Metabolic Engineering for the Heterologous Biosynthesis of Erythromycin A and Associated Polyketide Products in *Escherichia coli*

A dissertation

submitted by

Haoran Zhang

In partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

in

Chemical Engineering

Tufts University

February, 2011

Advisor: Blaine Pfeifer

Abstract

The natural product erythromycin is a potent and widely used antibiotic for bacterial infections. In structure, erythromycin consists of one 14-membered polyketide ring, 6-*deoxy*-erythronolide B (6dEB), and two sugar residues, cladinose and desosamine. Erythromycin is natively produced by a soil-dwelling bacterium *Saccharopolyspora erythraea* that is fastidious in its growth requirements and hard to manipulate through genetic and metabolic engineering approaches. In this thesis study, the production of the erythromycin polyketide precursor 6dEB was optimized by engineering the *Escherichia coli* native metabolic pathways to support biosynthesis of the required substrates. Furthermore, heterologous biosynthesis of the erythromycin sugar groups cladinose and desosamine was achieved to produce the final product erythromycin A in *E. coli*. The reported titer for erythromycin A was ~10 mg/L. The entire 55kb erythromycin producing gene cluster was then systematically transferred to *E. coli* to enable the complete heterologous biosynthesis of erythromycin from propionate. Lastly, novel erythromycin associated polyketide analogs were produced by modification of the polyketide and sugar moiety biosynthetic pathways in *E. coli*. The resulting products contained altered polyketide or sugar groups, as directed by the rationally designed new biosynthetic pathway. Success establishes *E. coli* as a viable option for the heterologous production of erythromycin A and, more broadly, as a platform for the directed production of erythromycin associated polyketide analogs.

Acknowledgement

I would like to thank everyone at Tufts University who helped me with my academic growth during my PhD study.

I would especially like to thank my advisor, Professor Blaine Pfeifer, for his continuous support for my thesis research. In addition to the disciplinary acknowledge, I also learned a lot from him about how to be detail-oriented, organized, and big-picture-minded. All of these skills will become the foundation for my future career development.

I was fortunate to work with many excellent colleagues in Professor Pfeifer's group. Yong Wang taught me molecular biology and metabolic engineering research hand-by-band. Brett Boghigian had been working with me in the same laboratory throughout my PhD program. I also thank Guangquan Shi, Janelle Lavoie, Michael Pistorino, Jiequn Wu and Sunghee Park for their encouragement and help.

I am thankful to Professor Andrew Wright, Dr. Marcia Osburne, and Professor Kyongbum Lee for being on my thesis committee and providing many insightful suggestions. I cannot imagine how my thesis research would have proceeded without their input.

Last but not least, I would like to thank my wife Cuixian Yang and other family members for their consistent encouragement and support throughout the past 5 years. I would not have been through difficult times and achieved my academic accomplishments without their love and support.

Table of Contents

LIST OF TABLES

LIST OF FIGURES

Chapter 1: Background and Research Goals

1.1 Research Background

Natural products are a group of chemical substances derived from living organisms. The various biological activities associated with natural products have made them an important source for new drug discovery (*1-3*). It has been reported that natural products or semi-synthetic derivatives compose 34% of small-molecule new chemical entities (NCE's) between 1981 and 2006. In particular, 68% of the antibacterial NCE's and 54% of the anticancer NCE's were of natural origins (*4*). However, natural products usually possesses complex chemical structures, which makes total synthesis and chemical modifications difficult (*5, 6*). In addition, despite their pharmaceutical values, natural products are often generated by the secondary metabolism of native producers, which makes the yield and productivity unsatisfactory (*7, 8*).

Polyketides are an important group of natural products as they posses a remarkably wide range of chemical structures that are often associated with particular medicinal values (*9, 10*). So far, many bioactive polyketide compounds have been identified and utilized to treat a variety of human diseases. Erythromycin is one of the most famous members of the polyketides. Since its discovery in 1952, erythromycin has been widely used for bacterial infection treatments (*11, 12*). More importantly, many of its analogs have been produced and utilized as antibiotics with a broader spectrum. For example, Azythromycin, an erythromycin analog, is one of the world's best selling antibiotics (*13*).

Like many other polyketides, erythromycin suppresses the growth of susceptible microbes by binding to ribosomal RNA and inhibiting protein synthesis. Such binding

occurs at the 50s subunit of the 70s ribosomal RNA complex and results in dissociation of peptidyl transfer RNA from the ribosome during the protein synthesis process (*14, 15*). This mechanism is similar to that of other antibiotics. Resistance to erythromycin was first reported in 1959 (*16*). Later on, three erythromycin resistance mechanisms were discovered in different microorganisms, including target ribosome binding site alteration, antibiotic modification, and altered antibiotic transport (*17*).

Although enormous efforts have gone into the study of chemical synthesis of erythromycin (*18-20*), challenges still exist because of the structural complexity of this macrolide. As a result, the overall yield of erythromycin remains low and large-scale chemical synthesis is not feasible. Therefore, biosynthesis is currently the main method for erythromycin mass production. The soil-dwelling bacterium *Saccharopolyspora erythraea* is a native producer of erythromycin (*11*). The gene cluster responsible for erythromycin biosynthesis in *S. erythraea*'s chromosome has been fully sequenced and the biosynthetic pathway has been well-studied and established (*21-24*). In structure, erythromycin consists of a 14-membered lactone ring and two sugars, L-cladinose and D-desosamine. The polyketide core of erythromycin, 6-deoxyerythronolide B (6dEB), is synthesized by 6-deoxyerythronolide B synthases (DEBS) whose crystal structure has been established (*6, 25, 26*). The 6dEB biosynthesis process is shown in Figure 1.1.

6dEB polyketide synthase is a large (>2 MDa) multi-domain enzyme complex consisting of 6 modules. Each module has at least three domains, including a ketosynthase (KS) domain, an acyltransferase (AT) domain, and an acyl carrier protein (ACP) domain.

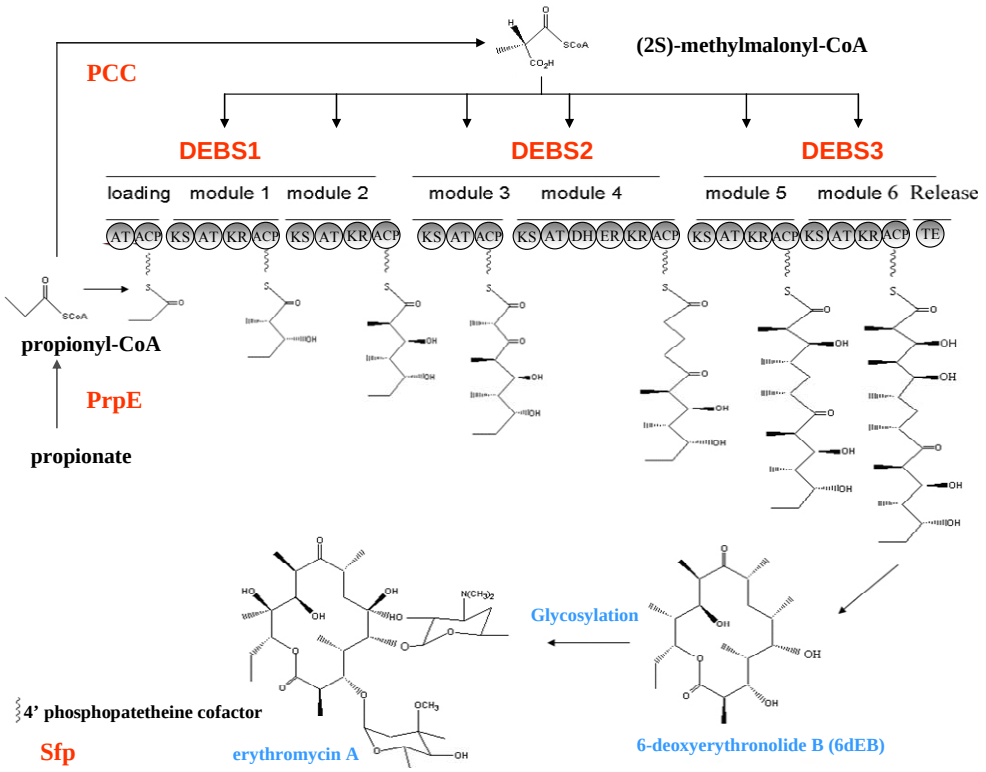


Figure 1.1 Biosynthesis of erythromycin polyketide precursor 6dEB.

The AT domain binds to specific substrates used for polyketide elongation. The ACP domain is responsible for binding the polyketide chain or intermediate during the biosynthesis process, and this domain needs to be modified by phosphopantetheinylation to be fully activated. The KS domain catalyzes the condensation reaction between polyketide chain intermediate and the extender unit. Some modules also contain a ketoreductase (KR) domain, a dehydratase (DH) domain, and an enoyl reductase (ER) domain, which are responsible for formation of certain functional groups during polyketide chain elongation. The last domain of erythromycin polyketide synthases is a thioesterase (TE) domain that catalyzes the cyclization of the final polyketide chain and release of the mature product 6-deoxyerythronolide B (6dEB) from the polyketide synthases complex. Substrates propionyl-CoA and (2S)-methylmalonyl-CoA are the

starter and extender units of polyketide biosynthesis, respectively. Propionyl-CoA can be produced from propionate by a propionyl-CoA synthetase (PrpE in *E. coli*), and can be further converted to (2S)-methylmalonyl-CoA by propionyl-CoA carboxylase (PCC).

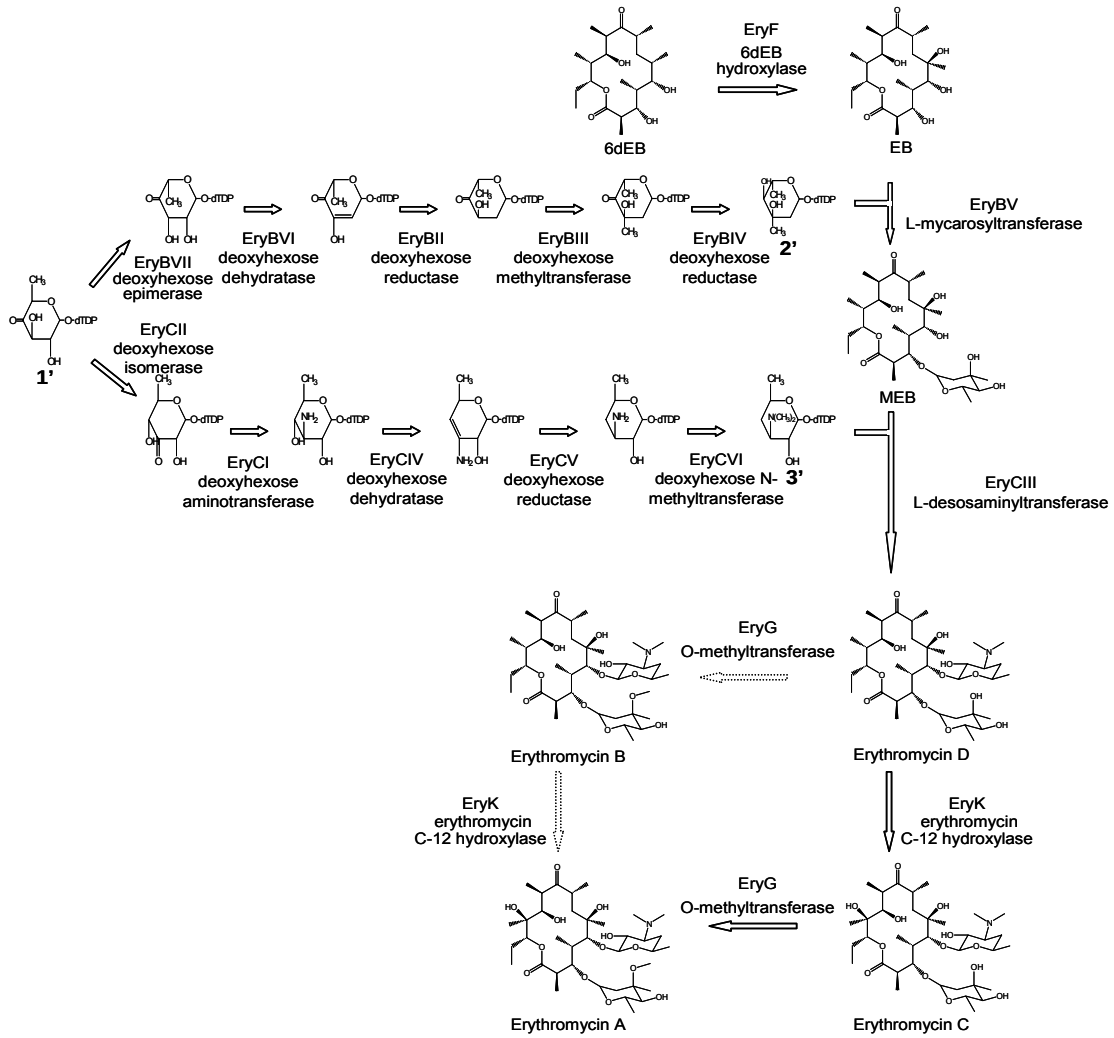


Figure 1.2 Schematic biosynthesis of erythromycin A. Highlighted in this Figure are the starting deoxysugar substrate (dTDP-4-keto-6-deoxy-hexose [**1'**], derived from glucose-1-phosphate), 6-deoxyerythronolide B (6dEB), erythronolide B (EB), dTDP-L-mycarose (**2'**), 3-mycarosyl-erythronolide B (MEB), dTDP-D-desosamine (**3'**), and the full names of the enzymes responsible for 6dEB to erythromycin A conversion. The final L-cladinose sugar attached to erythromycin A results from the methylation activity of EryG. The solid arrows present the expected preferred pathway from erythromycin D to erythromycin A in *S. erythraea*.

The polyketide 6dEB then undergoes a glycosylation process to become the active antibiotic erythromycin. Two sugar groups need to be attached at the C-3 and C-5 positions to accomplish the complete glycosylation. As shown in Figure 1.2, erythromycin sugar group biosynthesis starts from dTDP-4-keto-6-deoxyglucose (1'). Genes *eryBI-BVII* are responsible for formation and attachment of the first sugar group precursor mycarose (2') to make the intermediate 3-mycarosyl-erythronolide B (MEB). The mycarose sugar group is then modified by *eryG* to make the final sugar moiety cladinose. It should be noted that the presence of *eryBI* was not necessary for erythromycin production in the native producer *Saccharopolyspora erythraea*, and this enzyme's catalytic function is not clear (*27*). Genes ery*CI-CV* are responsible for formation and attachment of the second sugar moiety desosamine (3'). Gene *eryF* and *K* catalyze the hydroxylation of the polyketide structure at C-6 and C-12, respectively. The complete erythromycin biosynthesis from 6dEB is shown in Figure 1.2. Notably, the biosynthesis of erythromycin A through erythromycin C is preferred in the native producer *S. erythraea* (*28, 29*).

After years of genetic modification of the wild-type strain and continual improvement of the production process, the highest reported titer of erythromycin in *S. erythraea* reached 7 g/L (*30*). However, the native producer's fastidious nature and slow growth rate limit its utility for over-production of erythromycin. In addition, although many efforts have been dedicated to optimizing the production level in *S. erythraea* (*31-34*), lack of sophisticated DNA recombination techniques for this microorganism has hampered further improvement of the titer and the generation of erythromycin analogs with potentially novel bioactivity through genetic and metabolic engineering approaches.

Thus, establishing a heterologous biosynthesis of erythromycin in a more technically amenable system, such as *Escherichia coli*, is of great interest. In fact, numerous studies have been dedicated to the use of heterologous biosynthesis towards the production of complex natural products in recent years (*35-37*). For example, production of the erythromycin polyketide precursor in other hosts including *S. coelicolor* and *S. lividans* has already been achieved, although the production level is still below erythromycin titers associated with *S. erythraea* (*38, 39*).

When compared to *S. coelicolor* or *S. lividans*, *E. coli* offers unquestioned advantages in terms of growth kinetics and recombinant tools to aid in heterologous polyketide biosynthetic efforts. However, there are marked differences in cell morphology, physiology, metabolism, and regulatory mechanisms between *E. coli* and the actinomycetes responsible for the majority of polyketide compounds. Despite these possible disadvantages, *E. coli* has emerged as another option for heterologous biosynthesis, because (1) its genetic and metabolic traits are well-studied and characterized; (2) genetic and metabolic engineering tools are readily available; (3) *E. coli*'s native metabolic pathways have little crosstalk with the heterologous pathway, making it a clean host (i.e., there would be few native pathways to either contribute to or detract from targeted heterologous products); (4) it is easy to culture and exhibits relatively rapid growth kinetics; and (5) process engineering tools and scale-up techniques are well-established for *E. coli*. However, it should also be noted that in addition to common issues associated with inter-species genetic transfer such as codon bias, foreign enzyme folding, etc., *E. coli* lacks the machinery to support complex natural product biosynthesis. For example, *E. coli* does not possess enzymes such as the

phosphopantetheinyl transferase for polyketides synthase post-translational modification. Nor does *E. coli* provide certain substrates, such as (2S)-methylmalonyl-CoA, required for natural product biosynthesis. *E. coli* also lacks self-resistance mechanisms to protect itself from antibiotic production. Despite these problems, *E. coli* has been successfully utilized for production of the erythromycin precursor 6dEB, erythromycin C, and a variety of other polyketide products (*40-42*).

To achieve heterologous biosynthesis of erythromycin polyketide ring 6dEB, *E. coli* strain BAP1 was created to (1) disrupt the propionate catabolism operon, except the *prpE* gene which encodes a propionyl-CoA synthetase; (2) integrate into the chromosome a phosphopantetheinyl transferase gene *sfp* derived from *B. subtilis*. In addition, a propionyl-CoA carboxylase (PCC, isolated from *S. coelicolor*) was introduced to provide a metabolic pathway capable of converting exogenously fed propionate to propionyl-CoA and (2S)-methylmalonyl-CoA, the two substrates needed by the three deoxyerythronolide B synthase (DEBS) biosynthetic enzymes responsible for 6dEB biosynthesis. DEBS genes (DEBS1, 2, and 3) were also introduced by use of two plasmids named pBP130, pBP144.

Efforts have also been made towards the biosynthesis of erythromycin production in *E coli*. In a previous report of erythromycin C production, tailoring genes responsible for erythromycin sugar moiety biosynthesis, attachment and modification were transferred from *Micromonospora megalomicea* (not *S. erythraea*) to *E. coli*. Synthetic operons for erythromycin biosynthesis were designed, constructed and introduced into the *E. coli* host. After careful coordination of expression of the genes involved in the biosynthetic process, production of erythromycin C was achieved. Notably, 2 genes were

found to have low expression levels (not detectable by SDS-PAGE analysis), which may be a result of codon bias. Despite all the efforts dedicated to erythromycin biosynthesis in *E. coli*, the reported titer was only 0.4 mg/L, much lower than that of native producer *S. erythraea*. More importantly, heterologous biosynthesis of erythromycin A, the most potent form of erythromycin, has not been achieved.

Efforts have also been made to produce erythromycin analogs in *E. coli*. It has been found that compound 6-deoxyerythromycin D (erythromycin D without the C-6 hydroxyl group) could be produced by reconstitution of a hybrid pathway with tailoring biosynthetic genes from *S. fradiae*, *S. erythraea*, and *S. venezuelae* (*43*). Similar to the previous reports, plasmid vectors were utilized to achieve the heterologous biosynthesis of the proposed product. Furthermore, a bioassay-guided screening method was developed to identify an antibiotic over-producing mutation in the mycarose biosynthesis pathway, which resulted in 6-deoxyerythromycin D at around 5-10 mg/L. However, such a erythromycin analog is in fact a shunt metabolite of erythromycin biosynthesis. In other words, genetic manipulation was made only to change the sequence or the presence of certain biosynthetic steps, and no new or altered genetic materials were employed to fundamentally change the molecular structure of erythromycin. Therefore, more sophisticated studies are needed to produce better erythromycin analogs in the hope of finding next-generation unnatural products with novel bioactivity.

1.2 Research Goals and Challenges

This thesis research aims to establish *E. coli* as a viable option for heterologous biosynthesis of erythromycin A and other complex natural products. The production of

the erythromycin precursor 6dEB in *E. coli* needs successful transferring of three heterologous gene components, including genes for mega polyketide synthases, a gene for biosynthetic enzyme posttranslational modification, and genes for required CoA substrate biosynthesis. Moreover, the host's native metabolism and heterologous metabolic pathways need to be balanced to better produce the polyketide product. The effort for heterologous erythromycin A production requires the transfer of a 55 kb gene cluster and the coordinated expression of 26 genes, 23 of which are foreign. In order to achieve our goal, studies were conducted to address the challenges of cloning large gene clusters, coordinated gene expression, maintaining in vivo enzyme activity, and modification of *E. coli* native metabolism to support heterologous pathways. Metabolic engineering approaches were adapted to fully address these issues and achieve the direct production of the desired compound.

In addition, the erythromycin A modular polyketide and sugar group biosynthetic pathways were modified to produce rationally designed erythromycin analogs with novel chemical structures. This was accomplished using *E. coli* as a clean heterologous host, which could be hard to achieve in native natural product producers due to the unavailability of genetic and metabolic tools and undesired interference from native metabolism. Unlike previous reports (*43*), new biosynthetic enzymes from other microorganisms were introduced to reconstitute totally new erythromycin analog biosynthesis pathways in *E. coli*. Heterologous biosynthesis thus provides an even greater opportunity for molecular diversification as the modular polyketide and linear sugar biosynthetic scheme offers seemingly limitless options to reprogram biosynthesis for analog formation (*44-49*).

This thesis study is significant from several perspectives. First, erythromycin A is an important natural product with complex polyketide and sugar biosynthetic pathway. Its full heterologous production in *E. coli* has not been achieved yet. Second, successful heterologous biosynthesis requires addressing many important challenges in genetic, metabolic and process engineering, which will pave the way for future complex natural production biosynthesis in *E. coli*. Third, production of designed erythromycin analogs through directed manipulation of polyketide and sugar biosynthesis validates the use of *E. coli* for unnatural compound biosynthesis. Last but not least, the biosynthesis of novel erythromycin analogs offers the promise of generating a variety of new bioactive compounds. Supporting this view were the successful semi-synthetic erythromycin analogs that extended the molecule's utility in the face of acquired pathogenic drug resistance (*50, 51*). This potential then stands as an answer to the need for new antibiotics to combat the growing presence of pathogenic drug resistance (*52, 53*).

Chapter 2 Production Optimization of Erythromycin Precursor 6-Deoxyerythronolide B

The production of erythromycin's polyketide precursor 6-deoxyerythronolide B (6dEB) is of great research interest, as it is the first step towards complete erythromycin biosynthesis. In addition, the heterologous biosynthesis of polyketide products also offers potential access to structurally complex and medicinally relevant molecules. Natural product heterologous production requires successful introduction and reconstitution of foreign genes encoding a new metabolic pathway. For example, Khosla and coworkers developed a modified *S. coelicolor* strain CH999 which was used to accommodate the cloned polyketide 6dEB synthases and achieved production of 6dEB (*39*). Using similar technical approaches, *S. lividans* was also engineered to produce 6dEB. In particular, a system was implemented where three individual plasmids were used to introduce the three DEBS genes to *S. lividans* K4-114 (*38*). The 6dEB production in *S. lividans* was further improved through use of high- and low-copy number plasmids, compatible and incompatible plasmid origins of replication, and localizing multiple DEBS genes to the host chromosome.

As introduced in the previous chapter, the heterologous biosynthesis of the erythromycin polyketide ring, 6dEB, has also been achieved in *E. coli* (*40*). However, the titer of the target product remained low. To address this issue, in the following sections in the chapter, the production of polyketide 6dEB was optimized by (1) studying the relationship between three required foreign genes' dosage and 6dEB production levels and (2) manipulation of *E. coli* native metabolic pathways to better support 6dEB biosynthesis. The approaches used in these studies will help identify the rate-limiting

steps of both heterologous and native pathways for 6dEB biosynthesis and provide scientific insight towards the heterologous production process.

2.1 Optimization of *sfp, pcc,* and DEBS Genes' Copy Number for 6dEB Production

2.1.1 Introduction

As discussed previously, there are three foreign proteins needed for heterologous 6dEB biosynthesis in *E. coli*. They include Sfp (required for biosynthetic enzyme posttranslational modification via phosphopantetheinylation), PCC (propionyl-CoA carboxylase, required for biosynthetic substrate support), and DEBS (deoxyerythronolide B synthase, required for 6dEB biosynthesis) (Figure 1.1). Although the cellular design described for *E. coli*-derived 6dEB resulted in positive production from the new host, a simple mass balance confirmed that final 6dEB conversion fell well below 100%, typically only reaching 5-15% conversion (*54, 55*). Thus, efforts have begun to analyze the cellular level biosynthetic process with goals that include identifying bottlenecks to production, improving current conversion levels, and, more generally, offering a better understanding of the heterologous biosynthetic process. In this work, the heterologous components of *sfp*, *pcc*, and DEBS genes needed for *E. coli*-derived 6dEB biosynthesis were studied in the context of gene expression variation.

Chromosomal integration of foreign genes has become a very popular technique for genetic manipulation (*56-58*). The insertion of certain genes into the chromosome provides a stable localization method without the need for exogenous selective pressure. This is especially important when it comes to the compatibility and stability problems the use of plasmids often presents. In addition, chromosomal localization could also be used

to disrupt native genes that adversely affect desired metabolism. However, chromosomal integration in most cases reduces the copy number of the gene to be integrated, which correspondingly produces less gene product. It has been reported that the three DEBS genes were integrated to the chromosome of BAP1, although the resulting 6dEB titer was < 1 mg/L (*59*). In this section, we varied the gene copy number by either chromosomal integration or using multiple copy number plasmids. All genes studied were under the control of T7 promoters which maintained a relatively high gene transcription level.

2.1.2 Materials and Methods

(1) Reagents and Chemicals

The reagents and chemicals used in this study were purchased from Fisher Scientific and Sigma (USA). The restriction enzymes were from New England Biolab (USA). PCR primers were synthesized by Operon (USA). Phusion High-Fidelity PCR Master Mix for PCR was purchased from New England Biolab (USA).

(2) Plasmid and Strain Construction

E. coli strains and plasmids used in this paper are summarized in Table 2.1. Restriction enzyme digestions, transformations, SDS-PAGE, and other standard molecular biology techniques were carried out as described by Sambrook et al (*60*). Plasmid pBP130, pBP144, pYW1201 and pYW7317 were constructed in the previous work (*40, 59*). *B. subtilis sfp* gene was isolated as NdeI/XhoI fragment and inserted into pACYDuet-1 (purchased from novagen, USA) resulting the plasmid pACYC-*sfp*. *E. coli* BAP1, YW1 and YW9 were constructed previously by our group (*40, 59*).

Table 2.1 Strains and plasmids used for *sfp*, *pcc* and DEBS genes' copy number optimization

E. coli strain	description
BAP1	F-ompT hsdSB (rB-mB-) gal dcm (DE3) *prpRBCD*::T7prom-*sfp*, T7prom-*prpE*
BAP2	BAP1 Δ*sbm*
YW1	BAP1 *prpR*<>T7prom-*accA1-pccB*
YW9	BAP1 *ygfG*::T7prom-*accA1-pccB*-Kan^r-T7term, *araA*::T7 prom-DEBS1-T7term, *araB*::T7 prom-DEBS2-T7term, *araD*::T7 prom-DEBS3-T7term
Plasmid	
pBP130	pET21C derivative, T7prom-DEBS2-ribosome binding site-DEBS3-T7term
pBP144	pET28 derivative, T7prom-*pccB*-rbs-*accA1*-T7prom-DEBS1-T7term
pYW7317	pET28 derivative, T7prom-DEBS1-T7term
pACYC-sfp	pACYCDuet derivative, T7 prom-*sfp*-T7 term
pYW1200	pET21C derivative, T7prom- *pccB*-rbs-*accA1* -T7term

(3) Culture Condition

Glycerol stocks (15%) were used to start overnight cultures (2 mL) of strain-plasmid combinations at 30°C and 250 rpm in Luria-Bertani medium with antibiotic selection as appropriate (carbenicillin [100 mg/L], kanamycin [50 mg/L], chloramphenicol [34 mg/L]). The starter cultures were then used to inoculated 3 mL Luria–Bertani production cultures (at 2%, v/v) containing antibiotics (at the same concentrations referenced above), 100 μM isopropyl β-D-1-thiogalactopyranoside (IPTG), and 20 mM sodium propionate. The production cultures were incubated at 22^{o}C and 250 rpm for 3 days, and final culture optical density (OD) was measured at 600 nm.

