ABSTRACT

ZIRKLE, ROSS E. The genetics of *Sorangium cellulosum* So ce26 and the biosynthesis of the antifungal polyketide soraphen A. (Under the direction of James M. Ligon.)

Sorangium cellulosum is a gram-negative soil bacterium belonging to the order Myxococcales. The myxobacteria are distinguished by their complex morphological differentiation and ability to move on agar by gliding. Under conditions of starvation, myxobacteria form aggregates known as fruiting bodies in which vegetative cells are transformed into myxospores. The myxobacteria also produce a large number of diverse natural products. Many of these natural products have antimicrobial, antifungal and insectiscidal activities. The genus *Sorangium* is a particularly rich source of secondary metabolites.

While strains of *Sorangium* are a good source of secondary metabolites, relatively little has been accomplished in the development of genetic systems for these organisms. Due to poor growth characteristics and the relative lack of genetic tools, experiments with *S. cellulosum* have been especially difficult. The results presented here describe the cloning and disruption of *mglA*, a gene from *S. cellulosum* So ce 26. An *mglA* mutant strain of *S. cellulosum* So ce26 displays a non-swarming phenotype and allows the isolation of single colonies.

DNA gene transfer into *S. cellulosum* can only be achieved through conjugation, followed by homologous recombination of cloned DNA fragments on a transmissible plasmid and the corresponding chromosomal locus. The results presented here demonstrate an improvement of the efficiency of this system in *S. cellulosum* So ce26.

Soraphen A is a secondary metabolite produced by *S. cellulosum* So ce26. Soraphen A is an antifugal macrolide that is synthesized by a type I polyketide synthase. The genes responsible for the synthesis of the soraphen polyketide synthase have been previously cloned and sequenced. The role in soraphen A biosynthesis of genes upstream of the core soraphen polyketide synthase genes has been proposed, but not experimentally determined. These genes include *sorC*, which is proposed to be involved in the production of the unusual extender unit, methoxymalonate, utilized by the soraphen polyketide synthase. The gene *orf2*, which is upstream of *sorC*, is proposed to have no role in the production of soraphen A. In the results presented here, both *sorC* and *orf2* were disrupted and data obtained from the resulting strains demonstrates that *sorC* is required for the biosynthesis of soraphen A while *orf2* has no apparent role in soraphen A biosynthesis.

Due to the relative lack of satisfactory genetic systems in *S. cellulosum*, the heterologous production of soraphen A in a more amenable host was sought. Heterologous production would allow for detailed mechanistic studies of soraphen A biosynthesis as well as the production of engineered soraphen derivatives. The soraphen polyketide synthase genes *sorA* and *sorB* along with the methyltansferase *sorM* were cloned into expression vectors and integrated into the chromosome of *Streptomyces lividas* ZX7. *SorC*, *sorD*, *sorE* and *sorR*, all of which are genes upstream of *sorA*, were cloned into an expression plasmid with *badA*, a benzoate-coenzyme A ligase from *Rhodopseudomonas palustris*. This expression plasmid was then transformed into the *St. lividans* strain containing the integrated *sorA*, *sorB* and *sorM.* The resulting strains produced soraphen A when putative precursors for the

production of the benzoyl-coenzyme A starter unit, sodium benzoate or trans-cinnamic acid, were supplied to the fermentations.

THE GENETICS OF SORANGIUM CELLULOSUM SO CE26 AND THE BIOSYNTHSIS OF THE ANTIFUNGAL POLYKETIDE SORAPHEN A

by

ROSS ERIC ZIRKLE

A dissertation submitted to the Graduate Faculty of
the North Carolina State University
in partial fulfillment of the
requirements for the Degree of
Doctor of Philosophy

MICROBIOLOGY

Raleigh

2003

APPROVED BY:

_____	_____
Co-Chair of Advisory Committee	Co-Chair of Advisory Committee

BIOGRAPHY

In 1989 I graduated from Northeastern High School in Manchester PA. I then attended Virginia Tech and received my Bachelor's of Science degree in Biochemistry in 1993. I continued my education at Virginia Tech and earned a Master's of Science in Biology (Microbiology section) in 1995. My graduate studies were under the direction of Alumni Distinguished Professor Dr. Noel Kreig. Under his guidance, I studied various aspects of the microbiology of *Campylobacter* and other mircoaerophilic bacteria. My thesis and subsequent publication in the Journal of Applied Bacteriology (1996 81:133-138) was on a method to analyze DNA damage and repair in *Escherichia coli*. After completing my degree, I was hired as a temporary associate in the Combinatory Biosynthesis Group at Ciba in the Research Triangle Park in 1996. I am currently a Research Scientist at Syngenta (formally Ciba and Novartis) in the Natural Products Genetics Group. During my time at Syngenta, I have been an author on various patents and journal articles. In January 1999, under the encouragement of my supervisor at that time, Dr. James Ligon, I started the graduate program in Microbiology at North Carolina State University and a concurrent research project at Syngenta which is presented here.

ACKNOWLEGMENTS

There are many people who deserve recognition for having made it possible for me to complete this work. First and foremost is my family; Twilla, Eric and Andrew. Without the unwavering commitment and dedication of my wife, Twilla, achieving my goals would have be impossible. I also extend sincerest thanks to my parents, Doll and Helen Zirkle, for being terrific parents who sacrificed so much for all of my education. I would like to express my special thanks to my former supervisor and co-chair, James M. Ligon for his confidence in me and encouragement. I would also like to thank István Molnár for his supervision in my scientific development, excellent technical expertise and assistance, and critical reading of this manuscript. I am also grateful to Dr. Eric Miller, Dr. Greg Upchurch and Dr. Hosni Hassan for their excellent advice and recommendations and willingness to serve on my committee. I would also like to thank colleagues that I have had the privilege working with at Syngenta; Todd Black, Phillip Hammer, Brian Thompson, Amber Gaudreau, Frank Gross, John Mayo and Brandon Mears. I would also like to acknowledge Steve Hill's initial efforts that were critical for the success of this project. I would like to thank members of the Syngenta Analytical Chemistry and Sequencing groups for their assistance with analysis of samples; James Pomes, Belle Abrera, Lina Summerfelt, Makoto Ono, An Hu and Racella McNair. I would lastly like to thank Alumni Distinguished Professor Dr. Noel Kreig at Virginia Tech for affording me my first opportunity to study science under his kind tutelage.

TABLE OF CONTENTS

TABLE OF CONTENTS (continued)

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1. LITERATURE REVIEW: POLYKETIDES AND *SORANGIUM CELLULOSUM*

CHAPTER 2. GENERATION OF A NON-SWARMING STRAIN OF *SORANGIUM CELLULOSUM*

CHAPTER 3. CHARACTERIZATION OF THE BIOSYNTHETIC GENE CLUSTER FOR THE ANTIFUNGAL POLYKETIDE SORAPHEN A FROM *SORANGIUM CELLULOSUM* SO CE26

LIST OF FIGURES (continued)

CHAPTER 4. ROLE OF GENES LOCATED UPSTEAM OF THE SORAPHEN A POLYKETIDE SYNTHASE GENES

CHAPTER 5. HETEROLOGOUS EXPRESSION OF SORAPHEN A IN *STREPTOMYCES LIVIDANS*

CHAPTER 1. LITERATURE REVIEW: POLYKETIDES AND *SORANGUIM CELLULOSUM*

INTRODUCTION

Microorganisms produce a collection of natural product molecules that have a variety of roles including self-defence and intercellular communication. These secondary metabolites have been successfully exploited for use in a wide range of applications such as pharmaceutical, veterinary and agricultural products. These secondary metabolites, or their derivatives, have been valuable in these applications, but have also contributed to our understanding of novel targets and modes of action for further active ingredient discovery. Although the *Streptomyces* have been the traditional reservoir of screening and characterization of novel natural products, the gram negative myxobacteria have also been shown to be prolific producers of bioactive secondary metabolites (reviewed by Reichenbach and Hofle, 1993; Reichenbach, 2001). The secondary metabolites produced by myxobacterial strains often display an impressive diversity in chemical structure and biological activity that have not been previously discovered in other natural product screening efforts. Polyketides are an important group of secondary metabolites produced by a number of organisms. Polyketides display a wide variety of structural diversity as well as a broad range of biological activities and have been the focus of investigation in a number of industrial and academic laboratories. Strains of the myxobacterium *Sorangium cellulosum*, have been particularly proficient in the number and diversity of polyketides produced (Reichenbach and Hoefle, 1993; Reichenbach, 2001). The enthusiasm for the production of secondary metabolites from strains of *S. cellulosum* has, however, been tempered by the poor growth characteristics and difficulties in the genetic manipulation in *Sorangium*.

The Myxobacteria

The myxobacteria are a group of gram-negative, mainly soil-dwelling bacteria belonging to the delta subdivision of proteobacteria (Figure 1) and have a high G + C content (~70%) in their genomic DNA (reviewed by McCudry, 1989; Reichenbach, 1999). Myxobacteria are usually isolated from soil, rotting plant material, animal feces and tree bark, and produce extracellular hydrolytic enzymes that are able to degrade proteins, nucleic acids and some polysaccharides. The myxobacteria have two characteristics that distinguish them from other bacteria. First, myxobacterial cells move by gliding motility in which cells move in the direction of their long axis on agar surfaces without the use of flagella. This gliding motility is utilized to accomplish swarming under vegetative conditions, during which cells of myxobacteria secrete hydrolytic enzymes that lyse other cells, and convert their proteins into amino acids which are then metabolized. Second, myxobacteria can undergo complex cellular morphogenenesis to form aggregates called fruiting bodies in which dormant myxospores are generated in response to starvation. The fruiting bodies can be very simple in shape and structure but also may occur as extremely complex designs.

The taxonomy of myxobacteria has been somewhat debatable. The classical taxonomic classification in the order of myxobacteria from Bergey's Manual of Systematic Bacteriology is listed in Table 1 (McCudry, 1989). An alternate taxonomic classification which is in wide use was suggested by Reichenbach and is also shown in Table 1 for comparison (Reichenbach and Dworkin, 1981; Reichenbach and Hoefle, 1993). This alternate classification is based mainly on morphological

characteristics such as shape of vegetative cells, myxosprores, and fruiting bodies, but it is also corroborated by 16S rRNA studies (Reichenbach and Hoefle, 1993).

The most well-studied myxobacterium is *Myxococcus xanthus*, for which a number of genetic and molecular biological techniques are available. A focus of study of *M. xanthus* has been their social behavior and development (reviewed by Dworkin, 1996; Spormann, 1999). One of the key elements in the development of *M. xanthus* is its gliding motility. In *M. xanthus*, wild-type gliding motility and cell swarming is the result of two separate systems. The S-, or social system, is responsible for the motility of groups of cells and is equivalent to twitching motility as seen in *Neisseria* and *Pseudomonas* (Comolli et al., 1999; Wall and Kaiser, 1999; Merz et al., 2000; Beatson et al., 2002). Twitching motility is mediated by the extension and retraction of type IV pili. The A-, or adventurous system, is responsible for the motility of single cells and is related to gliding motility. Secretion-mediated gliding displayed by some cyanobacteria is due to the expulsion of a polyelectrolyte gel through membrane pores and is proposed to provide the thrust for A- type gliding (McBride, 2001; Thomasson et al., 2002; Wolgemuth et al., 2002). Many genes in *M. xanthus* have been shown to effect A- and S- type gliding separately, and a few have been shown to effect both systems.

The gene *mglA*, from *M. xanthus*, has been shown to effect both A- and S- motility. The *mglA* gene was discovered in a screen for non-motile isolates which displayed non-swarming phenotype and compact, discrete colonies in *M. xanthus* (Hodgkin and Kaiser, 1979b; Hodgkin and Kaiser, 1979a). It was originally thought that MglA was a component of the gliding apparatus. It was subsequently discovered that *mglA* mutant strains were motile, but cells had a high frequency of reversal in

their direction of motility, yielding no net movement, thereby resulting in a non-motile phenotype (Spormann and Kaiser, 1999). Unlike A^-S^- double mutants, *mglA* mutant strains are still capable of movement and express pili, fibrils and other components required for A- and S- motility (MacNeil et al., 1994). This suggests that MglA is involved in the coordination of motility systems. MglA has been shown to be a GTPase (Hartzell and Kaiser, 1991), and the *Saccharomyces cerevisiea* GTPase *SAR*1, was able to partially complement a *mglA* mutant strain of M. *xanthus* (Hartzell, 1997). It has been recently shown that MglA interacts with a tyrosine kinase, MasK, in *M. xanthus* and is proposed to be an intracellular switch that coordinates A- and S-motility (Thomasson et al., 2002).

Genetic Systems in Myxobacteria

Apart from *M. xanthus*, the number of published reports on the molecular manipulations in myxobacteria have been relatively few. The main reason for this is the generally poor growth characteristics of these bacteria including long generation times and swarming of cells that makes colony isolation difficult. Although *S. cellulosum* strains produce a variety of structurally diverse and potentially valuable secondary metabolites, there few genetic tools available to investigate the cellular and molecular biology of these strains. No plasmids have been found to replicate in *S. cellulosum* strains and phleomycin resistance has been the only selectable marker described in *S. cellulosum* until a recent report on the use of hygromycin resistance in *S. cellulosum* So ce56 (Pradella et al., 2002). The only method available for stable maintenance of transferred DNA in *S. cellulosum* involves homologous recombination between DNA fragments cloned on a transmissible plasmid and the chromosomal

locus (Jaoua et al., 1992). Improvements in this system and discovery of new genetic tools would facilitate the investigation of biosynthesis of secondary metabolites and other studies.

Myxobacterial Secondary Metabolites

The secondary metabolites that are produced by myxobacteria display a wide diversity of structures and biological activities. One research group reported the isolation of ~80 different basic structures and 450 variants of molecules from myxobacteria (Reichenbach, 2001). Many of these molecules are polyketides or polyketides with incorporated amino acids. About 40% of these structures were completely novel while the others posses structural elements that have been described from other organisms (Reichenbach, 2001). Only a few of the molecules isolated were identical to previously known compounds including pyrrolnitrin (from *Pseudomonas*), althiomycin and saframycin (from *Streptomyes*). Examples of secondary metabolites isolated from myxobacteria are shown in Figure 2 and the mechanisms of action of a number of natural products isolated from myxobacteria are listed in Table 2.

S. cellulosum is one of the myxobacteria that has been of considerable interest due to the secondary metabolites that have been isolated from its numerous strains. In one screening report, ~95% of *Sorangium* strains were found to display activity in comprehensive biological screens while only ~20-50% of other myxobacterial strains produced some activity in these screens (Reichenbach and Hoefle, 1993). Two secondary metabolites, soraphen A and epothilone, that are produced by *S. cellulosum* have attracted considerable research attention. Soraphen A is an

antifungal polyketide that is produced by the strain So ce26 (Gerth et al., 1994). The mode of action of soraphen A was determined to be the inhibition of the acetyl-coenzyme A carboxylase in a wide variety of fungi (Vahlensieck et al., 1994). This unique mode of action generated considerable interest due to its potential value in agriculture for products against plant pathogenic fungi, until testing revealed that it was a teratogen. Epothilone is a polyketide/peptide hybrid molecule that is produced by the strain So ce90 (Gerth et al., 1996) and is currently in clinical trials as an anticancer drug. There was a substantial effort among different research groups to sequence the biosynthetic genes of epothilone due to its activity and potential commercial value (Julien et al., 2000; Molnar et al., 2000).

Polyketide Biosynthesis

Polyketide synthases (PKS) are biosynthetic enzymes that catalyze the stepwise assembly of polyketides (reviewed by Khosla et al., 1999). Type II PKSs are responsible for the biosynthesis of aromatic polyketides such as tetracycline and doxorubicin (reviewed by Hutchinson, 1997). Type II PKSs are similar to bacterial fatty acid synthases in which the active sites on monofunctional polypeptides are used in an iterative fashion to synthesize the final polyketide molecule. Modular polyketides synthases (type I PKSs) are large multifunctional enzymes that synthesize macrocyclic polyketides such as the antibiotics erythromycin and rifamycin (Donadio et al., 1991; Schupp et al., 1998; Tang et al., 1998). Type I PKS multienzyme complexes contain multiple separate active sites or catalytic domains that are used sequentially and non-iteratively during the stepwise reactions of the biosynthetic pathway (reviewed by Hopwood, 1997; Cane et al., 1998; Rawlings, 2001b;

Rawlings, 2001a). The catalytic domains that are collectively responsible for one round of the chain extension and β-keto processing in the biosynthesis are organized and referred to as modules. The catalytic domains present in each module determine substrate recognition for chain-extension and the extent of β-keto reduction. The building units of polyketides include acetyl-CoA, propionyl-CoA, butyryl-CoA and their dicarboxy counterparts, malonyl-CoA, methylmalonyl-CoA and ethylmalonyl-CoA. Each extension unit contributes two carbon atoms to the backbone chain of the polyketide product, in which the β-carbon is initially a keto group, hence the name polyketide. The key step in the assembly of the polyketide chain is the decarboxylative condensation (Figure 3) which is analogous to the elongation step in fatty acid synthesis. In fact, polyketide synthesis and fatty acid synthesis show similarities in genetics and mechanisms. The β-keto group can then be processed by the reductive domains ketoreductase, dehydratase, and enoylreductase, encoded in the reductive loops of PKS modules (Figure 4). The remarkable structural diversity seen in polyketides is due to several factors. The first is the diversity of side groups on the second carbon donated by each extender unit to the growing polyketide chain. These side groups include hydrogen, methyl, ethyl or other rarer groups. Second is the optional level of reduction that occurs at each β-keto group, as determined by the reductive loop domains in each module (Figure 4). If no reduction occurs, the keto group is retained in the polyketide. The keto can be reduced to a hydroxyl group, or further to an enoyl group or fully reduced to the alkyl depending on the reductive loop domains. A third way that variety is introduced into polyketide biosynthesis is the total length of the chain determined by the number of modules in the PKS. Other degrees of complexity also arise from the variation of starter and extender units as

well as the chirality of the carbon atoms incorporated in to the polyketide. Polyketides can also have various post-PKS modifications such as glycosylation and methylation.

Soraphen A biosynthesis

The isolation, structure and antifungal activity of soraphen A was first described in the literature in 1994 (Gerth et al., 1994). Subsequent publications further characterized the production of soraphen A in *S. cellulosum* So ce 26 as well as the mode of action and the sequence of the genes responsible for the synthesis of the PKS (Vahlensieck et al., 1994; Chapter 3).

Two properties of soraphen A that are relatively uncommon in the area of PKS biosynthesis is the incorporation of an aromatic starter unit and the utilization of methoxymalonyl-CoA as an extender unit. There is interest in unusual properties within polyketide biosynthesis by a number of research groups investigating the engineering of "unnatural" natural products (Tsoi and Khosla, 1995). The modularity of type I PKS systems has led to the idea of combining modules from different PKS biosynthetic gene clusters to create novel molecules. The practice of this combinatorial approach has yielded some impressive results in generating novel polyketide-based combinatorial libraries (Xue et al., 1999). The unusual features of the soraphen PKS become important in the light of combinatorial efforts, since these features could be utilized to increase the diversity of biosynthetic mechanisms used to produce these combinatorial libraries.

The starter unit for most type I polyketides are derived from acetate or propionate but there are also some examples of incorporation of aromatic starter units

into polyketides (Figure 5). The aromatic starter unit 3-amino-5-hydroxybenzoic acid (AHBA) is synthesized in the aminoshikimate pathway and the corresponding genes have been cloned in the rifamycin, mitomycin, ansatrienin and napthomycin PKS clusters (August et al., 1998; Chen et al., 1999; Mao et al., 1999; Yu et al., 2001). AHBA is not the starter unit in soraphen A due to the lack of an amino group on the aromatic moiety in soraphen A and the lack of aminoshikimate biosynthetic genes in the soraphen A biosynthetic cluster. In cases such as rifamycin and ansamitocin, a coenzyme A ligase is used to activate and attach the AHBA starter unit to the loading ACP domain (August et al., 1998; Schupp et al., 1998; Tang et al., 1998; Yu et al., 2002). A carboxylic acid:conenzyme A ligase is found in the loading domain in these examples which is also lacking in the soraphen PKS.

A further example for the incorporation of an aromatic precursor into a polyketide is seen in enterocin, a type II polyketide produced by "*Streptomyces maritimus"* (Piel et al., 2000). Benzoyl-CoA has been shown to serve as a starter unit for the biosynthesis of enterocin (Xiang and Moore, 2002; Xiang and Moore, 2003). In enterocin biosynthesis, the starter unit formation includes the activity of a unique bacterial phenylalanine ammonia-lyase and β-oxidation reactions to synthesize benzoyl-CoA from phenylalanine (Figure 6). Although the biosynthesis of the starter unit in the soraphen PKS may resemble the pathway seen in enterocin biosynthesis, a benzoate-CoA ligase or a dedicated pathway for benzoyl-CoA biosynthesis was not found with the soraphen A biosynthetic gene cluster. The soraphen PKS has an acyltransferase and acyl carrier protein loading domains, which have been shown to incorporate benzoyl-CoA in a model system (Wilkinson et al., 2001). When the loading domains from the soraphen PKS were fused to extension modules 1 and 2 and

thioesterase from the erythromycin PKS, a triketide product was recovered with benzoyl-CoA derived started unit incorporated into the final molecule (Figure 7).

The second unusual feature of soraphen A biosynthesis is the utilization of methoxymalonate as an extender unit. Methoxymalonate is predicted to be the extender unit for both module 3 and 7 of the soraphen PKS (Figure 8b and Chapter 3). Chain extension in PKS biosynthesis occurs most often by the addition of malonyl-CoA or methylmalonyl-CoA. In some cases, labelling experiments and positioning of hydroxyl or methoxyl groups in the polyketide, suggests the utilization of an extender unit derived from glucose or glycerol (Haber et al., 1977; Omura et al., 1983; Hill et al., 1998; Ono et al., 1998). Such "glycolate" units are seen in ansamitocin, geldanamycin (Haber et al., 1977), leucomycin (Omura et al., 1983), FK520 and FK506 (Byrne et al., 1993), and soraphen A (Hill et al., 1998). The genes and biosynthetic pathway proposed for the incorporation of methoxymalonyl-CoA in FK520 are presented in Figure 8a (Wu et al., 2000). A group of homologous genes proposed to be involved in the biosynthesis of the "glycolate" extender in ansamitocin, *asm*13-17, were disrupted and the resulting strains showed no ansamitocin production (Yu et al., 2002). Recently, *asm*13-17 were co-expressed in *Streptomyces lividans* with a cassette of *eryABC* encoding the erythromycin PKS, in which one of the AT domains had been replaced with the methoxymalonate AT from the FK520 PKS (Kato et al., 2002). The engineered *St. lividans* strain produced a 6-deoxyerythronolide derivative with a methoxymalonate incorporated in the predicted area of the molecule (Figure 9).

The incorporation of the aromatic starter unit and methoxymalonate extender units into soraphen A are unusual features that merit further investigation. But due to

the limited tools available for experimentation with *S. cellulosum*, these studies have been difficult or in some cases impossible. A system in which genetic manipulation allows for gene replacements and other molecular procedures would be required for the detailed studies of these and other features of soraphen A biosynthesis.

Heterologous expression of polyketides

Heterologous expression of polyketides is often proposed to facilitate the cloning and characterization of biosynthetic gene clusters, as well as the production of polyketides on an industrial scale. Some examples of the heterologous production of polyketides are presented in Table 3. *Streptomyces lividans* and *Streptomyces coelicolor* are the two strains most commonly used for the heterologous expression of polyketides. Both strains have well developed genetic systems including well characterized promoters, plasmids and integration vectors (Kieser et al., 2000). Both strains have endogenous PKSs and are well suited for the posttranslational modifications and production of substrates required by the heterologously expressed PKSs (Kieser et al., 2000; Pfeifer and Khosla, 2001).

There are several reasons that the heterologous expression of polyketides is sought. First, isolated microorganisms that produce polyketides are often virtually impossible to ferment or are recalcitrant to laboratory manipulations. Examples of these type of organisms in which polyketides are isolated include: isolates of myxobacteria and dinoflagellates, and marine sponges in which the production of the secondary metabolite is probably due to a microbial symbiont. Difficulty in manipulation of the natural isolate that produces the polyketide hinders the characterization of the biological activity of the polyketide as well as the isolation of

the genes responsible for the synthesis of the PKS. The successful heterologous expression in a manageable host affords the ability to produce sufficient quantities of polyketide to investigate the biological activity as well as the molecular mechanisms of biosynthesis.

The production of polyketides in *Escherichia coli* or *Saccharomyces cerevisiae* is very desirable due to the advanced knowledge and technology in these organisms. Since combinatorial approaches require high-throughput molecular manipulation and small-scale fermentations, a heterologous host that can be engineered to be more amenable to these conditions is indispensable. Recently the PKS responsible for the synthesis of the macrocyclic core of the antibiotic erythromycin, 6-deoxyerythronolide B (6dEB), was expressed in *E. coli* (Pfeifer et al., 2001). A number of technical challenges were overcome to produce 6dEB in *E. coli*. First, the 6dEB megasynthase needed to be properly folded and posttranslantionally modified in *E. coli*, including the pantetheinylation of the active site acyl carrier proteins. Second, it was necessary to engineer the biosynthesis of the extender unit, (2S)-methylmalonyl-CoA, since this has not been observed as an *E. coli* metabolite (Haller et al., 2000). The production of 6-dEB required up to 25 genetic modification of the host *E. coli* with most alterations requiring proprietary genes or knowledge (Pfeifer and Khosla, 2001).

There is one previous example of the heterologous production of a *S. cellulosum* polyketide. The biosynthetic cluster from *S. cellulosum* So ce90 responsible for the production of the polyketide/peptide hybrid metabolite epothilone has been expressed in both *S. coelicolor* and *M. xanthus* (Tang et al., 2000; Arslanian et al., 2002). The potential value of epothilone and its relatively low production and

poor genetic systems in *S. cellulosum* So ce90 were the driving forces for these experiments. The expression of epothilone in *M. xanthus* was the first report of a polyketide heterologously produced in *M. xanthus* and the yields reported were extremely low (Julien and Shah, 2002).

The heterologous production of polyketides is a valuable tool in the progression of our understanding of polyketide synthesis and production of novel molecules. Areas such as the biochemistry of precursor supply and metabolic flux analysis will contribute to the engineered production of polyketides and secondary metabolites in heterologous hosts.

ACKNOWLEDGMENTS

I would like to thank István Molnár for critical reading of this manuscript and helpful discussions.

Table 1. Taxonomy of the myxobacteria

Reichenbach	**Bergy's Manual of Systematic Bacteriology**
Order: ***Myxococcales***	Order: ***Myxococcales***
Suborder: ***Cystobacterineae***	Family: ***Myxococcaceae***
Family: ***Myxococcacea***	Genus: ***Myxcoccus***
Genera: ***Myxococcus***	Family: ***Archangiaceae***
Corallococcus	Genus: ***Archangium***
Angiococcus	Family: ***Cystobacteraceae***
Family: ***Cystobacteraceae***	Genera: ***Cystobacter***
Genera***: Archangium***	***Melittangium***
Cystobacter	***Stigmatella***
Melittangium	Family: ***Polyangiaceae***
Stigmatella	Genera: ***Polyangium***
Suborder: ***Sorangineae***	***Nannocystis***
Family: ***Polyangiaceae***	***Chondromyces***
Genera: ***Polyangium***	
Haploangium	
Chondromyces	
Sorangium	
Family: ***Nannocystanceae***	
Genus: ***Nannocystis***	

Table 1 references:

McCudry, H.: Fruiting Gliding Bacteria. In: Holt, J. (Ed.), Bergey's Manual of Systematic Bacteriology. Williams & Wilkins, Baltimore, 1989, pp. 2139-2143.

Reichenbach, H. and Dworkin, M.: The order Myxobacterales. In: Starr, M. (Ed.), Prokaryotes; A handbook on habitats, isolation, and identification of bacteria. Springer-Verlag, Berlin, 1981, pp. 315-327.

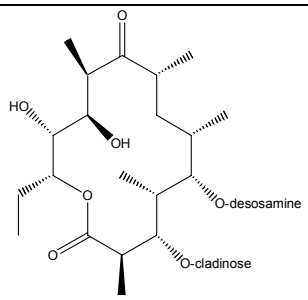
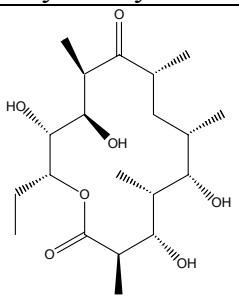
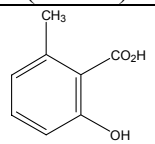
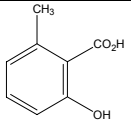
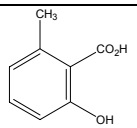
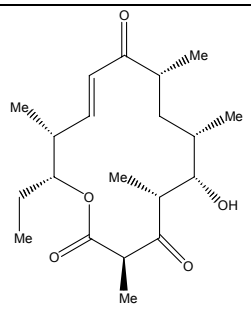
Reichenbach, H. and Hoefle, G.: Biologically active secondary metabolites from myxobacteria. Biotechnology Advances 11 (1993) 219-77.

Table 2. Compounds from myxobacteria and mode of action.

Compound	Mode of action
Myxovalargin	Protein synthesis inhibitor (prokaryotic)
Gephyronic acid	Protein synthesis inhibitor (eukaryotic)
Saframycin	Binds to DNA
Vioprolid	Cytotoxic
Leupyrin	Antifungal
Myxothiazol	Electron transport inhibitor
Stigmetellin	Electron transport inhibitor
Sorangicin	Eubacterial RNA polymerase inhibitor
Myxopyronin	RNA polymerase inhibitor
Ripostatin	RNA polymerase inhibitor
Chondramide A	Binds to actin
Epothilone	Stabilizes microtubules
Soraphen A	Fungal acetly-CoA carboxylase inhibitor

Reference for table 2:
Reichenbach, H.: Myxobacteria, producers of novel bioactive substances. J Ind Microbiol Biotechnol 27 (2001) 149-56.

Table 3. Examples of heterologous expression of polyketides

Polyketide	Native host	Herterologous host	Year
 erythromycin A	*Saccharopolyspora erythraea*	*Streptomyces lividans*	1986[a]
 6-deoxyerythronolide (6-dEB)	*S. erythraea*	*Streptomyces coelicolor*	1994[b]
 6-methylsalicylic acid (6-MSA)	*Penicillium patulum*	*S. coelicolor*	1995[c]
 6-MSA	*P. patulum*	*Escherichia coli*	1998[d]
 6-MSA	*P. patulum*	*Saccharomyces cerevisiae*	1998[e]
 narbonolide	*Streptomyces venezuelae*	*St. lividans*	1999[f]

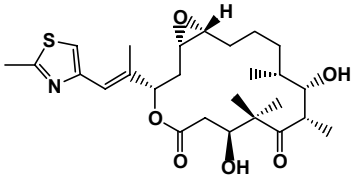
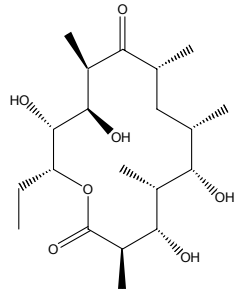
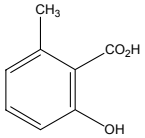
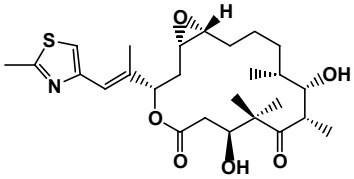
Polyketide	Native host	Herterologous host	Year
 epothilone	*Sorangium cellulosum* So ce90	*S. coelicolor*	2000[g]
 6-dEB	S. *erythrea*	*E. coli*	2001[h]
 6-MSA	*P. patulum*	*Nicotiana tabacum*	2001[i]
 epothilone	*S. cellulosum* So ce90	*Myxococcus xanthus*	2002[j]

Table 3 references:

[a]Stanzak, P., Matsushima, P., Baltz, R.H. and Rao, R.N.: Cloning and expression in *Streptomyces lividans* of clustered erythromycin biosynthesis genes from *Streptomyces erythreus*. Bio/Technology 4 (1986) 229-232.

[b]Kao, C.M., Katz, L. and Khosla, C.: Engineered biosynthesis of a complete macrolactone in a heterologous host. Science 265 (1994) 509-12.

[c]Bedford, D.J., Schweizer, E., Hopwood, D.A. and Khosla, C.: Expression of a functional fungal polyketide synthase in the bacterium *Streptomyces coelicolor* A3(2). J Bacteriol 177 (1995) 4544-8.

[d,e]Kealey, J.T., Liu, L., Santi, D.V., Betlach, M.C. and Barr, P.J.: Production of a polyketide natural product in nonpolyketide-producing prokaryotic and eukaryotic hosts. Proc Natl Acad Sci U S A 95 (1998) 505-9.

[f]Tang, L., Fu, H., Betlach, M.C. and McDaniel, R.: Elucidating the mechanism of chain termination switching in the picromycin/methymycin polyketide synthase. Chem Biol 6 (1999) 553-8.

[g]Tang, L., Shah, S., Chung, L., Carney, J., Katz, L., Khosla, C. and Julien, B.: Cloning and heterologous expression of the epothilone gene cluster. Science 287 (2000) 640-2.

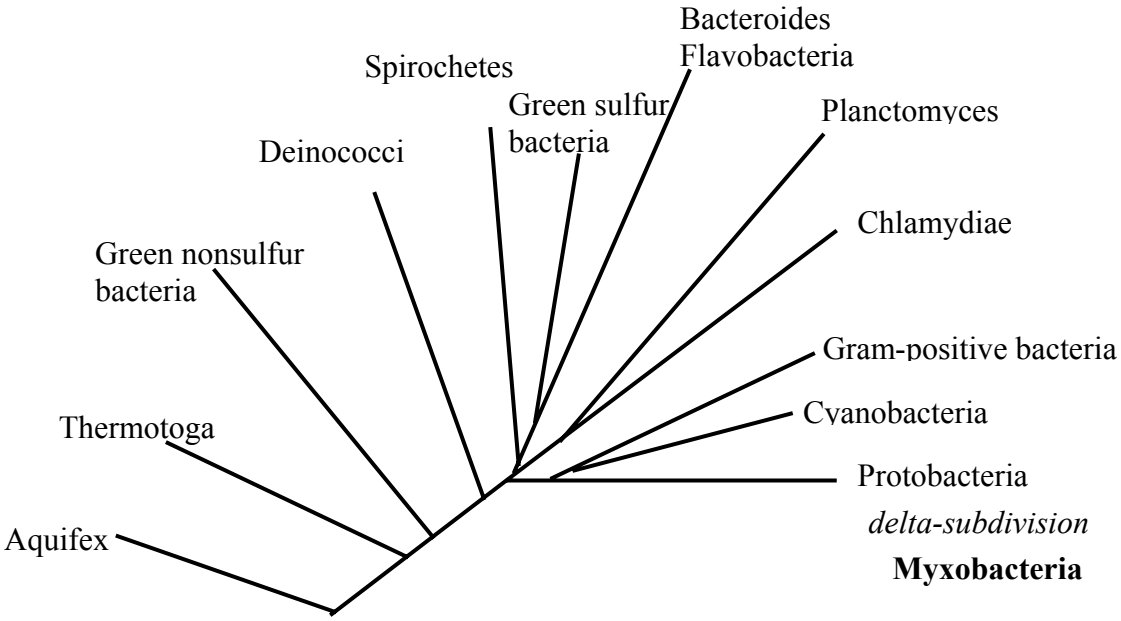
Table 3 references continued:

[h]Pfeifer, B.A., Admiraal, S.J., Gramajo, H., Cane, D.E. and Khosla, C.: Biosynthesis of complex polyketides in a metabolically engineered strain of *E. coli*. Science 291 (2001) 1790-2.

[i]Yalpani, N., Altier, D.J., Barbour, E., Cigan, A.L. and Scelonge, C.J.: Production of 6-methylsalicylic acid by expression of a fungal polyketide synthase activates disease resistance in tobacco. Plant Cell 13 (2001) 1401-9.

[j]Arslanian, R.L., Parker, C.D., Wang, P.K., McIntire, J.R., Lau, J., Starks, C. and Licari, P.J.: Large-scale isolation and crystallization of epothilone D from *Myxococcus xanthus* cultures. J Nat Prod 65 (2002) 570-2.

A.



B.

alpha	beta	gamma	delta	epsilon
Caulobacter	*Neisseria*	*Escherichia*	**Myxobacteria**	*Helicobacter*
Agrobacterium	*Burkholderia*	*Salmonella*	*Desulfovibro*	*Camplylobacter*
Rickettsia	*Bordetella*	*Pseudomonas*		

Figure 1. Phylogeny of myxobacteria. (A); Phylogenetic tree based on 16S rRNA of the eubacteria. The myxobacteria are in the delta division of the proteobacteria. Figure redrawn from Spormann. (B); Examples of the subdivisions of the proteobacteria

Figure 1 reference:

(A) Spormann, A.M.: Gliding motility in bacteria: insights from studies of *Myxococcus xanthus*. Microbiol Mol Biol Rev 63 (1999) 621-41.

(B) Information according to National Center for Biotechnology Information. http://www.ncbi.nlm.nih.gov/Taxonomy/taxonomyhome.html/

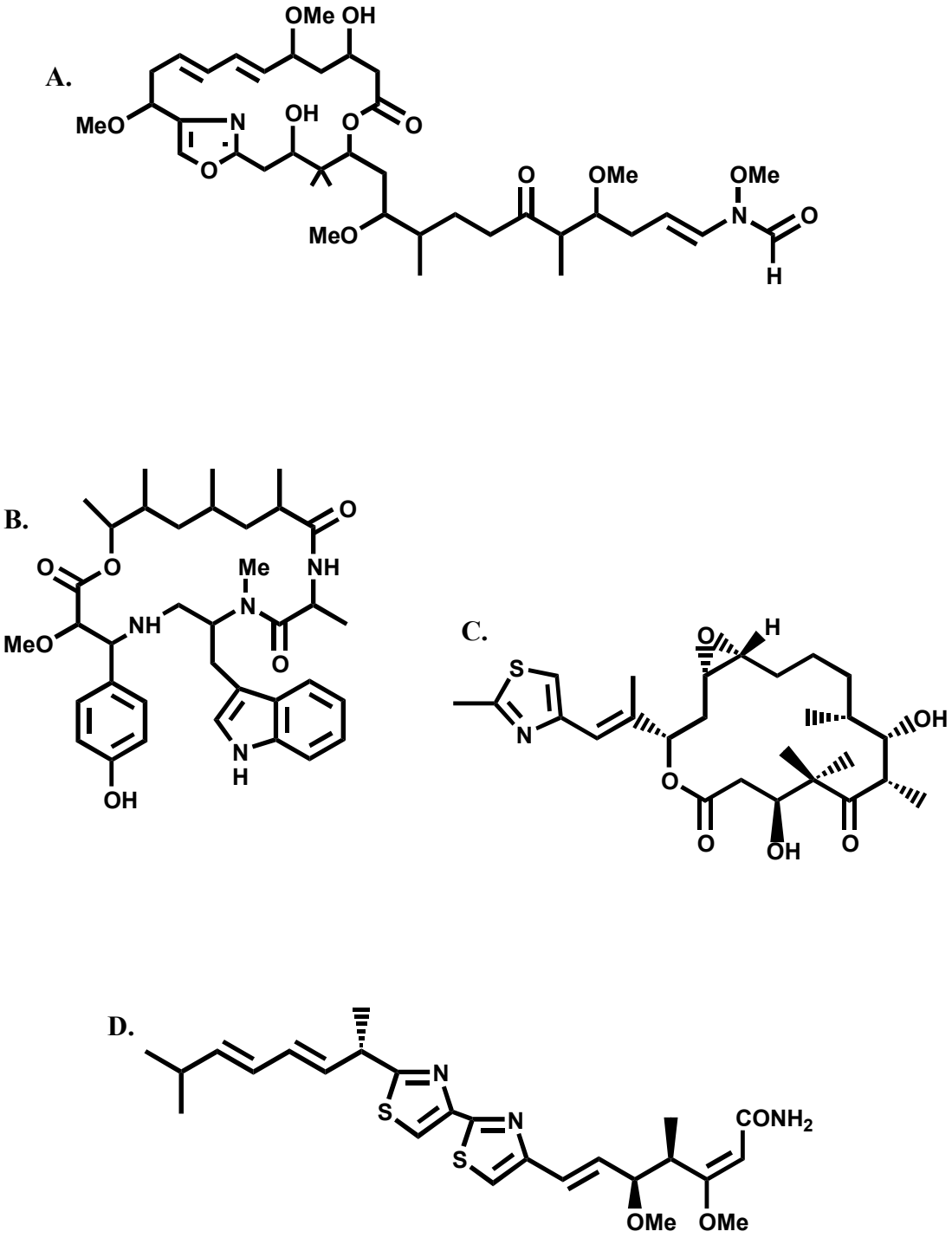


Figure 2. Secondary metabolites from myxobacteria. (A); Rhizopodin. (B); Chondramid A. (C); Epothilon A. (D); Myxothiazol

Figure 2 reference:

Reichenbach, H.: Myxobacteria, producers of novel bioactive substances. J Ind Microbiol Biotechnol 27 (2001) 149-56.

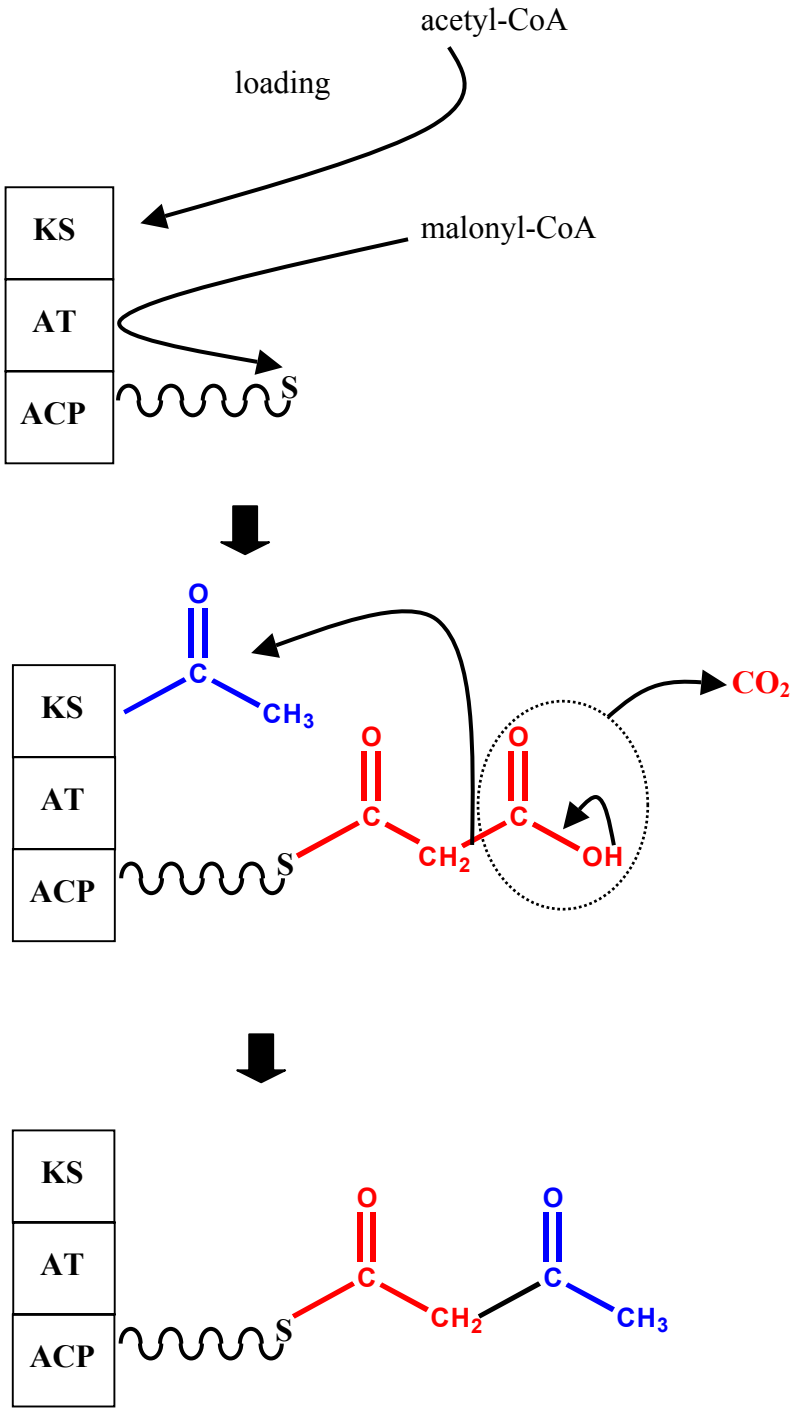


Figure 3. The decarboxylative condensation reaction for chain elongation in PKS. In the example above, acetyl-CoA is transferred to the ketosynthase (KS) domain by the loading domain. A malonyl-CoA is transferred to the pantetheinyl arm of the acyl carrier protein (ACP) by the acyl transferase (AT). The elongation reaction is accompanied by decarboxylation. The β-keto group (blue) can now undergo reduction.

