2011

COMBINATORIAL BIOSYNTHETIC DERIVATIZATION OF THE ANTITUMORAL AGENT GILVOCARCIN V

Micah Douglas Shepherd
University of Kentucky, shepherd.micah@gmail.com

ABSTRACT OF DISSERTATION

Micah Douglas Shepherd

The Graduate School
University of Kentucky
2011

COMBINATORIAL BIOSYNTHETIC DERIVATIZATION OF THE ANTITUMOR AGENT GILVOCARCIN V

ABSTRACT OF DISSERTATION

A dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the College of Pharmacy at the University of Kentucky

By
Micah Douglas Shepherd

Stanton, Kentucky

Director: Dr. Jürgen Rohr, Professor of Pharmaceutical Sciences

Lexington, Kentucky

2011

ABSTRACT OF DISSERTATION

COMBINATORIAL BIOSYNTHETIC DERIVATIZATION OF THE ANTITUMORAL AGENT GILVOCARCIN V

Gilvocarcin V (GV), the principal product of *Streptomyces griseoflavus* Gö 3592 and other *Streptomyces* spp., is the most prominent member of a distinct class of antitumor antibiotics that share a polyketide derived coumarin-based aromatic core. GV and other members of this class including polycarcin V from *Streptomyces polyformus*, often referred to as gilvocarcin-like aryl *C*-glycosides, are particularly interesting because of their potent bactericidal, virucidal and antitumor activities at low concentrations while maintaining low *in vivo* toxicity. Although the precise molecular mechanism of GV bioactivity is unknown, gilvocarcin V has been shown to undergo a photoactivated [2+2] cycloaddition of its vinyl side chain with thymine residues of DNA in near-UV or visible blue light. In addition, GV was shown to selectively crosslink histone H3 with DNA, thereby effectively disrupting normal cellular processes such as transcription. Furthermore, GVs ability to inhibit topoisomerase II has also been attributed as a mechanism of action for gilvocarcin V activity. The excellent antitumor activity, as well as an unprecedented structural architecture, has made GV an ideal candidate for biosynthetic studies toward the development of novel analogues with improved pharmacological properties. Previous biosynthetic research has identified several candidate genes responsible for key steps during the biosynthesis of gilvocarcin V including an oxygenase cascade leading to C-C bond cleavage, methylations, lactone formation, *C*-glycosylation and vinyl side chain formation.

In this study, we further examined two critical biosynthetic transformations essential for the bioactivity of gilvocarcin V, namely starter unit incorporation and *C*-glycosylation, through the following specific aims: 1) creation of functional chimeric *C*-glycosyltransferases through domain swapping of gilvocarcin-like glycosyltransferases and identification and evaluation of the donor substrate flexibility of PlcGT, the polycarcin V pathway specific *C*-glycosyltransferase; 2) creation of a library of *O*-methylated-L-rhamnose analogues of polycarcin V for structure activity relationship studies; 3) identification of the role of GilP and GilQ in starter unit specificity during gilvocarcin V biosynthesis; and 4) creation of a plasmid based approach in which selective gilvocarcin biosynthetic genes were utilized to produce important gilvocarcin intermediates for further *in vivo* and *in vitro* experimentation.

KEYWORDS: gilvocarcin V, polycarcin V, combinatorial biosynthesis, glycosyltransferase, methyltransferase

Micah Shepherd

2011

COMBINATORIAL BIOSYNTHETIC DERIVATIZATION OF THE ANTITUMORAL AGENT GILVOCARCIN V

By

Micah Douglas Shepherd

Dr. Jürgen Rohr
(Director of Dissertation)

Dr. Jim Pauly
(Director of Graduate Studies)

2011

RULES FOR THE USE OF DISSERTATIONS

Unpublished dissertations submitted for the Doctor's degree and deposited in the University of Kentucky Library are as a rule open for inspections, but are to be used only with due regard to the rights of the authors. Bibliographical references may be noted, but quotations or summaries of parts may be published only with the permission of the author, and with the usual scholarly acknowledgments.

Extensive copying or publication of the dissertation in whole or in part also requires the consent of the Dean of the Graduate School of the University of Kentucky

A library that borrows this dissertation for use by its patrons is expected to secure the signature of each user.

Name	Date

DISSERTATION

Micah Douglas Shepherd

The Graduate School

University of Kentucky

2011

COMBINATORIAL BIOSYNTHETIC DERIVATIZATION OF THE ANTITUMORAL AGENT GILVOCARCIN V

DISSERTATION

A dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the College of Pharmacy at the University of Kentucky

By
Micah Douglas Shepherd

Stanton, Kentucky

Director: Dr. Jürgen Rohr, Professor of Pharmaceutical Sciences

Lexington, Kentucky

2011

ACKNOWLEDGMENTS

The following dissertation was made possible by the guidance and support from several people. First, I would like to acknowledge my mentor, Dr. Jürgen Rohr, for his intellectual and personal guidance, encouragement and unwavering support throughout my graduate studies. I would also like to thank my committee members Dr. Steven Van Lanen, Dr. Hsin-Hsiung Tai and Dr. Robert Perry for their invaluable advice and assistance throughout my studies. I am very thankful for the generosity of Dr. Yi Tang, Dr. Hung-Wen Liu, Dr. Jose Salas, Dr. Andreas Bechthold, Dr. Chaitan Khosla, Dr. Christian Hertweck and Dr. Mark Zabriskie for supplying important plasmids and genes used throughout my studies.

In addition, I would like to thank all the past and present members of Dr. Rohr's lab who have given me technical advice and support over the years. These people have made studying at the University of Kentucky a wonderful experience and include Dr. Madan Kharel, Dr. Pallab Pahari, Dr. Miranda Beam, Mike Smith, Eric Nybo, Dr. Guo-Jun Wang, Dr. Khaled Shaaban, Dr. Mary Bosserman, Dr. Lily Zhu, Dr. Irfan Baig, Dr. Tao Liu and Dr. Mohamed Abdelfattah.

Finally, I would like to acknowledge my amazing family who has always been supportive of my ambitions. My sincere thanks go to my parents, Doug and Kay Shepherd, as well as my sister, Brittany Shepherd, for their encouragement and helping hand. Last but certainly not least, I wish to express my deepest gratitude to my wife, Megan Shepherd, for her love, understanding and unparalleled support throughout these past years. Without her strength and patience this would not have been possible.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

4,6-DH	4,6-dehydratase
aa	amino acid
ACoA	acetyl-coenzyme A
ACP	acyl carrier protein
apo-ACP	inactive acyl carrier protein
ARO	aromatase
AT	acyltransferase
bp	base pair
BSA	bovine serum albumin
cGT	chimeric glycosyltransferase
CLF	chain length factor (KS_β)
CoA	coenzyme A
CYC	cyclase
CYP450	cytochrome P-450 monooxygenases
Da	dalton
DH	dehydratase
DNA	deoxyribonucleic acid
ER	enoyl reductase
ESI-MS	electrospray ionization mass spectrometry
FAD	flavin adenine dinucleotide
FAS	fatty acid synthase
FLP	flippase
FMN	flavin mononucleotide
FRT	flippase recognition site
GT	glycosyltransferase
GV	gilvocarcin V
HMBC	heteronuclear multiple bond coherence
holo-ACP	activated acyl carrier protein
HPLC-MS	high-performance liquid chromatography-mass spectroscopy
IMAC	immobilized metal ion affinity chromatography
IPTG	isopropyl β-D-1-thiogalactopyranoside
KR	keto reductase
KS	ketoacyl synthase
KS III	β-ketoacyl:acyl carrier protein synthase III
LB	lysogeny broth
MAT	malonyl-coenzyme A:acyl carrier protein transacylase
MCAT	malonyl-coenzyme A:acyl carrier protein transacylase
MCoA	malonyl-coenzyme A
MCS	multiple cloning site
MMCoA	methylmalonyl-CoA
NADPH	nicotinamide adenine dinucleotide phosphate, reduced form
NDP	nucleotidyl diphosphate
NMR	nuclear magnetic resonance
NT	nucleotidyltransferase
ORF	open reading frame

OTC	over-the-counter
PCoA	propionyl-coenzyme A
PCR	polymerase chain reaction
PKS	polyketide synthase
post-PKS	post-polyketide synthase
PPTase	phosphopantetheinyl transferase
RBS	ribosomal binding site
SAH	*S*-adenosylhomocysteine
SAM	*S*-adenosylmethionine
SAR	Structure activity relationship
SDS-PAGE	sodium dodecylsulfate-polyacrylamide gel electrophoresis
STS	stilbene synthase
TCA	trichloroacetic acid
TFA	trifluoroacetic acid
UV-Vis	ultraviolet-visible spectroscopy

Chapter 1: Introduction

Natural product sources, whether from flora or fauna, have long been explored for their medicinal properties. Prehistoric medicine, for example, involved a primitive form of phytotherapy in which through trial and error materials from flora were recognized for their beneficial effects. The natural mummy of a prehistoric man known as the Similaun Man was found with *Piptoporus betulinus*, a type of bracket fungus whose oil shows antiparasitic properties, in his possessions.[1] It is believed that the Similaun Man, who walked over 5,300 years ago, was ingesting *Piptoporus betulinus* to combat an intestinal whipworm infection found upon examination of his body. The benefits of such natural products were passed down, by word of mouth, through generations until the advent of writing, in which we can trace medicinal natural product use back to ancient civilizations such as the Babylonian, Egyptian and Chinese cultures. In this context, modern medicine is the product of an amalgamation of thousands of years worth of natural product research.

Today, medicinal natural products are classified into six broad categories based on their building blocks and biosynthetic routes including: fatty acids and polyketides (**1** and **2**), terpenoids and steroids (**3** and **4**), alkaloids (**5** and **6**), phenylpropanoids (**7** and **8**), specialized amino acids and peptides (**9** and **10**), and specialized carbohydrates (**11** and **12**) (**Figure 1**).[2] Typically, bioactive natural products are found as secondary metabolites of specific organisms or groups of organisms. Secondary metabolites are natural products that do not directly affect the normal growth, development or reproduction of the producing organism. This is in contrast to primary metabolites which are characterized as compounds that occur in all cells and play a central role in metabolism and reproduction of those cells. Examples of primary metabolites include nucleic acids, proteinogenic amino acids (aa) and sugars.

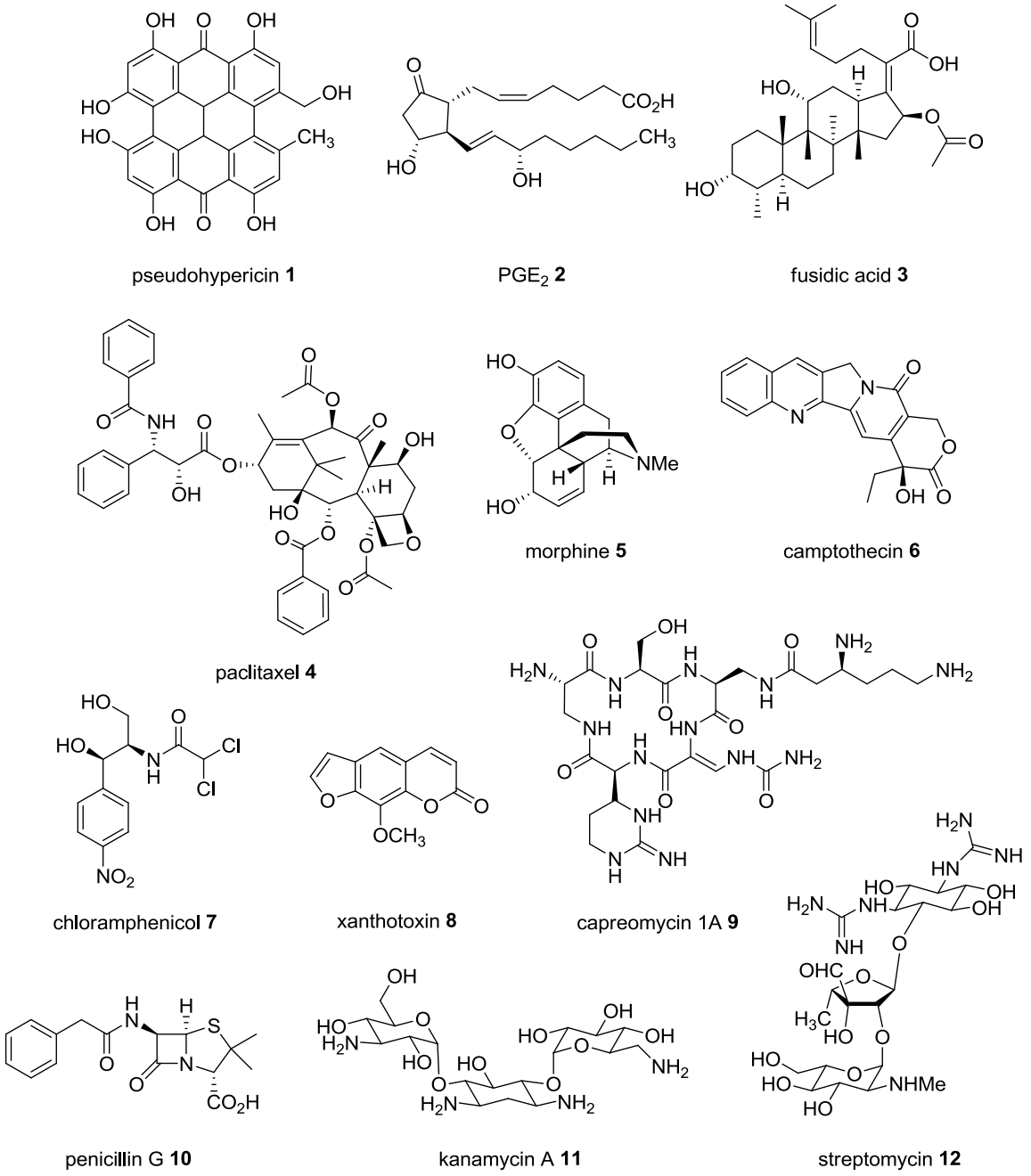


Figure 1. Medicinal natural products.

Polyketides

Interestingly, about half of all discovered biologically active secondary metabolites are produced by a group of Gram-positive bacteria known as actinomycetes.[3] Particularly, the genus *Streptomyces* is the most prolific producer of pharmaceutically and agriculturally relevant natural products covering all structural categories of secondary

metabolites. Of these, the polyketides have been the most extensively studied class of secondary metabolites. Polyketides comprise one of the largest groups of bioactive natural products and encompass several structurally diverse groups including polyethers, polyphenols, polyenes, macrolides and enediynes as exemplified by monensin (**13**), benastatin A (**14**), candicidin (**15**), spiramycin (**16**) and C-1027 (**17**), respectively (**Figure 2**).[4]

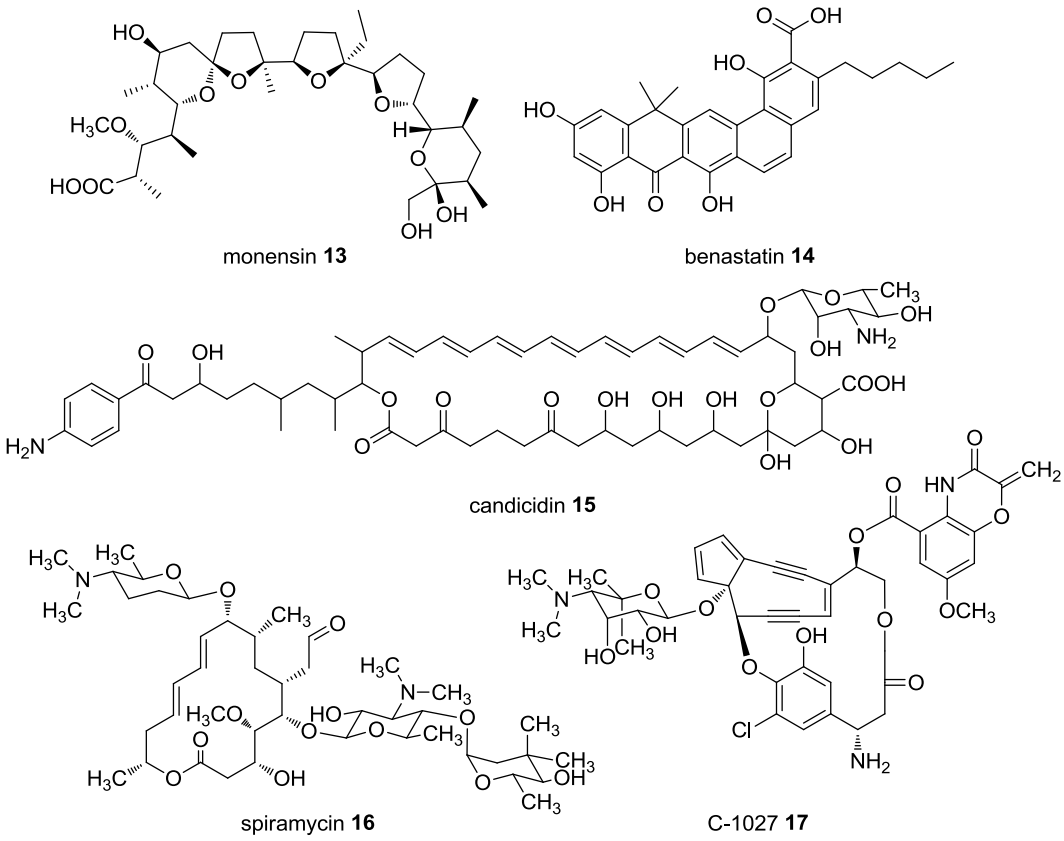


Figure 2. Structural diversity of polyketides.

Streptomycetes are not the only producer of polyketides. In fact, examples of polyketides can be found throughout prokaryotic and eukaryotic life including bacteria, fungi, plants, and protists. Some polyketides have been reported from insects, mollusks and sponges; however, these have been subsequently attributed to bacterial symbionts. The exact roles of these polyketides in their original biological hosts are widely

unknown, but it is believed that some may be involved in pigmentation, pathogen-defense, immune-response and symbiosis.[5] Many polyketides, or their derivatives, have become important commercial chemicals such as the antibiotics oxytetracycline (**18**), erythromycin A (**19**) and rifampicin (**20**); the immunosuppressant fujimycin (**21**), the antifungal amphotericin B (**22**); the antitumoral agents mithramycin (**23**) and epirubicin (**24**); the antiparasitic and insecticide avermectin B_{1a} (**25**) and spinosyn (**26**); and the anticholesterolemic lovastatin (**27**).[2] In contrast, other polyketides have been found to serve as virulence factors and potent toxins such as mycolactone (**28**) and aflatoxin B1 (**29**) (**Figure 3**).

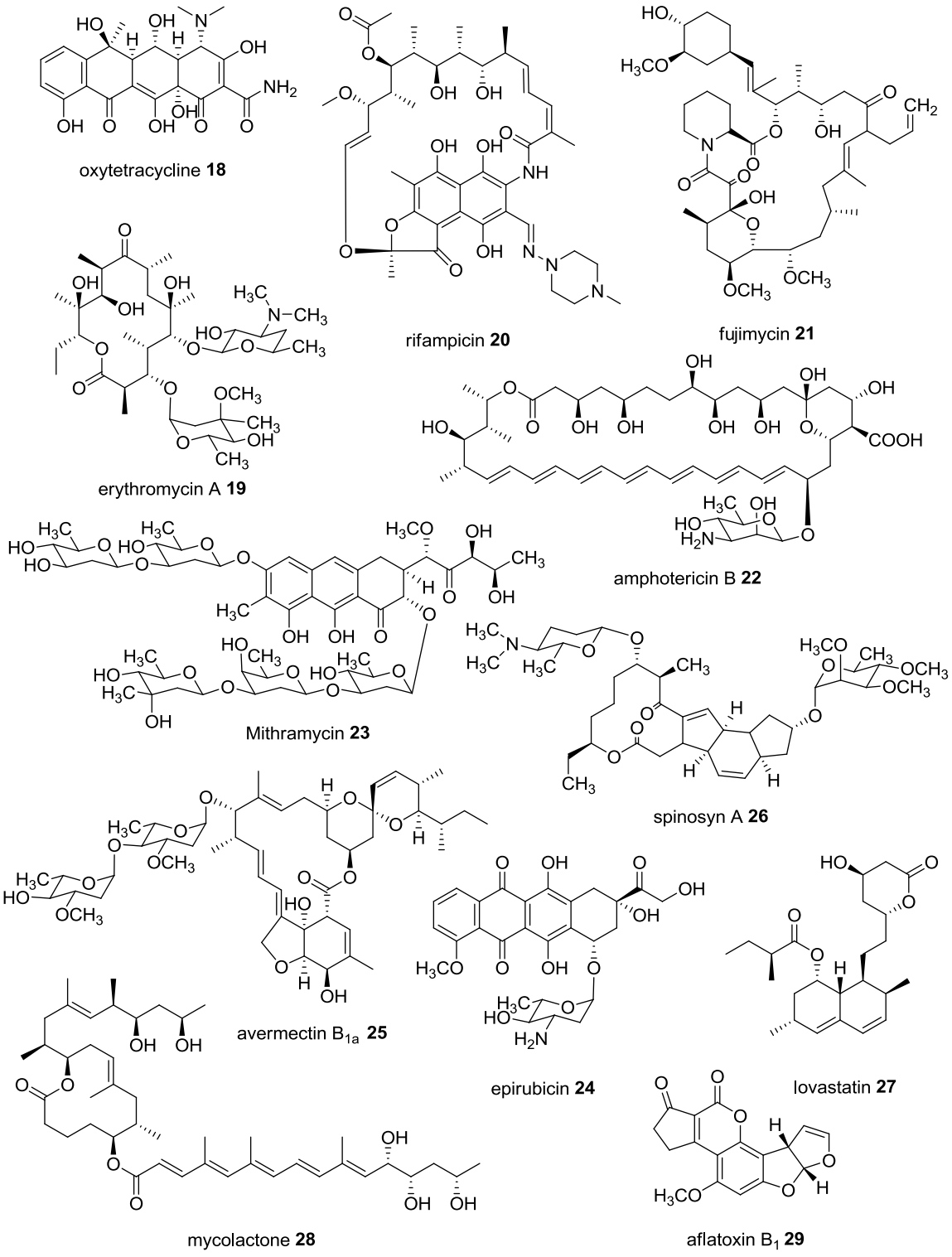


Figure 3. Bioactive polyketides.

Despite the enormous success of pharmaceutical polyketides, there is a growing need for the generation of novel drug leads to combat the rapid development of drug-resistant pathogens and emerging infectious microorganisms. The structural complexity of polyketides, however, presents an extremely difficult challenge for total synthetic strategies toward a majority of polyketide scaffolding despite their biogenesis through simple acetate derived building blocks.[4] Extensive research on polyketide natural product biosyntheses have provided a plethora of genetic and biochemical information that allow for the engineered generation of novel "unnatural" polyketide derivatives through a process known as combinatorial biosynthesis. This approach utilizes genetic engineering and various strategies including gene inactivation, heterologous gene expression and mutasynthesis to modify the natural biosynthetic pathway toward the generation of novel compounds.[6-7]

As combinatorial biosynthesis requires a deep fundamental understanding of the processes involved in the chemical modifications of natural products, the biosynthesis and common decorating reactions of polyketides, typically referred to as polyketide synthase (PKS) and post-polyketide synthase (post-PKS) reactions, will be described below.

Biosynthesis of polyketides

Polyketide biosynthesis is reminiscent of fatty acid biosynthesis as both pathways share a common biosynthetic logic. Generally derived from simple precursors such as acetyl-coenzyme A (CoA) and malonyl-CoA (MCoA), both polyketides and fatty acids are generated through repetitive decarboxylative Claisen thioester condensations of an activated acyl starter unit with MCoA or MCoA-derived extender units.[8] This process typically requires the activity of a β-ketoacyl synthase (KS), an activated acyl carrier protein (*holo*-ACP) and a malonyl/acyltransferase (MAT/AT). In fatty acid biosynthesis, every chain elongation step catalyzed by the fatty acid synthase (FAS) is followed by β-ketoreduction, dehydration and enoyl reduction steps to produce a fully saturated backbone (**Figure 4**, pathway A). This β-keto processing is referred to as the reduction cycle and is formed through the concerted action of a ketoreductase (KR), dehydratase (DH) and an enoyl reductase (ER), respectively. In polyketide biosynthesis, however, the

reduction cycle is optional, and can be partially (**Figure 4**, pathway B) or fully (**Figure 4**, pathway C) omitted before the next round of elongation. This in addition to the ability of the polyketide synthases (PKSs) to utilize a larger pool of acyl-CoA substrates set FAS and PKS systems apart. Nevertheless, both pathways repeat their respective elongation/reduction cycles until a predetermined chain length is obtained at which point the thioester-bound substrates are removed from the enzyme complex and are further modified through post-PKS reactions.

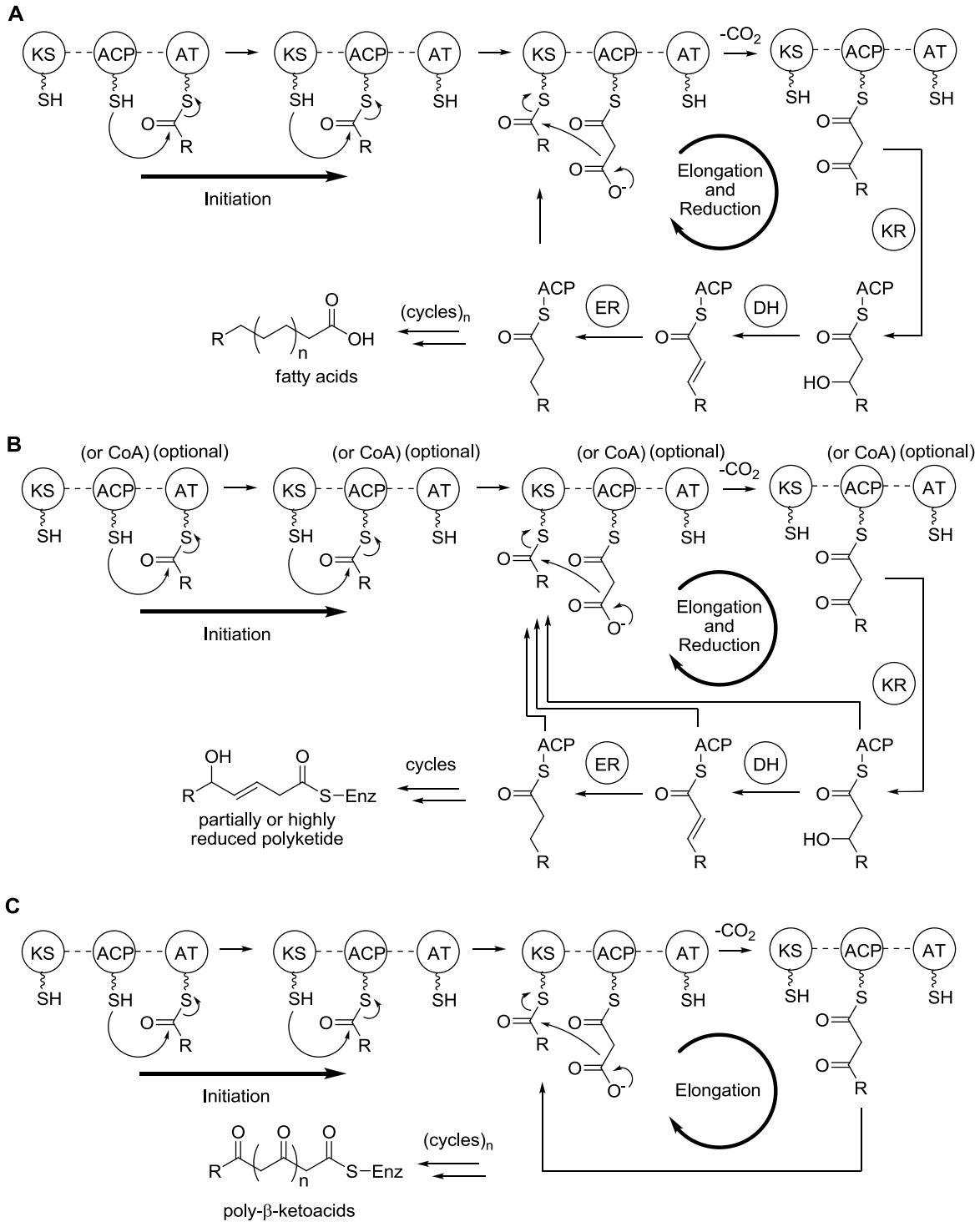


Figure 4. Biosynthetic route toward fatty acids (A), partially or highly reduced polyketides (B) and unreduced polyketides (C).

Polyketide synthases are classified into three categories, type I, type II and type III, based on their enzymatic properties and architecture.[9-10] Type I PKSs are large multifunctional enzymes consisting of distinct modules harboring a set of activities responsible for one (modular) or more (iterative) cycles of chain elongation.[11] The modular type I PKS resembles an assembly line in which the growing polyketide is passed from one module to the next. Each module contains an individual KS, AT and ACP domain responsible for the stepwise decarboxylative extension described above. In addition, each module also contains optional β-keto processing domains (KR, DH and ER) that will determine the extent to which the newly formed β-keto group is modified. Each module is used once as the growing chain is passed along the PKS assembly line. In this context, through bioinformatical analysis of the modules encoded within the multifunctional enzyme it is possible to predict the core structure, including the degree of β-keto reduction, produced by a modular type I PKS.

The biosynthesis of pikromycin (**30**) follows that of a typical modular type I PKS (**Figure 5**). The pikromycin (*pik*) polyketide synthase locus consists of five open reading frames (ORFs), *pikAI*, *pikAII*, *piKAIII*, *pikAIV* and *pikAV*, consisting of six modules, a loading module and an independent thioesterase.[12] The loading module initiates pikromycin biosynthesis by loading the starter unit propionyl-CoA (PCoA). The ketosynthase domain of module one self acylates with the extender unit methylmalonyl-CoA (MMCoA) and undergoes decarboxylative condensation with the propionyl-ACP of module L. The resulting diketide undergoes ketoreduction by the KR domain of module one installing the hydroxyl group at the C-3 position of the growing polyketide. The AT domain then transfers the nascent polyketide to the ACP domain of module one and the process continues down the subsequent modules. The amount of reduction varies from module to module as does the choice for extender unit. The polyketide is elongated until it reaches module six which carries a thioesterase responsible for catalyzing an intramolecular cyclization which forms narbonolide (**31**), a 14-membered macrolide. Interestingly, the pikromycin pathway contains an additional thioesterase, PikAV, that was found to be non-essential for lactone formation.[13] Further post-PKS modifications by a glycosyltransferase (DesVII, produces **32**) and cytochrome P-450 (CYP450) monooxygenase (PikC) creates pikromycin.[14]

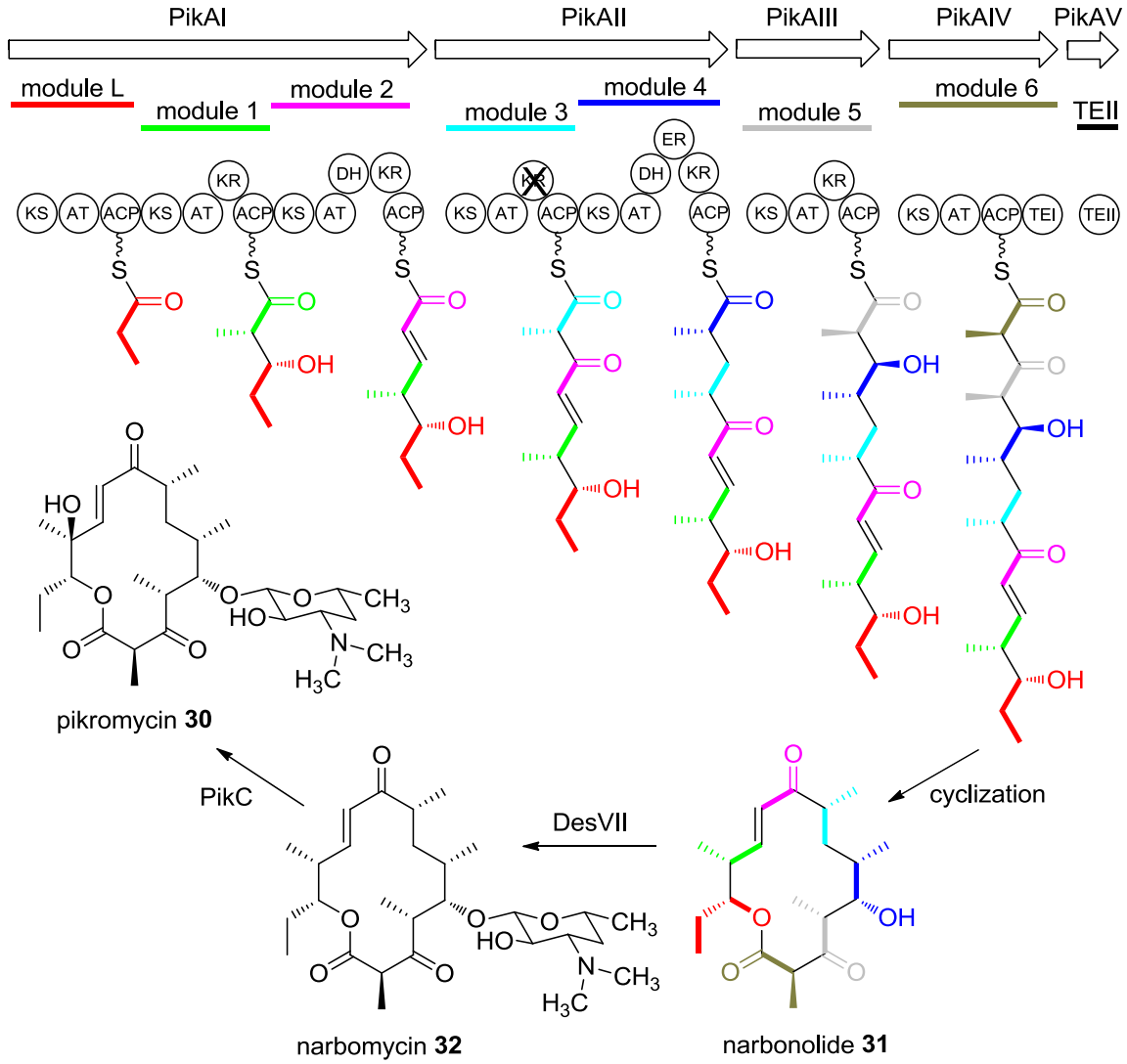


Figure 5. The biosynthetic pathway of pikromycin.

Not all type I PKSs behave the same, as examples of skipping and stuttering events within this modular type I PKS system have been reported which deviates from the linearity with which it is usually associated.[15] In addition there are examples of AT-less, and iteratively acting type I PKSs.[16] The iteratively acting type I PKS, usually found in fungi, differs from the modular system in that a single multifunctional enzyme consisting of a KS, AT, ACP, KR, DH and ER domains is used repeatedly throughout the elongation of the polyketide. As in modular type I PKSs, the degree of β-ketoreduction

can vary in each extended unit. The ability of iterative PKSs to control varying degrees of β-ketoreduction remains to be elucidated. The anticholesterolemic lovastatin is an example of an iteratively acting type I PKS. In lovastatin biosynthesis, one large multifunctional enzyme containing KS, AT, DH, MT, KR and ACP domains (LovB), in conjunction with the free standing ER (LovC), is responsible for polyketide propagation and reduction (**Figure 6**).[17]

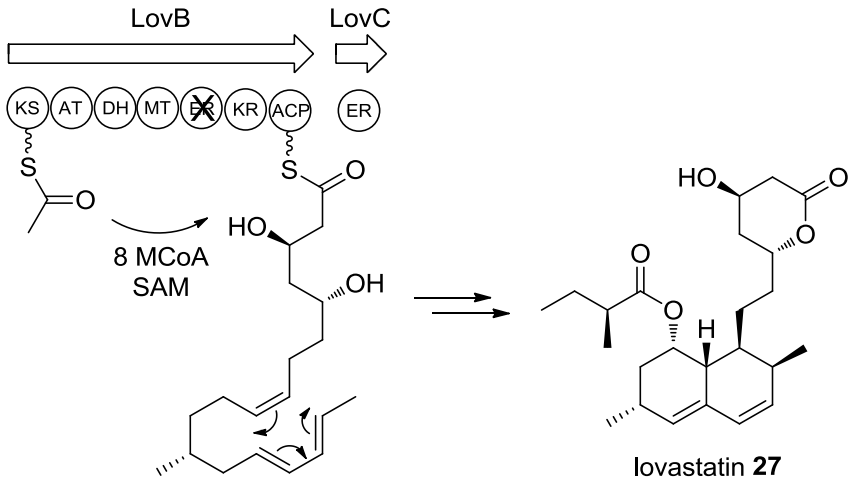


Figure 6. Biosynthetic pathway of lovastatin.

Type II PKS systems are found exclusively in prokaryotes.[4] They are usually observed in actinomycetes; however, a few examples of Gram-negative bacteria containing type II PKSs are known.[18-19] Type II PKSs are made up of individual iteratively acting enzymes, expressed from distinct genes, that form a multi-enzyme complex for polyketide assembly. Type II PKSs differ from type I PKSs in that the former does not include a β-ketoreduction cycle, and therefore produces a highly reactive poly-β-keto-thioester intermediate. The minimal proteins required for type II PKS functionality typically consists of two ketosynthase units (KS_α and KS_β or chain length factor (CLF)) and an ACP which make up the "minimal PKS".[20-21] Typically in type II PKSs polyketide formation is initiated by the loading of an activated ACP with malonyl-CoA by a malonyl-CoA:ACP transacylase (MCAT), borrowed from fatty acid

biosynthesis, and subsequent decarboxylation to form an acetate-ACP species. The acetate starter unit is then transferred from the ACP to the KS_α subunit (**Figure 4**, C). Again, an additional unit of malonyl-CoA is loaded to the ACP via an MCAT which undergoes a decarboxylative Claisen condensation with the acetate primed KS_α catalyzing the first elongation step. The ACP therefore serves as the anchor for the growing polyketide. The newly formed diketide is then transferred from the ACP back to the KS_α for additional rounds of elongation. This process is repeated until the desired polyketide length is achieved, which is believed to be determined by the KS_β subunit, hence the chain length factor (CLF).[22] The minimal PKS is also accompanied by additional subunits including ketoreductases, cyclases (CYC) and aromatases (ARO) that determine the folding, cyclization and aromatization pattern of the nascent poly-β-keto-thioester intermediate. These reactions provide the basis for structural diversity of type II PKS derived polyketide cores as observed by linear (anthracyclines, benzoisochromanequinones, tetracyclines, aureolic acids, pluramycins), angular (angucyclines and pentangular polyphenols) and discoid (resistomycins) polyphenols.[4]

The biosynthesis of tetracenomycin C (**33**) for example is a model type II PKS (TcmK, L and M) that utilizes one acetate and nine malonate subunits to form a decaketide intermediate that is further modified by the cyclase TcmN to produce TCM F2 (**34**) (**Figure 7**).[21,23] The inclusion of TcmJ, believed to be a cyclase, was shown to increase the production of TCM F2 in cell free assays, but was found to be non-essential for the continued production of **34**. TCM F2 is converted to TCM F1 (**35**) by the cyclase TcmI to form the fourth fused ring.[24-25] Additional modifications by methyltransferases and oxygenases convert TCM F1 into tetracenomycin C (**33**).[26-28]

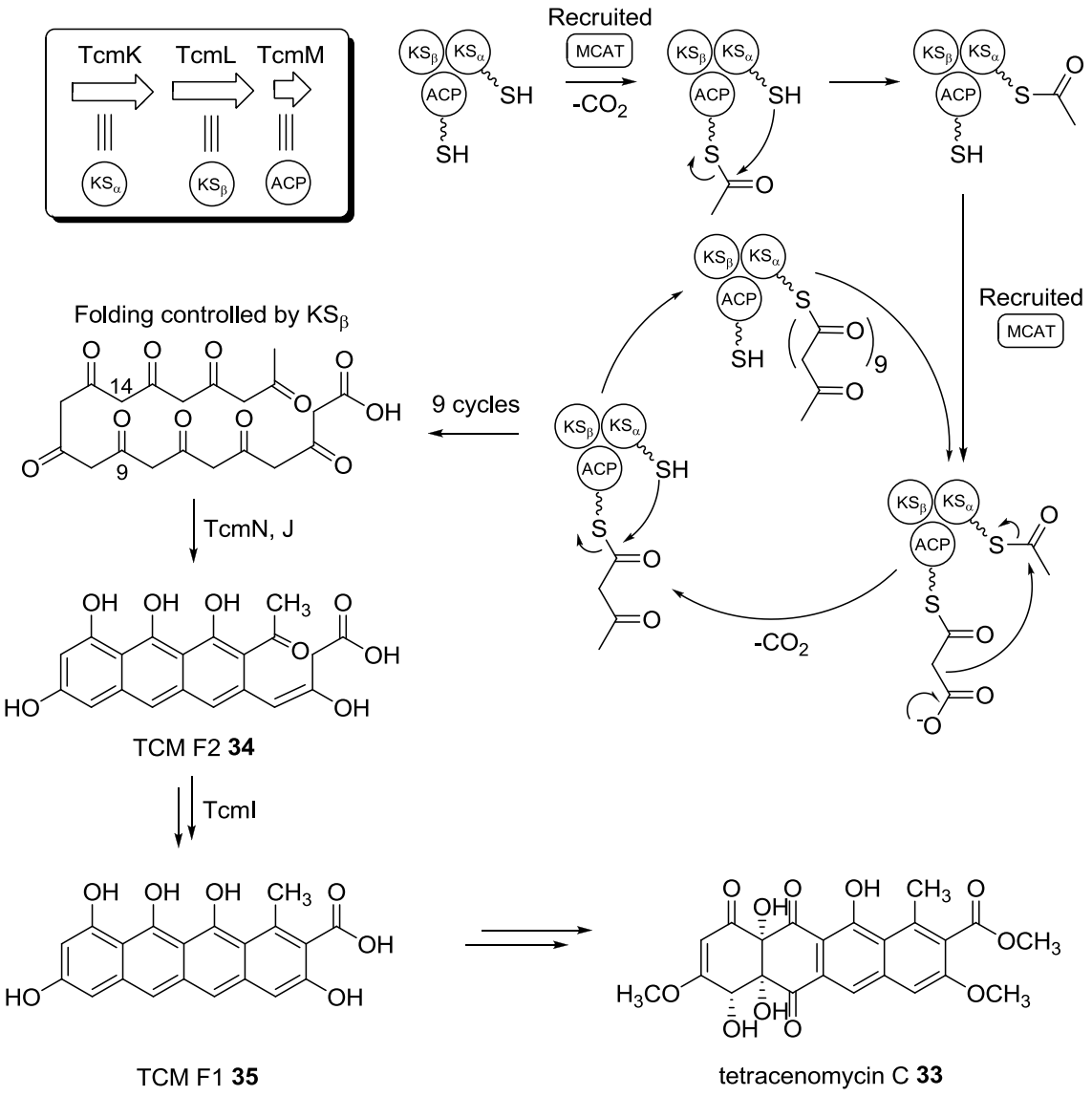


Figure 7. Biosynthetic pathway of tetracenomycin C.

The last class of PKS enzymes is the type III PKSs, or chalcone/stilbene synthase-like PKSs. Type III PKSs can be found in plants, bacteria and fungi as multifunctional condensing enzymes that do not utilize ACP-bound substrates for extension.[4] Instead, they utilize acyl-CoA substrates directly, as observed in resveratrol (**36**) biosynthesis in which p-coumaroyl-CoA is extended with three MCoA subunits (**Figure 8**).[29] The type III PKS then catalyzes a specific C2-C7 cyclization producing **36**.

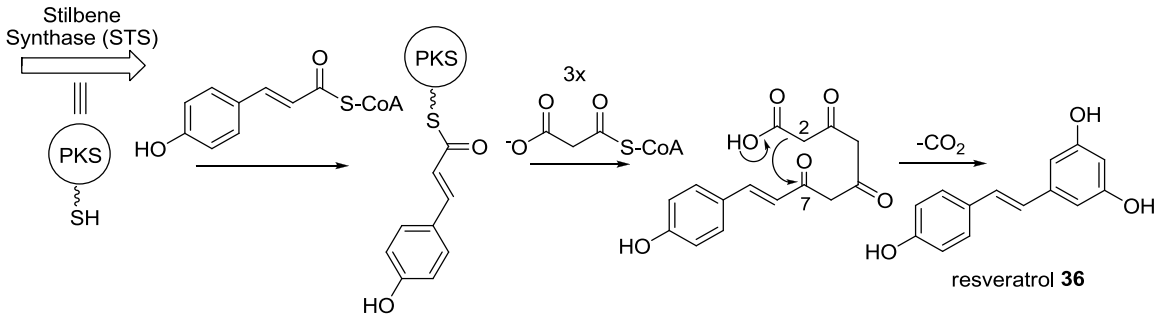


Figure 8. Biosynthetic pathway of resveratrol.

In addition to the three general classes of PKSs, there are also examples of mixed PKS systems. These include type I-type II, type I-type III, FAS-PKS and PKS-NRPS (non-ribosomal peptide synthetase) hybrids as exemplified by hedamycin (**37**), kendomycin (**38**), DIF-1 (**39**) and cryptophycin (**40**), respectively (**Figure 9**).[4,30]

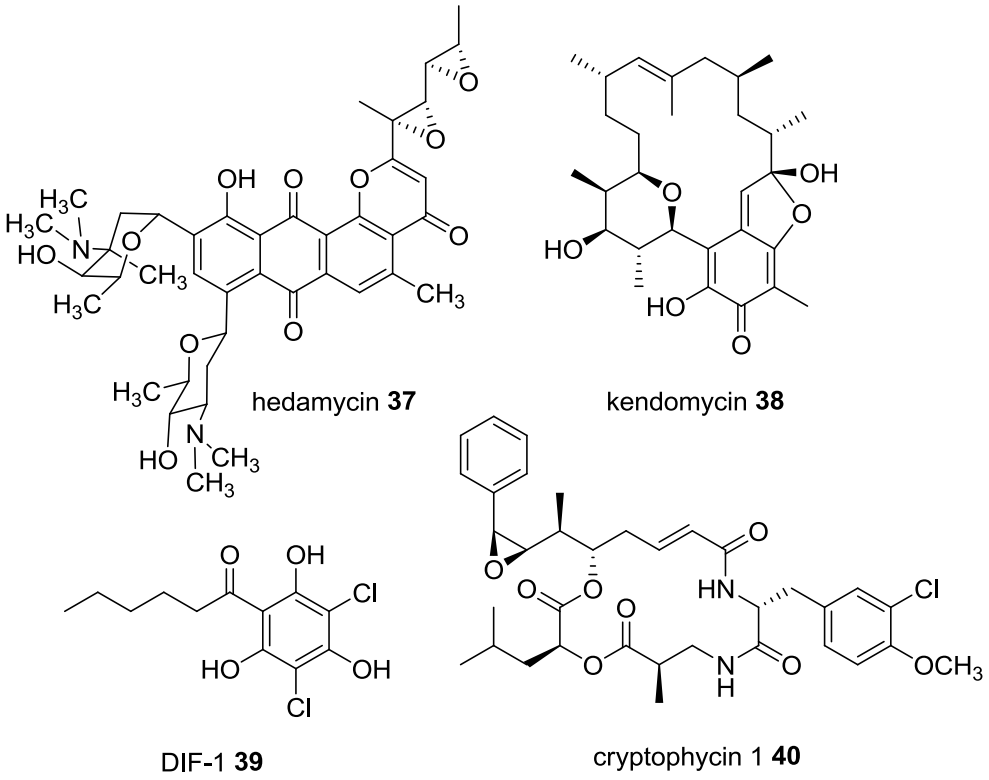


Figure 9. Polyketides derived from mixed PKS systems.

Polyketide tailoring enzymes

A majority of polyketides are further functionalized after their initial propagation by a series of modifying enzymes referred to as post-PKS enzymes including oxygenases, glycosyltransferases, methyltransferases and deoxysugar biosynthetic genes.[6-7] These tailoring reactions can lead to drastic alterations in the physio-chemical properties of compounds and are often responsible for "activating" polyketides. Many examples exist in which the polyketide intermediates produced by their respective PKSs are proven inactive until modifying enzymes, most commonly oxygenases and glycosyltransferases, tailor the polyketide into a bioactive compound. Specifically, the early PKS intermediates of erythromycin A (**19**) and epirubicin (**24**) have little to no activity until they are further modified by their post-PKS associated enzymes.[31] Given the importance of these post-PKS reactions they have been the subject of extensive combinatorial biosynthetic investigations to allow for the production of modified polyketides with improved pharmacological properties. This approach has led to the development of hundreds of modified polyketide derivatives, many with improved bioactivity profiles.[6-7] Common post-PKS tailoring reactions, which are observed in many natural product biosynthetic pathways, are briefly discussed below.

Glycosylation

Glycosylation is recognized as one of the most important post-PKS reactions because of its ability to install bioactivity to an otherwise inactive biomolecule. In addition it may serve as a crucial self-resistance mechanism as observed in macrolide biosynthesis.[32-33] Enzymatic glycosylation involves a glycosyltransferase (GT) which attaches an NDP-activated sugar, as the donor substrate, to an acceptor substrate.[34] There are a few examples of GTs requiring a second enzyme known as a helper enzyme to efficiently glycosylate an acceptor substrate.[35] In these systems it remains unknown the exact role these auxiliary proteins play in order to facilitate glycosylation.

Successful glycosylation of an acceptor substrate produces a glycosidic bond between the acceptor substrate and the donor substrate. These can be *O*-, *N*-, *S*- or *C*-glycosidic bonds depending on the functionality of the GT. *O*-glycosides are by far the most prevalent with *N*-, *S*- and *C*-glycosides making up only a minority of all glycosides

produced by glycosyltransferases in secondary metabolites. Conventionally, glycosylation is thought of as being unidirectional; however, some *O*-glycosyltransferases were recently shown to be reversible.[36-37]

Extensive bioinformatical analysis of glycosyltransferases have allowed them to be grouped into ~90 sequence-based families.[34] Despite the divergence in sequence homology of glycosyltransferases, all reported three-dimensional structures of GTs, thus far, fall into two major folds, GT-A and GT-B. Both folds contain two β/α/β Rossmann domains, however, in GT-A class GTs the Rossmann folds are abutting one another while in GT-B type GTs the Rossmann folds are facing each other. In addition, glycosyltransferases are further divided by their ability to retain or invert stereochemistry of the anomeric carbon of the donor substrate while forming their respective glycosidic bond (**Figure 10**). Mechanistically, inverting glycosylation is believed to involve a GT-mediated direct displacement S_N2-like reaction in which an active-site base catalyst deprotonates the nucleophile of the acceptor substrate facilitating direct displacement of the nucleotidyl diphosphate (NDP) leaving group. (**Figure 10**, A).[34,38] In contrast, retaining GTs first attack the donor substrate thereby removing the NDP activating group and tethering the donor substrate to the enzyme. A second base catalyst within the GT then deprotonates the acceptor substrate allowing for nucleophilic attack of the enzyme-bound donor substrate forming a glycosidic bond retaining the original stereochemistry of the anomeric carbon (**Figure 10**, B). Typically, glycosylation of secondary metabolites involves inverting-GTs; however, there is one example from chromomycin biosynthesis involving an α-1,4 glycosidic linkage that suggests the activity of a retaining glycosyltransferase.[39-40]

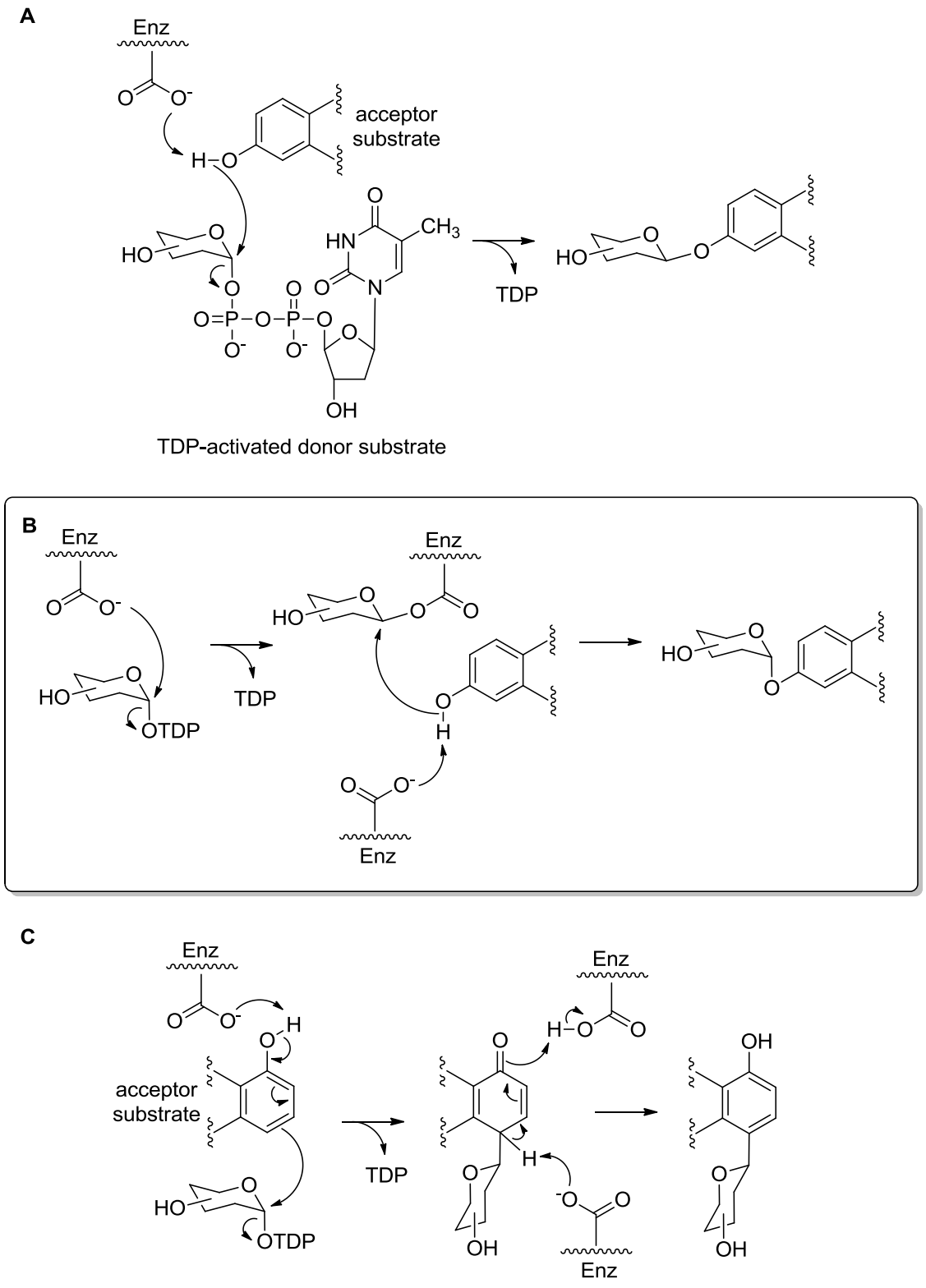


Figure 10. Hypothetical mechanisms for inverting (A) and retaining (B) *O*-glycoside formation as well as inverting *C*-glycoside formation (C).

