GENETIC STUDIES ON A SOIL *STREPTOMYCES* SP. THAT PRODUCES AN ANTIFUNGAL COMPOUND

NACHAMMA SOCKALINGAM

BSc (Hons), NUS

NATIONAL UNIVERSITY OF SINGAPORE
2002

GENETIC STUDIES ON A SOIL *STREPTOMYCES* SP. THAT PRODUCES AN ANTIFUNGAL COMPOUND

NACHAMMA SOCKALINGAM

BSc (Hons), NUS

A THESIS SUBMITTED
FOR THE DEGREE OF MASTER OF SCIENCE
DEPARTMENT OF MICROBIOLOGY
NATIONAL UNIVERSITY OF SINGAPORE
2002

ACKNOWLEDGEMENTS

I would like to thank my supervisors A/P Nga Been Hen and A/P Vincent Chow Tak Wong for their supervision, guidance and motivation

My thanks to all the faculty members of Department of Microbiology

My heartfelt gratitude to Dr. Fiona Flett and Dr. Colin Smith of UMIST for their kind gift of the *E.coli* strainET12567

I would also like to thank my family, with a special mention of Vignes and Ramesh for their endless support.

I would also like to thank my wonderful friends who have been there to discuss science,life and for fun, just about everything else. Special Thanks to Baskar,Dhira,Karen,Kokila, Kahmeng and Sunita.

INTRODUCTION

LITERATURE REVIEW

MATERIALS AND METHODS

RESULTS

DISCUSSION

REFERENCES

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ABBREVIATIONS

ACP	Acyl carrier protein
Ap^R	Apramycin resistance
AT	Acyl transferase
bp	Base-pair(s)
BSA	Bovine serum albumin
CIP	Calf intestinal phosphate
CoA	Coenzyme A
°C	Degree Celsius
DEBS	Deoxyerythronolide B synthase
DH	Dehydratase
DNA	Deoxyrinonucleic acid
ECL	Enhanced Chemiluminescence
ER	Enoyl reductase
ery	Erythromycin biosynthetic gene
FAS	Fatty acid synthase
g	Gram(s)
h	Hour(s)
kb	Kilobases
KR	Ketoreductase
KS	Ketosynthase
l	Litre(s)
ml	Millilitre(s)
M	Molarity
min	Minute(s)

mol	Mole(s)
OD	Optical density
ORF	Open reading frame
PKS I	Polyketide synthase I
PKS II	Polyketide synthase II
RNA	Ribonucleic Acid
RNAaseA	RibonucleaseA
rDNA	DNA of Ribosomal RNA
rpm	Revolutions per minute
s	Second(s)
SDS	Sodium dodecyl sulfate
TAE	Tris-acetae/EDTA
TE	Thioesterase
TLC	Thin layer chromatography
U	Units of enzyme activity
UV	Ultraviolet
V	Volt(s)
v/v	Volume/Volume
w/v	Weight/Volume

SUMMARY

In an effort to identify novel antifungal compounds, soil isolates from different parts of Singapore were screened. One such soil isolate named 98- 62, identified as a *Streptomyces* sp. based on 16S rDNA sequence analysis, was shown to produce antifungal compound that inhibited *Aspergillus niger* on primary screening. Thin layer chromatography separation of the antifungal compound compared to Rf values of complex polyketides rapamycin and FK506. Complex polyketides are molecules that are synthesized by large multifunctional enzymes called modular polyketide synthases (PKS I) via repeated condensation of carboxylic acids.

Genes encoding the polyketide synthase I (PKS I) enzymes in the genomic DNA of the soil isolate 98- 62 were identified with PKS I specific eryKSII probe of *Saccaropolyspora erythraea*. Degenerate primers based on conserved sequences of PKS I genes were used to amplify a KS–AT genes from the genomic DNA of the soil isolate 98- 62. This 850 bp DNA fragment was subsequently used as a probe to identify a 7-8kb *BamHI* fragment of the genomic DNA of the soil isolate 98- 62 to contain the smaller fragment. The larger fragment was then cloned from a subgenomic library by PCR screening. By chromosomal walking, three contiguous clones of a total length of 11.6kb of DNA were identified. Analysis of the 11.6 kb DNA sequence revealed the presence of two partial open reading frames encoding one complete module and two partial modules. The enzymatic motifs identified within each module occur in the order as has been reported for other known modular PKS modules of actinomycete strains. Comparison of the sequence of the cloned fragments with that of information from the database revealed that the genes contained therein were highly similar to other known PKS I genes.

To determine if the cloned PKS I genes were involved in the synthesis of antifungal compound, gene disruption of specific genes of the cloned PKS genes was carried out. Disruption of the internal modules of the PKS coding region in the soil isolate 98-62 eliminated the synthesis of the antifungal compound, demonstrating that the cloned genes are essentially involved in the biosynthesis of this compound. Disruption study has also established that the 11.6 kb sequence is of two different open reading frames (ORF) as the disruption of a contiguous gene fragment of both the ORFs in the soil isolate did not affect its ability to produce the antifungal compound.

Surprisingly, in addition to disrupting the antifungal compound synthesis, gene disruption of the internal fragments of the PKS I genes of the soil isolate 98- 62 also eliminated its ability to produce aerial mycelium, giving rise to phenotypically bald mutants. As far as we are aware, this is the first report of a case in which the PKS type I genes are involved in the morphological differentiation of *Streptomyces*.

In conclusion, this work has

1) confirmed that the soil isolate 98- 62, which produces a novel antifungal compound is of *Streptomyces* species.

2) identified and partially characterised a PKS I gene cluster from the soil isolate 98- 62.

3) provided functional evidence that the cloned PKS I genes from the soil isolate 98- 62 are involved in the synthesis of a novel antifungal compound.

4) demonstrated the involvement of PKS I genes in morphological differentiation of the strain.

Further work on identifying and sequencing the remaining genes of the complete polyketide synthase gene cluster will provide a better understanding of the organization of the gene cluster. Combined information from such genetic work and chemical analysis of the antifungal compound using NMR and mass spectroscopy would allow for elucidation of the chemical structure of the antifungal compound

produced by the soil isolate 98- 62. Structural information on the nature of chemical compound would assist in an understanding of the mode of action of the antifungal compound.

INTRODUCTION

Molecular genetics of antibiotic production is currently one of the most exciting and challenging areas of research on antimicrobials. Dramatic developments in gene technologies in the last decade have made it possible to clone antibiotic biosynthetic genes of an organism, which in turn has led to remarkable insights into their structure, organization, regulation and evolution of the biosynthetic genes. These studies have paved the way for radically new approaches such as engineering the enzymes to produce novel hybrid antibiotics.

Classical gene technologies such as obtaining defective mutants that do not synthesise or that overproduce antibiotics have played an important role in antibiotic production. These approaches have been used to define the biosynthetic pathway or to increase the antibiotic yields in industrial strains. However, with the invent of new methodologies and technologies, molecular tools are so advanced that the entire genome of an organism can be sequenced, let alone the antibiotic gene cluster. The current trend in understanding antibiotic production is to clone, sequence and express antibiotic genes in widening our knowledge on antibiotic production.

Several strategies are available for cloning antibiotic biosynthetic genes. They include,

1) complementation of blocked mutants,
2) search for homologous genes,
3) reverse cloning,
4) expression of genes in a heterologous host and
5) genome sequencing.

Sequencing of the cloned genes and analysis allow the understanding of the organization and evolution of the genes. Disruption or replacement of an antibiotic specific gene in vivo is the frequently used rigorous way of analysing its function in

the producing organism. As such, establishment of methodologies to transfer genes to allow disruption or replacement is therefore indispensable in the study of antibiotic biosynthetic genes.

The scope of this project is to study the genes responsible for the biosynthesis of an antifungal compound, produced by the soil isolate 98- 62. This would require

1) identification of the soil isolate 98- 62 to allow for a rational approach in establishing gene transfer methodologies specific for this organism,

2) identification of the type of antifungal compound it produces through the use of gene specific probes,

3) cloning of the genes based on homology,

4) chromosomal walking to obtain more genes of the antibiotic gene cluster,

5) sequencing and analysis of the cloned genes

6) establishment of gene disruption method for the soil isolate 98- 62 and finally

7) gene disruption to determine the function of the cloned genes in the antifungal compound synthesis.

For a more indepth understanding of the idea behind and approach to this project, the literature review section of this thesis is included herein.

2.1 ANTIBIOTICS

Antibiotics are defined as low molecular weight microbial secondary metabolites that inhibit the growth of other microorganisms at low concentration. A molecule with defined chemical structure having a relative mass of at most a few thousand is considered to be of low molecular weight. As such, enzymes such as lysozyme and complex proteins such as colchicine are not considered as antibiotics, although they are antibacterial.

Although by the given definition, only substances produced as natural products are considered as antibiotics, products obtained by chemical modification of microbial metabolites are also accepted as antibiotics and are called as semisynthetic antibiotics. Natural products from plants with antimicrobial activity are also sometimes referred to as antibiotic products from plants.

The key word " at low concentration" in the definition is to be highlighted as even essential and normal cellular components can be detrimental and cause damage if present at excessive concentrations. For example, glycine, one of the constituents of every protein has a strong bactericidal effect on some bacteria when present in the culture medium in a high concentration.

Inhibition of growth of other microorganism may be permanent or temporary. When inhibition is lost once the antibiotic is removed from its medium, the antibiotic is said to have a static action. If however inhibition is permanent, the antibiotic is said to have a cidal action. Antibiotics are frequently grouped according to the spectrum of activity. That is according to the classes of microorganisms they inhibit. There are, therefore, antiviral, antibacterial, antifungal and antiprotozoal antibiotics.

Another scheme of classification is based on the chemical structure of the compound. Currently, natural or semisynthetic antibiotics that share a basic chemical

structure are grouped into one "class" and named after the member first discovered or after a principal chemical property. Antibiotics can be therefore classified as β-lactams, tetracyclines, aminoglycosides, macrolides, ansamycins, peptide antibiotics and glycopeptide antibiotics based on their chemical structure. β- lactams, tetracyclines, aminoglycosides, macrolides and ansamycins fall under the group of compounds called polyketides, based on the chemical nature of these compounds. Although various classification schemes of antibiotics have been proposed, there is no one universally adopted scheme to date.

2.2 ANTIFUNGAL COMPOUNDS

2.2.1 NEED FOR ANTIFUNGAL AGENTS

Human and animal fungal infections pose serious medical and veterinary issues, whereas fungal infections of the plants result in significant losses of agricultural products. According to Bodey and Anaissi (1989), there has been a dramatic increase in the frequency of fungal infections, especially disseminated systemic mycoses in immunodeficient hosts in the last three decades. Antineoplastic chemotherapy, organ transplants, congenital defects, leukemia, Hodgkin's disease, and AIDS may cause immune deficiencies. These render an immunocompromised host more susceptible to a variety of fungal, bacterial, protozoal and viral diseases. Species of *Candida, Coccidioides, Histoplasma,* and *Aspergillus* are important causative agents. Of these, *Candida* species, especially *albicans* are clearly the most important causative agents (Holmberg & Mayer, 1986). Candidiasis has a wide range of clinical presentations, ranging from cutaneous to disseminated systemic infections, which include thrush, bronchitis, meningitis, septicaemia, asthma, gastritis and endocarditis.

2.2.2 EXISTING ANTIFUNGAL COMPOUNDS

Amphotericin B has been the choice of antifungal drug for 30 years (Medoff *et al*., 1983; Bodey, 1988). However amphotericin is toxic to human cells and has many side effects, which include renal dysfunction, fever, chills, hypotension and even cardiac failure. The mode of action of amphotericin is to complex with the membrane sterols, resulting in membrane distortion and leakage of intracellular contents.

Other clinically used antifungal drugs are nystatin which also complexes with ergosterol in fungal plasma membrane and imidazoles and triazoles which inhibit ergosterol biosynthesis in the fungi. 5- Fluorocytosine acts by inhibiting DNA and RNA synthesis. Griseofluvin interferes with microtubule formation. Nikkomycin is a peptidyl nucleoside, which is a chitin synthase inhibitor

One of the fundamental requirements for effective antimicrobial therapy is to inhibit the pathogen without affecting the infected host. This can be achieved by targeting a molecular process of the pathogen that is lacking or sufficiently different from the host mammalian cells, so that the host metabolism will be minimally affected. In the case of fungal and mammalian cells, both are eukaryotic and therefore share a great deal of enzymatic and biochemical machinery. This is one of the reasons for the obvious lag in the development of antifungal compounds compared to antibacterial compounds.

Thus, even though there is an extensive list of available antifungal compounds, new antifungal compounds that are more effective, less toxic and showing broader activity are still required.

2.2.3 SEARCH FOR NOVEL ANTIFUNGAL COMPOUNDS

Some of the traditional approaches in finding novel secondary metabolites include

1) screening of microrganisms that produce new, structurally and functionally different antibiotics,
2) mutation of microorganisms to produce new activities,
3) directed biosynthesis by biochemical modification of structures synthesised chemically,
4) chemical or biochemical modifications of a backbone molecule produced by a microorganism,
5) chemical synthesis of new compounds using structures produced in nature as templates for enhanced or more desired activities and
6) fusion of protoplasts of two microorganisms, each producing a desired trait, followed by selection for recombinants, which have desired traits (Strohl *et al.*, 1991).

In screening for microrganisms that produce new, structurally and functionally different antibiotics, microbial screens are first set up to evaluate a compound, or a mixture of compounds (secondary metabolites) on a " target". The aim of the screen is to act as a filter to narrow down to a small number of potential antimicrobial compound producers from a large number. The screen can be for microorganisms that produce antifungals, antibacterials or others.

In searching for novel secondary metabolites that is antifungal, the target used in the microbial screen can be an intact fungal pathogen in vitro or in vivo, or an indispensable enzyme activity or process.

Historically the main source of antimicrobial compounds has been from soil microorganisms. However, new sources of microorganisms, for example, marine invertebrates, plants, halophiles, thermophiles, bacteria are receiving increasing

attention. There is a wide spread belief that new sources of materials will bring new drugs. Correspondingly, there have been extensive programs to isolate microorganisms from exotic environments (de Souza *et al.,* 1982).

Antimicrobial screens of soil samples from diverse and untapped geographical location would also be one approach to identify new antimicrobial compounds. Asia represents one of the many regions in the world where the pool of natural diversity is untapped. Southeast Asia is well known for its species rich tropical rainforests (Bull *et al*., 1992; Myers, 1988). In Singapore, high actinomycete diversity is found in the tropical rainforest at both genus and subgenus levels, which could represent an excellent source for the discovery of novel bioactive compounds (Wang *et al.*, 1999). A total of 35 genera were isolated from primary and secondary rainforests of Singapore, compared to 29 genera in the whole of Yunnan province of China, an area known as the " Kingdom of plants and Animals" (Xu *et al*., 1996; Jiang & Xu, 1996).

2.3 ANTIBIOTICS PRODUCING ORGANISMS

2.3.1 ACTINOMYCETES: GROWTH AND NUTRIENT REQUIREMENTS

Most antibiotics are products of the secondary metabolism of three main groups of microorganisms: eubacteria, actinomycetes and filamentous fungi. The actinomycetes produce the largest number and greatest variety of antibiotics (Waksman, 1950). The actinomycetes comprise a group of branching unicellular gram-positive bacterial organisms, with DNA rich in Guanine and Cytosine (70%). They are widespread in nature, occurring typically in soil, composts, and aquatic habitats. Most species are free-living and saprotrophic, but some may form symbiotic associations, whilst others are pathogenic in man, animals and plants.

The growth of actinomycetes is filamentous. Their growth on a solid or liquid medium results in the formation of a mass of growth usually designated as "colony".

This is a mass of branching filaments that originated from a spore or from a bit of vegetative mycelium. The vegetative growth of the actinomycetes, or stroma is usually shiny, gel like, or lichnoid in appearance and varies in shape, size and thickness. Actinomycetes reproduce either by fission or by means of special conidia.

The actinomycetes are often characterised by the production of a variety of pigments, both on organic and on synthetic media. The variation of colour depends upon many factors, such as the nature and age of the culture. Acids and alkalis are known to have a marked effect upon the nature and integrity of the pigment produced (Waksman, 1950). The colour of the pigment produced varies from strain to strain. Some may be whitish or cream coloured, others may appear yellow, red, pink, orange, green, violet or brown.

The actinomycetes vary greatly in their nutritional requirements. They are able to utilise a great variety of simple and complex organic compounds as sources of carbon and energy. These compounds include organic acids, sugars, starches, hemicelluloses, celluloses, proteins, polypeptides, amino acids, nitrogen base and others. Certain actinomycetes can also utilise, to a more limited extent, fats, hydrocarbons, benzene ring compounds, and even more resistant substances, such as lignin, tannin and rubber.

2.3.2 ACTINOMYCETES: CLASSIFICATION

Many systems of classifying the actinomycetes have been suggested. Traditionally, classification of the actinomycetes has been based upon the morphological and physiological characteristics of the organism. Useful morphological characters for this purpose include the types of mycelium (substrate/aerial), the stability of this mycelium, the mode of division of hyphae; types, number and the

arrangement of spores; formation of flagellate elements and their mobility etc. However, phenotypic characteristics vary with growth conditions and have not been precise enough for distinguishing superficially similar organisms or for determining phylogenetic relationships among the actinomycetes. Physiological tests too have been unreliable as they give variable or unstable data, varying considerably with the growth conditions of microorganisms.

The development and application of new and reliable biochemical, chemical and molecular biology techniques are revolutionizing actinomycete systematics (Goodfellow, 1986). Chemotaxanomy is the study of chemical variation in living organisms and the use of selected chemical characters in classification and identification of organisms (Goodfellow & Minnikin, 1985). In chemotaxanomy, chemical information such as types of peptidoglycan, phospholipids, cell wall sugar and fatty acids are analysed.

Actinomycete taxonomists are well accustomed to " wall types", introduced by Lehevalier & Lechevalier, 1970. This particular chemotaxanomic marker has played an important role in the establishment of actinomycete taxa (Stackebrandt, 1986). This simple analysis of the composition of walls allowed actinomycetes and related organisms to be classified into nine groups of chemotypes based on the cell walls amino acid and sugar composition. Fatty acid composition of microorganism is also an important taxonomic character (Goodfellow & Minnikin, 1985). It has been demonstrated that fatty acid profiles can be analysed quantitatively (Drucker, 1974; Saddler *et al*., 1987) to provide useful taxonomic information at species and in some cases, subspecies level (O' Donnell, A.G., 1985). However, it is important that the environmental factors influencing the chemical composition of microorganisms grown in the laboratory are carefully controlled. It was found that different growth media

gave fatty acid profiles that were both qualitatively and quantitatively different (Farshtchi & Mc Clung, 1970).

Rapid accumulation in the knowledge of molecular biology and the recent advancement of nucleic acid analyses techniques such as the determination of G + C ratio, DNA-DNA hybridisation and 16S rDNA sequencing have provided an important alternative in differentiating the strains of a particular species and allowed the investigations of the evolution of the actinomycetes.

In general, the G + C content of the DNA of the actinomycetes is high. The M*ycobacteria* and *Nocardia* are on the low side of this spectrum (60-70%) while streptomycetes are on the high side (70-75%). DNA-DNA hybridisation was only used to study the species level relationships within a few actinomycete groups. But these studies made little impact on the understanding of higher-level phylogeny among actinomycetes.

The primary structure of rDNA is more conserved than the primary structure of the whole genome. The analyses of the 16/23S rDNAs have made the determination of moderate to even more remote relationships possible. 16S rRNA gene sequence based analyses have been used to resolve phylogenetic relationships between organisms at virtually all taxonomic levels (Stackenbrandt, 1985). Currently, 16S rDNA sequencing has been used to identify culturable as well as non-culturable bacteria (Amann *et al*., 1995; Stackenbrandt, 1997).

2.3.3 STREPTOMYCETES

One actinomycete genus, *Streptomyces* has become pre-eminent for genetic research. This could be attributed to not only the ability of the organism to produce a vast number and wide variety of antibiotics but also to the ease of isolating the organism from the soil and the convenience of cultivating them in the laboratory.

Streptomycetes are aerobic gram-positive soil bacteria that grow vegetatively as a branching and generally non-fragmenting mycelium. Individual branches are called hyphae. Occasional cross walls are formed in the hypha, with irregular spacing. After a certain amount of growth, some unknown stimulus, usually considered to be nutrient depletion, causes aerial branches to arise from the 'vegetative' substrate mycelium of surface grown colonies. The aerial mycelial branches eventually differentiate into chains of spores. Aerial hyphae appear to grow partially by utilising the degraded substrate mycelium.

Streptomyces colonies grown in laboratory conditions are sometimes visible as colonies with alternating surface colour associated with that of the spores and the white fluffiness typical of aerial mycelium. This is due to multiple rounds of germination and sporulation in the laboratory culture. (Dowding, 1973).

Germinated spores, vegetative hyphal fragments, aerial hyphal fragments produced by mutants blocked at any stage of differentiation are all capable of initiating a new colony.

2.3.4 STREPTOMYCETES: SECONDARY METABOLISM & DIFFERENTIATION

Most *Streptomyces* do not produce antibiotics during the period of vegetative growth. Instead, they produce antibiotics as their growth rate slows down. Hence production of the secondary metabolites is considered as inessential for vegetative

growth of the producing organisms. In *Streptomyces* colonies growing on solid surface, this slowdown occurs as the aerial mycelium starts to develop from the substrate mycelium. In liquid grown culture, it takes place at a 'transition stage' as biomass changes from the quasi-exponential toward the stationary phase.

It has been suggested that such timing of antibiotic production and differentiation is adaptive in helping to prevent invasion of microorganisms that could otherwise steal the nutrients released by the lysis of the substrate mycelium, which are meant to supply nutrients for the development of the aerial mycelium.

The genetic and physiological determinants of the switch between primary and secondary metabolism are still largely obscure. Two kinds of approaches are currently used to understand the switch mechanism. Physiological factors such as carbon and nitrogen sources or inorganic phosphate are being studied with reference to differentiation and antibiotic production to elucidate the role of these factors in the switching from primary metabolism to secondary metabolism. Such studies have led to the understanding that these physiological factors above a threshold concentration are potential switching devices.

The second approach has been to identify pleiotrophic mutants which are defective in the production of more than one antibiotic in the organism, or to isolate DNA fragments having a pleiotrophic effect on antibiotic production when they are over expressed or when the genomic copy of the gene is knocked out. This approach has led to the identification of several genes in *Streptomyces*, which affect just secondary metabolism or both the secondary metabolism and differentiation processes. Tables 1, 2 and 3 show some of the identified genes that affect secondary metabolism and their predicted functions.

Gene	Gene product
afsB	Transcriptional regulatory protein
afsR	Phosphoprotein similar to eukaryotic signal transduction pathways
absA1-absA 2	Similar to two component regulatory systems (negative regulator)
cutR-cutS	Similar to two component regulatory systems (negative regulator)
farA	Butyrolactone autoregulator receptor (negative regulator)

Table 1: Genes affecting secondary metabolism in *Streptomyces*

Several of the *bld* (bald) genes from *Streptomyces coelicolor A(3)* were capable of affecting both the secondary metabolism and differentiation processes. Mutants of *Streptomyces coelicolor A(3),* which lack an obvious aerial mycelium were designated as bald (*bld*). Most of the *bld* mutants turned out to be regulatory proteins. Following the identification of Bld mutants, several other genes have been identified in Bld mutant hunts, as having a pleiotropic effect on differentiation as well as secondary metabolism.

Tables 2 and 3 show some of the identified genes that affect secondary metabolism as well as differentiation; and their predicted functions.

Gene	Gene product
bld A	$tRNA^{Leu}$
bld B	Small DNA binding protein
bld C	?
bld D	Small DNA binding protein
bld G	Likely anti- sigma factor
bld H	?
bld I	?
bld J	?
bld K	Oligopeptide transporter
bld L	?
bld M	Similar to response-regulator
bld N	ECF sigma factor

Table2: *bld* genes and their predicted functions, "?" indicates unidentified functional role.

Gene	Function	Phenotype of knock out mutant
citA	Citrate synthase	Bald;
acoA	Aconitase	Bald;
cya	Adenylate cyclase	Bald; (suppressed by buffering);
obg	GTP binding protein	Mutational lethal (multiple copies inhibit aerial growth)
relA	(p)ppGpp synthetase	Retarded aerial growth on low nitrogen medium (overexpression accelerates growth)
catB	Catalase	Bald
brgA	Unknown	Bald; resistant to inhibitor of ADP ribosyl transferase

Table 3: Other genes capable of influencing secondary metabolism and differentiation in *Streptomyces*

Thus, regulatory elements governing the development in *Streptomyces* seem to be determined by nutritional, and physiological as well as genetic factors. These regulatory factors could either act at the secondary metabolism alone or at a level that affects both the differentiation and secondary metabolism processes. However, the exact role of differentiation in relation to secondary metabolism remains obscure and is yet to be worked out. Most of the work pertaining to this subject has been conducted using *Streptomyces coelicolor* and therefore the relevance to other *Streptomyces* is also to be confirmed.

2.3.5 STREPTOMYCES: GENOME AND ANTIBIOTICS SYNTHESIS

All the essential genes of *Streptomyces coelicolor* lie on a chromosome that is about ~8 mb in size. *S. ambofaciens* has a similar genome size (Leblond *et al*., 1990). Pulsed field gel electrophoresis (PFGE) of the streptomycete genome revealed a linear chromosome in all the species studied (Lin *et al*., 1993). Terminal structures on the chromosome of *Streptomyces lividans* (Lin *et al*., 1993), *S. griseus* (Lezhava *et al.*, 1995) and *S.ambofaciens* (Leblond *et al*., 1996) were identified as long inverted

repeats. The chromosomal ends of adjacent regions of *Streptomyces* chromosome tend to be highly unstable and could undergo frequent deletions of up to 2Mb. The deletion mutants of various species may show phenotype changes, especially affecting aerial mycelium formation, pigment and antibiotic production, and resistance to antibiotics (Hutter *et al*., 1988).

All of the antibiotic genes studied so far are chromosomally located with the exception of methylenomycin gene cluster, which is on the SCP1 plasmid of *S. coelicolor* (Kirby & Hopwood, 1977). More than one antibiotic cluster may be found in a single *Streptomyces sp*. Gene clusters for actinorhodin, undodecylprodigisin and CDA (Calcium dependent antibiotic) are encoded by *S.coelicolor* genome. In some cases, partial clusters are also found as in the case of rapamycin producer *S. hygroscopicus* (Ruan *et al*., 1997). The genes for each individual antibiotic biosynthesis are clustered together in a series of contiguous operons, which can range from 15 to 100kb size. The clusters usually also include pathway specific regulatory genes and one or more genes for resistance to the organism's own antibiotic (Chater *et al*., 1992).

2.4 POLYKETIDES

2.4.1 WHAT ARE POLYKETIDES?

Polyketide compounds are a large group of structurally diverse metabolites that are synthesized by repetitive condensations of small carbon precursors; typically acetate or propionate acyl groups derived from malonyl or methylmalonyl coenzyme A thioesters, respectively. In other words, polyketides are polymers of ketide units linked together. Polyketides fall into two structural classes: aromatic and complex depending on the building blocks of carbon acyl units and the extent of reduction after each round of condensation reaction.

2.4.2 AROMATIC AND COMPLEX POLYKETIDES

Aromatic polyketides are built mainly from condensation of acetate acyl units and the β– carbonyl group after each condensation step is left largely unreduced. The polyketide chain is rearranged immediately after synthesis to produce an aromatic product. Examples of these aromatic products are polycyclic aromatic compounds such as oxytetracycline, actinorhodin and anthracycline compounds such as daunorubicin. The enzymes responsible for the biosynthesis of the aromatic polyketides are encoded by genes called aromatic polyketide synthases or otherwise known as polyketide synthase type II (PKS II).

Complex polyketides can be built by condensation from acetate, propionate and butyrate acyl units. The extent of the β– carbonyl reduction in complex polyketide synthesis can vary from one condensation cycle to the next. The polyketide chain continues to grow until the desired length is reached, upon which the chain is cyclized to form the end product. The enzymes responsible for the biosynthesis of the complex polyketides are encoded by genes called modular polyketide synthases or otherwise known as polyketide synthase type I (PKS I).

2.4.3 STRUCTURE AND FUNCTION OF POLYKETIDES

Polyketides are diverse in structures. Structural diversity of the polyketides is reflected in the diversity in their biological activity. Examples of polyketide chemical class include macrolides, tetracyclines, anthracylclines, avermectins, and many others. Polyketides encompass bacterial metabolites such as antibiotics, fungal aflatoxins, plant flavonoids and hundreds of compounds of different structures that exhibit anti bacterial, antifungal, antihelminthic and antitumor properties (Fig. 1).

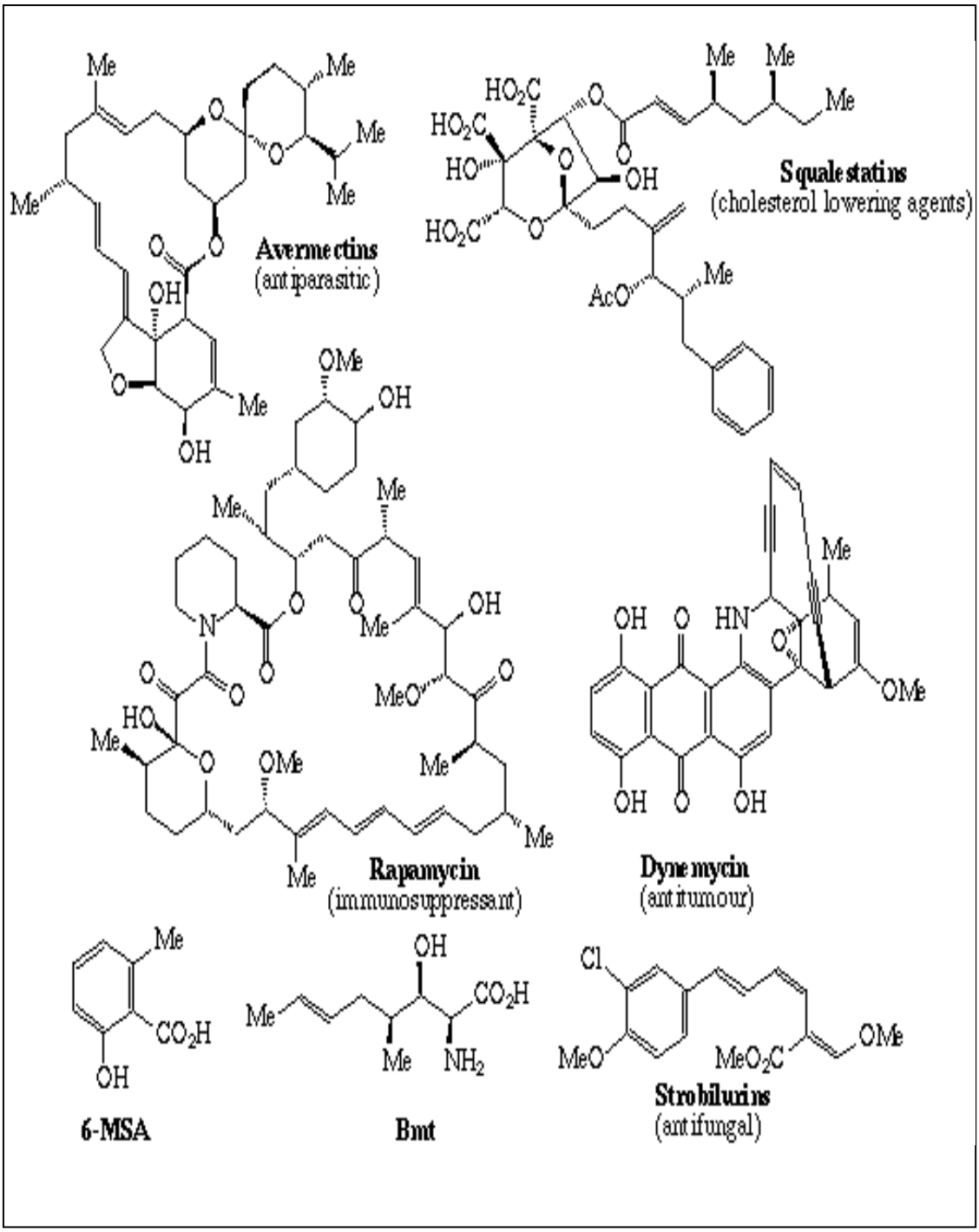


Figure 1: Diverse structures of polyketides and their functions. Polyketide biosynthesis" (Staunton, J and Weissman, J. K., 2001)

2.4.4 HISTORICAL PERSPECTIVE OF POLYKETIDES

The term "polyketide" was introduced into the chemical literature in 1907 by John Norman Collie in a paper entitled "Derivatives of the multiple ketene group"

(Collie, 1907). Collie proposed that the basic building block for a large number of aromatic plant compound was ketene (CH_2=C=O) or its hydrolysis product. Hence, the designation of the compound as polyketide. He provided experimental evidence in support of the hypothesis that the aromatic compounds were formed by condensation of acetic acid, acetoacetate and higher homologues of acetate.

In the 1940s, Rittenberg and Birch proved that acetate was involved in fatty acid biosynthesis by using radio labelled acetate (Rittenberg *et al*., 1944, 1945). Following this, Feodor Lynen led to the discovery that acetyl CoA that acted as active acetate was in fact the precursor in fatty acid synthesis.

Birch was stimulated with this new development and went on to systematically analyse and confirmed that structures of many aromatic compounds were compatible with a biosynthetic origin from the folding of extended β- ketochains from acetate. Birch was also the first to study biosynthetic experiments on fungal compounds, which placed the role of acetate in polyketide synthesis beyond doubt. These findings established the origin of polyketides (Birch *et al*., 1953a, 1953b, 1958).

Historically, significant developments in fatty acid synthesis have paved way for a better understanding of polyketide synthesis. This continues to be so even today.

2.5 FATTY ACID AND POLYKETIDE SYNTHASES

2.5.1 FATTY ACID SYNTHASES

Polyketide synthesis is similar in many respects to bacterial and mammalian fatty acid synthesis. Therefore before introducing the enzymes of the polyketide biosynthesis, it is appropriate to digress briefly into the closely related field of fatty acid biosynthesis.

Fatty acids are assembled from C2 units by repeated head to tail linkage. This assembly process is catalysed by enzymes known as fatty acid synthase (FAS). A

starter acyl unit, usually acetyl is condensed with a malonyl unit to form a carbon - carbon double bond by decarboxylation. The starter acetyl unit is attached to keto synthase (KS), whilst the malonate is attached to an acyl carrier protein (ACP). The condensation reaction is catalysed by the ketosynthase. The resulting β- keto ester, which is attached to the acyl carrier protein (ACP), is successively reduced by ketoreductase (KR), dehydrated by dehydratase(DH) and reduced once again by enoyl reductase(ER) to give a saturated chain longer than the original by two methylene groups. This sequence of reaction completes the first round of chain extension. The elongated chain is then transacylated to the KS, and a new cycle is initiated. This process is repeated until the desired chain length is reached (usually 14, 16 or 18 carbons). At this stage, the chain is transferred to a thioesterase (TE) which catalyses the release of the assembled product as a free acid or an acyl ester.

The structural organization of the FAS depends on the type of organism. Bacterial fatty acid synthases consist of discrete set of proteins that can be isolated separately and are designated as type II FAS enzymes. In contrast, mammalian FAS are large multifunctional proteins and are designated as type I FAS enzymes. Various intermediate stages of organization are found in other organisms.

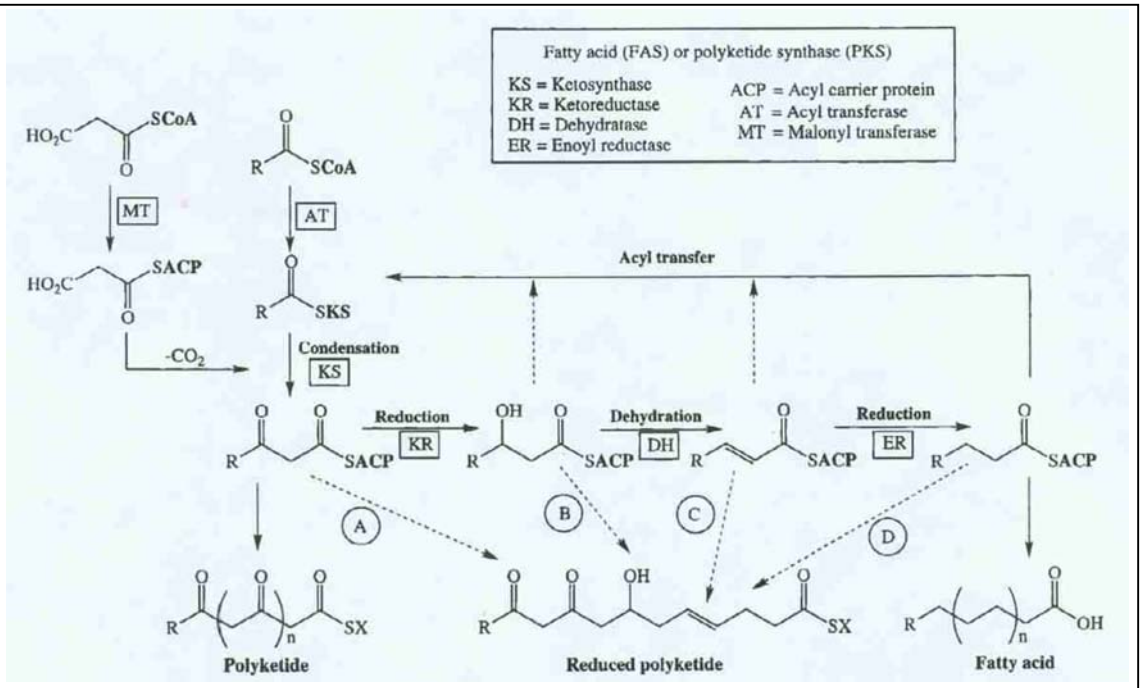


Figure 2: Mechanism of Fatty acid and Polyketide synthesis. The above diagram shows the various activities catalysed by the different domains of the fatty acid or polyketide synthases. A- D represents the possible reductive cycles that can lead to keto, hydroxyl, enoyl or methylene functional groups in the endproduct. Adapted from "Genetic contributions to understanding polyketide synthases" (Hopwood, D. A., 1997).

2.5.2 POLYKETIDE SYNTHASES

Like fatty acid synthases, two types of synthases also catalyse the polymerisation process of polyketide synthesis, type I polyketide synthase (PKS I) and type II polyketide synthase (PKS II). Polyketide synthesis however differs from long chain fatty acid synthesis in several aspects. For example, different starter units (linear or branched carboxylic acids, aromatic and aliphatic rings etc) are used for polyketides, whereas acetate or occasionally propionate or branched chain carboxylic acids are employed for long chain fatty acid synthesis. Secondly, the extent of processing in polyketide synthesis may not be constant through out unlike that of the

fatty acid synthesis. A new cycle in polyketide synthesis may initiate with an acyl group containing a β- keto, β-hydroxy, β-ene or fully reduced β-carbon (Fig. 2).

Polyketide synthase I genes catalyse the biosynthesis of complex polyketides and they typically consist of several modules of 3 to 6 domains, encoding large multifunctional polypeptides. The modules are termed as loading module, extender module and releasing module, depending on their role in the biosynthesis of the encoded polyketide. Each module catalyses a single step in the biosynthesis of the complex compound. Synthesis begins at the first module, loading module, located at one end of the PKS, and continues to the end through multiple extender modules, each of which extends the growing polyketide chain by two carbon units and modifies. Finally, the polyketide chain is transferred to the releasing module which catalyses the cyclisation and release of the polyketide.

Each module contains three essential enzymatic activities (domains) responsible for the polymerisation of the ketide units. They are the Keto Synthase (KS), Acyl Transferase (AT) and Acyl Carrier Protein (ACP) domains. AT domain selects the building block while KS and ACP are involved in the linking of the building block to the growing chain. A module may also contain 1-3 additional enzymatic activities involved in the modification of the growing polyketide chain. Dehydratase(DH), Keto Reductase (KR) and Enoyl Reductase (ER) are the domains which catalyse the modification of the growing polyketide chain.

Therefore, the structure of the complex polyketide is determined by the number of modules, the specificity of the AT domain in the modules and the variation in the modifying enzymatic activities of the modules. The number of modules would determine the size of the polyketide as each module catalyses a single step in the biosynthesis of the complex compound. The specificity of the AT domain would

determine the type of building block used to synthesise the polyketide. The varying modifying enzymatic activity would result in different level of reductive processing of the growing polyketide chain after each step of extension. These factors contribute to the numerous variety of complex polyketide structures found in nature although the biosynthetic machinery (enzymes) and mechanism of biosynthesis of the several different complex polyketides are typically similar. Sequence analysis of the PKS I DNA encoding the biosynthesis of different compounds reveals extensive similarity of enzymes to KS, AT, DH, KR, ER and ACP with regard to the fatty acid biosynthesis and also for the different producing organisms. Organisation of a few PKS I gene clusters is shown in Fig. 3.

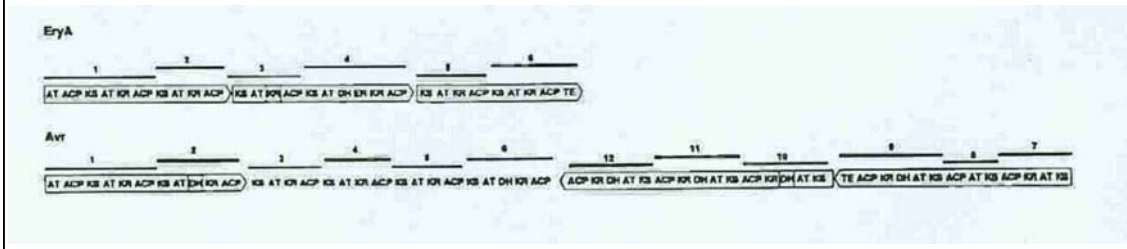


Figure 3: Organisation of various PKS I genes encoding large multifunctional polypeptides. Arrows indicate ORFs. Adapted from "Polyketide synthesis" (Katz, L. and Donadio, S. 1993)

Polyketide synthase II genes catalyse the biosynthesis of aromatic polyketides and they typically consist of 4 to 6 genes encoding mono or bifunctional enzymes. This set of enzymes is used repeatedly to synthesise the entire aromatic compound. Sequence analysis of the PKS II DNA encoding the different compounds reveals extensive similarity of enzymes to KS, ACP and KR enzymes of the fatty acid biosynthesis and also for the different producing organisms. Organisation of various PKS II gene clusters is shown in Fig. 4:

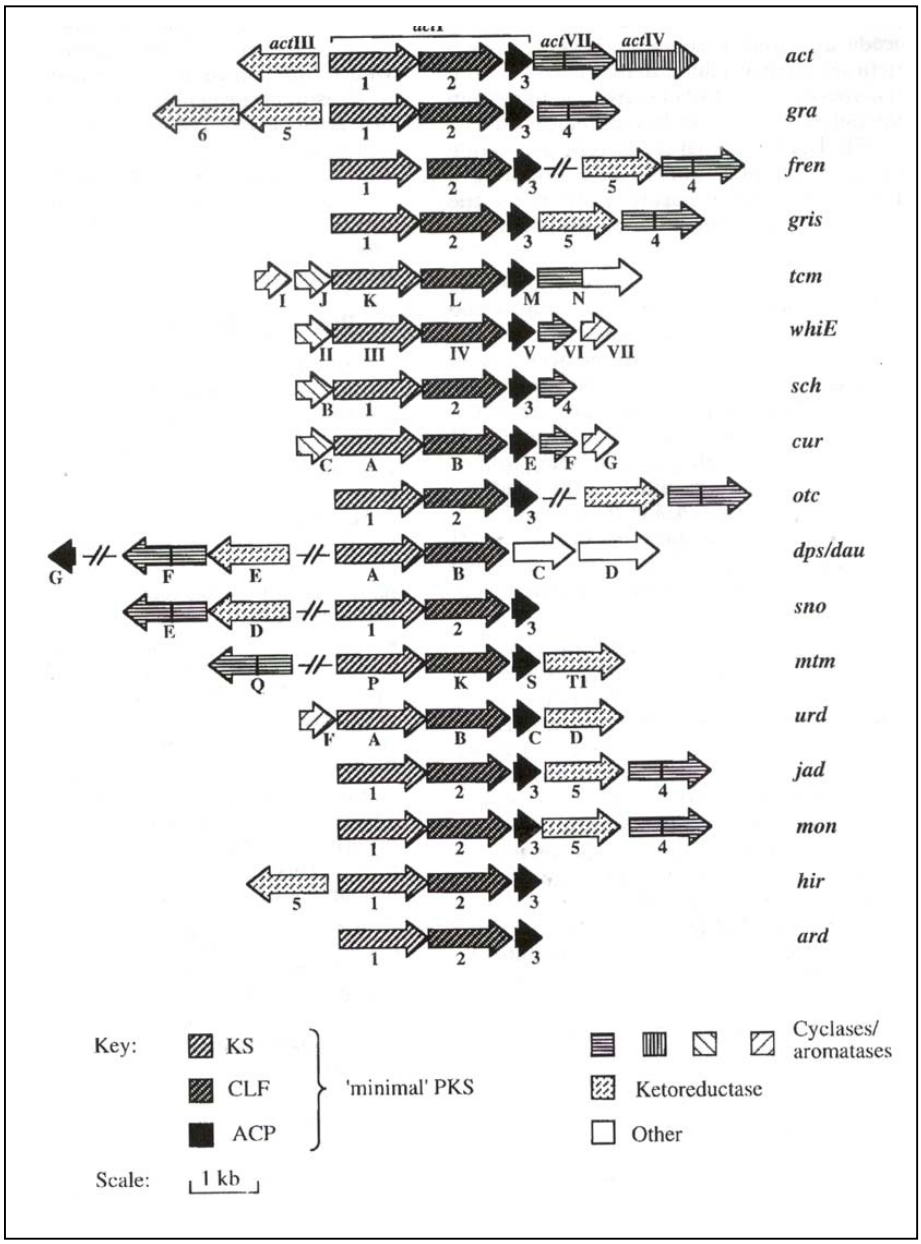


Figure 4: Organisation of various PKS II genes encoding mono or bifunctional proteins. Arrows indicate ORFs. Adapted from "Genetic contributions to understanding polyketide synthases" (Hopwood, D. A., 1997).

2.6 DISCOVERY OF TYPE I POLYKETIDE SYNTHASES

2.6.1 ERYTHROMYCIN POLYKETIDE SYNTHASE GENES

Long before the molecular biology revolution, indirect information regarding the biosynthetic properties of the modular PKSs was gained through incorporation experiments with ^{14}C, ^{13}C, ^{18}O and ^{2}H labelled substrates and intermediate analogs. For example isotope-labelling studies demonstrated that the carbon chain backbones of

natural products such as erythromycin, tylosin, monesin and avermectin are derived through C-C bond formation from acetate, butyrate building blocks.

More recently, the incorporation of exogenously added analogs of the putative biosynthetic intermediates into corresponding polyketides have proven without doubt that the modular PKSs act via a processive mechanism in which the oxidation level and stereochemistry of the growing polyketide chain is adjusted after each step of polyketide chain elongation. However, in depth knowledge of the biochemical basis for the processive assembly has only become possible with the advancement of molecular genetic tools.

The first modular PKS (PKS I) genes to be cloned were that of *S. erythraea*, encoding proteins for the biosynthesis of the complex polyketide, 6-deoxyerythronolide. This polyketide gene cluster is designated as *eryA.* The polyketide, 6-deoxyerythronolide is the aglycone moiety of erythromycin, which has to be oxidised and glycosylated to yield erythromycin.

Two separate groups identified the *eryA* gene cluster using different but complementary approaches. A gene fragment conferring resistance to erythromycin was cloned by Thomson *et al* in 1980. Peter Leadlay's group at Cambridge University used this resistance gene denoted as ermE as a hybridisation probe to clone genes for erythromycin biosynthesis from the genome of *S. erythraea,* based on the assumption that resistance genes and biosynthetic genes are clustered together. DNA fragments isolated were sequenced and used in gene disruption and complementation experiments to prove their function. Further chromosomal walking led to the identification of the eryA genes (Leadlay *et al*., 1990).

Meanwhile, Leonard Katz and co-workers at Abbott laboratories had also cloned the genes of the *eryA* cluster. They cloned the *eryA* by complementation of an

erythromycin non-producing mutant with a cosmid library generated from the *S. erythraea* genes. The insert fragment of the complementing clone was then sequenced. Partial sequence information was published by both groups in 1990. Detailed analyses of the gene cluster followed shortly thereafter (Bevitt *et al*., 1992, Donadio *et al*., 1991, Tuan, 1990).

The structural genes responsible for the biosynthesis of the first macrolide intermediate are three enormous open reading frames (ORFs), eryAI, eryAII and eryAIII, encoding the three gigantic multienzyme polypeptides. Each ORF is approximately 10 kb and consists of two repeating units designated as modules. Sequence analyses revealed that eryAI encoded a loading domain and 2 extender modules, eryAII encoded 2 extender modules and eryAIII encoded 2 extender modules and a final thioesterase domain (Fig. 5). Further sequence comparisons also showed that each of these modules consisted of 4 to 5 domains with considerable similarity to enzymes responsible for each of the individual steps of fatty acid synthesis.

Gene disruption experiments confirmed the predicted boundaries of eryAI, eryAII and eryAIII ORFs and proved the involvement of eryA genes in the synthesis of 6-deoxyerythronolide.

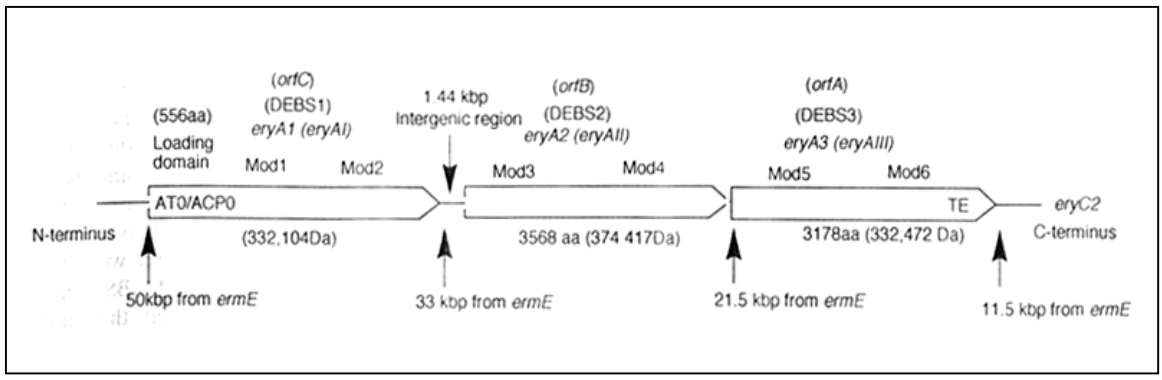


Figure 5: Open reading frames of erythromycin biosynthetic gene cluster. Adapted from "Type I polyketide biosynthesis in bacteria" (Rawlings, J. B., 2001)

2.6.2 DOMAIN IDENTIFICATION OF ERYTHROMYCIN POLYKETIDE SYNTHASE GENES

In 1991, Donadio *et al* sequenced 35kb of DNA which includes all of eryA. Analysis revealed that eryAI and eryAII were separated by an insertion element of 1.44kb whilst eryAII and eryAIII were contiguous. Each of the eryA ORF consisted of 2 repeated units (modules), which ranged from 4.3 kb to 6.5 kb in size. The sequences had a similarity of 64% or higher. The deduced amino acid sequences of the three ORFs were compared to FAS and other PKS systems, and the catalytic activities /domains in each module were assigned. A total of seven ACPs, six KSs, eight ATs, six KRs, one DH and one ER have been identified from the six repeated units.

Each domain was presumed to catalyse a single step of the processive assembly. The putative active sites of the domains were identified: Predicted active site motif GHS*SG motif was located in all the 8 AT domains, keto synthase active site motif GPXXXXXTAC*SS was identified in all of the 6 KS domains, signature sequence of ACP active site was found in all the 7 ACP domains, and a GXXGXXAXXA motif proposed as a common fingerprint region in NADPH reductase was identified in the 6 KR domains. One such KR domain in module 3 had a gap in the sequence corresponding to the highly conserved VSRRG motif, and therefore was proposed to be nonfunctional.

DH and ER domains were proposed in module 4 by comparison with that proposed for rat FAS, but the extent, exact location and limits of the individual domains were not described.

Leadlay *et al* (Bevitt, 1992) proposed the active site of DH to have a HXXXGXXXXP motif by sequence analogy with the *E. coli* FabA, which is 3 hydroxydeconyl thioester dehydratase and the active site of ER to have a histidine

residue at the active site by sequence analogy with rat and chicken FAS.

2.6.3 ENZYMOLOGY OF ERYTHROMYCIN POLYKETIDE SYNTHASE GENES

Most of our knowledge about modular PKSs arises from studies of the eryA system. DNA sequence analysis of the genes led to the postulation of the widely accepted model presented in the Fig. 6 (Donadio *et al*., 1991).

Here the acyl transferase (AT) domain of the loading domain in eryAI initiates the polyketide chain building process by transferring the propionyl CoA primer unit via the pantetheinyl residue of the loading domain acyl carrier protein (ACP) to the active site cysteine of the ketosynthase of module 1 (KS1). The acyl transferase in module 1 (AT1) loads methylmalonyl CoA onto the thiol terminus of the ACP domain of module 1. KS1 then catalyses the first polyketide chain elongation reaction by decarboxylative condensation between the methylmalonyl and propionyl residues resulting in the formation of a 2 methyl 3 keto pentanoyl- ACP thioester. The latter intermediate is reduced by the keto reductase of module 1 (KR1) giving rise to ACP bound (2S, 3R)-2 methyl, 3 hydroxy pentanoyl ACP. At this point, module 1 has completed its function and the diketide product is transferred to the core cysteine of KS 2, whereupon it undergoes another round of reduction, resulting in the formation of the corresponding triketide. This process is repeated six times, with each module being responsible for a separate round of chain elongation and reduction, as appropriate, of the resulting β- ketoacyl thioester. Finally the thioesterase at the C- terminus of ery AIII is thought to catalyse the release of the finished polyketide chain by lactonisation of the product generated by module 6. In summary, six methylmalonyl CoA acyl units are converted to 6- deoxyerythronolide by the catalytic activity of the eryA encoded PKS I enzymes.

It should be noted that although the domain organization of the eryA was in complete agreement with the information available from the earlier isotopic labelling studies, *unequivocal a priori* deduction of the product structure would not have been possible from the sequence information alone. This is because, firstly, the stereochemical features of the end product cannot be deduced from the primary structure alone. Secondly, the regio- specificity of cyclisation is not overtly dictated in the organization or sequence of eryA domains. Finally, occurrence of a domain would not necessarily indicate its functionality. For example, module 3 of eryA cluster contained a KR domain, albeit a non-functional one, whose amino acid was found to deviate significantly from that of the other KR domains. However, this deduction would not have been possible without the prior knowledge of the polyketide structure.

