Studies on the Biosynthesis of Antibiotics Mupirocin and Thiomarinol

A thesis submitted to the University of Birmingham for the degree of
DOCTOR OF PHILOSOPHY

By

Mukul Yadav

Institute of Microbiology and Infection
School of Biosciences
The University of Birmingham

March 2016

ABSTRACT

Many biosynthetic steps in the mupirocin (pseudomonic acids) biosynthetic pathway of *Pseudomonas fluorescens* NCIMB 10586 have already been deduced. Putative functions of most of the genes of the *mup* cluster have been assigned, although, exact sequence of steps in the pathway and their timings are not yet known. Thiomarinols are another group of anti-bacterials produced by *Pseudoalteromonas sp.* SANK73390. Very little is known about the biosynthesis of thiomarinols that share striking structural similarity with pseudomonic acids in their polyketide and fatty acid moieties. This similarity is reflected at genetic level as significant similarity in amino acid identity between the products of at least 27 open reading frames in the biosynthetic clusters.

This project aimed to learn more of biosynthetic steps in the biosynthesis of mupirocin and thiomarinol antibiotics by testing for functional cross-complementation between pair of genes or a group of genes whose products show significant homology. Surprisingly only two genes *tmlJ* and *tmlS* out of nine studied showed complementation in the *mup* system. Findings suggested protein-protein interactions limited interchangeability of equivalent functions between two biosynthetic systems. It was shown by expressing related genes as groups for complementation in the *mup* system that it was possible to confirm specificities of such interactions.

Acknowledgements

I would like to thank my lead supervisor, Prof. Chris M. Thomas, for accepting me for doctoral supervision and for providing me support, encouragement, criticism and guidance throughout my stay. I am particularly thankful to him for waiving my bench fee. Thanks to my co-supervisor, Joe Hothersall, for her support and guidance/advice whenever it was needed.

I am thankful to, Tony, for his patient listening to my problems and giving the advice accordingly whenever I approached him. I am greatly thankful to, Elton, who was always there to help me in all my laboratory work and whenever I got stuck in it. I am thankful to all the staff of school of biosciences who worked in the background to provide me with continuous supply of services, chemicals and particularly to the staff that regularly supplied clean glasswares and sterilized media that helped me in hassle-free working.

I am thankful to the members of S101 laboratory past and present, members of past for providing directions to my work and present ones for their company. Some of them have been of great help when I really needed it! I am particularly thankful to Ahmed M. Omer-bali for his timely help. It was pleasing to meet Daisuke during his visit to our laboratory whose work laid foundations to this work

Thanks are also due to collaborators for this project Dr Russel Cox group from Bristol University for their inputs and particularly to Dr Zhongshu Song of this group for carrying out LCMS studies and related data analysis. Thanks to my sponsor government of M.P. (India) for funding my studies and providing partial maintenance and to Indian Consulate, Birmingham, UK for helping me.

This work was not possible without the constant encouragement and support from my family particularly from my wife who let me go abroad and have such a long leave for my studies.

Table of Contents

LIST OF FIGURES

Chapter 1

Chapter 2

Chapter 3

Chapter 4

Chapter 5

LIST OF TABLES

Chapter 1

Chapter 2

LIST OF ABBREVIATIONS

3-HP	3-hydroxypropionate
6-dEB	6-deoxyerythronolide B
9-HN	9-hydroxynonanoic acid
A, C, G, T	nucleotides: adenine, cytosine, guanine, thymine
aa	amino acid
ACP	acyl carrier protein
Amp	ampicillin
AMP	adenosine monophosphate
ApR	ampicillin resistant
AT	acyl transferase
ATP	adenosine triphosphate
AR	Adapter domain
C	condensation
CHS	Chalcone synthase
CLF	chain length factor
CoA	coenzyme A
Cryo EM	cryo electron microscopy
CsCl	Caesium chloride
CYCS	cyclases
Cys	cysteine
DEBS	deoxyerythronolide B synthase
DH	dehydratase
DMSO	dimethyl sulfoxide
DNA	deoxyribonucleic acid
dNTP	deoxynucleotide triphosphate
E	epimerisation
EDTA	ethylene-diamine-tetra acetic acid
ELSD	evaporative light scattering detector
ER	enoyl reductase
FASs	fatty acid synthases
GTs	glycosyltransferases
GC	nucleotides: cytosine and guanine
HCS	HMG-CoA synthase
His	histidine
HMG	3-hydroxy-3-methyl glutaric acid
HPLC	high performance liquid chromatography
HR h	homologous recombination
9-HN	9-hydroxynonanoic acid
8-HO	8-hydroxyoctanoic acid
Ile	isoleucine
IleRS	isoleucyl tRNA synthase
IPTG	isopropyl- β-D-thiogalactopyranoside
Kan	kanamycin
KmR	kanamycin resistant
KR	β-ketoacyl reductase
KS	β-ketoacyl synthase

L agar	Luria-Bertani agar
L broth	Luria-Bertani broth
LC	liquid chromatography
LM	loading module
Macp	Mupirocin acyl carrier protein
macp	Gene for mupirocin acyl carrier protein
M broth	Marine broth
M agar	Marine agar
MA	monic acid
MAT/MCAT	malonyl-CoA: ACP transacylase
MeOH	Methanol
MCS	Multiple cloning site
MIC	Minimum inhibitory concentration
Mmp	mupirocin multifunctional protein
MmpEOR	mupirocin multifunctional protein oxidoreductase domain
MPM	mupirocin production media
mRNA	messenger RNA
MRSA	methicillin resistance *Staphylococcus aureus*
MS	mass spectrometry
MT	methyl transferase
NMR	nuclear magnetic resonance
NRPS	non-ribosomal peptide synthetase
OD	optical density
ORF	open reading frame
oriT	origin of conjugal DNA transfer
oriV	origin of vegetative replication
PA	pseudomonic acid
PCP	peptidyl carrier protein
PCR	polymerase chain reaction
Phe	phenylalanine
PKS	polyketide synthases
polA1	DNA polymerase I
PPtase	phosphopantetheinyl transferase
QS	quorum sensing
ΔQD	quadruple mutant
qsc	quorum sensing controlled
rbs	ribosomal binding site
RNase	ribonuclease
SAM	*S*-adenosylmethionine
SAXS	Small angle X-ray scattering
SDS	sodium dodecyl sulphate
SDW	sterile distilled water
SSM	secondary stage medium
T	thiolation
TcR	tetracycline resistant
TE	thioesterase
Tet	tetracycline
THP	tetrahydropyran ring
TM	thiomarinol

TNE	*tris*-sodium-EDTA
trfA	*trans*-acting replication function
Tris	*tris* (hydroxymethyl) amino methane
tsp	transcriptional start point
TTC	2, 3, 5-triphenyltetrazolium chloride
UV	ultraviolet
WT	wild type
X-gal	5-bromo-4-chloro-3-indolyl- β-D-galactopyranoside

UNITS

bp	Base pairs
deg	Degrees
g	Grams
h	Hours
kb	Kilobases
l	Litres
LSU	Light scattering units
M	Molar
min	Minutes
ºC	Degrees Celsius
rpm	Revolutions per minute
s	Seconds
S	Svedberg
V	Volts
v/v	Volume used per final
w/v	Weight used per final volume

CHAPTER 1

1.1 Introduction

The discovery of antibiotics has been one of the most important historical milestones for improving the health conditions of humans and animals since their discovery decades ago. Any compound that either arrests the growth of microorganisms or kills them is called an antibiotic. Antibiotics are not only synthesised by bacteria, fungi and some plants but also reported from organisms as diverse as insects, dinoflagellates, molluscs etc. Though the first discovered antibiotic-penicillin was a peptide antibiotic, naturally produced antibiotics have diverse chemical nature with majority of these being polyketides that are produced as a result of secondary metabolism by producing organisms. Bacteria and fungi make antibiotics to protect themselves from surrounding organisms with which they face competition for survival (for space and/or nutrition etc.). It is hypothesised that these chemical compounds are exported out of the cell to kill competitors or to serve other functions like signalling.

Today hundreds of antibiotics are used therapeutically, the majority of which are produced by bacteria belonging to actinomycetes or *Streptomyces* (Hopwood, 2007, Mahajan and Balachandran, 2012, de Lima Procópio et al., 2012). But because microorganisms have the inherent ability to develop resistance over a period against all the microbial agents whether natural or synthetic, today there is the growing problem of bacterial resistance shown by pathogenic microorganisms to many of the known antibiotics (Levy and Marshall, 2004, Ventola, 2015). Bacteria employ various molecular mechanisms like random mutations or acquisition of genes from other bacteria by horizontal transfer to evolve resistance to specific antibiotics (Kuroda et al., 2001). The problem of bacterial resistance has also increased due to unregulated use of antibiotics in certain parts of the world (Ventola, 2015). Some of the pathogenic bacteria over the period of time have developed resistance to all known antibiotics. Hence, they are classified as “superbugs” (Morris et al., 1998, Davies and Davies,

2010). One such bacterium is Methicillin-resistant *Staphylococcus aureus* (MRSA) which causes surgical wound infection, pneumonia and systemic infections. This bacterium has become resistant to all β-lactam antibiotics as well as to most of the non β-lactam antibiotics and therefore poses a serious health threat (Hiramatsu et al., 2001). Mupirocin, a polyketide antibiotic which is produced by bacterium *Pseudomonas fluorescens*, until recently has been very effective against MRSA until the development of resistance to it was reported by Rahman et al. in 1987. Various *S. aureus* strains have developed resistance to mupirocin. While low-level resistance can be developed as a result of point mutations in its gene, *ileS,* that encodes enzyme isoleucine t-RNA synthetase (IleRS), high level resistance is caused as a result of it acquiring *mup*A gene that encodes a different eukaryotic type IleRS protein which is not inhibited by mupirocin (Hodgson et al., 1994). Isoleucine t-RNA synthetase (IleRS) is the target of mupirocin for its antibacterial action (Antonio et al., 2002).

