4-20-2011

SELEX-Based Identification of Synthetic Binding Sites for Streptomyces acidiscabies 84.104 γ-Butyrolactone Receptors: SabS and SabR.

Caitin Nicole Spaulding
Trinity University, cspauldi@trinity.edu

SELEX-Based Identification of Synthetic Binding Sites for *Streptomyces acidiscabies 84.104* γ-Butyrolactone Receptors: SabS and SabR.

Caitlin Nicole Spaulding

A DEPARTMENT HONORS THESIS SUBMITTED TO THE
DEPARTMENT OF BIOLOGY AT TRINITY UNIVERSITY
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR GRADUATION WITH
DEPARTMENTAL HONORS

DATE _____

_____ _____
THESIS ADVISOR DEPARTMENT CHAIR

_____
ASSOCIATE VICE PRESIDENT FOR ACADEMIC AFFAIRS,
CURRICULUM AND STUDENT ISSUES

SELEX-Based Identification of Binding Sites for *Streptomyces acidiscabies* gamma-Butyrolactone Receptors SabS and SabR

Caitlin Nicole Spaulding

Dr. Frank G. Healy

April 20, 2011

Table of Contents

Acknowledgments:

I would like to thank my research mentor, Dr. Frank Healy - thank you for your guidance and mentorship throughout this project and for sharing your passion for molecular microbiology with me. Many, many thanks to Dr. Kevin Livingstone and Dr. James Shinkle for your insightful comments in review of this work. A special thanks to Dr. Jonathan King and Dr. Jessica Hollenbeck - I am truly grateful for the knowledge, support, and mentorship that you both provided me throughout my time at Trinity. I would also like to whole heartedly thank my wonderful parents, Larry and Angela Spaulding, and brother, Aaron Spaulding - this research and honors thesis would not have been possible without your unwavering love, support, advice and guidance. I dedicate this thesis to you.

This project was generously supported with funding from: the Trinity University Biology Department, Howard Hughes Medical Institute, and the American Society of Microbiology.

Abstract

Bacteria belonging to the Gram positive genus *Streptomyces* produce many secondary metabolite natural products of pharmaceutical importance and undergo a complex program of morphological differentiation. The expression of genes involved in these processes is often regulated by γ-butyrolactone (GBL) interactions with cognate GBL receptors. GBL receptors regulate gene expression by binding specific AT-rich DNA sequences known as autoregulatory elements (ARE) in response to available intracellular GBL pools. We previously identified two *S. acidiscabies* GBL receptor genes, *sabR* and *sabS,* flanking a gene encoding a GBL synthase homolog *sabA*. Genetic and biochemical studies revealed that *sabR* negatively regulates *sabS* expression by binding an ARE located in the 5' region of *sabS*, while *sabS* mutants overproduce the type II aromatic polyketide WS5995B and exhibit a conditional impairment in morphological differentiation. In order to further understand the role of GBL receptors in regulating cellular behavior, we have applied synthetic and genomic SELEX (Systematic Evolution of Ligands by Exponential Enrichment) methods to identify AREs for recombinant SabR and SabS. Following variable rounds of iterative SELEX, DNA substrates bound to the proteins were cloned and sequenced. Sequence analysis of cloned products revealed DNA sequences that resemble AT-rich ARE sequences. This method can now be used to identify target genes regulated by these GBL receptors and compare the properties of binding interactions of GBL receptors with physiological and synthetic DNA substrates.

Introduction

Transcriptional Regulators. Bacteria in the environment are living in unpredictable, competitive, and often extremely variable surroundings. In order to survive and thrive in less than optimal conditions, these organisms have developed adaptive responses that allow them to quickly adjust to the many challenges they face in their surroundings. These adjustments are the result of rapid changes in gene expression, made possible by transcriptional regulators (Ramos et al. 2005).

Transcriptional regulators play a central role in the alteration of gene expression by controlling levels of transcription and relative concentrations of RNA. Transcriptional regulators typically describe proteins that can interact with a specific DNA binding site, usually around the promoter region of a gene, and can act to enhance or repress transcription through interactions with RNA polymerase. Repressor proteins decrease transcription of a gene by binding specific DNA sequences and preventing or impeding RNA polymerases ability to transcribe a complimentary mRNA strand.

Most microbial transcriptional regulators are two-domain proteins, containing one signaling domain and another DNA binding (Henikoff et al. 1988, Hitomi et al. 203, Karinol et al. 2004, Pao et al. 1994, Wosten et al, 2004). Binding in these two domains is mutually exclusive: binding in one domain confers a conformational change that prevents the binding of the second domain. The signaling domain is bound by a small, low-molecular weight, chemical ligand, and unlike the DNA binding domain, the signaling domain lacks conservation between transcriptional proteins reflecting the differences in signals sensed by different regulators (Ramos et al. 2005). The DNA binding domain of a transcriptional protein is often identified by a well conserved

structural helix-turn-helix (HTH) signature. The use of this HTH domain is seen in ~95% of all prokaryotic transcription factor DNA binding domains (Branden and Tooze 1991, Brennan et al. 1990, Harrison 1991, Perez-Rueda et al. 2001, Rousseau et al. 2004).

Named after its most well studied member, one of the largest families of prokaryotic transcriptional repressors is the TetR family. This family contains 85 members which are characterized by an N-terminal HTH motif used in DNA binding (Ramos et al. 2005, Lee et al. 2007, Cho et al. 2003, Dodd et al. 1990). Most repressors belonging to this family have similar molecular masses ranging from 21 to 27 kDA and contain a conserved region beginning four residues before their DNA binding HTH motif and continue past the motif by six residues (Kojic and Venturi et al. 2001). TetR repressors use this conserved HTH domain to bind, as a homodimer, directly to the promoter region upstream of target genes within a symmetric DNA binding site (Engohang-Ndong et al. 2004, MacEachran et al. 2008, Schumacher et al. 2002). These repressors are well established in many prokaryotic species, including *Streptomyces*, to impact gene expression by negatively regulating processes like secondary metabolite production, morphological differentiation, solute transport, and stress response (Ramos et al. 2005, Christen et al. 2006).

The genus *Streptomyces*. *Streptomyces* represents the largest genus of Actinobacteria, a phylum of gram positive bacteria. Actinobacteria are characterized by their high GC content genomes (~70% guanine and cytosine base-pairs) and are found in both terrestrial and aquatic environments. These organisms are saprophytic, playing an important role in the turnover of organic material and the environmental carbon cycle.

Members of the genus *Streptomyces* are among the most common soil bacteria. Research on *Streptomyces* began when it was uncovered that organisms belonging to this genus are responsible for the distinctive odor of dirt through production of the chemical, geosmin. The analyses of the odorous chemical first occurred in 1891 when two French scientists discovered that the substance responsible for the smell could be extracted from a soil sample by steam (Berthelot and Andre 1891). The culprit substance for this odor was extracted from *S. griseus* using gas chromatography and given the name geosmin (from the Greek ‘ge’ meaning earth and ‘osme’ meaning oder) (Gerber et al. 1965). Further research revealed that all members of this genus are capable of geosmin production. This discovery introduced the scientific world to *Streptomyces* and more extensive study on these bacteria began to reveal some interesting facts about their ecology and molecular make-up.

***Streptomyces*: Morphological Differentiation.** One gram of soil can contain over 1.5×10^{10} microbial cells (Torsvik et al 1990). It has been estimated through the use of mathematical models that the number of different bacterial strains contained in this amount of dirt can range from 40,000 (Curtis et al. 2002) to several million (Gans et al. 2005). This massive population of organisms living in one gram of soil is forced to compete for the limited resources available in the surrounding environment. *Streptomyces* species have evolved extraordinarily complex life cycles, by bacterial standards, that make them well adapted to live in such a competitive and quickly changing environment. One important component of this life cycle includes complex morphological differentiation processes.

As *Streptomyces* species enter the vegetative growth phase, they begin to develop tubule structures called substrate mycelium that slowly extended outward and downward into the surrounding environment to gather nutrients. Mycelium is composed of loose hyphae that branch and can rapidly increase in surface area. As nutrients in the area closest to the colony are depleted, the bacterium begins to extend its mycelium farther into the environment to gather nutrients farther from its location. As the resources around the colony are completely depleted, aerial hyphae formation commences and extends upwards above the colony into the air or pockets in the soil.

As aerial hyphae form, the bacterium up-regulates the expression of DNA replication machinery, allowing copies of DNA to migrate up the hyphae structures (Wildermuth 1970). At the end of aerial hyphae formation, the sporulation process begins. This process induces coiling in the hyphal tips which then triggers the formation of a cross-wall (or sporulation-septa) that segregate individual DNA strands into compartments which will become spores (Wildermuth et al. 1970, Ellard et al. 1966, Glauerta et al. 1960, Hoenige et al. 1968). The result is the generation of long, distinctive, filamentous spore chains that can be dispersed by wind, rain, animals, ect. and grow. In his study in 1969, Wildermuth found that in some spores begin to germinate as soon as the sporulation process ends while others do not germinate until they are removed from the spore-chain and relocated to an environment with ample resources.

***Streptomyces*: Production of Secondary Metabolites.** Secondary metabolite production is another piece of their complex life cycle that allow individuals of the genus *Streptomyces* to cope with the competition in their environment. Secondary metabolites

are organic compounds that act as growth inhibitors, microbicidal compounds, antifungals, antivirals, and metabolites that play a role in transport of scarce nutrients (Alvarez & Steinbuchel, 2002). These compounds are not necessary for immediate survival, growth or development, however; loss of these compounds could result in impaired growth and survival over time. The same environmental stresses that initiate the formation of aerial hyphae in *Streptomyces* usually trigger the initiation of secondary metabolite production as well. The role of these secondary metabolite products is to harm nearby organisms who compete with *Streptomyces* for available nutrients (Arabolaza et al., 2008; Olukoshi et al., 1994).

Antibiotic Research: Past and Present. Antibiotic research began in 1928 when Alexander Fleming discovered penicillin. While the story behind his discovery is unclear, it is often told that he discovered the compound when he found a mold growing on agar plates that contained colonies of *Staphylococcus aureus*. He identified a ring of inhibition around the locations where the mold was growing, clearly inhibiting the growth and survival of the *S. aureus* colonies. Intrigued, Fleming performed several experiments that demonstrated that this mold, *Penicillium notatum*, was producing and secreting a chemical, which he isolated, responsible for killing the surrounding bacterial colonies. (Bentley 2005, Hare 1970). The potential medicinal impact of this finding were quickly realized and Fleming received the Nobel Prize in Physiology or Medicine for his discovery.

About 10 years after Fleming's discovery, the first series of successful antibiotic treatments on humans occurred - the potential of naturally produced antibiotics as powerful pharmacological agents was evident (Kallio 2008). Around the world,

researchers began to screen soil samples in hopes of identifying new microbial produced compounds to use as antibiotics. *Streptomyces* was quickly recognized as an abundant producer of these types of chemicals (Kallio 2008).

***Streptomyces* are antibiotic farms.** In 1944, Selman Waksman, the 'father of antibiotics', isolated the aminoglycoside compound streptomycin from *Streptomyces griseus* (Nobelprize.org). This compound was one of the first natural antibiotics discovered that could target and kill *Mycobacterium tuberculosis*, the causative agent of tuberculosis (Hinshaw et al. 1947, Waksman 1944, Waksman 1947). A few years later in 1949, Waksman isolated another aminoglycoside antibiotic capable of targeting *M. tuberculosis*, neomycin, from *Streptomyces fradiae*. Today neomycin is commonly found in common topical preparations such as Neosporin (Waksman et al. 1949, Rose et al. 2007, Iverson et al. 1952).

The discovery of these two medicinally valuable compounds opened the floodgates on *Streptomyces* research. In 2002 it was estimated that out of ~22,000 known naturally produced clinically relevant antibiotics, ~7,600 were derived from *Streptomyces* species, meaning that approximately a third of all microbial bioactive compounds come from this genus (Berdy 2005). Because of their pharmacological potential *Streptomyces* has continued to be a well-studied genus.

Molecular Background of *Streptomyces* genus. As our understanding of molecular processes increases along with development of techniques to study them, researchers have begun to strive to better understand the origin of complex life-cycle processes like secondary metabolism and morphological differentiation. It was quickly uncovered that the *Streptomyces* chromosome is unusual in both structure, having a

linear chromosome rather than the traditional circular chromosome found in prokaryotes, and size. These apparent oddities were first discovered in a pulsed-field gel electrophoresis (PFGE) experiment using DNA from *Streptomyces lividans*. This finding has since been confirmed in a number of other species from this genus (Lin et al. 1993, Gravius et al. 1994, Lezhava et al 1995, Leblond et al 1996).

A typical *Streptomyces* chromosome is around 8 Mb and replicates bidirectionally from a centrally located origin of replication toward the telomeres in a process involving terminal protein mediated circularization (Calcutt et al 1992, Kieser et al 1992, Zakrzewska et al 1992, Kallio 2008). The *Streptomyces* genome is exceptionally large containing approximately twice the number of open reading frames than found in the gram-positive soil bacterium *Bacillus subtilis*. To date its one of the largest bacterial genomes known ranging from 6.5 Mb in *S. ambofaciens* to 10 Mb in *S. scabies* (Leblond et al 1990, Goffeau et al 1996, Kunst et al, 1997). The unusually large size of this bacterial genome allows *Streptomyces* species to undergo the complex life cycle processes necessary for adapting and surviving in competitive soil environment.

Genomic Distribution of *Streptomyces*. The arrangement of genes on the *Streptomyces* linear chromosome is another distinctive characteristic of the genus. The central portion of DNA (~6.5Mb long) is highly conserved among *Streptomyces* and contains most of the essential and housekeeping genes, including genes that encode: DNA replication machinery, cell division, and transcription and translation machinery (Ikeda et al. 2003, Karoonuthaisiri et al. 2005, and Volff et al. 2000). The telomeric regions of the chromosomes are more variable and less conserved. Vloff et al. 1998 found that extensive deletions and amplifications of the telemetric ends do not affect the

viability of the organism *in vitro*. Based on this knowledge, it was initially assumed and now confirmed that these regions of the chromosome tend to hold the genes responsible for secondary metabolism processes.

Genes responsible for secondary metabolism processes are often grouped together based on the specific biosynthetic pathway they encode. These groupings of genes are often referred to as 'biosynthetic gene clusters', and the genes found in this cluster usually represent all the genes required to construct a specific metabolite product. The presence of these biosynthetic clusters in *Streptomyces* species simplifies detection and isolation of complete biosynthetic pathways and are extremely useful to researchers. By understanding these pathways at a molecular level, scientists can work to genetically engineer organisms to produce antibiotics with desired structural changes, increasing the amount of useful antibiotics available and overcoming antibiotic resistance.

Evidence for the 'biosynthetic gene clusters' system was revealed in 2002 as the first *Streptomyces* genome, *Streptomyces coelicolor* A3(2), was sequenced from an ordered cosmid library. The sequence data revealed 20 clusters that code for enzymes characteristic of secondary metabolites in the ends of the chromosomal arms, including the antibiotic actinorhodin (Bentley et al, 2002). The actinorhodin producing gene cluster had previously been identified, but the availability of the complete genome sequence allowed for an in depth analysis of this pathway as well as an analysis of the location and interaction of this cluster relative to the other 19 found (Hopwood et al., 1997, Chong et al. 1998). Today several *Streptomyces* genomes have been sequenced, *S. avermitilis* was completed in 2003 uncovering 30 biosynthetic pathways and *S. griseus*,

completed in 2008, revealed 34 biosynthetic gene clusters, including the cluster of the well established antibiotic streptomycin (Ikeda et al, 2003, Ohnishi et al, 2008).

Production of Polyketide Secondary Metabolites. A large portion of secondary metabolite products in *Streptomyces* are assembled by a polyketide synthase (PKS). The role of the PKS is to construct, from a starter unit, a small organic compound, a polyketide, which is to be secreted by the organism at the onset of sporulation and can target and inhibit the growth of neighboring competing organisms (Davis et al. 1990, Jez et al. 2000). Polyketides are often characterized as antibiotics, antifungals, insecticides, antimicrobials, and so on (Bedford et al. 1995, Funa et al. 1999, Sherman et al. 1989).

The term polyketide was coined in 1893 by John Norman Collie and was used to identify a natural product that contained multiple carbonyl or hydroxyl groups that were separated by at least one carbon atom (Collie et al. 1983, Bentley et al. 1999). Today this term is defined as a carbon skeleton formed by iterative decarboxylative condensations of malonic acid thioesters and assembled by a PKS (Bu'Lock et al. 1979, Bu'Lock 1961, O'Hagan et al. 1991).

PKS Assembly Derived from Fatty-Acid Biosynthesis. It has been well established that the production of polyketides very closely mimics that of fatty-acid biosynthesis (Hopwood et al. 1990, Mayorga et al. 1992). Two types of PKSs are categorized based on architectural similarities to type I or type II fatty acid synthases. The most defining difference between polyketides and fatty acids is the enormous amount of diversity found in polyketide structure, a consequence of post-PKS

modifications and variability of PKS enzymes involved in production (Hutchinson et al. 1997, Katz et al. 1993).

Type I PKSs consist of large modular proteins, usually greater than 300 kDa, that contain numerous active site domains. These domains are grouped into modules that control the addition of starter units to the growing chain and the subsequent modifications of the products (**Figure 1**, Seow et al. 1997). Examples of compounds produced by this type I PKSs are erythromycin, avermectin, rapamycin, and soraphen (Cortes et al. 1990, Donadio et al. 1991, MacNeil et al. 1994, Schwecke et al. 1995, Schupp et al. 1995) Type II PKS complexes are aggregates of three to five smaller mono- or bi-functional polypeptides whose active sites are used iteratively throughout the biosynthetic process of the polyketide chain assembly (Khosla et al. 1999, O'Hagan et al. 1991). Type II PKS produce mainly aromatic compounds, i.e. actinorhodin and tetracenomycin (Fernandez-Moreno et al. 1992, Bibb et al. 1989, Summers et al. 1992, Summers et al. 1993). In 1999 a new class of PKSs was discovered, type III PKSs. Type III PKSs are distinct from both type I and II PKS, and are classified as members of the chalcone synthase (CHS) and stillbene synthase superfamily (Funa et al. 1999, Bangera et al. 1999, Schroder et al. 1999, Moore et al. 2001).

The Minimal PKS. Type II PKSs contain a complex of three genes encoding the "minimal PKS". This minimal PKS is responsible for constructing the initial polyketide product and includes two β-ketoacyl synthase subunits, KSα and KSβ or chain length factor (CLF), and an acyl carrier protein (ACP) (McDaniel et al. 1993, Kim et al. 1995, Rajgarhia et al. 1997). The growth of a polyketide chain from a type II PKS occurs through iterative Claisen condensations of a starter unit, usually acetyl-CoA or

propionyl-CoA, with extender units of malonyl-CoA and methylmalonyl-CoA (Rawlings et al. 1999, Shen et al. 2000, Reeves et al. 2003, Moore et al. 2002). Each subunit of the minimal PKS plays an invaluable role in the production of a specific polyketide.

The KSα subunit, which dimerizes with the KSβ subunit, is responsible for catalyzing the condensation between starter and extender units while KSβ plays a role in determining the length of the polyketide chain backbone. Removal of the KSβ or KSα genes from an organism results in inactivation of the minimal PKS (Seow et al. 1997). The ACP is a small holoprotein that contains a pantethenic cofactor site allowing for priming and transfer of individual malonate extender units (through thioesterification) to generate nucleophiles for the Claisen condensations performed by the ketosynthase-α subunit (Khosla et al. 1993, Wakil et al. 1989). Secondary metabolite products produced by type II PKSs are generally classified as anthracyclines, angucyclines, aureolic acids, tetracyclines, and tertacenomycins (Hertweck et al. 2006).

Post-PKS Tailoring Enzymes. A source of enormous variation in the polyketide products of PKS gene clusters is the presences or absence of various post-PKS tailoring enzymes within the biosynthetic pathway (Remsing et al. 2001). These enzymes are often found in the immediate vicinity of the minimal PKS and allow for modifications to be made to the carbon chain created there. Some of the most common tailoring enzyme processes are: hydration, dehydration, cyclization, oxidation, reduction, methylation, glycosylation, and decarboxylation (Bentley et al. 1999). Genetic engineering of the genes that encode minimal PKS subunits or genes that control post-PKS tailoring enzymes could create novel pathways and therefore novel polyketide

products. The ability to engineer these novel natural products could be extremely beneficial in helping to develop antibiotics to overcome bacterial resistance.

γ-Butyrolactone Regulation of Secondary Metabolism and Morphological Differentiation. To cope with the fierce competition for nutrients and space in their soil habitat, *Streptomyces* have developed a complex life cycle including morphological differentiation processes as well as the production of secondary metabolites products. These two processes often act together in order to maximize the organism's ability to successfully gather enough nutrients, to grow, and then sporulate. Secondary metabolite production typically occurs after the onset of sporulation, as aerial hyphae begin to extend and allow for the organism to kill neighboring competitors in order to acquire larger amounts of available nutrients. The morphological differentiation and secondary metabolite production processes are both extremely important to the survival of the organism and each must occur within a strictly defined window of time during growth to be effective. In order to ensure that these processes occur within the correct time frame, they are tightly regulated through the interaction of low-molecular weight, hydrophobic, membrane diffuse-able gamma-butyrolactone (GBL) auto regulator compound with specific GBL receptor proteins (Healy et al. 2009).

Identification of GBLs as regulators of antibiotic production and morphological differentiation in *Streptomyces* began with the identification of A-factor, a GBL responsible for regulating production of streptomycin production and sporulation in *S. griseus* (Khokhlov et al. 1967). In their study, Khokhlov et al (1967). identified a small chemical transcription factor that could, in nanomolar concentrations, alter streptomycin production and aerial hyphae formation in mutant strains of *S. gresius*. This chemical, A-

factor, is produced by a GBL synthase in a growth-dependent manner as a part of the growth cycle of the organism, reaching its maximum concentration (~100 nM) in the middle of the exponential growth phase. A-factor's GBL receptor protein, ArpA, is also produced in higher concentrations as the organisms grows.

This transcription factor contains two highly specific binding domains, one for A-factor and another for a specific DNA binding site (referred to as an auto-regulatory Element or ARE) that can be bound by the receptor protein. As ArpA is produced it binds to an ARE site in the 5' promoter region of the *adpA* gene and prevents transcription from occurring by blocking RNA polymerase access to the promoter. *adpA*, when transcribed, produces a transcription factor that activates multiple genes necessary for the production or secondary metabolites (like streptomycin) and morphological differentiation, thus when bound by ArpA, secondary metabolite production ceases and morphological differentiation is altered. At its maximum concentration, A-factor moves through the cell and binds to ArpA's GBL binding domain. When bound by A-factor, the protein undergoes a conformational change that decreases binding affinity at the ARE site and dissociates the protein-DNA complex, allowing transcription to occur (Horinouchi et al 2007, Ohnishi et al 2004, Ohnishi et al 2005).

Fourteen different A-factor homologs in seven different *Streptomyces* species, having a GBL-like structure and regulatory function have since been identified (reviewed in Takano et al, 2006). To date there are 14 identified GBL receptor proteins whose target ARE sequence has been characterized (**Figure 23**).