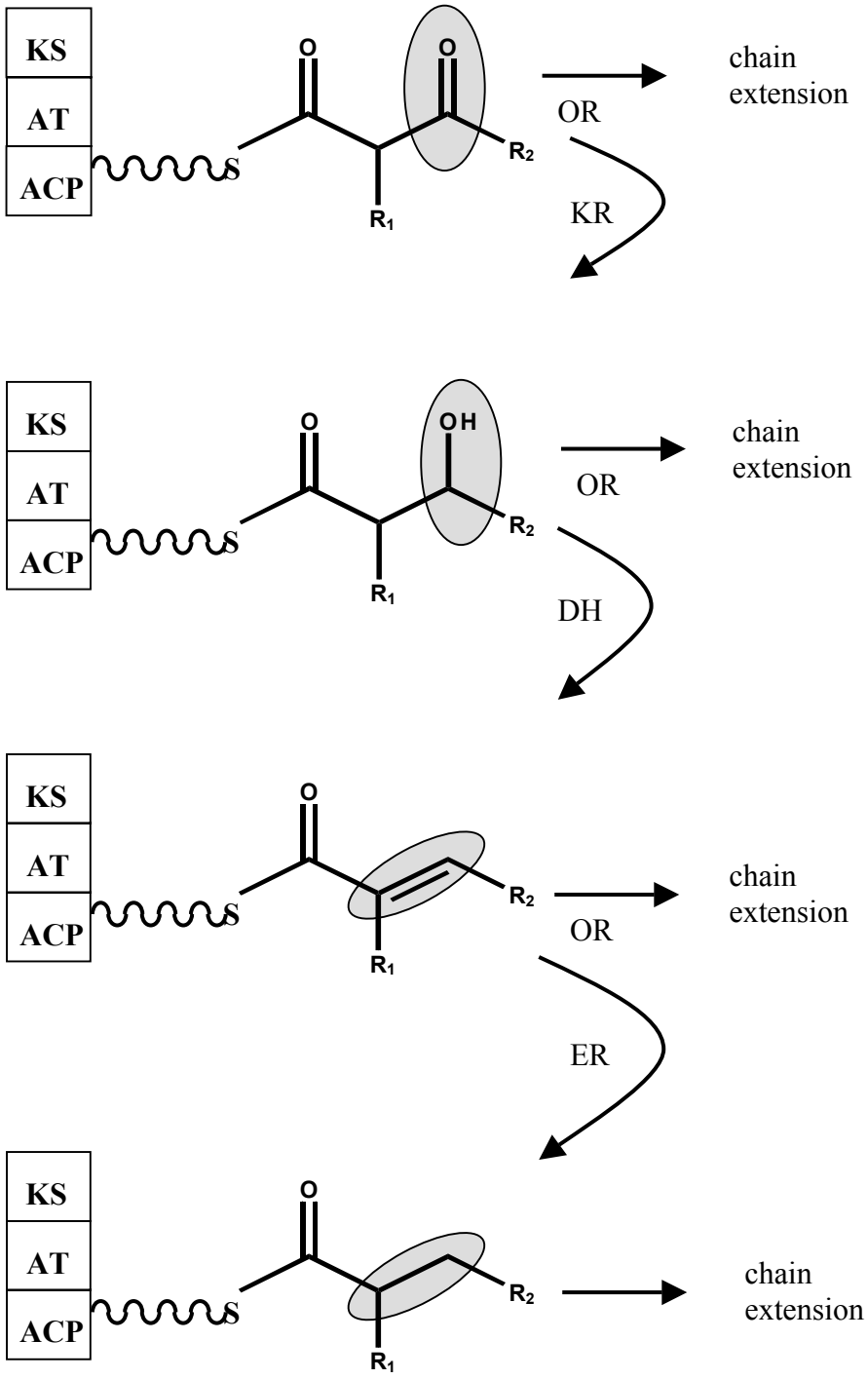


Figure 4. The optional step-wise reduction of the β-keto group. After the condensation reaction in Figure 3, the β-keto group can undergo reduction, depending on the presence of reductive domains. If no reduction occurs, a keto group is incorporated into the polyketide chain. Ketoreductase (KR), dehyratase (DH), or enyol reductase (ER) reduce the keto group (in shaded oval).

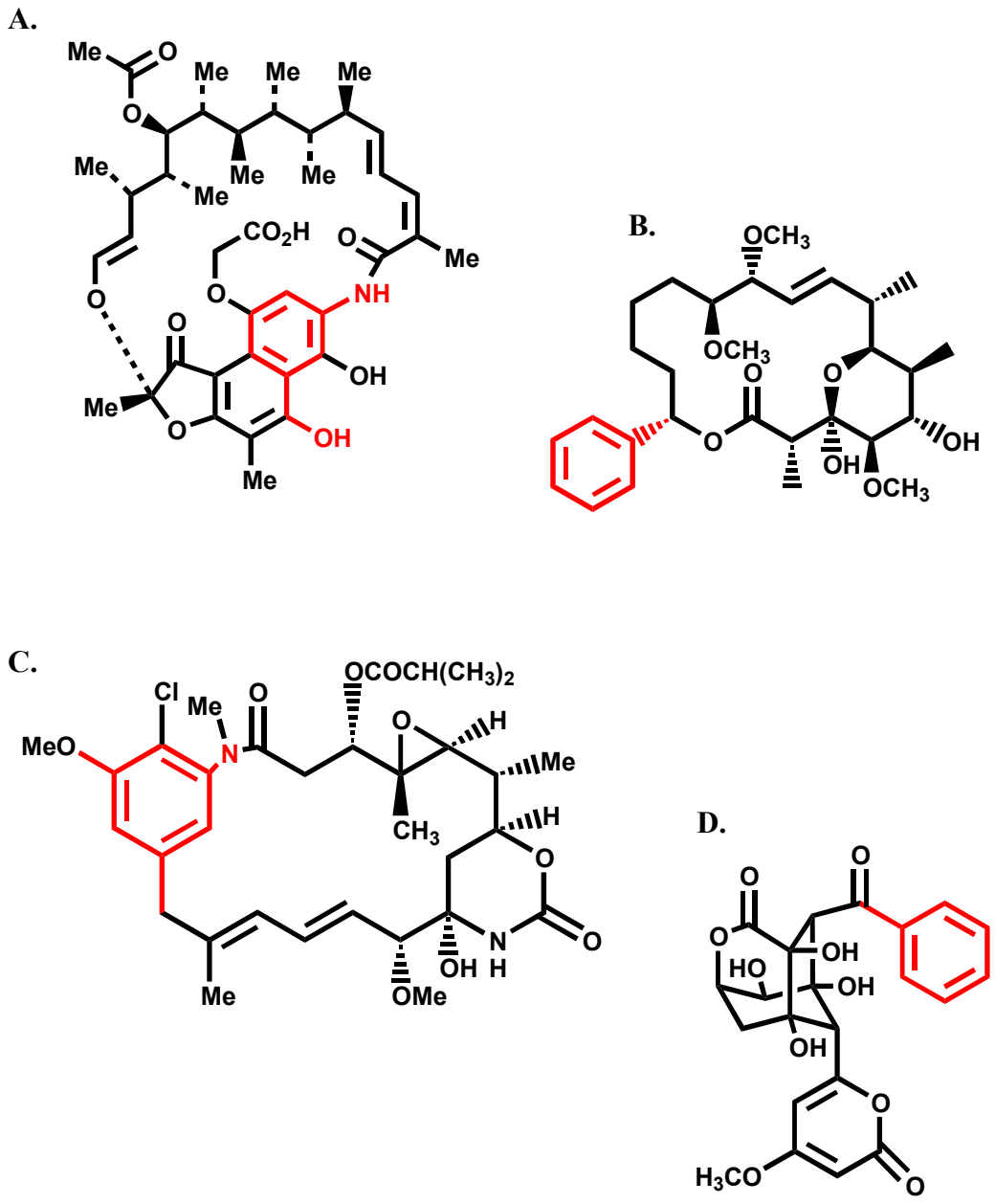


Figure 5. Aromatic starter units incorporated into polyketides. The aromatic starter units in polyketides are shown in red above. (A); Rifamycin B, (B); Soraphen A, (C); Ansamitocin P-3, (D) Enterocin.

Figure 5 references:

Piel, J., Hertweck, C., Shipley, P.R., Hunt, D.M., Newman, M.S. and Moore, B.S.: Cloning, sequencing and analysis of the enterocin biosynthesis gene cluster from the marine isolate '*Streptomyces maritimus*': evidence for the derailment of an aromatic polyketide synthase. Chem Biol 7 (2000) 943-55.

Rawlings, B.J.: Type I polyketide biosynthesis in bacteria (part B). Nat Prod Rep 18 (2001b) 231-81.

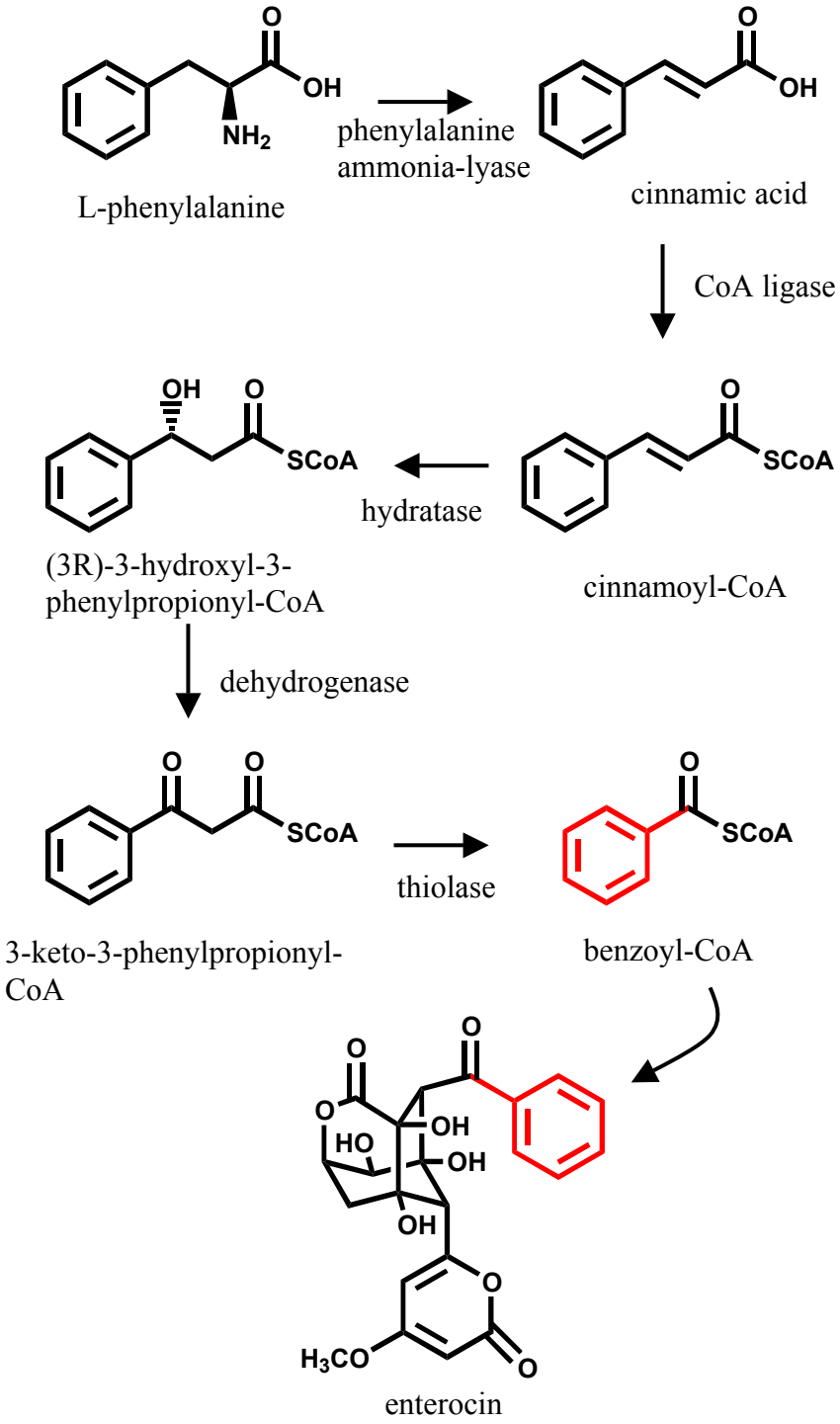


Figure 6. Proposed pathway for the biosynthesis of benzoyl-CoA derived starter unit in enterocin biosynthesis. Adapted from Xiang and Moore.

Figure 6 reference:

Xiang, L. and Moore, B.S.: Characterization of benzoyl coenzyme A biosynthesis genes in the enterocin-producing bacterium "*Streptomyces maritimus*". J Bacteriol 185 (2003) 399-404.

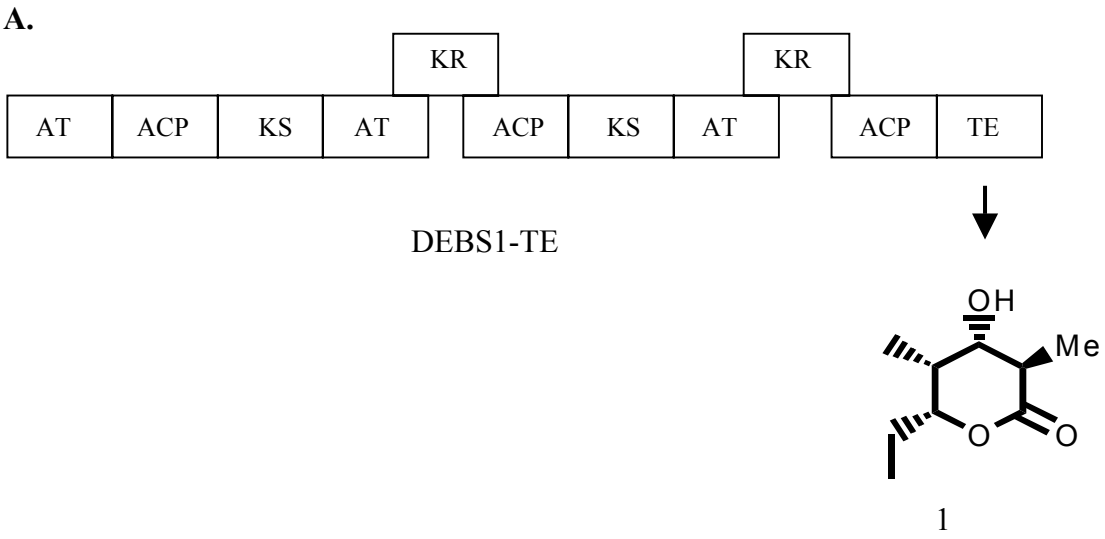
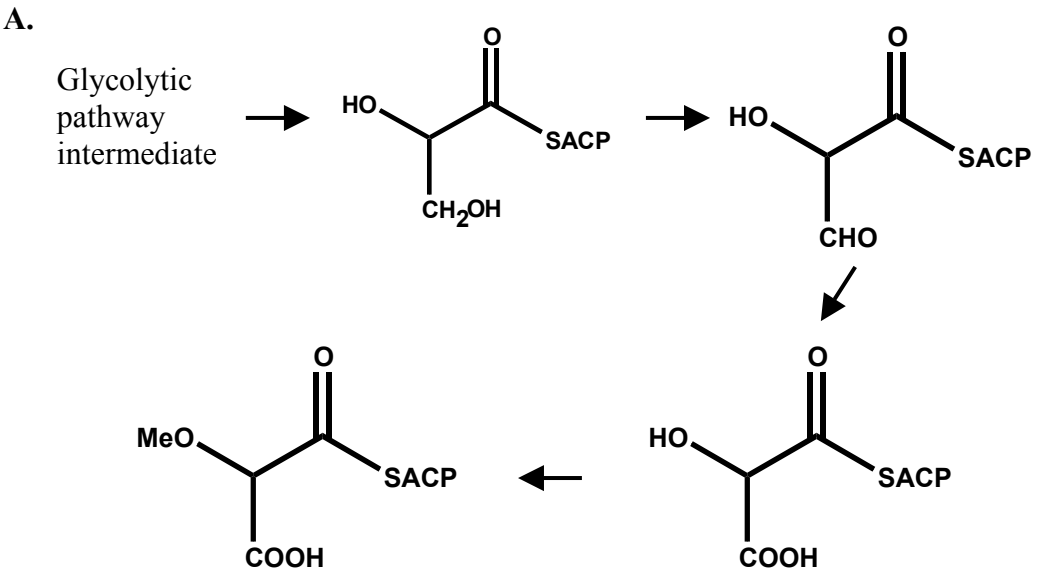


Figure 7. Soraphen loading domains. (A); A bimodular PKS based on the DEBS PKS expressed in *Streptomyces lividans* produces the triketide product 1. (B); When the loading domains from the soraphen PKS (in red) are fused to the DEBS extension domains and expressed in *St. lividans*, product 2 is produced with the benzoyl-CoA incorporated (red). Adapted from Wilkinson et al.

Figure 7 reference:

Wilkinson, C.J., Frost, E.J., Staunton, J. and Leadlay, P.F.: Chain initiation on the soraphen-producing modular polyketide synthase from *Sorangium cellulosum*. Chem Biol 8 (2001) 1197-208.

A.



B.

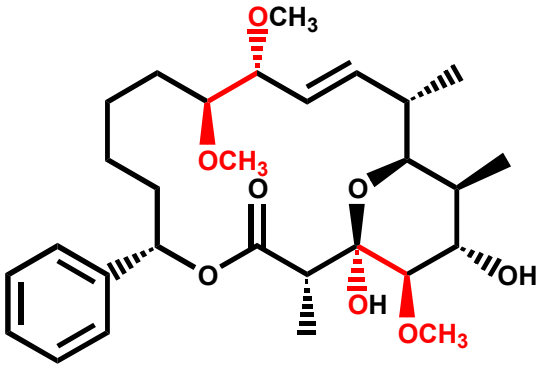
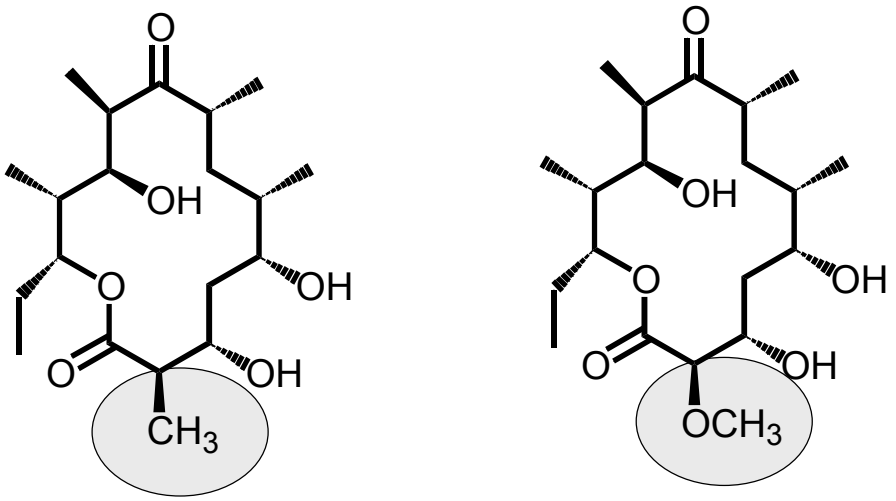


Figure 8. Methyoxymalonate as an extender unit in PKS biosynthesis. (A); The proposed pathway for methoxmalonyl biosynthesis for incorporation into FK520. Modified from Wu et al. (B); The proposed incorporation of methoxymalonyl into soraphen A. The methoxymalonyl-derived units are shown in red.

Figure 8A reference:

Wu, K., Chung, L., Revill, W.P., Katz, L. and Reeves, C.D.: The FK520 gene cluster of *Streptomyces hygroscopicus* var. ascomyceticus (ATCC 14891) contains genes for biosynthesis of unusual polyketide extender units. Gene 251 (2000) 81-90.



S. *lividans* + 6dEB plasmid

S. *livandans* + 6dEB plasmid with AT6:*fkb*A-AT8 + *asm*13-17

Figure 9. Incorporation of methoxymalonate into 6-deoxyerthronolide B (6-dEB). When a plasmid containing the 6-dEB PKS from *Saccharopolyspora erythraea* is expressed in *Streptomyces lividans*, 6-dEB is produced (molecule on the left). When the acyltransferase (AT) from module 8 from the FK520 is used to replace AT6 in 6-dEB and co-expressed with the genes *asm*13-17, a methoxymalonate is incorporated into the 6-dEB. *Asm*13-17 are the genes responsible for the biosynthesis of methoxymalonyl-CoA in the ansamitocin PKS cluster.

Figure 9 reference:

Kato, Y., Bai, L., Xue, Q., Revill, W.P., Yu, T.W. and Floss, H.G.: Functional expression of genes involved in the biosynthesis of the novel polyketide chain extension unit, methoxymalonyl-acyl carrier protein, and engineered biosynthesis of 2-desmethyl-2-methoxy-6-deoxyerythronolide B. J Am Chem Soc 124 (2002) 5268-9.

REFERENCES:

Arslanian, R.L., Tang, L., Blough, S., Ma, W., Qiu, R.G., Katz, L. and Carney, J.R.: A new cytotoxic epothilone from modified polyketide synthases heterologously expressed in *Myxococcus xanthus*. J Nat Prod 65 (2002) 1061-4.

August, P.R., Tang, L., Yoon, Y.J., Ning, S., Muller, R., Yu, T.W., Taylor, M., Hoffmann, D., Kim, C.G., Zhang, X., Hutchinson, C.R. and Floss, H.G.: Biosynthesis of the ansamycin antibiotic rifamycin: deductions from the molecular analysis of the rif biosynthetic gene cluster of *Amycolatopsis mediterranei* S699. Chem Biol 5 (1998) 69-79.

Beatson, S.A., Whitchurch, C.B., Sargent, J.L., Levesque, R.C. and Mattick, J.S.: Differential regulation of twitching motility and elastase production by Vfr in *Pseudomonas aeruginosa*. J Bacteriol 184 (2002) 3605-13.

Byrne, K.M., Shafiee, A., Nielsen, J.B., Arison, B., Monaghan, R.L. and Kaplan, L.: The Biosynthesis and Enzymology of an Immunosuppressant, Immunomycin, Produced by *Streptomyces hygroscopicus* var. *ascomyceticus*. In: R., C. (Ed.), Developments in Industrial Microbiology: Microbial Metabolites. McGraw-Hill Companies, Columbus, OH, 1993, pp. 29-45.

Cane, D.E., Walsh, C.T. and Khosla, C.: Harnessing the biosynthetic code: combinations, permutations, and mutations. Science 282 (1998) 63-8.

Chen, S., von Bamberg, D., Hale, V., Breuer, M., Hardt, B., Muller, R., Floss, H.G., Reynolds, K.A. and Leistner, E.: Biosynthesis of ansatrienin (mycotrienin) and naphthomycin. Identification and analysis of two separate biosynthetic gene clusters in *Streptomyces collinus* Tu 1892. Eur J Biochem 261 (1999) 98-107.

Comolli, J.C., Hauser, A.R., Waite, L., Whitchurch, C.B., Mattick, J.S. and Engel, J.N.: *Pseudomonas aeruginosa* gene products PilT and PilU are required for cytotoxicity in vitro and virulence in a mouse model of acute pneumonia. Infect Immun 67 (1999) 3625-30.

Donadio, S., Staver, M.J., McAlpine, J.B., Swanson, S.J. and Katz, L.: Modular organization of genes required for complex polyketide biosynthesis. Science 252 (1991) 675-9.

Dworkin, M.: Recent advances in the social and developmental biology of the myxobacteria. Microbiol Rev 60 (1996) 70-102.

Gerth, K., Bedorf, N., Irschik, H., Hofle, G. and Reichenbach, H.: The soraphens: a family of novel antifungal compounds from *Sorangium cellulosum* (Myxobacteria). I. Soraphen A1 alpha: fermentation, isolation, biological properties. J Antibiot (Tokyo) 47 (1994) 23-31.

Gerth, K., Bedorf, N., Hofle, G., Irschik, H. and Reichenbach, H.: Epothilons A and B: antifungal and cytotoxic compounds from *Sorangium cellulosum* (Myxobacteria). Production, physico-chemical and biological properties. J Antibiot (Tokyo) 49 (1996) 560-3.

Haber, A., Johnson, R.D. and Rinehart, K.L., Jr.: Biosynthetic origin of the C2 units of geldanamycin and distribution of label from D-[6-13C]glucose. J Am Chem Soc 99 (1977) 3541-4.

Haller, T., Buckel, T., Retey, J. and Gerlt, J.A.: Discovering new enzymes and metabolic pathways: conversion of succinate to propionate by *Escherichia coli*. Biochemistry 39 (2000) 4622-9.

Hartzell, P. and Kaiser, D.: Function of MglA, a 22-kilodalton protein essential for gliding in *Myxococcus xanthus*. J Bacteriol 173 (1991) 7615-24.

Hartzell, P.L.: Complementation of sporulation and motility defects in a prokaryote by a eukaryotic GTPase. Proc Natl Acad Sci U S A 94 (1997) 9881-6.

Hill, A.M., Harris, J.P. and Siskos, A.P.: Investigation of glycerol incorporation into soraphen A. Chemical Communications (Cambridge) (1998) 2361-2362.

Hodgkin, J. and Kaiser, D.: Genetics of gliding motility in *Myxococcus xanthus* (Myxobacterales): two gene systems control movement. Mol Gen Genet 171 (1979a) 177-191.

Hodgkin, J. and Kaiser, D.: Genetics of gliding motility in *Myxococcus xanthus* (Myxobacterales): genes controling movement of single cells. Mol Gen Genet 171 (1979b) 167-171.

Hopwood, D.A.: Genetic Contributions to Understanding Polyketide Synthases. Chem Rev 97 (1997) 2465-2498.

Hutchinson, C.R.: Biosynthetic Studies of Daunorubicin and Tetracenomycin C. Chem Rev 97 (1997) 2525-2536.

Jaoua, S., Neff, S. and Schupp, T.: Transfer of mobilizable plasmids to *Sorangium cellulosum* and evidence for their integration into the chromosome. Plasmid 28 (1992) 157-65.

Julien, B. and Shah, S.: Heterologous Expression of Epothilone Biosynthetic Genes in *Myxococcus xanthus*. Antimicrob Agents Chemother 46 (2002) 2772-8.

Julien, B., Shah, S., Ziermann, R., Goldman, R., Katz, L. and Khosla, C.: Isolation and characterization of the epothilone biosynthetic gene cluster from *Sorangium cellulosum*. Gene 249 (2000) 153-60.

Kato, Y., Bai, L., Xue, Q., Revill, W.P., Yu, T.W. and Floss, H.G.: Functional expression of genes involved in the biosynthesis of the novel polyketide chain extension unit, methoxymalonyl-acyl carrier protein, and engineered biosynthesis of 2-desmethyl-2-methoxy-6-deoxyerythronolide B. J Am Chem Soc 124 (2002) 5268-9.

Khosla, C., Gokhale, R.S., Jacobsen, J.R. and Cane, D.E.: Tolerance and specificity of polyketide synthases. Annu Rev Biochem 68 (1999) 219-53.

Kieser, T., Bibb, M.J., Buttner, M.J., Chater, K.F. and Hopwood, D.A.: Practical *Streptomyces* Genetics. The John Innes Foundation, Norwich, England, 2000.

MacNeil, S.D., Calara, F. and Hartzell, P.L.: New clusters of genes required for gliding motility in *Myxococcus xanthus*. Mol Microbiol 14 (1994) 61-71.

Mao, Y., Varoglu, M. and Sherman, D.H.: Molecular characterization and analysis of the biosynthetic gene cluster for the antitumor antibiotic mitomycin C from *Streptomyces lavendulae* NRRL 2564. Chem Biol 6 (1999) 251-63.

McBride, M.J.: Bacterial gliding motility: multiple mechanisms for cell movement over surfaces. Annu Rev Microbiol 55 (2001) 49-75.

McCudry, H.: Fruiting Gliding Bacteria. In: Holt, J. (Ed.), Bergey's Manual of Systematic Bacteriology. Williams & Wilkins, Baltimore, 1989, pp. 2139-2143.

Merz, A.J., So, M. and Sheetz, M.P.: Pilus retraction powers bacterial twitching motility. Nature 407 (2000) 98-102.

Molnar, I., Schupp, T., Ono, M., Zirkle, R., Milnamow, M., Nowak-Thompson, B., Engel, N., Toupet, C., Stratmann, A., Cyr, D.D., Gorlach, J., Mayo, J.M., Hu, A., Goff, S., Schmid, J. and Ligon, J.M.: The biosynthetic gene cluster for the microtubule-stabilizing agents epothilones A and B from *Sorangium cellulosum* So ce90. Chem Biol 7 (2000) 97-109.

Omura, S., Tsuzuki, K., Nakagawa, A. and Lukacs, G.: Biosynthetic origin of carbons 3 and 4 of leucomycin aglycone. J Antibiot (Tokyo) 36 (1983) 611-3.

Ono, M., Sakuda, S., Ikeda, H., Furihata, K., Nakayama, J., Suzuki, A. and Isogai, A.: Structures and biosynthesis of aflastatins: novel inhibitors of aflatoxin production by *Aspergillus parasiticus*. J Antibiot (Tokyo) 51 (1998) 1019-28.

Pfeifer, B.A., Admiraal, S.J., Gramajo, H., Cane, D.E. and Khosla, C.: Biosynthesis of complex polyketides in a metabolically engineered strain of *E. coli*. Science 291 (2001) 1790-2.

Pfeifer, B.A. and Khosla, C.: Biosynthesis of polyketides in heterologous hosts. Microbiol Mol Biol Rev 65 (2001) 106-18.

Piel, J., Hertweck, C., Shipley, P.R., Hunt, D.M., Newman, M.S. and Moore, B.S.: Cloning, sequencing and analysis of the enterocin biosynthesis gene cluster from the marine isolate '*Streptomyces maritimus*': evidence for the derailment of an aromatic polyketide synthase. Chem Biol 7 (2000) 943-55.

Pradella, S., Hans, A., Sproer, C., Reichenbach, H., Gerth, K. and Beyer, S.: Characterisation, genome size and genetic manipulation of the myxobacterium *Sorangium cellulosum* So ce56. Arch Microbiol 178 (2002) 484-92.

Rawlings, B.J.: Type I polyketide biosynthesis in bacteria (Part A--erythromycin biosynthesis). Nat Prod Rep 18 (2001a) 190-227.

Rawlings, B.J.: Type I polyketide biosynthesis in bacteria (part B). Nat Prod Rep 18 (2001b) 231-81.

Reichenbach, H. and Dworkin, M.: The order *Myxobacterales*. In: Starr, M. (Ed.), Prokaryotes; A handbook on habitats, isolation, and identification of bacteria. Springer-Verlag, Berlin, 1981, pp. 315-327.

Reichenbach, H.: The ecology of the myxobacteria. Environ Microbiol 1 (1999) 15-21.

Reichenbach, H.: Myxobacteria, producers of novel bioactive substances. J Ind Microbiol Biotechnol 27 (2001) 149-56.

Reichenbach, H. and Hoefle, G.: Biologically active secondary metabolites from myxobacteria. Biotechnology Advances 11 (1993) 219-77.

Reichenbach, H. and Hofle, G.: Production of Bioactive Secondary Metabolites. In: Dworkin, M. and Kaiser, A.D. (Eds.), Myxobacteria II. American Society for Microbiology, Washington D.C., 1993, pp. 347-397.

Schupp, T., Toupet, C., Cluzel, B., Neff, S., Hill, S., Beck, J.J. and Ligon, J.M.: A *Sorangium cellulosum* (myxobacterium) gene cluster for the biosynthesis of the macrolide antibiotic soraphen A: cloning, characterization, and homology to polyketide synthase genes from actinomycetes. J Bacteriol 177 (1995) 3673-9.

Schupp, T., Toupet, C., Engel, N. and Goff, S.: Cloning and sequence analysis of the putative rifamycin polyketide synthase gene cluster from *Amycolatopsis mediterranei*. FEMS Microbiol Lett 159 (1998) 201-7.

Spormann, A.M.: Gliding motility in bacteria: insights from studies of *Myxococcus xanthus*. Microbiol Mol Biol Rev 63 (1999) 621-41.

Spormann, A.M. and Kaiser, D.: Gliding mutants of *Myxococcus xanthus* with high reversal frequencies and small displacements. J Bacteriol 181 (1999) 2593-601.

Tang, L., Shah, S., Chung, L., Carney, J., Katz, L., Khosla, C. and Julien, B.: Cloning and heterologous expression of the epothilone gene cluster. Science 287 (2000) 640-2.

Tang, L., Yoon, Y.J., Choi, C.Y. and Hutchinson, C.R.: Characterization of the enzymatic domains in the modular polyketide synthase involved in rifamycin B biosynthesis by *Amycolatopsis mediterranei*. Gene 216 (1998) 255-65.

Thomasson, B., Link, J., Stassinopoulos, A.G., Burke, N., Plamann, L. and Hartzell, P.L.: MglA, a small GTPase, interacts with a tyrosine kinase to control type IV pili-mediated motility and development of *Myxococcus xanthus*. Mol Microbiol 46 (2002) 1399-413.

Tsoi, C.J. and Khosla, C.: Combinatorial biosynthesis of 'unnatural' natural products: the polyketide example. Chem Biol 2 (1995) 355-62.

Vahlensieck, H.F., Pridzun, L., Reichenbach, H. and Hinnen, A.: Identification of the yeast ACC1 gene product (acetyl-CoA carboxylase) as the target of the polyketide fungicide soraphen A. Curr Genet 25 (1994) 95-100.

Wall, D. and Kaiser, D.: Type IV pili and cell motility. Mol Microbiol 32 (1999) 1-10.

Wilkinson, C.J., Frost, E.J., Staunton, J. and Leadlay, P.F.: Chain initiation on the soraphen-producing modular polyketide synthase from *Sorangium cellulosum*. Chem Biol 8 (2001) 1197-208.

Wolgemuth, C., Hoiczyk, E., Kaiser, D. and Oster, G.: How myxobacteria glide. Curr Biol 12 (2002) 369-77.

Wu, K., Chung, L., Revill, W.P., Katz, L. and Reeves, C.D.: The FK520 gene cluster of *Streptomyces hygroscopicus* var. ascomyceticus (ATCC 14891) contains genes for biosynthesis of unusual polyketide extender units. Gene 251 (2000) 81-90.

Xiang, L. and Moore, B.S.: Inactivation, Complementation, and Heterologous Expression of *enc*P, a Novel Bacterial Phenylalanine Ammonia-Lyase Gene. J Biol Chem 277 (2002) 32505-9.

Xiang, L. and Moore, B.S.: Characterization of benzoyl coenzyme A biosynthesis genes in the enterocin-producing bacterium "*Streptomyces maritimus*". J Bacteriol 185 (2003) 399-404.

Xue, Q., Ashley, G., Hutchinson, C.R. and Santi, D.V.: A multiplasmid approach to preparing large libraries of polyketides. Proc Natl Acad Sci U S A 96 (1999) 11740-5.

Yu, T.W., Bai, L., Clade, D., Hoffmann, D., Toelzer, S., Trinh, K.Q., Xu, J., Moss, S.J., Leistner, E. and Floss, H.G.: The biosynthetic gene cluster of the maytansinoid antitumor agent ansamitocin from *Actinosynnema pretiosum*. Proc Natl Acad Sci U S A 99 (2002) 7968-73.

Yu, T.W., Muller, R., Muller, M., Zhang, X., Draeger, G., Kim, C.G., Leistner, E. and Floss, H.G.: Mutational analysis and reconstituted expression of the biosynthetic genes involved in the formation of 3-amino-5-hydroxybenzoic acid, the starter unit of rifamycin biosynthesis in *Amycolatopsis mediterranei* S699. J Biol Chem 276 (2001) 12546-55.

CHAPETER 2. GENERATION OF A NON-SWARMING STRAIN OF *SORANGIUM CELLULOSUM*

ABSTRACT

S. cellulosum is a gram-negative soil bacterium that belongs to the order Myxococcales that displays gliding motility and swarming on agar surfaces and does not form single colonies. This inability to form colonies on agar surfaces has a negative impact on the development of genetic systems since single cells cannot be separated and propagated as clones. Important basic investigations that require the isolation of single colonies including, the enumeration of colonies and screening and isolation of discrete genetic events, are difficult to perform. These significant limitations due to swarming of *S. cellulosum* led to the need for a strain in which single colonies could be isolated. In the closely related bacterium *Myxococcus xanthus*, the *mglA* gene has been shown to be involved in gliding motility and swarming and was used in this study to clone and sequence a homologous gene in *S. cellulosum*. The *S. cellulosum mglA* homolog was disrupted resulting in a strain with a non-swarming phenotype in which single colonies can be isolated.

INTRODUCTION

One of the most important properties of bacterial cells for cellular and molecular biology investigations is the ability to propagate colonies from single-cells. As an example, the growth of cells in liquid and their subsequent dilution and enumeration on agar media requires the formation of colonies. Also, the generation of mutant strains by either random or molecular means, requires that spatially separated single cells form colonies allowing the isolation of the genetic event.

Myxobacteria, such as *S. cellulosum*, display gliding motility on solid surfaces and undergo complex multicellular development. When starved for nutrients, myxobacteria aggregate to form a multicellular structure called a fruiting body. Within the fruiting body, cells differentiate to form myxospores as part of a complex series of developmental events. Gliding motility plays a key role in coordinate swarming required for the formation of the high cell density fruiting bodies. The most closely related bacterium to *S. cellulosum* in which gliding motility has been studied and characterized is *M. xanthus* (reviewed in Spormann, 1999). In *M. xanthus*, motility is controlled by two multigene families: the A- or adventurous system and the S- or social motility system (Hodgkin and Kaiser, 1979b; Hodgkin and Kaiser, 1979a; Wu and Kaiser, 1995). The A-system controls motility of single cells while the S-system is essential for coordinated cell movement in swarms and fruiting body formation. These two mechanisms of motility are independent of one another since mutations that inactivate either the A- or S-system reduce, but do not abolish motility. A double mutation (A^-S^-) is required to render the cells completely non-motile.

The S-motility of *M. xanthus* is equivalent to twitching motility displayed by *Neisseria* and *Pseudomonas* which is mediated by the extension and retraction of type IV pili (Whitchurch et al., 1991; Dworkin, 1999; Merz et al., 2000; Sun et al., 2000). The A-motility of *M. xanthus* is related to the secretion-mediated gliding displayed by some cyanobacteria (Hoiczyk and Baumeister, 1998; Wolgemuth et al., 2002). Expulsion of polyelectrolyte gel through membrane pores is proposed to provide the thrust for A-type gliding (Wolgemuth et al., 2002). In *M. xanthus*, many mutations have been characterized that lead to defects in one system or the other but a single locus, *mglA* (mutual function for gliding), was identified that renders colonies

completely non-swarming (Hodgkin and Kaiser, 1979a; Stephens et al., 1989). It was originally thought that the *mglA* locus might encode for components of the gliding motor since it appeared to be essential for both A- and S- motility. High-resolution motion analysis of *mglA* mutants showed that individual cells are still motile despite the non-swarming colony phenotype (Spormann and Kaiser, 1999). The *mglA* mutant cells of *M. xanthus* display abrupt, twitchy displacements and cell reversal frequencies at 10 fold higher rates than in wild-type cells (Spormann and Kaiser, 1999).

The *mglA* gene in *M. xanthus* encodes a 22kDa protein that is homologous to small eukaryotic GTPases of the Ras superfamily (Hartzell and Kaiser, 1991; Hartzell, 1997). The Ras superfamily of GTPases regulate signal transduction pathways and is involved in various cellular functions such as cytoskeletal organization and protein trafficking (Macara et al., 1996; Corbett and Alber, 2001). MglA is unique due to its lack of a C-terminal RNA-binding domain that is characteristic of prokaryotic GTPases (Caldon et al., 2001). In *M. xanthus*, MglA has been shown to interact with a protein of the serine-threonine and tyrosine (STY) kinase family and is proposed to be the intracellular switch that coordinates A- and S-motility (Thomasson et al., 2002). The interaction of a GTPase and a protein kinase of the STY family is well established in eukaryotes, but has not been previously identified in prokaryotes and demonstrates that *M. xanthus* regulates aspects of its life cycle by utilizing a signal transduction pathway with eukryotic characteristics (Thomasson et al., 2002). The gene *mglB*, is transcribed with *mglA*, and its protein product is predicted to interact and stabilize MglA (Hartzell and Kaiser, 1991).

Due to the complete non-swarming phenotype of the *mglA* mutants in *M. xanthus*, this gene seemed an attractive target for gene inactivation in *S. cellulosum*. In this investigation, the *mglA* gene was cloned from *M. xanthus* and used as a radioactive hybridization probe against *S. cellulosum* genomic DNA. A hybridizing DNA fragment from *S. cellulosum* was cloned and sequenced. The *S. cellulosum* DNA fragment contained two open-reading frames with sequences that are highly homologous to the *mglA* and *mglB* genes in *M. xanthus*. The *mglA* and *mglB* homologs in *S. cellulosum* have identical organization to the genes in *M. xanthus*. A fragment of the *S. cellulosum mglA* gene was then used to disrupt the *mglA* genomic locus through conjugation and homologous recombination. The disruption was verified genetically and the resulting strains displayed a non-swarming phenotype. The colony-forming phenotype of the mutant strains demonstrates that the *mglA* homolog in *S. cellulosum* is involved in motility and swarming.

MATERIALS AND METHODS

Strains and plasmids. The *Sorangium cellulosum* strain used was SJ3 (Jaoua et al., 1992), a streptomycin-resistant mutant of the wild-type strain *S. cellulosum* So ce26 (Gerth et al., 1994). The *Myxococcus xanthus* wild-type strain DK1622 (Kaiser, 1979) was used. *Escherichia coli* DH10B (F^- *mcr*A Δ(*mrr-hsd*RMS-*mcr*BC) ϕ80d*lac*ZΔM15 Δ*lac*X74 *deo*R *rec*A1 *end*A1 *ara*D139 Δ(*ara*, *leu*)7697 *gal*U *gal*K λ- *rps*L *nup*G) (Invitrogen, Carlsbad, CA) was used for routine cloning and *E. coli* ET 12567 (F^- *dam*13:Tn*9* *dcm*6 *hsd*M *hsd*R *rec*F143 *zjj*201:Tn*10* *gal*K2 *gal*T22 *ara*14 *lac*Y1 *xyl*5 *leu*B6 *thi*1 *ton*A31 *rps*L136 *his*G4 *tsx*78 *mtli* *gln*V44) with plasmid pUZ8002 (MacNeil et al., 1992) was used for conjugative plasmid transfer (from John Innes Centre, Norwich, UK). *E. coli* cloning vector pBluescript II SK (+) (Stratagene,

La Jolla, CA) and pCR-BluntII-Topo (Invitrogen, Carlsbad, CA) were used. The mobilizable plasmid pCIB132 was used for conjugative plasmid transfer (Schupp et al., 1995).

Media and growth conditions. *E. coli* was grown at 37°C in Luria broth or on Luria broth agar with the appropriate antibiotics. *S. cellulosum* was grown at 30°C on S42 agar (Jaoua et al., 1992) that contained: 0.5g/L tryptone (Difco, Sparks, MD), 1.5g/L $MgSO_4$ x $7H_2O$, 12g/L HEPES (N-2-hydroxyethylpiperazine-N'-2-ethanesulfonic acid), 15g/L agar, and pH adjusted to 7.4 with NaOH. After sterilization, the following sterile solution were added separately (final concentration): 0.05% $(NH_4)_2SO_4$, 0.1% $CaCl_2 \cdot 2H_2O$, 0.006% K_2HPO_4, 0.01% $Na_2S_2O_4$ (filter sterilized), 0.0008% Fe(III)-EDTA, 0.35% glucose (filter sterilized), and 3.5% (v/v) supernatant of autoclaved stationary culture of *S. cellulosum* also grown in G51t. *S. cellulosum* was grown at 30°C in G51t broth that contained; 5g/L potato starch (Sigma, St. Louis, Mo), 2g/L typtone, 1g/L probion S (Hoechst, Frankfort, Germany) , 0.5g/l $CaCl_2 \cdot 2H_2O$, 0.5g/L $MgSO_4 \cdot 7H_2O$, 12 g/L HEPES and 2g/L glucose. The pH was adjusted to 7.4 prior to autoclaving. *M. xanthus* was grown at 30°C in CTT medium: 1% Casitone, 8 mM $MgSO_4$, 10 mM Tris-HCl, 1mM KH_2PO_4 and pH adjusted to 7.6 prior to sterilization and on CTT agar plates.

DNA manipulations. Routine cloning and transformation procedures for *E. coli* were as previously described (Sambrook and Russel, 2001). PCR was preformed using *Pfu* polymerase (Stratagene, La Jolla, CA) according to the manufacturer's instructions with the modification of addition of 10% dimethyl sulfoxide in reactions, using a Robocycler Gradient 40 (Stratagene, La Jolla, CA). Restriction enzymes and T4 DNA ligase were from New England Biolabs (Beverly, MA) and used to

manufacturer's instructions. Oligonucleotides were obtained from Integrated DNA Technologies (Coralville, IA). Genomic DNA was isolated from *S. cellulosum* and *M. xanthus* with Puregene Kit according to the manufacturer's instructions (Gentra, Minneapolis, MN).

Cloning of the *S. cellulosum mglA* homolog. Genomic DNA isolated from *M. xanthus* was used in a PCR reaction to amplify a region of DNA containing the *mglA* gene using the primers 5'-GGATCCGTGGCCTCCACCCCACAGT and 5'-GAGGAGCGGGAAGGTGGCGGCGACA. The resulting 1.4-kb product was cloned into pCR-Blunt II-Topo (Invitrogen, Carlsbad, CA) according to the manufacturer's instructions creating plasmid pTMxmglA. This plasmid was then digested with *Bsp*120I and *Fsp*I to generate a 0.4-kb DNA fragment from the *mglA* gene that was radioactively labelled with Random Primers DNA labelling system (GibcoBRL, Carlsbad, CA). Genomic DNA was transferred from agarose gel to a Hybond-N+ membrane by a vacuum blotting system (Amersham Biosciences, Piscataway, NJ). Southern hybridizations were performed as described previously (Church and Gilbert, 1984) using the *mglA* gene fragment as a probe. The hybridizations and washes were carried out at 65^{o}C. The images from the Southern hybridizations were captured on a phosphor screen and visualized using the Storm 860 optical scanner and ImageQuant software (both from Amersham Biosciences, Piscataway, NJ). *S. cellulosum* genomic DNA was digested with *Bam*HI and fragments between 3 and 4-kb were isolated from 0.7% agarose gel, purified with QIAquick gel extraction kit (Qiagen, Valencia, CA) and ligated into *Bam*HI digested pBluescript SK II (+) vector. About 300 clones were picked and incubated in 96-well plates containing Luria broth with 100 ug/ml ampicillin, overnight at 37^{o}C. The

cultures in the 96-well plates were then replicated onto Hybond-N+ membrane supported on Luria broth agar with 100 ug/ml ampicillin and incubated overnight at 30^{o}C. DNA was isolated from the colonies by saturating the membrane, with the colonies facing up, with 0.5 N NaOH, 1.5 M NaCl for 5 minutes followed by neutralization with 1 M Tris-HCl, pH 8.0 for 5 minutes. The membranes were then briefly washed in 2 X SSC (1 X SSC is 0.15 M NaCl and 15 mM sodium citrate). The DNA was immobilized to the membrane by UV crosslinking at 120mJ/cm^2 (Stratalinker 2400, Stratagene, La Jolla, CA). Colony debris was removed by wiping the membrane with kimwipes wetted with 0.1M Tris-HCl, pH 7.5/2 X SSC and the blot was washed in 2 X SSC. Three positive clones were detected in the colony blot hybridization with the *M. xanthus mglA* fragment and these were sequenced with universal primers specific for each side of the cloning site and all three were determined to be identical. Clone 2C9 was completely sequenced by primer walking.

Analysis of *S. cellulosum* clone 2C9. DNA sequencing was carried out by using a BigDye terminator cycle sequencing reaction with ABI 377 and ABI 3700 sequencers (Applied Biosystems, Foster City, CA). Primary DNA sequence was analyzed using Sequencher software (Gene Codes Corporation, Ann Arbor, MI). Codon usage and open reading frame analysis was done by using the web-based program FramePlot 2.3.2 (http://www.nih.go.jp/~jun/cgi-bin/frameplot.pl). Protein sequences were analyzed with Vector NTI 7 (InforMax, Fredrick, MD). Database searches were performed using the standard BLAST 2.0 algorithm at the National Center for Biotechnology Information (NCBI) homepage (http://www.ncbi.nlm.nih.gov/BLAST/).

