INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6" x 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.

Bell & Howell Information and Learning
300 North Zeeb Road, Ann Arbor, MI 48106-1346 USA
800-521-0600

UMI®

Studies on the Biosynthesis of Oudenone and Verucopeptin

Xianshu Yang

A Thesis

In

The Department

of

Chemistry and Biochemistry

Presented in Partial Fulfilment of the Requirements
For the Degree of Doctor of Philosophy at
Concordia University
Montreal, Quebec, Canada

March 2000

National Library
of Canada

Acquisitions and
Bibliographic Services

395 Wellington Street
Ottawa ON K1A 0N4
Canada

Bibliothèque nationale
du Canada

Acquisitions et
services bibliographiques

395, rue Wellington
Ottawa ON K1A 0N4
Canada

Your file Votre référence

Our file Notre référence

0-612-47694-4

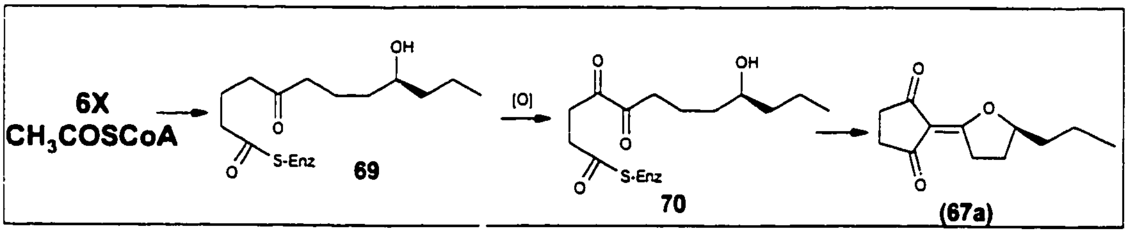
Canada

ABSTRACT

Studies on the Biosynthesis of Oudenone and Verucopeptin

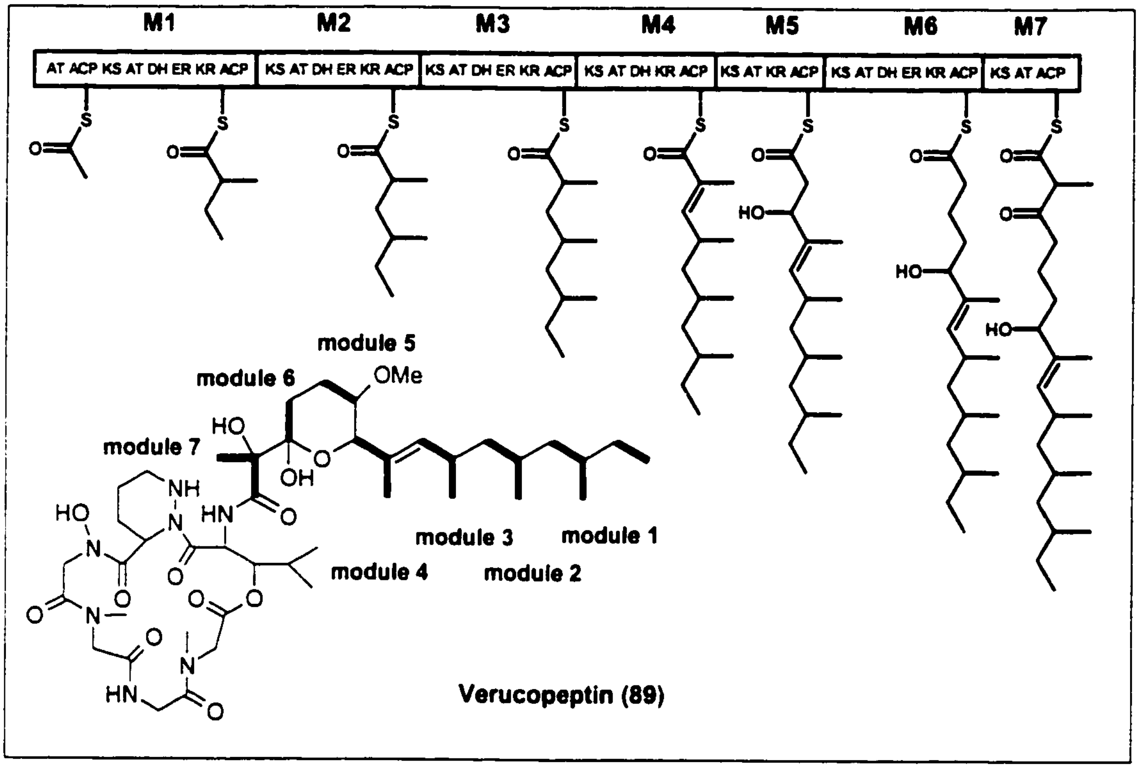
Xianshu Yang, Ph.D.
Concordia University, 2000

The biosynthesis of the fungal metabolite oudenone (**67**) was investigated. The α-diketone (**70**) was synthesized and shown to be the open-chain biosynthetic precursor of oudenone. Intact incorporation of **70** into **67** was achieved upon incubation of ^{2}H-labeled, *N*-acetylcysteamine thioester derivative of **70** with growing cultures of *Oudemansiella radicata*. The proposed mechanism for the cyclization of **70** into **67** is analogous to the "polyepoxide cascade" model, which has been previously implicated in the biosynthesis of polyether antibiotics.



The antitumor metabolite verucopeptin (**89**) is structurally characterized by a polyketide-derived tetrahydropyranyl side chain and a 19-membered cyclodepsipeptide. A DNA fragment of 72.5 kb was isolated from a cosmid library of *Actinomadura verrucosospora* chromosomal DNA. Partial analysis and sequencing of this gene cluster revealed the presence of several open reading frames (ORFs) coding for modular polyketide synthase (PKS) type I enzymes and multifunctional, non-ribosomal peptide synthetase enzymes

(NRPS). Based on the chemical structure of verucopeptin, the enzymatic motifs identified are consistent with those expected to be associated with its biosynthesis. A gene disruption experiment was performed with *A. verrucosospora* in order to demonstrate that the cloned 72.5 kb DNA fragment is associated with the biosynthesis of verucopeptin.



Verucopeptin (89)

Acknowledgements

I am fully indebted to my supervisor, Dr. Youla S. Tsantrizos, for her total commitment to my education. She made certain that my training included modern techniques in novel areas of research at all cost. She showed confidence in me when I lacked it, especially, when I was alone at Stanford University and KOSAN Biosciences Inc. Without her support, encouragement and confidence, I would have never begun, and certainly never have completed, the two projects in my thesis which involved both organic chemistry and molecular biology.

Special thanks to Professor Chaitan Khosla (Stanford University) and Dr.Gary Ashley (KOSAN Biosciences Inc) for their support and useful suggestions on the project dealing with molecular biology. I would also like to thank Professor C. Richard Hutchinson (University of Wisconsin, Madison) for his valuable advice on molecular biology.

In addition, my sincere thanks to my committee members, Dr. L. Colebrook, Dr J.Turnbull, Dr. J. Powlowski and Dr P. Joyce for their patience, help and useful suggestions and Mrs. Carole Coutts for her friendly help during last 5 years.

Many thanks to Dr. L. Liu, Dr. L. Tang, Dr. P. Hevezi, Dr. C. Gustafsson, Dr. A. Thamchaipenet, Dr. R. McDaniel, Dr. H. Fu, and Mr. X. Zhang for their help during my stay at KOSAN Biosciencres.

I am also indebted to the people in my lab, Fei Zhou, Lee Fader, AndrewMcClory, Dr. Warry Chew, Jacqueline F. Lunetta, Michael Boyd, Dr J.Shen.

I gratefully acknowledge financial support from KOSAN Biosciences Inc and Boehringer Ingelheim (Canada) Ltd, the Quebec Government in the form of a international tuition fee waiver, department of chemistry and biochemistry for a teaching assistant fellowship, NSERC and FCAR through grants to Dr.Youla S.Tsantrizos.

My deep appreciation is given to my family members, specially, my wife Qun Fan and our daughter. Their love, encouragement and understanding have been an unfailing spiritual support; without their support, I would have not been able to finish my projects.

Table of Contents:

Chapter 2: **Studies on the biosynthesis of the fungal metabolite oudenone II. synthesis and enzymatic cyclization of a α-diketone, open-chain precursor into oudenone in culture of *O. radicata***

Chapter 3: **Characterization of the genes involved in the biosynthesis of verucopeptin, a cyclodepsipeptide metabolite of *Actinomadura verrucosospora***

List of Figures

List of Tables:

List of Schemes:

List of Compounds (with structure)*:

*: Only noted the first page in which the compound structure was given. Some compounds may be shown more than once in the text following.

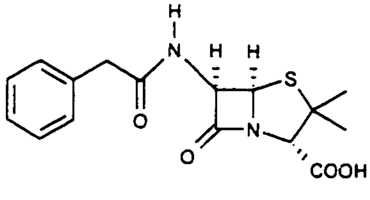
Chapter 1: Introduction

Studies of natural products often include investigations of the biosynthetic intermediates and pathways which lead to the formation of these compounds in a biological system.[1] In recent years, recombinant manipulations have been used to clone the entire gene clusters which are associated with the enzymes catalyzing the biosynthesis of natural products and to construct recombinants carrying hybrid functional genes coding for novel enzymes. Such enzymes have been expressed in genetically engineered microorganisms, plants or a variety of cells in tissue culture to produce new "unnatural" natural products with more desirable biological properties. The insights that can be gained from this type of work can be used to design more efficient routes for the total synthesis of complex natural products.

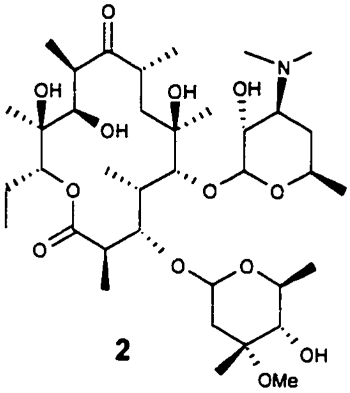
1.1 General introduction

1.1.1 Primary and secondary metabolism

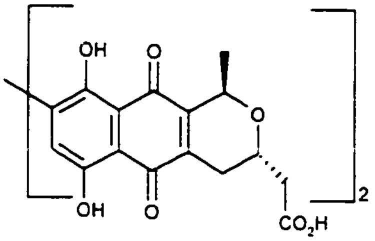
The discovery of penicillin G (**1**) launched a new era in medicine and the chemistry of natural products. In modern times, secondary metabolites isolated from cultures of microorganisms, plants, insects or marine organisms, have played a very important role in the discovery of therapeutic agents. Recent reports have indicated that secondary metabolites constitute approximately 60% of all the antitumor and antiinfective agents on the market, as well as a significant portion of the new compounds undergoing clinical testing or development.[2] Microorganisms are the largest source of these pharmaceutical agents, producing a variety of antibiotics [e.g., erythromycin A (**2**)], immunosuppressants [e.g., FK 506 (**3**) and rapamycin (**4**)] and antitumor agents [e.g., actinorhodin (**5**)].



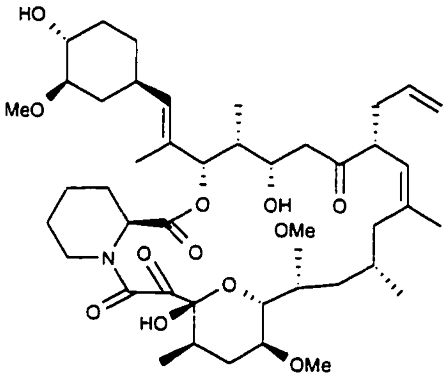
1



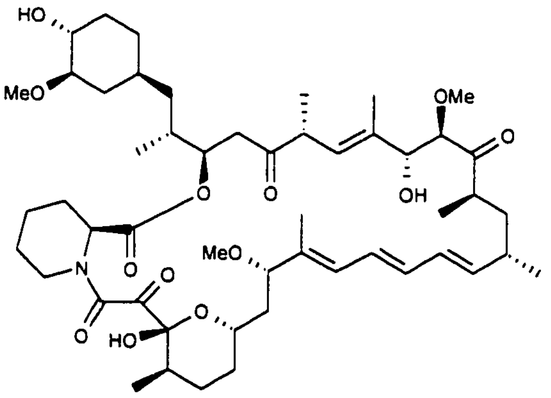
2



5



3



4

Unlike primary metabolites which are vital to every living cell (e.g., amino acids, acetyl-coenzyme A, fatty acids, sugars, nucleotides and the polymers derived from them such as polysaccharides, proteins, lipids, RNA and DNA), the exact role of secondary metabolites is not well understood. Nonetheless, after years of investigations into the field of natural products, it has become apparent that most secondary metabolites share the following key features:[3]

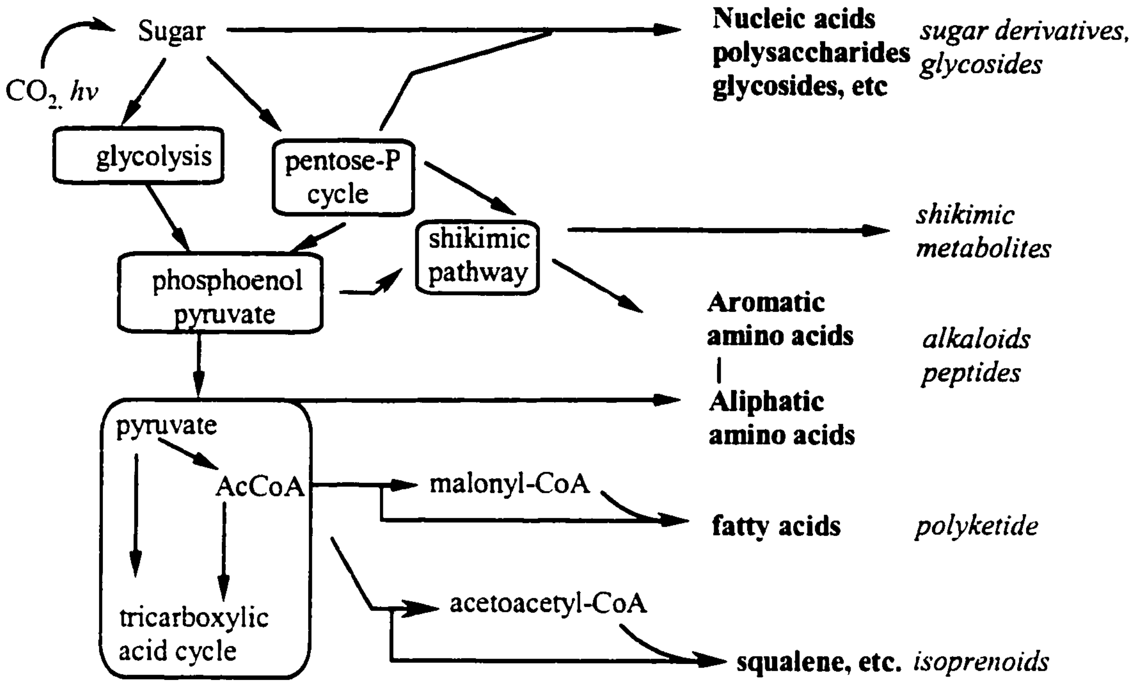
1. Secondary metabolites tend to be strain specific.

2. Although they have a wide range of chemical structures and biological activities, they belong to a small number of families which are defined by their biosynthetic origins.
3. Secondary metabolites are derived from unique and complex biosynthetic pathways, using primary metabolites as starter units. Each step in their biosynthetic pathway is usually catalyzed by an enzyme which has unique substrate specificity; this is in contrast to the biogenesis of primary metabolites which involves many shared enzymatic pathways.
4. The formation of secondary metabolites is directed by organized sets of genes associated with special regulatory mechanisms that seem to control both the timing and the level of gene expression. The control mechanisms of these genes are well integrated with the physiology of the producing organisms.

It is important to note that primary metabolism provides a number of small molecules which are employed as starting materials for all of the secondary metabolic pathways in a cell.[4] A simplified diagram of the overall relationship between primary and secondary metabolism is depicted in Scheme 1. As shown, there are three principal starting materials (or building blocks) utilized for the biosynthesis of secondary metabolites:

1. Shikimic acid, the precursor of many aromatic compounds including the aromatic amino acids, cinnamic acids and certain polyphenols;
2. Amino acids, leading to alkaloids and peptide antibiotics including the penicillins and cephalosporins;
3. Acetate, precursor of polyketides and the isoprenoids *via* two entirely separate biosynthetic pathways.

Scheme 1: Flow of carbon metabolism (left); the main products of primary biosynthesis (bold), and the corresponding categories of secondary metabolite (Italics)[4]



1.1.2 Techniques used in biosynthetic studies

1.1.2.1 *Isolation of secondary metabolites*

The isolation and purification of secondary metabolites is one of the most important steps in all studies of natural products. Most secondary metabolites are produced in minute quantities and they are usually extracted as a minor component in a large mixture of other compounds. Thus, a general purification scheme usually involves the initial isolation of a key metabolite as a mixture of many metabolites in an aqueous medium. For example, the extracellular microbial metabolites are released in an aqueous fermentation broth, whereas plant metabolites may be extracted into water by homogenizing, or even boiling, the plant tissues in water. On the basis of polarity and solubility, the different components of the mixture can then be partitioned by sequential extraction of water with immiscible organic solvents; a typical solvent sequence of increasing polarity could be hexane, chloroform, ethyl acetate and *n*-butanol. Metabolites can also be extracted into an organic solvent directly from the solid materials by using a continuous liquid-solid extraction method.

After the initial extraction of a metabolite mixture, separation of its components can be carried out by:

1. Adsorption chromatography, which depends on the adsorption-desorption equilibrium between a compound adsorbed on the surface of a solid stationary phase and the moving solvent.
2. Ion-exchange chromatography, which separates molecules on the basis of their net charge.
3. Size-exclusion chromatography, which involves separation of molecules on the basis of their molecular size.

1.1.2.2 *Nuclear magnetic resonance (NMR)*

NMR is an indispensable tool in any study involving the structure determination of natural products or the identification of intermediates in the biosynthetic pathways of these compounds.[5] In most cases, the one-dimensional (1D) ^{1}H NMR and ^{13}C NMR spectra of a complex, unknown metabolite are rarely sufficient to allow its unambiguous structural assignment. However, additional information can be easily obtained from two-dimensional (2D) NMR experiments. Among the most useful types of experiments are the COSY (*CO*rrelated *S*pectroscop*Y* for homonuclear correlation of ^{1}H chemical shifts), INADEQUATE (*I*ncredible *N*atural *A*bundance *D*oubl*E* *QUA*ntum *T*ransfer *E*xperiment, for homonuclear correlation of ^{13}C chemical shift), DEPT (*D*istortionless *E*nhanced *P*olarization *E*xperiment for the determination of the number of hydrogen atoms attached to each carbon), and HETCOR (*HET*eronuclear *COR*relation for the determination of the connectivities between protons and carbons).

In recent years, inverse correlation spectroscopy with heteronuclear multiple quantum phenomena and inverse probe heads have been used to detect insensitive nuclei (^{13}C, ^{15}N) by detecting the abundant nuclei (^{1}H, ^{19}F, ^{31}P). The new technique could improve the sensitivity of a low abundance nucleus with a few milligrams of sample. Examples of this type of experiment include the HMQC (*H*eteronuclear *M*ultiple *Q*uantum *C*oherence) experiment for nuclei with at least one directly attached proton and HMBC (*H*eteronuclear *M*ultiple *B*ond *C*oherence) experiment for nuclei connected to protons two or more bonds away.[6]

1.1.2.3 *Isotopic labeling*

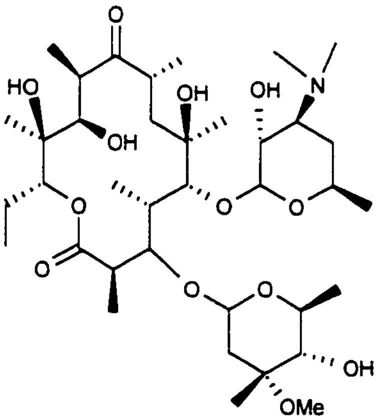
A multitude of different approaches is used to study the biosynthesis of secondary metabolites. These approaches usually involve the "feeding" of various isotopically labeled precursors, or intermediates, to an intact microorganism in order to assess the role of each labeled compound in the biosynthetic pathway. Such studies may be complemented with experiments using a cell-free system or purified enzymes.

Radioactive isotopes, usually ^{14}C and/or ^{3}H were widely used around 20 years ago to label putative precursors or intermediates of biosynthetic pathways. However, the relationship between the precursor and the final product can only be assessed by examining the percent incorporation of label, the percent specific incorporation, or the dilution factor. Thus, it is very difficult to draw definitive conclusions on the precise location of the radioactive label in the structure of the final metabolite without chemical degradation and further analysis of the labeled products.

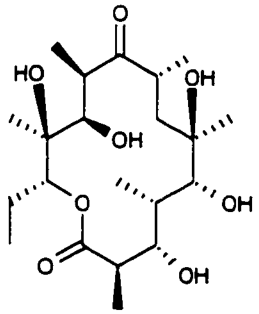
A far simpler approach, which also provides more information about a biosynthetic pathway, involves the use of stable isotopes.[7] The stable isotopes widely used are ^{13}C, ^{2}H, ^{15}N, and ^{18}O. Although the latter (^{18}O) does not give rise to NMR signals, its presence and exact location can be identified from the chemical shift of the carbon atom to which it is attached, because the presence of ^{18}O on a particular carbon atom causes an upfield shift of the ^{13}C resonance of that carbon atom relative to its chemical shift when it was connected to ^{16}O in the same compound.[8] The subsequent analysis of the labeled final metabolite by NMR spectroscopy can provide detailed information on the origin of a natural product, as well as some insight into the mechanisms associated with the enzymes catalyzing the metabolite. A precursor with a single stable isotope can be traced through a metabolic reaction sequence to the final product. The ^{13}C NMR signal corresponding to the atom (or atoms) in the product that originate(s) from the ^{13}C-labeled atom of the precursor will show an increased intensity in the ^{13}C NMR spectrum. ^{2}H-NMR directly shows the labeled position(s) in the final product when ^{2}H-labeled precursor(s) are fed into the investigated organism. A precursor with two adjacent ^{13}C-stable isotopes can be used to observe the fate of a bond through a metabolic reaction sequence by carbon-carbon coupling in the ^{13}C NMR spectrum. The main advantage that radioactive tracer experiments provide over experiments employing stable isotopes is in the sensitivity of detection, which can be a serious limitation in the use of stable isotopes if the amount of incorporation of labeled precursors or the amount of produced metabolite is very low. Nonetheless, labeling of biosynthetic precursors with stable isotopes, in conjunction with high-field NMR spectroscopy, has become one of the most commonly employed methods for exploring biosynthetic pathways.

1.1.2.4 Examples of the use of stable isotopes in investigating the biosynthetic pathway of erythromycin A.

Erythromycin A (**2**), a macrolide antibiotic produced by *Saccharopolyspora erythraea*, is widely used to treat infections caused by Gram-positive bacteria. It is composed of the 14-membered macrolactone ring 6-deoxyerythronolide B (6-dEB, **6**), to which two deoxysugars are attached.[9]



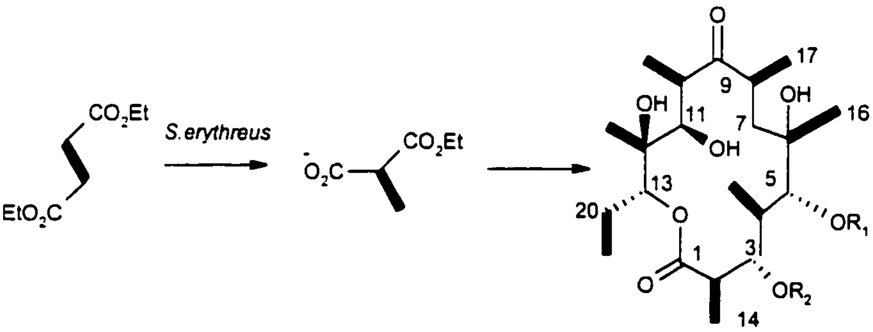
2



6

Corcoran, Grisebach and their co-workers fed radioactive propionate and methylmalonate to *S. erythreus* followed by partial degradation of the labeled macrolide and identified propionate as the origin of erythromycin A (**2**).[10] In the early 1980s, Cane and co-workers systematically studied the biosynthetic pathway of this polyketide using stable isotope precursors. Feeding experiments using [1-^{13}C] and [2-^{13}C] propionate further revealed that the carbon skeleton of **6** is derived from seven propionate units.[11] Diethyl [2,3-$^{13}C_2$] succinate was used as a source of [2,2'-$^{13}C_2$] methylmalonyl CoA produced *in vivo* by the action of methylmalonyl CoA mutase and considered as biologically equivalent to [2,3-$^{13}C_2$] propionate. Incorporation of the diethyl [2,3-$^{13}C_2$] succinate gave rise to seven pairs of enhanced and coupled doublets in the ^{13}C-NMR spectra of the labeled erythromycin A (**2**, Scheme 2).[11]

Scheme 2: Carbon origins of erythromycin A (**2**) by feeding ^{13}C precursors

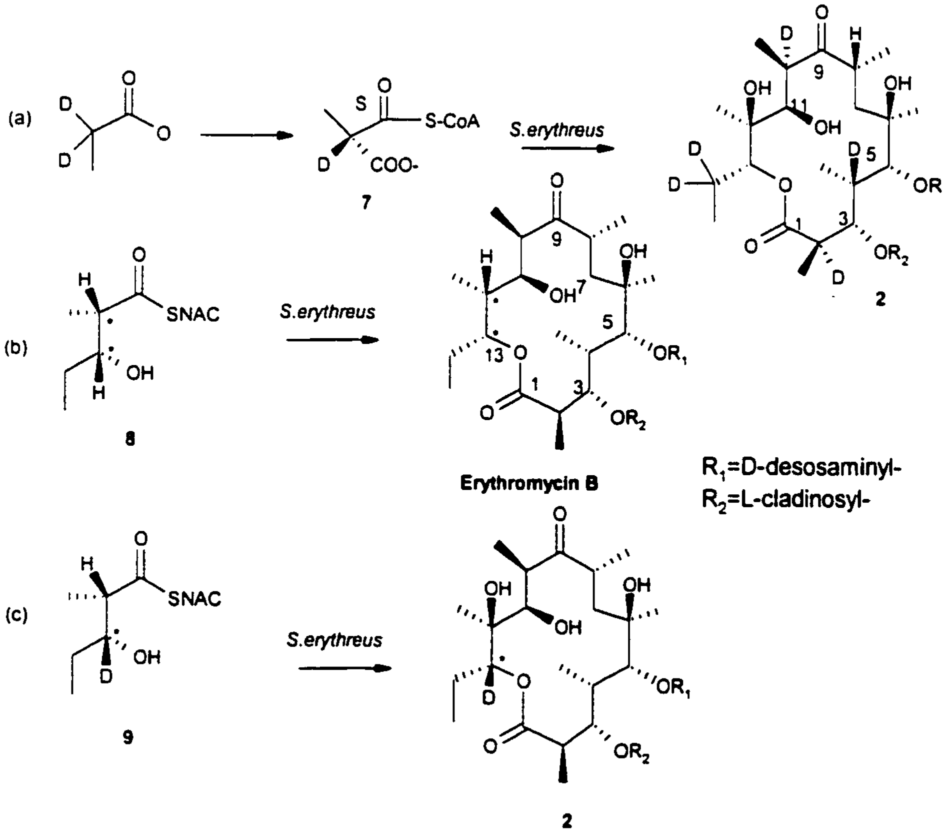


2

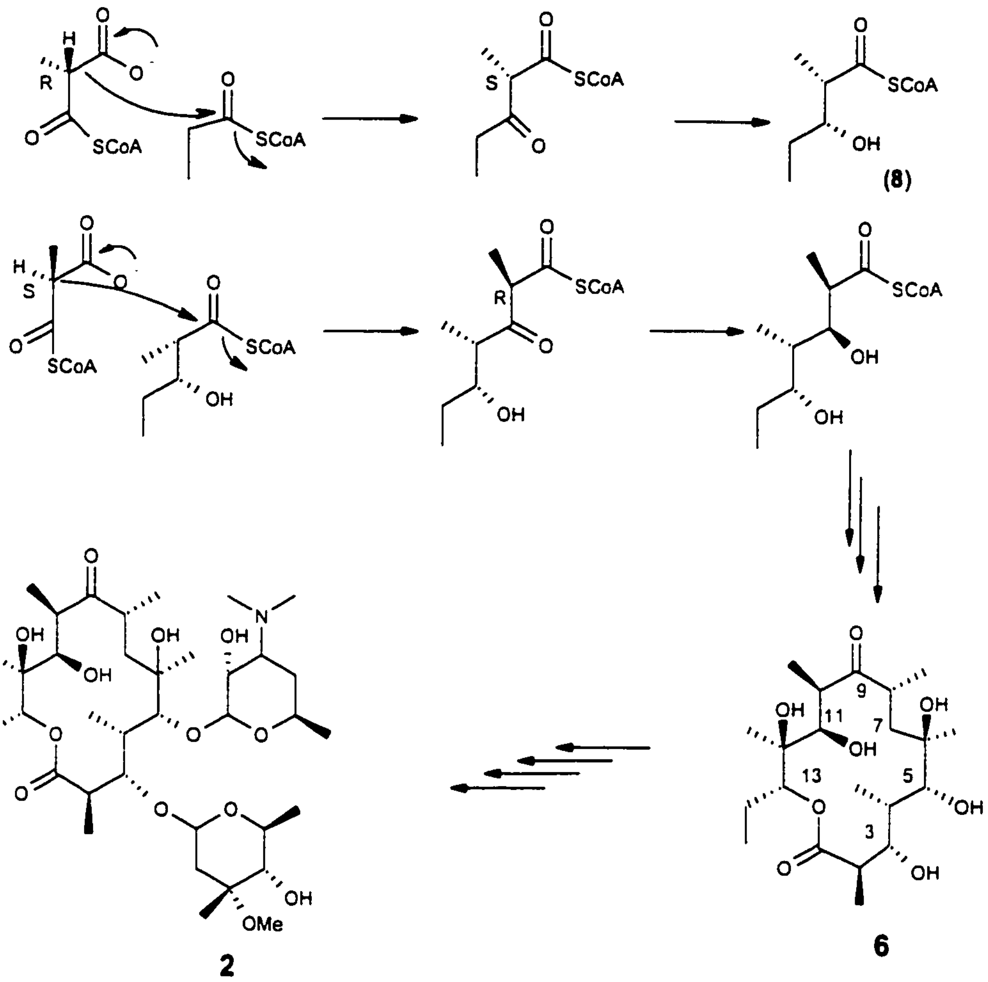
R1=D-desosaminyl-
R2=L-cladinosyl-

Furthermore, from feeding experiments using [1-$^{18}O_2$, 1-^{13}C] acetate, they were able to show that almost all of the oxygen atoms of **6** are also derived from propionate, with the exception of the hydroxyl groups at C-6 and C-12.[12] The D-methyl substituents at C-2, -4, and C-10 of erythromycin result the stereospecific decarboxylation of propionyl CoA to give (2S)-methylmalonyl-CoA (**7**), which then undergoes condensation with the appropriate enzyme-bound acyl thioester. NMR data obtained from the incorporation of [2-$^{2}H_2$, 2-^{13}C] propionate into 6-dEB (**6**) confirmed that the latter step involves decarboxylative inversion of the stereochemistry at the C-2 center of propionate (Scheme 3a).[13] Cane's group was also successful in achieving the intact incorporation of labeled chain-elongation intermediates such as compound **8** and **9** into erythromycin B and erythromycin A (**2**), respectively (Scheme 3b and 3c). These results established the precise order of biochemical events leading to the formation of erythromycin A (**2**).[14] The overall biosynthetic pathway of erythromycin A is now well established to follow the sequence outlined in Scheme 4.

Scheme 3: Advanced precursors for biosynthesis of erythromycin A (**2**) and B. *^{13}C - labeled positions



Scheme 4: Proposed biosynthetic pathway of erythromycin A (**2**)



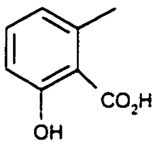
Feeding of advanced precursors has also been used to study the biosynthesis of numerous natural products including methymycin,[15] nargenicin,[16] aspyrone[17] and nonactin.[18]

1.2 Polyketides

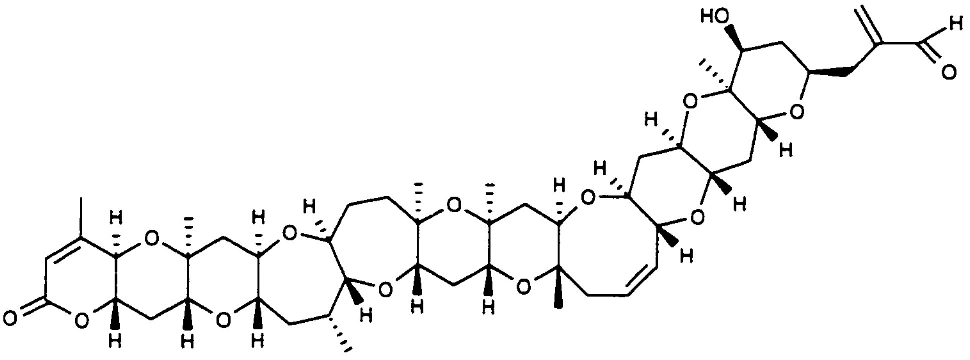
Polyketides are a class of metabolites produced by bacteria, fungi and plants, which have been used widely in human and veterinary medicine, agriculture and nutrition (Table 1). Polyketides can vary widely in structure, from the eight-carbon compound 6-methylsalicylic acid (**10**) to brevitoxin B (**11**) which contains 50 carbon atoms in its chain. The main part of the carbon skeleton of all of these compounds is synthesized through the successive condensation of simple carboxylic acids, such as acetate, propionate and butyrate.

Table 1: Major pharmaceutical polyketides and their activities

Product	Therapeutic Area	Product	Therapeutic Area
Clarithromycin	Antibacterial	Idarubicin (Idamycin)	Anticancer
Azithromycin	Antibacterial	Amphotericin B	Antifungal
Erythromycin	Antibacterial	Candicidin	Antifungal
Dalfopristin	Antibacterial	Griseofulvin	Antifungal
Josamycin	Antibacterial	Nystatin/Mycostatin	Antifungal
Minocycline (Dynacil)	Antibacterial	Spiramycin	Antirickettsial
Miokamycin	Antibacterial	Mevacor	Cholesterol-lowering
Mycinamicin	Antibacterial	Mevastatin	Cholesterol-lowering
Oleandomycin	Antibacterial	Pravastatin	Cholesterol-lowering
Pristinamycin	Antibacterial	Zocor	Cholesterol-lowering
Pseudomonic acid	Antibacterial	Zearalenone	Estrogenic activity
Rifamycins (Rifampin)	Antibacterial	Ascomycin	Immunosuppressant
Rokitamycin	Antibacterial	FK506	Immunosuppressant
Roxithromycin	Antibacterial	Rapamycin	Immunosuppressant
Tetracyclines	Antibacterial	Spinosad	Insecticide
Aclarubicin	Anticancer	Avermectin	Veterinary Product
Adriamycin	Anticancer	Doramectin	Veterinary Product
Chromomycin	Anticancer	Lasalocid A	Veterinary Product
Daunorubicin	Anticancer	Milbemycin	Veterinary Product
Enediynes	Anticancer	Monensin	Veterinary Product
Tylosin	Veterinary Product		



6-Methylsalicylic acid (10)

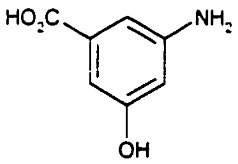


Brevetoxin B (11)

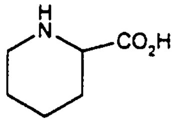
1.2.1 Polyketide Biosynthesis

Although diverse in structure and properties, polyketides can be grouped into two overall classes: the aromatic and the complex polyketides (e.g., polyether, and macrolides).[19] Aromatic polyketides are produced through the condensation of acetate starter and malonate extender which generate (β-) keto groups that remain largely unreduced during and after growth of the acyl chain. The acyl chain, either during or after completion of its synthesis, undergoes enzymatic, programmed or spontaneous, folding to allow aldol condensations to take place resulting in the formation of 6-membered rings. These rings are subsequently converted to aromatic systems through enzymatically controlled dehydration. In the majority of aromatic polyketides, the methyl substituents are often derived from S-adensoyl methionine (SAM).

The complex polyketides are structurally more diverse than the aromatic compounds and, unlike the mainly acetate-derived aromatic metabolites, are composed of acetate, propionates and butyrates in varying ratios. Because of a fundamental difference from the aromatic group in the chemistry of their biosynthesis, processive β-carbonyl reduction, as well as the structural constraints imposed by the presence of methyl side chains, these molecules can not undergo folding and aromatization. These molecules consist of long alkyl carboxylates carrying various functional groups at odd-numbered carbon atoms (keto, hydroxyl, enoyl, or alkene) and different side chains at even-numbered carbons (hydrogen, methyl, ethyl, or propyl) [e.g., niddamycin (**12**) and tylosin (**14**)]. In addition, the polyketide chain may incorporate an unusual moiety [e.g., mC_7N unit at the α-position in rifamycin B (**13**). The mC_7N unit consists of a six-membered carbocyclic ring carrying an additional carbon and a nitrogen substituent in a 1,3 (*meta*) arrangement. One example, 3-amino-5-hydroxy-benzoic acid, is shown below. This moiety may be as linked to the carboxylate [e.g., pipecolate shown below in FK506 (**3**) and rapamycin (**4**)]. Scheme 5 shows a selected number of complex polyketides and their primary building blocks.

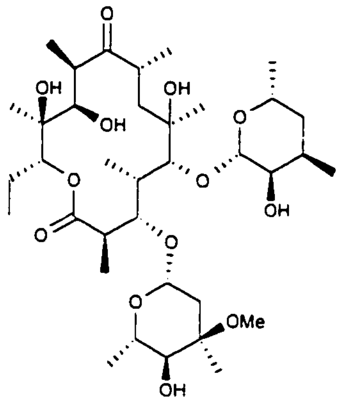


3-Amino-5-hydroxy-benzoic acid

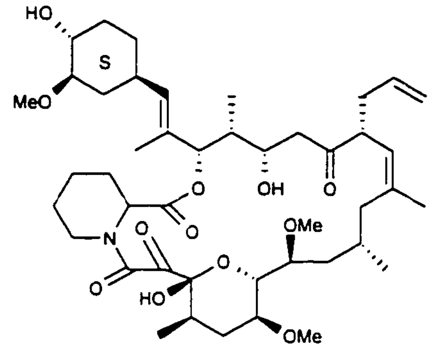


Pipecolate

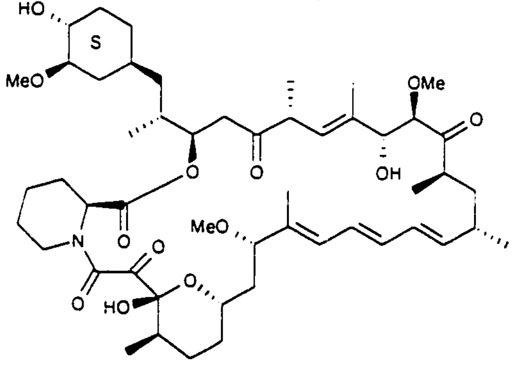
Scheme 5: Structures of some complex polyketides produced by bacteria.
The sequences of letters below the name of each compound indicate the sequence of residues contained in each polyketide chain. **A**: acetate, **P**: propionate, **B**: butyrate, **S**: unusual starter unit



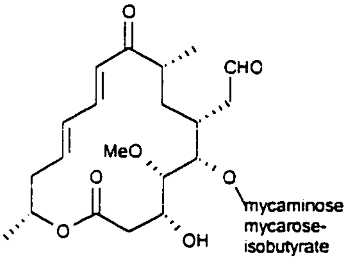
Erythromycin A (2)
(PPPPPPP)



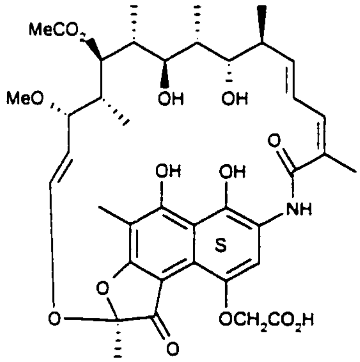
FK506 (3)
(SPPAPPPAAPA)



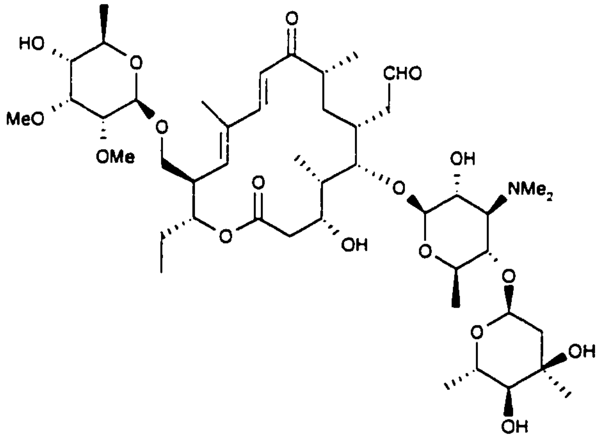
Rapamycin (4)
(SPAPPAPPAAPAAPA)



Niddamycin (12)
(AAAAPBAA)



Rifamycin B(13)
(SPAPPPPPPAP)



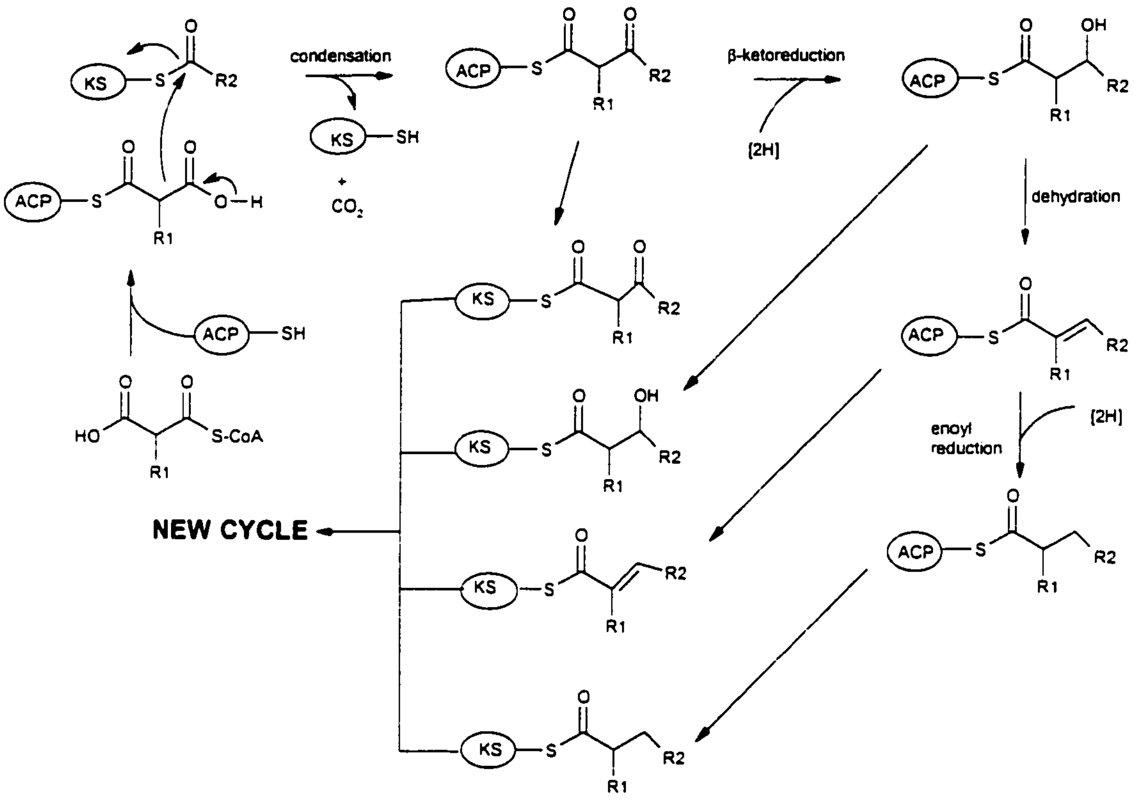
Tylosin (14)
(PPPAPBPA)

1.2.2 Enzymology and genetics associated with fatty acid and polyketide biosynthesis

In 1953, Birch and Donovan proposed that polyketides were formed through the condensation of acetate units to a hypothetical poly-β-ketone.[20] Since that first proposal, substantial similarities between the formation of long chain fatty acids (LCFA) and polyketides has been found through biosynthetic studies of numerous natural products. All of the enzymes involved in fatty acid biosynthesis (FAS) fall into two distinct classes.[21] The *type I* systems consist of multifunctional polypeptides carrying the required active sites as domains and are typical of yeast and animal systems, whereas the *type II* systems are characteristic of plant and bacterial enzymes and are comprised of several discrete peptides, each carrying a distinct activity, loosely associated in a complex.

Scheme 6 shows the general mechanism of fatty acid and polyketide formation. For every chain elongation cycle, the process begins when malonate is transferred from coenzyme A (CoA) to the phosphopantotheine arm of the acyl carrier protein (ACP) catalyzed by the enzyme acetyltransferase (AT). Decarboxylative condensation occurs between the ACP-bound malonate and the growing chain, which is attached through a thioester linkage to the active-site cysteine residue of the β-ketoacylsynthase (KS), the condensing enzyme. The resulting ACP-bound β-ketoacyl chain can then undergo three successive processing steps, a β-ketoreduction, a dehydration and an enoylreduction, by the action of the β-ketoreductase (KR), dehydratase (DH) and enoylreductase (ER), respectively. These processes are repeated until the desired chain length and functional group modifications are reached.

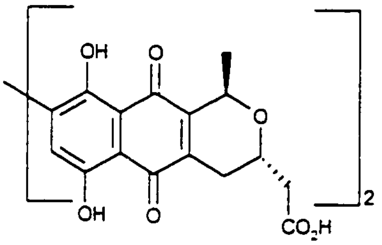
Scheme 6: Catalytic reactions involved in fatty acid and polyketide biosynthesis.



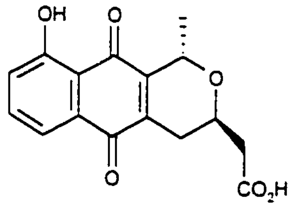
1.2.3 The polyketide synthase (PKS) enzymes

Polyketides are produced by the sequential activity of a large number of enzymes (5 to 50), collectively known as "polyketide synthase" (PKS). Once the parent polyketide is synthesized, it can be further modified by "tailoring enzymes" which add sugar group(s), some methyl group(s), or make other modification(s) on the polyketide core molecule. [22]

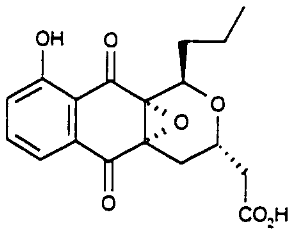
There are two types of PKSs in bacterial systems: "modular" and "iterative".[23] “Iterative" PKSs are analogous to bacterial fatty acid synthases and composed of active sites that are used more than once in the biosynthetic pathway of aromatic polyketides. These enzymes are similar to the PKS type I enzymes. The main difference is that they are composed of several separate, largely monofunctional proteins that catalyze formation of cyclic aromatic antibiotics that do not require extensive reduction and/ or dehydration cycles such as actinorhodin (**5**), nanaomycin (**15**) and frenolicin B (**16**). A detailed discussion of iterative PKS enzymes is beyond the scope of this thesis. A number of excellent reviews in this field are available for more details in [*Chemical Reviews*].[24]



5



15



16

Modular PKSs are assemblies of large multifunctional proteins with a different active-site for each catalytic step associated with the carbon chain assembly of complex polyketides such as FK506 (**3**) and rapamycin (**4**). They are organized into groups of active sites known as modules, in which each module is responsible for one cycle of polyketide chain extension and functional group modification. Each module contains the three essential domains of KS, AT and ACP, plus one or the full complement of the domains of KR, DH, and ER that are analogous in both function and sequence to the individual enzymes of fatty acid biosynthesis. Modular PKSs can be segregated from amino to carboxyl termini into a "loading" domain, multiple "modules", and a "releasing domain". The detailed explanation is given in **1.2.5.1.**

1. 2. 4 General strategies for cloning and manipulating genes associated with the biosynthesis of polyketides

Soon after methods for gene cloning in *Streptomyces* species were published in 1980,[25] it became possible to isolate the genes involved in the biosynthesis of antibiotics by a variety of procedures.[26] Examples include the complementation of mutants with DNA fragments from wild type chromosomal DNA. These DNA fragments could confer new biological properties to the mutants such as the production of novel biosynthetic intermediates or novel metabolites. Table 2 shows some examples of PKS gene clusters which have been cloned and sequenced.