The Discovery of *Streptomyces acidiscabies*. *Streptomyces acidiscabies* was first identified as a new bacterial species in scientific literature in 1989. Isolated from a

diseased potato, it quickly became clear that this organism was a plant pathogen and a causative agent of potato scab. Potato scab is characterized by tuber lesions on the skin of the potato, either superficial or pitted, which do not harm the starch content of the vegetable but give it an unfavorable appearance and make them hard for farmers to sell (Manzer et al. 1977).

There are two distinct types of the scab disease: common scab and acid scab. These two scab types, while indistinguishable based on their appearance on the infected potato, are differentiated by both the *Streptomyces* strain responsible for formation as well as the surrounding soil environment (Takeuchi et al, 1996, Lambert et al. 1989). *Streptomyces scabies* predominates in causing common scab disease and is usually found in soil above pH 5.2 while *Streptomyces acidiscabies* strains are responsible for the acid scab disease found in soil with a pH between 4.5 and 5.2 (Lambert et al. 1989, Bonde et al. 1968, Mazer et al. 1977, Hughes et al. 1971, Loria et al. 1986). While the symptoms caused by both of these organisms are indistinguishable, the two species are easily distinguishable in culture and on growth media.

Virulence Factors of *S. acidiscabies.* Organisms within the *Streptomyces* genus are infrequent pathogens, in fact only four to date have been identified as pathogens and each is a causal agent of scab disease in various plants (Loria et al. 1997, Miyajima et al. 1998). The mechanism of pathogenicity in these strains was first identified as thaxtomins, these chemicals were isolated from extracts of *S. scabies*-infected potato tissues (King et al. 1989, Lawrence et al. 1990). Work done by Healy et al. 2000 determined that two peptide synthases genes (txtA and txtB) are responsible for producing thaxtomin A in *S. acidiscabies* 84.104 through the transition to the

avirulent strain during gene disruption assays. Determination and isolation of the bacterial toxin was done through screening cultures in different mediums. The thaxtomin A toxin was found to be secreted in starch medium. A novel secondary metabolite product was also found during the screening process. This polyketide, WS5995B, was produced in the mannitol rich medium, SGM, *Streptomyces* growth medium (**Figure 2**)

Characterization of a GBL regulated Type II PKS in *S. acidiscabies* (Healy et al. 2009.) In order isolate the type II PKS responsible for WS5995B production in *S. acidiscabies* 84.104, degenerate PCR primers were designed to target and amplify the unique 5' end of genes encoding regulatory GBL receptor proteins. Amplification with these primers resulted in isolation of a ~150 bp fragment that displayed a high degree of similarity to previously characterized GBL receptor proteins. This isolated fragment was used to fish out the remaining genomic sequence as well as the sequence of flanking genes. The complete sequence was identified, through sequence analysis, as the GBL receptor gene, SabR. The flanking sequences that were uncovered through a radio-labeling assay were GBL synthase, *sabA*, and GBL receptor protein, *sabS*. SabR and SabS both contained distinct N-terminal helix-turn-helix motifs, hallmarks of GBL receptor proteins. Protein sequence alignments using SabA revealed a high degree of similarity with other previously characterized GBL synthases (Kato et al. 2007).

In-frame deletion mutants of each gene in the SabRAS cluster were created in order to determine the roles of each gene in secondary metabolism processes in the bacterium. Based on pigment production assays with these mutant strains it was observed that the absence of *sabS* and *sabA* genes resulted in an increase in the

production of WS5995B, while deletion of *sabR* did not alter production noticeably from the wild type phenotype (**Figure 3**). *sabS* and *sabA* mutants grown in liquid culture and filtered using solid-phase methods also showed this pigment overproduction trend. It was also observed that *sabS* mutants, when grown on ISP2 media, altered the normal morphological growth of the colonies - failing to produce aerial hyphae.

Pattern searching techniques identified an ARE site in the 5' promoter regions of *sabS*. This ARE site was bound by purified SabR protein in mobility gel shift analyses. No ARE regions bound by SabS were determined. The pathway of action based on the data gathered by Healy et al. 2009 suggests that SabR binds an ARE site in the 5' region of SabS and represses its expression. To confirm these finding, qRT-PCR analysis of *sabS* expression in wild-type and *sabR* mutant strains was completed. The result of these tests clearly demonstrated that relative expression of *sabS* was greater in the *sabR* mutant strain (**Figure 4**).

The hypothetical pathway for biosynthesis and regulation of secondary metabolism and morphological differentiation in *S. acidiscabies* 84.104 can be seen in **Figure 5**. It appears that SabR acts as a negative regulator of SabS by binding the ARE site in the 5' promoter region of the gene. This repressor is dissociated from the complex when bound by a GBL produced by SabA, allowing for expression of the sabS transcription factor protein. Expression of SabS is able to decrease the production of the WS5995B polyketide and has an impact on morphological differentiation processes.

Materials and Methods

Bacterial Strains and Plasmids. All bacterial strains and plasmids used in this study are listed in **Table 1**. BL21λDE3 *E. coli* strains containing pET25b expression vectors were grown in Luria-Bertani (LB) Broth supplemented with kanamycin (50 µg/mL, Sigma). E. coli NEB5-α and TOP10 were used primarily during cloning experiments. Transformants were grown on LB agar medium supplemented with ampicillin (100 µg/mL, Sigma) and 40 mg/mL X-gal in Dimethylformamide (Fischer Scientific).

Bacterial Media. LB broth: 10 g/L tryptone, 5 g/L yeast extract, 5 g/L sodium chloride. 15 g/L granulated agar (Difco) added for plates. All media supplemented as required with the following antibiotics for plasmid retention selection: kanamycin (Sigma) 50 µg/mL; and ampicillin (Sigma) 100 µg/mL.

Generation of a Genomic DNA Library. Genomic DNA was extracted from *S. acidiscabies* 84.104 cell cultures following the genomic DNA extraction protocol used by Healy and Lambert 1991. Extracted genomic DNA was amplified in a 50 µL PCR reaction by random N_9 oligonucleotide primer 229 and the Klenow fragment (**Table 2**). This amplification was performed using the (GenSelex) program (**Table 3**). 2 µL aliquots of these amplification products were subsequently re-amplified in five separate 50 µL reactions, with 25 µL of Taq Master Mix (Qiagen), using the GenSELEX program with primer 230. (**Tables 2 and 3.**) At each step, amplification products were confirmed through gel electrophoresis, loading 5 µL aliquots into a 2% agarose gel running at 100 volts in 1X TAE buffer. Each individual reaction containing amplified DNA after the second round of PCR was purified using Mini-elute PCR purification kits (Qiagen). The

purified amplification products were pooled and the concentration was determined using a NanoDrop Spectrophotometer. Pooled products were then concentrated to ~30 µL in a Speed Vacuum (ThermoSavant).

Generation of a Synthetic DNA Library. Random pools of double-stranded synthetic DNA substrates were generated through PCR amplification of a degenerate oligonucleotide (N_{28}) template by primers 225 and 226 (**Table 2**). Eight separate 50 µL PCR reactions, each containing 2 µL DNA template, were amplified using the SELEX2 program to set up the initial, N_0, synthetic library (**Table 3**). After amplification 5 µL aliquots of each reaction were analyzed through gel electrophoresis (2% agarose gel) running at 100 volts. Amplification reactions that contained the expected ~120 bp product on the analytical gel were individually purified, pooled, and concentrated following the same protocol used to generate the genomic library.

Expression of recombinant SabR and SabS proteins. GBL receptor genes, *sabR* and *sabS,* were cloned into the expression vector pET25b (pET25b::*sabS* or pET25b::*sabR)* and expressed in BL21λDE3 strain *Escherichia coli* following the protocol used by Healy et al. 2009 (**Table 1**). Cell cultures containing plasmids were grown up in 100mL cultures of LB broth containing kanamycin. Absorbance measurements were taken every 30 minutes by measuring a 0.7 mL aliquot of each culture in a spectrophotometer (250 nm). At ~0.6 nm absorbance, overexpression of the gene contained within the vector was induced through the addition of 1 mM IPTG (Gold Bio Technologies Inc.). Prior to addition of the inducer, a 1 mL aliquot of each culture was removed, pelleted and stored at -20°C. After 12 additional hours of growth, another 1 mL aliquot from each induced culture was removed and stored. Over-

expression of the protein was observed by comparing pre- and post-induction aliquots through sodium dodecyl-sulfate polyacrylamide gel electrophoresis (SDS-PAGE). Cultures that demonstrated over-expression of either *sabR* or *sabS* were pelleted and stored at -80°C.

Purification of recombinant SabR and SabS proteins. IPTG-induced BL21λDE3/pET25b cell pellets were retrieved from -80°C and resuspended in 10 mL of 50 mM Tris-Cl ph 7.5. Resuspended cells were washed and at 4°C for 10 minutes at 8000 rpm in an Avanti J-20 XP centrifuge using the JA25.50 rotor (Beckman-Coulter). Washed cells were resuspended in 1.5 mL of buffer 1 (20mM imidazole PO_4 (GE Healthcare), 1 mM dithiothreitol (Sigma)), followed by the addition of ~1 mg DNAse I (Sigma) and 2 mg PMSF (Fisher Scientific) to the suspension. Cell suspensions were disrupted by passage through a French pressure minicell (Thermo) at 16,000 lb/in². The lysate was centrifuged for 5 minutes at 4°C. The supernatant was loaded onto a 1-mL HisGraviTrap column (GE Healthcare) and washed consecutively with 10 mL of buffer 1, and twice with buffer 2 (100 mM imidazole PO_4 (GE Healthcare), 1 mM dithiothreitol (Sigma)). Protein bound to the column after the wash steps was eluted with 10 mL of buffer 3 (500 mM imidazole PO_4 (GE Healthcare), 1 mM dithiothreitol (Sigma)) and collected in eight 0.5 mL fractions. Collected fractions were analyzed using 15% non-denaturing SDS-PAGE. The five fractions containing the highest concentrations of SabR or SabS were pooled and loaded onto a PD-10 desalting column (GE Healthcare). Eight 0.5 mL fractions were collected in 4 mL 2X Storage Buffer (100 mM Tris-Cl pH 7.4, 400 mM NaCl (Fischer Scientific), 40 mM KCl (Fischer Scientific), 2 mM $MgCl_2$ (Fluka BioChemika)) and analyzed using 15% non-denaturing SDS-PAGE. The five fractions

with the highest SabR or SabS concentration after desalting were pooled and stored at -20°C in 1:1 glycerol (Fischer Scientific).

Set up of SELEX assay. Ni-NTA agarose resin beads (Qiagen) were washed three times successively with 1 mL of SELEX binding buffer (SBB) for 1 minute at 4°C. SELEX binding buffer: 1 mM Tris-Cl: pH 7.4, 150 mM NaCl, 20 mM KCl, 1 mM $MgCl_2$, 1 mM DTT. Purified recombinant SabR or SabS protein (~100 µL) and 300 µL SBB were added to washed Ni-NTA agarose and placed on a Rollordrum rotator (New Brunswick Scientific Co.) at a power setting of 5 for 30 minutes at a temperature of 4°C. Excess unbound protein was removed through three successive washes with 1 mL SBB. Concentrated synthetic or genomic DNA libraries and 300 µL SBB were added to protein-resin complex at room temperature and spun for 30 minutes at 5 power setting on a Rollordrum rotator. Excess unbound DNA was removed from the complex through three successive one minute washes with 1 mL SBB at room temperature**.** A final 1 minute room temperature spin allowed for the removal of excess DNA and buffer from the agarose:protein:DNA complex. DNA was eluted from the complex with 30 µL of buffer 3 and collected from the supernatant after centrifugation for 1 minute. Eluted DNA was PCR amplified in five separate 100 µL reactions (containing 2 µL of SELEX product) and analyzed by electrophoresis in a 2% agarose gel running at 100 volts. Amplification products were purified using the Mini-elute PCR purification kit and collected in 25 µL of elution buffer. Individual reactions were pooled and the relative nucleotide concentration of the sample was determined by a NanoDrop Spectrophotometer. Samples were then concentrated into ~30 µL in a Speed Vacuum

prior to use in the next round of SELEX. SELEX was repeated at least five times for synthetic sequences and at least three times for genomic sequences.

Counter-Selection of Synthetic Libraries. Synthetic libraries underwent counter-selection prior to DNA addition during the fourth round of SELEX. This step acts as a negative control for non-specific DNA:Ni-NTA agarose binding. The concentrated synthetic DNA library after three rounds of SELEX and 300 µL of SELEX binding buffer were added directly to 65 µL of washed Ni-NTA agarose beads and allowed to incubate at room temperature for 30 minutes on a Rollordrum rotator at a power setting of 5. Unbound DNA was collected from supernatant after centrifugation for 2 minutes at room temperature and used in the fourth round of SELEX.

Cloning and Sequencing DNA sequences bound by SabR or SabS. Genomic and synthetic sequences recovered after the final round of SELEX were cloned into the plasmid vector pCR 2.1 Topo (**Figure 6)** using the Topo TA cloning kit (Invitrogen). Clones containing inserts from synthetic DNA recovered from SabS and from genomic DNA recovered from both GBL receptor proteins was transformed into electro-competent TOP10 *E. coli* cells using the TA cloning kit One Shot Electroporation protocol. Clones containing inserts from synthetic DNA bound by SabR were transformed into New England BioLabs NEB-α chemically competent *E. coli* cells following NEB High Efficiency Transformation protocol. In all cases, transformed cells were grown overnight at 37°C on LB agar plates containing ampicillin and X-gal. Twelve to twenty white colonies and one blue negative control colony were selected and used to inoculate 1.5 mL of LB broth supplemented with ampicillin. Inoculated cultures were placed onto a Rollordrum rotator at power setting 5 and grown overnight at 37°C . After

12 hours, 5μL aliquots of each inoculated culture were lysed in 50 μL of DNAzol Direct (Molecular Research Center) and 2 μL of each lysate was amplified by PCR using primer 230 for genomic sequences and primers 225/226 for synthetic sequences.. Approximately 5 μL of each amplified reaction was run in an analytical 2% agarose gel to determine the relative size of cloned inserts. Plasmids were selected for extraction from inoculated culture based on the size of the insert. Plasmids were extracted using a High-Speed Plasmid Mini Kit (IBI). Relative nucleotide concentrations of extracted plasmids were determined and sent for sequencing at the University of Texas Institute for Cellular and Molecular Biology (UT-ICMB) core.

Analysis of Recovered Sequences. Sequences returned from UT-ICMB were aligned in a text editor program. Once aligned, a sequence logo of these sequences was created using the multiple sequence alignment function of the WebLogo program. (http://weblogo/berkeley.edu).

Analysis of GBL residues involved in DNA binding. Previously identified GBL receptors and their corresponding ARE sites are given in **Figure 23**. The amino acid sequences of each of these proteins were retrieved from the NCBI protein database. Sequences were aligned against TetR using the multiple sequence alignment function of ClustalW. The residues involved in DNA contact in TetR have been elucidated since the crystal structure of the protein has been solved (Hinrichs et al. 1994). Alignment of identified GBL receptors with TetR allowed for the determination of the residues most likely used in DNA binding.

Results

Purification of GBL Receptor Proteins, SabR and SabS. SabR and SabS were purified from a BL21λDE3 *E. coli* culture containing pET26b::*sabS* or pET26b::*sabR* expression vectors. Over-expression of recombinant protein occurred when the culture reached an optical density of ~0.6 nm and was induced by addition of IPTG. (**Figure 7**). The relative amounts of recombinant protein in cultures after IPTG induction was determined through analysis of a 15% acrylamyde, denaturing, discontinuous SDS PAGE (**Figure 8, 9**). Polyhistidine-tagged SabS and SabR were isolated from other cellular proteins through a HisGraviTrap column and fractions containing the protein of interest were determined utilizing SDS-PAGE (**Figure 10, 11**). Selected fractions were then pooled, desalted and stored at -20°C (**Figure 12, 13**). Stored proteins were then used in each round of SELEX assay (**Figure 14**).

Optimization of the SELEX assay. The amplification step of the SELEX assay and subsequent electrophoretic analysis of amplified products provided insight into the relative size of enriched DNA sequences. Through the first three rounds of SELEX with SabS, enriched synthetic products were exclusively found at the expected ~120 bp size range. However, after four rounds of SELEX amplification products were found at the expected size as well as at molecular weights up to 650 bp (**Figure 15**). An analysis of the effect of two variables on the amplification process, cycle number and annealing temperature, on the presence of these unwanted, higher molecular weight products was completed. Archived samples of SabS enriched synthetic DNA libraries from SELEX rounds three and five were used as DNA templates to create five identical 50 µL PCR

reactions for each template. PCR reactions were amplified using the SELEX1 program (**Table 2**). During the amplification process one tube from each library was removed from the thermal cycler after rounds 14, 18, 22, 26, and 30. A 5 µL aliquot from each reaction was analyzed through gel electrophoresis. The reaction containing exclusively the products of the expected size was removed after 14 cycles (**Figure 16**). Based on these results, the SELEX1 program was altered to repeat each cycle only 14 times rather than 30.

In order to determine the optimal annealing temperature, a new N_0 synthetic DNA library was generated and used for one round of SELEX. During the amplification step, eleven 50 µL PCR reactions were set up and placed into a Biometra Tgradient thermal cycler with annealing temperatures ranging from 45°C to 84°C. Electrophoretic analysis of amplification reactions revealed exclusively products of the expected size at each temperature (**Figure 17**). The annealing temperature of the amplification step of the assay was then set to 80°C rather than 52°C.

SELEX identification of synthetic sequences bound by SabS and SabR. A synthetic library was generated through amplification of a degenerate oligonucleotide (N_{28}) template, yielding a library containing ~10^{16} different sequences **(Figure 18**). To identify sequences bound by the two GBL receptor proteins, recombinant SabS or SabR was bound to Ni-NTA agarose and incubated with synthetic DNA substrates over at least five iterative rounds of SELEX.

Products bound by His_6-SabS after the 7th round were recovered, cloned into plasmid vector pCR 2.1 Topo, and sequenced (**Figures 19 and 20**). From 20 cloned inserts, nine unique sequences were identified (**Figure 21**). The

corresponding sequence logo of these retrieved sequences highlights the similarity between all of the retrieved sequences (**Figure 22**). These nine unique sequences also share significant similarities with previously identified ARE sites (**Figures 23 and 24,** Healy et al. 2009).

A new, N_0, synthetic DNA library was generated and enriched through binding to His_6-SabR over 5 consecutive rounds of SELEX (**Figure 25**). These products have been recovered and are in the cloning and sequencing stage of the assay.

SELEX identification of genomic sequences bound by SabR and SabS. Genomic libraries incubated with both GBL receptor proteins were taken through three rounds of SELEX (**Figure 26**). Enriched genomic sequences eluted from SabR or SabS after the third round of SELEX were subsequently cloned into the pCR 2.1 Topo vector. The approximate size of genomic inserts was determined through enzymatic digestion of extracted plasmids containing cloned inserts with *Hind III* (**Figure 27**). In an attempt to recover sequence data from all unique genomic DNA enriched during SELEX, plasmids to undergo sequencing were selected based on variability of size after enzymatic digestion. Four clones representing each GBL receptor protein were sent to the UT-ICMB core for sequencing. Based on a BlastX analysis of sequenced products, three out of four sequences bound by both SabR and SabS were identified as *E. coli* DNA. The last sequence bound by SabR returned a low quality sequence not analyzed here and the last sequence bound by SabS was similar to sequences from *S. scabies*. (**Figures 28 and 29**).

Analysis of residues important in GBL receptor protein DNA binding. There are currently 14 identified GBL receptor proteins whose target ARE sequences have been characterized (**Figure 23**). The amino acid sequences for each of these proteins was retrieved from NCBI and aligned in the seqeuence alignment program ClustalW. These sequences were compared to TetR, a member of the TetR repressor family whose crystal structure has been solved (**Figure 30,** Hinrichs et al. 1994). The HTH motif has been well charcterized in TetR which allowed for the identification of residues in the HTH motif of other identified GBL receptors that play a role in DNA bidning within their specified ARE site (**Table 4**).

Discussion

Optimization of SELEX. SELEX is a well-established molecular technique, however; optimization of this assay had to occur in order to isolate and identify enriched sequences with *S. acidiscabies* GBL proteins and synthetic and genomic DNA (Stoltenburg et al. 2007, Ellington et al. 1990, Tuerk et al. 1990). The PCR amplification step of the assay was targeted for customization to our system after the appearance of high molecular weight products began to appear on analytical gels containing aliquots of the amplified synthetic library, enriched after several rounds of SELEX with *sabS*.

Cycle number and annealing temperature represent the two most variable steps in the PCR process. Previously successful use of the SELEX assay in identifying DNA binding sites of a protein ligand used annealing temperatures ranging from 50°C to 62°C and cycle numbers extending from 8 to 35 cycles (Tsai and Reed 1998, Wang et al. 1998, Mohibullah et al. 1999). Knowing that a wide range of cycle numbers and annealing temperatures had been used successfully in identifying DNA binding sites we decided to perform several experiments that would determine the optimal annealing temperature and cycle number that would amplify exclusively the expected sized products and eliminate any higher-molecular weight products. Our results here show that 14 cycles is the optimum for exclusively isolating the expected ~120 bp synthetic ARE site plus fixed primer regions. The annealing temperature used to isolate and enrich for the desired sized product is quite variable and can range from 65°C to 83°C. We chose to use 80°C as our annealing temperature to ensure that primers were able to form a stable association with denatured target DNA.

Identification of synthetic sequences bound by SabS. After undergoing seven rounds of SELEX, the final enriched library was cloned into the 2.1 Topo vector and transformed into electro-competent TOP-10 *E. coli* cells. The plasmids of twenty colonies that exhibited a white phenotype indicating successful insertion of synthetic DNA were extracted and sent to UT-ICMB core for sequencing. Of these twenty sequences, nine unique sequences were returned (**Figure 21**). One of the nine unique SELEX generated sequences, 5'-GTGAAAACAGGTCACGCGCCCCGTTTTG -3', was isolated from seven out of the twenty sequenced clones. Because this clone was recovered from almost half of our sequenced clones we consider this to be representative of the synthetic ARE site bound by SabS. This also represents the only sequence that did not exhibit partial binding of the 3' region of the 5' fixed primer by SabS.

Creation of a sequence logo for synthetic sequences bound by SabS. The nine unique sequences bound by SabS were aligned in a text editor and submitted into the WebLogo program created by University of California at Berkeley. This program returned a sequence logo demonstrating a consensus sequence that took into account all nine unique aligned SELEX products (**Figure 22**) The sequence logo shows two extremely important and equally exciting results. The logo indicates a high level of consensus between each of the sequences used to create it. This result established that our newly customized assay had allowed us to, over seven iterative rounds of SELEX, isolate specific DNA sequences that share commonalities that promote their ability to be bound by *sabS* and enrich a synthetic library for the presence of only these sequences over time.