Gene disruption. A 0.5-kb internal fragment of the *S. cellulosum mglA* homolog was amplified with the PCR primers 5'-GGGCCCTCTA TTACGGCCCGGGTCTCTGC and 5'-GGGCCCTTGAGCGTGTCGAACACCCC which contained *Psp*OMI sites (underlined) and cloned into pCR-Blunt II-Topo vector. The resulting plasmid, pTSmglAKO, was sequence verified and digested with *Psp*OMI. The 0.5-kb fragment was isolated and cloned into the *Not*I digested pCIB132 (*Not*I and *Psp*OMI generated ends are compatible). The resulting plasmid pCIB132mglAKO was sequence verified and transformed into *E. coli* ET12567 containing the helper plasmid pUZ8002 for conjugation with *S. cellulosum*.

Conjugations. Conjugations were carried out by a method described previously (Jaoua et al., 1992) with modifications described below. *S. cellulosum* was grown for 5 days in 15 mLs S42 broth with shaking at 200 rpm at 30^{o}C after a 10% inoculation from a stationary phase culture. *E. coli* containing the mobilizable plasmid pCIB132mglAKO was inoculated from a single colony and grown overnight at 37^{o}C in 10mLs of Luria broth with the appropriate antibiotics. The *E. coli* cells were washed 2 times in 10 mLs of Luria broth and resuspended in a final volume of 0.5 mL of G51t. The *S. cellulosum* cells were pelleted and resuspened in 0.5 mL of G51t. *S. cellulosum* cells were then mixed with the *E. coli* cells and spread onto a S42 agar plate. The cell mixture was incubated at 30^{o}C for 40 hours, the cell mixture scraped off the plates, and resuspended in 1mL of G51t. Dilutions were spread on plates of S42 containing 25 ug/mL of phleomycin (Cayla, Toulouse, France) for selection of transconjugates and 50 ug/mL of apramycin (Melford, Ispwich, UK) for counterselection against the donor *E. coli*. After 3 weeks of growth on the S42 plates containing phelomycin and apramycin, single colonies were picked with a sterile

toothpick and spread onto S42 plate containing 25 ug/mL phleomycin as approximately 1-cm^2 patches. These patches were incubated at 30^{o}C for 3 weeks and then the cells from approximately one fourth of the patch were used to inoculate 5mLs of G51t supplemented with 25 ug/mL phleomycin. These cultures were grown for 5 days at 30^{o}C and then used for genomic DNA preparations.

Microscopy of colonies. Edges of colonies formed by mutants were compared with the wild-type strain SJ3 using a Nikon E600 microscope (Melville, NY). The strains were grown in G51 liquid media for 5 days and streaked onto S42 agar and incubated for four weeks at 30^{o}C. Images were captured using the Spot RT Slider imaging system (Diagnostic Instruments Inc., Sterling Heights, MI).

RESULTS

Cloning of the *S. cellulosum mglA* homolog. In *M. xanthus*, mutations in *mglA* have been shown to yield non-swarming cells that form compact single colonies. A 1.5-kb region of the *M. xanthus* chromosome containing the *mglA* gene was amplified by PCR. A 0.5-kb *Bsp*120I-*Fsp*I fragment containing a portion of the *mglA* open reading frame (ORF) was isolated and radioactively labeled. To demonstrate that *S. cellulosum* contained a *mglA* homolog, a Southern hybridization to *S. cellulosum* chromosomal DNA was preformed using the *Bsp*120I-*Fsp*I fragment. The *Bam*HI-, *Kpn*I- and *Bgl*II-digested *S. cellulosum* DNA displayed a single hybridizing fragment from *S. cellulosum* while the *Pst*I digest displayed two fragments (Figure 1). This experiment demonstrated that *S. cellulosum* contained a discrete region of DNA with homology to the *mglA* gene from *M. xanthus*.

S. cellulosum genomic DNA was digested with *Bam*HI, electrophoresed on an agarose gel and the 3 to 4-kb fragments were isolated. Genomic DNA was purified

from the gel and ligated into a *Bam*HI digested pBluescript SK II (+) vector and transformed into *E. coli* DH10B. About 300 white colonies were picked and screened by hybridization with the *Bsp*120I-*Fsp*I probe. Three colonies displayed strong hybridization to the probe (Figure 2). The three hybridizing clones, 1C10, 2C9 and 3F2 were sequenced from universal primer sites on the pBluescript SK II (+) vector and it was determined that the three clones contained the same DNA insert.

Sequence analysis of *Bam*HI clone 2C9. In order to characterize clone 2C9, the nucleotide sequence of the cloned fragment was determined. The fragment consists of 3546 bp and has a G+C content of 68.5%. Analysis of the sequence revealed one ORF consisting of 90.7% G+C base in the third position of codons within the ORF. BLAST analysis using the deduced amino acid sequence of this ORF (160 amino acids) revealed a high level of homology to *mglB* from *M. xanthus*. A second ORF within the fragment, with a 92.9% G+C content in the third position of codons was also found. Using the deduced 196 amino acids from this ORF in BLAST analysis showed that the ORF has a high level of homology to *mglA* from *M. xanthus*. The DNA sequence for the *S. cellulosum mglA* homolog and its deduced amino acid sequence are shown in Figure 3. The protein encoded by the *mglA* homolog from *S. cellulosum* is 82.5% identical and 89.2% similar to MglA of *M. xanthus* (Figure 4). The protein encoded by the *mglB* homolog from *S. cellulosum* is 66.8% identical and 82.5% similar to MglB of *M. xanthus*.

Gene disruption of the *S. cellulosum mglA* homolog. The role of the *S. cellulosum mglA* homolog in swarming motility was investigated by gene disruption. Gene disruptions were carried out based on a method developed for *S. cellulosum* (Jaoua et al., 1992). An internal fragment of the *S. cellulosum mglA* ORF was cloned

into the mobilizable plasmid pCIB132 and transformed into *E. coli* ET12567 containing the helper plasmid pUZ8002, which supplies the transfer functions for plasmid mobilization. This *E. coli* strain was used as a donor for mating experiments with *S. cellulosum*. After matings, phleomycin resistant transconjugates were isolated while apramycin was used for counterselection of the donor *E. coli*. No plasmids have been found that are able to replicate in *S. cellulosum* and therefore transconjugates from pCIB132-derived plasmids were maintained in *S. cellulosum* only after homologous recombination between the cloned insert and the homologous region on the chromosome (Jaoua et al., 1992). Integration of the plasmid into the chromosomal *mglA* locus of phleomycin-resistant exconjugates was confirmed with a Southern hybridization (Figure 5A). Hybridization of the *S. cellulosum mglA* fragment to *Bam*HI digested genomic DNA from wild type *S. cellulosum* revealed the expected 3.5-kb band. Hybridization of *Bam*HI-*Hind*III digested genomic DNA from transconjugates shows two hybridizing bands of 6.6 and 4.7-kb as a result of the gene duplication that arises from the integration of the plasmid at the *mglA* locus (Figure 5B).

The colony morphology of the *mglA* transconjugates was different from those of *S. cellulosum* strains derived from unrelated integration events into the biosynthetic gene cluster of the polyketide soraphen A (data not shown). The *mglA*-specific transconjugates displayed a small colony type with compact edges and appeared to be deficient in swarming motility. In contrast, those from unrelated integration events displayed the normal swarming phenotype, with spreading edges and larger size.

The *mglA* disruption strain grew similarly in generation time and density to wild type strains in liquid media. The *mglA* disruption strain displayed single,

compact, non-swarming colonies when streaked onto S42 medium compared to the non-colony forming, swarming phenotype seen in the SJ3 (wild-type swarming) strain (Figure 6). The morphology of the edge of a colony is shown in Figure 7. Strain SJ3 displays an irregular, undulating edge with flares of cells, which is indicative of wild type swarming. The *mglA* mutant strain displays a smooth and distinct edge morphology associated with non-swarming cells. These results indicate that the *mglA* gene from *S. cellulosum* has a role in swarming motility.

DISCUSSION

The *mglA* gene that had been shown to be involved in the swarming motility phenotype of *M. xanthus* was used as a probe to clone the *mglA* homolog from *S. cellulosum*. A portion of the *mglA* homolog from *S. cellulosum* was used in a gene disruption experiment and the resulting *mglA* disruption strain displayed a non-swarming phenotype.

S. cellulosum is of considerable interest due to the abundance of secondary metabolites it produces and the spectrum of activities some of these molecules possess (Reichenbach, 2001). The main focus of research with *S. cellulosum* has been the isolation, structure determination and activity profiling of the secondary metabolites recovered from cultured strains. Little progress in strain development of *S. cellulosum* has been achieved due to the difficulties in the growth and genetic manipulation of *S. cellulosum*. With the large number of bioactive metabolites recovered from *S. cellulosum*, it seems somewhat surprising that more effort has not been exerted in the development of genetic systems in this bacterium. Few papers have been published on the molecular genetic manipulation of *S. cellulosum* (Jaoua et

al., 1992; Pradella et al., 2002) and any further improvements in the growth characteristic or genetic systems would be an important development.

One major obstacle in genetic and microbiological experimentation with *S. cellulosum* is its swarming motility. When streaked or plated on solid growth media, bacterial cells swarm and aggregate so that single, distinct colonies do not form. Considerable effort has been made in the development of suitable growth media for *S. cellulosum* yet, *S. cellulosum* still grows poorly on the optimized media and fails to form colonies. The inability to isolate single colonies is a major limitation in a number of areas including enumerating colonies to isolation and propagation of genetic events.

As a consequence of investigations of its complex life cycle, motility has been well studied in *M. xanthus* (reviewed by Spormann, 1999). Cells must coordinate motility and swarming for aggregation and myxospore formation. In *M. xanthus*, the *mglA* gene has been shown to be involved in swarming and may be the cellular switch that controls A- and S-system motility (Hartzell and Kaiser, 1991; Hartzell, 1997; Thomasson et al., 2002). *S. cellulosum* also displays swarming and gliding motility but no detailed investigation of A- and S- motility systems in *S. cellulosum* has been reported. The presence of an *mglA* homolog and its similar role in swarming motility in *S. cellulosum* suggests that the regulation and the mechanisms of motility are similar in *S. cellulosum* to that of *M. xanthus*.

The generation of a non-swarming strain of *S. cellulosum* should now expand our ability to conduct genetic investigations of *S. cellulosum*. One interesting area of investigation would be to determine if the *S. cellulosum mglA* mutant strains are deficient in aggregation, fruiting body and myxospore formation. One would suspect

these same deficiencies seen in *M. xanthus mglA* mutants would also occur in *S. cellulosum mglA* mutants. The basic microbiological aspect of reliable colony enumeration is another area in which the *mglA* mutant strain of *S. cellulosum* would be useful. Until now, there has not been a way to determine cell numbers related to colony forming units and various aspects of growth rate and number. Since only one resistance marker (phleomycin resistance) is available in *S. cellulosum* So ce 26, an *mglA* mutant strain in which the phleomycin resistance marker is removed, is needed for further molecular genetic experiments. Attempts to carry out gene replacement using a strategy with two different counterselection markers were unsuccessful in *S. cellulosum*.

ACKNOWLEDGMENTS

I would like to thank István Molnár for critical reading of this manuscript and helpful discussions. I would also like to thank members of the Syngenta Sequencing group for their assistance with analysis of samples; Makoto Ono, An Hu and Racella McNair. I would also like to thank Dale Kaiser for the kind gift of *Myxococcus xanthus* strain DK1622.

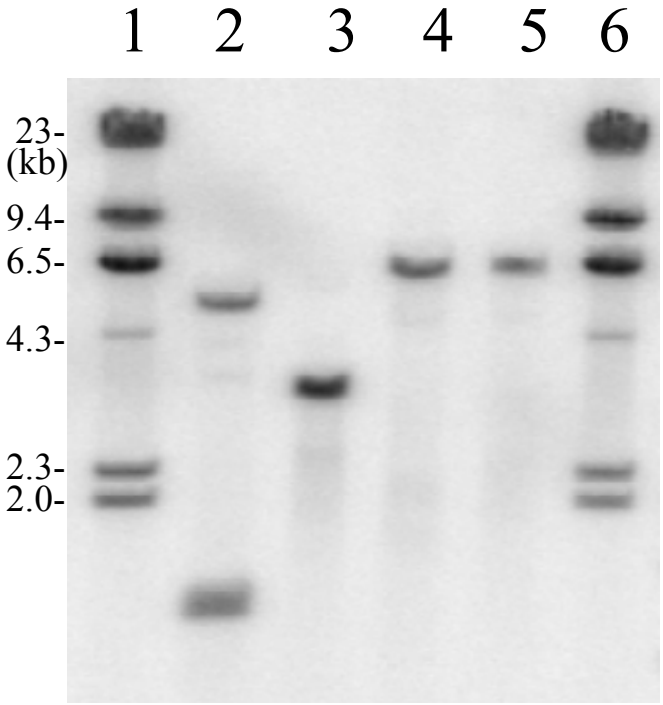


Figure 1. Southern hybridization of genomic DNA from *S. cellulosum* to *mglA* from *M. xanthus*. The genomic DNA was digested with *Pst*I, *Bam*HI, *Kpn*I or *Bgl*II (lanes 2, 3, 4, and 5 respectively) and hybridized with ^{32}P-labelled *Fsp*I-*Not*I fragment of *mglA* from *M. xanthus*. Markers (lanes 1 and 6) are lambda DNA digested with *Hind*III.

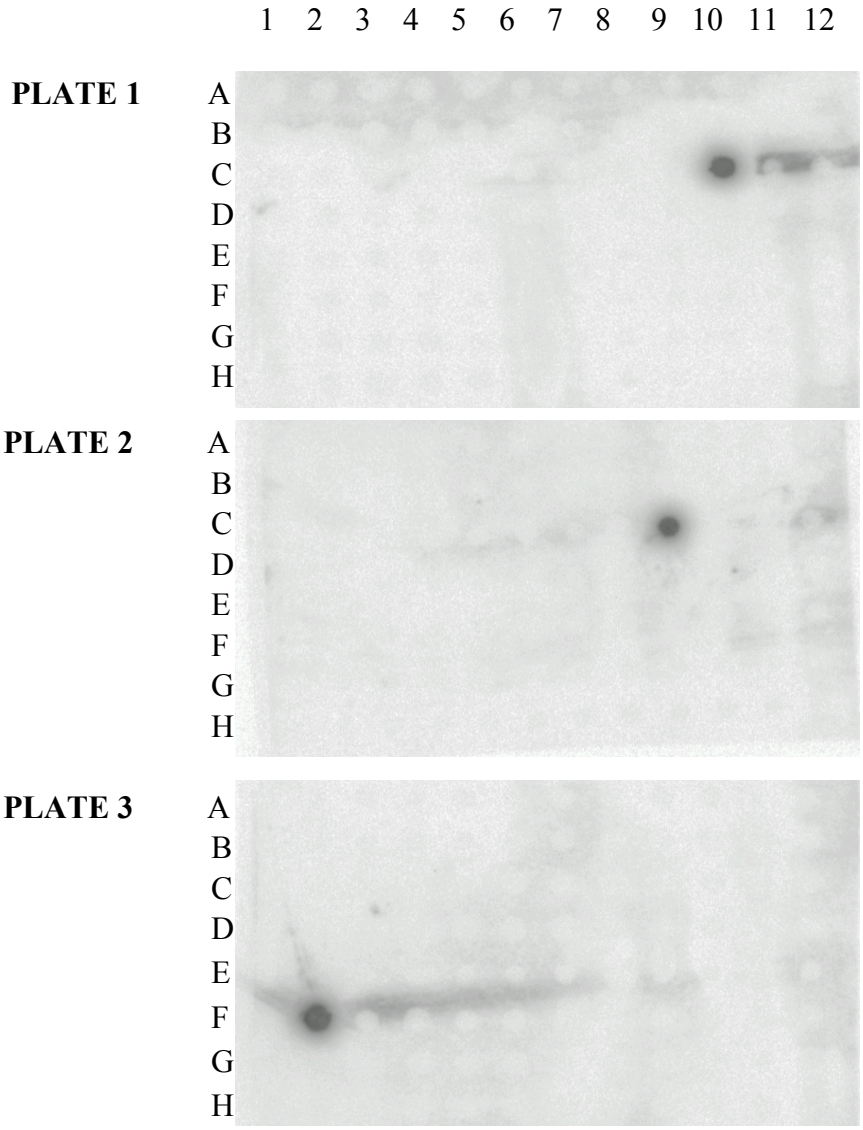
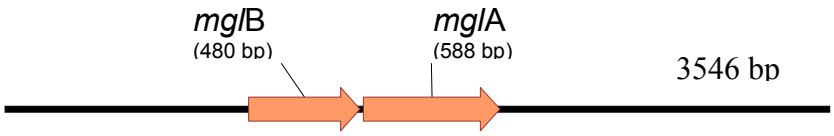


Figure 2. Colony blot hybridization of a *S. cellulosum* library to *mglA* of *M. xanthus*. Clones 1C10, 2C9 and 3F2 showed hybridization to the probe. All three positive clones were determined to have identical inserts and 2C9 was completely sequenced.

A.



B.

```
  1  ATGAGCTTCA TCAATTACAT GGCGCGCGAG ATCAACTGCA AGATCGTCTA
       M  S  F   I  N  Y   M  A  R  E   I  N  C   K  I  V

 51  TTACGGCCCG GGTCTCTGCG GCAAGACGAC GAACCTGCAG TACATCTACG
     Y  Y  G  P   G  L  C   G  K  T   T  N  L  Q   Y  I  Y

101  AGCGCACCAA TCCGGACGCG AAGGGGAAGA TGATCTCCCT CGCCACGGAG
      E  R  T   N  P  D  A   K  G  K   M  I  S   L  A  T  E

151  ACCGACCGGA CACTCTTCTT CGACTTCCTC CCGCTCGCGC TGGGCGAGAT
       T  D  R   T  L  F   F  D  F  L   P  L  A   L  G  E

201  CCGCGGCTTC AAGACGCGGT TCCACCTGTA CACGGTGCCG GGGCAGGTCT
     I  R  G  F   K  T  R   F  H  L   Y  T  V  P   G  Q  V

251  TCTACGACGC GAGCCGCAAG CTCATCCCGA AGGGCGTCGA CGGCGTGGTC
      F  Y  D   A  S  R  K   L  I  P   K  G  V   D  G  V  V

301  TTCGTGGCCG ACTCACAGCT CGGCCGCATC GAGGCGAACC AGGAGTCGGT
       F  V  A   D  S  Q   L  G  R  I   E  A  N   Q  E  S

351  GGAGAACCTC CGCACGAACC TGGCCGAGCA AGGCTACTCG CTCGACAAGA
     V  E  N  L   R  T  N   L  A  E   Q  G  Y  S   L  D  K

401  TCCCGTACGT CATCCAGTAC AACAAGCGGG ATCTGCCGGA GATCGGCACG
      I  P  Y   V  I  Q  Y   N  K  R   D  L  P   E  I  G  T

451  GTCGAGGAGC TCCGGGCGCT GCTCAACCCG ACCAATGTTC CCGATTTTGA
       V  E  E   L  R  A   L  L  N  P   T  N  V   P  D  F

501  AGGGGTGGCG CGGACAGGCG TCGGGGTGTT CGACACGCTC AAGGCCGTCG
     E  G  V  A   R  T  G   V  G  V   F  D  T  L   K  A  V

551  CGAAGCTCGT GCTGACCGAG CTGAAGAAGG GCGGCCCCTA G
      A  K  L   V  L  T  E   L  K  K   G  G  P   -
```

Figure 3. *S. cellulosum mglA* genomic region, nucleotide and protein sequence. (A); Genomic region of *S. cellulosum* containing *mglA* and *mglB* (B); The nucleotide sequence of the *mglA* open reading frame and deduced amino acid sequence are from sequence derived from subclone 2C9.

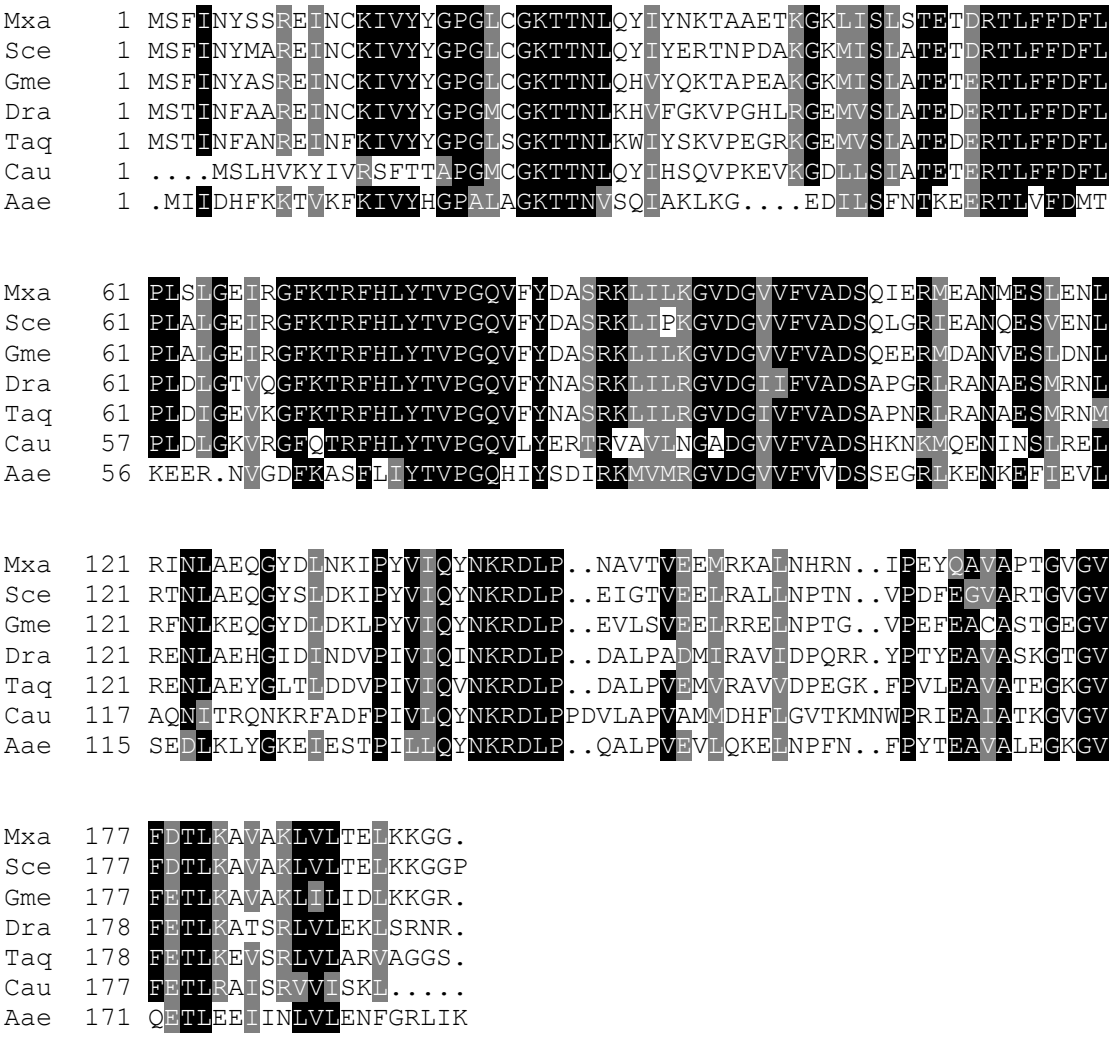


Figure 4. Alignment of *mglA* homologs. Sequences were aligned with ClustalW using the BLOSUM62mt scoring matrix. Identical amino acids are shaded black and similar amino acids gray with a 80% threshold. Abbreviations (Accession numbers are in parentheses): Mxa: *M. xanthus* MglA (AAA25389); Sce: *S. cellulosum* mglA protein; Gme: *Geobacter metallireducens* Hypothetical protein (ZP_00080324); Dra: *Deinococcus radiodurans* gliding motility protein (NP_294577); Taq: *Thermus aquaticus* probable mglA (T52479); Cau: *Chloroflexus aurantiacus* hypothetical protein (ZP_00017611); Aae; *Aquifex aeolicus* gilding motility protein (NP_214074).

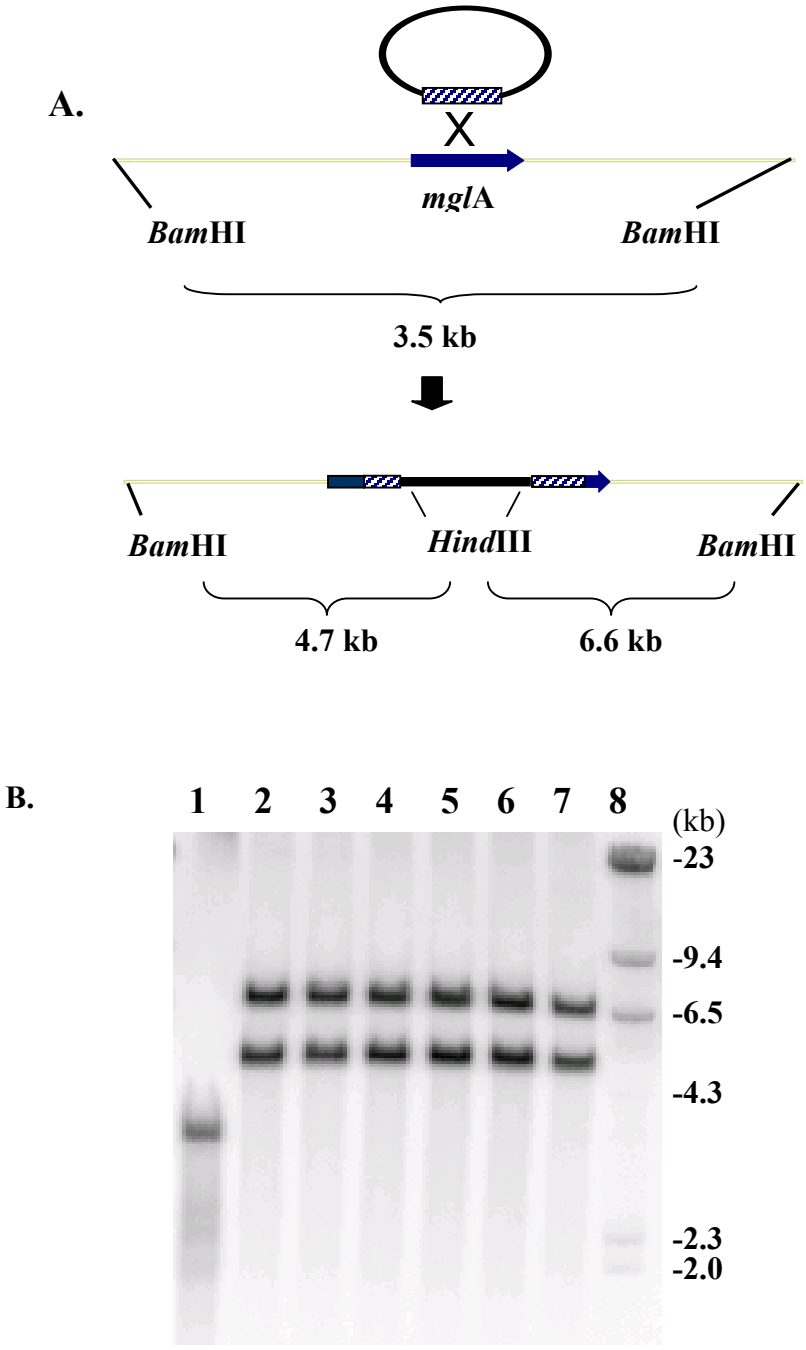


Figure 5. Homologous recombination and integration into *S. cellulosum mglA*. (A); Schematic represention of the *S. cellulosum mglA* integration event. A fragment of the *S. cellulosum mglA* gene (hatched blue box) was used to disrupt the chromosomal copy of *mglA* (solid blue arrow) through homologous recombination. The result is two truncated copies of the *mglA* gene and the interspersed vector (black line). B. Southern hybridization of *S. cellulosum* chromosomal DNA with the ^{32}P-labelled *mglA* disruption fragment. *Bam*HI digested genomic DNA from strain SJ3 is in lane 1. DNA digested with *Bam*HI-*Hind*III from 6 isolates are shown in lanes 2 through 7. Lane 8 is DNA size ladder Lamda digested with *Hind*III

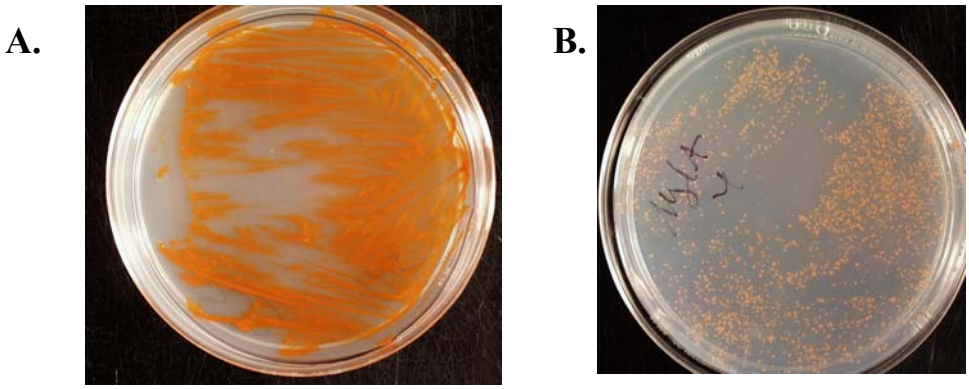


Figure 6. Non-swarming phenotype of the *mglA* mutant strain. Strains of *S. cellulosum* were grown for 5 days at 30°C in G51t media. Cells were streaked out on S42 media and plates incubated for four weeks at 30°C. (A); Swarming phenotype of *S. cellulosum* wild-type strain SJ3 lacking discrete colonies. (B); Non-swarming phenotype of *S. cellulosum mglA* mutant with discrete colonies.

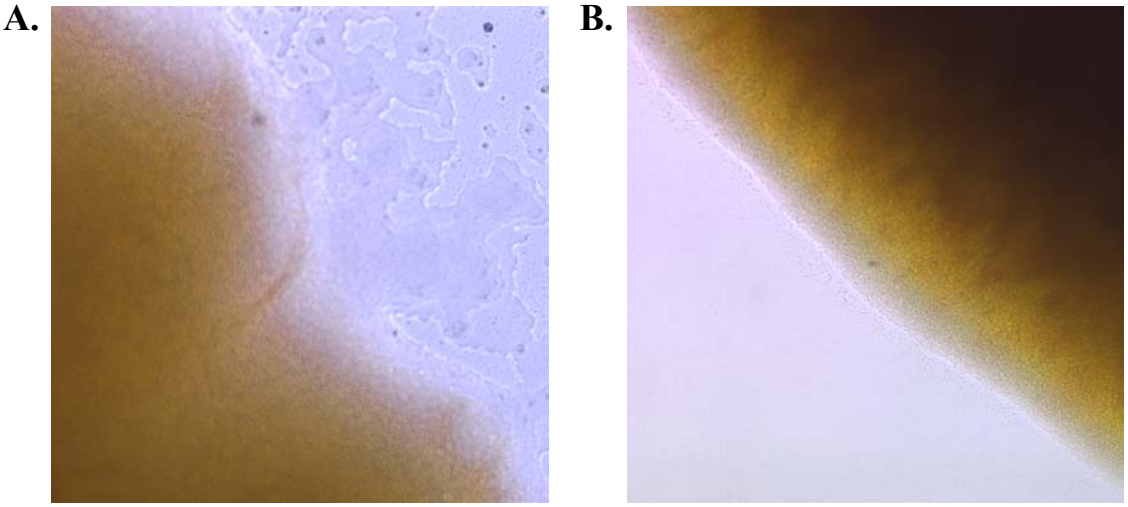


Figure 7. Morphology of the edge of *S. cellulosum* colonies**.** Strains of *S. cellulosum* were grown for 5 days at 30°C in G51t liquid media. Cell were streaked on S42 and incubated for four weeks. Colonies were illuminated from below and images captured through a Nikon E600 microscope. A. Wild-type edge morphology of strain SJ3 colony. B. Non-swarming edge morphology of a *mglA* mutant strain colony.

REFRENCES

Caldon, C.E., Yoong, P. and March, P.E.: Evolution of a molecular switch: universal bacterial GTPases regulate ribosome function. Mol Microbiol 41 (2001) 289-97.

Church, G.M. and Gilbert, W.: Genomic sequencing. Proc Natl Acad Sci U S A 81 (1984) 1991-5.

Corbett, K.D. and Alber, T.: The many faces of Ras: recognition of small GTP-binding proteins. Trends Biochem Sci 26 (2001) 710-6.

Dworkin, M.: Fibrils as extracellular appendages of bacteria: their role in contact-mediated cell-cell interactions in *Myxococcus xanthus*. Bioessays 21 (1999) 590-5.

Gerth, K., Bedorf, N., Irschik, H., Hofle, G. and Reichenbach, H.: The soraphens: a family of novel antifungal compounds from *Sorangium cellulosum* (Myxobacteria). I. Soraphen A1 alpha: fermentation, isolation, biological properties. J Antibiot (Tokyo) 47 (1994) 23-31.

Hartzell, P. and Kaiser, D.: Function of MglA, a 22-kilodalton protein essential for gliding in *Myxococcus xanthus*. J Bacteriol 173 (1991) 7615-24.

Hartzell, P.L.: Complementation of sporulation and motility defects in a prokaryote by a eukaryotic GTPase. Proc Natl Acad Sci U S A 94 (1997) 9881-6.

Hodgkin, J. and Kaiser, D.: Genetics of gliding motility in *Myxococcus xanthus* (Myxobacterales): two gene systems control movement. Mol Gen Genet 171 (1979a) 177-191.

Hodgkin, J. and Kaiser, D.: Genetics of gliding motility in *Myxococcus xanthus* (Myxobacterales): genes controling movement of single cells. Mol Gen Genet 171 (1979b) 167-171.

Hoiczyk, E. and Baumeister, W.: The junctional pore complex, a prokaryotic secretion organelle, is the molecular motor underlying gliding motility in cyanobacteria. Curr Biol 8 (1998) 1161-8.

Jaoua, S., Neff, S. and Schupp, T.: Transfer of mobilizable plasmids to *Sorangium cellulosum* and evidence for their integration into the chromosome. Plasmid 28 (1992) 157-65.

Kaiser, D.: Social gliding is correlated with the presence of pili in *Myxococcus xanthus*. Proc Natl Acad Sci U S A 76 (1979) 5952-6.

Macara, I.G., Lounsbury, K.M., Richards, S.A., McKiernan, C. and Bar-Sagi, D.: The Ras superfamily of GTPases. Faseb J 10 (1996) 625-30.

MacNeil, D.J., Occi, J.L., Gewain, K.M., MacNeil, T., Gibbons, P.H., Ruby, C.L. and Danis, S.J.: Complex organization of the *Streptomyces avermitilis* genes encoding the avermectin polyketide synthase. Gene 115 (1992) 119-25.

Merz, A.J., So, M. and Sheetz, M.P.: Pilus retraction powers bacterial twitching motility. Nature 407 (2000) 98-102.

Pradella, S., Hans, A., Sproer, C., Reichenbach, H., Gerth, K. and Beyer, S.: Characterisation, genome size and genetic manipulation of the myxobacterium *Sorangium cellulosum* So ce56. Arch Microbiol 178 (2002) 484-92.

Reichenbach, H.: Myxobacteria, producers of novel bioactive substances. J Ind Microbiol Biotechnol 27 (2001) 149-56.

Sambrook, J. and Russel, D.W.: Molecular Cloning: a laboratory manual, 3rd ed. Cold Spring Habor Laboratory Press, Cold Spring Habor, N.Y., 2001.

Schupp, T., Toupet, C., Cluzel, B., Neff, S., Hill, S., Beck, J.J. and Ligon, J.M.: A *Sorangium cellulosum* (myxobacterium) gene cluster for the biosynthesis of the macrolide antibiotic soraphen A: cloning, characterization, and homology to polyketide synthase genes from actinomycetes. J Bacteriol 177 (1995) 3673-9.

Spormann, A.M.: Gliding motility in bacteria: insights from studies of *Myxococcus xanthus*. Microbiol Mol Biol Rev 63 (1999) 621-41.

Spormann, A.M. and Kaiser, D.: Gliding mutants of *Myxococcus xanthus* with high reversal frequencies and small displacements. J Bacteriol 181 (1999) 2593-601.

Stephens, K., Hartzell, P. and Kaiser, D.: Gliding motility in *Myxococcus xanthus*: *mgl* locus, RNA, and predicted protein products. J Bacteriol 171 (1989) 819-30.

Sun, H., Zusman, D.R. and Shi, W.: Type IV pilus of *Myxococcus xanthus* is a motility apparatus controlled by the *frz* chemosensory system. Curr Biol 10 (2000) 1143-6.

Thomasson, B., Link, J., Stassinopoulos, A.G., Burke, N., Plamann, L. and Hartzell, P.L.: MglA, a small GTPase, interacts with a tyrosine kinase to control type IV pili-mediated motility and development of *Myxococcus xanthus*. Mol Microbiol 46 (2002) 1399-413.

Whitchurch, C.B., Hobbs, M., Livingston, S.P., Krishnapillai, V. and Mattick, J.S.: Characterisation of a *Pseudomonas aeruginosa* twitching motility gene and

evidence for a specialised protein export system widespread in eubacteria. Gene 101 (1991) 33-44.

Wolgemuth, C., Hoiczyk, E., Kaiser, D. and Oster, G.: How myxobacteria glide. Curr Biol 12 (2002) 369-77.

Wu, S.S. and Kaiser, D.: Genetic and functional evidence that Type IV pili are required for social gliding motility in *Myxococcus xanthus*. Mol Microbiol 18 (1995) 547-58.

CHAPTER 3. CHARACTERIZATION OF THE BIOSYNTHETIC GENE CLUSTER FOR THE ANTIFUNGAL POLYKETIDE SORAPHEN A FROM *SORANGIUM CELLULOSUM* SO CE26

This chapter is included directly as a reprint from Gene 2002 volume 285, pages 257-267. This paper is primarily the work of James Ligon and co-authors. I (Ross Zirkle) contributed to the gene disruption experiments and a portion of the DNA sequencing and analysis.

ELSEVIER

Gene 285 (2002) 257–267

GENE
AN INTERNATIONAL JOURNAL ON GENES AND GENOMES

www.elsevier.com/locate/gene

Characterization of the biosynthetic gene cluster for the antifungal polyketide soraphen A from *Sorangium cellulosum* So ce26

James Ligon[a,*], Steven Hill[a], James Beck[a], Ross Zirkle[a], István Molnár[a], Jennifer Zawodny[a], Stephanie Money[a], Thomas Schupp[b]

[a]*Syngenta Biotechnology, Inc., P.O. Box 12257, 3054 Cornwallis Road, Research Triangle Park, NC 27709, USA*
[b]*Novartis Pharma AG, CH 4002, Basel, Switzerland*

Received 19 September 2001; received in revised form 13 December 2001; accepted 2 January 2002
Received by A.M. Campbell

Abstract

A genomic DNA region of over 80 kb that contains the complete biosynthetic gene cluster for the synthesis of the antifungal polyketide metabolite soraphen A was cloned from *Sorangium cellulosum* So ce26. The nucleotide sequence of the soraphen A gene region, including 67,523 bp was determined. Examination of this sequence led to the identification of two adjacent type I polyketide synthase (PKS) genes that encode the soraphen synthase. One of the soraphen A PKS genes includes three biosynthetic modules and the second contains five additional modules for a total of eight. The predicted substrate specificities of the acyltransferase (AT) domains, as well as the reductive loop domains identified within each module, are consistent with expectations from the structure of soraphen A. Genes were identified in the regions flanking the two soraphen synthase genes that are proposed to have roles in the biosynthesis of soraphen A. Downstream of the soraphen PKS genes is an *O*-methyltransferase (OMT) gene. Upstream of the soraphen PKS genes there is a gene encoding a reductase and a group of genes that are postulated to have roles in the synthesis of methoxymalonyl-acyl carrier protein (ACP). This unusual extender unit is proposed to be incorporated in two positions of the soraphen polyketide chain. One of the genes in this group contains distinct domains for an AT, an ACP, and an OMT.

Keywords: Polyketide synthase; Methyltransferase; Methoxymalonyl-CoA; Methoxymalonyl-CoA synthase

1. Introduction

Although more than two-thirds of the known biologically active microbial natural products are produced by the actinomycetes, more recently the myxobacteria have been found to be a rich source of interesting new natural products. The myxobacteria are a group of Gram-negative eubacteria that are micropredators or saprophytes in soil environments. They have a gliding means of motility, complex social interactions, and morphological differentiation that results in the production of fruiting bodies and myxospora. In a screening program, analysis of 800 wild-type myxobacterial isolates revealed that over 60% produced metabolites with antimicrobial activity (Reichenbach et al., 1988). Strains of *Sorangium cellulosum* are slow growing, cellulolytic members of the myxobacteria that have been found to produce a diverse range of natural products. These include sorangicin that inhibits the initiation of transcription in eubacteria (Irschik et al., 1986), epothilones A and B (Gerth et al., 1996a) that have antitumor and cytotoxic activity, and the antifungal compounds jerangolid A and ambruticin (Gerth et al., 1996b) and soraphen A (Gerth et al., 1994). Soraphen A, produced by *S. cellulosum* strain So ce26, has a unique mode of action among fungicides. It is a strong inhibitor of fungal acetyl coenzyme A carboxylase and therefore disrupts the synthesis of fatty acids (Vahlensieck et al., 1994). Soraphen A is active against a wide range of fungal species, especially plant pathogenic fungi, and was being developed as an agricultural fungicide until it was discovered that it is a weak teratogen. Elucidation of the structure of soraphen A demonstrated that it is a macrolide polyketide containing an 18-membered lactone ring with an unsubstituted phenyl side ring (Bedorf et al., 1993).

A cosmid clone containing about 40 kilobase pairs (kb) of DNA from *S. cellulosum* So ce26 that contains genes involved in the synthesis of soraphen A was isolated by our

Abbreviations: ACP, acyl carrier protein; AT, acyltransferase; CL, coenzyme A ligase; DH, dehydratase; ER, enoylreductase; kb, kilobase pairs; KR, ketoreductase; KS, β-ketoacyl synthase; OMT, *O*-methyltransferase; PKS, polyketide synthase

* Corresponding author. Tel.: +1-919-541-8645; fax: +1-919-541-8557.
E-mail address: james.ligon@syngenta.com (J. Ligon).

0378-1119/02/$ - see front matter
PII: S0378-1119(02)00396-7

group and 6.4 kb of nucleotide sequence from this clone was reported (Schupp et al., 1995). This sequence revealed that, as expected, a modular type I polyketide synthase (PKS) is responsible for the synthesis of soraphen A. The sequence contained a single open reading frame (ORF) larger than 6.4 kb, whose beginning and end were not present in the sequenced region, and that coded for the end of one type I PKS module and the beginning of a second. Examination of the soraphen A structure suggests that eight PKS modules should be required for the synthesis of the soraphen A macrolide ring and that other post-PKS activities such as *O*-methylation might also be necessary. An interesting feature of soraphen A is the presence at C3–4 and C11–12 of adjacent oxygenated carbon atoms in the macrolide ring. In each case, the adjacent carbons are predicted to be derived from a single two-carbon extender molecule. Typically, polyketide biosynthesis involves incorporation of two-carbon extender molecules oxygenated only at the β-carbon. Therefore, the identity of the extender unit at C3–4 and C11–12 is intriguing. Another unusual feature of soraphen A is the phenyl group that is predicted to be incorporated at the beginning of the macrolide synthesis as the starter unit. These features can be studied by superimposing on the structure of soraphen A the deduced functions of the proteins encoded in the biosynthetic gene cluster, the results of precursor feeding assays and the characteristics of biosynthetic mutants. In this report, we describe the isolation, characterization, and sequence analysis of the complete soraphen A biosynthetic gene cluster and describe some of the interesting relationships between the genetics and structure of this molecule. Prior to this report, the only other description of an entire gene cluster for the biosynthesis of a polyketide from *Sorangium* is the recent characterization of the genes for the biosynthesis of epothilones A and B (Molnár et al., 2000).

2. Materials and methods

2.1. Strains, plasmids, and growth conditions

The *S. cellulosum* strain used in this study was the wild-type strain So ce26 (Gerth et al., 1994). *Escherichia coli* strains HB101 and JM101 were used as cloning hosts and vectors used for cloning and sequencing are described in Schupp et al. (1995). *Escherichia coli* and *S. cellulosum* were cultured as described by Schupp et al. (1995).

2.2. Genetic procedures

Standard genetic techniques with *E. coli* and for in vitro DNA manipulations were according to Maniatis et al. (1982).

2.3. Construction of genomic libraries and cloning of the soraphen A genes

The construction of a genomic library of DNA from strain So ce26 in cosmid vector pHC79 and the isolation of cosmid p98/1 containing a portion of the soraphen A gene cluster were previously described (Schupp et al., 1995). Cosmid pJL3 was isolated using a 1.3 kb *Bam*HI

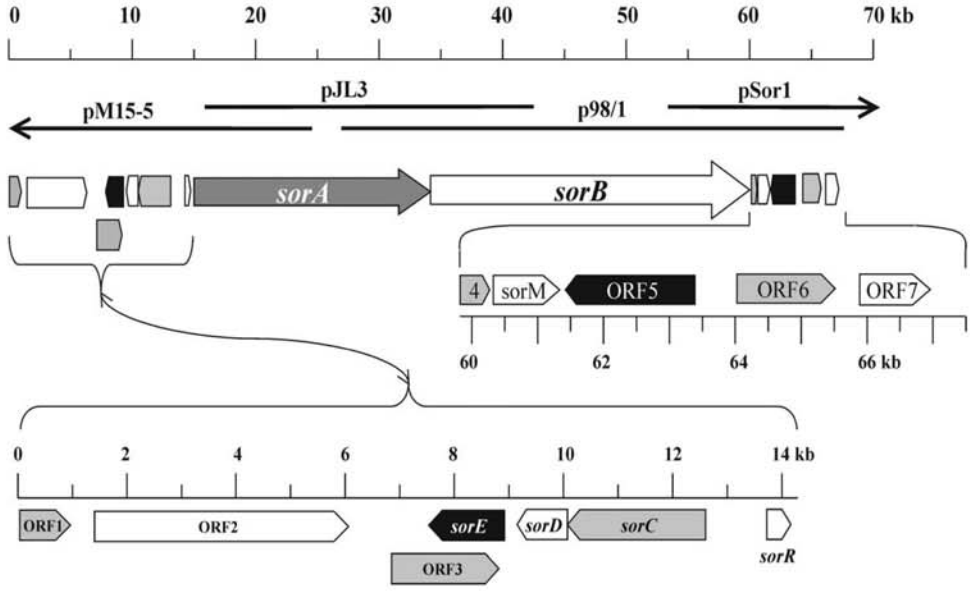


Fig. 1. Organization of the soraphen A gene cluster. The location of genes and ORFs identified in the soraphen gene region are shown as arrows. The regions flanking the *sorA* and *sorB* genes are shown below in larger scale with their respective size scales. The overlapping cosmid clones that cover the region are depicted at the top as solid lines.

fragment from the extreme left end (relative to the map in Fig. 1) of p98/1 as a hybridization probe. Similarly, a 375 bp DNA fragment amplified from the left end of pJL3 was used as a probe to identify cosmid pM15-5, and a 3.2 kb *Bgl*II fragment at the right end of p98/1 was employed to isolate pSor1 (Fig. 1).