Each of the enzymes in a PKS is encoded by contiguous genes found as an integrated unit in the host organism's genome and referred to as a "gene cluster".[27] Thus, when any component of the polyketide gene cluster is identified, the entire pathway can be obtained by sequencing in either direction of the chromosomal DNA such that the genes associated with the biosynthesis of a metabolite can be moved to plasmid vectors usually in their entirety.

Table 2: Examples of PKS gene clusters which have been cloned and sequenced

Host	Polyketide (aglycon)	PKS genes	Cloning strategy	Evidence for cloning of correct genes
Sac.Erythraea	Erythromycin(**2**) (6-dEB)	*ery*	Resistance, followed by walking and complementation	1,2,3,4,5,
S.Sp.MA 6548	FK506 (**3**)[28]	*fkb*	Reverse genetics for a tailoring step, followed by walking and sequencing	2
S.hygroscopicus	Rapamycin (**4**)[29]	*rap*	*ery* probe	2.,5
S.glaucescenc	Teteacenomycin C [30]	*tcm*	complementation	1,2,3
S. antibiotius	Oleandomycin [31]	*ole*	*ery* and resistance gene probes	
S.fradiae	Tylosin (**14**)[32]	*tyl*	Reverse genetics for a tailoring step, followed by walking and complementation	1,2
S.ambofacienes	Spiramycin [33]	*srm*	Resistance, followed by walking and complementation	1,5
Sorangium cellulosum	Soraphen A [34]		*gra* (type II PKS) probes	2
S.avermitillis	AvermectinBla[35]	*avr*	Complementation of blocked mutants followed by walking	1,2
Amycolatopsis mediterranei S699	Rifamycin (**13**)[36]	*rif*	Reverse genetics for a tailoring step, followed by walking and sequencing	2
S.venezuelae	Methymycin, neomethymycin,narbomycin, pikromycin37	*pik*	Type I PKS and tylAI a-D-glucose-1-phosphate thymidylyltransferase.	2
S.caelestis	Niddamycin (**12**)[38]	*nid*	KS and AT –specific oligonucleotide PCR product	2
S. coelicolor	Actinorhodin (**5**)[39]	*act*	complementation	1,2,3
S.roseofulvus	Nanaomycin(**15**)/ frenolicin (**16**) [40]	*fren*	*act* probe	4
S.griseus	Griseusin [41]	*gris*	*act* probe	2
S.sp.C5	Daunorubicin [42]	*dau*	*act* probe	1

Key: 1, complementation of pathway-blocked mutants; 2. Gene disruption; 3, production of antibiotic after transfer of cloned genes to *S.lividans*; 4. Production of relevant compounds in recombinant strains; 5. Agreement of module number, starting module content of protein subunits.

1.2.5. Cloning and sequencing of the erythromycin PKS genes

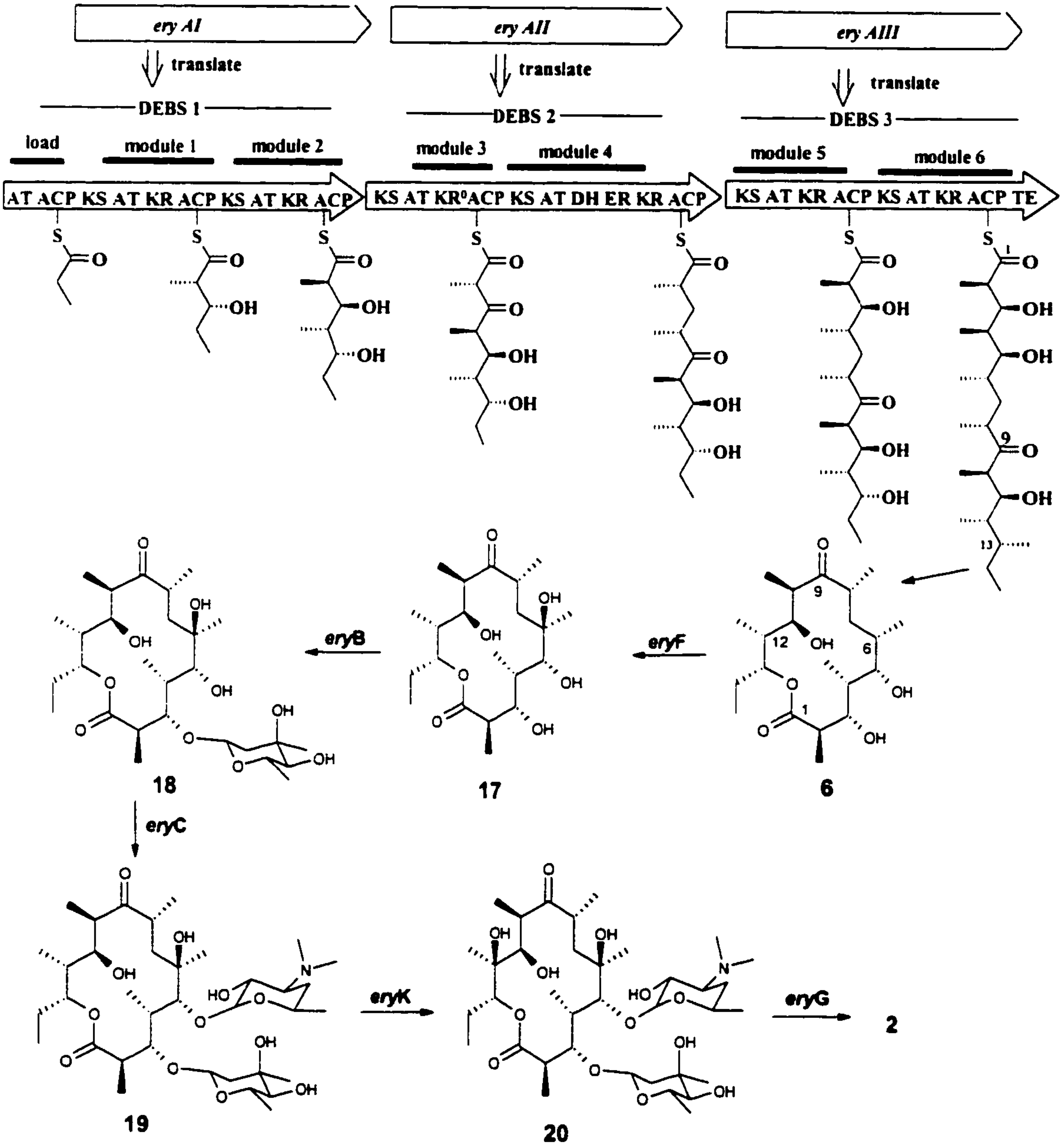
The best-studied example of a complex polyketide is the antibiotic erythromycin A (**2**) produced by *S. erythraea*.[43] The genes coding for the erythromycin PKS enzymes have been cloned and expressed in a heterologous host and have been used as a model to pioneer the field of combinatorial biosynthesis.

Leadlay,[44] Katz[45] and their co-workers cloned the gene cluster of *ery*A which is associated with the biosynthesis of erythromycin A (**2**) from the genomic library of *S. erythraea* using the resistance gene (*erm*E) against erythromycin as probe.[46] The gene clusters from *S. erythraea* and its roles in the biosynthesis of erythromycin A (**2**) are shown in Scheme 7. The three genes identified in this gene cluster were named *eryAI*, *eryAII*, and *eryAIII*. Each gene is approximately 10.0 kb in length and encodes one polypeptide named *de*oxyerythronolide *B s*ynthase (DEBS) 1, 2 or 3. Each of the DEBS proteins is over 3,300 amino acids in length and is composed of the enzymatic domains AT, ACP, KS, DH, KR and ER. A single domain with thioesterase (TE) activity is also found near the C-terminus of DEBS 3. DEBS1 also carries an additional AT and ACP loading domains near the N-terminus of the polypeptide. Thus, the DEBSs are multifunctional proteins resembling the type I fatty acid synthase found in vertebrates which carry a single copy of each domain. In addition, each DEBS protein carries two modules of the domains KS, AT, ACP, six in all, and a varying combination of DH, ER, and KR domains within each module in Scheme 7.

It was proposed that the AT at the loading domain of the DEBS1 binds propionyl CoA and transfers it to the pantotheine arm of the adjacent ACP domain and then to the active site cysteine residue of the KS domain of module 1. The AT domain in module 1 binds methylmalonyl CoA and transfers it to the ACP domain in module 1. The diketide is formed through decarboxylative condensation of the extender unit to the transfer unit catalyzed by KS. Then reduction of the β–carbonyl to a hydroxyl, catalyzed by the KR1 domain in module 1 takes place. Since no other functional domains are present, the diketide is transferred from the ACP to the KS domain of module 2 where it undergoes condensation with the second methylmalonyl moiety attached to the ACP domain in the same module. Again, only β-ketoreduction takes place due to the absence of functional domains other than the KR in module 2. After completion of this cycle, the triketide is transferred to the KS domain of module 3. Sequence analysis indicated that the KR domain in this module would be dysfunctional (denoted KR^{0}) due to the absence of a functional NADPH binding site (Scheme 7). Thus, after the third condensation, the β-keto group is retained through the succeeding biosynthetic steps and appears at C-9 in metabolite **6**. In contrast, ketoreduction is followed by dehydration and enoyl reduction to produce the methylene moiety which is found at C-7 of **6**, due to the presence of active DH and ER domains in module 4. The reduction cycle in modules 5 and 6 similar to that in modules 1 and 2. After six condensations, the 13-carbon chain is released from the ACP domain of DEBS3 through the action of the TE domain and forms the macrolactone **6** by attack of the C-13- hydroxyl group on the C1 thioester.

After synthesis of **6** by the three DEBS enzymes, the intermediate **6** is hydroxylated at C-6 by the P450 hydroxylase EryF, encoded by the *eryF* gene, to produce the compound erythronolide B (**17**).[47, 44e] The next steps involve addition of the two deoxysugars, first L-mycarose to yield 3-α-mycarosylerythronolide (**18**), and then desosamine to produce erythromycin D (**19**).[48] Erythromycin D (**19**) is converted to erythromycin C (**20**) by another P450 hydroxylase, the product of the *eryK* gene, acting at C-12.[49] Erythromycin C (**20**) is methylated at C-3' on the mycarose moiety to form erythromycin A (**2**).[50]

Scheme 7: The erythromycin A (**2**) biosynthesis gene clusters in *S.erythraea* and the role in erythromycin A (**2**) biosynthesis (detail see text)



From the above discussion of the catalytic mechanisms involved in the biosynthesis of **6** as well as many other known type I PKSs, the KS, AT, and ACP domains are essential for chain extension. The AT domains determine which extender unit is incorporated at each step of the polyketide chain, while the KS domains are responsible for catalyzing the actual condensation reactions. The role of ACP domains is to tether the growing polyketide chain to the PKS between condensations and to accept extender units from the AT domains in preparation for each condensation step.[51]

1.2.6 Combinatorial approaches to polyketide biosynthesis

In general, the cornerstone of combinatorial chemistry is the ability to produce large numbers of related structures (chemical libraries) using synthetic or biosynthetic chemistry. It is perhaps most remarkable that microorganisms which produce polyketide-type metabolites have the ability to do so by using the simplest of starting materials (e.g. acetate, malonate, propionate, 2-methylmalonate, butyrate) and a set of PKS enzymes.[52] The mode of assembly of complex polyketides by modular PKSs leads readily to the idea of combinatorial biosynthesis, since a separate active site is involved in the catalysis of each step.[53] When targeted alterations (like deletion, substitution, or addition of active sites) are made in the structural genes, a clear prediction can be made for the precise point at which chain assembly or structural modification will be affected, assuming that all the constituent enzymes act autonomously. The genetic engineering of the gene cluster associated with biosynthesis of erythromycin A (**2**) has been used as a model which can be used to vastly expand the molecular diversity of pharmacologically important compounds.[54]

A special host-vector system has been used in the manipulation of the gene clusters associated with biosynthesis of erythromycin A (**2**) and has generated a lot of "unnatural" natural products. This system was originally designed for the modification of the more thoroughly investigated type II PKSs in Khosla's group,[55] but has turned out to be suitable for modular PKSs as well.[56] In the host-vector system, the host is a *Streptomyces coelicolor* derivative strain (CH 999) from which the entire set *act* of genes (except for *act*VI-ORFA at the extreme left-hand end of the cluster) coding for the biosynthesis of the aromatic polyketide actinorhodin (**5**) has been deleted and replaced by a convenient marker gene for erythromycin resistance.[57] A shuttle vector (for example, pRM5) carrying both an *E.coli* origin of replication and the ampicillin resistance gene has been used for rapid genetic engineering of different biosynthetic pathways in *E.coli*. This vector also carries a *Streptomyces* origin of replication and the thiostrepton resistance gene for selection in *Streptomyces*. Thus, a series of plasmids carrying the desired sets of PKS subunit genes can be introduced into the CH999 host by standard procedures of protoplast transformation.[58] This host-vector system has been extremely important for work aimed at exploring the creation of hybrid PKSs, due to the high efficiency with which mutations can be made and the enhanced levels of PKS expression that can be achieved.

There are four degrees of freedom in polyketide biosynthesis that can be independently manipulated by genetic engineering:

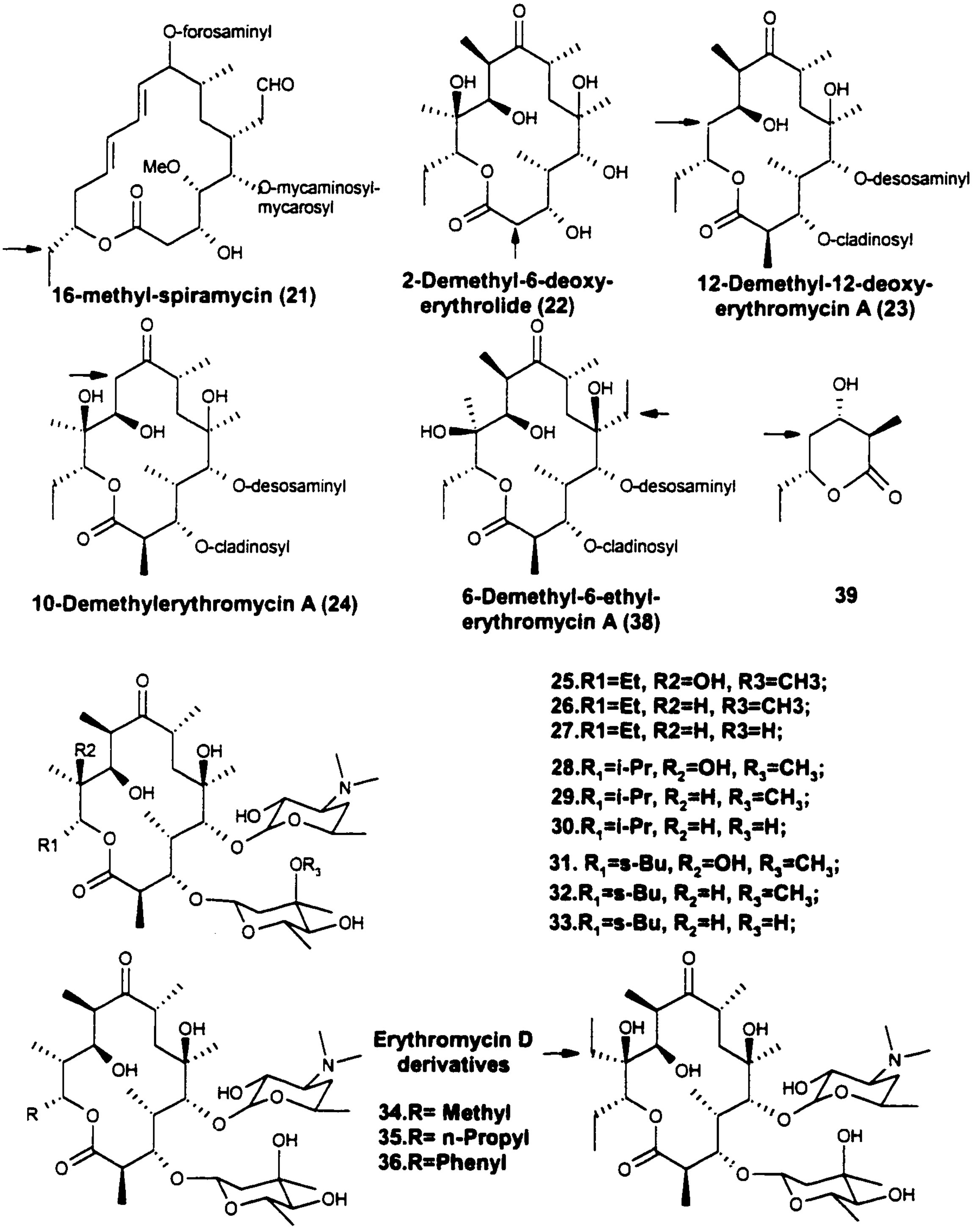
1) Substitution of AT domains to alter starter and extender unit incorporation (Scheme 8).
2) The length of the polyketide chain, which is determined by the number of modules that comprise the polyketide synthases (Scheme 9).
3) The degree of reduction of the polyketide backbone, which is determined by the set of enzyme domains present in each module (Scheme 10).
4) The stereochemistry at centers carrying alkyl and hydroxyl substituents, which is locally controlled by enzyme domains (Scheme 11).

The choice of starter unit can be varied by introducing heterologous loading modules into one of the multienzyme subunits of mutants. This first demonstrated for the biosynthesis of spiramycin when the native loading domain specific for acetyl-CoA was substituted with a loading domain that uses propionyl-CoA in the PKS enzymes for tylosin (**14**) biosynthesis was performed. The hybrid enzyme from this experiment efficiently produced the predicted homolog of spiramycin, 16-methyl-spiramycin (**21**, Scheme 8), showing that the altered chain was accommodated by all other activities of the PKS enzymes.[32b] In a more radical demonstration of this strategy, the loading domain of the erythromycin PKS was replaced by the equivalent domain from the avermectin PKS that has a broad specificity for different carboxylic acid starter units. Erythromycin derivatives **25-33** were made from branched carboxylic acids in addition to propionate in these mutants.[59]

An equally important development is the use of DEBS with an inactivated KS1 domain to make novel erythromycin analogs by biotransformation. Katz and co-workers found that the first step in erythromycin biosynthesis is the charging of KS1 with propionate directly from propionyl-CoA and that the loading domain of erythromycin polyketide synthase is not essential for erythromycin biosynthesis.[60] Some NAC derivatives of non-natural diketides were converted to novel 6-dEB (**6**) analogs with null mutants of KS1 in the mentioned host-vector system. These novel 6-dEB (**6**) analogs were then biologically transformed to novel erythromycin D derivatives **34-36**[61] and 12-ethyl-12-desmethyl erythromycin C (**37**) (Scheme 8).[62]

As already described for the loading domain, the replacement of the extender AT domains with heterologous AT domains of the PKS specific for a different type unit can generate novel natural products. The methylmalonyl-CoA-specific AT domain of module 6 of the DEBS3 was replaced with AT2 of rapamycin PKS which is specific for malonyl-CoA and led to a new compound **22** lacking a methyl group at C-2 of **6** (Scheme 8).[63] The methylmalonyl-CoA-specific AT domains of module 1 and 2 of the DEBS1 were replaced with malonyl-CoA of three hetologous AT domains and led to formation of compound **23** lacking the methyl group at C-12 of erythromycin A and compound **24** lacking the methyl group at C-10 of erythromycin A, respectively (Scheme 8).[64] The methylmalonyl-CoA-specific AT domain of module 4 of the DEBS2 was replaced with ethylmalonyl-CoA of the heterologous AT domain of niddamycin (**13**) PKS and led to the formation of 6-desmethyl-6-ethyl-erythromyicn A (**38**, Scheme 8).[65] In the mutant (DEBS1+TE), the AT domain of module 1 was replaced with the AT2 domain of rapamycin PKS, leading to the formation of the expected chimeric product, lactone **39** which lacked a specific methyl group.[66]

Scheme 8: Some new compounds made by replacement of the AT domain and KS1^0 null mutation.



16-methyl-spiramycin (21)

2-Demethyl-6-deoxy-erythrolide (22)

12-Demethyl-12-deoxy-erythromycin A (23)

10-Demethylerythromycin A (24)

6-Demethyl-6-ethyl-erythromycin A (38)

39

25.R1=Et, R2=OH, R3=CH3;
26.R1=Et, R2=H, R3=CH3;
27.R1=Et, R2=H, R3=H;

28.R$_1$=i-Pr, R$_2$=OH, R$_3$=CH$_3$;
29.R$_1$=i-Pr, R$_2$=H, R$_3$=CH$_3$;
30.R$_1$=i-Pr, R$_2$=H, R$_3$=H;

31. R$_1$=s-Bu, R$_2$=OH, R$_3$=CH$_3$;
32.R$_1$=s-Bu, R$_2$=H, R$_3$=CH$_3$;
33.R$_1$=s-Bu, R$_2$=H, R$_3$=H;

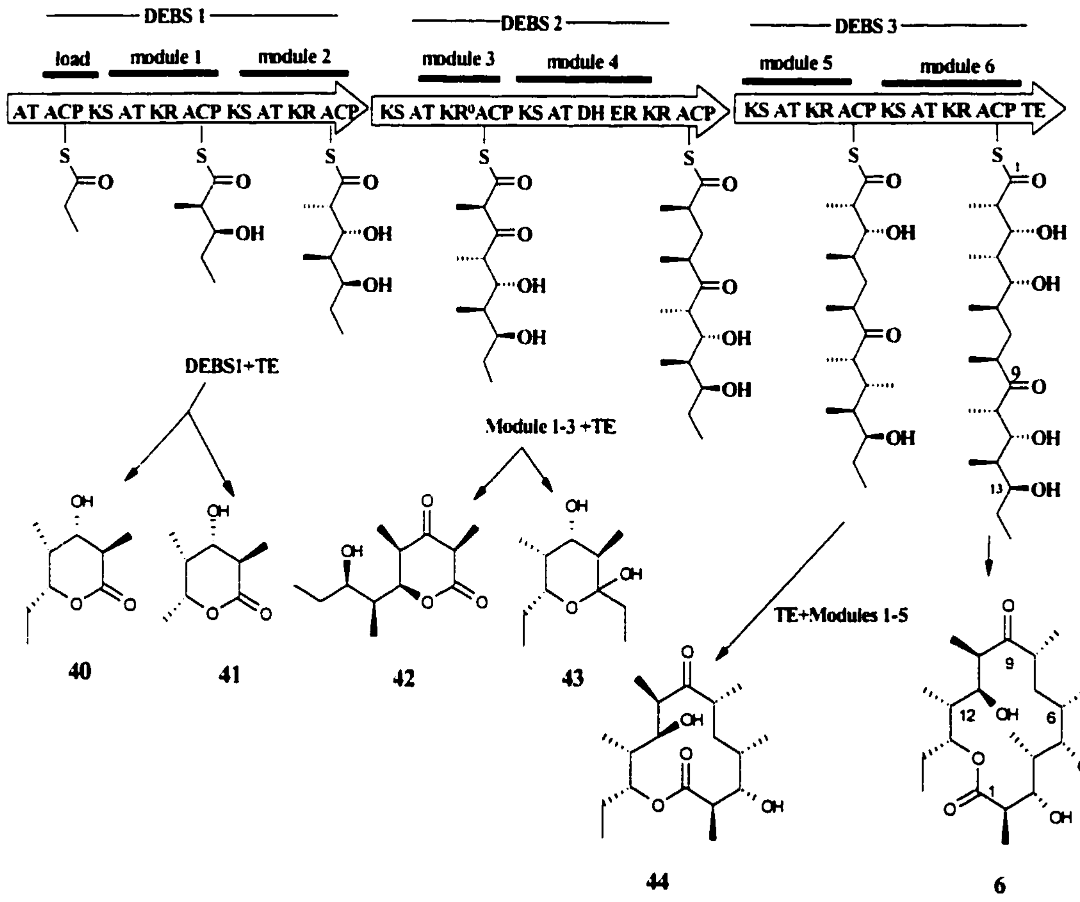
Erythromycin D derivatives

34.R= Methyl
35.R= n-Propyl
36.R=Phenyl

12-Ethyl 12-desmethyl-erythromycin C (37)

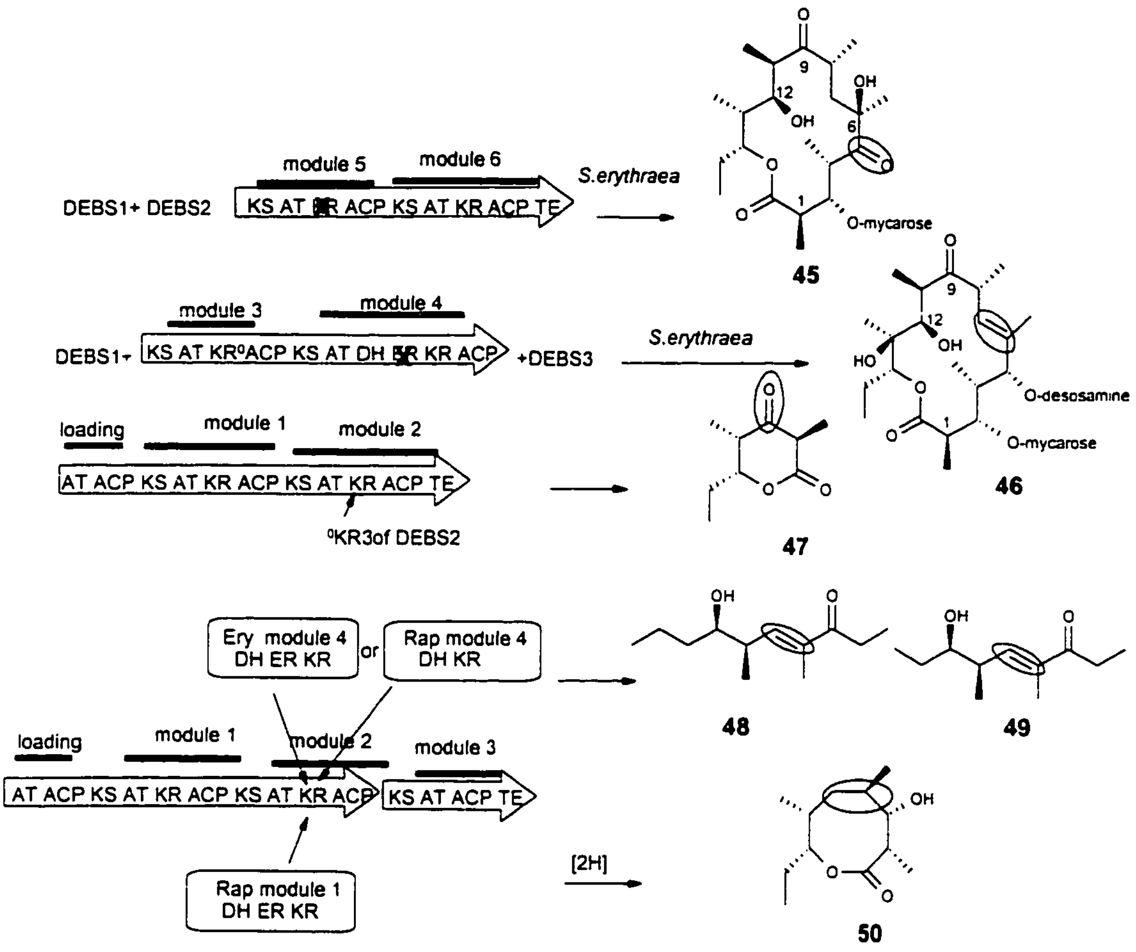
As mentioned earlier, the number of modules controls the polyketide chain length. Several truncated DEBS mutants were generated in the host/vector system.[56] For example, DEBS1+TE produced triketide lactone **40** and **41**,[67] whereas when TE was fused with modules 1-3, the tetraketides **42** and **43** were produced (Scheme 9).[68] TE was fused with modules 1-5 to produce an analog of 10-deoxymethynolide **44** (Scheme 9).[69]

Scheme 9: New compounds generated by combined different modules



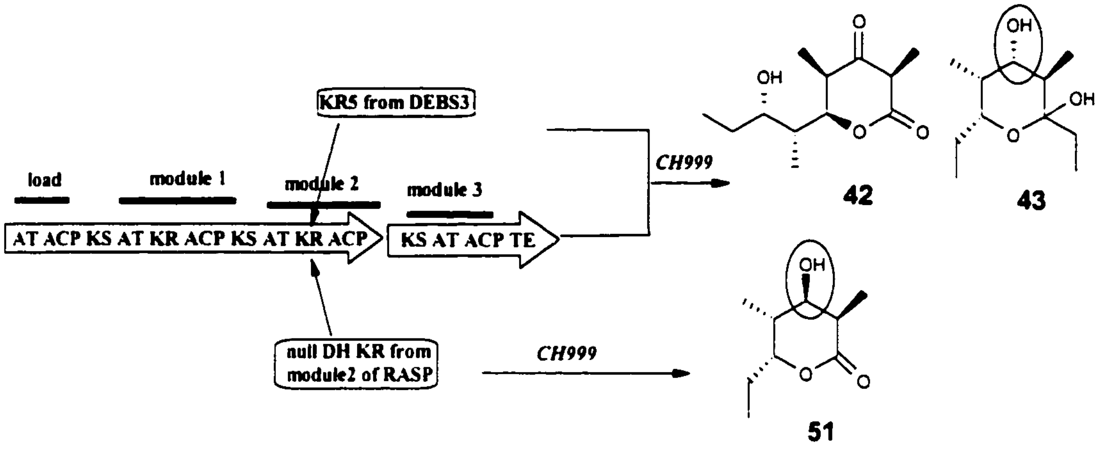
Loss and gain of β-carbonyl-processing functions will produce new compounds. Introduction of the first change into a complex PKS was reported for the KR5 domain of erythromycin PKS which was deleted by targeted alteration of the primary sequence of the protein responsible for the KR activity. This resulted in the formation of a 5-keto macrolide **45** instead of the natural 5-hydroxy analogue.[70] The ER domain in the fourth module of DEBS was deactivated by targeted alteration of the amino acid sequence of the putative NADPH binding site and the mutant produced the metabolite $\Delta^{6,7}$-anhydroerythromycin C **46** (Scheme 10).[45a] When the KR2 domain of truncated DEBS1+TE protein was replaced with dysfunctional KR^0 domain of module 3 in DEBS2, the triketide ketolactone **47** was produced (Scheme 10).[71] The different combinations of β-carbonyl-processing functions can be used to produce novel compounds, whereas when the KR2 domain in the truncated protein of module 1-3 +TE was replaced with the DH and KR domains of module 4 of rapamycin PKS or DH, ER, and KR domains of DEBS module 4, new tetraketides **48** and **49** were produced,[72] however when the KR2 domain was replaced with DH, ER and KR domains of module 1 of rapamycin PKS, the strain produced the novel eight-membered ring lactone **50** (Scheme 10).[73] This experiment indicated that the DEBS DH4-ER4-KR4 domains are only partially functional, whereas RAPS DH1-ER1-KR1 active sites are all functional in the heterologous context of DEBS's module 2. The natural truncated protein of DEBS module 1-3 +TE produces compounds **42** and **43** (Scheme 9).[67]

Scheme 10: Examples from the degree of reduction of the polyketide backbone



Khosla and coworkers reported that the KR domains of PKS enzymes control the absolute stereochemistry of the hydroxyl substituents in a growing polyketide backbone. For example, compounds **42** and **43** are the metabolites of a mutant of CH999 in which the KR2 domain was replaced with DEBS KR5 and generated the wild-type triketide since both KR2 and KR5 naturally generate (3S)-hydroxy intermediates. However, a different mutant in which the KR2 domain was replaced with null DH and KR of module 5 of rapamycin PKS, produced the (3R)-hydroxy triketide intermediate **51**, consistent with the expected stereochemistry of the rapamycin. This experiment also indicated that KS3 domain has a stringent specificity for (3S)-hydroxyl intermediate (Scheme 11).[74]

Scheme 11: Polyketides produced by replaced KR domains



A number of different mutations in different domains can be combined to generate new compounds with broad structural diversity. McDaniel *et al.* have shown how single mutations in any one module can be combined to create double or triple mutations in the 6-dEB gene cluster and generate over 100 new analogs of 6-deoxyerythronolide B (**6**). The library of 6-dEB which they were able to produce from these mutants represents about 3% of all the known polyketides. More importantly, this number exceeds the total number of different macrolide ring structures yet discovered in nature.[75] Most recently, Xue *et al.* reported that they used a three-plasmid system for the heterologous expression of 6-deoxyerythronolide B synthases (DEBSs) to facilitate combinatorial biosynthesis of polyketides. The three DEBS subunits carrying single, double or/and triple mutations were individually cloned into three compatible *Streptomyces* vectors and generated 59 analogs of **6**.[76] It is believed that these new genetic techniques will generate an even larger polyketide library in the near future.

1.2.7 Enzymes of the erythromycin polyketide synthase

Genetic approaches have been crucial in deducing the basic features of the programming rules for the modular PKSs, however, understanding how these rules are interpreted by the cellular machinery will have an even greater biochemical impact. Some very significant steps have already been taken.

Following the characterization of the DEBS genes,[44-45] Leadlay and coworkers succeed in purifying and characterizing the corresponding three DEBS proteins from the natural erythromycin producer, *S. erythraea*.[77] Recently, Khosla, Cane and coworkers used their host/vector system *S. coelicolor* CH999 to study extensively the biochemistry of DEBSs. They reported a method for cell-free enzymatic synthesis of 6-dEB and triketide using recombinant DEBSs.[78] Cane's group[79] and Leadlay's group[80] systematically studied the catalytic properties of the DEBS1 +TE with several synthesized diketide analogues because the starter residue is an attractive target for combinatorial manipulation. As mentioned earlier, the purified DEBS 1+TE multienzyme complex catalyzes synthesis of the triketide lactones **40** and **41** *in vitro* as in *vivo*, however *in vitro* it has a more relaxed specificity for the starter unit than *in vivo*. From these experiments, the following *conclusions* were drawn: 1) only diketides with (2S, 3R)-2-methyl-3-hydroxyl configuration can act as substrates *in vitro*; 2) increased bulk at C-2 is poorly accommodated by the enzymes; 3) the starter unit could be acetyl-CoA, *n*-butyryl-CoA or *iso*-butyryl-CoA.[81]

The PKS enzymes are a remarkable example of how natural synthetic methodologies can be harnessed in order to increase the pool of natural products of potential commercial value. However, there is still a significant body of knowledge that needs to be accumulated before the full power of combinatorial biosynthesis can be exploited.

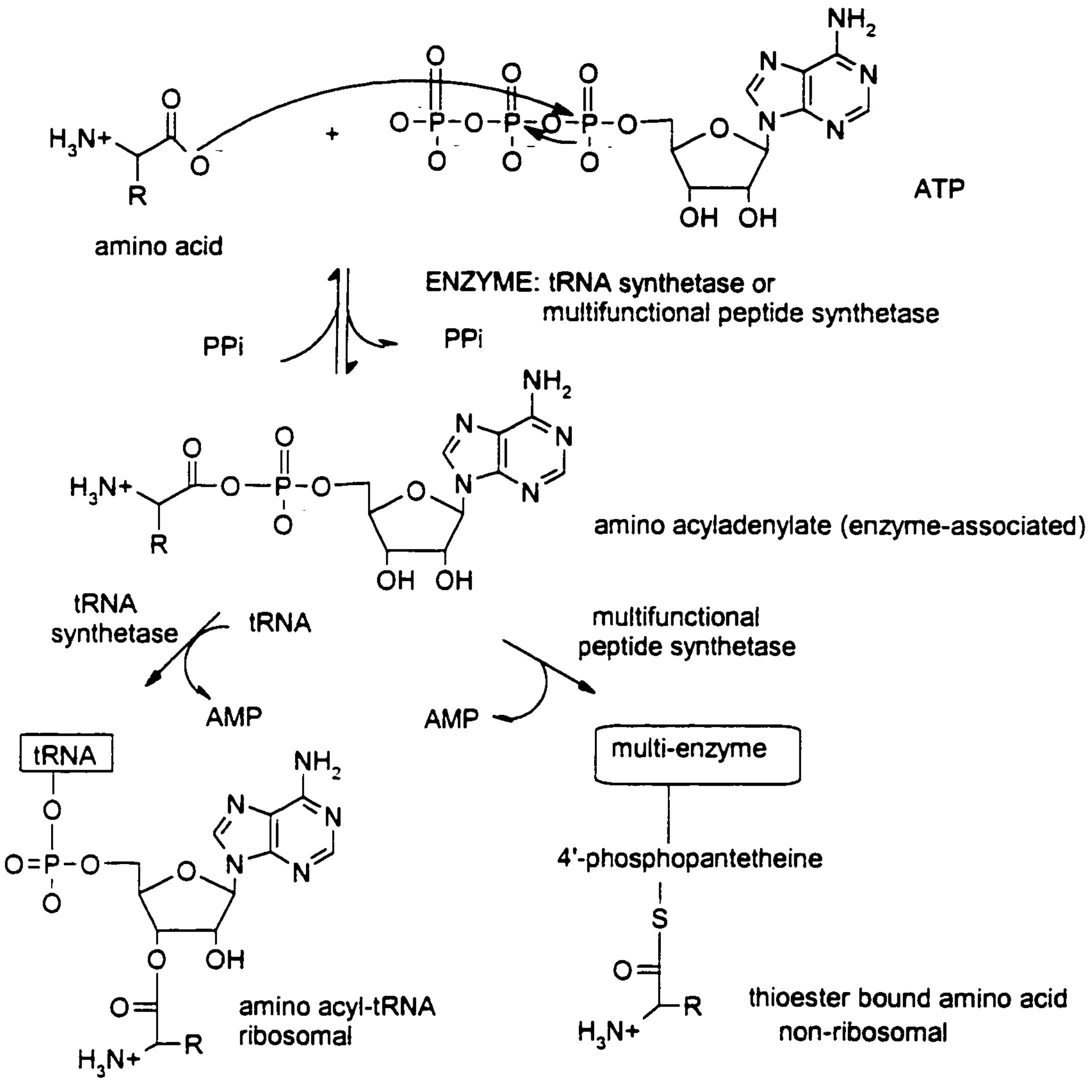
1.3 Non-ribosomal Peptides

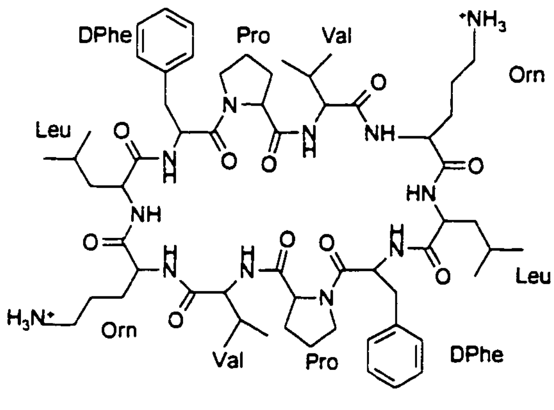
1.3.1 Introduction

Peptide biosynthesis is based on two different strategies of amino acid activation during polymerization. Most of mammalian peptide biosynthesis involves the ribosomal machinery, whereas non-ribosomal peptides are synthesized by multienzyme systems and are commonly involved in the formation of microbial secondary metabolites. The former system uses primarily the twenty proteinogenic amino acids and tRNA ligases. In the case of non-ribosomal peptide formation, more than 300 different amino acids or hydroxy acid units are known to serve as precursors. The formation of non-ribosomal peptide chains is catalyzed by large thio-template directed proteins called non-ribosomal peptide synthetases (NRPS).[82] The two known types of amino acid activation mechanisms are shown Scheme 12.[83]

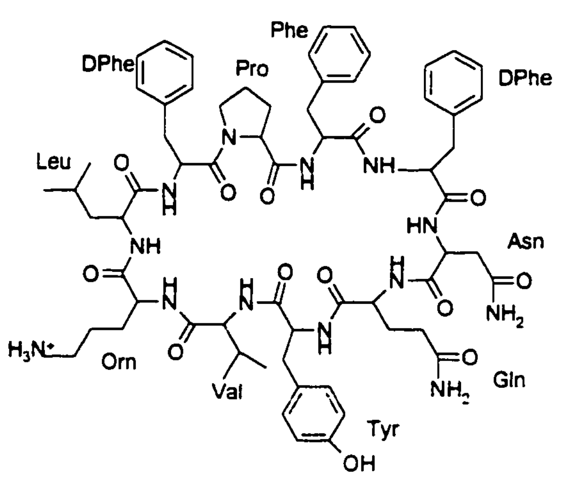
Like polyketides, non-ribosomally synthesized peptides exhibiting an impressive range of biological activities and structural variations have been isolated from bacteria, fungi, and plants.[84] Biological activities range from antimicrobial agents, such as gramicidin S (**52**), β-lactams [e.g., penicillin G (**1**)], tyrocidine A (**53**), biosurfactants [e.g., surfactin (**54**)], to immunosuppressants [e.g., cyclosporin A (**55**)].

Scheme 12: Amino acid activation mechanisms

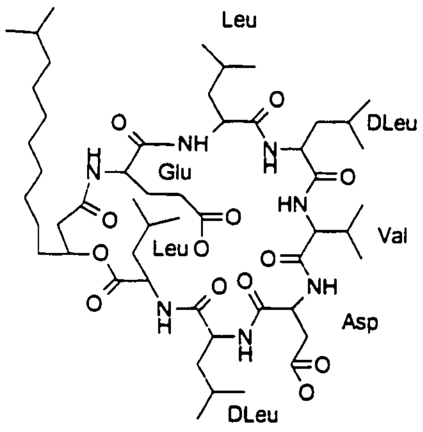




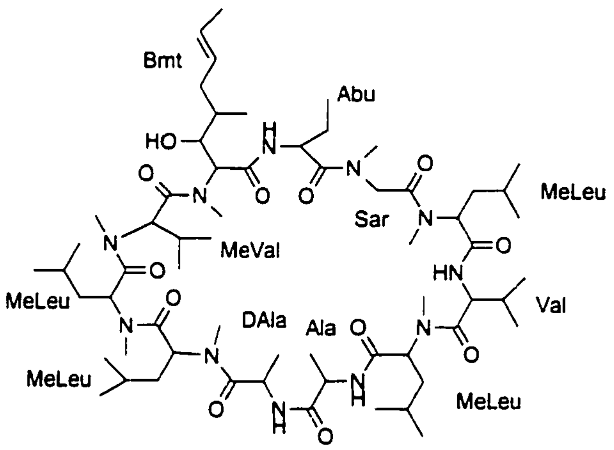
52



53



54



55

1.3.2 Genes encoding module NRPS enzymes

In contrast to their diverse structures, most non-ribosomally synthesized peptides share a uniform mode of biosynthesis: a multiple carrier thiotemplate mechanism similar to the PKS system. Analogous to the PKS, the NRPS enzymes are composed of multiple catalytic sites and a peptide of *n* amino acids would be catalyzed by *n* modules of NRPS.

A large number of bacterial operons and fungal genes encoding NRPSs have been cloned and sequenced (Table 3). Based on all known DNA sequences and their corresponding protein sequences, Marahel[85] and von Dohern[86] have identified some highly conserved motifs of NRPS. Recently, utilization of PCR technology to amplify specific sequences from genomic DNA established a convenient general approach for the identification and cloning of putative genes encoding NRPS multienzyme systems.[87] Scheme 13 and Table 4 show some examples of NRPS modules and their highly conserved amino acid sequences.

Table 3: Selection of currently studied non-ribosomal peptides.

Compounds	Gene Names	Producer Organism	Gene Cloning Method
Gramicidin (**52**)	*grs*	*Bacillus brevis*	Expression libraries by antibodies[88]
Tyrocidine (**53**)	*tyc*	*Bacillus brevis*	*grs* as probe[89]
Surfactin(**54**)	*srf*	*Bacillus subtils*	Complementation [90]
Cyclosporin(**55**)	*cssa*	*Tolypocladium niveum*	From peptide synthetases [91]
balhimycin	*aps*	*Amycolatopsis mediterranei*	PCR product of A2 .3.6.8&T[92]
Chloroeremomycin	*cep*	*Amycolatopsis orientalis*	PCR products based on TDP-glucose:aglycosyl vancomycin glucosyltransferase [93]
Bacitracin	*bac*	*Bacillus licheniforms*	PCR product of A7 &T [94]
Pristinamycins	*snb*	*Streptomyce spristinaspiralis*	From peptide synthetases[95]
lichenysin	*lic*	*Bacillus licheniformis*	PCR product of A7& T [96]
Fengycin	*fen*	*Bacillus subtils F29-3*	From peptide synthetases[97]
Actinomycin	*acm*	*Streptomyces chrysomallus*	From peptide synthetases [98]

Scheme 13: Schematic diagram showing the structure of a peptide synthetase module.

