the gene encoding the largest nonribosomal peptide synthetase in CDA biosynthesis). Total mRNA was isolated at three time points representing different phases of *S. coelicolor* growth. They are 18 h (lag phase), 24 h (exponential phase) and 48 h (stationary phase). We found that the transcriptions of *trpE1*, *trpE2*, *trpE3* and *cdaPSI* were all growth-phase dependent (Fig. 4.7).

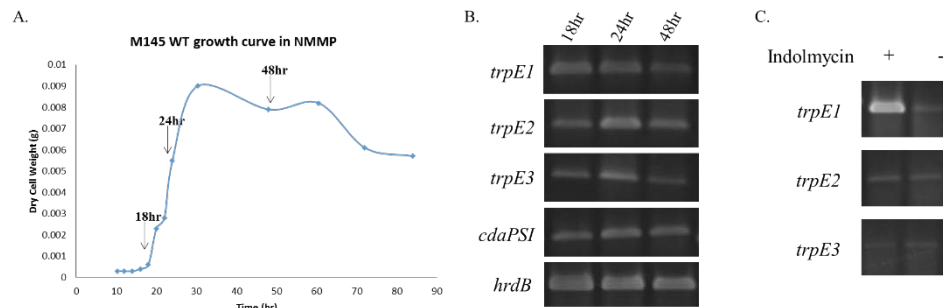


Figure 4.7 Expression of the *trpE* genes and CDA biosynthetic gene in wild-type *S. coelicolor* M145. RT-PCR was used to detect the transcription of the *trpE* genes and *cdaPSI*, one of CDA biosynthetic genes. The *hrdB* control transcript was detected in all of the RNA sample. The *hrdB* gene, which encodes a vegetative sigma factor, is constitutively transcribed. (A) Growth curve of wild-type *S. coelicolor* M145. Total RNA was isolated at three indicated time points, which represent three different growth phases, namely lag phase (18 h), exponential phase (24 h) and stationary phase (48 h). (B) The transcription profile of the *trpE* genes and *cdaPSI* during the growth of wild-type *S. coelicolor*. 1800 ng RNA template was employed for all the RT-PCRs. The transcriptions of *trpE* genes and *cdaPSI* were growth-phase dependent. (C) Transcriptional response of *trpE* genes to indolmycin, a tryptophanyl tRNA-synthetase inhibitor, which can trigger a physiological response that is analogous to tryptophan starvation by reducing the levels of charged tryptophanyl-tRNA. 1500 ng RNA template was employed for all the RT-PCRs. Indolmycin up-regulated the transcription of *trpE1*, whereas it had no influence on the transcription of *trpE2* and *trpE3*.

In nearly all microorganisms, anabolism of tryptophan is regulated through feedback inhibition of anthranilate synthase (TrpE enzyme) (17,28). For example, Crawford and co-workers reported that

trpE(G) gene in *Rhizobium meliloti* was regulated by feedback inhibition via ribosome-mediated attenuation (29). In *S. venezuelae*, it was found that feedback inhibition of *trpEG* is mediated by a ribosome-mediated attenuator (17). Curiously, based on genome-wide bioinformatics analyses in actinobacteria, Seliverstov *et al.* speculated that transcription of *trpE1* in *S. coelicolor* could be controlled by a ribosome-mediated attenuator (30). To test the hypothesis that the *trpE1* was regulated through feedback repression, RT-PCR was performed to assess changes in *trpE1* transcription in response to intracellular levels of charged tryptophanyl-tRNA (tRNA^{Trp}). To perturb levels of charged tRNA^{Trp} in a fashion mimicking tryptophan starvation, we used the competitive tRNA^{Trp} synthetase inhibitor indolmycin as a chemical tool. We found a robust transcription level of *trpE1* when the culture was treated with indolmycin, while there was only basal level transcription of *trpE1* for untreated culture (Fig. 4.7C). This observation is consistent with attenuator-mediated regulation proposed by Seliverstov *et al.* In parallel, *trpE2* and *trpE3* transcription were also assessed from the same RNA samples. However, induction of gene expression was observed for neither of the two genes when *S. coelicolor* was treated with indolmycin (Fig. 4.7C).