C-glycosyltransferases are particularly interesting because of the chemical stability of their resulting *C*-glycosidic bond. In addition, unlike *O*-glycosides, *C*-glycosides are resistant to *in vivo* glycosidase activity. *C*-glycosides are typically found with sugar attachment ortho- or para- to an electron rich aromatic functional group. The resonance of these particular structures facilitates ortho- or para- attack of the anomeric carbon of the donor substrate as observed in urdamycin and gilvocarcin V (**49**) glycosylation reactions.[41-42] The exact mechanism of *C*- and *N*-glycoside formation, however, remains unclear (**Figure 10**, C).

Glycosyltransferases play an important role in combinatorial biosynthetic strategies toward producing novel polyketide derivatives. Novel steffimycins, aranciamycins, elloramycins, gilvocarcins, mithramycins, urdamycins and landomycins are just a few examples in which glycosyltransferases were responsible for creating analogues of a particular natural product. These results clearly show the tremendous potential displayed by GTs in the rational design of natural product glycosides.

Deoxysugar biosynthesis

Sugar biosyntheses in microbial hosts are extremely important as the resulting donor substrates are utilized by GTs in glycosylation reactions described above. Even though several natural products contain fully oxygenated sugar molecules such as glucose, a majority of secondary metabolites include sugar moieties that are deoxygenated to various degrees prior to their utilization by a glycosyltransferase. In either case, a dedicated NDP-hexose-nucleotidyltransferase (NT) is responsible for appending an NDP species to hexose-1-phosphate, thereby creating an "activated" NDP-hexose species.[38] This reaction requires NTP (ATP, CTP, GTP, UTP or TTP) and in microbial deoxysugar biosynthesis TTP and glucose-1-phoshpate are by far the most prominently utilized substrates which results in the formation of TDP-D-glucose (**41**). There are a few examples, however, of GDP- and UDP- activated deoxysugars derived from glucose and mannose.[43-44] The activated TDP-D-glucose (**41**) may be utilized directly by GTs as in vancomycin biosynthesis, but **41** is typically deoxygenated by a TDP-D-glucose 4,6-dehydratase (4,6-DH) producing TDP-4-keto-6-deoxy-D-glucose (**42**), a common branching point for all 6-deoxysugar biosynthetic pathways (**Figure**

11).[38,45] Additional deoxysugar biosynthetic enzymes including dehydratases, epimerases, group transferases and ketoreductases determine the extent of deoxygenation and decoration of the final deoxysugar proceeding from **42**.

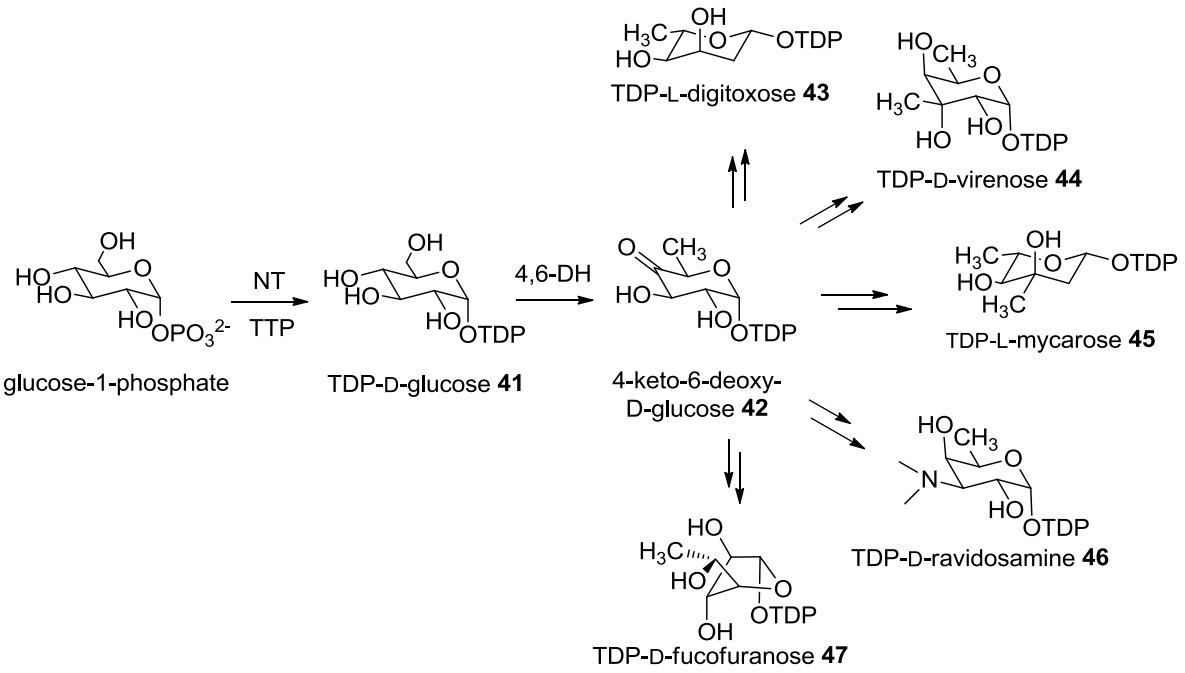


Figure 11. Common biosynthetic route toward the production of various 6-deoxysugars.

Deoxysugar biosynthetic genes offer additional resources for combinatorial biosynthesis. It is possible to predict the deoxysugar produced from a set of deoxysugar biosynthetic genes. In this context, it is therefore possible to combine various deoxysugar genes, sometimes from different pathways, in order to create a specific NDP-deoxysugar. These types of experiments have been successfully utilized with flexible GTs to rationally design specific deoxysugar containing natural product analogues.[46-50]

Oxygenation

The products of PKSs are often further functionalized through a series of oxidation reactions. In some instances these oxidation reactions are responsible for rearranging the polyketide core which provides an additional source for the structural

diversity observed in polyketide derived natural products. As with glycosylation, oxygenation is often required for the biological activity exhibited by many polyketides. Typically, these reactions are catalyzed by a large family of enzymes known as oxygenases, most commonly cytochrome P-450 monooxygenases (CYP450s), flavin (FAD or FMN)-dependent monooxygenases, anthrone oxygenases, and dioxygenases.[6-7] These enzymes generally use flavin, heme or metal ions to activate molecular oxygen, or the substrate, to mediate electron transfer during hydroxylation, epoxidation, anthrone oxidation, peroxide formation, dioxetane formation, desaturation and oxidative cleavage. Baeyer-Villiger monooxygenases are a class of flavin dependent monooxygenases that differ mechanistically from typical monooxygenase catalyzed hydroxylation by attacking keto groups rather than non-oxygen bearing carbon atoms. The classification of monooxygenase refers to the incorporation of one oxygen atom from molecular oxygen, while dioxygenases utilize both oxygen atoms. The additional oxygen atom remaining from a monooxygenase reaction is typically removed as H_2O or H_2O_2 by NADPH assisted reduction. Some examples of post-PKS oxygenation reactions in the biosynthesis of epirubicin (**24**), oleandomycin (**48**) and mithramycin (**23**) are shown in **Figure 12**.

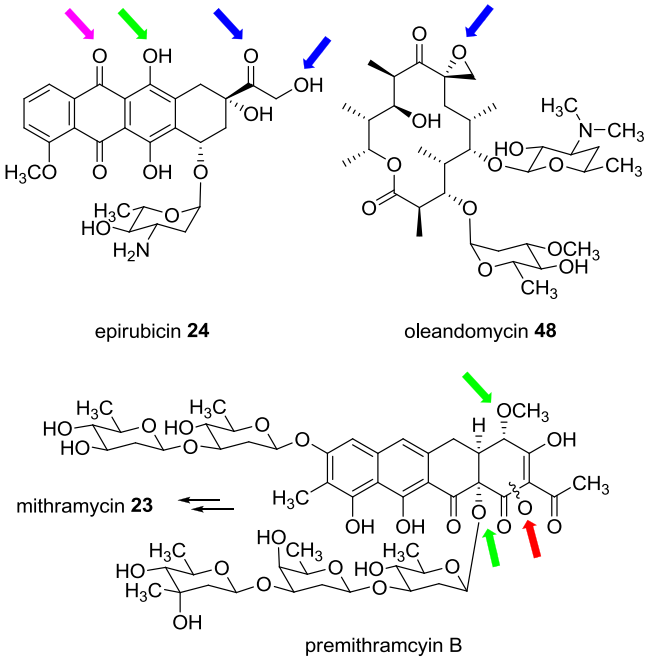


Figure 12. Oxygenase catalyzed modifications during epirubicin (**24**), oleandomycin (**48**) and mithramycin (**23**) biosynthesis. The oxygen atoms were introduced via a cofactor free anthrone oxygenase (purple), FAD dependant monooxygenase (green), CYTP450 monooxygenase (blue) or Baeyer-Villiger monooxygenase (red).

Oxygenases have been utilized heavily in combinatorial biosynthetic strategies including gene inactivation experiments to produce polyketide pathway intermediates and shunt products.[6-7] In addition, oxygenases with substrate promiscuity have been utilized in macrolide biosynthesis to create novel erythromycin and pikromycin derivatives.[51-52]

Methylation

Post-PKS tailoring events commonly involve methylation. These reactions are catalyzed by methyltransferases (MTs) that typically utilize *S*-adenosyl methionine (SAM) as a methyl donor to produce *N*-, *S*-, *C*- and *O*-tethered methyl groups. These groups can be carboxyl, phenol and hydroxyl groups, aliphatic and aromatic amines, thiols and thioethers, as well as alkenes and ring carbons.[7] MTs are also prominent in the deoxysugar biosynthetic pathways of L-mycarose, L-axenose, L-nogalose, L-oleandrose, D-virenose and many others.[38] The methylation of some deoxysugar species in secondary metabolites have been implicated in attributing biological activity. This is clearly

demonstrated in spinosyn and elloramycin derivatives containing various degrees of methylated L-rhamnose.[53-54]

Gilvocarcin

Gilvocarcin V (**49**, GV**, Figure 13**) is a unique angucyclinone born anticancer antibiotic produced by several *Streptomyces* species. This structurally unique molecule was first reported without complete structural characterization by Mizuno and coworkers as toromycin in 1980, and was found to be the principal product of *S. collinus*.[55-57] Shortly thereafter, GV as well as the closely related analogues gilvocarcin M (**50**) and gilvocarcin E (**51**) were fully characterized and found to be concomitant products of **49** production in *S. gilvotanareus* and *S. anandii*, respectively.[58-60] Together **49**, **50** and **51** comprise the gilvocarcins and are the most prominent members of a unique family of anticancer antibiotics that share a polyketide derived coumarin-based benzo[*d*]naptho[1,2-*b*]pyran-6-one moiety. Additional members of this family of natural products are often referred to as gilvocarcin-like aryl *C*-glycosides and include chrysomycin V (**52**) ravidomycin V (**53**), deacetylravidomycin V (**54**), FE35A (**55**) and B (**56**), Mer1020 dC (**57**) and dD (**58**), BE-12406A (**59**) and B (**60**) and polycarcin V (**61**) (**Figure 13**).[61-66] Members predominantly posses a *C*-glycosidically linked 6-deoxy-D-hexose moiety in C4 position while variants within this group exist as either furanose (**I** and **V, Figure 13**) or pyranose sugars (**II, III, IV and VI)**. Notably, BE-12406A (**59**) and B (**60**) as well as polycarcin V (**61**) are the only examples of 6-deoxy-L-sugars found in gilvocarcin-type compounds, with **59** and **60** representing the only *O*-glycosidically linked analogues reported thus far. Additional variations at the C8 side chain are restricted to a single methyl, ethyl or vinyl functional group and are the basis for the M, E and V abbreviations used in gilvocarcin nomenclature.

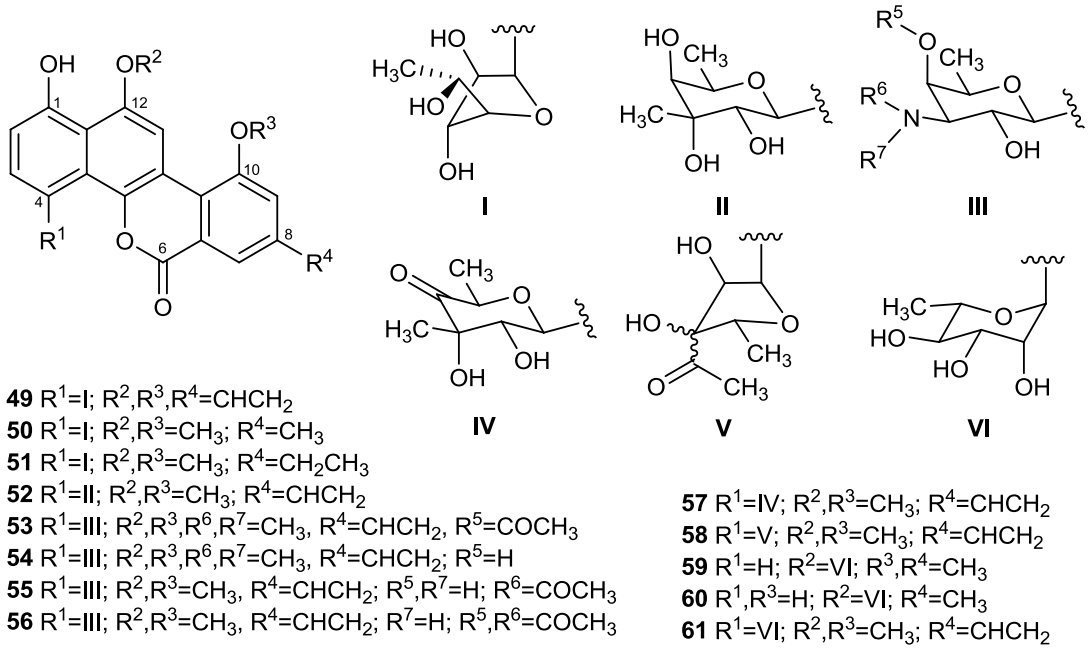


Figure 13. Glycodiversity of gilvocarcin-type anticancer drugs. Gilvocarcins V, M and E (**49**-**51**), chrysomycin V (**52**), ravidomycin V (**53**), deacetylravidomycin V (**54**), FE35A (**55**) and B (**56**), Mer1020 dC (**57**) and dD (**58**), BE-12406A (**59**) and B (**60**) and polycarcin V (**61**).

Gilvocarcin V has spurred particular interest because of its potent bactericidal, virucidal and antitumor activities at low concentrations while maintaining low *in vivo* toxicity.[58,67-70] **49** has been reported to undergo photoactivated [2+2] cycloaddition of its vinyl side chain with thymine residues of DNA in near-UV or visible blue light which results in single strand scissions leading to covalent binding with DNA.[71-76] This unique photo-activation of gilvocarcin V explains the lack of activity exhibited by gilvocarcin M and gilvocarcin E which do not contain the crucial vinyl functional group.[77-79] GV's activity is also attributed to a unique selective cross linking of DNA and histone H3, a core component of the histone complex that plays an important role for DNA replication and transcription.[80-84] The saccharide moiety, D-fucofuranose, of GV is essential for this activity as it is believed to facilitate binding of histone H3.[42,85] In addition, **49** has also been reported as an inhibitor of topoisomerase II.[86]

Biosynthetic Highlights of Gilvocarcin V

Initial incorporation studies of gilvocarcin, as well as ravidomycin and chrysomycin, revealed their unique backbone to be derived from the oxidative rearrangement of an angucyclinone core originating from acetate and propionate (**Figure 14**).[87-90] This unusual rearrangement in addition to other intriguing structural features including a *C*-glycosidically linked D-fucofuranose moiety and the use of a rare propionate starter unit was the driving force for further in-depth characterization of GV biosynthesis.

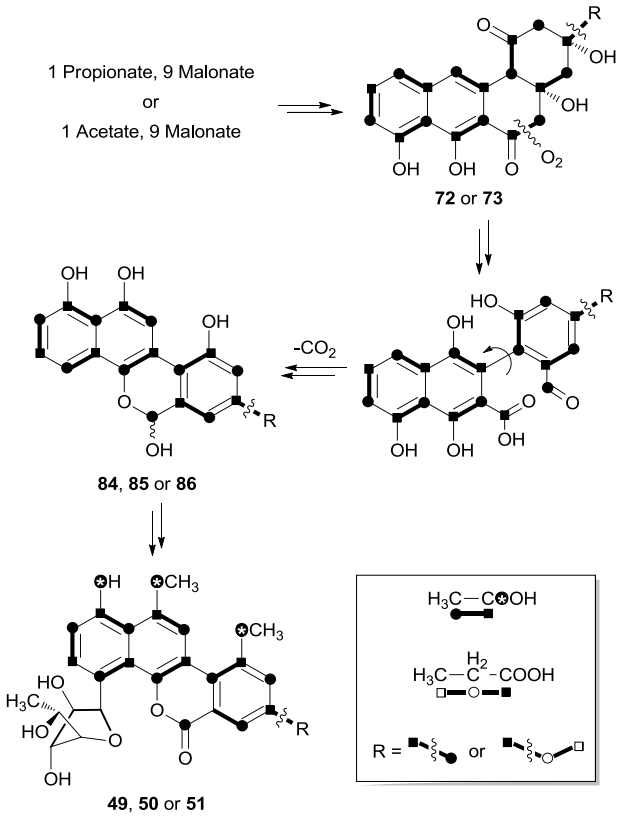


Figure 14. Labeling pattern and proposed ring opening during gilvocarcin biosynthesis.

Isolation of the complete GV biosynthetic gene cluster on a single cosmid, cosG9B3, from *S. griseoflavus* Gö3592 by Rohr et al. established the genetic foundation

for further biochemical analysis of the gilvocarcin biosynthetic pathway. The presence of the complete gilvocarcin biosynthetic locus (*gil*) was confirmed through heterologous reconstitution of the GV pathway by expressing cosG9B3 in *S. lividans* TK24.[91] Sequencing and bioinformatic analysis of cosG9B3 revealed the presence of genes encoding a type II PKS (*gilA, B* and *C*), a malonyl CoA:ACP transacylase (MCAT) (*gilP*), an acyltransferase (*gilQ*), a PKS associated ketoreductase (*gilF*), a cyclase/dehydratase (*gilK*), four oxygenases (*gilOI, OII, OIII* and *OIV*), a *C*-glycosyltransferase (*gilGT*), putative methyltransferases (*gilMT* and *gilM*), an oxidoreductase (*gilR*), deoxysugar biosynthetic genes (*gilE, D* and *U*), putative regulatory and resistance genes as well as several other genes with unknown functions including *gilN, L* and *V* (**Figure 15**).[91] Interestingly, the presence of an MCAT as well as an acyltransferase in the gilvocarcin cluster was unusual. Often, the minimal type II PKS ($KS_{\alpha/\beta}$ and ACP) does not require a dedicated MCAT, and instead recruits the endogenous FAS associated MCAT.[21,92-93] As gilvocarcin V and E biosynthesis utilizes a unique starter unit, namely propionate, it was hypothesized that GilQ might play a role in starter unit specificity (propionate vs. acetate).

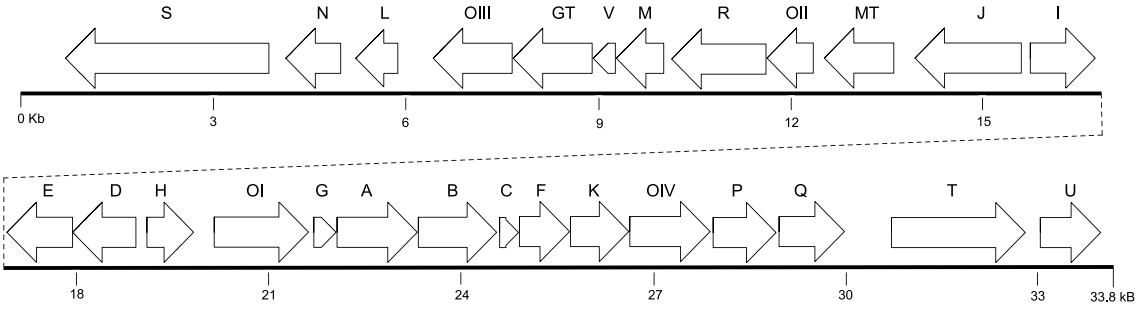


Figure 15. Biosynthetic gene cluster of gilvocarcin V and M.

The first biosynthetic experiments of GV, after the isolation of cosG9B3, involved determining the mode of oxidative C-C bond cleavage. To determine the genes responsible for oxidative cleavage, all four pathway oxygenases were inactivated on cosG9B3, individually or sequentially, and the generated mutant cosmids were heterologously expressed in *S. lividans* TK24. Accumulation of gilvocarcin E from the Δ*gilOIII* mutant clearly showed GilOIII did not play a role in oxidative ring cleavage, but instead was responsible for oxidizing the ethyl C8 side chain to the vinyl functional group

observed in gilvocarcin V.[42] The mutant cosmids Δ*gilOIV* and Δ*gilOIV, OII* both produced rabelomycins (**62** and **63**) as their major metabolites while Δ*gilOII* accumulated dehydro-rabelomycins (**64-66**) (**Figure 16**).[90,94] Similarly, Δ*gilOI* and Δ*gilOI, OII* mutually produced the 2,3-dehydro-UWM6s (**67** and **68**) and pregilvocarcin-*o*-quinones (**69** and **70**).[90,94] All of the metabolites produced by single or sequential inactivations of *gilOI*, *gilOII* and/or *gilOIV* were not the products of an oxidative rearrangement, therefore it was hypothesized that all three oxygenases were required for C5-C6 bond cleavage. Initial bioinformatic analysis of *gilOI*, *gilOII* of *gilOIV* revealed high sequence similarity to *jadH*, *jadG* and *jadF*, respectively from the jadomycin pathway (**Figure 17**). Jadomycin B (**71**) is also derived from the oxidative rearrangement of an angucyclinone core and has been extensively studied.[94-100]

Unlike gilvocarcin V, jadomycin B requires non-enzymatic incorporation of L-isoleucine and *O*-glycosylation (JadS) after the oxidative C-C bond cleavage (**Figure 17**). Interestingly, JadS may glycosylate before the incorporation of L-isoleucine.[97,169,170] Cross complementation of the individual gilvocarcin oxygenase mutants with their corresponding jadomycin pathway homologues restored gilvocarcin V production, except for Δ*gilOII*/*jadG* complementation.[94] These complementation experiments indicated that the gilvocarcin and jadomycin pathways share at least two functionally identical oxygenases while the third, *gilOII* and *jadG*, may serve unique functions for their respective pathways. Furthermore, co-expression of *gilOI*, *gilOII* and *gilOIV* with pWHM1238, a plasmid directing biosynthesis toward the putative pathway intermediate UWM6 (**72**), was able to restore jadomycin A (**74**) production when the fermentation media were supplemented with L-isoleucine.[94] Together, these results suggest that GilOI, GilOII and GilOIV together form a complex that is responsible for oxidative ring cleavage during gilvocarcin biosynthesis, and that both gilvocarcin and jadomycin oxygenase complexes act upon identical initial intermediates. Despite extensive experimentation, however, the ring cleavage mechanisms for both gilvocarcin and jadomycin have yet to be elucidated.

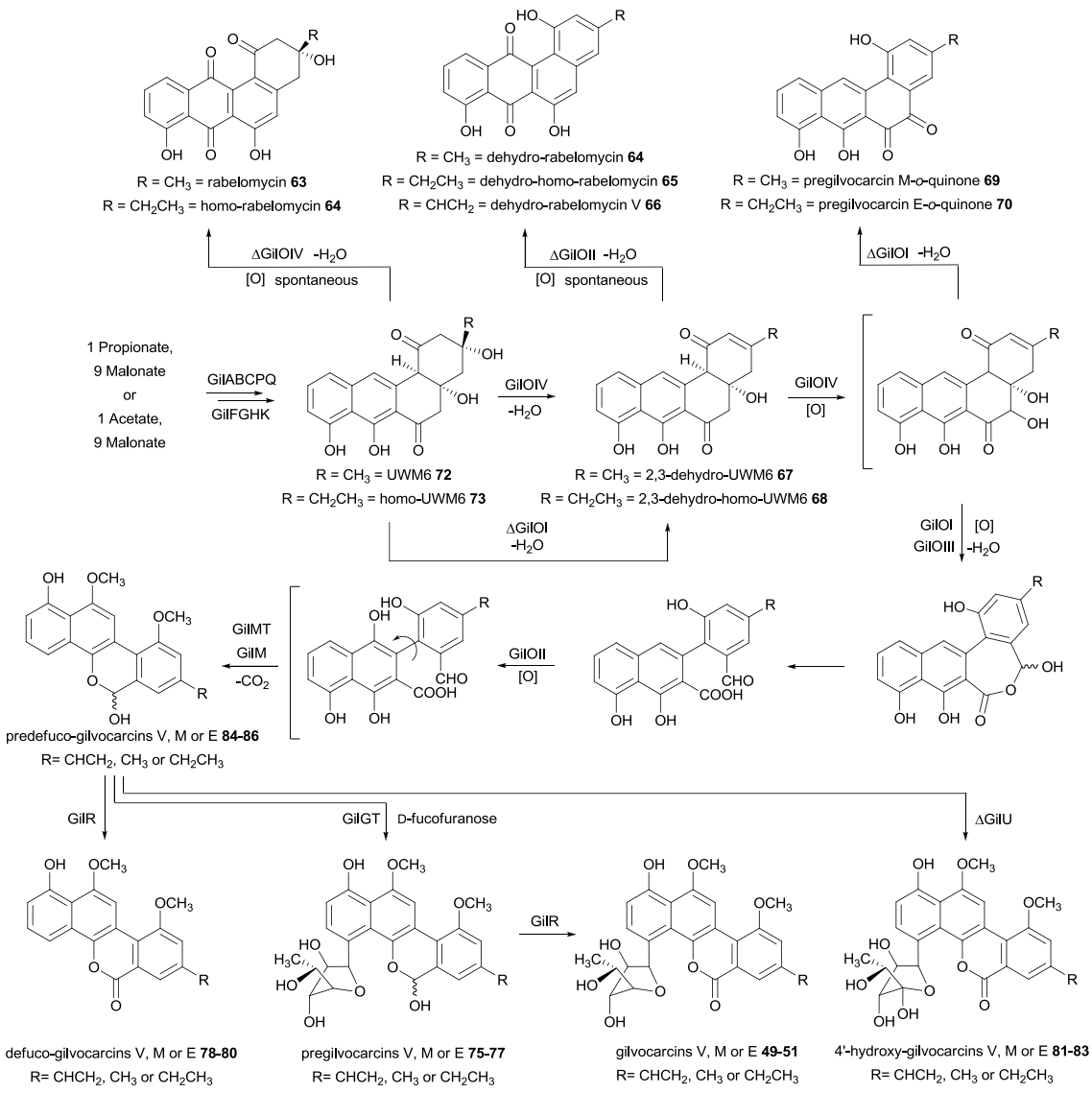


Figure 16. Proposed biosynthetic pathway for gilvocarcin V.

Additional post-PKS tailoring genes were inactivated by Rohr et al. in order to determine their function during GV biosynthesis. The inactivation of *gilR* led to the accumulation of glycosylated hemi-acetal intermediates referred to as pregilvocarcins (**75-77**) which indicated GilR acted as the last biosynthetic step toward GV production.[94] The presence of a hemi-acetal also provided evidence of an acid-aldehyde intermediate after oxidative ring cleavage. Subsequent *in vitro* analysis concluded that GilR was in fact an oxidoreductase responsible for the oxidation of the glycosylated hemi-acetal intermediate to form the lactone observed in **49**-**51**.[101] The function of *gilGT* was

similarly investigated by heterologous expression of the glycosyltransferase deficient cosmid Δ*gilGT*. The major metabolites produced by Δ*gilGT* were the defuco-gilvocarcins (**78-80**) which lack the expected D-fucofuranose moiety, and therefore implicated GilGT as the pathway specific *C*-glycosyltransferase (**Figure 16**).[42] Compounds **78-80** contain a lactone instead of the predicted hemi-acetal and are shunt products produced by the somewhat substrate flexible GilR.[42,101]

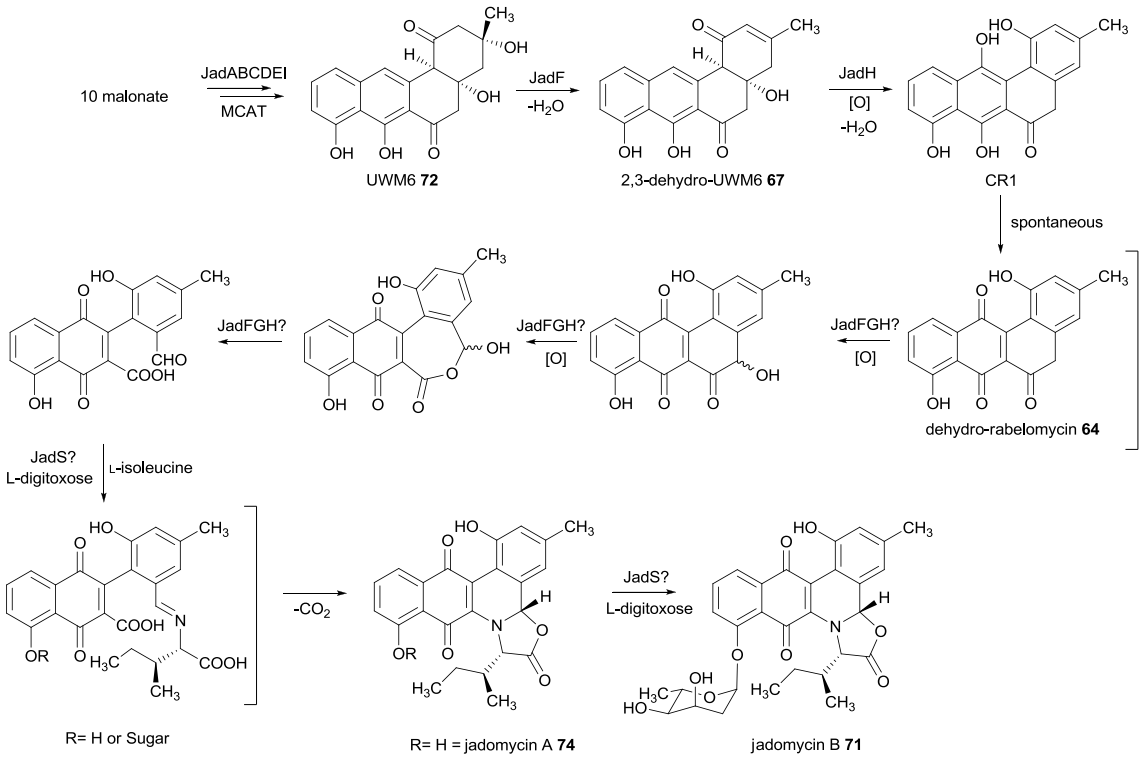


Figure 17. Proposed biosynthetic pathway for jadomycin B.

The sugar moiety of gilvocarcins, D-fucofuranose, is rarely found in natural products. Given that the sugar moiety is likely a crucial component for the biological activity of GV, the biosynthesis of this unique deoxysugar and its attachment to the gilvocarcin acceptor substrate was of particular interest. Inactivation of *gilU*, a putative deoxysugar 4-ketoreductase, resulted in the accumulation of a new gilvocarcin derivative, 4′-hydroxy-GV (**81, Figure 16**), with improved biological activity.[85] These results provided clear evidence that GilU was responsible for 4-ketoreduction of **42** during the biosynthesis toward TDP-D-fucofuranose (**47**) (**Figure 18**). More importantly, the

accumulation of **81-83** showed the natural glycosyltransferase, GilGT, retained donor substrate flexibility, and that modification of the sugar moiety of GV can lead to analogues with improved biological activity. Combinatorial approaches to probe the substrate flexibility of GilGT have been met with limited success and have resulted in the production of only D-olivosyl-gilvocarcins (**87-89**) and polycarcins (**61, 90 and 91**) containing L-rhamnose (**92**) and D-olivose (**93**), respectively (**Figures 19 and 34**).[48]

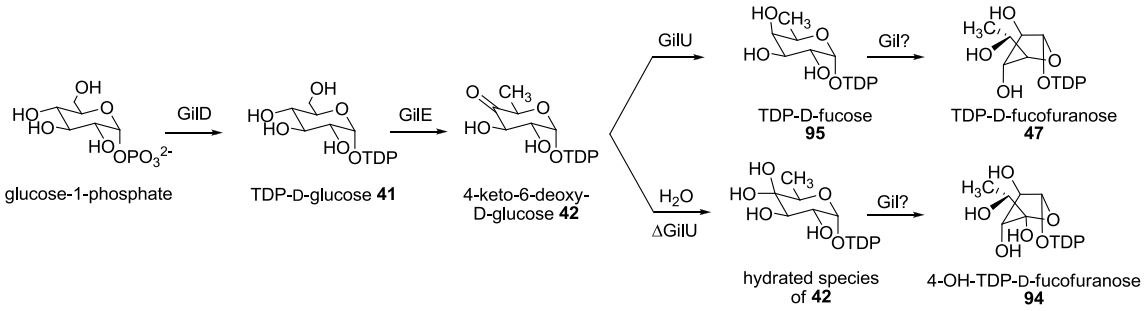


Figure 18. TDP-D-fucofuranose (**47**) and 4-OH-TDP-D-fucofuranose (**94**) biosynthetic pathways.

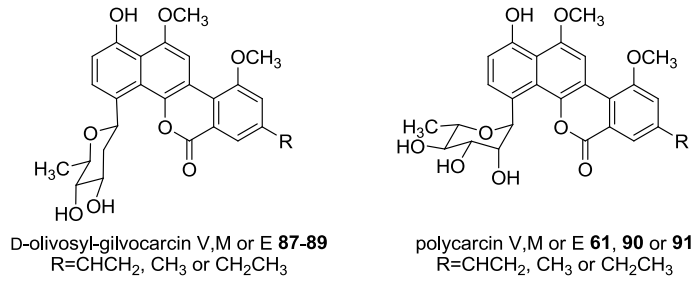


Figure 19. Gilvocarcin analogues produced through the moderately flexible glycosyltransferase GilGT.

Based on previous *in vivo* and *in vitro* experimentation one possible biosynthetic pathway of gilvocarcin has been proposed (**Figure 16**). The *gil*PKS is responsible for catalyzing the formation of the earliest isolated intermediates of GV biosynthesis, **72** and **73**, which are identical except for their C8 side chains which arise from the use of acetate or propionate, respectively during PKS initiation. As discussed above, GilOIV, GilOI and GilOII catalyze an oxygenase cascade leading to an acid-aldehyde intermediate while GilOIII installs the vinyl side chain by oxidizing the C8 ethyl side chain. Further

modification by *O*-methyltransferases producing **84-86** and subsequent *C*-glycosylation by GilGT creates the pre-gilvocarcins (**75-77**). Finally, oxidation of the hemi-acetal by GilR produces the lactone found in gilvocarcins (**49-51**).

Summary

Polyketides comprise one of the largest and most structurally diverse groups of natural products with examples found in use throughout the pharmaceutical and agricultural industries. Biosynthetic research is focused on understanding the genetic and biochemical basis for unique architecture found among natural products. This understanding has paved the way for combinatorial biosynthetic approaches to rationally design natural products with modified structural features not possible or extremely difficult through traditional chemical synthesis. Synthetic routes toward large natural products are extremely difficult and are made even more complex when the natural product contains one or more sugar moieties, as in gilvocarcin V (**49**) or mithramycin (**23**). Our lab has been primarily focused on understanding and modifying, through combinatorial biosynthesis, the biosynthetic pathway of gilvocarcin V. This has led to extensive *in vivo* and *in vitro* experimentations which have allowed the functional roles of many individual genes to be determined. Almost all of the biosynthetic knowledge of gilvocarcin has come through inactivation experiments. This has led to several shunt products which give indirect support for the role of the inactivated gene. Despite rigorous efforts, however, there are still many elusive biosynthetic steps that remain to be solved. For example, the genetic determinant for utilization of either acetate or propionate during gilvocarcin biosynthesis remains unknown. As gilvocarcin V is the active congener among gilvocarcins, it would be ideal to engineer a strain that produces only gilvocarcin V and therefore does not waste valuable biosynthetic building blocks toward the production of gilvocarcin M and E. A complete characterization of the putative genes, *gilP* and *gilQ*, involved in acyl-unit incorporation during GV biosynthesis would provide a means to rationally engineer a strain with increased GV production.

A complete understanding of the biosynthetic machinery of the gilvocarcin pathway provides the foundation for our ultimate goal, which is to create gilvocarcin V analogues with improved pharmaceutical properties. In this context, we have focused on the

glycodiversification of gilvocarcin V, as discussed earlier, through probing the substrate flexibility of GilGT. This approach has produced a small library of gilvocarcin V analogues with comparable or improved biological activity. The narrow donor substrate flexibility of GilGT limits its use as a combinatorial biosynthetic tool towards further expanding the library of glycosylated GV analogues. The creation of an engineered glycosyltransferase or the identification of a GilGT-like glycosyltransferase with improved substrate flexibility would therefore be greatly beneficial toward our combinatorial biosynthetic aims. In this context, we are also interested in the sugar moieties effect on the biological activity of GV as well as of and any other members of the gilvocarcin-like aryl *C*-glycosides. To date, there has not been an extensive structure-activity-relationship (SAR) study concerning each individual position of any gilvocarcin like aryl *C*-glycoside sugar moiety, despite their required presence for antitumor activity. The only such study involved synthetic derivatization of the 2′ and 4′ positions of ravidomycin V (**53**).[102]

Specific Aims

The objectives of this research study were to, (a) expand the glycodiversity of GV analogues by identifying/engineering gilvocarcin-like glycosyltransferases with improved substrate flexibility; (b) to create *O*-methylated-L-rhamnose derivatives of polycarcin V (**61**) for sugar oriented SAR studies; *c*) to determine the enzymes involved in unique starter unit incorporation during GV (**49**) biosynthesis; and *d*) to design a plasmid based system in which to produce gilvocarcin pathway intermediates. To achieve these goals, the following four specific aims were addressed:

Specific Aim 1a: Create engineered *C*-glycosyltransferases through domain swapping using *gilGT*, *chryGT* and *ravGT*.

Specific Aim 1b: Identify the polycarcin V pathway specific glycosyltransferase and explore its donor substrate flexibility.

Creation or identification of a *C*-glycosyltransferase with improved donor substrate flexibility will provide a powerful biosynthetic tool for further glycodiversification of GV

and will shed light on the viability of domain swapping as a useful approach in engineering *C*-glycosyltransferases.

Specific Aim 2: Utilize L-rhamnose-*O*-methyltransferases from elloramycin (ElmMII and ElmMIII) and steffimycin (StfMII) biosynthesis to create *O*-methylated-L-rhamnose analogues of polycarcin V.

Modification of L-rhamnose in polycarcin V will allow for an SAR study focused on the individual functional groups of the sugar moiety and their role in conferring biological activity. Insights gained through such a study would provide information for more rationally guided attempts to improve the biological activities of gilvocarcin-like aryl *C*-glycosides.

Specific Aim 3a: Characterize the putative MCAT, *gilP*, and acyltransferase, g*ilQ*, from gilvocarcin V biosynthesis *in vivo*.

Specific Aim 3b: Characterize the putative MCAT, GilP, and acyltransferase, GilQ, from gilvocarcin V biosynthesis *in vitro*.

Understanding the role of GilP and GilQ in starter unit specificity may lead to metabolic engineering of strains that produce only the active gilvocarcin V congener. In addition, the use of a propionate starter unit, as in GV biosynthesis, is rare among type II PKS natural products and understanding the enzyme/s responsible for this incorporation may be used as a combinatorial biosynthetic tool to engineer other type II PKS pathways to utilize a propionate starter unit.

Specific Aim 4: Design a plasmid based approach to produce proposed gilvocarcin V pathway intermediates using genes from the gilvocarcin biosynthetic cluster.

An *in vivo* method to produce proposed gilvocarcin V pathway intermediates could provide valuable insights into the biosynthesis of GV as well as a means to obtain substrates for *in vitro* characterization of putative enzyme remaining in the GV cluster. These results could also provide a means to delineate the entire gilvocarcin pathway by shuffling various biosynthetic genes and determining the consequences of adding or removing an individual gene.

Chapter 2: Glycodiversification of Gilvocarcin V

2.1 Engineering chimeric glycosyltransferases

As discussed previously, glycosylation is often a crucial requirement for the activity of bioactive natural products. Glycosyltransferases therefore hold significant promise as a combinatorial biosynthetic tool. Presently, there are over 40,000 putative glycosyltransferases within the NCBI gene databank, but only a small portion of these GTs have been experimentally investigated. The biggest restriction for the use of glycosyltransferases in combinatorial approaches is their common strict inherent substrate specificity. This has led to an explosion in glycosyltransferase structural biology and enzymology research, focused on ultimately broadening this substrate specificity.[103]

The glycosylation reaction during the biosynthesis of gilvocarcin is catalyzed by GilGT, a *C*-glycosyltransferase utilizing presumably TDP-D-fucofuranose (**47**) as its donor substrate. It is still possible that TDP-D-fucose (**95**) may first be attached and converted to **47** after the glycosyltransferase reaction. Previous inactivation of the deoxysugar 4-ketoreductase encoding gene, *gilU*, resulted in the formation of 4′-hydroxy gilvocarcin V (**81**) which was found to be more active than the parent gilvocarcin V compound (**49**).[85] These results automatically placed GilGT in an elite class of *C*-glycosyltransferases, accompanied only by UrdGT2 (from urdamycin biosynthesis), with marginal substrate flexibility. The donor substrate specificity of GilGT was further explored by co-expressing plasmids encoding deoxysugar genes toward to the biosynthesis of a particular deoxysugar (referred to as deoxysugar plasmids) with the Δ*gilU* mutant cosmid, cosG9B3-GilU$^-$. The Δ*gilU* strain was used in lieu of the unmodified cosG9B3 cosmid because it does not produce the natural donor substrate, TDP-D-fucofuranose (**47**), and instead produces TDP-4-OH-fucofuranose (**94**). The low production yields of **81-83** accumulated through heterologous expression of cosG9B3-GilU$^-$ revealed that **94** was in fact accepted by GilGT, but at a much reduced rate when compared to the natural donor substrate, **47**. This provided a suitable, albeit not ideal, host in which exogenous deoxysugar genes could be expressed to produce specific deoxysugars that could then compete with **47** for GilGT activity. These experiments resulted in the successful glycosylation with only TDP-D-olivose (**109**) and TDP-L-

rhamnose (**112**) producing D-olivosyl gilvocarcins (**87-89**) and polycarcins (**61, 90 and 91)**, respectively; despite attempting a range of various deoxysugars including D- and L-branched, amino and neutral deoxysugars.[48] These results exhausted the use of GilGT to further glycodiversify gilvocarcins as it seems to lack the broad substrate specificity needed to truly expand the library of gilvocarcin analogues. A possible route to overcome this problem would be to engineer substrate flexibility into GilGT.

There are only a handful of examples showing the successful generation of functional engineered *O*-glycosyltransferases through various methods including directed evolution, "hot spot" mutagenesis and domain swapping.[104-111] Directed evolution was used to improve the substrate flexibility of the naturally flexible glycosyltransferase OleD, involved in self resistance in the oleandomycin (**48**) pathway, by increasing its tolerance for the fluorescent surrogate acceptor 4-methylumbelliferone (**96**) (**Figure 20**).[33,104,111-112] Specifically, a small library of OleD variants were identified with improved ability to produce 4-methyl-umbelliferyl β-D-glucopyranoside (**97**). Three mutants were identified and the mutations were subsequently recombined to produce a triple mutant with marked improvement in substrate flexibility. Later a "hot spot" saturation mutagenesis approach was used in which the previously identified hot spots underwent single-site saturation mutagenesis producing several libraries modified at the individual hot spot positions.[106] High-throughput screening with the fluorescent **96** identified mutations showing improved activity. The most active mutations were combined to produce an OleD variant showing 150-fold improvement compared to wild type OleD.[106] These results show the promise of directed evolution and "hot spot" mutagenesis for improving GT activity, however, the apparent downside is the lack of high-throughput screening possibilities for a majority of natural product glycosyltransferases, including GilGT. The successful engineering of OleD was facilitated by the ability to visualize, through fluorescence, the glycosyltransferase reaction. The activity of OleD could quickly be assessed by the loss of fluorescence exhibited by following the OleD catalyzed conversion of **96** to **97**. Unfortunately, this procedure can not be broadly applied to other GTs due to their inability to glycosylate compounds applicable for colorimetric assays.

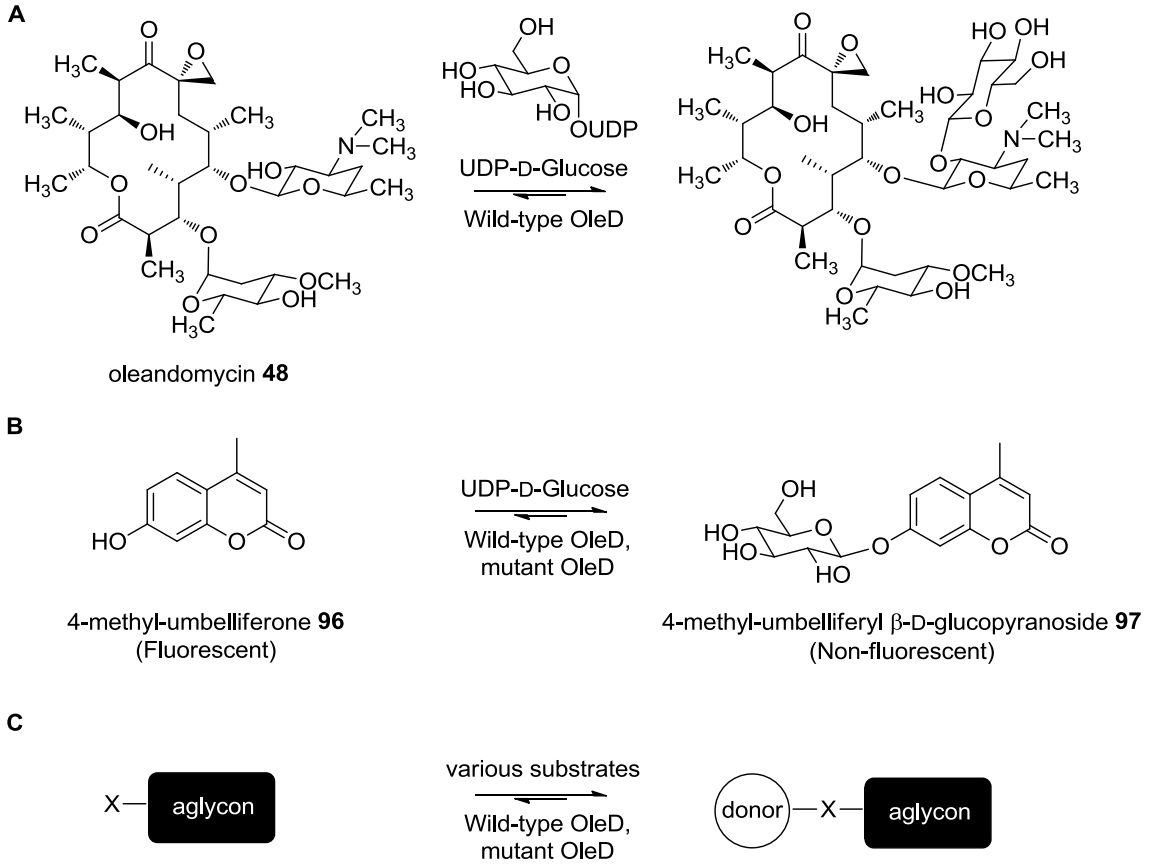


Figure 20. Broadening OleD substrate specificity. (A) Wild-type role of OleD during oleandomycin biosynthesis. (B) Directed evolution of OleD towards increased tolerance toward **96**. (C) Further optimization by "hot spot" saturation mutagenesis increases substrate flexibility.

Domain swapping is an approach in which the substrate binding domains of a particular glycosyltransferase are swapped with an alternative substrate binding domain from a second glycosyltransferase. This approach is based on the findings that glycosyltransferases have distinct acceptor (*N*-terminal) and donor (*C*-terminal) recognition domains which are connected by a linker loop region.[113-116] Successful glycosyltransferase domain swapping experiments have produced novel natural products including urdamycin P (**98**) and several vancomycin derivatives (**99**-**106**). [107,109] The production of hybrid vancomycin derivatives were achieved by swapping the *N*- and *C*-termini between GtfA (chloroorienticin) and Orf1 (teicoplanin) which are responsible for the attachment of TDP-*epi*-vancosamine and UDP-*N*-acetylglucosamine to desvancosaminyl vancomycin (**107**) and teicoplanin glucosaminyl-pseudoaglycone (**108**), respectively. One resultant chimeric glycosyltransferase, GtfAH1 (*N*-terminal GtfA, *C*-

terminal Orf1) was shown to attach UDP-glucose and UDP-*N*-acetylglucosamine to an array of acceptor substrates producing novel vancomycin derivatives (**Figure 21**).[109] These results showed that domain swapping was able to successfully increase substrate tolerance compared to wild type GtfA and Orf1.

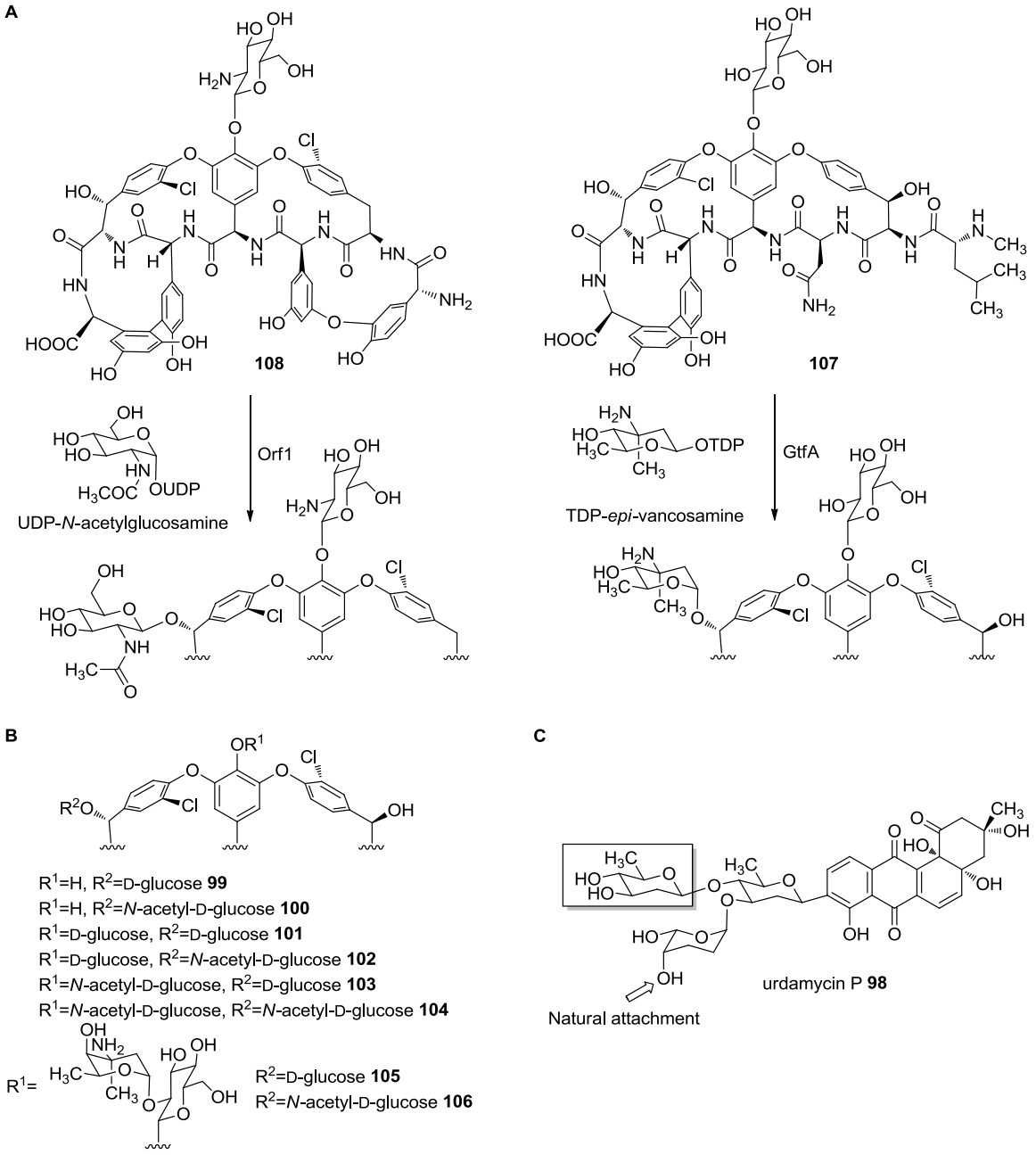


Figure 21. Examples of domain swapping to increase GT substrate flexibility. (A) Natural glycosylation function of Orf1 and GtfA in teicoplanin and chloroorienticin biosynthesis, respectively. (B) Novel vancomycin analogues produced through the chimeric glycosyltransferase GtfAH1 (N-term. GtfA, C-term. Orf1). (C) Structure of urdamycin P, a new natural product produced through an engineered glycosyltransferase.