Triplicate samples were used for each batch and 3 batches were conducted to ensure statistical analysis and reproducibility.

(4) SDS-PAGE Analysis

Culture conditions were identical to those described for 6dEB product formation. After the centrifugation, cells were washed and re-suspended in TE buffer. Cell densities were equalized between samples and sonication was performed with a Fisher Scientific Sonic Dismembrator Model 100 at maximum setting for a tip probe for three 10 s intervals. Lysates were then centrifuged and the whole cell lysate and soluble fractions were analyzed by SDS-PAGE. Only the soluble fraction was presented.

(5) 6dEB Production Analysis

The cultures of the *E. coli* strain BAP1(pBP130/pBP144), BAP1(pBP130/pBP144/pACYC-sfp), BAP1(pBP130/pBP144/pACYCDuet-1) were centrifuged at 10000 rpm for 5 min to spin down the cell pellet. The supernatant was analyzed for 6dEB titer using Agilent (Palo Alto, CA, USA) 1100 series HPLC coupled with an Alltech (Deerfield, IL, USA) 800 series evaporative light-scattering detector (ELSD). The high performance liquid chromatography (HPLC) method for 6-dEB separation and quantification has been described previously (*61*). The guard column used was an Inertsil ODS3 C18 5 μm, 4.6×10 mm while the analytical column used was an Inertsil ODS3 C18 5 μm, 4.6×150 mm, both purchased from GL Sciences (Sinjuku-Ku, Tokyo, Japan).

Since the 6dEB titer in the culture of YW1(pBP130/pYW7317) and YW9(pYW1200) was below the lower limit of HPLC detection, Mass Spectrometry was used to measure 6dEB concentration. Specifically, erythromycin (purchased from

Belgium) was added into samples as internal standard for 6dEB quantification. 1.5 mL of ethyl acetate was added into 3 mL *E. coli* culture and mixed well by vortex. After centrifugation at 2500 rpm for 5 min, 0.75 mL of ethyl acetate on the upper layer was transferred into eppendorf tube and dried overnight. 50 microliter of methanol containing 5 mg/L erythromycin was added into the eppendorf tube to dissolve the 6dEB. The 6dEB standard samples were prepared in the same way. After 3 days' culture of BAP1 without any plasmid, a series of 6dEB standards were added into the culture. The ethyl acetate extraction and erythromycin-containing methanol resuspension were then performed.

The prepared samples were injected into LTQ XL Linear Ion Trap Mass Spectrometer (Thermo Electron Corporation, Waltham, MA) by a 250 microliter syringe at a speed of 10 microliter/min. The 6dEB standard samples were used to developing the calibration curve. The ratio of peak intensity of 6dEB over erythromycin internal standards was correlated to the 6dEB concentration by using a calibration curve that was made every time before the real culture samples. The MS quantification was performed 3 times and each time 5 samples were measured.

2.1.3. Results

A series of *E. coli* strains were constructed to study the impact of copy number of *sfp*, *pcc*, and DEBS genes on heterologous biosynthesis of 6-deoxyerythronolide B (6dEB) in *E. coli*. The previously constructed *E. coli* strain BAP1 has a *Bacillus* subtilis *sfp* gene which encodes a phosphopantetheinyl transferase required to activate the DEBS proteins. Plasmids pBP130 and pBP144 contained high copy number of DEBS genes and *pcc* genes (containing *accA1* and *pccB* subunits). To investigate the effect of *sfp* gene dosage,

a high copy plasmid, pACYC-*sfp*, was constructed and introduced into *E. coli* BAP1. The 6dEB production level of BAP1(pBP130/pBP144/pACYCDuet-1) and BAP1(pBP130/pBP144/pACYC-*sfp*) were then compared to investigate how much the increased *sfp* copy number would help improve 6dEB production in *E. coli*. Empty plasmid pACYC was included here to facilitate the comparison with the strain harboring pACYC-*sfp*. Similarly, *E. coli* YW1 and YW9 were constructed for production comparison (Table 2.1). YW1(pBP130/pYW7317) and BAP1(pBP130/pBP144) have the single and multiple copy number of *pcc* genes, respectively. YW9(pYW1200) and BAP1(pBP130/pBP144) have the single and multiple copy number of DEBS genes, respectively. The comparison of the final titer levels (mg/L of 6dEB) in these strains would be correlated to gene dosage and expression of the necessary heterologous enzymes needed for 6dEB biosynthesis.

After the construction of the strains harboring varying copies of heterologous genes, SDS-PAGE analysis was first performed to determine the expression levels of the *sfp*, *pcc*, and DEBS genes. As shown in Figure 2.1.1, protein levels correlated with gene dosage design. A semi-quantitative densitometry analysis of protein level differences was assessed using ImageJ software version 1.40g (http://rsb.info.nih.gov/ij/). The analysis provided the ratio of densitometry measured band densities between respective Sfp, PCC, and DEBS strain pairs (as defined in Table 2.2). Specifically, *sfp* expression level in BAP1 (pBP130/pBP144/pACY-*sfp*) was higher than when *sfp* gene was integrated into the chromosome. The integration of *pcc* genes into the *E. coli* chromosome greatly reduced the amount of protein PCC. Similarly, the DEBS genes are observed when expressed from the pET expression plasmids but are not observed when expressed from

low copy number within the chromosome. It is interesting to note that DEBS genes expression was actually better when *pcc* genes had a low copy number (YW1 (pBP130/pYW7317) vs. BAP1 (pBP130/pBP144)). This could be because less metabolic burden for *pcc* over-expression led to a better supply of substrate for heterologous biosynthesis of DEBS. Densitometry analysis was further used to assess and confirm expression differences within the SDS-PAGE data.

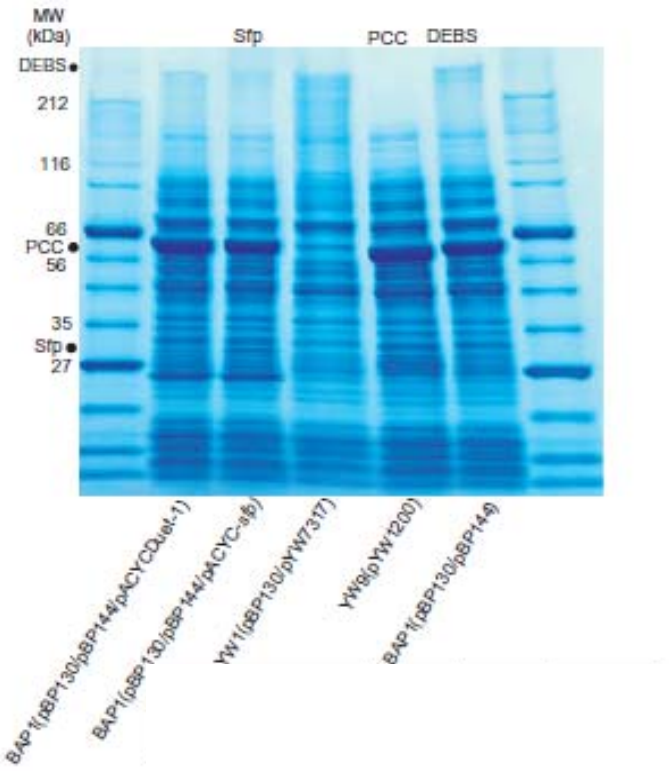


Figure 2.1.1 SDS-PAGE analysis of *sfp*, *pcc* and DEBS gene expression with varying copy numbers. The size and location of the Sfp, DEBS and PCC enzymes have been denoted by black-filled dots within the molecular weight marker notations.

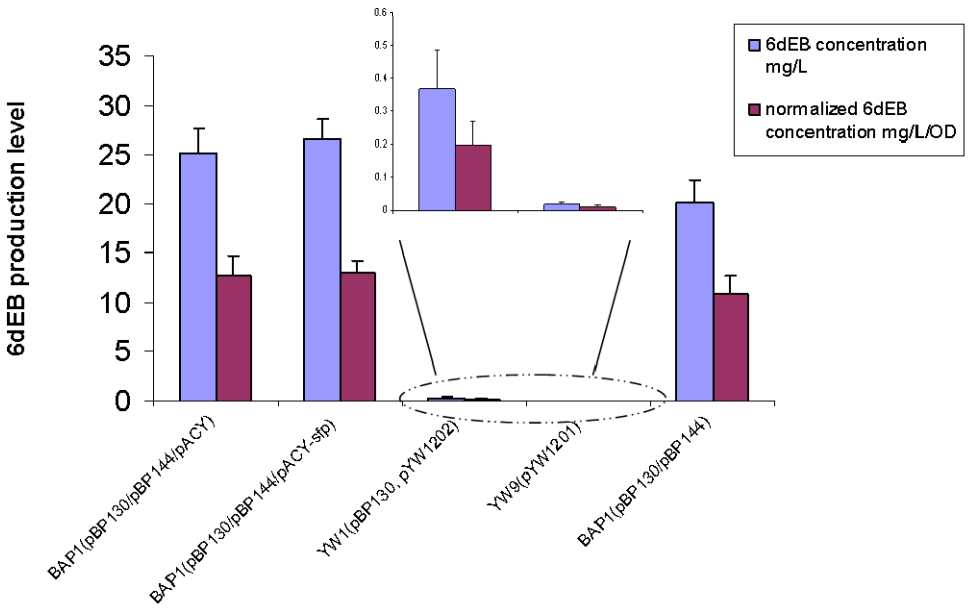


Figure 2.1.2 6dEB production level in *E. coli* strains with different heterologous gene copy numbers.

Each strain was then tested for 6dEB biosynthesis with results presented in Figure 2.1.2. The 6dEB production level was evaluated as the final concentration in the culture after 3 days' growth and it was normalized by the culture's OD values as well. It should be noted that the replicons of pET (pBP130 and 144 plasmid backbone) and pACYDuet-1 vectors used in this study were compatible, which stabilized the 3-plasmid system in *E. coli*. When the *sfp* gene copy number increased, the 6dEB production level variation was not significant (the first and the second column). The over-expression of *pcc* genes showed a tremendous increase of 6dEB concentration compared to the case with low *pcc* genes dosage (the third and the last column). Production change associated with DEBS protein level variation was also statistically significant (the forth and the last column).

Table 2.2 6dEB production Comparison between Sfp, PCC and DEBS pairs

Component	Sfp	PCC	DEBS
6dEB titer Ratio (High to low gene copy)	1.1	164.1	3308.2
Protein level ration (High to low gene copy)	1.6	6.2	6.59
Sensitivity Parameter: 6dEB Ratio divided by Protein Level Ratio	0.6	26.7	501.9

Statistical analysis was then conducted for each comparison pair presented in Table 2.2, which was a final summary collected from the densitometry and 6dEB titer data of Figures 2.1.1 and 2.1.2. A sensitivity parameter has been defined as the ratio of high and low gene copy 6dEB titers (Figure 2.1.2) divided by the ratio of high and low gene copy protein levels (Figure 2.1.1) for each heterologous component varied. The variation of the sensitive parameter thus reflected to what extent 6dEB heterologous production could be affected by the three gene/protein components.

Combined data from the SDS-PAGE and 6dEB production analyses (Table 2.2) suggested that 6dEB biosynthesis was most significantly affected by DEBS gene and protein levels. It was also indicated that 6dEB production would improve by further enhancement of PCC protein levels. It was found that 6dEB titers showed a smaller response to changes in Sfp protein levels. In summary, the effect on *E. coli* 6dEB biosynthesis caused by heterologous component copy number variation was ranked as DEBS>PCC>Sfp.

2.1.4. Discussion

The section's study focused on the effect of heterologous gene components' copy number and thus the corresponding protein levels on 6dEB production. This was achieved by cloning the genes into either the *E. coli* chromosome (single copy) or plasmids (multiple copies). Interestingly, when the *sfp* gene had high copy number, 6dEB production was not improved. Since DEBS is a large protein complex, heterologous expression in *E. coli* has been found to be at a low level. This can be indicated by the SDS-PAGE (Figure 1). Therefore, chromosomal *sfp* expression (BAP1) may already be sufficient for the posttranslational modification of DEBS. The further increase of *sfp* gene copy number did not improve 6-dEB heterologous production to any extent. In fact, it has been found that the BAP1 strain used in this study could achieve the phosphopantetheinylation of polyketide synthases at a high level (>99%) with only one copy of *sfp* gene in the chromosome (*62*).

The integration of the *pcc* genes into *E. coli* BAP1 chromosome had been studied before (*63, 64*). In this section, we constructed another *E. coli* strain with the *prpR* gene replaced by the *pcc* genes. As shown by the production level analysis, the use of a chromosomal copy of the *pcc* genes greatly reduced the 6dEB concentration. This could be attributed to the decrease of the genes' copy number. This observation is in contrast to results seen previously where the *pcc* genes with two subunits under the control of separate T7 promoters were integrated into the *ygfG* region of the *E. coli* chromosome (*64*). The strain reported in that case showed improved 6dEB levels; whereas, our strain showed a noticeable reduction in 6dEB. The chromosomal location and the use of two T7 promoters could be the reason why 6dEB production was different.

The use of *E. coli* YW9 with DEBS and *pcc* genes integration to produce 6dEB has also been reported previously (*59*). A plasmid carrying extra copies of the *pcc* genes was introduced into YW9 to maintain the high level of PCC proteins. This strategy resulted in a low 6dEB titer, indicating that chromosomal copies of the DEBS genes only produced low levels of polyketide synthases when compared with the plasmid-borne DEBS genes.

Based on the results above, it was concluded that DEBS and *pcc* genes' expression at a high level was essential for 6dEB production. Furthermore, when these two genes both had high copy number, the further increase of *sfp* gene dosage did not lead to a significantly higher production as shown in Figure 2.1.2. Between DEBS and PCC, the former was considered to be more limiting for 6dEB heterologous production, as highlighted by the production sensitivity parameter shown in Figure 2.1.2. In summary, it was concluded that the copy number of *sfp*, *pcc*, and DEBS genes had varying level of influence on 6dEB heterologous production in *E. coli*; and among them, the DEBS genes were found to be the most important components, followed by the *pcc* genes and the *sfp* gene.

2.2 6dEB Production Optimization by Engineering *E. coli* Native Metabolic Pathways

2.2.1 Introduction

Polyketide heterologous biosynthesis requires efficient introduction and expression of required genes responsible for establishing a new biosynthetic metabolic pathway. However, once accomplished, improving heterologous titers remains a

challenge since it is often unknown how native and heterologous metabolism will interact. Many strategies have been adopted to improve natural product titers, including codon optimization (*65*), auto-inducer utilization (*61*), promoter enhancement (*66*), and gene copy number control (*67*). However, these efforts primarily focus upon the reconstitution of a heterologous pathway which may be compromised by a metabolic bottleneck resulting from native metabolism limiting the availability of required substrates or cofactors. It is therefore essential to characterize native host metabolism in order to identify and remove undesired pathways and/or enhance those pathways favorably supporting heterologous biosynthesis.

A similar situation exists for heterologous biosynthesis of the erythromycin polyketide precursor 6dEB in *E. coli*, in which questions regarding how the heterologous biosynthetic pathways interact with native metabolism remain unclear. One good example of this unknown interaction is that propionate to 6dEB conversion was typically only between 5-15%, implying the activity of native metabolism detracting from desired 6dEB biosynthesis. The objective of the experiments described in this section is to further improve 6dEB production in *E. coli* by modification of the host's native pathway(s) to support heterologous biosynthesis.

To achieve this objective, three enzymes (Sbm (a methylmalonyl-CoA mutase), YgfG (a methylmalonyl-CoA decarboxylase), and YgfH (a propionyl-CoA:succinate-CoA transferase)) were identified as being capable of connecting native *E. coli* metabolism to heterologous 6dEB production, through the propionyl-CoA and (2S)-methylmalonyl-CoA substrates (Figure 2.2.1) (*68*) (*69*).

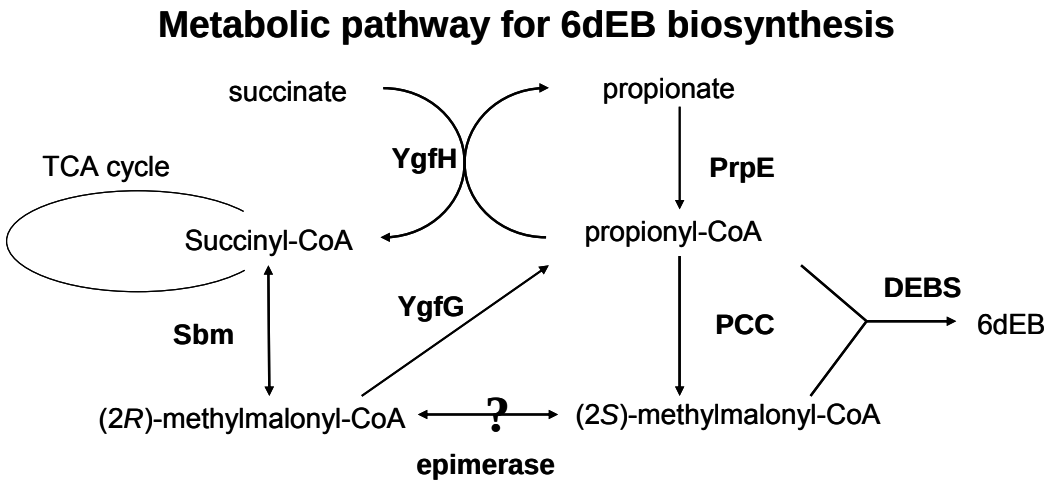


Figure 2.2.1 Metabolic pathways connecting native metabolism to 6dEB heterologous production in *E. coli*. The map has been constructed from previous studies *(40, 68, 70, 71).*

Dayem et al. has reported that the *E. coli* methylmalonyl-CoA mutase encoded by the *sbm* gene specifically catalyzes the conversion of succinyl-CoA to (2R)-methylmalonyl-CoA. It has been found that co-expression of *P. shermanii* methylmalonyl-CoA mutase and epimerase in *E. coli* harboring the DEBS genes produced 6dEB at a level of around 1 mg/L (*70*), but to our knowledge, the *E. coli* epimerase which catalyzes the interconversion of (2R) and (2S)-methylmalonyl-CoA has not been identified. *E. coli* methylmalonyl-CoA decarboxylase encoded by the *ygfG* gene removes the carboxyl group of methylmalonyl-CoA to make propionyl-CoA. This conversion is reverse to that of the *S. coelicolor* PCC pathway introduced into *E. coli*, which produces (2S)-methylmalonyl-CoA from propionyl-CoA. Notably, only (2R)-methylmalonyl-CoA was found to be utilized by the *ygfG* gene product. Whether (2S)-methylmalonyl-CoA can be decarboxylated by *E. coli* YgfG is still unknown. *E. coli* YgfH is responsible for the CoA moiety transfer between propionyl-CoA and succinyl-CoA. Since this is a reversible reaction, *E. coli* native and heterologous metabolism

competes for propionyl-CoA to produce succinyl-CoA or (2S)-methylmalonyl-CoA. In this section, the expression level of Sbm, YgfG and YgfH was varied by either deleting or over-expressing the corresponding genes in an attempt to investigate their impact on 6dEB biosynthesis. The lambda-Red recombination technique was used for the gene deletion mutant strain construction. And plasmid pACYCDuet, containing a compatible replicon with the pET plasmid containing DEBS and *pcc* genes, was utilized for gene over-expression.

Heterologous *pcc* genes encoding a propionyl-CoA carboxylase and exogenous propionate were believed to be required for successful polyketides 6dEB production. To further explore the potential of *E. coli* native metabolism to support complex natural product biosynthesis, the production of 6dEB was analyzed in the absence of either heterologous PCC or exogenous propionate. The motivation of this study is to investigate whether *E. coli* native metabolism can produce the required (2S)-methylmalonyl-CoA and propionyl-CoA for 6dEB heterologous biosynthesis. The constructed *sbm*, *ygfG* and *ygfH* mutants were also utilized to characterize the native metabolism for such biosynthesis. In doing so, the study in this section also makes efforts to more thoroughly explore, characterize, and engineer native *E. coli* metabolism to support the production of the complex polyketide 6dEB.

2.2.2 Materials and Methods

(1) Reagents and Chemicals

The reagents and chemicals used in this study were purchased from Fisher Scientific (Pittsburgh, PA, USA) or Sigma (St. Louis, MO, USA). All restriction enzymes

and the Phusion High-Fidelity PCR Master Mix were purchased from New England Biolabs (Ipswich, MA, USA). PCR primers were synthesized by Operon (Huntsville, AL, USA).

(2) Strains and Plasmid Construction

Restriction enzyme digestions, SDS-PAGE, and other standard molecular biology techniques were performed as described by Sambrook et al. (*60*). *E. coli* BAP1 (a derivative of BL21(DE3)，(*72*)) was constructed previously (*40*). BAP1 *sbm* and *ygfG* deletions were conducted by directly disrupting the target genes by kanamycin marker insertion using λ-Red recombination followed by excision of the kanamycin resistance cassette through recombination of flanking flippase recognition targets mediated by the flippase recombination enzyme harbored by pCP20 (*58*). Inserting the kanamycin cassette to the *ygfH* location was accomplished through P1 transduction from *E. coli* ECK2916 to BAP1 (*73*). The kanamycin cassette was removed in the same way as described for the *sbm* and *ygfG* deletions.

For gene over-expression, *sbm* was PCR amplified using primers 5'-GCCATATGATGTCTAACGTGCAGGAGT-3' and 5'-GCGAATTCTTAATCATGATGCTGGCTT-3'. The *Nde*I/*Eco*RI digested fragment was then inserted into pACYCDuet-1 digested with the same restriction enzymes. The *ygfG* gene was PCR amplified using primers 5'-GCCTGCAGATGTCTTATCAGTATGT-3' and 5'-GCCTGCAGTTAATGACCAACGAAATT-3' and the PCR product was digested by *Pst*I and inserted into the *Pst*I site of pACYCDuet-1. The *ygfH* gene was PCR amplified using primers 5'-GCCCATATGGAAACTCAGTGGACAAG-3' and 5'-GCCCTCGAGTTAACCCAGCATCGAGCCG-3'. It has been reported that the 6×-

hisitine-tagged YgfH lacked *in vitro* catalytic activity (*68*). Therefore, the *Nde*I/*Xho*I digested *ygfH* gene fragment was inserted into the second MCS of pACYCDuet-1 which does not produce a 6×-histidine-tag product. The *sbm*, *ygfG*, and *ygfH* inserations into pACYCDuet-1 led to plasmids pHZ1, pHZ2, and pHZ3, respectively. These constructs were verified by Sanger sequencing at the Tufts University Core Facility. Table 2.3 summarizes the *E. coli* strains and plasmids utilized in this study.

Table 2.3 Plasmid and *E. coli* strains used for native metabolism manipulation

	Description
BAP1	F- ompT hsdSB (rB-mB-) gal dcm (DE3) prpRBCD::T7prom-*sfp*,T7prom-*prpE*
TB1	BAP1 Δ*sbm*
TB2	BAP1 Δ*ygfG*
TB3	BAP1 Δ*ygfH*
pYW7317	pET28 (Kan^r) derivative, T7prom-DEBS1-T7term
pBP130	pET21c ($Carb^r$) derivative, T7prom-DEBS2-DEBS3-T7term
pBP144	pET28 (Kan^r) derivative, T7prom-*pccB*-rbs-*accA1*-T7prom-DEBS1-T7term
pHZ1	pACYCDuet-1 (Cm^r) derivative, T7prom-*sbm*-T7term
pHZ2	pACYCDuet-1 (Cm^r) derivative, T7prom-*ygfG*-T7term
pHZ3	pACYCDuet-1 (Cm^r) derivative, T7prom-*ygfH*-T7term

(3) 6dEB Production and Analysis

HPLC coupled to an evaporative light scattering detector (ELSD) was utilized to measure the 6dEB concentration in cultures of *E. coli* strains harboring the *pcc* genes

(carried by pBP144), since these strains consistently produced easily detectable (>5 mg/L) concentrations of 6dEB. *E. coli* BAP1 harboring pBP130, pBP144, and pHZ1, pHZ2, or pHZ3 were grown in 15 mL production medium containing 100 mg/L carbenicillin, 50 mg/L kanamycin, 34 mg/L chloramphenicol, 100 μM IPTG, and 20 mM propionate at 22°C and 250 rpm for five days. One liter of production medium contained 5 g yeast extract, 10 g tryptone, 15 g glycerol, 10 g sodium chloride, 3 mL 50% v/v Antifoam B, 100 mM 4-(2-Hydroxyethyl)-1-piperazineethanesuffonic acid (HEPES) buffer, and was adjusted to pH 7.6 by NaOH before use. Since hydroxocobalamin is required for Sbm activity (*68, 74*), BAP1 (pBP130, pBP144, pHZ1) was cultured with 10 μM of hydroxocobalamin (*70*). After five days, 1 mL aliquots of the cultures were centrifuged for 10 min at 10,000 rpm and the supernatants were analyzed by an Agilent (Palo Alto, CA, USA) 1100 series HPLC coupled with an Alltech (Deerfield, IL, USA) 800 series ELSD. The cell pellet collected after centrifugation was used for gene expression SDS-PAGE analysis. Cell density was assessed after the five day culturing period by OD_{600} measurements.

For bioreactor experiments, 50 mL overnight flask cultures of either BAP1 (pBP130/pBP144) or TB3 (pBP130/pBP144) were used to seed a 3 L bioreactor (New Brunswick Scientific, BioFlo110, Edison, NJ) containing 1 L bioreactor-production medium (with the same composition as the previously introduced production medium except that glycerol concentration is 45 g/L). Culture conditions were maintained as follows: air flow rate, 3 l/min; pH value, 7.6; temperature, 22°C; and agitation, 500 rpm. Samples were withdrawn every 12 hr for OD_{600} measurements and 6dEB concentration analysis by HPLC-ELSD.

Due to the detection limitations of the HPLC-ELSD system, mass spectrometry was used for the measurement of 6dEB production levels in the absence of the *pcc* genes or propionate. BAP1, TB1, TB2 and TB3 containing pBP130 and pYW7317 (together carrying the DEBS1, 2 and 3 genes) were grown in 3 mL LB medium containing 100 mg/L carbenicillin, 50 mg/L kanamycin, 100 μM IPTG, and 20 mM propionate at 22°C for 3 days. After culturing, 6dEB was extracted with 1.5 mL of ethyl acetate and dried. The extract was dissolved in 50 μl of methanol for analysis. To prepare a suitable calibration curve, the culture procedure was repeated using BAP1 without harboring any plasmids and 6dEB standards added at different concentrations at the end of the culture period prior to ethyl acetate extraction. The standard samples were then used to prepare the calibration curve that correlated 6dEB MS peak intensity and 6dEB concentration. Erythromycin was used as an internal standard during the MS analysis to account for internal measurement drift of the instrument; erythromycin was added to the final methanol solution at a concentration of 5 mg/L. The calibration curve was generated for every batch run and experiments were repeated four separate times using a Thermo Electron Corporation (Waltham, MA, USA) LTQ XL Linear Ion Trap Mass Spectrometer. For each experiment, there were at least three replicate samples for each *E. coli* strain.