As a result of the increasing problem of drug resistance among pathogens, in general, and in particular to this antibiotic due to their indiscriminate and unregulated use, demand for novel antibiotic molecules with novel and improved/useful activities is ever increasing (Walsh, 2003, Ventola, 2015). To tackle this urgent need for molecules with novel activity, new molecules are being explored and tested for novel activities that are produced by diverse organisms from different habitats ranging from extreme climates; deserts to deep sea. Other approaches involve improving the activity of existing molecules by modifying them chemically or biosynthetically. Attempts are being made to develop novel hybrid molecules by integrating the components from different biosynthetic pathways by applying the knowledge of genetic engineering and approaches of the combinatorial biosynthesis.

It has been shown that biosynthesis of polyketide antibiotic molecules which is catalysed by enzymes called polyketide synthases (PKS), can often be manipulated at the

genetic level of the producer organism to make them produce their altered derivatives which may have novel and useful/improved activity towards the pathogens.

Polyketides are present in nature in diverse forms like polyphenols, macrolides, enediynes and polyethers. They have distinct functions which include a wealth of clinically important activities such as antibiotics (erythromycin, tetracycline, rifampicin, picromycin, kirramycin etc.), antifungal, anticancer (epothilone B), antiparasitic (avermectin) and immunosuppressive properties (rapamycin), anti-cholesterol (lovastatin) and have great pharmacological potential. All polyketides are known to have been derived from the molecules that have the characteristic repetitive "-CH_2-CO-" moiety.

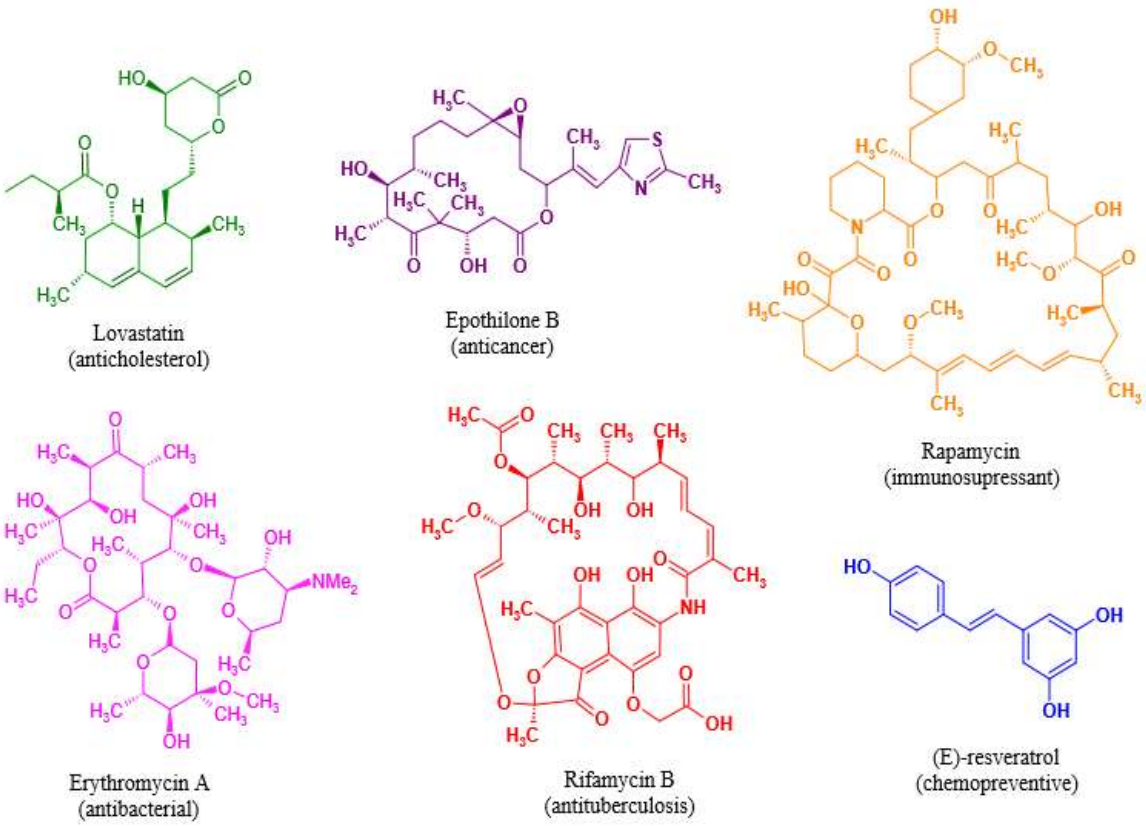


Figure 1.1 Structures of some polyketides of clinical importance.

Thiomarinols is another group of compounds produced by marine bacterium designated as *Alteromonas rava sp. nov*. SANK 73390 that has strong antimicrobial activity (Shiozawa et al., 1993) and therefore has potential to be developed as a strong antibiotic. It has been shown that these molecules are hybrid of two types of existing antibiotics; pseudomonic acids and pyrrothines (Shiozawa et al., 1993). Holomycins belong to pyrrothine class of antibiotics. Thus, thiomarinols are hybrid molecules that are coded by a hybrid of polyketide synthase and non-ribosomal peptide synthase gene cluster known as the *tml* cluster (Fukuda et al., 2011).

1.2 Mechanism of polyketide and fatty acid biosynthesis

Polyketides like fatty acid biosynthesis are produced as a result of simple sequential addition/condensation of 2-Carbon keto units, in repeated/recurring decarboxylative condensation reactions chemically known as Claisen condensation that are catalysed by gigantic multifunctional enzymes called as polyketide synthetases (PKS). PKSs also have similarity with the enzymes of fatty acid synthesis- fatty acid synthases (FAS). The PKS and FAS have also been shown to be evolutionary related (Smith and Tsai, 2007). Biosynthesis of both, polyketides and fatty acids, is initiated by an initial reaction in which a starter carboxylic acid unit condenses with a dicarboxylic acid.

In both, PKS and FAS, the priming/initiator (activated) carboxylic acid moiety is attached to the cysteine residue at the active site of the ketosynthase (KS) domain of synthase, which is condensed with the chain extender substrate. During the biosynthesis of polyketides for each round of condensation the acyl transferase (AT) region selects and loads the extender unit on the serine residue present in its active site which is then followed by transfer and attachment of the selected unit onto the thiol group of phosphopantetheine arm of Acyl Carrier Protein (ACP). Each condensation results in an extended chain with 2-carbons

which remains bound to ACP as β-keto ester which is subjected to β carbon processing/reduction before starting another round of elongation. In both, the reductive steps catalysed by ER/KR domains involve NADPH. This way when the chain of required length is reached by such successive addition of two carbon atoms followed by reduction, it is finally released from phosphopantetheine by the action of the thioesterase (TE) domain (Hopwood and Sherman, 1990). In both the synthases, all reaction intermediates remain covalently attached to them.

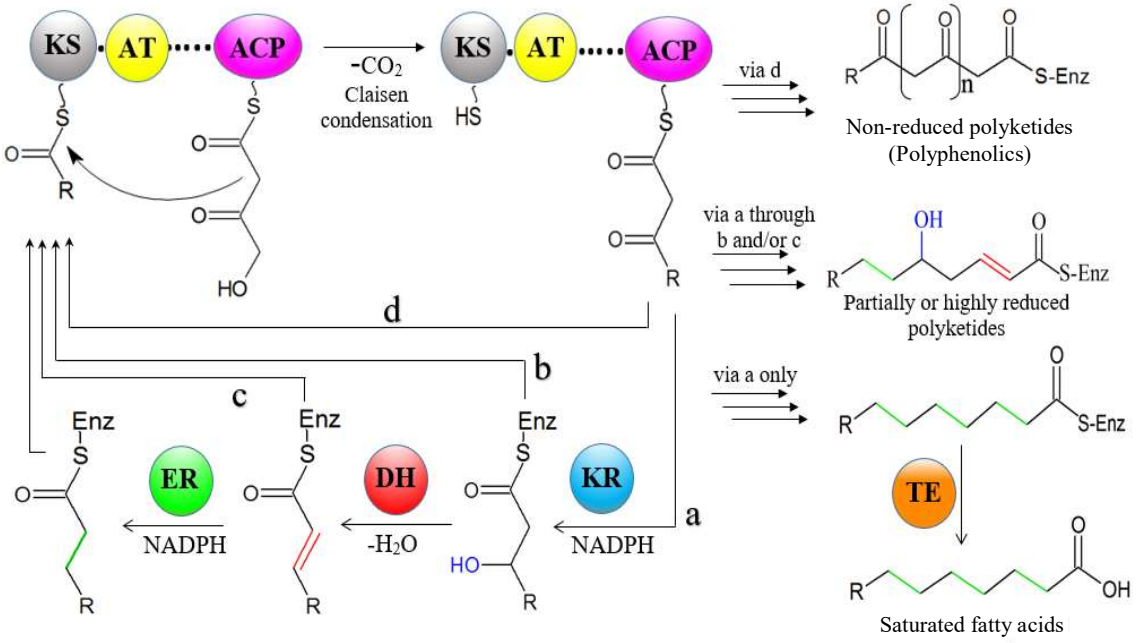


Figure 1.2 Basic mechanisms involved in fatty acid and polyketide biosynthesis.KS, ketosynthase; AT, acyl transferase; DH, dehydratase; ER, enoyl reductase; KR, ketoreductase; ACP, acyl carrier protein (Adapted from Hertweck, 2009).