The recent isolation and characterization of the biosynthetic gene clusters for ravidomycin V (**53)** and chrysomycin V (**52**) provided two additional gilvocarcin-like *C*-glycosyltransferases that could be utilized for a domain swapping approach to create

substrate flexible *C*-glycosyltransferases.[117] RavGT (ravidomycin) and ChryGT (chrysomycin) are responsible for the attachment of an amino sugar and branched sugar, respectively (**Figure 22**). This is important as ChryGT is the only known *C*-glycosyltransferase responsible for branched sugar (D-virenose) attachment and RavGT is only the second known amino sugar (D-ravidosamine) transferring *C*-glycosyltransferase besides Med-ORF8 (D-angolosamine) from medermycin biosynthesis.[118] Amino sugar attachment is particularly intriguing as amino sugars can improve solubility as well as provide useful functionality for drug formulations. Surprisingly, RavGT and not ChryGT was able to restore gilvocarcin production when complementing the GilGT deficient mutant *S. lividans* TK24/cosG9B3-GilGT$^-$.[117]

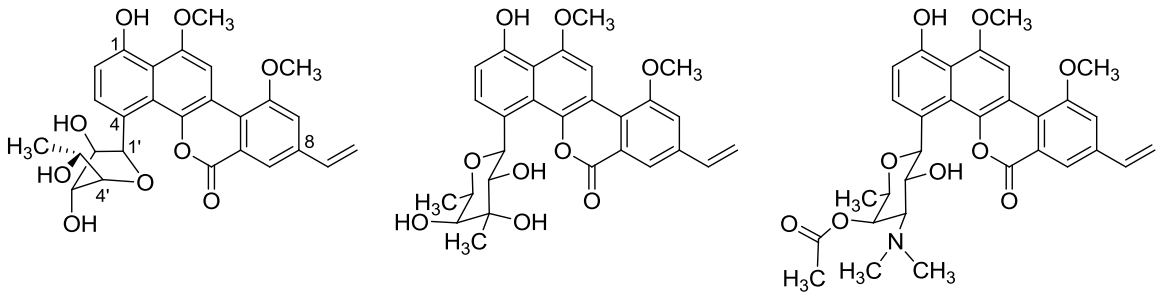


Figure 22. Structures of gilvocarcin, chrysomycin and ravidomycin.

Experimental design

In **specific aim 1a**, six chimeric glycosyltransferases will be constructed through domain swapping between *gilGT*, *chryGT* and *ravGT*. Resulting chimeras will initially be screened for functionality by co-expressing them with the cosG9B3-GilGT$^-$ mutant in *S. lividans* TK24 and accessing their ability to restore gilvocarcin E (**51**) production. Successful complementation of the *gilGT* deletion mutant does not produce the expected gilvocarcin V (**49**) congener, and instead produces **51** because of a polar effect on the downstream *gilOIII* gene responsible for converting the ethyl side chain to the vinyl side chain.[42] The chimeric glycosyltransferases will also be cloned into nine deoxysugar plasmids directing the biosynthesis of various structurally diverse deoxysugars. These plasmids will then be co-expressed with cosG9B3-GilGT$^-$ in *S. lividans* TK24 to probe

their substrate flexibility. All recombinant strains will be fermented and metabolites will first be screened on a small scale using high-performance liquid chromatography - mass spectrometry (HPLC-MS). Promising candidates will then be scaled up and metabolites will be isolated through chromatographic techniques and their structures will be elucidated through spectroscopic characterization using nuclear magnetic resonance (NMR), MS and ultraviolet-visible spectroscopy (UV-Vis).

Results

Construction of chimeric glycosyltransferases

Initial amino acid sequence comparison of *gilGT*, *chryGT* and *ravGT* revealed *chryGT* and *ravGT* to be extremely similar (88%/73% amino acid similarity/identity). The identity was reduced to 39% when *gilGT* was added to the alignment, possibly due to the donor substrate preferences of GilGT (furanose) vs. ChryGT and RavGT (pyranose). This was further explored by aligning the individual termini of each glycosyltransferase. Again the addition of *gilGT* domains reduced the overall identity as *ravGT* and *chryGT* identities were 77% and 68% compared to 44% and 35% with *gilGT* for *N*- and *C*-terminals, respectively. As GilGT, ChryGT and RavGT utilize different donor substrates for the same aglycone acceptor it is expected that the *N*-terminals will show higher similarities than the *C*-terminals; however, the differences are significantly higher when comparing two GTs that transfer a furanose and a pyranose sugar (RavGT or ChryGT and GilGT) as opposed to two pyranose sugar transferring GTs (RavGT and ChryGT). Further bioinformatical analysis using SEARCHGTr (http://www.nii.res.in/searchgtr.html) revealed the putative linker region for each glycosyltransferase (solid box in **Figure 23**).[119]

Figure 23. Amino acid alignment of *gilGT*, *chryGT* and *ravGT*. The figure shows the putative linker region (solid boxes) and the shuffling point (dotted box) for preparing the chimeric glycosyltransferases.

The linker region was the original site of interest for introducing a unique restriction site for domain swapping. The introduction of a restriction site would inevitably alter one or more amino acid residues, and due to the highly conserved nature of the linker region it was decided to instead focus on a small upstream region (toward *N*-terminal) with less conserved residues (dotted box in **Figure 23**). Using the polymerase chain reaction (PCR), each glycosyltransferase was amplified as two fragments individually covering the *N*- and *C*-terminals, respectively (**Figure 24**, step 1-2). Each fragment was designed to contain an engineered *Nhe*I restriction site at the *N*- and *C*-terminal interface by altering the nucleotide sequence of the amino acid represented in the dotted box shown in **Figure 24**. These modifications altered only two amino acids in the GilGT sequence (ATG to ALA), one amino acid in the ChryGT sequence (PGA to PLA) and one amino acid in the RavGT sequence (SFA to SLA). The resulting PCR amplified fragments were cloned into PCR-Blunt II-TOPO (Invitrogen) and subsequently combined to form intact glycosyltransferases as depicted in **Figure 24**, steps 3-4 (see experimental section for further details). The successful combination of domains from three individual glycosyltransferases produced six chimeric glycosyltransferases designated cGT-CG, cGT-CR, cGT-GC, cGT-GR, cGT-RC and cGT-RG. The parent glycosyltransferase contributing the *N*-terminal of the newly generated chimeric glycosyltransferase (cGT) is signified by C (ChryGT), G (GilGT) or R (RavGT) followed by the parent glycosyltransferase composing the *C*-terminal. For example, cGT-CG is comprised of an *N*-terminal and *C*-terminal from ChryGT and GilGT, respectively. Sequencing confirmed the expected sequence for each engineered glycosyltransferase.

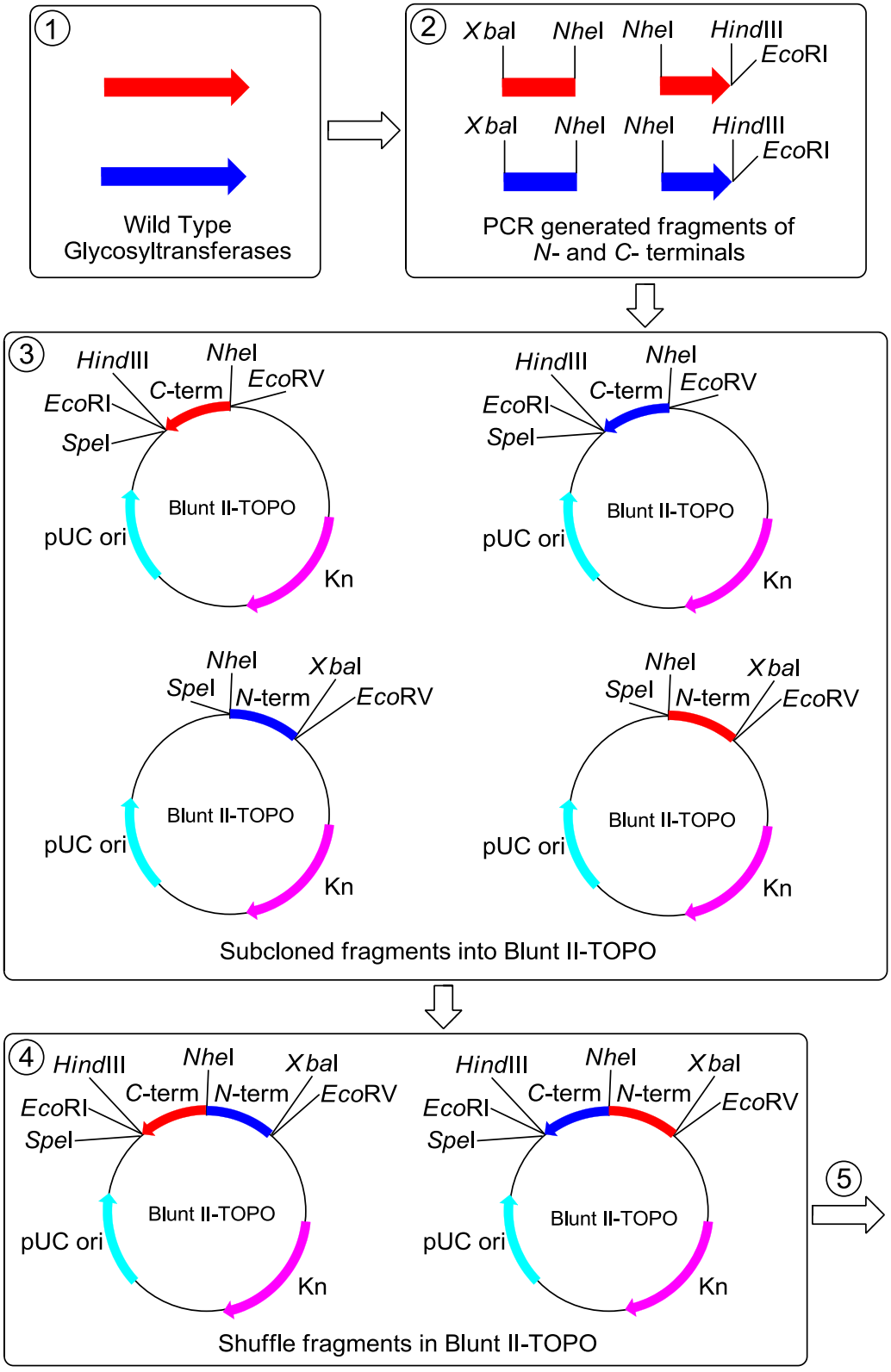


Figure 24. Representation of chimeric glycosyltransferase preparation (steps 1-4).

In order to express the newly constructed glycosyltransferases in *Streptomyces*, the GTs must be moved into a *Streptomyces* expression vector, such as pEM4.[112] Through restriction digests the chimeric GTs were transferred into pEM4 downstream of the constitutively activated erythromycin resistance promoter *ermE**p (**Figure 25**, step 5). Finally, the intact chimeric glycosyltransferases were removed from pEM4 with *ermE**p and placed into nine individual sugar plasmids (**Figure 25**, step 6).

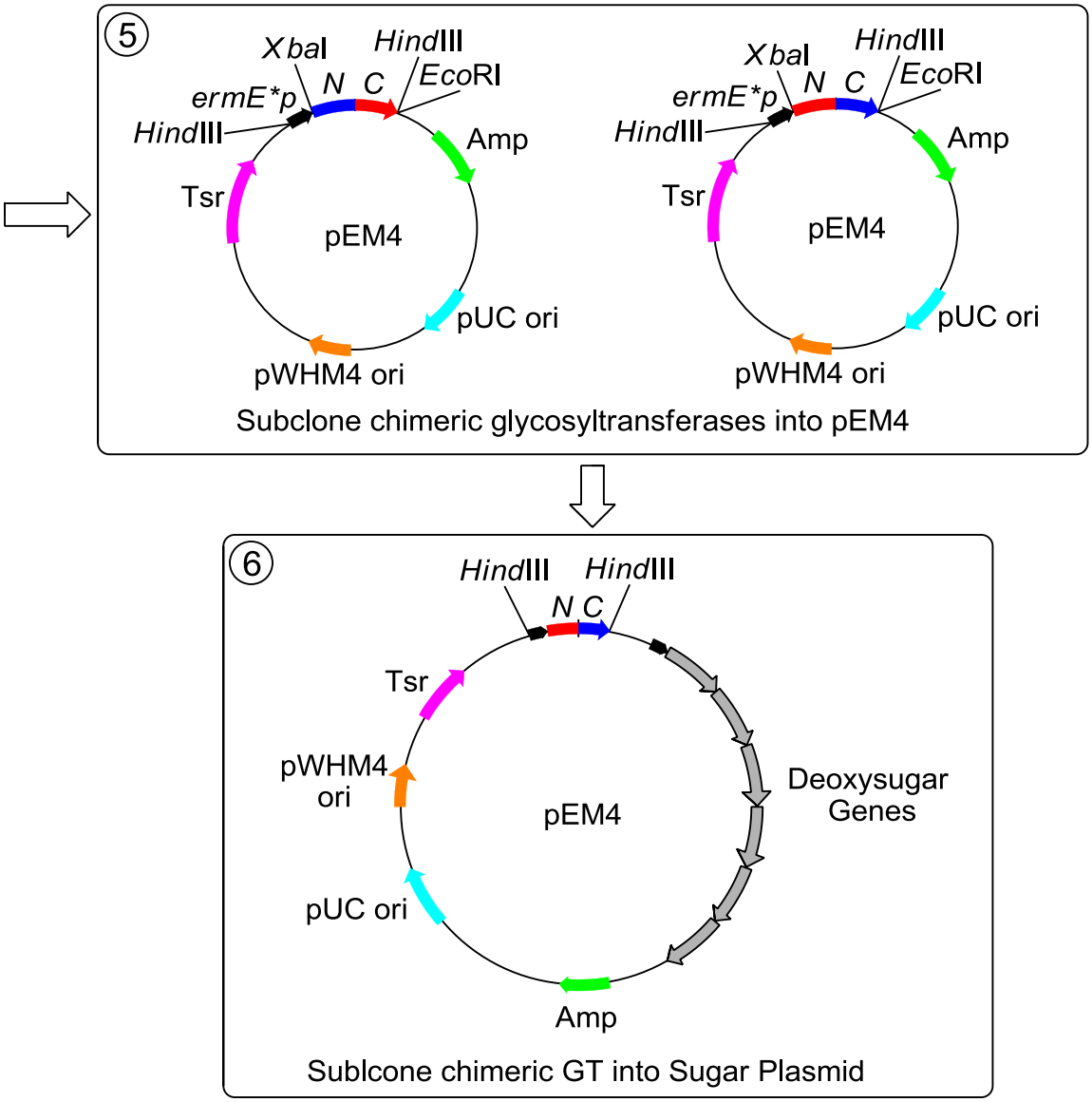


Figure 25. Representation of chimeric glycosyltransferase preparation (steps 5-6).

Initial chimeric glycosyltransferase screening

Each engineered glycosyltransferase, cGT-CG, cGT-CR, cGT-GC, cGT-GR, cGT-RC and cGT-RG was transformed via protoplast transformation into *S. lividans* TK24/cosG9B3-GilGT⁻ and screened for the reconstitution of gilvocarcin E (**51**) production. Ideally, the engineered glycosyltransferases should be tested in RavGT- and ChryGT-minus mutants. Unfortunately, these mutants do not exist making it nearly impossible to truly test if the newly constructed GTs retain parental functionality. By using *S. lividans* TK24/cosG9B3-GilGT⁻, parental functionality can only be tested for cGT-CG and cGT-RG which both contain the sugar binding domain of GilGT. To some extent this can also be tested for cGT-CR and cGT-GR as RavGT was observed to accept TDP-D-fucofuranose, albeit at a much reduced rate. In this context, reconstitution of **51** biosynthesis by any of the cGTs, other than cGT-CG and cGT-RG, will indicate improved donor substrate flexibility.

The fermentation of all recombinant strains revealed only *S. lividans* TK24/cosG9B3-GilGT⁻/cGT-CR was able to transfer TDP-D-fucofuranose (**47**) (**Figure 26**). This was confirmed by the presence of **50** and **51** through UV-Vis and electrospray ionization mass spectrometry (ESI-MS). Surprisingly, both constructs which contained the *C*-terminal domain of GilGT failed to complement the strain. In addition, the activity of cGT-CR and not cGT-GR was unexpected. Both constructs contain the *C*-terminal domain of RavGT, which is known to accept TDP-D-fucofuranose, and differ in only their *N*-terminal domains. It was expected that cGT-GR would restore **51** production as both domains are from glycosyltransferases that can themselves restore gilvocarcin biosynthesis. Likewise, cGT-RG contains the opposite shuffling pattern of cGT-GR and was also unable to transfer TDP-D-fucofuranose.

Complete characterization of the products from *S. lividans* TK24/cosG9B3-GilGT⁻/cGT-CR confirmed the structures to be that of **50** and **51**. The structures were confirmed by NMR to eliminate the possibility of TDP-D-fucose (**95**) transfer, and not TDP-D-fucofuranose (**47**) whose resulting products would be indiscernible from mass and UV-Vis data alone. TDP-D-fucose (**95**) undergoes ring contraction by an as of yet undetermined enzyme to produce TDP-D-fucofuranose (**47**) (**Figure 18**). The lack of a

putative ring contracting enzyme in the *gil* cluster has prompted several hypotheses to explain the generation of TDP-D-fucofuranose, including the hypothesis that GilGT may itself catalyze ring contraction of TDP-D-fucose before glycosylation. These results suggest ring contraction occurs separately from GilGT, and is more likely catalyzed by one of the few remaining uncharacterized *gil* genes or from an endogenous *Streptomyces* enzyme that is recruited by the gilvocarcin biosynthetic machinery.

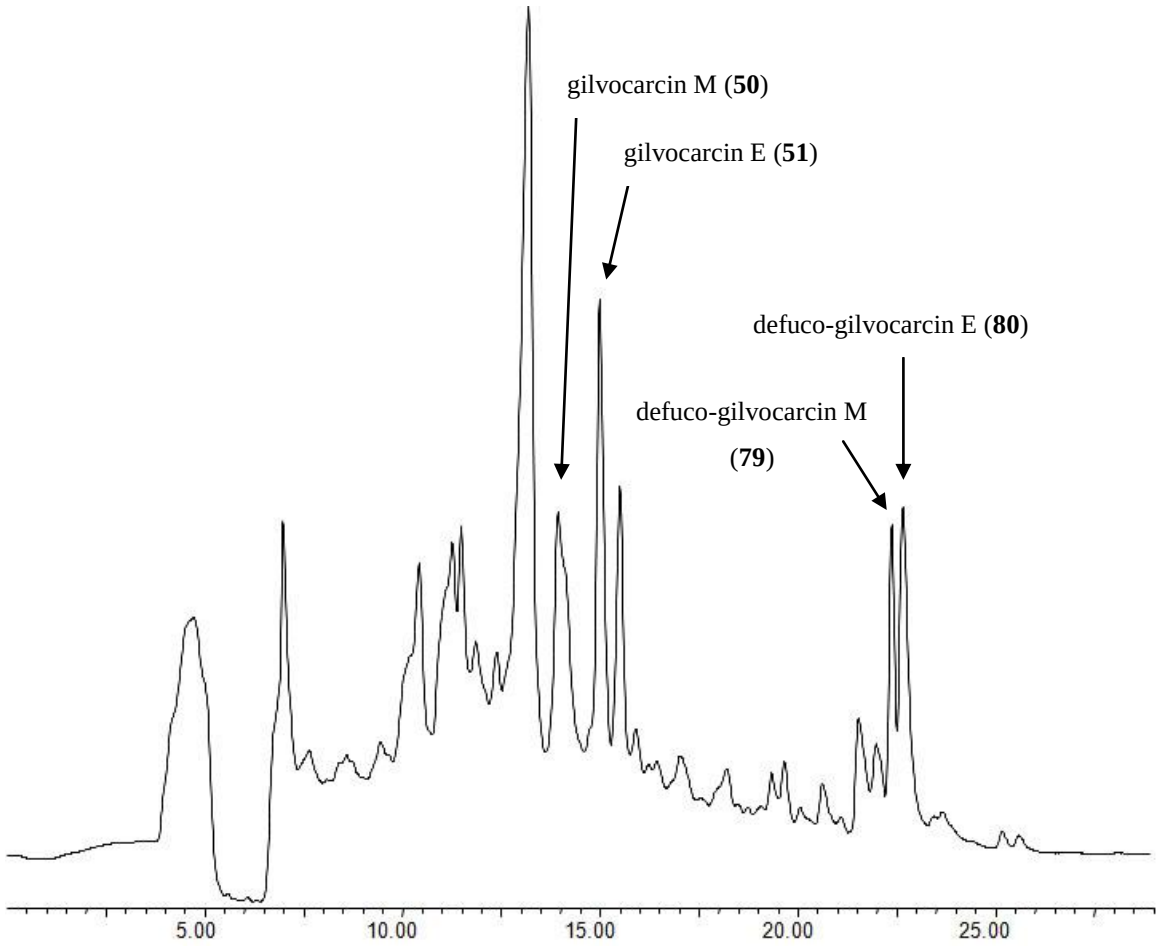


Figure 26. HPLC chromatogram trace of *S. lividans* TK24/cosG9B3-GilGT⁻/cGT-CR.

Chimeric glycosyltransferase and donor substrate flexibility

Each chimeric GT was transferred into individual sugar plasmids designed to produce TDP-D-olivose (**109**), TDP-L-olivose (**110**), TDP-L-digitoxose (**43**), TDP-L-rhodinose (**111**), TDP-L-rhamnose (**112**), TDP-4-amino-4,6-dideoxy-D-glucose (**113**),

TDP-L-daunosamine (**114**), *N*,*N*-didemethyl-TDP-D-desosamine (**115**) or TDP-3-keto-4,6,-dideoxy-D-glucose (**116**) (**Figure 27**). The resulting 54 constructs (see **Table 1**) were transformed into *S. lividans* TK24/cosG9B3-GilGT$^-$ via protoplast transformation and were subsequently screened for the production of novel gilvocarcin analogues.

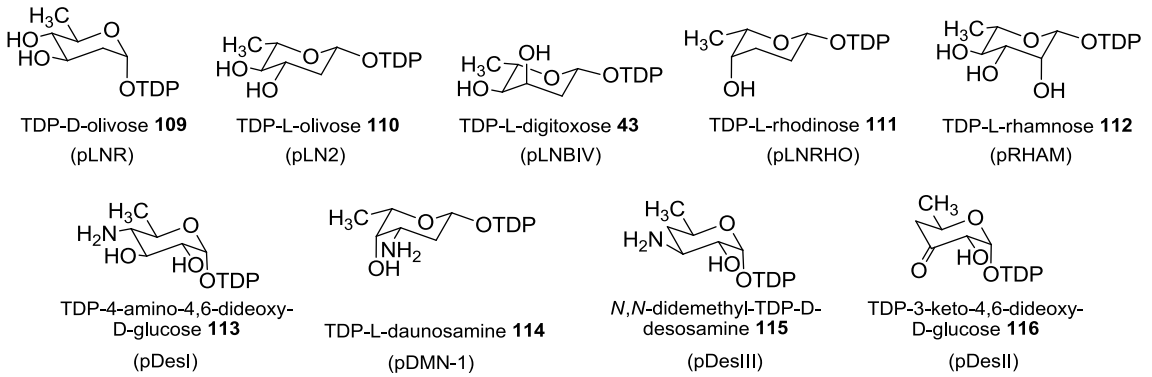


Figure 27. Deoxysugars produced by various sugar plasmids utilized in this study.

As expected, several cGT-CR containing sugar plasmids successfully produced gilvocarcin E (*m*/*z* = 498, observed *m*/*z* = 497 in negative mode ESI-MS) as shown in *S. lividans* TK24/cosG9B3-GilGT$^-$/pLNRHO-CR (**Figure 28**). Interestingly, the analysis of *S. lividans* TK24/cosG9B3-GilGT$^-$/pDesI-CR revealed three new peaks with UV-Vis profiles consistent with the gilvocarcin chromophore (**Figure 29**). Due to the low production yields of the three new peaks, we were unable to obtain clear mass data from the initial fermentation sample. To further characterize the new compounds the fermentation of *S. lividans* TK24/cosG9B3-GilGT$^-$/pDesI-CR was repeated in large scale. Unfortunately, subsequent fermentations did not contain the previously observed gilvocarcin analogues. The sugar plasmid pDesI contains the deoxysugar genes responsible for the production of the amino sugar NDP-4-amino-4,6-dideoxy-D-glucose (**113**). If successfully transferred, it would be expected to have a shorter retention time than that of **49** because of the added amino functionality of **113**. As the *gilGT* deficient strain can only produce gilvocarcin compounds with methyl and ethyl side chains, the presence of three peaks in *S. lividans* TK24/cosG9B3-GilGT$^-$/pDesI-CR was unexpected. It is possible that the peak at 14.0 minutes does not actually correspond to a gilvocarcin analogue at all and the UV-Vis signature observed is from an overlap in the previous

peak at 13.8 minutes. Alternatively, the close proximity of two peaks at 13.8 and 14.0 minutes may be hiding a smaller peak that elutes during this time frame which would indicate the transfer of two different donor substrates to two different aglycones, namely gilvocarcin M and gilvocarcin E.

Interestingly, two additional sugar plasmid/chimeric GT constructs reconstituted the natural gilvocarcin pathway, even though preliminary results indicated they were unable to utilize TDP-D-fucofuranose. *S. lividans* TK24/cosG9B3-GilGT$^-$/pLNR-CG (**Figure 30**) and *S. lividans* TK24/cosG9B3-GilGT$^-$/pLN2-RG (**Figure 31**) both produced a single compound with the same retention time (~14.7 minutes), UV-Vis spectrum and mass of gilvocarcin E (m/z = 498, observed m/z = 497 in negative mode ESI-MS). These results prove two additional chimeric glycosyltransferases are indeed functional; however, it is unclear why the cGT-CG and cGT-RG constructs were unable to complement cosG9B3-GilGT$^-$ mutant upon initial screening, when they obviously retained the ability to do so.

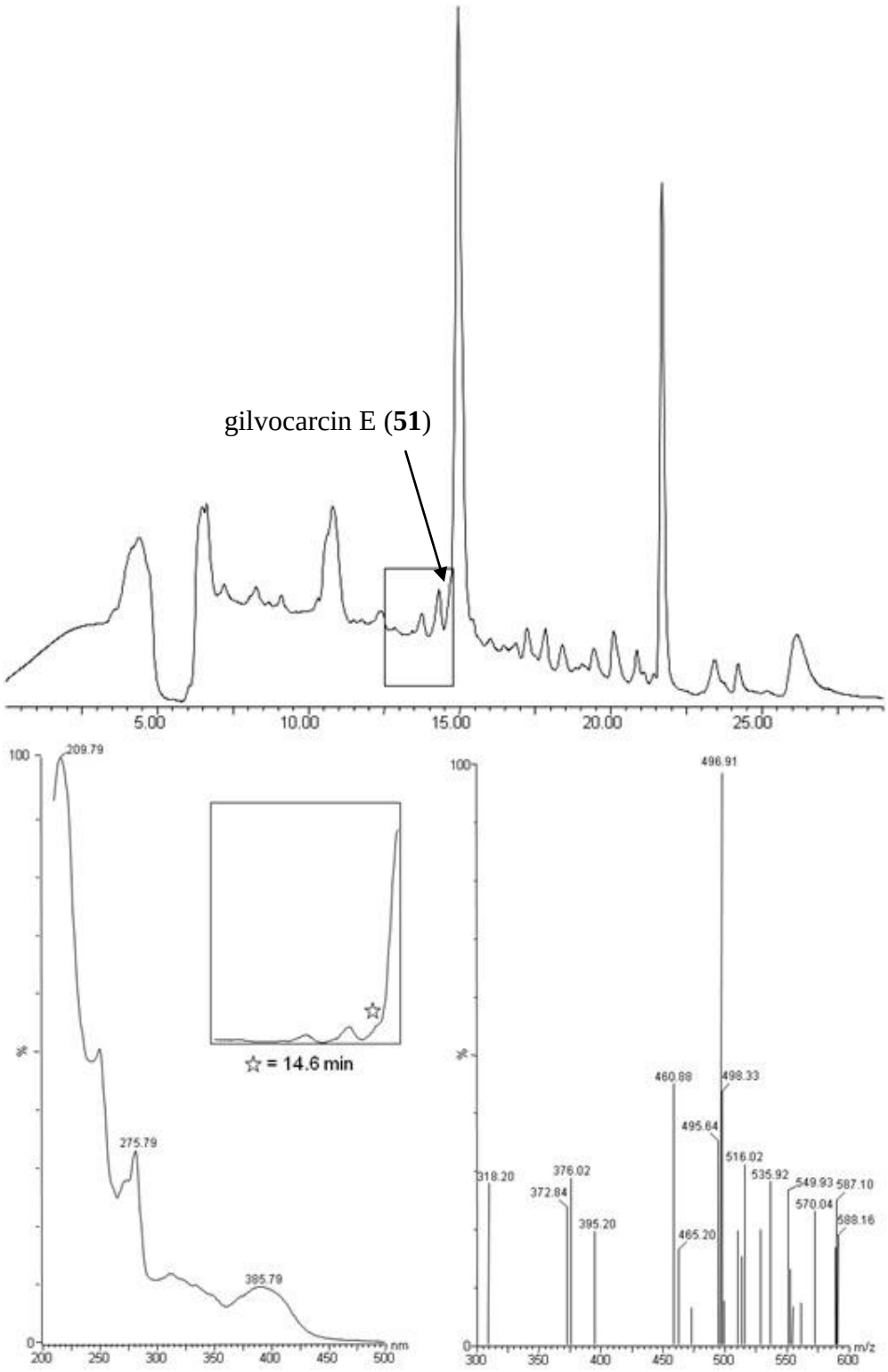


Figure 28. HPLC chromatogram trace of S. lividans TK24/cosG9B3-GilGT-/pLNRHO-CR. This figure also includes the UV-Vis spectrum and mass data of a single peak at 14.6 min.

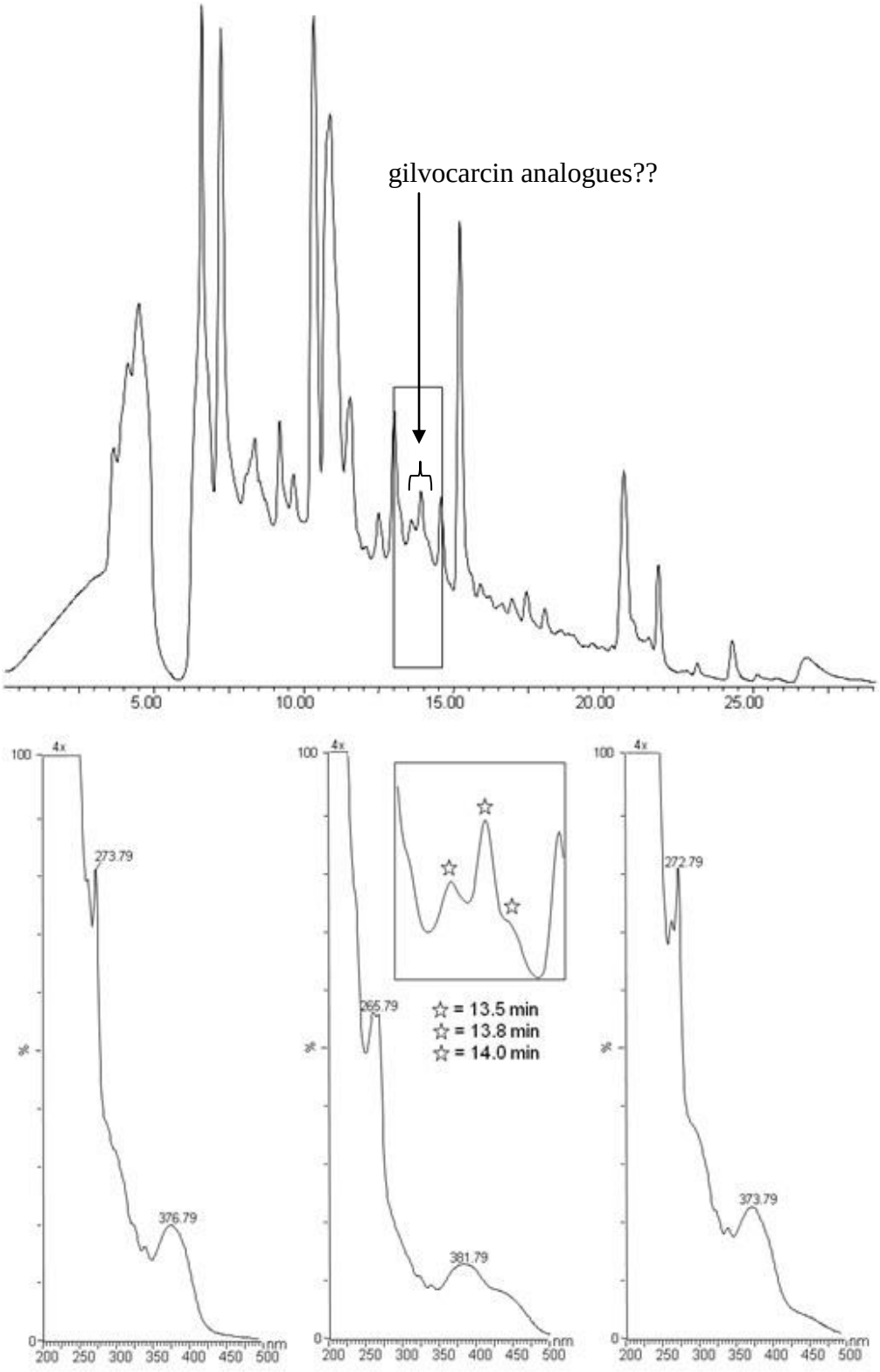


Figure 29. HPLC chromatogram trace of S. lividans TK24/cosG9B3-GilGT-/pDesI-CR. This figure also shows the UV-Vis spectrum of three individual peaks designated with stars.

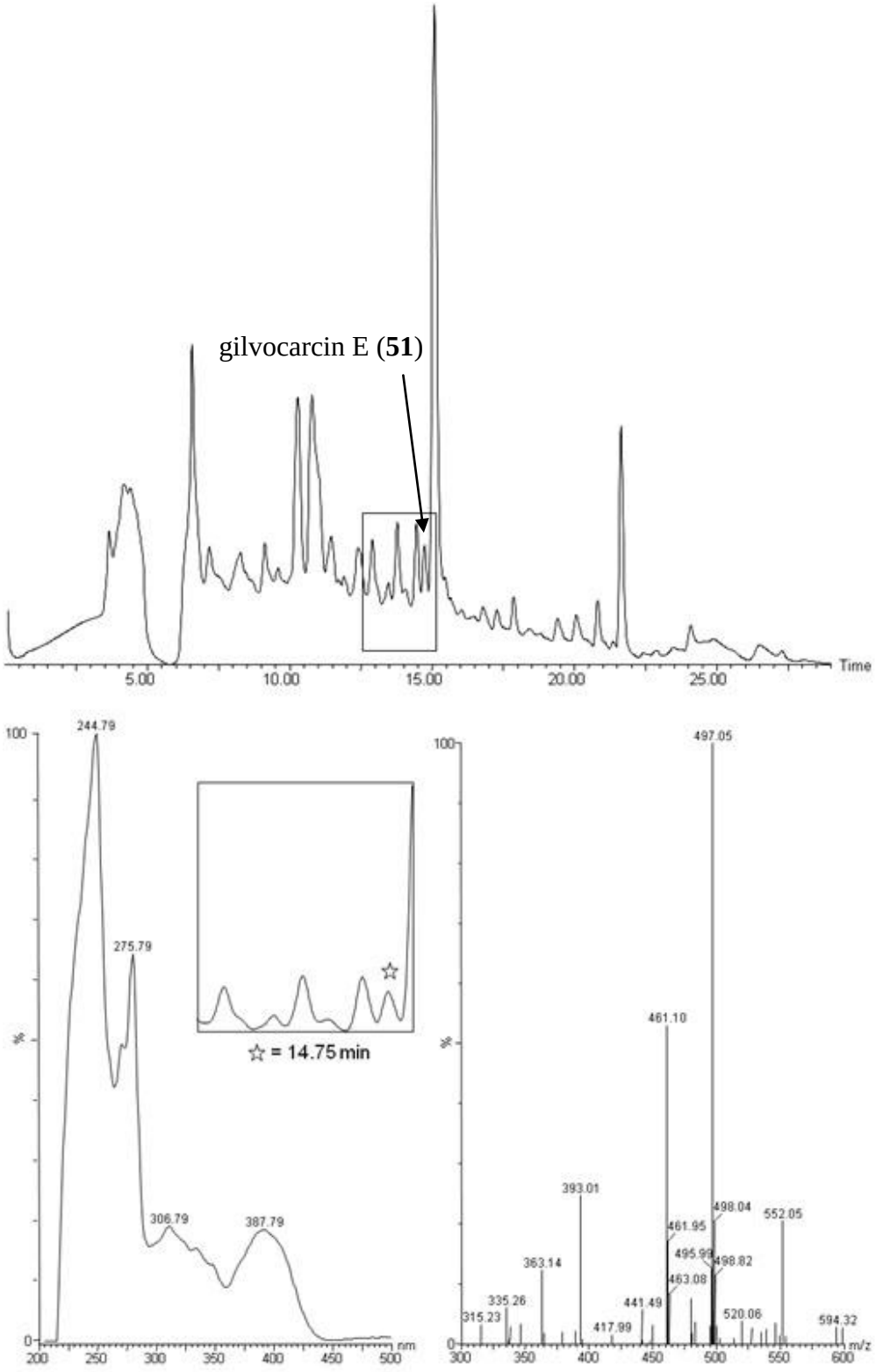


Figure 30. HPLC chromatogram trace of S. lividans TK24/cosG9B3-GilGT-/pLNR-CG. This figure also includes the UV-Vis spectrum and mass data of a single peak at 14.75 min.

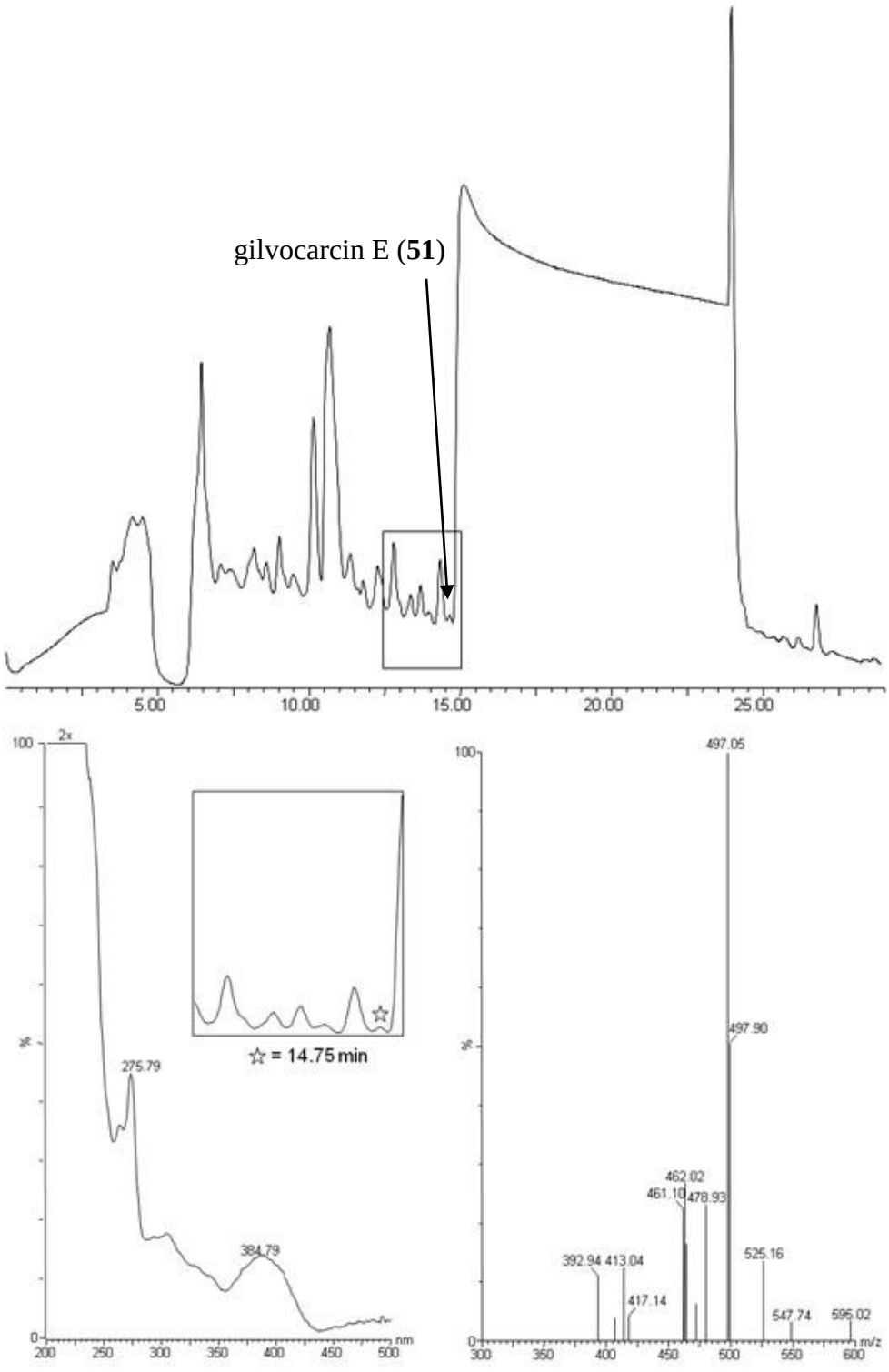


Figure 31. HPLC chromatogram trace of S. lividans TK24/cosG9B3-GilGT-/pLN2-RG. This figure includes the UV-Vis spectrum and mass data of a single peak at 14.75 min.

One underlying issue throughout these experiments were the unpredictable expression levels of gilvocarcin intermediates produced by the host *S. lividans* TK24/cosG9B3-GilGT^{-}. This can be monitored by the accumulation of defuco-gilvocarcin M (**79**) and E (**80**) as these are shunt products of the biosynthetic pathway. In most cases, multiple colonies of the chimeric GT containing recombinant strains had to be screened in order to find a single colony producing **79** and **80** in observable quantities. The typical production of **79** and **80** produced from these strains were low, indicating poor expression of the cosmid cosG9B3-GilGT^{-}.

A double inactivation experiment was designed to increase the expression levels of cosG9B3-GilGT^{-} in which the repressor gene *gilI* was inactivated in addition to *gilGT* through Redirect PCR Targeting.[120] Inactivation of *gilI* was previously found to increase the production yields of gilvocarcins (unpublished results by L. Zhu), therefore it was hypothesized that inactivation of both *gilI* and *gilGT* would similarly increase the production of defuco-gilvocarcins (**78-80**). Construction and subsequent expression of the double inactivated cosmid, cosG9B3-GilI^{-}GilGT^{-} (**Figure 32**), in *S. lividans* TK24, failed to produce defuco-gilvocarcins despite several colony screening attempts (data not shown).

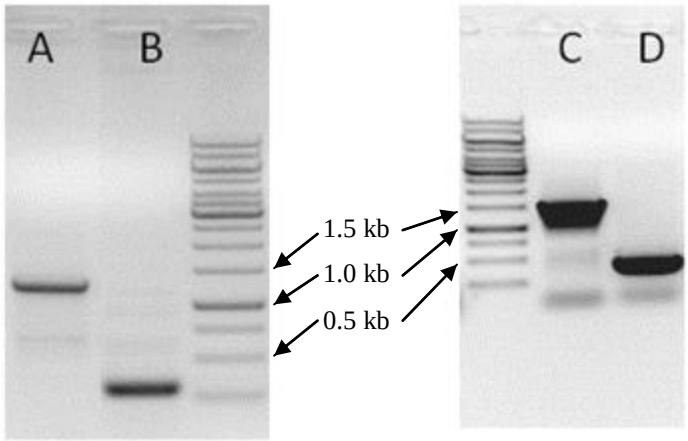


Figure 32. PCR confirmation of double inactivation in cosG9B3-GilI^{-}GilGT^{-}. The inactivation of *gilI* (~1 kb) was confirmed using primers that were flanking its coding sequence by ~100 bp. These primers were used to amplify *gilI* from (A) cosG9B3 and (B) cosG9B3-GilI^{-}GilGT^{-} templates. The small band in lane B is indicative of an FRT scar left after gene inactivation. This process was repeated using primers that flank the coding sequence of *gilU* by ~100 bp. The results are shown in lanes (C) cosG9B3 and (D) cosG9B3-GilI^{-}GilGT^{-}.

Discussion

In order to further broaden the glycodiversity of gilvocarcin-like aryl *C*-glycosides, a domain swapping approach was used to create six chimeric glycosyltransferases utilizing the *N*- and *C*- terminal domains from GilGT, ChryGT and RavGT. The resulting engineered glycosyltransferases, cGT-CR, cGT-CG, cGT-GC, cGT-GR, cGT-RC and cGT-RG, were transformed into *S. lividans* TK24/cosG9B3-GilGT$^-$ to evaluate their ability to transfer TDP-D-fucofuranose, the natural donor substrate for GilGT. In addition, each engineered glycosyltransferase was also placed into nine sugar plasmids and again co-expressed with cosG9B3-GilGT$^-$ in *S. lividans* TK24 to investigate their donor substrate promiscuity. Together these experiments clearly identified three functional chimeric *C*-glycosyltransferases, cGT-CR, cGT-CG and cGT-RG, able to transfer TDP-D-fucofuranose. Preliminary results suggest that cGT-CR may also be able to transfer NDP-4-amino-4,6-dideoxy-D-glucose (**113**). Unfortunately, this study did not result in the isolation of novel gilvocarcin analogues; however, cGT-CR, cGT-CG and cGT-RG are the only examples, to date, of functional chimeric *C*-glycosyltransferases. Ultimately, the ability to test the natural donor substrates for RavGT (TDP-D-ravidosamine (**46**)) and ChryGT (TDP-D-virenose (**44**)) would have been beneficial as it would have allowed for complete parental functionality testing of the chimeric GTs, and would have provided further evidence that rational glycosyltransferase engineering is possible through domain swapping.

As previously discussed the expression levels of *S. lividans*/cosG9B3-GilGT$^-$ were unpredictable and poor. This provided a restricted host for the screening of possible substrate flexible glycosyltransferases. The prospect of an overproducing GilGT deficient strain was exciting as it would have provided much more acceptor substrate for the chimeric GTs to utilize, therefore producing larger quantities of glycosylated product. This led to the generation of cosG9B3-GilI$^-$GilGT$^-$, a double mutant lacking the gilvocarcin biosynthetic repressor gene, *gilI*, and *gilGT* in hopes of increasing defuco-gilvocarcin yields. The expression of cosG9B3-GilI$^-$GilGT$^-$ in *S. lividans* TK24, however, did not produce defuco-gilvocarcin or any gilvocarcin intermediates. This limits the practicality of further chimeric glycosyltransferase testing utilizing this screening

method. The model *in vivo* vehicle for continued GT screening would only produce the acceptor substrate of gilvocarcin V and lack genes for the biosynthesis of TDP-D-fucofuranose, thereby removing donor substrate competition when endogenous deoxysugars are presented to the system. The competing deoxysugar pathways of TDP-D-fucofuranose and those encoded by deoxysugar biosynthetic genes on the sugar plasmids could be an important factor in the inability of the chimeric glycosyltransferases to produce novel gilvocarcin analogues. The biosynthesis of TDP-D-fucofuranose only requires two steps, 4-ketoreduction and ring contraction, beyond the production of TDP-4-keto-6-deoxy-D-glucose (**42, Figure 18). 42** is the branching point for all deoxysugar biosynthetic pathways utilized in this study. If the 4-ketoreductase, GilU, present in cosG9B3-GilGT$^-$ is able to sequester a majority of **42** before the exogenous deoxysugar biosynthetic genes can catalyze their pathway specific reactions, little to no exogenous deoxysugars will be present for the chimeric glycosyltransferase to transfer. This underscores the need for an improved screening host and is a contributing factor for designing **specific aim 4**.

Materials and Methods

Bacterial strains, culture conditions and plasmids

All complementation experiments were carried out in the heterologous host *S. lividans* TK24.[121] The mutant cosmid cosG9B3-GilGT$^-$ was introduced into *S. lividans* TK24 through conjugal transfer according to standard protocols producing *S. lividans* TK24/cosG9B3-GilGT$^-$.[122] Conjugation was carried out on MS agar (20 g/L mannitol, 20 g/L soya flour and 20 g/L agar) supplemented with 10 mM $MgCl_2$ and overlaid with nalidixic acid with appropriate antibiotics after 18 hours. Ex-conjugates were grown on solid M2 media (4 g/L glucose, 10 g/L malt extract, 4 g/L yeast extract, 1g/L $CaCO_3$ and 15 g/L agar) supplemented with appropriate antibiotics. *S. lividans* TK24/cosG9B3-GilGT$^-$ was transformed via protoplast transformation with cGT-CR, cGT-CG, cGT-GC, cGT-GR, cGT-RC, cGT-RG and all 54 sugar plasmid/chimeric glycosyltransferase constructs according to standard protocols.[122] Protoplasts were regenerated on R2YE agar media and overlaid after 18 hours with R3 soft agar supplemented with appropriate antibiotics.[122] Regenerated protoplasts were transferred to solid M2 agar supplemented

with appropriate antibiotics. For a comprehensive list of strains and plasmids used in this study refer to **Table 1**. Lysogeny broth (LB) was utilized for growing *Escerichia coli* (*E. coli*) strains throughout this study.

When antibiotics/antifungals were required for strain selection, 100 µg/mL of ampicillin, 50 µg/mL of apramycin, 25 µg/mL of thiostrepton, 50 µg/mL of chloramphenicol, 50 µg/mL of kanamycin and 25 µg/mL nalidixic acid was used.

Table 1. Strains and plasmids used in the chimeric glycosyltransferase study.

Strain/Plasmid	Characteristics and relevance	References
E. coli XL1-Blue-MRF	Host for routine cloning	Stratagene
E. coli ET12567/pUZ8002	Host for conjugal transfer	MacNeil, D. et al.[123-124]
PCR-Blunt II-TOPO	PCR fragment cloning vector	Invitrogen
pEM4	Streptomyces expression vector	Quiros, L. et al.[112]
cosG9B3	Template for *gilGT*	Fischer, C. et al.[91]
cosRav32	Template for *ravGT*	Kharel, M. et al.[117]
cosChry1-1	Template for *chryGT*	Kharel, M. et al.[117]
cosG9B3-GilGT$^-$	*gilGT* deficient mutant	Liu, T. et al.[42]
S. lividans TK24 (SLTK24)	Streptomyces heterologous host	Kieser, T. et al.[121]
S. lividans TK24/cosG9B3-GilGT$^-$	Host for screening chimeric GTs	Liu, T. et al.[42]
cGT-CG	Chimeric GT in pEM4	This study
cGT-CR	Chimeric GT in pEM4	This study
cGT-GC	Chimeric GT in pEM4	This study
cGT-GR	Chimeric GT in pEM4	This study
cGT-RC	Chimeric GT in pEM4	This study
cGT-RG	Chimeric GT in pEM4	This study
pLNR-CG, CR, GC, GR, RC, RG	(6x) Chimeric GT in pLNR	This study
pLN2-CG, CR, GC, GR, RC, RG	(6x) Chimeric GT in pLN2	This study
pRHAM-CG, CR, GC, GR, RC, RG	(6x) Chimeric GT in pRHAM	This study
pLNBIV-CG, CR, GC, GR, RC, RG	(6x) Chimeric GT in pLNBIV	This study
pDMN-1-CG, CR, GC, GR, RC, RG	(6x) Chimeric GT in pDMN-1	This study
pLNRHO-CG, CR, GC, GR, RC, RG	(6x) Chimeric GT in pLNRHO	This study
pDesI-CG, CR, GC, GR, RC, RG	(6x) Chimeric GT in pDesI	This study
pDesII-CG, CR, GC, GR, RC, RG	(6x) Chimeric GT in pDesII	This study
pDesIII-CG, CR, GC, GR, RC, RG	(6x) Chimeric GT in pDesIII	This study
SLTK24/cosG9B3-GilGT$^-$/all GTs	(60x) recombinant strains	This study

Fermentation and metabolite screening

Each recombinant strain was grown in a single 250 mL baffled Erlenmeyer flask containing 100 mL of liquid SG media (20 g/L glucose, 10 g/L soy peptone, 2 g/L $CaCO_3$, 0.001 g/L cobalt-II-chloride, pH 7.2) with appropriate antibiotics. After five days of fermentation at 28 ^oC with reciprocal shaking (250 rpm) 25 mL of the fermentation broth was extracted 1:1 with ethyl acetate. The organic phase was separated

and removed under vacuum. The dried extract was reconstituted with 500 µL of methanol and screened on an HPLC-MS consisting of a Micromass ZQ 2000 (Waters) equipped with HPLC (Waters Alliance 2695) and a photodiode array detector (Waters 2996). The HPLC utilized a Symmetry C_{18} (4.6 x 250 mm, 5 µm) column running a linear gradient of acetonitrile and acidified water at a flow rate of 0.5 mL/min (solvent A= 0.1% formic acid in H_2O; solvent V = acetonitrile; 0-15 min. 25% B to 100% B; 16-24 min. 100% B; 25-26 min. 100% to 25% B; 27-29 min. 25% B).