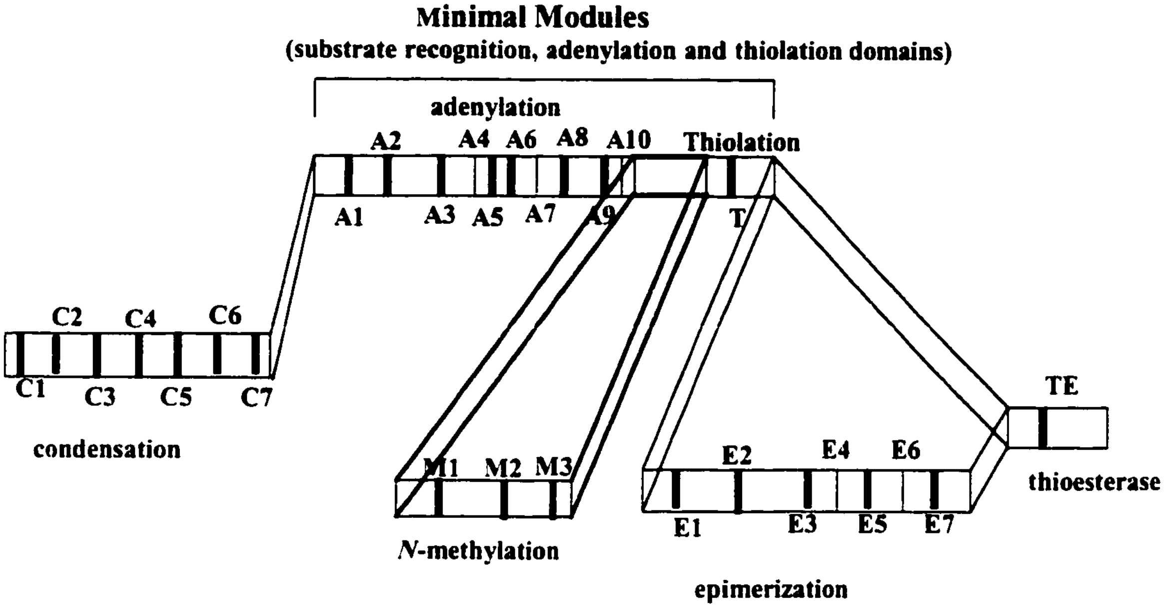


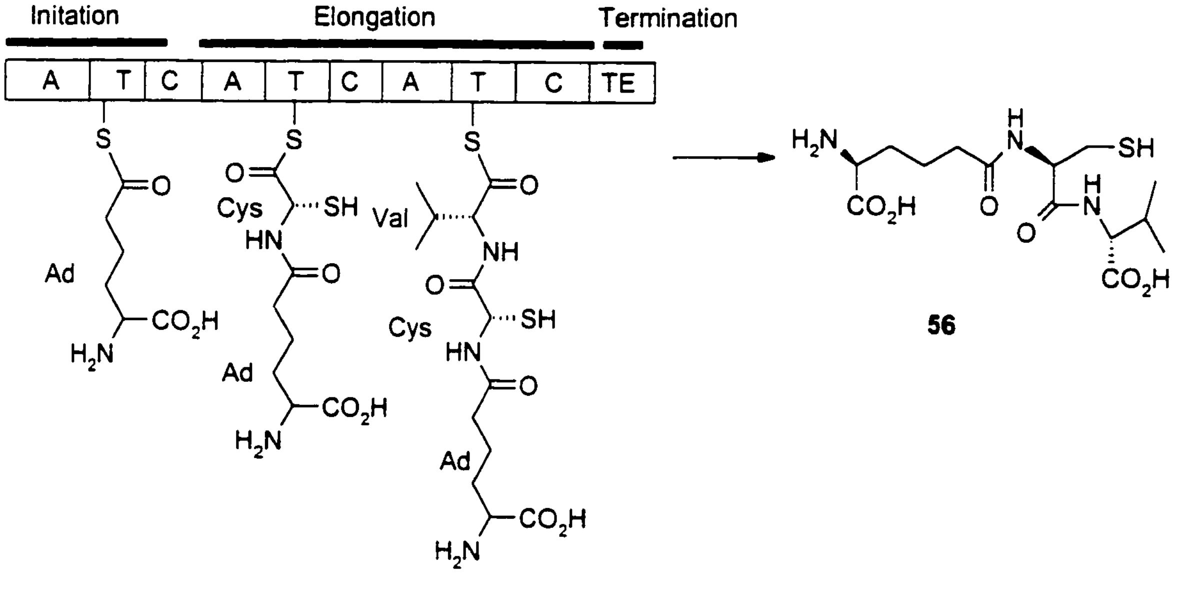
Table 4: Highly conserved core motifs of the catalytic domains of NRPS

Domain	Core	Consensus Sequence	Domain	Core	Consensus Sequence
Adenylation About 550 aa	A1	L(TS)YxEL	Condensation About 450 aa	C1	SxAQxR(LM)(WY)xL
	A2	LKAGxAYL(VL)P(LI)D		C2	RHExLRTxE
	A3	LAYxxYTSG(ST)TGxPKG		C3	MHHxISDG(WV)S
	A4	FDxS		C4	YxD(FY)AVW
	A5	NxYGPTE		C5	(IV)GxFVNT(QL)(CA)xR
	A6	GELxIxGxG(VL)ARGYL		C6	(HN)QD(YV)PFE
	A7	Y(RK)TGDL		C7	RDxSRNPL
	A8	GRxDxQVKIRGxRIELGEIE	Epimerization about 430 aa	E1	PIQxWF
	A9	LPxYM(IV)P		E2	HHxISDG(WV)S
	A10	NGK(VL)DR		E3	DxLLxAxG
Thiolation	T	DxFFxxLGG(HD)S(LI)		E4	EGHGRE
Thioesterase	TE	G(HY)SxG		E5	RTVGWFTxxYP(YV)PFE
N-methylation about 420 aa	M1	VL(DE)GxGxG		E6	PxxGxGYG
	M2	NELSxYRYxAV		E7	FNYLC(QR)
	M3	VExSxARQxGxLD	Note: thiolation about 100 aa. thioesterase about 250 aa.		

1.3.3 The functional domains of modular NRPS

Like polyketides, the mechanism for biosynthesis of non-ribosomal peptides involves an initiation step, several elongation cycles and a chain termination step. For example, Scheme 14 shows the biosynthesis of the isopenicillinic acid (**56**), a common precursor to all β-lactam antibiotics, catalyzed by ACV enzyme.[54]

Scheme 14: The biosynthesis of isopenicillic acid (**56**)



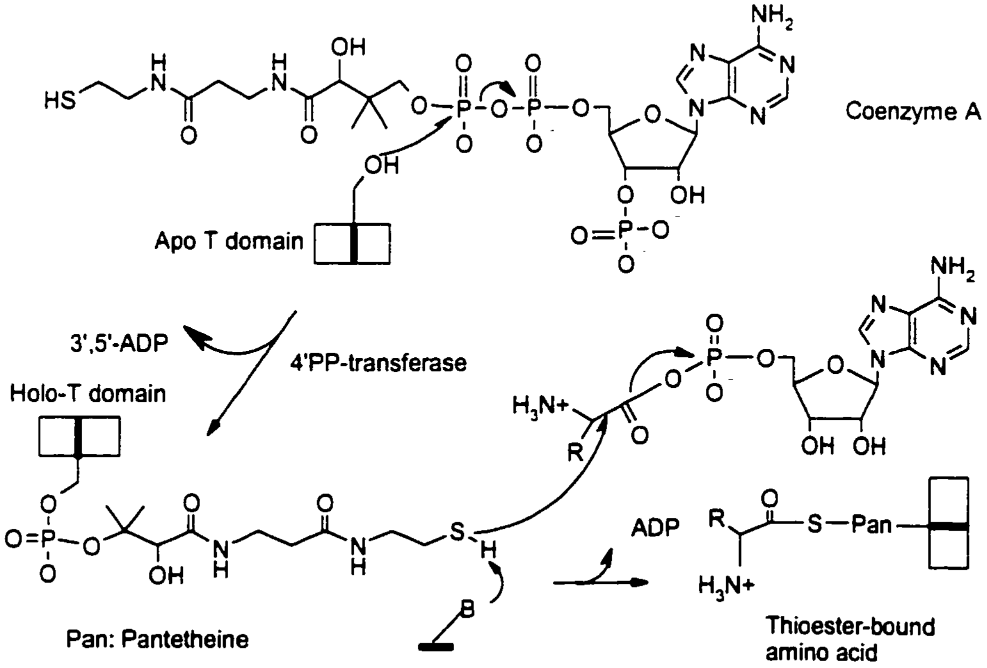
1. Adenylation (A) Domain

The adenylation domains represent the central point of action in all multifunctional peptide synthetases. Each amino acid is activated by dedicated adenylation (A) domains (Scheme 12). The core motifs of the A domain (Scheme 13, and Table 4) are the most highly conserved amino acid sequences throughout the superfamily of adenylate-formation enzymes.[99] The A3 motif is mostly disordered in the structure, but its orientation and distance to the AMP suggest an interaction with the pyrophosphateleaving group. A4 and A5 interact with the amino group via the main chain carbonyl oxygen.[100] A7 interacts with the oxygen atoms of the nucleotide ribose moiety via hydrogen bonds,[101] whereas A8 has been shown to be essential for adenylation.[102]

2. Thiolation Domain and Its Post-translational Modification (T)

The thiolation (T) domain of NRPSs (Scheme 13), also called peptidyl carrier protein (PCP), is approximately 100 amino acids long and follows the A domain. The phosphopantetheinyl transferases (PPT) promote the nucleophilic attack of the invariant serine hydroxyl group onto the pyrophosphate bridge of CoA, resulting in the transfer of the 4'-PP cofactor to the Apo-T domain and forming the Holo-T domain.[103] The activated amino acyladenylate substrates on the A domain are then transferred to the terminal cysteamine thiol group of the Holo-T domain 4'-PP cofactor (Scheme 15).[104]

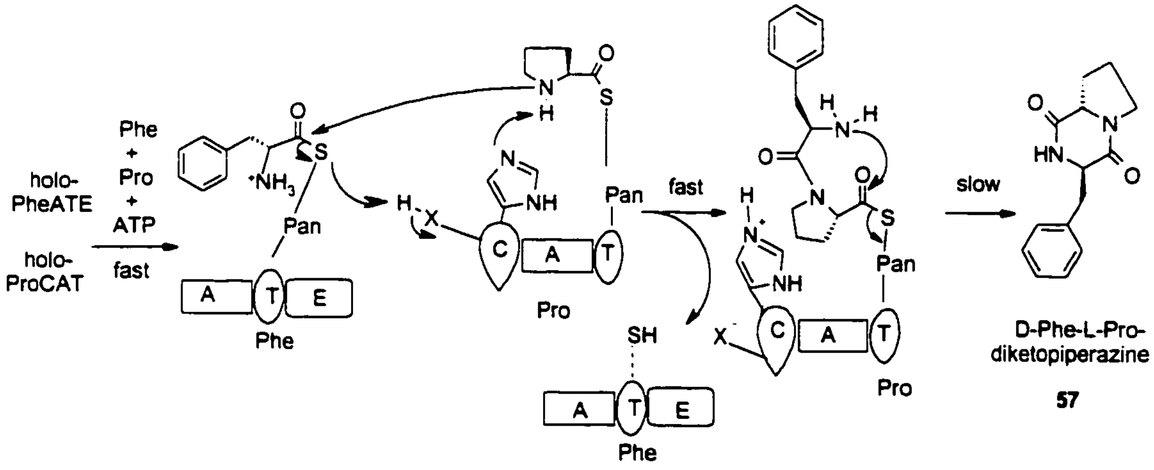
Scheme 15: Amino acid acylation of holo-T domain



3. Condensation Domain (Peptide Bond Formation)

The condensation (C) domains (Scheme 13) are about 450 amino acids in length and they are inserted between each consecutive pair of activating units within the polypeptide chain of NRPS. The highly conserved sequence C3 (HHxxxDG) is critical for peptide-bond formation. The second histidine is believed to serve as a base for deprotonation of the NH_3^+ moiety of the thioester-bound nucleophiles before peptide bond formation. Marahiel and coworkers demonstrated that *in vitro* conversion of the second histidine to valine abolishes the formation of dipeptide **57** in the biosynthesis of D-Phe-L-Pro shown in Scheme 16.[105] Besides the 'normal' position of the C domain between two modules mediating peptide-chain elongation, an extra C domain is found in several NRPS systems, such as the N-terminus of NRPS of cyclosporin (**55**). Their functions are unknown.

Scheme 16: Peptide bond formation



4. Thioesterase domain (TE)

The thioesterase domain (Scheme 13) is about 250 amino acid residues and is located near the C-terminal end of modules that are involved in adding the last amino acid to the linear peptide. It is thought that the full-length peptide bound to the last TE domain is transferred to the hydroxyl group of the highly conserved serine residue within the TE domain to generate a transient acyl-O-enzyme intermediate. This covalent species is then cleaved by an acyltransfer to water, resulting in a linear peptide, or to a functional group of a peptide side-chain liberating a cyclic or branched cyclic product. Like the TE domains in PKS systems, the TE domains in NRPSs have been shown to be important, but not always essential, for peptide synthesis, and their specific function is poorly understood.[106]

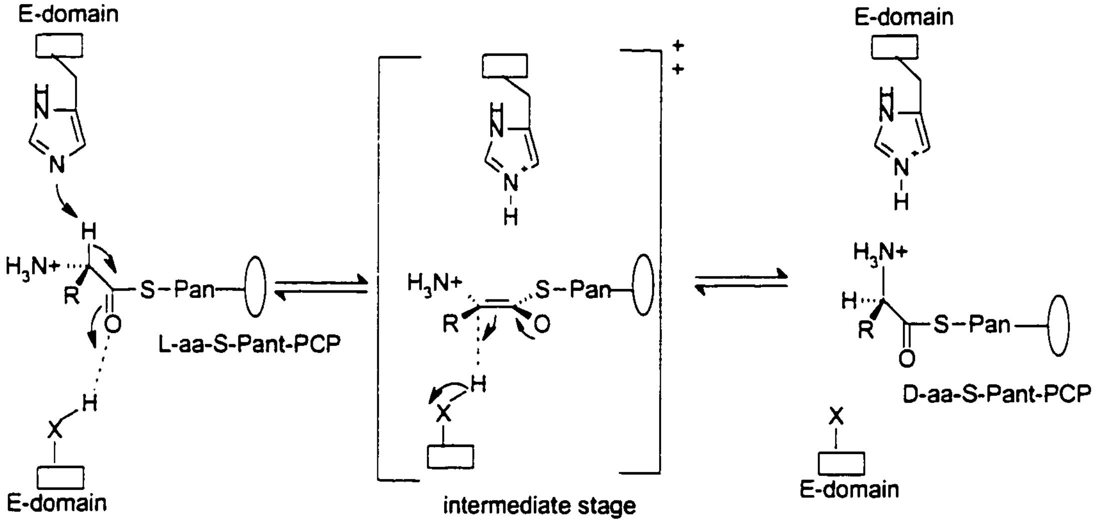
5. Modifying Domain

In addition to catalyzing the incorporation of a variety of amino acids and hydroxy acids into peptides, NRPSs can also carry out numerous modifications including *N*-methylation and site-specific epimerization. Thus, the modifying domains of peptide synthetases dramatically increase the versatility and biological activity of non-ribosomally synthesized peptides.[85]

a) Epimerization Domain (E)

Epimerization domains have been identified at different locations of some open reading frames, including at the C-terminal end of the corresponding NRPS (e.g. *grsA*[90], *srfA and* *srfB*[91]), or in the middle of an ORF (e.g. *cssa*[92]and *bac*[96]). The E2 motif [HHxISDG(I/V)VS] has a similar conserved sequence to C3 [MHHxISDG(I/V)S] (Scheme 13, Table 4). This group of enzymes may share a similar catalytic mechanism based on the acid/base properties of the second histidine residue. Epimerization involves a proton abstraction and readdition from the Cα proton of the aminoacyl or peptidyl moiety link to the cofactor 4'-PP (Scheme 17).

Scheme 17: Epimerization mechanism for an amino acid



b) ***N*-Methyltransferase Domain (M)**

N-Methylation is another modification of non-ribosomally synthesized peptides. For example, the *N*-methyltransferase domains of the enzyme involved in the biosynthesis of cyclosporin A (**55**) (seven *N*-methylated residues), enniatin (one *N*-methylated residue) and actinomycin (one *N*-methylated residue) possess 420 amino acids between the A and T domains. The glycine-rich M1 sequence shows significant similarity to the common *S*-

adenosylmethionine (SAM) binding site. The methyl groups are derived from the methyl donor SAM.

Like polyketides, new non-ribosomal peptides could be made by module organization of multifunctional NRPS which are assembled from multifunctional building domains. These advances have enabled the development of techniques for the rational design of bioactive peptides. However, there are only a few novel compounds which have been generated by genetically engineered mutants due to the structural complexity of most the non-ribosomal peptides and the few known complete DNA sequences of gene clusters associated with the biosynthesis of these peptides.[107]

1.4. References:

---

[1] Floss, H.G.; Keller, P.J.; Beale, J.M. *J. Nat. Prod.* **1986**, *49*, 957

[2] Cragg, G.M.; Newman, D.J.; Snader, K.M. *J. Nat. Prod.* **1997**, *60*, 52

[3] Vining, L.C. *Gene* **1992**, 115, 135.

[4] (a)Mann, J. *Secondary Metabolism;* Oxford Science Publications. **1987**
(b) Bu'Lock, J.D. *The biosynthesis of natural products. An introduction to secondary metabolism*; MCGraw-Hill Publishing Company Limited. **1965**

[5] (a) Croasmun, W.R.; Carlson, R.M.K. *Two-Dimensional NMR spectroscopy. Application for chemists and biochemists;* VCH Publishers. **1987**
(b) Sanders, J.K.M.; Hunter, B.K. *Modern NMR Spectroscopy*; 2nd. Oxford University Press. **1993**

[6] Gunther, H. *NMR Spectroscopy: Basic principles, concepts, and applications in chemistry.* Second ed. John Wiley & Sons. **1995** .

[7] (a) Vederas, J.C. *Nat. Prod. Rep.* **1987**, *4*, 277
(b) Floss, H.G.; Beale, J.M. *Angew. Chem. Int. Ed. Engl.* **1989**, *28*, 146

[8] Lane, M.P.; Nakashima, T.T.; Vederas, J.C. *J. Am. Chem. Soc.* **1982**, *104*, 913

[9] (a) Flynn, E.H.; Sigal, M.V.; Wily, P.F.; Gerzon, K. *J. Am. Chem. Soc.* **1954**, *76*, 3121
(b) Vasquez, D. *" Antibiotics III. Mechanism of Action of Antimicrobial and Antitumor Agents "*, Corcoran, J. W. and Hahn, F.E. ed. Springer, New York, **1975**

[10] Corcoran, J.W.; Kaneda, T.; Butte, J.C. *J. Biol. Chem.* **1960,** *235*, PC29
Grisebach, H.; Achenbach, H.; Hofheinz, W. *Z.Naturforsch. B.* **1960**, *15B*, 560

[11] Cane, D.E.; Hasler, H.; Taylor, P. B.; Liang, T. C. *Tetrahedron* **1983**, *39*, 3449

[12] Cane, D.E.; Hasler, H.; Liang, T.C. *J. Am. Chem. Soc.* **1981**, *103*, 5960

[13] Cane, D.E.; Liang, T.C. ; Taylor, P.B.; Chang, C.; Yang, C-C. *J. Am. Chem. Soc.* **1986**, *108*, 4957

[14] (a) Cane, D.E.; Yang, C-C. *J. Am. Chem. Soc.* **1987**, *109*, 1255
(b) Cane, D.E.; Prabhakaran, P.C.; Tan, W.; Ott, W.R. *Tetrahdron. Lett.* **1991**, *32*, 5457

[15] Cane, D.E.; Lambalot, R.H.; Prabhakaran, P.C.; Ott, W.R. *J. Am. Chem. Soc.* **1993**, *115*, 522

[16] (a) Cane, D.E.; Ott, W.R. *J. Am. Chem. Soc.* **1988,***110*, 4840
(b) Cane, D.E.; Luo, G. *J. Am. Chem. Soc.* **1995,***117*, 6633

[17] Staunton, J.; Sutkowski, A.C. *J. Chem. Soc. Chem. Commun.* **1991**, 1110

[18] Spavold, Z.M.; Robinson, J.A. *J. Chem. Soc. Chem. Commun.* **1988**, 4

[19] Katz, L. *Chem. Rev.* **1997**, *97*, 2557

[20] Birch, A.J.; Donovan, F.W. *Aust. J. Chem.* **1953**, *6*, 360

[21] Wakil, S.J. *Biochemistry* **1989**. *28*, 4523

[22] Walsh, C.T.; Gehring, A.M.; Weinreb,P.H.; Quadri, L.E.; Flugel, R.S. *Curr. Opin, in Chem. Biol.* **1997**, *1*, 309

[23] Hutchinson, C.R.; Fujii , I. *Annu. Rev. Microbiol.* **1995**, *49*, 201

[24] (a) Gould, S.J. *Chem.Rev.* **1997**, *97*, 2499
(b) Fuiji, S.; Ebizuka, Y. *Chem.Rev.* **1997**, *97*, 2511
(c) Hutchinson, C.R. *Chem.Rev.* **1997**, *97*, 2525

[25] (a) Bibb, M.J.; Schottel, J.L.; Cohen, S.N. *Nature* **1980,** *284*, 526
(b)Thompson, C.J.; Ward, J.M.; Hopwood, D.A. *Nature* **1980**, *286*, 525

(c) Suarez,J.E.; Chaterr, K.F. *Nature* **1980**, *286*, 527

[26] Hopwood, D.A.; Bibb, M.J.; Bruton,C.J.; Chater, K.F.; Feitelson, J.S.; Gil, J.A. *Trends Biotechnol.* **1983**, *1*, 42

[27] Hopwood, D.A. *Chemical Reviews* **1997**. *97*, 2463.

[28] (a) Motamedi, H.; Shafiee,A. *Eur. J. Biochem.* **1998**, *256*, 528

(b) Motamedi, H.; Cai, S-J.; Shafiee, A.; Elliston, K.O. *Eur. J. Biochem.* **1997**, *244*, 74

[29] Schwecke, T.; Aparicio, J. F.; Molnár, I.; König, A.; Khaw, L. E.; Haydock, S. F.; Oliynyk, M.; Caffrey, P.; Cortés, J.; Lester, J. B.; Böhm, G. A.; Staunton, J.; Leadlay. P. F. *Proc. Natl. Acad. Sci. U.S.A.* **1995**, *92*, 7839

[30] (a) Motamedi, H.; Hutchinson, C.R. *Proc .Natl. Acad. Sci.USA.* **1987**, *84*, 4445

(b) Summers, R.G.; Wendt-Pienkowski, E.; Motamedi, H.; Hutchinson, C.R. *J. Bacteriol.* **1992**, *174*, 1810

[31] (a)Swan, D.G.; Rodriguez, A.M.; Vilches, C.; Mendez, C.; Salas, J.A. *Mol. Gen. Genet.* **1994**, *242*, 358

(b) Quiros, L.M.; Aguirrezabalaga, I.; Olano, C.; Mendez, C.; Salas, J.A. *Mol. Microbiol.* **1998**, *28*, 1177

[32] (a) Fishman, S.E.; Cox, K.; Larson, J.L.; Reynolds, P.A.; Seno, E.T.; Yeh, W.K.; van-Frank, R.; Hershberger, C.L. *Proc. Natl. Acad. Sci. USA.* **1987**, *84*, 8248

(b) Kuhstoss, S.; Huber, M.; Turner, J.R.; Paschal, J.W.; Rao, R.N. *Gene* **1996**, *183*, 231

[33] Richardson, M.A.; Kuhstoss, S.; Huber, M.; Ford, L.; Godfrey, O.; Turner, J.R.; Rao, R.N. in *Genetics and Molecular Biology of industrial Microorganisms*; Hersherger, C.L., Queener, S.W., Hegeman, G., Eds., American Society for Microbiology. Washington, DC, **1989**, 40

[34] Schupp, T.; Toupet, C.; Cluzel, R.; Neff, S.; Hill, S.; Beck, J.J.; Ligon, J.M. *J. Bacteriol.* **1995**, *177*, 3673

[35] (a) MacNeil, D.J.; Occi, J.L.; Gewain, K.M.; MacNeil, T.; Gibbons, P.H.; Ruby, C.L.; Danis, S.J.; *Gene* **1992**, *115*, 119

(b) Ikeda, H.; Nonomiya, T.; Usami, M.; Ohta, T.; Omura, S. *Proc. Natl. Acad. Sci. USA.* **1999**, *96*, 9506

[36] (a) August, P.R.; Tang, L.; Yoon,Y.J.; Ning, S.; Muller, R.;Yu, T-W.; Taylor, M.; Hoffmann, D.; Kim, C-G.; Zhang, X.; Hutchinson, C.R.; Floss, H.G. *Chem. Biol.* **1998**, *5*, 69

(b) Tang, L.; Yoon,Y.J.; Choi, C-Y.; Hutchinson, C.R. *Gene* **1998**, *216*, 255

[37] Xue, Y.; Zhao, L.; Liu, H-W.; Sherman, D.H. *Proc. Natl. Acad. Sci. USA.* **1998**, *95*, 12111

[38] Kakavas, S.J.; Katz, L.; Stassi, D. *J.Bacteriol.* **1997**, *179*, 7515

[39] (a) Malpartida, F.; Hopwood, D.A. *Nature* **1984**, *309*, 462

(b) Hallam, S.E.; Malpartida, F.;Hopwood, D.A. *Gene* **1988**, *74*, 305

(c) Fernandez-Moreno, M.A.; Martinez, E.; Boto, L.; Hopwood, D.A.; Malpartida, F. *J.Biol. Chem.* **1992**, *267*, 19278

[40] Bibb, M.J.; Sherman, D.H.; Omura, S.; Hopwood, D.A. *Gene* **1994**, *142*, 31

[41] Yu, T-W.; Bibb, M.J.; Revill, W.P.; Hopwood, D.A. *J. Bacteriol.* **1994**, *176*, 2627

[42] Ye, J.; Dickens, M.L.; Plater, R.; Li, Y.; Lawrence, J.; Strohl, W.R. *J. Bacteriol.* **1994**, *176*, 6270

---

[43] (a) McGuire,J.M.; Bunch, R.L.; Aderson, R.C.; Boaz, H.E.; Flynn, E.H.; Powell, M.; Smith, J.W. *Antibiot, Chemother.* **1952,** *2,* 281
(b)Wiley, P.F.; Gerzon, K.; Flynn, E.H.; Sigal, M.V. Jr.; Weaver, Q.; Quarck,U.C.; Chauvette, R.R.; Monahan, R. *J. Am. Chem. Soc.* **1957,** *79,* 6062

[44] (a) Caffrey, P.; Green, B.; Packman, L. C.; Rawlings, B. J.; Staunton, J.; Leadlay, P. F. *Eur. J. Biochem.* **1991,** *195,* 823
(b) Caffrey, P.; Bevitt, D.J.; Staunton, J.; Leadlay, P.F. *FEBS. Lett.* **1992,** *304,* 225
(c) Bevitt, D.J.; Cortes, J.; Haydock, S.F.; Leadlay, P.F. *Eur. J. Biochem.* **1992** *204,* 39
(d) Cortes, J.; Haydock , S.F.; Roberts, G.A.; Bevitt, D.J.; Leadlay, P.F. *Nature* **1990,** *348,* 176
(e) Haydock , S.F.; Dowson, J.A.; Dhillon, N.; Roberts, G.A.; Cortes, J.; Leadlay, P.F. *Mol. Gen. Genet.* **1991,** *230,* 120

[45] (a) Donadio, S.; Staver, M.J.; McAlpine, J. B.; Swanson, S.J.; Katz, C. *Science* **1991,** *252,* 675
(b) Donadio, S.; Katz, C. *Gene* **1992,** *111,* 51

[46] Weber, J.M.; Leung, J.O.; Maine, G.T.; Potent, P.B.H.; Paulus, T.J.; Dewitt, J.P. *J. Bacteriol.* **1990,** 172, 2372

[47] Weber, J.M.; Leung, J.O.; Swanson, S.J.; Idler, K.B.; McAlpoine, J.B. *Science,* **1991,** *252,* 114

[48] (a) Summers, R.G.; Donadio, S.; Staver, M.J.; Wendt-Pienkowski, E.; Hutchinson, C.R.; Katz, L. *Microbiology.* **1997,** *143,* 3251
(b) Gaisser, S.; Bohm, G.A.; Cortes, J.; Leadlay, P.F. *Mol. Gen. Genet.* **1997,** *256,* 239
(c) Gaisser, S.; Bohm, G.A.; Doumith, M.; Raynal, M.-C.; Dhillon, N.; Cortes, J.; Leadlay, P.F. *Mol. Gen. Genet.* **1998,** *258,* 78

[49] Stassi, D.L.; Donadio, S.; Staver, M.J.; Katz, L. *J. Bacteriol.* **1993,** *175,* 182

[50] (a) Weber, J.M.; Schoner, B.; Losick, R. *Gene* **1989,***75,* 235
(b) Paulus, T.J.; Tuan, J.S.; Luebke, V.E.; Maine, G.T.; DeWitt, J.P.; Katz, L. *J. Bacteriol.* **1990,** *172,* 2541

[51] (a) Katz, L.; Donadio, S. *Ann. Rev. Microbiol.* **1993,** *47,* 875
(b) Haydock, S.; Aparico, J.F.; Molnar, I.; Schwecke, T.; Konig, A.; Marsden, A.F.A.; Galloway, I.S.; Staunton, J.; Leadlay, P.F. *FEMS Lett.* **1995,** *374,* 246

[52] Hutchinson, C.R. *Curr. Opin. Microbiol.* **1998.** *1.* 319

[53] Leadlay, P.F. *Curr. Opin. Chem. Biol.* **1997.** *1.* 162

[54] Cane, D.E.; Walsh, C.T.; Khosla, C. *Science* **1998,** *282,* 63

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

[56] Kao, C.M.; Katz, L.; Khosla, C. *Science* **1994,** *265,* 509

[57] Khosla, C.; Ebert-Khosla, S.; Hopwood, D.A. *Mol. Microbiol.* **1992,** *6,* 3237

[58] Hopwood, D. A.; Bibb, M. J.; Chatter, K. F.; Kisser, T.; Burton, C. J.; Kisser, H. M. K.; Lydia, D. J.; Smith, C. P.; Ward, J. M.; Schrempf, H. Genetic Manipulation of *Streptomyces*: a laboratory manual, The John Innes Foundation. 1985.

[59] Marsden, A.F.A.; Wilkinson, B.; Cortes, J.; Dunster, N.J.; Staunton, J.; Leadlay, P.F. *Science* **1998,** *279,* 199

[60] Pereda, A.;Summers, R.G.; Stassi, D.L.; Ruan, X.; Katz, L. *Microbiology* **1998,***144,*

543

[61] Jacobsen, J.R.; Hutchinson, C.R.; Cane, D.E.; Khosla, C. *Science* **1997**, *277*, 367

[62] Jacobsen, J.R.;Keatinge-Clay,A.T.;Cane, D.E.; Khosla, C. *Bioorg. Med. Chem.* **1998**, *6*, 1171

[63] Liu, L.; Thamchaipenet, A.; Fu, H.;Beltach, M.; Ashley, G. *J. Am.Chem.Soc.* **1997**, *119*, 10553

[64] Ruan, X.; Pereda, A.; Stassi, D.L.; Zeidner, D.;Summers, R.; Jackson, M.; Shivakumar, A.; Kkakavas, S.; Staver, M.J.; Donadio, S.; Katz, L. *J. Bacteriol.* **1997**, *179*, 6416

[65] Stassi, D.L.; Kakavas, S.J.; Reynold, K.A.; Gunawardana, G.; Swanson, S.; Zeidner, S.; Jackson,M.; Buko, H.L.; Katz, L. *Proc. Natl. Acad. Sci.USA.* **1998**, *95*, 7305

[66] Oliynyk, M.; Brown, M.J.B.; Cortes, J.; Staunton, J.; Leadlay, P.F. *Chem. Biol.* **1996**, *3*, 833

[67] (a) Cortes, J.; Wiesmann, K.E.H.; Roberts, G.A.; Brown, M.J.B.; Staunton, J.; Leadlay, P.F. *Science* **1995**, *268*, 1487

(b) Kao, C.M.; Luo, G.; Katz, L.; Cane, D.E.; Khosla,C. *J. Am. Chem. Soc.* **1994**, *116*,11612

[68] Kao, C.M.; Luo, G.; Katz, L.; Cane, D.E.; Khosla, C. *J. Am. Chem. Soc.* **1996**, *118*, 9184

[69] Kao, C.M.; Luo, G.; Katz, L.; Cane, D.E.; Khosla, C. *J. Am. Chem. Soc.* **1995**, *117*, 9105

[70] Donadio,S.; Staver,M.J.; McAlpine,J.B.; Swanson, S.J.; Katz, L. *Science* **1991**, *252*, 675

[71] Bedford,D.; Jacobsen, J.R.; Luo,G.; Cane,D.E.; Khosla. C. *Chem. Biol.* **1996**, *3*, 827

[72] McDaniel, R.; Kao,C.M.; Fu, H.;Hevezi, P.; Gustafsson, C.; Betlach, M.; Ashley,G.,Cane, D.E.; Khosla,C. *J. Am.Chem.Soc.* **1997**, *119*, 4309

[73] Kao,C.M.; McPherson,M.; McDaniel,R.; Fu,H.; Cane, D.E.; Khosla,C. *J.Am.Chem.Soc.* **1997**, *119*, 11339

[74] Kao,C.M.; McPherson,M.; McDaniel,R.; Fu,H.; Cane, D.E.; Khosla, C. *J. Am. Chem. Soc.* **1998**, *120, 2478*

[75] McDaniel, R. M.; Thamchaipenet, A..; Gustafsson, C.; Fu, H.; Betlach, M..; Betlach, M.; Ashley, G. *Proc. Natl. Acad. Sci. USA* , **1999**, *96*, 1846

[76] Xue, Q.; Ashley, G.; Hutchinson, C. R.; Santi, D. V. *Proc. Natl. Acad. Sci. USA.* **1999**, *96*: 11740

[77] (a) Apricio, J.F.; Caffrey, P.; Marsden, A.F.A.; Staunton, J.; Leadlay, P. *J. Biol. Chem.* **1994**, *269*, 8524

(b) Roberts, G.A.; Staunton, J.; Leadlay, P.F.*Eur. J. Biochem.* **1993**, *214*, 305

[78] (a) Pieper, R.; Luo, G.; Cane, D.E.; Khosla, C. *Nature* **1995**, *378*, 263

(b) Pieper, R.; Ebert-Khosla, S.; Cane, D.E.; Khosla, C. *Biochemistry* **1996**, *35*, 2054

[79] (a) Chuck, J-A.; McPherson, M.; Huang, H.; Jacobsen, J.R.; Khosla, C.; Cane, D.E. *Chem. Biol.* **1997**, *4*, *757*

(b) Luo, G.; Pieper, R.; Rosa, A.; Khosla, C.; Cane, D.E. *Bioorg. Med. Chem.* **1996**, *4*, 995

[80] Weissman, K.J.; Bycroft, M.; Cutter, A.L.; Hanefeld, U.; Frost, E.J.; Timoney, M.C.; Harris, R.; Handa, S.; Roddis M.; Staunton, J.; Leadlay, P.F. *Chem. Biol.***1998**, *5,743*

---

[81] Pieper, R.;Luo, G.; Cane, D.E.; Khosla, C. *J. Am. Chem. Soc.* **1995**, *117*, 11373

[82] Kleinkauf, H.; von Dohren, H. *Eur. J. Biochem.* **1996**, *236*, 335

[83] Stachelhaus,T.; Schneider, A.; Marahiel, M. A. *Biochemical Pharmacology* **1996**, *52*,177

[84] (a) Marahiel, M.A.; Stachelhaus, T.; Mootz, H.D., *Chem. Rev.* **1997**, *97*,2651
(b) von Dohren, H.;Keller, U.; Vater, J.; Zocher, R. *Chem. Rev.* **1997**, *97*,2675

[85] (a) Stachelhaus, T.; Marahiel, M.A. *FEMS Microbiol. Lett.* **1995**, *125*, 3
(b) Marahiel, M.A. *Febs J*, **1992**, *307*, 40

[86] Kleinkauf, H.; von dohern, H. *Eur. J. Biochem.* **1990**, *192*,1

[87] (a) Borchert, S.; Patil, S.S.; Marahiel, M.A. *FEMS Microbiol. Lett.* **1992**, *92*, 175
(b) Turgay, K.; Marahiel, M. *Peptide Res.* **1994**, *7*, 238

[88] Turgay, K.; Krause, M.; Marahiel, M.A. *Mol.Microbiol.* **1992**, *6*. 529

[89] Mootz, H.D.; Marahiel, M.A. *J. Bacteriol.* **1997**, *179*, 6843

[90] (a) Nakano, M.M.; Zuber, P. *J. Bacteriol.* **1989**, *171*, 5347
(b) Cosmina, P.; Rodriguez, F.; de Ferra, F.; Grandi, G.; Perego, M.; Venema,G.; van Sinderen, D. *Mol. Microbiol.* **1993**, *8*, 821

[91] Weber, G.; Schorgendorfer, K.; Schnider-Scherzer, E.; Leitner, E. *Curr. Genet.* **1994**, *26*, 120

[92] Pelzer, S.; Reichert, W.; Huppert, W.; Heckmann, D.; Wohlleben, W. *J. Biotechnol.* **1997**, *56*, 115

[93] (a) van Wageningen, A.M.; Kirkpatrick, P.N.; Williams, D.H.; Harris, B.R.; Kershaw, J.K.; Lennard, N.J.; Jones, M.; Jones, S.JM.; Solenberg, P.J. *Chem. Biol.* **1998**, *5*, 155
(b) Solenberg, P.; Matsushima, P.; Stack, D.R.; Wilkie, S.C.; Thompson, R.C.; Baltz, R.H. *Chem. Biol.* **1997**, *4*, 195

[94] Konz, D.; Klens, A.; Schorgendorfer, K.; Marahiel, M.A. *Chem. Biol.* **1997**, *4*, 927

[95] de Crecy-Lagard, V.; Blanc, V.; Gil, P.; Naudin, L.; Lorenzon, S.; Famechon, A.; Bamas-Jacques, N.;Crouzet, J.; Thibaut, D. *J. Bacteriol.* **1997**, *179*, 705

[96] Konz, D.; Doekel, S.; Marahiel, M.A. *J.Bacteriol*, **1999**, *181*, 133

[97] Steller, S.; Vollenbroich, D.; Leenders, F.; Stein, T.; Conrad, B.; Hofemeister, J.; Jacques, P.; Thonart, P., Vater, *Chem. Biol.* **1998**, *6*, 31

[98] Schauwecker, F.; Pfennig, F.; Schroder,W.; Keller, U. *J. Bacteriol.* **1998**,*180*, 2468

[99] Conti, E.; Stachelhaus, T.; Marahiel, M.A., Brick, P. *EMBO J*, **1997**, *16*, 4174

[100] Saito, M.; Hori, K.; Kurotsu, T.; Kanda, M.; Saito, Y. *J. Biochem.* **1995**, *117*, 276

[101] Gocht, M.; Marahiel, M. *J. Bacteriol.* **1994**, *176*, 2654

[102] Tokita, K.; Hori, K.; Kurotsu, T.; Kanda, M.; Saito,Y. *J. Biochem.* **1993**, *114*, 522

[103] Leenders, F.; Vater, J.; Stein, T.; Frank, P. *J. Biol. Chem.* **1998**, *273*, 18011

[104] Lambalot, T.R.H.; Gehring, A. M.; Flugel, R.S.; Zuber, P.; LaCelle, M.; Mariahiel, M.A.; Reid, R.; Khosla, C.; Walsh, C.T. *Chem. Biol*, **1996**, *3*, 923

[105] Stachelhaus, T.; Mootz, H.D.; Bergendahl. V.; Marahiel, M.A. *J. Biol. Chem.* **1998**, *273*, 22773

[106] Guenzi, E.; Gall, G.; Grgurina, I.; Pace, E.; Ferranti, P.; Grandi, G. *J. Biol. Chem.* **1998**, *273*, 14403

[107] (a) Stachelhaus, T.; Schneider, A.; Marahiel, A.M. *Science.* **1995**, *269*, 69
(b) Schneider, A.; Stachelhaus, T.; Marahiel, A *Mol.Gen. Genet.* **1998**, *257*, 308

Chapter 2: Studies on the Biosynthesis of the Fungal Metabolite Oudenone II. Synthesis and Enzymatic Cyclization of an α-Diketone, Open-Chain Precursor into Oudenone in Cultures of *Oudemansiella radicata*.

[This part includes two published papers: 1) Tsantrizos, Y.S.; Yang, X.-S.; McClory, A. *J. Org.Chem.* **1999,** 64, 6609-6614 and 2)Tsantrizos Y.S.; Zhou, F.; Famili, P.; Yang, X.-S. *J. Org.Chem.* **1995.** 60. 6922-6969]

2.1 General Introduction

Over the last ten years, polyketide synthase enzymes associated with the biosynthesis of macrolides [PKS Type I, e.g., erythromycin A (**2**),[44-45] FK506 (**3**),[28] rapamycin (**4**),[29] niddamycin (**11**),[38] FR-008[108]], and aromatic polyketides [PKS Type II, e.g., tetracenomycin C,[30] actinorhodin (**5**),[39] frenolicin (**16**),[40] aflatoxins[109]] have been identified and studied extensively. Furthermore, major advances in the field of microbial genetics have illuminated our knowledge and ability to manipulate the gene clusters encoding these enzymes.[54]

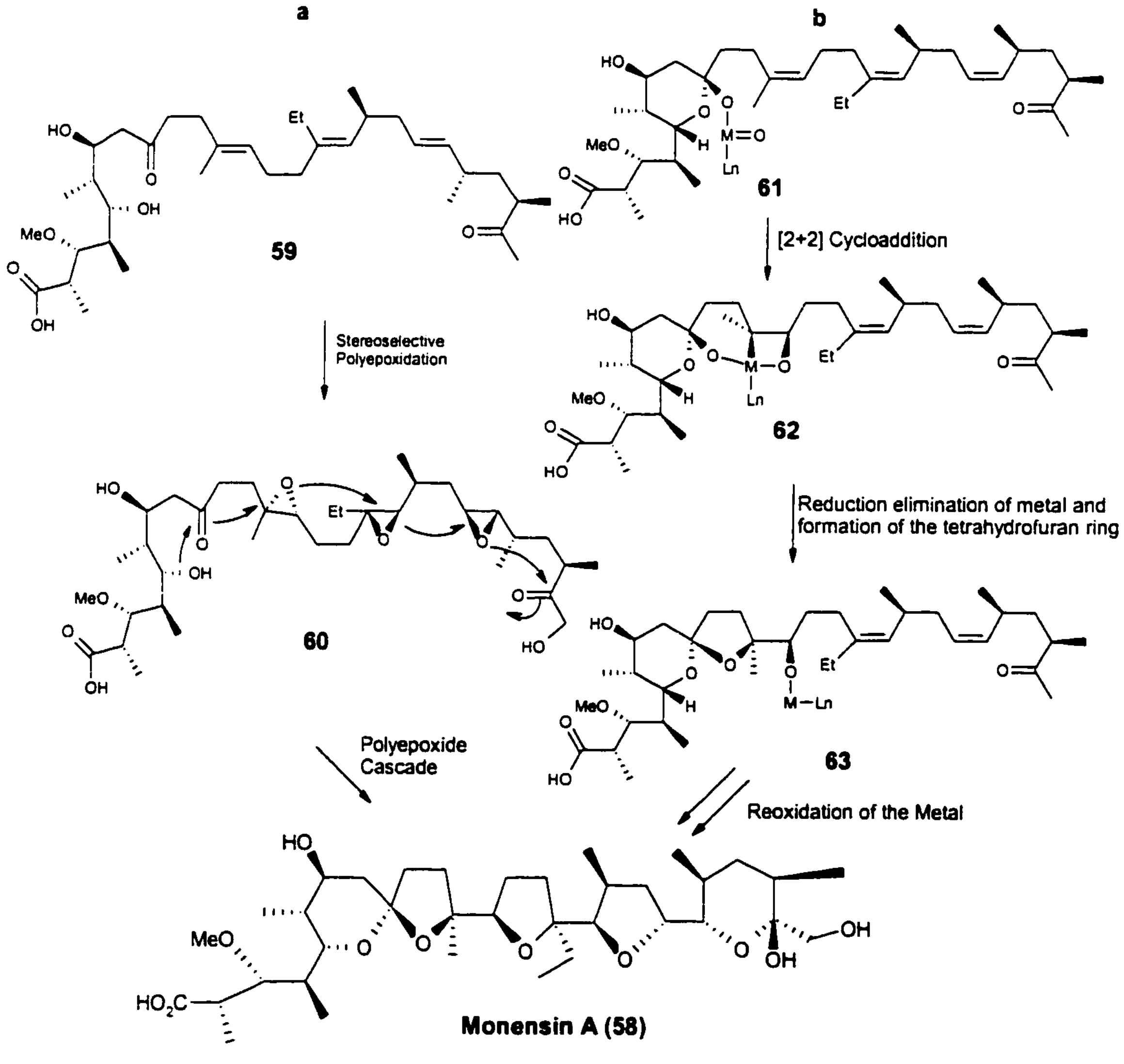
Although it is generally believed that the biosynthetic steps associated with the skeletal construction of polyether antibiotics are catalyzed by modular, multifunctional PKS Type I enzymes analogous to those known in the cases of macrolides, little is yet reported on the specific enzymes and genetics associated with this family of natural products.

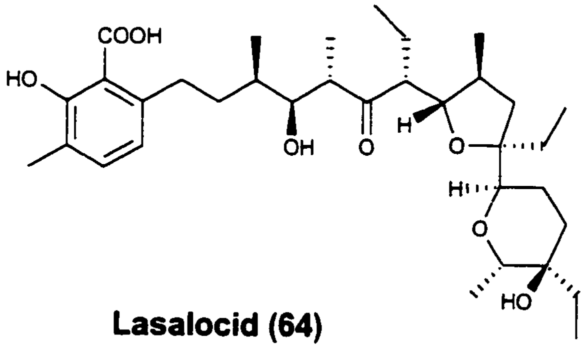
An idea which was initially proposed by Westley for the biogenesis of lasalocid (**64**)[110] and later modified by Cane, Celmer and Westley as a unified hypothesis for the biosynthesis of all polyether polyketides [e.g., monensin A (**58**)] is the "polyepoxide cascade" mechanism (Scheme 18).[111] In this three-step process, the open-chain oligoketide precursor **59** is formed from acetate, propionate, and butyrate *via* repeated decarboxylative Claisen condensations and subsequent modifications of the β-keto group; all of these steps are believed to be catalyzed by the multifunctional PKS enzyme(s). The second step was proposed to be an enzymatic polyepoxidation reaction of the acyclic hydroxypolyene precursor **59**, followed by a cascade of intramolecular nucleophilic attacks on the keto and epoxide moieties of intermediate **60**, eventually leading to the formation of the tetrahydrofuran and tetrahydropyran rings of the final product **58** (Scheme 18a).

More recently, the alternative mechanism of "*syn*-oxidative polycyclization" of the hydroxypolyene precursor **61** was proposed by Townsend, Basak, and McDonald.[112] In this later model, an alkoxy-bound oxo metal derivative **62** was proposed to undergo a series of consecutive intramolecular [2+2] cycloadditions with the double bonds to give the corresponding metallaoxetane intermediate **63**, followed by reductive elimination of the metal and closure of the tetrahydrofuran or tetrahydropyran ring (Scheme 18b).

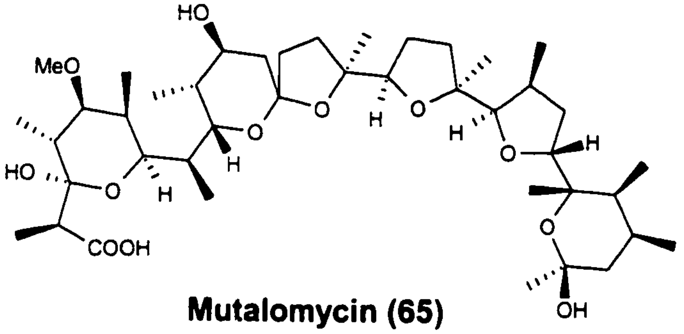
In addition to the polyether antibiotics, both of these elegant models may also be relevant in the biogenesis of some marine natural products which share common structural features with the polyether antibiotics, for example, lasalocid (**64**), mutalomycin (**65**) and the family of marine toxins typified by brevetoxin A (**66**). Nonetheless, to date both of these proposals remain unproven, mainly due to the chemical complexity and instability of the precursors and intermediates associated with the biosynthesis of such structurally complex molecules as monensin A (**58**) and brevetoxin A (**66**).