The polyketide biosynthesis differs from fatty acid biosynthesis on two counts. Firstly, in the set of reactions catalysing reduction of β-carbons which is optional in PKS biosynthesis compared to in the fatty acid biosynthesis. This is on account of the variable presence (or complete absence) of different reducing domains in PKSs compared to in FASs

all of which have complete set of the reducing domains and therefore always catalyse a uniform set of reductions. Secondly, PKSs and FASs differ in their choice of carbon-substrates used for biosynthesis. PKSs can use a variety of carbon substrates compared to FASs which utilize only acetyl Co-A and malonyl Co-A as starter and extender units, respectively (Figure 1.4). PKSs can use primer units with 2, 3 or 4 carbon atoms and can accept β-keto or β-hydroxy or enoyl or saturated substrates (Figure 1.4). It is on account of these two factors polyketides can have many different types of structures compared to that of fatty acids. Further diversity is added to the polyketides by the action of another set of enzymes called tailoring enzymes which may cause further oxidative-reduction, the addition of moieties or various functional groups or cyclization. Diversity is also present in PKS on account of the stereochemistry of hydroxy and alkyl side groups and in the way various carbon atoms of the newly formed molecule cyclise. Normally in PKSs, different dedicated AT modules are used for loading of primer and extender units unlike in FASs. This results in the flexibility of utilization of different substrates for each condensation cycle in PKS biosynthesis that imparts additional diversity to them (Figure 1.3). Furthermore, the difference between PKSs and FASs is in the manner by which they extend the chain. FASs function in an iterative manner by which the same ACP is used in all the chain condensation reactions for transferring of substrates. While in modular PKSs for such role each module has got its own ACP, which may be present on the same or another polypeptide. In PKS, an ACP on one hand coordinates with KS region of its module for chain elongation and on the other hand also coordinates with the KS domain of next module 'downstream' to facilitate chain transfer of processed intermediate (condensed and β-carbon processed). Same role in the downstream module is then performed by the ACP of that module.

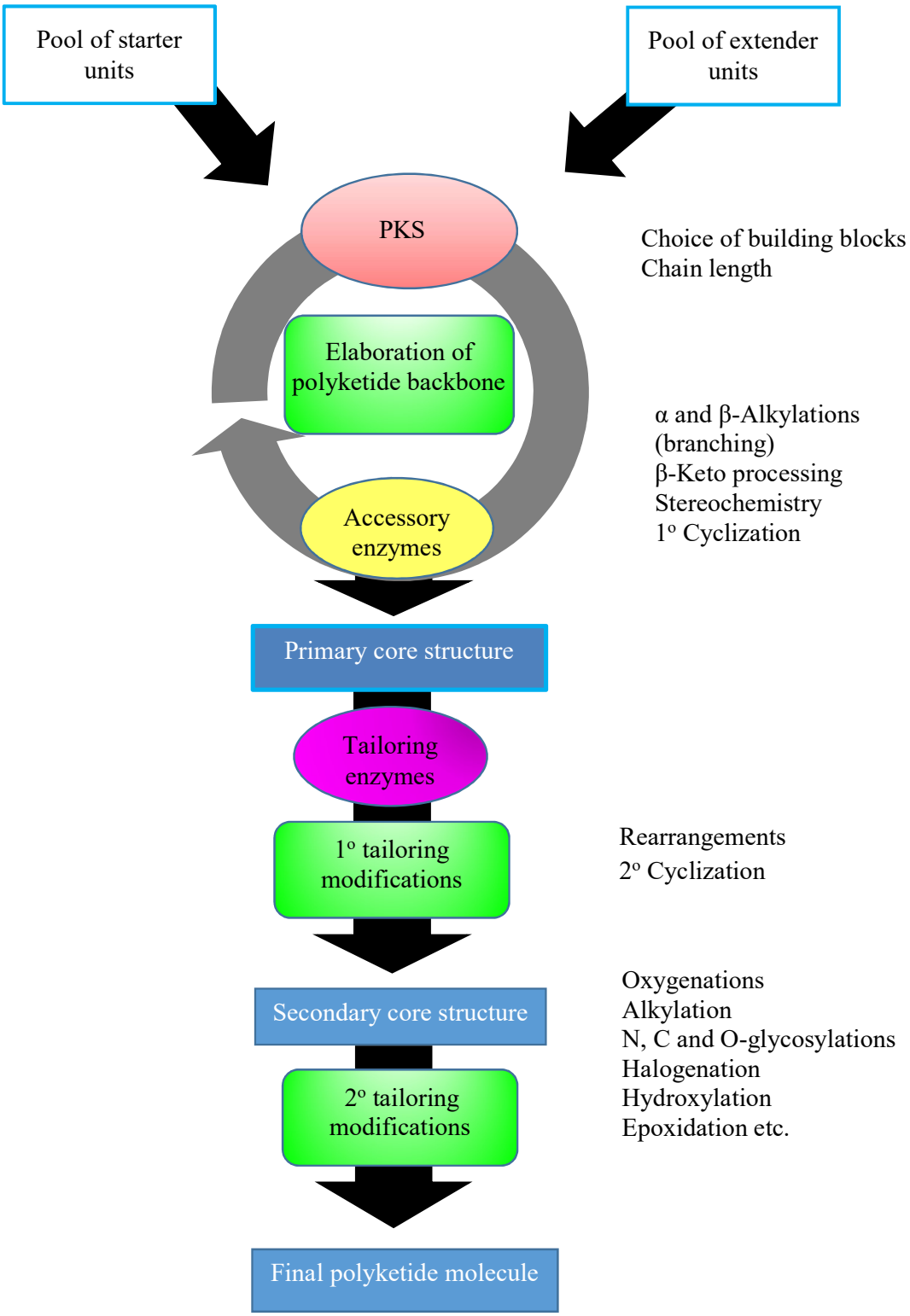


Figure 1.3 Schematic representation of levels in the biosynthetic pathway of polyketides where diversity can be introduced(Adapted from Hertweck, 2009).