4.3.4 Identification of the *trpE1* transcription start site revealed the presence of a putative ribosome-mediated transcriptional attenuator

The observed indolmycin-induced transcription of *trpE1* was consistent with the bioinformatic prediction that it was regulated by a *cis*-acting ribosome-mediated attenuator. To further investigate the transcription of *trpE1* and confirm the presence of the attenuator, we used 5' rapid amplification of cDNA ends (5'-RACE) to map the transcription start site of the *trpE1* gene. The transcription start site of *trpE1* was mapped at -163 nucleotide relative to its predicated translational initiation site (Fig. 4.8A). Within the 5' leader of *trpE1*, we identified a small ORF encoding 19 amino acids including three adjacent Trp residues. The translational initiation site of the small ORF is coincident with the *trpE1* transcription start site. In addition, the sequence contains inverted repeats that are

capable of forming mutually exclusive RNA secondary structures (Fig. 4.8B). In one of the hypothetical RNA secondary structures, stem 3 and stem 4 bind to each other and form a hairpin (Fig. 4.9A). As the hairpin is G+C rich and immediately upstream of an A+U-rich tract, the 3:4 hairpin resembles a *rho*-independent transcription terminator. By analogy to well-characterized transcription attenuators, the 3:4 hairpin is likely to mediate premature transcription. In the other proposed RNA secondary structure, stem 3 pairs with stem 2 to form the 2:3 hairpin (Fig. 4.9B). As this structure precludes the formation of the *rho*-independent terminator, it is likely that the 2:3 hairpin is an antiterminator structure permissive for the transcription of the *trpE1* ORF. This organization of attenuator elements is similar to what has been observed in other characterized *trp* attenuators identified either in *E. coli* (31) or in *S. venezuelae* (17).

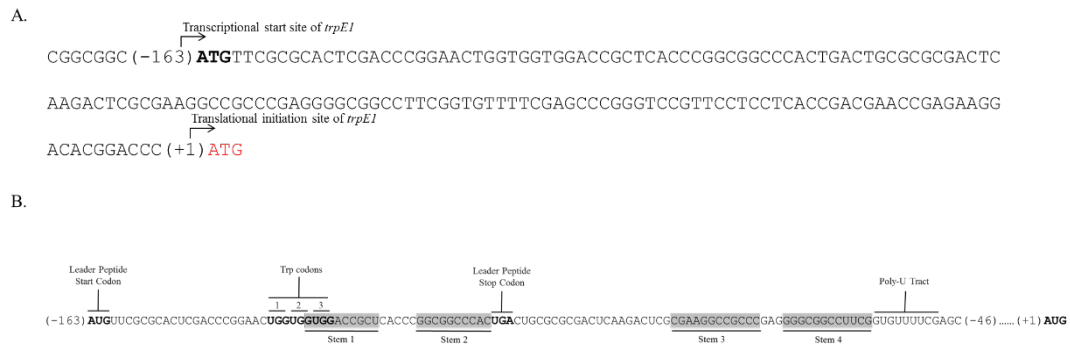


Figure 4.8 Analyses the region upstream of *trpE1*. (A) The overall sequence of the upstream region of *trpE1* coding strand, which indicates the transcriptional start site of *trpE1* (in bold) and the translational initiation site of *trpE1* (colored in red). (B) An abridged sequence of the 5' leader of the *trpE1* gene in *S. coelicolor* with various attenuator elements highlighted, including the leader peptide start and stop codons, the stem sequences (shaded in grey) and the tryptophan codon (UGG). Stem 2 and stem 4 compete to form hairpin structure with stem 3.

The presumptive switch between the transcriptional terminator and antiterminator structure is mediated by the translation of the small ORF including three tryptophan codons in succession. When there is sufficient charged tRNA^{Trp} in the bacteria cell to allow rapid completion of synthesis

of the leader peptide, the translating ribosome moves along sequences corresponding to stem 1 and stem 2, which promote the formation of the 3:4 hairpin (the apparent *rho*-independent terminator). When there is deficiency of charged tRNA^{Trp}, the stalling of the ribosome at tryptophan codon physically sequesters stem 1 and permit the formation of 2:3 hairpin (the RNA antiterminator structure). The stalling precludes the formation of terminator (3:4 hairpin). Therefore, the ribosome-mediated transcriptional attenuation model can explain the transcriptional response of *trpE1* to indolmycin (inhibitor of tRNA^{Trp} synthetase [TrpRS]). Specifically, the inhibition of TrpRS by indolmycin decreased intracellular levels of charged tRNA^{Trp}. Under this condition, the RNA antiterminator structure formed, leading to the transcription of the *trpE1* ORF.