Scheme 18: Putative biosynthesis of Monensin A
a: polyepoxide cascade mechanism
b: [2+2] *syn*-oxidative polycyclization mechanism

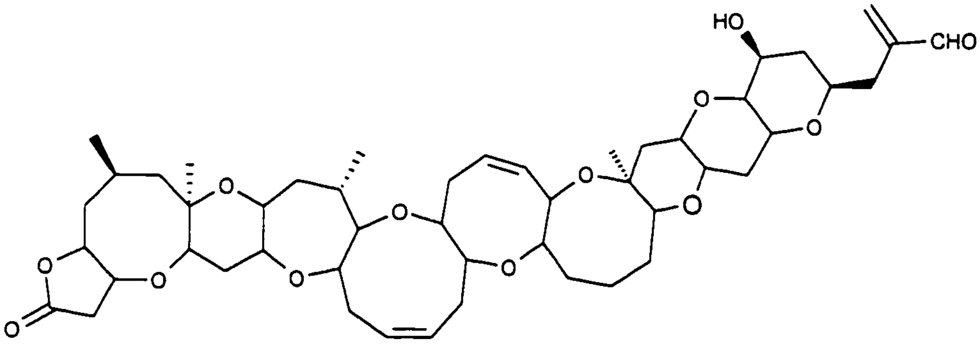




Lasalocid (64)



Mutalomycin (65)



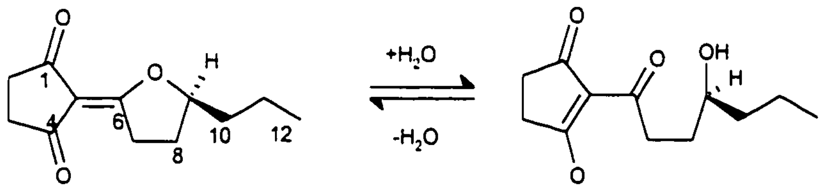
66

2.2 Studies on the biosynthesis of Oudenone

Oudenone (**67a** & **67b**),[113] a metabolite of the fungus *Oudemansiella radicata* is a strong inhibitor of catecholamine biosynthesis.[114] The biochemical mode of action of **67** is associated with inhibition of the enzymes tyrosine hydroxylase and phenylalanine hydroxylase. Kinetic studies using phenylalanine hydroxylase indicated that this inhibition is competitive with respect to the tetrahydropterin cofactor of the enzyme and noncompetitive with respect to phenylalanine and oxygen.[115] The physiological effects of oudenone include the reduction of blood pressure, as demonstrated with spontaneously hypertensive test animals.[116]

The structure of oudenone is unique among fungal metabolites (Scheme 19). In dry organic solvents, it is characterized by a tetrahydrofuran and 1,3-cyclopentadione moiety (**67a**). However, in aqueous solvents, structure **67a** is in dynamic equilibrium with the β-trione anion **67b**, via the simple addition or elimination of water (pKa=4.1)

Scheme 19: Structural features of oudenone and pH-dependence



(67a) UV_{max}=284, pH<3

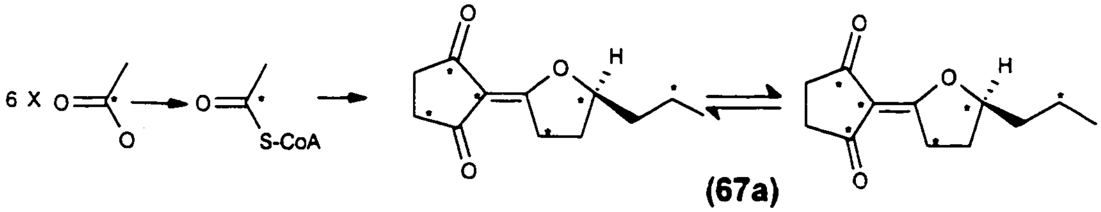
(67b) UV_{max}=246, pH>7

2.3 Results and discussion

2.3.1 Incorporation of primary precursors and preliminary studies on the mechanism of cyclization of the tetrahydrofuran moiety of oudenone

We have speculated that oudenone could serve as a model in our exploration into the general biogenesis of tetrahydrofuranyl and tetrahydropyranyl moieties of polyether-type metabolites. In our earlier studies, we confirmed that oudenone (**67**) is derived from a hexaketide, formed from the head-to-tail condensation of six acetate units (Scheme 20).[117]

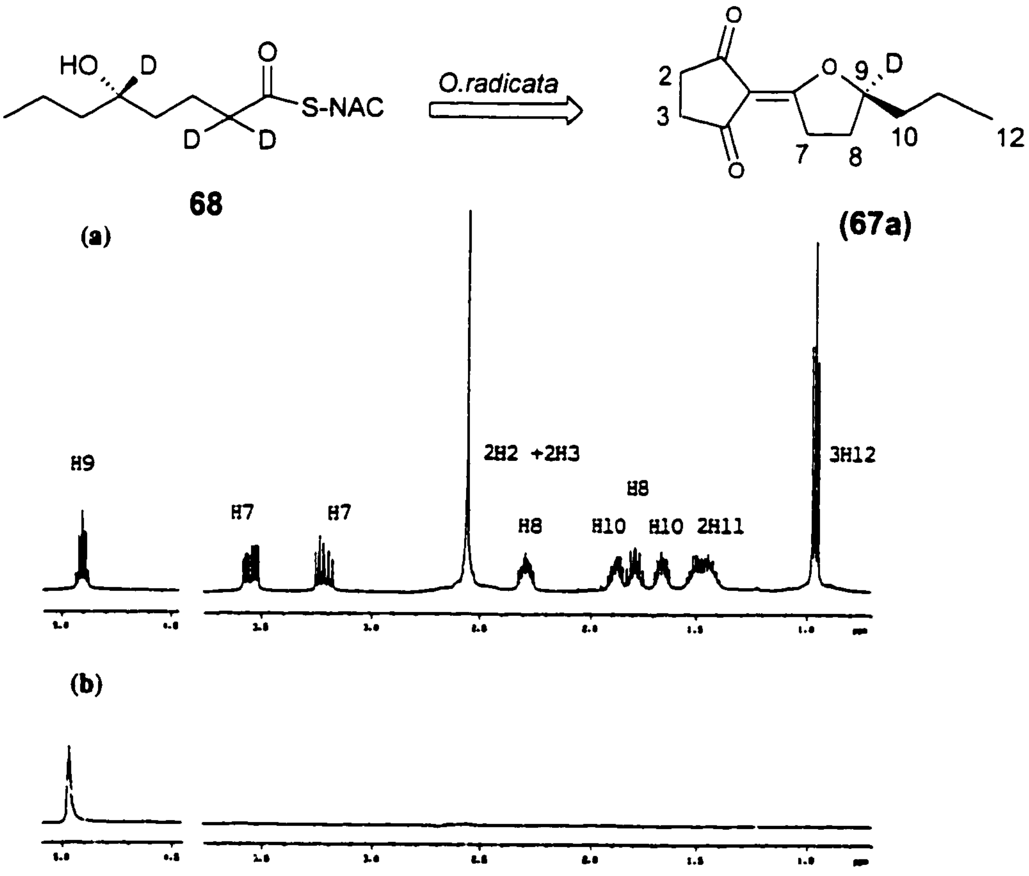
Scheme 20: ^{13}C Incorporation from primary precursors into oudenone



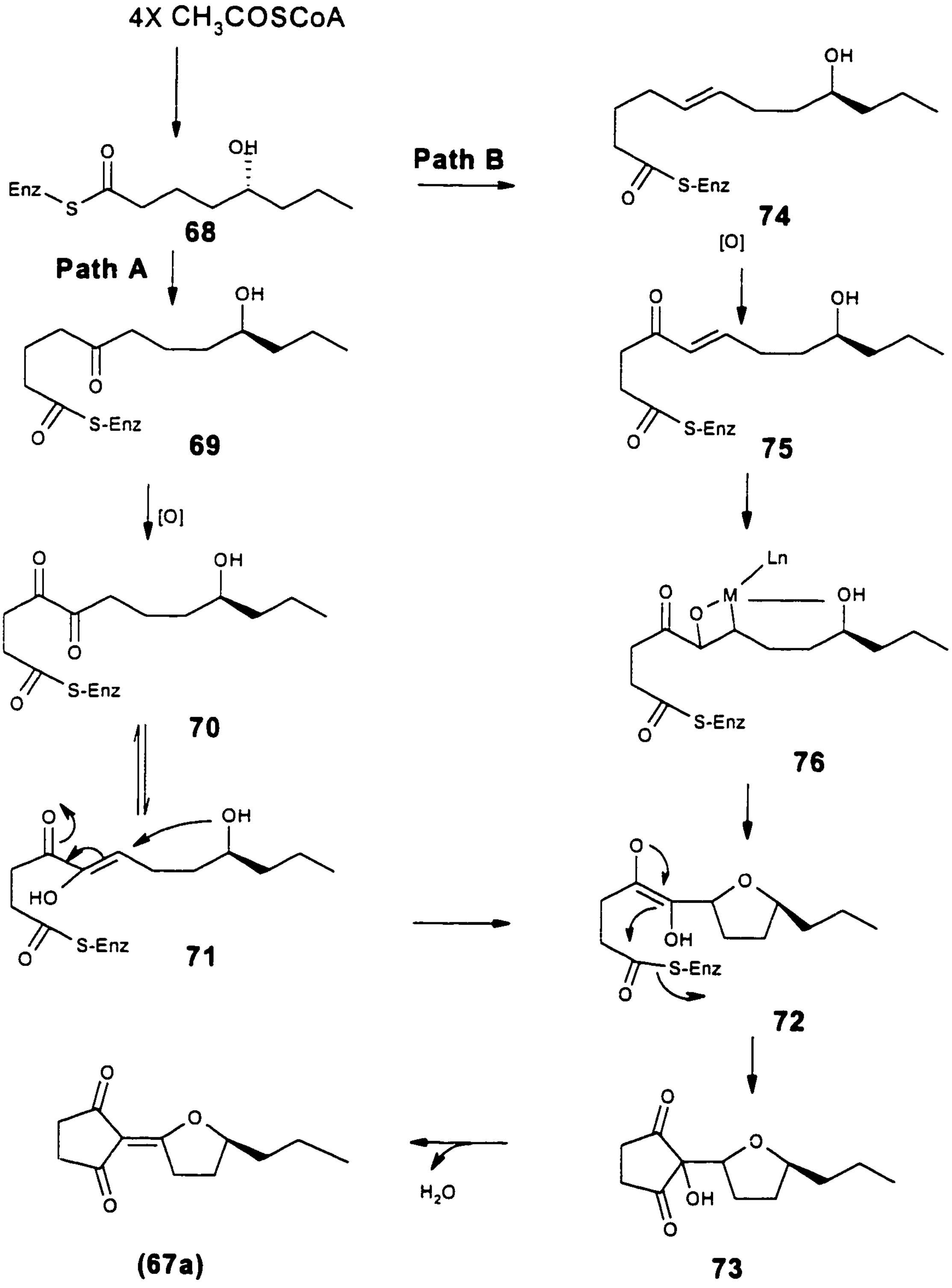
(67a)

Subsequently, we have shown that the *N*-acetylcysteamine thioester derivative of (5S)-5-hydroxyoctanoic acid **68** (Scheme 21) could serve as a substrate for the PKS enzyme(s) catalyzing the biosynthesis of **67** in cultures of *O.radicata*.[118] Furthermore, the enzymatic incorporation of **68** into the tetrahydrofuran moiety of **67** was shown to proceed without any change in the absolute stereochemistry of the C9 chiral center. This was demonstrated by the incorporation of a deuterium label into the expected carbon of oudenone (i.e. C9) and with the expected absolute stereochemistry upon feeding of the deuterium labeled NAC thioester of **68** to cultures of *O.radicata* (Figure 2). On the basis of these results, we have speculated on the structure and cyclization mechanism of two plausible pathways. Cyclization of the α-diketone **70** to oudenone (**67**) would be analogous to the "polyepoxide cascade" model (Scheme 21, path A). Alternatively, the "oxidative polycyclization" was also considered (Scheme 21, path B).[118] However, it seems less likely that the carbonyl moiety at C5 will be reduced and eliminated (intermediate **74**). We believe that the α-diketone **70** is the most likely precursor; it is reasonable to assume that oxidation at C4 of the open-chain hexaketide **69** occurs before cyclization. If the α-diketone **70** is the cyclase substrate, a favorable mechanism could involve an intramolecular Michael addition to form intermediate **72**, followed by a Claisen type of intramolecular condensation to generate intermediate **73**, and dehydration of intermediate **73** to give oudenone (**67**). In order to further explore our hypothesis, the synthesis and *in vivo* transformation of the α-diketone **70** into oudenone (**67**) was pursued.

Figure 2: (a) 500 MHz ^{1}H NMR spectrum of oudenone (**67a**); (b) 76.6 MHz ^{2}H NMR spectrum of oudenone (**67a**) derived from the intact incorporation of ^{2}H- labelled compound **68**



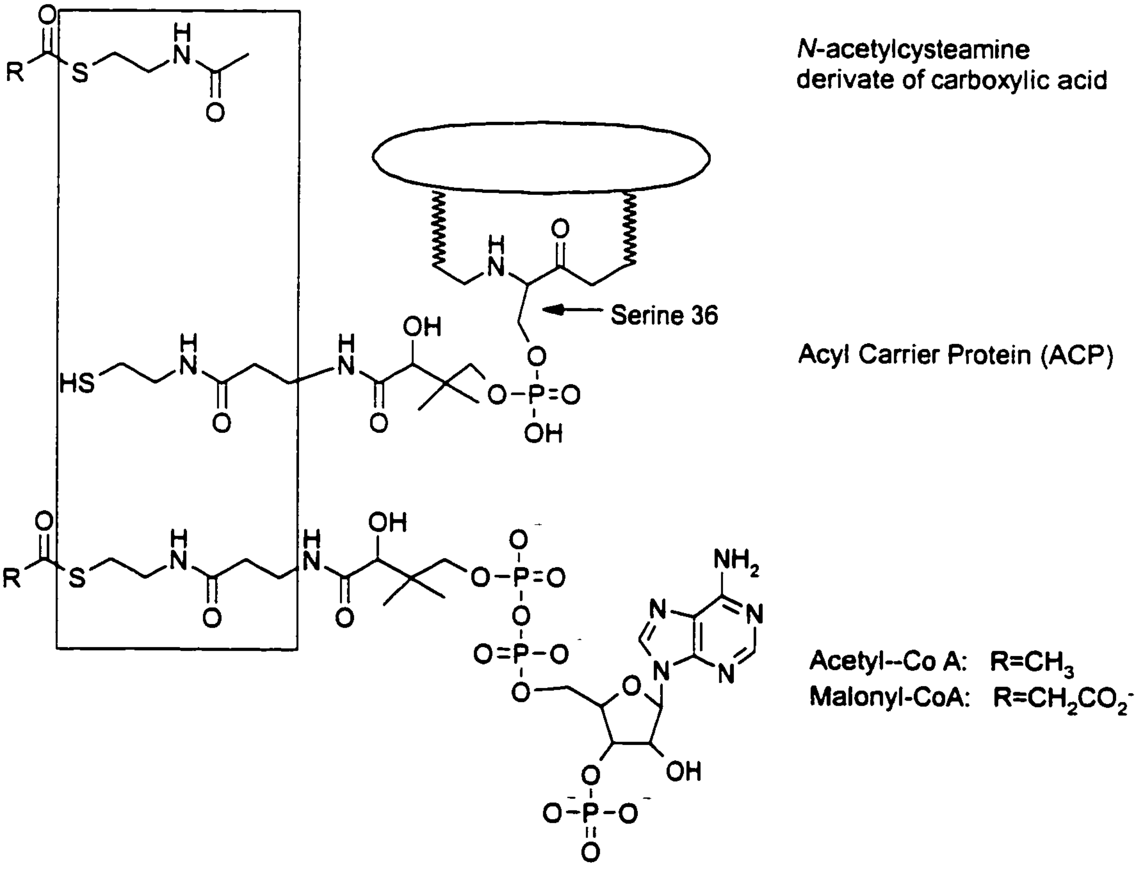
Scheme 21: Plausible mechanisms for the biogenesis of oudenone (**67a**)



2.3.2 Synthesis of α-diketide NAC thioester 86

N-acetylcysteamine (NAC) thioester derivatives are believed to serve as mimics of the biologically relevant coenzyme A (CoA) or acyl carrier protein (ACP) thioesters involved in the biosynthesis of polyketides due to their high degree of structural homology with the 4'-phosphopantetheine side chain of the active ACP of PKSs, as well as that of Coenzyme A (Scheme 23). Thus thioester derivatives are often better substrates for PKS enzymes than their corresponding acids.[16a, 119] Thus, we decided to synthesize the α-diketone NAC thioster **86**.

Scheme 22: Comparison of similar parts of ACP, CoA and NAC

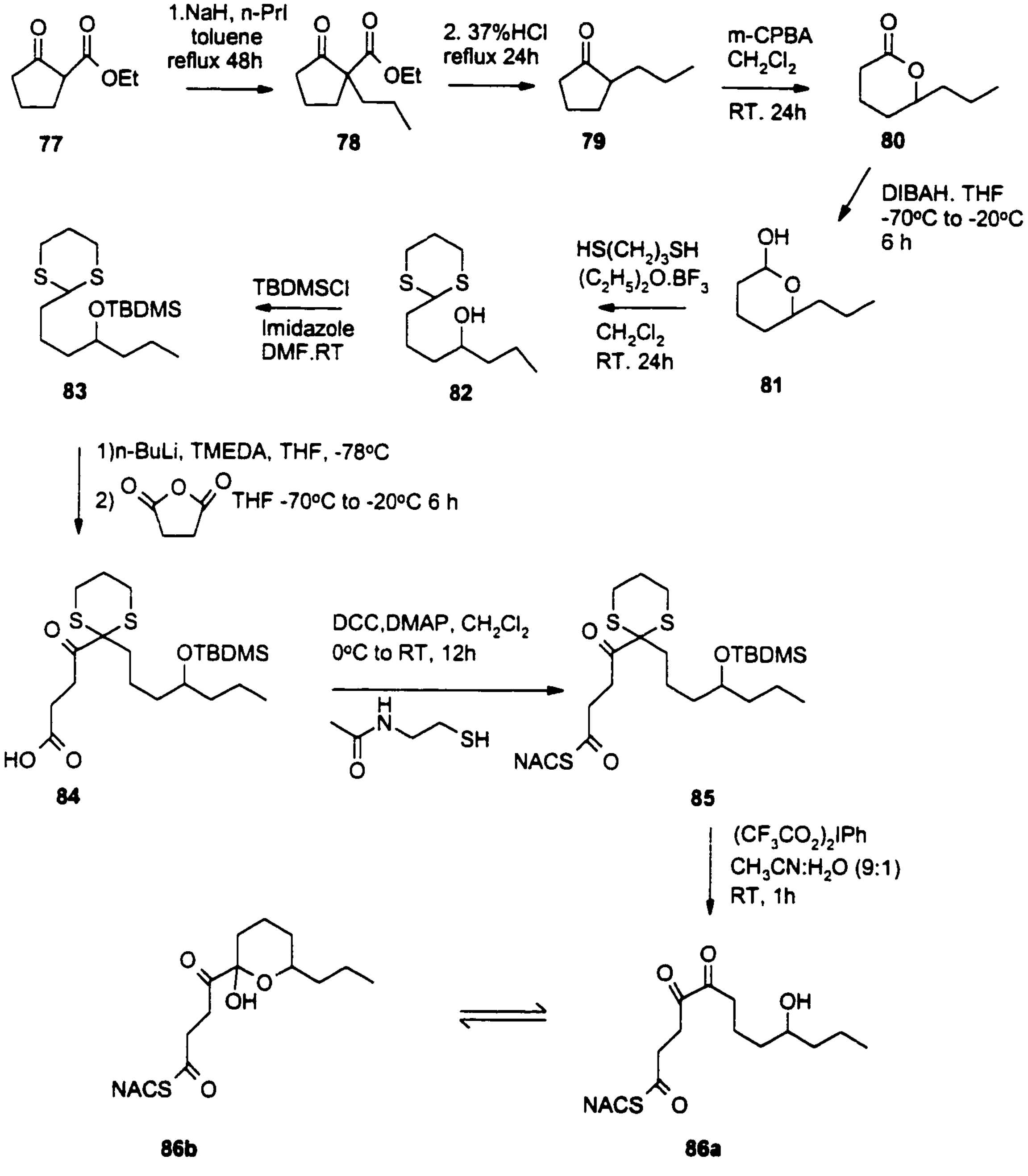


In our initial studies on the biosynthesis of oudenone (**67**),[119] we showed that its carbon skeleton is derived from six acetate units, and we proposed that the α-diketone **70** was the most likely open-chain hexaketide intermediate (Scheme 21). In order to test this hypothesis and further probe the mechanism of cyclization, the synthesis of deuterium labeled **70** as its *N*-acetylcysteamine (NAC) thioester derivative **86** (Scheme 22) was undertaken. Commercially available ethyl-2-oxocyclopentanecarboxylate (**77**) was first alkylated with *n*-propyl iodide at the α carbon to form compound **78** which was then hydrolyzed and decarboxylated with HCl to give racemic *n*-propylcyclopentanone (**79**). Compound **79** was oxidized using the Baeyer-Villiger method to give 3-propyl-β-valerolactone (**80**). Racemic 3-propyl-β-valerolactol (**81**) was obtained from the DIBAL-H reduction of the corresponding lactone **80**.[120] Lactol **81** was subsequently reacted with 1,3-propanedithiol in the presence of BF_3-etherate in order to protect the masked aldehyde,[121] and the 4-hydroxy group of the crude dithiane **82** was converted to the *t*-butyldimethylsilyl ether **83** in greater than 90% overall yield (Scheme 22).

Deprotonation of compound **83** (at C1) was carried out with the careful addition of 1 equivalent of *n*-butyllithium at –78 °C to –20 °C.[122] In order to avoid the addition of excess base, extra care was taken to introduce the solution of *n*-butyllithium into the reaction mixture in small aliquots. The progress of the reaction was monitored by ^{1}H NMR of the crude reaction mixture after quenching with D_2O. Once the deprotonation was complete, the anionic solution of **83** was transferred through a pre-cooled cannula at –78°C to an anhydrous solution of succinic anhydride (or [2-$^2H_2$, 3-$^2H_2$]succinic anhydride) at the same temperature. Both the deuterium labelled and unlabelled compounds **84** were prepared using the same protocol, and fully characterized by COSY, DEPT, HETCOR, ^{1}H, ^{13}C and ^{2}H NMR, and MS. The yield of this condensation ranged between 35-55%, based on the isolated pure compound **84** (>80% based on the recovery of unreacted compound **83**, however, the unreacted [2-$^{2}H_2$, 3-$^{2}H_2$]succinic anhydride could not be recovered). It is worth mentioning that any attempt to remove the 1,3-dithiane protecting group of **84** using standard literature procedures led primarily to decomposition of the starting material and the isolation of the TBDMS silyl ether of 4,5-dioxododecanoic acid in less than 15% yield. Thus, we coupled the protected carboxylic acid **84** with *N*-acetylcysteamine in the presence of DCC and DMAP to obtain the thioester **85** in 78% yield after chromatography.

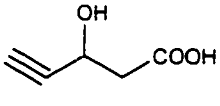
Finally, the removal of both the 1,3-dithiane and the silyl ether protecting groups were achieved in one step with [bis(trifluoroacetoxy)iodo]benzene following Stork's methodology.[123] The final product **86** (**86a** & **86b**) was isolated in 40% yield after chromatography and, not surprisingly, was found to be very chemically unstable. Therefore, feeding experiments had to be timed carefully so that the labeled final product, [2-$^{2}H_2$, 3-$^{2}H_2$]-α-diketone NAC thioster **86**, could be used as soon as it was prepared.

Scheme 23: Synthesis of α-Diketone NAC thioester **86**

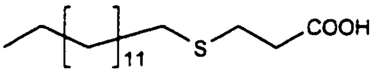


2.3.3. β-Oxidation inhibitors

Advanced precursors like di- and triketides as their carboxylic acids often rapidly degrade to acetate (or propionate) by efficient β-oxidation when they are fed into cultures.[124] In order to reduce degradation of our biosynthetic intermediates, two β-oxidation inhibitors, 3-hydroxypentynoic acid (**87**) and 3-(tetradecylthio)propanoic acid (**88**) were separately added to the culture of *O. radicata*. The addition of 3-(tetradecylthio)propanoic acid (**88**) up to a concentration of 6.0 mg/100 ml/day did not seem to affect the growth of culture and the production of oudenone, whereas 3-hydroxypentynoic acid (**87**) exhibited toxic effects on *O.radicata*. The maximum amount of each inhibitor which could be tolerated by the culture of *O.radicata* without significantly reducing the production of oudenone was determined by monitoring the UV changes of the culture medium during fermentation.



87



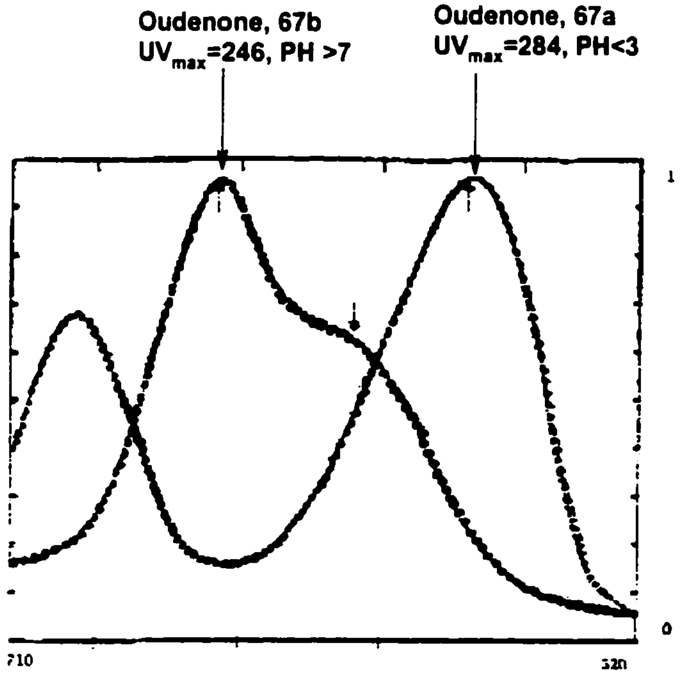
88

2.3.4 UV method for monitoring oudenone production and purification

The production of oudenone (**67**) in growing cultures of *O. radicata*, the appropriate timing for the feeding of labeled **86**, as well as purification of labeled oudenone (**67**) were monitored by the changes in UV absorption of the liquid culture. A shift in UV_{max} from 285 nm in acidic solution to 246 nm in neutral or basic solution was observed which is associated with the structural change of oudenone from **67a** to **67b**, respectively. (Figure

2). Therefore, the concentration of oudenone in fermentation broths was estimated by measuring the ΔUV_{max} at 246 nm of 1 ml aliquots of the fermentation broth, diluted with a constant volume of pH 7.0 phosphate buffer or 0.1N HCl. The presence of **67** in the growing cultures was first observed after 8-9 days of incubation from the time of inoculation; it reached a maximum concentration at 16-19 days (ΔUV_{max} = 0.8-1.2), and it began to decrease after 20-21 days of incubation. Incorporation of stable isotopes into metabolite **67** was usually successful when labeled precursors were fed to the cultures shortly after the presence of oudenone could be detected (day 9-11, ΔUV_{max} = 0.2-0.3).[118] Initially, unlabeled **86** was fed to cultures of *O. radicata* at various concentrations in order to establish the maximum amount (6.0 mg/100 ml/day) tolerated by the culture without substantially inhibiting the production of oudenone (**67**).

Figure 2: The ultraviolet spectrum of oudenone in neutral (phosphate buffer, pH 7) and acidic solution (0.1N HCl)



2.3.5 Incorporation of α-Diketone NAC thioester 86

Pulse feeding of [2-$^2H_2$, 3-$^2H_2$]-α-diketone NAC thioster **86** to three cultures of *O. radicata*, in the presence of the β-oxidation inhibitor 3-(tetradecylthio)propanoic acid (**88**), led to the isolation of metabolite **67**, labeled with deuterium at C2/C3 as predicted (Figure 3). All three samples of labeled oudenone (**67**) were purified by semi-preparative C18 reversed phase HPLC and analyzed by analytical HPLC and NMR.

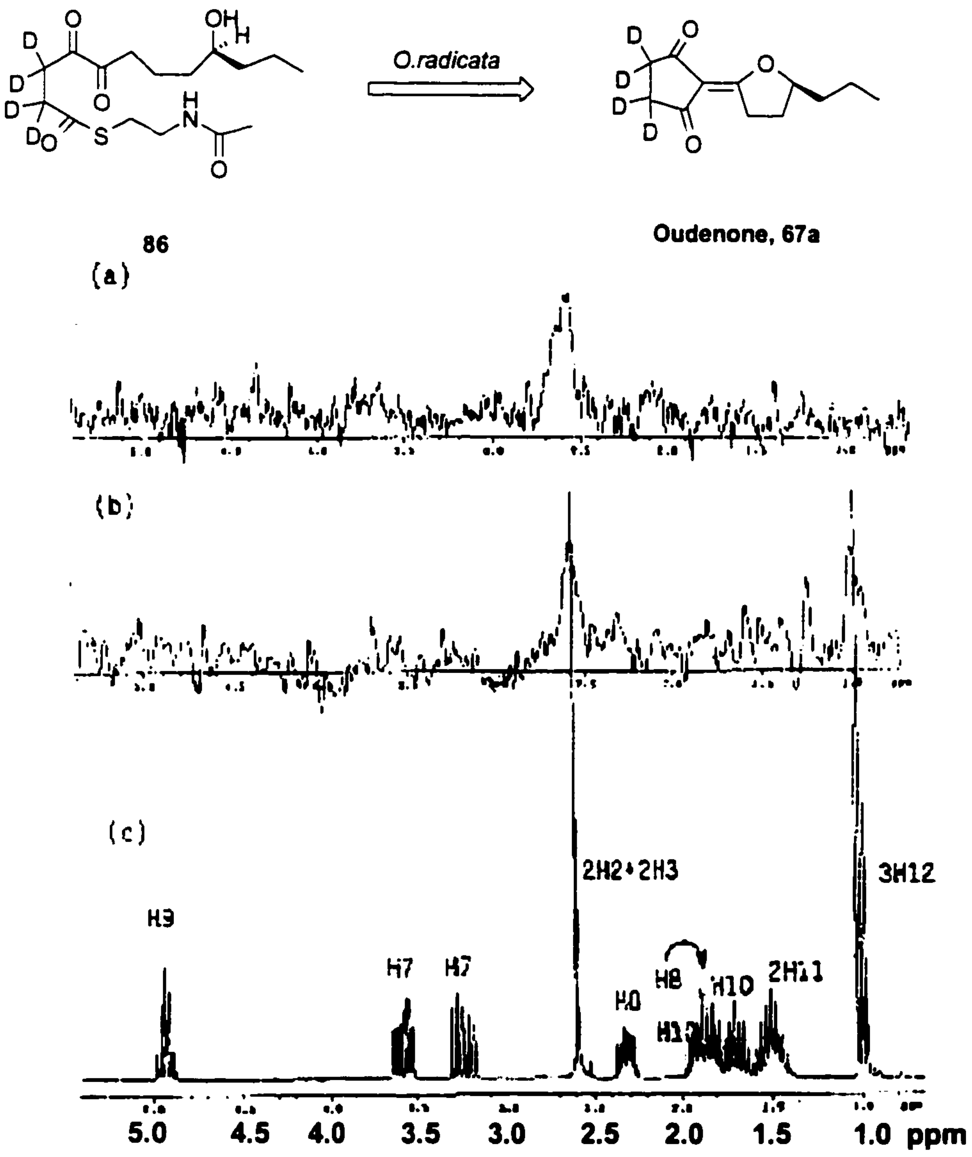
In one of the feeding experiments, deuterium incorporation (albeit with a very low % incorporation) was observed *exclusively* at the characteristic chemical shift of the H2/H3 protons (δ 2.55, Figure 3a). In the other two cases (Figure 3b), deuterium incorporation was also observed at C12 due to the inevitable degradation of compound **86** to [2-$^2H_2$]acetate by *in vivo* β-oxidation. However, the level of deuterium incorporation at C2/C3 was at least 2-4 fold higher than at C12. The *in vivo* production of [2-$^2H_2$]acetate would be expected to result in the incorporation of ^2H at C2, C8, C10 and C12 of metabolite **67**. However, we believe that the level of ^2H incorporation at C8 and C10 was significantly lower than at C12 and below detection limits.[118] Furthermore, the labeling pattern in the structure of **67** was scrambled between C2 and C3 due to the dynamic equilibrium of **67a** and **67b** in the growth culture of *O. radicata* at the pH of the fermentation broth (pH: 4.0-5.5).

Therefore, these results strongly suggest that the α-diketone **70** is the open-chain precursor of oudenone (**67**). Consequently, it is reasonable to believe that the cyclization of **70** proceeds *via* an intramolecular Michael addition, followed by a Claisen-type intramolecular condensation and dehydration to give **67** (Scheme 21, path A). This type of mechanism bears many similarities to the "polyepoxide cascade" model, as well as the catalytic mechanism leading to the formation of the 2-amino-3-hydroxycyclopent-2-enone, a unique structural unit found in several antibiotics including reductiomycin,[125] asukamycin,[126] moenomycin[127] and L-155,175.[128] A pyridoxal phosphate-dependent cyclization mechanism leading to the formation of this moiety was proposed by Floss in the biosynthesis of reductiomycin.

2.4. Summary

The results presented in this study provide strong evidence for our proposed pathway and mechanism of cyclization of the open-chain hexaketide precursor **70** into oudenone (Scheme 21). The proposed formation of the tetrahydrofuran and 1,3-cyclopentanedione moieties of **67** is consistent with the mechanisms previously proposed for microbial polyether metabolites and natural products characterized by the 2-amino-3-hydroxy cyclopent-2-enone structural unit,[127] respectively.

Figure 3: (a) 46.6 MHz ^{2}H NMR spectrum of oudenone (**67a**) derived from the intact incorporation of ^{2}H- labelled compound **86**; feeding experiment I. (b) 46.6 MHz ^{2}H NMR spectrum of oudenone (**67a**) derived from feeding of ^{2}H-labelled compound **86**; experiments II. (c) 300 MHz ^{1}H NMR spectrum of oudenone (**67a**).



2.5. References:

---

[108] Hu, Z.; Bao, K.; Zhou, X.; Zhou, Q.; Hopwood, D.A.; Kieser, T.; Deng, Z. *Mol. Microbiol.* **1994**, *14*, 163.

[109] (a) Minto, R.E.; Townsend, C.A. *Chem Rev.* **1997**, *97*, 2537.
(b) Brown, D.W.; Yu, J.-H.; Kelkar, H.S.; Fernandes, M.; Nesbitt, T.C.; Keller, N.P.; Adams, T.H.; Leonard, T.J. *Proc. Natl. Acad. Sci. USA* **1996**, *93*, 1418.

[110] Westley, J.W. In "Antibiotics IV:Biosynthesis"; Corcoran, J.W., Ed.; Springer-Verlag: New York, **1981**; pp 41-7.

[111] Cane, D.E.; Clemer, W.D.; Westley, J.W. *J.Am.Chem.Soc.* **1983**, *105*, 3594

[112] (a) Koert, U. *Angew. Chem. Int. Ed. Engl.* **1995**, *34*, 298.
(b) McDonald, F.E.; Towne, T.B. *J. Am. Chem. Soc.* **1994**, *116*, 7921.
(c) Townsend, C.A.; Basak, A. *Tetrahedron* **1991**, *47*, 2591.

[113] Umezawa, H.; Takeuchi, T.; Iinum, H.; Suzuki, J.; Ito, M.; Matsuzaki, M. *J. Antibiot.* **1970**, *23*, 514

[114] Nagatsu, K.; Mitzutani, K.; Nagatsu, I.; Umezawa, H.; Matsuzaki, M.;Takeuchi, T. *Mol. Cell. Biochem.* **1973**, *1*, 107

[115] Koizumi, S.; Nagatsu, T.; Iinuma, H.; Ohno, M.; Takeuchi, T.; Umezawa, H. *J. Antibiot.* **1982**, *35*, 458

[116] Nagatsu, K.; Nagatsu, I.; Umezawa, H.; Matsuzaki, M.; Takeuchi, T. *Nature* **1971**, *230*, 381

[117] Famili, P. The Biosynthesis of Oudenone, a hypotensive agent from *Oudemansiella radicata. M.Sc thesis.* **1993.** Concordia University.

[118] Tsantrizos Y.S.; Zhou, F.; Famili, P.; Yang, X.-S. *J.Org. Chem.* **1995**, *60*, 6922

[119] Yue, S.; Duncan, J.S. ;Yamamoto, Y.; Hutchinson, C.R *J. Am. Chem. Soc.* **1987**, *109*, 1253.

[120] Saucy, G.; Borer, R. *Helv. Chim. Acta* **1971**, *54*, 2121.

[121] Seebach, D.; Corey, E.J. *J. Org. Chem.* **1975**, *40*, 231

[122] Corey, E.J.; Hua, D.H.; Pan, B.-C.; Seitz, S.P. *J. Am. Chem. Soc.* **1982**, *104*, 6818.

[123] Stork, G.; Zhao, K. *Tetrahedron Lett.* **1989**, *30*, 287.

[124] Li, Z.; Martin, F.M.; Vederas, J.C. *J. Am. Chem. Soc.* **1992**, *114*, 1531

[125] Cho, H.; Beale, J.M.; Graff, C.; Mocek, U.; Nakagawa, A.; Omura, S.; Floss, H.G. *J. Chem. Soc., Chem. Comm.*, **1993**, *115*, 12296.

[126] Kakinuma, K.; Ikekawa, N.; Nakagawa, A.; Ômura, S. *J. Am. Chem. Soc.* **1979**, *101*, 3402.

[127] Welzel, P.; Witteler, F.J.; Muller, D.; Riemer, W. *Angew. Chem.* **1981**, *93*, 130.

[128] Goetz, M.A.; McCormick, P.A.; Monaghan, R.L.; Ostlind, D.A.; Hensens, O.D.; Liesch, J.M.; Albers-Schönberg, G. *J. Antibiot.* **1985**, *38*, 161.

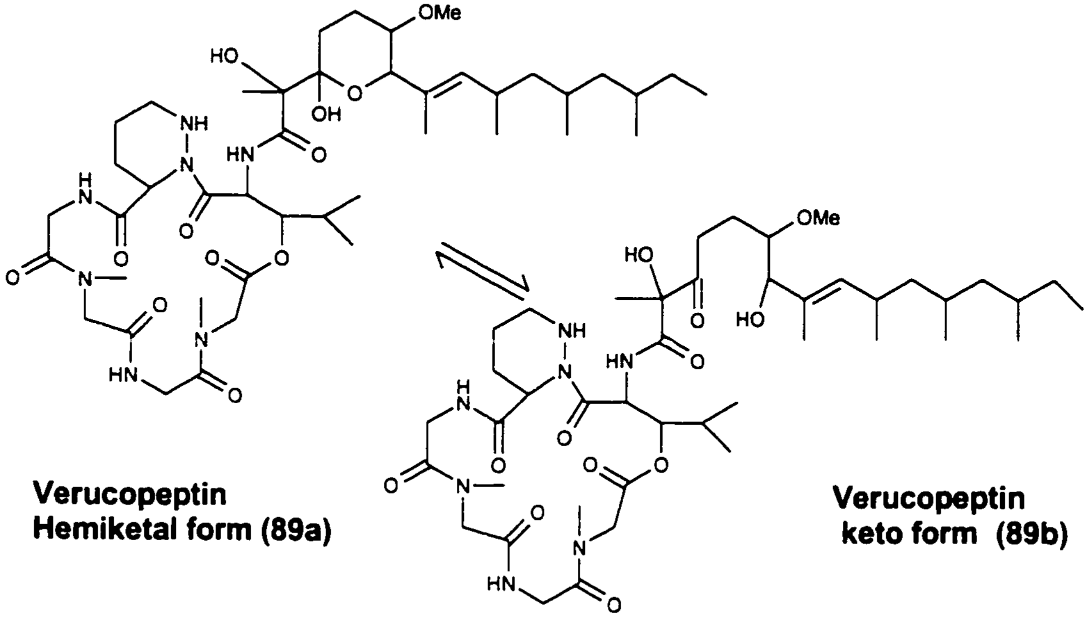
Chapter 3 Characterization of the Genes Involved in the Biosynthesis of Verucopeptin, a Cyclodepsipeptide Metabolite of *Actinomadura verrucosospora*

3.1 Introduction

The metabolite verucopeptin (**89**, Scheme 24) was isolated from cultures of *Actinomadura verrucosospora* and shown to exhibit specific *in vivo* activity against B16 melanoma.[130] It is structurally related to the antimicrobial agents A834586C (**90**),[131] azinothricin (**91**),[132] citropeptin (**92**) and variapeptin (**93**),[133] the extracellular matrix antagonist IC101 (**94**),[134] the anti-inflammatory agent L-156,602 (**95**),[135] and the antitumor metabolite GE3 (**96**).[136] This unique family of natural products is structurally characterized by a tetrahydropyranyl side chain and a 19-membered cyclodepsipeptide containing units of D- and/or L-piperazic acid, β-hydroxyleucine, *N*-hydroxyl- and *N*-methyl amino acids.

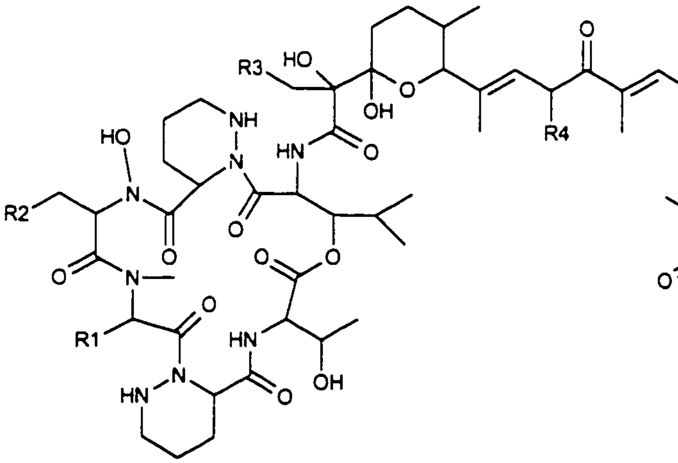
Verucopeptin **89** exists in a dynamic equilibrium between the hemiketal **89a** and keto **89b** forms (3:1) in solution (Scheme 24). The structure of metabolite **89** was originally deduced from spectroscopic analysis of a reduced analog (formed upon treatment of **89** with $NaBH_4$) and its chemical degradation products.[130] However, we were able to assign all of the ^1H and ^{13}C chemical shifts of the natural product **89** based on the data from high-field ^1H, ^{13}C, DEPT, COSY, TOCSY, HMQC, HMBC and NOESY NMR experiments for the first time, in order to pursue the biosynthesis of verucopeptin.[137]

Scheme 24: Two forms of verucopeptin (89)

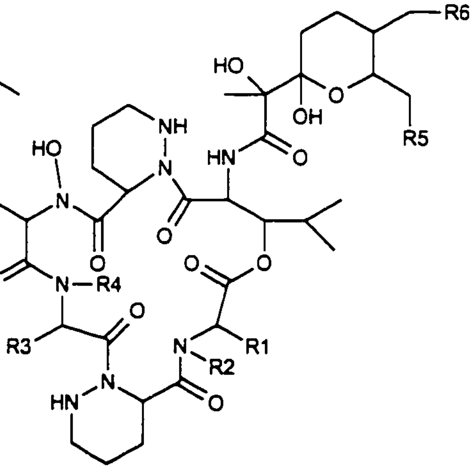


Verucopeptin Hemiketal form (89a)

Verucopeptin keto form (89b)



A83586C (90)
R1=R3=R4=CH_3, R2=H
Azinothricin (91)
R1=$CH_2CH(CH_3)_2$, R2=OCH_3, R3=CH_3, R4=CH_2CH_3
Citropeptin (92)
R_1=$CH_2CH(CH_3)_2$, R2=OCH_3, R3=H, R4=CH_3
GE3 (96)
R1=$CH_2CH(CH_3)_2$, R2= R3=H, R4=CH_3

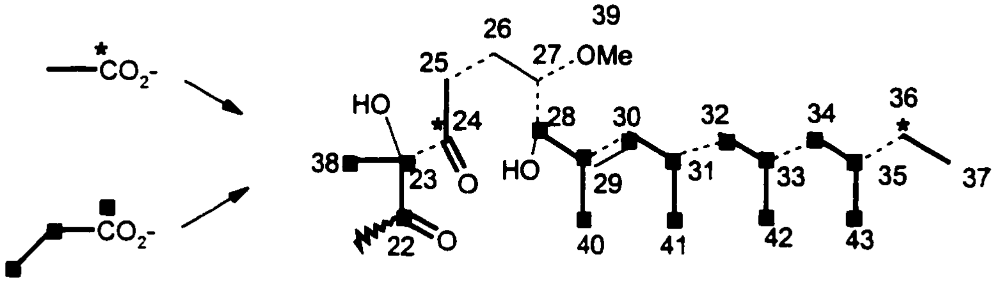


Variapetin (93)
R1=CH_2OH, R2=R5=H,
R3=$CH_2C_6H_5$, R_4=R6=CH_3
IC 101, (94)
R1=R5=CH_3, R2=OH,
R3=R4=H, R6=$CH_2CH(CH_3)_2$
L-156602 (95)
R1=CH_3, R2=OH,
R3=R4=R5=H, R6=$CH_2CH(CH_3)_2$

Feeding of ^{13}C-labeled precursors to cultures of *A. verrucosospora* led to the isolation of ^{13}C-labeled verucopeptin (Scheme 25). Because of the complex structure of this metabolite, it was difficult to identify the ^{13}C-labeled carbons from the 1D ^{13}C NMR spectra. However, the coupling patterns and $J_{C\text{-}C}$ values of ^{13}C-labeled verucopeptin were obtained from 1D-INADEQUATE NMR experiments and a series of 1D ^{13}C COSY NMR experiments which were carried out using optimized delay values (1/2J) between the selective excitation pulse and the mixing pulse for each set of carbons.[137]

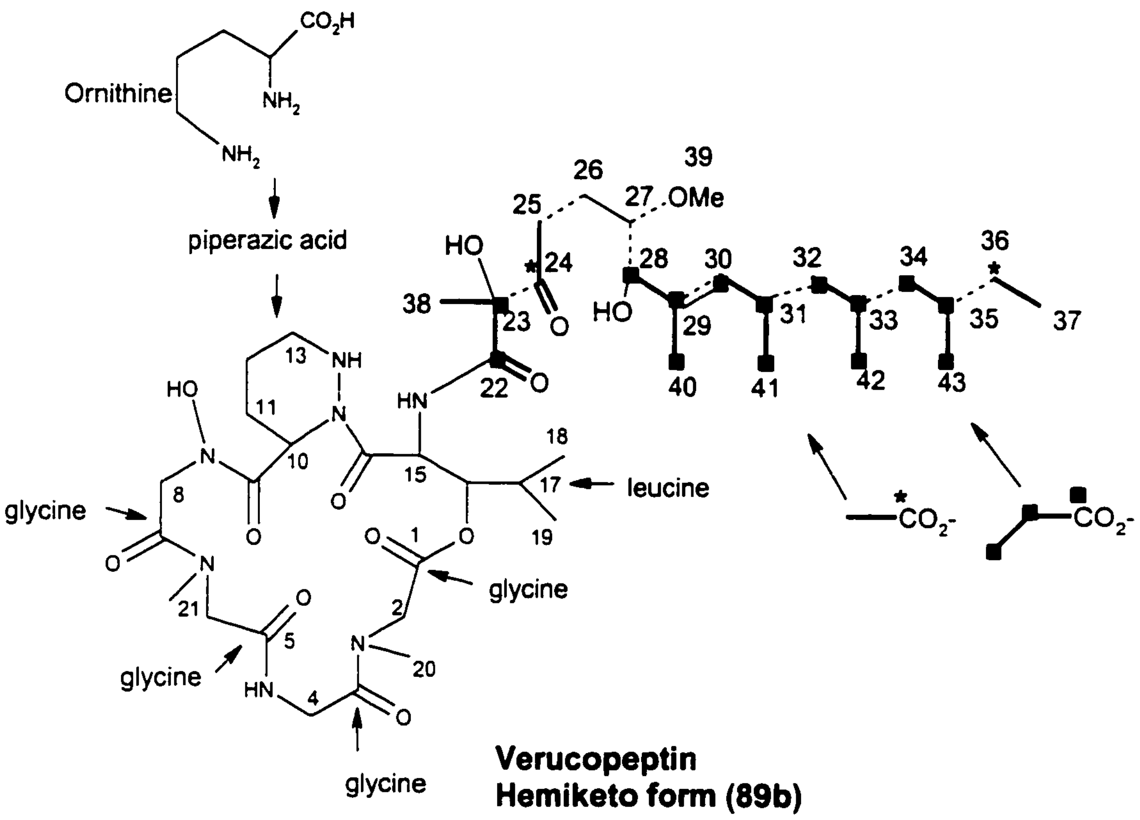
Feeding [2,3-$^{13}C_2$] propionate clearly showed the incorporation of five propionate units as shown in Scheme 25. ^{13}C enrichment at C22, C28, C30, C32 and C34 were observed from the incorporation of [1-^{13}C] propionate, which further confirmed the expected 5 propionate units involved in the biosynthesis of verucopeptin (Scheme 25). A feeding experiments with [1,2-$^{13}C_2$] acetate and [1-^{13}C] acetate clearly showed that C24-C25 and C36-C37 were derived from acetate (Scheme 25). The expected incorporation of acetate at the C26-C27 unit could not be confirmed from the INADEQUATE NMR spectrum, possibly due to the very low level of incorporation. From our general knowledge of the biosynthesis of complex polyketides, it is most likely that C26-C27 is derived from acetate. Therefore, we conclude that the tetrahydropyranyl side chain of verucopeptin is derived from three acetate and five propionate units.