2.4. DNA sequencing and analysis

DNA sequencing was accomplished as previously described (Schupp et al., 1995). In all cases, the complete sequence of both DNA strands was determined. Primary DNA sequence data were assembled using Sequencher software (Gene Codes Corp., Ann Arbor, MI) and analyzed using the web-based program FramePlot 2.3 (http://www.nih.go.jp/~jun/cgi-bin/frameplot.pl). Protein sequences were analyzed with SeqWeb version 1.2 (Genetics Computer Group, Madison, WI). The annotated sequence appears in GenBank under the Accession number U24241.

3. Results

3.1. Organization of the soraphen A biosynthetic gene cluster

A 67,523 bp segment of the soraphen A gene locus was sequenced and 14 putative ORFs were identified. The deduced amino acid sequences of each of these were compared to protein databases and matches for each ORF are listed in Table 1.

3.1.1. sorA and sorB

Two ORFs encode large, multifunctional type I PKSs that together make up the soraphen synthase. One gene, *sorA*, encodes a putative PKS with three biosynthetic modules, while the second, *sorB*, encodes five modules (Table 1, Fig. 2). Each of these modules contains one β-ketoacyl synthase (KS) domain that contains the signature active site sequence (TACSS) with the invariant cysteine. All modules also contain at least one acyl carrier protein

Table 1
Deduced functions of ORFs in the soraphen A biosynthetic gene cluster

Gene/ORF	Length (amino acids)	% G/C in 3rd base	Proposed function or sequence homologies detected
SorA	6315	80.3	PKS
			Mod1 ACP KS ATcb[a] ATa[a] KR ACP
			Mod2 KS ATa DH ER KR ACP
			Mod3 KS ATm[a] DH ER KR ACP
SorB	8815	83.0	PKS
			Mod4 KS ATa KR ACP
			Mod5 KS ATp[a] DH ER KR ACP
			Mod6 KS ATp KR ACP
			Mod7 KS ATm KR ACP
			Mod8 KS ATp DH[b] KR ACP TE
SorC	863	90.0	Proposed methoxymalonyl-CoA synthase
			AT ACP MT
			NosB polyketide synthase, AT domain, *Nostoc* sp. GSV224 ($E = 9 \times 10^{-72}$, AF204805)
			EpoD epothilone synthase, AT domain, *Sorangium cellulosum* So ce90 ($E = 2 \times 10^{-63}$, AF210843)
			FkbJ ACP, *Streptomyces hygroscopicus* ($E = 1 \times 10^{-5}$, AF235504)
			FkbM 31-*O*-methyltransferase, *Streptomyces hygroscopicus* ($E = 6 \times 10^{-50}$, AF235504)
			FkbG 31-*O*-demethyl-FK506-methyltransferase, *Streptomyces* sp. ($E = 2 \times 10^{-49}$, U65940)
SorD	295	93.1	β-Hydroxybutyryl-CoA dehydrogenase, *Clostridium acetobutylicum* ($E = 7 \times 10^{-62}$, A43723)
SorE	463	91.4	Acyl-CoA dehydrogenase, *Streptomyces coelicolor* ($E = 8 \times 10^{-59}$, CAB46799)
SorM	346	75.8	RapM methyltransferase, *Streptomyces hygroscopicus* ($E = 4.6 \times 10^{-56}$, X86780)
SorR	170	69.5	Spinosin synthase, KR domain, *Saccharopolyspora spinosa* ($E = 6 \times 10^{-20}$, AX089460)
Orf1	327	81.7	None
Orf2	1567	84.0	None
Orf3	677	91.2	Residues 6–375: conserved hypothetical protein, *Pseudomonas aeruginosa* ($E = 1 \times 10^{-36}$, AE004939)
			Residues 89–368: VgrG protein, unknown function, *Vibrio cholerae* ($E = 7 \times 10^{-27}$, AAF95932) and *E. coli* ($E = 8 \times 10^{-22}$, AF044503)
Orf4	167	76.2	Hypothetical protein MmyY, unknown function, *Streptomyces coelicolor* ($E = 1 \times 10^{-10}$, AJ276673)
Orf5	663	93.5	Conserved hypothetical protein YesW, unknown function, *Bacillus subtilis* ($E = 1 \times 10^{-177}$, CAB12524)
Orf6	502	94.2	β-Mannanase, *Dictyoglomus thermophilum* ($E = 7 \times 10^{-49}$, AF013989)
Orf7	335	93.8	Xylanase-arabinofuranosidase bifunctional enzyme, *Streptomyces chattanoogensis* ($E = 3 \times 10^{-15}$, AF121864)

[a] ATa, ATm, ATp, acyltransferases that incorporate acetate, methoxymalonate, and propionate, respectively. ATcb, acyltransferase that incorporates the carboxybenzoate starting unit.
[b] Predicted to be inactive.

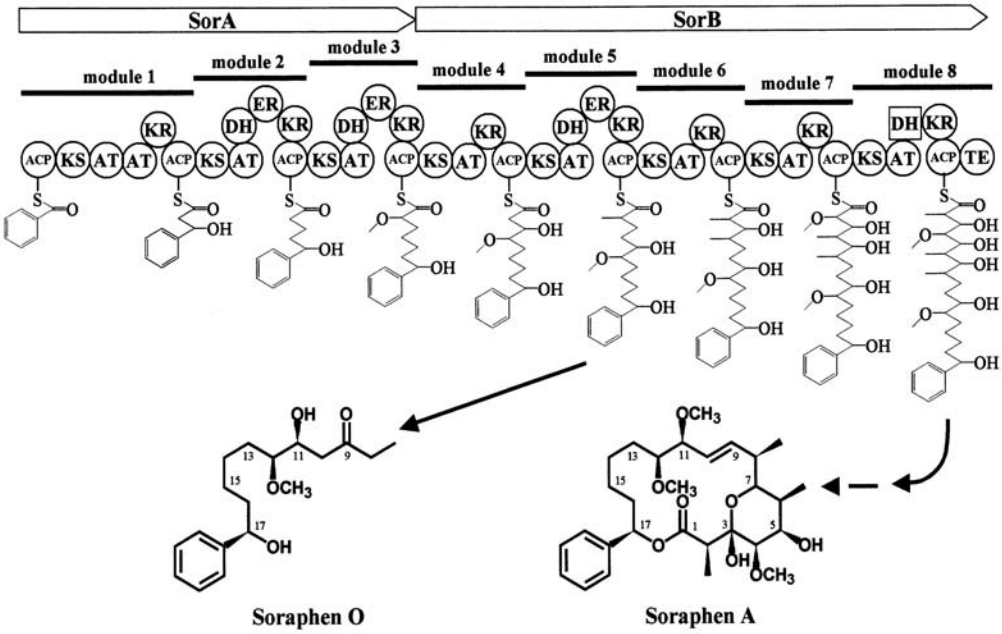


Fig. 2. The soraphen PKS and the chemical structures of soraphen A and soraphen O. The structures of the growing polyketide chains as deduced from the predicted activities of the constituent domains of the PKS are shown attached to the synthase. The DH domain of module 8 that is predicted to be inactive is shown as a square. The numbering scheme for the carbons of soraphen O corresponds to the carbons of soraphen A to facilitate comparisons of the two structures.

(ACP) domain that is recognized by the signature 4′-phosphopantetheine binding site LGxxS. The first module of *sorA* contains two ACP domains and two acyltransferase (AT) domains, as well as KS and ketoreductase (KR) domains (Table 1). This suggests that module 1 of *sorA* performs the functions of both the loading module and the first extension module. The other two modules of *sorA* and all of the modules in *sorB* contain single KS, AT, KR, and ACP domains (Table 1). Some of the modules also contain dehydratase (DH) and/or enoylreductase (ER) domains (Table 1). Module 8, the last module in the *sorB* gene, includes a thioesterase (TE) domain at the end of the module. The deduced amino acid sequence of the TE domain of SorB includes the invariant GxSxG motif common to ATs, serine proteases, and TEs. It also contains the PGdH motif located near the C-terminus that has been shown by site-directed mutagenesis to be essential for activity in TE domains (Witkowski et al., 1991).

In a previous report, eight DNA fragments from the soraphen A gene region were used in gene disruption experiments in So ce26 and six of these were shown to result in a soraphen A-non-producing phenotype (Schupp et al., 1995). By analysis of the complete DNA sequence of the region, it was determined that the six fragments that caused disruption of soraphen A biosynthesis include portions of the *sorA* and/or *sorB* genes. These results clearly demonstrate that the two type I PKS genes, *sorA* and *sorB*, are required for the biosynthesis of soraphen A.

3.1.2. sorR

The ORF designated *sorR* is located adjacent to *sorA* (Fig. 1) and is transcribed in the same direction as *sorA* and *sorB* and it encodes a protein that is highly homologous to the KR domains of type I PKSs (Table 1). SorR contains the consensus motif, GxGxxGxxxA, that with the motif Lx(S,G)Rx(G,T,A) has been implicated in the formation of the NADP(H) binding site of functional KR domains. In *sorR*, the latter motif is present as LrRrG and it contains the invariant arginine (underlined). Based on these results, we assign this ORF the genetic designation *sorR* and propose that this gene encodes a reductase.

3.1.3. sorC

The ORF designated *sorC* encodes a protein that contains three distinct domains. The N-terminal domain encompassing residues 100–452 is highly homologous to AT domains of type I PKSs (Table 1). The conserved AT-active site motif GySlG, compared to the consensus AT-active site motif GHSxG, is present in this region (Fig. 3A). The C-terminal domain of SorC, comprising amino acid residues 598–863, is highly homologous to *O*-methyltransferases (OMT) (Table 1). There are three highly degenerate motifs that are commonly found, at characteristic spacings, in methyltransferases (Kagan and Clarke, 1994). The C-terminal domain of SorC contains sequences with strong (motifs I and III) or plausible (motif II) similarities to these motifs at the appropriate distances from each

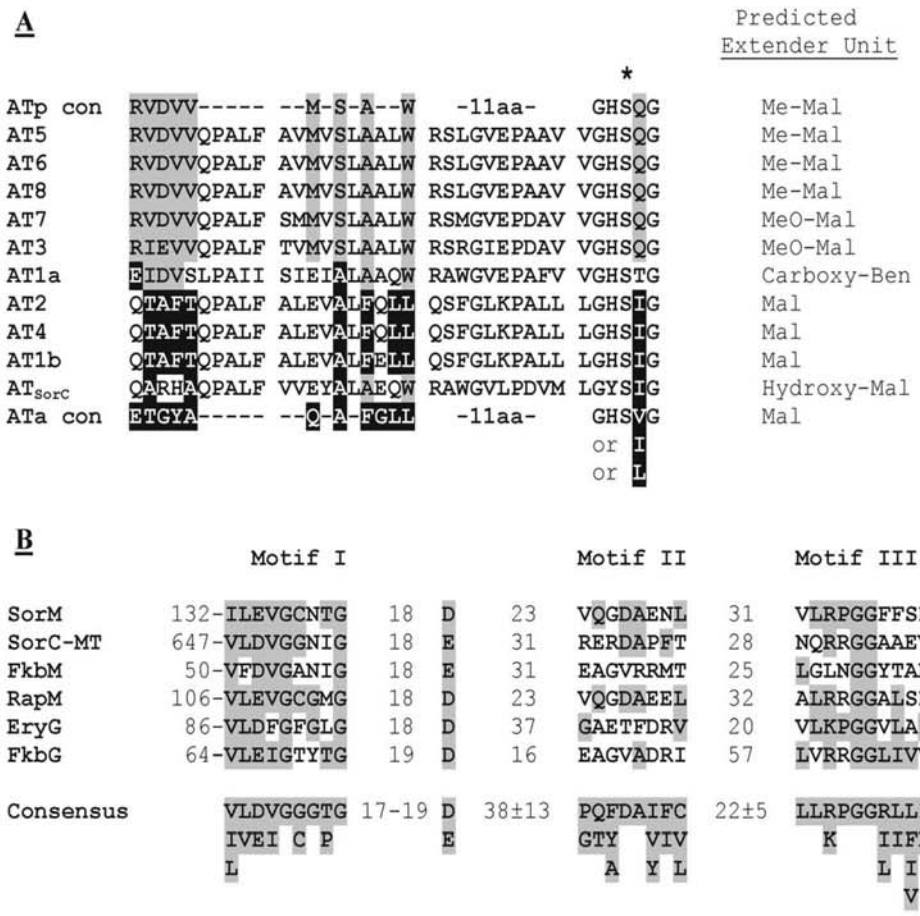


Fig. 3. (A) Multiple sequence alignment of the substrate specificity motifs of the AT domains of the soraphen synthase using the PILEUP program. The motifs are ordered by their relatedness to one another. The predicted substrate specificities for each AT are listed (methylmalonyl-CoA, Me-Mal; malonyl-CoA, Mal; methoxymalonyl-CoA, MeO-Mal; hydroxymalonyl-CoA, Hydroxy-Mal; and carboxybenzoyl-CoA, Carboxy-Ben). The consensus motifs for ATs with specificity for Me-Mal (ATp con) and Mal (ATa con) are shown above and below, respectively, the alignment of the soraphen ATs. The residues in the ATs that are conserved with the Me-Mal consensus are shown in a shaded background, while those conserved with the Mal consensus motif are shown as white text in a black background. The active site serine is marked with an asterisk. (B) Alignment of the methyltransferase signature motifs (Kagan and Clarke, 1994) with SorM and the OMT domain of SorC, and with OMTs from other polyketide biosynthetic systems, including FkbM from *Streptomyces* sp. (AF235504), RapM from *Streptomyces hygroscopicus* (X86780), EryG from *Saccharopolyspora erythraea* (S18533), and FkbG from *S. hygroscopicus* (U65940). Shaded residues are identical to the proposed consensus (Kagan and Clarke, 1994).

other within this predicted domain (Fig. 3B). Based on these observations, we propose that this domain of the SorC protein is an OMT.

In the middle of SorC, between the AT- and OMT-homologous domains, there is a short region consisting of residues 453–597 that is homologous to ACP domains of PKSs. This domain contains the sequence DANfFRLGGNSl-LATRLIDxxV that is very similar to the consensus phosphopantetheine attachment site (D,E,Q,G,S,T,A,L,M,K,R,-H)-(L,I,V,M,F,Y,S,T,A,C) (G,N,Q) (L,I,V,M,F,Y)-R-L-G-G-N-S-L-(P,C,F,Y)-A-T-R-L-I-x-x-V (underlined amino acids are conserved in the SorC motif; the bold serine residue is the attachment site for the 4′-phosphopantetheine prosthetic group that is an essential component of the ACP domains in PKSs). Based on these results, SorC appears to be a unique, multifunctional protein containing AT, ACP, and OMT domains.

3.1.4. sorD

Immediately downstream of the *sorC* gene is an ORF, designated *sorD*, that is highly homologous to several β-hydroxybutyryl-CoA dehydrogenases (Table 1). β-Hydroxybutyryl-CoA dehydrogenases are NAD-dependent enzymes and contain the consensus NAD(P) binding motif GxGxxGxxxA near the N-terminus. The *sorD* gene also contains this motif near the N-terminus.

3.1.5. sorE

An ORF located downstream of the *sorD* gene and named *sorE* was found to be highly similar to acyl-CoA dehydro-

genases (Table 1). Two highly conserved sequence patterns that are in the PROSITE database and are typically present in acyl-CoA dehydrogenases were found in the SorE peptide. The first motif is represented by the sequence LSEPGVGSDA including residues 124–133 of SorE (underlined amino acids are conserved in all acyl-CoA dehydrogenases) and the second motif is represented by the sequence LVIDGEKRWITY that is present in residues 147–158 of SorE.

3.1.6. ORFs 1, 2, and 3

Other ORFs located to the left of *sorE* (Fig. 1) were identified. BlastP analysis of the deduced amino acid sequences of ORFs 1 and 2 did not identify any close matches in the databases. ORF1 is a partial ORF as its start codon is not present in the sequenced region and ORF2 does not have a plausible ribosome binding site preceding it. The first half of ORF3 was found to have high similarity to a conserved hypothetical protein from *Pseudomonas aeruginosa* and to the VgrG proteins from *Vibrio cholerae* and *E. coli* (Table 1). ORF3 overlaps and is transcribed divergently from the *sorE* gene (Fig. 1).

3.1.7. ORF4

In the region downstream, and to the right of the *sorB* gene, five ORFs were identified. BlastP analysis of the deduced amino acid sequence of ORF4 revealed that it is closely related to the hypothetical protein MmyY from *Streptomyces coelicolor* (Table 1). The *mmyY* gene in *S. coelicolor* is located in a cluster of genes that are involved in the synthesis of the antibiotic methylenomycin, but its function is unknown.

3.1.8. sorM

Immediately downstream from ORF4 (29 bp) and transcribed in the same direction is an ORF with high similarity to methyltransferases (Fig. 1, Table 1). Similar to the OMT domain in SorC described in Section 3.1.3, SorM contains the three signature motifs common to most methyltransferases (Fig. 3B). Based on these features, we propose that this ORF encodes an OMT and we have given it the genetic designation of *sorM*.

3.1.9. ORFs 5, 6, and 7

To the right of *sorM* is ORF5 whose direction of transcription is opposite that of *sorM* (Fig. 1). BlastP analysis of ORF5 revealed that it is highly similar to the hypothetical protein YesW from *Bacillus subtilis* (Table 1) whose function is unknown. ORF6 and ORF7 are transcribed divergently from ORF5. ORF6 was found to be homologous to the β-mannanase of *Dictyoglumus thermophilum* and ORF7 was found to be homologous to a xylanase-arabinofuranosidase bifunctional enzyme from *Streptomyces chattanoogensis* (Table 1).

3.2. Chain initiation and elongation

As is the case with erythromycin (Donadio and Katz, 1992), rifamycin (Schupp et al., 1998), and epothilone (Molnár et al., 2000), the soraphen synthase exhibits colinearity between the biosynthetic sequence and the order of the genes. Biosynthesis of soraphen A begins with the phenyl side group at C17 that is derived from phenylalanine and is probably incorporated as carboxybenzoyl-CoA (Reichenbach and Höfle, 1994). The loading of this unit that initiates soraphen A chain formation is most likely performed by module 1, the first module of the SorA PKS (Fig. 2). This module has two ACP and two AT domains and is proposed here to be a functional hybrid between a conventional loading module and the first extension module, termed here the 'starter module'. Thus, the two ATs would load the carboxybenzoyl-CoA starter and the first malonyl-CoA extender units onto their respective ACPs within module 1, followed by the condensation of the two units catalyzed by the KS within the same module. The SorA PKS includes two other modules in addition to module 1 and SorB contains five additional modules for a total of eight PKS modules (Table 1). This is precisely the number of modules that are predicted to be required to synthesize a molecule such as soraphen A.

In the synthesis of polyketides, the ATs within each module are responsible for selecting the appropriate extender unit and transferring them to the ACP within the same module. An examination of AT domains from many PKSs has led to the ability to predict the substrate specificity for an AT based on the presence of certain consensus motifs (Haydock et al., 1995) and groups of conserved amino acids (August et al., 1998). The ATs that load malonyl-CoA and those that load longer chain alkyl units, for example methylmalonyl-CoA, or hydroxy- or methoxymalonyl-CoA, fall into two distinct classes. The ATs from the soraphen PKS can also be assigned to these two categories based on these criteria. The second AT in module 1, AT1b, and the ATs from modules 2 and 4 are predicted to direct the loading of malonyl-CoA extender units (Fig. 3A). These predictions are consistent with the structure of soraphen A since there are no methyl side groups present on the carbons that are incorporated by modules 1, 2, or 4 (Fig. 2), and also with isotopic labeling data showing that the corresponding carbons are derived from acetate (malonyl-CoA) units (Hill et al., 1998). The ATs from modules 3, 5, 6, 7, and 8 are predicted to direct the loading of methylmalonyl-CoA or longer chain alkyl-, hydroxy-, or methoxymalonyl-CoA extender units (Fig. 3A). The methyl side groups on C2, C6 and C8 of the soraphen A macrolactone ring (Fig. 2) were shown to be derived from the incorporation of methylmalonyl-CoA extender units by modules 8, 6, and 5, respectively. There are methoxy side groups at C12 and C4 of the soraphen A macrolactone ring (Fig. 2), and the corresponding extender units were shown by isotopic labeling studies to be derived from glycolate, with the methyl groups origi-

nating from SAM (Hill et al., 2000). Thus, incorporation of either hydroxymalonyl-CoA or methoxymalonyl-CoA is expected from modules 3 and 7, respectively. If hydroxymalonyl-CoA were the extender unit for these modules, post-PKS *O*-methylation by dedicated OMTs would be required to form the methoxy groups. The incorporation of hydroxymalonyl-CoA or methoxymalonyl-CoA extender units has been proposed for niddamycin (Kakavas et al., 1997), FK506 (Motamedi et al., 1996), and FK520 (Wu et al., 2000). The niddamycin AT6 responsible for loading this extender unit was found to be similar to those that normally load methylmalonyl-CoA, while the FK506 and FK520 ATs 7 and 8 were similar to malonyl-CoA-specific ATs. The substrate specificity motifs for the ATs of modules 3 and 7 of the soraphen synthase resemble those of the methylmalonyl-CoA-specific ATs (Fig. 3A). Thus, the predictions of AT substrate specificity for the AT domains in the soraphen synthase are completely consistent with those predicted from the structure of soraphen A.

The first AT domain in module 1, AT1a, is expected to direct the incorporation of an unusual starter unit. This was proposed to be carboxybenzoyl-CoA derived from phenylalanine since feeding of radiolabeled phenylalanine has been shown to label the phenyl side group of soraphen A (Reichenbach and Höfle, 1994). Examination of the substrate specificity motifs of AT1a shows that it has some features common to both malonyl-CoA- and methylmalonyl-CoA-specific ATs (Fig. 3A). In a recent experiment, the *sorA* AT1a was substituted for the loading AT in the truncated PKS DEBS1-TE (P.F. Leadlay, pers. commun.) to produce a triketide lactone derivative with a benzoyl starter unit, providing experimental proof for the role of AT1a as proposed above. SorC also contains an AT domain that clusters more closely with malonyl-CoA-specific ATs, but it also has some features common to methylmalonyl-CoA-specific ATs (Fig. 3A).

3.3. Ketone processing activities

Domains that are predicted to modify and process the ketones formed by the addition of each extender unit are present in all modules of the soraphen synthase. All of the modules contain KR domains with a consensus NAD(P)H binding motif GxGxxGxxxA and the Lx(S,G)Rx(G,T,A,) motif with the invariant arginine that together are found in all KR domains. This suggests that all of the KR domains within the soraphen synthase are active. Since all of the β-keto groups formed during the synthesis of soraphen A, with the exception of the terminal ketone at C1, are eliminated or reduced to hydroxyls, it is predicted that all modules of the soraphen synthase should contain active KR domains. Therefore, the presence of apparently active KR domains in each of the modules of the soraphen synthase is consistent with expectations from the structure of soraphen A.

In addition to the KR domain, modules 2, 3, and 5 of the soraphen synthase also contain DH and ER domains, while module 8 has only a DH domain (Table 1). The DH domains of modules 2, 3, and 5 contain the conserved DH active site motif LxxHxxxGxxxxP (Donadio and Katz, 1992). The DH in module 8 has a similar motif but the C-terminal proline residue is one residue closer to the conserved glycines and it is followed closely by three arginine residues. Examination of the soraphen A structure shows that C3 that is predicted to be processed by module 8 is not dehydrated (Fig. 2). This strongly suggests that the DH in module 8 of the soraphen A synthase is not functional.

The ER domains in modules 2, 3, and 5 contain the ER consensus motif GGVGxAAxQxA that has been proposed to have a role in the formation of the NADP(H)-binding site. The presence of DH and ER domains in these modules, in addition to a KR domain, suggests that they completely reduce the β-keto group and eliminate the ketone oxygen derived from the condensation of the nascent polyketide chain and the in-coming extender unit. It is predicted that the reductive loop domains present in modules 2, 3, and 5 process carbons 15, 13, and 9 of the soraphen A macrolactone ring, respectively. As expected from the complete complement of reductive activities present in modules 2 and 3, C13 and C15 of soraphen A are completely reduced (Fig. 2). Although C9 is also expected to be fully reduced due to the presence of KR, DH, and ER domains in module 5, it is not. There is a double bond in the soraphen A macrolactone ring between C9 and C10 (Fig. 2). The synthesis of double bonds by PKSs is normally due to the presence of active KR and DH domains without a functional ER within a PKS module. However, double bonds produced in this manner are located between the α-carbon of an extender unit and the β-carbon of the previous extender unit. The double bond in the soraphen macrolactone ring cannot be derived in this manner since it is between the α- and β-carbons of the same extender unit. Therefore, we postulate that the nascent soraphen macrolactone synthesized by the soraphen PKS is fully reduced at C9 due to the presence of apparently functional KR, DH, and ER domains in module 5 and that the double bond between C9 and C10 of soraphens C (Fig. 4A) and A is introduced by a post-PKS modification. The detection of small amounts of soraphen B (Fig. 4A) in *S. cellulosum* So ce26 fermentations which lacks this double bond and is completely reduced at C9 (Reichenbach and Höfle, 1994) supports this hypothesis.

3.4. Derivation of the methoxy groups

As noted above, the methoxy groups at C4 and C12 of soraphen A could be derived either from the incorporation of hydroxymalonyl-CoA with subsequent post-PKS methylation, or alternatively, from the direct incorporation of methoxymalonyl-CoA. On the other hand, the site of the third methoxy group of soraphen A, C11, is a β-carbon, so it must be derived by the *O*-methylation of a hydroxyl group at C11 by post-PKS modification. This is consistent with the finding that module 4 that is responsible for the

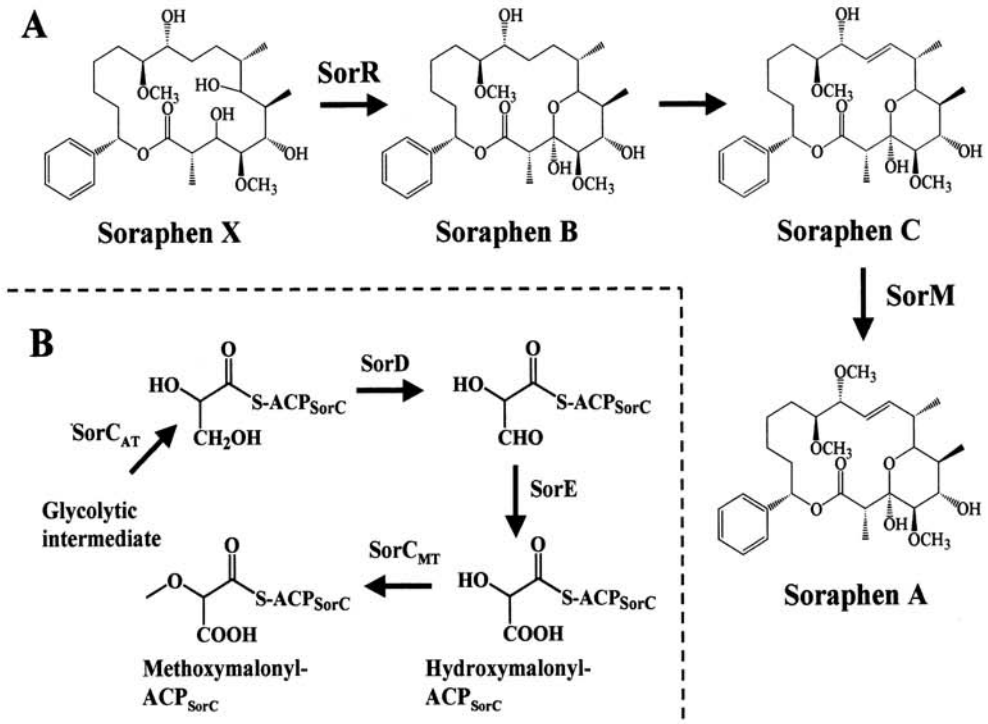


Fig. 4. Proposed pathways for post-PKS processing and methoxymalonyl-CoA biosynthesis. (A) Proposed post-PKS steps from the presumed soraphen A precursor that is the product of the soraphen PKS to soraphen A. (B) The proposed pathway for the synthesis of methoxymalonyl-ACP extender units (adapted from Wu et al., 2000) that are predicted to be incorporated by modules 3 and 7.

processing of C11 contains a KR domain but does not contain DH or ER domains (Fig. 2). Thus, the nascent β-keto group of C11 might be reduced to a hydroxyl by the PKS, and methylated in a post-PKS tailoring step. Correspondingly, the C11-hydroxyl form of soraphen A, named soraphen C (Fig. 4A), has been detected in small amounts in *S. cellulosum* So ce26 cultures (Reichenbach and Höfle, 1994). Further support comes from the finding that the C11-hydroxyl of soraphen C is efficiently methylated by a cell-free extract of *S. cellulosum* So ce26 (H. Reichenbach, pers. commun.). Examination of the DNA sequence in the regions flanking the soraphen synthase genes revealed the presence of only one gene, *sorM*, that encodes a conventional OMT. We propose that the SorM OMT is responsible for the post-PKS methylation of the hydroxyl at C11 (Fig. 4A).

In a survey of UV-induced mutants of *S. cellulosum* strain So ce26 that failed to produce soraphen A, a mutant was identified that produced soraphen O (Fig. 2), a linear intermediate of soraphen A biosynthesis (H. Reichenbach and G. Höfle, pers. commun.). The carbon of soraphen O that corresponds to C12 of soraphen A has a methoxy side group and C11 is hydroxylated (Fig. 2). These results support the hypothesis that module 3 directs the incorporation of methoxymalonyl-CoA rather than hydroxymalonyl-CoA and, conversely, that the methoxy group of C11 of soraphen A is derived by the post-PKS methylation of a hydroxyl formed by the soraphen synthase. If soraphen O were synthesized with a hydroxyl at the carbon corresponding to C12 of soraphen A, it is unlikely that this structure would be recognized by an OMT whose role is to methylate this group. This would result in the presence of a hydroxyl side group in the corresponding position of soraphen O. Fittingly, the hydroxyl group at the carbon of soraphen O that corresponds to C11 of soraphen A has a hydroxyl side group rather than a methoxy group as is the case for soraphen A. This is likely due to the inability of soraphen O to be recognized by SorM as a substrate for post-PKS methylation due to the gross structural differences between soraphen O and soraphen C.

4. Discussion

In a previous study, we reported the cloning of a cosmid clone from *S. cellulosum* strain So ce26 that contained genomic DNA involved in the biosynthesis of the antifungal polyketide soraphen A (Schupp et al., 1995). In that report, the sequence of a 6.4 kb DNA fragment was described. Here we report the cloning and sequence analysis of the complete soraphen A biosynthetic gene cluster.

The incorporation of six-membered rings as starting units

in type I polyketides is established in the case of rifamycin (Schupp et al., 1998), rapamycin (Aparicio et al., 1996), FK506 (Motamedi and Shafiee, 1998), and FK520 (Wu et al., 2000). In all of these examples, a domain encoding a carboxylic acid:coenzyme A ligase (CL) is found within the loading module of the respective PKS gene. The type I polyketide ansatrienin has a side chain derived from cyclohexane carboxylic acid and alanine and the *ansK* gene encoding a CL is required for the incorporation of this moiety (Chen et al., 1999). In contrast, a CL, although present, is not essential for the biosynthesis and incorporation of the benzoyl-CoA starter unit of the type II polyketide enterocin (Hertweck and Moore, 2000). Since the biosynthesis of soraphen A is thought to begin with a carboxybenzoyl-CoA starter unit, it would be expected that the soraphen A PKS would have a CL domain in the loading module to produce the CoA ester of carboxybenzoate. Instead of a CL domain, there are two AT and two ACP domains within a single 'starter module' of the soraphen A PKS (Fig. 2). The only other PKS that is known to contain two AT and two ACP domains similar to the 'starter module' of the soraphen synthase is the myxothiazol PKS from the myxobacterium *Stigmatella aurantiaca* (Silakowski et al., 1999). We propose that the first AT domain in module 1 of the soraphen PKS, AT1a, recognizes carboxybenzoyl-CoA and loads it onto ACP1a within the same module. The other AT domain in module 1, AT1b, recognizes malonyl-CoA and loads it onto the second ACP, ACP1b, of module 1. The two acyl groups are condensed by KS1 to form the first diketide intermediate bound to ACP1b. Presumably, a CL function is still required to form the carboxybenzoyl-CoA starter unit, but the gene encoding this function is not part of the soraphen PKS and is not located within the soraphen A biosynthetic gene cluster described herein.

Based on the features of the sequence of this region, we propose a scheme for the biosynthesis of soraphen A shown in Fig. 4A. The product of the soraphen synthase is predicted to lack the hemiketal ring of soraphen A. The hemiketal ring could be formed by the oxidation of the hydroxyl at C3 to the ketone that would collapse to the hemiketal after attack of C3 by the hydroxyl at C7, to form soraphen B that was identified from culture extracts of So ce26. A plausible candidate to catalyze this activity is SorR. Although SorR is homologous to reductase domains of PKSs, it could conceivably function in the reverse as an oxidase to catalyze this reaction. Alternatively, a ketone at C3 would be produced if the KR domain in module 8 were not functional. Examination of the KR8 amino acid sequence did not provide any basis for concluding that this domain is nonfunctional. The second post-PKS step in the proposed synthetic scheme is the introduction of the double bond at C9–C10. A gene encoding an enzyme that might be responsible for this activity is not apparent in the soraphen A gene region. This activity could be performed by an oxidoreductase and presumably a gene encoding this type of enzyme would be readily identified by homology searches of the database. It is still possible though that one of the ORFs in the soraphen gene region with an unknown function could encode this activity. A good candidate for this is ORF4 since its location in between the *sorB* and *sorM* genes suggests that it is part of the soraphen A gene cluster. Alternatively, SorR could be responsible for the introduction of the double bond, while the role of ORF4 is the formation of a ketone at C3 as a prelude to the formation of the hemiketal ring. Since soraphen C has been identified in the culture extracts of So ce26 and it differs from soraphen A only by the hydroxyl side group at C11 versus a methoxy group in soraphen A (Fig. 4A), soraphen C is likely to be the penultimate intermediate in the synthesis of soraphen A. This suggests that the methylation of the hydroxyl at C11 is the last step in soraphen A biosynthesis. We propose that the OMT encoded by the *sorM* gene is responsible for this activity.

An interesting feature of soraphen A is the presence in two locations of methoxy side groups on carbons that are predicted to be derived from the α-carbons of their two-carbon extender units. Recent labeling studies to determine the origin of the extender units from which the oxygenated carbons C4 and C12 of soraphen A are derived suggest that they originate from glycerate or glycolate (Hill et al., 1998). Other polyketides having hydroxy or methoxy groups at α-carbons have also shown unexpected labeling results that suggest these carbons are not derived from acetate, but from metabolites such as glycolate, glycerate, or glycerol (Sakuda et al., 1998). A recent report on the characterization of the gene cluster involved in the synthesis of the polyketide FK520 proposes that one of the extender units in the synthesis of the immunosuppressant FK520 is methoxymalonyl-CoA (Wu et al., 2000). A UV-mutant strain of *S. cellulosum* So ce26 that produces soraphen O, a truncated soraphen A intermediate (Fig. 2), has been identified. The carbon in soraphen O that corresponds to C12 of soraphen A has a methoxy side group, thereby supporting the hypothesis that this methoxy group is part of the extender unit that is incorporated by the soraphen synthase and that it is not added by a post-PKS methylation. Wu et al. (2000) identified a set of genes in the FK520 gene cluster that they propose to be responsible for the synthesis of methoxymalonyl-ACP. The gene *fkbH* that has no homologies to known genes is proposed to encode an enzyme that produces glyceryl-ACP from an unknown three-carbon glycolytic intermediate. They further postulated that glyceryl-ACP is oxidized to glyceraldehyde-ACP and subsequently to hydroxymalonyl-ACP by the action of the protein products of the *fkbK* and *fkbI* genes, respectively. Since the SorD and SorE proteins are highly similar to FkbK and FkbI, respectively, we propose that the SorD and SorE proteins have similar functions and are involved in the biosynthesis of the methoxymalonyl-CoA extender units required for the biosynthesis of soraphen A (Fig. 4B). Although no homolog of *fkbH* was found in the soraphen A gene region, the AT

domain in SorC could function to load a three-carbon glycolytic metabolite onto the 4′-phosphopantetheine prosthetic group of the ACP domain of SorC, thereby representing a biochemical alternative to FkbH. It is possible that the SorD and SorE proteins oxidize the precursor molecule while it is attached to the ACP of SorC. Indeed, the *fkbKI* gene products were also proposed to act on their substrates while they are bound to the ACP encoded by *fkbJ* (Wu et al., 2000). Finally, we propose that the OMT domain of SorC completes the pathway by transferring a methyl group from a methyl donor to the hydroxyl oxygen of hydroxymalonyl-ACP_{SorC} to produce methoxymalonyl-ACP_{SorC} (Fig. 4B). There is not a TE domain present in SorC nor are there separate TE genes in the soraphen A gene cluster. Therefore, it is likely that SorC interacts directly with the ATs of modules 3 and 7 of the soraphen synthase to transfer the methoxymalonyl group from the ACP_{SorC} to the ACPs of modules 3 and 7 of the soraphen synthase. A protein similar to SorC that includes AT, ACP and OMT domains in a single enzyme has not heretofore been described and so, SorC appears to represent an unusual protein with a unique function.

Apart from furthering our understanding of the biosynthesis of soraphen, the characterization of this biosynthetic gene cluster also provides unique building blocks for combinatorial biosynthetic approaches. The starter module of the *sorA* gene could be used to initiate polyketide biosynthesis or to cap nonribosomal peptides with an unsubstituted phenyl ring, while the combined activities of SorC, D, and E in conjunction with the AT domains of the soraphen modules 3 and 7 could be exploited to direct the incorporation of methoxymalonyl extender units into 'unnatural' natural products.

Acknowledgements

We would like to thank Hans Reichenbach and Gerhard Höfle of the Gesellschaft für Biotechnologische Forschung, Braunschweig, Germany and Christopher J. Wilkinson and Peter F. Leadlay of Cambridge University, Cambridge, UK for sharing their unpublished results with us. We also thank Paul Pachlatko for critically reading the manuscript.

References

Aparicio, J.F., Molnár, I., Schwecke, T., König, A., Haydock, S.F., Khaw, L.E., Staunton, J., Leadlay, P.F., 1996. Organization of the biosynthetic gene cluster for rapamycin in *Streptomyces hygroscopicus*: analysis of the enzymatic domains in the modular polyketide synthase. Gene 169, 9–16.

August, P.R., Tang, L., Yoon, Y.J., Ning, S., Muller, R., Yu, T.W., Taylor, M., Hoffmann, D., Kim, C.G., Zhang, X., Hutchinson, C.R., Floss, H.G., 1998. Biosynthesis of the ansamycin antibiotic rifamycin: deductions from the molecular analysis of the *rif* biosynthetic gene cluster of *Amycolatopsis mediterranei* S699. Chem. Biol. 5, 69–79.

Bedorf, N., Schonburg, D., Gerth, K., Reichenbach, H., Höfle, G., 1993. Isolation and structure elucidation of soraphen $A_{1\alpha}$, a novel macrolide from *Sorangium cellulosum*. Liebigs Ann. Chem. 9, 1017–1021.

Chen, S., von Bamberg, D., Hale, V., Breuer, M., Hardt, B., Müller, R., Floss, H.G., Reynolds, K.A., Leistner, E., 1999. Biosynthesis of ansatrienin (mycotrienin) and naphthomycin. Identification and analysis of two separate biosynthetic gene clusters in *Streptomyces collinus* Tü 1892. Eur. J. Biochem. 261, 98–107.

Donadio, S., Katz, L., 1992. Organization of the enzymatic domains in the multifunctional polyketide synthase involved in erythromycin formation in *Saccharopolyspora erythraea*. Gene 111, 51–60.

Gerth, K., Bedorf, N., Irschik, H., Höfle, G., Reichenbach, H., 1994. The Soraphens: a family of novel antifungal compounds from *Sorangium cellulosum* (Myxobacteria). J. Antibiot. 47, 23–31.

Gerth, K., Bedorf, N., Höfle, G., Irschik, H., Reichenbach, H., 1996a. Epothilones A and B: antifungal and cytotoxic compounds from *Sorangium cellulosum* (Myxobacteria). J. Antibiot. 49, 560–563.

Gerth, K., Washausen, P., Höfle, G., Irschik, H., Reichenbach, H., 1996b. The jerangolids: a family of new antifungal compounds from *Sorangium cellulosum* (Myxobacteria): production, physio-chemical and biological properties of jerangolid A. J. Antibiot. 49, 71–75.

Haydock, S.F., Aparicio, J.F., Molnár, I., Schwecke, T., Khaw, L.E., König, A., Marsden, A.F.A., Galloway, I.S., Staunton, J., Leadlay, P.F., 1995. Divergent sequence motifs correlated with the substrate specificity of (methyl)malonyl-CoA: acyl carrier protein transacylase domains in modular polyketide synthase. FEBS Lett. 374, 246–248.

Hertweck, C., Moore, B.S., 2000. A plant-like biosynthesis of benzoyl-CoA in the marine bacterium '*Streptomyces maritimus*'. Tetrahedron 56, 9115–9120.

Hill, A.M., Harris, J.P., Siskos, A.P., 1998. Investigation of glycerol incorporation into soraphen A. Chem. Commun. 21, 2361–2362.

Hill, A.M., Harris, J.P., Siskos, A.P., 2000. Elucidation of soraphen A biosynthesis using ^{2}H, ^{13}C, and ^{18}O labelled precursors. Conference on Synthesis and Applications of Isotopically Labelled Compounds, Conference Proceedings.

Irschik, H., Janson, R., Gerth, K., Höfle, G., Reichenbach, H., 1986. The sorangicins: novel and powerful inhibitors of eubacterial RNA polymerase isolated from myxobacteria. J. Antibiot. 40, 7–13.

Kagan, R.M., Clarke, S., 1994. Widespread occurrence of three sequence motifs in diverse S-adenosylmethionine-dependent methyltransferases suggests a common structure for these enzymes. Arch. Biochem. Biophys. 310, 417–427.

Kakavas, S.J., Katz, L., Stazzi, D., 1997. Identification and characterization of the niddamycin polyketide synthase genes from *Streptomyces caelestis*. J. Bacteriol. 179, 7515–7522.

Maniatis, T., Fritsch, E.F., Sambrook, J., 1982. Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Molnár, I., Schupp, T., Ono, M., Zirkle, R., Milnamow, M., Nowak-Thompson, B., Engel, N., Toupet, C., Stratmann, A., Cyr, D., Görlach, J., Mayo, J., Hu, A., Goff, S., Schmid, J., Ligon, J., 2000. The biosynthetic gene cluster for the microtubule-stabilizing agents epothilones A and B from *Sorangium cellulosum* So ce90. Chem. Biol. 7, 97–109.

Motamedi, H., Shafiee, A., 1998. The biosynthetic gene cluster for the macrolactone ring of the immunosuppressant FK506. Eur. J. Biochem. 256, 528–534.

Motamedi, H., Shafiee, A., Cai, S.-J., Streicher, S.L., Arison, B.H., Miller, R.R., 1996. Characterization of methyltransferase and hydroxylase genes involved in the biosynthesis of the immunosuppressants FK506 and FK520. J. Bacteriol. 178, 5243–5248.

Reichenbach, H., Höfle, G., 1994. Discovery of a new antifungal mechanism of action: soraphen – an almost-success story. In: Gesellschaft für Biotechnologische Forschung mbH, Scientific Annual Report, Braunschweig, Germany, pp. 5–22.

Reichenbach, H., Gerth, K., Irschik, H., Kunze, B., Höfle, G., 1988. Myxobacteria, a source for new antibiotics. Trends Biotechnol. 6, 115–121.

Sakuda, S., Makoto, O., Ikeda, H., Sakurada, M., Tsukigi, H., Furihata, K., Nakayama, J., Suzuki, A., Isogai, A., 1998. Studies on the inhibitor of

aflatoxin production, aflastatin A. 9th International Congress of Pesticide Chemistry, London.

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

Schupp, T., Toupet, C., Engel, N., Goff, S., 1998. Cloning and sequence analysis of the putative rifamycin polyketide synthase gene from *Amycolatopsis mediterranei*. FEMS Microbiol. Lett. 159, 201–207.

Silakowski, B., Schairer, H.U., Ehret, H., Kunze, B., Weinig, S., Nordsiek, G., Brandt, P., Blöcker, H., Höfle, G., Beyer, S., Müller, R., 1999. New lessons for combinatorial biosynthesis from Myxobacteria. The myxothiazol biosynthetic gene cluster of *Stigmatella aurantiaca* DW4/3-1. J. Biol. Chem. 274, 37391–37399.

Vahlensieck, H.F., Pridzun, L., Reichenbach, H., Hinnen, A., 1994. Identification of the yeast ACC1 gene product (acetyl-CoA carboxylase) as the target of the polyketide fungicide soraphen A. Curr. Genet. 25, 95–100.

Witkowski, A., Naggert, J., Wessa, B., Smith, S., 1991. A catalytic role for histidine-237 in rat mammary gland thioesterase II. J. Biol. Chem. 266, 18514–18519.

Wu, K., Chung, L., Revill, W.P., Katz, L., Reeves, C.D., 2000. The FK-520 gene cluster of *Streptomyces hygroscopicus* var. *ascomyeticus* (ATCC 14891) contains genes for biosynthesis of unusual polyketide extender units. Gene 251, 81–90.