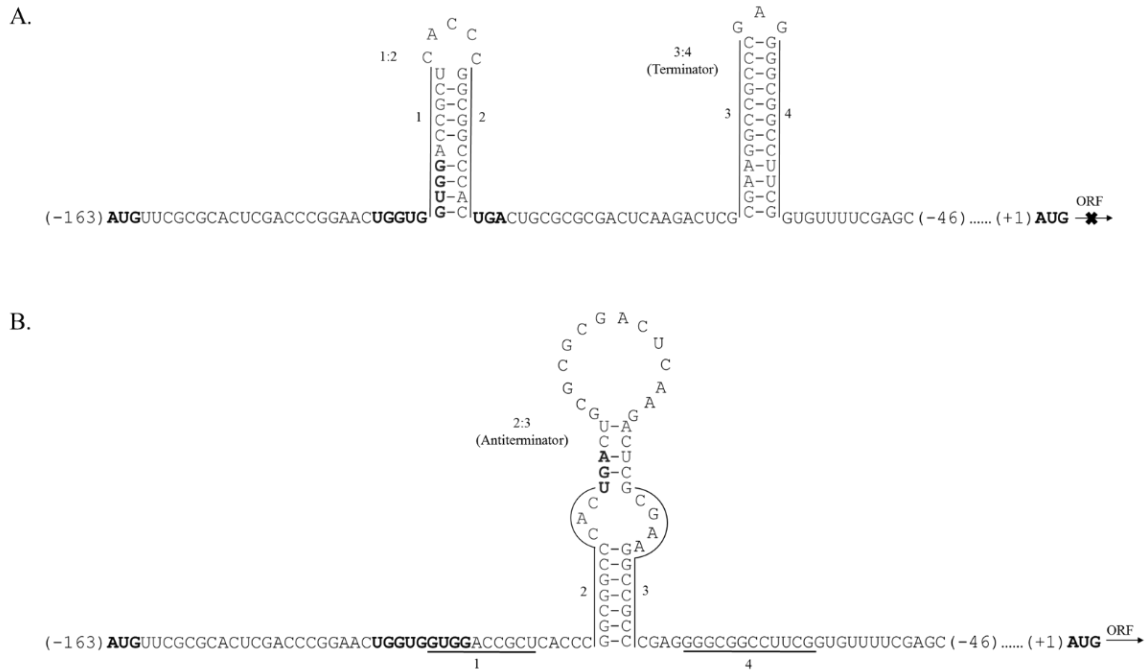


Figure 4.9 Schematic representation of the mutually exclusive secondary structures formed by the 163-nucleotide leader of the *trpE1* transcript. The start codon (AUG), tryptophan codon (UGG), and stop codon (UGA) of the *trpE1* leader peptide ORF and the start codon (AUG) of the *trpE1* ORF are in bold. (A) The proposed terminator structure, in which stem 1 binds with stem 2 to form hairpin and stem 3 binds with stem 4 to form hairpin. (B) The proposed antiterminator structure, in which stem 3 binds with stem 2 and form hairpin.

4.3.5 Missense point mutations within the *S. coelicolor trpE1* leader deregulate transcription

The presence of the small ORF in the *trpE1* leader is the central element in our prediction that the *trpE1* gene is under regulation of ribosome-mediated transcriptional attenuator. Our model for antitermination is based on the indolmycin-induced stalling of the ribosome at the tryptophan codons of the leader peptide ORF. Thus, we hypothesized that missense point mutation affecting this small ORF especially the tryptophan codons would deregulate the transcription of *trpE1*. In addition, there are three adjacent tryptophan codons. It was reported that the number of regulatory codons in the leader peptide ORF determines the sensitivity of the attenuator to amino acid availability (32). On this basis, we would also like to determine the relative impact of the three tryptophan codons on the leader's sensitivity to changes in charged tRNA^{T_{rp}} availability.