Comparing SabS enriched synthetic sequences with previously identified ARE sites. Using the newly formed sequence logo we were able to directly compare our synthetic logo and a logo representing previously identified ARE sites (**Figure 24**). Similarities between both logo sequences demonstrates that the DNA enriched in our library not only look similar to each other, but they also hold a high degree of consensus with other previously identified ARE sites. The SabS synthetic sequence logo, like that of previously identified ARE sites, is an AT rich sequence, a defining characteristic of ARE sites. Another hallmark of ARE sites is the symmetry of the sequence. The 5' region of the ARE site tends to contain high adenine concentrations (~7 bp) while the last 5 to 7 base pairs of the site contain mostly thymidine. This symmetry may allow for the GBL regulator to as a bind homodimer. This symmetry can be observed in the logo created from our synthetic sites as well as the one constructed from previously identified ARE sites. Another feature of identified ARE sites is the presence of a few cytosine and guanine base pairs in the middle portion of the sequence, especially an almost invariable CnG pattern found immediately prior to the 3' string of thymidine base pairs. This pattern is seen in both logos.

Using SELEX to identify synthetic ARE sites bound by SabR. At this point no synthetic ARE sites bound by *sabR* have been identified. An initial N_0 library was taken through 5 iterative rounds of SELEX and the final enriched library was cloned. From these clones, four were chosen for plasmid extraction and sent to UT-ICMB for sequencing. However, unlike sequences bound by *sabS*, sequences returned contained high proportions of guanine and cytosine. Analysis of the electropharogram file of each sequence revealed high quality sequences for the vector and primer regions of each

sample but the multiple cloning site contained several overlapping sequences (**Figure 31**). It appears that multiple transformed clones were selected during the inoculation process. The high GC content of the sequences found in the MCS could be attributed to selection of small ampicillin sensitive satellite colonies that were able to grow on the plates. It is apparent in several of these files that one or more of the overlapping sequences is fairly AT rich and could contain an ARE site enriched though the SELEX assay. The sequences bound after 5 rounds of SELEX will be re-amplified and re-sequenced in an attempt to identify these AT rich background sequences.

Using SELEX to identify genomic ARE sites bound by both GBL receptor proteins. In order to uncover the ARE site bound by the transcriptional regulators SabS and SabR, genomic libraries were generated using random-N_9 primers. The use of SabR, which binds the ARE site in the promoter region of *sabS*, in this assay is a positive control. The first attempt to identify ARE sites bound by these receptor proteins resulted in eight sequences (four clones from each protein) that were identified through a NCBI BlastX search as *E. coli* sequences. These results were attributed to the possible generation of a *E. coli* library rather than one containing *S. acidiscabies* DNA. In order to test this hypothesis, a new DNA library was constructed from pelleted 84.104 cultures.

New genomic libraries underwent three rounds of SELEX with SabR and SabS prior to cloning and sequencing. Four phenotypically selected clones for each GBL receptor were sent to UT-ICMB for sequencing, and the returned sequences were analyzed using a NCBI BlastX search. This search revealed that three of the the four clones for each protein contained an insert that was identical to DNA found in *E. coli*.

The fourth sequence bound by SabR was a low quality sequence that was not analyzed, while the fourth sequence bound by SabS was similar to a conserved hypothetical *S. scabies* protein when translated.

We hypothesize several possible reasons for the recovery of *E. coli* sequences after the second set of SELEX assays. The generation of an *E. coli* library is possible but is made much less probable by the acquisition of a *S. acidiscabies* sequences similar to *S. scabies*. Unlike the *S. scabies*, the genome of *S. acidiscabies* has not been sequenced and would not be a possible BLASTX match. While it has been established that *S. acidiscabies* is more closely related to other *Streptomyces* strains than *S. scabies*, the similarity between the two genomes is ~64% (Lambert and Loria 1989) and the BLASTX match was less than 100%, indicating that our sequence was most likely *S. acidiscabies*.

The presence of a *S. acidiscabies* sequence leads us to consider several hypotheses involving contamination of our library during or after its generation as the source of our *E. coli* sequences. We know this contamination event is not occurring during the cloning process due to the appearance of fixed primer region, 230, found in each sequence. It is possible that our initial library was contaminated with *E. coli*. Another, possibility is that contamination is occurring during the protein purification step. As protein are over-expressed through addition of inducer to BL21λDE3 *E. coli*, it may be possible that our proteins are binding to stretches of the AT rich *E. coli* sequences, protecting these regions from DNAse1 activity and preventing binding by *S. acidiscabies* sequences in the SELEX assay.

Analysis of residues involved in GBL protein binding to ARE sites. Table 4 shows DNA binding domains of 14 GBL receptors and TetR. Based on comparative analyses of the residues involved in DNA binding from each protein, general patters are immediately obvious. The first residue involved in the interaction contains either a polar, uncharged side chain (N, S, T, Q) or one that is electrically charged (R, K). It is also apparent that in over half of the sequences the second, fifth, and sixth residues are G, F, and H respectively. While the exact interaction between these residues and the ARE site has not been determined, the crystal structure of TetR regulator bound to the operator region of an efflux pump is known (Le et al. 2011). Swiss PDB viewer analysis of this structure reveals the six DNA contacting residues interacting with a nucleotide sequence, CCTATC, in the 5' region of the operator complex. It seems likely that members of the TetR family of repressors that contain similar protein sequences would interact with DNA strands of similar sequences. The ends of an ARE site contain high proportions of adenine and thymine preceded by the high GC content of the *Streptomyces* genome which would closely mimic the operator sequence bound by TetR in its crystal structure. The exact protein-DNA interaction could be elucidated through site directed mutagenesis of the six residue region found in a GBL receptor protein.

Conclusion

The comparison between our SabS enriched synthetic logo and the logo created based on previously identified ARE sites indicates that our recovered sequences display a high degree of similarity to that of an ARE site. Based on these results we know that our SELEX assay has been successful in allowing us to isolate synthetic sequences bound by *sabS* and enrich a synthetic library for only these sequences over time. The SELEX assay, optimized for use in *S. acidiscbies*, allowed for the recovery of nine unique sequences from an initial input of 10^{16} unique sequences a feat showing incredible specificity.

The development of an optimized assay for *S. acidiscabies* will allow for the recovery of genomic ARE sites bound by SabR and SabS. In order to accomplish this task, a contaminant free genomic library must be completed. SELEX with a genomic library will allow us to identify genomic ARE sites present within *S. acidiscabies*, indicating stretches of DNA transcriptionally regulated by SabR and SabS. In theory this task should be easier for the protein ligands than identification of synthetic sequences. The genomic library will contain an initial amounts of ~8 million base-pairs rather than ~10^{16}.

The genetic location of these ARE sites and the determination of the flanking sequences will allow us to further understand and characterize the regulatory role of our GBL synthase and receptor proteins in WS5995B production and morphological differentiation. Further analysis of the α-helix residues involved in DNA contact with the

ARE site will provide more insight into the exact mechanism of action between a GBL receptor and an ARE site.

References

Alvarez, H.M. & Steinbuchel, A. (2002) **Triacylglycerols in prokaryotic microorganisms.** *Appl. Microbiol. Biotechnol.* 60: 367-376.

Arabolaza, A., Rodriguez, E., Altabe, S., Alvarez, H. & Gramajo, H. (2008) **Multiple pathways for triacylglycerol biosynthesis in *Streptomyces coelicolor*.** *Appl. Environ. Microbiol.* 74: 2573- 2582.

Bangera, M. G. & Thomashow, L. S. (1999) **Identification and characterization of a gene cluster for synthesis of the polyketide antibiotic 2,4-diacetylphloroglucinol from *Pseudomonas fluorescens* Q2-87.** *J. Bacteriol.* 181: 3155-3163.

Bedford, D. J., Schweizer E., Hopwood, D. A. & Khosla C. (1995) **Expression of a function fungal polyketide synthase in the bacterium *Streptomyces coelicolor A3(2)*.** *J. Bacteriol.* 177: 4544-4548.

Bentley, R. (2005) **The Development of Penicillin: Genesis of a Famous Antibiotic.** *Perspectives in Biology and Medicine.* 48: 444-452.

Bentley, S. D, Chater, K. F., Cerdeno-Tarranga, A. M., Challis, G. L., Thomson N. R. & James, K. D. (2002) **Complete genome sequence of the model actinomycete *Streptomyces coelicolor* A3(2).** *Nature.* 417: 141-147.

Bentley, R. & Bennett, J. W. (1999) **Constructing Polyketides: From Collie to Combinatorial Biosynthesis.** *Annu. Rev. Microbiol.* 53: 411-446.

Berthelot, M. & Andre, G. (1891) **The Odor of the Land.** *Compt. Rend.* 112: 598–599.

Bibb, M. J., Biro, S., Motamedi, H., Colins, J. F. & Hutchinsin, C. R. (1989) **Analysis of the nucleotide sequence of the *Streptomyces glaucescens* tcmI genes provides key information about the enzymology of polyketide tetracenomycin C antibiotic biosynthesis.** *EMBO J.* 8: 2727-2736.

Bonde, M. R. & McIntyre, G. A. (1968) **Isolation and biology of a *Streptomyces* sp. causing potato scab in soils below pH 5.0.** *Am. Potato J.* 45:223-233.

Branden, C., and J. Tooze. 1991. **Introduction to protein structure.** Garland Publishing Co., New York, N.Y.

Brennan, R. G., Roderick, S. L., Takeda, Y. & Matthews, B. W. (1990) **Protein-DNA conformational changes in the crystal structure of a lambda Cro-operator complex.** *Proc. Natl. Acad. Sci.* 87: 8165–8169.

Bu'Lock, J. D. (1979) **Polyketide biosynthesis. in Comprehensive Organic Chemistry**, ed. Barton, D., Ollis, W. D. pp 927-87. Oxford, UK: Pergamon

Bu'Lock, J. D. & Smalley H. M. (1961) **Biosynthesis of aromatic substances from acetyl and malonyl-CoA.** *Proc. Chem. Soc.* 30:209-11.

Calcutt, M. J. & Schmidt, F. J. (1992) **Conserved gene arrangement in the origin region of the *Streptomyces coelicolor* chromosome.** *J. Bacteriol.* 174: 3220-3226.

Cho, Y., Kim, E., Chung, H., Choi, J., Chater, K. F., Ahn, B., Shin, J. & Roe, J. (2003) **The *parAB* operon is responsible for paraquat resistance in *Streptomyces coelicolor*.** *J. Bacteriol.* 185: 6756-6763.

Chong, P. P., Podmore, S., Kieser, H., Redenbach, M., Turday, K., Marahiel, M., Hopwood, D. A. & Smith, C. (1998) **Physical Identification of a chromosomal locus encoding biosynthetic genes for the lipopeptide calcium-dependent antibiotic (CDA) of *Streptomyces coelicolor* A3(2).** *Microbiology.* 144: 193-199.

Christen, S., Srinivas, A., Bahler, P., Zeller, A., Pridmore, D., Bieniossek, C., Baumann, U. & Erni, B. (2006) **Regulation of the Dha operon of Lactococcus lactis: a deviation from the rule followed by the TetR family of transcriptional regulators.** *J. Biol. Chem.* 281: 23129-23137.

Collie, J. N. (1983) **The formation of orcinol and other condensation products from dehydracetic acid.** *J. Chem. Soc.* 70: 122-126.

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

Curtis, T. P., Sloan, W. T. & Scannell, J. W. (2002) **Estimating prokaryotic diversity and its limits.** *Proc. Natl. Acad. Sci.* 99: 10494-10499.

Davis, N. K. & Chater, K. F. (1990) **Spore color in *Streptomyces coelicolor* A3(2) involves the developmentally regulated synthesis of a compound biosynthetically related to polyketide antibiotics.** *Mol Microbiol.* 4:1679-1691.

Dodd, I. B. & Egan, J. B. (1990) **Improved detection of helix-turn-helix DNA-binding motifs in protein sequences.** *Nucleic Acids Res.* 18: 5019-5026.

Donadio, S., Staver, M. J., McAlpine, J. B., Swanson, S. J. & Katz, L. (1991) **Modular organization of genes required for complex polyketide biosynthesis.** *Science.* 759: 675-679.

Ellard, J. & Lundgrend, G. (1966) **Fine structure of sporulation in *Bacillus cereus* grown in a chemically defined medium.** *J. Bacteriol.* 92: 1748-1764.

Ellington, A. D. & Szostak, J. W. (1990) **In vitro selection of RNA molecules that bind specific ligands.** *Nature.* 346: 818-822.

Engohang-Ndong, J., Baillat, D., Aumercier, M., Bellefontaine, F., Besra, G. S., Locht, C. & Baulard, A. R. **EthR, a repressor of the TetR/CamR family implicated in ethionamide resistance in mycobacteria, octamerizes cooperatively on its operator.** *Molec. Microbiol.* 51: 175-188.

Fernandez-Moreno, M. A., Martinez, E., Boto, L., Hopwood, D. A. & Malpartida, F. (1992) **Nucleotide sequence and deduced functions of a set of co-transcribed genes of *Streptmyces coelicolor* A3(2) including the polyketide synthase for the antibiotic actinorhodin.** *J. Biol. Chem.* 267: 19278-19290.

Fleming , B. A. (1929) **On the antibacterial action of cultures of a Penicillium, with special reference to their use in the isolation.** *J. Exper. Path.* 79: 780-790.

Fox, H. H. (1952) **The Chemical Approach to the Control of Tuberculosis.** *Science.* 8: 129-134.

Funa, N., Ohnishi, Y., Shibuya, M. & Ebizuka H. (1999) **A new pathway for polyketide synthesis in microorganisms.** *Nature.* 400: 897-899.

Gans, J., Wolinsky, M. & Dunbar, J. (2005) **Computational improvements reveal great bacterial diversity and high metal toxicity in soil.** *Science.* 309, 1387-1390.

Gerber, N. N., Lechevalier, H. A., (1965) **Geosmin, an earthy-smelling substance isolated from *Actinomycetes*.** *Appl. Microbiol.* 13, 935−938.

Glauerta,. M. & Hopwood, D. A. (1960). **The fine structure of *Streptomyces coelicolor.* I. The cytoplasmic membrane system.** *J. Biophys. and Biochem. Cytology.* 7: 479-490.

Goffeau, A., Barrell, B. G., Bussey, H., Davis, R. W., Dujon, B., Feldmann, H., Galibert, F., Hoheisel, J. D., Jacq, C., Johnston, M., Louis, E. J., Mewes, H. W., Murakami, Y., Philippsen, P., Tettelin, H. & Oliver, S. G. (1996) **Life with 6000 genes.** *Science* 27: 563-567.

Gravius, B., Glocker, D., Pandza, K., Hranueli, D. & Cullum, J. (1994) **The 387 kb linear plasmid pPZG101 of Streptomyces rirnosus and its interactions with the chromosome.** *Microbiology.* 140: 2271-2277.

Hare, R. ***The birth of penicillin and the disarming of microbes.*** London: Allen and Unwin, 1970. Print.

Harrison, S. C. (1991) **A structural taxonomy of DNA-binding domains**. *Nature* 353: 715–719.

Healy, F. G. & Lambert D. H. (1991) **Relationship among *Streptomyces* spp. Causing Potato Scab.** *Int. J. Syst. Bacteriol.* 41: 479-482.

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

Healy, F. G., Eaton, K. P., Limsirichai, P., Aldrich, J. F., Plowman, A. K., & King, R. R. (2009) **Characterization of gamma-butyrolactone autoregulatory signaling gene homologs in the angucyclinone polyketide WS5995B producer *Streptomyces acidiscabies.*** *J. Bacteriol*. 19: 4786-4797.

Henikoff, S., Haughn, G. W., Calvo, J. M. & Wallace J. C. (1988) **A large family of bacterial activator proteins.** *Proc. Natl. Acad. Sci.* 85: 6602– 6606.

Hertwick, C., Luzhetskyy, A., Rebets, Y. & Bechthold, A. (2006) **Type II polyketide synthases: gaining a deeper insight into enzymatic teamwork.** *Nat. Prod. Rep.* 24: 162-190.

Hinrichs, W., Kisker, C., Duvel, M., Muller, A., Tovar, K., Hillen, W. & Saenger, W. (1994) **Structure of the TetR repressor-tetracycline complex and regulation of antibiotic resistance.** *Science*. 264: 418-420.

Hinshaw, H. C., Feldman, W. H. & Pfuetze, K. H. (1947) **Streptomycin in treatment of clinical tuberculosis.** *Miss Valley Med J.* 69:160-6.

Hitomi, M., Nishimura, H., Tsujimoto, Y., Matsui, H. & Watanabe, K. (2003) **Identification of a helix-turn-helix motif of Bacillus thermoglucosida- sius HrcA essential for binding to the CIRCE element and thermostability of the HrcA-CIRCE complex, indicating a role as a thermosensor.** *J. Bacteriol.* 185: 381–385.

Hoenigejr, F. M., Stuartp, F. & Holt, S. C. (1968) **Cytology of spore formation in *Clostridium perfringens.*** *J. Bacteriol.* 96: I8I8.

Hopwood, D. A., Chater, K. F. & Bibb, M. J. (1995) **Genetics and Biochemistry of Antibiotic Production.** *Bacterol. Rev.* 41:65-102.

Hopwood, D. A., Sherman, D. H. (1990) **Molecular genetics of polyketides and its comparison to fatty acid biosynthesis.** *Annu. Rev. Genet.* 24: 37-66.

Horinouchi, S. (2007) **Mining and polishing of the treasure trove in the bacterial genus *Streptomyces*.** *Biosci. Biotechnol. Biochem.* 71: 283-299.

Hughes, C., McCrum, R. C. & Manzer, F. E. (1971) **Protein and esterase electrophoresis patterns of *Streptomyces* spp. isolated from Maine potatoes.** *Am. Potato J.* 48:206-213.

Hutchinson, C. R. (1997) **Antibiotics from genetically engineered microorganisms**, p. 659-682. In W. R. Strohl (ed.), Biotechnology of industrial antibiotics. Marcel Dekker, Baltimore, Md

Ikeda, H., Ishikawa, J., Hanamoto, A., Shinose, M., Kikuchi, H., Shiba, T., Sakaki, Y., Hattori, M. & Omura, S. (2003) **Complete genome sequence and comparative analysis of the industrial microorganism *Streptomyces avermitilis*.** *Nat. Biotechnol.* 21: 526-531.

Iverson, W. P. & Waksman, S. (1952) **Natural and Acquired Resistance of *Escherichia coli* to Neomycin.** *PNAS*. 38: 843-849.

Jez, J. M., Ferrer, J. L., Bowman, M. E., Dixon, R. A. & Noel, J. P. (2000) **Dissection of malonyl-coenzyme A decarboxylation from polyketide formation in the reaction mechanism of a plant polyketide synthase.** *Biochemistry* 39: 890-902.

Kallio, P. (2008) **Type II aromatic polyketide biosynthetic tailoring enzymes diversity and adaptation in *Streptomyces* secondary metabolism.** Diss: University of Turkey. Online.

Karniol, B. & Vierstra, R. D. (2004) T**he HWE histidine kinases, a new family of bacterial two-component sensor kinases with potentially diverse roles in environmental signaling.** *J. Bacteriol.* 186: 445–453.

Karoonuthaisiri, N., Weaver, D., Huang, J., Cohen, S.N. & Kao, C.M. (2005) **Regional organization of gene expression in *Streptomyces coelicolor*.** *Gene* 353: 53-66.

Kato, J. Y., Funa, N., Watanabe, Y., Ohnishi, Y. & Horinouchi, S. (2007) **Biosynthesis of gamma-butyrolactone autoregulators that switch on secondary metabolism and morphological development in *Streptomyces*.** *Proc. Natl. Acad. Sci.*104: 2378-2383.

Katz, L. & Donadio, S. (1993) **Polyketide Synthesis: Prospects for hybrid antibiotics.** *Annu. Rev. Microbiol.* 47: 875-912.

Keiser, H.M., Keiser, T. & Hopwood, D.A., (1992) **A combined genetic and physical map of the *Streptomyces coelicolor* A3(2) chromosome.** *J. Bacteriol.* 174: 5496-5507.

Khosla, C., Gokhale, R. S., Jacobsen, J. R. & Cane, D. E. (1999) **Tolerance and Specificity of Polyketide Synthases.** *Annu. Rev. Biochem.* 68: 219-253.

Khokhlov, A. S., Tovarova, L. L., Borisova, L. N., Pliner, S. A., Shevchenka, L. A., Kornitskaya, E. Y. A., Ivkina, N. S. & Rapoport, I. A. (1967) **The A-factor, responsible for streptomycin biosynthesis by mutant strains of Actinomyces streptomycin.** *Doki Akad Nauk SSSR*. 177: 232-235.

Khosla, C., McDaniel, R., Ebert-Khosla, S., Torres, R., Sherman, D. H., Bibb, M. J. & Hopwood, D. A. (1993) **Genetic construction and functional analysis of hybrid polyketide synthases containing heterologous acyl carrier proteins.** *J. Bacteriol.* 175: 2197-2204.

Kim, E. S., Cramer, K. D., Shreve, A. L. & Sherman, D. H. (1995) **Heterologous expression of an engineered biosynthetic pathway: functional dissection of type II polyketide synthase components in *Streptomyces* species.** *J. Bacteriol.* 177: 1202-1207.

King, R. R. & C. H. Lawrence. (1996) **Isolation and identification of pigments generated in vitro by *Streptomyces acidiscabies.*** *J. Agric. Food Chem.* 44: 2849–2851.

King, R. R., Lawrence, C. H., Clark, M.C. & Calhoun, L.A. (1989) **Isolation and characterization of phytotoxins associated with *Streptomyces scabies*.** *J. Chem. Soc. Commun.* 13: 849-850.

Kojic, M. & Venturi, V. (2001) **Regulation of *rpoS* Gene Expression in *Pseudomonas*: Involvement of a TetR Family Regulator**. *J. Bacteriol.* 183: 3712-3720.

Kunst, F., Ogasawara, N., Moszer, I., Albertini, A. M., Alloni, G., Azevedo, V., Bertero, M. G., Bessieres, P., Bolotin, A., Borchert, S., Borriss, R., Boursier, L., Brans, A., Braun, M., Brignell, S. C., Bron, S., Brouillet, S., Bruschi, C. V., Caldwell, B., Capuano, V., Carter, N. M., Choi, S. K., Codani, J. J., Connerton, I. F. & Danchin, A. (1997) **The complete genome sequence of the gram-positive bacterium *Bacillus subtilis*.** *Nature* 390: 249-256.

Lambert, D. H. & Loria, R. (1989) ***Streptomyces acidiscabies* sp. nov.** *Inter. J. of Sys. Bacteriol.* 39: 393-396.

Lawrence, C. H., Clark, M. C. & King, R. R. (1990) **Induction of common scab symptoms in aseptically cultured potato tubers by the vivotoxin, thaxtomin**. *Phytopathology* 80: 606-608.

Le, T. B., Stevenson, C. E., Buttner, M. J. & Lawson, D. M. (2011) **Crystallization and preliminary X-ray analysis of the TetR-like efflux pump regulator SimR.** *Acta. Cryst.* 67: 307-309.

Leblond, P., Fischer, G., Francou, F., Berger, F., Guerineau, M., & Decaris, B. (1996) **The unstable region of *Streptomyces ambofaciens* includes 210 kb terminal inverted repeats flanking the extremities of the linear chromosomal DNA.** *Mol. Microbiol.* 19: 261-271.

Leblond, P., Francou, F.X., Simonet, J.M. & Decaris, B. (1990) **Pulsed-field gel electrophoresis analysis of the genome of *Streptomyces ambofaciens* strains.** *FEMS Microbiol. Lett.* 60, 79-88.

