# NEW STRATEGIES IN THE LOCALIZATION OF NATURAL PRODUCT BIOSYNTHETIC PATHWAYS AND ACHIEVING HETEROLOGOUS EXPRESSION

A Dissertation

by

EUN JIN KIM

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

December 2009

Major Subject: Chemistry

# NEW STRATEGIES IN THE LOCALIZATION OF NATURAL PRODUCT BIOSYNTHETIC PATHWAYS AND ACHIEVING HETEROLOGOUS EXPRESSION

A Dissertation

by

EUN JIN KIM

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Approved by:

| | |
|---|---|
| Chair of Committee, | Coran M. H. Watanabe |
| Committee Members, | Kevin Burgess |
| | Daniel Romo |
| | Gregory D. Reinhart |
| Head of Department, | David H. Russell |

December 2009

Major Subject: Chemistry

# ABSTRACT

New Strategies in the Localization of Natural Product Biosynthetic Pathways and Achieving Heterologous Expression. (December 2009)

Eun Jin Kim,

B.S., M.S., Korea National University of Education;

M.S., Texas A&M University,

Chair of Advisory Committee: Dr. Coran M. H. Watanabe

Natural products have long furnished medical science playing a significant role in drug discovery and development. Their importance notwithstanding, it is estimated that less than 1% of microorganisms can be cultivated from environmental sources using standard laboratory techniques. It is therefore necessary to develop biochemical and genetic techniques to access these uncultivable genomes.

Here as a point of departure toward this goal, two cDNA libraries of two microorganisms were constructed along with five fosmid libraries with DNA isolated from marine environmental samples. We describe the construction of cDNA libraries from marine microbial species and detail experiments to exploit these libraries for their natural product biosynthetic pathways and other metabolic enzymes they harbor. However, no useful biosynthetic pathways were detected within the cDNA libraries.

Genetic selection by complementation was additionally explored as a method to identify and localize biosynthetic gene clusters within marine microbial DNA libraries. Genetic selection is a fast and economic method which utilizes selection of a part of a

pathway to represent the presence of an entire pathway for the complementation of known mutant strains. We describe genetic selection to localize biotin biosynthetic pathways of Hon6 (*Chromohalobacter sp*.) as a proof of principle experiment for the identification and localization of biosynthetic pathways in general.

Instead of developing purification methods or manipulating cultivation conditions, large fragments of non-culturable bacterial genomes can be cloned and expressed using recombinant DNA technology. A strong transcriptional promoter to control high-level gene expression is required in recombinant expression plasmids. We aimed to develop new tools to control gene expression through the use of riboswitches. Riboswitches are metabolite-sensing ribonucleic acid (RNA) elements that possess the remarkable ability to control gene expression. The thiamine pyrophosphate (TPP) riboswitch system was chosen as it will enable use of *E. coli* as a suitable host strain. This system is particularly attractive because it has one of the simplest structures among the riboswitches elucidated to date. The use of the TPP riboswitch will also enable modulation of pathway gene expression by varying the TPP coccentration as many gene products are toxic. The violacein gene cluster from *Chromobacterium violaceum* was selected and placed under the control of this riboswitch. We describe modulation of heterologous gene expression by the ThiC/Riboswitch and detail experiments to investigate the expression and manipulation of the gene cluster under various promoters.

## ACKNOWLEDGEMENTS

I would like to express my gratitude to my committee chair, Prof. Coran M. H. Watanabe, for her guidance, supervision and support during the course of my graduate program. I would like to thank my committee members, Prof. Kevin Burgess, Prof. Daniel Romo, and Prof. Gregory Reinhart, for serving on my advisory committee and for their guidance and support throughout the course of this research.

Thanks also go to my friends and colleagues, and the department faculty and staff for making my time at Texas A&M University a great experience: specifically, Dr. Scott Angell, Dr. Thom Kelly, Dr. Bennie J. Bench, Jennifer Foulke-Abel, Dr. Huitu Zhang, and Hillary Agbo. I really appreciate their encouragement and being my friends. I will not forget their friendship and the memories of my time in the Watanabe group. I will always remember them as my second family.

In addition, special thanks are due to Prof. Tadhg Begeley for sharing strains as a generous gift, to Prof. Dong Hyun Hong for valuable advice of cloning, to Dr. Charlie Roessner for sharing strains and valuable advice for this research, and to Dr. Howard Williams for running NMR samples. Special thanks go to Ming-chien Li and Roy Estrada for running analytical HPLC and setting up the condition for prep-HPLC.

The value of family and friends who supported me throughout the graduate program has been immeasurable. My thanks go to my friends, Chang Suk Lee, Eun Hwa Ko, Dr. Seong Wook Cho, Dr. Mi Ok Kim, Dr. Hyun Sook Lim, Dr. Cheol Woo Rhi, Dr. Yong Gu Kim, Dr. Seong Ho Shin, Ann Hwang, Dr. Benjamin Philmus and Vision Mission

Church members, for their support and sharing scientific experiences, or emotional encouragement and prayers.

I thank my parents and parents-in-law for their encouragement and prayers. I thank my husband Daewon Cho and my daughter Victoria Minyoung Cho for their patience, companionship, support and more love than I could ever have asked for, or could have hoped for. They have my eternal gratitude and love.

Finally, I would like to express my heartfelt gratitude and eternal love to the God. I always repeated your words in my mind when I had difficult and hard times. I know that you always are with me and walking by me.

*"I can do everything through him who gives me strength." (Philippians 4:13)*

# TABLE OF CONTENTS

Page

CHAPTER Page

## LIST OF FIGURES

Page

Page

Page

Page

Page

Page

Page

# LIST OF TABLES

Page

# LIST OF SCHEMES

Page

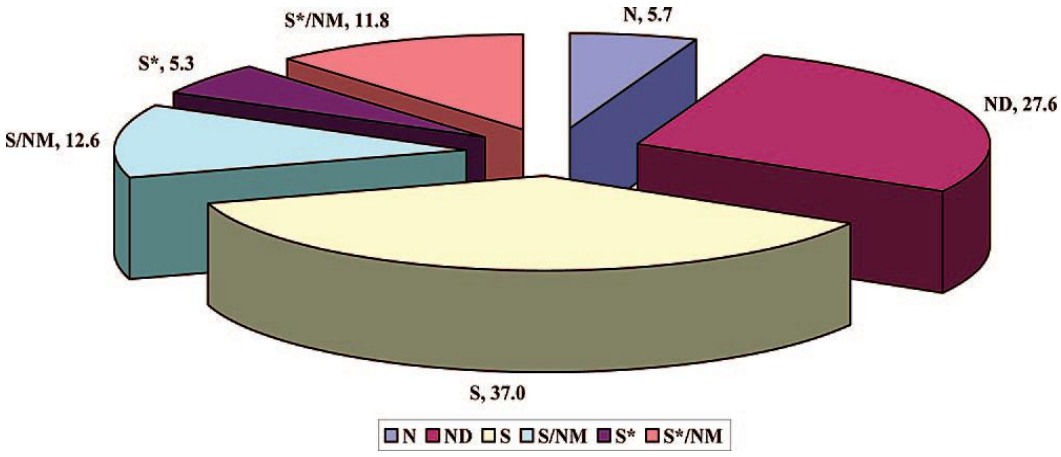
# CHAPTER I

## INTRODUCTION

Natural products are chemicals made by living organisms involving primary metabolites (such as encoded amino acids, nucleotides, protein, enzyme, oligonucleotides, compounds used for energy supply or cell homeostasis, lipids, hormones, pheromones) and secondary metabolites ('unnatural' amino acids and nucleotides, glycosides, compounds used for defense mechanisms). Many of the currently used pharmaceutical agents are natural products or were derived from natural product leads (*1*). For example, many natural products play a significant role in the treatment of the cancer and infectious disease (*2-4*). Many secondary metabolites are in use today as antibiotics, antitumor agents, antifungals, immunosuppressive agents, antivirals, and enzyme inhibitors. Indeed, approximately 60% of new drugs were derived from natural products and their structural analogs over the past 26 years as shown in Scheme 1 (*4, 5*). One of the most successful anti-cancer agents is Taxol, shown in Figure 1, which is produced by the Pacific Yew tree. Another natural product, vancomycin, isolated from *Amycolaptosis orientalis*, is used for the treatment of drug-resistant *Staphylococcus aureus* infections.

---

This dissertation follows the style of *Biochemistry*.




Scheme 1: All new chemical entities, 1981-2006, by source (N=1031) after Newman (*5*).
S=Totally synthetic drug; N= Natural product, ND=Natural product derivative, usually semi-synthetic, S*=Made by total synthesis but pharmacophore is a natural product, NM=Natural product mimic

Both drugs have each generated $1 billion in pharmaceutical sales per year (*6*, *7*). Therefore, natural products are very important not only as a primary source of drug leads but also as an economic source for the pharmaceutical market.

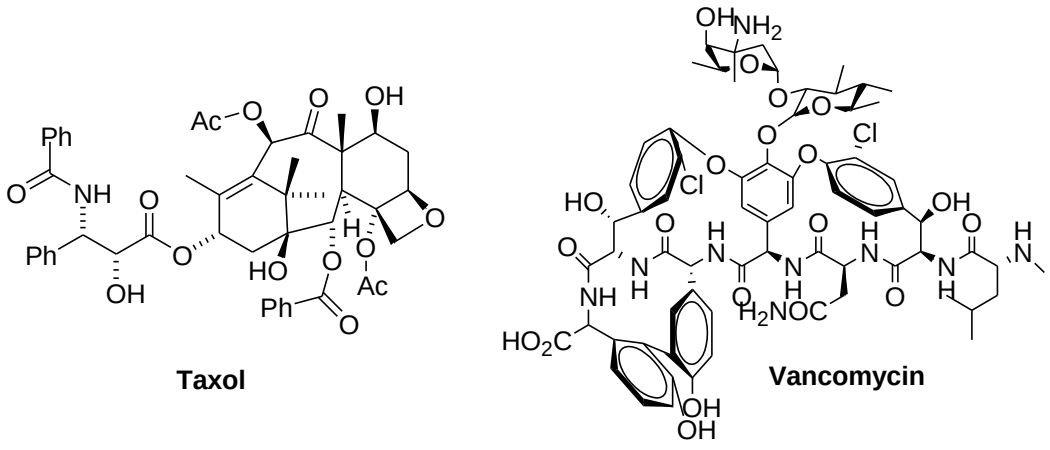
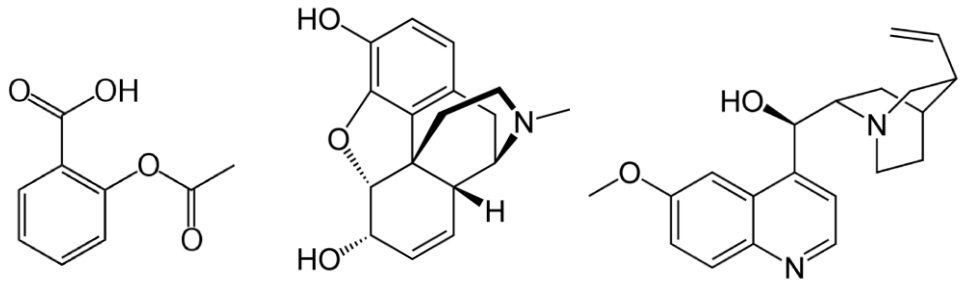



Figure 1: Structure of Taxol and Vancomycin.

## THE HISTORY OF NATURAL PRODUCT FOR DRUG DESIGN

What are the origins of drugs? Where did they originate? Evidence suggests that they originated tens of thousands of years ago. Research discoveries indicate that the medicinal use of natural products has been in use long since man evolved (6, 9). Palaeoanthropological studies have revealed pollen deposits in one of the graves at the cave site of Shanidar in Iraq where Neanderthals lived more than 60,000 years ago. This suggested that the Neanderthals might have been aware of the medicinal properties of numerous plants (*8*). We can find many examples of natural medicines that were used for human diseases and/or disorder in the traditional Ayurvedic and Chinese herbal treatments (*9*). The ancients took medicine for pain relief from the bark of trees, leaves, fruits, the extract of the seeds and many other sources in nature. People in Asia are still using these natural sources for treatments. Some over-the-counter drugs were started from extraction of natural sources. One famous example is Aspirin, which was extracted from willow bark containing salicyclic acid. The extract of willow bark has been used for mild to severe pain relief until the French chemist Charles Frederic Gerhardt synthesized acetylsalicylic acid (shown in Figure 2) in 1856 (*10*). The first semi-synthetic pure drug based on the natural product, aspirin, was synthesized by Bayer in 1899 (*10, 11*). Morphine and Quinine are also plant extracts that have been used in medicine for a long time. Morphine was discovered as the first active alkaloid extracted from the opium poppy plant in 1804 (*12*). The first commercial pure natural product, morphine, was available by E. Merck in 1826 (*11*). Quinine was only obtained from

Cinchona trees and was used as a muscle relaxant until American chemists R.B.Woodward and W.E. Doering synthesized the products in 1944 (*13*).




A B C

Figure 2: Structure of aspirin (A), morphine (B) and quinine (C).

The "Golden age" of antibiotic discovery was from 1945 to 1960. Many antibiotics of natural product origin were discovered such as daptomycin, erythromycin, chloramphenicol, vancomycin, the penicillins, the cephalosporins, the carbapenems, ramoplanin, tetracycline, streptomycin, ceftrizxone, and azithromycin (*9, 14*). The first ten antibiotic structures are shown in Figure 3. Daptomycin is a novel lipopeptide antibiotic used in the treatment of certain infections caused by Gram-positive organisms (*15, 16*). Erythromycin is a macrolide antibiotic (*17*). Chloramphenicol is a bacteriostatic antimicrobial originally derived from the bacterium *Streptomyces venezuelae*, isolated by David Gottlieb, and introduced into clinical practice in 1949 (*18*). It was the first antibiotic to be manufactured synthetically on a large scale, and is considered the prototypical broad-spectrum antibiotic (*19*). Vancomycin is a glycopeptide antibiotic used in the prophylaxis and treatment of infections caused by Gram-positive bacteria

(*20*). The penicillins are a group of antibiotics derived from *Penicillium* fungi (*21, 22*). Cephalosporins are a class of β-lactam antibiotics originally derived from *Acremonium*, which was previously known as "*Cephalosporium*" (*23*). Carbapenems are also a class of β-lactam antibiotics with a broad spectrum of antibacterial activity (*24, 25*). Ramoplanin is a glycolipodepsipeptide antibiotic drug derived from *Actinoplanes* strain ATCC 33076 (*26*). Tetracycline is a broad-spectrum polyketide antibiotic produced by the *Streptomyces* genus of Actinobacteria (*27*). Streptomycin is an antibiotic drug, the first of a class of drugs called aminoglycosides to be discovered, and was the first antibiotic remedy for tuberculosis (*28, 29*). Ceftriaxone is a third-generation cephalosporin antibiotic; it possesses broad spectrum activity against Gram positive and Gram negative bacteria (*30*). Azithromycin is an azalide, a subclass of macrolide antibiotics (*31*).

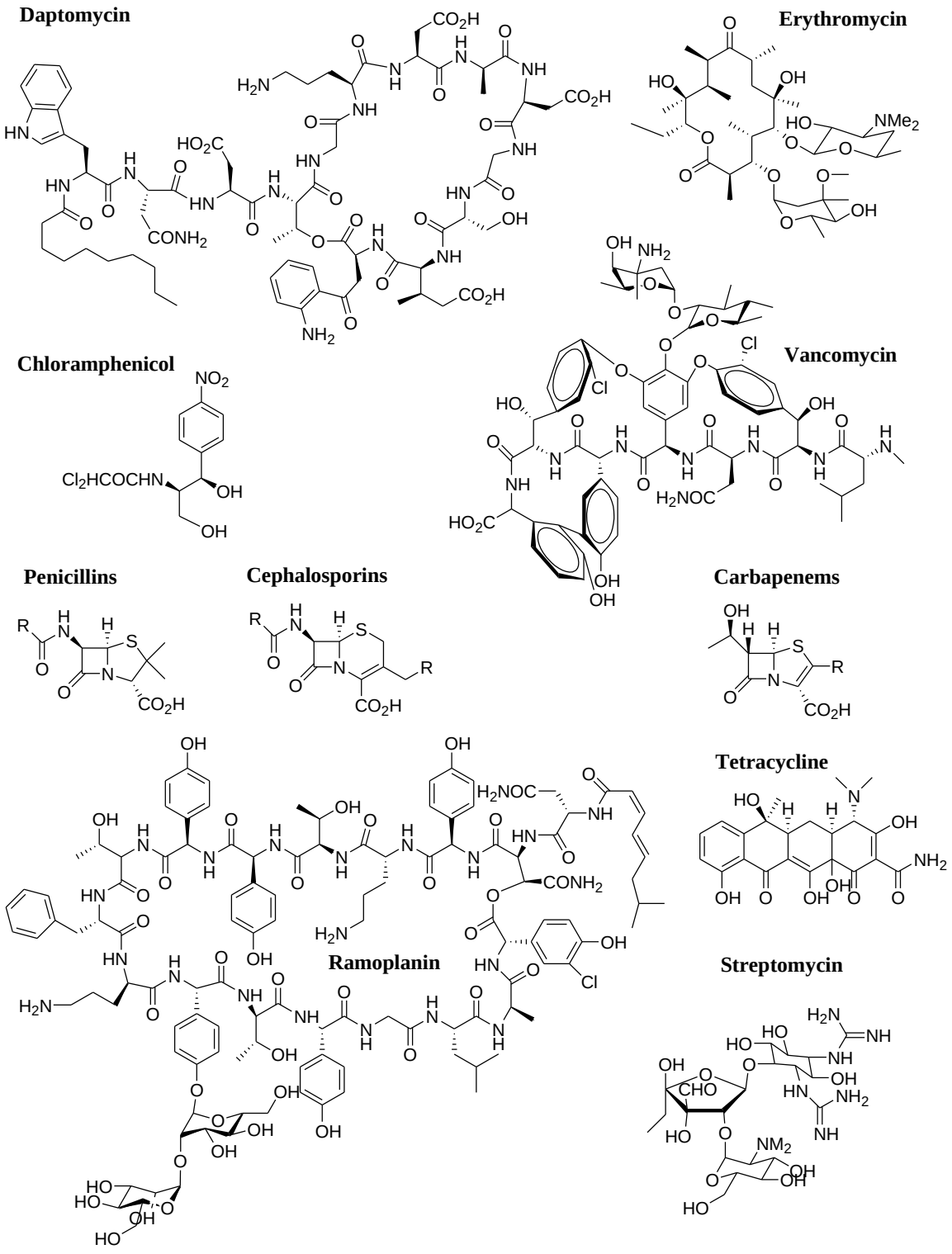



Figure 3: Structures of antibiotics of natural product origin.

Microbiologists developed assays for the identification and characterization of the antibacterial activity of natural products (*14*). Natural products can be derived from bacteria, fungi, marine organisms, terrestrial plants and other sources. A crude extract from any one of these sources contains novel, structurally diverse chemical compounds. Chemical diversity in nature is based on biological and geographical diversity. Researchers therefore travel around the world to obtain samples which may contain novel compounds. Plants are particularly interesting because they have the broadest spectrum of biosynthetic capability and produce a wide variety of compounds. Plants use simple starting materials such as water, carbon dioxide, nitrogen, phosphorus compounds and salts. Animal samples are not as easy to extract. There are still many potential sources of natural products in nature, even though 10,000 bioactive substances have been isolated from bacteria, 450 from mollusks and about 250 from insects and worms.(4, 11, 12) The latest version of "The Dictionary of Natural Product" has just over 214,000 entries (DNP; http://dnp.chemnetbase.com) (6). Even though there are many possible sources of natural products in nature, marine metabolites have not been rapidly utilized as clinical agents because of poor metabolite production. For example, bryostatin which has 20 different known structures has antineoplastic activity (the structure is shown in Figure 4), so this class of compound can be used chemotherapeutic cancer drug. However, we can only obtain about 1.5g of bryostatin1, 2g of bryostatin2, 0.0056g of bryostatin 8 and 0.0086g of bryostatin 15 from about 1,000kg of damp *Bugula neritina* (collection from the Southern California Coast) (*32*).

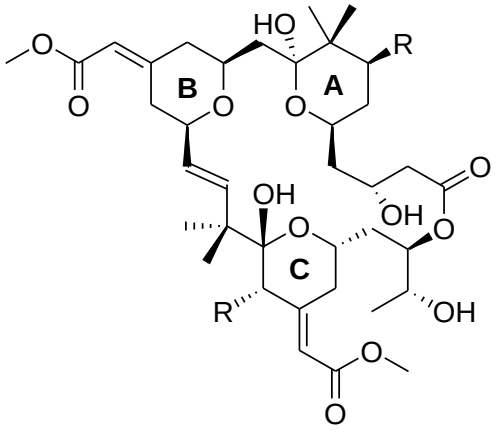



Figure 4: Structure of bryostatin.

Moreover, the vast majority of sponge species can only be found in small quantities. While these amounts might be sufficient to determine bioactivity, it is not typically adequate for structure determination studies. For the limited sample supply, researchers attempt to cultivate or synthesize the natural product source. Unfortunately, less than 1.0% of the marine microorganism can be cultivated shown in Table 1 (*33, 34*). Therefore, there has been wide interest in the development of new methodologies and culture techniques for the production of marine-derived natural products to obtain new drug sources or drug leads.

Table 1: Culturability of bacteria from different habitats. From Amman, *et al*. (*34*)

| Habitat | Culturability (%) |
|---|---|
| Seawater | 0.001 to 0.1 |
| Freshwater | 0.25 |
| Mesotrophic lake | 0.1 to 1 |
| Unpolluted estuarine | 0.1 to 3 |
| Activated Sludge | 1 to 15 |
| Sediments | 0.25 |
| Soil | 0.3 |

As the field of combinatorial chemistry emerged in the early 1980's (*5*), the interests in natural products moved from extraction of natural sources to synthesis of the natural products. Borman described combinatorial chemistry as a mainstream tool in drug discovery to advance synthesis, purification and analysis (*35*). New characterization methods using modern chromatography and spectroscopy, including one of the only available comprehensive treatments of the use of nuclear maganetic resonance (NMR) spectroscopy for the characterization of natural products, made great impetus on the synthetic endeavors to make the natural products with already-known biological activity, antimicrobial and antiviral properties or anticancer activity. Researchers attempted to find lead compounds which contain bioactivity. A lead compound can be produced by total synthesis or a precursor from a semisynthetic compound. It can serve as a template for a structurally different synthetic compound. As mentioned earlier, natural products from secondary metabolites often have very complex structures. Its complex structure makes it unique but may also be difficult to synthesize. The next popular technique was diversity-oriented synthesis (DOS) which is a strategy for quick access to molecule libraries with an emphasis on skeletal diversity (*5*). It was used in Schreiber group in the late 1990's (*5, 36*). In the time frame from 2000 to date, over 300 articles have been published with the DOS approach.

In the late 1990's, synthetic chemists realized that the synthetic compounds, such as nucleosides, peptides, and carbohydrates, lacked the complexity normally related to bioactive natural products which have multiple chiral centers, polycyclic structures and heterocyclic substituents (*5*). To discover and develop new drugs, multidisciplinary approaches can be combined with total, diversity oriented synthesis approach and

combinatorial synthetic methodologies (*4*) according to the development of characteristic methods. In the case of drug discovery efforts, structure elucidation of all components that are active *in vitro* is typically the end goal. Therefore, the structures of natural products were elucidated and characterized by modern spectroscopy; UV (ultraviolet 1930's) IR (infrared 1940's), NMR (nuclear magnetic resonance 1950's), MS (mass spectrometry 1960's) and ESR (electron spin resonance), and by other methods; ORD (optical rotator dispersion), CD (circular dichroism), acidity and basicity measurements (pK), advanced synthetic and biosynthetic technology and x-ray crystallography. Modern methods reduce the necessity of chemical degradation methods, so much less material is required to elucidate the structure. Classical structural eludication is done by determination of functional groups, determination of the carbon skeleton and the location of the functional groups, degradation to smaller fragments, elemental analysis and stereochemistry, synthesis of the smaller fragments and the entire molecule, and classification of the compound into a biogenetic family of compounds. Natural products can be classified by origin of biosynthetic pathways. A few biosynthetic pathways will be discussed to introduce the origin of natural products briefly.

## NATURAL PRODUCT BIOSYNTHETIC PATHWAYS

Isolated natural products contain a wide variety of structural features, yet the origin of these molecules flows from basic pathways related to primary metabolism. Metabolism is the set of chemical reactions occurring in living organisms involved in normal growth,

reproduction, maintenance of their structures, and response to their environments. Primary metabolites include amino acids, nucleotides, fatty acids, and vitamin complexes which produce proteins, nucleic acids, and cell membranes required for the survival of the organism. Secondary metabolites are typically classified upon the basic origin of their construction from both the metabolic building blocks and the molecular machinery that assembles the units. These natural products are produced by a few pathways including the fatty acid synthases (FAS), the polyketide synthases (PKS), the non-ribosomal peptide synthetases (NRPS), as well as using the terpene and alkaloid biosynthetic pathways. Polyketide synthases (PKS) are similar to fatty acid synthases (FAS), producing carbon chains of repeating units in various states of oxidation. NRPS pathways provide a route for the formation of polypepetide units from natural and unnatural amino acids. PKS pathways are constructed with structurally diverse starting units with subsequent two carbon (C2) extension units also called acetate units. Alkaloids are organic, nitrogenous bases found mostly in plants, but also to some extent in microorganisms.

## THE ACETATE PATHWAYS: FATTY ACIDS AND POLYKETIDES

*Fatty Acids.* Fatty acid biosynthesis is catalysed by a fatty acid synthase by the polymerization of subunits through a thio-Claisen condensation reaction. Figure 5 depicts the processive nature of fatty acid biosynthesis. Fatty acid construction begins with an acetate unit (keto starting unit) bound to the fatty acid synthase enzyme giving enzyme-bound thioesters (SEnz). After decarboxylation of a malonyl unit, the C2 unit is

added via attack of the carbonyl carbon on the starter unit creating a β-ketothioester by the keto synthase (KS) domain. The starter unit, which is enzyme-bound, is released from the enzyme. The extended unit is converted to a β-hydroxy thioester by the ketoreductase (KR) domain. The loss of a water molecule by the action of the dehydratase (DH) domain, turns the alcohol the α,β-unsaturated thioester. The final product of fatty acid biosynthesis can be obtained after reduction by the enoyl reductase (ER) domain resulting in the saturated product. The process is repeated until the desired lenghth is obtained after transfer of the molecule to the enzyme from the phosphopantetheinyl arm. As shown in Figure 5, path B releases the fatty acid by way of the thioesterase domain (TE) while further extension can occur through path A.

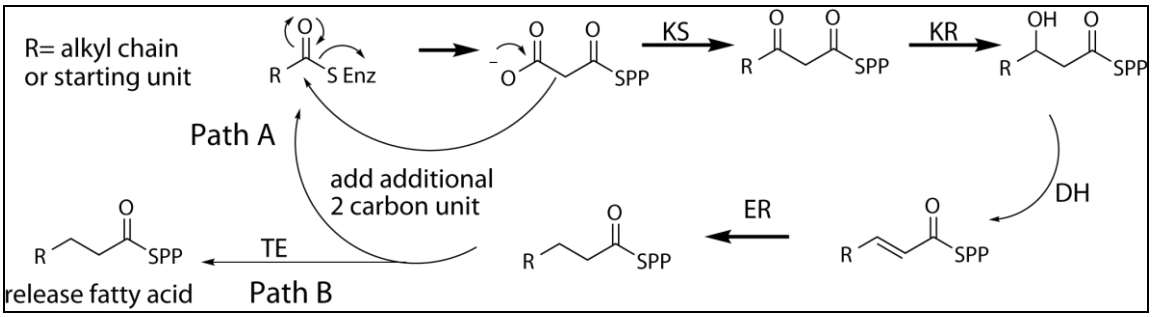



Figure 5: Fatty acid biosynthesis. Fatty acid biosynthesis proceeds through a cyclic mechanism, beginning with a Claisen condensation-like carbon-carbon bond forming reaction, then proceeding through a reduction, dehydration, and a second reduction to generate a saturated chain. KS - ketosynthase, KR - ketoreducatse, DH - dehydratase, ER -enoyl reductase, SPP - phosphopantetheine thiol, SEnz - enzyme bound cysteine. Path A adds additional 2 carbon units to the growing chain. Path B releases the fatty acid by way of the TE – thioesterase domain. Adapted from Dewick, 2002 (*37*).

*Polyketides*. Polyketides are the building blocks for secondary metabolites from a variety sources. Polyketides are biosynthesized by the polymerization of propionyl and acetyl subunits by a thio-Claisen condensation similar to that seen in fatty acid biosynthesis. Polyketides have diverse structures with a variety of biological activities and pharmacological properties. Polyketides are produced by the enzyme known as polyketide synthases (PKS). PKSs are an important source of naturally-occurring small molecules used for chemotheraphy (*38*) like epothilone B (*39*). Many antibiotics such as tetracycline and the macrolides are also produced by PKSs (*39*). Polyketide synthases can be divided into three classes: Type I, Type II and Type III. Type I PKSs are further subdivided into iterative PKSs and modular PKSs. Type I polyketide synthases are large and highly modular proteins. If the PKSs repeat domains in a cyclic fashion, it belongs to the iterative Type I PKSs class. If the PKS does not reuse domains and contains a sequence of discrete separate modules, it is classified as a modular Type I PKS. Type II polyketide synthases are aggregates of monofunctional proteins and Type III polyketides synthases do not use the acyl carrier protein (ACP) domains. Type III PKSs do not share domains with fatty acid synthease, so it is a unique polyketides synthase. Type I and Type II are illustrated in Figure 6.

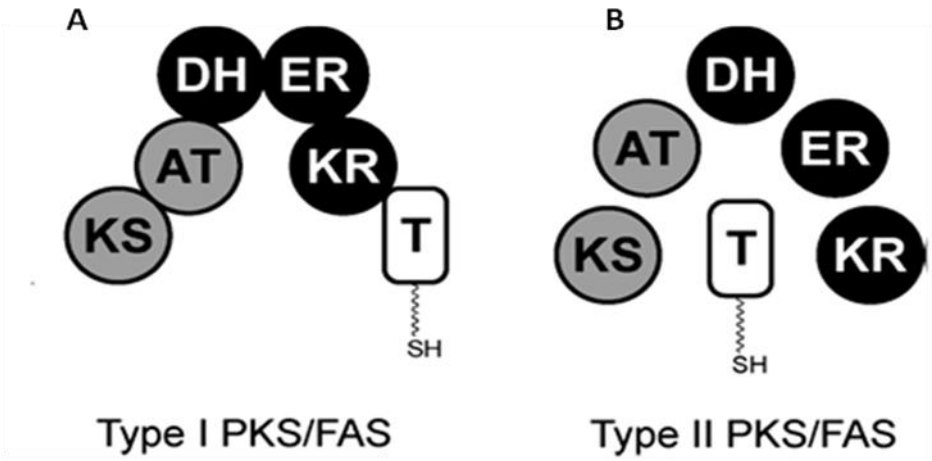



Figure 6: Comparision of Type I and Type II PKSs and FASs. (A) In Type I PKS/FAS systems, protein domains are connected in cis, (B) In Type II PKS/FAS systems, protein domains interact in trans. After Walsh *et al.*, 2006 (*40*).

The domains typically found in PKSs are the AT, ACP, KS, KR, DH, ER and TE domains; AT: Acyltransferase, ACP: Acyl carrier protein with an SH group on the cofactor, a serine-attached 4'-phosphopantetheine, KS: Keto-synthase with a conserved cysteine residue, KR: Ketoreductase, DH: Dehydratase, ER: Enoylreductase, and TE: Thioesterase. Starting modules typically consist of only on AT and ACP domains while the TE domain is involved in the termination or release module. Elongation or extending modules are combinations of -KS-AT-[DH-ER-KR]-ACP- domains. As an example, six different modules were used for the elongation to synthesize 6-deoxyerythronolide (DEBS). The modular structure of DEBS is shown in Figure 7. Each module consists of several domains with defined functions.

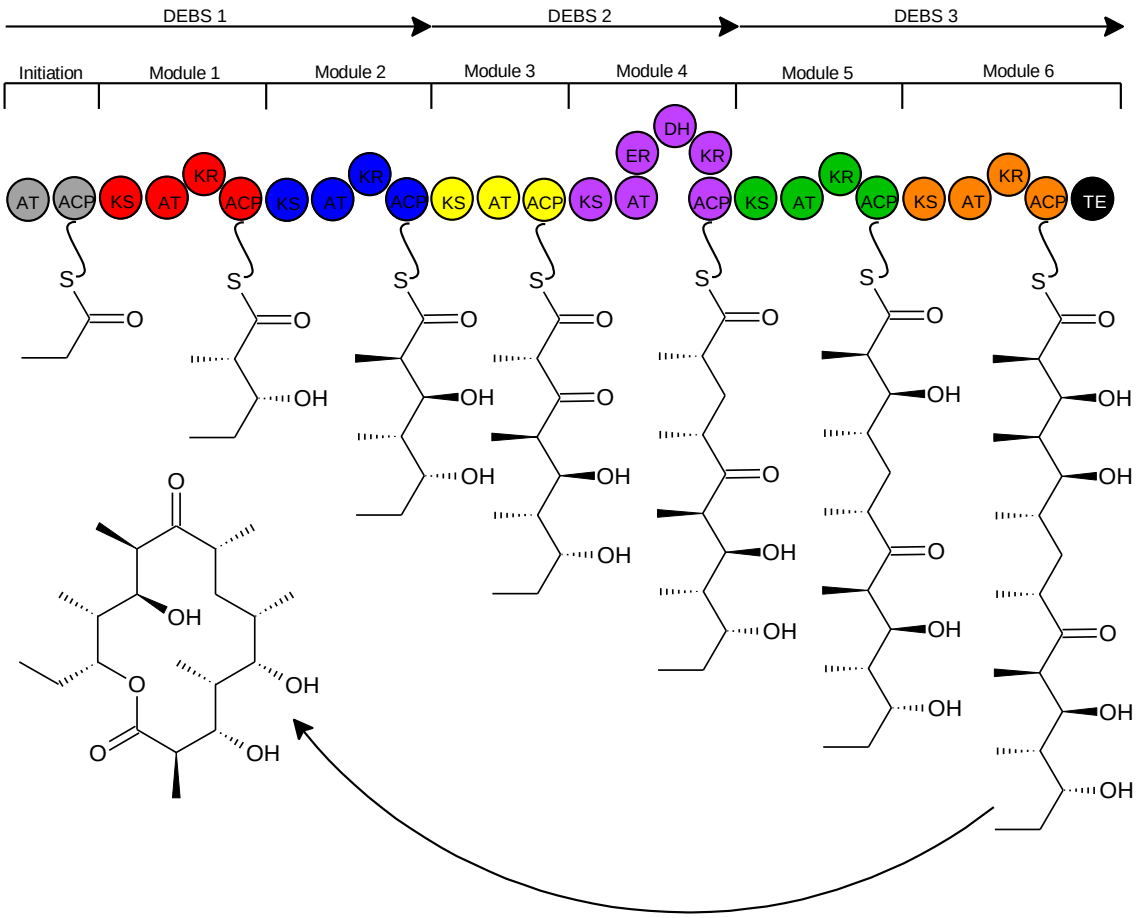



Figure 7: Modular structure of DEBS (6-Deoxyerythronolide) synthetase. After Staunton, *et al.* (*41*).

Polyketides are initiated by loading a starter unit, usually acetyl-CoA or malonyl-CoA, onto the ACP domain of the starter module catalyzed by the AT domain of the starter module. The polyketide chain is transferred from the ACP domain of the initiation module to the KS domain of the module I. The elongation group, usually malonyl-CoA or methylmalonyl-CoA, is loaded onto the ACP domain of module I, a reaction catalyzed by the AT domain of module I as shown in Figure 7. The ACP-bound elongation group reacts in a thio-Claisen condensation with the KS-bound polyketide

chain under $CO_2$ evolution, leaving a free KS domain and an ACP-bound elongated polyketide chain. The reaction takes place at the desired number (n) of $KS_n$-bound end of the chain, so that the chain moves out one position and the elongated group becomes the new bound group. The fragment of the polyketide chain can be altered stepwise by the action of additional domains. The KR domain reduces the β-keto group to a β-hydroxy group, the DH domain eliminates water and the resulting α-β-unsaturated alkene can be further reduced by the ER domain. It is important to note that these modification domains actually affect the previous addition to the chain, not the component recruited to the ACP domain of the module containing the modification domain. This cycle is repeated for each elongation module with different combination of the domains.

The TE domain hydrolyzes or cyclizes the completed polyketide chain from the ACP-domain of the previous module.

*Non-Ribosomal Peptides*. Another class of peptide secondary metabolites produced by microorganisms is the nonribosomal peptides (NRPs). NRPs are synthesized by one or more nonribosomal peptide synthetases (NRPS) enzymes which are independent of messenger RNA. NRPs are often dimers and trimers of identical sequences chained, cyclized, or even branched structures (*42*). NRPs are a very diverse family of natural products with an extremely broad range of biological activities and pharmacological properties. Many antibiotics (vancomycin, bacitracin, actinomicin and etc) and immunosuppressants (cyclosporine A) are produced from nonribosomal peptides. The modular structure of the bacitracin synthetase complex is shown in Figure 8 (*43*).

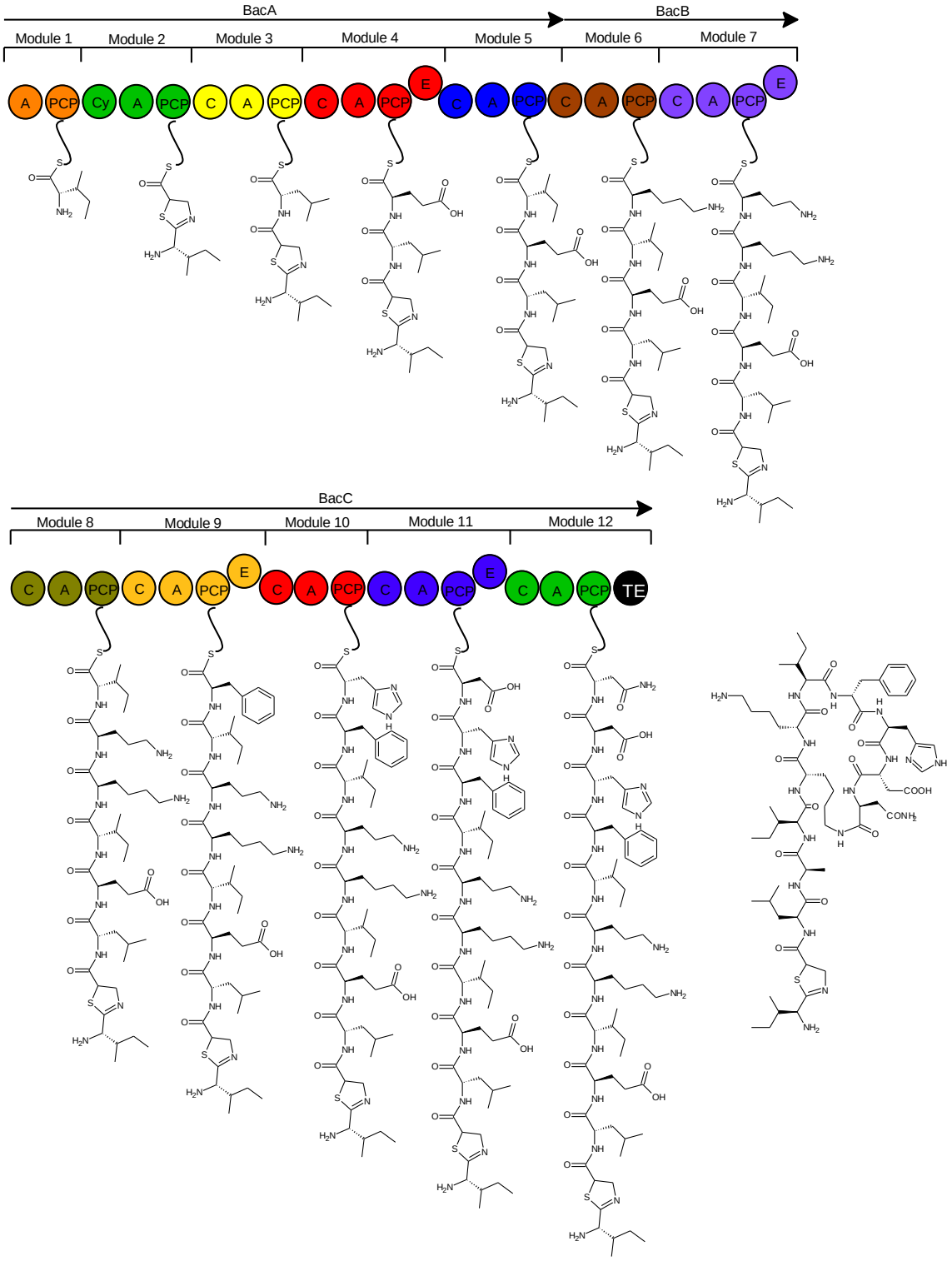



Figure 8: Modular structure of the bacitracin synthetase complex. After Schwarzer, D., *et al*.

The make up of an NRPS consists of initiation, elongation, and termination modules. The initiation module contains an adenylation domain (A) and thiolation domain (T or PCP). The T domain, also called peptide carrier protein (PCP), is required in a module for thiolation. Elongation modules contain condensation domains (C) in addition to A and T domains. Termination modules typically contain a thioesterase domain (TE). Besides required domains (*42*), many optional domains are available; F (formylation), Ox (Oxidation), Red (Reduction of thiaolines or oxazolines to thiazolidines or oxazolidines), NMT (*N*-methylation), R (Reduction to terminal aldehyde or alcohol), Cy (Cylization) and E (epimerization domain). The biosynthesis of NRPs shares characteristics with fatty acid biosynthesis and polyketide biosynthesis. Due to these structural and mechanistic similarities, some NRPS contain polyketide synthase modules for the insertion of acetate or propionate derived subunits into the peptide chain.

*Terpen.* Terpenes are derived biosynthetically from isoprene units as one of nature's building blocks. Terpenes are multiple linked isoprene units. The isoprene units may be linked "head to tail" or "tail to tail" to form linear chains or they may be arranged to form rings as shown in Figure 9 (*37*).

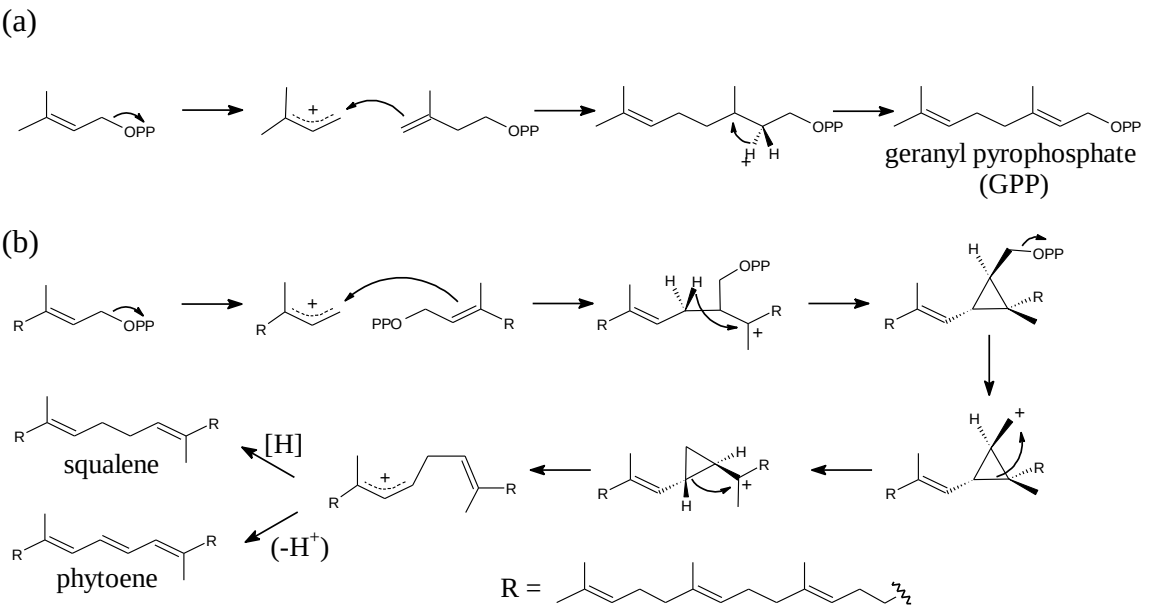



Figure 9: Head-to-tail and tail-to-tail condensation of isoprene units. After Dewick (*37*). (a) Head-to-tail condensation of dimethylallyl pyrophosphate and isopentenyl pyrophosphate to produce geranyl pyrophosphate. (b) Tail-to-tail condensation of geranylgeranyl pyrophosphate to generate squalene and phytoene.

Isoprene ($C_5H_8$) itself does not undergo the building process, but rather activated forms, isopentenyl pyrophosphate (IPP or also isopentenyl diphosphate) and dimethylallyl pyrophosphate (DMAPP or also dimethylallyl diphosphate), are the components in the biosynthetic pathway as shown in the first step of (a) and (b) of Figure 9.

Terpenes are classified sequentially by size as hemiterpenes ($C_5H_{8,}$, isovaleric acid, prenol), monoterpenes ($C_{10}H_{16,}$ geraniol, camphor, limonene, menthol), sesquiterpenes ($C_{15}H_{24}$, patchoulol, pentalenolactone, farnesol, farnesenes), diterpenes ($C_{20}H_{32}$, taxol, retinal, retinol, phytol), sesterterpenes ($C_{25}H_{40}$, geranylfarnesol), triterpenes ($C_{30}H_{48}$, squalene, lanosterol, cycloartenol), tetraterpenes ($C_{40}H_{64}$, acyclic lycopene, the

monocyclic gamma-carotene, the bicyclic α-carotenes, the bicyclic β-carotenes), and polyterpenes which consist of long chains of many isoprene units such as rubber (*44*). The biosynthesis of Lanosterol, a tetracyclic triterpenoid which is a precusor compound from which all steroids are derived, is shown in Figure 10 (*44*).

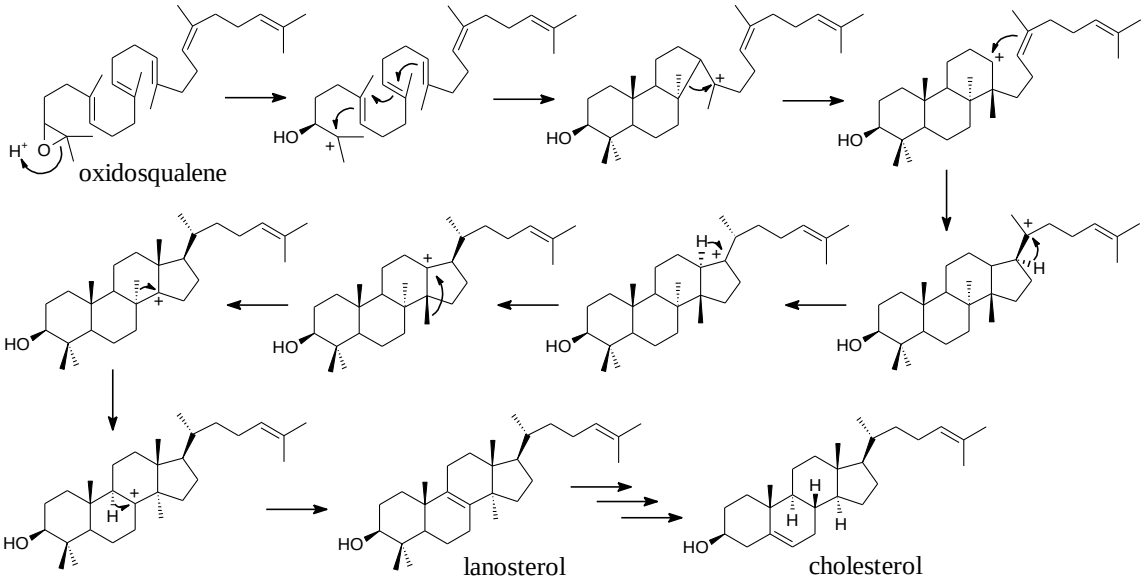



Figure 10: Polyene cyclization cascade of oxidosqualene to produce lanosterol. After Davis, *et al.* (*44*)

*Alkaloids.* Alkaloids are naturally occurring secondary metabolites containing basic nitrogen atoms defined by in International Union of Pure and Applied Chemistry (IUPAC). Many alkaloids are toxic to other organisms and have pharmacological effects so they can be used as medications or recreational drugs. Caffeine, nicotine, cocaine, morphine, and quinine belong to the alkaloid classification. Alkaloids are found in

plants, animals and fungi such as potatoes, tomatoes, shellfish and mushrooms. The biosynthesis of the plant derived alkaloid cocaine is shown in Figure 11 (*45*).

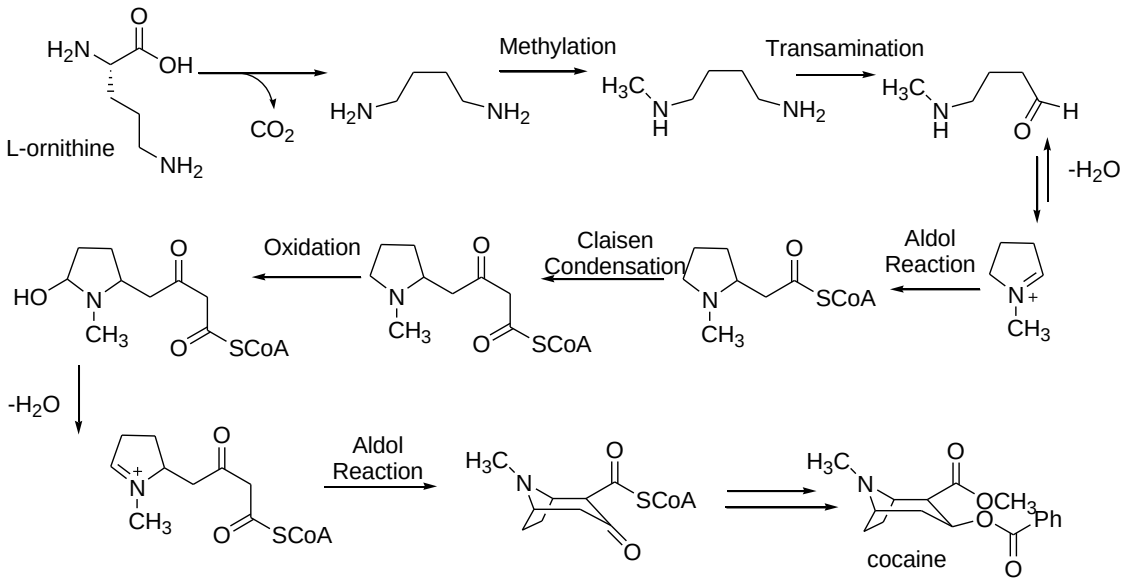



Figure 11: The biosynthesis of the plant derived alkaloid cocaine. The biosynthesis of cocaine from the coca plant, *Erythroxylon coca* .

## METABOLITES TO GENES

*DNA Library.* As already mentioned, only a small portion of microbes are culturable by present methods and it is assumed that 90-99% of the metagenome is not being sampled (*14*, *34*, *46*). Even marine microbial communities, including the cyanobacteria, have not been systematically collected and fully evaluated for bioactive compounds. Instead of developing purification methods or manipulating cultivation conditions, large fragments of non-culturable bacterial genomes can be cloned and expressed using recombinant DNA technology (*47-51*). The main scheme of the technique is that the

DNA of the coding region is extracted, inserted into a vector and this is transfected into a culturable bacterium or other host that expresses biosynthetic genes or clusters.

Construction of recombinant DNA molecules by simply ligating vector DNA and a fragment of interest is a straightforward process as described in Current Protocols in Molecular Biology (*52*). Often, however, molecular biologists do not yet posess the particular DNA they are interested in - they must isolate it from a large pool of DNA molecules. To do this, a modified form of the DNA cloning scheme is utilized, called 'the DNA library'. DNA libraries are mixtures of different restriction enzyme-digested DNA molecules ligated into vectors. Libraries are constructed using the same general approach, except that instead of inserting one type of DNA fragment into the vector, there are thousands or even millions of different DNA fragments inserted into vector molecules. There are essentially two types of DNA libraries, based on the source of the DNA: Genomic DNA libraries made from genomic DNA, all of the DNA found in the organism's nuclei and complementary DNA (cDNA) libraries which are made from DNA copies of mRNA molecules.

Genomic DNA molecules are very large (each chromosome in the nucleus is one such DNA molecule), so they must be fragmented into small enough pieces to insert into vectors. This is typically done through digestion with one or more appropriate restriction endonucleases, mechanical shearing, or a combination of the two processes. The DNA is then ligated into the vector, which could be a plasmid, but is more often a cosmid or a viral chromosome.

To generate cDNA, mRNA is isolated from a tissue or whole organism, and DNA is copied from the mRNA template using reverse transcriptase. This enzyme functions like

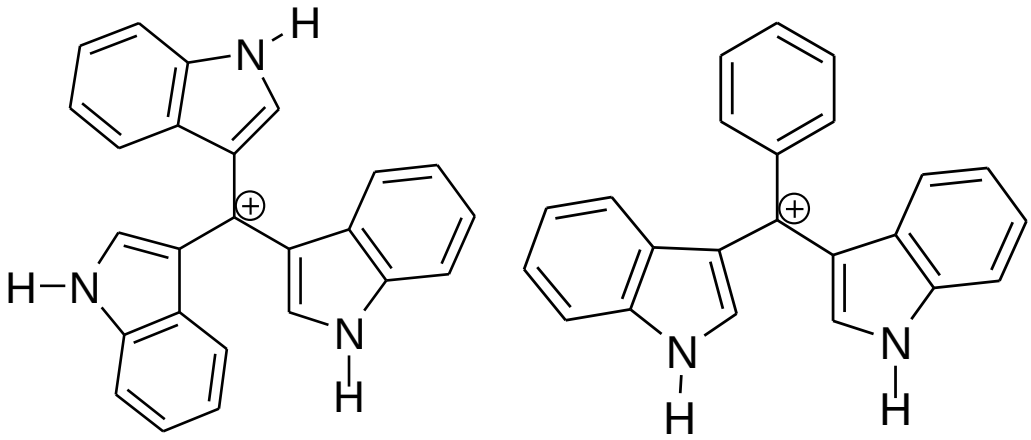
a DNA polymerase, except that is uses RNA as a template instead of DNA. The resulting cDNA molecules are then engineered so that they have restriction enzyme recognition sites at each end of every molecule, which allows them to be digested and inserted into a vector.

The difference between these two libraries is the nature of the DNA found in the library. Therefore, the choice of library type depends on what type of DNA sequence a researcher wishes to study.

*cDNA Libraries.* cDNA libraries are a pool of complementary DNA that are synthesized from the messenger RNA (mRNA) population of an organism. "Directional" cDNA expression libraries (DNA fragments are cloned in one specific direction) will be constructed from marine microorganisms for the purpose of mining biosynthetic genes of a pathway. The primary advantage of the approach is that no prior knowledge of the natural product's structure or origin (e.g. polyketide, terpene, non-ribosomal peptide, or alkaloid subclass identification) is necessary, nor does the method rely upon the host machinery to recognize foreign promoters. While cDNA library expression could support production of all enzymes of a biosynthetic pathway necessary for natural product production, identification of the actual gene sequences is important for future manipulation of the pathway genes in the production of second-generation compounds with modified structures and enhanced activities or simply for the improvement of natural product yields. In this regard, identification of a single biosynthetic gene by our cDNA library approach will be sufficient for identification of the remainder of the gene

cluster (except in plants, most biosynthetic genes have been found to be clustered) by the screening of a genomic fosmid or a bacterial artificial chromosome (BAC) library (*53*).

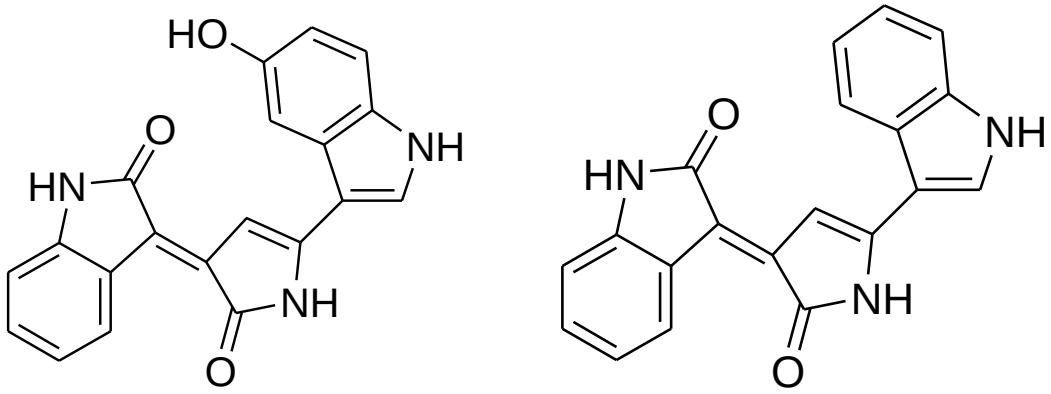
*Fosmid or BAC Libraries.* Fosmid and BAC libraries both allow for the cloning of large segments of DNA. BAC vectors are capable of carrying insert sizes ~100 kb in size while fosmid clones house fragments within the 30 - 50 kb range (*54*, *55*). In many instances, their primary use is to fully represent the genomic DNA of an organism. In this regard, BAC libraries of genomic DNA have been constructed from a variety of plant, animal, and fungal species and have been used to facilitate large-scale sequencing projects (*56*). Both vectors also have, in principle, the means to house an entire biosynthetic cluster(*56*). In this respect, great strides have been made with the randomized cloning of soil bacterial genomes giving rise to natural product expression without further manipulation (primarily through the use of fosmid libraries: (*57-61*). The antibiotics turbomycin A and B (Figure 12) for example, were both isolated and produced from a 24,546-member fosmid library (*61*).



Turbomycin A (Orange) Turbomycin B (Red)

Figure 12: Chemical structures of turbomycin A and turbomycin B.

Similarly, violacein and deoxyviolacein (Figure 13) were also isolated from an environmental fosmid library (*58*). Fosmid and BAC libraries will now be employed for the harnessing of natural product biosynthetic pathways from the marine environment.



Violacein Deoxyviolacein

Figure 13: The structures of violacein and deoxyviolacein.

*Pathway-based Screening: Genetic Selections of the Biotin Biosynthetic Pathway.* Traditional approaches to natural product discovery involve cell-based screening of natural product extracts followed by compound isolation and characterization. Their importance notwithstanding, continued mining leads to depletion of natural resources and the reisolation of previously identified metabolites. Metagenomic strategies aimed at localizing the biosynthetic cluster genes and expressing them in surrogate hosts offers one possible alternative. A fundamental question that naturally arises when pursuing such a strategy is, how large must the genomic library be to effectively represent the genome of an organism(s) and the biosynthetic gene clusters they harbor? Such an issue is certainly augmented in the absence of expensive robotics to expedite colony picking and/or screening of clones. The Watanabe group has developed an algorithm, named BPC (biosynthetic pathway coverage), supported by molecular simulations to deduce the number of BAC clones required to achieve proper coverage of the genome and their respective biosynthetic pathways. The strategy has been applied to the construction of a large-insert BAC library from a marine microorganism, Hon6 (isolated from Honokohau, Maui) thought to represent a new species. The genomic library is constructed with a BAC yeast shuttle vector pClasper *lacZ* paving the way for the culturing of libraries in both prokaryotic and eukaryotic hosts. Flow cytometric methods are utilized to estimate the genome size of the organism and BPC implemented to assess P-coverage or percent coverage. A genetic selection strategy is illustrated, applications of which could expedite screening efforts in the identification and localization of biosynthetic pathways from marine microbial consortia, offering a powerful complement

to genome sequencing and degenerate probe strategies. Implementing this approach, we report on the biotin biosynthetic pathway from the marine microorganism Hon6.

*Modulation of Gene Expression by Riboswitches*. Natural product biosynthetic pathways requiring further genetic manipulation to achieve enzymatic expression will be modified utilizing two independent but complimentary approaches. In the first strategy, we will put genes under the control of a riboswitch. Such an approach is attractive as *E. coli* could be utilized as a heterologous expression host, which eases genetic manipulations and future biochemical studies, where high levels of protein expression might be achieved.

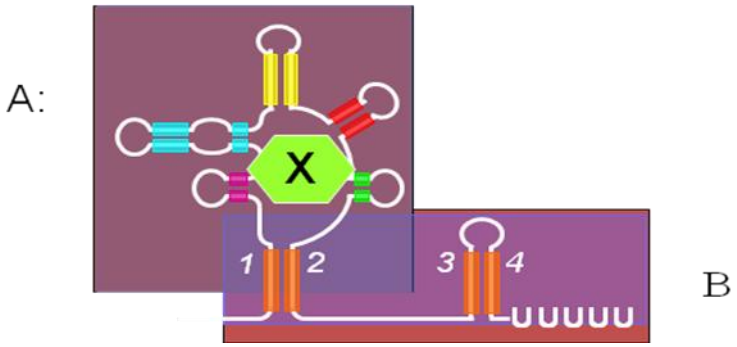



Figure 14: Riboswitch domains. A: a metabolite-binding 'aptamer' domain B: an 'expressionplateform' for gene regulation. These two activities are accomplished by two functionally separate domains on the RNA.

Riboswitches are metabolite-sensing RNA elements that possess the remarkable ability of controlling gene expression. A variety of riboswitches have been identified in nature responding to such metabolites as purines (adenine and guanine), glucosamine-6-phosphate, as well as the co-factors TPP (thiamine pyrophosphate), FMN (flavin

mononucleotide), SAM (S-adenosylmethionine), and B12 (*62-72*). Each riboswitch is typically comprised of an aptamer domain and an expression platform that either activates or disrupts gene expression transduced through ligand binding that causes allosteric changes in both the aptamer and expression domains (*66, 68*). Because of the modular nature of RNA structures, different types of expression platform can be linked to the conserved aptamer domain and leads to variations in the riboswitch mechanism. The structure is shown in figure 15 (*73*). Riboswitches are frequently located within the 5'-untranslated region (UTR) of mRNAs and interface with transcriptional and translational molecular machinery to modulate gene expression (*68*). Like proteins, riboswitches also demonstrate high specificity and can discriminate against other small molecules that display subtle variations on the natural ligand.

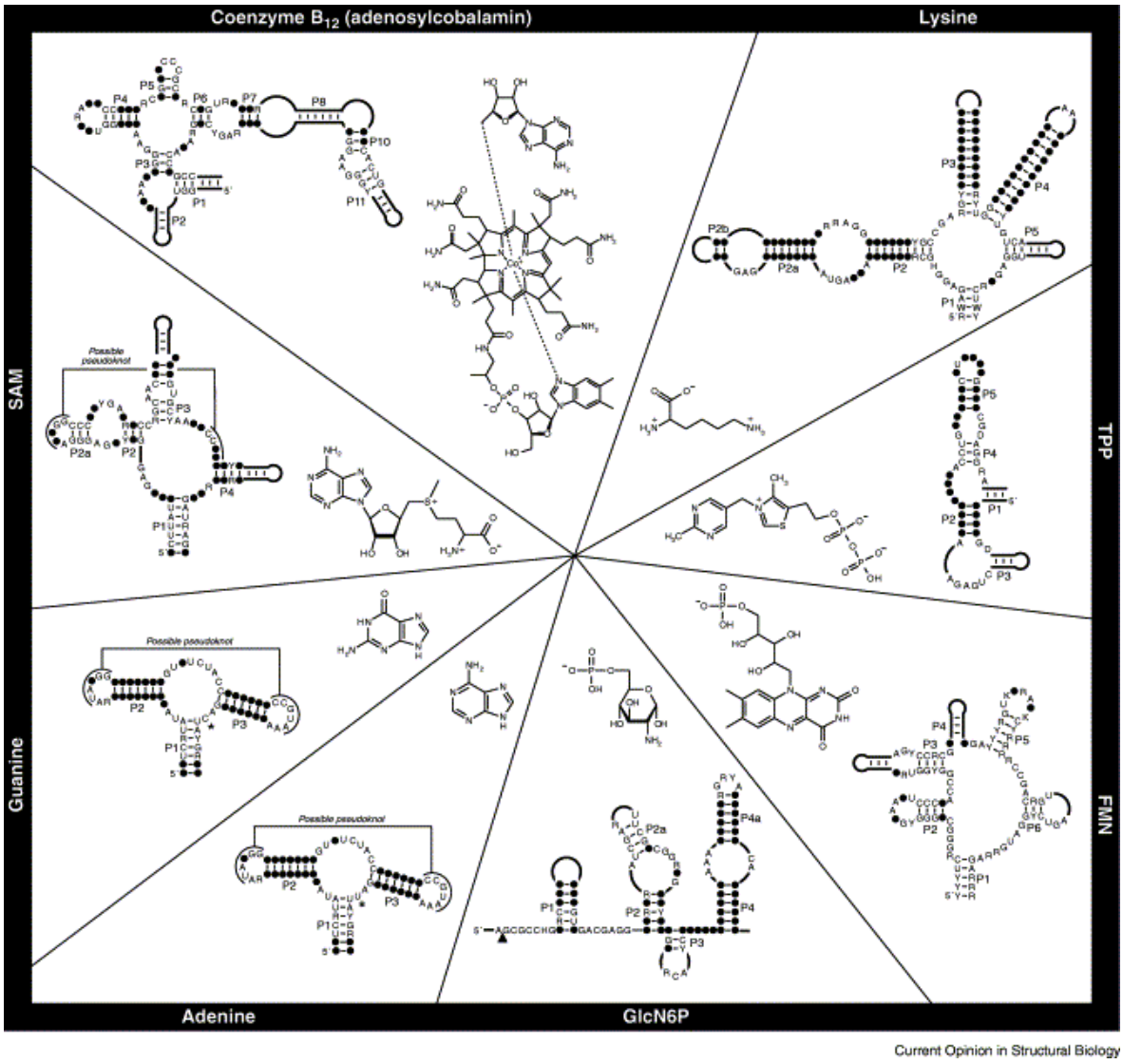



Figure 15: The known ribowitch aptamers. After Soukup and Soukup (*73*).

Crystal structures have been reported on a variety of riboswitch systems including the purine riboswitch, the TPP riboswitch, and the SAM riboswitch (*74*) as shown in figure 15. The guanine and adenine variants of the purine sensing riboswitch consist of three paried regions (P1, P2, and P3), which are arranged as an inverted 'h' forming three helices. The purine ligand binds via hydrogen bonding interactions at the three-way

junction of the riboswitch. The TPP riboswitch resembles that of the purine riboswitch, with an inverted h-architecture, but the helices are more tightly compressed. Docking of the TPP ligand bridges the two parallel helices as opposed to burying within the helical junction. In contrast to both of these systems, the SAM riboswitch adopts a completely different structural motif. The riboswitch consists of a pseudoknot coupled to a kink-turn where SAM binds at the interface of the minor grooves of P1 and P3(*75*). In at least all of these cases, riboswitch regulated gene modulation results from sequestration of a switch segment into either a metabolite-bound or a metabolite-free conformational fold. The model of the coupling between TPP binding and structural rearrangements in the TPP riboswitch is shown in Figure 16. Transcription is completed when metabolite is limited, while metabolite is abundant, transcription will be terminated via structural rearrangement. The metabolite is thiamine pyrophosphate (TPP) in TPP riboswitch.

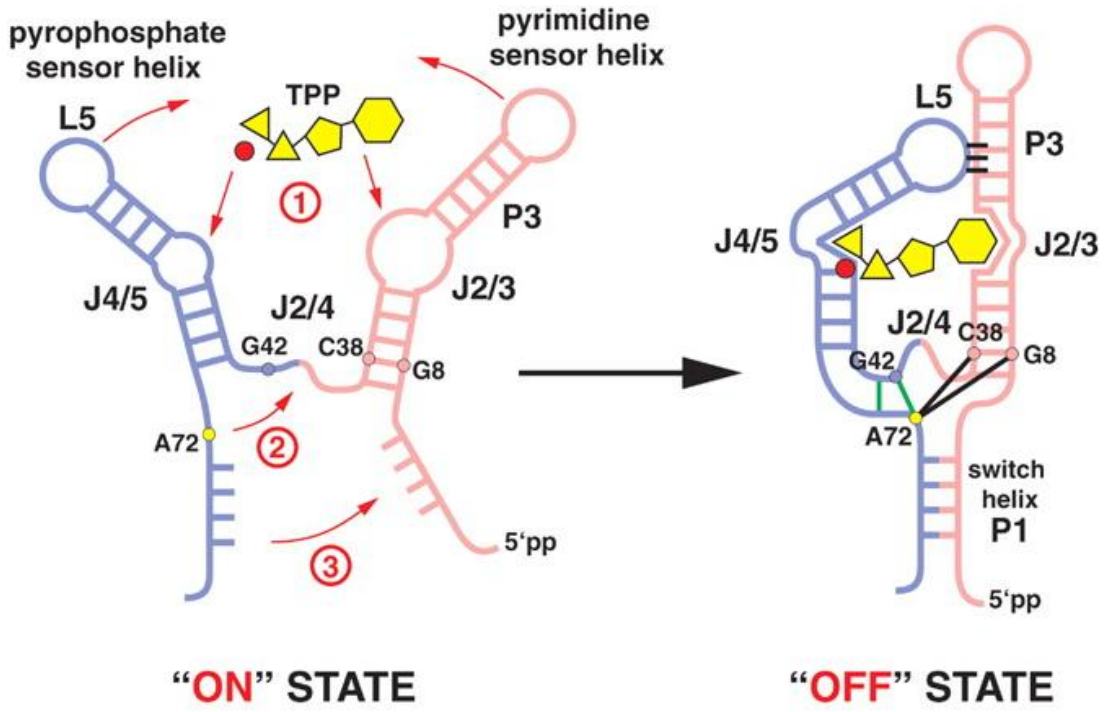



Figure 16: Model of the coupling between TPP binding and structural rearrangements in the TPP riboswitch. Upon TPP binding (1), bulges J2/3 and J4/5 rearrange, thereby closing the sensor helix clamp and forming new hydrogen bonds at the tip of loop L5. The TPP-induced parallel positioning of the sensor helices stabilizes the three-way junction with A72 at the core of the interaction platform (2) and promotes the formation of the switch helix P1 (3), which turns the riboswitch "off." After Thore, S. *et al.* (*76*).

Traditional methods for the identification of natural products are: isolation and purification, characterization, optimization of culture conditions to maximize metabolite production, direct inquiry of the pathway with isotopically labeled precursors, fishing experiment with protein extracts from DNA libraries using Ni-NTA column, sequencing with degenerate DNA primers such as NRPS, PKS, and etc, construction and screening of the genomic libraries including BAC, FOS or COS and cDNA libraries, genetic selection (pathway-based screeing) and commercial organism sequencing which is

available now due to low cost for genome sequencing. In the watanabe group, most of these methods have been in use for the identification of natural products or genes of biosynthetic pathways.

## STATEMENT OF PURPOSE

The overall aim of my research is to develop new strategies in the localization of natural product biosynthetic pathways and achieving heterologous expression.

The specific aims of my research are: (1) creation of cDNA expression libraries from pup14A/B and Hon6 for the identification of the natural product(s) biosynthetic genes; (2) construction of a fosmid library from environmental samples; (3) genetic selection of the biosynthetic pathways; (4) genetic manipulation of the production of natural product; (5) modulation of gene expression by ThiC/riboswitch promoter

# CHAPTER II

# CONSTRUCTION AND ANALYSIS OF COMPLEMENTARY DNA LIBRARIES OF HETEROGENEOUS MARINE ORGANISMS AND *CHROMOHALOBACTER SP.*

## INTRODUCTION

There are two general sources of genetic information for the construction of expression libraries of an organism; genomic DNA and messenser RNA (mRNA). The genomic content is the same in all tissues, so it does not matter which tissue is used to isolate the genomic information. However, the coding region for a gene of interest may be interrupted by one or more intron regions, and thus the complete coding region could be quite long. Introns are spliced out of mRNA which thus contains a contiguous coding region. Another advantage of mRNA is that tissue specific expression of the protein of interest may allow isolation of the appropriate mRNA at enhanced levels. Bacterial Artificial Chromosome (BAC) libraries can be constructed from genomic DNA (*77*), and complementary DNA (cDNA) libraries from mRNA in DNA form.

A general overview for the cDNA approach is as follows. Isolated mRNA serves as the template for reverse transcriptase, which synthesizes a strand of cDNA. The mRNA template is then destroyed, leaving a single–stranded cDNA molecule. The cDNA is then used as a template by DNA polymerase to synthesize a complementary strand. The net result is the formation of a double - stranded DNA molecule that contains the same

genetic information as the mRNA. In effect, a gene has been made using the information from the mRNA.

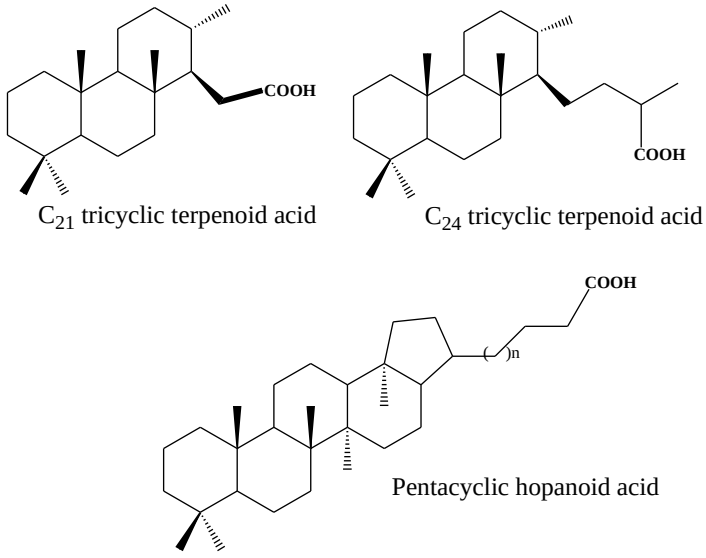
Chapter II will focus on the construction and use of cDNA expression libraries for the identification of the biosynthetic genes of a pathway. Construction of a BAC library was accomplished by one of the Watanabe members, Scott Angell. BAC libraries enable the cloning of large DNA fragments (≥100 kb) or gene clusters but relies very heavily on the host to recognize foreign promoters. Conversely, cDNA libraries express only a single to a few genes at a time, but involve the use of artificial promoters, for example, promoters that are specifically designed for that host strain. In other words, the method does not rely on the host machinery to recognize a foreign promoter. So, we can use a commercial vector which has its own promoter recognized by *E. coli.*

The primary advantage of a cDNA library approach for expressing genes for synthesis of natural products is that does not necessitate prior knowledge of the natural product's structure or biogenetic origin, for example, fatty acids, polyketide, non-ribosomal peptide, terpene, or alkaloid subclass identification. Indeed, to know what the actual gene sequences are is important for future applications. We can manipulate the sequences and modify the structures for enhancing the bioactivities or improvement of natural product production yield. In this regard, our cDNA library approach will give sufficient information for identification of a single biosynthetic gene and the remainder of the gene cluster by the screening of a genomic BAC library.

The appeal of the approach is the ability to identify natural product genes of a pathway without having prior knowledge of the biosynthetic origins of the metabolites. Moreover, as it is estimated that less than 1% of marine microorganisms can be

cultivated in the laboratory, (*33*, *34*) these methodologies might offer the means to access previously untapped resources.

Once the cDNA libraries were successfully generated, they were not only screened for the natural product biosynthetic pathway genes they harbor, but also for metabolic enzymes that could "neutralize" the deleterious effects of acid crude oil (*78-81*). Crude oil corrosivity is a central issue that plagues the oil refinement industry presenting a significant economic burden and challenge. Acid crude oils are found in many parts of the world resulting from microbial decomposition within the reservoir. While the resulting crude oils are cheaper, they are generally more viscous and high in naphthenic or terpenoid acids (Figure 17).




$C_{21}$ tricyclic terpenoid acid

$C_{24}$ tricyclic terpenoid acid

Pentacyclic hopanoid acid

Figure 17: Examples of naphthenic acids.

The compounds react with metal surfaces leading to corrosion of refinery pipework and vessels. Given this problem, we aim to identify biosynthetic or metabolic enzymes

from the marine environment that could be employed to degrade these acids (*78-81*). To complement these experiments, offering an orthogonal approach, a fosmid library of environmental DNA will be generated and diversely screened for natural product production.

## EXPERIMENTAL PROCEDURES

*General*. The SuperScriptTM Choice System for cDNA Synthesis kit was obtained from Invitrogen life technologies. E.coli cells were routinely grown in LB medium containing 1% Bacto Tryptone (Difco, Sparks, MD), 0.5% yeast extract (Difco, Sparks, MD), and 0.5% NaCl with antibiotics at 50μg/ml for ampicillin. RNeasy Mini Kit, Quiapep Spin Mini Kit, QIAGEN plasmid kit, QIAquick Nucleotide Removal kit, and QIAquick Gel Extraction Kit were obtained from QIAGEN Inc (Valencia, CA). Restriction enzymes and alkaline phosphatase enzymes were obtained from New England Biolabs (Ipswich, MA).

*Organisms.* All marine microorganisms utilized in these experiments were obtained from field collections by Watanabe's group (Table 2) (*82*). Hon6 (isolate 6) was acquired from Honokohau, Maui, Pup16 (isolate 16) from Pupukea, Oahu. Pup14A and Pup14B were also obtained from Pupukea (sample 14). The identity of each of the microorganisms was assigned by the Centralbureau voor Schimmelcultures (CBS) in the Netherlands. *E. coli* Zappers was obtained from Novagen (Gibbstown, NJ 08027), and

EC100 was from Epicentre (Madison, WI). pETBlue-2 was obtained from Novagen (Gibbstown, NJ 08027), and pYES2/CT/LacZ from Invitrogen (Carlsbad, CA)

Table 2: Bacterial strains and plasmids used in Chapter II

| Strain or Plasmid | Description [a] | Reference/Source |
|---|---|---|
| Strains | | |
| EC100 | Host strain for DNA manipulations | Epicentre |
| NovaXG™ Zappers | Host strain for DNA manipulations | Novagen |
| Pup14 | Marine microbial assemblage isolated in Pupukea, HI | This study (*82*) |
| Pup14A | Identified as Pseudomonas aeruginosa by CBS | This study (*82*) |
| Pup14B | Identified as Enterobacter sp. by CBS | This study (*82*) |
| Hon6 | Identified as Chromohalobacter sp. by CBS | This study (*82*) |
| Pup16 | Identified as Rhodotorula mucilaginosa by CBS | This study |
| Plasmids | | |
| pETBlue-2 | $Amp^r$ | Novagen |
| pYES2/CT/LacZ | $Amp^r$ | Invitrogen |

*Field Collection of Marine Microbial Samples.* This work was done by Dr. Scott Angell, a former member of the Watanabe group (*82*). Marine agar plates were prepared with slight modification to a standard literature protocol (*83*). Yeast extract (2 g), bactopeptone (2 g), glucose (4 g), and bacto-agar (Difco, 20 g) were combined in 1 liter of seawater, prepared by mixing Instant Ocean (38 g, Petco) with deionized distilled water. The marine agar plates were further supplemented with either 10 mg/L of miconazole or a combination of chloramphenicol (10 mg/L) and kanamycin (10 mg/L) for the culture of bacteria or fungi, respectively. Culture swabs (Becton Dickinson, supplier VWR International; West Chester, PA) were used to swab submerged rocks, and autoclaved vials were used to collect ocean floor sediments. Samples were applied

to both types of marine agar plates and incubated at room temperature for one week. Single colonies were generated by repeated streaking of microorganisms onto fresh plates. To generate marine microbial populations, cell scrapers (Costar, West Chester, PA) were applied to the original plates, and the cells resuspended and cultured in marine broth.

*Preparation of Total RNA from Marine Sources.* Pup 14 cultures were prepared by inoculating the salt-based medium with equal amounts of Pup14A and Pup14B. Following 72 hours of growth and induction of the blue phenotype, the cells were harvested by centrifugation and the cell pellets subsequently flash frozen in liquid nitrogen. Total RNA was extracted by hot phenol extraction and purified with a Qiagen RNeasy Kit as described in reference (*84*). Hon6 and Pup16 were also prepared with a traditional method (*83*). Hon6 glycerol stocks streaked on a marine agar plate as shown in Figure 18. Hon6 was inoculated into Marine broth containing bacto pepetone 1g, bacto Yeast Extract 1g, glucose 2g, and sea salt 19g (*83*).

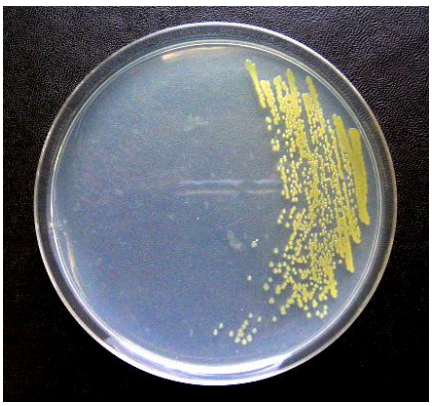


Figure 18: Photograph of the Hon6 strain cultured on a marine agar plate.

*Preparation of DNA Samples*. DNA samples of pETBlue-2 and pYES2/CT/LacZ were prepared by transformation into EC100 electrocompetent cells. After growing on agar plates containing ampicillin, they were inoculated in media for amplification. Plasmid purification was performed with a Qiagen plasmid purification kit followed by gel extraction to get the desired DNA. pYES2/CT/LacZ was sequentially digested with XhoI and Hind III and pETBlue-2 with EcoRI and HindIII respectively. Dephosphorylation was performed with Shrimp Alkaline Phosphatase (SAP) or Calf intestinal Alkaline Phosphatase (CIAP) after gel extraction of double digested DNA.