To test our hypotheses, we performed site-directed mutagenesis on the start codon of the small ORF as well as the three tryptophan codons. Variants of the *trpE1* leader with targeted point mutation(s) were cloned into the reporter plasmid, pMU1, which is specifically optimized for utilization in high G+C bacteria like *Streptomyces* species. The plasmid carries genes, *luxAB*, encoding the luciferase enzyme and also genes, *luxCDE*, encoding enzymes that catalyze the biosynthesis of decanol, the substrate of the luciferase enzyme (26). Fusion of the WT and mutated *trpE1* leaders to the luciferase genes enable a bioluminescence-based, quantifiable readout of transcriptional activity. All the reported constructs containing mutation(s) in the *trpE1* leader were introduced into *S. coelicolor* M600 (entry 3-9). The empty vector without the leader (pMU1) and the vector with the wild-type *trpE1* wild-type leader (pJS361) were also transformed into the same strain to serve as the negative (B798) and positive (B783) controls, respectively. The constructed reporter systems with various missense point mutation(s) were summarized in Table 4.4. Transcriptional activation is conveyed by the bioluminescent readout. Bioluminescence was measure through an Envision Multilabel Plate Reader. The amount of bioluminescence emitted by strain B798 harboring the empty vector was measured to establish the background signal to noise

ratio. The amount of bioluminescence emitted by strain B783, which contains the reporter construct with the unmodified *trpE1* leader (pJS361), grown in the presence of indolmycin was measured to define maximal signal. All strains were grown as surface-grown cultures in 96-well plates at 30°C for three days.

Table 4.4 Bioluminescence of *S. coelicolor* M600 WT with various reporters

Entry	strain	plasmid	Leader mutation	Mutated Codon(s)	Bioluminescence signal	
					(-) indolmycin	(+) indolmycin
1	B783	pJS361	WT <i>trpE1</i> leader		50	56
2	B798	pMU1	no <i>trpE1</i> leader		35	76
3	B784	pJS362	A(-163)G	start codon	40	37
4	B785	pJS363	G(-132)C	1-2-X (Trp)	50	261
5	B786	pJS364	G(-135)C	1-X-3 (Trp)	35	599
6	B787	pJS365	G(-138)C	X-2-3 (Trp)	40	3287
7	B788	pJS367	G(-132-135)C	1-X-X (Trp)	32	40
8	B789	pJS368	G(-135-138)C	X-X-3 (Trp)	38	134
9	B790	pJS369	G(-132-135-138)C	X-X-X (Trp)	43	47

We first introduced a missense mutation (AUG to GUG) that converted the canonical methionine start codon to a leucine codon, at which translation cannot be initiated in *S. coelicolor*. It was found that the corresponding strain B784 did not experience an upregulation in bioluminescence when grown on medium supplemented with indolmycin.

The tryptophan codon (UGG) was replaced by serine codons (UCG). Given the presence of three adjacent tryptophan codons, we initially introduced missense point mutations at each tryptophan codon individually. Interestingly, the strains B787, B786, and B785, which contains the single point mutation of the first, second, and third tryptophan codons, respectively, revealed different levels of bioluminescent upregulation in the presence of indolmycin. The trend is that the closer the tryptophan codon is to the start codon of the leader peptide, the less influence it has over luciferase activity. Strain with the third tryptophan (closest to the 1:2 hairpin) mutated showed the least activity in the presence of indolmycin (entry 4), while mutation of the first tryptophan (farthest

from the 1:2 hairpin) showed the greatest activity with an 80-fold increment in luciferase activity (entry 6). The mutation of the second tryptophan had a shift in between these two extremes with a nearly 20-fold increment in bioluminescence (entry 5). We also constructed reporter systems with two tryptophan codons mutated to serine codons, strain B788 (only contains the first tryptophan codon) and strain B789 (only contain the third tryptophan codon). Again, the two strains showed different luciferase activity when treated with indolmycin. There was almost no signal change in strain B788, whereas strain B789 exhibited a 3-fold increment, which is slightly higher than the negative control. Finally, mutations introduced at all the three tryptophan codons within the *trpEI* attenuator completely abolished its ability of sensing intracellular levels of the charged tRNA^{Trp} and there was no significant increment in bioluminescence in the presence of indolmycin.

Based on these observations, we concluded that the tryptophan codon closest to the hairpins exerts the greatest control over their formation. However, there was still one minor problem left. The positive control strain B783 exhibited unexpectedly low luciferase activity in the presence of indolmycin, which made the direct comparisons between the strains containing mutated leaders and the strain with the wild-type leader difficult. To address this issue, this positive control strain is being reconstructed.