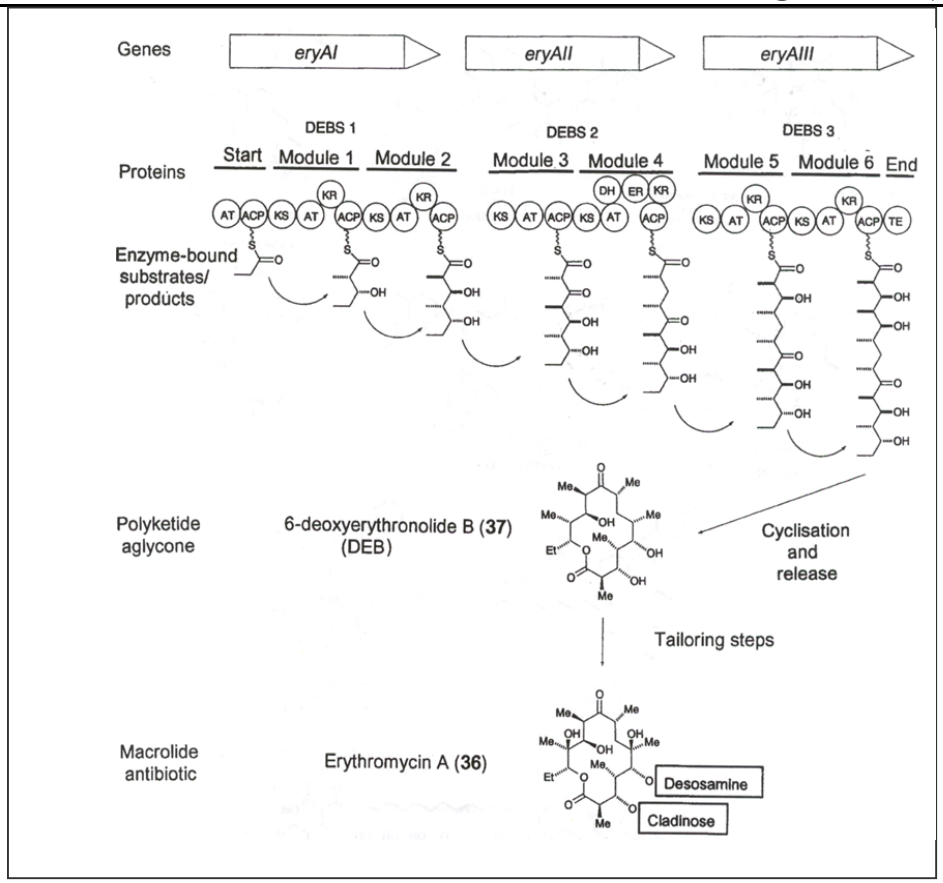


Figure 6: The proposed mechanism of erythromycin biosynthesis. Adapted from "Genetic contributions to understanding polyketide synthases" (Hopwood, D. A., 1997).

2.6.4 THE PROGRAMMING MODEL AND PROOF OF FUNCT ION

Initial evidence for the assembly line model for the programming of erythromycin was provided by the sequence itself. Firstly the number of modules of putative catalytic sites corresponded in number to the number of condensation steps needed to build the 6-deoxyerythronolide heptaketide. Secondly, the features of specific modules could be related to their proposed functions.

eryAI had extra N- terminal AT and ACP domains before module1 for loading of the propionyl CoA starter unit onto the KS domain of module 1.

eyAIII was unique in carrying a putative thioesterase domain after module 6 for hydrolysis of the final thioester bond between the completed polyketide chain and the 4',-phospho-pentathiene prosthetic group of the last ACP domain to release the carbon

chain.

Module 3 of eryAII lacked all three reductive functions DH, KR and ER agreeing with the presence of an unreduced keto group after the third condensation, while module 4 was unique in carrying candidate domains for all three such functions, as expected in view of the reduction of the keto group right through to a methylene after the fourth condensation.

Several line of experiments involving domain mutagenesis (by deletion, inactivation or swapping), synthetic precursor feeding to blocked mutants and module swapping have been performed by different groups, which have provided substantial evidence for the deduced programming model of the eryA PKS I cluster. As this subject is a specialised field on its own accord and beyond the scope of this review, it is not discussed in depth (Khosla, 1997).

2.7 OTHER MODULAR POLYKETIDE SYNTHASES

Since the discovery of the eryA genes, the involvement of modular PKSs in the biosynthesis of several other complex polyketides has been reported. Although some variations have been observed in the content and organization of the different systems, the key features of the modular hypothesis remain unchanged in whole or in part. PKS clusters encoding for complex polyketides rapamycin, FK506, spiramycin, oleandomycin, avermectins, niddamycin, methymycin, picromycin, pimaricin, nystatin and tylosin have been cloned. The cloning strategy and features of some of this polyketide gene cluster are discussed in the following section.

2.7.1 SPIRAMYCIN

The genes encoding the biosynthesis of the polyketide precursor of the 16-membered macrolide spiramycin have been cloned. Spiramycin is a 16-membered polyketide produced by *Streptomyces ambofaciens*. The gene cluster encoding for spiramycin biosynthesis was identified by cloning the spiramycin resistance gene. The biosynthetic genes linked to the resistance gene were then identified by complementation of blocked mutants.

This modular PKS includes seven modules whose organization is colinear with the biosynthetic order as in the eryA gene cluster. However three of the ORFs are unimodular in the spiramycin gene cluster. Furthermore, the loading domain of the spiramycin gene cluster also includes a ketosynthase domain in addition to the AT and ACP domains. However, the amino acid sequence of this ketosynthase domain deviates in the active site motif. The cysteine residue of the active site motif is replaced with a glutamine residue and therefore the KS domain is presumed to be inactive. Some of the AT domains of the spiramycin gene cluster are deduced to have specificity for ethyl malonyl CoA and malonyl CoA substrates in addition to the usual methylmalonyl CoA substrates (Yue *et al.*, 1987).

2.7.2 RAPAMYCIN

The entire gene cluster for rapamycin biosynthesis has been cloned and sequenced from *Streptomyces hygroscopicus* (Schwecke *et al.*, 1995). The 32-membered rapamycin structure, the PKS encoding for rapamycin is comprised of 14 modules. The gene cluster encoding for rapamycin biosynthesis was cloned using eryA gene probes. The gene cluster was identified to be 107 kb in size. Gene disruption studies have been used to prove the involvement of the cloned genes in the

biosynthesis of rapamycin (Hutchinson *et al.*, 1997).

In contrast to the eryA gene cluster, organization of the genes of rapamycin gene cluster is not colinear with the biosynthetic order. The ORFs of rapamycin gene cluster are large, consisting of 4 to 6 modules. The loading domain of rapamycin gene cluster is unusual and is comprised of a putative acyl CoA ligase and enoyl reductase. Typical thioesterase domain is also not found in the rapamycin gene cluster. Instead, a pipecolate-incorporating enzyme, which presumably completes the rapamycin macrocycle, is present.

Until the rapamycin PKS was sequenced the database of modular PKS only included sequences of AT domain with specificity for methylmalonyl CoA. The rapamycin module includes seven AT domains each with specificity toward a malonyl CoA or methylmalonyl CoA. Comparative analyses of the AT domains from rapamycin, erythromycin and oleandomycin PKSs revealed the substrate specificity of AT domain could be unambiguously predicted from two short consensus sequences of 5- 8 residues.

2.7.3 CANDICIDIN

Candicidin is a 38-membered polyene polyketide. The aglycone moiety of candicidin is identical to a related compound FR-008. FR-008 is produced by *Streptomyces* sp. FR-008. A gene cluster involved in FR-008 biosynthesis was isolated by hybridisation, initially using a gene involved in the biosynthesis of *para* -amino benzoic acid starter unit and later with several eryA gene probes. The hybridisation patterns with the eryA probes revealed that the DNA encoding the modular PKS extended approximately 105kb. It was pointed out that on the assumption that each PKS module was approximately 5 kb, 21 modules were predicted in the gene cluster to

be encoding for the synthesis of the FR-008 aglycone. This prediction was in striking agreement with the number of condensation steps required for the synthesis of the FR-008 aglycone. This finding was significant in implying a one to one relationship between modules and rounds of condensation. Furthermore, consistent with the presence of a *para*-amino benzoic acid primer unit in the polypeptide backbone, one end of the gene cluster appeared to encode the *para*-amino benzoic acid synthase and ligase genes. Gene disruption experiments were done to confirm the involvement of the cloned genes in the synthesis of the FR-008 aglycone (Deng *et al*., 1994).

2.7.4 SORAPHEN

Soraphen A is a 18-membered compound produced by the myxobacterium *Sorangium cellulosum*. It is the first example of a functional modular PKS so far known outside the actinomycetes. Interestingly, the DNA encoding part of the soraphen gene cluster was cloned by the use of a PKS II specific probe (graI). This is the first and only example of PKS I genes cloned by the use of PKS II specific gene probe. Gene disruption proved the involvement of the cloned DNA in soraphen biosynthesis and sequencing revealed part of a gene encoding one complete module of PKS active sites and an incomplete second module (Schupp *et al.*, 1995).

2.8 ELUCIDATION OF BIOSYNTHETIC PROCESS OF POLYKETIDES

The complete study of biosynthesis of polyketides would consist of

1) identifying the primary metabolites from which the polyketide is derived,
2) isolating the intermediate metabolites of the pathway which would give a better understanding of the sequence of reactions by which primary metabolites are converted onto the final molecule,
3) identifying the enzymes that catalyse this conversion process and

determining the organization and regulation of the governing genes.

4) identifying the genes of the polyketide biosynthesis,
5) identifying the regulatory factors of the polyketide biosynthesis and
6) working out the regulatory mechanism of the polyketide biosynthesis.

Although various experiments pertaining to the different steps could be performed in the above order, it is common to obtain relevant information through genetic methods before any biochemical evidence is obtained. The following section gives a brief review on the some of the approaches used in studying the biosynthesis of polyketides.

2.8.1 IDENTIFICATION OF BUILDING BLOCKS BY TRACER TECHNIQUES

Feeding the culture of the polyketide-producing organism at the end of their growth phase with radiolabelled presumptive precursor of the polyketide aids in the identification of the building blocks of the polyketide. After incubating the culture for an appropriate period of time to allow for the synthesis of the polyketide, solvents are added to the fermentation broth to extract the end product polyketide. Extracted compound is then purified and analysed by NMR to determine the incorporation of isotope in the end product polyketide. If the isotopic label is detected in the polyketide end product then it is concluded that the radiolabelled presumptive precursor is indeed the building block of that polyketide.

2.8.2 *ISOLATION OF INTERMEDIATES*

Identification of intermediates of the polyketide biosynthetic pathway is another approach to studying the polyketide biosynthetic pathway. A common procedure suitable for identifying intermediates of biosynthetic pathway is based on

the isolation of mutants blocked in one of the enzymatic reactions of the biosynthetic reaction. Blocked mutants are generally obtained by random mutation by subjecting the producing strain to mutagens such as UV treatment and screening the UV treated clones to isolate non-producers.

The " blocked mutants" often accumulate the substrate of the blocked reaction in the medium. Blocked mutants that do not produce the polyketide on their own but produce when grown together are mutants blocked in two different but complementing steps of the biosynthetic pathway. In this case, the inability of one mutant to produce an intermediate is complemented by the ability of the other mutant to accumulate it.

The accumulated intermediate product can be isolated and identified to verify that it is indeed the intermediate of the biosynthetic pathway. The original strains is assessed for conversion of this intermediate into the end product. This is done by feeding the intermediate metabolite to the producing strain for a specific time and determining the amount of polyketide produced in comparison to producing strain that is not fed.

2.8.3 IDENTIFICATION OF ENZYMES

Comparing enzymatic activities in producing and non-producing variants of the polyketide producing strain can identify enzymes of the polyketide biosynthetic pathway. The presence of an active enzyme in a producer and the absence of that enzyme in the non-producing variant are taken as an indication of the involvement of that enzyme in the biosynthetic pathway of the polyketide.

2.8.4 *IDENTIFICATION OF GENES*

Identifying the genes encoding the enzymes that catalyse the various steps of the biosynthetic pathway of the polyketide is the most commonly used approach to studying the polyketide biosynthesis. Initial studies have shown that particularly in actinomycetes, the biosynthetic genes are clustered together, usually on the chromosome. Regulatory genes and self-resistance genes are also usually present as part of this cluster. These features of the polyketide synthases in addition to the developments in molecular biology have made isolation and sequencing of entire polyketide synthase clusters more feasible.

Identification of building blocks by precursor feeding or identification of intermediates by complementation of non-producing mutants only allows for the elucidation of the biosynthetic pathway one step at a time. Identification of enzymes of the polyketide biosynthetic pathway by isolation of proteins is also not very feasible as the polyketide synthases are very large and isolation of these large proteins without affecting the function is not easy. Isolation of the genes involved in the polyketide biosynthesis not only allows for the elucidation of the entire pathway but also allows us to harness the potential of these genes in proving the function of these genes by gene disruption studies or in producing hybrid polyketides by domain swapping etc. Therefore, identifying the genes for polyketide biosynthesis is considered to be more beneficial in elucidating the biosynthetic process of polyketides.

There are several approaches to cloning the polyketide biosynthetic genes, which is discussed in the next section. Once a biosynthetic gene has been cloned and sequenced, the nucleotide sequence of the gene could be compared to those available in the data banks, which would give useful information on the nature and function of the gene product. Sequence analysis of the organization of the genes would give

suggestive idea of the mechanism of the biosynthetic pathway.

2.9 STRATEGIES FOR CLONING POLYKETIDE SYNTHASE GENES

2.9.1 COMPLEMENTATION OF MUTATIONS BLOCKED IN PRODUCTION

In the earlier part of the review, identification of the intermediates in polyketide biosynthesis by complementation of blocked mutants was discussed. Here, we see that complementation can be used to clone the polyketide biosynthetic genes. However the complementation experiments for this purpose is considered to be *in vivo.* The aim of the *in vivo* complementation experiment is to restore polyketide production in a non-producing mutant by introducing DNA from the wild type organism into the non-producing mutant. A shotgun library of DNA from the wild type producer is introduced into the non-producing mutant and screened for restoration of the polyketide biosynthesis. Sub-cloning experiments are then used to identify the smallest piece of DNA that restores production. Sequencing of the insert fragment followed by sequence analysis to characterize the physical limits of the gene and by comparison with the well-understood proteins, provide an inkling of the role that its protein product plays; if this has not already been revealed by the effect of the mutation.

This approach was used to clone all the genes for actinorhodin production from *S. coelicolor* thereby demonstrating the clustering of secondary metabolism, structural, resistance and regulatory genes (Malpartida *et al*., 1984). Shortly, thereafter clusters of genes for the production of tetracenomycin, streptomycin and bialaphos were cloned in the same way.

2.9.2 SEARCH OF HOMOLOGOUS GENES

Now that a large number of antibiotic production genes and gene clusters have been cloned from actinomycetes and fungi, it is often possible to use a known gene as

a probe to clone the genes for newly discovered metabolites.

As there are only a few different types of secondary metabolic pathways, and considerable homology exists among genes encoding functionally related enzymes, a known gene can be used to hybridise to homologous genomic DNA, to clone and characterize the homologous gene.

Comparison of the sequences of the various polyketide synthases revealed that the sequences of the polyketide synthases are highly conserved, especially around their active site regions. However, it was also noted that the sequences of PKS I were sufficiently different from that of PKS II and both were divergent enough from fatty acid synthases. Based on this knowledge, it seemed possible to use specific probes to accurately identify the two different polyketide synthases.

The genes most often used are the *Saccharopolyspora erythraea* eryA genes for type I polyketide synthases (PKSs) (Leadlay *et al.*, 1990, Donadio *et al.*, 1991) or the *S. glaucescens* tcmKL homologs for the type II PKS (Malpartida *et al*., 1987).

If this method is not successful, degenerate primers designed from the highly conserved regions of PKS can be used to amplify the corresponding region of the genomic DNA from the newly discovered polyketide-producing organism. Polyketide gene cluster encoding the genes for niddamycin biosynthesis was discovered by this approach (Kakavas *et al.*, 1997).

Once the desired gene is obtained which can be established by the loss of metabolite formation as a result of targeted disruption, then, the remaining genes for the metabolite biosynthesis can be found in the surrounding DNA by chromosomal walking. Although cloning by DNA homology provides less initial information about the biosynthetic pathway than the isolation of blocked mutants, it often is the faster way to identify and characterize the production genes. This approach has been most

successful for the polyketide metabolites, oligopeptide antibiotics.

With either approach, the involvement of the cloned DNA must be established by gene disruption or enzymatic assay of the gene product; because microorganisms often contain more than one set of PKS genes. For instance, *Streptomyces peuticus* contains four chromosomal DNA fragments, which hybridise to the act I and tcmKL genes, but only one of these fragments is responsible for doxorubin production.

In view of the structural differences between the type I and type II PKS enzymes, eryA should logically be used to clone a new type I PKS gene and an act I probe should be used to clone a type II PKS gene cluster as illustrated by cloning of rapamycin and donorubicin genes. However, Schupp *et al* (1995) were able to clone a type I PKS gene for soraphenA biosynthesis from *Sorangium cellulosum* by using the graI gene and actI homolouge from the granticin producing *Streptomyces violaceoruber* as probes.

2.9.3 PROTEIN ISOLATION FOLLOWED BY GENE CLONING

Sequence information of a purified enzyme from a secondary metabolite pathway or the availability of antibodies to that enzyme would provide a secure way of identifying the corresponding gene as well as the rest of the gene cluster by reverse genetics. This is the least used method compared to the previous two methods. This could be explained by the low titres of such enzymes in wild type organisms and the difficulty in working with large polypeptides. However the gene probes for actinomycete PKS genes synthesized in accord with the biased codon usage have much less degeneracy than *E. coli* or human proteins and therefore often give clear cut results in DNA hybridisation experiments. The first genes for the biosynthesis of penicillins and cephalosporins, macrolide antibiotics were cloned in this way

(Sampsons *et al*., 1985).

2.9.4 EXPRESSION OF SECONDARY METABOLISM GENES AND GENE CLUSTERS

Expression of several sets of genes in a suitable host followed by detection of the metabolite formed is yet another approach to cloning a particular cluster of secondary metabolism genes. For this approach to be successful, several requirements have to be first met.

1) the cloning vector must be able to accept large DNA segments and be able to replicate autonomously or integrate into the host genome, stably.

2) the host must be able to express all the genes.

3) expressed enzymes have to be post translationally modified or supplied with necessary cofactors by the host organism and

4) the product formed must not be toxic to the host, or a resistance gene has to be cloned together with the structural genes.

This approach was used to shotgun clone the cephamycin C production genes from *Streptomyces cattelya* into *Streptomyces lividans* (Chen *et al.*, 1988).

2.9.5 GENOME SEQUENCING

With the advancement in molecular biology and sequencing technology, it is now possible to sequence entire genomes. The resulting data can be analysed for the presence of putative antibiotic producing genes by searching for homologues of PKS genes. The "red genes" for the biosynthesis of undecylprodigiosin and related red pigments of *S. coelicolor* were identified in this way.

Although there are several approaches to cloning the polyketide synthase genes, the choice of method depends not just on the overall purpose of the project but also the availability of DNA probes, cloning vectors, host organisms as well as

established DNA transfer techniques. Genome sequencing, gene expression and reverse genetics are less commonly used approaches. Cloning by complementation or using homology-based approaches is a well-established approach. The advantage in cloning by complementation is that function of the cloned gene is determined simultaneously. Cloning by homology-based approach would however require additional experiments to determine the function of the cloned gene.

2.10 PROOF OF FUNCTION OF CLONED POLYKETIDE SYNTHASE GENES

Mere cloning of polyketide synthase genes is not sufficient to prove the involvement of genes in the polyketide synthesis. In several instances, a single *Streptomyces* species produces more than one polyketide antibiotics and therefore would carry more than one biosynthetic gene cluster. Thus, it is necessary to determine which one of the many polyketide biosynthetic pathways encodes for the biosynthesis of a particular polyketide.

Some of the strategies discussed above to clone the polyketide synthase genes not only allow the cloning of the gene but also throws light on the function of the genes. Complementation of mutants blocked in production and expression of secondary metabolism genes and gene clusters in heterologous hosts allow for both cloning and elucidation of the function of polyketide genes. The other strategies discussed earlier only allow for the cloning of the polyketide synthase genes and therefore require additional experiments to determine the function of the cloned genes.

Gene expression and gene inactivation are two different but complementary ways to elucidate the function of the cloned genes. Gene expression of large polyketide synthase genes often pose lots of difficulty, especially when the polyketide synthase gene is isolated from a not so well understood producing strain. Therefore gene

inactivation is of particular importance to establishing the function of the cloned polyketide synthase genes.

The idea behind gene inactivation is to functionally inactivate the genomic copy of the cloned gene by inserting a vector backbone into the genomic copy of the cloned gene so that the open reading frame of the gene is disrupted or to replace the genomic " good " copy of the gene with a corrupted and non- functional " bad" copy of the gene. Gene replacement can also be done with a marker gene. The first approach is called gene disruption and the second approach is called gene replacement.

If functional inactivation of the genomic copy of the polyketide synthase gene results in absence of polyketide production, then the experiment has provided functional proof for the involvement of the cloned polyketide synthase gene in the production of that particular polyketide.

2.10.1 GENE DISRUPTION

To inactivate genes by gene disruption, internal gene fragments without translational start or stop sites are cloned into plasmids. This disrupted construct is then transferred into the producing strain by a suitable DNA transfer technique. In the producing strain, homologous recombination between the disrupted construct and intact chromosome takes place and results in the integration of the whole disrupted construct into the chromosome such that there is duplication of the gene, albeit non-functional. This is because one copy of the gene is truncated at the 5' end, that is, it lacks ribosomal binding site, start codon and a region coding the 5' end amino acids. Therefore this copy is unlikely to produce a functional gene product. The second copy of the gene is truncated at the 3' end and therefore lacks the stop codon as well as a region coding the 3' end amino acid. This copy would also most likely produce a non-

functional gene product, as the gene product is truncated. However in some cases, the truncated copy might still possess some residual activity, especially in large multifunctional proteins.

Due to the presence of duplicated copies of the genes, disruption mutants tend to be highly unstable and undergo homologous recombination at the duplicated region, resulting in excision out of the disrupted construct from the genome. Therefore, the disrupted mutants need to be grown in antibiotic selection medium so as to prevent the integrated disrupted construct from excising out. Excision of the disrupted construct will result in the restoration of the chromosomal gene. Thus, it is important to grow the disruptants under antibiotic selection pressure, to maintain gene inactivation.

2.10.2 GENE REPLACEMENT

In order to obtain stable gene-inactivated mutants, gene replacement would be the preferable method of choice. In gene replacement, the intact chromosomal copy of the gene is replaced in part or in whole with a defective gene or an antibiotic resistance gene by a double crossing over event. As gene replacement does not result in duplication of genes, disrupted mutants obtained by gene replacement are more stable than disrupted mutants obtained by gene disruption. However, the efficiency of gene replacement is considerably lower that gene disruption as the later only involves one crossing over event.

2.10.3 GENE DISRUPTION VECTORS

For convenient DNA transfer and subsequent selection of recombination events resulting in gene disruption or replacement, several vector systems have been established for *Streptomyces*. They include 1) replicative plasmids, 2) phage

derivatives, 3) non-replicative suicide plasmids and 4) temperature sensitive plasmids. Non-replicative suicide plasmids and temperature sensitive plasmids are commonly used vectors.

Replicative plasmids are *E. coli* and *Streptomyces* shuttle plasmids capable of replicating autonomously in *Streptomyces*. One example of replicative plasmids is pWHM3, which is a pIJ101 derivative lacking the minus origin (Vara *et al.*, 1989). Very often, such shuttle plasmids are only maintained under selection and are lost at a high frequency when the plasmids replicates by rolling circle mode.

Phage vectors are vectors of phage origin as the name indicates. Such vectors are integrative and integrate into the chromosome by homologous recombination if they share homologous region with the chromosome. Several derivatives vectors that lack *att* site have been developed from the actinophage ϕC31 (Bruton *et al.*, 1991).

Non-replicative plasmids

For convenient DNA transfer and subsequent selection of recombination events resulting in the gene disruption or gene replacement of chromosomal genes, non - replicative suicide vectors are used. *E. coli* vectors not capable of replicating autonomously in *Streptomyces* and carrying a marker gene that can be selected in S*treptomyces,* can be used as a non-replicative plasmid. However, success of such experiment is often limited by the poor transformation efficiencies caused by the potent restriction systems of *Streptomyces* strains. To overcome this barrier, single stranded DNA of the disruption construct is used for transformation. Single stranded DNA, used for transformation or subsequent integration into the chromosome is up to 100 times more effective than double stranded DNA. Another way to overcome this barrier is to prepare the DNA in methylation deficient *E.coli* strain such as *E.coli*

ET12567, and to introduce the disruption vector by intergeneric conjugation (MacNeil, 1992).

Temperature sensitive plasmids

Temperature sensitive plasmids represent the most successful suicide vector systems for *Streptomyces.* Plasmids are maintained stably at permissive temperature, to allow for recombination and integration of plasmid into the chromosome to occur. Then the recombinants are conveniently selected by increasing the temperature to the non-permissive temperatures. Derivatives of the naturally occurring temperature sensitive plasmid pSG5 from *Streptomyces ghanaensis* DSM2932 (Muth *et al*., 1988) have been widely used. Plasmid pSG5 replicates stably at 35ºC but not at an elevated temperature of 37ºC.

The vector of choice for gene inactivation depends on many factors such as the availability of the vectors, the size of DNA fragment to be replaced or disrupted, protoplasting efficiency of the *Streptomyces* strain, temperature sensitivity of the *Streptomyces* strain, restriction system of the *Streptomyces* strain, to name a few.

2.10.4 DNA MANIPULATION IN GENE DISRUPTION

In order to transfer polyketide synthase genes into homologous or heterologous *Streptomyces* host so as to either disrupt the genes or express the genes, DNA transfer techniques are necessary. In general conjugation and transformation are the two most common techniques used. Transducing phages, electroporation and electroduction are other available methods, which are not discussed in this review.

Conjugation

Conjugation is the only way to transfer very large fragments between *Streptomyces* strains. As in other bacteria, self-transmissible plasmids are responsible for conjugation in streptomycetes. Such plasmids encode their own transfer functions or they have to be provided with transfer functions in trans. Generally, this method of gene transfer is used for expression studies rather than for gene disruption.

Intergeneric Conjugation

Intergeneric conjugation has proved to be a convenient method to transfer DNA into *Streptomyces* in gene disruption/ replacement studies. Intergeneric conjugation involves conjugation between *E. coli* and *Streptomyces* (Mazodier *et al.*, 1989, Bierman *et al*., 1992, Flett *et al*., 1997). The protocol for intergeneric conjugation from *E. coli* to *Streptomyces* does not require any strain specific optimisation of protoplasting conditions etc.

Intergeneric conjugation uses the broad host range transfer system of IncP plasmid RK2. The mobilizable vector carries the oriT region of RK2, which allows for replication in *E.coli*. The vector does not carry the genes for transfer functions and therefore requires the transfer functions to be supplied in *trans* by the *E. coli* donor strain. The methylation deficient *E.coli* strain ET12567 is a commonly used donor strain (MacNeil, 1988), which carries a plasmid pUB307. The plasmid pUB307 is a derivative of RP1 (Richmond, 1976), which encodes the transfer function, *tra*. Upon transfer of the conjugation compatible plasmid construct from *E. coli* into the *Streptomyces* strain, the plasmid construct can integrate into the homologous region of the chromosome by homologous recombination.

Polyethlene glycol (PEG) protoplast transformation

PEG mediated introduction of DNA into protoplasts is the standard procedure for transfer of naked DNA into *Streptomyces. Streptomyces* mycelium grown in 0.5%-1.0% glycine can be treated with lysozyme to protoplast it. PEG is known to mediate the efficient transfer of DNA into *Streptomyces* protoplasts (Bibb *et al*., 1978). Transformation efficiency has been reported to vary with different suppliers and therefore requires optimisation. Transformation efficiency also depends on the *Streptomyces* host strain. Following PEG mediated transformation, protoplasts are allowed to regenerate in isotonic media and transformants are selected by overlaying with the desired antibiotic. Upon transfer of the naked DNA into the *Streptomyces* strain by PEG mediated transformation, the plasmid construct can integrate into the homologous region of the chromosome by homologous recombination or replicate autonomously depending on the type of plasmid used. Generally, non-replicative suicide plasmid constructs are transferred into *Streptomyces* by PEG mediated transformation for gene disruption studies.

3.1 PREPARATION OF ORGANISMS

3.1.1 STREPTOMYCES

STREPTOMYCES CULTURES ON AGAR

S. hygroscopicus ATCC 29253 *, S. ascomyceticus* ATCC 55098 and the soil isolate 98- 62 were grown and maintained on ISP2 or oatmeal agar at 30°C. To obtain uniform cultures on agar surface, the organisms have to be inoculated throughout the entire surface of the agar. Using a wet loop to collect spores/ mycelium from existing cultures, which is then used to streak the entire surface of the agar plate, does this. Point inoculation does not yield a good culture as the colonies only spread a limited distance within a reasonable time of incubation.

GROWTH OF STREPTOMYCES MYCELIUM FOR ISOLATION OF TOTAL DNA

For purposes such as total DNA extraction, *Streptomyces* organisms are grown as mycelium in liquid culture from an inoculum of spores. *Streptomyces* is highly aerobic and therefore requires shaking during incubation in vessels that allows for good aeration. For example, a 25 ml culture is grown in a 250 ml or 500 ml conical flask. Preinoculum medium is generally 10 ml of ISP2. Preinoculum is grown at 28°C, 200 rpm for 1 - 2 days, depending on the growth.

GROWTH OF STREPTOMYCES FOR CONJUGATION

The soil isolate 98- 62 was used for conjugation experiment. In order to obtain a good spore suspension of soil isolate 98- 62, the soil isolate was grown on AS-1 agar for 2 days at 28°C, until the surface of the plate looks grey. Spore suspension preparation is as described below.

MAKING A STREPTOMYCES SPORE SUSPENSION

To obtain a good spore suspension, the surface of the sporulating agar culture is scraped with a wet loop and the spores are suspended in sterile water in a universal bottle. The crude suspension is vortexed for 2 min and filtered through cotton wool to remove mycelial fragments and pieces of agar medium. The spores are then pelleted by centrifuging at 8000 rpm for 10 min at 4°C, in order to remove compounds dissolved from the growth medium. The spores are resuspended in LB medium and counted using a hemocytometer and resuspended in LB medium such that the final spore density is $3x10^8$spores per 100 µl LB. For each intergeneric conjugation reaction, $3x10^8$ cells of spores per 100 µl of LB were used. Spore suspension of the soil isolate sp 98- 62 was prepared in the described manner for use in intergeneric conjugation.

GROWTH OF STREPTOMYCES FOR COMPOUND EXTRACTION

In order to extract antifungal compounds from culture broths, *Streptomyces* sp. or the soil isolate 98- 62 were inoculated into 10 ml ISP2 media, and incubated at 28° C, 200 rpm for 24 h. 500µl of this preinoculum was then used to inoculate into 25 ml of FK media in a 250ml flask and incubated at 28°C ,200 rpm for a further 4 days.

PRESERVATION OF STREPTOMYCES STRAINS

Streptomyces strains were maintained by subculturing periodically onto ISP2 or oatmeal agar plates. Short time storage was at 4°C for 3-4 weeks. For long-term storage, liquid cultures of *Streptomyces* strains or the soil isolate 98- 62 in ISP2 broth were stored in a equal amount of 50% glycerol at -80°C.

3.1.2 ESCHERICHIA COLI
GROWTH OF E. COLI FOR PLASMID ISOLATION

Transformed *E. coli* DH5α and Top 10 were maintained on LB agar supplemented with appropriate antibiotics. For mini prep isolation of plasmid DNA, single isolated colonies of transformed *E. coli* was inoculated into 5 ml LB with appropriate antibiotic selection and allowed to grow at 37°C, 200 rpm, overnight. Cells were harvested by centrifugation the following day for plasmid isolation.

PREPARATION OF COMPETENT E. COLI CELLS
SOLUTIONS FOR THE PREPARATION OF COMPETENT E. COLI CELLS

i) 0.1 M Calcium chloride

$CaCl_2$ 11.1g/ l

ii) 0.1M Magnesium chloride

$MgCl_2$ 20.3g/ l

Both the solutions were autoclaved at 15 psi for 15 min and stored at 4°C.

E. coli DH5α and Top 10 was used as a host for transformation and for preparation of plasmid DNA used in this study according to the method of Cohen *et al* (1972) with modifications.

A single colony of *E. coli* was precultured in 5 ml of LB broth in a universal bottle at 37°C at 200 rpm overnight. On the following day, a 0.8 ml of the overnight culture was transferred to 40 ml LB broth in 250 ml conical flask and grown at 37°C, 200 rpm for another 2hours. The flask was then left to stand on ice for 10 min before transferring the culture into a 50 ml centrifuge tube.

Cells were harvested at 8000 rpm, 10 min at 4°C. The supernatant was discarded and *E. coli* cell pellet was resuspended in 4 ml of ice- cold 0.1 M $MgCl_2$. This was centrifuged again at 8000 rpm, 10 min at 4°C. The supernatant was discarded

and *E. coli* cell pellet was resuspended in 4 ml of ice- cold 0.1 M $CaCl_2$. After a final centrifugation, the pellet was resuspended in 0.8 ml of ice cold 0.1 M $CaCl_2$ solution and left on ice for 1 h to obtain transformation competent *E. coli* cells.

Aliquots of 40 µl competent cells with 40 µl of 97% glycerol were stored at -80°C for later use. Each tube was used for a single transformation reaction after thawing out. For immediate use of competent cells for transformation, 40 µl of fresh competent cells were used as described in the later section.

PREPARATION OF COMPETENT CELLS OF E. COLI ET12567 FOR ONE STEP TRANSFORMATION
SOLUTIONS FOR THE PREPARATION OF COMPETENT E. COLI CELLS

i) 2X YT medium

Tryptone	16g/ l
Yeast extract	10g/ l
Sodium chloride	5g/ l

This was autoclaved at 15 psi for 15 min.

ii) 2X TSS medium, pH 6.5

Tryptone	16g/ l
Yeast extract	10g/ l
Sodium chloride	5g/ l
PEG 6500/8000	100g/ l

The pH of this was medium was adjusted to pH 6.5 and medium was autoclaved at 15 psi for 15 min. After autoclaving the following solutions were added to 10 ml of 2 X TSS

1M $MgCl_2$	200µl
DMSO	500µl

E. coli ET12567 was grown in 5ml of 2X YT medium supplemented with

kanamycin and chloramphenicol overnight at 37°C, 200 rpm. The following day, 500 µl of the preculture was used to inoculate a fresh 100 ml of 2X YT medium. The culture was grown at 37°C, 200 rpm for 3-4 h until the absorbance at 600nm reached 0.3. The cells were harvested by centrifugation at 5000 rpm for 10 min and resuspended in 2ml of ice cold 2 X TSS. Aliquots of 100 µl were frozen in liquid nitrogen and stored at -80°C for later use.

GROWTH OF TRANSFORMED E. COLI ET12567 FOR INTERGENERIC CONJUGATION

E. coli ET12567 transformant was grown in 5ml of LB medium supplemented with kanamycin ,chloramphenicol and apramycin, overnight at 37°C, 200 rpm. The following day, 500 µl of the pre-culture was used to inoculate a fresh 5ml of LB medium supplemented with kanamycin, chloramphenicol and apramycin, at 37°C, 200 rpm, for 1-2 h until the absorbance at 600nm reached 0.3.The cells were then counted using a hemacytometer and resuspended in LB medium such that the final cell density is $1x10^8$ cells per 100 µl LB. For each intergeneric conjugation reaction, $1x10^8$ cells of *E. coli* ET12567 transformant per 100 µl of LB were used.

PRESERVATION OF E. COLI CULTURES

The bacterial strains were maintained by subculturing periodically onto LB agar, with antibiotic selection when necessary. Short time storage was at 4°C for 3-4 weeks. For long-term storage, liquid cultures of *E. coli* strains in LB broth, with antibiotic selection when necessary, were stored in an equal amount of 97% glycerol at -80°C.

3.1.3 ASPERGILLUS NIGER

GROWTH OF ASPERGILLUS NIGER FOR BIOASSAY ON TLC

Aspergillus niger were grown and maintained on Sabouraud agar (oxoid) at 28°C.

PRESERVATION OF ASPERGILLUS NIGER

Spores of *Aspergillus niger* from a confluent plate of SAB agar were collected with a wet loop and resuspended well in 1ml sterile water. Short time storage of spore suspension was at 4°C for 3 - 4 weeks. For long-term storage, spore were stored in an equal amount of 97% glycerol at -80°C.

3.2 PREPARATION OF CHROMOSOMAL AND PLASMID DNA

3.2.1 ISOLATION OF STREPTOMYCES TOTAL DNA

SOLUTIONS FOR ISOLATION OF ACTINOMYCETE TOTAL CHROMOSOMAL DNA

i) TS buffer, pH 8.0

50mM Tris-HCl	7.88g/ l
0.7M Sucrose	256.73g/ l

This was adjusted to pH 8.0, before autoclaving at 10 psi for 10 min.

ii) Lysozyme solution

This solution was prepared fresh just before use, by addition of 50 mg lysozyme

(Sigma) to 1 ml of TS buffer, pH 8.0. The solution was filter sterilized using a 0.22µM disposable filter unit.

iii) Proteinase K

Proteinase K (Sigma) was dissolved in sterile water at 10mg/ ml and filter

sterilized using a 0.22μM disposable filter unit.

iv) Phenol: chloroform: isoamylalcohol

Buffer saturated phenol, pH 6.7 ± 0.2 (Sigma), chloroform (Merck) and isoamyl alcohol (Ajax Chemicals) were mixed in a ratio of 25: 24:1 and then allowed to separate slowly. Phenol: chloroform: isoamylalcohol were stored in aluminium foil covered bottle at 4°C.

The cells were harvested at 8000 rpm for 10 min at 4°C in a 50 ml centrifuge tube. 0.5 g of the cell pellet was first washed in 5 ml TS buffer before being resuspended in 6 ml of the same buffer. 0.6 ml of freshly prepared lysozyme solution and 1.2 ml of 0.5 M EDTA were added into the cell suspension. The suspension was incubated with slight agitation in a 37°C water bath for 1 h. Then 0.6 ml of proteinase K (2mg/ml) was added to the mixture which was incubated at 37°C for a further 15 min. A 3.6 ml of 3.3% SDS was then added to this and the mixture was incubated first at 70°C for 15 min and then at 37°C for 1 h. To this, an equal volume of phenol: chloroform: isoamylalcohol was added in a 50 ml Teflon tube and the contents were mixed gently by inverting the tubes 40 to 50 times. The tube was then centrifuged at 12,000 rpm for 10 min at 4°C. The top aqueous layer containing the chromosomal DNA was removed and transferred into a clean tube. To this, three volume of ice cold absolute ethanol was added to precipitate the chromosomal DNA. Precipitated chromosomal DNA was spooled with a pasteur pipette, and washed in 70% ethanol and air-dried. Semi-dried chromosomal DNA was dissolved in 800μl of sterile water in a 2 ml screw cap microfuge tube. RNase A was added to a final concentration of 50μg/ml to the dissolved DNA and this was incubated at 65°C for 1 h, following which an equal volume of phenol: chloroform: isoamylalcohol was added and mixed well. This was centrifuged at 12, 000 rpm for 10 min. The top aqueous layer was removed

and transferred to a new microfuge tube. The chromosomal DNA was re- precipitated with 3 volumes of ice cold absolute ethanol, spooled, washed in 70% ethanol, air dried and dissolved again in 200 to 500µl of sterile water.

3.2.2 PLASMID ISOLATION FROM E. COLI

Plasmid isolation was performed using Promega Wizard® *Plus* SV Minipreps DNA purification Kit according to the manufacturer's recommendation.

3.2.3 SPECTROPHOTOMETRIC DETERMINATION OF DNA

DNA samples were diluted 100 times in sterile water and placed in quartz cuvettes (Hellma). The absorbance at wavelengths of 260 nm and 280 nm were determined on a spectrophotometer (LKB Biochrom Ultrospec II). Taking an absorbance of 1 unit at 260 nm to be equivalent to 50µg/ml of double stranded DNA, the concentration of DNA samples were calculated, taking into account the dilution factor as well. The ratio of the absorbance at 260 nm to 280 nm gives an indication of the purity of the DNA sample. A ratio of 1.8 indicates pure double stranded DNA. A value significantly greater than 1.8 indicates RNA contamination, while a ratio significantly lower than 1.8 indicates protein contamination.

3.2.4 AGAROSE GEL ELECTROPHORESIS OF DNA
BUFFERS AND STOCK SOLUTIONS FOR AGAROSE GEL ELECTROPHORESIS

i) 10 X Gel loading buffer

10mM Tis-HCl	0.1ml of 1M Tris-HCl, pH 7.5
20mM EDTA	0.4ml of 0.5M EDTA, pH 7.5
40% Glycerol (v/v)	4.0ml of glycerol

The volume was made up to 10 ml with sterile water. Tiny amounts of bromophenol

blue (Sigma) and Xylene Cyanol EF (Sigma) dyes were added to the mixture.

ii) 10X Tris- acetate/ EDTA (TAE) buffer, pH 8.0

0.4 M Tris base (Promega)	48.44g
Glacial acetic acid	11.42ml
0.01M EDTA	3.72g or
	20ml of 0.5 EDTA, pH 8.0

The pH was adjusted to 8.0 using glacial acetic acid before making up to 1 litre with distilled water. The buffer was autoclaved at 15 psi for 15 min. The working concentration was 1 X

iii) Ethidium bromide (EtBr) stock solution

EtBr (Sigma) was dissolved in sterile water at a concentration of 10mg/ml.

Agarose gel electrophoresis was carried out in submerged horizontal agarose gel tanks (Hoefer Scientific instruments). Agarose (Hispanagar D1 LE) was dissolved in 1X TAE buffer, pH 8.0, at a concentration of 1.0% (w/v), by heating in a microwave oven.

For gel electrophoresis of restricted chromosomal DNA in preparation for Southern transfer, 0.7% (w/v) gel was prepared. Molten gel was then cooled to 50°C. When required, RNase A was added to the gel solution at a final concentration of 1μg/ ml to remove RNA contamination from chromosomal DNA samples. 1 μl of 10mg/ml ethidium bromide was then added to 40 ml of molten gel and mixed well before casting the gel.

Once the gel had set, the comb was removed gently and the gel was transferred to the electrophoresis tank and submerged in 1X TAE buffer, pH 8.0. 2μl of gel loading buffer was added to 18 μl of DNA sample and loaded in the wells. The DNA fragments were separated by electrophoresis at a constant voltage of 80V. If the

separated DNA fragments were to be transferred onto nitrocellulose filter, electrophoresis was carried out at 15 V to allow for better resolution in separation. The mobility of the DNA fragment is inversely proportional to the logarithm of its molecular weight. Electrophoresis was terminated when the bromophenol blue dye front reached the edge of the gel.

The size of the separated fragments was determined by comparing the mobility of the fragments with the standard marker fragments. The gel was viewed and photographed under ultraviolet light from a UV transilluminator (UVP, Inc. TM- 36) using a Polaroid MP4 camera (model 4-32), fitted with a red filter and Polaroid T665 instant film.

Preparative agarose gels containing DNA fragments required for cloning were observed and photographed only under a long range UV transilluminator (365 nm) to minimize damage to the DNA. DNA bands of interest were excised from the gel using alcohol- flamed cover slips and eluted out of the gel using Geneclean II® (Bio 101 Inc., la Jolla, CA).

3.3 IN VITRO MANIPULATION OF DNA AND CLONING

3.3.1 RESTRICTION OF DNA

Restriction enzymes (5 to 20 Units/ µl) were from New England Biolabs (NEB). They were used with recommended buffers supplied by the manufacturer.

3.3.2 ALKALINE PHOSPHATASE TREATMENT

Restriction enzyme digested vector DNA fragments with compatible cohesive ends were treated with calf intestinal alkaline phosphatase (CIAP) to prevent re circularization of the vectors during ligation. 1µl of (1U) CIAP was added to the linearized vector in a final concentration of 1X CIAP buffer. The reaction was

incubated at 37°C for about 2 h. The CIAP treated vector was gel electrophoresed and recovered from the gel.

3.3.3 RECOVERY OF DNA FRAGMENTS FROM GEL USING THE GENECLEAN II® KIT

This protocol was slightly modified from that of the manufacturer and was used for the recovery of DNA from agarose gels for cloning purpose or as hybridisation probes. This kit is convenient for the purification of DNA fragments with sizes between 200 bp and 20 kb. The Geneclean II® kit was obtained from Bio 101 Inc., La Jolla, CA).

To recover DNA from an agarose gel, the gel slice was weighed. A volume of 6M sodium iodide, equivalent to three times the weight of the gel slice, was then added to the gel slice in a microfuge tube and incubated at 55ºC to melt the gel slice. 5 to 7µl of Glassmilk® was then added, the contents were inverted to mix well and incubated on ice for 5 min to allow the binding of Glassmilk to DNA. This was followed by centrifugation at 12,000 rpm for 10sec.

The Glassmilk- DNA pellet was washed in 400µl of NEW wash buffer three times. After the final wash, traces of remaining buffer were removed before resuspending in 20µl sterile water. This was then incubated at 55°C for 5 min to elute the DNA from the Glassmilk. The eluted DNA was separated from the Glassmilk by centrifugation at 12,000 rpm for 30s. The supernatant containing the eluted DNA was transferred to a clean microfuge tube.

3.3.4 LIGATION

An approximate molar ratio of 3: 1 of insert DNA to vector was used for each ligation in a total volume of 20 µl. Ligation was carried out overnight at 16°C in a

multi temp (LKB Bromma 2219 MultitempII Thermostatic Circulator). 3µl of the ligation mixture was used to transform *E. coli* cells

3.3.5 pGEMT® - T EASY VECTOR SYSTEM

This cloning system was obtained from Promega Corp., Madison, USA. This vector was used for the cloning of the PCR products especially those without a tag on the restriction site of the PCR products. The reaction was performed according to manufacturer's recommendation.

3.3.6 TRANSFORMATION AND SELECTION OF COMPETENT DH5α OR TOP10 E. COLI CELLS

SOLUTIONS FOR THE TRANSFORMATION OF COMPETENT DH5α OR TOP10 E COLI CELLS

i) 0.1 M Calcium chloride

$CaCl_2$	11.1 g/ l

This was autoclaved at 15 psi for 15 min and stored at 4°C.

ii) 100mM isopropyl- β- D- thio- galactopyranoside (IPTG)

IPTG (Promega)	0.24g/ 10 ml

This was filter sterilized using a 0.22µM disposable filter unit.

iii) 5-Bromo-4-chloro-3-indolyl- β-D-galactoside (X-gal)

X-gal (Bio Rad)	50 mg/ ml

X-Gal was dissolved in N, N'dimethyl formamide (DMF).

Transformation reaction was set up as follows: 80µl of ice cold 0.1 M $CaCl_2$ solution was pipetted into 80 µl of thawed out competent cells. To this 1 µl of plasmid or 3 µl of ligation mixture was added. This was kept on ice for 20 min and heat shocked at 42°C for 90 sec, followed by incubating in ice for a further 3 min.

900µl of LB broth was then added to the cells and incubated at 37°C for 1 h

with shaking, to allow for recovery. 100 µl of the transformed culture were then plated onto LB agar with appropriate antibiotic selection. The plates were incubated at 37°C overnight.

If the cloning procedure involved the insertional inactivation of the *lac Z'* gene in pUC18 vector, 100µl of 100mM IPTG and 20 µl of 50mg/ml X-gal were spread on the LB agar plate with appropriate antibiotic selection and incubated at 37°C for 1 h prior to plating out the transformed culture. Recombinant plasmids inserted at the multiple cloning site of pUC18 would give white colonies whereas non-recombinant plasmids or pUC18 would give blue colonies because of the induction of the lac operon by IPTG and the subsequent conversion of the X- gal to a blue product by the functional β- galactosidase.

3.3.7 TRANSFORMATION AND SELECTION OF COMPETENT ET12567 *E. COLI CELLS*

SOLUTIONS FOR THE TRANSFORMATION OF COMPETENT ET12567 *E. COLI CELLS*

i) 2X TSS medium, pH 6.5 + 20 mM glucose

Tryptone	16g/ l
Yeast extract	10g/ l
Sodium Chloride	5g/ l
PEG 6500/ 8000	100g/ l

The pH of this was medium was adjusted to pH 6.5 and the medium was autoclaved at 15 psi for 15 min. After autoclaving, the following solutions were added to 10 ml of 2 X TSS

1M $MgCl_2$	200µl
DMSO	500µl
1M Glucose	200µl

To 100µl of thawed out or freshly prepared competent cells, 1 µl of plasmid DNA was added and the mixture was kept on ice for 30 min. To this mixture, 900µl of glucose supplemented 2X TSS medium, pH 6.5 and 20mM glucose were added and the culture was incubated at 37˚ for 1 h to recover. At the end of 1 h, 100 µl or 900 µl were then plated out onto LB agar plates supplemented with kanamycin, chloramphenicol and apramycin. Plates were incubated overnight at 37°C.

3.3.8 ANALYSIS OF RECOMBINANT CLONES

Single transformant colonies were streaked out onto antibiotics supplemented plates or inoculated into 5ml of LB supplemented with antibiotics and grown at 37°C overnight. Plasmid DNA was extracted from the culture and subjected to restriction analysis and sequencing if necessary.

3.4 INTERGENERIC CONJUGATION

3.4.1 CONJUGATION

SOLUTIONS REQUIRED FOR INTERGENERIC CONJUGATION

i) S MEDIUM

Peptone	0.4g
Yeast extract	0.4g
K_2HPO_4	0.4g
KH_2PO_4	0.4g
Water	79.5ml

This was autoclaved at 10 psi for 10 min. To the autoclaved medium, 0.5 ml of 10% $MgSO_4$ and 20.0 ml of 5% glucose was added before use.

$1x10^8$ cells of *E. coli* ET12567 transformants per 100 µl of LB and $3x10^8$ *Streptomyces* spores per 100 µl of LB were used for each intergeneric conjugation. *Streptomyces* spores in 100 µl of LB were centrifuged and resuspended in 100 µl of S

medium and this was heated at 50°C for 10 min to allow the germ tubes to form.

At the end of 10 min, the spores were mixed with 100 µl aliquot of *E. coli* ET12567 transformant and mixed well. 100 µl of this mixture was plated out onto AS-1 agar supplemented with $MgCl_2$ and incubated at 30°C overnight for 5 days. The agar was overlaid with nalidixic acid and apramycin on the second day to select for resistant conjugants.

3.4.2 SOFT AGAR OVERLAY TO SELECT FOR RESISTANT CONJUGANTS

5ml of Simple Nutrient Agar supplemented with nalidixic acid and apramycin were carefully poured onto the agar surface of AS-1 agar supplemented with $MgCl_2$ such that the overlaying agar was equally spread with no bubbles. Nalidixic acid was used to kill off the *E. coli* cells whereas apramycin was to select for *Streptomyces* conjugants.

3.4.3 ANALYSIS OF CONJUGANTS

Streptomyces conjugants were streaked out onto AS-1 agar supplemented with apramycin and grown at 30°C for 5 days to obtain single colonies. Single colonies were then used to inoculate 10 ml of ISP2 medium and grown at 30°C, 200 rpm for 2 days to be used as pre-inoculum for genomic DNA extraction and compound extraction.

3.5 TECHNIQUES USING DNA

3.5.1 SOUTHERN HYBRIDISATION

TRANSFER OF DNA FROM AGAROSE GELS TO NITROCELLULOSE FILTERS (SOUTHERN TRANSFER)

i) 0.25M Hydrochloric acid (HCl)

7.66 ml of concentrated HCl (Merck) (32.64 M) was made up to 1 litre with

distilled water. The solution was autoclaved at 15 psi for 15 min.

ii) 3M Sodium chloride (NaCl)

NaCl 175.32g/l

This was sterilized by autoclaving at 15 psi for 15 min.

iii) Denaturing solution

1 M NaOH; 1.5 M NaCl was prepared by mixing equal volumes of 2 M NaOH and 3 M NaCl.

iv) Neutralising solution

0.5 M Tris-HCl , pH 7.5 ;1.5M NaCl was prepared by mixing equal volumes of 1 M Tris- HCl, pH 7.5 and 3 M NaCl.

DNA from an agarose gel was transferred to a Hybond™ -N nylon membrane (Amersham) by means of the LKB Bromma 2016 Vacugene vacuum blotting pump, which used low pressure to vacuum transfer DNA from the gel onto the nylon membrane.

A sheet of plastic mask with a window just slightly smaller than the gel was placed over a porous support in the vacuum chamber. The nylon membrane, with the top left corner cut for orientation, was placed under the window, covering it completely. After the membrane was pre-wetted with sterile water, the gel was placed on the membrane, with the DNA side facing up. After ensuring there was no bubble or leakage, the vacuum pump was switched on to a constant suction pressure of 40 cm. H_2O. The following solutions were added to the gel in the stated order, covering the gel completely:

1) 0.25M HCl for 8 min (depurination)

2) 1.0M NaOH; 1.5M NaCl for 10 min (denaturation)

3) 0.5 M Tris - HCl, pH 7.5; 1.5 M NaCl for 8 min (neutralisation) and

4) 20 X SSC, pH 7.0 for 45 min.

After the transfer process, position of the wells was marked with a pencil. The gel was removed and checked under UV for any untransferred DNA. The membrane was dried on a blotting paper, wrapped with Saran wrap and UV cross-linked on both sides for 3 min each. DNA side was cross-linked first. The membrane was then used for DNA- DNA hybridisation or stored in a desiccator in between two blotting papers.

DNA- DNA HYBRIDISATION ON NITROCELLULOSE FLTERS

i) Prehybridisation buffer

ECL Gold Hybridisation buffer (Amersham) was prepared according to manufacturer's recommendation.

ii) Primary wash buffer

6M Urea	360 g
0.4% SDS	40 ml of 10% SDS

To this, 20 X SSC (pH 7. 0), was added to give a desired final concentration of SSC and made up to a final volume of 1 litre with distilled water.

Final concentration of SSC	Volume of 20 XSSC to use
0.5 XSSC	25 ml
0.3 XSSC	15 ml
0.1 XSSC	5 ml

iii) Secondary wash buffer

2X SSC	100 ml of 20X SSC, pH 7.0
Sterile water	900ml

PREPARATION OF PROBES

Labelling of probes was done by a non - radioactive labelling method, using ECL kit. The probes were labelled by conjugating denatured probe DNA to

horseradish peroxidase in ECL labelling mixture provided according to manufacturer's recommendation. At least 0.1µg of the DNA probe in 10 to 20 µl of sterile water was first denatured by boiling it for 10 min, and immediately cooled on ice for 5 min. the tube was pulsed briefly to collect all the contents at the bottom of the tube. An equivalent volume of DNA labelling reagent and glutaraldehyde were added to denature the DNA . The mixture was then incubated at 37°C for 15 min to label the probe. At the end of the 15 min, contents of the tube were pulsed briefly and added to the prehybridisation buffer to hybridise as described below.

PREHYBRIDISATION

The nylon blot was prehybridized in 10 ml prehybridiztion buffer for 2 h at 42ºC. The nylon membrane was placed with the DNA side up, allowing for maximal contact with the prehybridisation buffer in a hybridisation bottle and rotated in Hybaid rotisserie.

HYBRIDISATION

The labelled probe was added to the prehybridisation buffer and allowed to rotate at 42ºC overnight.

STRINGENCY WASHES

On the following day, the hybridisation buffer was decanted away carefully. The membrane was placed in a clean Tupperware with preheated 50 ml of 0.5X SSC (42ºC). The membrane was washed at 42ºC for 20 min with agitation, after which the buffer was replaced with preheated 50 ml of 0.3X SSC (42ºC). The membrane was then washed in secondary wash buffer for 5 min each time at room temperature, with

gentle agitation.

