Developing Heterologous Expression Platforms for the Production of Polyketides from Microbial Hosts

David Cole Stevens

Thesis submitted to the

Faculty of Graduate and Postdoctoral Studies

In partial fulfillment of the requirements

For the PhD degree in Chemistry

Chemistry Department

Faculty of Science

University of Ottawa

Table of Contents

List of Figures

List of Tables

List of Abbreviations

6-dEB, 6-deoxyerythronolide B

ACP, Acyl Carrier Protein

act, Actinorhodin

AHBA, 3-amino-5-hydroxybenzoic acid

ARO, Aromatase

AT, Acyltransferase

BAC, Bacterial Artificial Chromosome

cDNA, Complementary Deoxyribonucleic Acid

CLF, Chain Length Factor

CoA, Coenzyme A

COG, Clusters of Orthologous Groups

CTTYE, Casitone Yeast Extract Media

CYC, Cyclase

DH, Dehydratase

DMAC, 3,8-dihydroxy-1-methyl anthraquinone-2-carboxylic acid

DNA, Deoxyribonucleic Acid

eDNA, Environmental Deoxyribonucleic Acid

ER, Enoylreductase

ESI, Electrospray Ionization

fren, Frenolicin

gDNA, Genomic Deoxyribonucleic Acid

IPTG, isopropyl β-D-1-thiogalactopyranoside

KR, Ketoreductase

KS, Ketosynthase

LB, Luria Broth

LC-MS, Liquid Chromotography-Mass Spectroscopy

mDNA, Metagenomic Deoxyribonucleic Acid

MRM, Multiple Reaction Monitoring

mRNA, Messenger Ribonucleic Acid

NRPS, Nonribosomal Peptide Synthase

O.D., Optical Density

oxy, Oxytetracycline

PCR, Polymerase Chain Reaction

PKS, Polyketide Synthase

(p)ppGpp, Guanosine Pentaphosphate

qPCR, Quantitative Polymerase Chain Reaction

RNA, Ribonucleic Acid

RNAP, RNA Polymerase

tcm, Tetracenomycin

TE, Thioesterase

tRNA, Transfer Ribonucleic Acid

TPM, Tris Phosphate Magnesium Media

WT, Wild-type

ABSTRACT

Bacterial polyketides possess an enormous range of chemical diversity and biological function. Many polyketides such as tetracycline, epothilone, and rapamycin have been developed into key clinical pharmaceuticals in a broad range of therapeutic areas. Sequencing of bacterial genomes has shown that there are many more polyketide biosynthetic pathways than there are polyketides isolated from standard cultivation techniques. These genetically encoded polyketide natural products from cultivatable and uncultivatable bacteria represent one of the greatest remaining untapped reservoirs of new natural product diversity. To access this untapped diversity of polyketide products, a general method for heterologous expression of these pathways is needed. Heterologous expression has proven to be a valuable asset in the discovery, production, engineering, and characterization of bacterial secondary metabolites and the complex enzymology involved in their biosynthesis. Herein we discuss the development and investigation of two unique heterologous expression platforms utilizing host strains of *Myxococcus xanthus* and *Escherichia coli*. Using our developed heterologous hosts, we were able to produce the *Streptomyces rimosus* polyketide oxytetracycline. Through production of oxytetracycline in *E .coli* we have identified the potential of alternative transcription factors as regulators of secondary metabolism. Further investigation and development of alternative transcription factors as regulators of secondary metabolism in heterologous hosts could benefit the development of robust general methodology for the heterologous expression of polyketides.

Chapter 1: Heterologous expression of polyketides in bacterial hosts

1.1 Introduction

Heterologous expression has proven to be a valuable asset in the discovery, production, engineering, and characterization of bacterial secondary metabolites and the complex enzymology involved in their biosynthesis.[1-4] Successful heterologous production of a secondary metabolite requires insertion and transcription of large multi-enzyme segments of DNA, translation and post-translational modification of the subsequent enzymes, and in many cases, the production of specific substrates required for the biosynthesis of the natural product.[3-5] These stringent requirements limit both the strains capable of being used as heterologous hosts and the natural products that can be heterologously produced to a relatively small subset.

From early experiments in the 1980s to modern day, heterologous expression has been utilized to improve several areas of research.[3,4] The investigation and manipulation of biosynthetic pathways in heterologous hosts has drastically increased our understanding of the mechanism of the enzymes involved.[4] Heterologous expression has also become an asset to combinatorial biosynthesis as the engineering of biosynthetic pathways in the chromosome of the native producers is very challenging.[4,6-8] Medicinal value intrinsic to natural products has driven the desire to development heterologous production platforms capable of producing larger quantities, decreasing production costs, and lowering production times when compared to the native metabolite producer.[4,9] The heterologous expression of biosynthetic pathways discovered through sequencing but producing no known metabolite, known as cryptic or silent pathways, has also become a

desirable area of research.[4,6] The development of a culture independent method for the discovery of secondary metabolites from metagenomic or environmental DNA (eDNA) libraries has become an attractive application of heterologous expression.[10] Though varied success in each of these areas of research has been observed heterologous expression techniques have changed little since their inception. Heterologous production of polyketides represents one of first and remains one of the most successful areas of secondary metabolite production.[1,2]

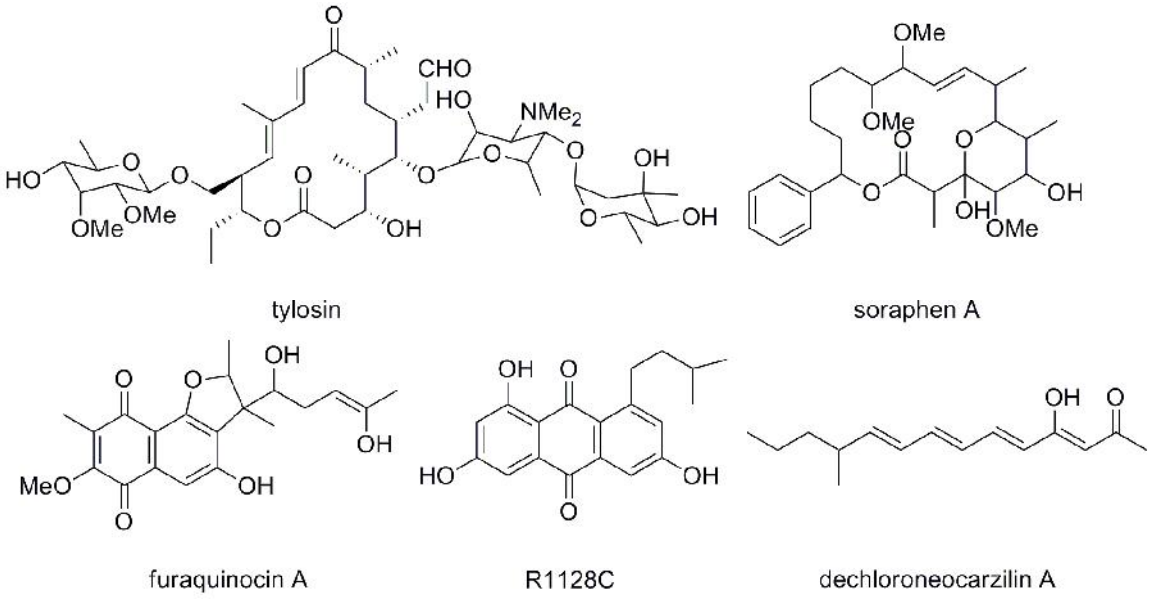


Figure 1.1.1: Examples of polyketides produced by heterologous hosts

Polyketides are a major class of secondary metabolites possessing an expansive range of chemical diversity and biological functions (Fig. 1.1.1). The requirement for bioactivity dictated by nature has lead many polyketides or hybrid polyketides to be developed into clinical drugs such as tetracycline, epothilone, rapamycin, erythromycin, FK506, doxorubicin, and bleomycin.[1,11,12] These complex molecules are produced from polyketides synthases (PKSs). PKSs are a class of proteins that construct broad ranges of

macrolactone or carboxylic acid scaffolds from simple substrates from primary metabolism such as malonyl-CoA and methylmalonyl-CoA. The malonyl- or methylmalonyl-building blocks are tethered as thioesters onto phosphopantetheinylated acyl carrier domains (ACPs). The CoA units attached to specific ACPs are dictated by acyl-transferase (AT) domains. Elongation of the tethered thioester through decarboxylative Claisen condenstation is catalyzed by a ketosynthase (KS) domain.[11] The ACP, AT, and KS domains represent the minimal PKS units necessary to produce an extending thioester. Intermediate catalytic domains, ketoreductase (KR), dehydratase (DH), and enoyl reductase (ER) domains, may be introduced between each module prior to the next elongation dictating the oxidation state of the newly added ketide. The length of the thioester is dictated through a thioesterase (TE) necessary for chain termination. The excised macrolactone or carboxylic acid may then be subjected to post-tailoring enzymes including: cyclases (CYC), aromatases (ARO), oxidations, reductions, acylations, and glycosylations. These post-tailoring enzymes afford a dramatic increase in the chemical diversity afforded from PKS biosynthesis (Fig 1.1.2).[11]

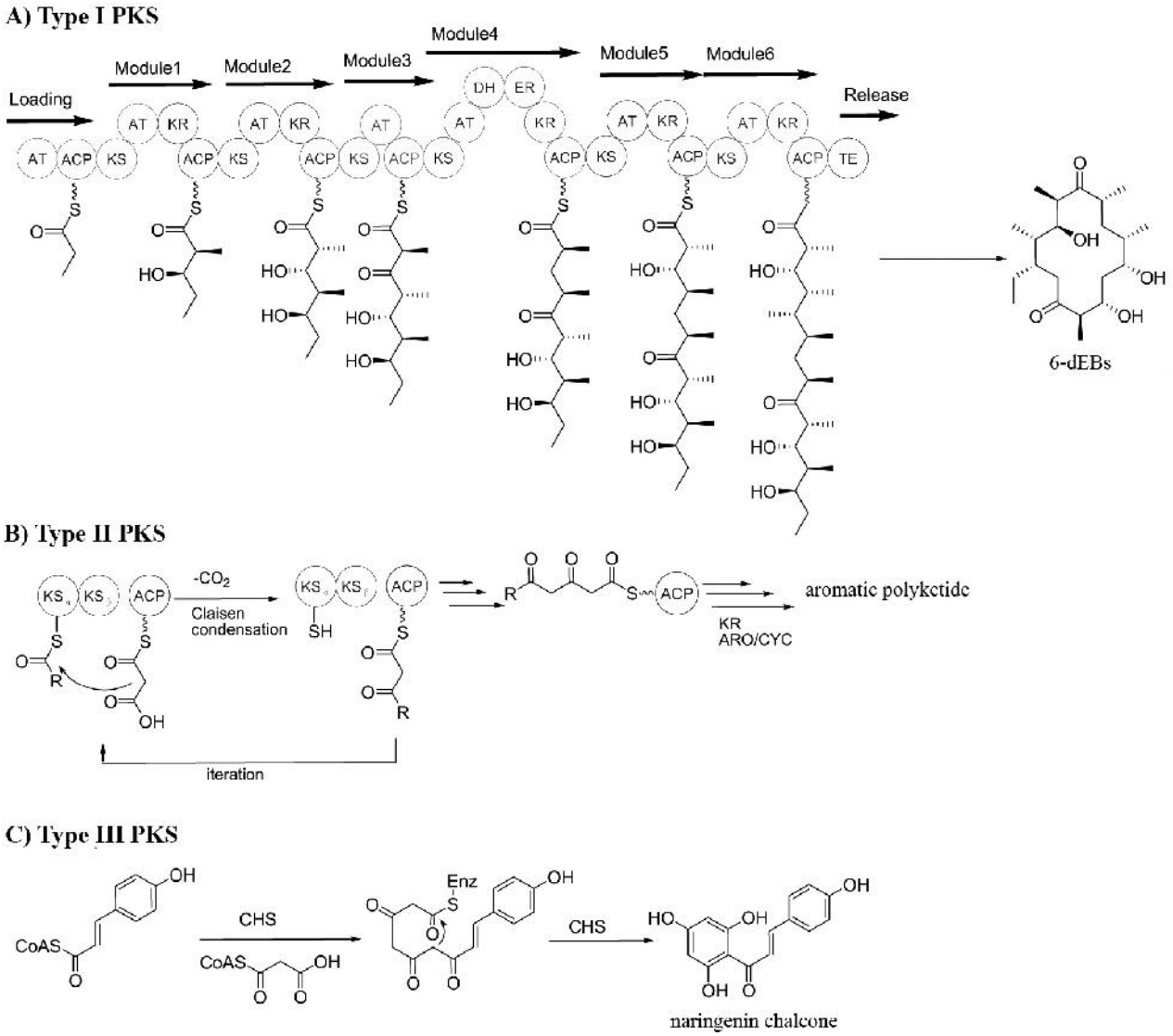


Figure 1.1.2: Types of PKS biosynthetic pathways A) Type I PKS pathway from the erythromycin biosynthetic pathway. B) Typical type II PKS biosynthetic pathway. C) Type III PKS pathway from the chalcone biosynthetic pathway.

The conserved characteristics across the thousands of known PKS biosynthetic pathways make these pathways excellent candidates for heterologous expression. The primary advantage that PKS pathways possess is that most discovered bacterial and fungal minimal PKS to date has been observed to exist as a cluster.[4] The majority of clusters contain not only the required PKS assembly enzymes but often when necessary

the resistance enzymes specific to the produced metabolite.[3,4] This clustering alleviates much of the difficulty in isolating the entire biosynthetic pathway from genomic DNA. Another attribute contributing to successful heterologous production of polyketides is that PKS proteins are functional in the cytosol and have activity independent from intracellular substructure.[4] This allows successfully expressed enzymes to be functional in non-native cellular environments. The limited variety in acyl-CoA substrates necessary for polyketide production and their wide availability in the native metabolism of most bacterium also benefit the heterologous production of polyketides. Even with these advantages in mind, successful production of a polyketide from a heterologous host remains challenging.

The commonly held consensus is that for successful heterologous expression of polyketides, the heterologous host needs to have a similar codon usage bias, to properly fold proteins, to be capable of post-translational modifications, and to possess precursors from primary metabolism.[3,4] Many of these requirements have been investigated and viable solutions have been provided. The production of 6-deoxyerythronilide (6-dEB), an intermediate from erythromycin biosynthesis native to *Sachropolyspora erythryae,* in *Escherichia coli*provides an excellent example of an engineered heterologous host meeting each of these requirements.[13-18] The requirement of post-translational activated ACP proteins was addressed through the development of an *E. coli* strain (BAP1) capable of post-translationally modifying ACP domains by inserting the phosphopantetheinyl transferase *sfp* from *Bacillus subtilis* which lead to the successful production of activated holoACP domains necessary to produce 6-dEBs.[13,19,20]Although *E. coli* and *S.erythryae* both have distinct differences in codon usage bias with the dEBs minimal PKS having

73% GC-content and 17% of the codons being rare in *E. coli*, through the use of lowered temperature and alternate promoters Pfeifer *et al* were able to produce 6-dEB from their engineered strain of *E. coli*.[13] The unavailability of necessary precursor pools has been addressed by several labs through either the insertion of precursor biosynthesis enzymes or the removal of precursor degradation pathways.[13,21-29] Pfeifer *et al* ensure the necessary amount of propionyl-CoA in their experiments by removing the propionate catabolism pathway from their strain of *E. coli*, feeding propionate, and expressing the propionyl-CoA carboxylase genes, *pccA* and *pccB*, necessary to produce methyl-malonyl-CoA.[13] Though solutions have been engineered for many of these requirements, the requirement for a similar codon usage bias to ensure transcription does not have an immediate solution. As in the production of 6-dEB in *E. coli* codon usage bias has not prevented the heterologous production of polyketides in strains possessing significant differences in codon usage bias.[13] Although the general consensus holds that the heterologous host must have a similar codon usage bias with that of the native strain to ensure successful heterologous production of polyketides this appears to be unfounded and unsupported by the literature.[4]

The more appropriate requirement for successful transcription of polyketide biosynthetic pathways observed in this review requires the heterologous host to possess transcriptional machinery orthologous or functionally comparable to the native strain. This can be observed in the literature by the insertion of proven heterologous promoters to ensure transcription of biosynthetic pathways. Heterologous promoters allow the assurance that the heterologous host is capable of recognizing and utilizing biosynthetic pathways by using promoters known to have transcriptional activators in the host or by

heterologously expressing the promoter's transcription factor along with the pathway. The use of heterologous promoters has provided the majority of examples of successful heterologous polyketide production. This review will present many of the successful examples of heterologous polyketide production and highlight the methodology used to garner success by focusing on the techniques employed to ensure transcription.

1.2 Heterologous promoters as a methodology to ensure transcription

Table 1.2.1: Polyketides produced using heterologous promoters to ensure transcription in the heterologous host.

Product	Type	Native Producer	Promoter	Heterologous Host	Reference
Mutacin/SEKs	Type II PKS	*S. glaucescens* *S. coelicolor* *S. roseofulvus*	actI actI T7	*S. coelicolor* *S. lividans* *E. coli*	[9,30,31]
Epothilones	Mixed PKS/NRPS	*S. cellulosum*	actI pikAI T7	*S. coelicolor* *S. venezuelae* *E. coli*	[32-34]
6-MSA	Fungal PKS	*P. patulum*	actI T7	*S. coelicolor* *E. coli*	[35,36]
6-dEB	Type I PKS	*S. erythrae*	actI T7	*S. lividans* *E. coli*	[8,13]
Tylosins	Type I PKS	*S. fradiae*	actI pikAI	*S. venezuelae*	[37]
Narbonolides	Type I PKS	*S. venezuelae* *S. erythrae*[B] *S.venezuelae*[B] *S. antibioticus*[B]	actI actI[B]	*S. lividans* *S. lividans*	[38,39]
Wailupemycins	Type II PKS	*S. maritimus*	actI	*S. lividans*	[40]
Aloesoponarin II	Type II PKS	*S. roseofulvus*	actI	*S. coelicolor*	[41]
R1128	Type II PKS	*S. sp* R1128	actI	*S. lividans*	[42]
Deoxyoleandolide	Type I PKS	*S. antibioticus*	actI	*S. lividans*	[43]
Macrotetrolides	Type III PKS	*S. griseus*	actI	*S. lividans*	[44]
Soraphen A	Type I PKS	*S. cellulosum*	tipA	*S. lividans*	[45]
Dechloroneocarzilin A	Type I PKS	*S. carzinostaticus*	tipA	*S. coelicolor*	[46]
Meridamycin	Type I PKS	*S. sp* NRRL 30748	ermE	*S lividans*	[47]
Flavanols/Flavones	Type III PKS	*P. crispum* *C. siensis* *C. unshins*	ermE	*S. venezuelae*	[48]
Phenylpropanoids	Type III PKS	*S. coelicolor* *A. thaliana* *M. sativa*	ermE	*S. venezuelae*	[49]

Napthoic acid	Type I PKS	*S. carzinostaticus*	ermE	*S. lividans* *S. coelicolor*	[50]
Fredericamycin	Type II PKS	*S. griseus*	ermE	*S. albus*	[51]
Landomycin	Type II PKS	*S. cyanogenus*	ermE	*S. fradiae*	[52]
Erythromycins	Type I PKS	*S. erythrae*	T7	*E. coli*	[53]
P8_1-OG	Mixed PKS/NRPS	*A. mediterranei*	T7	*E. coli*	[54]
Yersiniabactin	Mixed PKS/NRPS	*Y. pestis*	T7	*E. coli*	[55]
SMA76/SMA93	Fungal PKS	*G. fujikuroi*	T7	*E. coli*	[9]
Midecamycin	Type I PKS	*S. hygroscopicuc*	tylGI	*S. fradiae*	[24]
Chalcolactones	Type I PKS	*S. bikiniensis*	tylGI	*S. fradiae*	[56]
Tylactone	Type I PKS	*S. fradiae*	pikAI	*S. venezuelae*	[37]
Orsellinic acid	Iterative Type I PKS	*S. viridochromogenes*	tipA	*S. lividans* *S. coelicolor*	[57]
Myxothiazol	Mixed PKS/NRPS	*S. cellulosum*	Pm	*P. putida*	[26]
Myxochromide S	Mixed PKS/NRPS	*S. aurantica*	Pm Tn5 aphII	*P. putida* *M. xanthus* *C. macrosporus* GT-2	[58-60]
Flaviolins	Type III PKS	*S. cellulosum*	Pm	*P. putida* *P. syringae* *P. stutzeri*	[61]

A – Produced when co-expressed with PKS4 from *G. fujikuroi*

B – Expressed as chimeric PKS mixtures using the erythromycin, pimaricin, and oleandomycin pathways

1.2.1 The actI/actIII- actII-ORF4 promoter-activator expression system

The most successful approach towards the heterologous production of polyketides involves the incorporation of heterologous promoters known to be functional in the host in front of the transcriptional start site of the foreign biosynthetic pathway. The logic behind this methodology assumes that the native host is not capable of recognizing or utilizing promoters native to the pathway and thus not capable of producing the desired compound.

In 1984 Malpartida and Hopwood successfully produced actinorhodin, a type II polyketide native to *Streptomyces coelicolor,* in the heterologous host *Streptomyces parvulus* by inserting a SCP2-based plasmid containing the actinorhodin biosynthetic pathway into their host strain.[62] Following this study, the actinorhodin biosynthetic pathway was inserted into several other streptomycetal hosts allowing production of actinorhodin and several shunt products.[4-6] In the early 1990s regulatory mechanisms native to the actinorhodin pathway were first explored. The regulatory actII-ORF4 gene was discovered, and its ability to regulate transcription from the actI/actIII promoter leading to an increase in actinorhodin production in *S. coelicolor* when overexpressed was explored.[63-65] This established the actI/actIII-actII-ORF4 promoter-activator expression system as a successful technique for the heterologous production of polyketides in *streptomyces*.

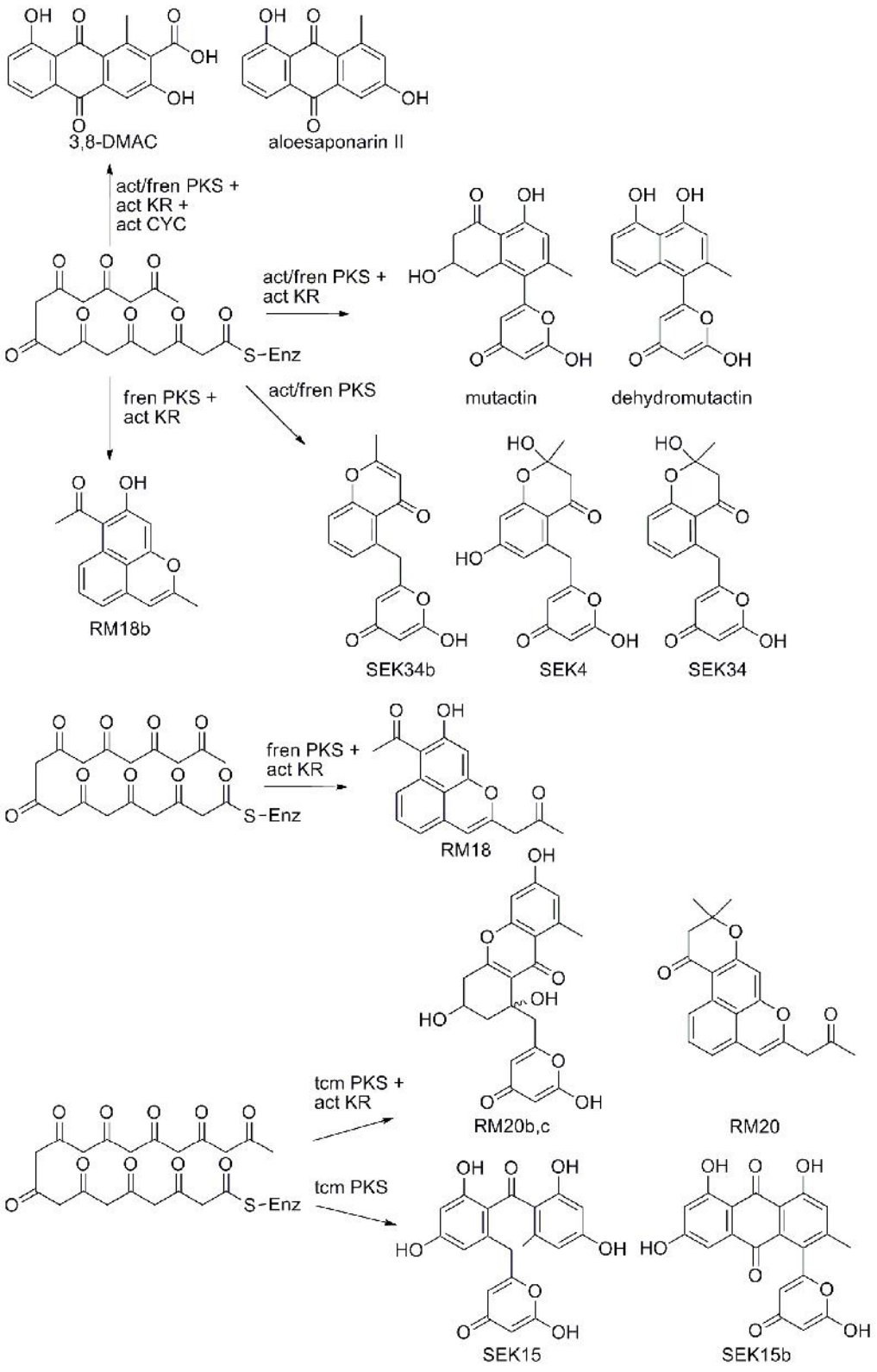


Figure 1.2.1.1: Structures of recombinant polyketides generated using the actI/actIII expression system in heterologous host *S. coelicolor*.

The Khosla group provides the first examples of production of polyketides using heterologous promoter exchange and insertion of the engineered biosynthetic pathway into a streptomycetal host.[30,31,66-68] Each of the examples involve the insertion of a type II polyketide pathway into a SCP2 based vector downstream of the actI/actIII promoter. The actII-ORF4 regulatory gene necessary for activation of the actI/actIII promoter was also cloned into the vector pRM5 to allow transcription of the inserted pathway.[66] The previous production of actinorhodin and other shunt products from heterologous hosts throughout the 1980s established the actI promoter as a robust promoter for an expression vector designed to produce type II PKS products.

Using their expression platform, the Khosla group began exploring polyketide production across several combinations of ACPs, KSs, and CLFs from the actinorhodin pathway from *S. coelicolor,* the granatacin pathway from *Streptomyces violaceoruber*, the tetracenomycin pathway from *Streptomyces glaucescens*, and the frenolicin pathway from *Streptomyces roseofulvus*.[68] Later experiments used the same expression systems and biosynthetic pathways to look at similar combinations of PKS elements including the introduction of various ketoredutases.[31] Within one year of its introduction the actI/actIII-actII-ORF4 promoter activator expression system produced at least 10 previously uncharacterized molecules, increased the understanding of the cyclization mechanisms in type II polyketide biosynthesis, and established the ground work for the use of heterologous expression as a platform for combinatorial chemistry development (Fig 1.2.1.1). Use of the actI/actIII promoter has provided several unique compounds

from type I, II, III, fungal, and mixed NRPS/PKS pathways across three strains of *streptomyces* since its development (Table 1.2.1).

Currently the actI promoter remains the most successful heterologous expression platform for the production of polyketides. However there have been only two proven heterologous hosts, *S. coelicolor* and *S. lividans*, capable of utilizing the actI/actIII-actII-ORF4 promoter activator expression system. Although several derivatives to the original SCP2 derived vector have been developed, each expression platform requires the heterologous host to be capable of expressing functional actII-ORF4 as no promoters are engineered at the start of the gene (Fig 1.2.1.2).[69-71] This limitation provides an explanation for the expression system only being functional in the actII-ORF4 native strain *S. coelicolor* and its close relative *S. lividans*. The type I PKS tylosin was shown to be produced from an actI/actIII developed vector however the authors acknowledge that the strain was capable of transcribing the tylosin pathway in the absence of the promoter and conclude that the promoter intrinsic to the tylosin pathway may have been utilized instead.[37] Other strains not capable of utilizing the actI/actIII-actII-ORF4 promoter activator expression system include *E. coli* and *B. subtilis*. With only two proven heterologous hosts and no further attempts to provide functionality in non-*streptomyces* methodology utilizing the actI/actIII-actII-ORF4 promoter activator expression system remains limited.

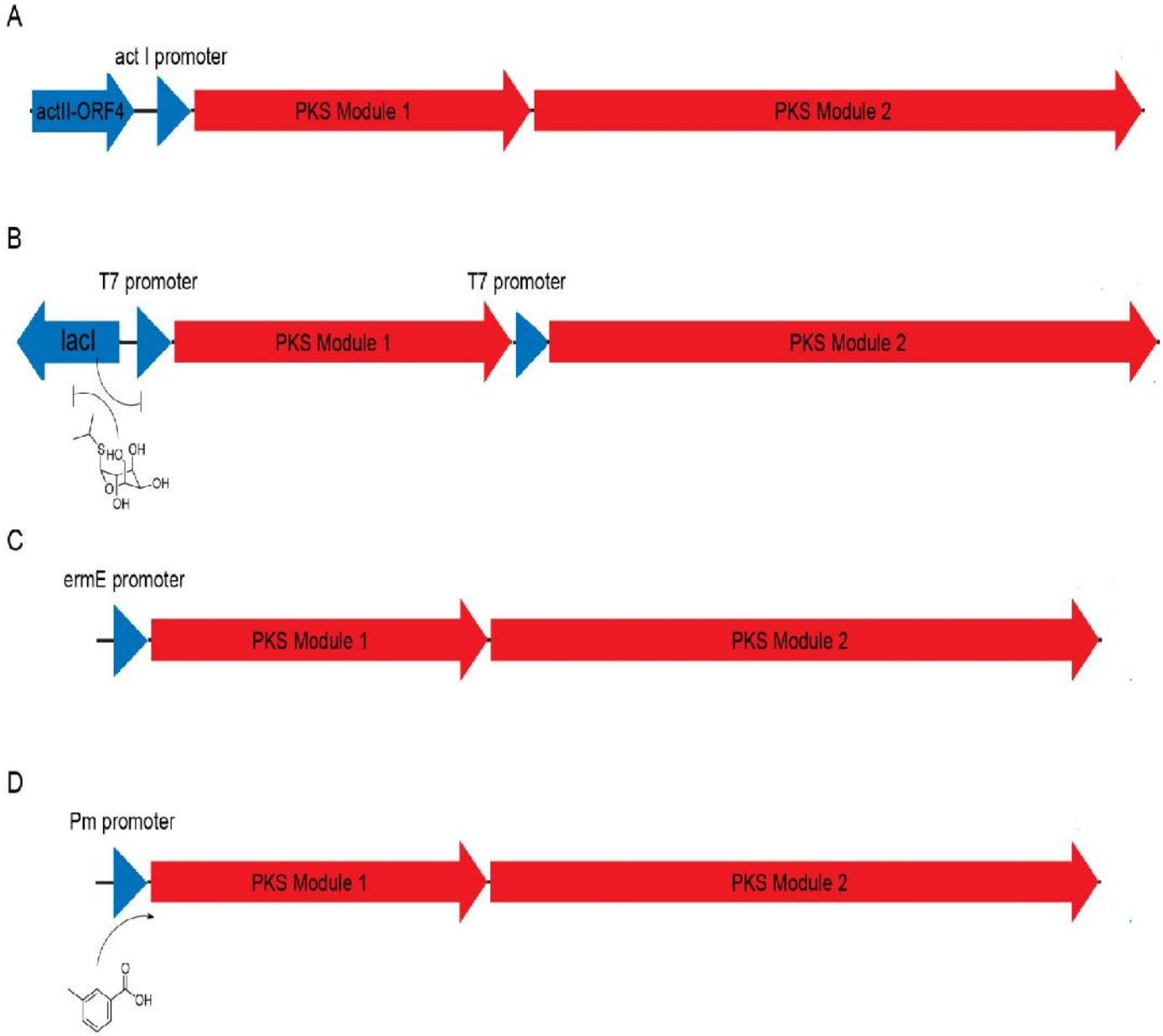


Figure 1.2.1.2: A) actII-ORF4/actI promoter platform for expression in *streptomyces*. B) T7 promoter platform for expression in *E. coli*. C) ermE promoter platform for expression in *streptomyces*. D) Pm promoter platform for expression in *pseudomonads*.

1.2.2 The T7 expression system

Although the area of heterologous expression has been dominated by *streptomyces* since its introduction, the combination of the T7 expression system and the previously described BAP1 strain has provided a viable alternative. The T7 expression system requires the host strain of *E. coli* to possess a chromosomal copy of the T7 polymerase necessary for transcription from the T7 promoter. The vector designed for heterologous expression must place T7 promoters at the start of each coding region in the PKS pathway. Transcription from these promoters is inhibited by the negative regulator *lacI* which is also on the inserted vector. Transcription is induced by the addition of lactose or more commonly isopropyl β-D-1-thiogalactopyranoside (IPTG) which alleviates *lacI* inhibition.

E. coli as a heterologous host possesses several beneficial characteristics including: fast growth rates, amenable culture conditions, a vast genetic toolset, characterized primary metabolism, and the lack of any native PKS pathways.[72]Following the production of 6-dEBs in *E. coli* BAP1 several groups produced various dEBs-based analogs by replacing, inactivating, or changing the available starter units of the dEBs promiscuous loading didomain.[13,18,22,23] In 2010 Zhang et al successfully heterologously expressed the complete erythromycin biosynthetic pathway including all tailoring enzymes producing erythromycin A from *E. coli* by placing the T7 promoter in front of multiple operons throughout the entire pathway across several expression vectors.[53]

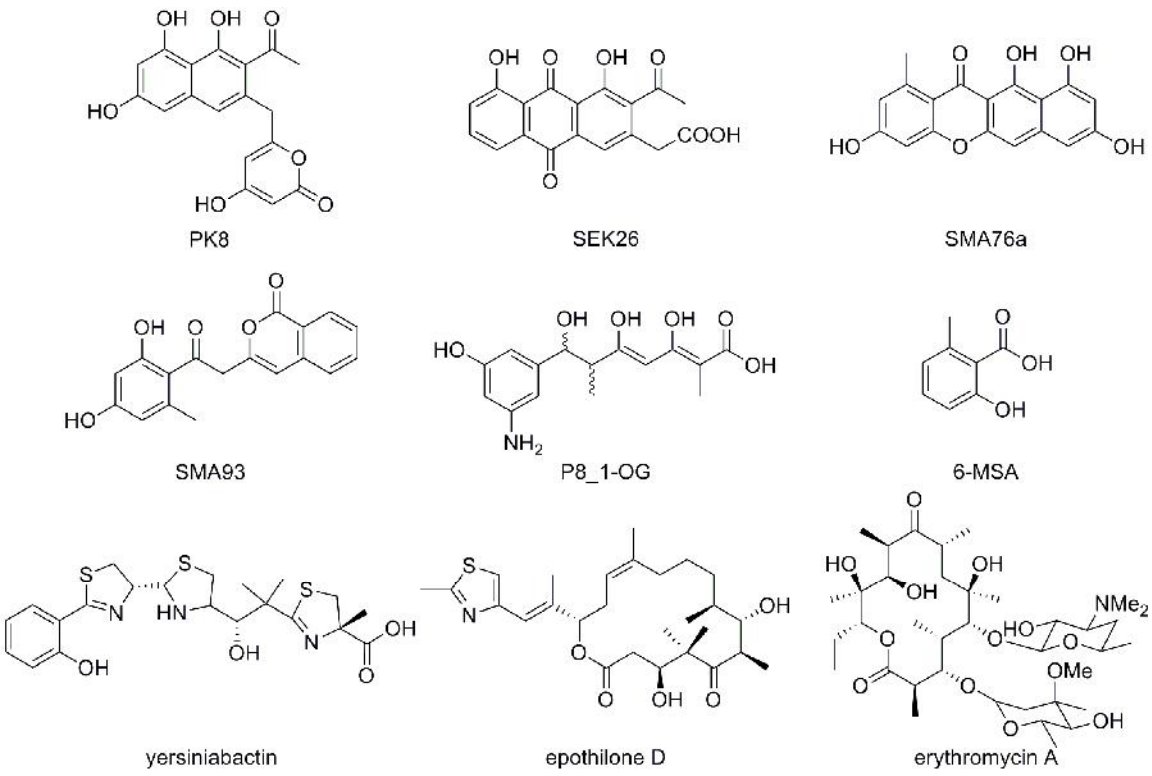


Figure 1.2.2.1: Polyketides produced using the T7 expression system in *E. coli*

The successful production of P8_1-OG from the rifamycin hybrid NRPS/PKS pathway native to *Amycolatopsis mediterranei* provides an excellent example of the versatility of the T7 promoter when used in *E. coli* BAP1.[54] The introduction of the seven genes needed to produce the 3-amino-5-hydroxybenzoic acid (AHBA) primer unit into *E. coli* resulted in poor expression of two of the required proteins. The genes encoding these proteins were replaced with the homologs present in the ansamycin biosynthetic pathway from *Actionsynnema pretiosum*. This enabled detectable levels of the starter unit AHBA though the authors do not address potential explanations for the poor expression of the AHBA pathway from *A. mediterranei*. The initial hybrid NRPS/PKS, RifA, proved another hurdle due to its size of ~530 kDa, which prevented soluble expression in *E. coli*. This was remedied by splitting the RifA protein into two

fragments of similar size and introducing the terminal domains from the docking regions between DEBS modules 2 and 3 onto their respective C- and N- termini. Through this novel linkage the successful production of P8_1-OG, a key intermediate from the rifamycin NRPS/PKS pathway, was observed.[54]

The use of the T7 promoter has provided successful production of various type I, fungal, and mixed NRPS/PKS pathways in *E. coli* and remains the most successful expression technique for the production of polyketides in *E. coli* (Fig 1.2.2.1).The T7 expression system has however only been successful for the production of polyketides in *E. coli.* Use of the T7 expression system also requires engineered variants of *E. coli* as the presence of the T7 polymerase and a functional phosphopantetheinyl transferase are required.[72] The T7 expression system also requires that T7 promoters be inserted at the start of each individual biosynthetic transcript to ensure proper transcription (Fig 1.2.1.2).[72] This limitation requires each biosynthetic pathway to undergo rigorous engineering to enable polyketide production. This handicap has limited the production of polyketide biosynthetic pathways that possess several operons with multiple changes in direction such as type II PKS pathways.[72] The required intensive cloning, limitations on the classes of polyketides, and dependence on a specific heterologous host has attributed to the low number of metabolites successfully produced in *E. coli* using the T7 expression system.

1.2.3 The ermE expression system

The ermE promoter from *S. erythrae* erythromycin resistance gene *ermE* has proven to be a successful promoter for the production of type I, II, and III PKS products

in *streptomyces*hosts. The ermE promoter was first recognized as an asset due to its ability to successfully express functional glycotransferases involved in polyketide tailoring enzymes in heterologous hosts.[73-75] The production of the heavily glycosylated type II polyketide 5-6-anhydro-landomycin A, endogenous to *Streptomyces cyanogenus,* by the heterologous host *Streptomyces fradiae* was the first utilization of the ermE promoter for transcription of an entire PKS pathway.[52] Mulert *et al* hypothesized that previous unsuccessful attempts to produce landomycins in *S. fradiae* were hampered due to either competition between native biosynthetic enzymes from the structurally similar urdamycin PKS pathway or absence of regulatory enzymes necessary for the production of landomycin. The authors constructed a *S. fradiae* strain incapable of producing urdamycin and co-expressed the landomycin PKS downstream of the ermE promoter along with a previously undiscovered regulatory gene, *lndI*, and were able to detect 5,6-anhydro-landomycin A.[52]

Lui et al observed that the mixed NRPS/PKS product meridamycin from the native host *Streptomyces* sp. NRRL 30748 could not be produced in *S. lividans*.[47] The mederimycin pathway was isolated from a BAC library and possessed all of its necessary promoters. The authors identified the transcript levels from the first gene in the biosynthetic pathway *merP* encoding the NRPS section of the biosynthesis were much lower when compared to the native producer. The replacement of the merP promoter with the ermE promoter led to detectable levels of the merP transcript and detectable levels of meridamycin. This result represents one of the few examples in which the authors report the unsuccessful production using native promoters and investigate which of the promoters could not be utilized in the heterologous host.[47] Other applications of

the ermE promoter include its success with the production of flavones and phenylpropanoids from type III PKS synthase pathways in the heterologous host *S. venezuelae*.[48,49]

Although the ermE promoter has proven to be capable of ensuring the transcription of polyketide biosynthetic pathways in several strains of *streptomyces* the number of successfully produced metabolites is much fewer than the metabolites produced using the actI/actIII-ORF4 expression system (Fig 1.2.3.1). All reported examples of successful production of non-*streptomyces* natural products using the ermE promoter have also required the codon optimization of the inserted biosynthetic pathway.[48,49] Although a potential alternative to the use of the actI/actIII-ORF4 expression system, the ermE promoter has also only been capable of producing heterologous polyketides in *streptomyces* hosts, and its limitations are much less detailed in the literature.

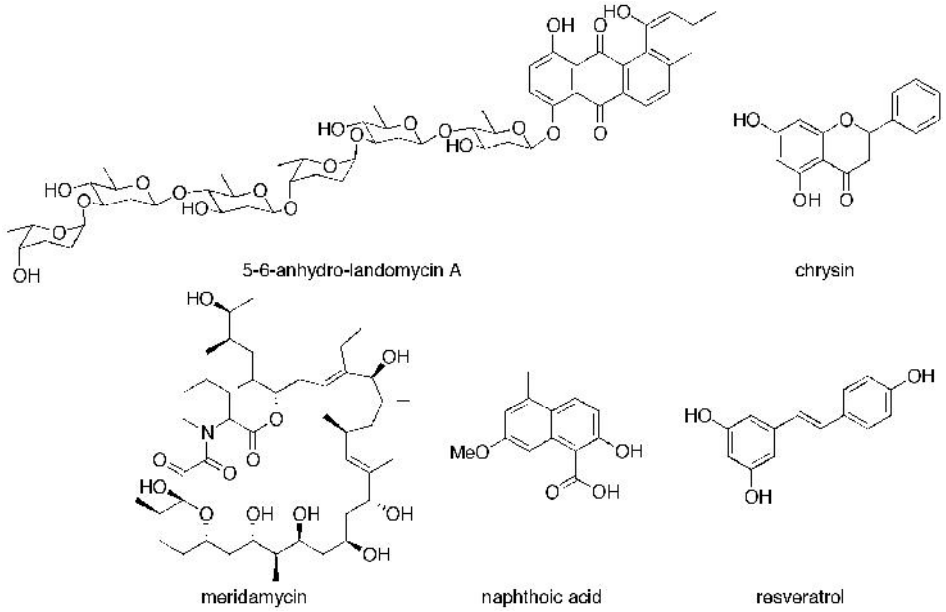


Figure 1.2.3.1: Polyketides produced using the ermE promoter in *S. venezuelae*

1.2.4 The Pm expression system

In 2006 Gross et al engineered a *Pseudomonas putida* strain capable of generating pools of the commonly used PKS precursor methyl-malonyl-CoA.[26] Through the combined use of this *P. putida* strain and the introduction of the known *pseudomonas* toluic acid inducible Pm promoter into the myxothiazol mixed NRPS/PKS pathway from *Stigmatella aurantiaca*, the heterologous host was capable of producing the methyl-malonyl-CoA dependent metabolite myxothiazol.[26] Induction of the Pm promoter by addition of toluic acid increases the levels of the Pm promoter activator *xylS* which is natively transcribed by the transcription factors σ^H and σ^S in the native *pseudomonas* host.[76-78] Previously the use of the Pm promoter as a suitable expression method for the production of myxobacterial metabolites was established through the production of the mixed NRPS/PKS product myxochromide S from the native host *S. aurantiaca*.[58] Using *P. putida* as a host, maximum production of myxochromide S was observed in half of the time required to get proportional production levels from the native host.

The production of metabolites from pathways that are cryptic or silent in their native hosts represents one of the attractive benefits of heterologous expression. Utilizing the Pm promoter,*P. putida,Pseudomonas syringae*, and *Pseudomonas stutzeri* were capable of producing flaviolins from a cryptic type III PKS native to the myxobacteria, *S. cellulosum*.[58,61] This example establishes *pseudomonas* utilizing the Pm promoter as suitable heterologous hosts for the production of cryptic myxobacterial PKS pathways.

The use of the Pm promoter expression methodology remains the only reported approach towards the production of complex myxobacterial metabolites in *pseudomonas* (Fig 1.2.4.1). However the Pm promoter has only been proven to be capable of producing myxobacterial metabolites and is specific to *pseudomonas* heterologous hosts. Although myxobacterial polyketides represent an interesting subset of natural products until a greater diversity of polyketides are succesfully produced, the Pm promoter will remain limited and poorly represented in the literature.

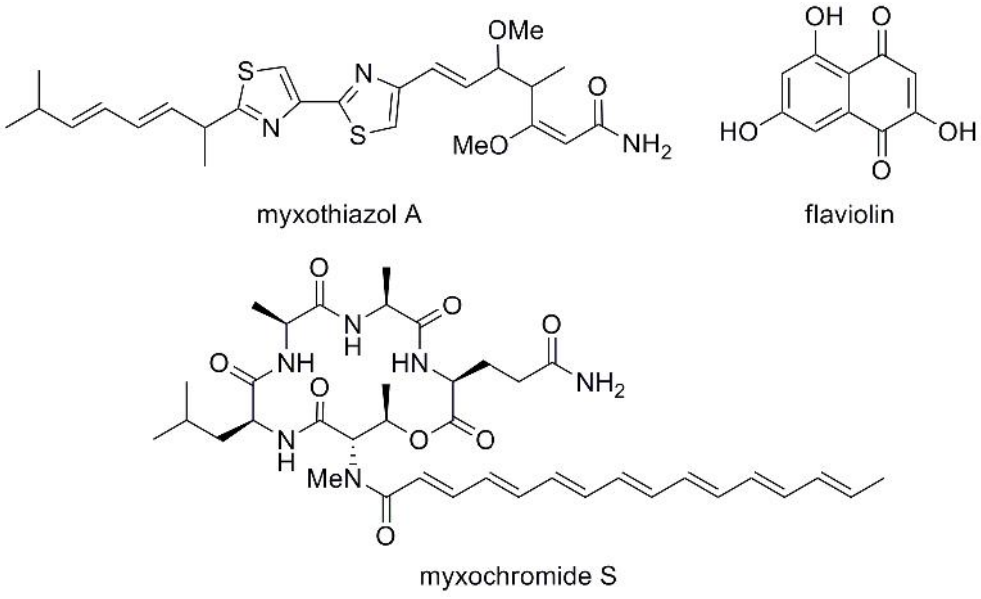


Figure 1.2.4.1: Myxobacterial polyketides produced using the Pm promoter in *P. putida*

1.2.5 Other heterologous expression systems

Several promoters beyond the previously mentioned examples have proven to be successful in the production of polyketides from heterologous hosts. Use of the pikAI promoter from the pikromycin biosynthetic pathway native to *S. venezuelae* has been successful in the production of the mixed NRPS/PKS epothilone and the production of tylosins from type I PKS pathways in *S. venezuelae*.[37] The thiostrepton inducible tipA

promoter allows production of the type I polyketides soraphen A from *S. cellulosum* in *S. lividans*, dechloroneocarzilin A from *Streptomyces carzinostaticus* in the heterologous *S. coelicolor*, and orsellinic acid from *Streptomyces viridochromogenes* in both *S. lividans* and *S. coelicolor*.[45,46,57] When utilized in the native host *S. fradiae*, the tylGI promoter from the tylosin biosynthetic pathway has been found to provide production of the type I polyketides midecamycin from *Streptomyces hygroscopicuc* and chalcolactone intermediates from *Streptomyces bikiniensis*.[24,56] Other promoters endogenous to their respective strains have been proven to be useful towards the heterologous production of polyketides such as the Tn5 promoter from *Myxococcus xanthus* and the aphII promoter from the thermophile *Corallococcus macroporus* GT-2.[59,60] However each of these techniques have been used to produce only a few or a single polyketide product in very specific heterologous hosts. Their utility and general use has yet to be determined.

Table 1.3.1: Polyketides produced in heterologous hosts using the promoters native to the biosynthetic clusters.

Product	Type	Native Producer	Heterologous Host	Reference
Oxytetracycline	Type II PKS	*S. rimosus*	*S. lividans* *S. albus* *M. xanthus* *E. coli*	[79,80]
Tetracenomycins	Type II PKS	*S. glacescens*	*S. cinnamonesis* *S. lividans*	[81]
Granaticin	Type II PKS	*S. violaceoruber*	*S. coelicolor* *S. lividans*	[82,83]
6-MSA	Type I PKS	*S. pactum*	*S. lividans*	[84]
3,5-dihydroxybenzoic acid	Type I PKS	*S. violaceoruber*	*S. coelicolor*	[85]
A-74528	Type II PKS	*S. sp* SANK 61196	*S. lividans*	[86]
Fredericamycin	Type II PKS	*S. sp* SANK 61196	*S. lividans*	[86]
Actinorhodin	Type II PKS	*S. coelicolor*	*S. parvulus*	[62]
Aloesaponarin II	Type II PKS	*S. coelicolor*	*S. parvulus* *S. lividans*	[41]
Aureothin	Type I PKS	*S. thioluteus*	*S. lividans*	[87]
Benastatin	Type II PKS	*S. sp* A2991200	*S. albus* *S. lividans*	[88]
Cytosylglucuronic acid	Type I PKS	*S. griseochromogenes*	*S. lividans*	[89]
Dehydrorabelomycin	Type II PKS	*S. murayamaensis*	*S. lividans*	[90]

Epothilone	Mixed NRPS/PKS	*S. cellulosum*	*M. xanthus*	[91,92]
Fluorostatins	Type II PKS	Environmental DNA	*S. albus*	[10]
Furaquinocin	Mixed PKS/Isoprenoid	*S. sp* strain KO-3988	*S. lividans*	[93]
Griseorhodin A	Type II PKS	*S. sp* JP95	*S. lividans*	[94]
Lysolipin	Type II PKS	*S. tendae* Tu 4042	*S. albus*	[95]
Medermycin	Type II PKS	*S. sp* AM-7161	*S. coelicolor*	[96]
PD116740	Type II PKS	*S. sp* strain WP 4669	*S. lividans*	[97]
Fridamycin E	Type II PKS	*S. rimosus*	*S. lividans*	[97]
Ravidomycins	Type II PKS	*S. ravidus*	*S. lividans*	[98]
Spectinabilin	Type I PKS	*S. spectinabilis*	*S. lividans*	[99]
Steffimycin	Type II PKS	*S. steffisburgensis*	*S. albus*	[100]
Myxothiazol	Mixed PKS/NRPS	*S. aurantiaca*	*M. xanthus*	[101]
Albicidin	Mixed PKS/NRPS	*X. albineaus*	*X. axonopodis*	[102]

1.3 Heterologous production hosts capable of utilizing native promoters

Although use of heterologous promoters has provided the majority of the examples of heterologous expression to date, their use provides little investigation into the transcriptional capabilities of heterologous hosts. Many examples of successful polyketide production from inserted, heterologous promoters observed throughout the literature do not report the unsuccessful production when using the promoters intrinsic to the PKS pathway. This lack of investigation has led to the use of a select few strain dependent promoters known to be active in the host in which they are incorporated. The rapid adoption of heterologous promoter systems into the preferred methodology has afforded little scientific investigation into the capabilities of heterologous hosts to recognize and utilize foreign promoters and the limits for each heterologous host.

Throughout the 1980s several type II PKS pathways were isolated from various *streptomyces* strains and inserted into closely related strains of *streptomyces*.[62,79,103-105] These experiments, such as the previously mentioned actinorhodin production from the heterologous host *S. parvulus*, provided the first examples of heterologous polyketide production and provided much insight into the evolution, mechanism, regulation, and isolation of PKS pathways. These experiments also established *streptomyces* as the bacterium of choice for the production of heterologous metabolites.

1.3.1 *Streptomyces* as heterologous hosts

From the initial success of *S. parvulus* to the incredibly versatile *S. lividans* the *streptomyces* remain the powerhouse of heterologous hosts capable of utilizing heterologous promoters systems and producing the subsequent natural product. This success has led to the construction of several engineered strains of *streptomyces* optimized for production and detection of heterologous metabolites. Engineered strains of *S. coelicolor*, *S. lividans*, and *S. albus* incapable of producing the majority of their native polyketides have been developed to eliminate background competition for metabolites and also to provide a much less complex matrix for the detection and isolation of produced metabolites.[4,5]

The most successful engineered strain *S. lividans* has proven to be capable of recognizing the endogenous promoters from a variety of PKS pathways native to *streptomyces* and producing over 15 different metabolites since its development as a heterologous host. In 1999 Ziermann et al engineered a *S. lividans* strain, T4-114, incapable of producing its native polyketide actinorhodin.[106] Similar work had been done

by Kao et al to construct the actinorhodin negative strain *S. coelicolor* CH999 however transformation of *S. coelicolor* requires non-methylated DNA and also exhibits low transformation efficiencies.[4,5,107] Recently Zaleta-Rivera et al transformed the newly discovered A-74528 type II PKS pathway from *Streptomyces* sp. SANK 61196 into *S. lividans* T4-114 to explore the exceptional sequence identity to the fredericamycin biosynthetic pathway from *S. rimosus*.[86] *S. lividans* T4-114 strains transformed with the entire A-74528 biosynthetic pathway were capable of producing the native metabolite, A-74528, and fredericamycin A. The authors hypothesize that the minimal difference between the two biosynthetic pathways could be attributed to the transference of fredericamycin PKS into *Streptomyces* sp. SANK 61196 or vice-versa followed by further recombination and rearrangements.[86] This snap shot of the evolution between the two pathways also allows the observation that the evolution between the two pathways had minimal effect on the promoters of the two pathways and that both of the native hosts *S. rimosus* and *S.* sp. SANK 61196 share the ability to recognize promoters native to the two biosynthetic pathways. A further observation is that *S. lividans* T4-114 also has the ability to recognize the native promoters as no introduction of a heterologous promoter into the pathway was necessary.

The heterologous production of the rare nitrophenyl-substituted polyketides aureothin and spectinabilin from native strains *Streptomyces thioluteus* and *Streptomyces spectabilis* respectively from *S. lividans* represents another example of the production of two metabolites from evolutionary divergent PKS pathways.[87,99] The heterologous production of aureothin from *S. lividans* provided the first strong evidence for iteration from a type I PKS module.[87] The heterologous production of spectinabilin further

demonstrated the iterative ability of the conserved type I PKS module and also explored the high level of homology between the two PKS pathways. The authors hypothesize that this homology is due to the transference of the PKS pathway between the two strains and the differences arise from further evolution within the respective native strain.[99] The ability to produce both metabolites aureothin and spectinabilin from the same *S. lividans* host without the introduction of non-native promoters into either pathway further exemplifies *S. lividans* ability to recognize a diverse range of streptomycetal promoters.

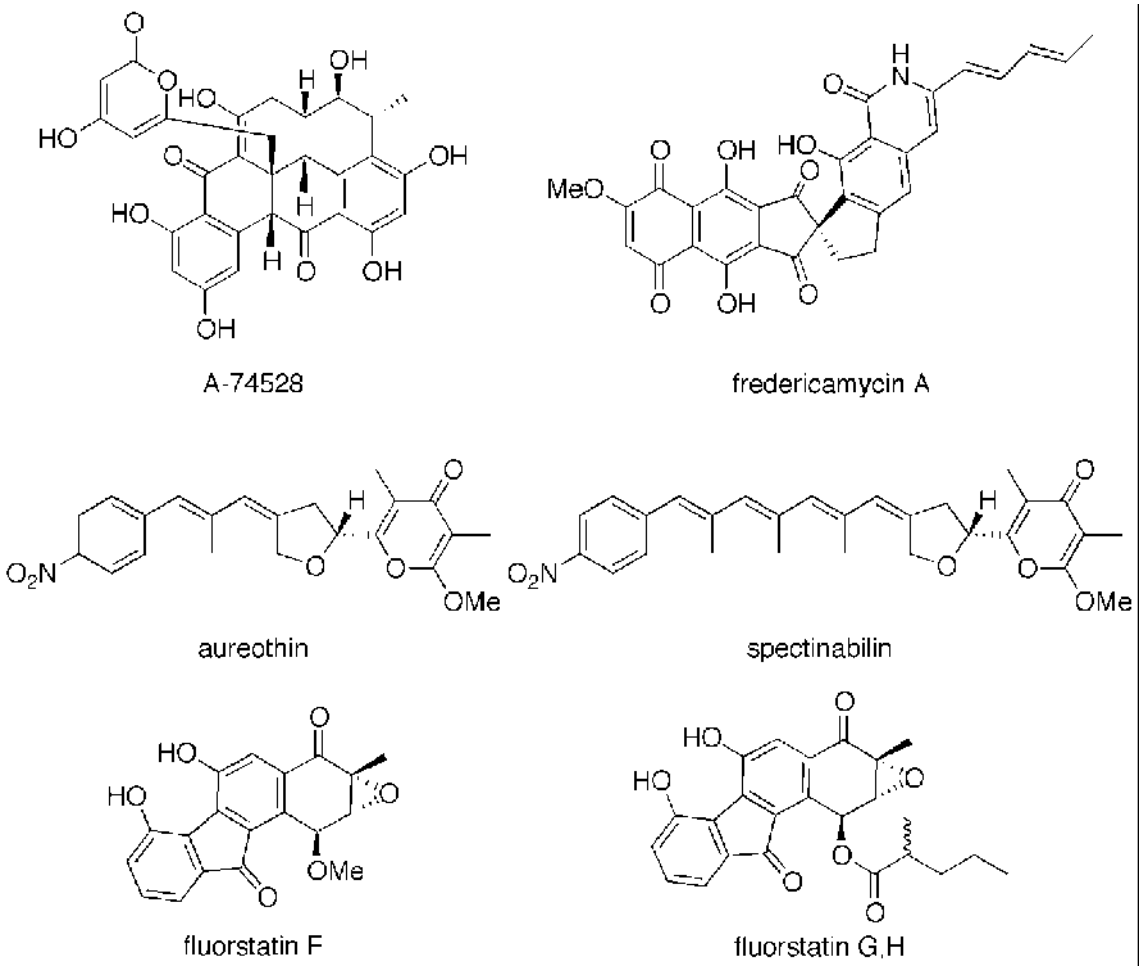


Figure 1.3.1.1: Polyketides produced by heterologous expression in *S. lividans*

The Brady lab has provided the only example of the production of previously undiscovered natural products from a BAC library constructed from eDNA.[10] The heterologous host *S. albus* was able to produce multiple fluorstatin compounds previously uncharacterized without need of an introduced promoter. This work provides the first successful example of a culture independent methodology for the production of metabolites from eDNA. However the fluorstatin compounds and their precursor metabolite rabelomycin are known to be produced from streptomycetal strains such as *Streptomyces lavendulae* and *Streptomyces olivaceus*, and the authors do not report the production of any other isolated metabolites known to be produced by non-streptomycetal strains.[10]

*Streptomyces*as heterologoushosts have proven to be capable of producing type I, II, and mixed isoprenoid/PKS products utilizing the native promoters from approximately 20 various streptomycetal biosynthetic pathways (Fig 1.3.1.1). However no successful production of a polyketide has been reported in a *streptomyces* host utilizing the PKS promoters native to a pathway from a non-streptomycetal strain (Table 1.3.1).

1.3.2 *Myxobacteria*as heterologous hosts

Although the genetic toolset available when working with *myxobacteria* pales in comparison to that of *streptomyces*, much work has been done towards the establishment of the *myxobacteria* as a viable heterologous host. When compared to *streptomyces*, the amount of heterologous production reported in myxobacterial strains lacks the range of metabolites; however the examples provided present potential that has not been reported from streptomycetal hosts. Initial work with the myxobacterial host *Myxococcus xanthus*

demonstrated the hosts ability to produce the mixed NRPS/PKS metabolite epothilone from the closely related native strain *S. cellulosum*.[91] This work, along with the production of myxothiazol from the native *myxobacteria,Stigmatella aurantiaca* demonstrates*M. xanthus* ability to produce natural products from related strains of *myxobacteria* without the need to introduce heterologous promoters into the inserted PKS pathways and is similar to the methodology observed from heterologous hosts of *streptomyces.*[101]

However production of the type II PKS product oxytetracycline from native host *Streptomyces rimosus* in *M. xanthus* represents the only reported example of a myxobacterial host producing a polyketide native to a streptomycetal host without the use ofa functional heterologous promoter.[80] The oxytetracycline biosynthetic pathway was inserted into the chromosome of *M. xanthus* through homologous recombination and mutants possessing the insertion were capable of producing oxytetracycline (Fig 1.3.2.1). This work established *M. xanthus* as the first reported strain capable of producing polyketides from a genetic source outside of its native family of bacterium without the requirement of heterologous promoters.This attribute makes *M. xanthus* an excellent candidate as a heterologous host when screening eDNA composed of genetic material from a variety of bacterial families. However, the unfamiliar genetics and techniques have limited the application of *M. xanthus* as a heterologous host present in the literature.[80]

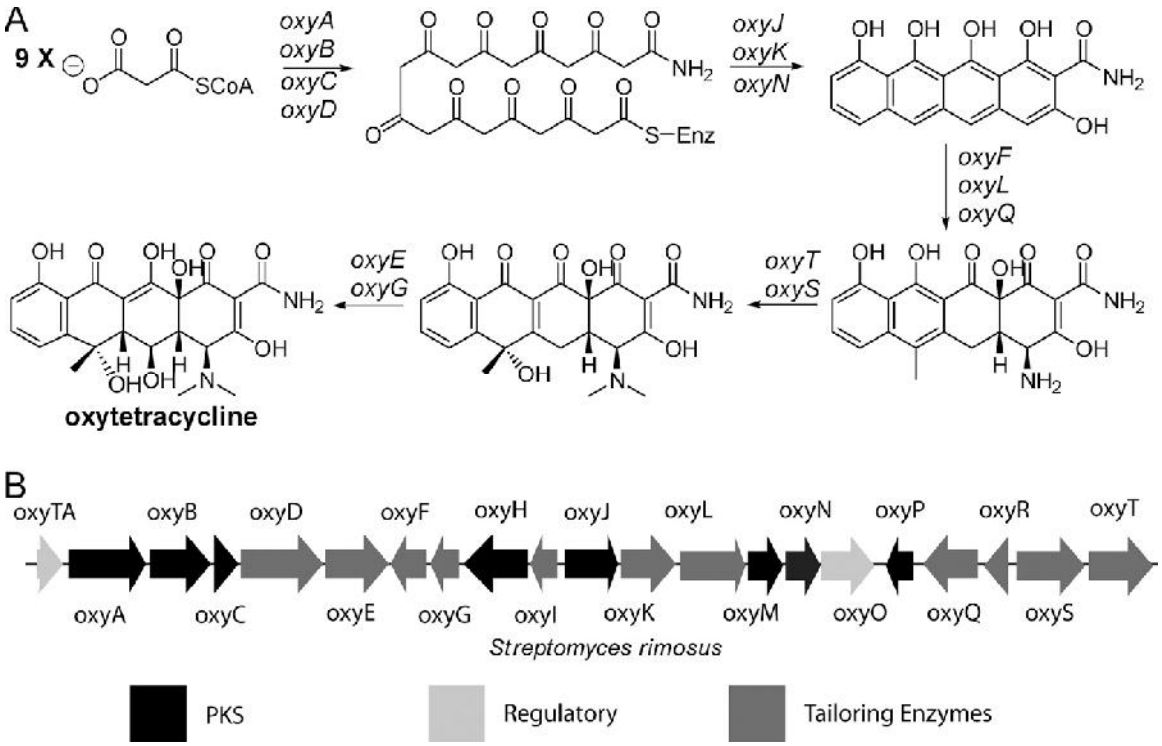


Figure 1.3.2.1: Oxytetracycline biosynthetic pathway. (A) Enzymatic pathway for formation of oxytetracycline. (B) Oxytetracycline biosynthetic gene cluster from *S. rimosus*.

1.3.3 *E. coli* as a heterologous host

The use of *E. coli* as a heterologous host has many attractive advantages. Advantages include the use of chaperone proteins and inclusion of rare tRNA synthases to ensure functional proteins, the ability to insert and produce all necessary starter metabolites, the superious genetic toolset, the ability to produce phophopantethienylated proteins, and the well characterized primary metabolism.[72] Though much work has been done in the heterologous host *E. coli* utilizing the T7 expression system, the range of the specie's ability to recognize and transcribe promoters from foreign DNA has not been reported.

The production of the *S. rimosus* type II polyketide oxytetracycline provided the first example of polyketide production without the insertion of non-native promoters in *E. coli*. This work hypothesized that a general mechanism for the transcription of PKS pathways should be present in bacteria because horizontal transfer of biosynthetic pathways is a key mechanism for their proliferation among species. This general transcription mechanism was further hypothesized to be regulated by alternative transcription factors downstream of the stringent response. Under native culturing conditions it was observed that *E. coli* BAP1 was not capable of producing transcripts from the *oxyABCDE* minimal PKS operon. However transcripts for all other putative operons were detected under standard culturing conditions. A functional promoter recognized by the alternative sigma factor σ^{54} was discovered appropriately placed upstream of the oxytetracycline minimal PKS operon *oxyABCDE* and found to be functional in *E. coli*. Through the introduction of the oxytetracycline PKS pathway and over-expression of σ^{54}*E. coli* BAP1 was capable of producing oxytetracycline. This work provides evidence for the native inability of *E. coli* to transcribe foreign type II minimal PKS genes and establishes alternative sigma factors as a potential means to remedy this limitation. The application of alternative sigma factors as general activators of PKS pathways in other bacterium is an interesting area of research that has yet to be explored further and could potentially strengthen the transcription ability of other heterologous hosts such as *S. lividans* and *M. xanthus*. The production of oxytetracycline from a streptomycetal PKS in *E. coli* represents the second example of a strain capable of heterologous production from genetic material outside of the hosts native family.

1.4 Conclusion

This review focuses on highlighting the techniques employed in the numerous successes observed from the heterologous production of polyketides since its establishment in the 1980s. Throughout observing the examples of heterologous expression of polyketides reported in the literature to date, limitations in each approach and each heterologous host become apparent. The majority of the requirements that need to be met by each heterologous host, such as precursor availability, proper folding, and activation of proteins, have been satisfied by a variety of strain dependent techniques. Many of the successful approaches focus on ensuring transcription of the foreign genetic material. The introduction of heterologous promoters to ensure transcription of heterologous PKS pathways has provided the most success in the production of polyketides and has drastically increased our understanding of the biosynthesis of polyketides. However the amount of success and the rapid adoption of heterologous promoter expression platforms have afforded little scientific investigation into how heterologous hosts recognize and utilize intrinsic promoters and the limits for each heterologous host. Although the *streptomyces* represent the most successful candidate to date, their ability to produce polyketides from non-streptomycetal genomic material has yet to be reported. The unique benefits presented by *M. xanthus* and *E. coli* along with the success from other strains such as *P. putida* ensure their further development as heterologous hosts.

1.5 References

1. Pfeifer, B. A.; Khosla, C. *Microbiol Mol Biol Rev***2001**, *65*, 106.
2. Park, S. R.; Han, A. R.; Ban, Y. H.; Yoo, Y. J.; Kim, E. J.; Yoon, Y. J. *Appl Microbiol Biotechnol***2010**, *85*, 1227.
3. Wenzel, S. C.; Müller, R. *Curr Opin Biotechnol***2005**, *16*, 594.
4. Rodriguez, E.; Menzella, H. G.; Gramajo, H. *Methods Enzymol***2009**, *459*, 339.
5. Zhang, H.; Wang, Y.; Pfeifer, B. A. *Mol Pharm***2008**, *5*, 212.
6. Wilkinson, B.; Micklefield, J. *Nat Chem Biol***2007**, *3*, 379.
7. Khosla, C.; Zawada, R. J. *Trends Biotechnol***1996**, *14*, 335.
8. Xue, Q.; Ashley, G.; Hutchinson, C. R.; Santi, D. V. *Proc Natl Acad Sci U S A***1999**, *96*, 11740.
9. Zhang, W.; Li, Y.; Tang, Y. *Proc Natl Acad Sci U S A***2008**, *105*, 20683.
10. Feng, Z.; Kim, J. H.; Brady, S. F. *J Am Chem Soc***2010**, *132*, 11902.
11. Fischbach, M. A.; Walsh, C. T. *Chem Rev***2006**, *106*, 3468.
12. Clardy, J.; Fischbach, M. A.; Walsh, C. T. *Nat Biotechnol***2006**, *24*, 1541.
13. Pfeifer, B. A.; Admiraal, S. J.; Gramajo, H.; Cane, D. E.; Khosla, C. *Science***2001**, *291*, 1790.
14. Roberts, G. A.; Staunton, J.; Leadlay, P. F. *Eur J Biochem***1993**, *214*, 305.
15. Murli, S.; Kennedy, J.; Dayem, L. C.; Carney, J. R.; Kealey, J. T. *J Ind Microbiol Biotechnol***2003**, *30*, 500.
16. Lau, J.; Tran, C.; Licari, P.; Galazzo, J. *J Biotechnol***2004**, *110*, 95.
17. Wang, Y.; Pfeifer, B. A. *Metab Eng***2008**, *10*, 33.
18. Pistorino, M.; Pfeifer, B. A. *Biotechnol Prog***2009**, *25*, 1364.

19. Quadri, L. E.; Weinreb, P. H.; Lei, M.; Nakano, M. M.; Zuber, P.; Walsh, C. T. *Biochemistry***1998**, *37*, 1585.

20. Mofid, M. R.; Finking, R.; Marahiel, M. A. *J Biol Chem***2002**, *277*, 17023.

21. Lombó, F.; Pfeifer, B.; Leaf, T.; Ou, S.; Kim, Y. S.; Cane, D. E.; Licari, P.; Khosla, C. *Biotechnol Prog***2001**, *17*, 612.

22. Dayem, L. C.; Carney, J. R.; Santi, D. V.; Pfeifer, B. A.; Khosla, C.; Kealey, J. T. *Biochemistry***2002**, *41*, 5193.

23. Kennedy, J.; Murli, S.; Kealey, J. T. *Biochemistry***2003**, *42*, 14342.

24. Rodriguez, E.; Ward, S.; Fu, H.; Revill, W. P.; McDaniel, R.; Katz, L. *Appl Microbiol Biotechnol***2004**, *66*, 85.

25. Kennedy, J.; Kealey, J. T. *Anal Biochem***2004**, *327*, 91.

26. Gross, F.; Ring, M. W.; Perlova, O.; Fu, J.; Schneider, S.; Gerth, K.; Kuhlmann, S.; Stewart, A. F.; Zhang, Y.; Müller, R. *Chem Biol***2006**, *13*, 1253.

27. Wang, Y.; Boghigian, B. A.; Pfeifer, B. A. *Appl Microbiol Biotechnol***2007**, *77*, 367.

28. Zhang, H.; Boghigian, B. A.; Pfeifer, B. A. *Biotechnol Bioeng***2010**, *105*, 567.

29. Maharjan, S.; Park, J. W.; Yoon, Y. J.; Lee, H. C.; Sohng, J. K. *Biotechnol Lett***2010**, *32*, 277.

30. McDaniel, R.; Ebert-Khosla, S.; Fu, H.; Hopwood, D. A.; Khosla, C. *Proc Natl Acad Sci U S A***1994**, *91*, 11542.

31. Fu, H.; Hopwood, D. A.; Khosla, C. *Chem Biol***1994**, *1*, 205.

32. Tang, L.; Shah, S.; Chung, L.; Carney, J.; Katz, L.; Khosla, C.; Julien, B. *Science***2000**, *287*, 640.

33. Park, S. R.; Park, J. W.; Jung, W. S.; Han, A. R.; Ban, Y. H.; Kim, E. J.; Sohng, J. K.; Sim, S. J.; Yoon, Y. J. *Appl Microbiol Biotechnol***2008**, *81*, 109.