*Preparation of the Random Hexamers for Priming First Strand Synthesis*. Dr. Watanabe ordered random hexamer oligonucleotides from SIGMA GENOSYS (The Woodlands, TX)

For pYES2/CT/LacZ (HindIII/XhoI) system (used with Pup16)

DAG HEXAMER: AGNNNNNN

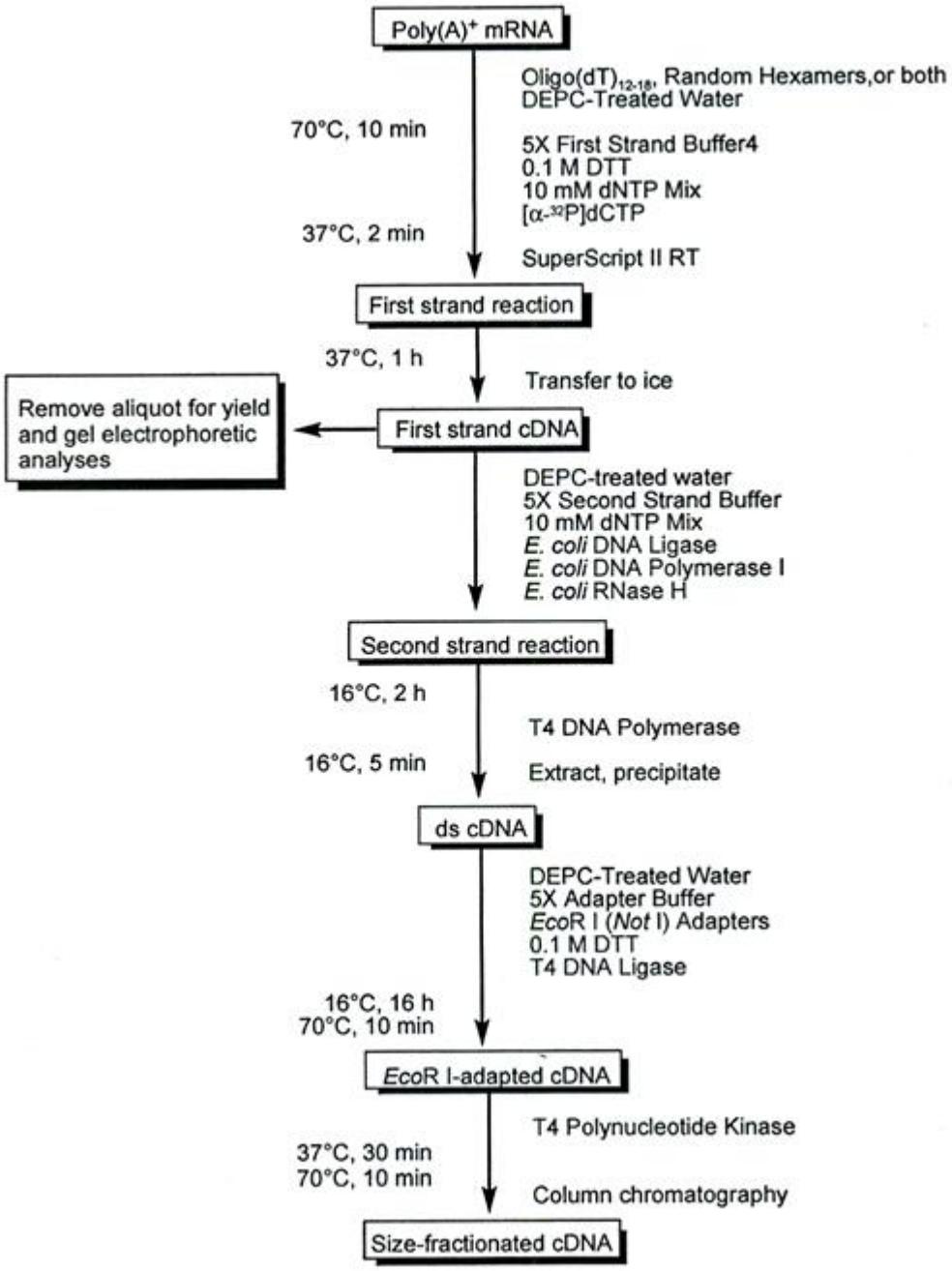
HindIII XhoI ARMS: CGAGAAGCTTCTCG

For pETBlue-2(EcoRI/HindIII) system (used with Hon 6 and Pup14)

DTT HEXAMER: TTNNNNNN

EcoRI HindIII ARMS: GCTTGAATTCAAGC

cDNA libraies were prepared with a slight modification to the standard Invitrogen instruction manual of the SuperScriptTM Choice System for cDNA Synthesis kit. Detailed protocol flow diagram shown in Scheme 2 (*85*).




Scheme 2: Detailed protocol flow diagram of cDNA synthesis (after Invitrogen manual p13) (*85*).

*Preparation of the First Strand cDNA to Optimize First Strand Primer Concentration.* 10µg of Hon6 mRNA was mixed with either10ng, 25ng, 50ng, 100ng, or 150ng of DTT Hexamer oligonucleotide, and then heated at 70°C for 10min. After incubation on ice for 5min, a quick spin was done to get the samples to the bottom of the tubes. First strand reaction mixture (4µl 5x first strand buffer, 2µl 0.1M DTT, 1µl 10mM dNTP, and 1µl [α-$^{32}$P] dCTP (1µCi/µl)) was added to each tube and mixed. After 2 min at 37°C, 1 µl superscript$^{TM}$ II RT (reverse transcriptase) was added to each tube and mixed well. A 1 hr incubation at 37°C was performed after a quick centrifugation. After analysis of the reaction by agarose gel electrophoresis and phosphoimaging, optimal conditions for the first strand reaction were used for further reactions. Samples of Pup14 and Pup16 were processed indentically to Hon6 except the DAG Hexamer oligonucleotide was substituted for the DTT Hexamer oligonucleotide in Pup16. After first strand synthesis, a nucleotide removal kit was used to purify the first strand cDNA and to remove $^{32}$P. After running the gel, the dried gel was developed with Phosphorimage plate for 2 hrs to overnight depending on the intensity of the signal to optimize the concentration random Hexamer oligonucleotides used in the reactions. Once optimized conditions were determined, second strand synthesis was followed.

*Preparation of the Double-Strand cDNA via Second Strand Reaction.* For second strand synthesis, the following reagents were added on ice to the first-strand reaction tube in the order shown: 89µl DEPC-treated water, 30µl 5x second-strand sysnthesis buffer, 3µl dCTP mixture ($CH_3$-dCTP:dCTP, used 1:0, 1:1, 1:3, 3:1), 1µl [γ-$^{32}$P] dCTP (1 µCi/ µl), 1µl *E. coli* DNA ligase (10units/µl), 4µl *E. coli* DNA polymerase I

(10units/µl), and 1µl *E. coli* RNase H (2units/µl). All tubes were gently vortexed to mix and then incubated at 16°C for 2 hours. After that 2µl (10units) of T4 DNA polymerase was added and incubation was continued for an additional 5 min at 16°C. The same procedure was performed to optimize the dCTP concentration by analysis with phophorimage reader. The phosphorimage is attached in Appendix A.

*EcoRI HindIII Adapter Addition.* First-strand and second-strand synthesis with optimized conditions was repeated, and the cDNA extracted with phenol: chloroform: isoamyl alcohol (25: 24: 1). Prepared cDNA was dried at 37°C for 5 min. The following reagents were added to the cDNA: 18µl DEPC-treated water, 10µl 5x adapter buffer, 10µl EcoRI HindIII adapters, 7µl 0.1M DTT, and 5µl T4 DNA ligase. The Adapter addition reaction was incubated at 16°C for a minimum of 16 hrs. After precipitation with EtOH, sequential digestion with HindIII followed by EcoRI was performed and the cDNA precipitated again.

*Column Chromatography.* A scintillation counter (facility of Biology Department at Texas A&M University) was used for checking specific activity after running a cDNA fractional column, which was provided with kit. All procedures were performed as described in the manual (*85*). Each fraction was evaluated for Cerenkov Counts (CPM) by Scintillation counter. The amount of double strand cDNA was calculated with equation (1) after inserting the SA value calculated with equation (2).

$$Amount\ of\ ds\ cDNA\ (ng) = \frac{(\quad)CPM \times (50\mu l/10\mu l) \times (20\mu l/2\mu l) \times (4pmol\ dNTP/pmol\ \ dCTP)}{SA\ (CPM/pmol\ dCTP) \times (3030\ pmol\ dNTP/\mu g\ cDNA)} \qquad (eq.1)$$

$$SA\ (CPM/pmol\ dCTP) = \frac{(\quad)CPM/10\mu l}{200pmol\ dCTP/10\mu l} \qquad (eq.2)$$

*Ligation of cDNA.* After column chromatography, the desired amount of double strand cDNA was precipitated with EtOH. The ligation reaction consisted of the following: 4µl 5x T4 DNA ligase buffer, 2µl plasmid (dephosphorylated), 8µl cDNA, 5µl DEPC-treated Water, and 1µl T4 DNA ligase. Dephosphorylated pET/HindIII/EcoRI plasmid was used for Pup14 and Hon6 and dephosphorylated PYES2/HindIII/XhoI plasmid for Pup16. The reaction was performed at 4°C overnight. Electroporation of cells was carried out with a BioRad MicroPulser. Electroporation for the transformation into competent cells (NovaXG™ Zappers and EC100) was performed after desalting the ligation reactions. The cells were plated on LB agar +Ampicilin, and then incubated for 16 hrs at 37°C.

*Analysis of cDNA Library.* Colonies were inculated in 5ml of LB media containing Ampicilin and incubation for 16 hrs at 37 °C. Grown cells were harvested and then purified with QIAprep mini-prep. EcoRI, NotI, XhoI and NcoI restriction enzymes were used for analysis of cDNA libraries to check insertion size.

*Screening of cDNA Library.* We calculated that the number of colonies we need to collect from the agar plates using equation 3.

$$N = \frac{\ln(1-P)}{\ln(1-\frac{Insertion\ size}{GS})} \qquad (eq.3)$$

GS for Pup14 or Pseudomonas pathway site = 1.96Mb = $1.96\times 10^6$

P = 99% = 0.99

Insertion size = 1.5kb = $1.5 \times 10^3$

Explanation of the derived equation is described in Chapter IV. For Pup14, 10,000 colonies were required while for Hon6 12,000 colonies were required based on the blue and white screening and insertion size. Transformants were collected and stored at -80°C (All 40 tubes for Pup14 and 60 tubes for Hon6 each tube contains 250 colonies).

*Screening of cDNA Libraries for Naphthenic Acid Converting Enzymes.* Using NCBI BLAST search, primers of SAM-dependent methyltransferase and Carboxylic Acid Reductase were designed (*86*). Eight organisms were used to design the primers of Carboxylic Acid Reductase: *Nacardia farcininca* IFM 10152, *Nocardia* sp NRPL 5646, *Mycobacterium avium* subsp, *Paratuberculosis* str. K10, *Mycobacterium tuberculosis* H37Rv, *Mycobacterium bovis* AF2122/97, *Mucobacterium tuberculosis* CDC155, and *Mycobacterium lepraae* TN. Thirteen organisms were used to design the primers of SAM-dependent Carboxylic Methyltrasferase: *Salmonbella enteric* subsp, *Enterica*

*Serovar choleraesuis* str. SC-B67, *Salmonella typhimurium* LT2, *Enterica serovar Typhi* Ty2, *Enterica serovar Typhi* str. CT18, *Enterica serovar Paratyphi* A. str. ATCC9150, *Escherichia coli* K12, *Escherichia coli*, Yjhp from *E. coli*, *Geobacter sulfurrecucens* PCA, *Anaeromyxobacter dehalogenans* 2CP-C, *Geobacter metallireducucens* GS-15, and *Rhodosprillum rubrum*.

## RESULTS AND DISCUSSION

*Isolation of Pup14* (*82*). Samples were collected from the north shore of Oahu, Hawaii. Ocean floor sediments (up to a depth of 40 ft) were obtained as well as culture swabs of submerged rocks. The microorganism specimens were plated on bacterial or fungal plates employing standard literature protocols (*83*). Microbial communities were inoculated into liquid media and cultured in shake flasks. Among the microorganism samples examined, Pup14 (Pupukea sample 14, bacterial plate) gave an unusual green phenotype following ~72 hr of culturing; while the cells remained white, the medium turned green (Figure 19). Extraction of the medium resulted in an aqua blue color. The Pup14 sample showed that cells propagated from single colonies and grown in culture could not generate the green phenotype, suggesting that natural product production was in some way coupled with microbial interactions.

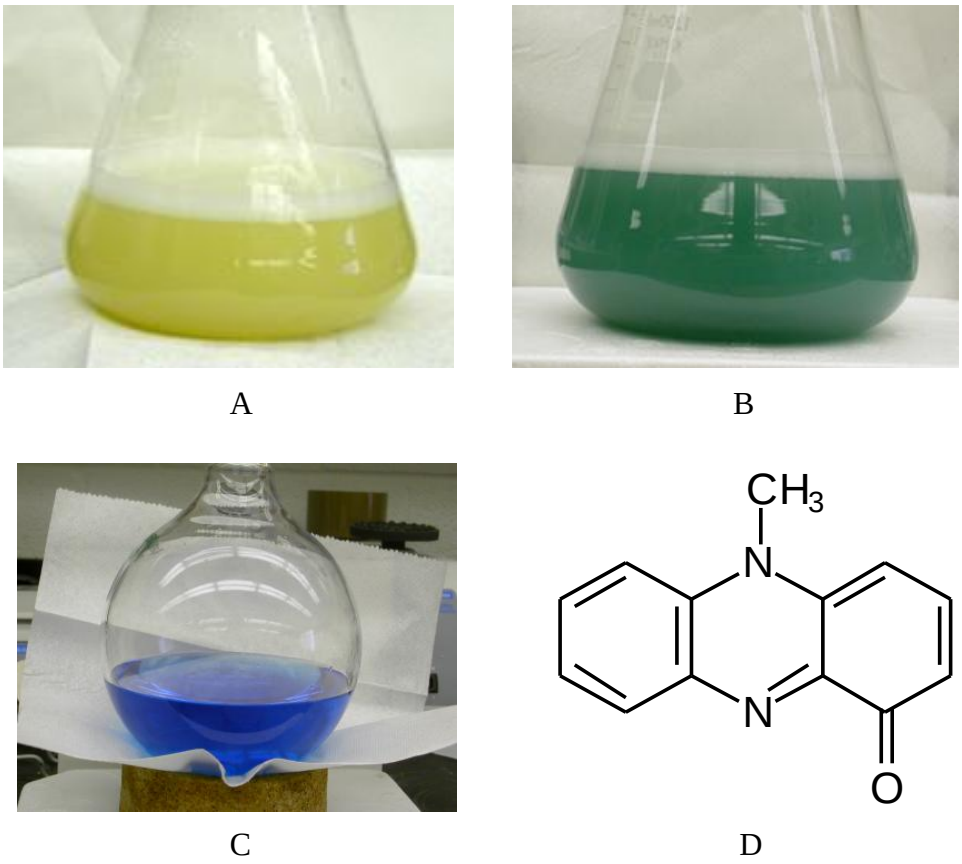
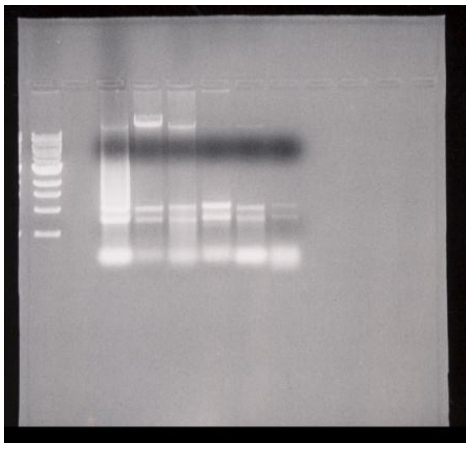



Figure 19: Induction of pyocyanin production. (A) Uninduced co-culture of Pup14A and Pup14B after 24 hours of growth. (B) Induced co-culture after 72 hours of growth. (C) Methylene chloride extract of induced co-culture. (D) Pyocyanin Structure

Pup14 A and B, two marine microorganisms were retrieved from the north shore of Oahu, Hawaii. Pup14 or *Pseudomonas aeruginosa* is a bacterium that was identified through examination of mixed cultures of marine microbial populations for bioactivity. The bacterium was shown to produce a blue metabolite, pyocyanin, when activated with another microorganism, identified as *Enterobacter* (Figure 19D). In 2005, the paper was published that addresses the synergistic role of the microorganisms in the production of the blue metabolite, pyocyanin.

As the techniques employed were quite general, they can be easily applied to address synergism observed with other organisms. Full characterization for the compound is also provided, including missing structural information, 2D-NMR data, X-ray analysis, and DNA microarray analysis, which suggest that the compound promotes oxidative damage and stress (*82*).

*Construction of cDNA Libraries from Marine Microorganisms.* cDNA libraries have been successfully generated from both Hon6 and the synergistic microorganisms Pup14 A/B. The directional libraries were constructed as follows. Total RNA was isolated by hot phenol extraction and subsequently purified with a Qiagen RNeasy Kit (Figure 20) (5).



Samples From Left to Right
[1] 1 kb DNA ladder (NEB)
[2]Yeast total RNA control
[3] Gal 2PB
[4] Gal 1 PC
[5] Pup 16
[6] Hon 17
[7] Hon 1B

Figure 20: Total RNA isolated from a variety of marine bacteria and fungi.

cDNA synthesis is initiated by employing an RNA-dependent DNA polymerase. Since these enzymes are not capable of initiating cDNA synthesis *de novo*, primers were used to which the nucleotides can add. Current priming strategies include the use of poly-dT primers or random hexamers. While oligo-dT primed synthesis should theoretically result in full length cDNA, in practice the yield of such reactions are typically less than 50%, also resulting in significant 3'-end bias (*87*). In contrast, with random priming control over the average cDNA size can be achieved by adjusting primer template ratios. At low ratios, priming events become a lot less frequent and larger cDNA sequences are produced. Therefore, we selected to construct our libraries with random hexamers (Figure 7, optimization of primer concentration with the marine microorganism Pup14). More specifically, directionality of the library could be achieved by using directional random hexamer primers that consist of a random hexamer sequence preceded by two defined bases at the 5'-end (Figure 21) (*88*). Having only two bases at this position allows proper priming while defining the 5'-end of the first strand to allow directional cloning of the cDNA.

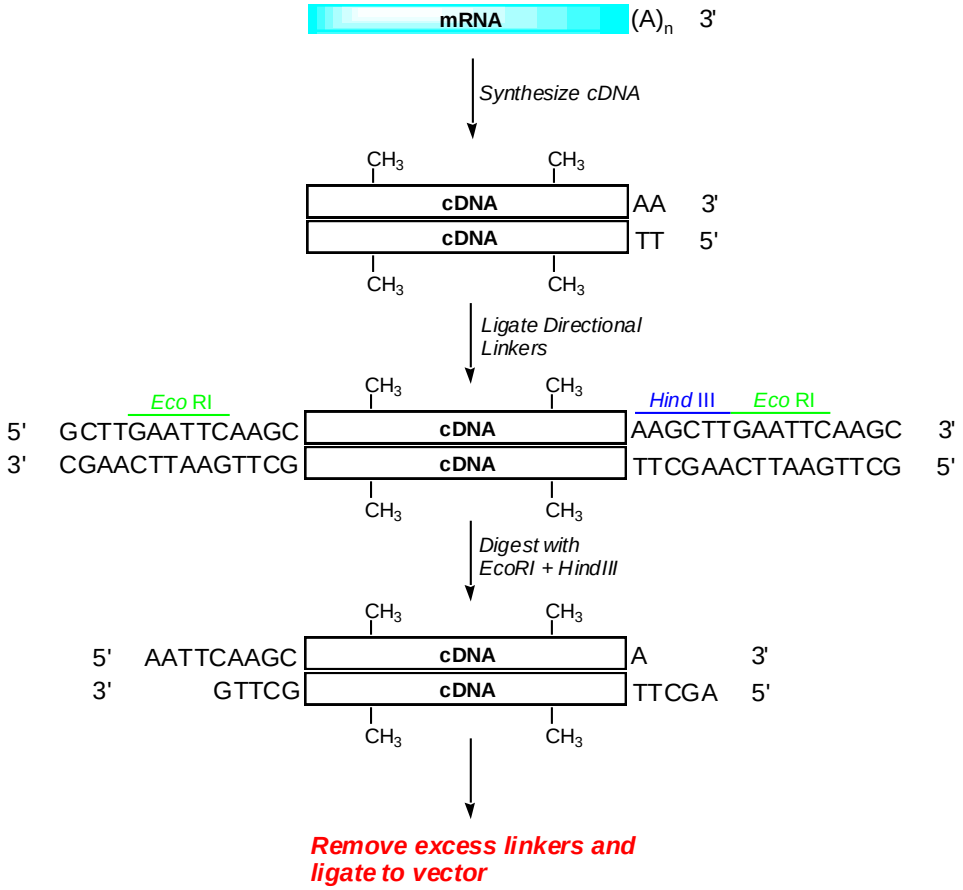



Figure 21: Directional cDNA cloning strategy with tethered random hexamers. Modified from Morris *et al.(88)*.

In the configuration shown here, the two defined bases are d(TT). Directional cloning of the cDNA is achieved upon ligation with an EcoRI/HindIII directional linker [5'd(GCTTGAATTCAAGC)] including an internal EcoR I flanked on each side by four of the six bases defining a Hind III site followed by digestion with EcoRI and HindIII. Since HindIII can only cut the linker if it ligates next to a d(AA):d(TT) dinucleoside base pair, heterogeneous ends are produced and cloning achieved in a unidirectional manner. To protect internal EcoRI and HindIII sites upon digestion, 5-methyl dCTP is

used in both first strand and second strand syntheses and the resulting cDNA clones propagated in a non-methylating host so that their methylation patterns are no longer retained (*88*).

First strand cDNA was obtained after optimization of the random hexamer concentration. 100ng DTT Hexamer was used with 10µg of Pup14 after analysis of the phophrous image as shown in Figure 22. 50ng DTT hexamer was used with 10µg of Hon6 and 50ng DAG hexamer was chosen for Pup16 by a phosphrous image analysis, more data is attached on AppendixA (Figure A.1).

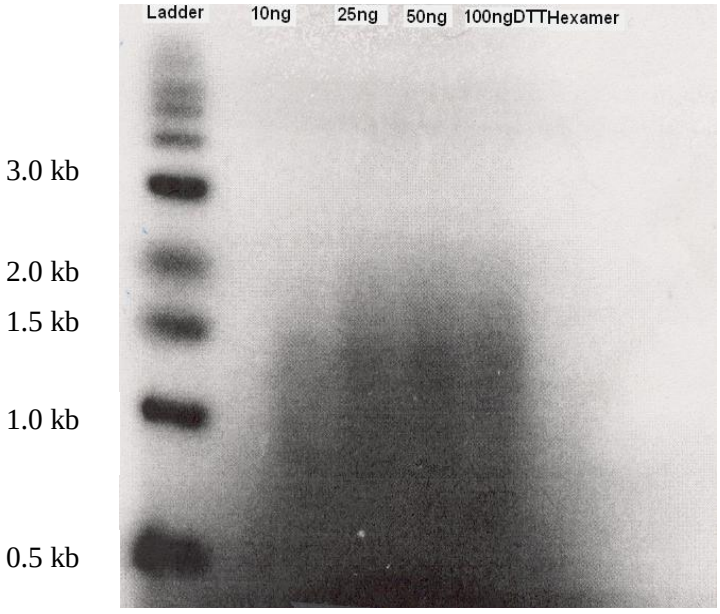
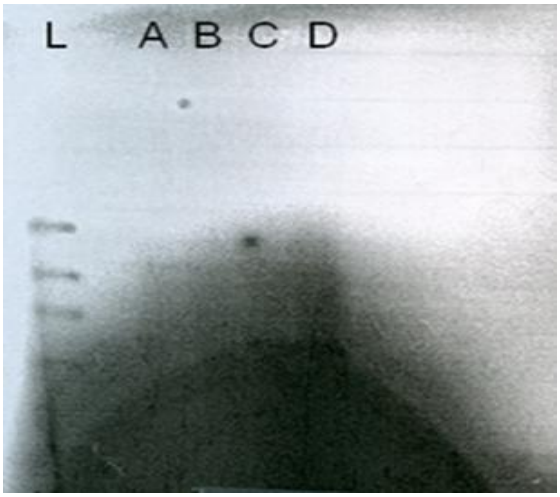



Figure 22: Optimization of cDNA synthesis of Pup14A/B by varying the concentration of the random DTT Hexamer.

Double strand cDNA was prepared via 2$^{nd}$ strand reaction. Second strand reaction was performed with a different ratio of methyl-dCTP to dCTP. After analysis of

phosphrous image, one to one ratio of methyl-dCTP to dCTP was employed for the second strand reaction. (Shown in Figure 23)




| | $CH_3$-dCTP | : | dCTP |
|---|---|---|---|
| A | 1 | : | 3 |
| B | 1 | : | 1 |
| C | 1 | : | 3 |
| D | 3 | : | 1 |

Figure 23: Phosphrous image of second strand cDNA with Pup14A/B.

After fractionating the cDNA using a fractional column, 4 to 5 fractions were combined to get about 10ng of cDNA at a concentration of ≥1ng/µl for use in the ligation reaction. DNA was precipitated following the ligation reaction and transformated into EC100 or NovaXGTM Zappers competent cells as described in the experimental procedure. Base on blue and white screening the first fraction group after column appeared to give the best results. The selected group was subjected to check insertion size with different restriction enzymes. As shown in Figure 24, more than 80% of selected clones of Pup14 A/B and Hon6 cDNA libraries have bigger than 3 kb size insertion after digestion with EcoRI. Pup16 cDNA library showed about 30% clones with an insert larger than 3 kb. More detailed results are attached at Appendix A (Figure A.2).

By the approach described above, a cDNA library was successfully generated from the marine microorganism Pup14A/B and Hon6 (Figure 24, representative clones of the libraries). Generated clones were collected and stored as described in experimental procedures. The resulting cDNA library in *E. coli* strain NovaXGTM Zappers was subdivided into 40 pools of 10,000 clones for Pup14A/B and 60 pools of 15,000 clones for Hon6. Frozen glycerol stocks of each pool also were prepared.

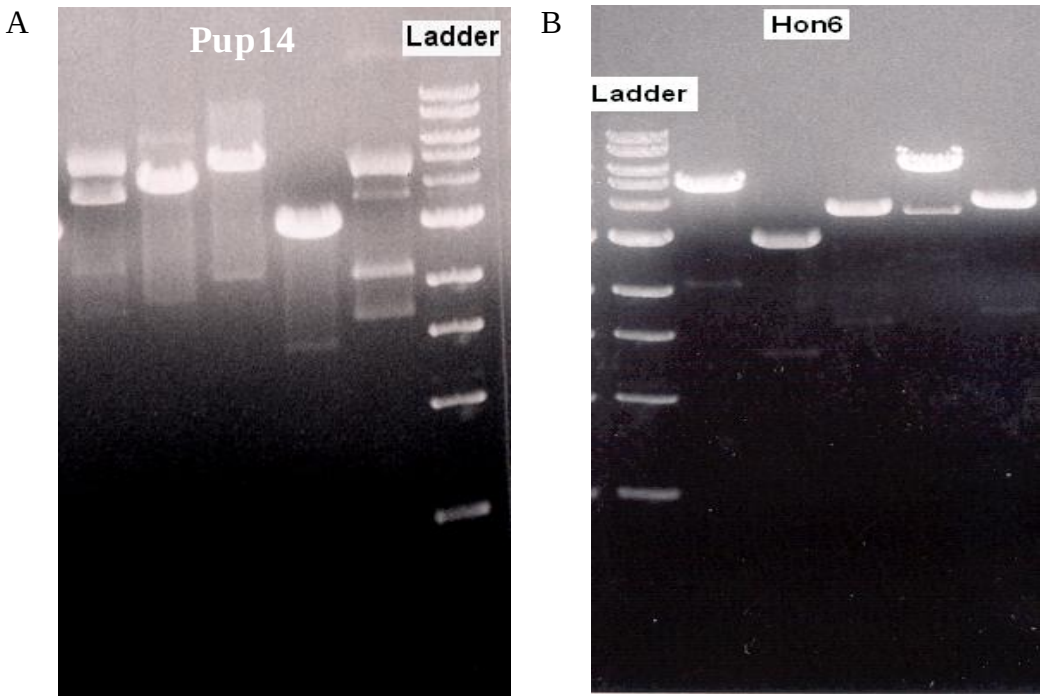



Figure 24: Representative clones from the Pup14 cDNA library (A) and Hon 6 cDNA library (B) cut with the restriction enzyme EcoRI. Each clone contains an insert greater than 3kb size. Ladder is 1kb DNA ladder from NEB. More data as attached on Appendix A.

*Screening of cDNA Libraries for the Identification of Natural Product Biosynthetic Genes.* Directional cDNA expression libraries offer an attractive complement to BAC library expression cloning (overview of approach Figures 20 to 23). Partial purification

of crude protein extracts can offer whole genome representation of proteins. Coupled with a selection strategy involving the use of rare enzyme cutters, the procedure could find general use for bio-prospecting of natural product biosynthetic pathways. A protein extract of the cDNA library was prepared and the expressed proteins purified with a Ni-NTA column. The protein extract was incubated with radio-labeled acetate or radio–labeled S-adenosylmethionine in addition to co-factors (NADPH, SAM, FAD) and other cellular components (prepared by filtration of a crude protein extract from E. coli, Pup14A/B or Hon6 through a stirred cell apparatus. Following incubation, the extract was extracted with organics. The extract was run on the TLC plate in 15 $CH_2Cl_2$: 1 methanol solution and then was exposed on phosphorous image plate. The result was analyzed by phosphorous imager. (Shown in Appendix Figure A.3) This approach failed to yield any intresting results. Another approach is analysis of the extract by HPLC. All peaks will be collected and analyzed for radioactivity with a scintillation counter. Identification of the natural product biosynthetic genes will be achieved by using rare enzyme cutters to deplete the library while not disturbing vector sequence (refer to Figure 25). The depleted library will be re-introduced into *E. coli*. If natural product production is affected, cut sequences can be distinguished from the remainder of the library population by ligating and cloning them into a vector containing a different selectable marker, the DNA will be isolated and sequenced. Finally, the identified sequences could then be used to probe a BAC library of the organism for elucidation of the remainder of the pathway genes. We however realized that it would require significant resources to complete the screening. Therefore, the prepared samples have been stored at -20 °C until more funds are available for sequencing and MS analysis

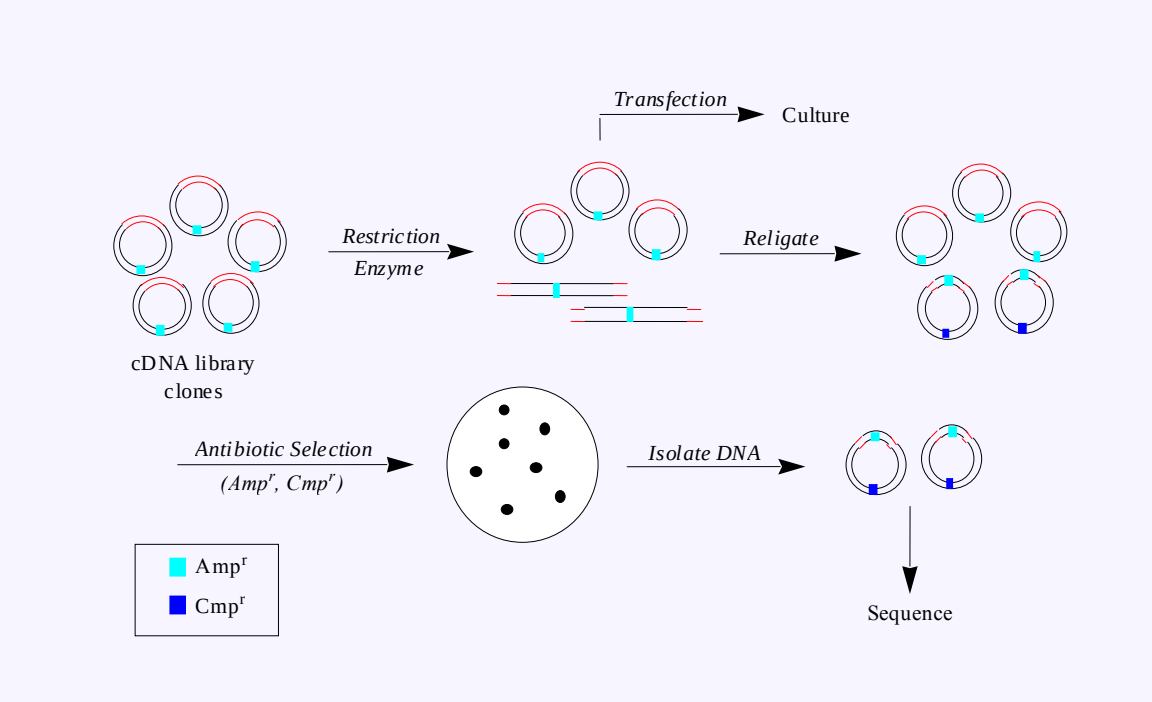



Figure 25: cDNA library depletion approach with rare enzyme cutters.

*Screening of cDNA Libraries for Naphthenic Acid Converting Enzymes.* The cDNA libraries were also screened for metabolic enzymes that can neutralize the effects of naphthenic acids (Figure 17). Degenerate oligonucleotide probes generated through sequence comparisons will be employed to identify S-adenosylmethionine dependent methyltransferase genes (converting carboxylic acids to esters), reductases (transforming carboxylic acids to aldehydes), and decarboxylases (*89*, *90*).

To identify enzymes from the marine environment to "neutralize" the deleterious effects of High Acid Crude (HAC), cDNA libraries from marine sources will be screened with degenerate oligonucleotide probes to identify appropriate metabolic enzymes. SAM-dependent methyltransferase primers were as following; M1F: 5' GAY TTY AAY CAR ATY TTY CA-3'; M2F: 5'-AAY CAN GRN CAY TTY TTY CA-3';

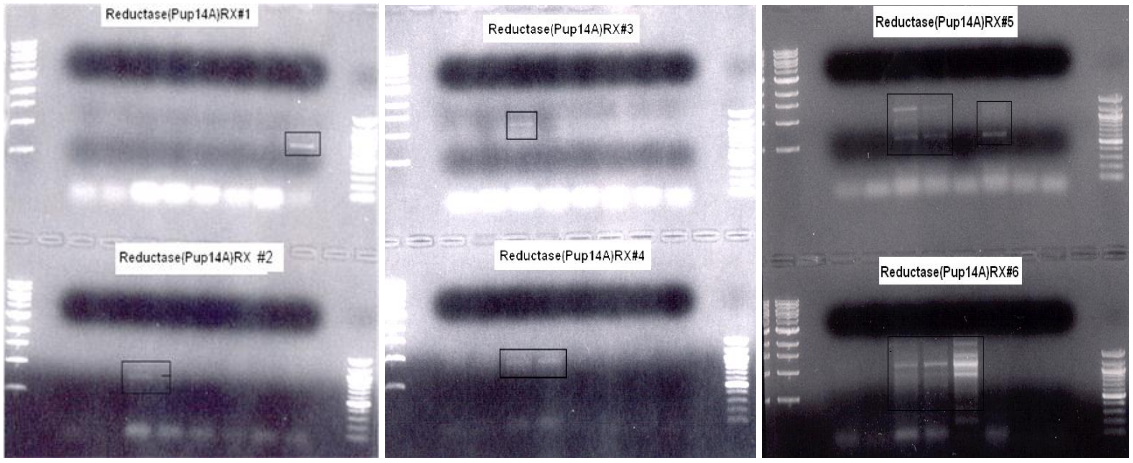
M2R: 5’-TGR AAR AAR TGN CAN GTR TT-3’; M3F: 5’-GAR AAR GAY GAY TTY ATG CAR TT-3’; M3R: 5’-AAY TGC ATR AAR TCY TTY TC-3’; M4R: 5’-CAR-TAD ATN GGR ATR TTR AA-3’. To get optimized conditions for PCR, M1F-M2R (252bp), M1F-M3R (540bp), M1F-M4R (948bp), M2F-M3R (258bp), M2F-M4R (666bp), and M3F-M4R (378bp) primer sets were performed with Pup14A, Pup14B and Hon6 genomic DNA.

Carboxylic acid reductase primers are R1F: 5’-CCN AAR GGN GCN ATG TAY AC-3’, R2F: 5’-TTY GAY GAR GAY GGN TTY TA-3’, R2R: 5’- TAR AAN CCR TCY TCR TCR AA-3’, R3F: 5’-TAY GCN AAY GGN RAY GGN AA-3’, R3R: 5’-TTN CCR TAN CCR TTN GCR TA-3’ and R4R: 5’-ARC CAR TCN ACR AAN GTR TC-3’. For the optimization of PCR condition, the following primer sets were used for PCR reaction; R1F-R2R, R1F-R3R, R1F-R4R, R2F-R3R, R3F-R4R and R3F-R4R. Even though Carboxylic acid reductase primers were designed with conserved region by multiple alignments, there are significant variations of non-conserved regions observered in different organisms. Any clear PCR product therefore was subjected for sequencing analysis. Degenerate oligonucleotide probes have been designed as shown in Table. 3.

Table 3: Designed degenerate primers of metabolic enzymes

| Carboxylic acid reductase | SAM-dependent carboxylic acid methyltransferase |
|---|---|
| 5’- CCN AAR GGN GCN ATG TAY AC-3’ | 5’- GAY TTY AAY CAR ATY TTY CA-3’ |
| 5’- TTY GAY GAR GAY GGN TTY TA-3’ | 5’- AAY CAN GRN CAY TTY TTY CA-3’ |
| 5’- TAR AAN CCR TCY TCR TCR AA-3’ | 5’- TGR AAR AAR TGN CAN GTR TT-3’ |
| 5’- TAY GCN AAY GGN TAY GGN AA-3’ | 5’- GAR AAR GAY GAY TTY ATG CAR TT-3’ |
| 5’- TTN CCR TAN CCR TTN GCR TA-3’ | 5’- AAY TGC ATR AAR TCY TTY TC-3’ |
| 5’- ARC CAR TCN ACR AAN GTR TC-3’ | 5’- CAR TAD ATN GGR ATR TTR AA-3’ |

PCRs were performed with genomic DNA of Pup14A, Pup14B, and Hon 6 to optimize PCR conditions and primer sets for each reaction as described in expreimetal procedure. The PCRs were analyzed using an agarose gel. To screen SAM-dependent methyltransferase primer sets, the reaction M1F-M2R, M2F-M3R and M3F-M4R with Pup14A resulted in PCR products of the expected size and the reaction with M2F-M3R primer set resulted in a product of the correct size from Pup14B. However, no PCR product of the expected size was found with Hon6. Detailed results are shown in Appendix Figure 4. For Carboxylic acid Reductase screening, all primer sets generated PCR product with Pup14A (Figure 26) and M1F-M2R, M1F-M4R and M3F-M4R reaction gave PCR product with Pup14B. Hon6 gave no PCR product with Carboxylic acid Reductase primiers.




A B C

Figure 26: Gel analysis with PCR product of carboxylic acid reductase on Pup14A. Ladder is 1kb DNA ladder from NEB. A: top R1F-R2R primer bottom: R1F-R3R; B: top R1F-R4R bottom: R2F-R3R; C: top R2F-R4R bottom: R3F-R4R primer set was used.

In conclusion, cDNA libraries of Pup14 A/B and Hon6 were successfully constructed. We attempted to find a desired clone for indentification of a natural product biosynthetic pathway with various methods. However, no useful biosynthetic pathways were detected with cDNA libraries so far. Screening of the cDNA libraries involved lots of sequencing with different primer sets. I did not include a chromosome walking experiment we used to identify valuable clones which was not completed due to lack of funds. Sequencing is quite expensive because numerous samples are required to obtain desired sequences. Ideally, the identified clone of the cDNA libraries can be utilized as a probe for identification of the remainder of the pathway genes of a genomic library.

# CHAPTER III

# CONSTRUCTION AND ANALYSIS OF FOSMID LIBRARY OF ENVIRONMENTAL DNA IDENTIFICATION AND SCREENING OF NATURAL PRODUCT OF BIOSYNTHETIC PATHWAYS

## INTRODUCTION

Traditional approaches to natural product discovery involve cell-based screening of natural product extracts followed by compound isolation and characterization. Their importance notwithstanding, continued mining leads to depletion of natural resources and the reisolation of previously identified metabolites. Metagenomic strategies aimed at localizing the biosynthetic cluster genes and expressing them in surrogate hosts offers one possible alternative. Environmental libraries have been extremely valuable as a source of information about microbial diverisity in natural ecosystems, even though little information about the physiology of unknown organisms (*91*). The construction of fosmid libraries from uncultured organisms is one method that theoretically can yield more insight into the metabolism and genetics from the sample (*91, 92*).

Great strides have been made with the randomized cloning of soil bacterial genomes giving rise to natural product expression without further manipulation (primarily through the use of fosmid libraries) (*57-61*). Antibiotics turbomycin A and B,were both isolated and produced from a 24,546-member fosmid library (*61*).

The most popular systems for less than 50 kb-insert libraries are cosmid libraries containing cos sites (cohesive ends) which are recognized by the viral proteins for packaging. A fosmid library is a modification of the cosmid with an F-factor origin of replication to lower the copy number of the generated plasmid, increasing its stability. Because the minimal DNA sequence to encode these functions requires about 10kb, these vectors are capable of carrying approximately 40 kb of foreign DNA. Construction of λ-bacteriophage-based libraries is a popular technique for molecular biologists (*93-95*). A wide variety of vector systems have been constructed based on this system which primarily differ in the amount of the original viral genome and function still contained in the DNA fraction of the system (*93*). The viral head, which carries the DNA construct, is limited by volume to contain no more than 50 kb of DNA, so the less viral sequence contained in the DNA the larger the foreign insert can be. While the fosmid cannot compete with the BAC (bacterial artificial chromosome) library or YAC (yeast artificial chromosome) library in terms of the size of the insert it can propagate, the fosmid gives wider information in fewer samples than a cDNA library. Fosmids are commonly used for preparing genomic libraries when a smaller insert size is desired. The inserts have an average size of 40 kb and are produced by random shearing, yielding more uniform coverage of the genome than other library types. Fosmids are excellent candidates for closing gaps in a whole genome sequence and physical maps due to their uniform coverage. Additionally, they are popular for metagenomic and expression studies (*53, 58, 91, 96*) . The entire 40 kb insert can be shuttled into an expression vector more easily than a larger insert (*92*) and the inducible high copy number per cell allows for easier

manipulation and cheaper downstream applications. Because of these advantages, we attempted to construct fosmid libraries from environmental samples.

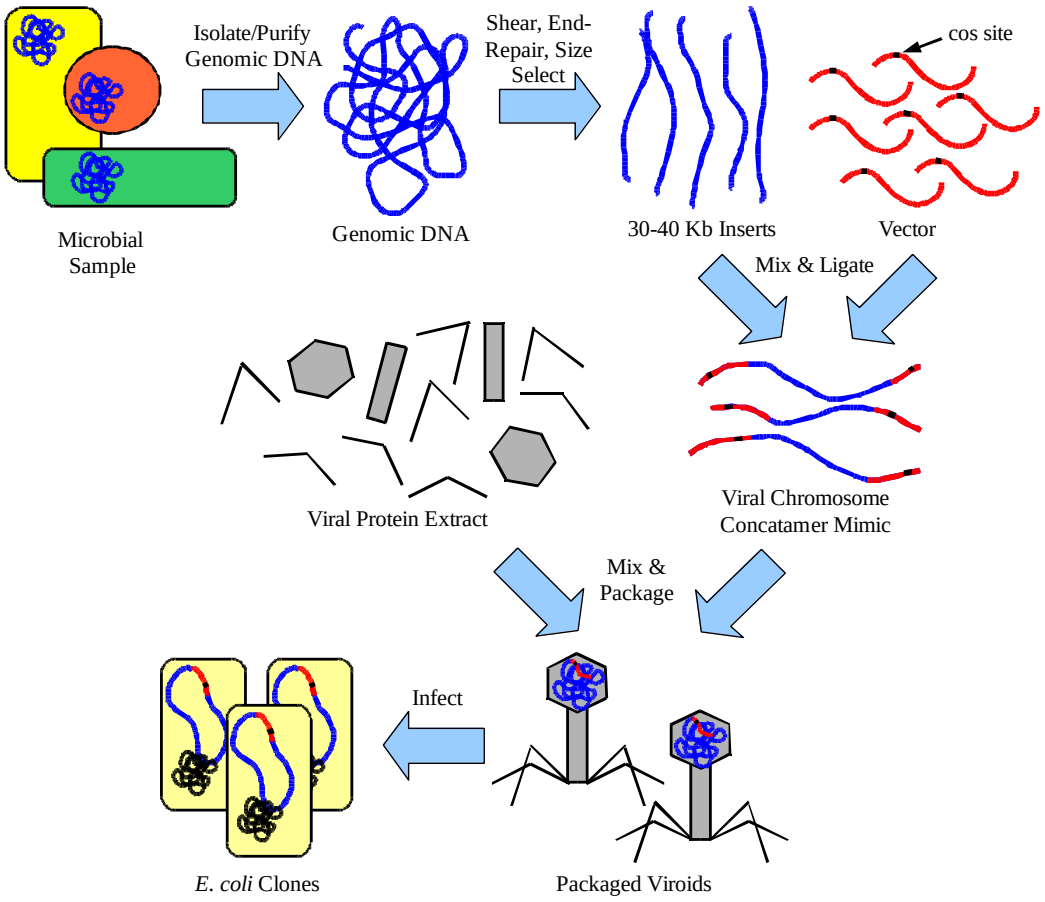



Figure 27: Overview of the fosmid library construction.

I described the method used to construct a fosmid library of CD151 and CD179 in the shuttle cosmid vector pOJ446. This included (A) preparation of pOJ446 vector (B) preparation of Genomic DNA from CD151 (C) preparation of Recombinants (D) isolation and analysis of the clones. Each of these is discussed in detail below.

## EXPERIMENTAL PROCEDURES

*General.* Cells were routinely grown in LB medium containing 1% Bacto Tryptone (Difco), 0.5% yeast extract (Difco), and 0.5% NaCl with 50μg/ml for ampicillin. UltraClean Mega Prep Soil DNA Kit was purchased from MO BIO Laboratories, Inc. (Carksbad, CA). Qiaprep Spin Mini Kit, QIAGEN plasmid kit, QIAquick Nucleotide Removal kit, and QIAquick Gel Extraction Kit were obtained from QIAGEN Inc (Valencia, CA). Restriction enzymes and alkaline phosphatase enzymes were obtained from New England Biolabs (Ipswich, MA 01938-2723). Pulsed Field Gel Electrophoresis (PFGE) was performed on a BioRad CHEF DR II apparatus (Hercules, CA). Electroporation of cells was carried out with a BioRad MicroPulser (*97*). The *E. coli* strain DH10B was obtained from Prof. Dennis Gross, Texas A&M University, Department of Plant Pathology.

*Constructions of Fosmid Libraries from Environmental DNA with pCCIFOS Vector (eDNA, Biofilm).* A fosmid library has been constructed from the samples collected from Galveston Beach. Sheared DNA of approximately 40 kb was isolated, end-repaired, packaged, titered, and plated (Figure 1). The CopyControl™ pCC1FOS™ cloning vector for CopyControl Fosmid library production is supplied linearized at its Eco72 I (blunt) site and then dephosphorylated and is ready for cloning end-repaired (blunt end) genomic DNA of approximately 40 kb (Epicentre). End- repaired eDNA is shown in Figure 28. Biofilm was obtained from a field collection by Dr. Dan Thornton, Department of Oceanography, Texas A&M University, in the Gulf of Mexico. The

biofilm was initially washed with 19g/L instant ocean seawater to remove dirt and debris, and then cut with a razor blade into small pieces. The pieces were flash frozen in liquid nitrogen, then ground with a mortar and pestle. (This part was performed by Dr. Angell) The Biofilm fosmid library was also constructed as described in manual provided by Epicentre (*98*).

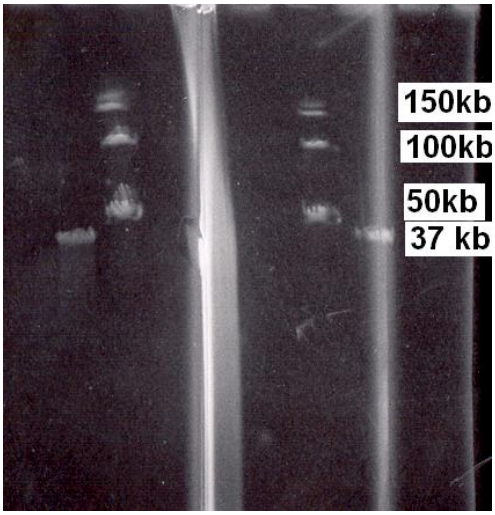



Figure 28: End- repaired eDNA ready for size selection on pulsed field electrophoresis (CHEF gel).

*Constructions of Fosmid Libraries from Environmental Samples with pOJ446 Vector (C3, Bahama, CD151, CD179).* Fosmid libraries were constructed with pOJ446, which is a cosmid shuttle vector for cloning (*99*) (shown in Figure 29) instead of Copycontrol pCC1FOS vector provided from EPICENTRE. pOJ446 can be maintained in both *E. coli* and Streptomyces. pOJ446 vector has Apramycin–resistance as an antibiotic selectable marker. In this study we also used methylation-deficient strain $dam^-/dcm^-$ from NEB. In wild-type *E. coli*, dam methylase transfers a methyl group to the adenine in the sequence GATC, and dcm methylase does to internal cytosine in the sequences CCAGG and

CCTGG (*100*). The strain dam-/dcm- was employed for transformation in streptomyces for future experiments.

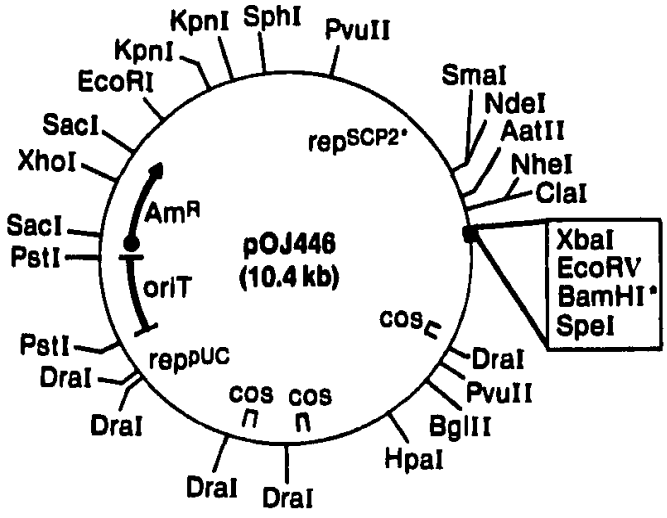



Figure 29: pOJ446 vector map. After Bierman, M. *et al* (*99*).

*Isolation of pOJ446 Vector DNA from E. coli.* pOJ446 in strain DH10B was grown in 5ml of LB+Apramycin (50μg/ml) for overnight then inoculated in 500ml LB + Apramycin media for about 16hrs. The cells were spun down at 9000 rpm for 10 min then purified with a QIAGEN Plasmid Maxi kit. The purification procedure was followed as described in the protocol provided in kit.

*Cleavage of Vector DNA and the Preparation of Linearized DNA.* Double digestion, which is cleaving a DNA substrate with two restriction enzymes simultaneously, was performed to obtain linearized vector DNA. HpaI and EcoRV were chosen as restriction enzymes to get the blunt end site for pOJ446 vector DNA. The remaining two cos sites of pOJ446 vector would make it possible to package into bacteriophage λ particles. The

restriction digestion included EcoRV (20units), HpaI (5units), NEB buffer#4, 100X BSA (10mg/ml), pOJ446 vector DNA, and water was performed at 37°C for 1.5hr, and then heat inactivation at 80°C was followed to kill the enzyme. Linearized pOJ446 vector was purified by agarose gel electrophoresis (shown in Figure 30 (A)). ~9 kb fragment was cut from the gel and isolated using QIAquick Gel Extraction Kit. Final concentration of linearized pOJ446 vector is 0.5µg/µl shown in Figure 30(B). After 1st elute obtained, the column was eluted 4 times more with deionized water to improve the yield of DNA recovery. 5th elute was detected on 1% agarose gel (Figure 30(B)).

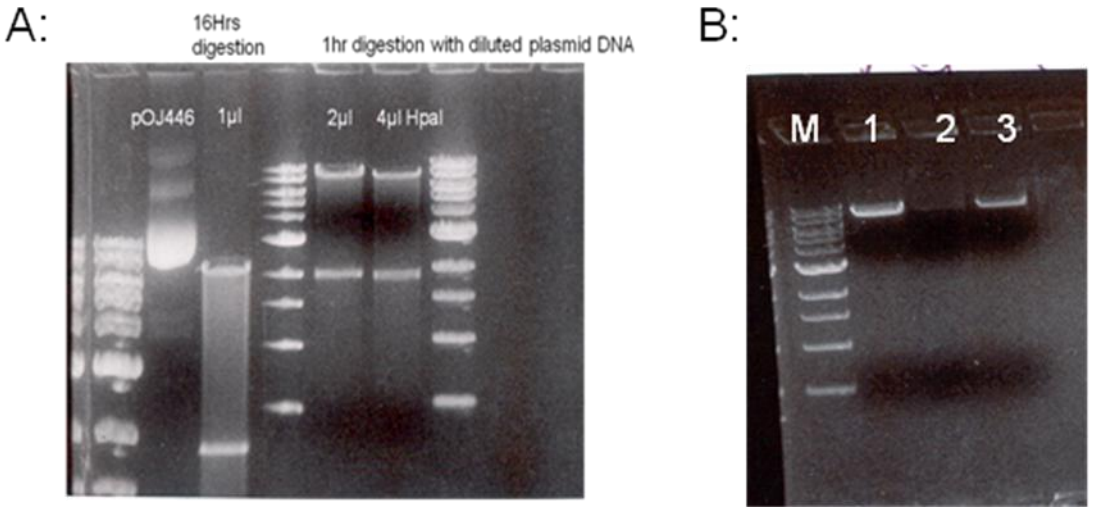



Figure 30: (A) After double digestion of pOJ446 vector DNA. (B) Recovered DNA after gel extraction. (1: 1ST elution 2: 5th elution 3: total)

*Dephosphorylation of Blunt–Ended pOJ446 Vector. DNA* 10µg of double digested pOJ446 vector DNA was dephosphorylated with 5units of CIP in NEB buffer #3 for 15 min at 37°C, then another 5units of CIP was added into the mixture to ensure complete

dephosphorylation. After another 45 min incubation at 55°C, the reaction was terminated by adding 4µl of 0.5 M EDTA and heat inactivation at 75°C for 10 min. Dephosphorylated pOJ446 vector DNA was concentrated with isopropanol precipitation followed by washing by 80% ethanol. Final concentration of pOJ446 vector DNA was 0.5µg/µl shown in Figure 31.

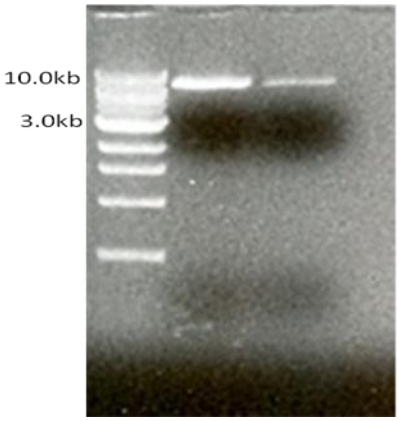



Figure 31: Dephosphorylated pOJ446 vector DNA with two arms. Lane 1:1kb ladder (NEB), lane 2: 33µg/µl, lane 3: 5µg/µl.

*Preparation of Genomic DNA from CD151.* CD151 sample was obtained from Arabian Sea, Pakistan Margin by RRS Charles Darwin. The core was taken on September, 27, 2003 from actual depth 263m. PowerMax Soil DNA Isolation kit from MO Bio Laboratories, Inc was used to extract sample DNA of CD151. After obtaining ~20µg purified DNA, we performed end repair reaction, which converts 3' and 5' overhangs into blunt ends with DNA terminator End Repair kit from Lucigen Corporation to get blunt ended CD151 DNA. We were interested in DNA bigger than 10kb size. Pulsed Field Electrophoresis (CHEF gel) was run for size selection shown in

Figure 32. We usually performed Gel elution using Pulsed Field Elctrophoresis, but the yield was quite low. In this study, we used gel extraction kit from QIAGEN to get the size selected DNA after running the CHEF gel. After gel extraction, we got the size selected DNA with ~7ng/µl. Isopropanol precipitation was performed to get 20ng/µl for ligation reaction.

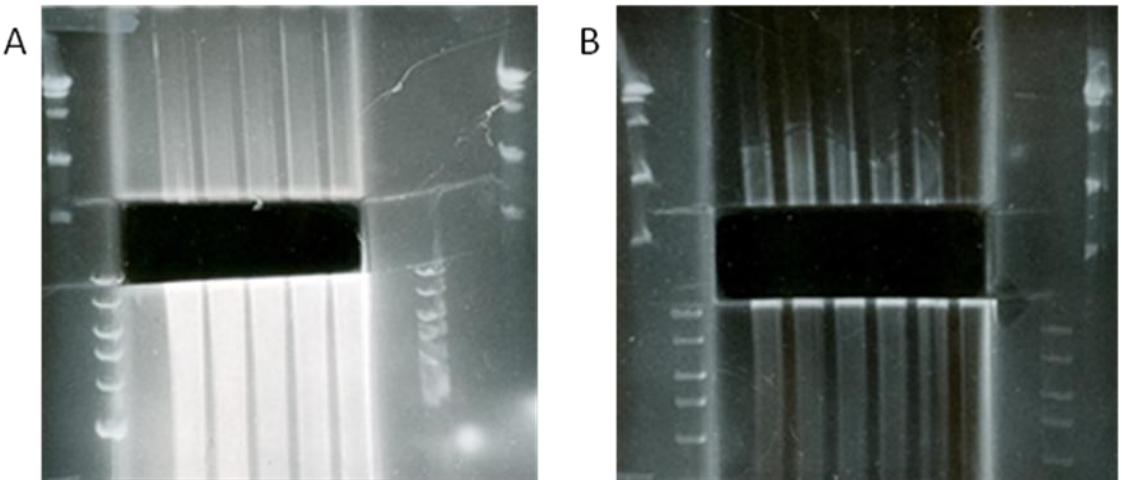



Figure 32: (A) Pulsed field electrophoresis of CD151 end-repaired DNA. (B) Pulsed field electrophoresis of CD179 end-repaired DNA CD151 from Arabian Sea, Pakistan Margin and CD179 from Nazare canyon, Portuguese magargin.

*Preparation of Recombinants.* Ligation reaction was performed with following condition; 1 µl 10X ligation buffer, 1 µl 10mM ATP, 1 µl pOJ446 (0.5µg/µl), 6 µl insert DNA, and 1 µl Fast-Link DNA Ligase in total 10 µl reaction followed by 2 hours incubation at room temperature and then heat inactivation at 70°C for 10min. Amount of insert DNA used was 1:5 and 1:10 ratio of vector to insert DNA. Electroporation instead of packaging and titering was chosen for these library constructions (CD151 and

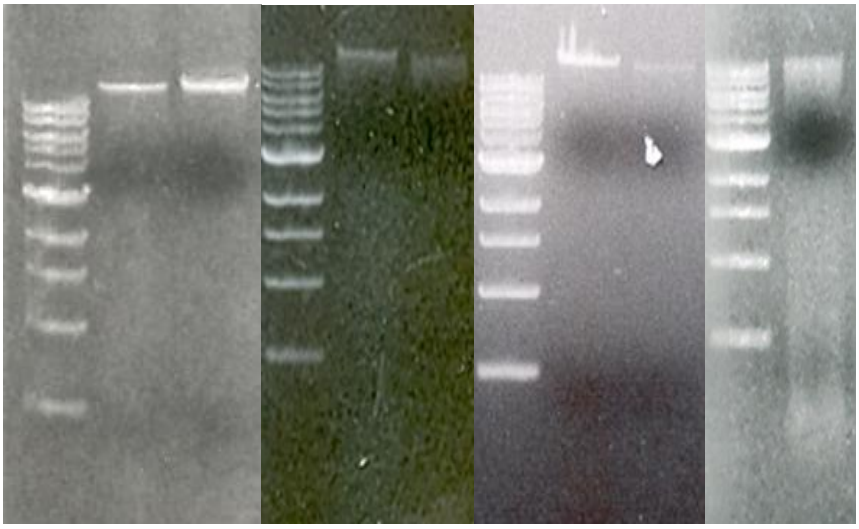
CD179). The ligation mixture was desalting using Agar-cone on ice for 1hr, followed by electroporation into *E. coli* EPI300-T1$^{R}$ competent cells provided with the Fosmid library construction kit.

*Isolation and Analysis of the Clone of the Constructed Library.* After transformation of the library into *E. coli* EPI300-T1$^{R}$, ten randomly selected clones were selected from 1:5 and 1:10 ratio ligation conditions. Plasmid DNA was isolated by alkaline lysis using QIAGEN Plasmid mini kits according to standard procedures (*101*). The plasmids were digested with EcoRI and analyzed by PFGE (Figure 33 and 34). Colonies were selected and arrayed on 384-well plates. In total, thirteen plates were arrayed for each environmental library, giving a total of 4992 members in each library.

*Smaple Preparation for the Ames Test.* CD151 and CD179 libraries were used for the Ames test. Each plate was grown in 500 ml LB+Apramycin (10μg/ml) media. The supernatant was removed by centrifugation, the cell pellet was resuspend in 50ml of 1:1 (Methanol: Dichloromethane) mixture. After 4 hrs stirring, the cell debris was removed by filtration. The filtrate was dry out with Rotory vaporizer and then purged with nitrogen gas to completely dry out. The sample was dissolved in 500μl of DMSO, ethanol can be used in the case of the sample does not completely dissolved in DMSO. For the Ames test, *E. coli* WP2 (CGSC#5378) and AB2437(CGSC#2437) strain were used.

## RESULTS AND DISCUSSION

*Isolation of DNA from Environmental Samples.* DNA purification from environmental samples was the most difficult, because the yield is too low and the DNA is susceptible to degradation by many enzymes and microbes present in the environment (*97*). After we evalutated several methods to isolate DNA from environmental samples, the UltraClean Mega Prep Soil DNA Kit was determined to be the best method. DNA was isolated by following the protocol provided with the UltraClean Mega Prep Soil DNA Kit. Some representative isolated samples are shown in Figure 33.



A B C D

Figure 33: Isolated DNA from environmental samples. All lane 1 is 1kb DNA ladder (NEB) A: lane 2: control DNA (1µg/µl) lane 3: eDNA, B: lane 2: control DNA (1µg/µl) lane 3: Sittee river, C: lane 2: control DNA (1µg/µl) lane 3:C3, D: lane 2: CD179

*Construction of the Fosmid Library from Environmental Samples.* Fosmid library construction was attempted with 13 different samples. (12687, 12704, 12723, Biofilm, eDNA from Galveston beach (Texas), A: Mouth of Mississippi, B: Atchafalaya Bay, C3: Louisiana/Texas border Jones lagoon, M: Mystery cay, BH: Bahamas, SR: Sittee River, CD151, CD179).

Fosmid libraries were successfully constructed with traditional packaging method from the following samples: the Biofilm from Gulf of Mexico, eDNA from Galveston beach at Texas, C3 from Louisiana/Texas border, and "Bahamas" from the Bahamas. However, we did not obtain enough clones of eDNA fosmid library from Galveston beach. The first 7 clones from eDNA forsmid library after digestion have bigger than 35kb size insertion as shown in Figure 34. Control DNA is 35kb size. Unfortunately, purification of eDNA resulted in low DNA yield and ligation efficiency is not enough to get many colonies. The number of clones obtained was insufficient to produce a library. The number of colonies is required to achieve sufficient coverage of putative biosynthetic pathways within the library.

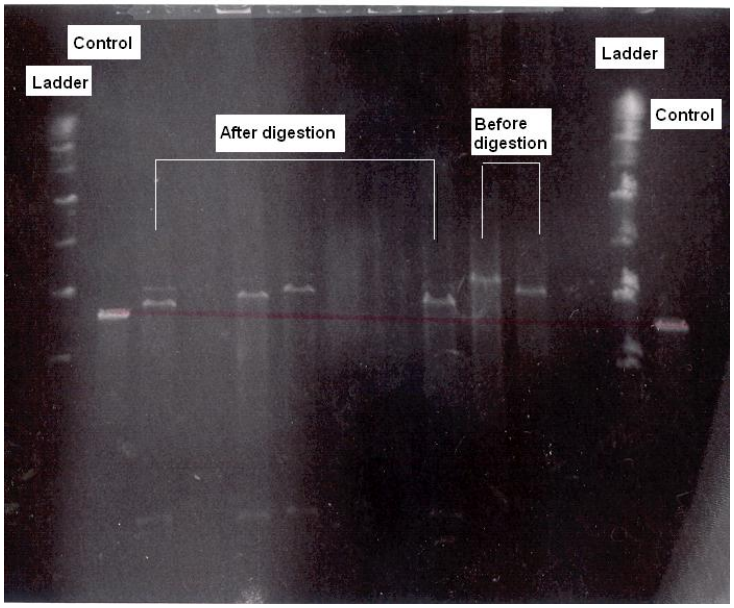



Figure 34: Pulsed field gel electorphoresis (PFGE) sizing of fosmid library of eDNA after digestion with EcoRI. Control DNA size is 35kb. Lane 1, 13: NEB midrange II PFG marker. Lane 2, 14: control DNA. Lane 3-9: EcoRI digests of fosmind clones. Lane 10,11: undigested lane 3 and 5.

CD151 fosmid library and CD179 library were constructed with a slightly different method. A *Streptomyces* - *E. coli* shuttle vector was used in place of a fosmid. I tried a packaging method and a transformation with electroporation method. Based on the investigation of insertion size, I decided to use the transformation method. The method had been successfully applied to different systems in our lab. Sample CD151 and CD179 were obtained from the Pakistan margin in Arabian Sea and the Portuguese margin in the Atlantic Ocean respectively. These 2 samples originate in the earth of deep depth under the ocean which may contain different microorganisms than the surface and terrestrial soil.

To get the clones with bigger insertion size, ligation reaction was performed with different ratio of the vector and insert DNA. CD179 fosmid clones were ramdomly chosen to check the insertion size with various enzymes. The pulsed field gel electrophoresis was performed after digestion with EcoRI of the representative clones as shown in Figure 35. 60% of greater than 10kb size insertion was found in 1:5 and 40% from 1:10 ratio of the ligation. The ligation from the ratio of 1:5 (vector to insert) has large-size inserts compare to from 1:10 ratio ligation. However, 1:10 has better variation of the insertion size. Therefore, 4600 clones were selected from the ligation of 1:10 ratio and 2300 clones were selected from the 1:5 ratio.

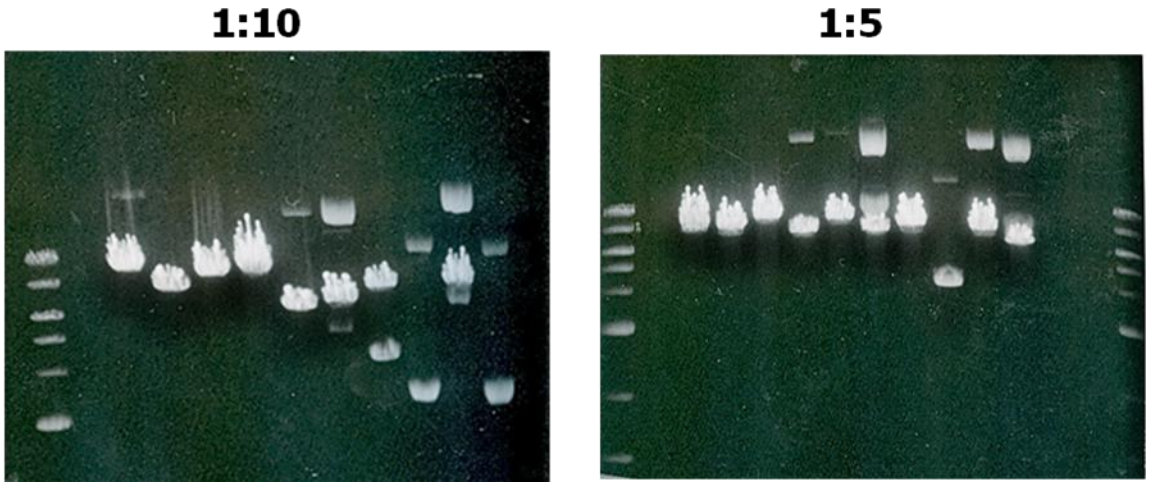


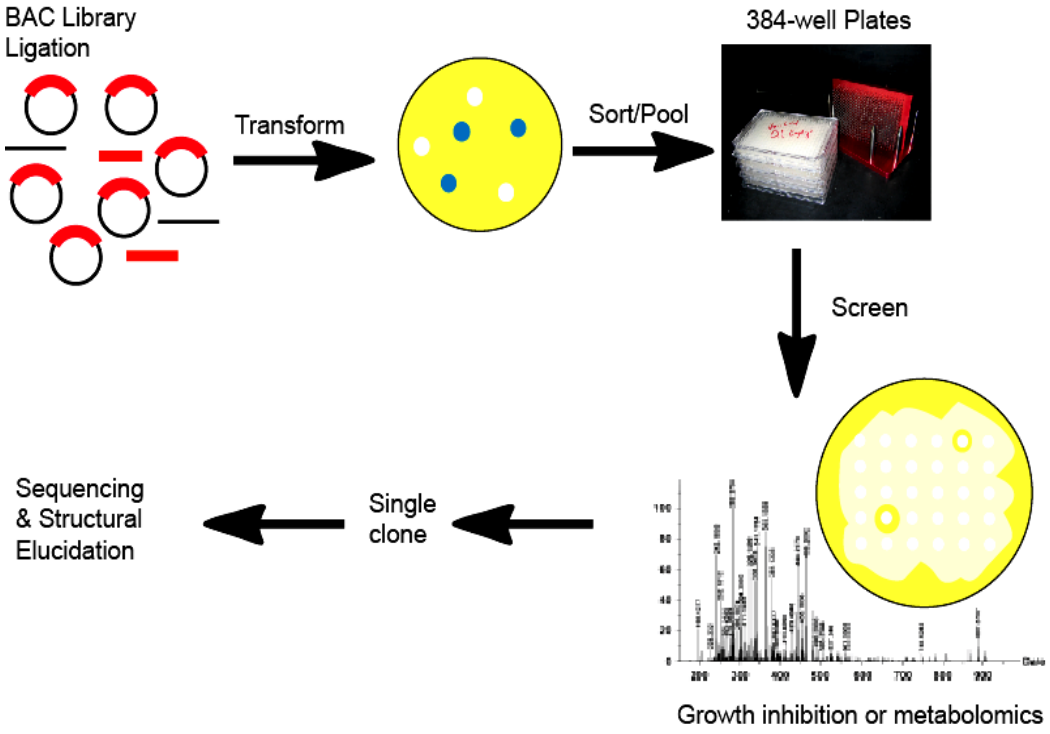

Figure 35: PFGE sizing of fosmid inserts (CD179). Lane 1,14-NEB 10kb markers. Lane 3-12 – EcoRI digests of CD179 fosmid clones. 1:5 (vector: insert) has large-size inserts, so using this ratio 6 plates of clones were selected. 12 plates of clones were selected from 1:10 which has better variation of the size.

*Screening of an Environmental Fosmid Library.* The Construction of fosmid library has been performed with many environmental samples obtained from the various

locations as described above. A total of ~5,000 clones was collected and screened for natural product production. Metabolite production was assayed by two independent methods including zone of inhibition assays and metabolic profiling by LC/MS spec analysis. An overview of the entire approach is shown in Scheme 3.

*Zone of Inhibition Assays.* Fosmid libraries were examined for natural product production by assaying for growth inhibition against a lawn of cells such as *B. subtilis* or yeast (*53*). Two strategies could be employed. Either the cells could be plated with pins in a 96–well format on a lawn of cells, or the fosmid library clones could be cultured and extracted with methanol or dichloromethane, and the organic extracts applied to filter disks and laid on the cultured lawn of cells (Kirby-Bauer technique,(*102*)). Inhibition would be observed by the presence of a halo around either the cell colony or filter disk. We chose the second method for screening of fosmid libraries (Biofilm, CD151, and CD179); methanol or dichloromethane extracted sample was dried with rotor vaporizer and then resuspended in however no interesting results were obtained.

*Metabolomics Approach*. The metabolic "blue prints" of the fosmid library colonies will also be compared to control colonies harboring the vector alone. The fosmid library colonies were grown, pooling sixteen colonies at a time, which will subsequently be extracted with ethyl acetate. The extracted colonies will be directly compared to extracts generated from the control colonies utilizing LCMS for analysis. All extracts will be examined in duplicate. 16 colonies combined to one sample for the LC/MS analysis. 312 LC/MS samples from Biofilm fosmid clones were prepared and then stored at -20°C for the LC/MS analysis.




Scheme 3: Overview of fosmid library screening: fosmid libraries are transformed into *E. coli*, the white colonies picked and cultured in 384-well plates. Clones are then examined in zone of inhibition assays or their extracts examined by mass spec (metabolomics approach). The fosmid clone is then identified, sequenced and the compound(s) of interest characterized.

*Screening with $^{32}P$ Labeled Hybridization Probe.* Natural product biosynthetic genes are typically clustered in prokaryotic genomes and are often clustered in fungi as well (*103*). Hence, detection of a single gene is often sufficient for the identification of the remainder of the gene cluster. Fosmid library screening with radiolabeled degenerate probes has thus been extensively utilized to target specific biosynthetic genes, such as polyketide synthases (PKS), and non-ribosomal peptide synthetases (NRPS). For the screening of the Biofilm fosmid library with hybridization probe, we employed nitrocellulose filters which can be used for replica plating and *in situ* hybridization. The

procedure was followed as described in literature (*104*). To test a large number of clones at one time, the library was spread out on agar plates, and then the clones were transferred to filter membranes. The clones could be simultaneously hybridized to a particular probe. All 13 plates of Biofilm fosmid library were screened with a $^{32}P$ labeled hybridization probe ($^{32}P$ dATP). The best hit (#4E10 from Biofilm fosmid library, Plate #4, row E and column 10) was chosen by analysis with a Phosporimage reader. Ninety positive clones were found, forty clones with the most intense signal selected, then sixteen positive colonies were sequenced with PKS primers. PCR reaction results and conditions are shown in Appendix B. After sequencing analysis by NCBI blast search (*86*), 5 PKS (such as PKS type I, Modular PKS, PKS/NRPS), NRPS, and many Biotin synthase genes (bioB, bioC, bioF) were found. On the way of this analysis, we decided to try screening for biotin biosynthetic pathways within a fosmid library. Detailed data is shown in Appendix B.

*Genetic Selection with Environmental Samples.* We identified the biotin biosynthetic genes within Hon6 to evaluate the potential use of genetic selection in the elucidation of biosynthetic pathways from marine microbial consortia. (Detailed in Chapter IV) As mentioned earlier, we constructed many fosmid libraries from environmental samples. We employed the strategy in a proof of principle experiment using the Hon6 system. We used the genetic selection for the biotin and Ames test. The Ames test is a biological assay to assess the mutagenic potential of chemical compounds (*105*). A positive test indicates that the chemical might act as a carcinogen. The test uses *Salmonella typhimurium* strains that carry mutations in genes involved in histidine synthesis, so that

they require histidine for growth. The variable being tested is the mutagen's ability to cause a reversion to growth on a histidine-free medium. The bacteria are spread on an agar plate with a small amount of histidine. This small amount of histidine in the growth medium allows the bacteria to grow for an initial time and have the opportunity to mutate. When the histidine is depleted, only bacteria will survive which can produce histidine by itself with mutation. The mutagenicity of a substance is proportional to the number of colonies observed.

CD151 and CD179 fosmid libraries were screened by Ames test. The plate was incubated for at least 48 hours. We used *E. coli* WP2 (CGSC#5378) and AB2437(CGSC#2437) histidine deficient strains. Azinomycin B in ethanol (1μg) and Methyl methanesulfonate (MMS, 12.99mg) were used in a serial dilution for positive control experiments. Compared to the positive control, we did not get more or an equal number of colonies grown on the agar plate. The plasmids from CD151 and CD179 fosmid libraries were transformed into *E. coli* WP2 and AB2437 competent cell, no survival colony was detected. CD151 and CD179 fosmid libraries were also screened with biotin deficient strains. Many positive results were obtained. Data for biotin screening with the Biofilm fosmid library is shown in Appendix B. C3 samples were also screened with bioA, bioB, bioC, bioD, bioF, and bioH. Only bioA, bioB, and bioD were detected. Data is shown in Appendix B. We published the data obtained with Hon6, therefore we attempted to find a novel biotin biosynthetic pathway within the positive ones. We did not get desired results for this attempt so far.

In conclusion, we successfully constructed 5 different fosmid libraries among 16 trials. The main issue for the failure to construct libraries from environmental samples is poor DNA yield. DNA isolation from environmental samples is a major bottleneck for most studies requiring large sample sizes. Specifically, marine environmental samples do not contain as much DNA when compared to a sample from surface of the earth. If we can improve the method to isolate environmental DNA, the rate of success of construction of the DNA library will be dramatically increased. Construction of the libraries is labor intensive and also a time consuming procedure; however it is a required procedure to screen for the identification of desired clones. We attempted to identify NRPS and PKS DNA fragments by $^{32}P$ labeled hybridization probe and PCR. Based on the sequencing results of the identified clone, as I mentioned above, our interest was inclined to a genetic selection method to identify a biotin biosynthetic pathway with Hon6 library. The result is shown in Chapter IV. Genetic selection with the Ames test was interesting and a fast method mutant clones.

Clones from the constructed libraries can be screened by genetic selection for several other biosynthetic pathway such as thiamine synthase (ThiC), pyrimidine synthase (nadA), heme (hemN), lipoyltransferase (LipB), and thiamine biosynthesis (ThiH).

# CHAPTER IV

# ESTIMATING P-COVERAGE OF BIOSYNTHETIC PATHWAYS IN DNA LIBRARIES AND SCREENING BY GENETIC SELECTION: BIOTIN BIOSYNTHESIS IN THE MARINE MICROORGANISM CHROMOHALOBACTER*

## INTRODUCTION

Recent advances in molecular microbial ecology have revealed that microbial diversity in nature far exceeds that reflected in laboratory strain collections since the majority of naturally occurring species either cannot withstand the conditions imposed by traditional culture media, or do not grow sufficiently to produce observable colonies under these conditions(*106-108*). Cultivation-independent analyses such as microbial rRNA sequencing have dramatically altered phylogenic definitions of major microbial classes and suggest that a large fraction of indigenous microbial communities are composed of new but uncultivable microbial groups(*34, 109-111*). Therefore, improving organism culturability could unveil previously untapped resources and natural products with novel structural frameworks. Some success has been achieved using high-throughput culturing

---

* Reprinted with permission from "Estimating p-coverage of biosynthetic pathways in DNA libraries and screening by genetic selection: Biotin biosynthesis in the marine microorganism chromohalobacter" by Eun Jin Kim, Scott Angelle, Coran M. Watanabe 2008. Mol. BioSyst. 4, 606-613. Reprinted by permission of The Royal Society of Chemistry.

procedures that utilize the concept of extinction culturing to isolate cells in very low nutrient medium and small volumes(*112*). Moreover, recent work involving microbial challenge experiments (*82, 113*) and small molecule activation studies(*114*) has yielded promising results in the induction of "silent" natural product pathways.

An alternative strategy would be to "shotgun" clone environmental genomic DNA and express these whole genomic bacterial communities within heterologous hosts. Recent studies with soil samples and bacterial symbiont genomes are beginning to reveal the potential of this technology for both natural product and enzyme discovery (*57-59, 96, 109, 115-117*). Molecular screening of recombinant environmental libraries with degenerate probes recognizing specific biosynthetic gene sequences have also proved quite fruitful in the identification of natural product biosynthetic pathways (*96, 118-120*). In some instances, heterologous expression of these clones has resulted in metabolite production (*121*).

Here we provide details on the construction of a large construct BAC library from a previously undescribed species of *Chromohalobacter*. An algorism is put forth to estimate the number of colonies required to achieve sufficient coverage of putative biosynthetic pathways within the library, "Biosynthetic P coverage," where genome size of the microorganism is estimated by flow cytometry. Genetic selection by complementation is additionally explored as a method to identify and localize biosynthetic gene clusters within marine microbial DNA libraries.

## EXPERIMENTAL PROCEDURES

*General.* Pulsed Field Gel Electrophoresis (PFGE) was performed on a BioRad CHEF DR II apparatus (Hercules, CA). Electroporation of cells was carried out with a BioRad MicroPulser. The *E. coli* strain DH10B was obtained from Prof. Dennis Gross, Texas A&M University, Department of Plant Pathology. *N. crassa* was obtained from the lab of Prof. Deborah Bell-Pederson (Texas A&M University, Department of Biology). The *E. coli* bioB (R875, CGSCII 6954), bioC (R878, CGSCII 7087), and bioF (Mec5, CGSCII 7261) knockout strains were obtained from Dr Mary K.B. Berlyn at the E. coli Genetics Stock Center, Yale University, MCD Biology Department. The E. coli bioA (R879, CGSCII 6952) and bioD (R877, CGSCII 6956) strains were obtained from OmniGene Bioproducts. Flow cytometry experiments were carried out on a Becton-Dickinson FACS Calibur flow cytometer (Franklin Lakes, NJ). Molecular simulations were performed on a Intel Pentium 4 running Linux 2.4.

*Construction of pClasper lacZ.* The lacZ multiple cloning site and its promoter were amplified from the vector pBeloBAC11 with primers, ASC1LACZF and MLUILACZR, tagged with restriction enzyme sites Asc I and Mlu I, respectively. The resulting DNA fragment was cloned into the cloning vector pCR 2.1 TOPO (Invitrogen; Carlsbad, CA). The plasmid was miniprepped and digested with the enzymes Asc I and Mlu I, respectively. The pClasper vector was also digested with Asc I and Mlu I, phosphatase treated, and subsequently ligated to the lacZ construct with T4 ligase. The resulting

colonies were miniprepped, screened by digestion with AscI and MluI, and the respective clone confirmed through sequencing.

ASCILACZF: 5′-ACGCGTTTGGCGGGTGTCGGGGCTGGCTTAA-3′

MLUILACZR: 5′-GGCGCGCCAATTTTTTTAAGGCAGTTATTGGTG-3′

*Isolation of Hon6.* The Hon6 strain was identified from marine microbial isolates obtained from Honokohau, Maui. Seawater and sediment were plated onto marine bacterial plates. Marine agar plates were prepared with slight modification to standard literature protocols (*83*). Yeast extract (1g; Difco), Bactopeptone (1g; Difco), glucose (2g), and Bactoagar (10g; Difco) were combined in 500 mL of seawater, prepared by mixing instant ocean (19g, Petco) with deionized distilled water. The agar suspension was autoclaved, cooled to 50°C, to which miconazole (10mg/L) was added. Hon6 was identified from single colonies generated through repeated streaking of microorganisms onto fresh plates. Taxonomy of the strain was assigned by the Centrallbureau voor Schimmelcultures (CBS) in the Netherlands.

*Library Construction.* DH10B cells containing pClasper lacZ were grown in a 2L culture and harvested by centrifugation(*122*). The plasmid was isolated using a Qiagen Maxiprep kit (Valencia, CA). The vector (10μg) was digested and dephosphorylated (*122*). Following dephosphorylation, the vector was further purified by self-ligation and electrophoresis (*117*) and isolated from the surrounding agarose with a Qiagen gel-extraction kit.