4.4 Discussion

Though bacterial genomes often have paralogous genes, their physiological roles are not always clear. The Sello group has long-standing interests in paralogous genes that are important in antibiotic resistance (8-10). In this case, we were focused on paralogous genes in *S. coelicolor* that appeared to have a role in metabolism. Specifically, our focus was genetic loci containing genes encoding enzymes for tryptophan biosynthesis. The tryptophan biosynthetic pathway is generally defined as an unbranched pathway that begins with chorismate and produces tryptophan as a substrate for general protein synthesis (3). The pathway appears to have evolved only once because

the enzymatic reactions involved in Trp biosynthesis are conserved in virtually all the microorganisms with this synthetic capacity (3,33). It has been shown that the *trp* operon must have been present in early prokaryote ancestors and the putative common ancestral whole-pathway *trp* operon is consisting of seven genes with a consensus gene order of *trpEGDCFAB* in bacteria (Fig. 4.10A). The current variety of *trp* operons is likely the result of several deduced evolutionary events including the fusion of *trpC* and *trpF*, separation of *trpB/trpA*, insertion between *trpG* and *trpD*, insertion between *trpD* and *trpC·trpF*, separation of *trpE/trpG*, fusion of *trpG* and *trpD*, and the complete loss of *trp* operon (3).

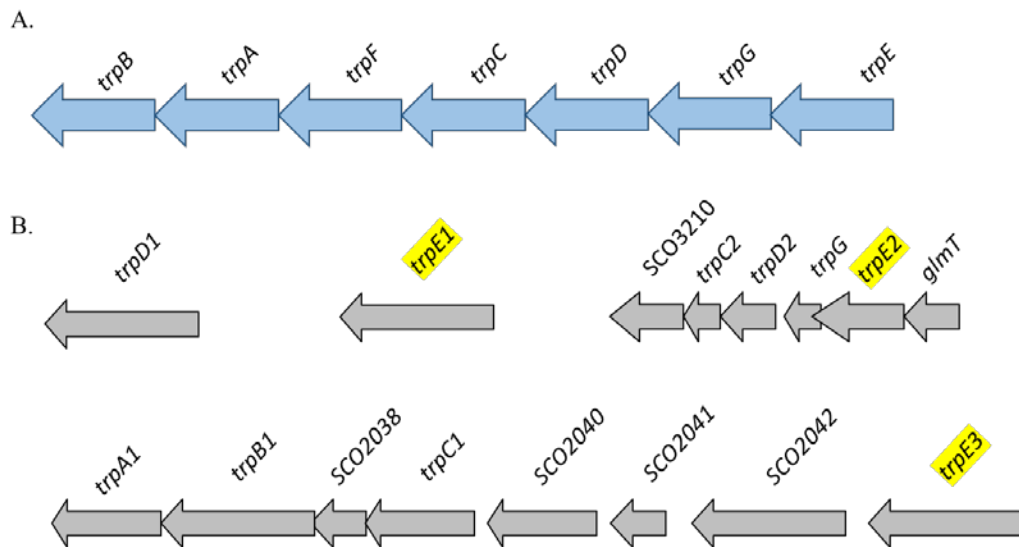


Figure 4.10 Representative *trp* operon. (A) Deduced common ancestral whole-pathway *trp* operon. (B) The four *trp* operons identified in *S.coelicolor*. Two of the operons are organized in polycistronic units and two of them are single genes.

In *Streptomyces*, the Trp metabolism can be considered divergently branched, with Trp being guided to different molecular fates, such as protein synthesis and antibiotic production (3,33,34). It was reported that *S. coelicolor* possesses several *trp* operons, which was further confirmed as four loci based on the analyses of the complete genome sequence (Fig. 4.10B). The primary *trpE3C1BA*

(SCO2036-2043) biosynthetic operon can be identified by phylogenetic analysis. There are also free-standing copies of *trpE* (SCO2117, regarded as *trpE1*) and *trpD* (SCO2147, regarded as *trpD1*) in the organism. The last operon *trpE2GD2C2* was found nested within a large cluster of genes that dictate synthesis of CDA and this operon was proposed to have a specialized role in antibiotic production that is unique to *Streptomyces* (3).