34. Mutka, S. C.; Carney, J. R.; Liu, Y.; Kennedy, J. *Biochemistry***2006**, *45*, 1321.

35. Bedford, D. J.; Schweizer, E.; Hopwood, D. A.; Khosla, C. *J Bacteriol***1995**, *177*, 4544.

36. Kealey, J. T.; Liu, L.; Santi, D. V.; Betlach, M. C.; Barr, P. J. *Proc Natl Acad Sci U S A***1998**, *95*, 505.

37. Jung, W. S.; Lee, S. K.; Hong, J. S.; Park, S. R.; Jeong, S. J.; Han, A. R.; Sohng, J. K.; Kim, B. G.; Choi, C. Y.; Sherman, D. H.; Yoon, Y. J. *Appl Microbiol Biotechnol***2006**, *72*, 763.

38. Tang, L.; Fu, H.; Betlach, M. C.; McDaniel, R. *Chem Biol***1999**, *6*, 553.

39. Tang, L.; Fu, H.; McDaniel, R. *Chem Biol***2000**, *7*, 77.

40. Hertweck, C.; Xiang, L.; Kalaitzis, J. A.; Cheng, Q.; Palzer, M.; Moore, B. S. *Chem Biol***2004**, *11*, 461.

41. Kim, E. S.; Cramer, K. D.; Shreve, A. L.; Sherman, D. H. *J Bacteriol***1995**, *177*, 1202.

42. Marti, T.; Hu, Z.; Pohl, N. L.; Shah, A. N.; Khosla, C. *J Biol Chem***2000**, *275*, 33443.

43. Shah, S.; Xue, Q.; Tang, L.; Carney, J. R.; Betlach, M.; McDaniel, R. *J Antibiot (Tokyo)***2000**, *53*, 502.

44. Kwon, H. J.; Smith, W. C.; Xiang, L.; Shen, B. *J Am Chem Soc***2001**, *123*, 3385.

45. Zirkle, R.; Ligon, J. M.; Molnár, I. *Microbiology***2004**, *150*, 2761.

46. Otsuka, M.; Ichinose, K.; Fujii, I.; Ebizuka, Y. *Antimicrob Agents Chemother***2004**, *48*, 3468.

47. Liu, H.; Jiang, H.; Haltli, B.; Kulowski, K.; Muszynska, E.; Feng, X.; Summers, M.; Young, M.; Graziani, E.; Koehn, F.; Carter, G. T.; He, M. *J Nat Prod***2009**, *72*, 389.

48. Park, S. R.; Paik, J. H.; Ahn, M. S.; Park, J. W.; Yoon, Y. J. *J Microbiol Biotechnol***2010**, *20*, 1295.

49. Park, S. R.; Yoon, J. A.; Paik, J. H.; Park, J. W.; Jung, W. S.; Ban, Y. H.; Kim, E. J.; Yoo, Y. J.; Han, A. R.; Yoon, Y. J. *J Biotechnol***2009**, *141*, 181.

50. Sthapit, B.; Oh, T. J.; Lamichhane, R.; Liou, K.; Lee, H. C.; Kim, C. G.; Sohng, J. K. *FEBS Lett***2004**, *566*, 201.

51. Wendt-Pienkowski, E.; Huang, Y.; Zhang, J.; Li, B.; Jiang, H.; Kwon, H.; Hutchinson, C. R.; Shen, B. *J Am Chem Soc***2005**, *127*, 16442.

52. von Mulert, U.; Luzhetskyy, A.; Hofmann, C.; Mayer, A.; Bechthold, A. *FEMS Microbiol Lett***2004**, *230*, 91.

53. Zhang, H.; Wang, Y.; Wu, J.; Skalina, K.; Pfeifer, B. A. *Chem Biol***2010**, *17*, 1232.

54. Watanabe, K.; Rude, M. A.; Walsh, C. T.; Khosla, C. *Proc Natl Acad Sci U S A***2003**, *100*, 9774.

55. Pfeifer, B. A.; Wang, C. C.; Walsh, C. T.; Khosla, C. *Appl Environ Microbiol***2003**, *69*, 6698.

56. Ward, S. L.; Hu, Z.; Schirmer, A.; Reid, R.; Revill, W. P.; Reeves, C. D.; Petrakovsky, O. V.; Dong, S. D.; Katz, L. *Antimicrob Agents Chemother***2004**, *48*, 4703.

57. Gaisser, S.; Trefzer, A.; Stockert, S.; Kirschning, A.; Bechthold, A. *J Bacteriol***1997**, *179*, 6271.

58. Wenzel, S. C.; Gross, F.; Zhang, Y.; Fu, J.; Stewart, A. F.; Müller, R. *Chem Biol***2005**, *12*, 349.

59. Fu, J.; Wenzel, S. C.; Perlova, O.; Wang, J.; Gross, F.; Tang, Z.; Yin, Y.; Stewart, A. F.; Müller, R.; Zhang, Y. *Nucleic Acids Res***2008**, *36*, e113.

60. Perlova, O.; Gerth, K.; Kuhlmann, S.; Zhang, Y.; Müller, R. *Microb Cell Fact***2009**, *8*, 1.

61. Gross, F.; Luniak, N.; Perlova, O.; Gaitatzis, N.; Jenke-Kodama, H.; Gerth, K.; Gottschalk, D.; Dittmann, E.; Müller, R. *Arch Microbiol***2006**, *185*, 28.

62. Malpartida, F.; Hopwood, D. A. *Nature***1984**, *309*, 462.

63. Takano, E.; Gramajo, H. C.; Strauch, E.; Andres, N.; White, J.; Bibb, M. J. *Mol Microbiol***1992**, *6*, 2797.

64. Gramajo, H. C.; Takano, E.; Bibb, M. J. *Mol Microbiol***1993**, *7*, 837.

65. Fujii, T.; Gramajo, H. C.; Takano, E.; Bibb, M. J. *J Bacteriol***1996**, *178*, 3402.

66. McDaniel, R.; Ebert-Khosla, S.; Hopwood, D. A.; Khosla, C. *Science***1993**, *262*, 1546.

67. Khosla, C.; McDaniel, R.; Ebert-Khosla, S.; Torres, R.; Sherman, D. H.; Bibb, M. J.; Hopwood, D. A. *J Bacteriol***1993**, *175*, 2197.

68. Fu, H.; McDaniel, R.; Hopwood, D. A.; Khosla, C. *Biochemistry***1994**, *33*, 9321.

69. Hu, Z.; Hopwood, D. A.; Hutchinson, C. R. *J Ind Microbiol Biotechnol***2003**, *30*, 516.

70. Bierman, M.; Logan, R.; O'Brien, K.; Seno, E. T.; Rao, R. N.; Schoner, B. E. *Gene***1992**, *116*, 43.

71. Rowe, C. J.; Cortés, J.; Gaisser, S.; Staunton, J.; Leadlay, P. F. *Gene***1998**, *216*, 215.

72. Gao, X.; Wang, P.; Tang, Y. *Appl Microbiol Biotechnol***2010**, *88*, 1233.

73. Trefzer, A.; Hoffmeister, D.; Künzel, E.; Stockert, S.; Weitnauer, G.; Westrich, L.; Rix, U.; Fuchser, J.; Bindseil, K. U.; Rohr, J.; Bechthold, A. *Chem Biol***2000**, *7*, 133.

74. Trefzer, A.; Fischer, C.; Stockert, S.; Westrich, L.; Künzel, E.; Girreser, U.; Rohr, J.; Bechthold, A. *Chem Biol***2001**, *8*, 1239.

75. Hoffmeister, D.; Wilkinson, B.; Foster, G.; Sidebottom, P. J.; Ichinose, K.; Bechthold, A. *Chem Biol***2002**, *9*, 287.

76. Marqués, S.; Gallegos, M. T.; Ramos, J. L. *Mol Microbiol***1995**, *18*, 851.

77. Marqués, S.; Gallegos, M. T.; Manzanera, M.; Holtel, A.; Timmis, K. N.; Ramos, J. L. *J Bacteriol***1998**, *180*, 2889.

78. Marqués, S.; Manzanera, M.; González-Pérez, M. M.; Gallegos, M. T.; Ramos, J. L. *Mol Microbiol***1999**, *31*, 1105.

79. Binnie, C.; Warren, M.; Butler, M. J. *J Bacteriol***1989**, *171*, 887.

80. Stevens, D. C.; Henry, M. R.; Murphy, K. A.; Boddy, C. N. *Appl Environ Microbiol***2010**, *76*, 2681.

81. Li, C.; Hazzard, C.; Florova, G.; Reynolds, K. A. *Metab Eng***2009**, *11*, 319.

82. Ichinose, K.; Bedford, D. J.; Tornus, D.; Bechthold, A.; Bibb, M. J.; Revill, W. P.; Floss, H. G.; Hopwood, D. A. *Chem Biol***1998**, *5*, 647.

83. Martinez, A.; Kolvek, S. J.; Yip, C. L.; Hopke, J.; Brown, K. A.; MacNeil, I. A.; Osburne, M. S. *Appl Environ Microbiol***2004**, *70*, 2452.

84. Ito, T.; Roongsawang, N.; Shirasaka, N.; Lu, W.; Flatt, P. M.; Kasanah, N.; Miranda, C.; Mahmud, T. *Chembiochem***2009**, *10*, 2253.

85. Wenzel, S. C.; Bode, H. B.; Kochems, I.; Müller, R. *Chembiochem***2008**, *9*, 2711.

86. Zaleta-Rivera, K.; Charkoudian, L. K.; Ridley, C. P.; Khosla, C. *J Am Chem Soc***2010**, *132*, 9122.

87. He, J.; Hertweck, C. *Chem Biol***2003**, *10*, 1225.

88. Xu, Z.; Schenk, A.; Hertweck, C. *J Am Chem Soc***2007**, *129*, 6022.

89. Cone, M. C.; Petrich, A. K.; Gould, S. J.; Zabriskie, T. M. *J Antibiot (Tokyo)***1998**, *51*, 570.

90. Gould, S. J.; Hong, S. T.; Carney, J. R. *J Antibiot (Tokyo)***1998**, *51*, 50.

91. Julien, B.; Shah, S. *Antimicrob Agents Chemother***2002**, *46*, 2772.

92. Arslanian, R. L.; Parker, C. D.; Wang, P. K.; McIntire, J. R.; Lau, J.; Starks, C.; Licari, P. J. *J Nat Prod***2002**, *65*, 570.

93. Kawasaki, T.; Hayashi, Y.; Kuzuyama, T.; Furihata, K.; Itoh, N.; Seto, H.; Dairi, T. *J Bacteriol***2006**, *188*, 1236.

94. Li, A.; Piel, J. *Chem Biol***2002**, *9*, 1017.

95. Lopez, P.; Hornung, A.; Welzel, K.; Unsin, C.; Wohlleben, W.; Weber, T.; Pelzer, S. *Gene***2010**, *461*, 5.

96. Ichinose, K.; Ozawa, M.; Itou, K.; Kunieda, K.; Ebizuka, Y. *Microbiology***2003**, *149*, 1633.

97. Hong, S. T.; Carney, J. R.; Gould, S. J. *J Bacteriol***1997**, *179*, 470.

98. Kharel, M. K.; Nybo, S. E.; Shepherd, M. D.; Rohr, J. *Chembiochem***2010**, *11*, 523.

99. Choi, Y. S.; Johannes, T. W.; Simurdiak, M.; Shao, Z.; Lu, H.; Zhao, H. *Mol Biosyst***2010**, *6*, 336.

100. Gullón, S.; Olano, C.; Abdelfattah, M. S.; Braña, A. F.; Rohr, J.; Méndez, C.; Salas, J. A. *Appl Environ Microbiol***2006**, *72*, 4172.

101. Perlova, O.; Fu, J.; Kuhlmann, S.; Krug, D.; Stewart, A. F.; Zhang, Y.; Müller, R. *Appl Environ Microbiol***2006**, *72*, 7485.

102. Vivien, E.; Pitorre, D.; Cociancich, S.; Pieretti, I.; Gabriel, D. W.; Rott, P. C.; Royer, M. *Antimicrob Agents Chemother***2007**, *51*, 1549.

103. Hopwood, D. A.; Malpartida, F.; Kieser, H. M.; Ikeda, H.; Duncan, J.; Fujii, I.; Rudd, B. A.; Floss, H. G.; Omura, S. *Nature***1985**, *314*, 642.

104. Omura, S.; Ikeda, H.; Malpartida, F.; Kieser, H. M.; Hopwood, D. A. *Antimicrob Agents Chemother***1986**, *29*, 13.

105. Sherman, D. H.; Malpartida, F.; Bibb, M. J.; Kieser, H. M.; Hopwood, D. A. *EMBO J***1989**, *8*, 2717.

106. Ziermann, R.; Betlach, M. C. *Biotechniques***1999**, *26*, 106.

107. Kao, C. M.; Katz, L.; Khosla, C. *Science***1994**, *265*, 509.

Chapter 2: Heterologous Expression of the Oxytetracycline Biosynthetic Pathway in *Myxococcus xanthus*

2.1 Introduction

Microbial natural products are excellent lead compounds for drug discovery and have played major roles in the development of pharmaceutical agents in nearly all therapeutic areas.[1-3] Unfortunately, the rate of discovery of new bacterial natural products has slowed, due in part to frequent rediscovery of known compounds.[2] An enormous and currently inaccessible reservoir of new natural products is encoded in the biosynthetic pathways found in the genomes of uncultivated bacteria.[4] Heterologous expression of these biosynthetic gene clusters represents a powerful tool for discovering new natural products.[5,6] Herein we demonstrate that the δ-proteobacteria *Myxococcus xanthus* is an effective host for heterologous expression of aromatic polyketide biosynthetic pathways. This work expands the scope of polyketide biosynthetic pathways which can be heterologously expressed in *M. xanthus* and suggests that *M. xanthus* may be a suitable general host for heterologous expression.

Molecular phylogenic studies have shown that bacterial diversity is enormous and the vast majority of the diversity is found in uncultivated bacterial species.[4] Estimates suggest that 99 % of soil bacteria from the environment are uncultivatable using standard techniques.[7-9] Culture-independent analysis of metagenomic DNA (mDNA) libraries from soil and marine environments indicate that there is a wealth of natural product diversity encoded in these uncultivated strains. For example, analysis of a soil metagenome for a highly conserved region of polyketide synthase genes showed that

none of sequences found were present in the known public databases.[10] Polyketide synthases (PKS) are key enzymes responsible for the production of the polyketide family of natural products in proteobacteria, actinobacteria and "low G + C gram positive bacteria".[11-13] Polyketide natural products have been developed into antibiotic, anticancer and immunosuppressant clinical agents.[1-3] Based on these observations, metagenomic DNA libraries are expected to possess a large number of new polyketide biosynthetic pathways, representing substantial new chemical diversity for drug discovery.

The utilization of heterologous expression towards the screening of metagenomic DNA (mDNA) libraries for the presence of novel PKS pathways is an attractive prospect. Heterologous production of polyketides in hosts such as *Streptomyces coelicolor* and *Streptomyces lividans* is an important tool in the identification and characterization of PKS pathways.[14-16] Results from these studies have shown that *Streptomyces* strains are good hosts for the heterologous production of many actinomycetal polyketides. However, *streptomyces* strains have proved to be poor hosts for expression of δ-proteobacterial PKS pathways, such as those found in *myxobacteria* and require the introduction of heterologous promoters to ensure transcription of the pathways.[16,17] As polyketide biosynthetic pathways in mDNA libraries contain both actinomycete-and δ-proteobactrium derived pathways along with pathways from other polyketide producing bacterium, a heterologous expression host competent to express pathways of varying origins without relying on the manual introduction of non-native promoters is needed.

We examined the ability of the δ-proteobacteria *Myxococcus xanthus* to act as a general heterologous expression host. *M. xanthus* is a predatory bacteria, which undergoes multicellular development in response to nutrient starvation. During

development *M. xanthus* is known to be an effective host for the heterologous expression of the δ-proteobacterial derived epothilone D biosynthetic pathway and has been used for the production of the epothilone D for clinical trial.[16] We demonstrate that *M. xanthus* can also heterologously express the *Streptomyces rimosus* oxytetracycline biosynthetic pathways producing oxytetracycline. This is the first example of a polyketide from a non-myxobacterial species heterologously expressed in *myxobacteria*. This also presents *M. xanthus* as the first heterologous host capable of producing polyketides from the genetic material of another family of bacterium without the engineering of non-native promoters to ensure transcription.

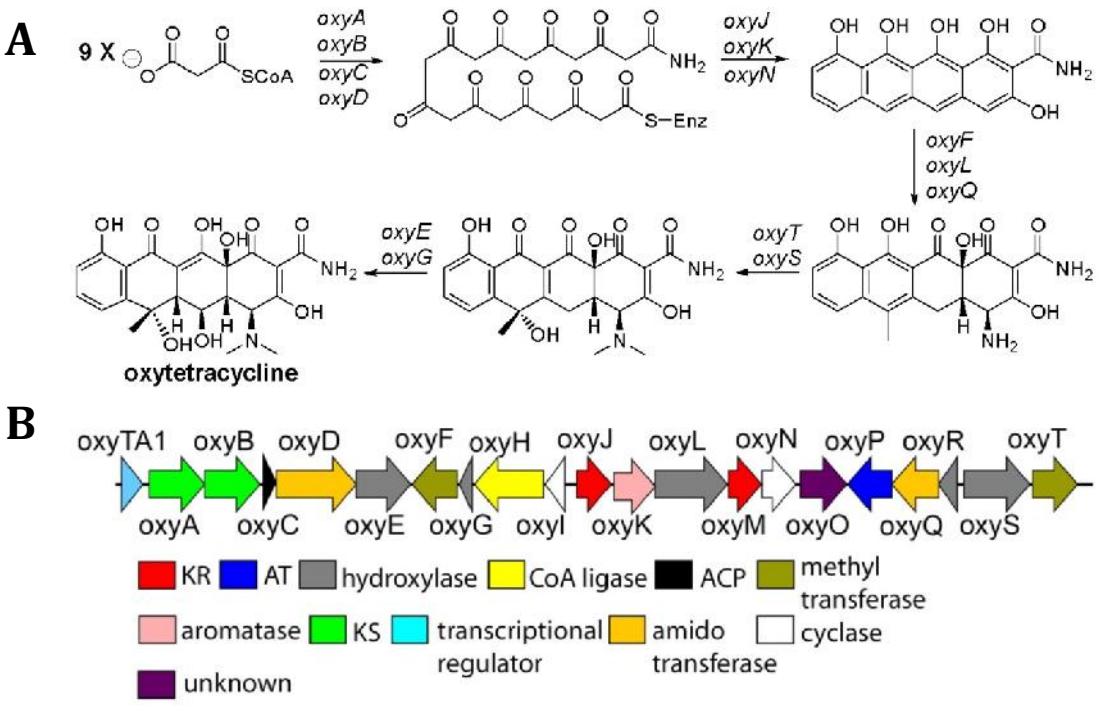


Figure 2.1.1:Oxytetracycline biosynthetic pathway. A) The enzymatic pathway responsible for formation of oxytetracycline is shown above. B) The oxytetracycline biosynthesis gene cluster from *S. rimosus* is shown below.

2.2 Results

To generate a *M. xanthus* strain capable of heterologously expressing oxytetracycline, the *S. rimosus* oxytetracycline biosynthetic pathway (Figure 2.1.1) was inserted via homologous recombination into the *asgE* locus of *M. xanthus*. The a*sgE* locus from *M. xanthus* was amplified and inserted into the *Bgl*II site of pET28b to produce pMRH02. The *asgE* locus provides a chromosomal region for heterologous recombination onto *M. xanthus* chromosome. The 32 kb oxytetracycline pathway from *S. rimosus* was amplified from pYT264 and cloned into the *Eco*RI site of pMRH02 to produce pMRH08.[18] *M. xanthus* DK1622 was electroporated with pMRH08 to provide a *M. xanthus*Δ*asgE* kan^R mutant. This large genomic insertion significantly increased doubling time for this strain (approx. 10hrs). The *asgE* locus has been reported to be involved in *M. xanthus* morphological shift from exponential growth to swarming fruiting body formation.[19] Growth assays to determine defects in morphology in the *M. xanthus* Δ*asgE* kan^R mutant were performed. The *M. xanthus* Δ*asgE* kan^R mutant was observed to have a defect in its ability to form fruiting bodies further confirming successful insertion of the oxytetracycline biosynthetic cluster into the *asgE* locus (Figure 2.5.2-2.5.4).

Oxytetracycline was heterologously produced in *M. xanthus* under standard rich media culture conditions and detected in culture broths by liquid chromatography-mass spectrometry (LC-MS). A liquid culture of the mutant strain containing the oxytetracycline gene cluster was cultured for 10 days at 33 °C in CTTYE (100 mL). Acetone (10 % v/v) was added to the culture, which was then vigorously mixed. The resulting mixture was extracted with three volumes of ethyl acetate to remove the organic

soluble materials, including oxytetracycline. The organic extracts were concentrated *in vacuo* and resuspended in methanol (100 μL). LC-MS analysis using reverse phase C18 stationary phase with a linear acetonitrile/water mobile phase gradient followed by electrospray positive ionization and analysis with a single quadrupole mass spectrometer indicated that oxytetracycline was present in the fermentation (Figure 2.2.1). The titer of oxytetracycline was determined to be approximately 10 mg per liter of fermentation broth. Quantification was performed in triplicate by LC-MS analysis using a standard curve generated from commercial oxytetracycline. Negative controls of *M. xanthus* DK1622 cultures processed under identical conditions did not contain detectable levels of oxytetracycline.

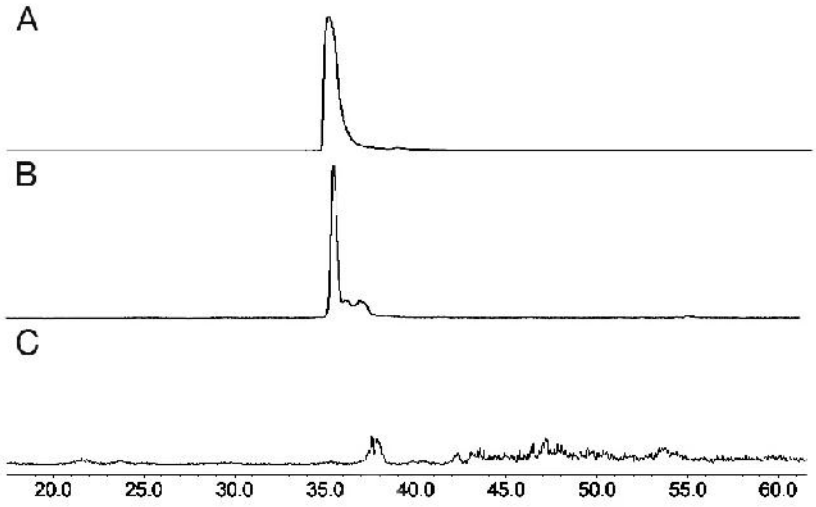


Figure 2.2.1:LC-MS ion extraction analysis, [M+H], of standard and culture extracts. (A) Oxytetracycline standard. (B) *M. xanthus*ΔasgE kan^R mutant containing the oxytetracycline biosynthetic pathway. (C) Wild-type *M. xanthus* DK1622.

2.3 Discussion

These data indicate that *M. xanthus* can heterologously express the oxytetracycline polyketide synthase biosynthetic pathway from *S. rimosus*. Several factors affect the successful heterologous production of polyketide synthase pathways including codon usage bias, mRNA stability, functionality of regulatory elements, and the presence of all necessary starter and extender units.[20] As codon usage between *M. xanthus* and the genus *streptomyces* is very similar, and *myxobacteria* are known to produce polyketide products requiring a wide diversity of starter and extender units, neither were considered likely to impact the ability of *M. xanthus* to heterologously express streptomycetal biosynthetic pathways. As *streptomyces* strains do not appear to be effective at heterologous expression of myxobacterial biosynthetic pathways, we were concerned that *myxobacteria* and *streptomyces* may possess substantially different regulatory elements. Our data indicates that the regulatory elements present in streptomycetal biosynthetic pathways are sufficient to enable expression of the biosynthetic genes in *M. xanthus*. Interestingly, a time course for oxytetracycline production indicates that maximum productivity correlates with mid to late stationary phase, suggesting a hypothesis where expression of the oxytetracycline gene cluster is regulated in *M. xanthus*. Further work exploring the regulatory elements present in myxobacterial polyketide biosynthetic gene clusters is needed to evaluate this hypothesis.

2.4 Conclusion

This study demonstrates that *M. xanthus* can heterologously express streptomycetal polyketide biosynthetic pathways in addition to myxobacterial polyketide

biosynthetic pathways. This result makes *M. xanthus* an excellent candidate for a "universal" host to facilitate heterologous expression of polyketide biosynthetic pathways derived from environmental samples of mDNA.

2.5 References

1. Berdy, J.*J. Antibiot. (Tokyo)***2005**, *58*, 1-26.
2. Lam, K. S.*Trends Microbiol*. **2007**, *15*, 279-89.
3. Newman, D. J.; Cragg, G. M.*J. Nat. Prod.***2007**, *70*, 461-77.
4. Van Lanen, S. G.; Shen, B.*Curr. Opin. Microbiol.***2006**, *9*, 252-60.
5. Watanabe, K.; Oikawa, H.*Org. Biomol. Chem.***2007**, *5*, 593-602.
6. Wenzel, S. C.; Gross, F.; Zhang, Y.; Fu, J.; Stewart, A. F.; Muller, R.*Chem. Biol.***2005**, *12*, 349-56.
7. Davis, K. E.; Joseph, S. J.; Janssen, P. H.*Appl. Environ.Microbiol.* **2005**, *71*, 826-34.
8. Sait, M.; Hugenholtz, P.; Janssen, P.H. *Environ. Microbiol.***2002**, *4*, 654-66.
9. Staley, J. T.; Konopka, A.*Annu. Rev. Microbiol.***1985**, *39*, 321-46.
10. Ginolhac, A.; Jarrin, C.; Gillet, B.; Robe, P.; Pujic, P.; Tuphile, K.; Bertrand, H.; Vogel, T.M.; Perriere, G.: Simonet, P.; Nalin, R.*Appl. Environ. Microbiol.* **2004**, *70*, 5522-7.
11. Gerth, K.; Pradella, S.; Perlova, O.; Beyer, S.; Müller, R.*J. Biotechnol.***2003**, *106*, 233-253.
12. Petkovic, H.; Cullum, J.; Hranueli, D.; Hunter, I. S.; Peric-Concha, N.; Pigac, J.; Thamchaipenet, A.; Vujaklija, D.; Long, P.F.*Microbiol. Mol. Biol. Rev.***2006**, *70,*704-28.
13. Walsh, C. T.*Science* **2004**, *303*, 1805-10.
14. Kao, C. M.; Katz, L.; Khosla, C.*Science***1994**,*265*, 509-12.

15. Liu, H.; Jiang, H.; Haltli, B.; Kulowski, K.; Muszynska, E.; Feng, X.; Summers, M.; Young, M.; Graziani, E.; Koehn, F.; Carter, G. T.; He, M.*J. Nat. Prod.***2009**,*72*, 389-395.

16. Tang, L.; Shah, S.; Chung, L.; Carney, J.; Katz, L.; Khosla, C.; Julien, B.*Science* **2000**, *287*, 640-2.

17. Park, S. R.; Park, J.W.; Jung, W. S.; Han, A.R.; Ban, Y. H.; Kim, E. J.; Sohng, J. K.; Sim, S. J.; Yoon, Y.J.*Appl. Microbiol. Biotechnol.* **2008**,*81*, 109-17.

18. Zhang, W.; Ames, B. D.; Tsai, S. C.; Tang, Y. *Appl. Environ. Microbiol.***2006**, *72*, 2573-80.

19. Garza, A. G.; Harris, B. Z.; Greenberg, B. M.; Singer, M. *J. Bacteriol.***2000**, *182*, 6622-9.

20. Rodriguez, E.; Menzella, H. G.; Gramajo, H.*Methods Enzymol.***2009**,*459*, 339-65.

2.6Experimental

General. *Escherichia coli* XL1 Blue (Stratagene) and *E. coli* TOP10 (Invitrogen) cells were used for routine cloning and plasmid preparations. *Myxococcus xanthus* DK1633 was used in all heterologous production experiments. *Myxococcus xanthus* strains were grown in CTTYE (1.0% Casitone, 0.5% yeast extract, 10.0 mM Tris-HCl, 1.0 mM KH_2PO_4, and 8.0 mM $MgSO_4$).*M. xanthus* morphological assays were performed on 0.5% agar TPM plates for S-motility analysis and 1.5% agar TPM plates for A-motility.*E. coli* cells were grown in LB media (Fisher) supplemented with the appropriate antibiotics when necessary. Oxytetracyline standard and IPTG were purchased from Sigma-Aldrich. Antibiotics and media components were purchased from Fisher. All strains were chemically or electronically transformed using standard transformation protocols.

Plasmid Construction. Genomic DNA prepared from *M. xanthus*DK162was prepared with a Wizard gDNA isolation kit (Promega) and used as template for amplification of *asgE*. Pfu Turbo HS polymerase (Agilent) and the Thermocycler Mastercycler personal (Eppendorf) were used for all PCR cloning reactions. The primers used for the amplification of the *asgE* locus were 5' – GACG*AGATCT*GTTGGAAGGTCGGCAACTGG – 3' and 5' – CTTA*AGATCT*TCCGTGAAGTACTGGCGCAC – 3'. The *asgE* locus was cloned into pCR-Blunt (Invitrogen) following the manufacturer's instructions. Cloned genes were confirmed by sequencing. Genes were subsequently sub-cloned using the introduced *Bgl*II site into pET28b (kanamycin resistance marker) to produce pMRH02 using standard cloning techniques. The oxytetracycline biosynthetic pathway was excised from

pYT264, provided by Prof. Yi Tang, using bookending *Eco*RI sites and introduced into pMRH02.

Oxytetracycline Production, Isolation, Characterization. Strains capable of producing oxytetracycline were prepared by electroporating *M. xanthus* DK1622 with pMRH08 using standard techniques. Production strains were grown in 100 ml CTTYE medium in shake flasks at 33 °C for 10 days. Following this, cells were treated with 10% acetone (v/v) and vigorous vortexing. Cell debris was removed by centrifugation and the clarified lysate was extracted with three volumes of ethyl acetate to remove the organic soluble materials, including oxytetracycline, and concentrated *in vacuo*.Extracts were then resuspended in 100 µl MeOH for LC-MS analysis. All LC-MS analysis collected on an Shimadzu 2010Asingle quadrupole mass spectrometer (Shimadzu) equipped with a turbo-ion spray ESI probe interfaced with a Prominence UFLC (Shimadzu). A reverse phase BDS Hypersil C18 column (dimensions 100 mm x 2.1 mm I.D., particle size 3 µm, Thermo Scientific) was employed. Mobile phases throughout experimental were: mobile phase A: 5% MeCN:95% H_2O with 0.05% formic acid and mobile phase B: 95% MeCN:5% H_2O with 0.05% formic acid. For initial determination of oxytetracycline production a gradient program (5 min mobile phase A hold, 60 min ramp to mobile phase B, 10 min mobile phase B hold, 5 min ramp to mobile phase A, 10 min mobile phase A hold) was developed with a flow rate of 0.20 ml/min into the mass spectrometer (Ion spray 5,500 V, mass range 250-550 m/z). Quantification was performed in triplicate by LC-MS/MS MRM analysis using a standard curve generated from commercial oxytetracycline (Aldrich). Other settings, including gas flow, lens and quadrupole

voltages and parameters for mass resolution of the separating quadrupole were used as obtained during routine optimization of the instrument.

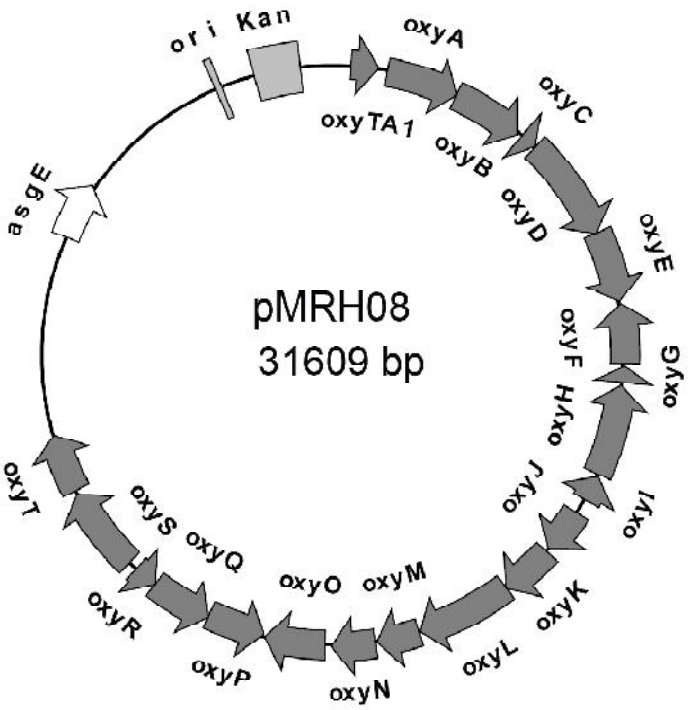


Figure 2.5.1: pMRH08 plasmid map

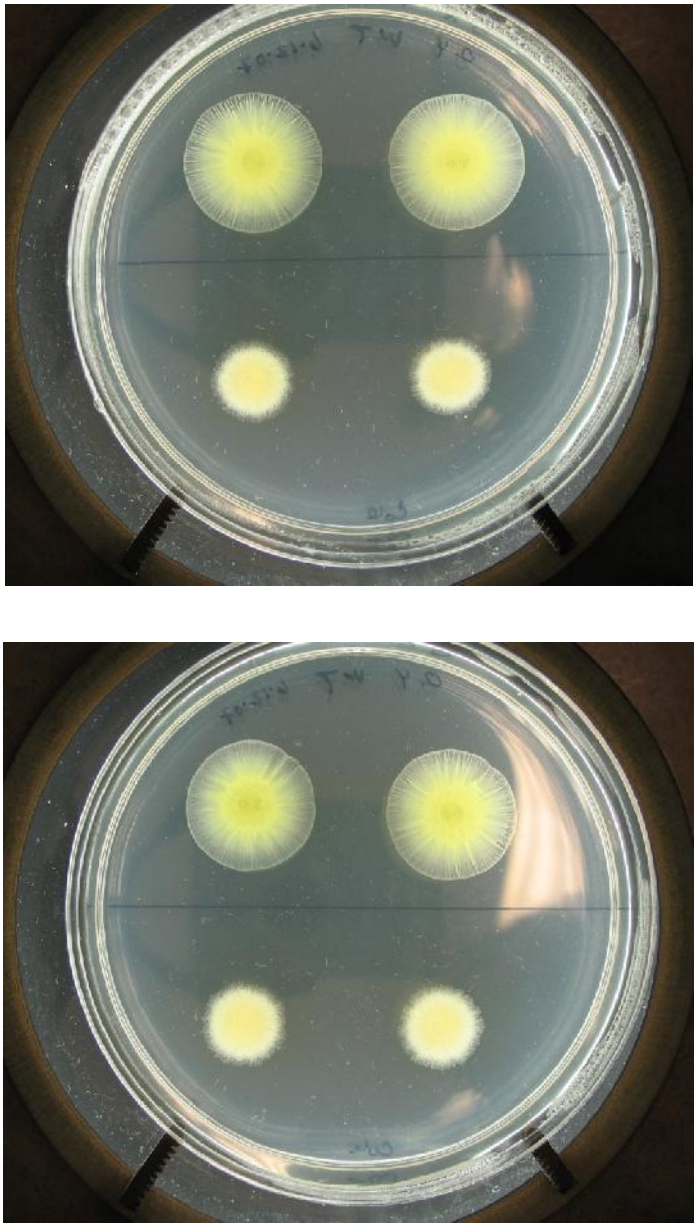


Figure 2.5.2: S-motility assay on *M. xanthus DK1622* (WT, top) and *M. xanthus* pMRH08 mutant (bottom). *M. xanthus* pMRH08 mutant swarm diameter is 61% of the WT diameter confirming a S-motility defect. Replicate data shown.

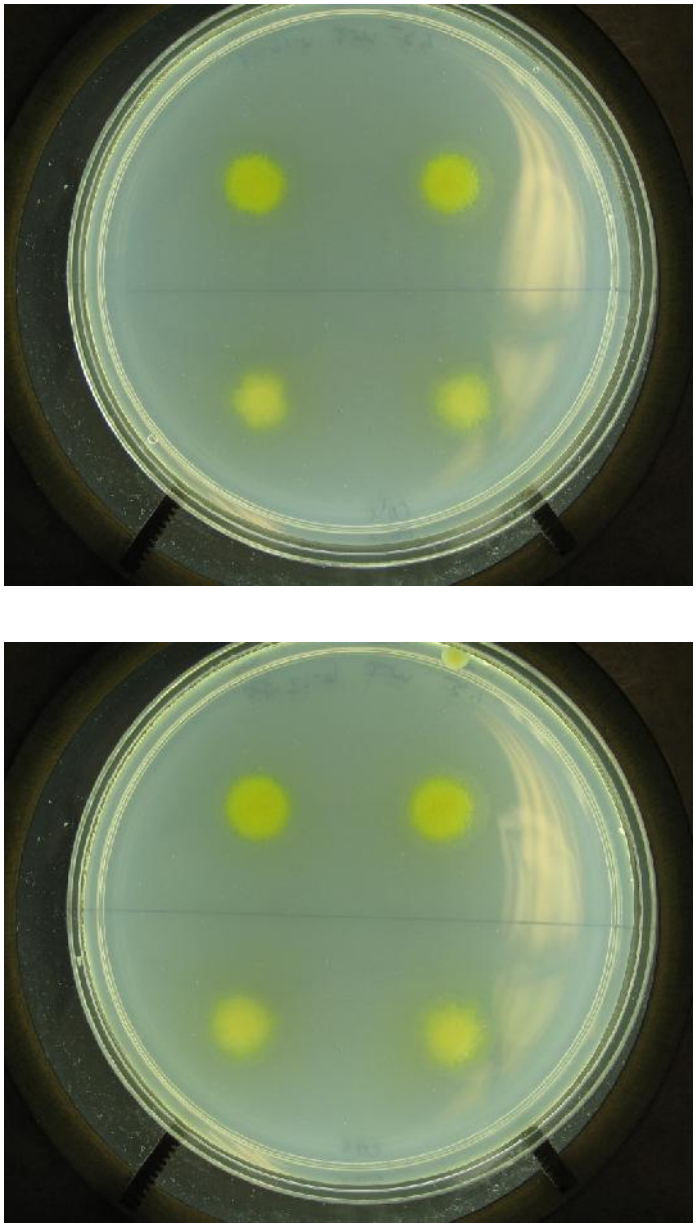


Figure 2.5.3: A-motility assay on *M. xanthus DK1622* (WT, top) and *M. xanthus* pMRH08 mutant (bottom). *M. xanthus* pMRH08 mutant swarm diameter is 93% of the WT diameter confirming there is not an A-motility defect. Replicate data shown.

Chapter 3: Alternative sigma factor over-expression enables heterologous expression of a type II polyketide biosynthetic pathway in *Escherichia coli*

3.1 Introduction

Bacterial polyketides possess an enormous range of chemical diversity and biological function. Many polyketides like tetracycline, epothilone, and rapamycin have been developed into key clinical pharmaceuticals in a broad range of therapeutic areas.[1-4] Sequencing of bacterial genomes, especially those of major polyketide producers such as *actinomycetes* and δ-*proteobacteria*, have shown that there are many more polyketide biosynthetic pathways than there are polyketides isolated from standard cultivation techniques.[5-7] These genetically encoded polyketide natural products from cultivatable and uncultivatable bacteria represent one of the greatest remaining untapped reservoirs of new natural product diversity. Methods to effectively access this diversity will have a major impact on drug discovery.[8]

To access this untapped diversity of polyketide products, a general method for heterologous expression of these pathways is needed. The selection of a heterologous host is contingent upon its ability to efficiently transcribe inserted pathways, translate the often GC rich transcripts, and possess all the starter and extender units necessary for polyketide production.[9] Hosts highly related to the native polyketide producing organism often meet these requirements. Host such as *Streptomyces coelicolor*, *Streptomyces lividans* and *Myxococcus xanthus*, have proved successful for heterologous production of a number of different polyketide products.[10-12] Screening of multiple hosts related to the producing organism is often necessary to identify a successful heterologous host.[13,14] In

addition, various culture conditions also need to be screened to identify conditions that produce the desired compounds.[15] The lack of a general heterologous expression system has made heterologous expression incompatible with screening bacterial genomic DNA libraries for new polyketide products.

Escherichia coli has a number of advantages that make it an appealing host for a heterologous expression system. The ease of genetic manipulation and culturing as compared to other heterologous hosts makes *E. coli* far more useful for bioprospecting from genomic DNA libraries. Codon usage as well as starter and extender unit availability have proven not to be obstacles for heterologous production in *E. coli*.[16-18] The principle impediment to the use of *E. coli* as a heterologous host is its inability to effectively transcribe heterologous pathways. In all examples of heterologous expression of polyketides in *E. coli*, all biosynthetic genes have been placed under the control of the T7 promoter.[16,17] The native promoters present in these heterologous pathways are not sufficient to ensure expression of all the genes in a given pathway under standard *E. coli* culture conditions.

To convert *E. coli* into a general heterologous expression system, a mechanism for ensuring transcription of all the biosynthetic genes in a foreign pathway is required. We hypothesized that a general transcriptional regulator of polyketide biosynthesis should be present in bacteria because horizontal transfer of biosynthetic pathways is a key mechanism for their proliferation among species.[19] Strong evidence for substantial horizontal transfer of pathways can be seen in the large number of diverse *Streptomyces* strains which produce the tetracycline family of antibiotics, the presence of yersiniabactin-like gene clusters in diverse *Actinomycetes*, and the presence of highly

related pederin, onnamide A and psymberin gene clusters in unidentified bacterial strains from the beetle *Paederus fuscipes* and the sponges *Theonella swinhoei* and *Psammocinia* aff. *bulbosa* respectively.[6,20] Because polyketide production is positively regulated in most bacteria by stress, we further hypothesized that alternative sigma factors downstream of the stringent response should positively regulate this general transcriptional mechanism.[21]

Herein we show that this fundamentally new mechanism for heterologous expression of polyketide biosynthetic pathways, where high-level pleoitropic sigma factors from the heterologous host are used to positively regulate transcription of the non-native biosynthetic gene clusters, is highly effective. We demonstrate that one of the six alternative sigma factors in *E. coli* can selectively, directly, and positively regulate heterologous expression from the *Streptomyces rimosus* oxytetracycline biosynthetic gene cluster in *E. coli*.[1] This is the first successful heterologous production of an aromatic polyketide in *E. coli*, demonstrating the utility of this method.[22]

3.2 Results

3.2.1 Standard *E. coli* culture conditions do not enable heterologous expression

To test our hypothesis that an alternative sigma factor downstream of the stringent response could positively regulate transcription of polyketide biosynthetic pathways in *E. coli*, we investigated the ability of *E. coli* to heterologously express the oxytetracycline biosynthetic pathway from *Streptomyces rimosus*[1]. Oxytetracycline is produced by a 32-kb type II polyketide synthase gene cluster consisting of 21 genes. Thirteen of these genes encode proteins required for the biosynthesis of oxytetracycline and thus must be transcribed for successful heterologous production of oxytetracycline.[23] The small

intergenic spaces between open reading frames and the changes in direction of transcription in the oxytetracycline gene cluster suggest that the entire pathway is expressed as, at minimum, five putative operons with at least one biosynthetic gene on each transcript (Figure 3.2.1.1a).

To determine if *E. coli* could produce oxytetracycline under standard culture conditions, *E. coli* BAP1 was transformed with the *S. rimosus* oxytetracycline pathway and cultured in rich liquid media.[12,16] BAP1 was used because it possesses a chromosomal insertion of the phosphopantetheinyl transferase from *Bacillus subtillis*, *sfp*, which ensures that polyketide synthase acyl carrier proteins (ACPs) are post-translationally modified.[16] LC-MS/MS analysis of the organic extracts from the culture media did not show any detectable oxytetracycline in the culture broth (Figure 3.2.2.1b). LC-MS/MS analysis was performed using multiple reaction monitoring (MRM) mode with a lower limit of detection of 100 μg/L. These results indicate that *E. coli*, under standard culture conditions, cannot heterologously produce oxytetracycline from the native *S. rimosus* gene cluster.

To determine if transcription of the genes in the oxytetracycline gene cluster was limiting production of oxytetracycline in *E. coli*, we quantified mRNA levels using quantitative PCR (qPCR) for five genes in the pathway. The five genes *oxyB*, *oxyF*, *oxyK*, *oxyP*, and *oxyT,* one per putative operon, were selected. *oxyB* encodes the β-ketosynthase, *oxyF* and *oxyT* encode methyltransferases and *oxyK* encodes an aromatase all of which are required for oxytetracycline biosynthesis.[23] *oxyP* encodes a malonyl-ACP transferase, which may not be required for biosynthesis of oxytetracycline.[23] Transcripts for four of the genes, *oxyF*, *oxyK*, *oxyP* and *oxyT* could be easily detected and

quantified (Fig. 3.2.1.1c). Transcript levels for *oxyB* were indistinguishable from negative control samples that lacked the oxytetracycline gene cluster, indicating that *oxyB* was not transcribed (Fig. 3.2.1.1b). As OxyB is required for generation of the polyketide backbone of oxytetracycline, the lack of transcription of this gene is sufficient to account for the inability of *E. coli* to produce oxytetracycline under standard culture conditions. These results, in conjunction with published work, confirm that transcription of heterologous polyketide pathways in *E. coli* limits the ability of the heterologous host to produce polyketide products when using the native promoters.[16,17]

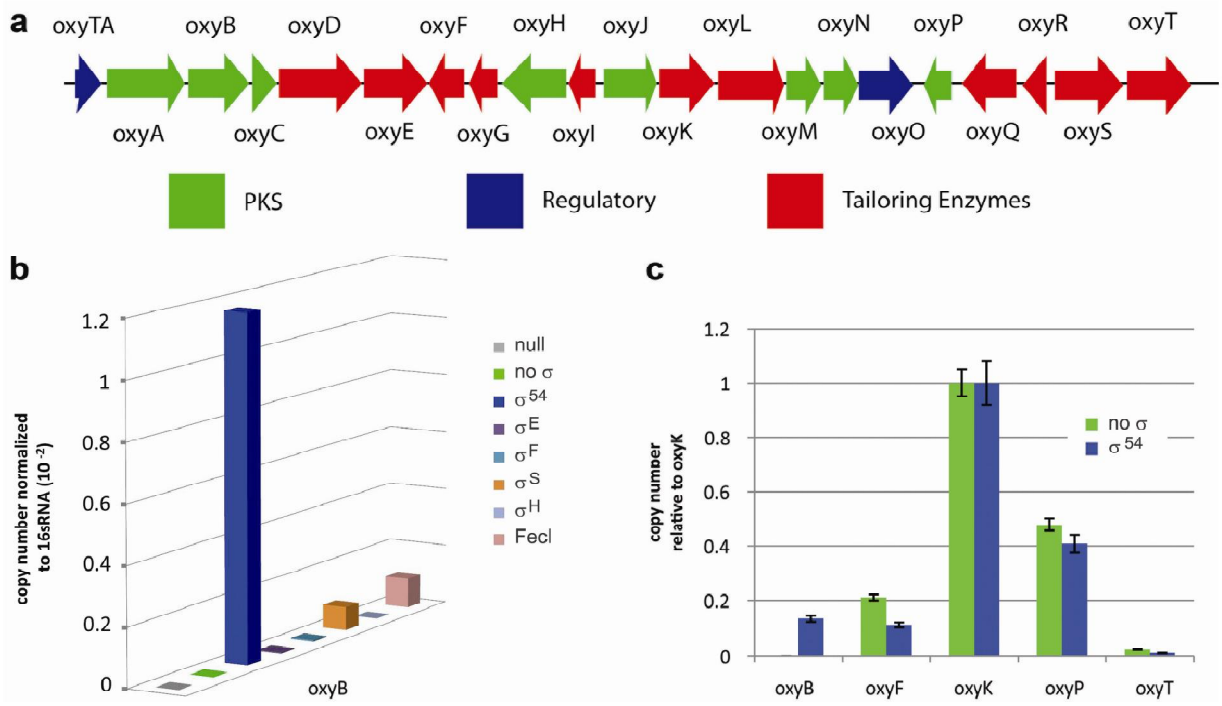


Figure 3.2.1.1: SYBR-green based qPCR analysis shows that transcription limits heterologous production of oxytetracycline in *E. coli.* (**a**) The 32 kb oxytetracycline biosynthetic gene cluster is shown. Five putative operons, *oxyABCDE*, *oxyIHGF*, *oxyJKLMNO*, *oxyRQP*, and *oxyST* are predicted for this gene cluster (**b**) qPCR analysis shows that over-expression of the alternative sigma factors σ^{54}, σ^{S} and FecI enable detectable levels of the *oxyB* transcript to be produced. Over-expression of no sigma factor,

σ^E, σ^F and σ^H do not lead to detectable levels of the *oxyB* transcript. (**c**) qPCR analysis shows that over-expression of the alternative sigma factor σ^{54} lead to detectable levels of transcripts for all five putative operons in the oxytetracycline biosynthetic pathway. In the absence of σ^{54} over-expression, the *oxyB* transcript cannot be detected.

3.2.2 Over-expression of alternative sigma factors σ^{54}, σ^S, and FecI positively regulate transcription of *oxyB* in *E. coli*.

To test our hypothesis that an alternative sigma factor downstream of the stringent response could positively regulate transcription of polyketide biosynthetic pathways in *E. coli*, we investigated the impact of alternative sigma factor over-expression on transcription from the oxytetracycline gene cluster. In their native hosts, polyketide biosynthetic gene clusters are principally regulated by one or more pathway specific regulators, which can respond to a diversity of stimuli.[21,25] The most general observed positive regulator of polyketide biosynthesis is nutrient limitation, which initiates the stringent response in bacteria. During stringent response the levels of the pleoitropic regulators phosphorylated guanosine nucleotides, ppGpp and pppGpp, increase, leading to an increase in alternative sigma factor-mediated transcription.[21,25] As increases in (p)ppGpp has been shown to correlate with polyketide production in diverse bacterial species, we hypothesized that this positive regulation could be mediated by alternative sigma factors.[26,27]

Sigma factors control the specificity of gene transcription by binding to RNA polymerase (RNAP) and recruiting it to promoter sequences upstream of the gene to be

transcribed.[28] In addition to the household sigma factor σ^{70}, *E. coli* has six alternative sigma factors, σ^{54} which is involved in nitrogen assimilation, σ^{H} which controls heat shock promoters, σ^{S} which controls stationary phase promoters, σ^{F} which controls flagellum related genes, σ^{E} controls response to extracytoplasmic stress, and FecI which is involved in iron transport.[28] Over-expression of sigma factors has been shown to increase expression of genes in their regulon in many bacteria, including *E. coli.*[29-31] If an alternative sigma factor is a positive regulator of polyketide biosynthesis, over-expression of that sigma factor will increase transcription from genes in the oxytetracycline gene cluster.

To determine if σ^{54}, σ^{H}, σ^{S}, σ^{F}, σ^{E}, and FecI could positively regulate polyketide biosynthesis in *E. coli,* we quantified mRNA levels using quantitative PCR for *oxyB*. Six *E. coli* BAP1 strains possessing the oxytetracycline biosynthetic gene cluster and over-expressing either σ^{54}, σ^{H}, σ^{S}, σ^{F}, σ^{E}, or FecI were investigated. No *oxyB* transcripts could be detected when σ^{H}, σ^{F}, and σ^{E} were over-expressed. Over-expression of σ^{S} and FecI led to detectable but low levels of *oxyB* transcription. When σ^{54} was over-expressed substantial levels of *oxyB* transcripts could be detected (Fig 3.2.1.1b). Quantification of mRNA levels for *oxyB*, *oxyF*, *oxyK*, *oxyP* and *oxyT* in *E. coli* BAP1 possessing the oxytetracycline biosynthetic gene cluster and over-expressing *E. coli* σ^{54} showed that we were able to detect all five transcripts (Fig 3.2.1.1c). This data demonstrates that all of the putative transcripts required for oxytetracycline are present when σ^{54} is over-expressed.

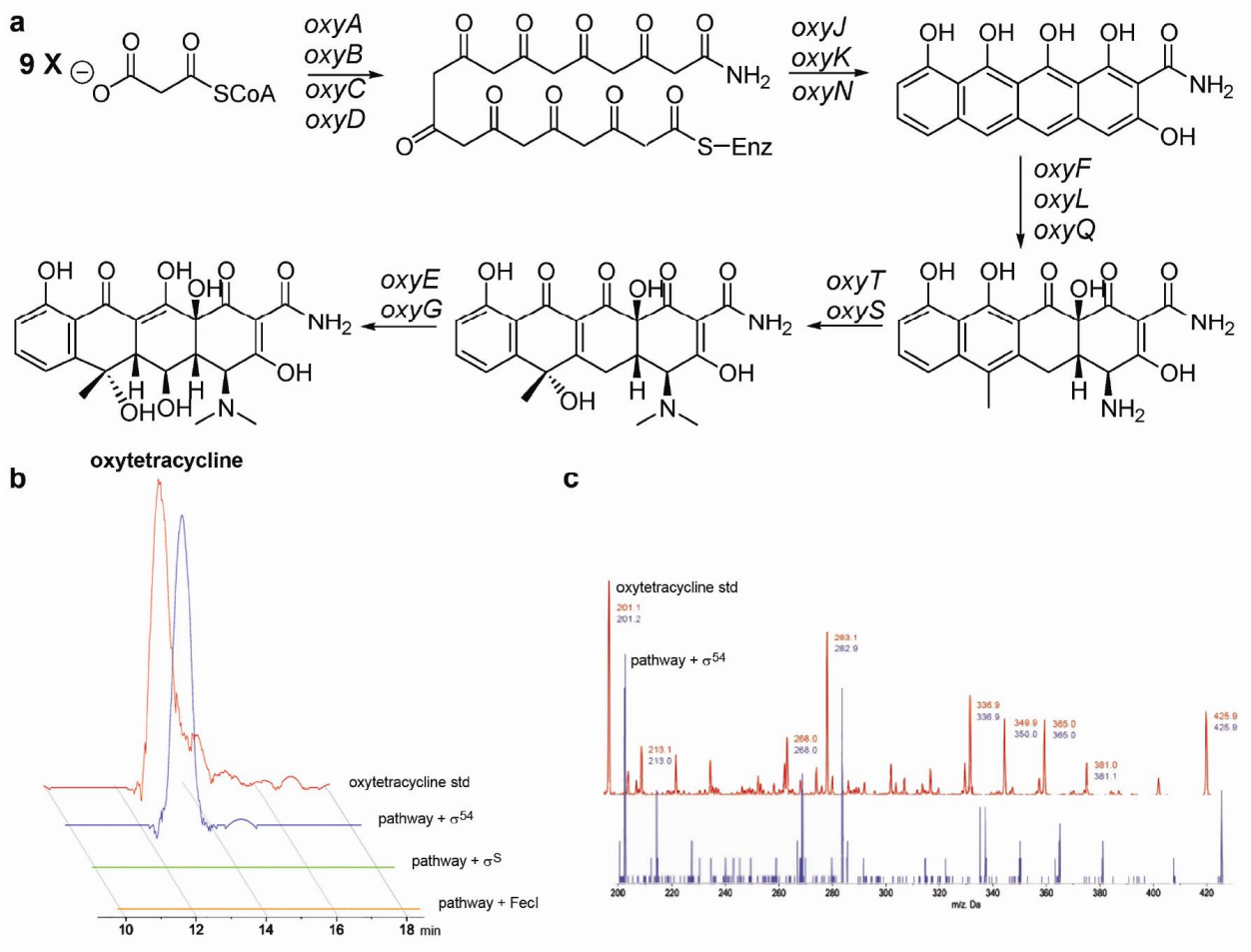


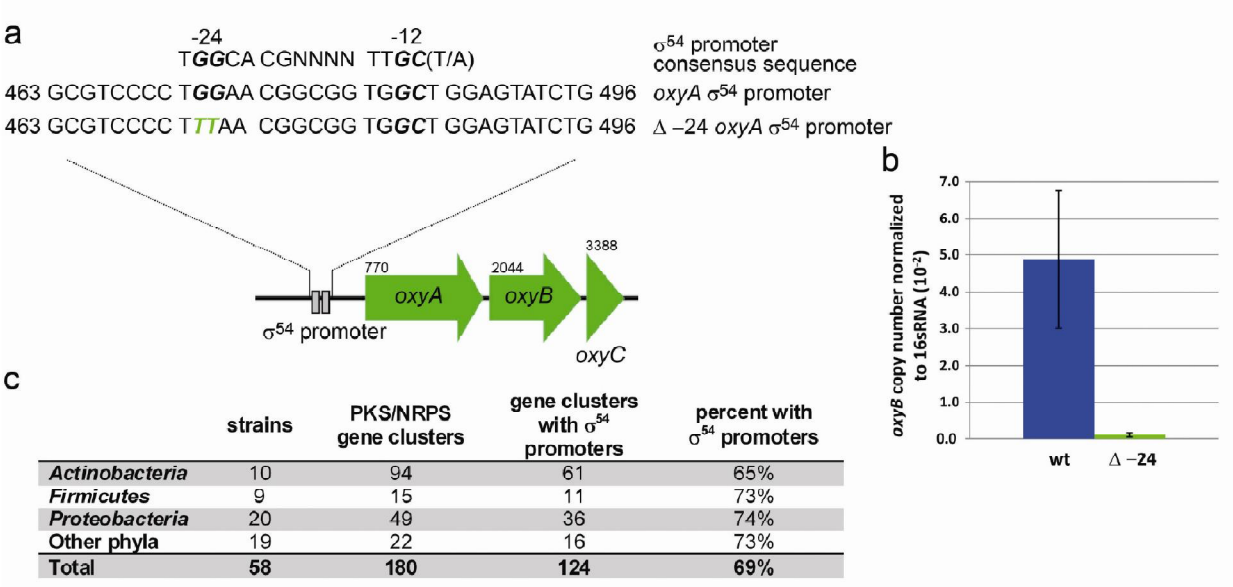
Figure 3.2.2.1: Over-expression of σ^{54} enables *E. coli* to heterologously produce oxytetracycline. (**a**) The enzymatic pathway responsible for the biosynthesis of oxytetracycline. (**b**). ESI-LC-MS/MS analysis of an oxytetracycline standard and the organic extracts from *E. coli* cultures containing the oxytetracycline gene cluster and over-expressing σ^{54}, σ^{S}, or FecI. These traces show the ion extraction data from the Q3 scan of MS/MS experiments. Q1 was used to select the $[MH]^+$ ion for oxytetracycline (m/z = 461). Ion extractions were performed for the oxytetracycline fragment m/z = 283 from the Q3 scan. Signals with a peak width of less than 0.1 s were regarded as noise and removed with the noise filter application. (**c**). The MS^2 spectrum of the m/z = 461 peak for the oxytetracycline authentic standard and heterologously produced oxytetracycline.

3.2.3 Over-expression of σ^{54} enables oxytetracycline production in *E. coli*

If transcription limits heterologous production of polyketides in *E. coli*, then over-expression of σ^{54} in the presence of the oxytetracycline gene cluster is predicted to lead to the production of oxytetracycline. To test this hypothesis we over-expressed σ^{54} in *E. coli* BAP1 containing the oxytetracycline gene cluster and analyzed the organic extracts with LC-MS/MS to determine if oxytetracycline was detectable. LC-MS/MS analysis using the MS^2 scan mode clearly identified the presence of oxytetracycline (Figure 3.2.2.1b). The LC retention time and MS^2 spectrum of our heterologously produced oxytetracycline was indistinguishable from authentic oxytetracycline standards (Figure 3.2.2.1c). Using MRM mode and a standard curve generated from authentic oxytetracycline, the titer of heterologously expressed oxytetracycline was determined to be 2.0 ± 0.1 mg per L of culture broth. These results demonstrate that when σ^{54} is over-expressed in *E. coli* containing the oxytetracycline gene cluster, all the biosynthetic genes are transcribed and translated, all the necessary proteins are functional *in vivo*, and the required starter and extender units are available, enabling oxytetracycline production.

Because over-expression of σ^S and FecI also produced detectable *oxyB* transcripts, it is possible that these alternative sigma factors may also enable heterologous production of oxytetracycline. We therefore investigated the ability of *E. coli* BAP1 cultures over-expressing σ^S and FecI to heterologously produce oxytetracycline. LC-MS/MS analysis showed that no detectable oxytetracycline was present in these culture broths (Figure 3.2.2.1b). As expected from our transcription data, over-expression of σ^H, σ^F, and σ^E did not led to heterologous production of detectable levels of oxytetracycline.

These data demonstrate that only the alternative sigma factor, σ^{54}, can positively regulate transcription of the oxytetracycline biosynthetic pathway, leading to heterologous production of oxytetracycline in *E. coli*.



	strains	PKS/NRPS gene clusters	gene clusters with σ^{54} promoters	percent with σ^{54} promoters
Actinobacteria	10	94	61	65%
Firmicutes	9	15	11	73%
Proteobacteria	20	49	36	74%
Other phyla	19	22	16	73%
Total	**58**	**180**	**124**	**69%**

Figure 3.2.3.1: Putative σ^{54} promoters are found in the majority of polyketide and non-ribosomal peptide biosynthetic gene clusters. (**a**) The *oxyABCDEF* operon from the *S. rimosus* oxytetracycline biosynthetic pathways contains a putative σ^{54} promoter. The promoter shows high homology to the σ^{54} promoter consensus sequence especially at the key -12 and -24 positions. This promoter could be responsible for direct, positive transcriptional regulation of the *oxyB* gene by σ^{54} over-expression. (**b**) Transcription of the *oxyABCDEF* operon is directly controlled by the σ^{54} promoter upstream of *oxyA*. Mutation of the highly conserved GG residues -24 from the transcriptional start site to TT leads to a greater than 40 fold decrease in *oxyB* transcript levels as determined by qPCR. (**c**) Results of a bioinformatics analysis of 58 bacterial genomes containing 180 polyketide synthase (PKS) and non-ribosomal peptide synthetase (NRPS) biosynthetic gene clusters show that the majority of gene clusters contain putative σ^{54} promoters. It is particularly intriguing that *Actinobacteria* possess gene clusters with σ^{54} promoters since they lack the gene encoding σ^{54}.

3.2.4 σ^{54} directly regulates heterologous production in *E. coli*.

σ^{54} can positively regulate transcription directly or indirectly. For direct regulation, the σ^{54}-RNAP complex binds to a σ^{54} consensus promoter upstream of the operon under σ^{54} control. The σ^{54} consensus promoter is unique among bacterial promoters with highly conserved residues –12 and –24 from the transcriptional start site (T*GG*CACG-N4-TT*GC*(T/A)) and can reliably be identified from sequence data.[32] If σ^{54} is directly regulating transcription of *oxyB*, a σ^{54} promoter consensus sequence must be present upstream of the operon containing *oxyB*. To determine if σ^{54} promoters were present in the oxytetracycline biosynthetic pathway the entire gene cluster was examined for σ^{54} promoter consensus sequences using the web-based tool PromScan (http://molbiol-tools.ca/promscan/).[33] A high scoring consensus promoter was located upstream of the operon containing *oxyB* in a location that would allow direct transcription of the putative operon (Fig. 3.2.3.1a). The location and consensus sequence of the identified promoter provides strong evidence for direct regulation by σ^{54} of the heterologous production of oxytetracycline in *E. coli*.

To confirm that the predicted σ^{54} promoter is functional and involved in the transcriptional regulation of *oxyB*, the highly conserved GG –24 from the transcriptional start site was mutated to TT. The –24 region plays a major role in anchoring the σ^{54}-RNAP complex to the promoter and mutation of the conserved GG is known to decrease transcription of genes under σ^{54} control by ten to one hundred-fold.[34-35] qPCR analysis showed that *oxyB* transcript levels from the TT mutant σ^{54} promoter were greater than

forty-fold lower as compared to the wild type (Fig. 3.2.3.1b). This data confirms that the predicted promoter is functional and that *oxyB* is directly regulated by σ^{54}.

3.2.5 σ^{54} promoters are common in polyketide and non-ribosomal peptide biosynthetic pathways.

If σ^{54} promoters are commonly found in polyketide and non-ribosomal peptide biosynthetic pathways, σ^{54} over-expression may be a general route to heterologous expression of polyketides and non-ribosomal peptides. To evaluate this possibility, we examined the genomes of 58 sequenced bacteria for polyketide and non-ribosomal peptide biosynthetic gene clusters and σ^{54} promoters. The 58 species included major secondary metabolite producers such as 10 *Actinobacteria*, 20 *Proteobacteria*, 9 *Firmicutes*, as well as 19 species from diverse phyla. This selection of species contains representative examples of most of the sequenced bacterial phylogenic diversity.

Manual examination of all genes annotated as polyketide synthase, 3-oxoacyl acyl carrier protein synthase, and non-ribosomal peptide synthetase (COG3321, COG0304, and COG1020 respectively) and their adjacent genes identified 180 polyketide and non-ribosomal peptide biosynthetic gene clusters (Figure 3.2.3.1c, see supplementary information Table 3.6.6 for all gene clusters identified).[36]

Each bacterial genome was analyzed to identify all putative σ^{54} promoters. To identify σ^{54} promoters, a positional weighted matrix describing the σ^{54} promoter consensus sequence[34] was aligned with all sites on the forward and reverse strands of the genome. The fit at each position was scored using the published algorithm from the bioinformatics tool PromScan. Sequences with a fit scoring 75 or greater and on the same strand, less than 500 bases upstream of a start codon were identified as putative σ^{54}

promoters.[33] Experimentally validated σ^{54} promoters in the *E. coli* genome have scores ranging from the 80 to high 90s.[31] In GC rich organisms such as *M. xanthus* a number of identified σ^{54} promoters have scores in the low to mid 70s.[37,38] 75 is thus a compromise between the low false positives identification, as seen in *E. coli,* and the lower sensitivity needed to avoid false negatives, as seen in some high GC organisms.

Of the 180 polyketide and non-ribosomal peptide biosynthetic gene clusters identified, 124 (69%) contained one or more σ^{54} promoters appropriately positioned to regulate transcription (see supporting information for a list of all the biosynthetic genes with computationally predicted σ^{54} promoters). 24 gene clusters possessed a single σ^{54} promoter, however the vast majority had two or more, with some pathways possessing up to two dozen σ^{54} promoters. These results clearly demonstrate that a majority (69%) of polyketide and non-ribosomal peptide biosynthetic gene clusters possess σ^{54} promoters and that these promoters are appropriately positioned to regulate transcription of at least one operon in these gene clusters. These results suggest that σ^{54} over-expression may be a general method for ensuring transcription of polyketide and non-ribosomal peptide biosynthetic gene clusters in heterologous hosts, such as *E. coli*.

3.3 Discussion

Heterologous expression of polyketide and non-ribosomal peptide biosynthetic gene clusters will play a major role in the discovery of new natural products from metagenomic and environmental DNA samples. Currently no general heterologous expression method appropriate for screening DNA libraries exists. Herein, we describe a fundamentally new mechanism for heterologous expression of a polyketide biosynthetic pathway, where a high-level, pleiotropic alternative sigma factor from the heterologous

host positively regulates transcription of the biosynthetic gene cluster. In contrast, known methods for heterologous expression rely on either replacing each native promoter with known, well-characterized promoter from the heterologous host, such as the T7 promoter in *E. coli*, or rely on the heterologous host to constitutively express each gene from the native promoters. Our approach which actively induces transcription of gene clusters by over-expression of alternative sigma factors may provide a general solution to the heterologous expression problem that is compatible with screening DNA libraries.

Our results demonstrate that over-expression of the sigma factor, σ^{54}, enables efficient heterologous expression of oxytetracycline biosynthetic gene cluster in *E. coli*. Using qPCR, we have demonstrated that no transcript is detectable under standard culture conditions for the key oxytetracycline biosynthetic gene *oxyB*. LC-MS/MS analysis of these cultures shows no detectable oxytetracycline production in the absence of detectable *oxyB* transcripts. Over-expression of σ^{54} positively regulates transcription of *oxyB*, generating detectable levels of *oxyB* transcripts, enabling production of oxytetracycline. Bioinformatics analysis of the oxytetracycline biosynthetic gene cluster shows a σ^{54} promoter sequence appropriately positioned to directly regulate *oxyB* transcription. Loss of function mutagenesis confirms that this promoter is functional. Further analysis of 180 polyketide and non-ribosomal peptide biosynthetic pathways shows that the majority of these pathways possess σ^{54} promoter sequences (69%), suggesting that σ^{54}-mediated heterologous expression may be a general phenomenon.

Substantial evidence exists to support the hypothesis that over-expression of alternative sigma factors, such as σ^{54}, can positively regulate diverse polyketide biosynthetic pathways. Nutrient limitation, which is one of the only known general

regulators of polyketide biosynthesis, initiates the stringent response in bacteria. During stringent response the levels of the global bacterial regulator (p)ppGpp increases.[21] Increases in (p)ppGpp have been shown to correlate with polyketide production in diverse bacterial species.[26,27] For example *relA*, the gene encoding the (p)ppGpp synthetase, in *S. coelicolor* is required to trigger polyketide biosynthesis.[26] (p)ppGpp disrupts the interaction between the RNA polymerase (RNAP) and the principle sigma factor, σ^{70}, enabling RNAP to interact with alternative sigma factors, which increases alternative sigma factor-mediated transcription.[25] Thus nutritional limitation and (p)ppGpp exert their regulatory effects by increasing alternative sigma factor-mediated transcription, suggesting that polyketide biosynthetic pathways are generally under either direct or indirect alternative sigma factor transcriptional control.

We investigated the effect of over-expression of all the alternative sigma factor from *E. coli* on the transcription of *oxyB* and on the production of oxytetracycline. Only σ^{54} over-expression led to both detectable levels of the *oxyB* transcript and detectable levels of oxytetracycline production. σ^{54} is known to play a major role in response to nitrogen limitation in most bacteria and sporulation in *δ-proteobacteria*.[28,39] Interestingly both nitrogen limitation and sporulation correlate with polyketide production, suggesting σ^{54} may play a role in regulation of polyketide biosynthesis in some bacteria.[21,26]

Over-expression of σ^{S} and FecI showed an increase in *oxyB* transcription but did not lead to detectable levels of oxytetracycline production. *oxyB* transcript levels were at least 10 fold lower for σ^{S} and FecI over-expression than when compared to σ^{54} over-expression. Presumably the low levels of the *oxyB* transcript were insufficient to generate enough OxyB to produce detectable levels of oxytetracycline. A plausible

explanation for the transcription of *oxyB* during σ^S and FecI over-expression is cross-talk between these alternative sigma factors and σ^{54}, leading to low levels of σ^{54}-mediated transcription. This is supported by the observation that σ^{54}, σ^S, and FecI are known to be linked by the polyamine response.[40,41]

The identification of a functional σ^{54} promoter in the *S. rimosus* oxytetracycline biosynthetic gene cluster and 236 putative σ^{54} promoters in 61 gene clusters from seven different actinobacterial genomes is highly unexpected. *Actinobacteria* do not contain a gene encoding σ^{54}. Two hypotheses could account for this observation. These promoters are non-functional in *Actinobacteria* and are remnants of horizontal biosynthetic pathway transfer from non-*Actinobacteria* where these σ^{54} promoters are functional. Alternatively, a functional equivalent of σ^{54} could be present. Extensive research efforts into regulation of polyketide biosynthesis in *Streptomyces* has not uncovered a functional equivalent to σ^{54}, however the vast majority of these studies have been carried out in *S. coelicolor*, the only *Streptomyces* in our genome analysis to contain no σ^{54} promoters in any of its 10 polyketide and non-ribosomal peptide biosynthetic gene clusters (see supporting information). Understanding the origins and roles of these σ^{54} promoter sequences is a major new direction in molecular microbiology of bacterial natural product biosynthesis.