Hon6 cultures (500mL; 1L Erlenmeyer flask) were grown for 20 h at 30°C and 250 rpm. The cell pellet was harvested by centrifugation, washed with TE buffer (*97*), and resuspended in an equal volume of TE. To the suspension was added an equal volume of molten agarose (1.5% m/v agarose, 10% m/v sucrose, and 40% v/v glycerol). This suspension was streaked out onto a Petri dish cooled on ice and allowed to solidify. The solidified agarose was immersed in wash buffer (*122*) containing 0.15% v/v beta-mercaptoethanol and incubated on ice for 20 h. The wash buffer was drained and the agarose incubated with fresh wash buffer lacking beta-mercaptoethanol for one hour at 37°C. The wash buffer was replaced with lysis buffer (*122*) and lysed at 50°C for 48 h. The lysed agarose material was embedded in a pulsed field agarose gel and run with settings of 15 s initial and final switch times, 6 V/cm, 14°C, and a 12 h run time in 0.5X TAE buffer (*97*). The agarose material was cut out with a razor blade and stored in 0.5 M EDTA, pH = 8 at 4 °C until use.

Restriction digest optimization was performed (*117*). Once optimal conditions were established, a preparative digest and double size fractionation was carried out (*117*) employing eight agarose plugs (150 μL), 250 U of Hind III per digest, and 0.5× TAE buffer instead of 0.5× TBE. The resulting high molecular weight Hon6 genomic DNA was isolated by electroelution with dialysis tubing (*122*).

The genomic DNA was ligated to the dephosphorylated pClasper lacZ with T4 ligase (two 300ng scale ligation reactions were necessary to produce the required number of clones, 1536) and electroporated (2kV; 0.2cm gapped cuvette, BioRad) into electrocompetent DH10B cells. Electrocompetent DH10B cells were prepared as

described in Current Protocols in Molecular Biology (*101*). Colonies (1536) were arrayed by hand into four 384 well microtiter plates and stored at −80°C.

*Insert Size Characterization.* Individual clones (24 clones) were cultured in LB broth (3mL) supplemented with chloramphenicol (15mg/L) and miconazole (3mg/mL). Clones were mini-prepped and the nucleic acid pellet dissolved in TE buffer (*101*). Samples (10 μL) were digested with 2 U of I–SceI and analyzed by PFGE (initial switch time: 20 sec.; final switch time: 6 V cm; 14°C; run time: 12 h). MidRange PFG Marker II from New England Biolabs (Beverly, MA) was utilized as a size reference.

*FACS Analysis of Genome Size.* Overnight cultures (3mL) were prepared for all species except Neurospora, which was cultured on an agar slant. *E. coli* and *S. cerevesiae* were grown for 16 h according to standard procedures in LB and YPD liquid media, respectively (*97, 101*). Neurospora crassa was propagated on an agar slant of Vogel's media according to standard procedures (*123*). Liquid cultures were pelleted by centrifugation. Pellets (50μL portions) were resuspended in 1mL of 60% Galbraith buffer in EtOH (*124*). Neurospora multinucleate spores were harvested by scraping the surface cultures with a spatula. The cells were resuspended in 60% Galbraith buffer as detailed above and filtered through a 20mμm nylon mesh. Samples were incubated at 4°C overnight, pelleted by centrifugation and resuspended in 500ml Galbraith's buffer. The samples were subsequently incubated at 37 °C for 1 h, pelleted by centrifugation and resuspended in 500ml of TE buffer containing propidium iodide (0.1mg/mL).

Following 1 h of incubation, the cells were again pelleted, resuspended in 500ml of TE and analyzed by flow cytometry.

*Analysis of Hon6 Biotin Pathway by PCR with Degenerate Probes.* The Hon6 BAC library was bulk cultured in LB medium supplemented with chloramphenicol (15mg/L). Specifically, the BAC library (from glycerol stocks) was stamped onto agar plates with 384 well pins. Following 24 h of growth, the agar was sliced into pieces and inoculated into 1L flasks containing LB medium and chloramphenicol (15mg/L). The cells were centrifuged and the DNA isolated using a Qiagen maxiprep kit. The isolated DNA was subsequently transformed into the *E. coli* bioB− strain. Transformed cells were spread onto M9 + chloramphenicol (15mg/L) agar plates. The resulting transformants were individually cultured in M9 + chloramphenicol (15mg/L) and the DNA isolated with a Qiagen miniprep kit. The isolated DNA was employed in PCR reactions with bioB degenerate probes. All subsequent PCR reactions employed DNA isolated from bioA−, bioC−, bioD−, bioF−, and bioH− *E. coli* transformants that were generated by isolating the bioB− transformant DNA and retransforming them into each knockout strain. The PCR cycles were as follows for each of the reactions detailed above: step 1= 95°C for 5 min; step 2=95°C for 1 min; step 3=annealing temp (−5°C below the melting temperature of the oligonucleotide pair) for 1 min; step 4=extension at 72°C for 10 min. Steps 2–4 were repeated 50 times and reaction stored at 4°C. The bands were evaluated by agarose gel analysis and sequenced. See supplemental information for photos of agarose gels and BLAST data.

Degenerate primers

BioA1F: 5′-CAY ATH TGG CAY CCN TAY AC-3′
BioA2R: 5′-CCR AAN GTR TCN CCR TGR TA-3′
BioB1F: 5′-GAR GAY TGY AAR TAY TGY CC-3′
BioB5R: 5′-CCN GCC ATR AAR CAC ATN GCY TG-3′
BioC6F: 5′-TNG CNG TNC ART GGT GYG G-3′
BioC7R: 5′-GTN GCN CCD ATN CCY TT -3′
BioD1F: 5′-GAY ACN GAR GTN GGN AA-3′
BioD3R: 5′-GCR TGR TTD ATR CAN CC-3′
BioF10F: 5′-ATG GTN GAY GAY GCN CAY GG -3′
BioF11R: 5′-GCY TCN GCN GGN GGC AT-3′

PCR conditions

BioA: 1 µl of Hon6 plasmid DNA, 1 µl of BioA1F primer, 1 µl of BioA2R primer, 0.5 µl of Taq polymerase, 5 µl of 10× thermal solution, 0.5 µl of dNTP, 3 µl of MgCl2, and 38 µl of deionized water.

BioB: 1 µl of Hon6 plasmid DNA, 1 µl of BioB1F primer, 1 µl of BioB5R primer, 0.5 µl of Taq polymerase, 5 µl of 10× thermal solution, 0.5 µl of dNTP, 2.5 µl of DMSO, and 38.5 µl of deionized water.

BioC: 1 µl of Hon6 plasmid DNA, 1 µl of BioC6F primer, 1 µl of BioC7R primer, 0.5 µl of Taq polymerase, 5 µl of 10× thermal solution, 0.5 µl of dNTP, 2.5 µl of DMSO, and 38.5 µl of deionized water]

BioD: 1 µl of Hon6 plasmid DNA, 1 µl of BioD1F primer, 1 µl of BioD3R primer, 0.5 µl of Taq polymerase, 5 µl of 10× thermal solution, 0.5 µl of dNTP, 3 µl of MgCl2, and 38 µl of deionized water.

BioF: 1 µl of Hon6 plasmid DNA, 1 µl of BioF1F primer, 1 µl of BioF2R primer, 0.5 µl of Taq polymerase, 5 µl of 10× thermal solution, 0.5 µl of dNTP, 2.5 µl of DMSO, and 38.5 µl of deionized water.

BioH: 1 µl of Hon6 plasmid DNA, 1 µl of BioH1F primer, 1 µl of BioH2R primer, 0.5 µl of Taq polymerase, 5 µl of 10× thermal solution, 0.5 µl of dNTP, 3 µl of MgCl2, and 38 µl of deionized water.

*Hon6 Extract ELISA.* The Hon6 clones (Hon6 #1-1, Hon6 #1-2 and Hon6 #1-3, corresponding to plate # and colony #) were each inoculated in M9 minimal salt medium supplemented with chloramphenicol (15mg/L). The M9 minimal salt medium (1L) consists of 33.9g of $Na_2HPO_4$, 3g of $KH_2PO_4$, 0.5g of NaCl, 1g of $NH_4Cl$, 20mL of 20% glucose, 2mL of 1.0 M $MgSO_4$ and 0.1mL 1.0 M $CaCl_2$. After growth in M9 liquid medium for ~41 h at 37 °C with shaking, 50µL of cells were inoculated again in fresh M9 liquid medium with chloramphenicol (15mg/L). After 7 h of growth, they were spun down for 5 min at 14 K rpm. The supernatant (50mL) from each sample was used for analysis with the vitamin H (Biotin) ELISA kit from MD Biosciences. Analyses were performed in duplicate following the manufacture's protocol. Briefly, 50µL of standard, serum control or diluted sample was transferred into the wells of the plate and incubated for 30 min. The diluted conjugate solution (50µL) was added into each well, and incubated for 30 min. After 4 washes with wash solution, 50µL of Chromogen was added to each well and incubated in the dark for 30 min. Stop solution (100mL) was added to each well and the absorbance of the solution read with a microplate reader (µQuant, BIO-TEK instruments, INC) set to 450 nm. All procedures were performed at room temperature. M9 medium (with chloramphenicol) was employed as an additional

control and showed that the medium contained no biotin. See supplemental information for details.

## RESULTS AND DISCUSSION

At the outset of constructing a genomic library of a microorganism(s) with the intent of mining the biosynthetic genes contained within, one must consider the size of the genomic library to effectively represent the genome(s) and their respective gene clusters. To address this issue, we have devised the algorism BPC (biosynthetic pathway coverage) and apply it to the construction of a BAC library of the marine microorganism Hon6.

*Achieving "Biosynthetic P Coverage" of the BAC Library (Performed by Dr. Scott Angell).* The probability of cloning an intact biosynthetic pathway of an organism, and/or determining the required number of clones *N* to ensure proper coverage of the genomic library can be calculated from eqn (1),

$$N = \frac{In(1-p)}{In(1-cf)} \qquad (eq.1)$$

where *P* represents the probability that a given pathway will be represented in its entirety by at least one clone in the library, *f* represents the fraction of *G* (the genome size of the organism) contained within an individual subclone of length *I*, or $f = I/G$, and *c* is a coefficient that defines the probability that an insert will contain the sequence of interest in its entirety. The eqn (1) is a manifestation of eqn (2),

$$N = \frac{In(1-p)}{In(1-f)} \qquad (eq.2)$$

which has been widely employed to calculate the number of clones required to ensure that any given gene is represented at least once in a library with probability *P(97, 125).* Eqn (2) is a rearrangement of eqn (3),

$$(1-P) = (1-f)^N \qquad (eq.3)$$

which states that the probability of not cloning a fragment, $(1 - P)$, is equal to $(1 - f)^N$. Recall that $f$ represents the fraction of the genome contained in *I*, *i.e.* it reflects the probability that a sequence of interest will be contained in *I*, and $(1 - f)$ is the probability that it will not, where *N* is the total number of clones. Eqn (2) and (3), however, only hold true if the sequence of interest is negligibly small in comparison to the insert size of the subclones. In the event that the desired sequence is in the same size range of the insert, the probability that the sequence will be cloned intact is considerably less than the probability of *f*. To correct for this scenario we inserted the variable *c*, where *c* is defined as (eqn (4)),

$$c = \frac{(I - B + 1)}{I} \qquad (eq.4)$$

where *B* is the length of the sequence (in our case the biosynthetic pathway) contained within the insert *I*. Hence, as the size of the gene cluster increases, the coefficient *c* decreases lowering the overall probability of cloning the pathway intact, and increasing the number of clones required to ensure that the biosynthetic pathway would be represented within the library of clones.

*Hon6 BAC Library Construction (Performed by Dr. Scott Angell).* Hon6 was identified from marine microbial isolates obtained from the coast of Honokohau, Maui. The Centrallbureau voor Schimmelcultures (CBS) in the Netherlands identified the strain as a new species, 98% identical to *Chromohalobacter*. High molecular weight genomic DNA was isolated utilizing agarose as a DNA stabilizer. Cultures of Hon6 were pelleted by centrifugation, embedded in molten agarose, washed and lysed with detergents. DNA degradation occurred to a lesser extent when the embedded cells were either streaked in a thin layer or cut into thin sheets prior to lysis. We attribute this to an increased surface area, which enhances the diffusion of nuclease inhibitors, detergents, and proteases into the agarose matrix. Isolation of megabase DNA was achieved by pulsed field gel electrophoresis, which eliminated small fragments of DNA from the gel plugs. The agarose plugs containing the megabase DNA were fragmented and digested with the restriction enzyme Hind III. A double size selection by pulsed field gel electrophoresis was carried out to generate a narrow insert size distribution (see supplemental information for representative gels on library construction).

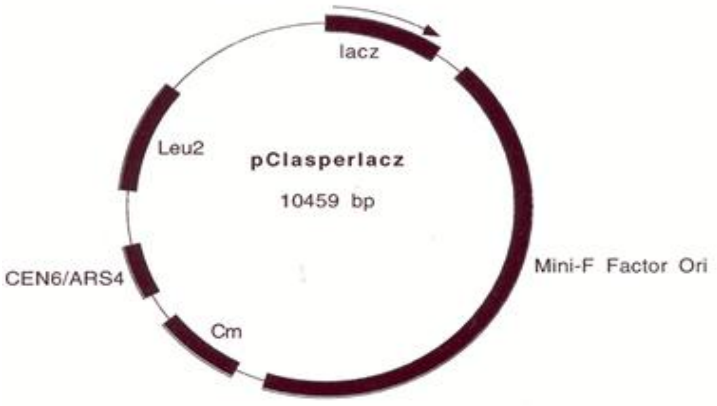



Figure 36: Basic plasmid map of the BAC yeast shuttle vector, pClasper *lacZ.*

The DNA was isol The DNA was isolated and ligated to the vector pClasper lacZ. PClasper (*126*), a vector that can maintain large DNA fragments in both bacteria and yeast was modified to install the lacz gene within the multiple cloning site (MCS region) of the vector (Figure 36). The lacZ modification allows for blue/white colony selection and elimination of empty constructs during library construction. In the long term, the vector might allow for the housing and expression of biosynthetic pathways in both

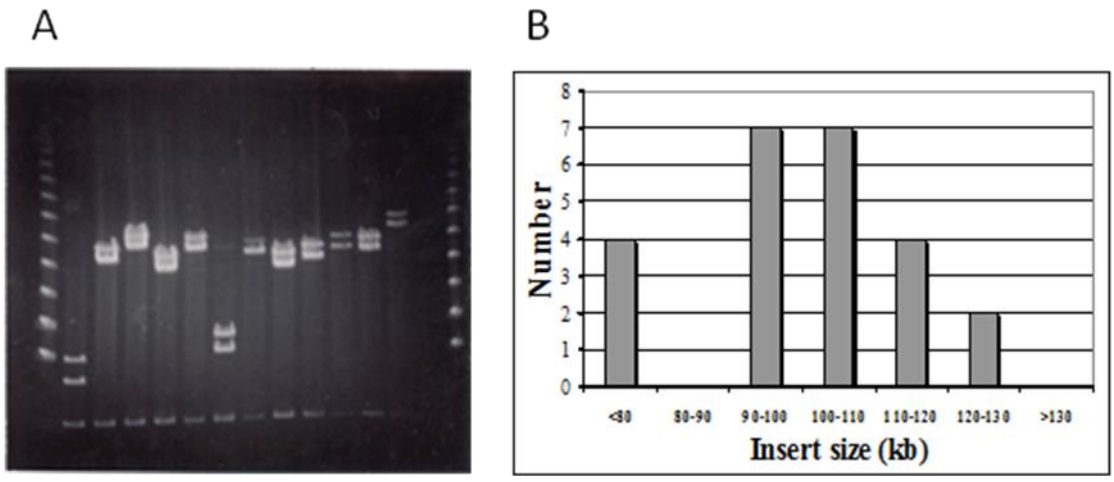


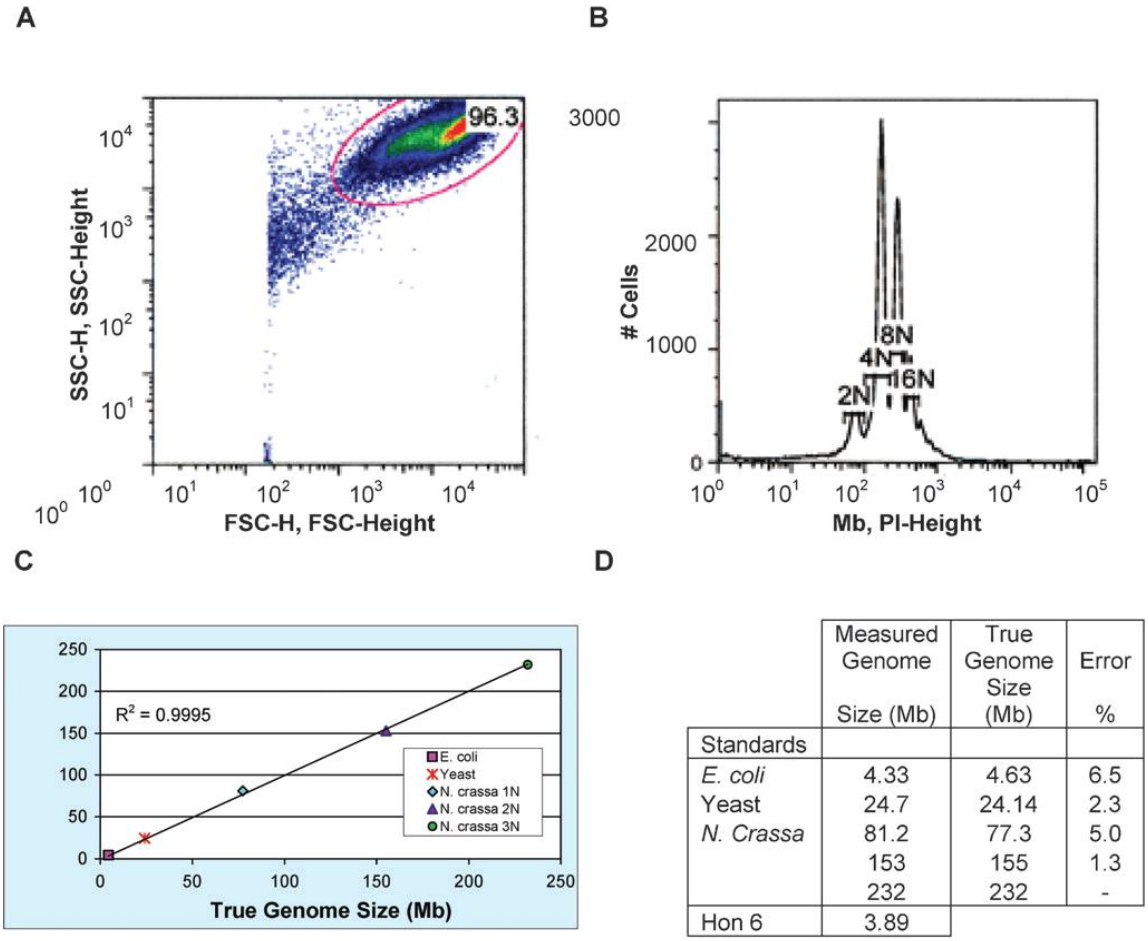

Figure 37: (A) Representative clones from the Hon6 BAC library cut with I-SecI. (B) Graph depicting the number of clones with specific insert size ranges (of 24 clones analyzed).

prokaryotic and eukaryotic hosts.

Following ligation and transformation into DH10B cells the clones were analyzed (Figure 37A). Four (16%) were found to contain no insert or an insert of negligible size, while those containing inserts, averaged 105 kb. The size distribution shows a large percentage of clones in the 100 kb size range with a few scattered individuals at sizes below 50 kb and a strong, roughly exponential decrease in the population of clones as

the size increases beyond 110 kb (Figure 37B).

*Estimation of Genome Size by Flow Cytometric Analysis* (Performed by Dr. Scott Angell). Calculating the number of clones required to allow good coverage of a genomic library requires prior knowledge of the genome size of an organism. We therefore measured the genome size of Hon6 by flow cytometric analysis. The nuclear DNA content of the Hon6 strain was evaluated by flow cytometry after PI staining relative to Neurospora crassa, which served as a standard (Figure 38 (A) and 38 (B)). Genome sizes of *E. coli* and *S. cerevisiae* were also measured to assess the degree of accuracy. The technique can be employed to estimate the genome size of an organism since fluorescence intensity is directly proportional to the DNA content of the cells (*127*, *128*). A calibration curve was generated resulting in the linear graph shown in Figure 38 (C) and was used to calculate the genome size of S. cerevisiae and *E. coli*, which are also plotted on the graph for comparison. Measured values were accurate to within 7% of the known standards (Figure. 38 (D)). For example, the genome size of *E. coli* was calculated to be 4.33 Mb while its true genome size is 4.63 Mb. Likewise, the genome of *S. cerevisiae* was measured to be 24.7 Mb while its established genome size is 24.14 Mb. The nuclear DNA content of Hon6 was estimated by this flow cytometric approach to be 3.89 Mb ± 7%.




| | Measured Genome Size (Mb) | True Genome Size (Mb) | Error % |
|---|---|---|---|
| Standards | | | |
| *E. coli* | 4.33 | 4.63 | 6.5 |
| Yeast | 24.7 | 24.14 | 2.3 |
| *N. Crassa* | 81.2 | 77.3 | 5.0 |
| | 153 | 155 | 1.3 |
| | 232 | 232 | - |
| Hon 6 | 3.89 | | |

Figure 38: (A) *N. crassa* cell population. (B) Genome size measurements for *N. crassa* in its various nuclear states. (C) Standardization curve of organisms with established genome sizes. (D) Genome size estimates for marine microorganisms based upon *N. crassa.*

*"Biosynthetic P coverage" of the Hon6 Library and Molecular Simulation Analysis (Performed by Dr. Scott Angell).* Utilizing the BPC algorism, we determined that a genome size of ~3.89 Mb, a library consisting of 1536 colonies or clones, with the vast majority of inserts in the 100 kb range (of which ~86% contain inserts), we can expect a greater than 99% probability of cloning an 85 kb pathway intact. This high probability is

achieved due to the depth of coverage. Our library contains about 132 Mb of DNA, or about 34 fold coverage of this organism.

To account for the heterogeneity in clone length, a molecular cloning simulation was conducted in the Perl programming language. For each step of simulation: a circular genome of 3.89 Mb was virtually constructed to contain a randomly-placed pathway of 85 kb; 1536 clones derived from this genome were simulated by taking a random "insert" from the genome, with the insert lengths following the distribution of sizes shown in Fig. 38B; the simulation step was scored as a success if at least one of the 1536 clones contained the pathway in its entirety, in either orientation. The simulation was carried out for over 400000 steps (Figure 39) and rapidly stabilized on a success rate of 99.88%. The success rate increases dramatically for pathways less than 85 kb. The simulation does not address bias that could be introduced by differential clonability of different genomic regions due to stability or the nonrandom distribution of restriction sites.

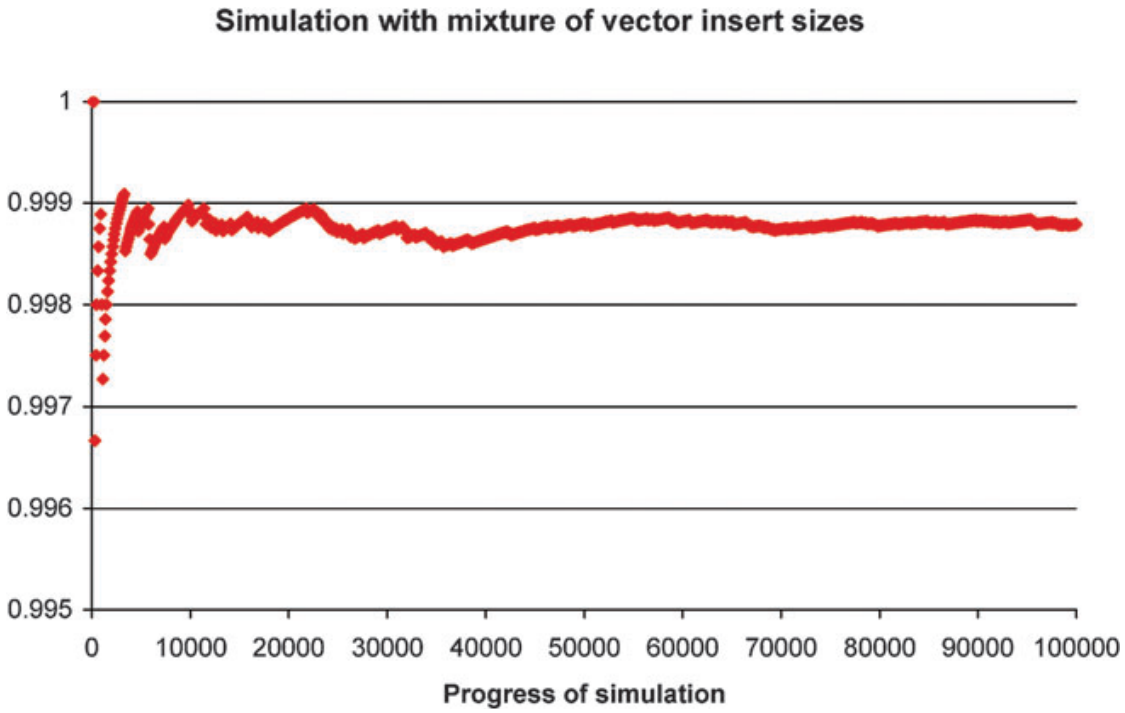



Figure 39: A simulation of cloning success rate on a pathway of 85,000 bases using the empirical distribution of clone insert sizes. The success rate is cumulative over previous steps in the simulation. The first 100,000 simulation steps are shown. Each marker represents 100 simulation steps.

*Screening by Genetic Selection: Biotin Biosynthesis in Hon6.* To explore the potential use of genetic selection in the identification and localization of biosynthetic pathways from marine microbial consortia, as a proof of principle experiment, we identified the biotin biosynthetic genes within Hon6. Biotin (vitamin H, a co-factor involved in $CO_2$ transfer) biosynthesis has been well characterized in terrestrial bacterial systems, notably that of *E. coli*, *B. subtilis*, and *B. sphaericus* (Figure 40) (*129-132*). In contrast, understanding of biotin formation within the marine environment is limited to information obtained from a few sequenced genomes (*133, 134*). The Hon6 library was

directly transformed into bioB− *E. coli* and the clones evaluated by PCR analysis with degenerate probes. The cells can only survive if clones contain an expressed gene or genes capable of complementing the knockout, i.e. cells cannot survive without some mode of $CO_2$ transfer. Sequence analysis of the PCR products revealed the presence of the bioB gene. As BioB represents the rate-limiting step in biotin biosynthesis, we measured biotin production levels of three representative transformants. Biotin production in these transformants ranged from 666ng/L to 753ng/L. While many microbes synthesize biotin, it is produced at low levels (*135*).

The bioB− containing BAC clones were then individually transformed into bioC−, bioF−, bioA−, bioD−, and bioH− *E. coli*. The resulting transformants were again analyzed by PCR with degenerate probes. The genetic selection results revealed the presence of bioC, bioF, bioA, and bioD biosynthetic genes. The bioW and bioH genes were not detected in either the bioB clones, or Hon6 genomic DNA. Therefore, from an evolutionary perspective, the Hon6 biotin biosynthetic pathway is quite similar to that of the *E. coli* bio operon. The results of these experiments additionally suggest that complementation analysis might not only be useful in the mining of biotin biosynthetic pathways from terrestrial bacterial genomes34 but also in the analysis of marine microbial consortia at various depths, locations, and environmental conditions. The approach might ultimately lead to the discovery of a novel biotin biosynthetic pathway(s) or even better, a new mode of $CO_2$ transfer; the resulting enzymology would certainly be of considerable interest. In this regard, genomic clones lacking key genes of the canonical pathways would be of principal interest.

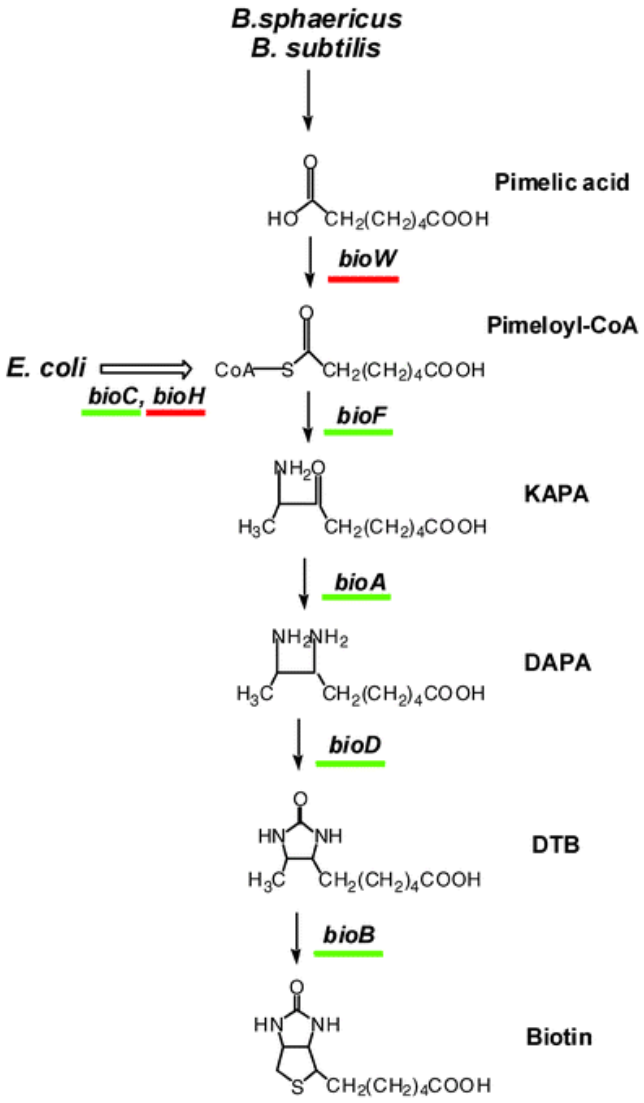



Figure 40: Schematic diagram depicting biotin biosynthetic pathways in *E. coli*, *B. sphaericus*, and *B. subtilis* the bioF, bioA, bioD, and bioB genes are common to all three pathways. The Hon6 biotin biosynthetic pathway genes are demarcated by the green underline. The red underline specifies genes not common to the Hon6 biotin biosynthetic pathway.

*Conclusions.* Current research indicates that less than 1% of microbes in the environment can be cultured by any given culture method. This suggests that countless biosynthetic pathways have escaped analysis for either natural product production, or

mechanistic investigations. Metagenomics, which involves the culture-independent genomic analysis of a microbial population aims to access this buried but highly valued information.

In this study, we have constructed a large insert BAC library of the marine microorganism Hon6, a new species of the genus *Chromohalobacter*, isolated from Honokohau, Maui. The size of the library was constructed on the basis of an algorism that we generated, referred to as BPC or "Biosynthetic P-Coverage." Based on this algorism, our library consisting of 1536 clones covering 132 Mb of DNA (34-fold coverage), ensures a greater than 99% probability of cloning an 85 kb pathway intact. While the approaches described herein focused on a single organism, image densitometry (*136*) might in the future offer the means to estimate genome sizes of microbial populations and expansion of the technique to enable genome quantification of heterogeneous populations and sizing of their respective libraries.

Additionally, to evaluate the potential use of genetic selection in the elucidation of biosynthetic pathways from marine microbial consortia, we identified the biotin biosynthetic genes within Hon6. As the statistical probability of observing heterologous expression of a single gene is much higher than that of an entire biosynthetic pathway, genetic selection could prove to be an important tool in the localization of novel biosynthetic pathways. Coupled with promoter replacement technologies, the approach could lead to the discovery of a new generation of bioactive natural products. In this regard, we are currently applying genetic selection to reveal novel terpene biosynthetic pathways in marine microbial consortia.

# CHAPTER V

# CONTROL OF HETEROLOGOUS GENE EXPRESSION BY THE THIAMIN PYROPHOSPHATE RIBOSWITCH

## INTRODUCTION

As already mentioned in Chapter I, only a small portion of microbes are culturable by present methods and 90-99% of the metagenome is not being sampled (*14*, *34*, *46*). Even marine microbial communities, including the cyanobacteria, have not been systematically collected and fully evaluated for bioactive compounds. Instead of developing purification methods or manipulating cultivation conditions, large fragments of non-culturable bacterial genomes can be cloned and expressed using recombinant DNA technology (*47-51*). *E. coli* is the most commonly used host for heterologous gene expression (*137*). A strong transcriptional promoter to control high-level gene expression is required in recombinant expression plasmids (*138*).

We aimed to develop new tools to control gene expression through the use of riboswitches. The strategy could facilitate the functional characterization of biosynthetic genes in a facile fashion by effecting gene modulation, thus adjusting natural product levels and permitting analysis of natural product flux. While pBAD and lac promoters are commonly employed inducible promoters, both promoters exhibit an all-or-none activity in wild-type *E. coli* strains and are not titratable (*139-144*). There is therefore a need for the development of new, enabling methodologies.

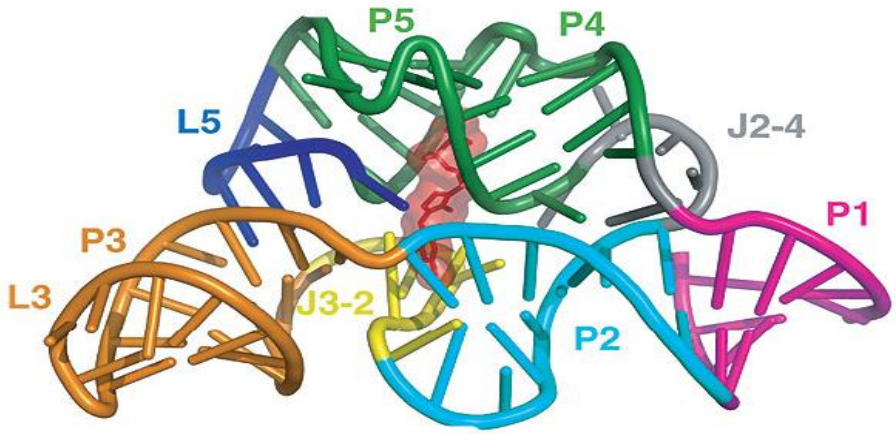


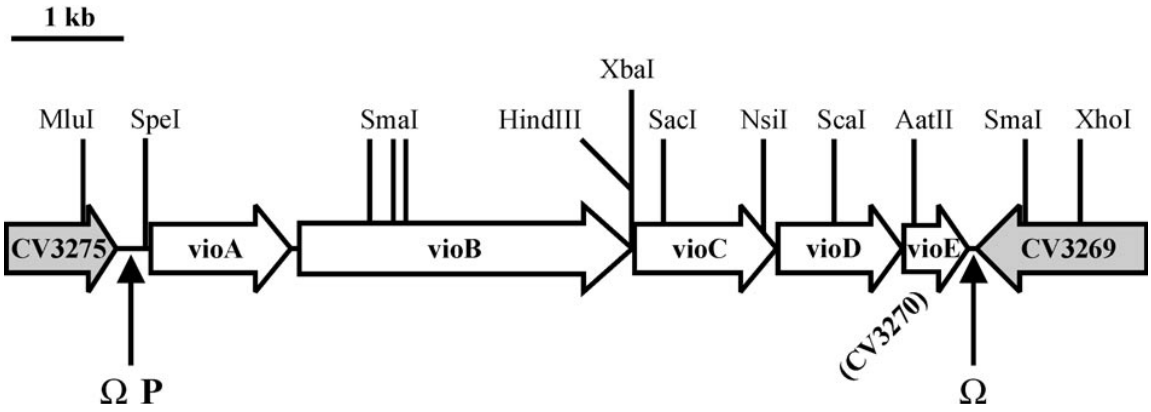

Figure 41: Crystal structure of TPP-bound sensing domain front view. The RNA is in a stick-and- ribbon representation, with bound TPP in red. after Serganov, A. et al. (*145*).

The TPP riboswitch system was chosen to upregulate biosynthetic genes that exhibit little to no innate gene expression upon initial analysis in *E. coli*. This system is particularly attractive for these investigations for the following reasons. The TPP riboswitch, like the purine riboswitch, adopts a compact inverted h-structure, and is among the simplest of riboswitches. The structure of the TPP riboswitch is shown in Figure 41(*145*).

In a β-galactosidase reporter assay, a translational fusion construct that encompassed the 5'-UTR of ThiC in *E. coli* exhibited a thiamine-dependent suppression of β-galactosidase activity of 110-fold when the cells where grown on minimal medium (*146*). Genetic control appears to occur primarily at the level of transcription as a transcriptional fusion containing the ThiC leader yields a modest 16-fold effect with added thiamine (*146*). Significant fold change differences are observed between the metabolite-bound and unbound states. Moreover, the metabolite-binding aptamer

domain of the TPP riboswitch has been shown to be essentially identical in both prokaryotic and eukaryotic systems, *E. coli* and *Arabidosis thaliana*, respectively (*76, 145, 147*). Additionally, not only will the use of the TPP riboswitch enable use of *E. coli* as a suitable host strain, but also as many gene products are toxic, the use of the TPP riboswitch will enable modulation of pathway gene expression by varying the TPP concentration. Having the ability to control metabolite flux in pathways might result in the accumulation of intermediates that could not otherwise be isolated or characterized. While inducible promoters such as the popular arabinose-inducible promoter PBAD and the isopropyl-β-D-thiogalactopyranoside (IPTG)-inducible promoter Plac come to mind, they are subject to all-or-none induction in which intermediate concentrations of IPTG or arabinose give rise to subpopulations of cells that are fully induced or uninduced (*139, 140, 144*). In contrast, the sensitivity of TPP detection by the RNA riboswitch as measured by monitoring the extent of spontaneous cleavage at several ligand-sensitive sites within the RNA over a range of ligand concentrations as well as measuring levels of β-galactosidase reporter gene expression reveals its dependence on metabolite concentration and the tunable nature of these regulatory switches.

Recently, the violacein operon has been studied in *E. coli* and other bacteria after isolation, sequencing, and heterologous expression of the gene cluster (*58, 148-150*). The biosynthetic pathway, which was mapped to a 5 kb region of DNA, has been functionally expressed in *E. coli.* The biosynthetic pathway was identified from a soil environmental library (*58*). This violacein gene cluster from *Chromobacterium violaceum* was selected to be placed under the control of the ThiC riboswitch. The entire *vio* operon is shown in Scheme 4 (*151*).




Scheme 4: DNA region encoding violacein biosynthesis from *Chromobacterium violaceum* ATCC12472 including putative transcriptional terminator (Ω) and promoter (P) regions. After Sánchez, C. *et al.* (*151*)

Besides evaluating the function of a riboswitch to express violacein in *E. coli*, the system was also applied to the expression of insoluble protein to evaluate the solubility change. Many heterologous proteins which are overexpressed lead to the formation of insoluble aggregates known as inclusion bodies (*152-154*). Inclusion body formation might potentially be minimized by a riboswitch promoter. The yield and solubility of recombinant protein are highly dependent on the specific protein sequences, as well as on the vector, host cell, and culture conditions used (*154*). Predicted flavin reductase (RutF) in a novel pyrimidine catabolism pathway and β-lacto-Globulin (BLG) protein were employed to verify the improvement of expression and solubility under the ThiC riboswitch.

## EXPERIMENTAL PROCEDURES

*Materials and General Methods.* Competent Top10, BL21 (DE3) and BL21-CodonPlus (DE3) RIL E. coli strains were from Invitrogen (Carlsbad, CA). The DH10B *E. coli* strain was obtained from Prof. Dennis Gross, Texas A&M University, Department of Plant Pathology. Oligonucleotide primers and linkers were from Integrated DNA Technologies (Coralville, IA). TaKaRa LA TaqTM DNA polymerase was from Takara Bio Inc. (Otsu, Shiga, Japan). Phusion DNA polymerase, restriction enzymes, and T4 DNA ligase were from New England Biolabs. Plasmid pETBlue-2 was from Novagen, pCCIFOS was from Epicentre, and TOPO XL 2.1 was from Invitrogen. DNA sequencing to verify PCR fidelity was performed on double-stranded DNA by the Gene Technology Lab Core Facility of Texas A&M University (College Station, TX). Plasmid DNA preparation was performed with the Qiaprep kit from Qiagen (Valencia, CA). Gel extraction of DNA fragments was done using QIAquick Gel Extraction kit from Qiagen. Thiamine Pyrophosphate Chloride (TPP salt) was purchased from Sigma. ONPG (o-Nitrophenyl-beta-galactopyranoside) ELISA β-galactosidase substrate was purchased from Biocompare.

Pulsed Field Gel Electrophoresis (PFGE) was performed on a BioRad CHEF DR II apparatus (Hercules, CA). Electroporation of cells was carried out with a BioRad MicroPulser. SDS-PAGE gels were from BioRad. High performance liquid chromatography (HPLC) purification of violacein was performed on a Gemini instrument from Phenomenex using a C18 column (3μm, 110A, 75 x 4.60 mm). Preparative HPLC was performed on a Biochrom 1010 ODS instrument from Regis with

a C18 column (10μm, 25 x 10.0 mm) and Analytical HPLC was performed on a Gemini instrument from Phenomenex using a C18 column (10μm, 150 x 4.60 mm). Liquid chromatography mass spectrometry (LCMS) identification was carried out on a Gemini-NX instrument from Phenomenex using C18 column (3μm, 110A, 150 x 2.00 mm) and TIC (ESI-APCI-positive mode) at UV 220 nm. Standard recombinant DNA, molecular cloning, and microbiological procedures were performed as described (*97*).

*Cloning of the Violacein Operon by Mediated PCR.* The entire violacein operon was obtained by PCR amplification of genomic DNA from *Chromobacterium violaceum* ATCC 12472 with LA Taq polymerase. Mediated PCR reaction (*155*) was performed with the following primers: VioB-R: 5'- GCC TCT CTA GAA AGC TTT CCA CAA GCC AAA TCC AG-3' and 3275RLR 4900 bp (Seq: VioBCR: 5'- GCA GCC CAG CGC GTA GTC GTC GTA GCT GCG G-3' and 3275RLS) and 3369LFR: 5'- CGG CGA ATC GGC GAC GCC GGA TGT CGC CGG CCT GCT CGA G-3' and VioBCF: 5'- CCG CAG CTA CGA CGA CTA CGC GCT GGG CTG C-3' 4450 bp (Seq: VioC-F: 5'- GAA AAG AGC AAT CAT AGT CGG AGG CGG G-3' and 3269LFS). All PCR products which contained the whole operon were digested with MluI and XhoI and ligated into dephosphorylated Topo XL vector after digestion with MluI and XhoI restriction enzymes.

Each gene in the cluster (VioA, VioB, VioC, VioD, and VioE) was detected with following primers: VioA: 5'- CGC GGC GAT GCG CTG CAG CAG CAG-3' and 5'- GAA GCA TTC TTC CGA TAT CTG CAT TGT CGG-3'; VioB: 5'- GCC TCT CTA GAA AGC TTT CCA CAA GCC AAA TCC AG-3' and 5'- GCA TTC TGG ATT TTC

CAC GCA TCC ATT TCC G-3’; VioC: 5’- TCA GTT GAC CCT CCC TAT CTT GTA CC-3’ and 5’- GAA AAG AGC AAT CAT AGT CGG AGG CGG G-3’ ; VioD: 5’- TCA GCG TTG CAG CGC GTA GCG C-3’ and 5’- ATG AAG ATT CTG GTC ATC GGC GCG G-3’; VioE: 5’- TAT ATC TAG ACT AGC GCT TGG CGG CGA AG-3’and 5’-ATA TAA GCT TAC TAG TGA GGA GGC CGC ATG GAA AAC-3’. All PCR products were confirmed through DNA sequencing. Sequencing reactions were performed with the Perkin Elmer ABI Bigdye Terminator kit according to the manufacturer’s instructions and analyzed on an ABI31303/Genetic Analyzer by the Gene Technology Lab Core Facility of Texas A&M University.

Initially genomic DNA from C. violaceum ATCC 12462 was digested with MluI and XhoI and run on a 1% agarose gel by Pulsed Field Gel Electrophoresis (PFGE) to isolate those fragments between 5 kbp and 12 kbp with little modification to the previously reported method (*149*). After gel extraction and purification, end-repaired DNA fragments were ligated into linearized pCCIFOS, which was blunt-ended by PmlI restriction enzyme. The resulting plasmid was transformed into BL21-Codon plus (DE3) RIL *E. coli* strain. However, we did not get any colored colonies from this construction. pETBlue-2 vector was also employed to clone the whole operon using the sized DNA fragments. pETBlue-2 was digested with MluI and XhoI and then dephosphorylated after gel extraction and purification. The plasmid was transformed into Top10 and DH10B.

Vector maps for pETBlue-2, pCCIFOS, and TOPO XL are shown in Appendix D. The described expression plasmids were transformed into Top10, BL21 (DE3), and

BL21-CodonPlus (DE3) RIL competent cells according to the vector used for the each construction.

*Expression, Extraction, and Purification of Violacein in Heterologous Systems.* Violacein was extracted from *Chromobacterium violaceum* ATCC12472 as described in references (*150, 151, 156-160*). For violacein expression in *E. coli,* cells were grown at 28 °C in Luria-Bertani (LB) media after transformation into Top10 by electroporation and then plated on LB agar plates with appropriate antibiotics. After incubation for 48hrs at 28 °C, only one purple colony appeared which was confirmed to contain the whole vio operon by performing digest screening and sequencing of miniprepped DNA. The colony from the construction under Lac/Native promoter was shown in Figure 2. The selected clone was inoculated into LB media with appropriate antibiotics (constructions with Topo system used 10μg/ml Kanamycin). The bacterial cultures were spun down by centrifugation and then separated to cell and supernatant. Both samples were extracted with ethyl acetate or methanol according to previously described methods (*151, 157, 160, 161*). The extracted violacein was concentrated in vacuo to dryness. The dried samples were subjected to high performance Liquid Chromatography (HPLC) and mass spectrometry (MS) analysis. Crude MS samples were prepared in Methanol. Crude samples from the cell extract and supernatant were confirmed by Electrospray Ionization (ESI) in positive mode for mass analysis. Purification was performed with Reverse-phase HPLC using a C18 column (75mm×4.6mm, 3micron) with Acetonitrile and water as solvent according to previously published data (*151*). The sample was dissolved in 10% Acetonitrile (ACN). Elution started at 10% ACN for 4 min, followed by a linear

gradient up to 88% at 30 min and a final isocratic hold at 100% for 5 min, at a flow rate 0.5ml/min with 168-575 nm high wavelength detection. The purified sample eluted with ACN was subjected to analyze by MS, ESI data shows only M minus Hydrogen. Therefore, the HPLC condition was changed to a gradient 50% MeOH / 50% water to 100% MeOH in 15 min. For prep-HPLC MeOH/water and Acetone/water were used to obtain purified violacein. The change in solvent polarity only results in a difference in elution time. Purified violacein was identified based on HPLC/MS, UV-visible absorption characteristic, mass spectra and NMR, compared to previously published data (*148, 151, 156, 157, 161-163*).

*Characterization and Analysis of Violacein.* Violacein was characterized and analyzed by proton ($^{1}$H) NMR (300MHz, 500MHz), Carbon ($^{13}$C) and $^{13}$C DEPT135 (500MHz) NMR spectra. NMR spectra were acquired on a Bruker ARX 500 instrument using a 3mm Broad band (BB)/H probe with WALTZ decoupling for the $^{13}$C experiment. Ultraviolet-visible (UV-Vis) spectroscopy was performed with a ThermoSpectronic, model Genesys 2 spectrophotometer, and mass spectrometry with a QStar from PE Sciex for ESI and 5Voyager DE-STR from Applied Biosystems (for MALDI).

All primers and linkers used in Chapter V are shown in Table 4.

Table 4: Primers and linkers used in Chapter V

| Primer Name | Squence | Melting T. |
|---|---|---|
| **Violacein cluster detection primers** | | |
| VioA-F reverse | 5’- GAA GCA TTC TTC CGA TAT CTG CAT TGT CGG-3’ | 60.8 |
| VioA-R forward | 5’- CGC GGC GAT GCG CTG CAG CAG CAG-3’ | 71.2 |
| VioB-F | 5’- GCA TTC TGG ATT TTC CAC GCA TCC ATT TCC G-3’ | 63.1 |
| VioB-R | 5’- GCC TCT CTA GAA AGC TTT CCA CAA GCC AAA TCC AG-3’ | 64.2 |
| VioC-F reverse | 5’- GAA AAG AGC AAT CAT AGT CGG AGG CGG G-3’ | 62.5 |
| VioC-R forward | 5’- TCA GTT GAC CCT CCC TAT CTT GTA CC-3’ | 59.3 |
| VioD-F reverse | 5’- ATG AAG ATT CTG GTC ATC GGC GCG G-3’ | 63.1 |
| VioD-R forward | 5’-TCA GCG TTG CAG CGC GTA GCG C-3’ | 67.3 |
| VioE-F reverse | 5’-ATA TAA GCT TAC TAG TGA GGA GGC CGC ATG GAA AAC-3’ | |
| VioE-R forward | 5’- TAT ATC TAG ACT AGC GCT TGG CGG CGA AG-3’ | |
| 3269F | 5’- GCC GCT ACC CGG TGG CGA TAG CCG -3’ | 69.7 |
| 3275R | 5’-GGC GCA GTC GGC GGT GGC GAA GCA GAC C-3’ | 73.2 |
| VioBCF | 5’ CCG CAG CTA CGA CGA CTA CGC GCT GGG CTG C-3’ | 72.2 |
| VioBCR | 5’- GCA GCC CAG CGC GTA GTC GTC GTA GCT GCG G-3’ | 72.2 |
| 3269LFR | 5’- CGG CGA ATC GGC GAC GCC GGA TGT CGC CGG CCT GCT CGA G-3’ | 77.0 |
| 3269RLF | 5’- GCT CGA GCG CCT GGG CGC CGG CCA TGC GCT GGA CGG CCA GCC-3’ | 80.9 |
| 3269LFS | 5’- GGA CGG CCA GCC GGG CTA CCA CGG CAA CCT CTC G-3’ | 74.7 |
| 3275RLR | 5’-GTA CGC GTA CTA CAC CGC CCA GGG GCC GCT CAA TCG TTC CGC C-3’ | 74.4 |
| 3275RLS | 5’- CGT TCC GCC GAG CTG CGC CTG ATC AAT CAC AGC CG-3’ | 72.0 |
| 3275FLR | 5’- GTA CGC GTA CTG GTT GCC CCC CAC GCC GCC CCA CCA GGC-3’ | 77.1 |
| 3275FLS | 5’- GGC CCA GAA CCA GCC GGG ACG GTC GGC GCG CTC CAG C-3’ | 78.2 |
| **TPP/Riboswitch construction primers** | | |
| F: TPP+P+Ribos | 5’-GC**ACGCGT**TTGA GTT CTG CGC TGT TAA GGC GTA ATT TAC-3’ | 66.4 |
| R:TPP+P+Ribos | 5’-CG**ACTAGT**AGCTC ATT CCA AAA AGT TAA GGA CGT GG-3’ | 62.2 |
| **Oligonucleotides for the linkers** | | |
| F: MluI-T7-SpeI (For Lac+T7) | 5’- CGC GTT AAT ACG ACT CAC TAT AGG GA-3’ | 57.5 |
| R: MluI-T7-SpeI | 5’- CTA GTC CCT ATA GTG AGT CGT ATT AA-3’ | 52.8 |
| F: MluI-T7-MluI (For Lac+T7+Ribos) | 5’- CGC GTT AAT ACG ACT CAC TAT AGG GA-3’ | 57.5 |
| R: MluI-T7-MluI | 5’- CGC GTC CCT ATA GTG AGT CGT ATT AA-3’ | 57.5 |
| F: AsnI-T7-SpeI (For T7) | 5’- TAA TTA ATA GCA CTC ACT ATA GGG A-3’ | 50.5 |
| R: AsnI-T7-SpeI | 5’- CTA GTC CCT ATA GTG AGT CGT ATT AAT-3’ | 53.0 |
| F: AsnI-HpaI-HindIII-MluI | 5’- TAA TGT TAA CAA GCT TA-3’ | 39.0 |
| R: AsnI-HpaI-HindIII-MluI | 5’- CGG GTA AGC TTG TTA ACA T-3’ | 50.1 |
| F: AsnI-T7-HpaI-HindIII-MluI (For T7+Riboswitch) | 5’-TAA TTA ATA CGA CTC ACT ATA GGG GTT AAC AAG CTT A-3’ | 58.3 |
| R: AsnI-T7-HpaI-HindIII-MluI | 5’- CGC GTA AGC TTG TTA ACC CCT ATA GTG AGT CGT ATT AAT-3’ | 62.4 |
| F: MluI-HpaI-HindIII-SpeI (For Lac) | 5’- CGC GTG TTA ACA AGC TTA-3’ | 50.5 |
| R: MluI-HpaI-HindIII-SpeI | 5’-CTA GTA AGC TTG TTA ACA-3’ | 43.3 |
| F: MluI-XhoI linker (For pCCIFOS) | 5’- GCA CGC GTC GAT CTC GAG C -3’ | 61.3 |
| R: MluI-XhoI linker | 5’-GCT CGA GAT CGA CGC GTG C-3’ | 61.3 |
| F: EcoRV+MluI+SpeI (For pETBlue-2) | 5’- CGA TAT CGG ACG CGT CCA CTA GTC-3’ | 60.6 |
| R: EcoRV+MluI+SpeI | 5’- GAC TAG TGG ACG CGT CCG ATA TCG-3’ | 60.6 |
| F: EcoRV+MluI+SpeI sticky | 5’- ATC GGA CGC GTC CAC TAG TC-3’ | 58.9 |

Table 4: continued

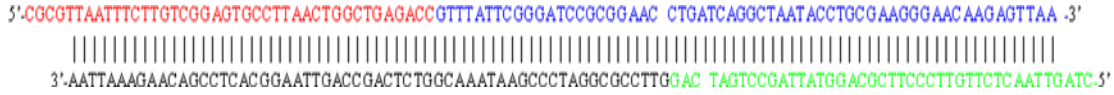
| Primer Name | Squence | Melting T. |
|---|---|---|
| R: EcoRV+MluI+SpeI sticky | 5'- CTA GTG GAC GCG TCC GAT-3' | 56.8 |
| Riboswitch linker | | |
| Ribos1 | 5'-CGC GTT AAT TTC TTG TCG GAG TGC CTT AAC TGG CTG AGA CC-3' | 67.3 |
| Ribos2 | 5'-GTTTATTCGGGATCCGCGGAACCTGATCAGGCTAATACCTGCGAAGGGAACAAGAGTTAA-3' | 69.9 |
| Ribos3 | 5'-GTTCCGCGGATCCCGAATAAACGGTCTCAGCCAGTTAAGGCACTCCGACAAGAAATTAA-3' | 70.3 |
| Ribos4 | 5' –CTA GTT AAC TCT TGT TCC CTT CGC AGG TAT TAG CCT GAT CAG-3' | 64.2 |
| LacZ construction and sequencing primers | | |
| PCR | | |
| LacZ F (SpeI) | 5'-CGC G**AC TAG T**GT TTA TCA CAG TTA AAT TG-3' | 56.3 |
| LacZ R (XbaI) | 5'- CCC C**TC TAG A**CT TAC AAT TTC C-3' | 52.4 |
| LacZ F (SacI-pBAD33) | 5'- CGC G**GA GCT C**GT TTA TCA CAG TTA AAT TG-3' | 59.8 |
| Linker | | |
| F: XbaI-Nsil-SpeI | 5'-CTA GAG CGA TGC ATG CGA-3' | 54.9 |
| R: XbaI-Nsil-SpeI | 5'-CTA GTC GCA TGC ATC GCT-3' | 55.2 |
| F: XbaI-Nsil-XhoI | 5'-CTA GAG CGA TGC ATG CGC-3' | 56.7 |
| R: XbaI-Nsil-XhoI | 5'-TCG AGC GCA TGC ATC GCT-3' | 60.5 |
| Sequencing | | |
| T7 pETBlue-2F | 5'-GTG CGG CGA CGA CCG GTG-3' | 64.8 |
| pETBlue-2 T7up Primer | 5'- TCA TAA CGT CCC GCG AAA-3' | 54.1 |
| FPCRXL-TOPO | 5'-GCA AAC CGC CTC TCC CCG CGC G-3' | 70.0 |
| RPCRXL-TOPO | 5'-CGA CGG CCA GTG AAT TG-3' | 54.0 |
| pET FP | 5'-GGA TGT GCT GCA AGG CGA TTA AGT TGG-3' | 62.0 |
| pET RP | 5'-CTC GTA TGT TGT GTG GAA TTG TGA GC-3' | 58.1 |
| RutF construction and sequencing primers | | |
| F:RutF (SpeI) | 5'-GCC **ACT AGT** ATG AAC ATT GTC GAT CAA C-3' | 57.6 |
| R:RutF (XhoI) | 5'-GGC **CTC GAG** TTA ACA AGC AGG GCG CAT CAG-3' | 67.2 |
| BLG construction primers | | |
| BLGThiCRibo(f)SpeI | 5'-GCG C**AC TAG T**AT GAA ATG TCT GCT GC-3' | 60.0 |
| BLGThiCRibo(R)XhoI | 5'- CGC G**CT CGA G**TT A<u>AT GAT GAT GAT GAT GAT G</u>AA TGT GAC AGT GTT CTT CCA GCT GGG-3' | 68.8 |

**Bold**: inserted restriction enzyme site, <u>underlined</u>: inserted 6-His tagged site

*Construction of the Violacein Gene Cluster under the Lac/Native, Native and Lac Promoter.* Lac/Native promoter was constructed first with the mediated PCR product which was confirmed by sequencing. The cloned whole operon was digested with MluI and XhoI restriction enzymes and then purified by gel extraction. The double digested PCR product was inserted into the Topo XL vector which has a Lac operon. This is the p(v4(MluI-XhoI) Topo) vector which contains Lac/Native + vio operon. To construct the vector containing Native promoter only, the p(v4(MluI-XhoI) Topo) vector was digested with AsnI and MluI restriction enzymes (removing the Lac promoter) and then treated with phosphatase (CIP). Lac operon is located at base pairs 95-216, AsnI cuts at base pairs 48 and 107 and MluI cuts at site 248. All vector maps including manipulating sites are shown in Appendix. Gel purified digested vector was ligated to the synthesized linker (F: AsnI-HpaI-HindIII-MluI: 5'- TAA TGT TAA CAA GCT TA-3' and R: AsnI-HpaI-HindIII-MluI: 5'- CGG GTA AGC TTG TTA ACA T-3') with T4 DNA ligase. Construction of violacein gene cluster under Lac promoter control was performed with ligation of linearlized p(v4(MluI-XhoI) Topo) vector which was digested with MluI and SpeI restriction enzymes (removes the native promoter) and then treated with phosphatase, and the linker which was synthesized with F: MluI-HpaI-HindIII-SpeI :5'-CGC GTG TTA ACA AGC TTA-3' and R: MluI-HpaI-HindIII-SpeI :5'-CTA GTA AGC TTG TTA ACA-3'. All constructions were transformed into Top10 *E. coli* strain except Lac promoter controlled vector which was transformed into Top10 and BL21 (DE3) *E. coli* strain.

*Construction of Violacein Gene Cluster under the Lac/T7 Promoter and T7 Promoter in Topo System.* The vector containing Lac/T7 promoter was constructed with MluI and SpeI double digested p(v4(MluI-XhoI) Topo) vector which was dephosphorylated and the linker which was synthesized with F: MluI-T7-SpeI: 5'- CGCGTTAATACGAC TCACTATAGGGA-3' and R: MluI-T7-SpeI: 5'- CTAGTCCCTATAGTGAGTCGT ATTAA-3'. To construct T7 promoter control vector, pv4(MluI-XhoI) Topo vector was linearized by digestion with AsnI and SpeI restriction enzymes. Ligation was followed with the linker which was produced with F: AsnI-T7-SpeI: 5'- TAA TTA ATA GCA CTC ACT ATA GGG A-3' and R: AsnI-T7-SpeI: 5'- CTA GTC CCT ATA GTG AGT CGT ATT AAT-3'. All constructions were transformed into Top10 *E. coli* strain.

S*ynthesis of Riboswitch Linker*. The sequences of the TPP riboswitch sequence were compared with *E. coli* str. K12 substr.DH10B (CP000948) of the reference paper Cheah *et al*. (8). The sequence is shown in Appendix D. Riboswitch was synthesized with oligonucleotides as shown in below (Ribos1: 5'-CGC GTT AAT TTC TTG TCG GAG TGC CTT AAC TGG CTG AGA CC-3'; Ribos2 : 5'- GTT TAT TCG GGA TCC GCG GAA CCT GAT CAG GCT AAT ACC TGC GAA GGG AAC AAG AGT TAA-3' Ribos3 : 5'-GTT CCG CGG ATC CCG AAT AAA CGG TCT CAG CCA GTT AAG GCA CTC CGA CAA GAA ATT AA-3'; Ribos4: 5' –CTA GTT AAC TCT TGT TCC CTT CGC AGG TAT TAG CCT GAT CAG-3' ). Reaction conditions for linker synthesis are described in Appendix D (Figure A.14).



*Construction of Violacein Gene Cluster under the Lac/Riboswitch, Lac/T7/Riboswitch and T7/Riboswitch Promoter.* To construct Lac/Riboswitch controlled vector, a synthesized riboswitch linker was ligated to digested p(v4(MluI-XhoI) Topo) vector with MluI and SpeI restriction enzymes. The construction was transformed into Top10 *E. coli* strain. This vector was used for manipulation of riboswitch containing vector, named as a p(RV8topo) vector. To prepare the Lac/T7/Riboswitch controlled vector, p(RV8topo) vector was digested with MluI restriction enzyme and then purified with gel extraction kit after dephosphorylation. The linearized p(RV8topo) vector was ligated to the linker which was synthesized with F: MluI-T7-MluI: 5'- CGC GTT AAT ACG ACT CAC TAT AGG GA-3'and R: MluI-T7-MluI: 5'- CGC GTC CCT ATA GTG AGT CGT ATT AA-3'. Transformation followed after ligation reaction. Top 10 *E. coli* strain was used for this construction. Construction of the violacein gene cluster under the T7/Riboswitch promoter was performed with ligation of AsnI and MluI restriction digested p(RV8topo) vector, which was dephosphorylated, and the linker. The linker was synthesized with F: AsnI-T7-HpaI-HindIII-MluI: 5'-TAA TTA ATA CGA CTC ACT ATA GGG GTT AAC AAG CTT A-3' and R: AsnI-T7-HpaI-HindIII-MluI: 5'-CGC GTA AGC TTG TTA ACC CCT ATA GTG AGT CGT ATT AAT-3'. The plasmid was transformed into BL21 *E. coli* strain (DE3 and CP RIL respectively).

*PCR Product of the ThiC/Riboswitch Promoter*. The gene sequence of the ThiC promoter for thiamine pyrophosphate protein was identified from the region 4291721 to 4294220 in the DH10B *Escherichia coli* strain after generation of complementary sequence (*164, 165*). The sequence was analyzed to design the primers for PCR reaction

as shown in Appendix with indication of Riboswitch part in red color, promoter region in blue color, start codon in pink italic font, and transcription part in green. Mini-prepped DH10B was used as template for PCR to obtain the ThiC/Riboswitch PCR product (288bp) with primers; 5'-GCA CGC GTT TGA GTT CTG CGC TGT TAA GGC GTA ATT TAC-3' and 5'-CGA CTA GTA GCT CAT TCC AAA AAG TTA AGG ACG TGG-3'. Red colored sequence containing MluI restriction site was modified (C to G) to protect from digestion with MluI restriction enzyme. PCR conditions are described in Appendix D. After gel extraction, the PCR product was confirmed by sequencing.

*Construction of Violacein Gene Cluster under the Lac/ThiC/Riboswitch and ThiC/Riboswitch.* ThiC+Riboswitch PCR product was digested with MluI and SpeI restriction enzyme and then heat inactivation followed at 80°C for 20 min. Dephosphorylated p(v4(MluI-XhoI) Topo) vector after double digestion with MluI and SpeI was employed for construction of the Lac/ThiC/Riboswitch controlled vector. Ligation reaction was performed at 16°C overnight and then 2 hrs more incubation at room temperature. The plasmid was transformed into Top 10 and BL21 *E. coli* strains. Construction including ThiC/Riboswitch promoter was performed with dephosphorylated Lac/ThiC/Riboswitch controlled vector after double digestion with AsnI and MluI restriction enzymes (removes the Lac promoter portion), and then ligation with a linker. The linker was synthesized with following oligonucleotides; F: AsnI-HpaI-HindIII-MluI 5'- TAA TGT TAA CAA GCT TA-3' and R: AsnI-HpaI-HindIII-MluI 5'- CGG GTA AGC TTG TTA ACA T-3'. The construction was transformed into Top 10 *E. coli* strain and BL21 *E. coli* strains (DE3 and CP RIL).

*Construction of Violacein Gene Cluster under the T7/Native, T7, Native and T7/Riboswitch Promoter in pCCIFOS System.* T7 promoter in pCC1FOS was used to manipulate the violacein operon containing vector. pCC1FOS vector map is attached in Appendix D. PmlI (an isoschizomer of Eco72 I) was employed to generate linearized pCCIFOS vector with a blunt end. HpaI and PmlI were used to remove T7promoter region. For the T7/Native promoter system, genomic DNA of *Chromobacterium violaceum* and was digested with MluI and XhoI and then end repaired as described earlier. Ligation reaction was performed with fast-link DNA ligase. The resulting plasmid was transformed into BL21 *E. coli* strain (DE3 and CP RIL respectively). Many colonies grew after 16hrs, but no violet - colored colonies were observed. An identical procedure was performed for construction of the T7 promoter controlled vector with SpeI and XhoI digested *vio* operon fragment which was end repaired after digestion. For the Native promoter, pCCIFOS vector was digested with HpaI and PmlI restriction enzymes to remove T7 promoter. The linearized vector with HpaI and PmlI was ligated with the linker (synthesized with 5'- GCA CGC GTC GAT CTC GAG C -3' and 5'- GCT CGA GAT CGA CGC GTG C-3'). The resulting plasmid was transformed into BL21 *E. coli* strain (CP RIL) and Top 10 *E. coli* to amplify the vector containing MluI and XhoI linker. Mini-prepped plasmid was dephosphorylated after digestion with MluI and XhoI restriction enzymes, and then ligated with MluI and XhoI digested DNA fragments of violacein operon from different sources, such as genomic DNA, PCR product and MluI and XhoI digested DNA fragment of p(v4(MluI-XhoI) Topo) vector. An alternative method for construction of native promoter controlled vector containing *vio* operon was usage of end-it repair enzyme mix. After removing the T7 promoter, end

repaired insert DNA was ligated to the vector. For preparation of T7/Riboswitch controlled vector, p(RV8topo) vector was digested with MluI and XhoI restriction enzymes to get riboswitch and *vio* operon without native promoter. The resulting DNA fragment was end repaired to remove sticky ends. pCC1FOS vector was digested with PmlI restriction enzyme and dephosphorylated. Ligation reaction was performed at 16°C overnight and then transformation followed.

*Construction of Violacein Gene Cluster under the T7/Native, T7, and T7/Riboswitch Promoter in pETBlue-2 System.* pETBlue-2 vector was manipulated with restriction site of EcoRV, MluI, and XhoI in multiple cloning site. The map of pETBlue-2 vector is attached in Appendix D. For construction of T7/Native promoter controlled vector, MluI and XhoI double digested pETBlue-2 vector was prepared as described earlier and then ligation was performed with 5-12kbp sized genomic DNA, PCR product and MluI and XhoI digested DNA fragment of p(v4(MluI-XhoI) Topo) vector, respectively. The resulting plasmid was transformation into Top10 *E. coli* strain. After digestion of pETBlue-2 vector with EcoRV and XhoI restriction enzyme, the synthesized linker and SpeI and XhoI digested DNA fragment of p(v4(MluI-XhoI) Topo) vector were ligated to it for construction of T7 promoter controlled vector containing violacein operon without the native promoter. The linker was synthesized with following oligonucleotides; 5'-ATC GGA CGC GTC CAC TAG TC-3' and 5'- CTA GTG GAC GCG TCC GAT-3'. For preparation of T7/Riboswitch controlled vector, Lac/Riboswitch promoter, p(RV8topo) vector was digested with MluI and XhoI restriction enzyme to obtain riboswitch and vio operon without native promoter. pETBlue-2 was digested with MluI

and XhoI restriction enzymes. Ligation reaction was performed at 16°C overnight and then transformation followed.

*Construction of ThiC Riboswitch/β-galactosidase System (LacZ+ThiC+Riboswitch) and Arabinose/β-galactosidase System (LacZ+pBAD33).* The LacZ multiple cloning site and open reading frame site were amplified from the plasmid pETBlue-2 with primers, LacZF (SpeI) (5'-CGC GAC TAG TGT TTA TCA CAG TTA AAT TG-3', the SpeI site is underlined) and LacZR (XbaI) (5'- CCC CTC TAG ACT TAC AAT TTC C-3', the XbaI site is underlined), and were tagged with restriction enzyme sites SpeI and XbaI, respectively. The desired PCR product was obtained after gel extraction with QIAquick gel extraction kit from Qiagen. The DNA fragment was digested with SpeI and XbaI, gel extracted, and subsequently ligated to the ThiC/Riboswitch promoter containing TOPO vector, which was linearized with SpeI and XhoI restriction enzymes and then phosphatase treated. The phosphatase treated vector was ligated to the liker with T4 DNA ligase. The sticky end linker was synthesized with F: XbaI-Nsil-XhoI (5'-CTA GAG CGA TGC ATG CGC-3', the XbaI site and XhoI site are underlined) and R: XbaI-Nsil-XhoI (5'-TCG AGC GCA TGC ATC GCT-3' the XbaI site and XhoI site are underlined). The resulting plasmid was transformed into Top10 *E. coli* strain. The resulting colonies were miniprepped, screened by blue and white colony, and the blue clone confirmed through sequencing.

LacZ construct containing LacZ+pBAD33 promoter was performed as described above. The vector pBAD33 was digested with XbaI and SacI and phosphatase treated with CIP. The phosphatase treated vector was ligated to the PCR product, which was

digested with XbaI and SpeI, and linker with T4 DNA ligase. PCR reaction was performed with LacZ F (SacI-pBAD33) (5'- CGC GGA GCT CGT TTA TCA CAG TTA AAT TG-3', the SacI site is underlined) and LacZ R (XbaI) (5'- CCC CTC TAG ACT TAC AAT TTC C-3', the XbaI site is underlined). This combination did not work well; pBAD33 cut with XbaI only due to SacI can not cut properly in the system. The DNA fragment for LacZ+ThiC+Riboswitch was also used for pBAD33. The sticky end linker was synthesized with F: XbaI-Nsil-SpeI (5'-CTA GAG CGA TGC ATG CGA-3', the part of XbaI and SpeI sites are underlined) and R: XbaI-Nsil-SpeI (5'-CTA GTC GCA TGC ATC GCT-3', the part of SpeI and XhoI sites are underlined). The insertion was confirmed by sequencing.

*Construction of RutF Protein Expression Vector containing ThiC/Riboswitch Promoter.* The PCR product was obtained with F:RutF(SpeI) and R:RutF(XhoI) primers:

F:RutF(SpeI) 5'-GCC **ACT AGT** ATG AAC ATT GTC GAT CAA C-3'

R:RutF(XhoI) 5'-GGC **CTC GAG** TTA ACA AGC AGG GCG CAT CAG-3'

The desired PCR product was obtained after gel extraction with QIAquick gel extraction kit from Qiagen. The DNA fragment was digested with SpeI and XhoI, gel extracted, and subsequently ligated with T4 DNA ligase to the ThiC/Riboswitch promoter-containing Topo vector, which was linearized with SpeI and XhoI restriction enzymes and then phosphatase treated**.** The resulting plasmid was transformed into Top10 *E. coli* cells. The resulting colonies were miniprepped and confirmed through sequencing.

*Construction of β-γ-Globulin Protein Expression Vector containing ThiC/Riboswitch Promoter.* The PCR product was obtained with BLGThiCRibo(F)SpeI and BLGThiCRibo(R)XhoI;

F: 5'-GCG CAC TAG TAT GAA ATGTCTGCTGC-3'

R: 5'-CGCGCTCGAGTTAATGATGATGATGATGATGAATGTGACAGTGTTCTTCCAGCTGGG-3'

Phusion Taq polymerase and Taq polymerase were employed to produce optimal PCR product. The desired PCR product was obtained after gel extraction with QIAquick gel extraction kit from Qiagen. The DNA fragment was digested with SpeI and XhoI, gel extracted, and subsequently ligated with T4 DNA ligase to the ThiC/Riboswitch promoter containing vector which was linearized with SpeI and XhoI restriction enzymes and then phosphatase treated**.** The resulting plasmid was transformed into Top10 *E. coli* strain. The resulting colonies were miniprepped, screened by blue and white colony, and the blue clone confirmed through sequencing.

*β-Galactosidase Assay.* The assay was modified from the protocol of S. Maloy lab by J. Boylason. The cells were grown in LB media with appropriate antibiotics in a 37°C incubator at 250 rpm. After incubation for an appropriate time (see the result part), the culture OD at 630 nm was measured. After centrifugation, the cells were resuspended in Z-buffer which contained $Na_2HPO_4$, $NaH_2PO_4{\cdot}H_2O$, KCl and $MgSO_4{\cdot}7H_2O$ (pH=7.0) to an $OD_{630}$ = 0.3~0.4. Z-buffer was supplemented with β-mercaptoethanol just before use. To permeabilize the cells, 10µl of 0.1% SDS and 20µl of chloroform were added to 750µl of cells resuspended in Z-buffer. After 15-20 sec vortexing, the cells were incubated at room temperature for 15 min. 250µl of permeabilized cells and 250µl of Z-

buffer were added into epp-tubes in triplicate and then 125µl of 10mg/ml o-nitrophenyl-β-D-galactopyranoside (ONPG) in Z-buffer was added and incubated until a light yellow color developed. To stop the reaction 250µl of 1M of $Na_2CO_3$ was added to each reaction tube and the time recorded. Centrifugation was performed at 13000 rpm for 10 min at 4°C. Absorbance was checked at 450 nm. Units of β-galactosidase were calculated by following equation.

$$units\ of\ \beta - galactosidase\ = \frac{1000 \times OD450}{Min \times 0.25ml \times OD630}$$

*TPP Titrability Test with Constructions Controlled under Lac/Riboswitch, ThiC/Riboswitch, LacZ/ThiC/Riboswitch and LacZ/pBAD33*. Thianine Pyrophospate Chloride (TPP) was used to check titrability of Riboswitch, ThiC/Riboswitch promoter with various promoters and operons. TPP was added with following final concentration; 0.0 mM, 0.5 mM, 1.0 mM, 1.5 mM, 2.0 mM, 2.5 mM, 3.0 mM, and 5.0 mM. After appropriate day culture, 100µl of culture was inoculated into fresh LB containing appropriate antibiotic and different amount of TPP. For the violacein production level with various TPP concentrations, HPLC was employed for the detection. The β-galactosidase assay was used to check production of β-galactosidase by LacZ with ThiC/Riboswitch or pBAD33 promoter. β-galactosidase assay was modified by adding additional TPP in Z-buffer.

*Solubility Test with SDS-PAGE Gel and Western Blot Gel*. To prepare soluble sample, 1ml of culture was centrifugated for 1min at 13,000 rpm. The supernatant was discarded

and replaced with 1ml of Tris (0.2mM, pH8.0) and 100µl of 10% SDS. After adding a small amount of glass beads, the tube was vortexed for 5min. After 10 min centrifugation the supernatant containing solubilized protein sample was obtained. After washing the cell pellet with Tris solution, 100µl Tris buffer was added to the same volume of protein loading dye (2x) and heated for 10min at 90°C. The cell sample which contained non-soluble protein for the SDS-PAGE gel after centrifugation for 10 min at 14,000 rpm. SDS- Polyacrylamide Gel Electrophoresis (SDS-PAGE) gel was run using PAGE-SDS Laemmli standard protocol (*166*). To run Western blot gel proteins were resolved on 12% SDS-PAGE gels and electrophoretically transferred to nitrocellulose and subjected to western immunoblot analysis using goat anti-mouse IgG/ HisTag monoclonal antibody detection system (from Novagen) as previously described (*167*).

**RESULTS AND DISCUSSION**

*Cloning of the Known Violacein Cluster from Chromobacterium violaceum.* The v*io* operon was obtained from PCR products. As shown in Figure 42, the plasmid with Lac/Native promoter was used to induce the violacein production. The *vio* operon has high G+C content (average : 68%) (*58, 168*). This feature was a major obstacle to obtaining the *vio* operon by PCR. A PCR product was obtained with LaTaq DNA polymerase along with GC buffer I. The entire *vio* operon (~8.9kb size) was achieved by mediated PCR (*155*). Mediated PCR products were obtained with the primer BR and 3275RLR (4.9kb) and 3269LFR and BCF (4.45Kb) as described in experimental procedures. PCR conditions were the following: 0.5µl of Takara La Taq polymerase,

25µl of 2x GC buffer II, 8µl of dNTP, 4µl of Genomic DNA, 1µl of Forward primer, 1µl of Reverse primer and 10.5µl of deionized water in total 50µl reaction volume. PCR was run at step 1= 94°C 1min, step 2=94°C for 30sec, step 3=60°C for 30sec, step 4=72°C for 10min step 5= go to step 2, 30cycles step 6=72°C for 5min step 7= 4°C hold. The resulting PCR products were confirmed by sequencing as shown in Appendix D.

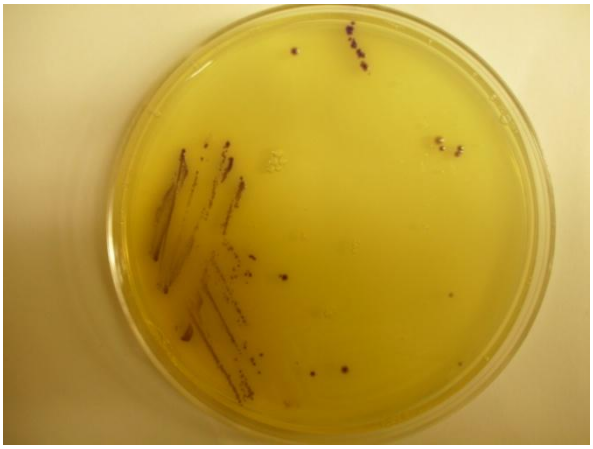


Figure 42: Violacein expression in *E. coli* with Lac/ Native promoter. One purple colony showed up at 28°C on LB+ 10µg/ml Kanamycin agar plate, and then streaked onto new agar plate to verify expression.

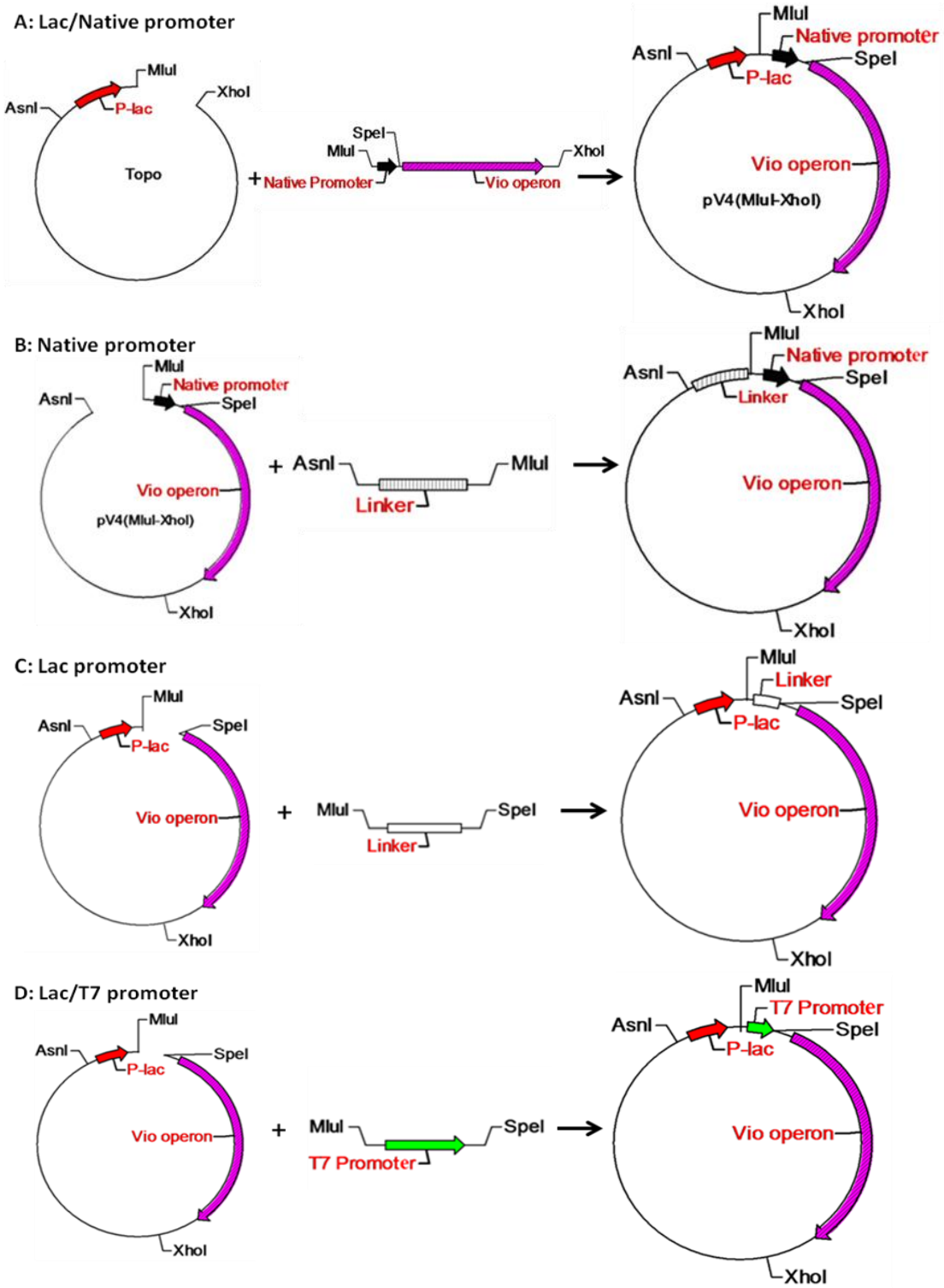
After PCR products were obtained (4.9 Kb (M1) and 4.45 kb (M2) size), gel purified PCR products were used as templates to obtain the entire *vio* operon with 8.9 kb size. To run the mediated PCR with M1 and M2, PCR conditions were as follows: 0.5µl of Takara LaTaq DNA polymerase, 25µl of 2x GC buffer II, 8µl of dNTP, 4µl of M1 PCR product, 4µl of M2 PCR product, 0.4µl of Forward (3269LFR) primer, 0.4µl of Reverse (3275RLR) primer and 7.7 µl of deionized water in total 50µl reaction volume. PCR was run utilizing the following conditions: step 1=98°C for 1 min, step 2=98°C for 20

sec, step 3=60°C for 30 sec, step 4=67°C for 15 min step 5= go to step 2, 30cycles step 6=72°C for 10 min step 7= 4°C hold. Previously reported results, from the Walsh group indicated only 2-violet colored colonies were obtained among thousands of colonies with the vector pET24b (*169*). Another group also reported the *vio* locus without native promoter, the SpeI-XhoI fragment, which expressed violacein under the ermE*p promoter in *S. albus (151)*. However, no violacein cluster was obtained with the methods under T7 promoter of pCC1FOS and pETBlue-2 in *E .coli*. Mediated PCR product was confirmed by sequencing with several combinations of primers to evolve the product contained vioA, vioB, vioC, vioD and vioE genes in their entirety. Only one violet-colored colony was obtained from the mediated PCR product after ligation with Topo vector and then transformation into Top 10 *E. coli* strain. This construction was used as the mother vector, named as pV4(MluI-XhoI) (Scheme 5(A)).