DNA isolation, DNA manipulation and PCR

Plasmid DNA isolations were conducted using GeneJet Plasmid mini-prep kits (Fermentas). All restriction endonuclease digestions, alkaline phosphate treatments, ligations, and other DNA manipulations were performed according to standard protocols.[125] Native *Pfu* polymerase (Stratagene) was used to amplify each domain from *gilGT*, *ravGT* and *chryGT* from cosG9B3, cosRav32 and cosChry1-1, respectively.[91,117] Oligonucleotide primers used in this study are summarized in **Table 3**. A typical reaction consisted of the following:

1 µL DNA template (~50-100 ng)
1 µL Forward primer (~200-250 ng)
1 µL Reverse primer (~200-250 ng)
1 µL dNTPs (10 mM)
5 µL Native *Pfu* buffer (10x)
2.5 µL DMSO
37.5 µL Distilled water (dH_2O)
1 µL Native *Pfu* polymerase (2.5 U/µl)

PCR reactions were carried out on a TechGene thermal cycler (Techne) with typical cycling conditions summarized in **Table 2**.

Table 2. PCR cycling conditions for Native *pfu* polymerase amplification.

Step	Temperature	Duration	Cycles
Initial denaturation	98 ^oC	5 min.	1
Denaturation	94 ^oC	45 sec.	25
Annealing	Primer T_M – (5-10)	45 sec.	
Extension	72 ^oC	1.5 min./kb of target	
Final extension	72 ^oC	10 min.	1

Each glycosyltransferase was amplified as two fragments corresponding to the *N*- and *C*-terminal domains using forward (F) and middle reverse (MR) primer pairs as well as middle forward (MF) and reverse (R) primer pairs, respectively. The resulting PCR amplified fragments were gel purified using a QIAquick Gel Extraction Kit (Qiagen) and cloned into PCR-Blunt II-TOPO (Invitrogen).

Table 3. Oligonucleotide sequence of primers used to amplify *gilGT*, *ravGT* and *chryGT* domains.

Primer	Oligonucleotide sequence (5′-3′)
GilGT-F	CTTCTCTAGACGCGACAAGGACCGGTCACGGGAGGG
GilGT-MR	GGTGGCTAGCGCCCCGACCAGGCCGGGCGGGCAGAT
GilGT-MF	GCGCTAGCCACCTTCATGCGCTGGACCCCGCAC
GilGT-R	TGGGGATGCATCGAATTCGAAAGCTTCGCTTCATCCTTTCCGCGCGAGGAG
RavGT-F	CTGTCTAGACGACGGCCCCGACCCTACGGAGAA
RavGT-MR	CGTCGCTAGCGACGGGAGGAGGCTCGGGGGCAGAT
RavGT-MF	TCGCTAGCGACGCCCGGCCGCCCCATGCGCTGG
RavGT-R	CGTCGGATGCATTGGAATTCGAAGCTTGCTGGCTCACCGAGAGGGTGCTCATCC
ChryGT-F	CGCGCTCTAGAGGGTGGCCAGCCCCGAAGGAGCAGC
ChryGT-MR	CTCGGCTAGCGGCGGGATCAGGCTCGGCGGGCAGAT
ChryGT-MF	CCGCCGCTAGCCGAGCGCGGCCGGCCGATGAGATGG
ChryGT-R	CCGGATGCATCGAATTCCAAGCTTTGGTGGGCCGGTGGCACGGGGTCAGCG

Cloning and preparation of chimeric glycosyltransferases

A general cloning strategy used for domain swapping of *gilGT*, *ravGT* and *chryGT* is summarized in **Figures 24** and **25**. Through primer design, the amplification of each *N*-terminal domain created a *Xba*I site 30-40 nucleotides upstream of their start codons which ensured the amplification of their putative ribosomal binding sites (RBS). On the opposite side, a unique in-frame *Nhe*I restriction site was engineered into the sequences of each glycosyltransferase as described previously. The *N*- and *C*-terminals of a given GT were amplified as two overlapping fragments that shared this *Nhe*I site designed as the shuffling point for domain swapping. In this context, all *C*-terminal domains also contained the in-frame *Nhe*I site. Immediately downstream of the stop codon of each *C*-terminal a *Hind*III and *Eco*RI site were inserted for use in later cloning strategies.

The resulting PCR amplified fragments were cloned into PCR-Blunt II-TOPO vector which is used to quickly and easily clone and amplify blunt DNA fragments, typically from PCR reactions. Several TOPO clones were screened for fragment

insertion as well as fragment orientation by *Spe*I/*Nhe*I (for C-terminal clones) and *Nhe*I/*Eco*RV (for N-terminal clones) double digestion analysis. Orientation was confirmed because shuffling would occur by transferring the *C*-terminal domain from their respective TOPO clones to various *N*-terminal containing TOPO clones. If the orientations are in opposite directions the resulting shuffling would not create a continuous open reading frame (ORF). These restriction analysis revealed six clones with appropriate orientation; C1.1, G1.1, R1.19, C2.7, G2.2 and R2.4. The naming reflected the parent GT, the domain and the exact colony screened. For example, R1.19 is R = RavGT, 1 = N-terminal domain and 19 = colony number.

Shuffling was achieved by removing the *C*-terminal domains from C2.7, G2.2 and R2.4 through *Nhe*I/*Eco*RV double digestion and ligating them into similarly prepared C1.1, G1.1 and R1.19. Screening the resulting constructs by *Eco*RI/*Xba*I double digestion revealed 6 chimeric glycosyltransferases; CG.4, CR.6, GR.12, GC.6, RC.20 and RG.2. These were transferred to the *Streptomyces* expression vector pEM4 by *Eco*RI/*Xba*I restriction digest and subsequent ligation into pEM4. Screening through *Hind*III digestion analysis revealed six completed constructs; CG4.4, CR6.16, GR12.7, GC6.16, RC20.7 and RG2.15. These pEM4 constructs were renamed to cGT-CG, cGT-CR, cGT-GR, cGT-GC, cGT-RC and cGT-RG, respectively.

The six chimeric glycosyltransferases were individually integrated into nine sugar plasmids directing the biosynthesis towards nine structurally distinct deoxysugars (see **Table 4** and **Figure 27**). Each sugar plasmid was found to contain a single *Hind*III restriction site which was used to transfer the intact chimeric glycosyltransferase including the constitutively active promoter, $ermE^{*}$p, from cGT-CG, cGT-CR, cGT-GR, cGT-GC, cGT-RC and cGT-RG. The resulting 54 constructs contained a single chimeric glycosyltransferase and deoxysugar biosynthetic genes for the generation of an activated deoxysugar (see **Table 4**).

Table 4. Sugar plasmids used in this study, and their expected deoxysugar product.

Sugar	Plasmid	Sugar	Plasmid
NDP-D-olivose (**109**)	pLNR[126]	NDP-L-rhodinose (**111**)	pLNRHO[126]
NDP-L-olivose (**110**)	pLN2[126]	NDP-4-amino-4,6-dideoxy-D-glucose (**113**)	pDesI[a]
NDP-L-rhamnose (**112**)	pRHAM[46]	NDP-3-keto-4,6-dideoxy-D-glucose (**116**)	pDesII[a]
NDP-L-digitoxose (**43**)	pLNBIV[126]	*N*,*N*-didemethyl-NDP-D-desosamine (**115**)	pDesIII[a]
NDP-L-daunosamine (**114**)	pDMN-1[a]		

[a]Unpublished plasmids constructed by M. Kharel

Nucleotide sequences of generated chimeric glycosyltransferases

The following sequences (5′-3′) represent the entire chimeric glycosyltransferases as found in pEM4 between the restriction sites *Xba*I and *Eco*RI. The shuffling point (*Nhe*I) is indicated as ***NNN*** while the remaining restriction sites, start codons and stop codons are designated by *NNN*, **NNN** and NNN, respectively.

cGT-CG

*TCTAGA*GGGTGGCCAGCCCCGAAGGAGCAGC**ATG**AAAGTCCTCTTCATCGCC
GCGGGAACGAGCCCGGCGGGGGTCTTCGCCCTCGCCCCTCTCGCGACGGCGG
TGCGCAACGCAGGGCACGAGATCCTGGTGGCGGCCTTCGACGAGCTGACCCC
GGCCGTCGCGTCCGTCGGACTGCCCGTCGTGGCCGTCGCCGACGGCCACACG
ACCGAGAGCATCAAGGGACTCGACCGGCCGGGCGGACCGATCGGGTTCCCCT
GGGCACCCGAGCAGGAACTGCCGTACGTGGGACGTTGGTTCGGCCGCCAGGC
GGCCGTCGCCATGGACGGACTCCTGCGGCTCGCCGACGTGTGGCGCCCCGAC
CTCGTGGTCGGTGGCACGGACGCCCACGCCGCCGCCCTGCTCGGCTCCCGGC
TGGGCATCCCGTACGTCCGCCAGGCGTGGGACTGGCTCCACTTCGACGGAGC
GGAGCGCTACGCGAACGACGAACTGGCCCCCGAACTGGCGCAGGCCGGACT
CGAACGGCTCCCGGCACCCGACCTGTTCATCGACATCTGCCCGCCGAGCCTG
ATCCCGCC***GCTAGC***CACCTTCATGCGCTGGACCCCGCACAACATGCAGCGGG
CGATCGAGCCGTGGATGCTGACGGCTCCGGACGCCGGGCGCGTGTGCCTGAC
GATGGGAAGCTTCCGGTACGCCTTCCCCGGCGCGATGGACCGCATCTCGGCC
ATCGTCGAAGGGCTGCTGGAGCTCGAGGTCGAGGTCGTCGTGGCCATCGGCG
AGGCGGAGGGGCAGCGGCTGCAGGAGAAGTACCCCCGGGTGCGCGCCGGCT
GGATCCCGCTGGAGGCCATCCTCCCGACCTGTGAGGTGATCATCCATCCGGC
GGGCGGACTGACGGCCATCAACGCCATCAACACGGCGACTCCGCAGCTGATC

CTCAACCCCTTCGAGGCCTTCGTCCCGAGGCTGAAGCACCTCACGGACTACG
GGTGCGCGCGGACGCTCTACCGCGAGGAGGGCACCCCGGAGGCGATCACGC
AGGTGGTCAAGGAGATGCTCGGGGATCCGTCCTACTCCTCCAGGGCCCGGAG
ACTGGCGGAGCAGGGCGCGACCGCGCCGACGGCCGTGGGCATGGTGCCGCT
GATCGAAGACCTCCTCGCGCGGAAAGGATGAAGCG*AAGCTT*TC*GAATTC*

cGT-CR

*TCTAGA*GGGTGGCCAGCCCCGAAGGAGCAGC**ATG**AAAGTCCTCTTCATCGCC
GCGGGAACGAGCCCGGCGGGGGTCTTCGCCCTCGCCCCTCTCGCGACGGCGG
TGCGCAACGCAGGGCACGAGATCCTGGTGGCGGCCTTCGACGAGCTGACCCC
GGCCGTCGCGTCCGTCGGACTGCCCGTCGTGGCCGTCGCCGACGGCCACACG
ACCGAGAGCATCAAGGGACTCGACCGGCCGGGCGGACCGATCGGGTTCCCCT
GGGCACCCGAGCAGGAACTGCCGTACGTGGGACGTTGGTTCGGCCGCCAGGC
GGCCGTCGCCATGGACGGACTCCTGCGGCTCGCCGACGTGTGGCGCCCCGAC
CTCGTGGTCGGTGGCACGGACGCCCACGCCGCCGCCCTGCTCGGCTCCCGGC
TGGGCATCCCGTACGTCCGCCAGGCGTGGGACTGGCTCCACTTCGACGGAGC
GGAGCGCTACGCGAACGACGAACTGGCCCCCGAACTGGCGCAGGCCGGACT
CGAACGGCTCCCGGCACCCGACCTGTTCATCGACATCTGCCCGCCGAGCCTG
ATCCCGCC***GCTAGC***GACGCCCGGCCGCCCCATGCGCTGGGTGCCCGGCAACC
GCCAGCGGTCGCTGGAGCCGTGGATGTACACCAAGGGCGAGCGCCCCCGGAT
CTGCGTCACCTACGGCAGCTTCCGTACCGCCATGCCGCAGATCTTCGAGCACC
TGTGCGCCCTGCTGTCCCGGCTCGTGGACCTGGACGCCGAGATCGTGGTCGC
GGCGAACGAGGCCGCGACGGAGAAGCTGCGCGAGCGCTTCCCGCAGGTCCG
CGCCGGCTGGGTGCCGCTGGAGTTCCTGCTGCGCACCTGCGACGGCATCGTG
CACAGCGGCGGGCTGACCGCGCTGAACGCCATGTCCGCCGGCATCCCGCAGG
TGGTGCTCAACCAGTTCGTCGCCTTCGAGCCGTCGCTGGCCCTGTTGCAGCAG
CAGGGCTCCGCCGTCGTGCTGCACCGCGAGGAGGGCTCCCCCGACGGCACCT
TCGAAGCCTGCCGCAAGGTCCTGTCCGACGGGAGCTACGGCCAGCAGGCCCG
GGTCCTCGCCGACGAGCTCAACTCCCTGCCCACCCCGAACGACGTGGTGAAG
GACCTGGAAGGACTGGTGCGCGGATGAGCACCCTCTCGGTGAGCCAGC*AAGC*
*TT*C*GAATTC*

cGT-GC

*TCTAGA*CGCGACAAGGACCGGTCACGGGAGGGCGCCGC**GTG**AAGGCCCTCTT
CTACGCCGCGGGCACGAGCCCGGCCAGCGCGTTCGCCATCGGACCCCTCGCC
TCGGCACTCCGGTTGTCCGGGCACGACGTCCTGGTGGCGTCCTTCGAGGAGA
TGTCCGGCGCCGTCACCGGCATCGGTCTGCCGTCCCTCCCGGTGGCCCGGGG
GCACACGACCGAGAGCATCAAGGCGGCCGCCGGCGGCAAGCCGGCGATCGA
GTACCCCCACCGGCCCGAACAGGAGATGCCCTACCTGGGCCACTGGTTCGGG
CGTCAGGGAAGCCACGTCTTCGACGACCTGGTGGACGTCGCCCGGACGTGGG
GCGCCGACGTCCTGATCGCGGGGAGCCAGGGACACGGTGCCGAGATCGCCG
CCCGGTTCCTCGGCATCCCCTTCGTCCGGCAGTCCTGGGACCTGTTCGACGTC
GGCGGCTACGAGGAGTACCTCCTCGAGGAGATGGCGGACGAGCTGGCCAGG
ATCGGCTCGGACGCACTGCCCGACCCCTCACTGAAGATCGACATCTGCCCGC
CCGGCCTGGTCGGGGC***GCTAGC***CGAGCGCGGCCGGCCGATGAGATGGGTGCC
GGGCAACCGCCAGCGCAAGCTGGAGCCGTGGATGTACGGCAAGGGCGAGCG
CGCCCGGGTGTGCGTCACGCTCGGCAGCTTCCGCACCGCCATGCCGGAGATG
TTCGCCTACCTGTGCGCACTGGTGGAGCGGCTGACGGCACTCGACGCCGAGA
TCGTCGTCGCCGCCGACGAGACGGCCTCCCGGAAGATCCGGGAGCGCTTCCC
CGGCGTACGCGCCGACTGGGTGCCGATGGAGTTCCTCGTGCGCACCTGCGAC
ACGATCGTGCACGCGGGCGGCCTCACCACTCTCAACGCCATGGCGGCCGGCA
CCCCGCAGGTGGTGGTGAACCAGTTCCAGGCGTTCGAACCCTCGCTGCGGCT
GCTCGAGCGGCAGGGATGCAGCGTGGTCCTGCACCGGGACGACAAGTCCCCG
GACAACACCTTCGACGCCTGCGAGCGCATCCTCTCCGACGACGGGTACGCGC
ACCGGGCAGGTGAGCTGGCCCGCGAACTGGGCGCCCTGCCCCCGCCGGCCAC
GGTCGTCGGCGACCTGGAGGCGCTGGCCCGCCGCTGACCCCGTGCCACCGGC
CCACCA*AAGCTT*G*GAATTC*

cGT-GR

*TCTAGA*CGCGACAAGGACCGGTCACGGGAGGGCGCCGC**GTG**AAGGCCCTCTT
CTACGCCGCGGGCACGAGCCCGGCCAGCGCGTTCGCCATCGGACCCCTCGCC
TCGGCACTCCGGTTGTCCGGGCACGACGTCCTGGTGGCGTCCTTCGAGGAGA

TGTCCGGCGCCGTCACCGGCATCGGTCTGCCGTCCCTCCCGGTGGCCCGGGG
GCACACGACCGAGAGCATCAAGGCGGCCGCCGGCGGCAAGCCGGCGATCGA
GTACCCCCACCGGCCCGAACAGGAGATGCCCTACCTGGGCCACTGGTTCGGG
CGTCAGGGAAGCCACGTCTTCGACGACCTGGTGGACGTCGCCCGGACGTGGG
GCGCCGACGTCCTGATCGCGGGGAGCCAGGGACACGGTGCCGAGATCGCCG
CCCGGTTCCTCGGCATCCCCTTCGTCCGGCAGTCCTGGGACCTGTTCGACGTC
GGCGGCTACGAGGAGTACCTCCTCGAGGAGATGGCGGACGAGCTGGCCAGG
ATCGGCTCGGACGCACTGCCCGACCCCTCACTGAAGATCGACATCTGCCCGC
CCGGCCTGGTCGGGGC***GCTAGC***GACGCCCGGCCGCCCCATGCGCTGGGTGCC
CGGCAACCGCCAGCGGTCGCTGGAGCCGTGGATGTACACCAAGGGCGAGCG
CCCCCGGATCTGCGTCACCTACGGCAGCTTCCGTACCGCCATGCCGCAGATCT
TCGAGCACCTGTGCGCCCTGCTGTCCCGGCTCGTGGACCTGGACGCCGAGAT
CGTGGTCGCGGCGAACGAGGCCGCGACGGAGAAGCTGCGCGAGCGCTTCCC
GCAGGTCCGCGCCGGCTGGGTGCCGCTGGAGTTCCTGCTGCGCACCTGCGAC
GGCATCGTGCACAGCGGCGGGCTGACCGCGCTGAACGCCATGTCCGCCGGCA
TCCCGCAGGTGGTGCTCAACCAGTTCGTCGCCTTCGAGCCGTCGCTGGCCCTG
TTGCAGCAGCAGGGCTCCGCCGTCGTGCTGCACCGCGAGGAGGGCTCCCCCG
ACGGCACCTTCGAAGCCTGCCGCAAGGTCCTGTCCGACGGGAGCTACGGCCA
GCAGGCCCGGGTCCTCGCCGACGAGCTCAACTCCCTGCCCACCCCGAACGAC
GTGGTGAAGGACCTGGAAGGACTGGTGCGCGGATGAGCACCCTCTCGGTGAG
CCAGC*AAGCTT*C*GAATTC*

cGT-RC

*TCTAGA*CGACGGCCCCGACCCTACGGAGAAGCC**ATG**AAAGTCCTGTTCATCG
CAGCGGGAACCAGCCCCGCGGGCGTCTTCGCCCTCGCCCCGCTCGCCACGGC
GGTGCGCAACGCCGGGCACGAGATCCTGGTCGCCTCCTTCGACGAGTTGACC
TCCAGCATCGAGGCGATCGGGCTGCCTCCGGTGGCCGTGGTCACCGAGCACA
CCACCGAGAGCATCAAGCAGCTGGACCGCCCCGGCGGCCCGATCGAGTTCCC
CTGGTCGCCCGACCAGGAACTGCCTTACGTGGGCCGTTGGTTCGGCCGGCAG
GCCGCCGTCAGCCTCGACGGACTCCTCGAACTGACCCGGCACTGGCGCCCCG
ACCTCGTCATCGGCGGCACCGACGCGCACGCCGCCGGGCTCGTCGCGGCCCA

CCTCGGCGTCCCGCACGTCCGCCAGGCGTGGGACTGGCTGCACTTCGGCGGC
GCCGAGCAGTACGCCGCCGAAGAGCTCGCCCCCGAGCTGGAGCGGCTGGGC
CTGGCCGAACTGCCCGCCCCGGACATGTTCTTGGACATCTGCCCGCCGAGCCT
CCTCCCGTC***GCTAGC***CGAGCGCGGCCGGCCGATGAGATGGGTGCCGGGCAAC
CGCCAGCGCAAGCTGGAGCCGTGGATGTACGGCAAGGGCGAGCGCGCCCGG
GTGTGCGTCACGCTCGGCAGCTTCCGCACCGCCATGCCGGAGATGTTCGCCT
ACCTGTGCGCACTGGTGGAGCGGCTGACGGCACTCGACGCCGAGATCGTCGT
CGCCGCCGACGAGACGGCCTCCCGGAAGATCCGGGAGCGCTTCCCCGGCGTA
CGCGCCGACTGGGTGCCGATGGAGTTCCTCGTGCGCACCTGCGACACGATCG
TGCACGCGGGCGGCCTCACCACTCTCAACGCCATGGCGGCCGGCACCCCGCA
GGTGGTGGTGAACCAGTTCCAGGCGTTCGAACCCTCGCTGCGGCTGCTCGAG
CGGCAGGGATGCAGCGTGGTCCTGCACCGGGACGACAAGTCCCCGGACAAC
ACCTTCGACGCCTGCGAGCGCATCCTCTCCGACGACGGGTACGCGCACCGGG
CAGGTGAGCTGGCCCGCGAACTGGGCGCCCTGCCCCCGCCGGCCACGGTCGT
CGGCGACCTGGAGGCGCTGGCCCGCCGCTGACCCCGTGCCACCGGCCCACCA
*AAGCTT*G*GAATTC*

cGT-RG

*TCTAGA*CGACGGCCCCGACCCTACGGAGAAGCC**ATG**AAAGTCCTGTTCATCG
CAGCGGGAACCAGCCCCGCGGGCGTCTTCGCCCTCGCCCCGCTCGCCACGGC
GGTGCGCAACGCCGGGCACGAGATCCTGGTCGCCTCCTTCGACGAGTTGACC
TCCAGCATCGAGGCGATCGGGCTGCCTCCGGTGGCCGTGGTCACCGAGCACA
CCACCGAGAGCATCAAGCAGCTGGACCGCCCCGGCGGCCCGATCGAGTTCCC
CTGGTCGCCCGACCAGGAACTGCCTTACGTGGGCCGTTGGTTCGGCCGGCAG
GCCGCCGTCAGCCTCGACGGACTCCTCGAACTGACCCGGCACTGGCGCCCCG
ACCTCGTCATCGGCGGCACCGACGCGCACGCCGCCGGGCTCGTCGCGGCCCA
CCTCGGCGTCCCGCACGTCCGCCAGGCGTGGGACTGGCTGCACTTCGGCGGC
GCCGAGCAGTACGCCGCCGAAGAGCTCGCCCCCGAGCTGGAGCGGCTGGGC
CTGGCCGAACTGCCCGCCCCGGACATGTTCTTGGACATCTGCCCGCCGAGCCT
CCTCCCGTC***GCTAGC***CACCTTCATGCGCTGGACCCCGCACAACATGCAGCGG
GCGATCGAGCCGTGGATGCTGACGGCTCCGGACGCCGGGCGCGTGTGCCTGA

CGATGGGAAGCTTCCGGTACGCCTTCCCCGGCGCGATGGACCGCATCTCGGC
CATCGTCGAAGGGCTGCTGGAGCTCGAGGTCGAGGTCGTCGTGGCCATCGGC
GAGGCGGAGGGGCAGCGGCTGCAGGAGAAGTACCCCCGGGTGCGCGCCGGC
TGGATCCCGCTGGAGGCCATCCTCCCGACCTGTGAGGTGATCATCCATCCGG
CGGGCGGACTGACGGCCATCAACGCCATCAACACGGCGACTCCGCAGCTGAT
CCTCAACCCCTTCGAGGCCTTCGTCCCGAGGCTGAAGCACCTCACGGACTAC
GGGTGCGCGCGGACGCTCTACCGCGAGGAGGGCACCCCGGAGGCGATCACG
CAGGTGGTCAAGGAGATGCTCGGGGATCCGTCCTACTCCTCCAGGGCCCGGA
GACTGGCGGAGCAGGGCGCGACCGCGCCGACGGCCGTGGGCATGGTGCCGC
TGATCGAAGACCTCCTCGCGCGGAAAGGATGAAGCG*AAGCTT*TC*GAATTC*

Double inactivation of *gilI* and *gilGT*

The double inactivation experiment was carried out using a modified PCR-targeting REDIRECT protocol.[127] An overview of this process is described in section 4.1: Inactivation of *gilP* and *gilQ*.

2.2 Genetic isolation of glycosyltransferase involved in polycarcin biosynthesis

Polycarcin V (**61**) is a gilvocarcin like aryl *C*-glycoside produced by *Streptomyces polyformus*.[61] **61** differs from **49** only by its sugar moiety, namely L-rhamnose (**92**). Interestingly, *S. polyformus* produces **61** and **49** as a concomitant 1:1 mixture, meaning the cluster either contains two individual glycosyltransferases responsible for the transfer of each deoxysugar or a single glycosyltransferase with donor substrate flexibility transfers both L-rhamnose and D-fucofuranose.

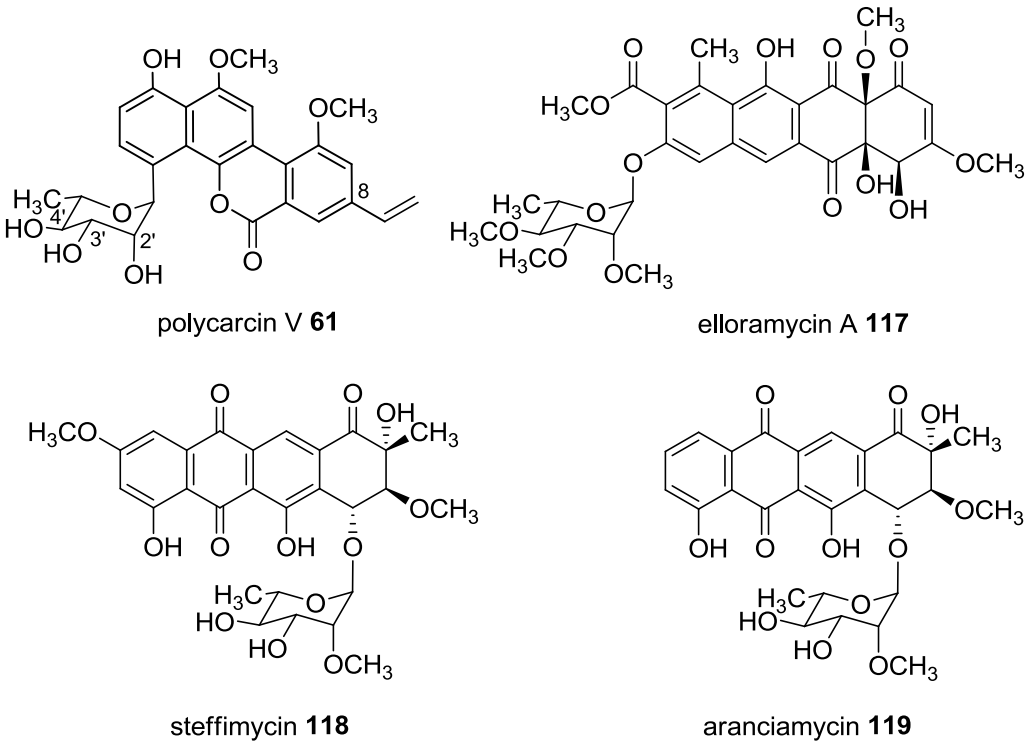


Figure 33. Structures of polycarcin V (**61**), elloramycin (**117**), steffimycin (**118**) and aranciamycin (**119**).

Rhamnose transferring glycosyltransferases are commonly referred to as rhamnosyltransferases, and in general seem to contain inherent substrate promiscuity. In this context, rhamnosyltransferases are an exciting combinatorial biosynthetic tool for creating glycosylated natural product analogues. The type II PKS derived natural products elloramycin (**117**), steffimycin (**118**) and aranciamycin (**119**) all contain an *O*-glycosidically linked L-rhamnose moiety, and have had their respective rhamnosyltransferase donor substrate specificities investigated (**Figure 33**).[46,49-50,128-132,171] Together, AraGT (aranciamycin), StfGT (steffimycin) and ElmGT (elloramycin) have been found to transfer 15 individual donor substrates to their respective aglycones (**Figure 34**). The donor substrates transferred included both D- and L- sugars and incorporated several classes of sugars such as non-deoxy (**120**), deoxy (**92** and **121**), dideoxy (**93** and **122- 128, 156**) and trideoxysugars (**129-131**).

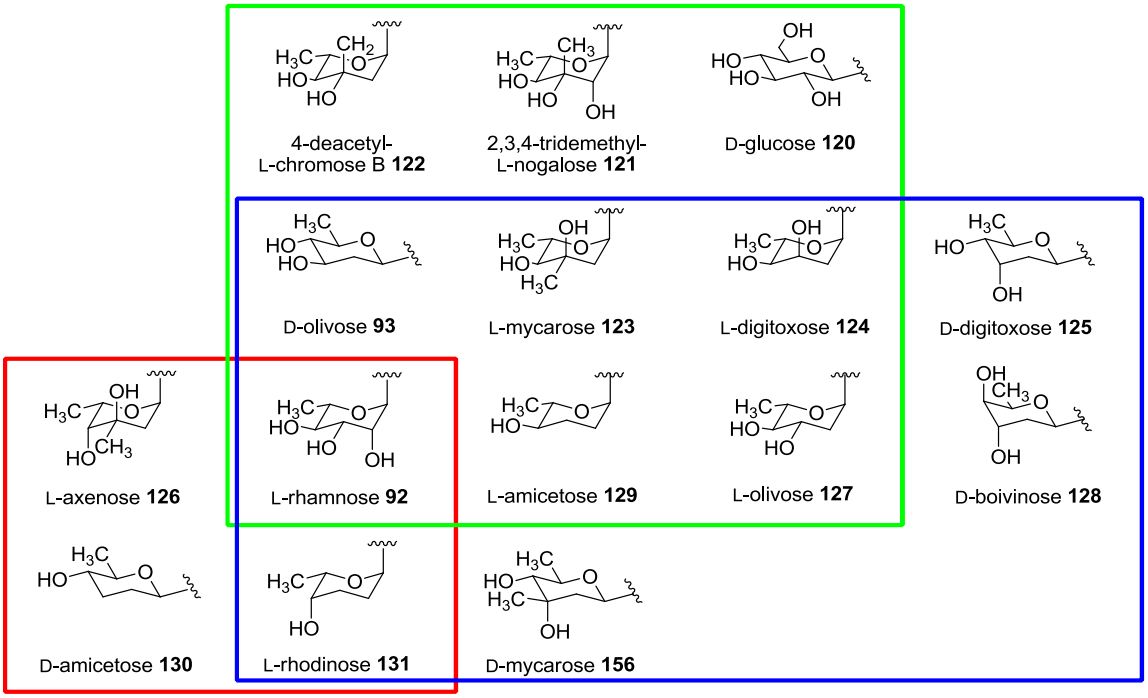


Figure 34. Sugar moieties transferred by the rhamnosyltransferases AraGT (red), StfGT (green) and ElmGT (blue).

The relaxed donor substrate flexibility of several rhamnosyltransferases makes the polycarcin pathway specific glycosyltransferase extremely desirable. The identification of a rhamnosyltransferase that catalyzes a *C*-glycosylation and exhibits donor substrate promiscuity would allow for an additional route to the production of gilvocarcin-like aryl *C*-glycoside analogues.

Experimental design

In **specific aim 1b**, a genomic cosmid library of *Streptomyces polyformus* will be constructed and screened for the presence of gilvocarcin-like glycosyltransferases. Candidate glycosyltransferases will be sequenced and their gene products compared to known gilvocarcin-like GTs (*gilGT*, *chryGT* and *ravGT*). The identified glycosyltransferase/s will then be confirmed as the gilvocarcin V and/or polycarcin V pathway specific glycosyltransferase. Heterologous expression and successful complementation of the newly identified GT in *S. lividans* TK24/cosG9B3-GilGT$^-$ will

indicate a TDP-D-fucofuranose transferring GT. Similarly, the same GT cloned into an L-rhamnose generating sugar plasmid and subsequent complementation in *S. lividans* TK24/cosG9B3-GilGT$^-$ will allow for rhamnosyltransferase activity screening. The identified rhamnosyltransferase will then be cloned into various sugar plasmids and transformed into *S. lividans* TK24/cosG9B3-GilGT$^-$ to evaluate its donor substrate promiscuity.

Results

S. polyformus cosmid library and screening

The genomic cosmid library of *S. polyformus* was constructed in the *E. coli-Streptomyces* shuttle vector, pOJ446.[133] Typically, two probes are used to screen the genomic library to increase the probability of capturing an entire gene cluster within a single cosmid. Probing the *S. polyformus* genomic library for polycarcin biosynthetic genes would normally involve a keto-acyl synthase (KS_α) and a NDP-glucose-4,6-dehydratase (4,6-DH) probe as polycarcin biosynthesis requires the catalytic activity of both of these enzymes for its backbone and sugar moiety, respectively.[91,117] For this study, however, we were interested in only the glycosyltransferase/s present in the cluster, therefore probing with a KS_α and 4,6-DH probe was unnecessary. Instead, degenerate primers (PlcGT-Deg-F1/PlcGT-Deg-R1, see **Table 7**) were constructed from conserved *N*-terminal regions (responsible for acceptor binding) of *gilGT*, *ravGT* and *chryGT* and used to amplify partial candidate glycosyltransferase sequences directly from *S. polyformus* genomic DNA (**Figure 35**). The resulting glycosyltransferase probe was used to screen the *S. polyformus* genomic library as a digoxigenin (DIG)-labeled glycosyltransferase gene probe. Initial colony hybridization led to over one hundred positively hybridized colonies (**Figure 36,** A). These were taken for a second round of hybridization in which each colony was scratched out in order to positively confirm hybridization of individual colonies (**Figure 36,** B). The second round of hybridization narrowed the colony count to twenty nine. Further restriction digest and PCR analysis identified two distinct cosmids, cosPlc47 and cosPlc75, to be investigated further (see section 2: *S. polyformus* cosmid library construction and screening).

Figure 35. Nucleotide sequence alignment of the N-terminal region of *gilGT*, *ravGT* and *chryGT*. This figure shows the conserved residues (boxed) used for generating degenerate gilvocarcin-like glycosyltransferase primers.

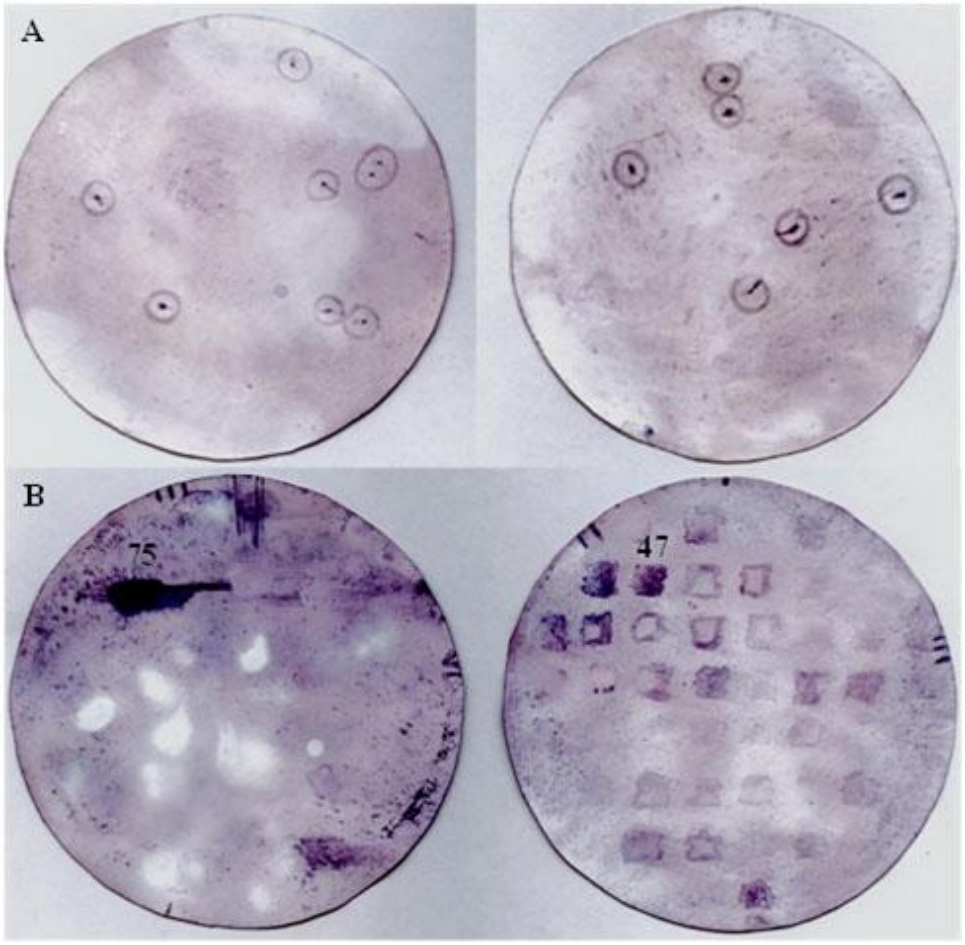


Figure 36. Selection of (A) initial and (B) conformational colony hybridization discs used in screening the *S. polyformus* genomic library.

Partial sequence analysis of cosPlc75

The degenerate glycosyltransferase primers used to probe the cosmid library were designed to amplify a small *N*-terminal portion of gilvocarcin-like glycosyltransferases. Amplification of this sequence from cosPlc47 and cosPlc75 revealed both cosmids

contained an identical glycosyltransferase with high sequence homology to GilGT. To get a complete gene sequence, cosPlc75 was used for primer walking experiments (SeqWright) which extended the sequenced region roughly 500 nucleotides flanking the internal glycosyltransferase sequence. This resulted in the full gene sequence of the initially identified glycosyltransferase, now designated *plcGT*.

Sequence analysis of *plcGT* revealed a very high sequence homology to *gilGT* (84% aa identity). When compared to *ravGT* and *chryGT*, however, the amino acid identity dropped to roughly 52% suggesting PlcGT may act on the furanose sugar D-fucofuranose. Primer walking also allowed for the identification of genes flanking *plcGT*, which included the complete sequence of *plcV* (*gilV* homologue, 84% aa identity) and the partial sequence of *plcM* (*gilM* homologue) and *plcOIII* (*gilOIII* homologue) (**Figure 37**). The expression of both cosPlc75 and cosPlc47 in S. lividans TK24 failed to produce gilvocarcin, polycarcin or any known gilvocarcin pathway intermediates suggesting incomplete clusters (data not shown).

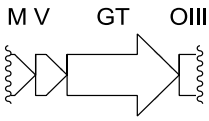


Figure 37. Gene sequences identified from partial sequencing of cosPlc75.

Complementation studies in *S. lividans* TK24/cosG9B3-GilGT$^-$

With the complete sequence of *plcGT* determined, complementation of *S. lividans* TK24/cosG9B3-GilGT$^-$ was carried out to determine the ability of PlcGT to transfer TDP-D-fucofuranose (**47**). Transformation of pPlcGT into *S. lividans* TK24/cosG9B3-GilGT$^-$ resulted in successful reconstitution of the gilvocarcin pathway as observed by the production of gilvocarcin M and E (**Figure 38**). These results support the role of PlcGT as the glycosyltransferase responsible for the production of gilvocarcins in *S. polyformus*. As only a single gilvocarcin-like glycosyltransferase was identified through cosmid library screening, it is possible that PlcGT also transfers TDP-L-rhamnose (**92**) toward the production of polycarcins. To investigate this hypothesis, PlcGT was cloned into the sugar plasmid pRHAM (pRHAM-PlcGT) and transformed into *S. lividans* TK24/cosG9B3-GilGT$^-$. Interestingly, the resultant recombinant strain

failed to produce either polycarcins or gilvocarcins (**Figure 39**). This result was surprising as PlcGT was previously shown to produce gilvocarcins. In addition, the inability to transfer L-rhamnose suggests that PlcGT is not involved in polycarcin biosynthesis. Alternatively, it is possible that TDP-L-rhamnose is not being produced in the fermentations from pRHAM, or that PlcGT utilizes an NDP-L-rhamnose substrate other than TDP-L-rhamnose.

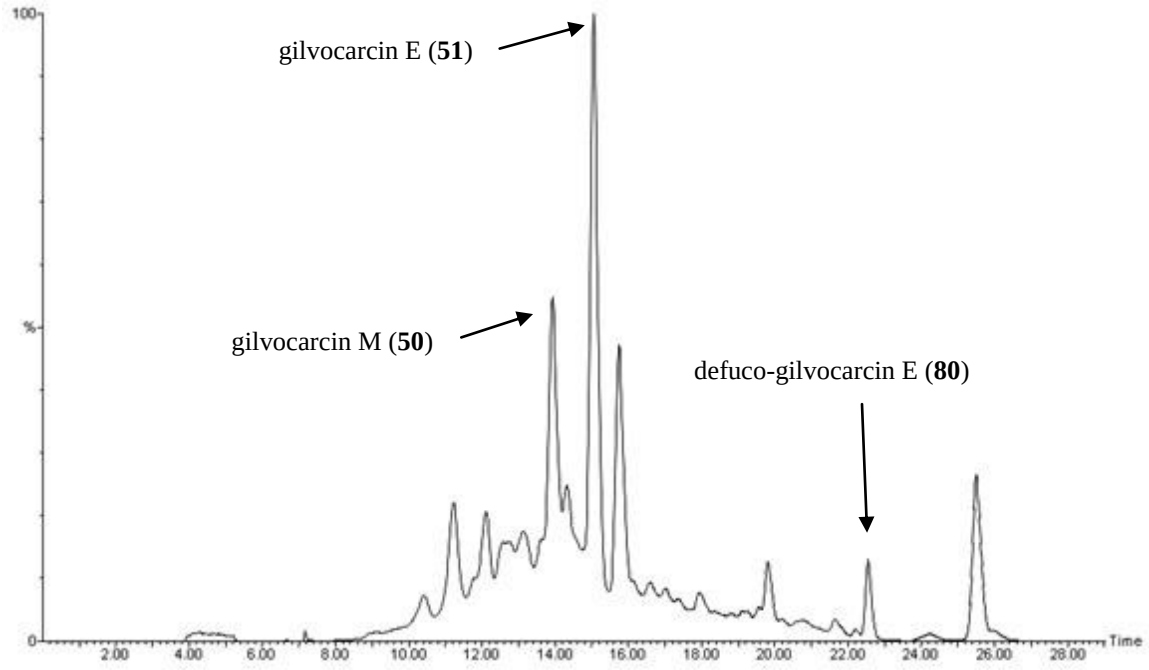


Figure 38. HPLC chromatogram trace of *S. lividans* TK24/cosG9B3-GilGT⁻/pPlcGT.

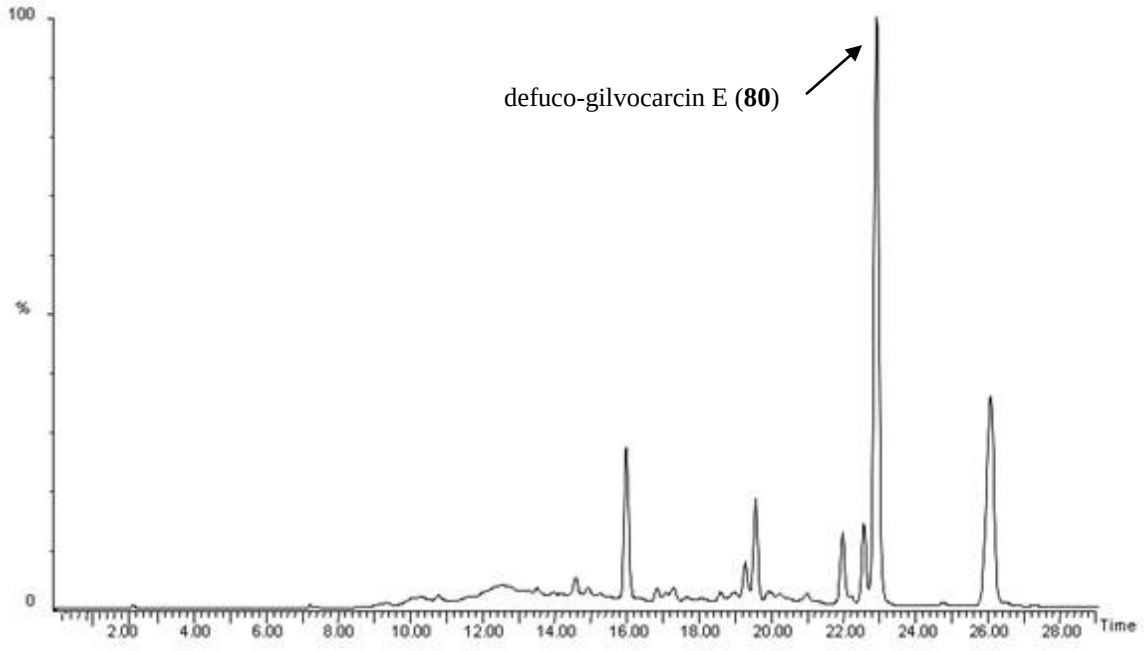


Figure 39. HPLC chromatogram trace of *S. lividans* TK24/cosG9B3-GilGT⁻/pRHAM-PlcGT.

Donor substrate specificity of PlcGT

To evaluate the donor substrate promiscuity of PlcGT, two constructs were created in which *plcGT* was placed into the sugar plasmids pLNBIV and pLNRHO, respectively. These sugar plasmids were chosen for two reasons; 1) their representative sugars have been transferred by other rhamnosyltransferases and 2) they produce L-sugars which in general are underrepresented in gilvocarcin type aryl *C*-glycosides. The resulting constructs, pLNBIV-PlcGT and pLNRHO-PlcGT, were transformed into *S. lividans* TK24/cosG9B3-GilGT⁻ and their metabolites were screened as previously described. No new metabolites were observed from the resulting recombinant strains, however, gilvocarcin E production was found to be restored in *S. lividans* TK24/cosG9B3-GilGT⁻/pLNRHO-PlcGT (**Figure 40 and 41**). These results suggest PlcGT exhibits a rigid donor substrate preference, and may not be a suitable glycosyltransferase for glycodiversification studies.

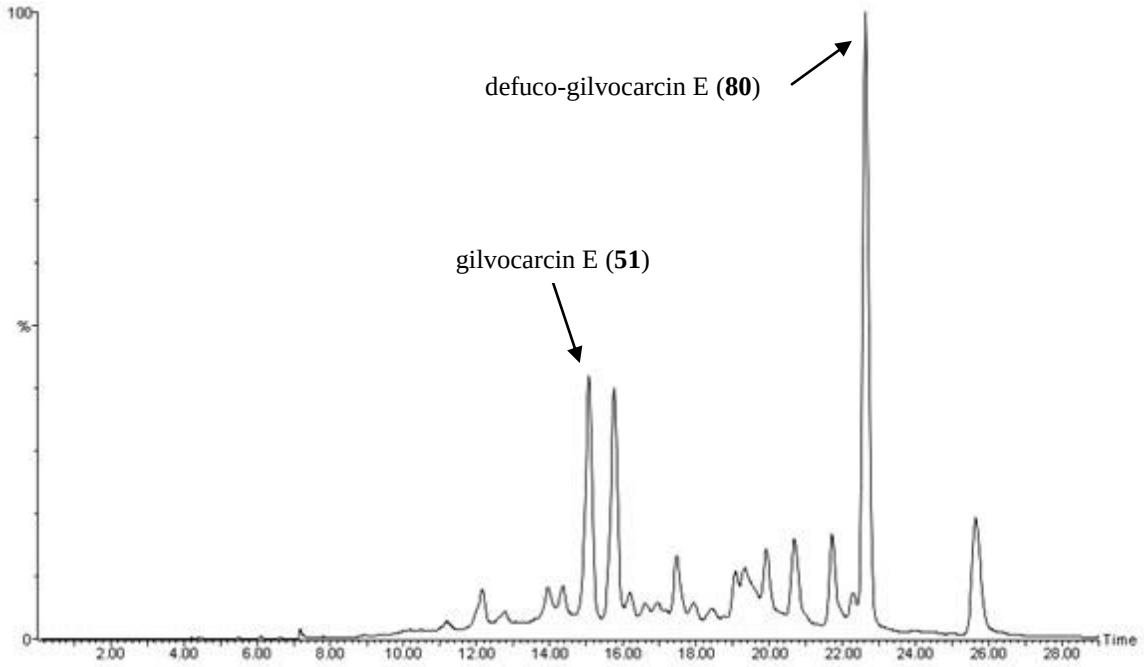


Figure 40. HPLC chromatogram trace of *S. lividans* TK24/cosG9B3-GilGT⁻/pLNRHO-PlcGT.

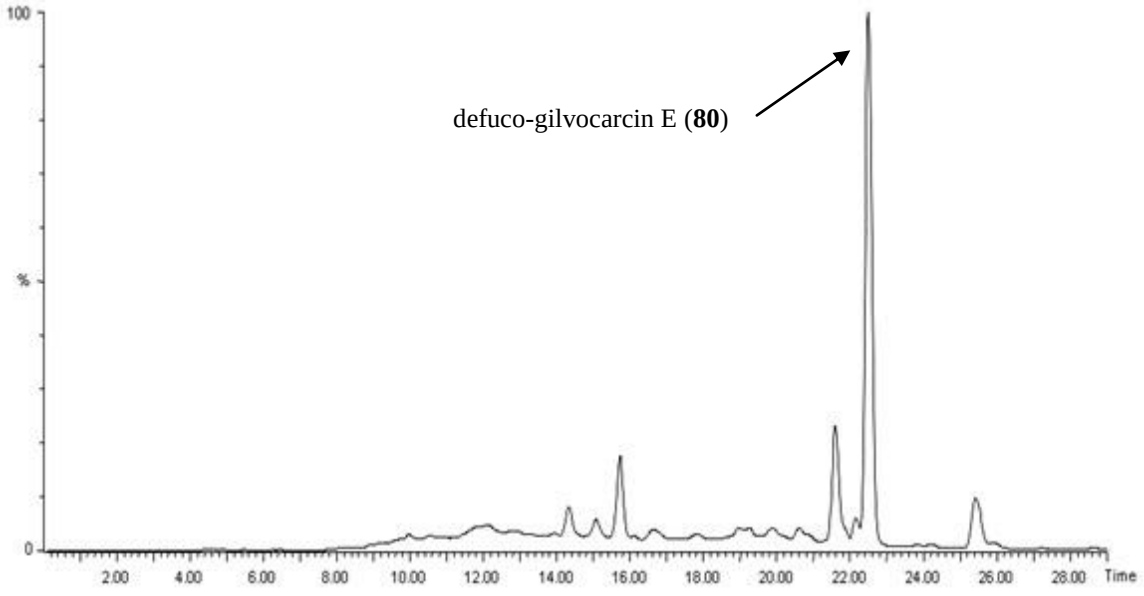


Figure 41. HPLC chromatogram trace of *S. lividans* TK24/cosG9B3-GilGT⁻/pLNBIV-PlcGT.

Discussion

The natural concomitant production of gilvocarcins as well as polycarcins in *S. polyformus* could arise from the activity of a single donor substrate flexible glycosyltransferase or from two distinct glycosyltransferases responsible for transferring D-fucofuranose and L-rhamnose, respectively. Particular interest was given to the isolation of a rhamnosyltransferase as previous studies have established rhamnosyltransferases as powerful tools for the glycodiversification of natural products.[46,49-50,128-132,171]

In an attempt to identify a substrate flexible *C*-rhamnosyltransferase, a genomic library of *S. polyformus* was generated and screened using degenerate primers based on conserved *N*-terminal regions of gilvocarcin-like glycosyltransferases. Screening of the library resulted in the isolation of two distinct cosmids, cosPlc47 and cosPlc75, which were shown through preliminary sequencing to contain an identical glycosyltransferase. A single cosmid, cosPlc75, was chosen for primer walking in order to obtain the entire gene sequence of the identified glycosyltransferase. The resulting ORF was designated *plcGT*, and was used in various complementation studies to determine its parental functionality as well as donor substrate flexibility.