This study opens the door to characterizing and engineering polyketide biosynthetic pathways in *E. coli.* The vast majority of genetic experiments, such as deletion or complementation experiments, used to characterize these biosynthetic pathways have been carried out in either native producing organisms or in *Streptomyces* heterologous hosts.[42-46] By developing an *E. coli*-based heterologous expression system,

these experiments can be carried out in the highly genetically malleable *E. coli*, simplifying these studies. The power of this heterologous expression tool can already be seen. This study represents the first example of successfully expressing functional type II polyketide synthases in *E. coli*.[22] This result will greatly facilitate studying the construction of aromatic polyketide backbones as well as the following oxidative tailoring chemistries.

3.4 Conclusion

In summary, we have developed a fundamentally new *E. coli*-based heterologous expression system for polyketide biosynthetic gene clusters. We have demonstrated the over-expression of the alternative sigma factor σ^{54} directly and positively regulates heterologous expression of the oxytetracycline biosynthetic gene cluster in *E. coli*. Bioinformatics analysis indicates that σ^{54} promoters are present in nearly 70% of polyketide and non-ribosomal peptide biosynthetic pathways suggesting that σ^{54}-mediated heterologous expression may be an effective, general approach. If sufficiently general, this approach will facilitate heterologous expression of new polyketides from metagenomic and environmental DNA samples. Lastly, it is not unreasonable to suggest that over-expression of other sigma factors, such as σ^{H}, σ^{S} or FecI, may positively regulate transcription of other biosynthetic gene clusters, potentially expanding the scope of alternative sigma factor-mediated heterologous expression.

3.5 References

1. Zhang, W.; Ames, B.; Tsai, S.; Tang, Y. *Appl. Environ. Microbiol.* **2006**, *72*, 2573-2580.
2. Tang, L.; Shah, S.; Chung, L.; Carney, J.; Katz, L.; Khosla, C.; Julien, B. *Science* **2000**, *287*, 640-642.
3. Schwecke, T.; Aparicio, J.F.; Molnár, I.; König, A.; Khaw, L.E.; Haydock, S.F.; Oliynyk, M.; Caffrey, P.; Cortés, J.; Lester J.B. *Proc. Natl. Acad. Sci. USA* **1995**, *92*, 7839–7843.
4. Newman, D.J.; Cragg, G.M. *J. Nat .Prod.* **2007**, *70*, 461-477.
5. Bode, H.; Müller, R. *J. Ind. Microbiol. Biotechnol.* **2006**, *33*, 577-588.
6. Udwary, D.W.; Zeigler, L.; Asolkar, R.N.; Singan, V.; Lapidus, A.; Fenical, W.; Jensen, P.R.; Moore, B.S. *Proc. Natl. Acad. Sci. USA* **2007**, *104*, 10376-10381.
7. Van Lansen, S.G.; Shen, B. *Curr. Opin. Microbiol.* **2006**, *9*, 252-260.
8. Li, J.W.-H.; Vederas, J.C. *Science* **2009**, *325*, 161-165.
9. Rodriguez, E.; Menzella H.G.; Gramajo H. *Methods Enzymol.* **2009**, *459*, 339-365.
10. Kao, C.M.; Katz, L.; Khosla, C. *Science* **1994**, *265*, 509-512.
11. Binz T.M.; Wenzel, S.C.; Apotheke, H.-J.S.; Bechthold, A.; Müller, R. *ChemBioChem* **2008**, *9*, 447-454.
12. Stevens D.C.; Henry M.R.; Murphy K.A.; Boddy C.N. *Appl. Environ. Microbiol.* **2010**, *76*, 2681-2684.
13. Baltz, R.H. *J. Ind. Microbiol. Biotechnol.* **2010**, *37*, 759-772.
14. Chen, Y.; Wendt-Pienkowski, E.; Shen, B. *J. Bacteriol.* **2008**, *190*, 5587–5596.

15. Gomez-Escribano, J.P.; Bibb, M.J. *Microbial Biotech.* **2011** doi: 10.1111/j.1751-7915.2010.00219.x**.**

16. Pfeifer, B.A.; Admiraal, S.J.; Gramajo, H.; Cane, D.E.; Khosla, C. *Science* **2001**, *291*, 1790-1792.

17. Boddy, C.N.; Hotta, K.; Tse, M.L.; Watts, R.E.; Khosla, C. *J. Am. Chem. Soc.* **2004**, *126*, 7436-7437.

18. Watanabe, K.; Hotta, K.; Praseuth, A.P.; Koketsu, K.; Migita, A.; Boddy, C.N.; Wang, C.C.C.; Oguri, H.; Oikawa, H. *Nat. Chem. Biol.* **2006**, *2*, 423-428.

19. Jørgensen, H.; Fjærvik, E.; Hakvåg, S.; Bruheim, P.; Bredholt, H.; Klinkenberg, G.; Ellingsen, T.E.; Zotchev, S.B. *Applied Environ. Microbiol.* **2009**, *75*, 3296-3303.

20. Fisch, K.M.; Gurgui1, C.; Heycke, N.; van der Sar, S.A.; Anderson, S.A.; Webb, V.L.; Taudien, S.; Platzer, M.; Rubio, B.K.; Robinson, S.J.; Crews, P.; Piel, J. *Nat. Chem. Biol.* **2009**, *5*, 494-501.

21. Bibb, M.J. *Curr Opin Microbiol.* **2005**, *8*, 208-215.

22. Zhang, W.; Li, Y.; Tang, Y. *Proc. Natl. Acad. Sci. USA* **2008***, 105*, 20683-20688.

23. Pickens, L.B.; Tang, Y. *J. Biol. Chem.* **2010***, 285*, 27509-27515.

24. van Wezel, G.P.; McKenzie, N.L.; Nodwell, J. R. *Methods Enzymol.* **2009**, *458*, 117-141.

25. Potrykus, K.; Cashel, M. (p)ppGpp: Still Magical? *Annu. Rev. Microbiol.* **2008**, *62*, 35-51.

26. Hesketh, A.; Sun, J.; Bibb, M. J. *Mol. Microbiol.* **2001**, *150*, 1485-1493.

27. Knauber, T.; Doss, S. D.; Gerth, K.; Perlova, O.; Müller, R.; Treuner-Lange, A. *Mol. Microbiol.* **2008**, *69*, 254-266.

28. Wösten, M. M. *FEMS Microbiol. Rev.* **1998**, *22*, 127-150.

29. Britton, R. A.; Eichenberger, P.; Gonzalez-Pastor, J. E.; Fawcett, P.; Monson, R.; Losick, R.; Grossman, A. D*J. Bacteriol*. **184**, 4881-4890 (2002).

30**.** Chen, G. & Schellhorn**,** H. E. Controlled induction of the RpoS regulon in *Escherichia coli*, using an RpoS-expressing plasmid *Can. J. Microbiol.* **2003**, *49*, 733-740.

31. Zhoa, K.; Liu, M.; Burgess, R.R. *Nucleic Acid Res.* **2010**, *38*, 1273-1283.

32. Buck, M.; Gallegos, M.T.; Studholme, D.J.; Guo, Y.; Gralla, J.D. *J. Bacteriol.* **2000**, *182*, 4129–4136.

33. Studholme, D.J.; Buck, M.; Nixon, T. *Microbiology* **2000**, *146*, 3021-3023.

34. Barrios, H.; Valderrama, B.; Morett, E. *Nucleic Acids Res.* **1999**, *27*, 4305–4313.

35. Gronewold, T. M. A.; Kaiser, D. *J. Bacteriol.* **2007**, *189*, 1836-1844.

36. Tatusov, R.L.; Galperin, M.Y.; Natale, D.A.; Koonin, E.V. *Nucleic Acid Res.* **2000**, *28*, 33-36.

37. Ossa, F.; Diodati, M.E.; Caberoy, N.B.; Giglio, K.M.; Edmonds, M.; Singer, M.; Garza, A.G. *J. Bacteriol.* **2007**, *189*, 8474-8483.

38. Giglio, K.M.; Eisenstatt, J.; Garza, A.G. *J. Bacteriol*. **2010**, *192*, 360-364.

39. Jakobsen, J.S.; Jelsbak, L.; Jelsbak, L.; Welch, R.D.; Cummings, C.; Goldman, B.; Stark, E.; Slater, S.; Kaiser, D. *J. Bacteriol.* **2004**, *186*, 4361-4368.

40. Terui, Y.; Higashi, K.; Taniguchi, S.; Shigemasa, A.; Nishimura, K.; Yamamoto, K.; Kashiwagi, K.; Ishihama, A.; Igarashi, K. *J. Bacteriol.* **2007**, *189***,** 2359-2368.

41. Terui Y.; Higashi, K.; Tabei, Y.; Tomitori, H.; Yamamoto, K.; Ishihama, A.; Igarashi, K.; Kashiwagi, K. *J. Bacteriol.* **2009**, *191*, 5348-5357.

42. Lackner, G.; Schenk, A.; Xu, Z.; Reinhardt, K.; Yunt, Z.S.; Piel,J.; Hertweck, C. *J. Am. Chem. Soc.* **2007**, *129*, 9306-9312.

43. Zhang, J.; Van Lanen, S.G.; Ju, J.; Liu, W.; Dorrestein, P.C.; Li, W.; Kelleher, N.L.; Shen, B. *Proc. Natl. Acad. Sci. USA* **2008**, *105*, 1460-1465.

44. Buntin, K.; Irschik, H.; Weissman, K.J.; Luxenburger, E.; Blöcker H.; Rolf Müller, R. *Chem. Biol.* **2010**, *17*, 342-356.

45. Carlson, J.C.; Fortman, J.L.; Anzai, Y.; Li, S.; Burr, D.A.; Sherman D.H. *ChemBioChem* **2010**, *11*, 564-572.

46. Zaleta-Rivera, K.; Charkoudian, L.K.; Ridley, C.P.; Khosla, C. *J. Am. Chem. Soc.* **2010**, *132,* 9122–9128.

3.6 Experimental

General. *Escherichia coli* XL1 Blue (Stratagene) and *E. coli* TOP10 (Invitrogen) cells were used for routine cloning and plasmid preparations. *E. coli* BAP1 was used in all heterologous production experiments. *E. coli* cells were grown in LB media supplemented with the appropriate antibiotics. Oxytetracyline standard and IPTG were purchased from Sigma-Aldrich. Antibiotics and media components were purchased from Fisher. All strains were chemically or electronically transformed using standard transformation protocols.

Plasmid Construction. PrimeSTAR HS polymerase and the Thermocycler Mastercycler personal were used for all PCR reactions. Pfu Ultra II polymerase and the Thermocycler Mastercycler personal were used for mutagenesis reactions. Primers can be found in the supplementary information (Table S1). *E. coli* MG1655 genomic DNA was used as the template. Each sigma factor was cloned into pCR-Blunt and confirmed by sequencing. Genes were sub-cloned using *Nde*I and *Eco*RI restriction sites into pET28b or pKH22, except for *rpoE* and *rpoD* where *Hind*III instead of *Eco*RI and *Nhe*I instead of *Nde*I were used to due presence of native restriction sites. Plasmid names and descriptions can be found in the supplementary information (Table S3).

qRT-PCR Analysis. ABsolute SYBR Green QPCR mix (Thermo Scientific) and the Mx3000 qPCR (Agilent) were used for all qRT-PCR experiments. Thermocycler conditions were as follows: 1 cycle 95 °C for 15 min, 40 cycles of 95 °C for 15 s, 55 °C for 30 s, and 72 °C for 30 s, followed by denaturation for 1 cycle at 95 °C for 1 min, 55 °C for 30 s, and 95 °C for 30 s. The SV Total RNA Isolation System (Promega) was used to isolate RNA for the construction of cDNA. cDNA from each strain was

generated from isolated RNA using AMV Reverse Transcriptase (Promega) using the manufacturer's instructions. Strains, prepared by electroporating *E. coli* BAP1, were BAP1 + pDCS11 (null), BAP1 + pMRH08 (no σ factor), BAP1 + pDCS11 + pDCS61, BAP1 + pDCS11 + pDCS62 (oxyAp mutant), and BAP1 + pMRH08 + either pDCS11 (σ^{54}), pDCS57 (σ^{F}), pDCS58 (σ^{E}), pDCS59 (FecI), pNDP6 (σ^{S}), or pNDP7 (σ^{H}). Strains were grown in 25 mL LB medium in shake flasks at 37 °C, at $O.D._{600}$ = 0.4 cultures were induced with 1 mM IPTG and incubated at 20 °C for 24 hours after which the cells were harvested for RNA isolation. In concert with RNA isolation the broth from each culture was extracted using a solid phase extraction resin (XAD-16) and the organic extracts were analysed by ESI-LC-MS/MS for oxytetracyline production as described below. qRT-PCR experiments were performed in triplicate for each strain with a 10-fold dilution series of the 16s RNA acting as an internal standard. Primers were designed (see supplementary information, Table S1) to provide 180-210 bp amplicons of the *rpoN, oxyB, oxyF, oxyK, oxyP,* and *oxyT* transcripts. The qRT-PCR reactions contained 1.0 µM of each primer, 12.5 µL ABsolute SYBR Green QPCR mix, 0.5 µL prepared cDNA, and 10.5 µL filtered deionized H_2O to give a total reaction volume of 20 µL. Standard curve R^2 values and amplification efficiency values ranged from 0.991 - 1.0 and 90.7% - >100% respectively. The amplification efficiency was calculated using the formula $A = 10^{(-1/slope)}$ in which the slope was calculated by regression analysis obtained from the C_t values versus calculating log number of cells in serial dilutions.

Oxytetracycline Production, Isolation, Characterization. The negative control strains used in this experiment were prepared by electroporating *E. coli* BAP1 with either pDCS11, pDCS57, pDCS58, pDCS59, pNDP6, or pNDP7. These strains lacked the

oxyteteracycline pathway and thus cannot produce oxytetracycline. Test strains were prepared by electroporation of *E. coli* BAP1 with pMRH08, the oxytetracycline gene cluster, and either no additional plasmid, pDCS11, pDCS57, pDCS58, pDCS59, pNDP6, or pNDP7. All strains were grown in 25 mL LB medium in shake flasks at 37 °C. At $O.D._{600}$ = 0.4, cultures were induced with 1 mM IPTG and grown at 20 °C for 48 hours. Cultures were treated with acetone (1 mL) and were vigorously vortexed. Cell debris was removed by centrifugation and Amberlite XAD-16 resin (0.50 g) was added to the clarified media. The resin was incubated with the media overnight and collected by filtration. The organic extracts were eluted from the resin with 20 mL MeOH over 2 hours and concentrated *in vacuo*. Extracts were then resuspended in 100 μL MeOH for ESI-LC-MS/MS analysis. All ESI-LC-MS/MS analysis collected on an API2000 LC/MS/MS System (Applied Biosystems) equipped with a turbo-ion spray ESI probe interfaced with a Prominence UFLC (Shimadzu). A reverse phase BDS Hypersil C18 column (dimensions 100 mm x 2.1 mm I.D., particle size 3 μm, Thermo Scientific) was employed. Mobile phases throughout experimental were mobile phase A: 5% MeCN:95% H_2O with 0.05% formic acid and mobile phase B: 95% MeCN:5% H_2O with 0.05% formic acid. For initial determination of oxytetracycline production a gradient program (3 min 100% A hold, 40 min linear gradient to 100% B, 10 min 100% B hold, 2 min linear gradient to 100% A, 10 min 100% A hold) was developed with a flow rate of 0.250 mL/min into the mass spectrometer (Ion spray 5,500 V, mass range 250-550 m/z). For oxytetracycline MS^2 fragmentation scans a gradient program (3 min 100% A hold, 30 min linear gradient to 100% B, 2 min 100% B hold, 0.5 min linear gradient to 100% A, 4.5 min 100% A hold) was developed and the mass spectrometer settings were optimized

for oxytetracycline (Ion spray 5,500 V, collision energy 50 eV, Q3 scan 400-500 m/z, MS^2 parent ion 461.3 m/z, MS^2 product ion range 198-462 m/z). For MRM quantification of oxytetracycline production, a gradient program identical to that of the MS^2 fragmentation scan gradient program was developed for the determined mass spectrometer settings (Ion spray 5,500 V, collision energy 33 eV, MRM Q1 461.3 m/z, MRM Q3 483.2 m/z). Quantification was performed in triplicate by ESI-LC-MS/MS MRM analysis using a standard curve generated from commercial oxytetracycline (Aldrich). Other settings, including gas flow, lens and quadrupole voltages and parameters for mass resolution of the separating quadrupole were used as obtained during routine optimization of the instrument.

Gene cluster identification. To identify polyketide and non-ribosomal peptide biosynthetic gene clusters from bacterial genomes, all genes with cluster of orthologous groups (COG) annotations of polyketide synthase (COG3321), 3-oxoacyl acyl carrier protein synthase (COG0304), and non-ribosomal peptide synthetase (COG1020) were identified from 58 bacterial genomes. COG0304 annotated genes with gene names containing *fab* were discarded as they are involved in fatty acid biosynthesis, not type II polyketide biosynthesis. Operons containing genes commonly associated with secondary metabolite biosynthesis were included in the gene cluster. Transposons and housekeeping genes were considered to occur outside of the biosynthetic gene cluster. All identified polyketide and non-ribosomal peptide biosynthetic gene clusters can be found in the supplementary information (Table 3.6.6).

σ^{54} promoter identification. To identify putative σ^{54} promoter sequences in bacterial genomes, the PromScan Perl script was modified to output all hits within a

genome, including intragenic hits, with a score of 65 or higher. The script was run on Windows using Strawberry Perl (http://strawberryperl.com/). The inputs were bacterial genome DNA sequence files (FNA files from NCBI's Genome Database, http://www.ncbi.nlm.nih.gov/) and the σ^{54} promoter positional weighted matrix file which is based on 186 known σ^{54} promoter sites (see supplementary information, Table S5).

The results were uploaded to an SQL database which was created using SQL Server 2008 Express (http://www.microsoft.com/express/). PTT files (from NCBI's Genome Database, http://www.ncbi.nlm.nih.gov/) containing annotation data were uploaded to the database. Using LINQ to SQL, hits generated from the modified PromScan algorithm were linked to genes located within 500 bp upstream and in the same coding direction as the gene start site.

Table 3.6.1: Primer sequences with engineered restriction sites (bold) used for cloning of sigma factors. Restriction sites indicated in bold were used to clone the respected gene into either pKH22 or pET28b.

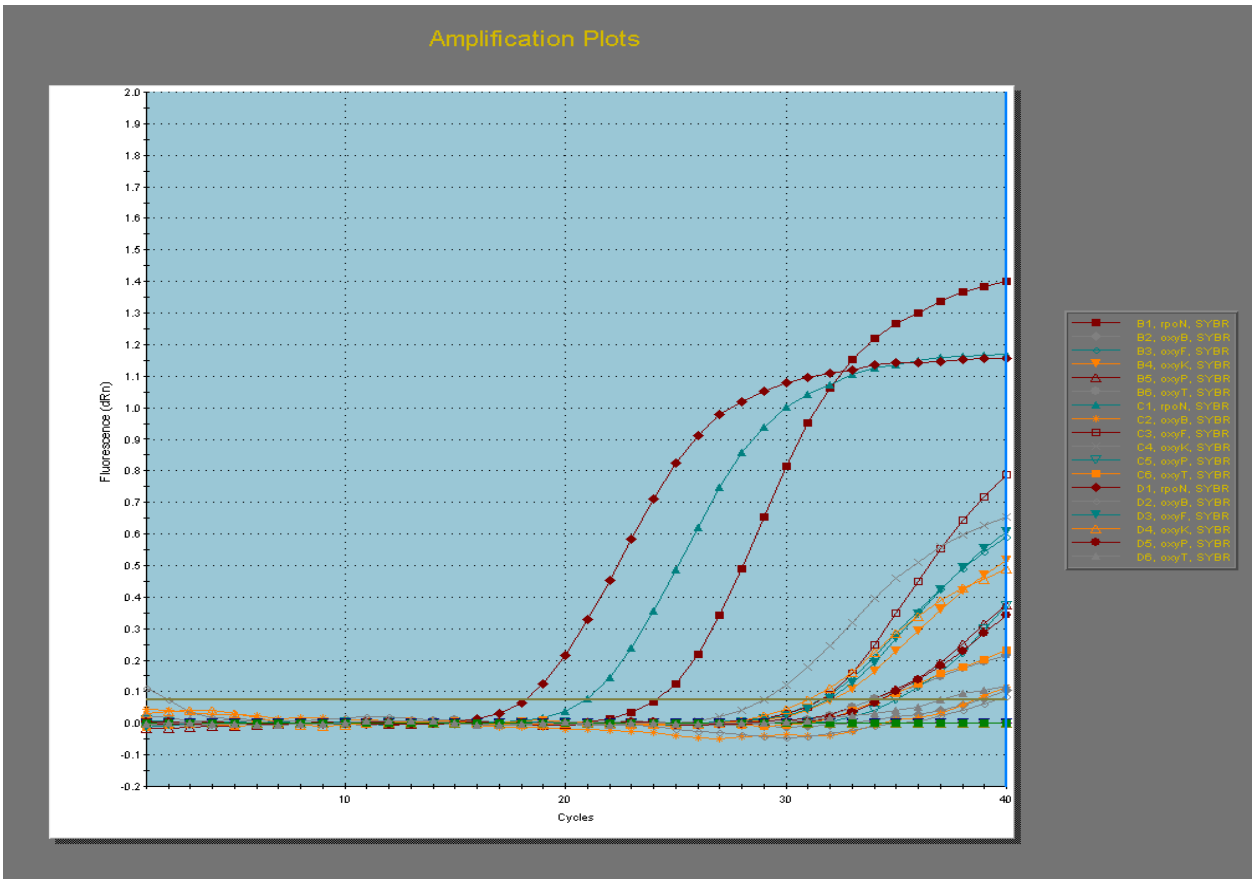
GENE	FORWARD	REVERSE	PROTEIN
E. coli rpoN	aatta**catatg**aacgaacctttgcaactca	aatta**gaattc**tcaaacgagtgtttacg	σ^{54}
E. coli rpoD	aatt**gctagc**atggagcaaaacccgc	aatt**gaattc**ttaattcgtccaggaag	σ^{70}
E. coli rpoE	aatt**catatg**agcgagcagttaacgg	aatt**gaattc**tcaacgcctgataagc	σ^{E}
E. coli rpoF	aatt**catatg**gtgaattcactctatac	aatt**aagctt**ttataacttacccagtt	σ^{F}
E. coli fecI	aatt**catatg**tctgaccgcgccacta	aatt**gaattc**tcaacgcctgataagc	FecI
E. coli rpoH	gtga**catatg**actgacaaaatgcaaa	cacg**gaattc**ttacgcttcaatggca	σ^{H}
E. coli rpoS	ctga**catatg**agtcagaatacgctga	cagt**gaattc**ttactcgcggaacagc	σ^{S}

Table 3.6.2: Primer sequences used for qRT-PCR from cDNA. Primers were used at 0.1 μM for PCR reactions.

TRANSCRIPT	FORWARD	REVERSE
16s RNA	aactgcctgatggaggggg	gtctcagttccagtgtggct
rpoN	gtctgttgcagttgtcgacg	cggtgtaaatggtgtcccaa
oxyB	tcaccggcctcggtgtggtc	gcgcgagccgggtcatcagg
oxyF	cccgcctgctccaccagcgc	cctggagtgggacgacgaga
oxyK	ccctgatcagggacgccgcc	tcgcggccaccggaggcgaa
oxyP	ggcccatccgcccgcttagg	gcaggtcgccgaccccgcgt
oxyT	aggcgaaggccctgcacagc	cggacaccaggaacgcctgg

Table 3.6.3: Plasmid Table.

PLASMID	DESCRIPTION
DCS08	*E. coli rpoN* in pCR-BLUNT
DCS11	*E. coli rpoN* in pKH22
DCS53	*E. coli rpoF* in pCR-BLUNT
DCS54	*E. coli rpoE* in pCR-BLUNT
DCS55	*E. coli fecI* in pCR-BLUNT
DCS57	*E. coli rpoF* in pKH22
DCS58	*E. coli rpoE* in pKH22
DCS59	*E. coli fecI* in pKH22
NDP4	*E. coli rpoS* in pCR-BLUNT
NDP5	*E. coli rpoH* in pCR-BLUNT
NDP6	*E. coli rpoS* in pKH22
NDP7	*E. coli rpoH* in pKH22
MRH08	*S. rimosus* oxytetraycline biosynthetic pathway in pET28b



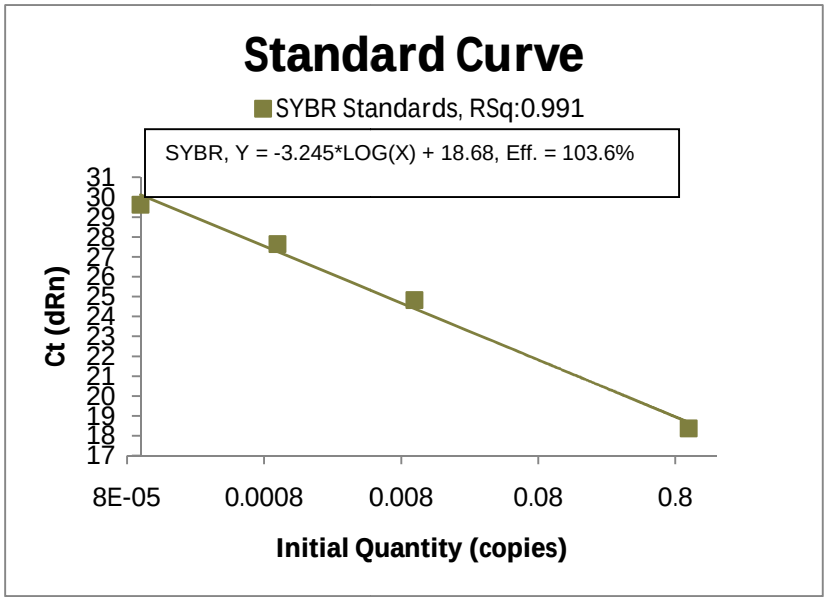
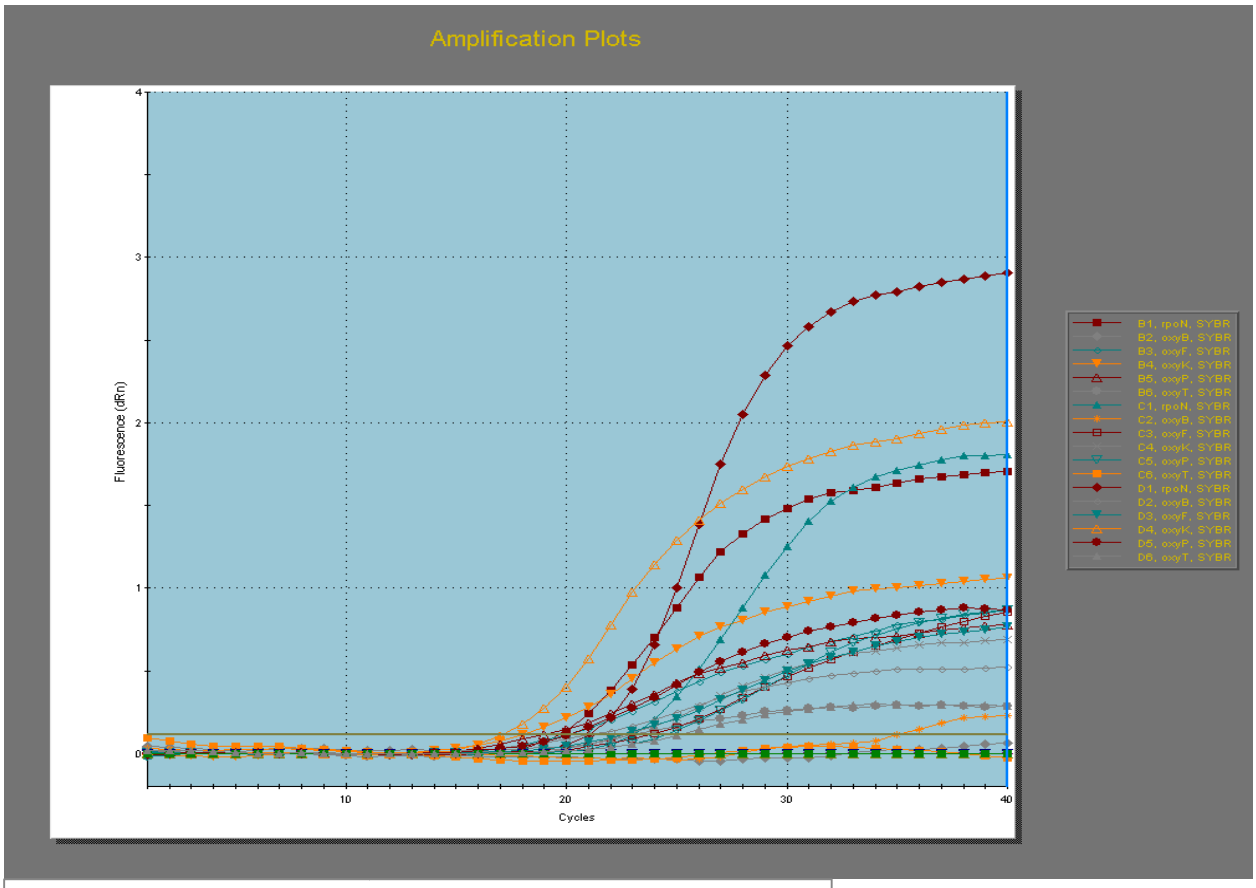


Figure 3.6.1: *E. coli* BAP1 + σ^{54} overexpression rtPCR analysis with standard curve



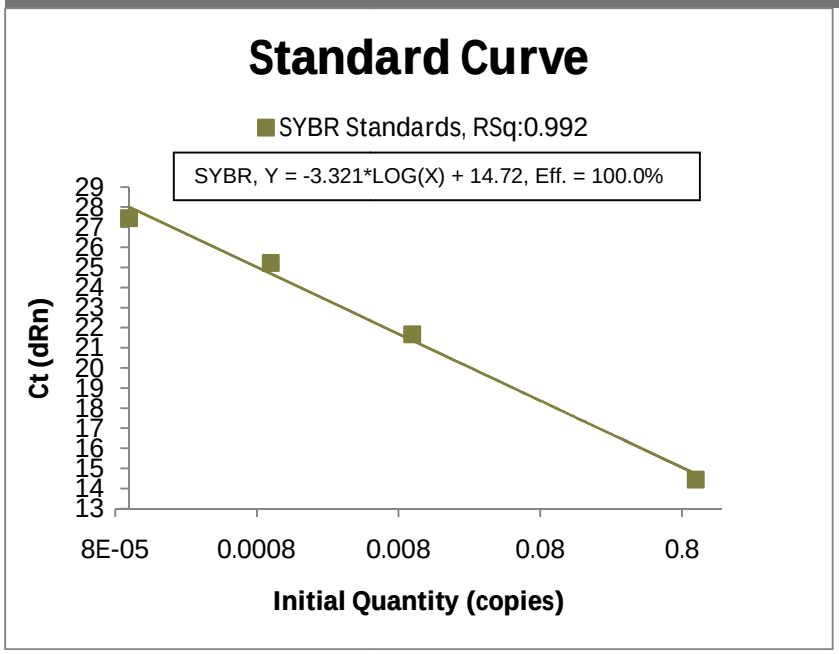
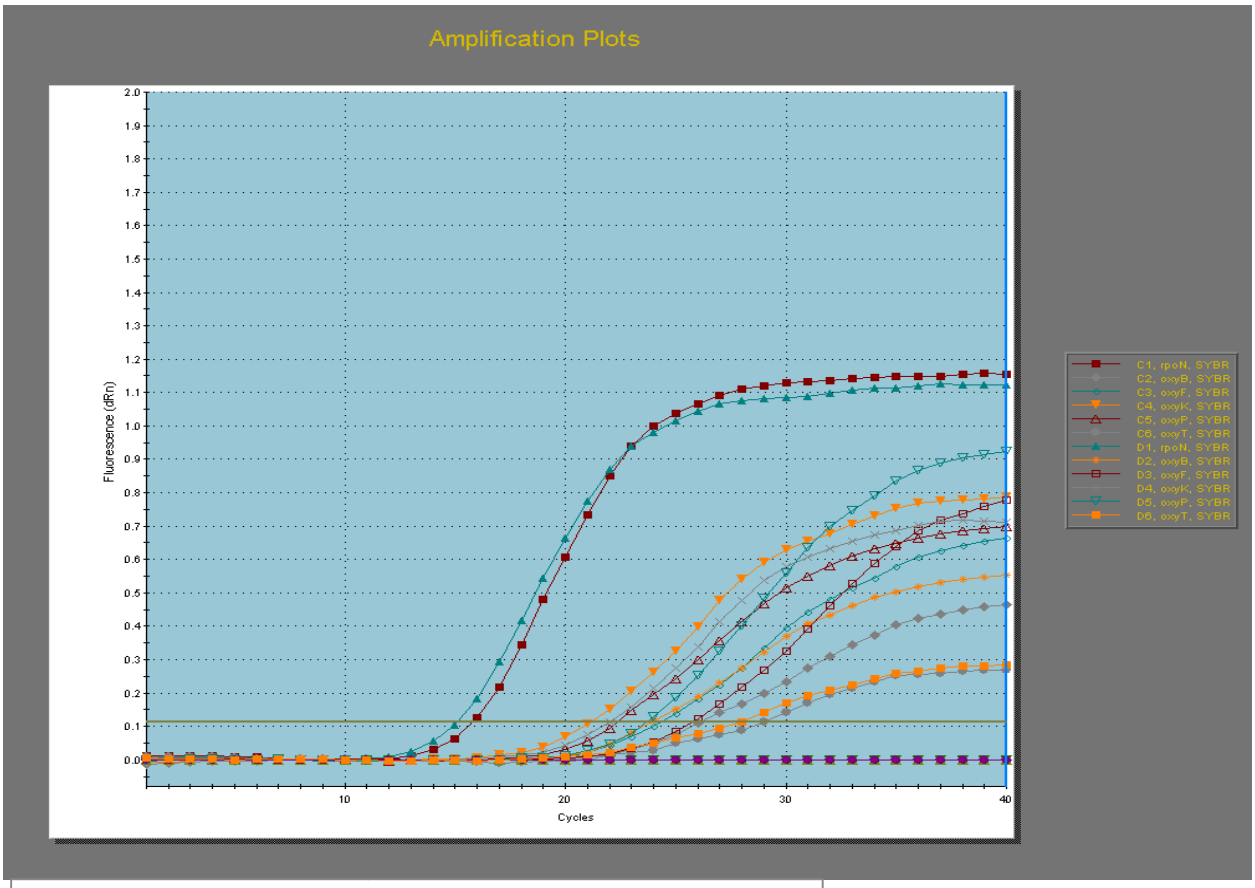


Figure 3.6.2: *E. coli* BAP1 + oxytetracycline pathway rtPCR analysis with standard curve



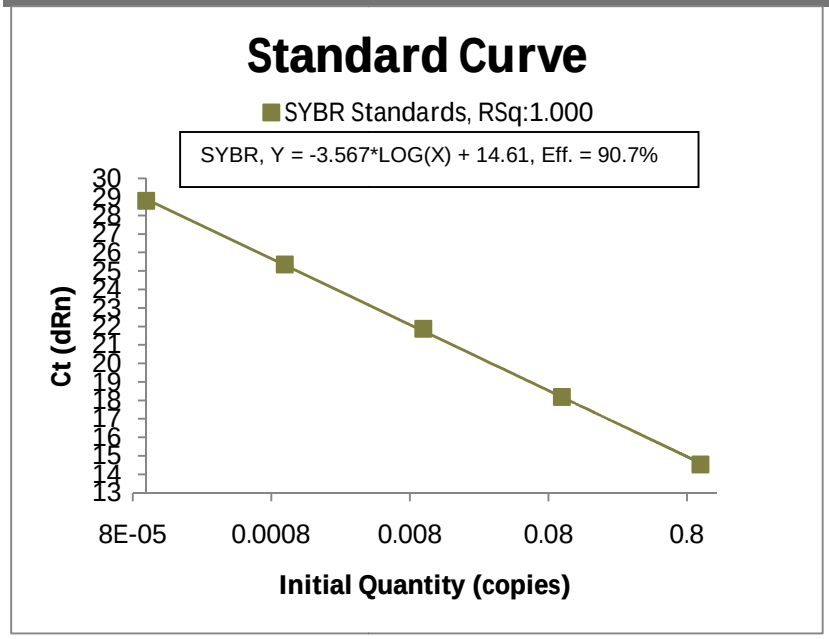


Figure 3.6.3: *E. coli* BAP1 + σ^{54} overexpression + oxytetracycline pathway rtPCR analysis with standard curve

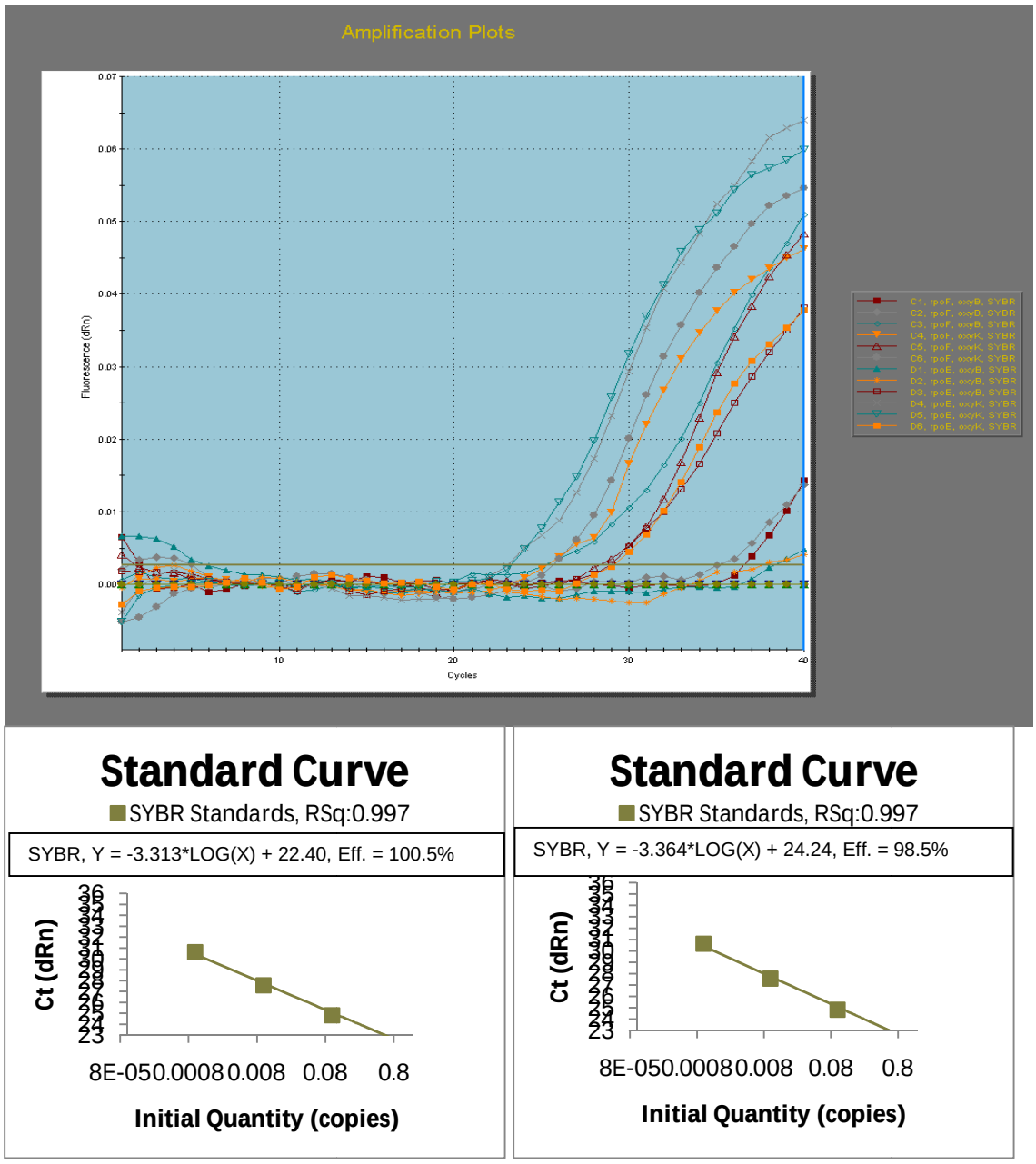


Figure 3.6.4: *E. coli* BAP1 + σ^F overexpression + oxytetracycline pathway and *E. coli* BAP1 + σ^E overexpression + oxytetracycline rtPCR analysis with standard curves

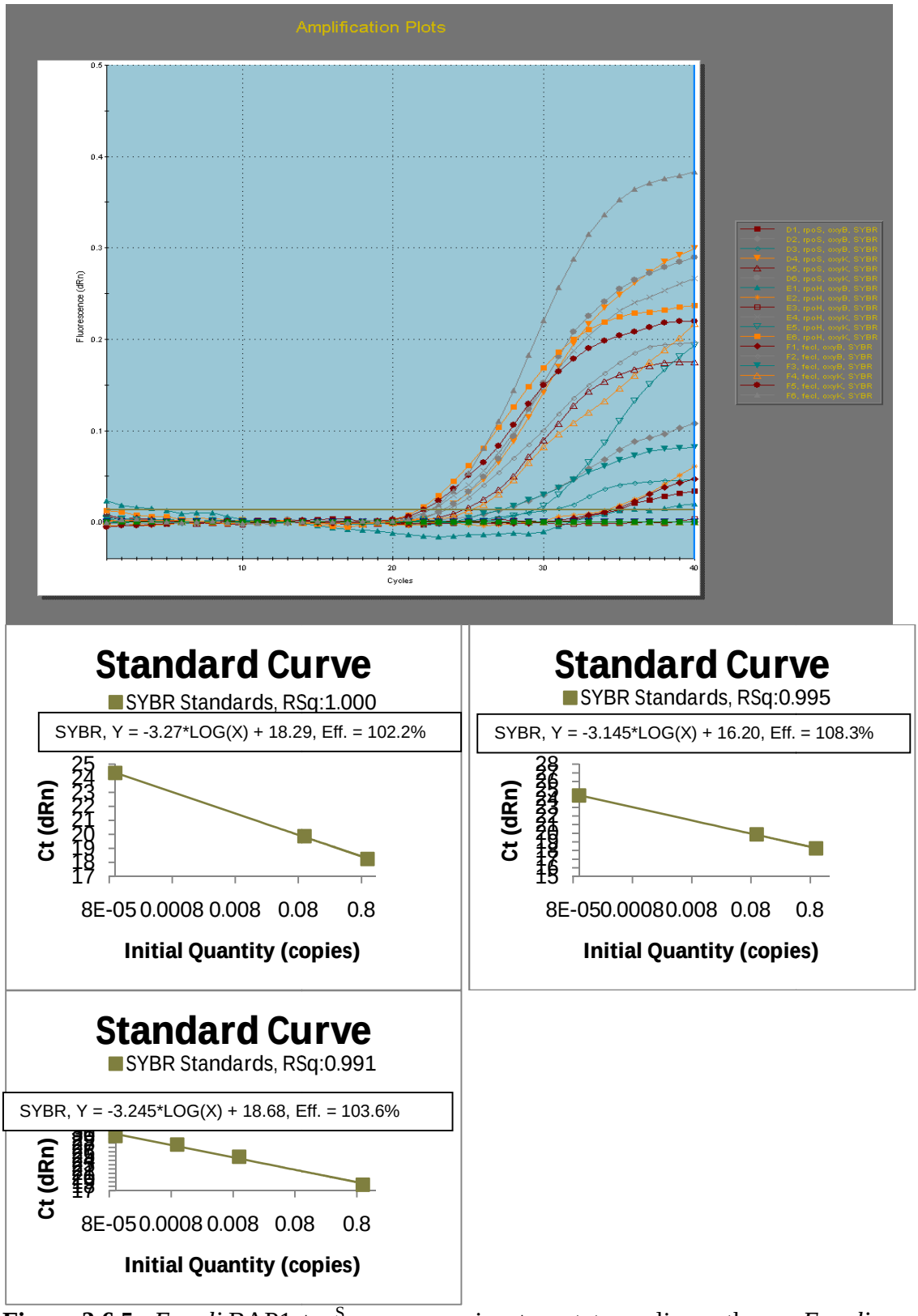


Figure 3.6.5: *E. coli* BAP1 + σ^S overexpression + oxytetracycline pathway, *E. coli* BAP1 + σ^H overexpression + oxytetracycline, and *E. coli* BAP1 + FecI overexpression + oxytetracycline rtPCR analysis with standard curves

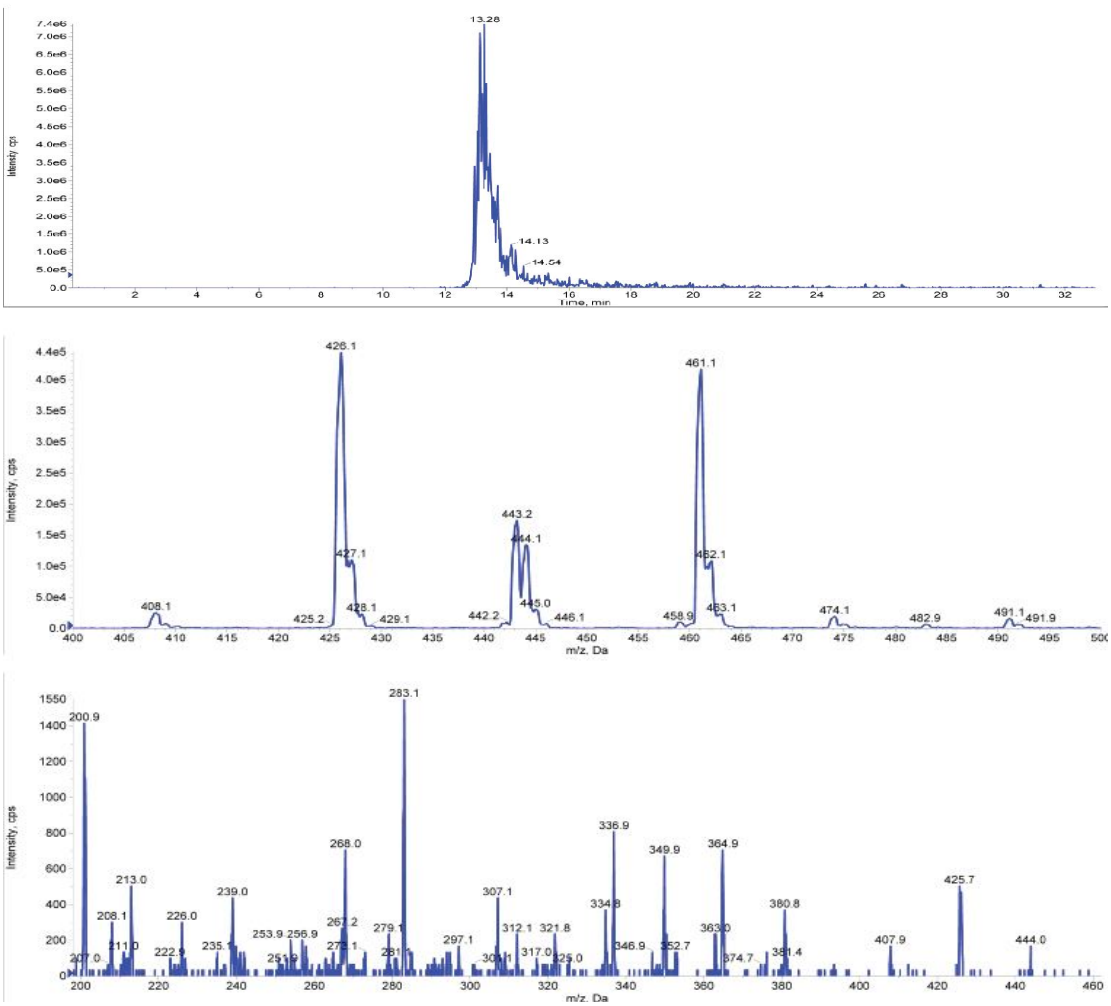


Figure 3.6.6: ESI-LC-MS/MS analysis of an oxytetracycline standard. The top trace shows the total ion count (TIC) chromatogram. Oxytetracycline elutes at approximately 13 min and is detectable in the TIC trace. The middle spectrum shows the mass spectrum from 12 min to 14 min. The $[M+H]^+$ of oxytetracycline (461) can be clearly seen in the mass spectrum. The bottom spectrum shows the MS^2 trace between 12 and 14 min for the 461 ion. Key fragments of oxytetracycline visible in the MS^2 spectrum are 201, 283, 337, 350, 365, 382, and 426.

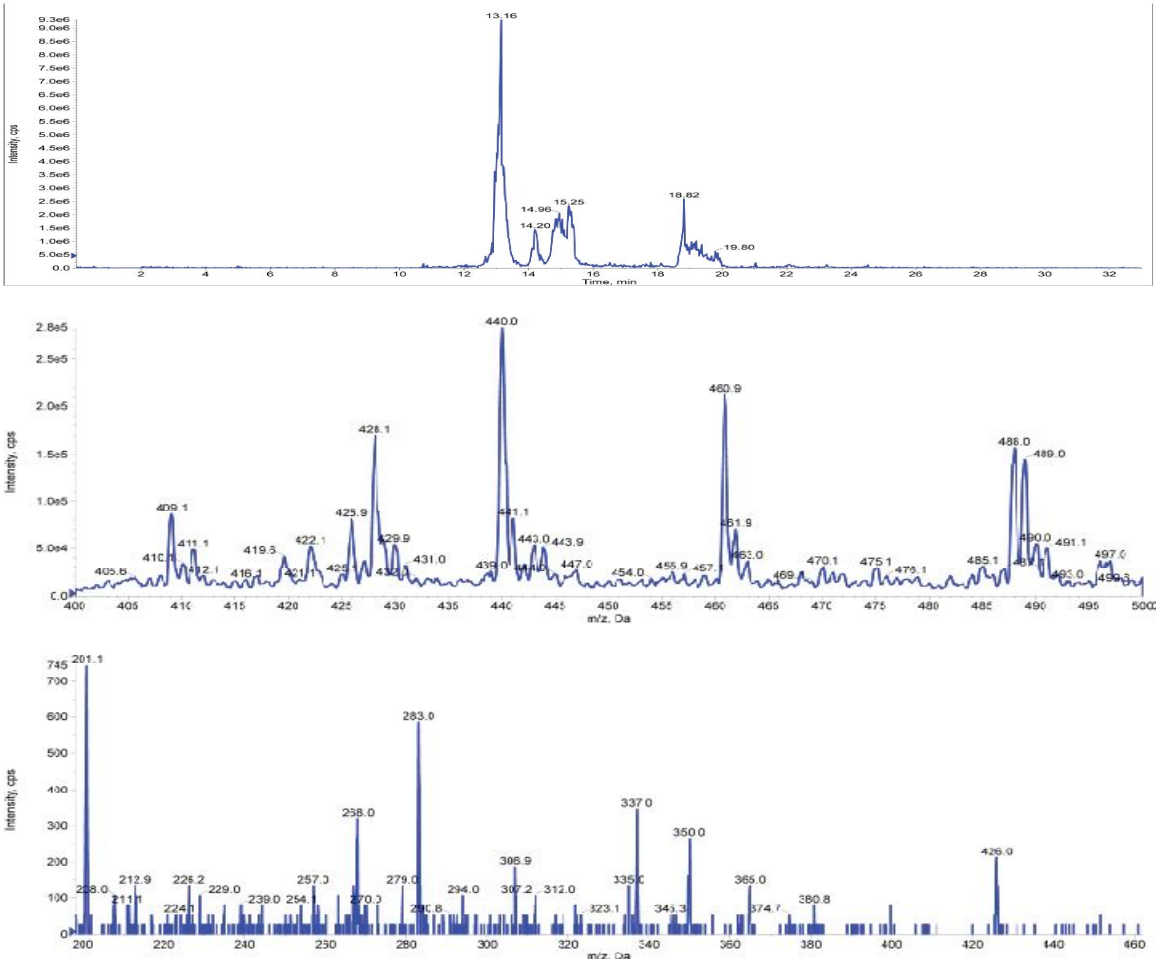


Figure 3.6.7: ESI-LC-MS/MS analysis of the organic extracts from a culture of BAP1 with the oxytetracycline biosynthetic gene cluster over-expressing σ^{54}. The top trace shows the total ion count (TIC) chromatogram. Oxytetracycline elutes at approximately 13 min and is detectable in the TIC trace. The middle spectrum shows the mass spectrum from 12 min to 14 min. The $[M+H]^+$ of oxytetracycline (461) can be clearly seen in the mass spectrum. The bottom spectrum shows the MS^2 trace between 12 and 14 min for the 461 ion. Key fragments of oxytetracycline visible in the MS^2 spectrum are 201, 283, 337, 350, 365, 382, and 426.

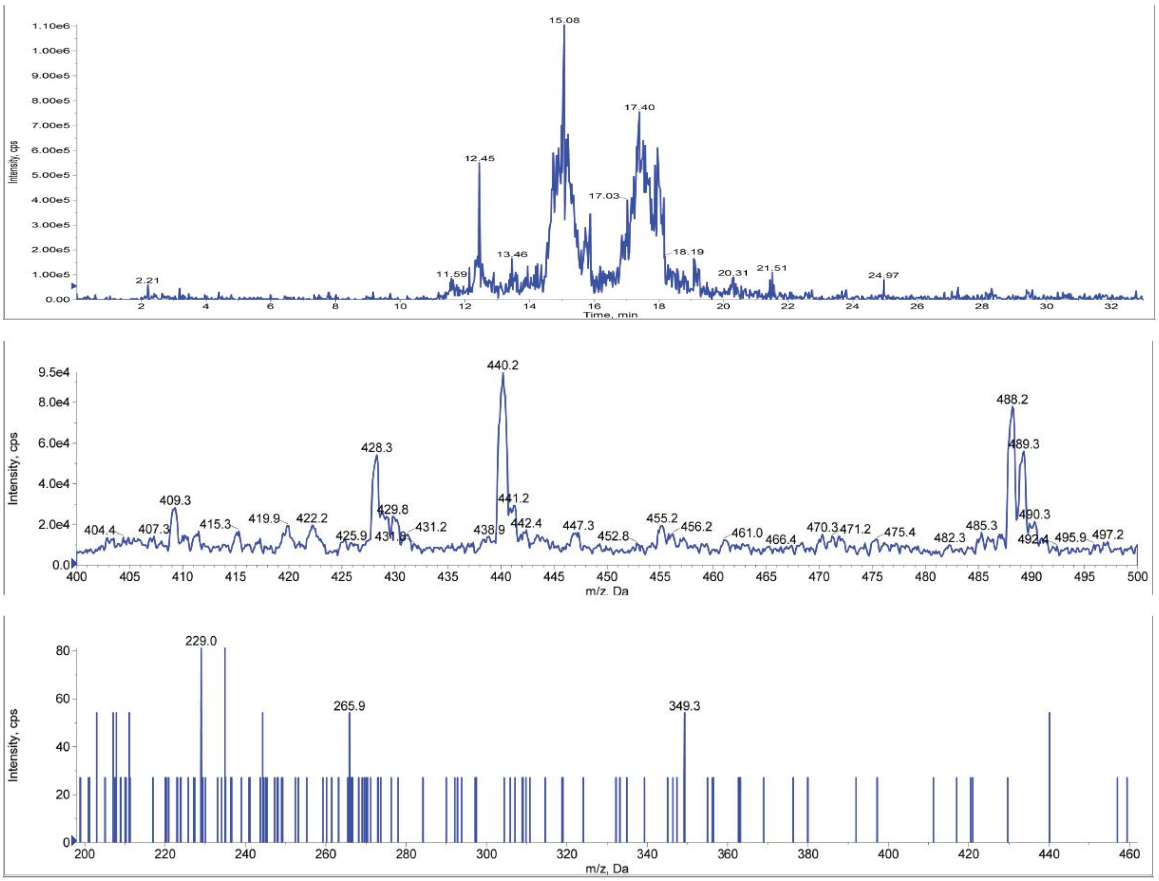


Figure 3.6.8: ESI-LC-MS/MS analysis of the organic extracts from a culture of BAP1 with the oxytetracycline biosynthetic gene cluster. The top trace shows the total ion count (TIC) chromatogram. Oxytetracycline elutes at approximately 13 min and is not detectable in the TIC trace. The middle spectrum shows the mass spectrum from 12 min to 14 min. The $[M+H]^+$ of oxytetracycline (461) cannot be seen in the mass spectrum. The bottom spectrum shows the MS^2 trace between 12 and 14 min for the 461 ion. None of the key fragments of oxytetracycline are visible in the MS^2 spectrum.

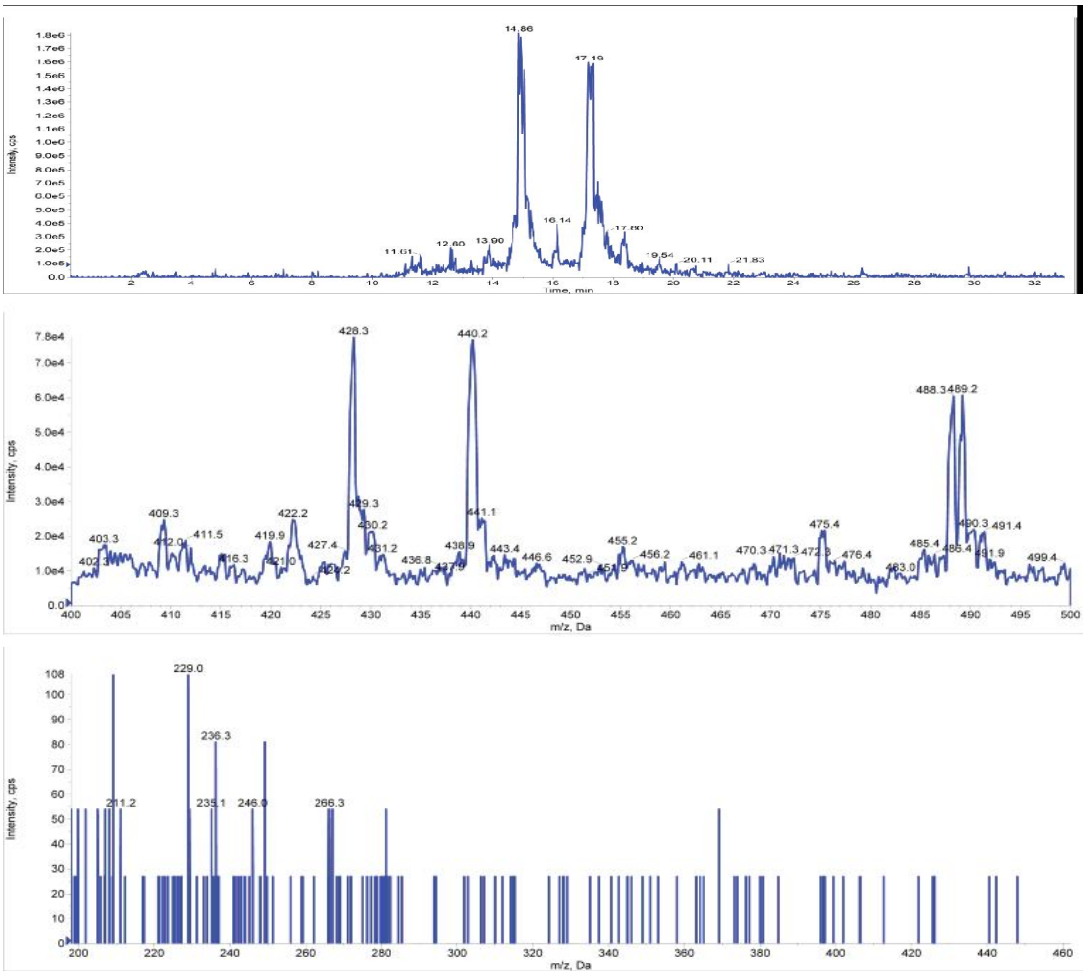


Figure 3.6.9: ESI-LC-MS/MS analysis of the organic extracts from a culture of BAP1 over-expressing σ^{54}. The top trace shows the total ion count (TIC) chromatogram. Oxytetracycline elutes at approximately 13 min and is not detectable in the TIC trace. The middle spectrum shows the mass spectrum from 12 min to 14 min. The $[M+H]^+$ of oxytetracycline (461) cannot be seen in the mass spectrum. The bottom spectrum shows the MS^2 trace between 12 and 14 min for the 461 ion. None of the key fragments of oxytetracycline are visible in the MS^2 spectrum.

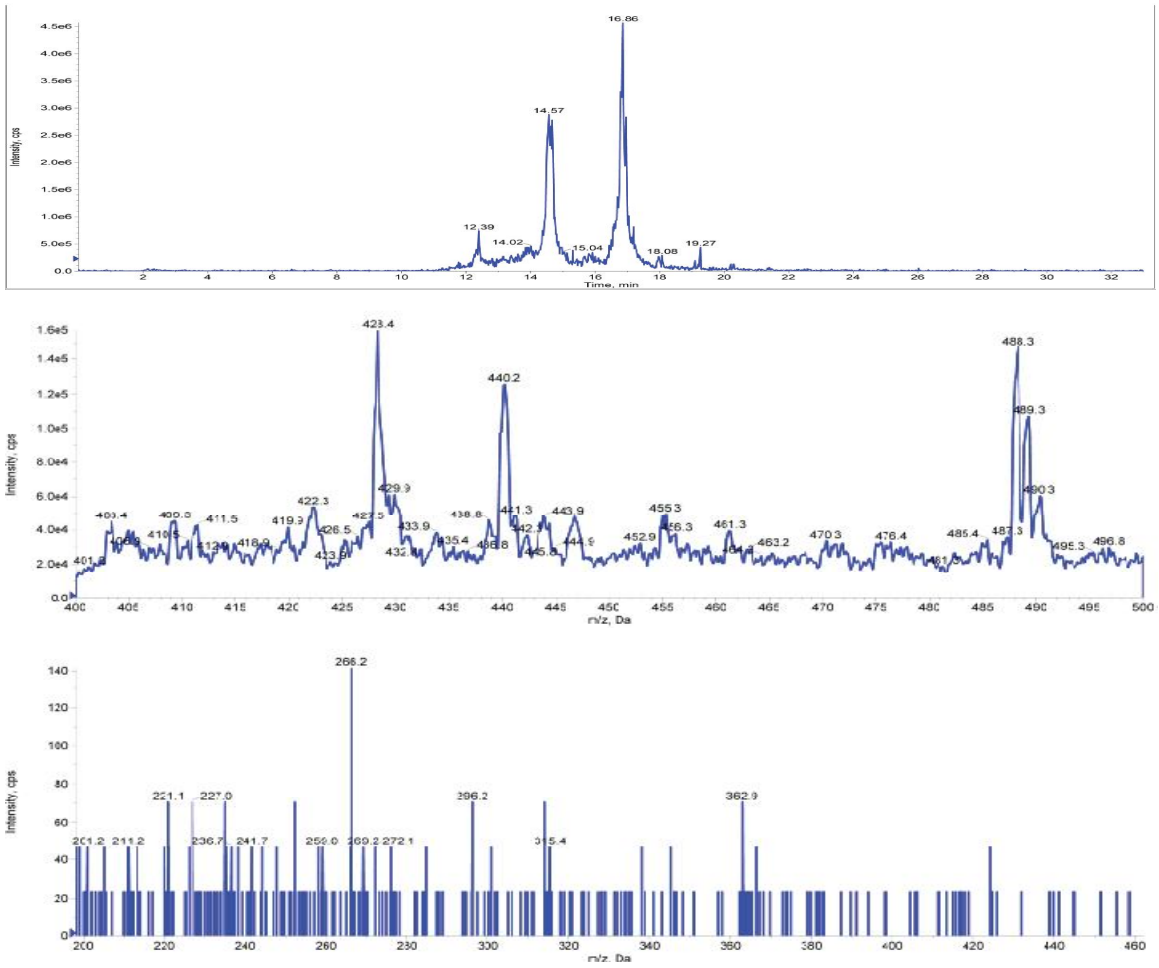


Figure 3.6.10: ESI-LC-MS/MS analysis of the organic extracts from a culture of BAP1 over-expressing FecI. The top trace shows the total ion count (TIC) chromatogram. Oxytetracycline elutes at approximately 13 min and is not readily detectable in the TIC trace. The middle spectrum shows the mass spectrum from 12 min to 14 min. The $[M+H]^+$ of oxytetracycline (461) cannot be seen in the mass spectrum. The bottom spectrum shows the MS^2 trace between 12 and 14 min for the 461 ion. None of the key fragments of oxytetracycline are visible in the MS^2 spectrum.

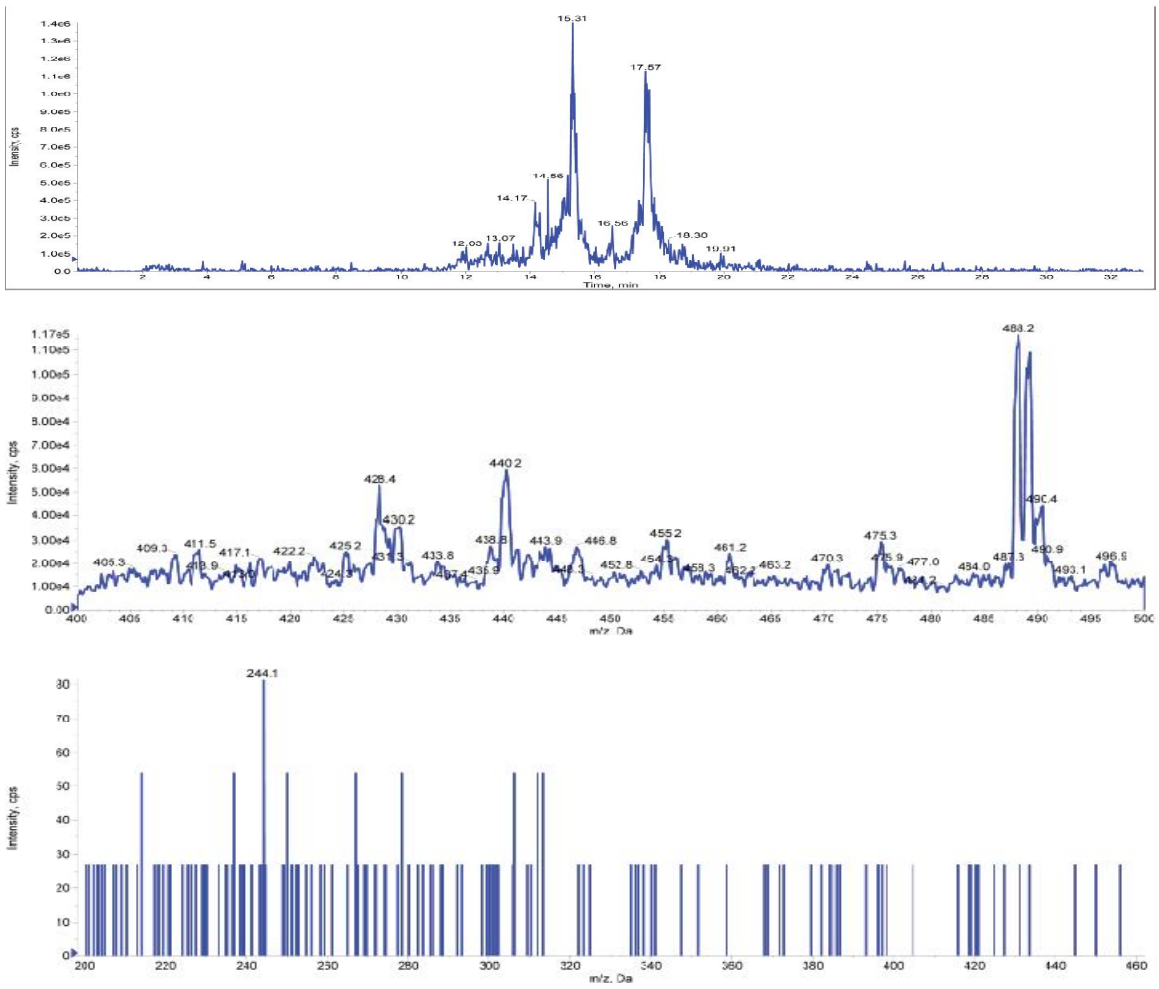


Figure 3.6.11: ESI-LC-MS/MS analysis of the organic extracts from a culture of BAP1 over-expressing σ^S. The top trace shows the total ion count (TIC) chromatogram. Oxytetracycline elutes at approximately 13 min and is not detectable in the TIC trace. The middle spectrum shows the mass spectrum from 12 min to 14 min. The $[M+H]^+$ of oxytetracycline (461) cannot be seen in the mass spectrum. The bottom spectrum shows the MS^2 trace between 12 and 14 min for the 461 ion. None of the key fragments of oxytetracycline are visible in the MS^2 spectrum.

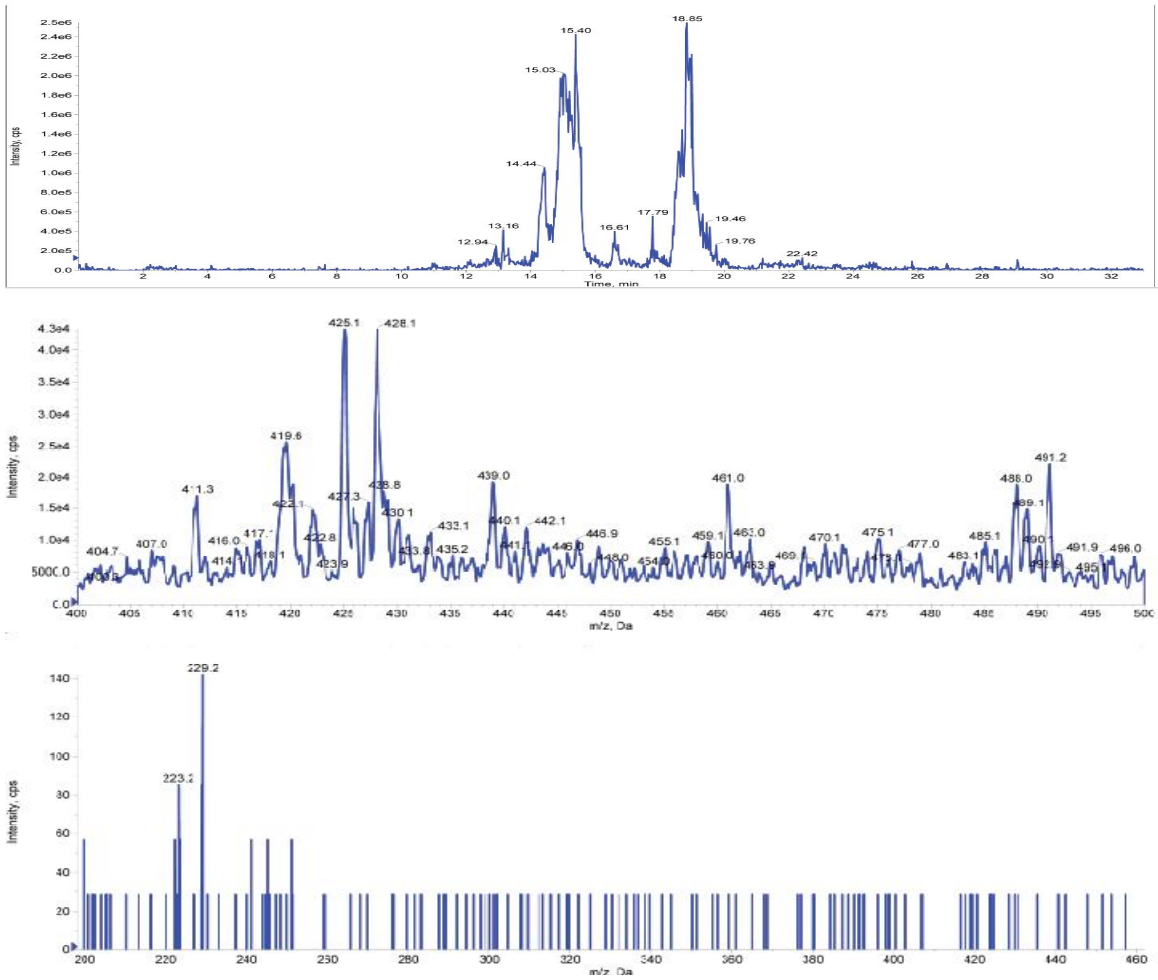


Figure 3.6.12: ESI-LC-MS/MS analysis of the organic extracts from a culture of BAP1 over-expressing σ^E. The top trace shows the total ion count (TIC) chromatogram. Oxytetracycline elutes at approximately 13 min and is not detectable in the TIC trace. The middle spectrum shows the mass spectrum from 12 min to 14 min. The $[M+H]^+$ of oxytetracycline (461) cannot be seen in the mass spectrum. The bottom spectrum shows the MS^2 trace between 12 and 14 min for the 461 ion. None of the key fragments of oxytetracycline are visible in the MS^2 spectrum.