SIGNAL GENERATION AND DETECTION

Signal generation and detection was performed according to manufacturer's instructions.

3.5.2 POLYMERASE CHAIN REACTION
dNTP MIX

Appropriate amounts of 1M dATP, dTTP, dCTP, dGTP were mixed with sterile water to give a final dNTP concentration of 2.5mM. dNTPs were purchased from Promega (USA).

AMPLIFICATION OF 16S rDNA OF THE SOIL ISOLATE 98- 62
PCR PRIMERS

Forward primer: RNAFORS AAG TGA CGG TAC CTG CAG
Reverse primer: RNAREVS ACA GCC ATG CAC CAC CTG

PCR CYCLING CONDITIONS

95°C	95°C	62°C	72°C	72°C	4°C
10MIN	1MIN	45SEC	1MIN	10MIN	INF
1x		35x		1x	

AMPLIFICATION OF THE KS/AT REGION OF THE SOIL ISOLATE 98- 62
PCR PRIMERS

Forward primer: NKSFOR CGG TSA AGT CSA ACA TCG G (19)
Reverse primer: NKSREV GCR ATC TCR CCC TGC GAR TG (20)

PCR CYCLING CONDITIONS

95°C	95°C	60°C	72°C	72°C	4°C
10MIN	30SEC	45SEC	1MIN30SEC	10MIN	INF
1x		35x		1x	

SCREENING PRIMERS FOR DOWNSTREAM CLONE TO C170

Forward primer: 1.5FOR CTG CCC ACG TAT CCC TTC (18)
Reverse primer: 1.5 REV CTG GGA GGC GGG CCC GTA (18)

PCR CYCLING CONDITIONS

95°C	95°C	56°C	72°C	72°C	4°C
10MIN	30SEC	45SEC	1MIN	10MIN	INF
1x		35x		1x	

SCREENING PRIMERS FOR UPSTREAM CLONE TO C170

Forward primer: R7SFOR ATT CCT CCA CGA CGC ACC (18)
Reverse primer: R7SREV AAG TCG ATG AAG GTG TCC (18)

PCR CYCLING CONDITIONS

95°C	95°C	50°C	72°C	72°C	4°C
10MIN	30SEC	45SEC	1MIN	10MIN	INF
1x		35x		1x	

3.5.3 SEQUENCING

All the DNA sequencing reactions were performed using the ABI PRISM® BIGDYE™ Terminator Cycle Sequencing Ready Reaction Kit (PE Applied Biosystems, USA) in a geneAmp PCR system 9600 (PE Applied Biosystems, USA) according to manufacturer's recommendation.

Sequencing primers M13/ pUC18 forward primer (5' CCC AGT CAC GAC GTT GTA AAA CG 3') and M13/ pUC reverse primer (5' AGC GGA TAA CAA TTT CAC ACA GG 3') were used to sequence inserts cloned into the pUC18 vector.

SEQUENCING PRIMERS FOR CLONE C170
3.7kb of C170

Primer name	Sequence
C1706.1KBF1S	GGT GTC AAC GTG CAC GGA
C170FLR1S	ACA CCG ACG GCC TCT ACG
C1706.1KBR2S	GTC GAG GAC GCG CCG CTC
C1706.1KBR3S	CGG ATC GTC CTT GTC GGC
C1706.1KBR4S	ACT GCA CCT CGA CCG GCC
C1706.1KBR3B	AAG CCT CGC CGA CGC CGC
C1706.1KBR5S	GCC GAC CAC GAG CAC ACC
C1706.1KBR6S	ATA CGG GCG GAG CAC CTC
C1706.1KBR7S	CAT CTA CGA TCC CGA CCC
C1706.1KBFC1S	CTC CAC CTG GCC GTG CAG

1.5kb of C170

Primer name	Sequence
C1701.5KBF1S	GGC GCG GCA GTC CAG GTC
C1701.5KBF3S	CTC CAG GCC GGT CGA CCC
C1701.5KBF4S	CAG CTG GCC CTG CGC GAG
C1701.5KBF2BS	TCG AAC TCC CCC GGT GAG

2.0kb of C170

Primer name	Sequence
C1702.0KBF1S	CGA GGA CGC TGC ACG CCG
C1702.0KBF2S	GAA CTG CTC GAC GGC TCA
C1702.0KBR1S	GTC AGC GCG GTG GTG TCC
C1702.0KBR2S	GTC GAG GAC GCG CCG CTC
C1702.0RC1S	GGA GAC CGC CGA CGC CGT

S*EQUENCING PRIMERS FOR CLONE C2*

Primer name	Sequence
C2F1S	TCG ACA TCA CGG ACA CGC
C2F2S	GCG TCG TAG AGG AAT CCG
C2PR1S	GCT TCG ACC TCG CGC AGT
C2PR2S	GCG TAC GCC GTT CTG GAC
C2P3RS	CAC CTG GCC ACC GAG CAC

SEQUENCING PRIMERS FOR 2.3 kb CLONE E27

Primer name	Sequence
5.2F1S	CTC CCA CCA GGT CGA CTG
5.2F2S	CCG GGA CTG GTA CGA CA
5.2R1S	CGC TGA CGA AGG GGT GGT C
5.2R2S	GTG CCG TAC CCA GTA GTC
5.2R1CS	GCT CGG ATC GGT GCT GGT

SEQUENCING PRIMERS FOR CLONE C2

Primer name	Sequence
C2F1S	TCG ACA TCA CGG ACA CGC
C2F2S	GCG TCG TAG AGG AAT CCG
C2PR1S	GCT TCG ACC TCG CGC AGT
C2PR2S	GCG TAC GCC GTT CTG GAC
C2P3RS	CAC CTG GCC ACC GAG CAC

SEQUENCING PRIMERS FOR 16S rDNA

Primer name	Sequence
RNAF1S	AAT TAT TGG CGT AAA GAG
RNAR1S	GTC GAA TTA AGC CAC ATG

3.6 BIOCOMPUTING SOFTWARE

Purpose	Software/Website
General sequence analysis (DNA and Protein)	DNASTAR
Checking designed primer	Oligotech
Nucleotide/ aminoacid search against a database	BLAST program at http://www.ncbi.nlm.nih.gov
Open reading frame	ORF Finder at http://www.ncbi.nlm.nih.gov
Multiple sequence Alignment	http://www.ebi.ac.uk

3.7 COMPOUND EXTRACTION AND ANALYSIS

3.7.1 COMPOUND EXTRACTION

i) Chemical extraction solvent

Ethyl acetate (Merck) was used to extract rapamycin from *Streptomyces hygroscopicus* ATCC 29253, FK 506 from *Streptomyces ascomyceticus* ATCC 55098 and antifungal compound from the soil isolate 98- 62.

ii) TLC plate

Silica gel 60 F- 254 TLC plate (Merck)

CHEMICAL EXTRACTION OF ANTIFUNGAL COMPOUNDS FROM CULTURE BROTHS

In order to extract antifungal compounds from culture broths, *Streptomyces* sp. or the soil isolate 98- 62 were inoculated into 10 ml ISP2 media, and incubated at 28°C, 200 rpm for 24 hr. 500µl of this preinoculum was then used to inoculate 25 ml of FK media and incubated at 28°C, 200 rpm for 4 days.

After 4 days, an equal volume of ethyl acetate (25 ml) was added to the culture broth and allowed to mix well on a 37°C shaker for 3 h. This mixture was then transferred to a centrifuge tube and centrifuged at 8000rpm for 10 min at 4°C to separate the organic and aqueous layer. The top layer containing the chemical compounds was transferred to a round - bottomed flask and concentrated by vacuum freeze drying. 1 ml of ethyl acetate was added to dissolve the dried chemical compound. Extracted compounds were then transferred to a small glass bottle and used for further analysis by TLC. Extracted compounds were stored at 4°C.

3.7.2 THIN LAYER CHROMATOGRAPHY

i) TLC separation solvent mixture

Chloroform: Methanol (95: 5, v/v) was used to separate chemical extracts on TLC plates.

3.7.3 BIOASSAY

Spore suspension of *Aspergillus niger* in 1ml water was added to 100 ml of autoclaved warm MHA agar. For a single TLC plate, spores of *Aspergillus niger* from half an agar plate were used. This spore suspended agar was then overlaid onto taped TLC plate. The overlaid TLC plate in an aluminium foil chamber was then incubated overnight at 28°C.

3.8 BACTERIAL STRAINS AND MEDIA

3.8.1 AGAR/ LIQUID MEDIA

LURIA - BERTANI MEDIUM (LB)

Tryptone(Difco)	10g/ l
Yeast extract (BBL)	5 g/l
Sodium chloride (Merck)	10g/l

This was sterilised by autoclaving at 15 psi for 15 min.

For solid media, agar (granulated, BBL) was added at a final concentration of 1.5% (wt./ vol.) prior to autoclaving.

ISP2 MEDIUM

0.4% Yeast extract (BBL)	4g/l
1.0% Malt extract (Oxoid)	10 g/l
0.4% Glucose (Merck)	4g/l

This was sterilised by autoclaving at 10 psi for 10 min. For solid media, agar (granulated, BBL) was added at a final concentration of 1.5% (wt./vol.) prior to autoclaving.

R2YE MEDIUM

10.3% Sucrose (BDH)	103g/l
0.025% K_2SO_4 (BDH)	0.25g/l
1.012% $MgCl_2.6H_2O$ (Merck)	10.12g/l
1.0% Glucose (Merck)	10.0g/
0.01% Casamino acid (Difco)	0.1g/l
0.2% Trace element solution	2ml/l
0.5% Yeast extract (BBL)	5g/l
0.573% TES buffer (Sigma)	5.73g/l

This was sterilised by autoclaving at 10 psi for 10 min. The following solutions , which were individually autoclaved at 15 psi for 15 min, except for L- proline, which was filter sterilized, were added to autoclaved R2YE medium before use.

0.5% KH_2PO_4 (Merck)	1.0 ml/l
5M $CaCl_2$.$2H_2O$ (Sigma)	0.4 ml/l
20% L- proline (Sigma)	1.5 ml/l
1N NaOH (Merck)	0.7 ml/l

For solid media, agar (granulated, BBL) was added at a final concentration of 1.5% (wt. / vol.) prior to autoclaving.

TRACE ELEMENT SOLUTION

$ZnCl_2$ (Merck)	40mg/l
$FeCl_3$. $6H_2O$(Merck)	200mg/l
$CuCl_3$. $2H_2O$(Merck)	10mg/l
$MnCl_3$. $4H_2O$(Merck)	10mg/l
$Na_2B_4O_7$.$10H_2O$(Merck)	10mg/l
$(NH_4)_6$ Mo_7. O_{24}. $4H_2O$(Merck)	10mg/l

This was sterilised by autoclaving at 15 psi for 15 min.

FK MEDIUM

Glucose	45g/l
Corn steep liquor	10g/l
Yeast extract	10g/l
Corn starch	10g/l
Cotton seed meal	10g/l
$CaCO_3$	1g/l

This was sterilised by autoclaving at 10 psi for 10 min.

OATMEAL AGAR

Oatmeal agar (Oxoid)	72.5g/l

This was sterilised by autoclaving at 15 psi for 15 min.

AS- 1 AGAR

Yeast extract	1g/l
Soluble starch	5g/l
Sodium chloride	2. 5g/l
Sodium sulphate	10 g/l
Agar	20g/l
Arginine (0.1g/ml)	2ml
Alanine (0.1g/ml)	0.8 ml

This was sterilised by autoclaving at 10 psi for 10 min. Magnesium chloride or antibiotics were added to the agar after autoclaving and cooling down to 50˚C.

SNA AGAR

Simple nutrient broth	13g/l
Agar	3g/l

This was sterilised by autoclaving at 15 psi for 15 min.

MULLER - HINTON AGAR (MHA)

MHA (Oxoid)	38g/l

This was sterilised by autoclaving at 15 psi for 15 min.

SABOURAUD AGAR

SAB (Oxoid)	38g/l

This was sterilised by autoclaving at 15 psi for 15 min.

3.8.2 ANTIBIOTIC CONCENTRATIONS

Antibiotic	Concentration of stock solution (mg/ ml)	Final concentration (µg/ ml)
Ampicillin	100	10
Apramycin	100	10
Chloramphenicol	15	25
Kanamycin	50	50
Nalidixic acid	100	10

3.8.3 STRAINS OF STREPTOMYCES, E. COLI AND ASPERGILLUS NIGER USED

Strain	Genotype	Phenotype	Use	Source
Escherichia coli DH5α	F^- ϕ80d lacZ ΔM15 (lacZyA - argF)U169 deoR recA1 endA1 hsdR1 (r_k^- m_k^+) supE44λ^- thi-1 gyrA96 relA1	Ampicillin sensitive	General cloning	Bethesda Research Laboratories
Escherichia coli Top10	F^- mcrA Δ(mrr- hsdRMS-mcrBC) ϕ80lacZΔM15 ΔlacX74 deoR recA1 araD139 Δ(are-leu) 7697 galU galK rpsL (Str^R) end A1 nupG	Ampicillin sensitive	General cloning	Invitrogen
Escherichia coli ET12567		Methylation deficient strain	Intergeneric conjugation	Dr Fiona Flett and Dr Colin Smith, UMIST
Streptomyces ascomyceticus ATCC55098		FK 506 producer	Positive control for PKS I genes	American Type Culture Collection (ATCC)
Soil Isolate 98-62 from Singapore		Novel antifungal compound producer	Source of PKS I gene for this study	Laboratory of A/P Nga B. H., Dept of Microbiology, NUS
Aspergillus niger			Test organism for antifungal compound	Laboratory of A/P Nga B. H., Dept of Microbiology, NUS

3.8.4 PLASMIDS USED

Plasmid	Characteristics	Source/Reference
pUC18	Carries β- lactamase gene conferring ampicillin resistance (Amp^r); bacterial origin of replication (ori); *E. coli* lac I' OPZ'; α–peptide of the β–galactosidase gene (lacZ') at its multiple cloning site (MCS) which allows for blue /white selection. Recombinant clones are white on IPTG and X-gal selection .	Bethesda Research Laboratories Yanisch-Perron *et al* (1985)
pSOK201	apramycin resistance(Ap^r); bacterial origin of replication (ori);	Sergey Zotchev *et al* (2000)

3.8.5 PROBES USED

DNA probe	Fragment size	Source	Reference
PKS – I probe	1.4kb	KS2 of ery gene cluster	Dr Soong Tuck Wah. IMCB, Professor CR Hutchinson, University of Wisconsin
PKS-I gene probe 1 from the soil isolate 98 -62	850 bp	KS-AT gene	This study
PKS-I gene probe 2 from the soil isolate 98 -62	7.2kb	DH-KR-ACP-KS-AT-DH of module 1 and 2	This study
PKS-I gene probe 3 from the soil isolate 98 -62	3.7kb	DH-KR-ACP-KS of module 1 and 2	This study
PKS-I gene probe 4 from the soil isolate 98 -62	1.5 kb	DH of module 2	This study
pSOK201 vector probe	3.0 kb	Vector back bone	Sergey Zotchev *et al* (2000)

3.8.6 DNA MODIFYING ENZYMES USED

DNA modifying enzyme	Concentration	Manufacturer
Calf intestinal alkaline phosphatase (CIAP)	1 unit/ µl	Promega
T4 DNA ligase	1 unit/ µl	BRL

DNA modifying enzymes were used with recommended buffers supplied by the manufacturer.

3.8.7 DNA SIZE STANDARDS

Marker	Concentration	Supplier
λHindIII	0.5µg/ µl	Promega
1kb ladder	1 µg/ µl	BRL
100bp plus	1.0µg/ µl	Fermentas

3.8.6 COMMON SOLUTIONS AND BUFFERS
COMMON SOLUTIONS

i) 2 M Sodium hydroxide (NaOH)

NaOH pellets (Merck) 80.0 g/l

ii) 10% Sodium dodecyl sulphate (SDS)

SDS (Merck) 100.0 g/l

iii) 0.5 M Ethylenediamine tetraacetate (EDTA), pH 7.5 or pH 8.0

$Na_2EDTA.\ 2H_2O$(Sigma) 186.1g/l

iv) 1M Tris (hydroxymethyl) aminomethane hydrochloride (Tris- Hl), pH 7.5 or pH 8.0

Tris- HCL (Sigma) 157.6g/l

v) 20XSSC, pH 7.0

3.0 M Sodium chloride	175.3 g/l
0.3 M Sodium citrate	88.2 g/l

vi) 50% Glycerol

Glycerol	50 ml
Distilled water	50 ml

Where necessary, the pH of each solution was adjusted to the desired one, followed by autoclaving at 15 psi for 15 min, except for 10% SDS, which was filter sterilized using a 0.22μm disposable filter unit.

vii) Ribonuclease A (RNaseA)

RNase A (Sigma) was dissolved in sterile water at a concentration of 10mg/ ml. This was then boiled for 15 min, cooled to room temperature and stored at - 20°C.

4.1 IDENTIFICATION OF THE STREPTOMYCES SP. *98- 62*

4.1.1 POLYMERASE CHAIN REACTION OF 16S rDNA FROM THE STREPTOMYCES SP. *98-62*

In order to identify novel antibiotics produced by microorganisms, random screening of indigenous soil microorganisms has been widely carried out. Selective methods for detecting and identifying these microorganisms are needed in order to gain an in depth knowledge of the organism. Actinomycetes are well known organisms that are responsible for producing a number of bioactive compounds such as antibiotics.

A number of methods such as morphological study, study of cell wall peptidoglycan has been instrumental in identifying and classifying the *Streptomyces* sp. A promising method for selective identification of soil bacteria is the amplification of 16S ribosomal DNA or ribosomal RNA using PCR.

Sequence comparisons of small subunit rRNA have been used as a source for determining phylogenetic and evolutionary relationships among organisms of the three kingdoms Archaea, Eukarya, Bacteria. The 16S rDNA are highly conserved, sharing common three-dimensional structural element of similar function. The primary structures are well conserved and variable regions have been determined (Woese, 1987). Primers located in highly conserved regions have been published, allowing the amplification of the 16S rDNA.

A pair of primers p27f (AGA GTT TGA TCM TGG CTC AG) as the forward primer and p1492r (TAC GGY TAC CTT GTT ACG ACT T) as the reverse primer were used to amplify the 16S rDNA from the genomic DNA of the *Streptomyces* sp. 98- 62. This pair of primers were designed based on the consensus sequence of bacterial 16S rDNA genes (Medlin *et al*., 1988)

An amplification product of 1500bp upon gel electrophoresis was obtained which was cloned into the pGEMT vector. The insert was sequenced using the vector primers T7 and SP6. Additional sequencing primers were designed to allow for complete sequencing of the insert. Nucleotide sequences were aligned using BLAST2 program. Searching database using BLAST program elucidated identity of the complete nucleotide sequence. The sequence of 16S rDNA from the *Streptomyces* sp. 98- 62 was determined to be 1490 bp long (Fig. 7). This nucleotide sequence was approximately 99% similar to that of the other *Streptomyces* 16S rDNA (Fig. 8).

4.1.2 SEQUENCE OF 16S rDNA FROM THE STREPTOMYCES SP. *98- 62*

agagtttgatcctggctcaggacgaacgctggcggcgtgcttaacacatgcaagtcgaacgatgaagcccttcggggtgg
attagtggcgaacgggtgagtaacacgtgggcaatctgcccttcactctgggacaagccctggaaacggggtctaataacc
ggataacactctgtcccgcatgggacggggttgaaagctccggcggtgaaggatgagcccgcggcctatcagcttgttgg
tggggtgatggcctaccaaggcgacgacgggtagccggcctgagagggcgaccggccacactgggactgagacacggccc
agactcctacgggaggcagcagtggggaatattgcacaatgggcgcaagcctgatgcagcgacgccgcgtgagggatgac
ggccttcgggttgtaaacctctttcagcagggaagaagcgcaagtgacggtacctgcagaagaagcgccggctaactacg
tgccagcagccgcggtaatacgtagggcgcaagcgttgtccggaattattgggcgtaaagagctcgtaggcggcttgtcg
cgtcggttgtgaaagcccggggcttaaccccgggtctgcagtcgatacgggcaggctagagtgtggtaggggagatcgga
attcctggtgtagcggtgaaatgcgcagatatcaggaggaacaccggtggcgaaggcggatctctgggccattactgacg
ctgaggagcgaaagcgtggggagcgaacaggattagataccctggtagtccacgccgtaaacgttgggaactaggtgttg
gcgacattccacgtcgtcg**gtgccgcagctaacgcattaagttccccgcctggggagtacggccgcaaggctaaaactca**
aaggaattgacgggggcccgcacaagcagcggagcatgtggcttaattcgacgcaacgcgaagaaccttaccaaggcttg
acatacaccggaaagcatcagagatggtgccccccttgtggtcggtgtacaggtggtgcatggctgtcgtcagctcgtgt
cgtgagatgttgggttaagtcccgcaacgagcgcaacccttgttctgtgttgccagcatgcctttcggggtgatggggac
tcacaggagactgccggggtcaactcggaggaaggtggggacgacgtcaagtcatcatgcccccttatgtcttgggctgca
cacgtgctacaatggccggtacaatgagctgcgatgtcgtgaggcggagcgaatctcaaaaagccggtctcagttcggat
tggggtctgcaactcgaccccatgaagtcggagttgctagtaatcgcagatcagcattgctgcggtgaatacgttcccgg
gccttgtacacaccgcccgtcacgtcacgaaagtcggtaacacccgaagccggtggcccaaccccttgtgggagggagct
gtcgaaggtgggactggcgattgggacgaagtcgtaacaaggtagccgta

Figure 7: 16S rDNA nucleotide sequence of the *Streptomyces* sp. 98- 62. The nucleotides in red represent the characteristic signature sequence of streptomycetes.

```
Sequences producing significant alignments:                      (bits) Value

gi|2832351|emb|Y10842.1|SSPY10842  Streptomyces sp. 16S rRNA...  2863   0.0
gi|21742834|emb|AJ494864.1|SFL494864  Streptomyces flavogris...  2859   0.0
gi|16611977|gb|AF429390.1|  Streptomyces sp. VTT E-99-1326 (...  2851   0.0
gi|16611989|gb|AF429398.1|  Streptomyces sp. VTT E-99-1334 (...  2843   0.0
gi|16611984|gb|AF429395.1|  Streptomyces sp. VTT E-99-1331 (...  2843   0.0
gi|14719240|gb|AF389344.1|AF389344  Streptomyces sp. YIM8 16...  2835   0.0
gi|733430|gb|U22972.1|SSU22972  Streptomyces sp., strain GP ...  2835   0.0
gi|733432|gb|U22974.1|SSU22974  Streptomyces sp., strain GP ...  2831   0.0
gi|16611990|gb|AF429399.1|  Streptomyces sp. VTT E-99-1335 (...  2827   0.0
gi|16611986|gb|AF429396.1|  Streptomyces sp. VTT E-99-1332 (...  2827   0.0
gi|16611980|gb|AF429392.1|  Streptomyces sp. VTT E-99-1328 (...  2827   0.0
gi|16611978|gb|AF429391.1|  Streptomyces sp. VTT E-99-1327 (...  2827   0.0
gi|14530936|gb|AY029698.1|  Streptomyces sp. KN-0479 16S rib...  2827   0.0
gi|6979922|gb|AF221837.1|AF221837  Streptomyces sp. AA8321 1...  2827   0.0
gi|733431|gb|U22973.1|SSU22973  Streptomyces sp., strain GP ...  2827   0.0
gi|5672637|dbj|AB030572.1|  Streptomyces griseus ribosomal R...  2827   0.0
gi|5672635|dbj|AB030570.1|  Streptomyces griseus ribosomal R...  2827   0.0
gi|5672633|dbj|AB030569.1|  Streptomyces griseus ribosomal R...  2827   0.0
gi|5672632|dbj|AB030568.1|  Streptomyces griseus ribosomal R...  2827   0.0
gi|5672630|dbj|AB030567.1|  Streptomyces griseus ribosomal R...  2827   0.0
gi|14582970|gb|AF361784.1|AF361784  Streptomyces sp. S63 16S...  2819   0.0
gi|5672636|dbj|AB030571.1|  Streptomyces griseus ribosomal R...  2819   0.0
gi|7715013|gb|AF112160.1|AF112160  Streptomyces caviscabies ...  2815   0.0
gi|971126|dbj|D63872.1|  Streptomyces setonii 16S ribosomal ...  2815   0.0
gi|153245|gb|M76388.1|STMDRNA  S.griseus 16S, 23S, and 5S rR...  2811   0.0
gi|14717423|gb|AF112174.2|AF112174  Streptomyces sp. EF-91 1...  2807   0.0
gi|13276861|emb|AJ308577.1|SSP308577  Streptomyces sp. Nm5 p...  2807   0.0
gi|2290506|gb|U93336.1|SSU93336  Streptomyces sp. JCM7249 16...  2807   0.0
gi|10039263|dbj|AB045872.1|  Streptomyces argenteolus gene f...  2807   0.0
gi|3550671|emb|Y15498.1|SY15498  Streptomyces sp. 16S rRNA g...  2803   0.0
gi|2290508|gb|U93338.1|SSU93338  Streptomyces sp. JCM 7250 1...  2799   0.0
gi|14717425|gb|AF112179.2|AF112179  Streptomyces sp. OB-35 1...  2797   0.0
gi|14717424|gb|AF112175.2|AF112175  Streptomyces sp. EF-93 1...  2797   0.0
gi|13276859|emb|AJ308575.1|SSP308575  Streptomyces sp. So10 ...  2791   0.0
gi|3550675|emb|Y15502.1|SY15502  Streptomyces griseus 16S rR...  2787   0.0
gi|3550674|emb|Y15501.1|SY15501  Streptomyces griseus 16S rR...  2787   0.0
gi|525283|emb|X79326.1|SO16SRN  S.ornatus (DSM 40307) 16S rR...  2775   0.0
gi|20513089|gb|AY094364.1|  Streptomyces sanglieri A-CR2 16S...  2752   0.0
gi|20513088|gb|AY094363.1|  Streptomyces sanglieri A5843 16S...  2750   0.0
gi|4079698|gb|AF012736.1|AF012736  Streptomyces sp. ASSF13 1...  2750   0.0
gi|4079700|gb|AF012738.1|AF012738  Streptomyces sp. ASSF22 1...  2738   0.0
gi|1359994|emb|X92614.1|MMM16RRNA  M.megalomicea 16S rRNA gene  2734   0.0
gi|7106037|emb|AJ399460.1|SCY399460  Streptomyces cyaneus pa...  2730   0.0
gi|10038689|dbj|AB045887.1|  Streptomyces peucetius gene for...  2724   0.0
gi|10038692|dbj|AB045890.1|  Streptomyces venezuelae gene fo...  2720   0.0
gi|1418300|emb|X87309.1|AS16SR119  Streptomycetaceae 16S rRN...  2714   0.0
gi|20513096|gb|AY094371.1|  Streptomyces griseus subsp. gris...  2712   0.0
gi|14719239|gb|AF389343.1|AF389343  Streptomyces sp. YIM26 1...  2710   0.0
gi|4079702|gb|AF012740.1|AF012740  Streptomyces sp. ASB33 16...  2700   0.0
gi|7715012|gb|AF112159.1|AF112159  Streptomyces sp. EF-52 16...  2692   0.0
gi|10038680|dbj|AB045878.1|  Streptomyces galilaeus gene for...  2682   0.0
gi|587558|emb|X80825.1|SSRDNA16S  S.subrutilus 16S rRNA gene     2678   0.0
```

Alignments

```
>gi|2832351|emb|Y10842.1|SSPY10842   Streptomyces sp. 16S rRNA gene, strain A46R62
          Length = 1476

 Score = 2863 bits (1444), Expect = 0.0
 Identities = 1462/1468 (99%)
 Strand = Plus / Plus
```

Figure 8: Sequence comparison of the 16S rDNA amplified from the *Streptomyces* sp. 98- 62 with the Genbank sequences

In the Atlas of actinomycetes (Yokota, 1997), a phylogenetic tree based on the 16S rRNA sequence of 90 genera had been drawn out to represent the phylogenetic relationship of actinomycetes. The 16S rDNA of the *Streptomyces* sp. 98– 62 was compared with that of the 16S rRNA sequences of at least one representative organisms from the various sections (all sections except section 4) of the actimomycete phylogenetic tree and represented in a phylogenetic tree using the phylip method in the ClustalW package.

The phylogenetic analysis showed that the *Streptomyces* sp. 98- 62 belongs to the genus *Streptomyces* (Fig. 9).

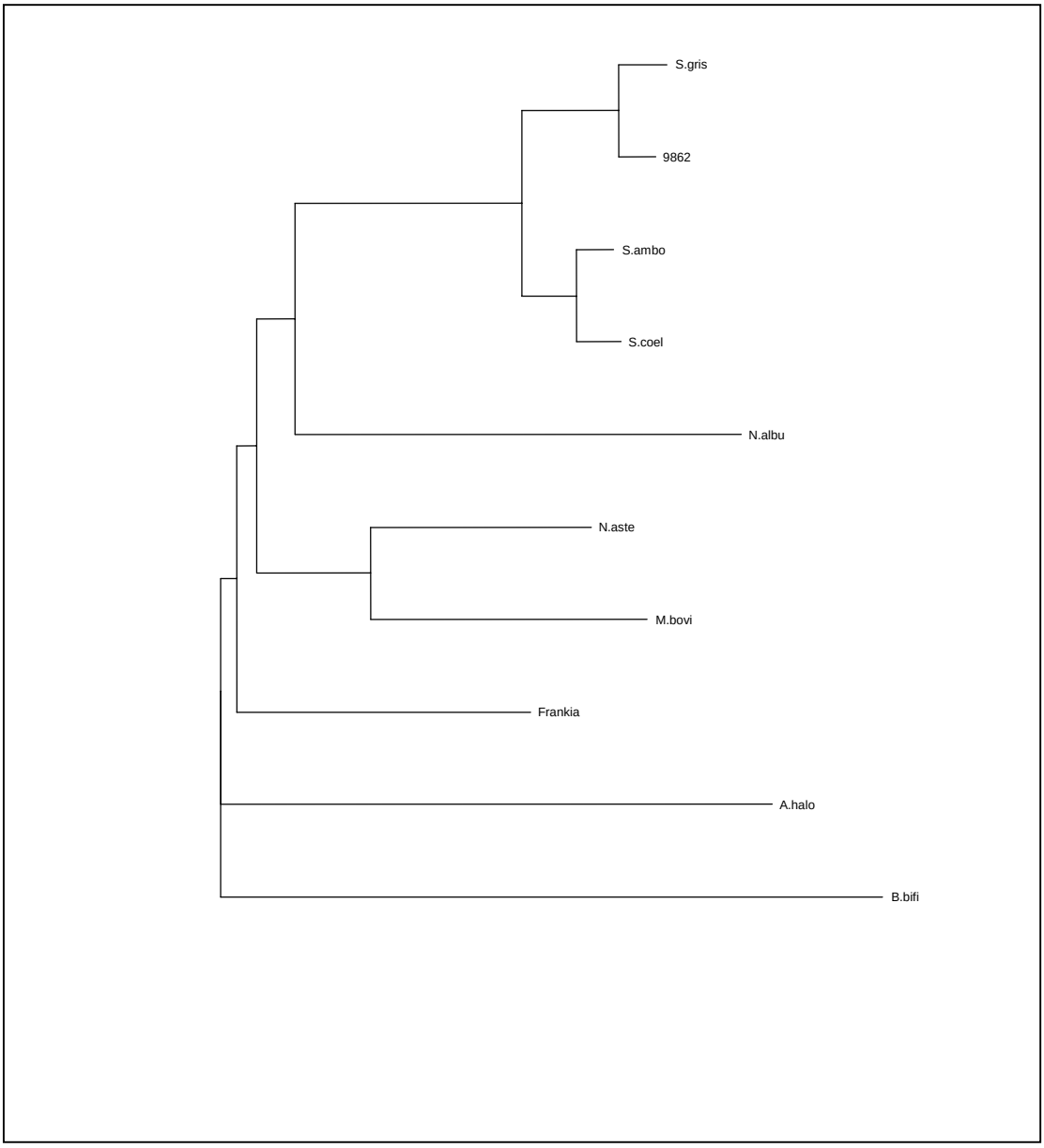


Figure 9: Phylogenetic analysis of 16S rDNA. Sequences of *Streptomyces griseus* (X61478), the *Streptomyces* sp. 98– 62, *Streptomyces ambofaciens* (M27245), *Streptomyces coelicolor* (M35377), *Nocardia albus* (X53211), *Mycobacterium bovis* (X55589), *Frankia* sp. (L11306), *Actinopolyspora halophila* (X 54287) and *Bifidobacterium bifidum* (M38018) were used for the phylogenetic analysis. The first five letters of these names are denoted in the phylogenetic tree. The tree was constructed using the CLUSTALW program. The relatedness between different actinomycetes is indicated by the length of the horizontal line. The shorter the horizontal line, the more closely related the actinomycetes. The length of the vertical lines are not significant.

4.2 PRELIMINARY EVIDENCE OF PKS I COMPOUND PRODUCTION BY THE STREPTOMYCES SP. *98- 62*

4.2.1 SCREENING FOR THE PRESENCE OF KETOACYL SYNTHASE GENE USING eryKSII GENE PROBE OF SACHHAROPOLYSPORA ERYTHRAEA

A number of antifungal polyketide compounds are synthesized by bacterial strains, by enzymes encoded by PKS I genes. As such it was postulated that the antifungal compound produced by the *Streptomyces* sp. 98- 62 could also be encoded by PKS type I genes.

In order to determine the characteristics of the compounds produced by the *Streptomyces* sp. 98- 62, genomic DNA of the *Streptomyces* sp. 98- 62 was restricted with different restriction enzymes, gel electrophoresed (Fig. 10) and Southern blotted with PKS I specific eryKSII probe from the erythromycin producer *Saccharopolyspora erythraea* (Fig. 11). *BamHI* restricted chromosomal DNA of FK506 producer *S. ascomyceticus* ATCC 55098 was used as positive control for PKS I genes.

The results showed strong hybridising bands of the genomic DNA of the *Streptomyces* sp. 98- 62 with the PKS I specific probe eryKSII, as when the genomic DNA of the positive control *S. ascomyceticus* ATCC 55098 was used. This evidence is suggestive that the antifungal compound produced by the *Streptomyces* sp. 98- 62 could be accounted for by the occurrence of PKS I specific genes in the genomic DNA of the strain.

The result has also shown that although both the *Streptomyces* sp. 98- 62 and *S. hygroscopicus var. ascomyceticus* ATCC55098 share homology with the eryKSII gene probe, they differ in the hybridization pattern obtained with the eryKSII probe.

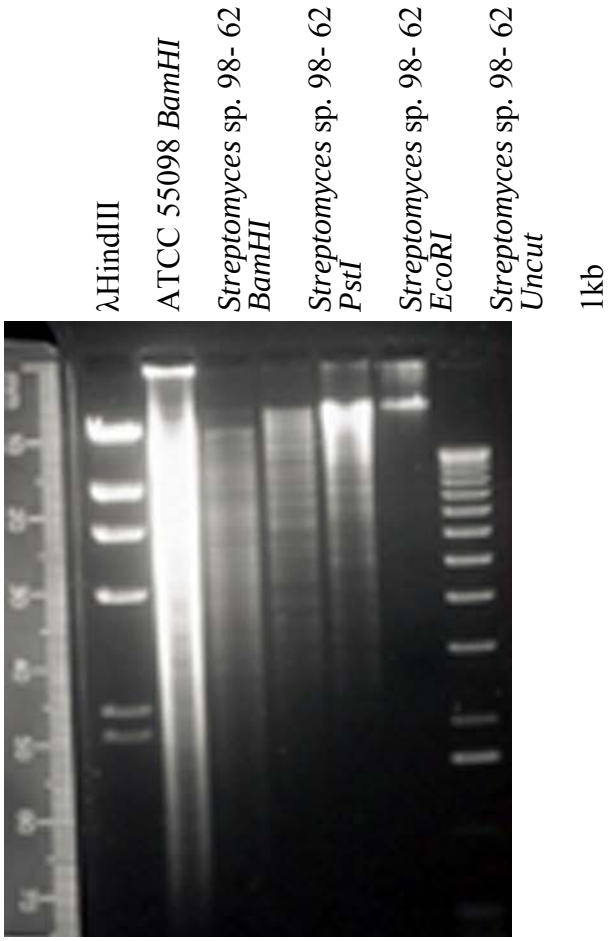


Figure 10: Electrophoretic profile of restriction endonuclease digested chromosomal DNA samples of *Streptomyces hygroscopicus var. ascomyceticus* ATCC55098 and the *Streptomyces* sp. 98– 62.

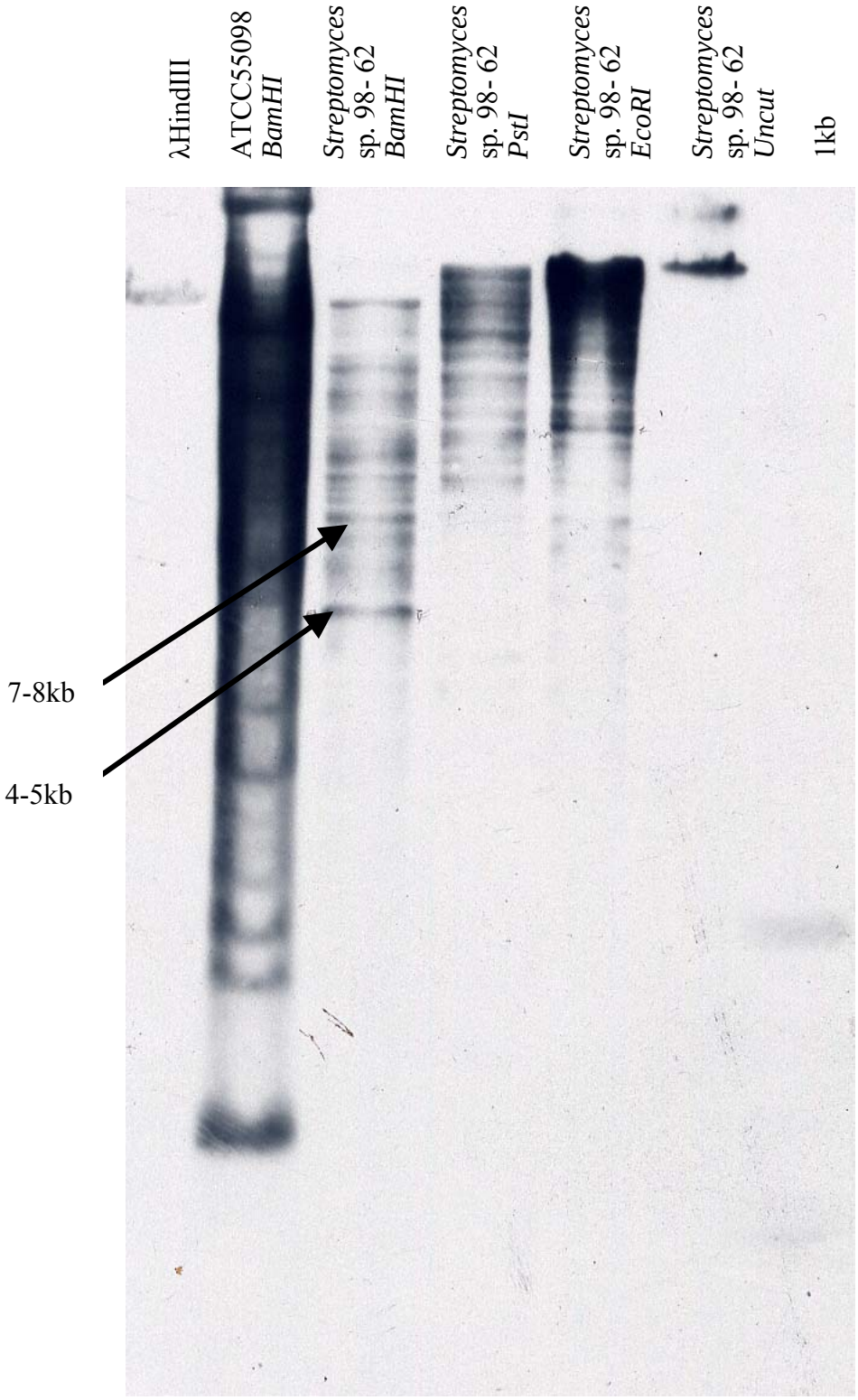


Figure 11: Southern blot of restriction endonuclease digested chromosomal DNA samples of *Streptomyces hygroscopicus var. ascomyceticus* ATCC55098 and the *Streptomyces* sp. 98– 62 probed with eryKSII probe. Genomic DNA of the *Streptomyces* sp. digested with *BamHI gave a 4-5kb* fragment that hybridised strongly to the eryKSII probe.

4.2.2 ANALYSIS OF SECONDARY METABOLITES PRODUCED BY THE STREPTOMYCES SP. *98- 62*

To determine if the antifungal compound produced by the *Streptomyces* sp. 98–62 was similar to PKS I antifungal compounds rapamycin and FK506, secondary metabolites from *Streptomyces* sp. 98- 62 grown in FK medium were subjected to TLC followed by a bioassay against *Aspergillus niger*. Ethyl acetate extract of the 96 h cultures of the *Streptomyces* sp. 98- 62 was analysed by TLC followed by a bioassay against *Aspergillus niger*. Pure rapamycin and FK506 were used as positive controls (Fig. 12). A zone of growth inhibition corresponding to the TLC spot of Rf 0.69 was observed in the case of the extract of the *Streptomyces* sp. 98- 62. Pure rapamycin and FK 506 gave a zone of inhibition at an Rf value of 0.80. From this observation it was concluded that the *Streptomyces* sp. 98- 62 indeed produced an antifungal compound. However, the antifungal compound produced by the *Streptomyces* sp. 98- 62 was determined to be different in its chromatographic separation from that of the PKS I compounds rapamycin and FK506.

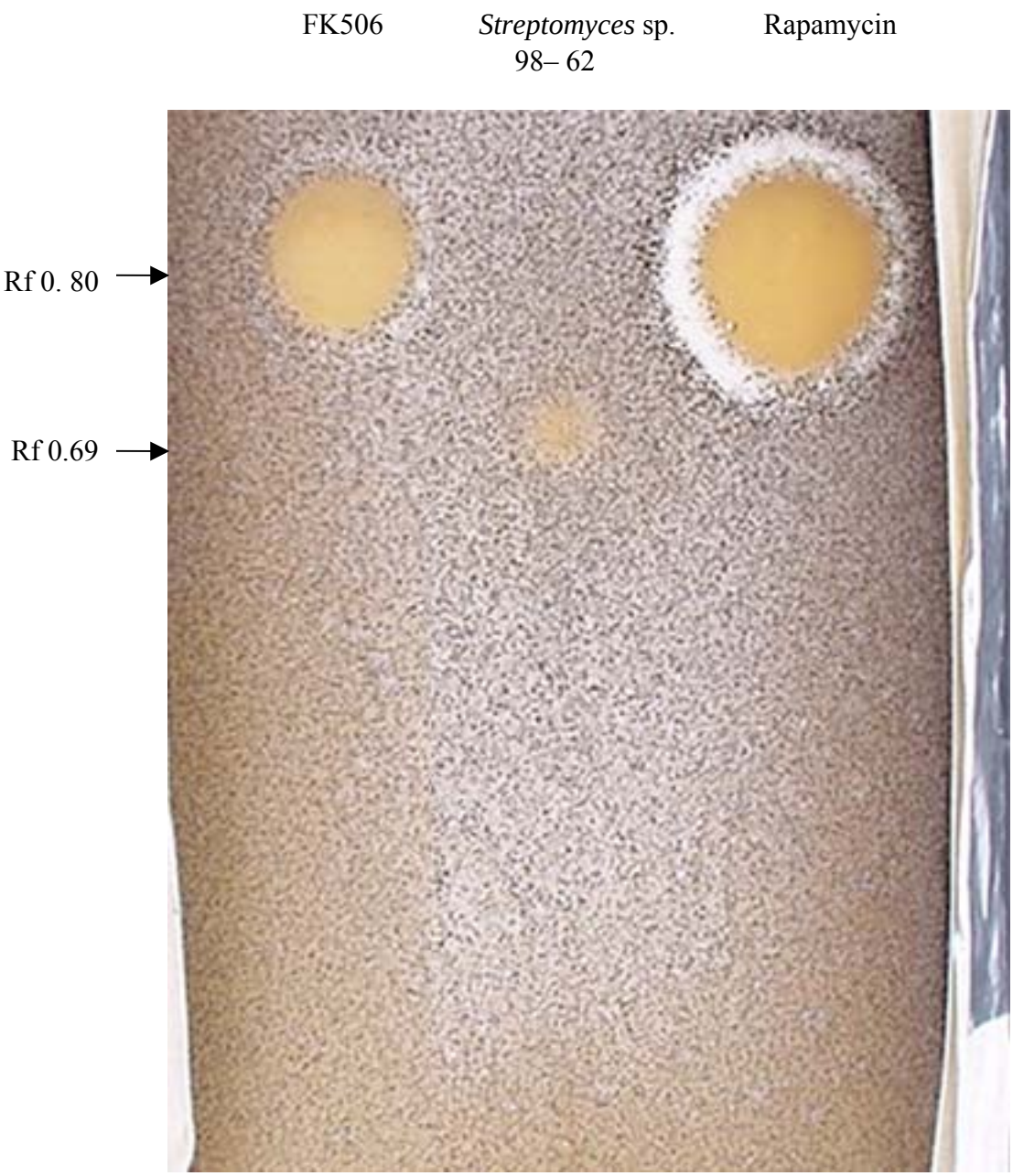


Figure 12: TLC Chromatogram and overlay assay of the extracts of pure FK506, the *Streptomyces* sp. 98- 62 and pure rapamycin. The cleared area represents the zone of inhibition. Test organism used was *Aspergillus niger.*

4.3 CLONING OF THE KETOACYL SYNTHASE-ACYL TRANSFERASE GENE FROM THE STREPTOMYCES SP. *98 -62*

4.3.1 AMPLIFICATION, CLONING AND SEQUENCING OF KETOACYL SYNTHASE- ACYL TRANSFERASE GENE FROM THE STREPTOMYCES SP. *98- 62*

The strategy to isolate the PKS I genes of the *Streptomyces* sp. 98- 62 was to amplify the KS and AT regions using primers targeted at conserved sequences in previously sequenced PKS genes. A pair of degenerate primers spanning conserved regions of KS and AT genes has been used successfully to identify niddamycin cluster (Kakavas *et al*, 1997). As the primer sequences were expected to be highly conserved in most PKS I genes, the same set of primers were used to amplify the KS/AT region from the chromosomal DNA of the *Streptomyces* sp. 98- 62. The PCR product ran as a 850 bp fragment on agarose gel and was subsequently cloned into the pGEMT vector (Promega®) and sequenced using vector primers T7 and SP6. Additional sequencing primers were designed to allow for complete sequencing of the nucleotides of the insert fragment. The nucleotide sequences were aligned using BLAST2 program. By searching database using the BLAST program, the identity of the complete nucleotide sequence was elucidated. The sequence of KS/ AT region from the *Streptomyces* sp. 98- 62 was found to be 843 bp (Fig. 13) and was approximately 50% similar to the other *Streptomyces* PKS I KS/ AT region. The 843bp sequence had the highest similarity to sequences of the antihelminthic avermectin compound producer S*treptomyces avermitilis.* The percentage similarity at the protein level is 54% and the percentage of positives is 62% (Fig. 14). The deduced protein product encoded by the 843 bp is the keto synthase - acyl transferase genes of the PKS type I system (Table 4).

4.3.2 SEQUENCE OF THE KS/AT GENES OF THE STREPTOMYCES SP. *98- 62*

```
CGGTCAAGTCCAACATCGGGCACACCCAGGCCGCCGCCGGGGTCGCCGGCGTCATCAAGATGGTGATGGCCATGCGCCGCGGCAGGC
TGCCGAGGACGCTGCACGCCGAACACCCCACCACCCGGGTCGACTGGGAGTCCGGCGCCGTCGAACTGCTCGGCGAGGCCCGCGACT
GGCCGGACGCGGGGGAGCCCCGCCGCGCCGCCGTGTCCTCCTTCGGCATCTCCGGCACCAACGCCCACGTCATCGTCGAGGCGGCCC
CCGACCCCGAGCCGCGCACCGGGGAACCCGTCTGGGACCGGCCGCTGCCGCTGGTGCTCTCCGCCCGAGACGAACCGGCCCTGGCCG
CCCAGGCACGCCGCATCCTCGACCACCTGGAGACCGGCGCCGACCTCGTCCCCGACATCGCCTACGCCCTGGCCACCACCCGCGCCG
CCCTGGACCGGCGGGCCGTCGTCATCGGCGCCGACCCGGCCACGATCACCGCGCGGCTCGCCGCCCTGGCCGAGGACGATCCGGCGT
CCGACGTGGTGCGCGGCGCACCGGCGGGGGAGTCCCGCATCGCGTTCGTCTTCCCCGGGCAGGGCTCCCAGTGGGCCGGCATGGCCG
CCGAACTGCTCGACGGCTCACCGGTGTTCGCGGCGGCATGGCCGACTGCGCCGAGGCGCTCGCCCCCTTCACCGACTGGGACCTCGT
CGACACCGTCCGGGAGCGCCGCCCCATGGAGCGGGTGGACGTGGTCCAGCCCGCGCTGTGGGCGATCATGGTCTCGCTGGCCGAGGT
GTGGCGCGCGCACGGGGTGCGGCCCGCCGCCGTCATTGGGCACTCCCAGGGCGAGATCGC
```

Figure 13: Sequence of the amplification product from the *Streptomyces* sp. 98- 62 with the primers specific for KS/AT genes of the PKS I systems. The sequence in red represents the deduced primer-binding site.

```
Score    E
Sequences producing significant alignments:                          (bits) Value

gi|15823982|dbj|BAB69199.1|  (AB070940) modular polyketide s...    211   6e-74
gi|15823978|dbj|BAB69195.1|  (AB070940) modular polyketide s...    207   3e-73
gi|15823977|dbj|BAB69194.1|  (AB070940) modular polyketide s...    217   9e-73
gi|478634|pir||S23070  erythronolide synthase (EC 2.3.1.94) ...    194   1e-70
gi|416966|sp|Q03132|ERY2_SACER  Erythronolide synthase, modu...    194   1e-70
gi|581651|emb|CAA44448.1|  (X62569) 6-deoxyerythronolide B  ...    194   1e-70
gi|10179853|gb|AAG13918.1|AF263245_14  (AF263245) megalomici...    196   2e-70
gi|14794893|gb|AAK73501.1|AF357202_4  (AF357202) AmphI [Stre...    197   5e-70
gi|12055072|emb|CAC20921.1|  (AJ278573) PimS2 protein [Strep...    206   5e-70
gi|15823981|dbj|BAB69198.1|  (AB070940) modular polyketide s...    199   2e-69
gi|7522143|pir||T17466  rifamycin polyketide synthase module...    204   2e-69
gi|9049536|gb|AAF82409.1|AF220951_2  (AF220951) 8,8a-deoxyol...    198   6e-69
gi|2506137|sp|Q03133|ERY3_SACER  Erythronolide synthase, mod...    187   7e-69
>gi|15823982|dbj|BAB69199.1|     (AB070940) modular polyketide synthase [Streptomyces
avermitilis]
          Length = 3970

 Score =  211 bits (537), Expect(2) = 6e-74
 Identities = 119/219 (54%), Positives = 138/219 (62%), Gaps = 7/219 (3%)
 Frame = +3

Query: 3    VKSNIGHTQAAAGVAGVIKMVMAMRRGRLPRTLHAEHPTTRVDWESGAVELLGEARDWPD 182
             VKSN+GHTQAAAG AG+IKM+MAMR G LPRTLH + P+  VDW  G VELL E R+WPD
Sbjct: 2178 VKSNLGHTQAAAGAAGIIKMIMAMRYGVLPRTLHVDRPSPEVDWSPGTVELLTEEREWPD 2237

Query: 183  AGEPRRAAVSSFGISGTNAHVIVEAAP-DPEPRTGEPVWDRPLPLVLSARDEPALAAQAR 359
             AG PRRAAVSSFGISGTNAHVI+E  P D EP T   V    +P VLS  D  AL AQA
Sbjct: 2238 AGRPRRAAVSSFGISGTNAHVILEQPPADDEPGTSGTVPGGVVPWVLSGHDRAALYAQAE 2297

Query: 360  RILDHLETGADLVPDIXXXXXXXXXXXXXXXVVIGADPATITARLAALA----EDDPA-- 521
             R++ H+    +L                   VV+G D   + A  A LA    E D A
Sbjct: 2298 RLVAHVAARPELSVADVGRTLTGRARLSHRAVVLGGDRDELLAAAAGLARRAGEPDEALP 2357

Query: 522  SDVVRGAPAGESRIAFVFPGQGSQWAGMAAELLDGSPVF 638
               VV G+  G+ R+ FVFPGQG+QWAGMAAELL  +PVF
Sbjct: 2358 PGVVEGSVLGDDRVVFVFPGQGAQWAGMAAELLVSAPVF 2396
```

Figure 14: Sequence comparison of the KS/AT genes amplified from the *Streptomyces* sp.
98- 62 with the Genbank sequences.

4.3.3 AMINOACID SEQUENCE COMPARISON OF THE KS/AT GENES OF THE STREPTOMYCES SP. *98- 62*

Deduced protein product	Ketosynthase(KS)	
Comparison of amino acid sequence	% Identity	% Similarity
Streptomyces avermitilis	73	80
Streptomyces hygroscopicus var. ascomyceticus	72	79

Deduced protein product	Acyl transferase	
Comparison of amino acid sequence	% Identity	%Similarity
Saccharopolyspora erythrae	58	66
Streptomyces avermitilis	56	65

Table 4: Compilation of the BLASTP result of the deduced KS/ AT genes of the *Streptomyces* sp. 98- 62 with the other PKS I genes in the Genbank.

4.4 SOUTHERN HYBRIDISATION OF CHROMOSOMAL DNA OF THE STREPTOMYCES SP. *98- 62 USING HOMOLOGOUS KETOACYL SYNTHASE-ACYL TRANSFERASE GENE PROBE*

The DNA fragment representing the KS/AT region from the *Streptomyces* sp. 98- 62 was used as a probe for Southern hybridisation experiments using the restriction enzyme digested chromosomal DNA fragments from it to determine if there is multiple KS/AT genes in the *Streptomyces* sp. 98- 62 as is expected of the PKS I system. The DNA from the *Streptomyces* sp. 98- 62 was restricted with different restriction enzymes and probed with the KS/AT genes probe (Fig. 15a, b). When genomic DNA

of the *Streptomyces* sp. 98- 62 was restricted with Sp*hI* and probed with the KS/AT genes, eleven hybridising bands were visible. This result showed that indeed the genome of the *Streptomyces* sp. 98- 62 contained a number of different DNA fragments, which contained homologous KS/AT genes to the KS/AT genes probe of the *Streptomyces* sp. 98- 62. This indicated strongly that the *Streptomyces* sp. 98- 62 indeed contained multiple KS/AT genes as is characteristic of the PKS I system. Cloning and sequencing of the repeated PKS I genes from the *Streptomyces* sp. 98- 62 would provide the conclusive evidence that KS/AT genes are part of a PKS I cluster.

The *BamHI* restricted genome of the *Streptomyces* sp. 98- 62 when probed with the KS/AT genes shared some common features as well as some differences to those obtained when probed with the eryKSII gene probe from the erythromycin producer *Saccharopolyspora erythraea*. These results suggest that some of the PKS I genes from *Streptomyces* sp. have higher homology to the KS II gene from the erythromycin producer *Saccharopolyspora erythraea* whilst some others have a higher homology to the KS/AT genes from the *Streptomyces* sp. 98- 62.