2.2.3 Results

(1) Deletion and Over-expression of the *sbm*, *ygfG*, and *ygfH* Genes

The *sbm*, *ygfG*, and *ygfH* genes are found located within the same operon. Individual deletion of the *sbm*, *ygfG,* and *ygfH* genes was conducted by λ-Red recombination or P1 transduction. The constructed strains are summarized in Table 2.3.

Over-expression of these genes together with the genes encoding for the PCC and DEBS enzymes was tested by SDS-PAGE analysis. As shown in Figure 2.2.2, Sbm, YgfG, and YgfH protein products were clearly visible after IPTG induction. PCC and DEBS were also expressed in *E. coli* as indicated by the arrows. Together, the gene deletion and over-expression experiments confirmed efforts to generate strains ranging in the levels of native enzymes potentially influencing the heterologous 6dEB pathway.

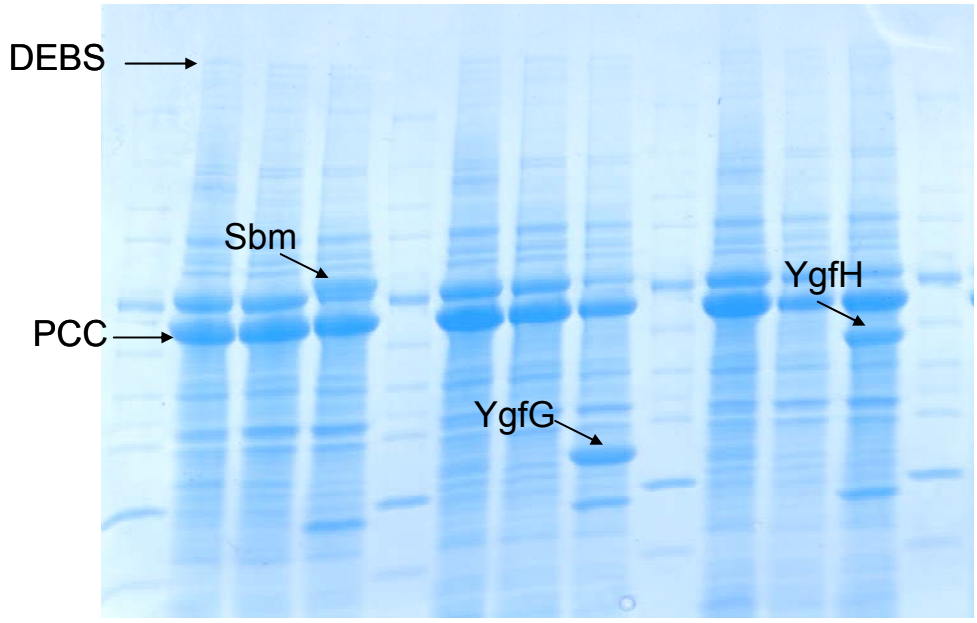


Figure 2.2.2 SDS-PAGE analysis of PCC, DEBS, *sbm*, *ygfG*, and *ygfH* over-expression. Lane 1: marker; lane 2: BAP1 (pBP130/pBP144); lane 3: TB1 (pBP130/pBP144); lane 4: BAP1 (pBP130/pBP144/pHZ1); lane 5: maker; lane 6: BAP1 (pBP130/pBP144); lane 7: TB2 (pBP130/pBP144); lane 8: BAP1 (pBP130/144/pHZ2); lane 9: marker; lane 10: BAP1 (pBP130/pBP144); lane 11: TB3 (pBP130/pBP144); lane 12: BAP1 (pBP130/144/pHZ3). Lane 13: marker

(2) 6dEB Production in the Presence of *pcc* Genes and Propionate

The production of 6dEB was analyzed by HPLC-ELSD when both the *pcc* genes and exogenous propionate were utilized for heterologous biosynthesis. The 6dEB production levels of different *E. coli* strains were compared in Figure 2.2.3. To take into

account the effect of strain-associated cell growth, specific 6dEB production levels are also indicated. The results indicated that Sbm had no impact on the heterologous 6dEB pathway as a result of either gene deletion or over-expression influencing 6dEB production. The deletion of *ygfG* also did not significantly impact 6dEB production levels as compared with the control BAP1 harboring pBP130/pBP144 (the forth column in Figure 2.2.3). However, the over-expression of *ygfG* led to a 4-fold decrease in 6dEB production. The deletion of *ygfH* increased 6dEB production levels by 2-fold (from 65 to 129 mg/L); whereas, *ygfH* over-expression did not have a significant affect on production. Specific 6dEB production showed the same trends as described above.

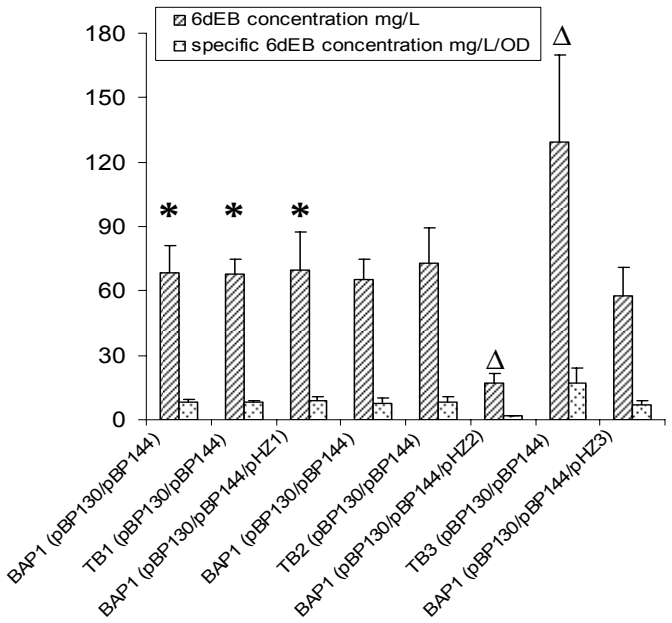
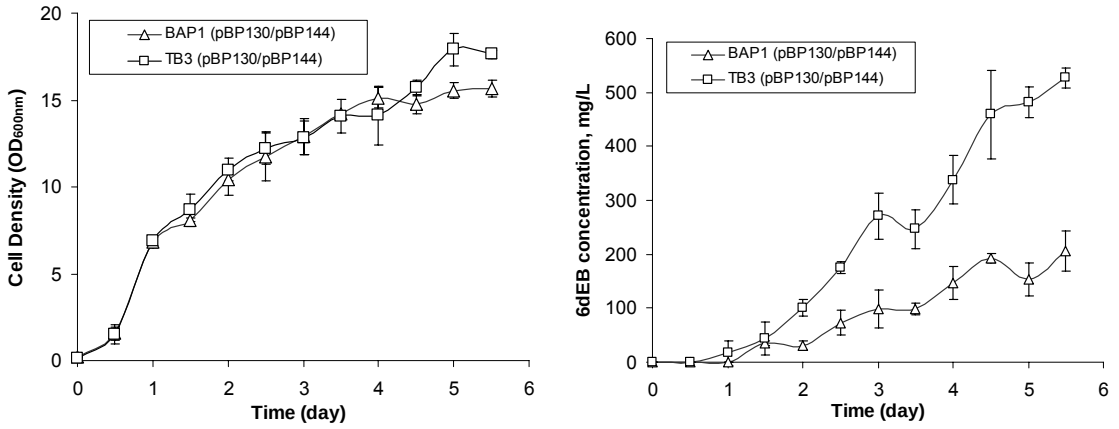


Figure 2.2.3 6dEB production comparisons in *E. coli* strains with *sbm*, *ygfG*, and *ygfH* genes deleted or over-expressed. Symbol * denotes that hydroxocobalamin was added. Symbol Δ denotes that 6dEB production levels are significantly different (95% confidence) from the control BAP1 (pBP130/pBP144).

To confirm the trends presented in the previous paragraph, production cultures were extended by testing production in a bioreactor system. BAP1 (pBP130/pBP144) and

TB3 (pBP130/pBP144) were compared, and as shown in Figure 2.2.4 (a), the growth curves of these two strains were similar. The production of 6dEB was observed after 1.5 day for BAP1 (pBP130/pBP144) and 1 day for TB3 (pBP130/pBP144). Final 6dEB titers reached approximately 206 mg/L in BAP1 (pBP130/pBP144) after 5.5 days; whereas, TB3 (pBP130/pBP144) produced approximately 527 mg/L 6dEB after the same time. In this scaled-up comparison, the 6dEB production level was increased by approximately 2.5 fold, slightly beyond the difference observed when using shake flask cultures.



A B

Figure 2.2.4 *E. coli* batch bioreactor comparison. (A) Growth curves of BAP1 (pBP130/pBP144) and TB3 (pBP130/pBP144). (B) 6dEB production levels of BAP1 (pBP130/pBP144) and TB3 (pBP130/pBP144).

(3) 6dEB Production in the Presence of Propionate but without *pcc* Genes

Although exogenous propionate and *S. coelicolor* propionyl-CoA carboxylase (PCC) were both required for high level 6dEB production, in the course of this study, it was also observed that *E. coli* metabolism can support heterologous 6dEB biosynthesis in the absence of exogenously fed propionate or expression of the *S. coelicolor* propionyl-CoA carboxylase.

Without the *pcc* genes, 6dEB production levels were greatly reduced (from 65 to 0.04 mg/L). The absence of the *pcc* genes eliminates the heterologously introduced conversion of propionyl-CoA to (2*S*)-methylmalonyl-CoA (Figure 2.2.1). However, 6dEB biosynthesis could still be observed, implying that native *E. coli* metabolism was capable of supplying (2*S*)-methylmalonyl-CoA. To confirm whether this provision of (2*S*)-methylmalonyl-CoA is relevant to the pathway encompassing Sbm, YgfG, and YgfH, the corresponding BAP1 deletion mutants were tested for 6dEB production in the absence of the *pcc* genes. As shown in Figure 2.2.5, the deletions of *sbm*, *ygfG,* and *ygfH* did not change the 6dEB production levels. The result suggested the availability of alternative native pathways (unrelated to Sbm, YgfG, and YgfH) capable of providing (2*S*)-methylmalonyl-CoA for 6dEB biosynthesis.

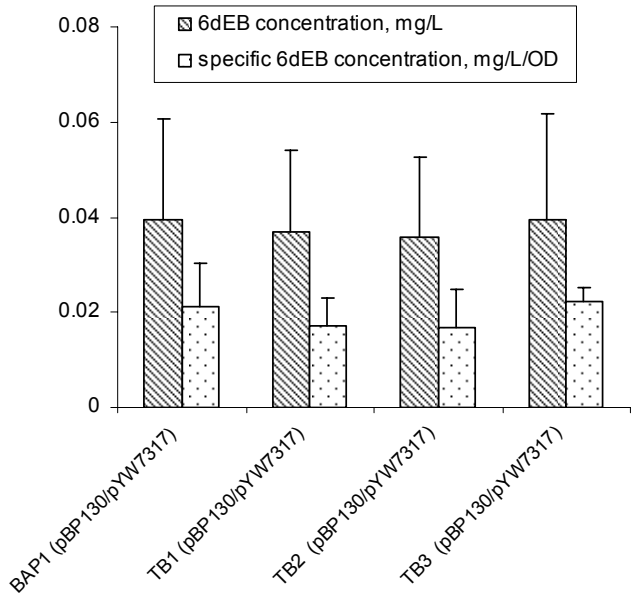


Figure 2.2.5 6dEB production levels in BAP1 and BAP1 mutant strains in the absence of the *S. coelicolor pcc* genes.

(4) 6dEB Production in the Presence of *pcc* Genes but without Propionate

Propionate is the substrate required for the engineered production of propionyl-CoA and (2*S*)-methylmalonyl-CoA. In the absence of exogenous propionate, 6dEB production was greatly reduced. The 6dEB production level of different *E. coli* strains was compared in Figure 2.2.6. When neither propionate nor *pcc* genes were present (using strain BAP1 harboring pBP130 and pYW7317), no 6dEB production was detected (<0.01 mg/L). In the presence of *pcc* genes but no propionate, *E. coli* BAP1 containing the DEBS genes produced 0.11 mg/L 6dEB. When this condition was tested with the strains containing the *sbm*, *ygfG*, and *ygfH* deletions, a statistically significant decrease in final 6dEB production was observed for each strain. These findings indicated that without exogenous propionate, *E. coli* native metabolism was able to synthesize propionyl-CoA which was later converted to (2*S*)-methylmalonyl-CoA (facilitated by the PCC enzyme) for 6dEB production. Furthermore, the observation of propionyl-CoA biosynthesis without exogenously fed propionate could be partly attributed to the Sbm, YgfG, and YgfH pathway since removal of the encoding genes decreased the heterologous 6dEB production levels.

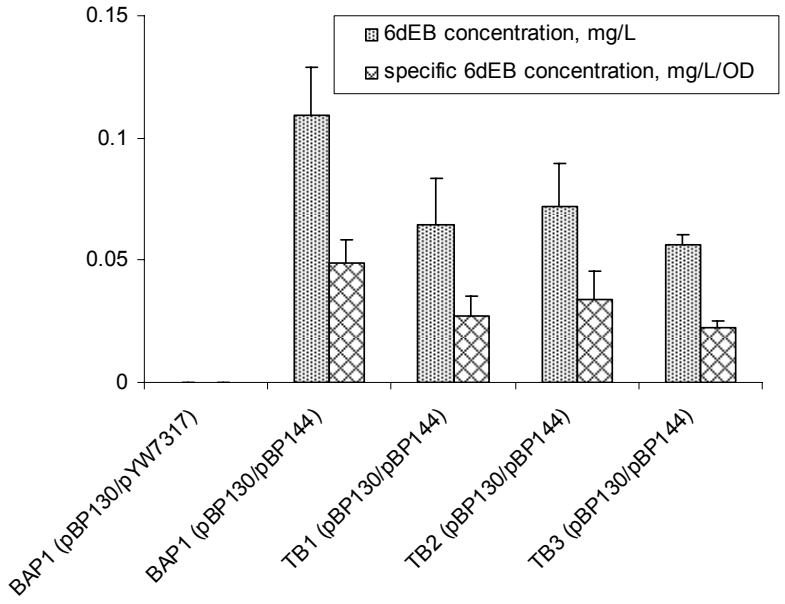


Figure 2.2.6 6dEB production levels in BAP1 and BAP1 mutant strains in the absence of propionate.

2.2.4 Discussion

(1) 6dEB Production in the Presence of *pcc* Genes and Propionate

Propionyl-CoA and (2*S*)-methylmalonyl-CoA are the two required substrates for heterologous production of 6dEB in *E. coli*. The 6dEB heterologous biosynthetic pathway and *E. coli* native metabolism are inter-connected by Sbm, YgfG, and YgfH. Though the design of the 6dEB pathway sought to maximize the flux of propionyl-CoA and (2*S*)-methylmalonyl-CoA to 6dEB, it was unknown whether native pathways like those associated with Sbm, YgfG, and YgfH would positively or negatively affect 6dEB production. Therefore, the deletion and over-expression of the *sbm*, *ygfG*, and *ygfH* genes were performed to investigate the impact on 6dEB production. The effect was first characterized under conditions meant to improve 6dEB production, namely, when both the *S. coelicolor pcc* genes and exogenous propionate were utilized.

Neither the deletion nor over-expression of *sbm* had any observable effect on 6dEB production as indicated in Figure 2.2.3. This result suggested that there should be no metabolic flux from (2*S*)-methylmalonyl-CoA to succinyl-CoA (or vice versa) that significantly influences the substrate availability for 6dEB production. It has previously been confirmed that native *E. coli* methylmalonyl-CoA mutase could only catalyze (2*R*)-methylmalonyl-CoA/propionyl-CoA conversion (*70*). Therefore, lack of influence upon 6dEB production may imply the lack of a suitable (2*R*)/(2*S*)-methylmalonyl-CoA epimerase in *E. coli*. This possibility was supported by the absence of a clear epimerase homolog encoded within the *E. coli* genome. The over-expression of *ygfG* reduced 6dEB production supporting a model that favored YgfG-catalyzed conversion of (2*S*)-methylmalonyl-CoA to propionyl-CoA. This result suggested that YgfG may also be capable of accepting (2*S*)-methylmalonyl-CoA as a substrate or there exists enzymatic activity capable of converting (2*R*)-methylmalonyl-CoA to (2*S*)-methylmalonyl-CoA followed by conversion to propionyl-CoA by YgfG (*68*). Although *ygfG* over-expression was not favorable for 6dEB biosynthesis, the deletion of *ygfG* did not significantly impact 6dEB production levels, perhaps because the expression of the chromosomal *ygfG* gene provides sub-threshold levels of YgfG enzyme needed to observe the effects seen with over-expression. The 2-fold improvement of 6dEB titer through the deletion of *ygfH* indicated that the propionyl-CoA:succinate-CoA transferase encoded by *ygfH* produced succinyl-CoA from propionyl-CoA, reducing substrate level for 6dEB production. Over-expression of *ygfH* showed no statistically significant reduction in 6dEB, indicating that basal YgfH levels and those resulting from over-expression had a similar impact upon heterologous 6dEB biosynthesis.

The impact of *ygfH* deletion was further analyzed in a batch bioreactor setting. The scaled and well-controlled environment provided by the bioreactor allowed significantly improved 6dEB levels and helped to emphasize the positive impact provided by the *ygfH* deletion. Compared to the control strain BAP1 (pBP130/pBP144), the *ygfH* mutant strain TB3 harboring pBP130 and pBP144 confirmed the smaller-scale shake flask comparisons with 6dEB titer increasing from 206 to 527 mg/L. This study indicated how straightforward metabolic engineering techniques, coupled to modest process scale-up, could be used to significantly increase titers in complex compound biosynthesis even without extension to fed-batch bioreactor configurations previously used to boost titers to > 1 g/L (*54*, *75*).

(2) 6dEB Production in the Presence of Propionate but without *pcc* Genes

In the course of this study, lower levels of 6dEB could be quantified in the context of production efforts without the PCC enzyme complex. This observation (and the similar situation in which production is observed without exogenous propionate addition) was of interest because it was unknown to what extent *E. coli* native metabolism may be able to support heterologous 6dEB biosynthesis. The *S. coelicolor* PCC is responsible for the production of (2*S*)-methylmalonyl-CoA in the engineered *E. coli* strain used to generate 6dEB. However, it was found that an *E. coli* BAP1 strain harboring DEBS but no PCC complex could still produce low levels of 6dEB in the presence of propionate. The significant decrease in production level was consistent with our previous results that suggested the relative importance of the heterologous PCC in the *E. coli* 6dEB production scheme (*67*).

The biosynthesis of 6dEB in the absence of the *pcc* genes indicated that *E. coli* was able to produce (2*S*)-methylmalonyl-CoA through native metabolism. The methylmalonyl-CoA mutase, Sbm, and methylmalonyl-CoA decarboxylase, YgfG, are involved in (2*R*)-methylmalonyl-CoA metabolism. However, 6dEB production was unchanged after deletion of the *sbm* and *ygfG* genes, suggesting that the required (2*S*)-methylmalonyl-CoA was not produced through the activity of these enzymes. The data then support other yet-unknown native *E. coli* pathways capable of directing the biosynthesis of the (2*S*)-methylmalonyl-CoA substrate for 6dEB production. These pathways, however, are dependent upon exogenous propionate since no 6dEB was produced in the absence of both the *pcc* genes and propionate (Figure 2.2.6).

Previous work has reported that the *E. coli* methylmalonyl-CoA mutase encoded by the *sbm* gene specifically catalyzes the conversion of succinyl-CoA to (2*R*)-methylmalonyl-CoA. In the same study, co-expression of the *P. shermanii* methylmalonyl-CoA mutase and epimerase in *E. coli* allowed (2*S*)-methylmalonyl formation and production of 6dEB at approximately 1 mg/L (*70*). To our knowledge, *E. coli* does not possess a methylmalonyl-CoA epimerase capable of interconversion of (2*R*)- and (2*S*)-methylmalonyl-CoA. This may then explain the lack of impact upon 6dEB levels caused by deleting the *sbm* gene in the absence of the *pcc* genes since it is uncertain whether *E. coli* Sbm can accept or produce the needed (2*S*)-methylmalonyl-CoA substrate. As shown in Figure 2.2.1, the *E. coli* YgfH is responsible for the CoA moiety transfer between propionyl-CoA and succinyl-CoA. Without PCC, the *ygfH* deletions did not impact 6dEB production, suggesting that these enzymes primarily affect

propionyl-CoA levels and that (2*S*)-methylmalonyl-CoA provision was limiting under these conditions.

(3) 6dEB Production in the Presence of *pcc* Genes but without Propionate

Exogenous propionate was believed to be required for heterologous 6dEB production in *E. coli*. It was found in this study that *E. coli* supports 6dEB synthesis without the addition of exogenous propionate. In fact, the results indicated that *E. coli* synthesized propionyl-CoA from its native metabolic pathways. Data in Figure 2.2.6 support a metabolic model where succinyl-CoA can be converted to (2*R*)-methylmalonyl-CoA by Sbm and then to propionyl-CoA by YgfG, since the deletion of *sbm* and *ygfG* led to a decrease of 6dEB production levels. Similar results were observed by Aldor et al. who used the *E. coli* Sbm-YgfG pathway to produce propionyl-CoA in recombinant *Salmonella enterica* serovar Typhimurium (*76*). However, such a model would still require an epimerase activity to provide the (2*S*) isomer of methylmalonyl-CoA. As such, at this stage, we can only note that the results of Figure 2.2.6 implicate the influence of Sbm and YgfG in 6dEB production under the given conditions. The deletion of *ygfH* caused a similar decrease in 6dEB levels, which suggests that propionyl-CoA may be directly produced from succinyl-CoA by YgfH. Similar to the case of *pcc* gene removal, in the absence of exogenous propionate, there may also be other metabolic pathways involved in propionyl-CoA biosynthesis.

Chapter 3 Heterologous Biosynthesis of Erythromycin in *E. coli*

3.1 Introduction

Erythromycin was first discovered from a soil bacterium eventually termed *Saccharopolyspora erythraea* (*11*). After decades of engineering, the production of erythromycin by *S. erythraea* reached as high as 7 g/L (*30*). However, efforts to engineer the biosynthetic process were hampered by the lack of genetic and metabolic engineering tools available for this microorganism as well as the complicated metabolic background. With the identification and sequencing of the gene cluster responsible for erythromycin (*21-24*), heterologous production of erythromycin became feasible. To date, *S. lividans* and *S. coelicolor* have been utilized to make the polyketide precursor 6dEB. However the glycosylation of 6dEB to produce the final product erythromycin was still implemented by the native producer *S. erythraea* (*77, 78*).

Erythromycin production was also attempted using *E. coli* as the host. Two separate efforts began using either an analogous pathway from *Micromonospora megalomicea* or a hybrid pathway with tailoring biosynthetic genes from *S. fradiae*, *S. erythraea*, and *S. venezuelae* (*41, 43*). Similar to the approaches used for 6dEB production (*40*), several plasmids were constructed to contain the genes required for the glycosylation process. Compared with *S. coelicolor* or *S. lividans*, *E. coli* is much easier to engineer using existing protocols. Indeed, the use of an *E. coli*-plasmid system resulted in the production of the desired erythromycin products in both cases. In the effort using biosynthetic genes from *Micromonospora megalomicea*, erythromycin C was produced based on the functional expression of erythromycin glycosylation genes. In the other

study utilizing genes from *S. fradiae*, *S. erythraea*, and *S. venezuelae*, biosynthesis of an erythromycin analog 6-deoxyerythromycin D, was accomplished. It should be noted that both efforts also took advantage of accompanying commercially-available chaperone protein folding systems available to *E. coli* to support eventual biosynthesis. However, the product titer achieved for heterologous biosynthesis was still low. More importantly, erythromycin A, the most potent product, has not been produced by use of *E. coli* as a surrogate host. The previous efforts for incomplete biosynthesis of erythromycin A indicated that there must be challenges with successful reconstitution of one or a few key biosynthetic pathways involved in erythromycin A production in *E. coli*. In addition, the use of glycosylation genes from other microorganisms, instead of *S. erythraea*, in these studies suggested that active gene expression of the original *S. erythraea* glycosylation genes in *E. coli* was challenging.

Based on the previous successful heterologous production of the polyketide 6dEB, complete biosynthesis of erythromycin A in *E. coli* was achieved in this chapter. Specifically, sixteen erythromycin glycosylation genes from *Saccharopolyspora erythraea* (*eryBI-BVII*, *eryCI-CVI*, *eryF*, *eryK*, *eryG*) were employed for heterologous production of erythromycin (Figure 3.1). These genes, together with the erythromycin resistance gene (*ermE*), were introduced via plasmid vectors into *E. coli*. Several genetic, metabolic and process engineering approaches were adapted to facilitate the successful reconstitution of the biosynthetic pathway and to make the full erythromycin A product.

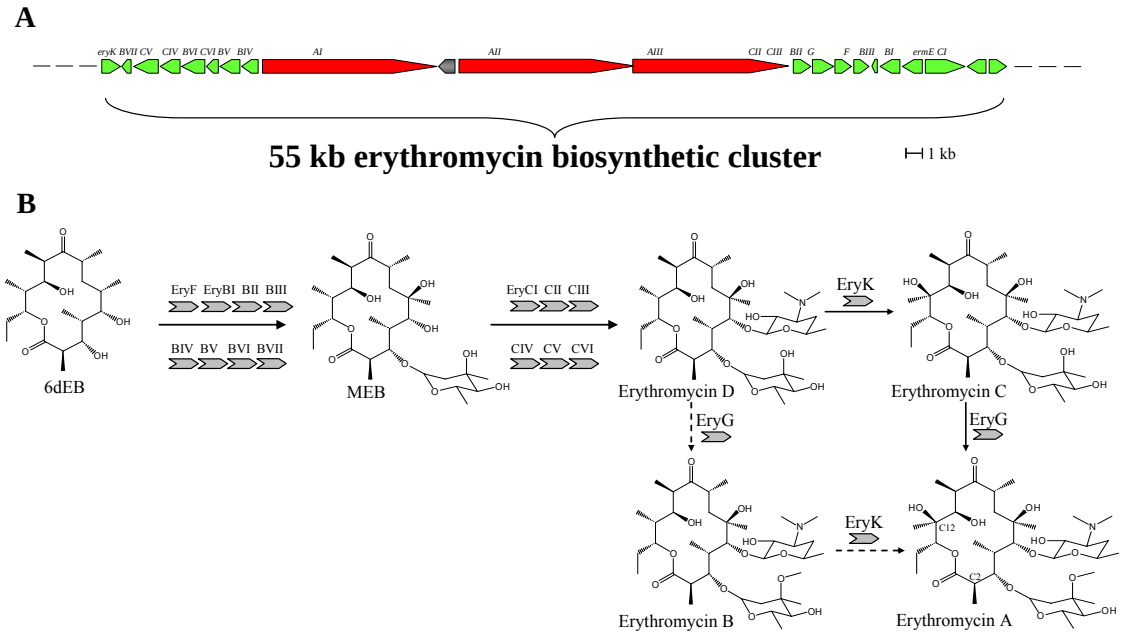


Figure 3.1 Biosynthesis of erythromycin A from polyketides precursor 6dEB. (A) Organization of erythromycin biosynthetic cluster on the *S. erythraea* chromosome. (B) Production of erythromycin A from 6-deoxyerythronolide B (6dEB) through intermediates (3-mycarosyl-erythronolide B) MEB, erythromycin D, C, and B. Solid arrows represent the favored route for erythromycin A biosynthesis in *S. erythraea*.

3.2 Materials and Methods

(1) Reagents and Chemicals

The reagents and chemicals used in this study were purchased from Fisher Scientific (Pittsburgh, PA, USA) or Sigma (St. Louis, MO, USA). All restriction enzymes and the Phusion High-Fidelity PCR Master Mix were purchased from New England Biolabs (Ipswich, MA, USA). PCR primers were synthesized by Operon (Huntsville, AL, USA). The chaperone plasmid kit and the pCDFDuet vector were purchased from Takara (Madison, WI, USA) and EMD Chemicals Inc. (Gibbstown, NJ, USA), respectively.