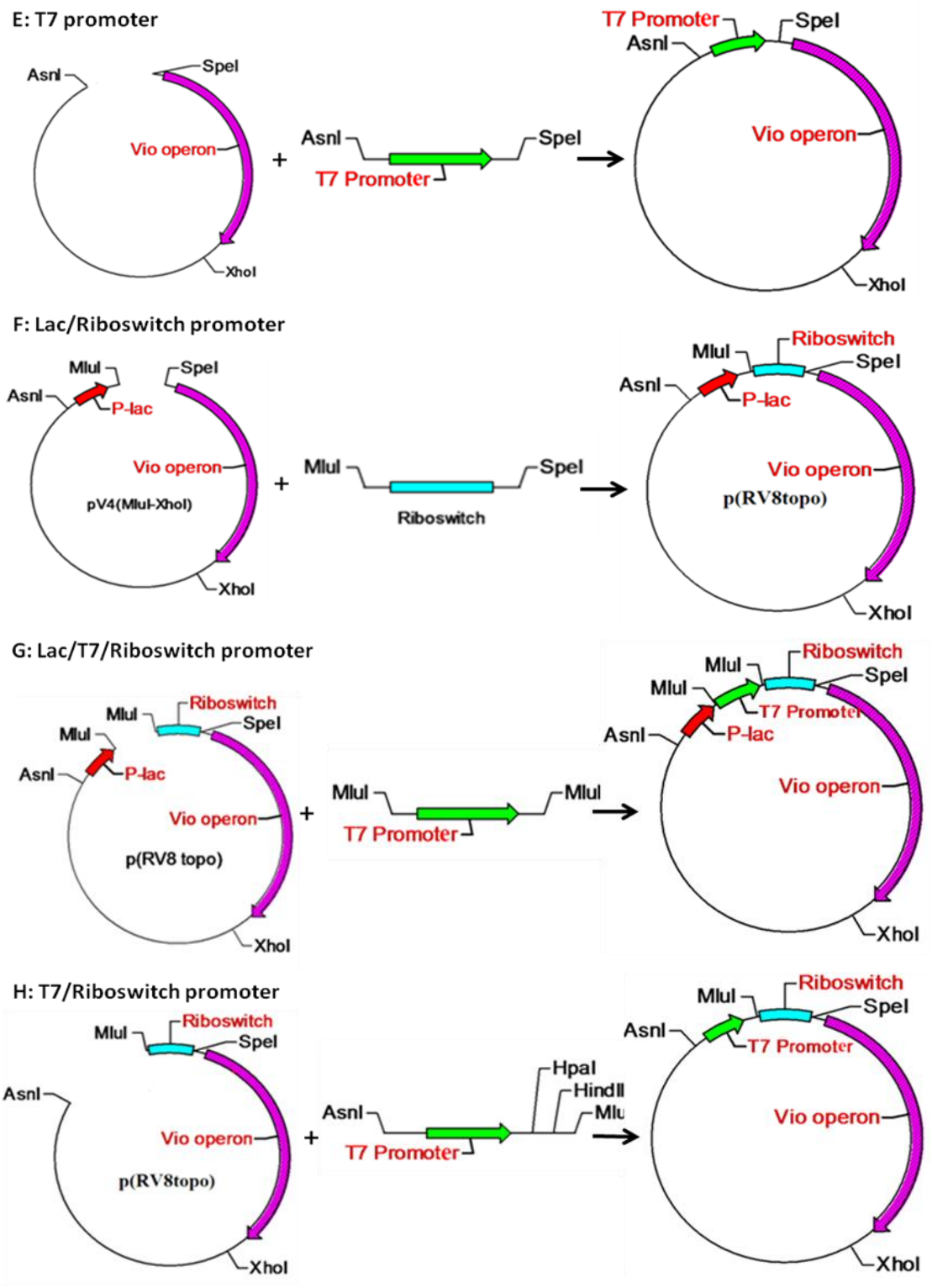
*Construction of Violacein Gene Cluster under Various Promoters.* All 17 different constructions were done by manipulation of the three different vector systems Topo XL, pETBlue-2 and pCC1FOS. The illustration of the constructions is shown in Scheme 5. Only six constructions in the Topo system expressed the violacein violet-colored colonies: Lac/Native, Lac/Riboswitch, Lac/T7, Lac/ThiC/Riboswitch, ThiC/Riboswitch and Lac promoter. T7 promoter failed to express violacein in *E. coli* but a combination of Lac/T7 promoter did. Because Lac promoter can express violacein by itself, Lac/T7 promoter construction can be controlled by Lac promoter not the T7 promoter. We wondered why the T7 promoter failed to express violacein in 3 different common vector systems. One possible explanation is that the T7 promoter is blocked by unidentified

genes. To address this issue, the sequence of the violacein operon was checked for a T7 terminator. A terminator-like sequence was found in bp7932 to 7978 (*170*), which is located after the native promoter site. As shown in Figure 1, the native promoter is located in between an MluI and SpeI site. This means the terminator-like sequence remained after removing the native promoter. The sequence could form a hairpin-like shape, which is the common shape of the terminator. Detailed data is shown in Appendix D.

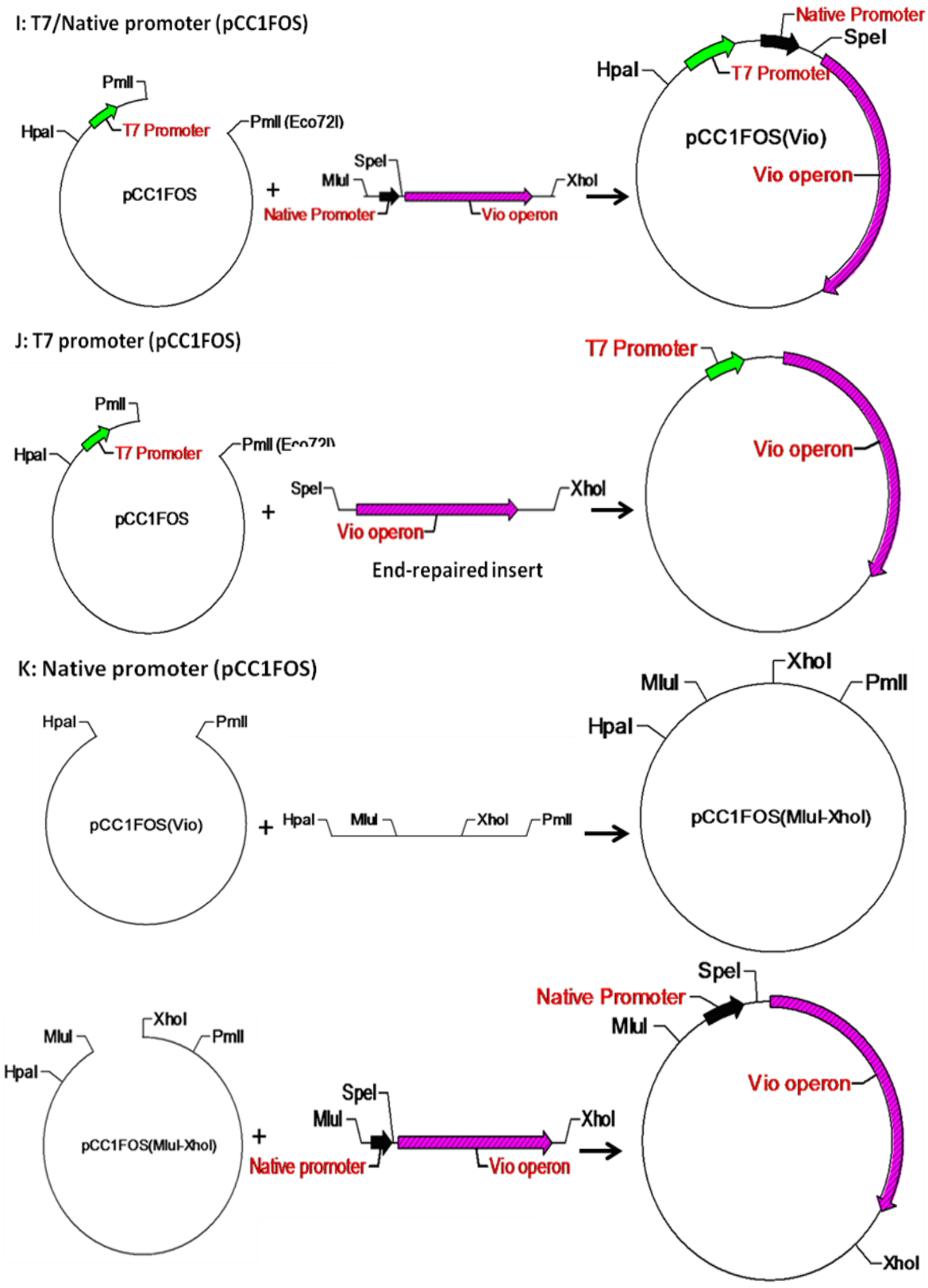
Complementary experiments were performed with RutF protein to check expression of the protein under a vector in which expression is controlled by T7 promoter. The RutF gene was obtained by PCR. PCR product was purified and digested with XhoI and SpeI. Double digested PCR product was ligated with linker (MluI-SpeI) into pETBlue-2 vector which was digested with MluI and XhoI. Transformation into Top 10 *E. coli* strain was followed. A protein of the expected size was observed on SDS-PAGE gel. The data is shown in Appendix D. Based on the experiment, we can conclude that our construction strategy and vector promoter is functional. However, further experiments need to be performed to confirm the hypothesis that the violacein sequence contains a similar sequence to the T7 terminator which can block the initiation of transcription by a T7 promoter.



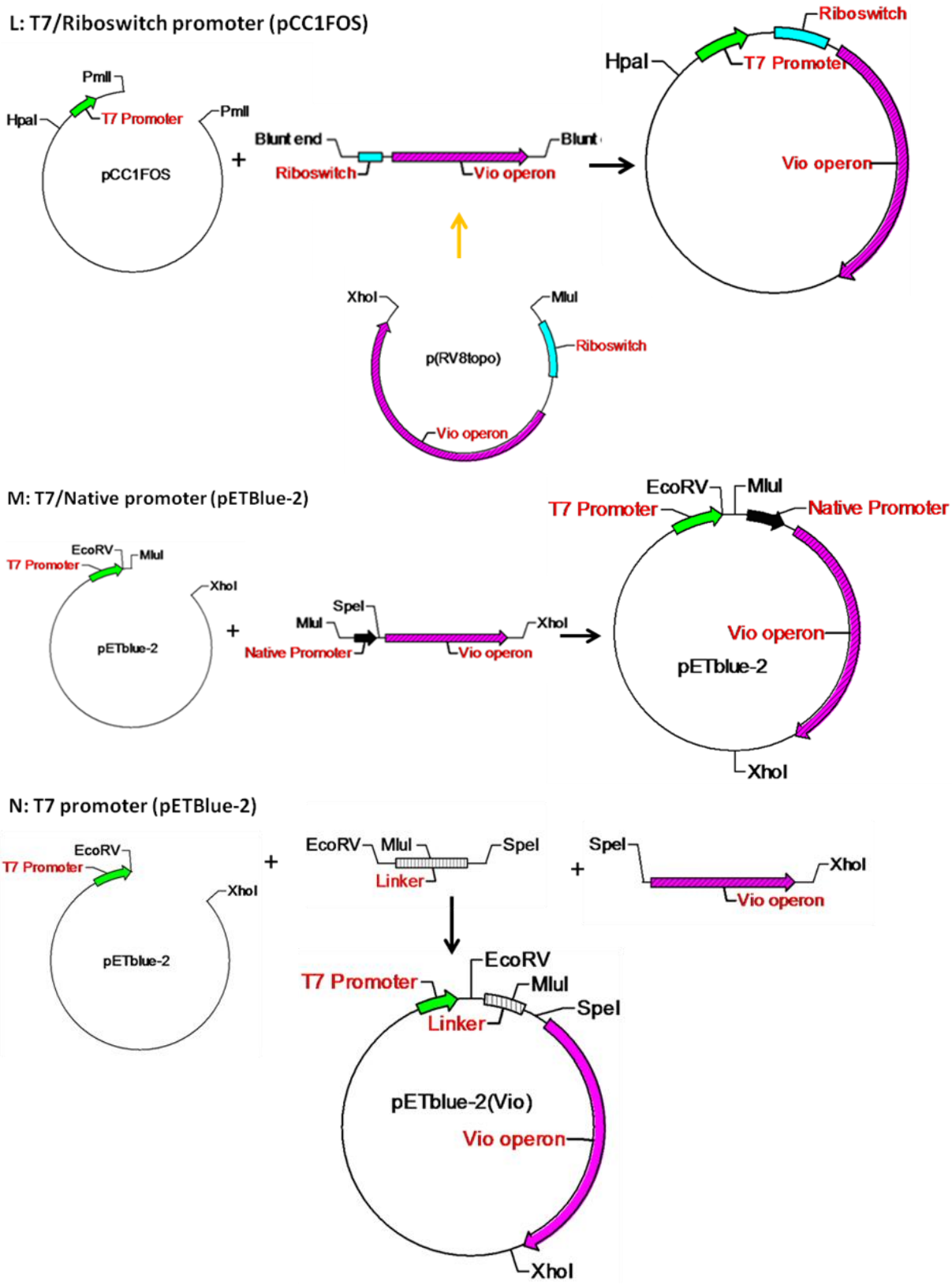

Scheme 5: Constructions of violacein cluster under the various promoters in Topo, pCC1FOS, and pETblue-2 vector. Construction M (T7/Native promoter in pETblue-2) used 3 different DNA fragments.



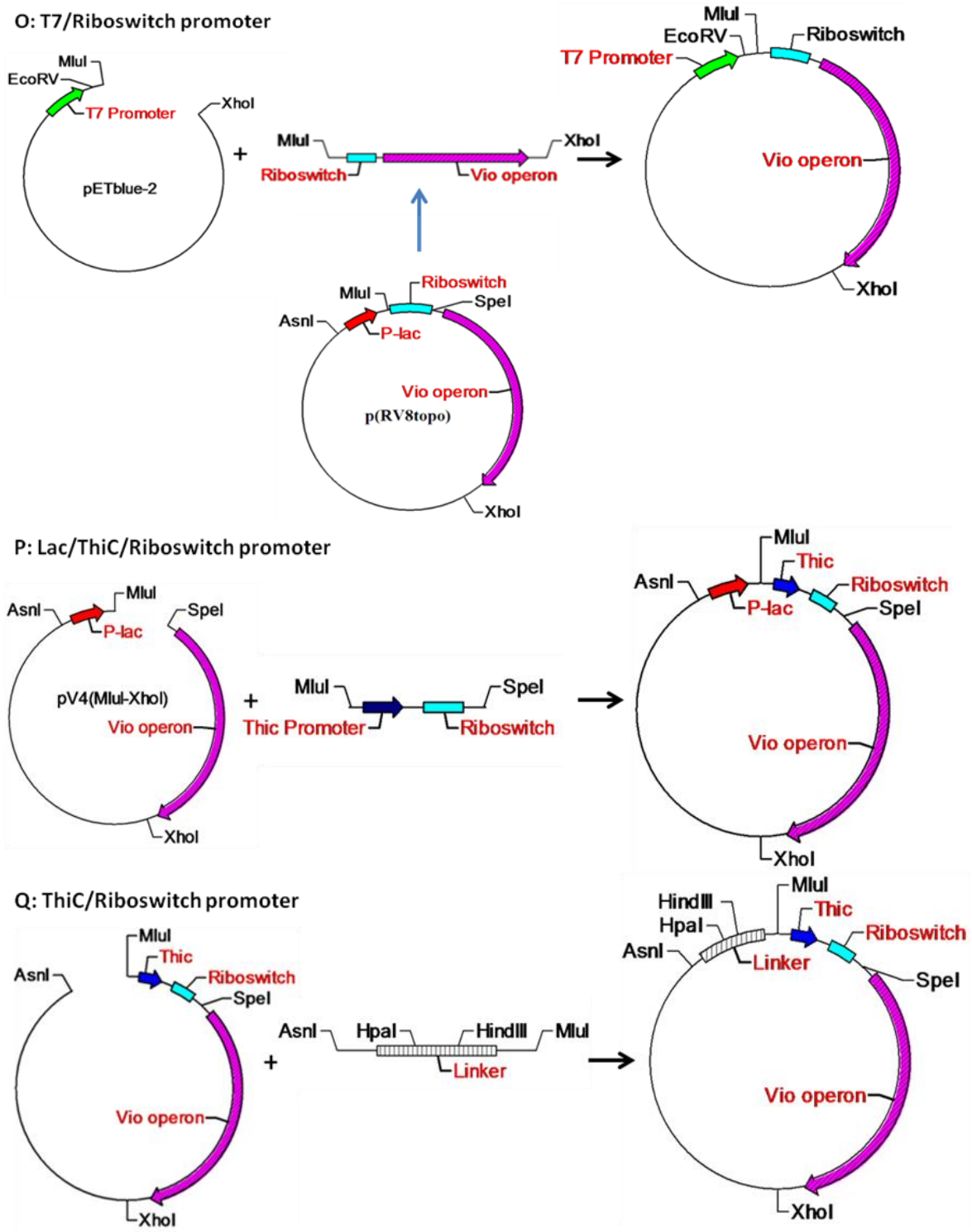

Scheme 5: continued




Scheme 5: continued




Scheme 5: continued




Scheme 5: continued

*Purification and Characterization of Violacein (UV-VIS/MS/ HPLC/LCMS/NMR).* Violacein was extracted as described in the experimental procedure section. *C. violaceum* was grown in Nutrient Broth (Difco) (*171*) at 28°C. The violacein-containing butanol phase was obtained by extraction with water-saturated butanol as previously described in the literature (*157, 160, 161*). The extract was scanned over the visible wavelengths of light; UV-VIS spectrum shows strong absorption at the visible region (585nm) due to resonance of violacein. The UV-VIS spectrum of violacein after purification with preparative HPLC is shown in Figure 43. The UV-VIS spectrums of crude violacein expressed in *E. coli* and *C. violaceum* are shown in Appendix D. All UV-VIS data matched previously reported data (*157, 160*). Crude violacein was subject to mass spectrometry as shown in Figure 44.

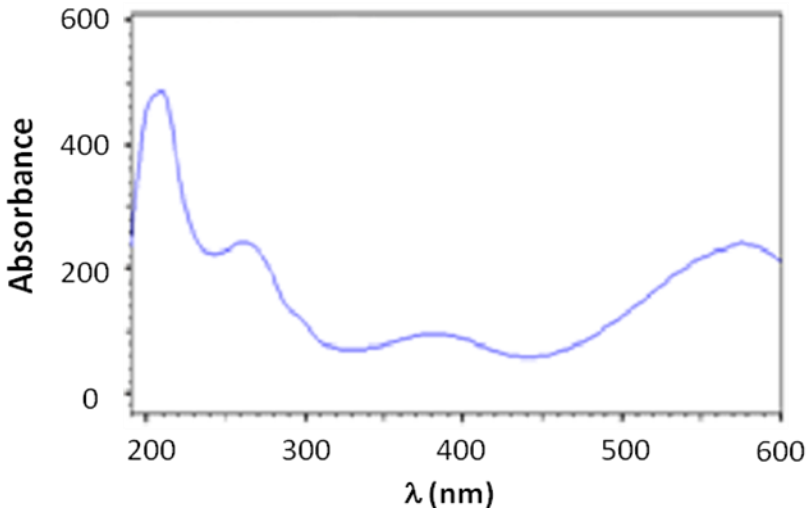



Figure 43: UV-VIS spectrum of violacein. After ran HPLC with major peak at 6.83min.

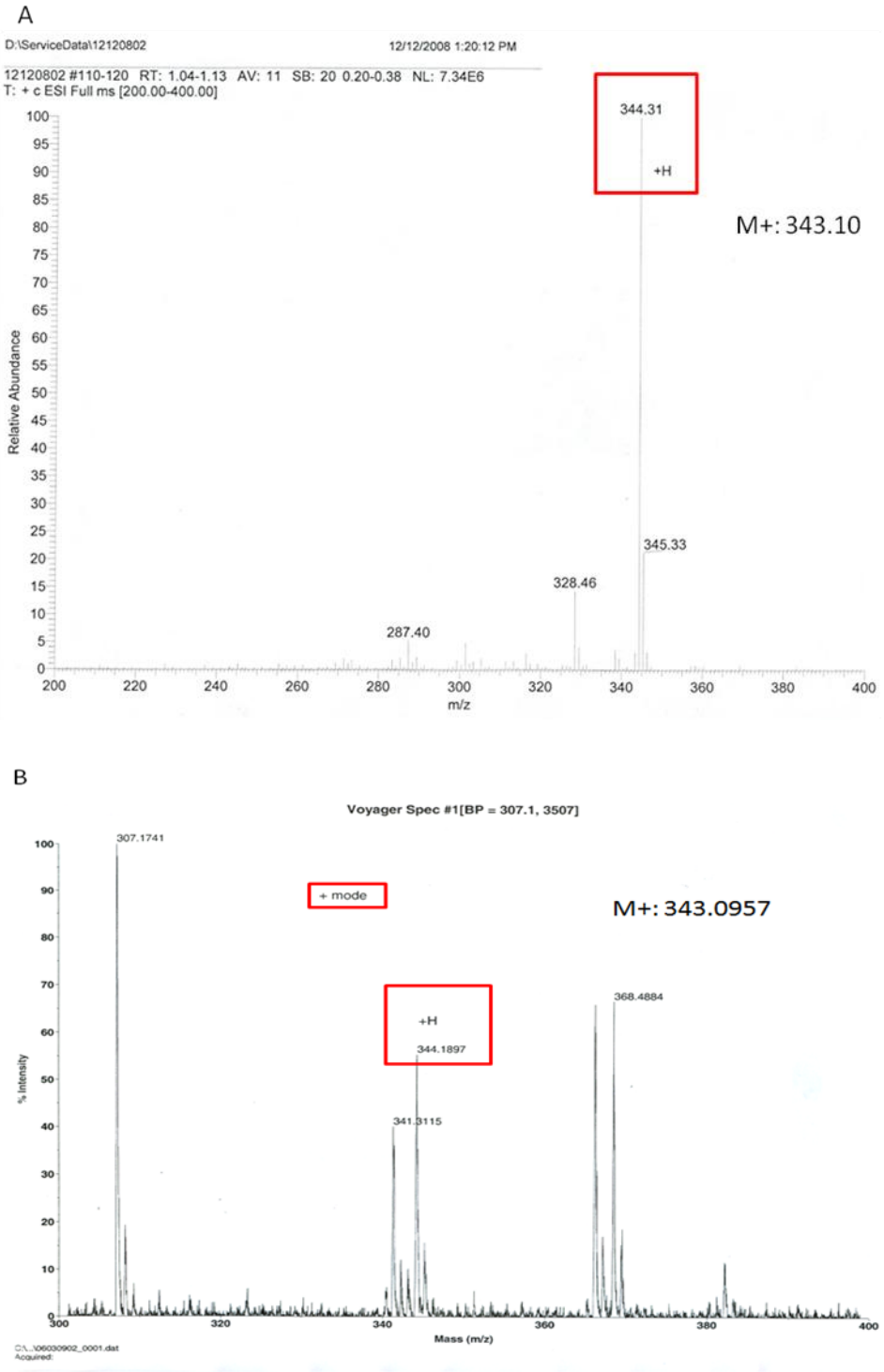


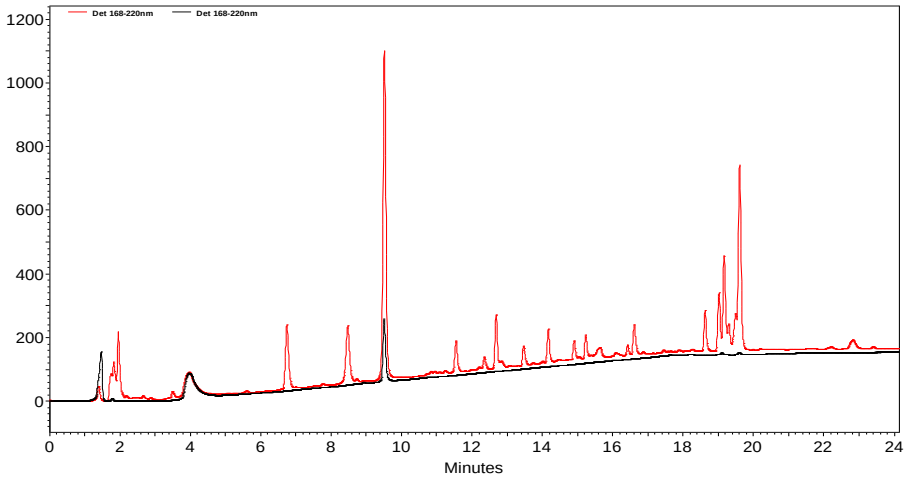

Figure 44: Mass spectrometry data. (A)MS data of violacein from cell extract with ESI in positive mode. (Expressed in *E. coli*) (B) MS data of violacein, after purification of HPLC in Acetonitrile, with MALDI in positive mode.

MS data from methanol extraction shows the expected M+H peak, but sample dissolved in Acetonitrile following HPLC purification, shows M+ in positive mode with ESI Mass spectrometry. To obtain the correct MS data from the sample dissolved in Acetonitrile, it was dried and resubjected to scan ESI MS in methanol. The data shows that the mass of the hydrogen in positive mode was still missing. Based on the MS data, HPLC solution was changed to Methanol or Acetone. The missing proton was detected by MALDI MS as shown in Figure 44 (B).

UV-VIS and MS data confirmed the violet pigments are violacein. The methanol extract was further purified using reverse phase preparative HPLC.

To optimize the HPLC condition, reported conditions were first applied (4, 6, 10, 13). Optimized conditions were applied to preparative HPLC. HPLC spectra of violacein are shown in Figure 45. Crude sample and pure sample were run on an analytical HPLC with detection at 220 nm and 575 nm. Due to the color of violacein, long wavelength was required to distinguish between violacein and deoxyviolacein. Violacein was eluted first and then a mixture of trace violacein and deoxyviolacein was eluted (Figure 45B). Each fraction was analyzed by LCMS. To further enrich samples for violacein, alternative solvents to dissolve violacein were tested. From tests with acetone and 4,5,5-trimethylfuran-2-one (TMF), TMF was shown to be the best solvent to increase solubility of violacein.

A




B

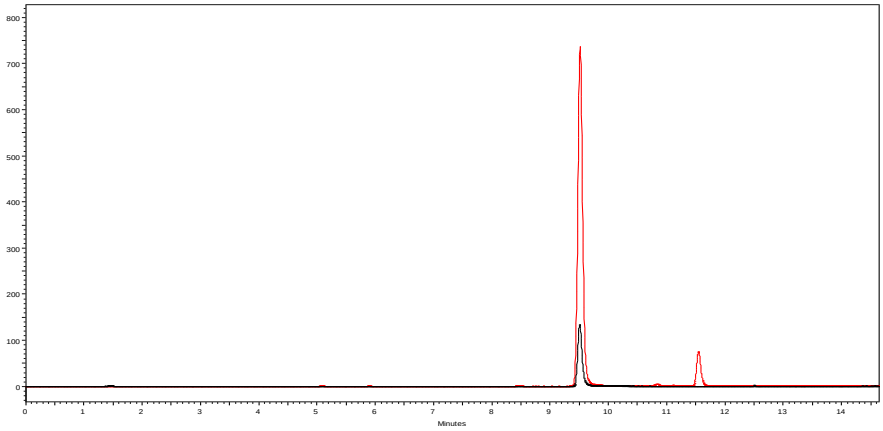



Figure 45: HPLC spectrums of violacein crude sample (red trace) and violacein after preparative HPLC (black trace) column: Gemini C18, 150X4.6mm (MeOH/water) Gradient: 50%B (MeOH) to 100% in 15min, (A) the column effluent was monitored at 220nm. Violacein was eluted at 9.53min. Peak around 4min most probable is a system peak, not from the sample itself. (B) the column effluent was monitored at 575nm. More data are shown in Appendix D.

A mixture of TMF and methanol was adapted as solvent to dissolve violacein and HPLC was run with MeOH/water gradient or acetone/water gradient. Acetone/water gradient eluted violacein earlier than methanol/water gradient because acetone is more

polar than methanol. More detailed data including LCMS data is shown in Appendix D. According to the literature (*157*), crystallized violacein was analyzed by analytical HPLC. Crystallized violacein contained more deoxyviolacein compared to the fraction after preparative HPLC. Therefore preparative HPLC is required to isolate pure violacein.

The samples obtained from preparative HPLC were combined and lyophilized to dryness after removal of the solvent by rotary evaporation. Dried violacein was dissolved in deuterated dimethylsulphoxide (DMSO-$d_6$). DMSO is not easy to remove for the recovery of the sample. Another concern is the hygroscopic property of DMSO. A very intense water peak was observed at 3.3 ppm. We attempted to use $CDCl_3$, but solubility of violacein was poor. $^{1}H$ and $^{13}C$ NMR spectra were measured at 300 and 500 MHz. NMR spectra were acquired on a Bruker ARX 500 instrument using a 3 mm Broad band (BB)/H probe with WALTZ decoupling for $^{13}C$ experiment. To get high intensity data, 3mm NMR probe was used as well as 5mm NMR probe. Comparing our $^{1}H$ NMR and $^{13}C$ NMR spectra with the data in the literature (*162, 171, 172*), we obtained highly purified violacein. $^{1}H$ NMR , $^{13}C$ NMR, and $^{13}C$ DEPT135 NMR spectra were shown in Figure 46. Tabulated data of $^{1}H$ NMR and $^{13}C$ NMR spectra is shown in Table 5. In $^{13}C$ DEPT (Distortionless Enhancement by Polarization Transfer), $CH_2$ yields negative peaks, whereas CH and $CH_3$ are positive. Two-dimensional $^{1}H$-$^{1}H$ Correlation Spectroscopy (2D $^{1}H$-$^{1}H$ COSY) NMR spectrum and more NMR spectra are shown in Appendix D.

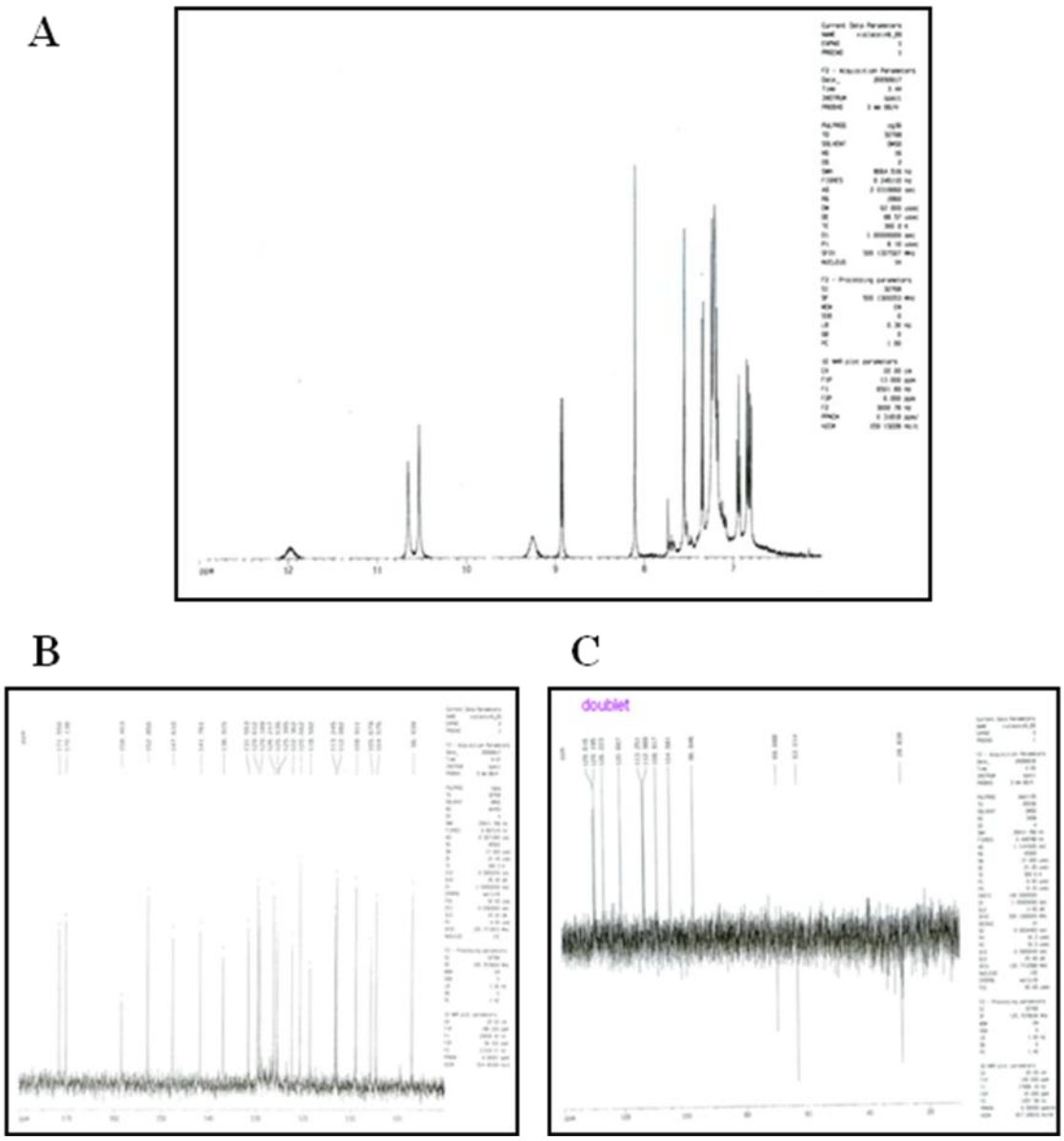
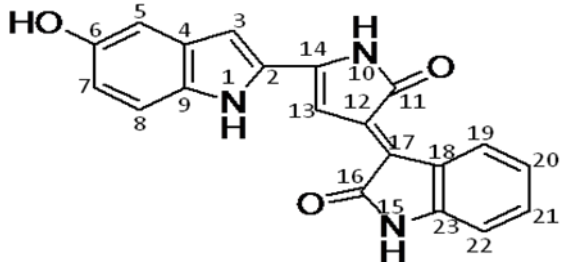



Figure 46: NMR spectra of purified violacein using preparative HPLC. NMR spectra were acquired on a Bruker ARX 500 instrument using a 3mm Broad band (BB)/H probe with WALTZ decoupling for $^{13}C$ experiment.(A) NMR-$^{1}H$ spectrum of violacein dissolved in DMSO-d6. (B) NMR-$^{13}C$ spectrum (C) $^{13}C$ DEPT135 NMR spectrum: doublet (CH, positive peak) and triplet ($CH_2$, negative peak). More NMR spectra and 2D $^{1}H$-$^{1}H$ COSY spectrum are shown in Appendix D.

Table 5: Identification of violacein using $^{1}H$ NMR and $^{13}C$ NMR spectroscopy (500MHz, DMSO-$d_6$)

| Position | $^{1}H$ NMR | | | $^{13}C$ NMR | | |
|---|---|---|---|---|---|---|
| | Sample | violacein a | violacein b | Sample | violacein a | violacein b |
| 1 | 11.95 | 11.95 | 11.89 | | | |
| 2 | 8.09 | 8.08 | 8.07 | 129.6 | 129.5 | 129.5 |
| 3 | | | | 105.7 | | 105.8 |
| 4 | | | | 125.5 | | 125.5 |
| 5 | 7.25 | 7.25 | 7.24 | 104.6 | 104.5 | 104.5 |
| 6 | 9.35 | | 9.35 | 152.9 | | 152.9 |
| 7 | 6.79 | 6.79 | 6.78 | 113 | 113.2 | 113.2 |
| 8 | 7.35 | 7.35 | 7.35 | 113.3 | 113.6 | 113.4 |
| 9 | | | | 131.6 | | 131.5 |
| 10 | 10.56 | 10.65 | 10.74 | | | |
| 11 | | | | 171.6 | | 171.5 |
| 12 | | | | 136.9 | | 136.9 |
| 13 | 7.55 | 7.56 | 7.55 | 96.9 | 97.2 | 96.9 |
| 14 | | | | 147.6 | | 147.5 |
| 15 | 10.65 | 10.75 | 10.64 | | | |
| 16 | | | | 170.1 | | 170.1 |
| 17 | | | | 118.6 | | 118.7 |
| 18 | | | | 122.4 | | 122.3 |
| 19 | 8.93 | 8.93 | 8.93 | 126.2 | 126.3 | 126.3 |
| 20 | 6.95 | 6.95 | 6.95 | 120.7 | 121 | 120.8 |
| 21 | 7.2 | 7.2 | 7.2 | 129.2 | 129.1 | 129.3 |
| 22 | 6.82 | 6.82 | 6.82 | 108.9 | 109 | 108.9 |
| 23 | | | | 141.8 | | 141.8 |




ESI MS m/z: 344.1033
MW: 343.0957 (violacein)

NMR data of violacein a in DMSO-d6 were cited from Nakamura, Y. *et al.* (*172*) and violacein b from Hoshino, T. *et al*.(*171*)

*Evaluation of Violacein Production Levels under the Various Promoters.* As mentioned above only six constructions in the Topo system expressed the violet-colored colonies: Lac/Native, Lac/Riboswitch, Lac/T7, Lac/ThiC/Riboswitch, ThiC/Riboswitch and Lac promoter. The expressed violet- colored colonies on Agar plates are shown in Figure 47.

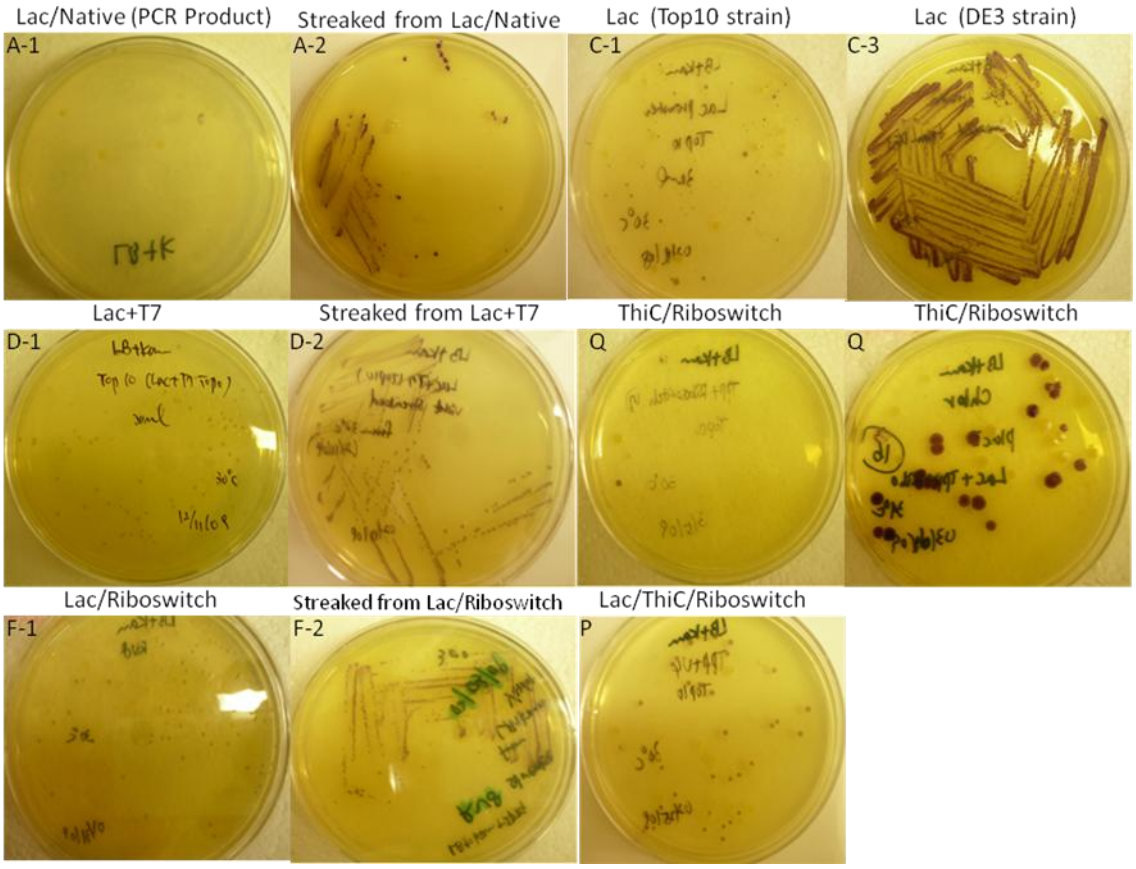



Figure 47: Constructions for the expression of violacein. 1 is original clone in Top10, 2 is streaked from 1, 3 is BL21 (DE3) strain. Labeled constructions as A, C, D, F, P, and Q are matched with Table 1. (A) Lac/Native promoter construction with PCR product, (C) Lac promoter (D) Lac/T7 promoter (F) Lac/Riboswitch (P) Lac/ThiC/Riboswitch (Q) ThiC/Riboswitch

The construction of thiC/Riboswitch promoter was shown to successfully express violacein. According to the results, we chose interesting constructions (such as Lac/Native, Native, Lac, T7, ThiC/Riboswitch and Lac/Riboswitch promoter) to evaluate violacein production. Native promoter and T7 promoter did not express violacein by itself in the *E. coli* system on agar plates. The constructions containing Native and T7 promoter were chosen because they cannot initiate the transcription of *vio* operon or production level is too low to detect by unaided eye. All chosen constructions were grown in LB including 10μg/ml kanamycin at 30°C for 5 days. The longer incubation time is due to original Topo XL vector being a high copy number plasmid (as well as pETBlue-2) and violacein having antibiotic properties (*149*). Violacein from *C. violaceum* was purified with preparative HPLC and then used to make a standard curve for quantification of violacein production in *E. coli* by various promoters. Standard curve is shown in Appendix D.

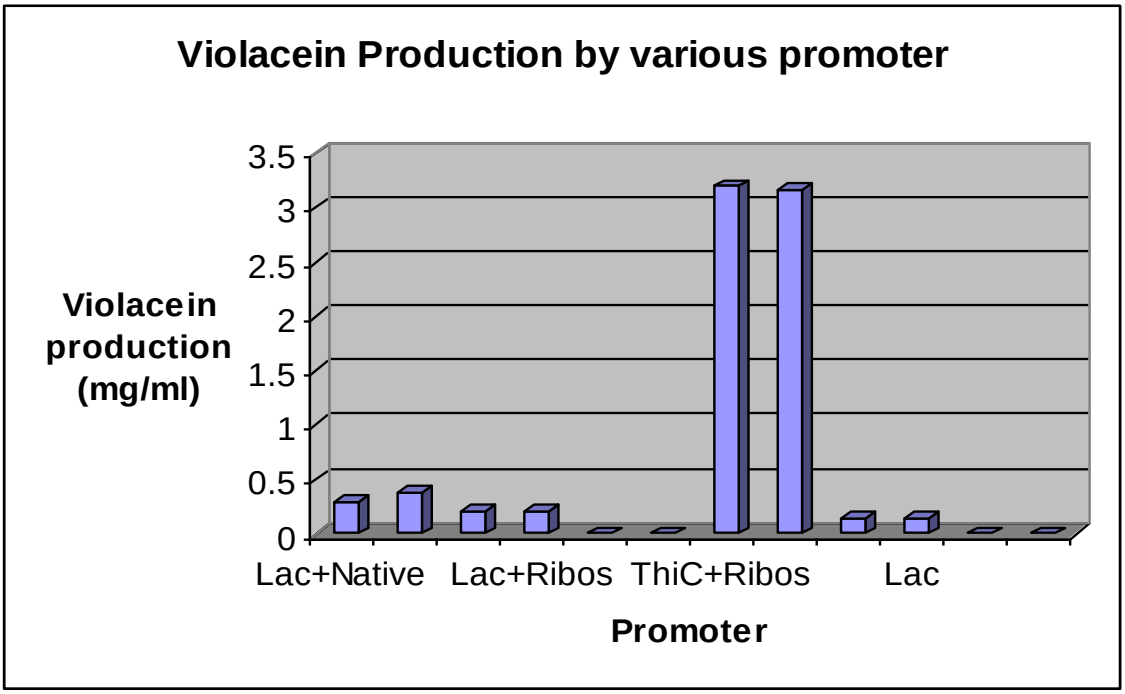


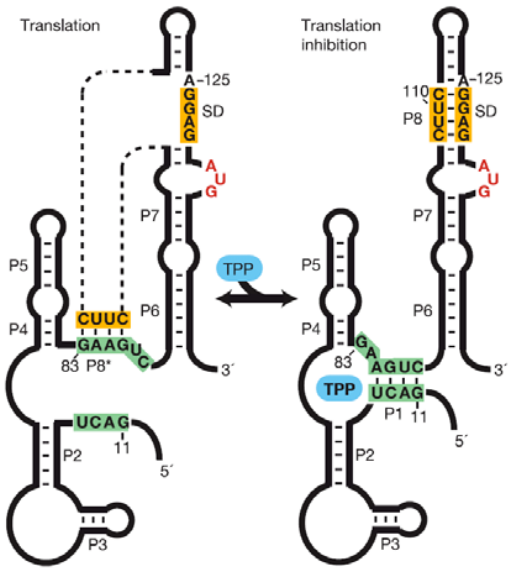

Figure 48: Violacein production by Lac/Native, Lac/Riboswitch, Native, ThiC/Riboswitch, Lac and T7 promoters. ThiC/Riboswitch promoter produced almost 3 times more violacein than other promoters.

As shown in Figure 48, the production of violacein by various promoters increased as follows Native ≈ T7 < Lac < Lac/Riboswitch < Lac/Native < ThiC/Riboswitch. ThiC and Riboswitch had a synergistic effect on the production of violacein. Lac/Native promoter also showed a synergistic effect compared to their individual effects on the production of violacein. Violacein production by Native promoter and T7 promoter was still too low to detect by HPLC. According to the standard curve, the minimum amount should be more than 0.03906mg/ml; the two promoter system might produce less than 0.03906mg/ml of violacein. Tabulated data is shown in Appendix D. This result suggests that the use of ThiC/Riboswitch promoter can be utilized as a general strategy

for improving the expression efficiency of a recombinant protein. Therefore, the system was evaluated to express β-galactosidase from LacZ, a predicted flavin reductase (RutF) in a novel pyrimidine catabolism pathway, and β-lacto-Globulin (BLG) protein. β-galactosidase was successfully expressed in a LacZ/ThiC/Riboswitch system with high efficacy. RutF and BLG were expressed in ThiC/Riboswitch system, but high expression level was not obtained in the system. The details of the data will be disclosed in subsequent sections.

*Evaluate Metabolite Profiles under the ThiC/Riboswitch Promoter and Lac/Riboswitch Promoter.* As mentioned above, translation of the *vio* operon under the ThiC/Riboswitch promoter control will be inhibited in the presence of high concentrations of TPP.

Scheme 6 is a schematic representation of the proposed mechanism for TPP-dependent deactivation of thiM translation. The conformation of TPP riboswitch will rearrange after TPP is bound. Rearranged TPP riboswitch will not initiate the translation.




Scheme 6: Schematic representation of the proposed mechanism for TPP-dependent deactivation of thiM translation. After Serganov, A. *et al*. (*145*)

The production of violacein under the ThiC/Riboswitch promoter and Lac/Riboswitch promoter were evaluated by analytical HPLC. Inhibition of violacein expression in *E. coli* under the ThiC/Riboswitch and the Lac/Riboswitch promoter was monitored in the presence of TPP. TPP concentrations were 0mM, 0.5 mM, 1.0mM, 1.5 mM, 2.0 mM, 2.5 mM, 3 mM and 5 mM as shown in Figure 49. As a result of the construction containing ThiC/Riboswitch promoter the expression of violacein can be tuned by the concentration of TPP as expected.

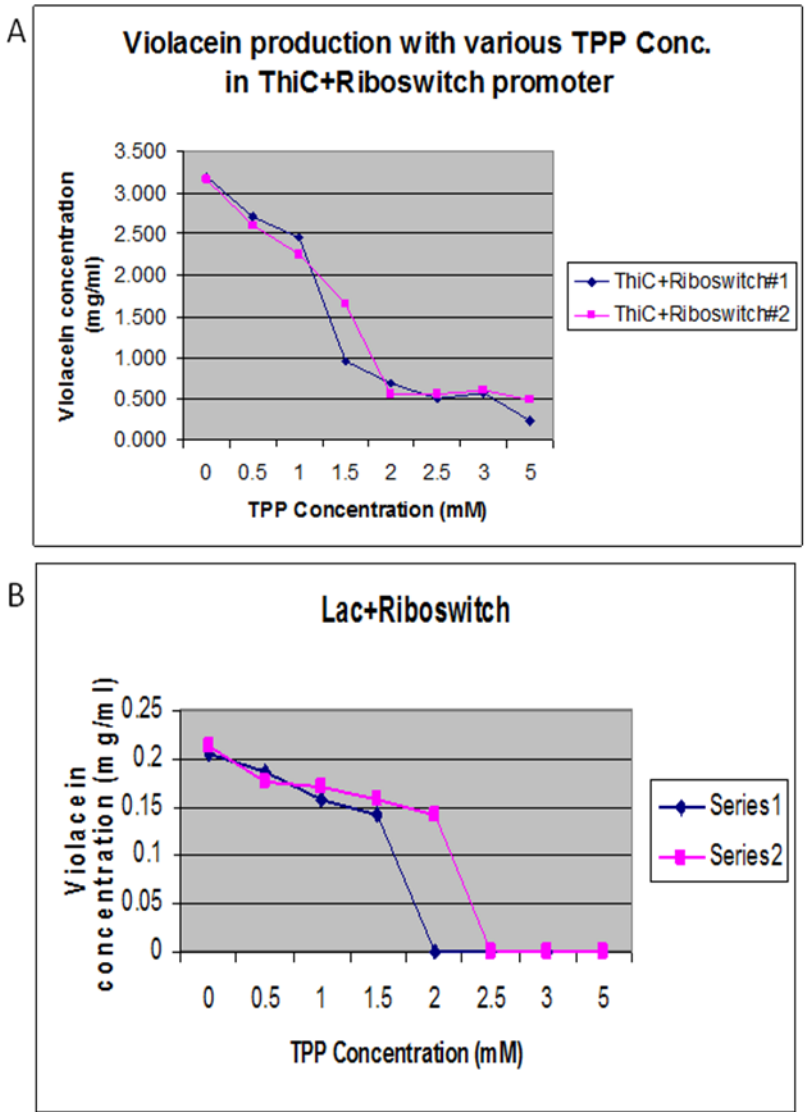



Figure 49: Inhibition of violacein expression in *E. coli* due to the presence of thiamine pyrophosphate chloride (TPP salt). A: under ThiC/Riboswitch promoter control B: under Lac/Riboswitch promoter control.

The effect of TPP on transcription was also monitored by HPLC analysis with the construction under Lac/Riboswitch. Violacein production was not detected in the presence of 2.25 mM TPP concentration. Therefore, the transcription was terminated or

inhibited by TPP concentration. Based on these results, the function of ThiC/Riboswitch and riboswitch can be controlled by TPP concentration to express violacein. The TPP riboswitch is believed to modulate gene expression primarily at the level of transcription where affinity for TPP is thought to vary as the RNA transcript is released from the RNA polymerase. Some modulation does occur at the translational level, where affinity for TPP varies as the protein emerges from the ribosome. The ability of the TPP riboswitch to accept heterologous templates was exemplified by its ability to modulate β-galactosidase expression.

*Investigation of Titratability of the Thiamin Riboswitch.* ThiC/Riboswitch/β-galactosidase system and Arabinose/β-galactosidase system were employed to investigate titrability of the thiamin riboswitch compared to the pBAD promoter. The LacZ gene used in both systems was obtained from PCR and insertion was confirmed by sequencing. In the case of the ThiC/Riboswitch/ β-galactosidase system, ONPG/ β-galactosidase assay was modified. After removing LB media containing various TPP concentrations, we again added TPP to permeabilized cells in Z-buffer. As a control experiment, we checked O.D at 450 nm with ONPG-treated permeabilized cells in Z-buffer without TPP. When TPP was not included in the Z-buffer we obtained results which were not consistent. Therefore, ThiC/Riboswitch is very sensitive to the concentration of TPP. The results showed that the expression of β-galactosidase with the ThiC/Riboswitch promoter was inhibited by TPP as shown in Figure 50 (A). This result corresponds to the results of the expression of violacein controlled by TPP. Violacein production and β-galactosidase production were controlled by the TPP concentration.

Therefore, it demonstrates that ThiC/Riboswitch promoter is titratable using TPP. As many gene products are toxic, the use of the ThiC/Riboswitch is able to modulate gene expression pathways by varying the TPP concentration.

The pBAD33 promoter was also monitored to check titratability using arabinose and glucose. Arabinose is the inducer and glucose is the repressor in the pBAD system. In the presence of 0.2% arabinose or 0.2% of glucose, gene expression can be fully induced or repressed as described in previously reported literature (*141, 173*). We investigated arabinose's effect on the expression of β-galactosidase with arabinose concentrations less than 0.2%. Results showed that β-galactosidase units increased proportional to the concentration of arabinose. The result is shown in Figure 50 (B). We did not check with high concentration of arabinose because 0.2% arabinose fully induces gene expression under the pBAD promoter (*141, 173*), and the pBAD33 promoter was monitored as a tunable promoter with lower concentration of arabinose. However, the data obtained with glucose titration showed the pBAD33 promoter is not a titratable promoter as shown in Figure 50 (C). The result corresponds to the previously reported data in which 0.2% glucose can fully repress the gene expression even though it is not titratable.

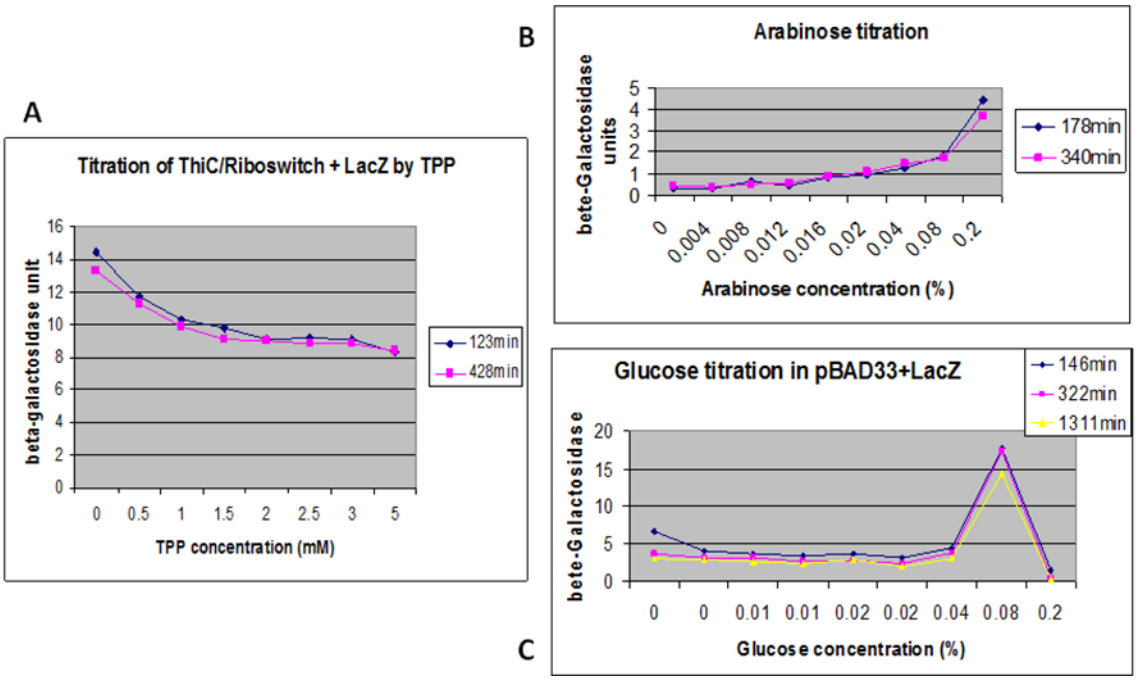



Figure 50: Titratability check with ThiC/Riboswitch promoter by TPP (A) pBAD33 promoter by arabinose (B) and glucose (C).

*Comparison of Expression Profiles of RutF Protein under the T5-Lac Promoter vs TPP/Riboswitch Promoter.* RutF protein is a predicted flavin reductase in a novel pyrimidine catabolism pathway. It is an insoluble protein under standard conditions. RutF DNA was obtained from Prof. Tadhg Begley's group, Department of Chemistry, Texas A&M University. The PCR product of RutF was ligated to the vector containing the ThiC/Riboswitch promoter as described in the experimental section. Transformation into Top 10 *E. coli* strain followed. RutF under a T5-Lac promoter of pCA24N vector was used as a control. The soluble and insoluble samples were loaded onto 12% SDS-PAGE gel (Figure 51). An 18kDa protein corresponding to the size of the RutF was

expressed by the construction under the ThiC/Riboswitch promoter. Compared to the T5-Lac promoter, the level of RutF protein expression was low and solubility of the protein was not increased under ThiC/Riboswitch promoter. Interestingly, the solubility of RutF protein was improved by transformation into Top10 *E. coli* strain. This indicates that recombinant protein solubility depends on the host cell used.

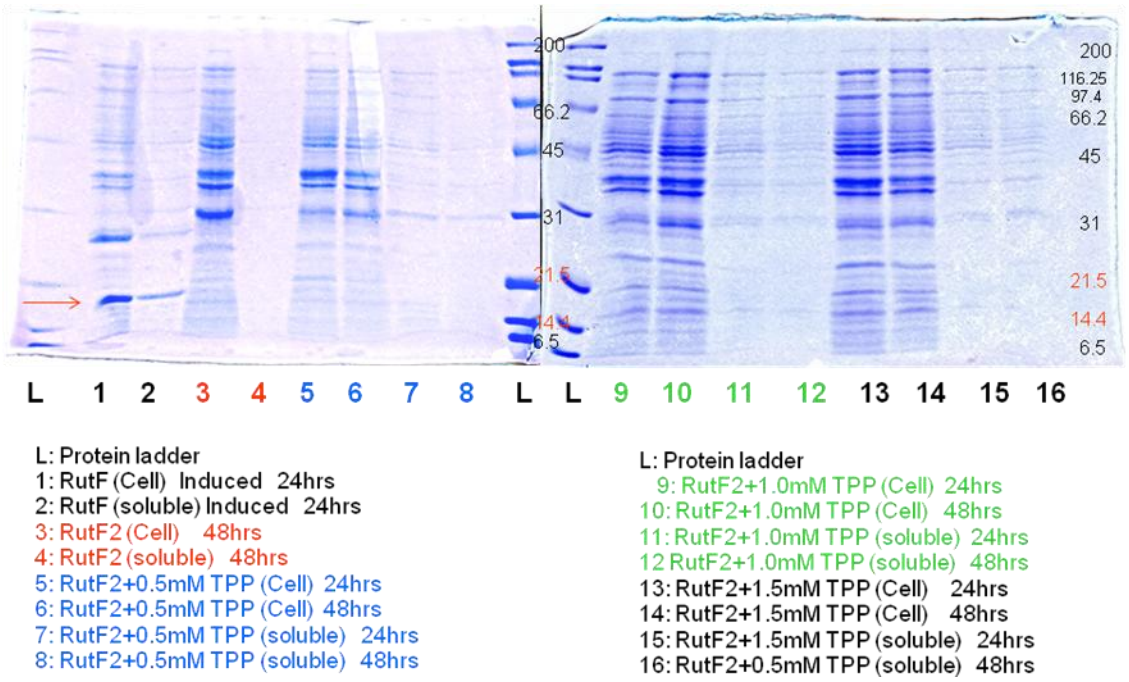



Figure 51: SDS-PAGE gel of RutF protein. Comparison of solubility of RutF under T5-Lac promoter *vs* RutF2 under TPP/Riboswitch promoter.

*Evaluate β-lactoglubin (BLG) Protein Expression by SDS/PAGE under TPP/Riboswitch Promoter*. BLG protein was chosen as a complementary experiment for another project in the Watanabe group. The BLG gene was inserted into the vector under control of the TPP/Riboswitch promoter as described in experimental procedure. Protein expression was analyzed on SDS/PAGE gel as shown in Figure 51. Phusion DNA

polymerase and Taq DNA polymerase were used to obtain the PCR product. Both polymerases produced the desired PCR product which contained the BLG gene. PH 1 protein was from the PCR product using phusion DNA polymerase. The protein was expressed in Top 10 *E. coli* host strain as shown in Figure 52.

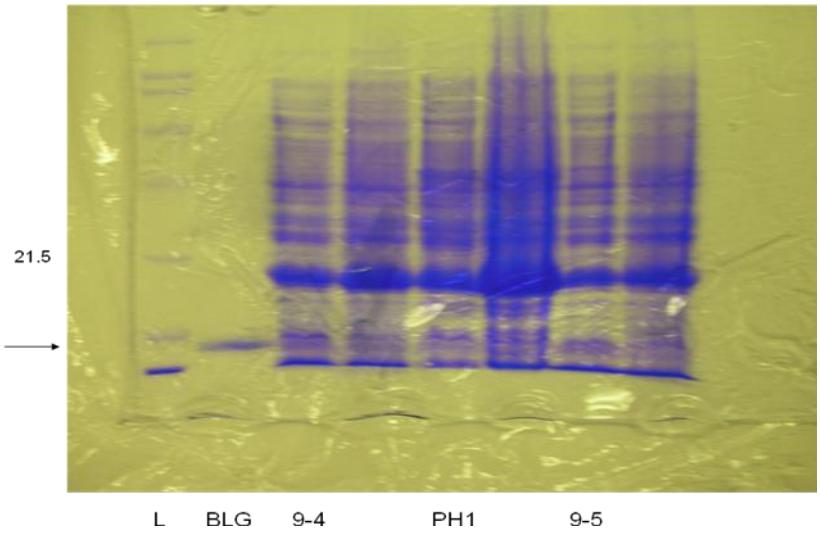



Figure 52: SDS –PAGE gel of BLG protein. PH1 from the PCR product using phusion DNA polymerase, 9-4 and 9-5 from the PCR product using Taq DNA polymerase.

In conclusion, the production of violacein was improved by employing the ThiC/Riboswitch as shown earlier. Violacein is commercially available from the Sigma/Aldrich Company, but the price is expensive at $249/mg, and purity is only 85%, with the remainder as deoxyviolacein. Our strategy can improve the production of violacein in *E. coli* and an optimized HPLC condition can be applied to purify violacein from deoxyviolacein.

We have successfully used the thiC riboswitch which is regulated by thiamin pyrophosphate concentration, to modulate gene expression of the natural product violacein. Additionally, the use of the ThiC/Riboswitch promoter as a titratable promoter was investigated with the LacZ system and compared to the pBAD promoter. Based on the data, we can conclude that the ThiC/Riboswitch promoter is tunable using varied thiamin pyrophosphate concentration, while the pBAD promoter is not tunable.

The ThiC/Riboswitch promoter was also employed in an attempt to improve the solubility of proteins, however we did not observe improvement of in solubility. Solubility of recombinant protein is highly dependent on the specific protein sequences, some proteins would increase solubility but not in proteins used in this study.

Therefore, the use of the ThiC/Riboswitch promoter is a general strategy for improving the expression efficiency of a recombinant protein or a metabolite, and for modulation of pathway gene expression by varying the TPP concentration.

# CHAPTER VI

# CONCLUSIONS

Current purification methods or cultivation conditions provide a limited source of marine microbial samples. We attemped to construct DNA libraries to clone and express non-culturable bacterial genomes using recombinant DNA technology instead of developing time consuming purification methods or manipulating cultivation conditions. Two different cDNA libraries from different microorganisms (Pup14A/B, Hon6) and five different fosmid libraries from environmental samples were successfully constructed. From cDNA libraries, we did not complete the screening of the libraries. Therefore, no useful clones of biosynthetic pathways have been detected so far. Fosmid libraries from marine environmental samples were created with 30 percent success (five libraries were constructed out of sixteen samples). Some samples originated from the deep depth of the ocean, so it might contain different organisms which may use different biosynthetic pathways compared to teresterial samples. However, the samples from marine environments contained too little DNA to construct libraries. DNA isolatation was the main obstacle in increasing the chance of successful construction of the libraries. In most cases, we failed to construct the library because of we could not isolate enough DNA.

For screening of the fosmid libraries, genetic selection is a practical and economic method. To evaluate the potential use of genetic selection in the elucidation of biosynthetic pathways from marine microbial consortia, we identified the biotin

biosynthetic genes within Hon6. As the statistical probability of observing heterologous expression of a single gene is much higher than that of an entire biosynthetic pathway, genetic selection could prove to be an important tool in the localization of novel biosynthetic pathways. Coupled with promoter replacement technologies, the approach could lead to the discovery of a new generation of bioactive natural products. In this regard, clones from the constructed libraries can be screened by genetic selection to reveal novel biosynthetic pathways, such as thiamine synthase (ThiC), pyrimidine synthase (NadA), heme (HemN), lipoyltransferase (LipB), and thiamine biosynthesis (ThiH), in marine microbial consortia.

The utility of the ThiC riboswitch was evaluated to control violacein production using promoter replacement technology. The production of violacein was increased by the ThiC riboswitch promoter compared to other promoters. The "switch" function of the riboswitch was evaluated by changing the thiamin pyrophosphate concentration in *E. coli*. We have successfully used the ThiC riboswitch to regulate gene expression of the natural product violacein. Since this work has been very successful, we are now interested in pursuing this work further to develop it as a tool to carry out genetic knockouts of biosynthetic pathways like the violacein gene cluster, or to analyze the metabolic flux of pathways. The violacein gene cluster is composed of five genes (vioA, vioB, vioC, vioD, and vioE) as shown in scheme 4 in Chapter V. A different riboswitch can also be substituted to control the gene cluster. A variety of riboswitches are made available by nature such as lysine, flavin, adenine, etc. (shown in Figure 53) which could be further exploited for orthogonal production.

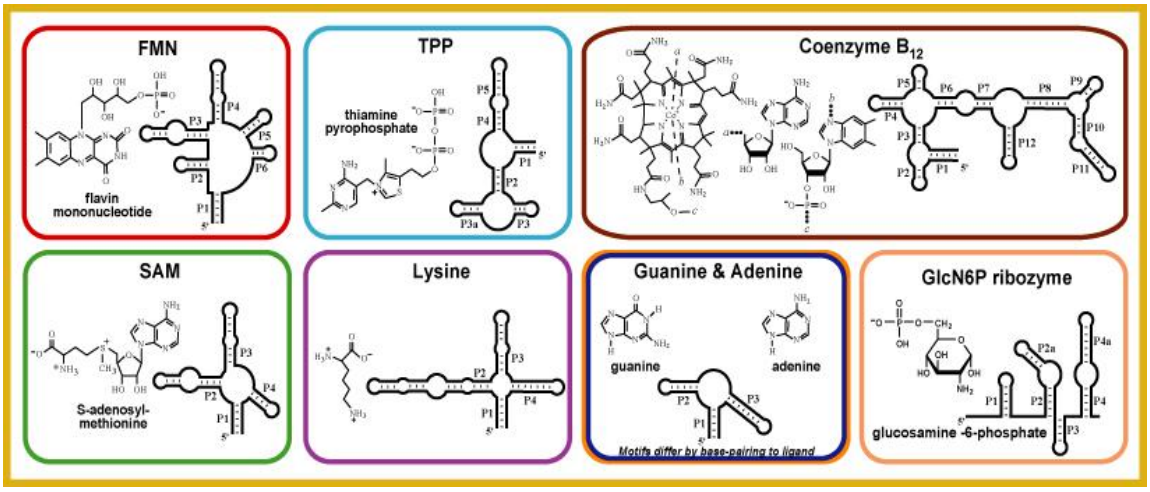



Figure 53: The chemical structures of riboswitch ligands and schematics of conserved secondary structure in riboswitches. There are 8 confirmed riboswitches with unique metabolite ligands. Many more conserved RNA motifs are currently under investigation. After Courtesy: Ron Breaker Lab (Yale University) (*174*)

In conclusion, our research goals were achieved. First we successfully created cDNA libraries from Pup14A/B and Hon6 and fosmid libraries from environmental samples for the identification of the natural product biosynthetic genes. The construction of DNA libraries could be an alternative method to identify genes from unculturable microorganisms. Second, genetic selection of the biosynthetic pathways was applied to identify biotin biosynthetic genes in Hon6. Third, genetic manipulations of the production of a natural product were successfully investigated with violacein controlled by various promoters. DNA libraries can be screened by genetic selection coupled with promoter replacement technologies. ThiC riboswitch was a powerful promoter in the heterologous production of violacein in *E. coli*. The method could be applied to increase the production of other natural products in *E. coli*. Fourth, modulation of gene expression by the ThiC riboswitch promoter was achieved using TPP concentrations. As

many gene products are toxic, the use of the ThiC Riboswitch is able to modulate gene expression pathways by varing the TPP concentration.

## REFERENCES


1. Fesher, M., and Schmidt, J. M. (2003) Property distributions: differences between drugs, natural products, and molecules from combinational chemistry, *J. Chem. Inf. Comput. Sci. 43*, 218-227.
2. Bull, A. D., Ward, A. C., and Goodfellow, M. (2000) Search and discovery strategies for biotechnology: the paradigm shift, *Microb. Mol. Biol. Rev 64*, 573-606.
3. Rouhi, A. M. (2003) Rediscovering natural products, *Chem. & Eng. News 81*, 77-91.
4. Newman, D. J., Cragg, G. M., and K.M., S. (2003) Natural products as sources of new drugs over the period 1981-2002, *J. Nat.Prod. 66*, 1022-1037.
5. Newman, D. J. (2008) Natural products as leads to potential drugs: an old process or the new hope for drug discovery? *J. Med. Chem. 51*, 2589-2599.
6. Kingston, D. G. I. (2001) Taxol, a molecule for all seasons, *Chem. Comm.*, 867-880.
7. Abad, F., Calbo, F., Zapater, P., Rodrı´guez-Vilanova, F., Garcı´a-Pe´rez, L.-E., and Sacrista´n, J. A. (2000) Comparative pharmacoeconomic study of vancomycin and teicoplanin in intensive care patients, *Int. J. Antimicrob. Agents 15*, 65-71.
8. Ji, H.-F., Li, X.-J., and Zhang, H.-Y. (2009) Natural products and drug discovery, *European Molecular Biology Organization Reports 10*, 194-200.

9. Singh, S. B., and Barrett, J. F. (2006) Empirical antibacterial drug discovery-Foundation in natural products, *Biochemical Pharmacology 71*, 1006-1015.

10. BAYER. (2009) "1899" Wonders of Aspirin, <http://www.wonderdrug.com/pain/asp_history.htm.>

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

12. Morimoto, S., Suemori, K., Moriwaki, J., Taura, F., Tanaka, H., Aso, M., Tanaka, M., Suemune, H., Shimohigashi, Y., and Shoyama, Y. (2001) Morphine metabolism in the opium poppy and its possible physiological function. biochemical characterization of the morphine metabolite, bismorphine, *J. Biol. Chem. 276*, 38179-38184.

13. Woodward, R. B., and Doering, W. E. (1944) The total synthesis of quinine, *J. Am. Chem. Soc. 66*, 849-849.

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

15. Tally, F. P., and DeBruin, M. F. (2000) Development of daptomycin for gram-positive infections, *J. Antimicrob. Chemother. 46*, 523-526.

16. Fowler, V. G., Boucher, H. W., Corey, G. R., Abrutyn, E., Karchmer, A. W., Rupp, M. E., Levine, D. P., Chambers, H. F., Tally, F. P., Vigliani, G. A., Cabell, C. H., Link, A. S., DeMeyer, I., Filler, S. G., Zervos, M., Cook, P., Parsonnet, J., Bernstein, J. M., Price, C. S., Forrest, G. N., Fätkenheuer, G., Gareca, M., Rehm, S. J., Brodt, H. R., Tice, A., and Cosgrove, S. E. (2006) Daptomycin versus

standard therapy for bacteremia and endocarditis caused by *Staphylococcus aureus*, *N. Engl. J. Med. 355*, 653-665.

17. Maheshwai, N. (2007) Are young infants treated with erythromycin at risk for developing hypertrophic pyloric stenosis? *Arch. Dis. Child. 92,* 271-273.
18. Feder, H. (1986) Chloramphenicol: what we have learned in the last decade, *South. Med. J. 79*, 1129-1134.
19. Wallerstein, R., Condit, P., Kasper, C., Brown, J., and Morrison, F. (1969) Statewide study of chloramphenicol therapy and fatal aplastic anemia, *JAMA 208*, 2045-2050.
20. Levine, D. P. (2006) Vancomycin: a history, *Clin Infect Dis 42*, (Suppl 1): S5-12.
21. Drews, J. (2000) Drug discovery: a historical perspective, *Science 287*, 1960-1964.
22. Antunez, C., Blanca-Lopez, N., Torres, M. J., Mayorga, C., Perez-Inestrosa, E., Montañez, M. I., Fernandez, T., and Blanca, M. (2006) Immediate allergic reactions to cephalosporins: evaluation of cross-reactivity with a panel of penicillins and cephalosporins, *J. Allergy Clin. Immunol. 117*, 404-410.
23. Pegle, R. S., and Healy, B. (2007) In patients allergic to penicillin, consider second and third generation cephalosporins for life threatening infections, *BMJ 335*, 991.
24. Birnbaum, J., Kahan, F. M., Kropp, H., and MacDonald, J. S. (1985) Carbapenems, a new class of beta-lactam antibiotics. Discovery and development of imipenem/cilastatin, *Am. J. Med. 78*, 3-21.

25. Kumagai, T., Tamai, S., Abe, T., and Hikda, M. (2002) Current status of oral carbapenem development, *Current Medicinal Chemistry -Anti-Infective Agents 1*, 1-14.

26. Farver, D. K., Hedge, D. D., and Lee, S. C. (2005) Ramoplanin: a lipoglycodepsipeptide antibiotic, *Annals of Pharmacotherapy 5*, 863-868.

27. Zakeri, B., and Wright, G. D. (2008) Chemical biology of tetracycline antibiotics, *Biochem. Cell Biol. 86*, 124-136.

28. Min, Z., Burman, W. J., Jaresko, G. S., Berning, S. E., Jelliffe, R. W., and Pwloquin, C. A. (2001) Streptomycin in patients with tuberculosis, *Pharmacotherapy 21*, 1037-1045.

29. Singh, B., and Mitchison, D. A. (1954) Bactericidal activity of streptomycin and isoniazid against tubercle bacilli, *British Medical Journal 1*, 130-132

30. Bradley, J. S., Wassel, R. T., Lee, L., and Nambiar, S. (2009) Intravenous ceftriaxone and calcium in the neonate: assessing the risk for cardiopulmonary adverse events, *Pediatrics 123*, e609-613.

31. Hahn, D. L. (1995) Treatment of chlamydia pneumoniae infection in adult asthma: a before-after trial, *J. Fam. Pract. 41*, 345-351.

32. Mutter, R., and Wills, M. (2000) Chemistry and clinical biology or the bryostatins, *Bioorganic Med. Chem. 8*, 1841-1860.

33. Reviglione, M. C., Gupta, R., Dye, C. M., and Espinal, M. A. (2001) The burden of drug-resistant tuberculosis and mechanisms for its control, *Ann. N. Y. Acad. Sci 953*, 88-97.

34. Amann, R. I., Ludwig, W., and Schleifer, K. H. (1995) Phylogenetic identification and *in-situ* detection of individual microbial-cells without cltivation, *Microbiol. Rev. 59*, 143-169.

35. Borman, S. (2002) Combinatorial chemistry: advances in synthesis, purification, and analysis further refine the combinatorial approach, now a mainstream tool in drug discovery, *C&EN Washington 45*, 43-57.

36. Kumagai, N., Muncipinto, G., and Schreiber, S. L. (2006) Short synthesis of skeletally and stereochemically diverse small molecules by coupling petasis condensation reactions to cyclization reactions, *Angew. Chem., Int. Ed. 45*, 3635 - 3638.

37. Dewick, P. M. (2002) *Medicinal natural products: a biosynthetic approach*, 2nd ed., Wiley, Chichester, West Sussex, England ; New York.

38. Koehn, F. E., and Carter, G. T. (2005) The evolving role of natural products in drug discovery, *Nat Rev Drug Discov. 4*, 206-220.

39. Wawrik, B., Zylstra, G., Kerkhof, L., and Kukor, J. (2005) Identification of unique type II polyketide synthase genes in soil, *Appl. Environ. Microbiol. 71*, 2232-2238.

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

41. Staunton, J., and Weissman, K. J. (2001) Polyketide biosynthesis: a millennium review, *Natural Product Reports 18*, 380-416.

42. Marahiel, M. A., Stachelhaus, T., and Mootz, H. D. (1997) Modular peptide synthetases involved in nonribosomal peptide synthesis, *Chem. Rev. 97*, 2651-2673.

43. Schwarzer, D., Finking, R., and Marahiel, M. A. (2003) Nonribosomal peptides: from genes to products, *Natural Product Reports 20*, 175-187.

44. Davis, E. M., and Croteau, R. (2000) Cyclization enzymes in the biosynthesis of monoterpenes, sesquiterpenes, and diterpenes, *Biosynthesis: Aromatic Polyketides, Isoprenoids, Alkaloids 209*, 53-95.

45. Humphrey, A. J., and O'Hagan, D. (2001) Tropane alkaloid biosynthesis. a century old problem unresolved, *Nat. Prod. Rep. 18*, 494-502.

46. MacNeil, I. A., Tiong, C. L., Minor, C., August, P. R., Grossman, T. H., Loiacono, K. A., Lynch, B. A., Phillip, T., Narula, S., Sundaramoorthi, R., Tyler, A., Aldredge, T., Long, H., Gilman, M., Holt, D., and Osburne, M. S. (2001) Expression and isolation of antimicrobial small molecules from soil DNA libraries, *J. Mol. Microbiol. Biotechnol. 3*, 301-308.

47. Martín, H. G., Ivanova, N., Kunin, V., Warnecke, F., Barry, K. W., McHardy, A. C., Yeates, C., He, S., Salamov, A. A., Szeto, E., Dalin, E., Putnam, N. H., Shapiro, H. J., Pangilinan, J. L., Rigoutsos, I., Kyrpides, N. C., Blackall, L. L., McMahon, K. D., and Hugenholtz, P. (2006) Metagenomic analysis of two enhanced biological phosphorus removal (EBPR) sludge communities, *Nat. Biotechnol 24*, 1263-1269.

48. Jones, B. V., and Marchesi, J. R. (2007) Transposon-aided capture (TRACA) of plasmids resident in the human gut mobile metagenome, *Nat. Methods 4*, 55-61.

49. Lee, D.-G., Jeon, J. H., Jang, M. K., Kim, N. Y., Lee, J. H., Kim, S.-J., Kim, G.-D., and Lee, S.-H. (2007) Screening and characterization of a novel fibrinolytic metablloprotease from a metagenomic library, *Biotechnol. Lett. 29*, 465-472.

50. Akira, O., Ryo, M., Masahiro, S., Yoshiyuki, O., Yuji, N., and Masataka, T. (2007) Isolation and characterization of naphthalene-catabolic genes and plasmids from oil-contaminated soil by using two cultivation-independent approaches, *Appl. Environ. Microbiol. 74*, 501-510.

51. Turnbaugh, P. J., Ley, R. E., Mahowald, M. A., Magrini, V., Mardis, E. R., and Gordon, J. I. (2006) An obesity-associated gut microbiome with increased capacity for energy harvest, *Nature 444*, 1027-1031.

52. Ausubel, F. M., Brent, R., Kingston, R. E., Moore, D. D., Seidman, J. G., Smithe, J. A., and Struhl, K. (2002) *Current Protocols in Molcecular Biology 1*, John Wiley & Sons, Inc. New York, NY

53. Rondon, M. R., August, P. R., Bettermann, A. D., Brady, S. F., Grossman, T. H., Liles, M. R., Loiacono, K. A., Lynch, B. A., MacNeil, I. A., Minor, C., Tiong, C. L., Gilman, M., Osburne, M. S., Clardy, J., Handelsman, J., and Goodman, R. M. (2000) Cloning the soil metagenome: a strategy for accessing the genetic and functional diversity of uncultured microorganisms, *Appl. Environ. Microbiol. 66*, 2541-2547.

54. Jetty, A., Luo, M., Yu, Y., Kudrna, D., Kim, H., and Wing, R. (2004) Construction of a random sheared fosmid library and its utility in refining the rice (*Oryza sativa* ssp. *japonica* cv. Nipponbare) physical map, *CA Plant & Animal Genomes XII Conference*, January 10-14, 2004 Town & Country

Convention Center, San Diego <http://www.intl-pag.org/12/abstracts/P2a_ PAG2 _142.html - 4k>.

55. Sternberg, N. (1990) Bacteriophage P1 cloning system for the isolation, amplification, and recovery of DNA fragments as large as 100 kilobase pairs, *Proc. Natl. Acad. Sci. U.S.A. 87*, 103-107.

56. Shizuya, H., and Kouros-Mehr, H. (2001) The development and applications of the bacterial artificial chromosome cloning system, *Keio J. Med. 50 (1)*, 26-30.

57. Brady, S. F., Chao, C. J., and Clardy, J. (2002) New natural product families from an environmental DNA (eDNA) gene cluster, *J. Am. Chem. Soc. 124*, 9968-9969.

58. Brady, S. F., Chao, C. J., Handelsman, J., and Clardy, J. (2001) Cloning and heterologous expression of a natural product biosynthetic gene cluster from eDNA, *Org. Lett. 3*, 1981-1984.

59. Brady, S. F., and Clardy, J. (2000) Long-chain N-Acyl amino acid antibiotics isolated from heterologously expressed environmental DNA, *J. Am. Chem. Soc. 122*, 12903-12904.

60. Brady, S. F., and Clardy, J. (2004) Palmitoylputrescine, an antibiotic isolated from the heterologous expression of DNA extracted from bromeliad tank water, *J. Nat. Prod. 6*, 1283-1286.

61. Gillepsie, D. E., Brady, S. F., Bettermann, A. D., Cianciotto, N. P., Liles, M. R., Clardy, J., Goodman, R. M., and Handelsman, J. (2002) Isolation of antibiotics turbomycin A and B from a metagenomic library of soil microbial DNA, *Appl. Environ. Microbiol. 68*, 4301-4306.