From the experiments outlined in this chapter, we conclude that *trpE3* is engaged in primary Trp biosynthesis and is insensitive to feedback inhibition. Though we initially predicted *trpE2* had a role ensuring that *S. coelicolor* had sufficient tryptophan for CDA biosynthesis, we found that deletion of the operon *trpE2GD2C2* only slightly reduced the production of CDA, whereas disruption of *trpE3* dramatically resulted in CDA production defect. The observation negates the essential contribution of *trpE2GD2C2* to the production of CDA. As the gene organization within this *trp* loci is the same as the deduced common *trp* ancestor, it is possible that this redundant *trp* operon originated from horizontal gene transfer (HGT). This partial-pathway *trp* operon was inexplicably inserted into the chromosome of *S. coelicolor* and located next to CDA biosynthetic gene cluster. Our analyses indicate that this operon is not functionally related to the production of CDA. The origin of *trpE2GD2C2* by LGT has been mentioned (35), but a detailed analysis has not yet been done.

The *trpE3* gene, member of the primary *trp* operon, contributed the most to the production of CDA. The transcription of both *trpE2* and *trpE3* were growth-phase dependent and not sensitive to feedback inhibition. The transcription of single gene *trpE1* was found to be under the regulation of a ribosome-mediated transcriptional attenuator in the 5' leader of its mRNA region. Thus, we hypothesized that *trpE1* should be engaged in primary Trp biosynthesis and is sensitive to feedback inhibition. This proposal is also supported by the presence of three *trp* codons in the *trpE1* leader because attenuators with higher numbers of regulatory codons are more sensitive to amino acid

availability (or changes in the availability of charged tRNA) than are those with fewer numbers of regulatory codons (36).

As discussed above, an excess of tryptophan is required for the biosynthesis of many secondary metabolites. In the interest of genetic engineering of *S. coelicolor* for the over-production of CDA, two different strategies for increasing intracellular tryptophan concentration have been adopted. One is to overexpress tryptophan biosynthetic genes and the other is to impair tryptophan catabolism genes. It has been reported that alteration of tryptophan catabolism occurred in the $\Delta kynU$ mutant resulted in an increased production of the CDA (34). Additionally, *trpB* and *trpA* overexpressing mutants were found to produce early CDA compared to the wild-type strain (37). In this work, we demonstrated that overexpression of the *trpE3* gene caused the production of a higher amount of CDA. In addition, we found that addition of anthranilate and tryptophan in the growth medium promoted the production of CDA. Intuitively, the stimulatory effect on CDA production was due to a direct precursor feeding. However, a recent study conducted by Gallo *et al.* concerning the regulatory effect of tryptophan on gene expression revealed that Trp supplementation is beneficial for actinorhodin production as well. Thus, they assessed that the effect of Trp on CDA production should not be related to the direct increment of precursor supply in the medium, but it may be generally due to the activation of regulatory pathways modulating secondary metabolism in *S. coelicolor*. However, the molecular mechanism directly link Trp metabolism to CDA production and whether the regulation could occur at the enzymatic activity level are still unclear and further experiments have to be performed (33).

Finally, given that the *trpE1* leader contains three *trp* codons, we validated the relative importance of each one in the transcriptional regulation. It was found that the tryptophan codon closest to the hairpin exerts the greatest control over their formation. As shown in our proposed terminator structure, the third *trp* codon is part of the 1:2 hairpin (Fig. 4.9A). This observation also supports

our proposed attenuator structures. Overall, this study provides insight into the functions of paralogous *trpE* genes and how they are regulated in bacteria.

4.5 Scientific Contributions

In this chapter, Ethan Hickey 12' (Brown University) constructed *S. coelicolor* strain B1670 (*trpE1* null). Gabrielle Grandchamp 13' (Brown University) constructed *S. coelicolor* strain B1678 (*trpE3* null) and strain B1682 (*trpE1* and *trpE3* double deletion). Dr. James Vecchione (Department of Chemistry, Brown University) performed 5'-RACE experiment to map the transcription start site of *trpE1*. Jenica Shipley 11' (Brown University) did the site-directed mutagenesis in *trpE1* leader, created the reporter constructs, and performed the bioluminescence assay with the reporter constructs. All the other experiments and data analyses were performed by Rui Zhang.