Complementation studies showed PlcGT could only accommodate D-fucofuranose and was unable to transfer L-rhamnose, L-digitoxose or L-rhodinose to defuco-gilvocarcin. Together, these results suggest PlcGT to be a substrate specific glycosyltransferase involved in only the gilvocarcin biosynthetic pathway of *S. polyformus*. If this hypothesis is true, then an additional glycosyltransferase is present within the genomic library of *S. polyformus* responsible for the production of polycarcins.

It is possible that the second glycosyltransferase may not share conserved *N*-terminal sequences compared to gilvocarcin-like glycosyltransferases. This would explain the inability to amplify such a glycosyltransferase from the gilvocarcin-like glycosyltransferase based degenerate primers. In this context, two additional sets of degenerate primers based on conserved regions of several rhamnosyltransferases (PlcGT-Deg-F2/PlcGT-Deg-R2 and PlcGT-Deg-F2/ PlcGT-Deg-R3) were used in an attempt to amplify the missing glycosyltransferase from *S. polyformus* (**Table 7**). Despite several attempts, no PCR products were obtained from these experiments (data not shown). The

inability to isolate a second glycosyltransferase from all degenerate primers used in this study suggests, contrary to complementation studies, the lack of a second glycosyltransferase. Again, the complementation studies using PlcGT to transfer TDP-L-rhamnose are not conclusive as it is unknown if TDP-L-rhamnose was actually present in the fermentations. Furthermore, it is possible that the natural substrate of PlcGT may be an alternative nucleotide activated NDP-L-rhamnose.

As described in section 2.1, *S. lividans* TK24/cosG9B3-GilGT$^-$ is not an ideal host for screening glycosyltransferases for the transfer of sugars other than TDP-D-fucofuranose. In this context, negative complementation results can not be used as conclusive evidence to support the lack of glycosyltransferase activity towards a given donor substrate. There are several possible contributing factors to negative complementation results, especially when additional deoxysugar genes are expressed in *S. lividans* TK24/cosG9B3-GilGT$^-$. Without the use of real time quantitative reverse transcription PCR (real-time qRT-PCR), there is no way to ensure the glycosyltransferase or exogenous deoxysugar genes are being expressed *in vivo* unless you observe a glycosylated product. Additionally, the competing deoxysugar biosynthetic pathways could lead to the formation of an unwanted and incompatible donor substrate for the glycosyltransferase causing a possible false-negative result or inhibition.

It should be noted that several gilvocarcin glycodiversification experiments were being conducted in parallel to this study. It was found only after the isolation of the *plcGT* sequence that GilGT could naturally transfer L-rhamnose. Unlike the studies described herein, the aforementioned study was conducted in the biosynthetically compromised TDP-D-fucofuranose mutant, cosG9B3-GilU$^-$. Unfortunately, neither PlcGT nor any other GT can be tested in this host as it contains the natural gilvocarcin glycosyltransferase, GilGT. Additionally, attempts to express soluble PlcGT in *E. coli* failed despite several attempts (data not shown).

In regards to the gilvocarcin and polycarcin biosynthetic pathways in *S. polyformus*, taking into consideration all available evidence it is likely that PlcGT is responsible for transferring both D-fucofuranose and L-rhamnose despite negative complementation results suggesting otherwise. Taking into consideration that, (a) PlcGT was the only gilvocarcin-like glycosyltransferase amplified from *S. polyformus* genomic DNA; (b)

PlcGT and GilGT share 83% amino acid identity; and (c) GilGT can naturally transfer L-rhamnose it is unlikely that a second glycosyltransferase exists for the explicit transfer of L-rhamnose during polycarcin biosynthesis. In addition, the conserved orientation and high amino acid sequence identity of the genes found in the partial polycarcin cluster suggests the *plc* cluster to be highly similar to the *gil* cluster and only produces polycarcin because unlike *S. griseoflavus* Gö 3952 (*gil* cluster), *S. polyformus* carries the deoxysugar biosynthetic genes to produce activated L-rhamnose.

Materials and Methods

Bacterial strains, culture conditions and plasmids

All complementations, conjugations, protoplast transformations and culturing conditions were carried out as described in section 2.1: Bacterial strains, culture conditions and plasmids. A comprehensive list of strains and plasmids used in this study can be found in **Table 5.**

Table 5. Strains and plasmids used in the polycarcin glycosyltransferase study.

Strain/Plasmid	Characteristics and relevance	References
E. coli XL1-Blue-MRF	Host for routine cloning	Stratagene
E. coli ET12567/pUZ8002	Host for conjugal transfer	MacNeil, D. et al.[123-124]
PCR-Blunt II-TOPO	PCR fragment cloning vector	Invitrogen
pGEM-T Easy Vector	PCR fragment cloning vector	Promega
pEM4	*Streptomyces* expression vector	Quiros, L. et al.[112]
pOJ446	Used for constructing library	Bierman, M et al.[133]
cosG9B3-GilGT$^-$	*gilGT* deficient mutant	Liu, T. et al.[42]
S. lividans TK24 (SLTK24)	Streptomyces heterologous host	Kieser, T. et al.[121]
S. lividans TK24/cosG9B3-GilGT$^-$	Host for screening chimeric GTs	Liu, T. et al.[42]
S. polyformus	Used for genomic library	Li, Y. et al.[61]
pPlcGT (1)[a]	pEM4 containing *plcGT*	This study
pRHAM-PlcGT (2)[a]	pRHAM containing *plcGT*	This study
pLNBIV-PlcGT (3)[a]	pLNBIV containing *plcGT*	This study
pLNRHO-PlcGT (4)[a]	pLNRHO containing *plcGT*	This study
SLTK24/cosG9B3-GilGT$^-$/(1)[a]	Produces gilvocarcins	This study
SLTK24/cosG9B3-GilGT$^-$/(2)[a]	Produces defuco-gilvocarcins	This study
SLTK24/cosG9B3-GilGT$^-$/(3)[a]	Produces defuco-gilvocarcins	This study
SLTK24/cosG9B3-GilGT$^-$/(4)[a]	Produces gilvocarcins	This study

[a]Numbering used only for table simplification.

Fermentation and metabolite screening

The recombinant strains created in this study were fermented and screened as described in section 2.1: Fermentation and metabolite screening.

DNA isolation, DNA manipulation and PCR

Plasmid/cosmid DNA isolations and manipulations were carried out as described in section 2.1: DNA isolation, DNA manipulation and PCR. Advantage-GC 2 polymerase mix (Clonetech; Advantage GC-2) was used to amplify glycosyltransferase probes from *S. polyformus* using primers summarized in **Table 7**. A typical reaction consisted of the following:

1 µL DNA template (~50-100 ng)
1 µL Forward primer (~200-250 ng)
1 µL Reverse primer (~200-250 ng)
1 µL dNTPs (10 mM)
10 µL GC 2 PCR buffer (5x)
10 µL GC Melt (5M)
25 µL Distilled water (dH_2O)
1 µL Advantage-GC 2 polymerase mix (50x)

PCR reactions were carried out on a TechGene thermal cycler (Techne) with typical cycling conditions summarized in **Table 6**.

Table 6. PCR cycling conditions for Advantage-GC 2 polymerase mix amplification.

Step	Temperature	Duration	Cycles
Initial denaturation	94 ^{o}C	3 min.	1
Denaturation	94 ^{o}C	30 sec.	25
Annealing	Primer T_M – (5-10)	45 sec.	
Extension	68 ^{o}C	1.0 min./kb of target	
Final extension	68 ^{o}C	3 min.	1

Table 7. Oligonucleotide sequence of primers used to amplify glycosyltransferase probes and *plcGT* from *S. polyformus*. Restriction sites are represented by italicized sequences.

Primer	Oligonucleotide sequence (5′-3′)
PlcGT-Deg-F1	SCACACSACCGAGAGCATCAAG
PlcGT-Deg-R1	AGGCRSGGCGGGCAGATGTC
PlcGT-Deg-F2	CGSTWCGTSCCSTWCAACGG
PlcGT-Deg-R2	GTSCCSSCSCCSSCGTGGTG
PlcGT-Deg-R3	SSYSARBGYMGTKCCSSWGCCSCC
plcGT-F	CTA*TCTAGA*GGAGGAGCCCATGTGAAAGCCCTGTTCTAC
plcGT-R	TAG*GAATTCAAGCTT*TCACTGCTTCACCTTTTC

S. polyformus cosmid library construction and screening

The genomic DNA of *S. polyformus* was isolated and a genomic library was constructed according to standard protocols.[122] Specifically, the isolated genomic DNA was partially digested with *Bfu*CI (roughly 30-40 kb fragments), purified using phenol-chloroform and dephosphorylated with CIP (**Figure 42**, step 1). In parallel, the *E. coli* – *Streptomyces* shuttle vector pOJ446 was digested with *Hpa*I/*Bam*HI and purified using phenol-chloroform (**Figure 42**, step 2). *Bfu*CI and *Hpa*I create compatible overhangs which are utilized in the ligation step between the partially digested *S. polyformus* genomic DNA and fragmented pOJ446 creating a linear vector which is packaged and transduced into *E. coli* XL1-Blue MRF cells using a Gigapack III packaging extract (Stratagene) (**Figure 42**, steps 3-4). The resulting library contained over 8,000 colonies and was calculated to have >99.9% chance to collectively cover the entire genome of *S. polyformus*.

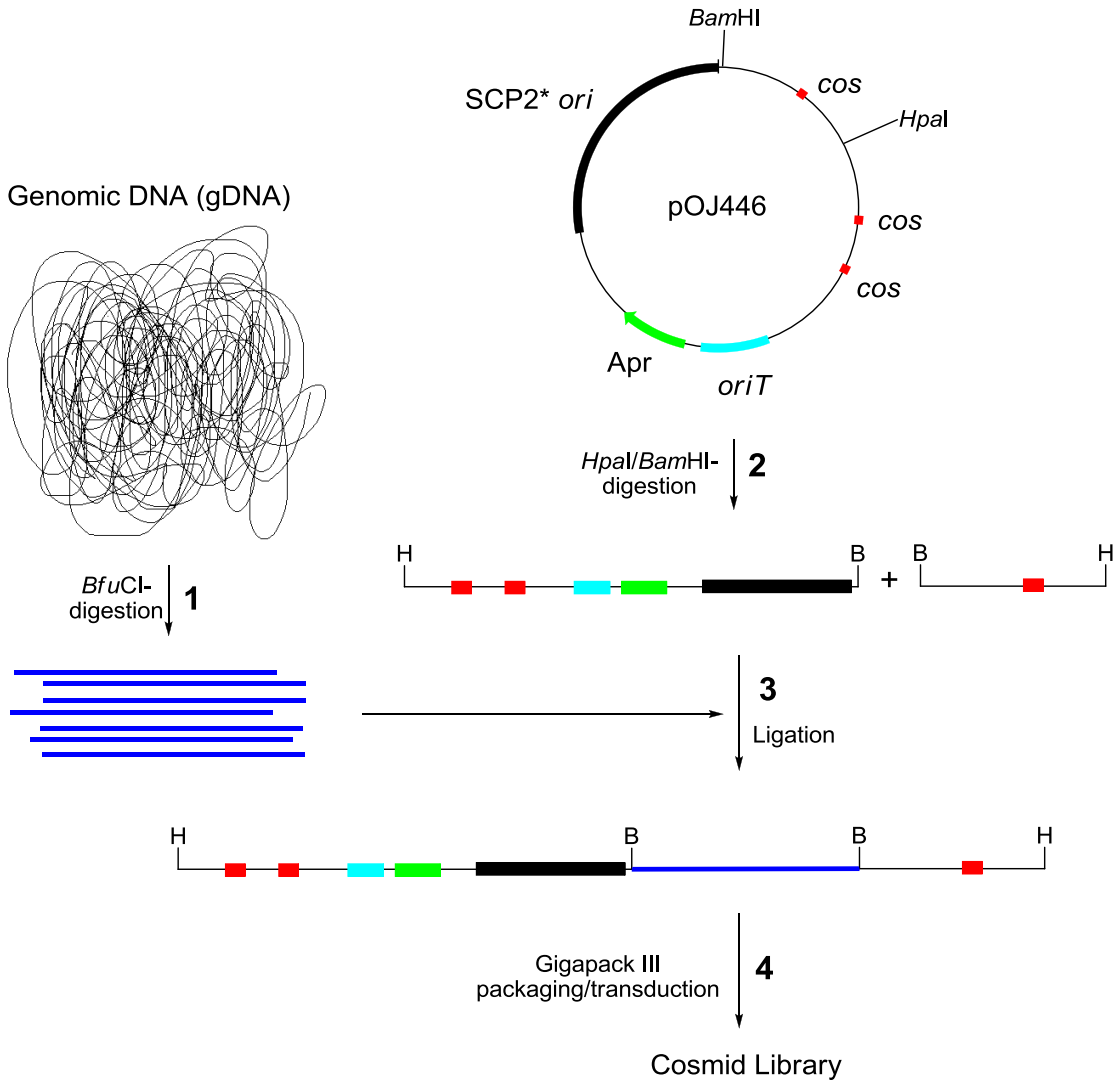


Figure 42. Schematic representation of the method used to produce the S. polyformus genomic library. 1 = BfuCI digestion of genomic DNA; 2 = HpaI/BamHI digestion of pOJ446; 3 = ligation of resulting fragments from steps 1 and 2; 4 = Gigapack III packaging and transduction into E. coli XL1-Blue MRF.

The degenerate primers PlcGT-Deg-F1and PlcGT-Deg-R1 were designed to amplify a small 400 nucleotide region from the *N*-terminal domain of gilvocarcin-like glycosyltransferases from the genomic DNA of *S. polyformus* (**Figure 41** and **Table 7**). The amplified DNA fragment was DIG-labeled and subsequently used as a probe to screen the *S. polyformus* genomic library using a DIG DNA labeling kit (Roche). Initial colony hybridization revealed roughly fifty candidate positive colonies (**Figure 36**, A). These colonies and surrounding colonies were individually scratched out and re-

hybridized. This conformational hybridization step identified 29 positively hybridized colonies (**Figure 36**, B). Cosmid DNA was isolated from each of the 29 colonies and digested with *Bam*HI to remove duplicate cosmids (**Figure 43**). The resulting 21 unique cosmids were used as templates for PCR using the gilvocarcin-like glycosyltransferase based degenerate primers. Only 12 of the 21 tested cosmids produced a strong PCR product at the expected ~400 base pair (bp) range (**Figure 44**). These 12 PCR products were gel purified using a QIAquick Gel Extraction Kit (Qiagen) and cloned into pGEM-T Easy Vector (Promega). The resulting clones were sequenced revealing only two clones with glycosyltransferase sequence homology. The sequence of these two PCR products, from cosPlc47 and cosPlc75, were found to be identical. The cosmid cosPlc75 was chosen for primer walking (SeqWright) and resulted in the complete nucleotide sequence of *plcGT* as well as flanking regions.

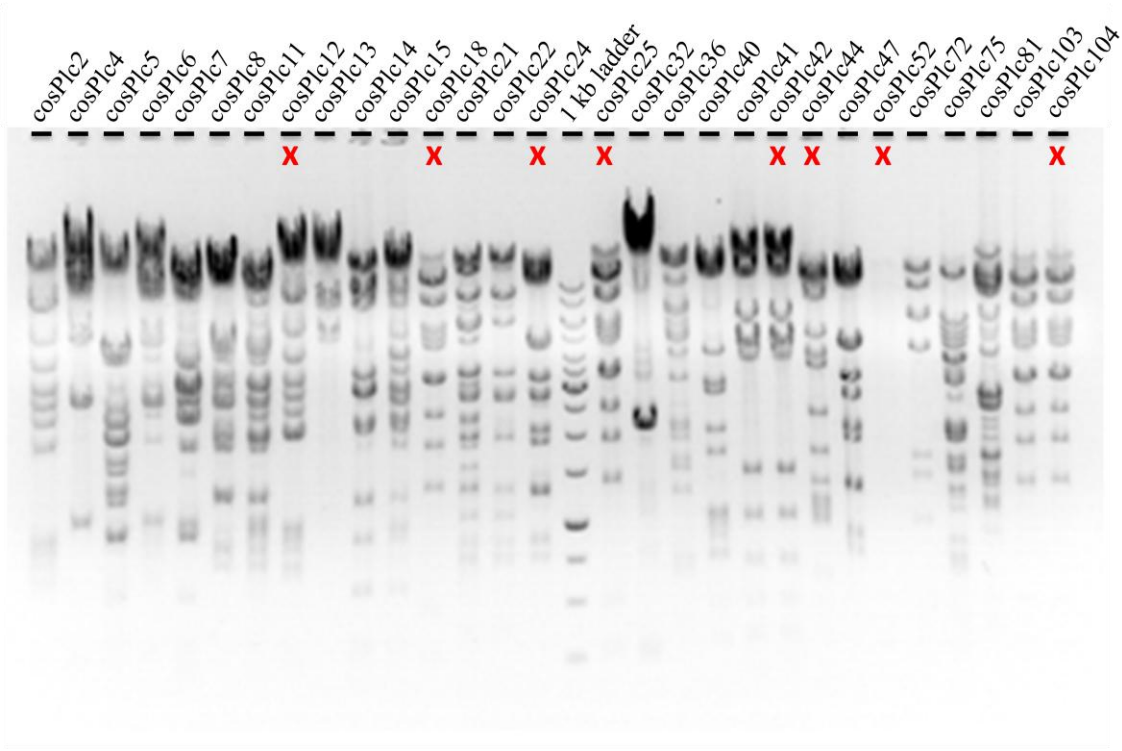


Figure 43. *Bam*HI digestion of 29 candidate polycarcin cosmids revealed 7 duplicates and 1 sample without DNA.

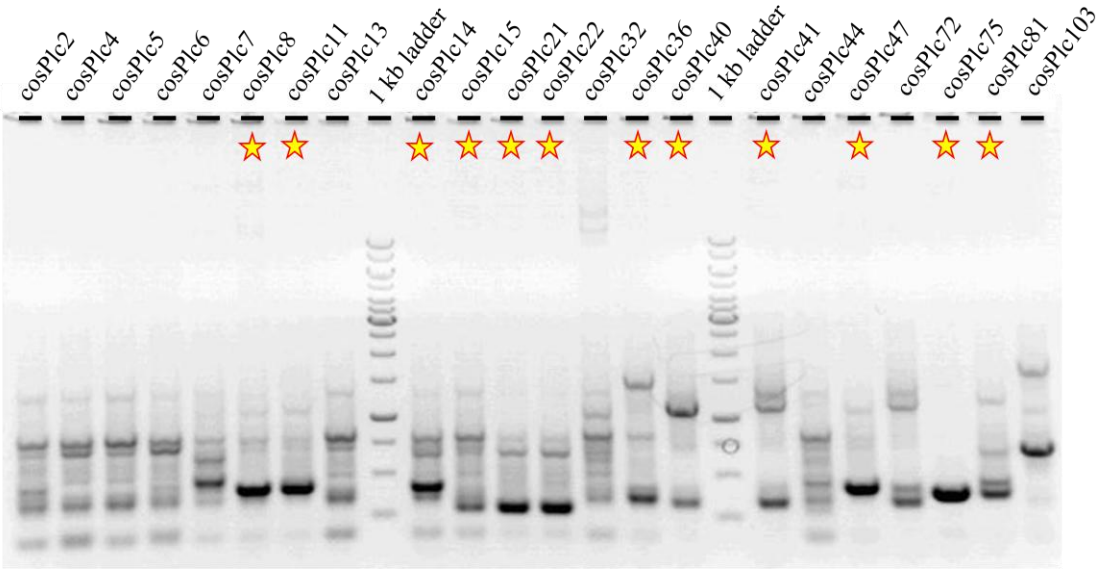


Figure 44. PCR products of the 21 unique polycarcin cosmids using the glycosyltransferase degenerate primers PlcGT-Deg-F1 and PlcGT-Deg-R1.

Plasmid construction

The newly identified glycosyltransferase, *plcGT*, was amplified from *S. polyformus* using Native *Pfu* polymerase and cloned into PCR-Blunt II-TOPO as described in section 2.1: DNA isolation, DNA manipulation and PCR. PCR was used to introduce a unique *Xba*I site upstream of the *plcGT* start codon and *Eco*RI/*Hind*III sites directly downstream of the stop codon. *Xba*I/*Eco*RI double digestion was used to transfer *plcGT* from TOPO to pEM4 producing pPlcGT. *plcGT* along with *ermE**p was removed from pPlcGT using *Hind*III and ligated into the sugar plasmids, pLNBIV and pLNRHO.

***plcGT* sequence**

The complete sequence (5′-3′) obtained from primer walking including *plcV*, *plcGT* and flanking regions is shown below. Start codons and stop codons are designated by **NNN** and **NNN**, respectively.

AGCCGGAAACGTCGTCTCCCTCATCTGGCACATGCACGCCGTACTGCGTCCG
CAGGAGACCCGATCGGAGATACCGGC**ATG**AAGTTCGCCGCCGTCATCGTCGT
CAACTACCAGACTCCCTGGTACCAGCTCTCGCTAGACGAGCGCGTCGCCATC

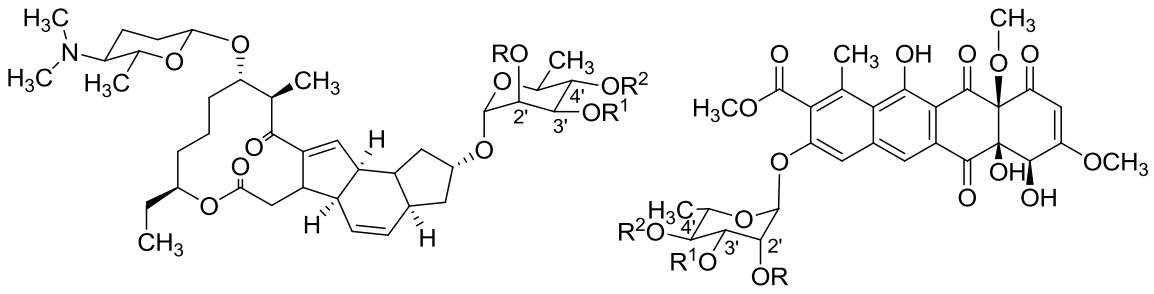
GAGCACCAGCACGTCTTCCCCAGCATCGGCGCCTTCGTGCAGCGGGGCGGCG
AGATCACCCCGGTCCGCGCCATCAAGTACGGCCGTGAGGCGGAGAGCTTCTT
CCTCCTCGGCTTCGACGACATGGAGGACTACCTGGACCTGGTCCGGGAGCTT
CGCTCAAGCCACCTGCTGACCTCGGGTCTCGCCGCCATCGAGTTCGAGACCG
TGGGGCTGAAGGAAGCCTACTTCGCGGAACGCGACAAAGACCGGTCACCGG
plcGT ⟶
CGGTGGCCTC**GTGA**AAGCCCTGTTCTACGCCGCGGGCACCAGCCCGGCCAGC
GCCTTCGCGATCGGCCCGCTCGCCTCCGCCCTCCGGGCGTCCGGGCACGACA
TCCTGGTCGCCTCCTTCGAGGAGATGTCCGGCGCCGTCACCGGCATCGGACT
GCCGTCCATCCCCGTGGCCCGTGGCCACTCGACCGAGAGCATCAAGGCCGCC
GTCGACGGCAGGCCGGCCATCGAGTACCCCCACCGGCCCGAACAAGAGATG
CCCTACCTGGGCCACTGGTTCGGCCGCCAGGGAAGCTATGTCTTCGACGACC
TGGTGGACATCGCCCGGACCTGGGGCGCGGACGTCCTGATCGCGGGCAGCCA
GGGACACGGCGCGGAGATCGCCGCCCGCCTCGTCGGCATCCCCTTCGTCCGC
CAGTCCTGGGACCTGTTCGACATCCACGGCTACGAGGAACACCTCCACGGCG
AACTGGCCGCGCAGTTGGCCCGCATCGGCGCGGACTCCCTGCCCGAACCGTC
GCTGCGCATCGACATCTGCCCTCCCGGACTGACGGACCTCACCGGCGGCACC
TTCATGCGCTGGACCCCGCACAACAAGCAGCGCCAGATCGAGCCGTGGATGC
TGACGGCCCCGGACCGGGGGCGGGTGTGTCTGACGATGGGCAGCTTCCGGTA
CGCCTTCCCGGGCGCGATGGACCGGATCTCGGCGATCGTGGAGCGGCTCCAG
GAACTTGAGACCGAGGTCGTCGTCGCCATCGGCGAGGCCGAAGGACAGCAG
ATCGAGGAGCGGTTCCCCGGGGTCCGGGCGGGCTGGATCCCGCTGGAGGCGA
TCCTCCCCACCTGTGAGGTGATCATCCATCCGGCGGGCGGCCTGACGGCGGT
CAACGCCATCAACACCGCCACACCCCAGCTCATCCTCAACCCCTTCGAGGCC
TTCGAACCCGCCCTGAAACGCCTCACCGACTACGGCTGCGCGCGGACGCTCT
ACCGGGAGGAGGGCACCCCCGCGGCGATAGCGCGGGTCGTCAAGGAGATGC
TCGGGGACTCCTCGTACTGCTCAAGGGCCCGGGATCTGGCGGCGCGGGCCGC
GACCGCGCCCACCGCCGTGGGCATGGTTCCGCTGATCGAAGANCTCGTCGCA
CGGGAGAAGCGGGAAAAGGTGAAGCAG**TGA**CTTCGACGTTGACGATTCCTC
AGACCGACCTCGATCCGTACACCGACGCGTCGATCACCGACCCGTATCCGCT
CTACGGCGCCCTG

Chapter 3: Polycarcin Sugar Residue Modification

In vitro methylation of L-rhamnosyl moiety of polycarcin V

The L-rhamnosyl moiety of polycarcin V (**61**) provides a unique opportunity to employ a combinatorial biosynthetic approach to selectively modify the sugar moiety of a gilvocarcin-like aryl *C*-glycoside. This study is particularly exciting for several reasons; (a) no extensive sugar based SAR studies exist for a single gilvocarcin-like aryl *C*-glycoside; (b) this study may produce several polycarcin V analogues with improved bioactivity or pharmaceutical properties; and (c) this study may provide a unique combinatorial biosynthetic methodology for modifying sugar moieties of several natural products.

Spinosyn A (**26**) and elloramycin A (**117**) are the only examples, to our knowledge, of polyketide derived natural products containing permethylated L-rhamnose moieties (**Figure 45**). L-rhamnosyl-*O*-methylation has been extensively studied in the spinosyn biosynthetic pathway, and has been shown to be extremely important in conferring spinosyn bioactivity.[53] Through *in vitro* experimentation it was found that every possible methylation pattern was achievable through the activity of three pathway specific *O*-methyltransferases, SpnH, SpnI and SpnK, and is the biosynthetic rational behind spinosyn H (**132**), J (**133**), K (**134**), P (**135**), T (**136**) and U (**137**) production in *Saccharopolyspora spinosa* (**Figure 45**).[134] Similarly, the L-rhamnosyl methylation steps of elloramycin biosynthesis have been investigated and have been shown to effect the bioactivity of elloramycin.[54] Unlike spinosyn, the elloramycin L-rhamnosyl-*O*-methyltransferases, ElmMI, ElmMII and ElmMIII, were not as flexible as spinosyn methyltransferases allowing for the production of only two permutations (**138** and **139**) regarding the *O*-methylation pattern of L-rhamnose (**Figure 45**). It is important to note that in both biosynthetic pathways, the modification of L-rhamnose occurred only after glycosylation to their respective aglycones.



Spinosyn A: R=R^1=R^2=CH_3 **26**
Spinosyn H: R^1=R^2=CH_3, R=H **132**
Spinosyn J: R=R^2=CH_3, R^1=H **133**
Spinosyn K: R=R^1=CH_3, R^2=H **134**
Spinosyn P: R=CH_3, R^1=R^2=H **135**
Spinosyn T: R^2=CH_3, R=R^1=H **136**
Spinoysn U: R^1=CH_3, R=R^2=H **137**

Elloramycin A: R=R^1=R^2=CH_3 **117**
Elloramycin C: R=R^2=CH_3, R^1=H **138**
Elloramycin D: R^1=R^2=CH_3, R=H **139**

Figure 45. Spinosyn A and elloramycin A are two examples of natural products containing permethylated L-rhamnose. This figure also shows their naturally produced methylated or partially methylated analogues, with respect to the L-rhamnose moiety.

The substrate flexibility of L-rhamnosyl-*O*-methyltransferases have not been extensively studied, and it is unknown if these methyltransferases are specific to the entire L-rhamnosyl-natural produce or only the L-rhamnose moiety itself. The recent observation that OleY, an *O*-methyltransferase involved in methylating L-olivose (**127**) after attachment to a macrolide core during oleandomycin biosynthesis, could recognize and methylate various deoxysugars attached to a tetracyclic core suggests these deoxysugar *O*-methyltransferases to be quite substrate promiscuous.[49] In this context, we plan to use ElmMI, ElmMII and ElmMIII to create a library of polycarcin V derivatives carrying various L-rhamnose-*O*-methylation patterns, which in turn will be used for sugar based SAR studies.

Experimental design

In **specific aim 2**, the elloramycin *O*-methyltransferases ElmMI, ElmMII and ElmMIII will be cloned and expressed as soluble proteins in *E. coli*. The resulting purified enzymes will be used in an *in vitro* assay with polycarcin V to ascertain their ability to accept the unnatural dibenzochromenone backbone of **61**. Reactions leading to

the production of new polycarcin V analogues will be scaled up and their respective products will be purified and structurally characterized using spectroscopic techniques. The bioactivity of fully characterized polycarcin V analogues will then be evaluated to determine the importance of individual sugar functional groups with respect to the activity of polycarcin V.

Results

In vitro activity of ElmMI, ElmMII, ElmMIII and StfMII

To test the polycarcin methylation ability of these enzymes, *N*-terminal $(His)_6$-tagged proteins were expressed in *E. coli* and purified using immobilized metal ion affinity chromatography (IMAC). Soluble purified proteins were analyzed through sodium dodecylsulfate-polyacrylamide gel electrophoresis (SDS-PAGE) to confirm the correct size of each protein (**Figure 46**, lanes 1-3). The expected size of His-tagged ElmMI, ElmMII and ElmMIII were calculated to be 43.3 kDa, 45.0 kDa and 31.2 kDa, respectively. Additionally, the 2′-*O*-methyltransferase from steffimycin (**118**) biosynthesis, StfMII, was prepared as an *N*-terminal His-tagged protein and purified using IMAC. The expected size of His-tagged StfMII was calculated to be 45.6 kDa which was observed in the SDS-PAGE gel containing purified StfMII (**Figure 46**, lane 4). StfMII was prepared after initial experiments showed ElmMI to be unable to efficiently accommodate polycarcins (discussed below).

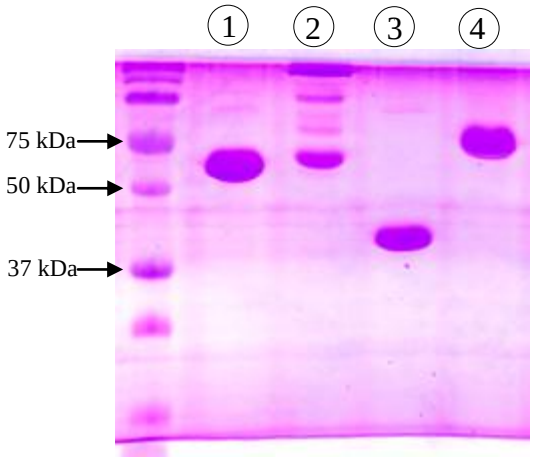


Figure 46. SDS-PAGE analysis of *N*-terminal $(His)_6$-tagged (1) ElmMI, (2) ElmMII, (3) ElmMIII and (4) StfMII.

The initial *in vitro* assay was conducted using a mixture of polycarcin V, M and E purified from *S. polyformus*. The purified mixture is roughly 75% polycarcin V (**61**), 20% polycarcin M (**90**) and 5% polycarcin E (**91**) (**Figure 47, A**). To test the ability of ElmMI to methylate the 2′-OH-position of polycarcin V we incubated ElmMI with polycarcins and S-adenosylmethionine (SAM). This reaction resulted in the formation of a new, yet very small, peak in HPLC (**Figure 47**, B). The new peak had a UV-Vis spectrum typical to that of polycarcins, and eluted (16.8 min) slightly after the control polycarcins (**Figure 47**, A) suggesting an increase in lipophilicity. With these data and the parental functionality of ElmMI, we propose **140** to be the 2′-*O*-methylated analogue of polycarcin V, designated polycarcin A.

Similarly, we tested the ability of ElmMII to methylate the 3′-OH-position of polycarcin V by incubating ElmMII with polycarcins and SAM (**Figure 47**, C). Interestingly, the major new peak, **143**, had an identical retention time (16.8 min) and UV-Vis spectrum to that of polycarcin A (**140**). The minor new peaks, **144** and **145**, were analogous to the elution pattern of polycarcin M and E with respect to polycarcin V. That is to say, the methyl congener elutes before the vinyl congener, while the ethyl congener elutes immediately after the vinyl congener. In this context, we propose **143**, **144** and **145** to be the 3′-*O*-methylated analogues of polycarcin V, M and E, respectively. As we are only interested in the vinyl analogues, we have designated **143** as polycarcin

B. The identical retention time of **140** and **143** can be explained by their structural similarities. **140** and **143** only differ in the placement of a single methyl group at either the 2′ or 3′ position, respectively; and this difference may not be reflected in the HPLC profile of these two compounds.

The last enzyme, ElmMIII, has been proposed as the 4′-OH-methyltransferase in elloramycin biosynthesis; however, it was shown to be unable to methylate the 4′-OH position without the presence of at least a 2′-methoxy or 3′-methoxy group.[135] This substrate specificity was also observed when ElmMIII was incubated with polycarcins and SAM (**Figure 47**, D). The reaction of polycarcin with ElmMIII produced no discernable new compounds, indicating an inability to methylate the naked L-rhamnose moiety of polycarcins.

Lastly, sequential methylations were tested by incubating ElmMI, ElmMII and ElmMIII with polycarcins and SAM (**Figure 47**, E). A strong peak at 16.8 min was proposed to be polycarcin B (**143**) and not polycarcin A (**140**) simply due to the relative inactivity exhibited previously by ElmMI. In addition, a new peak, **146**, was observed at 20.6 min indicating even further methylation by ElmMIII (this was corroborated in further experimentation). **146** is proposed to contain both 3′ and 4′-methoxy groups and is designated as polycarcin C.

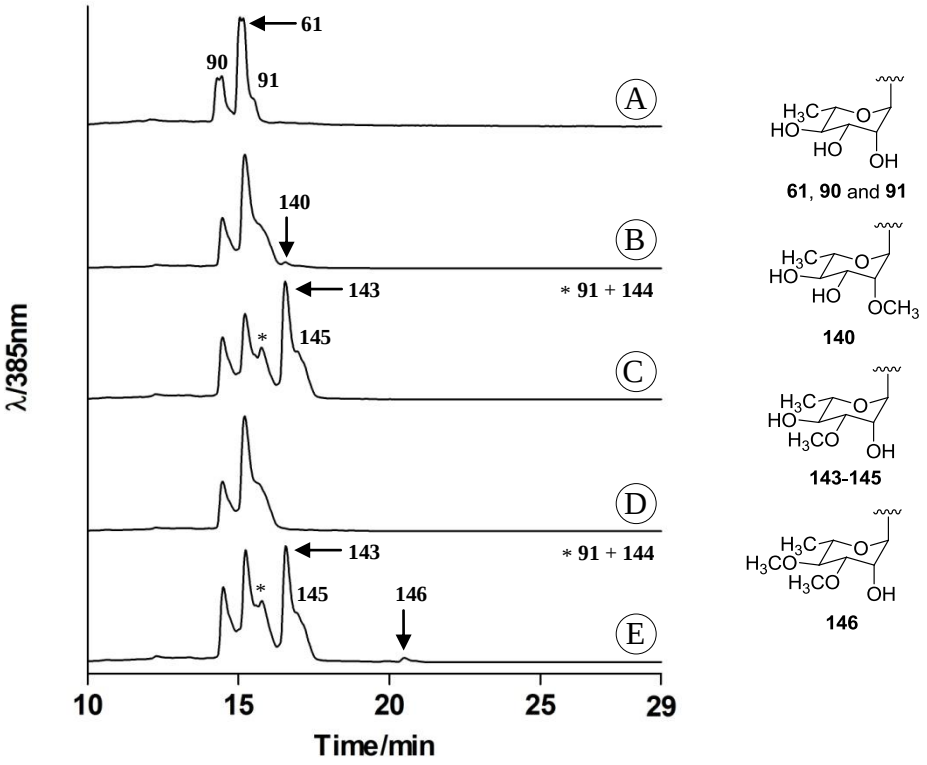


Figure 47. HPLC chromatogram trace of methyltransferase assays using ElmMI, ElmMII and ElmMIII. Reactions contained (A) polycarcins (**61**, **90** and **91**), (B) ElmMI + polycarcins, (C) ElmMII + polycarcins, (D) ElmMIII + polycarcins and (E) ElmMI + ElmMII + ElmMIII + polycarcins. **140**, **143** and **146** are polycarcin V analogues, while **144** and **145** are polycarcin M and E analogues, respectively.

From preliminary *in vitro* assays using ElmMI, ElmMII and ElmMIII it is apparent that only two polycarcin V analogues (**143** and **146**) can be created utilizing ElmMII and ElmMIII. The poor activity of ElmMI makes isolation of **140** virtually impossible. To remedy this problem and to restore the possibility of permethylating the L-rhamnosyl moiety of polycarcin V, StfMII from steffimycin biosynthesis was cloned, expressed and purified. Incubation of StfMII with polycarcins and SAM resulted in the formation of a major peak in HPLC with a retention time of 16.8 min (**Figure 48**, B). With the specific 2′-*O*-methyltransferase parental activity of StfMII we proposed this new peak to be identical to polycarcin A (**140)**. In addition to polycarcin A, two additional peaks **141** and **142** were observed and believed to be the methyl and ethyl congener of polycarcin A, respectively.

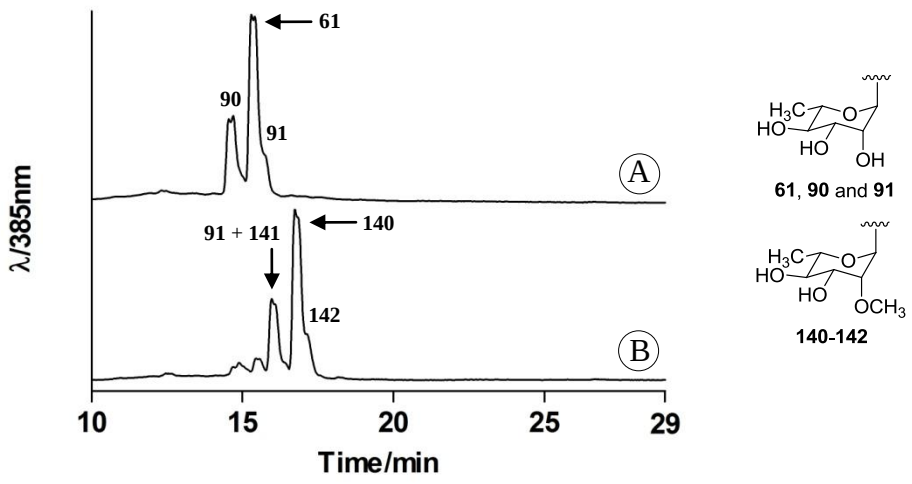


Figure 48. HPLC chromatogram trace of methyltransferase assay using StfMII. The reactions contained (A) polycarcins **61**, **90** and **91**) and (B) StfMII + polycarcins. **140** is a polycarcin V analogue, while **141** and **142** are polycarcin M and polycarcin E analogues.

With successful 2′-*O*-methylation by StfMII, it was possible to combine StfMII, ElmMII and ElmMIII in different combinations to produce various methylated derivatives of polycarcin V. Incubation of StfMII, ElmMII, polycarcins and SAM resulted in the formation of **140-142**, as expected from the activity of StfMII, and a new peak with a retention time of 18.7 min in HPLC. The new peak, **147**, had a typical UV-Vis spectrum to that of polycarcins and its increased retention time indicated increased lipophilicity. With these results, we propose **147** to be the 2′,3′-dimethoxy analogue of polycarcin V, designated polycarcin D. Interestingly, **146** and **147** have different retention times despite their only difference being the position of a second methoxy group at either the 4′-OH or 2′-OH, respectively.

Co-incubation of StfMII and ElmMIII only produced the products of StfMII activity, namely **140-142** (**Figure 49**, C). Unlike in elloramycin biosynthesis where ElmMIII can act on an elloramycin substrate containing only a 2′-methoxy group, ElmMIII cannot methylate the analogous polycarcin V substrate, polycarcin A. We know ElmMIII is active by its ability to methylate polycarcin B, as observed in **Figure 49** trace D, producing **146**.

Finally, we attempted to permethylate polycarcin V by incubating StfMII, ElmMII, ElmMIII, polycarcins and SAM (**Figure 49**, E). The HPLC analysis of this

reaction revealed a major peak at 16.8 min indicating either **140** or **143**. The accumulation of both **147** and to **146** indicates that both StfMII and ElmMII are competing for **61** and producing **140** and **143** (**Figure 50**). Most importantly, however, is the formation of a new peak, **148**, with a retention time of 22.8 min. This new peak had a typical UV-Vis spectrum to that of polycarcins and its increased retention time suggests a further methylated product. With previous *in vitro* results and the fact that **147** is present in the reaction mixture for the first time with ElmMIII, we propose **148** to be the permethylated L-rhamnose analogue of polycarcin V, designated polycarcin F.

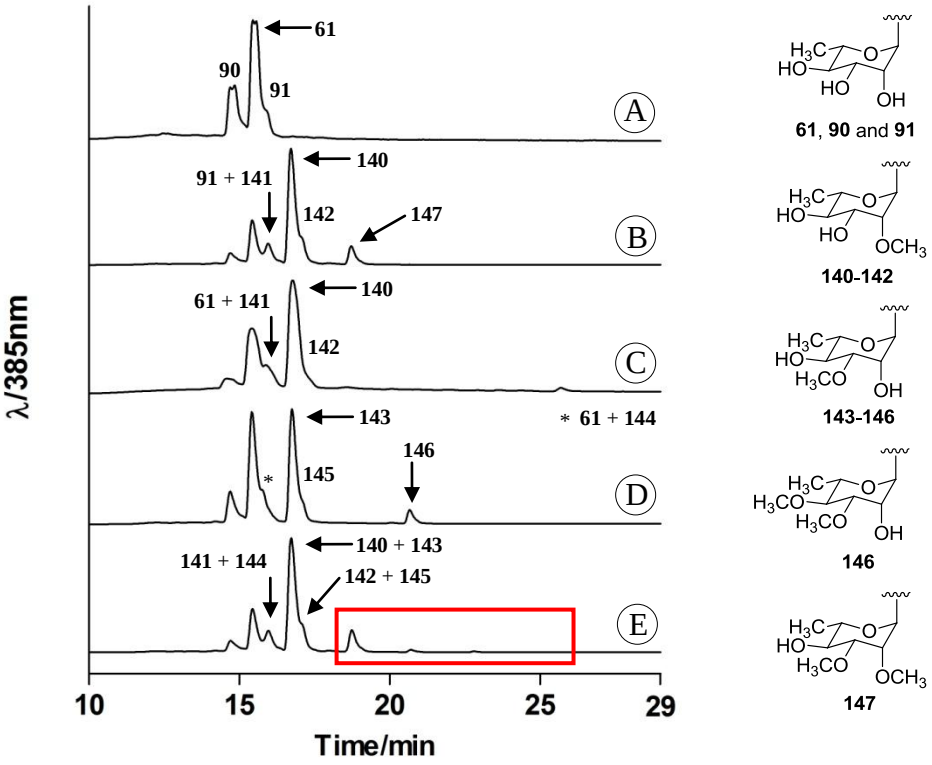


Figure 49. HPLC chromatogram trace of methyltransferase assays using StfMII, ElmMII and ElmMIII. Reactions contained (A) polycarcins (**61**, **90** and **91**), (B) StfMII + ElmMII + polycarcins, (C) StfMII + ElmMIII + polycarcins, (D) ElmMII + ElmMIII + polycarcins and (E) StfMII + ElmMII + ElmMIII + polycarcins. **140**, **143**, **146** and **147** are polycarcin V analogues, while **141** and **144** are polycarcin M analogues. The remaining compounds, **142** and **145**, are analogues of polycarcin E.

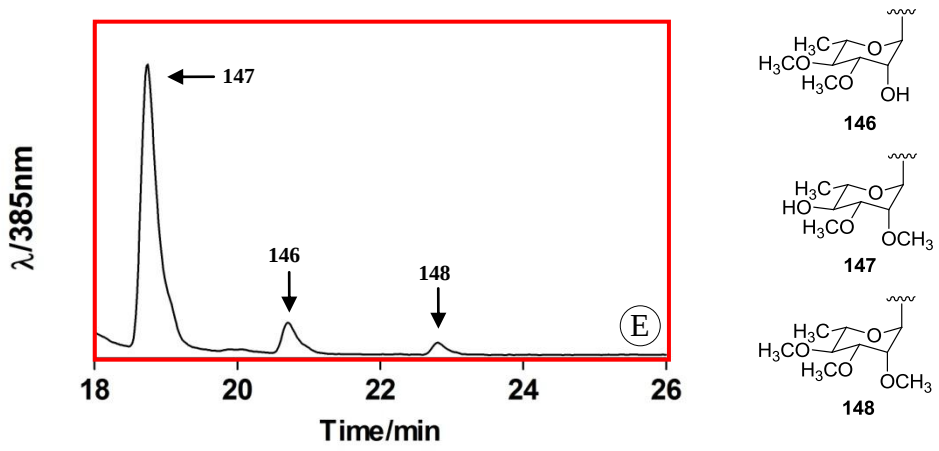


Figure 50. Expanded region of trace E from **Figure 49** showing the polycarcin V analogues **146**, **147** and **148**.

Preliminary structural characterization of polycarcin analogues

The reactions discussed above were conducted on a very small scale and mass data were not collected. To confirm the proposed structure of polycarcin A (**140**), the *in vitro* reaction was scaled up and the peak corresponding to polycarcin A was purified using analytical HPLC. To date, only around 2 mg of polycarcin A has been purified. This has allowed for partial structural characterization of polycarcin A using ^{1}H-NMR. For comparison, the ^{1}H-NMR of polycarcin V was taken to show the signals corresponding to the unmodified 2′-H, 3′-H, 4′-H, 2′-OH, 3′-OH and 4′-OH positions (**Figure 51**). When comparing the ^{1}H-NMR spectrums of polycarcin A and polycarcin V it was observed that in polycarcin A the 2′-OH signal had been lost and a new signal had appeared directly downfield of the 3′-H signal (**Figure 52**). This downfield shift is characteristic of a proton connected to a carbon with a methoxy group versus a hydroxyl group, and is what we would expect if polycarcin A contained a 2′-OCH_3. Further confirmation would require 2D NMR experiments, such as Heteronuclear Multiple Bond Coherence (HMBC), to connect the methoxy protons to the 2′ position of polycarcin A. Unfortunately, the amount of purified polycarcin A does not allow for such 2D experiments to be conducted. These results, however, give convincing evidence that polycarcin A is in fact the 2′-*O*-methoxy analogue of polycarcin V. It should be noted

the additional methoxy group of polycarcin A was not observed in the ^{1}H-NMR spectrum and is believed to be hidden under the water signal.

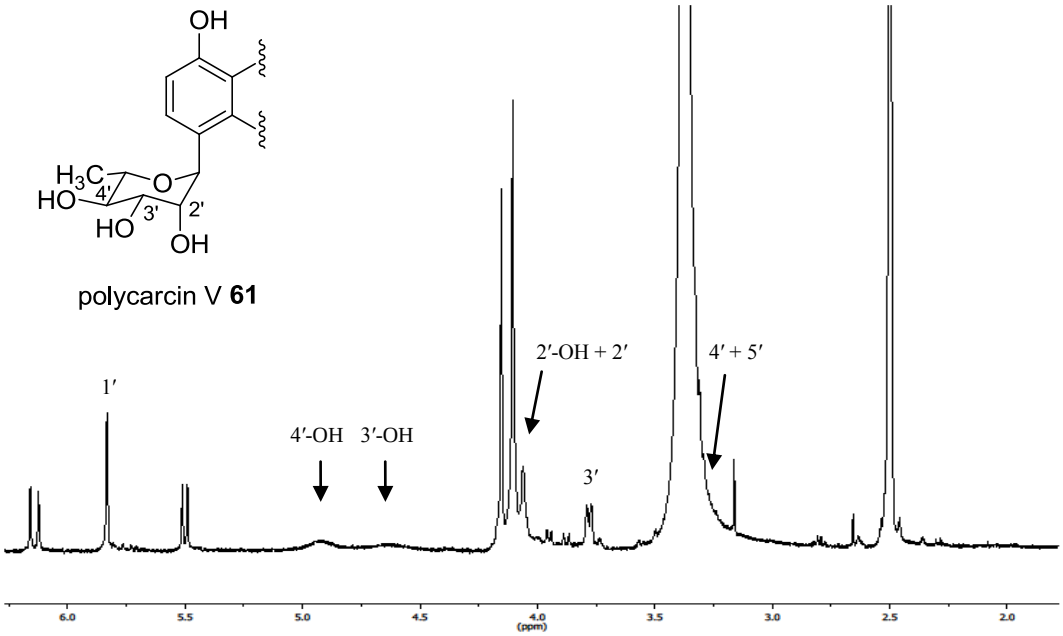


Figure 51. Upfield ^{1}H NMR spectrum for polycarcin V (**61**) highlighting the L-rhamnose signals (500 MHz, DMSO-d_6).

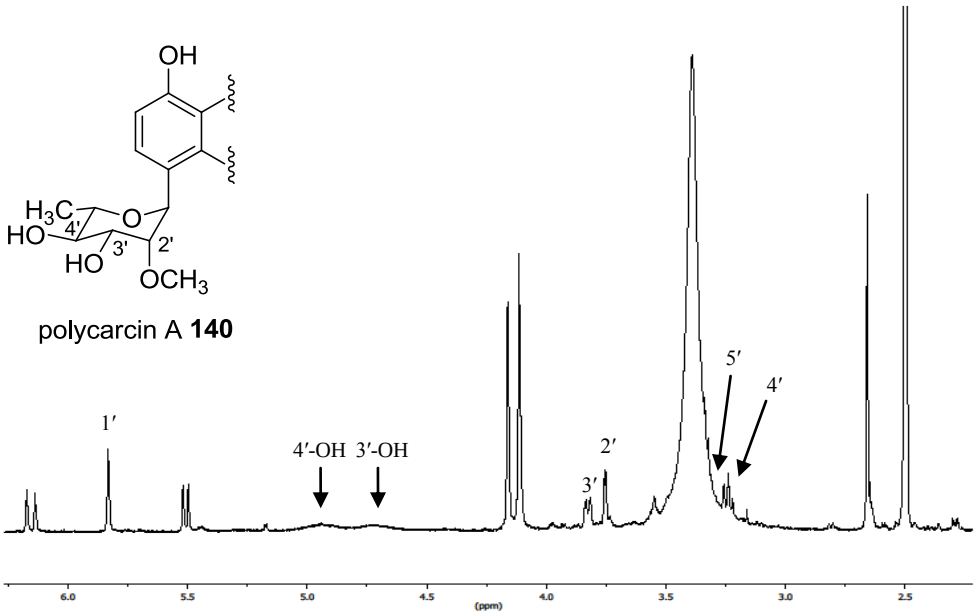


Figure 52. Upfield ^{1}H NMR spectrum for polycarcin A (**140**) highlighting the L-rhamnose signals (500 MHz, DMSO-d_6).

Discussion

The ability to produce methylated analogues of polycarcin V is extremely useful in understanding the unique role sugar moieties play in the bioactivity of gilvocarcin-like aryl *C*-glycosides. Specific modification of individual sugar functional groups would allow for an opportunity to analyze the importance of these individual functional groups toward the bioactivity of the given compound. This information could then be used in the rational design of gilvocarcin-like aryl *C*-glycosides with improved activity.

The results presented here exemplify a method in which L-rhamnosyl methyltransferases from various biosynthetic pathways could be used to specifically modify individual functional groups of polycarcin V. Specifically, StfMII, ElmMII and ElmMIII were used to methylate the 2′-OH, 3′-OH and 4′-OH of polycarcin V, respectively. Used individually or in combination, these enzymes successfully catalyzed the formation of five new polycarcin V analogues proposed to be polycarcin A (**140**), B (**143**), C (**146**), D (**147**) and F (**148**) (**Figure 53**).

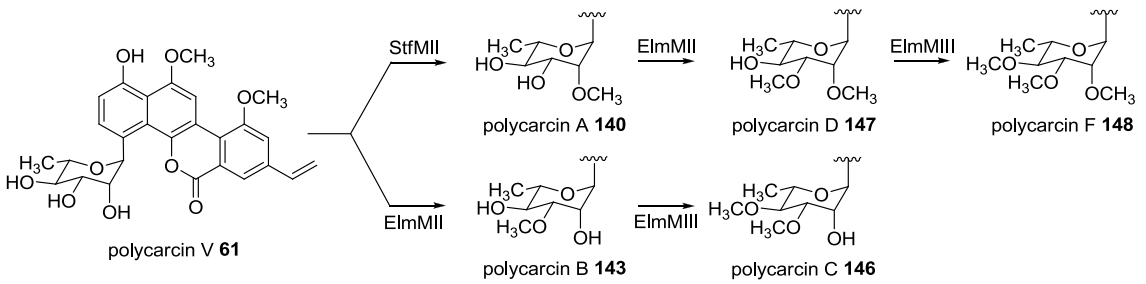


Figure 53. Engineered biosynthetic pathway leading to the production of five polycarcin V analogues.