Scheme 25: ^{13}C-Incorporation from primary precursors into tetrahydropyranyl side chain of verucopeptin



In summary, the results of our feeding experiments clearly indicate that the formation of the tetrahydropyranyl moiety of verucopeptin (**89**) may be catalyzed by modular type I polyketide synthases (PKS) enzymes.[19,26] The 19-membered cyclodepsipeptide is composed of six amino acids: one glycine, one β-hydroxyleucine, two sarcosines, and two unusual amino acids (one *N*-hydroxyglycine and one piperazic acid). The precursor for glycine, sarcosine and *N*-hydroxyglycine is believed to be glycine, whereas precursors for β-hydroxyleucine and one piperazic acid are believed to be leucine and ornithine, respectively. The non-ribosomal peptide synthetase (NRPS) enzymes and other enzymes will take these substrates to generate the 19-membered cyclodepsipeptide motif of verucopeptin (Scheme 26).[84]

Scheme 26: The putative biosynthetic origins of verucopeptin



Verucopeptin Hemiketo form (89b)

As discussed in the introduction of this thesis, PKSs and NRPSs systems catalyze the biosynthesis of polyketides and non-ribosomal peptides. In contrast to our knowledge of polyketide and non-ribosomal peptide biosynthesis, little is yet known about the genetics and biochemical mechanisms involved in the formation of natural products having a mixed biosynthetic origin. The gene clusters associated with the biosynthesis of FK506,[28] rapamycin,[29] and TA,[138] metabolites which have an amino acid moiety in a predominantly polyketide structural framework, are among the few reported. Furthermore, to date, very little has yet been reported on the genetics of secondary metabolites from the genus *Actinomadura,* yet *Actinomadura* is the most predominant genus of the rare actinomycetes.[139] Its DNA possesses a G+C content of 60-72 %, which is similar to that of the genus *Streptomyces*.[140] We speculated that highly conserved core regions encoding PKS type I and NRPS enzymes in *Streptomyces* could be used as hybridization probes in the screening of a genomic library of *A. verrucosospora* and the identification of genes associated with the biosynthesis of verucopeptin (**89**).

3.2 Results and discussion

3.2.1 Isolation and cloning of the PKS and RNPS genes associated with the biosynthesis of verucopeptin in *A. verrucosospora*.

A cosmid library of the chromosomal DNA of *A. verrucosospora* was constructed in the cosmid vector Supercos I (Appendix 6). A DIG-labeled KS1/ery fragment of 1,040 bp from the gene cluster of erythromycin A (**2**) was used to screen 60,000 colonies from the original genomic library, leading to the isolation of 2,000 positive colonies with 70 among those giving a strong hybridization signal. The DNA content of all 70 colonies was further screened with a set of PKS probes from *A. verrucosospora* (probes **KS/V-1**) to show that 53 of those contained DNA fragments which hybridized strongly with the KS/V probe under hybridization conditions of high stringency (0.1XSSC-0.1% SDS, 75^{0}C for 30 min, Figure 4). The PKS specific probes, **KS/V-1**, were generated by PCR amplification of the genomic DNA of *A. verrucosospora* using a set of KS degenerate primers (Table 5). The design of these primers was based on the highly conserved amino acid sequences of the KS domains associated with the biosynthesis of erythromycin (modules 1-6)[45], rapamycin (modules 1-14)[28] oleandomycin (modules 1 and 2)[31] and soraphen A[34] (Table 6). A single PCR amplified fragment of about 750 bp was obtained from these experiments (Figure 5a).

The DNA from each of the 53 KS-positive colonies was subsequently isolated, digested with *Bam*HI and further analyzed by Southern blot hybridization once again with the **KS/V-1** probes (Figure 6). A number of strongly positive fragments were observed, consistent with the presence of PKS type I genes in at least 49 out of the 53 colonies identified. Positive bands of the same size were also present in the *Bam*HI digestion mixtures of the chromosomal DNA of *A. verrucosospora* (Figure 6). Furthermore, overlapping regions of DNA (originating from chromosomal DNA) were clearly evident among the cosmids containing identical KS-positive bands; the fragments of 2.1 kb, 2.4 kb, 3.8 kb and 5.4 kb were the most predominant. The cosmids pYT24 and pYT27 were found to contain four identical KS-positive regions, fragments of 2.1 kb, 2.4 kb, 3.8 kb and 5.4 kb (Figure 6). In addition, pYT27 contained 2 larger bands of approximately 10 kb and 12 kb; these two cosmids were selected for further investigations.

Subsequently, we focused our search on the isolation of a cosmid containing both PKS and NRPS genes. NRPS degenerate primers (26-22A and 26-22B, Table 7) were designed based on the highly conserved regions of the adenylation domain A7 and thiolation T motifs (Table 8, Scheme 13 and Table 4).[87] PCR amplification of NRPS genes using these primers and genomic DNA as the template, led to the isolation of 500 bp DIG-labeled product [probe(s) **Pep/V**, Figure 5b]

Table 5: Primers for amplifying KS fragment as probe(s) **(KS/V-1)**

YXS1: GACACVGCNTGYTCBTCV	YXS2: TCRCCNTGVTTRCGSGTR
D T A C S S	S G T N A H
Primers 5' → 3'	3' ← 5'

(R=A+G. S=G+C. V=G+A+C. N= A+G+C+T. Y=C+T. B=G+T+C, K=G+T, W=A+T)

Table 6: Highly conserved region in KS domains of PKSs type I

PKS type I/KS Domains	Amino Acid Sequences
DEBS-KS1 and KS3 [45]	... D T A C S S S.....-[~250 AA]-.....S G T N A H a...
DEBS-KS2. 4.5.6[45]	... D T A C S S S.....-[~250 AA]-.....S G T N A H v...
Ole-KS1 and KS2[31]	... D T A C S S S.....-[~250 AA]-.....S G T N A H v...
Rap-KS1.2.3.5.6.8-14[28]	... D T A C S S S.....-[~250 AA]-.....S G T N A H v...
Rap-KS 4.7[28]	... D T A C S S S.....-[~250 AA]-.....S G T N A H v...
Sor-KS[34]	... D T A C S S S.....-[~250 AA]-.....S G T N A H v...

Note: DEBS: 6-deoxyerythronolide B synthase; Ole:Oleandomycin; Rap:Rapamycin; Sor:Soraphen A;

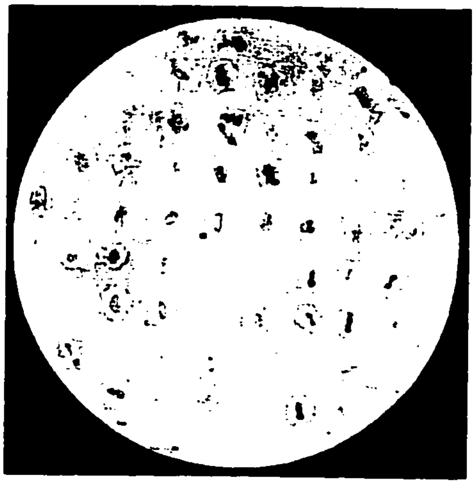
Table 7: Primers for making probe (s) **Pep/V**

	5' → 3'
N26-22A	CGGAATTC TAC CGC ACS GGC GAC STC GYC CG
N26-22B	CGGGATCC GA GTG GCC GCC SAG SKY GAA GAA

Table 8: Conserved A7 and Thiolation regions and designed primers for NRPS

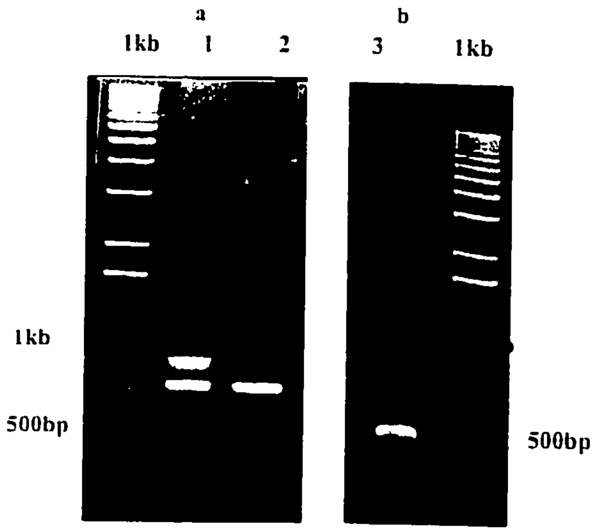
Chosen motifs	Highly Conserved Amino acid
Adenylation domain A7 [87]	Y(F) **RTGDL** (R.W)V(A)R
26-22A (related to A7)	Y **RTGDL** (V) V(A)R
Thiolation T Motifs [87]	FFL(F.R) **L(P)GGS(P)LY(F)**
26-22B (related to T motifs)	FFA(E,T,S.D)**L** **GGHS**

Figure 4: Positive colonies from hybridization with **KS/V-1** Probe.



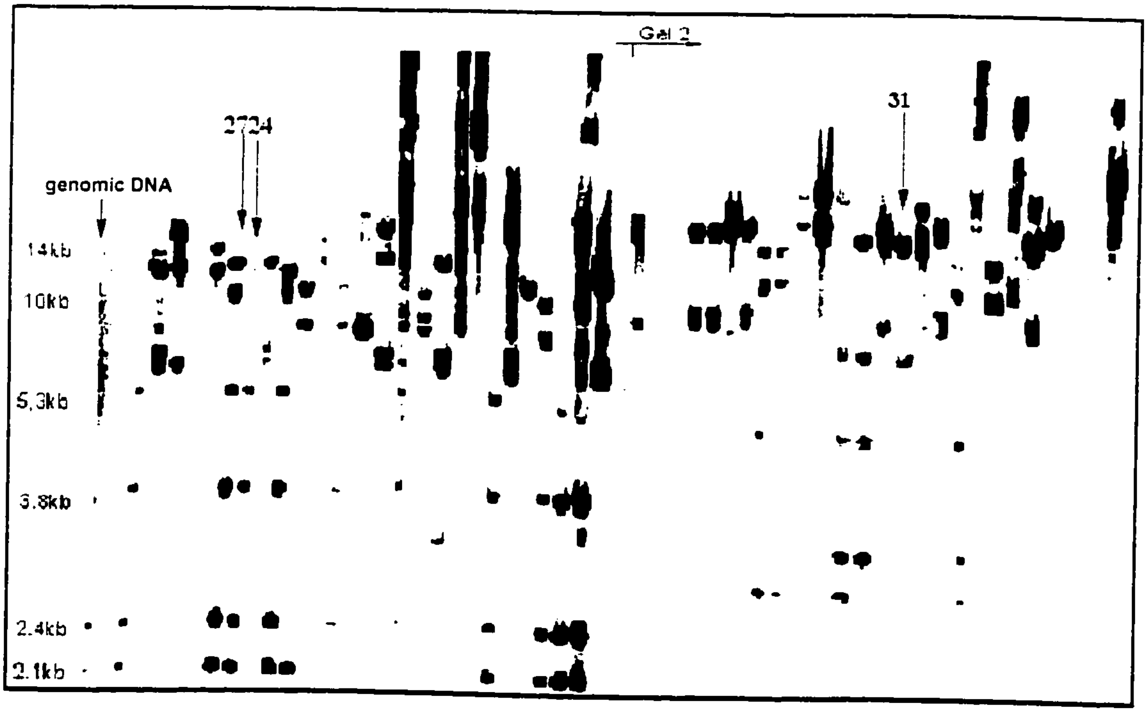
Washing conditions:.
1XSSC-0.1%SDS at RT twice for 30 min and
0.1XSSC-0.1%SDS at 75°C twice for 30 min.

Figure 5: PCR products for probes from genomic DNA of *A. verrucosospora*



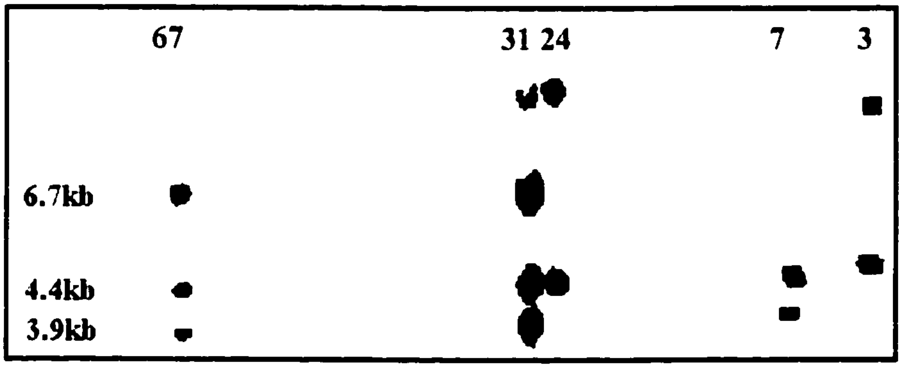
a: YXS1 and YXS2 as primers for making **KS/V-1** probes;
b: N26-22A and B as primers for making **Pep/V** probe.
Lane 1: pKAO127 containing DEBSs as a positive control
Lane 2: **KS/V-1** PCR products
Lane 3: **Pep/V** PCR products

Figure 6: Southern analysis of KS positively hybridizing cosmids using **KS/V-1** probe. Lanes of specifically labeled cosmids are discussed in the text.



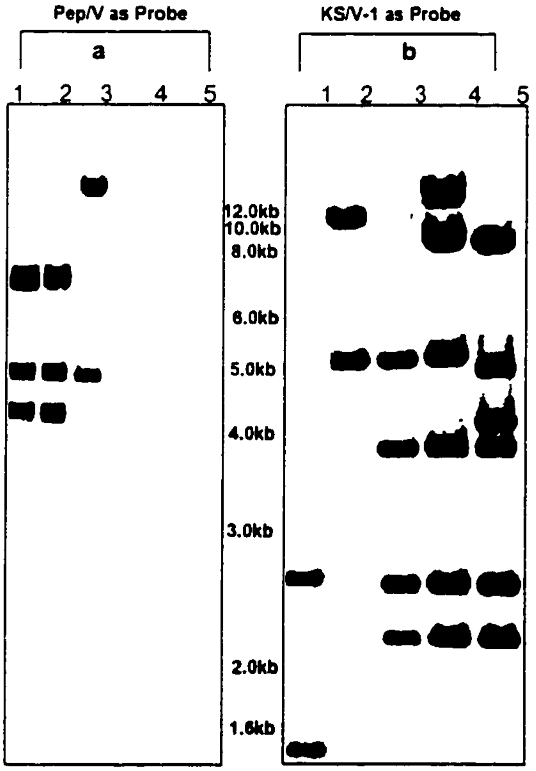
Southern blot hybridization experiments of the *Bam*HI digestion mixtures obtained previously from the 49 KS/V-1 positive cosmids (Figure 6), led to the identification of 5 cosmids which contain **Pep/V** positive bands (Figure 7). Two of these cosmids, pYT24 and pYT31 contained 2 and 3 **Pep/V** positive bands, respectively (Figure 7).

Figure 7: Southern blot results using **Pep/V** as a probe to screen all of the 49 KS/V positive cosmids



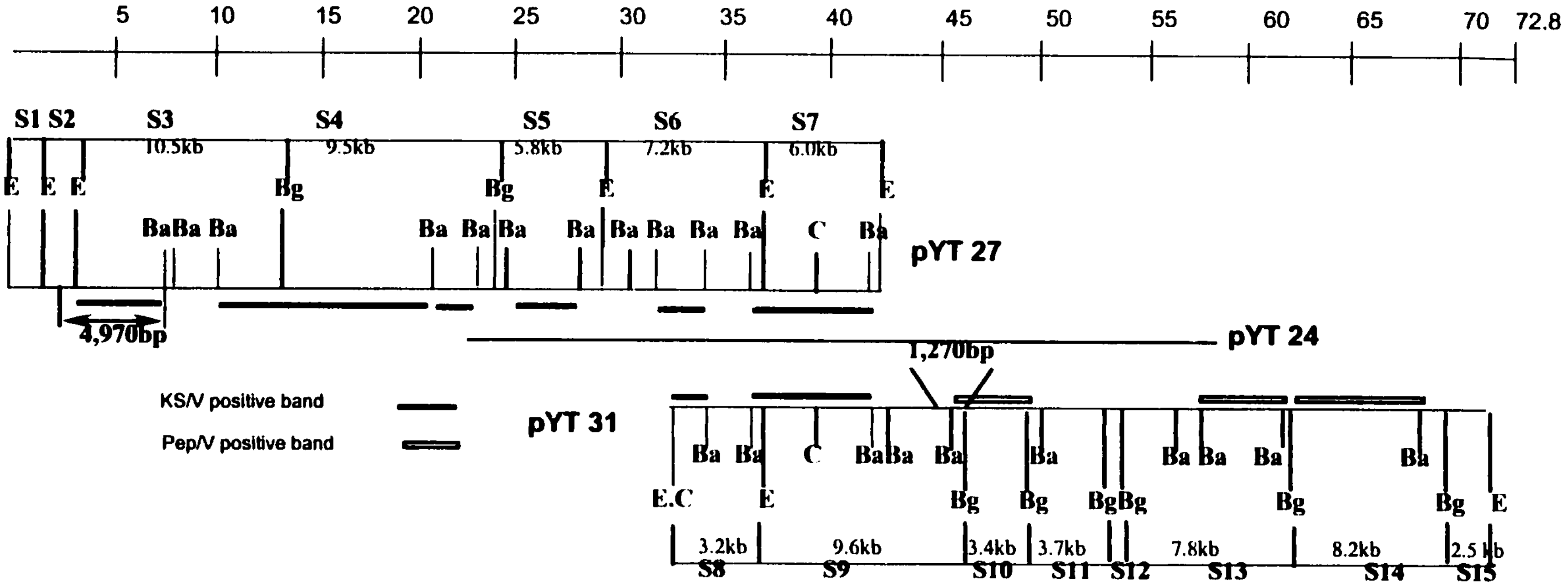
The three cosmids pYT27, pYT24 and pYT31 were further digested with different restriction enzymes and examined by Southern blot analysis using the **KS/V-1** and **Pep/V** probes (Figure 8). As mentioned previously, the *Bam*HI digestion mixtures of cosmids pYT27 and pYT24 contained 4 identical KS/V positive bands (2.1 kb, 2.4 kb, 3.8 kb and 5.4 kb),one of which was also present in the digestion mixture of pYT31 (5.4 kb). In addition, cosmids pYT24 and pYT31 contained one identical *Bam*HI Pep/V positive band of 4.4 kb. Thus, we concluded that cosmids pYT24, pYT27 and pYT31 contained overlapping regions of chromosomal DNA from *A. verrucosospora* (Figure 6). The restriction maps of pYT27 and pYT31 were based on multiple enzyme digestions and Southern blot analysis (Figure 9, Appendix I and II). Cosmids pYT27 and pYT31 contain approximately 42.4 kb and 39.9 kb of insert chromosomal DNA, respectively, with an overlapping fragment of approximately 9.7 kb. Based on the structural features and our knowledge of genetics of polyketide and non-ribosomally formed microbial peptides, the sum of the chromosomal DNA contained in the two cosmids pYT27 and pYT31 (72.5 kb in total) is approximately what would be expected for the total PKS and NRPS gene clusters associated with its biosynthesis.

Figure 8: Southern analysis of pYT27, pYT31, and pYT24



a: **Pep/V** as probe
b: **KS/V-1** as probe.
Lane1: pYT31 digested with *Bam*HI/*Cla*I/*Eco*RI;
Lane2: pYT31 digested with *Bam*HI;
Lane3: pYT24 digested with *Bam*HI;
Lane4: pYT27 digested with *Bam*HI;
Lane5: pYT27 digested with *Bam*HI/*Eco*RI.

Figure 9: Restriction maps of the genomic region of *A. verrucosospora* cloned in cosmids pYT27, pYT24 and pYT31.



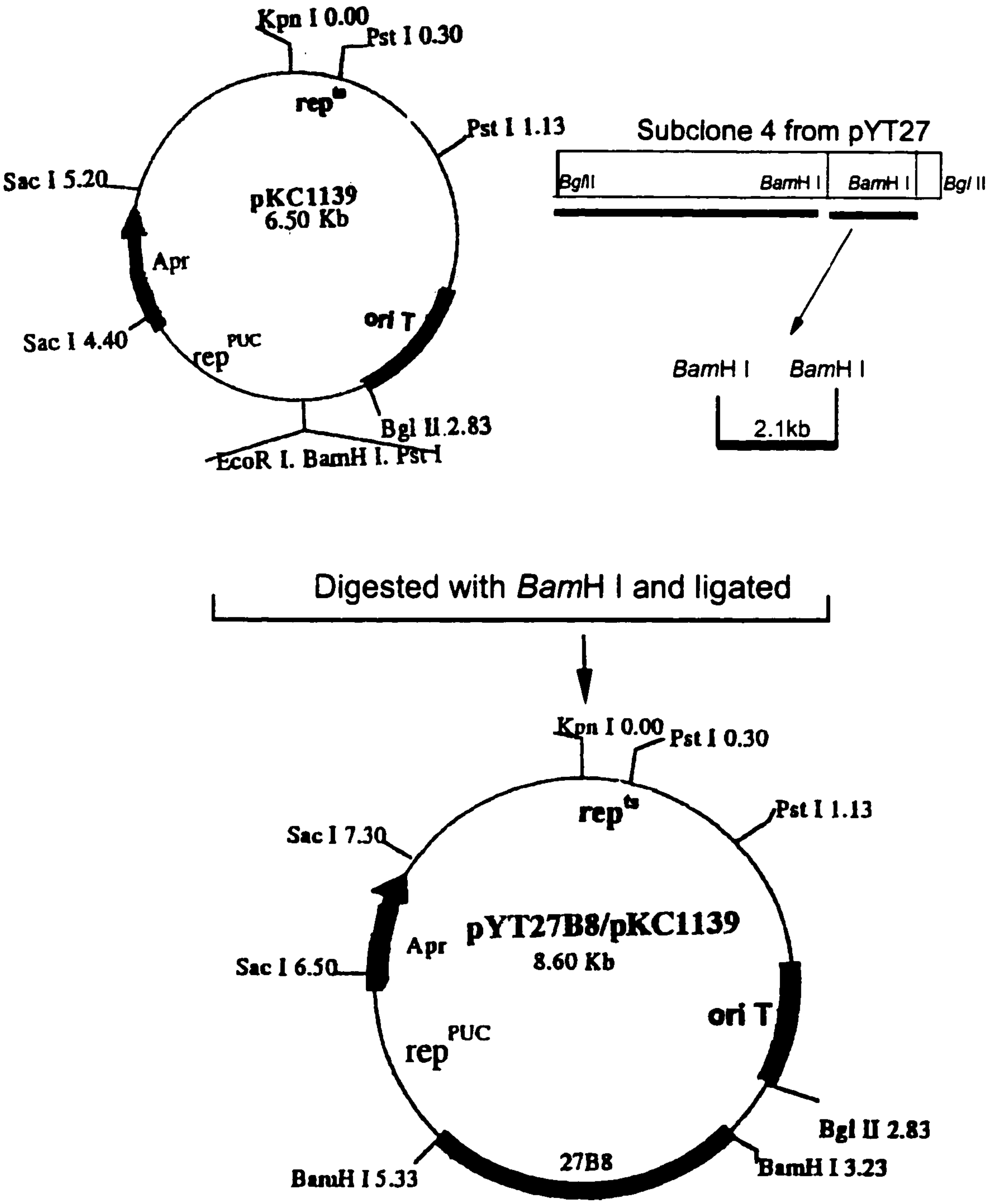
Ba: *Bam*HI; E: *Eco*RI; Bg: *Bgl*II; C: *Cla*I; S: subclone names;
4965 bp and 1270 bp regions were sequenced which are discussed in the text in detail.

3.2.2 Gene disruption of the PKS gene cluster associated with the biosynthesis of verucopeptin.

In order to confirm that the PKS genes cloned into the pYT27 cosmid were involved in the biosynthesis of verucopeptin (**89**) in cultures of *A. verrucosospora* Q886-2, a gene disruption experiment was designed. We chose to work with the conjugation vector pKC1139 (Figure 10)[141], which had been used previously in studies on the biosynthesis of methymycin[37] and niddamycin.[38] This plasmid contains an apramycin resistance selection marker which would allow selection in *E.coli* S17-1 and *A. verrucosospora* since both are sensitive to apramycin. It also contains an *ori*T which is a specific site for transferring plasmid DNA during bacterial conjugation and both the *E.coli* and *Streptomyces* origins of replication. Furthermore, pKC1139 contains a temperature-sensitive replicon from *S. ghanaenis* which functions only at temperatures below 34°C, and so under selective pressure at a nonpermissive temperature, the plasmid must integrate into the chromosome through homologous recombination to be maintained (Figure 10).[142] Thus, this plasmid constitutes a useful delivery system for experiments requiring homologous recombination between plasmid-borne sequences and the chromosome. The donor strain of pKC1139, *E.coli* S17-1 carries the conjugation locus RP4 in which plasmids could be mobilized. A 2.1 kb KS/V *Bam*HI-digested positive fragment (Figure 10) isolated from subclone 4 of the cosmid pYT27, was cloned into the *Bam*HI site of pKC1139 to generate the plasmid pYT27B8/pKC1139 (Figure 10). Plasmid pYT27B8/pKC1139 and pKC1139 were first transferred into *E.coli* S17-1. The S17-1 transformants were incubated with wild-type *A. verrucosospora* to allow for the transfer of the pYT27B8/pKC1139 plasmid. As a negative control, S17-1 cells with the original pKC1139 plasmid were also incubated with *A. verrucosospora.* The new cultures were challenged with apramycin to eliminate non-plasmid-containing *A. verrucosospora* cells, since the wild-type *A. verrucosospora* strain is resistant to carbenicillin and sensitive to apramycin. In addition, carbenicillin was used to eliminate the original *E. coli* donor cells of the S17-1 plasmid. Finally, after an incubation period of 18 days at 30°C, a single colony of *A. verrucosospora* was recovered which exhibited the phenotype expected of a mutant carrying the pYT27B8/pKC1139 plasmid.

The putative transconjugant, designated *A.verrucosospora:*pYT27B8/pKC1139, was grown in liquid seed culture for a period of 3 days per generation, at 37°C in the presence of apramycin (60 μg/ml). After 4 generations, the cultures were transferred to SY plates containing apramycin (100 μg/ml) and incubated at 42°C, in order to further select the mutants in which the pYT27B8/pKC1139 temperature-sensitive plasmid had integrated into the chromosomal DNA *via* homologous recombination. Once this culture of *A. verrucosospora:* pYT27B8/pKC1139 had sporulated, the spores were streaked onto several new SY plates containing apramycin (60 μg/ml) and re-incubated at 42°C for 3 days. After inoculation and growth of this culture for four generations, several single colonies of the *A.verrucosospora:*pYT27B8/pKC1139 mutant were isolated and examined for their ability to produce verucopeptin, as well as for the presence of the pYT27B8/pKC1139 plasmid in their chromosomal DNA.

Figure 10: Construction of pYT27B8/pKC1139 for gene disruption



A DIG-labeled KS fragment **KS/V-2** (750 bp) was produced by PCR amplification of the 2.1 kb KS-positive insert (Figure 10) using the same degenerate primers previously designed for the amplification of the **KS/V-1** probe(s). The genomic DNA of 2 colonies of *A. verrucosospora* and 5 from the *A. verrucosospora:*pYT27B8/pKC1139 mutant, were isolated and exposed to a single digestion by *Sac*I or a double digestion by *SacI/Bgl*II. All digestion mixtures were subsequently exposed to hybridization conditions with the **KS/V-2** probe and analyzed by Southern blot. As a control, one set of samples (wild type and mutant) were also exposed to a DIG-labeled linear pKC1139 plasmid and analyzed by Southern blot.

The results of these experiments clearly showed the integration of the pYT27B8/pKC1139 plasmid in the expected chromosomal position of *A. verrucosospora:* pYT27B8/pKC1139 mutant. As expected, the DNA of the mutant resulted in different hybridization positive bands than the DNA of wild type *A. verrucosospora* (Figure 11). In the *Sac*I digestion mixture of the mutant, hybridization with the **KS/V-2** probe, the results show two KS positive bands at 6.6 and 7.2 kb (Figure 12, lanes 1 and 3). As expected, these bands were absent from the digestion mixture of the wild type, whereas a KS positive band of 6.0 kb was present (Figure 12, lane 2). These differences are completely consistent with those expected if the integration of the plasmid was to occur at the appropriate location of the chromosomal DNA (Figure 11). The results obtained from the *SacI/Bgl*II digestion mixtures were also consistent with those expected, showing only one positive band of 3.6 kb for the wild-type (Figure 12, lane 5) and two positive bands of 3.0 kb and 4.0 kb for the mutant (Figure 12, lanes 4 and

6). It should be noted that the KS positive band at 3.6 kb observed in all samples, indicates the presence of other KS domain(s) in the chromosomal DNA with identical or nearly identical sequence to that found in the 2.1 kb fragment used in these experiments (Appendix II). When the DIG-labeled pKC1139 vector was used as the probe, only the DNA of the mutant showed hybridization positive bands. Two of the expected bands at 6.6 kb and 7.2 kb are clearly shown in Figure 12 lane 7, an additional band of 0.8 kb was also formed (not shown in this gel) as was expected (Figure 11). Table 9 summarizes the expected positive bands from this Southern blot.

Two additional experiments were performed in order to confirm the characterization of genomic DNA of the transconjugant mutant *A. verrucosospora:*pYT27B8/pKC1139. First the genomic DNA of the mutants was used to successfully amplify the 1.2 kb of the apramycin gene cassette by PCR (Figure 13). Second, the genomic DNA from this mutant could not transforms of *E-coli* XL-blue cells, which indicated the absence of free plasmid DNA as a contaminant of the isolated genomic DNA of the mutant. All of these results are consistent with the successful integration of the pYT27B8/pKC1139 into the chromosomal DNA of *A. verrucosospora* and provide a mechanism by which the involvement of the isolated gene clusters (pYT27 and pYT31) in the biosynthesis of verucopeptin can be explored further.

Figure11: Diagrammatic representation of expected Southern blot results from genomic DNA of wild type and mutant with different enzyme digestions

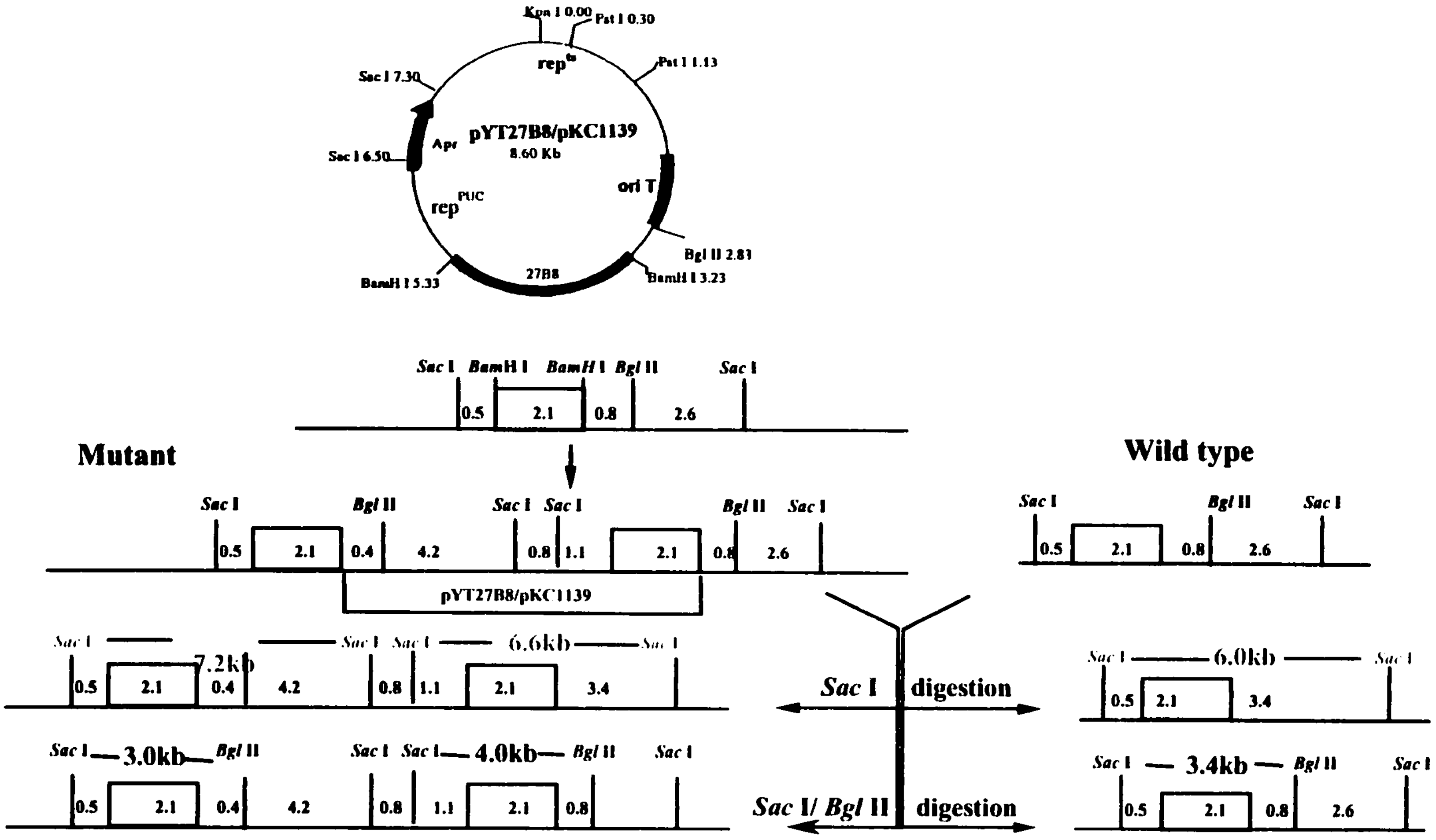
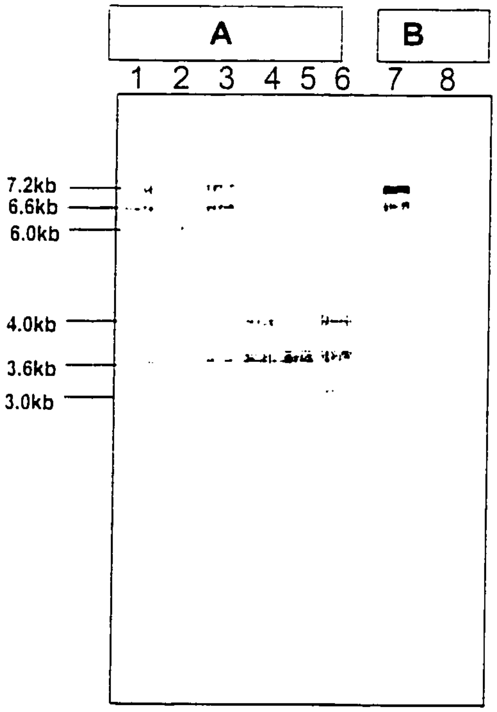


Table 9: Expected positive bands from Southern blot experiments

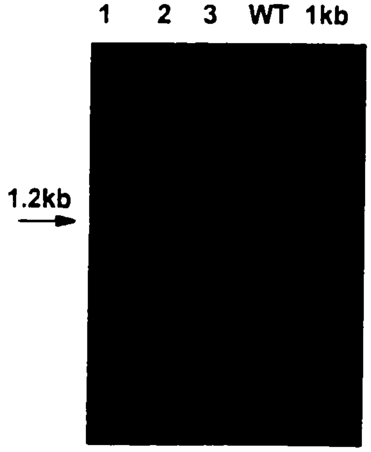
Digestion Enzyme	Genomic DNA	KS/V-2 as probe	pKC1139 as probe
*Sac*I	Wild type	6.0 kb	NO
	Mutant	6.6, 7.2 kb	0.8kb, 6.6kb, 7.2kb
*Sac*I/*Bgl*II	Wild Type	3.6kb	
	Mutant	3.0kb, 4.0kb, 3.6kb	

Figure12: Southern analysis of genomic DNA from wild type and transconjugant mutant.



A: Hybridization with KS/V-2
B: Hybridization with labeled pKC1139
Lane 1, 3, 4, 6 and 7: Genomic DNA of mutants
Lane 2, 5, and 8 : Genomic DNA of Wild type
Lane 1, 2, 3, 7 and 8: Digested with *Sac* I
Lane 4, 5 and 6 : Digested with *Sac* I/*Bgl* II
1-kb ladder (BRL) : Indicated at the left

Figure 13: PCR products of apramycin resistance gene



Lane 1 and 2: genomic DNA of mutants.
Lane 3 : plasmid pKC1139 as positive control
Lane 4 : genomic DNA of wild type.

3.2.3 Production of verucopeptin.

Five single colonies of the *A. verrucosospora*:pYT27B8/pKC1139 mutant and two single colonies of the wild type strain were cultured in liquid production medium at two different temperatures, 30^{0}C and 37^{0}C. Each set of experiments was performed in duplicate in order to minimize experimental error and this experiment was repeated three times with different single colonies. All of the liquid cultures were extracted following the previously developed protocol for the isolation of verucopeptin,[130] and the extracts were analyzed by LC-MS equipped with an evaporative light scattering detection ("ELSD") digraph (Figure 14 and 15). Both the hemiketal and keto forms (14.85 min and 15.98 min) of verucopeptin were observed by LC in the extracts of the wild-type cultures of *A. verrucosospora* and were confirmed by the presence of the appropriate ions in the mass spectra of both LC peaks. Using an authentic sample of verucopeptin, the MS fragmentation pattern was previously shown to contain 4 predominant ions: *m/z* 879 (M+H-H$_2$O)$^+$, 861 (M+H-2H$_2$O)$^+$, 847 (M+H-H$_2$O-CH$_3$OH)$^+$, and 829 (M+H-2H$_2$O-CH$_3$OH)$^+$ (Figure 14 and 15). Although the amount of verucopeptin produced by the wild-type cultures incubated at 37^{0}C was significantly lower than produced by cultures incubated at 30^{0}C, the presence of this metabolite was clearly observed in both cases. In contrast, the presence of verucopeptin could not be detected in any of the mutant cultures, although mycelial growth appeared to be normal. The common minimum amount for detecting a secondary metabolite in crude extract is about 1 ng in this LC-MS equipment. These results strongly suggest that the disrupted gene cluster is indeed associated with the biosynthesis of verucopeptin.

Figure 14: LC-MS Spectra of crude extracts of *A. verrucosospora* and its mutants at 30°C.
I: LC spectra; **II:** MS spectra of the sample eluted at 15.98 min; **III:** MS spectra of the sample eluted at 14.86 min;
A: Wild Type **B: Mutant**

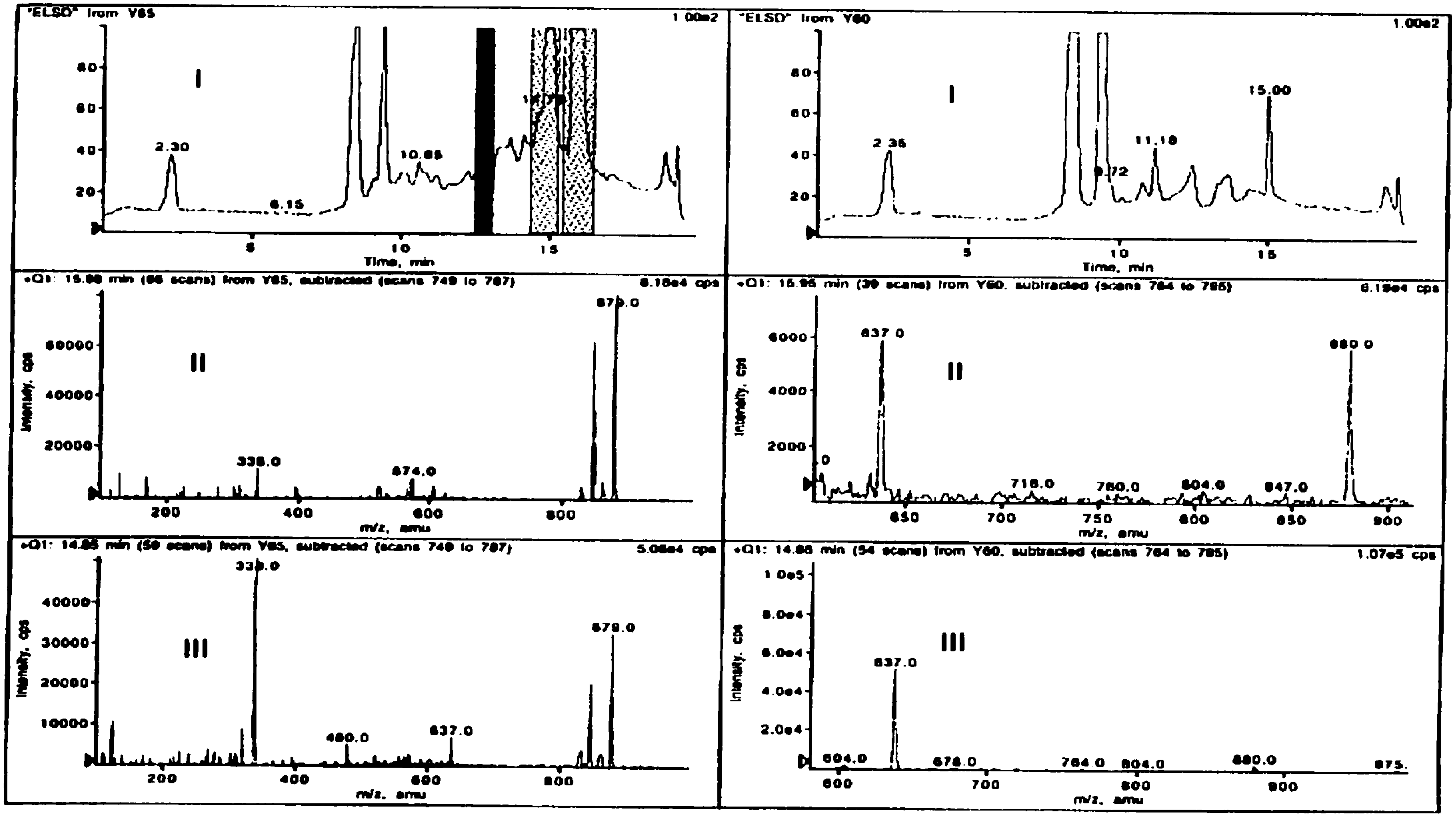
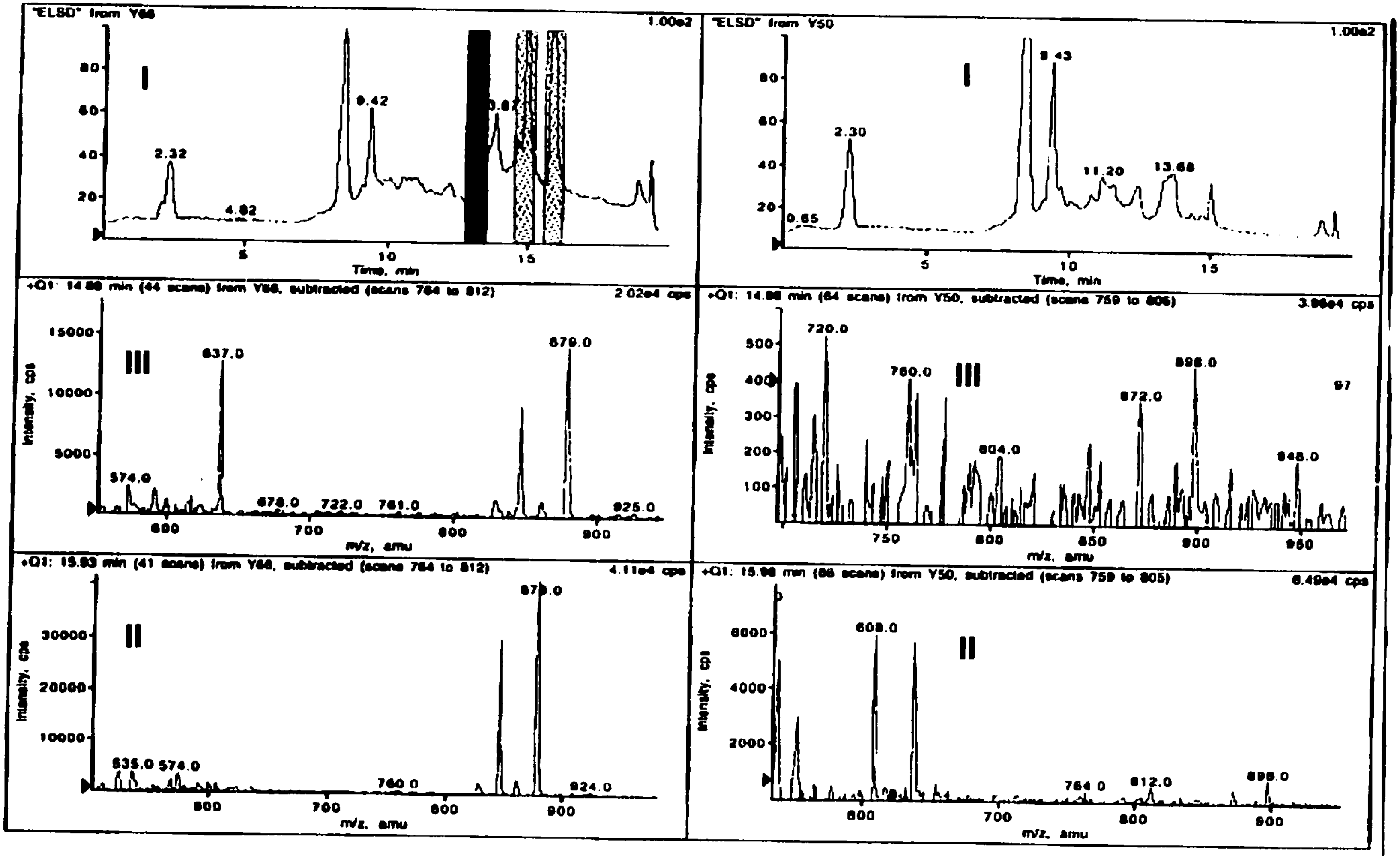


Figure 15: LC-MS Spectra of crude extracts of *A. verrrucosospora* and its mutants at 37°C.
I: LC spectra; **II:** MS spectra of the sample eluted at 15.98 min; **III:** MS spectra of the sample eluted at 14.86 min;
A: Wild Type **B: Mutant**

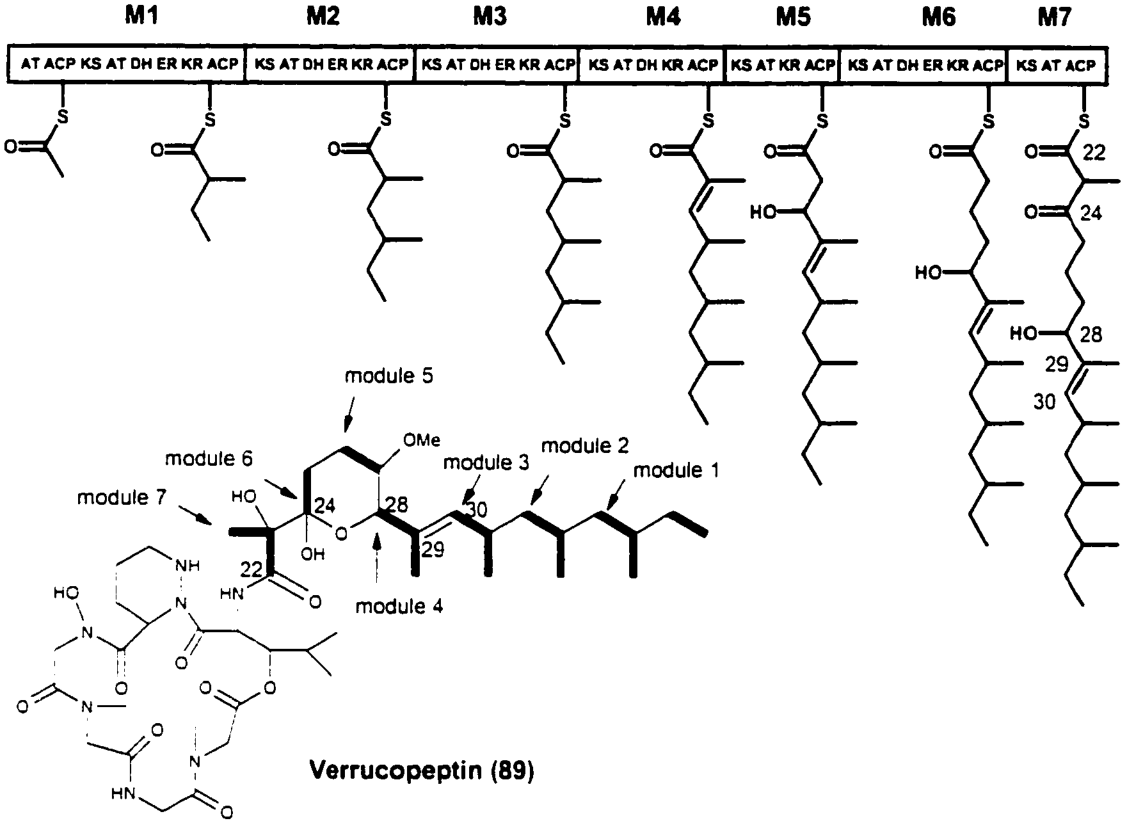


3.2.4. Putative PKS enzymes for the biosynthesis of the tetrahydropyranyl side chain of verucopeptin (89)

From the chemical structure of verucopeptin (**89**, Scheme 24), our feeding experiments and our general knowledge of PKS type I enzyme systems, we believe that seven modules of PKS type I are involved in the biosynthesis of the tetrahydropyranyl side chain of verucopeptin (**89**, Scheme 27). Module 1 should contain a loading domain and a full β-keto functional domain (KR, DH and ER), thus this module contains AT,ACP, KS, AT, DH, ER, KR, and ACP domains. Like module 1, modules 2, 3 and 6 should also contain a full β-keto functional domain (KR, DH and ER), whereas module 4 should contain KR and DH domains to form the double-bond between carbons 29 and 30 (Scheme 27). Module 5 should contain a KR domain to add the hydroxyl group at carbon 28, whereas module 7 should not contain any β-keto functional domain and should leave the carbonyl alone at carbon 24 so that this carbonyl with the hydroxyl group at carbon 28 can form a six-membered hemiketal ring (Scheme 27). The PKS enzyme of verucopeptin, like FK506 and rapamycin, may not contain a TE domain which is usually located in the last module. We assume that entire polyketide chain may be directly transferred to the NRPS enzyme to form the verucopeptin backbone skeleton.