Lee, L., Chen, Y., Kirby, R., Chen, C. & Chen, C. (2007) **A multidrug efflux system is involved in colony growth in *Streptomyces lividans.*** *Microbiol.* 153: 924-934.

Lezhava, A. L., Mizukami, T., Kajitani, T., Kameoka, D., Redenback, M. & Shinkawa, H. (1995) **Physical map of the linear chromosome of *Streptomyces griseus.*** *J. Bacteriol.* 177: 6492-6498.

Lin, Y. S., Kieser, H. M., Hopwood, D. A. & Chen, C. W. (1993) **The chromosomal DNA of *Streptomyces lividans* 66 in linear.** *Mol. Microbiol.* 10: 923-933.

Loria, R. & Kempter, B. A. (1986) **Relative resistance of potato tubers produced from stem cuttings and seed-piece-propogated plants to *Streptomyces scabies*.** *Plant Dis.* 70:1146-1148.

Loria, R., Buckhalid, R. A., Fry, B. A. & King, R. R. (1997) **Plant pathogenicity in the genus *Streptomyces*.** *Plant Dis.* 81: 836-846.

MacEachran, D. P., Stanton, B. A. & O'Toole, G. A. (2008) **Cif is negatively regulated by the TetR family repressor CifR.** *Infect. & Immun.* 76: 3197-3206.

MacNeil, D. J., Occi, J. L., Gewain, K. M. & MacNeil, T. (1994) **Correlation of the avermectin polyketide synthase genes to the avermectine structure: implications for designing novel avermectins.** *Ann. N. Y. Acad. Sci.* 721: 123-132.

Manzer, F. E., McIntyre, G. A. & Merriam, D. C. (1977) **A new potato scab problem in Maine.** *Marine Life Sci. Agric. Exp. Stn. Bull.* 85:1-24.

Mayorga, M. E. & Timberlake, W. E. (1992) **The developmentally revulared *Asperigillus nidulans* WA gene encodes a polypeptide homologous to polyketide and fatty acid synthases.** *Mol. Gen. Genet.* 235: 205-212.

McDaniel, R., Ebert-Khosla, S., Hopwood, D. A. & Khosla, C. (1993) **Engineered biosynthesis of novel polyketides.** *Science* 262: 1546-1550.

Miyajima, K., Tanaka, F. & Kuninaga, S. (1998) ***Streptomyces turgidiscabies* sp. nov.** *Int. J. Syst. Bacteriol.* 48: 495-502.

Mohibullah, N., Donner, A., Ippolito, J. A. & Williams, T. (1999) **SELEX and missing phosphate contact analysis reveal flexibility within the AP-2α protien:DNA binding complex.** *NAR.* 27: 2760-2769.

Moore, B. S. & Hopke, J. N. (2001) **Discovery of a New Bacterial Polyketide Biosynthetic Pathway.** *J. Bio Chem*. 2: 35-38.

Moore, B. S. & Hertweck, C. (2002) **Biosynthesis and Attachment of Novel Bacterial Polyketide Synthase Starter Units.** *Nat. Prod. Rep*. 19: 70-99.

O'Hagan, D. & Cullum, J. (2001) **New Polyketide Biosynthesis.** *Chem. Annu. Rev.* 50: 381-412.

O'Hagan, D. (1991) **The polyketide metabolites.** Horwood, Chichester, United Kingdom

Ohnishi, Y., Ishikawa, J., Hara, H., Suzuki, H., Ikenoya, M., Ikeda, H., Yamashita, A., Hattori, M. & Horinouchi, S. (2008) **The genome sequence of the streptomycin-producing microorganism *Streptomyces griseus* IFO 13350.** *J. Bacteriol.* 190, 4050-4060.

Ohnishi, Y., & Horinouchi, S. (2004) **The A-factor regulatory cascade that leads to morphological development and secondary metabolism in *Streptomyces*.** *Biofilms.* 1: 319-328.

Ohnishi, Y., Yamazaki, H., Kato, J., Tomono, A., & Horinouchi, S. (2005) ***AdpA*, a central transcription regulator in the A-factor cascade that leads to morphological development and secondary metabolism in *Streptomyces griseus*.** *Biosci. Biotechnol. Biochem.* 69: 431-439.

Olukoshi, E. R. & Packter, N. M. (1994) **Importance of stored triacylglycerols in *Streptomyces*: Possible carbon source for antibiotics.** *Microbiology* 140, 931- 943.

Pao, G. M.,Tam, R., Lipschitz, L. J. & Saier, M. H. (1994) **Response regulators: structure, function and evolution.** *Res. Microbiol.* 145: 356–362.

Perez-Rueda, E. & Collado-Vides, J. (2001) **Common history at the origin of the position–function correlation in transcriptional regulators in archaea and bacteria.** *J. Mol. Evol.* 53: 172–179.

Rajgarhia, V. B. & Strohl, W. R. (1997) **Minimal *Streptomyces* sp. strain C5 daunorubicin polyketide biosynthesis genes required for aklanonic acid biosynthesis.** *J. Bacteriol.* 179: 2690-2696.

Ramos, J. L., Martinez-Bueno, M, Monlina-Henares, A. J., Terna, W., Watanabe, K., Zhang, X., Gallegos, M. T., Brennan, R. & Tobes, R. (2005) **The TetR Family of Transcriptional Repressors.** *Microbiol. and Molec. Biol. Rev.* 69: 326-356.

Rawlings, B. J. (1999) **Type I polyketide biosynthesis in bacteria (part A — erythromycin biosynthesis)** *Nat. Prod. Rep.* 16: 425-484.

Reeves, C. D., (2003) **The enzymology of combinatorial biosynthesis.** *Crit. Rev. Biotechnol.* 23: 95-147.

Remsing L. L., Garcia-Bernardo, J., Gonzalez, A., Kunzel, E., Rix, U., Brana, A. F., Bearden, D. W., Mendez, C., Salas, J. A. & Rohr, J. (2001) **Ketopremthramycins and Ketomithramycins, Four New Aureolic Acid-type compounds obtained upon inactivation of two genes involved in the biosynthesis of the deoxysugar moieties of the antitumor drug mithramycin by *Streptomyces argillaceus*, Reveal novel insights into Post-PKS tailoring steps of the amthramycin biosynthetic pathway.** *JACS.* 124: 1606-1614

Rose, M., Balakumar, P. & Singh, M. (2007) **Ameliorative Effect of Combination of Fenofibrate and Rosiglitazone in Pressure Overload-Induced Cardiac Hypertrophy in Rats.** *Pharmacol.* 80: 177-184.

Rousseau, P., Guegen, E., Duval-Valentin, G. & Chandler, M. (2004) **The helix-turn-helix motif of bacterial insertion sequence IS911 transposase is required for DNA binding.** *Nucleic Acids Res.* 32: 1335–1344.

Rubin, R. P. (2007) **A Brief History of Great Discoveries in Pharmacology: In Celebration of the Centennial Anniversary of the Founding of the American Society of Pharmacology and Experimental Therapeutics.** *Pharmacol. Rev.* 59: 289-359.

Schroder, J.(1999) in Comprehensive Natural Products Chemistry, Vol 1 (eds.: D. Barton, K. Nakanishi, O. Meth-Cohn), Elsevier, Amsterdan, pp. 749-771.

Schumacher, M. A., Miller, M. C., Grkovic, S., Brown, M. H., Skurray, R. A. & Brennan, R. G. (2002) **Structural basis for cooperative DNA binding by two dimers of the multidrug-binding protein QacR**. *EMBO. J.* 21: 1210-1218

Schupp, T. Toupet, C., Cluzel, B., Neff, S., Hill, S., Beck, J. J. & Ligon, J. (1995) **A *Sorangium cellulosum* (Myxobacterium) gene cluster fro the biosynthesis of the macrolide antibiotic soraphen A: cloning, characterization, and homology to polyketide synthase genes from actinomycetes.** *J. Bacteriol.* 177: 3673-3679.

Schwecke, T., Aparicio, J. F., Molnar, I., Konig, A., Khaw, L. E., Haydock, S. F., Oliynyk, M., Caffrey, P., Cortes, J., Lester, J. B., Bohm, G. A., Staunton, J. & Leadlay, P. F. (1995) **The biosynthesis gene cluster for the polyketide immunosuppressant rapamycin.** *Proc. Natl. Acad. Sci.* 92: 7839-7843.

“Selman A. Waksman - Biography". Nobelprize.org. 2 Mar 2011 http://nobelprize.org/nobel_prizes/medicine/laureates/1952/waksman-bio.html

Seow, K., Meurer, G., Gerlitz, M., Wendt-Pienkowski, E., Hutchinson, C. R. & Davies, J. (1997) **A study of iterative type II polyketide synthases, using bacterial genes cloned from soil DNA: a means to access and use genes from uncultured microorganisms.** *J. Bacteriol.* 179: 7360-7368.

Shen, B. (2000) **The biosynthesis of aromatic polyketides.** *Top. Curr. Chem.* 209: 1-51.

Sherman, D. H., Malpartida F., Bibb, M. J., Kieser H. M. & Hopwood, D. A. (1989) **Structure and decided function of the granaticin-producing polyketide synthase gene cluster of *Streptomyces violaceoruber* TU22.** *J. EMBO.* 8: 2717-2725.

Stoltenburg, R., Reinemann, C. & Strehlitz, B. (2007) **SELEX- an evolutionary method to generate high affinity nucleic acid ligands.** *Biomolec. Engineer.* 24: 381-403.

Summers, R. G., Wendt-Pienkowski, E., Mtamedi, H. & Hutchinson, C. R. (1992) **Nucleotide sequence of the tcmNO region of the tetracenomycin C biosynthetic gene cluster of *Streptomyces glaucescens* and evidence that the *tcmN* gene encodes a multifunctional cyclase/dehydrogenase/O-methyltransferase.** *J. Bacteriol.* 174: 1810-1820.

Summers, R. G., Wendt-Pienkowski, E., Motamedi, H. & Hutchinson, C. R. (1993) **The tcmVI region of the tetracenomycin C biosynthesis gene cluster of *Streptomyces glaucenscens* encodes the tetracenomycin F1 monooxygenase, tetracenomycin F2 cyclase, and, most likely, a second cyclase.** *J. Bacteriol.* 175: 7571-7580.

Takeuchi, T., Sawada, H., Tanaka, F. & Matsuda, I. (1996) **Phylogenetic analysis of *Streptomyces* spp. causing potato scab based on 16S rRNA Sequences.** *Inter. J of Sys. Bacteriol.* 46: 476-479.

Takono, E. (2006) **Gamma-Butyrolactones: Streptomyces signaling molecules regulating antibiotic production and differentiation.** *Curr. Opin. Microbiol.* 41: 1015-1028.

Torsvik, V., Goksoyr, J. & Daae, F. L. (1990) **High diversity in DNA of soil bacteria.** *Appl. Environ. Microbiol.* 56: 782-787.

Tsai, R. & Reed, R. (1998) **Identification of DNA Recognition Sequences and Protein Interaction Domains of the Multiple-Zn-Finger Protein Roaz.** *Molec. and Cell. Biol.* 11: 6447-6456.

Tuerk, C. & Gold, L. (1990) **Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase.** *Science.* 249: 505–510.

Volff, J. N. & Altenbuchner, J. (2000) **A new beginning with new ends: Linearisation of circular chromosomes during bacterial evolution.** *FEMS Microbiol. Lett.* 186: 143-150.

Vloff, J. N., & Altenbuchner, J. (1998) **Genetic instability of the Streptomyces chromosome.** *Mol. Microbiol.* 27: 2389-246.

Wakil, S. J. (1989) **Fatty acid synthase, a proficient multifunctional enzyme.** *Biochemistry.* 28: 4523-4530.

Waksman, S. A. (1944) **Effect of streptomycin and other antibiotic substances upon Mycobacterium tuberculosis and related organisms**. *Proc. Soc. Exp. Biol. Med.* 57, 244-248.

Waksman, S. A. (1947) **Tuberculostatic and tuberculocidal properties of streptomycin**. *J. Bacteriol.* 54: 253-261.

Waksman, S. A. & Lechevalier, H. A. (1949) **Neomycin, a New Antibiotic Active against Streptomycin-Resistant Bacteria, including Tuberculosis Organisms** *Science.* 25: 305-307.

Wang, Z. Yang, P., Fan, B. & Chen, Z. (1998) **An oligo selection procedure for identification of sequence specific DNA-binding activities associated with the plants defense response.** *The Plant J.* 16: 515-522.

Wildermuth, H. & Hopwood D. (1970) **Septation during sporulation in *Streptomyces coelicolor.*** *J. Gen. Microbiol.* 60: 51-59.

Wosten, M., Wagenaar, J. A. & van Putten, J. P. (2004) **The FlgS/FlgR two component signal transduction system regulates the fla regulon in *Campylobacter jejuni*.** *J. Biol. Chem.* 279: 16214–16222.

Zakrzewska-Czerwinska, J. & Schrempf, H. (1992) **Characterization of an autonomously replicating region from the *Streptomyces lividans* chromosome.** *J Bacteriol.* 147: 2688-2693.

Figure Legends

Figure 1. Construction of a polyketide by a type II polyketide synthase. The type II PKS is made up of a 'minimal PKS' contain three protein components: two ketoacyl synthases and an acyl carrier protein. The minimal PKS is responsible for constructing the polyketide by catalyzing iterative Claisen condensation reactions between a starter unit and an extender unit. After construction the polyketide chain interacts with tailoring enzymes and is altered into its final structure.

Figure 2. Structure of WS5995B. WS5995B is a decaketide secondary metabolite product produced by a type II PKS and secreted by *S. acidiscabies*.

Figure 3. Phenotypic properties of *sabR*, *sabA* and *sabS* mutants. (A) Pigment production characteristics of *sabR*, *sabA* and *sabS* deletion mutants on SGM; wt denotes wild type strain *S. acidiscabies* 84.104. (B) Complementation of *sabS* deletion mutant on SGM with plasmid pIJ86::*sabS*. (C) Complementation of *sabA* deletion mutant on SGM with plasmid pIJ86::*sabA*. (D) Complementation of conditional morphological defect of *sabS* mutant on ISP2 medium with plasmid pIJ86::*sabS*.

Figure 4. Quantitative RT-PCR of *sabS* target gene expression in 84.104 and *ΔsabR*. *sabS* C_t values were normalized to *hrdB*, and the mean relative expression ratios were calculated by the $\Delta\Delta C_t$ Graph shows relative fold-difference in *sabS* expression in wild type 84.104 and *ΔsabR* strains. Significance was calculated by ANOVA and the error bars represent standard error. From Healy. et al, 2009.

Figure 5. Proposed model for the regulation of WS5995B biosynthesis and morphological differentiation by GBL components in *S. acidiscabies* 84.104.

Figure 6. Map of pCR 2.1-Topo Cloning Vector.

Figure 7. Growth of BL21λDE3/pet25b::SabS and Expression of His_6-SabS. Cultures of BL21λDE3/pET25b::SabS are grown at 30°C to OD_{600}, and *sabS* expression is induced with addition of IPTG at 0.6 nm. Plot shows growth of three independent cultures.

Figure 8. SDS-PAGE analysis of IPTG-dependent His_6-SabS expression. Total cell protein from cultures with (+) or without (-) added IPTG were electrophoresed through 15% polyacrylamide gel. Molecular weights of protein standards are shown at left. Predicted anhydrous molecular weight of *sabS* is 24 kDa

Figure 9. SDS-PAGE analysis of IPTG-dependent His_6-SabR expression. Total cell protein from cultures with (+) or without (-) added IPTG were electrophoresed through 15% polyacrylamide gel. Molecular weights of protein standards are shown at left. Anhydrous molecular weight of *sabR* is 24 kDa.

Figure 10. SDS-PAGE analysis of isolated SabS recombinant protein. Polyhistidine-tagged SabS is isolated from other cellular proteins after passing extracted protein through a HisGraviTrap column. Lane 1 contains a protein molecular weight marker while lane two contains an aliquot of the original IPTG-induced BL21λDE3/ pET25b::*sabS* culture followed by seven lanes each containing a fraction from the HisGraviTrap column.

Figure 11. SDS-PAGE analysis of isolated recombinant protein. Polyhistidine-tagged SabR is isolated from other cellular proteins after passing extracted protein samples through a HisGraviTrap column. Lane 1 contains a protein molecular weight marker while lane two contains an aliquot of the original IPTG-induced BL21λDE3/ pET25b::*sabR* culture followed by six lanes each containing a fraction from the HisGraviTrap column.

Figure 12. SDS-PAGE analysis of SabS recombinant protein after desalting. Fractions of isolated, purified recombinant SabS are collected after pooled isolated fractions are passed through a PD-10 desalting column. Lane 1 contains a protein molecular weight marker while lane two contains an aliquot of the original IPTG-induced BL21λDE3/ pET25b::*sabS* culture followed by five lanes each containing a fraction from the desalting column.

Figure 13. SDS-PAGE analysis of SabR recombinant protein after desalting. Fractions of isolated, purified recombinant SabR are collected after pooled isolated fractions are passed through a PD-10 desalting column. Lane 1 contains a protein molecular weight marker while lane two contains an aliquot of the original IPTG-induced BL21λDE3/ pET25b::*sabR* culture followed by five lanes each containing a fraction from the desalting column.

Figure 14. Systematic Evolution of Ligands by Exponential Enrichment (SELEX) assay. Purified His_6-SabS is bound to Ni-NTA agarose and unbound protein is removed by washing. Synthetic SELEX library is incubated with agarose-bound protein and unbound DNA sequences are removed by washing. Bound DNA sequences are eluted in PO_4 buffer containing 0.5 M imidazole PO_4 and amplified by PCR. The process is repeated to enrich for DNA sequences which specifically bind protein. After several rounds, sequences are cloned and sequenced.

Figure 15. High molecular weight products after amplification step of SELEX assay. Unwanted high molecular weight products were observed after amplification of synthetic DNA that bound to SabS in round six of SELEX. Expected product bands, representing the 224 primer template can be seen at ~120 bp. Lane 1 represents a kb ladder. Lanes 2-9 represent 5 µL aliquots of PCR products amplified from enriched synthetic DNA after 6 rounds of SELEX.

Figure 16. Determination of optimal cycle number for eluted SELEX products. SabS fractions F3 and F7 used in SELEX assay, were eluted and amplified for the 14, 18, 22, 26 or 30 cycles, then electrophoresed through a 2% agarose gel. Based on these results the number of cycles used in the amplification step of SELEX was 14. Lane 1 contains a kb ladder while lanes 2-6 contain 5 µL aliquots of F3 and lanes 7-11 contain 5 µL aliquots of F7.

Figure 17. Determination of annealing temperature for eluted SELEX products. SELEX products were amplified in a Tgradient thermal cycler with annealing temperatures ranging from 45ºC to 83º.

Figure 18. Construction of synthetic SELEX library. A) PCR amplification of SELEX library with a degenerate oligonucleotide (N_{28}) template.

Figure 19. SELEX enriched sequences bound by SabS using a synthetic library. Electrophoretic analysis of synthetic DNA library after seven rounds of SELEX. The sequences seen here represent the synthetic library prior to the cloning and sequencing step of SELEX.

Figure 20. PCR amplification of SELEX clones. After seven rounds of SELEX with SabS, products were into pCR 2.1 Topo vector and transformed into electrocompetent cells. 29 transformants (1-22 on the top gel, 23-29 on the bottom gel) were selected for analysis. Plasmid was extracted and amplified to assess presence of insert. Samples with inserts were sequenced. The last lane on the bottom gel represents a negative control (blue colony).

Figure 21. ARE-like sequences bound by SabS. From 20 inserts, nine unique sequences were identified as binding sites for SabS. The fixed primer portion of each sequence is underlined. Sequences 3 and 4 represent the reverse compliment of the original sequence.

Figure 22. ARE-like sequences bound by SabS. Sequence logo illustrating conservation of bases within aligned sequences.

Figure 23. Summary of known AREs. Nucleotide sequence alignment of predicted ARE sequence from *S. acidiscabies* (sabS-ARE) with previously characterized *Streptomyces* ARE boxes.

Figure 24. Summary of known AREs. Sequence logo illustrating conservation of bases within aligned ARE sequences.

Figure 25. Electrophoretic analysis of synthetic DNA library after five rounds of SELEX. The sequences seen here represent the SabR enriched synthetic library prior to the cloning and sequencing step of SELEX.

Figure 26. SELEX enriched sequences bound by SabR and SabS using a genomic library and SabR using a synthetic library. A) Enriched sequences for SabS with a genomic library after two rounds of SELEX. B) Enriched sequences for SabR with a genomic library (lanes 2-6) and a synthetic library (lanes 9-13).

Figure 27. Genomic digestion plasmids containing genomic inserts with Hind III. Genomic DNA bound by SabR and SabS after three rounds of SELEX was cloned into a Topo 2.1 plasmid. Plasmids were extracted and digested by Hind III in order to determine the relative size of genomic inserts. A) Hind III digestion of plasmids containing inserts bound by SabS. B) Hind III digestion of plasmids containing inserts bound by SabR.

Figure 28. Genomic sequences bound by SabR after three rounds of SELEX. Sequences bound by SabR. NCBI BlastX analysis indicates that each of these sequences is from *E. coli*.

Figure 29. Genomic sequences bound SabS after three rounds of SELEX. NCBI BlastX analysis indicates that sequence numbers 05, 06, and 08 are from *E. coli* while 07 is similar to sequences found in *S. scabies.*

Figure 30. Identification of GBL receptor protein residues involved in DNA contacts with their specified ARE site. CLUSTALW 2.1 multiple sequence alignment of 14 previously identified GBL receptors with known ARE binding sites. TetR is used to align sequences and identify conserved HTH motif in the DNA binding domain. Highlighted column represent α-helix residues involved in DNA contacts based on the crystal structure of TetR.

Figure 31. Analysis of sequence quality of sabR synthetic sequence. Electropherogram of R03 plasmid intended to contain a synthetic insert enriched by SabR over 5 rounds of SELEX. The multiple cloning site highlighted by the red box demonstrates several overlapping insert sequences.

Table Legends

Table 1. Bacterial strains and plasmids used in this study.

Table 2. Oligonucleotide primers and templates used in this study.

Table 3. Description of PCR amplification programs used in this study.

Table 4. Comparison of GBL receptor protein α-helix residues involved in DNA contact with the ARE site bound by each of protein

Figures:

Figure 1.