62. Corbino, K. A., Barrick, J. E., Lim, J., Welz, R., Tucker, B. J., Puskarz, I., Mandal, M., Rudnick, N. D., and Breaker, R. R. (2005) Evidence for a second class of S-adenosylmethionine riboswitches and other regulatory RNA motifs in alpha-proteobacteria, *Genome Biol. 6*, R70-R70.9

63. Mandal, M., and Braker, R. R. (2004) Gene regulation by riboswitches, *Nature Reviews 5*, 451-463.

64. Wachter, A., Tunc-Ozdemir, M., Grove, B. C., Green, P. J., Shintani, D. K., and Breaker, R. R. (2007) Riboswitch control of gene expression in plants by splicing and alternative 3' end processing of mRNAs, *Plant Cell 19*, 3437-3450.

65. Cochrane, J. C., and Strobel, S. A. (2008) Riboswitch effectors as protein enzyme cofactors, *RNA 14*, 993-1002.

66. Tucker, B. J., and Breaker, R. R. (2005) Riboswitches as versatile gene control elements, *Curr Opin Struct Biol 15*, 342-348.

67. Vitreschak, A. G., Rodionov, D. A., Mironov, A. A., and Gelfand, M. S. (2004) Riboswitches: the oldest mechanism for the regulation of gene expression? *Trends Genet. 20*, 44-50.

68. Batey, R. T. (2006) Structures of regulatory elements in mRNAs, *Curr Opin Struct Biol 16*, 299-306.

69. Mironov, A. S., Gusarov, I., Rafikov, R., Lopez, L. E., Shatalin, K., Kreneva, R. A., Perumov, D. A., and Nudler, E. (2002) Sensing small molecules by nascent RNA: a mechanism to control transcription in bacteria, *Cell 111*, 747-756.

70. Winkler, W. C., Cohen-Chalamish, S., and Breaker, R. R. (2002) An mRNA structure that controls gene expression by binding FMN, *Proc. Natl. Acad. Sci. U.S.A. 99*, 15908-15913.

71. Sudarsan, N., Barrick, J. E., and Breaker, R. R. (2003) Metabolite-binding RNA domains are present in the genes of eukaryotes, *RNA 9*, 644-647.

72. Barrick, J. E., Corbino, K. A., Winkler, W. C., Nahvi, A., Mandal, M., Collins, J., Lee, M., Roth, A., Sudarsan, N., Jona, I., Wickiser, J. K., and Breaker, R. R. (2004) New RNA motifs suggest an expanded scope for riboswitches in bacterial genetic control, *Proc. Natl. Acad. Sci. U.S A. 101*, 6421-6426.

73. Soukup, J. K., and Soukup, G. A. (2004) Riboswitches exert genetic control through metabolite-induced conformational change, *Curr Opin Struct Biol 14*, 344-349.

74. Montange, R., K,, and Batey, R. T. (2006) Structure of the S-adenosylmethionine riboswitch regulatory mRNA element, *Nature 441*, 1172-1175.

75. Heppell, B., and Lafontaine, D. A. (2008) Folding of the SAM aptamer is determined by the formation of a K-turn-dependant pseudoknot, *Biochemistry 47*, 1490-1499.

76. Thore, S., Leibundgut, M., and Ban, N. (2006) Structure of the eukaryotic thiamine pyrophosphate riboswitch with its regulatory ligand, *Science 312*, 1208 - 1211.

77. Osoegawa, K., Woon, P. Y., Zhao, B. H., Frengen, E., Tateno, M., Catanese, J. J., and de Jong, P. J. (1998) An improved approach for construction of bacterial artificial chromosome libraries, *Genomics 52*, 1-8.

78. Kibala, B. (1993) Naphthenic acid corrosion in a refinery setting, *NACE Corrosion Annual Conference and Expo, paper #631*.

79. Craig Jr, H. L. (1995) Naphthenic acid corrosion in the refinery, *NACE Corrosion Annual Conference and Expo, paper # 333*.

80. Derungs, W. A. (1965) Naphthenic acid corrosion - an old enemy of the petroleum industry, *Corrosion*, *12*, 41-46.

81. Lewis, K. R., Daane, M. L., and Shelling, R. (1999) Processing corrosive crude oils, *NACE Corrosion Annual Conference and Expo, paper #377*.

82. Angell, S., Bench, B. J., Williams, H., and Watanabe, C. M. (2006) Pyocyanin isolated from a marine microbial population: synergistic production between two distinct bacterial species and mode of action, *Chem. Biol. 13*, 1349-1359.

83. Cueto, M., Jensen, P. R., and Fenical, W. (2002) Aspergilloxide, a novel sesterterpene epoxide from a marine-derived fungus of the genus *Aspergillus*, *Org. Lett 4*, 1583-1585.

84. Watanabe, C. M. H., Supekova, L., and Schultz, P. G. (2002) Transcriptional effects to the potent enediyne anti-cancer agent clicheamicin $g_1$I, *Chem. Biol. 9*, 245-251. .

85. Invitrogen Life Technologies. (2004) *Superscript™ choice system for cDNA synthesis, Manual I*, Invitrogen, Inc. Carlsbad, CA.

86. Piel, J., Hui, D., Wen, G., Butzke, D., Platzer, M., Fusetani, N., and Matsunaga, S. (2004) Antitumor polyketide biosynthesis by an uncultivated bacterial symbiont of the marine sponge *Theonella swinhoei*, *Proc. Natl. Acad. Sci. U.S.A. 101*, 16222-16227.

87. Kimmel, A. R., and Berger, S. L. (1987) Preparation of cDNA and the generation of cDNA libraries: overview, *Meth. Enzymol. 152*, 307-316.

88. Morris, B., Hammer, B., and Mierendorf, R. (1995) A novel strategy for directional cloning of random primed cDNA, *Newsletter of Novagen, Inc. 3*, 1-4.

89. He, A., Li, T., Daniels, L., Fotheringham, I., and Rosazza, J. P. N. (2004) *Nocardia* sp. carboxylic acid reductase: cloning, expression, and characterization of a new aldehyde oxidoreductase family, *Appl. Environ. Microbiol. 70*, 1874-1881.

90. Hendricks, C. L., Ross, J. R., Pickersky, E., Noel, J. P., and Zhou, Z. S. (2004) An enzyme-coupled colorimetric assay for S-adenoxylmethionine dependent methyltrasferases, *Anal. Biochem. 326*, 100-105.

91. Vergin, K. L., Urbach, E., A. Stein, J. L., Delong, E. F., Lanoil, B. D., and Giovannoni, S. J. (1998) Screening of a fosmid library of marine environmental genomic DNA fragments reveals four clones related to members of the order *Planctomycetales*, *Appl. Environ. Microbiol. 64*, 3075-3078.

92. Stein, J. L., Marsh, T. L., Wu, K. Y., Shizuya, H., and DeLong, E. F. (1996) Characterization of uncultivated prokaryotes: isolation and analysis of a 40-kilobase-pair genome fragment from a planktonic arine archaeon, *J. Bacteriol. 178*, 591-599.

93. Chauthaiwale, V. M., Therwath, A., and Deshpande, V. V. (1992) Bacteriophage lambda as a cloning vector, *Microb. Mol. Biol. Rev 56*, 577-591.

94. Seed, B. (1982) Theoretical-study of the fraction of a long-chain DNA that can be incorporated in a recombinant DNA partial-digest library, *Biopolymers 21*, 1793-1810.

95. Murrray, M. E. (1983) *Lambda Vectors, in Lambda II*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

96. Courtois, S., Cappellano, C. M., Ball, M., Francou, F., Normand, P., Helynck, G., Martinez, A., Kolvek, S. J., Hopke, J., Osburne, M. S., August, P. R., Nalin, R., Guerineau, M., Jeannin, P., Simonet, P., and Pernodet, J. L. (2003) Recombinant environmental libraries provide access to microbial diversity for drug discovery from natural products, *Appl. Environ. Microbiol. 69*, 49-55.

97. Sambrook, J., Fritsch, E. F., and Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual*, 2nd ed., Cold Spring Harbor Press, Plainview, NY.

98. Epincentre Technologies. (2005) *CopyControl™ Fosmid Library Porduction Kit with Phage T1-Resistant EPI300™-T1R E. coli Plating Strain manual*, Madison, WI.

99. Bierman, M., Logan, R., O'Brien, K., Seno, E. T., Nagaraja, R. R., and Schoner, B. E. (1992) Plasmid cloning vectors for the conjugal transfer of DNA from *Escherichia coli* to *Streptomyces* spp., *Gene 116*, 43-49.

100. Epicentre Technologies. (2005) *CopyControl™ Fosmid Library Production Kit with pCC1FOS™ Vector and Phage T1-Resistant EPI300™-T1R E. coli Plating Strain*, Madison, WI.

101. Ausubel, F. M., Brent, R., Kingston, R. E., Moore, D. D., Seidman, J. G., and Struhl, K. (1998) *Current Protocols in Molecular Biology*, John Wiley and Sons, New York, NY.

102. Mitchell, J. K., and Carter, W. E. (2000) Modeling antimicrobial activity of cloroxTM using an agar-diffusion test: a new twist on an old experiment, *Bioscene 26*, 9-13.

103. Watanabe, C. M. H., and Townsend, C. A. (1996) Incorporation of molecular oxygen in aflatoxin B-1 biosynthesis, *J. Org. Chem. 61*, 1990-1993.

104. Ausubel, F. M., Brent, R., Kingston, R. E., Moore, D. D., Seidman, J. G., Smith, J. A., and Struhl, K. (1987) *Current Protocols in Molecular Biology1*, John Wiley and Sons, New York, NY.

105. Mortelmans, K., and Zeiger, E. (2000) The ames salmonella/microsome mutagenicity assay, *Mutat. Res 455*, 29-60.

106. Floyd, M. M., Tang, J., Kane, M., and Emerson, D. (2005) Captured diversity in a culture collection: case study of the geographic and habitat distributions of environmental isolates held at the american type culture collection, *Appl. Environ. Microbiol. 71*, 2813-2823.

107. Keller, M., and Zengler, K. (2004) Tapping into microbial diversity, *Nat. Rev. Microbiol. 2*, 141-150.

108. Zhang, L., An, R., Wang, J., Sun, N., Zhang, S., Hu, J., and Kuai, J. (2005) Exploring novel bioactive compounds from marine microbes, *Curr. Opin. Microbio. 8*, 276-281

109. Schloss, P., and Handelsman, J. (2004) Status of the microbial census, *Microbiol. Mol. Rev. 68*, 686-691.

110. Torsvik, V., Øvreås, L., and Thingstad, T. F. (2002) Prokaryotic diversity-magnitude, dynamics, and controlling factors, *Science 296*, 1064-1066.

111. Beja, O., Suzuki, M. T., Koonin, E. V., Aravind, L., Hadd, A., Nguyen, L. P., Villacorta, R., Amjadi, M., Garrigues, C., Jovanovich, S. B., Feldman, R. A., and DeLong, E. F. (2000) Construction and analysis of bacterial artificial chromosome libraries from a marine microbial assemblage, *Environ. Microbiol. 2*, 516-529.

112. Connon, S. A., and Giovannoni, S. J. (2002) High-throughput methods for culturing microorganisms in very-low-nutrient media yield diverse new marine isolates, *Appl. Environ. Microbiol. 68*, 3878-3885.

113. Oh, D. C., Kauffman, C. A., Jensen, P. R., and Fenical, W. (2007) Induced production of the emericellamides A and B, from the marine-derived fungus *Emericella sp*. in competing co-culture, *J. Nat.Prod. 70*, 515-520.

114. Christian, O. E., Compton, J., Christian, K. R., Mooberry, S. L., Valeriote, F. A., and Crews, P. (2005) Using jasplakinolide to turn on pathways that enable the isolation of new chaetoglobosins from *Phomospis asparagi*, *J. Nat.Prod. 68*, 1592-1597.

115. Brady, S. F., and Clardy, J. (2005) Cloning and heterologous expression of isocyanide biosynthetic genes from environmental DNA, *Angew. Chem., Int. Ed. 44*, 7063-7065.

116. Gillespie, D. E., Brady, S. F., Bettermann, A. D., Cianciotto, N. P., Liles, M. R., Rondon, M. R., Clardy, J., Goodman, R. M., and Handelsman, J. (2002) Isolation of antibiotics turbomycin A and B from a metagenomic library of soil microbial DNA, *Appl. Environ. Microbiol. 68*, 4301-4306.

117. Peterson, D. G., Tomkins, J. P., Frisch, D. A., Wing, R. A., and Patterson, A. H. (2000) Construction of plant bacterial artificial chromosome libraries: an illustrated guide, *J. Agric. Genomics. 5*, 1-100.

118. Piel, J. (2002) A polyketide synthase-peptide synthetase gene cluster from an uncultured bacterial symbiont of Paederus beetles, *Proc. Natl. Acad. Sci. U.S.A. 99*, 14002-14007.

119. Piel, J., Hui, D., Wen, G., Butzke, D., Platzer, M., Fusetani, N., and Matsunaga, S. (2004) Antitumor polyketide biosynthesis by an uncultivated bacterial symbiont of the marine sponge *Theonella swinhoei*, *Proc. Natl. Acad. Sci. U.S.A. 101*, 16222-16227.

120. Schirmer, A., Gadkari, R., Reeves, C. D., Ibrahim, F., DeLong, E. F., and Hutchinson, C. R. (2005) Metagenomic analysis reveals diverse polyketide synthase gene clusters in microorganisms associated with the marine sponge *discodermia dissolute*, *Appl. Environ. Microbiol. 71*, 4840-4849.

121. Schmidt, E. W., Nelson, J. T., Rasko, D. A., Sudek, S., Eisen, J. A., Haygood, M. G., and Ravel, J. (2005) Patellamide A and C biosynthesis by a microcin-like pathway in *Prochloron didemni*, the cyanobacterial symbiont of *Lissoclinum patella, Proc. Natl. Acad. Sci. U.S.A. 102*, 7315-7320.

122. Zhang, H. B. (2000) *Construction and Manipulation of Large-insert Bacterial Clone Libraries-Manual*, Texas A&M University, College Station, <http://hbz7. tamu. edu/homelinks/tool/bac_content.htm.>

123. Davis, R. H., and De Serres, F. J. (1970) Genetic and microbial research techniques for *Neurospora crassa*, *Methods Enzymol. 17A,* 79-143.

124. Johnston, J. S., Bennett, M. D., Rayburn, A. L., Galbraith, D. W., and Price, H. J. (1999) Reference standards for determination of DNA content of plant nuclei, *Am. J. Bot. 86*, 609-613.

125. Clarke, L., and Carbon, J. (1976) A colony bank containing synthetic Col EI hybrid plasmids representative of the entire *E. coli* genome, *Cell 9*, 91-99.

126. Bradshaw, M. S., Bollekens, J. A., and Ruddle, F. H. (1995) A new vector for recombination based cloning of large DNA fragments from yeast artificial chromosomes, *Nucleic Acids Res. 23*, 4850-4856.

127. Bennett, M. D., and Leitch, I. J. (1995) Nuclear DNA amounts in angiosperms, *Ann. Bot. 76,* 113-176

128. Hoffman, R. A. (2001) Standardization and quantitation in flow cytometry, *Meth. Cell Biol. 63*, 299-340.

129. Begley, T. P., Xi, J., Kinsland, C., Taylor, S., and McLafferty, F. (1999) The enzymology of sulfur activation during thiamin and biotin biosynthesis, *Curr. Opin. Chem. Biol. 3,* 623-629.

130. Bower, S., Perkins, J. B., Yocum, R. R., Howitt, C. L., Rahaim, P., and Pero, J. (1996) Cloning, sequencing, and characterization of the *Bacillus subtilis* biotin biosynthetic operon, *J. Bact. 178*, 4122-4130.

131. Cronan, J. J. E. (1989) The *E. coli* bio operon: transcriptional repression by an essential protein modification enzyme, *Cell 58*, 427-429.

132. Izumi, Y., Kano, Y., Inagaski, K., Kawase, N., Tani, Y., and Yamada, H. (1981) Characterization of biotin biosynthetic enzymes of *Bacillus sphaericus*: a desthiobiotin producing bacterium, *Agric. Biol. Chem. 45*, 1983-1989.

133. Glöckner, F. O., Kube, M., Bauer, M., Teeling, H., Lombardot, T., Ludwig, W., Gade, D., Beck, A., Borzym, K., Heitmann, K., Rabus, R., Schlesner, H., Amann, R., and Reinhardt, R. (2003) Complete genome sequence of the marine planctomycete *Pirellula sp.* strain 1, *Proc. Natl. Acad. Sci. U.S.A. 100*, 8298-8303.

134. Hou, S., Saw, J. H., Lee, K. S., Freitas, T. A., Belisle, C., Kawarabayasi, Y., Donachie, S. P., Pikina, A., Galperin, M. Y., Koonin, E. V., Makarova, K. S., Omelchenko, M. V., Sorokin, A., Wolf, Y. I., Li, Q. X., Keum, Y. S., Campbell, S., Denery, J., Aizawa, S., Shibata, S., Malahoff, A., and Alam, M. (2004) Genome sequence of the deep-sea -proteobacterium *Idiomarina loihiensis* reveals amino acid fermentation as a source of carbon and energy, *Proc. Natl. Acad. Sci. U. S. A. 101*, 18036-18041

135. Entcheva, P., Liebl, W., Johann, A., Hartsch , T., and Streit, W. R. (2001) Direct cloning from enrichment cultures, a reliable strategy for isolation of complete operons and genes from microbial consortia, *Appl. Environ. Microbiol. 67*, 89-99.

136. Hardie, D. C., Gregory, T. R., and Hebert, P. D. N. (2002) From pixels to picograms: a beginners' guide to genome quantification by feulgen image analysis densitometry, *J. Hist. Cytochem. 50*, 735-749.

137. Nishikubo, T., Nakgawa, N., Kuramitsu, S., and Mashi, R. (2005) Improved heterologous gene expression in *Escherichia coli* by optimization of the AT-content of codons immediately downstream of the initiation codon, *J. Biotec. 120*, 341-346.

138. Sorensen, H. P., and Mortensen, K. K. (2005) Advanced genetic strategies for recombinant protein expression in *Escherichia coli*, *J. Biotec 115*, 113-128.

139. Jensen, P. R., Westerhoff, H. V., and Michelsen, O. (1993) The use of lac-type promoters in control analysis, *Eur. J. Biochem. 211*, 181-191.

140. Lee, S. K., Chou, H. H., Pfleger, B. F., Newman, D. J., Yoshikuni, Y., and Keasling, J. D. (2007) Directed evolution of AraC for improved compatibility of arabinose- and lactose-inducible promoters, *Appl. Environ. Microbiol. 73*, 5711-5715.

141. Guzman, L.-M., Belin, D., Carson, M. J., and Beckwith, J. (1995) Tight regulation, modulation, and high-level expression by vectors containing the arabinose $P_{BAD}$ promoter, *J. Bacteriol. 177*, 4121-4130.

142. Lobell, R. B., and Schleif, R. F. (1990) DNA looping and unlooping by AraC protein, *Science 250*, 528-532.

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

144. Novick, A., and Weiner, M. (1957) Enzymeinduction as an all-or-none phenomenon, *Biochemistry 43*, 553-566.

145. Serganov, A., Polonskaia, A., Phan, A. T., Breaker, R. R., and Patel, D. J. (2006) Structural basis for gene regulation by a thiamine pyrophosphate-sensing riboswitch, *Nature 441*, 1167-1171.

146. Winkler, W., Nahvi, A., and Breaker, R. R. (2002) Thiamine derivatives bind messenger RNAs directly to regulate bacterial gene expression, *Nature 419*, 952-956.

147. Sudarsan, N., Barrick, J. E., and Breaker, R. R. (2003) Metabolite-binding RNA domains are present in the genes of eukaryotes, *RNA 9*, 644-647

148. August, P. R., Grossman, T. H., Minor, C., Draper, M. P., MacNeil, I. A., Pemberton, J. M., Call, K. M., Holt, D., and Osburne, M. S. (2000) Sequence analysis and functional characterization of the violacein biosynthetic pathway from *Chromobacterium violaceum*, *J. Mol. Microbiol. Biotechnol. 2*, 513-519.

149. Balibar, C. J., and Walsh, C. T. (2006) *In vitro* biosynthesis of violacein from l-tryptophan by the enzymes VioA−E from *Chromobacterium violaceum*, *Biochemistry 45*, 15444-15457.

150. Pemberton, J. M., Vincent, K. M., and Penfold, R. J. (1991) Cloning and heterologous expression of the violacein biosynthesis gene cluster from *Chromobacterium violaceum*, *Curr. Microbiol. 22*, 355-358.

151. Sánchez, C., Braña, A. F., Méndez, C., and Salas, J. A. (2006) Reevaluation of the violacein biosynthetic pathway and its relationship to indolocarbazole biosynthesis *Chem. Bio. Chem. 7*, 1231-1240.

152. Berne, P. F., Doublies, S., and Carter, C. W. J. (1999) *Crystallization of nucleic acids and proteins. a practical approach*, Oxford University Press, Oxford.

153. Mitraki, A., and King, J. (1989) Protein folding intermediates and inclusion body formation, *BioTechnol. 7*, 690-697.

154. Stevens, R. C. (2000) Design of high-throughput methods of protein production of structural biology, *Structure 8*, R177-R185.

155. Heckman, K. L., and Pease, L. R. (2007) Gene splicing and mutagenesis by PCR-driven overlap extension, *Nature Protocols 2*, 924-932.

156. Pantanella, F., Berlutti, F., Passariello, C., Sarli, S., Morea, C., and Schippa, S. (2006) Violacein and biofilm production in *Janthinobacterium lividum*, *J. Appl. Microbiol 102*, 992-999.

157. Rettori, D., and Duran, N. (1998) Production, extraction and purification of violacein: an antibiotic pigment produced by *Chromobacterium violaceum*, *World Journal of Microbiology & Biotechnology 14*, 685-688.

158. Moss, M. O., Ryall, C., and Logan, N. A. (1978) The classification and characterization of chromobacteria from a lowland river, *J. Gen. Microbiol. 105*, 11-21.

159. Tobie, W. C. (1934) The pigment of *bacillus violaceus* I. The production, extraction, and purification of violacein, *J. Bact. 29*, 223-227.

160. Blosser, R. S., and Gray, K. M. (2000) Extraction of violacein from *Chromobacterium violaceum* provides a new quantitatvie bioassay for N-acyl homoserine lactone autoinducers, *J. Microbiol. Methods 40*, 47-55.

161. Morohosi, T., Kato, M., Fukamachi, K., Kato, N., and Ikeda, T. (2008) *N*-Acylhomoserine lactone regulates violacein production in *chromobacterium violaceum* type strain ATCC12472, *FEMS Microbiol Lett 279*, 124-130.

162. Lu, Y., Wang, J., Xue, Y., Zhang, C., Xing, X., Lou, K., Zhang, Z., Li, Y., Zhang, G., Bi, J., and Su, Z. (2009) Production of violet pigment by a newly isolated psychrotrophic bacterium from a glacier in Xinjiang, China, *Biochemical Engineering Journal 43*, 135-141.

163. Mendes, A., Carvalho, J. E., Duarte, M. C. T., Duran, N., and Bruns, R. E. (2001) Factorial design and response surface optimization of crude violacein for *Chromobacterium violaceum* production, *Biotechnol. Lett. 23*, 1963-1969.

164. Sudarsan, N., E., B. J., and Breaker, R. R. (2003) Metabolite-binding RNA domains are present in the tenes of eukaryotes, *RNA 9*, 644-647.

165. Cheah, M. T., Wachter, A., Sudarsan, N., and Breaker, R. R. (2007) Control of alternative RNA splicing and gene expression by eukaryotic riboswitches, *Nature 447*, 497-500.

166. Laemmli, U. K. (1970) Cleavage of structural proteins during the assembly of the head of bacteriophage T4, *Nature 227*, 680-685.

167. Egelhoff, T. T., and Long, S. R. (1985) *Rhizobium meliloti* nodulation genes: identification of *nodDABC* gene products, purification of *nodA* protein and expression of *nodA* in *Rhizobium meliloti*, *J. Bact. 164*, 591-599.

168. Brazillian National Genome Project Consortium: Edited by Haselkorn, R. (2003) The complete genome sequence of *Chromobacterium violaceum* reveals remarkable and exploitable bacterial adaptability, *PNAS 100*, 11660-11665.

169. Walsh, C. T., and Balibar, C. J. (2006) In vitro biosynthesis of violacein from L-tryptophan by the enzymes VioA-E from *Chromobacterium violaceum*, *Biochemistry 45*, 15444-15457.

170. Novagen, Inc. (1997) *T7 Terminator Sequence from Novagene Catalog*, Novagen Inc., Gibbstown, NJ

171. Hoshino, T., Kondo, T., Uchiyama, T., and Ogasawara, N. (1987) Biosynthesis of violacein: a novel rearrangement in Tryptophan metabolism with 1,2-Shift of the indole ring, *Agric. Biol. Chem. 51*, 965-968.

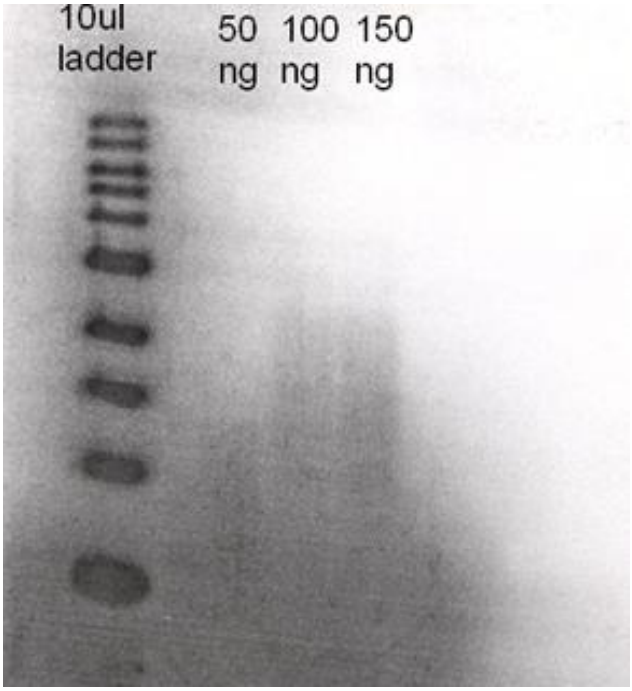
172. Nakamura, Y., Sawada, T., Morita, Y., and Tamiya, E. (2002) Isolation of a psychrotrophic bacterium from the organic residue of a water tank keeping rainbow trout and antibacterial effect of violet pigment produced the strain, *Biochemical Engineeing Journal 12*, 79-86.

173. Huang, W., McKevitt, M., and Palzkill, T. (2000) Use of the arabinose pbad promoter for tightly regulated display of proteins on bacteriophage, *Gene 251*, 187-197.

174. Blount, K., and Breaker, R. R. (2006) Riboswitches as antibacterial drug targets, *Nat. Biotechol. 24*, 1558-1564.

# APPENDIX A

**A: Pup16 1st strand**




**B: Pup16 2nd strand**

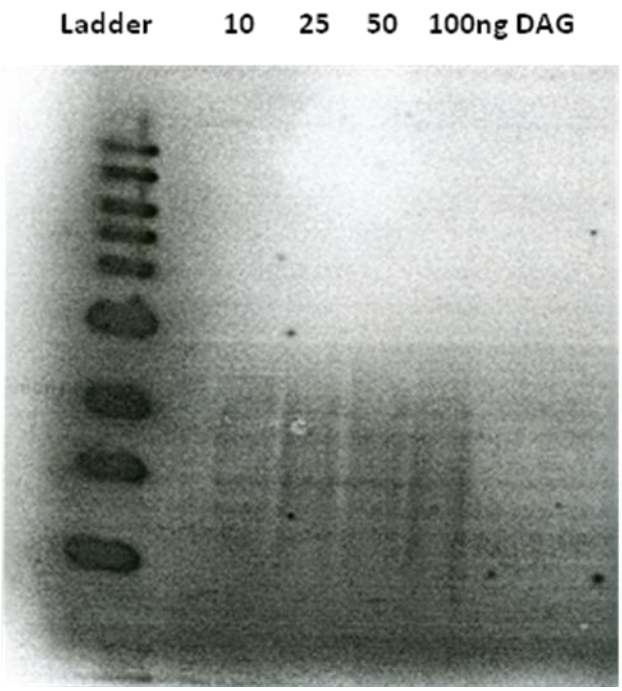
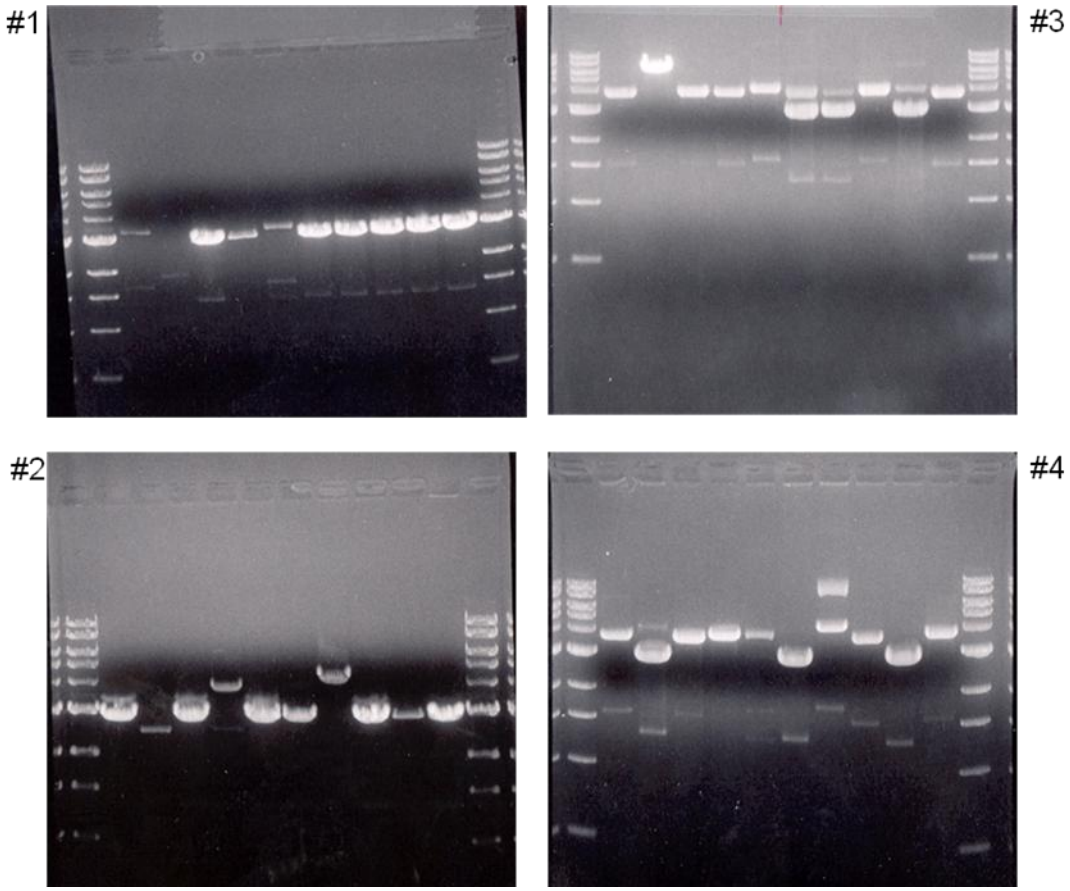


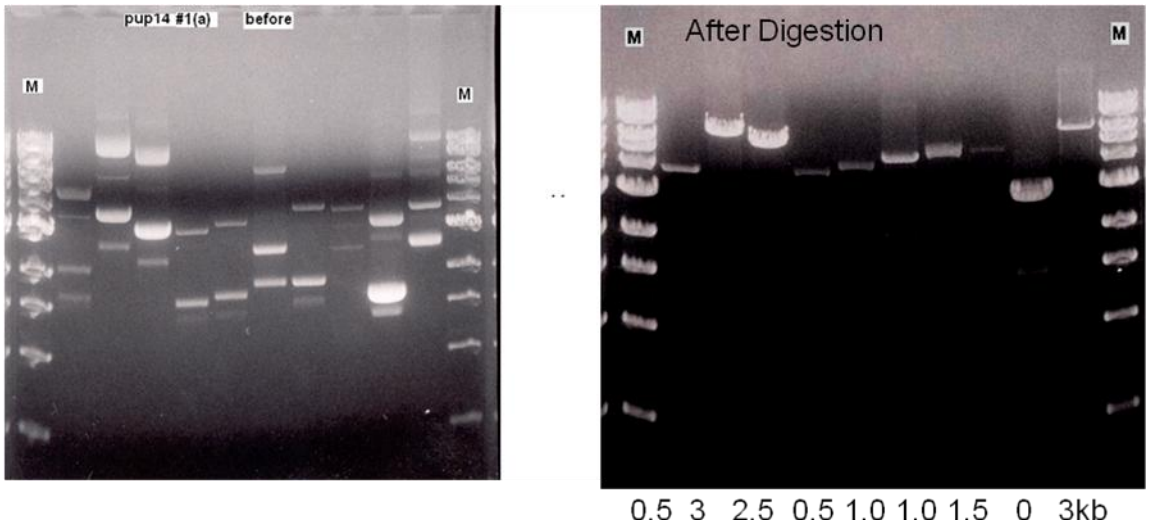

**Figure A.1: Phosphorimage A: Pup16 1st strand B: Pup16 2nd strand**

## A: Pup 14

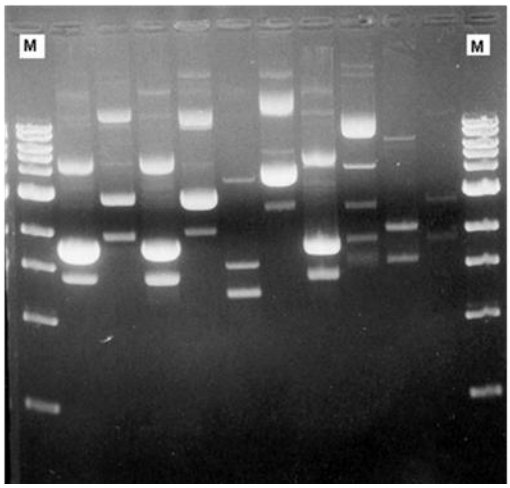
**Pup14 (EC100)**

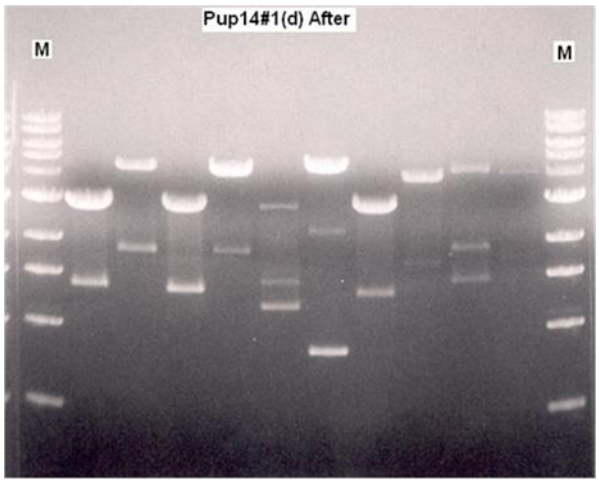



**Pup14 (Zapper) before and after digestion**

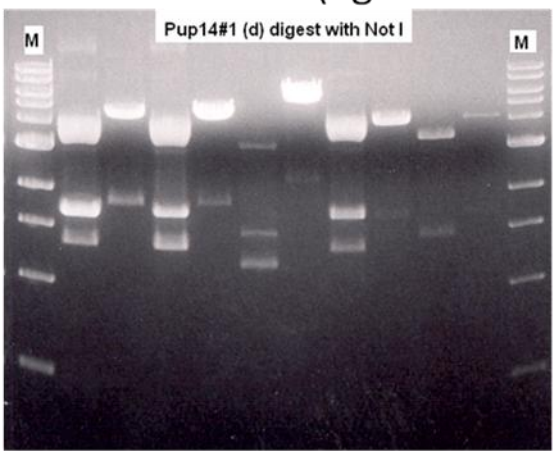


Pup14#1 (d) Before Digestion

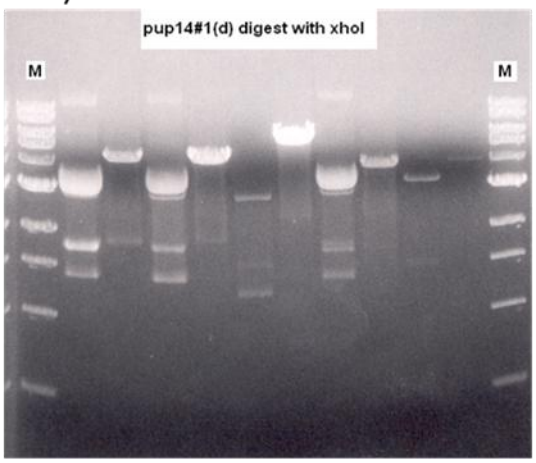


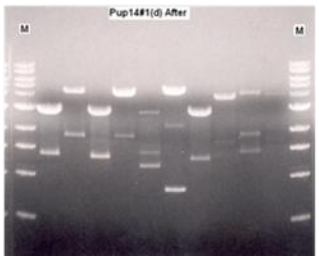

- Digestion with other enzyme which does not contain the arm site. (eg. NotI or XhoI)




**NotI**



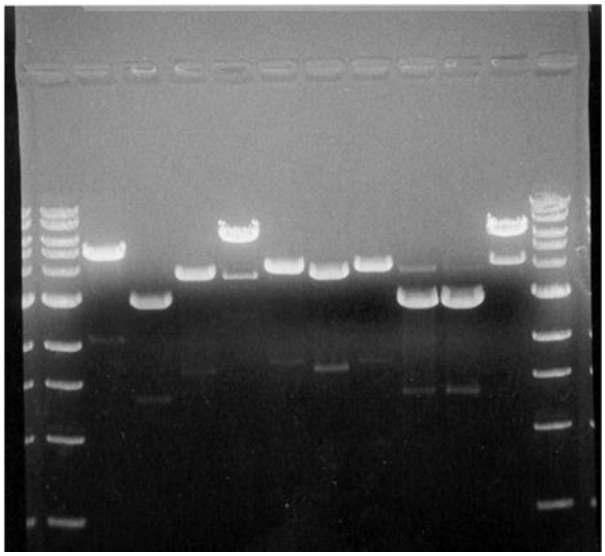

**xhol**




**EcoRI**

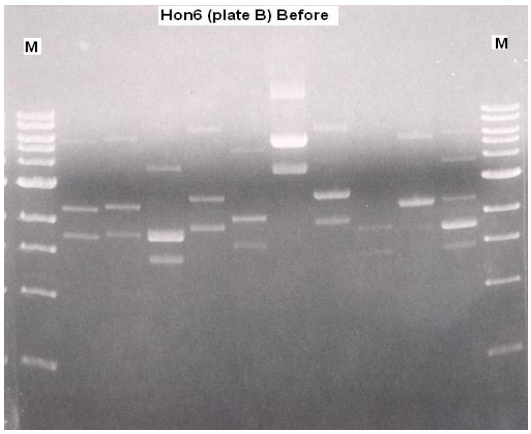
## B: Hon 6

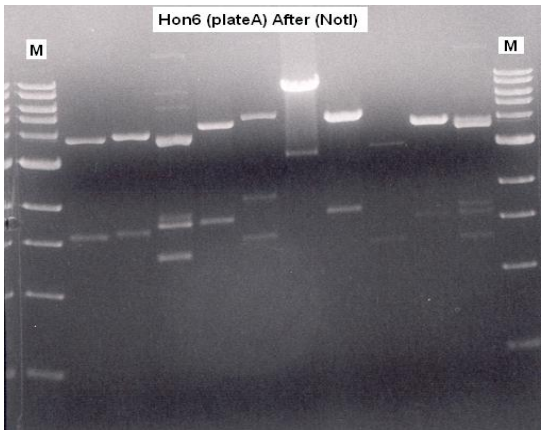
### Hon6 ( EC100)

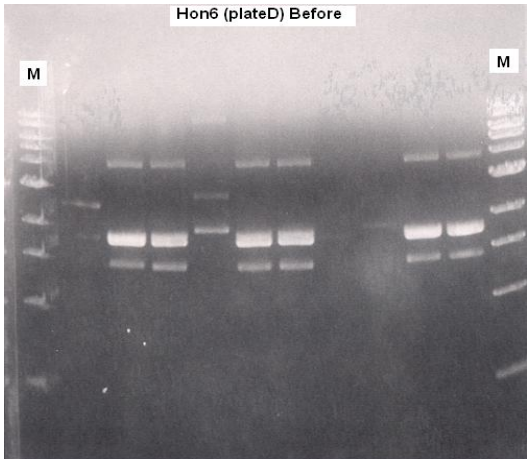


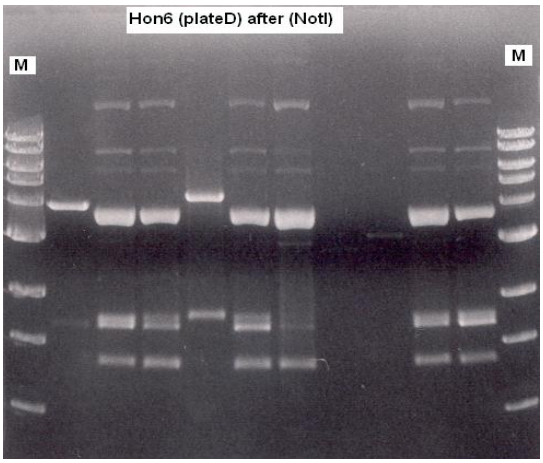
Insertion: 8/11
SIZE: 0.5Kb 7
1.5kb 1
3.0kb 2

### Hon 6 before and after digestion with NotI



## C: Pup16

**Pup 16 (Zapper) before and after digestion**

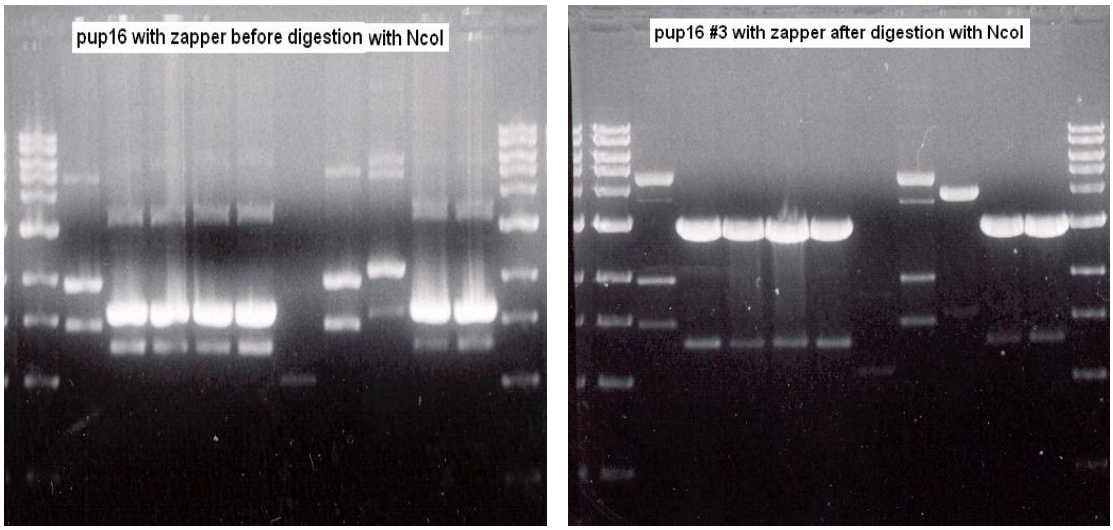
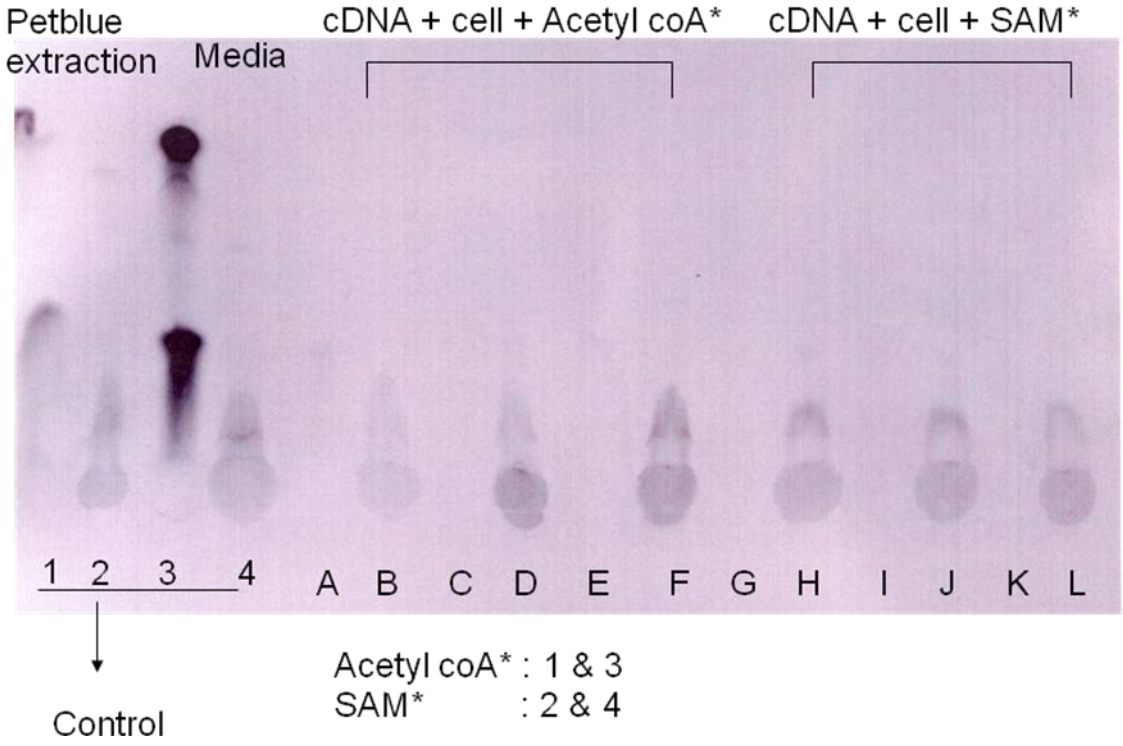



**Figure A.2: Insertion size**




**Control: pETBlue+Zapper**

**1. pETBlue protein extract + Acetyl CoA enzyme(radioactive) + cofactors (NADPH, FAD, SAM)**
**2. pETBlue protein extract + SAM (radioactive) + cofactors (NADPH, FAD)**
**3.pETBlue Lyophilized media + Acetyl CoA enzyme(radioactive) + cofactors (NADPH, FAD, SAM)**
**4. pETBlue Lyophilized media + SAM(radioactive) + cofactors (NADPH, FAD)**

**A: cDNA protein extract + unactivated media + Acetyl CoA*+ NADPH+FAD+SAM**
**B: cDNA protein extract + Cell + Acetyl CoA*+ NADPH+FAD+SAM**
**C: cDNA protein extract + half-activated media + Acetyl CoA*+ NADPH+FAD+SAM**
**D: cDNA protein extract + Cell + Acetyl CoA*+ NADPH+FAD+SAM**
**E: cDNA protein extract + fully-activated media + Acetyl CoA*+ NADPH+FAD+SAM**
**F: cDNA protein extract + Cell + Acetyl CoA*+ NADPH+FAD+SAM**
**G: cDNA protein extract + unactivated media + SAM*+ NADPH+FAD**
**H: cDNA protein extract + Cell + SAM*+ NADPH+FAD**
**I: cDNA protein extract + half-activated media + SAM*+ NADPH+FAD**
**J: cDNA protein extract + Cell + SAM*+ NADPH+FAD**
**K: cDNA protein extract + fully-activated media + SAM*+ NADPH+FAD**
**L: cDNA protein extract + Cell + SAM*+ NADPH+FAD**

**Figure A.3: Phosphorous image of the cDNA protein Extract (Pup14)**

## A: SAM-dependent Methyltransferase Pup14A

**RX#1 ---expected PCR Product size (252bp)**

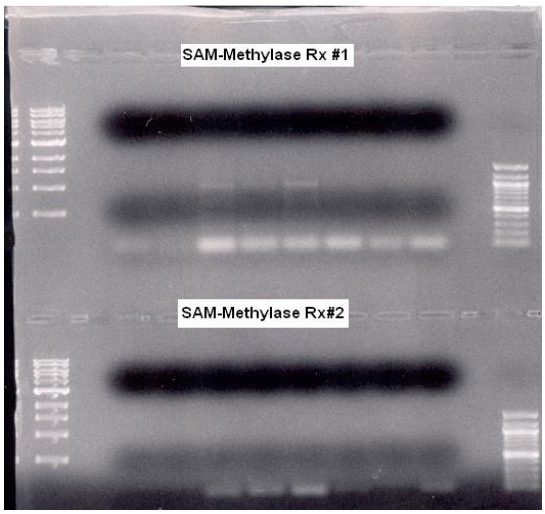
**RX#2 ---expected PCR Product size (540bp)**

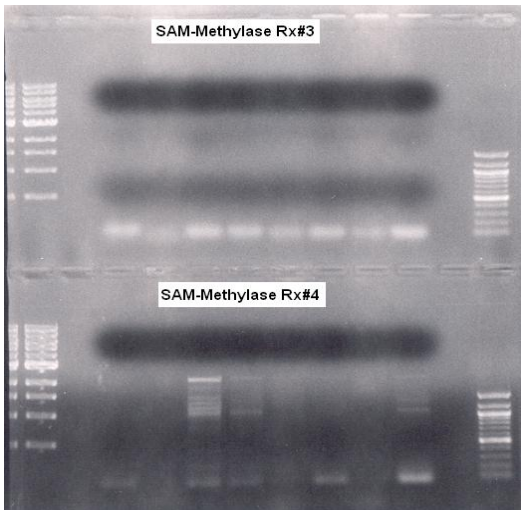
**RX#3 ---expected PCR Product size (948bp)**

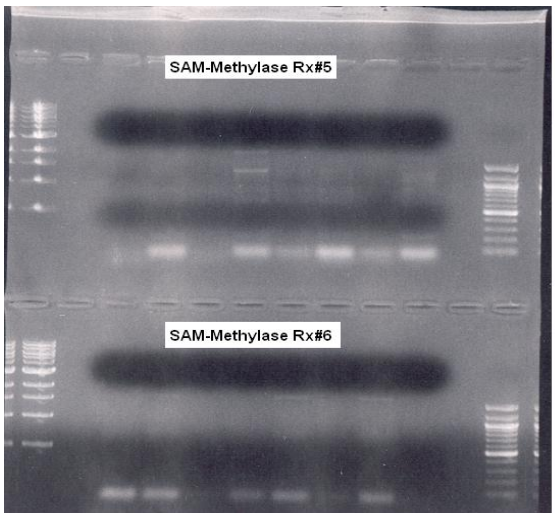
**RX#4 --- expected PCR Product size (258bp)**

**RX#5 --- expected PCR Product size (666bp)**

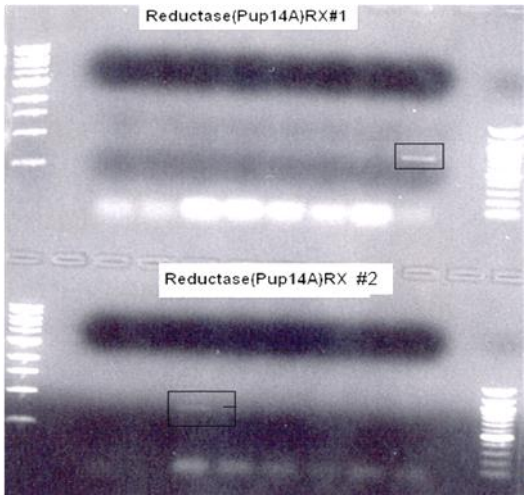
**RX#6 --- expected PCR Product size (378bp)**

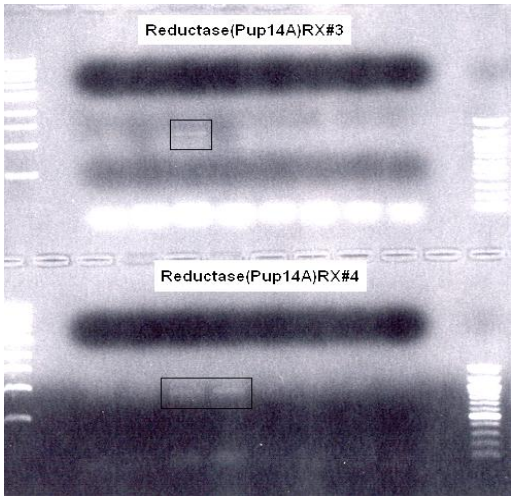


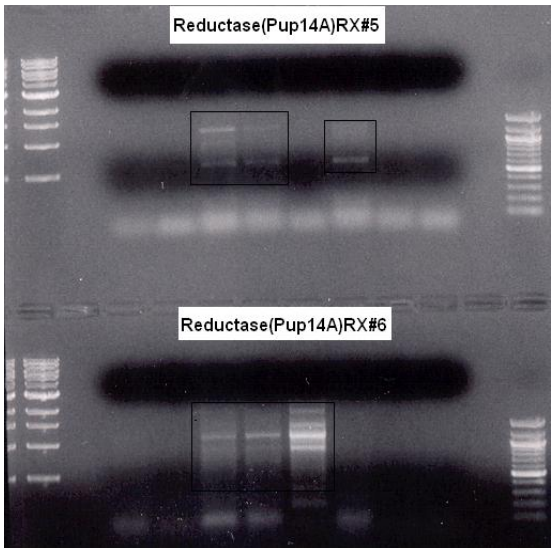



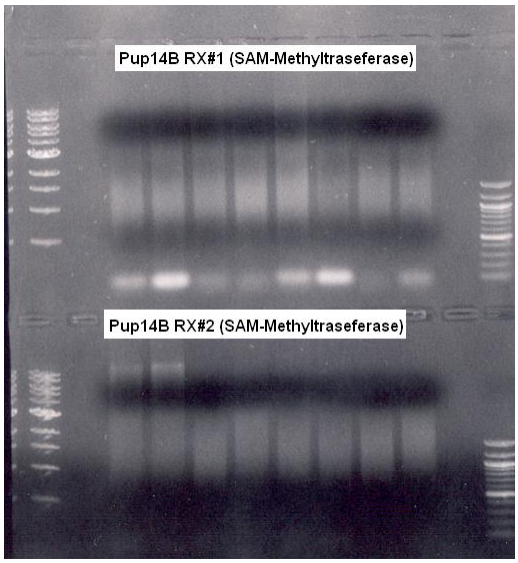
## B : Reductase Pup14A

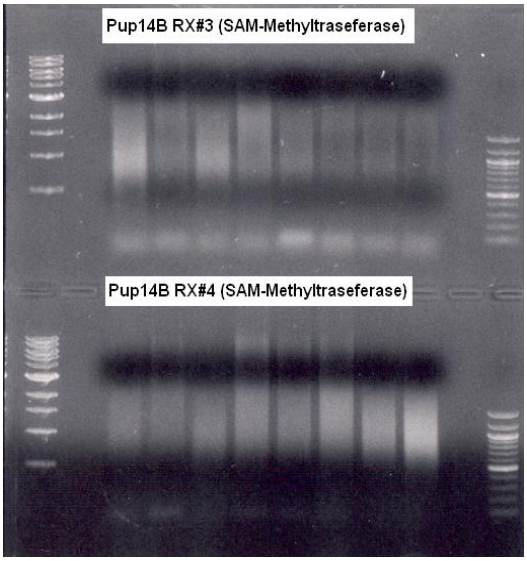


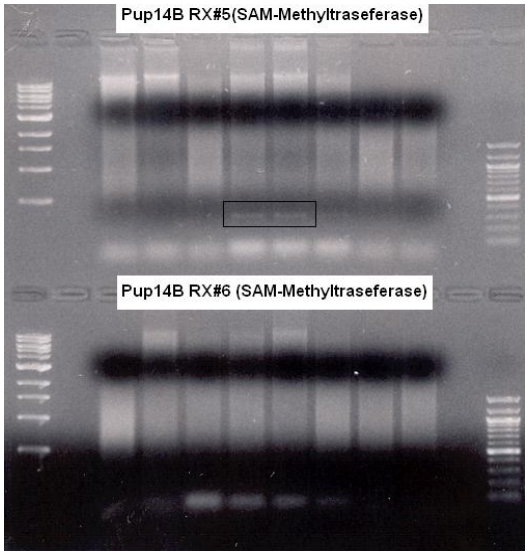



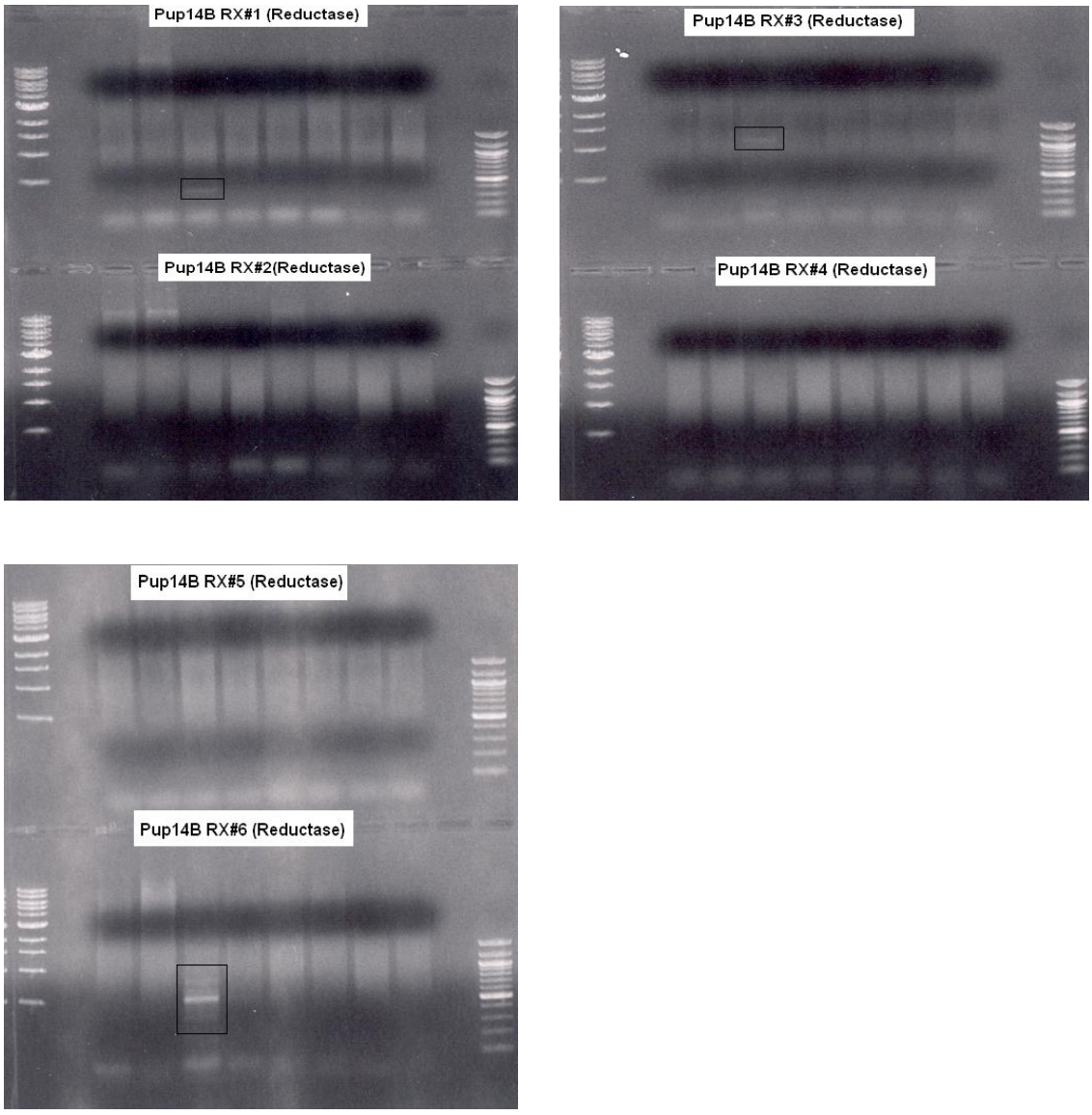
## C: Pup 14B SAM-dependent Methyltransferase



## D: Pup 14B Reductase

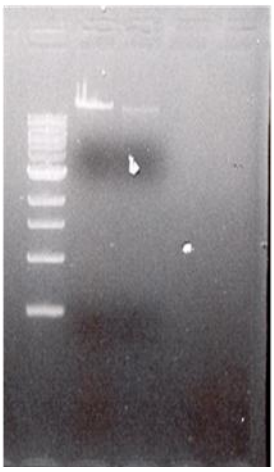



- **PCR results**
  - **SAM-dependent Methyltransferase**
    **Pup14A—Rx 1,4,5,6**
    **Pup14B— Rx 5**
    **Hon6— no hit**
  - **Reductase (Pup14A)**
    **Pup14A— Rx 1,2,3,4,5,6**
    **Pup14B— Rx 1,3,6**
    **Hon6 — no hit**

**Figure A.4: PCR**

**APPENDIX B**

**Supplementary Data for Chapter III**

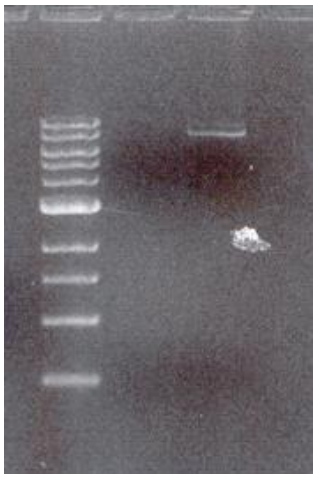


1st: ladder

2nd: control DNA

3rd: C3 (after 5 times pipetting)

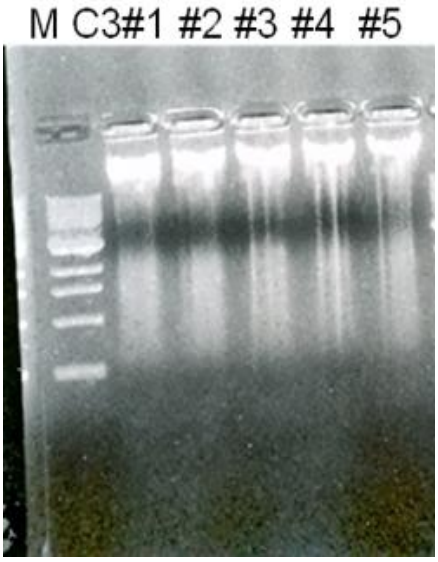
No more shearing needed!!



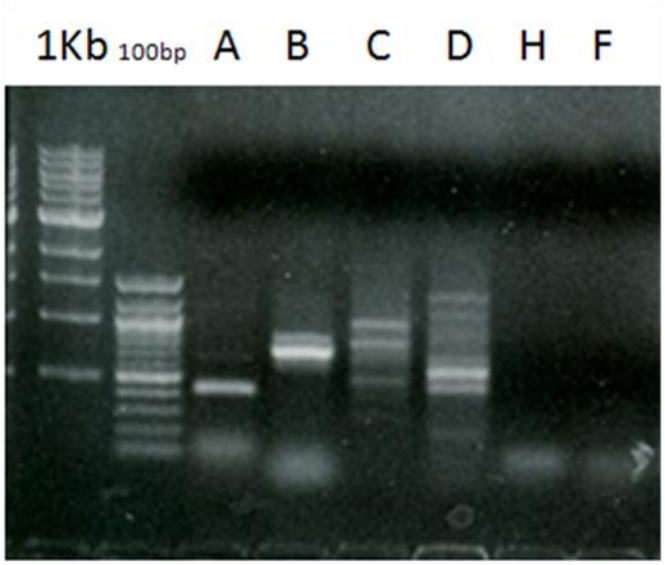
DNA extract

C3 samples after dephosphorylation (25ng)

Plasmid isolated from C3 fosmid library



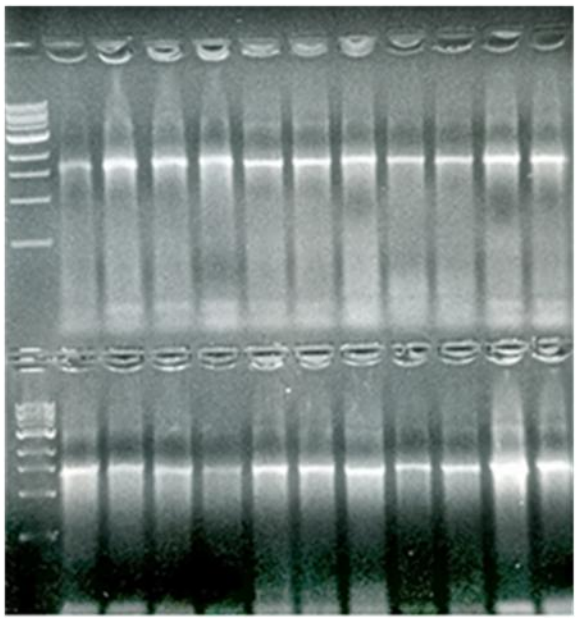
## Biotin biosynthetic pathway screening of C3 Fosmid library (plate#1): PCR



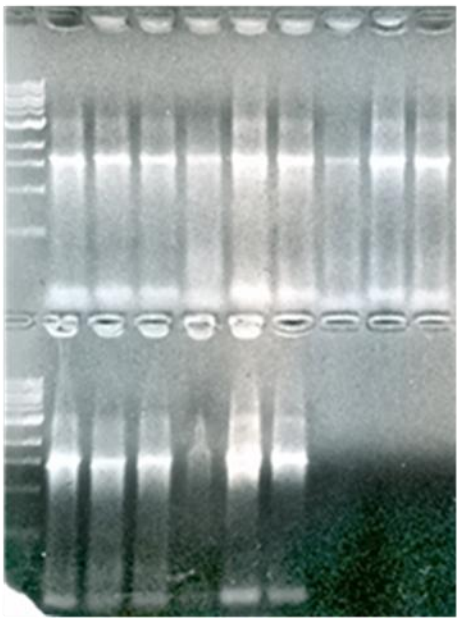

1kb: 1kb ladder
100bp: 100bp ladder
A: BioA PCR product
B: BioA PCR product
C: BioC PCR product
D: BioD PCR product
H: BioH PCR product
F: BioF PCR product

## PCR with 16S rDNA

Meso fosmid library



Bahama fosmid library



**C3 fosmid library**

- **Extract PCR products**
  - **Subsequecing PCR results**
  - **Bahama samples from salt pond (Foward: No *E.coli* Reverse: *E. coli*)**
  - **C3 (Foward: No *E. coli*, Reverse: Containing 95~98% *E. coli*)**
  - **Meso (Forward: No *E. coli*, Reverse: Containing 95~98% *E. coli*)**

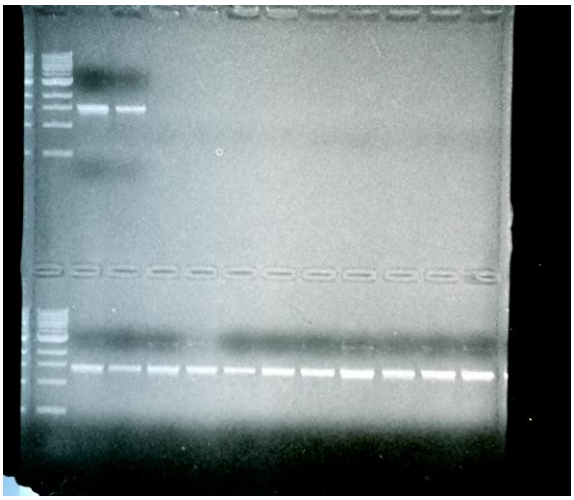


**Figure B.1: Supplementary data of C3 Fosmid library**

**Biofilm NRPS and PKS Screening Conditions:**

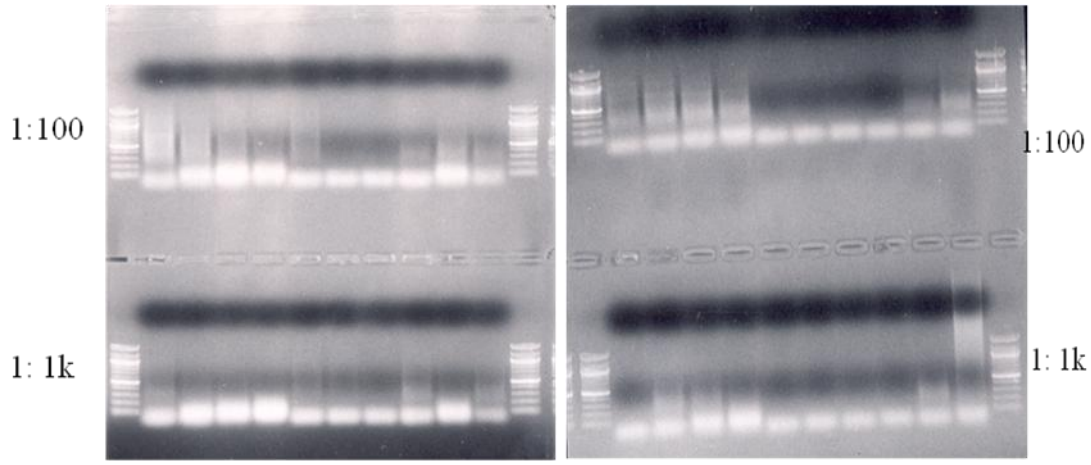
| | |
|---|---|
| Template (20 ng/ul) | 1 ul |
| Primer 1 (100 pmol/ul) | 3 ul |
| Primer 2 (100 pmol/ul) | 3 ul |
| Taq (1 U/ul) | 1 ul |
| ThermoPol Buffer (10X) | 10 ul |
| dNTP's (25 mM) | 2 ul |
| $H_2O$ | 80 ul |
| Total | 100 ul |

Thermocycler settings: Touch-down program

| | | |
|---|---|---|
| 96 °C | 2 min. | |
| 96 °C | 30 sec | 20 cycles |
| 55 °C -.5/cycle | 30 sec | |
| 68 °C | 1 min | |
| 96 °C | 30 sec | 20 cycles |
| 50 °C | 30 sec | |
| 68 °C | 1 min | |
| 68 °C | 7 min | |
| 4 °C | forever | |

**PCR reaction with PKS primers (Biofilm)**

PKSfrom BioRIGHT: 5'-ACA CTG GCG GCC GTT ACT A-3' 58.9°C
PKSfromBioLEFT: 5'-CTG TCC AAC CCC ACT GTT G-3' 56.1°C




**12 conditions for PCR**

#1: 1µl of DNA, 1µl of Forward primer, 1µl of Reverse primer, 0.5µl of Taq DNA polymerase, 5µl of 10x buffer, 0.5µl of dNTP, 41µl of $H_2O$

#2: 1µl of DNA, 1µl of Forward primer, 1µl of Reverse primer, 0.5µl of Taq DNA polymerase, 5µl of 10x buffer, 0.5µl of dNTP, 38µl of $H_2O$, 3µl of $MgCl_2$

#3: 1µl of DNA, 1µl of Forward primer, 1µl of Reverse primer, 0.5µl of Taq DNA polymerase, 5µl of 10x buffer, 0.5µl of dNTP, 36µl of $H_2O$, 5µl of $MgCl_2$

#4: 1µl of DNA, 1µl of Forward primer, 1µl of Reverse primer, 0.5µl of Taq DNA polymerase, 5µl of 10x buffer, 0.5µl of dNTP, 34µl of $H_2O$, 7µl of $MgCl_2$

#5: 1µl of DNA, 1µl of Forward primer, 1µl of Reverse primer, 0.5µl of Taq DNA polymerase, 5µl of 10x buffer, 0.5µl of dNTP, 38.5µl of $H_2O$, 2.5µ of glycerol

#6: 1µl of DNA, 1µl of Forward primer, 1µl of Reverse primer, 0.5µl of Taq DNA polymerase, 5µl of 10x buffer, 0.5µl of dNTP, 35.5µl of $H_2O$, 2.5µ of glycerol, 3µl of $MgCl_2$

#7: 1µl of DNA, 1µl of Forward primer, 1µl of Reverse primer, 0.5µl of Taq DNA polymerase, 5µl of 10x buffer, 0.5µl of dNTP, 33.5µl of $H_2O$, 2.5µ of glycerol, 5µl of $MgCl_2$

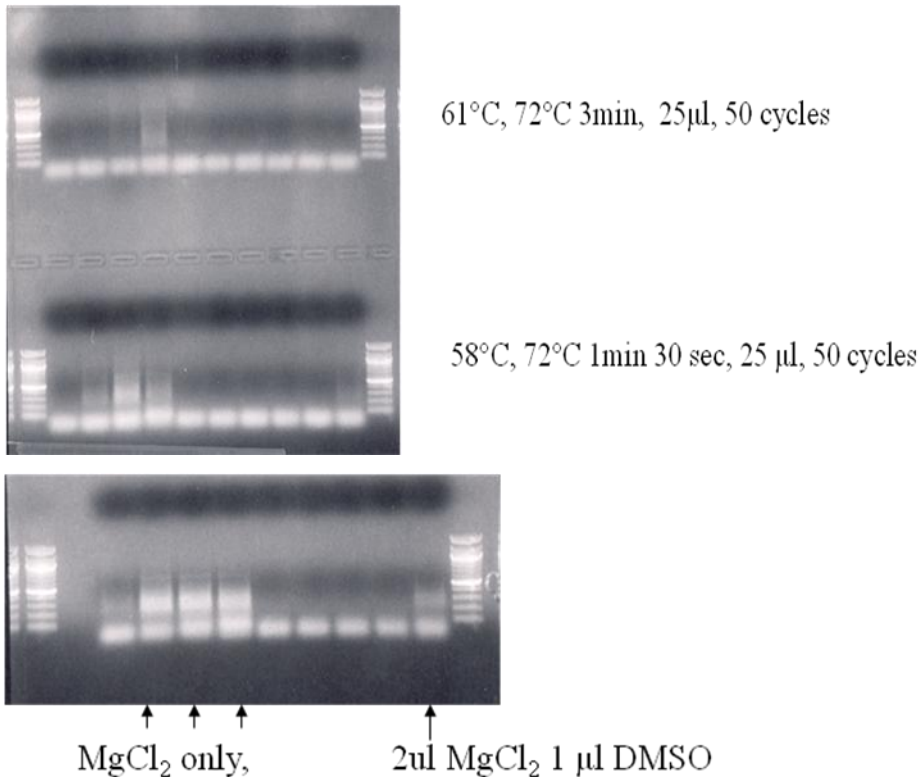
#8: 1µl of DNA, 1µl of Forward primer, 1µl of Reverse primer, 0.5µl of Taq DNA polymerase, 5µl of 10x buffer, 0.5µl of dNTP, 31.5µl of $H_2O$, 2.5µ of glycerol, 7µl of $MgCl_2$

#9: 1µl of DNA, 1µl of Forward primer, 1µl of Reverse primer, 0.5µl of Taq DNA polymerase, 5µl of 10x buffer, 0.5µl of dNTP, 38.5µl of $H_2O$, 2.5µ of DMSO

#10: 1µl of DNA, 1µl of Forward primer, 1µl of Reverse primer, 0.5µl of Taq DNA polymerase, 5µl of 10x buffer, 0.5µl of dNTP, 35.5µl of $H_2O$, 2.5µ of DMSO, 3µl of $MgCl_2$

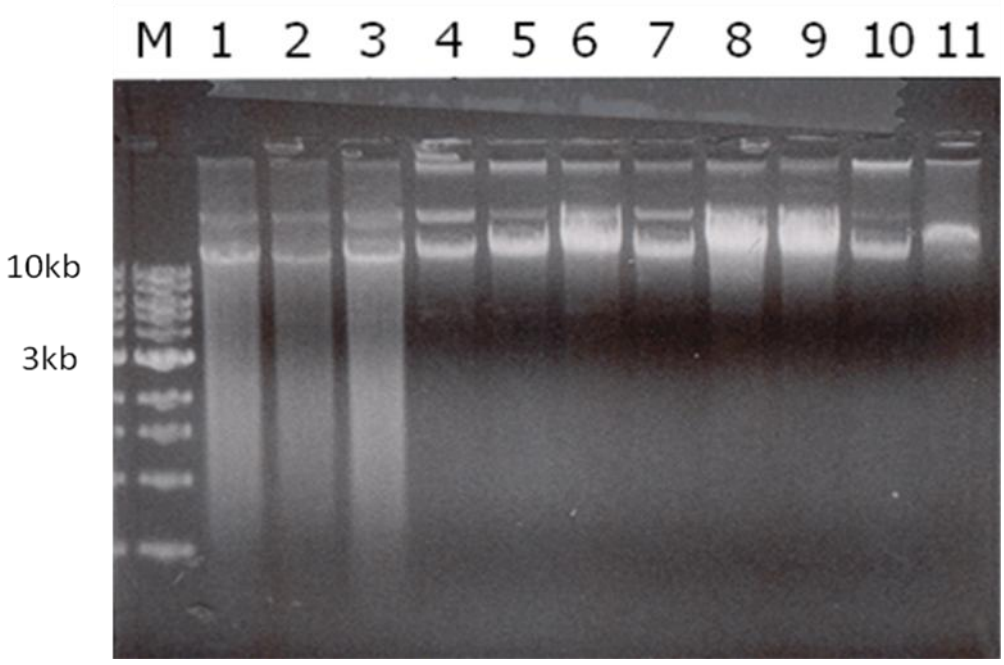
#11: 1µl of DNA, 1µl of Forward primer, 1µl of Reverse primer, 0.5µl of Taq DNA polymerase, 5µl of 10x buffer, 0.5µl of dNTP, 33.5µl of $H_2O$, 2.5µ of DMSO, 5µl of $MgCl_2$

#12: 1µl of DNA, 1µl of Forward primer, 1µl of Reverse primer, 0.5µl of Taq DNA polymerase, 5µl of 10x buffer, 0.5µl of dNTP, 31.5µl of $H_2O$, 2.5µ of DMSO, 7µl of $MgCl_2$

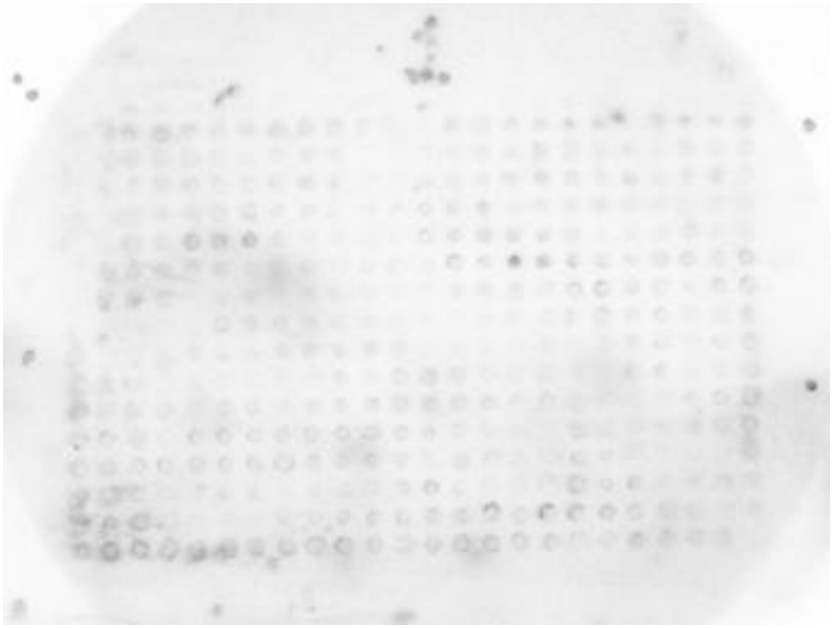


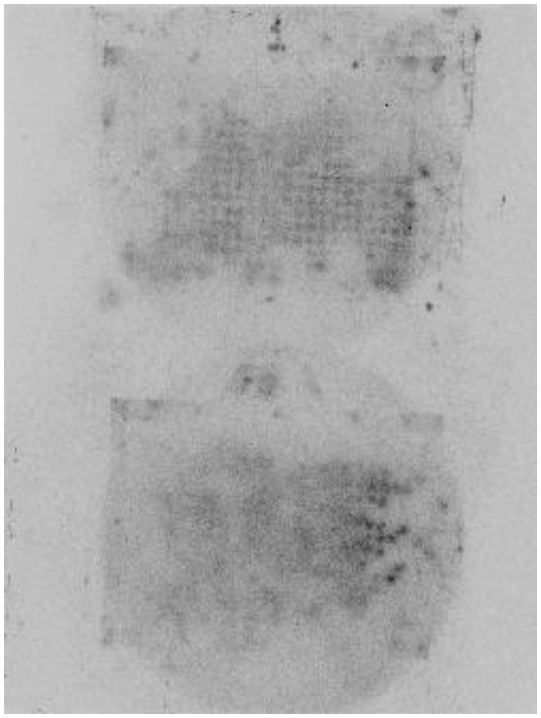

**Biofilm fosmid DNA from each plate before digestion**

**(#1-#11 shown in here total 13 plates)**

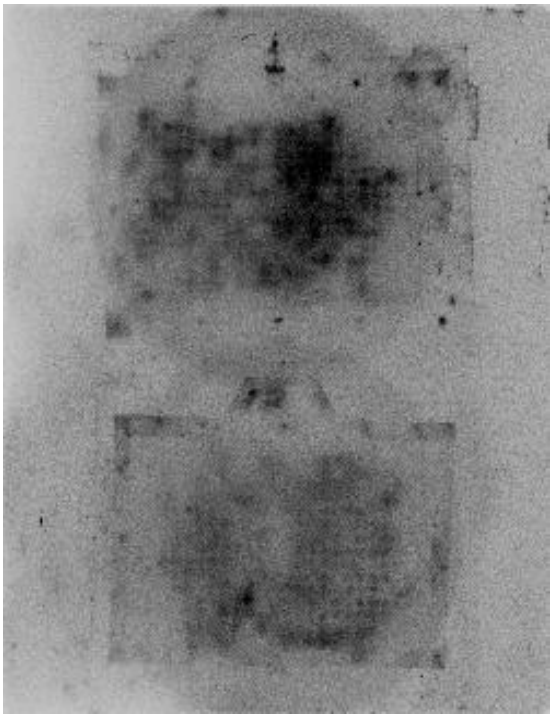


**Autoradiograph of Immobilized Biofilm Fosmid Library Probed With Radioactive PKS Fragment**





8hr developed image



overnight developed image

**Best hit from Biofilm fosmid library by screening with $^{32}$P labeled hybridization probe: #4E10**

**PCR Probe Primer Sequences:**

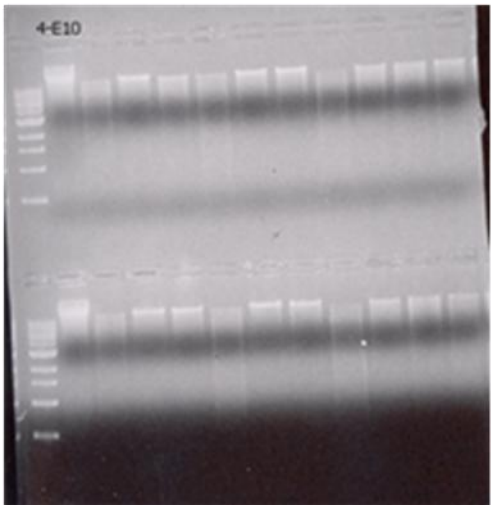
PKS and NRPS Primers

| | |
|---|---|
| PaltNRPS_for | GGYTTTCGYRTNGARYTRGGYGA |
| PaltNRPS_rev | TGRCCYARYKSAAAAAARTYNGC |
| PaltPKS_for | TWYVTNNNNRCNCAYGGYAC |
| PaltPKS rev | CCRCCAAARCCAAADSWRTT |