From our feeding experiments, the polyketide side chain of verucopeptin is derived from five propionate and three acetate units. We believe that the AT domains of modules 1, 2, 3, 4 and 7 should use methylmalonyl-CoA as their substrate, whereas the AT domains of modules 5, 6 and the loading domain should use malonyl-CoA as their substrate.

Scheme 27: The proposed modules for the biosynthesis of the tetrahydropyranyl side chain of verucopeptin (**89**)



Verrucopeptin (89)

3.2.5 DNA sequence analysis:

A summary of the restriction maps of pYT27 and pYT31 are shown in Figure 9. The overlapping region of pYT27 and pYT31 was also confirmed by the independent sequencing of the 300 bp DNA fragments of subclone S7 of pYT27 and subclone S9 of pYT31, respectively (Appendix III).

3.2.5.1 **Polyketide Synthase type I Encoding Regions: Partial DNA Analysis and Characterization**

In order to further characterize the cloned region of *A. verrucosospora* and to identify any similarities between the PKS genes associated with the biosynthesis of verucopeptin and those of known PKS type I enzymes, we sequenced DNA starting at the beginning of subclone S2. An region of 4,970 bp from pYT27 (Figure 9) was sequenced. The details of the nucleotide sequence are given in Appendix IV. As expected, a high G+C content (74.68%), typical for both *Streptomyces* and *Actinomadura*, was observed. Sequence analysis using the GCG and PC/Gene programs, and comparison with other known PKS type I genes, revealed the presence of one complete open reading frame (ORF) of 3,123 bp, which we assigned to module 7 of the PKS gene cluster associated with verucopeptin (Scheme 27). This ORF has a start codon (ATG) at nucleotide 1,786 which is in the plasmid pS3-1, and ends with a stop codon (TGA) at 4,909 bp which is located in the middle of subclone S2 (Figure 9). PS3-1 was a 4.1 kb DNA fragment from subclone 4 digested with *Eco*RI/*Bam*HI which was cloned into pLitmus 28. The expected putative ribosomal binding site (RBS) GGAAG was also observed at 1,777 bp (9 bp upstream of start codon).

Identification of the enzymatic motifs encoded by these genes was delineated using the BLAST program and the type I PKS enzyme primary sequences deposited in GenBank, EMBL, and Swissprot databases. Several regions were identified with a high degree of similarity to PKS type I enzymes from the biosynthetic pathways of well known microbial metabolites, such as erythromycin, FK506 and rapamycin. The complete set of the three essential PKS domains, KS, AT and ACP, was present as would be expected in the proposed module 7 (Scheme 27); module 7 is not expected to contain any of the enzymes required for the structural modification of a β-carbonyl moiety (i.e. KR, ER, DH domains). It should be noted that a TE domain, commonly found as the last module of the PKS enzymes involved in the biosynthesis of most macrolide antibiotics, is not present in module 7. This is analogous to the last PKS module of metabolites having a mixed biosynthetic origin, such as FK506.[28] In addition, this observation is consistent with our hypothesis that the entire polyketide product is perhaps loaded onto the peptide synthetase thiotemplate for further assembly of the verucopeptin backbone skeleton.

Detailed analysis of the 424 amino acids of the putative KS domain in module 7 of the verucopeptin PKS enzyme system revealed an average of 65% sequence identity with many other KS domains of PKS type I enzymes. For example, 58.01% and 68.66% sequence homology were observed between the KS domain of module 7 of verucopeptin PKS and eryKS2 and rapKS1, respectively (Figure 16). The sequences corresponding to the two KS primers, YXS1 (DTACSS) and YXS2 (SGTNAH), initially used to screen our cosmid library (Table 5), were also present and are shown in Figure 16. As expected, the amino acid sequence ^{804}DTACSSSL which is perfectly conserved around the active site of all PKS type I enzymes, contains the ^{807}Cys required for catalysis in the formation of a thioester linkage to the growing acyl chain.[143] Furthermore, the two amino acid residues ^{943}His and ^{982}His are consistent with the 2 His residues which are invariant at the C-termini of PKS type I enzymes. It has been proposed by Leadlay that one of these His residues may increase the nucleophilicity of the active site Cys by acting as a general base.[142]

The AT domains of known PKS type I enzymes can be mainly divided into two subgroups depending on their substrate specificity for methylmalonyl-CoA or malonyl-CoA chain extender units.[144] The sequence homology with known malonyl AT domains and that of module 7 was found to be less than 30%. In contrast, this AT domain was homologous to other methylmalonyl AT domains, such as the eryAT3 (55.8% homology) and the eryAT4 (49% homology) (Figure 16). This domain contains the consensus sequence motif ^{1228}RVDVVP-6-MVSmAalW for incorporating the propionate extender units.[143] Two other regions, ^{1259}GHSQG and^{1361}YASH, that allow the unambiguous assignment of the methylmalonyl-CoA specificity were also indicated in Figure 16.[56] The GHSQG is the active site signature sequence in the formation of the acyl-enzyme intermediate. It is also worth noting that our earlier incorporation experiments of ^{13}C-labeled propionate into verucopeptin are completely consistent with the results obtained from the sequence analysis of the AT domain in module 7, indicating that a propionate unit serves as the substrate of the AT domain in this module (Scheme 27).

The amino acid sequence corresponding to the ACP domains identified in module 7 was also consistent with PKS type I enzymes, exhibiting sequence homology from 42% (nid ACP6) to 65% (nid ACP5).

In addition to the complete ORF associated with module 7, the 4,965 bp DNA fragment (Appendix IV) also contains the terminal portion of another ORF (stop codon ^{1782}TGA) which we have tentatively assigned to module 6 of the PKS gene cluster associated with the biosynthesis of verucopeptin (Scheme 27). The start codon of this gene lay outside of

4.9 kb DNA fragment and has not yet been sequenced. The nucleotide sequence found in the overlapping region of the two ORFs, encodes a peptide sequence having a high percentage of hydrophilic and charged (bold) amino acid residues (^{594}M**DD**AA**DR**APA**D**GA**DR**V**ER**AL**R**ALL**EERDR**L**RRE**N**DD**L**K**AG**R**G). This amino acid pattern fits well with Khosla's hypothesis for interpolypeptide linker sequences which are usually longer and more hydrophilic than those found in intermodular peptide sequences.[145]

Based on the molecular structure of verucopeptin (Scheme 24 and 27) and the results of our original precursor feeding experiments (scheme 25), module 6 should contain the essential domains (KS, AT and ACP) and complete β-carbonyl modification domains (KR, DH and ER). From the sequence analysis of the partial ORF of module 6, the ER, KR, and ACP domains were identified (Appendix IV).

The putative KR domain in module 6 (scheme 27) was found to be typical of PKS type I KR domains, having a sequence homology ranging from 31.25% (nid KR6) to 56.13% (tyl KR2). As expected, an active NADPH-binding site, ^{192}GxGxxGxxxA was observed and is highlighted in Figure 16.

A sequence identity of 53.52% was observed between the two ACP domains of modules 6 and 7. They both contain the amino acid motif LGfDS associated with the active site of all PKS type I ACP domains (Figure 16). The phosphopantotheine-binding residue,

^{504}Ser and ^{1588}Ser in module 6 and 7 respectively, was also observed in the common GFDSL motif of the two ACP domains (Figure 16).[45]

The partial ER domain containing the invariant sequences ^{7}As**GVGMAAVQLA** for the NADP^{+}(H) binding was also found in module 6 (Figure 16). However, a serine (^{8}Ser) residue is present at the first position of the NADP^{+}(H) binding site rather than a glycine (^{8}Gly) residue. A replacement of the first glycine residue in the NADP^{+}(H) binding-site of the ER domain of module 9 by a threonine residue, without any loss in catalytic functions, was previously observed in the biosynthesis of FK506.[29] It is not yet known if these minor differences in the key residues of the ER domain have any effect on its catalytic function.

Figure 16: Alignments of sequences of conserved motifs found in PKS domains.

1. Enoyl reductase (ER)

```
VerER6  3   LIHAAASGVGMAAVQLAQH----GVDVVLNSL----GKTDIRD----RHVGKLVL-177
EryER4  114 LIHAAAGGVGMAAVALARR----GVDVVLNSL----GKTDIRD----RHVGKLVL-285
NidER5  168 LVHAAAGGVGMAAVQIARH----GVDVVLNSL----GRTDVRD----RHVGKLVL-349
RapER1  112 LIHAAAGGVGMAATQIARH----GVDVVLNSL----GKTDIRD----RHVGKLVL-289
RapER13 112 LIHAAAGGVGMAATQIARH----GVDVVLNSL----GKTDIRD----RHVGKLVL-288
RapER7  112 LIHAAAGGVGMAATQIARH----GVDVVLNSL----GKTDIRD----RHIGKIVL-289
consensus   L.HAAA.GVGMAA...A..    GVDVVLNSL    G.TD.RD    RH.GK.VL
```

(Note: only partial DNA sequences available of VerER6)

2. β-Ketoreductase (KR)

```
VerKR6 186 GTVLVTGGTGGVGSLVARHLAAEHGVRSLVL---WGWW--369
EryKR4     GTVLITGGTGTLGRLLARHLVTEHGVRHLLL---WGLW--182
NidKR5     GTVLITGGTGTIGSRIARHLVTRHGVRHLLL---WGLW--184
EryKR1     GTVLVTGGTGGVGGQIARWL-ARRGAPHLLL---WGLW--182
RapKR1     GTILITGGSGVLAGILARHLAAEHGARHLLL---WGLW--182
RapKR13    GTILITGGSGVLAGILARHLAAEHGARHLLL---WGLW--184
RapKR4     GTVLITGGSGVLAGIAARHLVAERGVRHLLL---WGLW--184
RapKR7     GTVLVTGGSGVLAGIAARHLVAEQGVRHLLL---WGLW--184
NidKR6     -TILITGGTGALGTHIATWL-AHKGAKHLIL---WGLW--183
Consensus   T.L.TGG.G .G/A A  L . .G.  L L   WG W
```

3. Acyl-carrier protein (ACP: VerACP6 from 471-541 and VerACP7 from 1555-1625)

```
VerACP6   VLGHT---FTHLGFDSL---TGLRLPSTLVFSYPTPRELGRHIL 71
VerACP7   VLGHA---FKDLGFDSL---TGVRLKPTLVLRHPSPRLIAGHIA 71
EryACP1   VLGHA---FAELGVDSL---TGVRLPTTTVFDHPDVRTLAAHLA 70
EryACP4   VSGYG---FKDLGFDSL---TGVRLPSTLVFDHPTPLAVAEHLR 71
EryACP2   VLGHD---FKELGFDSL---TGLRLPSTLVFDHPNASAVAGFLD 86
NidACP5   VLGHG---FREAGFDSL---TGLRLPATLVFDHPTPTALAGRLD 86
RapACP1   VLGHA---FRDLGVDST---TGLRLPATLVFDYPTPAALAARLE 73
RapACP7   VLGHA---FKDLGIDSL---TGLRLPATLVFDYPTPTALAARLG 74
RapACP4   VLGHA---FKDLGVDSL---TGLRLPPTLVFDYPTPTALAARLD 74
EryACP3   VLGHE---FSELGLDSL---TGLRLPASLVFDHPTVTALAQHL- 70
NidACP6   VLGHD---FKDLGFDSL---TGLDLPATVIFDHPTV----DRLA 98
RapaCP13  VLGHS---FKDLGMDSL---TGLQLPATMVFDYPTANALAAHLL 72
Consensus V G..   F . G DSL   TG. L .. .. .P.
                         *
```

4. β-ketoacyl acyl carrier protein synthetase (KS)

```
           1                173        308        418
VerKS7 636 PIAVVAMACRFPG----DTACSSSL----VEAHGTGT----SGTNAH--1060
RapKS1     PLAIVGMACRLPG----DTACSSSL----VEAHGTGT----SGTNAH--418
RapKS13    PLAVVGMACRLPG----DTACSSSL----VEAHGTGT----SGTNAH--417
RAPKS4     PLAIVGMACRLPG----DTACSSSL----VEAHGTGT----SGTNAH--417
RAPKS7     PLAIVGMACRLPG----DTACSSSL----VEAHGTGT----SGTNAH--417
EryKS1     PVAVVAMACRLPG----DTACSSSL----VEAHGTGT----SGTNAH--459
EryKS4     PIAIVGIGCRFPG----DTACSSSL----VEAHGTGT----SGTNAH--469
ErrKS3     PIAIVSMACRLPG----DTACSSSL----VEAHGTGT----SGTNAH--462
NidKS5     PIAIIGMACRLPG----DTACSASL----VEAHGTGT----SGTNAH--430
NidKs6     -IAIVGMACRLPG----DTACSSSL----VEAHGTGT----SGTNAH--425
EryKS2     PIAIVGMACRLPG----DTACSSSL----VEAHGTGT----SGTNAH--462
Consensus   .A.....CR.PG    DTACS.SL    VEAHGTGT    SGTNAH
```

5.Acyltransferase (AT) for propionate VerAT7 from 1169 to 1489

```
VerAT7  FMFPGQGSQW--RVDVVQP--MVSMAALW---GHSQGEIAAA--RSR--VDYASHSHEV-
EryAT3  FLFPGQGSQW--RVDVVQP--MVSLAELW---GHSQGEIAAA--RSR--VDYASHSPQI-320
NidAT4  FVFPGQGAQW--RVDVVQP--VVALAAMW---GHSQGEIAAA--RSQ--VDYASHSRHV-330
EryAT2  FVFPGQGAQW--RVDVVQP--MVSLARLW---GHSQGEIAAA--RSR--VDYASHTAHV-346
EryAT4  LVFPGQGAQW--RVDVLQP--MVSLAELW---GHSQGEIAAA--RSQ--VRYASHSPEV-322
NidAT5  FVFPGQGSQW--RVDVVQP--MVSLAAVW---GHSQGEIAAA--RSR--VDTAGHSPQV-328
NidAT6  FVFPGQGAQW--RADVVQP--MVSLAALW---GHSQGEVAAA--RSR--VDFASHSPQV-345
EryAT1  FVFPGQGWQW--RVDVVQP--MVSLASMW---GHSQGEIAAA--RSR--VDYASHSSHV-325
RapAT1  FVFPGQGWQW--RVDVVQP--MVSLAAVW---GHSQGEIAAA--RSQ--VDYASHTPHV-307
RapAT7  FVFPGQGWQW--RVDVVQP--MVSLAAVW---GHSQGEIAAA--RSQ--VDYASHTPHV-307
RapAT4  FVFPGQGWQW--RVDVVQP--MVSLAAVW---GHSQGEIAAA--RSQ--VDYASHTPHV-306
RapAT13 FVFPGQGWQW--RVDVVQP--MVSLAAVW---GHSQGEIAAA--RSQ--VDYASHTPHV-307
consensus FPGQG QW--R.DV.QP. .V..A .W---GHSQGE.AAA  RS.  V YASH. .
             *                            *         *         *
```

Sequences were aligned by using the PCGENE program:

1). Perfectly conserved amino acid residues in all the sequences are marked in the bottom line.

2). Well-conserved amino acids in all sequences are marked as "."

3). Active-site residues are marked with asterisks. An amino acid indicated in bold is a conserved residue or motif important for function.

4). The numbers on the right side of other known PKS domains represent the amino acid number of each domain.
5). Ver: verucopeptin; Ery: erythromycin; Rap: rapamycin; Nid:niddamycin.

3. 2. 5. 2. NRPS Encoding Regions: Partial DNA Analysis and Characterization

A DNA fragment of 1,270 bp, adjacent to the 3.9 kb **Pep/V** positive band (Figure 9), was subcloned and sequenced (Appendix V). The subsequent analysis of both the DNA sequence and its expected translation products by GCG and PC/GENE programs and BLAST, revealed the presence of two partial open reading frames (ORFs). The protein encoded by one of the ORFs (70 amino acids from the *Bgl*II restriction site, Appendix V) showed 30%-50% sequence identity to thiolation domains of several known peptide synthetases, including those of gramicidin S (**52**) synthetase,[90] cyclosporin C (**55**) synthetase[92] and pristinamycin I_A synthetase[95] (Table 10). The amino acid sequence of the thiolation domain contains the well-known invariant sequence ^{18}DXFFXLGGXSLI, which is the binding site of the 4'-phosphopantetheine transferase cofactor. The primer 26-22B contained the amino acid sequence FFA(E,T,S,D)**LGG**HS which could be found as ^{20}FFALGGNS in the sequenced region. A sequence comparison of this region with other known thiolation domains is shown in Table 10. The identification of a stop codon at nucleotides ^{213}TGA, suggested that the start codon for this ORF lay outside the sequenced DNA region (Appendix V).

Table10: Sequence alignment around the highly conserved cofactor 4'pp binding site:

Enzyme	Organism	Position (aa)	Sequences
Ver-THIOL	*A. verrucosospora*	18	**D D F F A L G G N S L**
FkbL[28]	*S. sp. MA6548*	1022	**D E F F A L G G H S L**
Rap L[29]	*S. hygroscopicus*	1028	**D D F F A L G G H S L**
Ta1[138]	*Myxococcus xanthus*	951	**D R F F E V G G D SV**
SnbC[95]	*S.pristinaespiralis*	3464	**D N F F D L G G H S L**
CpsB[92]	*S.chrysomallus*	1022	**D E F F A L G G H S L**
TycA[89]	*Bacillus brevis*	553	**D N F Y S L G G H S I**
GrsB[88]	*Bacillus brevis*	2033	**D N F F E L G G H S L**
Cssa[91]	*Tolypocladium inflatum*	5528	**D N F F E L G G H S L**
Consensus			**D F LGG S**

Note: FkbP:pipecolate-incorporating enzyme in FK506;
RapP: pipecolate-incorporating enzyme in rapamycin;
Tal:TA 1 synthetase; *SnbC*: Pristinamycin synthetase;
CpsB: Daptomycin synthetase.
TycA: tyrocidine synthetase A;
GrsB: gramicidin S synthetase *A*;
Cssa: cyclosporin A synthetase;

A start codon for the second ORF (^{317}ATG) was identified (Appendix V), which was preceed by a putative RBS (GGACG) 7 bp upstream. The expected 318 amino acid translation product of this ORF showed an approximate 30% identity to other known cytochrome P450 enzymes (Table 11). The ORF encoding the putative cytochrome P450 enzyme follows the ORF encoding the peptide synthetase, but, its stop codon lies outside of the sequenced DNA region.

Most P450 enzymes are proteins of 400-530 amino acids, with the exception of the *Bacillus* BM-3 (CYP102)[146] which is a protein composed of 1048 amino acids. Although the putative cytochrome P450 enzyme identified from our analysis of the *A. verrucosospora* chromosomal DNA (Figure 9) is incomplete, it contains all of the necessary catalytic elements for a functional P450 monooxygenase. For example, the likely O_2 binding site with a conserved threonine residue ^{253}Thr, believed to be involved in O_2 scission, was identified (Appendix V, ^{248}LLGANV^{253}TT). In addition, the expected heme-binding site with its conserved cysteine residue ^{359}Cys was also observed (^{352}FGYGPHY^{359}CVGA). Table 11 shows a sequence comparison between this enzyme and the amino acid sequences of several other known P450 monooxygenases associated with the biosynthesis of macrolide and peptide metabolites from microbial sources.

Table 11: Highly conserved amino acid sequences of active-site of P450s

Enzyme	O_2 binding pocket amino acids	Heme-binding pocket amino acids
VerP450	ILLGANVTTP	FGYGPHYCVGAPLARLT
PicK[147]	ILLVAGHETT	FGHGIHFCIGAPLARLE
OleP[148]	SLLIAGHETS	FGHGAHHCIGAQLGRLE
TylI[149]	LLLTAGHISS	FGHGIHHCLGSFLARLE
MycG[150]	GLLVAGYETT	FGHGVHHCLGAPLARVE
EryK[77]	ALLLAGHITT	FGHGVHFCLGAPLARLE
EryF[69,73]	VLLLAGFEAS	FGQGIHFCMGRPLAKLE
CarP450[151]	TFVTAGNETT	FGFGPHYCLGAALAHQE
TaH[198]	TLLMAGHETS	FGGGIRKCIGTSFAYYE
consensus	L A **T**	FG G **C** G

Note: PicK: narbomycin C-12 and YC-17 C-10 hydroxylase.
OleP: putative oleandomycin 8, 8a-epoxidase.
TylI: tylosin C-20 oxidase.
MycG: mycinamicin oxidase.
EryK: erythromycin C-12 hydroxylase.
EryF: erythromycin C-6 hydroxylase.
CarP450: putative carbomycin P450 oxidase.
TaH: putative TA C-20 hydroxylase.

3.3. Summary

We have demonstrated that the polyketide portion of verucopeptin is derived from three acetate and five propionate units which suggests the requirement of seven PKS type I modules.[137] The expected modules should contain these active domains as shown in Scheme 27.

Our results, based on partial sequencing of the cloned DNA and its translation products (proteins), are completely consistent with the expected catalytic functions for modules 6 and 7 (Scheme 27). Some specific tailoring enzymes must also be involved in the post-translational modifications of both the polyketide and the depsipeptide portion of verucopeptin. The production of the hydroxyl groups at C-23 and C-27 in verucopeptin may be catalyzed by P450 enzymes. The methyl moiety of the methoxy group at C-27 could be derived from methionine, catalyzed by *S*-adenosylmethionine methyltransferase in a manner analogous to that involved in the biosynthesis of erythromycin and rapamycin.[69,152] The source of the two unusual amino acids, β-hydroxyleucine and *N*-hydroxyglycine, in the depsipeptide portion of verucopeptin is not known with certainty. Currently, the mechanism of hydroxylation in the modular NRPS systems is not known. Nonetheless, it would be reasonable to expect that the genes coding the six NRPS modules associated with the construction of the depsipeptide portion of verucopeptin would be clustered in the chromosomal DNA of *A. verrucosospora*. A DNA fragment of approximately 26 kb, such that identified in cosmid pYT31, should contain most of the genetic information required for the catalytic functions needed for the biosynthesis of the depsipeptide portion of verucopeptin.

To the best of our knowledge, this work represents the first analysis of a gene cluster encoding both PKS and NRPS enzymes which are contained at a discreet location on the chromosomal DNA of a microorganism. In contrast, the genes coding for the RapP and FkbP enzymes are in the middle of the PKS gene clusters involved in the biosynthesis of rapamycin and FK506. This work is also the first report of a gene disruption experiment involving a PKS type I gene cluster responsible for the biosynthesis of a secondary metabolite from the genus *Actinomadura*. Thus, it offers the first opportunity to compare this system with well-characterized PKS systems from the *Streptomyces*

Further work is needed in order to find the border of the gene region coding for NRPS and to completely characterize the complete genes involved in the biosynthesis of verucopeptin.

3.4 Significance

We have reported the isolation and partial DNA sequences of the PKS and NRPS genes involved in the biosynthesis of the antitumor antibiotic verucopeptin. A KS DNA fragment of 1.1 kb from the gene cluster associated with the biosynthesis of erythromycin A (*ery* module 1) was initially used to screen the genomic library of *A. verrucosospora*. The positive colonies identified were further screened with two types of selective probes, KS/V-1 and Pep/V, generated by PCR amplification of the polyketide and non-ribosomal peptide synthetase regions, respectively, in the chromosomal DNA of *A. verrucosospora*. The primers used for these PCR experiments were designed based on highly conserved amino acid sequences of PKS type I KS domains (e.g. KS domains of erythromycin and rapamycin) and the known highly conserved regions of the adenylation (motif 7) and the thiolation domains of non-ribosomally formed microbial peptides.

A DNA fragment of 72.5 kb was isolated from the cosmid library of *A. verrucosospora* chromosomal DNA (cosmid pYT27 and cosmid pYT31). Partial analysis and sequencing of this gene cluster revealed the presence of several open reading frames (ORFs) encoding modular polyketide synthase (PKS) type I enzymes and multifunctional, non-ribosomal peptide synthetase enzymes (NRPS). Based on the chemical structure of verucopeptin, the enzymatic motifs identified are consistent with those expected to be associated with its biosynthesis.

In order to confirm that the cloned genes encode proteins involved in the biosynthesis of verucopeptin, a gene disruption experiment was designed to eliminate the production of verucopeptin. The commercially available conjugate vector pKC1139 was used to construct the plasmid pYT27B8/pKC1139 in which a 2.1 kb KS positive band from pYT27 was inserted into its *Bam*HI site. Comparison of the chromosomal DNAs from wild-type and mutant *A. verrucosospora* by Southern blot analysis clearly showed the integration of the pYT27B8/pKC1139 plasmid into the expected location of chromosomal DNA in the mutant microorganism. Furthermore, LC-MS analysis clearly showed that this mutant was unable to produce the key metabolite, verucopeptin, under the usual fermentation conditions and growth medium. Thus, we have concluded that the PKS gene cluster associated with the biosynthesis of verucopeptin had been successfully disrupted at the chromosomal level. At this point, the involvement of the genes found in cosmid pYT31 remains to be confirmed. However, the assumption that these genes are associated with the required non-ribosomal peptide synthetase and cytochrome P450 enzymes is consistent with the common observation that the genes associated with the biosynthesis of microbial polyketides and peptides are usually closely clustered on the chromosomal DNA of the producing microorganism.

3.5. Reference:

---

[130] (a) Nishiyama,Y.; Sugawara, K.; Tomita, K.; Yananoto, H.; Kami, H.; Oki, T. *J. Antibiot.* **1993**, *46*, 921

(b) Sugawara, K.;Toida, S.;Moriyama, T.; Konishi, M.; Oki, T. *J. Antibiot.* **1993**, *46*, 928

[131] Smitka, T.A.; Deeter, J. B.; Hunt, A. H.; Mertz, F. P.; Ellis, R. M. L.; Boeck, D.; Rao, C.Y. *J. Antibiot.* **1988**, *41*,726

[132] Maehr, H.; Liu, C-M.; Palleroni, N. J.; Smallheer, J.; Todaro, L.; Williams, T. H.; Blount, J. F. *J. Antibiot.* **1986**, *39*,17

[133] Nakagawa, M.; Hayakawa, Y.; Furihata, K.; Seto, H. *J. Antibiot.***1990**, *43*,477

[134] Ueno, M.; Amemiya, M.; Someno, T.; Masuda, T.; Iinuma, H.; Naganawa, H.; Hamada, M.; Ishizuka, M.; Takeuchi, T. *J. Antibiot.* **1993**, *46*, 1658

[135] Hensens, O. D.; Borris, R. P.; Koupal, L. R.; Caldwell, C. G.; Currie, S. A.; Haidri, A. A.; Homnick, C. F.; Honeycutt, S. S.; Lindenmayer, S. M.; Schwartz, C. D.; Weissberger, B. A.; Woodruff, H. B.; Zink, D. L.; Zitano, L.; Fieldhouse, J. M.; Rollins, T.; Springer, M. S.; Springer, J. P. *J. Antibiot.* **1991**, *44*,249.

[136] (a) Sakai, Y.; Yoshida, T.; Tsujita, T.; Ochiai, K.; Agatsuma, T.; Saitoh, Y.; Tanaka, F.; Akiyama, T.; Akinaga, S.; Mizukami, T. *J. Antibiot.* **1997**, 50,659

(b) Agatsuma, T.; Sakai, Y.; Mizukami, T.; Saitoh, Y. *J. Antibiot.* **1997**, *50*, 704

[137] Tsantrizos, Y.S.; Shen, J.; Trimble, L.A. *Tetra . Let.* **1997**, *38*, 7033

[138] Paitan, Y.; Alon, G.; Orr, E.; Ron, E.Z.; Rosenberg, E. *J. Mol. Biol.* **1999**, 286, 465

[139] (a) Dairi, T.; Hamano, Y.; Igarashi, Y.; Furumai, T.; Oki T. *Actinomycetol.* **1997** 11,1

(b) Dairi, T.; Hamano, Y.; Igarashi, Y.; Furumai, T.; Oki T. *Biosci. Biotech. Biochem.* **1997**, *61*,1445

(c) Dairi, T.; Hamano, Y.; Furumai, T.; Oki. *Appl. Environ. Microbiol.* **1999**, ***65***, 2703

[140] (a) Fischer, A.; Kroppenstedt, R. M.; Stackebrandt. **E.** *J. Gen. Microbiol.* **1983**, 129,3433

(b) P iron-Fraipont, C.; Duez, C.; Matagne, A.; Molitor, C.; Dusart, J.; Frere, J-M.; Ghuysen. J-M. *Biochem. J.* **1989**, *262*, 849

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

[142] Muth, G; Nubbaumer, B; Wohlleben, W; Puhler, A. *Mol. Gen. Genet,* **1989**, *219*, 341

[143] Aparicio,J. F.; Molnar, I.; Schwecke, A.; Konig, A.;Haydock, S.F.; Khaw, L. E.; Staunton, J.; Leadlay, P.F. *Gene,* **1996**, *169*, 9

[144] Haydock,S.F.; Aparicio, J. F.; Molnar, I.; Schwecke, T.; Khaw, L. F.; Konig, A.; Marsden, F. A.; Galloway I. S.; Staunton, J.; Leadlay, P.F. *FEBS Letters* **1995**, *374*, 246

[145] Gokhale, R. S.; Tsuji, S.Y.; Cane, D. E.; Khosla, C. *Science* **1999**. *284*, 482

[146] Paitan, Y.; Orr, E.; Ron, E.Z.; Rosenberg, E. *Gene,* **1999**. *288*,147

[147] (a) Betlach, M. C.; Kealey, J. T.; Betlach, M. C.; Ashley, G. W.; McDaniel, R. *Biochemistry* **1998**. *37*,4937

(b) Xue, Y.; Wilson, D.; Zhao, L.; Liu, H-W.; Sherman, D.H. *Chem. Biol.* **1998**. *5*, 561

[148] Rodriguez, A. M.; Olano, C.; Mendez, C.; Hutchinson, C. R.; Salas, J. A. *FEMS*

Microbiol. Lett. **1995**. *127*, 117

[149] Merson-Davies, L. A.; Cundliffe, E. *Mol. Microbiol.* **1994**. *13*, 349

[150] Inouye, M.; Takada, Y.; Muto, N.; Beppu, T.; Horinouchi, S. *Mol. Gen. Genet.* **1994**. *245*, 465

[151] Arisawa, A.; Tsunekawa, H.; Okamura, K.; Okamoto, R. *Biosci. Biotechnol. Biochem* **1995**, 59, 582

[152] Molnar, I.; Aparico, J. F.; Haydock, S. F.; Khaw, L. E.; Schwecke, T.; Konig, A.; Staunton, J.; Leadlay, P. *Gene* **1996**. *169*, 1

Chapter 4: Experimental

A. Experimental: Studies on the biosynthesis of oudenone

4.1 Materials

NMR spectra were obtained at 20-22 °C. ^{1}H, ^{2}H, and ^{13}C NMR chemical shifts are given in ppm and are referenced to the internal deuterated solvent (^{2}H NMR spectra were recorded in $CHCl_3$). All reactions carried out under anhydrous conditions were performed under a nitrogen atmosphere using oven-dried syringes and glassware except hydrolysis and decarboxylation of compound **78** and the final step for compound **86**.

THF was distilled from sodium/benzophenone, CH_2Cl_2 was distilled from P_2O_5, MeOH was distilled from Mg turnings, toluene was distilled from $CaCl_2$, and DMF from CaO. Reagents and solvents were purchased from Aldrich Chemical Co. and VWR Scientific of Canada, respectively. Isotopically labeled reagents were purchased from Cambridge Isotope Laboratories.

Stock cultures of *O. radicata* ATCC 20295 were maintained on PDY (ATCC medium No 337) plus 1.5% agar in slant tubes at 4°C. All culture media and glassware were autoclaved prior to use, and all biological manipulations were conducted in a sterile hood. *O. radicata* cultures were grown at 27°C and 140 rpm, in a rotary incubator shaker.

Flash column chromatography was carried out on Merck Kieselgel 60, 230-400 mesh. Reversed-phase flash column chromatography was carried out on silica gel reacted with *n*-octadecyltrichlorosilane, following previously reported procedures.[153]

4.2 Fermentation of *O. radicata* ATCC 20295 for precursor feeding and isolation of oudenone.

Autoclaved medium (2 g of glucose, 5 g of Avicel microcrystalline cellulose, 0.5 g of yeast extract, 100 ml of distilled water, in a 500 ml flask) was inoculated with 10-15 small scrapings of mycelia from a PDY agar plate of *O. radicata.*[118] The culture was grown in an incubator shaker at 27^0C and 140 rpm for 15-18 days. Oudenone (**67**) production was usually observed after 7-9 days of growth, and a maximum concentration was reached after approximately 14-16 days. Before and after feeding of labeled precursors, oudenone production was monitored by the UV method (below). In all cases, ^{2}H-labeled compound **86** was dissolved in absolute ethanol (~1.5 mg sample/100 μl of ethanol) and administered in 100 μl aliquots, every 6-12 h (3x/day) for a period of 48 h, to growing cultures of *O. radicata* (100 ml of fermentation broth). Between feedings, the ethanolic solution of compound **86** was stored at -85°C to prevent decomposition. At the same time, the β-oxidation inhibitor 3-(tetradecylthio)propanoic acid (**88**) (1.5 mg in 30 μl of ethanol) was also administered in order to suppress degradation of compound **86**. All cultures were harvested after 24 hours from the last feeding and the labeled metabolite **67** was isolated.

4.3 UV Assay for monitoring oudenone (67) production

Fermentation broth aliquots (3 ml) were withdrawn from an actively growing culture every 12-24 h and filtered. Supernatant of 1.0 ml was diluted to 10 ml with distilled H_2O. The diluted solution of 1.0 ml was subsequently added to 9.0 ml of phosphate buffer (pH=7.0) and to 9.0 ml of 0.1N HCl respectively. The λ max of oudenone is at 246 nm in phosphate buffer (pH=7.0), whereas it shifts to 284 nm in acid (0.1 N HCl). Thus, the amount of oudenone present in the fermentation broth could be estimated by measuring the difference in UV absorbance between the phosphate buffer and HCl solutions of the broth at 246 nm (Figure 2).

4.4 Isolation and purification of oudenone (67)

The fermentation broth (100 ml) was filtered through cheesecloth and then extracted with *n*-butanol (3x 100 ml). Evaporation of the butanol under high vacuum gave a light brown syrup which was redissolved in EtOAc (10 ml) and extracted with H_2O (pH=7.0, 3x 10 ml). The pH of the aqueous layer was adjusted to 3.0 with 1.0 N HCl and extracted with EtOAc (3x 10 ml). The EtOAc layer was dried with anhydrous $MgSO_4$, filtered, and evaporated to dryness to give 10~20 mg of crude oudenone as a yellow oil. Oudenone (~3-5 mg) was isolated after reverse phase flash column chromatography on C18 silica. The column was eluted with a linear gradient from 0.1%AcOH/99.9% H_2O to 0.1% AcOH/49.95%H_2O/50%MeOH. The elution of oudenone from the column was monitored by UV absorbance. All samples of ^{2}H-labelled oudenone (**67**) were further purified by C_{18} reversed-phase HPLC to assure greater than 99% purity before they were analyzed by ^{2}H NMR.

4.5 Synthesis of α-Diketone NAC Thioester 86

2-Propylcyclopentanone (79)

Sodium hydride (2.52 g, 105.6 mmol) was suspended in dry toluene (300 ml) and cooled in an ice bath under N_2. Ethyl-2-oxocyclopentanecarboxylate (77) (15.0g, 96 mmol) dissolved in 50 ml of toluene was added dropwise via a syringe, over a period of 20 min. The reaction mixture was allowed to warm to RT, and was stirred for an additional 15 min. Propyl iodide (32.60 g, 192 mmol) was added, and the mixture was heated at reflux for 48 h. The reaction mixture was then cooled to RT, quenched with H_2O (~1.0 ml), and extracted with 200 ml of H_2O. The aqueous layer was further extracted with EtOAc (2x 100 ml). The organic layers were combined, dried over $MgSO_4$, and concentrated to give compound **78** as a yellow oil (22 g, ~100%). Compound **78** was nearly pure as judged by NMR and TLC. Thus, it was used in the synthesis of **79** without further purification.

TLC [silica, hexanes:EtOAc(10:1)]: R_f=0.22.

^{1}H NMR ($CDCl_3$, 270 MHz) δ: 0.9 (t, J = 7.3Hz, 3H), 1.25 (t, J=7.3Hz, 3H), 1.08-1.42(m, 2H), 1.48-1.63 (dt, 1H), 1.8-2.2 (m, 4H), 2.25-2.58 (m, 3H), 4.15 (q, J=7.3Hz, 2H).

^{13}C NMR ($CDCl_3$, 67.5 MHz) δ: 13.8,14.2, 18.0, 19.4, 32.5, 35.8, 37.7, 60.3, 61.0, 170.8, 214.7.

Concentrated HCl (37%, 100 ml) was added to compound **78** (22 g), and the mixture was allowed to head at reflux for 24 h. The reaction mixture was cooled, extracted with Et_2O (2x100 ml), washed with saturated $NaHCO_3$ (2x10 ml) and brine (100 ml), dried over $MgSO_4$, and concentrated to give **79** as a crude oil. After vacuum distillation, pure **79** was obtained in an 82% overall yield (9.9 g). TLC [silica, hexanes:EtOAc(4:1)]: R_f=0.71.

^{1}H NMR ($CDCl_3$, 270 MHz) δ: 0.8 (t, J = 7.3Hz, 3H), 1.08-2.30 (5m, 11H).

^{13}C NMR ($CDCl_3$, 67.5 MHz) δ: 13.8, 20.52, 20.55, 29.4, 31.7, 37.9, 48.7, 221.5.

3-Propyl-δ-valerolactone (80)

m-Chloroperbenzoic acid (1.7 g, 9.85 mmol) was added to a solution of compound **79** (1.2 g, 9.52 mmol) in CH_2Cl_2 (40 ml). The reaction mixture was allowed to stir at rt for 24 h. The solution was filtered, washed with saturated $NaHCO_3$ (3X30 ml), dried over $MgSO_4$, and concentrated to give a crude oil. After vacuum distillation, lactone **80** was isolated as colorless oil, in 85% yield (1.2 g). TLC [silica, hexanes:EtOAc(1:4)]: R_f=0.22.
^1H NMR ($CDCl_3$, 270 MHz) δ: 0.9 (t, J = 7.3Hz, 3H), 1.3-1.9 (m, 8H) 2.3-2.6 (m, 2H), 4.3(m, 1H)
^{13}C NMR ($CDCl_3$, 67.5 MHz) δ: 13.8, 18.0, 18.2, 27.5, 29.2, 37.7, 80.1, 171.

3-Propyl-β-valerolactol (81)

An anhydrous solution of 3-propyl-β-valerolactone (**80**) (12.51 g, 88.06 mmol) in THF was cooled to -70 °C for ~30 min before a solution of DIBAL-H (176.2 ml of 1M solution in anhydrous THF, 2 eq.) was added over a period of 30 min *via* a dropping funnel and under nitrogen.[120] The reaction mixture was stirred at -70°C for 1 h and then at -20°C for an additional period of 5 h. The reaction was quenched by pouring the solution (slowly) into a mixture of 10.0 g ice and 10 ml of acetic acid. The mixture was then extracted with diethyl ether (3x200 ml). The organic layers were combined, washed with brine (150 ml), saturated aqueous $NaHCO_3$ (3x150 ml), and brine (2x150 ml), dried over anhydrous $MgSO_4$, and the solvents were evaporated to obtain the crude product as an oil. The pure lactol **81** (7.998 g) was isolated in 63% yield after purification by flash column chromatography, using a solvent mixture of hexane-EtOAc (linear gradient from 15:1 to 10:1).

TLC [silica, hexane:EtOAc (4:1)]: $R_f = 0.22$.

^{1}H NMR (mixture of lactol conformer, $CDCl_3$, 270 MHz) δ: 0.80 (t, J = 6.1 Hz, 3H), 1.15-1.48 [m, 10H (several overlapping signals)], 3.29 & 3.85 (2m, 0.67H & 0.36, H3), 4.55 (dd, J = 1.7 Hz, 0.67H, $H1_{ax}$), 5.15 (bs, 0.36H, $H1_{eq}$).

MS [(NH_3) CI, direct inlet 150°C], *m/z* (% relative intensity, assignment): 162 [5, $(M+NH_4)^+$], 144 [9, M^+], 127 [100, $(M+H-H_2O)^+$], 109 [43, $(127-H_2O)^+$].

HRFAB+MS (glycerol/KCl at rt) *m/z*: 145.12290, calcd mass for $C_8H_{16}O_2+H^+$=145.122855

4-Hydroxyoctanal-1-(1,3)-dithiane (82)

To a solution of **81** (7.95 g, 55.2 mmol) in anhydrous CH_2Cl_2 (50 ml), 8.32 ml of 1,3-propanedithiol (82.8 mmol, 1.5 eq.) and 3 ml of BF_3 etherate (3.46 g, 24.4 mmol, 0.44 eq.) were added *via* a syringe. The reaction mixture was allowed to stir at RT under nitrogen for 24 h. The crude mixture was subsequently washed with equal volumes of H_2O, 10% aqueous KOH and again with H_2O. After drying the organic layer over K_2CO_3 and evaporation of the solvent, the dithiane derivative **82** was obtained in 90% yield and in high enough purity to be used as such in the subsequent reaction.

TLC [silica, hexane:EtOAc (4:1)]: $R_f = 0.23$

^{1}H NMR ($CDCl_3$, 270 MHz) δ: 0.8 (t, J = 6.1 Hz, 3H), 1.15-2.03 [m, 12H (several overlapping signals)], 2.64-2.87 (m, 4H), 3.55 (m, 1H), 4.00 (t, J = 6.1 Hz, 1H).

^{13}C NMR ($CDCl_3$, 67.5 MHz) δ: 14.2, 18.9, 22.8, 26.1, 30.5, 35.5, 37.0, 39.7, 47.6, and 71.5.

MS [(NH_3) CI, direct inlet 150°C], *m/z* (% relative intensity, assignment): 252 [9, $(M+NH_4)^+$], 234 [32, M^+], 217 [15, $(M+H-H_2O)^+$], 127 [100, $(M+H-SCH_2CH_2CH_2S)^+$].

HRFAB+MS (glycerol/KCl at rt) *m/z*: 235.11894, calcd mass for $C_{11}H_{22}O_2S_2+H^+$=235.119034

4-*t*-Butyldimethylsilyl ether derivative (83)

A mixture of *t*-butyldimethylsilyl chloride (17.0 g, 112.6 mmol, 4.5 eq.), imidazole (17.1 g, 250.5 mmol, 10 eq.) and dithiane derivative **82** (6.0 g, 50.1 mmol) was stirred in anhydrous DMF (2 ml), under nitrogen at RT for 24 h.[121] The reaction was subsequently quenched with the addition of Et_2O (15 ml) and brine (15 ml). The aqueous layer was further extracted with Et_2O (2x50 ml) and the combined ether layers were dried over $MgSO_4$ and concentrated under reduced pressure to give the crude product **83** as an oil. Pure compound **83** was obtained in quantitative yield (>98%) after purification by flash column chromatography using hexane / EtOAc (25:1) as the eluting solvent.

TLC [silica, hexane:EtOAc (19:1)]: $R_f = 0.66$

^1H NMR ($CDCl_3$, 300 MHz) δ: 0.01 (s, 6H), 0.85 (overlapping s & t, 12H), 1.15-2.03 (overlapping, m, 12H), 2.61-2.87 (m, 4H), 3.55 (m, 1H), 4.00 (t, J = 6.7 Hz, 1H).

^{13}C NMR ($CDCl_3$, 75 MHz) δ: -4.3, -4.2, 14.5, 18.3, 18.7, 22.5, 26.1, 26.2, 30.6, 35.9, 36.9, 39.4, 47.8, 71.9.

MS [(NH_3) CI, direct inlet 30°C], *m/z* (% relative intensity, assignment): 348 [5, M^+], 291 [60, $(M\text{-}tBu)^+$], 127 [100, $(M\text{-}TBDMS\text{-}SCH_2CH_2CH_2S)^+$].