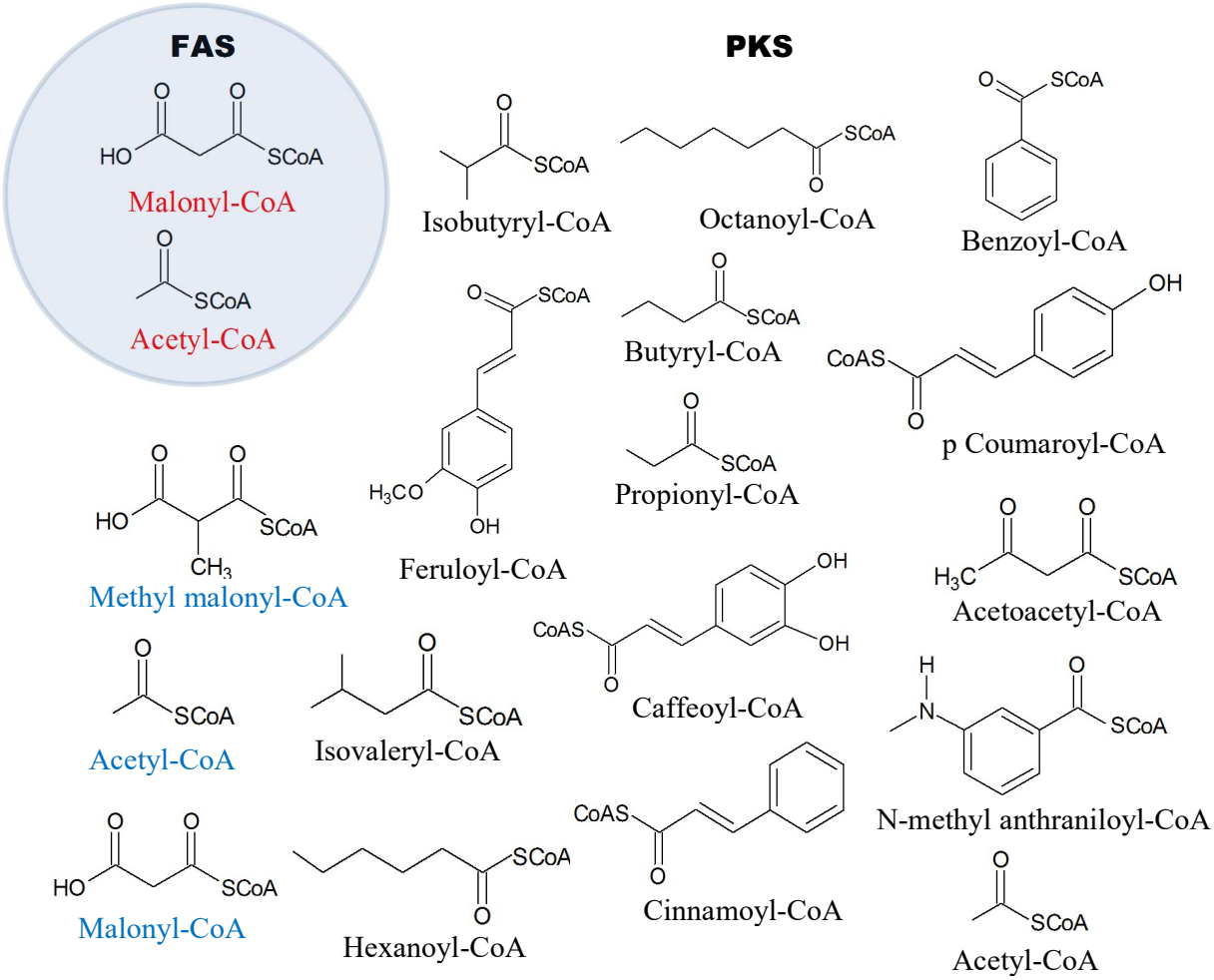


Figure 1.4 Various starter and extender units utilized by FASs and PKSs. Those used by FASs are labelled in red. A range of starter units used by PKSs are labelled in black while extender units used by them are labelled in blue.

1.3 Polyketide synthases (PKS)

These are multifunctional enzymes that are responsible for the biosynthesis of polyketide molecules and are of various types. A minimal type I PKS has three essential catalytic regions that are acyltransferase (AT), ketosynthase (KS) and acyl carrier protein (ACP) domain which are responsible for the extension of carbon chain; while terminal thioesterase domain (TE) catalyses the final release of polyketide molecule. In fact, there are various domains/mechanisms for the release of extended polyketide molecule. Apart from

these minimal catalytic regions a PKS can have any or all of many other catalytic regions like ketoreductase (KR), dehydratase (DH), enoyl reductase (ER) that catalyse the reduction of the β-keto groups in the polyketide, and other domains like methyltransferase that add a methyl group. As per the level of organization of protein structure (architectural organization) PKSs are of three major types viz type I PKS, type II PKS and type III PKS that are further sub-divided on the basis of their mode of action during biosynthesis which may be linear or iterative. These are described below.

1.3.1 Type I (modular) polyketide synthases

Type I PKS (modular) synthases are large (megadalton) multienzyme protein complexes made of several large polypeptides which have a set of separate catalytic active sites for each cycle of carbon chain elongation and modification in the polyketide biosynthetic pathway. Each active site is known as a "domain" and a set of domains is called a "module" for which reason these are also known as modular PKS. All the necessary domains that are KS, ACP and AT required to complete one elongation cycle along with any or all of the optional β-keto processing domains are all present in one module. This way one module is responsible for completing one elongation cycle. Several modules are present in these synthases that are covalently attached to each other making them as one integrated structure. The number of modules present in these synthases directly correlates with the number of carbon atoms in the polyketide products they synthesize. Similarly, the number and type of β-keto processing units that are present also directly relate to the nature and degree of reduction of the synthesized polyketide. As a result, there is direct correlation between composition of PKS architecture (one-to-one correspondence of the domain/module present in it) and the structure of product that it encodes or vice versa implying the principle of colinearity in PKS biosynthesis (Staunton and Weissman, 2001).

Various degrees of colinearity exist between PKS architecture and the product that it encodes depending upon whether the PKS is an iterative type or non-iterative and whether all the domains/modules present in them are functional. A non-iterative polyketide synthase uses its each module for chain elongation/processing during a biosynthetic cycle only once. Therefore, colinearity between PKS architecture and its product is highest in type I non-iterative (modular) PKS, which on account of their organization functions like an assembly line. As a result of this, usually the final number of ketide units synthesized by Type I (modular) PKS equals the number of carbon chain extension modules present in the PKS. In fact, in the biosynthesized product even level, nature, and position of reductions are also a clear-cut reflection of the type, nature, number and position of the various reducing domains present in the PKS structure (Keatinge-Clay, 2012). This way there is colinearity between PKS structure and the structure of the molecule that it encodes. Type I PKS may also contain a loading module to initiate biosynthesis. The final product is cleaved by thioesterase (TE) domain.

The type I PKS are further sub-divided into iterative and non-iterative based on the use of modules/domains for biosynthesis whether they are used for repeated chain extensions or are used for just one extension cycle. An example of a non-iterative type I PKS is 6-deoxyerythronolide B synthase that synthesises the core of erythromycin antibiotic. Other examples are avermectin and rapamycin. Modular type I (non-iterative) PKSs are analogous to vertebrate FAS and mostly present in bacteria while modular iterative PKSs are mainly found in fungi.

1.3.1.1 Erythromycin- a non-iterative type I PKS

The 6-deoxy erythronolide (DEBS) polyketide core of erythromycin antibiotic is the most widely studied type I modular PKS system which is synthesized by *Sacchropolyspora*

erythrae. Erythromycin polyketide gene cluster (*eryPKS*) consists of three genes *eryAI*, *eryAII* and *eryAIII* (Donadio and Katz, 1992) which encode three multifunctional proteins named deoxyerythronolide B synthase 1, 2 and 3 designated as DEBS1, DEBS2 and DEBS3 respectively (Figure 1.5). All non-iterative type I PKS of these three multifunctional proteins have two modules each of which in turn consists of three domains that are required to complete one cycle of chain elongation; ketosynthase, acyltransferase, and acyl carrier protein (Swan et al., 1994). The polyketide core of erythromycin-A molecule is thus biosynthesized by six successive condensation/chain elongation reactions with each module catalysing two condensation steps. The DEB core of erythromycin is derived from the condensation of one propionyl-CoA starter unit and six methyl malonyl-CoA extender units (Marsden et al., 1994). Apart from chain extender domains each of module 1, 2, 5 and 6 have one or more β-keto reductions domains that are ketoreductase (KR), dehydratase (DH) and enoyl reductase (ER) while no such domain is present in module 3. The N-terminal end of the DEBS-1 module has an extra module called a loading module which consists of an AT and an ACP domain. It has been proposed that this module plays a role in improving the efficiency of the PKS system. In this PKS system, the genetic organization is clearly reflected in the biochemical sequence of events that result in the biosynthesis of heptaketide intermediate (Donadio et al., 1991). That is finally cleaved by thioesterase (TE) domain present at the C-terminal end of the DEBS-3 module. This release by TE domain results in its cyclization to form 6-deoxyerythronolide (6-DEB) which is erythromycin aglycon.

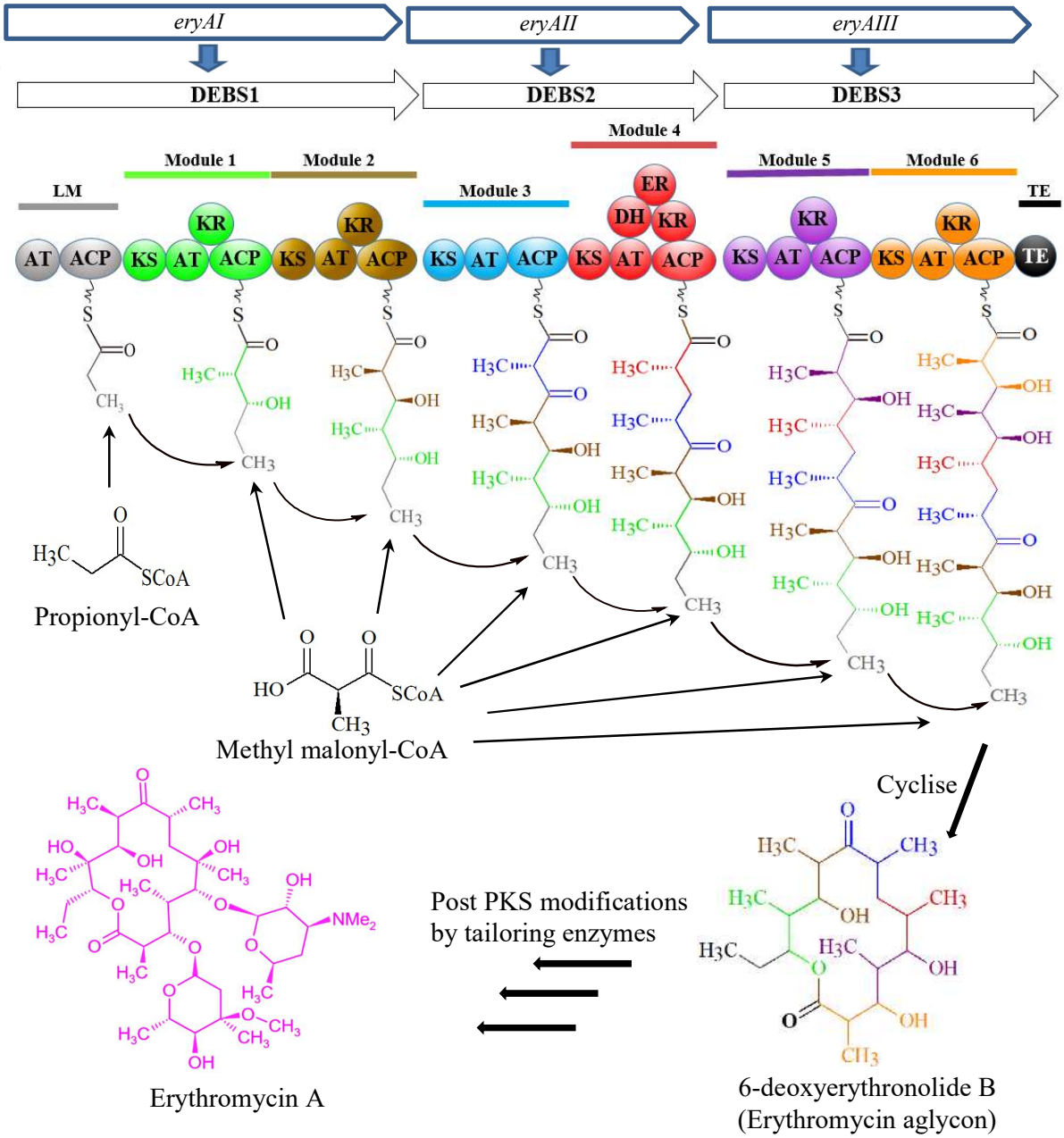
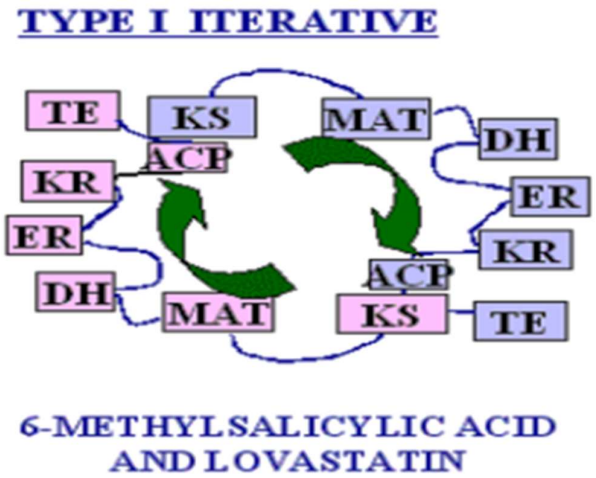