4.6 References

1. Zhang, J. (2012) Genetic redundancies and their evolutionary maintenance. *Adv Exp Med Biol*, **751**, 279-300.
2. Wagner, A. (1996) Genetic redundancy caused by gene duplications and its evolution in networks of transcriptional regulators. *Biol Cybern*, **74**, 557-567.
3. Xie, G., Keyhani, N.O., Bonner, C.A. and Jensen, R.A. (2003) Ancient origin of the tryptophan operon and the dynamics of evolutionary change. *Microbiol Mol Biol Rev*, **67**, 303-342, table of contents.
4. Bentley, S.D., Chater, K.F., Cerdeno-Tarraga, A.M., Challis, G.L., Thomson, N.R., James, K.D., Harris, D.E., Quail, M.A., Kieser, H., Harper, D. *et al.* (2002) Complete genome sequence of the model actinomycete *Streptomyces coelicolor* A3(2). *Nature*, **417**, 141-147.
5. Walsh, C. (2000) Molecular mechanisms that confer antibacterial drug resistance. *Nature*, **406**, 775-781.
6. Amrik S. Thiara, E.C. (1988) Cloning and characterization of a DNA gyrase B gene from *Streptomyces sphaeroides* that confers resistance to novobiocin. *EMBO J*, **7**, 2255-2259.

7. Gominet, M., Seghezzi, N. and Mazodier, P. (2011) Acyl depsipeptide (ADEP) resistance in *Streptomyces*. *Microbiology*, **157**, 2226-2234.
8. Compton, C.L., Schmitz, K.R., Sauer, R.T. and Sello, J.K. (2013) Antibacterial activity of and resistance to small molecule inhibitors of the ClpP peptidase. *ACS Chem Biol*, **8**, 2669-2677.
9. Vecchione, J.J. and Sello, J.K. (2008) Characterization of an inducible, antibiotic-resistant aminoacyl-tRNA synthetase gene in *Streptomyces coelicolor*. *J Bacteriol*, **190**, 6253-6257.
10. Vecchione, J.J. and Sello, J.K. (2009) A novel tryptophanyl-tRNA synthetase gene confers high-level resistance to indolmycin. *Antimicrob Agents Chemother*, **53**, 3972-3980.
11. Maurer, K.H., Pfeiffer, F., Zehender, H. and Mecke, D. (1983) Characterization of two glyceraldehyde-3-phosphate dehydrogenase isoenzymes from the pentalenolactone producer *Streptomyces arenae*. *J Bacteriol*, **153**, 930-936.
12. Schneider, D., Bruton, C.J. and Chater, K.F. (2000) Duplicated gene clusters suggest an interplay of glycogen and trehalose metabolism during sequential stages of aerial mycelium development in *Streptomyces coelicolor* A3(2). *Mol Gen Genet*, **263**, 543-553.
13. Yanofsky, C. (2007) RNA-based regulation of genes of tryptophan synthesis and degradation, in bacteria. *RNA*, **13**, 1141-1154.
14. Yanofsky, C. and Crawford, I.P. (1987) The tryptophan operon. In *Escherichia coli* and *Salmonella typhimurium*. *Cellular and molecular biology* (eds. F.C. Neidhardt et al., American Society for Microbiology, Washington D.C.), 1452-1472.
15. Yanofsky, C., Miles, E.W., Kirschner, K. and Bauerle, R. (1999) TRP operon. *The encyclopedia of molecular biology* (ed. T.C. Creighton, John Wiley & Sons, New York), **4**, 2676-2689.
16. Hojati, Z., Milne, C., Harvey, B., Gordon, L., Borg, M., Flett, F., Wilkinson, B., Sidebottom, P.J., Rudd, B.A.M., Hayes, M.A. et al. (2002) Structure, Biosynthetic Origin, and Engineered Biosynthesis of Calcium-Dependent Antibiotics from *Streptomyces coelicolor*. *Chemistry & Biology*, **9**, 1175-1187.
17. Lin, C., Paradkar, A.S. and Vining, L.C. (1998) Regulation of an anthranilate synthase gene in *Streptomyces venezuelae* by a trp attenuator. *Microbiology*, **144** (Pt 7), 1971-1980.
18. Sambrook J., F.E.F., T. Maniatis. (1989) *Molecular cloning: a laboratory manual* (2nd ed, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N. Y.).
19. Kieser T, B.M., Buttner MJ, Chater KF, Hopwood DA. (2000) *Practical Streptomyces Genetics* (John Innes Foundation, Norwich, England).
20. Hasuoka, A., Nakayama, Y., Adachi, M., Kamiguchi, H. and Kamiyama, K. (2001) Development of a stereoselective practical synthetic route to indolmycin, a candidate anti-*H. pylori* agent. *Chem Pharm Bull* (Tokyo), **49**, 1604-1608.