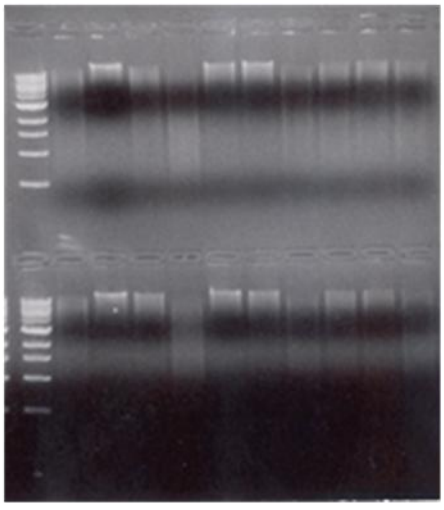
- Digestion of **#4E10** with series restriction enzyme with DMSO and w/o DMSO for 2 hrs, 4hrs, 9hrs, 24hrs (Shown in here 2hrs and 4hrs)

1. **EcoRI** 37°C
2. HpaI 37°C
3. PmlI 37°C
4. **SfoI** 37°C
5. BamHI 37°C
6. BglII 37°C
7. **EcoRV** 37°C
8. **KpnI** 37°C
9. NotI 37°C
10. PvuI 37°C

#4E10 After 2hrs digestion




#4E10 After 4hrs digestion



w/o DMSO

w/ DMSO

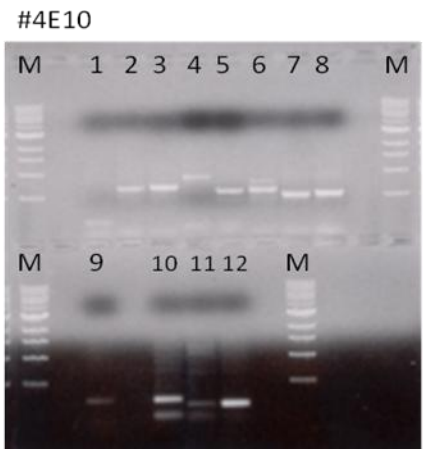
**Screening of Biotin biosynthetic pathway with best hit of Biofilm Fosmid library**

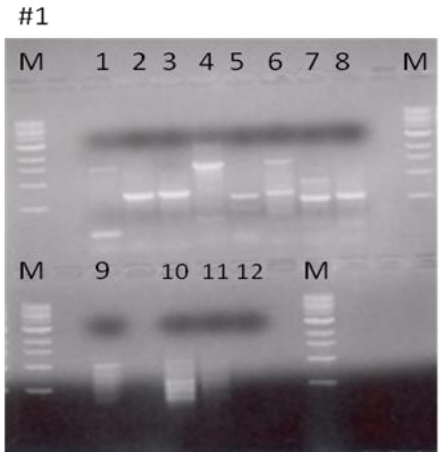
Used PCR conditions #9 : DMSO 2.5µl in total 50µl volume

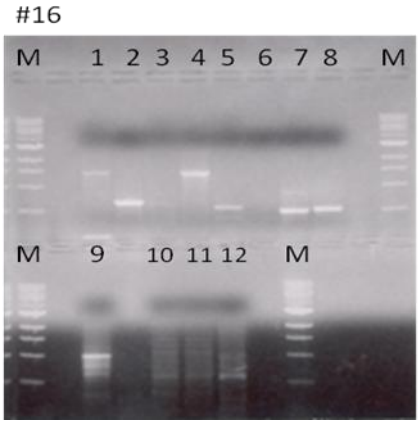
Selected PCR product: BioB: 1-8 (#3 1F-5R 678bp)

BioC: 9 (#9 6F-7R 324bp)

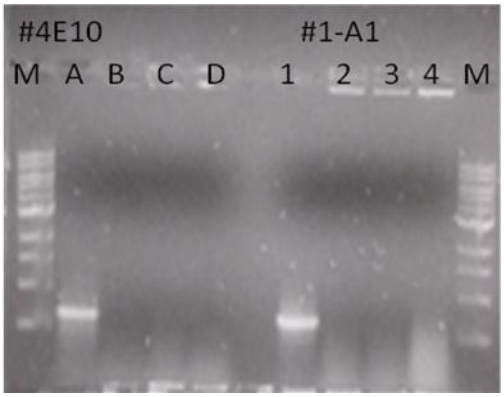
BioF: 10-12 (#12 10F-11R 225bp)




**Determination of optimal amount of DNA for PCR reaction**




A~D: #4E10 sample used
1~4: biofilm plate #1 (A1)

A,1: streaked bacteria clone once then microwaved, add 50µl of water, 1µl used for PCR (1/50)
B,2: streaked once
C,3: streaked twice
D,4: streaked 3times

**Screening of biotin biosynthetic pathway using Biofilm fosmid library**

BioB PCR Product with Biofilm Fosmid library (678bp)

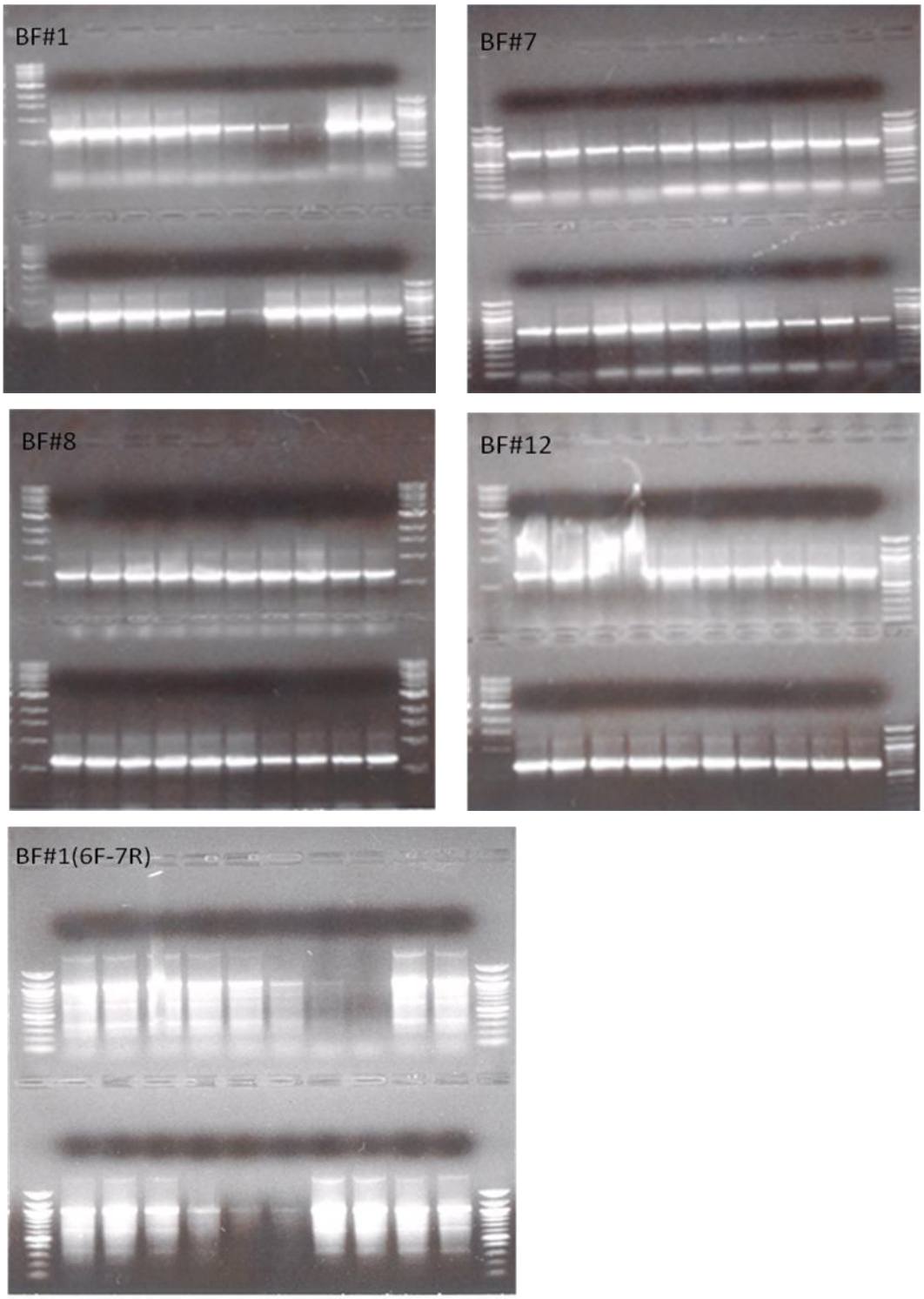
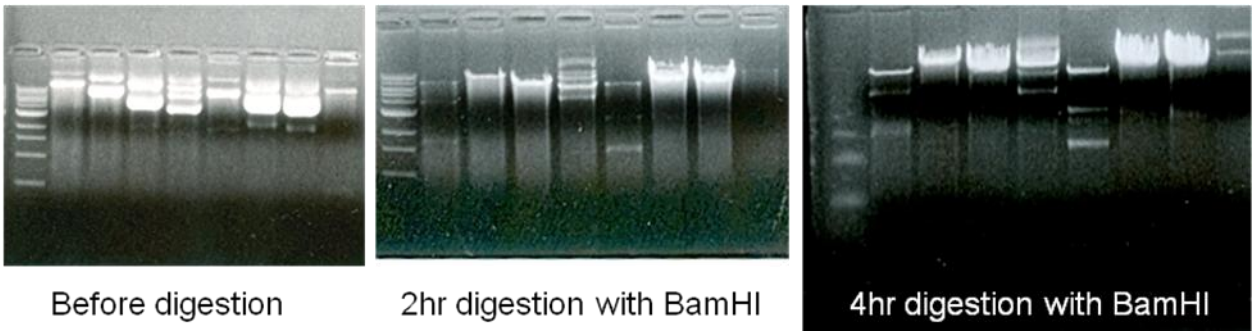


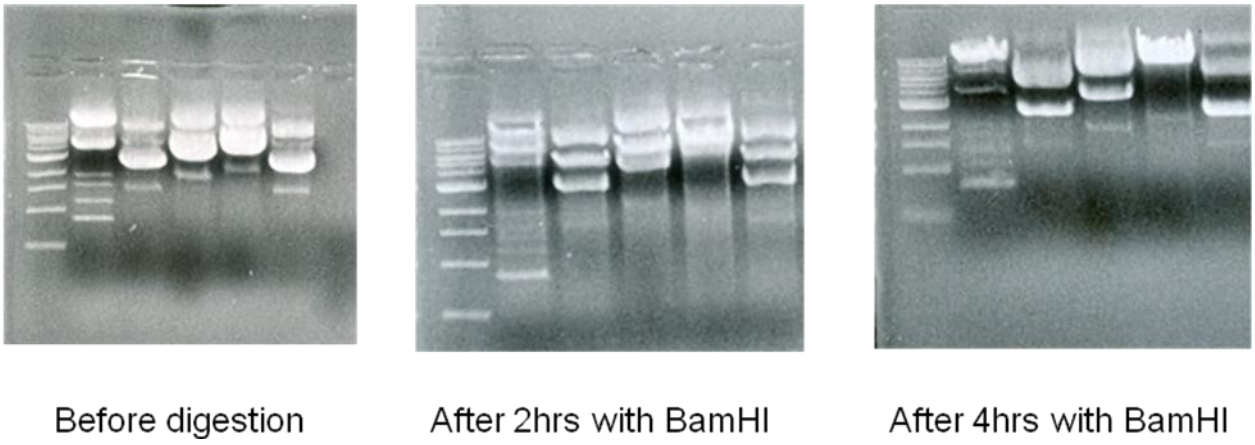

**Figure B.2: Supplementary data of Biofilm Fosmid library**

**Isolated plasmids from packaging and tittering method (transfection)**




**1st: Ladder**
**2nd: Big from dam-/dcm- (O.1)**
**3rd: Big from dam-/dcm- (O.1)**
**4th: Big from dam-/dcm- (O.5)**
**5th: Big from dam-/dcm- (O.5)**
**6th: Big from DH10B (0.1)**
**7th: Big from DH10B (0.1)**
**8th: Big from DH10B (0.5)**
**9th: Big from DH10B (0.5)**

**Electroporation into EPI300**



**CD151 isolated plasmids from each plate before digestion(using transformation method)**

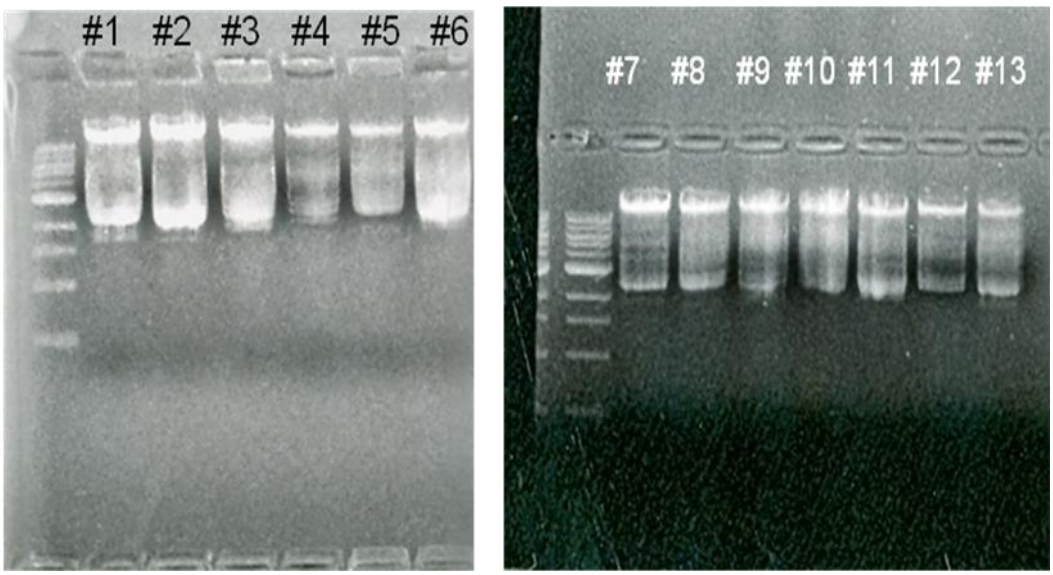



**Figure B.3: Supplementary data of CD151 fosmid library**

**CD179 before and after digestion (Regular gel)**

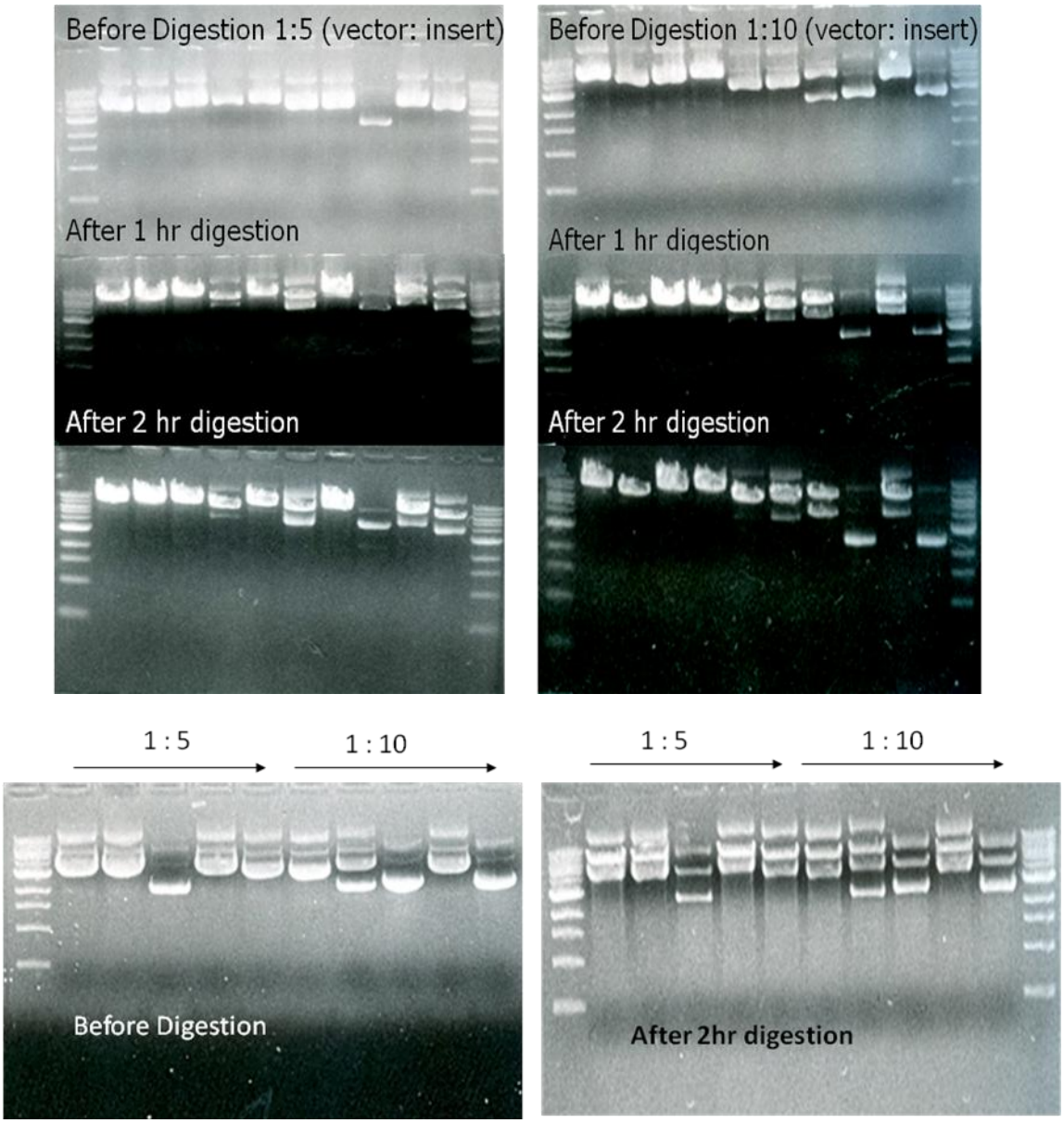
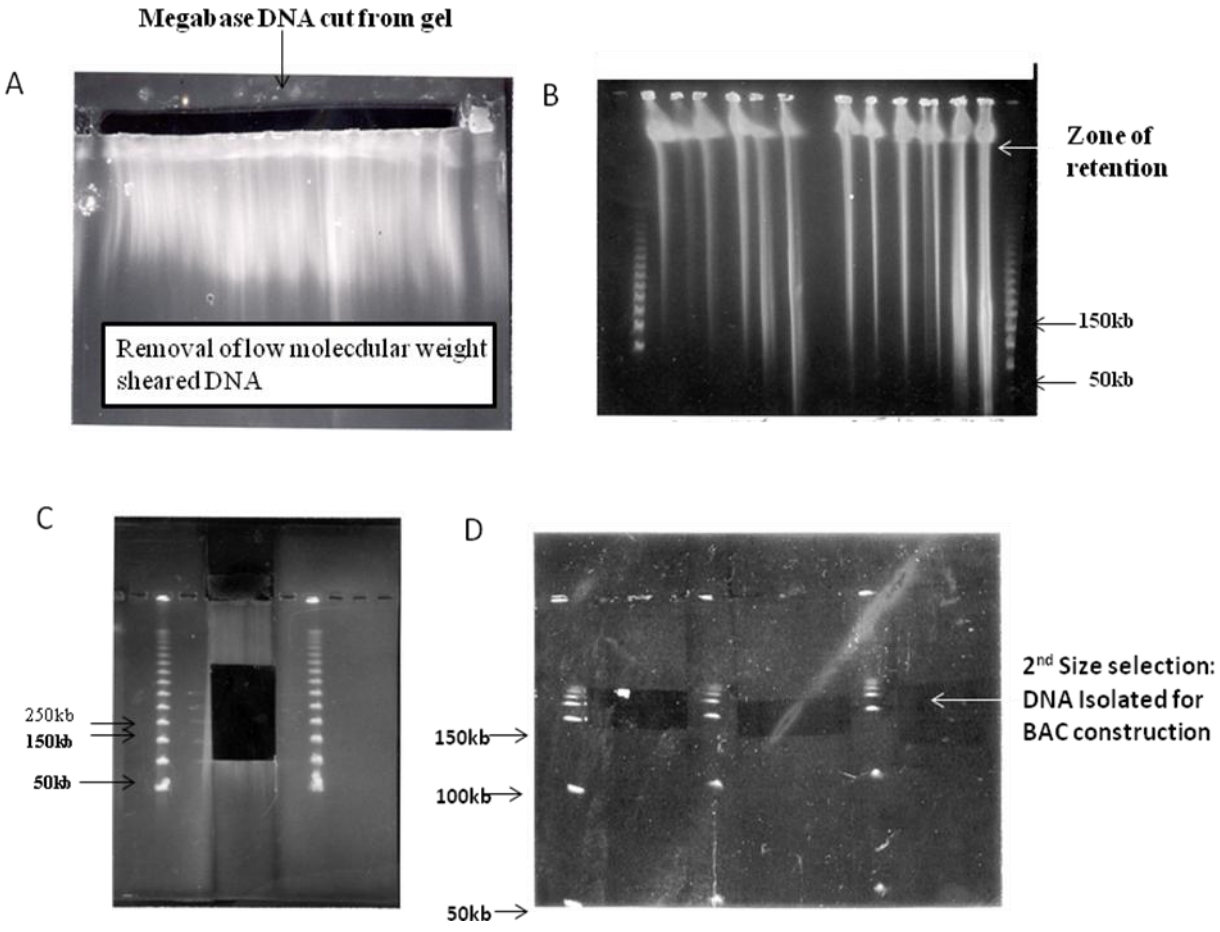



**Figure B.4: Supplementary data of CD179 fosmid library**

# APPENDIX C

Supplementary Data for Chapter IV




(A) High molecular weight genomic DNA isolated from Hon6. (B) Optimization of restriction enzyme digestion with *Hind* III (C)First size fraction with *Hind* III cut DNA (D) Second size selection to eliminate small fragments with altered CHEF gel conditions

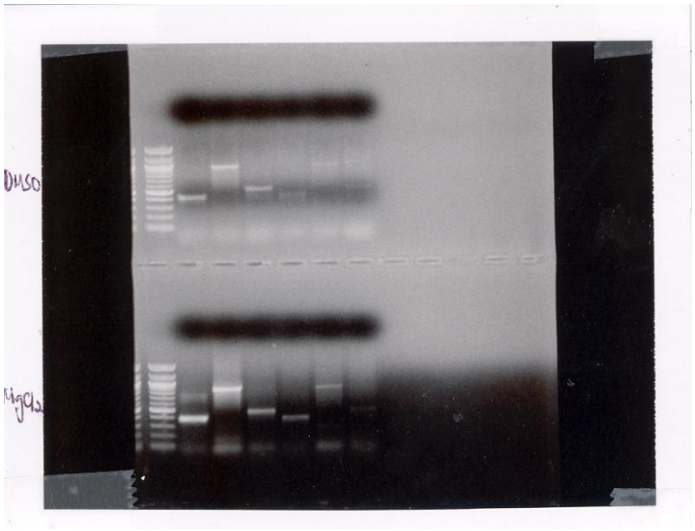
**Figure C.1: Representative gels on Hon6 library construction**

Biotin biosynthetic pathway degenerate probes were designed using multiple sequence alignments to search for regions of high homology. Protein sequences were downloaded from the NCBI website (http://www.ncbi.nlm.nih.gov.). Sequences (10) were chosen for for BioA, 17 sequences for BioB, 16 sequences for BioC, 16 sequences for BioD, 16 sequences for BioF, and 16 sequences for BioH. Sequences were chosen from multiple species and strains, which were identified by BLAST search analysis (http://ncbi.nlm.nih.gov/blst/Blst.cgi). Multiple sequence alignment was performed on the Baylor College of Medicine Multiple sequence alignment tool using Clustal X (http://searchlauncher.bcm.tmc.edu/multi-align/multi-align.html). Highly conserved sequenced were identified and manually reverse-translated into degenerate codons. Sequences were chosen which minimized degeneracy and maximized specificity. Oligonucleotides were ordered from the GTL (Texas A&M University, Gene Technological Lab). These oligonucleotides were used as PCR primers.

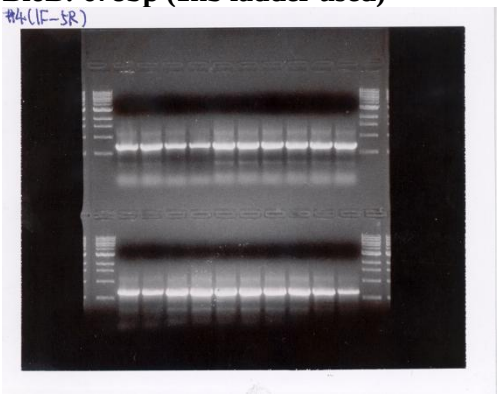
**Gel pictures and expected size of PCR products**

BioA: 1F-2R 420bp
BioB: 1F-5R 678bp
BioC: 6F-7R 324bp
BioD: 1F-3R 450bp
BioF: 10F-11R 225bp

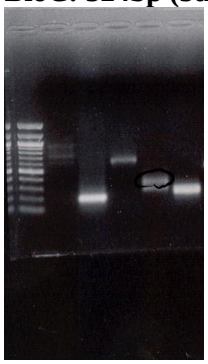
**BioA: 420bp (2nd lane)(100bp ladder used)**



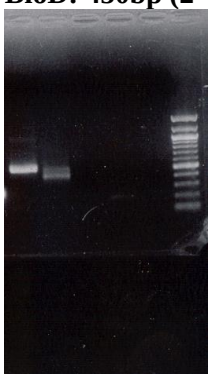
**BioB: 678bp (1kb ladder used)**




**BioC: 324bp (5th lane) (100bp ladder used)**



**BioD: 450bp (2nd lane)(100bp ladder used)**



**BioF:225bp (2nd lane)(100bp ladder used)**

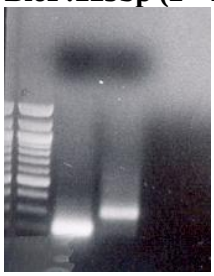
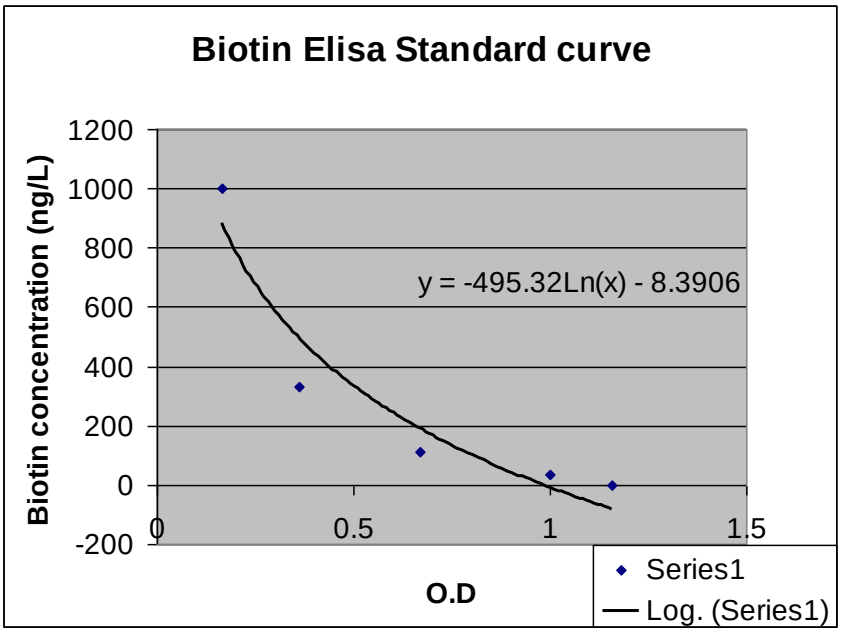


**Figure C.2: Description of Biotin Degenerate Probe Design**

| Standards and samples | Biotin Concentration(ng/L) | O.D. |
|---|---|---|
| **STD #1** | **0** | **1.158** |
| **STD #2** | **37** | **1.000** |
| **STD #3** | **111** | **0.668** |
| **STD #4** | **333** | **0.362** |
| **STD #5** | **1000** | **0.165** |
| **Hon6#2-1** | **753** | **0.215** |
| **Hon6#2-2** | **732** | **0.224** |
| **Hon6#2-3** | **667** | **0.256** |



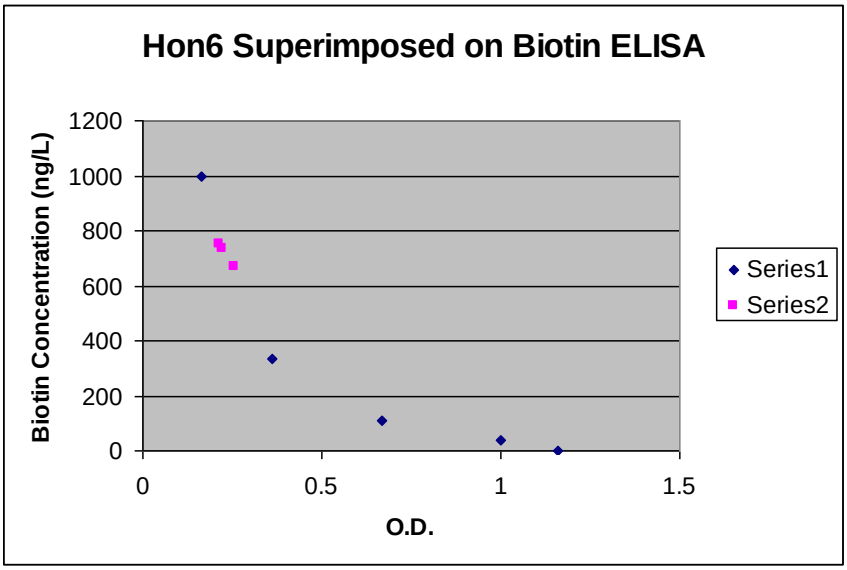




**Figure C.3: Biotin ELISA Tabulated Results**

NBioA1F: 5′-CAY ATH TGG CAY CCN TAY AC-3′
NBioA2R: 5′-CCR AAN GTR TCN CCR TGR TA-3′

CACAGTGTGTGTTTGCAGAGTATGTCGGTGCGAGAACCGAAGTGGCAGCGGACCTGTCTGACGGCAGACGC
CAGTGACGGTATGCGTCCTGGTGGGCAGAGATCCACGACTACAATCACCCGCAGCTTAATGCGTTTATGAA
GTCGTCATTGATGCCATGTCGCATGTGATGTTTGAAGATATCACGCATGCGCCAGCCAATGAGCTACTCCG
CAAACTGGTGGAGATGACGCCGAACCGCGCGGGTGTGCGTTTTTCCTACCAGCACCAGATACCGTAGCACC
AGGAAGTGGCCGTTGAAAAATAAAGCAGCTCGCTTGCTTAGGCCAAAGACGACGCGGATAAACCGTTTGAG
GACCTTCCCAAAAGATTTTATAACGGGTGCCACAATTTGG

| Sequences producing significant alignments: | Score (Bits) | E Value | |
|---|---|---|---|
| pdb\|1DTY\|A Chain A, Crystal Structure Of Adenosylmethionine-8... | 49.7 | 4e-13 | S |
| ref\|ZP_01702231.1\| adenosylmethionine-8-amino-7-oxononanoate ... | 48.5 | 8e-13 | |
| ref\|YP_668706.1\| adenosylmethionine-8-amino-7-oxononanoate am... | 48.5 | 8e-13 | G |
| ref\|YP_539789.1\| adenosylmethionine-8-amino-7-oxononanoate am... | 48.5 | 8e-13 | G |
| ref\|NP_752783.1\| adenosylmethionine--8-amino-7-oxononanoate t... | 48.5 | 8e-13 | G |
| ref\|YP_851870.1\| adenosylmethionine-8-amino-7-oxononanoate am... | 48.5 | 8e-13 | G |
| gb\|AAG60578.1\|AF250776_2 DAPA-aminotransferase BioA [uncultured | 48.5 | 8e-13 | |
| ref\|YP_407168.1\| 7,8-diaminopelargonic acid synthetase [Shige... | 48.5 | 8e-13 | G |
| ref\|NP_286537.1\| adenosylmethionine--8-amino-7-oxononanoate t... | 48.5 | 8e-13 | G |
| pdb\|1S08\|A Chain A, Crystal Structure Of The D147n Mutant Of ... | 48.5 | 8e-13 | S |
| pdb\|1S07\|A Chain A, Crystal Structure Of The R253a Mutant Of ... | 48.5 | 8e-13 | S |
| pdb\|1S09\|A Chain A, Crystal Structure Of The Y144f Mutant Of ... | 48.5 | 8e-13 | S |
| pdb\|1S06\|A Chain A, Crystal Structure Of The R253k Mutant Of ... | 48.5 | 8e-13 | S |
| pdb\|1S0A\|A Chain A, Crystal Structure Of The Y17f Mutant Of 7... | 48.5 | 8e-13 | S |
| ref\|NP_308879.1\| adenosylmethionine--8-amino-7-oxononanoate t... | 48.5 | 8e-13 | G |
| pdb\|1QJ3\|A Chain A, Crystal Structure Of 7,8-Diaminopelargoni... | 48.5 | 8e-13 | S |
| ref\|NP_415295.1\| 7,8-diaminopelargonic acid synthase, PLP-dep... | 48.5 | 8e-13 | G |
| pdb\|1MGV\|A Chain A, Crystal Structure Of The R391a Mutant Of ... | 48.5 | 8e-13 | S |
| ref\|ZP_00923501.1\| COG0161: Adenosylmethionine-8-amino-7-oxon... | 48.5 | 8e-13 | |
| ref\|ZP_00704996.1\| COG0161: Adenosylmethionine-8-amino-7-oxon... | 48.5 | 8e-13 | |
| ref\|ZP_00698063.1\| COG0161: Adenosylmethionine-8-amino-7-oxon... | 48.5 | 8e-13 | |
| ref\|ZP_00721164.1\| COG0161: Adenosylmethionine-8-amino-7-oxon... | 48.5 | 1e-12 | |
| ref\|NP_706652.1\| adenosylmethionine--8-amino-7-oxononanoate t... | 48.5 | 2e-12 | G |
| ref\|YP_402503.1\| 7,8-diaminopelargonic acid synthetase [Shige... | 48.5 | 2e-12 | G |
| ref\|NP_836430.1\| adenosylmethionine--8-amino-7-oxononanoate t... | 48.5 | 2e-12 | G |
| ref\|ZP_00709063.1\| COG0161: Adenosylmethionine-8-amino-7-oxon... | 48.5 | 2e-12 | |
| ref\|ZP_00703082.1\| COG0161: Adenosylmethionine-8-amino-7-oxon... | 47.4 | 2e-12 | |
| ref\|ZP_00715008.1\| COG0161: Adenosylmethionine-8-amino-7-oxon... | 47.4 | 7e-12 | |
| gb\|AAA23514.1\| 7,8-diamino-pelargonic acid aminotransferase | 47.4 | 9e-12 | |
| ref\|ZP_00720007.1\| COG0161: Adenosylmethionine-8-amino-7-oxon... | 47.0 | 1e-11 | |
| sp\|P53656\|BIOA_ESCVU Adenosylmethionine-8-amino-7-oxononanoat... | 45.1 | 8e-10 | |
| ref\|YP_151180.1\| adenosylmethionine--8-amino-7-oxononanoate t... | 43.1 | 8e-10 | G |
| ref\|NP_455335.1\| adenosylmethionine--8-amino-7-oxononanoate t... | 43.1 | 8e-10 | G |
| ref\|NP_459771.1\| adenosylmethionine--8-amino-7-oxononanoate t... | 43.1 | 8e-10 | G |
| ref\|YP_215778.1\| adenosylmethionine--8-amino-7-oxononanoate t... | 43.1 | 2e-09 | G |
| gb\|AAG53588.1\|AF248314_3 DAPA aminotransferase BioA [uncultured | 42.4 | 4e-09 | |
| gb\|AAG60558.1\| BioA-like protein [uncultured bacterium pCosFS1] | 41.6 | 5e-09 | |
| ref\|YP_001175997.1\| adenosylmethionine-8-amino-7-oxononanoate... | 41.2 | 8e-09 | G |
| gb\|AAG60563.1\|AF250770_1 DAPA-aminotransferase [uncultured bacte | 41.2 | 1e-08 | |

| | | | |
|---|---|---|---|
| ref\|ZP_01167088.1\| | adenosylmethionine--8-amino-7-oxononanoate... | 45.4 | 1e-08 |
| gb\|AAF04396.1\|AF191556_4 | BioA [Xenorhabdus nematophila] | 41.2 | 1e-08 |
| gb\|ABR76235.1\| | adenosylmethionine--8-amino-7-oxononanoate tra... | 40.8 | 3e-08 |
| ref\|ZP_00824685.1\| | COG0161: Adenosylmethionine-8-amino-7-oxon... | 42.7 | 5e-08 |
| ref\|ZP_01222495.1\| | adenosylmethionine--8-amino-7-oxononanoate... | 42.7 | 8e-08 |
| ref\|NP_777893.1\| | adenosylmethionine-8-amino-7-oxononanoate am... | 42.0 | 1e-07 |
| ref\|YP_130517.1\| | adenosylmethionine--8-amino-7-oxononanoate t... | 42.7 | 1e-07 |
| ref\|ZP_01065427.1\| | adenosylmethionine--8-amino-7-oxononanoate... | 41.2 | 1e-07 |
| ref\|YP_069716.1\| | adenosylmethionine--8-amino-7-oxononanoate t... | 40.8 | 1e-07 |
| ref\|ZP_01497422.1\| | hypothetical protein YpseI_02002586 [Yersi... | 40.8 | 1e-07 |
| ref\|NP_670332.1\| | adenosylmethionine--8-amino-7-oxononanoate t... | 40.8 | 1e-07 |
| ref\|NP_404759.1\| | adenosylmethionine--8-amino-7-oxononanoate t... | 40.8 | 1e-07 |
| ref\|ZP_01175286.1\| | COG0161: Adenosylmethionine-8-amino-7-oxon... | 40.8 | 1e-07 |
| ref\|ZP_00821421.1\| | COG0161: Adenosylmethionine-8-amino-7-oxon... | 40.8 | 2e-07 |
| ref\|YP_454582.1\| | adenosylmethionine-8-amino-7-oxononanoate am... | 40.4 | 2e-07 |
| ref\|ZP_00829478.1\| | COG0161: Adenosylmethionine-8-amino-7-oxon... | 39.7 | 2e-07 |
| ref\|ZP_01535731.1\| | adenosylmethionine-8-amino-7-oxononanoate ... | 41.6 | 2e-07 |
| ref\|YP_001131218.1\| | adenosylmethionine-8-amino-7-oxononanoate... | 40.8 | 2e-07 |
| ref\|YP_206706.1\| | adenosylmethionine--8-amino-7-oxononanoate t... | 42.7 | 2e-07 |
| ref\|ZP_00832231.1\| | COG0161: Adenosylmethionine-8-amino-7-oxon... | 39.7 | 2e-07 |
| ref\|YP_001007096.1\| | adenosylmethionine-8-amino-7-oxononanoate... | 38.9 | 3e-07 |
| ref\|ZP_00530485.1\| | Adenosylmethionine--8-amino-7-oxononanoate... | 39.3 | 5e-07 |
| ref\|ZP_00532661.1\| | Adenosylmethionine--8-amino-7-oxononanoate... | 39.3 | 5e-07 |
| ref\|ZP_01224606.1\| | putative adenosylmethionine-8-amino-7-oxon... | 42.0 | 5e-07 |
| ref\|ZP_01693278.1\| | adenosylmethionine-8-amino-7-oxononanoate ... | 42.0 | 7e-07 |
| ref\|NP_240116.1\| | adenosylmethionine-8-amino-7-oxononanoate am... | 39.3 | 7e-07 |
| ref\|ZP_00591569.1\| | Adenosylmethionine--8-amino-7-oxononanoate... | 38.9 | 7e-07 |
| ref\|YP_691976.1\| | 7,8-diaminononanoate transaminase [Alcanivor... | 40.4 | 7e-07 |
| ref\|NP_928779.1\| | adenosylmethionine--8-amino-7-oxononanoate t... | 38.5 | 7e-07 |
| ref\|ZP_01043920.1\| | Adenosylmethionine-8-amino-7-oxononanoate ... | 38.5 | 8e-07 |
| ref\|YP_375950.1\| | Adenosylmethionine--8-amino-7-oxononanoate a... | 41.2 | 8e-07 |
| ref\|YP_588695.1\| | adenosylmethionine-8-amino-7-oxononanoate am... | 38.5 | 1e-06 |
| ref\|ZP_00588811.1\| | Adenosylmethionine--8-amino-7-oxononanoate... | 40.8 | 1e-06 |
| ref\|ZP_01449474.1\| | adenosylmethionine--8-amino-7-oxononanoate... | 37.7 | 1e-06 |
| ref\|ZP_01613057.1\| | adenosylmethionine-8-amino-7-oxononanoate ... | 37.7 | 1e-06 |
| gb\|EDN14554.1\| | adenosylmethionine-8-amino-7-oxononanoate amin... | 39.3 | 1e-06 |
| gb\|EDM55468.1\| | adenosylmethionine-8-amino-7-oxononanoate amin... | 39.3 | 1e-06 |
| gb\|EAY41069.1\| | adenosylmethionine-8-amino-7-oxononanoate amin... | 39.3 | 1e-06 |
| ref\|ZP_01679552.1\| | adenosylmethionine-8-amino-7-oxononanoate ... | 39.3 | 1e-06 |
| ref\|NP_230756.1\| | adenosylmethionine--8-amino-7-oxononanoate t... | 39.3 | 1e-06 |
| ref\|YP_662128.1\| | adenosylmethionine-8-amino-7-oxononanoate am... | 40.0 | 1e-06 |
| sp\|P36568\|BIOA_SERMA | Adenosylmethionine-8-amino-7-oxononanoat... | 38.5 | 1e-06 |
| gb\|EAZ75883.1\| | adenosylmethionine-8-amino-7-oxononanoate amin... | 39.3 | 1e-06 |
| ref\|YP_340126.1\| | adenosylmethionine-8-amino-7-oxononanoate am... | 39.3 | 2e-06 |
| ref\|YP_910611.1\| | adenosylmethionine-8-amino-7-oxononanoate am... | 38.9 | 2e-06 |
| gb\|EAY34686.1\| | adenosylmethionine-8-amino-7-oxononanoate amin... | 38.9 | 2e-06 |
| ref\|YP_001303271.1\| | adenosylmethionine--8-amino-7-oxononanoat... | 39.3 | 2e-06 |
| ref\|ZP_00992911.1\| | adenosylmethionine--8-amino-7-oxononanoate... | 37.4 | 2e-06 |
| ref\|ZP_01484255.1\| | hypothetical protein VchoV5_02003173 [Vibrio | 38.9 | 2e-06 |
| ref\|YP_050917.1\| | adenosylmethionine--8-amino-7-oxononanoate t... | 40.0 | 2e-06 |
| ref\|ZP_01787556.1\| | adenosylmethionine--8-amino-7-oxononanoate... | 39.7 | 2e-06 |
| ref\|ZP_01785523.1\| | adenosylmethionine--8-amino-7-oxononanoate... | 39.7 | 2e-06 |
| ref\|YP_249043.1\| | adenosylmethionine--8-amino-7-oxononanoate t... | 39.7 | 2e-06 |
| ref\|NP_439703.1\| | adenosylmethionine--8-amino-7-oxononanoate t... | 39.7 | 2e-06 |
| ref\|ZP_00155127.2\| | COG0161: Adenosylmethionine-8-amino-7-oxon... | 39.7 | 2e-06 |
| ref\|ZP_00157150.1\| | COG0161: Adenosylmethionine-8-amino-7-oxon... | 39.7 | 2e-06 |

| | | | |
|---|---|---|---|
| ref\|YP_001291543.1\| | adenosylmethionine--8-amino-7-oxononanoat... | 39.7 | 2e-06 |
| ref\|YP_001290829.1\| | adenosylmethionine--8-amino-7-oxononanoat... | 39.7 | 2e-06 |
| ref\|ZP_01797619.1\| | adenosylmethionine--8-amino-7-oxononanoate... | 39.7 | 2e-06 |
| ref\|ZP_01794704.1\| | adenosylmethionine--8-amino-7-oxononanoate... | 39.7 | 2e-06 |
| ref\|ZP_01793207.1\| | adenosylmethionine--8-amino-7-oxononanoate... | 39.7 | 2e-06 |

TCTTCCGGCGCCGTTTGGGACCCGATGTGCGTGCGAGTTCGGTTGCGGTACTGAACTCTGTTTTATCGCGA
CTTCCACCGCTACGGTAACCGGACCCGCGAGAAAAACGCACTCCAGCACAAGCACCGTCAGCTTAATGCGT
TTGCGGCGTAGTCATTGACTCCATCATGGGTGATACCTTCGAACATCACATCCAACATGCAATCAATTTTG
CTACTTCATCGACTCATTAAGCTGAGGGTGATTGTAGCCGTGGATCGTTTCCTACCAGGACGACATACCGT
CAACCAGGCGTCTGCGGTCAGAAAAAATCAGCTCGCAACCTTCGGCGCTCACCACCGGATAAACCGGCAGA
GGGCAGGTCATGGATGTATACGGGTGCCAATTTTG

| Sequences producing significant alignments: | | Score (Bits) | E Value |
|---|---|---|---|
| pdb\|1DTY\|A | Chain A, Crystal Structure Of Adenosylmethionine-8... | 123 | 4e-27 |
| ref\|NP_415295.1\| | 7,8-diaminopelargonic acid synthase, PLP-dep... | 122 | 5e-27 |
| ref\|ZP_00923501.1\| | COG0161: Adenosylmethionine-8-amino-7-oxon... | 122 | 5e-27 |
| ref\|YP_407168.1\| | 7,8-diaminopelargonic acid synthetase [Shige... | 122 | 7e-27 |
| ref\|ZP_00698063.1\| | COG0161: Adenosylmethionine-8-amino-7-oxon... | 122 | 7e-27 |
| ref\|ZP_01702231.1\| | adenosylmethionine-8-amino-7-oxononanoate ... | 121 | 2e-26 |
| ref\|YP_851870.1\| | adenosylmethionine-8-amino-7-oxononanoate am... | 121 | 2e-26 |
| ref\|YP_668706.1\| | adenosylmethionine-8-amino-7-oxononanoate am... | 121 | 2e-26 |
| ref\|YP_539789.1\| | adenosylmethionine-8-amino-7-oxononanoate am... | 121 | 2e-26 |
| gb\|AAG60578.1\|AF250776_2 | DAPA-aminotransferase BioA [uncultured | 121 | 2e-26 |
| ref\|NP_706652.1\| | adenosylmethionine--8-amino-7-oxononanoate t... | 121 | 2e-26 |
| ref\|YP_402503.1\| | 7,8-diaminopelargonic acid synthetase [Shige... | 121 | 2e-26 |
| ref\|NP_836430.1\| | adenosylmethionine--8-amino-7-oxononanoate t... | 121 | 2e-26 |
| ref\|NP_286537.1\| | adenosylmethionine--8-amino-7-oxononanoate t... | 121 | 2e-26 |
| ref\|NP_752783.1\| | adenosylmethionine--8-amino-7-oxononanoate t... | 121 | 2e-26 |
| ref\|NP_308879.1\| | adenosylmethionine--8-amino-7-oxononanoate t... | 121 | 2e-26 |
| ref\|ZP_00704996.1\| | COG0161: Adenosylmethionine-8-amino-7-oxon... | 121 | 2e-26 |
| ref\|ZP_00703082.1\| | COG0161: Adenosylmethionine-8-amino-7-oxon... | 121 | 2e-26 |
| ref\|ZP_00709063.1\| | COG0161: Adenosylmethionine-8-amino-7-oxon... | 121 | 2e-26 |
| ref\|ZP_00721164.1\| | COG0161: Adenosylmethionine-8-amino-7-oxon... | 120 | 3e-26 |
| ref\|ZP_00720007.1\| | COG0161: Adenosylmethionine-8-amino-7-oxon... | 119 | 7e-26 |
| pdb\|1S08\|A | Chain A, Crystal Structure Of The D147n Mutant Of ... | 118 | 1e-25 |
| pdb\|1S07\|A | Chain A, Crystal Structure Of The R253a Mutant Of ... | 118 | 1e-25 |
| pdb\|1S09\|A | Chain A, Crystal Structure Of The Y144f Mutant Of ... | 118 | 1e-25 |
| pdb\|1S06\|A | Chain A, Crystal Structure Of The R253k Mutant Of ... | 118 | 1e-25 |
| pdb\|1QJ3\|A | Chain A, Crystal Structure Of 7,8-Diaminopelargoni... | 118 | 1e-25 |
| pdb\|1MGV\|A | Chain A, Crystal Structure Of The R391a Mutant Of ... | 118 | 1e-25 |
| ref\|ZP_00715008.1\| | COG0161: Adenosylmethionine-8-amino-7-oxon... | 118 | 1e-25 |
| gb\|AAA23514.1\| | 7,8-diamino-pelargonic acid aminotransferase | 117 | 3e-25 |
| pdb\|1S0A\|A | Chain A, Crystal Structure Of The Y17f Mutant Of 7... | 117 | 3e-25 |
| ref\|YP_454582.1\| | adenosylmethionine-8-amino-7-oxononanoate am... | 115 | 6e-25 |
| ref\|NP_459771.1\| | adenosylmethionine--8-amino-7-oxononanoate t... | 111 | 2e-23 |
| ref\|YP_215778.1\| | adenosylmethionine--8-amino-7-oxononanoate t... | 110 | 3e-23 |
| ref\|YP_151180.1\| | adenosylmethionine--8-amino-7-oxononanoate t... | 110 | 3e-23 |

| | | | |
|---|---|---|---|
| ref\|NP_455335.1\| | adenosylmethionine--8-amino-7-oxononanoate t... | 110 | 3e-23 |
| gb\|ABR76235.1\| | adenosylmethionine--8-amino-7-oxononanoate tra... | 108 | 1e-22 |
| sp\|P53656\|BIOA_ESCVU | Adenosylmethionine-8-amino-7-oxononanoat... | 108 | 1e-22 |
| gb\|AAG53588.1\|AF248314_3 | DAPA aminotransferase BioA [uncultured | 106 | 4e-22 |
| ref\|NP_928779.1\| | adenosylmethionine--8-amino-7-oxononanoate t... | 106 | 4e-22 |
| ref\|ZP_00829478.1\| | COG0161: Adenosylmethionine-8-amino-7-oxon... | 106 | 4e-22 |
| ref\|YP_001175997.1\| | adenosylmethionine-8-amino-7-oxononanoate... | 106 | 5e-22 |
| ref\|ZP_00824685.1\| | COG0161: Adenosylmethionine-8-amino-7-oxon... | 106 | 5e-22 |
| ref\|ZP_00821421.1\| | COG0161: Adenosylmethionine-8-amino-7-oxon... | 105 | 7e-22 |
| ref\|YP_001007096.1\| | adenosylmethionine-8-amino-7-oxononanoate... | 104 | 1e-21 |
| gb\|AAF04396.1\|AF191556_4 | BioA [Xenorhabdus nematophila] | 104 | 1e-21 |
| ref\|ZP_00832231.1\| | COG0161: Adenosylmethionine-8-amino-7-oxon... | 104 | 2e-21 |
| ref\|YP_588695.1\| | adenosylmethionine-8-amino-7-oxononanoate am... | 103 | 3e-21 |
| ref\|YP_050917.1\| | adenosylmethionine--8-amino-7-oxononanoate t... | 100 | 2e-20 |
| ref\|YP_069716.1\| | adenosylmethionine--8-amino-7-oxononanoate t... | 100 | 3e-20 |
| ref\|NP_404759.1\| | adenosylmethionine--8-amino-7-oxononanoate t... | 100 | 3e-20 |
| ref\|ZP_01175286.1\| | COG0161: Adenosylmethionine-8-amino-7-oxon... | 100 | 3e-20 |
| ref\|ZP_01497422.1\| | hypothetical protein YpseI_02002586 [Yersi... | 100 | 3e-20 |
| ref\|NP_670332.1\| | adenosylmethionine--8-amino-7-oxononanoate t... | 100 | 3e-20 |
| gb\|AAG60558.1\| | BioA-like protein [uncultured bacterium pCosFS1] | 99.0 | 8e-20 |
| ref\|ZP_01535731.1\| | adenosylmethionine-8-amino-7-oxononanoate ... | 97.8 | 2e-19 |
| sp\|P36568\|BIOA_SERMA | Adenosylmethionine-8-amino-7-oxononanoat... | 97.1 | 3e-19 |
| ref\|YP_691976.1\| | 7,8-diaminononanoate transaminase [Alcanivor... | 95.1 | 1e-18 |
| gb\|AAG60563.1\|AF250770_1 | DAPA-aminotransferase [uncultured bacte | 94.4 | 2e-18 |
| ref\|NP_660627.1\| | adenosylmethionine-8-amino-7-oxononanoate am... | 94.4 | 2e-18 |
| ref\|ZP_01222495.1\| | adenosylmethionine--8-amino-7-oxononanoate... | 94.0 | 3e-18 |
| ref\|YP_130517.1\| | adenosylmethionine--8-amino-7-oxononanoate t... | 94.0 | 3e-18 |
| ref\|ZP_01898736.1\| | adenosylmethionine-8-amino-7-oxononanoate ... | 92.8 | 6e-18 |
| ref\|ZP_01160450.1\| | adenosylmethionine--8-amino-7-oxononanoate... | 92.8 | 6e-18 |
| ref\|ZP_00530485.1\| | Adenosylmethionine--8-amino-7-oxononanoate... | 92.8 | 6e-18 |
| ref\|ZP_01043920.1\| | Adenosylmethionine-8-amino-7-oxononanoate ... | 92.4 | 8e-18 |
| ref\|ZP_01235384.1\| | adenosylmethionine--8-amino-7-oxononanoate... | 92.0 | 1e-17 |
| ref\|NP_240116.1\| | adenosylmethionine-8-amino-7-oxononanoate am... | 92.0 | 1e-17 |
| ref\|ZP_00532661.1\| | Adenosylmethionine--8-amino-7-oxononanoate... | 92.0 | 1e-17 |
| ref\|YP_001093844.1\| | adenosylmethionine-8-amino-7-oxononanoate... | 90.5 | 3e-17 |
| ref\|NP_871455.1\| | hypothetical protein WGLp452 [Wigglesworthia... | 89.7 | 5e-17 |
| ref\|ZP_01693278.1\| | adenosylmethionine-8-amino-7-oxononanoate ... | 89.4 | 6e-17 |
| ref\|NP_874062.1\| | adenosylmethionine--8-amino-7-oxononanoate t... | 89.4 | 6e-17 |
| ref\|YP_001131218.1\| | adenosylmethionine-8-amino-7-oxononanoate... | 89.4 | 6e-17 |
| ref\|YP_001303271.1\| | adenosylmethionine--8-amino-7-oxononanoat... | 89.0 | 8e-17 |
| ref\|YP_378405.1\| | Adenosylmethionine--8-amino-7-oxononanoate a... | 88.6 | 1e-16 |
| ref\|YP_910611.1\| | adenosylmethionine-8-amino-7-oxononanoate am... | 88.6 | 1e-16 |
| ref\|ZP_01224606.1\| | putative adenosylmethionine-8-amino-7-oxon... | 87.8 | 2e-16 |
| ref\|ZP_01167088.1\| | adenosylmethionine--8-amino-7-oxononanoate... | 87.8 | 2e-16 |
| ref\|ZP_00591569.1\| | Adenosylmethionine--8-amino-7-oxononanoate... | 87.8 | 2e-16 |
| ref\|ZP_00511607.1\| | Adenosylmethionine--8-amino-7-oxononanoate... | 87.0 | 3e-16 |
| ref\|ZP_01674408.1\| | adenosylmethionine-8-amino-7-oxononanoate ... | 86.7 | 4e-16 |
| ref\|ZP_01473993.1\| | hypothetical protein VEx2w_02003422 [Vibri... | 86.3 | 5e-16 |
| ref\|ZP_01385590.1\| | adenosylmethionine-8-amino-7-oxononanoate ... | 85.9 | 7e-16 |
| ref\|YP_375950.1\| | Adenosylmethionine--8-amino-7-oxononanoate a... | 85.9 | 7e-16 |
| ref\|YP_856026.1\| | adenosylmethionine-8-amino-7-oxononanoate tr... | 85.5 | 9e-16 |
| ref\|ZP_01921353.1\| | adenosylmethionine-8-amino-7-oxononanoate ... | 85.1 | 1e-15 |
| ref\|YP_206706.1\| | adenosylmethionine--8-amino-7-oxononanoate t... | 85.1 | 1e-15 |
| ref\|NP_797491.1\| | adenosylmethionine--8-amino-7-oxononanoate t... | 85.1 | 1e-15 |
| ref\|ZP_01065427.1\| | adenosylmethionine--8-amino-7-oxononanoate... | 85.1 | 1e-15 |

| | | | |
|---|---|---|---|
| ref\|ZP_01816621.1\| | adenosylmethionine--8-amino-7-oxononanoate... | 84.7 | 2e-15 |
| ref\|ZP_00588811.1\| | Adenosylmethionine--8-amino-7-oxononanoate... | 84.3 | 2e-15 |
| ref\|YP_750164.1\| | adenosylmethionine-8-amino-7-oxononanoate am... | 84.3 | 2e-15 |
| ref\|ZP_01133854.1\| | adenosylmethionine-8-amino-7-oxononanoate ... | 84.0 | 3e-15 |
| ref\|YP_870155.1\| | adenosylmethionine-8-amino-7-oxononanoate am... | 84.0 | 3e-15 |
| ref\|XP_718351.1\| | adenosylmethionine-8-amino-7-oxononanoate am... | 84.0 | 3e-15 |
| ref\|ZP_01597382.1\| | adenosylmethionine-8-amino-7-oxononanoate ... | 83.6 | 3e-15 |
| ref\|YP_943296.1\| | adenosylmethionine-8-amino-7-oxononanoate am... | 83.6 | 3e-15 |
| ref\|ZP_01260980.1\| | adenosylmethionine--8-amino-7-oxononanoate... | 83.6 | 3e-15 |
| ref\|YP_562664.1\| | adenosylmethionine-8-amino-7-oxononanoate am... | 83.6 | 3e-15 |
| ref\|ZP_00992911.1\| | adenosylmethionine--8-amino-7-oxononanoate... | 83.6 | 3e-15 |

>> gb|CP000038.1| Shigella sonnei Ss046, complete genome
Length=4825265

Features in this part of subject sequence:
7,8-diaminopelargonic acid synthetase

Score = 230 bits (124), Expect = 4e-57
Identities = 184/211 (87%), Gaps = 11/211 (5%)
Strand=Plus/Minus

```
Query  55      TGTCTGACGGCAGACGCC-AG-TGACGGTATG-CGTCCTGGTGGGCAGAGATCCACGACT  111
               ||||||||||||||||||  | |||||||||| |||||||||||| | |||||||| ||
Sbjct  798228  TGTCTGACGGCAGACGCCTGGTTGACGGTATGTCGTCCTGGTGGGCGGCGATCCACGGCT  798169

Query  112     ACAATCACCCGCAGCTTAATGCGTTTATGAAGTCGTC--ATTGATGCCATGTCGCATGTG  169
               |||||||||||||||||||||   ||||||||| |  ||||||||||||||||||||
Sbjct  798168  ACAATCACCCGCAGCTTAATGCGGCGATGAAGTCG-CAAATTGATGCCATGTCGCATGTG  798110

Query  170     ATGTTTGAAGATATCACGCATGCGCCAGCCAATGAGCTACT-CCGCAAACTGGTGGAGAT  228
               |||||||  | |||||| |||||||||||||| ||||||  | ||||||||||||| |||
Sbjct  798109  ATGTTTGGCGGTATCACCCATGCGCCAGCCATTGAGCTG-TGCCGCAAACTGGTGGCGAT  798051

Query  229     GACGCCG-AACCGCGCGGGTGTGCGTTTTTC  258
               ||||||| ||||||  ||  ||||||||||
Sbjct  798050  GACGCCGCAACCGCT-GGA-GTGCGTTTTTC  798022
```

> gb|CP000036.1| Shigella boydii Sb227, complete genome
Length=4519823

Features in this part of subject sequence:
7,8-diaminopelargonic acid synthetase

Score = 219 bits (118), Expect = 9e-54
Identities = 169/192 (88%), Gaps = 9/192 (4%)

```
 Strand=Plus/Minus

Query  55      TGTCTGACGGCAGACGCC--AGTGACGGTATG-CGTCCTGGTGGGCAGAGATCCACGACT  111
               ||||||||||||||||||  | |||||||||| |||||||||||| | |||||||| ||
Sbjct  675164  TGTCTGACGGCAGACGCCTGATTGACGGTATGTCGTCCTGGTGGGCGGCGATCCACGGCT  675105

Query  112     ACAATCACCCGCAGCTTAATGCGTTTATGAAGTCGTC--ATTGATGCCATGTCGCATGTG  169
               |||||||||||||||||||||||   ||||||||| |  ||||||||||||||||||||
Sbjct  675104  ACAATCACCCGCAGCTTAATGCGGCGATGAAGTCG-CAAATTGATGCCATGTCGCATGTG  675046

Query  170     ATGTTTGAAGATATCACGCATGCGCCAGCCAATGAGCTACT-CCGCAAACTGGTGGAGAT  228
               |||||||  | |||||| |||||||||||||| ||||||  | |||||||||||||| |||
Sbjct  675045  ATGTTTGGCGGTATCACCCATGCGCCAGCCATTGAGCTG-TGCCGCAAACTGGTGGCGAT  674987

Query  229     GACGCCG-AACC  239
               ||||||| ||||
Sbjct  674986  GACGCCGCAACC  674975
```

**BioB**

BioB1F: 5′-GAR GAY TGY AAR TAY TGY CC-3′
BioB5R: 5′-CCN GCC ATR AAR CAC ATN GCY TG-3′

ATCTCGGATGAACATACGTCTGCCGCCAGGAGGGTGACGTAAGTGTTAGATGATGATGCCCGCCAAGTGGT
GCGATAAAATCAAAGGCATCGACATCATCGTTATCGGCAAGCGGCGTGCTTTCACCTTCACCAGCATGTTG
ATTGGCACGCTTTCCGGCGCGCGGAGGGTTGGAGGTGAGCAATAATCCGGCACGATCTTTTACCGTTACGC
CTAAGCCCACAATGCCGCCAGAACAGACTTTGATCCCGGCATCGTGCACTTTTTCCAGCGTATCGAGGCGT
TCCTGATAAGCGCATGTCTTTGATGATATTGCCGTAAAACTCCGGCAAGAGGTCCAGGTTGTGGTTGAAGT
AATCCAGCTGGGCGTTCGGGAGGCGTTGGGAAAAAGATAAACTCAACGTGCCCAGCATTATACACGCTTCC
AGCCCCATCGCTGCTGCCGGAAGCGCTACTTATTCAAGGTACGGCGTATGGGAAAGGTACCCCGCTTTGCC
AATCACCGTGCATACAGTAGCTCGGAGTTTATGCGGATAATTGTTGGGGCGGGGACTCTAGTACCAGCTCA
CTTACGATCAATTTCTCGCGGACCAGCGCGGTTTTACCGAACAGACTTAGGCATTATGCTTAGGGACTCCT
TAAA

| Sequences producing significant alignments: | Score (Bits) | E Value | |
|---|---|---|---|
| pdb\|1R30\|A Chain A, The Crystal Structure Of Biotin Synthase,... | 94.4 | 7e-37 | S |
| gb\|AAA23515.1\| biotin synthetase | 94.4 | 7e-37 | |
| ref\|YP_402502.1\| biotin synthesis [Shigella dysenteriae Sd197... | 94.4 | 7e-37 | G |
| ref\|NP_706653.1\| biotin synthesis, sulfur insertion? [Shigell... | 94.4 | 7e-37 | G |
| ref\|NP_415296.1\| biotin synthase [Escherichia coli K12] >ref\|... | 94.4 | 7e-37 | G |
| ref\|NP_752785.1\| Biotin synthase [Escherichia coli CFT073] >r... | 92.8 | 2e-36 | G |
| ref\|ZP_00721165.1\| COG0502: Biotin synthase and related enzym... | 92.8 | 2e-36 | |
| ref\|NP_308880.1\| biotin synthetase [Escherichia coli O157:H7 ... | 94.4 | 4e-36 | G |
| gb\|AAG60579.1\|AF250776_3 biotin synthase BioB [uncultured bacter | 94.4 | 7e-36 | |
| ref\|ZP_00698064.1\| COG0502: Biotin synthase and related enzym... | 94.4 | 7e-36 | |
| ref\|NP_286538.1\| biotin synthesis, sulfur insertion? [Escheri... | 94.4 | 2e-35 | G |
| sp\|Q47862\|BIOB_ESCVU Biotin synthase (Biotin synthetase) >gb\|... | 93.2 | 7e-35 | |
| gb\|ABR76236.1\| biotin synthesis, sulfur insertion? [Klebsiell... | 91.7 | 1e-34 | |
| ref\|YP_215779.1\| biotin synthetase [Salmonella enterica subsp... | 90.5 | 4e-34 | G |

| | | | |
|---|---|---|---|
| ref\|YP_151179.1\| | biotin synthetase [Salmonella enterica subsp... | 90.5 | 4e-34 |
| ref\|NP_459772.1\| | biotin synthetase [Salmonella typhimurium LT... | 90.5 | 4e-34 |
| ref\|NP_455336.1\| | biotin synthetase [Salmonella enterica subsp... | 90.5 | 4e-34 |
| ref\|YP_001175998.1\| | biotin synthase [Enterobacter sp. 638] >g... | 90.1 | 2e-33 |
| gb\|AAG53589.1\|AF248314_4 | biotin synthase BioB [uncultured bacter | 82.0 | 7e-32 |
| sp\|P36569\|BIOB_SERMA | Biotin synthase (Biotin synthetase) >dbj... | 90.5 | 1e-31 |
| ref\|YP_001007095.1\| | biotin synthase [Yersinia enterocolitica ... | 84.7 | 7e-31 |
| ref\|ZP_00824684.1\| | COG0502: Biotin synthase and related enzym... | 84.7 | 7e-31 |
| ref\|ZP_01535732.1\| | biotin synthase [Serratia proteamaculans 5... | 84.3 | 2e-30 |
| ref\|ZP_00829477.1\| | COG0502: Biotin synthase and related enzym... | 83.2 | 3e-30 |
| gb\|AAG60559.1\| | BioB-like protein [uncultured bacterium pCosFS1] | 82.0 | 6e-30 |
| ref\|ZP_00821420.1\| | COG0502: Biotin synthase and related enzym... | 83.6 | 7e-30 |
| ref\|ZP_00832232.1\| | COG0502: Biotin synthase and related enzym... | 84.7 | 1e-29 |
| ref\|YP_050916.1\| | biotin synthase [Erwinia carotovora subsp. a... | 85.5 | 2e-29 |
| ref\|NP_928780.1\| | biotin synthetase (biotin synthase) [Photorh... | 79.7 | 5e-29 |
| ref\|NP_404760.1\| | biotin synthase [Yersinia pestis CO92] >ref\|... | 80.1 | 2e-28 |
| ref\|ZP_01497421.1\| | hypothetical protein YpseI_02002585 [Yersi... | 80.1 | 2e-28 |
| ref\|YP_454583.1\| | biotin synthase BioB [Sodalis glossinidius s... | 80.1 | 9e-26 |
| ref\|NP_797492.1\| | biotin synthase [Vibrio parahaemolyticus RIM... | 71.6 | 4e-24 |
| ref\|ZP_01661930.1\| | Biotin synthase [Ralstonia pickettii 12J] ... | 74.7 | 1e-23 |
| ref\|ZP_01914987.1\| | probable biotin synthase protein [Limnobac... | 77.4 | 1e-23 |
| gb\|EDN42426.1\| | biotin synthase [Ralstonia pickettii 12D] | 74.7 | 2e-23 |
| ref\|ZP_00967143.1\| | COG0502: Biotin synthase and related enzym... | 68.9 | 2e-23 |
| gb\|EDN56181.1\| | biotin synthase [Vibrio sp. Ex25] | 70.9 | 2e-23 |
| gb\|AAT51401.1\| | PA0500 [synthetic construct] | 68.9 | 2e-23 |
| ref\|YP_788670.1\| | biotin synthase [Pseudomonas aeruginosa UCBP... | 68.9 | 2e-23 |
| ref\|NP_249191.1\| | biotin synthase [Pseudomonas aeruginosa PAO1... | 68.9 | 2e-23 |
| ref\|ZP_01260979.1\| | biotin synthase [Vibrio alginolyticus 12G0... | 70.9 | 2e-23 |
| ref\|ZP_01473994.1\| | hypothetical protein VEx2w_02003423 [Vibrio s | 70.9 | 2e-23 |
| ref\|ZP_01640775.1\| | biotin synthase [Pseudomonas putida W619] ... | 69.3 | 4e-23 |
| ref\|YP_001189567.1\| | biotin synthase [Pseudomonas mendocina ym... | 68.9 | 4e-23 |
| ref\|ZP_01293752.1\| | hypothetical protein PaerP_01004394 [Pseud... | 68.2 | 4e-23 |
| ref\|YP_001174331.1\| | biotin synthase [Pseudomonas stutzeri A15... | 68.6 | 4e-23 |
| ref\|ZP_01869259.1\| | biotin synthase [Vibrio shilonii AK1] >gb\|... | 71.2 | 5e-23 |
| ref\|ZP_01484254.1\| | hypothetical protein VchoV5_02003172 [Vibr... | 70.1 | 6e-23 |
| ref\|NP_230757.1\| | biotin synthase [Vibrio cholerae O1 biovar e... | 70.1 | 6e-23 |
| ref\|YP_001265745.1\| | biotin synthase [Pseudomonas putida F1] >... | 69.3 | 8e-23 |
| ref\|NP_518387.1\| | PROBABLE BIOTIN SYNTHASE PROTEIN [Ralstonia ... | 68.6 | 1e-22 |
| ref\|NP_742529.1\| | biotin synthetase [Pseudomonas putida KT2440... | 69.3 | 1e-22 |
| ref\|ZP_01235385.1\| | putative biotin synthase [Vibrio angustum ... | 71.6 | 1e-22 |
| ref\|ZP_01160449.1\| | putative biotin synthase [Photobacterium s... | 71.6 | 1e-22 |
| ref\|YP_588696.1\| | biotin synthase [Baumannia cicadellinicola s... | 73.2 | 1e-22 |
| ref\|YP_262751.1\| | biotin synthetase [Pseudomonas fluorescens P... | 64.7 | 2e-22 |
| ref\|NP_790343.1\| | biotin synthetase [Pseudomonas syringae pv. ... | 65.5 | 2e-22 |
| ref\|YP_610533.1\| | biotin synthetase [Pseudomonas entomophila L... | 69.3 | 2e-22 |
| ref\|ZP_01816622.1\| | biotin synthase [Vibrionales bacterium SWA... | 70.5 | 2e-22 |
| ref\|NP_934120.1\| | biotin synthase [Vibrio vulnificus YJ016] >d... | 68.9 | 3e-22 |
| ref\|ZP_01065426.1\| | biotin synthase [Vibrio sp. MED222] >gb\|EA... | 70.1 | 3e-22 |
| ref\|NP_761747.1\| | Biotin synthase [Vibrio vulnificus CMCP6] >g... | 68.9 | 3e-22 |
| ref\|ZP_01717084.1\| | biotin synthase [Pseudomonas putida GB-1] ... | 68.6 | 4e-22 |
| ref\|YP_276822.1\| | biotin synthase [Pseudomonas syringae pv. ph... | 66.2 | 4e-22 |
| ref\|YP_237752.1\| | Biotin synthase [Pseudomonas syringae pv. sy... | 66.2 | 4e-22 |
| ref\|ZP_00992912.1\| | biotin synthase [Vibrio splendidus 12B01] ... | 70.5 | 4e-22 |
| gb\|EAY41068.1\| | biotin synthase [Vibrio cholerae MZO-3] | 67.4 | 4e-22 |

| | | | |
|---|---|---|---|
| ref\|ZP_00418345.1\| | Biotin synthase [Azotobacter vinelandii Av... | 66.6 | 4e-22 |
| ref\|YP_350904.1\| | Biotin synthase [Pseudomonas fluorescens PfO... | 65.1 | 8e-22 |
| ref\|ZP_00943498.1\| | Biotin synthase [Ralstonia solanacearum UW... | 65.9 | 1e-21 |
| ref\|YP_269309.1\| | biotin synthase [Colwellia psychrerythraea 3... | 69.7 | 1e-21 |
| ref\|YP_130516.1\| | putative biotin synthase [Photobacterium pro... | 72.0 | 1e-21 |
| ref\|YP_001142637.1\| | biotin synthase [Aeromonas salmonicida su... | 69.3 | 2e-21 |
| ref\|YP_943295.1\| | biotin synthase [Psychromonas ingrahamii 37]... | 67.4 | 2e-21 |
| ref\|ZP_01510809.1\| | biotin synthase [Burkholderia phytofirmans... | 68.6 | 2e-21 |
| ref\|YP_556786.1\| | Biotin synthase [Burkholderia xenovorans LB4... | 68.6 | 2e-21 |
| ref\|ZP_01224751.1\| | biotin synthetase [marine gamma proteobact... | 63.5 | 4e-21 |
| ref\|YP_528608.1\| | Biotin synthase [Saccharophagus degradans 2-... | 62.0 | 4e-21 |
| ref\|ZP_01222496.1\| | putative biotin synthase [Photobacterium p... | 70.1 | 5e-21 |
| ref\|YP_155713.1\| | Biotin synthase [Idiomarina loihiensis L2TR]... | 65.5 | 5e-21 |
| ref\|YP_856027.1\| | biotin synthase [Aeromonas hydrophila subsp.... | 67.0 | 6e-21 |
| ref\|ZP_01921354.1\| | biotin synthase [Shewanella sediminis HAW-... | 60.5 | 6e-21 |
| ref\|ZP_01597380.1\| | biotin synthase [Marinomonas sp. MWYL1] >g... | 68.6 | 8e-21 |
| ref\|ZP_01167078.1\| | biotin synthase; contains an iron-sulfur c... | 69.7 | 9e-21 |
| ref\|ZP_00987655.1\| | COG0502: Biotin synthase and related enzym... | 65.5 | 9e-21 |
| ref\|ZP_01540454.1\| | biotin synthase [Shewanella woodyi ATCC 51... | 58.2 | 1e-20 |
| ref\|YP_934320.1\| | biotin synthase [Azoarcus sp. BH72] >emb\|CAL... | 67.4 | 1e-20 |
| ref\|ZP_01499695.1\| | biotin synthase [Burkholderia phymatum STM... | 66.6 | 2e-20 |
| ref\|NP_871456.1\| | hypothetical protein WGLp453 [Wigglesworthia... | 63.9 | 2e-20 |
| ref\|YP_562665.1\| | biotin synthase [Shewanella denitrificans OS... | 63.2 | 2e-20 |
| ref\|YP_001093845.1\| | biotin synthase [Shewanella loihica PV-4]... | 61.6 | 2e-20 |
| ref\|YP_734487.1\| | biotin synthase [Shewanella sp. MR-4] >ref\|Y... | 59.7 | 2e-20 |
| ref\|YP_113595.1\| | biotin synthetase [Methylococcus capsulatus ... | 69.7 | 2e-20 |
| ref\|NP_881326.1\| | biotin synthase [Bordetella pertussis Tohama... | 68.6 | 2e-20 |
| ref\|ZP_01569499.1\| | biotin synthase [Burkholderia multivorans ... | 65.5 | 2e-20 |
| ref\|YP_927315.1\| | Biotin synthase [Shewanella amazonensis SB2B... | 60.8 | 2e-20 |
| ref\|ZP_01115221.1\| | biotin synthetase [Reinekea sp. MED297] >g... | 70.1 | 2e-20 |
| ref\|YP_963054.1\| | biotin synthase [Shewanella sp. W3-18-1] >re... | 59.7 | 2e-20 |
| ref\|NP_718324.1\| | biotin synthase [Shewanella oneidensis MR-1]... | 59.3 | 2e-20 |