To isolate, purify and characterize the polycarcin V analogues, the *in vitro* reactions were scaled up by increasing the overall volume of each reaction. Unfortunately, increasing the reaction volume resulted in extensive protein precipitation and therefore low yields were observed. Instead of increasing the reaction volume, scaling up consisted of increasing the number of small scale reactions, and combining them after the reaction was completed. This process is time consuming and still yields low quantities of compound. Furthermore, methylation reactions conducted *in vitro* are notorious for not going to completion. This is attributed to S-adenosylhomocysteine (SAH), the demethylated product of SAM, produced as a byproduct during methylation reactions and a known inhibitor of methyltransferases.[136-138] Despite these factors, roughly 2 mg of polycarcin A (**140**) was isolated and purified. Initial structural characterization of **140** through ^{1}H-NMR implies the presence of a 2′-methoxy group, however, this can not be unequivocally assigned without further 2D NMR experiments. ^{13}C-NMR was attempted but there was not enough **140** to obtain clear carbon signals.

Together, these results illustrate the power of L-rhamnosyl-*O*-methyltransferases in producing specific derivatives of natural products. This method could be used to create methylated derivatives of several natural products containing L-rhamnose including steffimycin (**118**) and aranciamycin (**119**). Additionally, similar studies could be conducted on natural products that contain sugars that slightly differ from L-rhamnose but retain the stereochemistry of one or more of the 2′-OH, 3′-OH or 4′-OH positions such as L-mycarose (**123**), L-digitoxose (**124**) and L-olivose (**127**). Furthermore, these studies could be conducted *in vivo* making the production of new analogues much easier than

using an *in vitro* approach. The polycarcin V producer, *S. polyformus*, is recalcitrant to DNA uptake making an *in vitro* approach impossible; however, several natural product producing strains can be manipulated or the natural product can be heterologously expressed. This would allow for *in vivo* expression of the exogenous methyltransferases directly in the producing strain or heterologous host, thereby removing the need for an *in vitro* assay.

Future Research

The goal of this study was to be able to conduct SAR experiments on the produced polycarcin V analogues. To date, the analogues have not been isolated in quantities to afford such studies. This will require continued *in vitro* reactions using the methyltransferases discussed herein. Several important factors have been identified from the small scale experiments that need to be considered to successfully produce these polycarcin V analogues *in vitro*. First, the small scale reactions consisted of roughly 75% polycarcin V, 20% polycarcin M and 5% polycarcin E. It was shown that the methyltransferases do not discriminate between the different polycarcin congeners allowing multiple substrates to be methylated. This wastes valuable SAM co-factor and complicates later purification processes. Since we are not interested in the M or E congeners, it would be beneficial for large scale reactions to contain only polycarcin V. Secondly, it was found that StfMII and ElmMII could utilize polycarcin V directly to produce **140** and **143**, respectively. For this reason, the production of **147** should not be catalyzed by a one-pot enzymatic synthesis containing StfMII and ElmMII. This reaction would allow for part of the starting material to be transformed into **143** and therefore will reduce the overall yield of **147**. Lastly, the ability to remove SAH from the reaction mixture would greatly increase the overall yield of methylated products. An *S*-adenosylhomocysteine nucleosidase has recently been identified from *E. coli* that converts SAH to *S*-ribosylhomocysteine and adenine, thereby eliminating its ability to inhibit the methyltransferase reaction.[139-140] An *S*-adenosylhomocysteine nucleosidase could be used in conjunction with the methyltransferases used in this study to prevent SAH inhibition and allow for higher yields of methylated products.

Only two L-rhamnosyl-*O*-methylated analogues of polycarcin V were not produced during this study. These two analogues would consist of a single methylation at the 4′-OH position and two methylations at the 2′-OH and 4′-OH positions, respectively. The inability to create these analogues can be attributed to the natural programming of ElmMIII, as discussed above. An analogous enzyme from spinosyn biosynthesis, SpnH, has been shown to be flexible enough to produce both of the aforementioned methylation patterns in spinosyn J (**133**) and T (**136**) biosynthesis.[141] It would be fascinating to see if SpnH could also utilize polycarcin V, which contains a drastically altered backbone compared to spinosyn. This experiment in combination with the previously described study could allow for the complete derivatization of the polycarcin V sugar moiety.

Materials and Methods

Bacterial strains, culture conditions and plasmids

All transformation and culturing conditions were carried out as described in section 2.1: Bacterial strains, culture conditions and plasmids. A comprehensive list of strains and plasmids used in this study can be found in **Table 8**.

Table 8. Strains and plasmids used in the expression of methyltransferases.

Strain/Plasmid	Characteristics and relevance	References
E. coli XL1-Blue-MRF	Host for routine cloning	Stratagene
E. coli BL21 (DE3)	Host for protein expression	Invitrogen
S. polyformus	Produces polycarcin V	Li, Y. et al.[61]
cos16F4	PCR template for ElmMI, ElmMII and ElmMIII	Decker, H. et al[142]
stfB3	PCR template for StfMII	Gullon, S. et al[143]
PCR-Blunt II-TOPO	PCR fragment cloning vector	Invitrogen
pET28a(+)	Expression vector	Novagen
pElmMI	*elmMI* cloned into pET28a(+)	This study
pElmMII	*elmMII* cloned into pET28a(+)	This study
pElmMIII	*elmMIII* cloned into pET28a(+)	This study
pStfMII	*stfMII* cloned into pET28a(+)	This study

DNA isolation, DNA manipulation and PCR

Plasmid DNA isolations, manipulations and PCR reactions were carried out as described in section 2.1: DNA isolation, DNA manipulation and PCR. Please find **Table 9** for a complete list of primers used in this study.

Table 9. Oligonucleotide sequence of primers used to amplify *elmMI*, *elmMII*, *elmMIII* and *stfMII*. Restriction sites are represented by italicized sequences.

Primer	Oligonucleotide sequence (5′-3′)
ElmMI-F	ATCG*CATATG*GACTCCCCACAGGTG
ElmMI-R	CGAT*GAATTC*AGGTCGCTTCCCGCAG
ElmMII-F	ATCG*CATATG*ACCACCCCTTCACCC
ElmMII-R	CGAT*GAATTC*ATACGTAGTCGATCTC
ElmMIII-F	ATCG*CATATG*ACGGAATACGCCCGC
ElmMIII-R	CGAT*GAATTC*AGACGGTGGCCGGCTG
StfMII-F	GGA*CATATG*TCCGCTGAGAGTGAC
StfMII-R	ATT*GAATTC*TCACTCCGCCGCCGG

Protein expression constructs

The primer pairs ElmMI-F and R, ElmMII-F and R, and ElmMIII-F and R were used to amplify *elmMI*, *elmMII* and *elmMIII* from cos16F4, respectively. Likewise, the primer pair StfMII-F and R was used to amplify *stfMII* from stfB3.[143] The *pfu* amplified PCR products were cloned into PCR-Blunt II-TOPO vector as described previously. All genes were removed from TOPO and ligated into pET28a(+) as *Nde*I/*Eco*RI fragments generating pElmMI, pElmMII, pElmMIII and pStfMII. Cloning of the genes in pET28a(+) results in a *N*-terminal $(His)_6$-tagged protein, which can be purified using IMAC.

Expression and purification of proteins

The expression constructs for ElmMI, ElmMII, ElmMIII and StfMII were transformed into *E. coli* BL21 (DE3) according to standard protocols.[125] Single colonies from each transformation was inoculated into 10 mL of LB containing appropriate antibiotics and grown for 5 hours at 37 ^{o}C to prepare seed cultures. One liter cultures (1 L Erlenmeyer baffled flasks) were then inoculated with each seed culture and grown with the appropriate antibiotics at 37 ^{o}C with reciprocal shaking at 250 rpm. When the cultures reached an OD_{600} of ~0.5, isopropyl β-D-1-thiogalactopyranoside (IPTG) was added to a

final concentration of 50 µM and allowed to continue shaking at 18 ^{o}C overnight, or 16 hours. After 16 hours, the pellet was collected by centrifugation (4000 x g, 15 min) and washed twice with 25 mL lysis buffer (50 mM KH_2PO_4, 300 mM NaCl, 10 mM imidazole, pH 7.6). The washed pellet was stored at -80 ^{o}C overnight and then used when needed. To recover soluble proteins, the frozen pellet was lysed with a French Press (Thermo Electron Corporation), and crude protein was recovered through centrifugation (18000 x g, 30 min). The cell free lysate was then passed through a 15 mL column containing 2 mL of Talon metal affinity resin (Clontech) pre-washed with 10 mL of lysis buffer. The flow through was discarded and the column was washed again with 10 mL of lysis buffer. After washing the column, 3 mL of elution buffer (50 mM KH_2PO_4, 300 mM NaCl, 200 mM imidazole, pH 7.6) was added to the column to elute the soluble protein. The elution was collected in an Amicon Ultra centrifugal filter (Millipore) and concentrated by centrifugation (12000 x g, 20-45 min). The concentrated soluble protein was then quantified using the Bradford protein assay method and the purity of each enzyme was visualized using SDS-PAGE analysis (**Figure 46**).[144]

Methylation reaction conditions

The methylation reaction assays were carried out in microcentrifuge tubes (USA Scientific) using the following conditions: 50 mM NaH_2PO_4 (pH 7.4), 20 mM $MgCl_2$, 2 mM SAM, 200 µM polycarcin V/M/E mixture, 20 µM methyltransferase/s in a final volume of 400 µL. Once the reaction mixture was complete, the microcentrifuge tubes were removed from light and allowed to incubate at 28 ^{o}C for 3 hours. After 3 hours, 400 µL of ethyl acetate was used to extract the polycarcin-like compounds from the reaction mixture. The organic phase of the extraction was removed and then evaporated using a SpeedVac (Savant). The dried extract was reconstituted using 50 µL of methanol and this solution was analyzed using an HPLC (Waters Alliance 2695) and photodiode array detector (Waters 2996) with the same column and linear gradient as described in section 2.1: Fermentation and metabolite screening.

The normally expensive co-factor SAM was purchased inexpensively as SAM-e (Nature Made), an over-the-counter (OTC) dietary supplement in 200 mg tablets. These tablets were divided into two equal portions and crushed using a mortar and pestle. The

crushed half was then added to a 15 mL screw top vial (USA Scientific) containing 750 µL of dH_2O, and the mixture was thoroughly mixed using a vortex (Fisher) for 1 min. The resulting mixture was filtered using a 0.45 µm syringe-driven filter (Millipore) and collected into a 1.5 mL microcentrifuge tube (USA Scientific). The filtrate SAM concentration was quantified using the SAM extinction coefficient ($15400 cm^{-1}M^{-1}$ at 260 nm). The concentration of SAM prepared as described had a concentration typically ~ 200 mM. These solutions were immediately used and/or stored for no more than 48 hours. This process was developed based on personal communications with Dr. S. Van Lanen.

Chapter 4: Starter Unit Incorporation during Gilvocarcin V Biosynthesis

4.1 *In vivo* characterization of GilP and GilQ

Initial bioinformatic analysis of the gilvocarcin gene cluster revealed the presence of a typical type II polyketide synthase consisting of two ketosynthase subunits ($KS_{\alpha/\beta}$) and an acyl carrier protein (ACP).[91] Interestingly, the *gil* cluster also contained two genes, *gilP* and *gilQ*, with high sequence similarities to malonyl CoA:ACP transacylases (MCATs) and acyltransferases, respectively. GilP shows 47/33% amino acid similarity/identity to FabD, a characterized *E. coli* MCAT. GilQ, however, shows sequence identity to several acyltransferases found in type II PKS systems with non-acetate starter unit incorporation (discussed below). The inclusion of an MCAT homologue in the *gil* cluster is unusual as it is often found that type II polyketide synthase clusters do not posses a pathway specific MCAT, and instead recruit the endogenous fatty acid synthase (FAS) MCAT.[92-93] In the actinorhodin and tetracenomycin pathways, this borrowed MCAT has been shown to be responsible for transferring both starter and extender units to the type II minimal PKS as acetate and malonate, respectively.[20-21]

Unlike actinorhodin and tetracenomycin, gilvocarcins utilize two distinct starter units. Early labeling studies revealed these starter units to be derived from acetate and propionate which condense with 9 malonate extender units to produce 20- and 21-carbon decaketides, respectively (**Figure 54**).[87-89] The formation of 20- and 21-carbon decaketides is followed by subsequent intramolecular aldol condensations and several complex post-PKS modifications to produce **50** and **51** (see Biosynthetic Highlights of Gilvocarcin V). At a yet unknown step during the biosynthesis of **51**, oxidation of its ethyl side chain by GilOIII produces the vinyl functional group of **49**.[42]

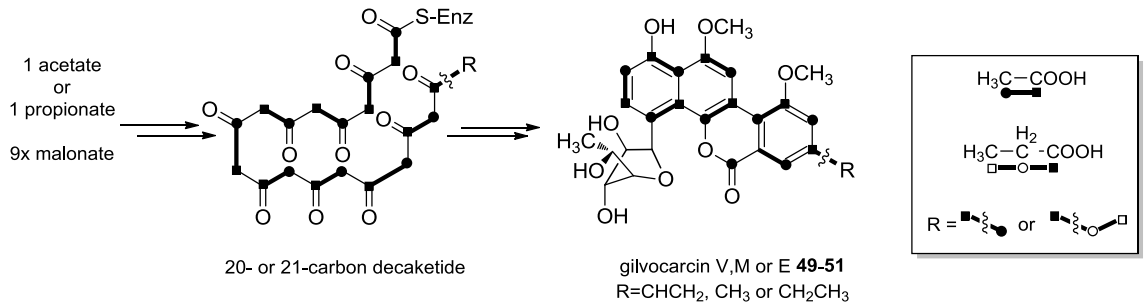


Figure 54. Labeling pattern of the 20- and 21- carbon decaketide formed by the minimal gilvocarcin PKS.

The use of propionate or any starter unit other than acetate, in a type II PKS is relatively uncommon and has been observed in only a handful of natural products belonging to this class of secondary metabolites. Most notably, daunorubicin (**149**), doxorubicin (**150**), hedamycin (**37**), R1128a-d (**151-154**) and fredericamycin A (**155**) are type II PKS derived natural products with unique starter unit incorporation (**Figure 55**). The daunorubicin and doxorubicin pathways, which share early PKS enzymes, were the first to be investigated in regard to starter unit specificity.[145-151] The daunorubicin cluster contains a unique β-ketoacyl:ACP synthase III (KS III) homologue (DpsC) usually found as a component of a type II FAS, and a single acyltransferase homologue (DpsD). A KS III protein is typically involved in the initiation of fatty acid biosynthesis, and the characterized KS III homologue, FabH, from *E. coli* K-12 has been shown to be involved in transacylating acetyl-CoA and subsequently condensing with malonyl-ACP to produce acetoacetyl-ACP.[152] Further investigation of DpsC and DpsD determined the KS III homologue, and not the acyltransferase homologue, to be responsible for conferring starter unit specificity, and therefore propionyl-CoA incorporation.[149,151] Similarly, both KS III and acyltransferase homologues have been found in the gene clusters of several type II PKS derived natural products utilizing unique starter units including hedamycin (2,4-hexadienoate), R1128A-D (short-chain fatty acids), fredericamycin (2,4-hexadienoate), frenolicin (butyrate), and aclacinomycin (propionate).[153-157] In each case, except hedamycin, the KS III homologue has been shown, or is believed, to be responsible for the incorporation of the unique starter unit in their respective pathways.[153-154,158-160] In this context, it is interesting that the *gil* cluster lacks a KS III

homologue and instead carries an acyltransferase that has been shown to be non-essential in similar pathways in regard to starter unit specificity. This indicates that the minimal gilvocarcin PKS likely incorporates propionate utilizing a protein with unique function not seen in similar pathways. The only candidate genes present in the *gil* cluster that may play a role in starter unit specificity are *gilP* and *gilQ*. The strong sequence similarity of GilP to FabD suggests GilP functions as a typical MCAT during the biosynthesis of gilvocarcin V, and does not determine starter unit selection. The acyltransferase, GilQ, is therefore the most likely candidate for conferring starter unit specificity based on bioinformatical analysis alone.

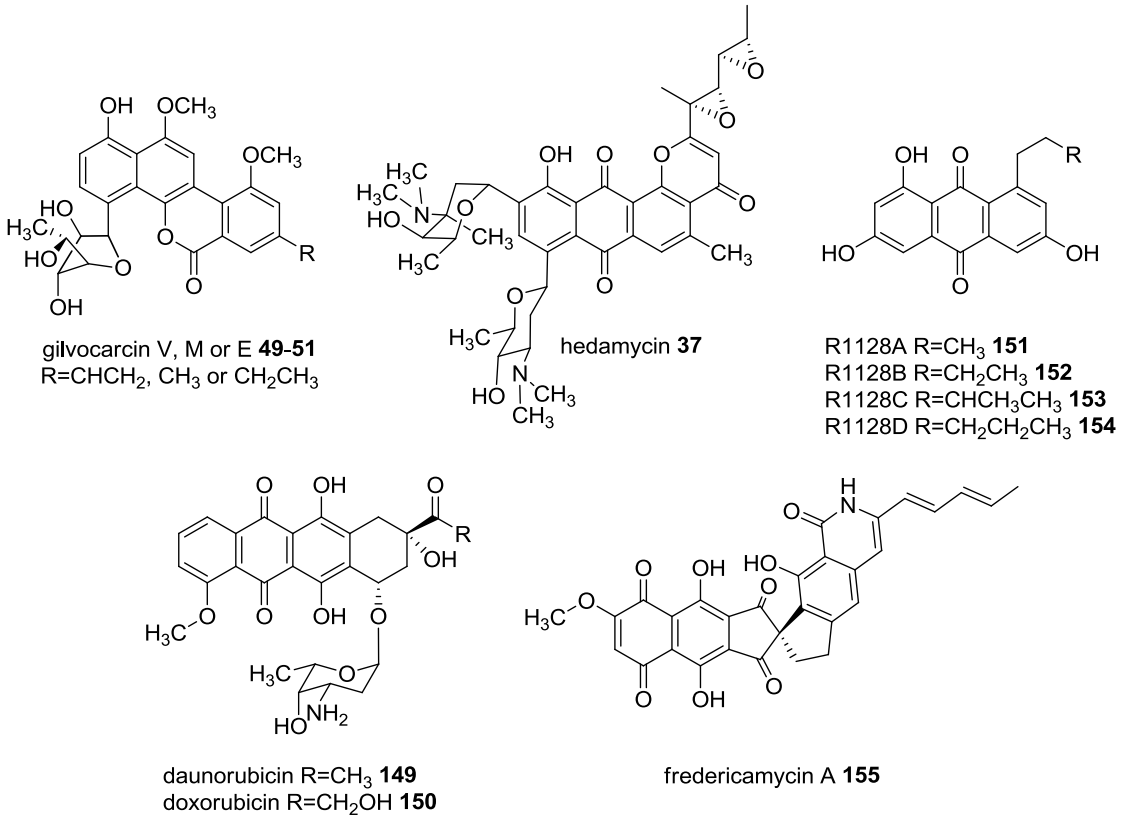


Figure 55. Examples of type II polyketide derived natural products with non-acetate starter unit incorporation.

The initiation of the gilvocarcin biosynthetic pathway with propionate is extremely important. The choice of starter unit directly influences the type of gilvocarcin analogue produced by the pathway, because the starter unit becomes the C-8 side chain in the final gilvocarcin compound. If acetate is incorporated as the starter unit the final

gilvocarcin will contain a single carbon C-8 side chain producing gilvocarcin M (**50**). However, if propionate is incorporated, the final gilvocarcin compound will contain a two carbon C-8 side chain producing gilvocarcin E (**51**). At an undetermined step during the biosynthesis of **51**, GilOIII oxidizes the two carbon C-8 side chain to give the vinyl group found in **49** (as discussed above). As the vinyl group is essential for the photo[2+2]cycloaddition reaction with DNA thymine base, gilvocarcin V is the most bioactive compound compared to the methyl and ethyl congeners. For this reason gilvocarcin M and gilvocarcin E are unwanted side products of gilvocarcin V biosynthesis. In order to engineer the gilvocarcin biosynthetic pathway towards only gilvocarcin V production, a deeper understanding of starter unit specificity must be obtained.

Experimental design

In **specific aim 3a**, the putative MCAT, *gilP*, and acyltransferase, *gilQ*, will be inactivated from cosG9B3 using PCR-targeting REDIRECT technology. The resulting mutants will be heterologously expressed in *S. lividans* TK24 and their metabolites will be analyzed using HPLC-MS. In addition, *gilQ* and *gilOIII* will be overexpressed in *S. lividans* TK24/cosG9B3 to evaluate the effect of increased intracellular GilQ and GilOIII concentrations on gilvocarcin V production.

Results

Production profile of engineered strains

Traditional *in vivo* methods for investigating gilvocarcin biosynthesis are severely hindered due to the inability to introduce genetic information into wild type *S. griseoflavus* Gö3592. The successful isolation of cosG9B3, containing all the genes required for gilvocarcin V production, has provided a useful tool for several successful biosynthetic studies (see Gilvocarcin V Biosynthetic Highlights). To determine the exact biosynthetic role of GilP and GilQ, two cosG9B3 mutants were constructed harboring an in-frame deletion of *gilP* (cosG9B3-GilP$^-$) and *gilQ* (cosG9B3-GilQ$^-$). The resulting cosmids were transformed, using conjugal transfer, into *S. lividans* TK24 producing *S. lividans* TK24/cosG9B3-GilP$^-$ and *S. lividans* TK24/cosG9B3-GilQ$^-$, respectively.

Fermentation and HPLC-MS analysis of the mutant strains, as well as *S. lividans* TK24/cosG9B3, were conducted to determine the percentage of gilvocarcin V, M and E produced by each strain (**Table 10**).

Table 10. Percentage of individual gilvocarcin congeners produced by engineered strains in this study.

Strain	% GM (50)[a]	% GV (49)[a]	% GE (51)[a]
SLTK24[b]/cosG9B3	34.6 ± 1.8	65.4 ± 1.8	0
SLTK24[b]/cosG9B3-GilP$^-$	36.5 ± 1.6	52.3 ± 2.2	11.2 ± 1.8
SLTK24[b]/cosG9B3-GilQ$^-$	77.2 ± 0.9	22.8 ± 1.2	0
SLTK24[b]/cosG9B3-pGilQGilOIII	16.5 ± 0.7	83.5 ± 0.7	0
SLTK24[b]/cosG9B3-pGilQ	11.6 ± 0.7	83.7 ± 0.6	4.7 ± 0.01

[a]Percentage of gilvocarcins calculated by the relative values of area under the curve taken from HPLC-MS chromatogram traces (n =3), [b]*S. lividans* TK24 (SLTK24)

The control strain, *S. lividans* TK24/cosG9B3, produced roughly 35% and 65% of gilvocarcin M and V, respectively. The *gilP* deficient mutant, *S. lividans* TK24/cosG9B3-GilP$^-$, produced gilvocarcins in percentages comparable to the control strain. Combining both gilvocarcin V and E percentages from the *gilP* mutant reveals 63% propionate incorporation. This observation clearly shows that the removal of GilP from the gilvocarcin biosynthetic machinery does not influence the overall production of gilvocarcins. This in not unexpected, as GilP shows high sequence similarity to MCATs, and pathway specific MCATs are typically not required for type II PKS systems. This result implies that in the absence of GilP, the endogenous FAS MCAT from *S. lividans* TK24 is able to successfully transfer malonyl-CoA to the *gil* PKS. Furthermore, this suggests GilP functions as a typical MCAT, and does not play an important role in the selection of propionate during gilvocarcin V biosynthesis.

In contrast, analysis of the *gilQ* deficient strain showed a severely altered gilvocarcin production profile (**Table 10**). *S. lividans* TK24/cosG9B3-GilQ$^-$ produced significantly more gilvocarcin M than the control strain, 77% versus 35%, respectively. Consequently the production of gilvocarcin V was reduced to roughly 23% compared to 65% in the control strain. The dramatic decrease in gilvocarcin V production suggests GilQ primarily influences the use of propionate over acetate. Interestingly, the complete abolishment of GV was not observed demonstrating GilP or an endogenous

acyltransferase from *S. lividans* TK24 may have the ability, albeit at a much reduced rate, to load and transfer propionate to the minimal gilvocarcin PKS.

Based on the above *in vivo* results, an additional strain was created in which GilQ was over-expressed in *S. lividans* TK24/cosG9B3 to increase the propensity for catalysis utilizing a two carbon starter unit. The proposed increase of propionate primed decaketides would lead to an increase in two carbon C-8 side chains that would need to be converted to the vinyl said chain by GilOIII in order to produce gilvocarcin V (refer to **Figure 16**). Normally, GilOIII can completely convert all two carbon C-8 side chains to the vinyl functional group as evident by the absence of gilvocarcin E in *S. lividans* TK24/cosG9B3 fermentations (**Table 10**). The possible increase in GilOIII substrate, created by over-expressing GilQ, may saturate GilOIII leading to an incomplete oxidation of the two carbon C-8 side chain preventing production of the ideal congener, gilvocarcin V. For this reason, a second construct was created in which GilQ and GilOIII could be over-expressed in *S lividans* TK24/cosG9B3 to completely flux all intermediates to **49**.

The resulting strains, *S. lividans* TK24/cosG9B3/pGilQ and *S. lividans* TK24/cosG9B3/pGilQGilOIII, were fermented and analyzed as discussed above (**Table 10**). Both strains increased **49** production to roughly 85% and reduced **50** production to approximately 15% compared to 65% and 35% in control, respectively. The inclusion of GilOIII in *S. lividans* TK24/cosG9B3/pGilQGilOIII was able to completely convert all gilvocarcin E intermediates to **49**, unlike *S. lividans* TK24/cosG9B3/pGilQ which produced around 5% of **51**. The increase of **49** production in strains containing additional GilQ further validates the unique role of this acyltransferase in starter unit selectivity during gilvocarcin V biosynthesis.

Surprisingly, strains over-expressing GilQ not only increased the percentage of **49** produced compared to **51**, but also significantly increased the overall production yield of gilvocarcins (**Table 11**). Incredibly, *S. lividans* TK24/cosG9B3/pGilQ and *S. lividans* TK24/cosG9B3/pGilQGilOIII increased gilvocarcin production to 450 mg/L and 200 mg/L, respectively. This correlates to a 10-20 fold increase when compared to the wild type strain. Together, these findings suggest propionate incorporation is the earliest rate limiting step of gilvocarcin biosynthesis and is primarily dependent upon the activity of GilQ.

Table 11. Total gilvocarcin yields produced by engineered strains over-expressing GilQ.

Strain	GM (50) (mg/L)[a]	GV (49) (mg/L)[a]	GE (51) (mg/L)[a]	Total (mg/L)[a]
SLTK24[b]/cosG9B3	7 ± 1	14 ± 3	0	21 ± 4
SLTK24[b]/cosG9B3/pGilQ	53 ± 11	376 ± 65	21 ± 4	450 ± 79
SLTK24[b]/cosG9B3/pGilQGilOIII	33 ± 3	168 ± 10	0	201 ± 13

[a]Production of gilvocarcins calculated by the relative values of area under the curve taken from HPLC-MS chromatogram traces (n =3), compared to a gilvocarcin V standard curve. [b]*S. lividans* TK24 (SLTK24)

Discussion

The *in vivo* investigation into the roles of GilP and GilQ regarding starter unit specificity revealed that only in the absence of GilQ did the production profile for gilvocarcins change. The role of GilP as a *gil* pathway associated MCAT was indirectly supported by the successful recruitment of an endogenous host FAS MCAT in the *gilP* deficient mutant strain. Inactivation of *gilQ* led to a marked decrease in gilvocarcin V production, and therefore a decrease in propionate priming during decaketide formation. Interestingly, gilvocarcin V production was partially retained in *S. lividans* TK24/cosG9B3-GilQ$^-$, suggesting GilQ is primarily, although not completely, responsible for propionate incorporation during **49** biosynthesis.

Utilizing this information, rationally designed recombinant strains, over-expressing GilQ, were created in which the preference for propionate incorporation was increased 23% compared to the control strain. In addition to increased propionate incorporation, these strains were also able to produce gilvocarcins in yields 10-20 times that of the wild type (*S. griseoflavus* Gö3592) or control strain (*S. lividans* TK24/cosG9B3), correlating to up to 375 mg/L of gilvocarcin V production.

Materials and Methods

Bacterial strains, culture conditions and plasmids

All complementations, conjugations, protoplast transformations and culturing conditions were carried out as described in section 2.1: Bacterial strains, culture conditions and plasmids. A comprehensive list of strains and plasmids used in this study can be found in **Table 12.**

Table 12. Strains and plasmids used in investigating the role of GilP and GilQ *in vivo*.

Strain/Plasmid	Characteristics and relevance	References
E. coli XL1Blue-MRF	Cloning host	Stratagene
E.coli BW25113/pKD20	Host for homologous recombination	Datsenko, K. et al.[161]
E. coli ET12567/pUZ8002	Host for conjugal transfer	MacNeil, D. et al.[123-124]
pIJ790	pKD20 derived plasmid	Gust, B. et al.[162]
PCR-Blunt II-TOPO	PCR fragment cloning vector	Invitrogen
pEM4	*Streptomyces* expression vector	Quiros, L. et al.[112]
cosG9B3 (1)[a]	PCR template for *gilP* and *gilQ*	Fischer, C. et al.[91]
pGilQ-1-TOPO	*gilQ* (*Pst*I/*Nhe*I) cloned into TOPO	This study
pGilQ-2-TOPO	*gilQ* (*Nhe*I/*Xba*I) cloned into TOPO	This study
pGilOIII-TOPO	*gilOIII* (*Xba*I/*Eco*RI*)* cloned into TOPO	This study
pGilQ (4)[a]	*gilQ* cloned into pEM4	This study
p1Q	*gilQ* cloned into pEM4	This study
pGilQGilOIII (5)[a]	*gilQ* and gilOIII cloned into pEM4	This study
cosG9B3-GilP$^-$ (2)[a]	*gilP* deletion mutant of cosG9B3	This study
cosG9B3-GilQ$^-$ (3)[a]	*gilQ* deletion mutant of cosG9B3	This study
S. lividans TK24 (SLTK24)	Heterologous expression host	Kieser, T. et al.[121]
S. lividans TK24/(1)[a]	SLTK24 transformed with (1)[a]	Fischer, C. et al.[91]
S. lividans TK24/(2)[a]	SLTK24 transformed with (2)[a]	This study
S. lividans TK24/(3)[a]	SLTK24 transformed with (3)[a]	This study
S. lividans TK24/(1)[a]/(4)[a]	SLTK24/cosG9B3 transformed with (4)[a]	This study
S. lividans TK24/(1)[a]/(5)[a]	SLTK24/cosG9B3 transformed with (5)[a]	This study

[a]Numbering used only for table simplification

Fermentation and metabolite screening

The recombinant strains created in this study were fermented and screened in triplicate as described in section 2.1: Fermentation and metabolite screening.

DNA isolation, DNA manipulation and PCR

Plasmid/cosmid DNA isolations, manipulations and PCR reactions were carried out as described in section 2.1: DNA isolation, DNA manipulation and PCR. Primers used for inactivation experiments as well as over-expressing constructs are summarized in **Table 13**.

Table 13. Oligonucleotide sequence of primers used to inactivate and amplify *gilQ* and *gilP* from cosG9B3. Restriction sites are represented by italicized sequences.

Primer	Oligonucleotide sequence (5′-3′)
GilQ-Inact-F	CGTCCTTCCTGGCCGGCGTGACGGGCTGGTGAGCCGGTGATTCCGGGGAT CCGTCGACC
GilQ-Inact-R	CATGGCCGGCCTTGCGCGTTGCACGTAGTGCAATGGTCATGTAGGCTGGA GCTGCTTC
GilQ-Cntrl-F	TCTAGAGCAAGGTCCTCTCGGGATTC
GilQ-Cntrl-R	ATGGGTGCCTCCCGGGAACTCG
GilP-Inact-F	GCGCCTCCACGACGACTTGGAACCCGAGGTGACGCGGTGATTCCGGGGAT CCGTCGACC
GilP-Inact-R	ATGCGGCACTGCGTGCTTCACCGGCTCACCAGCCCGTCATGTAGGCTGGA GCTGCTTC
GilP-Cntrl-F	TCTAGATCCGTCCCGATCACCACTG
GilP-Cntrl-R	GACTCCCATGGCAGCGAACTG
GilQ-F1	GGC*CTGCAG*GGCTGGTGAGCCGGTGAAGCACGCA
GilQ-R1	GCGT*GCTAGCTTCTAGA*ATGGTCAACAGAATTCCTCGGCGACCT
GilQ-F2	TG*GCTAGC*CGGAGAAGCACGCAATGCCGCATCAGGCAACC
GilQ-R2	CGG*TCTAGA*GCGTTGCACGTAGTGCAATGGTCA
GilOIII-F	CC*TCTAGA*GGAAAGGATGAAGCGATGATCTCCACA
GilOIII-R	GTC*GAATTC*GACCACCGTCACGTCCTCGACG

Inactivation of *gilP* and *gilQ*

The inactivation of *gilP* and *gilQ* from cosG9B3 (previously unpublished results by L. Zhu) was carried out using a modified PCR-targeting REDIRECT protocol described elsewhere.[90] Briefly, a chloramphenicol (cat^r) resistance gene with engineered flanking flippase recognition target (FRT) and flippase (FLP) sites was used as a template to amplify inactivation cassettes using the primer pairs GilQ-Inact-F + GilQ-Inact-R and GilP-Inact-F + GilP-Inact-R (**Table 13**). The resulting chloramphenicol inactivation cassettes were then introduced into *E. coli* BW25113/pKD20 (amp^r) harboring cosG9B3 (apr^r) by electroporation. Elevating the incubation temperature facilitates the loss of the temperature sensitive pKD20, and resulting chl and apr resistant colonies were collected and evaluated further. Sequences flanking the start and stop codons of both *gilP* and *gilQ* were used to create control primer pairs (GilP-Cntrl-F+ GilP-Cntrl-R and GilQ-Cntrl-F + GilQ-Cntrl-R) used to confirm the replacement of the targeted genes with the chloramphenicol cassette (**Table 13**). FLP-mediated excision of the inactivation cassette resulted in an 82 bp in-frame scar confirmed by the previously discussed control primer pairs. The final mutated cosmids, cosG9B3-$GilP^-$ and cosG9B3-$GilQ^-$, were transformed into *E. coli* ET12567/pUZ8002 and conjugally transferred into *S. lividans* TK24

producing *S. lividans* TK24/cosG9B3-GilP$^-$ and *S. lividans* TK24/cosG9B3/GilQ$^-$, respectively.

Over-expression constructs

The primer pair GilQ-F1 and GilQ-R1was used to amplify *gilQ*, including the natural putative RBS, from cosG9B3. The amplified product was gel purified and cloned into PCR-Blunt II-TOPO (Invitrogen), as described previously, creating pGilQ-1-TOPO. Restriction digests were used to remove *gilQ* from pGilQ-1-TOPO as a *Pst*I/*Nhe*I fragment and subsequently ligated into pEM4 producing pGilQ. Similarly, *gilQ* and *gilOIII* were amplified from cosG9B3 using the primer pairs GilQ-F2 + GilQ-R2 and GilOIII-F + GilOIII-R, respectively. As before, both PCR products were cloned into TOPO producing pGilQ-2-TOPO and pGilOIII-TOPO. In order to clone both genes into pEM4, *gilQ* was first removed from pGilQ-2-TOPO as a *Nhe*I/*Xba*I fragment and ligated in pEM4 digested with only *Xba*I producing p1Q. Additional restriction digest analysis using *Hind*III/*Xba*I confirmed the correct orientation of *gilQ* in p1Q. Finally, *gilOIII* was taken from pGilOIII-TOPO as an *Xba*I/*Eco*RI fragment and ligated in p1Q prepared accordingly. The resulting construct, pGilQGilOIII, contained both *gilQ* and *gilOIII* downstream of the constitutively active promoter, *ermE**p. Both over-expression constructs were then transformed via protoplast transformation into *S. lividans* TK24/cosG9B3 creating *S. lividans* TK24/cosG9B3/pGilQ and *S. lividans* TK24/cosG9B3/pGilQGilOIII.

4.2 *In vitro* characterization of GilP and GilQ

The above *in vivo* investigations into GilP and GilQ function suggests that GilQ is the primary determinant for propionate incorporation, however, in the absence of GilQ, the gilvocarcin biosynthetic machinery could continue to produce propionate primed decaketides. This may be due to partially relaxed substrate specificity of GilP towards acetyl-CoA (ACoA) and propionyl-CoA (PCoA). To investigate the possibility of GilP assisted propionate incorporation as well as to further confirm the unique activity of GilQ, we chose to examine the substrate flexibility of both GilP and GilQ towards transferring individual acyl-CoA substrates to the activated ACP, RavC, from the related

ravidomycin biosynthetic pathway.[117,163] The natural gilvocarcin pathway specific ACP, GilC, was not used in this study due to an inability to express it as a soluble protein.[163]

Experimental design

In **specific aim 3b**, the gilvocarcin genes, *gilP* and *gilQ*, will be expressed and purified as soluble proteins from *E. coli*. The resulting purified enzymes will be subjected to an acyl transfer assay where their specific activity towards transferring several acyl-CoA species will be evaluated. The acyl transfer assay will determine the ability of GilP and GilQ to transfer acetyl-, propionyl, malonyl- and methylmalonyl-CoA to the *holo*-RavC (activated RavC). In this context RavC and Svp (used to activate RavC) must also be prepared as soluble proteins.

Results

GilP and GilQ acyl transfer specific activity

To test the substrate specificity of GilP and GilQ, *N*-terminal $(His)_6$-tagged proteins of GilP, GilQ, RavC, $RavC_1$ and Svp were expressed in *E. coli* and purified using IMAC ($RavC_1$ will be discussed below). The correct size and purity of each protein was confirmed by SDS-PAGE analysis (**Figure 56**). The calculated sizes of GilP, GilQ, RavC, $RavC_1$ and Svp were 33.7 kDa, 38.9 kDa, 11.3 kDa, 10.8 kDa and 25.6 kDa, respectively. Svp appears to be larger than expected, but it has been reported to run high on SDS-PAGE gels.[164]

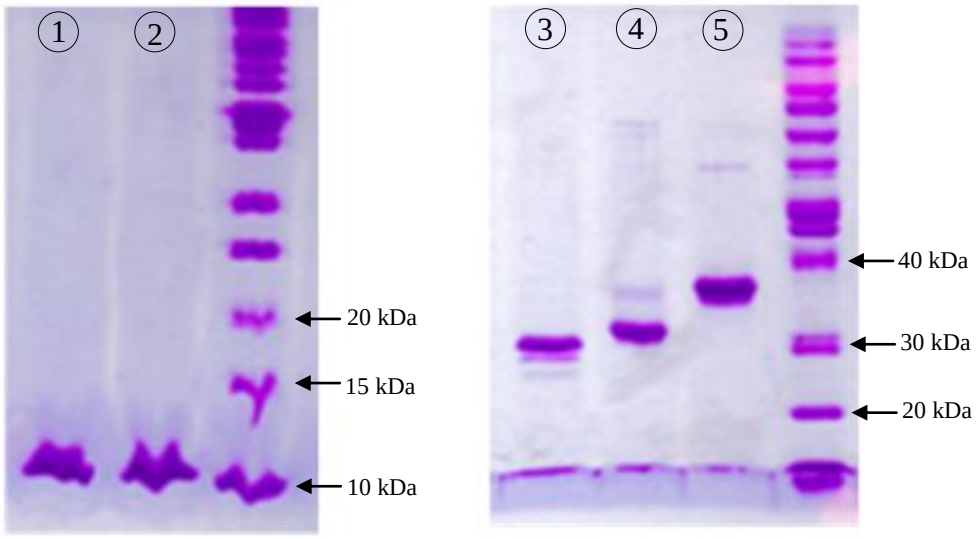


Figure 56. SDS-PAGE analysis of *N*-terminal $(His)_6$-tagged proteins (1) RavC, (2) $RavC_1$, (3) Svp, (4) GilP and (5) GilQ.

Expression of the *Streptomyces* ACPs, RavC and $RavC_1$, in *E. coli* resulted in almost the exclusive production of inactive or *apo*-ACPs. ACPs contain a conserved serine residue which must be phosphopantetheinylated, or activated, by a phosphopantetheinyl transferase (PPTase). The terminal thiol of the Coenzyme A (CoA) derived phosphopantetheinyl arm of the activated ACP (*holo*-ACP) is responsible for the catalytic activity of ACPs. Commonly, heterologous expression of ACPs results in predominantly the inactive form (*apo*-ACPs). To circumvent this problem, promiscuous PPTases have been identified, such as Svp, that can readily convert purified *apo*-ACPs to functional *holo*-ACPs (**Figure 57**).

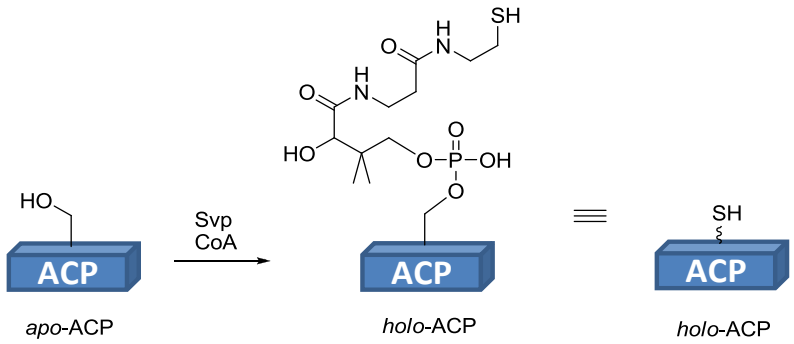


Figure 57. Activating the acyl carrier protein. The functional *holo*-ACP is produced by the transfer of the phosphopantetheinyl prosthetic group from CoA to *apo*-ACP via a phosphopantetheinyl transferase, such as Svp.

An acyl transfer assay was utilized in which purified GilP and GilQ were incubated separately with *holo*-RavC and individual ^{14}C-acyl-CoA species (^{14}C-acetyl-CoA, ^{14}C-malonyl-CoA, ^{14}C-propionyl-CoA, and ^{14}C-methylmalonyl-CoA). The reactions were conducted in triplicate and their compositions are summarized in **Table 14**, reactions 1-30. The ability of GilP and GilQ to transfer specific ^{14}C-acyl-CoAs to *holo*-RavC (^{14}C-acyl-CoA-RavC) were monitored by scintillation counting, and represented as specific activity in **Table 15**.

Table 14. List of reaction combinations for acyl transfer assays (1- 30) and self loading studies (31-33).

Reaction #	***holo*-RavC (8 µM)**	**GilP (0.8 µM)**	**GilQ (0.8 µM)**	**^{14}MCoA (8 µM)**	**^{14}ACoA (8 µM)**	**^{14}MMCoA (8 µM)**	**^{14}PCoA (8 µM)**
1,2,3	X	X	-	-	-	-	-
4,5,6	X	X	-	X	-	-	-
7,8,9	X	X	-	-	X	-	-
10,11,12	X	X	-	-	-	X	-
13,14,15	X	X	-	-	-	-	X
16,17,18	X	-	X	-	-	-	-
19,20,21	X	-	X	X	-	-	-
22,23,24	X	-	X	-	X	-	-
25,26,27	X	-	X	-	-	X	-
28,29,30	X	-	X	-	-	-	X
31	X	-	-	X	-	-	-
32	-	X	-	X	-	-	-
33	X	X	-	X	-	-	-

(X), added to reaction; (-), not added to reaction; ^{14}MCoA, ^{14}C-malonyl-CoA; ^{14}ACoA, ^{14}C-acetyl-CoA; ^{14}MMCoA, ^{14}C-methylmalonyl-CoA; ^{14}PCoA, ^{14}C-propionyl-CoA, final concentration for 100 µL reactions shown in parenthesis.

Table 15. Specific activity of GilP and GilQ

Acyltransferase	^{14}C-labeled Substrates	Specific Activity [a] (unit/mg)	Relative Activity
GilP	^{14}C-acetyl-CoA	16 x 10^3	1
GilP	^{14}C-malonyl-CoA	463 x 10^3	29
GilP	^{14}C-propionyl-CoA	86 x 10^3	5
GilP	^{14}C-methylmalonyl-CoA	301 x 10^3	19
GilQ	^{14}C-acetyl-CoA	16 x 10^3	1
GilQ	^{14}C-malonyl-CoA	27 x 10^3	2
GilQ	^{14}C-propionyl-CoA	51 x 10^3	3
GilQ	^{14}C-methylmalonyl-CoA	15 x 10^3	1

[a]One unit is defined as the amount of His_6-GilP or His_6-GilQ required to catalyze the synthesis of 1 *p*mol of acetyl, malonyl, propionyl or methylmalonyl-RavC per sec at pH 7.5 and 30 ^oC with 8 µM RavC and 8 µM ^{14}C-labeled substrates.

The specific activity of GilP clearly shows a strong preference toward malonyl-CoA, and is comparable to other characterized *Streptomyces* MCAT proteins.[165] Similarly, GilP shows a strong preference for methylmalonyl-CoA. This may be an artifact created by the way MCAT binds the carboxylic acid moiety of acyl-CoA molecules. As methylmalonyl-CoA and malonyl-CoA both contain a carboxylic group and differ only at their beta position, it is reasonable to observe high specific activity for methylmalonyl CoA. Interestingly, GilP shows significant activity for propionyl-CoA, with a relative activity of 1 to 5 when compared with malonyl-CoA. This suggests that GilP can in fact transfer propionyl-CoA, and explains the presence of gilvocarcin V in the strains lacking the critical starter unit determining enzyme, GilQ.

Comparing the specific activity of GilQ reveals a preference for propionyl-CoA over any other acyl-CoA species tested. Surprisingly, GilQ only shows a 2-fold higher preference for producing propionyl-RavC compared to malonyl-RavC, and is roughly 60% less active than GilP at utilizing propionyl-CoA.

In addition to RavC, $RavC_1$ was included in initial acyl transfer reactions. RavC and $RavC_1$ are ACPs found in the ravidomycin biosynthetic gene cluster. There are only a few examples of type II PKS systems containing two individual ACP proteins. One example, R1128A-D (**151**-**154**), has two unique ACPs that were found to be specific for initiation and elongation. It was hypothesized that RavC and $RavC_1$ may also serve a similar purpose in the biosynthesis of ravidomycin; however, initial results indicated that neither ACP exhibited substrate preference (data not shown). Based on these preliminary results, $RavC_1$ was removed from further experimentation and concluded that RavC and $RavC_1$ are most likely functional repeats.

***holo*-RavC self loading capability**

It is well documented that *holo*-ACPs contain an intrinsic ability to load acyl-CoA species in the absence of an acyltransferase.[166] It is therefore important to quantify the self loading ability of *holo*-RavC in order to validate the previous acyl transfer assay. This will allow confirmation that during the acyl transfer assay acyl-RavC species were created by acyltransferase activity and not by the self loading properties of *holo*-RavC. Under identical conditions as the acyl transfer assay, *holo*-RavC and GilP were incubated separately and together with ^{14}C-malonyl-CoA (**Table 14**, reactions 31-33). The presence of ^{14}C-malonyl-RavC and ^{14}C-malonyl-GilP were quantified using scintillation counting as before, and the results are summarized in **Figure 58**. The initial rate of *holo*-RavC (reaction 31) and GilP (reaction 32) self loading of ^{14}C-malonyl-CoA were calculated to be 0.03 μM min^{-1} and 0.67 μM min^{-1}, respectively. The production of ^{14}C-malonyl-RavC through ^{14}C-malonyl transfer by GilP (reaction 33) was calculated to have an initial rate of 3.81 μM min^{-1}. These results show the initial rate of self loading by *holo*-RavC or GilP could not significantly contribute to the acyl transfer assay ^{14}C-radiolabeled proteins accumulated in the time frame tested.

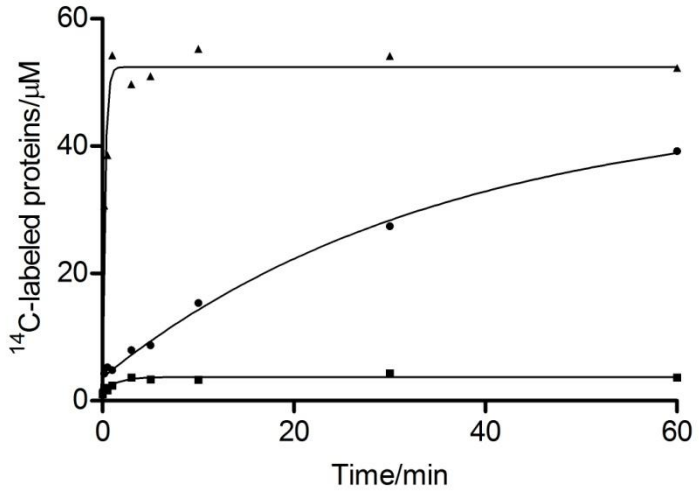


Figure 58. Non-linear regression curve of the self loading properties of *holo*- RavC (●) and GilP (■) compared to *holo*-RavC + GilP (▲).

Activity of GilPS$_{90}$A, GilQS$_{111}$A and RavCS$_{39}$A

Bioinformatical analysis of *gilP*, *gilQ* and *ravC* revealed the anticipated conserved active site motifs for acyltransferases (xGHSxGE) and ACPs (LGxDSLxxVE). To confirm the activity exhibited by GilP, GilQ and RavC were due to the expected active site sequences, the active Ser residue was replaced with Ala using site directed mutagenesis. The resulting proteins GilPS$_{90}$A, GilQS$_{111}$A, and RavCS$_{39}$A were substituted for their wild type counterparts in several acyl transfer reactions to determine their ability to transfer or load acyl-CoA substrates (**Table 16**).

Table 16. List of reaction combinations used to test the loading and transfer functionality of the mutated proteins GilPS$_{90}$A, GilQS$_{111}$A, and RavCS$_{39}$A.

Reaction #	*holo*-RavC (8 μM)	GilP (0.8 μM)	GilQ (0.8 μM)	RavCS$_{39}$A (8 μM)	GilPS$_{90}$A (0.8 μM)	GilQS$_{111}$A (0.8 μM)	M-CoA (8 μM)	P-CoA (8 μM)
34	X	X	-	-	-	-	X	-
35	-	X	-	X	-	-	X	-
36	X	-	-	-	X	-	X	-
37	-	-	-	X	X	-	X	-
38	-	-	X	-	-	-	-	X
39	-	-	-	-	-	X	-	X

(X), added to reaction; (-), not added to reaction; M-CoA, ^{14}C-malonyl-CoA; P-CoA, ^{14}C-propionyl-CoA; final concentration for 100 µL reactions shown in parenthesis.

Phosphor imaging was used to visualize the activity of the generated mutants (**Figure 59**). GilPS$_{90}$A and GilQS$_{111}$A, unlike GilP and GilQ, were both shown to be unable to load ^{14}C-malonyl-CoA or ^{14}C-propionyl-CoA to *holo*-RavC, respectively (**Figure 59**, reactions 34 and 36-39). The slight loading of holo-RavC seen in reaction 36 is due to the self loading property of *holo*-RavC and is not due to the presence of GilPS$_{90}$A. Similarly, the removal of the active site serine in RavCS$_{39}$A also inhibited loading of ^{14}C-malonyl-CoA by the functional GilP (**Figure 59**, reaction 35). The inactivity of all mutants generated suggests that their activity is due to the expected active site motifs.

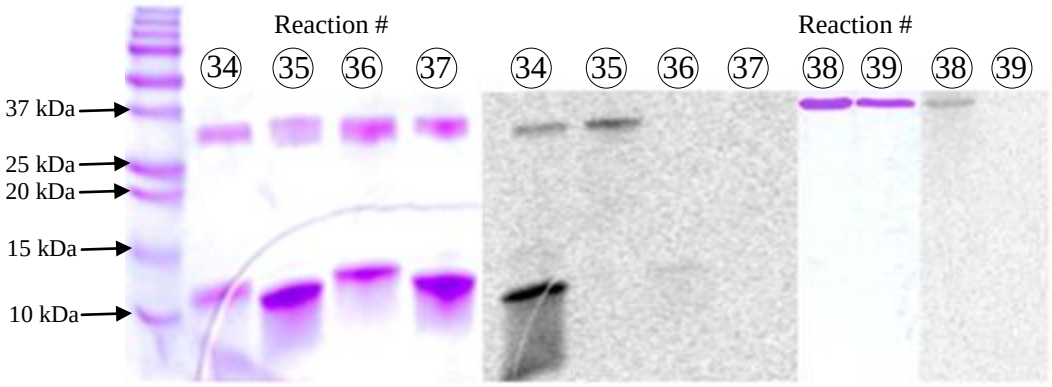


Figure 59. ^{14}C-acyl transfer assay using mutated GilPS$_{90}$A, GilQS$_{111}$A, and RavCS$_{39}$A.