A 7-8kb *BamHI* fragment of the *Streptomyces* sp. 98- 62 was identified to hybridise very strongly to the KS/AT genes probe. This 7-8kb *BamHI* fragment therefore was likely to be the genomic fragment of the *Streptomyces* sp. 98- 62 that contained the KS/AT genes used as a probe.

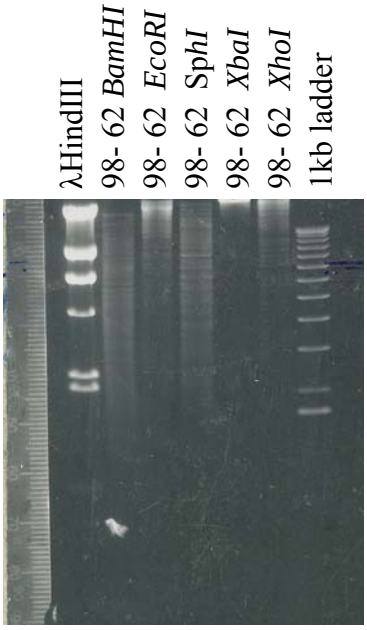


Figure 15a

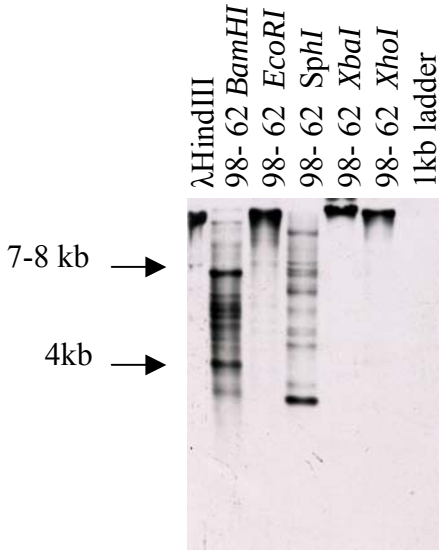


Figure 15b

Figure 15a: Electrophoretic profile of restriction endonuclease digested chromosomal DNA samples of the *Streptomyces* sp. 98– 62. Figure 15b: Southern blot of the restriction endonuclease digested chromosomal DNA samples of the *Streptomyces* sp. 98– 62 probed with KS/AT probe from the *Streptomyces* sp. 98– 62.
Genomic DNA of the *Streptomyces* sp. digested with Sp*hI* gave ~11 hybridising bands with KS/AT probe from the *Streptomyces* sp. 98- 62. Genomic DNA of the *Streptomyces* sp. digested with *BamHI gave a 7-8kb* fragment that hybridised strongly to the KS/AT probe from the *Streptomyces* sp. 98- 62.

4.5SUBGENOMIC LIBRARY CONSTRUCTION OF THE STRAIN 98-62 AND SCREENING OF THE RECOMBINANT CLONES BY PCR

4.5.1 SUBGENOMIC LIBRARY CONSTRUCTION

In order to clone the PKS I gene cluster of 98- 62 surrounding the KS/AT genes, a sub-genomic library of 98- 62 DNA fragments was constructed. This was done by isolating the total genomic DNA, digesting it with *BamHI*, and ligating the purified 7 to 8 kb fragments into the *BamHI* site of pUC18. The ligation mixture was then introduced into *E. coli* Top 10 competent cells. After overnight incubation at 37°C, 500 white and ampicillin resistant colonies were patched onto LB + ampicillin plates.

4.5.2 PCR SCREENING TO IDENTIFY CLONE CONTAINING KETOACYL SYNTHASE-ACYL TRANSFERASE GENE

The plasmid DNA from pools of 50 colonies were extracted and used as template for PCR amplification of the KS/AT genes, using the same degenerate primers used earlier to amplify the KS/AT genes from the genomic DNA of the *Streptomyces* sp. 98- 62. From the identified positive pool of 50 colonies, screening was narrowed to subpools of 10 colonies and thereafter to individual colonies. Eventually, one clone, C170 was identified to give the PCR product of the expected size (Fig. 16a, b).

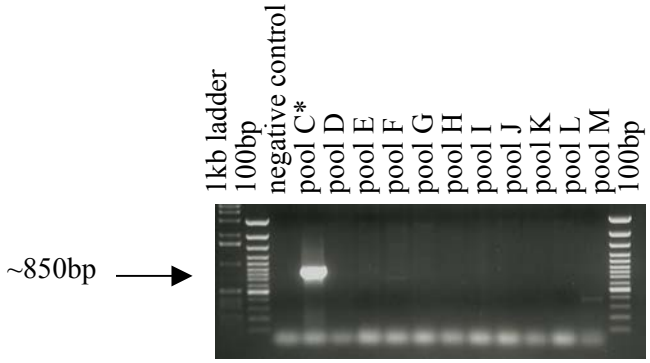


Figure 16a: PCR screening of pool DNA to identify a pool that gave an amplification of product size 850bp. Lane denoted with asterisk gave the expected size amplification product.

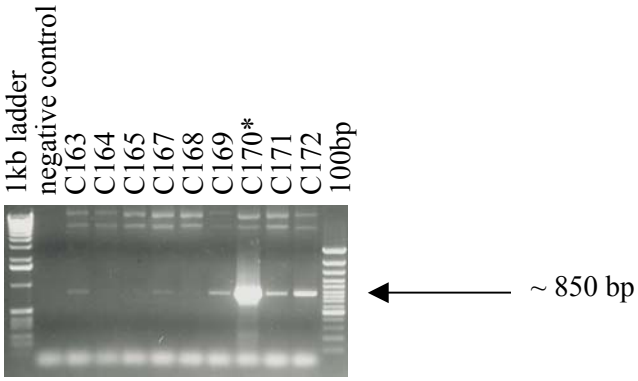


Figure 16b: PCR screening of individual clones to identify a pool that gave an amplification of product size 850bp. Lane denoted with asterisk gave the expected size amplification product.

4.6 RESTRICTION AND SEQUENCE ANALYSIS OF THE DNA INSERT IN THE RECOMBINANT CLONE C170 IDENTIFIED TO CONTAIN THE KETOACYL SYNTHASE-ACYL TRANSFERASE GENE

Restriction digestion of the C170 plasmid DNA with *BamHI* gave a DNA insert fragment of approximately 7-8kb. Restriction digestion of the C170 plasmid DNA with Sp*hI* gave three DNA fragments of the approximate sizes 1.5kb, 2.0kb and 6.5kb upon gel electrophoresis. Two Sp*hI* DNA fragments of sizes 1.5 kb and 2.0 kb were subcloned into the vector pUC18 at the Sp*hI* site. The subclones were designated as p1.6KBC170 and p2.0KBC170 respectively. The larger fragment, which is expected to contain the pUC18 vector, was self-ligated. This subclone was designated as p6.5KBC170. Subclones were sequenced using M13 forward and reverse primers. Complete sequence of the subclones were obtained from primer walking. The nucleotide sequences were aligned using the BLAST2 program. By searching the database using the BLAST program, the complete nucleotide sequence was elucidated.

The recombinant clone was restricted with *BamHI, EcoRI,* Sp*hI, BamHI+ EcoRI, BamhI*+Sp*hI and EcoRI*+ Sp*hI* in order to construct a restriction map for the clone. The restriction profile and the deduced restriction map are given (Fig. 17a, b). The complete sequence of the insert fragment of the recombinant clone C170 was determined to be 7177 bp. Analysis of the sequence for the restriction sites confirmed the predicted restriction profile. Conserved sequences for the restriction enzyme Sp*hI* occurred at nucleotide positions 3723 and 5634. Conserved sequences for the restriction enzyme *BamHI* occurred at the beginning and the end of the fragment.

The DNA sequence data obtained were analysed for open reading frames (ORFs). There were two partial open reading frames, in the same orientation (Fig. 27, 28). The ORFs were labelled ORF 1 and ORF 2 for convenience. ORF1 module was designated as module 1 for convenience, and it encodes a partial DH, a complete KR

and a complete ACP in the stated order. ORF 2 module was designated as module 2 for convenience and it encodes a complete KS, a complete AT which is methyl malonyl specific and a complete DH in the given order. The organization of the enzymatic domains within each module is consistent with other PKS type I genes.

ORF 1 is predicted to terminate with a stop codon TGA. A second stop codon TAG is predicted 372 bases downstream of the first stop codon. ORF 2 is predicted to initiate with a start codon ATG and lies 60 nucleotides downstream of the predicted second stop codon of ORF1. The sequence TGGACA which is located 38nt upstream of the predicted start codon of ORF2 is deduced to be the transcriptional promoter as the sequence is identical to ermE-P1 promoter (Strohl, 1992). The sequence GAGG which is located 14nt upstream of the predicted start codon of ORF2 is deduced to be the ribosomal binding site of ORF2 (Strohl, W. 1992). From the sequence analysis of clone C170 PKS I genes, it is proposed that the encoded ORFs are translationally uncoupled .

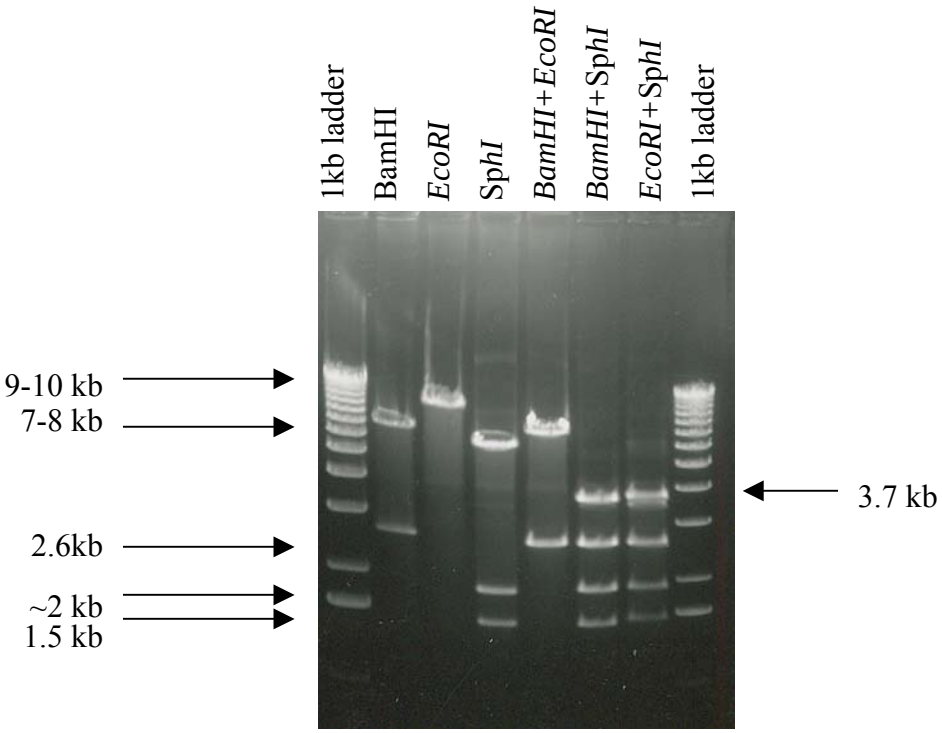


Figure 17a: Restriction profile of clone C170 of the *Streptomyces* sp. 98- 62 digested with different restriction enzymes.

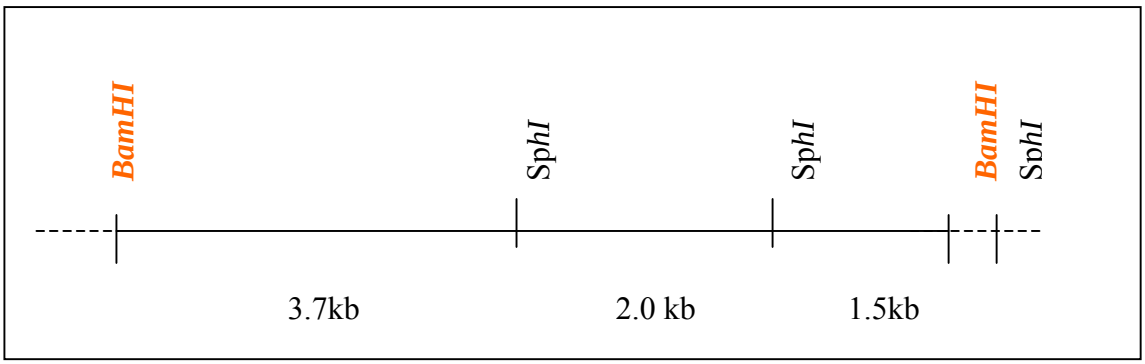


Figure 17b: Restriction map of the clone C170 of the *Streptomyces* sp. 98- 62. The dotted line represents the multiple cloning site of the vector pUC18. The picture is not drawn to scale. The *BamHI* site in red indicates the cloning site.

4.7 CHROMOSOMAL WALKING

SOUTHERN HYBRIDISATION OF THE CHROMOSOMAL DNA OF THE STREPTOMYCES SP. *98- 62 USING EXTERNAL FRAGMENTS OF CLONE C170 TO IDENTIFY ADJOINING UPSTREAM AND DOWNSTREAM GENES TO THE INSERT IN THE CLONE C170*

In order to identify the genomic fragments of the *Streptomyces* sp. 98- 62 that is adjacent to the genomic fragment of the clone C170, the genomic DNA of the *Streptomyces* sp. 98- 62 was restricted with different restriction enzymes and probed with the external sub-genomic fragments of the recombinant clone C170. A 3.7kb Sp*hI/BamHI* fragment of the clone C170 fragment was used as a probe to identify the adjoining upstream genes to the insert of the clone C170. A 1.5kb Sp*hI/BamHI* fragment of the clone C170 fragment was used as a probe to identify the adjoining downstream genes to the insert of the clone C170.

When probed with the 3.7 kb Sp*hI/BamHI* fragment of the clone C170 fragment, a 5.5-6.5kb Sp*hI* fragment of the *Streptomyces* sp. 98- 62 showed the strongest hybridisation. Hence this 5.5-6.5kb Sp*hI* fragment was deduced to contain the adjoining upstream genes to the insert of the clone C170 (Fig. 18a). When probed with the 1.5kb Sp*hI/BamHI* fragment of clone C170 fragment, a 3.5-4.5kb Sp*hI* fragment of the *Streptomyces* sp. 98-62 showed the strongest hybridisation (Fig. 18b). Hence this

3.5-4.5kb Sp*hI* fragment was deduced to contain the adjoining downstream genes to the insert of the clone C170.

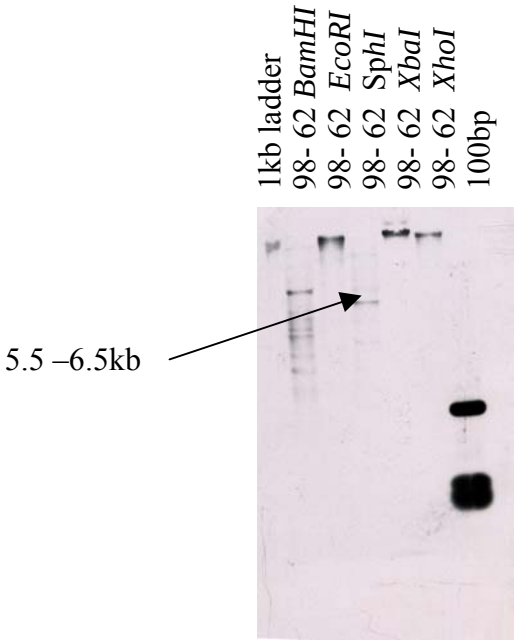


Figure 18a

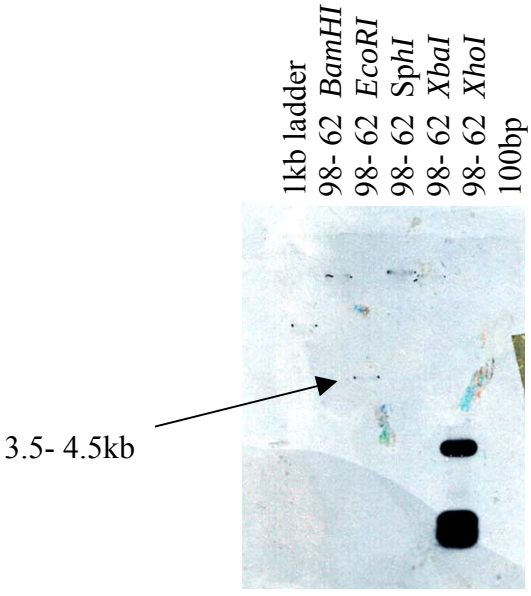


Figure 18b

Figure 18a & b: Southern blot of the restriction endonuclease digested chromosomal DNA samples of the *Streptomyces* sp. 98– 62 probed with the 3.7kb Sph*I/ BamHI* probe from the *Streptomyces* sp. 98– 62 and 1.5kb Sph*I/ BamHI* probe from the *Streptomyces* sp. 98– 62, respectively. Genomic DNA of the *Streptomyces* sp. digested with Sph*I* gave a 5.5-6.5 kb hybridizing band with the 3.7kb Sph*I/ BamHI* probe from the *Streptomyces* sp. 98– 62 and a 3.5-4.5 kb hybridizing band with the 1.5kb Sph*I/ BamHI* probe from the *Streptomyces* sp.
98– 62. See Fig. 15a for the electrophoretic profile of restriction endonuclease digested chromosomal DNA samples of the *Streptomyces* sp. 98– 62.

4.8 SUBGENOMIC LIBRARY CONSTRUCTION AND SCREENING OF THE RECOMBINANT CLONES BY PCR TO IDENTIFY THE ADJOINING DOWNSTREAM GENES TO THE INSERT OF THE CLONE C170; CLONE C2

4.8.1 SUBGENOMIC LIBRARY CONSTRUCTION

A library of the *Streptomyces* sp. 98- 62 DNA fragments was constructed by isolating the total genomic DNA, digesting it with Sp*hI*, and ligating the purified 3.5-4.5 kb fragments into the Sp*hI* site of the vector pUC18. The ligation mixture was then introduced to *E. coli* Top 10 competent cells. After overnight incubation at 37°C, 500 white and ampicillin resistant colonies were patched onto LB+ ampicillin plates.

4.8.2 PCR SCREENING TO IDENTIFY THE CLONE CONTAINING THE DOWNSTREAM GENES TO THE INSERT OF THE CLONE C170

The plasmid DNA from pools of 50 colonies were extracted and used as template for PCR screening to identify a clone containing the DNA fragment that overlaps and carries the downstream genes to the insert of the clone C170. From the deduced DNA sequence of the insert of the clone C170, a pair of primers depicting the DNA fragment spanning the deduced overlapping region of clone C170 and the putative downstream gene fragment, was designed. This pair of primers was expected to amplify a 573 bp product. From the identified positive pool of 50 colonies, screening was narrowed to subpools of 10 colonies and thereafter to individual colonies. One clone, C2 was identified to give a PCR product of the expected size (Fig. 19a, b).

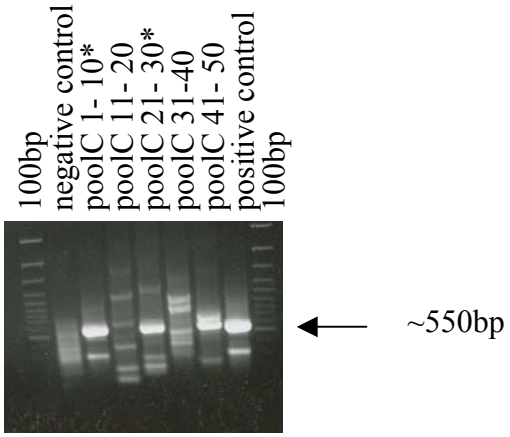


Figure 19a: PCR screening of pool DNA to identify the pool that contains the clone downstream to the insert fragment of clone C170. The expected PCR fragment size is 573bp. Lane denoted with asterisk gave the amplified product of the expected size.

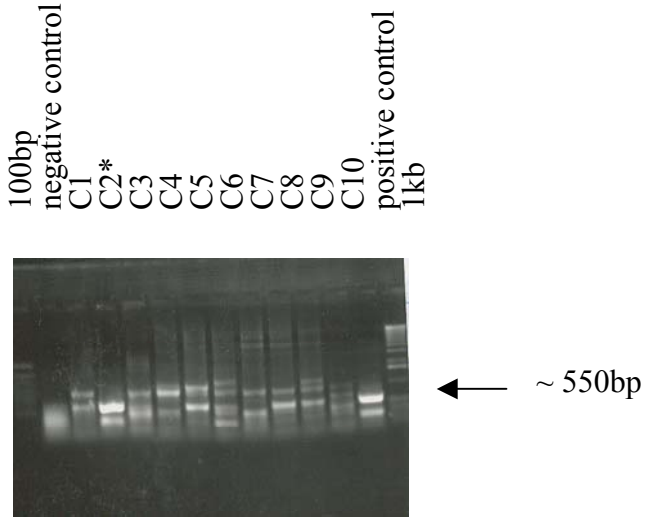


Figure 19b: PCR screening of individual clones to identify a clone that carried the insert downstream to the DNA insert fragment of clone C170. The expected PCR fragment size is 573bp. Lane denoted with asterisk gave the amplified product of the expected size.

4.9 RESTRICTION AND SEQUENCE ANALYSIS OF THE RECOMBINANT CLONE C2 IDENTIFIED TO CONTAIN THE FRAGMENT THAT CARRIED THE ADJOINING DOWNSTREAM GENES TO THE INSERT OF THE CLONE C170; CLONE C2

Restriction digestion of the C2 plasmid DNA with Sp*hI* gave an insert fragment of approximately 3.8kb. Restriction digestion of the C2 plasmid DNA with *BamHI* gave three fragments of the approximate sizes 1.5 kb, 2.1 kb and 2.6 kb, upon gel electrophoresis. Clone C2 was sequenced using M13 forward and reverse primers. Complete sequence of the clone C2 was obtained from primer walking. The nucleotide sequences were aligned using the BLAST2 program. By searching database using the BLAST program, the identity of the complete nucleotide sequence was elucidated.

The recombinant clone was restricted with *BamHI,* Sp*hI, and BamHI*+Sp*hI* in order to restriction map the clone. The restriction profile and deduced restriction map are given in the Fig. 20a, b. The complete sequence of the insert fragment of the recombinant clone C2 was determined to be 3682 bp. Analysis of the sequence for restriction sites confirmed the predicted restriction profile. Conserved sequences for the restriction enzyme *BamHI* occurred at the nucleotide position 1537 and 3139. Conserved sequences for the restriction enzyme Sp*hI* was only observed at the beginning end of the insert fragment. The end part of the clone was resistant to sequencing and therefore sequence information for the last 20-30 nucleotides was very noisy.

The nucleotide sequences were analysed for encoded protein products. The domains represented are a partial AT, a complete DH, a complete KR, a complete ACP and a partial KS in the stated order. The nucleotide sequence and the order of PKS I gene domains is in agreement with the deduction that the clone C2 overlaps and lies downstream of the clone C170.

The sequence analysis also revealed that the 3.8 kb fragment of the clone C2 encompasses 2 modules, module 2 and a downstream module designated for convenience as module 3. There is no stop/start codons or ribosomal binding sites or such regulatory sequences between the two modules. This suggests that module 2 and module 3 are translationally coupled and belong to the same ORF, ORF 2.

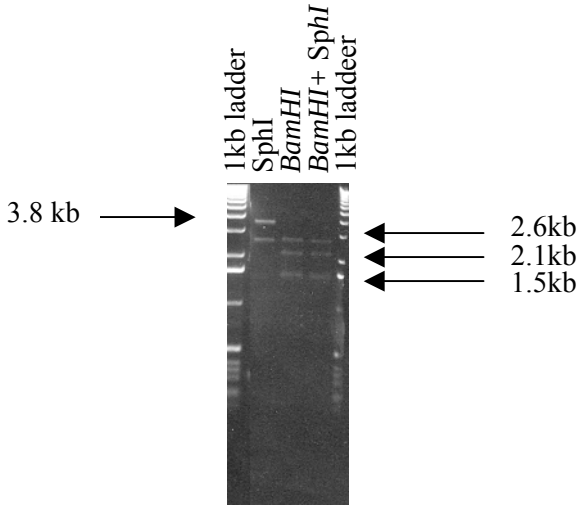


Figure 20a: Restriction profile of clone C2 digested with different restriction enzymes.

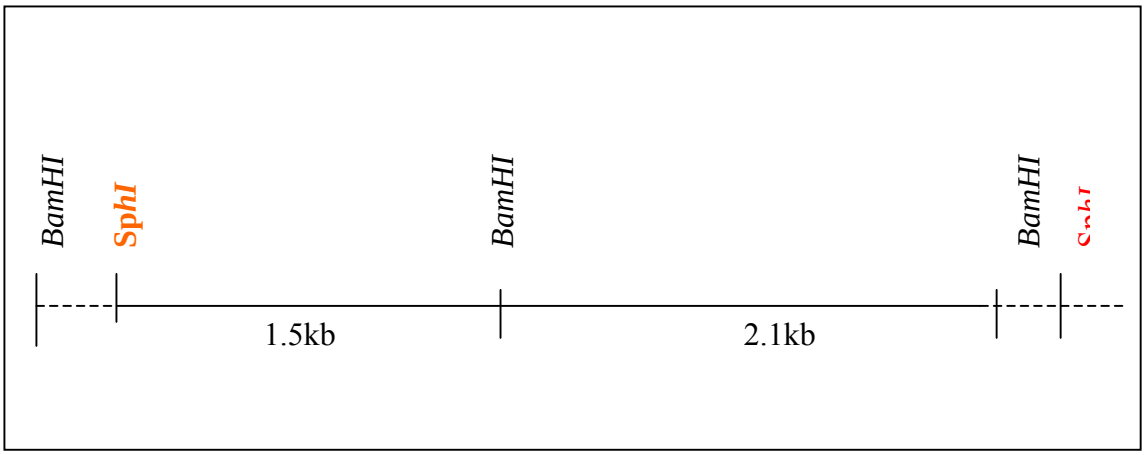


Figure 20b: Restriction map of the clone C2. The dotted line represents the multiple cloning site of the vector pUC18. The cloning site is indicated in red. The picture is not drawn to scale.

4.10 SUBGENOMIC LIBRARY CONSTRUCTION AND SCREENING OF THE RECOMBINANT CLONES BY PCR TO IDENTIFY THE ADJOINING UPSTREAM GENES TO THE INSERT OF THE CLONE C170; CLONE E27

4.10.1 SUBGENOMIC LIBRARY CONSTRUCTION

A genomic library of the *Streptomyces* sp. 98- 62 DNA fragments was constructed by isolating total genomic DNA, digesting it with Sp*hI*, and ligating the purified

5.5-6.5 kb fragments into the Sp*hI* site of the vector pUC18. The ligation mixture was then introduced to *E. coli* Top 10 competent cells. After overnight incubation at 37°C, 500 white and ampicillin resistant colonies were patched onto LB+ ampicillin plates.

4.10.2 PCR SCREENING TO IDENTIFY THE CLONE CONTAINING UPSTREAM GENES TO THE INSERT OF THE CLONE C170

The plasmid DNA from pools of 50 colonies were extracted and used as template for PCR screening to identify a clone containing DNA that overlaps and carries the upstream genes to the insert of the clone C170. From the deduced sequence of the clone C170, a pair of primers depicting the DNA fragment spanning the deduced overlapping region of clone C170 and the putative upstream gene fragment, was designed. This pair of primers was expected to amplify a 444 bp product. From the identified positive pool of 50 colonies, screening was narrowed to subpools of 10 colonies and thereafter to individual colonies. One clone, E27 was identified to give the PCR product of the expected size (Fig. 21a, b).

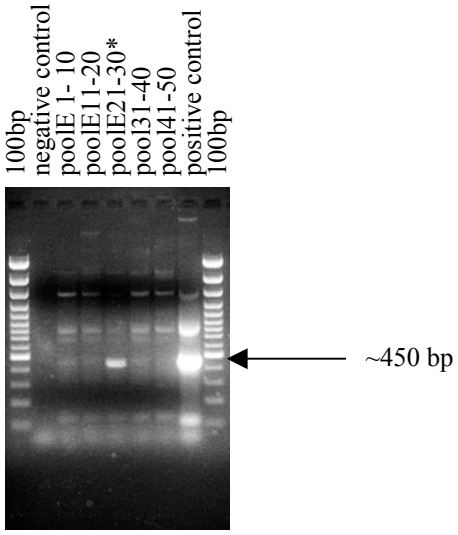


Figure 21a: PCR screening of pool DNA to identify a pool that contains a clone upstream to the insert fragment of clone C170. The expected PCR fragment size is 444bp. Lane denoted with asterisk gave the amplified product of the expected size.

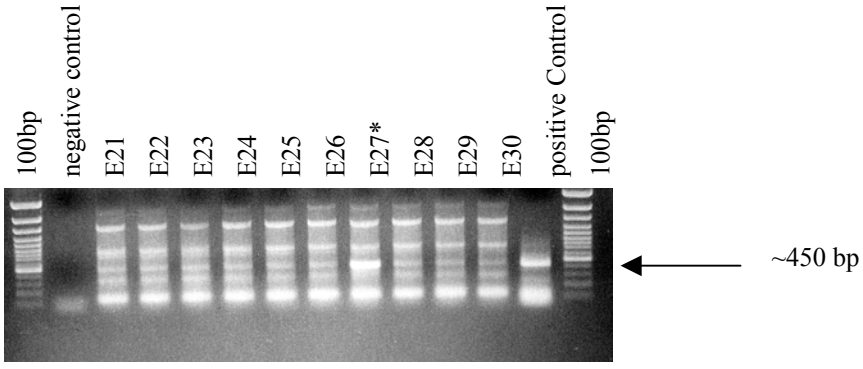


Figure 21b: Colony PCR screening of individual clones to identify a clone upstream to the insert fragment of clone C170. The expected PCR fragment size is 444bp. Lane denoted with asterisk gave the amplified product of the expected size.

4.11 RESTRICTION AND SEQUENCE ANALYSIS OF THE RECOMBINANT CLONE E27 IDENTIFIED TO CONTAIN THE DNA INSERT THAT CARRIED THE ADJOINING UPSTREAM STREAM GENES TO THE INSERT OF THE CLONE C170; CLONE E27

Restriction digestion of the E27 plasmid DNA with Sp*hI* gave an insert fragment of approximately 6.1kb. Restriction digestion of E27 plasmid DNA with *BamHI* gave two fragments of the approximate sizes 3.7kb and 5.4kb upon gel electrophoresis. The 3.7kb *BamHI* fragment was deduced to be the overlapping region between clone E27 and the clone C170. The larger fragment, which was expected to contain the pUC18 vector, was self-ligated and sequenced using M13 forward and reverse primers. This subclone was designated as p2.3KBE27. Complete sequence of the subclones was obtained from primer walking. The nucleotide sequences were aligned using the BLAST2 program. By searching database using the BLAST program, the identity of the complete nucleotide sequence was elucidated.

The recombinant clone was restricted with *BamHI,* Sp*hI, and BamHI*+Sp*hI* in order to restriction map the clone. The restriction profile and deduced restriction map is given in the Fig. 22a,b. The complete sequence of the insert fragment of the recombinant clone E27 was determined to be 6069 bp. Analysis of the sequence for restriction sites confirmed the predicted restriction profile. Conserved sequences for the restriction enzyme *BamHI* occurred at the nucleotide position 2340. Conserved sequences for the restriction enzyme Sp*hI* was observed at the beginning and the ending of the insert fragment.

The nucleotide sequence of the DNA insert in clone E27 was analysed for encoding protein products. The domains represented in the 6.1 kb sequence are a partial KS, a complete AT, a complete DH, a complete KR, a complete ACP and a

partial KS. The nucleotide sequence and the order of PKS I gene domains of the clone E27 is in agreement with the deduction that the clone E27 overlaps and lies upstream of clone C170.

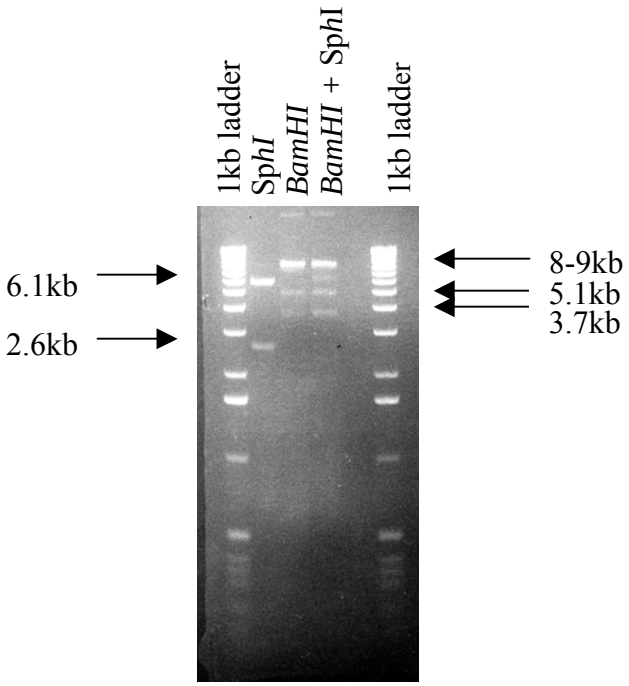


Figure 22a: Restriction profile of clone E27 digested with different restriction enzymes.

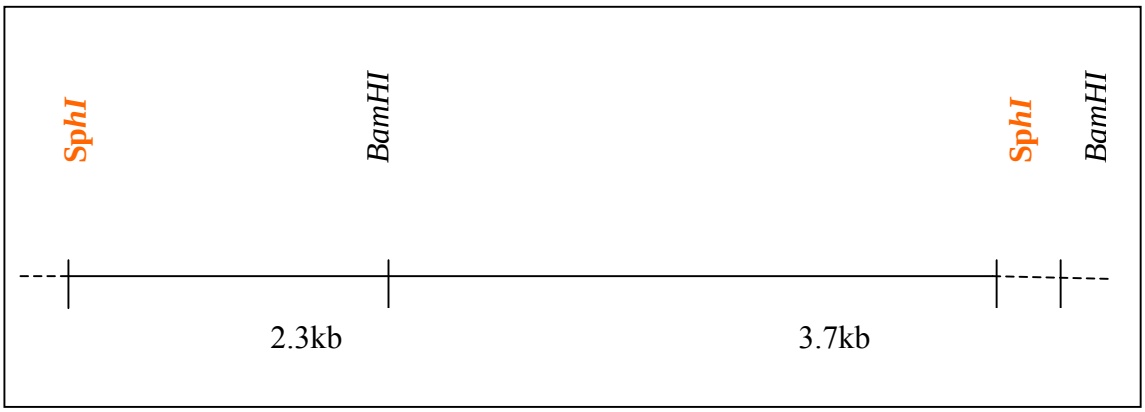


Figure 21b: Restriction map of clone E27. The dotted line represents the multiple cloning sites of the vector pUC18. The cloning site is indicated in red. The picture is not drawn to scale.

4.12 RESTRICTION ANALYSIS AND SEQUENCE ANALYSIS OF OVERLAPPING CLONES C2, C170 AND E27

In order to further characterize the cloned DNA region of the *Streptomyces* sp. 98- 62 and to analyse the potential similarities of these to PKS genes from actinomycetes, the nucleotide sequence of the 11656bp fragment was determined and the restriction profile elucidated. The nucleotide sequence is shown in Fig. 23. The restriction profile of the three contiguous clones are shown in Fig. 24.

4.12.1 SEQUENCE OF OVERLAPPING CLONES C2, C170 AND E27

GCATGCTCTTTGNNTAACGGTTCTCCGACGCCCGTCGCAACGGNCACCGGGTCCTGGCCGCGGTCCGTTNTTCCGCCGTCAACTCCG
ACGGCGCGTCCAACGGGCTGACCGCCCCCAACGGGCCCTCCCAGCAACGCGTCATCCGCGCCGCGCTCGCCGCCGCCCGCCTCGCCC
CGGCCGATGTCGACGCGGTCGAGGCGCACGGCACCGGCACCACGCTCGGCGACCCGATCGAGGCGCAGGCGCTGCTGGCCACGTACG
GCCAGGACCGGCCGGGCGACGAACCCCTCTGGCTCGGCTCCGTCAAGTCCAACATGGGCCACACCCAGGCCGCCGCCGGGGTGGCCG
GAATCATCAAGATGGTCATGGCGATGCGGCACGGCACCCTGCCCCGCACCCTGCACGTCGACACGCCCTCCCACCAGGTCGACTGGA
CGACGGGCGCGGTCCGCCTGCTCACGGAGGAGCGGCCCTGGCCGGGAGCGGCGGACCGTCCGCGCCGGGCGGGGGTGTCCTCGTTCG
GGATCAGCGGCACCAACGCCCATGTGATTCTTGAGGAGTTCGAGGAGTTCGAGGAGTTCGCGGGGGAGCCGGTCGGGACGGGGCCGC
GGACCGCCGGTCCGGACGCCGACGGGCACGACGGTGCGGCAGCGCACCCTCCCGCCACGCCGCCCGTACTCGCCCTTCCGGTCTCCG
CCCGCTCACCCGAGGCCCTGCGCGGCCAGGCGGCCCGCCTGCGGGAACTGACCGGCACCTCGGCCGCCGAACTCGGCCTCGCCCTGT
CCACCACCCGCACCACCCACCCGTACCGCGCCGTCGTCCTCGCCCCCGGTGAGGAGCGGGCCGACGAGGCCCTGGACGCCCTCGCCC
ACGGGCACGAGGCACCCGGCCTGCTCGTCAGCGGTTCCATCACCGACGGCACCCTGGCCTGTCTGTTCTCCGGGCAGGGCGCCCAGC
GGCCCGGCATGGGCCGGGACTGGTACGACACCTTCCCGGTCTACGCGGAGCACTTCGACCGCACGGGCGAACTCTTCGCCAAGCACC
TGGAACGGGCGCTCGCCGAAGTGGTCCTGGGCGACCACCCCGACGTACTGGAACGGACCGCCTACACCCAGGCCGCCCTCTTCACCA
CCCAGGTCGCCCTCTACCGACTGCTGGAGTCCTTCGGGCTGCGGCCCGACTGGCTGGCCGGCCACTCCGTCGGCGAGTTCGCCGCCG
CGCACGTCGCCGGTGTGTGGTCGCTCCAGGACGCCGTCACCGCCGTCGCGGCGCGCGGCAGGCTCATGCAGGCGCTTCCCGAGGGCG
GTGCGATGACCGCCGTACAGGCCGCCGAGGAGGAGGTGCGGCCGCTGCTGGACGAACGGTGCGACATCGCCGCGGTCAACGGCCCGC
GCGCCGTGGTCGTCTCCGGGGACGAGGACGCCGTCGCCGCCGTCGCCGCGCACTTCGCCACCACCCGGCGACTGCGCGTCTCGCACG
CCTTCCACTCGCCGCGCATGGAACCCGTGCTGGACGAGTTCCGCCGGGTCTTGGCCGCCCTGCCGGCCGGGGAACCGGCCCTGCCGA
TCGTCTCCACCCTCACCGGCGCCCGGGCCACCGCCGCCGAACTCGGCTCCGCCGACTACTGGGTACGGCACGTACGGGAGACCGTCC
GCTTCGCCGACGCCGTGGGGACGCTGGCCGCGCAGGGCGCCGACACCTTCCTCGAACTCGGCGCCGCTCCCGTCCTGACGGCCCTCG
GCCCGGACTGCCTCCCGGACGCGGACGCCGAGGAGGCCGCGTTCGTCCCCACCGCCCGCAAGGGCACCGCCGAGGTGCCCGGTCTGC
TGGCCGCCCTGGCCGCCGTGCACACCCGCGGTTCGGACGTCGACTGGGCGGTCCTCTACGACGGCCTCCCCGGGCACCGCGACCGAC
CCGGGCGCCGCGACGAACCCGGGCACCGCGACCAACCGGGGCGCCGTGACCAACCGGGGCGCCGCGTCGAACCGGGGCGTTGTGTCG
AGCTGCCTACCTACGCCTTCCAGCACCGCCGCTACTGGCTTCCCACGTCCACCGCCACCGCCAGGGGCGACGCTGCCGGTCACGGTC
TCGCGGCCGTCGACCACCCCTTCGTCAGCGCCCGCCTCGACCTGCCGGGCGACGGCGGAACCCTGCTCACCGGCCGGATCTCCACCG
CCACCCACCCGGTGCTCGCCCAGCACGCCGTGCTCGGATCGGTGCTGGTGCCCGGCGCCGCCCTCGTCGATCTCGCCCTGTACGCAA
GTGGGTTGACGGGACGCCCGGTGCTGGAGGAACTCACCCTCCAGGCCCCGTTGGCCCTGCCCGGGAACGGTGCCGTACGGATCCAGG
TCGCGCTCCGGCCCGACGGCGGTGTGGAGATCCACTCCCGGCCCGCCGATGCGCCCGAGGACGGGAGCTGGACCCGGCACGCCACCG
GCACCCTCACCGTCACCGACCCCGCCTCCGGACTTCCCGCGTCGTCCGTTCCGTCCGCCGCCTGGCCGCCGCCGGGTGCCGTGCCGC
TCGACACCGACGGCCTCTACGAGCGGCTGCGCGGCGAGGGTTACGACTACGGCCCCGTCTTCCAGGGCGTACGGGCCGCCTGGCGGC
ACGGCGACACGGTCCTCGCCGAACTCGAACTGCCCGCCGAGGCCCGGCAGGACGCCGCCCGGCACGTCCTGCACCCCGCGCTGCTGG
ACTCCGCCCTGCACACCACCGCCCTCGCCGACGCGGACGCCCGCGACGCGGTACCGGACGGCACGATCGCCCTGCCCTTCGCCTGGA
CCGGTGTCACCGTGCACGGACGGCCGTCGTCACGTACCACCCCGTCCCGCACGGGCGTCCCCTCCCGCGCAGCCGCCCCGGACCACA
CCGCAGCCCGGGTCCGCGTCACCCGGGGCGAGGAGGGCATCCGGCTCGATCTGACGGACACCGAGGGCGGGCCGCTGGCCACTGTCG
CGTCCTACGTCACCCGCCCCGTCACCGCCGACCGGCTCACCGGGCGGCAGCGTTCCCTGTACGTCGTCGAGGACGCGCCGCTCCCCG
AGTCCGCCGGGCGCCCCGAGCGCCGCACCTGGGCCGTGCTGGGCCCGGACGACCTCGGACTCGGCGTCCCGCACCACCCCGAACCGG
CCGCGATCGACGGCCCCGCACCCGACGTCGTCGTCCTTCCGGTGCACATCCCGGACGTCGCCGACGCGGACGCCGACGGCGAACGGG
TGCCGGGGGCCGTGCGTACCGCGCTGAACACGACGCTCACGACCCTCCGGGCCTGGCTGGACGACGAACGCCGGGCCGGTTCCACGC
TGCTGGTGCTCACCGAGGGAAGCCTCGCCGACGCCGCCGTGCACGGACTGGTGCGGGCCGCGCAGGCCGAACACCCGGGCCGGATCG
TCCTTGTCGGCCGGGCCGGGCCCGGCAGCCCCGTCCCGGACCGCGCAGCGCTGGCCGCCGTCCTCGACTCCGGTGAACCGGAGGTGC
GGTGGCGGGACGGCCGGGCCCACGCCCCGCGCCTGGTGCGCGCCGGGGAGCCGGACGCGCCGCGCACCGGGCGCCCCTGGGGCACCG
TCCTGATCACCGGCGGCACCGGCGGGCTCGGCGCCCTGGTGGCCCGGCACCTGGTGACCCGGCACGGCGTCACCCGCCTGATCCTGG
CGGGCCGTCGCGGACCCGCCGCCCCGGGCGCCGACGAACTGCGCGCGGACCTGGCCGGCCTGGGCGCCCAGGCCGATGTCGTCGCCT
GCGACGTCGCCGACCGCACGGCGCTCGCCGCGCTGCTGGCCGCCCACCCCGTCGACAGCGTCGTGCACACCGCGGGCGTCCTGGACG
ACGGACTGGTCACCTCGCTCGGCCCCGAACGCCTGGACACGGTCCTGCGCCCCAAGGCGGACGCCGCCTGGCACCTGCACGAACTGA
CCCTCGACCGGCCGCTGTCCCACTTCGTGCTGTTCTCCTCGGCAGCGGGCACCATCGACGCCTCCGGCCAGGGCAACTACGCCGCCG
CCAACGTCTTCCTCGACGCCCTGGCAGTCCACCGTGCCGCCCGGTACCTGCCGGCGCTCTCCCTCGCCTGGGGCCTGTGGTCCGGTG
GCGGCATGGGAGCCGGCCTCGACGAGAGCGGCGCCCGGCGCATCGAACGGTCCGGCATCGGCGCCCTCGACCCGGAGGAGGGCCTCG
AACTCTTCGACGCCGCCGTGGCGTCCGGCCGCCCCGCCCTGGTGCCGGTCCGGCTGGACACCACCGTGCTGCGCCGCCGGGGCGACG

ACGTACCGCCGGTGCTGCGCACCCTGGCCGGTGTCACCGCCCCCGCCGCACGGGAGGACCGGACCCGCGGCCTCGGCGAGCGCCTGG
CCGCCCTGCCCGCCGCCGACCACGAGCACACCGTGCTGGAGGCCGTCCGTACCGAGGTCGCCGCCGTCCTCGGCCACGACGGACCCG
CCGCGGTCGGGCCTCGGCGCGCTTTCACCGAGCTGGGATTCGACTCGCTCGCCGCGGTCGAACTGCGCAACCGGCTCAACGCGATCA
GCGGACTGCGCCTGCCGTCGACGCTCGTCTTCGACTACGCCACTCCCGTGGCGCTGGCGGGCCATCTGCTCGAACGGCTAGCCCCGG
ACGACGACACCGGCACCGGTGCGGCGCCCACCGACCCGAGGGGCGACGACGAGGTGCGGGCCCTCATCGACCGCATCCCGATCGCGC
GCATCCGCGACGCCGGACTGCTCGACGGGCTGCTGAGACTGTCCGAAGCGGCCCCGCCCGCACCGCCCGCCGCCGACCGGGTCATGG
ACATCAGGTCCATGGGGGTGGCCGATCTGGTGCGAGCCGCGCTGAACCGCACCAGCCCCGAGTGAGACCGCCCCGGTGCCGGACCGC
GCGGACCACGCGCCGCGTCCGGAACCGGCCGCACATCCGGCCCGCACATCCGGCCGGTACGACCGGCCGCACATGGTCGGCCCGTAC
GACCGCCCATACGGCCGGAGCACCTCAGCCGTACCTCCGCAGCACCTCCAGCGCGTCCCCCGCACATCCGCACCGCGTCACCAGCGC
CGCCGAGTCAGCCAGTGCTGCGACGGGAAAGGTTCACCGGCTCGCGACGCCCGGCACGGCACCTCGCCGCGTTGTCGCATCACCCGA
GTACGTCCCGTACGAGGGCGATCCTCCGCCTTACGACGCACCGCACCGGACGCCGCGAGCTTCCCGGCAAACCCTTCCGGCCACAGC
ACTAGGGAGCGATACCGACCGTGGACACATCCGTCGAGCAGATCGTCGAGGCGCTGCGCGAGGCCATGCTCGAGAACGAGCGGCTGC
GCCGGCAGAACGACCGGATCGCCGAAGCGGCGCACGAGCCCGTCGCCGTCGTCGCCATGAGCTGCCGCTACCCCGGCGGCGTCGGCA
CGCCCGAACAGCTGTGGCAACTGGTCGACGCCGGAGTGGACGCCGTGGGCGACTTCCCGGACGACCGGGACTGGGACGTCGACGCCA
TCTACGATCCCGACCCCGACGCCCCCGGCAGGACCCATGTCGCGAGGGCGGATTCCTCCACGACGCACCGCGGTTCGACCCGGGCT
TCTTCGGTATCAGCCCGCGTGAGGCCCTCGCCATGGACCCGCAGCAGCGGCTGCTGCTGGAGACCGCCTGGGAGGCGTTCGAACGCG
GCGGCATCGACCCGCACACCCTGCGCGGCAGCCGCACCGGCATCTACGCCGGGGTCATGTACCACGACTACGGCAGCTGGCTCACCG
ACGTACCGGAGGGCGTCGAGGGCTACCTCGGCAACGGCAACCTCGGCAGCGTCGCCTCCGGCCGCGTCTCCTACACGCTCGGCCTGG
AGGGCCCCGCCGTCACCGTCGACACCGCCTGCTCCTCCTCGCTGGTCGCCCTCCACCTGGCCGTGCAGGCCCTGCGCACCGGCGAGT
GCGCCCTCGCCCTGGCCGGGGGCGTGACCGTGATGTCCACCCCGGACACCTTCATCGACTTCTCCCGCCAGCGCGGGCTCGCCCTGG
ACGGGCGCTGCAAGTCCTTCGCGGAGGGCGCCGACGGCACCGGCTGGGGCGAGGGCGTCGGCATGCTCCTGCTGGAACGGCTCTCCG
ACGCCCGCCGCAACGGCCACCGCGTCCTCGCCGTCGTCGCGGCACCGCCGTCAACCAGGACGGCGCCTCGAACGGGCTGACCGCGC
CCAACGGCCCCTCCCAGCAACGCGTCATCCGCGCCGCGCTCGCCGACGCCCGCCTGGAACCCCACCAGGTGCACGCCGTGGAGGCGC
ACGGCACCGGCACCCCGCTCGGCGACCCCATCGAGGCCCAGGCCCTGCTCGCCACCTACGGGCAGGACCGGCAGGCCGGCGAACCGC
TGTGGCTGGGCTCGGTCAAGTCCAACATCGGGCACACCCAGGCCGCCGCCGGGGTCGCCGGCGTCATCAAGATGGTGATGGCCATGC
GCCGCGGCAGGCTGCCGAGGACGCTGCACGCCGAACACCCCACCACCCGGGTCGACTGGGAGTCCGGCGCCGTCGAACTGCTCGGCG
AGGCCCGCGACTGGCCGGACGCGGGGGAGCCCCGCCGCGCCGCCGTGTCCTCCTTCGGCATCTCCGGCACCAACGCCCACGTCATCG
TCGAGGCGGCCCCCGACCCCGAGCCGCGCACCGGGGAACCCGTCTGGGACCGGCCGCTGCCGCTGGTGCTCTCCGCCCGAGACGAAC
CGGCCCTGGCCGCCCAGGCACGCCGCATCCTCGACCACCTGGAGACCGGCGCCGACCTCGTCCCCGACATCGCCTACGCCCTGGCCA
CCACCCGCGCCGCCCTGGACCGGCGGGCCGTCGTCATCGGCGCCGACCCGGCCACGATCACCGCGCGGCTCGCCGCCCTGGCCGAGG
ACGATCCGGCGTCCGACGTGGTGCGCGGCGCACCGGCGGGGGAGTCCCGCATCGCGTTCGTCTTCCCCGGGCAGGGCTCCCAGTGGG
CCGGCATGGCCGCCGAACTGCTCGACGGCTCACCGGTGTTCGCGGCGGCCATGGCCGACTGCGCCGAGGCGCTCGCCCCCTTCACCG
ACTGGGACCTCGTCGACACCGTCCGGGAGCGCCGCCCCATGGAGCGGGTGGACGTGGTCCAGCCCGCGCTGTGGGCGATCATGGTCT
CGCTGGCCGAGGTGTGGCGCGCGCACGGGGTGCGGCCCGCCGCCGTCATTGGGCACTCCCAGGGCGAGATCGCCGCCGCGTGCGTGG
CGGGCGCGCTGAGCCTGTCCGACGGGGCCCGCGTGGTGGCCCTGCGCAGCCGGGCCATCGCGGAAGTGCTCTCCGGACCCGCCGATT
CCGGGACCGTTCCCGGGAAAGGTGCCTCCGGGCCCACCAATTCGGCGCGTGGCGCCTGTGGCCGCGGCGGGATGATGTCGGTGGCGC
TGCCCGAGTCCCGGGCGCGCGAACTCGTCGCCGCCCACGACGGGCGGGTCGCCGTGGCCGCGGTCAACGGCGCCTCGTCGGTGGTGC
TCTCCGGGGACGCCGAGGTGCTCGACGCGCTGCGCGAGAGGATCGTCGCGGACGGCGGCCGGGCCAAGCGGCTGCCGGTGGACTACG
CCTCGCACTGCGCCCATGTCGAGTCGATCCGCGAACGGCTGCTCACCGACCTCGCGGGCGTACGGGCCCGGGGGGCCGACGTACCGT
TCTACTCCACCGTCACCGGTGCAGTGCTGGACACCACCGCGCTGACCGCCGACTACTGGTACACGAACCTGCGCCGGAGCGTGTTGT
TCGAGCCGACCACCCGGGCCCTGCTCGATTCCGGATACGGGATCTTCGTCGAGTGCAGCCCGCACCCGGTGCTGCTGAACAGCATCG
AGGAGACCGCCGACGCCGTGGGCGCGACCGTCACCGGGCTGGGCTCGCTGCGCCGCGACGACGGCGGGGCCGAGCGCCTGCTCACCT
CGCTCGGCGAGGCGTTCGTGGCGGGTGTCCCGGTCGACTGGTCGGCGGTGTTCACGGGCATGCCGGTGCGCGCCGCCGATCTGCCCA
CGTATCCCTTCCAGCGCGAGCGCTACTGGCTGGGCCGGTCCGCGGCCTCCGGCGACGTCACCGCCGCCGGCTGCGGGCCACCACCC
ATCCGCTGCTGGGCGCGGCAGTCCAGGTCGCCGGGGGCGGCACCCTGTTCACCGGCCGGCTCTCCGTGTCCACCACGCCCTGGCTGG
CCGACCACGCGGTCTCGGGCACCCCCCTGCTGCCCGGCACCGCGCTGGTGGAGCTGGCGCTGAGCGCGGGCCACGAACTCGGGTACG
GGCACGTCGCCGAACTCACCCTCCAGGCGCCGCTGGTGCTGCCCGGCCGGGCGGCGGTCCAGTTCCAGGTACACGTGGCCGCCGCCG
ACGAGGACGGCCACCGCGCGCTGACCGTCCACTCCCGCCCCGAGGGCGCCGACGACACCGAGTGGACCGCGCACGCCACCGGGCTGC
TCGCCCCGCGGACCGCCCCGCCCGGCTTCGACCTCGCGCAGTGGCCGCCCCGGGGCGCGGAACCGGTGCTGGTGGACGACGCCTACG
ACACGCTGGCCGCGCTCGGCTACGACTACGGGCCCGCCTTCCAGGGCCTGCGCGCGGTCTGGCGGCGTGGCGACGAGACCTTCGCCG
AGGTCGAACTCCCCGGTGAGGCAGGTGCGTTCGGCCTGCACCCGGCCCTGTTCGACGCGGCCCTGCACGCCGACGGCCTGCGCACGG
CCCCGCCCGGCACCGACGGCCCCGGGGCGCGGGGGCAGGGGGCGGCGCGGCTGCCCTTCGTCTGGACCGGCGTGTCGTTGTATGCGT
CCGGGGCCACCGCCCTGCGGGTCCGCATCCGGGGCGGCGACACGCTCTCCCTGGACCTGGCCGACCCGACCGGCGCACCGGTCGCCG
CCGTGGAGGCCCTGGTCTCCAGGCCGGTCGACCCGGCGGCGCTGACCTCCCCGGTCCGGGACGACGACCTGTACCGGCTGGACTGGC
AGGCGCTGCCCGTACCCGTGGCGGACGCGCCCGCGTACGCCGTTCTGGACGAGCGGGGCACGGCCGCGGCGGACGCCGTGCCGGACT
GGGTGGTCCTGCCGGTGAGCGGTGACGGCGGCGACCCGGTGGGCGGGGTGCGCGCGGCGACCGGGCGGGTCCTCGCCGCCGTGCGCG
ACTGGCTGGCGGACGAGCGTACGGCCGGGGCCCGGGGGGCCCGGCTGGTGGTCCTGACCGGCGGCGCGGTCGCCACCGGCACGGAGG
ACGTCACCGACCTGGCGGGTGCCGCCGTATGGGGCCTGGTCCGGGCGGCCCAGGGCGAACACCCCGACCGCTTCGTCCTGGTGGACT
CCGTCGCCCACGACGGCGGCGGCGAAAGTGCCTCCGGCCCGGGTGTCTTTGCCACCGACCGGGTCACCGAGGCCGTGCGCGCCGCG
CGGCGAGCGGCGAACCGCAGCTGGCCCTGCGCGAGGGCACCGTACGGGTACCCCGGCTGGCCCGTGCCGCCGTAACGGGAACGGCCG
CCGTACCCGCTTTTGACGGCCCCGCGCCGGATCCTCACGGCACCGTGCTCATCACCGGCGGCACGGGAGTGCTCGGTGCCGTGGTCG
CCCGGCACCTGGCCACCGAGCACGGGGTGCGCCGTCTCGTCCTGGCCGGCCGCAGCGGCACCGCCTTCGACGACTTCGGCGATCTCG
CCGAACGCGGCACCGAGGTCGTCGTCGCCCGCTGCGACGCCGCCGAACGCGACCAACTGGCCGCGCTGCTGGCCGACATGCCCGCGG
AGCGCCCGCTGACCGCGGTGATCCACCTCGCCGGGGTCCTGGACGACGGACTGGTGACCGATCAGACACCCGGGCGACTGGACGCCG
TCCTGCGGCCCAAGGCGGACGCCGCCTGGAACCTGCACGAGCTGACCCGTGACCTGGACCTGTCGGCGTTCGTCCTCTTCTCCTCGG
CCGCGGGCACGATCGACGGCGCGGGCCAGTCCGGGTACGCCGCCGCCAACGCCTTCCTCGACGGCCTGGCCGCCCACCGCGCCGCCC
AGGGCCTGCCCGCGCTCTCCCTCGCCTGGGGCTTCTGGGAGCAGCGCACCGGGATGACCGCCCACCTCACCGACGCCGACGTGGAGC
GCATGGCACGTGCCGGGGTCCGGCCCCTGCCCACCGAGGAGGGGCTGAGGCTGCTGGACGCCGCGCTCGCCGCCGACGTACCGCTGC
TGCTGCCCGTCGGCCTGGACCCGCGCGCCCTGCGCGGTGCCGACGACGTCCCGCCCGTGCTTGCGCGCTCTGGCGCCCGCGCCCGTC
CGTCGTACGGCGGCCTCCCGCGCCACCGCCGTTCCGCCGCCGAACGGCTGGCCGCCCTCGGCGCCGCCGAACGCGAGGCGGCGCTCA
CGGAGCTGGTCCGCACCCATGTCGCGGCCGTTCTCGGGCACGGCGCGGACATGGTGCTCGACCCGCGCCGCTCCTTCCGCGAGGCCG
GTTTCGACTCGCTGACCGCGGTCGAGCTGCGCAACCGCCTCGGAAACGCCGTCGGCCTCCGGTTGCCCGCCACCCTCGTCTTCGACC
ACCCCGACGCCGAGGCCCTGGTCAGGTACCTGAAGACGGAACTCTTCGGCGCGGACCCCGAGGACGCCGAGGCCTCCACCGGGATCG
GGGCCGTCGTCCCCGGAGCGGGGTACGAACCGGACGAGCCGGTGGCGATCGTCGGGATGGCGTGCCGCTACCCCGGCGGCGTCACCA
CGCCCGAGGAGCTGTGGCGGCTCGTCGCGGACGGCGTGGACGGCATCGGCGCGTTCCCCGACGACCGGGGCTGGAACCTCGACACCC
TGTACGACCCGGAGCCCGGCAAGCCCGGCCACTGCTCCACCCGCGCGGGCGGATTCCTCTACGACGCCGCCGACTTCGACCACGACT
TCTTCGGCATCGGCCCCCGCGAGGCCCTCGCCATGGACCCGCAGCAGCGGTTGCTGCTGGAGACCTCCTGGGAGGCGCTGGAACGGG

CCGGCATCGATCCGCACTCCGTGCGCGGCAGCCGCACCGGCGTGTTCGCCGGGGTCATGTACCACGACTACGGCAGCAGGCTGCGCG
ACGTCCCCGAGGCCGTGCGCGACTACCTCGGCAACGGAAGCCTCGGCAGTATCGCCTCCGGCCGTATCGCCTACACCCTGGGTCTGG
AGGGCCCGGCGCTCACCGTGGACACGGCCTGCTCCTCGTCGCTGGTGGCGCTGCACCTGGCGGCGCAGGCACTGCGGCGGGGGGAGT
GCGGCCTGGCCCTGGCCGGTGGCGTGTCCGTGATGTCGACCGTCGACACGTTCGTGGACTTCAGCAGGCAGCGCAACCTCGCCGCCG
ACGGCCGCGCCAAGTCCTTCGCCGAGGCGGCGGACGGCACGGCGCTGTCCGAGGGCGTCGGTGTGTTGGTGTTGGAGCGGTTGTCGG
ATGCGCGGCGGTCGGGGCGTCGGGTGTGGGGGGTGGTGCGGGGTTCGGCGGTGAATCAGGATGGTGCGTCGAATGGGTTGACGGCGC
CGAATGGTCCGGCGCAGCAGCGGGTGATTCGTGAGGCGTGGGTGGCTGCGGGTGTGTCGGGTGGTGGGGTGGATGTGGTGGAGGCGC
ATGGGACGGGGACGGTGTTGGGTGATCCGATCGAGGCGCAGGCGTTGTTGTCTACGTACGGGCAGGGGCGTGGGGGTGGGGATCC

Figure 23: Nucleotide sequence of clones E27, C170 and C2. Clone E27 is represented in brown. Clone C170 is represented in blue and clone C2 is represented in green. The overlapping sequence of clones E27 and C170 is represented in maroon. The overlapping sequence of the clones C170 and C2 is represented in turquoise.