Figure 1.5 Biosynthesis of 6-dEB (erythromycin aglycon). The protein assembly line involved in the biosynthesis is present as 6-dEB synthase, a type I PKS that consists of three polypeptides DEBS1, DEBS2 and DEBS3 encoded by three ORFs *eryAI*, *eryAII* and *eryAIII*, respectively. Biosynthesis is initiated by loading of activated propionate (starter unit) on the loading domain and activated methyl malonate (extender unit) on the first module of DEBS1 that primes the first condensation reaction. The final molecule is released from assembly line by the action of TE domain after elongation of the polyketide chain by the remaining five modules of the assembly line each catalysing condensation reaction between the elongated polyketide chain and a molecule of methyl malonate. KS, ketosynthase; AT, acyl transferase; DH, dehydratase; ER, enoyl reductase; KR, ketoreductase; ACP, acyl carrier protein and TE, thioesterase domain.

1.3.1.2 Iterative type-I PKS

Iterative type I PKSs are characteristic of fungal polyketide biosynthesis that produces molecules like lovastatin, 6-MSA. These PKS enzymes have modular covalent architecture like that of type I non-iterative PKS but have only single module which functions repetitively for chain extension (Figure 1.6A) (Fujii et al., 2001). And for each round of chain extension, the degree of β-keto reduction may vary. It was first reported from the avilamycin cluster of *Streptomyces viridochromogens* (Gaisser et al., 1997b). Other reported iterative type I PKS examples are lovastatin nonaketide synthase (LNKS) and 6-methyl salicylic acid synthase. Recently this type of PKS has been identified in bacteria. One such example is C-1027 biosynthetic cluster reported from *Streptomyces globisporus* that codes for C-1027 an anti-tumour antibiotic (Liu et al., 2002).

(A)



(B)

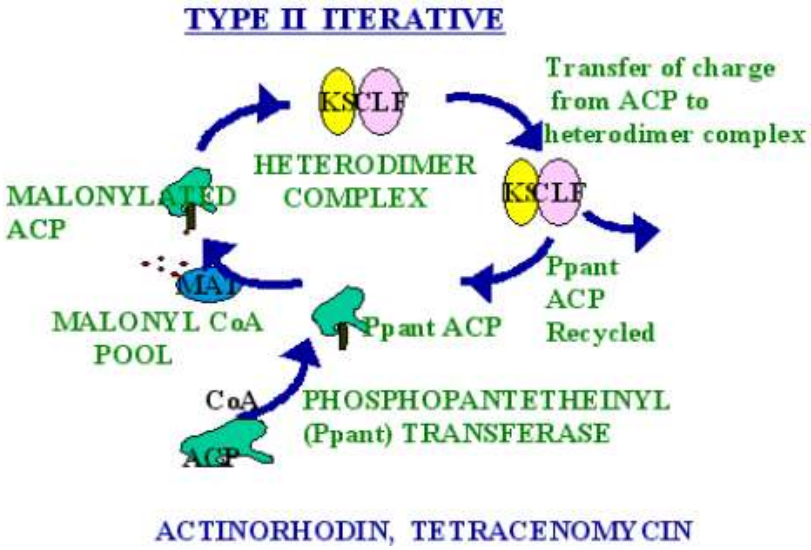


Figure 1.6 Schematic presentation of mode of action of iterative polyketides. (A) Type I iterative PKS (Source: NII, 2010) (B) Type II iterative PKS with examples (Reproduced from Gokhale and Tuteja, 2008).

1.3.2 Type II PKS (aromatic)

The organization of Type II PKSs differ from that of type I PKSs because these are present as discrete proteins that are not covalently linked to each other but are present as complexes of separate mono-functional units/polypeptides (Figure 1.6 B). They function in an iterative manner and catalyse synthesis of compounds that require aromatization and

cyclization. They are similar to bacterial fatty acid synthase. Type II PKS are also known as minimal PKS because they just have KS and ACP domains in which KS has two subunits KS_{α} and KS_{β} (McDaniel et al., 1995, Schneider, 2005). While KSα is responsible for repeated condensation of malonyl-CoA extender units, the length of growing chain is decided by KS_{β} (Bisang et al., 1999). Additional (accessory) domains that are found in them are cyclases (CYC) and aromatases (ARO) which do not play role in the elongation of polyketide chain but modify the poly β-ketoacyl carrier protein intermediate to give the aromatic polycyclic product (Bisang et al., 1999). This type of PKS is found in actinomycetes. A well-studied example of it is actinorhodin PKS which is produced by *Streptomyces coelicolor*.

1.3.3 Type III PKS system (chalcone/stilbene synthase)

Type III PKSs are chalcone synthase (CHS) (Hugler et al., 2002) (Hugler et al., 2002) (Hugler et al., 2002) (Hugler et al., 2002) (Hugler et al., 2002) (Hugler et al., 2002) (Hugler et al., 2002) (Hugler et al., 2002) or stilbene synthase (STS) like proteins that are comparatively small and have single polypeptide chains (Figure 1.7). These are usually found in plants (Schroder, 1999), and are unique in not having phosphopantetheine (P-pant) arm which is characteristic of all the PKS and on which the growing polyketide chain is tethered in all. Instead of using acyl carrier proteins for translocation of substrates and intermediates during biosynthesis CHS uses CoA thioesters. It acts as a unimodular PKS to carry out all the reactions like decarboxylative condensation, cyclization and aromatization. CHS catalyses biosynthesis of the precursor of flavonoids and isoflavonoids, which is 4, 2', 4', 6' tetrahydroxychalcone (Jez et al., 2001).

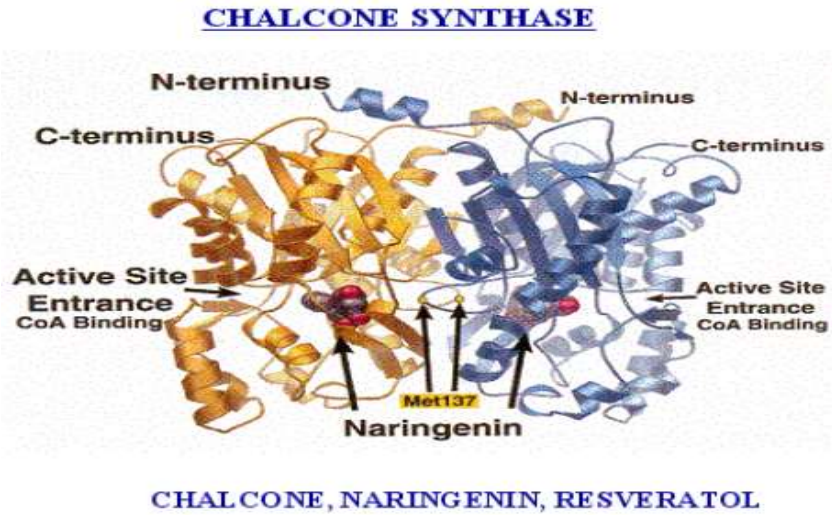


Figure 1.7 Structure of chalcone synthase-a type III PKS(Source: NII, 2010).