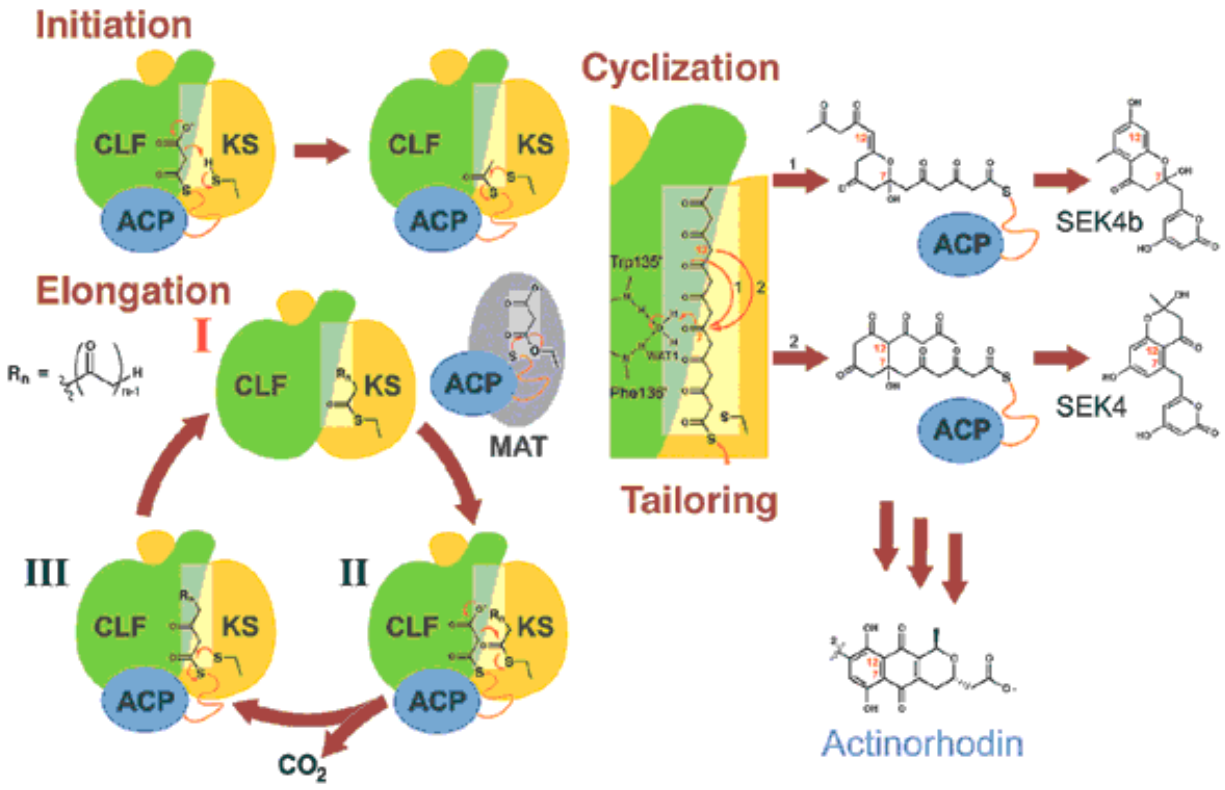


Figure 2.

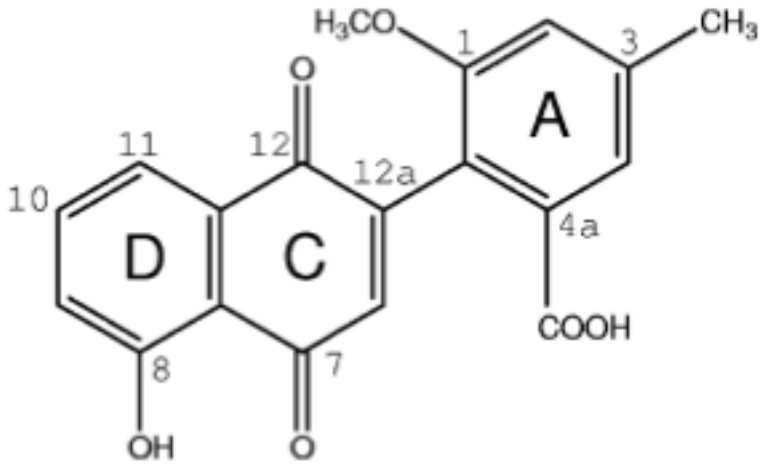


Figure 3.

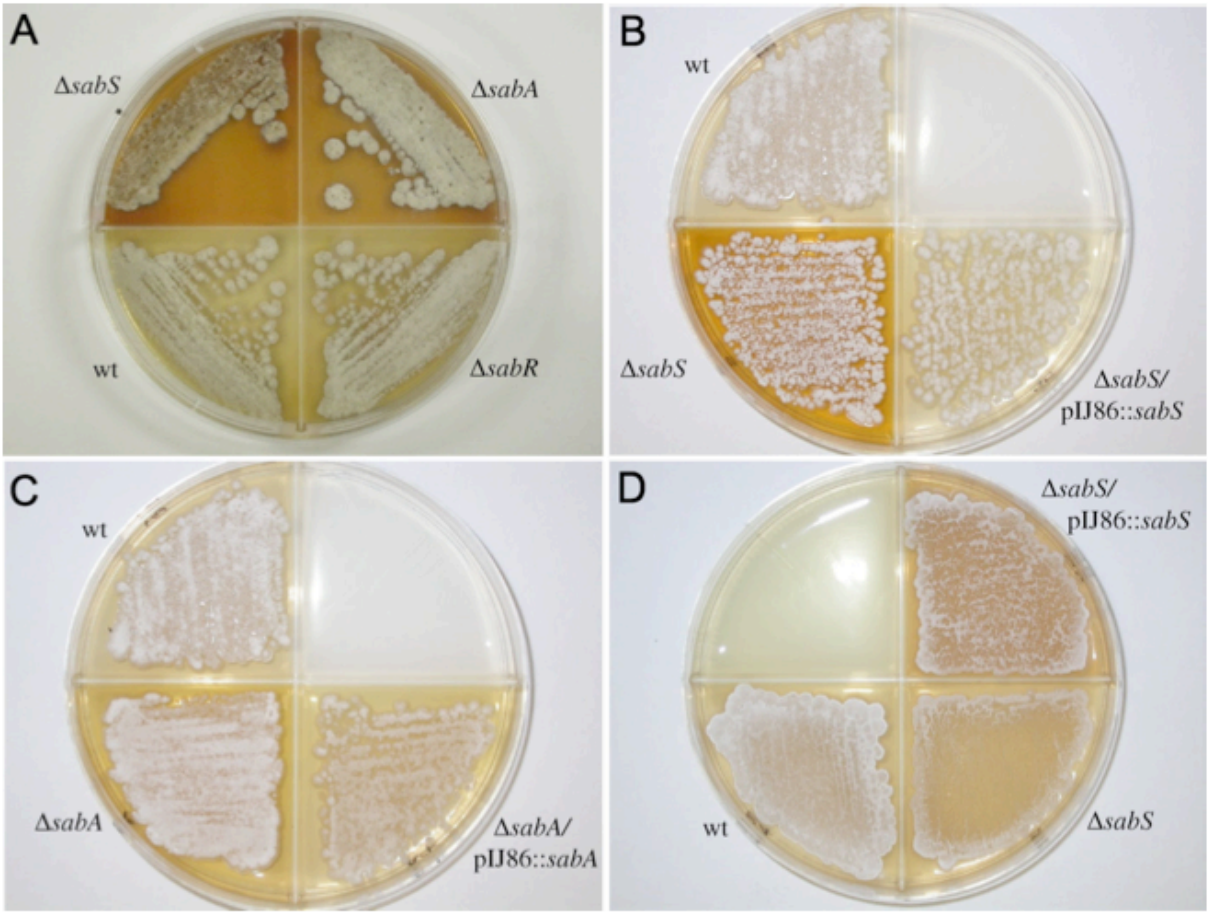

A
ΔsabS
ΔsabA
wt
ΔsabR
B
wt
ΔsabS
ΔsabS/
pIJ86::sabS
C
wt
ΔsabA
ΔsabA/
pIJ86::sabA
D
ΔsabS/
pIJ86::sabS
wt
ΔsabS

Figure 4.

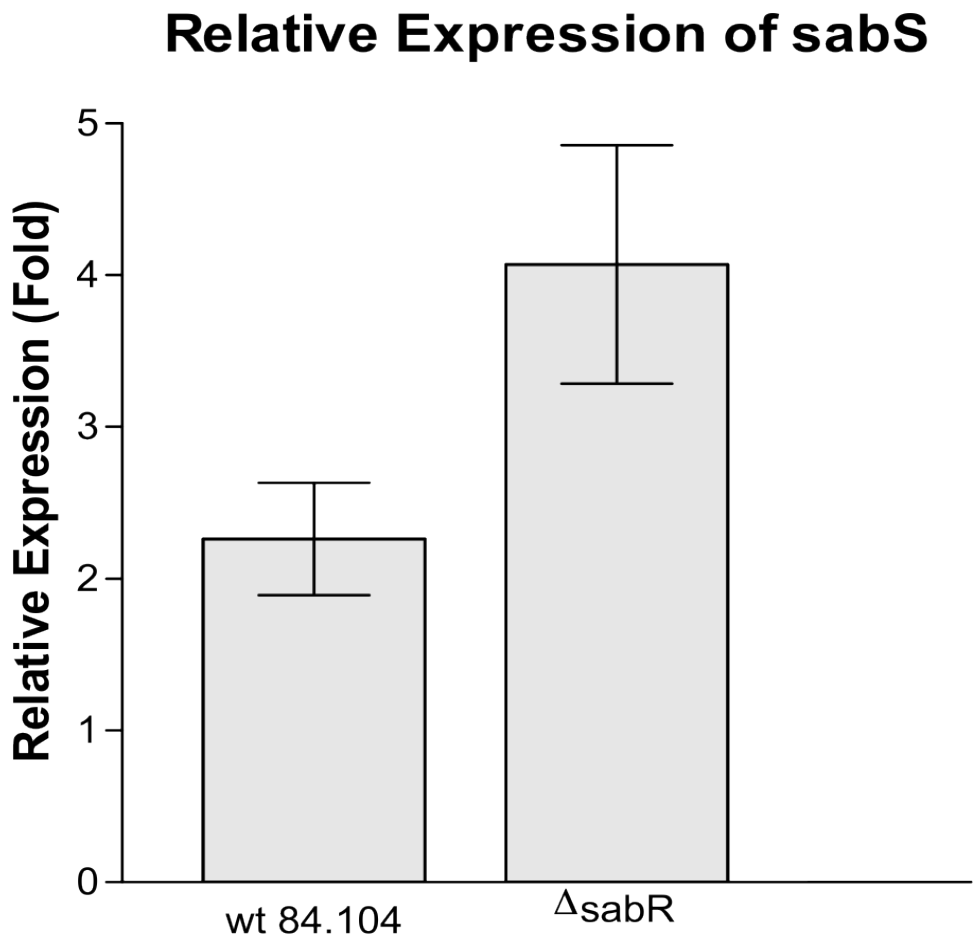

Relative Expression of sabS
Relative Expression (Fold)
5
4
3
2
1
0
wt 84.104
ΔsabR

Figure 5.

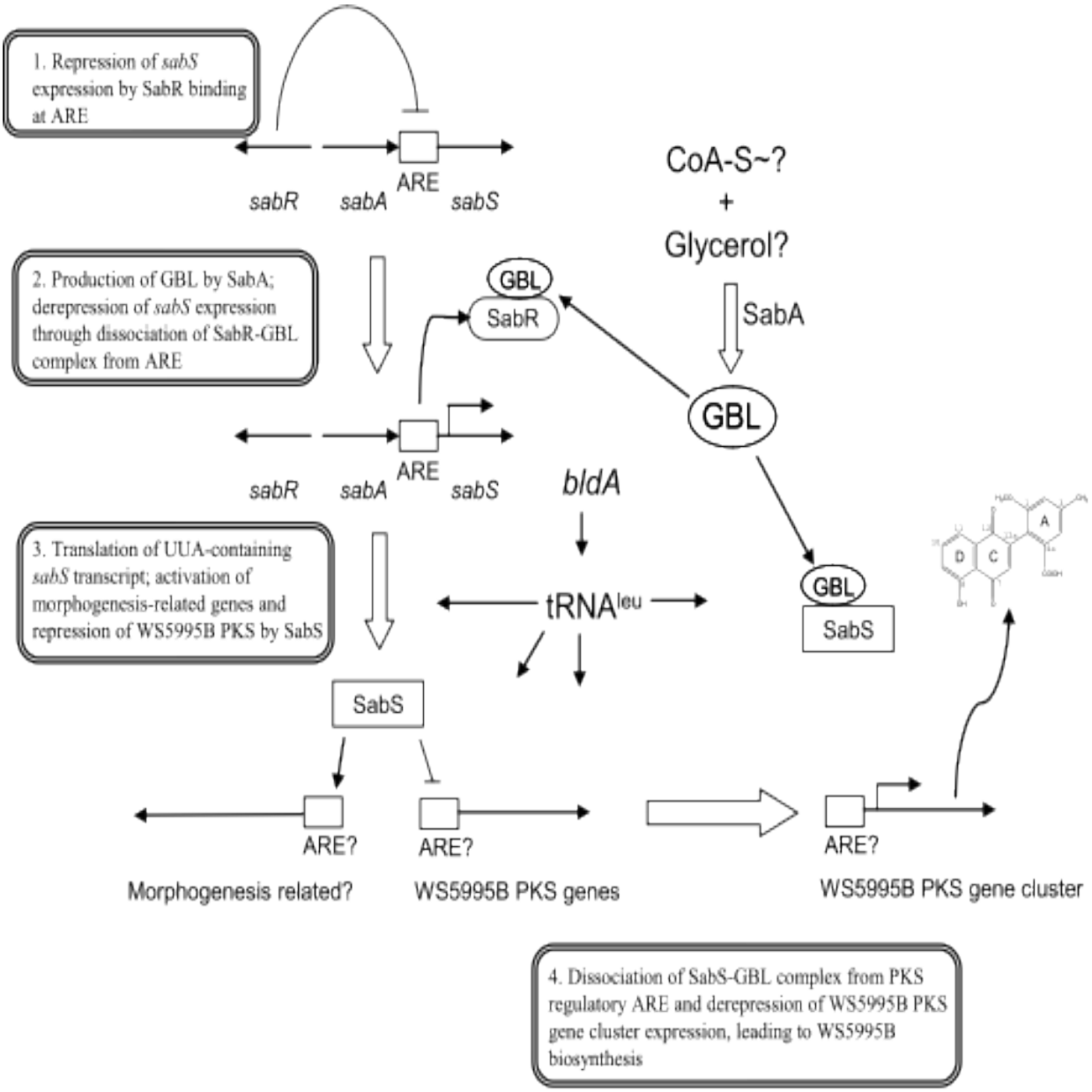
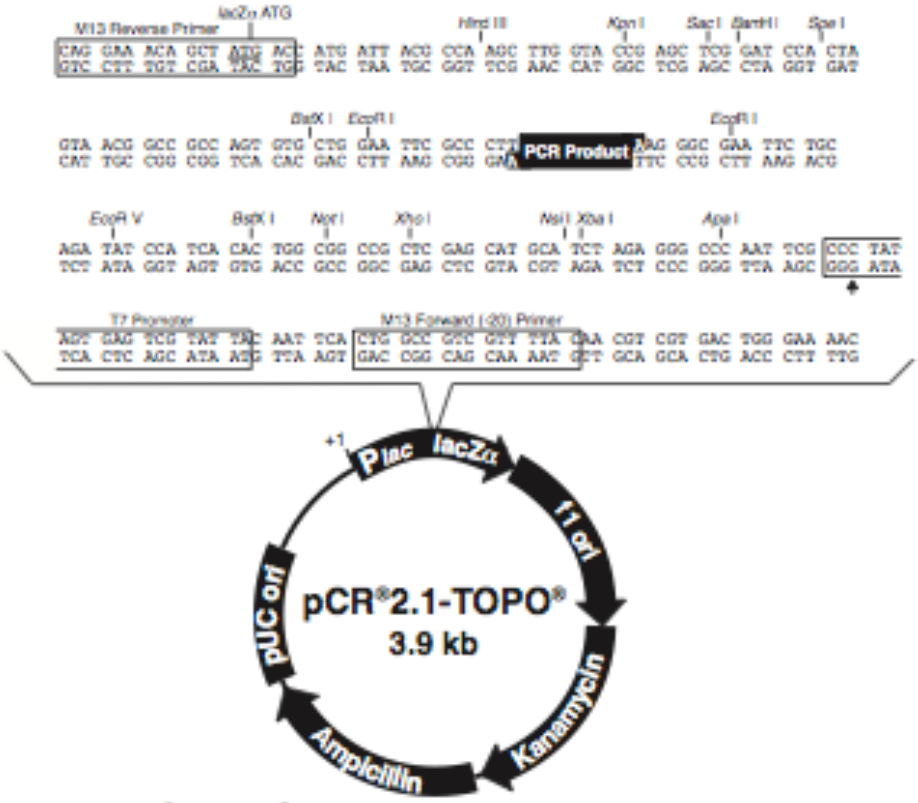

1. Repression of sabS expression by SabR binding at ARE
sabR
sabA
ARE
sabS
CoA-S~?
+
Glycerol?
SabA
2. Production of GBL by SabA; derepression of sabS expression through dissociation of SabR-GBL complex from ARE
GBL
SabR
GBL
sabR
sabA
ARE
sabS
bldA
3. Translation of UUA-containing sabS transcript; activation of morphogenesis-related genes and repression of WS5995B PKS by SabS
tRNALeu
GBL
SabS
SabS
ARE?
ARE?
ARE?
Morphogenesis related?
WS5995B PKS genes
WS5995B PKS gene cluster
4. Dissociation of SabS-GBL complex from PKS regulatory ARE and derepression of WS5995B PKS gene cluster expression, leading to WS5995B biosynthesis

Figure 6.



Comments for pCR®2.1-TOPO®
3931 nucleotides

LacZα fragment: bases 1-547
M13 reverse priming site: bases 205-221
Multiple cloning site: bases 234-357
T7 promoter/priming site: bases 364-383
M13 Forward (-20) priming site: bases 391-406
f1 origin: bases 548-985
Kanamycin resistance ORF: bases 1319-2113
Ampicillin resistance ORF: bases 2131-2991
pUC origin: bases 3136-3809

Figure 7.

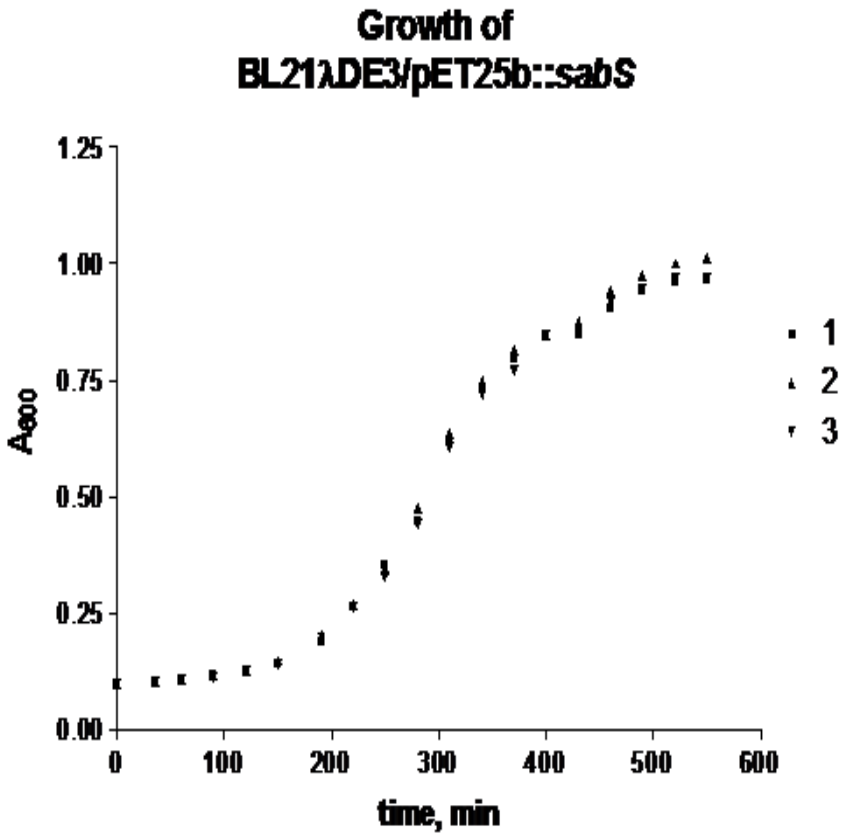

Growth of
BL21λDE3/pET25b::sabS
1.25
1.00
0.75
0.50
0.25
0.00
A600
0
100
200
300
400
500
600
time, min
1
2
3

Figure 8.

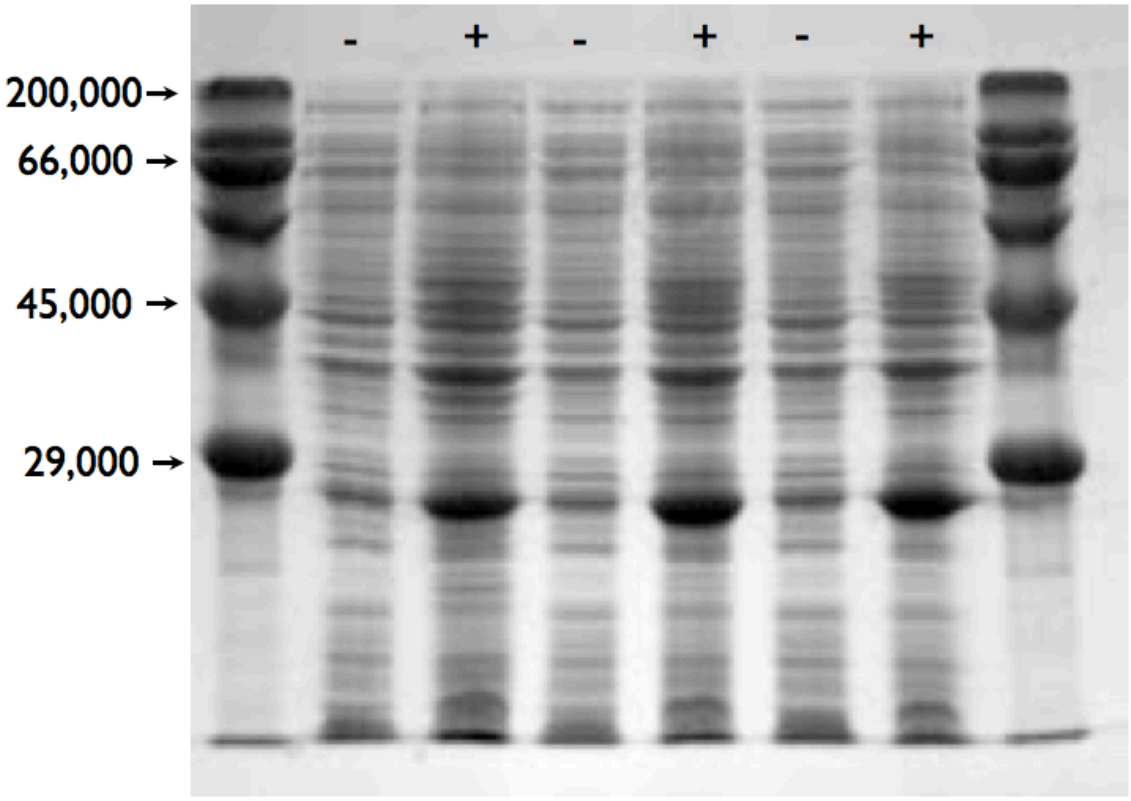


Figure 9.