(2) Cell Culture Media and Growth Conditions

All *E. coli* cell cultures for molecular biology and SDS-PAGE analysis were conducted in Luria-Bertani (LB) medium at 37°C and 250 rpm. Heterologous 6dEB and erythromycin biosynthesis were conducted in production medium at 22°C and 250 rpm. One liter of production medium contained 5 g yeast extract, 10 g tryptone, 15 g glycerol, 10 g sodium chloride, 3 mL 50% v/v Antifoam B, 100 mM 4-(2-hydroxyethyl)-1-piperazineethanesuffonic acid (HEPES) buffer, and was adjusted to pH 7.6 by 5 M NaOH before use.

(3) Plasmid Design, Construction, and Gene Expression

Restriction enzyme digestions, SDS-PAGE, and other standard molecular biology techniques were performed as described by Sambrook et al (*60*). Erythromycin tailoring and resistance genes were PCR-amplified from *S. erythraea* genomic DNA using the primers listed in Table 3.1. The *Nde*I/*Eco*RI digested fragments of the PCR products (except for genes *eryBIII*, *BV*, and *CIII*) were individually inserted into plasmid pET21c digested with the same restriction enzymes. The *eryBIII* PCR product was digested by *Nhe*I/*Eco*RI and inserted into pET21c using the corresponding restriction sites. Similarly, the *eryBV* and *eryCIII* PCR products were digested and inserted by *Nde*I/*Sac*I and *Nde*I/*Hin*dIII, respectively. After transformation of the individual pET21c expression cassettes into BL21(DE3), cultures were induced overnight at 37^{o}C followed by sonication cell disruption. Cell lysates were clarified by centrifugation and soluble fractions were analyzed by SDS-PAGE.

The tailoring genes that could not be expressed using pET21c (*eryBI*, *BIII*, *BVII*, and *CIV*) were transferred to plasmid pET28a by *Nhe*I/*Hin*dIII (for *eryBIII*) or *Nde*I/*Hin*dIII (for *eryBI*, *BVII*, and *CIV*) digestion and subsequent ligation. SDS-PAGE

was performed as previously described to investigate expression from pET28a. The insertion of all erythromycin tailoring and resistance genes into individual expression plasmids was verified by Sanger sequencing at the Tufts University Core Facility.

Table 3.1 PCR primers for tailoring and resistance gene isolation from *S. erythraea* genomic DNA and for alteration of plasmid pBP130 and pBP144

Gene	Oligonucleotide sequence
BI	Forward 5'-CCGGGCATATG ATGACTGGGGGTGAG-3' Reverse 5'-GGGAATTCACTAGTTCAGAGGTTGATGTC-3'
BII	Forward 5'-TGGTCATATG ATGACCACCGACGCCGCG-3' Reverse 5'-GGGAATTCACTAGTTCACTGCAACCAGGCTTC-3'
BIII	Forward 5'-GGCCGCTAGCATGATCTTCCTTGTGGGACT-3' Reverse 5'-GGGAATTCCCACTAGTTCATACGACTTCCAGTCGGG-3'
BIV	Forward 5'-CCGGGCATATG ATGAATGGGATCAGTGATTCC-3' Reverse 5'-GGGAATTCACTAGTCTAGTGCTCCTCGGTGGG-3'
BV	Forward 5'-GGGCCATATGATGCGGGTACTGCTGACGTC-3' Reverse 5'-GGGAATTCACTAGTCTAGCCGGCGTGGCG -3'
BVI	Forward 5'-CCGGGCATATG ATGGGTGATCGGACCGG-3' Reverse 5'-GGGAATTCACTAGTTCATCCGGCGGTCCT-3'
BVII	Forward 5'-CTGGCATATG GTGGCGGGCGGTTTCGA-3' Reverse 5'-GGGAATTCACTAGTTCACCTGCCGGTGCT-3'
ermE	Forward 5'-CCGGGCATATG ATGAGCAGTTCGGACGAGC-3' Reverse 5'-GGGAATTCACTAGTCTACCGCTGCCCGGG-3'
CI	Forward 5'-CCGGGCATATG ATG GACGTCCCCTTCC-3' Reverse 5'-GGGAATTCACTAGTTCAAGCCCCAGCCTTGAG-3'
CII	Forward 5'-CCTAGCATATG ATGACCACGACCGATCGCG-3' Reverse 5'-GGGAATTCACTAGTTCAGAGCTCGACGGGGCA-3'
CIII	Forward 5'-CGGCCATATGATGCGCGTCGTCTTCTCCTC-3' Reverse 5'-GGGAATTCACTAGTT CATCGTGGTTCTCTCCT-3'
CIV	Forward 5'-CCGGGCATATG ATGAAACGCGCGCTG-3' Reverse 5'-GGGAATTCACTAGTTCACGAACCGTTGCG-3'
CV	Forward 5'-CCGGGCATATG ATGAACACAACTCGTACGGCA-3' Reverse 5'-GGGAATTCACTAGTTCACCTTCCGCGCAG-3'
CVI	Forward 5'-CCGGGCATATG ATGTACGAGGGCGGGTTC-3' Reverse 5'-GGGAATTCACTAGTTCATCCGCGCACACC-3'
eryF	Forward 5'-CCGGTCATATG ATGACGACCGTTCCCGA-3' Reverse 5'-GGCGGAATTCACTAGTTCATCCGTCGAGCCG-3'
eryG	Forward 5'-CCGGTCATATG GAGCACAAGAGGAACCGA-3' Reverse 5'-GGCGGAATTCACTAGTGGTCACCGAGGTGGC-3'
eryK	Forward 5'-CCGGTCATATG TTGACCACCATCGACGA-3' Reverse 5'-GGCGGAATTCACTAGTCTACGCCGACTGCCT-3'
apr	Forward 5'-GAAGAGTATGAGTATTCAACATTTCCGTGTCGCCCTTATTCCCTTTGCTCACATGCAT-3' Reverse 5'-TTAATCAGTGAGGCACCTATCTCAGCGATCTGTCTATTTCGTTCATGAAGCTGAAAGC-3'
tet	Forward 5'-GCTTACATAAACAGTAATACAAGGGGTGTTATGAGCCATATTCTCATGTTTGACAGCTT-3' Forward 5'-ACCTGAATCAGGATATTCTTCTAATACCTGGAATGCTGTTTTCCCGAGATGCGCCGCGT-3'

After SDS-PAGE confirmation of gene expression, the *eryBI*, *BII*, *BIII*, *BIV*, *BV*, *BVI*, *BVII*, and *ermE* genes were combined (in order) into one operon by sequential *Xba*I/*Spe*I and *Sac*I digestion and ligation. The resulting plasmid pHZT1 was a derivative of the kanamycin-resistant pET28a plasmid. Genes *eryCI*, *CII*, *CIII*, *CIV*, *CV*, *CVI*, *eryF*, *eryG*, and *eryK* were combined (in order) into one operon by sequential *Xba*I/*Spe*I and *Hin*dIII digestion and ligation. The resulting plasmid pHZT2 was a derivative of the ampicillin-resistant pET21c plasmid.

In order to construct a plasmid carrying an additional *eryK* gene, PCR and *Nde*I/*Xho*I digestion preceded ligation into the streptomycin-resistant pCDFDuet-1 vector. The resulting plasmid was named pHZT4. In addition, the chaperone expression plasmids pG-KJE8, pGro7, pKJE7, pG-Tf2, and pTf16 were individually co-transformed with plasmids pHZT1, pHZT2, and (when indicated) pHZT4 into *E. coli* BL21(DE3) to test the effects each chaperone or chaperone combination had on erythromycin biosynthesis.

For one-cell production of erythromycin A, plasmid pBP130 containing the genes for DEBS2 and DEBS3 (*40*) was modified by replacing the ampicillin-resistance marker with an apramycin-resistance marker through λ-Red recombination (*58*). Specifically, the apramycin (apr) resistance gene was PCR-amplified using plasmid pSBAC as a template (*79*)(Table S1). Purified PCR product and plasmid pBP130 were co-transformed into *E. coli* TOP10 harboring pKD46 through electroporation, and the resulting culture was grown in LB medium overnight at 37°C. Plasmid product was then isolated from the overnight culture and transformed into *E. coli* TOP10 for selection on an LB-agar plate containing 50 mg/L apramycin. Positive colonies were cultured and the

resulting plasmids assessed by restriction analysis. The final plasmid was named pBPJW130.

In a similar fashion, the kanamycin resistance marker on plasmid pBP144 (containing the genes required for the PCC and DEBS1 (*40*)) was changed to a tetracycline resistance marker. Here, the tetracycline resistance gene was PCR-amplified from plasmid pBR322 (Table S1). The kanamycin marker of pBP144 was replaced analogous to the procedure to modify pBP130 and the resulting plasmid was named pBPJW144.

(4) Erythromycin Biosynthesis

For serial, two-cell erythromycin production, 6dEB was first generated using strain BAP1(pBP130/pBP144) as previously described. The 6dEB compound was quantified using an HPLC-Evaporative Light Scattering Detector and the 6dEB analogs were quantified by MS as previously described (with purified 6dEB as a standard) (*80*). All MS analyses used electrospray ionization and were conducted in positive ion mode.

To complete the two-cell production process, an overnight culture of *E. coli* BL21(DE3)(pHZT1/pHZT2/pHZT4/pGro7) was inoculated (5% v/v) into production 1 mL medium containing 100 mg/L carbenicillin, 50 mg/L kanamycin, 20 mg/L chloramphenicol, 50 mg/L streptomycin, 100 μM IPTG, and 2 mg/mL arabinose (pGro7 required arabinose induction) and cultured at 22^{o}C for 24 hours. The BAP1(pBP130/144) culture containing 6dEB (produced as described in the preceding paragraph) or the 6dEB (or 6dEB analog for erythromycin analog production) methanol extracts was then added to final polyketide concentrations of 50 mg/L, and erythromycin production commenced by culturing for an additional 3 days at 22^{o}C.

For one-cell erythromycin A production, an overnight culture of *E. coli* BAP1(pBPJW130/pBPJW144/pHZT1/pHZT2/pHZT4/pGro7) was inoculated (5% v/v) into 1 mL production medium containing 100 mg/L carbenicillin, 50 mg/L kanamycin, 20 mg/L chloramphenicol, 50 mg/L streptomycin, 50 mg/L apramycin, 5 mg/L tetracycline, 100 μM IPTG, 2 mg/mL arabinose, and 20 mM propionate and grown at 22°C for 7 days.

(5) *Bacillus subtilis* Growth Inhibition Bioassay

To confirm the antibiotic activity of *E. coli*-derived erythromycin A, a *B. subtilis* zone-of-inhibition bioassay was performed. Methanol solutions of erythromycin extracts were adjusted to specific concentrations based upon the results from quantification. This solution was then added to a filter paper disk and placed upon an LB agar plate prepared by mixing 20 μL of an overnight *B. subtilis* culture with 20 mL liquified LB agar (maintained at 45ºC). After overnight incubation at 37^{o}C, *B. subtilis* inhibition zones were assessed for positive control, negative control, and experimental filter disk samples. The assay was repeated a minimum of three times from independent culture experiments to ensure reproducibility.

(6) Selection of robust erythromycin producing strain

In order to further improve the erythromycin production, the erythromycin producing *E. coli* strain was subjected to high concentrations of erythromycin and inducers. Specifically, BL21(DE3, pHZT1/pHZT2/pGro7) was grown in LB medium at 37 ^{o}C overnight and then diluted 4×10^{4} fold by fresh LB medium. 100, 300, and 600 μl of diluted cultures were spread onto LB agar plates containing 100 mg/L ampinicillin, 50 mg/L kanamycin, 20 mg/L chloroamphenicol, 100 mM IPTG, 2 mg/mL arabinose and erythromycin ranging from 10-500 mg/L. After the incubation at 37 ^{o}C, the colonies on

the plates with different erythromycin concentrations were counted and this experiment was performed 4 times and the results were presented as average numbers. The colonies were then cultured for erythromycin production. Specifically, the colonies were picked and transformed with the plasmid pHZT4. The resulting strains were grown in LB medium for starter culture preparation. After overnight growth, the culture was inoculated (5% v/v) into 1 mL production medium. After 2 days growth, 6dEB extract was supplemented into the culture to make a final concentration of 50 mg/L. The mixed culture was then grown for additional 3 days before extracted by ethyl acetate and analyzed by LC-MS. 2.5 μg/L Roxithromycin was used as an internal standard for quantification.

(7) Erythromycin Production Analysis

One milliliter of the *E. coli* cultures containing erythromycin A or the erythromycin analogs was extracted with 0.5 mL ethyl acetate. After centrifugation at 12,000 rpm for 1 minute, the extract was transferred to a 1.5 mL micro-centrifuge tube, air-dried, and re-suspended in 100 μL of methanol. Samples were then analyzed using an LTQ XL Linear Ion Trap Mass Spectrometer (Thermo Electron Corporation) coupled with a Finnigan Surveyor LC system (Thermo Electron Corporation) with an Agilent ZORBAX Eclipse XDB-C18 HPLC column (Santa Clara, CA, USA). A linear gradient of 100% water to 100% acetonitrile over 15 minutes was used at a flow rate of 0.6 mL/min. The generated LC chromatograms were filtered with the molecular weight (with a width of ± 0.5 m/z) of the target compounds, including erythromycin A, B, C and D. Where indicated, MS/MS analysis was performed using a collision energy of 25V.

To quantify erythromycin A and analog production, an LC-MS calibration curve was prepared using commercially available erythromycin A as an external standard and roxithromycin as an internal standard. Known amounts of erythromycin A were added to *E. coli* BL21(DE3) or BAP1 cultures grown under the same conditions described above. After ethyl acetate extraction, the samples were dissolved in 100 μL of methanol containing 2.5 mg/L roxithromycin and subjected to LC-MS analysis to prepare the calibration curve. A portion of the extracts resulting from the erythromycin producing *E. coli* cultures was similarly mixed with roxithromycin and analyzed against the calibration curve. The ratio of the erythromycin A and roxithromycin standard peak areas was correlated to quantify experimental erythromycin production with a suitable calibration curve made before every experimental analysis. All reported titers represent at least three independent experiments. Negative controls included 1) experiments reliant on strain BAP1 without the required production plasmids, 2) un-induced cultures, and 3) the replacement of pHZT2 with an empty pET21c expression vector.

3.3 Results and Discussion

(1) Tailoring Biosynthetic Gene Isolation and Individual Expression

Erythromycin tailoring and resistance genes were first PCR amplified using erythromycin native producer *S. erythraea* chromosome as a template (Figure 3.2). Each gene was then placed into separate pET21c vectors to test individual gene expression upon induction. Though most genes expressed from pET21c (as assessed by SDS-PAGE, Figure 3.3), four proteins (EryBI, EryBIII, EryBVII, and EryCIV) were not observed when gene expression was induced from this vector. However, *eryBI*, *eryBIII*, *eryBVII*,

and *eryCIV* were successfully expressed from pET28a. The pET28a vector includes an N-terminal sequence (containing a 6×histidine tag) preceding the original start codon of the gene to be expressed, and this addition may account for the positive impact upon final protein production.

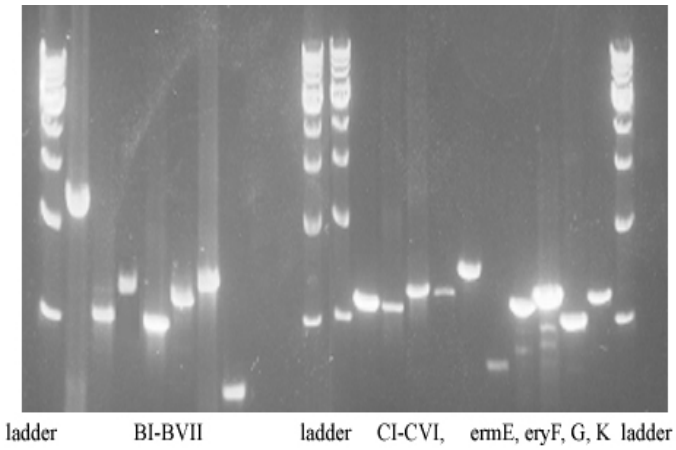


Figure 3.2 PCR amplification of erythromycin tailoring and resistance genes. Lane1-8: DNA ladder, *eryBI-BVII*; lane10-22: DNA ladder, DNA ladder, *eryCI-CVI*, *ermE*, *eryG*, *eryG*, *eryK* and DNA ladder.

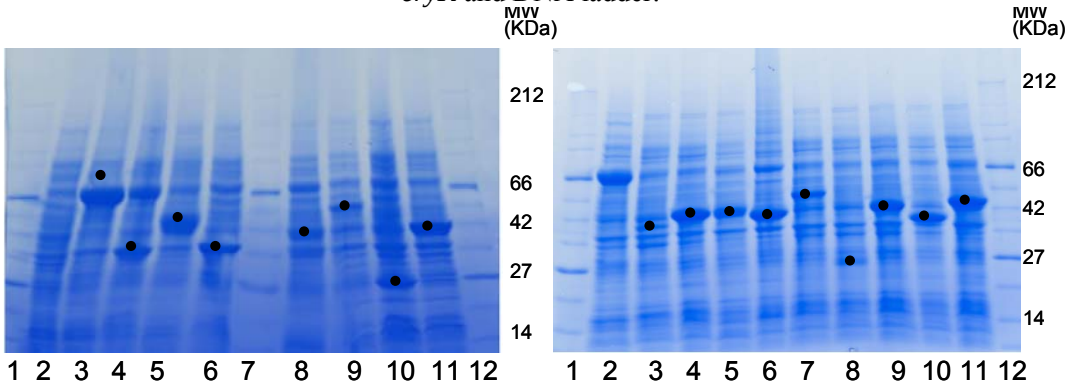


Figure 3.3 SDS-PAGE analysis of erythromycin tailoring and resistance gene expression. The protein products were indicated by black dots. The first gel, lane1-12: marker, BL21(DE3) control. BI-BIV, marker, BV-BVII ermE, Marker. The second gel lane1-12: marker, BAP1 control, CI-CVI, eryF, eryG, eryK, marker.

Convenient biosynthetic or phenotypic tests were then performed to confirm the activity of select enzymes, as shown in Figure 3.4. Specifically, the introduction of the

plasmid carrying *ermE* gene conferred the host cell resistance to erythromycin up to 400 mg/L. Similarly, 6dEB derivative, erythronolide B, was produced by feeding 6dEB substrate to the *E. coli* strain containing *eryF* gene.

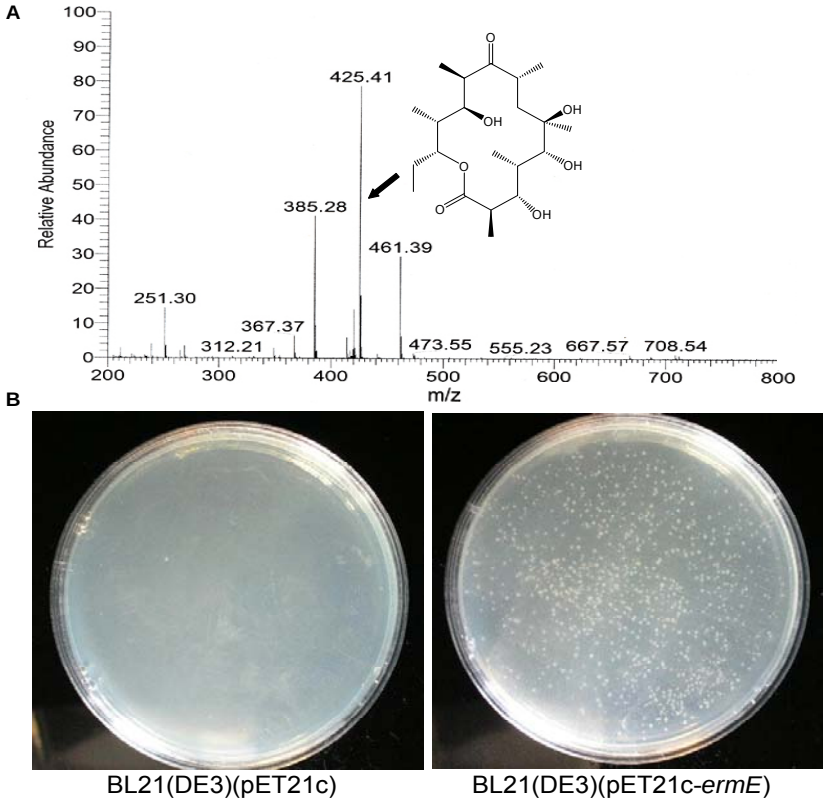


Figure 3.4 Assays to assess individual tailoring biosynthetic or resistance activity. (A) Mass spectrum for erythronolide B (425 m/z; sodium adduct) resulting from the conversion of 6dEB fed to cultures of BL21(DE3)(pET21c-*eryF*). (B) Resistance conferred by BL21(DE3)(pET21c-*ermE*). Both plates contain carbenicillin (100 mg/L), IPTG (100 μM), and erythromycin A (from Sigma, 50 mg/L).

Because most polyketide products (including erythromycin A) derive from hosts exhibiting a high G+C genetic content, alternative approaches to improving expression have included designed gene synthesis to eliminate rare codon usage within a new host or commercially-available hosts designed to accommodate G+C rich genes (*63, 81*).

However, our system relied totally on the original gene sequences from *S. erythraea* and highlighted the native capabilities of *E. coli*, at least in this case, to accommodate significantly foreign genes and proteins. In addition, our approach was in contrast to earlier efforts which used a combination of tailoring genes from multiple sources (*S. erythraea*, *S. fradiae*, and *S. venezuelae*) that each demonstrated sufficient heterologous expression within *E. coli* (*43*). Here, we were able to directly reconstitute full activity from the original *S. erythraea* erythromycin A pathway.

(2) Tailoring Biosynthetic Operon Design

With successful individual gene expression confirmed, our attention turned to packaging the tailoring and resistance genes for transfer to and coordinated expression within *E. coli*. Operons were designed using a strategy similar to earlier efforts at introducing modular polyketide and nonribosomal peptide pathways to heterologous hosts (*39-41, 43, 82, 83*). Namely, synthetic operons were constructed according to the diagram in Figure 2.

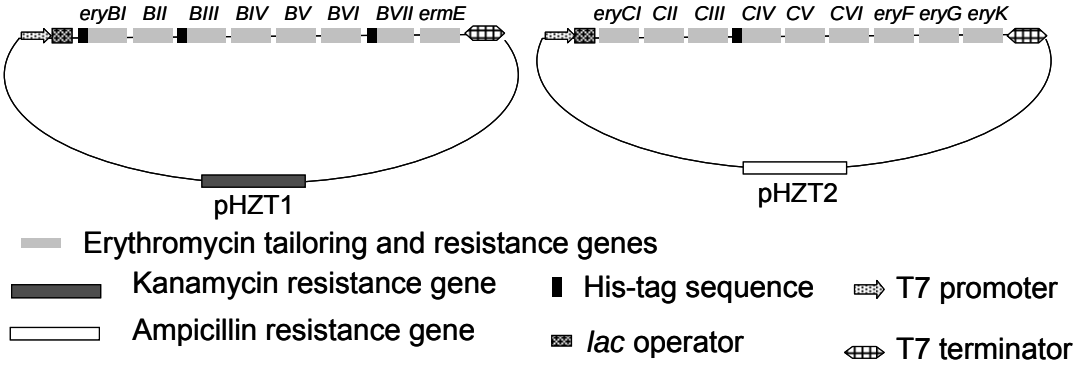


Figure 3.5 Operon design for the pHZT1 and pHZT2 plasmids.

Here, the cassettes demonstrating successful individual expression were used to construct two operons of eight and nine genes with each operon driven by a preceding T7 promoter under the control of a *lac* operator. It should be noted that no specific biosynthetic activity has been assigned to EryBI, and the removal of *eryBI* has been shown to have negligible effect on native *S. erythraea* erythromycin A production (*27, 84*). However, here, *eryBI* was included so as to retain any resulting beneficial activity within a heterologous environment. The erythromycin resistance gene *ermE* was similarly included to protect the *E. coli* host from nascent erythromycin A activity. The final operons were subsequently used in the following experiments to generate *E. coli*-derived erythromycin A, and coordinated gene expression and enzymatic activity were confirmed in the context of those experiments.

(3) *E. coli*-derived Erythromycin B and D Biosynthesis

Initial efforts at *E. coli* erythromycin biosynthesis consisted of producing 6dEB from a strain containing the required DEBS1, 2, and 3 enzymes followed by feeding the extracted 6dEB product to a separate *E. coli* strain containing the plasmids harboring the newly-constructed tailoring biosynthetic operons. LC-MS results showed that no erythromycin including A, B, C and D, was detected. However, we did see the accumulation of a large amount of erythronolide B (EB) which was converted from exogenous 6dEB by *eryF*.

The utilization of chaperone plasmids was then performed to help the folding of the glycosylation proteins and thus erythromycin production in *E. coli*. Chaperone plasmids pG-KJE8, pGro7, pKJE7, pG-Tf2 and pTf16 were transformed into BL21(DE3, pHZT1/pHZT2). The resulting strains were grown in the production medium and induced

by IPTG and arabinose. The culture was then extracted by ethyl acetate and subject to LC-MS analysis. The results showed that erythromycin B and D (peak at 718 and 704 m/z in Figure 3.6) were detected in both BL21(DE3) when pG-KJE8 and pGro7 were used. Further investigation demonstrated that chaperon vector pGro7, carrying the *E. coli* GroEL/ES chaperone system, helped produce more erythromycin B and D in *E. coli* than pG-KJE8. Therefore, pGro7 was used in all the following erythromycin heterologous biosynthesis studies. It should be noted that similar approaches to aid protein folding and/or association have been needed during previous complex natural product biosynthetic attempts (*40, 41, 43, 85, 86*), further highlighting the challenges associated with reconstituting such pathways.

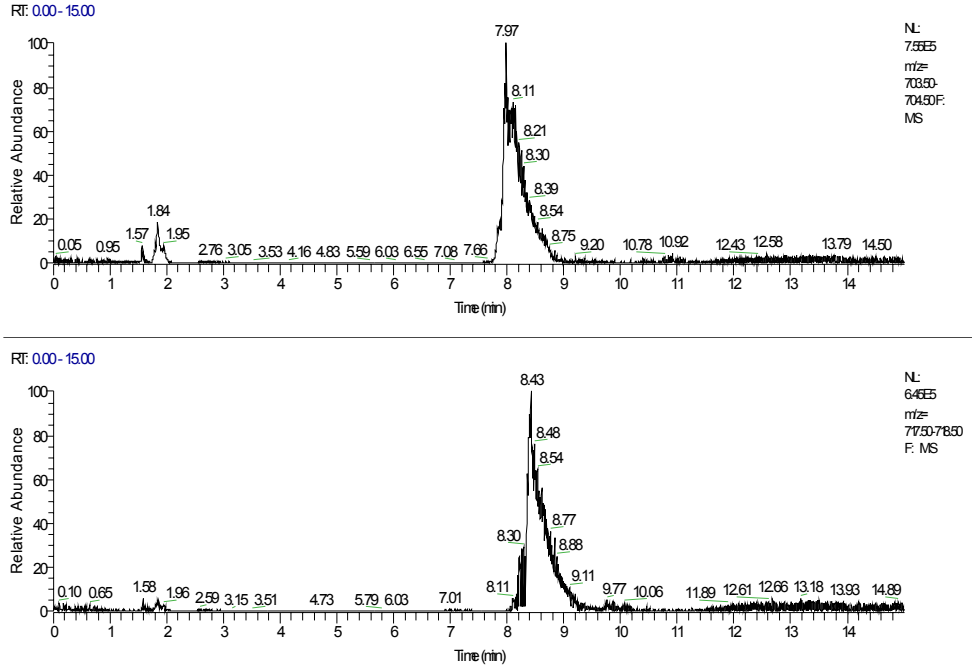


Figure 3.6 *E. coli*-erythromycin B and D production with the aid of chaperone plasmid. The first LC-MS chromatogram showed erythromycin D (704 m/z) production; the second LC-MS chromatogram showed erythromycin B (718 m/z) production. The LC peak intensity and m/z value are indicated on the right.