1.4 Initiation/priming of polyketide biosynthesis and various starter units of PKS systems

Biosynthesis of most of the polyketides is derived from acetate/malonate and propionate/ methylmalonate moieties that are usually available in the metabolic pool. It begins with the loading of acetyl-CoA onto synthases. Biosynthesis of erythromycin starts with the loading of propionyl-CoA starter units (Hertweck, 2009). Apart from these, several alternative starter units are also used for priming the biosynthesis of PKS that are utilized as CoA thioesters (Figure 1.4). Alternative initiation strategies using acetate starter units have been reported from *trans*-AT PKS systems like pederin and curacin. Loading modules of these systems have domains with distinct structural features for which these are known as $GNAT_L$ domains. These domains belong to GNAT superfamily of N-acyl transferase enzymes. But in systems like curacin it is involved in transfer of acyl group to thiol group of ACP which is a role that is normally performed in most of the PKS loading modules by specific AT_L domain that loads an acyl substrate to the neighbouring ACP_L domain. And this reaction is followed by decarboxylation of the loaded substrate by a KS_Q domain specific to the loading module

which is present at its N-terminal (Figure 1.8). The $GNAT_L$ domain is widespread in *trans*-acetyltransferase systems which may be present as a tridomain (along with other domains) or as didomain like in onnamide A biosynthetic pathway. These loading domains exhibit decarboxylase and S-acyl transferase activities essential for the loading of substrate (Figure 1.8) (Piel et al., 2004, Gu et al., 2007). Knowledge about the mechanism of supply, activation and loading of diverse starter units has created ways to broaden the horizons for the development of novel polyketide derivatives with novel/enhanced activity (Lee et al., 2005, Gregory et al., 2005, Liu et al., 2011, Liu et al., 2014).

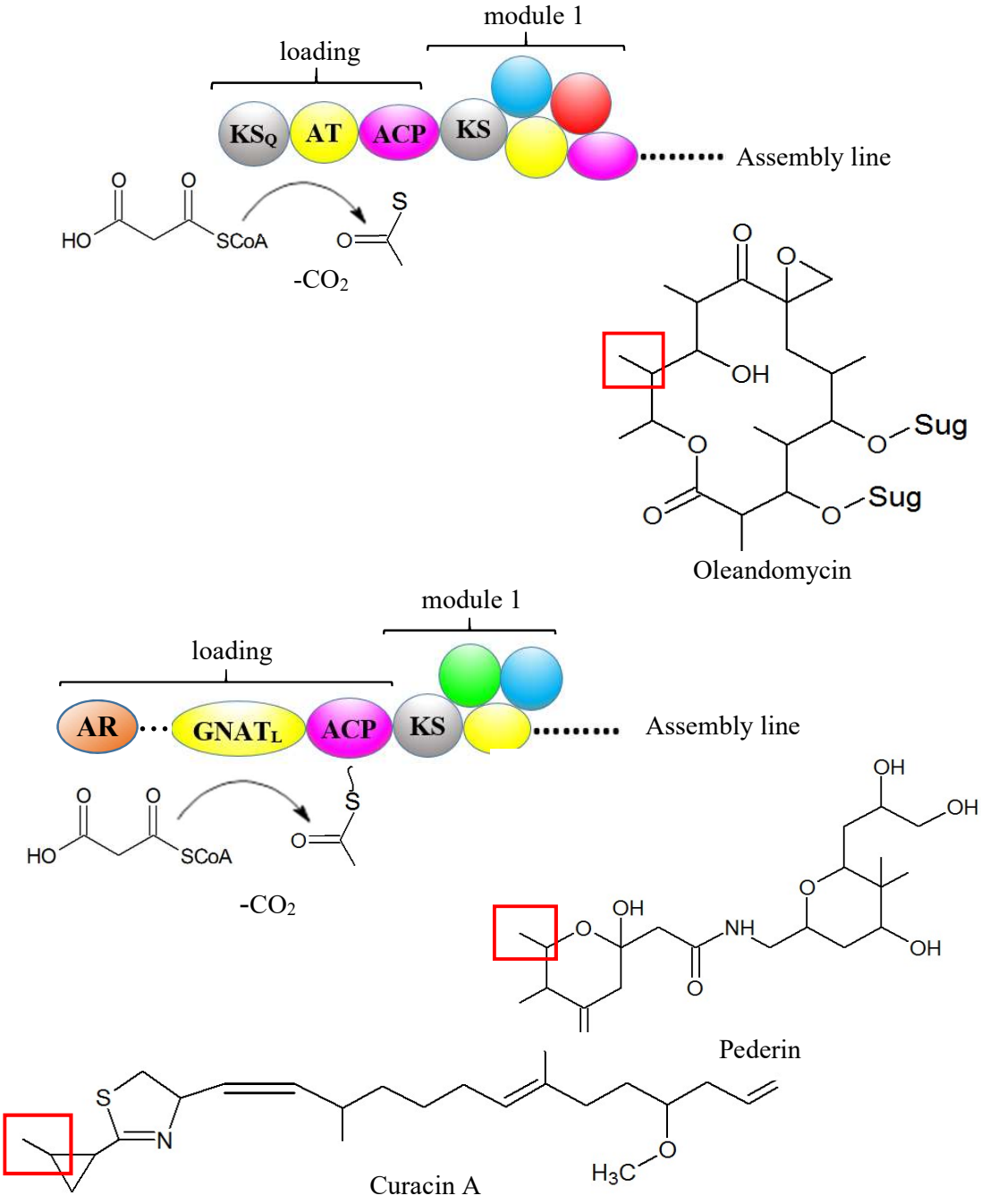


Figure 1.8 Priming of PKSs with acetyl starter units involves different loading domains in different biosynthetic clusters. Positions of groups derived from starter units in the final molecules are marked with red squares. KS, ketosynthase; KS_Q, ketosynthase domain in which active site cysteine is replaced with glutamine residue; AR, adapter domain; ACP, acyl carrier protein; AT, acyl transferase; AT_L, loading acyl transferase; $GNAT_L$, N-acetyl transferase of loading domain (Adapted from Hertweck, 2009).

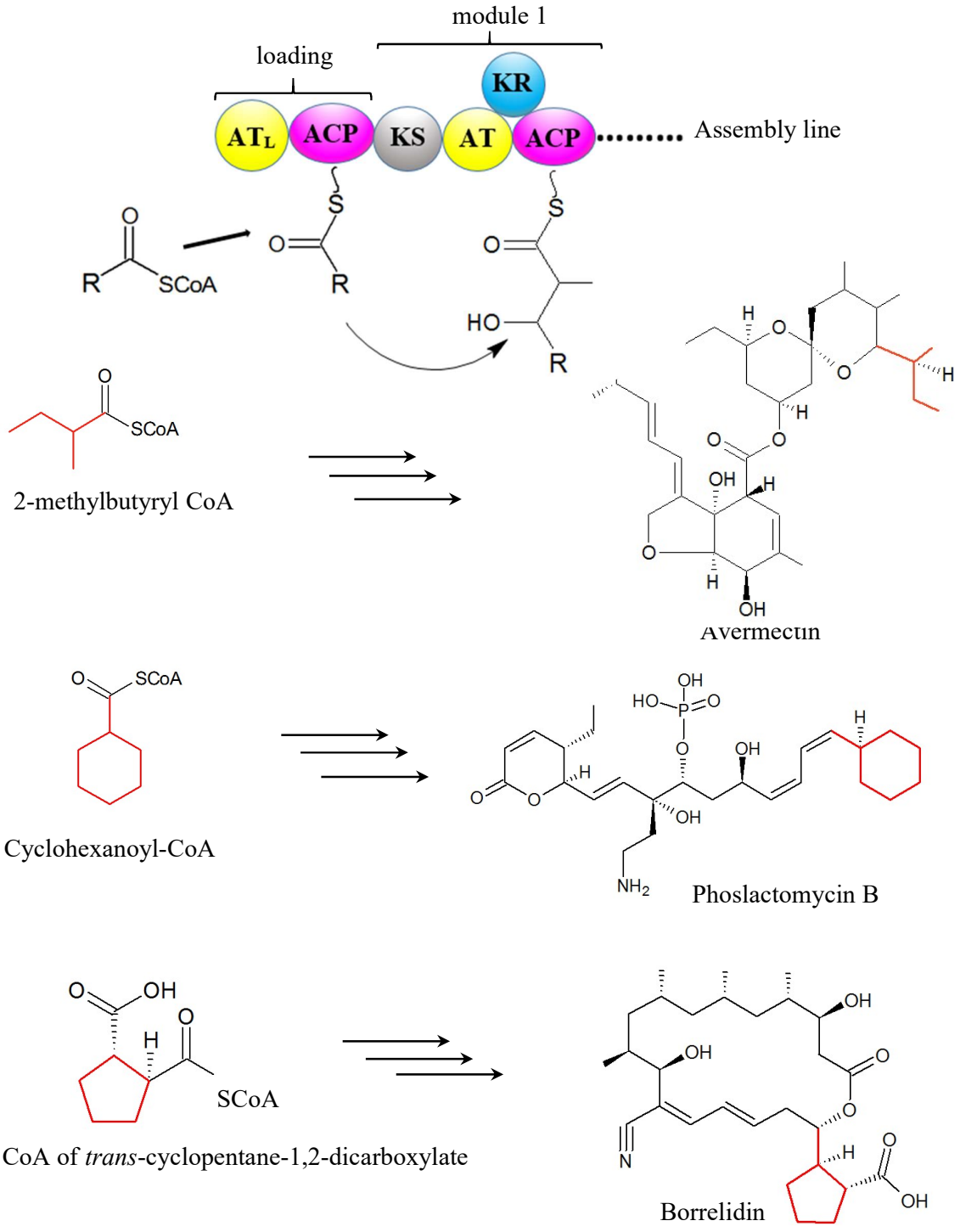


Figure 1.9 Priming by various non-acetate CoA thioesters is used to initiate biosynthesis of some polyketidesKS, ketosynthase; ACP, acyl carrier protein; AT, acyl transferase; AT_L, loading acyl transferase; KR, ketoreductase (Adapted from Hertweck, 2009).