4.12.2 SEQUENCE ANALYSIS OF OVERLAPPING CLONES C2, C170 AND E27

The cloned nucleotide sequence of 11656 nucleotides from the *Streptomyces* sp. 98- 62 has a high G+ C content of 75.3%, which is typical of strains of actinomycetes (Wright, 1992). The sequence was compared with the available nucleotide sequences from the Genbank/ EMBL databank by using BLAST search, which revealed significant similarities with other PKS genes from several *Streptomyces* sp. and other actinomycetes. The highest score of similarity was obtained with *Streptomyces* genes encoding the synthesis of the polyketides (Fig. 25).

Sequences producing significant alignments:	(bits)	Value
gi|8050835|gb|AF263912.1|AF263912 Streptomyces noursei ATCC...	432	e-117
gi|14794889|gb|AF357202.1|AF357202 Streptomyces nodosus amp...	365	8e-97
gi|12055067|emb|AJ278573.1|SNA278573 Streptomyces natalensi...	331	1e-86
gi|21449342|gb|AF453501.1| Actinosynnema pretiosum subsp. a...	305	6e-79
gi|3808326|gb|AF079138.1|AF079138 Streptomyces venezuelae m...	283	2e-72
gi|2558836|gb|AF016585.1|AF016585 Streptomyces caelestis cy...	278	1e-70
gi|15824136|dbj|AB070949.1| Streptomyces avermitilis polyen...	276	6e-70
gi|20520686|emb|AL591083.2|SC1G7 Streptomyces coelicolor co...	274	2e-69
gi|12231153|emb|AJ300302.1|SGR300302 Streptomyces griseus p...	264	2e-66
gi|2317859|gb|U78289.1|SFU78289 Streptomyces fradiae tylact...	260	3e-65
gi|4678702|emb|AJ132222.1|SNA132222 Streptomyces natalensis...	260	3e-65
gi|20520683|emb|AL512902.2|SC2C4 Streptomyces coelicolor co...	258	1e-64
gi|15823967|dbj|AB070940.1| Streptomyces avermitilis oligom...	248	1e-61
gi|21999182|gb|AY118081.1| Streptomyces sp. GERI155 putativ...	230	3e-56
gi|9049534|gb|AF220951.1|AF220951 Streptomyces antibioticus...	208	1e-49
gi|153407|gb|L09654.1|STMPKS3ORF Streptomyces antibioticus ...	208	1e-49
gi|9280381|gb|AF235504.1|AF235504 Streptomyces hygroscopicu...	200	3e-47

Figure 25: Sequence comparison of the 11.6 kb of cloned genes from the *Streptomyces* sp. 98- 62 with the Genbank sequences

The nucleotide sequence of the 11656kb fragment was analysed for open reading frames (ORFs), using the open reading frame finder at http://www.ncbi.nlm.nih.gov/. Two open reading frames spanning the 11.7 kb were elucidated. The two ORFs read in the same direction as the genes encoding the PKS domains. The ORFs are named ORF1 and 2 for convenience (Fig. 29). It is deduced that the ORF1 terminates with a stop codon TGA. A second stop codon TAG is predicted 372 nucleotides downstream of the first stop codon. The ORF2 is predicted to use ATG as the start codon, which occurs 60 nucleotides downstream of the

predicted second stop codon of ORF1. The sequence TGGACA, which is located 38 nucleotides upstream of the predicted start codon of ORF2 is deduced to be the transcriptional promoter as the sequence is identical to ermE-P1 promoter (Strohl, 1992). The sequence GAGG, which is located 14 nucleotides upstream of the predicted start codon of ORF2 is deduced to be the ribosomal binding site of ORF2 (Strohl, 1992). As such, it is predicted that ORF1 and ORF2 are probably translationally uncoupled.

The genes of the ORFs occur in a repeated modular fashion, as is characteristic of the PKS I genes of other actinomycetes. Three modules were identified in the 11656bp in the PKS genes of the *Streptomyces* sp. 98- 62. The modules are labelled module 1 to 3 for convenience in the order of their positions. (Fig.27, 28). The sequence data from the available clones only reveal a part of modules 1 and 3. ORF1 appeared to encompass at least one module, designated as module1 for convenience and ORF2 appeared to encompass at least two modules, designated as modules 2 and 3 for convenience. There was no stop codon or start codon observed in the intermodular region of module 2 and module 3. Therefore it is predicted that modules 2 and 3 are translationally coupled.

The limits of each domain within the modules were readily assigned by comparison with the modules of B-deoxyerythronolide synthase and rapamycin synthase (Fig. 27, 28) (Bevitt, 1992, Molnar, 1996). Module1 was found to encode enzymatic domains KS, AT, DH, KR and ACP. The KS domain of module 1 within the 11.7kb of PKS genes is only partial. Module 1 has the highest homology to pimaricin producer *S.natalensis.* The percentage positives at the amino acid level was 57%. The domains within module 2 also occur in the order characteristic of PKS I genes. Module 2 was found to encode the complete enzymatic domains KS, AT, DH,

KR and ACP, in the stated order. Module 2 has highest homology to avermectin producer *S.avermitilis.* The homology at the amino acid level was 72%. Module 3 within the 11.6 kb of PKS genes encodes the N terminal portion of the KS domain. This short region of the KS domain has the highest homology to the KS gene of the avermectin producer *S.avermitilis.* The percentage positives at the amino acid level was 80%.

The nucleotide sequence of the cloned putative PKS I genes, repeated occurrence of the genes as modules and domains within the modules and the organization of the PKS I genes was similar to other PKS I systems of streptomycetes. These results give strong evidence that PKS type I genes have been cloned from the novel anti fungal compound producer, the *Streptomyces* sp. 98- 62.

```
DNA: GCATGCTCTTTGNNTAACGGTTCTCCGACGCCCGTCGCAACGGNCACCGGG
 +3:   M  L  F  X  *  R  F  S  D  A  R  R  N  G  H  R  V
DNA: TCCTGGCCGCGGTCCGTTNTTCCGCCGTCAACTCCGACGGCGCGTCCAACG
 +3:   L  A  A  V  R  X  S  A  V  N  S  D  G  A  S  N  G
DNA: GGCTGACCGCCCCCAACGGGCCCTCCCAGCAACGCGTCATCCGCGCCGCGC
 +3:   L  T  A  P  N  G  P  S  Q  Q  R  V  I  R  A  A  L
DNA: TCGCCGCCGCCCGCCTCGCCCCGGCCGATGTCGACGCGGTCGAGGCGCACG
 +3:   A  A  A  R  L  A  P  A  D  V  D  A  V  E  A  H  G
DNA: GCACCGGCACCACGCTCGGCGACCCGATCGAGGCGCAGGCGCTGCTGGCCA
 +3:   T  G  T  T  L  G  D  P  I  E  A  Q  A  L  L  A  T
DNA: CGTACGGCCAGGACCGGCCGGGCGACGAACCCCTCTGGCTCGGCTCCGTCA
 +3:   Y  G  Q  D  R  P  G  D  E  P  L  W  L  G  S  V  K
DNA: AGTCCAACATGGGCCACACCCAGGCCGCCGCCGGGGTGGCCGGAATCATCA
 +3:   S  N  M  G  H  T  Q  A  A  A  G  V  A  G  I  I  K
DNA: AGATGGTCATGGCGATGCGGCACGGCACCCTGCCCCGCACCCTGCACGTCG
 +3:   M  V  M  A  M  R  H  G  T  L  P  R  T  L  H  V  D
DNA: ACACGCCCTCCCACCAGGTCGACTGGACGACGGGCGCGGTCCGCCTGCTCA
 +3:   T  P  S  H  Q  V  D  W  T  T  G  A  V  R  L  L  T
DNA: CGGAGGAGCGGCCCTGGCCGGGAGCGGCGGACCGTCCGCGCCGGGCGGGGG
 +3:   E  E  R  P  W  P  G  A  A  D  R  P  R  R  A  G  V
DNA: TGTCCTCGTTCGGGATCAGCGGCACCAACGCCCATGTGATTCTTGAGGAGT
 +3:   S  S  F  G  I  S  G  T  N  A  H  V  I  L  E  E  F
DNA: TCGAGGAGTTCGAGGAGTTCGCGGGGGAGCCGGTCGGGACGGGGCCGCGGA
 +3:   E  E  F  E  E  F  A  G  E  P  V  G  T  G  P  R  T
DNA: CCGCCGGTCCGGACGCCGACGGGCACGACGGTGCGGCAGCGCACCCTCCCG
 +3:   A  G  P  D  A  D  G  H  D  G  A  A  A  H  P  P  A
DNA: CCACGCCGCCCGTACTCGCCCTTCCGGTCTCCGCCCGCTCACCCGAGGCCC
 +3:   T  P  P  V  L  A  L  P  V  S  A  R  S  P  E  A  L
DNA: TGCGCGGCCAGGCGGCCCGCCTGCGGGAACTGACCGGCACCTCGGCCGCCG
 +3:   R  G  Q  A  A  R  L  R  E  L  T  G  T  S  A  A  E
DNA: AACTCGGCCTCGCCCTGTCCACCACCCGCACCACCCACCCGTACCGCGCCG
 +3:   L  G  L  A  L  S  T  T  R  T  T  H  P  Y  R  A  V
DNA: TCGTCCTCGCCCCCGGTGAGGAGCGGGCCGACGAGGCCCTGGACGCCCTCG
 +3:   V  L  A  P  G  E  E  R  A  D  E  A  L  D  A  L  A
DNA: CCCACGGGCACGAGGCACCCGGCCTGCTCGTCAGCGGTTCCATCACCGACG
 +3:   H  G  H  E  A  P  G  L  L  V  S  G  S  I  T  D  G
DNA: GCACCCTGGCCTGTCTGTTCTCCGGGCAGGGCGCCCAGCGGCCCGGCATGG
 +3:   T  L  A  C  L  F  S  G  Q  G  A  Q  R  P  G  M  G
DNA: GCCGGGACTGGTACGACACCTTCCCGGTCTACGCGGAGCACTTCGACCGCA
 +3:   R  D  W  Y  D  T  F  P  V  Y  A  E  H  F  D  R  T
DNA: CGGGCGAACTCTTCGCCAAGCACCTGGAACGGGCGCTCGCCGAAGTGGTCC
 +3:   G  E  L  F  A  K  H  L  E  R  A  L  A  E  V  V  L
DNA: TGGGCGACCACCCCGACGTACTGGAACGGACCGCCTACACCCAGGCCGCCC
 +3:   G  D  H  P  D  V  L  E  R  T  A  Y  T  Q  A  A  L
DNA: TCTTCACCACCCAGGTCGCCCTCTACCGACTGCTGGAGTCCTTCGGGCTGC
 +3:   F  T  T  Q  V  A  L  Y  R  L  L  E  S  F  G  L  R
DNA: GGCCCGACTGGCTGGCCGGCCACTCCGTCGGCGAGTTCGCCGCCGCGCACG
 +3:   P  D  W  L  A  G  H  S  V  G  E  F  A  A  A  H  V
DNA: TCGCCGGTGTGTGGTCGCTCCAGGACGCCGTCACCGCCGTCGCGGCGCGCG
 +3:   A  G  V  W  S  L  Q  D  A  V  T  A  V  A  A  R  G
DNA: GCAGGCTCATGCAGGCGCTTCCCGAGGGCGGTGCGATGACCGCCGTACAGG
 +3:   R  L  M  Q  A  L  P  E  G  G  A  M  T  A  V  Q  A
DNA: CCGCCGAGGAGGAGGTGCGGCCGCTGCTGGACGAACGGTGCGACATCGCCG
 +3:   A  E  E  E  V  R  P  L  L  D  E  R  C  D  I  A  A
DNA: CGGTCAACGGCCCGCGCGCCGTGGTCGTCTCCGGGGACGAGGACGCCGTCG
 +3:   V  N  G  P  R  A  V  V  V  S  G  D  E  D  A  V  A
DNA: CCGCCGTCGCCGCGCACTTCGCCACCACCCGGCGACTGCGCGTCTCGCACG
 +3:   A  V  A  A  H  F  A  T  T  R  R  L  R  V  S  H  A
DNA: CCTTCCACTCGCCGCGCATGGAACCCGTGCTGGACGAGTTCCGCCGGGTCT
 +3:   F  H  S  P  R  M  E  P  V  L  D  E  F  R  R  V  L
DNA: TGGCCGCCCTGCCGGCCGGGGAACCGGCCCTGCCGATCGTCTCCACCCTCA
```

```
 +3:   A  A  L  P  A  G  E  P  A  L  P  I  V  S  T  L  T
DNA: CCGGCGCCCGGGCCACCGCCGCCGAACTCGGCTCCGCCGACTACTGGGTAC
 +3:   G  A  R  A  T  A  A  E  L  G  S  A  D  Y  W  V  R
DNA: GGCACGTACGGGAGACCGTCCGCTTCGCCGACGCCGTGGGGACGCTGGCCG
 +3:   H  V  R  E  T  V  R  F  A  D  A  V  G  T  L  A  A
DNA: CGCAGGGCGCCGACACCTTCCTCGAACTCGGCGCCGCTCCCGTCCTGACGG
 +3:   Q  G  A  D  T  F  L  E  L  G  A  A  P  V  L  T  A
DNA: CCCTCGGCCCGGACTGCCTCCCGGACGCGGACGCCGAGGAGGCCGCGTTCG
 +3:   L  G  P  D  C  L  P  D  A  D  A  E  E  A  A  F  V
DNA: TCCCCACCGCCCGCAAGGGCACCGCCGAGGTGCCCGGTCTGCTGGCCGCCC
 +3:   P  T  A  R  K  G  T  A  E  V  P  G  L  L  A  A  L
DNA: TGGCCGCCGTGCACACCCGCGGTTCGGACGTCGACTGGGCGGTCCTCTACG
 +3:   A  A  V  H  T  R  G  S  D  V  D  W  A  V  L  Y  D
DNA: ACGGCCTCCCCGGGCACCGCGACCGACCCGGGCGCCGCGACGAACCCGGGC
 +3:   G  L  P  G  H  R  D  R  P  G  R  R  D  E  P  G  H
DNA: ACCGCGACCAACCGGGGCGCCGTGACCAACCGGGGCGCCGCGTCGAACCGG
 +3:   R  D  Q  P  G  R  R  D  Q  P  G  R  R  V  E  P  G
DNA: GGCGTTGTGTCGAGCTGCCTACCTACGCCTTCCAGCACCGCCGCTACTGGC
 +3:   R  C  V  E  L  P  T  Y  A  F  Q  H  R  R  Y  W  L
DNA: TTCCCACGTCCACCGCCACCGCCAGGGGCGACGCTGCCGGTCACGGTCTCG
 +3:   P  T  S  T  A  T  A  R  G  D  A  A  G  H  G  L  A
DNA: CGGCCGTCGACCACCCCTTCGTCAGCGCCCGCCTCGACCTGCCGGGCGACG
 +3:   A  V  D  H  P  F  V  S  A  R  L  D  L  P  G  D  G
DNA: GCGGAACCCTGCTCACCGGCCGGATCTCCACCGCCACCCACCCGGTGCTCG
 +3:   G  T  L  L  T  G  R  I  S  T  A  T  H  P  V  L  A
DNA: CCCAGCACGCCGTGCTCGGATCGGTGCTGGTGCCCGGCGCCGCCCTCGTCG
 +3:   Q  H  A  V  L  G  S  V  L  V  P  G  A  A  L  V  D
DNA: ATCTCGCCCTGTACGCAAGTGGGTTGACGGGACGCCCGGTGCTGGAGGAAC
 +3:   L  A  L  Y  A  S  G  L  T  G  R  P  V  L  E  E  L
DNA: TCACCCTCCAGGCCCCGTTGGCCCTGCCCGGGAACGGTGCCGTACGGATCC
 +3:   T  L  Q  A  P  L  A  L  P  G  N  G  A  V  R  I  Q
DNA: AGGTCGCGCTCCGGCCCGACGGCGGTGTGGAGATCCACTCCCGGCCCGCCG
 +3:   V  A  L  R  P  D  G  G  V  E  I  H  S  R  P  A  D
DNA: ATGCGCCCGAGGACGGGAGCTGGACCCGGCACGCCACCGGCACCCTCACCG
 +3:   A  P  E  D  G  S  W  T  R  H  A  T  G  T  L  T  V
DNA: TCACCGACCCCGCCTCCGGACTTCCCGCGTCGTCCGTTCCGTCCGCCGCCT
 +3:   T  D  P  A  S  G  L  P  A  S  S  V  P  S  A  A  W
DNA: GGCCGCCGCCGGGTGCCGTGCCGCTCGACACCGACGGCCTCTACGAGCGGC
 +3:   P  P  P  G  A  V  P  L  D  T  D  G  L  Y  E  R  L
DNA: TGCGCGGCGAGGGTTACGACTACGGCCCCGTCTTCCAGGGCGTACGGGCCG
 +3:   R  G  E  G  Y  D  Y  G  P  V  F  Q  G  V  R  A  A
DNA: CCTGGCGGCACGGCGACACGGTCCTCGCCGAACTCGAACTGCCCGCCGAGG
 +3:   W  R  H  G  D  T  V  L  A  E  L  E  L  P  A  E  A
DNA: CCCGGCAGGACGCCGCCCGGCACGTCCTGCACCCCGCGCTGCTGGACTCCG
 +3:   R  Q  D  A  A  R  H  V  L  H  P  A  L  L  D  S  A
DNA: CCCTGCACACCACCGCCCTCGCCGACGCGGACGCCCGCGACGCGGTACCGG
 +3:   L  H  T  T  A  L  A  D  A  D  A  R  D  A  V  P  D
DNA: ACGGCACGATCGCCCTGCCCTTCGCCTGGACCGGTGTCACCGTGCACGGAC
 +3:   G  T  I  A  L  P  F  A  W  T  G  V  T  V  H  G  R
DNA: GGCCGTCGTCACGTACCACCCCGTCCCGCACGGGCGTCCCCTCCCGCGCAG
 +3:   P  S  S  R  T  T  P  S  R  T  G  V  P  S  R  A  A
DNA: CCGCCCCGGACCACACCGCAGCCCGGGTCCGCGTCACCCGGGGCGAGGAGG
 +3:   A  P  D  H  T  A  A  R  V  R  V  T  R  G  E  E  G
DNA: GCATCCGGCTCGATCTGACGGACACCGAGGGCGGGCCGCTGGCCACTGTCG
 +3:   I  R  L  D  L  T  D  T  E  G  G  P  L  A  T  V  A
DNA: CGTCCTACGTCACCCGCCCCGTCACCGCCGACCGGCTCACCGGGCGGCAGC
 +3:   S  Y  V  T  R  P  V  T  A  D  R  L  T  G  R  Q  R
DNA: GTTCCCTGTACGTCGTCGAGGACGCGCCGCTCCCCGAGTCCGCCGGGCGCC
 +3:   S  L  Y  V  V  E  D  A  P  L  P  E  S  A  G  R  P
DNA: CCGAGCGCCGCACCTGGGCCGTGCTGGGCCCGGACGACCTCGGACTCGGCG
 +3:   E  R  R  T  W  A  V  L  G  P  D  D  L  G  L  G  V
```

```
DNA: TCCCGCACCACCCCGAACCGGCCGCGATCGACGGCCCCGCACCCGACGTCG
 +3:   P  H  H  P  E  P  A  A  I  D  G  P  A  P  D  V  V
DNA: TCGTCCTTCCGGTGCACATCCCGGACGTCGCCGACGCGGACGCCGACGGCG
 +3:   V  L  P  V  H  I  P  D  V  A  D  A  D  A  D  G  E
DNA: AACGGGTGCCGGGGGCCGTGCGTACCGCGCTGAACACGACGCTCACGACCC
 +3:   R  V  P  G  A  V  R  T  A  L  N  T  T  L  T  T  L
DNA: TCCGGGCCTGGCTGGACGACGAACGCCGGGCCGGTTCCACGCTGCTGGTGC
 +3:   R  A  W  L  D  D  E  R  R  A  G  S  T  L  L  V  L
DNA: TCACCGAGGGAAGCCTCGCCGACGCCGCCGTGCACGGACTGGTGCGGGCCG
 +3:   T  E  G  S  L  A  D  A  A  V  H  G  L  V  R  A  A
DNA: CGCAGGCCGAACACCCGGGCCGGATCGTCCTTGTCGGCCGGGCCGGGCCCG
 +3:   Q  A  E  H  P  G  R  I  V  L  V  G  R  A  G  P  G
DNA: GCAGCCCCGTCCCGGACCGCGCAGCGCTGGCCGCCGTCCTCGACTCCGGTG
 +3:   S  P  V  P  D  R  A  A  L  A  A  V  L  D  S  G  E
DNA: AACCGGAGGTGCGGTGGCGGGACGGCCGGGCCCACGCCCCGCGCCTGGTGC
 +3:   P  E  V  R  W  R  D  G  R  A  H  A  P  R  L  V  R
DNA: GCGCCGGGGAGCCGGACGCGCCGCGCACCGGGCGCCCTGGGGCACCGTCC
 +3:   A  G  E  P  D  A  P  R  T  G  R  P  W  G  T  V  L
DNA: TGATCACCGGCGGCACCGGCGGGCTCGGCGCCCTGGTGGCCCGGCACCTGG
 +3:   I  T  G  G  T  G  G  L  G  A  L  V  A  R  H  L  V
DNA: TGACCCGGCACGGCGTCACCCGCCTGATCCTGGCGGGCCGTCGCGGACCCG
 +3:   T  R  H  G  V  T  R  L  I  L  A  G  R  R  G  P  A
DNA: CCGCCCCGGGCGCCGACGAACTGCGCGCGGACCTGGCCGGCCTGGGCGCCC
 +3:   A  P  G  A  D  E  L  R  A  D  L  A  G  L  G  A  Q
DNA: AGGCCGATGTCGTCGCCTGCGACGTCGCCGACCGCACGGCGCTCGCCGCGC
 +3:   A  D  V  V  A  C  D  V  A  D  R  T  A  L  A  A  L
DNA: TGCTGGCCGCCCACCCCGTCGACAGCGTCGTGCACACCGCGGGCGTCCTGG
 +3:   L  A  A  H  P  V  D  S  V  V  H  T  A  G  V  L  D
DNA: ACGACGGACTGGTCACCTCGCTCGGCCCCGAACGCCTGGACACGGTCCTGC
 +3:   D  G  L  V  T  S  L  G  P  E  R  L  D  T  V  L  R
DNA: GCCCCAAGGCGGACGCCGCCTGGCACCTGCACGAACTGACCCTCGACCGGC
 +3:   P  K  A  D  A  A  W  H  L  H  E  L  T  L  D  R  P
DNA: CGCTGTCCCACTTCGTGCTGTTCTCCTCGGCAGCGGGCACCATCGACGCCT
 +3:   L  S  H  F  V  L  F  S  S  A  A  G  T  I  D  A  S
DNA: CCGGCCAGGGCAACTACGCCGCCGCCAACGTCTTCCTCGACGCCCTGGCAG
 +3:   G  Q  G  N  Y  A  A  A  N  V  F  L  D  A  L  A  V
DNA: TCCACCGTGCCGCCCGGTACCTGCCGGCGCTCTCCCTCGCCTGGGGCCTGT
 +3:   H  R  A  A  R  Y  L  P  A  L  S  L  A  W  G  L  W
DNA: GGTCCGGTGGCGGCATGGGAGCCGGCCTCGACGAGAGCGGCGCCCGGCGCA
 +3:   S  G  G  G  M  G  A  G  L  D  E  S  G  A  R  R  I
DNA: TCGAACGGTCCGGCATCGGCGCCCTCGACCCGGAGGAGGGCCTCGAACTCT
 +3:   E  R  S  G  I  G  A  L  D  P  E  E  G  L  E  L  F
DNA: TCGACGCCGCCGTGGCGTCCGGCCGCCCCGCCCTGGTGCCGGTCCGGCTGG
 +3:   D  A  A  V  A  S  G  R  P  A  L  V  P  V  R  L  D
DNA: ACACCACCGTGCTGCGCCGCCGGGGGCGACGACGTACCGCCGGTGCTGCGCA
 +3:   T  T  V  L  R  R  R  G  D  D  V  P  P  V  L  R  T
DNA: CCCTGGCCGGTGTCACCGCCCCCGCCGCACGGGAGGACCGGACCCGCGGCC
 +3:   L  A  G  V  T  A  P  A  A  R  E  D  R  T  R  G  L
DNA: TCGGCGAGCGCCTGGCCGCCCTGCCCGCCGCCGACCACGAGCACACCGTGC
 +3:   G  E  R  L  A  A  L  P  A  A  D  H  E  H  T  V  L
DNA: TGGAGGCCGTCCGTACCGAGGTCGCCGCCGTCCTCGGCCACGACGGACCCG
 +3:   E  A  V  R  T  E  V  A  A  V  L  G  H  D  G  P  A
DNA: CCGCGGTCGGGCCTCGGCGCGCTTTCACCGAGCTGGGATTCGACTCGCTCG
 +3:   A  V  G  P  R  R  A  F  T  E  L  G  F  D  S  L  A
DNA: CCGCGGTCGAACTGCGCAACCGGCTCAACGCGATCAGCGGACTGCGCCTGC
 +3:   A  V  E  L  R  N  R  L  N  A  I  S  G  L  R  L  P
DNA: CGTCGACGCTCGTCTTCGACTACGCCACTCCCGTGGCGCTGGCGGGCCATC
 +3:   S  T  L  V  F  D  Y  A  T  P  V  A  L  A  G  H  L
DNA: TGCTCGAACGGCTAGCCCCGGACGACGACACCGGCACCGGTGCGGCGCCCA
 +3:   L  E  R  L  A  P  D  D  D  T  G  T  G  A  A  P  T
DNA: CCGACCCGAGGGGCGACGACGAGGTGCGGGCCCTCATCGACCGCATCCCGA
```

```
 +3:   D  P  R  G  D  D  E  V  R  A  L  I  D  R  I  P  I
DNA: TCGCGCGCATCCGCGACGCCGGACTGCTCGACGGGCTGCTGAGACTGTCCG
 +3:   A  R  I  R  D  A  G  L  L  D  G  L  L  R  L  S  E
DNA: AAGCGGCCCCGCCCGCACCGCCCGCCGCCGACCGGGTCATGGACATCAGGT
 +3:   A  A  P  P  A  P  P  A  A  D  R  V  M  D  I  R  S
DNA: CCATGGGGGTGGCCGATCTGGTGCGAGCCGCGCTGAACCGCACCAGCCCCG
 +3:   M  G  V  A  D  L  V  R  A  A  L  N  R  T  S  P  E
DNA: AGTGAGACCGCCCCGGTGCCGGACCGCGCGGACCACGCGCCGCGTCCGGAA
 +3:   *  D  R  P  G  A  G  P  R  G  P  R  A  A  S  G  T
DNA: CCGGCCGCACATCCGGCCCGCACATCCGGCCGGTACGACCGGCCGCACATG
 +3:   G  R  T  S  G  P  H  I  R  P  V  R  P  A  A  H  G
DNA: GTCGGCCCGTACGACCGCCCATACGGCCGGAGCACCTCAGCCGTACCTCCG
 +3:   R  P  V  R  P  P  I  R  P  E  H  L  S  R  T  S  A
DNA: CAGCACCTCCAGCGCGTCCCCCGCACATCCGCACCGCGTCACCAGCGCCGC
 +3:   A  P  P  A  R  P  P  H  I  R  T  A  S  P  A  P  P
DNA: CGAGTCAGCCAGTGCTGCGACGGGAAAGGTTCACCGGCTCGCGACGCCCGG
 +3:   S  Q  P  V  L  R  R  E  R  F  T  G  S  R  R  P  A
DNA: CACGGCACCTCGCCGCGTTGTCGCATCACCCGAGTACGTCCCGTACGAGGG
 +3:   R  H  L  A  A  L  S  H  H  P  S  T  S  R  T  R  A
DNA: CGATCCTCCGCCTTACGACGCACCGCACCGGACGCCGCGAGCTTCCCGGCA
 +3:   I  L  R  L  T  T  H  R  T  G  R  R  E  L  P  G  K
DNA: AACCCTTCCGGCCACAGCACTAGGGAGCGATACCGACCGTGGACACATCCG
 +3:   P  F  R  P  Q  H  *  G  A  I  P  T  V  D  T  S  V
DNA: TCGAGCAGATCGTCGAGGCGCTGCGCGAGGCCATGCTCGAGAACGAGCGGC
 +3:   E  Q  I  V  E  A  L  R  E  A  M  L  E  N  E  R  L
DNA: TGCGCCGGCAGAACGACCGGATCGCCGAAGCGGCGCACGAGCCCGTCGCCG
 +3:   R  R  Q  N  D  R  I  A  E  A  A  H  E  P  V  A  V
DNA: TCGTCGCCATGAGCTGCCGCTACCCCGGCGGCGTCGGCACGCCCGAACAGC
 +3:   V  A  M  S  C  R  Y  P  G  G  V  G  T  P  E  Q  L
DNA: TGTGGCAACTGGTCGACGCCGGAGTGGACGCCGTGGGCGACTTCCCGGACG
 +3:   W  Q  L  V  D  A  G  V  D  A  V  G  D  F  P  D  D
DNA: ACCGGGACTGGGACGTCGACGCCATCTACGATCCCGACCCCGACGCCCCCG
 +3:   R  D  W  D  V  D  A  I  Y  D  P  D  P  D  A  P  G
DNA: GCAGGACCCATGTGCGCGAGGGCGGATTCCTCCACGACGCACCGCGGTTCG
 +3:   R  T  H  V  R  E  G  G  F  L  H  D  A  P  R  F  D
DNA: ACCCGGGCTTCTTCGGTATCAGCCCGCGTGAGGCCCTCGCCATGGACCCGC
 +3:   P  G  F  F  G  I  S  P  R  E  A  L  A  M  D  P  Q
DNA: AGCAGCGGCTGCTGCTGGAGACCGCCTGGGAGGCGTTCGAACGCGGCGGCA
 +3:   Q  R  L  L  L  E  T  A  W  E  A  F  E  R  G  G  I
DNA: TCGACCCGCACACCCTGCGCGGCAGCCGCACCGGCATCTACGCCGGGGTCA
 +3:   D  P  H  T  L  R  G  S  R  T  G  I  Y  A  G  V  M
DNA: TGTACCACGACTACGGCAGCTGGCTCACCGACGTACCGGAGGGCGTCGAGG
 +3:   Y  H  D  Y  G  S  W  L  T  D  V  P  E  G  V  E  G
DNA: GCTACCTCGGCAACGGCAACCTCGGCAGCGTCGCCTCCGGCCGCGTCTCCT
 +3:   Y  L  G  N  G  N  L  G  S  V  A  S  G  R  V  S  Y
DNA: ACACGCTCGGCCTGGAGGGCCCCGCCGTCACCGTCGACACCGCCTGCTCCT
 +3:   T  L  G  L  E  G  P  A  V  T  V  D  T  A  C  S  S
DNA: CCTCGCTGGTCGCCCTCCACCTGGCCGTGCAGGCCCTGCGCACCGGCGAGT
 +3:   S  L  V  A  L  H  L  A  V  Q  A  L  R  T  G  E  C
DNA: GCGCCCTCGCCCTGGCCGGGGGCGTGACCGTGATGTCCACCCCGGACACCT
 +3:   A  L  A  L  A  G  G  V  T  V  M  S  T  P  D  T  F
DNA: TCATCGACTTCTCCCGCCAGCGCGGGCTCGCCCTGGACGGGCGCTGCAAGT
 +3:   I  D  F  S  R  Q  R  G  L  A  L  D  G  R  C  K  S
DNA: CCTTCGCGGAGGGCGCCGACGGCACCGGCTGGGGCGAGGGCGTCGGCATGC
 +3:   F  A  E  G  A  D  G  T  G  W  G  E  G  V  G  M  L
DNA: TCCTGCTGGAACGGCTCTCCGACGCCCGCCGCAACGGCCACCGCGTCCTCG
 +3:   L  L  E  R  L  S  D  A  R  R  N  G  H  R  V  L  A
DNA: CCGTCGTCCGCGGCACCGCCGTCAACCAGGACGGCGCCTCGAACGGGCTGA
 +3:   V  V  R  G  T  A  V  N  Q  D  G  A  S  N  G  L  T
DNA: CCGCGCCCAACGGCCCCTCCCAGCAACGCGTCATCCGCGCCGCGCTCGCCG
 +3:   A  P  N  G  P  S  Q  Q  R  V  I  R  A  A  L  A  D
```

```
DNA: ACGCCCGCCTGGAACCCCACCAGGTGCACGCCGTGGAGGCGCACGGCACCG
 +3:   A  R  L  E  P  H  Q  V  H  A  V  E  A  H  G  T  G
DNA: GCACCCCGCTCGGCGACCCCATCGAGGCCCAGGCCCTGCTCGCCACCTACG
 +3:   T  P  L  G  D  P  I  E  A  Q  A  L  L  A  T  Y  G
DNA: GGCAGGACCGGCAGGCCGGCGAACCGCTGTGGCTGGGCTCGGTCAAGTCCA
 +3:   Q  D  R  Q  A  G  E  P  L  W  L  G  S  V  K  S  N
DNA: ACATCGGGCACACCCAGGCCGCCGCCGGGGTCGCCGGCGTCATCAAGATGG
 +3:   I  G  H  T  Q  A  A  A  G  V  A  G  V  I  K  M  V
DNA: TGATGGCCATGCGCCGCGGCAGGCTGCCGAGGACGCTGCACGCCGAACACC
 +3:   M  A  M  R  R  G  R  L  P  R  T  L  H  A  E  H  P
DNA: CCACCACCCGGGTCGACTGGGAGTCCGGCGCCGTCGAACTGCTCGGCGAGG
 +3:   T  T  R  V  D  W  E  S  G  A  V  E  L  L  G  E  A
DNA: CCCGCGACTGGCCGGACGCGGGGGAGCCCCGCCGCGCCGCCGTGTCCTCCT
 +3:   R  D  W  P  D  A  G  E  P  R  R  A  A  V  S  S  F
DNA: TCGGCATCTCCGGCACCAACGCCCACGTCATCGTCGAGGCGGCCCCCGACC
 +3:   G  I  S  G  T  N  A  H  V  I  V  E  A  A  P  D  P
DNA: CCGAGCCGCGCACCGGGGAACCCGTCTGGGACCGGCCGCTGCCGCTGGTGC
 +3:   E  P  R  T  G  E  P  V  W  D  R  P  L  P  L  V  L
DNA: TCTCCGCCCGAGACGAACCGGCCCTGGCCGCCCAGGCACGCCGCATCCTCG
 +3:   S  A  R  D  E  P  A  L  A  A  Q  A  R  R  I  L  D
DNA: ACCACCTGGAGACCGGCGCCGACCTCGTCCCCGACATCGCCTACGCCCTGG
 +3:   H  L  E  T  G  A  D  L  V  P  D  I  A  Y  A  L  A
DNA: CCACCACCCGCGCCGCCCTGGACCGGCGGGCCGTCGTCATCGGCGCCGACC
 +3:   T  T  R  A  A  L  D  R  R  A  V  V  I  G  A  D  P
DNA: CGGCCACGATCACCGCGCGGCTCGCCGCCCTGGCCGAGGACGATCCGGCGT
 +3:   A  T  I  T  A  R  L  A  A  L  A  E  D  D  P  A  S
DNA: CCGACGTGGTGCGCGGCGCACCGGCGGGGGAGTCCCGCATCGCGTTCGTCT
 +3:   D  V  V  R  G  A  P  A  G  E  S  R  I  A  F  V  F
DNA: TCCCCGGGCAGGGCTCCCAGTGGGCCGGCATGGCCGCCGAACTGCTCGACG
 +3:   P  G  Q  G  S  Q  W  A  G  M  A  A  E  L  L  D  G
DNA: GCTCACCGGTGTTCGCGGCGGCCATGGCCGACTGCGCCGAGGCGCTCGCCC
 +3:   S  P  V  F  A  A  A  M  A  D  C  A  E  A  L  A  P
DNA: CCTTCACCGACTGGGACCTCGTCGACACCGTCCGGGAGCGCCGCCCCATGG
 +3:   F  T  D  W  D  L  V  D  T  V  R  E  R  R  P  M  E
DNA: AGCGGGTGGACGTGGTCCAGCCCGCGCTGTGGGCGATCATGGTCTCGCTGG
 +3:   R  V  D  V  V  Q  P  A  L  W  A  I  M  V  S  L  A
DNA: CCGAGGTGTGGCGCGCGCACGGGGTGCGGCCCGCCGCCGTCATTGGGCACT
 +3:   E  V  W  R  A  H  G  V  R  P  A  A  V  I  G  H  S
DNA: CCCAGGGCGAGATCGCCGCCGCGTGCGTGGCGGGCGCGCTGAGCCTGTCCG
 +3:   Q  G  E  I  A  A  A  C  V  A  G  A  L  S  L  S  D
DNA: ACGGGGCCCGCGTGGTGGCCCTGCGCAGCCGGGCCATCGCGGAAGTGCTCT
 +3:   G  A  R  V  V  A  L  R  S  R  A  I  A  E  V  L  S
DNA: CCGGACCCGCCGATTCCGGGACCGTTCCCGGGAAAGGTGCCTCCGGGCCCA
 +3:   G  P  A  D  S  G  T  V  P  G  K  G  A  S  G  P  T
DNA: CCAATTCGGCGCGTGGCGCCTGTGGCCGCGGCGGGATGATGTCGGTGGCGC
 +3:   N  S  A  R  G  A  C  G  R  G  G  M  M  S  V  A  L
DNA: TGCCCGAGTCCCGGGCGCGCGAACTCGTCGCCGCCCACGACGGGCGGGTCG
 +3:   P  E  S  R  A  R  E  L  V  A  A  H  D  G  R  V  A
DNA: CCGTGGCCGCGGTCAACGGCGCCTCGTCGGTGGTGCTCTCCGGGGACGCCG
 +3:   V  A  A  V  N  G  A  S  S  V  V  L  S  G  D  A  E
DNA: AGGTGCTCGACGCGCTGCGCGAGAGGATCGTCGCGGACGGCGGCCGGGCCA
 +3:   V  L  D  A  L  R  E  R  I  V  A  D  G  G  R  A  K
DNA: AGCGGCTGCCGGTGGACTACGCCTCGCACTGCGCCCATGTCGAGTCGATCC
 +3:   R  L  P  V  D  Y  A  S  H  C  A  H  V  E  S  I  R
DNA: GCGAACGGCTGCTCACCGACCTCGCGGGCGTACGGGCCCGGGGGGCCGACG
 +3:   E  R  L  L  T  D  L  A  G  V  R  A  R  G  A  D  V
DNA: TACCGTTCTACTCCACCGTCACCGGTGCAGTGCTGGACACCACCGCGCTGA
 +3:   P  F  Y  S  T  V  T  G  A  V  L  D  T  T  A  L  T
DNA: CCGCCGACTACTGGTACACGAACCTGCGCCGGAGCGTGTTGTTCGAGCCGA
 +3:   A  D  Y  W  Y  T  N  L  R  R  S  V  L  F  E  P  T
DNA: CCACCCGGGCCCTGCTCGATTCCGGATACGGGATCTTCGTCGAGTGCAGCC
```

+3: T R A L L D S G Y G I F V E C S P
DNA: CGCACCCGGTGCTGCTGAACAGCATCGAGGAGACCGCCGACGCCGTGGGCG
+3: H P V L L N S I E E T A D A V G A
DNA: CGACCGTCACCGGGCTGGGCTCGCTGCGCCGCGACGACGGCGGGGCCGAGC
+3: T V T G L G S L R R D D G G A E R
DNA: GCCTGCTCACCTCGCTCGGCGAGGCGTTCGTGGCGGGTGTCCCGGTCGACT
+3: L L T S L G E A F V A G V P V D W
DNA: GGTCGGCGGTGTTCACGG*GCATGC*CGGTGCGCGCCGCCGATCTGCCCACGT
+3: S A V F T G M P V R A A D L P T Y
DNA: ATCCCTTCCAGCGCGAGCGCTACTGGCTGGGCCGGTCCGCGGCCTCCGGCG
+3: P F Q R E R Y W L G R S A A S G D
DNA: ACGTCACCGCCGCCGGGCTGCGGGCCACCACCCATCCGCTGCTGGGCGCGG
+3: V T A A G L R A T T H P L L G A A
DNA: CAGTCCAGGTCGCCGGGGGCGGCACCCTGTTCACCGGCCGGCTCTCCGTGT
+3: V Q V A G G G T L F T G R L S V S
DNA: CCACCACGCCCTGGCTGGCCGACCACGCGGTCTCGGGCACCCCCCTGCTGC
+3: T T P W L A D H A V S G T P L L P
DNA: CCGGCACCGCGCTGGTGGAGCTGGCGCTGAGCGCGGGCCACGAACTCGGGT
+3: G T A L V E L A L S A G H E L G Y
DNA: ACGGGCACGTCGCCGAACTCACCCTCCAGGCGCCGCTGGTGCTGCCCGGCC
+3: G H V A E L T L Q A P L V L P G R
DNA: GGGCGGCGGTCCAGTTCCAGGTACACGTGGCCGCCGCCGACGAGGACGGCC
+3: A A V Q F Q V H V A A A D E D G H
DNA: ACCGCGCGCTGACCGTCCACTCCCGCCCCGAGGGCGCCGACGACACCGAGT
+3: R A L T V H S R P E G A D D T E W
DNA: GGACCGCGCACGCCACCGGGCTGCTCGCCCCGCGGACCGCCCCGCCCGGCT
+3: T A H A T G L L A P R T A P P G F
DNA: TCGACCTCGCGCAGTGGCCGCCCCGGGGCGCGGAACCGGTGCTGGTGGACG
+3: D L A Q W P P R G A E P V L V D D
DNA: ACGCCTACGACACGCTGGCCGCGCTCGGCTACGACTACGGGCCCGCCTTCC
+3: A Y D T L A A L G Y D Y G P A F Q
DNA: AGGGCCTGCGCGCGGTCTGGCGGCGTGGCGACGAGACCTTCGCCGAGGTCG
+3: G L R A V W R R G D E T F A E V E
DNA: AACTCCCCGGTGAGGCAGGTGCGTTCGGCCTGCACCCGGCCCTGTTCGACG
+3: L P G E A G A F G L H P A L F D A
DNA: CGGCCCTGCACGCCGACGGCCTGCGCACGGCCCCGCCCGGCACCGACGGCC
+3: A L H A D G L R T A P P G T D G P
DNA: CCGGGGCGCGGGGGCAGGGGGCGGCGCGGCTGCCCTTCGTCTGGACCGGCG
+3: G A R G Q G A A R L P F V W T G V
DNA: TGTCGTTGTATGCGTCCGGGGCCACCGCCCTGCGGGTCCGCATCCGGGGCG
+3: S L Y A S G A T A L R V R I R G G
DNA: GCGACACGCTCTCCCTGGACCTGGCCGACCCGACCGGCGCACCGGTCGCCG
+3: D T L S L D L A D P T G A P V A A
DNA: CCGTGGAGGCCCTGGTCTCCAGGCCGGTCGACCCGGCGGCGCTGACCTCCC
+3: V E A L V S R P V D P A A L T S P
DNA: CGGTCCGGGACGACGACCTGTACCGGCTGGACTGGCAGGCGCTGCCCGTAC
+3: V R D D D L Y R L D W Q A L P V P
DNA: CCGTGGCGGACGCGCCCGCGTACGCCGTTCTGGACGAGCGGGGCACGGCCG
+3: V A D A P A Y A V L D E R G T A A
DNA: CGGCGGACGCCGTGCCGGACTGGGTGGTCCTGCCGGTGAGCGGTGACGGCG
+3: A D A V P D W V V L P V S G D G G
DNA: GCGACCCGGTGGGCGGGGTGCGCGCGGCGACCGGGCGGGTCCTCGCCGCCG
+3: D P V G G V R A A T G R V L A A V
DNA: TGCGCGACTGGCTGGCGGACGAGCGTACGGCCGGGGCCCGGGGGGCCCGGC
+3: R D W L A D E R T A G A R G A R L
DNA: TGGTGGTCCTGACCGGCGGCGCGGTCGCCACCGGCACGGAGGACGTCACCG
+3: V V L T G G A V A T G T E D V T D
DNA: ACCTGGCGGGTGCCGCCGTATGGGGCCTGGTCCGGGCGGCCCAGGGCGAAC
+3: L A G A A V W G L V R A A Q G E H
DNA: ACCCCGACCGCTTCGTCCTGGTGGACTCCGTCGCCCACGACGGCGGCGGCG
+3: P D R F V L V D S V A H D G G G E

```
DNA: AAAGTGCCTCCGGCCCGGGTGTCTTTGCCACCGACCGGGTCACCGAGGCCG
 +3:   S  A  S  G  P  G  V  F  A  T  D  R  V  T  E  A  V
DNA: TGCGCGCCGCCGCGGCGAGCGGCGAACCGCAGCTGGCCCTGCGCGAGGGCA
 +3:   R  A  A  A  A  S  G  E  P  Q  L  A  L  R  E  G  T
DNA: CCGTACGGGTACCCCGGCTGGCCCGTGCCGCCGTAACGGGAACGGCCGCCG
 +3:   V  R  V  P  R  L  A  R  A  A  V  T  G  T  A  A  V
DNA: TACCCGCTTTTGACGGCCCCGCGCCGGATCCTCACGGCACCGTGCTCATCA
 +3:   P  A  F  D  G  P  A  P  D  P  H  G  T  V  L  I  T
DNA: CCGGCGGCACGGGAGTGCTCGGTGCCGTGGTCGCCCGGCACCTGGCCACCG
 +3:   G  G  T  G  V  L  G  A  V  V  A  R  H  L  A  T  E
DNA: AGCACGGGGTGCGCCGTCTCGTCCTGGCCGGCCGCAGCGGCACCGCCTTCG
 +3:   H  G  V  R  R  L  V  L  A  G  R  S  G  T  A  F  D
DNA: ACGACTTCGGCGATCTCGCCGAACGCGGCACCGAGGTCGTCGTCGCCCGCT
 +3:   D  F  G  D  L  A  E  R  G  T  E  V  V  V  A  R  C
DNA: GCGACGCCGCCGAACGCGACCAACTGGCCGCGCTGCTGGCCGACATGCCCG
 +3:   D  A  A  E  R  D  Q  L  A  A  L  L  A  D  M  P  A
DNA: CGGAGCGCCCGCTGACCGCGGTGATCCACCTCGCCGGGGTCCTGGACGACG
 +3:   E  R  P  L  T  A  V  I  H  L  A  G  V  L  D  D  G
DNA: GACTGGTGACCGATCAGACACCCGGGCGACTGGACGCCGTCCTGCGGCCCA
 +3:   L  V  T  D  Q  T  P  G  R  L  D  A  V  L  R  P  K
DNA: AGGCGGACGCCGCCTGGAACCTGCACGAGCTGACCCGTGACCTGGACCTGT
 +3:   A  D  A  A  W  N  L  H  E  L  T  R  D  L  D  L  S
DNA: CGGCGTTCGTCCTCTTCTCCTCGGCCGCGGGCACGATCGACGGCGCGGGCC
 +3:   A  F  V  L  F  S  S  A  A  G  T  I  D  G  A  G  Q
DNA: AGTCCGGGTACGCCGCCGCCAACGCCTTCCTCGACGGCCTGGCCGCCCACC
 +3:   S  G  Y  A  A  A  N  A  F  L  D  G  L  A  A  H  R
DNA: GCGCCGCCCAGGGCCTGCCCGCGCTCTCCCTCGCCTGGGGCTTCTGGGAGC
 +3:   A  A  Q  G  L  P  A  L  S  L  A  W  G  F  W  E  Q
DNA: AGCGCACCGGGATGACCGCCCACCTCACCGACGCCGACGTGGAGCGCATGG
 +3:   R  T  G  M  T  A  H  L  T  D  A  D  V  E  R  M  A
DNA: CACGTGCCGGGGTCCGGCCCCTGCCCACCGAGGAGGGGCTGAGGCTGCTGG
 +3:   R  A  G  V  R  P  L  P  T  E  E  G  L  R  L  L  D
DNA: ACGCCGCGCTCGCCGCCGACGTACCGCTGCTGCTGCCCGTCGGCCTGGACC
 +3:   A  A  L  A  A  D  V  P  L  L  L  P  V  G  L  D  P
DNA: CGCGCGCCCTGCGCGGTGCCGACGACGTCCCGCCCGTGCTTGCGCGCTCTG
 +3:   R  A  L  R  G  A  D  D  V  P  P  V  L  A  R  S  G
DNA: GCGCCCGCGCCCGTCCGTCGTACGGCGGCCTCCCGCGCCACCGCCGTTCCG
 +3:   A  R  A  R  P  S  Y  G  G  L  P  R  H  R  R  S  A
DNA: CCGCCGAACGGCTGGCCGCCCTCGGCGCCGCCGAACGCGAGGCGGCGCTCA
 +3:   A  E  R  L  A  A  L  G  A  A  E  R  E  A  A  L  T
DNA: CGGAGCTGGTCCGCACCCATGTCGCGGCCGTTCTCGGGCACGGCGCGGACA
 +3:   E  L  V  R  T  H  V  A  A  V  L  G  H  G  A  D  M
DNA: TGGTGCTCGACCCGCGCCGCTCCTTCCGCGAGGCCGGTTTCGACTCGCTGA
 +3:   V  L  D  P  R  R  S  F  R  E  A  G  F  D  S  L  T
DNA: CCGCGGTCGAGCTGCGCAACCGCCTCGGAAACGCCGTCGGCCTCCGGTTGC
 +3:   A  V  E  L  R  N  R  L  G  N  A  V  G  L  R  L  P
DNA: CCGCCACCCTCGTCTTCGACCACCCCGACGCCGAGGCCCTGGTCAGGTACC
 +3:   A  T  L  V  F  D  H  P  D  A  E  A  L  V  R  Y  L
DNA: TGAAGACGGAACTCTTCGGCGCGGACCCCGAGGACGCCGAGGCCTCCACCG
 +3:   K  T  E  L  F  G  A  D  P  E  D  A  E  A  S  T  G
DNA: GGATCGGGGCCGTCGTCCCCGGAGCGGGGTACGAACCGGACGAGCCGGTGG
 +3:   I  G  A  V  V  P  G  A  G  Y  E  P  D  E  P  V  A
DNA: CGATCGTCGGGATGGCGTGCCGCTACCCCGGCGGCGTCACCACGCCCGAGG
 +3:   I  V  G  M  A  C  R  Y  P  G  G  V  T  T  P  E  E
DNA: AGCTGTGGCGGCTCGTCGCGGACGGCGTGGACGGCATCGGCGCGTTCCCCG
 +3:   L  W  R  L  V  A  D  G  V  D  G  I  G  A  F  P  D
DNA: ACGACCGGGGCTGGAACCTCGACACCCTGTACGACCCGGAGCCCGGCAAGC
 +3:   D  R  G  W  N  L  D  T  L  Y  D  P  E  P  G  K  P
DNA: CCGGCCACTGCTCCACCCGCGCGGGCGGATTCCTCTACGACGCCGCCGACT
 +3:   G  H  C  S  T  R  A  G  G  F  L  Y  D  A  A  D  F
DNA: TCGACCACGACTTCTTCGGCATCGGCCCCCGCGAGGCCCTCGCCATGGACC
```

```
 +3:   D  H  D  F  F  G  I  G  P  R  E  A  L  A  M  D  P
DNA: CGCAGCAGCGGTTGCTGCTGGAGACCTCCTGGGAGGCGCTGGAACGGGCCG
 +3:   Q  Q  R  L  L  L  E  T  S  W  E  A  L  E  R  A  G
DNA: GCATCGATCCGCACTCCGTGCGCGGCAGCCGCACCGGCGTGTTCGCCGGGG
 +3:   I  D  P  H  S  V  R  G  S  R  T  G  V  F  A  G  V
DNA: TCATGTACCACGACTACGGCAGCAGGCTGCGCGACGTCCCCGAGGCCGTGC
 +3:   M  Y  H  D  Y  G  S  R  L  R  D  V  P  E  A  V  R
DNA: GCGACTACCTCGGCAACGGAAGCCTCGGCAGTATCGCCTCCGGCCGTATCG
 +3:   D  Y  L  G  N  G  S  L  G  S  I  A  S  G  R  I  A
DNA: CCTACACCCTGGGTCTGGAGGGCCCGGCGCTCACCGTGGACACGGCCTGCT
 +3:   Y  T  L  G  L  E  G  P  A  L  T  V  D  T  A  C  S
DNA: CCTCGTCGCTGGTGGCGCTGCACCTGGCGGCGCAGGCACTGCGGCGGGGGG
 +3:   S  S  L  V  A  L  H  L  A  A  Q  A  L  R  R  G  E
DNA: AGTGCGGCCTGGCCCTGGCCGGTGGCGTGTCCGTGATGTCGACCGTCGACA
 +3:   C  G  L  A  L  A  G  G  V  S  V  M  S  T  V  D  T
DNA: CGTTCGTGGACTTCAGCAGGCAGCGCAACCTCGCCGCCGACGGCCGCGCCA
 +3:   F  V  D  F  S  R  Q  R  N  L  A  A  D  G  R  A  K
DNA: AGTCCTTCGCCGAGGCGGCGGACGGCACGGCGCTGTCCGAGGGCGTCGGTG
 +3:   S  F  A  E  A  A  D  G  T  A  L  S  E  G  V  G  V
DNA: TGTTGGTGTTGGAGCGGTTGTCGGATGCGCGGCGGTCGGGGCGTCGGGTGT
 +3:   L  V  L  E  R  L  S  D  A  R  R  S  G  R  R  V  W
DNA: GGGGGGTGGTGCGGGGTTCGGCGGTGAATCAGGATGGTGCGTCGAATGGGT
 +3:   G  V  V  R  G  S  A  V  N  Q  D  G  A  S  N  G  L
DNA: TGACGGCGCCGAATGGTCCGGCGCAGCAGCGGGTGATTCGTGAGGCGTGGG
 +3:   T  A  P  N  G  P  A  Q  Q  R  V  I  R  E  A  W  V
DNA: TGGCTGCGGGTGTGTCGGGTGGTGGGGTGGATGTGGTGGAGGCGCATGGGA
 +3:   A  A  G  V  S  G  G  G  V  D  V  V  E  A  H  G  T
DNA: CGGGGACGGTGTTGGGTGATCCGATCGAGGCGCAGGCGTTGTTGTCTACGT
 +3:   G  T  V  L  G  D  P  I  E  A  Q  A  L  L  S  T  Y
DNA: ACGGGCAGGGGCGTGGGGGTGGGGATCC
 +3:   G  Q  G  R  G  G  G  D
```

Figure 27: Nucleotide sequence of the 11.6 kb PKS I genes isolated from the *Streptomyces* sp. 98– 62 and the deduced amino acid sequence. The different modules are represented in different colours. Module 1 of ORF1 in blue, modules 2 and 3 of ORF 2 in black and maroon respectively. The various deduced domains of each module are indicated in bold. The deduced stop codons are in red. The deduced start codon is in green. The deduced promoter like sequence and ribosomal binding site are in pink.