(4) *E. coli*-derived Erythromycin A and C Biosynthesis

Although erythromycin B and D were successfully produced by reconstitution of the erythromycin tailoring pathway in *E. coli*, the final and most potent form of erythromycin, erythromycin A, was not detected using the method described previously. In addition, erythromycin C was not found, either. In fact, the enzyme required for the conversion from erythromycin B to erythromycin A, and from erythromycin D to erythromycin C is encoded by the gene *eryK* which has already been integrated into the operon carried by pHZT2. The absence of erythromycin A indicated that eryK activity needed to be enhanced in order to facilitate erythromycin B and A formation. Given this status, we analyzed the positioning of *eryK* in our operon design. The *eryK* gene was the last of nine genes in one of the tailoring operons. As such, we hypothesized a lack of gene expression of *eryK*. To address this issue, an extra copy of eryK gene was cloned into a streptomycin-resistant plasmid pCDFDuet-1. This eryK-carrying plasmid was called pHZT4.

The introduction of pHZT4 into BL21 (DE3, pHZT1, 2, pGro7) resulted in the production of erythromycin A. As illuminated in Figure 3.7, both the retention time and mass-to-charge ratio of the erythromycin standard and the *E. coli*-derived erythromycin were the same, which clearly demonstrated the biosynthesis of erythromycin A in our *E. coli* system.

Erythromycin C that derived from erythromycin D by eryK was also found to be produced. The successful production of erythromycin A and C confirmed our hypothesis that more eryK activity was required for the completion of the whole heterologous

production process. LC-MS analysis also showed that EB, MEB, erythromycin B, C and D, along with a large amount of unused 6dEB were present in the culture.

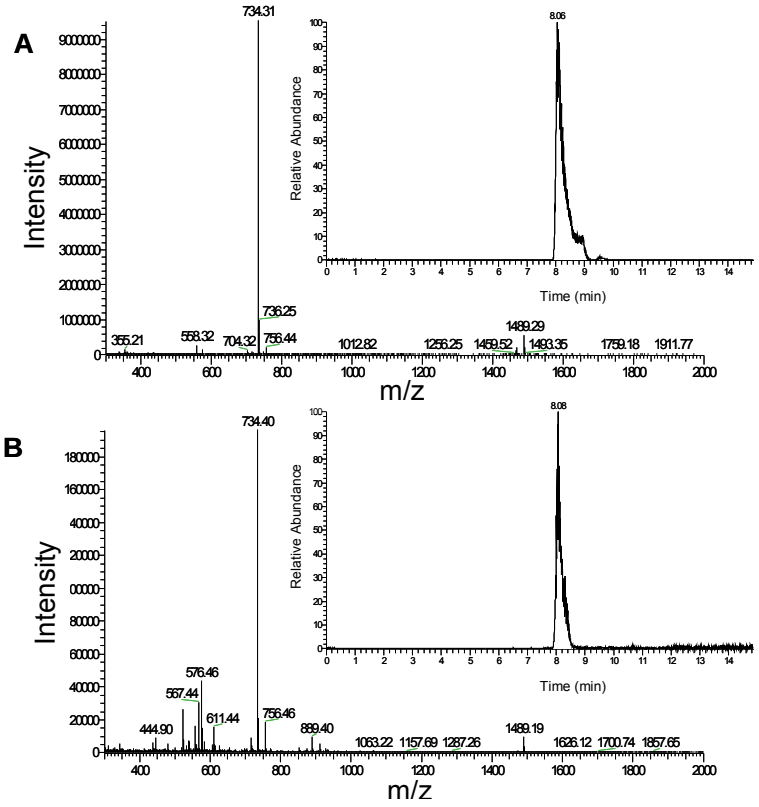


Figure 3.7 LC-MS analysis of erythromycin A production. (A) LC chromatogram and MS spectrum of erythromycin A standard. (B) LC chromatogram and MS spectrum of erythromycin A produced by the strain BL21(pHZT1/2/4/pGro7).

The erythromycin product was further characterized by MS/MS analysis. As shown in Figure 3.8 (A), *E. coli*-derived erythromycin A (756 m/z; sodium adduct) had the same MS/MS spectrum with that of the standard. This result further confirmed the production of the erythromycin A product in the developed *E. coli* system.

In order to study the antibiotic activity of *E. coli*-derived erythromycin product, a *Bacillus subtilis* growth inhibition assay was performed. As shown in Figure 3.8 (B),

erythromycin extract from an erythromycin-producing *E. coli* culture did show an inhibition activity against *Bacillus subtilis*, which was verified by negative and positive controls. Therefore, erythromycin A heterologous biosynthesis in *E. coli* was confirmed based on the above characterizations of the erythromycin A product.

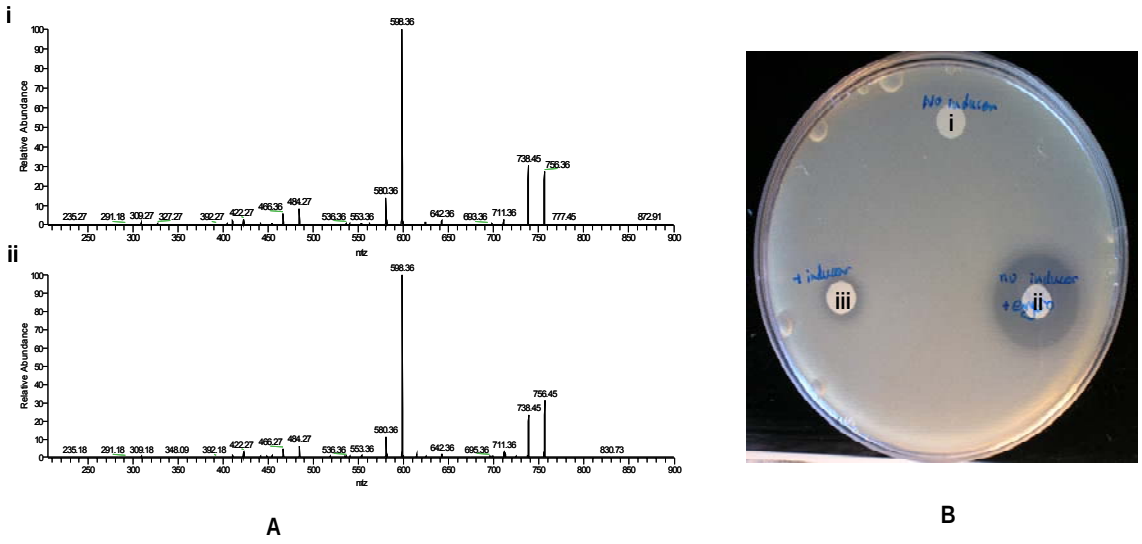


Figure 3.8 *E. coli*-derived erythromycin A characterization. (**A**) MS/MS analysis of *E. coli*-derived erythromycin A. (**i**) MS/MS spectrum of erythromycin A standard (756 m/z; sodium adduct). (**ii**) MS/MS spectrum of *E. coli*-derived erythromycin A. (**B**) *B. subtilis* filter disk zone-of-inhibition bioassay. The filter disks contain extracts from the following culture scenarios: (**i**) BL21(DE3)(pHZT1/pHZT2/pHZT4/pGro7) without IPTG or arabinose induction (negative control); (**ii**) BL21(DE3)(pHZT1/pHZT2/pHZT4/pGro7) without IPTG or arabinose induction but with the exogenous addition of 50 mg/L erythromycin A (from Sigma) (positive control); (**iii**) BL21(DE3)(pHZT1/pHZT2/pHZT4/pGro7) induced by IPTG and arabinose with erythromycin A production quantified at 10 mg/L by LC-MS.

(5) Selection of Robust Erythromycin-producing Strain

The finding that the enhanced *eryK* gene dosage was essential for production suggested that the gene at the end of the constructed operon may have an expression problem. Based on this hypothesis, it was expected that the expression of the *ermE* gene, placed at the end of the operon in pHZT1, could be improved as well. A potential

solution to this issue was to introduce a extra copy of the *ermE* gene to the *E. coli* host, as performed previously for the *eryK* gene. However, this approach may result in an even bigger metabolic burden on the host, as there were already 4 plasmids in the *E. coli* system for successful erythromycin A heterologous biosynthesis. Furthermore, once induced, the toxicity of the foreign gene products on the host was unknown.

Therefore, in order to further optimize the erythromycin production, a new approach was used. Specifically, directed evolution of erythromycin producing *E. coli* strain was performed by subjecting the parental erythromycin-producing *E. coli* strain to an environment containing inducers and different concentrations of erythromycin. It was expected that the strain, under the pressure of high inducers and erythromycin titer, would develop better protective/resistance mechanisms, such as enhanced product-pumping machinery and better tolerate the toxicity of the gene product or antibiotic product, and eventually produce more erythromycin.

Specifically, *E. coli* BL21(DE3, pHZT1/pHZT2/pGro7) was first grown in LB medium overnight and then diluted 4×10^4 fold with fresh LB medium. 100, 300, and 600 μL of diluted cultures were spread onto LB agar plates containing antibiotics, inducers and different concentrations of erythromycin. After overnight incubation, the colonies on the plates were counted and then cultured for erythromycin production.

As shown in Table 3.2, our results showed that there were a limited number of colonies on all the plates. Very interestingly, the colony number in this selection study did not change with the erythromycin concentration in the plates. Particularly, there was not a lawn of *E. coli* on the plate containing only inducers but no erythromycin. It was therefore suggested that the addition of exogenous erythromycin standard did not have

apparent effect on the colony numbers under the experimental conditions. However, whether this addition produced any genotype change was still unknown and needs to be further characterized. The finding also suggested that the presence of the inducers caused a toxicity issue on the cell, as only a few colonies grew on the plate containing only inducers and no erythromycin. It was then hypothesized that some *E. coli* BL21(DE3, pHZT12, pGro7) cells, when subjected to the inducers, evolved to adjust their metabolism to better tolerate the metabolic burden posed by the inducers.

Table 3.2 Number of colonies of selected robust strain
(The results are represented as the average number of 4 repeated experiments)

Spreading volume (μl)	Erythromycin concentration in the agar plate (mg/L)					
	0	50	100	200	300	500
100	0.3	1.25	0	0	0.25	0.75
300	2	2.66	4.25	2.75	1.5	2.5
600	7.5	2.66	6.3	3	3.33	5.66

To characterize cell growth of the strains before and after selection, one BL21 (DE3, pHZT1/pHZT2/pGro7) colony resistant to 500 mg/L of erythromycin was named "robust strain" and grown in LB medium containing inducers and erythromycin in comparison to the parental strain before selection (normal strain). As shown in Figure 3.9, the robust strain did show a faster growth rate than the normal strain in the presence of the inducers (IPTG and arabinose). Also, the addition of antibiotic erythromycin (100 mg/L) did not reduce the growth rate of either normal or robust strain.

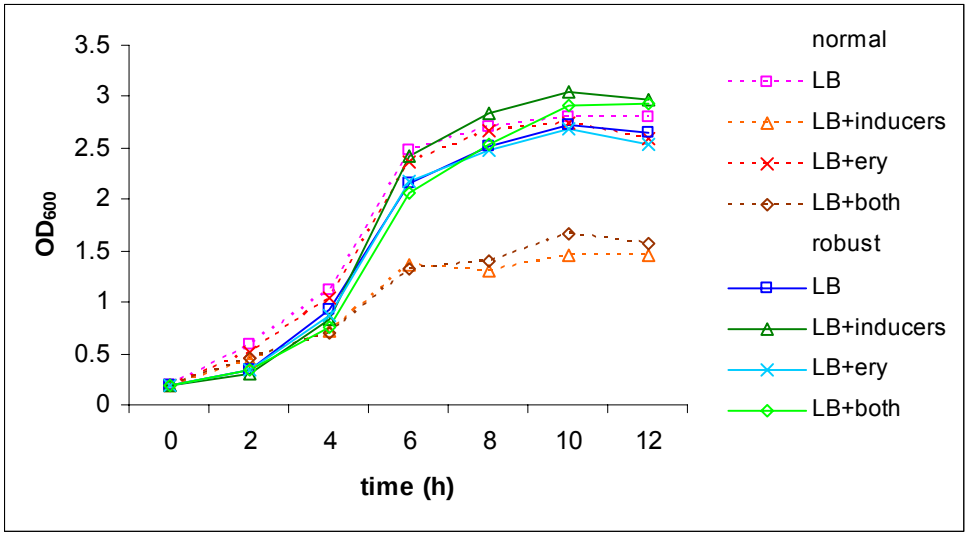


Figure 3.9 Growth curve comparison between normal and robust strain. LB: LB medium includes antibiotics for plasmid selection; Inducers: IPTG (100 μM) and arabinose (2 mg/mL); Ery: erythromycin A (from Sigma) at 200 mg/L.

Interestingly, the addition of inducers (alone or with erythromycin) inhibited the cell growth of the normal strain, further indicating that over-expression of heterologous genes exerted a heavy metabolic burden on the host cell. No obvious differences were observed with or without induction for the robust strain. The results demonstrated that the selected robust strain had a different phenotype that was influenced by inducers, and this phenotype change protected the robust strain against the heterologous gene products' toxicity that clearly inhibiting the growth of the normal strain.

A plasmid stability test was then conducted to further characterize the robust strain. As previously observed for those plasmids harboring the DEBS genes (*40*), plasmid instability was apparent for pHZT1 (61% retention 12 hours after induction) and pHZT2 (28% retention 12 hours after induction). This no doubt stemmed from both plasmids containing the same origin of replication, further supported by the 100% retention of pGro7 containing a different, compatible origin.

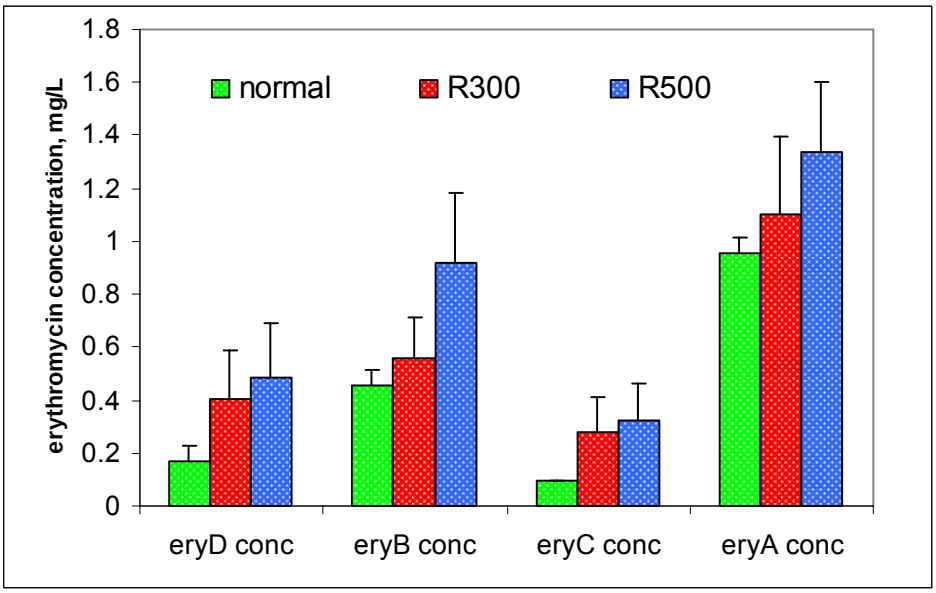


Figure 3.10 Erythromycin production by normal, R300, and R500 strain.

Erythromycin A production of the normal and robust strains was then analyzed and compared. Robust strains BL21(DE3, pHZT1/pHZT2/pGro7) picked from the plate containing 300 and 500 mg/L of erythromycin were named R300 and R500, respectively. They were first transformed with pHZT4 carrying an extra copy of *eryK*. The resulting strains were grown in the production medium supplemented with 50 mg/L 6dEB, and the production of erythromycin was analyzed by LC-MS. As shown in Figure 3.10, the R300 strain produced more erythromycin A than the normal strain, and the production of erythromycin A, B, C and D was further increased in strain R500. This result demonstrated that selected robust strains had better erythromycin-producing ability. The better the strain could resist the erythromycin's antibiotic activity, the more erythromycin product it could make (normal<R300<R500). Combined cell growth and erythromycin production results show that the developed method selected for *E. coli* strains that tolerate

not only the induced metabolic burden of heterologous pathway, but also the toxicity of the antibiotic product. These two effects together made the selected strain grow faster under the pressure of inducers and capable of producing more erythromycin product.

The next question to be addressed was whether the improvement of erythromycin production was due to the mutation of the erythromycin biosynthesis genes or *E. coli* chromosomal genes during the selection process. To answer that, the plasmids pHZT1/pHZT2/pHZT4/pGro7 contained by R500 strain were all transferred into a BL21(DE3) or normal strain without any plasmid. The production of erythromycin from the new normal strain was analyzed and compared with the original normal and robust strains. The result summarized in Figure 3.11 shows that the re-transformation of the erythromycin-producing plasmids back to the normal strain did not change the erythromycin production. The erythromycin titers of the normal and re-transformed strain were similar. The R500 strain still had the highest titers in all three strains. It was therefore suggested that the evolution or genotype change of the robust strain occurred in the native *E. coli* chromosome, not the introduced erythromycin biosynthetic gene operons. Although the exact information about the genotype change was not identified, it was considered that such change must enable the robust strains to tolerate high concentrations of inducers and erythromycin, and thus grow faster and produce more erythromycin product.

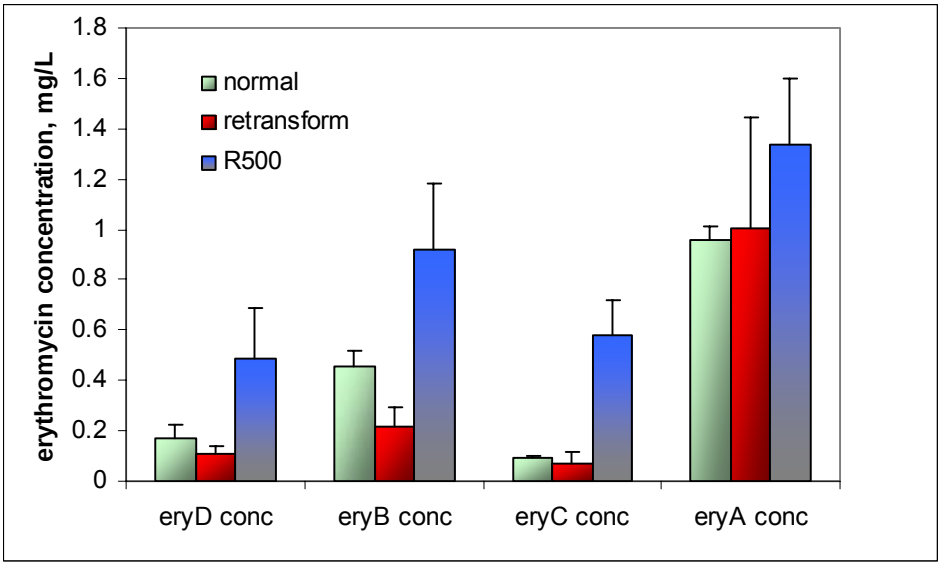


Figure 3.11 Erythromycin production comparison for re-transformed normal strain. Normal strain, normal strain retransformed with erythromycin biosynthetic plasmids isolated from R500 strain, and R500 strain were used for comparison.

Erythromycin biosynthesis in all previous efforts was achieved by feeding the BAP1(pBP130/144) culture containing 6dEB product to the *E. coli* strain harboring glycosylation gene operons. Next, the production of erythromycin was improved by using 25 g/L 6dEB extract, instead of the BAP1(pBP130/144) culture. This new strategy removed the use of the spent BAP1(pBP130/144) culture that would potentially contain potential unwanted metabolites. The titer of erythromycin produced by R500(pHZT1/pHZT2/pHZT4/pGro7) was then analyzed by LC-MS. Titers of erythromycin derivatives were found as follows: erythromycin A 11.1 mg/L; erythromycin B 3.5 mg/L; erythromycin C 2.3 mg/L; and erythromycin D 3.7 mg/L. The residual 6dEB was measured as 2.3 mg/L after the heterologous production process. Since the initial 6dEB concentration was 50 mg/L, around 95% of 6dEB added was utilized, among which 12.3% was converted to erythromycin A, 4.0% was converted to erythromycin B, 2.6% was converted to erythromycin C and 4.2% was converted to

erythromycin D (molar conversion). A significant amount of 6dEB was used to produce the intermediates including EB and MEB, which were both observed in LC-MS. In addition, it was also found that 90% of the erythromycin product was secreted to the surrounding medium, indicating the potential of our *E. coli* system for downstream product processing.

(6) One cell Erythromycin production

In an effort to consolidate the entire erythromycin A biosynthetic pathway to a single *E. coli* cell, the resistance markers of the plasmids responsible for 6dEB biosynthesis were altered such that one cell harbored all six plasmids required for both polyketide and tailoring biosynthesis. Specifically, plasmids pBPJW130 and pBPJW144 (responsible for 6dEB biosynthesis) together with glycosylation plasmids pHZT1, pHZT2, pHZT4, and pGro7, were introduced into *E. coli* BAP1. The feeding of the building block propionate to the constructed strain resulted in erythromycin A production at 0.6 mg/L (Figure 3.12).

Even though the final titer was reduced when compared to the sequential production strategy, one-cell production confirmed the capability of *E. coli* to coordinately express the 26 genes required for final erythromycin A biosynthesis from an inexpensive three-carbon building block of propionate. This strategy eliminates the need to obtain the polyketide precursor 6dEB from one *E. coli* strain, and feed it to the other *E. coli* strain for final erythromycin production formation. Thus, the production process was much simplified, saving the efforts for intermediate 6dEB extraction and purification, and reducing the required production time. Furthermore, titers are expected to improve with the use of the metabolic and process engineering approaches available for *E. coli*. For

example, *E. coli* fed-batch bioreactor protocols have previously been used to overproduce the 6dEB precursor to titers >1 g/L, a nearly 200-fold increase over original levels (*55*).

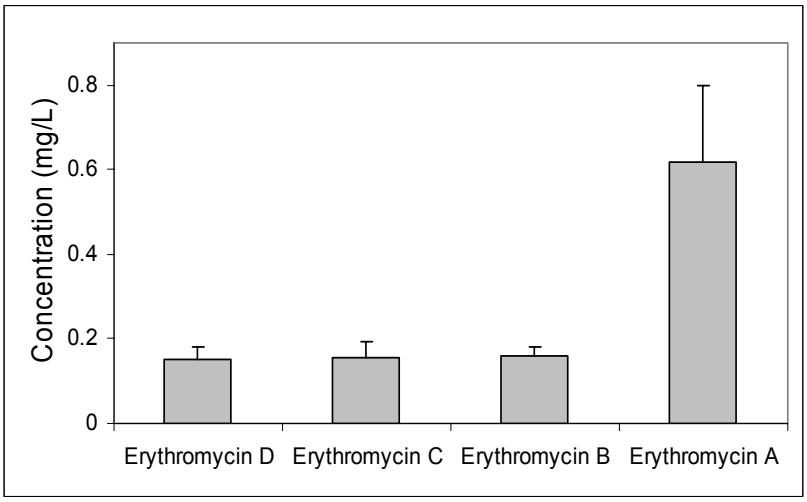


Figure 3.12 Erythromycin A titers from a one-cell *E. coli* production scheme. *E coli* BAP1(pBPJW130/pBPJW144/pHZT1/pHZT2/pHZT4/pGro7) was used in this strategy.

3.4 Discussion

There is no doubt about the positive impact antibiotics and other natural products have had on modern medicine. However, beneath this medicinal value lie the technical challenges of ready access to complex natural compounds. Production from intractable environmental sources limits the speed, economy, and structural diversity possible from reconstituted biosynthesis in technically-convenient heterologous hosts. Thus, heterologous biosynthesis offers the potential of rapid overproduction of original and analog forms of clinically-relevant natural compounds. The study in this chapter presents the complete production of erythromycin A using *E. coli* as a heterologous host. Specifically, heterologous biosynthetic pathway for erythromycin production was reconstituted by systematically transferring the required *S. erythraea* genes into *E. coli*.

Despite a potential codon bias problem, the introduced erythromycin biosynthetic genes were expressed in *E. coli*, facilitated by the use of a variety of commercially available plasmid vectors. The dilemma of harboring large genes or gene clusters has previously been addressed through the use of specialized plasmids capable of more stably maintaining large DNA fragments. Other options include cosmid or fosmid vectors (*87, 88*). In some cases, conventional cloning vectors have been redesigned to stably maintain up to 300 kb of foreign DNA(*89*). Alternatively, P1-derived artificial chromosome (PAC), bacterial artificial chromosome (BAC), or yeast artificial chromosome (YAC) plasmids have demonstrated the capability of stably maintaining 100-300 kb genetic inserts, sizes similar to many natural product gene cluster lengths(*90-93*). In this chapter, popularly used pET21c and pET28a plasmid vectors were used to contain all the glycosylation genes.

An *E. coli* chaperone system was then employed to address problems of protein folding. After coordinated expression of 23 foreign genes and 3 native genes, all required enzymes' activity was fully achieved, which enabled the conversion from starter molecule propionate to the final complex product erythromycin A. Analytical chemistry and bio-activity analyses were utilized to confirm the formation of the target product. In order to improve the production of erythromycin A, a strain selection method was developed to identify the *E. coli* strains that were able to resist the imposed metabolic burden and produce more erythromycin product. This method subjected the parental strain to high inducer and erythromycin concentrations and thus forced the cell to evolve for survival. A similar approach has been utilized to isolate *E. coli* strains that could tolerate high concentrations of ethanol (*94*). The obtained robust strain indeed showed

better erythromycin production, as well as better growth kinetics. The accomplishment indicated that the same technique may be adapted in other natural product biosynthesis system for production improvement.

The work in this chapter is significant for several reasons: 1) erythromycin A is an established antibiotic compound with a complex polyketide biosynthetic pathway; 2) the full heterologous production of this compound has not been reported and the steps to do so required overcoming several challenges in heterologous gene transfer, coordinated gene expression, metabolic engineering, and final biosynthetic activity; 3) un-optimized product titers have reached 10 mg/L, a starting point considered promising given the range of metabolic and process engineering options associated with *E. coli*. Besides providing access to the most potent form of erythromycin, the *E. coli* production platform offers numerous next-generation engineering opportunities. Optimal protein levels and activity, availability of required cofactors and precursors, conversion of substrate to product, and scaled bioreactor production are just of a few of the challenges to be addressed with the established engineering tools available to *E. coli*. Aiding this effort will be the range of molecular biology strategies to manipulate native and heterologous metabolism.

Chapter 4: Heterologous Biosynthesis of Erythromycin-Associated Polyketide Analogs in *E. coli*

4.1 Introduction

Biosynthesis of erythromycin associated polyketide analogs is of great research interest, as it provides a viable option for producing unnatural erythromycin products with potential novel bioactivity. The success of analog biosynthesis requires reliable methods to alter the chemical structure of erythromycin through genetic engineering means (*49, 95, 96*). In addition, a surrogate host is also needed to contain the constructed genetic materials and make the desired products directed by these materials. It has been reported that erythromycin analogs could be produced by feeding the polyketide precursor 6dEB analog to polyketide mutants of *S. erythraea*, resulting in the production of full erythromycin analogs. Using this design, numerous rationally designed 6dEB derivatives were generated using the *Streptomyces* strains and these analogs would then be converted to erythromycin derivatives via converter *S. erythraea* strains (*77, 78*). In addition, *S. lividans* has also been genetically modified such that the deoxysugar desosamine biosynthetic and auxiliary genes were integrated to the chromosome. The constructed *S. lividans* produced desosaminylated erythromycin derivatives from introduced polyketide precursor analogs. However, it should be noted that final product titer remained low and more importantly, only sugar desosamine was attached to make the incomplete erythromycin analog without the other sugar cladinose (*97*).