Discussion

Taken together, the *in vivo* and *in vitro* investigation into GilP and GilQ functionality allows us to delineate the earliest steps of gilvocarcin biosynthesis. From bioinformatical analysis as well as experimental data it is clear that GilP and GilQ function as acyltransferases, and work with the gilvocarcin PKS to produce 20- and 21-carbon decaketides primed with acetate or propionate starter units, respectively. The severe reduction in gilvocarcin V production by the GilQ deficient strain suggested that GilQ is the primary determinant for starter unit specificity; however, the presence of gilvocarcin V in the absence of GilQ hinted at an additional route for propionate incorporation. The specific activity of GilP revealed that it addition to strong malonyl-CoA preference, GilP could also accept propionyl-CoA explaining the presence of propionate primed gilvocarcin congeners in strains lacking GilQ. It is interesting that even though GilP has a higher specific activity towards loading propionyl-CoA than GilQ, the loss of GilQ, and not GilP, in an *in vivo* system drastically reduces **49** production. This can be explained by the primary MCAT functionality of GilP which is to load and transfer malonyl-CoA for the extension of the growing ketide. Even though GilP can load propionyl-CoA it does so at a much reduced rate compared to malonyl-CoA as evident in both *in vivo* and *in vitro* experimentation. GilQ on the other hand prefers propionyl-CoA over malonyl-CoA, and despite the fact that it has a lower specific activity towards loading propionyl-CoA; in an *in vivo* environment, GilP would almost

exclusively load malonyl-CoA while GilQ would be responsible for loading propionyl-CoA (**Figure 60**, route A and elongation).

The initiation of acetate in gilvocarcin M biosynthesis is likely facilitated by GilP; however, unlike traditional acetate initiation where malonyl-CoA is decarboxylated by the KS_β subunit to create acetyl-CoA, we propose direct loading and transfer of acetyl-CoA to *holo*-ACP facilitated by GilP. This hypothesis is supported by the recent observation that acetyl-CoA and malonyl-CoA are absolute requirements for the production of 20-carbon decaketides by the *gil* PKS.[163] Additionally, the highly conserved glutamine (or glutamate) residue of type II KS_β found to be responsible for decarboxylation is replaced by a shorter aspartate residue in the *gil* KS_β, GilB. This aspartate residue may be too short to initiate decarboxylation of the ACP-bound malonate necessitating the need for direct loading and transfer of acetyl-CoA (**Figure 60**, route B).

To the best of our knowledge, GilQ is the first and only characterized acyltransferase responsible for non-acetate initiation in a type II PKS pathway. Ravidomycin and chrysomycin are the only additional gilvocarcin-like aryl C-glycosides with isolated biosynthetic gene clusters, and both were found to contain a GilQ homologue.[117] In this context, GilQ as well as RavQ and ChryQ may comprise a unique category of type II PKS acyltransferases that control selection and utilization of the non-acetate starter unit propionyl-CoA.

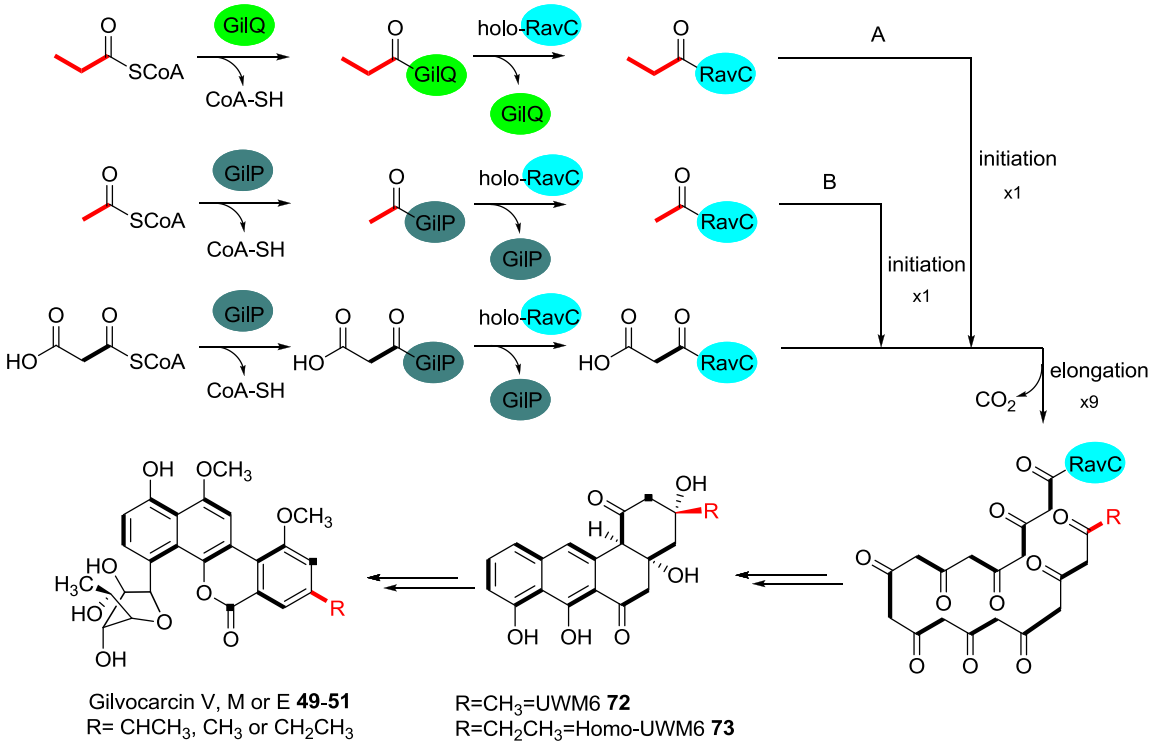


Figure 60. The role of GilP, GilQ and RavC in initiation and elongation during the biosynthesis of gilvocarcins.

Materials and Methods

Bacterial strains, culture conditions and plasmids

All transformation and culturing conditions were carried out as described in section 2.1: Bacterial strains, culture conditions and plasmids. A comprehensive list of strains and plasmids used in this study can be found in **Table 17**.

Table 17. Strains and plasmids used in the expression of GilP, GilQ, RavC, $RavC_1$ and Svp.

Strain/Plasmid	Characteristics and relevance	References
E. coli XL1Blue-MRF	Cloning host	Stratagene
E. coli BL21 (DE3)	Host for protein expression	Invitrogen
PCR-Blunt II-TOPO	PCR fragment cloning vector	Invitrogen
pET28a(+)	Protein expression vector	Novagen
cosG9B3	PCR template for *gilP* and *gilQ*	Fischer, C. et al.[91]
cosRav32	PCR template for *ravC* and *ravC$_1$*	Kharel, M. et al.[117]
pGilQ-3-TOPO	*gilQ* (*Nde*I/*Bam*HI) cloned into TOPO	This study
pRavC$_1$-TOPO	*ravC$_1$* (*Nde*I/*Eco*RI) cloned into TOPO	This study
pET-GilP	*gilP* cloned into pET28a(+)	Kharel, M. et al.[163]
pET-GilQ	*gilQ* cloned into pET28a(+)	This study
pET-RavC	*ravC* cloned into pET28a(+)	Kharel, M. et al.[163]
pET-RavC$_1$	*ravC$_1$* cloned into pET28a(+)	This study
pET-GilPS$_{90}$A	*gilPS$_{90}$A* in pET28a(+)	This study
pET-GilQS$_{111}$A	*gilQS$_{111}$A* in pET28a(+)	This study
pET-RavCS$_{39}$A	*ravCS$_{39}$A* in pET28a(+)	This study
pBS18	*svp* cloned into pQE-70 expression vector	Sanchez, C. et al.[164]

DNA isolation, DNA manipulation and PCR

Plasmid DNA isolations, manipulations and PCR reactions were carried out as described in section 2.1: DNA isolation, DNA manipulation and PCR. Please find **Table 18** for a complete list of primers used in this study.

Table 18. Oligonucleotide sequence of primers used to amplify *gilQ*, *ravC*, *gilP* and *ravC$_1$*. Restriction sites are represented by italicized sequences.

Primer	Oligonucleotide sequence (5′-3′)
GilQ-F3	AA*CATATG*GTGCCGCATCAGGCAACC
GilQ-R3	AA*GGATCC*TCAACAGAATTCCTCGGC
RavC1-F	ATT*CATATG*ACCACCGGCACGTTCACC
RavC1-R	ATT*GAATTC*TCACGCCGCGTTGACCAGCTC
GilP-S$_{90}$A-F	CATCGCGGGCCACGCTCTGGGCGAGTAC
GilP-S$_{90}$A-R	GTACTCGCCCAGAGCGTGGCCCGCGATG
GilQ-S$_{111}$A-F	GCTGGTCGGTCACGCCGTGGGCGAGCTG
GilQ-S$_{111}$A-R	CAGCTCGCCCACGGCGTGACCGACCAGC
RavC-S$_{39}$A-F	CTGGGGTACGACGCCCTGGCGCTGC
RavC-S$_{39}$A-R	GCAGCGCCAGGGCGTCGTACCCCAG

Protein expression constructs

The primer pairs GilQ-F3 + GilQ-R3 and $RavC_1$-F + $RavC_1$-R were used to amplify *gilQ* and *ravC₁* from cosG9B3 and cosRav32, respectively. The *pfu* amplified PCR products were gel purified and cloned into PCR-Blunt II-TOPO vector, producing pGilQ-3-TOPO and p$RavC_1$-TOPO, respectively. Restriction digests were used to remove *gilQ* and *ravC₁* as *Nde*I/*Bam*HI and *Nde*I/*Eco*RI fragments and subsequently ligated into pET28a(+) generating pET-GilQ and pET-$RavC_1$, respectively. The remaining protein expression constructs used in this study (GilP, RavC and Svp) were reported elsewhere.[163-164]

Primer sets GilP-S_{90}A-F + GilP-S_{90}A-R, GilQ-S_{111}A-F + GilQ-S_{111}A-R and RavC-S_{39}A-F + RavC-S_{39}A-R were used with the QuickChange lightning site-directed mutagenesis kit (Stratagene) to produce pET-GilPS_{90}A, pET-GilQS_{111}A, and pET-$RavCS_{39}$A from pET-GilQ, pET-GilP and pET-RavC, respectively.

Expression and purification of proteins

The expression of soluble GilP, GilQ, RavC, $RavC_1$, Svp, GilPS_{90}A, GilQS_{111}A, and $RavCS_{39}$A proteins were carried out as described in section 3.1: Expression and purification of proteins.

RavC and $RavC_1$ activation reaction conditions

The production of *holo*-ravC and *holo*-$RavC_1$ was obtained through a modified protocol presented elsewhere.[167] The reaction conditions were as follows: 50 mM KH_2PO_4 (pH 7.5), 12.5 mM $MgCl_2$, and a 1:20:200 molar ratio of Svp:*apo*-ACP:CoA. The reaction was incubated at 30 ^oC and 10 μL samples were taken at 60 min time intervals and analyzed by HPLC with a Platinum C4-EPS-300 column (250 x 4.6 mm, 5 µm; Grace, Deerfield, IL). Samples were eluted with linear gradients from solvent A (0.1% trifluoroacetic acid (TFA) in 10% acetonitrile) to solvent B (0.1% trifluoroacetic acid in 90% acetonitrile): 0 to 5 min, 5% B; 5 to 32 min, gradient from 5 to 95% B; 32-40 min, hold at 95% B; and 40 to 45 min, gradient from 95 to 5% B. Peaks were collected, dried under vacuum and analyzed by matrix assisted laser desorption ionization time of flight (MALDI-TOF) mass spectrometry. Under these conditions *apo*-RavC and *apo*-

$RavC_1$ were converted to their active *holo*-forms in 60 min (**Figure 61**). Peak 1 had a mass of 11187 Da and was representative of *apo*-RavC (cal. 11197 Da), and peak 2 had a mass of 11524 Da indicating the presence of *holo*-RavC (cal. 11537 Da).

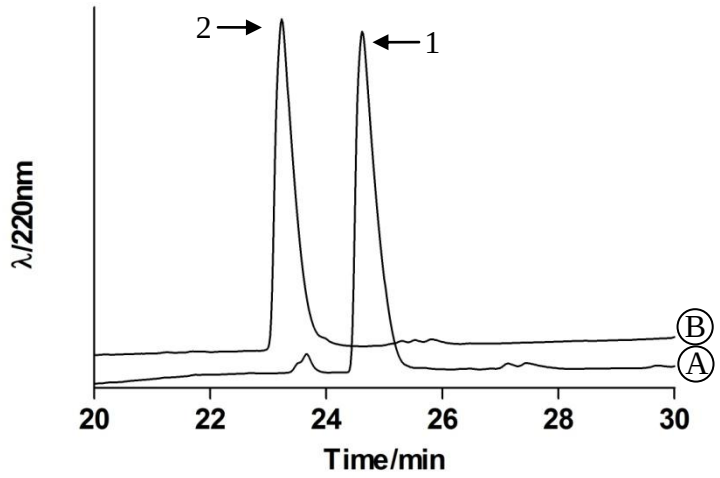


Figure 61. Conversion of *apo*-RavC (peak 1) to *holo*-RavC (peak 2) at time points (A) 0 min and (B) 60 min.

Acyl transfer assay conditions

An acyl transfer assay was set up to determine the specific activity of GilP and GilQ to form radiolabeled acyl-RavC. The radiolabeled acyl-CoAs used in this study were malonyl-CoA (malonyl-2-^{14}C, Perkin Elmer), acetyl-CoA (acetyl-1-^{14}C, Perkin Elmer), methylmalonyl-CoA (methyl-^{14}C, American Radiolabeled Chemicals), and propionyl-CoA (propionyl-1-^{14}C, American Radiolabeled Chemicals). 100 µL reactions were set up consisting of 50 mM KH_2PO_4 (pH 7.5), 12.5 mM $MgCl_2$, 1 mM *tris*(2-carboxyethyl)phosphine (TCEP), 8 µM *holo*-RavC, and 0.8 µM purified GilP or GilQ (see **Table 14**, reactions 1-30). This was allowed to equilibrate at 30 ^{o}C for 5 min and then the reaction was initiated by the addition of 8 µM radiolabeled acyl-CoA. The reaction was quenched with trichloroacetic acid (TCA) after 30 seconds at 30 ^{o}C to a final concentration of 7% TCA. 200 µg of bovine serum albumin (BSA) was added to the quenched reaction and was kept on ice for 15 min. Protein was collected by centrifugation (13000 x g, 10 min) and washed twice with cold 7% TCA. The pellet was

then dissolved with 100 µL of a 1:1 solution of 2 M NaOH and 2 M Tris Base. The reconstituted pellet was then combined with 500 µL of scintillation cocktail (Research Products International) and analyzed on a liquid scintillation analyzer (2200CA TRI-CARB, Packard).

GilP and RavC self loading assay conditions

The self loading study of GilP and RavC followed the same reaction conditions as the acyl transfer assay discussed above. However, instead of a single 30 second time point as discussed above, the self loading reactions (**Table 14**, reactions 33-33) were quenched at 0 sec, 10 sec, 30 sec, 1 min, 3 min, 5 min, 10 min, 30 min and 60 min. The ^{14}C-labeled proteins were collected and quantitated by liquid scintillation counting as described above. This data was then used to calculate the initial rate of self loading by GilP and RavC.

Mutant activity assay conditions

Activity of the mutant proteins, GilPS_{90}A, GilQS_{111}A, and RavCS_{39}A, were tested using the same reaction conditions as described above utilizing different combinations of wild type and mutant proteins (see **Table 16**).

Chapter 5: Production of Proposed Gilvocarcin Intermediates

Assembly of partial gilvocarcin biosynthetic gene clusters

Biosynthetic investigations into the enzymatic cascade leading to gilvocarcin V biosynthesis has primarily been elucidated through cosG9B3 based gene inactivations and complementation experiments. These experiments, combined with bioinformatical analyses, have provided evidence that allows us to predict the biosynthetic steps leading to gilvocarcin V production (**Figure 16**). However, despite rigorous experimentation, almost the entire post-PKS set of reactions remain unclear. The only exception is GilR, an oxidoreductase proven through *in vitro* experimentation to oxidize the hemi-acetal of pregilvocarcins (**75-77**) to the lactone in gilvocarcins (**49-51**).[101]

The lack of isolable substrates is the most prominent factor limiting the characterization of a majority of the post-PKS reactions. It was recently shown that all gilvocarcin enzymes believed to be involved in gilvocarcin post-PKS reactions, except GilGT, could be expressed as soluble proteins in *E. coli* (unpublished results by M. Kharel). In order to identify new intermediates from late stage biosynthetic reactions, and to further confirm the function of specific enzymes, a plasmid based approach was utilized in which partial gilvocarcin clusters could be assembled in various combinations to investigate unknown reactions, isolate unknown intermediates and provide substrates for further *in vitro* analyses.

Furthermore, this plasmid based approach may be used to create more stable gilvocarcin mutants. For example, the GilU mutant, cosG9B3-GilU$^-$, produces the most active gilvocarcin analogue to date, 4′-hydroxy-gilvocarcin V (**81**). Unfortunately, *S. lividans* TK24/cosG9B3-GilU$^-$ produces **81** in yields of 0.5 mg/L.[85] A synthetic construct could be assembled in which all genes responsible for **81** production are placed under the control of a strong promoter, such as ErmE*p, to increase the overall yield of this important analogue. Additionally, the production profile of *S. lividans* TK24/cosG9B3-GilGT$^-$ is notoriously unpredictable. This has created difficulties in further glycodiversification studies of gilvocarcin V (as discussed in section 2). Using this approach, it would be possible to place all genes needed for defuco-gilvocarcin V (**78**) production and omit all deoxysugar biosynthetic genes. This would provide an ideal host

for introducing exogenous glycosyltransferases and deoxysugars towards creating sugar derivatives of **49**.

Experimental design

In **specific aim 4**, genes from a large operon in cosG9B3, containing *gilH*, *OI*, *G*, *A*, *B*, *C*, *F*, *K*, *OIV*, *P* and *Q*, will be subcloned into an integrating vector (pSET152) and a self replicating expression vector (pUWL201PW) to determine their ability to produce dehydro-rabelomycins (**64** and **65**). The ErmE*p promoter will be incorporated into these plasmids to ensure strong expression of the biosynthetic genes. The placement and number of *ermE**p copies used will also be investigated to determine their optimal configuration for metabolite production. Once a successful construct is identified, additional post-PKS enzymes, *gilOII*, *M*, *MT*, *R* and *OIII* will be added to the base construct in various combinations and their metabolite production will be analyzed via HPLC-MS as describe previously.

Results

Metabolite production of partial cluster constructs

Analysis of the gilvocarcin gene cluster revealed that *gilG*/*gilA*, *gilA*/*gilB*, *gilC*/*gilF* and *gilOIV*/*gilP* share stop and start codons (**Figure 62**, red arrows). In order to retain the natural nucleotide sequence integrity of the gilvocarcin cluster, PCR was used to amplify *gilH-OI*, *gilG-A-B*, *gilC-F-K* and *gilOIV-P-Q* as four large fragments, allowing the overlapping ORFs to remain unaltered. Assembly of these fragments into pSET152 resulted in the addition of only 6 nucleotides between each fragment. This created only 18 nucleotide changes, dispersed among 3 positions, when compared to the natural sequence in cosG9B3.

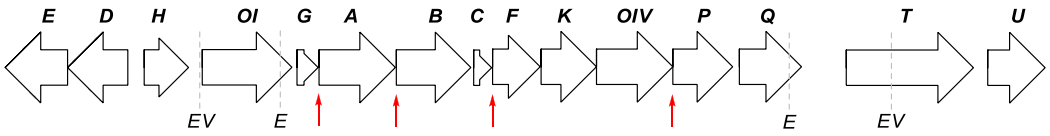


Figure 62. Partial gilvocarcin cluster highlighted with important features utilized in this study. *Eco*RV (*EV*) and *Eco*RI (*E*) sites are shown by the dotted line, and the red arrow indicates ORF overlaps.

Two pSET152 based constructs (pSET-GCOIV and pSET-HGCOIV) were created using the PCR fragments discussed above (**Figure 63)**. pSET152 carries the integration function and attachment site of ΦC31, and allows for site specific integration of the plasmid into the *attB* site of *S. lividans* TK24 chromosome.[121] Integration is advantageous because it produces a much more stable host than using a self replicating plasmid system. Both pSET-GCOIV and pSET-HGCOIV were transformed into *S. lividans TK24* through conjugation, and integration was confirmed through PCR. The resulting strains, *S. lividans* TK24/pSET-GCOIV and *S. lividans* TK24/pSET-HGCOIV were both unable to produce any gilvocarcin intermediate. Theoretically, both strains contain genes that should be able to produce 2,3-dehydro-UWM6 compounds (**67** and **68**) that will then be converted to dehydro-rabelomycins (**64** and **65**) spontaneously or through the activity of GilOI. The exact promoter regions of the *gil* cluster are unknown. pSET-GCOIV was tested without the addition of *ermE**p to see if a natural promoter contained within the cloned sequence or a promoter downstream of the *attB* site of *S. lividans* TK24 could induce transcription of the biosynthetic genes. It was therefore not surprising that pSET-GCOIV did not produce gilvocarcin metabolites. Alternatively, genes included in pSET-HGCOIV were cloned downstream of the strong constitutive promoter, *ermE**p. Successful integration and expression of this plasmid should have resulted in the formation of **64** and **65**.

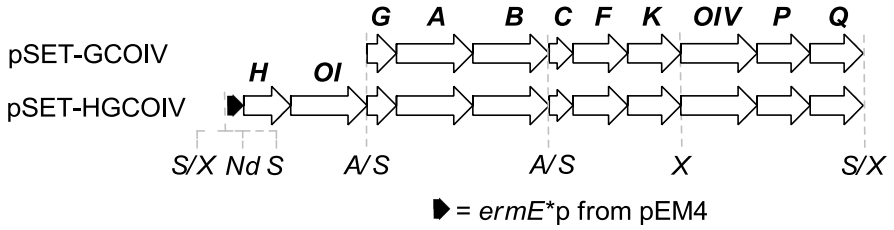


Figure 63. Gene placement and restriction sites used in creating pSET-GCOIV and pSET-HGCOIV. *Spe*I (*S*), *Xba*I (*X*), *Nde*I (*Nd*) and *Avr*II (*A*) sites are shown by the dotted line. Broken restriction sites are illustrated with a back slash.

An alternative cloning strategy was taken in which natural occurring restriction sites from cosG9B3 were utilized in conjunction with PCR to clone a completely unaltered fragment of cosG9B3covering *gilH* through *gilQ*. The sequence of *gilQ* was not

complete as the natural restriction site used in this study removed the last six nucleotides of its gene sequence (**Figure 62**). As discussed earlier, GilQ is involved in propionate incorporation during early gilvocarcin V biosynthesis. The absence of GilQ activity will only result in a loss or reduction of propionate primed intermediates produced by the engineered constructs.

The previous constructs were based on the integration vector pSET152. To test the effectiveness of self replicating plasmids, pUWL201PW was chosen for two reasons. First, pUWL201PW contains a copy of *ermE**p and an engineered RBS removing the need to clone the natural RBS; and secondly, the multiple cloning site (MCS) of pUWL201PW offered the appropriate restriction sites needed to design the cloning of several genes. The base construct, pSGC-HQ, was designed with a single *ermE**p promoter upstream of the large *gilH, OI, G, A, B, C, F, K, OIV, P* and *Q* fragment, and a second *ermE**p directly upstream of the first promoter (**Figure 64**). The cloning strategy was to introduce additional genes between the two promoters to ensure strong promotion of all genes. In this context *gilOII*, *gilM*, *gilMT* and *gilR* were cloned between these two promoters to create the constructs shown in **Figure 64**.

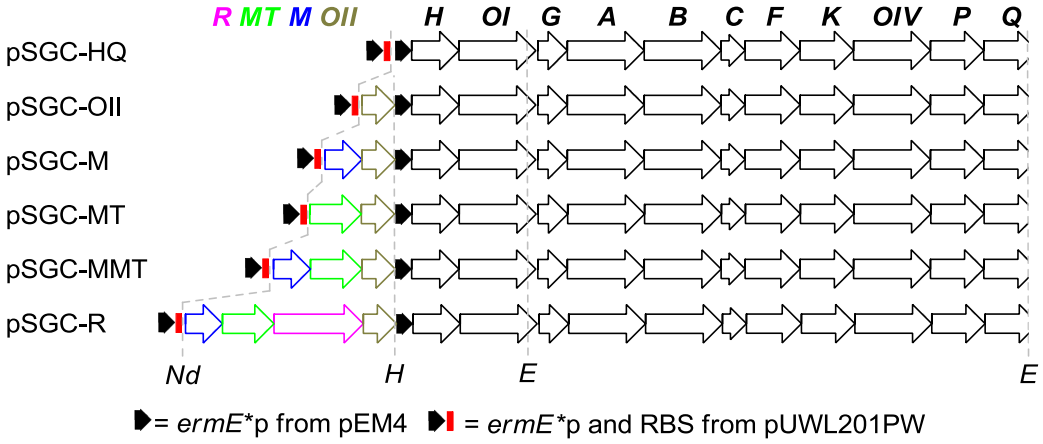


Figure 64. Gene placement and restriction sites used in creating pSGC constructs. *Eco*RI (*E*), *Hind*III (*H*) and *Nde*I (*Nd*) sites are shown by the dotted line.

Fermentation of *S. lividans* TK24/pSGC-HQ resulted in the accumulation of metabolites believed to be dehydro-rabelomycin (**64**) and dehydro-homo-rabelomycin (**65**) (**Figure 65**). This determination was made based on identical retention times and

UV-Vis spectra of the compounds accumulated in *S. lividans* TK24/pSGC-HQ compared to dehydro-rabelomycins (**64-66**). This result was exciting as it confirmed the successful expression of genes contained within the base pSGC-HQ construct. The production of both **64** and **65** was surprising. The production of such high quantities of a propionate primed intermediate compared to the acetate primed intermediate indicates the presence of a functional GilQ protein. It is possible that an in-frame stop codon is found downstream of the restriction site in which GilQ was cloned. If this in-frame stop codon is relatively close to this restriction site, the translated protein of the modified *gilQ* ORF may remain functional explaining the presence of propionate primed dehydro-homo-rabelomycin (**65**). Most fascinating, however, is the possible production of dehydro-rabelomycin V (**66**) in the absence of GilOIII. The oxidation of the ethyl side chain to form the vinyl functional group is known to involve GilOIII.[42] Furthermore, the inability of the GilOIII mutant to produce vinyl side chain congeners in *S. lividans* TK24 indicates endogenous host enzymes can not complement the reaction. To confirm that *gilOIII* was not accidentally incorporated into the construct, PCR was used in an attempt to amplify *gilOIII* from pSGC-HQ. This experiment resulted in no PCR fragments suggesting *gilOIII* was not present in the construct (data not shown). The elution pattern and intensity of peaks 1-3 in **Figure 65** are identical to that of expected M, V and E congeners, respectively. However, based on the lack of mass data to support the presence of a vinyl congener, it must be concluded that most likely peak two in **Figure 65** is actually dehydro-homo-rabelomycin and not dehydro-rabelomycin V. Furthermore, the UV-Vis of peak three can be explained by the overlap between peaks two and three, which may allow for the spectra of peak three to be similar to that of dehydro-homo-rabelomycin.

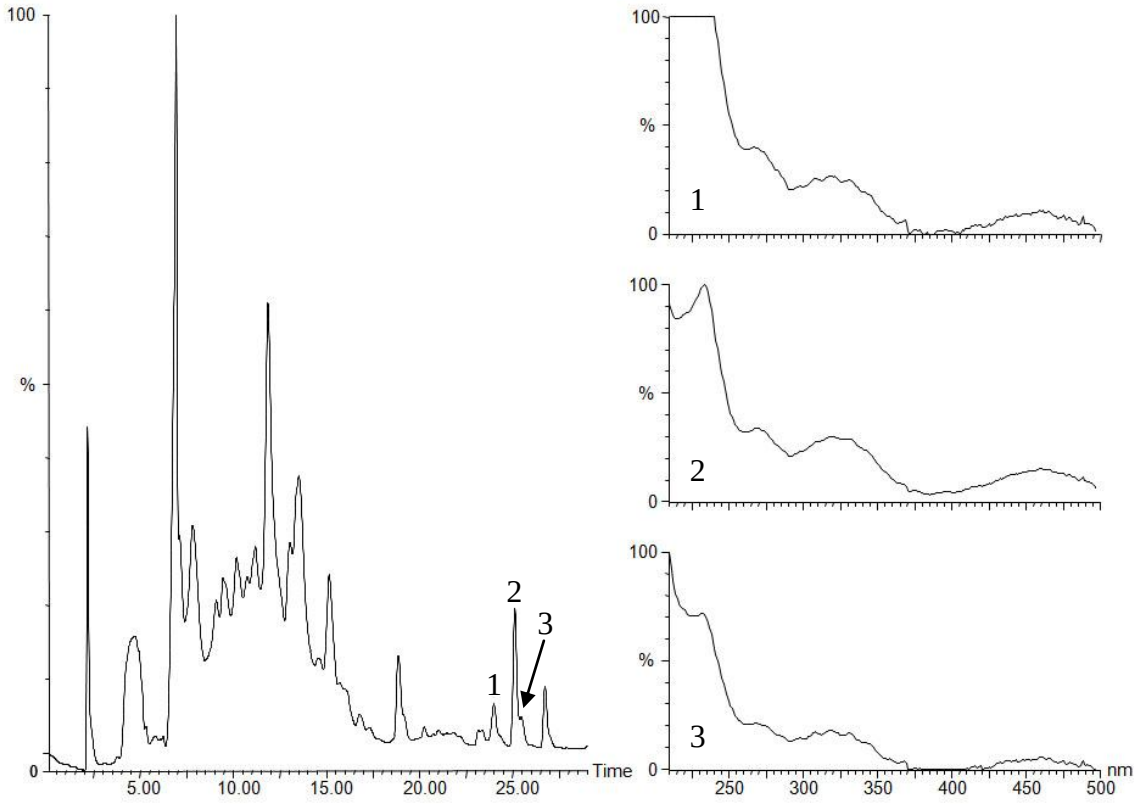


Figure 65. HPLC chromatogram trace of *S. lividans* TK24/pSGC-HQ including the UV-Vis spectrum of peaks 1, 2 and 3.

The successful production of gilvocarcin intermediates by *S. lividans* TK24/pSGC-HQ provided a working system in which additional post-PKS enzymes could be added to continue the biosynthesis toward gilvocarcin intermediates and/or defuco-gilvocarcin M and E. Unfortunately, the expression of pSGC-M, pSGC-MT, pSGC-MMT and pSGC-R in *S. lividans* TK24 did not produce observable amounts of any gilvocarcin intermediates. This suggested that the addition of genes to pSGC-HQ negatively affected the ability of the constructs to be expressed. For an unknown reason it seems that transcription is not proceeding as expected in constructs with genes inserted between the two promoters of pSGC-HQ. To try to eliminate any possible problems caused by including two promoters, an additional set of constructs were created in which only a single promoter from pUWL201PW was used to promote transcription of all downstream genes (**Figure 66**, pSGC* constructs). In addition, pSET-HQ was created in

which the same gene fragment found in pSGC-HQ was introduced into pSET152 (**Figure 66**). The successful integration and production of dehydro-rabelomycin from such a strain would be beneficial as this strain could be used as a host in which individual post-PKS genes could be introduced on smaller plasmid based gene constructs.

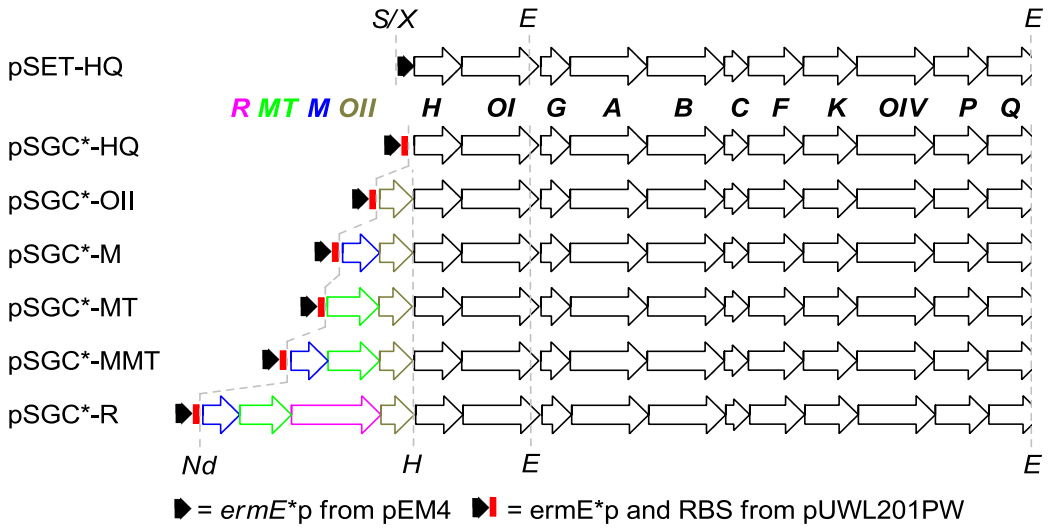


Figure 66. Gene placement and restriction sites used in creating pSET-HQ and the pSGC* constructs. *Spe*I (*S*), *Xba*I (*X*), *Nde*I (*Nd*), *Hind*III (*H*) and *Eco*RI (*E*) sites are shown by the dotted line. Broken restriction sites are illustrated with a back slash.

Unfortunately, all pSGC* constructs failed to express gilvocarcin intermediates in *S. lividans* TK24. Similarly, the integrated pSET-HQ strain was also unable to produce any gilvocarcin metabolites. From these results, it seems that the two *ermE**p system utilized previously was not contributing to the lack of transcription, however, the root cause remains unknown. In general, when cloning large consecutive genes it is typical to add additional genes downstream of the previous gene. This cloning strategy proved difficult for cloning the gilvocarcin genes due to a very limited amount of acceptable restriction sites that could be used. This resulted in previous approaches in which additional genes were added on the 5′ end of the fragment rather than the 3′ end.

In a final attempt to construct a partial gilvocarcin cluster that could be used to investigate gilvocarcin post-PKS reactions, a cloning strategy was designed for the addition of genes downstream of *gilQ*. The resulting base construct, pSGC**-HQ, included the entire unmodified nucleotide sequence of cosG9B3 from the start codon of

gilH through the stop codon of *gilQ*. The addition of *gilOII* to pSGC**-HQ produced pSGC**-OII and both of these constructs were transformed into *S. lividans* TK24 to screen for metabolite production (**Figure 67**). Again, both strains failed to produce gilvocarcin intermediates.

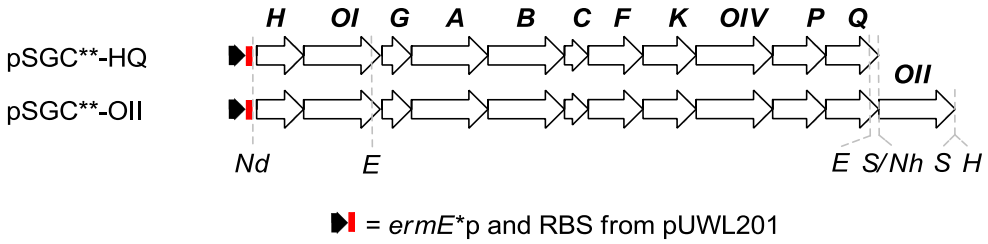


Figure 67. Gene placement and restriction sites used in creating pSGC** constructs. *Spe*I (*S*), *Eco*RI (*E*), *Nde*I (*Nd*), *Nhe*I (*Nh*) and *Hind*III (*H*) sites are shown by the dotted line. Broken restriction sites are illustrated with a back slash.

Discussion

A plasmid based approach was taken to assemble a minimal set of gilvocarcin genes responsible for producing dehydro-rabelomycins (**64** and **65**). This metabolite was chosen as all the genes needed for this biosynthetic product are conserved to large operon in the gilvocarcin cluster. This operon was cloned using several approaches and cloning strategies to produce six different constructs that could serve as the foundation for additional studies (**Figures 63**, **64**, **66** and **67**). In addition, the post-PKS genes, *gilM*, *gilMT*, *gilOII* and *gilR* were added to these base constructs in various combinations in order to investigate GilM and GilMT function, isolate new gilvocarcin intermediates as well as to produce a stable defuco-gilvocarcin producer.

One construct, pSGC-HQ, was able to successfully produce dehydro-rabelomycin, however, addition of post-PKS enzymes resulted in a failure to produce gilvocarcin associated metabolites in all cases despite altering the cloning strategy, promoter position, number of promoter copies and gene placement. The lack of production from these constructs suggest a problem with transcription, however it remains unknown why pSGC-HQ could be successfully transcribed and the similar construct pSET-HQ could not. Furthermore, it is surprising that pSGC-HQ worked well using the double *ermE**p system while the remaining pSGC constructs failed to produce

any gilvocarcin metabolites. Further experimentation using real-time qRT-PCR could determine if the problem truly lies with the *ermE**p promoter and transcription. Alternatively, a resistance gene could be cloned downstream of *gilQ* in pSGC**-HQ and the constructs ability to transmit resistance in *S. lividans* TK24 could be investigated. If the strain showed resistance it would suggest the genes downstream of *ermE**p are being transcribed properly and the problems encountered in this study are probably due to the inactivity of individual gilvocarcin proteins and not from transcription.

Materials and Methods

Bacterial strains and culture conditions

All transformation and culturing conditions were carried out as described in section 2.1: Bacterial strains, culture conditions and plasmids. A comprehensive list of strains and plasmids used in this study can be found in **Table 19**.

Table 19. Strains and plasmids used in the construction of partial gilvocarcin clusters.

Strain/Plasmid	Characteristics and relevance	References
E. coli XL1Blue-MRF	Cloning host	Stratagene
PCR-Blunt II-TOPO	PCR fragment cloning vector	Invitrogen
pSET152	*Streptomyces* integration vector	Kieser, T et al.[121]
pEM4	Used for cloning *ermE**p	Quiros, L. et al.[112]
pUWL201PW	*Streptomyces* expression vector	Doumith. M et al.[168]
pET28a(+)	Used as a cloning vector	Novagen
pBluescript II SK(+)	Used as a cloning vector	Fermentas
cosG9B3	Template for *gil* genes	Fischer, C. et al.[91]
TOPO-OIV	TOPO with *gilOIV,P,Q*	This study
TOPO-CFK	TOPO with *gilC,F,K*	This study
TOPO-GAB	TOPO with *gilG,A,B*	This study
TOPO-HOI	TOPO with *gilH,OI*	This study
pEM4-HOI	pEM4 with *gilH,OI*	This study
TOPO-2-HOI	TOPO with *ermE**p,*gilH,OI*	This study
pSET-OIV	pSET152 with *gilOIV,P,Q*	This study
pSET-COIV	pSET152 with *gilC,F,K,OIV,P,Q*	This study
pSET-GCOIV (1)[a]	pSET152 with *gilG,A,B,C,F,K,OIV,P,Q*	This study
pSET-HGCOIV (2)[a]	pSET152 with *ermE**p,*gilH,OI,G,A,B,C,F,K,OIV,P,Q*	This study
TOPO-OII	TOPO with *gilOII*	Kharel, M et al.[c]
TOPO-M	TOPO with *gilM*	Kharel, M et al.[c]
TOPO-MT	TOPO with *gilMT*	Kharel, M et al.[c]
TOPO-R	TOPO with *gilR*	Kharel, M et al.[c]
pET-OII	pET28a with *gilOII*	This study
pET-OIIM	pET28a with *gilOII,M*	This study
pET-OIIMT	pET28a with *gilOII,MT*	This study
pET-OIIMMT	pET28a with *gilOII,M,MT*	This study
pET-OIIMMTR	pET28a with *gilOII,M,MT,R*	This study
pBlu-G9B3-EV	pBluescript with partial *gil* cluster (*Eco*RV)	This study
pUWL-HOI	pUWL201 with *ermE**p,*gilH,OI*	This study
pSGC-HQ (3)[a]	pUWL201 with *ermE**p,*gilH,OI,G,A,B,C,F,K,OIV,P,Q*	This study
pSGC-OII (4)[a]	pUWL201 with *gilOII,ermE**p,*gilH,OI,G,A,B,C,F,K,OIV,P,Q*	This study
pSGC-M (5)[a]	pUWL201 with *gilM,OII,ermE**p,*gilH,OI,G,A,B,C,F,K,OIV,P,Q*	This study
pSGC-MT (6)[a]	pUWL201 with *gilMT,OII,ermE**p,*gilH,OI,G,A,B,C,F,K,OIV,P,Q*	This study
pSGC-MMT (7)[a]	pUWL201 with *gilM,MT,OII,ermE**p,*gilH,OI,G,A,B,C,F,K,OIV,P,Q*	This study
pSGC-R (8)[a]	pUWL201 with *gilM,MT,R,OII,ermE**p,*gilH,OI,G,A,B,C,F,K,OIV,P,Q*	This study
pSET-HOI	pSET152 with *ermE**p,*gilH,OI*	This study
pSET-HQ (9)[a]	pSET152 with *ermE**p,*gilH,OI,G,A,B,C,F,K,OIV,P,Q*	This study
TOPO-3-HOI	TOPO with *gilH,OI*	This study
pUWL-2-HOI	pUWL201PW with *gilH,OI*	This study
pSGC*-HQ (10)[a]	pUWL201 with *gilH,OI,G,A,B,C,F,K,OIV,P,Q*	This study
pSGC*-OII (11)[a]	pUWL201 with *gilOII,H,OI,G,A,B,C,F,K,OIV,P,Q*	This study
pSGC*-M (12)[a]	pUWL201 with *gilM,OII,H,OI,G,A,B,C,F,K,OIV,P,Q*	This study
pSGC*-MT (13)[a]	pUWL201 with *gilMT,OII,H,OI,G,A,B,C,F,K,OIV,P,Q*	This study
pSGC*-MMT (14)[a]	pUWL201 with *gilM,MT,OII,H,OI,G,A,B,C,F,K,OIV,P,Q*	This study
pSGC*-R (15)[a]	pUWL201 with *gilM,MT,R,OII,H,OI,G,A,B,C,F,K,OIV,P,Q*	This study
TOPO-Q	TOPO with partial *gilQ*	This study
TOPO-4-HOI	TOPO with *gilH,OI*	This study
pET-Q	pET28a with partial *gilQ*	This study
pET-HQ	pET28a with partial *gilH,OI,Q*	This study
pET**-HQ	pET28a with *gilH,OI,G,A,B,C,F,K,OIV,P,Q*	This study
TOPO-2-OII	TOPO with *gilOII*	This study
pSGC**-HQ (16)[a]	pUWL201PW with *gilH,OI,G,A,B,C,F,K,OIV,P,Q*	This study
pSGC**-OII (17)[a]	pUWL201PW with *gilH,OI,G,A,B,C,F,K,OIV,P,Q,OII*	This study
SLTK24[b]	Heterologous expression host	Kieser, T. et al.[121]
SLTK24[b]/(1-17)[a]	SLTK24[b] transformed with (1-21)[a]	This study

[a]Numbering used only for table simplification, [b]*S. lividans* TK24, [c]Unpublished work by M. Kharel

DNA isolation, DNA manipulation and transformations

Plasmid DNA isolations, manipulations and PCR reactions were carried out as described in section 2.1: DNA isolation, DNA manipulation and PCR. Please find **Table 20** for a complete list of primers used in this study.

Table 20. Oligonucleotide sequence of primers used in creating partial gilvocarcin clusters. Restriction sites are represented by italicized sequences.

Primer	Oligonucleotide sequence (5′-3′)
GilOIVPQ-F	TT*ACTAGTCATATGTCTAGA*GTGGGGACATGACGGAGC
GilOIVPQ-R	TT*ACTAGTT*CAACAGAATTCCTCGGCGACCTCA
GilCFK-F	AA*CATATGACTAGT*GAGAAGGAGCAAGGAATG
GilCFK-R	AA*TCTAGA*TCACCTCTGCGACACGGCGG
GilGAB-F	TT*CATATGACTAGT*CCAGAGGAGGAAGAACCCTTGT
GilGAB-R	TT*CCTAGG*TCATTTCCCGACGATCAG
GilHOI-F	TT*ACTAGT*CTCTACCGCTCATAATCTGGCCGCT
GilHOI-R	AA*TCTAGACCTAGG*TTACGCGGCGGGCTGACC
GilHOI-pEM4-F	TT*CATATGACTAGT*CAGGAAACAGCTATGAC
GilOII-F	TT*CATATG*TC*GCTAGC*TC*GAATTC*GACACCTCATGGAAGGCC
GilOII-R	AG*AAGCTT*TCACGACGCGTACCCCTC
GilM-F	AT*ACTAGT*GAGAGGAGTGGCGCTCGATGCCAA
GilM-R	AT*GAATTC*ATT*GCTAGC*GAACTTCATGCCGGGCTCTCCGA
GilMT-F	AT*ACTAGT*GTCGAAGGGACATCATGACCATTACTGCATCGGG
GilMT-R	AT*GAATTC*ATT*GCTAGC*GTCACCGGCTGCGGGGAGAGC
GilR-F	AT*ACTAGT*GAGGGGTACGCGTCGTGACCGCTT
GilR-R	AT*GAATTC*ATT*GCTAGC*TCCTCTCAGAGTCCTATGGACAT
GilHOI-F2	TT*AAGCTT*CTCTACCGCTCATAATCTGGCCGCT
GilQ-F	TC*GGATCC*GGAGGAGCTCAAGGGTG
GilQ-R	*AAGCTT*ATC*ACTAGT*TGGTCAACAGAATTCCTC
GilHOI-F3	GG*CATATG*ATCAGGATCGCCGTC
GilOII-F2	TAT*GCTAGC*GGAAGGCCTCATGCCGATCAT
GilOII-R2	AGT*AAGCTT*GTA*ACTAGT*CGGTCACGACGCGTACCCCTC

Preparing partial cluster constructs

pSET constructs

The primer pairs GilOIVPQ-F + GilOIVPQ-R, GilCFK-F + GilCFK-R, GilGAB-F + GilGAB-R, and GilHOI-F + GilHOI-R were used to amplify fragments of cosG9B3 consisting of the complete genes *gilOIV-P-Q*, *gilC-F-K*, *gilG-A-B*, and *gilH-OI*, respectively. The *gilOIV-P-Q*, *gilC-F-K*, *gilG-A-B* and *gilH-OI* PCR fragments were amplified with engineered restriction sites *Spe*I/*Nde*I/*Xba*I, *Nde*I/*Spe*I, *Nde*I/*Spe*I and *Spe*I on their 5′ ends, respectively. The 3′ ends were also engineered with *Spe*I, *Xba*I, *Avr*II and *Avr*II/*Xba*I, respectively. Each PCR amplified fragment was cloned into the

PCR-Blunt TOPO-II cloning vector producing TOPO-OIV, TOPO-CFK, TOPO-GAB and TOPO-HOI. Initially, the *gilOI-P-Q* fragment was removed from TOPO-OIV as a *Spe*I fragment and ligated into pSET152 (*Xba*I). The resulting plasmid, pSET-OIV, was used as the foundation for adding additional gilvocarcin genes. TOPO-CFK was digested with *Nde*I/*Xba*I, removing *gilC-F-K*, and ligated into pSET-OIV (*Nde*I/*Xba*I) producing pSET-COIV. Similarly, TOPO-GAB digested with *Avr*II/*Xba*I, removing *gilG-A-B*, and ligated into pSET-COIV (*Nde*I/*Spe*I) producing pSET-GCOIV. In order to include the constitutively active promoter *ermE**p, TOPO-HOI was digested with *Spe*I/*Xba*I, removing *gilH-OI*, and ligated into pEM4 (*Xba*I). The correct orientation was confirmed by a *Hind*III/*Xba*I digest. This plasmid, pEM4-HOI, was then used as template for PCR. The primers GilHOI-pEM4-F and GilHOI-R were used to amplify *ermE**p, *gilH* and *gilOI* from pEM4-HOI introducing *Nde*I/*Spe*I and *Avr*II/*Xba*I sites at the 5′ and 3′ end, respectively. The resulting PCR fragment was cloned into TOPO to create TOPO-2-HOI. Finally, TOPO-2-HOI was digested with *Nde*I/*Avr*II and ligated into pSET-GCOIV (*Nde*I/*Spe*I) to produce the final construct, pSET-HGCOIV.

The construct pSET-HOI was created by removing *ermE**p, *gilH* and *gilOI* from TOPO-2-HOI (discussed below) as a *Spe*I/*Eco*RI fragment and ligating it into pSET152 (*Xba*I/*Eco*RI). The large *Eco*RI fragment from pBlu-G9B3-EV was then placed into pSET-HOI (*Eco*RI) creating pSET-HQ. The orientation of the *Eco*RI fragment was confirmed as described below.

pSGC constructs

The previously created construct pEM4-HOI was digested with *Hind*III/*Eco*RI and ligated into pUWL201PW (*Hind*III/*Eco*RI). The gene sequence of *gilOI* contains an *Eco*RI site near its stop codon, therefore the resulting plasmid, pUWL-HOI, contained the complete sequence of *ermE**p and *gilH*, but only the partial sequence of *gilOI*. A large *Eco*RV fragment from cosG9B3 including *gilOI, G, A, B, C, F, K, OIV, P, Q* and *T* was subcloned into pBluescript II SK(+) creating pBlu-G9B3-EV (see **Figure 63**). An *Eco*RI fragment was taken from pBlu-G9B3-EV, cuts inside *gilOI* and *gilQ*, and ligated into pUWL-HOI (*Eco*RI). The correct orientation was confirmed by amplifying a *gilOI-G* fragment and confirming the correct size. The resulting plasmid pSGC-HQ, contained

the complete gene sequences of *ermE**p, *gilH, OI, G, A, B, C, F, K, OIV* and *P* as well as partial sequence of *gilQ*. This plasmid was used to further build the gilvocarcin cluster.

The PCR primer pairs GilOII-F + GilOII-R, GilM-F + GilM-R, GilMT-F + GilMT-R and GilR-F + GilR-R were used to amplify *gilOII*, *gilM*, *gilMT* and *gilR* from cosG9B3, respectively. The amplified product of *gilOII* contained the restriction sites *Nde*I/*Nhe*I/*Eco*RI and *Hind*III at its 5′ and 3′ ends; while *gilM*, *gilMT* and *gilR* were amplified containing *Nhe*I and *Spe*I/*Eco*RI restriction sites at their 5′ and 3′ ends, respectively. The resulting PCR products were cloned into PCR-Blunt II-TOPO vector creating TOPO-OII, TOPO-M, TOPO-MT and TOPO-R constructs. Initially, TOPO-OII was digested with *Nde*I/*Hind*III and ligated into pET28a(+) (*Nde*I/*Hind*III) to produce pET-OII. This construct was used to build part of the gilvocarcin cluster before being added to the main construct. A *Nhe*I/*Eco*RI fragment was removed from TOPO-M, containing *gilM*, and ligated into pET-OII (*Nhe*I/*Eco*RI) forming pET-M. Similarly, TOPO-MT was digested with *Nhe*I/*Eco*RI and ligated into pET-OII (*Nhe*I/*Eco*RI) forming pET-MT. The same *Nhe*I/*Eco*RI fragment from TOPO-MT was also ligated into pET-M (*Nhe*I/*Spe*I) creating pET-MMT. TOPO-R was also digested with *Nhe*I/*Eco*RI and the resulting fragment was ligated into pET-MMT (*Nhe*I/*Spe*I) to produce pET-R. The final constructs were created by digesting pET-M, pET-MT, pET-MMT and pET-R with *Nde*I/*Hind*III and ligating them into pSGC-HQ (*Nde*I/*Hind*III) producing pSGC-OII, pSGC-M, pSGC-MMT and pSGC-R, respectively.

pSGC* constructs

To create constructs containing only a single copy of *ermE**p, the primer pair GilHOI-F2 + GilHOI-R was used to amplify *gilH* and *gilOI* with unique *Hind*III restriction placed at the 5′ end. As before, *gilOI* contains an *Eco*RI site within its nucleotide sequence, and this *Eco*RI site is used for further cloning. The PCR product was cloned into TOPO producing TOPO-3-HOI. The gilvocarcin genes were removed through *Hind*III/*Eco*RI digestion and placed into pUWL201PW (*Hind*III/*Eco*RI) producing pUWL-2-HOI. The large *Eco*RI fragment from pBlu-G9B3-EV was ligated into pUWL-2-HOI forming pSGC*-HQ. The orientation of the *Eco*RI fragment was confirmed as described above. *Nde*I/*Hind*III fragments taken from pET-OII, pET-M,

pET-MT, pET-MMT and pET-R (described above) were ligated into pSGC*-HQ to produce pSGC*-OII, pSGC*-M, pSGC*-MT, pSGC*-MMT and pSGC*-R.

pSGC constructs**

Flanking regions of the *gilQ* stop codon were amplified from cosG9B3 using primers GilQ-F and GilQ-R. The amplified product contained a *Bam*HI site and *Spe*I/*Hind*III sites at its 5′ and 3′ prime ends, respectively. In pSGC and pSGC* constructs, only the partial *gilQ* sequence was included. This approach allows for the complete gene sequence of *gilQ* to be included in the engineered gilvocarcin cluster. The amplified PCR product was cloned into TOPO to produce TOPO-GilQ. This small fragment was removed from TOPO-GilQ and placed into pET28a(+) using *Bam*HI/*Hind*III creating pET-Q. The primer pair GilHOI-F3 + GilHOI-R was used to amplify *gilH* and *gilOI* from cosG9B3 with an *Nde*I site upstream of its natural RBS. Again, the natural *Eco*RI site within *gilOI* was used in further cloning steps. The resulting PCR product was cloned into TOPO producing TOPO-4-HOI. This construct was digested with *Nde*I/*Eco*RI, removing *gilH* and *gilOI*, and ligated into pET-Q (*Nde*I/*Eco*RI) forming pET-HQ. The large *Eco*RI fragment from pBlu-G9B3-EV was ligated into pET-HQ (*Eco*RI) creating a complete fragment, identical to cosG9B3, covering the genes *gilH, OI, G, A, B, C, F, K, OIV, P* and *Q*. This construct, pSGC**-HQ was used as the foundation for further expansion of the gilvocarcin cluster. The primer pair GilOII-F2 + GilOII-R2 was used to amplify *gilOII* from cosG9B3. The resulting PCR product included *Nhe*I and *Spe*I/*Hind*III sites at the 5′ and 3′ prime ends, respectively; and was cloned into TOPO vector. The resulting construct was designated TOPO-2-OII and was digested with *Nhe*I/*Hind*III and ligated into pSGC**-HQ (*Spe*I/*Hind*III) producing pSGC**-OII.