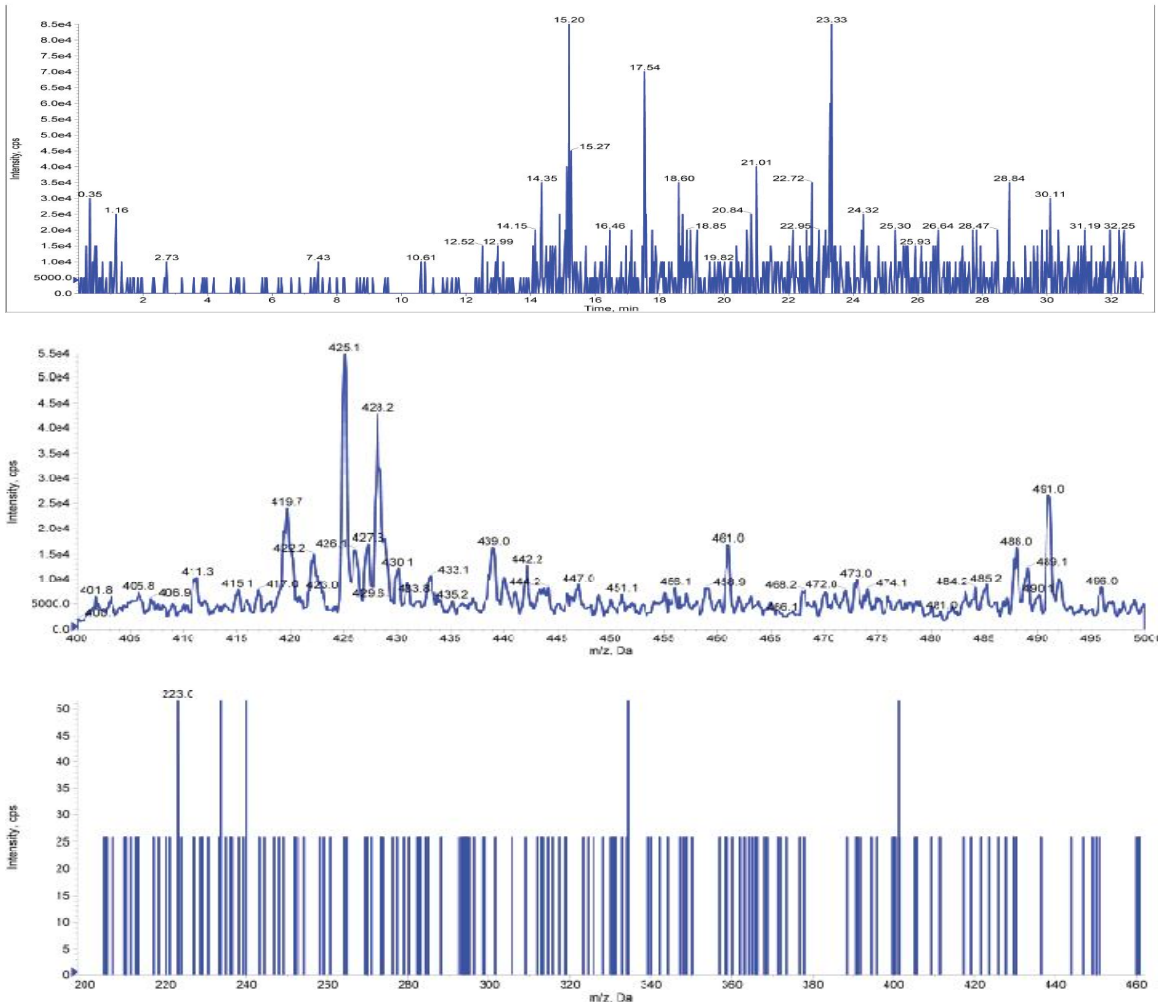


Figure 3.6.13: ESI-LC-MS/MS analysis of the organic extracts from a culture of BAP1 over-expressing σ^F. The top trace shows the total ion count (TIC) chromatogram. Oxytetracycline elutes at approximately 13 min and is not detectable in the TIC trace. The middle spectrum shows the mass spectrum from 12 min to 14 min. The $[M+H]^+$ of oxytetracycline (461) cannot be seen in the mass spectrum. The bottom spectrum shows the MS^2 trace between 12 and 14 min for the 461 ion. None of the key fragments of oxytetracycline are visible in the MS^2 spectrum.

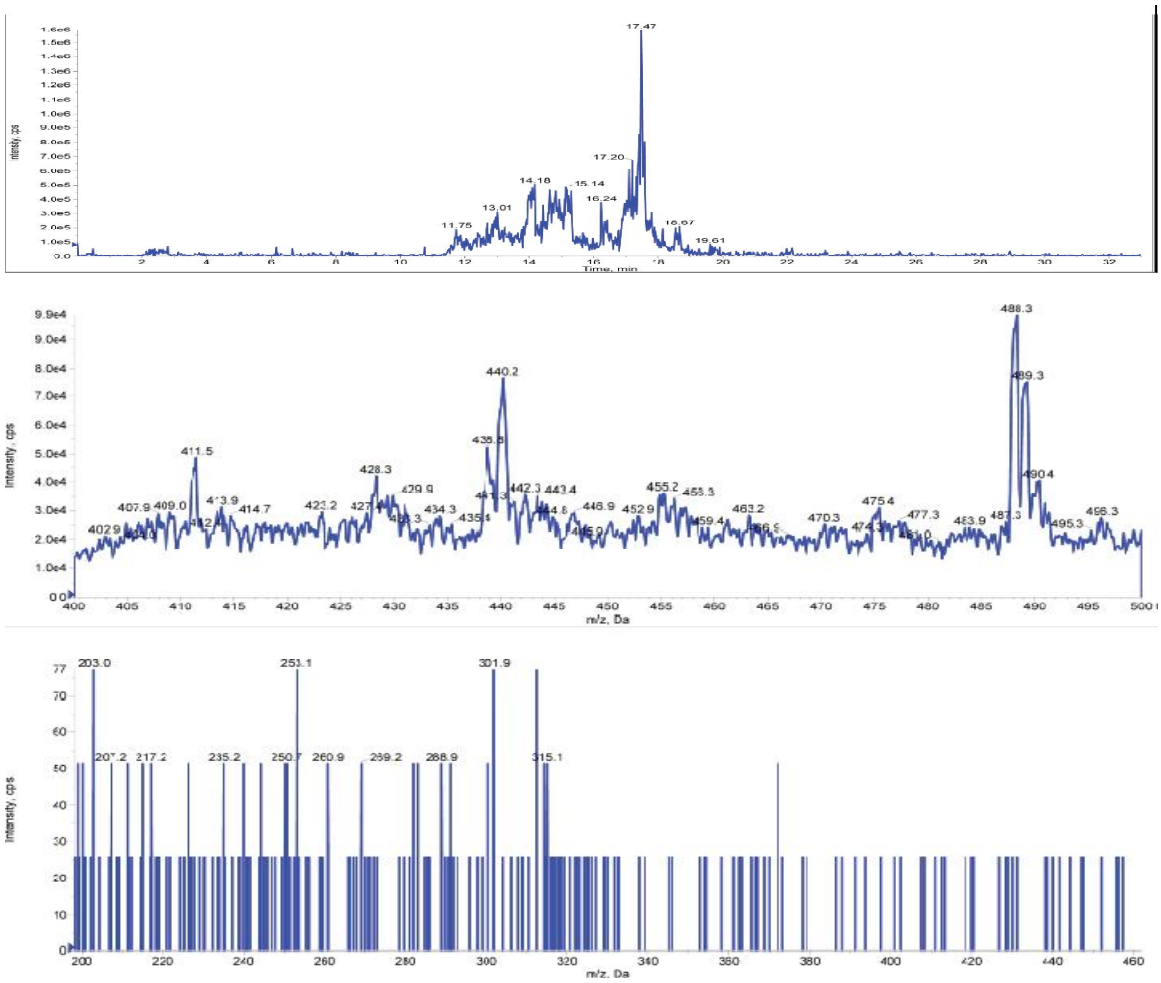


Figure 3.6.14: ESI-LC-MS/MS analysis of the organic extracts from a culture of BAP1 over-expressing σ^H. The top trace shows the total ion count (TIC) chromatogram. Oxytetracycline elutes at approximately 13 min and is not detectable in the TIC trace. The middle spectrum shows the mass spectrum from 12 min to 14 min. The $[M+H]^+$ of oxytetracycline (461) cannot be seen in the mass spectrum. The bottom spectrum shows the MS^2 trace between 12 and 14 min for the 461 ion. None of the key fragments of oxytetracycline are visible in the MS^2 spectrum.

Table 3.6.4: List of organisms used for bioinformatics analysis.

Phylum	Genome Name	Contains σ^{54}	RefSeq ID
Actinobacteria	Bifidobacterium longum NCC2705, complete genome	N	NC_004307.2
Actinobacteria	Mycobacterium tuberculosis CDC1551, complete genome	N	NC_002755.2
Actinobacteria	Mycobacterium ulcerans Agy99, complete genome	N	NC_008611.1
Actinobacteria	Nocardia farcinica IFM 10152, complete genome	N	NC_006361.1
Actinobacteria	Rubrobacter xylanophilus DSM 9941, complete genome	N	NC_008148.1
Actinobacteria	Saccharopolyspora erythraea NRRL 2338, complete genome	N	NC_009142.1
Actinobacteria	Salinispora tropica CNB-440, complete genome	N	NC_009380.1
Actinobacteria	Streptomyces avermitilis MA-4680, complete genome	N	NC_003155.4
Actinobacteria	Streptomyces coelicolor A3(2), complete genome	N	NC_003888.3
Actinobacteria	Streptomyces griseus subsp. griseus NBRC 13350, complete genome	N	NC_010572.1
Aquificae	Sulfurihydrogenibium sp. YO3AOP1, complete genome	Y	NC_010730.1
Bacteroidetes/Chlorobi	Bacteroides fragilis NCTC 9343, complete genome	Y	NC_003228.3
Bacteroidetes/Chlorobi	Candidatus Amoebophilus asiaticus 5a2, complete genome	Y	NC_010830.1
Bacteroidetes/Chlorobi	Gramella forsetii KT0803, complete genome	Y	NC_008571.1
Chlamydiae/Verrucomicrobia	Chlamydia trachomatis D/UW-3/CX, complete genome	Y	NC_000117.1
Chlamydiae/Verrucomicrobia	Chlamydophila pneumoniae AR39, complete genome	Y	NC_002179.2
Chlamydiae/Verrucomicrobia	Coraliomargarita akajimensis DSM 45221 chromosome, complete genome	N	NC_014008.1
Chloroflexi	Sphaerobacter thermophilus DSM 20745 chromosome 1, complete genome	Y	NC_013523.1
	Sphaerobacter thermophilus DSM 20745 chromosome 2, complete genome		NC_013524.1
Crenarchaeota	Pyrobaculum arsenaticum DSM 13514, complete genome	N	NC_009376.1
Cyanobacteria	Glocobacter violaceus PCC 7421, complete genome	N	NC_005125.1
Cyanobacteria	Microcystis aeruginosa NIES-843, complete genome	N	NC_010296.1
Cyanobacteria	Nostoc punctiforme PCC 73102, complete genome	N	NC_010628.1
Cyanobacteria	Synechocystis sp. PCC 6803, complete genome	N	NC_000911.1
Deinococcus-Thermus	Deinococcus deserti VCD115, complete genome	N	NC_012526.1
Euryarchaeota	Pyrococcus horikoshii OT3, complete genome	N	NC_000961.1
Fibrobacteres/Acidobacteria	Solibacter usitatus Ellin6076, complete genome	Y	NC_008536.1
Firmicutes	Bacillus cereus subsp. cytotoxis NVH 391-98, complete genome	Y	NC_009674.1
Firmicutes	Bacillus subtilis subsp. subtilis str. 168, complete genome	Y	NC_000964.3
Firmicutes	Clostridium acetobutylicum ATCC 824, complete genome	Y	NC_003030.1
Firmicutes	Clostridium kluyveri DSM 555, complete genome	Y	NC_009706.1
Firmicutes	Clostridium perfringens str. 13, complete genome	Y	NC_003366.1
Firmicutes	Lactobacillus brevis ATCC 367, complete genome	N	NC_008497.1
Firmicutes	Listeria monocytogenes str. 4b F2365, complete genome	Y	NC_002973.6

Firmicutes	Staphylococcus aureus subsp. aureus MRSA252, complete genome	N	NC_002952.2
Firmicutes	Streptococcus pyogenes M1 GAS, complete genome	N	NC_002737.1
Proteobacteria (Alpha)	Agrobacterium tumefaciens str. C58 chromosome circular, complete sequence	Y	NC_003062.2
	Agrobacterium tumefaciens str. C58 chromosome linear, complete sequence		NC_003063.2
Proteobacteria (Alpha)	Methylobacterium extorquens DM4, complete genome	Y	NC_012988.1
Proteobacteria (Alpha)	Rickettsia rickettsii str. Iowa, complete genome	N	NC_010263.1
Proteobacteria (Beta)	Neisseria meningitidis MC58, complete genome	Y	NC_003112.2
Proteobacteria (Beta)	Ralstonia solanacearum GMI1000 plasmid pGMI1000MP, complete sequence	Y	NC_003296.1
	Ralstonia solanacearum GMI1000, complete genome		NC_003295.1
Proteobacteria (Delta)	Geobacter sulfurreducens PCA, complete genome	Y	NC_002939.4
Proteobacteria (Delta)	Myxococcus xanthus DK 1622, complete genome	Y	NC_008095.1
Proteobacteria (Delta)	Sorangium cellulosum 'So ce 56', complete genome	Y	NC_010162.1
Proteobacteria (Epsilon)	Campylobacter jejuni subsp. jejuni NCTC 11168, complete genome	Y	NC_002163.1
Proteobacteria (Epsilon)	Helicobacter pylori J99, complete genome	Y	NC_000921.1
Proteobacteria (Gamma)	Candidatus Carsonella ruddii PV, complete genome	N	NC_008512.1
Proteobacteria (Gamma)	Escherichia coli str. K-12 substr. MG1655, complete genome	Y	NC_000913.2
Proteobacteria (Gamma)	Francisella tularensis subsp. tularensis SCHU S4, complete genome	N	NC_006570.2
Proteobacteria (Gamma)	Legionella pneumophila subsp. pneumophila str. Philadelphia 1, complete genome	Y	NC_002942.5
Proteobacteria (Gamma)	Pseudomonas aeruginosa PAO1, complete genome	Y	NC_002516.2
Proteobacteria (Gamma)	Pseudomonas putida KT2440, complete genome	Y	NC_002947.3
Proteobacteria (Gamma)	Pseudomonas syringae pv. tomato str. DC3000, complete genome	Y	NC_004578.1
Proteobacteria (Gamma)	Salmonella enterica subsp. enterica serovar Typhi str. CT18, complete genome	Y	NC_003198.1
Proteobacteria (Gamma)	Shewanella oncidensis MR-1, complete genome	Y	NC_004347.1
Proteobacteria (Gamma)	Yersinia pestis CO92, complete genome	Y	NC_003143.1
Spirochaetes	Borrelia burgdorferi B31, complete genome	Y	NC_001318.1
Spirochaetes	Treponema denticola ATCC 35405, complete genome	Y	NC_002967.9
Tenericutes	Candidatus Phytoplasma mali, complete genome	N	NC_011047.1

Table 3.6.5: σ^{54} promoter positional weighted matrix used to identify σ^{54} promoters.

Matrix derived from 186 σ^{54} promoter sites

A:	12	2	0	12	139	11	55	51	46	44	38	13	4	1	9	76
C:	14	0	0	147	23	122	17	48	64	42	62	22	18	2	173	5
G:	10	184	186	6	18	10	103	69	36	35	43	15	10	181	1	17
T:	150	0	0	21	6	43	11	18	40	65	43	136	154	2	3	88

Table 3.6.6: List of PKS, NRPS and NRPS/PKS gene clusters with the associated σ^{54} promoters identified from our bioinformatics analysis.

NC_002755.2 Mycobacterium tuberculosis CDC1551, complete genome

1	NRPS		105314-123821					
	Score	Distance From ORF	N	Strand	Sequence	Gene	Synonym	Protein Name
	75	430	106294	+	TTGGCAGCGCTTCGGCT	-	MT0106	dioxygenase, putative
	75	157	106567	+	CTGGCAGTCAAGGTGCC	-	MT0106	dioxygenase, putative
	81	323	109450	+	GTGGCGTGAACATTGCG	-	MT0109	hypothetical protein
	81	325	122488	-	GTGGAATAGCACTTGCC	-	MT0112	cation transporter E1-E2 family ATPase
	81	60	122247	+	ATGGCCCGCTGGATGCA	-	MT0113	hypothetical protein
2	PKS		485426-491388					
3	PKS		1313164-1328792					
	Score	Distance From ORF	N	Strand	Sequence	Gene	Synonym	Protein Name
	78	341	1327340	-	CTGGCTCAGACCCTGCG	-	MT1222	acyl-CoA synthetase
4	PKS		1865066-1886258					
	Score	Distance From ORF	N	Strand	Sequence	Gene	Synonym	Protein Name
	78	121	1866089	+	CTGGCACTGATGCTGGA	-	MT1701	polyketide synthase
	76	370	1878562	+	TTGCCACCGACCTTGAT	-	MT1704	polyketide synthase
5	PKS		2294295-2309319					
6	NRPS/ PKS		2647576-2673168					
	Score	Distance From ORF	N	Strand	Sequence	Gene	Synonym	Protein Name
	75	344	2649233	-	CCGGCCCCGCCGTTGCC	-	MT2439	hypothetical protein
	83	331	2651649	-	GAGGAACGCCTTTTGCT	dnaJ-2	MT2442	chaperone protein DnaJ
	78	334	2673502	-	CTGGCAGCCGTTTGGCA	pchE	MT2451	dihydroaeruginoic acid synthetase
7	PKS		3236513-3319822					
	Score	Distance From ORF	N	Strand	Sequence	Gene	Synonym	Protein Name
	76	385	3236128	+	TCGGCACGCCTCGGGCT	-	MT2998	thioesterase
	75	1	3237290	+	TTCGCTGGACATTTGCT	-	MT2998.1	hypothetical protein
	77	349	3256220	+	CTGGGACTCGACTCGCT	-	MT3004	polyketide synthase
	78	23	3292184	-	GTGGCCGACCGTTTGCG	-	MT3021.1	polyketide synthase
	77	323	3296988	-	ATCGCACGGCGCTGGCG	-	MT3023	acyl-CoA synthetase
	79	490	3298272	+	TTGGCATCGATCTTGGG	-	MT3026	methyltransferase, putative
	78	477	3298285	+	TGGGCACGAATTTCGCC	-	MT3026	methyltransferase, putative
	77	104	3298658	+	CTGCCATACTCCTTGCC	-	MT3026	methyltransferase, putative
	82	353	3303494	+	GTGGCATGCTCATTCCT	-	MT3031	glycosyl transferase
	76	44	3303803	+	GTGGCTGCGGTATTGCC	-	MT3031	glycosyl transferase

	79	276	3306596	-	CTGGCAAGATCTTCGCC	-	MT3034	UDP-glucoronosyl and UDP-glucosyltransferase family protein
	75	379	3308002	-	TTGGCCAGCTAGTTACT	-	MT3036	hypothetical protein
	76	106	3308155	-	GTGGCAGGGGCCGTGAT	-	MT3037	hypothetical protein
	89	439	3311348	+	CTGGCACGCTACATGCA	-	MT3041.1	hypothetical protein

NC_008611.1 Mycobacterium ulcerans Agy99, complete genome

1	PKS		1803923-1835953					
	Score	Distance From ORF	N	Strand	Sequence	Gene	Synonym	Protein Name
	77	280	1820828	+	CTGGACAAGTTATTGCT	pks11	MUL_1656	chalcone synthase, Pks11
	75	91	1821017	+	ATGTCACATGTGTTGAA	pks11	MUL_1656	chalcone synthase, Pks11
	80	234	1834666	-	ATGGTGCCGCAGTTGCT	-	MUL_1664	hypothetical protein
	75	443	1834875	-	ACGGCACAGCATTTTCG	-	MUL_1664	hypothetical protein
	75	322	1835436	-	CTGGATCCCGTTTTGCG	-	MUL_1665	antibiotic resistance ABC transporter, efflux protein
2	PKS		2211634-2271101					
	Score	Distance From ORF	N	Strand	Sequence	Gene	Synonym	Protein Name
	75	133	2213368	+	CGGGCACACCGAGTGCG	-	MUL_2003	chorismate pyruvate-lyase
	78	133	2216250	+	GTGGCCGAACGGTTGCG	pks15/1	MUL_2005	polyketide synthase Pks15/1
	79	305	2228469	+	GTGGCTCAACGGTGGCT	-	MUL_2009	methyltransferase
	75	67	2230104	+	GTGTCACACCGGTTGAA	mas	MUL_2010	multifunctional mycocerosic acid synthase membrane-associated Mas
	78	354	2240002	-	ACGGCAAGATCATTGCC	drrB	MUL_2013	daunorubicin-DIM-transport integral membrane protein ABC transporter DrrB
	76	153	2240793	-	CTGGCGCGGTCGTTCCC	drrA	MUL_2014	daunorubicin-DIM-transport ATP-binding protein ABC transporter DrrA
	77	288	2240928	-	TTGGCAATGCCGGTGCG	drrA	MUL_2014	daunorubicin-DIM-transport ATP-binding protein ABC transporter DrrA
	84	178	2245274	-	GTGACACTGATTTTGCA	ppsE	MUL_2015	phenolpthiocerol synthesis type-I polyketide synthase PpsE
	80	247	2250769	-	ATGGCCCTCGAGTTGCG	ppsD	MUL_2016	phenolpthiocerol synthesis type-I polyketide synthase PpsD
	82	34	2257161	-	TTGGAACAGCTTTCGCA	ppsC	MUL_2017	phenolpthiocerol synthesis type-I polyketide synthase PpsC
	76	121	2266630	-	GTGGCCGACGTCGTGCT	ppsA	MUL_2019	phenolpthiocerol synthesis type-I polyketide synthase PpsA
3	NRPS		2921595-2959529					
	Score	Distance From ORF	N	Strand	Sequence	Gene	Synonym	Protein Name
	79	76	2921519	+	TCGGCACGATAATTCCT	-	MUL_2620	hypothetical protein
	80	216	2929670	-	ATGTCGCGGGCATTGCT	-	MUL_2625	hypothetical protein
	81	92	2930909	-	CTGGCGCAGACCGTGCA	fadE25	MUL_2626	acyl-CoA dehydrogenase FadE25

	75	87	2934800	-	GTCGCGCGGGGATTGCC	birA	MUL_2631	bifunctional protein BirA
	76	286	2952542	-	GTGGCAGACCTGCGGCT	sigF	MUL_2640	RNA polymerase sigma factor SigF
	83	205	2953689	-	GTGGCACGGCGCCGGCT	usfY	MUL_2642	hypothetical protein
4	NRPS/ PKS		4029686-4059363					
	Score	Distance From ORF	N	Strand	Sequence	Gene	Synonym	Protein Name
	76	385	4033562	-	TTGGACTACGGCTTGCT	mbtB	MUL_3632	phenyloxazoline synthase MbtB
	75	397	4038148	-	CTGGCCCCACCGCTGCC	mbtF	MUL_3633	non-ribosomal peptide synthetase MbtF
	75	61	4042974	+	GTTGAATGGCTGTTGCG	-	MUL_3635	hypothetical protein
	76	245	4044260	+	CTGGCACTTCGCGTGAT	mbtC	MUL_3637	polyketide synthase MbtC
	76	56	4050456	-	ATGGCAACGCTATTACC	mbtG	MUL_3640	lysine-N-oxygenase MbtG
	80	31	4053325	+	ATCGCAGGCTCATTGCA	-	MUL_3643	short-chain membrane-associated dehydrogenase
	75	317	4057245	-	CTGGCTCGGCAGCGGCA	desA1_1	MUL_3646	acyl-[acyl-carrier protein] desaturase DesA1_1
5	NRPS		4833556-4838804					
6	NRPS		5377263-5405727					
	Score	Distance From ORF	N	Strand	Sequence	Gene	Synonym	Protein Name
	77	255	5378620	+	ACGGCACCAGCACTGCT	-	MUL_4854	oxidoreductase
	75	467	5379274	+	CTGGCCAAGGTTTGGCA	-	MUL_4855	hypothetical protein
	84	398	5391216	-	TCGGCCCGGATATTGCT	-	MUL_4861	hypothetical protein
	83	199	5396076	-	ATGGCAATGACGTTGCG	-	MUL_4867	hypothetical protein
7	PKS		5523988-5532849					

NC_006361.1 Nocardia farcinica IFM 10152, complete genome

1	PKS		184410-208907					
	Score	Distance From ORF	N	Strand	Sequence	Gene	Synonym	Protein Name
	78	160	207612	+	GTGGCGCCGTCGGTGCT	-	nfa1980	hypothetical protein
2	NRPS		742402-783087					
	Score	Distance From ORF	N	Strand	Sequence	Gene	Synonym	Protein Name
	78	5	782060	+	GTGGCACGCGTGCCGCG	-	nfa7210	putative esterase
3	NRPS/ PKS		809229-843740					
	Score	Distance From ORF	N	Strand	Sequence	Gene	Synonym	Protein Name
	75	426	815841	-	GTGGGCCAACAACTGCT	-	nfa7550	putative ATP-dependent protease
	78	324	816789	-	CTGGCCTGGGCGGTGCT	amiE	nfa7560	acylamide amidohydrolase
	78	239	820253	-	TTGGCGCAACACGTGCG	-	nfa7600	hypothetical protein
	80	307	820321	-	CTGGCGAGCCAGTTGCC	-	nfa7600	hypothetical protein
	76	231	827277	+	CTGGCCGAGCGGATGCT	nbtD	nfa7660	putative non-ribosomal peptide synthetase

	76	64	837122	+	CTGGCCGAGCTGTGGCA	nbtF	nfa7680	putative non-ribosomal peptide synthetase
4	NRPS		1219367-1235955					
	Score	Distance From ORF	N	Strand	Sequence	Gene	Synonym	Protein Name
	77	475	1218892	+	CTGGCCCGTTGCCTGCG	-	nfa11040	hypothetical protein
5	NRPS		2965856-2985159					
6	PKS		3189119-3223839					
	Score	Distance From ORF	N	Strand	Sequence	Gene	Synonym	Protein Name
	76	247	3191679	-	CTGGCGCTGGTGCTGCG	-	nfa30090	putative iron-sulfur oxidoreductase
	75	103	3193577	-	CTAGCGCGCGGATTGCG	-	nfa30110	putative monooxygenase
	75	380	3193854	-	GTGGCATCCTGGTGGCC	-	nfa30110	putative monooxygenase
	76	406	3195498	-	CTGGGGCAGTGATTGCG	-	nfa30120	acyl-CoA synthetase
	75	62	3198518	+	CCGGCAAGCGCGATGCT	-	nfa30170	hypothetical protein
	76	56	3202150	+	ATTGCTGGAATATTGCT	-	nfa30220	putative sigma factor
	75	459	3212616	-	CTGGCGGACCAACTGCT	-	nfa30250	putative polyketide synthase
	76	221	3220860	-	CTGGACTTCCTGTTGCT	-	nfa30320	hypothetical protein
7	NRPS		3297940-3312573					
	Score	Distance From ORF	N	Strand	Sequence	Gene	Synonym	Protein Name
	78	179	3299108	-	GTGGCCGGGACCGTGCA	-	nfa31160	putative ABC transporter ATP-binding protein
	77	296	3299225	-	CTGGCGCGGGACTTCCG	-	nfa31160	putative ABC transporter ATP-binding protein
8	PKS		4478704-4491458					
	Score	Distance From ORF	N	Strand	Sequence	Gene	Synonym	Protein Name
	76	338	4489377	-	ATGGCACAGGTGGCGCC	-	nfa43240	putative polyketide synthase
	77	493	4491951	-	CCGGCAACGCACTTGCC	-	nfa43260	putative transporter
9	NRPS		5244714-5291138					
10	NRPS		5318160-5361768					
11	PKS		5919882-5934805					
	Score	Distance From ORF	N	Strand	Sequence	Gene	Synonym	Protein Name
	80	97	5923205	+	GCGGCACGTGAGCTGCT	-	nfa55870	hypothetical protein
	79	153	5926963	-	GTGGCACGACCGTGCCA	-	nfa55900	long-chain-fatty-acid--CoA ligase
	78	189	5926962	+	ATGGCACGGTCGTGCCA	-	nfa55910	putative dioxygenase

NC_009142.1 Saccharopolyspora erythraea NRRL 2338, complete genome

1	PKS		14813-51921					
	Score	Distance From ORF	N	Strand	Sequence	Gene	Synonym	Protein Name

	81	175	15677	-	TTGGCCGGGCCGTTGCG	-	SACE_0012	TetR family transcriptional regulator
	76	129	49054	+	GCGGCCCGGCTCTTTCT	-	SACE_0029	RNA polymerase sigma factor
2	PKS		778214-832825					
	Score	Distance From ORF	N	Strand	Sequence	Gene	Synonym	Protein Name
	76	328	783452	-	CTGGCGCCTACGGTGCA	eryCIV	SACE_0716	eryCIV NDP-6-deoxyhexose 3,4-dehydratase
	76	128	784712	-	CTGGCACAGGTGATCCG	eryBVI	SACE_0717	NDP-4-keto-6-deoxy-glucose 2,3-dehydratase
	77	362	787113	-	GTGGCACGACGGCGGCG	eryBV	SACE_0719	6-DEB TDP-mycarosyl glycosyltransferase
	81	224	831762	-	GAGGTACGCGGCTTGCA	ermE	SACE_0733	N-6-aminoadenine-N-methyltransferase,erythromycin resistance
3	NRPS		1436967-1445722					
4	PKS		2528343-2549947					
	Score	Distance From ORF	N	Strand	Sequence	Gene	Synonym	Protein Name
	76	242	2528101	+	GTGGACCGGCAGTTGGT	cmtC	SACE_2339	trehalose corynomycolyl transferase C
5	PKS		2795362-2818679					
	Score	Distance From ORF	N	Strand	Sequence	Gene	Synonym	Protein Name
	76	132	2798967	-	TCGGCACGCGACTTGGG	-	SACE_2592	glucose-methanol-choline oxidoreductase
	80	116	2800190	-	CTGGCGCACCGGGTGCT	-	SACE_2593	glycopeptide antibiotics resistance protein
6	NRPS/ PKS		2838066-2886848					
	Score	Distance From ORF	N	Strand	Sequence	Gene	Synonym	Protein Name
	81	433	2844648	-	GTGGCGCAGGTGCTGCA	-	SACE_2619	non-ribosomal peptide synthetase
	79	322	2855647	-	GAGGCACTGCGCCTGCT	-	SACE_2627	isopentenyl-diphosphate delta-isomerase II 2
	77	316	2883545	-	GTGGCGCTGGCGCTGCA	-	SACE_2635	major facilitator superfamily transporter phthalate permease
7	NRPS		2944491-2974637					
	Score	Distance From ORF	N	Strand	Sequence	Gene	Synonym	Protein Name
	75	112	2962692	+	GTGGAACGGCCGCGGCA	-	SACE_2698	putative regulatory protein
	76	63	2974297	+	GTGGCACGCGCGGCGCC	-	SACE_2709	hypothetical protein
8	PKS		3127652-3157229					
	Score	Distance From ORF	N	Strand	Sequence	Gene	Synonym	Protein Name

	76	284	3131486	-	TCGGCACGGTGTTCGCC	rifI	SACE_2867	aminoquinate/shkimate dehydrogenase
	79	195	3138945	+	GTGCCATTCCGGTTGCT	-	SACE_2875	modular polyketide synthase
	77	158	3138982	+	CAGGCATCGAGGTTGCG	-	SACE_2875	modular polyketide synthase
	78	428	3151884	+	CTGGCGCACGTGTTCCT	mhpA	SACE_2877	3-(3-hydroxyphenyl)propionate hydroxylase
	77	240	3152072	+	CTGGCACGGCCTCTACC	mhpA	SACE_2877	3-(3-hydroxyphenyl)propionate hydroxylase
9	NRPS		3301269-3308199					
	Score	Distance From ORF	N	Strand	Sequence	Gene	Synonym	Protein Name
	77	55	3307140	+	ACGGCGCGGATCGTGCT	-	SACE_3016	SyrP-like protein
10	NRPS		3326161-3344964					
	Score	Distance From ORF	N	Strand	Sequence	Gene	Synonym	Protein Name
	85	242	3325919	+	GTGGCCAGGCGGTTGCT	-	SACE_3033	putative peptide monooxygenase
11	PKS		4545937-4616845					
	Score	Distance From ORF	N	Strand	Sequence	Gene	Synonym	Protein Name
	75	59	4554457	+	TTGACTAGAGTGTTGCT	-	SACE_4130	hypothetical protein
	78	290	4559173	+	ATGGCCTCGCGGTTGCC	-	SACE_4132	acyl carrier protein
	86	49	4563179	+	TGGGCACCGCCCTTGCT	-	SACE_4135	MaoC-like dehydratase
	80	259	4572145	-	CTGGCCGGGCACCTGCT	-	SACE_4138	type I PKS modular polyketide synthas
	78	417	4588385	-	GTGGAACAGGCCTTCCA	-	SACE_4139	type I modular polyketide synthase
	75	146	4612861	-	GTGGCGCTGCGCGTGCC	-	SACE_4142	cytochrome P450 monooxygenase
	75	251	4612966	-	CTGGCTCGGATGGAGCT	-	SACE_4142	cytochrome P450 monooxygenase
	83	27	4614155	-	ATGGCAGGGAAGTTCCT	-	SACE_4143	cytochrome P450 monooxygenase
12	NRPS		4759004-4809694					
	Score	Distance From ORF	N	Strand	Sequence	Gene	Synonym	Protein Name
	78	475	4772132	-	CTGGCCCGGGCGCTGCG	-	SACE_4285	AMP-dependent synthetase and ligase
	76	498	4772155	-	GTGGCGCGCTGCTGGCC	-	SACE_4285	AMP-dependent synthetase and ligase
13	PKS		4821866-4828194					
14	PKS		4992823-5006652					
	Score	Distance From ORF	N	Strand	Sequence	Gene	Synonym	Protein Name
	75	183	4996606	-	GTGGACCGCCCGGTGCA	-	SACE_4472	NAD-dependent epimerase/dehydratase
	79	82	4997888	-	GTGGCGCTGAAGCTGCT	qor	SACE_4474	quinone oxidoreductase

15	PKS		5102367-5117363					
	Score	Distance From ORF	N	Strand	Sequence	Gene	Synonym	Protein Name
	79	327	5106524	+	CTCGCACGCGGCCTGCA	-	SACE_4566	hypothetical protein
	78	359	5112790	+	GTGCCGCAACGTTTGCT	pteA1	SACE_4574	modular polyketide synthase
16	PKS		5932907-5941187					
	Score	Distance From ORF	N	Strand	Sequence	Gene	Synonym	Protein Name
	78	213	5933803	+	CGGCCACGCCGGTTGCA	chlB1	SACE_5308	iterative type I polyketide synthase
17	PKS		6225358-6248235					
	Score	Distance From ORF	N	Strand	Sequence	Gene	Synonym	Protein Name
	78	287	6239374	-	TCGGCACCGCTCTTCCA	-	SACE_5540	3-oxoacyl-(acyl carrier protein) synthase III

NC_009380.1 Salinispora tropica CNB-440, complete genome

1	PKS		671414-679577					
	Score	Distance From ORF	N	Strand	Sequence	Gene	Synonym	Protein Name
	75	214	671873	+	TTGGAACGAAACGAGCT	-	Strop_0597	hypothetical protein
	83	374	672939	+	ATGGCTGGCACGTTGCT	-	Strop_0598	beta-ketoacyl synthase
2	NRPS/ PKS		144505-1172682					
	Score	Distance From ORF	N	Strand	Sequence	Gene	Synonym	Protein Name
	79	296	1150310	-	CTGGCATGCCTCCGGCA	-	Strop_1020	MbtH domain-containing protein
	82	296	1164400	+	GTGGCATGGCCTATGCC	-	Strop_1028	hypothetical protein
	76	157	1164539	+	CTGGCCGGCCGATCGCT	-	Strop_1028	hypothetical protein
	78	220	1165260	+	CTGCCACGGCTGCTGCC	-	Strop_1029	cyclase family protein
	81	438	1167460	-	CTGGCACAGATAGCGCT	-	Strop_1030	regulatory protein, LuxR
3	PKS		2502319-2520690					
	Score	Distance From ORF	N	Strand	Sequence	Gene	Synonym	Protein Name
	76	33	2505275	-	TGGGCACAAAGCCTGCG	-	Strop_2212	TAP domain-containing protein
	75	447	2507023	-	GCGGCACCCTGGTTGAT	-	Strop_2213	hypothetical protein
	78	480	2513894	-	ATGGGACTGACCTCGCT	-	Strop_2220	hypothetical protein
	83	411	2514406	+	CTGGTAGAGCAGTTGCA	-	Strop_2222	dTDP-glucose 4,6-dehydratase
	75	1	2517088	+	TGGGCACGGGTCCGGCA	-	Strop_2224	beta-ketoacyl synthase
	76	463	2519253	+	GCGGCCGGCCCGTTGCT	-	Strop_2227	hypothetical protein
4	PKS		2795012-2816666					
	Score	Distance From ORF	N	Strand	Sequence	Gene	Synonym	Protein Name
	75	433	2798754	-	CTGGTCGAACTCTTGCG	-	Strop_2487	dehydrogenase, E1 component

	76	124	2804213	-	GAGGCCGACCTGTTGCT	-	Strop_2492	AMP-dependent synthetase and ligase
	81	84	2805809	-	CTGGCCTGGCTTGTGCA	-	Strop_2494	beta-ketoacyl synthase
	80	205	2809136	-	GTGGCGACCGTGTTGCT	-	Strop_2498	antibiotic biosynthesis monooxygenase
	77	357	2812320	-	GTGGCCGAGATCCTGCA	-	Strop_2501	cupin 2 domain-containing protein
	77	112	2814752	-	CTGGCGCACCTGGTGCC	-	Strop_2504	acetyl-CoA carboxylase, biotin carboxyl carrier protein
5	NRPS/ PKS		2956815-2996840					
	Score	Distance From ORF	N	Strand	Sequence	Gene	Synonym	Protein Name
	78	393	2974666	-	TTGGCGCAGCTCTGGCC	-	Strop_2648	methyltransferase type 12
	80	85	2977211	-	CTGGCACCCGCCGTGCC	-	Strop_2650	ABC transporter related
	75	466	2979339	-	CTGGCGTTGACCCTGCA	-	Strop_2652	binding-protein-dependent transport systems inner membrane component
	77	150	2996990	-	TTGGCCAAGCAAGTGCT	-	Strop_2659	hypothetical protein
6	PKS		3020222-3042851					
	Score	Distance From ORF	N	Strand	Sequence	Gene	Synonym	Protein Name
	76	204	3023263	-	GTGGCCTGGTTCTGGCC	-	Strop_2686	alpha amylase, catalytic region
	75	322	3025145	-	TTGGCATGTGGTGCGCT	-	Strop_2688	helix-turn-helix domain-containing protein
	75	330	3026030	-	GCGGCACCGGCGTTGTT	-	Strop_2689	hypothetical protein
	76	207	3027623	-	ACGGCCCGTCTCGTGCT	-	Strop_2691	hypothetical protein
	81	377	3031869	-	TTGGCAGGGCCGGTGCC	-	Strop_2695	flavin reductase domain-containing protein
	77	192	3037925	-	CTGGCGCTGGCCTCGCA	-	Strop_2697	beta-ketoacyl synthase
	76	235	3043086	-	ATGGCACGGGTGTTCTC	-	Strop_2701	hypothetical protein
7	NRPS/ PKS		3112454-3249126					
	Score	Distance From ORF	N	Strand	Sequence	Gene	Synonym	Protein Name
	75	422	3116359	+	ACGGCATGAGCGTCGCA	-	Strop_2766	regulatory protein, TetR
	75	117	3141159	-	CTGGTCCGGCTGGTGCC	-	Strop_2770	cytochrome P450
	80	282	3142490	+	CTGGCCCGGACCGTGCG	-	Strop_2772	aminotransferase, class I and II
	76	440	3143585	+	CTGGCCGGCCCGATGCA	-	Strop_2773	hypothetical protein
	76	344	3143681	+	GTGGCCGGCGCCGTGCA	-	Strop_2773	hypothetical protein
	77	410	3144589	+	TTGGCGCAGCGGCTGCG	-	Strop_2774	AMP-dependent synthetase and ligase
	79	25	3160851	-	CTGGCAACGAGTTTGGT	-	Strop_2778	beta-ketoacyl synthase
	85	327	3165966	-	CTGGCCCGGCACCTGCT	-	Strop_2779	acyl transferase domain-containing protein
	76	471	3188122	-	GCGGCACGCCGATCGCC	-	Strop_2785	hypothetical protein
	77	312	3190952	-	AGGGCAAGCGGATTGCC	-	Strop_2787	AMP-dependent synthetase and ligase
	75	112	3193950	-	CTGGCCGCGACCCTGCT	-	Strop_2791	alpha/beta hydrolase fold
	75	303	3194141	-	CTGGCCGACCCGCTGCT	-	Strop_2791	alpha/beta hydrolase fold
	78	427	3198676	-	TTGGCACTCTCATGGCC	-	Strop_2795	beta-ketoacyl synthase
	78	168	3198676	-	TTGGCACTCTCATGGCC	-	Strop_2796	phosphopantetheine-binding
	79	373	3198881	-	GCGGCACTGCTCTGGCT	-	Strop_2796	phosphopantetheine-binding

	75	470	3198978	-	CGGGAACGAGCCCTGCT	-	Strop_2796	phosphopantetheine-binding
	75	370	3201972	-	CAGGCACGGCTGCGGCA	-	Strop_2800	L-carnitine dehydratase/bile acid-inducible protein F
	76	7	3203132	+	CTGGCACTCTGGAGGCT	-	Strop_2803	AMP-dependent synthetase and ligase
	78	466	3212075	+	CTGGCGCTGGAGGTGCT	-	Strop_2809	hypothetical protein
	78	205	3222961	-	CGGGCACTGCGGGTGCT	-	Strop_2817	thioesterase
	75	410	3224293	-	GTGGCTCGACTGGTACT	-	Strop_2818	thiazolinyl imide reductase
	78	126	3241571	-	GCGCCACAGGACTTGCT	-	Strop_2825	cytochrome P450
8	PKS		3475927-3510099					
9	NRPS		4996301-5013814					
	Score	Distance From ORF	N	Strand	Sequence	Gene	Synonym	Protein Name
	80	426	4997797	-	GTGGCCCCGCTCTTCCT	-	Strop_4414	methyltransferase type 12
	84	166	5003350	-	ATGGCCCGCGCGTTGCG	-	Strop_4416	amino acid adenylation domain-containing protein
	79	496	5009133	-	CTGGCCCGGCGGCTGCC	-	Strop_4419	kynurenine 3-monooxygenase

NC_003155.4 Streptomyces avermitilis MA-4680, complete genome

1	PKS		113361-118594					
2	PKS		486648-567017					
	Score	Distance From ORF	N	Strand	Sequence	Gene	Synonym	Protein Name
	76	157	490476	-	TTGGCAAGAAAGCGGCA	pteF	SAV_409	LuxR family transcriptional regulator
	77	370	567387	-	CTGGCGAGATTTTCGCA	pteA1	SAV_419	modular polyketide synthase
3	NRPS		751402-762285					
4	NRPS/ PKS		991484-1042269					
	Score	Distance From ORF	N	Strand	Sequence	Gene	Synonym	Protein Name
	75	136	992855	-	CGGGTGCGTCTGTTGCT	cyp4	SAV_838	cytochrome P450
	77	39	1002360	-	CTGGCCCGGCGGTTGGC	-	SAV_845	modular polyketide synthase
	75	51	1021253	-	CTGGCACACAAAGAGCG	nrps7-8	SAV_855	non-ribosomal peptide synthetase
	76	223	1022254	-	TTGGCCGAGGACGTGCT	-	SAV_856	thioesterase
	87	293	1024835	+	CTGGCACGGGCCTTCCA	nrps7-10	SAV_859	non-ribosomal peptide synthetase
	81	433	1026118	+	TTGGCGCGGGCGGTGCA	nrps7-11	SAV_860	non-ribosomal peptide synthetase
	79	488	1036207	+	GTGGCAGGCGAACTGCC	-	SAV_867	ABC transporter ATP-binding protein
5	PKS		1132045-1212960					
	Score	Distance From ORF	N	Strand	Sequence	Gene	Synonym	Protein Name
	75	76	1168427	+	ATGGCACCGCACCCGCC	aveC	SAV_940	AveC
	75	111	1208710	-	CAGGCCCGCACCCTGCT	aveBV	SAV_949	dTDP-4-keto-6-deoxyhexose 3,5-epimerase
	76	410	1212380	-	ACGGCCCGGATACTGCA	aveBVIII	SAV_952	dTDP-4-keto-6-deoxy-L-hexose 2,3-reductase

	77	281	1211927	+	GGAGCACTGCTGTTGCT	aveG	SAV_953	thioesterase
	75	114	1212094	+	GTGGGACACGGGCTGCC	aveG	SAV_953	thioesterase
6	PKS		1549424-1554224					
7	PKS		1893266-1913284					
	Score	Distance From ORF	N	Strand	Sequence	Gene	Synonym	Protein Name
	78	263	1910520	-	GCGGTACGCCAGTTGCC	pks2-2	SAV_1551	modular polyketide synthase
	76	443	1912725	-	CTGGCCCTCCTTGTGCG	pks2-3	SAV_1552	non-ribosomal peptide synthetase
	76	354	1912250	+	CTGGTGCGAGGCGTGCT	-	SAV_1553	methyltransferase
8	PKS		2773878-2784841					
9	PKS		2877941-2894413					
10	PKS		2896543-2914291					
	Score	Distance From ORF	N	Strand	Sequence	Gene	Synonym	Protein Name
	81	339	2904983	-	CTGGCACCGCTCACGCT	-	SAV_2378	polyketide oxidase/hydroxylase
	75	296	2906293	+	ACGGCCCGCCCGCTGCT	-	SAV_2381	TetR family transcriptional regulator
11	PKS		3477658-3493243					
	Score	Distance From ORF	N	Strand	Sequence	Gene	Synonym	Protein Name
	77	112	3491806	-	CTGGCCCTGTTCCTGCC	-	SAV_2846	hypothetical protein
	75	159	3492602	-	GAGGCACCCCTTCTGCG	-	SAV_2847	hypothetical protein
	77	210	3493453	-	ACGGCCAGCACCTTGCT	-	SAV_2849	hypothetical protein
12	PKS		3534525-3634592					
	Score	Distance From ORF	N	Strand	Sequence	Gene	Synonym	Protein Name
	77	17	3630164	-	ATGGCGCGGCTCGCGCA	olmRII	SAV_2901	LuxR family transcriptional regulator
13	NRPS		3930306-3937493					
14	NRPS		3980213-3994940					
	Score	Distance From ORF	N	Strand	Sequence	Gene	Synonym	Protein Name
	76	11	3985530	-	CTGGAAGGTACCGTGCA	nrps1-1	SAV_3197	non-ribosomal peptide synthetase
	77	263	3989602	-	CGGGCACGGGACTGGCC	nrps1-2	SAV_3198	non-ribosomal peptide synthetase
15	NRPS		4494250-4540450					
	Score	Distance From ORF	N	Strand	Sequence	Gene	Synonym	Protein Name
	79	266	4496242	-	CTGGCGCAGATCCTGCC	pptA4	SAV_3637	phosphopantetheinyl transferase
	83	421	4500418	-	CTGGCCCGGGAACTGCT	-	SAV_3641	ornithine carbamoyltransferase

	80	163	4516448	-	CTGGCACGCCGAGGGCA	nrps2-2	SAV_3643	non-ribosomal peptide synthetase
	76	214	4521073	-	GCGGCACAGCTGTTCCC	cysK2	SAV_3648	cysteine synthase
	77	416	4521275	-	CTGGCGCGCAGCTCGCC	cysK2	SAV_3648	cysteine synthase
	76	211	4523419	-	CTGGCCAGGCTGCTGCG	fadE25	SAV_3650	acyl-CoA dehydrogenase
	75	154	4528046	-	GTGGCCAAGGAGATGCT	-	SAV_3652	isomerase
	75	388	4536085	-	ATGGCGCTGCGGGTGCG	-	SAV_3661	hypothetical protein
16	PKS		8553602-8561604					
	Score	Distance From ORF	N	Strand	Sequence	Gene	Synonym	Protein Name
	75	252	8560553	-	CTGGCCCGGACCGAGCT	-	SAV_7185	UDP-glucose:sterol glucosyltransferase
17	PKS		8776963-8789766					

NC_003888.3 Streptomyces coelicolor A3(2), complete genome

1	PKS	104989-119654
2	NRPS	513989-526783
3	PKS	1335793-1343779
4	NRPS	3543335-3585724
5	PKS	5508078-5508078
6	PKS	6432566-6465258
7	PKS	6895193-6961810
8	NRPS	7106284-7116513
9	PKS	7590412-7602007
10	NRPS	8506283-8523749

NC_010572.1 Streptomyces griseus subsp. griseus NBRC 13350, complete genome

1	PKS		294490-307047					
2	NRPS		480685-538065					
	Score	Distance From ORF	N	Strand	Sequence	Gene	Synonym	Protein Name
	80	142	535949	-	CTGGCCCCGGTGCTGCT	-	SGR_453	iron ABC transporter
	75	367	536174	-	CTGGCCCGGCCGCTGAA	-	SGR_453	iron ABC transporter
3	NRPS		659688-691515					
	Score	Distance From ORF	N	Strand	Sequence	Gene	Synonym	Protein Name
	80	48	664691	+	GTGGCGTGCCTTCTGCA	-	SGR_578	putative SyrP-like protein
	76	167	676251	+	CTGGCCCGGCTGGAGCT	-	SGR_583	putative NRPS
	82	298	685576	-	CTGGCCCGCGCTCTGCT	-	SGR_588	hypothetical protein
4	PKS		704021-717508					

	Score	Distance From ORF	N	Strand	Sequence	Gene	Synonym	Protein Name
	76	480	703541	+	GTGCGACGGCGGTTGCC	tebC	SGR_604	putative enediyne biosynthesis protein
	76	474	704168	+	AGGGCACCCGAGGTGCT	unbL1	SGR_605	putative enediyne biosynthesis protein
	79	329	708297	+	GTGGCACGCCAGTGGGA	unbU	SGR_608	putative enediyne biosynthesis protein
5	NRPS		762595-774309					
6	PKS		937762-968455					
	Score	Distance From ORF	N	Strand	Sequence	Gene	Synonym	Protein Name
	78	299	940053	+	CTGCCACAGCCGATGCA	-	SGR_803	putative large multi-functional protein
	81	130	956182	-	CTGGCACGGCAGCTCCA	-	SGR_813	putative FAD-dependent oxidoreductase
7	NRPS		1044652-1071469					
	Score	Distance From ORF	N	Strand	Sequence	Gene	Synonym	Protein Name
	75	294	1046802	-	GGGGCACGCACATCGCC	-	SGR_890	hypothetical protein
	75	437	1060513	-	CTGGCTCACCGGATGCG	-	SGR_896	putative O-methyltransferase
8	NRPS/ PKS		2924650-2941895					
	Score	Distance From ORF	N	Strand	Sequence	Gene	Synonym	Protein Name
	75	374	2930023	+	ATGGTGCGGCAGTTGGA	-	SGR_2485	putative NRPS
	75	470	2935806	+	ATGGCCCATGTGCTGCC	-	SGR_2486	putative aminopeptidase 2
	78	400	2937233	+	ATGGCGCACGCGGTGCA	-	SGR_2487	putative thioesterase
9	NRPS		3048501-3074603					
10	NRPS		3762124-3830096					
11	PKS		7102079-7171428					
	Score	Distance From ORF	N	Strand	Sequence	Gene	Synonym	Protein Name
	76	362	7161653	+	TCGGCACGCTCGGTGCC	pks1-7	SGR_6083	putative type-I PKS
12	PKS		7286041-7354154					
13	PKS		7551740-7619551					
	Score	Distance From ORF	N	Strand	Sequence	Gene	Synonym	Protein Name
	83	426	7551314	+	CTGGCAGGGCAGTTGGA	-	SGR_6359	hypothetical protein
	78	256	7570469	-	CTGGGACTGTCCTCGCT	-	SGR_6360	hypothetical protein
	75	422	7574873	-	GTGGCCGACGTAGTGCA	-	SGR_6365	hypothetical protein
	76	266	7596949	-	CTGGCCGGGCACCTGCG	pks3-3	SGR_6371	putative type-I PKS
	81	244	7619795	-	GTGGCCCAGCGGCTGCA	pks3-1	SGR_6373	putative type-I PKS
14	NRPS		8011976-8029693					

	Score	Distance From ORF	N	Strand	Sequence	Gene	Synonym	Protein Name
	81	164	8012818	+	GCGGCACGATCCGTGCA	-	SGR_6714	putative ABC-type Fe3+-siderophore transporter substrate-binding protein
	85	326	8015602	+	CTGGCCCGGCTGGTGCT	-	SGR_6716	putative NRPS
	77	424	8030117	-	CTGGCACATCGGCTGGT	-	SGR_6718	putative esterase
15	NRPS		8049140-8065069					
16	PKS		8100042-8154770					
	Score	Distance From ORF	N	Strand	Sequence	Gene	Synonym	Protein Name
	78	30	8100012	+	CGGGCAGAGTCCTTGCG	pks4-1	SGR_6776	putative NRPS-type-I PKS fusion protein
	77	383	8110593	+	TTCGCACGGACGGTGCG	pks4-2	SGR_6777	putative type-I PKS
	75	269	8124572	+	TTCGGACGCATCGTGCT	-	SGR_6780	putative malonyl-CoA:ACP transacylase

NC_003228.3 Bacteroides fragilis NCTC 9343, complete genome

1	NRPS		3294535-3306572					
2	NRPS		3635663-3639879					
	Score	Distance From ORF	N	Strand	Sequence	Gene	Synonym	Protein Name
	75	357	3638482	+	GCGGCCGACAGCTTGCT	-	BF3118	hypothetical protein

NC_005125.1 Gloeobacter violaceus PCC 7421, complete genome

1	PKS		2057662-2100787					
	Score	Distance From ORF	N	Strand	Sequence	Gene	Synonym	Protein Name
	76	304	2064191	-	ATGGTCCGCATCTTCCT	-	gll1939	cyclopropane-fatty-acyl-phospholipid synthase
	83	457	2064344	-	CTGGCCCGAGATCTGCT	-	gll1939	cyclopropane-fatty-acyl-phospholipid synthase
	77	280	2065156	-	ATGGCCCAGTTGTGGCC	-	gll1940	fatty acid desaturase
	77	423	2072927	-	GCGGCCCTGGGGTTGCT	-	gll1946	fatty acid desaturase
	81	329	2073836	-	GTGGGGCGGCTTTTGCC	-	gll1947	fatty acid desaturase
	75	166	2078085	-	CCGGCAGCACTCCTGCA	-	gll1950	long-chain fatty-acid-CoA ligase
	76	219	2078138	-	ATGGCTTCGCACTCGCT	-	gll1950	long-chain fatty-acid-CoA ligase
	77	466	2078385	-	GGGGCGCACCCCTTGCG	-	gll1950	long-chain fatty-acid-CoA ligase
	81	411	2079094	-	CTGGAGCGCTGTTTGCA	-	gll1951	phosphopantetheinyltransferase family protein
	76	461	2088597	-	GTGGCTGGACAGTTGAT	-	gll1954	polyketide synthase
	76	146	2088597	-	GTGGCTGGACAGTTGAT	-	gll1955	polyketide synthase
	75	317	2088768	-	ATGGCCCTGATCTTTCG	-	gll1955	polyketide synthase
	75	277	2094691	-	CTGGACCAGTTGCTGCA	-	gll1957	glycolipid synthase
2	PKS		2999221-3029646					
	Score	Distance From ORF	N	Strand	Sequence	Gene	Synonym	Protein Name

	77	396	2998825	+	CTGGCACCAGAAACGCT	-	glr2821	WD repeat-containing protein
	81	288	3009343	-	CTGGCGGGGCCTTTGCG	-	gll2826	hypothetical protein
	77	307	3010641	-	ACGGGACATGATTTGCT	-	gll2827	hypothetical protein
	80	157	3014822	-	CAGGCGCTGATCTTGCT	-	gll2829	polyketide synthase
	79	450	3023483	-	GTGGCACAGACCGTCCT	-	gll2836	alcohol dehydrogenase
	79	174	3023483	-	GTGGCACAGACCGTCCT	-	gsl2837	hypothetical protein
3	PKS		3033832-3050434					
	Score	Distance From ORF	N	Strand	Sequence	Gene	Synonym	Protein Name
	75	221	3040466	+	CTGGCGCAGTGTTTGAG	-	glr2850	thioesterase
	79	41	3040646	+	CTGGCAACTCTGGTGCA	-	glr2850	thioesterase
	78	393	3048049	+	GTGGCCCGTCCGATGCC	-	glr2859	oxidoreductase
4	PKS		4421532-4440489					
	Score	Distance From ORF	N	Strand	Sequence	Gene	Synonym	Protein Name
	75	215	4422391	-	CTGGAAAGTCCATCGCT	-	gll4219	hypothetical protein
	83	331	4422933	-	CTGGCCGGAGGATTGCT	-	gll4220	hypothetical protein
	75	154	4431563	-	CTGGTAGGCCCCATGCG	-	gll4225	glycolipid synthase

NC_010296.1 Microcystis aeruginosa NIES-843, complete genome

1	NRPS/ PKS		2507861-2539146					
	Score	Distance From ORF	N	Strand	Sequence	Gene	Synonym	Protein Name
	77	38	2518358	+	TGGGCGCAAGCCTTGCG	-	MAE_27820	amino acid adenylation
	78	118	2526583	+	TGGGCGCGAGCATTGCG	-	MAE_27860	short-chain dehydrogenase/reductase
	79	235	2537743	+	GTGGCATCGGTTGTGCG	-	MAE_27910	hypothetical protein
2	NRPS/ PKS		3486436-3541027					
	Score	Distance From ORF	N	Strand	Sequence	Gene	Synonym	Protein Name
	79	340	3528001	+	GTGGCAATTGTCCTGCA	mcyF	MAE_38620	McyF protein
	79	114	3540068	+	ATGGCAATATTTCTGCA	mcyJ	MAE_38660	McyJ protein
3	NRPS		5194435-5220124					
	Score	Distance From ORF	N	Strand	Sequence	Gene	Synonym	Protein Name
	75	130	5208380	-	TTGGCATGGTATTCTCA	-	MAE_56550	hypothetical protein
	80	148	5209406	-	TTGGCTCGACCGTTTCT	-	MAE_56560	bacilysin biosynthesis protein BacA-like protein
	80	198	5209456	-	ATGGCTTAGACCTTGCC	-	MAE_56560	bacilysin biosynthesis protein BacA-like protein
4	NRPS		5519975-5552181					

NC_010628.1 Nostoc punctiforme PCC 73102, complete genome

1	PKS		56243-74142					
	Score	Distance From ORF	N	Strand	Sequence	Gene	Synonym	Protein Name

	81	379	58235	-	GTGGCAATATAATTGCC	-	Npun_R0038	short-chain dehydrogenase/reductase SDR
	79	405	60343	-	TTCGCAAAACTATTGCT	-	Npun_R0039	polyketide synthase thioester reductase subunit HglB
	76	291	68964	-	ATGGCCCAAAGTTTTCA	-	Npun_R0042	KR
2	PKS		1549256-1571247					
3	PKS		2521301-2548634					
4	NRPS/ PKS		2648282-2708401					
	Score	Distance From ORF	N	Strand	Sequence	Gene	Synonym	Protein Name
	75	183	2660116	-	TTGTCATAGACATTGCC	-	Npun_R2175	KR
	79	484	2664987	-	ATGGCAGGTATAATGCC	-	Npun_R2178	WD-40 repeat-containing protein
	76	206	2704134	+	GTGATACAGTAATTGCA	-	Npun_F2186	alcohol dehydrogenase
	81	362	2705124	+	CAGGCATTCCTATTGCT	-	Npun_F2187	pyrroline-5-carboxylate reductase
5	NRPS		3054973-3081156					
	Score	Distance From ORF	N	Strand	Sequence	Gene	Synonym	Protein Name
	76	50	3055694	+	TGGGCAAGGATATAGCA	-	Npun_F2460	amino acid adenylation domain-containing protein
	76	93	3065615	+	GTGACTCAAGAGTTGCT	-	Npun_F2462	amino acid adenylation domain-containing protein
6	NRPS/ PKS		3722118-3780010					
7	NRPS/ PKS		3899401-3958883					
	Score	Distance From ORF	N	Strand	Sequence	Gene	Synonym	Protein Name
	75	191	3899210	+	TTCACACGTATGTTGCT	-	Npun_F3155	hypothetical protein
	86	415	3925855	+	ATGGCGTGAATTTTGCT	-	Npun_F3168	hypothetical protein
	76	296	3950844	+	GTGGCTAGGGTGCTGCG	-	Npun_F3173	amino acid adenylation domain-containing protein
	75	148	3950992	+	AAGGCCCTAGTATTGCC	-	Npun_F3173	amino acid adenylation domain-containing protein
	75	54	3955358	+	TAGGGAAGATAATTGCT	-	Npun_F3174	isoprenylcysteine carboxyl methyltransferase
8	PKS		4180429-4210477					
	Score	Distance From ORF	N	Strand	Sequence	Gene	Synonym	Protein Name
	75	482	4181756	-	CTCGAACATTTATTGCT	-	Npun_R3355	thioesterase
	75	474	4185955	+	TTGTCAGGAGGGTTGCC	-	Npun_F3359	beta-ketoacyl synthase
	76	364	4186065	+	CTGGCTAGAGAAATGCT	-	Npun_F3359	beta-ketoacyl synthase
	83	221	4201954	+	ATGGCAGTGATATTGCC	-	Npun_F3363	short-chain dehydrogenase/reductase SDR, HetN
9	NRPS/		4256267-4330313					

	PKS							
	Score	Distance From ORF	N	Strand	Sequence	Gene	Synonym	Protein Name
	81	468	4259675	-	ATGGCAAGAGTTTTGTA	-	Npun_R3419	thioesterase
	81	165	4259675	-	ATGGCAAGAGTTTTGTA	-	Npun_R3420	MbtH domain-containing protein
	76	491	4260403	-	ATTGCACGAGGCTTTCT	-	Npun_R3421	hypothetical protein
	77	499	4280094	-	ACGGCCCGACTGTAGCT	-	Npun_R3429	condensation domain-containing protein
	76	482	4286311	-	TTGGATCGGCGATCGCT	-	Npun_R3430	beta-ketoacyl synthase
	75	239	4307280	-	ATGGCTCTCTTATTGAA	-	Npun_R3436	amino acid adenylation domain-containing protein
	78	331	4307372	-	TTGGCATCCAGCTTGTT	-	Npun_R3436	amino acid adenylation domain-containing protein
	81	358	4321256	-	ATGGCAGCACTTGTGCA	-	Npun_R3448	taurine catabolism dioxygenase TauD/TfdA
10	NRPS/ PKS		8150738-8167006					

NC_008536.1 Solibacter usitatus Ellin6076, complete genome

1	PKS		2480784-2494608					
	Score	Distance From ORF	N	Strand	Sequence	Gene	Synonym	Protein Name
	80	22	2489341	+	CTGGCCGGGAGATTGCG	-	Acid_1973	beta-ketoacyl synthase
	77	463	2493837	+	ACGGCAGACCCATTGCC	-	Acid_1974	putative acyl carrier protein
2	PKS		3884932-3895290					
	Score	Distance From ORF	N	Strand	Sequence	Gene	Synonym	Protein Name
	75	213	3885870	-	GTGGATCGAGTTTCGCT	-	Acid_3076	4'-phosphopantetheinyl transferase
	78	300	3885957	-	ATGGCACCGCGGCTGAT	-	Acid_3076	4'-phosphopantetheinyl transferase
	77	50	3893616	-	CGGGCACGCCGCTTGAG	-	Acid_3077	beta-ketoacyl synthase

NC_009674.1 Bacillus cereus subsp. cytotoxis NVH 391-98, complete genome

1	NRPS		417602-429629					
	Score	Distance From ORF	N	Strand	Sequence	Gene	Synonym	Protein Name
	80	491	422474	+	CTGGTAGAGGATTTGCA	-	Bcer98_0368	amino acid adenylation domain-containing protein
2	NRPS		1197549-1205485					
	Score	Distance From ORF	N	Strand	Sequence	Gene	Synonym	Protein Name
	77	484	1198103	+	ACGTCACGAGGTTTGCA	-	Bcer98_1088	2-nitropropane dioxygenase NPD
3	NRPS		1853248-1878954					
	Score	Distance From ORF	N	Strand	Sequence	Gene	Synonym	Protein Name

	76	310	1855588	+	TTGGAGCGTCTTTGGCT	-	Bcer98_1744	AMP-dependent synthetase and ligase
	80	293	1856961	+	TAGGCGCTGGTGTTGCT	-	Bcer98_1745	hypothetical protein
4	PKS		3245067-3256221					
5	PKS		1781906-1859783					
	Score	Distance From ORF	N	Strand	Sequence	Gene	Synonym	Protein Name
	77	130	1785974	+	CCGGCACCTTAGCTGCT	pksE	BSU17120	enzyme involved in polyketide synthesis
	81	446	1791566	+	CTGCCACAGATATTGCC	pksI	BSU17170	polyketide biosynthesis enoyl-CoA hydratase
6	NRPS		1949682-2002351					
	Score	Distance From ORF	N	Strand	Sequence	Gene	Synonym	Protein Name
	81	138	1956312	-	GTGGAACGGATGCTGCG	yngI	BSU18250	AMP-binding domain protein
	82	303	1956477	-	CTGGCACGGCAAATGGT	yngI	BSU18250	AMP-binding domain protein
	79	340	1957700	-	CGGGCACTGCTCTTGTT	yngJ	BSU18260	acyl-CoA dehydrogenase, short-chain specific
	76	263	1960350	-	ATGGGAAGGCGCCTGCG	yngL	BSU18290	putative integral inner membrane protein
	77	161	1975017	-	ATTGAACACTTTTTGCT	ppsD	BSU18310	plipastatin synthetase
7	NRPS		3280519-3297919					
	Score	Distance From ORF	N	Strand	Sequence	Gene	Synonym	Protein Name
	75	391	3290651	-	GTGGCACCTGTCCAGCG	dhbE	BSU31980	2,3-dihydroxybenzoate-AMP ligase
	82	299	3291784	-	TTGGCCTTGAGCTTGCA	dhbC	BSU31990	isochorismate synthase DhbC
	78	267	3296537	-	CAGGCACAGCGACTGCC	yuiF	BSU32040	amino acid transporter
	80	328	3296598	-	ATGGCAGGAACGTTTCT	yuiF	BSU32040	amino acid transporter
8	NRPS		3952275-3958482					
	Score	Distance From ORF	N	Strand	Sequence	Gene	Synonym	Protein Name
	78	347	3957044	+	ATGGCACCGCCTATGGT	ywaA	BSU38550	branched-chain amino acid aminotransferase

NC_003030.1 Clostridium acetobutylicum ATCC 824, complete genome

1	PKS		3529432-3534819					

NC_009706.1 Clostridium kluyveri DSM 555, complete genome

1	NRPS		1554646-1602560					
	Score	Distance From ORF	N	Strand	Sequence	Gene	Synonym	Protein Name
	75	321	1564208	+	CTCGAATGGATATTGCT	-	CKL_1507	NRPS cyclization domain containing protein
	77	233	1578681	+	GAGGCATAGAACATGCA	-	CKL_1515	hypothetical protein
2	NRPS/ PKS		1818097-1833491					

3	NRPS/ PKS		2414655-2426709					
	Score	Distance From ORF	N	Strand	Sequence	Gene	Synonym	Protein Name
	76	361	2419626	-	TTGATACAGTTTTTGCA	-	CKL_2354	nonribosomal peptide synthetase

NC_008497.1 Lactobacillus brevis ATCC 367, complete genome

1	NRPS		1293616-1300116					
	Score	Distance From ORF	N	Strand	Sequence	Gene	Synonym	Protein Name
	75	122	1297040	-	GTGCAACTGTTGTTGCA	-	LVIS_1323	D-alanyl transfer protein
	81	491	1300607	-	GGGGCACCACGGTTGCG	-	LVIS_1327	D-Ala-teichoic acid biosynthesis protein (putative)

NC_002973.6 Listeria monocytogenes str. 4b F2365, complete genome

1	NRPS		1002948-1007380					

NC_002952.2 Staphylococcus aureus subsp. aureus MRSA252, complete genome

1	NRPS		198393-206225					
	Score	Distance From ORF	N	Strand	Sequence	Gene	Synonym	Protein Name
	81	487	197906	+	CTAGCATGCTTGTTGCT	-	SAR0180	putative non-ribosomal peptide synthetase

NC_003063.2 Agrobacterium tumefaciens str. C58 chromosome linear, complete sequence

1	NRPS		74687-80557					
	Score	Distance From ORF	N	Strand	Sequence	Gene	Synonym	Protein Name
	78	121	79699	-	ATGGCTGCTCGGTTGCT	-	Atu3073	aspartate racemase
	83	322	80879	-	CCGGCACGGCGCTCGCT	-	Atu3074	short chain dehydrogenase
2	NRPS/ PKS		716058-769035					
	Score	Distance From ORF	N	Strand	Sequence	Gene	Synonym	Protein Name
	78	463	729910	-	CTGGTCCGGATTTCGCT	-	Atu3673	siderophore biosynthesis protein
	78	187	729910	-	CTGGTCCGGATTTCGCT	-	Atu3674	hypothetical protein
	85	425	734796	+	ATGGCACGCTACTGGCG	-	Atu3677	polyketide synthase, siderophore biosynthesis protein
	75	293	734928	+	GCGGCGCGTTTCTCGCT	-	Atu3677	polyketide synthase, siderophore biosynthesis protein
	75	28	737448	+	ATGCAAGGCTTCTTGCA	-	Atu3680	putative siderophore biosynthesis protein

	80	62	742445	+	ATGGCAGCGCTGTTGAT	-	Atu3682	non-ribosomal peptide synthetase, siderophore biosynthesis protein
	86	328	765024	+	TTGGCACGGATGTCGCC	fecD	Atu3690	ABC transporter, membrane spanning protein (iron (III) dicitrate)
	79	247	765105	+	CCGGCAGCGCCGTTGCA	fecD	Atu3690	ABC transporter, membrane spanning protein (iron (III) dicitrate)
	75	470	767004	+	CTGGCGCCCCAGGTGCC	-	Atu3692	sigma factor