4.13 SETTING UP OF A GENE DISRUPTION EXPERIMENT

4.13.1 GENE DISRUPTION: CHOICE OF VECTOR AND DONOR E.COLI STRAIN

In order to prove the functions of the cloned PKS I genes of the *Streptomyces* sp. 98- 62 in the biosynthesis of the novel antifungal compound produced by the *Streptomyces* sp. 98- 62, the chosen strategy was to specifically inactivate the PKS I genes within the identified cluster and observe if the production of the antifungal compound by the *Streptomyces* sp. 98- 62 was negated. To perform gene inactivation in the *Streptomyces* sp. 98- 62, a gene transfer system for the *Streptomyces* sp. 98- 62 needed to be established first.

Since PEG -mediated protoplast transformation was generally not very efficient in addition to being time and labour intensive, intergeneric method of plasmid DNA transfer from *E.coli* to *Streptomyces* was attempted in order to transfer DNA into the *Streptomyces* sp. 98- 62. Since *E.coli/ Streptomyces* intergeneric conjugation was first reported by Mazodier *et al* (1989), this method has been successfully used with a number of streptomycete strains. As the *Streptomyces* sp. 98- 62 was identified to belong to *Streptomyces* sp., intergeneric conjugation was expected to be a feasible method of gene transfer into the *Streptomyces* sp. 98- 62.

The plasmid pSOK201 was the vector of choice to be used in intergeneric conjugation. The vector contains the *oriT* sequence from the Inc-P group plasmid RK2, which allows for replication in *E. coli*. However this vector is a nonreplicative vector in *Streptomyces*, and needs to be integrated into the streptomycete chromosome by homologous recombination between a cloned DNA fragment and the homologous sequence in the genome, to yield stable recombinant strains. The vector pSOK201 does not carry the genes for transfer functions and therefore requires the transfer functions to be supplied in *trans* by the *E. coli* donor strain (Zotchev, 2000).

The *E.coli* donor strain used is the methylation deficient strain ET12567 (MacNeil, 1992), which carries a plasmid pUB307. The plasmid pUB307 is a derivative of RP1 (Richmond, 1976), which encodes the transfer function, *tra*. The use of *E. coli* (pUB307) or equivalent strains may be more widely applicable since several *Streptomyces* species have been shown to possess a methyl specific restriction system (Macneil, 1988). Intergeneric conjugation has also been used in streptomycetes which do not possess methyl DNA restriction systems (Wohllben, 1993, Mazodier, 1993). Thus in the case of the *Streptomyces* sp. 98- 62 which has been identified to be a streptomycete, gene transfer by intergeneric conjugation seemed to the method of choice of gene transfer.

4.13.2 DISRUPTION CONSTRUCTS

The plasmid pSOK201 derivatives containing DNA fragments of the *Streptomyces* sp. 98- 62 DNA ranging in size from 1. 5 kb to 7. 0 kb were tested for their ability to integrate into the chromosome of the *Streptomyces* sp. 98- 62. Homologous recombination between the cloned DNA and the corresponding homologous chromosomal region would lead to the integration of the plasmid. Four of the gene disruption plasmid constructs contained different DNA fragments from the clones E27, C170 and C2.

The PKS I gene fragments of the *Streptomyces* sp. used in the gene disruption experiment are shown in Fig.29. Construction of the different disruption vectors and restriction map of the disruption constructs are also given in Fig. 30-33.

These PKS I gene fragments of the *Streptomyces* sp. 98- 62 were cloned into the *EcoRI/HindIII* site of pSOK201. As C170 had no restriction site for *EcoRI or HindIII*, the restriction sites of pUC18 were used to extract out the 7.2kb insert

fragment of C170 from pUC18 and cloned into pSOK201. It has to be noted that the 7.2k b insert fragment of this disruption construct, labelled pDC170FL, contained 2 stop codons and a start codon in the intermodular region. The insert fragment encoded the enzymes DH, KR, ACP, KS, AT and DH (Fig. 30).

From the subclone of E27, p2.3KBE27, which carried the 2.3 kb BamHI/Sp*hI* fragment of E27, the disruption construct pDE27 was constructed. The insert fragment had no unique restriction site for *EcoRI* and *HindIII.* Therefore *EcoRI* and *HindIII* restriction site of pUC18 were used to release the 2.3 kb insert fragment from pUC18 and cloned into pSOK201. The insert fragment encoded the enzymes KS, AT and DH (Fig. 31).

From the subclone of C170, which carried the 2.0 kb Sp*hI* fragment of C170, the disruption construct pD2KBC170 was constructed. The insert fragment contained no restriction site for *EcoRI or HindIII.* Therefore, restriction sites of pUC18 were used to release the 2.0 kb insert fragment from pUC18 and cloned into pSOK201. The insert fragment encoded the enzymes KS and AT (Fig. 32).

From the clone C2, which carried the 3.7 kb Sp*hI* fragment, the disruption construct pDC2 was constructed. As the insert fragment contained no restriction site for *EcoRI or HindIII*, restriction sites of pUC18 were used to release the 3.8 kb insert fragment of C2 from pUC18 and cloned into pSOK201. The insert fragment encoded the enzymes DH, KR, ACP and KS. As a consequence of this, the insert fragment of this construct spanned the inter modular region of two modules. However, there is no predicted stop/ start codon in the intermodular region (Fig. 33).

The integration of the pSOK201 disruption vector constructs would bring about the disruption of a gene or operon only if the cloned fragment lacked the start codon and/or stop codon of that gene or operon. Homologous recombination between the insert fragment of the disruption construct and the homologous region in the intact chromosome would result in the integration of the whole disruption vector backbone into the chromosome such that there is duplication of the homologous gene fragment on either side of the inserted vector backbone.

The duplicated genes would be non-functional only if the reading frame of the gene is disrupted such that a functional protein product cannot be produced. Such a situation would only arise if the homologous fragment in the gene disruption construct lacked the start codon and/or stop codon of that gene or operon. This is because insertion of such a disruption construct into the chromosome would result in one copy of the duplicated gene being truncated at the 3' end and therefore would lack the stop codon as well as a region coding the 3' end amino acid. This copy would also most likely produce a non-functional gene product, as the gene product would be truncated. However in some cases, the truncated copy might still possess some residual activity, especially in large multifunctional proteins. The second copy of the duplicated gene would be truncated at the 5' end lacking the ribosomal binding site, start codon and a region coding the 5' end amino acids. Therefore this copy would be unlikely to produce a functional gene product (Fig. 34).

GENE DISRUPTION USING DISRUPTION CONSTRUCT WITHOUT A STOP/ START CODON

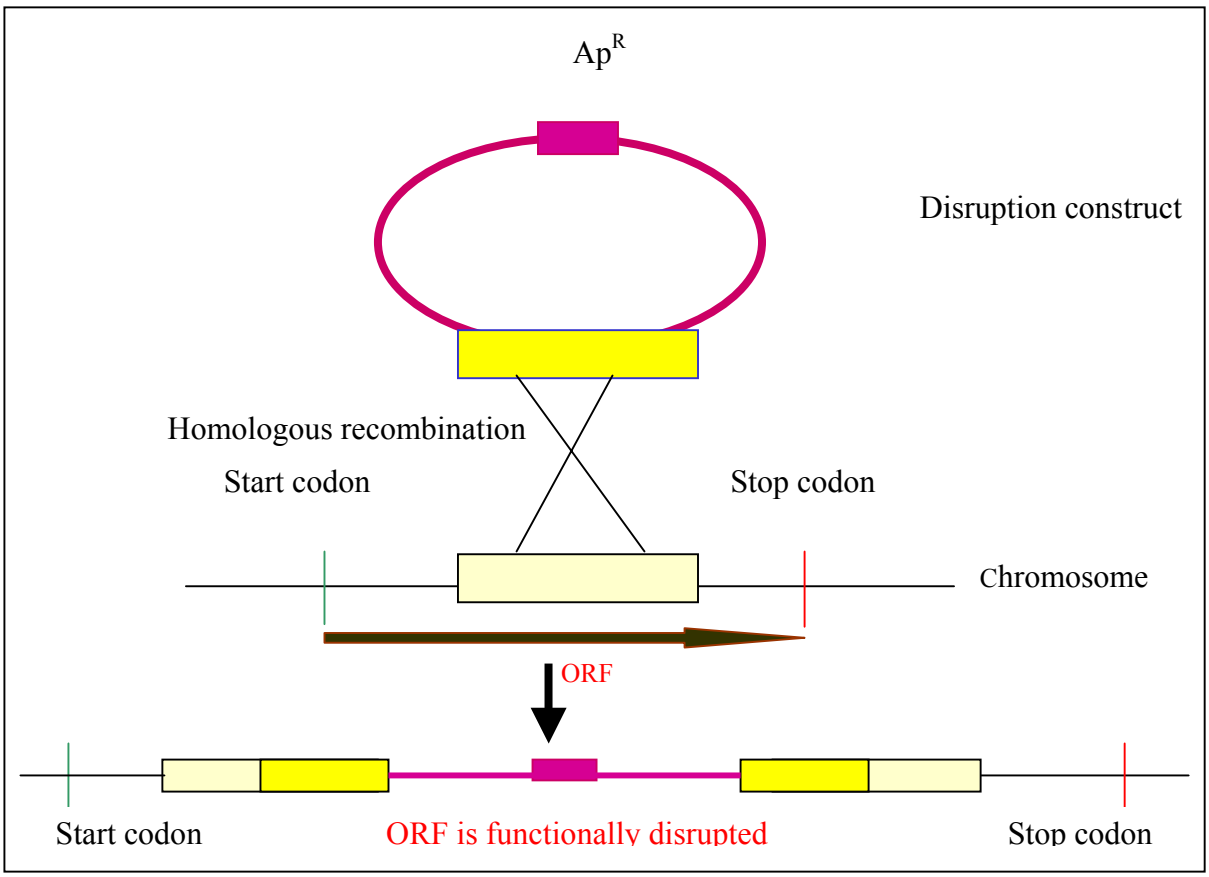


Figure 34: Gene disruption using a gene fragment without a stop/start codon.

If the homologous fragment in the gene disruption construct contained the start and/or stop codon of that gene or operon, insertion of the disruption construct into the intact chromosome would result in duplication of the genes without any functional change in the open reading frame although there is a physical separation of the open reading frames due to the insertion of the vector backbone in between the duplicated copies of the homologous fragment (Fig. 35).

GENE DISRUPTION USING DISRUPTION CONSTRUCT WITH A START/ STOP CODON

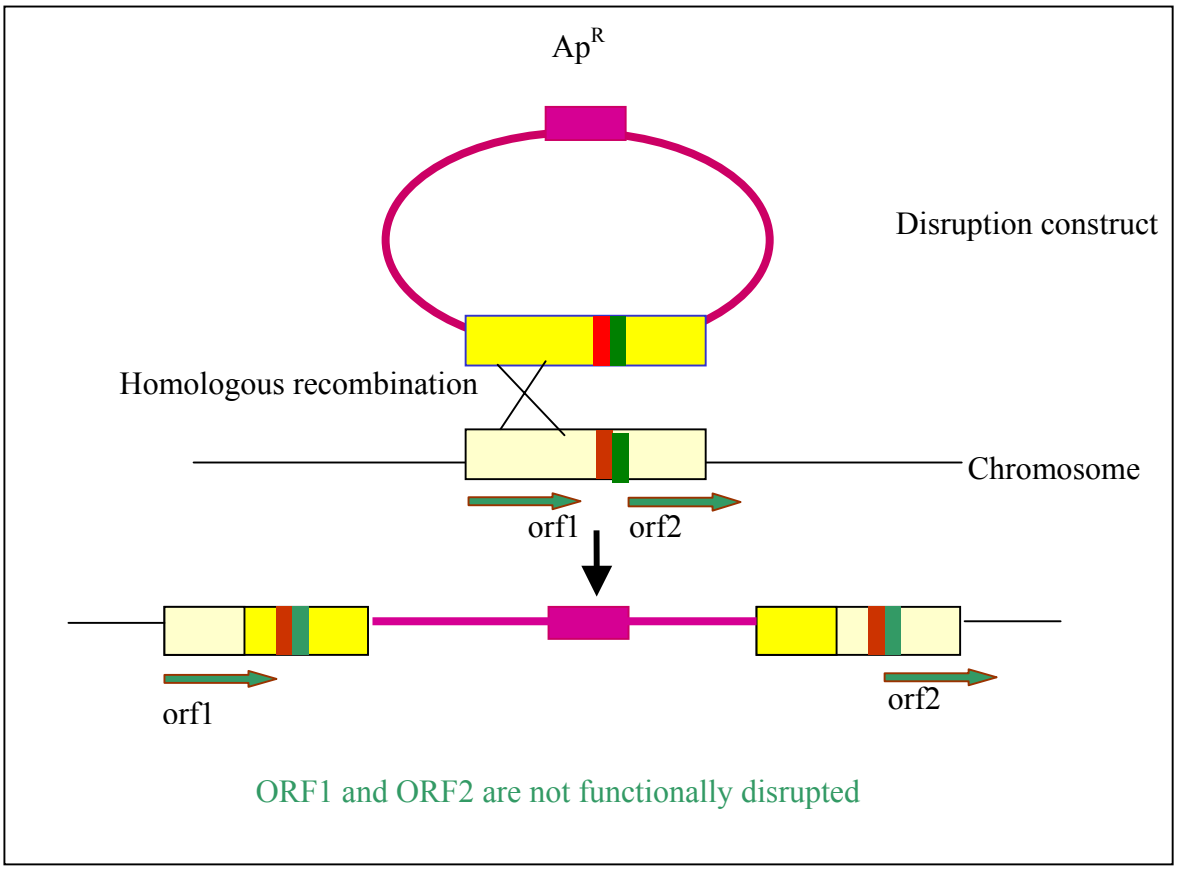


Figure 35: Gene disruption using a gene fragment with a start/ stop codon.

One of the gene disruption construct pDC170 contained 2 stop codons and a start codon. This construct was utilised to prove that the predicted stop/start codons were indeed functional and that the disrupting sequence of 7.2 kb constituted two different open reading frames as deduced.

Conjugation experiment with the different disruption constructs were set up as described in Materials and Methods.

4.14 GENE DISRUPTION USING pDC170, A DISRUPTION CONSTRUCT WITH STOP/START CODONS

Initial optimisation experiments were done using the C170 gene disruption construct. Optimisation experiments included incubating agar plates of conjugation mixes at two different temperatures for 5 days to determine the effect of temperature on conjugation. The conjugation mixes were incubated at either 30°C or 37°C for 5 days. Higher temperatures have been shown to increase conjugation frequency in *Streptomyces fradiae* (Schoner, 1992).

Intergeneric conjugation experiments with integrative pSOK201 gene disruption constructs from donor strain *E.coli* (pUB307) to the *Streptomyces* sp. 98- 62 were successful. No exconjugants were obtained in control experiment without the addition of *E. coli* cells.

Matings at 37°C gave a high number of small apramycin resistant colonies. However, the exconjugants did not grow well after 3 to 4 days. Matings at 30°C gave fewer number of apramycin resistant colonies compared to 37°C. The exconjugants from 30°C grew well even up to 12 days. Hence matings at 30°C appeared to be optimal for our purpose of study (Fig. 36, 37).

The exconjugants were streaked out onto a fresh AS- 1 plate and allowed to grow at 30°C for 5 days. The phenotype of the exconjugants was noted to be the same as the wild type *Streptomyces* sp. 98- 62. This was then used to inoculate ISP2 liquid broth, which served as the preculture for the secondary metabolite extraction as well as the genomic DNA extraction procedure. Two exconjugants, named 170D1 and 170D2 from pDC170 disruption experiment were analysed by Southern blot and TLC-bioassay to determine the physical and functional disruption.

4.14.1 SOUTHERN BLOT HYBRIDISATION TO PROVE PHYSICAL DISRUPTION OF THE GENES ENCODING THE PRODUCTION OF ANTIFUNGAL COMPOUND

The genomic DNA of disruptants 170D1 and 170D2 were restricted with *BamHI* and Sp*hI*, and Southern blotted and probed with the 3 kb vector sequence pSOK201 (Fig. 38a, b). The restriction profile of the disrupted chromosome is shown in the Fig. 30. A 3 kb *BamHI and* 6.8 kb Sp*hhI* of 170D1 and 170D2 genomic DNA hybridised to the 3 kb vector sequence pSOK201 as predicted of successful physical gene disruption. The wild type *Streptomyces* sp. 98- 62 genomic DNA did not hybridise to the vector probe at all. This Southern hybridisation result confirmed that 170D1 and 170D2 had undergone insertion of the gene disruption construct.

The same blot was stripped and probed with the 7.2 kb C170 PKS I fragment as well to confirm that the insertion of the gene disruption construct had occurred in the expected region in the genome (Fig. 38c). The 6.8 kb Sp*hI* fragment hybridised strongly to the 7.2 kb C170 PKS I probe. This 6.8 kb Sp*hI* fragment was expected to hybridise to both the vector sequence and 7.2 kb PKS I genes and it indeed hybridised to both the probes. Multiple hybridising bands were observed for the DNA of the wild type as well as the disruptant with the 7.2 kb C170 PKS I probe. The detection of

multiple hybridising bands are due to the repeated nature of the PKS I genes in the *Streptomyces* sp. 98- 62.These results clearly showed that the disruption construct pDC170 had inserted into the homologous 7.2 kb *BamHI* region in the genomic DNA of the *Streptomyces* sp. 98- 62.

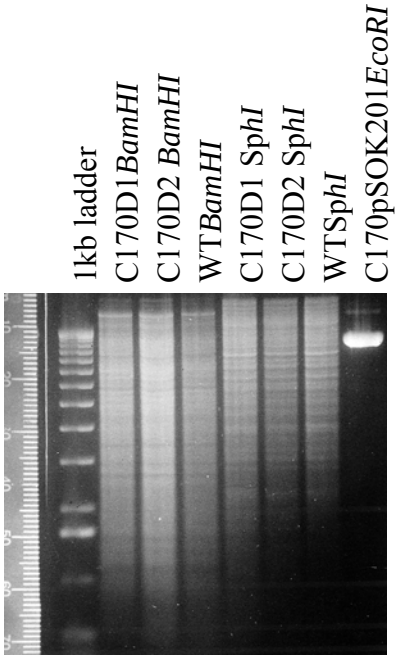


Figure 38a

Figure 38a: Electrophoretic profile of restriction endonuclease digested chromosomal DNA samples of disruptants C170D1, C170D2 and wild type *Streptomyces* sp. 98- 62

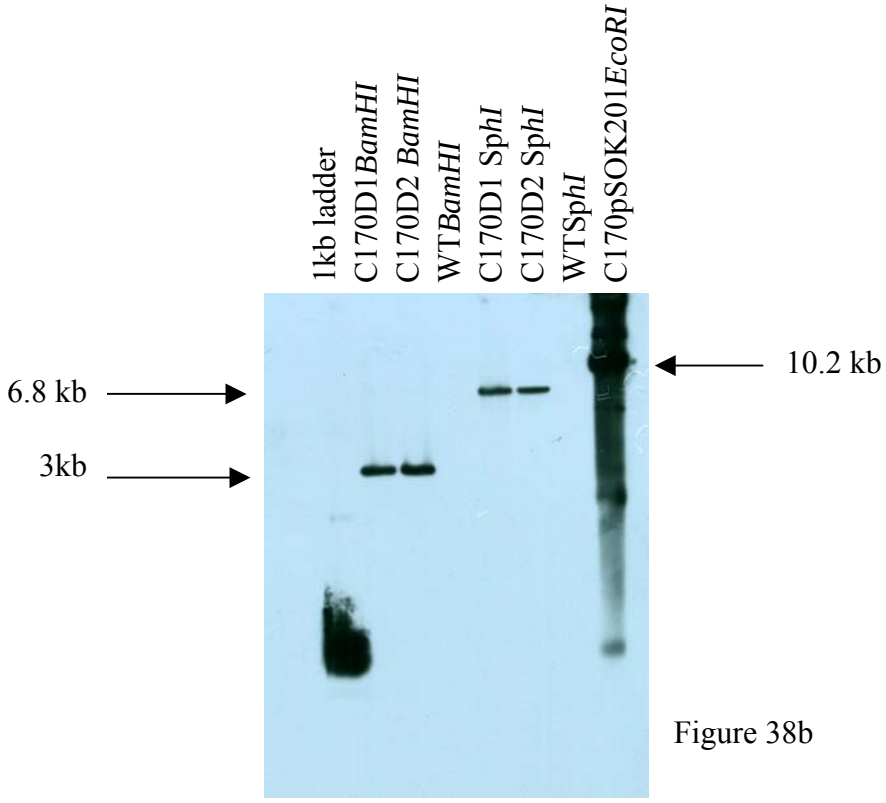


Figure 38b

Figure 38b: Southern blot of restriction endonuclease digested chromosomal DNA samples of disruptants C170D1, C170D2 and wild type *Streptomyces* sp. 98- 62, denoted as WT, probed with the vector backbone of disruption construct C170 pSOK201

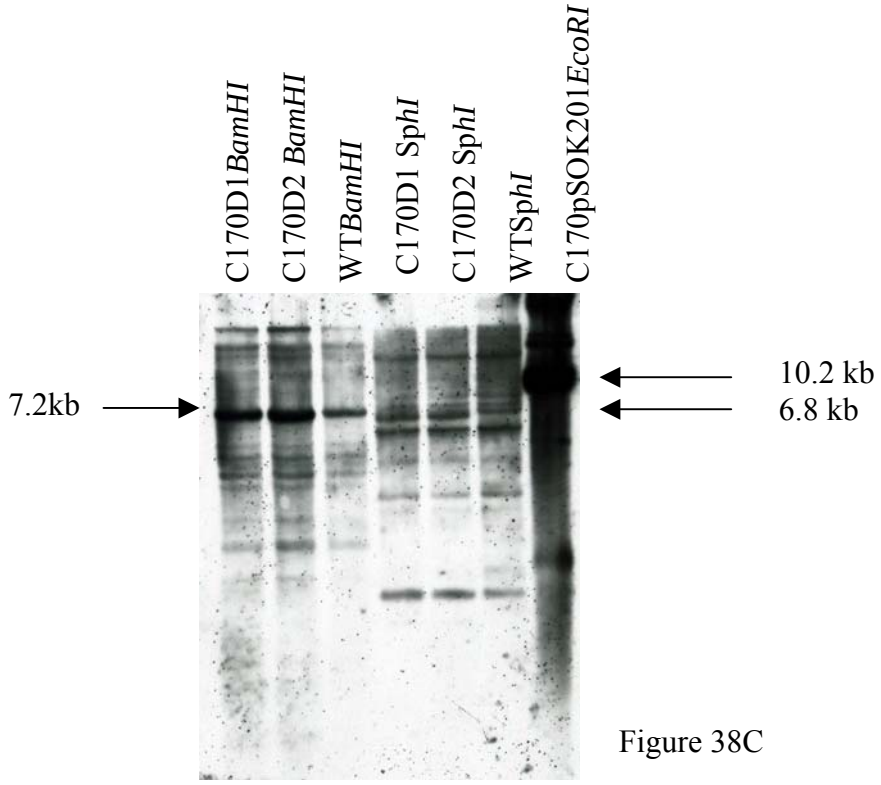


Figure 38C

Figure 38c: Southern blot of restriction endonuclease digested chromosomal DNA samples of disruptants C170D1, C170D2 and wild type *Streptomyces* sp. 98- 62, denoted as WT, probed with the 7.2 kb insert fragment of the disruption construct C170 pSOK201

4.14.2 COMPOUND EXTRACTION AND OVERLAY ASSASY TO PROVE FUNCTIONAL DISRUPTION OF THE GENES ENCODING THE PRODUCTION OF ANTIFUNGAL COMPOUND

Secondary metabolites of disruptants grown in FK medium were extracted and analysed by TLC and then bioassayed against *Aspergillus niger,* to determine if the antifungal compound biosynthesis by the *Streptomyces* sp. 98- 62 was affected by the gene disruption of the PKS I genes (Fig. 39).

The extracts of the disruptants 170D1 and 170D2 produced a spot on the TLC plate at the Rf value of 0.69, as was the case for the wild type. Bioassay by overlay of cells of *A. niger* on the TLC plate revealed that the secondary metabolite at the Rf

value of 0.69 retained its bioactivity. This result indicates that the disruptants 170D1 and 170D2 are not functionally disrupted.

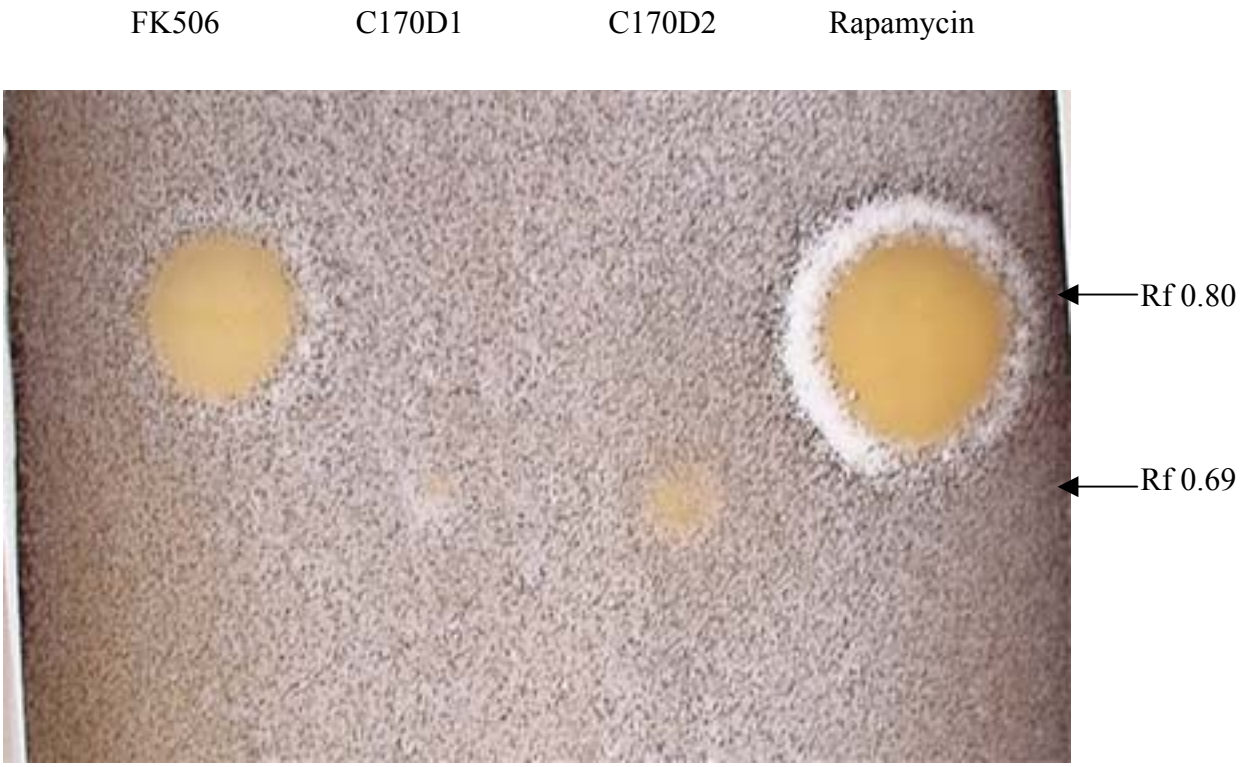


Figure 39: TLC Chromatogram and overlay assay of the extracts of pure FK506, disruptants C170D1 , C170D2 and rapamycin. The cleared area represents the zone of inhibition. The test organism used was *Aspergillus niger*.

This could be because the insert fragment carries 2 stop codons and a start codon between the modules. Fig. 35 explains why in such a case, physical disruption will not result in a functional disruption.

The observation of nonfunctional disruption does not however rule out the possibility for sure that the 7.2 kb PKS I gene used in gene disruption may not encode the enzymes involved in the biosynthesis of the antifungal compound. In order to clarify that the lack of functional disruption in disruptants 170D1 and 170D2 is due to the presence of stop/ start codon and to determine if the PKS genes of the 7.2 kb PKS I

fragment cloned from *Streptomyces* sp. is indeed involved in the biosynthesis of the antifungal compound, internal fragments of clone C170 as well as E27 and C2 were decided to be used for further gene disruption analysis.

4.15 GENE DISRUPTION USING INTERNAL FRAGMENTS OF PKS I GENES.

4.15.1 PHENOTYPE OF DISRUPTANTS

One representative disruptant each of the three different PKS I internal fragments, named 27D1, 2KBC170D1 and C2D1, were analysed by Southern blot and TLC and bioassay tests to determine the physical and functional disruption of the genes encoding the production of the antifungal compound.

In these experiments, the phenotype of the disruptants obtained upon disruption of the internal fragments of the PKS I gene was surprisingly very different from that of the wild type strain (Fig. 40a, b and c). The disruptants did not sporulate, were bald and were creamish white in colour. Single colonies were also much smaller when compared to the wild type colonies when grown for a similar period of time. This is the first report that the disruption of the PKS I cluster of genes in a strain results in the change in the phenotype of the producing streptomycete.

27I

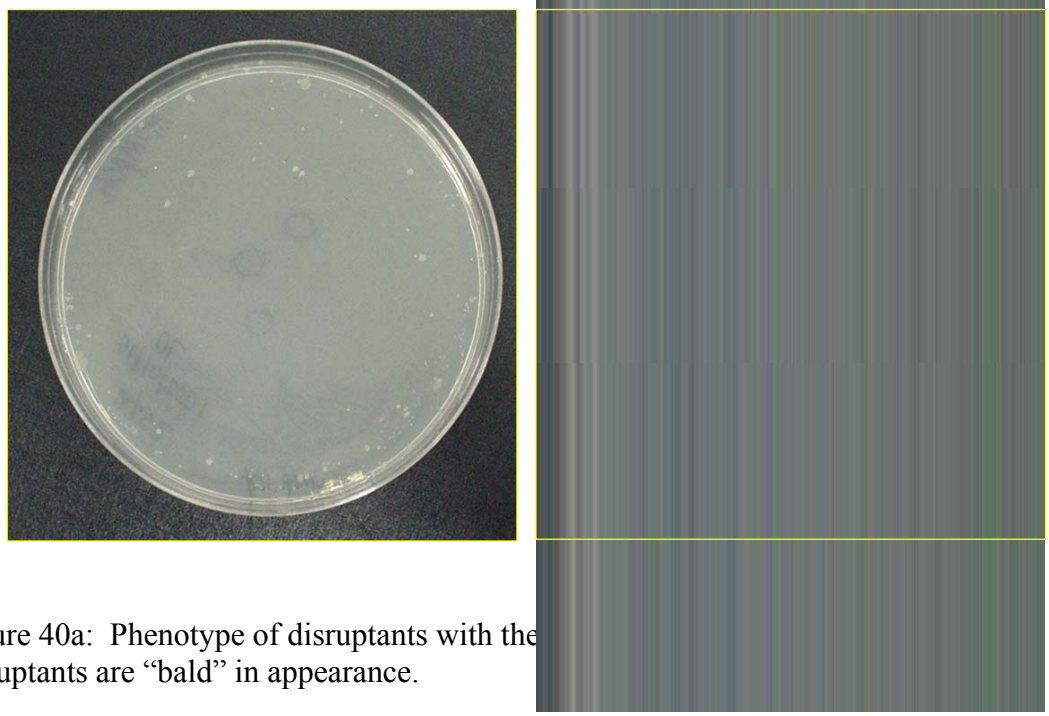


Figure 40a: Phenotype of disruptants with the
disruptants are “bald” in appearance.

C2D1

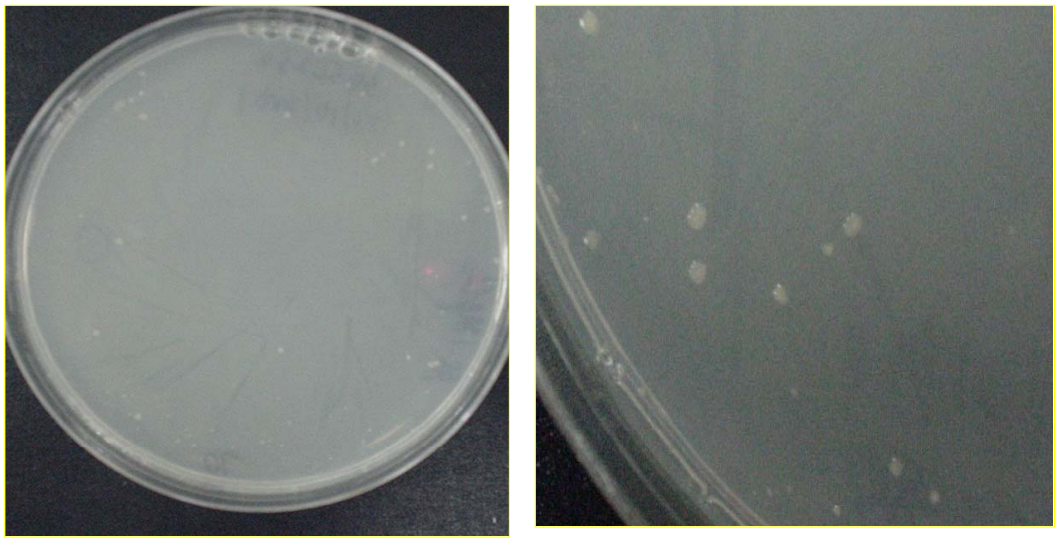


Figure 40b: Phenotype of disruptants with the disruption construct pDC2. The disruptants are “bald” in appearance.

2KBC170D1

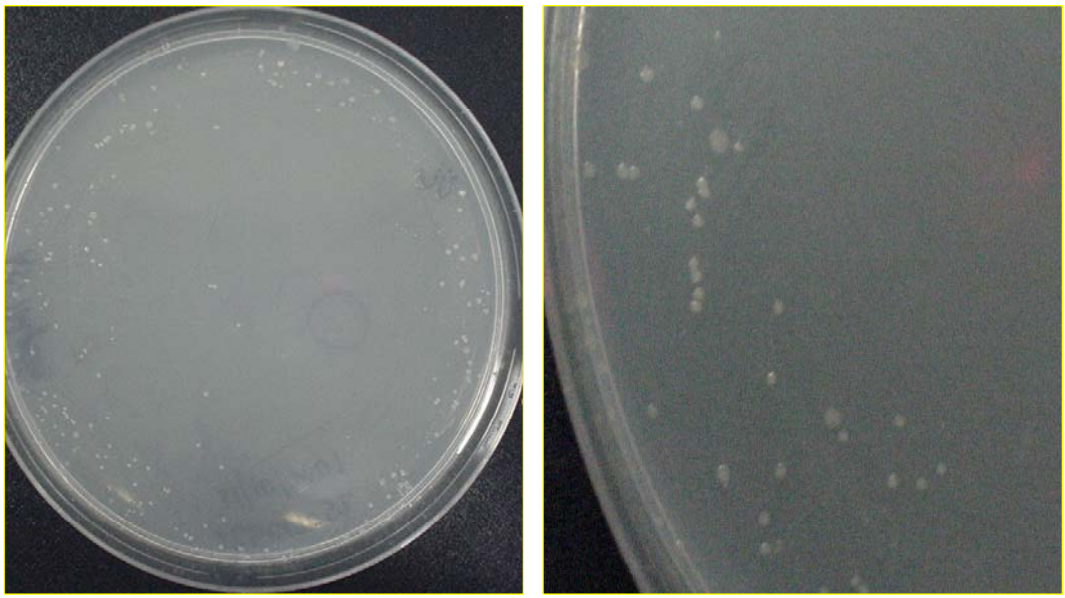


Figure 40C: Phenotype of disruptants with the disruption construct pD2KBC170. The disruptants are "bald" in appearance.

4.15.2 SOUTHERN BLOT HYBRIDISATION TO PROVE PHYSICAL DISRUPTION OF THE GENES ENCODING THE PRODUCTION OF ANTIFUNGAL COMPOUND

The genomic DNA of disruptants 27D1, 2KBC170D1 and C2D1 were restricted with Sp*hI*, and Southern blotted and probed with the 3 kb vector sequence pSOK201 (Fig. 41a,b). The expected restriction profiles of the disrupted chromosome are shown in the Fig. 31-33. A 3 kb Sp*hI* fragment of 2KBC170D1 and C2D1 genomic DNA hybridised to the 3 kb vector sequence pSOK201 and a 5.3 kb Sp*hI* fragment of 27D1 genomic DNA hybridised to the 3 kb vector sequence pSOK201, as expected of successful physical disruption in each case. The wild type *Streptomyces* sp. 98- 62 genomic DNA did not hybridise to the vector probe. This Southern hybridisation result confirmed that the disruption constructs pDE27D1, pD2KBC170 and pDC2 had undergone insertion into the expected region of the genomic DNA of the *Streptomyces* sp. 98 -62.

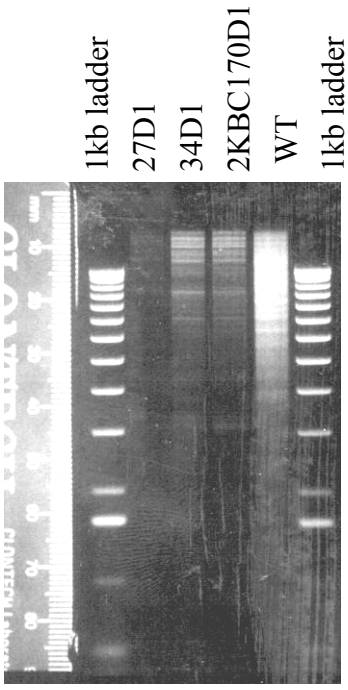


Figure 41a: Electrophoretic profile of digested chromosomal DNA samples of the disruptants 27D1, 34D1, 2KBC170D1 and wild type *Streptomyces* sp. 98- 62

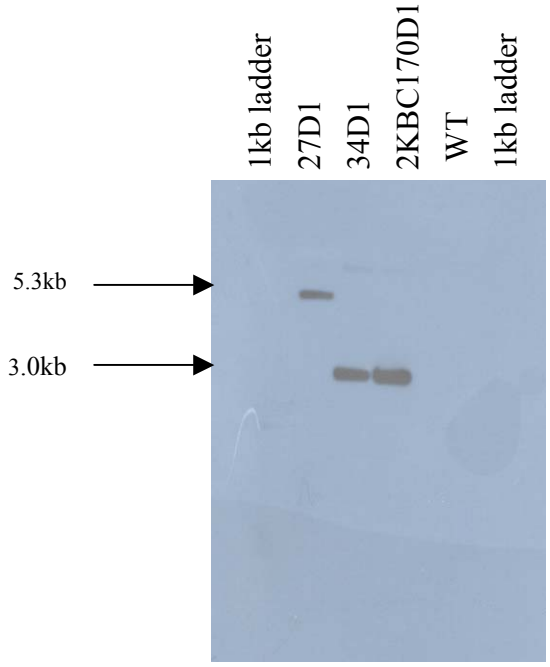


Figure 41b: Southern blot of Sp*hI* digested chromosomal DNA samples of the disruptants 27D1, 34D1, 2KBC170D1 and wild type *Streptomyces* sp. 98– 62, probed with the pSOK201 vector backbone of the disruption constructs.

4.15.3 COMPOUND EXTRACTION AND OVERLAY TO PROVE FUNCTIONAL DISRUPTION OF THE GENES ENCODING THE PRODUCTION OF ANTIFUNGAL COMPOUND

Secondary metabolites of disruptants 27D1, 2KBC170D1 and C2D1 grown in FK medium were extracted analysed by TLC and then bioassayed against *Aspergillus niger* to determine if the antifungal compound biosynthesis by the *Streptomyces* sp. 98-62 was affected by the gene disruption of the PKS I genes in these cases (Fig. 42).

The extracts of the disruptants 27D1, 2KBC170D1 and C2D1 failed to produce a spot on the TLC plate at the Rf value of 0.69, as compared with that of the wild type. Bioassay on the TLC plate revealed that the disruptants 27D1, 2KBC170D1 and C2D1 failed to show spots with any bioactivity. This result indicates that the genes of the PKS I system of the disruptants 27D1, 2KBC170D1 and C2D1 are functionally disrupted.

These results suggest that the cloned PKS I genes of *Streptomyces* sp. 98- 62 are responsible for the biosynthesis of the antifungal compound and that the lack of functional disruption in disruptants 170D1 and 170 D2 is indeed due to the presence of stop/start codons.

Figure 42: TLC chromatogram and overlay assay of the extracts of pure FK506, wild type *Streptomyces* sp. 98– 62,disruptants 27D1, 2KBC170D1, C2D1, C170D1 The cleared area represents the zone of inhibition. The test organism used was *Aspergillus niger.*

This is the first report ever showing that disruption of PKS I genes affects the phenotype of the producing organism. The exact reason is yet to be elucidated. However, it is enticing to postulate that the polyketide compound, encoded by the cloned PKS I genes, has dual functions. One function of the polyketide is to act as an antifungal compound and the other is to play a role in the differentiation of the producing organism.

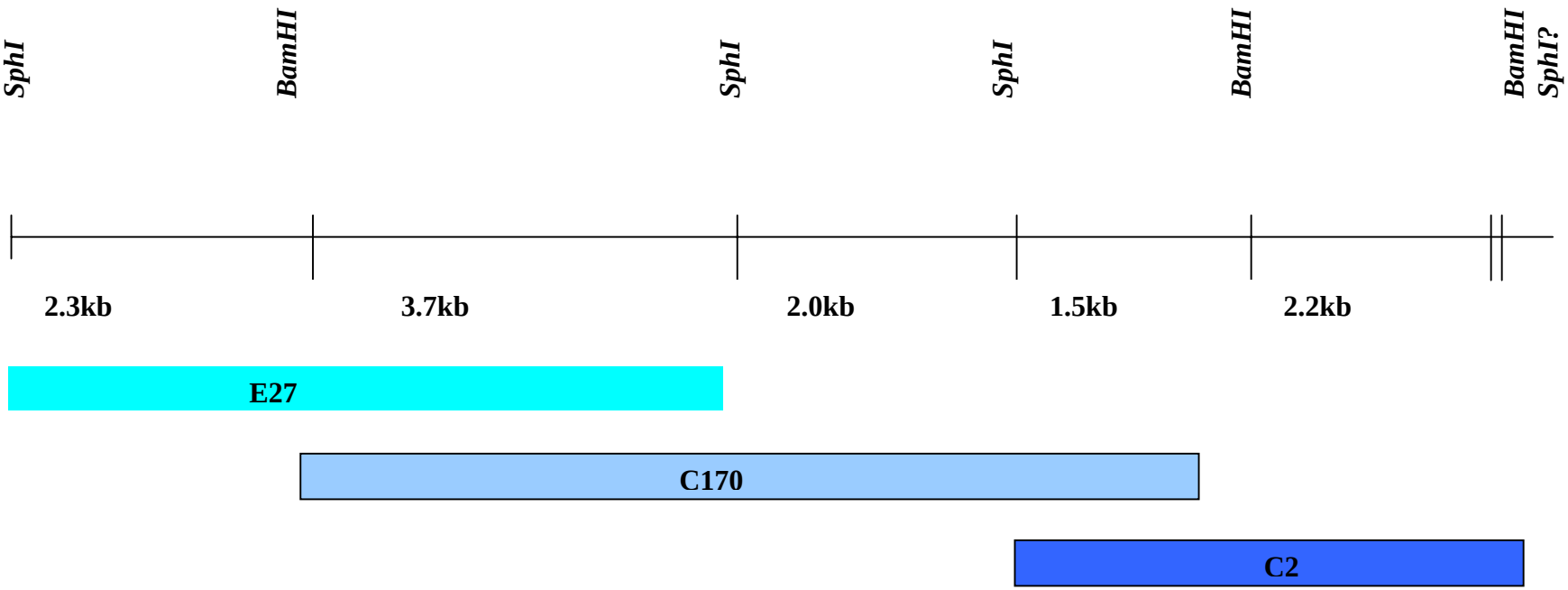


Figure 24: Restriction map of the genomic region of the *Streptomyces* sp. 98 –62 cloned in three contiguous segments in clone E 27, clone C170 and clone C2. The shaded bars below the restriction map indicates the three recombinant clones E27, C170 and C2 as deduced to occur in the genomic DNA of the *Streptomyces* sp. 98- 62. The picture is not drawn to scale. *SphI*? indicates the expected *SphI* site of clone C2.

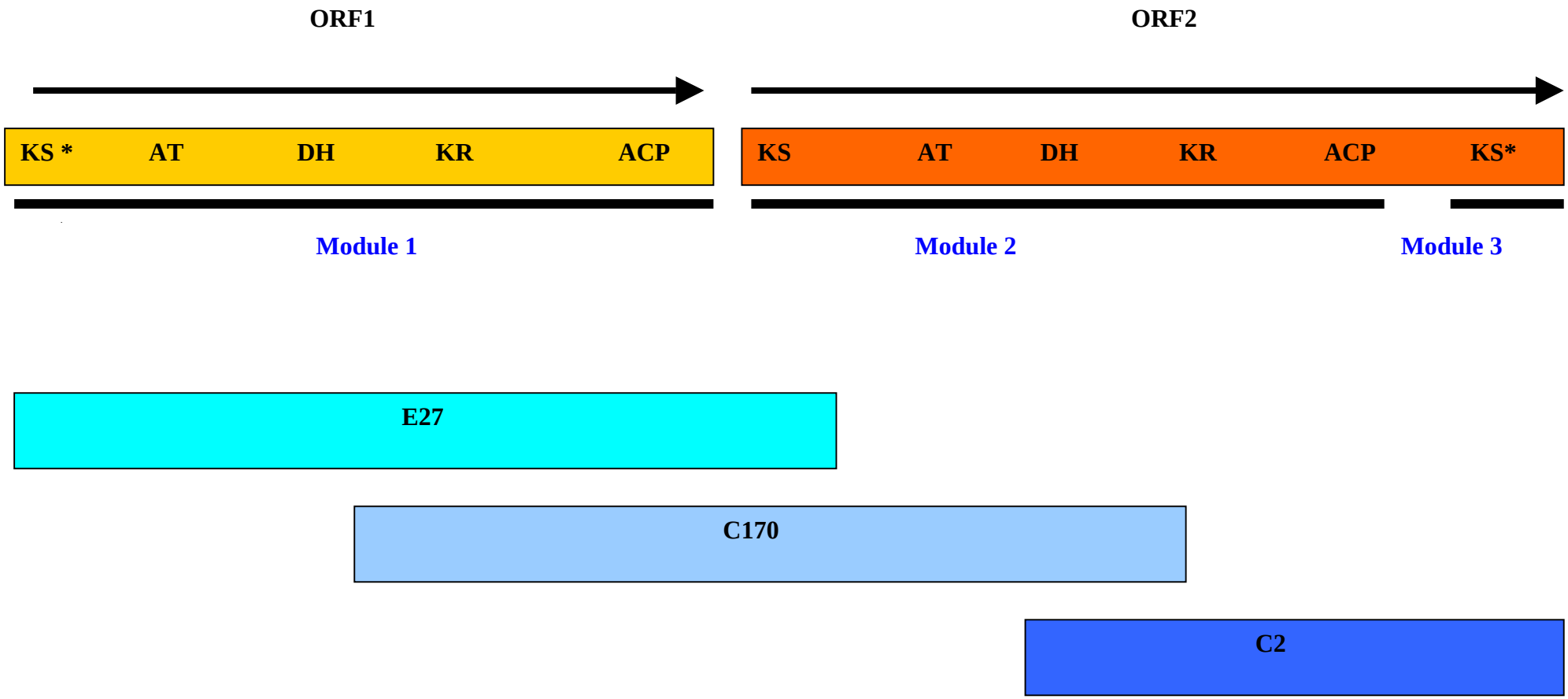


Figure 28: Organization of the PKSI genes from the *Streptomyces* sp. 98 –62. The shaded bars below the restriction map indicates the three recombinant clones E27, C170 and C2 as deduced to occur in the genomic DNA of the *Streptomyces* sp. 98- 62. ORFs,enzymatic domains and modules were identified from DNA sequence analysis. Domains labelled with asterisks are partial. The picture is not drawn to scale.