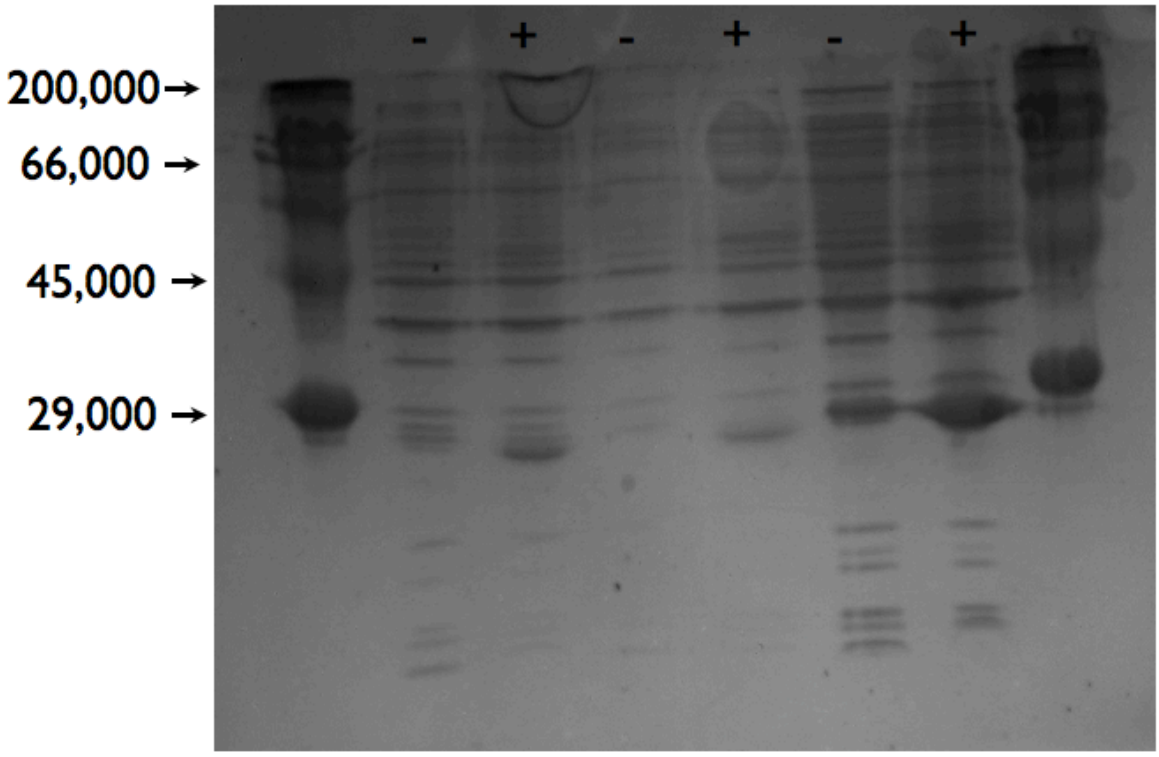


Figure 10.

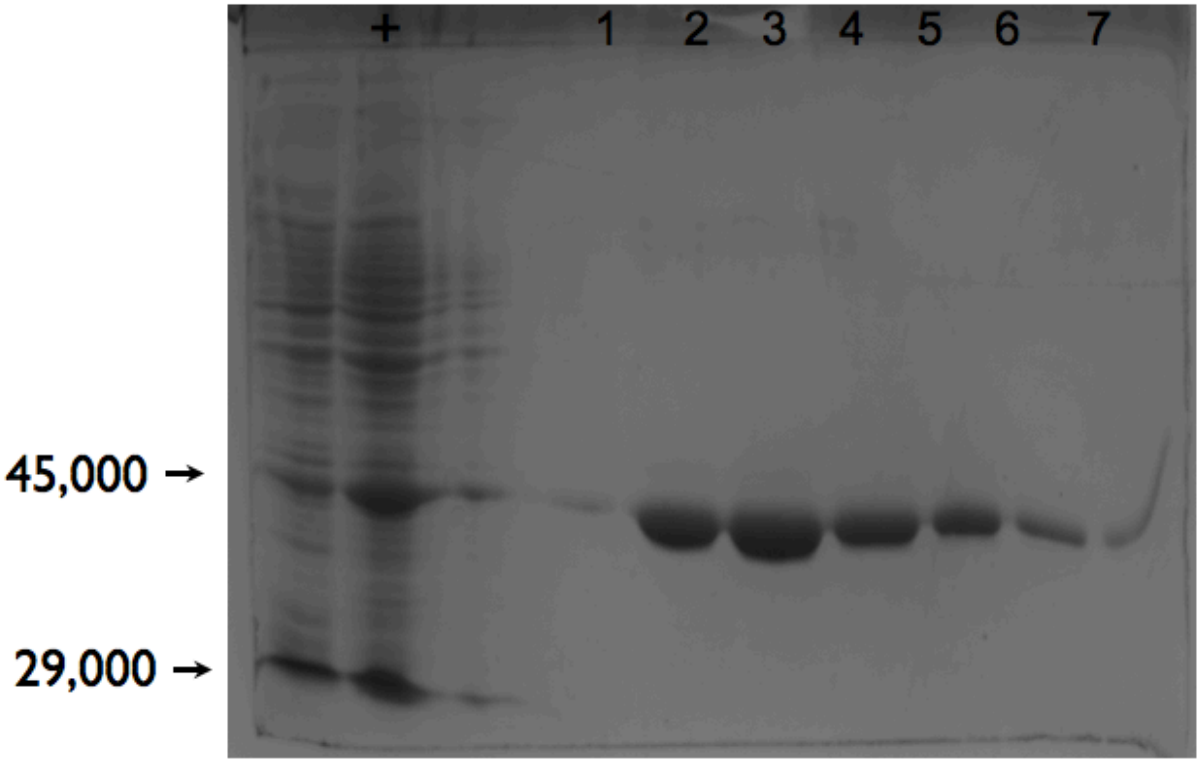


Figure 11.

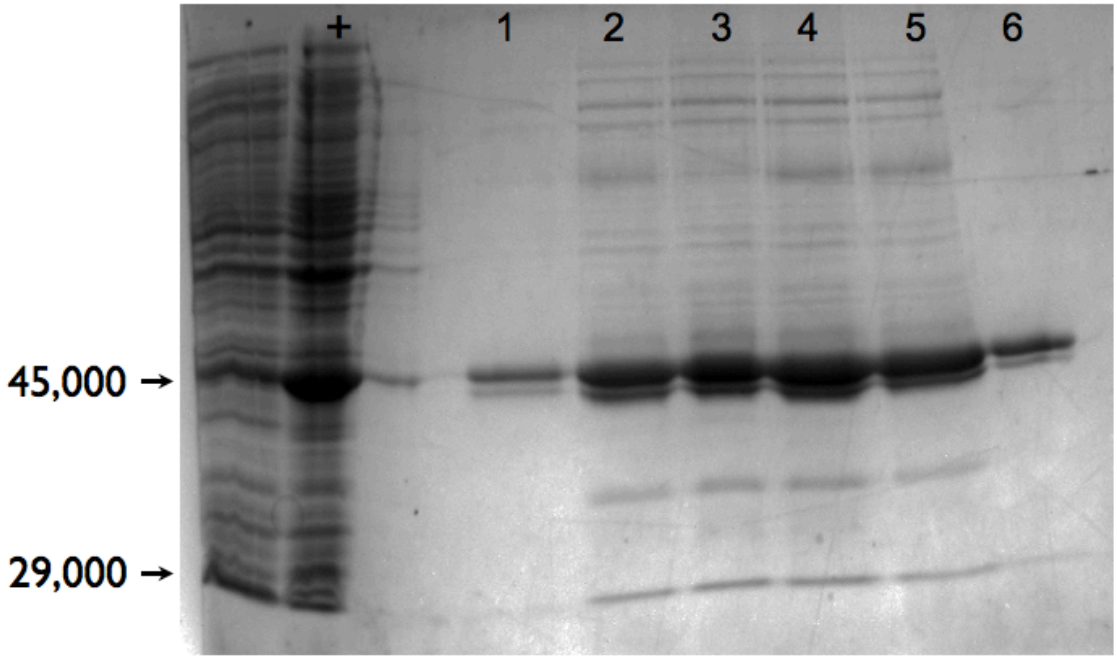


Figure 12.

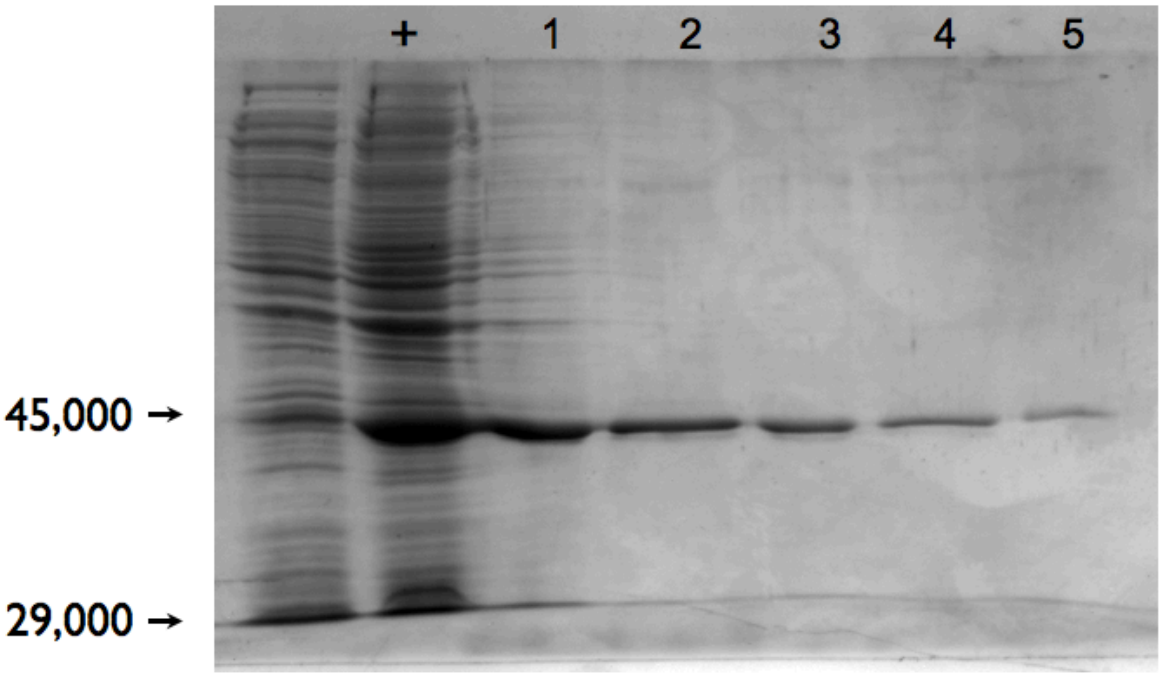


Figure 13.

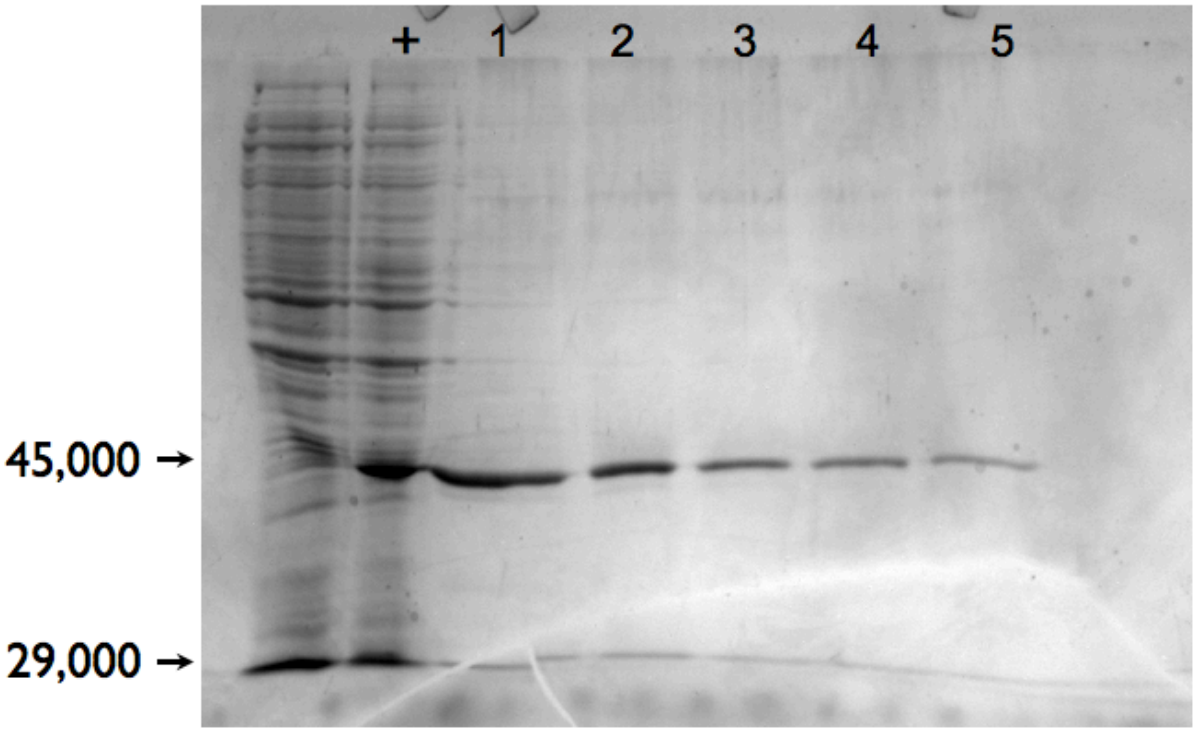


Figure 14.

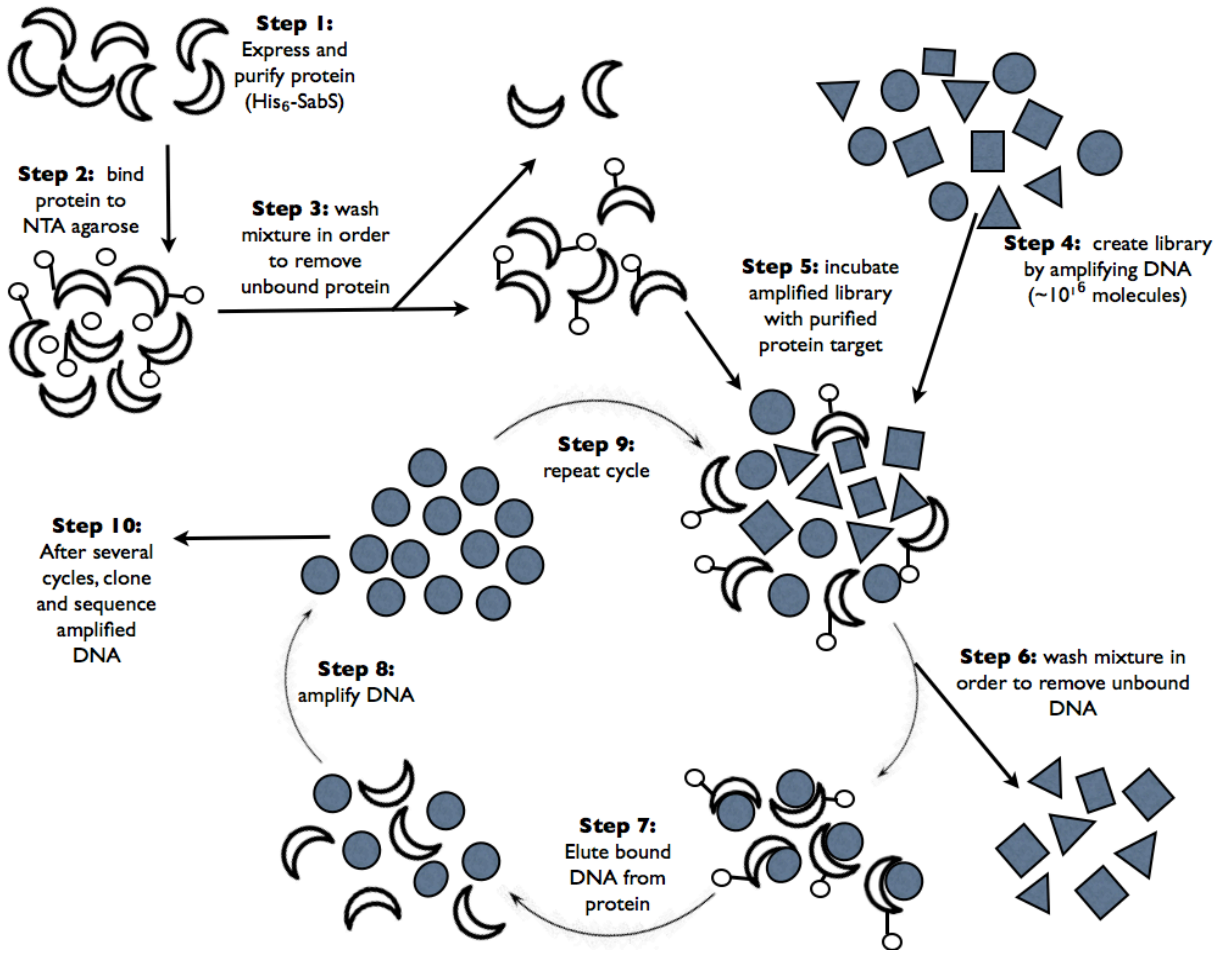

Step 1: Express and purify protein (His_6-SabS)
Step 2: bind protein to NTA agarose
Step 3: wash mixture in order to remove unbound protein
Step 4: create library by amplifying DNA (~10^{16} molecules)
Step 5: incubate amplified library with purified protein target
Step 6: wash mixture in order to remove unbound DNA
Step 7: Elute bound DNA from protein
Step 8: amplify DNA
Step 9: repeat cycle
Step 10: After several cycles, clone and sequence amplified DNA

Figure 15.

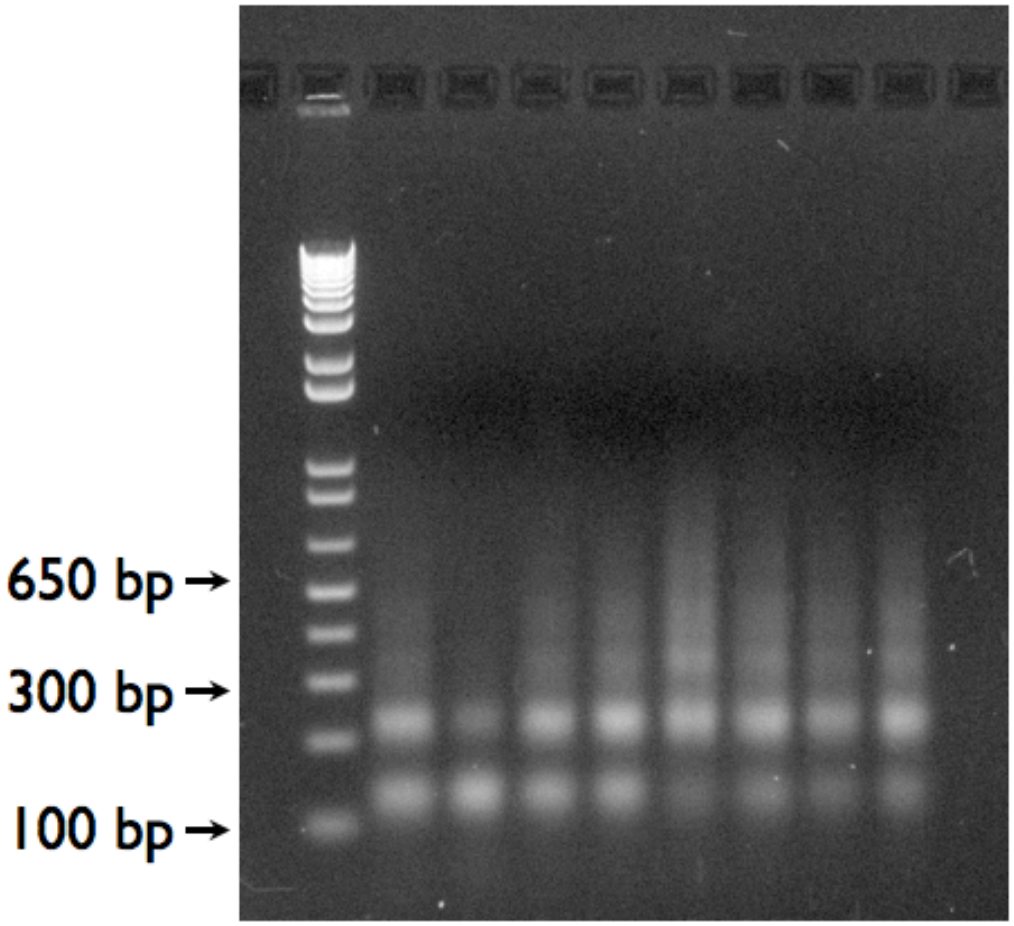


Figure 16.

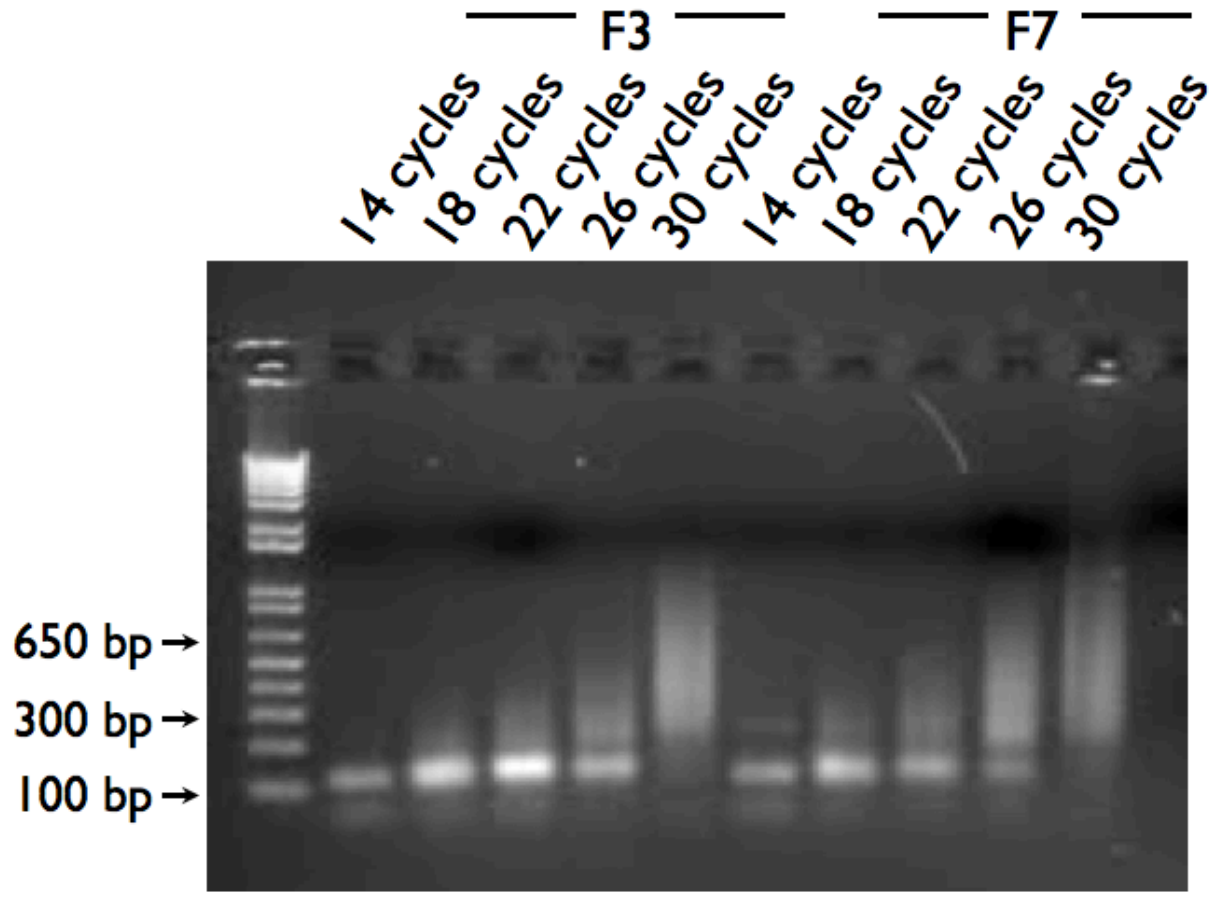

F3
F7
14 cycles
18 cycles
22 cycles
26 cycles
30 cycles
14 cycles
18 cycles
22 cycles
26 cycles
30 cycles
650 bp
300 bp
100 bp

Figure 17.