CHAPTER 4. ROLE OF GENES LOCATED UPSTEAM OF THE SORAPHEN A POLYKETIDE SYNTHASE GENES

ABSTRACT

The biosynthetic gene cluster of the antifungal polyketide soraphen A has been sequenced and characterized (Chapter 3). The core of the soraphen synthase consists of two adjacent ORFs, *sorA* and *sorB*, encoding large, multifunctional type one polyketide synthase (PKS). Genes have been identified flanking *sorA* and *sorB* that have proposed roles in the biosynthesis of the final polyketide product, soraphen A. Gene disruption has delineated the downstream border of the soraphen A PKS cluster but the upstream border of the cluster has yet to be determined. *SorC* is one of the genes upstream of soraphen PKS genes and its gene product is proposed to take part in the synthesis of the unusual extender unit methoxymalonyl-coenzyme A. A fragment of *sorC* was used to disrupt the chromosomal copy through homologous recombination and resulted in a strain of *S. cellulosum* in which soraphen A production was abolished. This result confirmed that *sorC* is required for the biosynthesis of soraphen A. A second gene designated *orf2*, lies further upstream of *sorC* and has no homologs in the databases. A fragment of *orf2* was used to disrupt the chromosomal copy and yielded a strain with no change in the production of soraphen A. This confirms that *orf2* is not required for soraphen A biosynthesis and helps define the upstream border of the soraphen PKS cluster.

INTRODUCTION

The myxobacteria are a group of gram-negative bacteria that can undergo a complex multicellular development resulting in the formation of fruiting bodies and

myxospore production. The myxobacteria are also a rich source of natural products with interesting and valuable biological activates (Reichenbach and Hofle, 1993b). While much effort in the area of natural product discovery has focused on the actinomycetes, the myxobacteria have also proven to be prolific in the number and diversity of natural products they produce. Many of the natural products isolated from myxobacteria have unique chemistries and modes of action. *S. cellulosum* has been one of the myxobacteria shown to produce a diverse array of natural products (Reichenbach, 2001). In one screening program between 1975 and 1991, 95% of the *Sorangium* strains screened produced bioactive compounds (Reichenbach and Hofle, 1993a). Three of the natural products isolated from strains of *S. cellulosum* include the polyketides epothilone (Gerth et al., 1996), sorangicin (Irschik et al., 1995b), and soraphen A (Gerth et al., 1994). Epothilone has cytotoxic and antitumor activity and is currently being tested as an anti-cancer drug. Sorangicin inhibits bacterial RNA polymerase and is effective at very low doses against gram-positive bacteria (Irschik et al., 1995a). Soraphen A has been shown to inhibit the activity of acetyl coenzyme A carboxylase in fungi and is unique in its mode of action (Vahlensieck et al., 1994). This inhibition disrupts fatty acid synthesis in a wide range of fungal species. Soraphen A was shown to be a macrolide polyketide (Gerth et al., 1994) and its biosynthetic gene cluster has been sequenced and characterized (Ligon et al., 2002).

SorC is a 2589 bp gene that lies approximately 1.5 kb upstream of the start of *sorA* and is transcribed in the opposite direction of *sorA*. The *sorC* gene contains three distinct domains; an N-terminal acyltransferase-homologous domain (AT), a C-terminal methlytransferase-homologous domain (OMT) and an acyl carrier protein-homologous domain (ACP). *SorC* has been proposed to take part in the synthesis of

methoxymalonyl-CoA, which is proposed to be an unusual extender unit required for the biosynthesis of soraphen A.

Upstream of *sorC* lies a 4701 bp ORF designated as *orf2*. Analysis of the deduced amino acid sequence of *orf2* did not identify any close matches in the database. *Orf2* has not been proposed to have a role in the biosynthesis of soraphen A. This investigation demonstrates that *sorC* is required for the production of soraphen A consistent with the hypothesis that it is involved in production of methoxymalonyl-CoA for soraphen A biosynthesis. It also demonstrated that *orf2* is not involved in the production of soraphen A, thereby defining the upstream border of the cluster to a region between *sorC* and *orf2*.

MATERIALS AND METHODS

Strains and plasmids. Strains of *S. cellulosum* and *Escherichia coli* were used as described in the "Materials and Methods" section of Chapter 2. *Candida albicans* (type strain NRRL Y-12983, Agricultural Research Service Culture Collection, Peoria, IL) is sensitive to soraphen A and was used as an indicator strain.

Media and growth conditions. Media and growth conditions for *S. cellulosum* and *E. coli* were as described in "Materials and Methods" section Chapter 2. *C. albicans* was grown at 30^{o}C in Bacto (Difco, Sparks, MD) potato dextrose broth (24g/L) or agar (1.5%).

DNA manipulations. Routine cloning, transformation, and PCR were carried out as described in the "Materials and Methods" section of Chapter 2. Restriction enzymes and T4 DNA ligase were from New England Biolabs (Beverly, MA) and were used according to manufactures instructions. Oligonucleotides were obtained

from Integrated DNA Technologies (Coralville, IA). Sequencing was done as described in the "Materials and Methods" section of Chapter 2.

Cloning of *sorC* and *orf2* fragments. Genomic DNA isolated from *S. cellulosum* was used as DNA template in a PCR reaction to amplify and clone an internal fragment of the *sorC* and *orf2* coding regions. The primers 5'-GCGGCCGCCGAATGCCGTGCTTCACTGA and 5'-GCGGCCGCTCGACATC CGCACGGTGATGCA, which contain *Not*I sites (underlined), were used to amplify a 1506 bp fragment internal to the coding region of *sorC*. The fragment was cloned into pCR-Blunt II –Topo vector (Invitogen, Carlsbad, CA) creating pTsorCko, and the sequence of the insert DNA was verified. Plasmid pTsorCko was digested with *Not*I, the *sorC* fragment was isolated on an agarose gel and the DNA was purified by QIAquick gel extraction kit (Qiagen, Valencia, CA) and ligated into a compatible, *Not*I digested pCIB132 vector. The resulting clone, pCIB132sorCko, was verified through restriction digest patterns and end sequencing of the inserted *sorC* fragment. Similarly, a fragment internal to the coding region of *orf2* was PCR amplified with the primers 5'-GCGGCCGCACATCACGGCTGCCAGCGCGTATCC and 5'-GCGGCCGCATCAGATCGTCGCCCCGTGGCT to generate a 1995 bp fragment that was cloned into pCR-Blunt-Topo to generate pTorf2ko. The *orf2* fragment was isolated and cloned into pCIB132 to create pCIB132orf2ko as described above. These plasmids were transformed into the conjugative strain of *E. coli*, ET 12567 (pUZ8002) for conjugation into *S. cellulosum*.

Conjugations. Introduction of plasmids pCIB132sorCko and pCIB132orf2ko into *S. cellulosum* by conjugation and selection of transconjugates were carried out as described in the "Methods and Materials" section of Chapter 2.

Gene disruptions. Genomic DNA isolated from putative transconjugates from pCIB132sorCko and pCIB132orf2ko were probed with ^{32}P-labeled fragments derived from pCIB132. These fragments were amplified by PCR from pCIB132 and contain regions that flank the *Not*I cloning site. One of these DNA fragments contained the kanamycin resistance gene from pCIB132. The primers used were 5'-TGATTGAACAAGATGGATG and 5'-AGAAGAACTCGTCAAGAAGG and the 792 bp product was isolated and ^{32}P-labeled as described in "Material and Methods" in Chapter 2. The primers 5'-AGGCCATGTTTGACAGCTT and 5'-CTCCTTTTCATCCGCATCG were used to amplify a second DNA fragment, flanking the *Not*I cloning site from pCIB132, and was also labelled with ^{32}P. The genomic DNA from the strains of *S. cellulosum* was isolated and hybridized as described in "Materials and Methods" in Chapter 2.

Bioassay for soraphen A production. The detection of soraphen A was carried out by bioassay. Strains of *S. cellulosum* were grown 12 days in G51t liquid medium. Cultures of the transconjugates were supplemented with 25 ug/mL of phleomycin. After 12 days of growth, 100 uL of culture was spotted onto PDA plates and allowed to dry. An overnight culture of *C. albicins* was used to seed (1% v/v) potato dextrose soft agar (PDA, 0.7% agar) at 55^{o}C. The *C. albicans* cells and soft PDA were mixed and 5mL per plate were used to overlay the PDA plates containing the dried *S. cellulosum* cultures. After the overlay solidified, the plates were incubated at 30^{o}C for 3 days and observed for the inhibition of growth of *C. albicans*. G51t supplemented with 25 ug/ml of phleomycin does not inhibit the growth of *C. albicans* (data not shown).

***S. cellulosum* extracts**. To further analyze strains of *S. cellulosum* for the production of soraphen A, culture extracts were prepared and analyzed. 250 mL of G51t was inoculated with 25 mLs of a 5-day culture of *S. cellulosum* and incubated in 1L-baffled flasks at 30°C with shaking at 225 rpm. After 12 days, the entire culture was freeze dried and approximately 100 mL of methanol was added to the dried material and slowly mixed overnight. The entire extract was then centrifuged at 9000 g for 30 minutes and the methanol removed. The methanol was dried in a SpeedVac centrifuge and redissolved in 5 mLs of methanol. Extracts were stored at –20°C.

HPLC and LC/MS analysis. Extracts from strains of *S. cellulosum* were analyzed by reverse phase high-performance liquid chromatography (HPLC) and liquid chromatography-mass spectrometry (LC/MS). HPLC was performed on 25 µl of extract on a HP1100 series HPLC machine (Agilent Technologies, Palo Alto, CA) fitted with a Synergi 4 micron MAX-PR 80Å 150x4.6 mm column (Phenomenex Torrance, CA) and a diode array detector (DAD). Mobile phase A was water with 0.01% trifluoroacetic acid (TFA) and mobile phase B was acetonitrile with 0.0075% TFA. Elution was performed as follows: 25% B to 85% B in 15 minutes at 40°C with a flow rate of 0.8 ml/minute. Soraphen A was detected by its absorption at 210 nm at a retention time of 14 minutes.

Extracts were also analyzed by LC/MS. LC analysis of 100 µl of extract was carried out on a HP1100 HPLC (Agilent Technologies, Palo Alto, CA) with the column described above. A 25% to 85% gradient of methanol, in water, was used for elution in the presence of 0.1% sodium formate and 0.1% formic acid, with a flow rate of 0.4 mL/min. The effluent from the LC was directed to a Finnigan LCQ Classic Mass Spectrometer (ThermoFinnigan, San Jose, CA). The analyses were done using

positive mode electrospray ionisation (ESI) (capillary temperature = 275, capillary voltage = 3.22 V, and spray needle voltage = 4.54 kV).

RESULTS

Generation of strains containing gene disruptions. The genes *orf2* and *sorC* lie upstream of the soraphen PKS genes *sorA* and *sorB*, and were targeted for gene disruption (Figure 1). Fragments of *sorC* and *orf2* were PCR amplified from *S. cellulosum* genomic DNA, cloned, and sequenced. The fragments of *sorC* and *orf2*, respectively, were then cloned into the mobliziable plasmid pCIB132 for conjugation into *S. cellulosum*. The gene disruptions were carried out based on a method developed for *S. cellulosum* (Jaoua et al., 1992) with some modifications as described in the "Materials and Methods" in Chapter 2. It has been shown that transconjugtes from pCIB132-derived plasmids were maintained in *S. cellulosum* only after homologous recombination between the cloned insert and the homologous region on the chromosome (Jaoua et al., 1992). After the conjugations, phleomycin resistant transconjugates were isolated and genomic DNA from these were prepared. The genomic DNA was digested and probed in a Southern hybridization with ^{32}P-label probes derived from the pCIB132 plasmid (Figures 2 and 3). Due to cross hybridization of probes derived from *sorC* and *orf2*, portions of the pCIB132 vector were used to confirm the integration of these plasmid. Southern hybridizations with a probe derived from the kanamycin resistance gene of pCIB132, produced the expected size DNA fragments encompassing the upstream border of the integration event, verifying the localization of the integration events at both *sorC* or *orf2* respectively. Similar Southern hybridizations probing the DNA region downstream of the integration events also produced DNA fragments of the expected size, again

confirming the localization of the integrations at *sorC* or *orf2*, respectively (data not shown).

Soraphen A assays. *C. albicans* is sensitive to soraphen A. To determine the production of soraphen A by strains of *S. cellulosum*, a *C. albicans* bioassay was used. *S. cellulosum* strains were grown for 12 days in G51t and 100 uL of cultures were spotted on PDA plates (1.5% agar) and allowed to dry. An overnight culture of *C. albicans* was used to seed tempered soft PDA and overlaid onto the plates with the *S. cellulosum* cultures spots. The plates were incubated for 3 days at 30^{o}C and observed for zones of inhibition. Strain SJ3 displays a large zone of inhibition in the bioassay due to the production of soraphen A (Figure 4). Strains of *S. cellulosum*, in which homologous recombination-mediated integration disrupted *sorC* displayed no zone of inhibition (Figure 4). These results are similar to results seen in strains with disruptions in *sorA* or *sorB* (data not shown). In strains of *S. cellulosum* in which *orf2* was insertionally inactivated, no difference in zones of inhibition as was observed compared to strain SJ3.

To further analyze the extracts, HPLC was performed with 25 uL of each extract and typical results shown in Figure 5. Extracts from strain SJ3 as well as *orf2* mutant strains displayed a predominant peak that co-migrates with the peak from the soraphen A standard at 14 minutes. Extracts from *sorC* mutant strains display no peak at the retention time for soraphen A. The extracts were further analyzed by LC/MS. The soraphen A standard displayed an UV absorption peak at 36 minutes corresponding to the extracted soraphen A mass ion (Figure 6). Conversely, the mass spectrum of the peak at 36 minutes displayed the expected mass for soraphen A (Figure 6). The extracts from SJ3, as well as the *orf2* disruption strain displayed a

peak at 36 minutes corresponding to the extracted soraphen A mass ion (Figure 7) and also the expected mass spectra corresponding to the UV absorption peak at 36 minutes (Figure 8). These results are consistent with the presence of soraphen A in the samples as indicated by the bioassay and HPLC data. The extract from the strain with the *sorC* disruption had no UV absorption peak at 36 minutes corresponding to the extracted soraphen A mass ion (Figure 7). Similarly, mass spectra consistent with soraphen A production were undetectable in the 36 minute area of the chromatogram, confirming the results seen in the bioassay and HPLC data.

DISCUSSION

The soraphen A gene cluster has been sequenced and characterized, and gene disruption was used to demonstrate the involvement of the PKS genes *sorA* and *sorB* in soraphen A biosynthesis. The genes flanking the PKS genes *sorA* and *sorB* have also been postulated to have a role in soraphen A biosynthesis but no experimental evidence has been supplied to support this hypothesis. One of these genes is *sorC*. The analysis of the domains and conserved residues of *sorC* are described in the "Results" section of Chapter 3. The "Discussion" section in Chapter 3 summarizes the evidence and rationale supporting the hypothesis that *sorC* is involved in the production of methoxymalonyl-CoA, but there has been no direct evidence available for *sorC* playing a role in soraphen A biosynthesis. The data presented here show that a functional *sorC* is required for the biosynthesis of soraphen A. This result in itself does not provide any information on the exact role of *sorC* and its involvement in the synthesis of methoxymalonyl-CoA. It is also possible that abolishment of soraphen A production is due to polar effects exerted by the *sorC* disruption on *sorD* and *sorE*,

genes downstream of *sorC*. This is not likely though, since *sorC*, *sorD* and *sorE* appear to be an operon and genes in operons tend to have similar functions.

The utilization of methoxymalonate as an extender unit is not common in type I polyketide biosynthesis. Polyketide synthesis involves chain extension of an acyl starter unit by the addition of extender units, usually malonyl-CoA or methylmalonyl-CoA (reviewed by Hopwood and Sherman, 1990; Khosla et al., 1999). In rare cases, labelling experiments and positioning of hydroxyl or methoxyl groups in the final polyketide, have suggested the utilization of another type of extender unit not derived from acetate, but rather from glucose or glycerol (Haber et al., 1977; Omura et al., 1983; Hill et al., 1998; Ono et al., 1998). Such "glycolate" units are seen in ansamitocin, geldanamycin (Haber et al., 1977), leucomycin (Omura et al., 1983), FK520 and FK506 (Byrne et al., 1993), as well as in soraphen A (Hill et al., 1998). The genes proposed to be involved in the biosynthesis of the "glycolate" extender in ansamitocin, *asm*13-17, were disrupted and the resulting strains showed no ansamitocin production (Yu et al., 2002). This group of genes, *asm*13-17 show high levels of homology to the gene cluster *fkb*GHIJK, proposed to be involved in the biosynthesis of the methoxylmalonate extender unit in FK520 (Wu et al., 2000; Yu et al., 2002). Recently, *asm*13-17 were co-expressed in *Streptomyces lividans* with a cassette of *ery*ABC encoding the erythromycin PKS, in which one of the AT domains had been replaced with the hydroxymalonate AT from the FK520 PKS (Kato et al., 2002). The engineered *St. lividans* strain produced a 6-deoxyerythronolide derivative with a methoxymalonate incorporated in the predicted area of the molecule (Kato et al., 2002). While *sorD* and *sorE* are homologous to genes in the *fbk* and *asm* gene clusters, and probably carry out similar functions in the biosynthesis of

methoxymalonate, *sorC* is unique due to the AT, ACP, and OMT domains residing on a single enzyme.

The probable participation of *sorC*, *sorD* and *sorE* in the biosynthesis of the extender unit methoxymalonyl-CoA is also of interest due to the desire for new building blocks for combinatorial biosynthesis of novel polyketides. Since polyketides have been well established as bioactive molecules, combinatorial biosynthesis strives to generate new, unnatural polyketides, through the "reprogramming" of PKSs. One of the issues in this reprogramming is the availability of unique extender units to generate greater diversity in a combinatorial biosynthesis approach.

The upstream border of the soraphen A biosynthetic gene cluster was previously undetermined. A large number of type I PKS clusters have been characterized in the literature leading to the ability to predict how PKS-associated domains and motifs relate to the final polyketide structure. While the analysis of DNA sequence of the soraphen A gene cluster suggests that no genes upstream of *sorE* may be involved in soraphen A biosynthesis (Chapter 3), no experimental evidence existed to support this conclusion. However, the experiments described here show that the production of soraphen A in strains containing the *orf2* disruption mutation is similar to that of strain SJ3, thereby indicating *orf2* has no apparent role in soraphen A biosynthesis.

Establishment of the extent of the soraphen A biosynthetic gene cluster is also important for devising heterologous expressions strategies. The right border of the soraphen A cluster had been previously established. The data presented here shows that *sorC* is required for the biosynthesis of soraphen A, but not *orf2*. This

demonstrates that the left border of the cluster does not extend as far as *orf2*, but lies between *orf2* and *sorC*.

ACKNOWLEDGMENTS

I would like to thank István Molnár for critical reading of this manuscript and helpful discussions. I would like to thank members of the Syngenta Analytical Chemistry and Sequencing groups for their assistance with analysis of samples; James Pomes, Belle Abrera, Lina Summerfelt, Makoto Ono, An Hu and Racella McNair.

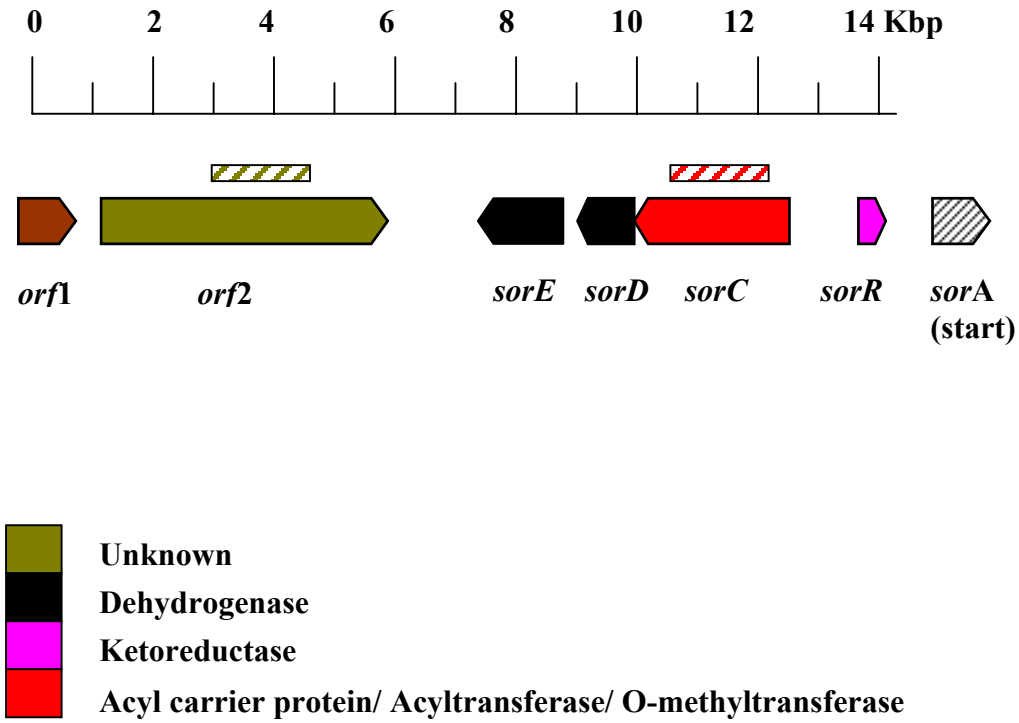
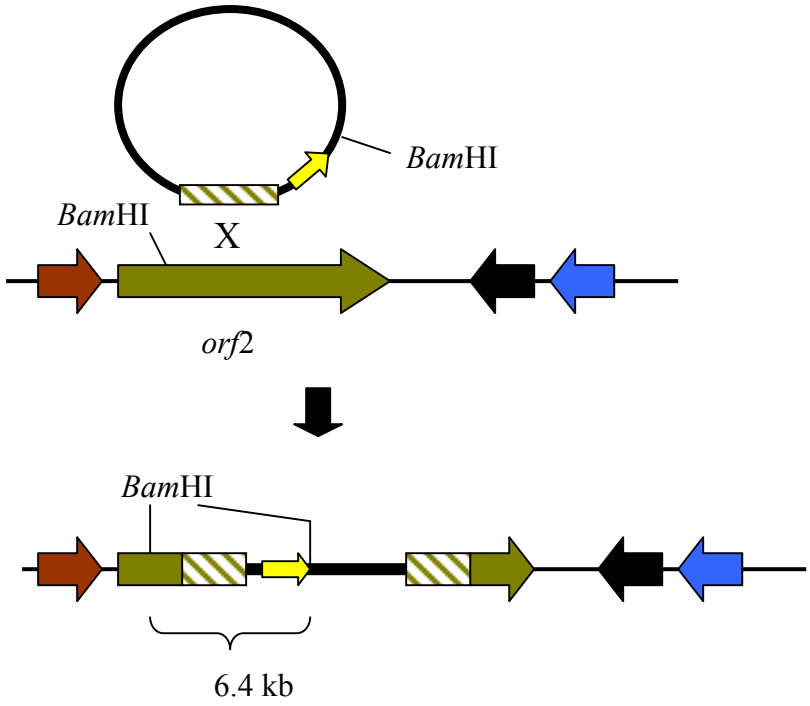


Figure 1. Organization of the upstream region of the soraphen A biosynthetic gene cluster. The orientation of the arrows indicates the direction of each gene and ORF. The black and white hatched box indicates the start of *sorA*. The hatched bars above *sorC* and *orf2* indicate the regions that were used to generate the homologous-recombination mediated gene disruptions.

A.



B.

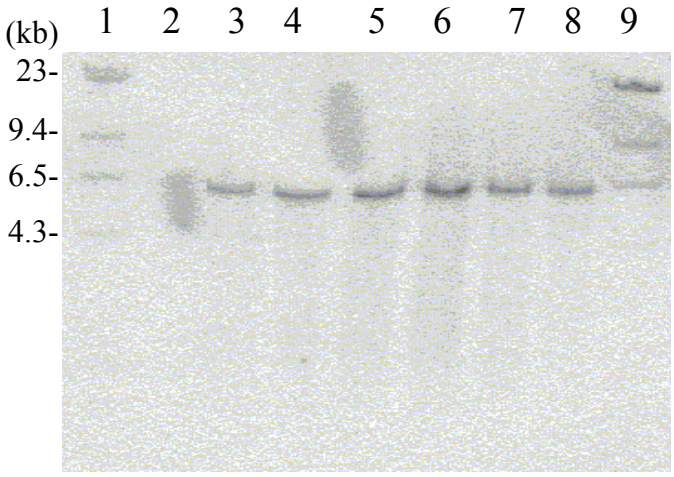
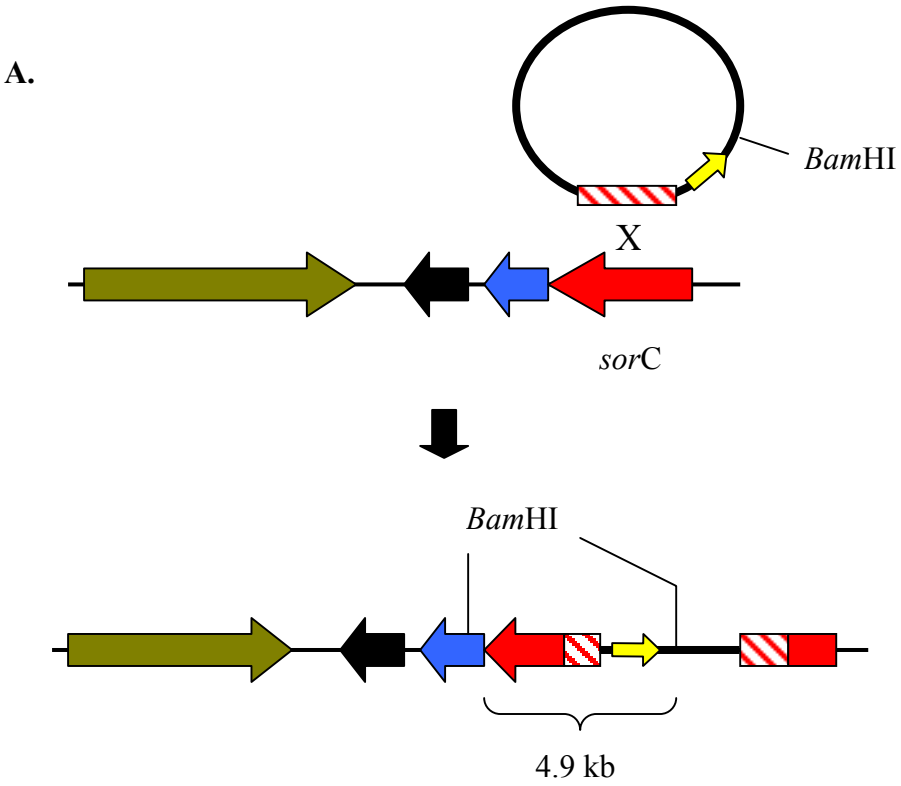


Figure 2. *Orf2* gene disruptions. (A): Diagram of the integration event disrupting *orf2*. The hatched box represents the homologous DNA fragment cloned into pCIB132. The kanamycin resistance gene is represented by the yellow arrow. (B): Southern hybridization of genomic DNA from *S. cellulosum* strains probed with ^{32}P labeled kanamycin resistance gene. All genomic DNA was digested with *Bam*HI. Lane 2 contains genomic DNA from *S. cellulosum* So ce 26 wild-type strain SJ3. Lanes 3-8 contain genomic DNA from six *orf2* disruption strains. DNA markers in lanes 1 and 9 are lambda DNA digested with *Hind*III.



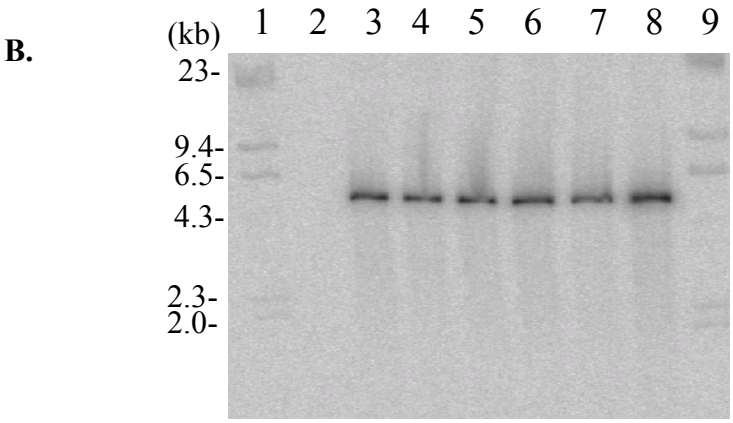


Figure 3. *SorC* gene disruptions. (A): Diagram of the integration event disrupting *sorC*. The hatched box represents the homologous DNA fragment cloned into pCIB132. The kanamycin resistance gene is represented by the yellow arrow. (B): Southern hybridization of genomic DNA from *S. cellulosum* strains probed with ^{32}P labeled kanamycin resistance gene. All genomic DNA was digested with *Bam*HI. Lane 2 contains genomic DNA from *S. cellulosum* So ce 26 wild-type strain SJ3. Lanes 3-8 contain genomic DNA from six *sorC* disruption strains. DNA markers in lanes 1 and 9 are lambda DNA digested with *Hind*III.

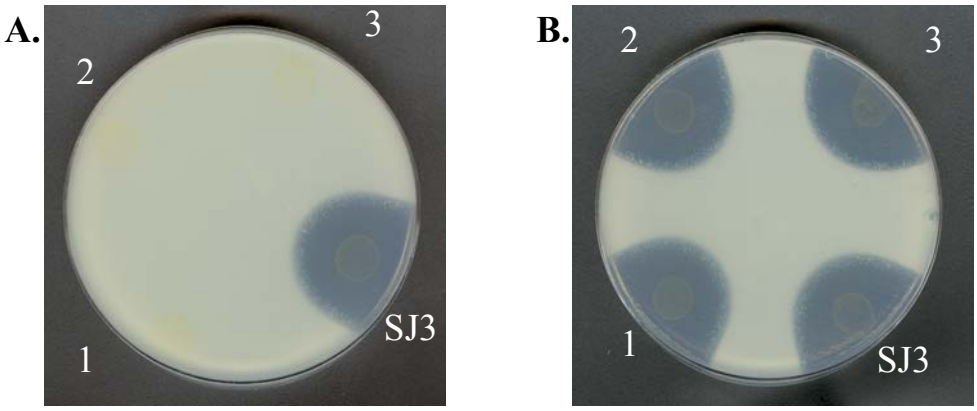


Figure 4. Soraphen A bioassay. From 12-day broth cultures of *S. cellulosum*, 100 ul of culture was spotted and dried on plates of PDA. *C. albicans* was seeded into soft PDA and overlaid on the plates. The plates were incubated for 3 days at 30^{o}C. (A): *S. cellulosum* strain SJ3 and the zone of inhibition due to soraphen A in bottom right portion of the plate. Three isolates of *S. cellulosum* mutant strains with *sorC* disruptions and abolishment of zone of inhibition and soraphen A production. (B): Strain SJ3 zone of inhibition shown with three mutant strains disrupted in *orf2*.

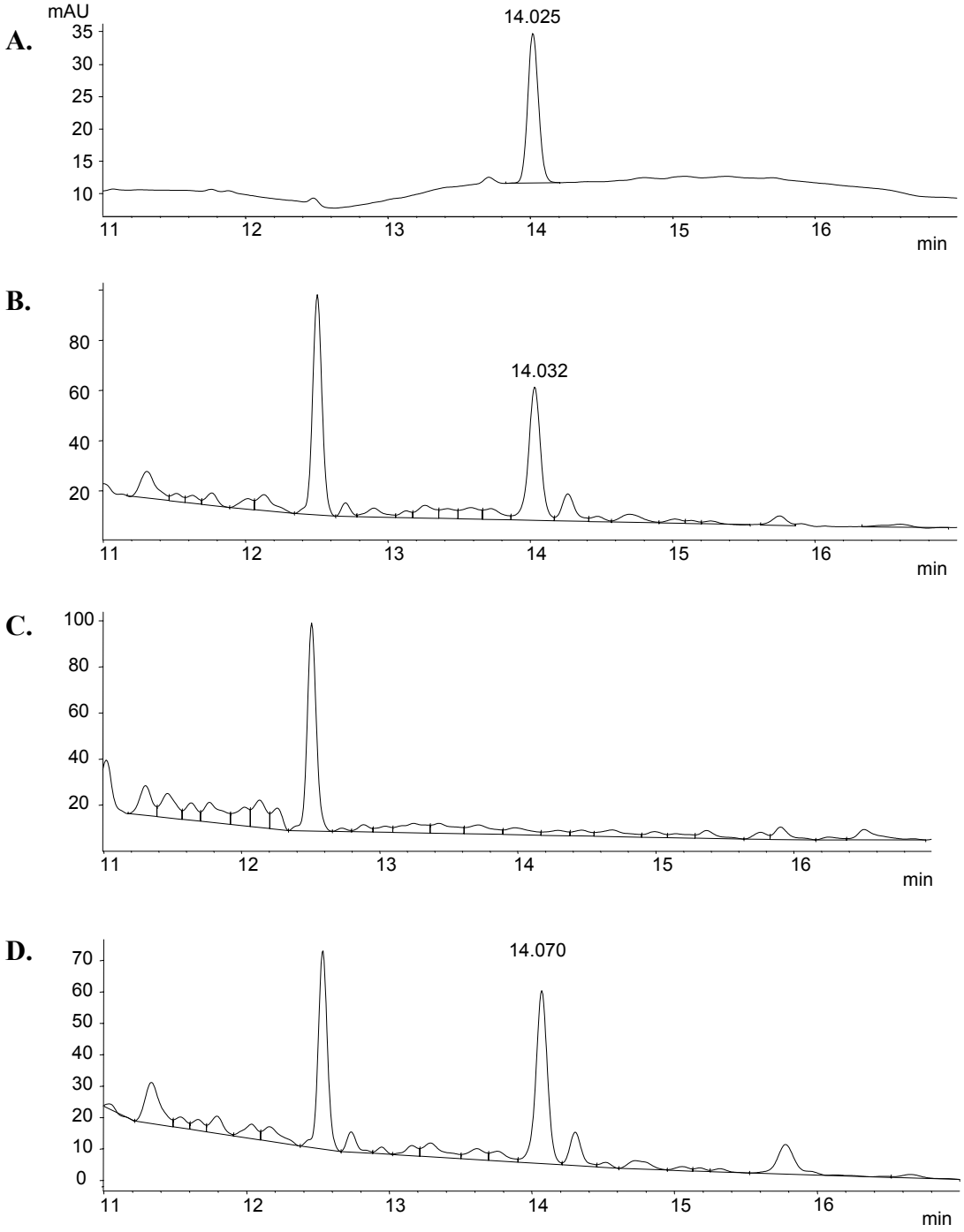


Figure 5. HPLC chromatograms. (A): Soraphen A standard (retention time of 14 minutes) (B): extract from *S. cellulosum* wild-type strain SJ3, (C): extract of mutant strain disrupted in *sorC*, (D): mutant strain disrupted in *orf2*.

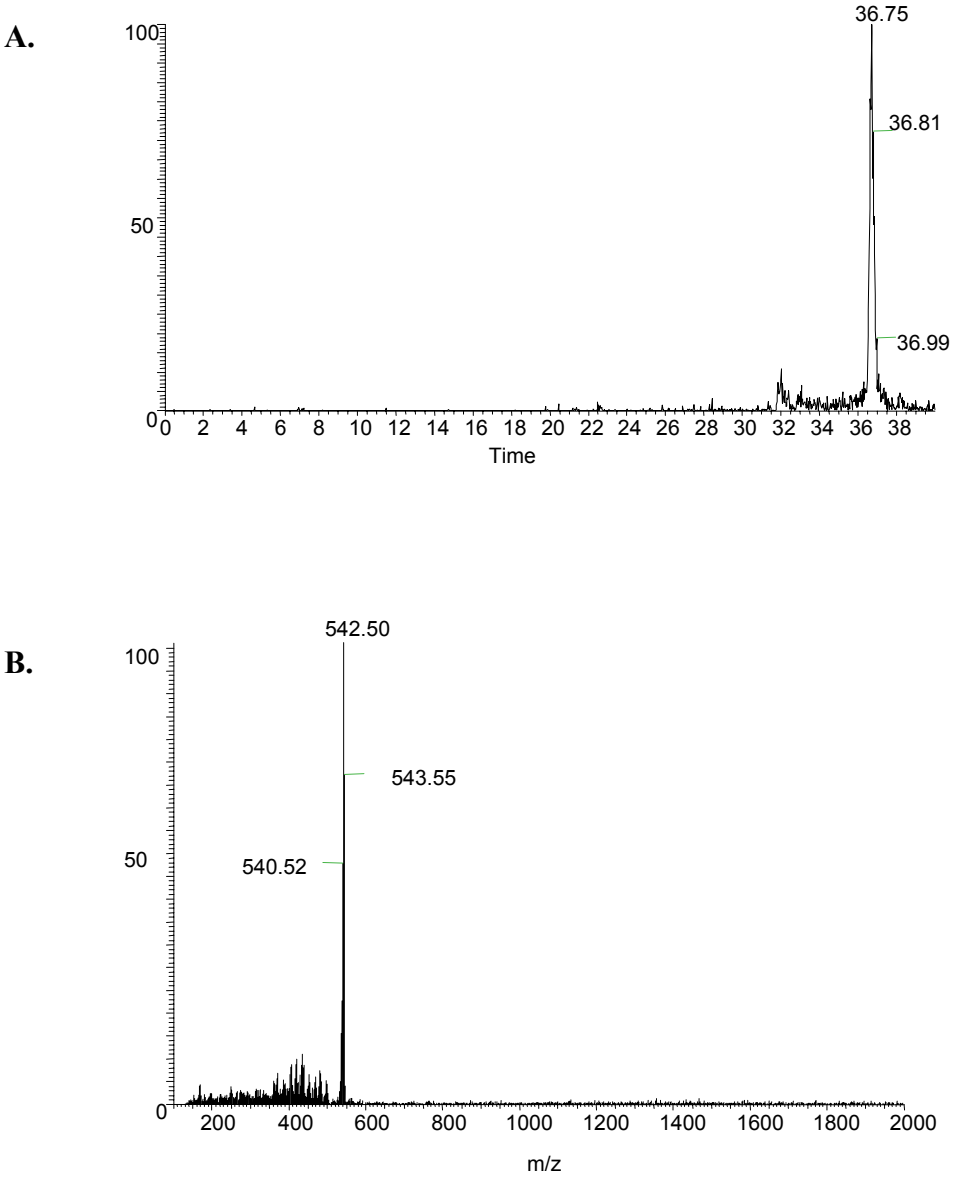


Figure 6. LC/MS chromatograms from soraphen A standard. Relative abundance is shown on the x-axis. (A): The chomatogram of the extracted mass ion (m/z=542-545) from the LC data showing the soraphen A peak at 36.7 minutes, (B): the mass spectra of the soraphen A peak in (A) showing the mass of soraphen A (M.W. of 520 + 23 for Na adduct).

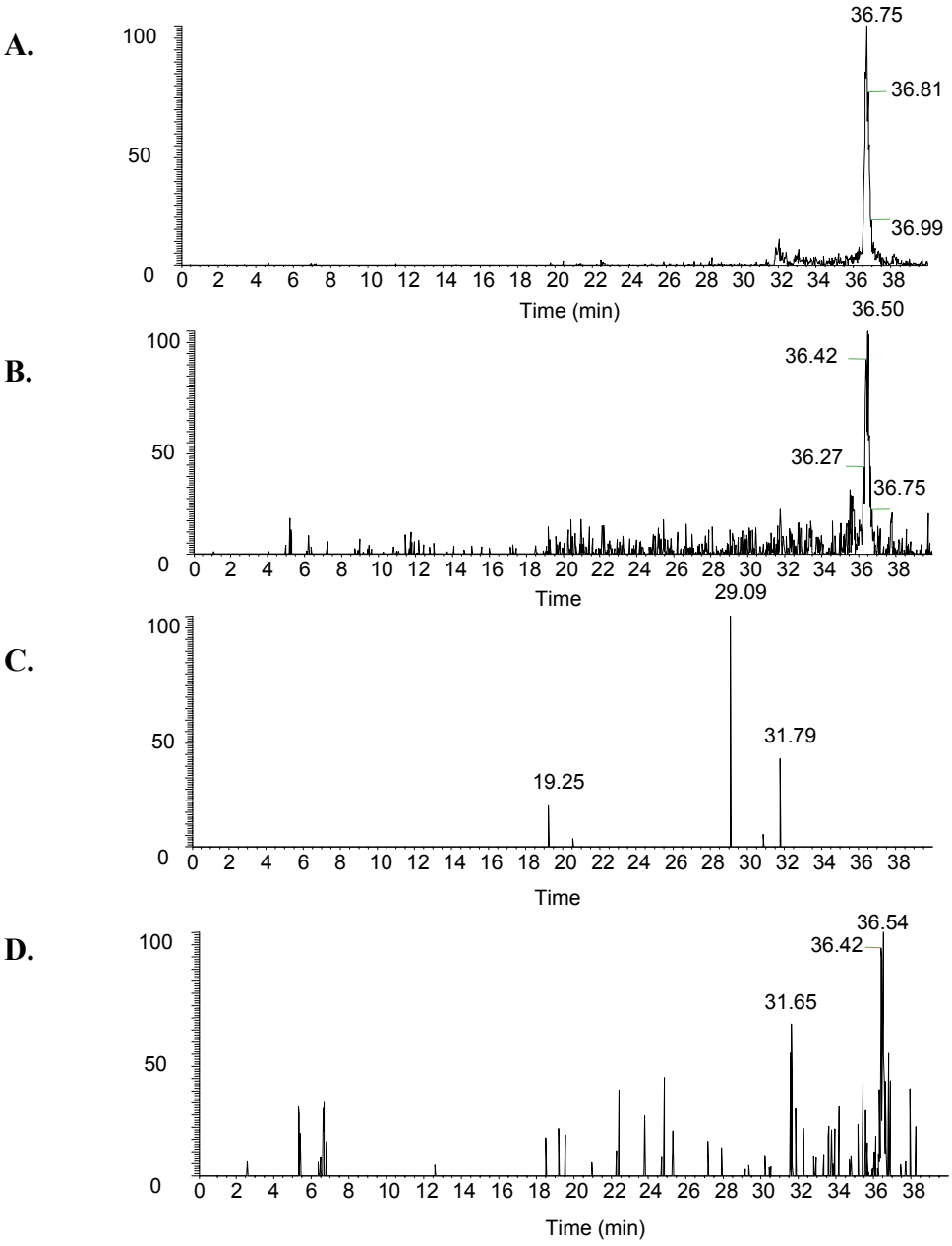


Figure 7. The soraphen A extracted mass ion (m/z 542-545) chromatograms from LC data. Relative abundance is shown on the x-axis. (A): Soraphen A standard, (B): extract of *S. cellulosum* So ce26 wild-type strain SJ3 with mass ion peak at ~36.5 minutes, (C): extract from a strain disrupted in *sorC* with no peak at the soraphen A retention time, (D): extract from a strain disrupted in *orf2* showing a peak at 36.5 minutes

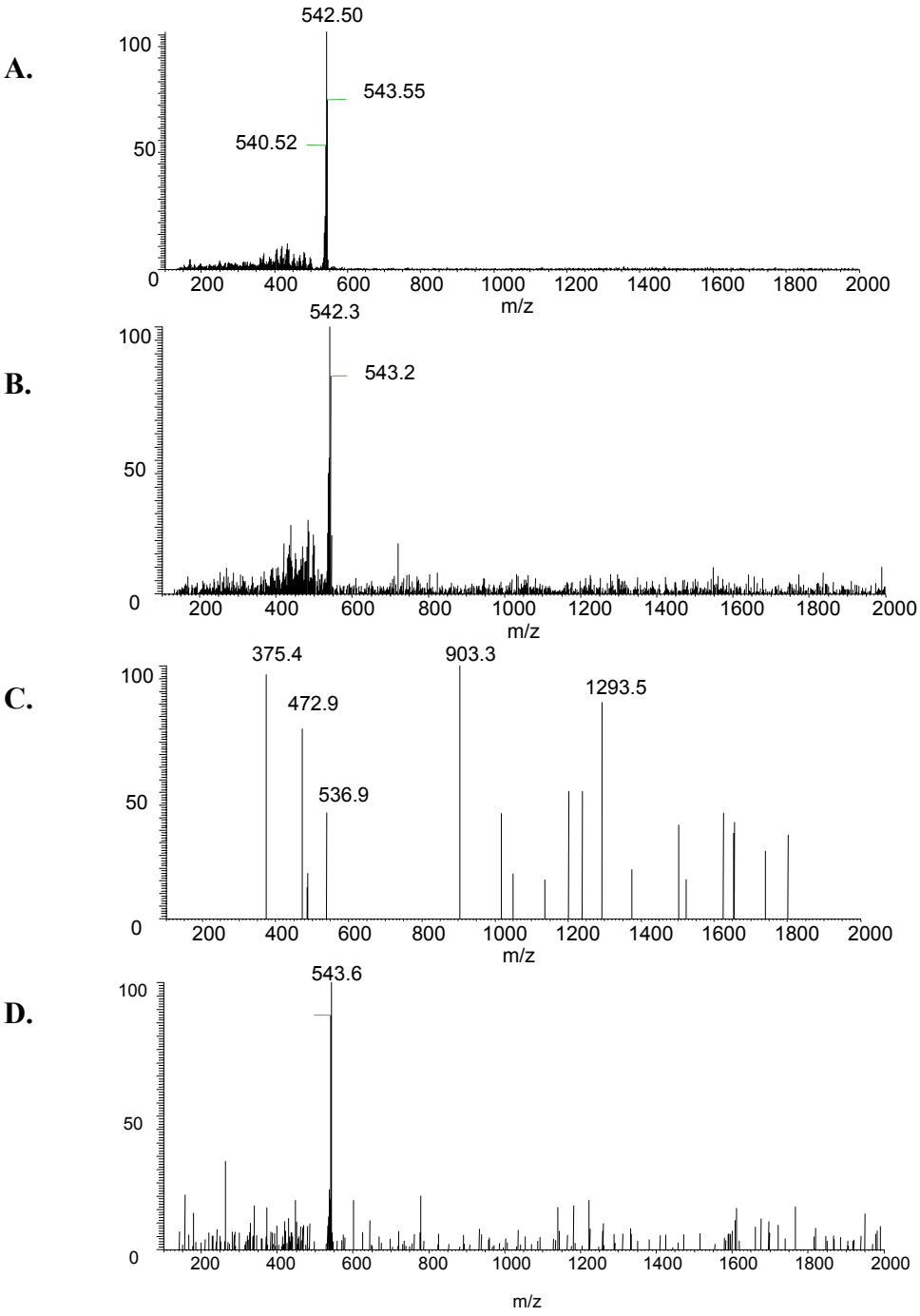


Figure 8. The mass spectra of peaks (with a retention time ~36 minutes). Relative abundance is on the x-axis. (A) The soraphen A standard, (B): extract of *S. cellulosum* So ce 26 wild-type strain SJ3 showing soraphen A (M.W. of 520 + 23 for Na adduct). (C): extract of *sorC* disruption strains with no detectable mass for soraphen A. (D): extract of *orf2* disruption strain showing the mass of soraphen A.