HRFAB+MS (glycerol at rt) *m/z*: 349.20553, calcd mass for $C_{17}H_{36}OS_2Si+H^+$=349.205513

9-(*t*-Butyldimethylsilyloxy)-5-(1,3-dithiane)-4-oxolauric acid (84)

To a cooled (-78 °C) solution of **83** (3.72 g, 10.7 mmol) in anhydrous THF (10 ml), a solution of *n*-butyllithium (1.6 M solution in hexanes, 16.0 mmol, 1.5 eq.) was added, followed by tetramethylethylenediamine (2.4 ml, 1.86 g, 16 mmol, 1.5 eq.).[122] The reaction mixture was allowed to warm to -20°C and stirred at that temperature for ~1 h. An aliquot of the reaction mixture was quenched with D_2O and analyzed by ^1H NMR to confirm complete deprotonation of the starting material before continuing with the addition of the electrophile. Succinic anhydride (1.6 g, 16.0 mmol, 1.5 eq.) was dissolved in anhydrous THF (18 ml) and cooled to -78 °C. The anionic solution of **83** was then cooled to -78 °C before it was transferred through a pre-cooled cannula to the solution of the deuterated succinic anhydride. The mixture was allowed to warm to -20 °C and stirred at that temperature for an additional 5 h. Finally, the reaction was quenched by the addition of acidified water with HCl (50 ml, pH = ~3). The crude product was extracted into EtOAc (3x150 ml), the organic layers were dried over $MgSO_4$ and the EtOAc was evaporated to dryness. The pure lauric acid derivative **84** was isolated after flash column chromatography (2.65 g, 82% yield based on the recovery of unreacted started material) using as the eluent a mixture of hexanes / EtOAc (25:1) containing 0.3% acetic acid.

TLC [silica, hexanes:EtOAc (1:1) plus 0.3% AcOH]: R_f = 0.37

^{1}H NMR ($CDCl_3$, 300 MHz) δ: -0.01 (s, 6H), 0.85 (overlapping s & t, 12H), 1.21-1.55 (m, 8H), 1.70-2.10 (m, 4H), 2.59-2.64 (m, 4H), 2.92-3.07 (m, 4H), 3.55 (m, 1H).

^{13}C NMR ($CDCl_3$, 75 MHz) δ: -4.4, -4.3, 14.2, 18.0, 18.6, 20.1, 24.8, 25.9, 27.8, 28.6, 30.7, 37.3, 39.0, 39.2, 60.8, 71.6, 178.7, 202.7.

MS [(NH_3) CI, direct inlet 300°C], *m/z* (% relative intensity, assignment): 449 [18, $(M+H)^+$], 347 [62, $(M-C_4O_3H_5)^+$], 317 [100, $(M\text{-TBDMS})^+$], 213 [28, $(M\text{-TBDMS-}C_4O_3H_5)^+$].

HRFAB+MS (glycerol at rt) *m/z*: 449.22158, calcd mass for $C_{21}H_{40}O_4S_2Si+H^+$=449.221557

[2-$^2H_2$, 3-$^2H_2$]-9-(*t*-Butyldimethylsilyloxy)-5-(1,3-dithiane)-4-oxolauric acid (84)

The deuterium labeled analog **84** was obtained following the same synthetic procedure as described for the unlabelled compound **84**, with the exception that the anion of **83** was coupled with [2-$^2H_2$, 3-$^2H_2$]succinic anhydride.

^{1}H NMR ($CDCl_3$, 300 MHz) δ: -0.01 (s, 6H), 0.85 (overlapping s & t, 12H), 1.21-1.55 (m, 8H), 1.70-2.10 (m, 4H), 2.66 (dt, J = 3.6 & 1.2 Hz, 2H), 2.97-3.07 (m, 2H), 3.55 (m, 1H).

^{2}H NMR ($CDCl_3$, 46.6 MHz) δ: 3.06 (s), 2.62 (s) ~1:1 ratio.

^{13}C NMR ($CDCl_3$, 75 MHz) δ: -4.3, -4.2, 14.5, 18.3, 18.8, 20.3, 24.9, 26.1, 27.9, (resonance at δ: 28.6 could not be clearly observed, ^{1}H and ^{13}C resonance at C2 and C3 were assigned from the combined ^{1}H, ^{13}C, COSY, HETCOR and DEPT NMR data of the unlabeled and labeled compound.), 30.5 (m), 37.5, 39.1, 39.4, 60.9, 71.8, 178.3, 202.8.

MALDI MS, *m/z* (assignment): 474 [$(M-H+Na)^+$], 453 (M^+), 320 [$(M\text{-OSiMe}_2\text{tBu})^+$].

HRFAB+MS (glycerol at rt) *m/z*: 453.24656, calcd mass for $C_{21}H_{36}O_4S_2SiD_4$+H^+ =453.246664

9-(*t*-Butyldimethylsilyloxy)-5-(1,3-dithiane)-4-oxolauryl NAC Thioester (85)

A solution of **84** (134 mg, 0.3 mmol) in anhydrous CH_2Cl_2 (4 ml) was cooled to 0 °C under nitrogen. 1,3-Dicyclohexylcarbodiimide (1 ml, 86 mg, 0.42 mmol, 1.4 eq.) and 4-dimethylaminopyridine (7.3 g, 0.06 mmol, 0.2 eq., dissolved in 0.5 ml anhydrous CH_2Cl_2) were added, followed by a solution of N-acetylcysteamine (142 mg, 1.2 mmol. 4 eq., dissolved in 7 ml of anhydrous CH_2Cl_2). The reaction mixture was stirred at 0°C for 30 min before it was allowed to warmup to RT and stirred for an additional 2 h. The precipitated urea was subsequently removed by filtration, the solid material was rinsed with EtOAc and the combined organic solvents were evaporated to dryness. The crude residue was chromatographed on silica gel using EtOAc / hexanes (1:1) as the eluent providing pure NAC-derivative **84** (128 mg) in 78% yield.

TLC [silica, EtOAc]: R_f = 0.17

^{1}H NMR ($CDCl_3$, 300 MHz) δ: -0.01 (s, 6H), 0.85 (overlapping s & t, 12H), 1.2-1.5 (m, 8H), 1.7-2.1 (m, 4H), 1.97 (s, 3H), 2.6 (dt, J = 3.6 & 1.2 Hz, 2H), 2.84 (t, J = 5.2 Hz, 2H, H3), 2.92 (d, J = 12.6 Hz, 2H), 3.04 (t, J = 6.0 Hz, 2H, H2), 3.08 (t, J = 5.2 Hz, 2H, H3), 3.38 (q, J = 5.7 Hz, 2H), 3.58 (m, 1H), 5.95 (bs, NH).

^{13}C NMR ($CDCl_3$, 75 MHz) δ: -4.3, -4.2, 14.5, 18.3, 18.8, 20.3, 23.4, 24.9, 26.1, 27.9, 28.8, 31.4 (C2), 37.4, 38.5 (C3), 39.1, 39.4, 39.7, 61.0, 71.8, 170.5, 198.6, 202.7.

MS [(NH_3) CI, direct inlet 300°C], *m/z* (% relative intensity, assignment): 550 [12, $(M+H)^+$], 418 [22, $(M-OSiMe2tBu)^+$], 347 [32, $(M-NAC-COCH_2CH_2CO)^+$], 120 [100, $(NAC)^+$].

HRFAB+MS (glycerol at rt) *m/z*: 550.25174, calcd mass for $C_{25}H_{47}NO_4S_3Si+H^+$=550.251478

[2-$^2H_2$, 3-$^2H_2$] 9-(*t*-Butyldimethylsilyloxy)-5-(1,3-dithiane)-4-oxolauryl NAC Thioester (85)

[2-$^2H_2$, 3-$^2H_2$]-9-(*t*-Butyldimethylsilyloxy)-5-(1,3-dithiane)-4-oxolauric acid (**84**) was converted to the deuterium labeled NAS-derivative **85** following the same synthetic procedure as described for the unlabelled compound **85**.

^{1}H NMR ($CDCl_3$, 300 MHz) δ: -0.01 (s, 6H), 0.85 (overlapping s & t, 12H), 1.2-1.5 (m, 8H), 1.7-2.1 (m, 4H), 1.97 (s, 3H), 2.66 (dt, J = 3.6 & 1.2 Hz, 2H), 2.92 (d, J = 12.6 Hz, 2H), 3.04 (t, J = 6.3 Hz, 2H), 3.38 (q, J = 6.3 Hz, 2H), 3.58 (m, 1H), 5.95 (s, NH).

^{2}H NMR ($CDCl_3$, 46.6 MHz) δ: 2.87 (s), 3.14 (s) ~1:1 ratio.

^{13}C NMR ($CDCl_3$, 75 MHz) δ: -4.3, -4.2, 14.5, 18.3, 18.8, 20.3, 23.4, 24.9, 26.1, 27.9, 28.8, 31.4 (m, C2), 37.4, (resonance at δ38.5 could not be clearly observed, C3), 39.1, 39.4, 39.7, 61.0, 71.8, 170.5, 198.6, 202.7.

HRFAB+MS (glycerol at rt) *m/z*: 554.27674 calcd mass for $C_{25}H_{43}NO_4S_3SiD_4+H^+$=554.276585

α-Diketone NAC Thioester 86 (a & b)

To a solution of compound **85** (360 mg, 0.65 mmol) in CH_3CN / H_2O (9:1 ratio, 10 ml), [bis(trifluoroacetoxy)iodo]benzene (566 mg, 1.3 mmol, 2 eq.) was added and the reaction mixture was stirred at RT for 1.5 h.[123] Most of the solvents were subsequently removed under high vacuum and the crude reaction mixture was mixed with saturated aqueous $NaHCO_3$ (10 ml). After extraction of the aqueous layer with EtOAc (3x100 ml), the organic layer was dried over $MgSO_4$ and evaporated to dryness. Pure α-diketone **86** was obtained after flash column chromatography using 25% hexanes in EtOAc in 40 % yield (~90 mg). ^1H and ^{13}C NMR indicated that the predominant form in solution was the hemiketal **86b**.

TLC [silica, EtOAc]: $R_f = 0.12$

^1H NMR ($CDCl_3$, 300 MHz) δ: 0.85 (t, J = 6.9 Hz, 3H), 1.2-2.0 (overlapping m, 10H), 1.97 (s, 3H), 2.75-3.15 (m, 6H), 3.38 (q, J = 6.3 Hz, 2H), 3.58 (m, 1H), 5.95 (bs, NH).

^{13}C NMR ($CDCl_3$, 75 MHz) δ: 14.3, 18.6, 18.7, 23.4, 28.9, 30.1 (C2), 30.8, 31.2, 37.6 (C3), 38.5, 39.7, 71.0, 96.7, 170.6, 198.9, and 207.3.

MALDI MS, *m/z* (assignment): 368 [$(M+Na)^+$], 328 [$(M-H_2O+H)^+$], 209 [$(328\text{-}NAC)^+$].

HRFAB+MS (glycerol at rt) *m/z*: 346.16867, calcd mass for $C_{16}H_{27}NO_5S+H^+$=346.168820

[2-$^2H_2$, 3-$^2H_2$] β-Diketone NAC Thioester 86

Deprotection of [2-$^2H_2$, 3-$^2H_2$] 9-(*t*-butyldimethylsilyloxy)-5-(1,3-dithiane)-4-oxolauryl NAC thioester (**86**) was carried out using the same procedure as described for the unlabelled compound.

^1H NMR ($CDCl_3$, 300 MHz) δ: 0.85 (t, J = 6.9 Hz, 3H), 1.2-2.0 (overlapping m, 10H), 1.97 (s, 3H), 3.04 (dt, J = 6.3 & 1.5 Hz, 2H), 3.38 (q, J = 6.3 Hz, 2H), 3.58 (m, 1H), 5.95 (bs, NH).

^2H NMR ($CDCl_3$, 46.6 MHz) δ: 2.9 (bs).

^{13}C NMR ($CDCl_3$, 75 MHz) δ: 14.3, 18.6, 18.7, 23.4, 28.9, 30.1 (m, C2), 30.8, 31.2, (resonance at δ 37.6 could not be clearly observed, C3) 38.5, 39.7, 71.0, 96.7, 170.6, 198.9, 207.3.

MS [(NH_3) CI, direct inlet 200°C], *m/z* (% relative intensity, assignment): 371 [2, $(M-H+Na)^+$], 331 [4, $(M-H_2O)^+$], 230 [46, $(M-NAC)^+$], 143 [36, $(230-C_4O_2D_4)^+$], 120 [100, $(NAC+H)^+$].

HRFAB+MS (glycerol at rt) *m/z*: 350.19392, calcd mass for $C_{16}H_{23}NO_5SD_4+H^+=350.19392$

B: Experimental: Studies on the biosynthesis and genetics of verucopeptin

4.6 Materials

4.6.1 Strains, plasmids and enzymes

Strains:

Actinomadura verrucosospora Q886-2 was used as wild type strain to produce verucopeptin;

E.coli XL 1-blue MRF' kan strain (Stratagene) was used as host for construction of the cosmid library of *A.verrucosospora*;

E.coli XL 1-blue was used for cloning (Stratagene);

E.coli S17-1 was used for conjugative plasmid transfer (USDA.ARS, Preoria, Illinois).

Plasmids:

Litmus 28 (New England, Bio-lab), pBluescript Sk^- for cloning and sequencing;

SuperCos I (Stratagene) for constructing cosmid library;

pKC1139 (Lilly Research Laboratory, Indianapolis, Indiana, USA); for conjugation in gene disruption;

pKAO127 containing *ery* gene cluster (Kosan Bioscience Inc. CA, USA).

Enzymes:

Restriction enzymes, Klenow fragment of DNA polymerase I, T4 DNA ligase, and T4 DNA polymerase from New England Bio-labs;

Lysozyme, pronase, and RNAse from Boehringer Mannheim;

Shrimp alkaline phosphatase (SAP) and DNA ladders from Amersham Life Science.

4.6.2 Media and reagents

SeaKem-agarose was obtained from FMC BioProducts, Marine, USA. Bacto Soytone, soluble starch, yeast extract, and Bacto Agar were obtained from Difco Laboratories, Detroit. Kanamycin, Carbenicillin, Apramycin and amino acids were purchased from Sigma (St. Louis, MO). Soy flour was purchased from Natural Foods Inc, Milwaukee, Oregon. Glycerol was obtained from ICN. Nylon Transfer Membranes were purchased from MagnaGraph, MSI, Westboro, MA.

Media:

Seed medium for ***A. verrucosospora*** (g/100ml H_2O): [140a]

peptone (0.5), yeast extract (0.2), sucrose (3.0), KNO_3 (0.2), K_2HPO_4 (0.1), $MgSO_4.7H_2O$ (0.05), KCl (0.05).

Production medium for *A. verrucosospora* (g/100ml $H_{2}0$):[130]

soy flour (1.0), glycerol (2.0) and $CaCO_3$ (0.5). The pH of the medium was adjusted to 7.5 with 0.1N HCl before autoclaving.

SY agar medium for selecting single colony of *A. verrucosospora* and its mutants (g/100 ml $H_{2}0$):
soluble starch (1.0), yeast extract (0.2), agar (1.5).

AS-1 medium[154] for generating transconjugant mutants in gene disruption experiments, (g/100 ml $H_{2}0$):

soluble starch (0.5), yeast extract (0.1), Na_2SO_4(2.0), NaCl (0.25), L-alanine (0.02), L-arginine (0.2) and L-asparagine (0.05). The pH of the medium was adjusted to 7.5.

Luria broth (LB) medium for *E.coli* growth (g/100ml $H_{2}0$): [155]

tryptone (1.0), yeast extract (0.5), and NaCl (1.0)

Luria broth (LB) agar:

LB medium plus agar (1.5g)

SM medium:[55]

LB broth, 10 mM $MgSO_4$, maltose 0.2% (w/v)

Buffer:

Lysis buffer:[156]

15% sucrose, 25 mM Tris-HCl pH 8.0, 25 mM EDTA, pH 8.0

TE buffer: [156]

10.0 mM Tris-HCl at pH 8.0, 1 mM EDTA at pH 8.0

1XSSC buffer: [156]

0.15 M NaCl, 0.015 M sodium citrate, pH 5.8

4.7. Experimental

E.coli was grown at 37°C in Luria broth or on Luria broth agar, with the appropriate antibiotics. The isolation of plasmids and cosmids from *E. coli* was performed by using a Qiagen Midi kit or Qiaprep spin plasmid kit (Qiagen GmbH, Hilden, Germany). DNA cloning, manipulations, colony hybridization, and Southern blots were performed using standard procedures.[155]

4.7.1 Isolation of genomic DNA[157]

A. verrucosospora was cultured on 2 X 500 ml sterilized seed media (500 ml medium in 2800 ml flask) which was inoculated with 5 ml of spore suspension and incubated for 3 days at 28°C. An average of 5.0 g of wet weight mycelia was obtained from the 500 ml culture.

The wet mycelia (10.0 g) were resuspended in 135 ml lysis buffer and incubated with 300 mg lysozyme at 37°C for 40 min, followed by addition of 30 mg of pronase and 15 ml of 10% SDS. The mixture was incubated at 50°C for overnight. The solution was cooled to room temperature and the DNA was extracted three times with an equal volume of phenol-chloroform (1:1). The upper aqueous layer was removed with a wide-bore pipette and treated with 0.6 mg RNAse at 37°C for 30 min and then a solution 15.0 ml of 3M NaOAc (pH 5.2) was added to the mixture solution. The DNA was extracted twice with an equal volume of phenol-chloroform (1:1). The upper aqueous layer was removed with a wide-bore pipette. The DNA was precipitated with 450 ml of 100% ethanol and spooled out using a sealed Pasteur pipette, washed with 75% ethanol, dried in the air, and finally dissolved in 10 ml TE buffer. The concentration of DNA was estimated to be approximately 3.52 mg/ml (ratio on A260 nm/280 nm 1.82).

4.7.2. Construction of cosmid library

Chromosomal DNA (4X25.0 µl DNA) was partially digested with 1U of *Sau*3AI at 37°C for 30 minutes in 500 µl of digestion solution. The DNA was extracted twice with an equal volume of phenol-chloroform (1:1). The upper aqueous layer (2.0 ml) was removed with a wide-bore pipette and 0.2 ml of 3M NaOAc (pH 5.2) was added. The DNA was precipitated with 100% ethanol, washed with 75% ethanol, dried in the air, and dissolved in 0.4 ml TE buffer. An aliquot of 200.0 µl (about 150.0 µg) of partially digested DNA was dephosphorylated with 15 U shrimp alkaline phosphatase (SAP) at 37°C for 1.5 h, and then the enzyme was inactivated at 65°C for 15 min. The DNA was extracted once with phenol-chloroform and once with chloroform. A solution of 3M NaOAc (10 µl) was added and gently mixed. The DNA was precipitated with ethanol, washed with 70% ethanol, and dissolved in 75.0 µl of TE buffer. The concentration was estimated to be 1.0 µg/µl.

Supercos I, a commercially vector with two cohesive end sequences (*cos)* sites and cloning capacity of 32-42 kb, was used to construct a cosmid library. SuperCos I (25 µg) was first digested with 5 U of *Xba*I, dephosphorylated by 5 U of SAP, and then digested with 5 U of *Bam*HI. After treatment with each enzyme each time, the DNA was extracted with phenol-chloroform-isoamyl alcohol (25:24:1) saturated with 50 mM Tris-

HCl (pH 8.0) and then once with chloroform. The DNA was precipitated with ethanol, washed with 70% ethanol, and finally dissolved in 15.0 µl TE buffer. The final concentration was estimated to be 1.0 µg/µl.

The 2.5 µg of insert DNA fragments were then ligated to the 1.0 µg of vector arms using 1 µl (2U) of T4 DNA ligase in final volume of 10.0 µl. The solution (3.0 µl) of ligated DNA was packaged into one Gigapack III packaging extract (Stratagene) *in vitro*. Only those molecules which contain the two *cos* sites 38-52 kb apart and in which the *cos* sites are oriented in the same direction will be encapsulated into λ phage heads. The packaged reaction was diluted with 500 µl of SM and 20 µl of chloroform was added.

The packaged cosmids were transdued into *E.coli* XL1-Bule MRF'kan strain. A saturated culture was grown in SM broth, then pelleted by centrifugation and the cells were resuspended to OD_{600}=1.0 with 10 mM $MgSO_4$. For titering the packaged cosmids, a small aliquot of package solution (2.5 µl) was diluted to 25.0 µl in 10 mM $MgSO_4$, absorbed onto 25 µl of diluted host cells (OD_{600}=0.5) and incubated at room temperature for 30 min. LB broth (200 µl) was added and incubated for 1h at 37°C, shaking the tube every 15.0 min. The pellet was collected using centrifugation at 13,000 rpm for 10 min and resuspended in 50 µl of fresh LB broth. The mixture was spread on an LB agar plate containing carbenicillin (100 µg/ml) and kanamycin (50 µg/ml) and incubated at 37°C overnight. From an aliquot (2.5 µl) package solution, the 636 colonies were obtained, indicating that the cosmid library was estimated at approximately 127,200 colonies (500 µl package solution). A small number of 10 colonies were picked up at random and the cosmids were isolated using a Qiaprep kit. The DNA of each colony was digested with *Pst*I and analyzed using 0.4 % agarose gel. The average insert size was found to be approximately 37.0 kb.

The rest of the package solution containing the chromosomal library was divided into three tubes. Each of them (165.5 µl) was mixed with 170.0 µl of XL1-Blue MRF'kan host cells (OD_{600}=0.5) in a 15.0 ml culture tube and incubated at room temperature for 30 min. LB broth (560.0 µl) was added to each tube and incubated at 37°C for 1 hour with shaking. The cells were harvested by centrifugation at 2,000 rpm for 10 min at room temperature and resuspended in 500 µl of LB broth. Aliquots of 50 µl were spreaded onto 10 LB agar plates containing carbenicillin (100µg/ml) and kanamycin (50 µg/ml) and incubated at 37°C overnight. LB broth (2.0 ml) was poured onto each plate. The colonies were scraped off and pipetted into a sterilized bottle. The plate was washed with an additional 2 ml of LB broth and then all the LB broth was combined. To this bottle, 40 ml of 50% sterile glycerol was added to give a final concentration of 20% and 2.8 g of carbenicillin was added to give a final concentration 50 mg/ml (total 150 ml). An aliquot of culture (100 ml) was kept at -80°C and another 50 ml of culture was kept at 4°C. Appendix VI summarized the construction of the cosmid library.

4.7.3 Probes for *A. verucosospora* PKS and NRPS

The KS1 of DEBS1 (3100-4140,GenBank) was amplified by PCR. The KS oligonucleotides (5' TTT AGA TCT AGG GCT AGC ACC GAG 3' and 5' CGC GGT CGC GCC CGT ACG CCG CG 3') were used as primers in a PCR mixture with the PCR DIG labeled probe synthesis kit (Boehringer Mannheim). The template DNA was plasmid pKAO127. The mixture was cycled 30 times in a GeneAmp-PCR system 2400 (Perkin Elmer) at 95°C for 30 s, 65°C for 1 min, 72°C for 90 s followed by a final incubation at 72°C for 5 min. The reaction generated a 1.1kbp DNA fragment which was purified from a 1% agarose gel using a Qiaquick Gel Extraction Kit (Qiagen GmbH, Hilden, Germany).

A PKS-specific probe and a non-ribosomal peptide synthase-specific probe were generated by PCR amplification of *A. verucosospora* genomic DNA. The degenerate primers used for these PCR reactions were designed from highly conserved amino acid sequences of KS domains and highly conserved adenylation (A7) and thiolation (T) motif in amino acid sequences in the GenBank database.

The sequences of the KS primers were:

XSY1: 5'--GAC ACV GCN TGY TCB TCV-3'

XSY2: 5'--RTG SGC RTT VGT NCC RCT-3'.

The sequences of peptide primers were:

N26A: 5'-CGG ATT TC TAC CGC ACS GGC GAC STC GYC CG -3'

N26B: 5'-CGG GAT CC GAG TGG CCG CCS AGS KYG AAG AA -3'.

These primers were used in a PCR mixture of the PCR DIG Probe Synthesis kit. The mixture was cycled 30 times in PCR system 2400 (Perkin Elmer) at 95°C for 30 s, 55°C for 30s, 72°C for 1 min, and followed by a final incubation at 72°C for 5 min. The probe **KS/V-1** was a 750-bp labeled DNA fragment and **Pep/V** was a 500-bp labeled DNA fragment. The labeled DNA fragments were purified from 1% agarose gel using Qiaquick Gel Extraction Kit (Qiagen GmbH, Hilden, Germany).

4.7.4 Hybridization

4.7.4.1 Colony Hybridization

E.coli XL 1-blue MRF' kan culture (5 µl) with cosmids was diluted to 1000 µl with LB broth and plated to five 150.0 mm plates. After incubation, each plate contained about 10,000 to 15,000 colonies. These colonies were transferred from the 150.0 mm plates to nylon transfer membranes and the membranes were treated by standard methods. The DNA fragments attached to the membranes were crosslinked with UV light using Stratalinker UV2400 crosslinker (Stratagene). Prehybridization at 42^{0}C for 2 hours, hybridization at 42^{0}C for 16 h and detections were carried out following the instructions in the DIG Nucleic Acid Detection Kit (Boehringer Mannheim). About 100 ng of DIG labeled KS1/ery probe was used in the hybridization solution. The membranes were washed twice with 2X SSC-0.1% sodium dodecyl sulfate (SDS) at room temperature for 15 min, and twice in 0.1 X SSC -0.1% SDS at 65^{O}C for 30 min.

A 2.0 mm region surrounding each strong hybridization signal from the 5 master plates was scraped and washed into 60.0 µl of LB broth (total 300 µl). An aliquot (1.0 µl) of this mixture was diluted to 1.0 ml with LB broth and plated onto four 135 mm plates. The rest of the steps were the same as above. A total of 70 colonies exhibiting strong hybridization with KS1/ery probe were picked up and transferred onto a new 100.0 mm plate. After incubation of this plate at 37°C overnight, the 70 colonies were further screened with the **KS/V-1** probe. The steps followed for the hybridization experiment were the same as those described above except that the second wash was carried out in 0.1 X SSC -0.1% SDS at 75°C for 30 min. The positive colonies (53 positive colonies) identified from this experiment were used in the subsequent studies.

4.7.4.2. The second screening with probes KS/V-1and Pep/V

The 53 cosmids isolated from the 53 colonies picked from the above hybridization experiment were digested with *Bam*HI and the DNA fragments were separated on a 0.8% agarose gel and transferred to a nylon transfer membrane. The membrane was treated by the standard methods as described before: prehybridization at 42°C for 2 hours, hybridization at 42°C for 16 hours and detection carried out following the instructions in the DIG Nucleic Acid Detection Kit (Boehringer Mannheim). Approximately 100 ng of the DIG labeled probe (s) **KS/V-1** was (were) used in the hybridization solution. The membranes were then washed twice in 2X SSC-0.1% SDS at room temperature for 15 min and twice in 0.1 X SSC -0.1% SDS at 75°C for 30 min.

From the original 53 KS positive cosmids, 29 cosmids were picked up for further screening with the **Pep/V** probe. Their DNA was digested with *Bam*HI and the hybridization protocol identical to that described above was followed. Approximately

100 ng of DIG labeled **Pep/V** probe was used in the hybridization solution. The nylon membranes were then washed twice in 2X SSC-0.1% SDS at room temperature for 15 min and twice in 0.1 X SSC -0.1% SDS at 65°C for 30 min. Five cosmids containing peptide positive bands were identified. Two cosmids, labeled as pYT27 and pYT31 were finally chosen to study further.

4.7. 5. Restriction map of pYT27 and pYT31

pYT27 and pYT31 were doubly digested with *Bgl*II and *EcoR*I respectively. The digested DNA fragments were cloned into Litmus 28 which was digested with *Bgl*II, *Eco*RI, or *Bgl*II and *Eco*RI respectively to form subclones from S1 to S13 (Figure 9). The restriction map of pYT27 was based on the multiple enzyme digestions (Table 12), digestion of its subclones and the Southern blot results (Appendix I and II). The restriction map of pYT31 was also based on the multiple enzyme digestions (Table 12). In general, the digestion mixtures consisted of 3.0 μl of purified DNA from miniprep (about 0.3 μg), 2.0 μl enzyme(s), 3.0 μl digestion buffer, 3.0 μl BSA and 19.0 μl water to a final volume of 30.0 μl. These digestion reactions were carried out at 37°C for 2h.

Table 12: The enzymes used to determine restriction map of pYT27

	*Bgl*II	*Cla*I	*Eco*RI	*Bam*HI
*Bgl*II	X			
*Cla*I	X	X		
*Eco*RI	X	X	X	
*Bam*HI	X	X	X	X
*Bgl*II/*Cla*I			X	
*Bam*HI/*Eco*RI	X	X		

X: The combination of enzymes in single, double, and triple digestions

4.7. 6 Antibiotic Marker

A. verrucosospora was tested with different antibiotics to determine the antibiotic resistance of this microorganism which was required as a marker for the gene disruption experiments. The 0.5 ml of spores of *A. verrucosospora* were spread onto 30.0 ml SY plates and overlaid with different concentrations of antibiotics in 2 ml water. These plates were incubated in 30^{0}C for two weeks. The results are summarized on the table 13.

Table 13: Antibiotic marker for *A. verrucosospora*

Antibiotic	concentration (μg/ml)	Results	Antibiotic	concentration (μg/ml)	Results
Apramycin	100	S	Lincomycin	100	S
Carbenicillin	100	R	Streptomycin	100	S
Chloramphenicol	25	S	Tetracycline	20	R
Hygromycin	100	S	Thiostrepton	50	S
Kanamycin	50	R			

R: resistant (growth); S: sensitive (no growth at all).

4.7.7. Gene disruption

A 2.1 kb *Bam*H I KS positive fragment from subclone 4 of pYT27 was cloned into the *Bam*HI site of pKC1139 (Figure 10). The recipient cells, wild type *A. verrucosospora,* were prepared by inoculating 25 ml seed media with 1 ml of spore suspension and incubating the culture overnight at 30^{0}C on a rotary shaker. The culture was then centrifuged at 3,700 rpm for 10 min, the cells were washed twice with seed medium, recentrifuged and resuspended in 2.0 ml of seed media. *E.coli* donor cells, S17-1, S17-1/(pKC1139) and S17-1 (pKC1139/27B8) were prepared by growing cultures in 3 ml LB-broth containing apramycin (60 μg/ml) at 37°C on a rotary shaker overnight. The cultures (300.0 µl) were inoculated into 35 ml of fresh LB medium and grown at 37°C to OD_{600} of 0.4-0.6. The cells were pelleted by centrifuging at 3,700 rpm for 10 min and washed twice with LB broth. After a final centrifugation, the pellets were resuspended in 2.0 ml of LB. The donor cell broth (2.0 ml) was mixed with 2.0 ml of spore broth and the mixture was incubated at 30°C for 1 hr. The cells were centrifuged for 10 min at 3,800 rpm, the pellets were resuspended in 0.5 ml of seed medium and the 250 µl of mixture was plated onto an AS-1 plate. After an incubation period of 12 h, the plates were overlaid with 2 ml of sterile deionized Millipore water with 1.5 mg of carbenicillin and 1.0 mg of apramycin. The plates were incubated at 30°C until transconjugates appeared (approximately 3 weeks).

The single transconjugant colony which was formed was selected and subsequently cultured in seed culture with 60 µg/ml of apramycin for four generations. Each generation was 3 days incubation period at 37°C. The culture was diluted, plated on SY plates with 100 µg/ml of apramycin, and incubated at 42°C. A single spore was further purified on a SY plate at 42°C after 4 generations.

4.7.8 Metabolite analysis

Several single spores of wild type and mutant *A. verrucosospora* were plated onto the SY agar plates (with apramycin 100 µg/ml for transconjugants) and incubated at 42°C for 4 days. Two single colonies of wild type and 5 single colonies of the transconjugants were inoculated into 50 ml seed culture (with apramycin 100 µg/ml for transconjugants) and incubated at 37°C for 3 days. An aliquot of 5.0 ml seed culture from each culture was transferred to 50 ml production medium (with apramycin 100 μg/ml for transconjugants) and incubated at 30°C and 37°C for 9 days, respectively. The supernatants from each 30 ml production culture were extracted three times with equal volumes of ethyl acetate. The organic phases were pooled and concentrated. The residues were dissolved in 200 μl of CH_3CN. The presence or absence of verucopeptin in the organic extract was confirmed by LC-MS assay with 50 μl of solution using API 100LC (Perkin-Elmer Sciex Instruments). The solvent gradient was from CH_3CN:H_2O (1:1 ratio) to 100% CH_3CN in 20 min. Within the detection limit of our instrument (1.0 ng from crude extract), verucopeptin could not be detected in the extract from mutant culture.

4.7.9 Analysis of genomic DNA of wild type and mutant

The 20 ml of seed cultures of two randomly picked mutants from five cultures and wild type were used to isolate genomic DNA, respectively. The procedure used was the same as that previously described for the extraction of genomic DNA from wild type culture. The genomic DNA samples were examined from each culture by Southern blot analysis, PCR analysis, and their ability transform *E. coli*; the latter experiment was required in order to confirm the absence of free plasmids in the mutant cells.

4.7.9.1 Southern blot assay for gene disruption

Two probes were used in the Southern blot analysis of the *A. verrucosospora* mutant locus. One was the labeled **KS/V-2** fragment from the PCR product of pYT27B8/pKC1139 using the primers of YXS1 and YXS2. The PCR conditions were the same as those used to make **KS/V-1**. The second probe was the labeled pKC1139 vector. Plasmid pKC1139 (5 μg) was digested with *Bam*HI, purified from a 1% agarose gel, and recovered from the gel using a Qiagen spin column. Denatured DNA was labeled using the DIG DNA labeling kit (Boehringer Mannheim)

The genomic DNA samples from mutants (3.0 μg) and wild type (3.0 μg) were digested with 5U of *Sac*I or *Sac*I/*Bgl*II at 37°C for 2 hrs. DNA fragments were separated in a 0.8% agarose gel. DNA transfer, hybridization, and detection were the same as before.

4.7.9.2. PCR product of apramycin resistance gene.

The 1.2 kb DNA fragment of apramycin resistance gene cluster was amplified using PCR. The primers were:

040-73A: 5' TCT <u>GATATC</u> GCC AGT TAC CTT CGG AAA AAAG-3'

040-73B: 5' AGA <u>AAGCTT</u> ATC CCC GAT CCG CTC CAC GTG-3'.

The underlined sequence of **040-73A** is the *Eco*RV site. The underlined sequence of 040-73B is the *Hind*III site. The PCR mixtures consisted of 2.0 μl of primers (each is 1.0μl of 100 pmol), 25.0 μl of Easy-Start micro 50 (Molecular Bio-Products Inc), 1.0 μl of template DNA, 1.0 μl of *Pfu* enzyme, and 21.0 μl of water. The template DNA used was pKC1139 and genomic DNA samples from mutants and the wild type of *A. verrucosospora*. The mixture was cycled 30 times in a GeneAmp-PCR system 2400 (Perkin Elmer) at 95°C for 1 min, 61°C for 30s, 72°C for 90s, and followed by a final incubation at 72°C for 5 min.

4.8 DNA sequencing and analysis

pYT27 was digested with *Eco*RI and DNA fragments were separated on a 1% agar gel. The DNA fragments of 1.7 kb, 1.6 kb, 7.2 kb and 6.0 kb were cloned into Litmus 28 which was digested with *Eco*RI and they were named as subclone (S) 1, 2, 6, and 7. Cosmid pYT27 was digested with *Eco*RI and *Bgl*II to generate three DNA fragments of 10.5 kb, 9.5 kb and 5.9 kb. These DNA fragments were cloned into Litmus 28 which was digested with *Eco*RI or *Eco*RI and *Bgl*II, respectively, which generated S3, S4 and S5. Subclone 3 (Figure 9) was further doubly digested with *BamHI* and *Eco*RI. The 4.1 kb DNA fragment was cloned into Litmus 28 which was digested with *Bam*HI and *EcoR*I.

As with subclones of pYT27, cosmid pYT31 was digested with *Eco*RI and *Bgl*II and its DNA fragments were separated on a 1% agarose gel. These DNA fragments were cloned into Litmus 28 which was digested with *Eco*RI or *Eco*RI and *Bgl*II to generate S8 to S15. The Subclone 9 (Figure 9) was digested with *Bam*HI. The vector which contained a 0.7 kb DNA fragment was self-ligated to form a small subclone. The 2.5 kb *Bam*HI DNA fragment from S9 was cloned into pBluescript vector which was digested with *Bam*HI.

Both strands of each DNA fragment were sequenced independently using the ABI PRISM Big Dye-Terminator sequencing system with ds DNA templates and run on an Applied Biosystem 373A or 377XL automated sequencer. Sequencing reactions were initially performed with universal forward and reverse primers, followed by specific primers generated from the initial sequence data. Sequence homology searches were performed using the NCBI world wide blast server. The nucleotide sequences and multiple sequence alignments were performed using the GCG and PC/GENE programs (University of Geneva, Switzerland. 1993). All DNA sequencing experiments were performed at the University Core DNA services of University of Calgary, Calgary, AB. Canada.

4.9 **References:**

---

[153] (a) Tsantrizos, Y.S.; Ogilvie, K.K.; Watson, A.K. *Can. J. Chem.* **1992**, *70*, 2276.
(b) Blunt, J.W.; Calder, V.L.; Fenwick, G.D.; Lake, R.J.; McCombs, J.D.; Munro, M.H.G.; Perry, N.B. *J. Nat. Prod.* **1987**, *50*, 290.

[154] Baltz, R.H. *Dev. Ind. Microbiol.* **1980**, *21*, 43

[155] Sambrook, J; Fritsch, E.F.; Maniatis, T. Molecular cloning: a laboratory manual, 2nd ed. **1989**, Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.

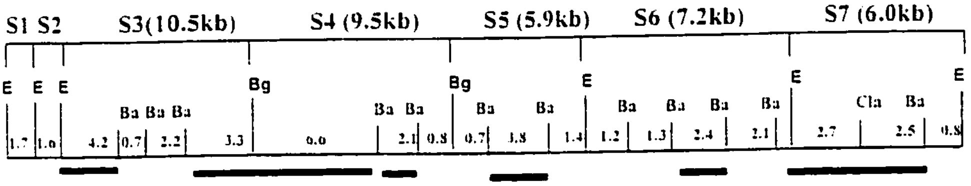
[156] Rao, R.N.; Richardson, Kuhstoss, S. *Methods Enzymolo.* **1987**, 153, 167-198

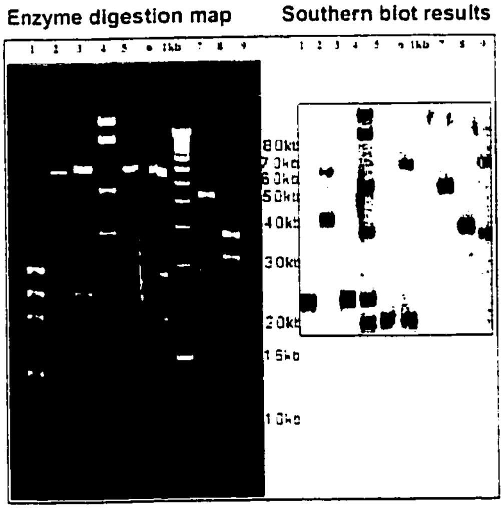
[157] *Instruction manual of SuperCos 1.* **1997**. Stratagene

Appendix I:

Restriction Map of pYT27 based on its subclone digestion with *Eco*RI, *Bam*HI and *Bgl*II.

pYT27 was double digested with *Eco*RI/*Bgl*II. The different fragments were purified from 1.0% agarose gel and cloned into Litmus 28 which was digested with *Eco*RI /*Bgl*II, *Eco*RI or *Bgl*II, respectively. Subsequently, these subclones were digested with different enzymes and checked by Southern blotting experiments as shown in the following figure.





KS positive band: ▬

Lane 1: S6 digested with *Eco*RI/*Bam*HI
Lane 2: S3 digested with *Eco*RI/*Bam*HI
Lane 3: pYT27B6/pKC1139 digested with *Bam*HI
Lane 4: pYT27 digested with *Bam*HI
Lane 5: pYT27B8/pKC1139 digested with *Bam*HI
Lane 6: S4 digested with *Bgl*II/*Bam*HI
Lane 7: S7 digested with *Eco*RI/*Bam*HI
Lane 8: S5 digested with *Eco*RI/*Bam*HI
Lane 9: pYT27B2 digested with *Bgl*II/*Bam*HI

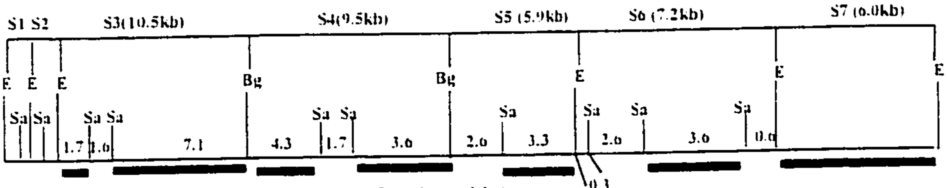
Table 14: Summary of fragments of different subclones: (**KS/V-1** positive bands were shown in bold)

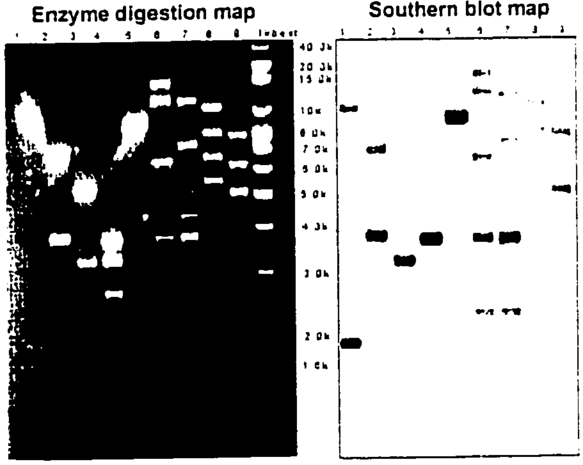
Subclone	Fragments (kb)
S3	0.7, 2.2, **4.2, 6.7** (with 2.8kb of Litmus 28)
S4	0.8, **2.1, 6.6,**
S5	0.7, 1.4, **3.8**
S6	0.3, 1.2, 1.3, 2.1, **2.4**
S7	0.8, **5.7**
pYT27B8/pKC1139	**2.1** (a KS/V-1 positive band from S4 was cloned into pKC1139)
PYT27B6/pKC1139	**2.4** (a KS/V-1 positive band from S6 was cloned into pKC1139)

Appendix II:

Restriction Map of pYT27 based on its subclones digested with *Sa* I

The restriction map of pYT27 digested with *Sac*I has to be done because genomic DNA of wild type and transconjugant mutant digested with the *Sac*I were used in Southern blotting assays. pYT27 and its subclones were digested with *Sac*I and the digested fragments were analyzed by Southern blot with **KS/V-1** as probe. The results follow.