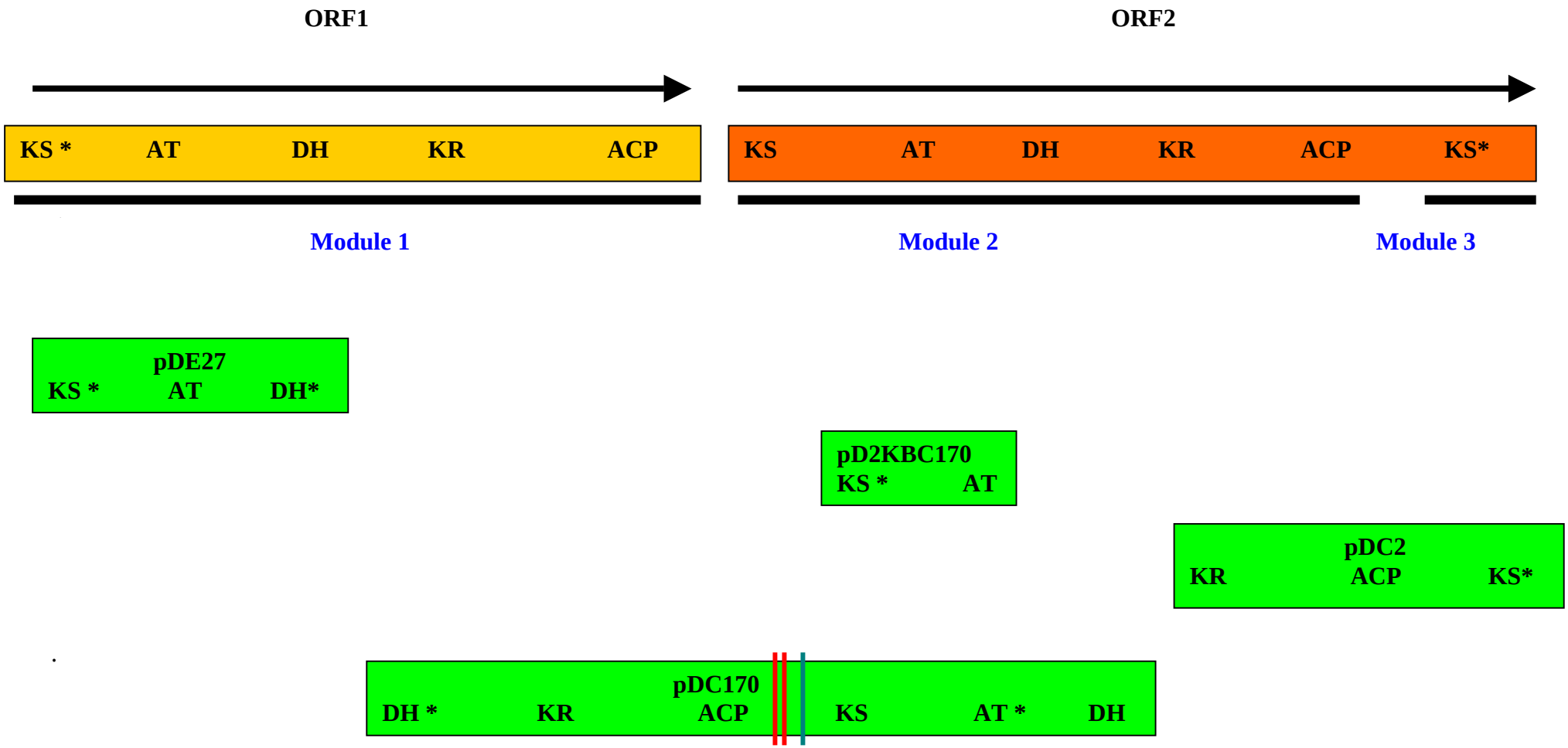


Figure 29: Organization of the gene fragments used in the construction of the disrupted constructs. Domains labelled with asterisk are partial. Red lines in the construct pDC170 indicates the predicted stop codons. Green line in the construct pDC170 indicates the predicted start codon. The picture is not drawn to scale.

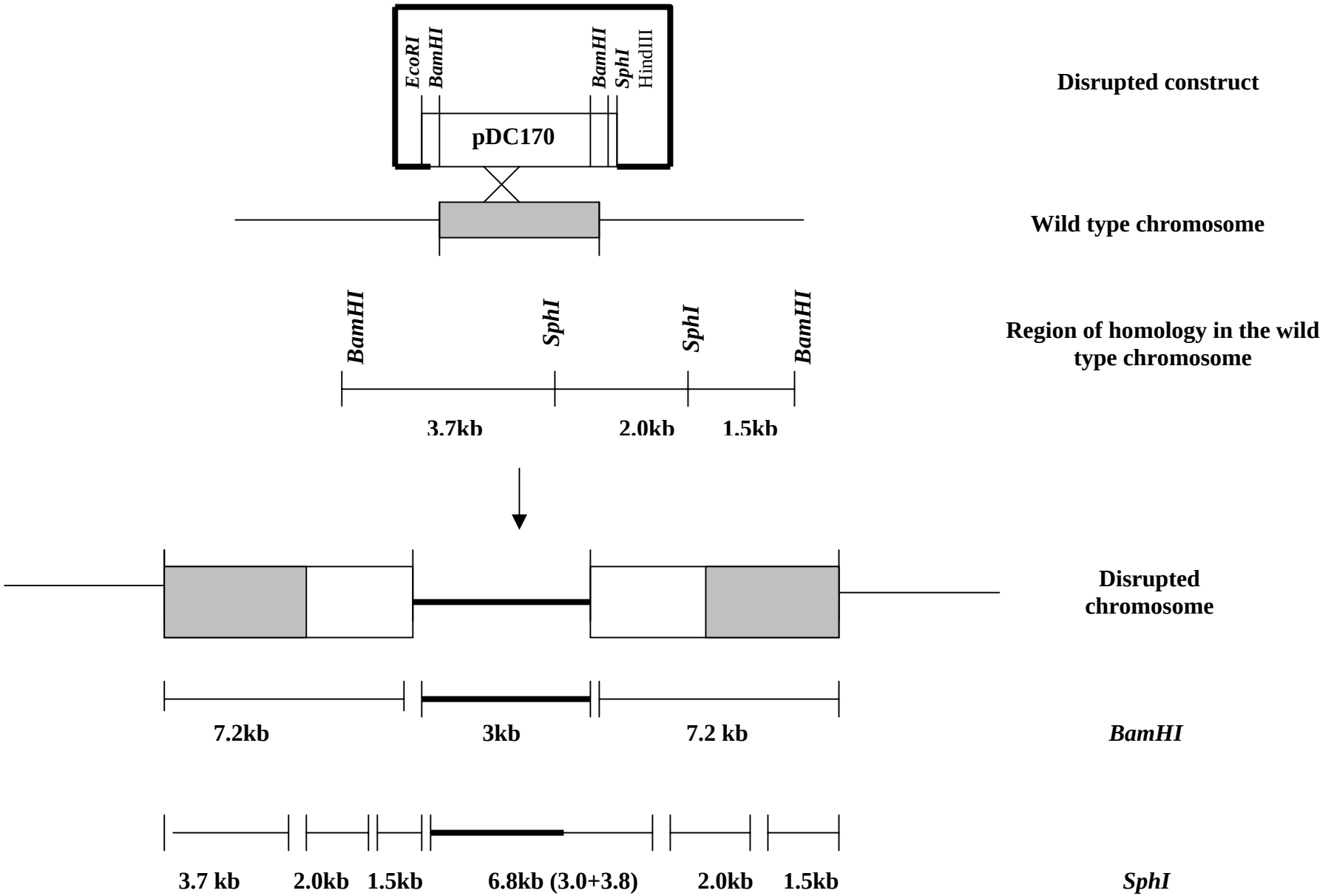


Figure 30: Disruption of the *Streptomyces* sp. 98-62 PKS type I gene using pDC170 disrupted construct. Diagrammatic representation of the integration of pDC170 into the *Streptomyces* sp. 98- 62 chromosome. Region of homology in the chromosome is represented by the shaded box. pSOK201 DNA is represented by the heavy line. *BamHI* and *SphI* sites are indictaed. The picture is not drawn to scale.

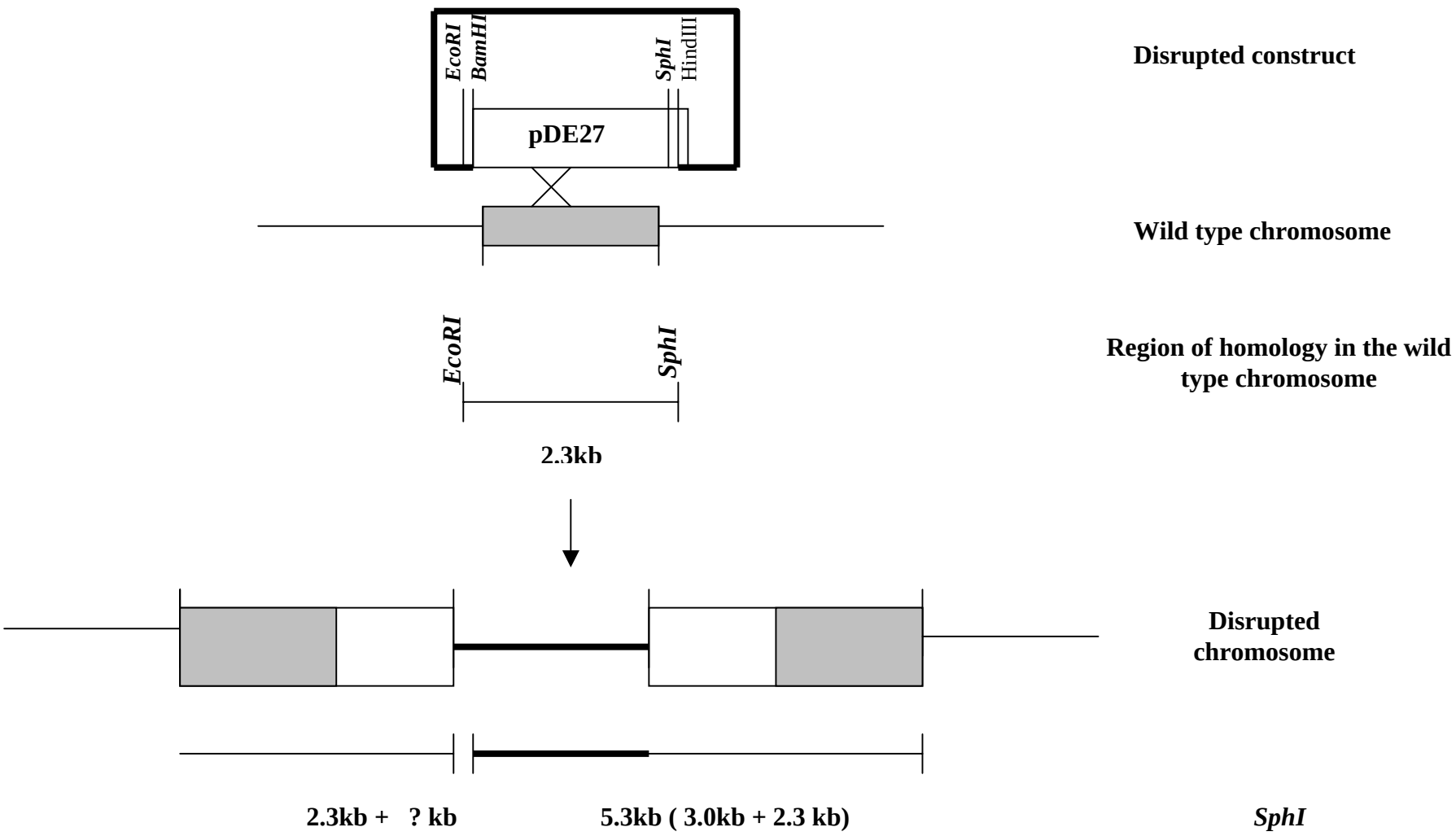


Figure 31 : Disruption of the *Streptomyces* sp. PKS type I gene using pDE27 disrupted construct. Diagrammatic representation of the integration of pDE27 into the *Streptomyces* sp. 98- 62 chromosome. Region of homology in the chromosome is represented by the shaded box. pSOK201 DNA is represented by the heavy line. *Eco RI* and *SphI* sites are indictaed. The picture is not drawn to scale.

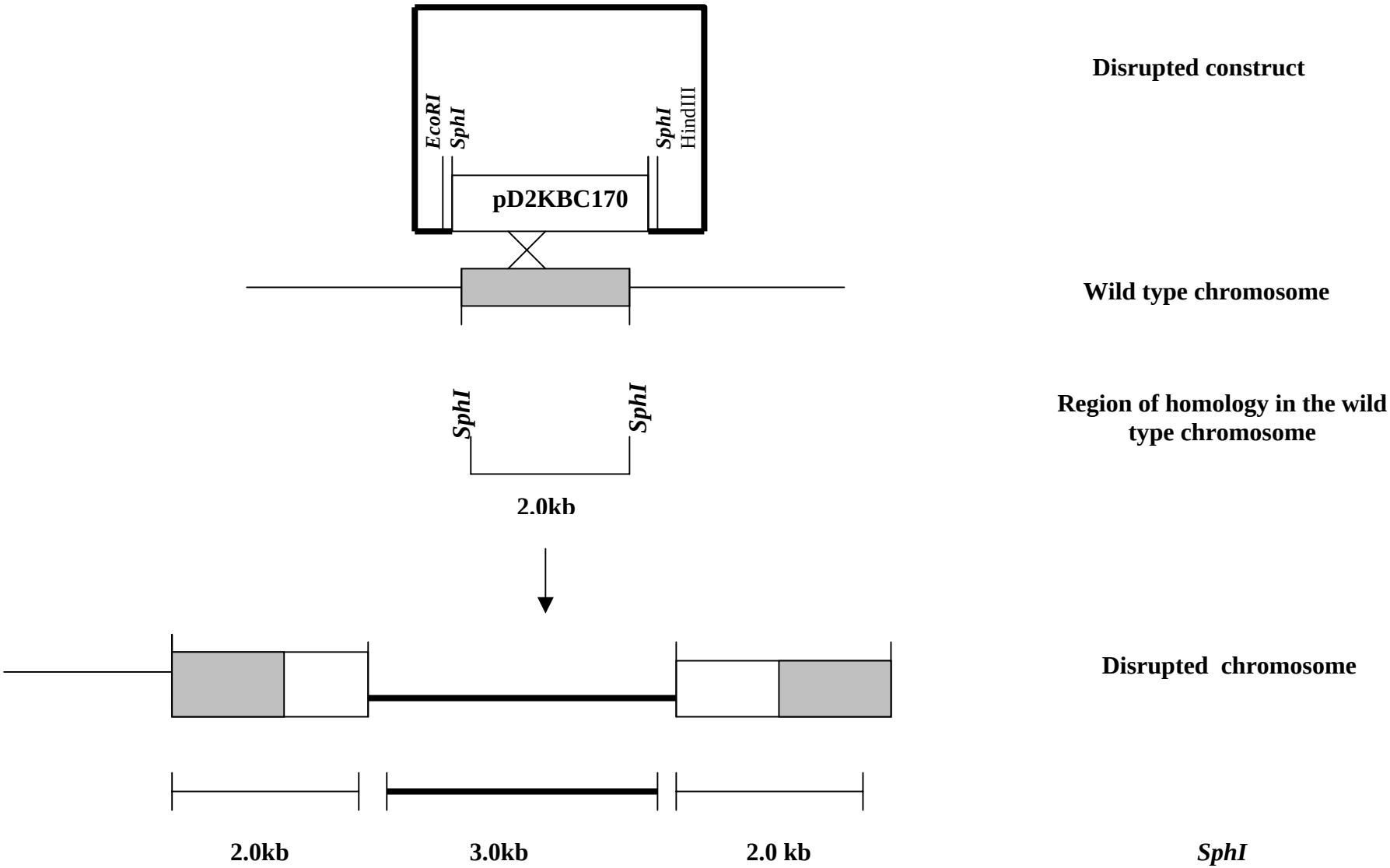


Figure 32: Disruption of the *Streptomyces* sp. 98-62 PKS type I gene using pD2KBC170 disrupted construct. Diagrammatic representation of the integration of pD2KBC170 into the *Streptomyces* sp. 98- 62 chromosome. Region of homology in the chromosome is represented by the shaded box. pSOK201 DNA is represented by the heavy line. *SphI* sites are indictaed. The picture is not drawn to scale.

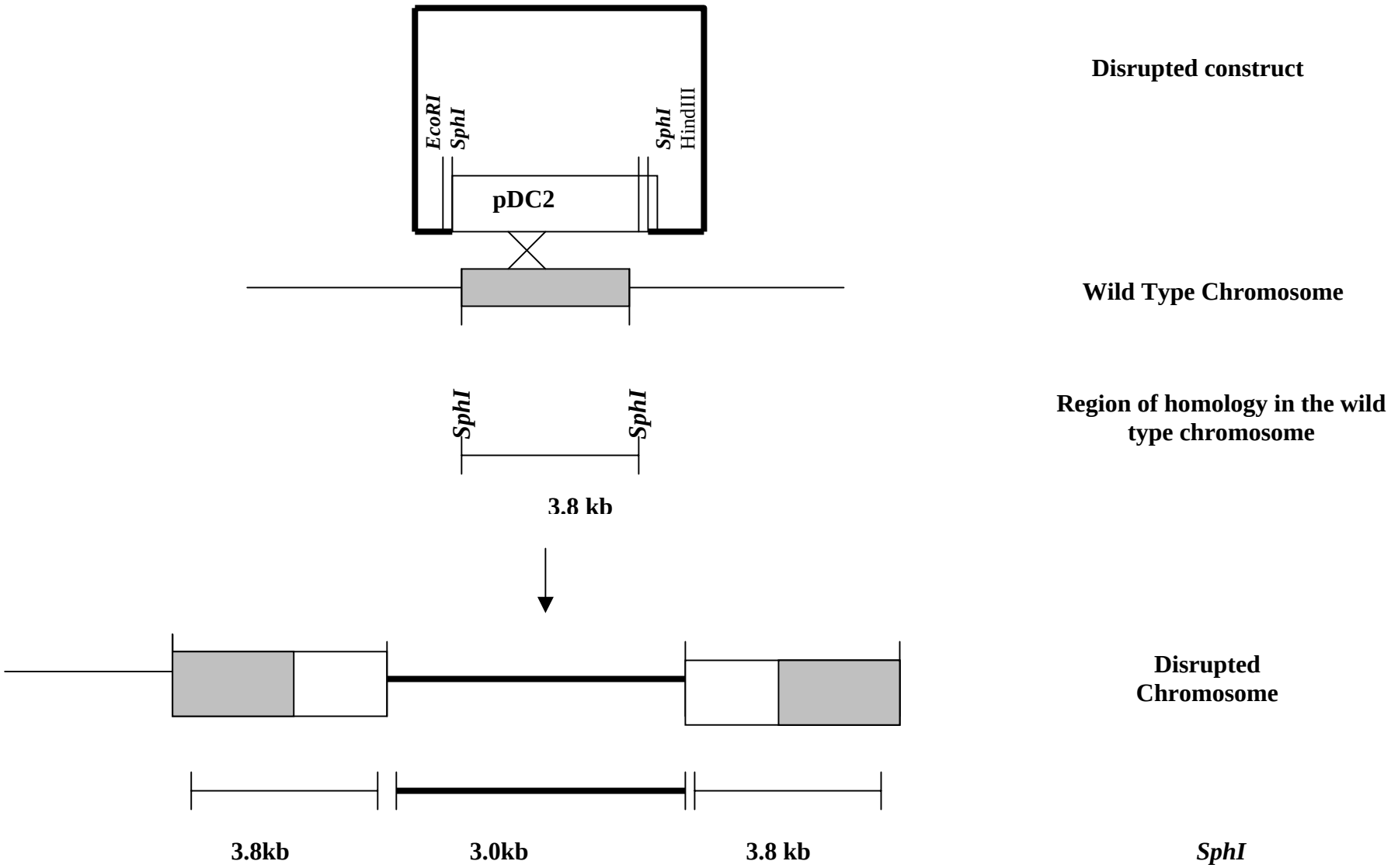
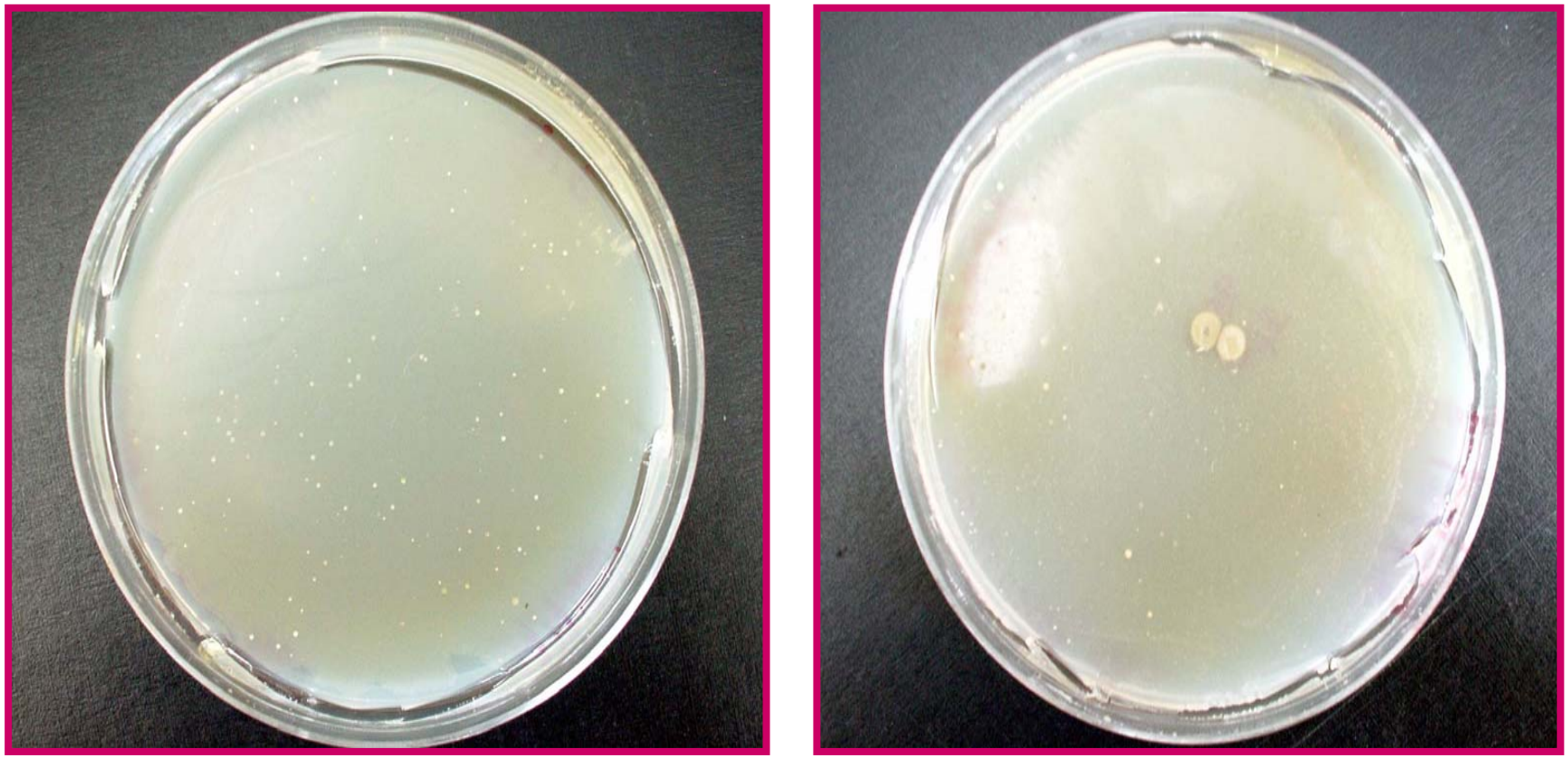


Figure 33: Disruption of the *Streptomyces* sp. 98-62 PKS type I gene using pDC2 disrupted construct. Diagrammatic representation of the integration of pDC2 into the *Streptomyces* sp. 98- 62 chromosome. Region of homology in the chromosome is represented by the shaded box. pSOK201 DNA is represented by the heavy line. *SphI* sites are indictaed. The picture is not drawn to scale.

WITHOUT SELECTION

WITH SELECTION

Figure 36: Conjugation and selection for exconjugants at 30˚C, 12 days



WITHOUT SELECTION

WITH SELECTION

Figure 37: Conjugation and selection for exconjugants at 37˚C, 5 days

In an effort to identify a novel antifungal compound, soil from Singapore were screened for isolates that show antifungal activity. From such a screen, a *Streptomyces* sp. designated 98- 62 was identified to have antifungal activity. The aim of current project is to identify, isolate and characterise the genes encoding the antifungal compound. The present study describes identification of the *Streptomyces* sp., cloning of a number of the PKS I genes from the *Streptomyces* sp. 98- 62, establishment of DNA transfer method for the *Streptomyces* sp. 98- 62 and gene disruption studies to determine the involvement of the cloned PKS I genes in the antifungal compound biosynthesis.

In working with a novel *Streptomyces* sp., it is critical to identify the organism, to allow for informed decision to be made regarding handling of the organism. For example, in studying the genes of the organism it is first necessary to know how to grow the organism for different purposes. It is also necessary to know how to manipulate the organism genetically as in introducing DNA into the organism and so on.

Sequence comparisons of 16S rDNA have been used as a source for determining phylogenetic and evolutionary relationships among organisms of the three kingdoms Archaea, Eukarya, Bacteria. Currently, 16S rDNA sequencing has been used to identify culturable as well as non-culturable bacteria (Amann *et al*., 1995; Stackenbrandt, 1997). A pair of primers designed based on the consensus sequence of bacterial 16S rRNA gene was used to amplify the 16S rDNA from the genomic DNA of the *Streptomyces* sp. 98- 62. Amplification product was cloned and sequenced.

The complete sequence of the cloned amplified product was 1490 bases in length and contained approximately 58.7% G+C nucleotide bases, which is in agreement with the estimated G+C content of the 16S rDNA sequences of

Streptomyces sp. (Wright, 1992). The sequences of the 16S rDNA from several other *Streptomyces* sp. contained approximately 55 to 59% G +C nucleotide bases (Mehling, 1995) although the G+ C content of the total genomic DNA was estimated to be 75%.

The sequence of the 16S rDNA from the *Streptomyces* sp. 98- 62 was highly related (over 95%) to previously published 16S rDNA sequences of other S*treptomyces* (Fig. 9). Sequence analysis of the 16S rDNA of the *Streptomyces* sp. 98-62 showed that there is a deletion of approximately 20nt around nucleotide position 450 when compared to *Escherichia coli* and *Bacillus subtilis*. This deletion can also be found in other actinomycete and related genera such as *Frankia* sp, *Mycobacterium bovis, Arthrobacter simplex, Dermatophilus congolensis* and *Kibdello*sp*orangium radium*. A further deletion of varying length is found within the region of nucleotide 70-90 in most gram-positive bacteria. These deletions were observed in the sequence of *Streptomyces* sp. 16S rDNA. Another region with prominent feature is located around nucleotide position 800 containing the sequence 5'ACATTCCACGTCGTCG-3' which is conserved only in the *Streptomyces* strains but not in the representatives of closely related taxa of actinomycetes or other bacteria. As can be seen in the Fig. 7, this sequence was conserved in the sequence of *Streptomyces* sp. 16S rDNA at nucleotide position 804 (Mehling, 1995). Furthermore, the phylogenetic analysis of the 16S rDNA of the *Streptomyces* sp. 98- 62 with the other actinomycetes showed that the *Streptomyces* sp. 98- 62 grouped together with the genus *Streptomyces* (Fig. 9).

These data provides strong evidence that the *Streptomyces* sp. 98- 62 belongs to the *Streptomyces* sp., making the *Streptomyces* sp. 98- 62 as yet another addition to the existing thousands of known *Streptomyces* sp. capable of producing antimicrobial compounds

Streptomyces studied so far possess varying numbers of rDNA gene clusters. Therefore it is most likely that the *Streptomyces* sp. 98- 62 also carries more than one 16S rDNA. However only one clone of amplified product was sequenced and analysed. This clone would represent one of the many 16S rDNA of the *Streptomyces* sp. 98- 62. Southern hybridisation of the genomic DNA of the *Streptomyces* sp. 98- 62 with this 16S rDNA probe will allow one to determine all the rDNA genes in *Streptomyces* sp. 98- 62. For the purpose of current study, data from a single 16S rDNA was deemed sufficient to identify the *Streptomyces* sp. 98- 62.

As the *Streptomyces* sp. 98- 62 was identified as *Streptomyces* sp., all of the protocols for the manipulation of the *Streptomyces* sp. sp 98- 62 was based on protocols dedicated to *Streptomyces.*

To determine if the antifungal compound produced by the *Streptomyces* sp. 98- 62 is encoded by PKS type I genes, genomic DNA from the *Streptomyces* sp. 98- 62 was subjected to probing with PKS type I specific probe, eryKS II. The gene KS II of eryA from *Saccharopoly*sp*ora erythraea* is usually used to identify type I polyketide synthases. The result from Southern hybridisation revealed that there is homology between the genomic DNA of the *Streptomyces* sp. 98- 62 which produces a novel antifungal compound and *Saccharopoly*sp*ora eryhthraea* gene coding for components of erythromycin PKS. The eryKS II gene from *Saccharopoly*sp*ora erythraea* has also been used to identify the PKS I gene clusters of rapamycin (Molnar, 1996). This experiment therefore has demonstrated that there are DNA regions in the novel antifungal compound producing *Streptomyces* sp. 98- 62, which show a degree of homology to eryKS II gene.

The result is also suggestive that the putative PKS I genes of the *Streptomyces* sp. 98- 62 is different from that of *S. hygrocopicus var. ascomyceticus* ATCC 55098

at the nucleotide level as the Southern hybridization band pattern obtained upon probing the *BamHI* restricted genomic DNA of the *Streptomyces* sp. 98- 62 and *S. hygrocopicus var ascomyceticus* ATCC55098 with eryKS II probe are different from each other. It is to be noted that genes encoding the synthesis of even structurally related PKS I compounds rapamycin and FK506 vary in their sequence at the nucleotide level (Molnar, 1996, Motamedi, 1997).

To determine if the antifungal compound produced by *Streptomyces* sp. 98- 62 was similar to PKS I antifungal compounds rapamycin and FK506, secondary metabolites from *Streptomyces* sp. 98- 62 grown in FK medium were subjected to TLC followed by a bioassay against *Aspergillus niger* to test for antifungal activity. A zone of growth inhibition corresponding to the TLC spot of Rf 0.69 was observed in the case of the *Streptomyces* sp. 98- 62. Positive controls rapamycin and FK506 gave growth inhibition corresponding to the TLC spot with Rf 0.80. Zone of inhibition corresponds to antifungal activity. The zone of inhibition Rf value of the compound produced by the *Streptomyces* sp. 98- 62 grown in FK medium differs from that of pure rapamycin and 63FK 506 compounds, indicating that the chemical nature of the compound produced by the *Streptomyces* sp. 98- 62 is likely to be different form that of the PKS I antifungal compounds rapamycin and FK506.

Rapamycin and FK506 are both macrocyclic polyketides with antifungal and immunosuppressive activity and have share certain degree of similarity in their structure (Fig. 43).

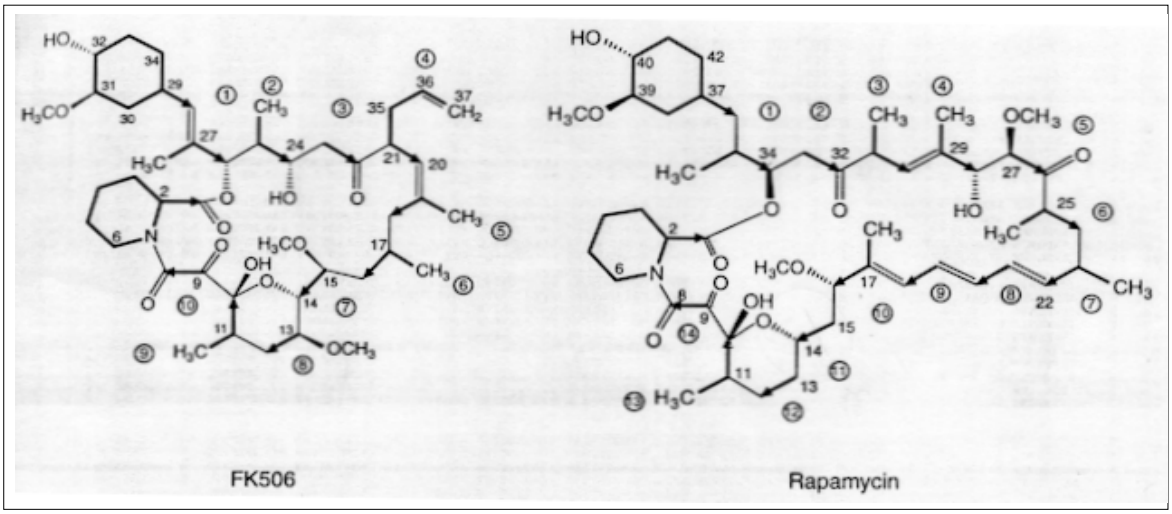


Figure 43: Structures of rapamycin and FK506 (Motamedi, H., 1996)

Both FK506 and rapamycin act via a regulatory domain known as immunophilin binding domain. This domain is the structurally similar region of FK506 and rapamycin. The effector domain of the compounds FK506 and rapamycin is specific for each drug and accounts for their different activities. The similarity in the structure of FK506 and rapamycin is consistent with the enzymology of the biosynthesis of the compounds. The mode of polyketide chain initiation and termination are similar in FK506 and rapamycin biosynthesis. In addition the two pathways are identical in the final three condensation steps (Motamedi, 1997). The homologous enzymes in FK506 and rapamycin biosynthesis are involved in the biosynthesis of the regulatory regions (Molnar, 1996, Motamedi, 1998)

The similarity in enzymology of the biosynthesis of rapamycin and FK506 can also be seen in the organization of the biosynthetic genes of the two compounds. A comparison of the gene clusters for rapamycin from *Streptomyces hygroscopicus* and FK506 from *Streptomyces* sp. *MA6548* reveals that the gene order and direction of transcript of the PKS and peptide synthetase genes, fkb C, B, P, and A and their

equivalents rap B, A, P and C are conserved between the two clusters (Molnar, 1996, Motamedi, 1998) (Fig. 44).

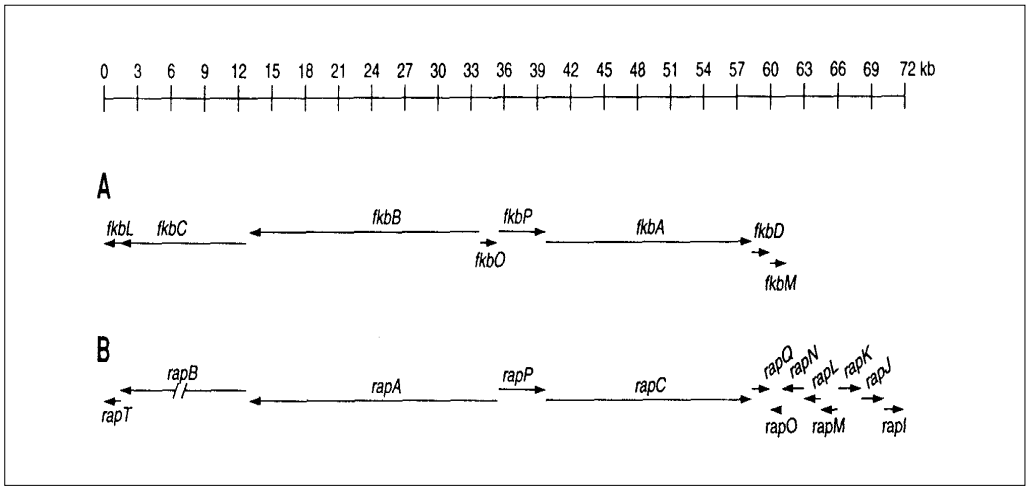


Figure 44: Organisation of the biosynthetic gene clusters of rapamycin and FK506. (Motamedi, H., 1998)

If the deduction from the TLC- bioassay experiment results that the chemical nature of the compound produced by the *Streptomyces* sp. 98- 62 is different from rapamycin and FK506 is correct, it is to be expected that the tertiary structure of the compound which determines the chemical nature is also to be different from that of FK506 and rapamycin. In that case, the enzymology of the biosynthesis of the compound produced by the *Streptomyces* sp. 98- 62 and the organization of the encoding genes of the compound produced by the *Streptomyces* sp. 98- 62 is also expected to differ significantly from those of the genes for rapamycin and FK506. This is to be confirmed upon identification, cloning and characterisation of the PKS I genes form the *Streptomyces* sp.

98- 62.

Based on the evidence that there is homology between the genomic DNA of the *Streptomyces* sp. 98- 62 and the PKS genes of *S. erythraea* which produces erythromycin, a pair of degenerate primers spanning conserved regions of type I PKS

genes, KS and AT gene was used to successfully amplify KS/AT region from the chromosomal DNA of the *Streptomyces* sp. 98- 62. The PCR product ran as a 850 bp fragment on agarose gel. The PCR primers used to amplify the KS/AT genes are designed to amplify the methylmalonyl specific AT gene downstream of the KS gene. AT genes can be malonyl specific or methylmalonyl specific, based on the amino acids in the conserved regions which determines the substrate specificity of the encoding AT enzyme (Haydock, 1995). In our attempt to amplify the KS/AT gene from the *Streptomyces* sp. 98- 62, only one such region was amplified. This could be because the degenerate primers are most suitable for amplifying this region only and no others even if these are present as a result of the degree of homology between the primer and template DNA. The other reason could be that there is only one methylmalonyl Co A specific AT domain in the organism. Different PKS I systems have varying number of methylmalonyl specific AT domains (Fig. 45). Erythromycin PKS I cluster has six methylmalonyl CoA specific AT domains and no malonyl Co A specific AT domains at all. Whereas rapamycin PKS I cluster has seven malonyl Co A specific AT domain and seven methylmalonyl Co A specific AT domains. As we do not know the chemical structure of the antifungal compound produced by the *Streptomyces* sp. 98- 62, it is not possible to estimate the minimum number of methylmalonyl CoA specific AT domain present in the antifungal compound producing *Streptomyces* sp. 98- 62.

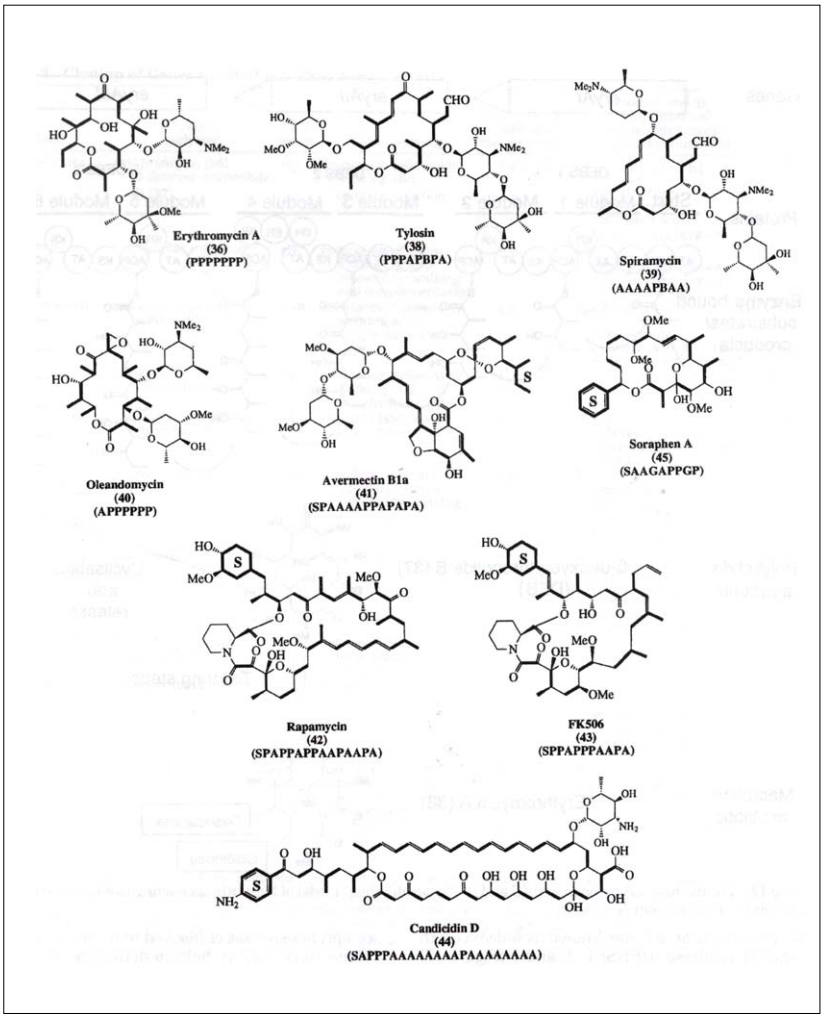


Figure 45: Structures of various complex polyketides built from different acyl units. A stands for acetyl acyl (malonyl CoA) units and P stands for Propionyl acyl (methylmalony Co A) units (Hopwood , D.A., 1997)

The PCR product was subsequently cloned and sequenced using vector primers T7 and SP6. Additional sequencing primers were designed to allow for complete sequencing of the 850 bp insert fragment. The sequence of the 850bp insert fragment from the *Streptomyces* sp. 98- 62 was found to be similar at the amino acid level to KS/AT genes of the other *Streptomyces* PKS I systems. The highest degree of similarity at the amino acid level was with *Streptomyces avermitilis*, which encodes PKS I gene cluster for the biosyhnthesis of the polyketide avermectin (Fig. 14). Successful cloning of the PKS I KS/AT genes form the *Streptomyces* sp. 98- 62 has

proved conclusively that the *Streptomyces* sp. 98- 62 indeed carried the PKS I genes and that the genes are sufficiently conserved enough to the other known PKS I systems that the KS/AT genes from the *Streptomyces* sp. 98-62 can be amplified based on conserved sequences.

To determine if the genome of the *Streptomyces* sp. 98- 62 carried repeated KS/AT genes as is characteristic of PKS I systems, the KS/AT genes of the *Streptomyces* sp. 98- 62 was then used to probe the restriction digested genomic DNA of the *Streptomyces* sp. 98- 62. Eleven Sp*hI* fragments hybridised to the probe, proving that indeed the genome of the *Streptomyces* sp. 98- 62 carried repeated KS/AT genes as is characteristic of PKS I systems.

A 7-8 kb *BamHI* genomic fragment of the *Streptomyces* sp. 98- 62 hybridised very strongly to the KS/AT genes probe of the *Streptomyces* sp. 98- 62. By comparison of this blot (Fig. 15b) with the earlier blot of *BamHI* restricted genomic fragment of the *Streptomyces* sp. 98- 62, probed with the eryKS II probe (Fig. 11), some common features as well as differences could be noticed. For example, A 4-5 kb *BamHI* genomic fragment of the *Streptomyces* sp. 98- 62 hybridised strongest to the eryKS II gene probe of

S. erythraea, whereas a 7- 8 kb *BamHI* genomic fragment of the *Streptomyces* sp. 98- 62 hybridised strongest to the KS/AT genes probe of the *Streptomyces* sp. 98- 62. However the eryKS II gene probe also hybridised to the 7-8 kb *BamHI* genomic fragments of the *Streptomyces* sp. 98- 62.

The eryKS II probe would be able to hybridise only to KS genes of the *Streptomyces* sp. 98- 62 PKS I gene cluster, while the KS/AT genes probe of the *Streptomyces* sp.

98- 62 would be able to hybridise to both KS and (methylmalonyl specific) AT genes of the *Streptomyces* sp. 98- 62 PKS I gene cluster. This could explain partially for the observation that there are some shared as well as different features in the Southern hybridisation experiments with the two probes. The other reason could be that PKS type I genes of the *Streptomyces* sp. 98- 62 may be of two kinds, one that is of higher similarity to that erythromycin genes and another which may be more unique to this specific *Streptomyces* sp.. The PKS genes of the *Streptomyces* sp. 98- 62 are to be cloned and sequenced completely to obtain more conclusive evidence for this deduction.

The variation as well as the numerous hybridisation bands revealed by the two different probes suggest that the *Streptomyces* sp. 98- 62 may carry more than one PKS I cluster. It is known that some strains of *Streptomyces* sp. such as *Streptomyces hygroscopicus* ATCC29253 which produces rapamycin, have multiple clusters of PKS I genes (Ruan, 1997, Lomovskaya, 1997). Evidence from this experiment and observations in other polyketide producers suggest that the occurrence of more than one kind of PKS type I genes in the *Streptomyces* sp. 98- 62 could be possible. Gene disruption would therefore be indispensable to identify the cluster responsible for the antifungal compound production.

In order to study the PKS I genes of the *Streptomyces* sp. 98- 62, it was deemed necessary to clone out a larger fragment/portion of the PKS I cluster form the *Streptomyces* sp. 98- 62. Therefore, attempts were made to obtain a cosmid library of the genomic DNA of the *Streptomyces* sp. using a shuttle cosmid vector pKC505. Intriguingly, the recombinant clones had undergone recombination and seemed unstable. Although there was no strong evidence as to the reason why the instability

was observed, the repeated nature of PKS I genes were thought to be one of the possible reasons resulting in homologous recombination.

In order to overcome this cloning problem, subgenomic library approach was undertaken. The reason being that insert fragments of subgenomic library would be typically less than 10kb. The average size of a PKS I module is approximately 5-6 kb. Therefore insert fragments of sizes below 10kb is likely to constitute 1 to 2 modules of PKS I genes and therefore would have a lesser chance of undergoing recombination within the 10kb (if at all), compared to a 30 kb insert as in the case of cosmid library. Moreover much time had been taken up in constructing the cosmid library and a less time consuming and more cost effective method was needed. Hence the subgenomic library was considered the best available choice of cloning the PKS I genes although it was understood that chromosomal walking in the later stages would be more tedious with this approach due to the repeated nature of the PKS I genes.

From the earlier Southern hybridisation blot of *Streptomyces* sp. 98- 62 genomic DNA probed with the homologous KS/AT genes, a single 7 to 8kb *BamHI* was identified to hybridise very strongly to the KS/AT probe. This 7-8kb *BamHI* fragment therefore was expected to be the genomic fragment of *Streptomyces* sp. 98-62 that contained the KS/AT gene used as a probe.

To clone the PKS I gene surrounding the KS/AT gene, a subgenomic library of *Streptomyces* sp. 98- 62 DNA fragments was constructed. This was done by isolating total genomic DNA, digesting it with *BamHI*, and ligating purified 7 to 8 kb fragments into the *BamHI* site of pUC18. A total of 500 recombinant clones were screened by PCR. One clone designated as C170, gave an amplification product of expected size ~850bp.

Restriction digestion of C170 plasmid DNA with *BamHI* gave an insert fragment of approximately 7-8 kb. This is the expected fragment size and suggests that recombinant clone is likely to contain the PKS I gene surrounding the KS/AT genes used as the probe. Restriction digestion of C170 plasmid DNA with Sp*hI* gave three fragments of approximate sizes 1.6 kb, 2.0 kb and 6.5 kb. The larger fragment of 6.5 kb was expected to contain the pUC18 vector and was confirmed to be so, by self ligation and sequencing.

Complete sequence of C170 was then determined. The entire sequence length of the recombinant clone C170 insert was 7177bp. The DNA sequence data obtained were analysed for open reading frames (ORFs). There were two partial open reading frames, in the same orientation (Fig. 27, 28). The ORFs were labelled ORF 1 and ORF 2 for convenience. ORF1 module was designated as module 1 for convenience, and it encodes a partial DH, a complete KR and a complete ACP in the stated order. ORF 2 module was designated as module 2 for convenience and it encodes a complete KS, a complete AT which is methyl malonyl specific and a complete DH in the given order. The organization of the enzymatic domains within each module is consistent with other PKS type I genes.

ORF 1 is predicted to terminate with a stop codon TGA. A second stop codon TAG is predicted 372 bases downstream of the first stop codon. ORF 2 is predicted to initiate with a start codon ATG and lies 60 nucloetides downstream of the predicted second stop codon of ORF1. The sequence TGGACA which is located 38nt upstream of the predicted start codon of ORF2 is deduced to be the transcriptional promoter as the sequence is identical to ermE-P1 promoter (Strohl, 1992). The sequence GAGG which is located 14nt upstream of the predicted start codon of ORF2 is deduced to be the ribosomal binding site of ORF2 (Strohl, 1992). From the sequence analysis of

clone C170 PKS I genes, it is proposed that the encoded ORFs are translationally uncoupled.

Most of the PKS I gene clusters are bi or tri modular. The maximal number of modules observed so far in a single PKS type I ORF is six. Hexamodularity is observed in the amphotericin B producer *S. nodosus* (Caffrey, P. 2001). The result so far suggests that module 1 of ORF1 is the last module of ORF1 and that module 2 is the starting module of ORF2. The available sequence information is insufficient to determine the modularity of the PKS I genes of the *Streptomyces* sp. 98- 62.

The adjacent genes upstream and downstream genes to the genomic DNA in recombinant clone C170 was identified by probing the genomic DNA of the *Streptomyces* sp. 98- 62 with the external subgenomic fragments of C170. From the result obtained it was deduced that a 5.5–6.5kb Sp*hI* fragment overlaps with and lies upstream to the 7.2kb *BamHI* fragment of C170, and that a 3.5–4. 5kb Sp*hI* fragment overlaps with and lies downstream to the 7.2kb *BamHI* fragment of C170.

Hybridisation of the restriction digested genomic DNA of the *Streptomyces* sp. 98- 62 with the 3.7 kb Sp*hI/BamHI* fragment of the clone C170 or the 1.5 kb Sp*hI/BamHI* of the clone C170 fragment as probes only showed single hybridising band each although multiple hybridising bands are expected as the genes are expected to belong to the repetitive PKS I genes. This could be due to the high stringency primary washes (0.1 X SSC instead of the usual 0.3 X SSC) and repetitive use of the blot after stripping and low concentration of probes used. Whatever the exact reason may be, the band that contains the complete gene sequence as the probe is expected to hybridise the strongest to the probe. Taking this into consideration, the result was taken to indicate that the 5.5-6.5kb Sp*hI* fragment contains the 3.7 kb probe sequence

(Fig. 18a) and that the 3.5-4. 5kb Sp*hI* fragment contains the 1.5kb probe sequence (Fig. 18b).

To clone the PKS I gene downstream of the 7.2kb *BamHI* fragment of clone C170, subgenomic library of *Streptomyces* sp. 98- 62 DNA fragments was constructed. This was done by isolating total genomic DNA, digesting it with Sp*hI*, and ligating purified 3.5 to 4.5 kb fragments into the Sp*hI* site of pUC18. A total of 500 recombinant clones were screened by PCR. One clone designated as C2, gave an amplification product of expected size of approximately 550bp .

Restriction digestion of C2 plasmid DNA with Sp*hI* gave an insert fragment of approximately 3.7kb. This is within the expected fragment size range and suggests that recombinant clone is likely to contain the PKS I gene upstream of the 7.2kb *BamHI* fragment of clone C170 gene. Restriction digestion of C2 plasmid DNA with Sp*hI* gave three fragments of approximate sizes 1.5 kb, 2.1 kb and 2.6 kb, upon gel electrophoresis. The restriction profile matched with the expected profile of a clone that has to overlap with the clone C170 in that the expected 1.5 kb Sp*hI/BamHI* fragment was also observed in the clone C2

The entire sequence length of the recombinant clone C2 insert was 3682bp. Sequence analysis revealed that the 1.5 kb SphI/BamHI fragment of C2 was identical to the external 1.5 kb SphI/BamHI fragment of C170. This confirms that clone C2 is indeed overlapping and upstream to clone C170.

The domains represented in the remaining 2.1 kb sequence are a complete KR, a complete ACP and a partial KS in the stated order. This order of PKS I gene domains is in agreement with the deduction that the 2.1 kb fragment lies downstream of the 7.2kb *BamHI* PKS I gene fragment. The sequence analysis reveals that the 3.8 kb fragment encompasses 2 modules, module2 and a downstream module designated for

convenience as module 3. There is no stop/start codon or ribosomal binding sites or such regulatory sequences between the two modules. This suggests that module 2 and module 3 are translationally coupled and belong to the same ORF, ORF 2.

To clone the PKS I gene upstream of the 7.2kb *BamHI* fragment of clone C170, subgenomic library of *Streptomyces* sp. 98- 62 DNA fragments was constructed. This was done by isolating total genomic DNA, digesting it with Sp*hI*, and ligating purified 5.5 to 6.5 kb fragments into the Sp*hI* site of pUC18. A total of 500 recombinant clones were screened by PCR. One clone designated as E27, gave an amplification product of expected size of approximately 450bp.

Restriction digestion of E27 plasmid DNA with Sp*hI* gave an insert fragment of approximately 6.1 kb. This is within the expected fragment size range and suggests that recombinant clone is likely to contain the PKS I gene upstream of the 7.2kb *BamHI* fragment of clone C170 gene. Restriction digestion of E27 plasmid DNA with Sp*hI* gave two fragments of approximate sizes 3.7kb and 5.4kb upon gel electrophoresis. The restriction profile matched with the expected profile of a clone that has to overlap with the clone C170 in that the expected 3.7.kb Sp*hI/BamHI* fragment was also observed in the clone E27

The entire sequence length of the recombinant clone E27 insert was 6069bp.Sequence analysis revealed that the 3.7kb Sp*hI/BamHI* fragment of E27 was identical to the external 3.7 kb Sp*hI/BamHI* fragment of C170. This confirms that clone E27 is indeed overlapping and up stream to clone C170.

The domains represented in the remaining 2.3kb sequence are a partial KS, a complete AT and a partial DH. This order of PKS I gene domains is in agreement with the deduction that the 2.3kb fragment lies upstream of the 7.2kb *BamHI* PKS I gene fragment, as part of the predicted module 1 of ORF1.

The DNA sequences from the three contiguous clones were aligned and analysed. The aligned nucleotide sequence is 11656 bp in length, and has a high G+C content of 75.3% as expected of *Streptomyces* sp. (Wright, 1992). Three modules of two separate ORFS oriented in the same direction were identified (Fig. 27, 28).

The distance between ORF1 and 2 of the PKS I genes from the *Streptomyces* sp. 98- 62 was 489 bases. Comparison with the erythromycin gene cluster reveals that the ORF1 and 2 of the erythromycin gene cluster was separated by 1.44kb whilst ORF2 and 3 were contiguous (Leadlay *et al.*, 1992).

A complete module (KS-AT-DH-KR-ACP) of the *Streptomyces* sp. 98-62, module 2 is of the size 1743 aa. This is in agreement with other PKS I gene clusters (Table 5). The limits of each domain within the modules were readily assigned by comparison with the modules of B-deoxyerythronolide synthase and rapamycin synthase (Fig. 42) (Bevitt, 1992, Molnar, 1996). Individual domains of the modules are also relatively similar to those of erythromycin and other PKS I clusters (Table 5). KS domain is approximately 421aa. AT domain is approximately 315 or 343 aa. DH domain is approximately 164 or 167aa. KR domain is approximately 233 or 234aa. ACP domain is approximately 77aa. Although module 1 is incomplete, the domain size is comparable to that of module 2 (Fig. 46).