NC_012988.1 Methylobacterium extorquens DM4, complete genome

1	PKS	2613383-2622159

NC_012988.1 Methylobacterium extorquens DM4, complete genome

1	NRPS		4859328-4866323					
2	NRPS		5649466-5678659					
	Score	Distance From ORF	N	Strand	Sequence	Gene	Synonym	Protein Name
	86	134	5649332	+	CAGGCACAGGCTTTGCT	-	METDI5670	putative UDP-N-acetyl-D-mannosaminuronic acid dehydrogenase
	77	53	5653547	+	ATGGCGCCCGTCTTCCT	-	METDI5674	putative glycosyl transferase
	75	105	5662917	-	CCGGCAACCACCTTGCC	-	METDI5678	hypothetical protein
	75	46	5667529	-	ATGGCAGGTCCGCCGCT	-	METDI5683	hypothetical protein
	75	148	5675379	-	CTGGCACCGCGCTTCGA	murE	METDI5688	UDP-N-acetylmuramoylalanyl-D-glutamate--2, 6-diaminopimelate ligase
	75	355	5677962	-	TTGGTGCGCGGCCTGCA	-	METDI5690	hypothetical protein

NC_003296.1 Ralstonia solanacearum GMI1000 plasmid pGMI1000MP, complete sequence

1	NRPS/ PKS		782247-820777					
	Score	Distance From ORF	N	Strand	Sequence	Gene	Synonym	Protein Name
	76	194	802722	+	CTGGCCGAACTCTGGCA	RSp0642	RS05859	peptide synthetase protein
2	NRPS		1782597-1797233					
	Score	Distance From ORF	N	Strand	Sequence	Gene	Synonym	Protein Name
	76	13	1787443	+	CTCCCACCGCTTTTGCA	RSp1421	RS03121	hypothetical protein
3	NRPS/ PKS		1945814-1979076					
	Score	Distance From ORF	N	Strand	Sequence	Gene	Synonym	Protein Name
	79	452	1949544	-	GTCGCACGCGATCTGCA	tISRso15	RSp1547	ISRSO15-transposase protein
	83	50	1950406	-	GTGGCACGTTCAATGCG	RSp1549	RS02105	hypothetical protein
	82	103	1951581	-	CCGGCATTGCTTTTGCA	RSp155	RS02107	hypothetical protein

					1		
80	248	1951726	-	GTGGCGCTGCTGGTGCA	RSp1551	RS02107	hypothetical protein
75	279	1957715	+	GTGGGATGGGATATGCC	hexR	RSp1556	putative transcription regulation repressor HEXR transcription regulator protein
75	60	1957934	+	GCGGCCCCGCGCTTGCC	hexR	RSp1556	putative transcription regulation repressor HEXR transcription regulator protein
77	289	1960873	-	GCGGCACTGGGCCTGCA	pgl	RSp1558	6-phosphogluconolactonase oxidoreductase protein
77	106	1962311	+	GAGGCACGCAATTCGCG	edd	RSp1560	phosphogluconate dehydratase
78	135	1964859	-	CGGGCATCGTTGTTGCG	RSp1561	RS02117	hypothetical protein
80	247	1964633	+	ATGGCGCGGATGGTGCC	RSp1562	RS02118	hypothetical protein
77	139	1965305	+	TGCGCATGGATTTTGCA	RSp1563	RS02119	hypothetical protein
75	312	1974525	-	CTGGAATGGCGTTAGCG	RSp1571	RS02127	transcriptional regulator transcription regulator protein

NC_008095.1 Myxococcus xanthus DK 1622, complete genome

1	NRPS		1488859-1520119					
	Score	Distance From ORF	N	Strand	Sequence	Gene	Synonym	Protein Name
	82	476	1490324	+	GTGGCGCGGCCCTTGAA	-	MXAN_1276	glutamate-cysteine ligase family 2 protein
	77	461	1497379	-	TCTGGAACGCGTGCAGCAGGGG	-	MXAN_1281	hypothetical protein
	75	1	1500305	+	AGGGCAAGGCATTTCCA	-	MXAN_1284	2-isopropylmalate synthase/homocitrate synthase family protein
	75	159	1503794	+	TTGGACCGGCTCGTGCG	abcA	MXAN_1286	ABC transporter permease/ATP-binding protein
	75	63	1503890	+	TTGGCGCGGAAGGCGCA	abcA	MXAN_1286	ABC transporter permease/ATP-binding protein
	76	304	1505775	+	CATTGCCCGCGCCCTGCTGCTG	-	MXAN_1287	hypothetical protein
	78	233	1505846	+	GAGGCCCAACTGTTGCG	-	MXAN_1287	hypothetical protein
	77	252	1517769	-	CCCTGGCACGTGACGTGTTCAA	-	MXAN_1291	non-ribosomal peptide synthetase
	79	252	1517769	-	CCCTGGCACGTGACGTGTTCA	-	MXAN_1291	non-ribosomal peptide synthetase
	80	250	1517767	-	CTGGCACGTGACGTGTT	-	MXAN_1291	non-ribosomal peptide synthetase
	76	460	1517977	-	GTGGTGCGGCGGATGCA	-	MXAN_1291	non-ribosomal peptide synthetase
	76	455	1518967	-	GCTGGCGCGCGAGGAACTGAAG	-	MXAN_1292	hypothetical protein
	75	300	1520419	-	ACCGGTACACACGTTGGTGGCG	hemG	MXAN_1293	protoporphyrinogen oxidase
	76	336	1520455	-	GGAGACACCGGCGTTGTTGATG	hemG	MXAN_1293	protoporphyrinogen oxidase

2	NRPS				1832662-1910318			
	Score	Distance From ORF	N	Strand	Sequence	Gene	Synonym	Protein Name
	75	298	1834249	-	GTGGCACGTCGCTGGTG	-	MXAN_1560	class I aminotransferase
	79	84	1836766	-	TCTGGCACGGTGACTGCTAAGC	ahpC	MXAN_1564	alkyl hydroperoxide reductase C
	89	85	1836767	-	TTCTGGCACGGTGACTGCTAAG	ahpC	MXAN_1564	alkyl hydroperoxide reductase C
	89	83	1836765	-	CTGGCACGGTGACTGCT	ahpC	MXAN_1564	alkyl hydroperoxide reductase C
	88	85	1836767	-	TTCTGGCACGGTGACTGCTAA	ahpC	MXAN_1564	alkyl hydroperoxide reductase C
	77	420	1841516	-	CGCTGGCGCGGTTGCTGCGCAT	-	MXAN_1567	urea amidolyase-like protein
	79	420	1841516	-	CGCTGGCGCGGTTGCTGCGCA	-	MXAN_1567	urea amidolyase-like protein
	80	418	1841514	-	CTGGCGCGGTTGCTGCG	-	MXAN_1567	urea amidolyase-like protein
	76	61	1842452	-	GGCTGGCACGCCAGCGTCTCCG	-	MXAN_1568	LamB/YcsF family/allophanate hydrolase family protein
	78	61	1842452	-	GGCTGGCACGCCAGCGTCTCC	-	MXAN_1568	LamB/YcsF family/allophanate hydrolase family protein
	75	461	1842852	-	AGCTGCCACGCGCCTGGCGCG	-	MXAN_1568	LamB/YcsF family/allophanate hydrolase family protein
	76	293	1844083	+	ATGACTCGCAGCTTGCG	-	MXAN_1570	class V aminotransferase
	75	376	1847137	+	GCTGGCCTCCACGGTGGTGACG	-	MXAN_1573	AMP-binding domain protein
	77	227	1849786	-	CCTGGCGCGCGCCGGGGTGAAG	-	MXAN_1574	TfoX domain-containing protein
	75	60	1854462	-	CACGGCGCAGCGCTTGCTAAGG	-	MXAN_1578	metallo-beta-lactamase family protein
	75	61	1854463	-	CCACGGCGCAGCGCTTGCTAAG	-	MXAN_1578	metallo-beta-lactamase family protein
	81	59	1854461	-	ACGGCGCAGCGCTTGCT	-	MXAN_1578	metallo-beta-lactamase family protein
	76	104	1854518	+	GGGGCGCAATCCTTGCG	-	MXAN_1579	hypothetical protein
	77	75	1854547	+	GGGGCGCAGGTCTTGCG	-	MXAN_1579	hypothetical protein
	80	180	1860605	-	AAGTGGCGCGGCCGCTGCTTGG	-	MXAN_1584	hypothetical protein
	80	180	1860605	-	AAGTGGCGCGGCCGCTGCTTG	-	MXAN_1584	hypothetical protein
	84	178	1860603	-	GTGGCGCGGCCGCTGCT	-	MXAN_1584	hypothetical protein
	75	146	1861734	+	GTGGGACATCCGGTGCG	-	MXAN_1587	hypothetical protein
	76	130	1866438	-	GGAGGCACGCCGCACGCTGGAA	-	MXAN_1590	putative para-aminobenzoate synthase, component I
	75	190	1871142	-	CCTGCTCTGCACGGAGCTGAAG	-	MXAN_1592	hypothetical protein
	75	203	1871155	-	GGCTGGAAGTGGACCTGCTCTG	-	MXAN_1592	hypothetical protein
	77	203	1871155	-	GGCTGGAAGTGGACCTGCTCT	-	MXAN_1592	hypothetical protein
	78	130	1875006	-	CATGGCACGGGGGACGAAGAGG	-	MXAN_1595	hypothetical protein
	75	428	1875304	-	GGAGGATGACCGGCTGCTGATG	-	MXAN_1595	hypothetical protein
	76	277	1877522	+	GGTGGCCAGCCGGGCGCGGACG	-	MXAN_1599	putative methyltransferase
	75	44	1877755	+	GTGGTGCCGCTGCTGCT	-	MXAN_1599	putative methyltransferase

	75	398	1878919	+	GGAGGCCCGCTCCGAGCTGTGG	-	MXAN_1600	class I aminotransferase
	79	267	1879050	+	GGCTGGCTCGGGCGTTGCGTCC	-	MXAN_1600	class I aminotransferase
	85	270	1879047	+	CTGGCTCGGGCGTTGCG	-	MXAN_1600	class I aminotransferase
	80	268	1879049	+	GGCTGGCTCGGGCGTTGCGTC	-	MXAN_1600	class I aminotransferase
	79	398	1881868	-	GTGGCCCGGGCACTGCG	-	MXAN_1601	fatty acid desaturase family protein
	76	297	1883277	+	GTGGCCTGGGTGGTGCC	-	MXAN_1603	putative non-ribosomal peptide synthetase
	78	262	1889174	+	TGCTGGAAGGGTCGCTGCTGGG	-	MXAN_1605	putative permease
	79	263	1889173	+	TGCTGGAAGGGTCGCTGCTGG	-	MXAN_1605	putative permease
	78	265	1889171	+	CTGGAAGGGTCGCTGCT	-	MXAN_1605	putative permease
	75	256	1890469	+	GGTGGCCGCCTTCATGCTGGAG	-	MXAN_1606	hypothetical protein
	75	260	1890465	+	GTGGCCGCCTTCATGCT	-	MXAN_1606	hypothetical protein
	75	294	1910612	-	GTAGGAACGCACGACCCTCAAG	-	MXAN_1608	hypothetical protein
3	NRPS/ PKS		3251570-3266656					
4	PKS		4029025-4040172					
	Score	Distance From ORF	N	Strand	Sequence	Gene	Synonym	Protein Name
	76	178	4029404	+	GCTGGTCCACATCATGCTCGAG	-	MXAN_3460	Gfo/Idh/MocA family oxidoreductase
	76	182	4029400	+	CTGGTCCACATCATGCT	-	MXAN_3460	Gfo/Idh/MocA family oxidoreductase
5	NRPS/ PKS		4220394-4350875					
	Score	Distance From ORF	N	Strand	Sequence	Gene	Synonym	Protein Name
	75	319	4254532	-	ATGGCGGAAGAGCTGCT	-	MXAN_3627	hypothetical protein
	75	398	4255390	-	ACTGCCACGCCAGCAGCTCGGG	-	MXAN_3628	non-ribosomal peptide biosynthesis thioesterase
	75	399	4255391	-	AACTGCCACGCCAGCAGCTCG	-	MXAN_3628	non-ribosomal peptide biosynthesis thioesterase
	75	308	4259689	-	CTGGCGGACATCTGGCA	-	MXAN_3630	polyketide synthase type I
	75	287	4264309	-	GAAGGAACAGCTGTTGCGCATG	-	MXAN_3631	polyketide synthase type I
	78	286	4264308	-	AAGGAACAGCTGTTGCG	-	MXAN_3631	polyketide synthase type I
	76	211	4284118	+	AAGGCGTGCGCATTGCT	-	MXAN_3635	non-ribosomal peptide synthase/polyketide synthase
	77	32	4334952	+	CTGGTACTTCGAGTGCA	-	MXAN_3638	M19 family peptidase
	76	311	4347401	-	GTGGAAAACCATCTGCT	-	MXAN_3644	isochorismatase
	75	268	4351143	-	GTGGCACGCGAAATCCC	-	MXAN_3647	2,3-dihydroxybenzoate-2,3-dehydrogenase
6	PKS		4498820-4544105					
	Score	Distance From ORF	N	Strand	Sequence	Gene	Synonym	Protein Name
	76	284	4498536	+	TTGAGACGGAGATTGCG	-	MXAN_3778	DnaK family protein
7	PKS		4729983-4809194					

	Score	Distance From ORF	N	Strand	Sequence	Gene	Synonym	Protein Name
	75	337	4731192	-	CGTGGGCCGGTTCCGGCTCAAG	-	MXAN_3931	hypothetical protein
	76	73	4740097	-	TCAGACACGCCCATCGCTGATC	-	MXAN_3932	polyketide synthase
	75	90	4740114	-	CTTGAAAAACCGGTTGCTCAGA	-	MXAN_3932	polyketide synthase
	75	103	4755572	-	AGGTGGTACTTCGCTTACTCG	-	MXAN_3933	mixed type I polyketide synthase - peptide synthetase
	76	101	4755570	-	GTGGTACTTCGCTTACT	-	MXAN_3933	mixed type I polyketide synthase - peptide synthetase
	75	428	4784222	-	ATTGGCTCGCGTCTGGCTGCGC	-	MXAN_3935	non-ribosomal peptide synthase/polyketide synthase Ta1
	77	221	4784015	-	TGCTGGCACGAACCCTGGGCG	-	MXAN_3935	non-ribosomal peptide synthase/polyketide synthase Ta1
	82	427	4784221	-	TTGGCTCGCGTCTGGCT	-	MXAN_3935	non-ribosomal peptide synthase/polyketide synthase Ta1
	78	219	4784013	-	CTGGCACGAACCCTGGG	-	MXAN_3935	non-ribosomal peptide synthase/polyketide synthase Ta1
	75	315	4784109	-	TTGGCGCTGAGCGTGCC	-	MXAN_3935	non-ribosomal peptide synthase/polyketide synthase Ta1
	78	333	4790662	-	GTTGGTCCGCCGGGCGGTGGAG	-	MXAN_3936	polyketide synthase
	75	155	4797857	-	GCTGTCGCTCGTGACGCTCAAG	-	MXAN_3938	polyketide synthase
	76	245	4800766	-	GGCTGGCTCTGCAACTGCGCAG	-	MXAN_3941	polyketide beta-ketoacyl:acyl carrier protein synthase
	76	245	4800766	-	GGCTGGCTCTGCAACTGCGCA	-	MXAN_3941	polyketide beta-ketoacyl:acyl carrier protein synthase
	75	243	4800764	-	CTGGCTCTGCAACTGCG	-	MXAN_3941	polyketide beta-ketoacyl:acyl carrier protein synthase
	77	318	4804366	-	GTGGCGCATGCCTTTCT	-	MXAN_3943	cytochrome P450 family protein
	77	97	4805951	-	CCTGGACAGCCTGCGGCTGCAG	taF	MXAN_3945	polyketide TA biosynthesis protein TaF
	76	464	4806318	-	CCTGTCGCGCGTGGTGCTGGAC	taF	MXAN_3945	polyketide TA biosynthesis protein TaF
	76	220	4806318	-	CCTGTCGCGCGTGGTGCTGGAC	-	MXAN_3946	putative acyl carier protein
8	NRPS/ PKS		4875470-4906976					
	Score	Distance From ORF	N	Strand	Sequence	Gene	Synonym	Protein Name
	79	135	4900530	-	CTCGCCCGCAGGTTGCA	-	MXAN_4001	non-ribosomal peptide synthase/polyketide synthase
	75	470	4906209	-	AGGGCACGGCCCGTGAT	-	MXAN_4002	nonribosomal peptide synthetase
9	NRPS/ PKS		4996865-5026357					
10	NRPS/ PKS		5252320-5298262					

	Score	Distance From ORF	N	Strand	Sequence	Gene	Synonym	Protein Name
	76	350	5253698	+	GCTGGAGCGCGTGCTGCTGGCC	-	MXAN_4290	putative thioesterase
	77	352	5253696	+	TGCTGGAGCGCGTGCTGCTGG	-	MXAN_4290	putative thioesterase
	76	354	5253694	+	CTGGAGCGCGTGCTGCT	-	MXAN_4290	putative thioesterase
	76	132	5263568	-	ATGTCGCCGTGCTTGCA	-	MXAN_4295	patatin-like phospholipase family protein
	77	150	5267893	-	AGCTGGGAGTGGACCTGCTCA	-	MXAN_4296	non-ribosomal peptide synthetase
	77	379	5279177	-	GATGCGTCGGCAGTTGCTGATG	-	MXAN_4298	polyketide synthase type I
	80	178	5287908	-	ATGGCGCTCGAGTTGCG	-	MXAN_4299	non-ribosomal peptide synthase/polyketide synthase
	77	183	5292178	-	CGCTGGCGAAGCGGCTGCTGT	-	MXAN_4300	polyketide synthase type I
	76	181	5292176	-	CTGGCGAAGCGGCTGCT	-	MXAN_4300	polyketide synthase type I
11	PKS		5407416-5440677					
12	NRPS/ PKS		5444807-5471033					
	Score	Distance From ORF	N	Strand	Sequence	Gene	Synonym	Protein Name
	75	182	5448069	+	CCTGGACGACCGGTTGGTGCTG	-	MXAN_4413	hypothetical protein
	78	328	5468596	+	CCTGGGCCACGGATTGCTGCGC	-	MXAN_4416	cephalosporin hydroxylase family protein
	79	332	5468592	+	CTGGGCCACGGATTGCT	-	MXAN_4416	cephalosporin hydroxylase family protein
13	NRPS/ PKS		5600883-5653548					
	Score	Distance From ORF	N	Strand	Sequence	Gene	Synonym	Protein Name
	75	259	5612126	-	ATGGCGCTGGAGCTGCG	-	MXAN_4526	polyketide synthase type I
	78	373	5612240	-	ATGGCGCTGCTGTTGGA	-	MXAN_4526	polyketide synthase type I
	81	466	5612333	-	GTGGCAGGGCAGGTGCG	-	MXAN_4526	polyketide synthase type I
	75	170	5627582	-	CTGGCTGAACAACTGCA	-	MXAN_4527	polyketide synthase
	79	187	5653735	-	GTGGCACAAGCTGCGCT	-	MXAN_4532	non-ribosomal peptide synthase
14	NRPS		5735538-5780049					
	Score	Distance From ORF	N	Strand	Sequence	Gene	Synonym	Protein Name
	75	39	5755794	+	CAATGGAACGGCGCATCCTCC	-	MXAN_4598	non-ribosomal peptide synthase
	75	41	5755792	+	ATGGAACGGCGCATCCT	-	MXAN_4598	non-ribosomal peptide synthase
	75	485	5759989	+	CCTGGGCTACGGGGTGCTGCGC	-	MXAN_4599	M28 family peptidase
	75	493	5765386	-	GGTGGCCTGGGCGTCGGTGGCG	-	MXAN_4600	radical SAM domain-containing protein
	75	40	5764933	-	GCGGTAAAGTCTTTGCT	-	MXAN_4600	radical SAM domain-containing protein
	78	237	5774788	+	CCTGGCCCGTCTGGAGCTCCAG	-	MXAN_4602	hypothetical protein

75	241	5774784	+	CTGGCCCGTCTGGAGCT	-	MXAN_4602	hypothetical protein
76	62	5777754	+	ATGTCGCTCGACTTGCT	-	MXAN_4604	hypothetical protein

NC_010162.1 Sorangium cellulosum 'So ce 56', complete genome

1	PKS		488146-499710				
2	PKS		1150403-1167578				
Score	Distance From ORF	N	Strand	Sequence	Gene	Synonym	Protein Name
76	300	1150103	+	CTGGCCTGCGCGTCGCT	-	sce0818	putative dioxygenase
75	299	1151780	+	ATGGCGCTGATCTTCCG	-	sce0819	polyketide synthase
3	NRPS		3204860-3216562				
4	PKS		4376667-4461851				
Score	Distance From ORF	N	Strand	Sequence	Gene	Synonym	Protein Name
84	488	4376179	+	CTGGCAGGATATCTGCT	-	sce3188	polyketide synthase
77	52	4440389	+	CTGGGGCGGATCTCGCT	-	sce3193	polyketide synthase
5	PKS		5761328-5853706				
6	PKS		6867494-6856887				
7	PKS		402480-9513516				
Score	Distance From ORF	N	Strand	Sequence	Gene	Synonym	Protein Name
81	172	9407529	-	TCGGCACGCTGCCTGCA	-	sce6747	hypothetical protein
90	68	9411437	-	GTGGCGCGCCGCTTGCT	-	sce6751	hypothetical protein
75	485	9425010	-	GTGGCGCCGCGGCTGCC	-	sce6759	hypothetical protein
75	4	9436354	+	ATGGCAGGAGGGGTGAT	-	sce6766	putative secreted protein
81	80	9440680	-	CGGGCACGGGCGCTGCT	-	sce6770	hypothetical protein
80	438	9441038	-	AGGGCATGATCGTTGCC	-	sce6770	hypothetical protein
77	38	9452036	-	GCGGTACACCCCTTGCG	-	sce6780	hypothetical protein
75	93	9452091	-	GAGGCGCGGCGGATGCT	-	sce6780	hypothetical protein
81	344	9452342	-	CTGGCTCGAGGCCTGCA	-	sce6780	hypothetical protein
75	212	9452057	+	GCGGCCCGCCCGCTGCT	-	sce6781	TetR family transcriptional regulator
77	479	9456589	+	ACGGCACGCGACGTGCG	-	sce6785	hypothetical protein
77	260	9460836	+	CTGGCACCGGGCGAGCT	-	sce6787	hypothetical protein
78	361	9466227	-	CTGGCGCTCGATCTGCT	-	sce6792	hypothetical protein
77	107	9470202	-	TTCGCATGGCTGTTGGA	-	sce6797	hypothetical protein
75	293	9479287	-	ATGGTTCGTTTCTCGCT	-	sce6805	hypothetical protein
77	345	9497084	-	GTGGATCGAGTCATGCT	-	sce6819	hypothetical protein
8	NRPS/ PKS		11424644-11529102				
Score	Distance From ORF	N	Strand	Sequence	Gene	Synonym	Protein Name
81	427	1.1E+07	-	GGGGCACGATCGCTGCT	-	sce8207	sigma-54 dependent transcriptional regulator
84	42	1.1E+07	-	CTGGCATCCCTCGTGCT	-	sce8210	putative DNA helicase
75	419	1.1E+07	+	AGGGCGCGGCGTCTGCA	-	sce8233	hypothetical protein
82	43	1.1E+07	+	AGGGCACGGCTCTTGGA	-	sce8234	hypothetical protein
76	173	1.1E+07	+	GTGGCCGACGTCCTGCT	yfbL2	sce8235	putative aminopeptidases

	75	388	1.1E+07	+	GTGGCGCGCTCCCCGCT	-	sce8244	putative protein phosphatase
	75	371	1.1E+07	-	GAGGCCCAGCTCGTGCA	-	sce8246	CobW/P47K family protein

NC_002163.1 Campylobacter jejuni subsp. jejuni NCTC 11168, complete genome

1	NRPS	1236809-1238584

NC_000913.2 Escherichia coli str. K-12 substr. MG1655, complete genome

1	NRPS		608682-631222					
	Score	Distance From ORF	N	Strand	Sequence	Gene	Synonym	Protein Name
	78	373	612792	+	ATCGCACCGTGGTTGCC	ybdZ	b4511	conserved protein
	80	361	619783	-	CTGGTACGCCGGATGCG	fepC	b0588	iron-enterobactin transporter subunit
	79	153	620564	-	CCGTCACGCTACTTGCT	fepG	b0589	iron-enterobactin transporter subunit
	79	138	621385	+	CGGGCACGGCAATGGCG	entS	b0591	enterobactin exporter, iron-regulated
	76	428	624161	-	TGGGCGCAAGTGTTGCC	fepB	b0592	iron-enterobactin transporter subunit
	79	291	625002	+	CTGGCCTGTCTGCTGCA	entE	b0594	2,3-dihydroxybenzoate-AMP ligase component of enterobactin synthase multienzyme complex

NC_002942.5 Legionella pneumophila subsp. pneumophila str. Philadelphia 1, complete genome

1	NRPS/ PKS	2166115-2171151
2	NRPS/ PKS	2442757-2470485

NC_002516.2 Pseudomonas aeruginosa PAO1, complete genome

1	NRPS		2532669-2549458					
	Score	Distance From ORF	N	Strand	Sequence	Gene	Synonym	Protein Name
	90	276	2539319	-	CTGGCCCGGGCCTTGCT	-	PA2302	non-ribosomal peptide synthetase
	79	154	2545100	-	ATGGCGCTGCTGTTGGT	-	PA2305	non-ribosomal peptide synthetase
	83	8	2545667	-	ATGGAACGAATGTCGCT	-	PA2306	hypothetical protein
	84	55	2547555	-	CTGGCAGGACTATTTCT	-	PA2308	ABC transporter ATP-binding protein
2	NRPS		2636517-2687178					
	Score	Distance From ORF	N	Strand	Sequence	Gene	Synonym	Protein Name
	75	125	2640910	+	ACGGCACTGTCCTTCCA	fpvR	PA2388	FpvR
	80	88	2651219	+	CTGGCGCCGAGCCTGCT	pvdO	PA2395	PvdO
	75	140	2653197	-	CTTCCACGTAAATTGCA	pvdF	PA2396	pyoverdine synthetase F
	77	421	2665565	-	CTGGCGCGGATGCCGCT	pvdD	PA2399	pyoverdine synthetase D
	77	479	2672108	-	CTGGCGCGGATGCCGCT	pvdJ	PA2400	PvdJ

3	NRPS			4724639-4746551				
	Score	Distance From ORF	N	Strand	Sequence	Gene	Synonym	Protein Name
	81	393	4727194	-	CTGGCTCGCGGGCTGCT	fptA	PA4221	Fe(III)-pyochelin outer membrane receptor precursor
	78	265	4728882	-	CCGGCGAGCCTGTTGCT	-	PA4222	ABC transporter ATP-binding protein
	80	286	4728903	-	CTGGCCCGTGCGCTGCT	-	PA4222	ABC transporter ATP-binding protein
	75	43	4731415	-	CTGGCGCAATCCTTGTC	pchG	PA4224	pyochelin biosynthetic protein PchG
	75	16	4736814	-	CTGGAAGAGGGCGTGCT	pchF	PA4225	pyochelin synthetase
	81	64	4736862	-	CTTGCCCGCCACTTGCA	pchF	PA4225	pyochelin synthetase
	77	465	4741576	-	CTGGTCGGCGCCTTGCA	pchE	PA4226	dihydroaeruginoic acid synthetase

NC_002947.3 Pseudomonas putida KT2440, complete genome

1	NRPS			4767855-4798662				
	Score	Distance From ORF	N	Strand	Sequence	Gene	Synonym	Protein Name
	79	278	4795256	-	TTGGCGCCGAAGGTGCA	-	PP_4222	syrP protein, putative
	77	431	4795409	-	ATGGGGCAGCGGTTGCG	-	PP_4222	syrP protein, putative
	80	413	4796855	-	ATGGGCCAGGCATTGCT	-	PP_4223	diaminobutyrate--2-oxoglutarate aminotransferase
2	NRPS			4818235-4832793				
	Score	Distance From ORF	N	Strand	Sequence	Gene	Synonym	Protein Name
	76	122	4832915	-	TTGGCCAGTGGCCTGCT	-	PP_4245	siderophore biosynthesis protein, putative

NC_004578.1 Pseudomonas syringae pv. tomato str. DC3000, complete genome

1	NRPS			2307009-2356936				
	Score	Distance From ORF	N	Strand	Sequence	Gene	Synonym	Protein Name
	81	315	2306694	+	ATGGAATTCACCTTGCT	-	PSPTO_2134	pyoverdine synthetase, thioesterase component
	76	381	2322147	+	CTGGTCGGTCATTTGCA	-	PSPTO_2137	MbtH-like protein
	78	274	2324978	-	CCGGCGCCGAAATTGCT	-	PSPTO_2139	cation ABC transporter, permease protein
	80	491	2327813	-	CTGGCACAAGTGGCGCT	-	PSPTO_2143	hypothetical protein
	75	384	2330265	+	CTGGAAAACGACGTGCT	-	PSPTO_2147	pyoverdine sidechain peptide synthetase I, epsilon-Lys module
	77	389	2333571	+	CCGACACACTTGTTGCT	-	PSPTO_2148	pyoverdine sidechain peptide synthetase II, D-Asp-L-Thr component
2	NRPS/ PKS			2863432-2890634				
	Score	Distance From	N	Strand	Sequence	Gene	Synonym	Protein Name

		ORF						
	76	322	2868091	-	TTGGCGCGGGTGCCGCA	-	PSPTO_2593	multidrug resistance protein, AcrA/AcrE family
	79	78	2872757	-	GTGGCCAGGCTCTGGCA	irp4	PSPTO_2598	yersiniabactin synthetase, thioesterase component
	81	235	2873968	-	CTGGCGCAACACTCGCA	irp3	PSPTO_2599	yersiniabactin synthetase, thiazolinyl reductase component
	80	112	2884576	-	CTGGCGCGGCACTTGAG	-	PSPTO_2601	hypothetical protein
	78	384	2884848	-	CTGGCGCTGACGTGGCT	-	PSPTO_2601	hypothetical protein
	75	89	2890723	-	CAGGCTCGGAACTGGCA	-	PSPTO_2602	yersiniabactin non-ribosomal peptide synthetase
3	NRPS		3151815-3185293					
	Score	Distance From ORF	N	Strand	Sequence	Gene	Synonym	Protein Name
	76	472	3152407	+	TTGATACGGCTGATGCT	syfA	PSPTO_2829	non-ribosomal peptide synthetase SyfA
4	NRPS		5087432-5106501					
	Score	Distance From ORF	N	Strand	Sequence	Gene	Synonym	Protein Name
	76	440	5088005	+	CTGGCAGCCTGGTTCCT	-	PSPTO_4518	non-ribosomal peptide synthetase, initiating component
5	PKS		5282597-5304804					
	Score	Distance From ORF	N	Strand	Sequence	Gene	Synonym	Protein Name
	75	23	5284944	+	CTGGTCCGCAAGTCGCT	cfa3	PSPTO_4683	coronafacic acid beta-ketoacyl synthetase component
	75	194	5286445	+	CTGCCAGGCCAACTGCT	cfa5	PSPTO_4685	coronafacic acid synthetase, ligase component
6	NRPS		5313602-5327197					

NC_003198.1 Salmonella enterica subsp. enterica serovar Typhi str. CT18, complete genome

1	NRPS		624926-644794					
	Score	Distance From ORF	N	Strand	Sequence	Gene	Synonym	Protein Name
	82	483	640548	-	CTGGCGCAATTTCTGCT	fepB	STY0638	iron-enterobactin transporter periplasmic binding protein
	79	288	641267	+	CTGGCCTGTCTGCTGCA	entE	STY0640	enterobactin synthase subunit E

NC_004347.1 Shewanella oneidensis MR-1, complete genome

1	PKS		1670313-1686159					
	Score	Distance From ORF	N	Strand	Sequence	Gene	Synonym	Protein Name
	78	441	1676645	-	ATAGCATGAGTATTGCC	-	SO_1599	beta keto-acyl synthase

NC_003143.1 Yersinia pestis CO92, complete genome

1	NRPS		839849-856936					
	Score	Distance From ORF	N	Strand	Sequence	Gene	Synonym	Protein Name
	79	186	841919	-	CGGGCACTATTGCTGCT	-	YPO0774	hypothetical protein
	81	353	857289	-	ATGGCCCCCCGCCTGCT	-	YPO0778	putative siderophore biosysnthesis protein
2	NRPS/ PKS		2140840-2169669					
	Score	Distance From ORF	N	Strand	Sequence	Gene	Synonym	Protein Name
	82	239	2145615	-	CTGGCGCGACTGCTGCG	irp4	YPO1908	yersiniabactin biosynthetic protein YbtT
	75	486	2166571	+	ATGACACGCTGCTGGCG	irp8	YPO1915	putative signal transducer
	83	441	2167924	+	CTGGCCGGCGGATTGCT	ybtS	YPO1916	salicylate synthase Irp9
	79	96	2168269	+	GGGGCATTCACGTTGCT	ybtS	YPO1916	salicylate synthase Irp9

Chapter 4: Interplay between *rpoN* and *rpoS* mediated by overexpression positively regulates transcription of the oxytetracycline biosynthetic gene cluster in *E. coli*

4.1 Introduction

The ability to adapt and survive under various stresses such as nutrient availability and shifts in environmental exposure observed in bacteria can be attributed to the complex alternative sigma factor regulatory cascade.[1,2]Stressed bacteria undergo a dramatic shift in gene expression mediated by alternative sigma factors that alter the RNA polymerase core specificity.[1-5] Two specific alternative sigma factors σ^S and σ^{54},transcribed by the genes *rpoS* and *rpoN*,have been recently found to possess an antagonistic overlap in 60% of their respective regulons.[6] This antagonistic relationship furthers the already complex regulatory competition between alternative sigma factors observed during stress response.

Previously,σ^{54} overexpression has been observed to afford heterologous expression of the oxytetracycline biosynthetic gene cluster from *Streptomyces rimosus* and production of the secondary metabolite oxytetracycline in the heterologous host *Escherichia coli*.[7]In this previous work,*E. coli* strain BAP1 was co-transformed with a plasmid containing the oxytetracycline biosynthetic gene cluster along with a second vector containing an alternative sigma factor under control of an inducible promoter.This production of oxytetracycline in the σ^{54} overexpression strain was attributed to the direct transcription of the *oxyABCD* minimal polyketide synthase (PKS) from a discovered σ^{54} promoter upstream of the operon.[7,8] Although no oxytetracycline production was observed with σ^S overexpression, transcription from the minimal PKS was observed.

Transcription from the minimal PKS in the absence of oxytetracycline production was also observed when the alternative sigma factor σ^{fecI}, transcribed by *fecI*, was overexpressed.[7]

The hypothesis that overexpression of both σ^{S} and σ^{fecI} leads to increased transcription from σ^{54} and subsequent direct transcription from the oxytetracycline minimal PKS would offer an explanation for the presence of minimal PKS transcripts in both σ^{S} and σ^{fecI} overexpression strains. The hypothesis that overexpression of both σ^{S} and σ^{fecI}leads to either direct or indirect transcription of the oxytetracyline minimal PKS decoupled from σ^{54} regulation also provides a viable explanation. In this work we explore the regulatory cascade between alternative sigma factors as it relates to the heterologous production of oxytetracycline in *E. coli*. Herein we present the existence of a previously undiscovered positive induction loop between σ^{54} and σ^{S} during overexpression along with the observation that σ^{S} overexpression affords transcription of the oxytetracycline minimal PKS in the absence of σ^{54}.

To explore the hypothesis that interplay between the respective sigma factors σ^{S} and σ^{fecI} increases transcription of σ^{54} and its regulon, a series of qPCR experiments were setup to monitor the transcriptional effects of various overexpression conditions. As in previous experiments, total RNA was isolated during stationary phase and cDNA was generated from *E. coli* BAP1 harboring an inducible vector containing either *rpoN, rpoS,* or *fecI* along with a vector containing the entire oxytetracycline biosynthetic gene cluster.[7]cDNA from a WT control and a negative control was also generated from *E. coli* BAP1 harboringempty expression vectors, pET21c and pET28b, for the WT control and

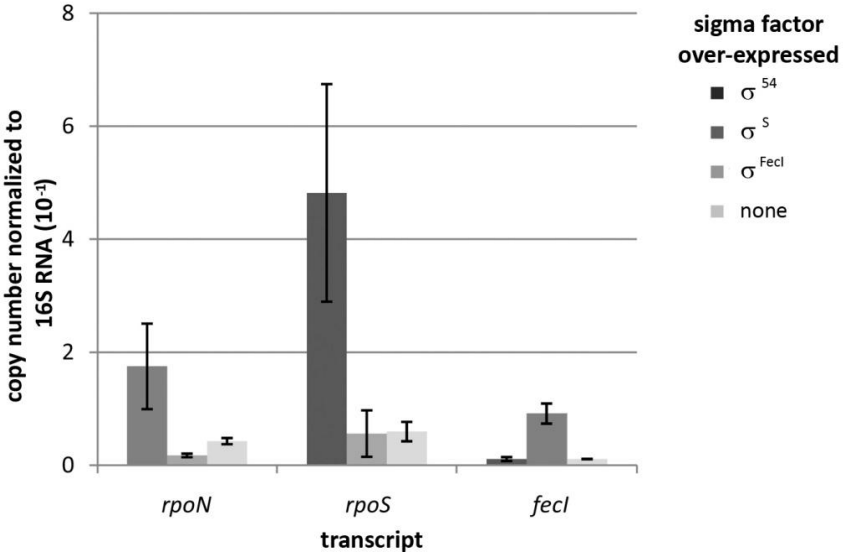
E. coli BAP1 harboring an empty expression vector, pET21c, and the entire oxytetracycline biosynthetic gene cluster for the negative control. Amplicons from each sigma factor σ^{54}, σ^{S}, and σ^{fecI} along with an amplicon from the regulon of each respective sigma factor *glnA*, *entD*, and *fecA* were quantitated.[3,4,9] Due to the global effect of ppGpp on alternative sigma factor transcription and the observation that overexpression of σ^{54}, σ^{S}, or σ^{fecI} could produce a positive feedback loop through a synergistic upregulation of ppGpp synthesis and its transcriptional effects, amplicons for *spoT,* involved in ppGpp synthesis during starvation, and *dksA,* involved in the ppGpp mediated shift in core polymerase specificity towards alternative sigma factors, were also quantitated.[10]

4.2 Results

When compared to the WT control, overexpression of σ^{S} led to expected increases in *rpoS* transcripts and transcripts of *entD* a known member of its regulon involved in enterobactin synthesis.[4] Increased transcription from *rpoN* was also observed in the σ^{S} overexpression strain and provides an observation of a potential amplification loop between σ^{S} and σ^{54} during overexpression of σ^{S}. Increased transcription from *fecI* as has been previously reported was also observed with σ^{S}overexpression.[4]No significant increase in transcription from *fecA*,*glnA*, or*spoT*were observed.[4] Overexpression of σ^{54} led to an increase in the transcription of *rpoN*, *glnA*, *rpoS*,*entD,* and *dksA* along with a slight increase in the transcription of *fecA* (Figure 4.2.1). Increased transcription of *rpoS* and *entD* from the σ^{54} overexpression strain provides further evidence towards the existence of an amplification loop between σ^{54} and σ^{S} during overexpresion of either sigma factor (Figure 4.2.1). Although *dksA* has been linked to transcription from bothσ^{54}

and σ^S, an increase in transcription of *dksA* from σ^{54} and σ^S overexpression has not been reported.[10]Overexpression of σ^{fecI} led to the expected increase in transcripts of *fecI* and *fecA*.[9] No other significant increase in any of the other transcripts were observed in the σ^{fecI} overexpression strains.[4] Therefore, interplay between σ^{54} and σ^{fecI}cannot justify the transcription of the oxytetracycline minimal PKS during σ^{fecI} overexpression.An amplification loop between σ^{54} and σ^S provides evidence for σ^{54} mediated transcription of the oxytetracycline minimal PKSduring σ^S overexpression. However these data do not discount the hypothesis that σ^S and σ^{fecI} overexpression provide direct or indirect production of oxytetracycline independent of σ^{54} in *E. coli*.[f]

A



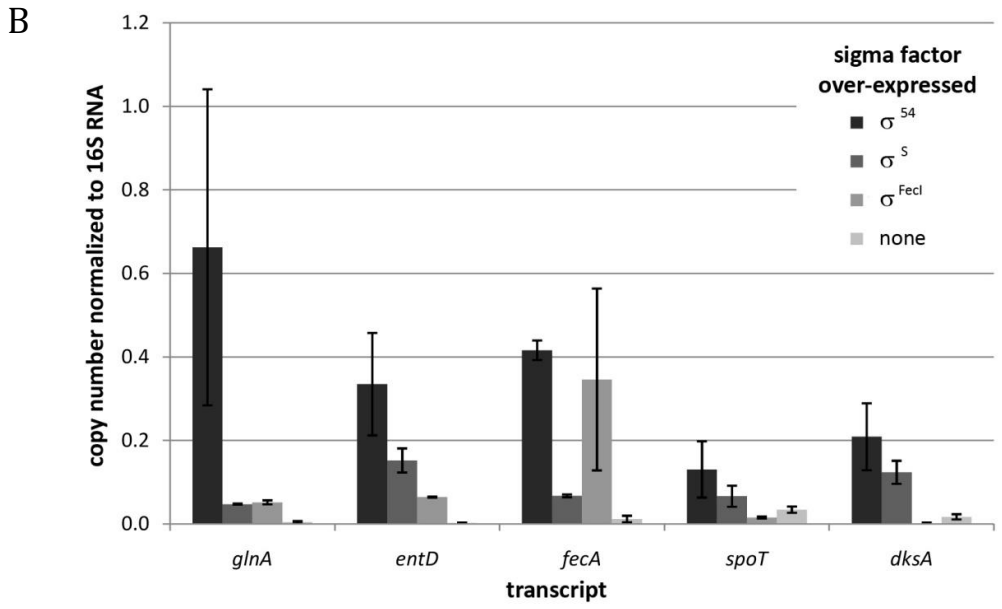


Figure 4.2.1: qPCR data for transcription of A) *rpoN, rpoS,* and *fecI* and B) *fecA, entD, spoT, glnA,* and *dksA* from sigma factor overexpression in *E. coli* BAP1.

To decouple σ^{54} regulation from σ^{S} and σ^{fecI} overexpression, a *rpoN*- strain was acquired from the Yale Stock Center's Keio collection.[11] The MG1655 derivative was subsequently transduced with the T7 RNA polymerase gene to allow inducible protein expression. cDNA was generated from the *rpoN*- strain as previously described *E. coli* BAP1 experiments monitoring the overexpresion of σ^{54}, σ^{S}, and σ^{fecI}. Transcripts previously used to observe transcription of the oxytetracycline biosynthetic gene cluster, *oxyB* from the KS in the minimal PKS and *oxyK* a post-tailoring cyclase, were also monitored to observe any changes in transcription associated with loss of σ^{54}.[7] Upon restoring and overexpressing σ^{54} in the *rpoN*- strain, results were comparable to the previous data aquired from *E. coli* BAP1 experiments including increased transcription

from *rpoS* and *entD*. Overexpression of σ^{fecI} in the *rpoN*- strain produced no detectable levels of transcript from *oxyB*. Although transcripts from *rpoN*duringσ^{S} overexpression could not be monitored in the *rpoN*- strain, the previously observed increase in *glnA* transcription was abolished. Interestingly, transcripts were detected for *oxyB* indicating that σ^{S} is capable of either direct or indirect transcription of the ketosynthase in *E. coli,*independent of σ^{54}.

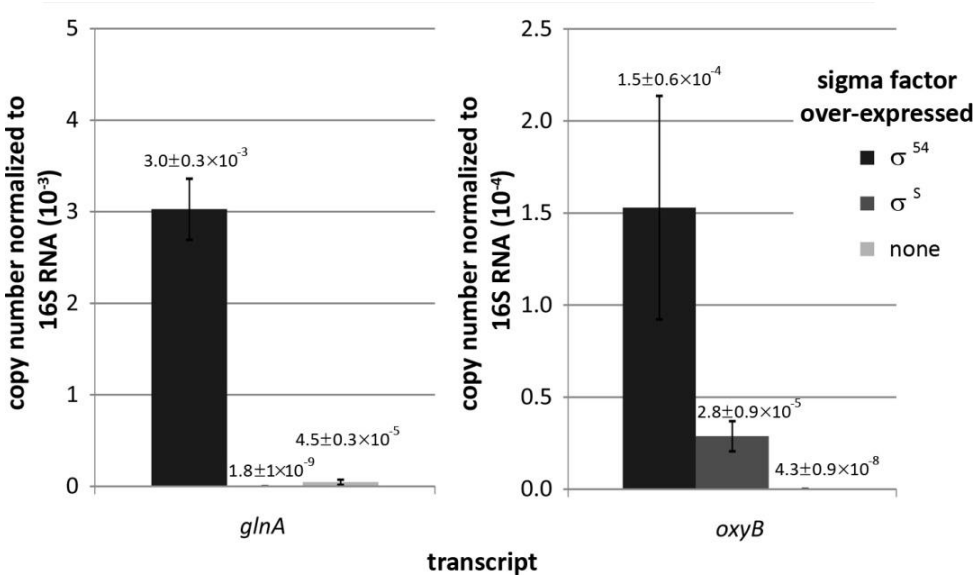


Figure 4.2.2: qPCR data for transcription of A) *oxyB* and *oxyK* from the oxytetracycline biosynthetic pathway and B) *glnA* and *entD*from sigma factor overexpression in *rpoN*- strain *E. coli* KC01.

4.3 Discussion

The ability of σ^{S} to heterologously transcribe the oxytetracycline minimal PKS during overexpression in the absence of σ^{54} presents σ^{S} as another link in alternative

sigma factors and secondary metabolism. This result also furthers the utility of alternative sigma factors as tools for the heterologous expression of secondary metabolites in *E. coli*. Although no oxytetracycline production was observed from the σ^S overexpression strain, presumably due to a 10-fold lower level of *oxyB* transcripts, future experiments involving biosynthetic pathways for other secondary metabolites could benefit from σ^S overexpression. This data presents the second example of a sigma factor not native to the oxytetracycline producing strain, *S. rimosus*,possessing a regulatory response when overexpressed in *E. coli*.[7,12] Although previous experiments depicted σ^{54} as directly transcribing *oxyB* in *E. coli*, the consensus promoter for σ^S remains difficult to determine and would require further experiments to determine whether this results occurs due to direct or indirect transcription.[4] However, no σ^{70}-like promoter region possessing the currently accepted σ^S consensus motif could be located upstream of *oxyB* from the oxytetracycline biosynthetic pathway suggesting an indirect means of transcription.[4]

4.4 Conclusion

This work presents the first reported observation of a transcriptional amplication loop existing between σ^{54} and σ^S. Although an antagonistic relationship between the regulons of the two sigma factors has been previously reported, the antagonistic interplay was discovered through null mutations of the sigma factors and thorough comparison between each of the mutant strains' regulons.[6]The existence of a antagonistic relationship correlates well with the existence of an amplification loop between σ^{54} and σ^S. Each sigma factor's ability to upregulate the basal expression of the other during

overexpression provides a mechanism for maintaining homeostasis of this complex regulatory system. Previously observed regulatory elements such as Crl, a regulatory protein that shifts core RNA specificity from σ^{70} towards σ^{S}, and Rsd, an anti-sigma factor for σ^{70}, have both been observed as members of the σ^{54} and σ^{S} regulons respectively.[6,13,14] This work also suggests that transcription of the stringent response regulatory element *dksA* is also linked to the σ^{54} and σ^{S} regulons. These regulatory elements offer a potential explanation for the indirect amplification loop observed in our data. However the regulatory cascade between σ^{54}, σ^{S}, and other sigma factors during stress response and other conditions, such as overexpression of individual sigma factors, requires further investigation towards understanding this complex network of regulatory elements.

4.5 References

1. Grigorova, I. L.; Phleger, N. J.; Mutalik, V. K.; Gross, C. A., *Proc. Natl. Acad. Sci. USA* **2006,***103* (14), 5332-7.

2. Martínez-Núñez, M. A.; Pérez-Rueda, E.; Gutiérrez-Ríos, R. M.; Merino, E., *Microbiology* **2010,***156* (Pt 1), 14-22.

3. Reitzer, L.; Schneider, B. L., *Microbiol. Mol. Biol. Rev.***2001,***65* (3), 422-444.

4. Hengge-Aronis, R., *Microbiol. Mol. Biol. Rev.***2002,***66* (3), 373-95.

5. Hengge, R., *Adv. Exp. Med. Biol.***2008,***631*, 40-53.

6. Dong, T.; Yu, R.; Schellhorn, H., *Mol. Microbiol.***2011,***79* (2), 375-86.

7. Stevens, D. C.; Conway, K.; Pearce, N.; Boddy, C. N., **2011** (Submitted).

8. Pickens, L. B.; Tang, Y., *J. Biol. Chem.***2010,***285* (36), 27509-15.

9. Brooks, B. E.; Buchanan, S. K., *Biochim. Biophys. Acta.***2008,***1778* (9), 1930-45.

10. Aberg, A.; Fernández-Vázquez, J.; Cabrer-Panes, J. D.; Sánchez, A.; Balsalobre, C., *J. Bacteriol.***2009,***191* (10), 3226-36.

11. Baba, T.; Ara, T.; Hasegawa, M.; Takai, Y.; Okumura, Y.; Baba, M.; Datsenko, K. A.; Tomita, M.; Wanner, B. L.; Mori, H., *Mol. Syst. Biol.***2006,***2*, 2006.0008.

12. Chiang, S. M.; Schellhorn, H. E., *J. Mol. Evol.***2010,***70* (6), 557-71.

13. Zhao, K.; Liu, M.; Burgess, R. R., *Nucleic Acids Res.***2010,***38* (4), 1273-83.

14. Typas, A.; Barembruch, C.; Possling, A.; Hengge, R., *EMBO J.***2007,***26* (6), 1569-78.

4.6 Experimental

General. *E. coli* BAP1 was used as the heterologous host for all transcription experiments not performed in the constructed *rpoN*- strain. *E. coli* cells were grown in LB media supplemented with the appropriate antibiotics. Oxytetracyline standard and IPTG were purchased from Sigma-Aldrich. Antibiotics and media components were purchased from Fisher. All strains were chemically or electronically transformed using standard transformation protocols.

Construction of *rpoN*- strain. *E. coli rpoN*- mutant JW3169 was requested from the Keio Collection deposited at the Yale Stock Center. The inserted kan^R markers was excised using the surrounding Flp sites and the temperature sensitive vector pCP20 using previously reported conditions. The produced strain was then transfected with the T7 polymerase using using the λDE3 Lysogenization Kit (Novagen).

qPCR Analysis. ABsolute SYBR Green QPCR mix (Thermo Scientific) and the Mx3000 qPCR (Agilent) were used for all qPCR experiments. Thermocycler conditions were as follows: 1 cycle 95 °C for 15 min, 40 cycles of 95 °C for 15 s, 55 °C for 30 s, and 72 °C for 30 s, followed by denaturation for 1 cycle at 95 °C for 1 min, 55 °C for 30 s, and 95 °C for 30 s. The SV Total RNA Isolation System (Promega) was used to isolate RNA for the construction of cDNA. cDNA from each strain was generated from isolated RNA using AMV Reverse Transcriptase (Promega) using the manufacturer's instructions. Strains, prepared by electroporating *E. coli* BAP1, were BAP1 + pET21c+ pET28b (WT), BAP1 + pET21c +

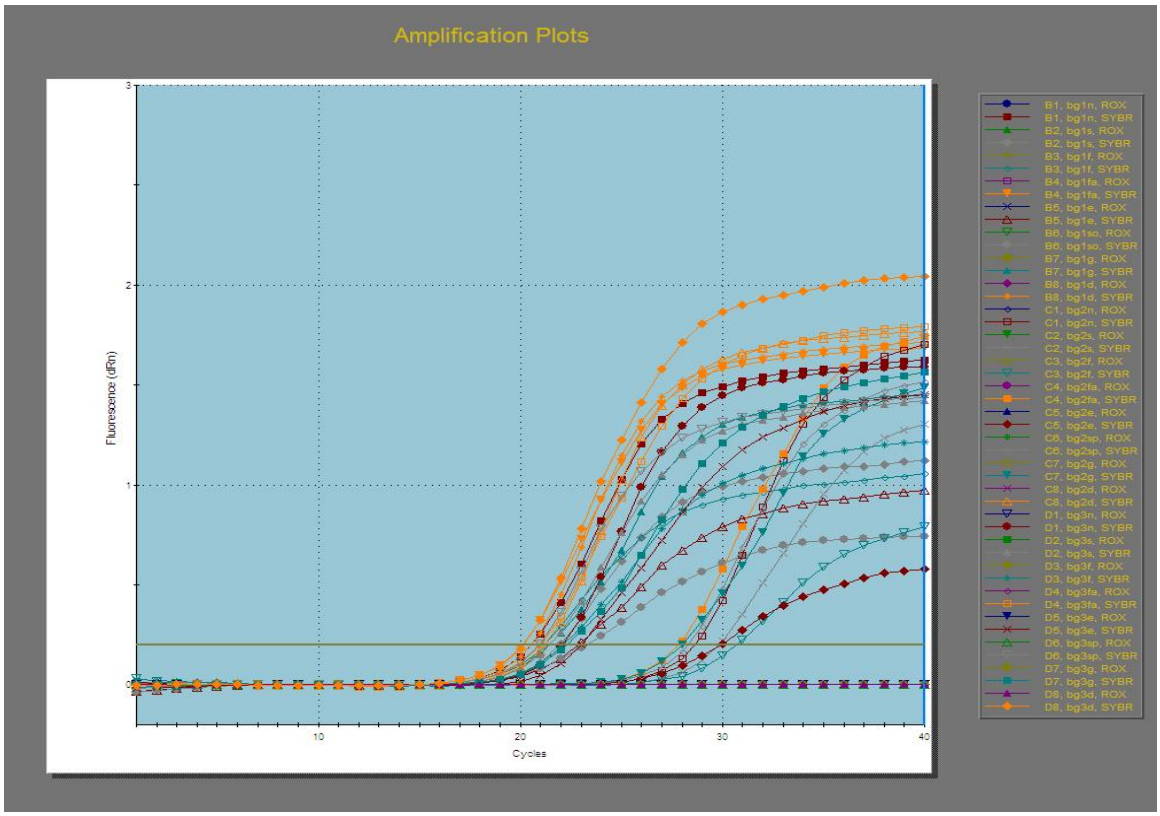
pMRH08 (negative control),BAP1 + pMRH08 + either pDCS11 (σ^{54}), pDCS59 (σ^{fecI}), or pNDP6 (σ^{S}), KC01 + pET21c + pMRH08 (negative control), and KC01 + pMRH08 + either pDCS11 (σ^{54}), pDCS59 (σ^{fecI}), or pNDP6 (σ^{S}). Strains were grown in 2 mL LB medium in shake flasks at 37 °C, at $O.D._{600}$ = 0.4 cultures were induced with 1 mM IPTG and incubated at 20 °C for 24 hours after which the cells were harvested for RNA isolation. In concert with RNA isolation the broth from each culture was extracted using a solid phase extraction resin (XAD-16) and the organic extracts were analysed by ESI-LC-MS/MS for oxytetracyline production as described below. qPCR experiments were performed in triplicate for each strain with a 10-fold dilution series of the 16s RNA acting as an internal standard. Primers were designed (see supplementary information, Table S1) to provide 180-210 bp amplicons of the *rpoN, rpoS, fecI, fecA, entD, glnA, spoT, dksA, oxyB, and oxyK* transcripts. The qPCR reactions contained 1.0 µM of each primer, 12.5 µL Absolute SYBR Green QPCR mix, 0.5 µL prepared cDNA, and 10.5 µL dH_2O to give a total reaction volume of 20 µL. Standard curve R^2 values and amplification efficiency values ranged from 0.991 - 1.0 and 90.7% - >100% respectively. The amplification efficiency was calculated using the formula $A = 10^{(-1/slope)}$ in which the slope was calculated by regression analysis obtained from the C_t values versus calculating log number of cells in serial dilutions.

Table 4.6.1: Primer sequences used for qRT-PCR from cDNA. Primers were used at 0.1μM for PCR reactions.

TRANSCRIPT	FORWARD	REVERSE
16s RNA	aactgcctgatggaggggg	gtctcagttccagtgtggct
rpoN	gtctgttgcagttgtcgacg	cggtgtaaatggtgtcccaa
oxyB	tcaccggcctcggtgtggtc	gcgcgagccgggtcatcagg
oxyK	ccctgatcagggacgccgcc	tcgcggccaccggaggcgaa
*rpoS*gacgaagcatacggctgacgt	acgcaacctggtggattc	
*fecI*cttcacgtgttttgccgtttagcccgtctccgcgatcctcgctccttcc		
*fecA*cttcttcgtgcgtgcct	atgaaatgcgttactacaccg	
*entD*accgcagtctgcgagtcgctctgtgtgcccgcaatcggcgag		
*glnA*ttagcgtgtttgattacg	gatctacaaatatgttgtgca	
*spoT*gagatgatccaaaaaatcctctctgaacatctggcccagcacgcgata		
*dksA*ctttttttcagcgtcttctcgatctttt agctggcgcacttccgt		

Table 4.6.2: Plasmid table.

PLASMID	DESCRIPTION
DCS11	*E. coli rpoN* in pKH22
DCS59	*E. coli fecI* in pKH22
NDP6	*E. coli rpoS* in pKH22
MRH08	*S. rimosus* oxytetraycline biosynthetic pathway in pET28b
pCP20	Temperature sensitive vector for removal of inserts possessing Flp sites



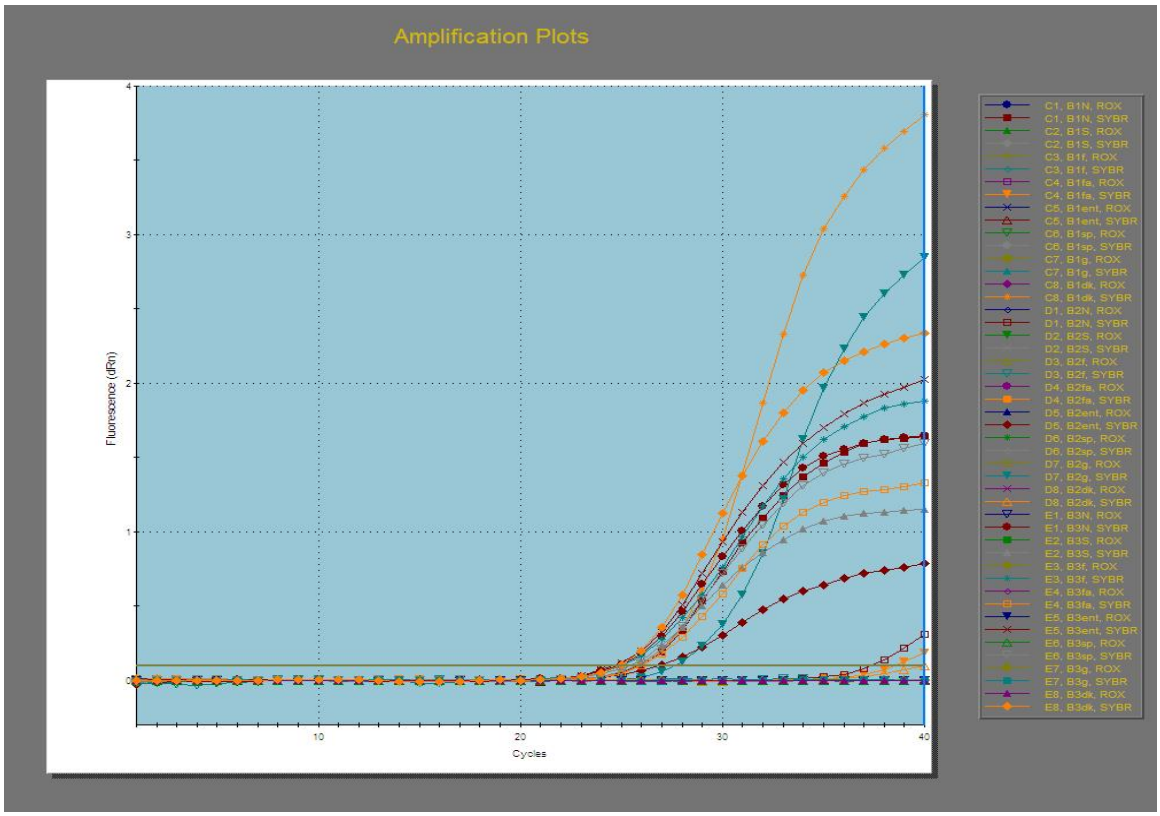
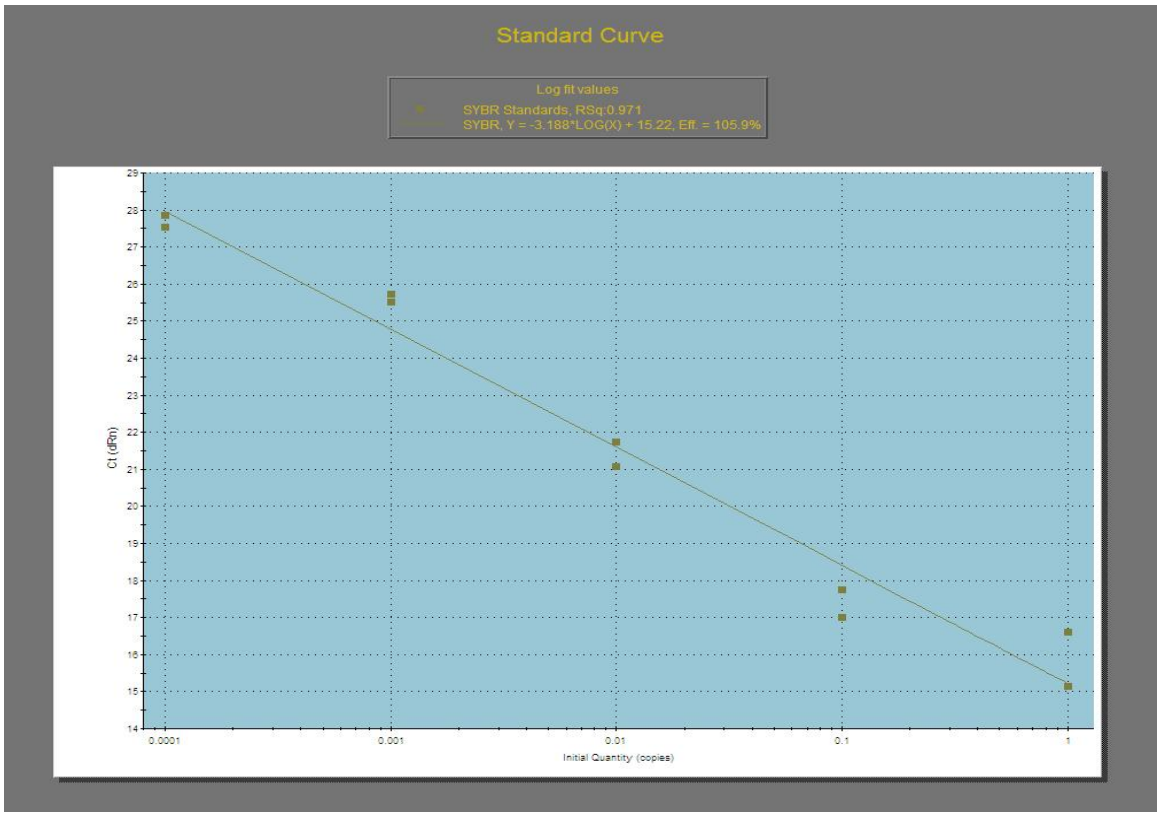


Figure 4.6.1: Amplification plots from *E. coli* BAP1+ pET21c + pET28b



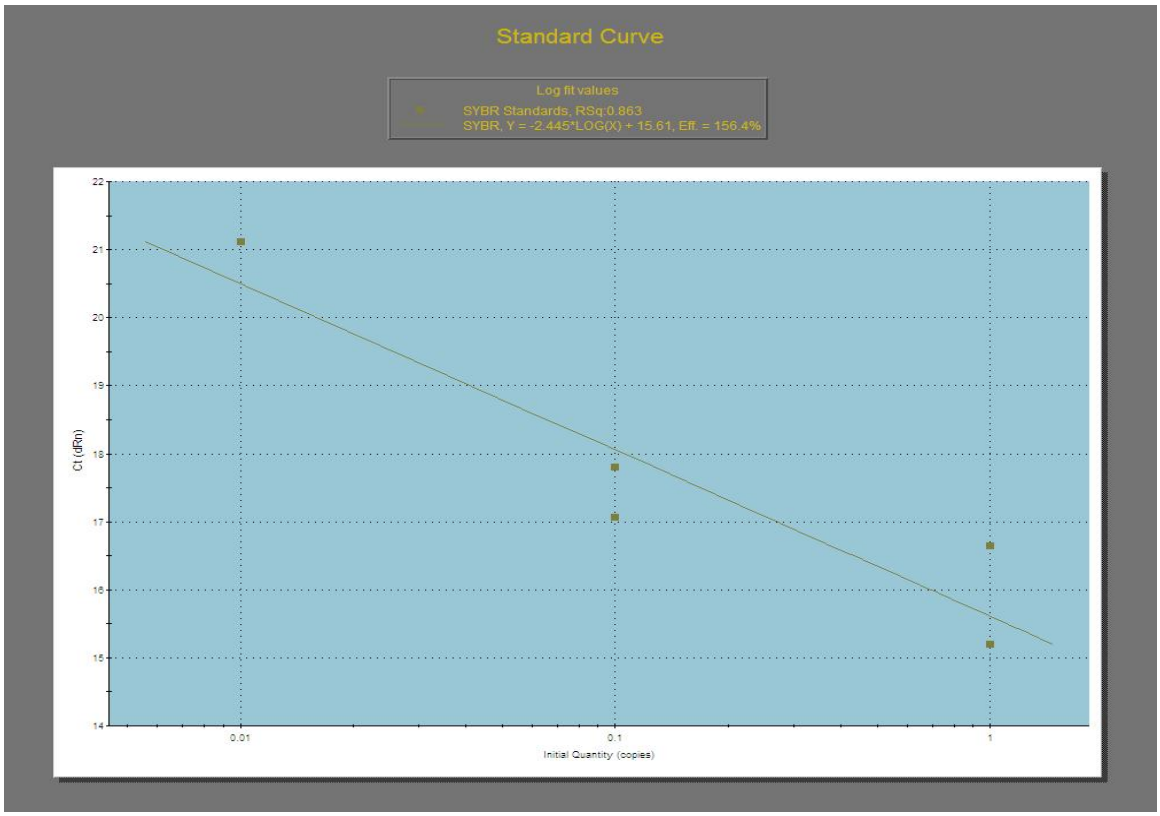


Figure 4.6.2: Standard curves from *E. coli* BAP1+ pET21c + pET28b

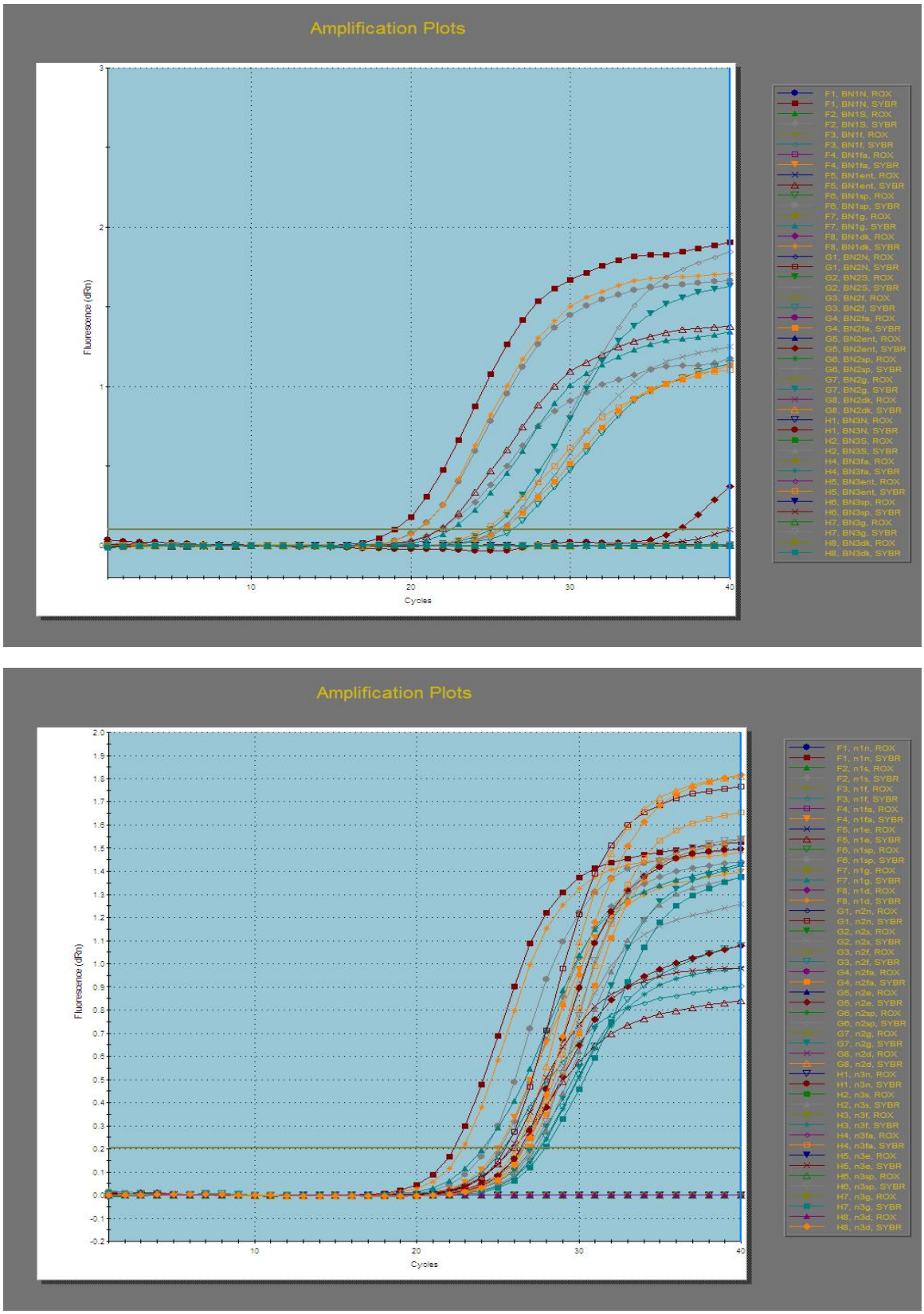
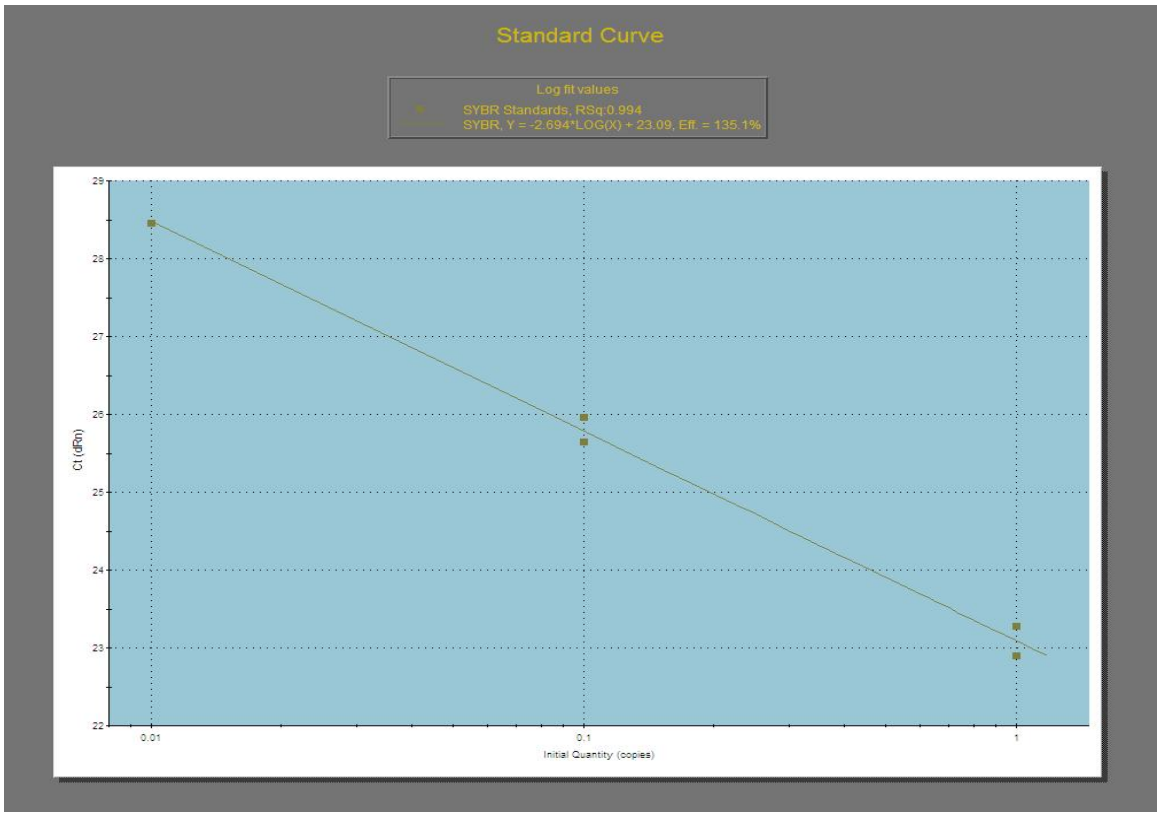