Nonetheless, as erythromycin native producer *S. erythraea* and previously constructed *S. lividans* strains are relatively hard to manipulate by genetic and metabolic engineering tools, interests have been spurred to rationally design erythromycin analogs

by use of the technically more amenable host *E. coli*. For example, it has been reported that 6dEB analogs could be produced in *E. coli* by engineering the biosynthetic pathway for polyketide formation (*98*).

After the establishment of an *E. coli* platform for the heterologous biosynthesis of native erythromycin, the production of designed erythromycin analogs was studied. The work in this chapter aims at the formation of novel erythromycin associated polyketide analogs by changing the molecular structure of erythromycin polyketide precursor and a sugar moiety in a designed biosynthesis fashion such that rationally designed erythromycin analogs can be produced in the *E. coli* host. The use of *E. coli* is of particular interest, because it is relatively easy to manipulate the heterologous genes in *E. coli*, compared with other microorganisms. In addition, *E. coli* native metabolism can be modified to meet the needs of the heterologous pathway, as shown in the previous chapters. More importantly, there is limited crosstalk between *E. coli* native and heterologous metabolic pathways, diminishing unwanted interference from native pathways. The success of erythromycin analog heterologous biosynthesis will provide a viable method to generate a variety of new "unnatural" natural compounds which could be used for production of compounds with novel biological activities.

In this chapter, novel erythromycin analogs are rationally designed and produced by changing the structure of the erythromycin polyketide ring and sugar group structure. This was achieved by modification of the enzymes or pathways responsible for corresponding structure biosynthesis.

The first study in this chapter focused on the biosynthesis of a erythromycin analog by changing the polyketide structure. It has been reported that a polyketide library

could be constructed by genetic modification of the erythromycin polyketide synthase (*44*). As a result, polyketide analogs containing a variety of structures were produced in *S. coelicolor* CH999 and *S. lividans* K4–114. However, to our knowledge, there is no report on glycosylation of these designed polyketide analogs to make the full erythromycin analog in *E. coli*. This chapter aims at designing and producing not only novel polyketide analogs, but also glycosylated erythromycin analogs in the surrogate host *E. coli*. The previously established erythromycin glycosylation system in *E. coli* was adapted to produce the complete erythromycin analog product with sugar moieties.

Specifically, a benzyl group was added to 6dEB and a methyl group was removed from the erythromycin polyketide precursor 6dEB. The structure of the designed analogs is shown in Figure 4.1. These modifications were achieved by altering the polyketide synthase to accept different substrates for analog biosynthesis. The altered polyketide synthase was then introduced into *E. coli* BAP1 capable of supporting natural product biosynthesis, and the resulting strains were utilized to produce the desired products. The obtained polyketide products were then glycosylated by the established *E. coli*-based erythromycin glycosylation platform to make the full erythromycin analogs. The substrate specificity of the enzymes responsible for sugar group transfer was also analyzed in the same study, as polyketide analogs possessed modified structures and can be challenging to be glycosylated.

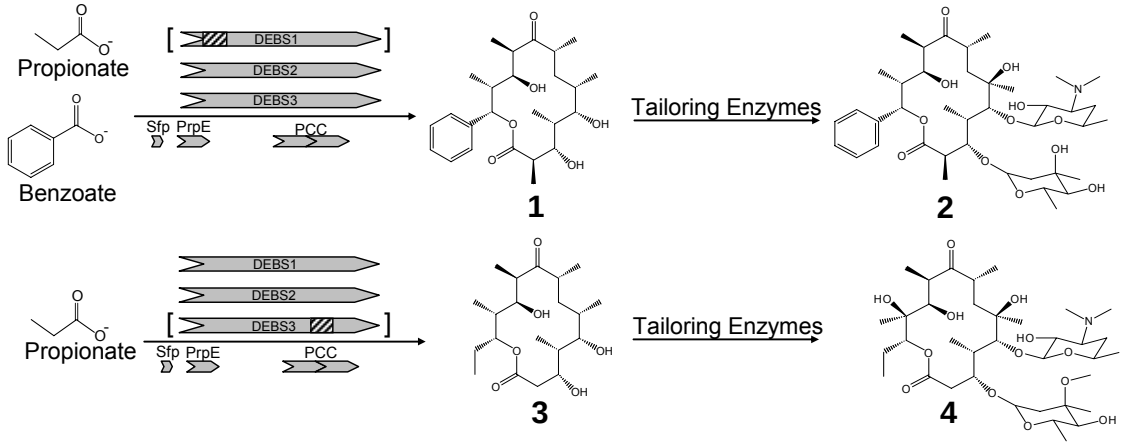


Figure 4.1 scheme of erythromycin analog biosynthesis by modification of polyketide structure.

Furthermore, erythromycin structure was modified by changing the erythromycin sugar groups. Native erythromycin contains two sugar groups, cladinose and desosamine. Intermediate 3-mycarosyl-erythronolide B (MEB, Figure 1.2) contains a native polyketide core and one native sugar group mycarose, and it was used as the substrate for the second sugar attachment. A heterologously biosynthesized novel sugar group was then utilized to attach to MEB, replacing the native sugar group desosamine and making an erythromycin analog. The sugar *N*,*N*-dimethyl pyrrolosamine was selected for the second sugar replacement. The rare amino sugar pyrrolosamine was first discovered as a component of β-pyrrolosporin (*99*). The *N*,*N*-dimethyl derivative of pyrrolosamine was also found to be the sugar group of the anticancer natural product lomaiviticin (*100*).

The *N*,*N*-dimethyl derivative of pyrrolosamine has a structure very similar to the erythromycin native sugar desosamine. They have the same molecular weight and the same functional groups, and they differ only in the position of the methyl, *N*,*N*-dimethyl and hydroxyl groups. To date, the biosynthetic pathways of pyrrolosamine and its *N*,*N*-dimethyl derivative have not been illuminated. However, several genes in the

lomaiviticin biosynthesis cluster were identified to be involved in the *N*,*N*-dimethyl pyrrolosamine formation. These genes include *lom32*, *lom33*, *lom36* and *lom30*, encoding a deoxyhexose epimerase, deoxyhexose reductase, deoxyhexose aminotransferase and deoxyhexose N-methyltransferase, respectively. Based on these gene products' putative function, a pyrrolosamine *N*,*N*-dimethyl derivative biosynthetic pathway was proposed in Figure 4.2. It should be noted that enzymes EryBVI and EryBVII for erythromycin native sugar group mycarose biosynthesis are also involved in the proposed pathway. The identified genes were then introduced in *E. coli* to make *N*,*N*-dimethyl pyrrolosamine sugar.

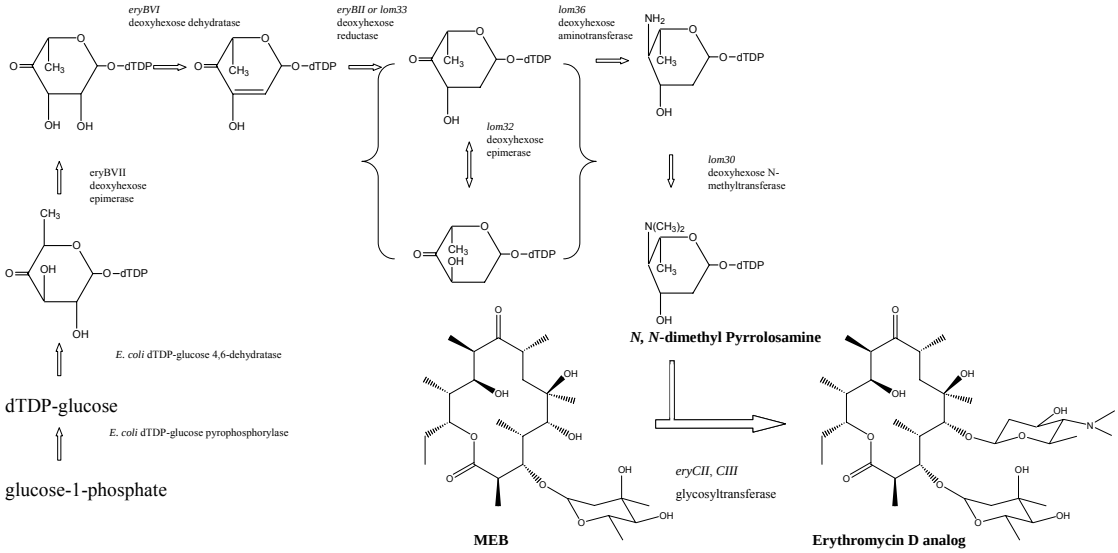


Figure 4.2 Proposed pyrrolosamine *N*,*N*-dimethyl derivative and erythromycin analog biosynthetic pathway.

Furthermore, sugar *N*,*N*-dimethyl pyrrolosamine was attached to the polyketide 6dEB to make a new erythromycin analog by two sets of glycosyltransferase pairs. The first set was glycosyltransferase DesVII and DesVIII isolated from *Streptomyces*

venezuelae. It has been reported that DesVII and DesVIII were flexible in recognizing different glycosylation substrates (*101*). In fact, it was found that DesVII and DesVIII were able to attach the erythromycin native sugar desosamine to the polyketide precursor MEB to make the complete erythromycin product (*102*). The second set of glycosyltransferase used in this study was the erythromycin native transferase pair EryCII and CIII. Successful reconstitution through *E. coli* thus offers an alternative route to the original compound and a platform for full manipulation of the biosynthetic pathway with the express purpose of expanding molecular diversity and antibiotic activity.

4.2 Materials and Methods

(1) Reagents and Chemicals

The reagents and chemicals used in this study were purchased from Fisher Scientific (Pittsburgh, PA, USA) or Sigma (St. Louis, MO, USA). All restriction enzymes and the Phusion High-Fidelity PCR Master Mix were purchased from New England Biolabs (Ipswich, MA, USA).

(2) Cell Culture Media and Growth Conditions

Heterologous 6dEB and erythromycin analog biosynthesis were conducted in production medium at 22°C and 250 rpm. One liter of production medium contained 5 g yeast extract, 10 g tryptone, 15 g glycerol, 10 g sodium chloride, 3 mL 50% v/v Antifoam B, 100 mM 4-(2-hydroxyethyl)-1-piperazineethanesuffonic acid (HEPES) buffer, and was adjusted to pH 7.6 by 5 M NaOH before use.

(3) Plasmid construction

For polyketide analog production, plasmids pBP165 (kanamycin-resistant) and pBP173 (ampicillin-resistant) were introduced to replace pBP144 and pBP130,

respectively. Plasmid BP165 was described previously (*40*) and contains the rifamycin loading di-domain instead of the native DEBS1 di-domain. Plasmid BP173 derives from the constructs presented in Liu et al. in which the acyltransferase domain of module 6 is replaced with the rapamycin module 2 acyltransferase such that the extender unit accepted at this position is malonyl-CoA (*103*). The plasmids pBP165 and pBP173 carrying the modified DEBS genes were constructed previously (*104*).

To reconstitute the *N*,*N*-dimethyl pyrrolosamine biosynthetic pathway, genes *lom*30, *lom*33, *lom*36 were PCR amplified using *Salinispora pacifica* genomic DNA, digested by *Nde*I and *Hind*III, and cloned into plasmid pET28a treated with the same endonucleases. To utilize reported flexible glycosyltransferases for pyrrolosamine *N*,*N*-dimethyl derivative attachment onto polyketide precursor (*101*), gene des*VII* and des*VIII* were PCR amplified using *Streptomyces venezuelae* genomic DNA, and cloned into pET28a by *Nde*I and *Hind*III digestion, followed by DNA ligation. The constructed plasmids were sequenced at the Tufts University Core Facility.

Table 4.1 PCR primers for *N*,*N*-dimethyl pyrrolosamine derivative biosynthetic gene and glycosyltransferase gene isolation

Gene	Oligonucleotide sequence
Lom30	Forward 5'-GGCCATATGACAGGCGTATACCGA-3' Reverse 5'-GGCAAGCTTACTAGT TCAGCCGCCACCGCCCGG-3'
Lom32	Forward 5'-GCCATATGGTGAAGGTCGAGAAGC-3' Reverse 5'-GCAAGCTTACTAGTTCAGTGGGCCACCTG-3'
Lom33	Forward 5'-GCCATATGATGCCCGCTCTCATCCGG -3' Reverse 5'-GCAAGCTTACTAGTTCACCCGCTCCGTGCGTCCCC -3'
Lom36	Forward 5'-GCCATATGATGAAAGACACGGTCGCC-3' Reverse 5'-GCAAGCTTACTAGTCTAGGAGGTGGCGAAGGG-3'
desVII	Forward 5'-GCCATATGATGCGCGTCCTGCTGACC-3' Reverse 5'-GCAAGCTTACTAGTTCAGTGCCGGGCGTCGGC -3'
desVIII	Forward 5'-GCCATATGATGACCGACGACCTGACG -3' Reverse 5'-GCAAGCTTACTAGTTCAGGAGCTGCTGACCGG-3'

The *lom30*, *lom32*, *lom33* and *lom36* genes, together with glycosyltransferase genes *desVII, desVIII* and erythromycin tailoring genes *eryCII, eryCIII, ery*F, *eryG* and *eryK*, were combined in order into one operon by sequential *Xba*I/*Spe*I and *Hind*III digestion and ligation. The resulting plasmid pHZT14 was a derivative of the ampicillin-resistant pET21c plasmid. The *lom30*, *33*, and *36* genes, together with erythromycin tailoring genes *eryCII, eryCIII, eryF, eryG* and *eryK*, were combined in order into one operon by sequential *Xba*I/*Spe*I and *Hind*III digestion and ligation. The resulting plasmid pHZT15 was a derivative of the ampicillin-resistant pET21c plasmid.

(4) Erythromycin Analog Biosynthesis and Analysis

For 6dEB analog **3**, BAP1(pBP173/pBP144) was cultured in 100 mL production medium containing 100 mg/L carbenicillin, 50 mg/L kanamycin, 100 μM isopropyl β-D-1-thiogalactopyranoside (IPTG), and 20 mM propionate at 22°C for 7 days. Analog **1** was produced similarly from BAP1(pBP130/pBP165) with 20 mM of benzoate (stock solution adjusted to pH 7) included with the propionate, IPTG, and selection antibiotics. Native 6dEB and both analogs were extracted with ethyl acetate, air dried, and re-suspended in methanol. The 6dEB compound was quantified using an HPLC-Evaporative Light Scattering Detector and the 6dEB analogs were quantified by LC-MS as previously described (with purified 6dEB as a standard) (*80*). All MS analyses used electrospray ionization and were conducted in positive ion mode.

To make erythromycin analogs by changing the polyketide structure, *E. coli* BL21(DE3, pHZT1/pHZT2/pHZT4/pGro7) was grown in production medium at 22^{o}C for 24 hrs after 5% v/v inoculation. Then the 6dEB analog substrate was fed into the culture, which was grown at 22^{o}C for an additional 3 days. The initial 6dEB analog **1** and **3**

concentrations were 3mg/L and 13mg/L in the culture, respectively. Erythromycin analog products were then extracted by ethyl acetate, air-dried and dissolved in 2.5 mg/L roxithromycin MeOH solution before being analyzed by LC-MS as introduced before.

To make erythromycin analogs by changing the sugar moiety, *E. coli* BL21(DE3, pHZT1/pHZT13/pGro7) or BL21(DE3, pHZT1/pHZT14/pGro7) was grown in the production medium at 22^{o}C for 24hrs after 5% v/v inoculation. Then the native EB or 6dEB substrate was fed into the culture, which was grown at 22^{o}C for an additional 3 days. Erythromycin analog product was then extracted by ethyl acetate, air-dried, and dissolved in 2.5 mg/L roxithromycin MeOH solution followed by LC-MS analysis as introduced in chapter 3, except that the a linear gradient of 100% water to 100% acetonitrile was used at a flow rate of 0.5 mL/min. .

4.3 Results

(1) Erythromycin Analog with C-13 Benzyl Group

The structure of the designed compound is shown in Figure 4.1. A 6dEB analog with a benzyl group attached to the C-13 location (designated as compound **1**) was biosynthesized by modification of polyketide synthase DEBS genes responsible for 6dEB production in *E. coli*. Specifically, the loading module of DEBS1 was replaced with the loading module from the rifamycin polyketide synthase such that a benzoate starter unit could be accepted, producing a benzyl addition to the erythromycin structure (*40, 105*). The polyketide analog **1** product was then extracted and analyzed by MS. Although there was no standard available for this analog's analysis, MS results showed the presence of expected analog **1** peak with the correct mass-to-charge ratio. No such peak was

identified in the negative control (*E. coli* strain BAP1(pBP130/pBP144) that could not produce any analog).

Once compound **1** production was confirmed, it was fed to the *E. coli* culture containing the erythromycin glycosylation system. The resulting erythromycin analog **2** possessed native erythromycin sugar moieties, cladinose and desosamine. The biosynthesis of the expected compound **2** was identified by LC-MS analysis (Figure 4.3). The results showed that a peak with expected mass/charge ratio 752 m/z was associated with the erythromycin analog **2**, and such a peak was not present in the native control sample (*E. coli* BL21(DE3, pHZT1/pHZT2/pHZT4/pGro7) grown under the same fermentation conditions except no polyketide analog **1** was fed. In addition, intermediates such as corresponding analog equivalents of erythronolide B and 3-mycarosyl-erythronolide B were also observed during final LC-MS analyses.

It should be noted that compound **2** was analogous to erythromycin D. However, no erythromycin A, B, or C analog was found by LC-MS. It was considered that the absence of these analog products could be attributed to the benzyl group hindering the activity of EryK, resulting in the lack of the erythromycin C analog and the following erythromycin A analog. Although the erythromycin B analog can be formed from the erythromycin D analog by the EryG enzyme, this pathway was not preferred as indicated in Figure 1.2, and the titer of the erythromycin B analog, if any, might be too low to be detected.

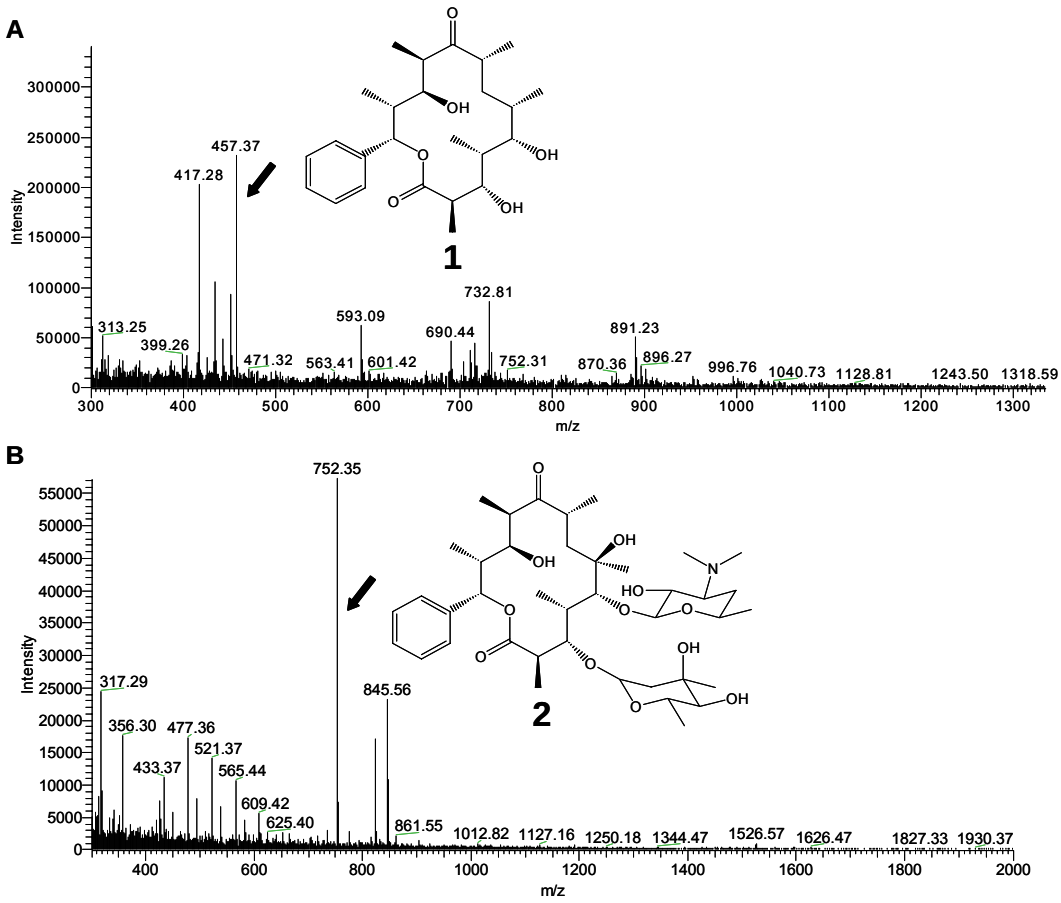


Figure 4.3 MS spectrum of polyketide analog **1** and erythromycin analog **2**.

Surprisingly, native polyketide precursor 6dEB was also produced by the constructed *E. coli* strain containing modified polyketide synthases, although the titer was low. As a result, after feeding the extract containing both native 6dEB and its analog to *E. coli* strain BL21(DE3, pHZT1/pHZT2/pHZT4/pGro7) for glycosylation, native erythromycin products were also produced in the same culturing process. Intermediates such as erythronolide B and 3-mycarosyl-erythronolide B were also observed by LC-MS analysis. The reason why modified polyketide synthases produced native 6dEB could be explained by propionyl-CoA being loaded directly to the ketosynthase domain of DEBS1, indicating the substrate flexibility of the constructed polyketide synthases that could still

utilize propionyl-CoA, in addition to benzoyl-CoA, as the starting unit and produce native 6dEB (*106-108*). The quantification of the erythromycin analogs was then performed based on the method developed for erythromycin production in the previous chapter. It was found that erythromycin D analog **2** was produced at a titer of 0.49 ± 0.19 mg/L. In addition, native erythromycin A was produced at a titer of 0.26±0.12 mg/L.

(2) Erythromycin Analog without C-2 Methyl Group

The structure of the second designed compound is shown in Figure 4.1. 6dEB analog without C-2 methyl (designated as compound **3**) was produced by altered DEBS genes responsible for 6dEB production. Specifically, module 5 of DEBS was modified to accept malonyl-CoA instead of (2S)-methylmalonyl-CoA as the extender unit (*103*). As a result, the corresponding product synthesized by the modified enzymes has no methyl group at the C-2 position, in contrast to the native 6dEB structure. The analog **3** was then extracted by ethyl acetate and analyzed by MS as discussed before. As shown in Figure 4.3 (A), the constructed *E. coli* BAP1(pBP173/pBP144) containing the modified DEBS genes produced a MS peak associated with analog **3**. No such peak in the MS spectrum was identified in the negative control (*E. coli* BAP1(pBP130/pBP144)) containing intact DEBS genes. Notably, low titers of native 6dEB were also found in the extract. The reason of this observation was not clear, but could be associated with the relaxed substrate specificity of the modified polyketide synthase module.

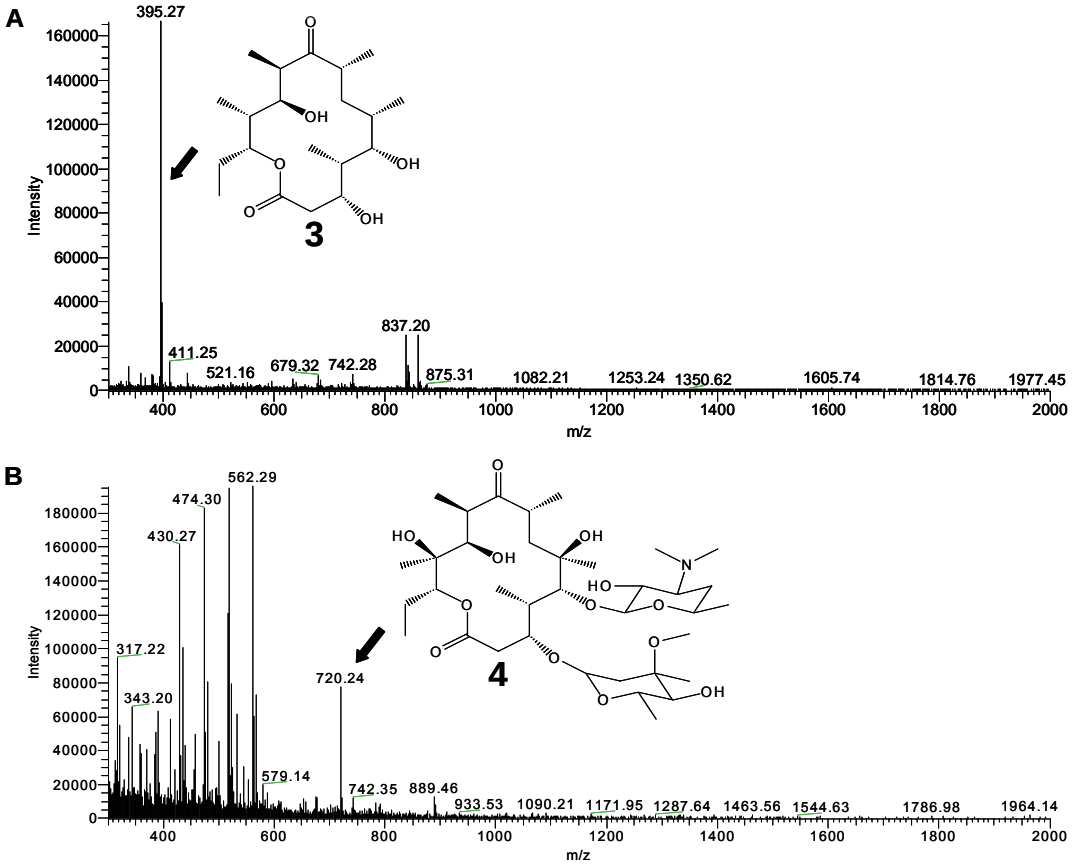


Figure 4.4 MS spectrum of polyketide analog **3** and erythromycin analog **4**.

The glycosylation of polyketide analog **3** was then attempted by using the erythromycin glycosylation system developed in the previous chapters. Specifically, compound **3** extract was fed into the BL21(DE3, pHZT1/pHZT2/pHZT4/pGro7) culture for sugar attachment. The glycosylated products were extracted from the culture by ethyl acetate, and subject to LC-MS analysis. As shown in Figure 4.4 (B), erythromycin A analog **4** without the methyl group at the C-2 position was successfully produced. There is no such peak in the native control sample (*E. coli* BL21(DE3, pHZT1/pHZT2/pHZT4/pGro7) grown under the same fermentation conditions except no

polyketide analog **3** was fed). In addition, erythromycin B, C, and D analogs were produced as well.

Similar to what was found for the first designed analog, native erythromycin products were also produced in the same culturing process as native 6dEB was also present in the feeding extract. It should be noted that erythromycin A analog **4** had the same molecular weight with native erythromycin C and these two compounds' MS peak overlapped in the spectrum. Similarly, erythromycin B analog had the same molecular weight with native erythromycin D. However, erythromycin D analog without the C-2 methyl group was clearly identified by LC-MS, as its molecular weight was not the same as any other erythromycin products and its associated LC-MS peak was not found in the negative control (*E. coli* BL21(DE3) culture without erythromycin biosynthesis genes). The LC-MS analysis also showed production of the erythromycin C and MEB analogs, whose molecular weight did not overlap with native erythromycin and intermediate MEB. Therefore, it was reasonable to argue that the erythromycin analog without the C-2 methyl group was produced, albeit that erythromycin A and B analogs' LC-MS peaks could be not separated from those of native erythromycin C and D.