1.5 Diversity of extender units in polyketide biosynthesis

Most of the PKS systems found in bacteria use malonyl-CoA (MCoA) or methylmalonyl-CoA (mMCoA) for chain elongation of the polyketide. It has been shown by domain swapping and mutagenesis experiments that specific motifs in particular AT domains govern the choice for selection of the type of extender unit (Liou and Khosla, 2003). A rare extender unit that is reported to be utilized in bacterial systems is 2 ethyl malonyl-CoA (eMCoA) which is used in the biosynthesis of antibiotic niddamycin by *Streptomyces caelestis* (Stassi et al., 1998) and ascomycin (immunosuppressant FK 520) from *Streptomyces hygroscopicus var. ascomyceticus* (Wu et al., 2000) as well as tylosin, concanamycin and kirramycin. Figure 1.10 shows various malonyl derived extender units that are utilized by modular PKSs. In some complex polyketide structures incorporation of heterosubstituted malonyl derivatives has been reported. Hydroxymalonylate (hoM) and aminomalonate (aM) are other extender units that are used in the biosynthesis of zwittermicin antibiotic produced by *Bacillus cereus* (Figure 1.11) (Emmert et al., 2004). A variety of other extender units have been reported to be incorporated into either polyketide systems or NRPS-PKS hybrid systems (Hertweck, 2009).

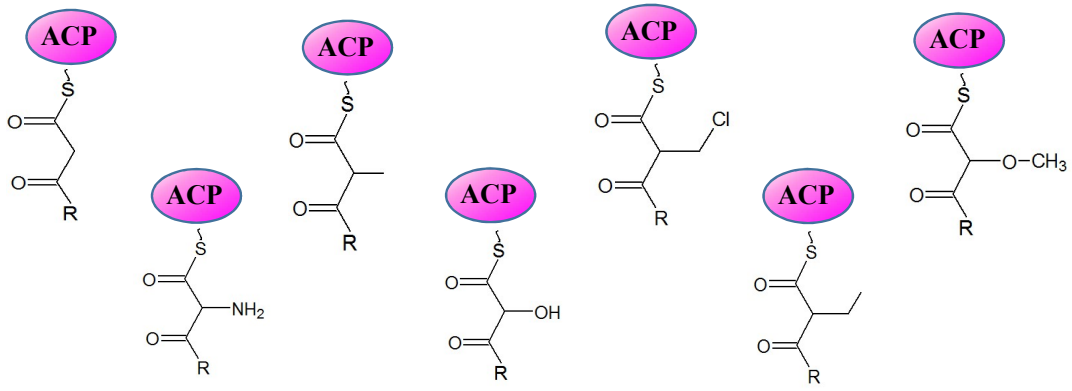


Figure 1.10 Malonyl derived extender units used by modular PKSs(Adapted from Hertweck, 2009).

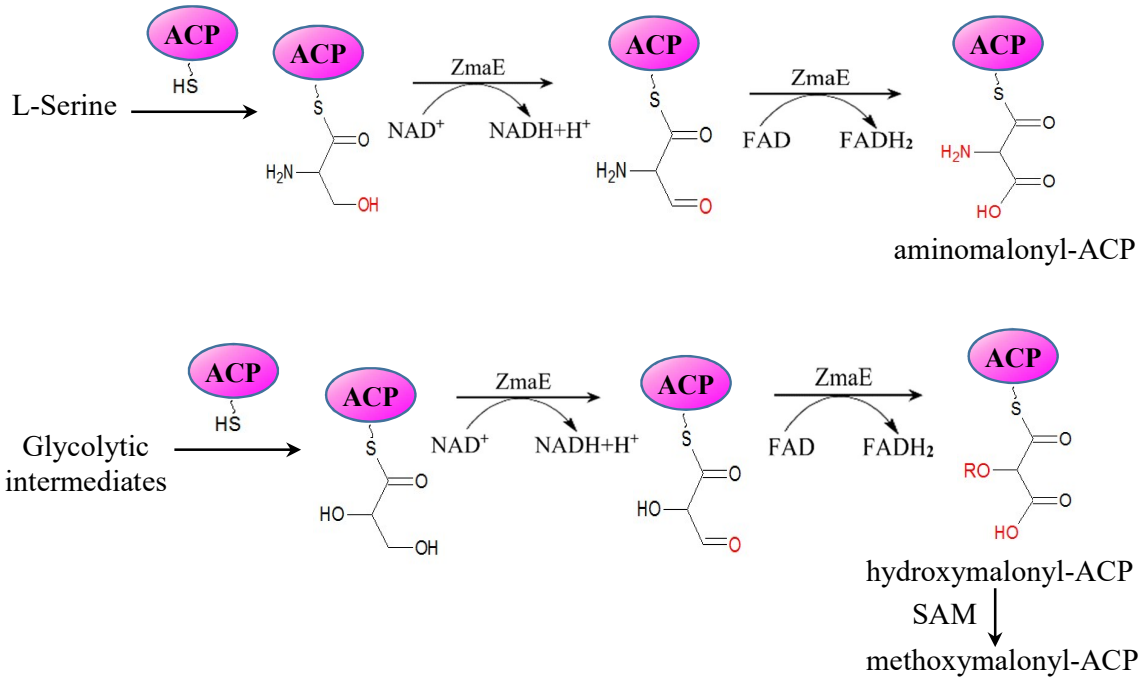


Figure 1.11 Biosynthesis of hydroxyl, methoxy and aminomalonyl-ACP extender units by zwittermicin biosynthetic pathway. Aminomalonyl-ACP is derived from serine while methoxymalonyl-ACP is derived from glycolyl intermediates as a result of methylation of hydroxymalonyl-ACP by SAM derived methyl group. ZmaE, oxygenase; SAM, *S*-adenosylmethionine (Adapted from Hertweck, 2009).

1.6 Auxiliary proteins in PKS systems

Apart from core PKS genes that are responsible for the biosynthesis of carbon backbone (carbon-carbon bonding) of polyketides, PKS systems also have other enzyme activities that are mostly mono-functional and which are crucial for the production and activity of the final PKS molecule. Domains for these auxiliary proteins encode enzymes for the biosynthesis of alternative or specialized starter and extender units, enzymes for antibiotic resistance or the enzymes for modification of intermediate PKS during polyketide biosynthesis as well as a number of tailoring enzymes responsible for the modification of core PKS moiety after it is synthesized etc. The auxiliary enzymes that modify the core polyketide moiety once after it is biosynthesized as a result of action of PKS, are called post-

PKS tailoring enzymes and the modifications caused by them are known as post-PKS modifications which are discussed in next section.

Domains for auxiliary functions like enoylreductase, dehydratase and ketoreductase that catalyse reduction of double bond, hydroxyl group and keto group, respectively, or for methyltransferases that are responsible for insertion of methyl group may be found within modules of type I PKSs. These modify the biosynthetic intermediates accordingly during the polyketide biosynthesis. Such modules are widely found in polyketide biosynthetic clusters like erythromycin gene cluster (Donadio and Katz, 1992), mupirocin biosynthetic cluster (Thomas et al., 2010), kalimantacin/batumin gene cluster (Mattheus et al., 2010), avermectin biosynthetic cluster (Ikeda et al., 1999) etc. Auxiliary proteins for these functions are also found as standalone proteins in the polyketide biosynthetic clusters where these are present as discrete proteins which act in *trans,* to modify the biosynthetic intermediates selectively. It has been shown that standalone auxiliary protein LovC having enoyl reductase activity is necessary for the proper functioning of LNKS-the nonaketide synthase of lovastatin biosynthetic cluster (Kennedy et al., 1999). In many biosynthetic clusters such standalone auxiliary proteins represent two or more activities that are fused together. For example, MmpC of mupirocin biosynthetic cluster in which enoyl reductase activity is present along with acyl transferase activity and both these functions are supplied in *trans,* for the reduction of biosynthetic intermediates and for the loading of extender units respectively during the biosynthesis (Hothersall et al., 2007, Thomas et al., 2010). A group of *trans* acting auxiliary proteins together known as HCS cassette has been shown to be involved in the β- carbon branching of the polyketides during their biosynthesis (Haines et al., 2013a). This cassette/group of genes has been found to be of wide distribution in *trans*-AT polyketide

biosynthetic clusters like pederin, mupirocin, bacillaene, thiomarinol, curacin, jamaicamide, and kalimantacin/batumin (Piel, 2010, Helfrich and Piel, 2016).