CAGCTTTTCGTTAATCTCCTCGGCCGGCTGAAGGCCACGTAAGAGGCTGGATCATGATCCGGCGCCCTATG
GTGCGATAAAATCAAAGGCATCGACATCATCGTTATCGGCAAGCGGTGTGCTTTCACCTTCACCAGCATGT
TGATTGGCACGCTTTCCGGCGCGTCGCAGGGTTGCAAGTTGCAGCAATAATCCGGCGCGATCTTTTACCGT
TTCGCCTAAGCCCACAATGCCGCCAGAACAGACTTTGATCCCGGCATCGGGACTTTTTCCAGGGGTTATCG
GGGGCGTTCCTGATAAGTGCGTGTGGTATGAGATGAATATTGCCGTAAAACTCCGGCGAGGTGTCCAGGTT
GTGGTTGAAGTAATCCAGCCCGGCGTTCGCGAGGCGTTGCCCAAAAGATTAACTCAACGTGCCCAGCGTCA
TACACGCCTCCAGCCCCATCGCTTTTCCCCATGCCCATTTGTTCAAGGTACGGCAAATCGCGAACGTGGGG
ATTCTCCCACGCCGACCAATACAGAACCGCGCATTCTTCCGCTTTCGCTTTGCGCGCGGCTCTGGACCTGT
TCATTCCACAACGCTGGGTCCAGCCGGTTTTGTGCCCAGTTTTGCGGTAGATTTCGATCTTAA

| Sequences producing significant alignments: | | Score (Bits) | E Value |
|---|---|---|---|
| ref\|NP_308880.1\| | biotin synthetase [Escherichia coli O157:H7 ... | 89.4 | 7e-35 |
| pdb\|1R30\|A | Chain A, The Crystal Structure Of Biotin Synthase,... | 89.0 | 9e-35 |
| gb\|AAA23515.1\| | biotin synthetase | 89.0 | 9e-35 |
| ref\|YP_402502.1\| | biotin synthesis [Shigella dysenteriae Sd197... | 89.0 | 9e-35 |
| ref\|NP_706653.1\| | biotin synthesis, sulfur insertion? [Shigell... | 89.0 | 9e-35 |

| | | | |
|---|---|---|---|
| ref\|NP_415296.1\| | biotin synthase [Escherichia coli K12] >ref\|... | 89.0 | 9e-35 |
| ref\|ZP_00721165.1\| | COG0502: Biotin synthase and related enzym... | 89.4 | 2e-34 |
| sp\|Q47862\|BIOB_ESCVU | Biotin synthase (Biotin synthetase) >gb\|... | 87.4 | 2e-34 |
| ref\|NP_752785.1\| | Biotin synthase [Escherichia coli CFT073] >r... | 89.0 | 3e-34 |
| ref\|NP_286538.1\| | biotin synthesis, sulfur insertion? [Escheri... | 87.0 | 3e-34 |
| gb\|AAG60579.1\|AF250776_3 | biotin synthase BioB [uncultured bacter | 86.3 | 6e-34 |
| ref\|ZP_00698064.1\| | COG0502: Biotin synthase and related enzym... | 86.3 | 6e-34 |
| ref\|YP_215779.1\| | biotin synthetase [Salmonella enterica subsp... | 89.0 | 1e-33 |
| ref\|NP_459772.1\| | biotin synthetase [Salmonella typhimurium LT... | 89.0 | 1e-33 |
| ref\|NP_455336.1\| | biotin synthetase [Salmonella enterica subsp... | 87.8 | 3e-33 |
| gb\|ABR76236.1\| | biotin synthesis, sulfur insertion? [Klebsiell... | 85.1 | 3e-33 |
| ref\|YP_151179.1\| | biotin synthetase [Salmonella enterica subsp... | 86.3 | 8e-33 |
| ref\|YP_001175998.1\| | biotin synthase [Enterobacter sp. 638] >g... | 84.7 | 3e-32 |
| sp\|P36569\|BIOB_SERMA | Biotin synthase (Biotin synthetase) >dbj... | 79.3 | 6e-31 |
| ref\|ZP_01535732.1\| | biotin synthase [Serratia proteamaculans 5... | 81.3 | 7e-31 |
| gb\|AAG60559.1\| | BioB-like protein [uncultured bacterium pCosFS1] | 82.0 | 2e-30 |
| ref\|YP_001007095.1\| | biotin synthase [Yersinia enterocolitica ... | 78.6 | 1e-29 |
| gb\|AAG53589.1\|AF248314_4 | biotin synthase BioB [uncultured bacter | 81.6 | 2e-29 |
| ref\|ZP_00821420.1\| | COG0502: Biotin synthase and related enzym... | 78.6 | 3e-29 |
| ref\|ZP_00824684.1\| | COG0502: Biotin synthase and related enzym... | 77.0 | 4e-29 |
| ref\|ZP_00832232.1\| | COG0502: Biotin synthase and related enzym... | 76.6 | 5e-29 |
| ref\|ZP_00829477.1\| | COG0502: Biotin synthase and related enzym... | 78.6 | 8e-29 |
| ref\|NP_928780.1\| | biotin synthetase (biotin synthase) [Photorh... | 79.0 | 3e-28 |
| ref\|YP_050916.1\| | biotin synthase [Erwinia carotovora subsp. a... | 74.3 | 5e-28 |
| ref\|YP_454583.1\| | biotin synthase BioB [Sodalis glossinidius s... | 76.6 | 1e-27 |
| ref\|NP_404760.1\| | biotin synthase [Yersinia pestis CO92] >ref\|... | 76.6 | 1e-27 |
| ref\|ZP_01497421.1\| | hypothetical protein YpseI_02002585 [Yersi... | 76.3 | 2e-27 |
| ref\|NP_797492.1\| | biotin synthase [Vibrio parahaemolyticus RIM... | 78.6 | 2e-25 |
| gb\|EAY41068.1\| | biotin synthase [Vibrio cholerae MZO-3] | 77.8 | 7e-25 |
| ref\|YP_001189567.1\| | biotin synthase [Pseudomonas mendocina ym... | 73.9 | 9e-25 |
| ref\|ZP_01640775.1\| | biotin synthase [Pseudomonas putida W619] ... | 73.2 | 9e-25 |
| ref\|ZP_01484254.1\| | hypothetical protein VchoV5_02003172 [Vibr... | 77.8 | 9e-25 |
| ref\|NP_230757.1\| | biotin synthase [Vibrio cholerae O1 biovar e... | 77.8 | 9e-25 |
| gb\|EDN56181.1\| | biotin synthase [Vibrio sp. Ex25] | 77.0 | 1e-24 |
| ref\|ZP_01260979.1\| | biotin synthase [Vibrio alginolyticus 12G0... | 77.0 | 1e-24 |
| ref\|ZP_01473994.1\| | hypothetical protein VEx2w_02003423 [Vibrio s | 77.0 | 1e-24 |
| ref\|ZP_01869259.1\| | biotin synthase [Vibrio shilonii AK1] >gb\|... | 76.3 | 1e-24 |
| ref\|YP_001174331.1\| | biotin synthase [Pseudomonas stutzeri A15... | 74.3 | 1e-24 |
| ref\|ZP_00418345.1\| | Biotin synthase [Azotobacter vinelandii Av... | 76.3 | 1e-24 |
| ref\|ZP_01293752.1\| | hypothetical protein PaerP_01004394 [Pseud... | 72.8 | 2e-24 |
| gb\|AAT51401.1\| | PA0500 [synthetic construct] | 72.8 | 2e-24 |
| ref\|ZP_00967143.1\| | COG0502: Biotin synthase and related enzym... | 72.8 | 2e-24 |
| ref\|YP_788670.1\| | biotin synthase [Pseudomonas aeruginosa UCBP... | 72.8 | 2e-24 |
| ref\|NP_249191.1\| | biotin synthase [Pseudomonas aeruginosa PAO1... | 72.8 | 2e-24 |
| ref\|YP_001265745.1\| | biotin synthase [Pseudomonas putida F1] >... | 72.0 | 2e-24 |
| ref\|NP_742529.1\| | biotin synthetase [Pseudomonas putida KT2440... | 71.6 | 2e-24 |
| ref\|ZP_01717084.1\| | biotin synthase [Pseudomonas putida GB-1] ... | 70.5 | 5e-24 |
| ref\|YP_610533.1\| | biotin synthetase [Pseudomonas entomophila L... | 70.5 | 5e-24 |
| ref\|NP_934120.1\| | biotin synthase [Vibrio vulnificus YJ016] >d... | 75.9 | 7e-24 |
| ref\|NP_761747.1\| | Biotin synthase [Vibrio vulnificus CMCP6] >g... | 75.9 | 7e-24 |
| ref\|ZP_01914987.1\| | probable biotin synthase protein [Limnobac... | 68.9 | 7e-24 |
| ref\|YP_350904.1\| | Biotin synthase [Pseudomonas fluorescens PfO... | 73.6 | 9e-24 |
| ref\|YP_262751.1\| | biotin synthetase [Pseudomonas fluorescens P... | 73.9 | 1e-23 |
| ref\|ZP_01816622.1\| | biotin synthase [Vibrionales bacterium SWA... | 73.9 | 1e-23 |

| | | | |
|---|---|---|---|
| ref\|NP_790343.1\| | biotin synthetase [Pseudomonas syringae pv. ... | 72.0 | 2e-23 |
| ref\|ZP_01065426.1\| | biotin synthase [Vibrio sp. MED222] >gb\|EA... | 73.9 | 2e-23 |
| ref\|YP_155713.1\| | Biotin synthase [Idiomarina loihiensis L2TR]... | 74.3 | 2e-23 |
| ref\|YP_276822.1\| | biotin synthase [Pseudomonas syringae pv. ph... | 71.2 | 3e-23 |
| ref\|YP_237752.1\| | Biotin synthase [Pseudomonas syringae pv. sy... | 71.2 | 3e-23 |
| ref\|ZP_00992912.1\| | biotin synthase [Vibrio splendidus 12B01] ... | 72.8 | 3e-23 |
| ref\|ZP_01115221.1\| | biotin synthetase [Reinekea sp. MED297] >g... | 69.7 | 3e-23 |
| ref\|YP_588696.1\| | biotin synthase [Baumannia cicadellinicola s... | 68.2 | 4e-23 |
| ref\|ZP_01102888.1\| | biotin synthase [gamma proteobacterium KT ... | 74.7 | 6e-23 |
| ref\|ZP_01235385.1\| | putative biotin synthase [Vibrio angustum ... | 70.5 | 6e-23 |
| ref\|ZP_01160449.1\| | putative biotin synthase [Photobacterium s... | 70.1 | 8e-23 |
| ref\|ZP_00987655.1\| | COG0502: Biotin synthase and related enzym... | 77.8 | 8e-23 |
| ref\|YP_693940.1\| | biotin synthase [Alcanivorax borkumensis SK2... | 70.5 | 1e-22 |
| ref\|ZP_01569499.1\| | biotin synthase [Burkholderia multivorans ... | 76.6 | 2e-22 |
| emb\|CAB56476.1\| | biotin synthetase [Pseudomonas stutzeri] | 73.2 | 2e-22 |
| gb\|EDN42426.1\| | biotin synthase [Ralstonia pickettii 12D] | 68.6 | 3e-22 |
| ref\|ZP_01661930.1\| | Biotin synthase [Ralstonia pickettii 12J] ... | 68.6 | 3e-22 |
| ref\|ZP_01558074.1\| | biotin synthase [Burkholderia ambifaria MC... | 77.8 | 3e-22 |
| ref\|ZP_01224751.1\| | biotin synthetase [marine gamma proteobact... | 70.9 | 4e-22 |
| ref\|YP_130516.1\| | putative biotin synthase [Photobacterium pro... | 65.1 | 4e-22 |
| ref\|YP_744320.1\| | biotin synthase [Granulibacter bethesdensis ... | 72.0 | 4e-22 |
| ref\|YP_774873.1\| | biotin synthase [Burkholderia cepacia AMMD] ... | 77.8 | 5e-22 |
| ref\|ZP_00943498.1\| | Biotin synthase [Ralstonia solanacearum UW... | 72.0 | 8e-22 |
| ref\|YP_001093845.1\| | biotin synthase [Shewanella loihica PV-4]... | 74.7 | 8e-22 |
| ref\|ZP_00980828.1\| | COG0502: Biotin synthase and related enzym... | 76.3 | 8e-22 |
| ref\|ZP_01043921.1\| | Biotin synthase [Idiomarina baltica OS145]... | 68.9 | 1e-21 |
| ref\|YP_927315.1\| | Biotin synthase [Shewanella amazonensis SB2B... | 73.6 | 1e-21 |
| ref\|ZP_01222496.1\| | putative biotin synthase [Photobacterium p... | 65.1 | 1e-21 |
| ref\|YP_370516.1\| | Biotin synthase [Burkholderia sp. 383] >gb\|A... | 75.9 | 2e-21 |
| ref\|YP_001120855.1\| | biotin synthase [Burkholderia vietnamiens... | 75.9 | 2e-21 |
| ref\|YP_934320.1\| | biotin synthase [Azoarcus sp. BH72] >emb\|CAL... | 69.7 | 2e-21 |
| ref\|ZP_01510809.1\| | biotin synthase [Burkholderia phytofirmans... | 70.9 | 2e-21 |
| ref\|YP_001142637.1\| | biotin synthase [Aeromonas salmonicida su... | 68.6 | 2e-21 |
| ref\|YP_856027.1\| | biotin synthase [Aeromonas hydrophila subsp.... | 67.8 | 2e-21 |
| ref\|NP_518387.1\| | PROBABLE BIOTIN SYNTHASE PROTEIN [Ralstonia ... | 71.6 | 2e-21 |
| ref\|NP_871456.1\| | hypothetical protein WGLp453 [Wigglesworthia... | 64.7 | 3e-21 |
| ref\|YP_269309.1\| | biotin synthase [Colwellia psychrerythraea 3... | 63.5 | 3e-21 |
| ref\|YP_192423.1\| | Biotin synthase [Gluconobacter oxydans 621H]... | 73.2 | 3e-21 |
| ref\|YP_556786.1\| | Biotin synthase [Burkholderia xenovorans LB4... | 70.9 | 4e-21 |
| ref\|YP_562665.1\| | biotin synthase [Shewanella denitrificans OS... | 74.3 | 4e-21 |
| ref\|ZP_01921354.1\| | biotin synthase [Shewanella sediminis HAW-... | 73.2 | 4e-21 |

**BioC**

BioC6F: 5’-TNG CNG TNC ART GGT GYG G-3’
BioC7R: 5’-GTN GCN CCD ATN CCY TT -3’

AAAAGTATGGACTGACGCATCATCAAACCCAGCGTGATGGGCTGAATATGATGTTGATAATGCACGCCGTT
CAGCGACTGTTCGATTTCATCTGGCGGTAAAAAGCGATTAGCATGCGGACGCTCGTCCACCGCCTGCCACG
CCTGATGCAGTTCGGGTAACGATCCCTGCACCAGCGTGGTAAACGCGACCACGCCTTTGGGGCGCACCACC
CGATACAGCTCGCGGAGTGCCGCGGATAAATTACCGCACCACTGCTCCGCCACACAATTGTACAGCTTGCT
TCCTTGCTGGCCTCACTCCCCAACGATAACTCCCAAAAATGCCCCGCCGATTAACTAACCACTTCCCTAGC
AAATACCCCCCCCCAATTTGCCCCCAGGTGCCCAAATACCTCTCTTTCACAGCGACGACCTGCTCTGACTCC
TTATCATCTCCTCCTTCGTCTACGATGCAGCCGCGACTAACCCCCGCCCGTTCAACCTGCCTTG

| Sequences producing significant alignments: | Score (Bits) | E Value | |
|---|---|---|---|
| ref\|NP_415298.1\| predicted methltransferase, enzyme of biotin... | 162 | 8e-39 | G |
| gb\|AAA23517.1\| bioC | 159 | 4e-38 | |
| ref\|ZP_00703085.1\| COG0500: SAM-dependent methyltransferases ... | 159 | 5e-38 | |
| ref\|YP_309740.1\| biotin biosynthesis protein [Shigella sonnei... | 159 | 5e-38 | G |
| gb\|AAG60581.1\|AF250776_5 biotin biosynthesis protein BioC [uncul | 159 | 7e-38 | |
| ref\|ZP_00721167.1\| COG0500: SAM-dependent methyltransferases ... | 157 | 2e-37 | |
| ref\|NP_286540.1\| biotin biosynthesis; reaction prior to pimel... | 156 | 5e-37 | G |
| ref\|YP_402500.1\| biotin biosynthesis protein [Shigella dysent... | 155 | 6e-37 | G |
| ref\|YP_539792.1\| biotin synthesis protein BioC [Escherichia c... | 154 | 1e-36 | G |
| ref\|ZP_00709066.1\| COG0500: SAM-dependent methyltransferases ... | 154 | 2e-36 | |
| ref\|NP_752787.1\| Biotin synthesis protein bioC [Escherichia c... | 152 | 5e-36 | G |
| ref\|NP_836433.1\| biotin biosynthesis protein BioC [Shigella f... | 145 | 6e-34 | G |
| ref\|YP_151177.1\| biotin synthesis protein BioC [Salmonella en... | 106 | 5e-22 | G |
| ref\|NP_455338.1\| biotin synthesis protein BioC [Salmonella en... | 106 | 5e-22 | G |
| ref\|YP_215781.1\| biotin biosynthesis; reaction prior to pimel... | 105 | 1e-21 | G |
| ref\|NP_459774.1\| biotin biosynthetic protein [Salmonella typh... | 105 | 1e-21 | G |
| ref\|YP_001176000.1\| biotin biosynthesis protein BioC [Enterob... | 102 | 1e-20 | G |
| sp\|O06898\|BIOC_ESCVU Biotin synthesis protein bioC >gb\|AAB529... | 97.8 | 2e-19 | |
| gb\|ABR76238.1\| biotin biosynthesis; reaction prior to pimeloy... | 91.7 | 1e-17 | |
| gb\|AAG53591.1\|AF248314_6 biotin biosynthesis BioC-like protei... | 89.7 | 5e-17 | |
| gb\|AAG60557.1\|AF250768_2 biotin biosynthesis protein BioC [uncul | 82.4 | 8e-15 | |
| ref\|NP_404762.1\| biotin synthesis protein BioC [Yersinia pest... | 82.4 | 8e-15 | G |
| ref\|YP_069719.1\| biotin synthesis protein BioC [Yersinia pseu... | 82.0 | 1e-14 | G |
| ref\|ZP_01497419.1\| hypothetical protein YpseI_02002583 [Yersi... | 82.0 | 1e-14 | |
| ref\|ZP_00824682.1\| COG0500: SAM-dependent methyltransferases ... | 80.5 | 3e-14 | |
| ref\|YP_050914.1\| biotin synthesis protein [Erwinia carotovora... | 79.7 | 5e-14 | G |
| ref\|NP_928782.1\| biotin synthesis protein [Photorhabdus lumin... | 79.7 | 5e-14 | G |
| ref\|ZP_00832234.1\| COG0500: SAM-dependent methyltransferases ... | 79.0 | 9e-14 | |
| ref\|ZP_00821418.1\| COG0500: SAM-dependent methyltransferases ... | 77.4 | 3e-13 | |
| ref\|YP_001007093.1\| biotin synthesis protein BioC [Yersinia e... | 76.3 | 6e-13 | G |
| ref\|ZP_01535734.1\| biotin biosynthesis protein BioC [Serratia... | 75.9 | 8e-13 | |
| ref\|YP_454585.1\| biotin synthesis protein BioC [Sodalis gloss... | 75.9 | 8e-13 | G |
| ref\|NP_761745.1\| Biotin synthesis protein [Vibrio vulnificus ... | 75.1 | 1e-12 | G |
| sp\|P36571\|BIOC_SERMA Biotin synthesis protein bioC >dbj\|BAA04... | 75.1 | 1e-12 | |
| ref\|YP_130514.1\| putative biotin synthesis protein BioC [Phot... | 73.9 | 3e-12 | G |
| ref\|ZP_00829475.1\| COG0500: SAM-dependent methyltransferases ... | 73.9 | 3e-12 | |
| ref\|ZP_01222498.1\| putative biotin synthesis protein BioC [Ph... | 73.6 | 4e-12 | |
| ref\|ZP_01235387.1\| putative biotin synthesis protein BioC [Vi... | 71.2 | 2e-11 | |
| ref\|ZP_00418348.1\| probable biotin synthesis protein BioC [Az... | 71.2 | 2e-11 | |
| ref\|ZP_01160447.1\| putative biotin synthesis protein BioC [Ph... | 70.9 | 3e-11 | |
| ref\|YP_276819.1\| biotin biosynthesis protein BioC [Pseudomona... | 70.1 | 4e-11 | G |
| ref\|ZP_01307867.1\| biotin synthesis protein BioC [Oceanobacte... | 69.3 | 7e-11 | |
| ref\|YP_588698.1\| biotin biosynthesis protein BioC [Baumannia ... | 68.9 | 1e-10 | G |
| ref\|YP_001189564.1\| biotin biosynthesis protein BioC [Pseudom... | 67.0 | 4e-10 | G |
| ref\|YP_237749.1\| biotin synthesis protein BioC [Pseudomonas s... | 67.0 | 4e-10 | G |
| ref\|YP_610530.1\| biotin biosynthesis protein BioC [Pseudomona... | 66.2 | 6e-10 | G |
| gb\|EAY41077.1\| biotin synthesis protein BioC [Vibrio cholerae MZ | 65.1 | 1e-09 | |
| ref\|NP_790346.1\| biotin synthesis protein BioC [Pseudomonas s... | 64.7 | 2e-09 | G |
| gb\|EDM52799.1\| biotin synthesis protein BioC [Vibrio cholerae MZ | 64.3 | 2e-09 | |
| gb\|EDN14557.1\| biotin synthesis protein BioC [Vibrio cholerae AM | 63.9 | 3e-09 | |
| gb\|EDL74527.1\| biotin synthesis protein BioC [Vibrio cholerae 62 | 63.9 | 3e-09 | |

| | | | |
|---|---|---|---|
| ref\|YP_001142635.1\| | biotin synthesis protein BioC [Aeromonas ... | 63.9 | 3e-09 |
| ref\|ZP_01678573.1\| | biotin synthesis protein BioC [Vibrio chol... | 63.9 | 3e-09 |
| ref\|ZP_01597377.1\| | Methyltransferase type 11 [Marinomonas sp.... | 63.9 | 3e-09 |
| ref\|NP_230759.1\| | biotin synthesis protein BioC [Vibrio choler... | 63.9 | 3e-09 |
| ref\|NP_871458.1\| | hypothetical protein WGLp455 [Wigglesworthia... | 63.5 | 4e-09 |
| ref\|ZP_01484411.1\| | hypothetical protein VchoV5_02003001 [Vibr... | 63.5 | 4e-09 |
| ref\|ZP_01717081.1\| | biotin biosynthesis protein BioC [Pseudomo... | 62.8 | 7e-09 |
| ref\|NP_742532.1\| | biotin biosynthesis protein BioC [Pseudomona... | 62.8 | 7e-09 |
| ref\|ZP_01898733.1\| | Biotin synthesis protein [Moritella sp. PE... | 62.4 | 9e-09 |
| gb\|EAY34736.1\| | biotin synthesis protein BioC [Vibrio cholerae 15 | 62.4 | 9e-09 |
| ref\|ZP_01293749.1\| | hypothetical protein PaerP_01004391 [Pseud... | 62.4 | 9e-09 |
| gb\|AAT51029.1\| | PA0503 [synthetic construct] | 62.4 | 9e-09 |
| ref\|YP_001265748.1\| | biotin biosynthesis protein BioC [Pseudom... | 62.4 | 9e-09 |
| ref\|ZP_01473996.1\| | hypothetical protein VEx2w_02003425 [Vibri... | 62.4 | 9e-09 |
| ref\|YP_788673.1\| | putative biotin synthesis protein BioC [Pseu... | 62.4 | 9e-09 |
| ref\|NP_249194.1\| | probable biotin synthesis protein BioC [Pseu... | 62.4 | 9e-09 |
| ref\|YP_262748.1\| | biotin synthesis protein BioC [Pseudomonas f... | 62.0 | 1e-08 |
| ref\|ZP_01640772.1\| | Biotin biosynthesis protein BioC [Pseudomo... | 61.6 | 2e-08 |
| ref\|ZP_01816624.1\| | biotin synthesis protein BioC [Vibrionales... | 61.2 | 2e-08 |
| ref\|YP_856029.1\| | biotin biosynthesis protein BioC [Aeromonas ... | 61.2 | 2e-08 |
| ref\|YP_155711.1\| | Biotin synthesis protein [Idiomarina loihien... | 60.8 | 3e-08 |
| ref\|ZP_00992914.1\| | biotin synthesis protein BioC [Vibrio sple... | 60.8 | 3e-08 |
| ref\|YP_113598.1\| | biotin synthesis protein BioC [Methylococcus... | 60.5 | 3e-08 |
| ref\|NP_797494.1\| | biotin synthesis protein BioC [Vibrio paraha... | 60.1 | 4e-08 |
| ref\|ZP_01065424.1\| | biotin synthesis protein BioC [Vibrio sp. ... | 60.1 | 4e-08 |
| ref\|YP_001174328.1\| | biotin synthesis protein BioC [Pseudomona... | 59.7 | 6e-08 |
| ref\|YP_350901.1\| | Biotin biosynthesis protein BioC [Pseudomona... | 59.7 | 6e-08 |
| ref\|YP_206704.1\| | biotin synthesis protein BioC [Vibrio fische... | 58.9 | 1e-07 |
| ref\|ZP_01260977.1\| | biotin synthesis protein BioC [Vibrio algi... | 58.5 | 1e-07 |
| ref\|ZP_01869261.1\| | biotin synthesis protein BioC [Vibrio shil... | 57.8 | 2e-07 |
| ref\|YP_392245.1\| | Biotin biosynthesis protein BioC [Thiomicros... | 57.4 | 3e-07 |
| ref\|YP_943293.1\| | biotin biosynthesis protein BioC [Psychromon... | 56.6 | 5e-07 |
| ref\|ZP_01133857.1\| | desthiobiotin biosynthesis; reaction prior... | 56.6 | 5e-07 |
| ref\|ZP_01613054.1\| | desthiobiotin biosynthesis; reaction prior... | 55.5 | 1e-06 |
| ref\|ZP_01167075.1\| | biotin synthesis protein BioC [Oceanospiri... | 54.7 | 2e-06 |
| ref\|YP_344088.1\| | Biotin biosynthesis protein BioC [Nitrosococ... | 53.9 | 3e-06 |
| ref\|YP_269311.1\| | biotin biosynthesis protein bioC [Colwellia ... | 53.9 | 3e-06 |
| ref\|ZP_01127180.1\| | biotin synthesis protein [Nitrococcus mobi... | 53.9 | 3e-06 |
| ref\|NP_777890.1\| | biotin synthesis protein BioC [Buchnera aphi... | 53.1 | 5e-06 |
| ref\|ZP_01737942.1\| | biotin synthesis protein [Marinobacter sp.... | 52.8 | 7e-06 |
| ref\|YP_903510.1\| | biotin biosynthesis protein BioC [Candidatus... | 52.8 | 7e-06 |
| ref\|YP_286099.1\| | Biotin biosynthesis protein BioC [Dechloromo... | 52.0 | 1e-05 |
| ref\|YP_314080.1\| | Biotin biosynthesis protein BioC [Thiobacill... | 51.6 | 2e-05 |
| ref\|YP_960016.1\| | biotin biosynthesis protein BioC [Marinobact... | 51.2 | 2e-05 |
| ref\|YP_001219097.1\| | biotin biosynthesis protein BioC [Candida... | 50.8 | 3e-05 |
| ref\|YP_743263.1\| | biotin biosynthesis protein BioC [Alkalilimn... | 50.8 | 3e-05 |
| ref\|ZP_01216783.1\| | biotin synthesis protein BioC [Psychromona... | 50.8 | 3e-05 |
| ref\|ZP_01895249.1\| | biotin synthesis protein BioC [Marinobacte... | 49.7 | 6e-05 |
| ref\|YP_748322.1\| | biotin biosynthesis protein BioC [Nitrosomon... | 49.7 | 6e-05 |

**BioD**

BioD1F: 5′-GAY ACN GAR GTN GGN AA-3′
BioD3R: 5′-GCR TGR TTD ATR CAN CC-3′

TTGTTGTATACGACTCCTATAGGGCGAACTGGGCCCTCTAGATGCAAGCTCGAGCGGCCGCCAGTGTGATG
GATATCTGCAGAATTCGCCCTTGATACAGAAGTGGGGAAAACCGTCGCCAGCAGCGCCTTGCTGCAGGCCG
CCAACCGGGCGGGCTACCGCAGCGCCGGCTATAAGCCGGTGGCCTCCGGCAGCGAGATGACCGCCGAGGGG
CTGCGCAACGGCGACGCGCTGGCGCTGCAGGCCAACAGCGGTGTGGCGCTGGATTACGACGAAGTGAACCC
TTACGTATTTGCCGAACCGACCTCGCCGCATATCGTCAGCGCCGATGAAGGCCGGCCGATCGACGCGGAGC
GGCTGTCCAACGGCCTGCGCCGGCTGGAGCAGCGCGCCGACTGGGTGCTGGTCGAGGGGGGCGCGGCGGGG
TGGTTTACCCCGTTGTCGGCGGAGTACACCCTTCGCCGACTGGGTGCGGCAAGAGCAACTGCCGGTGATCC
TGGTGGTGGGGCATCAAGCTGGGCTGCATCAACCACGCAAGGGCGAATTCCAGCACACTGGCGGCCGGTTA
CTATGGAATCCGAGCTCGGGTACCAGCTTGGCGTAATCATGGTCAT

| Sequences producing significant alignments: | Score (Bits) | E Value | |
|---|---|---|---|
| sp\|P36572\|BIOD_SERMA Dethiobiotin synthetase (Dethiobiotin sy... | 223 | 8e-64 | |
| ref\|ZP_01535735.1\| dethiobiotin synthase [Serratia proteamacu... | 199 | 2e-56 | |
| gb\|AAA23518.1\| dethiobiotin synthetase | 180 | 3e-50 | |
| gb\|AAG60582.1\|AF250776_6 dethiobiotin synthetase BioD [unculture | 180 | 3e-50 | |
| gb\|ABR76239.1\| dithiobiotin synthetase [Klebsiella pneumoniae... | 181 | 8e-50 | |
| ref\|ZP_01702227.1\| dethiobiotin synthase [Escherichia coli B]... | 177 | 1e-49 | |
| ref\|NP_415299.1\| dethiobiotin synthetase [Escherichia coli K1... | 177 | 1e-49 | G |
| ref\|ZP_00705000.1\| COG0132: Dethiobiotin synthetase [Escheric... | 177 | 1e-49 | |
| pdb\|1DBS\| Chain , Dethiobiotin Synthetase (E.C.6.3.3.3) >pd... | 177 | 1e-49 | S |
| pdb\|1DTS\| Chain , Dethiobiotin Synthase (E.C.6.3.3.3) | 177 | 1e-49 | S |
| ref\|ZP_00734066.1\| COG0132: Dethiobiotin synthetase [Escherichia | 177 | 1e-49 | |
| ref\|YP_539793.1\| dethiobiotin synthetase [Escherichia coli UT... | 177 | 2e-49 | G |
| ref\|NP_286541.1\| dithiobiotin synthetase [Escherichia coli O1... | 177 | 2e-49 | G |
| ref\|YP_668710.1\| dethiobiotin synthetase [Escherichia coli 53... | 176 | 3e-49 | G |
| ref\|YP_402499.1\| dethiobiotin synthetase [Shigella dysenteria... | 176 | 3e-49 | G |
| ref\|YP_309741.1\| dethiobiotin synthetase [Shigella sonnei Ss0... | 176 | 5e-49 | G |
| ref\|ZP_00698067.1\| COG0132: Dethiobiotin synthetase [Shigella bo | 175 | 8e-49 | |
| ref\|ZP_00721168.1\| COG0132: Dethiobiotin synthetase [Escherichia | 174 | 1e-48 | |
| ref\|YP_407172.1\| dethiobiotin synthetase [Shigella boydii Sb2... | 174 | 1e-48 | G |
| ref\|ZP_00709067.1\| COG0132: Dethiobiotin synthetase [Escherichia | 176 | 1e-48 | |
| ref\|NP_836434.1\| dithiobiotin synthetase [Shigella flexneri 2... | 172 | 4e-48 | G |
| sp\|O06899\|BIOD_ESCVU Dethiobiotin synthetase (Dethiobiotin sy... | 172 | 1e-47 | |
| ref\|NP_459775.1\| dithiobiotin synthetase [Salmonella typhimur... | 172 | 1e-47 | G |
| gb\|AAG53592.1\|AF248314_7 dethiobiotin synthase BioD [uncultured | 174 | 2e-47 | |
| ref\|YP_001176001.1\| dethiobiotin synthase [Enterobacter sp. 6... | 172 | 2e-47 | G |
| ref\|YP_215782.1\| dithiobiotin synthetase [Salmonella enterica... | 172 | 3e-47 | G |
| ref\|YP_151176.1\| dithiobiotin synthetase [Salmonella enterica... | 171 | 4e-47 | G |
| ref\|NP_455339.1\| dithiobiotin synthetase [Salmonella enterica... | 171 | 4e-47 | G |
| ref\|YP_050913.1\| dithiobiotin synthetase [Erwinia carotovora ... | 168 | 2e-46 | G |
| ref\|NP_752788.1\| dithiobiotin synthetase [Escherichia coli CF... | 166 | 3e-46 | G |
| ref\|ZP_00824681.1\| COG0132: Dethiobiotin synthetase [Yersinia... | 165 | 8e-46 | |
| ref\|ZP_01497418.1\| hypothetical protein YpseI_02002582 [Yersi... | 163 | 4e-45 | |
| ref\|NP_670328.1\| dithiobiotin synthetase [Yersinia pestis KIM... | 162 | 6e-45 | G |
| ref\|NP_404763.1\| dithiobiotin synthetase [Yersinia pestis CO9... | 162 | 6e-45 | G |
| ref\|ZP_00832235.1\| COG0132: Dethiobiotin synthetase [Yersinia... | 164 | 8e-45 | |
| ref\|ZP_00821417.1\| COG0132: Dethiobiotin synthetase [Yersinia... | 160 | 2e-44 | |
| ref\|NP_928783.1\| dithiobiotin synthetase [Photorhabdus lumine... | 159 | 2e-43 | G |
| ref\|ZP_00829474.1\| COG0132: Dethiobiotin synthetase [Yersinia... | 155 | 5e-43 | |
| ref\|YP_454586.1\| dethiobiotin synthase [Sodalis glossinidius ... | 155 | 1e-37 | G |

| | | | |
|---|---|---|---|
| ref\|YP_588699.1\| | dethiobiotin synthase [Baumannia cicadellini... | 144 | 1e-37 |
| ref\|YP_001007092.1\| | dethiobiotin synthetase [Yersinia enteroc... | 137 | 2e-37 |
| ref\|ZP_01235388.1\| | dithiobiotin synthetase [Vibrio angustum S... | 132 | 3e-33 |
| ref\|ZP_01160446.1\| | putative dethiobiotin synthetase [Photobac... | 131 | 3e-33 |
| ref\|YP_130513.1\| | dithiobiotin synthetase [Photobacterium prof... | 121 | 1e-29 |
| ref\|YP_206703.1\| | dithiobiotin synthetase [Vibrio fischeri ES1... | 118 | 3e-29 |
| ref\|ZP_01222499.1\| | dithiobiotin synthetase [Photobacterium pr... | 117 | 7e-29 |
| ref\|ZP_01869262.1\| | dithiobiotin synthetase [Vibrio shilonii A... | 113 | 4e-28 |
| ref\|NP_761744.1\| | dithiobiotin synthetase [Vibrio vulnificus C... | 111 | 2e-27 |
| gb\|EDN56178.1\| | dethiobiotin synthase [Vibrio sp. Ex25] | 110 | 5e-27 |
| ref\|ZP_01473997.1\| | hypothetical protein VEx2w_02003426 [Vibrio s | 110 | 5e-27 |
| ref\|NP_240114.1\| | dethiobiotin synthetase [Buchnera aphidicola... | 114 | 5e-27 |
| ref\|NP_797495.1\| | dithiobiotin synthetase [Vibrio parahaemolyt... | 108 | 8e-27 |
| sp\|Q87QN3\|BIOD_VIBPA | Dethiobiotin synthetase (Dethiobiotin sy... | 108 | 8e-27 |
| ref\|ZP_01260976.1\| | dithiobiotin synthetase [Vibrio alginolyti... | 107 | 2e-26 |
| ref\|NP_230760.1\| | dithiobiotin synthetase [Vibrio cholerae O1 ... | 106 | 5e-26 |
| ref\|ZP_01478768.1\| | hypothetical protein VchoM_02002132 [Vibri... | 106 | 5e-26 |
| gb\|EAY37851.1\| | dethiobiotin synthase [Vibrio cholerae MAK 757] | 106 | 5e-26 |
| gb\|EDN14558.1\| | dethiobiotin synthetase [Vibrio cholerae AM-19226 | 106 | 6e-26 |
| gb\|EDL74522.1\| | dethiobiotin synthetase [Vibrio cholerae 623-39] | 106 | 6e-26 |
| ref\|ZP_00820371.1\| | COG0132: Dethiobiotin synthetase [Yersinia... | 100 | 6e-26 |
| gb\|EAY34705.1\| | dethiobiotin synthase [Vibrio cholerae 1587] | 106 | 6e-26 |
| ref\|NP_669423.1\| | hypothetical protein y2111 [Yersinia pestis ... | 100 | 1e-25 |
| gb\|AAL14974.1\|AF418982_11 | putative dethiobiotin synthase [Yersin | 100 | 1e-25 |
| ref\|NP_405805.1\| | putative dethiobiotin synthetase [Yersinia p... | 100 | 1e-25 |
| ref\|ZP_01065423.1\| | dithiobiotin synthetase [Vibrio sp. MED222... | 106 | 1e-25 |
| ref\|YP_001142634.1\| | dethiobiotin synthetase [Aeromonas salmon... | 104 | 1e-25 |
| ref\|YP_856030.1\| | dethiobiotin synthase [Aeromonas hydrophila ... | 104 | 2e-25 |
| ref\|ZP_01816625.1\| | dithiobiotin synthetase [Vibrionales bacte... | 105 | 3e-25 |
| ref\|YP_365467.1\| | dithiobiotin synthetase [Xanthomonas campest... | 101 | 4e-25 |
| ref\|YP_662124.1\| | dethiobiotin synthase [Pseudoalteromonas atl... | 110 | 5e-25 |
| ref\|YP_001174327.1\| | dethiobiotin synthase [Pseudomonas stutze... | 111 | 6e-25 |
| ref\|ZP_00992915.1\| | dithiobiotin synthetase [Vibrio splendidus... | 103 | 7e-25 |
| ref\|YP_001008378.1\| | putative dethiobiotin synthetase [Yersini... | 97.4 | 7e-25 |
| ref\|ZP_00827097.1\| | COG0132: Dethiobiotin synthetase [Yersinia... | 97.4 | 7e-25 |
| ref\|YP_199406.1\| | dithiobiotin synthetase [Xanthomonas oryzae ... | 99.4 | 1e-24 |
| ref\|YP_449726.1\| | dethiobiotin synthetase [Xanthomonas oryzae ... | 99.0 | 2e-24 |
| ref\|ZP_01167074.1\| | dethiobiotin synthase [Oceanospirillum sp.... | 109 | 2e-24 |
| ref\|NP_790347.1\| | dithiobiotin synthetase [Pseudomonas syringa... | 110 | 2e-24 |
| ref\|ZP_00834859.1\| | COG0132: Dethiobiotin synthetase [Yersinia... | 95.5 | 2e-24 |
| ref\|YP_943292.1\| | dethiobiotin synthase [Psychromonas ingraham... | 105 | 3e-24 |
| ref\|YP_788674.1\| | dethiobiotin synthase [Pseudomonas aeruginos... | 109 | 3e-24 |
| ref\|NP_837279.1\| | dithiobiotin synthetase [Shigella flexneri 2... | 97.4 | 4e-24 |
| ref\|NP_460449.1\| | dithiobiotin synthetase [Salmonella typhimur... | 97.4 | 5e-24 |
| ref\|YP_150621.1\| | dithiobiotin synthetase [Salmonella enterica... | 97.1 | 7e-24 |
| ref\|ZP_01293748.1\| | hypothetical protein PaerP_01004390 [Pseud... | 108 | 7e-24 |
| ref\|ZP_01216784.1\| | dethiobiotin synthetase [Psychromonas sp. ... | 103 | 7e-24 |
| ref\|ZP_00828727.1\| | COG0132: Dethiobiotin synthetase [Yersinia... | 94.4 | 9e-24 |
| ref\|YP_403190.1\| | hypothetical protein SDY_1563 [Shigella dyse... | 95.9 | 1e-23 |
| ref\|ZP_00717407.1\| | COG0132: Dethiobiotin synthetase [Escherichia | 95.9 | 1e-23 |
| ref\|NP_288028.1\| | dithiobiotin synthetase [Escherichia coli O1... | 95.9 | 1e-23 |
| ref\|NP_455990.1\| | dithiobiotin synthetase [Salmonella enterica... | 96.3 | 1e-23 |
| ref\|NP_310326.2\| | dithiobiotin synthetase [Escherichia coli O1... | 95.9 | 1e-23 |

| | | | |
|---|---|---|---|
| ref\|NP_777889.1\| | dethiobiotin synthetase [Buchnera aphidicola... | 100 | 2e-23 |
| ref\|YP_244707.1\| | dithiobiotin synthetase [Xanthomonas campest... | 95.9 | 2e-23 |
| ref\|YP_310505.1\| | hypothetical protein SSON_1571 [Shigella son... | 95.1 | 2e-23 |
| ref\|NP_753879.1\| | dithiobiotin synthetase [Escherichia coli CF... | 95.1 | 2e-23 |
| ref\|YP_669445.1\| | putative dethiobiotin synthetase [Escherichi... | 95.1 | 2e-23 |
| gb\|AAT50325.1\| | PA0504 [synthetic construct] | 106 | 2e-23 |
| ref\|NP_249195.1\| | dithiobiotin synthetase [Pseudomonas aerugin... | 106 | 2e-23 |
| ref\|YP_216494.1\| | dithiobiotin synthetase [Salmonella enterica... | 95.1 | 3e-23 |

---

## BioF

BioF10F: 5'-ATG GTN GAY GAY GCN CAY GG -3'
BioF11R: 5'-GCY TCN GCN GGN GGC AT-3'

ACGATGCAGCTTGCTCCGGCTCGGACCCTAGTAACGGCCGCCAGTGTGCTGGAATTCGCCCTTATGGTTGA
AGACGCGCATGGCATCGGCGTGCGCGGCGAACAAGGCCGCGGCAGCTGCTGGCAGCAGGGCGTGCGCCCTG
AACTGCTGGTGGCGACCTTCGGCAAGGCGTTCGGCGTCAGCGGCGCGGCGGTGCTGTGCGATGAGGCGACC
GCCGAGTATCTGCTGCAGTTCGCCCGCCATCTGATCTACAGCACCGCGATGCCCCCAGCACAGGCAAGGGC
GAATTCTGCAGATATCCATCACACTGGCGGCCGCTCGAGCAGGCATCTAGAGGGCCCAATTCGCCCTATAG
TGAGTCGTATTACAATTCACTGGCCGTCGTTTTACAACGTCGTGACTGGGAAAACCCTGGCGTTACCCAAC
TTAATCGCCTTGCAGCACATCCCCCTTTCGCCAGCTGGCGTAATAGCGAAGAGGCCCGCACCGATCGCCCT
TCCCAACAGTTGCGCAGCCTGAATGGCGAATGGACGCGCCCTGTAGCGGCGCATTAAGCGCGGCGGGTGTG
GTGGTT

| Sequences producing significant alignments: | | Score (Bits) | E Value |
|---|---|---|---|
| sp\|P36570\|BIOF_SERMA | 8-amino-7-oxononanoate synthase (AONS) (... | 150 | 4e-35 |
| gb\|AAG60556.1\|AF250768_1 | KAPA-synthetase [uncultured bacterium p | 148 | 2e-34 |
| ref\|ZP_01535733.1\| | 8-amino-7-oxononanoate synthase [Serratia ... | 147 | 3e-34 |
| emb\|CAA71575.1\| | fused-ccdB [Cloning vector pZeRO-2T] | 136 | 8e-31 |
| gb\|AAS78488.1\| | LacZ alpha & ccdB lethal fusion protein [Cloni... | 136 | 8e-31 |
| ref\|NP_928781.1\| | 8-amino-7-oxononanoate synthase [Photorhabdu... | 135 | 1e-30 |
| ref\|YP_215780.1\| | 8-amino-7-oxononanoate synthase [Salmonella ... | 134 | 4e-30 |
| ref\|YP_151178.1\| | 8-amino-7-oxononanoate synthase [Salmonella ... | 134 | 4e-30 |
| ref\|NP_455337.1\| | 8-amino-7-oxononanoate synthase [Salmonella ... | 134 | 4e-30 |
| ref\|YP_069718.1\| | 8-amino-7-oxononanoate synthase [Yersinia ps... | 133 | 5e-30 |
| ref\|ZP_01497420.1\| | hypothetical protein YpseI_02002584 [Yersi... | 133 | 5e-30 |
| ref\|ZP_00829476.1\| | COG0156: 7-keto-8-aminopelargonate synthet... | 133 | 7e-30 |
| gb\|ABR76237.1\| | 8-amino-7-oxononanoate synthase [Klebsiella pn... | 132 | 9e-30 |
| ref\|ZP_00832233.1\| | COG0156: 7-keto-8-aminopelargonate synthet... | 132 | 1e-29 |
| emb\|CAA67127.1\| | ccdB [synthetic construct] | 132 | 1e-29 |
| ref\|NP_404761.1\| | 8-amino-7-oxononanoate synthase [Yersinia pe... | 132 | 1e-29 |
| ref\|NP_459773.1\| | 8-amino-7-oxononanoate synthase [Salmonella ... | 130 | 3e-29 |
| gb\|AAA53120.1\| | beta-galactosidase-complementation protein | 116 | 7e-29 |
| emb\|CAA75108.1\| | beta-galactosidase [Phagemid cloning vector pTZ1 | 116 | 7e-29 |
| ref\|YP_050915.1\| | 8-amino-7-oxononanoate synthase [Erwinia car... | 129 | 7e-29 |
| gb\|AAR19394.1\| | modified lacZ' [Cloning vector pTA6] | 114 | 3e-28 |
| gb\|AAF61634.1\| | lacZ [Cloning vector pTG8] | 114 | 3e-28 |
| gb\|ABI35978.1\| | LacZ alpha peptide [Cloning vector pYESW29'] | 114 | 3e-28 |
| gb\|AAP44986.1\| | LacZalpha [Shuttle vector pNG168] | 114 | 3e-28 |
| gb\|AAG34547.1\| | beta-galactosidase [Cloning vector pUG7] | 114 | 3e-28 |
| gb\|AAG34526.1\| | beta-galactosidase [Cloning vector pUG26] | 114 | 3e-28 |
| gb\|AAA53118.1\| | beta-galactosidase-complementation protein | 114 | 3e-28 |
| ref\|YP_539791.1\| | 8-amino-7-oxononanoate synthase [Escherichia... | 127 | 4e-28 |
| gb\|AAG60580.1\|AF250776_4 | KAPA synthetase BioF [uncultured bacter | 127 | 4e-28 |

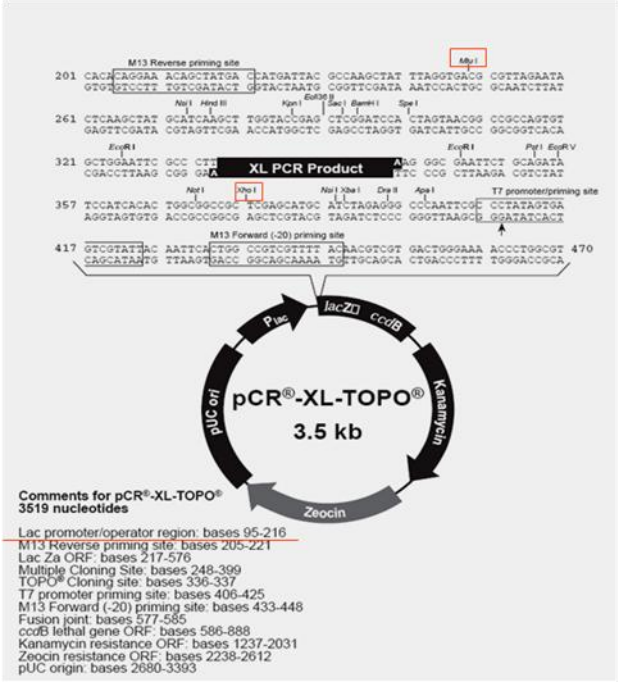
| | | | |
|---|---|---|---|
| ref\|YP_402501.1\| | 8-amino-7-oxononanoate synthase [Shigella dy... | 127 | 4e-28 |
| ref\|YP_309739.1\| | 8-amino-7-oxononanoate synthase [Shigella so... | 127 | 4e-28 |
| ref\|NP_706654.1\| | 8-amino-7-oxononanoate synthase [Shigella fl... | 127 | 4e-28 |
| ref\|NP_752786.1\| | 8-amino-7-oxononanoate synthase [Escherichia... | 127 | 4e-28 |
| ref\|NP_415297.1\| | 8-amino-7-oxononanoate synthase [Escherichia... | 127 | 4e-28 |
| pdb\|1DJ9\|A | Chain A, Crystal Structure Of 8-Amino-7-Oxonanoate... | 127 | 4e-28 |
| ref\|ZP_00721166.1\| | COG0156: 7-keto-8-aminopelargonate synthet... | 127 | 4e-28 |
| ref\|ZP_00734064.1\| | COG0156: 7-keto-8-aminopelargonate synthet... | 127 | 4e-28 |
| ref\|ZP_00715006.1\| | COG0156: 7-keto-8-aminopelargonate synthet... | 127 | 4e-28 |
| ref\|ZP_00703084.1\| | COG0156: 7-keto-8-aminopelargonate synthet... | 127 | 4e-28 |
| ref\|ZP_00709065.1\| | COG0156: 7-keto-8-aminopelargonate synthet... | 127 | 4e-28 |
| ref\|ZP_00704998.1\| | COG0156: 7-keto-8-aminopelargonate synthet... | 127 | 4e-28 |
| ref\|ZP_00821419.1\| | COG0156: 7-keto-8-aminopelargonate synthet... | 127 | 5e-28 |
| gb\|AAS77680.1\| | LacZ-alpha [Shuttle vector pELS100] >gb\|AAS776... | 126 | 6e-28 |
| gb\|AAS77687.2\| | LacZ-alpha [Shuttle vector pLPV111] | 126 | 6e-28 |
| gb\|AAD31805.1\|AF128862_1 | LacZ [Cloning vector pHIND2.2] | 126 | 6e-28 |
| emb\|CAB76939.1\| | alpha-peptide [Cloning vector pPW78] | 126 | 6e-28 |
| ref\|NP_286539.1\| | 8-amino-7-oxononanoate synthase [Escherichia... | 126 | 8e-28 |
| ref\|NP_308881.1\| | 8-amino-7-oxononanoate synthase [Escherichia... | 126 | 8e-28 |
| ref\|ZP_00698065.1\| | COG0156: 7-keto-8-aminopelargonate synthet... | 126 | 8e-28 |
| ref\|ZP_00824683.1\| | COG0156: 7-keto-8-aminopelargonate synthet... | 125 | 1e-27 |
| gb\|AAA56741.1\| | lacZ alpha peptide | 114 | 5e-27 |
| gb\|AAA56738.1\| | lacZ alpha peptide | 114 | 5e-27 |
| ref\|YP_001007094.1\| | 8-amino-7-oxononanoate synthase [Yersinia... | 122 | 1e-26 |
| gb\|AAF86670.1\|AF178449_1 | beta-galactosidase alpha peptide [In... | 122 | 2e-26 |
| gb\|AAF86672.1\|AF178450_1 | beta-galactosidase alpha peptide [In... | 122 | 2e-26 |
| gb\|AAG53590.1\|AF248314_5 | 7-KAPA synthetase [uncultured bacterium | 119 | 1e-25 |
| sp\|Q47829\|BIOF_ENTAG | 8-amino-7-oxononanoate synthase (AONS) (... | 119 | 1e-25 |
| ref\|YP_001175999.1\| | 8-amino-7-oxononanoate synthase [Enteroba... | 117 | 4e-25 |
| gb\|EDM59207.1\| | 8-amino-7-oxononanoate synthase [Vibrio paraha... | 115 | 1e-24 |
| ref\|ZP_01222497.1\| | 8-amino-7-oxononanoate synthase [Photobact... | 115 | 1e-24 |
| ref\|YP_130515.1\| | 8-amino-7-oxononanoate synthase [Photobacter... | 115 | 1e-24 |
| ref\|NP_797493.1\| | 8-amino-7-oxononanoate synthase [Vibrio para... | 115 | 1e-24 |
| gb\|AAK73367.1\|AF327711_1 | LacZ [Cloning vector pDN19] >emb\|CAA... | 114 | 2e-24 |
| ref\|ZP_01160448.1\| | 8-amino-7-oxononanoate synthase [Photobact... | 114 | 2e-24 |
| gb\|EDL74524.1\| | 8-amino-7-oxononanoate synthase [Vibrio cholerae | 114 | 3e-24 |
| ref\|ZP_01235386.1\| | 8-amino-7-oxononanoate synthase [Vibrio an... | 114 | 4e-24 |
| gb\|EDN14556.1\| | 8-amino-7-oxononanoate synthase [Vibrio cholerae | 113 | 5e-24 |
| gb\|EDM52798.1\| | 8-amino-7-oxononanoate synthase [Vibrio cholerae | 113 | 5e-24 |
| gb\|EAY41082.1\| | 8-amino-7-oxononanoate synthase [Vibrio cholerae | 113 | 5e-24 |
| gb\|AAA75561.1\| | LacOPZ-alpha peptide from pUC9; putative [unid... | 113 | 5e-24 |
| ref\|NP_230758.1\| | 8-amino-7-oxononanoate synthase [Vibrio chol... | 113 | 5e-24 |
| emb\|CAC14450.1\| | alfa-peptide [Cloning vector pBPSCat2] >emb\|C... | 113 | 5e-24 |
| ref\|ZP_01478821.1\| | hypothetical protein VchoM_02001980 [Vibri... | 113 | 5e-24 |
| gb\|EAY34702.1\| | 8-amino-7-oxononanoate synthase [Vibrio cholerae | 113 | 7e-24 |
| gb\|AAC43252.1\| | lacZ alpha peptide [Cloning vector Pk184] | 113 | 7e-24 |
| gb\|AAC48874.1\| | lacZ-alpha [Cloning vector pAL-Z] >gb\|AAC53715... | 113 | 7e-24 |
| ref\|YP_454584.1\| | 8-amino-7-oxononanoate synthase [Sodalis glo... | 113 | 7e-24 |

**Figure C.4: Sequence Alignment Data for Biotin from Hon6**

# APPENDIX D

Supplementary data of Chapter V

## Topo XL vector




MluI~MluI: ~80bp

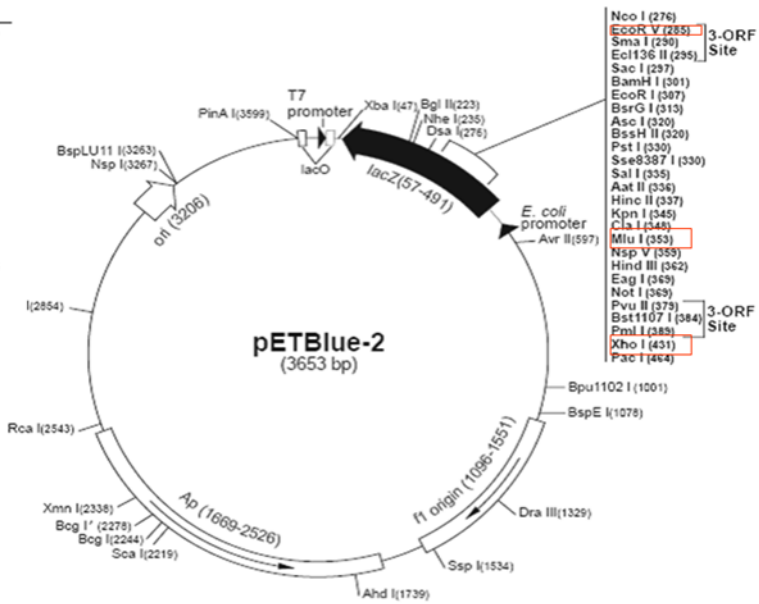
MluI~XhoI: ~8.9Kb

XhoI~XhoI: ~40bp

After digestion with MluI~XhoI:
TOPO: 3.38kb

## pETBlue-2 vector

| pETBlue-2 sequence landmarks | |
|---|---|
| *lac* operator | 3606–3625 |
| T7 promoter | 1–17 |
| *lac* operator | 22–42 |
| T7 transcription start | 18 |
| multiple cloning region (*Nco* I–*Pac* I) | 276–467 |
| His•Tag® coding sequence | 437–454 |
| HSV•Tag® coding sequence | 395–430 |
| *lacZ* start codon | 491 |
| *lacZ* α-peptide ORF | 57–491 |
| *E. coli* promoter | 541–569 |
| f1 origin | 1096–1551 |
| *bla* coding sequence | 1669–2526 |
| pUC origin | 3206 |




EcoRV~XhoI=146 bp

MluI~XhoI=78 bp

**pCC1FOS vector**

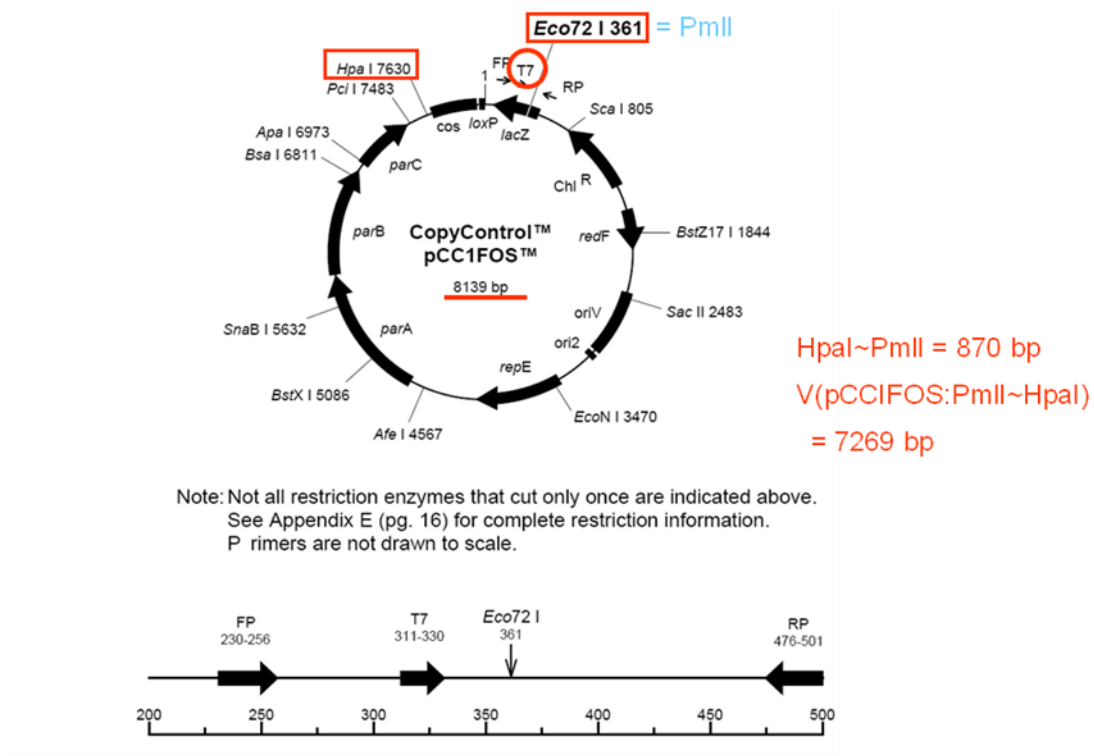



**Figure D.1: Vector Map**

**Reaction condition of Synthesis of linker**

PNK 10X buffer 1µl
2mmol ATP 1µl
Oligo 2µl
PNK 0.1µl
Deionized water 5.9µl
Total 10µl → incubation at 37°C for 1 hr then mix all oligo → incubation at 95°C for 5 min cooling it down to R.T.

**Ligation condition with vector and linker**

Vector 2µl
10mmol ATP 1µl
Oligo linker 5µl
T4 DNA ligase 1µl
10X buffer 1µl
Total 10µl -> incubation at 16°C for overnight.

**Expected PCR product size**

VioE-R - VioE-F 550bp
VioD-R - VioD-F 1060bp
VioC-R - VioC-F 1230bp
VioB-R - VioB-F 2920bp
VioA-R - VioA-F 1190bp
VioA-R – 3275R 2130bp
3269-F - VioE-F 1700bp
VioE-R - VioD-F 1650bp
VioD-R - VioC-F 2350bp
VioC-R - VioB-F 4230bp
VioB-RX – VioA-F 1865bp
VioB-RX – 3275R 2840bp
VioB-R - VioA-F 4270bp
VioB-R – 3275R 5230bp
3269-F - VioC-F 4400bp
3369-F - VioBCF 4550bp
VioB-R – 3275RLR 4900bp (Seq: VioBCR – 3275RLS)
3369LFR - VioBCF 4450bp (Seq: VioCF – 3269LFS)
3275RLR-3269LFR 9100bp (Seq: 3269LFS-3275RLS)
3275RLR-3269RLF 8920bp (Seq: 3269LFS-3275RLS)

**PCR conditions**:
VioA: 1 µl of C. Violacein genomic DNA or PCR product, 1 µl of VioA F primer, 1 µl of VioA R primer, 0.5 µl of Taq polymerase, 5 µl of 10× thermal solution, 0.5 µl of dNTP, 2.5 µl of glycerol, and 38.5 µl of deionized water.

VioC: 1 µl of C. Violacein genomic DNA or PCR product, 1 µl of VioC F primer, 1 µl of VioC R primer, 0.5 µl of Taq polymerase, 5 µl of 10× thermal solution, 0.5 µl of dNTP, 2.5 µl of DMSO and 38.5 µl of deionized water.

VioD: 1 µl of C. Violacein genomic DNA or PCR product, 1 µl of VioD F primer, 1 µl of VioD R primer, 0.5 µl of Taq polymerase, 5 µl of 10× thermal solution, 0.5 µl of dNTP, 2.5 µl of DMSO and 38.5 µl of deionized water.

VioE: 1 µl of C. Violacein genomic DNA or PCR product, 1 µl of VioE F primer, 1 µl of VioE R primer, 0.5 µl of Taq polymerase, 5 µl of 10× thermal solution, 0.5 µl of dNTP, 2.5 µl of DMSO and 38.5 µl of deionized water.

TPP+P+Riboswitch: 1 µl of DH10B plasmid DNA, 1 µl of TPP+Riboswitch F primer, 1 µl of TPP+Riboswitch R primer, 0.5 µl of Taq polymerase, 5 µl of 10× thermal solution, 0.5 µl of dNTP, 2.5 µl of glycerol, and 38.5 µl of deionized water.

BLG: 1 µl of pETBlue plasmid DNA, 0.4µl of BLGThiCRibo(f) SpeI F primer, 0.4 µl of BLGThiCRibo(R) XhoI R primer, 0.4 µl of phusion polymerase, 4 µl of 5× Buffer, 0.5 µl of dNTP, and 13.3 µl of deionized water.

**Figure D.2: Reaction conditions and PCR conditions**

**Riboswitch sequence**

sequence from Cheah *et al*. (*165*)

TAATTTCTTGTCGGAGTGCCTTAACTGGCTGAGACCGTTTATTCGGGATCCGCGGAACCTGATCAGGCTA
ATACCTGCGAAGGGAACAAGAGTTA

**TPP riboswitch aptamer sequence (Gene/Homolog:ThiC, ID:Eco2)**

E.coli str. K12 substr.DH10B (CP000948)

thiamin phosphate synthase (thiamin phosphate pyrophospho...

thiamin (pyrimidine moiety) biosynthesis protein

4291923-4293818

Sequence was found comparison of TPP riboswitch sequence from Cheah et al (*165*)

```
4291921 catcacgctt cctccttacg caggtagatt tctccgcctc tggcacggaa gttctccgac
4291981 atatccgcca ttcccatttc aatagtttgc gtggcggcgt aatcacgcac ttcctggctg
4292041 attttcatcg agcagaattt cggcccacac atggagcaaa aatgggcgac tttacctgac
4292101 tcttgcggca gggtttcatc gtgataagcg cgggcggtaa acgggtcgag ggccagatta
4292161 aactggtctt cccagcgaaa ttcgaagcgg gctttcgaca tggcgttatc gcgaatttgc
4292221 gcgcccggat gccctttcgc caggtcagcg gcgtgggcag caatcttata ggtgataagc
4292281 ccctgcttaa catcttcttt attgggcaga cccagatgct cttttggcgt tacgtaacag
4292341 agcatcgcgc agccaaacca gccaatcatc gccgcaccaa tccccgacgt gaagtggtca
4292401 tagcccggcg caatatcggt agttagcggc cccagagtgt aaaacggcgc ttcgtggcag
4292461 tgctctaact cctcggtcat attgcggcgg atcatctgca tcggcacgtg gcctgggcct
4292521 tcaatcatca cctgcacgtc atattcccag gcaattttgg tcagttcgcc cagcgtatgc
4292581 agctcggcaa actgcgcttc atcgttggcg tcctgaatag aaccggggcg cagaccgtcg
4292641 cccagcgaca gcgaaacgtc ataagcggca cagatttcac aaatttcgcg gaagtgttga
4292701 tagaggaaat tttcctgatg atgggagagg caccatttcg ccataatcga accgccgcga
4292761 gagacgatac cggtcaggcg tttcgcggtc atcggcacat agcgcagcag tacgcccgca
4292821 tggatagtga agtaatccac accttgctcg gcctgttcca gcagcgtgtc gcggaacgct
4292881 tcccaggtaa gatcttcggc gatcccgtta accttctcca gcgcctggta gatcggcact
4292941 gtaccgatcg gcaccgggct gttacgcaaa atccactcgc gggtttcgtg aatatagcga
4293001 ccggtggaga gatccatcac cgtatccgct ccccagcgcg tggaccatac cagcttttcc
4293061 acttcttctt cgatggaaga ggtgaccgcc gagttgccga tattggcgtt aacttttacc
4293121 aggaaattgc gaccaataat catcggctcc gattccggat gattaatgtt ggccgggata
4293181 atcgcacgtc cggcagcaac ttcatcacgg acaaattccg cagtgatatt ttccggcaga
4293241 tgtgcgccaa agctcattcc cggatgctgg tggcgtaaaa cctcgctacg gatgcgctcg
4293301 cggcccatat tctcgcggat ggcgatgaat tccatttccg gcgtgatgat gccctggcgg
4293361 gcgtagtgca gttgggtgac acggcgtcct gctttggcgc gttttggtgt tagtacgccg
4293421 ctaaaacgca gttcgtcgag gccatcatct gccagccgcg ctttagtgta atcggaactg
4293481 cgcacggtaa gttcttcggt atcgccgcgc gcatcgatcc acggctggcg tagttttgcc
4293541 agcccttgct gcacgttaat ggcaatctgc ggatcaccat acgggccgga ggtgtcgtag
4293601 accggaatcg cttcgttttc ttcgtactgc ggctgttctt tgctaccgcc aattagcgtc
4293661 gggctaagct ggatctcacg catcggcacg cgcacgccgg ggtgtgtgcc agtgatataa
4293721 atgcgttttg agttgggaaa ggcggtgcct tccagggtgt cgataaaatg ttgggcccgg
4293781 gcgcgttgtt cgcggcgggt cagttttgtt gcagacatag ctcattccaa aaagttaagg
4293841 acgtggcttg tcagacgacg gatgaagcaa gagacgatcg ccgcaggggc gatgcgatag
4293901 cagattaact cttgttccct tcgcaggtat tagcctgatc aggttccgcg gatcccgaat
4293961 aaacggtctc agccagttaa ggcactccga caagaaatta gccccgcaaa tggggcattg
```

**Complementary DNA Sequence: Bold and underlined sequences are Riboswitch part.**

```
LOCUS       nameless                2100 bp
ORIGIN
    1 CAATGCCCCA TTTGCGGGGC TAATTTCTTG TCGGAGTGCC TTAACTGGCT GAGACCGTTT
   61 ATTCGGGATC CGCGGAACCT GATCAGGCTA ATACCTGCGA AGGGAACAAG AGTTAATCTG
  121 CTATCGCATC GCCCCTGCGG CGATCGTCTC TTGCTTCATC CGTCGTCTGA CAAGCCACGT
  181 CCTTAACTTT TTGGAATGAG CTATGTCTGC AACAAAACTG ACCCGCCGCG AACAACGCGC
  241 CCGGGCCCAA CATTTTATCG ACACCCTGGA AGGCACCGCC TTTCCCAACT CAAAACGCAT
  301 TTATATCACT GGCACACACC CCGGCGTGCG CGTGCCGATG CGTGAGATCC AGCTTAGCCC
  361 GACGCTAATT GGCGGTAGCA AAGAACAGCC GCAGTACGAA GAAAACGAAG CGATTCCGGT
  421 CTACGACACC TCCGGCCCGT ATGGTGATCC GCAGATTGCC ATTAACGTGC AGCAAGGGCT
  481 GGCAAAACTA CGCCAGCCGT GGATCGATGC GCGCGGCGAT ACCGAAGAAC TTACCGTGCG
  541 CAGTTCCGAT TACACTAAAG CGCGGCTGGC AGATGATGGC CTCGACGAAC TGCGTTTTAG
  601 CGGCGTACTA ACACCAAAAC GCGCCAAAGC AGGACGCCGT GTCACCCAAC TGCACTACGC
  661 CCGCCAGGGC ATCATCACGC CGGAAATGGA ATTCATCGCC ATCCGCGAGA ATATGGGCCG
```

```
 721 CGAGCGCATC CGTAGCGAGG TTTTACGCCA CCAGCATCCG GGAATGAGCT TTGGCGCACA
 781 TCTGCCGGAA AATATCACTG CGGAATTTGT CCGTGATGAA GTTGCTGCCG GACGTGCGAT
 841 TATCCCGGCC AACATTAATC ATCCGGAATC GGAGCCGATG ATTATTGGTC GCAATTTCCT
 901 GGTAAAAGTT AACGCCAATA TCGGCAACTC GGCGGTCACC TCTTCCATCG AAGAAGAAGT
 961 GGAAAAGCTG GTATGGTCCA CGCGCTGGGG AGCGGATACG GTGATGGATC TCTCCACCGG
1021 TCGCTATATT CACGAAACCC GCGAGTGGAT TTTGCGTAAC AGCCCGGTGC CGATCGGTAC
1081 AGTGCCGATC TACCAGGCGC TGGAGAAGGT TAACGGGATC GCCGAAGATC TTACCTGGGA
1141 AGCGTTCCGC GACACGCTGC TGGAACAGGC CGAGCAAGGT GTGGATTACT TCACTATCCA
1201 TGCGGGCGTA CTGCTGCGCT ATGTGCCGAT GACCGCGAAA CGCCTGACCG GTATCGTCTC
1261 TCGCGGCGGT TCGATTATGG CGAAATGGTG CCTCTCCCAT CATCAGGAAA ATTTCCTCTA
1321 TCAACACTTC CGCGAAATTT GTGAAATCTG TGCCGCTTAT GACGTTTCGC TGTCGCTGGG
1381 CGACGGTCTG CGCCCCGGTT CTATTCAGGA CGCCAACGAT GAAGCGCAGT TTGCCGAGCT
1441 GCATACGCTG GGCGAACTGA CCAAAATTGC CTGGGAATAT GACGTGCAGG TGATGATTGA
1501 AGGCCCAGGC CACGTGCCGA TGCAGATGAT CCGCCGCAAT ATGACCGAGG AGTTAGAGCA
1561 CTGCCACGAA GCGCCGTTTT ACACTCTGGG GCCGCTAACT ACCGATATTG CGCCGGGCTA
1621 TGACCACTTC ACGTCGGGGA TTGGTGCGGC GATGATTGGC TGGTTTGGCT GCGCGATGCT
1681 CTGTTACGTA ACGCCAAAAG AGCATCTGGG TCTGCCCAAT AAAGAAGATG TTAAGCAGGG
1741 GCTTATCACC TATAAGATTG CTGCCCACGC CGCTGACCTG GCGAAAGGGC ATCCGGGCGC
1801 GCAAATTCGC GATAACGCCA TGTCGAAAGC CCGCTTCGAA TTTCGCTGGG AAGACCAGTT
1861 TAATCTGGCC CTCGACCCGT TTACCGCCCG CGCTTATCAC GATGAAACCC TGCCGCAAGA
1921 GTCAGGTAAA GTCGCCCATT TTTGCTCCAT GTGTGGGCCG AAATTCTGCT CGATGAAAAT
1981 CAGCCAGGAA GTGCGTGATT ACGCCGCCAC GCAAACTATT GAAATGGGAA TGGCGGATAT
2041 GTCGGAGAAC TTCCGTGCCA GAGGCGGAGA AATCTACCTG CGTAAGGAGG AAGCGTGATG
```

## Thiamin biosynthesis protein ThiC sequence

**>gi|169887498:4291721-4294220 Escherichia coli str. K12 substr. DH10B, complete genome**

```
ATAAACAATCGCGCGTTATAGCGGCGGCCCAGCGCAATTGCCGCCACGACATCGGCTTCCACCTCTTCAT
CGCGCCGATCTTTGATGCGTAGCTGGAGAGTACGTACGCCTGCATCCAACAGACGTTCGATCCACTGTAC
GCTGTCCACCACCGGGTACAGTCCTGAACGAAAAGGTACAGGAGGAAAATCAGGCTGATACATCACGCTT
CCTCCTTACGCAGGTAGATTTCTCCGCCTCTGGCACGGAAGTTCTCCGACATATCCGCCATTCCCATTTC
AATAGTTTGCGTGGCGGCGTAATCACGCACTTCCTGGCTGATTTTCATCGAGCAGAATTTCGGCCCACAC
ATGGAGCAAAAATGGGCGACTTTACCTGACTCTTGCGGCAGGGTTTCATCGTGATAAGCGCGGGCGGTAA
ACGGGTCGAGGGCCAGATTAAACTGGTCTTCCCAGCGAAATTCGAAGCGGGCTTTCGACATGGCGTTATC
GCGAATTTGCGCGCCCGGATGCCCTTTCGCCAGGTCAGCGGCGTGGGCAGCAATCTTATAGGTGATAAGC
CCCTGCTTAACATCTTCTTTATTGGGCAGACCCAGATGCTCTTTTGGCGTTACGTAACAGAGCATCGCGC
AGCCAAACCAGCCAATCATCGCCGCACCAATCCCCGACGTGAAGTGGTCATAGCCCGGCGCAATATCGGT
AGTTAGCGGCCCCAGAGTGTAAAACGGCGCTTCGTGGCAGTGCTCTAACTCCTCGGTCATATTGCGGCGG
ATCATCTGCATCGGCACGTGGCCTGGGCCTTCAATCATCACCTGCACGTCATATTCCCAGGCAATTTTGG
TCAGTTCGCCCAGCGTATGCAGCTCGGCAAACTGCGCTTCATCGTTGGCGTCCTGAATAGAACCGGGGCG
CAGACCGTCGCCCAGCGACAGCGAAACGTCATAAGCGGCACAGATTTCACAAATTTCGCGGAAGTGTTGA
TAGAGGAAATTTTCCTGATGATGGGAGAGGCACCATTTCGCCATAATCGAACCGCCGCGAGAGACGATAC
CGGTCAGGCGTTTCGCGGTCATCGGCACATAGCGCAGCAGTACGCCCGCATGGATAGTGAAGTAATCCAC
ACCTTGCTCGGCCTGTTCCAGCAGCGTGTCGCGGAACGCTTCCCAGGTAAGATCTTCGGCGATCCCGTTA
ACCTTCTCCAGCGCCTGGTAGATCGGCACTGTACCGATCGGCACCGGGCTGTTACGCAAAATCCACTCGC
GGGTTTCGTGAATATAGCGACCGGTGGAGAGATCCATCACCGTATCCGCTCCCCAGCGCGTGGACCATAC
CAGCTTTTCCACTTCTTCTTCGATGGAAGAGGTGACCGCCGAGTTGCCGATATTGGCGTTAACTTTTACC
AGGAAATTGCGACCAATAATCATCGGCTCCGATTCCGGATGATTAATGTTGGCCGGGATAATCGCACGTC
CGGCAGCAACTTCATCACGGACAAATTCCGCAGTGATATTTTCCGGCAGATGTGCGCCAAAGCTCATTCC
CGGATGCTGGTGGCGTAAAACCTCGCTACGGATGCGCTCGCGGCCCATATTCTCGCGGATGGCGATGAAT
TCCATTTCCGGCGTGATGATGCCCTGGCGGGCGTAGTGCAGTTGGGTGACACGGCGTCCTGCTTTGGCGC
GTTTTGGTGTTAGTACGCCGCTAAAACGCAGTTCGTCGAGGCCATCATCTGCCAGCCGCGCTTTAGTGTA
ATCGGAACTGCGCACGGTAAGTTCTTCGGTATCGCCGCGCGCATCGATCCACGGCTGGCGTAGTTTTGCC
AGCCCTTGCTGCACGTTAATGGCAATCTGCGGATCACCATACGGGCCGGAGGTGTCGTAGACCGGAATCG
CTTCGTTTTCTTCGTACTGCGGCTGTTCTTTGCTACCGCCAATTAGCGTCGGGCTAAGCTGGATCTCACG
CATCGGCACGCGCACGCCGGGGTGTGTGCCAGTGATATAAATGCGTTTTGAGTTGGGAAAGGCGGTGCCT
```

TCCAGGGTGTCGATAAAATGTTGGGCCCGGGCGCGTTGTTCGCGGCGGGTCAGTTTTGTTGCAGACATAG
CTCATTCCAAAAAGTTAAGGACGTGGCTTGTCAGACGACGGATGAAGCAAGAGACGATCGCCGCAGGGGC
GATGCGATAGCAGATTAACTCTTGTTCCCTTCGCAGGTATTAGCCTGATCAGGTTCCGCGGATCCCGAAT
AAACGGTCTCAGCCAGTTAAGGCACTCCGACAAGAAATTAGCCCCGCAAATGGGGCATTGAATGTAAATT
ACGCGTTAACAGCGCAGAACTCAAGCAGGATGTTTGACGCGGGCGGCGTCAAGCACCAGCAGAATCAGCT
TATCTTCCAGCACAAAGCGCGCTTCCAGCGCCTCGCCGATGTCAGATAAAACCTGTTGAAACTCAAGGTA
ATTATCATGATCGATGGCGGTTTCCAGACTGGAATCGTAGTAATCCATAA

**4291721-4294220 Escherichia coli str. K12 substr. DH10B**
**Complementary DNA sequence: Blue: -35 and -10 promoter region, Red: riboswitch, Pink: start codon for transcription, Green: transcriptional part for thiamine biosynthetic protein**

TTATGGATTACTACGATTCCAGTCTGGAAACCGCCATCGATCATGATAATTACCTTGAGTTTCAACAGGTT
TTATCTGACATCGGCGAGGCGCTGGAAGCGCGCTTTGTGCTGGAAGATAAGCTGATTCTGCTGGTGCTTGA
CGCCGCCCGCGTCAAACATCCTGCTTGAGTTCTGCGCTGTTAACGCGTAATTTACATTCAATGCCCCATTT
GCGGGGCTAATTTCTTGTCGGAGTGCCTTAACTGGCTGAGACCGTTTATTCGGGATCCGCGGAACCTGATC
AGGCTAATACCTGCGAAGGGAACAAGAGTTAATCTGCTATCGCATCGCCCCTGCGGCGATCGTCTCTTGCT
TCATCCGTCGTCTGACAAGCCACGTCCTTAACTTTTTGGAATGAGCT*ATG*TCTGCAACAAAACTGACCCGC
CGCGAACAACGCGCCCGGGCCCAACATTTTATCGACACCCTGGAAGGCACCGCCTTTCCCAACTCAAAACG
CATTTATATCACTGGCACACACCCCGGCGTGCGCGTGCCGATGCGTGAGATCCAGCTTAGCCCGACGCTAA
TTGGCGGTAGCAAAGAACAGCCGCAGTACGAAGAAAACGAAGCGATTCCGGTCTACGACACCTCCGGCCCG
TATGGTGATCCGCAGATTGCCATTAACGTGCAGCAAGGGCTGGCAAAACTACGCCAGCCGTGGATCGATGC
GCGCGGCGATACCGAAGAACTTACCGTGCGCAGTTCCGATTACACTAAAGCGCGGCTGGCAGATGATGGCC
TCGACGAACTGCGTTTTAGCGGCGTACTAACACCAAAACGCGCCAAAGCAGGACGCCGTGTCACCCAACTG
CACTACGCCCGCCAGGGCATCATCACGCCGGAAATGGAATTCATCGCCATCCGCGAGAATATGGGCCGCGA
GCGCATCCGTAGCGAGGTTTTACGCCACCAGCATCCGGGAATGAGCTTTGGCGCACATCTGCCGGAAAATA
TCACTGCGGAATTTGTCCGTGATGAAGTTGCTGCCGGACGTGCGATTATCCCGGCCAACATTAATCATCCG
GAATCGGAGCCGATGATTATTGGTCGCAATTTCCTGGTAAAAGTTAACGCCAATATCGGCAACTCGGCGGT
CACCTCTTCCATCGAAGAAGAAGTGGAAAAGCTGGTATGGTCCACGCGCTGGGGAGCGGATACGGTGATGG
ATCTCTCCACCGGTCGCTATATTCACGAAACCCGCGAGTGGATTTTGCGTAACAGCCCGGTGCCGATCGGT
ACAGTGCCGATCTACCAGGCGCTGGAGAAGGTTAACGGGATCGCCGAAGATCTTACCTGGGAAGCGTTCCG
CGACACGCTGCTGGAACAGGCCGAGCAAGGTGTGGATTACTTCACTATCCATGCGGGCGTACTGCTGCGCT
ATGTGCCGATGACCGCGAAACGCCTGACCGGTATCGTCTCTCGCGGCGGTTCGATTATGGCGAAATGGTGC
CTCTCCCATCATCAGGAAAATTTCCTCTATCAACACTTCCGCGAAATTTGTGAAATCTGTGCCGCTTATGA
CGTTTCGCTGTCGCTGGGCGACGGTCTGCGCCCCGGTTCTATTCAGGACGCCAACGATGAAGCGCAGTTTG
CCGAGCTGCATACGCTGGGCGAACTGACCAAAATTGCCTGGGAATATGACGTGCAGGTGATGATTGAAGGC
CCAGGCCACGTGCCGATGCAGATGATCCGCCGCAATATGACCGAGGAGTTAGAGCACTGCCACGAAGCGCC
GTTTTACACTCTGGGGCCGCTAACTACCGATATTGCGCCGGGCTATGACCACTTCACGTCGGGGATTGGTG
CGGCGATGATTGGCTGGTTTGGCTGCGCGATGCTCTGTTACGTAACGCCAAAAGAGCATCTGGGTCTGCCC
AATAAAGAAGATGTTAAGCAGGGGCTTATCACCTATAAGATTGCTGCCCACGCCGCTGACCTGGCGAAAGG
GCATCCGGGCGCGCAAATTCGCGATAACGCCATGTCGAAAGCCCGCTTCGAATTTCGCTGGGAAGACCAGT
TTAATCTGGCCCTCGACCCGTTTACCGCCCGCGCTTATCACGATGAAACCCTGCCGCAAGAGTCAGGTAAA
GTCGCCCATTTTTGCTCCATGTGTGGGCCGAAATTCTGCTCGATGAAAATCAGCCAGGAAGTGCGTGATTA
CGCCGCCACGCAAACTATTGAAATGGGAATGGCGGATATGTCGGAGAACTTCCGTGCCAGAGGCGGAGAAA
TCTACCTGCGTAAGGAGGAAGCGTGATGTATCAGCCTGATTTTCCTCCTGTACCTTTTCGTTCAGGACTGT
ACCCGGTGGTGGACAGCGTACAGTGGATCGAACGTCTGTTGGATGCAGGCGTACGTACTCTCCAGCTACGC
ATCAAAGATCGGCGCGATGAAGAGGTGGAAGCCGATGTCGTGGCGGCAATTGCGCTGGGCCGCCGCTATAA
CGCGCGATTGTTTAT