The quantification of the erythromycin analogs was then performed based on the method developed for erythromycin production in the previous chapter. As erythromycin B and A analogs could be not distinguished from native erythromycin D and C in LC-MS, the calculation of the titers is based on the assumption that all corresponding LC-MS peaks were attributed to the analog production. Although it is not accurate, this method roughly estimates the titers of the erythromycin analogs produced by *E. coli* as 0.10± 0.12, 0.12± 0.09, 0.07±0.10 and 0.10±0.1 mg/L for erythromycin D, C, B and A analogs,

respectively. Notably, native erythromycin A was produced at a titer of 0.53±0.25 mg/L, even thought the native 6dEB concentration in the feeding substrate was very low. This observation demonstrated that when compared with the native 6dEB substrate, 6dEB analog **3** had lower glycosylation efficiency.

(3) Biosynthesis of Erythromycin Analog by Modification of the sugar moiety

Heterologous biosynthesis of erythromycin analogs by changing the sugar moiety's structure was achieved in this section. The proposed analog structure contained a *N,N*-dimethyl pyrrolosamine sugar that is also the sugar structure of the natural product lomaiviticin. The genes identified for *N,N*-dimethyl pyrrolosamine biosynthesis were *lom30*, *lom32*, *lom33*, and *lom36*, encoding a deoxyhexose *N*-methyltransferase, deoxyhexose epimerase, deoxyhexose reductase, and deoxyhexose aminotransferase, respectively. These genes were first PCR amplified and cloned into plasmid vectors. Genes *eryBVI* and *eryBVII* involved in erythromycin native sugar mycarose biosynthesis, together with *lom30*, *lom32*, *lom33*, and *lom36*, were utilized to reconstitute the complete *N,N*-dimethyl pyrrolosamine heterologous formation in *E. coli* (Figure 4.2).

To transfer *N,N*-dimethyl pyrrolosamine onto a polyketide precursor to make the complete erythromycin analog, a flexible glycosyltransferase pair of DesVII and DesVIII isolated from *Streptomyces venezuelae* was used in this study (*101, 109*). DesVII and DesVIII have been shown to be able to attach different sugar substrates to polyketide 6dEB in vitro (*102*). Therefore, this glycosyltransferase pair was employed for novel sugar attachment.

A plasmid, named pHZT13, containing genes *desVII* and *desVIII* and lomaiviticin genes *lom30*, *lom32*, *lom33*, and *lom36* was constructed (Figure 4.5). This plasmid was

then introduced into *E. coli* BL21(DE3) together with plasmid pHZT1 and chaperone vector pGro7 to make the erythromycin analog. After feeding the substrate EB to the resulting strain, erythromycin analog production was analyzed by LC-MS. However, no analog product was identified, implying that the sugar transfer using DesVII and DesVIII was not successful. As demonstrated in the previous section, native erythromycin desosamine transferase pair, EryCII and CIII, also had a certain level of flexibility to recognize different polyketides substrates (6dEB analog **1** and **3**) and make erythromycin analogs. Their ability to recognize new sugar groups (desosamine analog) and attach it to MEB was studied here. To achieve this, the erythromycin native glycosyltransferase pair e*ryCII* and *eryCIII* was added to the operon carried by pHZT13, resulting in a new plasmid pHZT14. Erythromycin modification genes *eryF*, *eryG*, and *eryK* were also included in this synthetic operon to make the corresponding erythromycin B, C, and A analogs. The structure of plasmid pHZT14 was shown in Figure 4.5.

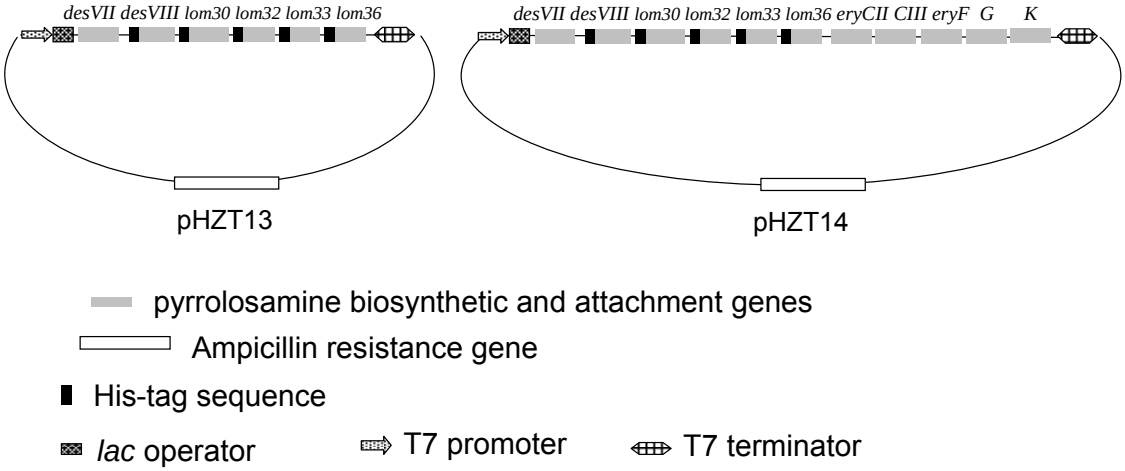


Figure 4.5 Operon design for the pHZT13 and pHZT14 plasmids.

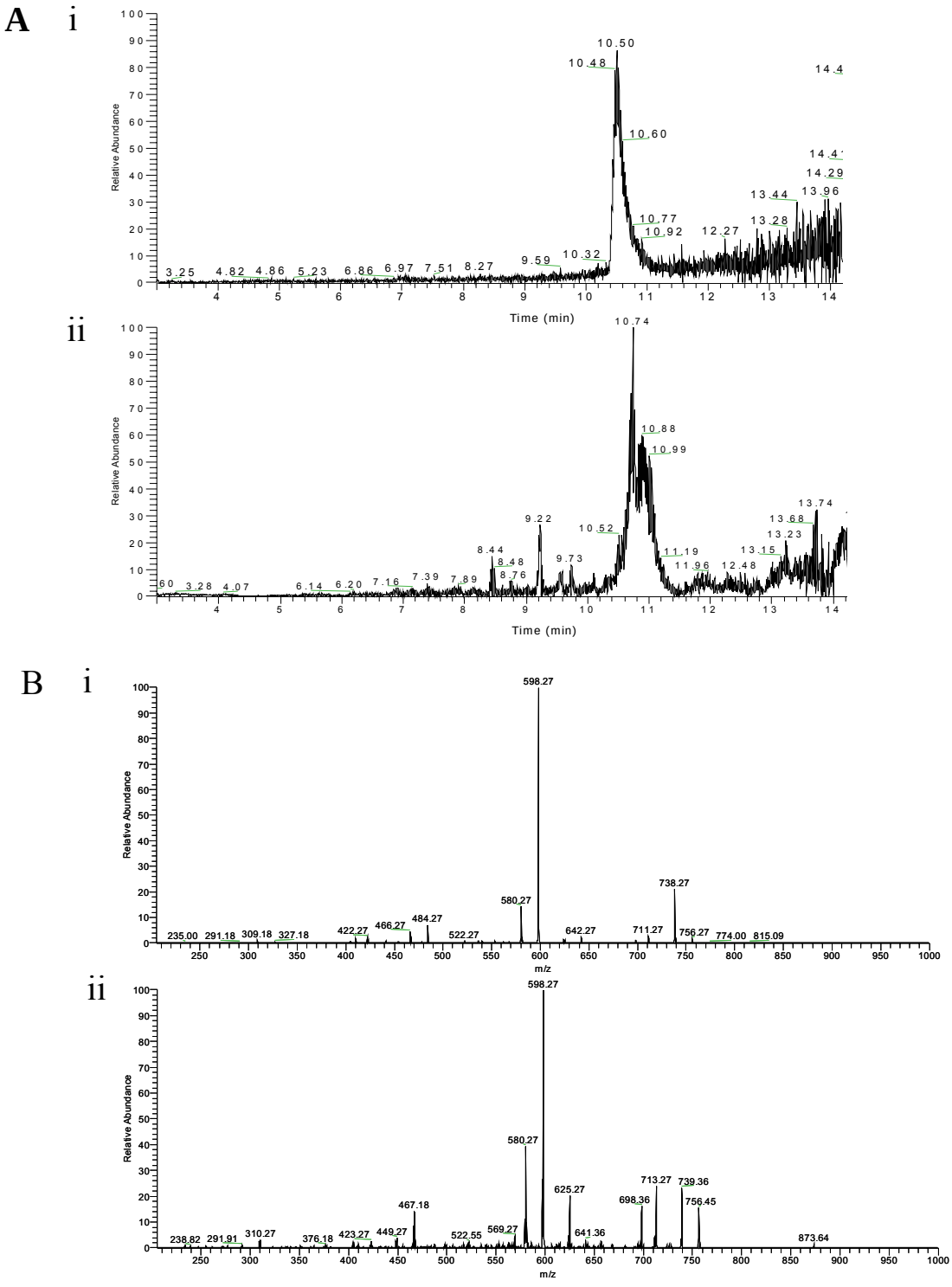


Figure 4.6 LC-MS analysis of erythromycin analog containing a *N*,*N*-dimethyl pyrrolosamine sugar group. (A) LC chromatogram analysis. (i) erythromycin A standard. (ii) *E. coli*-derived erythromycin A analog. (B) MS/MS analysis. (i) erythromycin A standard. (ii) *E. coli*-derived erythromycin A analog.

The *E. coli* strain BL21(DE3) harboring pHZT1, pHZT14, and pGro7 was grown at 22^{o}C and the fermentation product was then subjected to LC-MS analysis. It should be noted that, as the *eryF* gene was included in pHZT14, substrate 6dEB was directly fed to the *E. coli* culture for erythromycin analog production. As the pyrrolosamine *N,N*-dimethyl derivative has the same molecular weight with desosamine, the designed erythromycin analog containing pyrrolosamine *N,N*-dimethyl derivative sugar has the same molecular weight with native erythromycin. Therefore, native erythromycin was used as standard for LC-MS study.

Figure 4.6 showed that BL21(DE3, pHZT1/pHZT14/pGro7) extract produced a LC peak with the same retention time as that of erythromycin A standard. It should be noted that the retention time of the erythromycin A standard and the analog in the LC chromatogram here was different from the previous study, due to the change of the LC solvent elution flow rate. Further analysis showed that such a erythromycin analog had the same MS/MS spectrum with the erythromycin standard. In addition, the erythromycin D analog and the intermediate MEB were also identified. These results indicated that the designed erythromycin A analog could be produced in the constructed *E. coli* strain. Interestingly, erythromycin A analog production here did not rely on the use of the plasmid carrying an extra copy of *eryK* gene, as required for native erythromycin A biosynthesis in the previous section. The reason for this observation is unknown, although the genes contained in the constructed operon pHZT14 and pHZT2 are different. Further characterization of the produced erythromycin analog is needed to fully confirm the compound's biosynthesis in the future. Nevertheless, the data in this section

suggested that a erythromycin analog with a new sugar group *N*,*N*-dimethyl pyrrolosamine was produced in *E. coli* through a designed metabolic pathway. This was facilitated by the use of a native transferase pair EryCII and EryCIII, and it further demonstrated the substrate flexibility towards the sugar group to be attached, which has not been reported in an *E. coli* system before.

4.4 Discussion

Biosynthesis of erythromycin analogs was achieved in this section by modification of the polyketide core structure. One benzyl group was added to the C-13 position and one methyl group at the C-2 position was removed by rationally designing and altering the polyketide synthase functionality by genetic engineering. Although polyketide 6dEB analog biosynthesis has been reported before (*44*, *103*, *105*), it was achieved by use of *E. coli* as host in this study. Glycosylated erythromycin analogs were then produced by feeding the obtained polyketide analogs to an *E. coli* strain with the erythromycin glycosylation system. It was noted that erythromycin analog production demonstrated a level of flexibility associated with the glycosyltransferase enzymes responsible for attaching the deoxysugar residues to the modified polyketide structure. This flexibility has been observed with other systems and identified as a key step in realizing the full potential of altered structural design (*102*).

The relaxed substrate specificity of glycosyltransferase EryCII and CIII was also observed for the altered sugar moiety *N*,*N*-dimethyl pyrrolosamine, which was also found through an in vitro study (*110*). This study demonstrated that the reconstituted glycosyltransferase pair EryCII and EryCIII remained flexible in vivo using *E. coli* as a

host. As a result, a erythromycin analog with native desosamine replaced by the *N,N*-dimethyl pyrrolosamine was found produced using the constructed synthetic operons. Under the same experimental condition, however, the previously studied glycosyltransferase pair DesVII and DesVIII was not able to attach the sugar *N,N*-dimethyl pyrrolosamine to make the corresponding erythromycin analog. This could be because the enzymatic activity of DesVII and DesVIII pair requires a high pH value (*102*). Furthermore, sugar *N,N*-dimethyl pyrrolosamine's biosynthetic pathway was also studied in this chapter by its heterologous reconstitution in *E coli*. Identified genes *lom30*, *lom32*, *lom33*, and *lom36* from the natural product lomaiviticin gene cluster were used to make the pyrrolosamine *N,N*-dimethyl derivative sugar in *E. coli*. However, whether all these genes are required for *N,N*-dimethyl pyrrolosamine biosynthesis, or whether the proposed pathway is correct, needs future studies for confirmation. For example, *lom33 eryBII* genes both encode a deoxyhexose reductase, and *lom32* and *eryBVII* genes both encode a deoxyhexose epimerase. It is likely that *lom33* and *lom32* gene products are not active in the constructed *E. coli* stain, and the enzymatic reactions they are responsible for are catalyzed by EryBII and EryBVII for the completion of the proposed pathway. Nevertheless, our current results demonstrated that *E. coli* is a powerful host for new biosynthetic pathway discovery. Traditionally, one needs to knock-out and over-express hypothetic genes in a native natural product producer to investigate which genes are involved. However, when the native producer is hard to manipulate genetically and metabolically, the overall discovery process can be slow. Lomaiviticin producer *Salinispora pacifica* has this problem, which hampered the study of the lomaiviticin biosynthesis pathway, including that of *N,N*-dimethyl pyrrolosamine. Therefore, this

thesis study used an *E. coli* host to reconstitute the hypothetic pathway using a bottom-up approach. As a result, the genes for *N*,*N*-dimethyl pyrrolosamine were identified much more rapidly. Although further study needs to be performed to figure out the authentic pathway, our method showed the potential of the use of *E. coli* for natural product biosynthesis pathway reconstitution and discovery. As there is diminished crosstalk between *E. coli*'s native and heterologous pathways, decreasing the formation of undesired byproducts or shunt metabolites, the pathway discovery process is especially simplified in *E. coli*.

Success of heterologous erythromycin analog biosynthesis in *E. coli* supports the capabilities of the native glycosyltransferase enzymes to accommodate similar modifications to the polyketide and tailoring steps anticipated in analog biosynthetic approaches. To my knowledge, this is the first report about erythromycin analog production by use of *E. coli*. This accomplishment enables us to more easily manipulate or change erythromycin structure to discover unnatural products with potential novel activity, as the host *E. coli* provides a relatively clean metabolic background given the crosstalk between native and heterologous pathways is minimized. In addition, the modular, multi-catalytic nature of the DEBS polyketide synthase, in addition to the engineering opportunities within tailoring biosynthesis, promises a wide range of new compounds. The established *E. coli* platform equipped with the new polyketide formation system and the erythromycin native and novel sugar transferring system will hopefully be employed to speed up the progress for natural product heterologous biosynthesis.

Chapter 5 Summary and Future Directions

This thesis work achieved heterologous biosynthesis of erythromycin and its associated polyketide analogs. *E. coli*, as a surrogate host, was utilized for the production of a desired product. As was discussed previously, *E. coli* has the advantages of a well characterized genetic and metabolic setup and readily available engineering tools; fast growth kinetics and easy culturing techniques; and established process engineering and scale-up approaches. More importantly, there is diminished crosstalk between *E. coli*'s native metabolic pathways and heterologous pathways, reducing the undesired metabolic flux detracting from heterologous products. Therefore, *E. coli* was chosen as the production host for this thesis study, despite the issues of codon bias, foreign enzyme folding and activity, and limited supporting metabolic pathways for complex natural product biosynthesis.

Erythromycin was selected as our target product, as it is an important polyketide natural product and has been widely used for bacterial infections. The erythromycin native producer *Saccharopolyspora erythraea* has a fastidious growth requirement and it is relatively hard to manipulate erythromycin biosynthesis due to the lack of metabolic engineering tools available for this microbe. However, after decades of study and characterization, the erythromycin biosynthesis mechanism has been well-illuminated and its biosynthetic cluster in the chromosome of *S. erythraea* has been fully sequenced. The progress thus enabled the transferring of erythromycin biosynthesis machinery to a heterologous host. Meanwhile, the need for improving the erythromycin production level as well as generating novel erythromycin analogs calls for a surrogate host other than *Saccharopolyspora erythraea*.

Therefore, my thesis research focused on heterologous biosynthesis of erythromycin and associated polyketide analogs in *E. coli*. The aims included (1) improving the polyketide precursor production; (2) making the complete erythromycin A product; and (3) generating new erythromycin associated polyketide analogs. To achieve our goal, genetic, metabolic, and process engineering approaches were employed to address the challenges of cloning large gene clusters, coordinated gene expression, maintaining enzyme activity, and modification of *E. coli* native metabolism.

To improve the polyketide 6dEB production in *E. coli*, the copy number of *sfp*, *pcc*, and DEBS genes that are required for heterologous biosynthesis was varied. This was accomplished by cloning the corresponding genes into either plasmid vectors with high copy numbers or the *E. coli* chromosome with one single copy. Through careful analysis of the gene product level and associated 6dEB titer level, it was concluded that *sfp*, *pcc*, and DEBS genes' dosage had different levels of influence on 6dEB production. Between the studied genes, 6dEB heterologous production was found the most sensitive to DEBS genes, followed by the *pcc* genes and the *sfp* gene.

In an effort to further optimize 6dEB production, the *E. coli* native metabolic pathways were manipulated to coordinate with the heterologous pathway. Propionyl-CoA and (2S)-methylmalonyl-CoA are two required substrates for polyketide 6dEB biosynthesis. Three *E. coli* native genes, including *sbm* (a methylmalonyl-CoA mutase), *ygfG* (a methylmalonyl-CoA decarboxylase), and *ygfH* (a propionyl-CoA:succinate-CoA transferase) were identified to be involved in propionyl-CoA and (2S)-methylmalonyl-CoA metabolism in *E. coli*. These genes were then either over-expressed by use of high copy number plasmids, or deleted from the chromosome through DNA recombination.

By analysis of the 6dEB titer of the resulting *E. coli* strains, it was clearly demonstrated that the Sbm pathway did not have much interaction with the heterologous pathways. The enhancement of the YgfG pathway, however, was found to be disadvantageous, as it detracted the metabolic flux from 6dEB biosynthesis. The most important finding was the elimination of the YgfH pathway increased the 6dEB titer by 2 fold, due to the removal of undesired propionyl-CoA:succinate-CoA conversion. Such improvement was further magnified by use of a batch bioreactor, with the final production reaching more than 500 mg/L.

E. coli native metabolism was then further explored to support polyketide 6dEB heterologous biosynthesis. Interestingly, 6dEB could be produced in the absence of exogenous propionate or the foreign *pcc* genes for (2S)-methylmalonyl-CoA provision. It was then found that this unexpected 6dEB production could be associated with a unknown *E. coli* native metabolic pathway. This finding was of importance, as it showed that, although it is limited, *E. coli* native metabolism does have the potential to support complex polyketide production.

The next outcome of this study was the complete biosynthesis of erythromycin A in *E. coli*. It required the introduction of a > 55 kb gene cluster and the coordinated expression of 26 genes, 23 of which were foreign. Efforts were first made to isolate all the required heterologous genes and express them in *E. coli*. Aided by the advance of genetic and protein engineering, coordinated gene expression was achieved by use of commercially available plasmids. After combining all cloned genes into two synthetic operons, erythromycin heterologous biosynthesis was carried out in *E. coli* by feeding the polyketide precursor 6dEB into the culture. It was later found out that the production of

erythromycin B and D could be identified only at the presence of a chaperone vector carrying the *E. coli* GroEL/ES chaperone system. This chaperone system was believed to facilitate the correct folding of the introduced foreign enzymes and help reconstitute enzyme activity, and thus it was used for all the following studies. The lack of erythromycin A and C production indicated the lack of EryK activity. As such, one plasmid carrying an extra copy of *eryK* was constructed and introduced into the *E. coli* host. This strategy resulted in the successful biosynthesis of erythromycin A and C. In addition to LC-MS and MS/MS analysis, a *Bacillus subtilis* growth inhibition assay was performed to confirm the successful production of the erythromycin product. The reported erythromycin A titer was 10 mg/L, demonstrating the potential of a *E. coli*-derived erythromycin production process. Furthermore, a six-plasmid system was established in *E. coli* to allow for erythromycin biosynthesis directly from propionate. This achievement not only transferred the whole erythromycin biosynthetic gene cluster of 55 kb into *E. coli*, highlighting the power of the established heterologous biosynthesis system, but also simplified the culturing and downstream processes.

Lastly, novel erythromycin analogs were successfully produced by changing either the polyketide ring or sugar group structure through modification of biosynthetic pathways in *E. coli*. The first designed analog has the feature of a benzyl group attached to the C-13 position. To make this product, the loading module of the polyketide synthase DEBS1 was replaced with the loading module from the rifamycin polyketide synthase to utilize benzoate as the starter unit. The modified polyketide synthase was then introduced into *E. coli* to make the product. The second designed analog does not have a C-2 methyl group. This was achieved by changing the module 5 of DEBS to accept malonyl-CoA

instead of (2S)-methylmalonyl-CoA as the extender unit. Again, the desired product was produced by use of an *E. coli* strain harboring the altered DEBS genes. The designed polyketide precursor analogs were then glycosylated by the developed *E. coli* platform that attached the two erythromycin sugar groups onto the polyketide precursors to make the erythromycin analogs. The native glycosyltransferase pair EryCII and EryCIII was employed for this accomplishment. And its success clearly demonstrated the substrate flexibility of the glycosyltransferase pair, which could be more widely used to produce more erythromycin analogs.

The third designed erythromycin analog has a novel sugar group pyrrolosamine *N*,*N*-dimethyl derivative that is also one of the sugar components of the natural product lomaiviticin. First, a biosynthetic pathway for *N*,*N*-dimethyl pyrrolosamine was proposed based on putative functions of the hypothetic genes in the lomaiviticin biosynthesis gene cluster. This pathway was then reconstituted in *E. coli* by introduction of synthetic operons containing the proposed genes. Second, the native erythromycin glycosyltransferase pair EryCII and EryCIII was again utilized to attach *N*,*N*-dimethyl pyrrolosamine to a polyketide precursor *in vivo*. LC-MS analysis showed that the designed analog was produced in the developed *E. coli* system, although more characterization for the compound is needed. This result further highlighted the flexibility of the native glycosyltransferase pair EryCII and EryCIII in recognizing not only a polyketide precursor substrate but also the sugar substrate. Furthermore, this progress showed the great potential of adapting *E. coli* to make novel erythromycin analogs by changing the polyketide ring structure, the sugar group structure, or both. This

establishment of the technique would facilitate the effort in developing more unnatural products with improved antibiotic or other biological activities.

The findings of this thesis study present several future research opportunities. First of all, the production levels of erythromycin A, B, C and D can be further improved by use of processing engineering approaches that are readily available for *E. coli*, including medium optimization, process scale-up, culturing time optimization, etc. As had been observed previously for polyketide biosynthesis in *E. coli* (*54, 55, 111*), these efforts could help boost the production to a high level. Furthermore, the erythromycin biosynthesis system can be explored to increase the activities of some key enzymes for better conversion from propionate to erythromycin A. For example, the accumulation of intermediates, such as EB and MEB, were observed at the end of the production process, which indicated that production efficiency could be improved. To address this issue, the activity of glycosyltransferase pairs for the first sugar mycarose and the second sugar desosamine can be enhanced by introduction of more copies of these genes or better designed synthetic operons. Advances in protein engineering techniques would also help to improve these enzymes' activities and thus improve the production levels.

Second, the results in this study present engineering opportunities for making novel erythromycin analogs. More structural modifications of the polyketide substrate 6dEB can be achieved by genetic alteration of the DEBS genes, followed by glycosylation in *E. coli*. In addition to pyrrolosamine *N,N*-dimethyl derivative, several other sugar groups with similar molecular structures should be considered to be utilized for sugar group replacement. These include D-olivose, L-olivose, orleandrose, mycaminose, and rhamnose. As opposed to the pyrrolosamine *N,N*-dimethyl derivative,

the biosynthetic pathways for these sugars have been well-studied and characterized. Therefore, it may be relatively easier to clone and introduce the required biosynthetic gene materials to *E. coli* for analog production using the techniques established in this study. With the observed relaxed substrate specificity, it will be reasonable to further utilize the erythromycin native glycosyltransferases for sugar attachment and corresponding product biosynthesis. Analog quantification and bio-activity analysis can be used to carefully characterize the designed products, as well as assess the *E. coli* platform's validity for new complex unnatural product biosynthesis. All these efforts will contribute to the generation of new erythromycin analogs, and hopefully some of them will become options for fighting existing and future antibiotic resistance.

Reference

1. **Demain, A. L. (2006) From natural products discovery to commercialization: a success story, *J Ind Microbiol Biotechnol 33*, 486-495.**
2. **Gullo, V. P., and Hughes, D. E. (2005) Exploiting new approaches for natural product drug discovery in the biotechnology industry, *Drug Discovery Today: Technologies 2*, 281-286.**
3. **Paterson, I., and Anderson, E. A. (2005) The Renaissance of Natural Products as Drug Candidates, in *Science*, pp 451-453.**
4. **Newman, D. J., and Cragg, G. M. (2007) Natural products as sources of new drugs over the last 25 years, *J Nat Prod 70*, 461-477.**
5. **Clardy, J., and Walsh, C. (2004) Lessons from natural molecules, *Nature 432*, 829-837.**
6. **Cane, D. E., Walsh, C. T., and Khosla, C. (1998) Harnessing the biosynthetic code: combinations, permutations, and mutations, *Science 282*, 63-68.**
7. **Li, J. W. H., and Vederas, J. C. (2009) Drug Discovery and Natural Products: End of an Era or an Endless Frontier?, in *Science*, pp 161-165.**
8. **Scott, A. I. (1974) Biosynthesis of Natural Products, in *Science*, pp 760-764.**
9. **Staunton, J., and Weissman, K. J. (2001) Polyketide biosynthesis: a millennium review, *Nat Prod Rep 18*, 380-416.**
10. **Walsh, C. T. (2004) Polyketide and nonribosomal peptide antibiotics: modularity and versatility, *Science 303*, 1805-1810.**
11. **McGuire, J. M., Bunch, R.L., Anderson, R.C., Boaz, H.E., Flynn, E.H., Powell, H.M., Smith, J.W. (1952) "Ilotycin," a new antibiotic., *Antibiot & Chemother 2*, 281-284.**
12. **Heilman, F. R., Herrell, W. E., Wellman, W. E., and Geraci, J. E. (1952) Some laboratory and clinical observations on a new antibiotic, erythromycin (ilotycin), *Proceedings of the staff meetings. Mayo Clinic 27*, 285-304.**
13. **Organization, T. W. I. P. (2004) (Azythromycin: A world best-selling antibiotic – Pliva, *http://www.wipo.int/sme/en/case_studies/pliva.htm*.**
14. **Brock, T., and BROCK, M. (1959) Similarity in mode of action of chloramphenicol and erythromycin, *Biochimica et Biophysica Acta 33*, 274-275.**
15. **Taubman, S., Jones, N., Young, F., and Corcoran, J. (1966) Sensitivity and resistance to erythromycin in Bacillus subtilis 168: the ribosomal binding of erythromycin and chloramphenicol, *Biochimica et Biophysica Acta 123*, 438-440.**
16. **Lowbury, E., and Hurst, L. (1959) The sensitivity of staphylococci and other wound bacteria to erythromycin, oleandomycin and spiramycin, *Journal of Clinical Pathology 12*, 163-164.**
17. **Weisblum, B. (1995) Erythromycin resistance by ribosome modification, *Antimicrobial Agents and Chemotherapy 39*, 577-585.**
18. **Woodward, R. B., Logusch, E., Nambiar, K. P., Sakan, K., Ward, D. E., Auyeung, B. W., Balaram, P., Browne, L. J., Card, P. J., Chen, C. H.,**

Chenevert, R. B., Fliri, A., Frobel, K., Gais, H. J., Garratt, D. G., Hayakawa, K., Heggie, W., Hesson, D. P., Hoppe, D., Hoppe, I., Hyatt, J. A., Ikeda, D., Jacobi, P. A., Kim, K. S., Kobuke, Y., Kojima, K., Krowicki, K., Lee, V. J., Leutert, T., Malchenko, S., Martens, J., Matthews, R. S., Ong, B. S., Press, J. B., Rajanbabu, T. V., Rousseau, G., Sauter, H. M., Suzuki, M., Tatsuta, K., Tolbert, L. M., Truesdale, E. A., Uchida, I., Ueda, Y., Uyehara, T., Vasella, A. T., Vladuchick, W. C., Wade, P. A., Williams, R. M., and Wong, H. N. C. (1981) Asymmetric Total Synthesis of Erythromycin .1. Synthesis of an Erythronolide a Seco Acid-Derivative Via Asymmetric Induction, *Journal of the American Chemical Society 103*, 3210-3213.