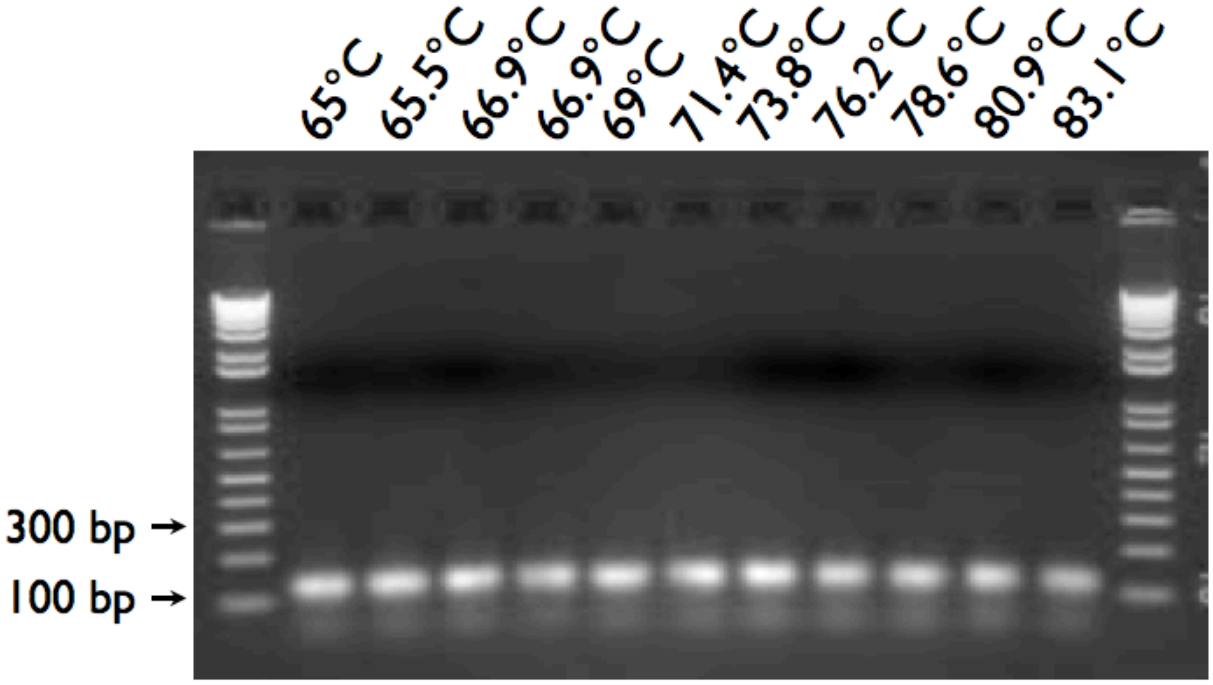

65°C
65.5°C
66.9°C
66.9°C
69°C
71.4°C
73.8°C
76.2°C
78.6°C
80.9°C
83.1°C
300 bp
100 bp

Figure 18.

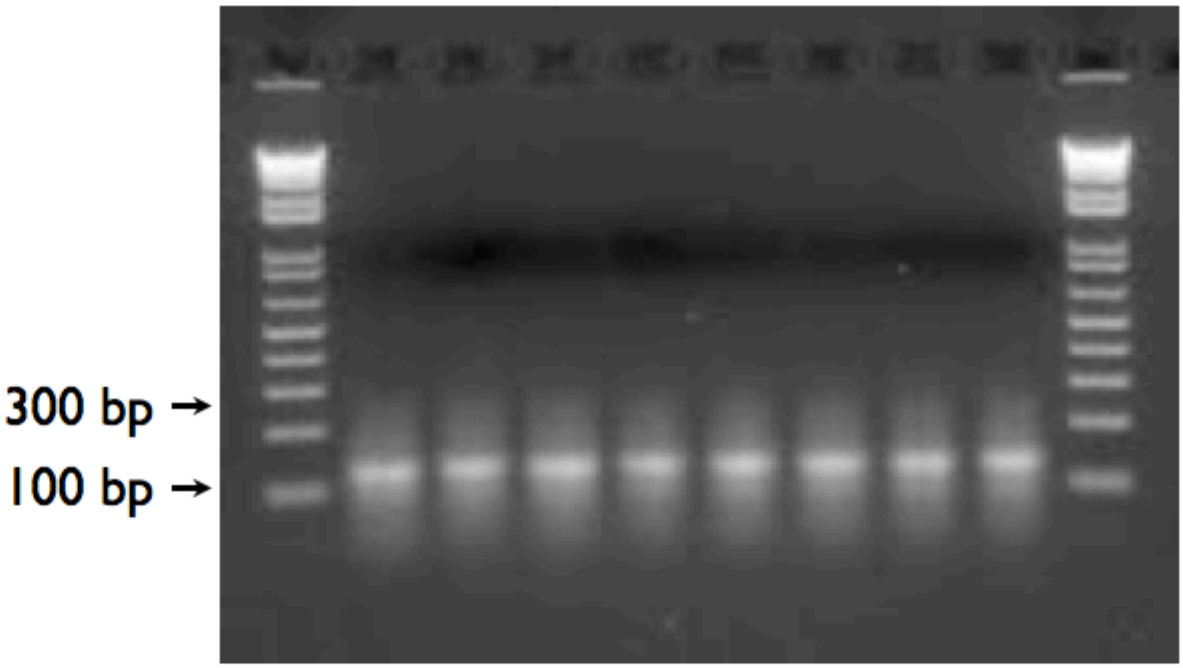


Figure 19.

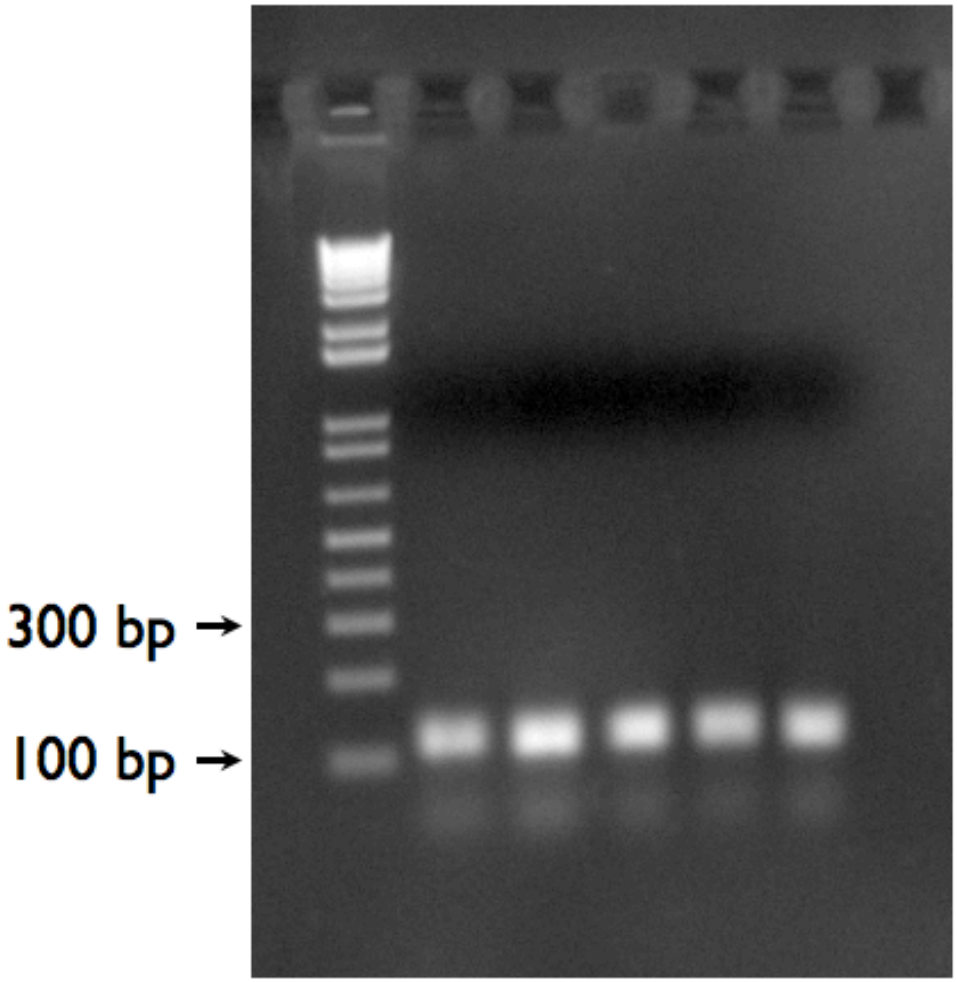


Figure 20.

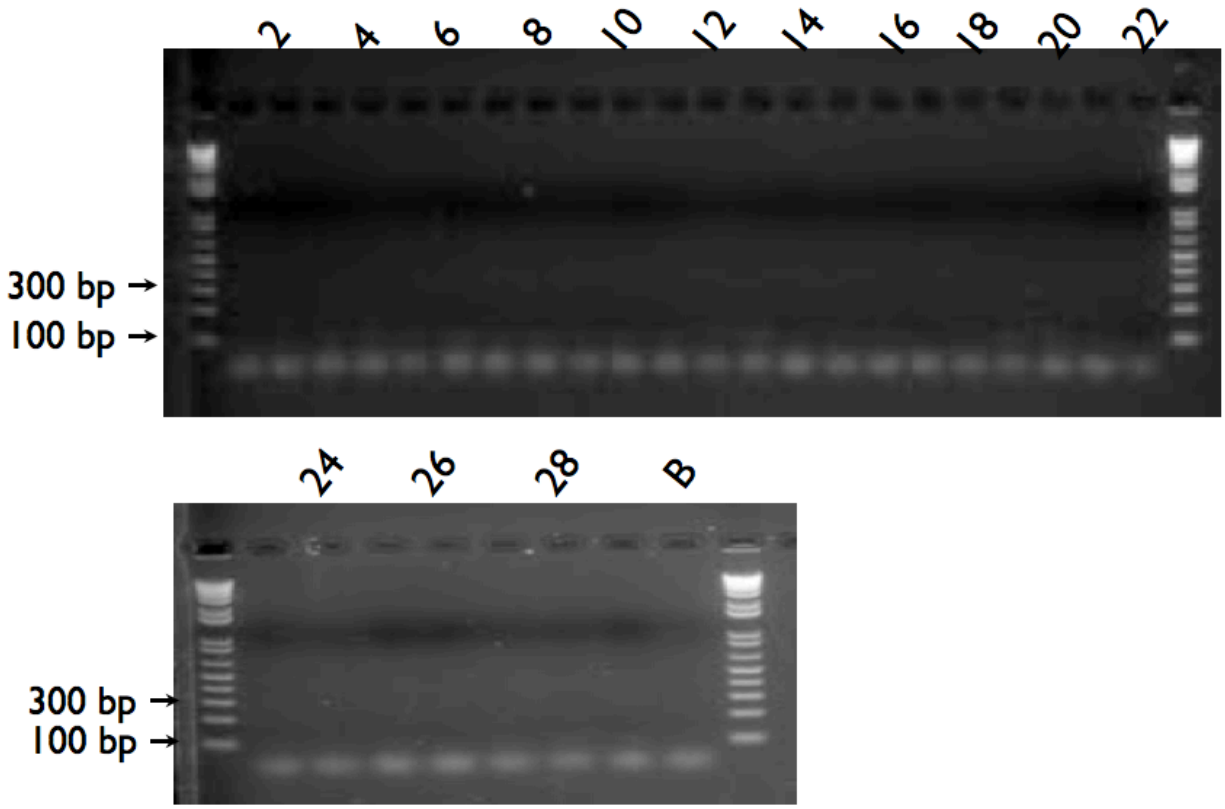


Figure 21.

Number of Clones Generated	SELEX Generated Sequence	Sequence Number
7	GTGAAAACAGGTCACGCGCCCCGTTTTG	1
2	ATATCAACAGGTCACGCGACCTGTTTTT	2
1	ATCAACACATACCGGCCGACCTGTTTCC	3
1	ACAACAACATACCGACCGACCCGTTTTT	4
2	AACGAAACAGGTCGGGCGTCCTGTTTTC	5
1	ATATCAACAGGTCAGGTGACCTGTTTTG	9
1	ACCAAAACGGGGTGGATGACCGGTTTTT	12
4	ATATCAACAGGGCGGACGACCTGTTTTG	15
1	ATATCAACGGGGTGGGTGACCGGTTTTA	27

Figure 22.

Figure 23.

sabS-ARE	AAACAAACGGCATGACCCGTTCTTTT
afsA	CAAGATACAGAATAATCGGTTTTTTT
scbA	TAAGATACAGACTGAGCGGTTTTTTT
barA-BARE3	TAAGATACATACCAACCGGTTCTTTT
scbR	ATCGGAACCGGCAATGCGGTTTGTTC
barB-BARE1	AGGCAAGCGAACCGCTCGGTTTGCTG
barB-BARE2	CCAAAAACAAGGCAACCGGTCTGGTT
papR1	TGACAAACCGACCGTGCCGTTTTTTT
spbR-ARE1	AAAGATACGTACCCACCGGTTTTGTT
spbR-ARE2	TAAGATACGAACCCGCCGGTTTCTTT
strR-ARE1	CAGGATCGCGCCACGGCGGCTGATTC
strR-ARE2	TTCGGCAATCAAACTGCGGTTTATTT
tylS	TGACAAACCGTCCGCTCCGTTTTTTT
vmsR	TCACAAACCGTATAGTCTGTTTTCAT
ccaR	GGAAAAACGTACCCCGGGGTCGGTTT
farX	AAAAATATATACCAACCGGTTTTTTC
farA	TAAGATACGAACGGGACGGACGGTTT

Figure 24.

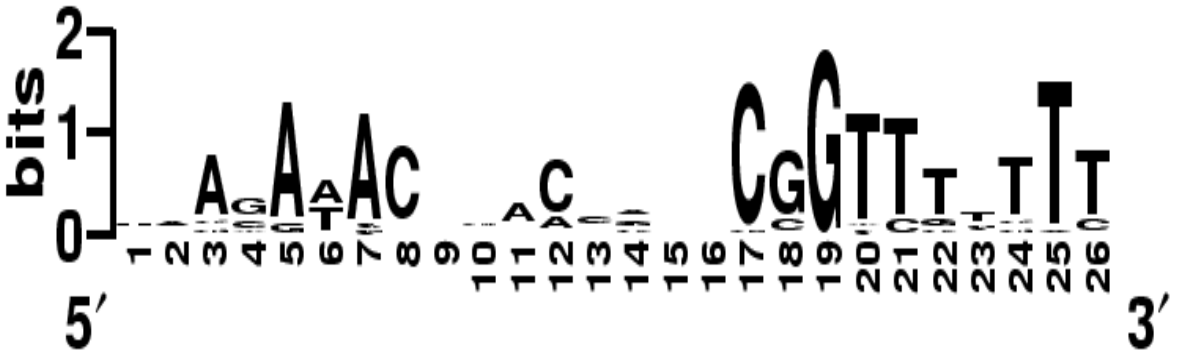

bits
2
1
0
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26
5′
3′

Figure 25.

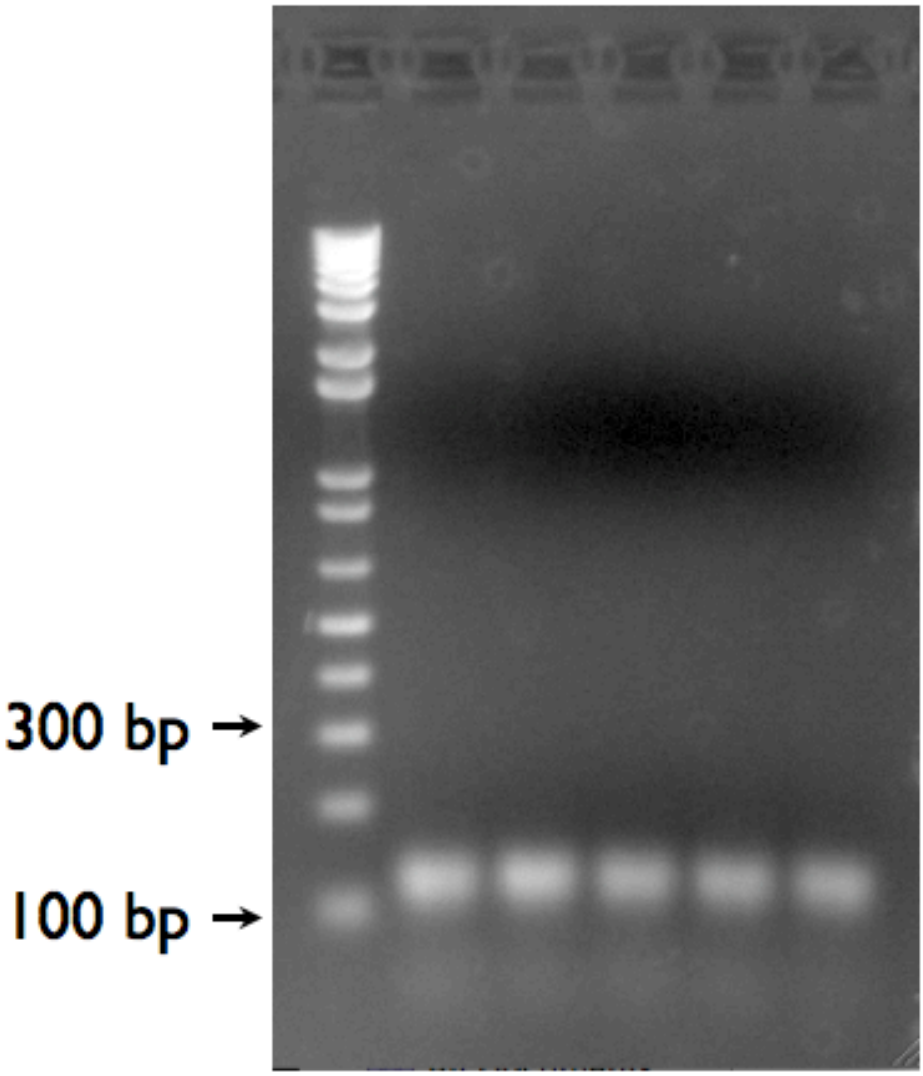


Figure 26.

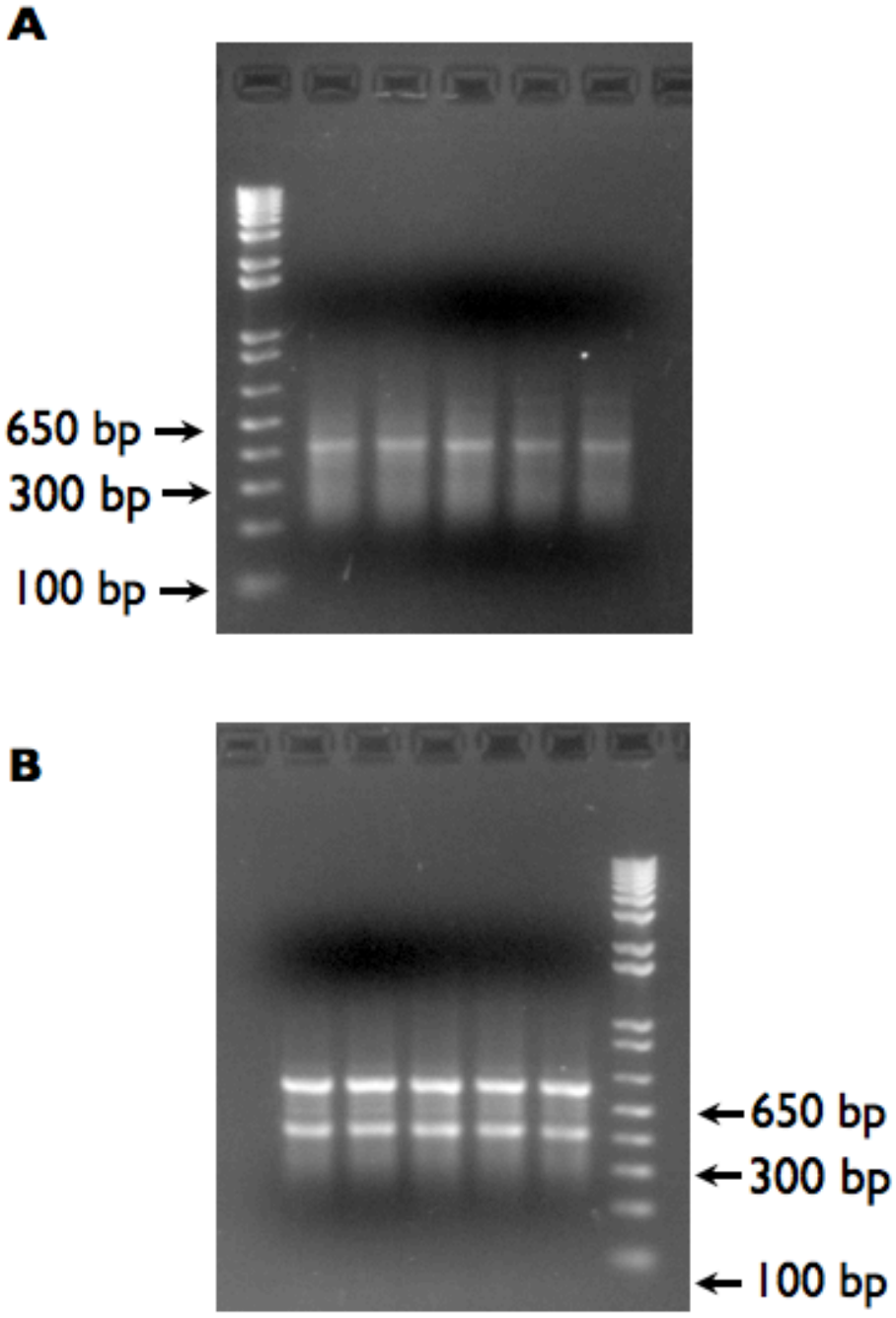


Figure 27.