**pBAD33 Sequence**

```
source          1..5376
ORIGIN      BASE COUNT     1355 a   1365 c   1276 g   1380 t
   1 atcgatgcat aatgtgcctg tcaaatggac gaagcaggga ttctgcaaac cctatgctac
  61 tccgtcaagc cgtcaattgt ctgattcgtt accaattatg acaacttgac ggctacatca
 121 ttcacttttt cttcacaacc ggcacggaac tcgctcgggc tggccccggt gcatttttta
 181 aatacccgcg agaaatagag ttgatcgtca aaaccaacat tgcgaccgac ggtggcgata
 241 ggcatccggg tggtgctcaa aagcagcttc gcctggctga tacgttggtc ctcgcgccag
 301 cttaagacgc taatccctaa ctgctggcgg aaaagatgtg acagacgcga cggcgacaag
 361 caaacatgct gtgcgacgct ggcgatatca aaattgctgt ctgccaggtg atcgctgatg
 421 tactgacaag cctcgcgtac ccgattatcc atcggtggat ggagcgactc gttaatcgct
 481 tccatgcgcc gcagtaacaa ttgctcaagc agatttatcg ccagcagctc cgaatagcgc
 541 ccttcccctt gcccggcgtt aatgatttgc ccaaacaggt cgctgaaatg cggctggtgc
 601 gcttcatccg ggcgaaagaa ccccgtattg gcaaatattg acggccagtt aagccattca
 661 tgccagtagg cgcgcggacg aaagtaaacc cactggtgat accattcgcg agcctccgga
 721 tgacgaccgt agtgatgaat ctctcctggc gggaacagca aaatatcacc cggtcggcaa
 781 acaaattctc gtccctgatt tttcaccacc ccctgaccgc gaatggtgag attgagaata
 841 taacctttca ttcccagcgg tcggtcgata aaaaaatcga gataaccgtt ggcctcaatc
 901 ggcgttaaac ccgccaccag atgggcatta aacgagtatc ccggcagcag gggatcattt
 961 tgcgcttcag ccatactttt catactcccg ccattcagag aagaaaccaa ttgtccatat
1021 tgcatcagac attgccgtca ctgcgtcttt tactggctct tctcgctaac caaaccggta
1081 accccgctta ttaaaagcat tctgtaacaa agcgggacca aagccatgac aaaaacgcgt
1141 aacaaaagtg tctataatca cggcagaaaa gtccacattg attatttgca cggcgtcaca
1201 ctttgctatg ccatagcatt tttatccata agattagcgg atcctacctg acgcttttta
1261 tcgcaactct ctactgtttc tccatacccg ttttttgggg ctagcgaatt cgagctcggt
1321 acccggggat cctctagagt cgacctgcag gcatgcaagc ttggctgttt tggcggatga
1381 gagaagattt tcagcctgat acagattaaa tcagaacgca gaagcggtct gataaaacag
1441 aatttgcctg gcggcagtag cgcggtggtc ccacctgacc ccatgccgaa ctcagaagtg
1501 aaacgccgta gcgccgatgg tagtgtgggg tctccccatg cgagagtagg gaactgccag
1561 gcatcaaata aaacgaaagg ctcagtcgaa agactgggcc tttcgtttta tctgttgttt
1621 gtcggtgaac gctctcctga gtaggacaaa tccgccgggа gcggatttga acgttgcgaa
1681 gcaacggccc ggagggtggc gggcaggacg cccgccataa actgccaggc atcaaattaa
1741 gcagaaggcc atcctgacgg atggcctttt tgcgtttcta caaactcttt tgtttatttt
1801 tctaaataca ttcaaatatg tatccgctca tgagacaata accctgataa atgcttcaat
1861 aatattgaaa aaggaagagt atgagtattc aacatttccg tgtcgccctt attccctttt
1921 ttgcggcatt ttgccttcct gtttttgctc acccagaaac gctggtgaaa gtaaaagatg
1981 ctgaagatca gttgggtgca ctgcgcaaac tattaactgg cgaactactt actctagctt
2041 cccggcaaca attaatagac tggatggagg cggataaagt tgcaggacca cttctgcgct
2101 cggcccttcc ggctggctgg tttattgctg ataaatctgg agccggtgag cgtgggtctc
2161 gcggtatcat tgcagcactg gggccagatg gtaagccctc ccgtatcgta gttatctaca
2221 cgacggggag tcaggcaact atggatgaac gaaatagaca gatcgctgag ataggtgcct
2281 cactgattaa gcattggtaa ctgtcagacc aagtttactc atatatactt tagattgatt
2341 tacgcgccct gtagcggcgc attaagcgcg gcgggtgtgg tggttacgcg cagcgtgacc
2401 gctacacttg ccagcgccct agcgcccgct cctttcgctt tcttcccttc ctttctcgcc
2461 acgttcgccg gctttccccg tcaagctcta aatcgggggc tcccttttagg gttccgattt
2521 agtgctttac ggcacctcga ccccaaaaaa cttgatttgg gtgatggttc acgtagtggg
2581 ccatcgccct gatagacggt ttttcgccct ttgacgttgg agtccacgtt ctttaatagt
2641 ggactcttgt tccaaacttg aacaacactc aaccctatct cgggctattc ttttgattta
2701 taagggattt tgccgatttc ggcctattgg ttaaaaaatg agctgattta acaaaaattt
2761 aacgcgaatt ttaacaaaat attaacgttt acaatttaaa aggatctagg tgaagatcct
2821 ttttgataat ctcatgacca aaatccctta acgtgagttt tcgttccact gagcgtcaga
2881 ccccgtagaa aagatcaaag gatcttcttg agatcctttt tttctgcgcg taatctgctg
2941 cttgcaaaca aaaaaaccac cgctaccagc ggtggtttgt ttgccggatc aagagctacc
3001 aactcttttt ccgaaggtaa ctggcttcag cagagcgcag ataccaaata ctgtccttct
3061 agtgtagccg tagttaggcc accacttcaa gaactctgta gcaccgccta catacctcgc
```

```
3121 tctgctaatc ctgttaccag tggctgcctc aggcatttga gaagcacacg gtcacactgc
3181 ttccggtagt caataaaccg gtaaaccagc aatagacata agcggctatt taacgaccct
3241 gccctgaacc gacgaccggg tcgaatttgc tttcgaattt ctgccattca tccgcttatt
3301 atcacttatt caggcgtagc accaggcgtt taagggcacc aataactgcc ttaaaaaaat
3361 tacgccccgc cctgccactc atcgcagtac tgttgtaatt cattaagcat tctgccgaca
3421 tggaagccat cacagacggc atgatgaacc tgaatcgcca gcggcatcag caccttgtcg
3481 ccttgcgtat aatatttgcc catggtgaaa acgggggcga agaagttgtc catattggcc
3541 acgtttaaat caaaactggt gaaactcacc cagggattgg ctgagacgaa aaacatattc
3601 tcaataaacc ctttagggaa ataggccagg ttttcaccgt aacacgccac atcttgcgaa
3661 tatatgtgta gaaactgccg gaaatcgtcg tggtattcac tccagagcga tgaaaacgtt
3721 tcagtttgct catggaaaac ggtgtaacaa gggtgaacac tatcccatat caccagctca
3781 ccgtctttca ttgccatacg gaattccgga tgagcattca tcaggcgggc aagaatgtga
3841 ataaaggccg gataaaactt gtgcttattt ttctttacgg tctttaaaaa ggccgtaata
3901 tccagctgaa cggtctggtt ataggtacat tgagcaactg actgaaatgc ctcaaaatgt
3961 tctttacgat gccattggga tatatcaacg gtggtatatc cagtgatttt tttctccatt
4021 ttagcttcct tagctcctga aaatctcgat aactcaaaaa atacgcccgg tagtgatctt
4081 atttcattat ggtgaaagtt ggaacctctt acgtgccgat caacgtctca ttttcgccaa
4141 aagttggccc agggcttccc ggtatcaaca gggacaccag gatttattta ttctgcgaag
4201 tgatcttccg tcacaggtat ttattcggcg caaagtgcgt cgggtgatgc tgccaactta
4261 ctgatttagt gtatgatggt gtttttgagg tgctccagtg gcttctgttt ctatcagctg
4321 tccctcctgt tcagctactg acggggtggt gcgtaacggc aaaagcaccg ccggacatca
4381 gcgctagcgg agtgtatact ggcttactat gttggcactg atgagggtgt cagtgaagtg
4441 cttcatgtgg caggagaaaa aaggctgcac cggtgcgtca gcagaatatg tgatacagga
4501 tatattccgc ttcctcgctc actgactcgc tacgctcggt cgttcgactg cggcgagcgg
4561 aaatggctta cgaacggggc ggagatttcc tggaagatgc caggaagata cttaacaggg
4621 aagtgagagg gccgcggcaa agccgttttt ccataggctc cgccccctg acaagcatca
4681 cgaaatctga cgctcaaatc agtggtggcg aaacccgaca ggactataaa gataccaggc
4741 gtttcccccct ggcggctccc tcgtgcgctc tcctgttcct gcctttcggt ttaccggtgt
4801 cattccgctg ttatggccgc gtttgtctca ttccacgcct gacactcagt tccgggtagg
4861 cagttcgctc caagctggac tgtatgcacg aaccccccgt tcagtccgac cgctgcgcct
4921 tatccggtaa ctatcgtctt gagtccaacc cggaaagaca tgcaaaagca ccactggcag
4981 cagccactgg taattgattt agaggagtta gtcttgaagt catgcgccgg ttaaggctaa
5041 actgaaagga caagttttgg tgactgcgct cctccaagcc agttacctcg gttcaaagag
5101 ttggtagctc agagaacctt cgaaaaaccg ccctgcaagg cggtttttc gttttcagag
5161 caagagatta cgcgcagacc aaaacgatct caagaagatc atcttattaa tcagataaaa
5221 tatttctaga agcgctcatg agcccgaagt ggcgagcccg atcttcccca tcggtgatgt
5281 cggcgatata ggcgccagca accgcacctg tggcgccggt gatgccggcc acgatgcgtc
5341 cggcgtagag gatctgctca tgtttgacag cttatc
```

**Protein Sequences for Expression and solubility**

**RutF sequence**

**>fig|83333.1.peg.995 [Escherichia coli K12] [Predicted flavin reductase RutF in novel pyrimidine catabolism pathway]**

ATGAACATTGTCGATCAACAAACTTTTCGCGATGCGATGTCCTGCATGGGCGCGGCGGTCAATATTATCAC
CACGGACGGTCCAGCCGGGCGCGCCGGGTTCACCGCCAGCGCCGTCTGCAGTGTGACCGATACGCCGCCCA
CATTACTGGTGTGCCTGAATCGTGGGGCGTCCGTCTGGCCGGCATTCAATGAAAATCGAACGCTGTGTGTA
AATACACTTAGCGCCGGGCAGGAGCCGCTTTCAAACCTTTTTGGCGGCAAAACGCCCATGGAACACCGCTT
TGCCGCCGCCCGCTGGCAGACGGGGGTGACCGGATGCCCGCAACTGGAAGAGGCTCTGGTTTCGTTTGACT
GCCGTATCAGCCAGGTGGTGAGCGTCGGCACCCACGACATTCTGTTTTGCGCCATCGAAGCGATTCATCGT
CACACCACACCCTACGGGCTGGTGTGGTTCGATCGCAGTTATCACGCGCTGATGCGCCCTGCTTGTTAA

**BLG sequence**

ATGAAATGTCTGCTGCTGGCGCTGGCCCTGACCTGCGGTGCGCAGGCCCTGATTGTGACCCAGACCATGAA
AGGCCTGGATATTCAGAAAGTGGCGGGTACCTGGTATAGCCTGGCCATGGCGGCGAGCGATATTAGCCTGC
TGGATGCGCAGAGCGCGCCGCTGCGTGTGTATGTTGAAGAACTGAAACCGACCCCGGAAGGCGATCTGGAA
ATTCTGCTGCAGAAATGGGAAAACGGCGAATGCGCGCAGAAAAAAATTATCGCGGAAAAAACCAAAATTCC
GGCGGTGTTTAAAATTGATGCGCTGAATGAAAACAAAGTGCTGGTGCTGGATACCGATTATAAAAAATATC
TGCTGTTTTGCATGGAAAATAGCGCGGAACCGGAACAGAGCCTGGCGTGCCAGTGTCTGGTGCGCACCCCG
GAAGTGGATGATGAAGCCCTGGAAAAATTTGATAAAGCGCTGAAAGCGCTGCCGATGCATATTCGCCTGAG
CTTTAATCCGACCCAGCTGGAAGAACAGTGTCACATTTAA

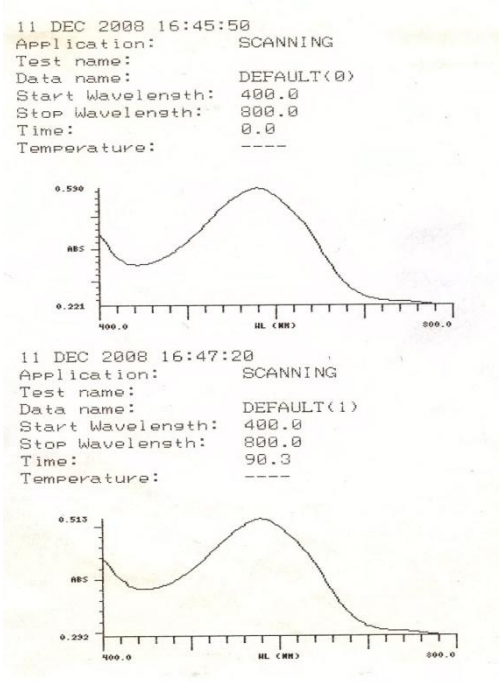
**Figure D.3: Sequences used in Chapter V**

- 10kb A PCR product (Whole vio opron)
  - VioA: 97% identity $3e^{-45}$
  - CV3269 92% identity, $7e^{-126}$ E value
- ~5kb PCR product
  - vioA 98% identity $4e^{-46}$
  - cv3269 84% identity $1e^{-8}$
  - VioBCF 91% identity $7e^{-17}$ (vioC 82%, VioB 97%)
  - vioB 97% identity $2e^{-172}$
- ~10kb size from ~5kb+~5kb (Whole vio operon)
  - Cv3269 97% identity $2e^{-120}$
  - VioA 97% identity $2e^{-55}$
- Original violet colored colony (mini-prep) (Lac+Native)
  - Oriv3269LFS : CV3269: 93% Identity $4e^{-122}$
  - O3276RLS: VioA, 90% identity $1e^{-96}$
- V1(best results) (Lac+Native)
  - V13269LFS: CV3269, 99% Identity $1e^{-122}$
  - V13275RLS: vioA, 99% Identity $8e^{-91}$
- V2 (Lac+Native)
  - V23269LFS: CV3269, 99% Identity $3e^{-120}$
  - V23275RLS: vioA, 73% Identity $9e^{-67}$
- V5 violet colored colony (mini-prep) (Lac+Native)
  - AF :vioA, 99% Identity $4e^{-122}$
  - BF: vioB, 99% Identity $1e^{-122}$
  - BR: vioB, 99% Identity $8e^{-91}$
  - CF: vioC, 99% Identity $3e^{-120}$
  - CR: vioC, 99% Identity $9e^{-67}$
  - DF: vioD, 99% Identity $1e^{-122}$
  - DR: vioD, 99% Identity $8e^{-91}$
  - EF: vioE, % Identity $3e^{-120}$
  - ER: vioE, % Identity $9e^{-67}$
  - All operon was in the vector (TOPO)
- RV8: colored clone from riboswitch in TOPO plasmid without native promoter (lac promoter + riboswitch)
  - M13R: $3e^{-152}$ 96% Identity VioA
  - M13F: $2e^{-153}$ 99% Identity CV3269
  - VioAF: $1e^{-138}$ 99% Identity VioA
  - CV3269: $8e^{-120}$ 98% Identity VioA
- V4D1 (pETBlue-2 + vio operon)
  - vioAF: $2e^{-118}$ 99% identity VioA
  - CV3269 : $4e^{-118}$ 99% identity CV3269 in chromobacterium violaceum ATCC12472
- RutF
  - FRutF: 98% identity
  - RRutF: 94% identity
- Riboswitch
  - 100% Identity

**Riboswitch sequence (100% match)**


**Figure D.4: Sequencing results with PCR products**

**At 585nm**



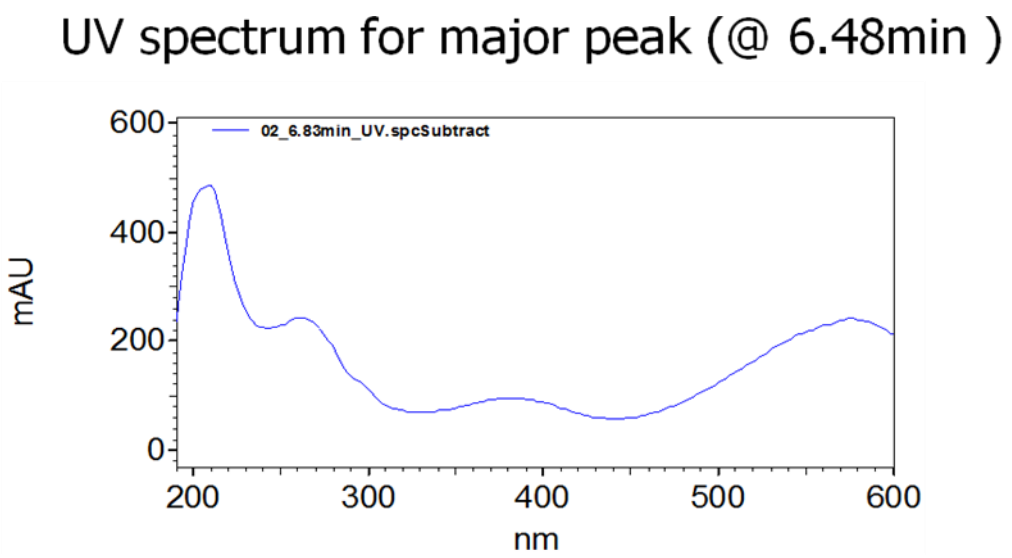
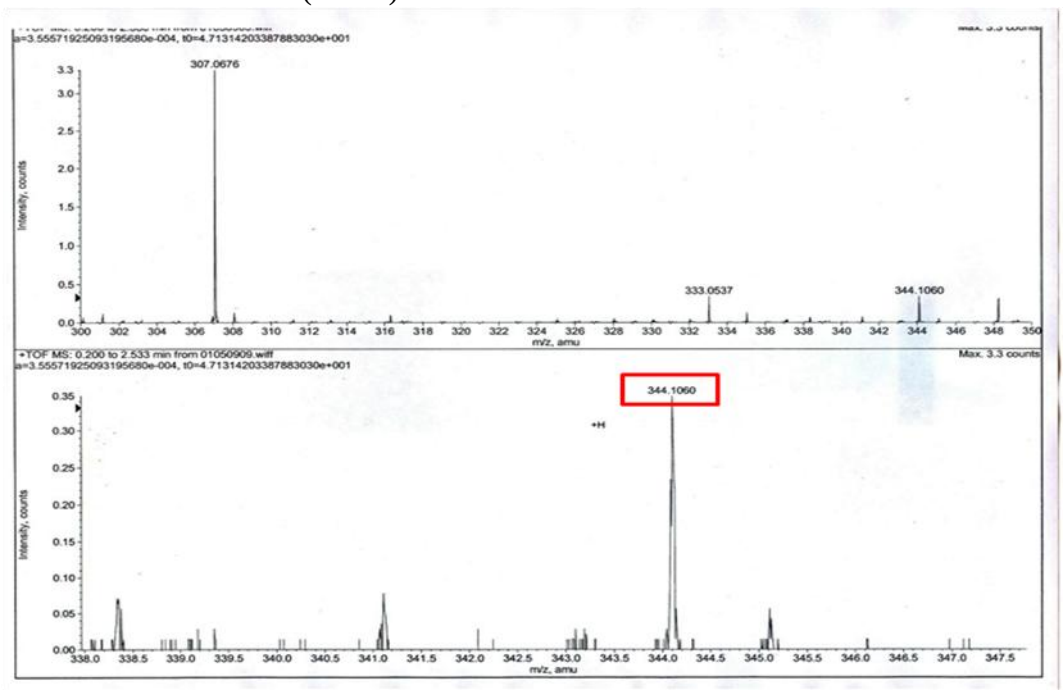




**Figure D.5: UV-Vis DATA**

V7 MS crude cell extract (*E. coli*)




V5 MS crude cell extract (*E. coli*)

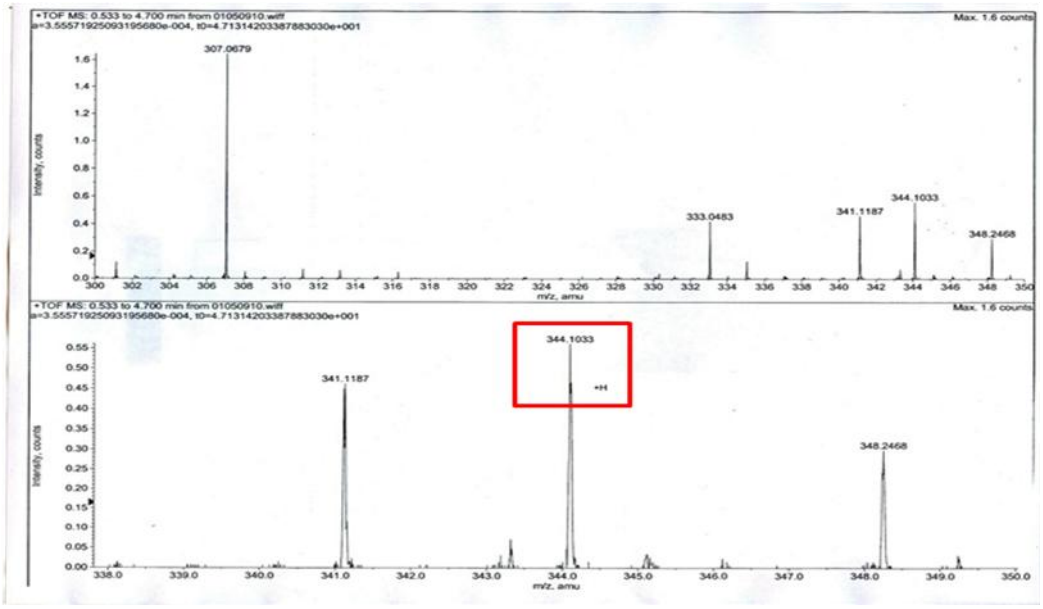
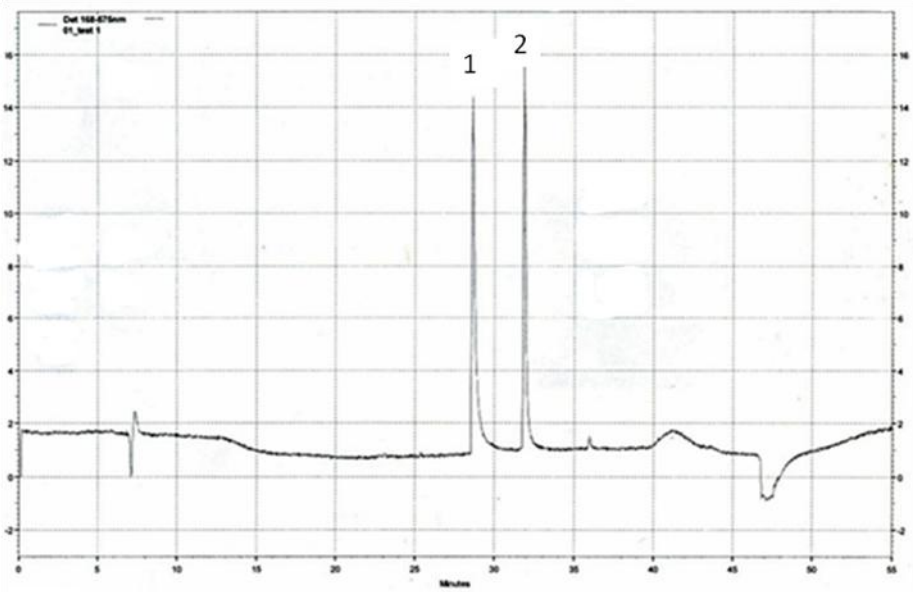
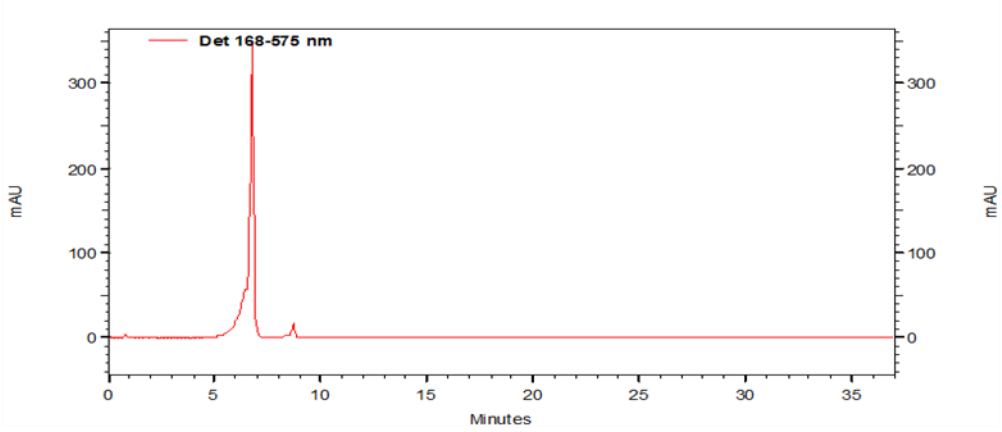


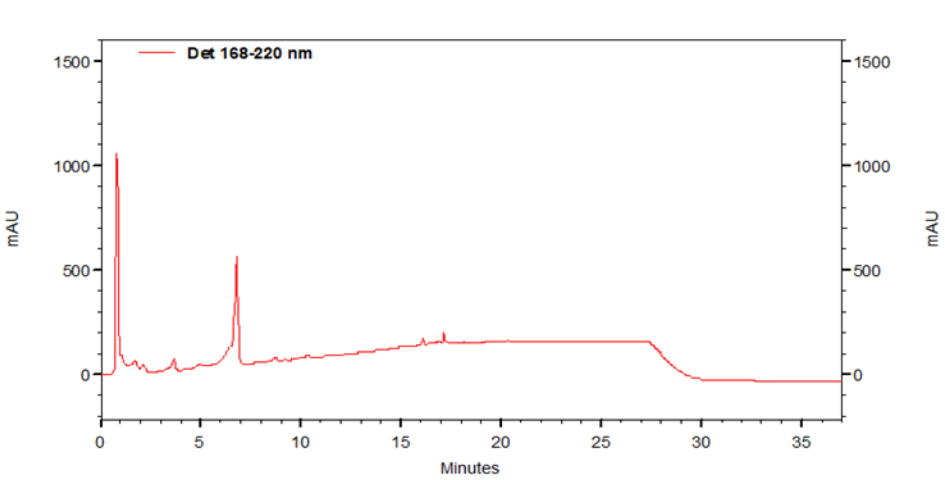

**Figure D.6: MS Spectra data of Violacein**




**Sampled was dissolved in 10% Acetnitile (ACN)**
**Detected at 168-575nm**
**Gradient 4 min 10% ACN 90% Water → 88% ACN in 30min then 100% CAN**
**Flow rate 0.5ml/min**

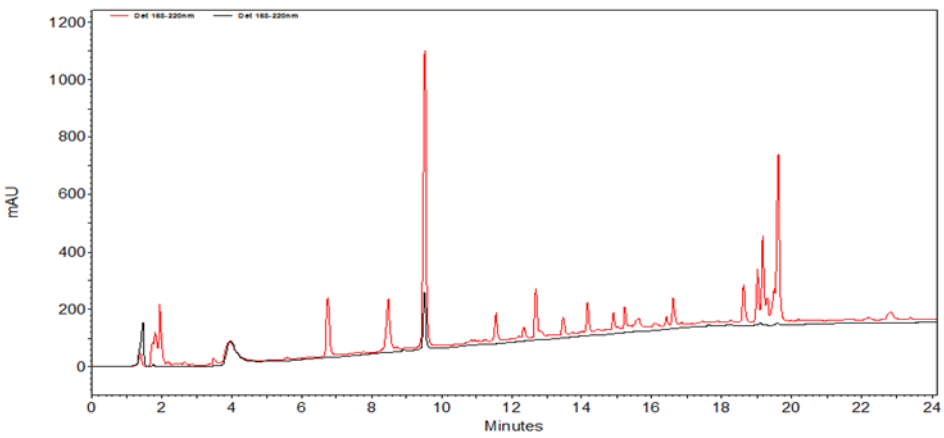



Detected@575nm
Gradient 50% MeOH-50%water → 100% MeOH in 15 min
Column Gmini C18 (75mmX4.6mm, 3micron)

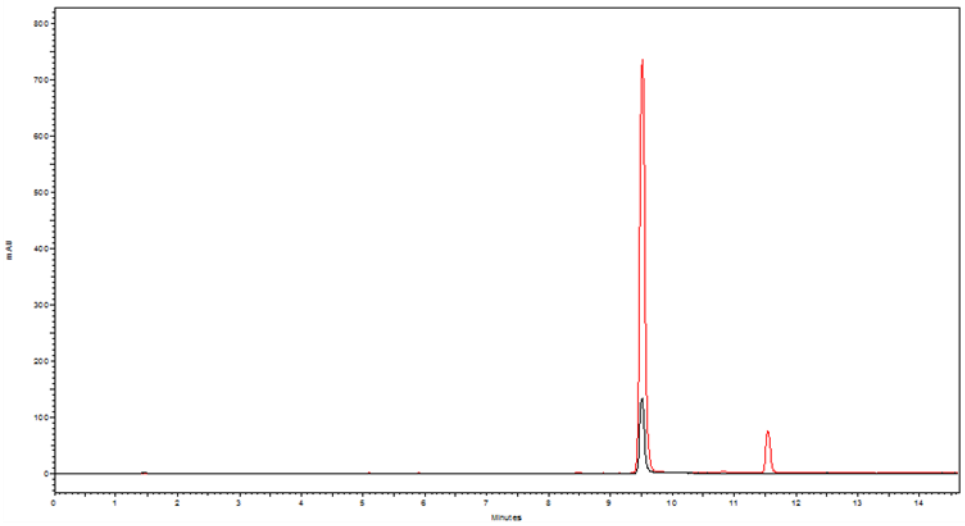



Detected@575nm
Gradient 50% MeOH-50%water → 100% MeOH in 15 min
Column Gmini C18 (75mmX4.6mm, 3micron)

**Anlytical vs prep HPLC**




Column: Gemini C18, 150X4.6mm (MeOH/water) Gradient: 50%B to 100% in 15min
Wavelength 220nm
Violacein @ 9.53min
**Red trace: violacein crude sample**
**Black trace: violacein prep. fraction**
Peak around 4min most probable is a system peak, not from the sample itself.



**Red trace: violacein crude sample**
**Black trace: violacein prep. fraction**

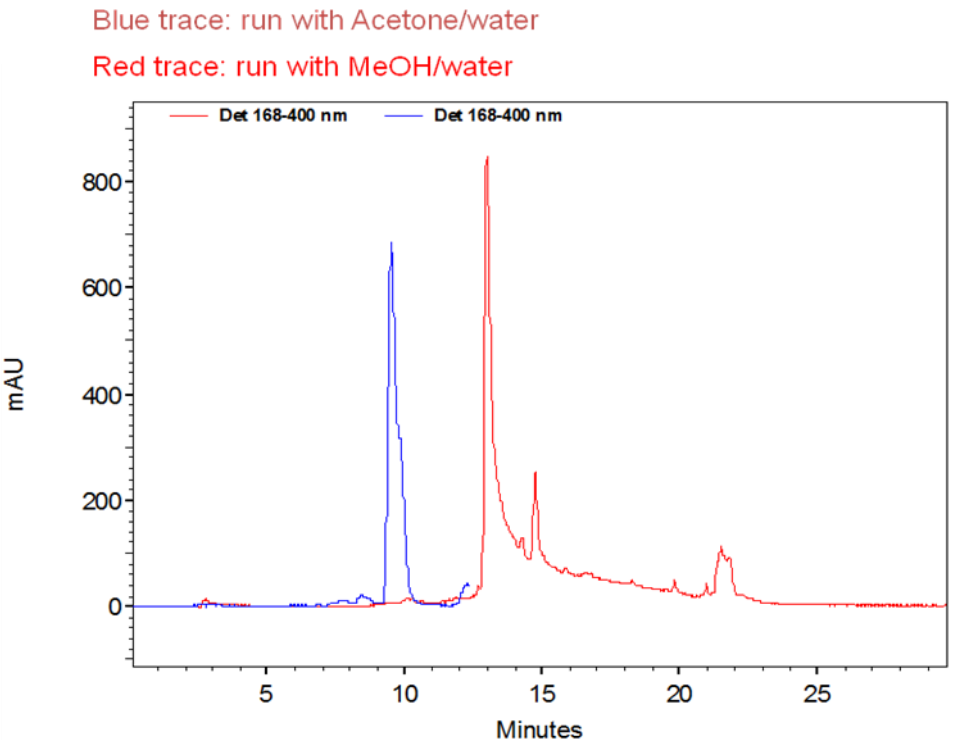
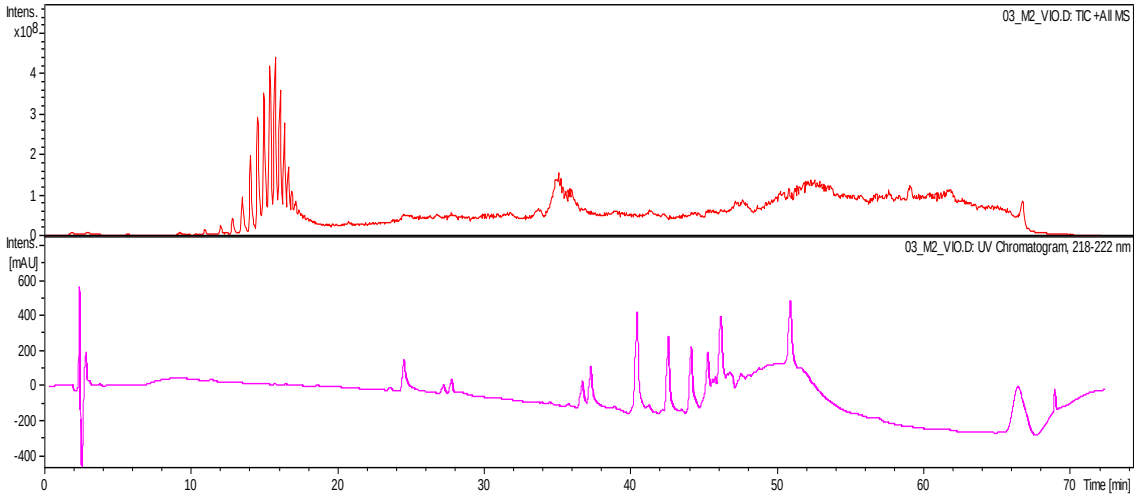
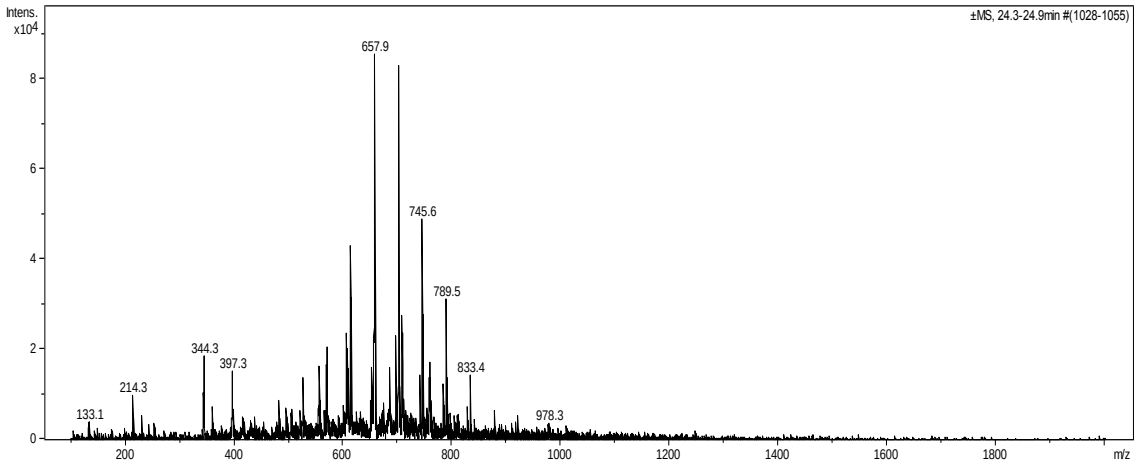



**Figure D.7: HPLC spectra of violacein**

**(1) TIC (ESI-APCI-Positive mode) (red trace) UV 220NM ( pink trace)**
**A: 0.2% Formic acid water B: 0.14% formic acid Acetonitrile**
**Gradient 5→ 100% B in 45min**




**(2)Average mass of peak 24.3-24.9 min**




**(3) Zoom-in at target mass range**

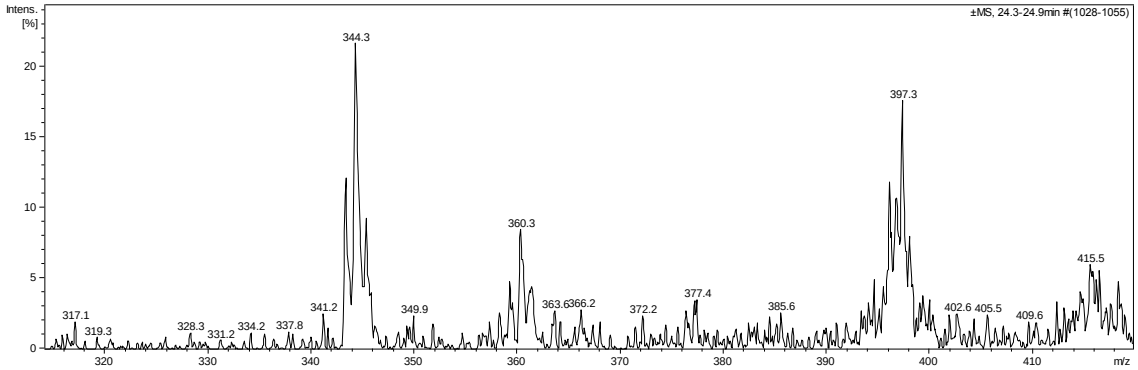



**Figure D.8: LCMS spectra of violacein**

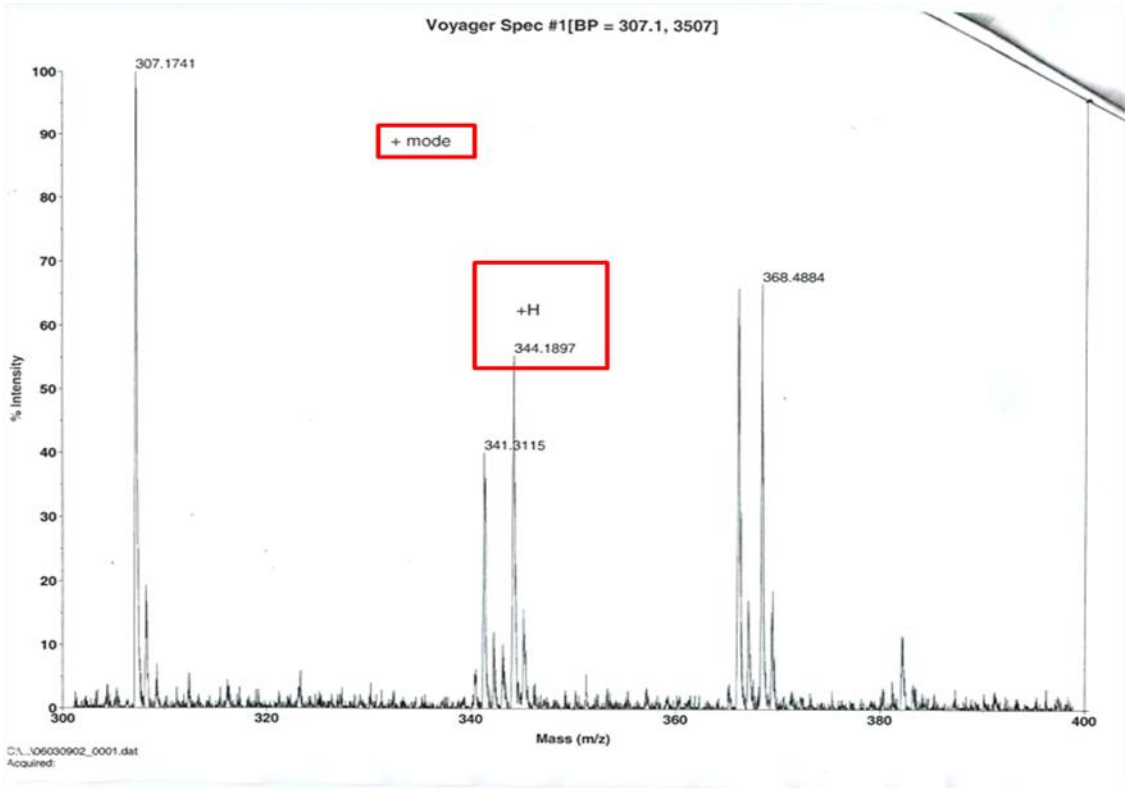
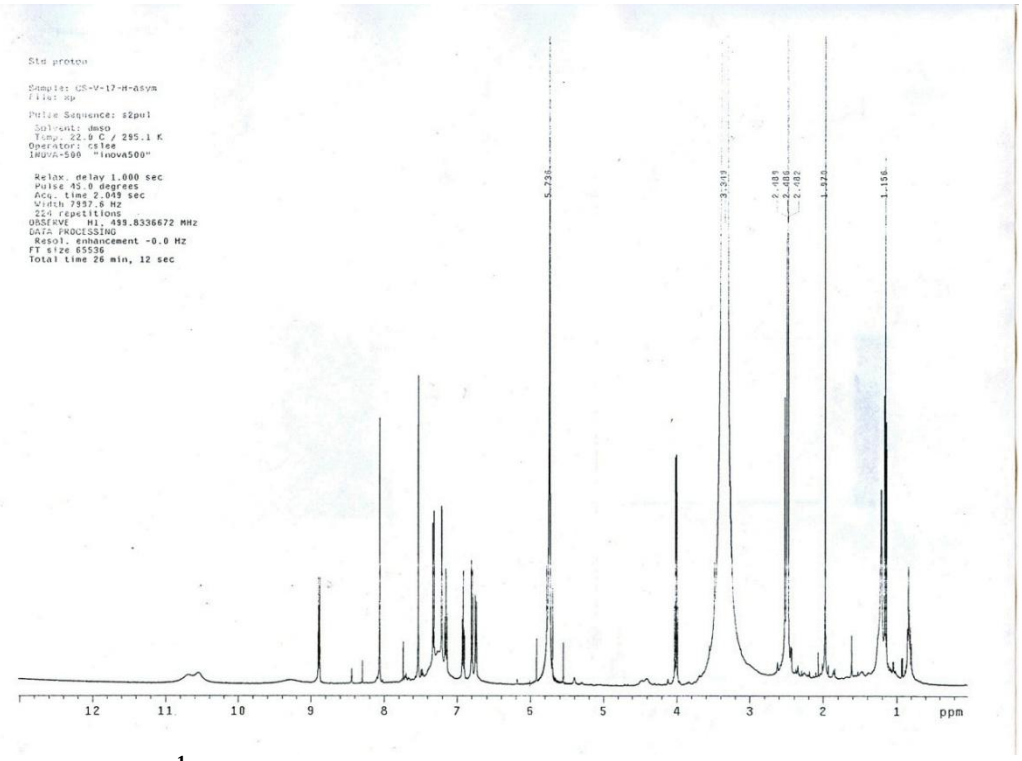
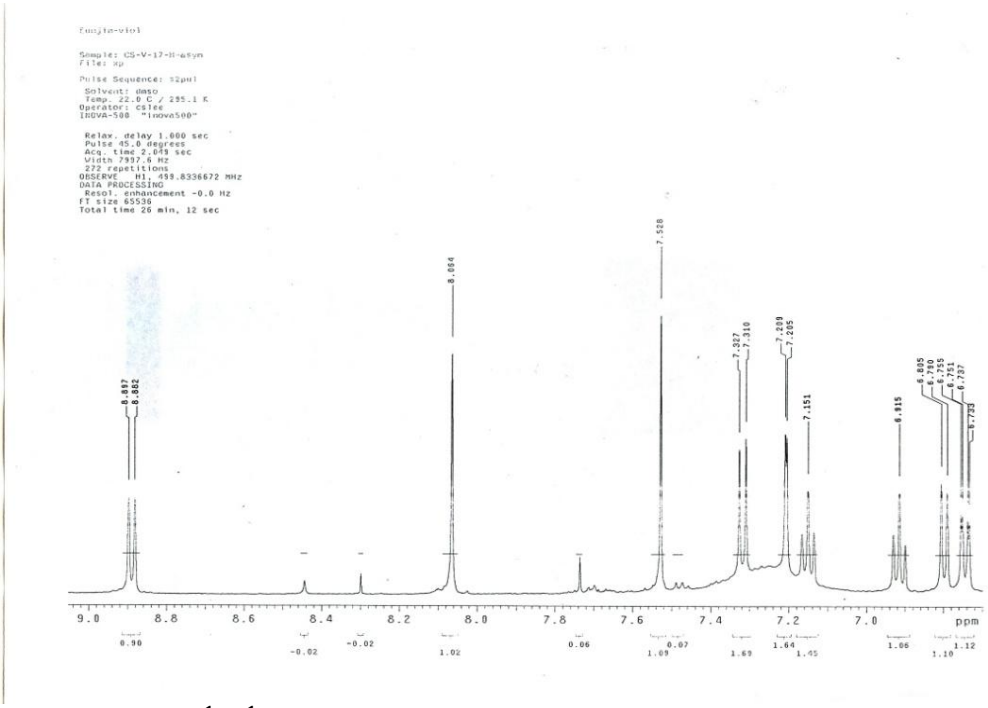



**Figure D.9: MALDI MS spectra of violacein**

**$^{1}$H NMR 500MHz (used 5mm BB/H probe)**




**$^{1}$H NMR 500MHz (used 3mm BB/H probe) >7ppm**




**$^{1}$H-$^{1}$H COSY 2D NMR (used 3mm tube) >7ppm**

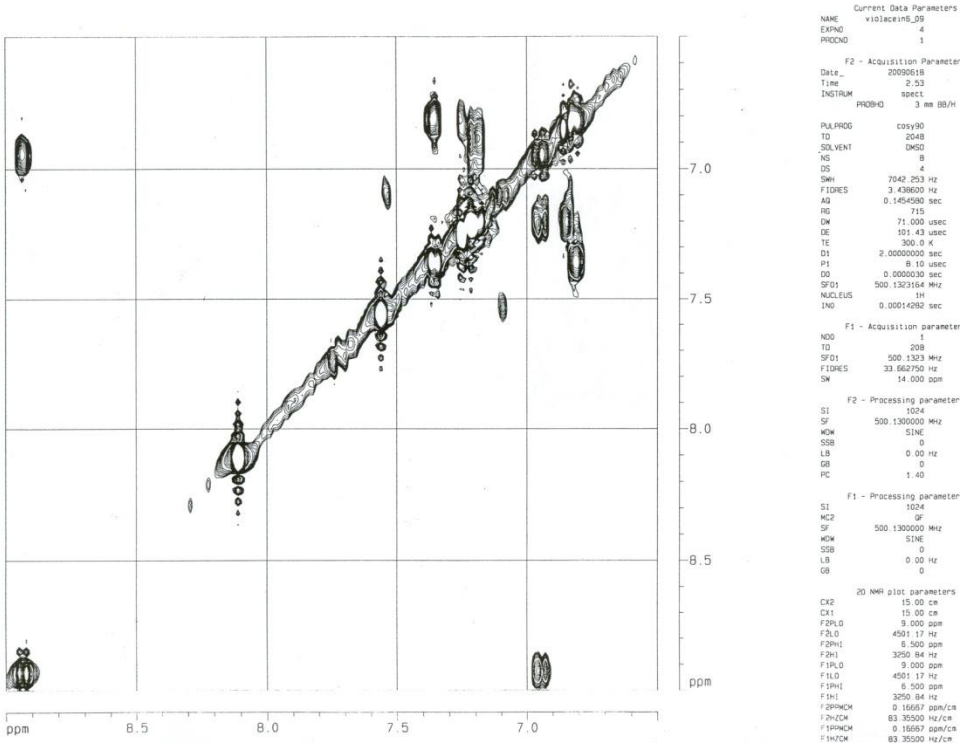
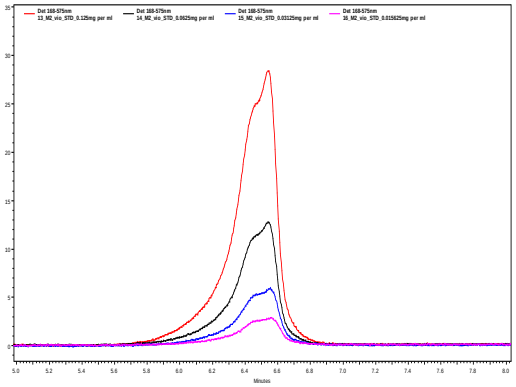



**Figure D.10: NMR spectra of violacein**

**Standard Curve**

| STD (mg/ml) in HPLC sample | STD(mg/ml) | peak area @ 575nm |
|---|---|---|
| 1 | 5 | 7866240 |
| 0.5 | 2.5 | 3176951 |
| 0.25 | 1.25 | 1348400 |
| 0.125 | 0.625 | 522324 |
| 0.0625 | 0.3125 | 243981 |
| 0.03125 | 0.15625 | 116922 |
| 0.015625 | 0.078125 | 59583 |
| 0.0078125 | 0.0390625 | 52023 |



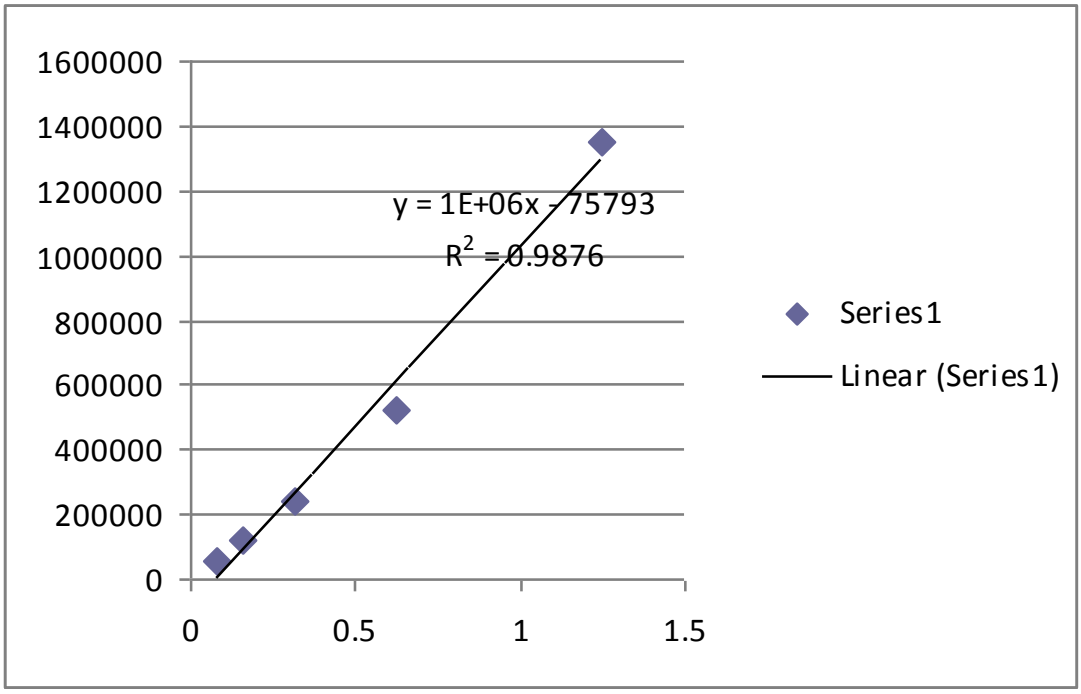



**Figure D.11: Violacein production supplementary data**

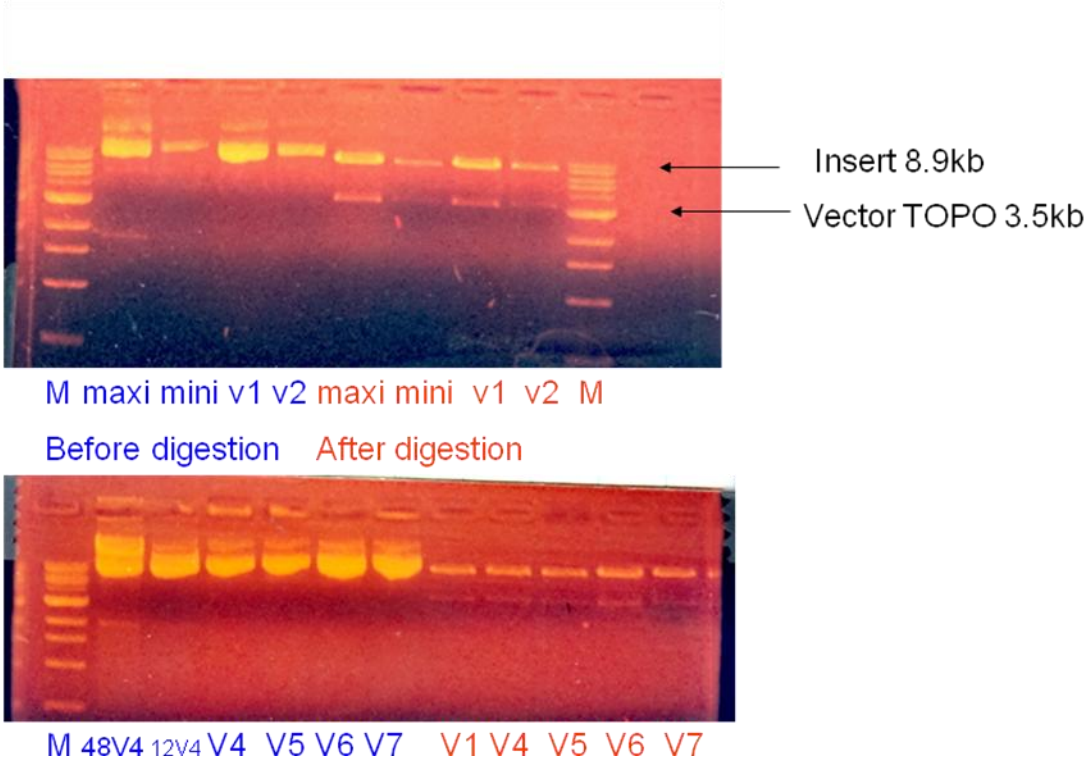
| Sample | Area | violacine conc. (mg/ml) |
|---|---|---|
| Lac+Native | 212930<br>298101 | 0.289<br>0.374 |
| Lac+Riboswitch | 129323<br>138304 | 0.205<br>0.214 |
| Native (Used 5times more samples) | Too low | 0.000<br>0.000 |
| T7 (Used 5times more samples) | Too low | 0.000<br>0.000 |
| ThiC+Riboswitch | 3121740<br>3089160 | 3.198<br>3.166 |
| Lac | 63047<br>70136 | 0.139<br>0.146 |

**Figure D.12: Tabulated data of violacein production under various promoter**

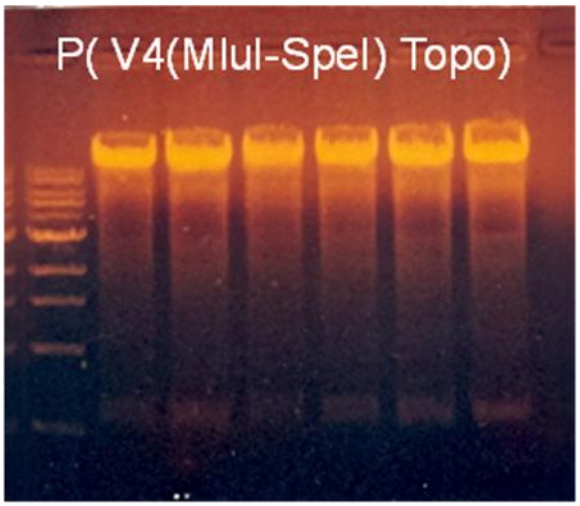
| | | |
|---|---|---|
| ThiC+Ribo | 3121740 | 3.198 |
| 0.5mM TPP | 2637494 | 2.713 |
| 1.0mM TPP | 2388635 | 2.464 |
| 1.5mM TPP | 887811 | 0.964 |
| 2.0mM TPP | 616711 | 0.693 |
| 2.5mM TPP | 430717 | 0.507 |
| 3.0mM TPP | 499072 | 0.575 |
| 5.0mM TPP | 155624 | 0.231 |
| ThiC+Ribo | 3089760 | 3.166 |
| 0.5mM TPP | 2526476 | 2.602 |
| 1.0mM TPP | 2184841 | 2.261 |
| 1.5mM TPP | 1577464 | 1.653 |
| 2.0mM TPP | 477161 | 0.553 |
| 2.5mM TPP | 473124 | 0.549 |
| 3.0mM TPP | 529531 | 0.605 |
| 5.0mM TPP | 417570 | 0.493 |

**Figure D.13: Tabulated Data for violacein production under ThiC/Riboswitch in the presence of TPP**

**Insertion check**

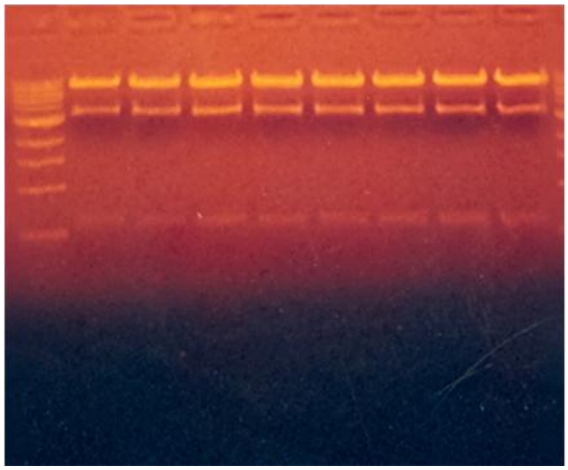


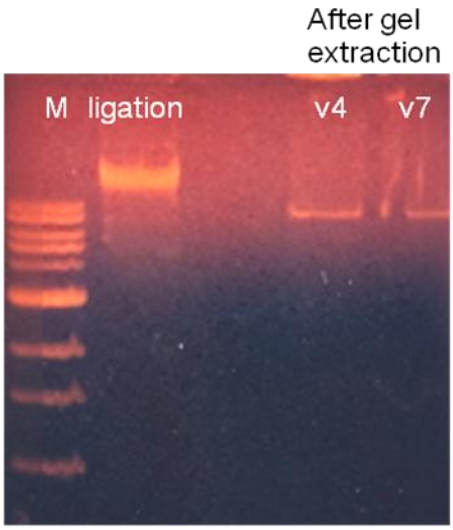

**pV4(MluI-SpeI)Topo construction**
**: (Lac promoter + Violacein without native promoter part)**




3.5kb + 8.9kb – 0.57kb
= 11.83 kb

**SpeI-XhoI violacein operon (no Native promoter region)**

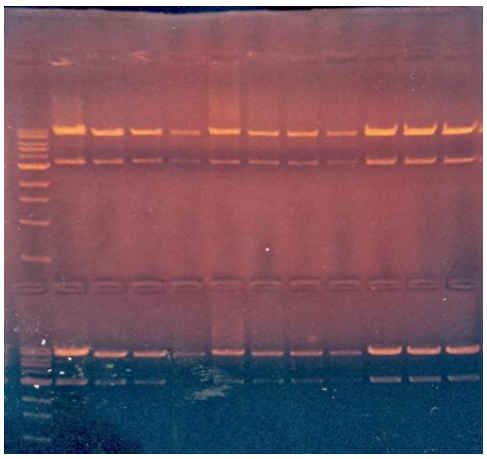





**Lac+Riboswitch construction:**

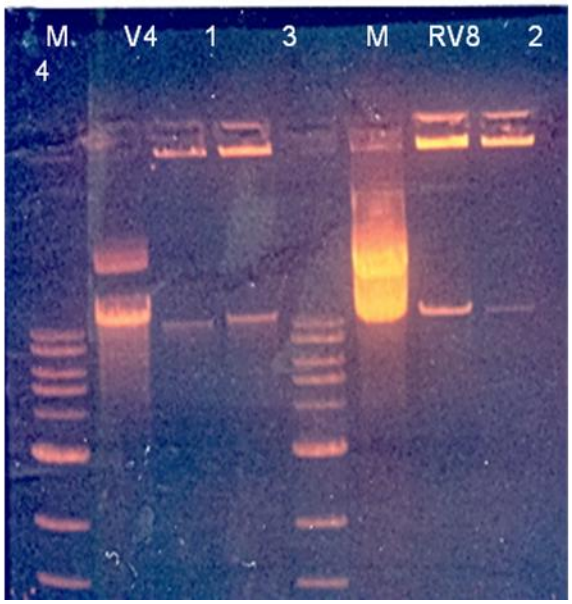
A: Riboswitch insert after MluI-SpeI digestion B: After ligation with Riboswitch and MluI-SpeI digested Topo XL vector which containing violacein operon to remove native promoter part (MluI-SpeI)

A:



B:

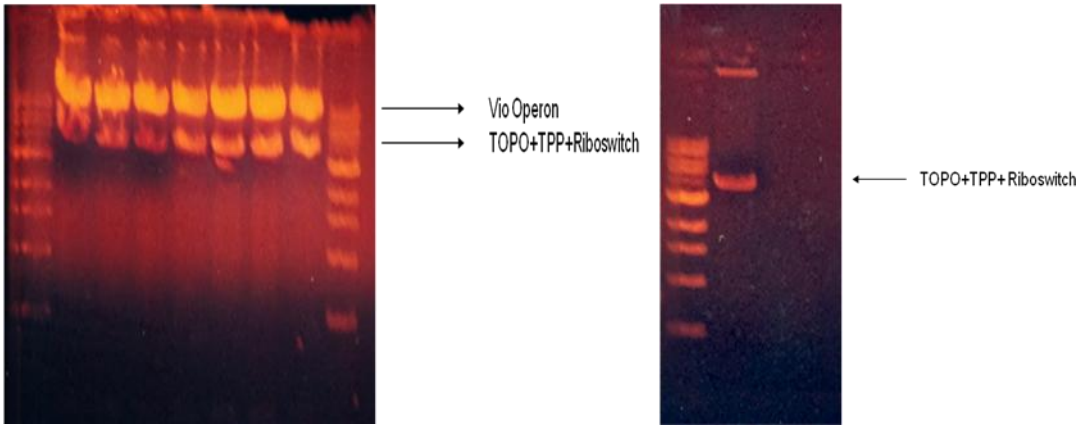
**Vector manipulation for different construction with various promoters**




V4: Before digestion
(TOPO+Vio operon)
1: V4 (AseI-SpeI) digestion (11.56kb)
3: V4 (AseI-MluI) digestion (12.197kb)

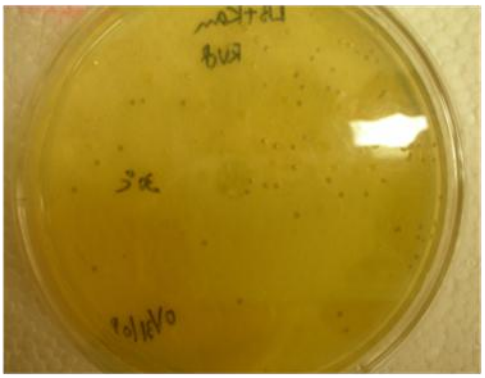
RV8: Before digestion
(TOPO+Vio operon+Riboswitch)
2: RV8 (AseI-MluI) digestion (12.297kb)
4: RV8 (MluI) digestion (12.31kb)

**TPP+Riboswitch+Vio operon (no Native promoter)**

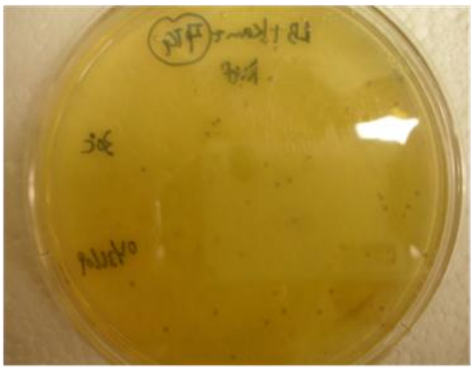



After gel extraction

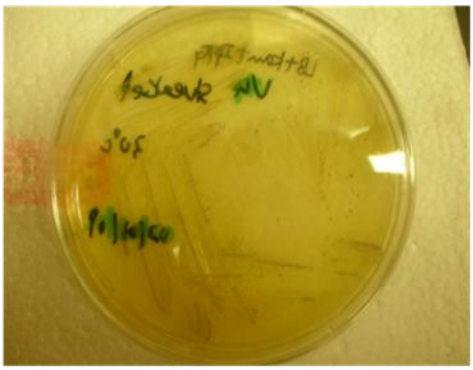
**IPTG effect check with Lac+Native and Lac+Riboswitch systems**



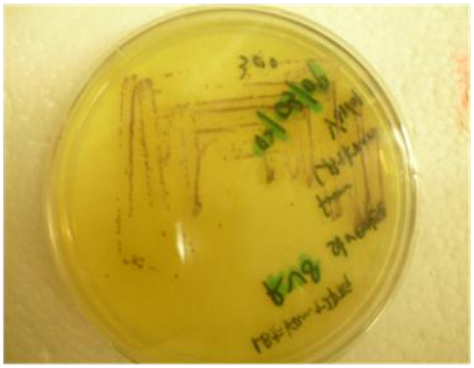
RV8 (LB+Kan)



RV8 (LB+Kan+IPTG)

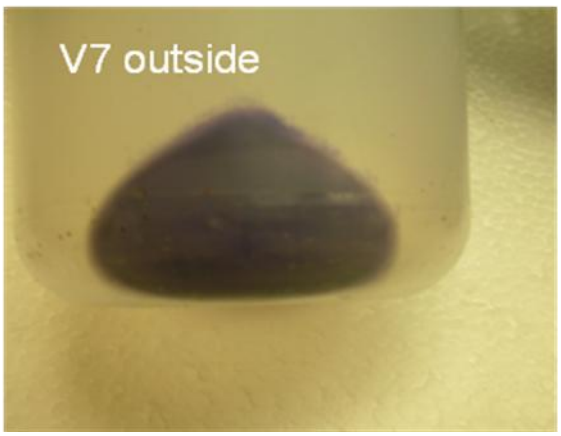


Streaked V4 from LB+Kan (LB+Kan+IPTG)

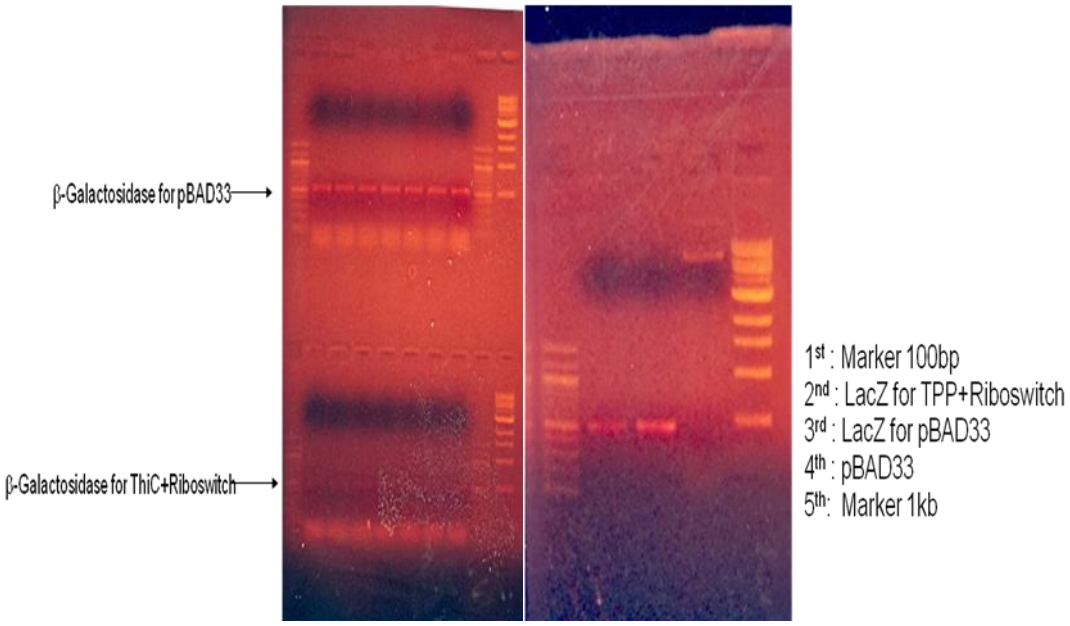


Streaked RV8 from LB+Kan (LB+Kan+IPTG)

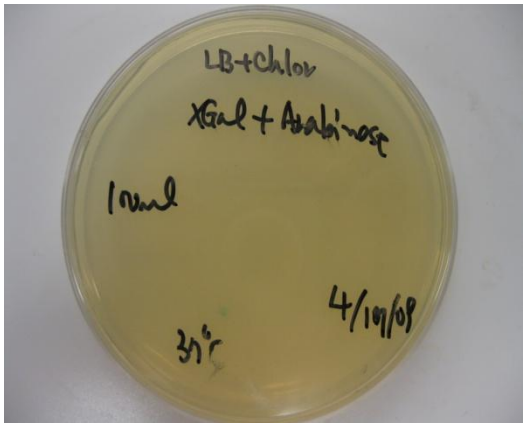
**Expression of violacein in E.coli with Lac+Native promoter**

**β-Galactosidase PCR product and DNA recovery after gel extraction**




After transformation into Top 10

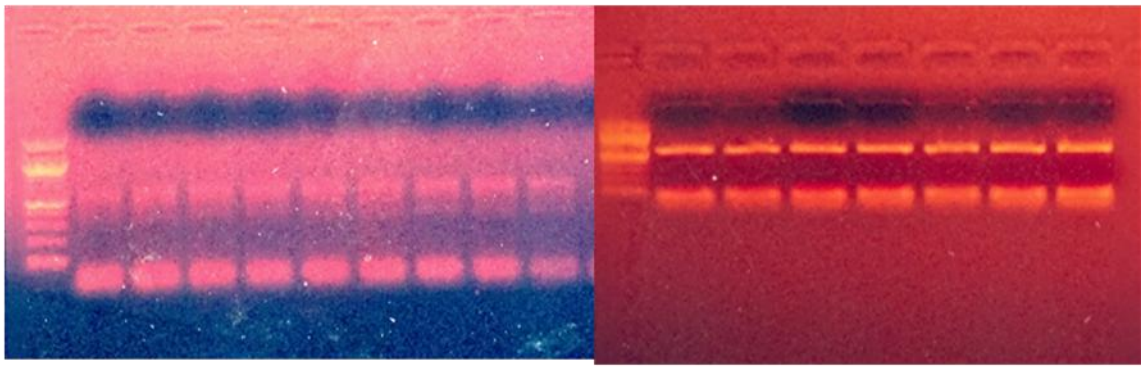




Sequencing results

- pBAD33+LacZ
  - 20ul PCR reaction :~60% identity
  - 3ul PCR reaction: 99% identity

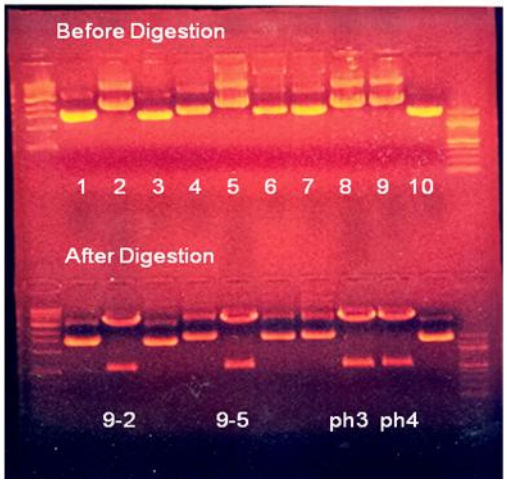
**BLG protein**

PCR



Condition #9 with Taq polymerase

With Phusion DNA polymerase




1~5: condition #9
6~10: phusion
Results:
**9-2, 9-5, ph3,** and **ph4** have insertion

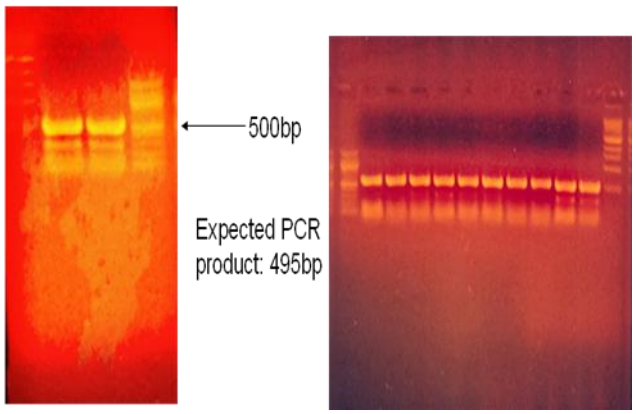
Sequencing results:

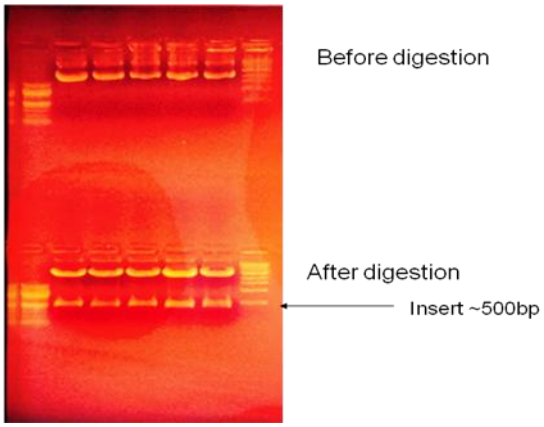
All got the BLG Insertion

> 92% Identity

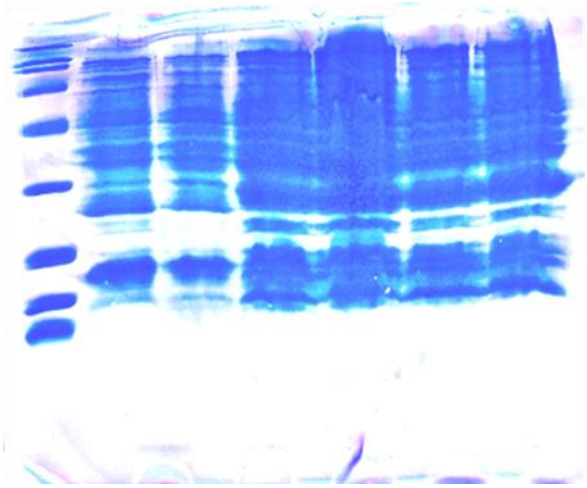
**Figure D.14: Supplementary data of Gene manipulation**

Predicted flavin reductase RutF in novel pyrimidine catabolism pathway



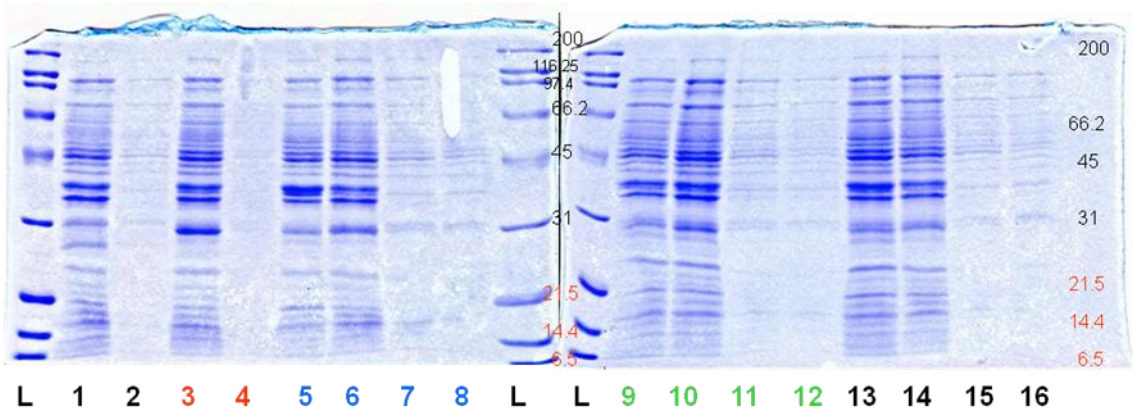




**15% SDS-PAGE gel**



M: Protein marker
1: RutF(top10) 24hr
2: RutF(top10) 42hr
3: RutF2+ThiC+Ribo 24hr
4: RutF2+ThiC+Ribo 42hr
5: RutF4+ThiC+Ribo 24hr
6: RutF4+ThiC+Ribo 42hr

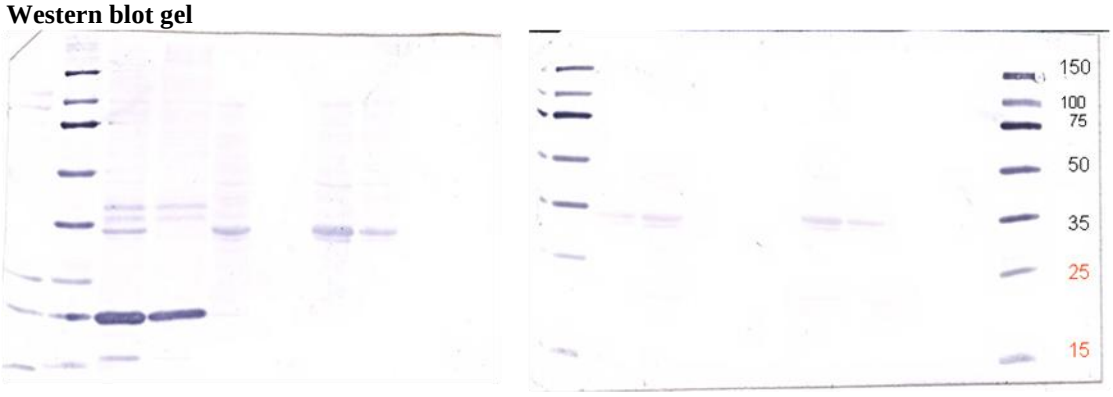
All samples were treated with heat (90°C)




L: Protein ladder
1: RutF (Cell) no IPTG 24hrs
2: RutF (soluble) no IPTG 24hrs
3: RutF2 (Cell) 48hrs
4: RutF2 (soluble) 48hrs
5: RutF2+0.5mM TPP (Cell) 24hrs
6: RutF2+0.5mM TPP (Cell) 48hrs
7: RutF2+0.5mM TPP (soluble) 24hrs
8: RutF2+0.5mM TPP (soluble) 48hrs

L: Protein ladder
9: RutF2+1.0mM TPP (Cell) 24hrs
10: RutF2+1.0mM TPP (Cell) 48hrs
11: RutF2+1.0mM TPP (soluble) 24hrs
12 RutF2+1.0mM TPP (soluble) 48hrs
13: RutF2+1.5mM TPP (Cell) 24hrs
14: RutF2+1.5mM TPP (Cell) 48hrs
15: RutF2+1.5mM TPP (soluble) 24hrs
16: RutF2+0.5mM TPP (soluble) 48hrs

**Western blot gel**

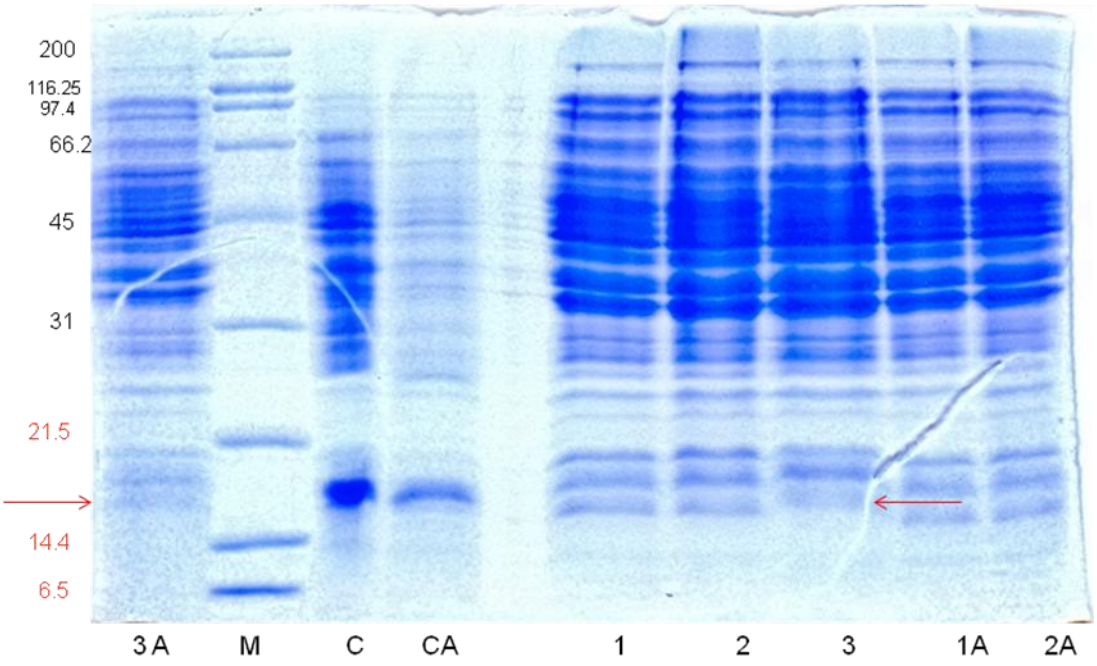



L: Western-blot ladder
1: RutF (Cell) Induced 24hrs
2: RutF (soluble) Induced 24hrs
3: RutF2 (Cell) 48hrs
4: RutF2 (soluble) 48hrs
5: RutF2+0.5mM TPP (Cell) 24hrs
6: RutF2+0.5mM TPP (Cell) 48hrs
7: RutF2+0.5mM TPP (soluble) 24hrs
8: RutF2+0.5mM TPP (soluble) 48hrs

L: Western-blot ladder
9: RutF2+1.0mM TPP (Cell) 24hrs
10: RutF2+1.0mM TPP (Cell) 48hrs
11: RutF2+1.0mM TPP (soluble) 24hrs
12 RutF2+1.0mM TPP (soluble) 48hrs
13: RutF2+1.5mM TPP (Cell) 24hrs
14: RutF2+1.5mM TPP (Cell) 48hrs
15: RutF2+1.5mM TPP (soluble) 24hrs
16: RutF2+0.5mM TPP (soluble) 48hrs

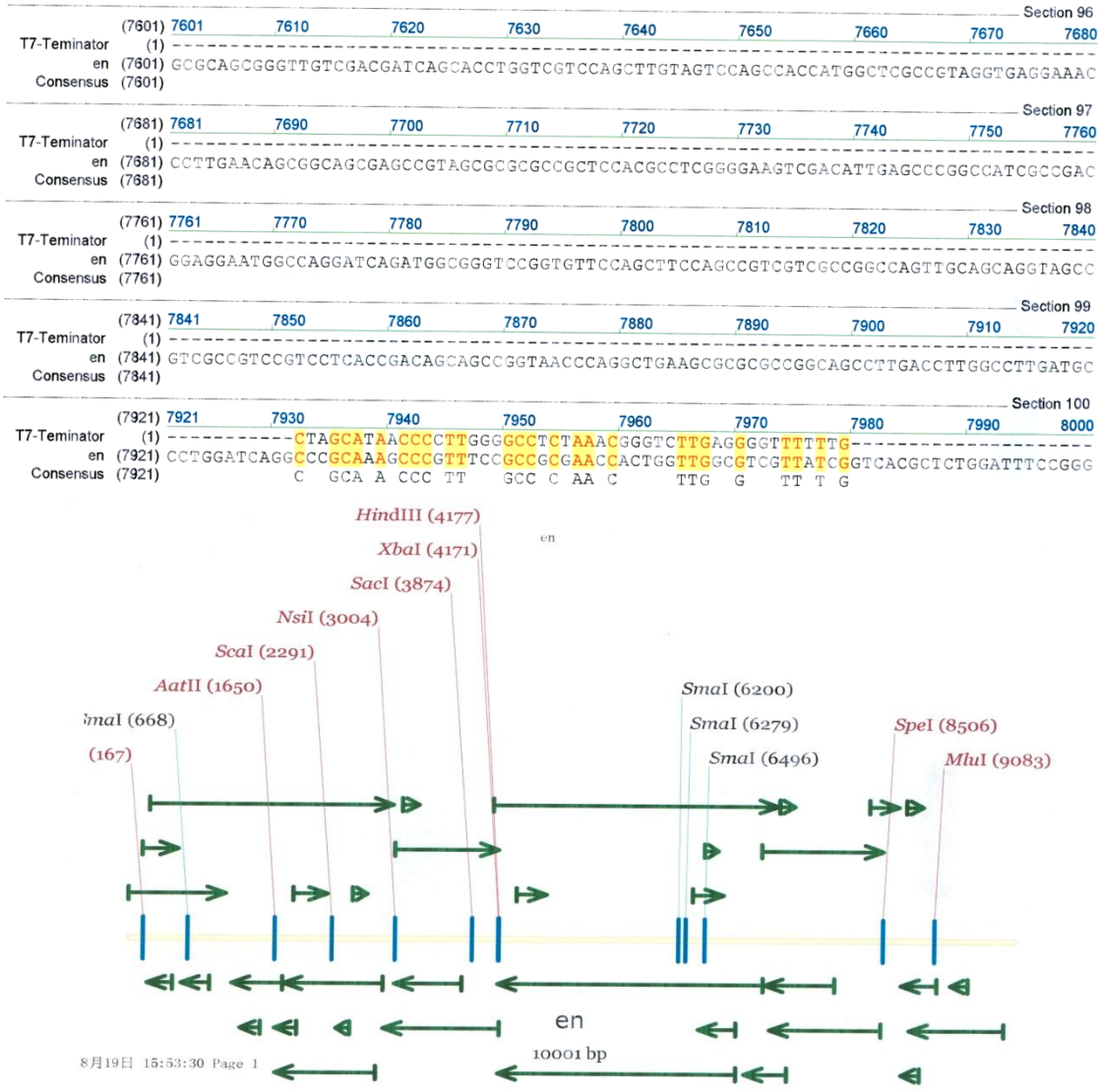
**Figure D.15: Supplementary Data for Gene manipulations of the Insoluble Proteins RutF**

SDS-PAGE gel (12%)




M: ladder
C : Control RutF sample CA:RutF (soluble) Induced 24hrs
1,1A: pETBlue/RutF (10hrs after induction with IPTG)
2, 2A: pETBlue/RutF (24hrs after induction with IPTG)
3, 3A: pETBlue/RutF (34hrs, Cell was grown in LB+Amp+IPTG)

- Violacein operon has been checked for T7 terminator
  - Found similar T7 terminator sequence (7932-7978) after native promotor (MluI~SpeI site)




**Looks like hair-pin shape**:

```
                GCTATTGCTGCGGTTGGT
                ++ +  ++ ||| ++++ C
CCCGCAAAGCCCGTTTCCGCCGCGAACCA
Stem Length = 3  Loop Length = 13
```

**Figure D.16: Evaluation of T7 Promoter function with RutF protein**

# VITA

Name: Eun Jin Kim

Address: Chemistry Department, Texas A&M University, College Station, TX 77843 Mail stop 3255

Email Address: kejin11@hotmail.com

Education: B. S. Chemical Education, Korea National Univeristy of Education, 1993

M. S. Chemistry, Korea National University of Education, 1995

M. S. Chemistry, Texas A&M University, 2004

Ph. D. Chemistry, Texas A&M University, 2009