Figure 4.6.3: Amplification plots from *E. coli* BAP1+ pDCS11 + pMRH08



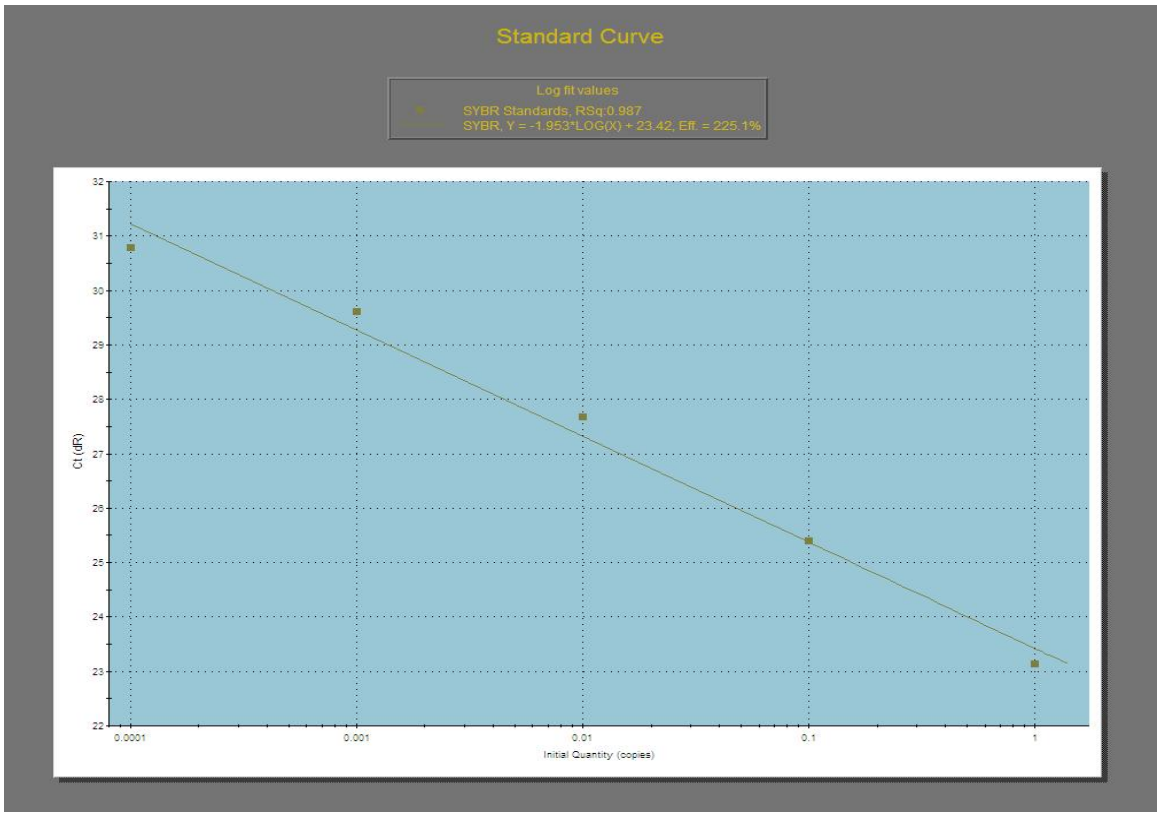
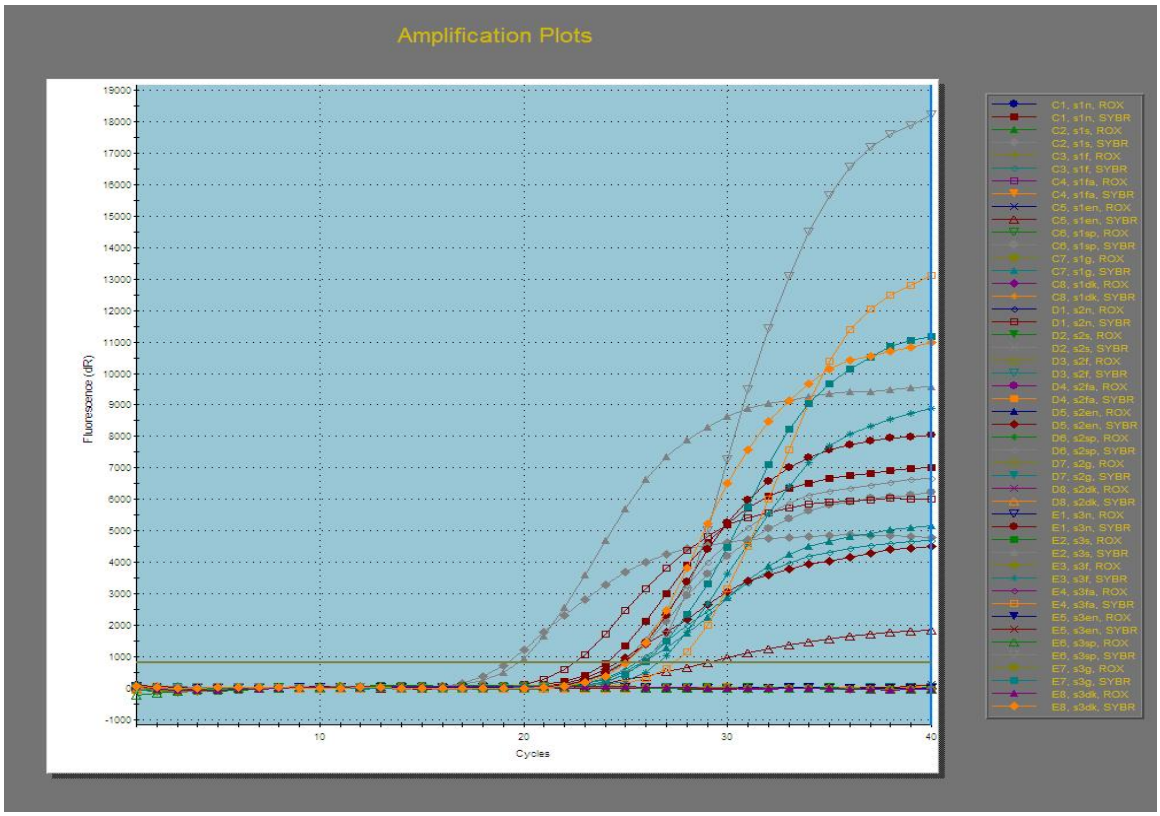


Figure 4.6.4: Standard curves from *E. coli* BAP1+ pDCS11 + pMRH08



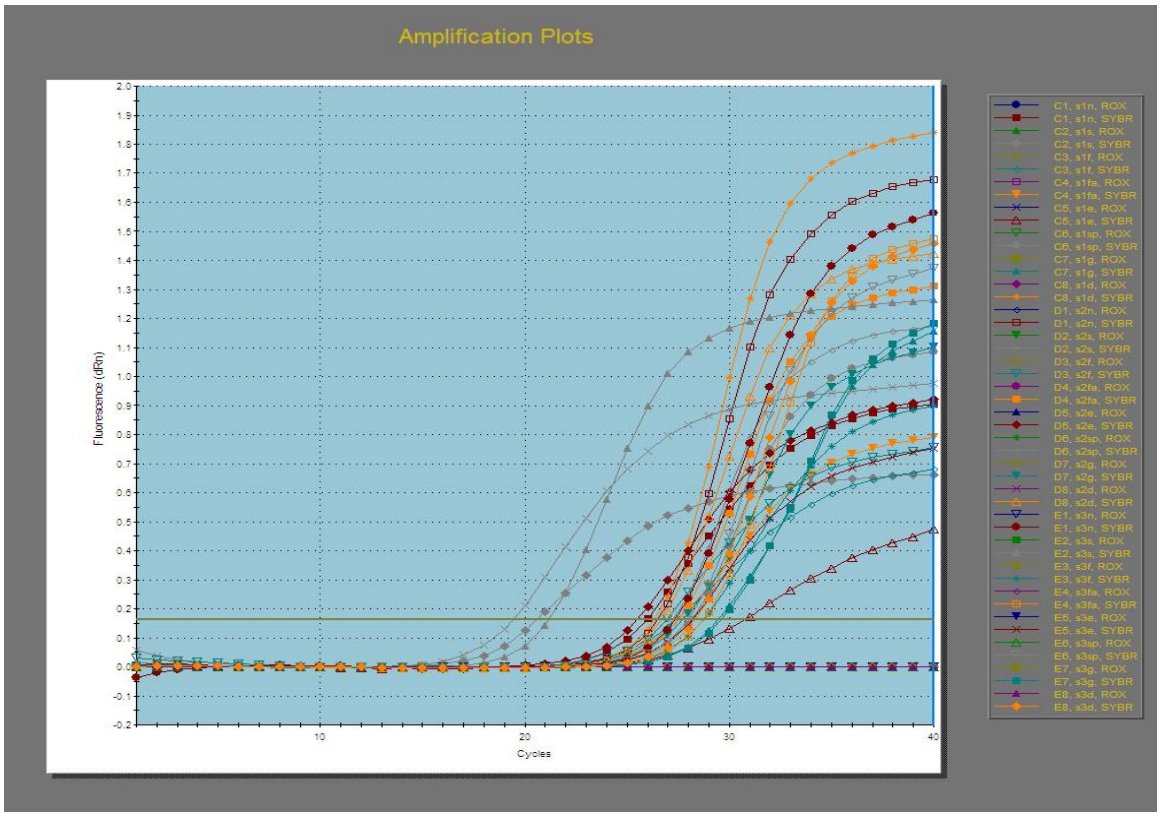
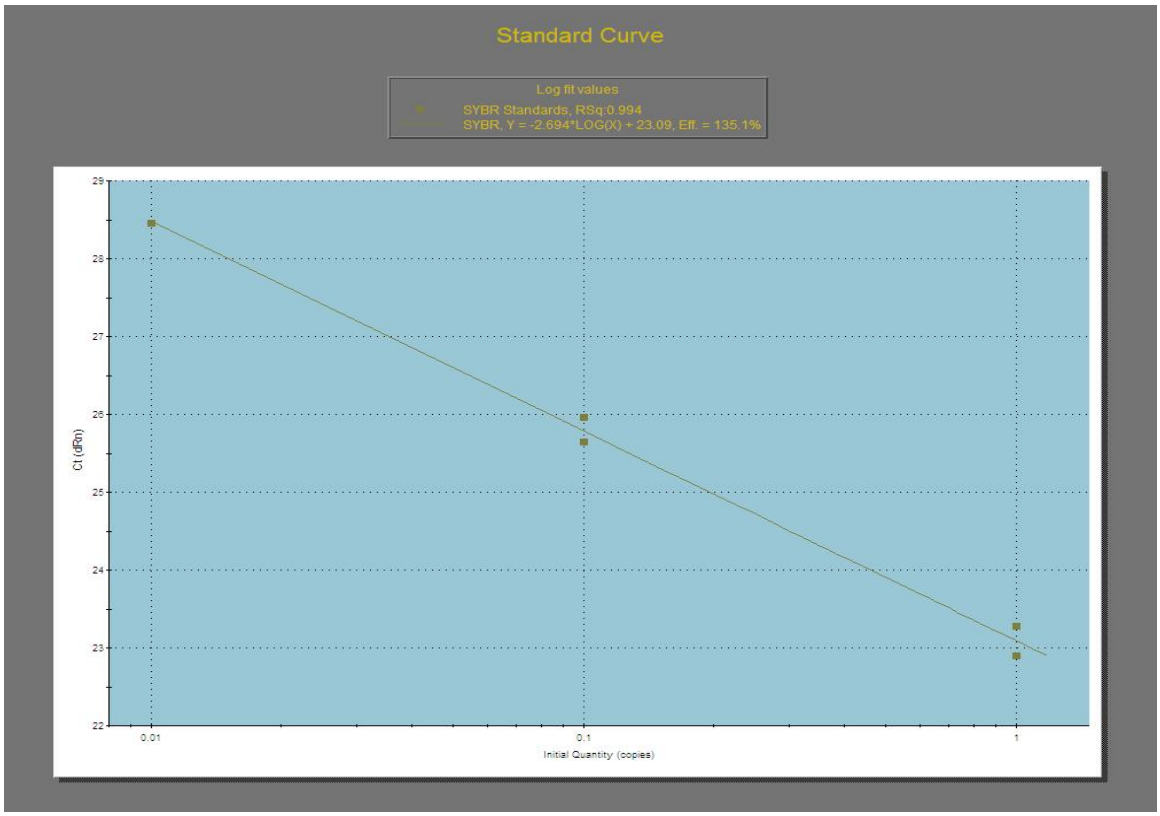


Figure 4.6.5: Amplification plots from *E. coli* BAP1+ pNDP6 + pMRH08



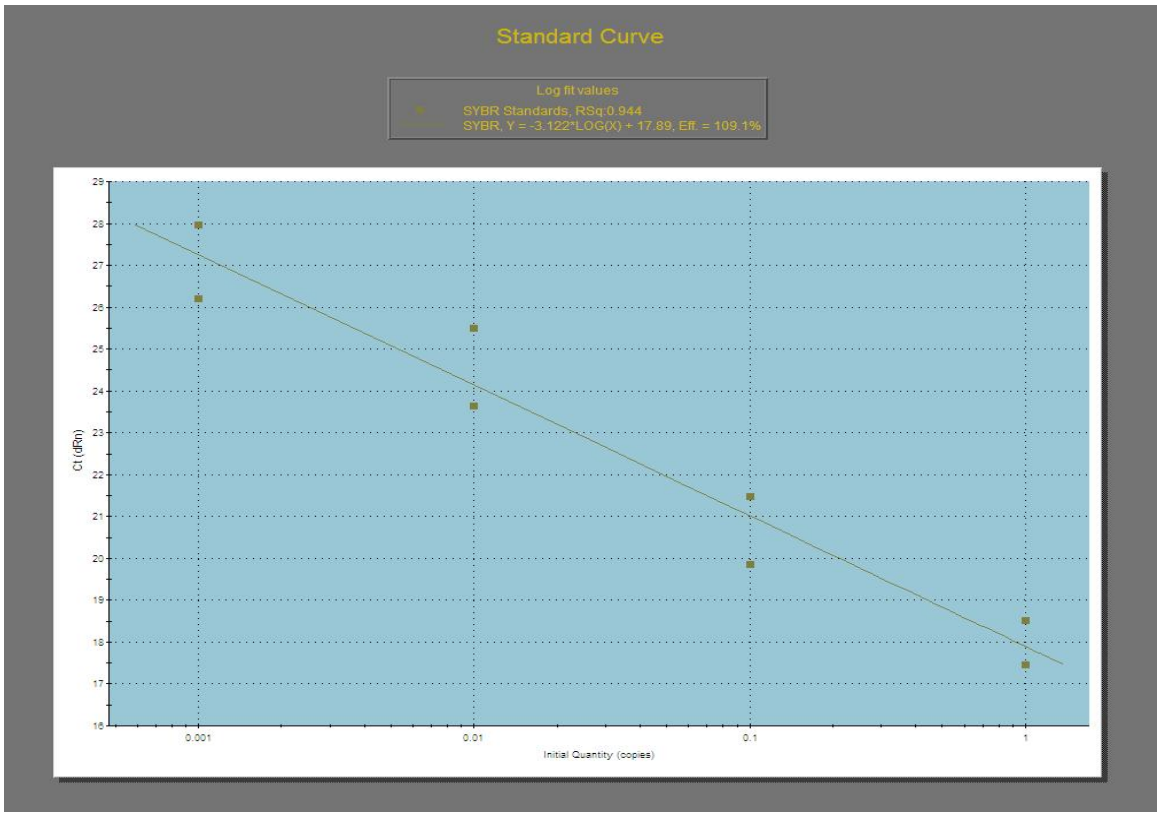
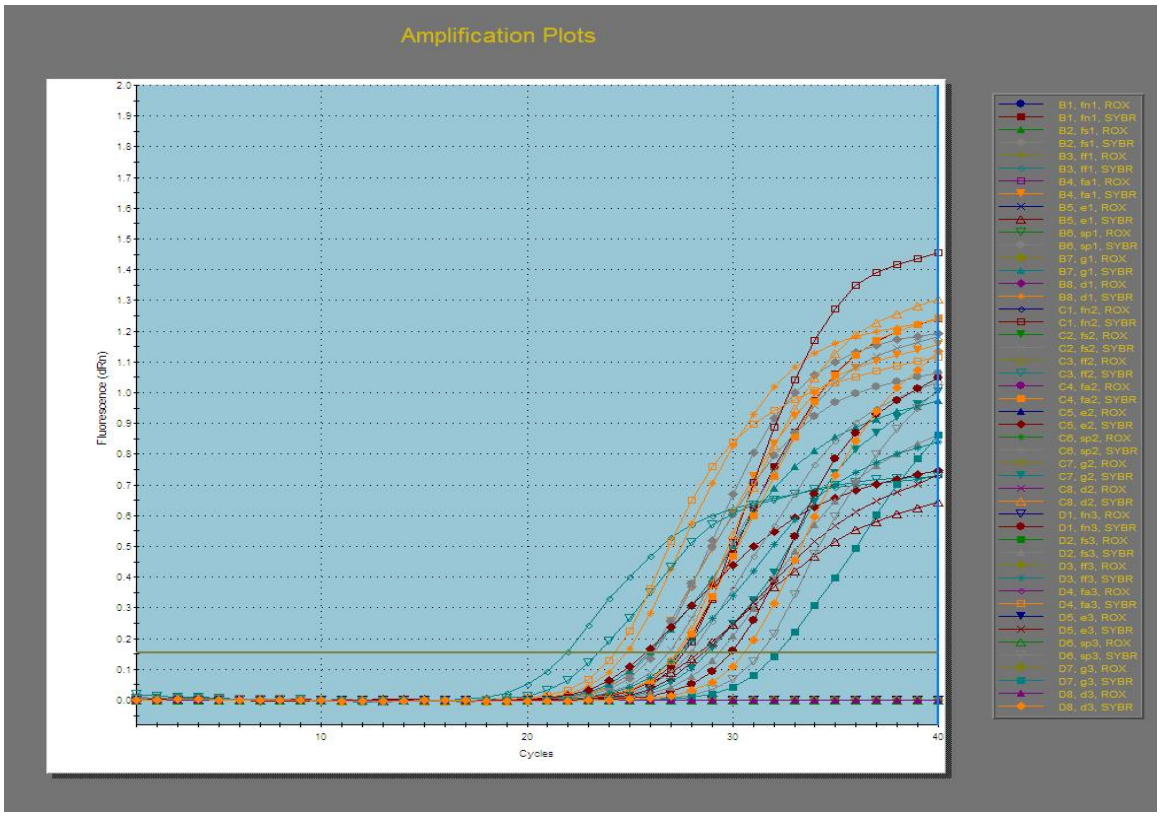


Figure 4.6.6: Standard curves from *E. coli* BAP1+ pNDP6 + pMRH08



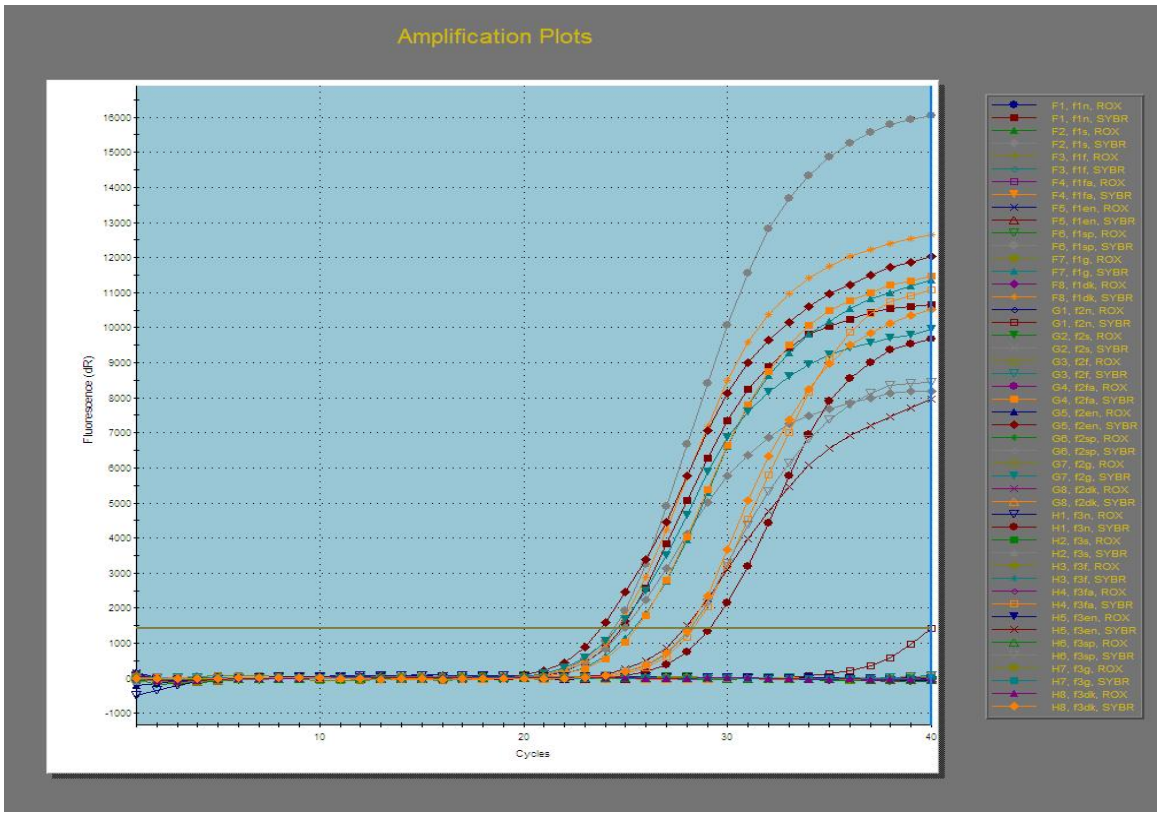
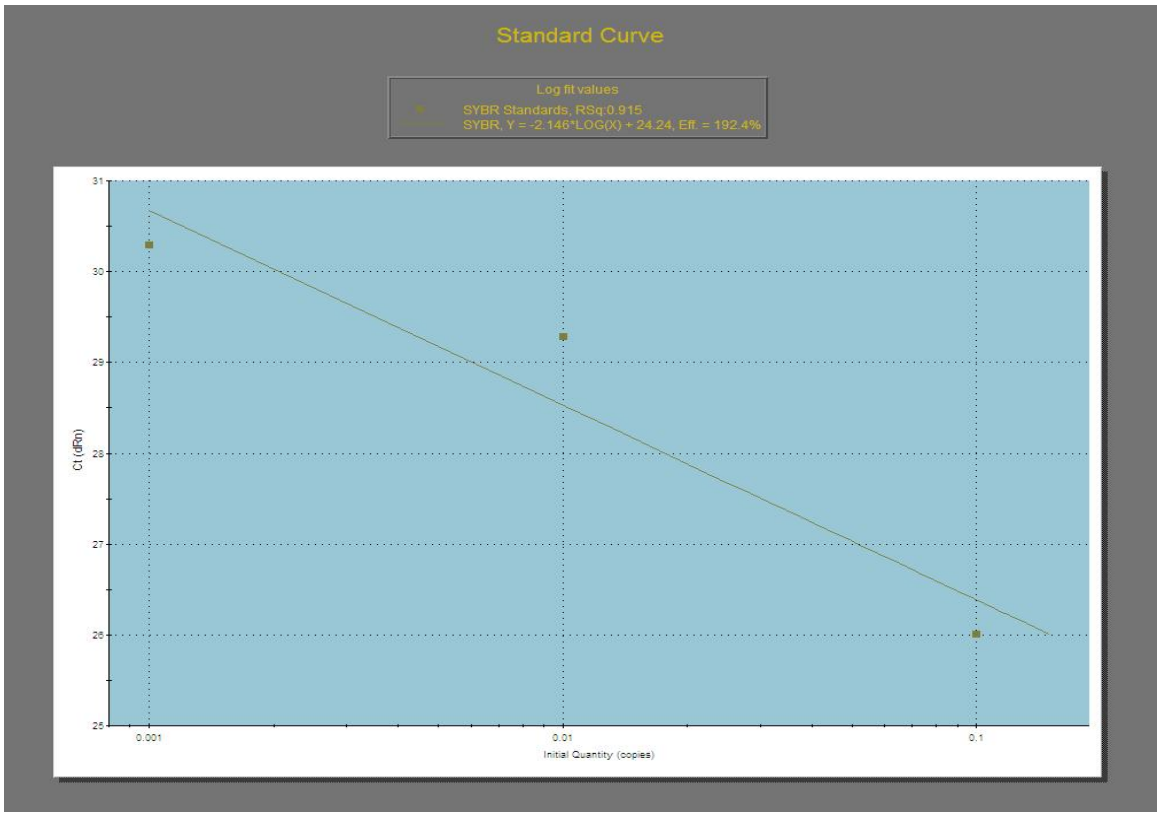


Figure 4.6.7: Amplification plots from *E. coli* BAP1+ pDCS59 + pMRH08



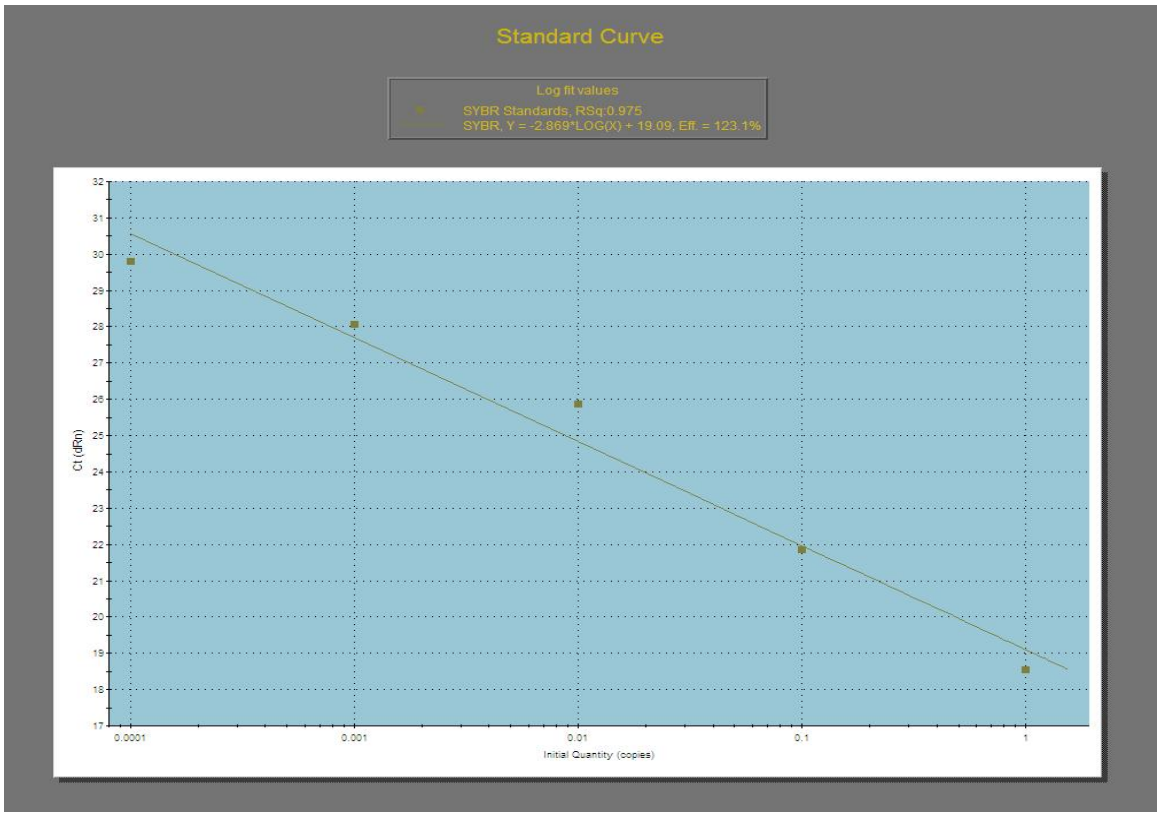


Figure 4.6.8: Standard curves from *E. coli* BAP1+ pDCS59 + pMRH08

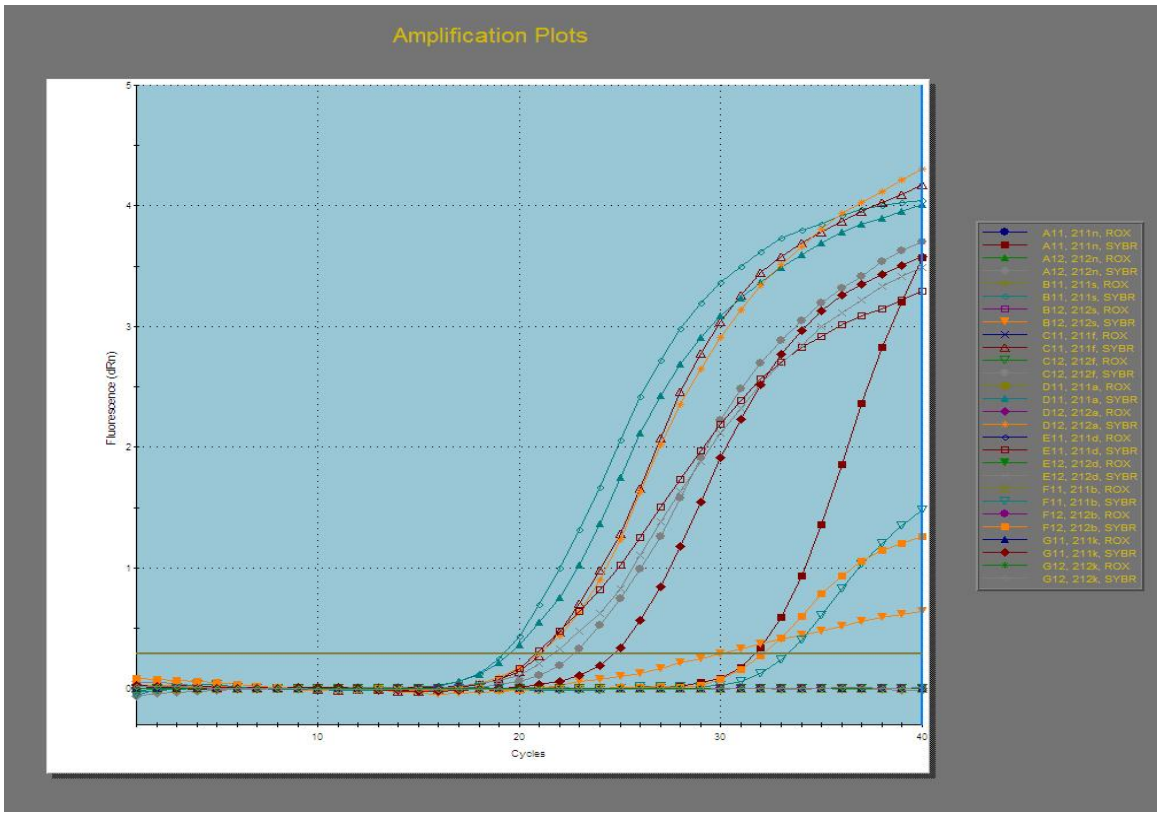


Figure 4.6.9: Amplification plots from *E. coli* KC01+ pET21c + pET28b

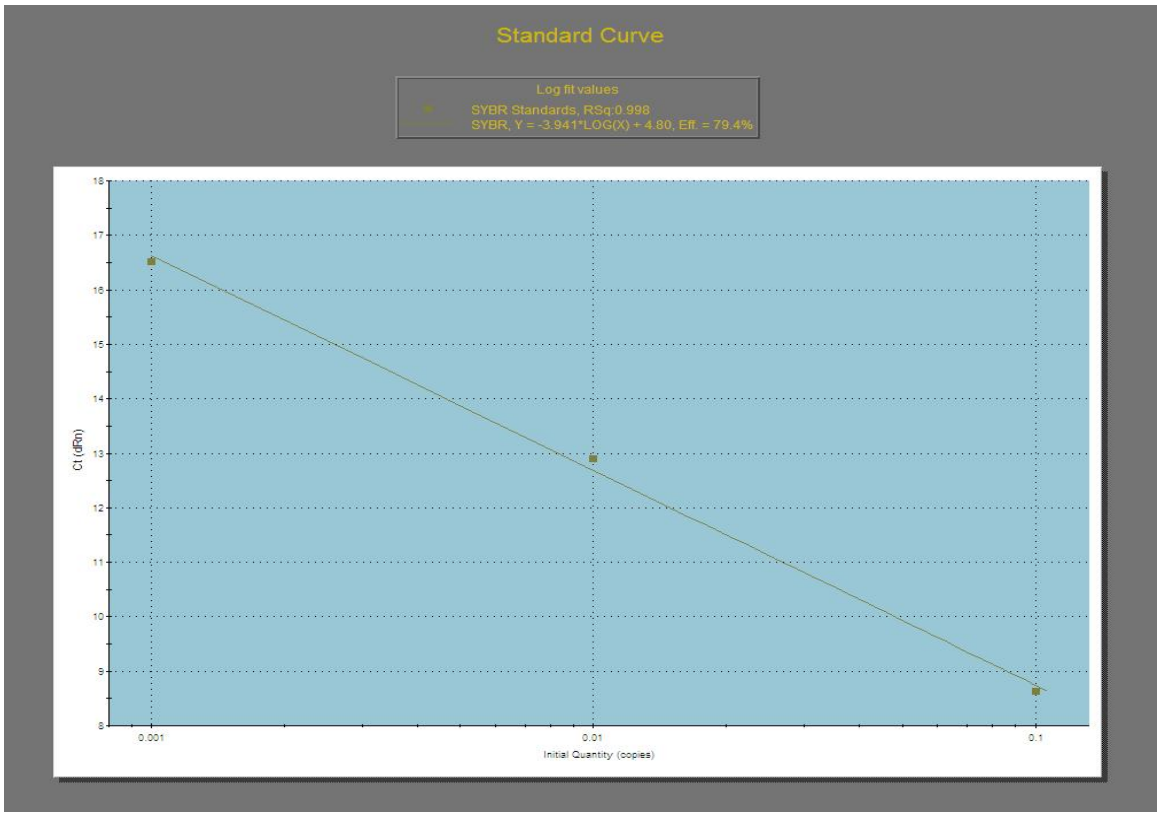


Figure 4.6.10: Standard curves from *E. coli* KC01+ pET21c + pET28b

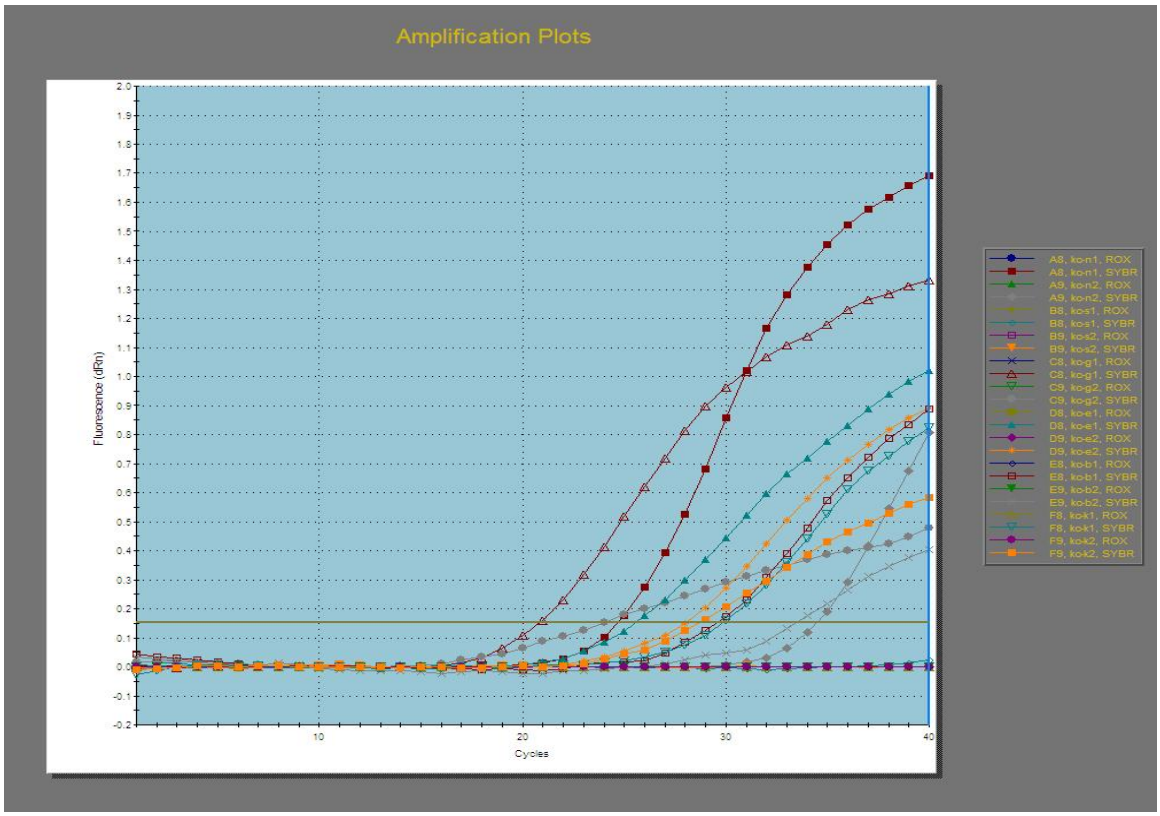


Figure 4.6.11: Amplification plots from *E. coli* KC01+ pET21c + pMRH08

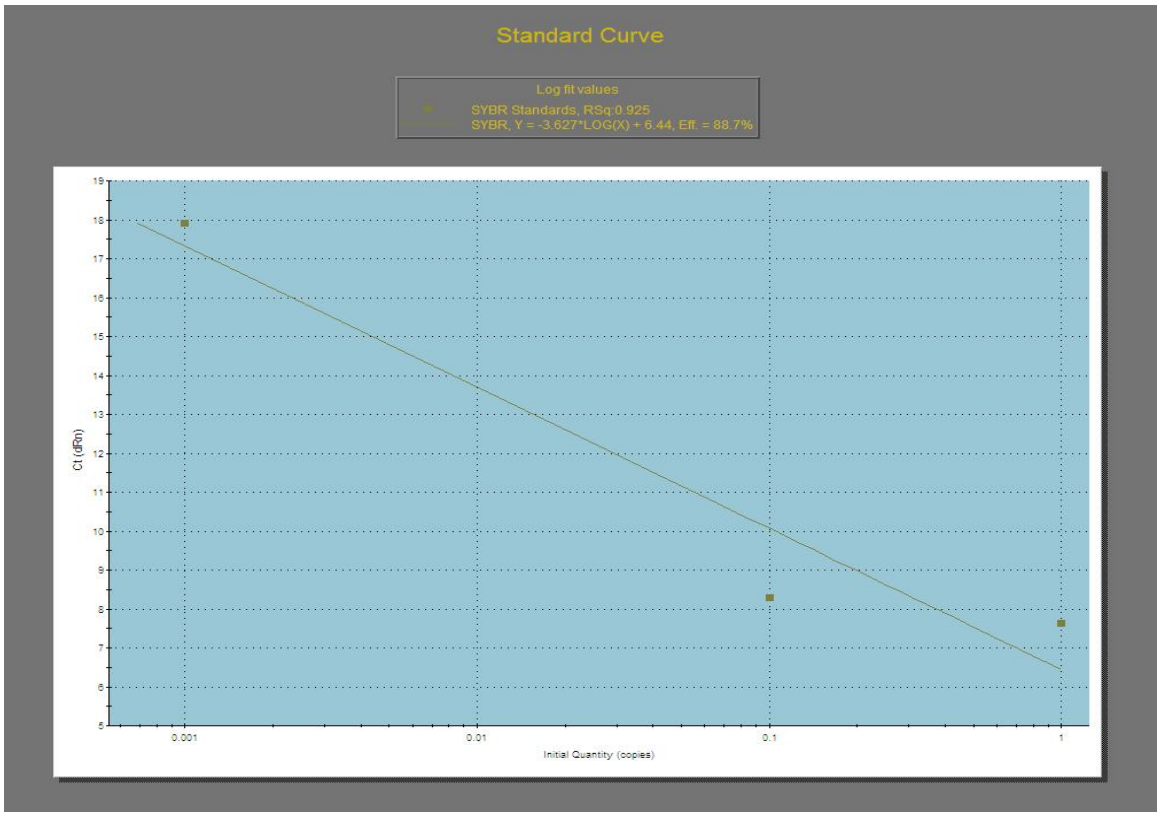
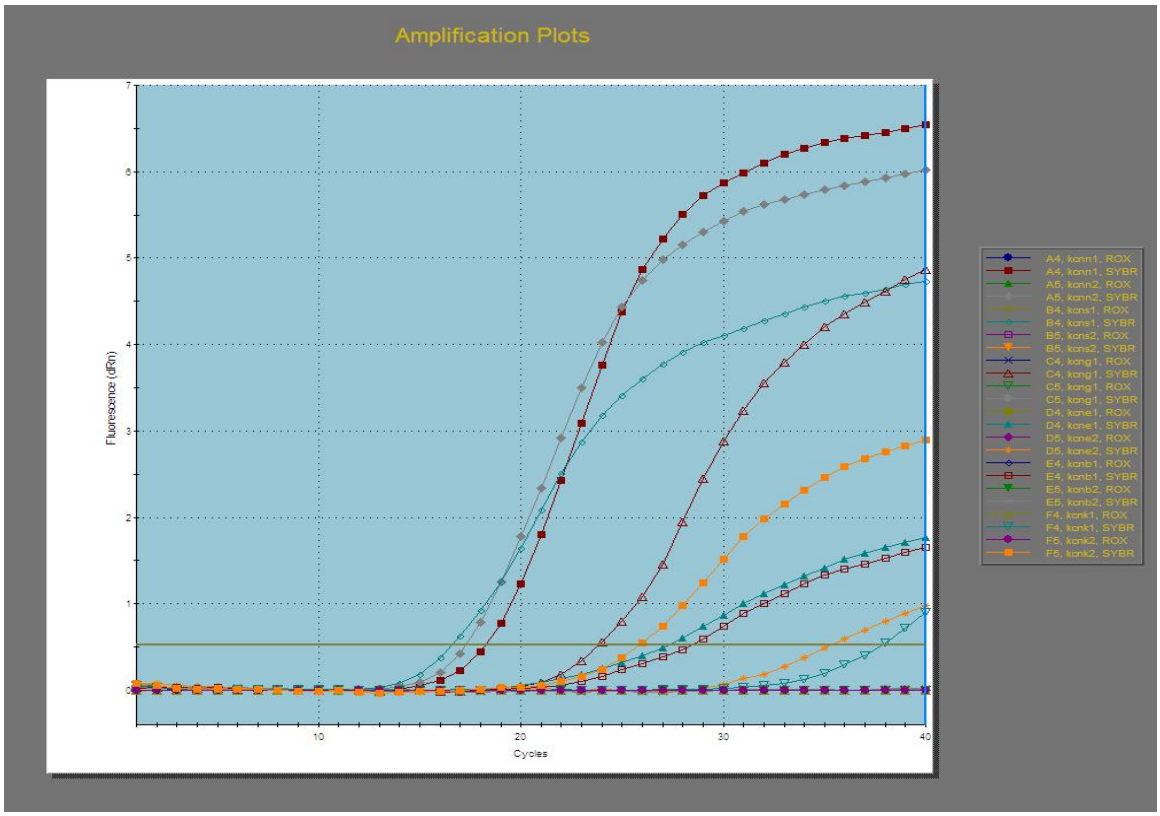


Figure 4.6.12: Standard curves from *E. coli* KC01+ pET21c + pMRH08



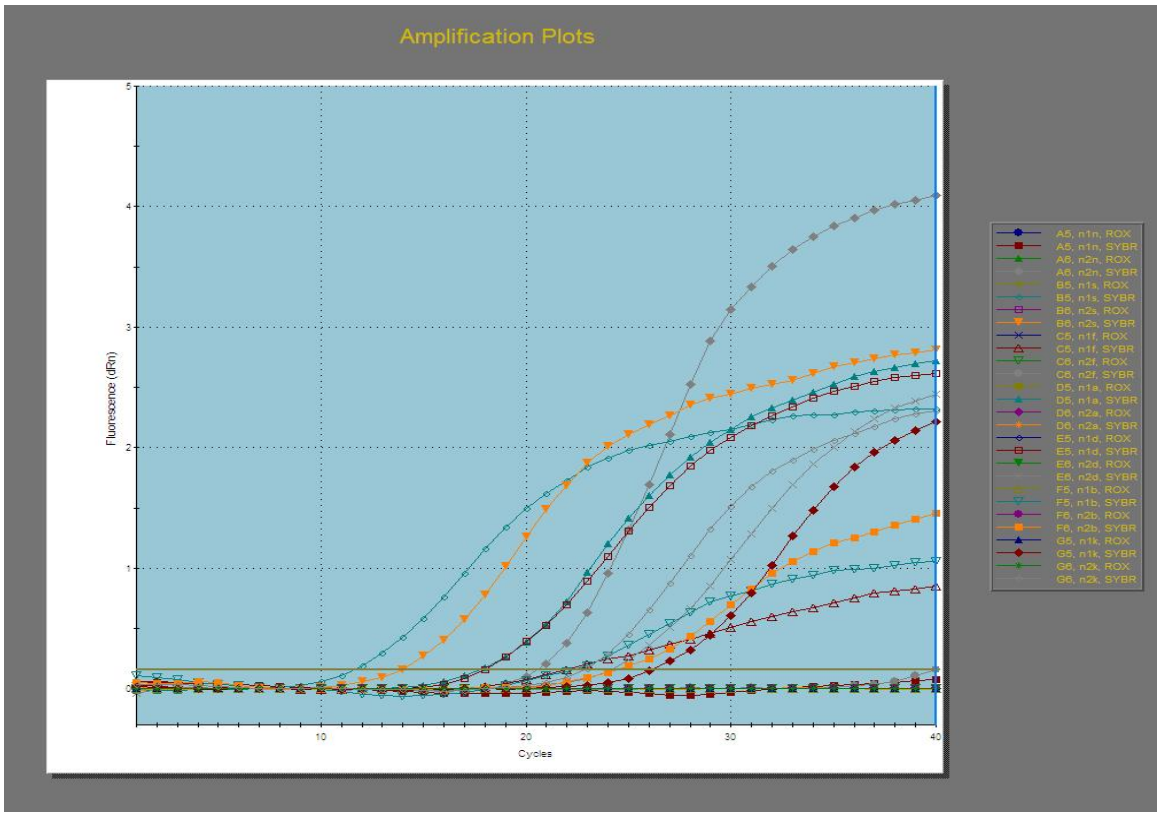
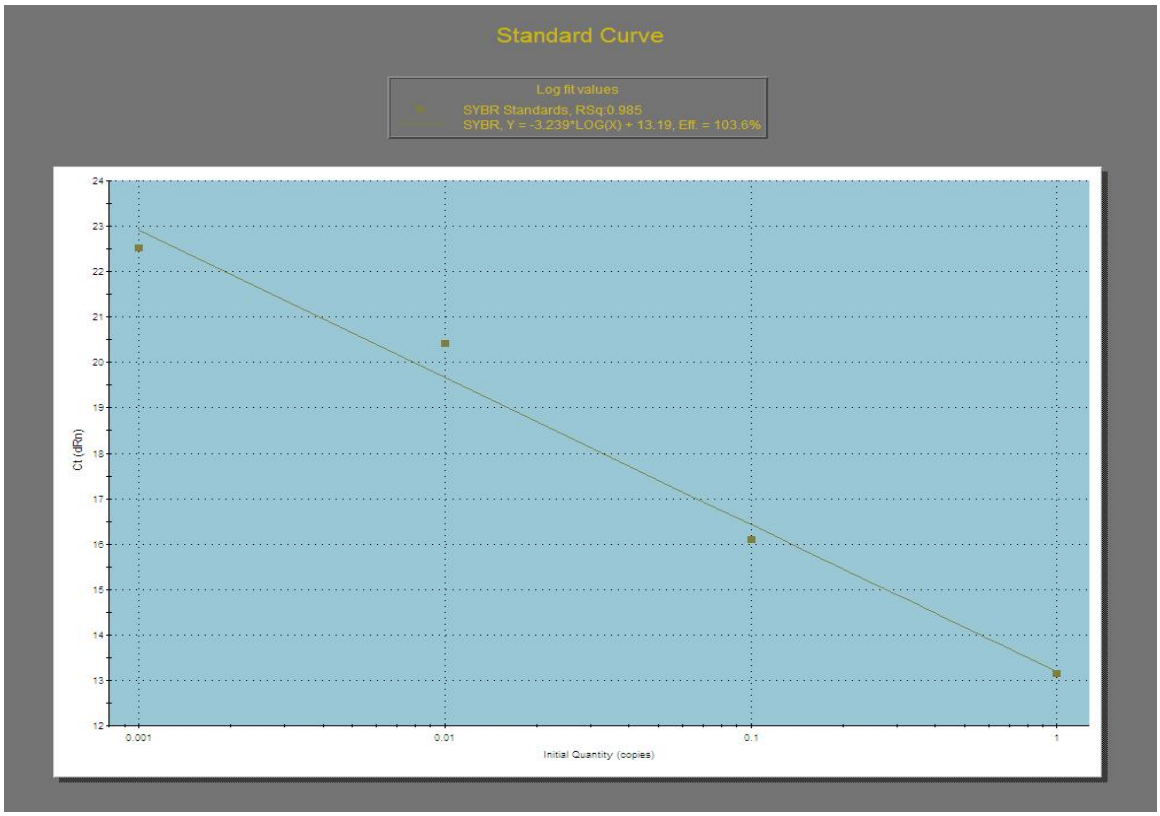


Figure 4.6.13: Amplification plots from *E. coli* KC01+ pDCS11 + pMRH08



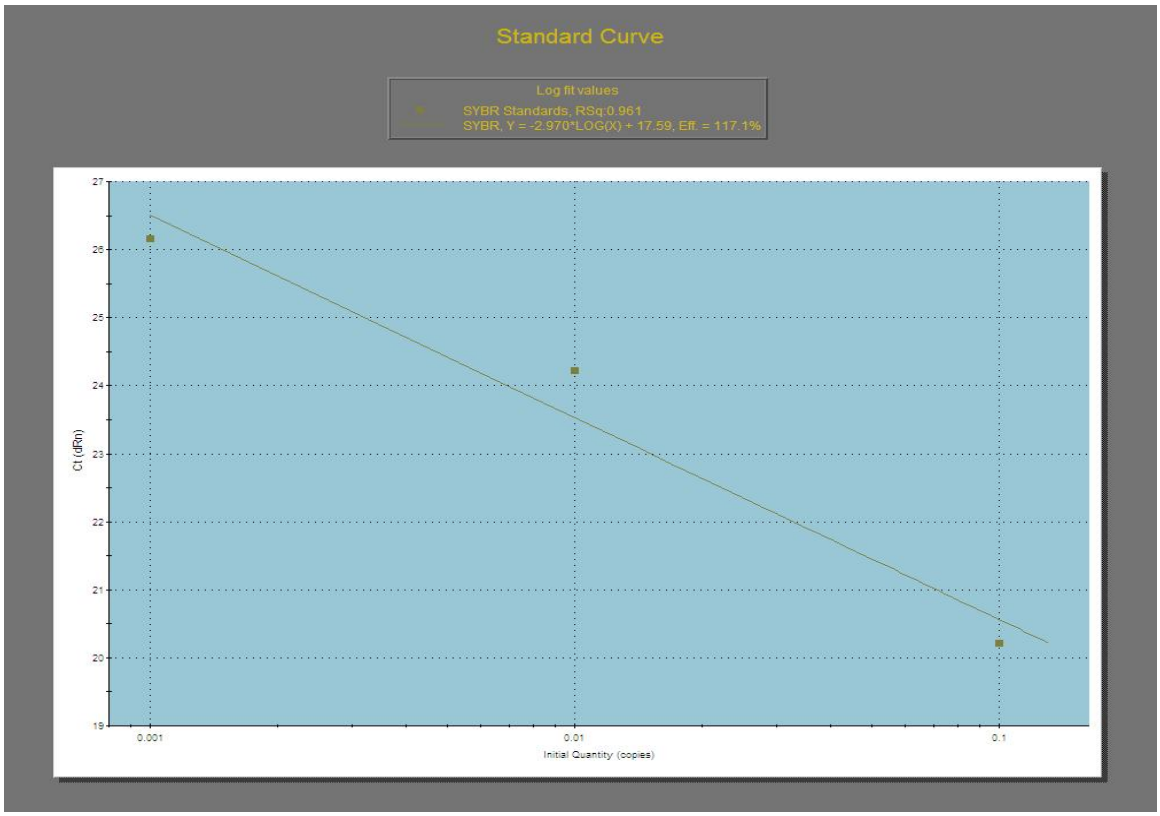
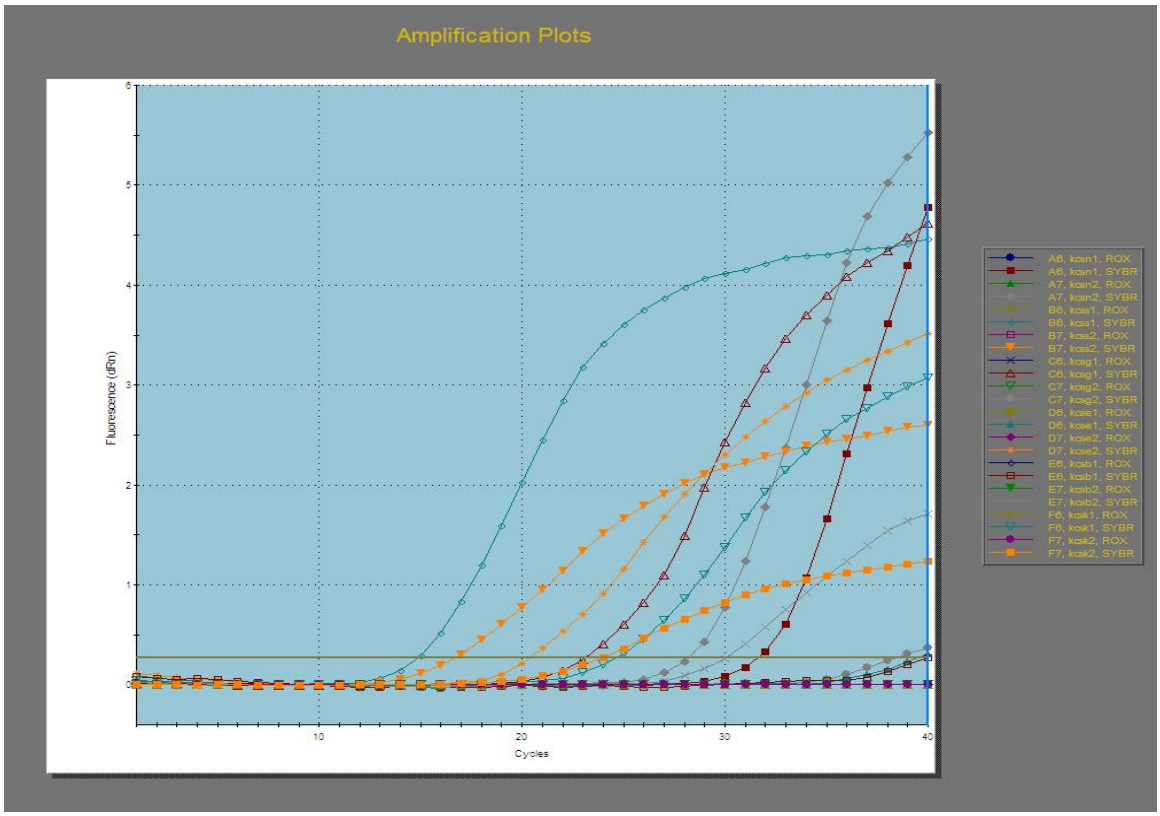


Figure 4.6.14: Standard curves from *E. coli* KC01+ pDCS11 + pMRH08



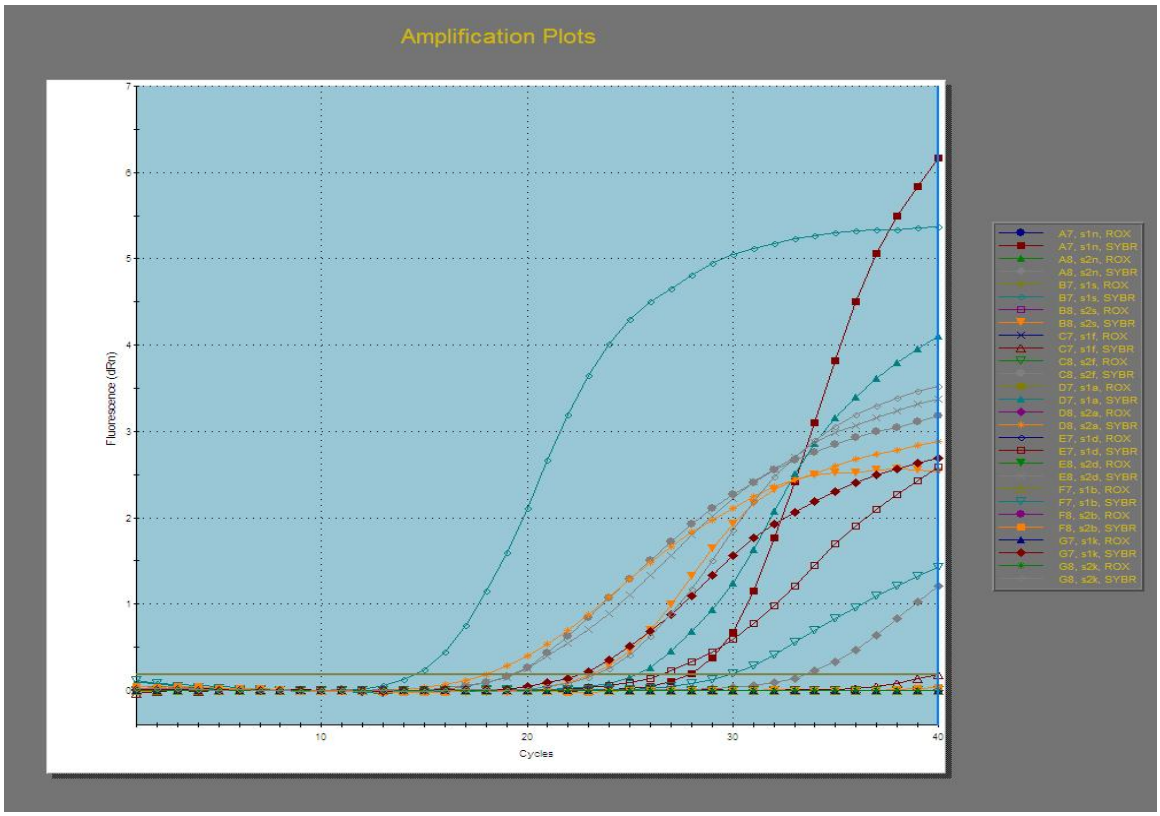
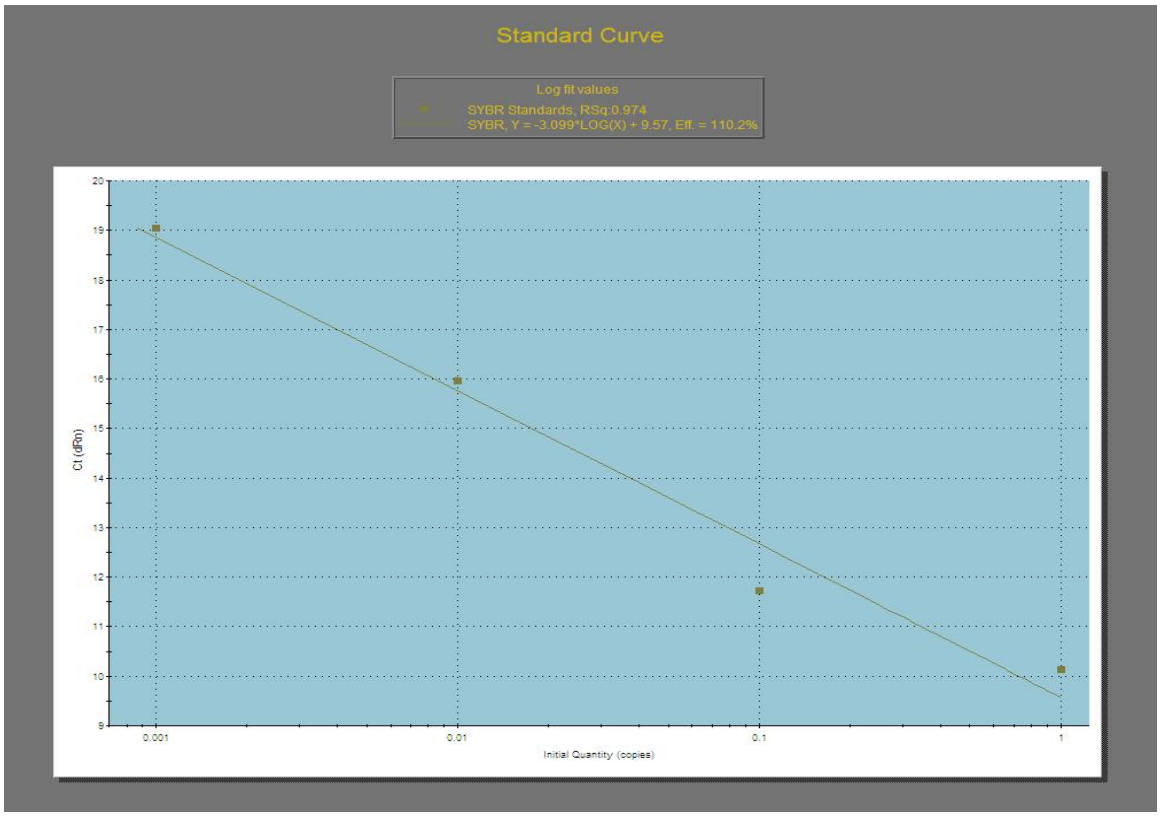


Figure 4.6.15: Amplification plots from *E. coli* KC01+ pNDP6 + pMRH08



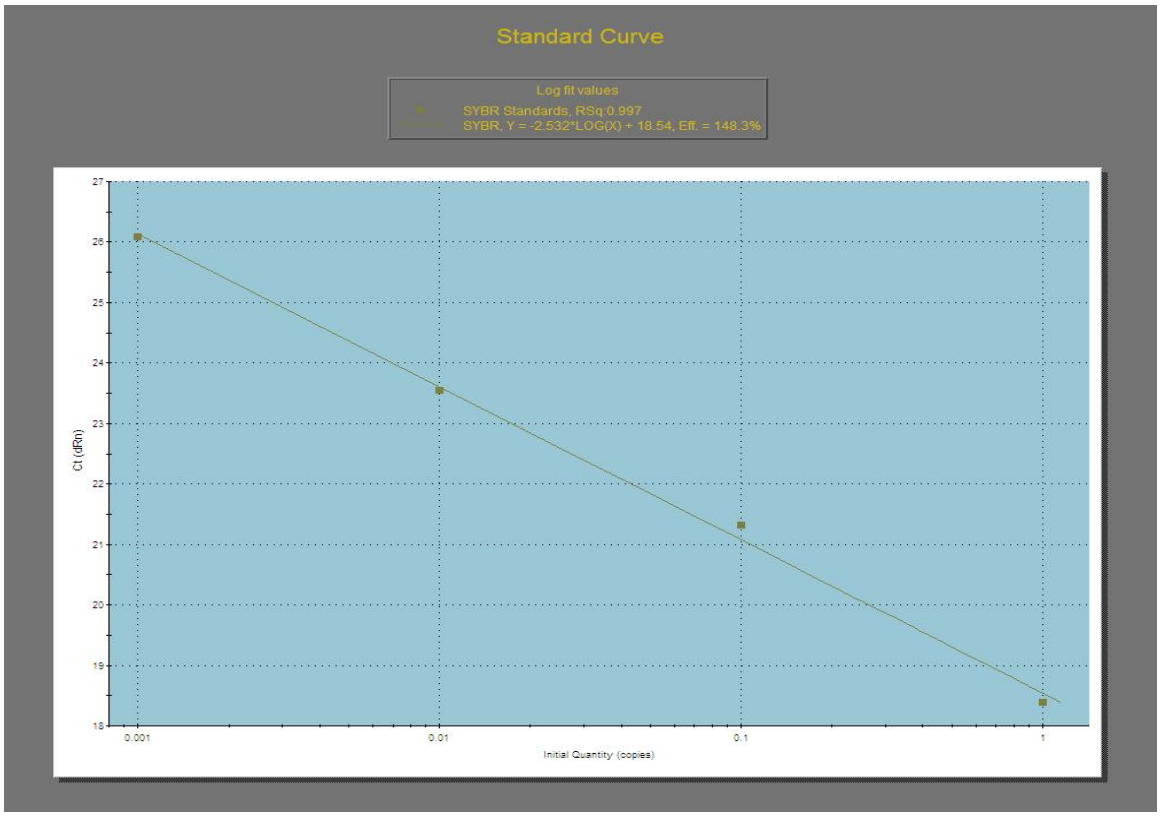


Figure 4.6.16: Standard curves from *E. coli* KC01+ pNDP6 + pMRH08

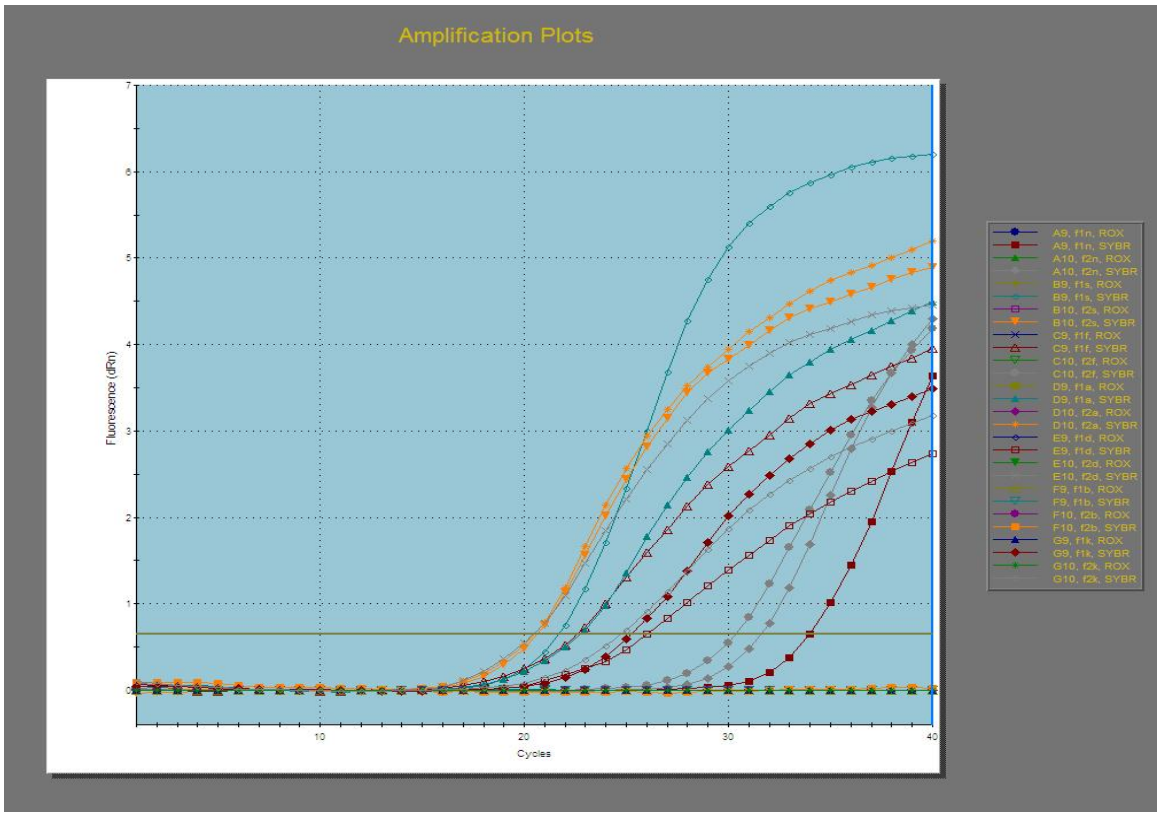


Figure 4.6.17: Amplification plots from *E. coli* KC01+ pDCS59 + pMRH08

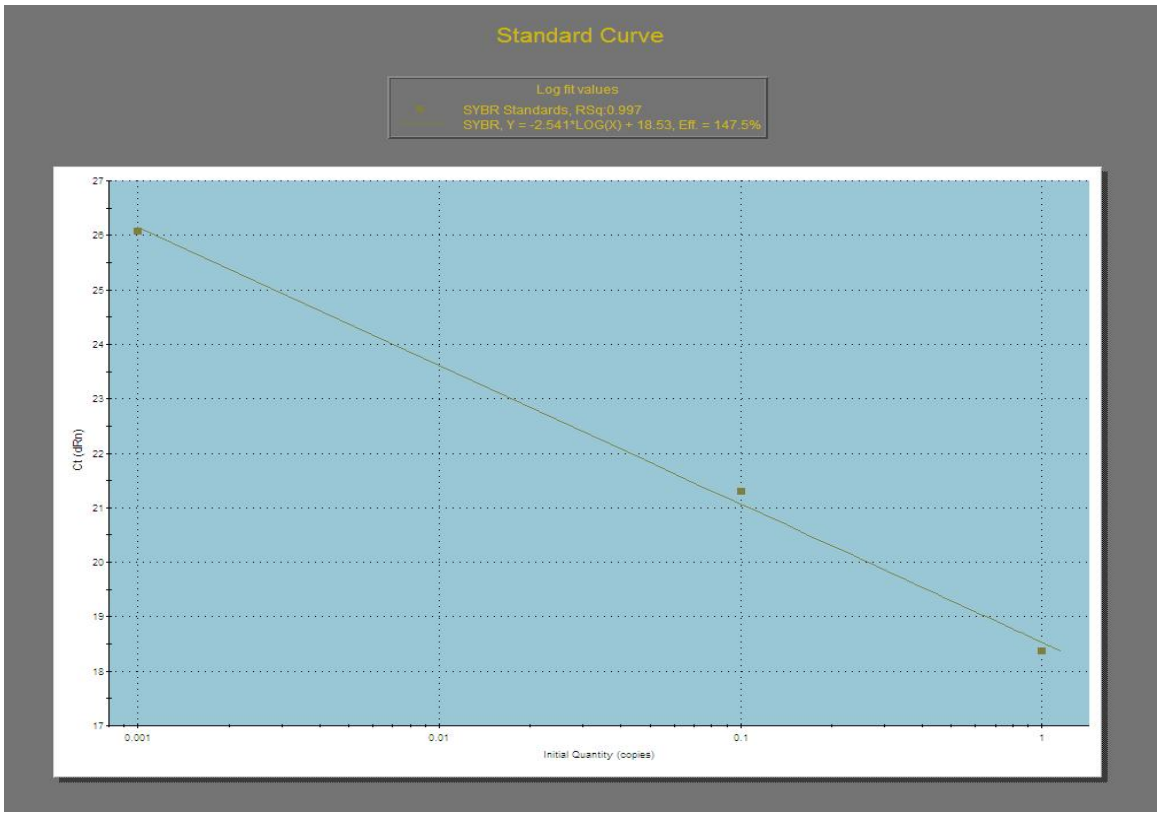


Figure 4.6.18: Standard curves from *E. coli* KC01+ pDCS59 + pMRH08

Chapter 5: Conclusion and Contributions

5.1 Introduction

Bacterial natural products are excellent lead compounds for drug discovery and have played major roles in the development of pharmaceutical agents in nearly all therapeutic areas. Bacterial polyketides possess an enormous range of chemical diversity and biological function. Sequencing of bacterial genomes, especially those of major polyketide producers have shown that there are many more polyketide biosynthetic pathways than there are polyketides isolated from standard cultivation techniques. These genetically encoded polyketide natural products from cultivatable and uncultivatable bacteria represent one of the greatest remaining untapped reservoirs of new natural product diversity. Methods to effectively access this diversity, such as heterologous expression, will have a major impact on drug discovery. Heterologous expression has proven to be a valuable asset in the discovery, production, engineering, and characterization of bacterial secondary metabolites and the complex enzymology involved in their biosynthesis. The goals of this research are to 1) contribute new methodology focusing on heterologous expression to allow access to this untapped diversity of polyketide products and 2) increase the understanding of native and heterologous regulatory mechanisms that occur during polyketide production.