KS positive band: ▬

Lane 1: S3 Digested with *SacI*
Lane 2: S4 Digested with *SacI*
Lane 3: S5 Digested with *SacI*
Lane 4: S6 Digested with *SacI*
Lane 5: S7 Digested with *SacI*
Lane 6: pYT27 Digested with *SacI*
Lane 7: pYT27 Digested with *SacI/BglII*
Lane 8: pYT31 Digested with *SacI*
Lane 9: PYT31 Digested with *SacI/BglII*

Table 15: Summary of fragments of different subclones: (**KS/V-1** positive bands are shown in bold)

Subclone	Fragments (kb)
S3	1.6, **1.7, 10.0** (with 2.8 kb of Litmus)
S4	1.7, **3.6, 7.1** (with 2.8 kb of Litmus)
S5	**3.3,** 5.4 (with 2.8 kb of Litmus)
S6	2.6, **3.6, 3.8** (with 2.8kb of Litmus)
S7	**8.9** (with 2.8kb of Litmus)
pYT27 (*Sac*I digestion)	1.5, 1.6, 1.7, **2.4,** 2.6, **3.6, 3.6, 6.2, 11.4, 13** (with 6.5 kb of Supercos I)
pYT27 (*SacI/Bgl*II digestion)	1.5, 1.6, 1.7, **2.4,** 2.5, 2.6, **3.6, 3.6, 3.6, 4.3,7.1, 9.6** (with 3.6 kb of Supercos I)

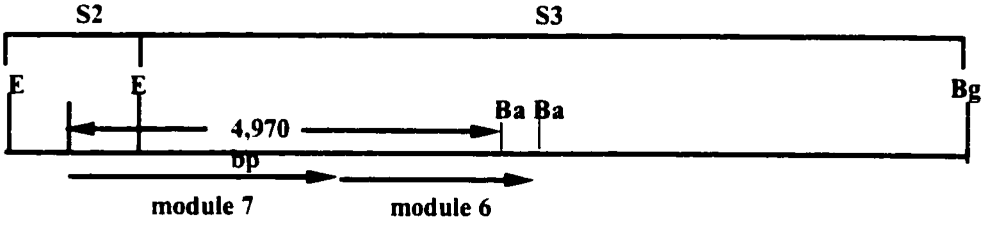
Appendix III: The overlapping region of cosmid pYT27 and pYT31 was confirmed by sequencing subclone S7 from pYT27 and subclone S9 from pYT31 (Figure 9).

```
S7R:    aattccccggccaggctgtccaggacgacgtccattccctggccgtccgtggcggaaagg 60
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
S9F:    aattccccggccaggctgtccaggacgacgtccattccctggccgtccgtggcggaaagg 60

S7R:    aaccgctcctcgaactcgacggtgcgtgacgacgcgatgtgctcgtcgtccagcccgatc 120
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
S9F:    aaccgctcctcgaactcgacggtgcgtgacgacgcgatgtgctcgtcgtccagcccgatc 120

S7R:    tcccgcagcacggcccacttgccggtgctcgccgtcccgaacacctcagcacccagatgc 180
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
S9F:    tcccgcagcacggcccacttgccggtgctcgccgtcccgaacacctcagcacccagatgc 180

S7R:    cgcgccaactgcaccgcggccatccccacaccaccggcggcgccgtgcacgaggaccttc 240
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
S9F:    cgcgccaactgcaccgcggccatccccacaccaccggcggcgccgtgcacgaggaccttc 240

S7R:    tcgcccgcctggagcccagccaggttcacgaggccgtggtacgcagtcgcgaagacgatc 300
        |||||||||||||||||| ||||||||||||||||||||||||||||| |||||||||||
S9F:    tcgcccgcctggagcccanccaggttcacgaggccgtggtacgcagtcncgaagacgatc 300

S7R:     ggcacggacgccgcgtcctcgaacgaccaggtccggggcatcggtacgaccaggcggtga 360
         ||||||| |||||| ||||| ||| |||||||||||||||||||||||||||||||||||
S9F:     ggcacgggcgccgcntcctctaacnaccaggtccggggcatcggtacgaccaggcggtga 360

S7R:     tcggtgaccgcgacgggcccgaacgctccgggcatcatccccatcacgcggtcgcccggc 420
         ||||||||||||||||||||||| ||||||||||||||||||||||||||||| ||||||
S9F:     tcggtgaccgcgacgggcccgaangctccgggcatcatccccatcacgcggtcncccggc 420

S7R:      cgcagcccggtcacgtccgaa 441
          ||||||||||||||||| |||
S9F:      cgcagcccggtcacgtcngaa 441
```

S7R: Universal reverse primer used for the sequence of S7
S9F: Universal forward primer used for the sequence of S9

Appendix IV:

Nucleotide sequence of the 4.9kb fragment from pYT27 (Figure 9) and the deduced amino acid sequences



```
     BamHI
     GGATCCTCATCCACGCCGCGGCGAGCGGCGTCGGGATGGCCGCCGTCCAACTGGCGCAGC
   1 ---------+---------+---------+---------+---------+---------+ 60
     CCTAGGAGTAGGTGCGGCGCCGCTCGCCGCAGCCCTACCGGCGGCAGGTTGACCGCGTCG
   1  I  L  I  H  A  A  A  S  G  V  G  M  A  A  V  Q  L  A  Q  H
      ER6
     ACGCCGGGGCCGAGGTCTACGCGACCGCGAGCCCGGCCAAGTGGCCCGTCCTGCGCGCCC
  61 ---------+---------+---------+---------+---------+---------+ 120
     TGCGGCCCCGGCTCCAGATGCGCTGGCGCTCGGGCCGGTTCACCGGGCAGGACGCGCGGG
  21  A  G  A  E  V  Y  A  T  A  S  P  A  K  W  P  V  L  R  A  L

     TGGGACTGGACGACGAGCACCTGGCGTCGTCGCGGGACCTGGCGTTCGCGGACGAGTTCC
 121 ---------+---------+---------+---------+---------+---------+ 180
     ACCCTGACCTGCTGCTCGTGGACCGCAGCAGCGCCCTGGACCGCAAGCGCCTGCTCAAGG
  41  G  L  D  D  E  H  L  A  S  S  R  D  L  A  F  A  D  E  F  L

     TGGAGGCCACGGGCGGGCGCGGGGTGGACGTCGTCCTCAACTCCCTCGCCCACGCGTTCG
 181 ---------+---------+---------+---------+---------+---------+ 240
     ACCTCCGGTGCCCGCCCGCGCCCCACCTGCAGCAGGAGTTGAGGGAGCGGGTGCGCAAGC
  61  E  A  T  G  G  R  G  V  D  V  V  L  N  S  L  A  H  A  F  V

     TCGACGCCTCCCTCCGGCTGCTCCCCGAGGGCGGGCACTTCCTCGAAATGGGCAAGACCG
 241 ---------+---------+---------+---------+---------+---------+ 300
     AGCTGCGGAGGGAGGCCGACGAGGGGCTCCCGCCCGTGAAGGAGCTTTACCCGTTCTGGC
  81  D  A  S  L  R  L  L  P  E  G  G  H  F  L  E  M  G  K  T  D

     ACATCCGCGATCCCCGGCGTGTGGCGGCCGACCACCCCGGCGTCGACTACCGCGCCTTCG
 301 ---------+---------+---------+---------+---------+---------+ 360
     TGTAGGCGCTAGGGGCCGCACACCGCCGGCTGGTGGGGCCGCAGCTGATGGCGCGGAAGC
 101  I  R  D  P  R  R  V  A  A  D  H  P  G  V  D  Y  R  A  F  D

     ACGTCTACGGATCGGCGGGCCCCGAGCTGCTCCAGGAGATGACCCGCGCGGTCATGGACC
 361 ---------+---------+---------+---------+---------+---------+ 420
     TGCAGATGCCTAGCCGCCCGGGGCTCGACGAGGTCCTCTACTGGGCGCGCCAGTACCTGG
 121  V  Y  G  S  A  G  P  E  L  L  Q  E  M  T  R  A  V  M  D  L

     TGTTCGCCGACGGGAGCCTGCGCCTCAACCCCGTCTCGGTCCGGGACGTCCGCGACGCGC
 421 ---------+---------+---------+---------+---------+---------+ 480
     ACAAGCGGCTGCCCTCGGACGCGGAGTTGGGGCAGAGCCAGGCCCTGCAGGCGCTGCGCG
 141  F  A  D  G  S  L  R  L  N  P  V  S  V  R  D  V  R  D  A  R
```

```
     GCAAGGCGTTCCGCGAGATGAGCCACGGCCGCCACGTCGGCAAGCTGGTCCTCGACGTGG
 481 ---------+---------+---------+---------+---------+---------+ 540
     CGTTCCGCAAGGCGCTCTACTCGGTGCCGGCGGTGCAGCCGTTCGACCAGGAGCTGCACC
 161  K  A  F  R  E  M  S  H  G  R  H  V  G  K  L  V  L  D  V  G

     GCGGCGGCTTCGGCGGCGGGACCGTGCTGGTCACGGGCGGCACCGGGGGAGTGGGCTCGC
 541 ---------+---------+---------+---------+---------+---------+ 600
     CGCCGCCGAAGCCGCCGCCCTGGCACGACCAGTGCCCGCCGTGGCCCCCTCACCCGAGCG
 181  G  G  F  G  G  G  T  V  L  V  T  G  G  T  G  G  V  G  S  L
                      KR6
     TCGTCGCCCGGCACCTCGCCGCCGAGCACGGCGTGCGGAGCCTGGTGCTCGCGAGCCGCC
 601 ---------+---------+---------+---------+---------+---------+ 660
     AGCAGCGGGCCGTGGAGCGGCGGCTCGTGCCGCACGCCTCGGACCACGAGCGCTCGGCGG
 201  V  A  R  H  L  A  A  E  H  G  V  R  S  L  V  L  A  S  R  R

     GGGGGACGGCGGCCGGCGGCGTCCGCGAGCTCGTCGCGGACCTGGAGGCGGCGGGCGCGG
 661 ---------+---------+---------+---------+---------+---------+ 720
     CCCCCTGCCGCCGGCCGCCGCAGGCGCTCGAGCAGCGCCTGGACCTCCGCCGCCCGCGCC
 221  G  T  A  A  G  G  V  R  E  L  V  A  D  L  E  A  A  G  A  A

     CCGTCACGGTCGTGGCCTGCGACGTCGCGGACCGGGCCGCCGTCGCCGACCTGATCGCGG
 721 ---------+---------+---------+---------+---------+---------- 780
     GGCAGTGCCAGCACCGGACGCTGCAGCGCCTGGCCCGGCGGCAGCGGCTGGACTAGCGCC
 241  V  T  V  V  A  C  D  V  A  D  R  A  A  V  A  D  L  I  A  D

     ACATGCCGGCGCGGTATCCGCTGACGGCGGTCGTGCACGCGGCGGGCGCGGTCGCGGACG
 781 ---------+---------+---------+---------+---------+---------+ 840
     TGTACGGCCGCGCCATAGGCGACTGCCGCCAGCACGTGCGCCGCCCGCGCCAGCGCCTGC
 261  M  P  A  R  Y  P  L  T  A  V  V  H  A  A  G  A  V  A  D  G

     GCACGGTCGAGTCGCTGACCGCCGAGGACCTCGACCACGCGCTGCGCGCCAAGGTCGGCG
 841 ---------+---------+---------+---------+---------+---------+ 900
     CGTGCCAGCTCAGCGACTGGCGGCTCCTGGAGCTGGTGCGCGACGCGCGGTTCCAGCCGC
 281  T  V  E  S  L  T  A  E  D  L  D  H  A  L  R  A  K  V  G  G

     GCGCGATCAACCTGCACGAGCTGACCAAGGGCCACGCTCTGTCGGCGTTCATCCTGTTCT
 901 ---------+---------+---------+---------+---------+---------+ 960
     CGCGCTAGTTGGACGTGCTCGACTGGTTCCCGGTGCGAGACAGCCGCAAGTAGGACAAGA
 301  A  I  N  L  H  E  L  T  K  G  H  A  L  S  A  F  I  L  F  S

     CCGCCCTCGCGGGCGCGCTCGGAGCGGGCGGCCAGGCGAACTACGCCGCGGCCAACACCT
 961 ---------+---------+---------+---------+---------+---------- 1020
     GGCGGGAGCGCCCGCGCGAGCCTCGCCCGCCGGTCCGCTTGATGCGGCGCCGGTTGTGGA
 321  A  L  A  G  A  L  G  A  G  G  Q  A  N  Y  A  A  A  N  T  F

     TCCTGGACGGCCTGGCGGCGCGGCGCAGGGCGTCCGGCCTGGCCTGCACCTCGCTGAGCT
1021 ---------+---------+---------+---------+---------+---------+ 1080
     AGGACCTGCCGGACCGCCGCGCCGCGTCCCGCAGGCCGGACCGGACGTGGAGCGACTCGA
 341  L  D  G  L  A  A  R  R  R  A  S  G  L  A  C  T  S  L  S  W

     GGGGCTGGTGGGAGCGGCGGGGCGGCATGACCGCCGGCCTGGACGAGGCCCGCCTCGCAC
1081 ---------+---------+---------+---------+---------+---------+ 1140
     CCCCGACCACCCTCGCCGCCCCGCCGTACTGGCGGCCGGACCTGCTCCGGGCGGAGCGTG
 361  G  W  W  E  R  R  G  G  M  T  A  G  L  D  E  A  R  L  A  R

     GCCTGCGGCGGCTGGGCGCCGTCGAGATGCCCGCGTCCGAGGCGCTGGCCCTGTTCGACG
1141 ---------+---------+---------+---------+---------+---------+ 1200
     CGGACGCCGCCGACCCGCGGCAGCTCTACGGGCGCAGGCTCCGCGACCGGGACAAGCTGC
 381  L  R  R  L  G  A  V  E  M  P  A  S  E  A  L  A  L  F  D  A
```

```
     CGGCGTGCGACCACGGCGGGCCCCTCCTCATCCCGGCCCGGGTGGACCTCGGCGCGCTGC
1201 ---------+---------+---------+---------+---------+---------+ 1260
     GCCGCACGCTGGTGCCGCCCGGGGAGGAGTAGGGCCGGGCCCACCTGGAGCCGCGCGACG
 401   A  C  D  H  G  G  P  L  L  I  P  A  R  V  D  L  G  A  L  R

     GCGGCGCGGCCCGCGGCGAACTGCCCCCCGTCCTGCGCGACCTGGTCCACGGCGGCCGCC
1261 ---------+---------+---------+---------+---------+---------+ 1320
     CGCCGCGCCGGGCGCCGCTTGACGGGGGGCAGGACGCGCTGGACCAGGTGCCGCCGGCGG
 421   G  A  A  R  G  E  L  P  P  V  L  R  D  L  V  H  G  G  R  P

     CGTCCCGCGGCCGGACGGGCCAGACCGGCGAGGGCCCGCGCGGCCTGACCGCCCGGTTGG
1321 ---------+---------+---------+---------+---------+---------+ 1380
     GCAGGGCGCCGGCCTGCCCGGTCTGGCCGCTCCCGGGCGCGCCGGACTGGCGGGCCAACC
 441   S  R  G  R  T  G  Q  T  G  E  G  P  R  G  L  T  A  R  L  A

     CCGCACTGGCGGACGGCGAGGCGGAGGCGGCCCTCCTCGACTGGGTCCGCGAGCAGACCG
1381 ---------+---------+---------+---------+---------+---------+ 1440
     GGCGTGACCGCCTGCCGCTCCGCCTCCGCCGGGAGGAGCTGACCCAGGCGCTCGTCTGGC
 461   A  L  A  D  G  E  A  E  A  A  L  L  D  W  V  R  E  Q  T  A
                                     ACP6
     CCATCGTGCTCGGCCACACCTCCGGCACGAGCGTCGACGCGGACCGGGCGTTCACCCACC
1441 ---------+---------+---------+---------+---------+---------+ 1500
     GGTAGCACGAGCCGGTGTGGAGGCCGTGCTCGCAGCTGCGCCTGGCCCGCAAGTGGGTGG
 481   I  V  L  G  H  T  S  G  T  S  V  D  A  D  R  A  F  T  H  L

     TCGGATTCGACTCGCTGACCTCCGTCGAGCTGTGCAACCGCGTCGCGTCCTCGACCGGGC
1501 ---------+---------+---------+---------+---------+---------+ 1560
     AGCCTAAGCTGAGCGACTGGAGGCAGCTCGACACGTTGGCGCAGCGCAGGAGCTGGCCCG
 501   G  F  D  S  L  T  S  V  E  L  C  N  R  V  A  S  S  T  G  L

     TGCGCCTCCCGTCCACCCTCGTGTTCAGCTACCCGACGCCCCGCGAGCTGGGCCGGCACA
1561 ---------+---------+---------+---------+---------+---------+ 1620
     ACGCGGAGGGCAGGTGGGAGCACAAGTCGATGGGCTGCGGGGCGCTCGACCCGGCCGTGT
 521   R  L  P  S  T  L  V  F  S  Y  P  T  P  R  E  L  G  R  H  I

     TCCTCGGCCTGCTGCGCCCCGCCTCCGCCGCGCCCGGCGCGGGACCGTCCGGGGGCGCGG
1621 ---------+---------+---------+---------+---------+---------+ 1680
     AGGAGCCGGACGACGCGGGGCGGAGGCGGCGCGGGCCGCGCCCTGGCAGGCCCCCGCGCC
 541   L  G  L  L  R  P  A  S  A  A  P  G  A  G  P  S  G  G  A  E

     AGGCCGGCGAGCGGGCACCGGACCCGGACGCGCGCGCGGGCGGACTCGCGGACCTGGACC
1681 ---------+---------+---------+---------+---------+---------+ 1740
     TCCGGCCGCTCGCCCGTGGCCTGGGCCTGCGCGCGCGCCCGCCTGAGCGCCTGGACCTGG
 561   A  G  E  R  A  P  D  P  D  A  R  A  G  G  L  A  D  L  D  L

     TGGAAGGGCTCCTCGACCTGGCGCTGGACGAGAAGGGGAAGTGACATGGACGATGCCGCG
1741 ---------+---------+---------+---------+---------+---------+ 1800
     ACCTTCCCGAGGAGCTGGACCGCGACCTGCTCTTCCCCTTCACTGTACCTGCTACGGCGC
 581   E  G  L  L  D  L  A  L  D  E  K  G  K  *
                                              594  M  D  D  A  A

     GACCGGGCGCCGGCGGACGGCGCGGACCGGGTCGAGCGGGCGCTCCGCGCCCTGCTGGAG
1801 ---------+---------+---------+---------+---------+---------+ 1860
     CTGGCCCGCGGCCGCCTGCCGCGCCTGGCCCAGCTCGCCCGCGAGGCGCGGGACGACCTC
 599 D  R  A  P  A  D  G  A  D  R  V  E  R  A  L  R  A  L  L  E

     GAGCGGGACCGCCTCCGGCGGAGAACGGACGACCTGAAGGCCGGGCGCGGCGAGCCGATC
1861 ---------+---------+---------+---------+---------+---------+ 1920
     CTCGCCCTGGCGGAGGCCGCCTCTTGCCTGCTGGACTTCCGGCCCGCGCCGCTCGGCTAG
 619 E  R  D  R  L  R  R  R  T  D  D  L  K  A  G  R  G  E  P  I
                                                          KS7
```

```
     GCCGTGGTGGCGATGGCGTGCCGCTTCCCCGGCGGCGTGTCGACGCCCGAGGGGCTGTGG
1921 ---------+---------+---------+---------+---------+---------+ 1980
     CGGCACCACCGCTACCGCACGGCGAAGGGGCCGCCGCACAGCTGCGGGCTCCCCGACACC
639  A  V  V  A  M  A  C  R  F  P  G  G  V  S  T  P  E  G  L  W

     GAGGCGGTCCGCGGCGGCCGGGACCTGGTGGAGGAGTTCCCCCGGGACCGGGGGTGGCGG
1981 ---------+---------+---------+---------+---------+---------+ 2040
     CTCCGCCAGGCGCCGCCGGCCCTGGACCACCTCCTCAAGGGGGCCCTGGCCCCCACCGCC
659  E  A  V  R  G  G  R  D  L  V  E  E  F  P  R  D  R  G  W  R

     GACGTGTTCGACGAGGACCCCGGCACGCTCGGCACCTCCAGCACCCGGCACGGGGGCTTC
2041 ---------+---------+---------+---------+---------+---------+ 2100
     CTGCACAAGCTGCTCCTGGGGCCGTGCGAGCCGTGGAGGTCGTGGGCCGTGCCCCCGAAG
679  D  V  F  D  E  D  P  G  T  L  G  T  S  S  T  R  H  G  G  F

     CTTACCGGCGTGGCGGGCTTCGACGCCGAGTTCTTCGGCATCAGCCCGCGGGAGGCGCTG
2101 ---------+---------+---------+---------+---------+---------+ 2160
     GAATGGCCGCACCGCCCGAAGCTGCGGCTCAAGAAGCCGTAGTCGGGCGCCCTCCGCGAC
699  L  T  G  V  A  G  F  D  A  E  F  F  G  I  S  P  R  E  A  L

     GCGATGGACCCGCAGCAGCGCCTGCTGCTGGAGGCGTCGTGGGAGGTCTTCGAACGGGCG
2161 ---------+---------+---------+---------+---------+---------+ 2220
     CGCTACCTGGGCGTCGTCGCGGACGACGACCTCCGCAGCACCCTCCAGAAGCTTGCCCGC
719  A  M  D  P  Q  Q  R  L  L  L  E  A  S  W  E  V  F  E  R  A

     GGCATCGTCCCCGTCGACGTGCGCGGCACCGACGCCGGCGTCTTCTCCGGGGTGAGCTCC
2221 ---------+---------+---------+---------+---------+---------+ 2280
     CCGTAGCAGGGGCAGCTGCACGCGCCGTGGCTGCGGCCGCAGAAGAGGCCCCACTCGAGG
739  G  I  V  P  V  D  V  R  G  T  D  A  G  V  F  S  G  V  S  S

     CCCGAGTACGGGCTGCGCTTCCTGGAGGTGGACCAGCACGAACTGGAGGGCCACCTGTCG
2281 ---------+---------+---------+---------+---------+---------+ 2340
     GGGCTCATGCCCGACGCGAAGGACCTCCACCTGGTCGTGCTTGACCTCCCGGTGGACAGC
759  P  E  Y  G  L  R  F  L  E  V  D  Q  H  E  L  E  G  H  L  S

     CTCGGGAGCGCGCTGAGCGTCGCGTCCGGCCGGGTCGCGTACCAGTTGGGGCTGACCGGC
2341 ---------+---------+---------+---------+---------+---------+ 2400
     GAGCCCTCGCGCGACTCGCAGCGCAGGCCGGCCCAGCGCATGGTCAACCCCGACTGGCCG
779  L  G  S  A  L  S  V  A  S  G  R  V  A  Y  Q  L  G  L  T  G

     CCTGCGATGACCGTGGACACGGCGTGCTCGTCGTCGCTGGTGGCCCTGCACCTGGCGGTG
2401 ---------+---------+---------+---------+---------+---------+ 2460
     GGACGCTACTGGCACCTGTGCCGCACGAGCAGCAGCGACCACCGGGACGTGGACCGCCAC
799  P  A  M  T  V  D  T  A  C  S  S  S  L  V  A  L  H  L  A  V

     CGGTCGCTGCGGTCCGGGGAGTGCTCGCTCGCGCTGGCGGGCGGGGCGGCGGTGATGACC
2461 ---------+---------+---------+---------+---------+---------+ 2520
     GCCAGCGACGCCAGGCCCCTCACGAGCGAGCGCGACCGCCCGCCCCGCCGCCACTACTGG
819  R  S  L  R  S  G  E  C  S  L  A  L  A  G  G  A  A  V  M  T

     ACCCCCGCCGTCTTCGTGGAGTTCTCCCGGCAGCGCGGCCTGGCGCCCGACGGCCGGTGC
2521 ---------+---------+---------+---------+---------+---------+ 2580
     TGGGGGCGGCAGAAGCACCTCAAGAGGGCCGTCGCGCCGGACCGCGGGCTGCCGGCCACG
839  T  P  A  V  F  V  E  F  S  R  Q  R  G  L  A  P  D  G  R  C

     AAGGCGTTCGCGGACGCCGCCGACGGGACCGGCTGGTCCGAGGGCGTCGGCGTGCTGCTG
2581 ---------+---------+---------+---------+---------+---------+ 2640
     TTCCGCAAGCGCCTGCGGCGGCTGCCCTGGCCGACCAGGCTCCCGCAGCCGCACGACGAC
859  K  A  F  A  D  A  A  D  G  T  G  W  S  E  G  V  G  V  L  L
```

```
     CTGGAGCGGCTGTCGGACGCCCAACGCAACGGCCACCCGGTACTGGCCGTCCTGCGCGGA
2641 ---------+---------+---------+---------+---------+---------+ 2700
     GACCTCGCCGACAGCCTGCGGGTTGCGTTGCCGGTGGGCCATGACCGGCAGGACGCGCCT
 879 L  E  R  L  S  D  A  Q  R  N  G  H  P  V  L  A  V  L  R  G

     TCGGCGGTGAACCAGGACGGGGCCAGCAACGGGCTGACCGCGCCCAGCGGCCGTGCCCAG
2701 ---------+---------+---------+---------+---------+---------+ 2760
     AGCCGCCACTTGGTCCTGCCCCGGTCGTTGCCCGACTGGCGCGGGTCGCCGGCACGGGTC
 899 S  A  V  N  Q  D  G  A  S  N  G  L  T  A  P  S  G  R  A  Q

     GAGGAAGTGATCCGCAAGGCGCTGGCGGACGCGGGGATCGCGGCGTCCGAGGTGGACCTG
2761 ---------+---------+---------+---------+---------+---------+ 2820
     CTCCTTCACTAGGCGTTCCGCGACCGCCTGCGCCCCTAGCGCCGCAGGCTCCACCTGGAC
 919 E  E  V  I  R  K  A  L  A  D  A  G  I  A  A  S  E  V  D  L

     GTCGAGGCCCACGGCACCGGAACCGTGCTGGGCGACCCCATCGAGGCCCACGCTCTGCTG
2821 ---------+---------+---------+---------+---------+---------+ 2880
     CAGCTCCGGGTGCCGTGGCCTTGGCACGACCCGCTGGGGTAGCTCCGGGTGCGAGACGAC
 939 V  E  A  H  G  T  G  T  V  L  G  D  P  I  E  A  H  A  L  L

     ATGACGTATGGCCAGGGGCGGGCCGAGGACAGGCCGCTCTGGCTGGGGTCCCTGAAGTCG
2881 ---------+---------+---------+---------+---------+---------+ 2940
     TACTGCATACCGGTCCCCGCCCGGCTCCTGTCCGGCGAGACCGACCCCAGGGACTTCAGC
 959 M  T  Y  G  Q  G  R  A  E  D  R  P  L  W  L  G  S  L  K  S

     AACCTTGGCCACGCCCAGGCCGCCGCCGGGGTCGCCGGTGTCATCAAGGCCGTGATGGCT
2941 ---------+---------+---------+---------+---------+---------+ 3000
     TTGGAACCGGTGCGGGTCCGGCGGCGGCCCCAGCGGCCACAGTAGTTCCGGCACTACCGA
 979 N  L  G  H  A  Q  A  A  A  G  V  A  G  V  I  K  A  V  M  A

     CTGCGCCACGGCTACCTGCCGAAGACCCTGCACCTGGACGCGCCTTCGCGGCACGTGGAC
3001 ---------+---------+---------+---------+---------+---------+ 3060
     GACGCGGTGCCGATGGACGGCTTCTGGGACGTGGACCTGCGCGGAAGCGCCGTGCACCTG
 999 L  R  H  G  Y  L  P  K  T  L  H  L  D  A  P  S  R  H  V  D

     TGGTCCTGCGGCGCGGTGGAACTGCTCGTGGACGGGCGCGAATGGGAGCGGGGCGACGGT
3061 ---------+---------+---------+---------+---------+---------+ 3120
     ACCAGGACGCCGCGCCACCTTGACGAGCACCTGCCCGCGCTTACCCTCGCCCCGCTGCCA
1019 W  S  C  G  A  V  E  L  L  V  D  G  R  E  W  E  R  G  D  G

     CCGCGCAGAGCCGGGGTGTCGTCGTTCGGGTTCAGCGGCACCAACGCGCACGTGATCCTG
3121 ---------+---------+---------+---------+---------+---------+ 3180
     GGCGCGTCTCGGCCCCACAGCAGCAAGCCCAAGTCGCCGTGGTTGCGCGTGCACTAGGAC
1039 P  R  R  A  G  V  S  S  F  G  F  S  G  T  N  A  H  V  I  L

     GAAGAGGCGCCGCGTAAGGGCGCGGCGGCCGAGCCCATCCCGGGCCCGGCCCTCGCCGGT
3181 ---------+---------+---------+---------+---------+---------+ 3240
     CTTCTCCGCGGCGCATTCCCGCGCCGCCGGCTCGGGTAGGGCCCGGGCCGGGAGCGGCCA
1059 E  E  A  P  R  K  G  A  A  A  E  P  I  P  G  P  A  L  A  G

     GGACTGGTGCCGTGGGCGGTGTCGGGAAAGAACCCGGCGGCGCTGGCGCGGCAGGCAGCC
3241 ---------+---------+---------+---------+---------+---------+ 3300
     CCTGACCACGGCACCCGCCACAGCCCTTTCTTGGGCCGCCGCGACCGCGCCGTCCGTCGG
1079 G  L  V  P  W  A  V  S  G  K  N  P  A  A  L  A  R  Q  A  A

     GCGCTCCGCGATTTCGTCACCACGTCGGTGGGTGCGCAGGCGGACGCCGCGGACATCGGC
3301 ---------+---------+---------+---------+---------+---------+ 3360
     CGCGAGGCGCTAAAGCAGTGGTGCAGCCACCCACGCGTCCGCCTGCGGCGCCTGTAGCCG
1099 A  L  R  D  F  V  T  T  S  V  G  A  Q  A  D  A  A  D  I  G
```

```
     TGGTCGCTGGTGTCGAGCCGGTCGGCGTTCGCCCATCGCGCGGTCGTGCTGGGCCGCGAT
3361 ---------+---------+---------+---------+---------+---------+ 3420
     ACCAGCGACCACAGCTCGGCCAGCCGCAAGCGGGTAGCGCGCCAGCACGACCCGGCGCTA
1119 W  S  L  V  S  S  R  S  A  F  A  H  R  A  V  V  L  G  R  D

     CGCGACGAGCTGCTGAGCGGCCTGGCCTCGCTGAGCGAGGGCGCGGAGCCCGCCGGGGTG
3421 ---------+---------+---------+---------+---------+---------+ 3480
     GCGCTGCTCGACGACTCGCCGGACCGGAGCGACTCGCTCCCGCGCCTCGGGCGGCCCCAC
1139 R  D  E  L  L  S  G  L  A  S  L  S  E  G  A  E  P  A  G  V

     GTGCGCGGTGTCGCCGGGGAACTCGGGGGAGTCGTCTTCATGTTCCCGGGCCAGGGCTCG
3481 ---------+---------+---------+---------+---------+---------+ 3540
     CACGCGCCACAGCGGCCCCTTGAGCCCCCTCAGCAGAAGTACAAGGGCCCGGTCCCGAGC
1159 V  R  G  V  A  G  E  L  G  G  V  V  F  M  F  P  G  Q  G  S
                                   AT7
     CAGTGGGCCGGTATGGGCCGGCGGCTGTACGACGCCTACCCCGTCTTCGCGCACGCCCTC
3541 ---------+---------+---------+---------+---------+---------+ 3600
     GTCACCCGGCCATACCCGGCCGCCGACATGCTGCGGATGGGGCAGAAGCGCGTGCGGGAG
1179 Q  W  A  G  M  G  R  R  L  Y  D  A  Y  P  V  F  A  H  A  L

     GACGAGTGCGCGGCCGCGCTCGACGAATGGGCCGGCTGGTCGCTGCTCGACGTGATCCGG
3601 ---------+---------+---------+---------+---------+---------+ 3660
     CTGCTCACGCGCCGGCGCGAGCTGCTTACCCGGCCGACCAGCGACGAGCTGCACTAGGCC
1199 D  E  C  A  A  A  L  D  E  W  A  G  W  S  L  L  D  V  I  R

     GGCGAGGCCGGGGCGCCGCCCCTCGAACGGGTGGACGTGGTCCAGCCCGCGCTGTTCTCG
3661 ---------+---------+---------+---------+---------+---------+ 3720
     CCGCTCCGGCCCCGCGGCGGGGAGCTTGCCCACCTGCACCAGGTCGGGCGCGACAAGAGC
1219 G  E  A  G  A  P  P  L  E  R  V  D  V  V  Q  P  A  L  F  S

     GTGATGGTGTCCATGGCCGCTCTGTGGCGGTCATGGGGTGTGGAACCGGCCGCGGTGGTG
3721 ---------+---------+---------+---------+---------+---------+ 3780
     CACTACCACAGGTACCGGCGAGACACCGCCAGTACCCCACACCTTGGCCGGCGCCACCAC
1239 V  M  V  S  M  A  A  L  W  R  S  W  G  V  E  P  A  A  V  V

     GGGCACAGCCAAGGCGAGATCGCCGCCGCCCACGTCTGCGGCGCGCTGTCGCTGCGCGAC
3781 ---------+---------+---------+---------+---------+---------+ 3840
     CCCGTGTCGGTTCCGCTCTAGCGGCGGCGGGTGCAGACGCCGCGCGACAGCGACGCGCTG
1259 G  H  S  Q  G  E  I  A  A  A  H  V  C  G  A  L  S  L  R  D

     GCCGCCAAGGTGGTCGCGCTCCGCAGCAGGGCGCTGGTGGACCTCGCAGGCCACGGCGGG
3841 ---------+---------+---------+---------+---------+---------+ 3900
     CGGCGGTTCCACCAGCGCGAGGCGTCGTCCCGCGACCACCTGGAGCGTCCGGTGCCGCCC
1279 A  A  K  V  V  A  L  R  S  R  A  L  V  D  L  A  G  H  G  G

     ATGGCGTCGGTCGCGGAGTCCGCGGACGCCGTGGCGGAGCAGCTCGCCCCGTGGAGCGGC
3901 ---------+---------+---------+---------+---------+---------+ 3960
     TACCGCAGCCAGCGCCTCAGGCGCCTGCGGCACCGCCTCGTCGAGCGGGGCACCTCGCCG
1299 M  A  S  V  A  E  S  A  D  A  V  A  E  Q  L  A  P  W  S  G

     CGGCTGGGCATCGCCGCCGTCAACGGGCCCCGCTCGGTGGTGGTGTCCGGAGCGCTGGAC
3961 ---------+---------+---------+---------+---------+---------+ 4020
     GCCGACCCGTAGCGGCGGCAGTTGCCCGGGGCGAGCCACCACCACAGGCCTCGCGACCTG
1319 R  L  G  I  A  A  V  N  G  P  R  S  V  V  V  S  G  A  L  D
              EcoRI
     GCACTCGACGAATTCCTTGAGAAGATGAACGCCGAAGGCGTCCGGGCCAGGCGGATATCG
4021 ---------+---------+---------+---------+---------+---------+ 4080
     CGTGAGCTGCTTAAGGAACTCTTCTACTTGCGGCTTCCGCAGGCCCGGTCCGCCTATAGC
1339 A  L  D  E  F  L  E  K  M  N  A  E  G  V  R  A  R  R  I  S
```

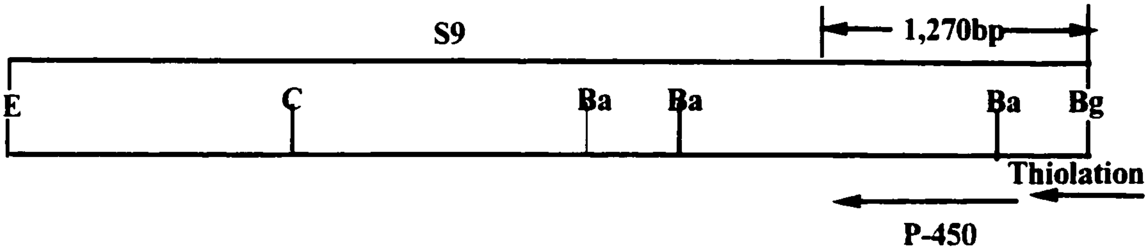
```
     GTCGATTACGCCTCCCATTCCCACGAGGTGGCCCGGGTTCGCGAACAGGTGATCGGGCCG
4081 ---------+---------+---------+---------+---------+---------+ 4140
     CAGCTAATGCGGAGGGTAAGGGTGCTCCACCGGGCCCAAGCGCTTGTCCACTAGCCCGGC
1359 V  D  Y  A  S  H  S  H  E  V  A  R  V  R  E  Q  V  I  G  P

     CTGTCCGATGTGGGGCCCAGAACGTCGGAACTGCCTTTCTATTCGACGCTGTACGGTGAG
4141 ---------+---------+---------+---------+---------+---------+ 4200
     GACAGGCTACACCCCGGGTCTTGCAGCCTTGACGGAAAGATAAGCTGCGACATGCCACTC
1379 L  S  D  V  G  P  R  T  S  E  L  P  F  Y  S  T  L  Y  G  E

     GTCATCGACACCGCCGAGCTGGACGGCGACTATTGGTACACCAACCTCCGTGAAAAGGTG
4201 ---------+---------+---------+---------+---------+---------+ 4260
     CAGTAGCTGTGGCGGCTCGACCTGCCGCTGATAACCATGTGGTTGGAGGCACTTTTCCAC
1399 V  I  D  T  A  E  L  D  G  D  Y  W  Y  T  N  L  R  E  K  V

     CTCTTCGAATCGTCCGTCCGCAGGCTGGCCGACGACGGTTACCGCGTCTTCATCGAGATG
4261 ---------+---------+---------+---------+---------+---------+ 4320
     GAGAAGCTTAGCAGGCAGGCGTCCGACCGGCTGCTGCCAATGGCGCAGAAGTAGCTCTAC
1419 L  F  E  S  S  V  R  R  L  A  D  D  G  Y  R  V  F  I  E  M

     AGCCCGCACCCCGTCCTGACCGTTCCGGTGCAGGAAATGGTCGAGGCGGACGGTGCGGTC
4321 ---------+---------+---------+---------+---------+---------+ 4380
     TCGGGCGTGGGGCAGGACTGGCAAGGCCACGTCCTTTACCAGCTCCGCCTGCCACGCCAG
1439 S  P  H  P  V  L  T  V  P  V  Q  E  M  V  E  A  D  G  A  V

     GTCCTGTCGTCCTCCCGCCGCGACCGCGGGGAGGTCGAGGCCATGGCCGACTCGCTGGCC
4381 ---------+---------+---------+---------+---------+---------+ 4440
     CAGGACAGCAGGAGGGCGGCGCTGGCGCCCCTCCAGCTCCGGTACCGGCTGAGCGACCGG
1459 V  L  S  S  S  R  R  D  R  G  E  V  E  A  M  A  D  S  L  A

     CAGCTGTACGTGCGGGGCGGAACCGTGGACTGGGCCGCCCTGTTCGGGACGCGCAGGCGG
4441 ---------+---------+---------+---------+---------+---------+ 4500
     GTCGACATGCACGCCCCGCCTTGGCACCTGACCCGGCGGGACAAGCCCTGCGCGTCCGCC
1479 Q  L  Y  V  R  G  G  T  V  D  W  A  A  L  F  G  T  R  R  R

     ACGGACCTCCCGACCTACGCGTTCCAGAGGCGGCGCTACTGGCTGAACTCCTCGCACACG
4501 ---------+---------+---------+---------+---------+---------+ 4560
     TGCCTGGAGGGCTGGATGCGCAAGGTGTCCGCCGCGATGACCGACTTGAGGAGCGTGTGC
1499 T  D  L  P  T  Y  A  F  Q  R  R  R  Y  W  L  N  S  S  H  T

     GCGGTGGCGATGGAGCTCCCCGCCGACGCGCCGGCCGAAGAGGACGAGCAGGTCACCCCG
4561 ---------+---------+---------+---------+---------+---------+ 4620
     CGCCACCGCTACCTCGAGGGGCGGCTGCGCGGCCGGCTTCTCCTGCTCGTCCAGTGGGGC
1519 A  V  A  M  E  L  P  A  D  A  P  A  E  E  D  E  Q  V  T  P

     CTGCCCGAACGGGTCGCGGCCCTCCCCGAGGGCGACGCGGAGGCGCTCGTCCTGGACCAC
4621 ---------+---------+---------+---------+---------+---------+ 4680
     GACGGGCTTGCCCAGCGCCGGGAGGGGCTCCCGCTGCGCCTCCGCGAGCAGGACCTGGTG
1539 L  P  E  R  V  A  A  L  P  E  G  D  A  E  A  L  V  L  D  H
                                                     ACP7
     GTCCTGGAGAAGATCGCCACCGTCCTCGGGCATGCCCCCGGCGAGAGCGTCGACCCCGAC
4681 ---------+---------+---------+---------+---------+---------+ 4740
     CAGGACCTCTTCTAGCGGTGGCAGGAGCCCGTACGGGGGCCGCTCTCGCAGCTGGGGCTG
1559 V  L  E  K  I  A  T  V  L  G  H  A  P  G  E  S  V  D  P  D

     CGGGAGTTCAAGGACCTCGGCTTCGACTCGCTGCTGGCGATGGACCTGAGCAGGCGCCTG
4741 ---------+---------+---------+---------+---------+---------+ 4800
     GCCCTCAAGTTCCTGGAGCCGAAGCTGAGCGACGACCGCTACCTGGACTCGTCCGCGGAC
1579 R  E  F  K  D  L  G  F  D  S  L  L  A  M  D  L  S  R  R  L
```

```
     GCCGCCTCGACGGGGGTCAGGCTGAAGCCGACGCTGGTGCTGCGGCACCCGTCACCGCGG
4801 ---------+---------+---------+---------+---------+---------+ 4860
     CGGCGGAGCTGCCCCCAGTCCGACTTCGGCTGCGACCACGACGCCGTGGGCAGTGGCGCC
1599 A  A  S  T  G  V  R  L  K  P  T  L  V  L  R  H  P  S  P  R

     CTGATCGCCGGGCACATCGCGGCCGTGATGCTCGACCGCCGTCCCGCCTGACCCCGCGGT
4861 ---------+---------+---------+---------+---------+---------+ 4920
     GACTAGCGGCCCGTGTAGCGCCGGCACTACGAGCTGGCGGCAGGGCGGACTGGGGCGCCA
1619 L  I  A  G  H  I  A  A  V  M  L  D  R  R  P  A  *  P  R  G

     GTCCCGCCGGCGGGACACCGCGCGCCCGGCGCCGGCATCGTGTTGACCGG
4921 ---------+---------+---------+---------+---------+ 4970
     CAGGGCGGCCGCCCTGTGGCGCGCGGGCCGCGGCCGTAGCACAACTGGCC
1640 V  P  P  A  G  H  R  A  P  G  A  G  I  V  L  T
```

Note: the approximate locations of the catalytic domains are indicated by underlining the amino acid sequences and designating its identity at beginning. The stop codon of the incomplete ORF is TGA at 1782. The start codon and stop codon of Module 7 are ATG at 1786 and TGA at 4909 respectively. The RBS site for Module 7 was GGAAG at 1777. The sequences for *Bam*HI and *Eco*RI were labeled as italic, below the enzyme respectively.

Appendix V:

Nucleotide sequence of the 1.2kb fragment from pYT31 and the deduced amino acid sequences.



```
      BglII
      AGATCTCGGCGCTGTGGGCGGAGCACCTCGACGGCCGCGACGTCGGCCCGGACGACGACT
   1  ---------+---------+---------+---------+---------+---------+  60
      TCTAGAGCCGCGACACCCGCCTCGTGGAGCTGCCGGCGCTGCAGCCGGGCCTGCTGCTGA
   1   I  S  A  L  W  A  E  H  L  D  G  R  D  V  G  P  D  D  D  F
       Thiolation
      TCTTCGCGCTCGGCGGCAACTCGCTGATCGGAATCAAGATCATCGAGCGGGTGGGCCGGG
  61  ---------+---------+---------+---------+---------+---------+  120
      AGAAGCGCGAGCCGCCGTTGAGCGACTAGCCTTAGTTCTAGTAGCTCGCCCACCCGGCCC
  21   F  A  L  G  G  N  S  L  I  G  I  K  I  I  E  R  V  G  R  E

      AGTACGGCGTGGAGCTGTCCGTGCGCGACTTCTACCTGGCGCAGACGCCGGCGCGGGTCG
 121  ---------+---------+---------+---------+---------+---------+  180
      TCATGCCGCACCTCGACAGGCACGCGCTGAAGATGGACCGCGTCTGCGGCCGCGCCCAGC
  41   Y  G  V  E  L  S  V  R  D  F  Y  L  A  Q  T  P  A  R  V  A

      CCGGGCTGGTCGAGCGGGGGAGGGCACGGACGTGAGGCTGACGCCGGGCCCGGCGCGCGA
 181  ---------+---------+---------+---------+---------+---------+  240
      GGCCCGACCAGCTCGCCCCCTCCCGTGCCTGCACTCCGACTGCGGCCCGGGCCGCGCGCT
  61   G  L  V  E  R  G  R  A  R  T  *

      CGTCGACCTCGACCGCGTCGACCTGTTCGACCTGGACCTGTACTCGGCGGGCGACCCGCA
 241  ---------+---------+---------+---------+---------+---------+  300
      GCAGCTGGAGCTGGCGCAGCTGGACAAGCTGGACCTGGACATGAGCCGCCCGCTGGGCGT


      CCCGATCTGGGACGTGATGCGGGCGAGGGCGCCCCTGCACCACCAGGTGCTGCCGGACGG
 301  ---------+---------+---------+---------+---------+---------+  360
      GGGCTAGACCCTGCACTACGCCCGCTCCCGCGGGGACGTGGTGGTCCACGACGGCCTGCC
  71                     M  R  A  R  A  P  L  H  H  Q  V  L  P  D  G
                         P450
      CCGGGAGTTCTGGTCGGTCACCCGCTACGACGACGTCTGCCGCGTCCTCGGCGAGCACCA
 361  ---------+---------+---------+---------+---------+---------+  420
      GGCCCTCAAGACCAGCCAGTGGGCGATGCTGCTGCAGACGGCGCAGGAGCCGCTCGTGGT
  86  R  E  F  W  S  V  T  R  Y  D  D  V  C  R  V  L  G  E  H  Q

      GCGGTTCACCTCGGAGCGCGGGACGGTGGTCACCCACCTCGGGGTGGACGACGTCGCGGC
 421  ---------+---------+---------+---------+---------+---------+  480
      CGCCAAGTGGAGCCTCGCGCCCTGCCACCAGTGGGTGGAGCCCCACCTGCTGCAGCGCCG
 106  R  F  T  S  E  R  G  T  V  V  T  H  L  G  V  D  D  V  A  A

      GGGCACGCTGCTGACCTCCACCGACCCGCCGCGGCACACCCTCGTCCGCCGGGCCATCGG
 481  ---------+---------+---------+---------+---------+---------+  540
      CCCGTGCGACGACTGGAGGTGGCTGGGCGGCGCCGTGTGGGAGCAGGCGGCCCGGTAGCC
 126  G  T  L  L  T  S  T  D  P  P  R  H  T  L  V  R  R  A  I  G
```

```
                                                 BamHI
     CGCCCGGCTCACGGCCCGCGCGGTCGCGCCGTGGCGGGAGCGGATCCCCGAGCGGGACTG
 541 ---------+---------+---------+---------+---------+---------+ 600
     GCGGGCCGAGTGCCGGGCGCGCCAGCGCGGCACCGCCCTCGCCTAGGGGCTCGCCCTGAC
 146  A  R  L  T  A  R  A  V  A  P  W  R  E  R  I  P  E  R  D  W

     GGACGAGCTCGTCCAGCTGACCGCGATGGTCACGGCCCCCTCGGACCCGCACTTCCGGCA
 601 ---------+---------+---------+---------+---------+---------+ 660
     CCTGCTCGAGCAGGTCGACTGGCGCTACCAGTGCCGGGGGAGCCTGGGCGTGAAGGCCGT
 166  D  E  L  V  Q  L  T  A  M  V  T  A  P  S  D  P  H  F  R  H

     CGGCAGCGAGGCCGCGACCCTGGCGATCGCCCACCACGAACTCGTCACCTACGTCAAGGA
 661 ---------+---------+---------+---------+---------+---------+ 720
     GCCGTCGCTCCGGCGCTGGGACCGCTAGCGGGTGGTGCTTGAGCAGTGGATGCAGTTCCT
 186  G  S  E  A  A  T  L  A  I  A  H  H  E  L  V  T  Y  V  K  E

     ATGGGCCGCCCGGCGCAGATCGGCGGGCGGCGACGACGGCAGCCTGCTCGACCACCTCAT
 721 ---------+---------+---------+---------+---------+---------+ 780
     TACCCGGCGGGCCGCGTCTAGCCGCCCGCCGCTGCTGCCGTCGGACGAGCTGGTGGAGTA
 206  W  A  A  R  R  R  S  A  G  G  D  D  G  S  L  L  D  H  L  M

     GACCGTCCGGGTGGCGGGCGCGCCGCTCACCGACGAGGAGATCGCCCTCGACGGCTACAG
 781 ---------+---------+---------+---------+---------+---------+ 840
     CTGGCAGGCCCACCGCCCGCGCGGCGAGTGGCTGCTCCTCTAGCGGGAGCTGCCGATGTC
 226  T  V  R  V  A  G  A  P  L  T  D  E  E  I  A  L  D  G  Y  S

     CATCCTCCTCGGCGCCAACGTGACGACGCCGCACACCGTCTCGGGCACGGTGCTCGCCCT
 841 ---------+---------+---------+---------+---------+---------+ 900
     GTAGGAGGAGCCGCGGTTGCACTGCTGCGGCGTGTGGCAGAGCCCGTGCCACGAGCGGGA
 246  I  L  L  G  A  N  V  T  T  P  H  T  V  S  G  T  V  L  A  L
                               •
     CATCGAGCGGCCCGAGCAGTTCGGGAAGGTCCAGGCCGACCCGTCGCTGGTCCCGAACCT
 901 ---------+---------+---------+---------+---------+---------+ 960
     GTAGCTCGCCGGGCTCGTCAAGCCCTTCCAGGTCCGGCTGGGCAGCGACCAGGGCTTGGA
 266  I  E  R  P  E  Q  F  G  K  V  Q  A  D  P  S  L  V  P  N  L

     GGTGGAGGAGGGGCTGCGCTGGACGTCGGCCGCCTGCAACTTCATGCGGTACGCGGTGGA
 961 ---------+---------+---------+---------+---------+---------+ 1020
     CCACCTCCTCCCCGACGCGACCTGCAGCCGGCGGACGTTGAAGTACGCCATGCGCCACCT
 286  V  E  E  G  L  R  W  T  S  A  A  C  N  F  M  R  Y  A  V  D

     CGACGTCCGCATCGCCGGCGGGACGATCCCGGCCCGCGGCGCGGTCGTCGCATGGATCGG
1021 ---------+---------+---------+---------+---------+---------+ 1080
     GCTGCAGGCGTAGCGGCCGCCCTGCTAGGGCCGGGCGCCGCGCCAGCAGCGTACCTAGCC
 306  D  V  R  I  A  G  G  T  I  P  A  R  G  A  V  V  A  W  I  G

     CTCGGCCAACCGCGACGAGTCCCAGTTCGCCGACCCGCACACGTTCGACGTCACGCGGAA
1081 ---------+---------+---------+---------+---------+---------+ 1140
     GAGCCGGTTGGCGCTGCTCAGGGTCAAGCGGCTGGGCGTGTGCAAGCTGCAGTGCGCCTT
 326  S  A  N  R  D  E  S  Q  F  A  D  P  H  T  F  D  V  T  R  N

     CGCGAGCCGCCAGGTCGCATTCGGGTACGGGCCGCACTACTGCGTCGGGGCGCCGCTGGC
1141 ---------+---------+---------+---------+---------+---------+ 1200
     GCGCTCGGCGGTCCAGCGTAAGCCCATGCCCGGCGTGATGACGCAGCCCCGCGGCGACCG
 346  A  S  R  Q  V  A  F  G  Y  G  P  H  Y  C  V  G  A  P  L  A
                                            *
     CCGGCTGACCCTCCGCGTCTTCTTCAAGGAGCTCCTCCGGCGCTTCGGGTCATTGAGCTC
1201 ---------+---------+---------+---------+---------+---------+ 1260
     GGCCGACTGGGAGGCGCAGAAGAAGTTCCTCGAGGAGGCCGCGAAGCCCAGTAACTCGAG
 366  R  L  T  L  R  V  F  F  K  E  L  L  R  R  F  G  S  L  S  S

     GGGGGGGAGC
1261 ---------+
```

```
    CCCCCCCTCG
386  G  G  S
```

Note: The stop codon of the incomplete ORF containing the thiolation domain is TGA at 213. The start codon of the other incomplete ORF containing P450 is ATG at 318 and the RBS for this ORF is GGACG at 310. Active-site residues are marked with asterisks. Boldfaced amino acid are conserved residues or motifs important for function. The sequences for *Bam*HI and *Bgl*II were labeled as italic below the enzyme respectively.

Appendix VI:

Construction of cosmid library of *A. verrucosospora*