REFERENCES

Byrne, K.M., Shafiee, A., Nielsen, J.B., Arison, B., Monaghan, R.L. and Kaplan, L.: The Biosynthesis and Enzymology of an Immunosuppressant, Immunomycin, Produced by *Streptomyces hygroscopicus* var. *ascomyceticus*. In: R., C. (Ed.), Developments in Industrial Microbiology: Microbial Metabolites. McGraw-Hill Companies, Columbus, OH, 1993, pp. 29-45.

Gerth, K., Bedorf, N., Hofle, G., Irschik, H. and Reichenbach, H.: Epothilons A and B: antifungal and cytotoxic compounds from *Sorangium cellulosum* (Myxobacteria). Production, physico-chemical and biological properties. J Antibiot (Tokyo) 49 (1996) 560-3.

Gerth, K., Bedorf, N., Irschik, H., Hofle, G. and Reichenbach, H.: The soraphens: a family of novel antifungal compounds from *Sorangium cellulosum* (Myxobacteria). I. Soraphen A1 alpha: fermentation, isolation, biological properties. J Antibiot (Tokyo) 47 (1994) 23-31.

Haber, A., Johnson, R.D. and Rinehart, K.L., Jr.: Biosynthetic origin of the C2 units of geldanamycin and distribution of label from D-[6-13C]glucose. J Am Chem Soc 99 (1977) 3541-4.

Hill, A.M., Harris, J.P. and Siskos, A.P.: Investigation of glycerol incorporation into soraphen A. Chemical Communications (Cambridge) (1998) 2361-2362.

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

Irschik, H., Augustiniak, H., Gerth, K., Hofle, G. and Reichenbach, H.: The ripostatins, novel inhibitors of eubacterial RNA polymerase isolated from myxobacteria. J Antibiot (Tokyo) 48 (1995a) 787-92.

Irschik, H., Jansen, R., Gerth, K., Hofle, G. and Reichenbach, H.: Disorazol A, an efficient inhibitor of eukaryotic organisms isolated from myxobacteria. J Antibiot (Tokyo) 48 (1995b) 31-5.

Jaoua, S., Neff, S. and Schupp, T.: Transfer of mobilizable plasmids to *Sorangium cellulosum* and evidence for their integration into the chromosome. Plasmid 28 (1992) 157-65.

Kato, Y., Bai, L., Xue, Q., Revill, W.P., Yu, T.W. and Floss, H.G.: Functional expression of genes involved in the biosynthesis of the novel polyketide chain extension unit, methoxymalonyl-acyl carrier protein, and engineered biosynthesis of 2-desmethyl-2-methoxy-6-deoxyerythronolide B. J Am Chem Soc 124 (2002) 5268-9.

Khosla, C., Gokhale, R.S., Jacobsen, J.R. and Cane, D.E.: Tolerance and specificity of polyketide synthases. Annu Rev Biochem 68 (1999) 219-53.

Ligon, J., Hill, S., Beck, J., Zirkle, R., Molnar, I., Zawodny, J., Money, S. and Schupp, T.: Characterization of the biosynthetic gene cluster for the antifungal polyketide soraphen A from *Sorangium cellulosum* So ce26. Gene 285 (2002) 257-67.

Omura, S., Tsuzuki, K., Nakagawa, A. and Lukacs, G.: Biosynthetic origin of carbons 3 and 4 of leucomycin aglycone. J Antibiot (Tokyo) 36 (1983) 611-3.

Ono, M., Sakuda, S., Ikeda, H., Furihata, K., Nakayama, J., Suzuki, A. and Isogai, A.: Structures and biosynthesis of aflastatins: novel inhibitors of aflatoxin production by *Aspergillus parasiticus*. J Antibiot (Tokyo) 51 (1998) 1019-28.

Reichenbach, H.: Myxobacteria, producers of novel bioactive substances. J Ind Microbiol Biotechnol 27 (2001) 149-56.

Reichenbach, H. and Hofle, G.: New secondary metabolites from myxobacteria. Developments in Industrial Microbiology Series 33 (1993a) 27-34.

Reichenbach, H. and Hofle, G.: Production of Bioactive Secondary Metabolites. In: Dworkin, M. and Kaiser, A.D. (Eds.), Myxobacteria II. American Society for Microbiology, Washington D.C., 1993b, pp. 347-397.

Vahlensieck, H.F., Pridzun, L., Reichenbach, H. and Hinnen, A.: Identification of the yeast ACC1 gene product (acetyl-CoA carboxylase) as the target of the polyketide fungicide soraphen A. Curr Genet 25 (1994) 95-100.

Wu, K., Chung, L., Revill, W.P., Katz, L. and Reeves, C.D.: The FK520 gene cluster of *Streptomyces hygroscopicus* var. ascomyceticus (ATCC 14891) contains genes for biosynthesis of unusual polyketide extender units. Gene 251 (2000) 81-90.

Yu, T.W., Bai, L., Clade, D., Hoffmann, D., Toelzer, S., Trinh, K.Q., Xu, J., Moss, S.J., Leistner, E. and Floss, H.G.: The biosynthetic gene cluster of the maytansinoid antitumor agent ansamitocin from *Actinosynnema pretiosum*. Proc Natl Acad Sci U S A 99 (2002) 7968-73.

CHAPTER 5. HETEROLOGOUS EXPRESSION OF SORAPHEN A IN *STREPTOMYCES LIVIDANS*.

ABSTRACT

Soraphen A is an antifungal polyketide produced by the myxobacterium *Sorangium cellulosum* So ce26. *S. cellulosum* grows slowly, and its genetic manipulations, strain development and large-scale fermentations are problematic. Consequently, a better host for soraphen A production was sought. The soraphen polyketide synthase genes, *sorA* and *sorB*, along with the methyltransferase *sorM*, were cloned into expression vectors and integrated into the chromosome of *Streptomyces lividans* ZX7. The genes upstream of *sorA* and a benzoate-coenzyme A ligase from *Rhodopseudomonas palustris* were cloned into a *Streptomyces* expression plasmid and transformed into the *St. lividans* ZX7 strain containing the chromosomally integrated *sorA*, *sorB* and *sorM* genes. The resulting strains produced soraphen A when putative precursors for the production of the benzoyl-coenzyme A starter unit, sodium benzoate or trans-cinnamic acid, were supplied to the fermentations.

INTRODUCTION

Soraphen A is a macrolide polyketide that has an unique mode of action in the inhibition of fungal acetyl-CoA carboxylase (Gerth et al., 1994; Vahlensieck et al., 1994) thereby blocking the synthesis of fatty acids in many fungi. The activity of soraphen A was of considerable commercial interest due to its activity against plant pathogenic fungi, until testing showed it to be a weak teratogen.

Soraphen A is produced by the myxobacterium *S. cellulosum* So ce26 and the gene cluster responsible for its biosynthesis has been sequenced and characterized (Chapter 3). The major obstacle to investigations into the biosynthesis of soraphen A

is the difficulty in molecular and microbiological manipulations and few genetic tools in *S. cellulosum* So ce26. The genetic manipulation of *S. cellulosum* is extremely difficult. No plasmids have been found to replicate in *S. cellulosum* strains. Stable maintenace of transfered DNA in *S. cellulosum* relies on homologous recombination between cloned DNA fragments, introduced by conjugation on a transmissible plasmid, and the homologous chromosomal locus (Jaoua et al., 1992). This procedure limits the scope of experimentation that can be done in *S. cellulosum* and is very inefficient. Phleomycin resistance has been the only effective selection marker described in *S. cellulosum* until a very recent report on the use of a hygromycin resistance marker in *S. cellulosum* So ce56 (Pradella et al., 2002). *S. cellulosum* So ce26 also does not form colonies when spread on agar surfaces and grows relativity slowly (12-16 hour generation time) requiring specialized media. These important obstacles limit experimentation on the molecular genetics of soraphen A biosynthesis and hinder the production of altered soraphen polyketide derviatives in *S. cellulosum*.

A more amenable host was sought for the heterologous expression of the soraphen A biosynthetic genes and *St. lividans* ZX7 was chosen. *St. lividans* ZX7 has well developed genetic systems that are easy to use, it produces polyketides and other secondary metabolites, and it has been used to heterologously express other type I polyketides (Stanzak et al., 1986; Xue et al., 1999; Ziermann and Betlach, 1999; Kieser et al., 2000; Shah et al., 2000; Tang et al., 2000a). In this study, the soraphen PKS genes *sorA* and *sorB*, and the methytransferase *sorM* were cloned into integrative *Streptomyces* expression vectors, transformed into *St. lividans* ZX7 and integrated into the chromosome. The genes upstream of *sorA*; *sorR*, *sorC*, *sorD* and *sorE* that are proposed to be involved in soraphen A production, were cloned on a

Streptomyces expression plasmid along with the *badA* gene. *BadA* is a benzoate-coenzyme A ligase from *Rhodopseudomonas palustris* (Egland et al., 1995). This expression plasmid was then transformed into *St. lividans* strains containing the integrated soraphen polyketide synthase genes. The resulting strains produced soraphen A when putative precursors for the production of the benzoyl-coenzyme A starter unit, sodium benzoate or trans-cinnamic acid, were supplied to the fermentations.

MATERIALS AND METHODS

Bacterial strains. Bacterial strains and plasmids used in this study are shown in Table 1. *Escherichia coli* strain DH10B and SURE (Stratagene La Jolla, CA) was used for plasmid construction. *Streptomyces lividans* ZX7 (John Innes Centre, Norwich, UK) is a derivative of *St. lividans* 66.

DNA manipulations. Routine cloning and transformation with *E. coli* was carried out as described in the "Materials and Methods" section of Chapter 2. Restriction enzymes and T4 DNA ligase were from New England Biolabs (Beverly, MA) and used according to manufacturer's instructions. PCR reactions for purposes of cloning DNA were carried out with *Pfu* polymerase (Stratagene, La Jolla, CA) as described in the "Materials and Methods" section of Chapter 2. PCR reactions for the analysis of strains was done with Herculase polymerase (Stratagene, La Jolla, Ca) according to the manufacture's instruction with the addition of 10% dimethyl sulfoxide per reaction. Oligonucleotides were obtained from Integrated DNA Technologies (Coralville, IA). Sequencing was done as described in the "Materials and Methods" section of Chapter 2. The isolation of *St. lividans* ZX7 genomic DNA was done by using DNAzol reagent according to the manufacturer's directions

(Invitrogen, Carlsbad, CA). The isolation of plasmid DNA from *St. lividans* ZX7 was done with Qiagen (Valecia, CA) tip-100 columns as previously described (Kieser et al., 2000).

Media and growth conditions. *E. coli* with plasmids were grown in Luria medium at 37°C, or at 30°C where noted, and supplemented with ampicillin (100 μg/ml) or kanamycin 50 (μg/ml) when appropriate (Sigma, St. Louis, MO).

St. lividans ZX7 was grown on R5, YEME, or ISP-2 media (Kieser et al., 2000). The media components were purchased from Sigma (St. Louis, MO) except where noted. R5 contains (per litre): 103 g of sucrose, 0.25 g K_2SO_4, 10.12 g of MgCl2 x $6H_20$, 10 g of glucose, 0.1 g Difco casaminoacids (Difco, Sparks, MD), 2 ml trace element solution, 5.73 g TES buffer and 20 g of agar (Difco, Sparks, MD). After sterilization the following sterile solutions were added (per litre): 10 ml of 0.5 % KH_2PO_4, 4 ml of 5M $CaCl_2$ x $2H_2O$, 15 ml of 20% L-proline and pH adjusted to 7 with NaOH. The trace elements solution contains (per litre): 40 mg of $ZnCl_2$, 200 mg of $FeCl_3$ x $6H_20$, 10 mg of $CuCl_2$ x $2H_20$, 10 mg of $MnCl_2$ x $4H_20$, 10 mg of $Na_2B_4O_7$ x $10H_20$ and 10 mg of $(NH_4)_6Mo_7O_{24}$ x $4H_20$. YEME contains (per litre): 3 g of yeast extract (Difco, Sparks, MD), 5 g of Bacto peptone (Difco, Sparks, MD), 10 g of glucose, 170 g of sucrose and $MgCl_2$ x $6H_20$ and pH adjusted to 7 with NaOH prior to sterilization. ISP-2 contains (per litre): 4 g of yeast extract, 10 g of malt extract (Difco, Sparks, MD), 4 g of glucose and 20 g of agar. The pH was adjusted to 7 with NaOH prior to sterilization. Erythromycin, kanamycin (Sigma, St. Louis, MO), and hygromycin (Calbiochem, La Jolla, CA) were used for selection at levels noted when appropriate.

Construction of expression vectors. The primers used in the construction of the following vectors are listed in Table 2. The integrative expression plasmid containing the *sorA* gene was constructed as follows and is outlined in Figure 1. A DNA fragment containing the region of the *sorA* gene from the start codon to the *Sma*I was PCR amplified from cosmid pM15-5 (Chapter 3). The 370 bp PCR product was cloned as a *Kpn*I-*Spe*I fragment by sites incorporated from the amplification primers (PS1) and cloned into *Kpn*I-*Spe*I digested pBluescript II SK (+) (Stratagene, La Jolla, CA) vector. The sequence of the PCR product was verified in the resulting plasmid pSorAs. A second PCR reaction was used to amplify the 3' end of *sorA*, from a *Hind*III site to the stop codon, from cosmid p98/1 (Chapter 3). The PCR product was cloned as a *Spe*I-*Xba*I fragment by sites incorporated from the amplification primers (PS2) into a *Spe*I-*Xba*I digested pBluescript II SK (+) vector. The sequence of the resulting plasmid, pSorAe, was verified, and the plasmid was digested with *Xba*I-*Spe*I. The 1.6 kb fragment was ligated to a *Spe*I digested pSorAs plasmid. The resulting plasmid, pSorAlink, was end sequenced and verified through multiple restriction digests. The plasmid pSorAlink was digested with *Sma*I and *Hind*III and ligated to a 6.3 kb *Sma*I-*Hind*III fragment isolated from cosmid pM15-5. The resulting plasmid, pSorA3m15, was verified by restriction digests. Plasmid pSorA3m15 was then digested with *Hind*III and used in a ligation with a 10.6 kb *Hind*III fragment from cosmid p98/1. The resulting plasmid, pSorA, was verified by restriction digests, end sequencing, and sequencing across the cloning junctions. The plasmid pSorA was digested with *Pac*I-*Pme*I, the19 kb fragment was isolated and ligated to the *Pac*I-*Pme*I digested vector, pTBK (István Molnár, Syngenta, RTP, NC), yielding pTBKsorA. The plasmid pTBK contains the thiostrepton-inducibe promoter

*ptip*A upstream of the cloning sites of *Pac*I and *Pme*I. The vector also contains a kanamycin resistance marker and the ϕC31 *int*/*att*P site for site-specific recombination into the chromosome of *St. lividans* ZX7. The plasmid also contains the pCU18 origin of replication and ampicillin resistance marker for propagation in *E. coli*.

A plasmid that contains the genes *sorB* and *sorM* was constructed as follows and outlined in Figure 2. A PCR fragment containing the *sorB* start codon to a *Bst*1107I site within *sorB* was amplified from cosmid p98/1. The 546 bp PCR product was cloned as a *Pac*I-*Pme*I fragment by incorporation of restriction sites from the amplification primers (PS3) into *Pac*I-*Pme*I digested pNEB193 (New England Biolabs, Beverly, MA). The sequence of the PCR product was verified in the resulting plasmid, pSorBs. This plasmid was digested with *Bst*1107I and *San*DI (present in PS3) and ligated to a 29.8 kb *Bst*1107I-*San*DI fragment isolated from cosmid p98/1. The 29.8 kb fragment contained the majority of *sorB*, the flanking gene *sorM* and 1.4 kb of DNA 3' to *sorM*. The fragment was isolated from a 0.4% agarose gel using Qiaex II gel extraction kit (Qiagen, Valencia, CA). The ligation was transformed into SURE competent cells (Statagene, La Jolla CA) and incubated at 30^{o}C. The resulting plasmid, pSorB was verified through restriction digests, and end sequencing. The pSorB plasmid was then digested with *Pac*I and *Pme*I, the 30.3 kb fragment was isolated as described above and ligated to *Pac*I-*Pme*I digested vector pTBBH (from István Molnár, Syngenta RTP, NC), yielding pTBBHsorB. The plasmid pTBBH contains the *ptipA* promoter upstream of the cloning sites of *PacI* and *Pme*I. The vector also contains a hygromycin resistance marker and the *IS*117 transposase and attachment site att_{M}, for site-specific integration into the chromosome of *St. lividans*

ZX7. The plasmid also contains the pCU18 origin of replication and an ampicillin resistance marker for propagation in *E. coli*.

The expression plasmid containing the genes upstream of *sorA*, together with the gene *badA* from *R. palustis* was constructed as follows and is outlined in Figure 3. The gene *sorR* was PCR amplified from cosmid pM15-5 and cloned as a *Nde*I-*Xba*I fragment, by sites introduced by the amplification primers (PS4), into *Xba*I-*Nde*I digested pNEB193. The sequence of the 510 bp PCR product was verified in the resulting plamid, pSorR. The genes *sorC*, *sorD* and *sorE* were PCR amplified from cosmid pM15-5 with primers (PS5) that introduced *Spe*I and *EcoR*I sites as well as a ribosome binding site (RBS), AGGAGG, and cloned into pCR-Blunt II–Topo vector (Invitogen, Carlsbad, CA) creating pTsorCDE. The sequence of the PCR amplification product was verified and plasmid pTsorCDE was then digested with *Spe*I and *Eco*RI (sites introduced by PS5) and the 6 kb fragment ligated into *Spe*I-*Eco*RI digested pSorR. The resulting plasmid, pSorRCDE was verified by restriction digests. The *badA* gene was PCR amplified (PS6) from plasmid pPE202 (Egland et al., 1995) with *Spe*I and *Xba*I restriction sites as well as the ribosome binding site AGGAGG incorporated into the PCR product, and were cloned into pCR-Blunt II–Topo vector creating pTbadA. The 1.5-kb PCR product was sequenced and pTbadA was digested with *Spe*I-*Xba*I and the fragment was cloned into *Spe*I-*Xba*I digested pSorRCDE to create pSorRCDEbadA. The plasmid pSorRCDEbadA was verified by restriction digests, and then digested with *Pac*I-*Pme*I and the fragment containing *sorR*, *sorC*, *sorD*, *sorE* and *badA* was cloned into the *Streptomyces* expression vector pTUE (from István Molnár, Syngenta, RTP, NC) yielding pTUEsor5. Plasmid pTUE contains the *Streptomyces* origin of replication (*ori*) and replication protein gene (*rep*)

from pIJ101, a plasmid from *St. lividans* ISP5434. The plasmid also contains the *ptipA* promoter, erythromycin resistance marker as well as the pUC18 origin of replication and ampicillin resistance gene for propagation in *E. coli*.

Construction of strains of *St. lividans* ZX7 producing soraphen A. Plasmids pTBKsorA and pTBBHsorB were co-transformed into protoplasts of *St. lividans* ZX7 by PEG-assisted transformation using the standard procedure (Kieser et al., 2000). The transformants were selected on R5 media overlaid with hygromycin (100 μg/ml) and kanamycin (25 μg/ml) to select for the co-integration event. Resistant strains were analyzed by Southern hybridization and it was determined that the resistant strains contained both integrated plasmids (data not shown). Proplasts were then prepared from this strain and transformed with the plasmid pTueSor5 and selected on R5 media containing hygromycin and kanamycin for selection of the previous integration events and erythromycin (200 μg/ml) for selection of strains containing the plasmid pTueSor5. These strains were analyzed for the intact integration of the *sorA*, *sorB* and *sorM* genes by a series of overlapping PCR reactions (Figures 5 and 6), and designated as SLS strains. Three strains SLS4, SLS5 and SLS7 were analyzed for the production of soraphen A.

Fermentation of SLS strains. To test for the production of soraphen A, spores from about 2 cm^2 patches of strains SLS4, SLS5, and SLS7 were inoculated into 25 ml of YEME with 20 μg/ml erythromycin and 10 grams of 3 mm glass beads to disperse cells. The cultures were incubated in a 500 ml baffled flask with 225 rpm shaking at 30^oC. After 2 days the 25 ml of culture was transferred to 2 L YEME containing 20 μg/ml erythromycin and 5 μg/ml thiostrepton in 4 L flasks. The media also contained about ten 5.0 cm long x 0.5 cm diameter strainless steel springs to

disperse cells. The cultures were grown at 30°C with shaking of 225 rpm for 3 days. Some fermentations were also supplemented with one of the following: 2.5 mM trans-cinnamic acid, 5 mM L-phenylalanine, 5 mM sodium benzoate, 5 mM phenylpyruvic acid (Sigma, St. Louis, MO).

Extraction of soraphen A. To extract soraphen A from the cells and medium, the entire fermentation broth was freeze dried. About 500 ml of methanol was then added to the dried cells and medium, and mixed overnight at room temperature. The mixture was then centrifuged at 9000 g for 1 hour and the methanol was transferred to new tubes. The methanol extracts were then dried in a SpeedVac and taken up in a final volume of 10 ml of methanol. The extracts were used in bioassays and also in HPLC and LC/MS analyses as described in Chapter 4 in the "Materials and Methods" section.

Growth of St. lividans ZX7 on aromatic compounds. *St. lividans* ZX7 growth was assayed on media containing aromatic compounds that are potential precursors of the soraphen A starter unit (Table 3). The following stock solutions were made and pH adjusted to 7.5: 0.5 M trans-cinnamic acid (in 50% methanol), 0.2 M L-phenylalanine, 1 M sodium benzoate, 0.25 M phenylpyruvic acid. Spores of *St. lividans* ZX7 from a 2 cm^2 patch on ISP-2 plates were used to inoculate 10 mls of ISP-2 liquid media and grown for 3 days at 30°C with 1 gram of glass beads as described above. R5 plates supplemented with the aromatic compounds were then inoculated with 100 µl of culture, and incubated at 30°C for seven days. After the incubation, growth was scored by visual inspection of relative amount of growth and sporulation with the following system: 4 pluses (+ + + +) indicates growth and sporulation similar to un-supplemented media, 3 pluses (+ + +) indicates a slight

reduction in growth with a delay in sporulation, 2 pluses (+ +) indicates a moderate reduction in growth and delay or no sporulation, and 1 plus(+) indicates poor growth with no sporulation.

RESULTS

Generation of *St. lividans* ZX7 strains containing the soraphen A gene cluster. To transfer the soraphen A gene cluster to *St. lividans* ZX7, the soraphen A PKS genes *sorA*, *sorB*, and the methyltransferase *sorM* were introduced into the chromosome of *St. lividans* ZX7. The genes for *sorA*, *sorB* and *sorM* were reconstructed on plasmids using PCR and restriction digests as outlined in Figures 1 and 2. The expression plasmid containing the *sorA* gene was integrated into the chromosome of *St. lividans* ZX7 by site-specific integration mediated by the *int*/*att*P system from the *Streptomyces* phage ϕC31. The system is in widespread use as an essential component of many integration vectors in *Streptomyces* (Henderson et al., 1990; Kuhstoss et al., 1991; Brunker et al., 1998; Kieser et al., 2000; Wu et al., 2000; Combes et al., 2002). The expression plasmid (Figure 2) containing the *sorB* and *sorM* genes was integrated into the chromosome of *St. lividans* ZX7 by site-specific integration mediated by the transposase and att_M sites from *IS*117. *IS*117 is a transposable element from *Streptomyces coelicolor* A3(2) and is widely used in integration vectors in *St. lividans* (Henderson et al., 1990; Smokvina and Hopwood, 1993; Smokvina et al., 1994; Wall et al., 1998; Bhatt and Kieser, 1999; Xue et al., 1999; Kieser et al., 2000). Both *sorA* and *sorB* were cloned under the control of the *ptipA* promoter. The *ptipA* is one of the most utilized promoters to drive inducible expression in *Streptomyces* (Kuhstoss and Rao, 1991; Takano et al., 1995; Brunker et al., 1998; Xue et al., 1999; Kieser et al., 2000; Ali et al., 2002; Wilkinson et al.,

2002). The *ptipA* promoter is induced by small amounts of thiostrepton. The integration of the expression plasmids containing *sorA* and *sorB*/*sorM* were confirmed by Southern hybridization analysis (data not shown).

The genes *sorR*, *sorC*, *sorD*, and *sorE* were cloned into an expression vector that replicates in *St. lividans* ZX7 (Figure 3). The *badA* gene from *R. palustris* was also cloned on this plasmid. The expression plasmid contains a *Streptomyces* origin of replication and replication protein gene derived from pIJ101, a plasmid originally isolated from *St. lividans* ISP5434 (Kieser et al., 2000). *SorR* was placed under the control of the *ptipA* promoter. A *Streptomyces* ribosome binding site (RBS), AGGAGG, was engineered 12 bases upstream of the start codon of the *sorC*-E cassette, and the *badA* gene. This RBS occurs in the *ptip*A promoter and is widely distributed in genes that are highly expressed in *Streptomyces* (Strohl, 1992; Kieser et al., 2000). *SorC*-E and *badA* were cloned downstream of *sorR* in this manner to create an artificial operon whose expression is controlled by the *ptipA* promoter.

The genes *sorR* and *sorC*, *sorD*, *sorE* were predicted to be involved in the biosynthesis of soraphen A (Chapter 3) and data presented in Chapter 4 shows that the gene *sorC* is required for soraphen A production in *S. cellulosum*. *Orf2*, which lies upstream of the gene *sorE*, was shown not to be required for soraphen A production. The expression plasmid, pTUEsor5, was transformed into strains of *St. lividans* ZX7 that were confirmed to have *sorA* and *sorB*/*sorM* integrated into the chromosome. These transformants were designated SLS strains.

The SLS strains were then used in a series of overlapping PCR reactions that spanned the integrated genes *sorA*, *sorB* and *sorM*. These overlapping PCR reactions were done to analyze the integration events and to confirm that these genes were

intact. The PCR reactions of the SLS strains produced the expected size bands (Figures 5 and 6). End sequencing of the PCR products isolated from agarose gels was also carried out using the amplification primers and yielded the expected sequence. The PKS domains in the soraphen PKS are highly homologous to one another and it was important to analyze the *sorA* and *sorB* genes in the SLS strains to ensure that no spurious recombination events disrupted the integrated genes. Southern hybridizations using genes derived from the soraphen PKS to analyze the SLS strains were of limited use due to the high level of homology in the cluster itself. Due to this fact and that *St. lividans* ZX7 also contains native PKS clusters, a method of analyzing the SLS strains based on PCR reactions rather than on Southern hybridizations was employed. To analyze the plasmid pTUEsor5 in the SLS strains, DNA preparations were analyzed on an agarose gel and bands visualized that co-migrated with the pTUEsor5 control plamid (data not shown). These DNA preparations were also transformed back into *E. coli* DH10B and the plasmid DNA was re-isolated and analyzed by restriction digests and shown to be identical to pTUEsor5 (data not shown).

Strains SLS 4, 5, and 7 were fermented in 2 L of YEME with the addition of thiostrepton to induce the promoter and erythromycin for maintenance of pTUEsor5. The pTBK and pTBBH based integration events have been shown to be substantially stable without selection during fermentations. As a control, *St. lividans* ZX7 carrying plasmid pTUE was also fermented in identical conditions. To some cultures, aromatic compounds were also added at the time of inoculation. These compounds were added as a possible precursor supply for the unusual, aromatic starter unit of soraphen A. The concentration of these compounds that had little effect on growth on *St. lividans*

ZX7 on agar plates was previously determined (Table 3) and these levels were supplemented to separate cultures.

Analysis of soraphen A production. The extracts of the fermentations were analyzed for soraphen A production by bioassay. 100 µl of extract was spotted and dried onto PDA plates. An overlay of soraphen A-sensitive *C. albicans* was then applied and the plates were incubated at 30^{o}C for 3 days. Slight, opaque zones of inhibition were seen in all extracts. Since this inhibition was seen in the controls, it probably was due to some unknown compound common to all the cultures and present in the highly concentrated extracts. Only these opaque, background zones were seen in extracts from all strains grown in YEME or YEME supplemented with L-phenylalanine or phenylpyruvate (Figure 7). In strains SLS 4, 5, and 7 supplemented with sodium benzoate or trans-cinnamic acids, however, zones of clear inhibition were seen while only the small, opaque, background zones were observed in the control strain containing pTUE grown in the same conditions.

The extracts were then analyzed by HPLC. The extracts from the control strain and all strains supplemented with L-phenylalanine and phenylpyruvate that showed no activity in the bioassay displayed no soraphen A peak at the retention time of 14 minutes (Figure 8). Extracts from strains SLS4, SLS5, and SLS7, when supplemented with sodium benzoate or trans-cinnamic acid, showed a small peak at 14 minutes (Figure 9).

The extracts were then analyzed by LC/MS. The extracted soraphen A mass ion of the soraphen A standard is detected at a retention time of 36 minutes, and the mass spectrum of its peak shows the expected mass of soraphen of 542 (F.W. 523 + 23 for Na^{+} adduct) (Figure 10). Soraphen A was not detected in either the extracted

mass ion from the LC data or in the mass spectra corresponding to the expected retention time for soraphen A for the strains and conditions that showed no activity in the bioassay and no soraphen in the HPLC (Figures 11 and 12). Extract from SLS4, SLS5 and SLS7 grown on sodium benzoate and trans-cinnamic acid, however, displayed a peak corresponding to the extracted soraphen A mass ion at 36 minutes (Figures 13 and 15). Similarly, mass spectra of the peaks at 36 minutes include the a mass ion of 542 (Figures 14 and 16), consistent with the presences of soraphen A. Taken together, the bioassay, HPLC and LC/MS data confirm that strains SLS4, SLS5 and SLS7 produced soraphen A in the fermentations supplemented with sodium benzoate or trans-cinnamic acid.

DISCUSSION

Soraphen A is a macrolide polyketide produced by the myxobacterium *S. cellulosum* So ce 26. Soraphen A has a broad spectrum of activity against fungi, including important plant pathogens and was of considerable commercial interest until the research program was abandoned due to a weak teratogenic activity of soraphen A. Due to its unique activity of inhibition of acetyl coenzyme carboxylase in fungi, the generation of derivatives of soraphen A would be of interest. Some of these derivatives might retain the antifungal activity and have reduced or zero teratogenicity. The re-engineering of type I PKSs to produce polyketide derivatives is technically difficult, but possible in *Streptomyces* strains that have developed genetic systems (Xue et al. 1999). This task becomes even more difficult with soraphen A since few genetic systems are available in *S. cellulosum* So ce26. *S. cellulosum* has a long generation time and no known plasmids, and DNA can only be introduced at a low frequency through conjugation followed by homologous recombination. These

factors led to the attempt to heterologously express the soraphen A gene cluster in *St. lividans* ZX7. *St. lividans* has been shown to be an acceptable host for the production of polyketides, it has developed genetic systems and it is easy to manipulate (Stanzak et al., 1986; Xue et al., 1999; Ziermann and Betlach, 1999; Kieser et al., 2000; Shah et al., 2000; Tang et al., 2000a).

There are several considerations for the heterologous expression of PKS genes (reviewed in; Pfeifer and Khosla, 2001). The acyl carrier protein (ACP) involved in the biosynthesis of polyketides requires posttranslational modification via the covalent attachment of a 4'-phosphopantetheine group to a conserved serine residue. Successful heterologous expression systems require that the enzyme that carries out this modification, phosphopanteteinyl transferase (PPTase), is able to activate the heterologously expressed ACP (Lambalot et al., 1996; Kealey et al., 1998). It was expected that the soraphen ACP domains would be activated by the native PPTases of *St. lividans* ZX7, since this strain produces polyketides, and also heterologous polyketides have been previously expressed in *St. lividans*.

A second consideration is the substrate availability for polyketide biosynthesis in the heterologous host. The substrates that are predicted to be required for soraphen A biosynthesis include malonyl-CoA, methylmalonyl-CoA, hydroxymalonyl-CoA (or its methylated counterpart, methoxymalonyl-CoA) and benzoyl-CoA (Chapter 3). Malonyl-CoA and methylmalonyl-CoA are common metabolites in *Streptomyces* with multiple synthetic routes existing for these precursors, and they are also utilized by the endogenous *St. lividans* PKSs (Pfeifer and Khosla, 2001). The genes *sorC*, *sorD* and *sorE* were included in the expression strategy to supply a biosynthetic route for the unusual extender unit methoxymalonate. Methoxymalonate is predicted to be

utilized as an extender unit by two different modules in the soraphen A PKS (Chapter 3). In a recent experiment, the genes predicted to be responsible for the biosynthesis of the methoxymalonate extender unit for the polyketide ansamitocin, *asm*13-17, were co-expressed in a *St. lividans* strain with a cassette of *ery*ABC encoding the erythromycin PKS, in which one of the acyltransferase domains had been replaced with the methoxymalonate acyltransferase from the FK520 PKS (Kato et al., 2002). The engineered *St. lividans* strain produced a 6-deoxyerythronolide derivative with a methoxymalonate incorporated in the predicted area of the molecule. This result indicates that engineered *St. lividans* strains are able to support the biosynthesis and incorporation of methoxymalonate into heterologous expressed polyketides.

The starter unit for most type I polyketides is derived from acetate or propionate. In other examples, aromatic starter units are used as in the cases of rifamycin and ansamycin (Aparicio et al., 1996; Yu et al., 2002). In these examples, a coenzyme A ligase is used to activate and attach the aromatic starter unit to the loading ACP domain. In enterocin, a type II polyketide in "*Streptomyces maritimus"*, benzoyl-CoA serves as a starter unit for the biosynthesis of the polyketide (Piel et al., 2000). In enterocin biosynthesis, the starter unit formation resembles an eukaryotic-type pathway including the activity of a unique bacterial phenylalanine ammonia-lyase and β-oxidation reactions to synthesize benzoyl-CoA from phenylalanine (Xiang and Moore, 2002; Xiang and Moore, 2003). A benzoate-CoA ligase or dedicated pathway for benzoyl-CoA biosynthesis was not found with the soraphen A gene cluster. There is no evidence that *St. lividans* ZX7 has benzoyl-CoA as part of its metabolism. In an effort to supply the heterologous strains with benzoyl-CoA to serve as a starter unit for soraphen A biosynthesis, the *badA* gene from *R. palustris*

was included in the expression plasmid. BadA, a benzoate-CoA ligase, is the first enzyme in the metabolic pathway that allows *R. palustris* to grow on benzoic acid as a sole carbon source (Egland et al., 1995). In a recent study, constructs were made which contained the loading module and most of the first extension module from the soraphen PKS fused to the first two extension modules of the erythromycin PKS and a thioesterease domain (Wilkinson et al., 2001). The production of hybrid aromatic triketides from this model system were not observed until the *badA* gene was also co-expressed in the presence of sodium benzoate in *Saccharopolyspora erythraea*. The combination of supplying benzoate to the fermentations and the expression of *badA* were used in an effort to supply the activated starter unit, benzoyl-CoA, for soraphen A biosynthesis. The production of soraphen A under these conditions indicates that this strategy was successful in *St. lividans* ZX7. Since sodium benzoate addition was required for soraphen A production in the fermentations, it appears that *St. lividans* ZX7 does not have an endogenous pool of benzoyl-CoA available for incorporation into soraphen A. Further experimentation will need to be done to characterize the role of the *badA* gene in the utilization of supplemented sodium benzoate for soraphen A production.

In further efforts to supply the benzoyl-CoA starter unit to the heterologously expressed soraphen A PKS genes, other compounds that might be activated through *St. lividans* ZX7 metabolism were also supplied in separate fermentations. Figure 17 outlines some of the pathways by which aromatic substrates could be incorporated into the soraphen A starter unit. The additions of phenylalanine and phenylpyruvate did not yield detectable amounts of soraphen A indicating that these compounds are not activated, or metabolized to active precursors for the biosynthesis of soraphen A

in this system. When trans-cinnaimic acid was supplied to the fermentations, however, soraphen A was produced. The biosynthesis of the benzoyl-CoA starter unit in enterocin proceeds through cinnamic acid to benzoyl-CoA by β-oxidation. The biosynthesis of soraphen A in the trans-cinnamic acid fed fermentations indicates that *St. lividans* ZX7 carries out the β-oxidation of trans-cinnamic acid to benzoyl-CoA.

The levels of soraphen A production in the SLS strains is quite low. The amount of soraphen A in the fermentations with sodium benzoate appear to be somewhat higher than those with trans-cinnamic acid. This could indicate that the incorporation of the starter unit is more efficient in the direct activation of the exogenous benzoate than via the β-oxidation of trans-cinnamic acid. While the levels of production of soraphen A are low in the SLS strains, improvements may be achieved through the optimization of fermentation media and incubation conditions. The addition of absorber resin to bind soraphen A in the fermentation may also increase the amount of soraphen A production.

A number of type I PKS gene clusters from *Streptomyces* have been successfully expressed in heterologous systems in other *Streptomyces* (Stanzak et al., 1986; Kao et al., 1994; Gould et al., 1998; Tang et al., 1999; Shah et al., 2000; Tang et al., 2000a; Tang et al., 2000b). More recently, the biosynthetic gene cluster for epothilone, a mixed PKS/non-ribosomal peptide synthetase, from *S. cellulosum* So ce90 was expressed in the gram negative myxobacterium *M. xanthus* (Arslanian et al., 2002) and epothilone was produced. The production of epothilone in *M. xanthus* was at extremely low levels (Julien and Shah, 2002). Also, the PKS responsible for the synthesis of the macrocyclic core of the antibiotic erythromycin, 6-deoxyerthronolide B (6dEB), was produced in *E. coli* (Pfeifer et al., 2001). The biosynthesis of 6dEB in

E. coli is unprecedented and necessitated up to 25 genetic alterations to the *E. coli* strain, and many of these modifications are proprietary (Pfeifer et al., 2001). This result represents the second type I polyketide cluster from a gram negative bacterium, to be successfully expressed in a Streptomyces heterologous system. The SLS strains can now be utilized to study the roles of *sorC*, *sorD* and *sorE* as well as other genes on the production of soraphen A. While the current levels of soraphen A production in the SLS strains may make the goal of engineering soraphen A derivatives difficult, the possible improvements in yield could generate an important strain for this application. The results of this heterologus expression in *St. lividans* also give insight into the metabolism and biosynthesis of the unusual starter unit benzoyl-CoA, and the extender unit methoxylmalonate involved in polyketide biosynthesis. This information is important in on-going programs of combinatorial biosynthesis with PKS systems for the production of "unnatural" natural products.

ACKNOWLEDGMENTS

I would like to thank István Molnár for critical reading of this manuscript, helpful discussions, expression plasmids, and technical assistance. I would like to thank members of the Syngenta Analytical Chemistry and Sequencing group for their assistance with analysis of samples; James Pomes, Belle Abrera, Lina Summerfelt, Makoto Ono, An Hu and Racella McNair. I would also like to thank Caroline Harwood for the kind gift of plasmid pPE202

Table 1. Bacterial strains and plasmids.

Strain or plasmid	Properties or product	Source or reference
Strains		
E. coli DH10B	*E. coli* cloning strain	Stratagene
E. coli SURE	*E. coli* cloning strain	Stratagene
St. lividans ZX7	Derivative of *St. lividans* 66	John Innes Center
St. lividans SLS4	ZX7 containing pTBKsorA, pTBBHsorB and pTUEsor5	This study
St. lividans SLS5	ZX7 containing pTBKsorA, pTBBHsorB and pTUEsor5	This study
St. lividans SLS7	ZX7 containing pTBKsorA, pTBBHsorB and pTUEsor5	This study
Plasmids		
pNEB193	Cloning vector	New England Biolabs
pBluescript SKII(+)	Cloning vector	Stratagene
pCR-BluntII-TOPO	PCR cloning vector	Invitrogen
pSorAs	370 bp amplified fragment of 5'end of *sorA* cloned into pBluescript SK II (+)	This study
pSorAe	1.6-kb amplified fragment of 3'end of *sorA* cloned into pBluescript SK II (+)	This study
pSorAlink	Plasmid from ligation of pSorAs and pSorAe	This study
pSorA3m15	Plasmid from ligation of pSorAlink and 6.3 kb *Sma*I-*Hind*III fragment from *sorA*	This study
pSorA	Plasmid from ligation of pSorA3m15 and 10.6 kb *Hind*III fragment of *sorA*	This study
pTBK	Thiostrepton inducible expression vector containing ϕC31 intergrase and attP for integration into the chromosome of ZX7; Kan^R, Amp^R	István Molnár
pTBKsorA	Plasmid from ligation of pTBK and pSorA	This study
pSorBs	546 bp amplified fragment of 5' end of *sorB* cloned into pNEB193	This study
pSorB	Plasmid from ligation of pSorBs and 29-kb *Bst*1107I-*San*DI fragment of *sorB*	This study
pTBBH	Thiostrepton inducible expression vector containing IS117 intergrase and attP for integration into the chromosome of ZX7; Hyg^R, Amp^R	István Molnár
pTBBHsorB	Plasmid from ligation of pTBBH and pSorB	This study
pSorR	510 bp amplified product of *sorR* cloned into pNEB193	This study
pTsorCDE	6-kb amplified product of *sorC*-E cloned into pCR-BluntII-TOPO	This study
pSorRCDE	Plasmid from the ligation of pSorR and pSorRCDE	This study
pPE202	Plasmid containing *badA* from *R. palustris*	C. Harwood
pTbadA	1.5-kb amplified product of *badA* cloned into pCR-BluntII-TOPO	This study
pSorRCDEbadA	Plasmid from the ligation of pSorRCDE and pTbadA	This study
pTUE	*E. coli* – Streptomyces thiostrepton inducible expression plasmid; Amp^R, Ery^R	István Molnár
pTUEsor5	Plasmid from ligation of pTUE and pSorRCDEbadA	This study
p98/1	Soraphen cosmid	Chapter 3
pM15-5	Soraphen cosmid	Chapter 3

Table 2. Oligonucleotide primers used in PCR reactions.

Name	Direction	Oligonucleotide sequence
PS1	Forward	GCC**GGTACCTTAATTAA**ATGACAAAGGAGTACACGCGTC
	Reverse	GCC**ACTAGT**CCCGGGAACCGGCAGGCCATCCCG
PS2	Forward	GCC**ACTAGTAAGCTT**TGTCGAGATGGGCAAGACGGAT
	Reverse	GCC**TCTAGAGTTTAAAC**TCATTTACCCAGGCCTCCGAACGC
PS3	Forward	GGG**TTAATTAA**ATGAATAACGACGAGAAGCTTGTCTCC
	Reverse	**AAACGGGACCC**TCCT**GTATAC**GCGATCCGGCCGGAGGCGACGCT
PS4	Forward	GGAATTC**CATATGTTAATTAA**GTGTCGCGGGACGACC
	Reverse	GC**TCTAGAACTAGTGAATTC**TCATCCCGGACAACGGC
PS5	Forward	CG**GAATTC***AGGAGG*TCTCGCATGACTACTGCTCTCCA
	Reverse	GG**ACTAGT**CTACGACTTCGAGAAGCCCGAG
PS6	Forward	GG**ACTAGT***AGGAGG*TCTCGCATGAATGCAGCCGCGG
	Reverse	GC**TCTAGA**TCAGCCCAACACACCCTCGCGC
SA1	Forward	ATGACAAAGGAGTACACGGGTCCG
	Reverse	AGCGCTGCGGCCTATGTGGATAT
SA2	Forward	CGACATCGACACACTCAGCGA
	Reverse	TTGGGGTAGGCTGTGAAGAGAGA
SA3	Forward	ACTCTCTCTTCACAGCCGACCCCA
	Reverse	TCGGATGTCGTGAGCAGTTGCTG
SA4	Forward	CAGCAACTGCTCACGACATCCGA
	Reverse	TTGGATTGACCCATACCGCCGAG
SB1	Forward	ATGAATAACGACGAGAACGCTTGTCTC
	Reverse	CATAAGCGTCGGAATCTTCGGTG
SB2	Forward	AGCACCGAAGATTCCGACGCTTAT
	Reverse	TGAGCACGTTGTAGTCGAAGGGG
SB3	Forward	CCCCCTTCGACTACAACGTCGTCA
	Reverse	AGACCATCATCGTCGAGAACAGCGCG
SB4	Forward	CTGCGAAACACGAGCTCCGAAGAG
	Reverse	AAGAAGCCTTGTCGGGGTCGAGCT
SB5	Forward	AGCTCGACCCCGACAAGGCTTCTT
	Reverse	AGGACAGGGTATGCCGAAGACGAC
SB6	Forward	GTCGTCTTCGGCATACCCTGTCTT
	Reverse	AGGTTGTAGCCGTCCTCGATGAGG

Restriction sites are in **bold**
Ribosome binding site is in *italic*

Table 3. Growth of *St. lividans* ZX7 supplemented with aromatic substrates

mM	trans-cinnamic acid	L-phenylalanine	sodium benzoate	phenylpyruvic acid
0	+ + + +	+ + + +	+ + + +	+ + + +
0.5	+ + + +	+ + + +	+ + + +	+ + + +
1	+ + + +	+ + + +	+ + + +	+ + + +
2.5	+ + +	+ + + +	+ + +	+ + + +
5	+ +	+ + + +	+ + +	+ + +
10	+	+ + + +	+ + +	+ + +

Plates were inoculated as described in the text. Growth was scored by visual inspection of relative amount of growth and sporulation with the following system: 4 pluses (+ + + +) indicates growth and sporulation similar to unsupplemented media, 3 pluses (+ + +) indicates a slight reduction in growth with a delay in sporulation, 2 pluses (+ +) indicates a moderate reduction in growth and delay or no sporulation, and 1 plus(+) indicates poor growth with no sporulation.