5.2 Results

Throughout this research we have developed two unique heterologous expression platforms allow future access to undiscovered chemical diversity. Although *Myxococcus xanthus* has been previously reported as a successful host in the literature, we have reported the first example of a *myxobacteria* capable of producing a *streptomycetal* secondary metabolite. *M. xanthus* ability to produce oxytetracycline presents *M. xanthus* as a potential culture independent

heterologous host capable of screening metagenomic libraries without the need of further library engineering. Our developed *Escherichia coli* platform for heterologous expression represents a fundamentally new mechanism for heterologous expression of polyketide biosynthetic pathways, where a high-level, pleiotropic alternative sigma factor from the heterologous host positively regulates transcription of the biosynthetic gene cluster. In contrast, known methods for heterologous expression rely on either replacing each native promoter with known, well-characterized promoter from the heterologous host, such as the T7 promoter in *E. coli*, or rely on the heterologous host to constitutively express each gene from the native promoters. Our approach which actively induces transcription of gene clusters by over-expression of alternative sigma factors may provide a general solution to the heterologous expression problem that is compatible with screening DNA libraries. This work also reports the first successful production of a type II polyketide product from an *E. coli* host. Through this research we have also suggested that over-expression of other sigma factors, such as σ^H, σ^S or FecI, may positively regulate transcription of other biosynthetic gene clusters, potentially expanding the scope of alternative sigma factor-mediated heterologous expression. Through investigation of the heterologous regulatory mechanisms occurring in our *E. coli* heterologous host we have reported the first evidence of a transcriptional amplication loop existing between σ^{54} and σ^S. These results further demonstrate the complex and subtle regulatory interplay between σ^{54}, σ^S, their regulons, and RNAP, which modulate key native cellular responses as well as non-native processes such as heterologous expression of secondary metabolite biosynthetic pathways.

5.3 Conclusion

Our work has contributed two unique heterologous hosts as capable of producing complex *streptomycetal* polyketides. This work has also presented the application of pleotropic alternative sigma factors as a fundamentally novel approach for the heterologous production of natural products. Through the application of our reported heterologous hosts and the further

investigation of alternative sigma factors as regulators of heterologous metabolite production, we have provided potential access to untapped chemical diversity and biological activities and contributed to the overall understanding of heterologous metabolite regulation and production.

Appendix

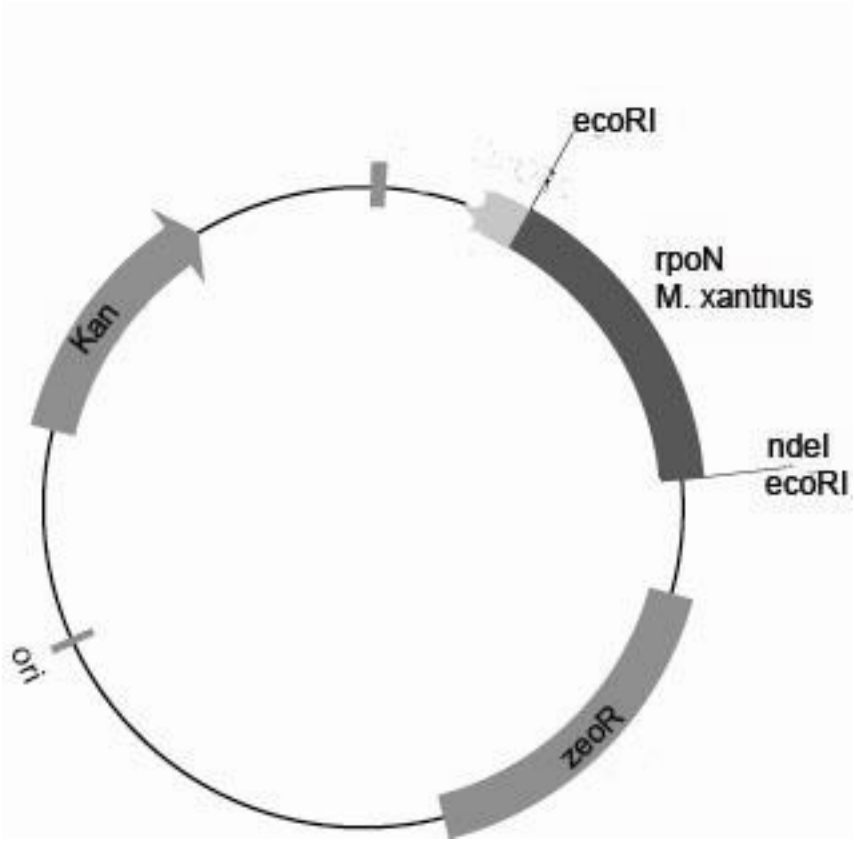
A1. Vector Tables

Table 5.1.1: Vectors constructed appendix

Name	Insert	Vector platform	Resistance
pDCS01	*M. xanthus rpoN*	pCR-Blunt	Kan
pDCS02	*M. xanthus rpoN*	pET21c	Amp
pDCS03	AmpR removed from pBR322	pBR322	Tet
pDCS04	carQ promoter	pCR-Blunt	Kan
pDCS05	carQ replaced T7 promoter	pET28b	Kan
pDCS06	asgE	pDCS05	Kan
pDCS07	*M. xanthus rpoN*	pET28b	Kan
pDCS09	Tetracycline repressor	pCR-Blunt	Kan
pDCS10	*E. coli rpoN*	pCR-Blunt	Kan
pDCS11	*E. coli rpoN*	pKH22	Amp
pDCS12	*M. xanthus rpoN*	pKH22	Amp
pDCS13	*B. subtilis sfp*	pCR-Blunt	Kan
pDCS14	*B. subtilis sfp*	pKH22	Amp
pDCS15	*B. subtilis sfp*	pDCS12	Amp
pDCS18	*E. coli rpoN*	pKH61	Cam
pDCS19	*M. xanthus rpoN*	pKH61	Cam
pDCS20	*pCC* complex	pET28b	Kan
pDCS21	*B. subtilis sfp*	pDCS11	Amp
pDCS23	*E. coli rpoN*	pDCS15	Amp
pDCS24	*E. coli pspA* promoter	pCR-Blunt	Kan
pDCS25	*E. coli lacZ*	pCR-Blunt	Kan
pDCS26	*E. coli pspA* promoter replaced T7 promoter	pET28b	Kan
pDCS27	*E. coli lacZ*	pET28b	Kan
pDCS29	*S. rimosus oxyI* promoter	pCR-Blunt	Kan
pDCS30	*S. rimosus oxyJ* promoter	pCR-Blunt	Kan
pDCS31	*E. coli lacZ*	pDCS26	Kan
pDCS33	*S. rimosus oxyJ* promoter replaced T7 promoter	pDCS27	Kan
pDCS34	*P. putida rpoN*	pCR-Blunt	Kan

pDCS35	*E. coli IHFα*	pCR-Blunt	Kan
pDCS36	*E. coli IHFβ*	pCR-Blunt	Kan
pDCS37	*E. coli pspF*	pCR-Blunt	Kan
pDCS38	*E. coli IHFα*	pET28b	Kan
pDCS39	*E. coli IHFβ*	pET28b	Kan
pDCS40	*P. putida rpoN*	pET28b	Kan
pDCS41	*P. putida rpoN*	pKH22	Amp
pDCS42	*E. coli pspF*	pKH22	Amp
pDCS43	*E. coli pspF*	pDCS11	Amp
pDCS44	*E. coli pspF*	pDCS12	Amp
pDCS45	*E. coli pspF*	pDCS41	Amp
pDCS46	*E. coli pspFΔHTH*	pCR-Blunt	Kan
pDCS47	*E. coli pspFΔHTH*	pET28b	Kan
pDCS48	*E. coli pspFΔHTH*	pKH22	Amp
pDCS49	*E. coli pspFΔHTH*	pDCS11	Amp
pDCS50	*E. coli pspFΔHTH*	pDCS12	Amp
pDCS51	*E. coli pspFΔHTH*	pDCS41	Amp
pDCS53	*E. coli rpoF*	pCR-Blunt	Kan
pDCS54	*E. coli rpoE*	pCR-Blunt	Kan
pDCS55	*E. coli fecI*	pCR-Blunt	Kan
pDCS56	*E. coli rpoD*	pCR-Blunt	Kan
pDCS57	*E. coli rpoF*	pKH22	Amp
pDCS58	*E. coli rpoE*	pKH22	Amp
pDCS59	*E. coli fecI*	pKH22	Amp
pDCS60	*E. coli rpoD*	pKH22	Amp
pDCS61	*S. rimosus oxyTA1ABCD* operon	pET28b	Kan
pDCS62	Promoter mutant of *oxyA* promoter	pDCS61	Kan