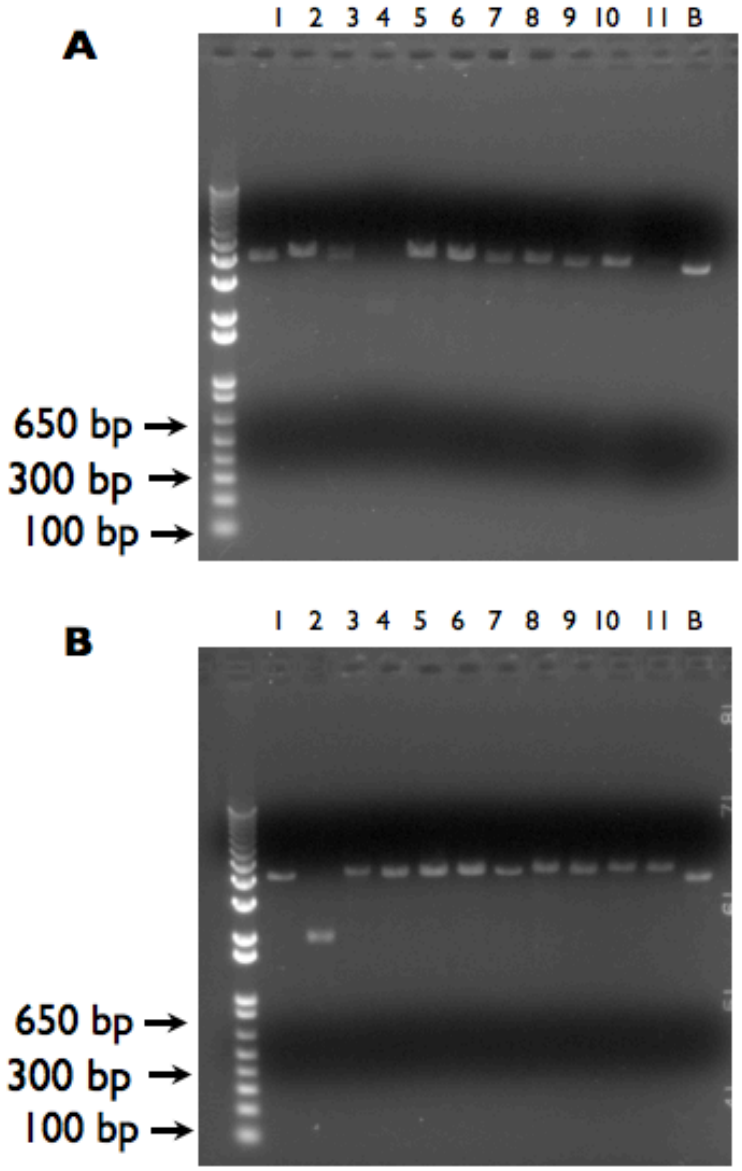

A
1 2 3 4 5 6 7 8 9 10 11 B
650 bp
300 bp
100 bp
B
1 2 3 4 5 6 7 8 9 10 11 B
650 bp
300 bp
100 bp

Figure 28.

```
>SabR 07 TGGGCACTATTTATATCAGGCGGAATGGAATAGACGCCGCTAAAATGAACAT
>SabR 08 TGGGCACTATTTATATCAGGGGGGCAACTGGAGGAATGGCCGCACCTGTATT
>SabR 10 TGGGCACTATTTATATCACCAACAAAACCAAAAAATATTGTACCTGTCAGTG


>SabR 07 CACCGAAAAATAATCCACAGCAAGACGTCACCACAATAAAACCAGCACCGCG
>SabR 08 AGCTACATACATTTCAGGATATTATGGCTTAGGTGTTTCATTATTCATTGTT
>SabR 10 TTCCTAAACCTGCTATTGTCATCAGAGTTGAAATGACCACTGTATTAACTCC


>SabR 07 GCATGAGAAATGACGTGATGACAACACGGTTATAATAAAAAACCCTAAAATA
>SabR 08 ACGGTTGCATTCTCTGCCTTATTAATTTTGTTAGTTGGTTTTGATATTCCAG
>SabR 10 ATTATCTGCCAGGTAGGAAGGAAGTAGTCCGTTTATCGGCCANTTGCCCCTA


>SabR 07 CCTAACAACACGCCGTGAGAACTTAAATCGCCAATCATCACCAGCGTGTCTT
>SabR 08 GTAAAATGTTATAAACTATCCGTGGCTAAATGATTGGAGGCTTTATGATTAA
>SabR 10 AGTTATAAATAAGACCGGTCCCTAATCCTCCTAATTTTGTTGGAAAGNANTC


>SabR 07 TATGGCGACAATAACGCCAGTATTTTTTAATCCCATTAAGGCGATAAATAAT
>SabR 08 TTATGGCGTGTTGTGATAAGATTGGTGTAAAAAAGCCTTTGTAGTCGGTCTA
>SabR 10 ATATNTNNATTTTGGAACCAACCCTGCNNTACCTANNTTTGNAAAAATTAAT


>SabR 07 CCAATTCCGCTGGTGATACCAATACGTAAATGAAAATGACGTGATGACAACA
>SabR 08 ATAACTTCATTTATTTTCCTTTGTCCTCTGTTGCATTCTCTGCCTTATTAAT
>SabR 10 CCAANNANNCATNNTCCNATANGAGAAGAATCTGCCAGGTAGGAAGGAAGTA
```

Figure 29.

```
>SabS 05 TGGGCACTATTTATATCACCAACGAGCTTCCGGACATCTTCTCGACTATCTC
>SabS 06 TGGGCACTATTTATATCAGCAGAGTATCGGACCGAGCTGGAGGACCTAGTTC
>SabS 07 TGGGCACTATTTATATCACGCCGATCTATCCGCGCCTCATTCGAGCCTTCGT
>SabS 08 TGGGCACTATTTATATCAACGCAAACTGGCAAGACCAAAAAGTGAACGGTAT


>SabS 05 CGACGGTTGCCTGGAACTGCGAGAGAGGGCCGGTTAGGTCTGCAATCTTCAT
>SabS 06 GTTCCTTCCTGCCGGAAGCGCCGTTGTGGGTCGCGGTTAATGAGCGGGATCA
>SabS 07 AGATCAGTGGACCAGGGATGAGAAGCTCAAGCAGTATGAGTCGCCGATCGCC
>SabS 08 TCCTCAAGCAACAATGAAAATTGTTTATCGTTATTATTTCCGGGATTAATTA


>SabS 05 GTCAATGAGAAAAGGAACGACATGTGCAATTTCGAACTCCTTGGCGAGTGAG
>SabS 06 GCCGGTTGGATTTATGTTGCTAAGTGGGCAGCATATGGATGCGCTGTTTATC
>SabS 07 AACCTTATCTACCAGTACTCATTGCTAGTCGATAACGAGTCGGCAGACTCTA
>SabS 08 CCAAAAGCGAACAATGGTTTAACCTGCTCCATTTATTCATTTTATCCAACCA


>SabS 05 CCTGCCTCGAAATTGATCCACTTCGACGCCTCGTTCTCTGGGGTTACGCAGA
>SabS 06 GATCCTGATGTGCGCGGCTGCGGCGTAGGTCGGGTGCTGGTGGAGCATGCGC
>SabS 07 AGCCGACGGACTTCAAGCCGGCGGCGCGGCAAAATGCAGGTGAAGGTGAAGG
>SabS 08 TGAGCGAAGCCGCTCGGCAGGAATGTTTTGCCATGCGTTATTTGCGCAAACA


>SabS 05 TGATCCCGAAATCTGTAGCGGCAAGTTCATGAGTAATCTCCTTGATCCCTCT
>SabS 06 TCTCGATGGCACCGGAACTGACAACCAATGTTAATGAGCAAAATGAGCAGGC
>SabS 07 CGACCTCATCAAGAGGTTCAATGAGATCCGCATCGGATTGACCAAGCTCTCG
>SabS 08 AGAATAAAAAGGTAATTATGGGGATCAATANAACACTGCAAATCACGGGTCA


>SabS 05 GCCTCCTTTGGCGATGTCCTGAGTCGAGACCCACGGATCAACCTCTTGGATA
>SabS 06 CGTTGGGTTCTATAAGAAGGTGGGTTTTAAGGTTACGGGACGCTCTGAGGTG
>SabS 07 CAGAGGAAAACTATCGAGGAGCTGGAGGATCTTATCCCTGATATAAATAGTG
>SabS 08 AATANTATAGACCTTTTTCATGATTTANCATTGAAAATAATTTTATTTTTTC
```

Figure 30.

```
afsA        ------------------------------------------------------MDA 3
scbA        ---------------------------------------------MPEAVVLINSASDANS 16
spbR        ------------------------------------------------------------
farA        ------------------------------------------------------------
scbR        ------------------------------------------------------------
barA        ------------------------------------------------------------
vmsR        ------------------------------------------------------------
papR1       ------------------------------------------------------------
tylS        ------------------------------------------------------------
sabS        ------------------------------------------------------MT 2
ccaR        ------------------------------------------------------------
barB        ------------------------------------------------------------
farX        ------------------------------------MSVVLDDSRTMKAPSMGATVE 21
strR        MELSSRRQFDMHDLSSRGVSELNGNSAKLSGVVMVPIASLRPSDSPRSAGEDPEHIRALA 60
tetR        ------------------------------------------------------------

afsA        EAEVVHPVGIEMVHRTRPEDAFPRNWVRLGRDRFAVEAVLPH------------------ 45
scbA        IEQTALPVPMALVHRTRVQDAFPVSWIPKGGDRFSVTAVLPH------------------ 58
spbR        -----------MARQERA--------VRTRRAILVAAAEVFD------------------ 23
farA        -----------MAQQERG--------NRSRRSILEAAARVFD------------------ 23
scbR        -----------MAKQDRA--------IRTRQTILDAAAQVFE------------------ 23
barA        ---MAVRHERVAVRQERA--------VRTRQAIVRAAASVFD------------------ 31
vmsR        ---MAVRHERVAVRQERA--------VRTRQAIVRAAASVFD------------------ 31
papR1       ---MDIDILGALSARENN------ITVTPTAPKPRQVLALLA------------------ 33
tylS        ---MDIAVLGPLDVRENG------LSVTPTAPKPRQVLALLA------------------ 33
sabS        RSFTGPGVSQAMARQLRA--------EQTRATIITAAADLFD------------------ 36
ccaR        ---MTIRLLGPVTLVKGS------VPIPIRGQRQRRFLASLA------------------ 33
barB        ----------MTPKQERA--------FRTRTQLVLSAAEAFD------------------ 24
farX        VLSFLQPVTRELVHRSSIAEVFVTDGVRTGENAFSVGAQWPR------------------ 63
strR        ESGAELPPIIVMASTMRVVDGMHRLRAAELRGESELAVRFFEGDEKEAFVLAVEANITHG 120
tetR        ----------MMSRLDKS------------KVINSALELLN------------------ 19

afsA        -----DHPFFAPVGDDLHDPLLV--AEAMRQAAMLAFHAGYGIPLGYHFLLTELDYVCHP 98
scbA        -----DHPFFAPVHGDRHDPLLI--AETLRQAAMLVFHAGYGVPVGYHFLMATLDYTCHL 111
spbR        -----EVGYEAAT----ISEILK--RSGLTKGALYFHFASKEELAQGVLAEQVHALPDLP 72
farA        -----ERGYDAAS----TNEILA--STGLTRGALYHHFPSKEAIASALVAAQSEAL-VVP 71
scbR        -----KQGYQAAT----ITEILK--VAGVTKGALYFHFQSKEELALGVFDAQEPPQ-AVP 71
barA        -----EYGFEAAT----VAEILS--RASVTKGAMYFHFASKEELARGVLAEQTLHV-AVP 79
vmsR        -----EYGFEAAT----VAEILS--RASVTKGAMYFHFASKEELARGVLAEQTLHV-AVP 79
papR1       -----LNAGQVVPMSALSEELWGPTPPRSARTTVQTYILQLRELITRALGHG---NTGRT 85
tylS        -----LHADQVVPVSALIEELWGERPPRSARTTLQTYVLQLRELISAAITNDPEEARPRT 88
sabS        -----RHGYESTS----LSDIVE--HAQVTKGALYFHFAAKEDLAHAIMELQSQASQRVA 85
ccaR        -----LRPGQVISKEAIIEDSWDGEPPLTVSGQLQTSAWMIR----TALAEAGLPRDALG 84
barB        -----RQGFATAS----LTAISN--SAGVSNGALHFHFESKEALAAAVEAEAAERMRTIV 73
farX        -----DHALYHPDENGLNDPLLF--AETLRQAHFYGAHTYFGVPVGSRFIGQDVSFEITD 116
strR        LPLGLQDRKDAAARILRSHPMWSDRAIGTATGLSAKTVAALRARSTEGLPQSNVRIGRDG 180
tetR        -----EVGIEGLT----TRKLAQ--KLGVEQPTLYWHVKNKRALLDALAIEMLDRHHTHF 68

afsA        EYL------------------GVGGEPTEIGLEVFCSDLKWRAGLPAQGRVGWAVHRGD 139
```

```
scbA        DHL-------------------GVSGEVAELEVEVACSQLKFRGGQPVQGQVDWAVRRAG 152
spbR        EGE-------------------LMLQTAVDRALLLAHLLRRDTGDPIVRGSVRLTVEQGA 113
farA        DQP-------------------VKLQAVIDLTLEFARRLQHD---PVLRASVRLAVEQTS 109
scbR        EQP-------------------LRLQELIDMGMLFCHRLRTN---VVARAGVRLSMDQQA 109
barA        ESG-------------------SKAQELVDLTMLVAHGMLHD---PILRAGTRLALDQGA 117
vmsR        ESG-------------------SKAQELVDLTMLVAHGMLHD---PILRAGTRLALDQGA 117
papR1       AKD-------------------VLTTGPGGYRLDVGGGTLDFREFERRAGAGYRAMDAGD 126
tylS        AKQ-------------------VLVTTPGGYVLDTSGGTSDVCEFEHMAGTGHRAMDAGD 129
sabS        GEIDNRGY--------------SSLEALMRLTFGLTRLSVEG---PIARAGLRLATGGVA 128
ccaR        SHD-------------------------RGYELRVLPDSIDLFVFREAVRAVRDLHARGQ 119
barB        DGAARRGA--------------SALQALVDTSHAVMLRLRQD---VVVRAGFRLSGDAAR 116
farX        PTALR-------------------VGAAPLAVVLNGTWTEERDRRGRPAGARLDVTLTVDG 158
strR        RARPVDPAEGRRLAGRLMQENPSAPLRQIAAQAGVSLGTVSDVRKRLQRGEGPVPERGRA 240
tetR        CPLEGE----------------SWQDFLRNNAKSFRCALLSHRDGAKVHLGTRPTEKQYE 112


afsA        RLAATGVAATRFSTPKAYRRMRGDVPVEGISLPETAPVPA-----SPAGRARVEDVVLSG 194
scbA        RLAATGTATTRFTSPQVYRRMRGDFATPTASVPGTAPVPA-----ARAGRTRDEDVVLSA 207
spbR        LRDGLDRRVPMQAWMEQTQDLFEQAQAAGEILPHVDLVGA-----AKTFVGAFTGVQVLS 168
farA        FSR--PAQTPYEQSNATILDLLRQAREQGEVLPGVDLQEA-----TSTIVGAFTGLQVMS 162
scbR        HGL--DRRGPFRRWHETLLKLLNQAKENGELLPHVVTTDS-----ADLYVGTFAGIQVVS 162
barA        VDF--SDANPFGEWGDICAQLLAEAQERGEVLPHVNPKKT-----GDFIVGCFTGLQAVS 170
vmsR        VDF--SDANPFGEWGDICAQLLAEAQERGEVLPHVNPKKT-----GDFIVGCFTGLQAVS 170
papR1       YPG---AAKRLRDALSLWTGSALADVPAGARLTLEARRLE-----EARLCALDQRIEADL 178
tylS        FAG---AARQLRDALGLWTGSAFSDIQTGLRLTMEARRLE-----ETRLCALDQRIEADL 181
sabS        VRP--PLRRPFSEWGELINRGLSGAVREADIHQDVDIEAV-----AHSLVCFFVGTRVVG 181
ccaR        HQE---ASERLDTALALWKGPAFADVTSS-RLRLRGETLE-----EERTAAVELRALIDV 170
barB        QAT----HDLPEHWRQSVVRLLERAGRDGSLTSAVTPSDV-----AGVVTATVLGFGVLA 167
farX        RPCGRGHTRGLMLDDRRYRLLRGRPAASGEVSPPRPAPDARIARPNRVGRLRWKDCVLER 218
strR        RTEPGTNPEPAHTPERPYDEAEAERPSRALILRHLSRDPS--VRLTEDGRTLLRWLTVAA 298
tetR        TLEN--QLAFLCQQGFSLENALYALSAVGHFTLGCVLEDQ---------EHQVAKEERET 161


afsA        TGREGVWELRVDTRHLTLFQRPNDHVPGM--------LLLEAARQAACLVAGPAGIVPVE 246
scbA        SSQQDTWRLRVDTSHPTLFQRPNDHVPGM--------LLLEAARQAACLVTGPAPFVPSI 259
spbR        NIMTGRQDMTERVADLYRFLMTAIAVPGV--------LVRLDFSPGRGV-LAYEEAVRRR 219
farA        QVYSNRQDLPDRVCALWRFLLPGLATPGM--------LARL-------------------- 195
scbR        QTVSDYQDLEHRYALLQKHILPAIAVPSV--------LAALDLSEERGARLAAELAPTGK 214
barA        RVTSDRQDLGHRISVMWNHVLPSIVPASM--------LTWIETGEERIGKVAAAAEAAEA 222
vmsR        RVTSDRQDLGHRISVMWNHVLPSIVPASM--------LTWIETGEERIGKVAAAAEAAEA 222
papR1       RLGRHRELLAELTVLVGRHRLHESLHGQF--------MLALHRSGRRGEALSVYQRLRTT 230
tylS        RLGRHRELLGELTMLVSRHRTHENLHGQL--------MLALHRSGRRSEALGVYQRLRTA 233
sabS        RSREASGRLPRRMAEMWYVLIRGLVPVPR--------RPRYLSLASRLEREITAACVS-- 231
ccaR        GLGYYGDAITRLSELVDHDPFREDLYVSL--------MKAYYAEGRQADAIQVFHRAKDI 222
barB        R-FDSAWLASGSLSGFWKLMLPMIAAGPVE-------RGELDCRPAVPADVRRAPAV--- 216
farX        DRPDQDWRLRVDRDHAVLFDHPTDHVPLMVMLEGFRQLGHLTVHEASRRTLGDRAFALAG 278
strR        VRPQDWERLLRNVPAHRVGAVAELARGCS---QTWQLVAEQLERARGSGAEQETGGAPGR 355
tetR        PTTDSMPPLLRQAIELFDHQGAEPAFLFG--------LELIICGLEKQLKCESGS----- 208
```

```
afsA        ARTRFHRYSEFGSPCWIGAV--VQPGTDEDTVTVRVTGHQDGETVFSTVLSGPRAHG--- 301
scbA        GGTRFVRYAEFDSPCWIQAT--VRPGPAAGLTTVRVTGHQDGSLVFLTTLSGPAFSG--- 314
spbR        DAAPQPAAH---------------------------------------------------- 228
farA        ----------------------------------------------------------
scbR        D----------------------------------------------------------- 215
barA        AEASEAASDE-------------------------------------------------- 232
vmsR        AEASEAASDE-------------------------------------------------- 232
papR1       LVTELGLEPSAQLSRLQRSL--LTAGPETPVPPAPAAAAAGHRDGRPAGHRPAATAG--- 285
tylS        LIRELGLEPSPALRRIQRYV--LMAGPE----PAAVGAAAGR-----AGHLSPAG----- 277
sabS        ----------------------------------------------------------
ccaR        LREQIGISPGERMTRVMQAI--LRQDEQVLRVGTPA------------------------ 256
barB        ----------------------------------------------------------
farX        LSLDCAAFGELGETILLSLEKGPSEGVPTEECALRVAAHQGERLLARADMTWKCVGSRAP 338
strR        QEAEGAGQSAAGMPRRAAALESRRTGEVTVQVPDAPDRPVPPADPCCDTVLRGPASAGAP 415
tetR        ----------------------------------------------------------
```

Figure 31.

Tables

Table 1.

Strain or plasmid	Relevant properties and/or function	Reference or Source
Strains		
S. acidiscabies 84.104	wild-type WS5995B producer	King et al. 1996
E. coli BL21λDE3	strain for recombinant protein expression	Novagen
Plasmids		
pCR 2.1 Topo	general cloning vector	Invitrogen
pET26b::sabR	pET26b carrying *sabR* encoding GBL receptor SabR	Healy et al. 2009
pET26b::sabS	pET26b carrying *sabS* encoding GBL receptor SabS	Healy et al. 2010

Table 2.

Primers and Templates	Sequence (5'-3')
224	TGGGCACTATTTATATCAAC(N_{28})AATGTCGTTGGTGGCCC
225	CGCGGATCCTAATACGACTCACTATAGGGGCCACCAACGACATT
226	CCCGACACCCGCGGATCCATGGGCACTATTTATATCAAC
229	TGGGCACTATTTATATCA(N_9)
230	TGGGCACTATTTATATCA

Table 3.

Program	Initialization Step	Denaturation Step	Annealing Step	Elongation Step	Number of Cycles	Final Elongation
GenSELEX	94°C, 2’	94°C, 1’	50°C, 1’	72°C, 2’	30	72°C, 6’
SELEX1	94°C, 2’	94°C, 1’	80°C, 0.45’	72°C, 1’	14	72°C, 10’
SELEX2	94°C, 2’	94°C, 1’	55°C, 0.45’	72°C, 1’	30	72°C, 10’

Table 4.

GBL receptor protein	Residues involved in DNA binding	ARE sequence
sabS	**KGA**L**YFH**	AAACAAACGGCATGACCCGTTCTTTT
scbR	**KGA**L**YFH**	ATCGGAACCGGCAATGCGGTTTGTTC
spbR	**KGA**L**YFH**	AAAGATACGTACCCACCGGTTTTGTT
spbR	**KGA**L**YFH**	TAAGATACGAACCCGCCGGTTTCTTT
barA	**KGA**M**YFH**	TAAGATACATACCAACCGGTTCTTTT
vmsR	**KGA**M**YFH**	TCACAAACCGTATAGTCTGTTTTCAT
barB	**NGA**L**HFH**	AGGCAAGCGAACCGCTCGGTTTGCTG
barB	**NGA**L**HFH**	CCAAAAACAAGGCAACCGGTCTGGTT
farA	**RGA**L**YHH**	TAAGATACGAACGGGACGGACGGTTT
strR	**TGL**S**AKT**	CAGGATCGCGCCACGGCGGCTGATTC
strR	**TGL**S**AKT**	TTCGGCAATCAAACTGCGGTTTATTT
ccaR	**SGQ**L**QTS**	GGAAAAACGTACCCCGGGGTCGGTTT
tylS	**RTT**L**QTY**	TGACAAACCGTCCGCTCCGTTTTTTT
papR1	**RTT**V**QTY**	TGACAAACCGACCGTGCCGTTTTTTT
afsA	**QAA**M**LAF**	CAAGATACAGAATAATCGGTTTTTTT
scbA	**QAA**N**LVF**	TAAGATACAGACTGAGCGGTTTTTTT
farX	**QAH**F**YGA**	AAAAATATATACCAACCGGTTTTTTC