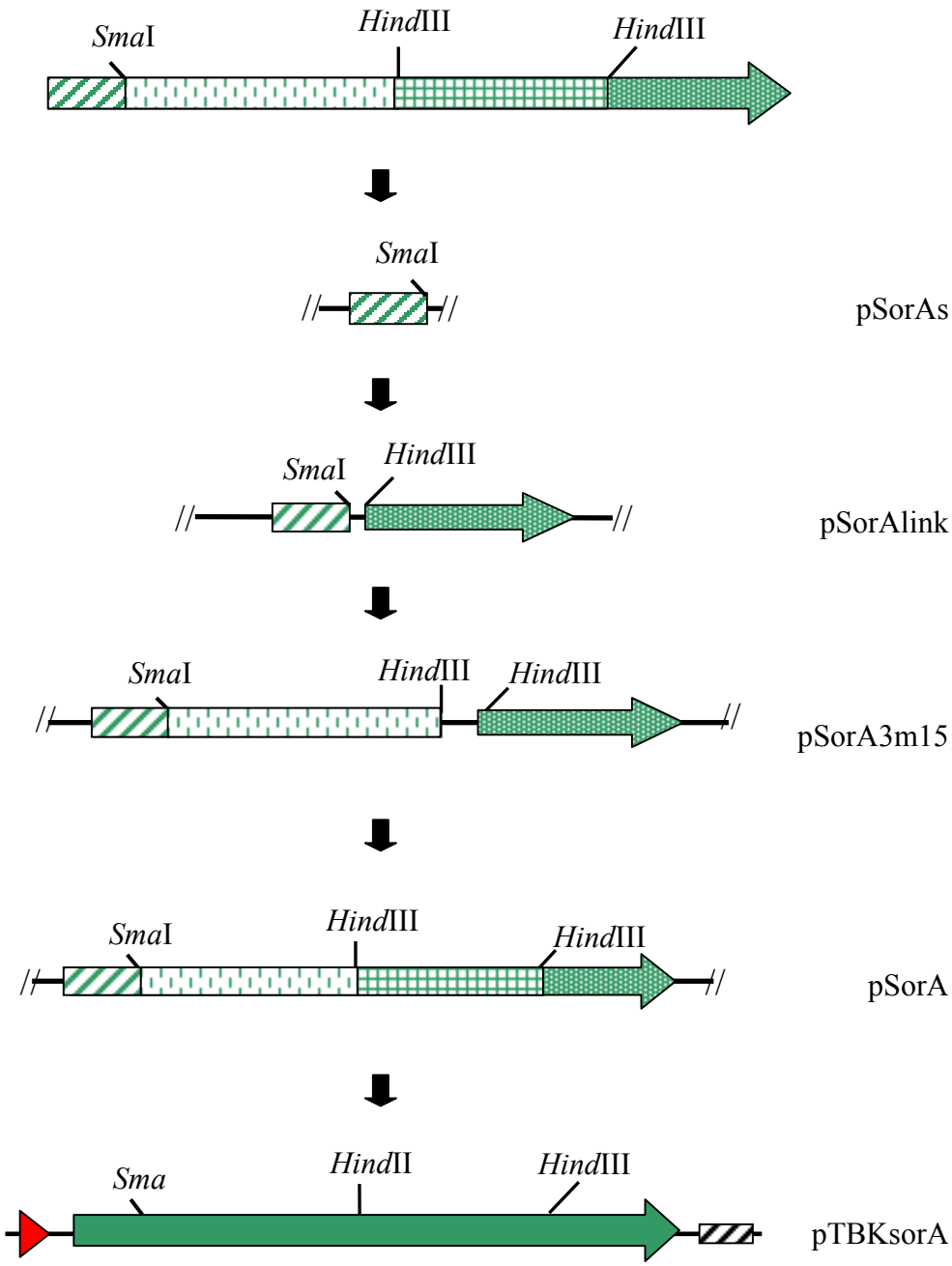


Figure 1. Linear representation of plasmids involved in the construction of the integrative expression vector for *sorA*. Details of construction are in the text. Plasmid names are listed on the right. The red arrow represents the *ptipA* promoter and the black and white hatched box represents the ϕC31 *int*/*att*P.

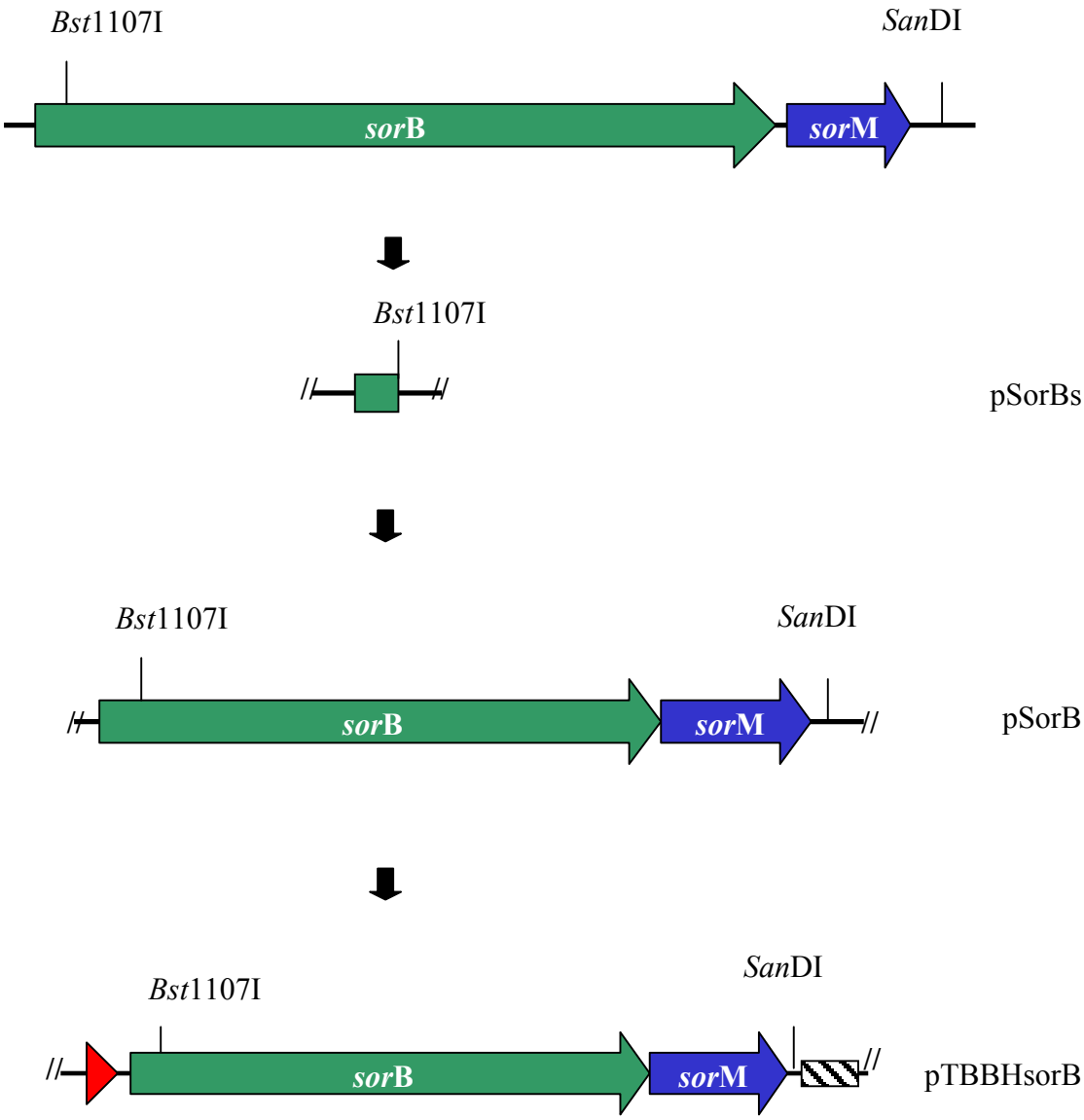


Figure 2. Linear representations of the plasmids involved in the construction of the integrative expression vector for *sorB*. Details of the construction are in the text. The red arrow represents the *ptipA* promoter and the black and white hatched box represents the transposase and *att*M attachment site from *IS*117

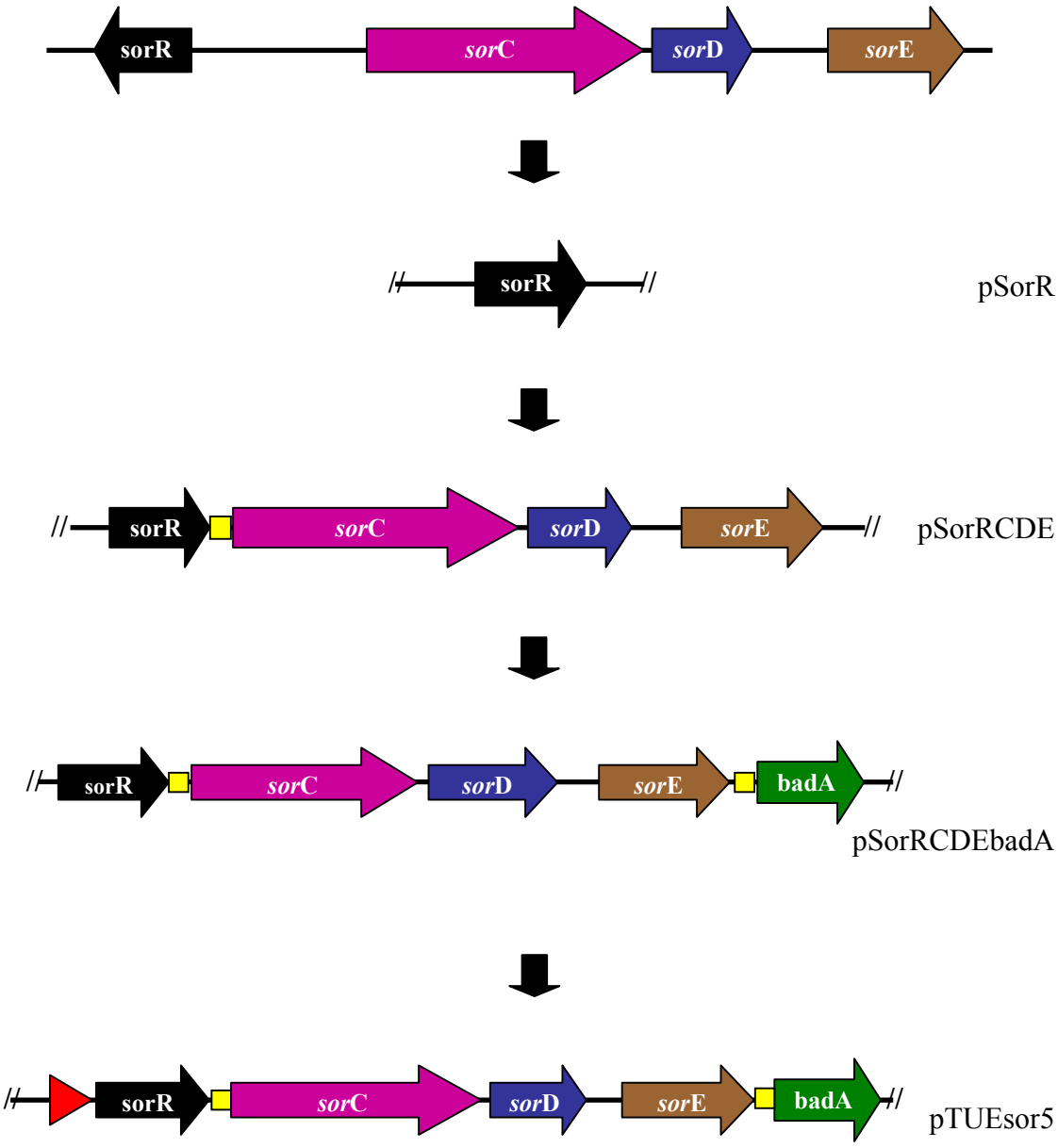


Figure 3. Linear representation of the construction of the *Streptomyces* expression plasmid containing the genes upstream of *sorA* and the *badA* gene. Details are in the text. The red arrow represents the p*tip*A promoter and the yellow boxes represent the engineered RBS "AGGAGG".

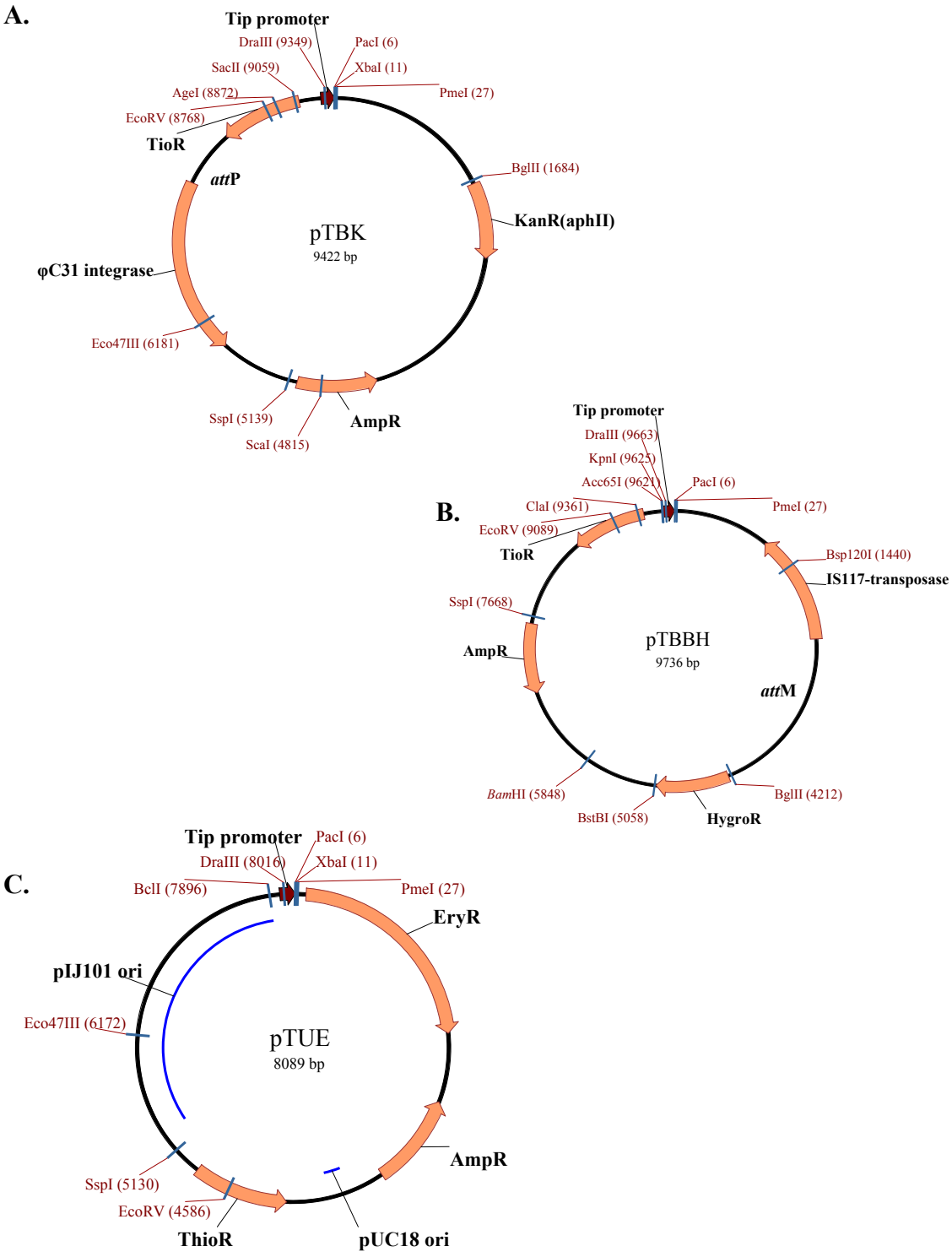


Figure 4. *Streptomyces* expression plasmids. (A); pTBK, (B); pTBBH, and (C); pTUE.

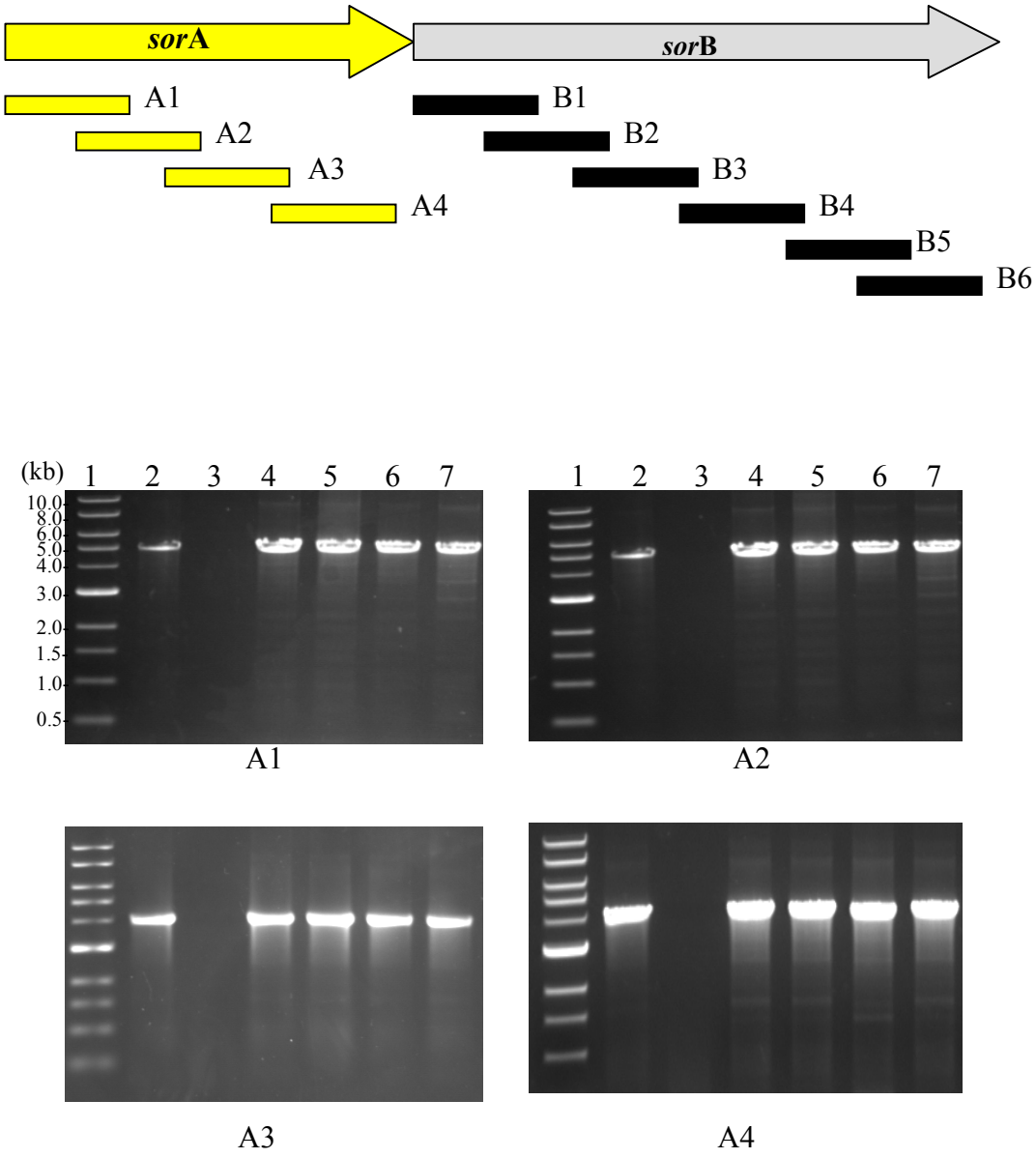


Figure 5. Overlapping PCR reactions verifying the integration of *sorA* into the chromosome of *St. lividans* ZX7. Panel A1 corresponds to the A1 fragment shown above and the SA1 primers used in the PCR reaction, listed in Table 2. 5 µl from a 50µl reaction was loaded in each lane. In each gel, 1-kb ladder is in lane 1. Lane 2 is the product of PCR reactions from genomic DNA from *S. cellulosum*. Lane 3 is the product of PCR reaction with *St. lividans* ZX7. Lanes 4, 5, 6 and 7 are product of PCR reactions from SLS 3, 4, 5 and 7 respectively.

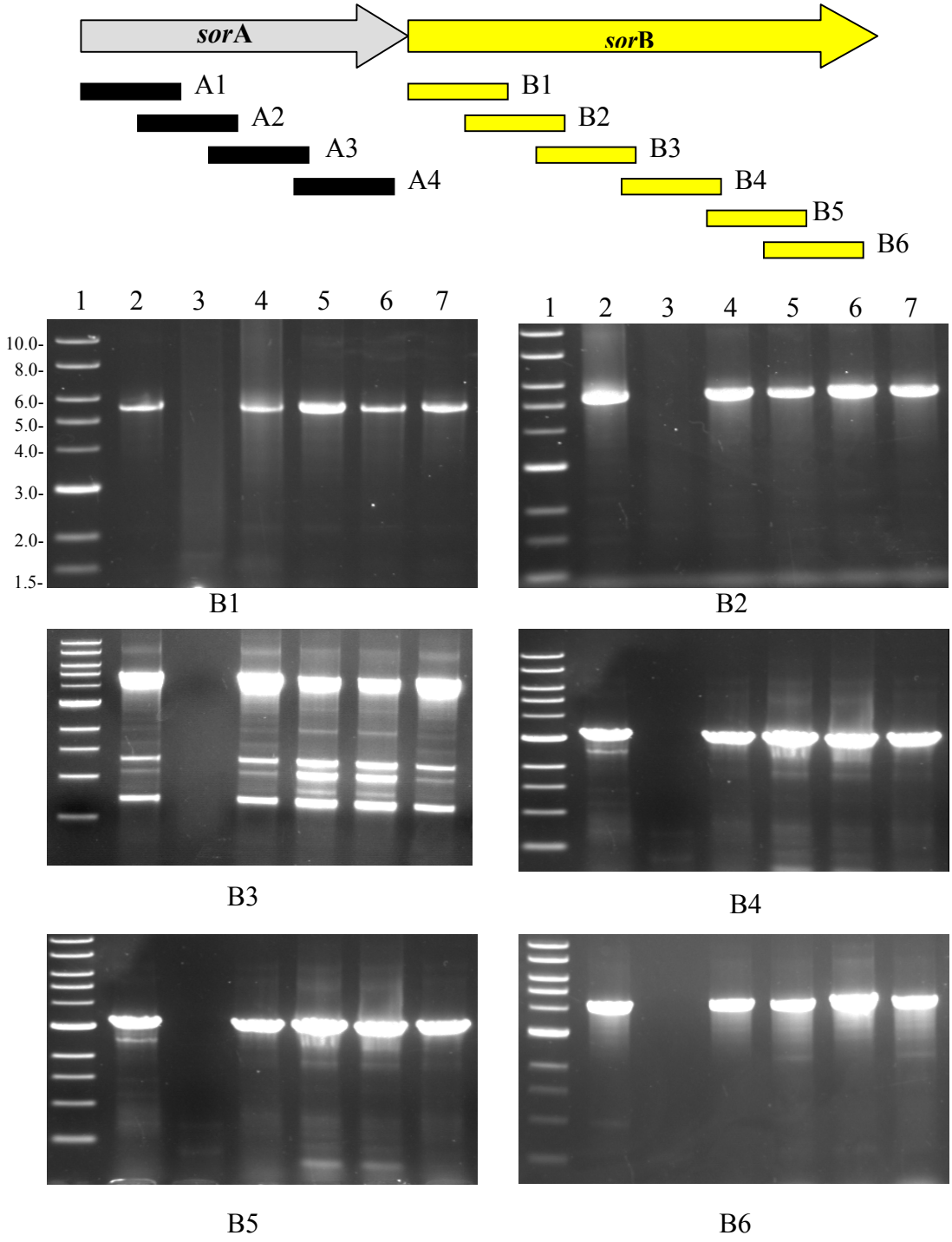


Figure 6. Overlapping PCR reactions verifying the integration of *sorB* into the chromosome of *St. lividans* ZX7. Panel B1 corresponds to the B1 fragment shown above and the SB1 primers used in the PCR reaction, listed in Table 2. 5 µl from a 50µ/ reaction was loaded in each lane. In each gel, 1-kb ladder is in the lane 1. Lane 2 is the product of PCR reactions from genomic DNA from *S. cellulosum*. Lane 3 is the product of PCR reaction with *St. lividans* ZX7. Lanes 4, 5, 6 and 7 are product of PCR reactions from SLS 3, 4, 5 and 7 respectively.

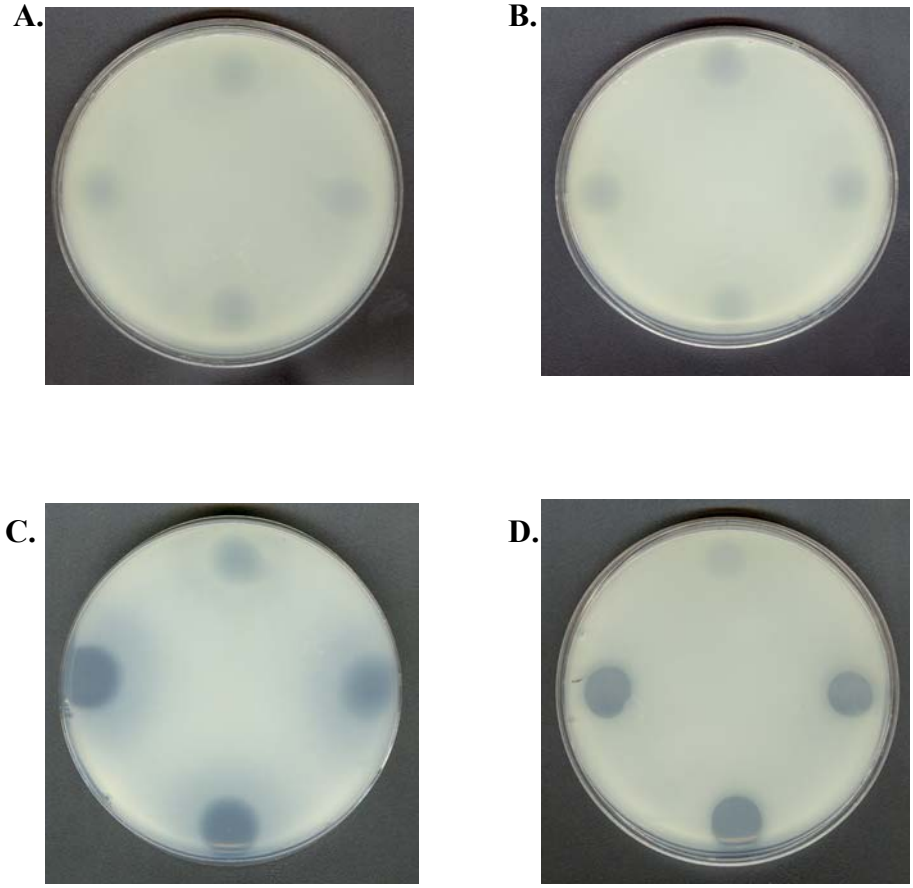


Figure 7. Bioassays of fermentation extracts. To PDA plates, 100 ul of extract was added and dried. *C. albicans* was used in an agar overlay and plates incubated for 3 days at 30^{o}C. Each plate is oriented in the same way. Extracts at the top of each plate are from *St. lividans* ZX7 control strain containing plasmid TUE. Extracts on the right of each plate are from fermentations of SLS 4. Extracts on the bottom of each plate are from fermentations of strain SLS 5. Extracts on the left of each plate are from fermentations of strain SLS 7. (A): extracts from fermentations in YEME. (B): Extracts from fermentations in YEME + 5mM phenylalanine. (C): Extracts from fermentations in YEME +5 mM sodium benzoate. (D): Extracts from fermentations in YEME + 2.5 mM trans-cinnamic acid.

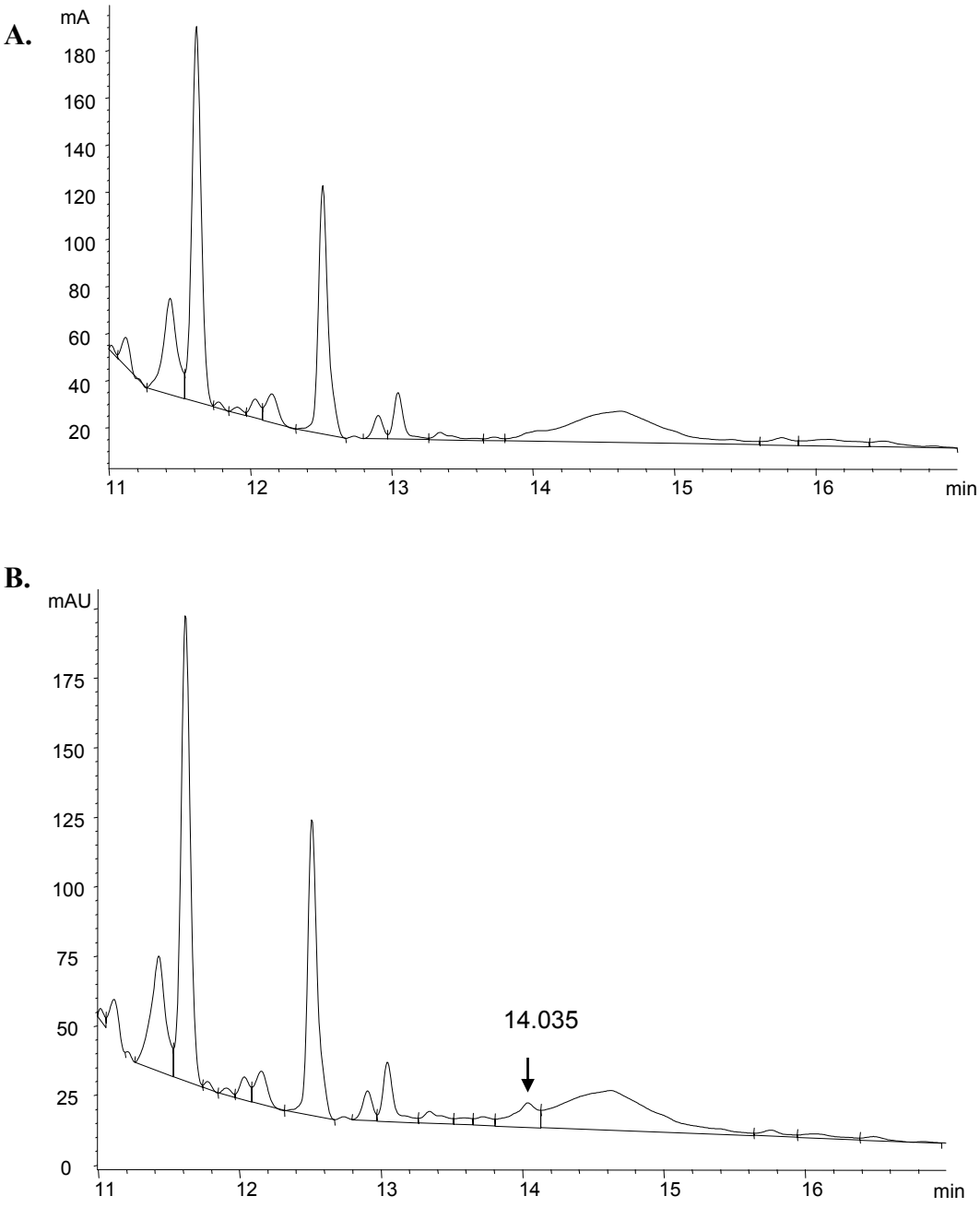


Figure 8 . HPLC of fermentation extracts. (A): Extract from *St. lividans* ZX7 with control plasmid pTUE in YEME + 5mM sodium benzoate. (B): Extract from SLS 7 in YEME + 5mM sodium benzoate displaying a peak migrating at the retention time of soraphen A standard (14 minutes).

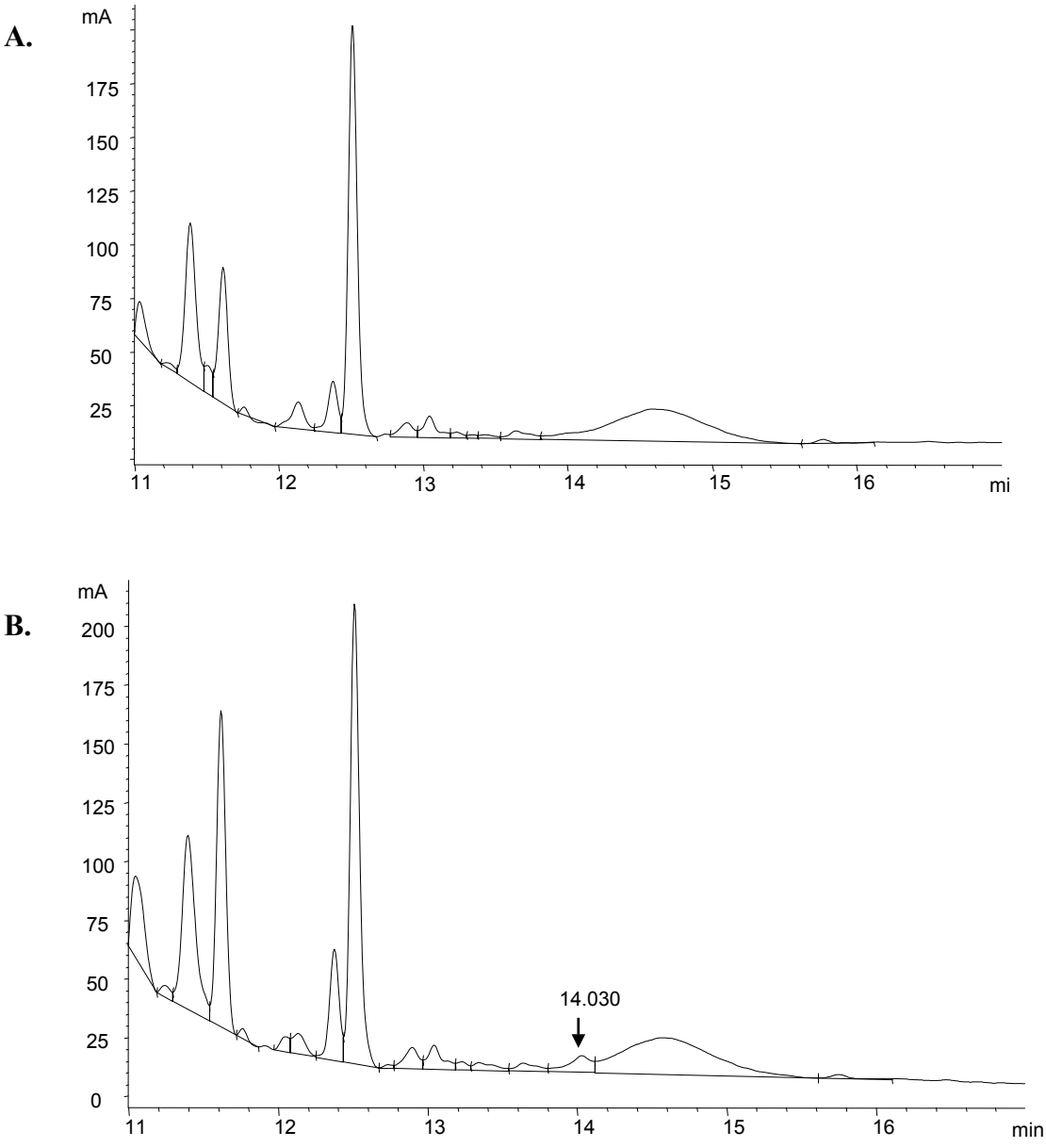


Figure 9. HPLC of fermentation extracts. (A): Strain SLS 7 in YEME. (B) Strain SLS 7 in YEME + 2.5 mM trans-cinnamic acid displaying a peak migrating to the retention time of soraphen A standard (14 minutes).

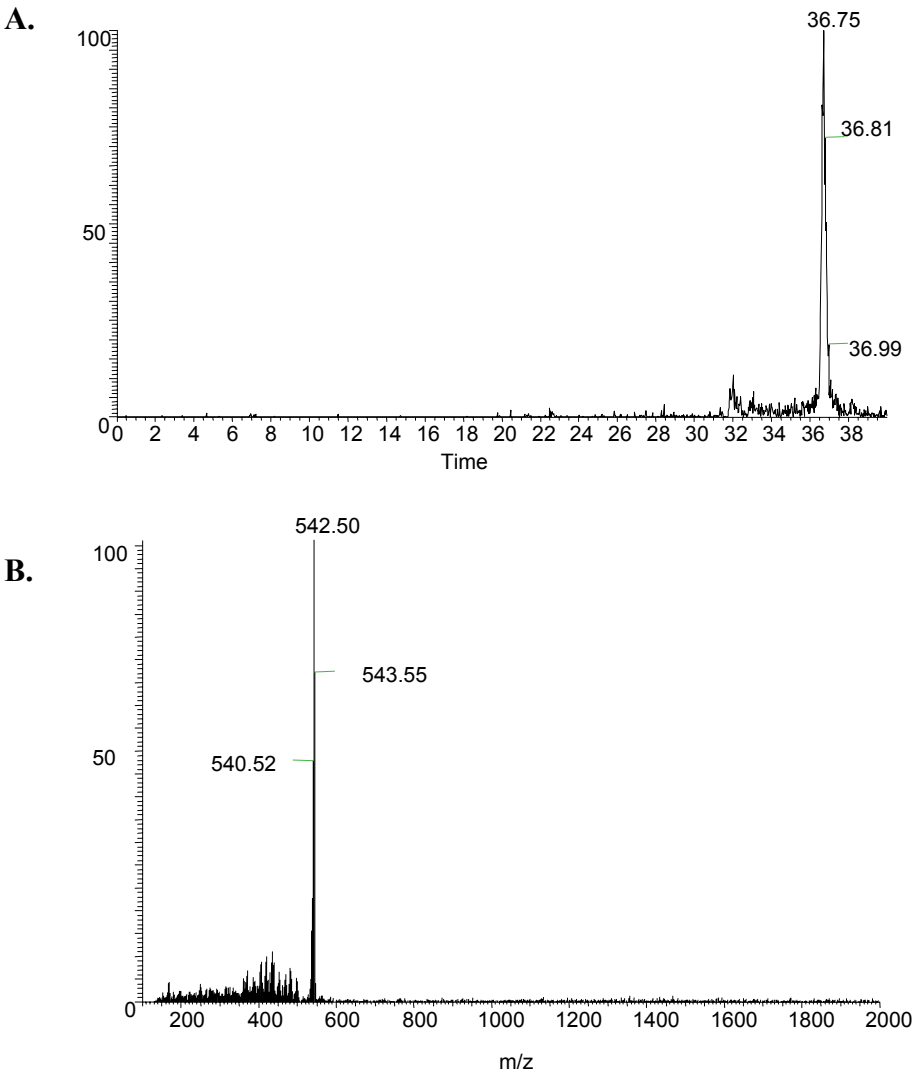


Figure 10. The extracted mass ion and mass spectra of the soraphen A standard . (A); Peaks corresponding to the extracted mass ion (m/z 542-545) from the LC data of soraphen A standard. (B); The mass spectra of the peak at 36 minutes from (A) of the soraphen A standard (M.W. of soraphen A is 520 + Na^+ adduct of 23 = 543). The relative abundance is shown on the x-axis

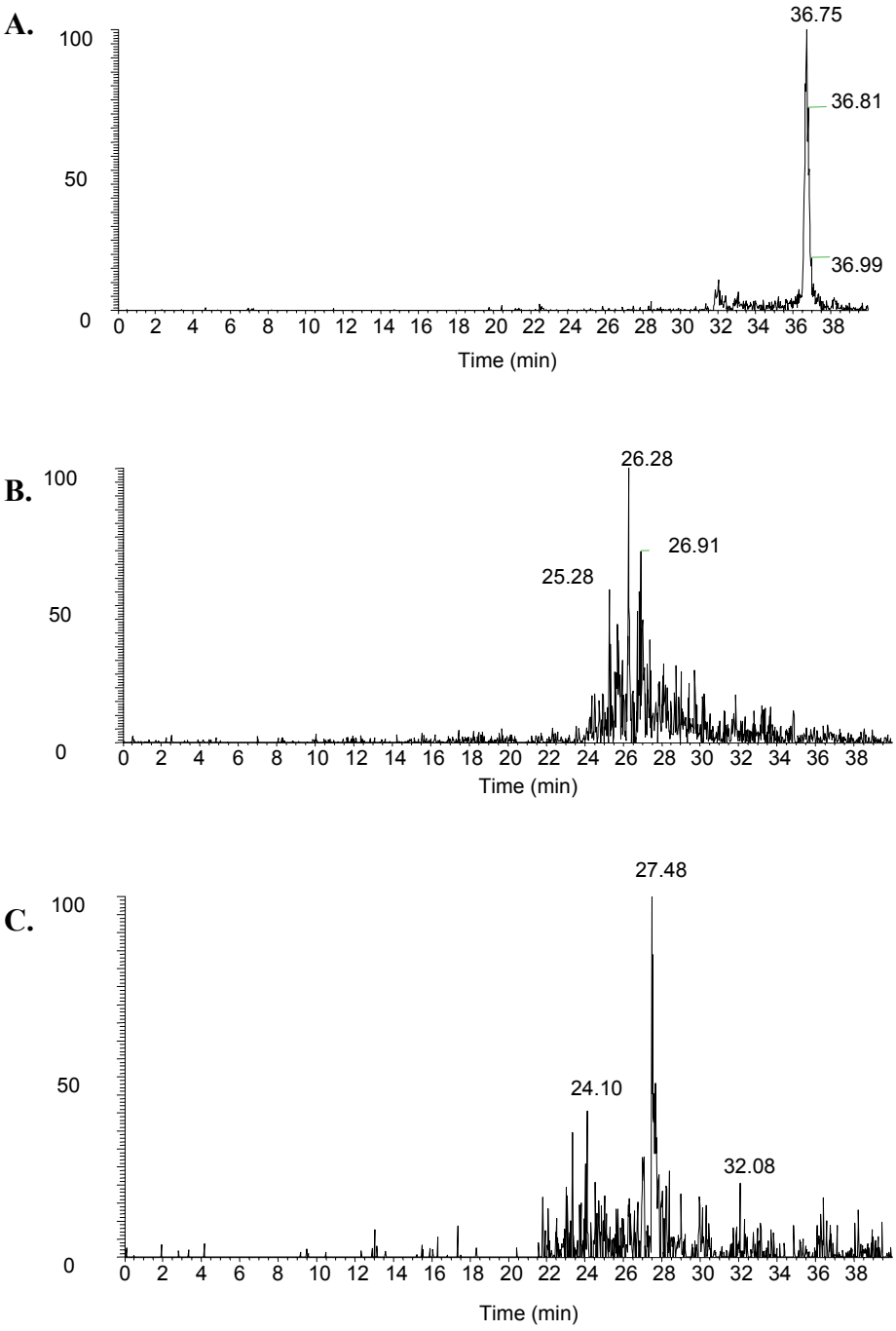


Figure 11. Peaks corresponding to the extracted mass ion (m/z 542-545) from the LC data of the (A); soraphen A standard, (B); extract of fermentation *St. lividans* ZX7 with control plasmid pTUE in YEME and (C); extract of fermentation of strain SLS 7 in YEME. No peaks are detected migrating at the retention time of the soraphen A standard. Relative abundance is shown on the x-axis.

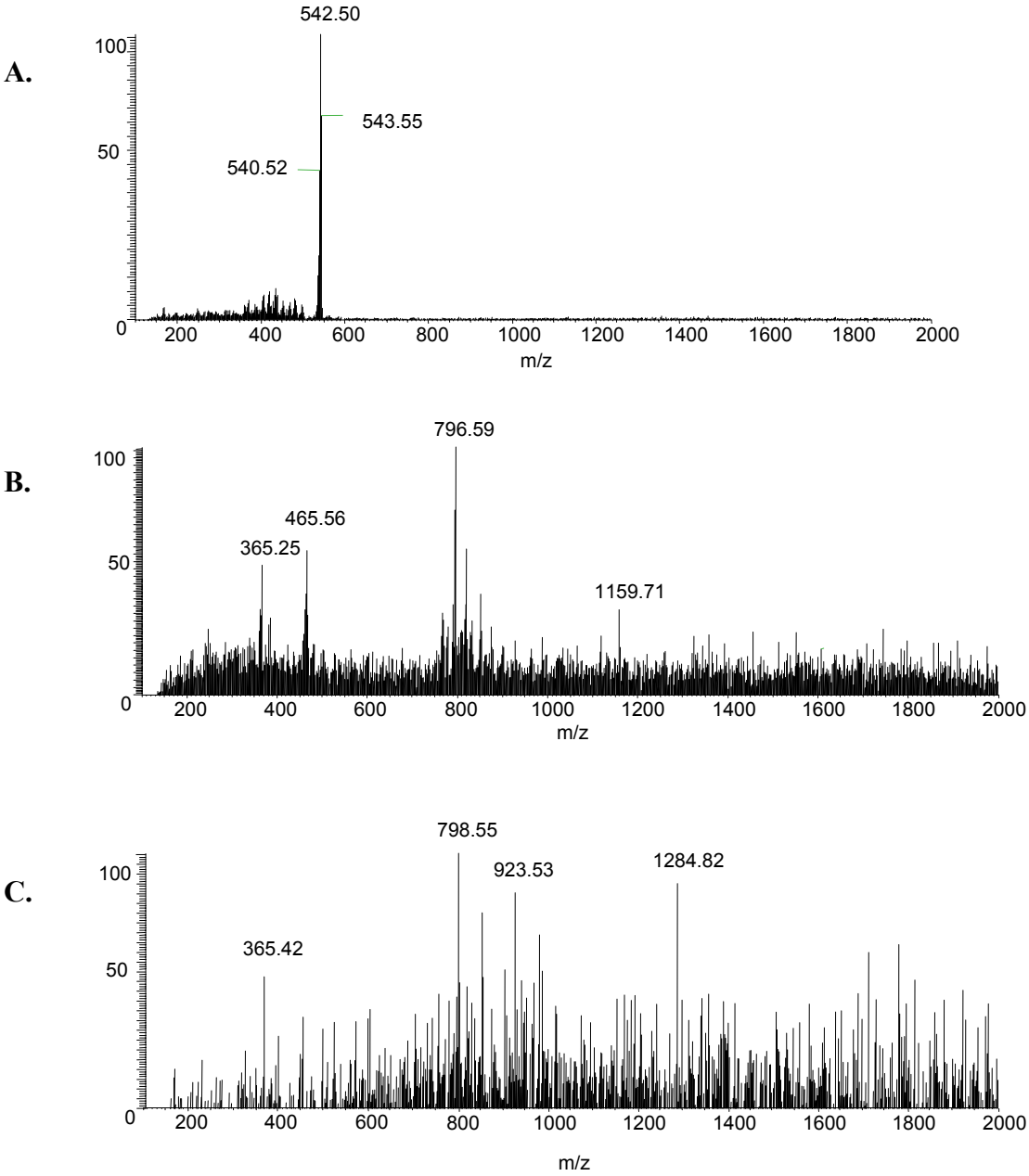


Figure 12. The mass spectra corresponding to the peaks at the expected retention time (~36.5 minutes) of soraphen A from the data in Figure 10. (A): Soraphen A standard with mass of 542 (soraphen A MW is 520 + 23 for Na+ adduct). (B): Mass spectra for fermentation extract of *St. lividans* ZX7 with control plasmid pTUE in YEME. (C): Mass spectra of fermentation extracts of strain SLS 7 in YEME. The extracts from YEME display no ion with the same mass as the soraphen A standard. Relative abundance is on the x-axis.