MULTIPLE SEQUENCE ALIGNMENTS OF THE 3 MODULES OF THE SOIL ISOLATE 98- 62

```
MOD1            .......... .......... .......... .......... ..........
MOD2            LREAMLENER LRRQNDRIAE AAHEPVAVVA MSCRYPGGVG TPEQLWQLVD
MOD3            .......... .......... ...EPVAIVG MACRYPGGVT TPEELWRLVA
                ═══════════════════════════════════════════════════════

MOD1            .......... .......... .......... .......... ..........
MOD2            AGVDAVGDFP DDRDWDVDAI YDPDPDAPGR THVREGGFLH DAPRFDPGFF
MOD3            DGVDGIGAFP DDRGWNLDTL YDPEPGKPGH CSTRAGGFLY DAADFDHDFF
                ═══════════════════════════════════════════════════════

MOD1            .......... .......... .......... .......... ..........
MOD2            GISPREALAM DPQQRLLLET AWEAFERGGI DPHTLRGSRT GIYAGVMYHD
MOD3            GIGPREALAM DPQQRLLLET SWEALERAGI DPHSVRGSRT GVFAGVMYHD
                ═══════════════════════════════════════════════════════

MOD1            .......... .......... .......... .......... ..........
MOD2            YGSWLTDVPE GVEGYLGNGN LGSVASGRVS YTLGLEGPAV TVDTACSSSL
MOD3            YGSRLRDVPE AVRDYLGNGS LGSIASGRIA YTLGLEGPAL TVDTACSSSL
                ═══════════════════════════════════════════════════════

MOD1            .......... .......... .......... .......... ..........
MOD2            VALHLAVQAL RTGECALALA GGVTVMSTPD TFIDFSRQRG LALDGRCKSF
MOD3            VALHLAAQAL RRGECGLALA GGVSVMSTVD TFVDFSRQRN LAADGRAKSF
                ═══════════════════ KS ═══════════════════════════════

MOD1            .......... .......... .......... .......... ..........
MOD2            AEGADGTGWG EGVGMLLLER LSDARRNGHR VLAVVRGTAV NQDGASNGLT
MOD3            AEAADGTALS EGVGVLVLER LSDARRSGRR VWGVVRGSAV NQDGASNGLT
                ═══════════════════════════════════════════════════════

MOD1            .......... .......... .......... .......... ..........
MOD2            APNGPSQQRV IRAALADARL EPHQVHAVEA HGTGTPLGDP IEAQALLATY
MOD3            APNGPAQQRV IREAWVAAGV SGGGVDVVEA HGTGTVLGDP IEAQALLSTY
                ═══════════════════════════════════════════════════════

MOD1            .......... ........MG HTQAAAGVAG IIKMVMAMRH GTLPRTLHVD
MOD2            GQDRQAGEPL WLGSVKSNIG HTQAAAGVAG VIKMVMAMRR GRLPRTLHAE
MOD3            GQGRGGGD.. .......... .......... .......... ..........
                ═══════════════════════════════════════════════════════

MOD1            TPSHQVDWTT GAVRLLTEER PWPGAADRPR RAGVSSFGIS GTNAHVILEE
MOD2            HPTTRVDWES GAVELLGEAR DWPDAGE.PR RAAVSSFGIS GTNAHVIVEA
MOD3            .......... .......... .......... .......... ..........
                ═══════════════════════════════════════════════════════

MOD1            FEEFEEFAGE PVGTGPRTAG PDADGHDGAA AHPPATPPVL ALPVSARSPE
MOD2            APDPEPRTGE PVWDRP.... .......... .........L PLVLSARDEP
MOD3            .......... .......... .......... .......... ..........

MOD1            ALRGQAARLR ELTGTSA... AELGLALSTT RTTHPYRAVV LAPGEERADE
MOD2            ALAAQARRIL DHLETGADLV PDIAYALATT RAALDRRAVV IGADPATITA
MOD3            .......... .......... .......... .......... ..........

MOD1            ALDALAHGHE APGLLVSGSI TDGTLACLFS GQGAQRPGMG RDWYDTFPVY
MOD2            RLAALAEDDP ASDVVRGAPA GESRIAFVFP GQGSQWAGMA AELLDGSPVF
MOD3            .......... .......... .......... .......... ..........
                                       ═══════════════════════════════

MOD1            AEHFDRTGEL FAKHLERALA EVVLGDHPDV LERTAYTQAA LFTTQVALYR
MOD2            AAAMADCAEA LAPFTDWDLV DTVRERRP.. MERVDVVQPA LWAIMVSLAE
MOD3            .......... .......... .......... .......... ..........
                ═══════════════════════════════════════════════════════

MOD1            LLESFGLRPD WLAGHSVGEF AAAHVAGVWS LQDAVTAVAA RGRLMQALPE
MOD2            VWRAHGVRPA AVIGHSQGEI AAACVAGALS LSDGARVVAL RSRAIAEVLS
MOD3            .......... .......... .......... .......... ..........
                ═══════════════════════════════════════════════════════

MOD1            G......... .......... ......GAMT AVQAAEEEVR PLL...DERC
MOD2            GPADSGTVPG KGASGPTNSA RGACGRGGMM SVALPESRAR ELVAAHDGRV
```

```
MOD3            .......... .......... .......... .......... ..........
                                      AT
MOD1            DIAAVNGPRA VVVSGDEDAV AAVAAHFAT. ...TRRLRVS HAFHSPRMEP
MOD2            AVAAVNGASS VVLSGDAEVL DALRERIVAD GGRAKRLPVD YASHCAHVES
MOD3            .......... .......... .......... .......... ..........

MOD1            VLDEFRRVLA ALPAGEPALP IVSTLTGARA TAAELGSADY WVRHVRETVR
MOD2            IRERLLTDLA GVRARGADVP FYSTVTGAVL DTTAL.TADY WYTNLRRSVL
MOD3            .......... .......... .......... .......... ..........

MOD1            FADAVGTLAA QGADTFLELG AAPVLTALGP DCLPDADAEE AAFVPTARKG
MOD2            FEPTTRALLD SGYGIFVECS PHPVLLNS.I EETADAVGAT VTGLGSLRRD
MOD3            .......... .......... .......... .......... ..........

MOD1            TAEVPGLLAA LAAVHTRGSD VDWAVLYDGL PGHRDRPGRR DEPGHRDQPG
MOD2            DGGAERLLTS LGEAFVAGVP VDWSAVFTGM P......... ..........
MOD3            .......... .......... .......... .......... ..........

MOD1            RRDQPGRRVE PGRCVELPTY AFQHRRYWLP TSTATARGDA AGHGLAAVDH
MOD2            .......... .VRAADLPTY PFQRERYWLG RSAASG..DV TAAGLRATTH
MOD3            .......... .......... .......... .......... ..........

MOD1            PFVSARLDLP GDGGTLLTGR ISTATHPVLA QHAVLGSVLV PGAALVDLAL
MOD2            PLLGAAVQVA G.GGTLFTGR LSVSTTPWLA DHAVSGTPLL PGTALVELAL
MOD3            .......... .......... .......... .......... ..........
                                      DH

MOD1            YASGLTGRPV LEELTLQAPL ALPGNGAVRI QVALRPDG.. ...GVEIHSR
MOD2            SAGHELGYGH VAELTLQAPL VLPGRAAVQF QVHVAAADED GHRALTVHSR
MOD3            .......... .......... .......... .......... ..........

MOD1            PADAPEDGSW TRHATGTLTV TDPASGLPAS SVPSAAWPPP GAVPLDTDGL
MOD2            PEGA.DDTEW TAHATGLLAP RTAPPGFDL. ....AQWPPR GAEPVLVDDA
MOD3            .......... .......... .......... .......... ..........

MOD1            YERLRGEGYD YGPVFQGVRA AWRHGDTVLA ELELPAEARQ DAARHVLHPA
MOD2            YDTLAALGYD YGPAFQGLRA VWRRGDETFA EVELPGEAGA FGLHPALFDA
MOD3            .......... .......... .......... .......... ..........

MOD1            LLDSALHTTA LADADARDAV PDGTIALPFA WTGVTVHGRP SSRTTPSRTG
MOD2            ALHADGLRTA PPGTDGPGAR GQGAARLPFV WTGVSLY... ..........
MOD3            .......... .......... .......... .......... ..........

MOD1            VPSRAAAPDH TAARVRVTRG EEGIRLDLTD TEGGPLATVA SYVTRPVTAD
MOD2            ......ASGA TALRVRIRGG D.TLSLDLAD PTGAPVAAVE ALVSRPVDPA
MOD3            .......... .......... .......... .......... ..........

MOD1            RLTGRQRSLY VVEDAPLPES AGRPERRTWA VLGPDDLGLG VPHHPEPAAI
MOD2            ALTSPVR... .......DDD LYRLDWQALP VPVADAPAYA VLDERGTAAA
MOD3            .......... .......... .......... .......... ..........

MOD1            DGPAPDVVVL PVHIPDVADA DADGERVPGA VRTALNTTLT TLRAWLDDER
MOD2            D.AVPDWVVL PVSG...... ..DGGDPVGG VRAATGRVLA AVRDWLADER
MOD3            .......... .......... .......... .......... ..........

MOD1            RAGS...TLL VLTEG..... .....SLADA AVHGLVRAAQ AEHPGRIVLV
MOD2            TAGARGARLV VLTGGAVATG TEDVTDLAGA AVWGLVRAAQ GEHPDRFVLV
MOD3            .......... .......... .......... .......... ..........
```

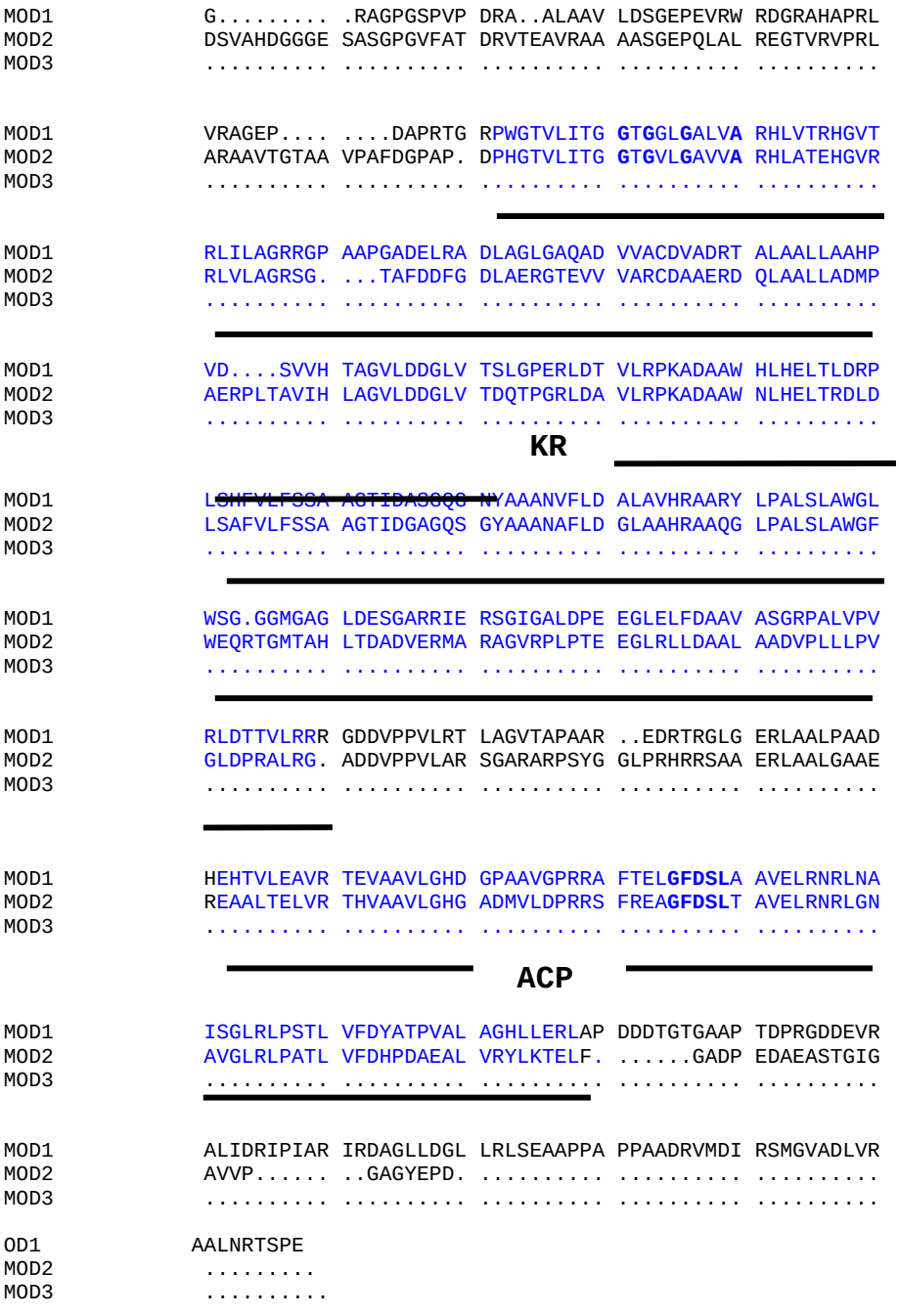


Figure 46: Alignments of the 3 modules of the *Streptomyces* sp. 98- 62. Domains are represented in blue colour. The identity of each the domain is indicated along the black bar underlining the sequences. The active sites of each domain are in bold.

	Streptomyces sp. 98- 62 module1 and module 2. Module 2 aa in brackets if different from module 1.	Niddamycin module 3 (Kakavas *et al*, 1997)	Rapamycin module 11 (Molnar *et al*, 1996)
Complete module	1743 aa	1839 aa	1629 aa
KS domain	421aa	424 aa	452 aa
AT domain	315 (343 aa)	334 aa	292 aa
DH doman	164 (167 aa)	190 aa	150 aa
KR domain	233 (234 aa)	185 aa	243 aa
ACP domain	77 aa	86 aa	76 aa

Table 5: Comparison of the number of aminoacids constituting the domains and modules of PKS I genes cloned from the *Streptomyces* sp. 98- 62 with that of the nidddamycin and rapamycin PKS I genes

The nucleotide sequence of the cloned genes and the repeated occurrence of the genes isolated from the *Streptomyces* sp. 98- 62 as modules and the organization of the domains within the modules provide strong evidence that they belong to the PKS type I genes.

From the domain organization of the cloned PKS I genes from the *Streptomyces* sp. 98- 62, it is predicted that the cloned modular genes are both extender modules. Loading modules of other streptomycete PKS I gene clusters generally have the essential catalytic domains (KS, AT and ACP). The cysteine residue of these PKS I loading module KS domain active sites are also typically replaced with serine or glutamine. Extender modules of the other streptomycete PKS I gene clusters typically

consist of the essential domains (KS, AT and ACP) and 1 to 3 of the modifying domains (DH, KR and ER). Releasing modules of other known streptomycete PKS I gene clusters typically contain an additional thioesterase domain to the essential and modifying domains. The two PKS I modules isolated from the *Streptomyces* sp. 98- 62 have both the essential domains as well as modifying domains (KS, AT, DH, KR and ACP) but no thioesterase domain. Moreover the active site cysteine residue of both the KS domains cloned from the *Streptomyces* sp. 98- 62 are conserved. This sequence information suggests that the two PKS I modules of the *Streptomyces* sp. 98- 62 are likely to be involved in the extension of the polyketide biosynthesis rather than initiation or termination of the polyketide biosynthesise.

Sequences of the individual domains of the modules were compared within the cluster and with other PKS I clusters for sequence homology (Table 6) to predict the activity of the deduced domain based on the conserved amino acid (Fig. 46). These conserved amino acid sequences are known to be required for the catalytic function of the encoded gene product.

MODULE	KS	AT	DH	KR	ACP
1	*S.hygroscopicus var. ascomyceticus* 89% +ves	*S.nodosus* 67% +ves	*S.avermitilis* 62% +ves	*Polyangium cellulosum* 69% +ves	*S.antibioticus* 72% +ves
2	*S.avermitilis* 84% +ves	*S.spinosa* 67% +ves	*S.avermitilis* 70 % +ves	*S.nodosus* 78% +ves	*S.avermitilis* 73% +ves
3	*S.avermitilis* 83% +ves				

Table 6: Comparison of domains of PKS I genes cloned from the *Streptomyces* sp. 98-62 with other PKS I genes of *Streptomyces* sp.

KS DOMAIN

Table 6 shows the percentage of homology with each individual domains of other PKS I clusters of *Streptomyces* sp. KS domain was the most conserved domain in the cluster. Homology of the KS domains within the cluster was determined to vary from 65-74% similarity by Multiple Sequence Alignment. Comparison of the KS domains with other type I PKS revealed that the conserved actives site motif TACSS is invariant in modules 2 and 3 of the *Streptomyces* sp. 98- 62. Sequence information is insufficient to determine that of module 1.The cysteine residue in the conserved sequenced is required for the KS to be active, and is required for the formation of a thio ester linkage to the growing acyl chain. As such modules 2 and 3 could be

predicted to be active. Two other His residues are also reported in other active KS genes. Module 2 contained both the His residues. Module 1 and 3 had one of the two His residues each. It is indeterminable from the available sequence as to the presence of the second His residue. Fig. 46 shows the conserved residues in bold.

AT DOMAIN

The AT domains of PKS genes from *Streptomyces* sp. 98- 62 show more sequence variability than the KS domains. It has been demonstrated the AT domains fall into two distinctive classes and this can be distinguished from the conserved motifs in the AT domain (Haydock, 1995). As a result, substrate specificity of the AT domain can be determined from the primary amino acid sequence. This analysis shows that the AT domain of module 1 PKS I genes from *Streptomyces* sp. 98- 62 has substrate specificity for malonyl CoA and the AT domain of module 2 PKS I genes from *Streptomyces* sp. 98- 62 has substrate specificity for methylmalonyl CoA (Fig. 47).

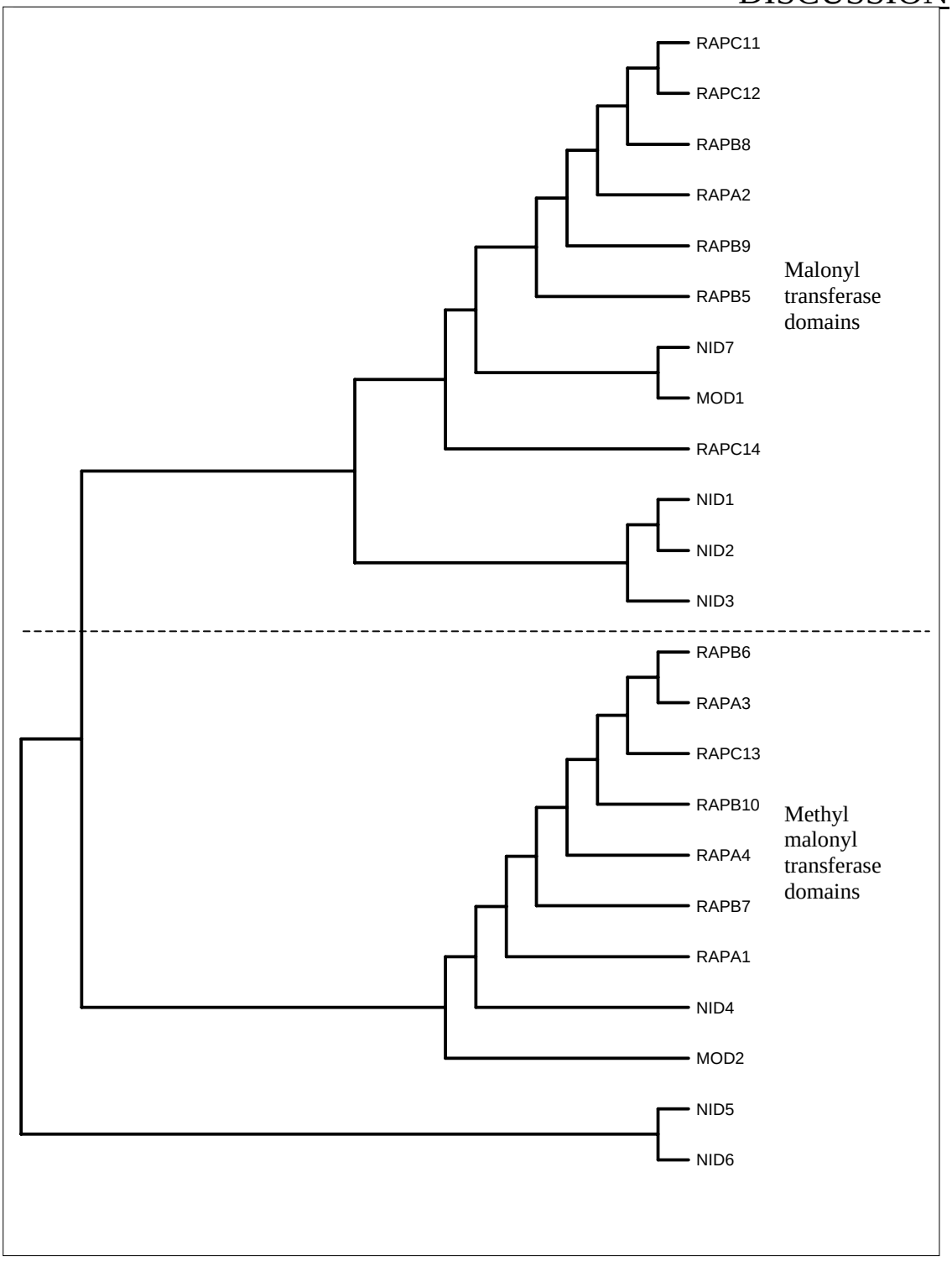


Figure 47: Phylogenetic analysis of acyltransferase domains. Phylogenetic tree of aminoacid sequences of acyl transferase domains from *Streptomyces* sp. type I PKS showing clustering of malonyl and methylmalonyl loading domain sequences. The PKS I genes used for comparison are that of rapamycin, denoted as RAP, niddamycin denoted as NID and that of the *Streptomyces* sp. 98- 62, denoted as MOD. The tree was constructed using the CLUSTALW program. The relatedness between different domains is indicated by the length of the horizontal line. The shorter the horizontal line, the more closely related the domains. The length of the vertical lines are not significant.

The AT domains of module 1 and module 2 were only 27.6% similar The AT domains of module 1 and module 2 were only 27.6% similar by Multiple Sequence Alignment. The sequence difference in malonyl CoA specific AT domain and methylmalonyl CoA specific AT domain would explain the low homology between the AT domains of the *Streptomyces* sp. 98- 62 modules. Similar observations have been noted between niddamycin AT2 and AT6 domains, where the similarity between the malonyl CoA specific AT domain AT2 and methymalonyl CoA specific AT domain AT 6 is about 30% (Kakavas *et al.*, 1997). It is predicted that if the same substrate specific AT domains of the *Streptomyces* sp. 98-62 are to be compared with each other, a higher homology between the domain sequences would be obtained. Niddamycin AT2 and AT3 domains are malonyl CoA specific and share an aminoacid identity of 95% (Kakavas *et al*., 1997).

Both the identified AT domains of the *Streptomyces* sp. 98- 62 retain the active site sequence GHSXG. The Ser residue of this consensus sequence is involved in the formation of the acyl enzyme intermediate. In addition there is also a conserved His residue about 100aa downstream of the active site XAXHX, which is invariant in other AT domains. This His residue is believed to be involved in the catalysis of acyltransferases. Two other Gln and Arg are invariant among all AT domains. These residues were also maintained in both the identified AT domains. Fig. 46 shows the conserved residues in bold. From the result, it can be predicted that the AT domains of the *Streptomyces* sp. 98- 62 are both active; module 1 AT domain being specific for malonyl CoA and module 2 AT domain being specific for methylmalonyl CoA.

The homology between the degenerate primer sequence used to amplify the KS/AT region from the *Streptomyces* sp. and the sequence of KS/AT region of the modules 1 and 2 sequnce of the *Streptomyces* sp. 98- 62 cloned were compared. The

comparison revealed that the degenerate forward and reverse primers were 100% identical to the sequences in module 2 KS/AT region. The degenerate primers however only had 16.67% identity with the forward primer and no identity with the reverse primer. This could be expected because the degenerate reverse primer is designed from the methyl malonyl CoA substrate specifying enzyme. AT domain of module 1 is however malonyl CoA substrate specific. Hence it could be concluded that sequence homology of the degenerate primers plays a significant role in amplifying a gene product.

From the data obtained that the malonyl CoA specific AT or methylmalonyl CoA specific AT domains are different, it will be useful to employ malonyl CoA specific AT or methylmalonyl CoA specific AT domains as probes to probe the genomic DNA of the *Streptomyces* sp. 98- 62, to determine the number of specific domains. These probes will be more specific than KS probes and will give strong hybridising bands to the respective homologous gene sequences in the genomic DNA.

ACP DOMAIN

The ACP domains of PKS genes from the *Streptomyces* sp. 98- 62 also show more sequence variability than the KS domains. ACP domains of module 1 and module 2 were 57% similar by Multiple Sequence Alignment. The pantothiene binding Ser residue in the GFDSL motif was present in the ACP domains of both modules 1 and 2, indicating that these domains of the *Streptomyces* sp. 98- 62 are likely to be functional

Fig. 46 shows the conserved residues in bold.

DH DOMAIN

Domains with predicted reductive functions are DH and KR domains. DH domains of module 1 and module 2 were 39.6% similar by Multiple Sequence Alignment. Highly conserved His, Gly and Pro residues of the HXXXGXXXXP conserved sequence were retained in both the modules 1 and 3, predicting that these domains of the *Streptomyces* sp. 98- 62 are likely to be functional. Fig. 46 shows the conserved residues in bold.

KR DOMAIN

KR domains of module 1 and module 2 were 57.9% similar by Multiple Sequence Alignment. Designating the limits of KR domain was difficult as the C terminal sequence of the KR domains varied slightly from that of the rapamycin domains. Active KR domains are expected to have a NADP (H) binding site GXGXXAXXXA. The first invariant Ala residue in the motif has been found to be replaced by Gly residue sometimes. The KR domains of module 1 and module 2 retain the predicted sequence of GXGXXGXXXA, where the first Ala residue is substituted with Gly residue. Therefore both module 1 and module 2 KR of the *Streptomyces* sp. 98- 62 are predicted to retain the activity. Fig. 46 shows the conserved residues in bold.

DH and KR functions and as the DH and KR domains of modules 1 and 2 are predicted to be active, the enzyme encoded are expected to be involved in the formation of a double bond in the PKS I compound that the enzymes biosynthesize. However, it has to be determined if the reduction functions are reflected in the structures of the PKS I product formed. Modules 3 and 6 of Rapamycin with the

predicted active sites for reduction are not reflected in the ultimate structure of Rapamycin.

It is noted that amino acids in the domain level are more conserved than in the modular level (Results). This could be because, modular level comparison includes the sequences of the domains as well as interdomains. Interdomain linker regions are required for the folding of the multifunctional polypeptide encoded by the modules and are typically less conserved. The variability of the interdomain region would therefore result in a lower modular homology than domain homology when compared to other PKS I genes.

Variability of domains within the PKS I cluster are also observed in all other PKS I clusters. Each module of the PKS I cluster catalyses a single step of the polyketide biosynthesis. As substrates for each step of the polyketide biosynthesis would be different in a PKS I system, enzymatic PKS I domains are likely to vary slightly in different PKS I systems. This could be one of the reasons for the variability of domains and the modules within the clusters.

The PKS I genes of the *Streptomyces* sp. 98- 62 studied in this work seem to have a higher similarity to the corresponding PKS I genes of the avermectin producer *Streptomyces avermitilis*. However, it should be noted that the available information of the PKS I cluster from the *Streptomyces* sp. 98- 62 is of insufficient detail to conclude on the gene organization in comparison to that of the avermectin PKS I gene cluster. Further sequencing work to determine the adjacent genes to span a region of at least one complete ORF would be required to put up a more comprehensive study for evolutionary origin of the PKS I genes.

It would also be premature at this juncture to draw firm conclusions on the exact nature of the chemical structure of the PKS I compound of the *Streptomyces* sp.

98- 62. In the case of rapamycin and FK506, the ORFs of the PKS I genes do not encode proteins that follow the order of KS, AT(A), DH, KR, ACP, KS, AT(P), DH, KR, ACP. This is also so for the ORF of the avermectin PKS I genes (Fig. 7). This suggests that there are some minor differences in the catalysis for the production of the PKS I compound of the *Streptomyces* sp. 98- 62 PKS I system from those of rapamycin, FK506 or avermectin. In view of this, it is expected that there are some minor difference in the structure of the PKS I compound of the *Streptomyces* sp. 98- 62 PKS I as compared to those of rapamycin, FK506 or avermectin. This evidence is in line evidence with the results of the TLC bioassay of the extracts of the *Streptomyces* sp. 98- 62 which predicted that the structure of the *Streptomyces* sp. 98- 62 PKS I antifungal compound was likely to be different from that of rapamycin and FK506.

In order to determine if the cloned PKS I genes of the *Streptomyces* sp. 98- 62 functioned in the antifungal compound synthesis, gene disruption of the genes were considered indispensable. To do this, a gene transfer system for the *Streptomyces* sp. 98- 62 to inactivate the genomic DNA had to be established. Intergeneric conjugation experiments with integrative pSOK201 gene disruption constructs from donor strain *E.coli* (pUB307) to the *Streptomyces* sp. 98- 62 were performed and demonstrated to be successful.

The integration of the disruption constructs into the homologous regions of the genome of the *Streptomyces* sp. 98- 62 by a single reciprocal recombination would be reflected by the presence of the vector backbone in the chromosome of the disruptant but not the wild type. Such a physical disruption by gene disruption would show functional disruption only if the homologous gene fragment of the disruption construct lacks the stop codon and/or start codon of that gene or operon.

Gene disruption with the disruption construct pDC170FL was used to establish the intergeneric conjugation experiment. This construct was also utilized to prove that the predicted stop codons and start codon in the 7.2 kb fragment of recombinant clone C170 was indeed functional and that the disrupting sequence of 7.2 kb constituted two different open reading frames as deduced. Southern analysis of the *BamHI* and Sp*hI* restricted DNA from the wildtype and disruptant using two probes, the 3.0 kb vector probe and the 7.2kb insert probe confirmed physical disruption. Secondary metabolites of disruptants grown analysed to determine if the antifungal compound biosynthesis by *Streptomyces* sp. 98- 62 was affected by the gene disruption of the PKS I genes. The result indicated that the disruptants 170D1 and 170 D2 were not functionally disrupted.

Although the results observed could be explained by the presence of the stop/start codons in the disruption construct, there was no direct evidence for the involvement of the cloned PKS I genes in the antifungal compound biosynthesis. Therefore, to determine if the PKS genes of the 7.2 kb PKS I fragment cloned from the *Streptomyces* sp. is indeed involved in the biosynthesis of the antifungal compound, internal fragments of DNA sequence of an individual ORF of the PKS I genes from clone C170 as well as E27 and C2 were decided to be used for further gene disruption analysis.

Upon gene disruption experiment with internal fragments of DNA sequences of an individual ORF, one representative disruptant each of the three different PKS I internal fragments, named 27D1, 2KBC170D1 and C2D1, were analysed by Southern blot and TLC- bioassay to determine physical and functional disruption. Southern hybridisation with vector probe result confirmed that the disruption constructs pDE27D1, pD2KBC170 and pDC2 had undergone insertion into the expected region

of the genomic DNA of the *Streptomyces* sp. 98- 62 to produce the disruptant transformants.

Secondary metabolites of disruptants 27D1, 2KBC170D1 and C2D1 failed to show any bioactivity. This result indicated that the disruptants 27D1, 2KBC170D1 and C2D1 were functionally disrupted.

Disruptant construct pDC2 contained the genes of module 2 and module 3. Disruption construct pDC2 is similar to disruption construct pD170FL in that both carry genes that span the parts of 2 modules. The construct pDC170FL contains the genes of modules 1 and 2. The key difference in the two constructs is that pDC170FL construct has 2 stop codons and a start codon in the intermodular region, but pDC2 construct does not contain any stop or start codons.

Comparison of the result from gene disruption experiment using constructs pDC170FL and pDC2 confirms the prediction that there are stop/start codons between module 1 and 2 but not between module 2 and 3. Thus module 1 is in a separate ORF from that of module 2 and 3. The result also shows that physical disruption of genes is not sufficient for functional disruption, and that it is important to use internal fragments of genes to observe functional disruption.

These results gave strong evidence that the cloned partial PKS I gene cluster of the *Streptomyces* sp. 98- 62 are responsible for the biosynthesis of the antifungal compound, and that the deduced partial ORF1 and ORF 2 of the cloned PKS I genes of *Streptomyces* sp. 98- 62 are indeed transcriptionally uncoupled.

It is intriguing to observe that functional disruption of the antifungal compound biosynthesis by disruption of the PKS I genes from the *Streptomyces* sp. 98- 62 had a pleiotropic effect on aerial mycelium formation. The exact cause as to this observation

is yet to be determined. It should be noted that this is the first ever report of PKS I genes having pleiotropic effect on differentiation of *Streptomyces.*

So far, only PKS II genes of the polyketide synthases have been implicated in differentiation of *Streptomyces,* albeit in spore colour formation. A PKS II gene designated *whi E* has been shown to be involved in the spore pigment formation of *S. coelicolor* (Keleman et al, 1998). Mutants of these PKS II gene *Whi E* were described as white (*Whi)* mutants. The white mutants of *Streptomyces coelicolor A(3)* produce an obvious aerial mycelim but not the normal spores. It is interesting to note that the PKS I gene disruptants of the *Streptomyces* sp. 98- 62 showed a " bald " phenotype rather than " white" phenotype. Mutants of *Streptomyces coelicolor A(3),* which lack an obvious aerial mycelium are called bald (*bld*). Most of the known *bld* mutants are regulatory proteins (Table 1 and 2). Such regulatory proteins are rather small and would be able to diffuse out of the cells to act as signals for differentiation process (Miyake *et al,* 1990). However in the case of the *Streptomyces* sp. 98- 62, the gene products of PKS I genes would be a large multifunctional polypeptide, that functions in the biocatalysis of a polyketide. Assuming that the polyketide rather than the polyketide synthase has a role in the differentiation process, the size of the polyketide would be large in comparison to the other known regulatory proteins of differentiation. Hence, it would be very interesting to determine how and why the PKS I genes are associated with differentiation of the *Streptomyces* sp. 98- 62.

CONCLUSION

In conclusion, the current work has identified PKS I genes in the novel antifungal compound producing *Streptomyces* sp. 98- 62, using PKS I specific probe eryKS II from *Saccharopolyspora erythraea*. The PKS I genes of the *Streptomyces* sp.

98- 62 were then cloned by homologous based approach whereby PCR primers from conserved sequence of PKS I genes were used to amplify the keto synthase-acyl transferase genes from the genomic DNA of the *Streptomyces* sp. 98- 62. Eventually the genomic copy of the keto synthase-acyl transferase genes of the *Streptomyces* sp. 98- 62 was isolated from a subgenomic library. Chromosomal walking aided in the isolation of clones that carried the adjacent fragments to that of the first clone isolated. The cloned DNA fragments of 11656 base pairs correspond to PKS I genes of streptomycetes, encompassing 3 modules. Module 1 was predicted to be a part of one open reading frame whilst module 2 and 3 of the cloned genes were predicted to be part of another open reading frame adjacent to the ORF encompassing module 1. The genes consist of repeated modules of ~5 kb and are characteristic of PKS I genes of other streptomycetes. The domains of the modules were also organised like those of the PKS I genes of other streptomycetes. All of the identified domains are predicted to be active based on sequence comparison with other known PKS I genes. The acyl transferase domain of module 1 was predicted to be specific for malonyl CoA specific substrates whilst the acyl transferase domain of module 2 was predicted to be specific for methylmalonyl CoA specific substrates. These results provided strong evidences that PKS I genes have been isolated from the novel antifungal producing *Streptomyces* sp. 98- 62. A gene transfer system for the *Streptomyces* sp. 98- 62 was then established and used to prove the function of the cloned genes in the biosynthesis of the antifungal compound. The gene disruption experiments established that indeed the cloned PKS I genes of the *Streptomyces* sp. 98- 62 were involved in the biosynthesis of the novel antifungal compound produced by the *Streptomyces* sp. 98- 62. The gene disruption experiments also confirmed the prediction that this work made a study of parts of two open reading frames in the total length of the cloned genes. In addition, the gene

disruption experiment highlighted the possible involvement of the cloned PKS I genes of the *Streptomyces* sp. 98- 62 in the morphological differentiation of the novel antifungal compound producing *Streptomyces* sp. 98- 62.

SIGNIFICANCE OF THIS PROJECT

According to Milind *et al.,* 2001, "it is becoming increasingly difficult to obtain novel compounds, and screening more often yields the same compounds again and again". However they suggest in their paper that the rate of decline in the rate of discovering new compounds is due to the decline in screening efforts rather than exhaustion of compounds. This opinion is resonated in an earlier paper by Hans Zahner and Hans- Peter Fiedler (1995). Several different approaches were described to identify new antibiotics in this paper. One of the many suggestions was to search for new antibiotics using new test methods, different microorganisms and varying culture conditions.

Given the difficulty of finding a new antibiotics, the finding of a novel antifungal compound from the *Streptomyces* sp. 98- 62 that is sufficiently different from the known polyketide antifungal compounds such as rapamycin and FK506 is of significance. Experimental data from the nucleotide sequences and TLC separation profile provide evidence that the antifungal complex polyketide compound produced by the *Streptomyces* sp. 98- 62 is different from the known antifungal polyketides such as rapamycin and FK506.

The isolation of the novel antifungal compound producing *Streptomyces* sp. 98- 62 goes to reiterate the point that with improved screening methods and using different microorganisms, new antibiotics could be identified. A rational screening approach to screen for antifungal compound producers form the pool of *Streptomyces* sp.s isolated

from various parts of the untapped Singapore soil has resulted in the successful identification of a *Streptomyces* sp. 98- 62 capable of producing a novel antifungal compound.

Cloning and characterisation of the polyketide synthase type I gene from the novel antifungal compound producing *Streptomyces* sp. 98- 62 as well as functional proof by gene disruption studies with the cloned genes have proved without doubt that the cloned PKS I genes are the biosynthetic genes that brought about the production of the novel antifungal compound by the *Streptomyces* sp. 98- 62.

Although it is too preliminary to suggest that the novelty of the antifungal compound would account for its usefulness as a potential pharmaceutical product, further study of this novel antifungal compound would be highly beneficial in understanding the natural evolution of polyketide synthase genes, use of the PKS I genes of the *Streptomyces* sp. in 98- 62 in combinatorial biosynthesis of novel hybrid polyketides.

This study has also for the first time led to the discovery of possible association of PKS I genes to the morphological differentiation of *Streptomyces* sp. Further work on this subject would be very useful in understanding the possible role of secondary metabolites in regulation of differentiation of the producing streptomycete.

FUTURE DIRECTIONS

Several directions can be taken for the further study of the novel antifungal compound produced by the *Streptomyces* sp. 98- 62. Cloning and sequencing the remaining PKS I genes of the *Streptomyces* sp. 98-62 are needed to characterise the complete gene cluster. Gene disruption studies of specific domains of the PKS I genes could be done to determine the function of the individual domain in the biosynthesis of

the antifungal compound. Determination of the chemical structure of the antifungal compound would be required to understand the correlation between the catalysis of the predicted PKS I genes and the compound structure. Structural analysis of the compound is then required to understand the role of chemical structure of the PKS I compound in relation to its mode of action as an antifungal compound and pleiotropic regulator of differentiation. Cloning the novel PKS I genes as that of the *Streptomyces* sp. 98- 62 would also increase the repertoire of available catalytic domains / modules that could be used to rationally engineer novel hybrid polyketide compounds.

FINAL REMARK

Search for a new antibiotic is a long road to success. Regardless of whether the destination of the search is as desired, the lessons to be learnt along the journey is as important as the destination itself, if not more. At this juncture, the results of this project is supportive of the potential of the novel antifungal compound produced by the *Streptomyces* sp. 98- 62 as promising in terms of the lessons to be learnt along the way as well as at the destination.

Amann, R., Ludwig, W. and Scheifer, K. H. 1995. Phylogenetic identfication and in situ detection of individual microbial cells without cultivation. Microbial. Rev. 59: 143 - 169.

Bodey, G. P. 1988. Topical and systemic antifungal agents. Med. Clin. North Am. 72 (3): 637- 659.

Bevitt, D. J., Cortes, J. Haydock, F. and Leadlay, P.F. 1992. 6-Deoxyerythronolide- B synthase 2 from *Saccharopolyspora erythraea*. Eur. J. Biochem. 204: 39-49

Bibb, M. J., Ward, M. J. and Hopwood, D. A. 1978. Transformation of plasmid DNA into *Streptomyces* at high frequency. Nature. 284: 526- 531.

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

Birch, A. J. and Elliot, P. 1953a. Studies in relation to biosynthesis II. The structure of Macrophone. Aust. J. Chem. 6: 369.

Birch, A. J. and Donovan, F. W. 1953b. Studies in relation to biosynthesis III. The structure of Eleutherinol. Aust. J. Chem. 6: 373.

Birch, A. J., Massey- Westropp, R. A, Rickards, R. W. and Smith, H. 1958. Studies in relation to biosynthesis. Part XIII. Griseofulvin. J.Chem. Soc. 360.

Bodey, G. P. and Anaissie, E.J. 1989. Chronic systemic candidiasis. Eur. J. Clin. Microbiol. Infect. Dis. 8 (10): 855- 857.

Bruton, C. J., Guthne, E. P. and Chater, K. F. 1991. Phage vectors that allow monitoring of transcription of secondary metabolism genes in *Streptomyces*. Bio Tech. 9: 652- 656.

Bull, A. T., Goodfellow, M. and Slater, J. H. 1992. Biodiversity as a main source of innovation in biotechnology. Ann. Rev. Microbiol. 46 : 219- 252.

Caffrey, P., Lunch, S., Flood, E., Finnan, S. and Oliynyk, M. 2001. Amphotericin biosynthesis in *Streptomyces nodosus*: deductions from analysis of polyketide synthase and late genes. Chem. and Biol. 8: 713-723.

Chater, K. F. 1992. Genetic regulation of secondary metabolic pathways in *Streptomyces*. CIBA Found. Symp. 171: 144- 162.

Chen, C. W., Lin, H. F., Kuo, C. L.,Tsai, H. L. and Tsai, J. F. Y. 1988. Cloning and expression of a DNA sequence conferring cephamycin C production. Bio Tech. 6: 1222- 1224.

Cohen, S.N., Chang, A.C.Y., and Hsu, L. 1972. Nonchromosomal antibiotic resistance in bacteria: genetic transformation of Escherichia coli by R-factor DNA. Proc. Natl. Acad. Sci. U.S.A. 69: 2110-2114.

Collie, J. N. 1907.Derivatives of the multiple ketene group. Proc. Chem. Soc. 91: 230.

Deng, Z., Hu, Z., Zhou, X., Zhou, Q., Hopwood, D. A. and Kieser, T. 1994. Repeated polyketide synthase modules involved in the biosynthesis of a heptaene macrolide by *Streptomyces* sp. FR008. Mol. Microbiol. 14(1): 163-172.

De Souza , N., Ganguli, B. and Reden, J. 1982. Strategies in the discovery of drugs from natural sources. 1982. Ann. Rep. Med. Chem. 17: 301-310.

Donadio, S., Staver, M. J., McAlpine, J. B., Swanson, S. J. and Katz, L. 1991. Modular organization of genes required for complex polyketide biosynthesis, Science. 252: 756- 679.

Dowding, J.E., Vivian, A., Chater, K.F. and Hopwood, D. A. 1973. Advances in *Streptomyces coelicolor* genetics. Bacteriol. Rev. 37: 371-405.

Drucker, D. B. 1974. Chemotaxonomic fatty acid fingerprints of some streptococci with subsequent statistical analysis. Canadian J. Microbiol. 20: 1723- 1728

Farshtchi, D. and McClung, N. M. 1970. Effects of substrate on fatty acid production in *Nocardia asteroids*. Canadian J. Microbiol. 16 : 213-217.

Flett, F. Mersinias, V. and Smith, C. P. 1997. High efficiency intergeneric conjugal transfer of plasmid DNA from *Escherichia coli* to methyl DNA restricting streptomycetes. FEMS Microbiol Lett. 155: 223- 229.

Goodfellow, M. 1986. Actinomycete systematics: Present state and future prospects. In: Szabo, G., Biro, S. and Goodfellow, M, (eds.). Biological, Biochemical and Biomedical Aspects of Actinomycetes. Proceeding of the sixth International Symposium on Actinomycetes Biology. Pp. 487- 496. Akademiai Kiaoto, Budapest.

Goodfellow, M. and Minnikin, D. E. 1985. Chemical methods in Bacterial Systematics. Academic Press, London.

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

Holmberg, K. and Meyer, R.D. 1986. Fungal infections in patients with AIDS and AIDS related complex. Scand . J. Infect. Dis. 18: 179- 192.

Hutchinson, R. C., Lomovskaya, N., Fonstein, L., Ruan, X., Stassi, D., and Katz, L. 1997. Gene disruption and replacement in the rapamycin producing *Streptomyces hygroscopicus* strain ATCC 29253. Microbiology. 143: 875-883.

Hutter, R. and Eckhardt, T. 1988. Genetic manipulation. In Goodfelow, M., Williams, S. T. and Mordarski, M. Actinomycetes in Biotechnology. pp: 89-184. Academic Press. London.

Jiang, C. L. and Xu, L. H. 1996. Diversity of aquatic actinomycetes in lakes of the middle plateau, Yunnan, China. App. Environ. Microbiol. 62: 249- 253.

Kakavas, S., Katz, L. and Stassi, D. 1997. Identification and characterization of the Niddamycin polyketide synthase genes from *Streptomyces caelestis.* J. Bacteriol. 179: 7515-7522.

Katz, L. and Donadio, S. 1993. Polyketide synthesis: prospects for hybrid antibiotics. Ann. Rev. Microbiol. 47: 875-912.

Kelemen, G. J., Brian, P., Flardh, K., Chamberlin, L ., Chater, K. F. and Butner, M. J. 1998. Developmental regulation of transcription of *WhiE*, a locus specifying the polyketide spore pigment in *Streptomyces coelicolor* A3(2). J. Bacteriol. 180: 2515-2521.

Kirby, R. and Hopwood, D.A. 1977. Genetic determination of methylenomycin synthesis by the SCP1 plasmid of *Streptomyces coelicolor A3(2).* J. Gen. Microbiol. 98: 239-252.

Khosla, C. 1997. Harnessing the biosynthetic potential of modular polyketide synthases. Chem Rev. 97: 2577-2590.

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

Leblond, P. F., Francou, F. X., Sinonet, J. M. and Devaris, B. 1990. Pulsed field gel electrophoresis analysis of the genome of the *Streptomyces ambofaciens* strains. FEMS Microbiol Lett. 72: 79-88.

Leblond, P., Fischer, G., Francou, F. X., Berger, F., Guerineau, M. and Decaris, B. 1996. The unstable region of *Streptomyces ambofaciens* includes 21kb- terminal inverted repeats flanking the extremities of the linear chromosomal DNA. Mol. Microbiol. 19: 261-271.

Lezhava, A., Mizukami, T., Kajitani, T., Kameoka, D., Redenbach, M., Shinkawa, H., Nimi, C. and Kinashi, H. 1995. Physical map of the linear chromosome of *Streptomyces griseus*. J. Bacteriol. 177: 6492-6498.

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

MacNeil, D. J. 1988. Characterization of a unique methyl specific restriction system in *Streptomyces avermitilis*. J. Bacteriol. 170: 5607-5612.

MacNeil, D. J., Gewain, K. M., Ruby, C. L., Dezeny, G., Gibbons, P. H. and MacNeil, T. 1992. Analysis of *Streptomyces avermitilis* genes required for avermectin biosynthesis utilising a novel integration vector. Gene. 111: 61-68.

Malpartida, F. and Hopwood, D. A. 1984. Molecular cloning of the whole biosynthetic pathway of a *Streptomyces* antibiotic and its expression in heterologous host. Nature. 309: 462-464.

Malpartida, F., Hallam, S. E., Kieser, H. M., Motamedi, H., Hutchinson, C. R., Butler, M. J., Sugden, D. A., Warren, M., Bailey, C. R. McKillop, C., Humphreys, G. O. and Hopwood, D. A. 1987. Sequence homology between *Streptomyces* genes coding for synthesis of different polyketides and its use to clone antibiotic biosyntehsis genes. Nature. 325: 818-821.

Mazodier, P., Petter, R. and Thompson, C. 1989. Intergeneric conjugation between *Escherichia coli* and *Streptomyces* species. J. Bacteriol. 171: 3583-3585.

Mazodier, P., Servant, P. and Thompson, C. 1993. Use of new *Escherichia coli Streptomyces* conjugative vectors to probe the functions of the two gro*EL*- like genes of *Streptomyces albus* G by gene disruption. Gene. 134: 25-32.

Medlin, L., Elwood, H.J., Stickel, S. and Sogin, M.L. 1988. The Characterization of enzymatically amplified eukaryotic 16S-like rRNA-coding region. Gene. 71: 491-499.

Medoff, G., Brajtburg, J., Kobayashi, G. S. and Bolard, J. 1983. Antifungal agents useful in therapy of systemic fungal infections. Ann. Rev. Pharmacol. Toxicol. 23: 303-330.

Mehling, A., Wehmier, U. F. and Pipseberg, W. 1995. Nucleotide sequences of *Streptomyces* 16S ribosomal DNA: Towards a specific identification system for streptomycetes using PCR. Microbiology. 141: 2139-2147.

Milind, G. W., Rashmi, T., Maithili, M. J. and Bhalachandra, D. B. 2001. How many antibiotics are produced by the genus *Streptomyces*. Arch. Microbiol. 176: 386-390.

Miyake, K., Kuzuyama, T., Horinouchi, S. and Beppu, T. 1990. The A- factor-binding protein of *Streptomyces griseus* negatively controls streptomycin production and sporulation. J. Bacteriol. 172: 3003-3008.

Molnar, I., Aparico, J. F., Haydock, S. F., Khaw, L. E., Schwecke, T., Konig, A., Stauton, J., and Leadlay, P. F. 1996. Organization of the biosynthetic gene cluster for rapamycin in *Streptomyces hygroscopicus*: Analysis of genes flanking the polyketide synthase. Gene 169: 1-7.

Motamedi, H., Shaffie, A., Cai, S. J., Streicher, S. L., Arison, B. H. and Miller, R.R. 1996. Characterisation of methyltransferase and hydroxylase genes involved in the biosynthesis of the immunosuppressants FK 506 and FK 520. J. Bacteriol. 178: 5243-5248.

Motamedi, H., Cai, S. J., Shafiee, A. and Elliston, K. O. 1997. Structural organization of the multifunctional polyketide synthsae involved in the biosynthesis of the macrolide immunosuppressant FK506. Eur. J. Biochem 244: 74-80.

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

Muth, G., Wohllenben, W. and Puhler, A. 1988. The minimal replicon of the *Streptomyces ghanaensis* plasmid pSG5 identified by subcloning and Tn5 mutagenesis. Mol. Gen. Genet.211 : 424-429.

Myers, N. 1988. Threatned biotas: “hotspots” in tropical forests. Environmentalist 8: 1-120.

Nakamura, Y., Gojobiri, T. and Ikemura, T. 1980. Codon usage tabulated from the international DNA sequence databases. Nucl. Acids Res. 26: 334.

O’Donnell, A. G. 1985. Numerical analysis of chemotaxanomic data. In: Goodfellow, M., Jones, D. and Priest, F. G. (eds). Computer – assisted Bacterial Systematics. pp: 403-414 . Academic Press. London.

Rawlings. J.B. 2001. Type I polyketide biosynthesis in bacteria. Nat. Prod. Rep. 18: 190-227.

Richmond, M. H., Bennet, P. M. and Grinsted, J. 1976. Transposition of Tn*A* does not generate deletions. Mol. Gen. Genetics. 154: 205-211.

Rittenberg, D. and Bloch, K. 1944. The utilisation of AcOH for fatty acid synthesis. J. Biochem, 154: 311.

Rittenberg, D. and Bloch, K. 1945. Utilisation of acetic acid for the synthesis of fatty acid synthesis. J. Biol. Chem. 160: 417.

Ruan, X., Stassi, D., Lax, S. A. and Katz, L. 1997. A second type I PKS gene cluster isolated from *Streptomyces hygroscopicus* ATCC 29253, a rapamycin producing strain. Gene 203: 1-9.

Saddler, G. S., O'Donnell, A. G., Goodfellow, M. and Minnikin, D. E. 1987. SIMCA pattern recognition in the analysis of streptomycete fatty acids. J. Gen. Microbiol. 133: 1137-1147.

Sampsons,S. M., Belgaga, R, Blankensnipo, D. T., Chapman, J. L. , Perry, D. , Skatrud, P. L., VanFrank, R. M., Abraham, E. D., Baldwin, J. E., Queener, S. W. and Ingolia, T. D. 1985. Isolation, sequence determination and expression in *Escherichia coli* of the isopenicillin N synthetase gene from the *Cephalosporium acremonium.* Nature. 318: 191-194.

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

Schwecke, T., Aparicio, J. F., Molnar, I., Konig. A., Caffrey, P., Cortes, J., Lester, J. B., Bohm, G. A., Stauton, J. and Ledalay, P. F. 1995. The biosynthetic gene cluster for the polyketide immunosuppressant rapamycin. Proc. Natl. Acad. Sci. USA 92: 7839-7843.

Stackenbrandt, E. 1985. Phylogeny and phylogenetic classification of prokaryotes. In: Schleifer, K. H. and Stackenbrandt, E. (eds.). Evolution of prokaryotes. pp: 309-314. Academic Press, London.

Stackenbrandt, E. 1986. The significance of " wall types" in phylogenetically based taxonomic studies on actinomycetes. In: Szabo, G., Biro, S. and Goodfellow, M, (eds). Biological, Biochemical and Biomedical Aspects of Actinomycetes. Proceeding of the sixth International Symposium on Actinomycetes Biology. Pp: 497-506. Akademiai Kiaoto, Budapest.

Stackenbrandt, E., Rainey, F. A. and Ward – Rainey, N. L. 1997. Proposal for a new hierarchic classification system, Actinobateria classis nov. Int. J. Syst. Bacteriol. 47: 479-491.

Strohl, W. R. 1992. Compilation and analysis of DNA sequences associated with apparent streptomycete promoters. Nuc. Acids Res. 20 (5): 961-974.

Strohl, W. R., Bartel, P. L., Li, Y., Connors, N. C. and Woodman, R. H. 1991. Expression of polyketide biosynthesis and regulatory genes in heterologous streptomycetes. J. Indus. Microbiol. 7: 163-174.

Thompson, C. J., Ward, J. M. and Hopwood, D. A. 1980. DNA cloning in *Streptomyces:* resistance genes from antibiotic-producing species. Nature.286: 525-527.

Tuan, J. S., Weber, J. M., Staver, M. J., Leung, J. O., Donadio, S. and Katz, L. 1990. Cloning of genes involved in erythromycin biosynthesis from *Saccharopolyspora erythraea* using a novel actinomycete- *Escherichia coli* cosmid. Gene. 90: 21-29.

Vara, J., Lewowska-Skarbek, M., Wang, Y. G., Donadio, S. and Hutchinson, C. R. 1989. Cloning of genes governing the deoxysugar portion of the erythromycin biosynthesis pathway in *Saccharoplyspora erythraea*. J Bacteriol. 171: 5872-5881.

Waksman, S. A. 1950. The Actinomyctes. Their Nature, Occurrence, Activities and Importance. Chronica Botanica Company, Waltham, Mass., USA.

Wang, Y., Ruan, J., Zhang,Z. S. and Wang, Y. M. 1999. High actinomycete diversity in the tropical rainforests of Singapore. In: Asian network on microbial research symposium microbial diversity: The Asian Scene. National University of Singapore.

Wright, F. and Bibb, B. 1992. Codon Usage in the G+C rich *Streptomyces* genome. Gene 113: 55-65.

Wohlloben, W. and Muth, G. 1993. *Streptomyces* plasmid vectors. In: Plasmids: A Practical Approach (Hardy, K. G. Ed.). pp: 147-175. IRL Press, Oxford.

Xu, L. H., Li, Q. R. and Jiang, C. L. 1996. Diversity of soil actinomycetes in Yunnan, China. App. Environ. Microbiol. 62: 244-248.

Yap, W. H. and Wang, Y. 1999. Molecular cloning and comparative sequence analysis of rRNA operons in *Streptomyces nodosus* ATCC 14899. Gene 232: 77-85.

Yokota, A. 1997. In: Atlas of Actinomycetes. A phylogenetic relationship of actinomycetes. pp 194-198. The Society of Actinomycetes, Japan.

Yue, S., Duncan, J. S., Yamamoto, Y. and Hutchinson, C. R. 1987. Chem Soc. 1253-1255.

Zotchev, S., Haugan, K., Sekurova, O., Sletta, H., Ellingsen, T. E. and Valla, S. 2000. Identification of a gene cluster for antibacterial polyketide derived antibiotic biosynthesis in the nystatin producer *Streptomyces nouresi* ATCC 11455. Microbiology, 146: 611-619.