21. Gust, B., Challis, G.L., Fowler, K., Kieser, T. and Chater, K.F. (2003) PCR-targeted *Streptomyces* gene replacement identifies a protein domain needed for biosynthesis of the sesquiterpene soil odor geosmin. *Proc Natl Acad Sci U S A*, **100**, 1541-1546.
22. Jones, G.H., Paget, M.S.B., Chamberlin, L. and Buttner, M.J. (1997) Sigma-E is required for the production of the antibiotic actinomycin in *Streptomyces antibioticus*. *Molecular Microbiology*, **23**, 169-178.
23. Gregory, M.A., Till, R. and Smith, M.C.M. (2003) Integration Site for *Streptomyces* Phage BT1 and Development of Site-Specific Integrating Vectors. *J Bacteriol*, **185**, 5320-5323.
24. Hong, H.J., Hutchings, M.I., Hill, L.M. and Buttner, M.J. (2005) The role of the novel Fem protein VanK in vancomycin resistance in *Streptomyces coelicolor*. *J Biol Chem*, **280**, 13055-13061.
25. Bibb, M.J., White, J., Ward, J.M. and Janssen, G.R. (1994) The mRNA for the 23S rRNA methylase encoded by the *ermE* gene of *Saccharopolyspora erythraea* is translated in the absence of a conventional ribosome-binding site. *Molecular Microbiology*, **14**, 533-545.
26. Craney, A., Hohenauer, T., Xu, Y., Navani, N.K., Li, Y. and Nodwell, J. (2007) A synthetic *luxCDABE* gene cluster optimized for expression in high-GC bacteria. *Nucleic Acids Res*, **35**, e46.
27. Datsenko, K.A. and Wanner, B.L. (2000) One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. *Proc Natl Acad Sci U S A*, **97**, 6640-6645.
28. Yanofsky, C. (2001) Advancing our knowledge in biochemistry, genetics, and microbiology through studies on tryptophan metabolism. *Annu Rev Biochem*, **70**, 1-37.
29. Bae, Y.M. and Crawford, I.P. (1990) The Rhizobium-Meliloti TrpE(G) Gene Is Regulated by Attenuation, and Its Product, Anthranilate Synthase, Is Regulated by Feedback Inhibition. *Journal of Bacteriology*, **172**, 3318-3327.
30. Seliverstov, A.V., Putzer, H., Gelfand, M.S. and Lyubetsky, V.A. (2005) Comparative analysis of RNA regulatory elements of amino acid metabolism genes in Actinobacteria. *BMC Microbiol*, **5**, 54.
31. Oxender, D.L., Zurawski, G. and Yanofsky, C. (1979) Attenuation in the *Escherichia coli* tryptophan operon: role of RNA secondary structure involving the tryptophan codon region. *Proc Natl Acad Sci U S A*, **76**, 5524-5528.
32. Johnston, H.M., Barnes, W.M., Chumley, F.G., Bossi, L. and Roth, J.R. (1980) Model for regulation of the histidine operon of *Salmonella*. *Proc Natl Acad Sci U S A*, **77**, 508-512.
33. Palazzotto, E., Renzone, G., Fontana, P., Botta, L., Scaloni, A., Puglia, A.M. and Gallo, G. (2015) Tryptophan promotes morphological and physiological differentiation in *Streptomyces coelicolor*. *Appl Microbiol Biotechnol*, **99**, 10177-10189.

34. Zummo, F.P., Marineo, S., Pace, A., Civiletti, F., Giardina, A. and Puglia, A.M. (2012) Tryptophan catabolism via kynurenine production in *Streptomyces coelicolor*: identification of three genes coding for the enzymes of tryptophan to anthranilate pathway. *Appl Microbiol Biotechnol*, **94**, 719-728.
35. Ryding, N.J., Anderson, T.B. and Champness, W.C. (2002) Regulation of the *Streptomyces coelicolor* calcium-dependent antibiotic by *absA*, encoding a cluster-linked two-component system. *J Bacteriol*, **184**, 794-805.
36. Yanofsky, C. (2003) Using studies on tryptophan metabolism to answer basic biological questions. *J Biol Chem*, **278**, 10859-10878.
37. Marineo S., G.A., Mandalà S., Bravatà V., Canepa A., Ledda M., Di Carlo E., Curcurù L. and Puglia A.M. (2007) Relation of primary metabolism on antibiotic production in *Streptomyces coelicolor*. *DBCS 2007*, 5°Congresso Annuale (poster abstract).