Chapter 6: Summary

In summary, the studies reported in this dissertation are aimed at producing glycodiversified analogues of gilvocarcin-like aryl *C*-glycosides and understanding the biosynthetic propensity towards propionate starter unit incorporation during gilvocarcin V (**49**) biosynthesis. The insights gained from these studies provide the framework for further combinatorial biosynthetic investigations to produce gilvocarcin analogues.

In **specific aim 1a**, we utilized the gene sequences from *gilGT*, *chryGT* and *ravGT* to create six rationally designed chimeric GTs with improved donor substrate flexibility compared to their wild type activity. This led to the identification of three functional chimeric *C*-glycosyltransferases able to produce **49**. Attempts to test the donor substrate flexibility of the chimeric GTs were inhibited by the lack of a suitable host, but preliminary results showed promising results for their use in producing gilvocarcin analogues with altered sugar moieties.

In **specific aim 1b**, the polycarcin V (**61**) rhamnosyltransferase, *plcGT*, was isolated from *S. polyformus* genomic DNA for its proposed ability to transfer both furanose and pyranose sugars. This in addition to being a rhamnosyltransferase, which are in general donor substrate promiscuous, made *plcGT* a valuable resource for further glycodiversification studies. Preliminary investigation into the activity as well as donor substrate promiscuity revealed that PlcGT was likely involved in transferring both furanose and pyranose sugars but was rather substrate specific.

In **specific aim 2**, the L-rhamnose moiety of **61** was modified using a unique combinatorial biosynthetic approach in which rhamnosyl-*O*-methyltransferases from the steffimycin and elloramycin pathways were used to produce a library of five *O*-methylated-L-rhamnose analogues of polycarcin. This library provides a unique opportunity for SAR studies regarding specific residues of the sugar moiety as well as providing additional analogues with possible increased pharmacological properties.

In **specific aim 3a** and **3b**, the acyltransferases GilP and GilQ were investigated to determine their role in starter unit specificity. The biosynthesis of **49** requires the incorporation of a propionate starter unit which is atypical of a type II PKS system. Furthermore, the concomitant production of **49** and **50** during gilvocarcin fermentation illustrates an ability to also initiate polyketide biosynthesis with acetate. *In vivo* and *in*

vitro characterization of GilP and GilQ revealed the primary determining factor for propionate incorporation during the biosynthesis of gilvocarcin V was GilQ. This led to genetically engineered strains producing 10-20 times more gilvocarcin V compared to the wild type strains.

In **specific aim 4**, a plasmid based approach to organize a minimal set of gilvocarcin genes to produce pathway intermediates was undertaken. The goal of this study was initially to design a plasmid that would facilitate further gene incorporation and could produce the early shunt products dehydro-rabelomycin (**64**) and dehydro-homo-rabelomycin (**65**). This would allow individual genes to be added to the growing plasmid to both evaluate genes role in gilvocarcin biosynthesis, but also allow for pathway intermediates to be isolated. Furthermore, the successful preparation of these plasmids would allow for their use in complementation experiments and would provide an ideal system for glycodiversification studies. A single construct, pSGC*-HQ, was identified that could produce **64** and **65**; however, the incorporation of additional genes resulted in the abolishment of metabolite production.

REFERENCES

(1) Capasso, L. 5300 years ago, the Ice Man used natural laxatives and antibiotics. *Lancet* **1998**, *352*, 1864.

(2) Dewick, P. *Medicinal Natural Products*; Second Edition ed.; John Wiley & Sons, Ltd, 2007.

(3) Berdy, J. Bioactive microbial metabolites. *J. Antibiot. (Tokyo)* **2005**, *58*, 1-26.

(4) Hertweck, C. The Biosynthetic Logic of Polyketide Diversity. *Angew. Chem. Int. Edit.* **2009**, *48*, 4688-716.

(5) O'Hagan, D. *The Polyketides Metabolites, Ellis Horwood, Chichester*. **1991**.

(6) Rix, U.; Fischer, C.; Remsing, L. L.; Rohr, J. Modification of post-PKS tailoring steps through combinatorial biosynthesis. *Nat. Prod. Rep.* **2002**, *19*, 542-80.

(7) Olano, C.; Mendez, C.; Salas, J. A. Post-PKS tailoring steps in natural product-producing actinomycetes from the perspective of combinatorial biosynthesis. *Nat. Prod. Rep.* **2010**, *27*, 571-616.

(8) Smith, S.; Tsai, S. C. The type I fatty acid and polyketide synthases: a tale of two megasynthases. *Nat. Prod. Rep.* **2007**, *24*, 1041-72.

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

(10) Hopwood, D. A. Genetic contributions to understanding polyketide synthases. *Chem. Rev.* **1997**, *97*, 2465-97.

(11) Fischbach, M. A.; Walsh, C. T. Assembly-line enzymology for polyketide and nonribosomal Peptide antibiotics: logic, machinery, and mechanisms. *Chem. Rev.* **2006**, *106*, 3468-96.

(12) Xue, Y.; Zhao, L.; Liu, H. W.; Sherman, D. H. A gene cluster for macrolide antibiotic biosynthesis in Streptomyces venezuelae: architecture of metabolic diversity. *Proc. Natl. Acad. Sci. USA* **1998**, *95*, 12111-6.

(13) Xue, Y. Q.; Wilson, D.; Sherman, D. H. Genetic architecture of the polyketide synthases for methymycin and pikromycin series macrolides. *Gene* **2000**, *245*, 203-11.

(14) Xue, Y.; Sherman, D. H. Biosynthesis and combinatorial biosynthesis of pikromycin-related macrolides in Streptomyces venezuelae. *Metab. Eng.* **2001**, *3*, 15-26.

(15) Wenzel, S. C.; Muller, R. Formation of novel secondary metabolites by bacterial multimodular assembly lines: deviations from textbook biosynthetic logic. *Curr. Opin. Chem. Biol.* **2005**, *9*, 447-58.

(16) Campbell, C. D.; Vederas, J. C. Biosynthesis of lovastatin and related metabolites formed by fungal iterative PKS enzymes. *Biopolymers* **2010**, *93*, 755-63.

(17) Ma, S. M.; Li, J. W.; Choi, J. W.; Zhou, H.; Lee, K. K.; Moorthie, V. A.; Xie, X.; Kealey, J. T.; Da Silva, N. A.; Vederas, J. C.; Tang, Y. Complete reconstitution of a highly reducing iterative polyketide synthase. *Science* **2009**, *326*, 589-92.

(18) Sandmann, A.; Dickschat, J.; Jenke-Kodama, H.; Kunze, B.; Dittmann, E.; Muller, R. A Type II polyketide synthase from the gram-negative Bacterium Stigmatella aurantiaca is involved in Aurachin alkaloid biosynthesis. *Angew. Chem. Int. Edit.* **2007**, *46*, 2712-6.

(19) Brachmann, A. O.; Joyce, S. A.; Jenke-Kodarna, H.; Schwar, G.; Clarke, D. J.; Bode, H. B. A type II polyketide synthase is responsible for anthraquinone biosynthesis in *Photorhabdus luminescens*. *Chembiochem* **2007**, *8*, 1721-8.

(20) Carreras, C. W.; Khosla, C. Purification and in vitro reconstitution of the essential protein components of an aromatic polyketide synthase. *Biochemistry* **1998**, *37*, 2084-8.

(21) Bao, W. L.; Wendt-Pienkowski, E.; Hutchinson, C. R. Reconstitution of the iterative type II polyketide synthase for tetracenomycin F2 biosynthesis. *Biochemistry* **1998**, *37*, 8132-8.

(22) Tang, Y.; Tsai, S. C.; Khosla, C. Polyketide chain length control by chain length factor. *J. Am. Chem. Soc.* **2003**, *125*, 12708-9.

(23) Shen, B.; Hutchinson, C. R. Enzymatic synthesis of a bacterial polyketide from acetyl and malonyl coenzyme A. *Science* **1993**, *262*, 1535-40.

(24) Shen, B.; Hutchinson, C. R. Tetracenomycin F2 cyclase: intramolecular aldol condensation in the biosynthesis of tetracenomycin C in Streptomyces glaucescens. *Biochemistry* **1993**, *32*, 11149-54.

(25) Thompson, T. B.; Katayama, K.; Watanabe, K.; Hutchinson, C. R.; Rayment, I. Structural and functional analysis of tetracenomycin F2 cyclase from Streptomyces glaucescens. A type II polyketide cyclase. *J. Biol. Chem.* **2004**, *279*, 37956-63.

(26) Summers, R. G.; Wendtpienkowski, E.; Motamedi, H.; Hutchinson, C. R. Nucleotide sequence of the *tcmII-tcmIV* region of the tetracenomycin C biosynthetic gene cluster of *Streptomyces glaucescens* and evidence that the *tcmN* gene encodes a multifunctional cyclase-dehydratase-O-methyl transferase. *J. Bacteriol.* **1992**, *174*, 1810-20.

(27) Shen, B.; Hutchinson, C. R. Tetracenomycin F1 monooxygenase: oxidation of a naphthacenone to a naphthacenequinone in the biosynthesis of tetracenomycin C in Streptomyces glaucescens. *Biochemistry* **1993**, *32*, 6656-63.

(28) Shen, B.; Hutchinson, C. R. Triple hydroxylation of tetracenomycin A2 to tetracenomycin C in Streptomyces glaucescens. Overexpression of the tcmG gene in Streptomyces lividans and characterization of the tetracenomycin A2 oxygenase. *J. Biol. Chem.* **1994**, *269*, 30726-33.

(29) Tsai, S. C. A fine balancing act of type III polyketide synthase. *Chem. Biol.* **2004**, *11*, 1177-8.

(30) Magarvey, N. A.; Beck, Z. Q.; Golakoti, T.; Ding, Y.; Huber, U.; Hemscheidt, T. K.; Abelson, D.; Moore, R. E.; Sherman, D. H. Biosynthetic characterization and chemoenzymatic assembly of the cryptophycins. Potent anticancer agents from cyanobionts. *Chem. Biol.* **2006**, *1*, 766-79.

(31) Staunton, J.; Wilkinson, B. Biosynthesis of Erythromycin and Rapamycin. *Chem. Rev.* **1997**, *97*, 2611-30.

(32) Zhao, L.; Beyer, N. J.; Borisova, S. A.; Liu, H. W. Beta-glucosylation as a part of self-resistance mechanism in methymycin/pikromycin producing strain Streptomyces venezuelae. *Biochemistry* **2003**, *42*, 14794-804.

(33) Vilches, C.; Hernandez, C.; Mendez, C.; Salas, J. A. Role of glycosylation and deglycosylation in biosynthesis of and resistance to oleandomycin in the producer organism, *Streptomyces antibioticus*. *J. Bacteriol.* **1992**, *174*, 161-5.

(34) Lairson, L. L.; Henrissat, B.; Davies, G. J.; Withers, S. G. Glycosyltransferases: structures, functions, and mechanisms. *Annu. Rev. Biochem.* **2008**, *77*, 521-55.
(35) Hong, J. S.; Park, S. J.; Parajuli, N.; Park, S. R.; Koh, H. S.; Jung, W. S.; Choi, C. Y.; Yoon, Y. J. Functional analysis of desVIII homologues involved in glycosylation of macrolide antibiotics by interspecies complementation. *Gene* **2007**, *386*, 123-30.
(36) Minami, A.; Uchida, R.; Eguchi, T.; Kakinuma, K. Enzymatic approach to unnatural glycosides with diverse aglycon scaffolds using glycosyltransferase VinC. *J. Am. Chem. Soc.* **2005**, *127*, 6148-9.
(37) Zhang, C. S.; Griffith, B. R.; Fu, Q.; Albermann, C.; Fu, X.; Lee, I. K.; Li, L. J.; Thorson, J. S. Exploiting the reversibility of natural product glycosyltransferase-catalyzed reactions. *Science* **2006**, *313*, 1291-4.
(38) Thibodeaux, C. J.; Melancon, C. E.; Liu, H. W. Natural-product sugar biosynthesis and enzymatic glycodiversification. *Angew. Chem.-Int. Ed.* **2008**, *47*, 9814-59.
(39) Menendez, N.; Nur-e-Alam, M.; Brana, A. F.; Rohr, J.; Salas, A. F.; Mendez, C. Biosynthesis of the antitumor chromomycin A(3) in *Streptomyces griseus*: Analysis of the gene cluster and rational design of novel chromomycin analogs. *Chem. Biol.* **2004**, *11*, 21-32.
(40) Menendez, N.; Nur-e-Alam, M.; Fischer, C.; Brana, A. F.; Salas, J. A.; Rohr, J.; Mendez, C. Deoxysugar transfer during chromomycin A3 biosynthesis in Streptomyces griseus subsp. griseus: new derivatives with antitumor activity. *Appl. Environ. Microb.* **2006**, *72*, 167-77.
(41) Durr, C.; Hoffmeister, D.; Wohlert, S. E.; Ichinose, K.; Weber, M.; von Mulert, U.; Thorson, J. S.; Bechthold, A. The glycosyltransferase UrdGT2 catalyzes both C- and O-glycosidic sugar transfers. *Angew. Chem. Int. Edit.* **2004**, *43*, 2962-5.
(42) Liu, T.; Kharel, M. K.; Fischer, C.; McCormick, A.; Rohr, J. Inactivation of gilGT, encoding a C-glycosyltransferase, and gilOIII, encoding a P450 enzyme, allows the details of the late biosynthetic pathway to gilvocarcin V to be delineated. *Chembiochem* **2006**, *7*, 1070-7.
(43) Bililign, T.; Shepard, E. M.; Ahlert, J.; Thorson, J. S. On the origin of deoxypentoses: evidence to support a glucose progenitor in the biosynthesis of calicheamicin. *Chembiochem* **2002**, *4*, 1143-6.
(44) Aparicio, J. F.; Caffrey, P.; Gil, J. A.; Zotchev, S. B. Polyene antibiotic biosynthesis gene clusters. *Appl. Microbiol. Biot.* **2003**, *61*, 179-88.
(45) Thibodeaux, C. J.; Melancon, C. E.; Liu, H. W. Unusual sugar biosynthesis and natural product glycodiversification. *Nature* **2007**, *446*, 1008-16.
(46) Rodriguez, L.; Oelkers, C.; Aguirrezabalaga, I.; Brana, A. F.; Rohr, J.; Mendez, C.; Salas, J. A. Generation of hybrid elloramycin analogs by combinatorial biosynthesis using genes from anthracycline-type and macrolide biosynthetic pathways. *J. Mol. Microb. Biotech.* **2000**, *2*, 271-6.
(47) Perez, M.; Baig, I.; Brana, A. F.; Salas, J. A.; Rohr, J.; Mendez, C. Generation of new derivatives of the antitumor antibiotic mithramycin by altering the glycosylation pattern through combinatorial biosynthesis. *Chembiochem* **2008**, *9*, 2295-304.

(48) Shepherd, M. D.; Liu, T.; Mendez, C.; Salas, J. A.; Rohr, J. Engineered Biosynthesis of Gilvocarcin Analogues with Altered Deoxyhexopyranose Moieties. *Appl. Environ. Microb.* **2011**, *77*, 435-41.

(49) Olano, C.; Abdelfattah, M. S.; Gullon, S.; Brana, A. F.; Rohr, J.; Mendez, C.; Salas, J. A. Glycosylated derivatives of steffimycin: Insights into the role of the sugar moieties for the biological activity. *Chembiochem* **2008**, *9*, 624-33.

(50) Luzhetskyy, A.; Hoffmann, J.; Pelzer, S.; Wohlert, S. E.; Vente, A.; Bechthold, A. Aranciamycin analogs generated by combinatorial biosynthesis show improved antitumor activity. *Appl. Microbiol. Biot.* **2008**, *80*, 15-9.

(51) Lee, S. K.; Park, J. W.; Kim, J. W.; Jung, W. S.; Park, S. R.; Choi, C. Y.; Kim, E. S.; Kim, B. S.; Ahn, J. S.; Sherman, D. H.; Yoon, Y. J. Neopikromycin and novapikromycin from the pikromycin biosynthetic pathway of *Streptomyces venezuelae*. *J. Nat. Prod.* **2006**, *69*, 847-9.

(52) Desai, R. P.; Rodriguez, E.; Galazzo, J. L.; Licari, P. Improved bioconversion of 15-fluoro-6-deoxyerythronolide B to 15-fluoro-erythromycin A by overexpression of the *eryK* gene in *Saccharopolyspora erythraea*. *Biotechnol. Progr.* **2004**, *20*, 1660-5.

(53) Sparks, T. C.; Crouse, G. D.; Durst, G. Natural products as insecticides: the biology, biochemistry and quantitative structure-activity relationships of spinosyns and spinosoids. *Pest. Manag. Sci.* **2001**, *57*, 896-905.

(54) Fiedler, H. P.; Rohr, J.; Zeeck, A. Elloramycins B, C, D, E and F: minor congeners of the elloramycin producer Streptomyces olivaceus. *J. Antibiot. (Tokyo)* **1986**, *39*, 856-9.

(55) Hatano, K.; Higashide, E.; Shibata, M.; Kameda, Y.; Horii, S.; Mizuno, K. Toromycin, a New Antibiotic Produced by Streptomyces-Collinus Subsp-Albescens Subsp-Nov. *Agr. Biol. Chem. Tokyo* **1980**, *44*, 1157-63.

(56) Horii, S.; Fukase, H.; Mizuta, E.; Hatano, K.; Mizuno, K. Chemistry of Toromycin. *Chem. Pharm. Bull.* **1980**, *28*, 3601-11.

(57) Jain, T. C.; Simolike, G. C.; Jackman, L. M. Structure and Stereochemistry of Toromycin - Studies of Its Acid-Catalyzed Rearrangement. *Tetrahedron* **1983**, *39*, 599-605.

(58) Nakano, H.; Matsuda, Y.; Ito, K.; Ohkubo, S.; Morimoto, M.; Tomita, F. Gilvocarcins, new antitumor antibiotics. 1.Taxonomy, fermentation, isolation and biological activities. *J. Antibiot. (Tokyo)* **1981**, *34*, 266-70.

(59) Takahashi, K.; Yoshida, M.; Tomita, F.; Shirahata, K. Gilvocarcins, new antitumor antibiotics. 2. Structural elucidation. *J. Antibiot. (Tokyo)* **1981**, *34*, 271-5.

(60) Balitz, D. M.; O'Herron, F. A.; Bush, J.; Vyas, D. M.; Nettleton, D. E.; Grulich, R. E.; Bradner, W. T.; Doyle, T. W.; Arnold, E.; Clardy, J. Antitumor agents from Streptomyces anandii: gilvocarcins V, M and E. *J. Antibiot. (Tokyo)* **1981**, *34*, 1544-55.

(61) Li, Y. Q.; Huang, X. S.; Ishida, K.; Maier, A.; Kelter, G.; Jiang, Y.; Peschel, G.; Menzel, K. D.; Li, M. G.; Wen, M. L.; Xu, L. H.; Grabley, S.; Fiebig, H. H.; Jiang, C. L.; Hertweck, C.; Sattler, I. Plasticity in gilvocarcin-type C-glycoside pathways: discovery and antitumoral evaluation of polycarcin V from *Streptomyces polyformus*. *Org. Biomol. Chem.* **2008**, *6*, 3601-5.

(62) Yamashita, N.; Shin-Ya, K.; Furihata, K.; Hayakawa, Y.; Seto, H. New ravidomycin analogues, FE35A and FE35B, apoptosis inducers produced by *Streptomyces rochei*. *J. Antibiot. (Tokyo)* **1998**, *51*, 1105-8.
(63) Weiss, U.; Yoshihira, K.; Highet, R. J.; White, R. J.; Wei, T. T. The chemistry of the antibiotics chrysomycin A and B. Antitumor activity of chrysomycin A. *J. Antibiot. (Tokyo)* **1982**, *35*, 1194-201.
(64) Sehgal, S. N.; Czerkawski, H.; Kudelski, A.; Pandev, K.; Saucier, R.; Vezina, C. Ravidomycin (AY-25,545), a new anti-tumor antibiotic. *J. Antibiot. (Tokyo)* **1983**, *36*, 355-61.
(65) Nakashima, T.; Fujii, T.; Sakai, K.; Tomohiro, S.; Kumagai, H.; Yoshioka, T. In *United States Patent 6,030,951*; Mercian Corporation (JP): United States, 2000.
(66) Nakajima, S.; Kojiri, K.; Suda, H.; Okanishi, M. New antitumor substances, BE-12406A and BE-12406B, produced by a *streptomycete*. II. Structure determination. *J. Antibiot. (Tokyo)* **1991**, *44*, 1061-4.
(67) Morimoto, M. O., S.; Tomita, F.; Marumo, H.; Gilvocarcins, new antitumor antibiotics. 3. Antitumor activity. *J. Antibiot. (Tokyo)* **1981**, *34*, 701-7.
(68) Bockstahler, L. E.; Elespuru, R. K.; Hitchins, V. M.; Carney, P. G.; Olvey, K. M.; Lytle, C. D. Inhibition of Herpesvirus Plaquing Capacity in Human-Diploid Fibroblasts Treated with Gilvocarcin-V Plus near UV-Radiation. *Photochem. Photobiol.* **1990**, *51*, 477-9.
(69) Lytle, C. D.; Wagner, S. J.; Prodouz, K. N. Antiviral activity of gilvocarcin V plus UVA radiation. *Photochem. Photobiol.* **1993**, *58*, 818-21.
(70) Lytle, C. D.; Routson, L. B.; Prodouz, K. N. Herpes-Virus Infection and Repair in Cells Pretreated with Gilvocarcin-V or Merocyanine-540 and Radiation. *J. Photochem. Photobio. B.* **1994**, *23*, 57-62.
(71) Elespuru, R. K.; Gonda, S. K. Activation of Antitumor Agent Gilvocarcins by Visible Light. *Science* **1984**, *223*, 69-71.
(72) Tsedinh, Y. C.; Mcgee, L. R. Light-Induced Modifications of DNA by Gilvocarcin V and Its Aglycone. *Biochem. Bioph. Res. Co.* **1987**, *143*, 808-12.
(73) Gasparro, F. P.; Knobler, R. M.; Edelson, R. L. The Effects of Gilvocarcin-V and Ultraviolet-Radiation on Pbr322 DNA and Lymphocytes. *Chem.-Biol. Interact.* **1988**, *67*, 255-65.
(74) Knobler, R. M.; Radlwimmer, F. B.; Lane, M. J. Gilvocarcin V exhibits both equilibrium DNA binding and UV light induced DNA adduct formation which is sequence context dependent. *Nucleic Acids Res.* **1992**, *20*, 4553-7.
(75) Knobler, R. M.; Radlwimmer, F. B.; Lane, M. J. Gilvocarcin-V exhibits both equilibrium DNA binding and UV light induced DNA adduct formation which is seqence context dependent. *Nucleic Acids Res.* **1993**, *21*, 3920-.
(76) McGee, L. R.; Misra, R. Gilvocarcin Photobiology. Isolation and Characterization of the DNA Photoadduct. *J. Am. Chem. Soc.* **1990**, *112*, 2386-9.
(77) Alegria, A. E.; Zayas, L.; Guevara, N. A comparative-study of the visible-light photochemistry of gilvocarcin-V and gilvocarcin-M. *Photochem. Photobiol.* **1995**, *62*, 409-15.
(78) Oyola, R.; Arce, R.; Alegria, A. E.; Garcia, C. Photophysical properties of gilvocarcins V and M and their binding constant to calf thymus DNA. *Photochem. Photobiol.* **1997**, *65*, 802-10.

(79) Arce, R.; Oyola, R.; Alegria, A. E. The photobiological differences of gilvocarcins V and M are not related to their transient intermediates and triplet yields. *Photochem. Photobiol.* **1998**, *68*, 25-31.
(80) Antczak, A. J.; Tsubota, T.; Kaufman, P. D.; Berger, J. M. Structure of the yeast histone H3-ASF1 interaction: implications for chaperone mechanism, species-specific interactions, and epigenetics. *BMC Struct. Biol.* **2006**, *6*, 26.
(81) Matsumoto, A.; Hanawalt, P. C. Histone H3 and heat shock protein GRP78 are selectively cross-linked to DNA by photoactivated gilvocarcin V in human fibroblasts. *Cancer Res.* **2000**, *60*, 3921-6.
(82) Eickbush, T. H.; Moudrianakis, E. N. The histone core complex: an octamer assembled by two sets of protein-protein interactions. *Biochemistry* **1978**, *17*, 4955-64.
(83) Matsumoto, A.; Fujiwara, Y.; Elespuru, R. K.; Hanawalt, P. C. Photoactivated Gilvocarcin-V Induces DNA-Protein Cross-Linking in Genes for Human Ribosomal-Rna and Dihydrofolate-Reductase. *Photochem. Photobiol.* **1994**, *60*, 225-30.
(84) Peak, M. J.; Peak, J. G.; Blaumueller, C. M.; Elespuru, R. K. Photosensitized DNA Breaks and DNA-to-Protein Crosslinks Induced in Human-Cells by Antitumor Agent Gilvocarcin-V. *Chem.-Biol. Interact.* **1988**, *67*, 267-74.
(85) Liu, T.; Kharel, M. K.; Zhu, L. L.; Bright, S. A.; Mattingly, C.; Adams, V. R.; Rohr, J. Inactivation of the Ketoreductase gilU Gene of the Gilvocarcin Biosynthetic Gene Cluster Yields New Analogues with Partly Improved Biological Activity. *Chembiochem* **2009**, *10*, 278-86.
(86) Lorico, A.; Long, B. H. Biochemical characterization of elsamicin and other coumarin-related antitumor agents as potent inhibitors of human topoisomerase II. *Eur. J. Cancer* **1993**, *29A*, 1985-91.
(87) Takahashi, K.; Tomita, F. Gilvocarcins, new antitumor antibiotics. 5. Biosynthesis of gilvocarcins: Incorporation of ^{13}C-labeled compounds into gilvocarcin aglycones. *J. Antibiot. (Tokyo)* **1983**, *36*, 1531-5.
(88) Carter, G. T.; Fantini, A. A.; James, J. C.; Borders, D. B.; White, R. J. Biosynthesis of chrysomycin-A and chrysomycin-B origin of the chromophore. *J. Antibiot. (Tokyo)* **1985**, *38*, 242-8.
(89) Carter, G. T.; Fantini, A. A.; James, J. C.; Borders, D. B.; White, R. J. Biosynthesis of ravidomycin- use of ^{13}C-^{13}C double quantum NMR to follow precursor incorporation. *Tetrahedron Lett* **1984**, *25*, 255-8.
(90) Liu, T.; Fischer, C.; Beninga, C.; Rohr, J. Oxidative rearrangement processes in the biosynthesis of gilvocarcin V. *J. Am. Chem. Soc.* **2004**, *126*, 12262-3.
(91) Fischer, C.; Lipata, F.; Rohr, J. The complete gene cluster of the antitumor agent gilvocarcin V and its implication for the biosynthesis of the gilvocarcins. *J. Am. Chem. Soc.* **2003**, *125*, 7818-9.
(92) Bibb, M. J. B., S.; Motamedi, H., Collins, J. F.; Hutchinson, C. R. Analysis of the nucleotide sequence of the Streptomyces glaucescens tcmI genes provides key information about the enzymology of polyketide antibiotic biosynthesis. *Embo J.* **1989**, *8*, 2727-36.
(93) Fernandezmoreno, M. A.; Martinez, E.; Boto, L.; Hopwood, D. A.; Malpartida, F. Nucleotide sequence and deduced functions of a set of cotranscribed genes of

Streptomyces coelicolor A3(2) including the polyketide synthase for the antibiotic antinorhodin. *J. Biol. Chem.* **1992**, *267*, 19278-90.

(94) Kharel, M. K.; Zhu, L. L.; Liu, T.; Rohr, J. Multi-oxygenase complexes of the gilvocarcin and jadomycin biosyntheses. *J. Am. Chem. Soc.* **2007**, *129*, 3780-1.

(95) Ayer, S. W.; Mcinnes, A. G.; Thibault, P.; Walter, J. A.; Doull, J. L.; Parnell, T.; Vining, L. C. Jadomycin, a Novel 8h-Benz[B]Oxazolo[3,2-F]Phenanthridine Antibiotic from Streptomyces-Venezuelae Isp5230. *Tetrahedron Lett* **1991**, *32*, 6301-4.

(96) Doull, J. L.; Ayer, S. W.; Singh, A. K.; Thibault, P. Production of a Novel Polyketide Antibiotic, Jadomycin-B, by Streptomyces-Venezuelae Following Heat-Shock. *J. Antibiot. (Tokyo)* **1993**, *46*, 869-71.

(97) Rix, U.; Wang, C.; Chen, Y.; Lipata, F. M.; Remsing Rix, L. L.; Greenwell, L. M.; Vining, L. C.; Yang, K.; Rohr, J. The oxidative ring cleavage in jadomycin biosynthesis: a multistep oxygenation cascade in a biosynthetic black box. *Chembiochem* **2005**, *6*, 838-45.

(98) Chen, Y. H.; Wang, C. C.; Greenwell, L.; Rix, U.; Hoffmeister, D.; Vining, L. C.; Rohr, J.; Yang, K. Q. Functional analyses of oxygenases in jadomycin biosynthesis and identification of JadH as a bifunctional oxygenase/dehydrase. *J. Biol. Chem.* **2005**, *280*, 22508-14.

(99) Yang, K. Q.; Han, L.; Ayer, S. W.; Vining, L. C. Accumulation of the angucycline antibiotic rabelomycin after disruption of an oxygenase gene in the jadomycin B biosynthetic gene cluster of *Streptomyces venezuelae*. *Microbiol.* **1996**, *142*, 123-32.

(100) Chen, Y.; Fan, K.; He, Y.; Xu, X.; Peng, Y.; Yu, T.; Jia, C.; Yang, K. Characterization of JadH as an FAD- and NAD(P)H-dependent bifunctional hydroxylase/dehydrase in jadomycin biosynthesis. *Chembiochem* **2010**, *11*, 1055-60.

(101) Kharel, M. K.; Pahari, P.; Lian, H.; Rohr, J. GilR, an unusual lactone-forming enzyme involved in gilvocarcin biosynthesis. *Chembiochem* **2009**, *10*, 1305-8.

(102) Rakhit, S.; Eng, C.; Baker, H.; Singh, K. Chemical modification of ravidomycin and evaluation of biological activities of its derivatives. *J. Antibiot. (Tokyo)* **1983**, *36*, 1490-4.

(103) Luzhetskyy, A.; Mendez, C.; Salas, J. A.; Bechthold, A. Glycosyltransferases, important tools for drug design. *Curr. Top. Med. Chem.* **2008**, *8*, 680-709.

(104) Williams, G. J.; Zhang, C.; Thorson, J. S. Expanding the promiscuity of a natural-product glycosyltransferase by directed evolution. *Nat. Chem. Biol.* **2007**, *3*, 657-62.

(105) Williams, G. J.; Thorson, J. S. A high-throughput fluorescence-based glycosyltransferase screen and its application in directed evolution. *Nat. Protoc.* **2008**, *3*, 357-62.

(106) Williams, G. J.; Goff, R. D.; Zhang, C.; Thorson, J. S. Optimizing glycosyltransferase specificity via "hot spot" saturation mutagenesis presents a catalyst for novobiocin glycorandomization. *Chem. Biol.* **2008**, *15*, 393-401.

(107) Hoffmeister, D.; Wilkinson, B.; Foster, G.; Sidebottom, P. J.; Ichinose, K.; Bechthold, A. Engineered urdamycin glycosyltransferases are broadened and altered in substrate specificity. *Chem. Biol.* **2002**, *9*, 287-95.

(108) Park, S. H.; Park, H. Y.; Sohng, J. K.; Lee, H. C.; Liou, K.; Yoon, Y. J.; Kim, B. G. Expanding substrate specificity of GT-B fold glycosyltransferase via domain swapping and high-throughput screening. *Biotechnol. Bioeng.* **2009**, *102*, 988-94.
(109) Truman, A. W.; Dias, M. V.; Wu, S.; Blundell, T. L.; Huang, F.; Spencer, J. B. Chimeric glycosyltransferases for the generation of hybrid glycopeptides. *Chem. Biol.* **2009**, *16*, 676-85.
(110) Krauth, C.; Fedoryshyn, M.; Schleberger, C.; Luzhetskyy, A.; Bechthold, A. Engineering a function into a glycosyltransferase. *Chem. Biol.* **2009**, *16*, 28-35.
(111) Gantt, R. W.; Goff, R. D.; Williams, G. J.; Thorson, J. S. Probing the Aglycon Promiscuity of an Engineered Glycosyltransferase. *Angew. Chem. Int. Edit.* **2008**, *47*, 8889-92.
(112) Quiros, L. M.; Aguirrezabalaga, I.; Olano, C.; Mendez, C.; Salas, J. A. Two glycosyltransferases and a glycosidase are involved in oleandomycin modification during its biosynthesis by *Streptomyces antibioticus*. *Mol. Microbiol.* **1998**, *28*, 1177-85.
(113) Mulichak, A. M.; Losey, H. C.; Walsh, C. T.; Garavito, R. M. Structure of the UDP-glucosyltransferase GtfB that modifies the heptapeptide aglycone in the biosynthesis of vancomycin group antibiotics. *Structure* **2001**, *9*, 547-57.
(114) Mulichak, A. M.; Losey, H. C.; Lu, W.; Wawrzak, Z.; Walsh, C. T.; Garavito, R. M. Structure of the TDP-epi-vancosaminyltransferase GtfA from the chloroeremomycin biosynthetic pathway. *Proc. Natl. Acad. Sci. USA* **2003**, *100*, 9238-43.
(115) Mulichak, A. M.; Lu, W.; Losey, H. C.; Walsh, C. T.; Garavito, R. M. Crystal structure of vancosaminyltransferase GtfD from the vancomycin biosynthetic pathway: Interactions with acceptor and nucleotide ligands. *Biochemistry* **2004**, *43*, 5170-80.
(116) Coutinho, P. M.; Deleury, E.; Davies, G. J.; Henrissat, B. An evolving hierarchical family classification for glycosyltransferases. *J. Mol. Biol.* **2003**, *328*, 307-17.
(117) Kharel, M. K.; Nybo, S. E.; Shepherd, M. D.; Rohr, J. Cloning and characterization of the ravidomycin and chrysomycin biosynthetic gene clusters. *Chembiochem* **2010**, *11*, 523-32.
(118) Ichinose, K.; Ozawa, M.; Itou, K.; Kunieda, K.; Ebizuka, Y. Cloning, sequencing and heterologous expression of the medermycin biosynthetic gene cluster of Streptomyces sp. AM-7161: towards comparative analysis of the benzoisochromanequinone gene clusters. *Microbiol.-SGM* **2003**, *149*, 1633-45.
(119) Kamra, P.; Gokhale, R. S.; Mohanty, D. SEARCHGTr: a program for analysis of glycosyltransferases involved in glycosylation of secondary metabolites. *Nucleic Acids Res.* **2005**, *33*, 220-5.
(120) Gust, B.; Kieser, T.; Chater, K. *PCR targeting system in Streptomyces coelicolor A3(2)*; John Innes Centre.
(121) Kieser, T.; Bibb, M.; Buttner, M. J.; Chater, K. F.; Hoopwood, D. A.; The John Innes Foundation, Norwich, 2000.
(122) Nybo, S. E.; Shepherd, M. D.; Bosserman, M. A.; Rohr, J. Genetic manipulation of streptomyces species. *Curr. Protoc. Microbiol.* **2010**, *Chapter 10*, Unit 10E 3.

(123) Macneil, D. J.; Gewain, K. M.; Ruby, C. L.; Dezeny, G.; Gibbons, P. H.; Macneil, T. Analysis of Streptomyces avermitilis genes required for avermectin biosynthesis utilizing a novel integration vector. *Gene* **1992**, *111*, 61-8.
(124) Macneil, D. J.; Occi, J. L.; Gewain, K. M.; Macneil, T.; Gibbons, P. H.; Ruby, C. L.; Danis, S. J. Complex organization of the *Streptomyces avermitilis* genes encoding the avermectin polyketide synthase. *Gene* **1992**, *115*, 119-25.
(125) Sambrook, J.; Russel, D. W. *Molecular Cloning, A Laboratory Manual, 3rd ed.*; 3rd Ed. ed.; Cold Spring Harbor Laboratory Press: Cold Spring Harbor, NY, 2001.
(126) Rodriguez, L.; Aguirrezabalaga, I.; Allende, N.; Brana, A. F.; Mendez, C.; Salas, J. A. Engineering deoxysugar biosynthetic pathways from antibiotic-producing microorganisms: A tool to produce novel glycosylated bioactive compounds. *Chem. Biol.* **2002**, *9*, 721-9.
(127) Lombo, F.; Brana, A. F.; Salas, J. A.; Mendez, C. Genetic organization of the biosynthetic gene cluster for the antitumor angucycline oviedomycin in *Streptomyces antibioticus* ATCC 11891. *Chembiochem* **2004**, *5*, 1181-7.
(128) Blanco, G.; Patallo, E. P.; Brana, A. F.; Trefzer, A.; Bechthold, A.; Rohr, J.; Mendez, C.; Salas, J. A. Identification of a sugar flexible glycosyltransferase from *Streptomyces olivaceus*, the producer of the antitumor polyketide elloramycin. *Chem. Biol.* **2001**, *8*, 253-63.
(129) Fischer, C.; Rodriguez, L.; Patallo, E. P.; Lipata, F.; Brana, A. F.; Mendez, C.; Salas, J. A.; Rohr, J. Digitoxosyltetracenomycin C and glucosyltetracenomycin C, two novel elloramycin analogues obtained by exploring the sugar donor substrate specificity of glycosyltransferase ElmGT. *J. Nat. Prod.* **2002**, *65*, 1685-9.
(130) Perez, M.; Lombo, F.; Zhu, L. L.; Gibson, M.; Brana, A. F.; Rohr, R.; Salas, J. A.; Mendez, C. Combining sugar biosynthesis genes for the generation of L- and D-amicetose and formation of two novel antitumor tetracenomycins. *Chem.Commun.* **2005**, 1604-6.
(131) Lombo, F.; Gibson, M.; Greenwell, L.; Brana, A. F.; Rohr, J.; Salas, J. A.; Mendez, C. Engineering biosynthetic pathways for deoxysugars: Branched-chain sugar pathways and derivatives from the antitumor tetracenomycin. *Chem. Biol.* **2004**, *11*, 1709-18.
(132) Chen, Y. L.; Chen, Y. H.; Lin, Y. C.; Tsai, K. C.; Chiu, H. T. Functional characterization and substrate specificity of spinosyn rhamnosyltransferase by in vitro reconstitution of spinosyn biosynthetic enzymes. *J. Biol. Chem.* **2009**, *284*, 7352-63.
(133) Bierman, M.; Logan, R.; O'Brien, K.; Seno, E. T.; Rao, R. N.; Schoner, B. E. Plasmid cloning vectors for the conjugal transfer of DNA from Escherichia coli to Streptomyces spp. *Gene* **1992**, *116*, 43-9.
(134) Kirst, H. A. The spinosyn family of insecticides: realizing the potential of natural products research. *J. Antibiot. (Tokyo)* **2010**, *63*, 101-11.
(135) Patallo, E. P.; Blanco, G.; Fischer, C.; Brana, A. F.; Rohr, J.; Mendez, C.; Salas, J. A. Deoxysugar methylation during biosynthesis of the antitumor polyketide elloramycin by Streptomyces olivaceus. Characterization of three methyltransferase genes. *J. Biol. Chem.* **2001**, *276*, 18765-74.

(136) Deguchi, T.; Barchas, J. Inhibition of transmethylations of biogenic amines by S-adenosylhomocysteine. Enhancement of transmethylation by adenosylhomocysteinase. *J. Biol. Chem.* **1971**, *246*, 3175-81.
(137) Zappia, V.; Zydek-Cwick, R.; Schlenk, F. The specificity of S-adenosylmethionine derivatives in methyl transfer reactions. *J. Biol. Chem.* **1969**, *244*, 4499-509.
(138) Akamatsu, Y.; Law, J. H. The enzymatic synthesis of fatty acid methyl esters by carboxyl group alkylation. *J. Biol. Chem.* **1970**, *245*, 709-13.
(139) Cornell, K. A.; Riscoe, M. K. Cloning and expression of Escherichia coli 5'-methylthioadenosine/S-adenosylhomocysteine nucleosidase: identification of the pfs gene product. *Biochim. Biophys. Acta.* **1998**, *1396*, 8-14.
(140) Cornell, K. A.; Swarts, W. E.; Barry, R. D.; Riscoe, M. K. Characterization of recombinant Eschericha coli 5'-methylthioadenosine/S-adenosylhomocysteine nucleosidase: analysis of enzymatic activity and substrate specificity. *Biochem. Bioph. Res. Co.* **1996**, *228*, 724-32.
(141) Kim, H. J.; White-Phillip, J. A.; Ogasawara, Y.; Shin, N.; Isiorho, E. A.; Liu, H. W. Biosynthesis of spinosyn in Saccharopolyspora spinosa: synthesis of permethylated rhamnose and characterization of the functions of SpnH, SpnI, and SpnK. *J. Am. Chem. Soc.* **2010**, *132*, 2901-3.
(142) Decker, H.; Rohr, J.; Motamedi, H.; Zahner, H.; Hutchinson, C. R. Identification of Streptomyces olivaceus Tu 2353 genes involved in the production of the polyketide elloramycin. *Gene* **1995**, *166*, 121-6.
(143) Gullon, S.; Olano, C.; Abdelfattah, M. S.; Brana, A. F.; Rohr, J.; Mendez, C.; Salas, J. A. Isolation, characterization, and heterologous expression of the biosynthesis gene cluster for the antitumor anthracycline steffimycin. *Appl. Environ. Microb.* **2006**, *72*, 4172-83.
(144) Bradford, M. M. Rapid and sensitive method for quantitation of microgram quantities of protein utilizing principle of protein-dye binding. *Anal. Biochem.* **1976**, *72*, 248-54.
(145) Grimm, A.; Madduri, K.; Ali, A.; Hutchinson, C. R. Characterization of the *Streptomyces peucetius* ATCC 29050 genes encoding doxorubicin polyketide synthase. *Gene* **1994**, *151*, 1-10.
(146) Ye, J. S.; Dickens, M. L.; Plater, R.; Li, Y.; Lawrence, J.; Strohl, W. R. Isolation and sequence analysis of polyketide synthase genes from the daunomycin producing Streptomyces sp. strain C5. *J. Bacteriol.* **1994**, *176*, 6270-80.
(147) Gerlitz, M.; Meurer, G.; WendtPienkowski, E.; Madduri, K.; Hutchinson, C. R. Effect of the daunorubicin dpsH gene on the choice of starter unit and cyclization pattern reveals that type II polyketide synthases can be unfaithful yet intriguing. *J. Am. Chem. Soc.* **1997**, *119*, 7392-3.
(148) Rajgarhia, V. B.; Strohl, W. R. Minimal Streptomyces sp. strain C5 daunorubicin polyketide biosynthesis genes required for aklanonic acid biosynthesis. *J. Bacteriol.* **1997**, *179*, 2690-6.
(149) Bao, W. L.; Sheldon, P. J.; Hutchinson, C. R. Purification and properties of the Streptomyces peucetius DpsC beta-ketoacyl : acyl carrier protein synthase III that specifies the propionate-starter unit for type II polyketide biosynthesis. *Biochemistry* **1999**, *38*, 9752-7.

(150) Bao, W. L.; Sheldon, P. J.; Wendt-Pienkowsi, E.; Hutchinson, C. R. The Streptomyces peucetius dpsC gene determines the choice of starter unit in biosynthesis of the daunorubicin polyketide. *J. Bacteriol.* **1999**, *181*, 4690-5.
(151) Rajgarhia, V. B.; Priestley, N. D.; Strohl, W. R. The product of dpsC confers starter unit fidelity upon the daunorubicin polyketide synthase of Streptomyces sp strain C5. *Metab. Eng.* **2001**, *3*, 49-63.
(152) Tsay, J. T.; Oh, W.; Larson, T. J.; Jackowski, S.; Rock, C. O. Isolation and characterization of the beta-ketoacyl-acyl carrier protein synthase-III gene (FabH) from Escherichia coli K-12. *J. Biol. Chem.* **1992**, *267*, 6807-14.
(153) Bibb, M. J.; Sherman, D. H.; Omura, S.; Hopwood, D. A. Cloning, sequencing and deduced functions of a cluster of Streptomyces genes probably encoding biosynthesis of the polyketide antibiotic frenolicin. *Gene* **1994**, *142*, 31-9.
(154) Raty, K.; Kantola, J.; Hautala, A.; Hakala, J.; Ylihonko, K.; Mantsala, P. Cloning and characterization of Streptomyces galilaeus aclacinomycins polyketide synthase (PKS) cluster. *Gene* **2002**, *293*, 115-22.
(155) Bililign, T.; Hyun, C.-G.; Williams, J. S.; Czisny, A. M.; Thorson, J. S. The Hedamycin Locus Implicates a Novel Aromatic PKS Priming Mechanism. *Chem. Biol.* **2004**, *11*, 959-69.
(156) Wendt-Pienkowski, E.; Huang, Y.; Zhang, J.; Li, B.; Jiang, H.; Kwon, H.; Hutchinson, C. R.; Shen, B. Cloning, sequencing, analysis, and heterologous expression of the fredericamycin biosynthetic gene cluster from Streptomyces griseus. *J. Am. Chem. Soc.* **2005**, *127*, 16442-52.
(157) Marti, T.; Hu, Z.; Pohl, N. L.; Shah, A. N.; Khosla, C. Cloning, nucleotide sequence, and heterologous expression of the biosynthetic gene cluster for R1128, a non-steroidal estrogen receptor antagonist. Insights into an unusual priming mechanism. *J. Biol. Chem.* **2000**, *275*, 33443-8.
(158) Das, A.; Khosla, C. In vivo and in vitro analysis of the hedamycin polyketide synthase. *Chem. Biol.* **2009**, *16*, 1197-207.
(159) Das, A.; Szu, P. H.; Fitzgerald, J. T.; Khosla, C. Mechanism and engineering of polyketide chain initiation in fredericamycin biosynthesis. *J. Am. Chem. Soc.* **2010**, *132*, 8831-3.
(160) Meadows, E. S.; Khosla, C. In vitro reconstitution and analysis of the chain initiating enzymes of the R1128 polyketide synthase. *Biochemistry* **2001**, *40*, 14855-61.
(161) Datsenko, K. A.; Wanner, B. L. One-step inactivation of chromosomal genes in Escherichia coli K-12 using PCR products. *Proc. Natl. Acad. Sci. USA* **2000**, *97*, 6640-5.
(162) Gust, B.; Challis, G. L.; Fowler, K.; Kieser, T.; Chater, K. F. PCR-targeted Streptomyces gene replacement identifies a protein domain needed for biosynthesis of the sesquiterpene soil odor geosmin. *Proc. Natl. Acad. Sci. USA* **2003**, *100*, 1541-6.
(163) Kharel, M. K.; Pahari, P.; Lian, H.; Rohr, J. Enzymatic total synthesis of rabelomycin, an angucycline group antibiotic. *Org. Lett.* **2010**, *12*, 2814-7.
(164) Sanchez, C.; Du, L. C.; Edwards, D. J.; Toney, M. D.; Shen, B. Cloning and characterization of a phosphopantetheinyl transferase from Streptomyces

verticillus ATCC15003, the producer of the hybrid peptide-polyketide antitumor drug bleomycin. *Chem. Biol.* **2001**, *8*, 725-38.

(165) Szafranska, A. E.; Hitchman, T. S.; Cox, R. J.; Crosby, J.; Simpson, T. J. Kinetic and mechanistic analysis of the malonyl CoA : ACP transacylase from Streptomyces coelicolor indicates a single catalytically competent serine nucleophile at the active site. *Biochemistry* **2002**, *41*, 1421-7.

(166) Arthur, C. J.; Szafranska, A.; Evans, S. E.; Findlow, S. C.; Burston, S. G.; Owen, P.; Clark-Lewis, I.; Simpson, T. J.; Crosby, J.; Crump, M. P. Self-malonylation is an intrinsic property of a chemically synthesized type II polyketide synthase acyl carrier protein. *Biochemistry* **2005**, *44*, 15414-21.

(167) Horsman, G. R.; Van Lanen, S. G.; Shen, B. In *Complex Enzymes in Microbial Natural Product Biosynthesis, Part B: Polyketides, Aminocoumarins and Carbohydrates*; Elsevier Academic Press Inc: San Diego, 2009; Vol. 459, p 97-112.

(168) Doumith, M.; Weingarten, P.; Wehmeier, U. F.; Salah-Bey, K.; Benhamou, B.; Capdevila, C.; Michel, J. M.; Piepersberg, W.; Raynal, M. C. Analysis of genes involved in 6-deoxyhexose biosynthesis and transfer in *Saccharopolyspora erythraea*. *Mol. Gen. Genet.* **2000**, *264*, 477-85.

(169) Shan, M.; Sharif, E. U.; O'Doherty, G. A. Total Synthesis of Jadomycin A and a Carbasugar Analogue of Jadomycin B. *Angew. Chem. Int. Ed.* **2010**, *49*, 9492-5.

(170) Borissow, C. N.; Graham, C. L.; Syvitski, R. T.; Reid, T. R.; Blay, J.; Jakeman, D. L. Stereochemical integrity of oxazolone ring-containing jadomycins. *Chembiochem* **2007**, *8*, 1198-203.

(171) Wohlert, S. E.; Blanco, G.; Lombó, F.; Fernández, E.; Braña, A. F.; Reich, S.; Udvarnoki, G.; Méndez, C.; Decker, H.; Frevert, J.; Salas, J. A.; Rohr, J. Novel Hybrid Tetracenomycins through Combinatorial Biosynthesis Using a Glycosyltransferase Encoded by the elm Genes in Cosmid 16F4 and Which Shows a Broad Sugar Substrate Specificity. *J. Am. Chem. Soc.* **1998**, *120*, 10596-601.

VITA

Micah Douglas Shepherd

DATE OF BIRTH:

11/09/1984 Winchester, KY

EDUCATION:

2006 B.S., Biology, University of Kentucky, Lexington, KY. (GPA: 3.95)

POSITIONS AND TRAINING:

2004-2005 Independent Undergraduate Research, Department of Biology, University of Kentucky, Lexington, KY. (Advisor: Brian Rymond, Ph.D.)

2005-2006 Research Assistantship, Department of Pharmaceutical Sciences, University of Kentucky, Lexington, KY. (Advisor: Jürgen Rohr, Ph.D.)

2006-2011 Graduate Student, Department of Pharmaceutical Sciences, University of Kentucky, Lexington, KY. (Advisor: Jürgen Rohr, Ph.D.)

AWARDS:

2002-2006 Robinson Scholar, University of Kentucky, Lexington, KY.

PUBLICATIONS:

Shepherd, M. D., Liu, T., Méndez, C., Salas J. A., Rohr, J. Engineered Biosynthesis of New Gilvocarcin Analogues with Deoxyhexopyranose Moieties. *App. and Env. Micro.*, **2011**, 2, 435-441.

Bosserman, M., **Shepherd, M. D.**, Kharel, M.K., Rohr, J. Laboratory Maintanance of *Streptomyces argillaceus*, *Streptomyces griseus*. *Curr. Protoc. Microbiol.*, **2010**, November, Chapter 10:Unit 10E.2.

Kharel M. K., **Shepherd, M. D.**, Bosserman, M., Rohr J. Isolation of *Streptomyces* species from soil. *Curr. Protoc. Microbiol.*, **2010**, November, Chapter 10:Unit 10E.4.

Nybo E. S., **Shepherd M. D.**, Kharel M. K., Bosserman, M., Rohr J. Genetic Manipulation of *Streptomyces* species. *Curr. Protoc. Microbiol.*, **2010**, November, Chapter 10:Unit 10E.3

Shepherd, M. D., Kharel M. K., Bosserman, M., Rohr J. Laboratory maintenance of *Streptomyces* species. *Curr. Protoc. Microbiol.*, **2010,** August, Chapter 10 Unit 10E.1**.**

Shepherd, M. D., Kharel M. K., Zhu, L. L., Rohr J. Delineating the Earliest Steps of Gilvocarcin Biosynthesis: Role of GilP and GilQ in Starter Unit Specificity. *Org. Biomol. Chem*., **2010**, 8, 3851-3856.

[a]Kharel, M. K., [a]Nybo, S. E., [a]**Shepherd, M. D**., and Rohr, J. Cloning and characterization of the ravidomycin and chrysomycin biosynthetic gene clusters. *Chembiochem*, **2010**, 11, 523-532.

CONFERENCE PROCEEDINGS:

Shepherd, M. D., Rohr J. Biosynthetic Strategies for Producing Gilvocarcin Analogues. 50th annual meeting of the American Society of Pharmacognosy (ASP), Honolulu, HI, USA, **2009**.

Kharel, M. K., Nybo, E., **Shepherd, M. D.**, Rohr, J. Genetic Investigation of chrysomycin and ravidomycin biosynthesis. American Chemical Society (ACS) Central Regional Meeting**,** Covington, Kentucky, USA, **2007**.

Shepherd, M. D., Rohr J. Generation of Chimeric *C*-glycosyltransferase for Combinatorial Biosynthesis. American Chemical Society (ACS) Central Regional Meeting, Covington, Kentucky, USA, **2007**.