19. Woodward, R. B., Logusch, E., Nambiar, K. P., Sakan, K., Ward, D. E., Auyeung, B. W., Balaram, P., Browne, L. J., Card, P. J., Chen, C. H., Chenevert, R. B., Fliri, A., Frobel, K., Gais, H. J., Garratt, D. G., Hayakawa, K., Heggie, W., Hesson, D. P., Hoppe, D., Hoppe, I., Hyatt, J. A., Ikeda, D., Jacobi, P. A., Kim, K. S., Kobuke, Y., Kojima, K., Krowicki, K., Lee, V. J., Leutert, T., Malchenko, S., Martens, J., Matthews, R. S., Ong, B. S., Press, J. B., Rajanbabu, T. V., Rousseau, G., Sauter, H. M., Suzuki, M., Tatsuta, K., Tolbert, L. M., Truesdale, E. A., Uchida, I., Ueda, Y., Uyehara, T., Vasella, A. T., Vladuchick, W. C., Wade, P. A., Williams, R. M., and Wong, H. N. C. (1981) Asymmetric Total Synthesis of Erythromycin .2. Synthesis of an Erythronolide a Lactone System, *Journal of the American Chemical Society 103*, 3213-3215.

20. Woodward, R. B., Logusch, E., Nambiar, K. P., Sakan, K., Ward, D. E., Auyeung, B. W., Balaram, P., Browne, L. J., Card, P. J., Chen, C. H., Chenevert, R. B., Fliri, A., Frobel, K., Gais, H. J., Garratt, D. G., Hayakawa, K., Heggie, W., Hesson, D. P., Hoppe, D., Hoppe, I., Hyatt, J. A., Ikeda, D., Jacobi, P. A., Kim, K. S., Kobuke, Y., Kojima, K., Krowicki, K., Lee, V. J., Leutert, T., Malchenko, S., Martens, J., Matthews, R. S., Ong, B. S., Press, J. B., Rajanbabu, T. V., Rousseau, G., Sauter, H. M., Suzuki, M., Tatsuta, K., Tolbert, L. M., Truesdale, E. A., Uchida, I., Ueda, Y., Uyehara, T., Vasella, A. T., Vladuchick, W. C., Wade, P. A., Williams, R. M., and Wong, H. N. C. (1981) Asymmetric Total Synthesis of Erythromycin .3. Total Synthesis of Erythromycin, *Journal of the American Chemical Society 103*, 3215-3217.

21. Cortes, J., Haydock, S. F., Roberts, G. A., Bevitt, D. J., and Leadlay, P. F. (1990) An unusually large multifunctional polypeptide in the erythromycin-producing polyketide synthase of Saccharopolyspora erythraea, *Nature 348*, 176-178.

22. Donadio, S., Staver, M. J., McAlpine, J. B., Swanson, S. J., and Katz, L. (1991) Modular organization of genes required for complex polyketide biosynthesis, *Science 252*, 675-679.

23. Summers, R. G., Donadio, S., Staver, M. J., Wendt-Pienkowski, E., Hutchinson, C. R., and Katz, L. (1997) Sequencing and mutagenesis of genes from the erythromycin biosynthetic gene cluster of Saccharopolyspora erythraea that are involved in L-mycarose and D-desosamine production, *Microbiology 143 (Pt 10)*, 3251-3262.

24. **Oliynyk, M., Samborskyy, M., Lester, J. B., Mironenko, T., Scott, N., Dickens, S., Haydock, S. F., and Leadlay, P. F. (2007) Complete genome sequence of the erythromycin-producing bacterium Saccharopolyspora erythraea NRRL23338,** ***Nat Biotechnol 25*****, 447-453.**

25. **Gokhale, R. S., Tsuji, S. Y., Cane, D. E., and Khosla, C. (1999) Dissecting and exploiting intermodular communication in polyketide synthases,** ***Science 284*****, 482-485.**

26. **Tang, Y., Kim, C.-Y., Mathews, I. I., Cane, D. E., and Khosla, C. (2006) The 2.7-Å crystal structure of a 194-kDa homodimeric fragment of the 6-deoxyerythronolide B synthase, in** ***The Proceedings of the National Academy of Sciences of the United States of America*****, pp 11124-11129.**

27. **Reeves, A. R., Seshadri, R., Brikun, I. A., Cernota, W. H., Gonzalez, M. C., and Weber, J. M. (2008) Knockout of the Erythromycin Biosynthetic Cluster Gene, eryBI, Blocks Isoflavone Glucoside Bioconversion during Erythromycin Fermentations in Aeromicrobium erythreum but Not in Saccharopolyspora erythraea, in** ***Applied and Environmental Microbiology*****, pp 7383-7390.**

28. **Lambalot, R. H., Cane, D. E., Aparicio, J. J., and Katz, L. (1995) Overproduction and characterization of the erythromycin C-12 hydroxylase, EryK,** ***Biochemistry 34*****, 1858-1866.**

29. **Paulus, T. J., Tuan, J. S., Luebke, V. E., Maine, G. T., DeWitt, J. P., and Katz, L. (1990) Mutation and cloning of eryG, the structural gene for erythromycin O-methyltransferase from Saccharopolyspora erythraea, and expression of eryG in Escherichia coli,** ***J Bacteriol 172*****, 2541-2546.**

30. **Minas, W., Brunker, P., Kallio, P. T., and Bailey, J. E. (1998) Improved erythromycin production in a genetically engineered industrial strain of Saccharopolyspora erythraea,** ***Biotechnol Prog 14*****, 561-566.**

31. **Wang, Y., Wang, Y., and Zhang, S. (2008) High frequency transformation of the industrial erythromycin-producing bacterium <i>Saccharopolyspora erythraea</i&gt,** ***Biotechnology Letters 30*****, 357-361.**

32. **Chng, C., Lum, A. M., Vroom, J. A., and Kao, C. M. (2008) A key developmental regulator controls the synthesis of the antibiotic erythromycin in Saccharopolyspora erythraea,** ***Proc Natl Acad Sci U S A 105*****, 11346-11351.**

33. **Reeves, A. R., Brikun, I. A., Cernota, W. H., Leach, B. I., Gonzalez, M. C., and Mark Weber, J. (2007) Engineering of the methylmalonyl-CoA metabolite node of Saccharopolyspora erythraea for increased erythromycin production,** ***Metabolic Engineering 9*****, 293-303.**

34. **Chen, Y., Deng, W., Wu, J., Qian, J., Chu, J., Zhuang, Y., Zhang, S., and Liu, W. (2008) Genetic Modulation of the Overexpression of Tailoring Genes eryK and eryG Leading to the Improvement of Erythromycin A Purity and Production in Fermentation of Saccharopolyspora erythraea, pp AEM.02770-02707.**

35. **Pfeifer, B. A., and Khosla, C. (2001) Biosynthesis of polyketides in heterologous hosts,** ***Microbiol Mol Biol Rev 65*****, 106-118.**

36. Zhang, H., Wang, Y., and Pfeifer, B. A. (2008) Bacterial hosts for natural product production, *Mol Pharm 5*, 212-225.

37. Zhang, H., Boghigian, B. A., Armando, J., and Pfeifer, B. A. (2010) Methods and options for the heterologous production of complex natural products, *Natural Product Reports*.

38. Xue, Q., Ashley, G., Hutchinson, C. R., and Santi, D. V. (1999) A multiplasmid approach to preparing large libraries of polyketides, *Proc Natl Acad Sci U S A 96*, 11740-11745.

39. Kao, C. M., Katz, L., and Khosla, C. (1994) Engineered biosynthesis of a complete macrolactone in a heterologous host, *Science 265*, 509-512.

40. Pfeifer, B. A., Admiraal, S. J., Gramajo, H., Cane, D. E., Khosla, C. (2001) Biosynthesis of complex polyketides in a metabolically engineered strain of E. coli, *Science 291*, 1790-1792.

41. Peiru, S., Menzella, H. G., Rodriguez, E., Carney, J., and Gramajo, H. (2005) Production of the potent antibacterial polyketide erythromycin C in Escherichia coli, *Appl Environ Microbiol 71*, 2539-2547.

42. Menzella, H. G., Reid, R., Carney, J. R., Chandran, S. S., Reisinger, S. J., Patel, K. G., Hopwood, D. A., and Santi, D. V. (2005) Combinatorial polyketide biosynthesis by de novo design and rearrangement of modular polyketide synthase genes, *Nat Biotechnol 23*, 1171-1176.

43. Lee, H. Y., and Khosla, C. (2007) Bioassay-guided evolution of glycosylated macrolide antibiotics in Escherichia coli, *PLoS Biol 5*, e45.

44. McDaniel, R., Thamchaipenet, A., Gustafsson, C., Fu, H., Betlach, M., and Ashley, G. (1999) Multiple genetic modifications of the erythromycin polyketide synthase to produce a library of novel "unnatural" natural products, *Proc Natl Acad Sci U S A 96*, 1846-1851.

45. Weissman, K. J., and Leadlay, P. F. (2005) Combinatorial biosynthesis of reduced polyketides, *Nat Rev Microbiol 3*, 925-936.

46. Nguyen, K. T., Ritz, D., Gu, J.-Q., Alexander, D., Chu, M., Miao, V., Brian, P., and Baltz, R. H. (2006) Combinatorial biosynthesis of novel antibiotics related to daptomycin, in *Proceedings of the National Academy of Sciences of the United States of America*, pp 17462-17467.

47. Menzella, H. G., and Reeves, C. D. (2007) Combinatorial biosynthesis for drug development, *Current Opinion in Microbiology 10*, 238-245.

48. McDaniel, R., Ebert-Khosla, S., Hopwood, D. A., and Khosla, C. (1995) Rational design of aromatic polyketide natural products by recombinant assembly of enzymatic subunits, *Nature 375*, 549-554.

49. Luo, G., Pieper, R., Rosa, A., Khosla, C., and Cane, D. E. (1996) Erythromycin biosynthesis: exploiting the catalytic versatility of the modular polyketide synthase, *Bioorg Med Chem 4*, 995-999.

50. Denis, A., Agouridas, C., Auger, J. M., Benedetti, Y., Bonnefoy, A., Bretin, F., Chantot, J. F., Dussarat, A., Fromentin, C., D'Ambrieres, S. G., Lachaud, S., Laurin, P., Le Martret, O., Loyau, V., Tessot, N., Pejac, J. M., and Perron, S. (1999) Synthesis and antibacterial activity of HMR 3647 a new ketolide highly potent against erythromycin-resistant and susceptible pathogens, *Bioorg Med Chem Lett 9*, 3075-3080.

51. **Ma, Z., Clark, R. F., Brazzale, A., Wang, S., Rupp, M. J., Li, L., Griesgraber, G., Zhang, S., Yong, H., Phan, L. T., Nemoto, P. A., Chu, D. T., Plattner, J. J., Zhang, X., Zhong, P., Cao, Z., Nilius, A. M., Shortridge, V. D., Flamm, R., Mitten, M., Meulbroek, J., Ewing, P., Alder, J., and Or, Y. S. (2001) Novel erythromycin derivatives with aryl groups tethered to the C-6 position are potent protein synthesis inhibitors and active against multidrug-resistant respiratory pathogens,** ***J Med Chem 44*****, 4137-4156.**

52. **Fischbach, M. A., and Walsh, C. T. (2009) Antibiotics for emerging pathogens,** ***Science 325*****, 1089-1093.**

53. **Walsh, C. (2003) Where will new antibiotics come from?,** ***Nat Rev Microbiol 1*****, 65-70.**

54. **Pfeifer, B., Hu, Z., Licari, P., and Khosla, C. (2002) Process and metabolic strategies for improved production of Escherichia coli-derived 6-deoxyerythronolide B,** ***Appl Environ Microbiol 68*****, 3287-3292.**

55. **Lau, J., Tran, C., Licari, P., Galazzo, J. (2004) Development of a high cell-density fed-batch bioprocess for the heterologous production of 6-deoxyerythronolide B in Escherichia coli,** ***J Biotechnol 110*****, 95-103.**

56. **Hamilton, C. M., Aldea, M., Washburn, B. K., Babitzke, P., Kushner, S. R. (1989) New method for generating deletions and gene replacements in Escherichia coli,** ***J Bacteriol 171*****, 4617-4622.**

57. **Yu, D., Ellis, H. M., Lee, E. C., Jenkins, N. A., Copeland, N. G., and Court, D. L. (2000) An efficient recombination system for chromosome engineering in Escherichia coli,** ***Proc Natl Acad Sci U S A 97*****, 5978-5983.**

58. **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.**

59. **Wang, Y., and Pfeifer, B. A. (2008) 6-Deoxyerythronolide B production through chromosomal localization of the deoxyerythronolide B synthase genes in E. coli,** ***Metab Eng 10*****, 33-38.**

60. **Sambrook, J., Fritsch, E.F., Maniatis, T. (1989)** ***Molecular cloning : a laboratory manual*****, Vol. 3 2nd ed., Cold Spring Harbor Laboratory Press., Cold Spring Harbor, N.Y.**

61. **Wang, Y., Boghigian, B. A., and Pfeifer, B. A. (2007) Improving heterologous polyketide production in Escherichia coli by overexpression of an S-adenosylmethionine synthetase gene,** ***Appl Microbiol Biotechnol 77*****, 367-373.**

62. **Lee, K. K., Da Silva, N. A., and Kealey, J. T. (2009) Determination of the extent of phosphopantetheinylation of polyketide synthases expressed in Escherichia coli and Saccharomyces cerevisiae,** ***Anal Biochem 394*****, 75-80.**

63. **Menzella, H. G., Reisinger, S. J., Welch, M., Kealey, J. T., Kennedy, J., Reid, R., Tran, C. Q., and Santi, D. V. (2006) Redesign, synthesis and functional expression of the 6-deoxyerythronolide B polyketide synthase gene cluster,** ***J Ind Microbiol Biotechnol 33*****, 22-28.**

64. **Murli, S., Kennedy, J., Dayem, L. C., Carney, J. R., and Kealey, J. T. (2003) Metabolic engineering of Escherichia coli for improved 6-deoxyerythronolide B production,** ***J Ind Microbiol Biotechnol 30*****, 500-509.**

65. Strohl, W. R., Bartel, P. L., Li, Y., Connors, N. C., and Woodman, R. H. (1991) Expression of polyketide biosynthesis and regulatory genes in heterologous streptomycetes *Journal of Industrial Microbiology & Biotechnology 7*, 163-174.

66. Wenzel, S. C., and Muller, R. (2005) Recent developments towards the heterologous expression of complex bacterial natural product biosynthetic pathways, *Curr Opin Biotechnol 16*, 594-606.

67. Zhang, H., Wang, Y, Boghigian, B, Pfeifer BA. (2009) Probing the Heterologous Metabolism Supporting 6-deoxyerythronolide B Biosynthesis in E. coli, *Microbial Biotechnology 2*, 390-394.

68. Haller, T., Buckel, T., Retey, J., Gerlt, J. A. (2000) Discovering new enzymes and metabolic pathways: conversion of succinate to propionate by Escherichia coli, *Biochemistry 39*, 4622-4629.

69. Froese, D. S., Dobson, C. M., White, A. P., Wu, X., Padovani, D., Banerjee, R., Haller, T., Gerlt, J. A., Surette, M. G., and Gravel, R. A. (2009) Sleeping beauty mutase (sbm) is expressed and interacts with ygfd in Escherichia coli, *Microbiological Research 164*, 1-8.

70. Dayem, L. C., Carney, J. R., Santi, D. V., Pfeifer, B. A., Khosla, C., Kealey, J. T. (2002) Metabolic engineering of a methylmalonyl-CoA mutase-epimerase pathway for complex polyketide biosynthesis in Escherichia coli, *Biochemistry 41*, 5193-5201.

71. Marsden, A., Caffrey, P., Aparicio, J., Loughran, M., Staunton, J., and Leadlay, P. (1994) Stereospecific acyl transfers on the erythromycin-producing polyketide synthase *Science 263*, 378 - 380.

72. Studier, F. W., and Moffatt, B. A. (1986) Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes, *J Mol Biol 189*, 113-130.

73. Baba, T., Ara, T., Hasegawa, M., Takai, Y., Okumura, Y., Baba, M., Datsenko, K. A., Tomita, M., Wanner, B. L., and Mori, H. (2006) Construction of Escherichia coli K-12 in-frame, single-gene knockout mutants: the Keio collection, *Mol Syst Biol 2*.

74. Kellermeyer, R. W., Allen, S. H. G., Stjernholm, R., and Wood, H. G. (1964) Methylmalonyl Isomerase, *Journal of Biological Chemistry 239*, 2562-2569.

75. Tsuruta, H., Paddon, C. J., Eng, D., Lenihan, J. R., Horning, T., Anthony, L. C., Regentin, R., Keasling, J. D., Renninger, N. S., and Newman, J. D. (2009) High-level production of amorpha-4,11-diene, a precursor of the antimalarial agent artemisinin, in Escherichia coli, *PLoS One 4*, e4489.

76. Aldor, I. S., Kim, S.-W., Prather, K. L. J., and Keasling, J. D. (2002) Metabolic Engineering of a Novel Propionate-Independent Pathway for the Production of Poly(3-Hydroxybutyrate-co-3-Hydroxyvalerate) in Recombinant Salmonella enterica Serovar Typhimurium, pp 3848-3854.

77. Jacobsen, J. R., Hutchinson, C. R., Cane, D. E., and Khosla, C. (1997) Precursor-directed biosynthesis of erythromycin analogs by an engineered polyketide synthase, *Science 277*, 367-369.

78. Carreras, C., Frykman, S., Ou, S., Cadapan, L., Zavala, S., Woo, E., Leaf, T., Carney, J., Burlingame, M., Patel, S., Ashley, G., and Licari, P. (2002)

Saccharopolyspora erythraea-catalyzed bioconversion of 6-deoxyerythronolide B analogs for production of novel erythromycins, *J Biotechnol 92*, 217-228.

79. Liu, H., Jiang, H., Haltli, B., Kulowski, K., Muszynska, E., Feng, X., Summers, M., Young, M., Graziani, E., Koehn, F., Carter, G. T., and He, M. (2009) Rapid cloning and heterologous expression of the meridamycin biosynthetic gene cluster using a versatile Escherichia coli-streptomyces artificial chromosome vector, pSBAC, *J Nat Prod 72*, 389-395.

80. Zhang, H., Boghigian, B. A., and Pfeifer, B. A. (2010) Investigating the role of native propionyl-CoA and methylmalonyl-CoA metabolism on heterologous polyketide production in Escherichia coli, *Biotechnol Bioeng 105*, 567-573.

81. Gouy, M., and Gautier, C. (1982) Codon usage in bacteria: correlation with gene expressivity, *Nucleic Acids Res 10*, 7055-7074.

82. Pfeifer, B. A., Wang, C. C., Walsh, C. T., and Khosla, C. (2003) Biosynthesis of Yersiniabactin, a complex polyketide-nonribosomal peptide, using Escherichia coli as a heterologous host, *Appl Environ Microbiol 69*, 6698-6702.

83. Watanabe, K., Hotta, K., Praseuth, A. P., Koketsu, K., Migita, A., Boddy, C. N., Wang, C. C., Oguri, H., and Oikawa, H. (2006) Total biosynthesis of antitumor nonribosomal peptides in Escherichia coli, *Nat Chem Biol*.

84. Gaisser, S., Bohm, G. A., Doumith, M., Raynal, M. C., Dhillon, N., Cortes, J., and Leadlay, P. F. (1998) Analysis of eryBI, eryBIII and eryBVII from the erythromycin biosynthetic gene cluster in Saccharopolyspora erythraea, *Mol Gen Genet 258*, 78-88.

85. Mutka, S. C., Carney, J. R., Liu, Y., and Kennedy, J. (2006) Heterologous Production of Epothilone C and D in Escherichia coli, *Biochemistry 45*, 1321-1330.

86. Betancor, L., Fernandez, M. J., Weissman, K. J., and Leadlay, P. F. (2008) Improved catalytic activity of a purified multienzyme from a modular polyketide synthase after coexpression with Streptomyces chaperonins in Escherichia coli, *Chembiochem 9*, 2962-2966.

87. Meyerowitz, E. M., Guild, G. M., Prestidge, L. S., and Hogness, D. S. (1980) A new high-capacity cosmid vector and its use, *Gene 11*, 271-282.

88. Kim, U. J., Shizuya, H., de Jong, P. J., Birren, B., and Simon, M. I. (1992) Stable propagation of cosmid sized human DNA inserts in an F factor based vector, *Nucleic Acids Res 20*, 1083-1085.

89. Tao, Q., and Zhang, H. B. (1998) Cloning and stable maintenance of DNA fragments over 300 kb in Escherichia coli with conventional plasmid-based vectors, *Nucleic Acids Res 26*, 4901-4909.

90. Pierce, J. C., Sauer, B., and Sternberg, N. (1992) A positive selection vector for cloning high molecular weight DNA by the bacteriophage P1 system: improved cloning efficacy, *Proc Natl Acad Sci U S A 89*, 2056-2060.

91. Shizuya, H., Birren, B., Kim, U. J., Mancino, V., Slepak, T., Tachiiri, Y., and Simon, M. (1992) Cloning and stable maintenance of 300-kilobase-pair fragments of human DNA in Escherichia coli using an F-factor-based vector, *Proc Natl Acad Sci U S A 89*, 8794-8797.

92. **Burke, D. T., Carle, G. F., and Olson, M. V. (1987) Cloning of large segments of exogenous DNA into yeast by means of artificial chromosome vectors, *Science 236*, 806-812.**
93. **Monaco, A. P., and Larin, Z. (1994) YACs, BACs, PACs and MACs: artificial chromosomes as research tools, *Trends Biotechnol 12*, 280-286.**
94. **Yomano, L. P., York, S. W., and Ingram, L. O. (1998) Isolation and characterization of ethanol-tolerant mutants of <i>Escherichia coli</i> KO11 for fuel ethanol production, *Journal of Industrial Microbiology & Biotechnology 20*, 132-138.**
95. **Pieper, R., Ebert-Khosla, S., Cane, D., and Khosla, C. (1996) Erythromycin biosynthesis: kinetic studies on a fully active modular polyketide synthase using natural and unnatural substrates, *Biochemistry 35*, 2054-2060.**
96. **Baerga-Ortiz, A., Popovic, B., Siskos, A. P., O'Hare, H. M., Spiteller, D., Williams, M. G., Campillo, N., Spencer, J. B., and Leadlay, P. F. (2006) Directed Mutagenesis Alters the Stereochemistry of Catalysis by Isolated Ketoreductase Domains from the Erythromycin Polyketide Synthase, *Chemistry & biology 13*, 277-285.**
97. **Tang, L., and McDaniel, R. (2001) Construction of desosamine containing polyketide libraries using a glycosyltransferase with broad substrate specificity, *Chemistry & Biology 8*, 547-555.**
98. **Kennedy, J., Murli, S., and Kealey, J. T. (2003) 6-Deoxyerythronolide B Analogue Production in Escherichia coli through Metabolic Pathway Engineering, *Biochemistry 42*, 14342–14348.**
99. **Schroeder DR, C. K., Klohr SE, Lee MS, Matson JA, Brinen LS, Clardy J. (1996) Pyrrolosporin A, a new antitumor antibiotic from Micromonospora sp. C39217-R109-7. II. Isolation, physico-chemical properties, spectroscopic study and X-ray analysis, *The Journal of Antibiotics 49*, 865-872.**
100. **He, H., Ding, W. D., Bernan, V. S., Richardson, A. D., Ireland, C. M., Greenstein, M., Ellestad, G. A., and Carter, G. T. (2001) Lomaiviticins A and B, potent antitumor antibiotics from Micromonospora lomaivitiensis, *J Am Chem Soc 123*, 5362-5363.**
101. **Borisova, S. A., Zhao, L., Melancon, I. C., Kao, C. L., and Liu, H. W. (2004) Characterization of the glycosyltransferase activity of desVII: analysis of and implications for the biosynthesis of macrolide antibiotics, *J Am Chem Soc 126*, 6534-6535.**
102. **Borisova, S. A., Zhang, C., Takahashi, H., Zhang, H., Wong, A. W., Thorson, J. S., and Liu, H. W. (2006) Substrate specificity of the macrolide-glycosylating enzyme pair DesVII/DesVIII: opportunities, limitations, and mechanistic hypotheses, *Angew Chem Int Ed Engl 45*, 2748-2753.**
103. **Liu, L., Thamchaipenet, A., Fu, H., Betlach, M., and Ashley G. (1997) Biosynthesis of 2-Nor-6-deoxyerythronolide B by Rationally Designed Domain Substitution, *J. Am. Chem. Soc. 119*, 10553-10554.**
104. **Pfeifer, B. (2005) Metabolic engineering for complex natural product biosynthesis utilizing Escherichia coli., *Stanford University PhD thesis*.**

105. Admiraal, S. J., Walsh, C. T., and Khosla, C. (2001) The loading module of rifamycin synthetase is an adenylation-thiolation didomain with substrate tolerance for substituted benzoates, *Biochemistry 40*, 6116-6123.

106. Weissman, K. J., Bycroft, M., Staunton, J., and Leadlay, P. F. (1998) Origin of starter units for erythromycin biosynthesis, *Biochemistry 37*, 11012-11017.

107. Pereda, A., Summers, R. G., Stassi, D. L., Ruan, X., and Katz, L. (1998) The loading domain of the erythromycin polyketide synthase is not essential for erythromycin biosynthesis in Saccharopolyspora erythraea, *Microbiology 144 (Pt 2)*, 543-553.

108. Long, P. F., Wilkinson, C. J., Bisang, C. P., Cortes, J., Dunster, N., Oliynyk, M., McCormick, E., McArthur, H., Mendez, C., Salas, J. A., Staunton, J., and Leadlay, P. F. (2002) Engineering specificity of starter unit selection by the erythromycin-producing polyketide synthase, *Mol Microbiol 43*, 1215-1225.

109. Xue, Y., Zhao, L., Liu, H. W., and Sherman, D. H. (1998) A gene cluster for macrolide antibiotic biosynthesis in Streptomyces venezuelae: architecture of metabolic diversity, *Proc Natl Acad Sci U S A 95*, 12111-12116.

110. Lee, H. Y., Chung, H. S., Hang, C., Khosla, C., Walsh, C. T., Kahne, D., and Walker, S. (2004) Reconstitution and characterization of a new desosaminyl transferase, EryCIII, from the erythromycin biosynthetic pathway, *Journal of the American Chemical Society 126*, 9924-9925.

111. Pistorino, M. (2008) Efficient Experimental Design and Micro-scale Medium Enhancement of 6-deoxyerythronolide B Production through *Escherichia coli*, *Tufts University Master Thesis*.