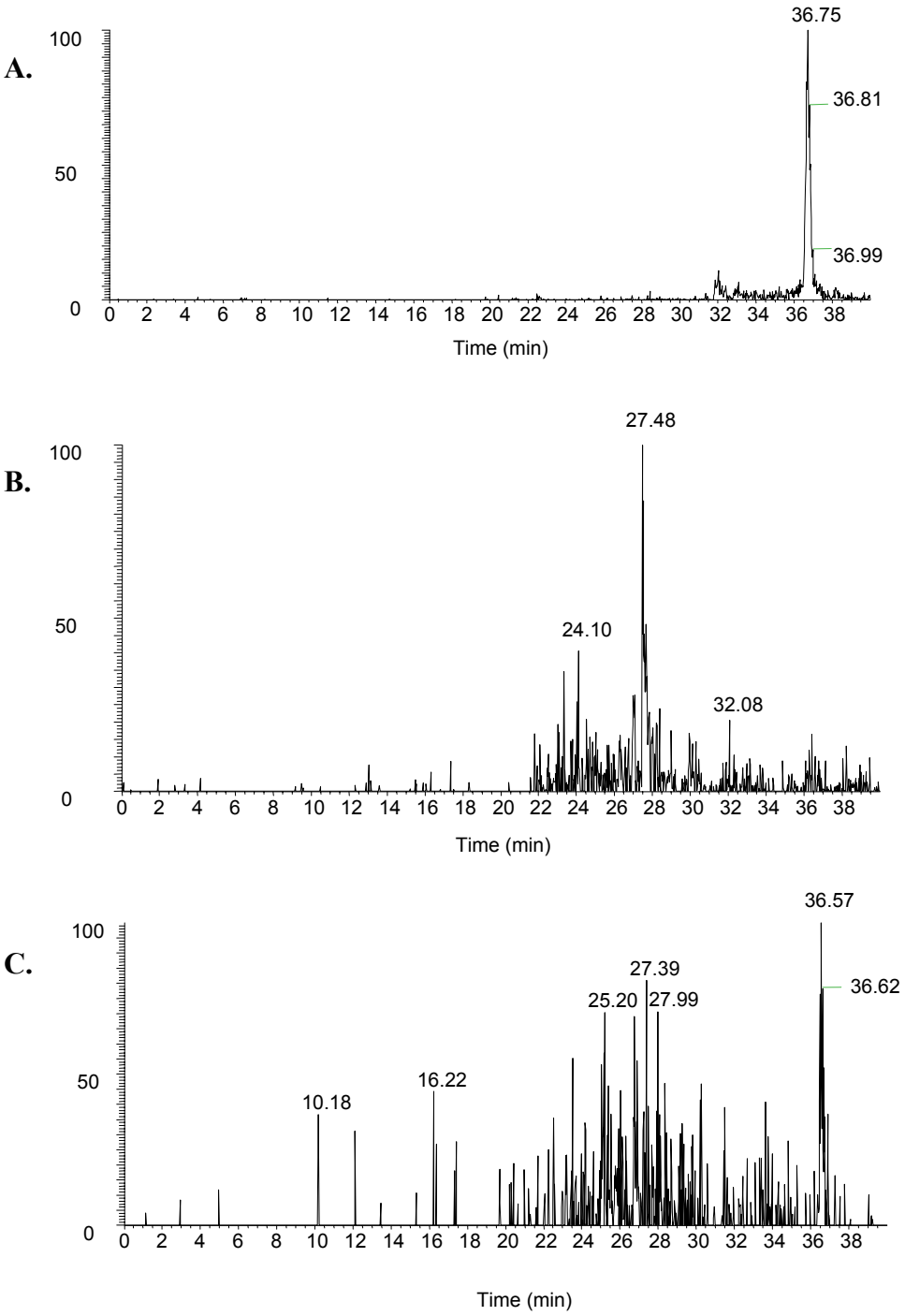


Figure 13. Peaks corresponding to the extracted mass ion (m/z 542-545) from the LC data of the (A): soraphen A standard showing the soraphen A at a retention time of ~36.5 , (B): extract of fermentation of strain SLS 7 in YEME and (C): extract of fermentation of strain SLS 7 in YEME + 5 mM sodium benzoate displaying a peak migrating at the same retention time of soraphen A (~36.5 minutes). Relative abundance is on the x-axis.

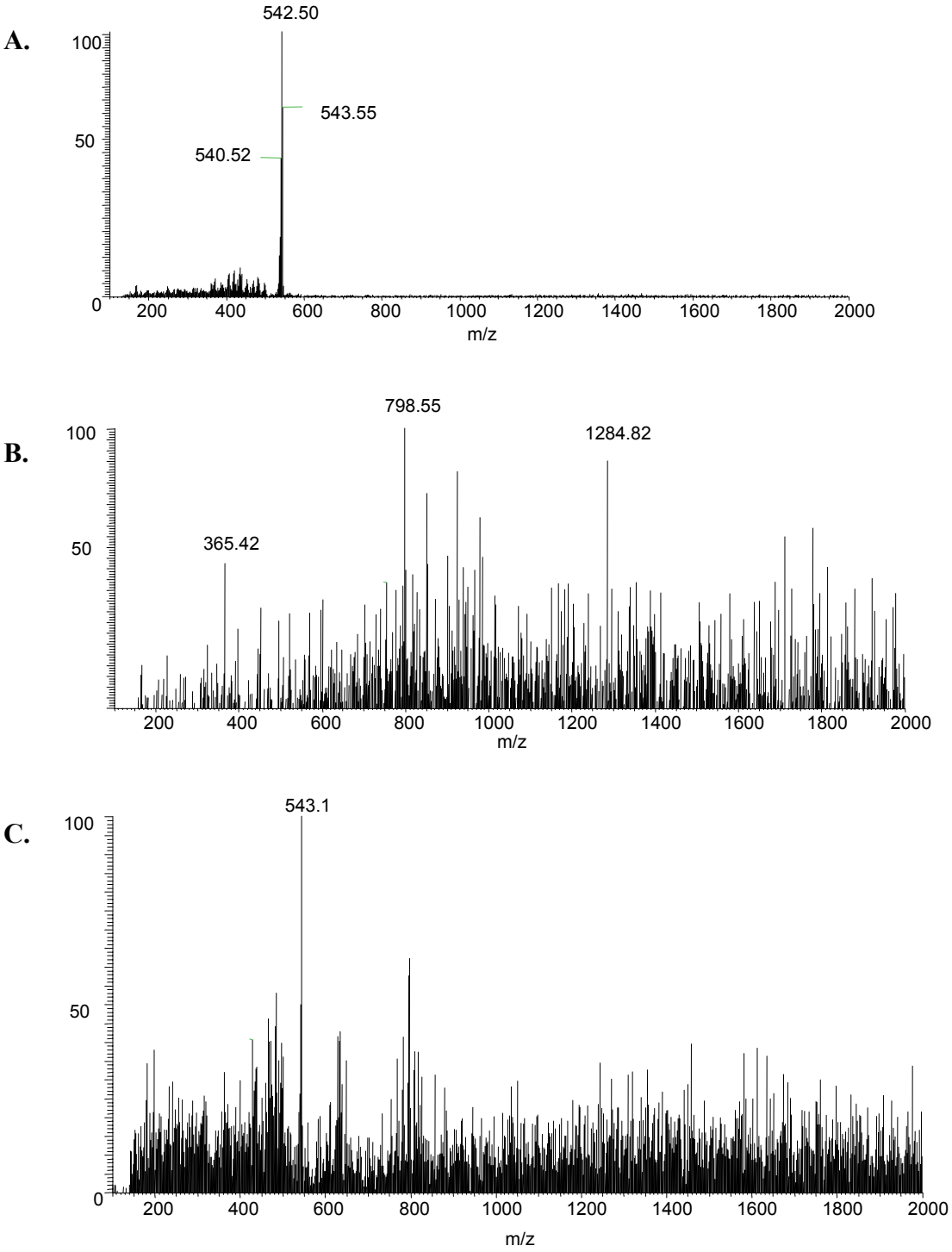


Figure 14. The mass spectra corresponding to the 36.5 minute peaks in Figure 12.(A); Soraphen A standard with mass of 542 (soraphen A MW is 520 + 23 for Na adduct). (B); Mass spectra for fermentation extract of strain SLS7 in YEME. (C); Mass spectra of fermentation extracts of strain SLS 7 in YEME + 5 mM sodium benzoate displaying the same mass ion as the soraphen A standard. Relative abundance is on the x-axis.

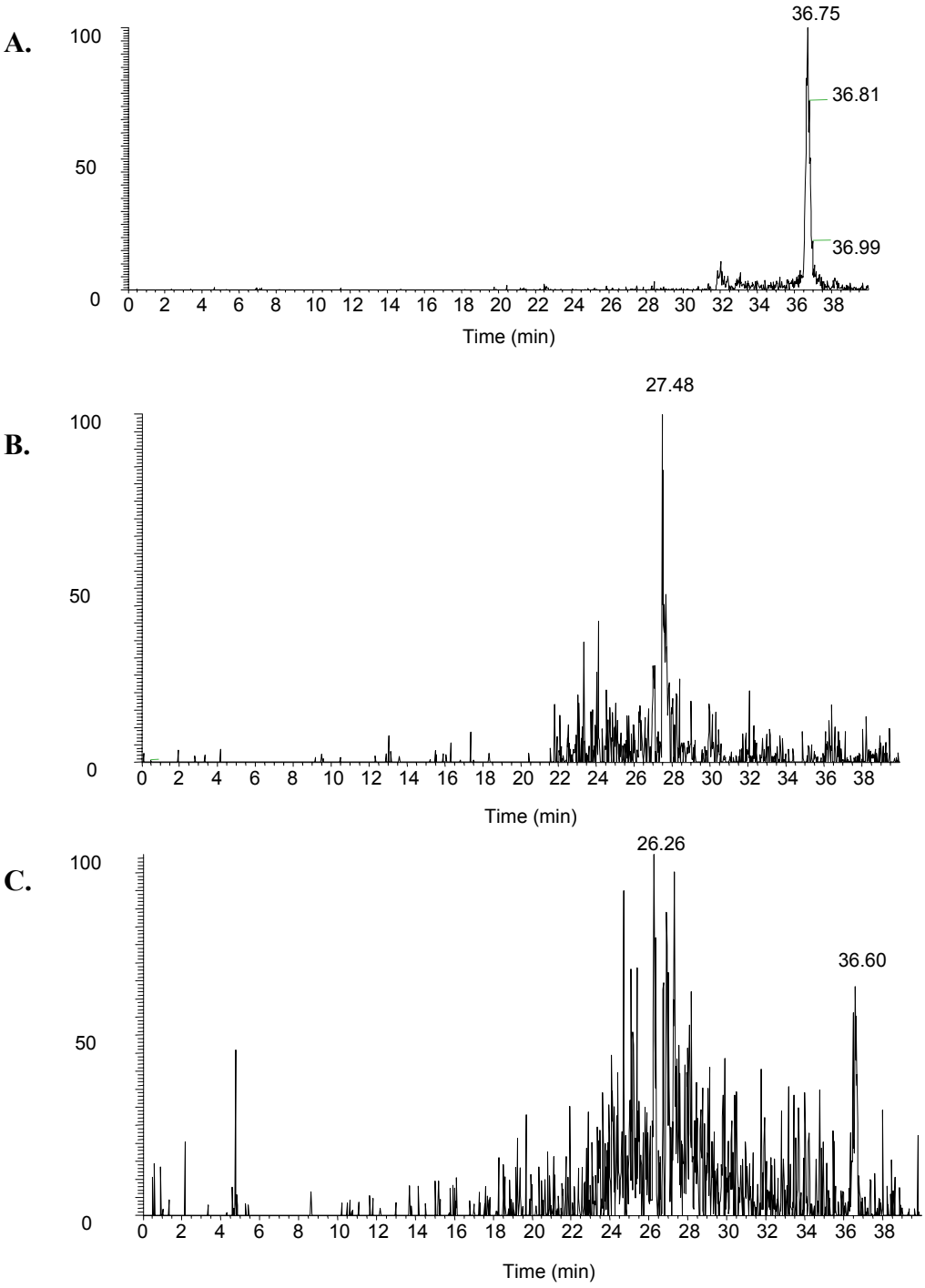


Figure 15. Peaks corresponding to the extracted mass ion (m/z 542-545) from the LC data of the (A); soraphen A standard showing the soraphen A at a retention time of ~36.5 , (B); extract of fermentation of *St. lividans* ZX7 with control plasmid pTUE in YEME + 2.5 mM trans-cinnamic acid and (C); extract of fermentation of strain SLS 7 in YEME + 2.5 mM trans-cinnamic acid displaying a peak co-migrating with soraphen A at ~36.5 minutes. Relative abundance is on the x-axis.

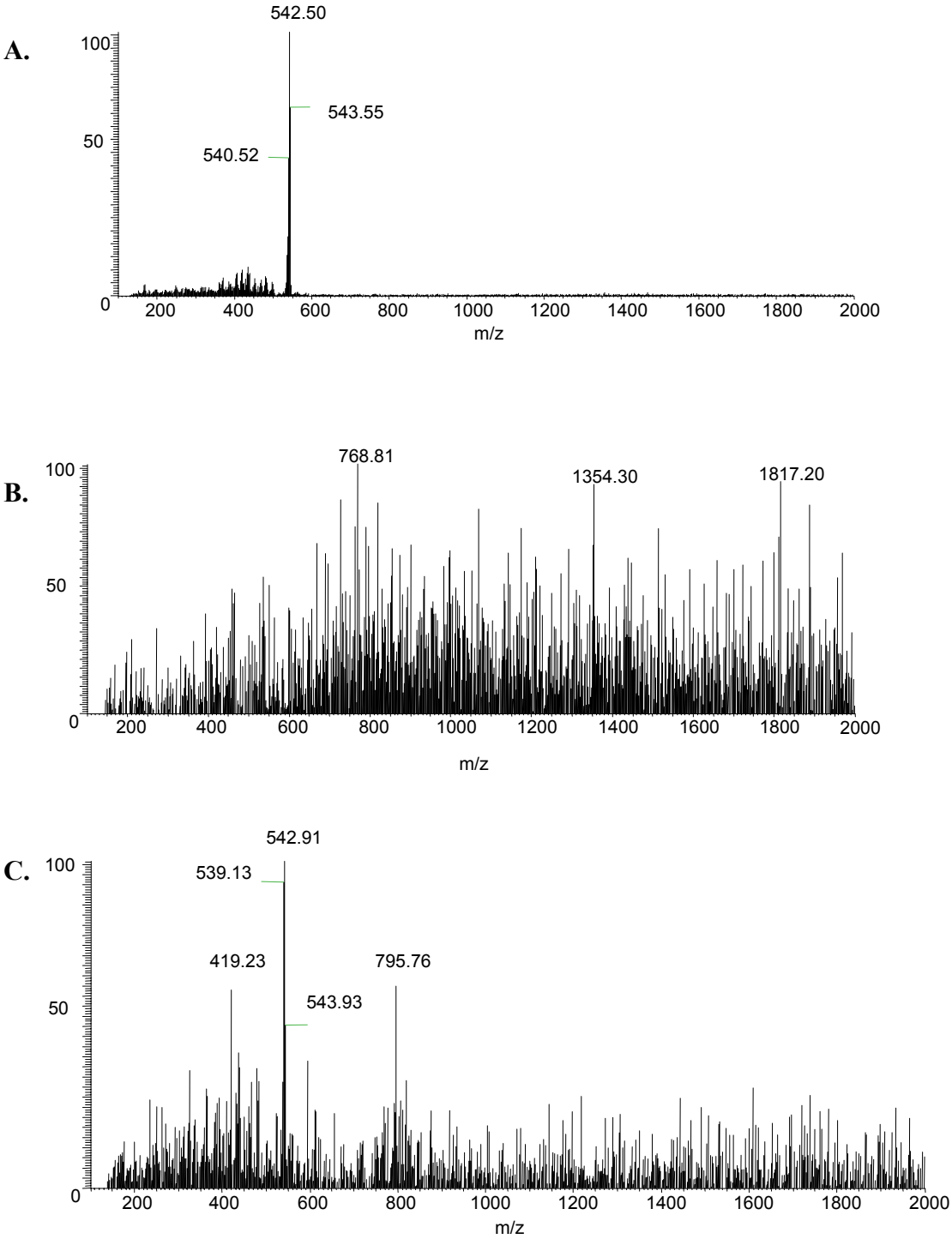


Figure 16. The mass spectra corresponding to the peaks at 36.5 minutes in Figure 14. (A); Soraphen A standard with mass of 542 (soraphen A MW is 520 + 23 for Na+ adduct). (B); Mass spectra for fermentation extract of *St. lividans* ZX7 with control plasmid pTUE in YEME + 2.5 mM trans-cinnamic acid. (C); Mass spectra of fermentation extracts of strain SLS 7 in YEME + 2.5 mM trans-cinnamic acid displaying a the same mass ion as the soraphen A standard. Relative abundance is on the x-axis.

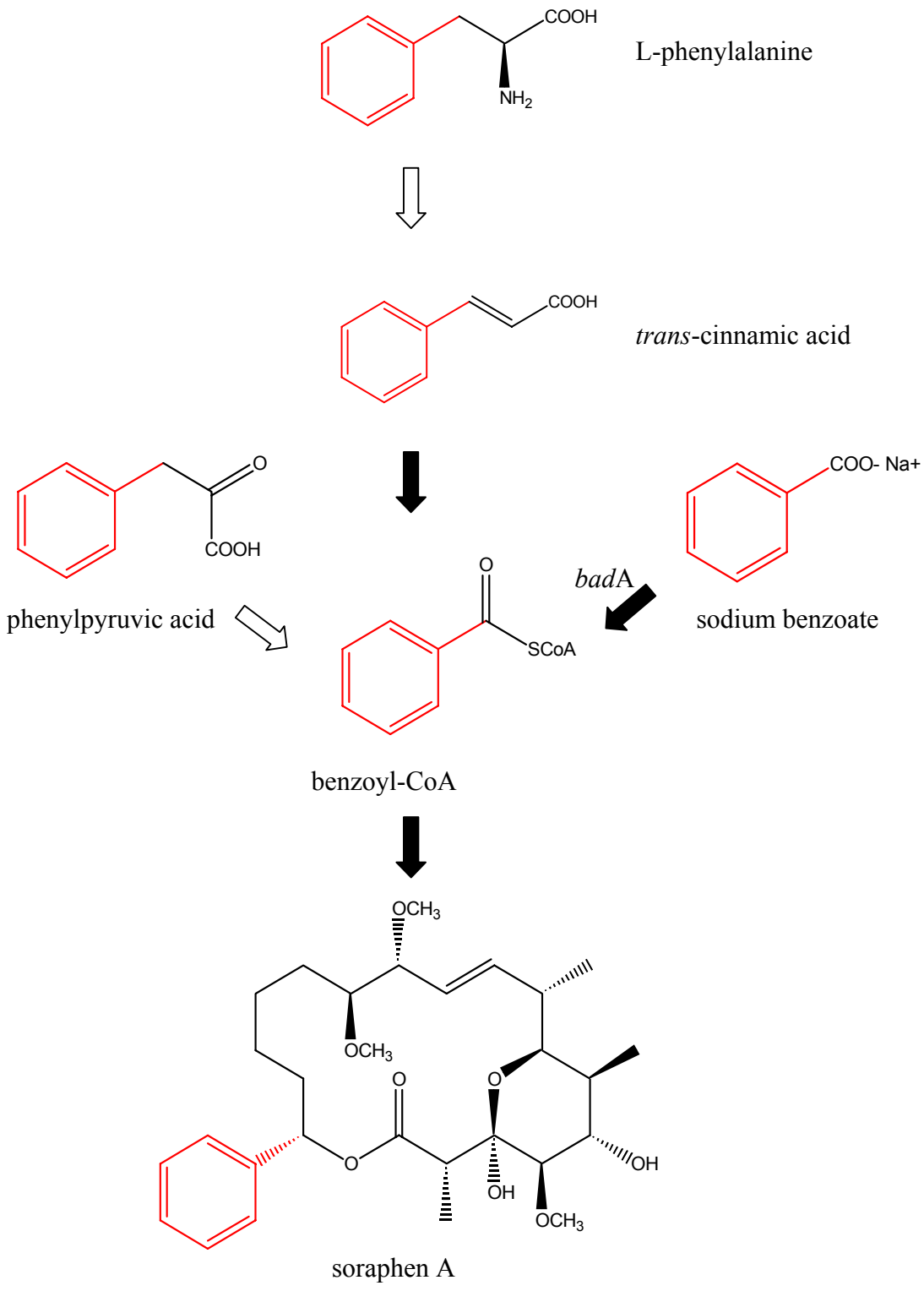


Figure 17. Possible pathways for aromatic precursor incorporation into soraphen A heterologously produced in *St. lividans* ZX7. The black arrows indicate pathways that produced soraphen A in the fermentation of the SLS strains. The white arrows indicate possible pathways that were not detectable in the experimental conditions.

REFRENCES

Ali, N., Herron, P.R., Evans, M.C. and Dyson, P.J.: Osmotic regulation of the *Streptomyces lividans* thiostrepton-inducible promoter, *ptip*A. Microbiology 148 (2002) 381-90.

Aparicio, J.F., Molnar, I., Schwecke, T., Konig, A., Haydock, S.F., Khaw, L.E., Staunton, J. and Leadlay, P.F.: Organization of the biosynthetic gene cluster for rapamycin in *Streptomyces hygroscopicus*: analysis of the enzymatic domains in the modular polyketide synthase. Gene 169 (1996) 9-16.

Arslanian, R.L., Parker, C.D., Wang, P.K., McIntire, J.R., Lau, J., Starks, C. and Licari, P.J.: Large-scale isolation and crystallization of epothilone D from *Myxococcus xanthus* cultures. J Nat Prod 65 (2002) 570-2.

Bhatt, A. and Kieser, T.: Transposition of IS117 of *Streptomyces coelicolor* A3(2) in *Mycobacterium smegmatis*. Microbiology 145 (1999) 1201-7.

Brunker, P., Minas, W., Kallio, P.T. and Bailey, J.E.: Genetic engineering of an industrial strain of *Saccharopolyspora erythraea* for stable expression of the *Vitreoscilla* haemoglobin gene (*vhb*). Microbiology 144 (1998) 2441-8.

Combes, P., Till, R., Bee, S. and Smith, M.C.: The streptomyces genome contains multiple pseudo-*att*B sites for the (phi)C31-encoded site-specific recombination system. J Bacteriol 184 (2002) 5746-52.

Egland, P.G., Gibson, J. and Harwood, C.S.: Benzoate-coenzyme A ligase, encoded by *badA*, is one of three ligases able to catalyze benzoyl-coenzyme A formation during anaerobic growth of *Rhodopseudomonas palustris* on benzoate. J Bacteriol 177 (1995) 6545-51.

Gerth, K., Bedorf, N., Irschik, H., Hofle, G. and Reichenbach, H.: The soraphens: a family of novel antifungal compounds from *Sorangium cellulosum* (Myxobacteria). I. Soraphen A1 alpha: fermentation, isolation, biological properties. J Antibiot (Tokyo) 47 (1994) 23-31.

Gould, S.J., Hong, S.T. and Carney, J.R.: Cloning and heterologous expression of genes from the kinamycin biosynthetic pathway of *Streptomyces murayamaensis*. J Antibiot (Tokyo) 51 (1998) 50-7.

Henderson, D.J., Brolle, D.F., Kieser, T., Melton, R.E. and Hopwood, D.A.: Transposition of *IS*117 (the *Streptomyces coelicolor* A 3 (2) mini-circle) to and from a cloned target site and into secondary chromosomal sites. Mol Gen Genet 224 (1990) 65-71.

Jaoua, S., Neff, S. and Schupp, T.: Transfer of mobilizable plasmids to *Sorangium cellulosum* and evidence for their integration into the chromosome. Plasmid 28 (1992) 157-65.

Julien, B. and Shah, S.: Heterologous Expression of Epothilone Biosynthetic Genes in *Myxococcus xanthus*. Antimicrob Agents Chemother 46 (2002) 2772-8.

Kao, C.M., Katz, L. and Khosla, C.: Engineered biosynthesis of a complete macrolactone in a heterologous host. Science 265 (1994) 509-12.

Kato, Y., Bai, L., Xue, Q., Revill, W.P., Yu, T.W. and Floss, H.G.: Functional expression of genes involved in the biosynthesis of the novel polyketide chain extension unit, methoxymalonyl-acyl carrier protein, and engineered biosynthesis of 2-desmethyl-2-methoxy-6-deoxyerythronolide B. J Am Chem Soc 124 (2002) 5268-9.

Kealey, J.T., Liu, L., Santi, D.V., Betlach, M.C. and Barr, P.J.: Production of a polyketide natural product in nonpolyketide-producing prokaryotic and eukaryotic hosts. Proc Natl Acad Sci U S A 95 (1998) 505-9.

Kieser, T., Bibb, M.J., Buttner, M.J., Chater, K.F. and Hopwood, D.A.: Practical *Streptomyces* Genetics. The John Innes Foundation, Norwich, England, 2000.

Kuhstoss, S. and Rao, R.N.: A thiostrepton-inducible expression vector for use in *Streptomyces* spp. Gene 103 (1991) 97-9.

Kuhstoss, S., Richardson, M.A. and Rao, R.N.: Plasmid cloning vectors that integrate site-specifically in *Streptomyces* spp. Gene 97 (1991) 143-6.

Lambalot, R.H., Gehring, A.M., Flugel, R.S., Zuber, P., LaCelle, M., Marahiel, M.A., Reid, R., Khosla, C. and Walsh, C.T.: A new enzyme superfamily - the phosphopantetheinyl transferases. Chem Biol 3 (1996) 923-36.

Pfeifer, B.A., Admiraal, S.J., Gramajo, H., Cane, D.E. and Khosla, C.: Biosynthesis of complex polyketides in a metabolically engineered strain of *E. coli*. Science 291 (2001) 1790-2.

Pfeifer, B.A. and Khosla, C.: Biosynthesis of polyketides in heterologous hosts. Microbiol Mol Biol Rev 65 (2001) 106-18.

Piel, J., Hertweck, C., Shipley, P.R., Hunt, D.M., Newman, M.S. and Moore, B.S.: Cloning, sequencing and analysis of the enterocin biosynthesis gene cluster from the marine isolate '*Streptomyces maritimus*': evidence for the derailment of an aromatic polyketide synthase. Chem Biol 7 (2000) 943-55.

Pradella, S., Hans, A., Sproer, C., Reichenbach, H., Gerth, K. and Beyer, S.: Characterisation, genome size and genetic manipulation of the myxobacterium *Sorangium cellulosum* So ce56. Arch Microbiol 178 (2002) 484-92.

Schupp, T., Toupet, C., Cluzel, B., Neff, S., Hill, S., Beck, J.J. and Ligon, J.M.: A *Sorangium cellulosum* (myxobacterium) gene cluster for the biosynthesis of

the macrolide antibiotic soraphen A: cloning, characterization, and homology to polyketide synthase genes from actinomycetes. J Bacteriol 177 (1995) 3673-9.

Schupp, T., Toupet, C., Engel, N. and Goff, S.: Cloning and sequence analysis of the putative rifamycin polyketide synthase gene cluster from *Amycolatopsis mediterranei*. FEMS Microbiol Lett 159 (1998) 201-7.

Shah, S., Xue, Q., Tang, L., Carney, J.R., Betlach, M. and McDaniel, R.: Cloning, characterization and heterologous expression of a polyketide synthase and P-450 oxidase involved in the biosynthesis of the antibiotic oleandomycin. J Antibiot (Tokyo) 53 (2000) 502-8.

Smokvina, T., Henderson, D.J., Melton, R.E., Brolle, D.F., Kieser, T. and Hopwood, D.A.: Transposition of *IS*117, the 2.5 kb *Streptomyces coelicolor* A3(2) 'minicircle': roles of open reading frames and origin of tandem insertions. Mol Microbiol 12 (1994) 459-68.

Smokvina, T. and Hopwood, D.A.: Analysis of secondary integration sites for *IS*117 in *Streptomyces lividans* and their role in the generation of chromosomal deletions. Mol Gen Genet 239 (1993) 90-6.

Stanzak, P., Matsushima, P., Baltz, R.H. and Rao, R.N.: Cloning and expression in *Streptomyces lividans* of clustered erythromycin biosunthesis genes from *Streptomyces erythreus*. Bio/Technology 4 (1986) 229-232.

Strohl, W.R.: Compilation and analysis of DNA sequences associated with apparent streptomycete promoters. Nucleic Acids Res 20 (1992) 961-74.

Takano, E., White, J., Thompson, C.J. and Bibb, M.J.: Construction of thiostrepton-inducible, high-copy-number expression vectors for use in *Streptomyces* spp. Gene 166 (1995) 133-7.

Tang, L., Fu, H., Betlach, M.C. and McDaniel, R.: Elucidating the mechanism of chain termination switching in the picromycin/methymycin polyketide synthase. Chem Biol 6 (1999) 553-8.

Tang, L., Fu, H. and McDaniel, R.: Formation of functional heterologous complexes using subunits from the picromycin, erythromycin and oleandomycin polyketide synthases. Chem Biol 7 (2000a) 77-84.

Tang, L., Shah, S., Chung, L., Carney, J., Katz, L., Khosla, C. and Julien, B.: Cloning and heterologous expression of the epothilone gene cluster. Science 287 (2000b) 640-2.

Tang, L., Yoon, Y.J., Choi, C.Y. and Hutchinson, C.R.: Characterization of the enzymatic domains in the modular polyketide synthase involved in rifamycin B biosynthesis by *Amycolatopsis mediterranei*. Gene 216 (1998) 255-65.

Vahlensieck, H.F., Pridzun, L., Reichenbach, H. and Hinnen, A.: Identification of the yeast ACC1 gene product (acetyl-CoA carboxylase) as the target of the polyketide fungicide soraphen A. Curr Genet 25 (1994) 95-100.

Wall, D., Wu, S.S. and Kaiser, D.: Contact stimulation of Tgl and type IV pili in *Myxococcus xanthus*. J Bacteriol 180 (1998) 759-61.

Wilkinson, C.J., Frost, E.J., Staunton, J. and Leadlay, P.F.: Chain initiation on the soraphen-producing modular polyketide synthase from *Sorangium cellulosum*. Chem Biol 8 (2001) 1197-208.

Wilkinson, C.J., Hughes-Thomas, Z.A., Martin, C.J., Bohm, I., Mironenko, T., Deacon, M., Wheatcroft, M., Wirtz, G., Staunton, J. and Leadlay, P.F.: Increasing the efficiency of heterologous promoters in actinomycetes. J Mol Microbiol Biotechnol 4 (2002) 417-26.

Wu, K., Chung, L., Revill, W.P., Katz, L. and Reeves, C.D.: The FK520 gene cluster of *Streptomyces hygroscopicus* var. ascomyceticus (ATCC 14891) contains genes for biosynthesis of unusual polyketide extender units. Gene 251 (2000) 81-90.

Xiang, L. and Moore, B.S.: Inactivation, Complementation, and Heterologous Expression of *enc*P, a Novel Bacterial Phenylalanine Ammonia-Lyase Gene. J Biol Chem 277 (2002) 32505-9.

Xiang, L. and Moore, B.S.: Characterization of benzoyl coenzyme A biosynthesis genes in the enterocin-producing bacterium "*Streptomyces maritimus*". J Bacteriol 185 (2003) 399-404.

Xue, Q., Ashley, G., Hutchinson, C.R. and Santi, D.V.: A multiplasmid approach to preparing large libraries of polyketides. Proc Natl Acad Sci U S A 96 (1999) 11740-5.

Ziermann, R. and Betlach, M.C.: Recombinant polyketide synthesis in *Streptomyces*: engineering of improved host strains. Biotechniques 26 (1999) 106-10.

APPENDIX I. IMPROVEMENT IN CONJUGATION EFFICIENCY BETWEEN *SORANGIUM CELLULOSUM* SO CE 26 AND *ESCHERICHA COLI*

ABSTRACT

Sorangium cellulosum is a gram negative myxobacterium that forms myxospores under starvation conditions and also produces a significant number of structurally diverse, bioactive secondary metabolites. Despite the considerable interest in their natural products, there are few genetic tools available to investigate the molecular- and microbiology of *S. cellulosum*. DNA transfer in *S. cellulosum* So ce26 relies on conjugation followed by homologous recombination between DNA fragments on a transmissible plasmid and their chromosomal locus. The data presented here demonstrates an improvement in this system.

INTRODUCTION

The gram negative myxobacterium, *S. cellulosum*, has been shown to produce a wide variety of natural products including polyketides with unique and valuable commercial activities (reviewed by Reichenbach, 2001). Natural products produced by *S. cellulosum* include: soraphen A, an antifungal polyketide (Gerth et al., 1994); epothilone, a cytotoxic polyketide/non-ribosomal peptide hybrid molecule (Gerth et al., 1996); sorangicin, an antibiotic polyketide (Irschik et al., 1987) and ratjadon, an antifungal polyketide (Gerth et al., 1995). Despite the interest and effort in screening of isolates of *S. cellulosum* for biologically active natural products, few genetic tools are available to study the micro- and molecular biology of *S. cellulosum*. *S. cellulosum* strains grow slowly, and often poorly, requiring specialized media. *S.*

cellulosum strains also often fail to form colonies that are required for genetic and microbiological experimentation. Also, no plasmids have been found to replicate in *S. cellulosum* strains. Phleomycin resistance has been the only marker described in *S. cellulosum* until a recent report on the use of hygromycin in *S. cellulosum* So ce56 (Pradella et al., 2002). The only method available for stable DNA transfer in *S. cellulosum* is by homologous recombination between DNA fragments cloned into a transmissible plasmid and the homologous chromosomal locus (Jaoua et al., 1992). This system of conjugation coupled with homologous recombination results in a low efficiency. The data presented here demonstrates an improvement in conjugation efficiencies in *S. cellulosum* So ce26 compared to previously reported methods of conjugation used with this strain.

MATERIALS AND METODS

Bacterial strains. The strains SJ3 of *S. cellulosum* So ce26, *Eschericha coli* DH10B and *E. coli* ET12567 used were described in the "Materials and Methods" section of Chapter 2. *E. coli* ED8767 (*lac*Y1 *gln*V44 *gal*K2 *tyr*T58 *rec*A56 *met*B1 *hsd*S3) with helper plasmid pME305 was a gift from Thomas Schupp (Novartis, Basel, Switzerland).

DNA manipulations. Routine cloning and transformation with *E. coli* was carried out as described in the "Materials and Methods" section of Chapter 2. Restriction enzymes and T4 DNA ligase were from New England Biolabs (Beverly, MA) and used according to manufacturer's instructions. PCR reactions were carried out with *Pfu* polymerase (Stratagene, La Jolla, CA) as described in the "Materials and Methods" section of Chapter 2. Oligonucleotides were obtained from Integrated

DNA Technologies (Coralville, IA). Sequencing was done as described in the "Materials and Methods" section of Chapter 2.

Plasmid construction. Plasmid pCIBm2dhko contains a deletion of the active site in the dehydratase domain of module 2 in the soraphen polyketide synthase gene *sorA*. The PCR primers used to amplify the two DNA fragments that were used to generate pCIBm2dhko, incorporated a *Hind*III (underlined) site in place of the deleted active site and *Psp*OMI sites (in bold, *Not*I compatible) for cloning into pCIB132. The PCR primers for the first amplified fragment were 5'-AGA**GGGCCC**CCAATGGCACCTCCCAGAAGAAGGT and 5'-GACAAGCTT GATCGGCGAGCCACGGATGCTCTGA and cosmid pJL3 (from Chapter 3) was used as the template. The PCR primers used for the second amplified product were 5'-GGGAAGCTTTGCCGGGCACTGCCTTTCTCGAGCTT and 5'-AGC**GGGCCC** TCGACGAACTCGCGTGCCAAGCAGTC with cosmid pJL3 used as the template. The PCR products were digested with *Hind*III and *Psp*OMI and then cloned in a 3-way ligation into the *Not*I site of pCIB132.

The plasmid pCIBmko containing the *mglA* gene with an internal deletion, was constructed by 3-way ligation of two PCR products amplified from clone 2C9 (from Chapter 2) with pCIB132. This first PCR reaction was carried out with primers 5'-GCGCGCCGTAGACCTC and 5'-GGGAAGCTTCGGCTCGCGTCGTAG AAGA. The second PRC reaction was carried out with primers 5'-GGGAAGCTTGGTGTTCGACACGCTCAAG and 5'-CGGCCTCCTCGAGCAG. The two PCR amplification products were digested with *Hind*III and then cloned separately into *Hind*III-*Sma*I digested pBluescript SK II (+) (Stratgene, La Jolla, CA)

and sequenced. The two fragments were isolated from pBluescript SK II (+) by *Hind*III-*Not*I digestion and cloned in a 3-way ligation into the *Not*I site of pCIB132.

Media and growth conditions. The media and growth conditions used were described in the "Materials and Methods section of Chapter 2.

Conjugations. The conjugations were carried out as described in the "Materials and Methods" section of Chapter 2.

Efficiency determination. *E. coli* cultures were diluted in Luria broth prior to conjugations, plated on Luria agar, incubated at 37^{o}C, and the number of colony forming units determined. Colony forming unit determinations cannot be made with *S. cellulosum* So ce26 since it does not form colonies. Total protein content of *S. cellulosum* prior to conjugation, was determined and used to calculate the relative conjugation efficiency. To assay the protein content of *S. cellulosum* cultures, 100 µl of cells were pelleted and resuspended in 500 µl of cold sample buffer (100 mM phosphate buffer pH 7.5, 20 mM NaEDTA pH 8.0 and 10% v/v acetone) and sonicated (Branson Sonifier S450A Danbury, CT) 4 X 15 seconds on ice. Then 50 µl of 10% triton X-100 solution was added and the cell lysate was kept on ice for 15 minutes. The samples were centrifuged at 14,000 g and the supernatant was transferred to a fresh tube. The samples were assayed for protein content by using Bradford reagent in a Bio-Rad (Hercules, CA) protein assay kit following manufacturer's directions.

RESULTS

Gene transfer in *S. cellulosum* So ce26 to date has been demonstrated only through conjugation and homologous recombination. In a previous report, low efficiency conjugation was carried out using the *E. coli* donor strains W3101 or

ED8767 containing the helper plasmid pME305, that supplies the transfer functions *in trans* for the mobilization of the transmissible plasmid (Jaoua et al., 1992; Schupp et al., 1995). In this study, conjugation efficiencies were compared in experiments using *E. coli* donor strains ED8768 (pME305) and ET12567 (pUZ8002) and the results are presented in Table 1 and Table 2. The results show that conjugation efficiencies were improved by using *E. coli* donor strain ET12567 (pUZ8002). Conjugations were carried out targeting two different chromosomal loci and the results were similar in both cases. In the first experiment, pM2dko (Table 1) contained a DNA fragment from the *S. cellulosum* genome with a deletion in the dehydratase domain of module 2 in the soraphen PKS cluster. Experiments under the same conditions (Table 2), with plamid pCIBmko, containing a DNA fragment from the *S. cellulosum* genome with a deletion in the *mglA* homologous gene, also showed an increase in the number of transconjugants when strain ET12567 (pUZ8002) was used.

DISSCUSSION

Few genetic tools are available for experimentation with *S. cellulosum* So ce 26, the strain that produces soraphen A. Introduction of DNA was reported with this strain through conjugation and homologous recombination using the transmissible plasmid pCIB132 and *E. coli* donor strain W3101 or ED8767 (Jaoua et al., 1992; Schupp et al., 1995). The utilization of the donor strain ET12567 (pUZ8002) provided an improvement in the efficiency of transconjugations in experiments involving two unrelated chromosomal regions.

E. coli strain ET12567 has been in wide use as a donor in *Streptomyces* conjugation experiments as well as a host strain for the preparation of non-methylated DNA for *Streptomyces* protoplast transformation (Kieser and Hopwood, 1991; Flett et

al., 1997; Kieser et al., 2000). *Streptomyces* often contain powerful methylation-restriction systems and non-methylated DNA from ET12567 has been shown to increase efficiencies of transformations (Kieser et al., 2000). ET12567 is mutated in both Dam methylase (*dam*-13:Tn*9*) and Dcm methylase (*dcm*-6). Strain ET12567 also contains a *recF* mutation which is important in stabilization of plasmid DNA by minimizing recombination. These properties prompted the experimentation with ET12567 (pUZ8002) as a donor strain in conjugation with *S. cellulosum* So ce26.

To calculate the relative efficiency of conjugation, the number of transconjugants relative to the total cell protein of the *S. cellulosum* recipients was determined. In some of the experiments, the homologous recombination event in the transconjugants was verified by PCR or Southern hybridization (data not shown). In our laboratory, we have not seen any spontaneous phleomycin resistance with *S. cellulosum* So ce 26 strains under these experimental conditions, therefore, all phleomycin resistant colonies were counted as transconjugants. Since *S. cellulosum* So ce 26 does not form colonies suitable for enumeration, total protein was extracted from a portion of the culture just prior to conjugation and the total protein content was used in the calculation of efficiency of conjugation.

Although the methylation-restriction systems have not been studied in *S. cellulosum*, the improvement of conjugation efficiency is probably due to the non-metylated state of DNA from ET12567. More experimentation would be needed to definitely show that the non-methalated DNA from ET12567 is the cause for the higher conjugation efficiency. These experiments could include conjugations using otherwise isogenic strains of *E. coli* carrying combinations of *dcm dam* and *hsd* mutations. An another explanation to account for the improvement in conjugation

efficiency could be the presence of the helper plasmid pUZ8002. Comparing the conjugation efficiencies in experiments using otherwise isogenic *E. coli* strains that carry either the helper plasmid pUZ8002 or pME305 would determine if configuration of the helper plasmid influences conjugation efficiencies.

Improvements in transconjugation efficiency with *S. cellulosum* is very important since this is the only method available for DNA transfer in *S. cellulosum*. No plasmids have been found to replicate in *S. cellulosum* and electorporation of linear DNA fragments isolated from ET12567, with up to 4-kb of homologous DNA, were found not to mediate transformation of *S. cellulosum* So ce26 (unpublished data). The increase in conjugation efficiency has allowed the recovery of transconjugants in experiments where a reduced length of homologous DNA is supplied. This reduction in length requirement of homologous DNA allows a greater number of genes to be inactivated through homologous recombination thereby facilitating the determination of the function of genes.

ACKNOWLEDGMENTS

I would like to thank István Molnár for critical reading of this manuscript and helpful discussions. I would like to thank members of the Syngenta Sequencing group for their assistance with analysis of samples; Makoto Ono, An Hu and Racella McNair.

Table 1. Conjugation efficiency with plasmid pCIBm2dhko.

E. coli ED8767 (pME305) as donor

E.coli (cfu/mL)	mg of *S. cellulosum* protein	Number of *S. cellulosum* transconjugants	Relative efficiency[a]
$1.9x10^9$	31	10	0.32
$2.1x10^9$	27	3	0.11
$2.3x10^9$	33	0	0
$2.6x10^9$	26	14	0.54
			0.24[b]

[a]Relative efficiency = (number *S. cellulosum* transconjugants)/(mg of *S. cellulosum* protein)
[b]Average of relative efficiencies

E. coli ET12567 (pUZ8002) as donor

E.coli (cfu/mL)	mg of *S. cellulosum* protein	Number of *S. cellulosum* transconjugants	Efficiency
$6.4x10^8$	21	62	3.0
$5.2x10^8$	27	58	2.1
$5.5x10^8$	33	39	1.2
$4.9x10^8$	30	70	2.3
			2.2

Table 2. Conjugaion efficiency with plasmid pCIBmko.

E. coli ED8767 (pME305)

E. coli (cfu/mL)	mg of *S. cellulosum* protein	Number of *S. cellulosum* transconjuagates	Efficiency
$1.3x10^9$	22	12	0.55
$1.9x10^9$	28	20	0.71
			0.63

E. coli ET12567 (pUZ8002)

E. coli (cfu/mL)	mg of *S. cellulosum* protein	Number of *S. cellulosum* transconjugants	Efficiency
$6.3x10^8$	30	112	3.7
$5.7x10^8$	31	121	3.9
			3.8

REFRENCES

Flett, F., Mersinias, V. and Smith, C.P.: High efficiency intergeneric conjugal transfer of plasmid DNA from *Escherichia coli* to methyl DNA-restricting streptomycetes. FEMS Microbiol Lett 155 (1997) 223-9.

Gerth, K., Bedorf, N., Hofle, G., Irschik, H. and Reichenbach, H.: Epothilons A and B: antifungal and cytotoxic compounds from *Sorangium cellulosum* (Myxobacteria). Production, physico-chemical and biological properties. J Antibiot (Tokyo) 49 (1996) 560-3.

Gerth, K., Bedorf, N., Irschik, H., Hofle, G. and Reichenbach, H.: The soraphens: a family of novel antifungal compounds from *Sorangium cellulosum* (Myxobacteria). I. Soraphen A1 alpha: fermentation, isolation, biological properties. J Antibiot (Tokyo) 47 (1994) 23-31.

Gerth, K., Schummer, D., Hofle, G., Irschik, H. and Reichenbach, H.: Ratjadon: a new antifungal compound from *Sorangium cellulosum* (myxobacteria) production, physio-chemical and biological properties. J Antibiot (Tokyo) 48 (1995) 973-6.

Irschik, H., Jansen, R., Gerth, K., Hofle, G. and Reichenbach, H.: The sorangicins, novel and powerful inhibitors of eubacterial RNA polymerase isolated from myxobacteria. J Antibiot (Tokyo) 40 (1987) 7-13.

Jaoua, S., Neff, S. and Schupp, T.: Transfer of mobilizable plasmids to *Sorangium cellulosum* and evidence for their integration into the chromosome. Plasmid 28 (1992) 157-65.

Kieser, T., Bibb, M.J., Buttner, M.J., Chater, K.F. and Hopwood, D.A.: Practical *Streptomyces* Genetics. The John Innes Foundation, Norwich, England, 2000.

Kieser, T. and Hopwood, D.A.: Genetic manipulation of *Streptomyces*: integrating vectors and gene replacement. Methods Enzymol 204 (1991) 430-58.

Pradella, S., Hans, A., Sproer, C., Reichenbach, H., Gerth, K. and Beyer, S.: Characterisation, genome size and genetic manipulation of the myxobacterium *Sorangium cellulosum* So ce56. Arch Microbiol 178 (2002) 484-92.

Reichenbach, H.: Myxobacteria, producers of novel bioactive substances. J Ind Microbiol Biotechnol 27 (2001) 149-56.

Schupp, T., Toupet, C., Cluzel, B., Neff, S., Hill, S., Beck, J.J. and Ligon, J.M.: A *Sorangium cellulosum* (myxobacterium) gene cluster for the biosynthesis of the macrolide antibiotic soraphen A: cloning, characterization, and homology to polyketide synthase genes from actinomycetes. J Bacteriol 177 (1995) 3673-9.