A.2 Vector maps

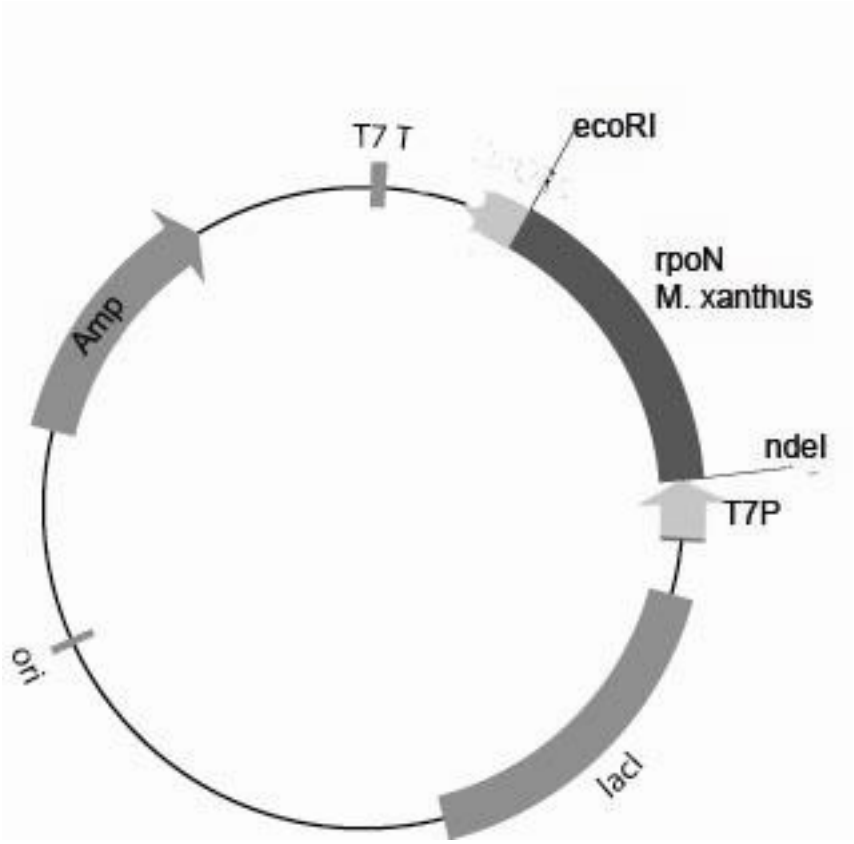


Vector Name: DCS01

Location: Notebook I

Construction: Blunt PCR insert

Note: n/a

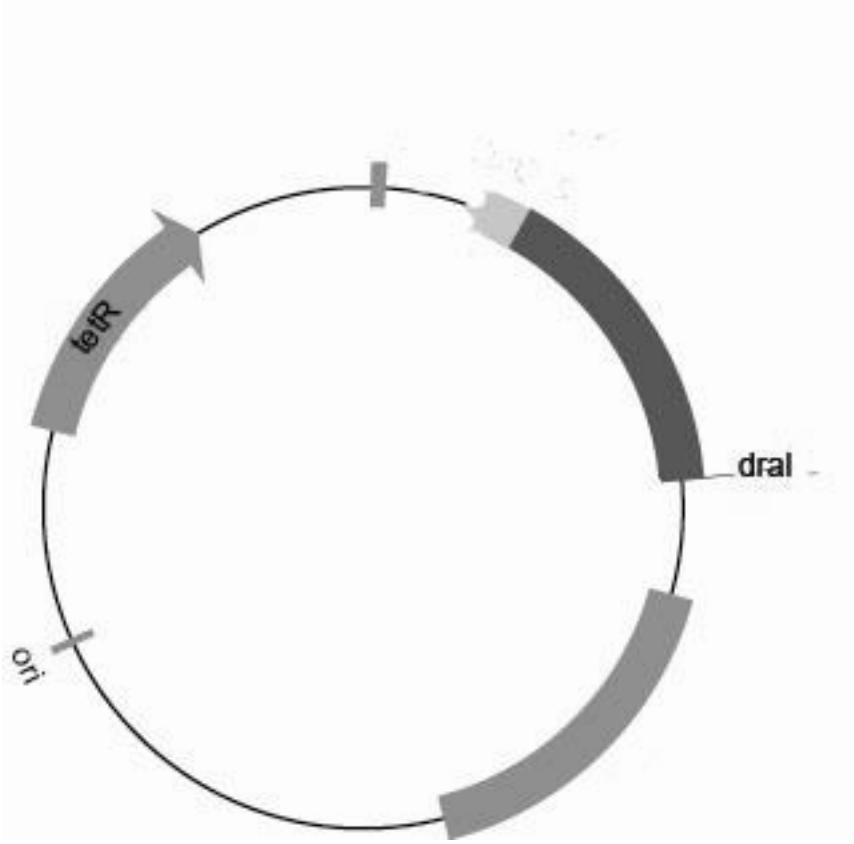


Vector Name: DCS02

Location: Notebook I

Construction: NdeI/EcoRI insertion from pDCS01 into pET21c

Note: n/a

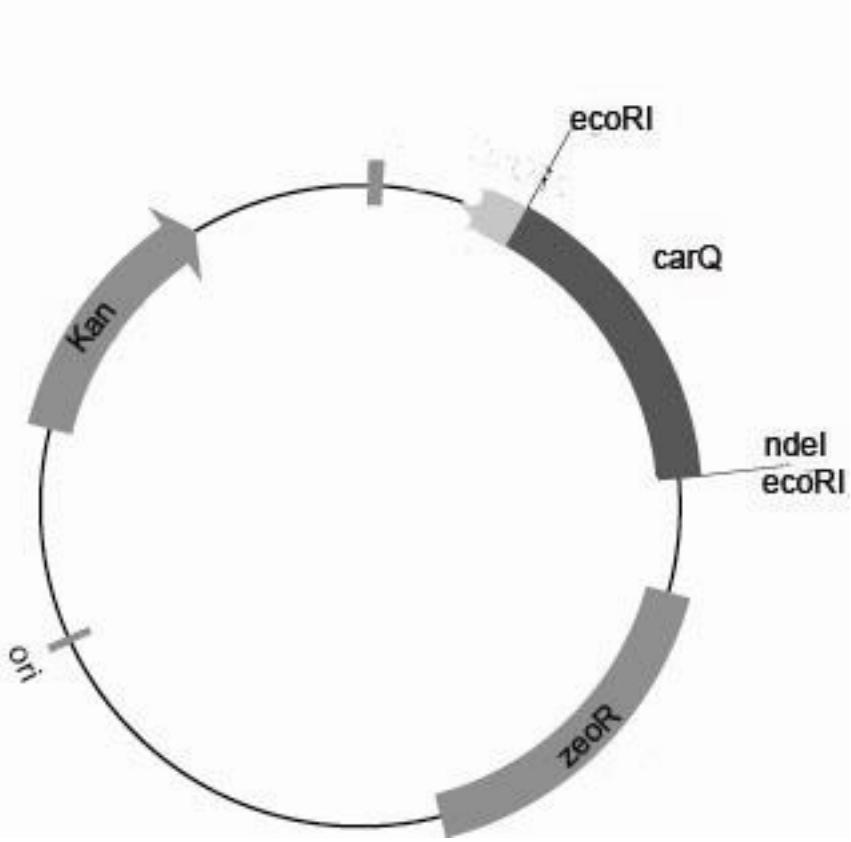


Vector Name: DCS03

Location: Notebook I

Construction: DraI removal of AmpR

Note: Needed TetR on vector without AmpR

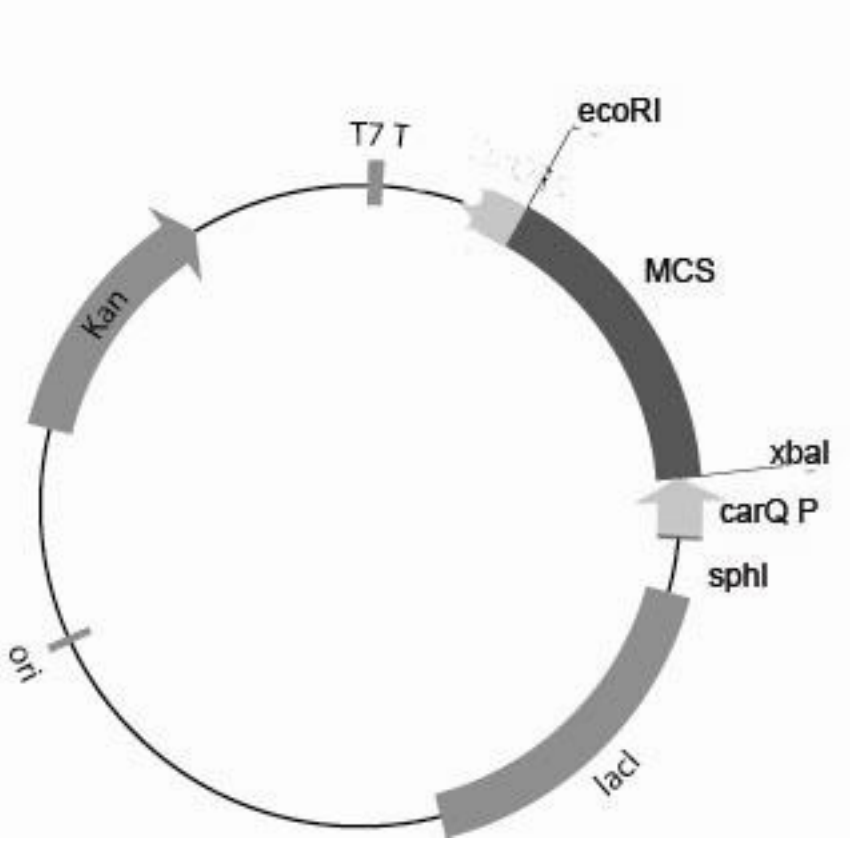


Vector Name: DCS04

Location: Notebook I

Construction: Blunt PCR insert

Note: n/a

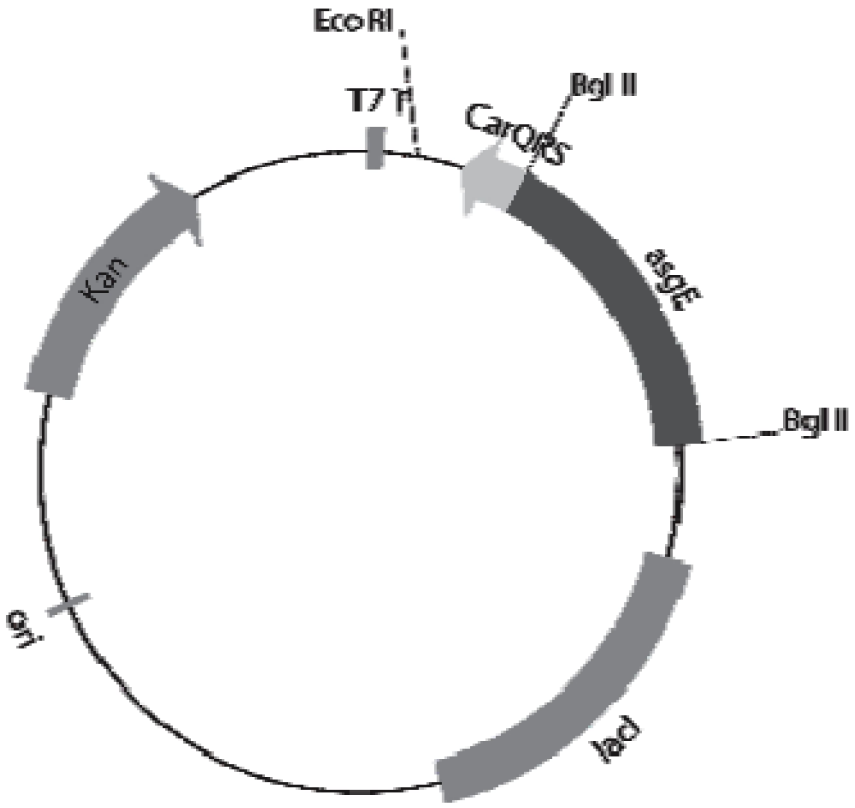


Vector Name: DCS05

Location: Notebook I

Construction: BglII/XbaI insertion of CarQ replacing T7P in pET28b

Note: n/a

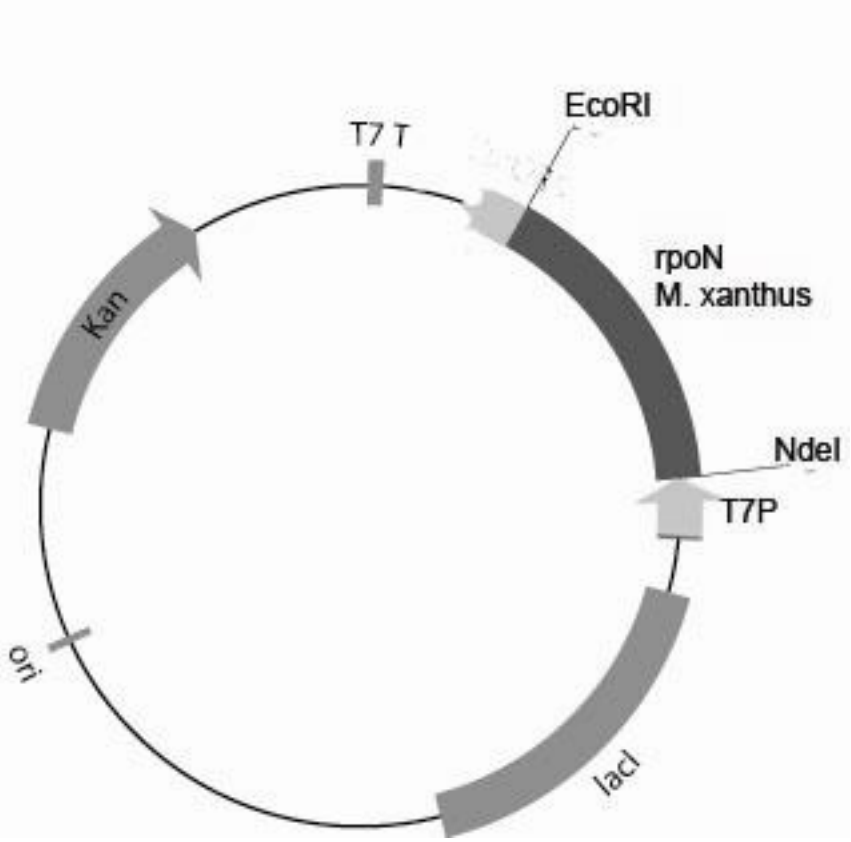


Vector Name: DCS06

Location: Notebook I

Construction: BglII insertion of AsgE from *M. xanthus* into DCS05

Note: Built for overexpression insertion into *M. xanthus,* inducible by light

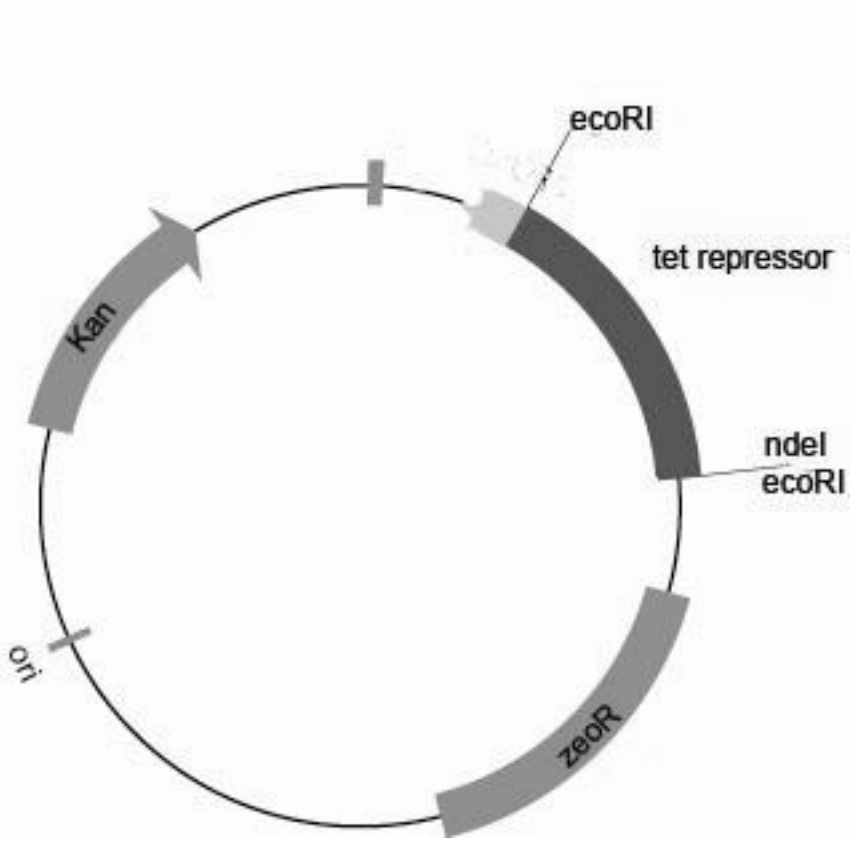


Vector Name: DCS07

Location: Notebook I

Construction: NdeI/EcoRI insertion of rpoN into pET28b

Note: Reading frame appropriate for protein expression

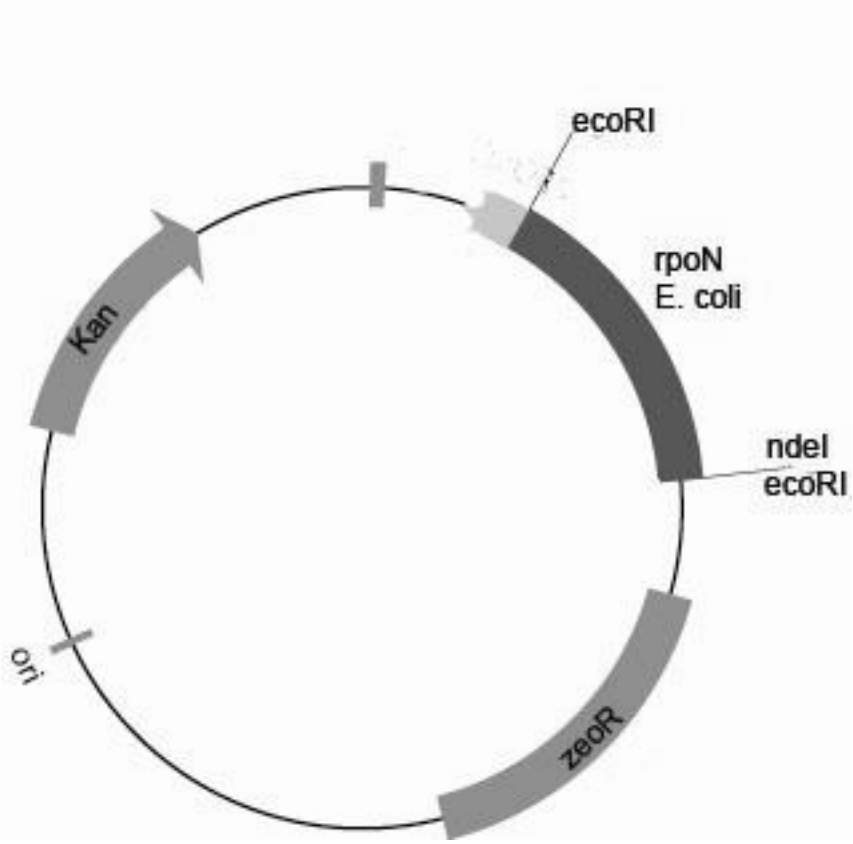


Vector Name: DCS09

Location: Notebook I

Construction: Blunt PCR insert

Note: Oxytetracycline purification attempt

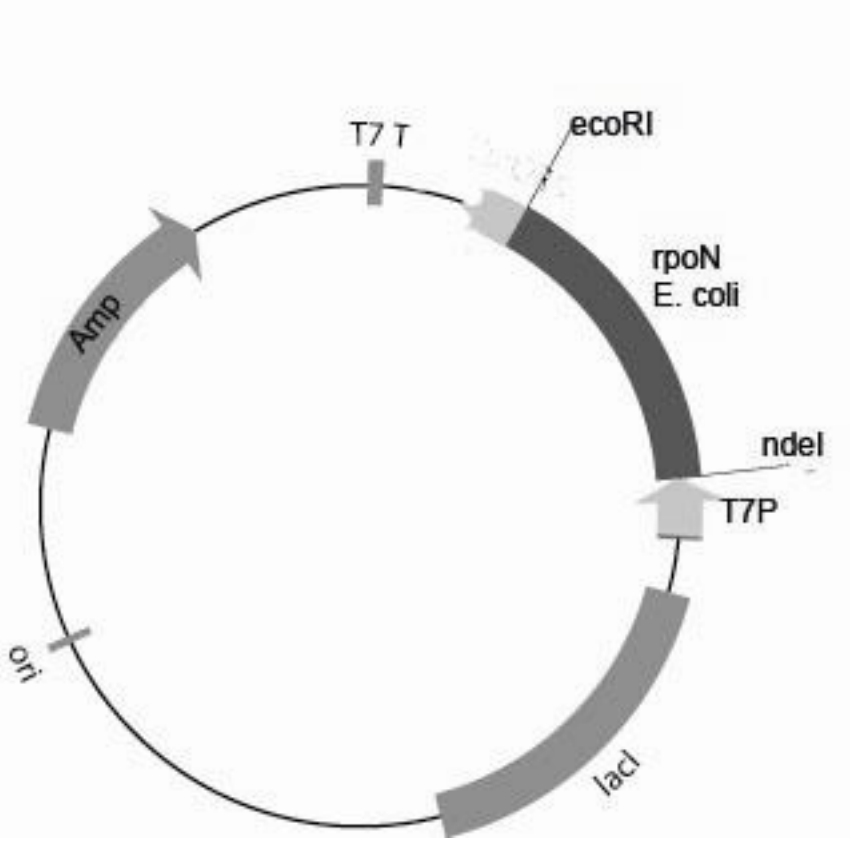


Vector Name: DCS10

Location: Notebook I

Construction: Blunt PCR Insert

Note: n/a

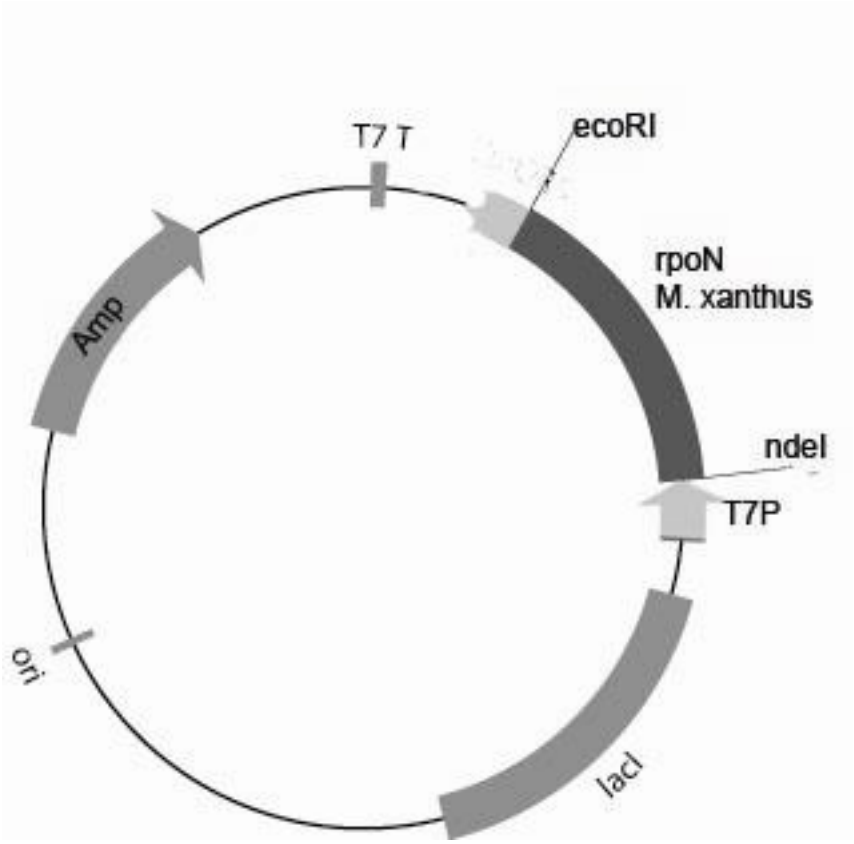


Vector Name: DCS11

Location: Notebook II

Construction: NdeI/EcoRI insertion

Note: Vector for all published oxytetracycline production experiments

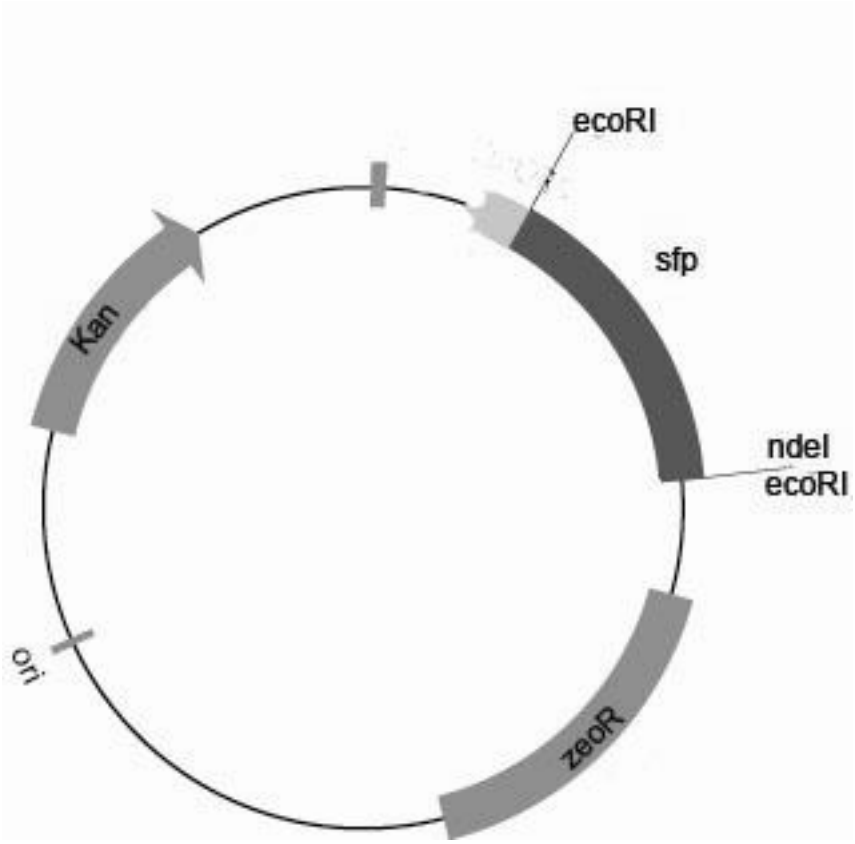


Vector Name: DCS12

Location: Notebook II

Construction: NdeI/EcoRI insertion

Note: n/a

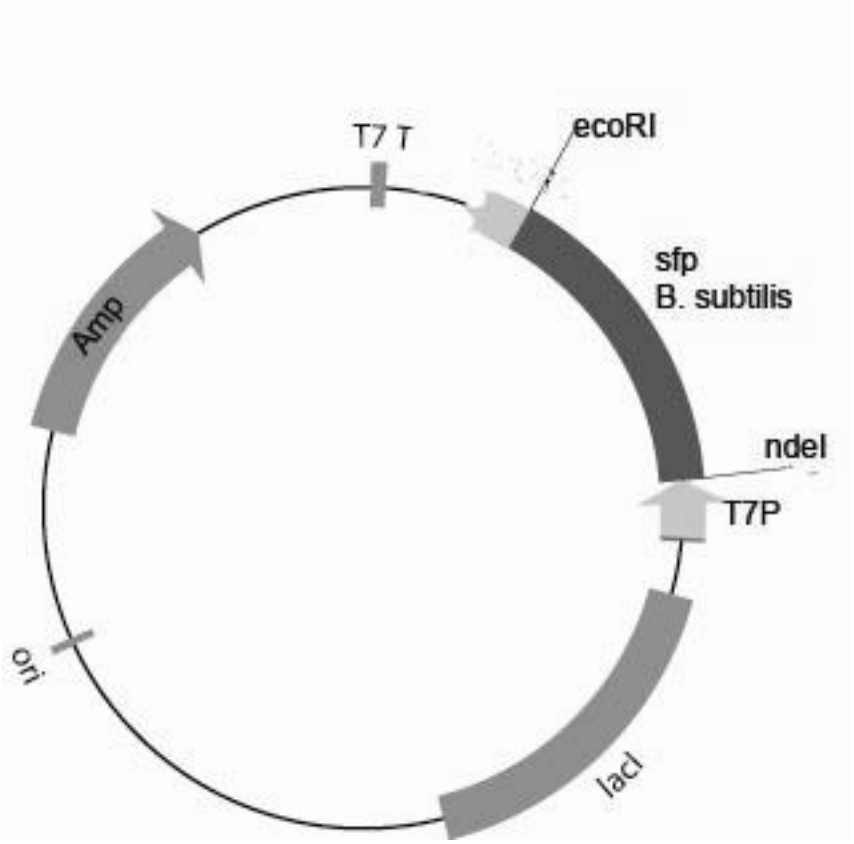


Vector Name: DCS13

Location: Notebook II

Construction: Blunt PCR insert

Note: *sfp* from *B. subtilis* but amplified from *E. coli* BAP1 gDNA

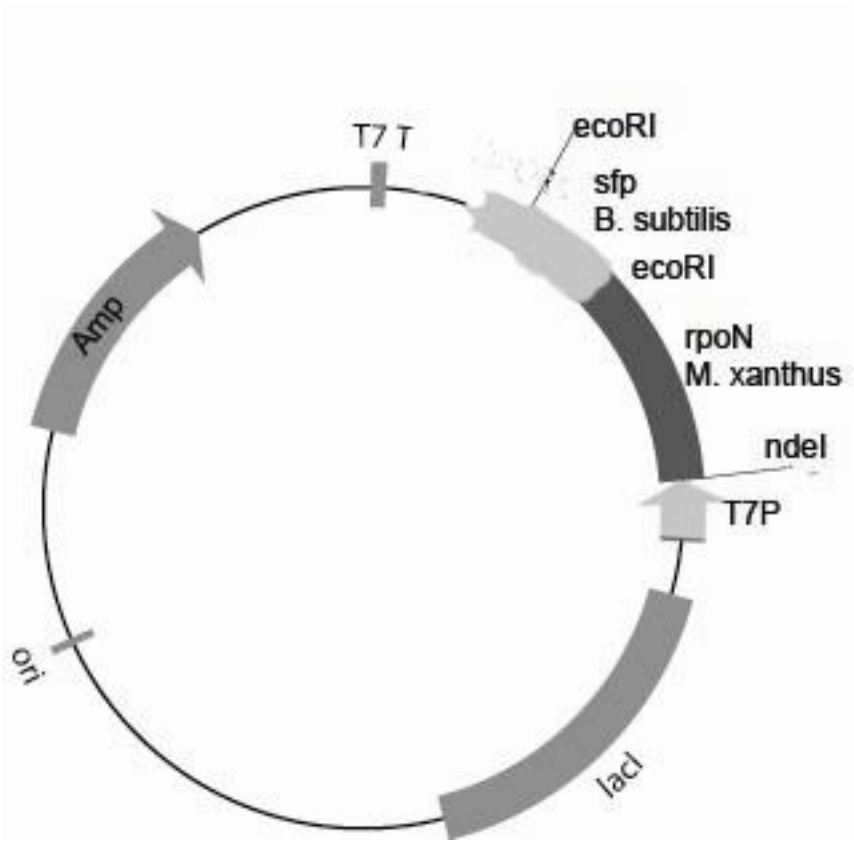


Vector Name: DCS14

Location: Notebook II

Construction: NdeI/EcoRI insertion

Note: n/a

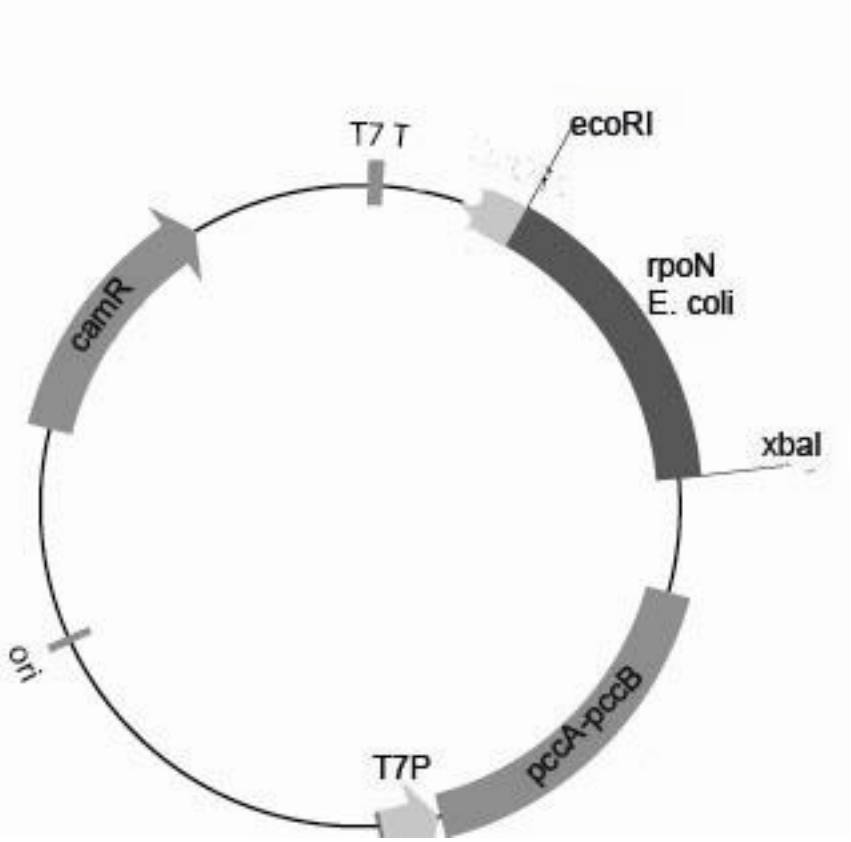


Vector Name: DCS15

Location: Notebook II

Construction: AvrII/XbaI Insertion

Note: n/a

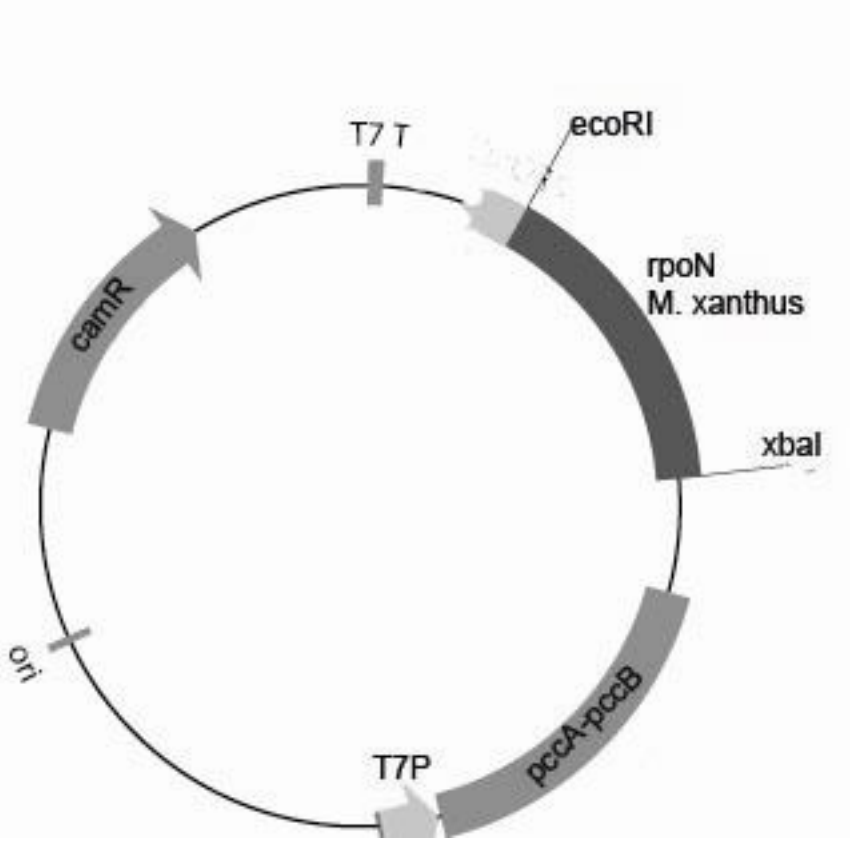


Vector Name: DCS18

Location: Notebook II

Construction: XbaI/EcoRI insertion into pKH61

Note: n/a

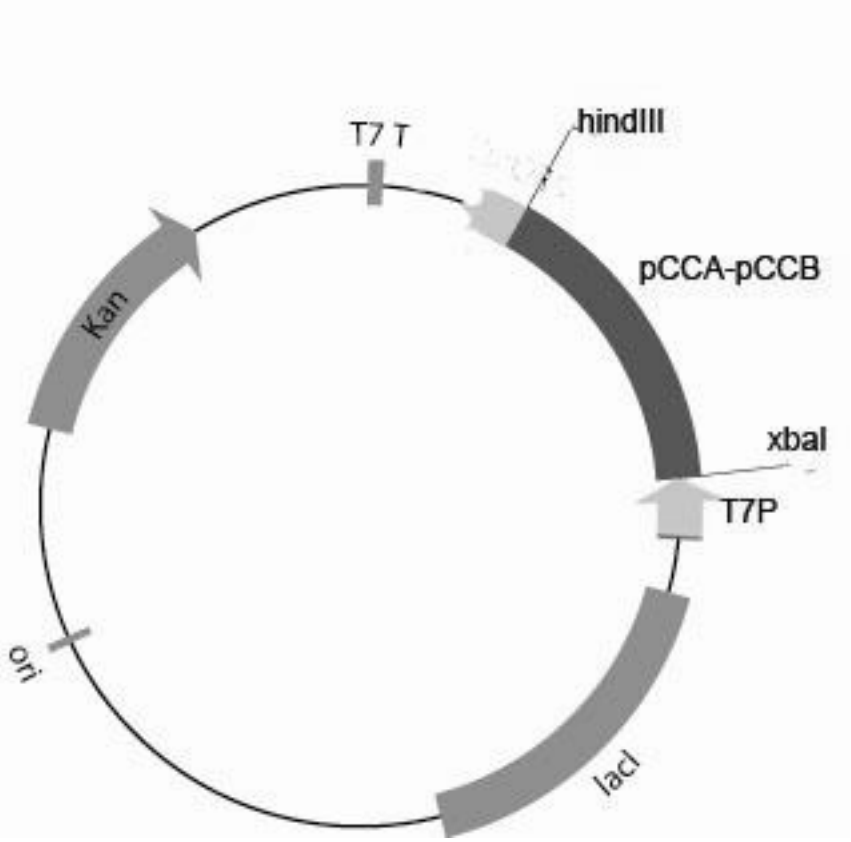


Vector Name: DCS19

Location: Notebook II

Construction: XbaI/EcoRI insertion into pKH61

Note: n/a

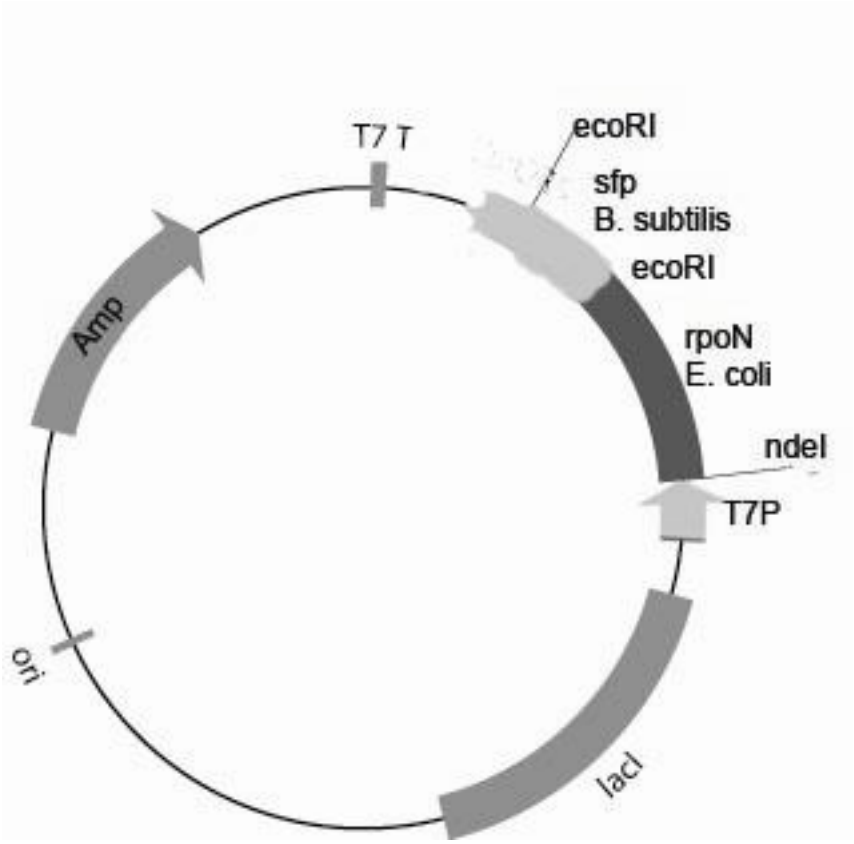


Vector Name: DCS20

Location: Notebook II

Construction: XbaI/HindIII insertion of pCC complex from pKH61

Note: n/a

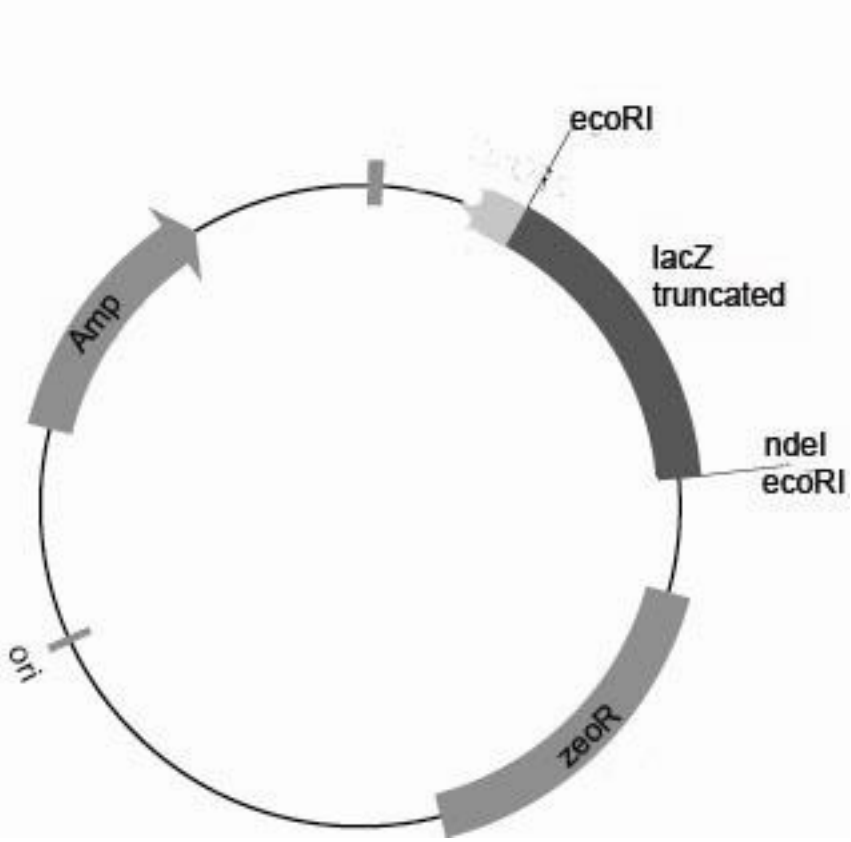


Vector Name: DCS21

Location: Notebook II

Construction: AvrII/XbaI insertion

Note: n/a

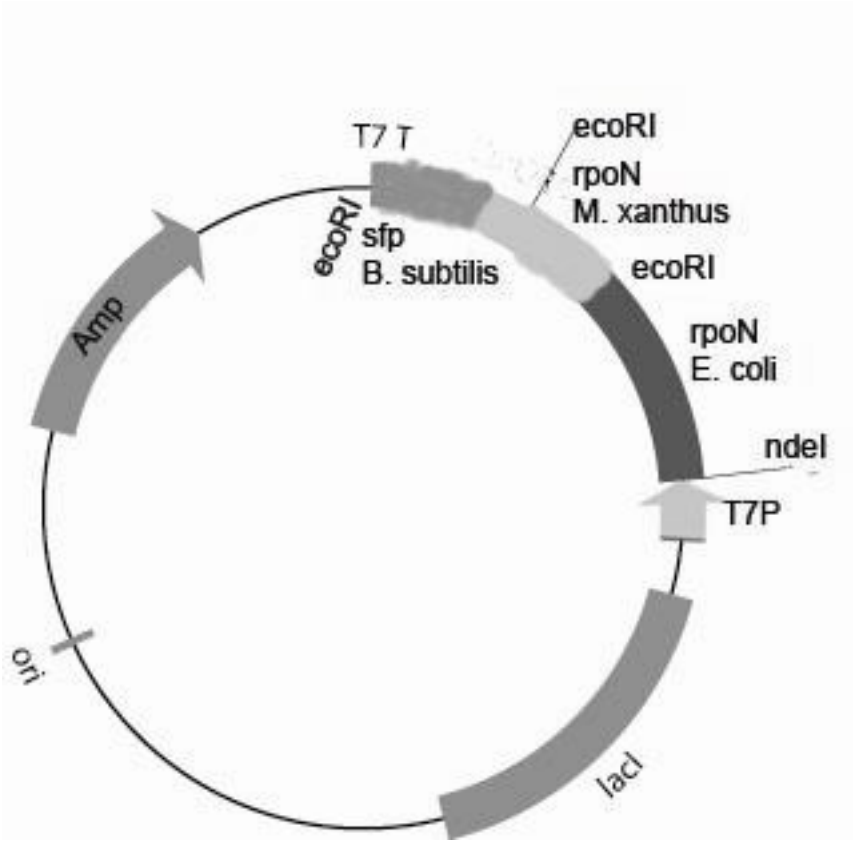


Vector Name: DCS22

Location: Notebook II

Construction: Blunt PCR insert

Note: Truncated lacZ

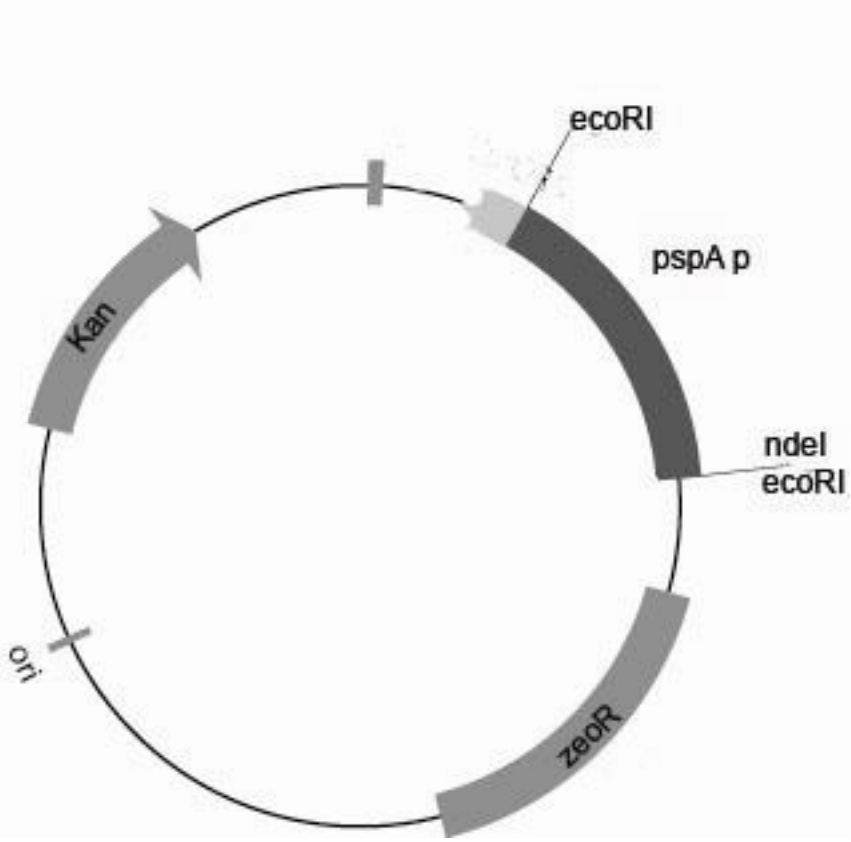


Vector Name: DCS23

Location: Notebook II

Construction: AvrII/XbaI insertion

Note: n/a

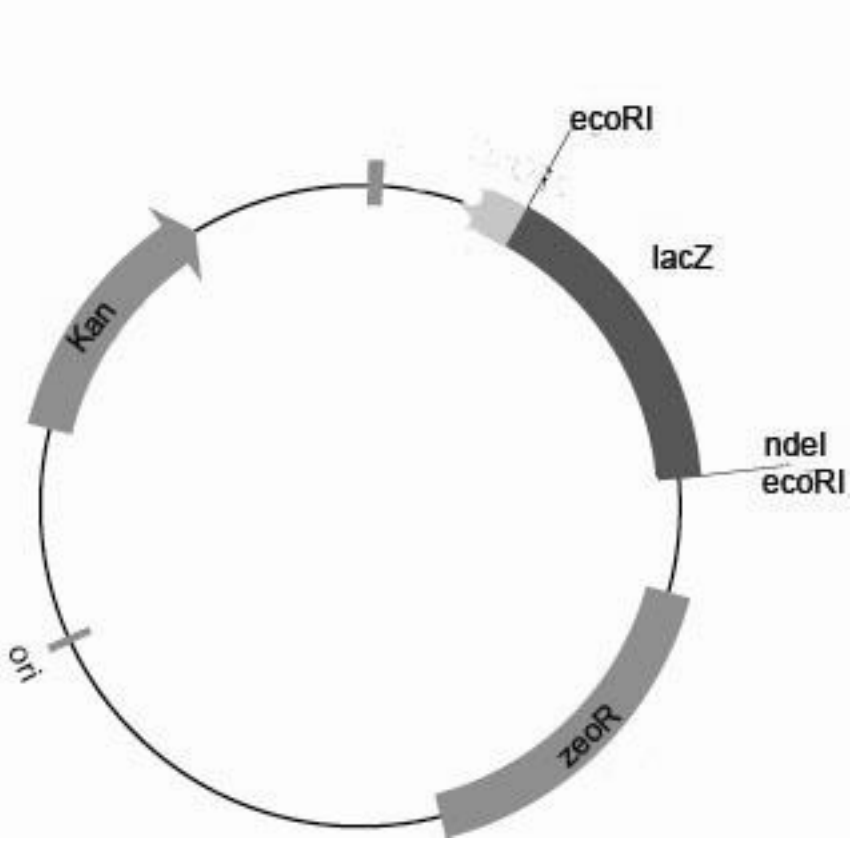


Vector Name: DCS24

Location: Notebook II

Construction: Blunt PCR insertion

Note: n/a

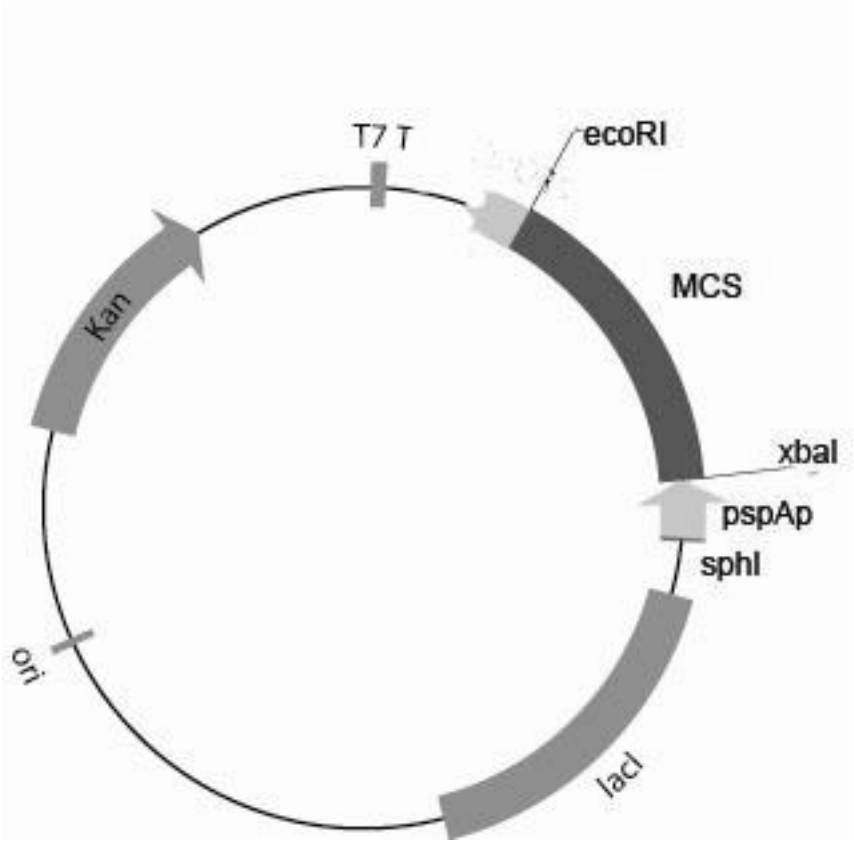


Vector Name: DCS25

Location: Notebook II

Construction: Blunt PCR insertion

Note: full length lacZ

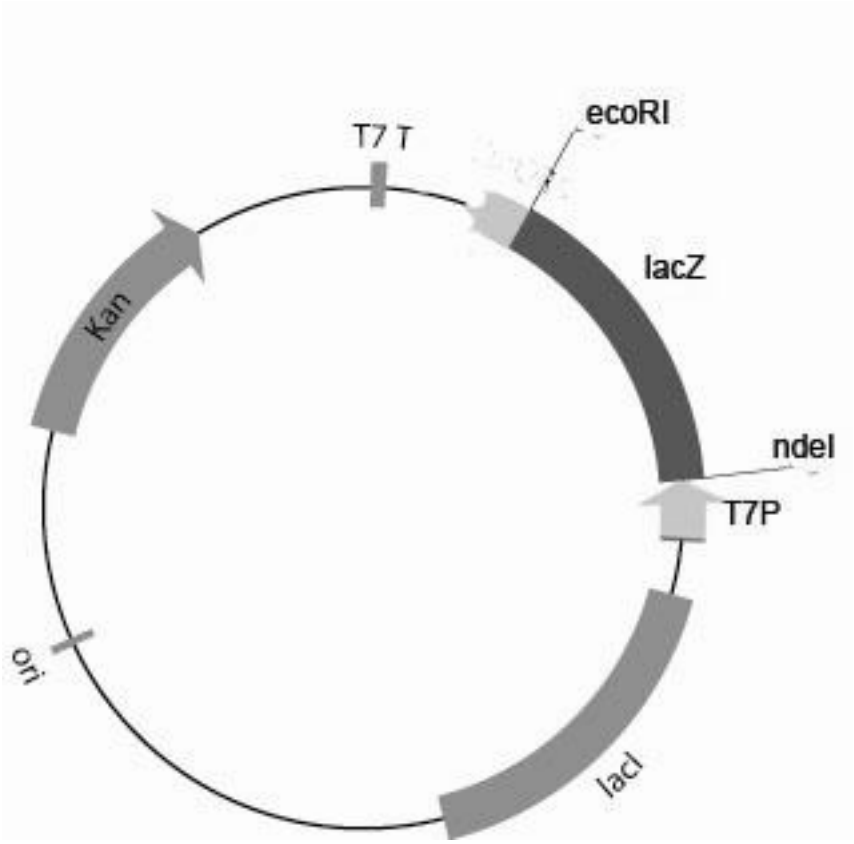


Vector Name: DCS26

Location: Notebook II

Construction: SphI/XbaI insertion

Note: replaced T7 promoter

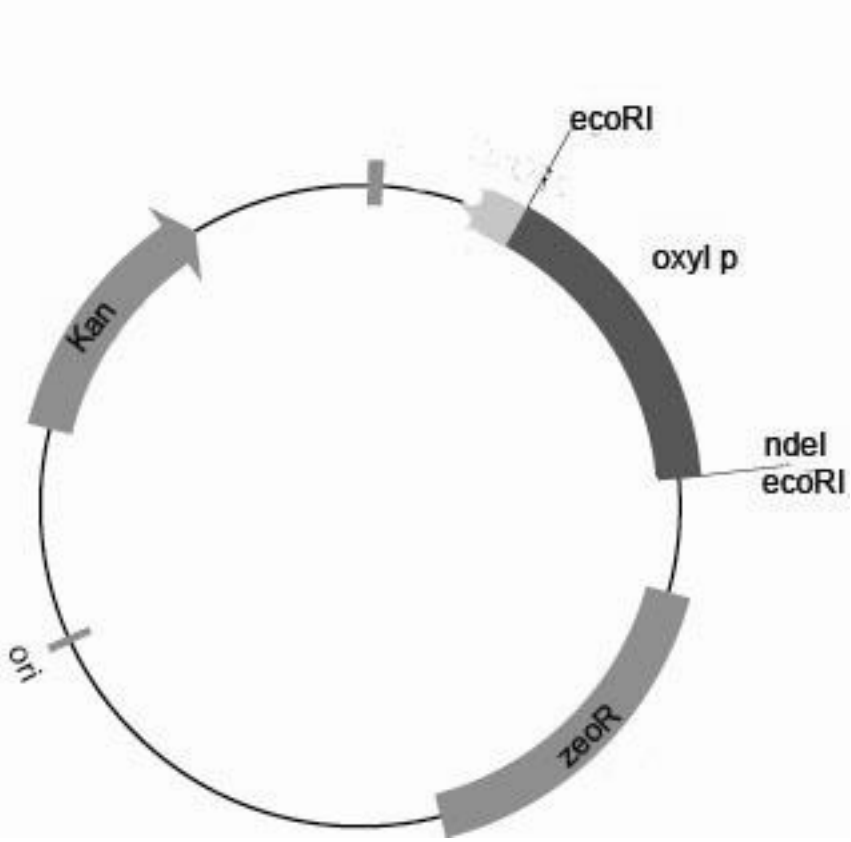


Vector Name: DCS27

Location: Notebook II

Construction: NdeI/EcoRI insertion

Note: n/a

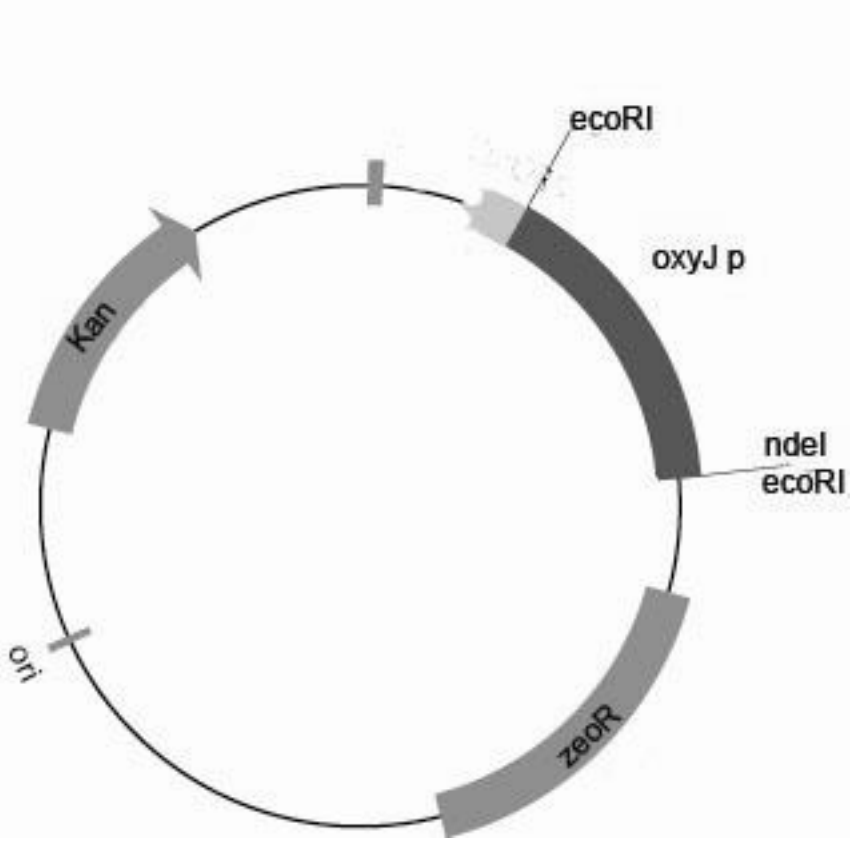


Vector Name: DCS29

Location: Notebook II

Construction: Blunt PCR insert

Note: n/a

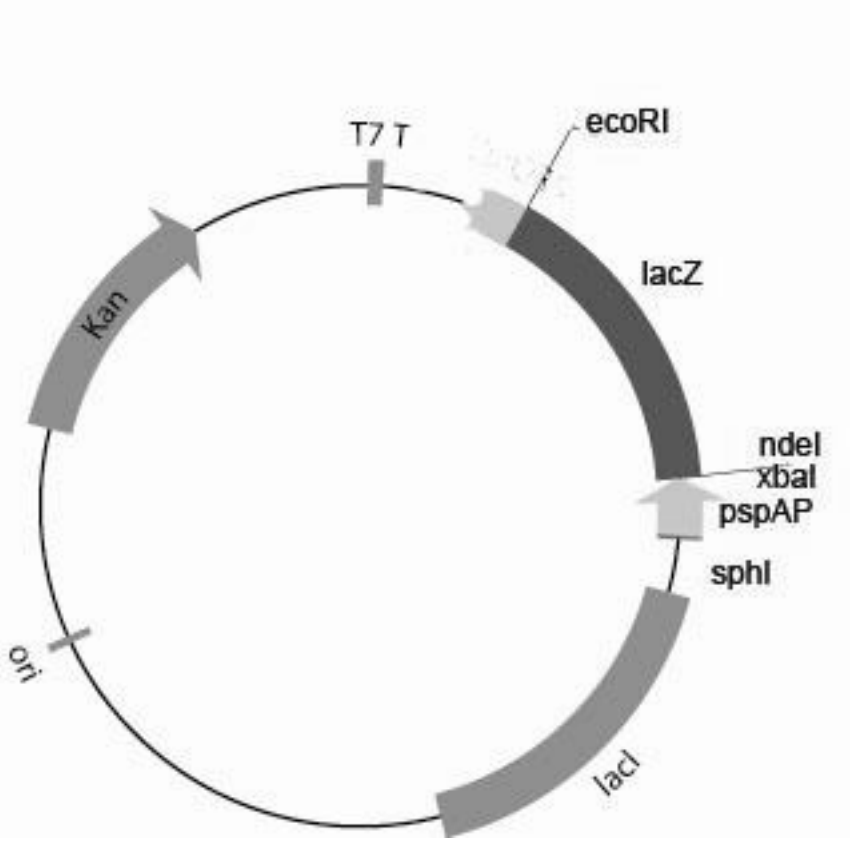


Vector Name: DCS30

Location: Notebook II

Construction: Blunt PCR insert

Note: n/a

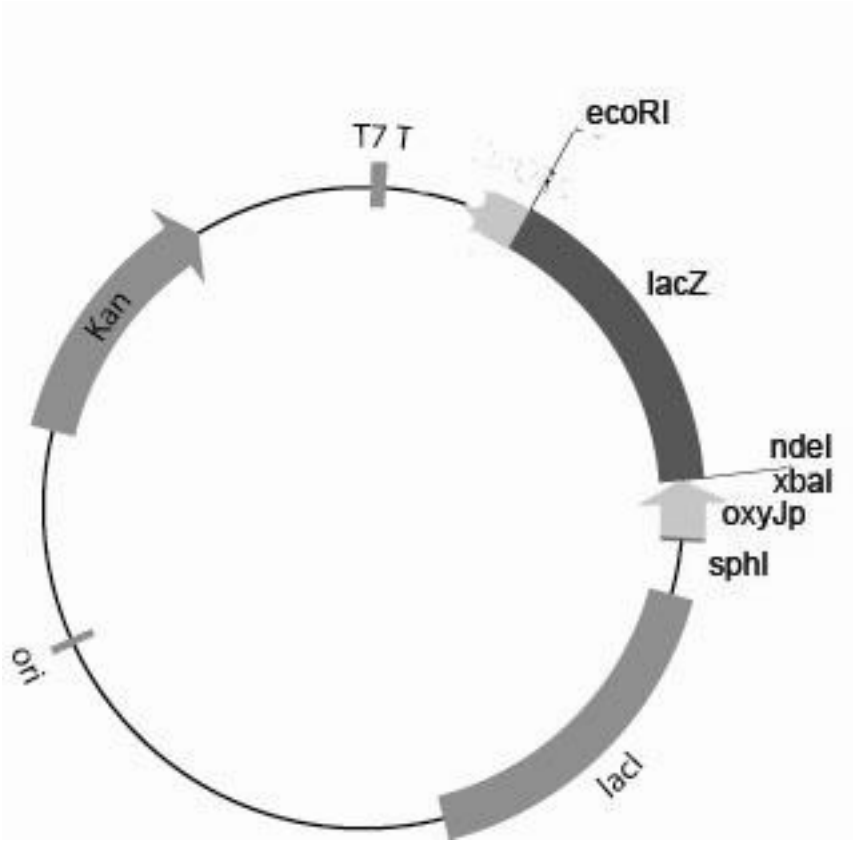


Vector Name: DCS31

Location: Notebook II

Construction: NdeI/EcoRI insertion

Note: n/a

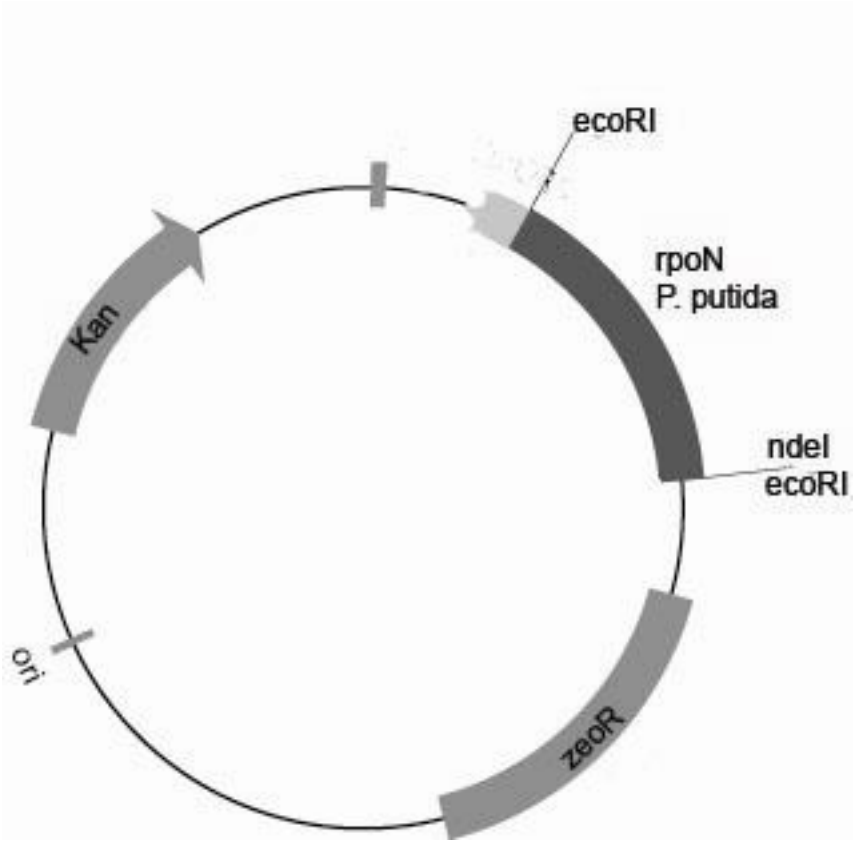


Vector Name: DCS33

Location: Notebook II

Construction: NdeI/EcoRI insertion

Note: n/a

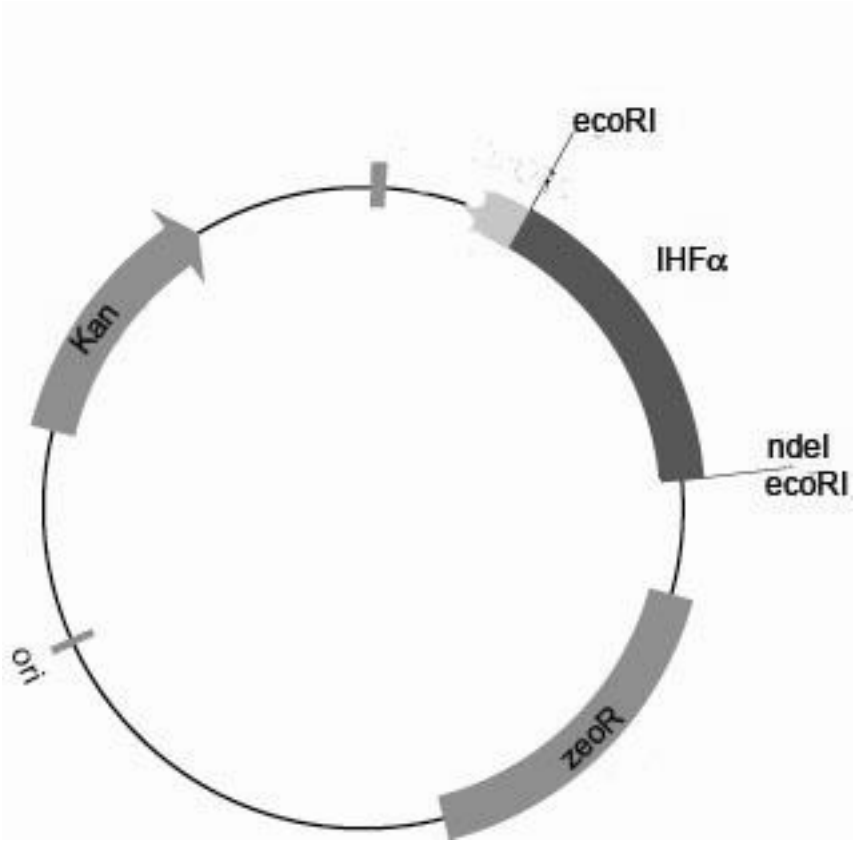


Vector Name: DCS34

Location: Notebook II

Construction: Blunt PCR insertion

Note: n/a

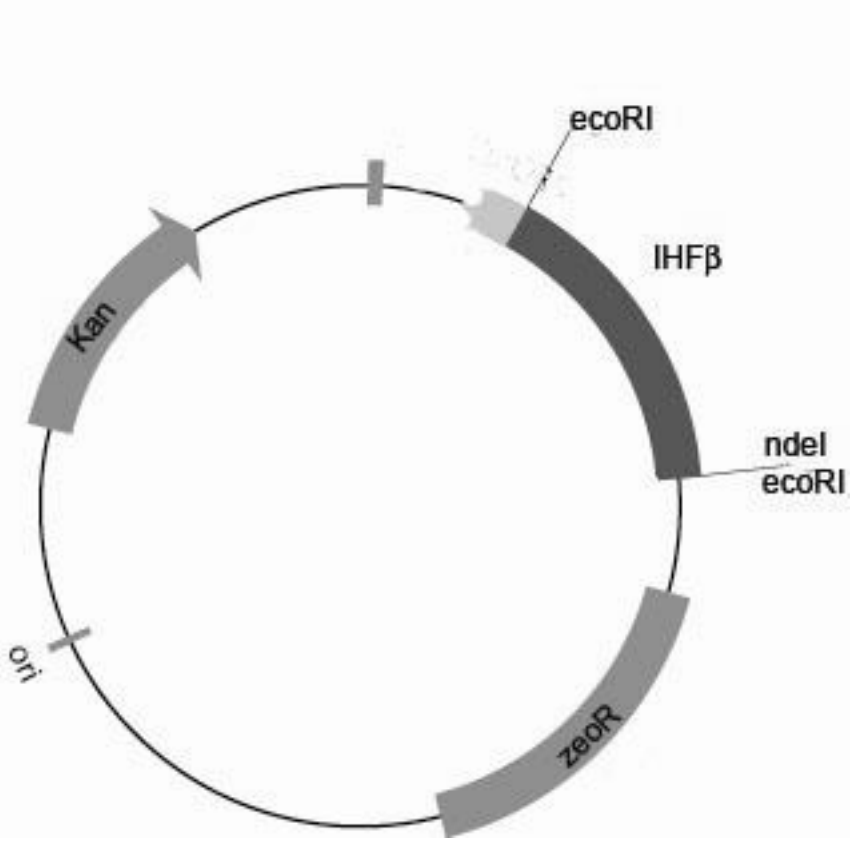


Vector Name: DCS35

Location: Notebook II

Construction: Blunt PCR insert

Note: n/a

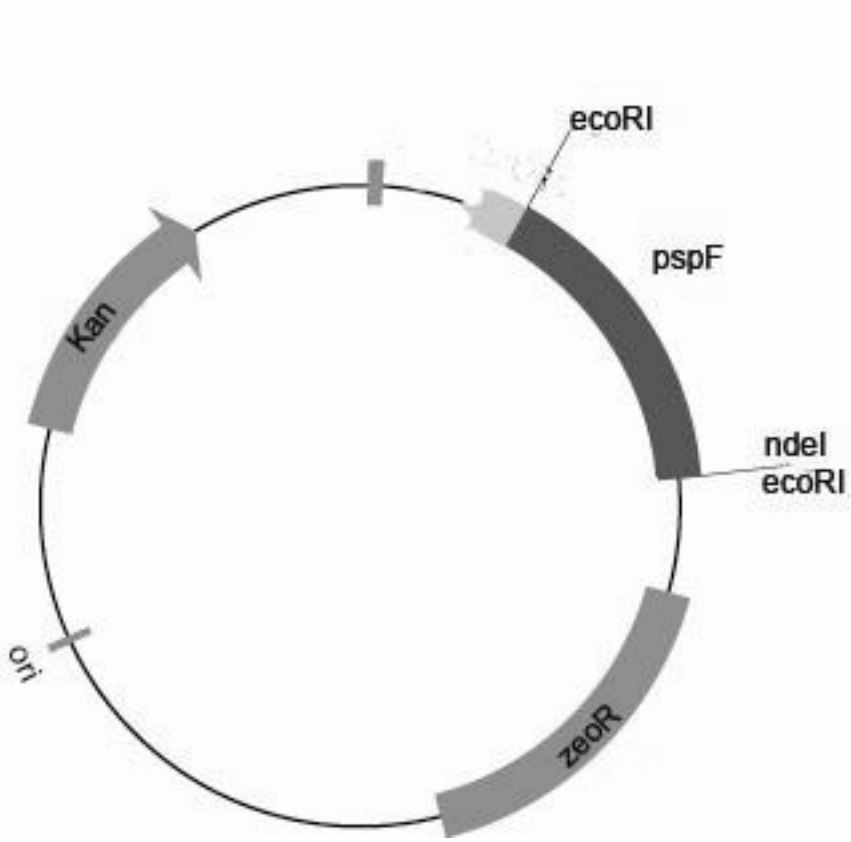


Vector Name: DCS36

Location: Notebook II

Construction: Blunt PCR insert

Note: n/a

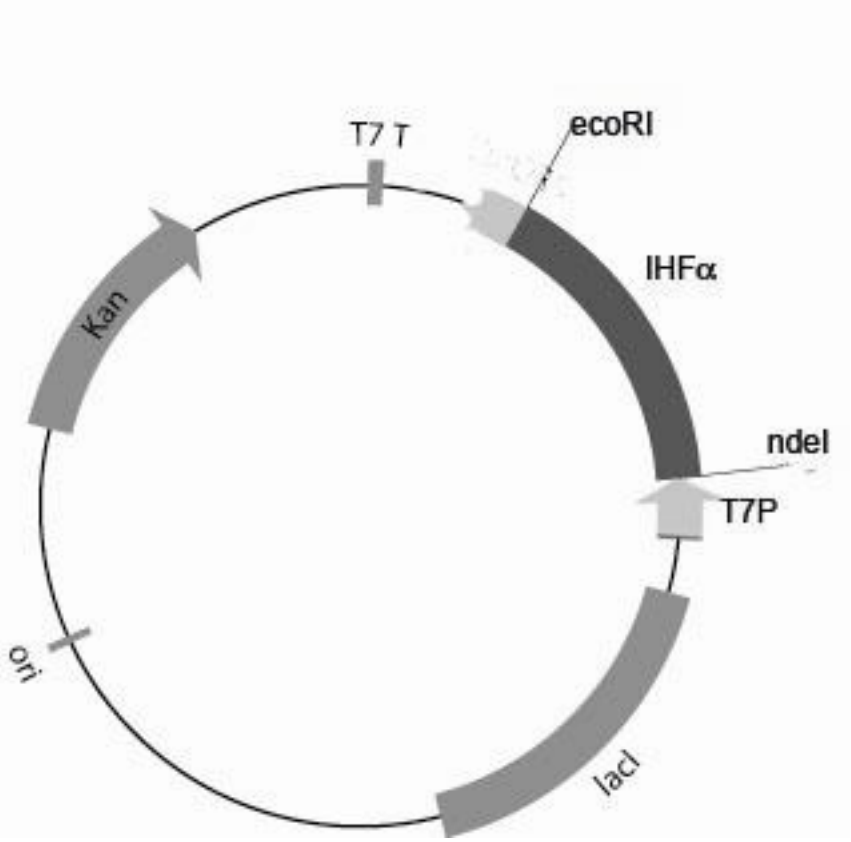


Vector Name: DCS37

Location: Notebook II

Construction: Blunt PCR insert

Note: n/a

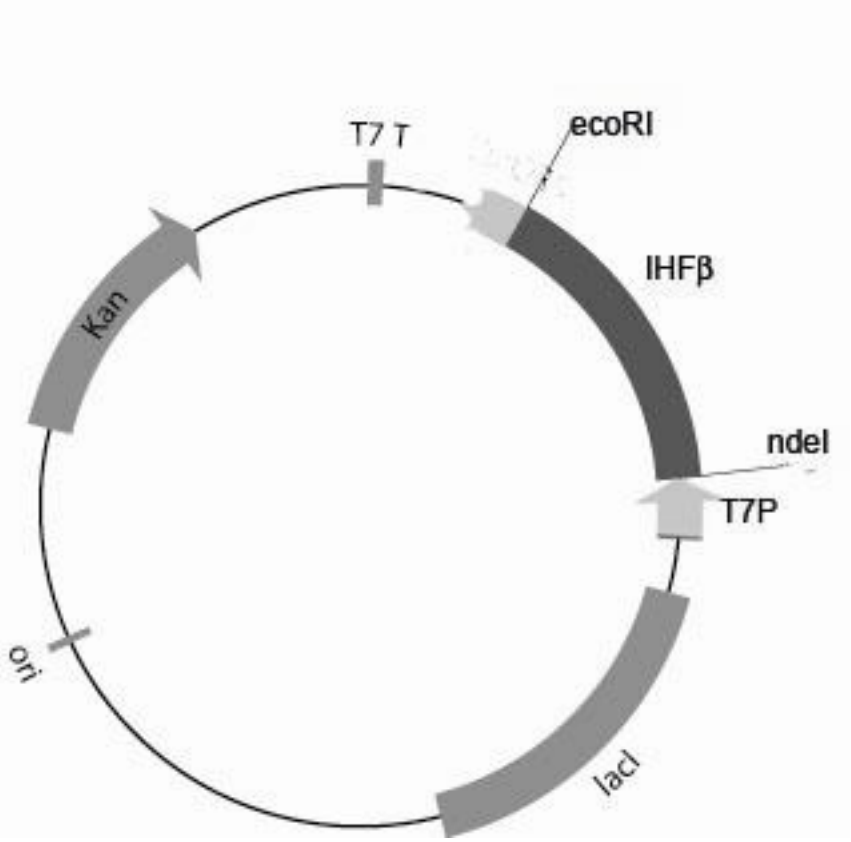


Vector Name: DCS38

Location: Notebook II

Construction: NdeI/EcoRI insertion

Note: Appropriate for protein isolation

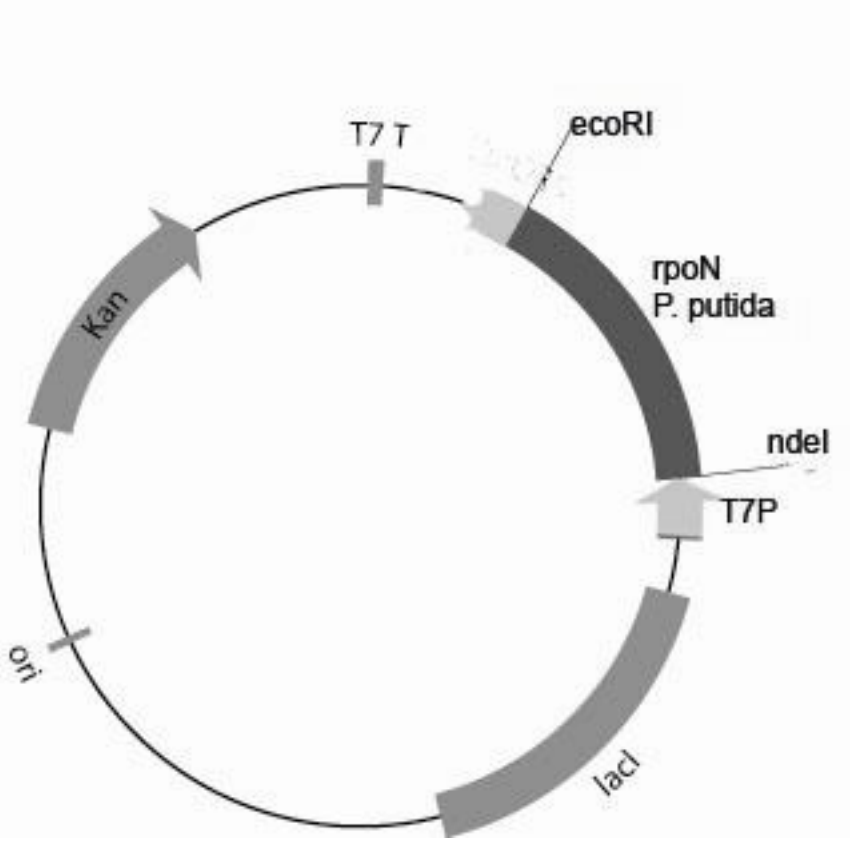


Vector Name: DCS39

Location: Notebook II

Construction: NdeI/EcoRI insertion

Note: Appropriate for protein isolation

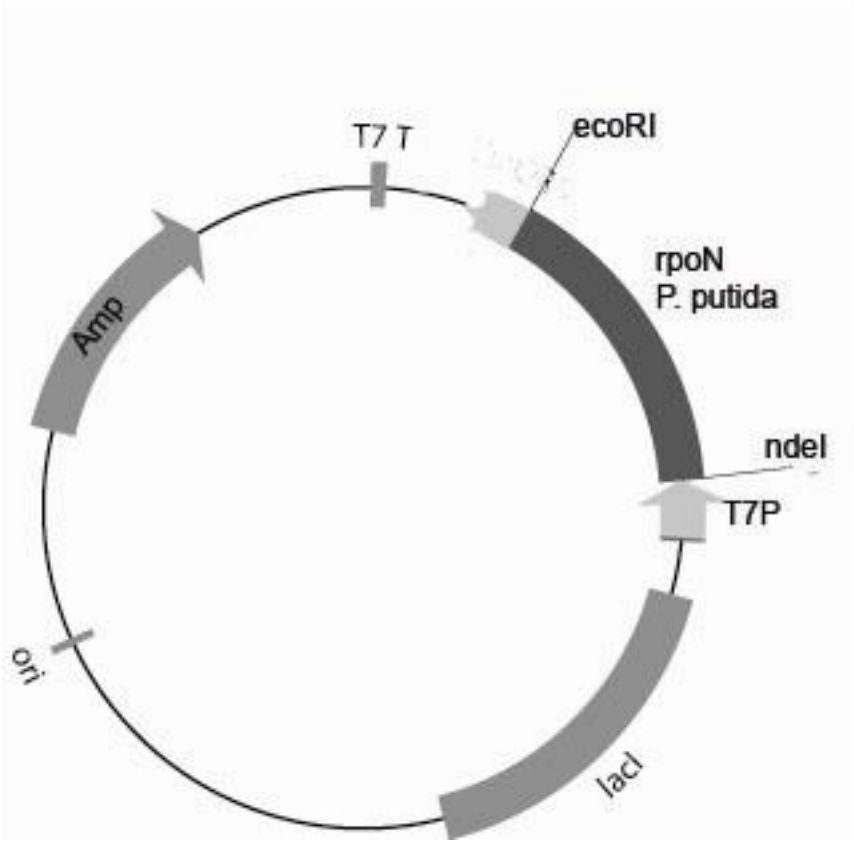


Vector Name: DCS40

Location: Notebook II

Construction: NdeI/EcoRI insertion

Note: Appropriate for protein isolation

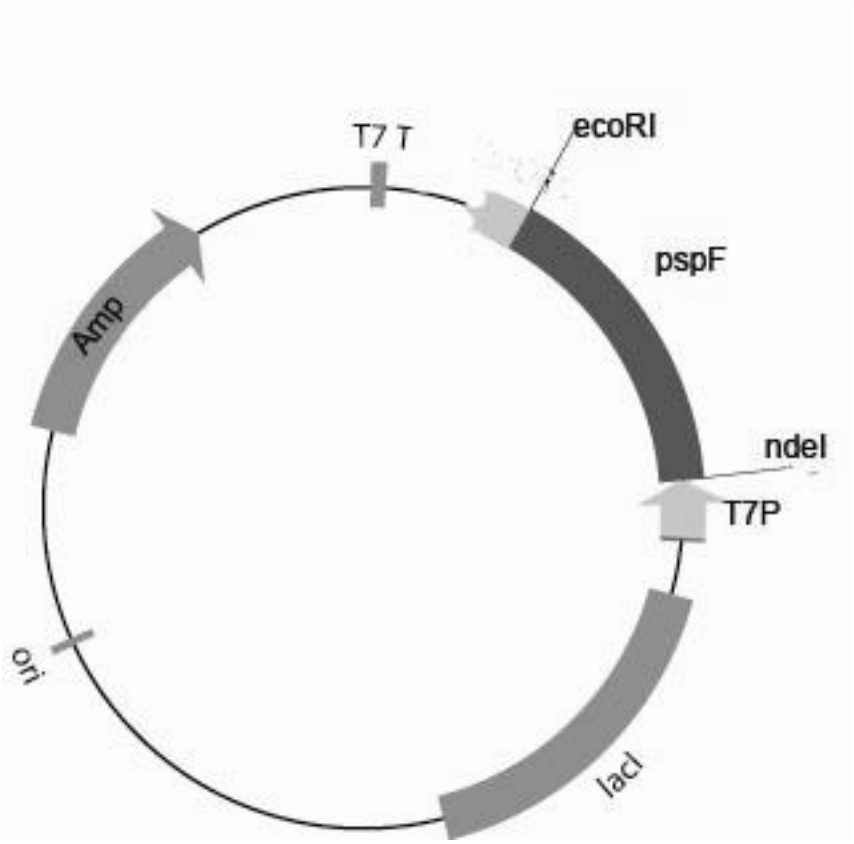


Vector Name: DCS41

Location: Notebook II

Construction: NdeI/EcoRI insertion

Note: n/a

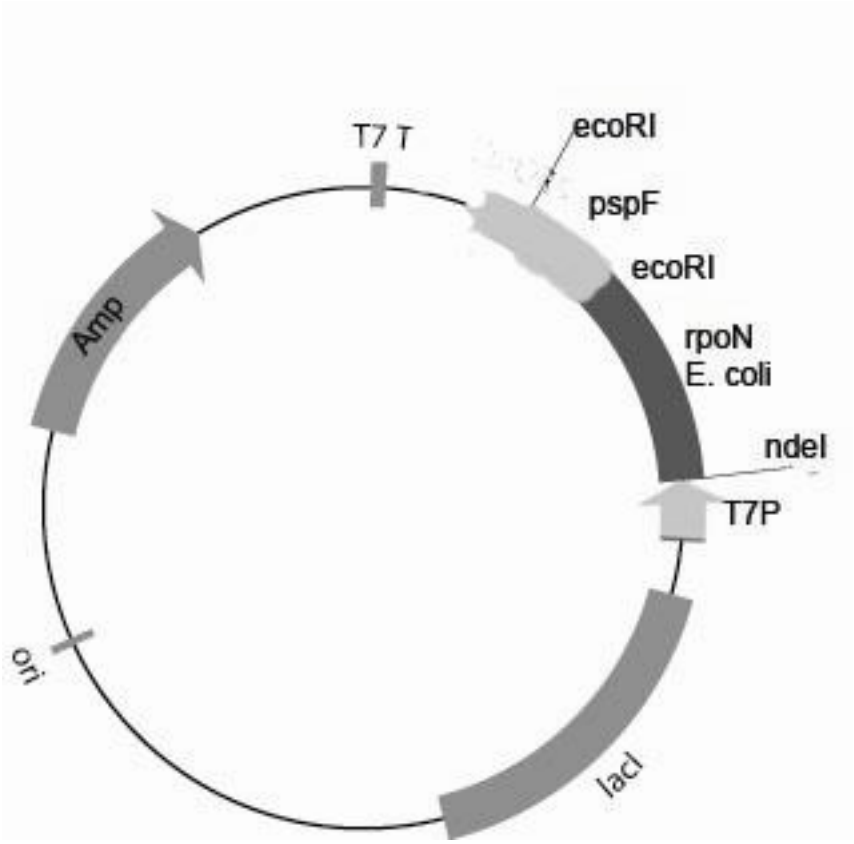


Vector Name: DCS42

Location: Notebook II

Construction: NdeI/EcoRI insertion

Note: n/a

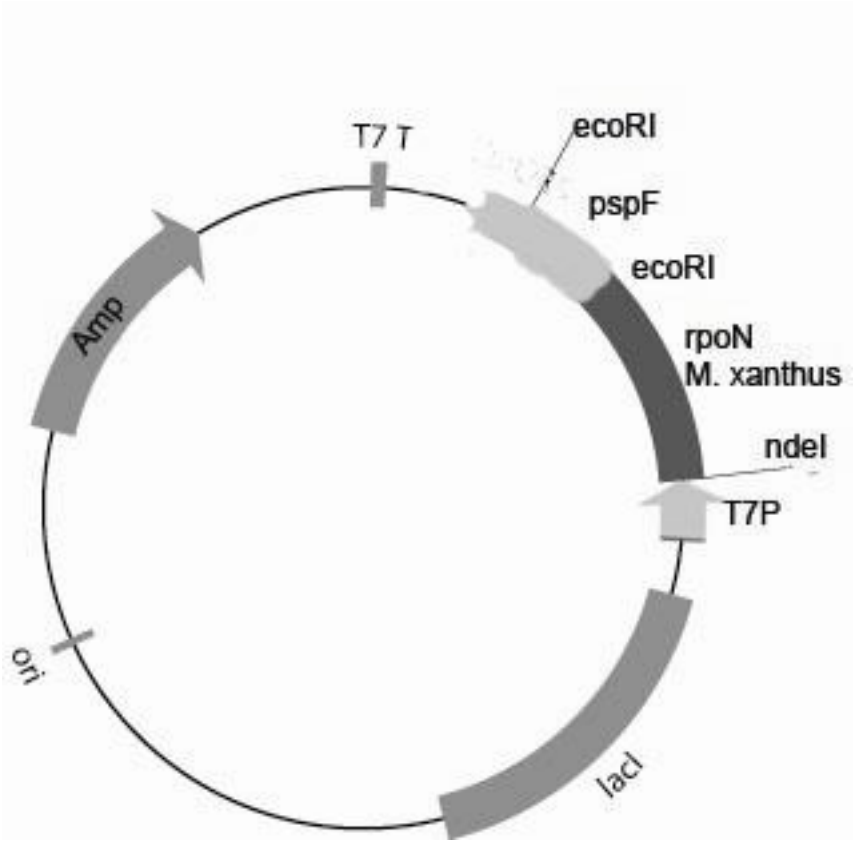


Vector Name: DCS43

Location: Notebook II

Construction: AvrII/XbaI insertion

Note: n/a

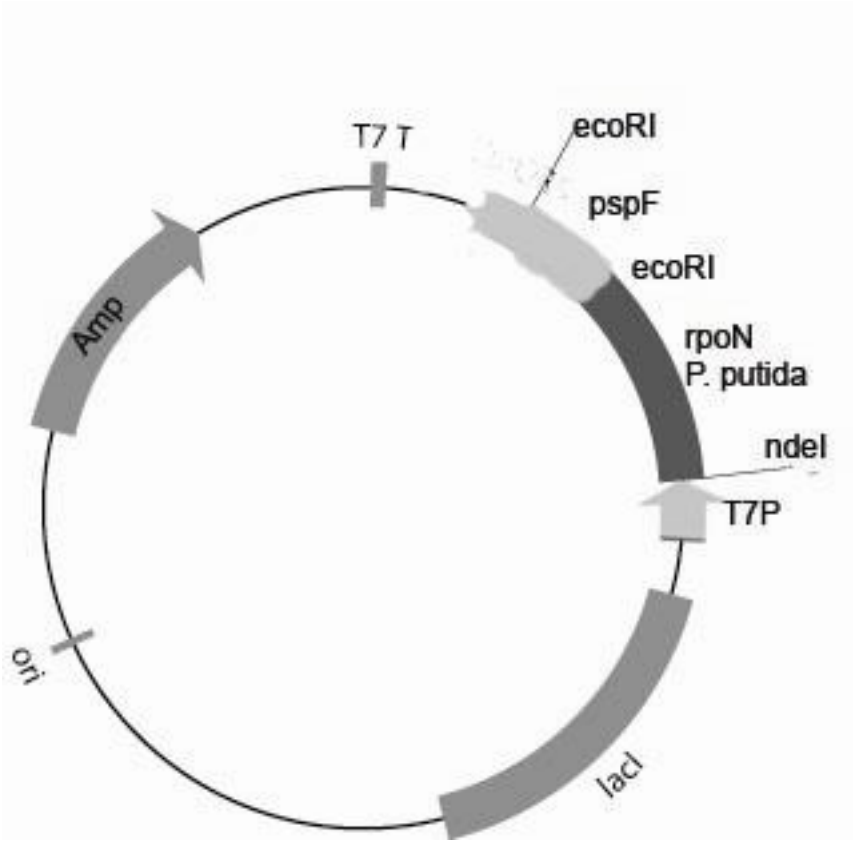


Vector Name: DCS44

Location: Notebook II

Construction: AvrII/XbaI insertion

Note: n/a

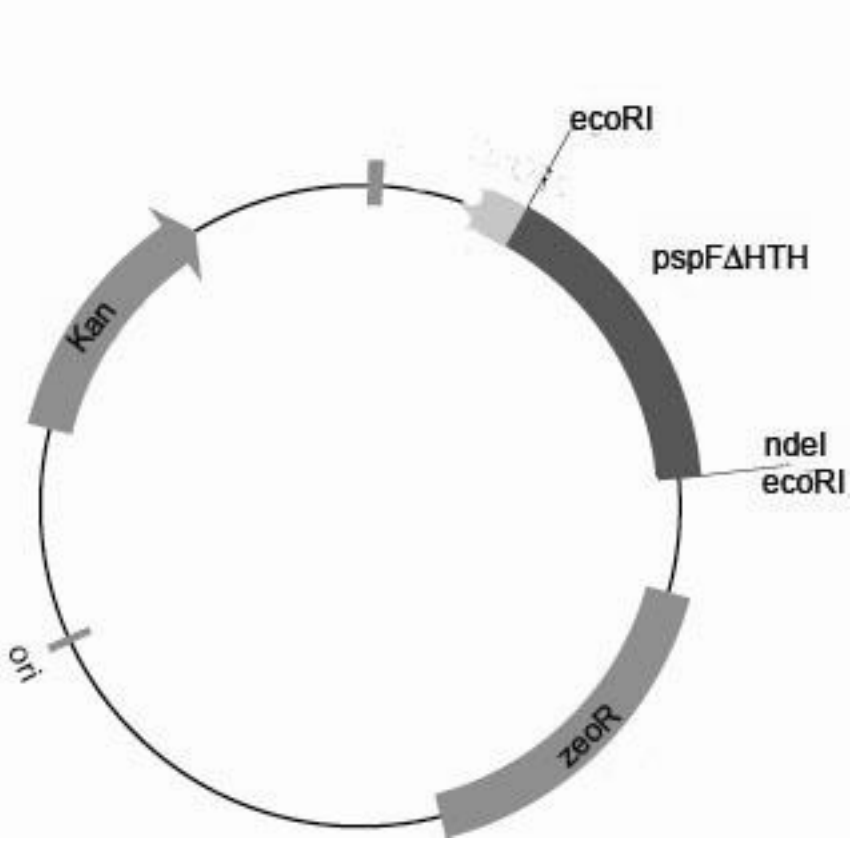


Vector Name: DCS45

Location: Notebook II

Construction: AvrII/XbaI insertion

Note: n/a

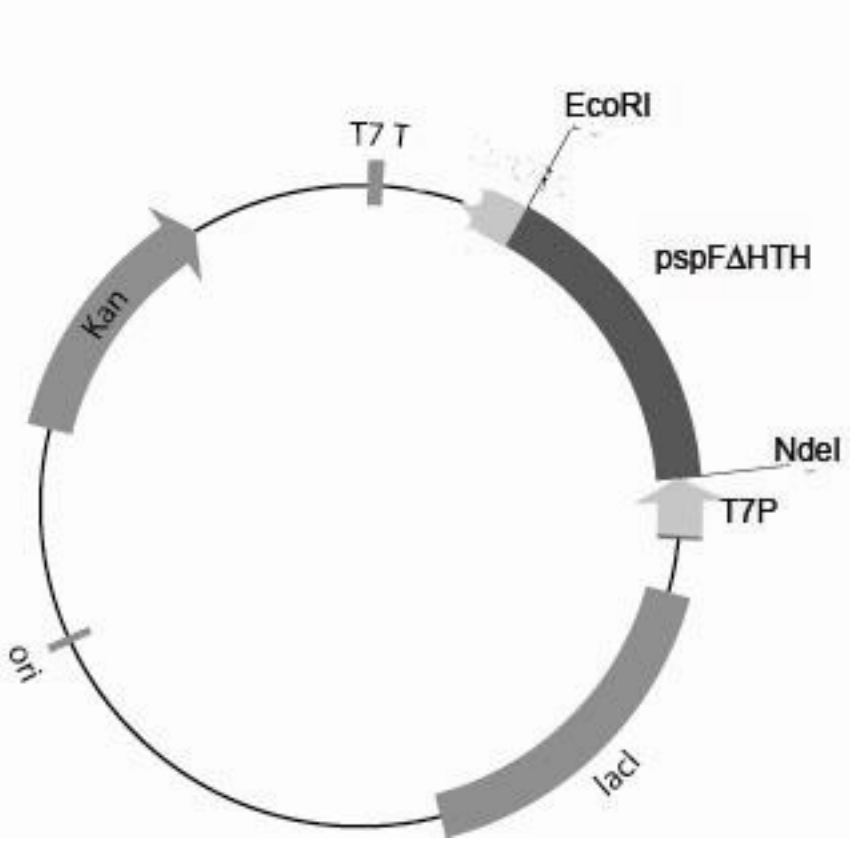


Vector Name: DCS46

Location: Notebook II

Construction: Blunt PCR insertion

Note: Removed DNA binding domain using a new reverse primer to truncate the gene

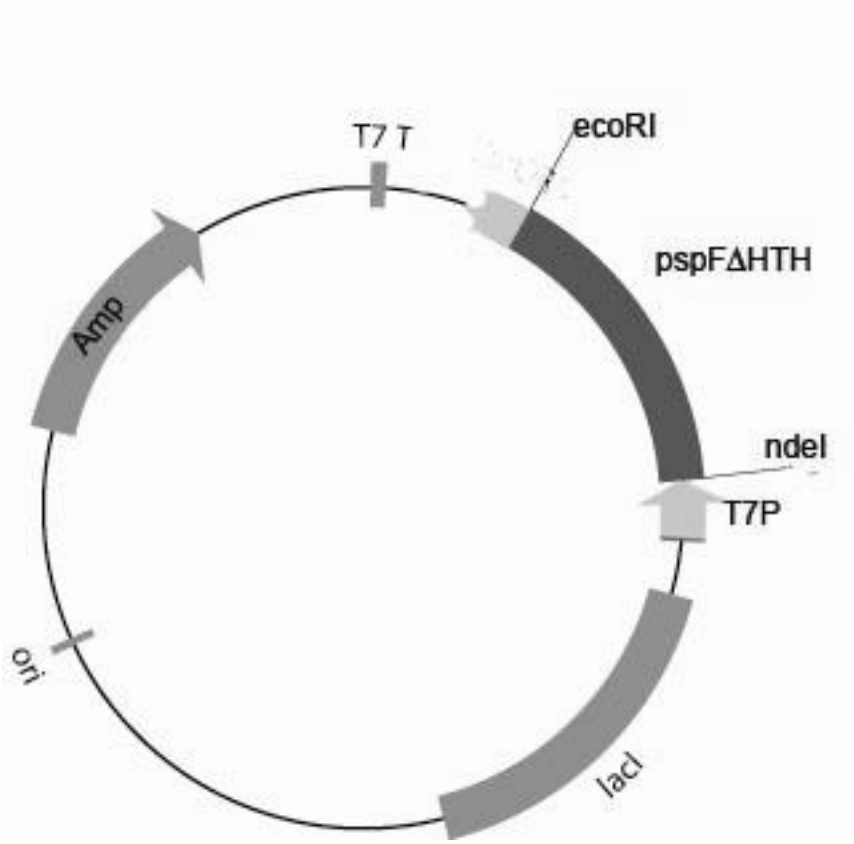


Vector Name: DCS47

Location: Notebook II

Construction: NdeI/EcoRI insertion

Note: Appropriate for protein isolation

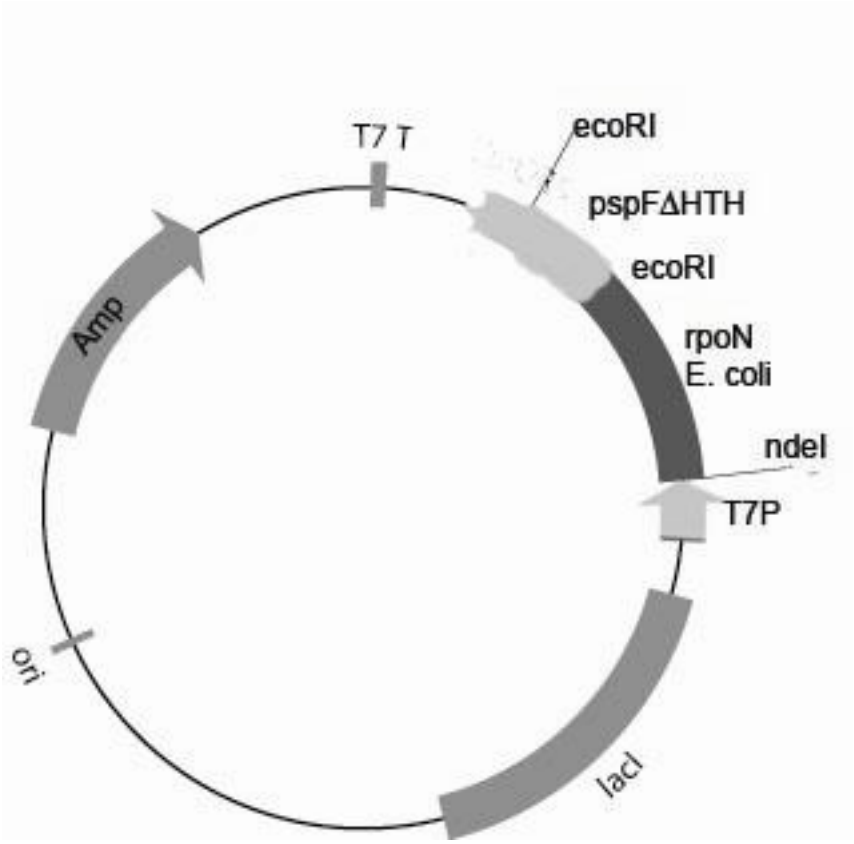


Vector Name: DCS48

Location: Notebook II

Construction: NdeI/EcoRI insertion

Note: n/a

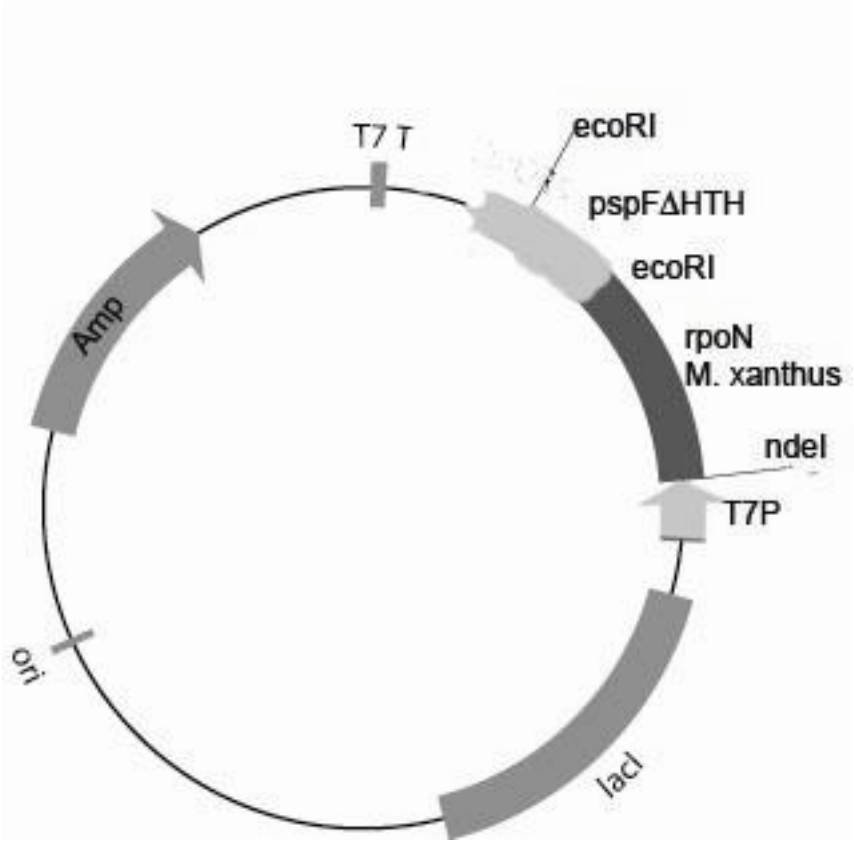


Vector Name: DCS49

Location: Notebook II

Construction: AvrII/XbaI insertion

Note: n/a

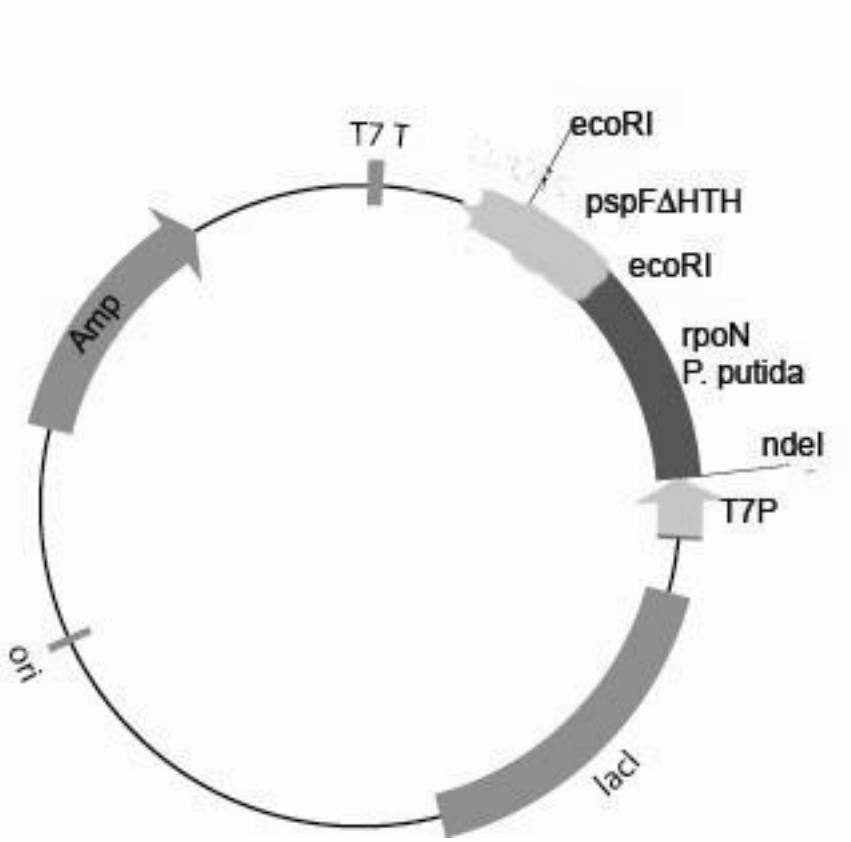


Vector Name: DCS50

Location: Notebook II

Construction: AvrII/XbaI insertion

Note: n/a

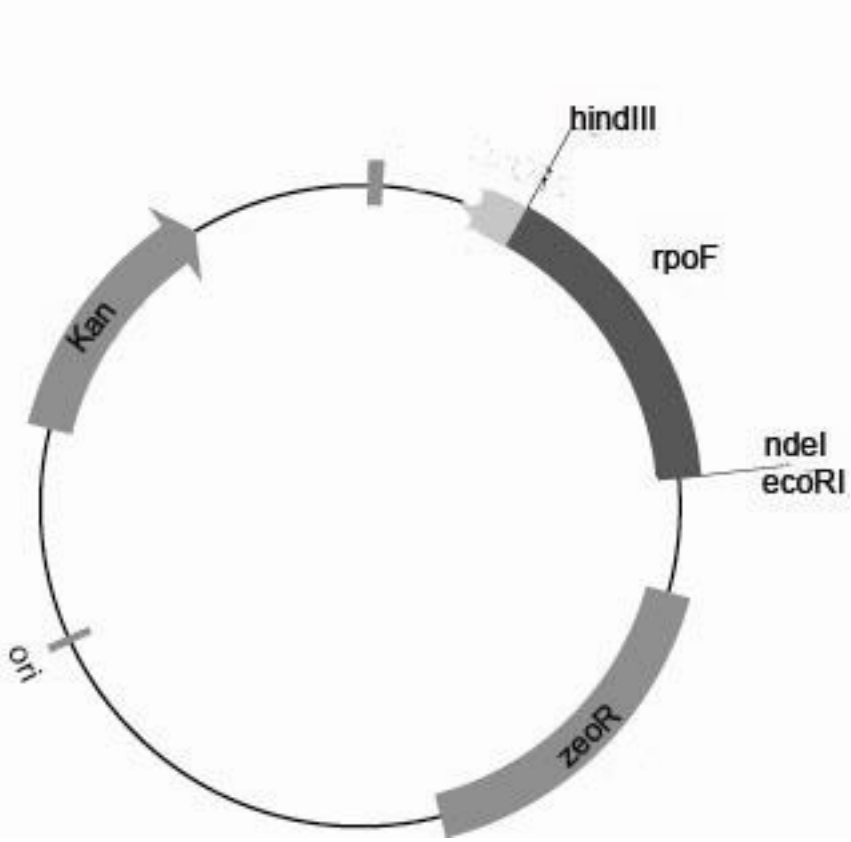


Vector Name: DCS51

Location: Notebook II

Construction: AvrII/XbaI insertion

Note: n/a

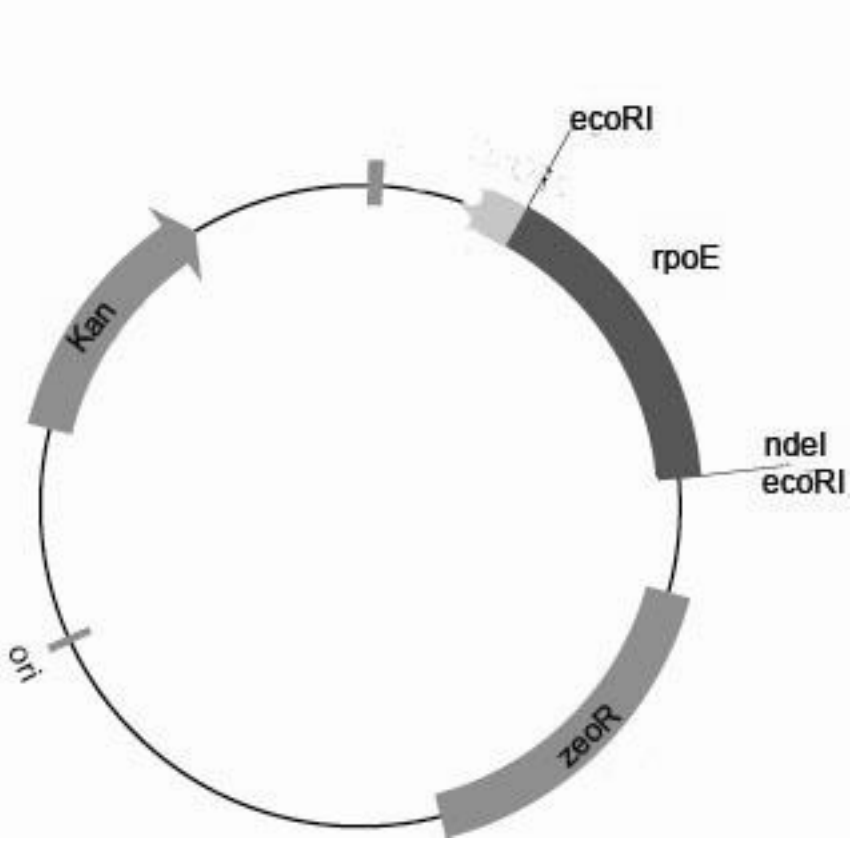


Vector Name: DCS53

Location: Notebook III

Construction: Blunt PCR insert

Note: n/a

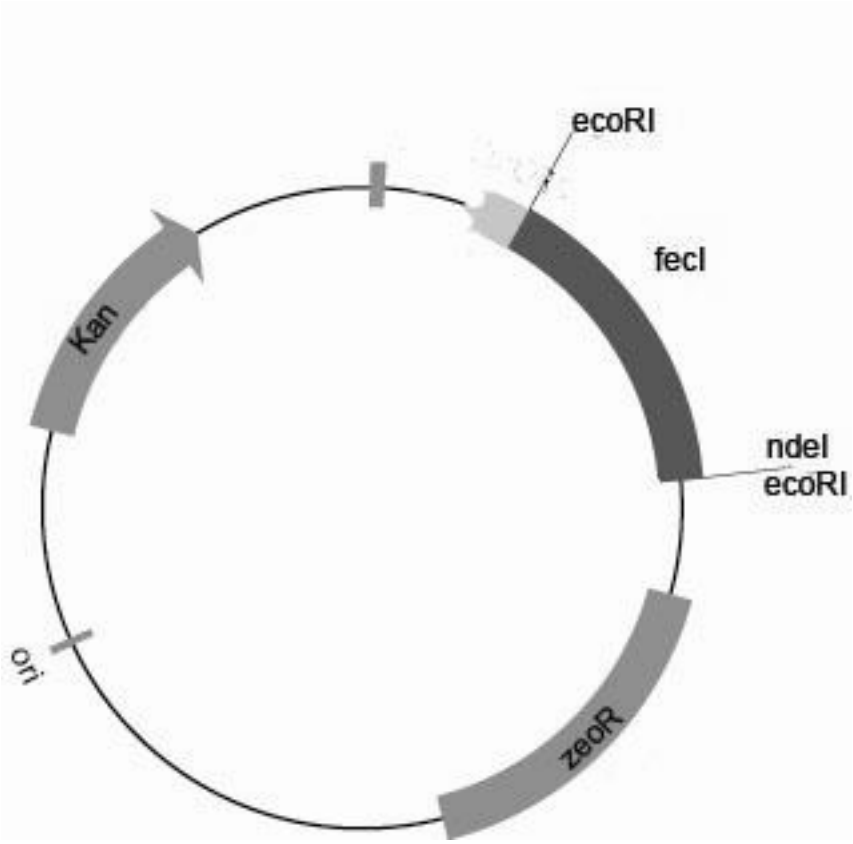


Vector Name: DCS54

Location: Notebook III

Construction: Blunt PCR insert

Note: n/a

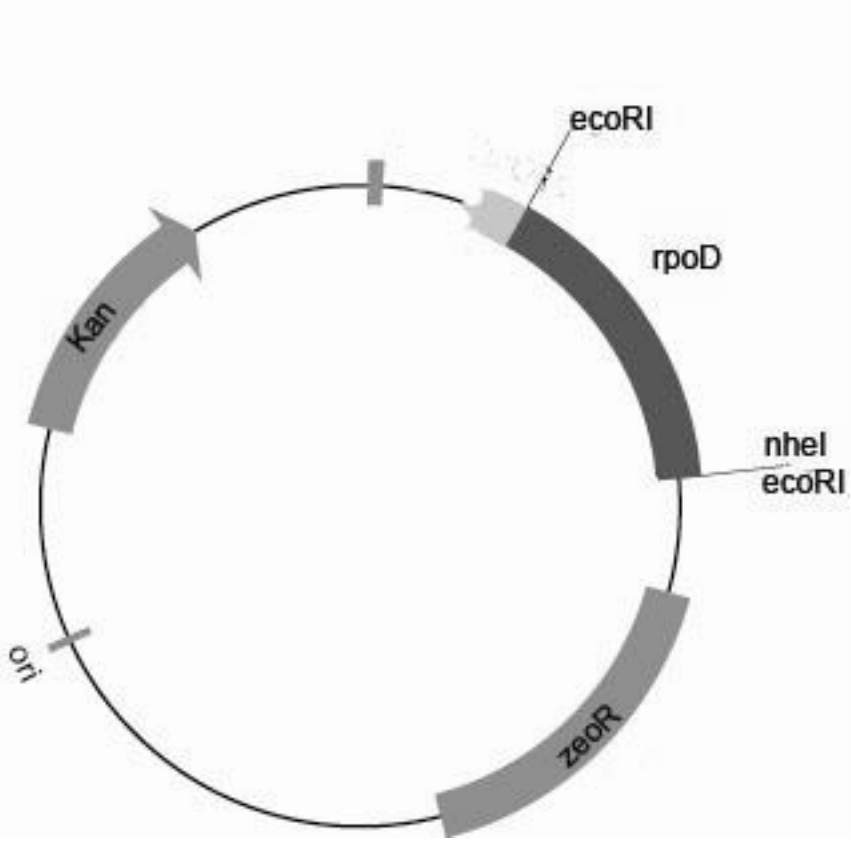


Vector Name: DCS55

Location: Notebook III

Construction: Blunt PCR insert

Note: n/a

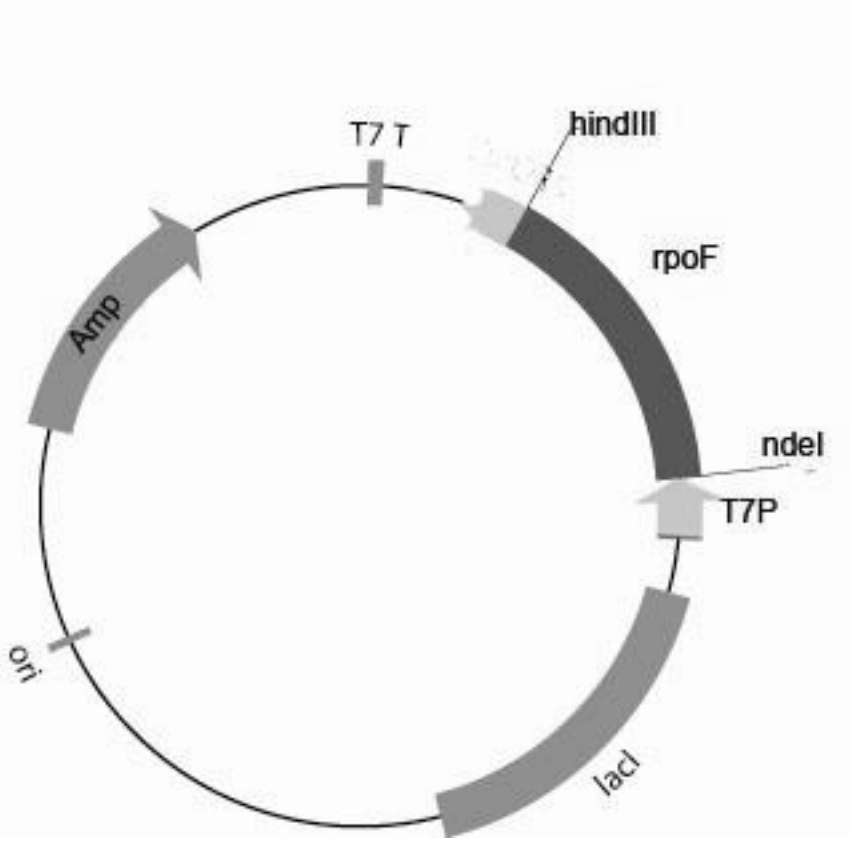


Vector Name: DCS56

Location: Notebook III

Construction: Blunt PCR insert

Note: n/a

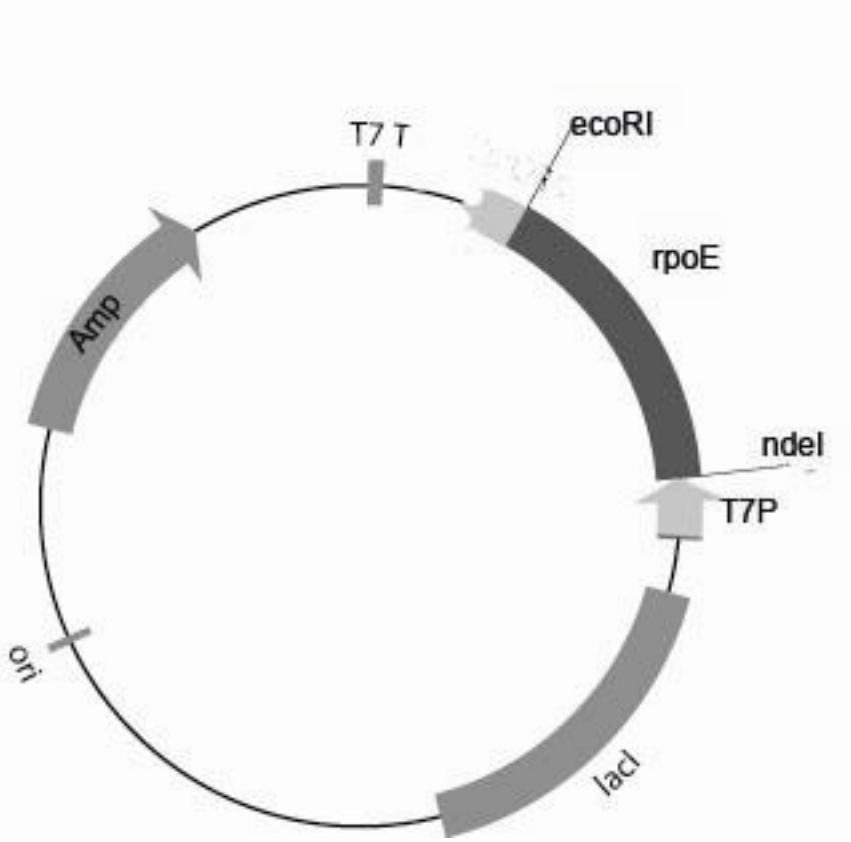


Vector Name: DCS57

Location: Notebook III

Construction: NdeI/HindIII insertion

Note: n/a

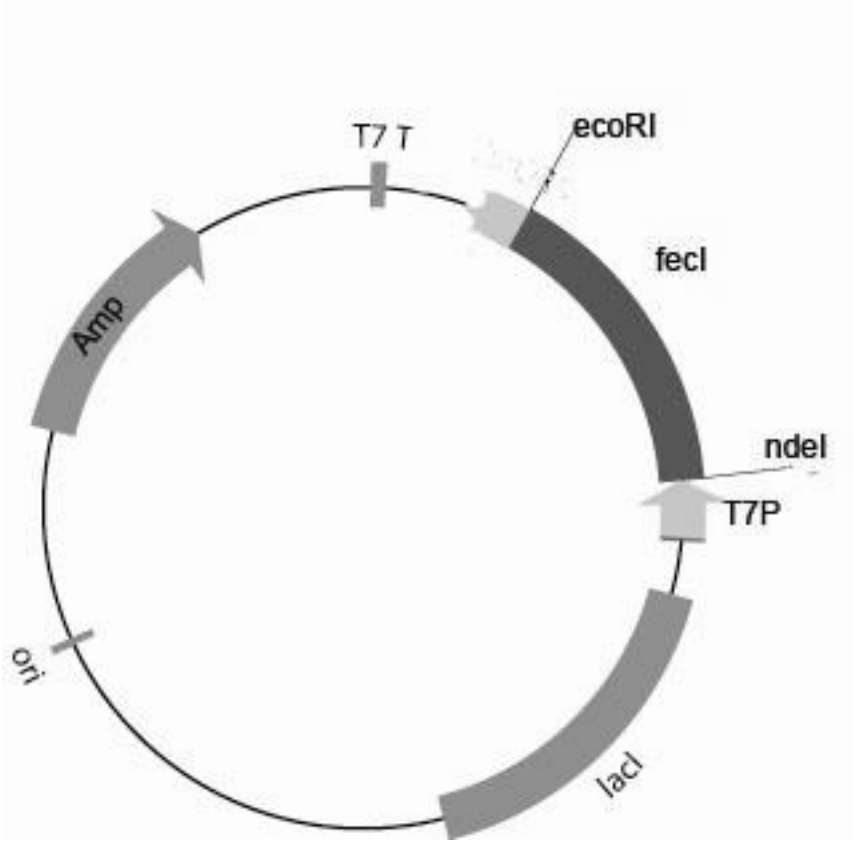


Vector Name: DCS58

Location: Notebook III

Construction: NdeI/EcoRI insertion

Note: n/a

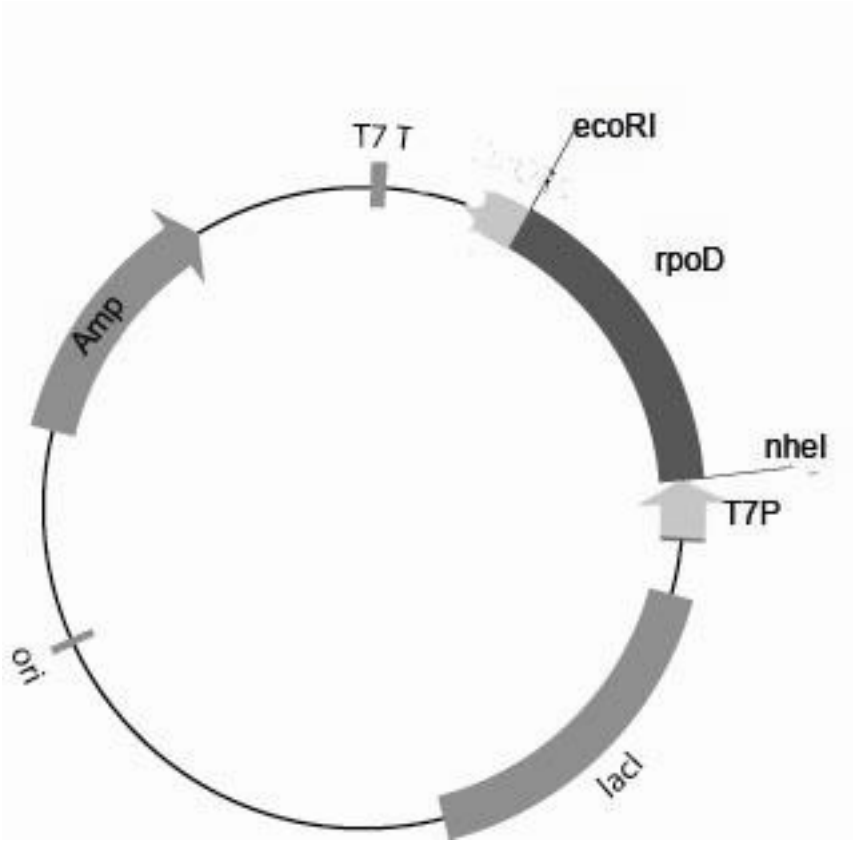


Vector Name: DCS59

Location: Notebook III

Construction: NdeI/EcoRI insertion

Note: n/a

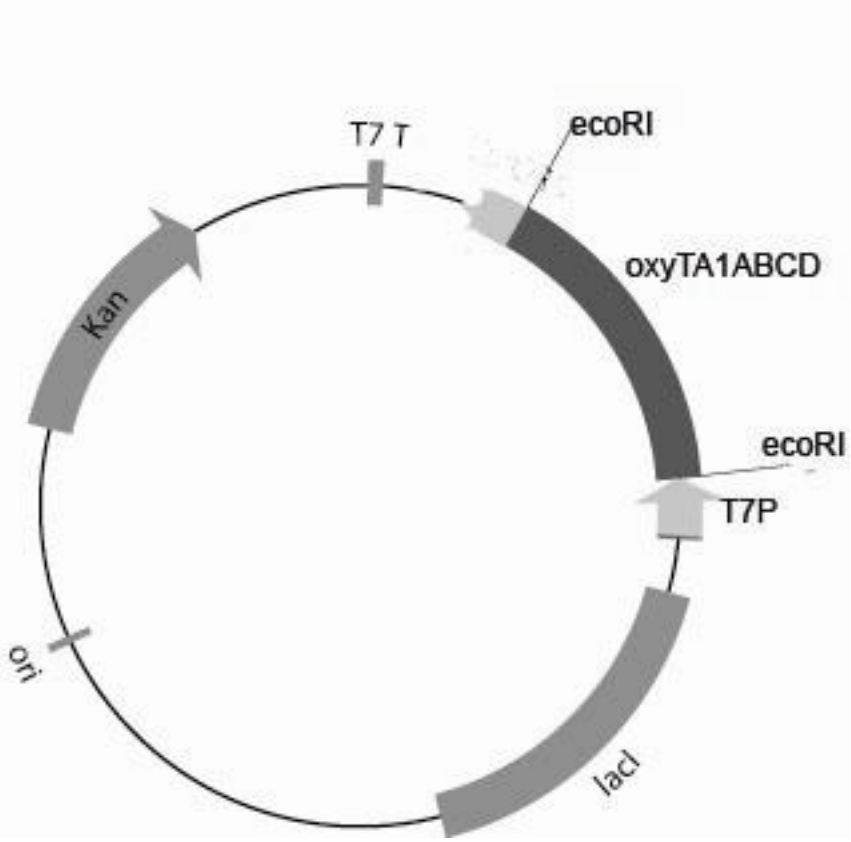


Vector Name: DCS60

Location: Notebook III

Construction: NheI/EcoRI insertion

Note: n/a

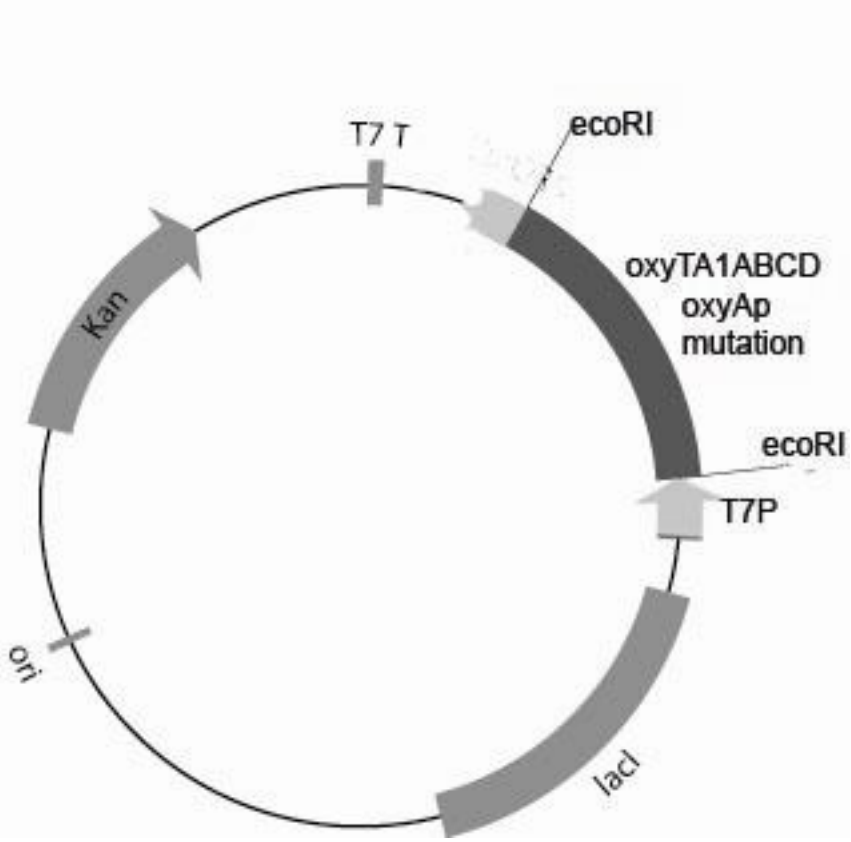


Vector Name: DCS61

Location: Notebook III

Construction: Splice by overlap extention using DK04r primer and T7p primer excised section with EcoRI

Note: n/a



Vector Name: DCS62

Location: Notebook III

Construction: Site-directed mutagenesis of pDCS61

Note: n/a