Apart from enzymatic roles that auxiliary proteins play in polyketide biosynthetic clusters these are also found to be involved in non-enzymatic functions. It has been shown in different polyketide biosynthetic systems that auxiliary proteins are essential for the proper functioning of glycosylating enzymes that are involved in the glycosylation of polyketide moieties. For example, DesVIII auxiliary protein improves transfer of TDP-D-desosamine by DesVII; a glycosyltransferase in the biosynthesis of macrolide antibiotics, neomethymycin, pikromycin and methymycin in *Streptomyces venezuelae* ATCC15439 (Hong et al., 2007). Several homologues of DesVIII auxiliary protein have been reported in different biosynthetic clusters like EryCII from erythromycin biosynthetic cluster of *Sacchropolyspora erythraea* (Hong et al., 2007), AknT from aclacinomycin A biosynthetic cluster of *Streptomyces galilaeus* (Lu et al., 2005), DnrQ from doxorubicin biosynthetic cluster of *Streptomyces peucetius* (Hong et al., 2007), MydC from mycinamycin biosynthetic cluster of *Micromonospora griseorubida* (Melancon et al., 2004), TylM3 from tylosin biosynthetic cluster of *Streptomyces fradiae* (Melancon et al., 2004) and OleP1 from the oleandomycin biosynthetic cluster of *Streptomyces antibioticus* (Doumith et al., 1999). Similarly, Srm6 and Srm28 auxiliary proteins are found in spiromycin biosynthetic cluster of *Streptomyces ambofaciens* that aid glycotransferases Srm5 and Srm29 respectively (Nguyen et al., 2010). These auxiliary proteins are required by respective glycotransferases for different reasons. Some auxiliary proteins are predicted to have chaperone like function with partner glycotransferases like EryCII which enhances the activity of EryCIII by allosteric binding (Moncrieffe et al., 2012), while others like DesVII/DesVIII have been found to be involved in the folding of glycotransferase, that are required to be tightly bound to it during catalysis

while playing no role in actual catalysis (Borisova and Liu, 2010). Auxiliary proteins sometimes are dispensable as well, like Srm6 of spiromycin cluster as it has been found that glycosylation of Srm5 is also supported by Srm28 (Nguyen et al., 2010).

Auxiliary proteins which include post PKS tailoring enzymes may act in *cis* or in *trans* during the biosynthesis of polyketides depending on the location of genes that encode them and where they finally act. Auxiliary proteins coded by genes that are found embedded with PKS genes (in modules) or by those that are found adjacent to them and which work alongside products of PKS genes act in *cis* during the biosynthesis, while *trans* acting auxiliary proteins are coded by gene functions located elsewhere as standalone or fused with other gene functions away from their site of action.

1.7 Post-PKS modification of polyketides

The core polyketide structure that is synthesised by PKS is subjected to further/secondary modifications by the action of various enzymes to modulate its functionality or to turn it into the bioactive compound. Post-PKS modifications are crucial for the final activity of PKS molecules (Weymouth-Wilson, 1997, Weissman and Leadlay, 2005). These post PKS-tailoring enzymes add diversity to polyketide molecules by acting on the core polyketide moiety after its biosynthesis by the PKS (Staunton and Weissman, 2001). The tailoring enzymes may oxidise the molecule with the introduction of new carbonyl or hydroxyl groups, or these may methylate the molecule at different atoms/centres like oxygen/nitrogen or carbon. These may also glycosylate or halogenate the polyketide molecule. In erythromycin biosynthesis, the 6-DEB molecule which is released from PKS is hydroxylated, glycosylated and acted upon by methyltransferase before giving rise to active erythromycin-A, the final product (Katz, 1997).

1.8 Post-polyketide modifications in erythromycin biosynthesis

Before erythromycin is produced as an active molecule and after its core polyketide structure, which is 6-deoxyerythronolide B- a macrolactone, is biosynthesized by type I PKS DBS it is acted upon by many processing enzymes. The first tailoring modification of the polyketide core of erythromycin is done by P450 hydroxylase enzyme EryF that causes hydroxylation of the C-6 carbon of the basic macrolactone ring. By mutating *eryF* it was possible to obtain erythromycin molecules without this modification (Weber et al., 1991). The next tailoring modification is the addition of two sugar molecules L-mycarose and D-desosamine. Products of genes *eryBII to eryBVII* with the exception *of eryBV* are involved in the biosynthesis of L-mycarose from TDP-4, 6-deoxyglucose. In addition, the product of *ery*BV participates in the attachment of L-mycarose at the C-3 hydroxyl group of core macrolactone unit. Similarly, D-desosamine is also synthesized from TDP-4, 6-deoxyglucose by the action of enzymes coded by *eryCI to eryCVI* except of *eryCIII* which is involved in its attachment to the C-5 hydroxyl group of the macrolactone unit (Summers et al., 1997). These three modifications result in the production of erythromycin D, a metabolic intermediate in the biosynthesis of active erythromycin A. Depending upon the action of further tailoring enzymes, erythromycin D is either converted into erythromycin C or erythromycin B. Erythromycin D is acted upon by products *eryK* and *eryG* which causes hydroxylation at C-12 and O-methylation of the C-3" hydroxyl group of the attached mycarose molecule respectively (Lambalot et al., 1995, Paulus et al., 1990). Depending upon the order in which either of these modifications occur (i.e. the order of action of enzymes), erythromycin D is converted into intermediate metabolites either erythromycin B or C. If erythromycin D acted upon by EryK firstly it was converted into erythromycin C, but if EryG acted upon it firstly this resulted in the formation of erythromycin B which was a shunt product. Finally, bioactive erythromycin A is produced by the action of either EryG or EryK. It has been

reported that EryG shows four-fold preference for erythromycin C over D as a substrate (Paulus et al., 1990) while EryK shows a 1200-1900 fold preference for erythromycin D over B as a substrate (Lambalot et al., 1995). The importance of tailoring regions in the production of the final bioactive molecule and its yield is clearly evident from this example.

Table 1.1 Functions of non-PKS genes involved in the biosynthesis of erythromycin (Modified from Katz, 1997).

Gene	Identified function	Role in erythromycin biosynthesis	References
eryBI	Not known	Except *eryBV* which is	Summers et al., 1997, Gaisser et al., 1998
eryBII	3-ketoreductase	responsible for the attachment of	Summers et al., 1997, Gaisser et al., 1998
eryBIII	Not known	L-mycarose and *eryBI* which is not	Summers et al., 1997, Gaisser et al., 1998
eryBIV	4-Keto-reductase	essential for biosynthesis, rest	Summers et al., 1997, Gaisser et al., 1997a
eryBV	Mycarosyl-transferase	all are involved in the biosynthesis of	Summers et al., 1997
eryBVI	2,3 dehydratase	L-mycarose.	Summers et al., 1997, Salah-Bey et al., 1998
eryBVII	5-epimerase		Summers et al., 1997
eryCI	3-aminotransferase	Except *eryCIII* which is	Summers et al., 1997
eryCII	3,4 isomerase	responsible for attachment of	Summers et al., 1997, Salah-Bey et al., 1998
eryCIII	Desomethyltransferase	D-desoamine, all others are involved	Summers et al., 1997
eryCIV	3,4-dehydratase	in its biosynthesis	Summers et al., 1997
eryCV	3,4 reductase		Summers et al., 1997, Salah-Bey et al., 1998
eryCVI	3-aminodimethyl-transferase		Summers et al., 1997
eryF	P_{450} monoxygenase	C-6 hydroxylation	Weber et al., 1991, Haydock et al., 1991
eryG	O-methyltransferase	C"-3 O-methylation	Weber et al., 1989, Paulus et al., 1990
eryI	thioesterase	Unknown	Weber et al., 1991, Haydock et al., 1991
eryK	P_{450} monooxygenase	C-12 hydroxylation	Stassi et al., 1993, Lambalot et al., 1995
ermE	N-methyltransferase	Resistance	Uchiyama and Weisblum, 1985

1.9 Exploitation of tailoring enzymes for the development of new polyketide molecules with novel/ improved bioactivities

Tailoring enzymes provide another way of getting novel molecules with improved bioactivity as they can be easily manipulated with minimum invasion of the PKS clusters along which these are typically located. Because they are often present as discrete ORFs and their activity is not because of being part of a multifunctional polypeptide, they can be easily genetically manipulated. Manipulation of just one or two tailoring enzymes that modify a functional group on the core polyketide framework like hydroxylation or addition or deletion of sugar moiety or its modification may result in novel /altered activity of basically the same polyketide molecule. In fact, genetic manipulation of tailoring enzymes provides the best and potentially easiest way of producing novel polyketides with new/enhanced activities (Katz and Donadio, 1993, Olano et al., 2010). By manipulating the *carE* tailoring region gene of biosynthetic pathway of carbomycin antibiotic a novel polyketide molecule was created. The *carE* gene codes for acylase whose proposed activity is to attach the isovaleryl or butyryl to mycarose moiety of this antibiotic. Expression of the *carE* gene in *Streptomyces ambofaciens* that endogenously produces antibiotic spiramycin led to the isolation of modified antibiotic isovaleryl-spiramycin (Epp et al., 1989). Similarly, another tailoring region gene of the carbomycin biosynthetic cluster from *Streptomyces thermotolerans*, namely *acyA,* that encodes 3-O-acyltransferase activity, was expressed in a tylosin producer *Streptomyces fradiae* which resulted in the isolation of 3-O-acetyltylosin (Arisawa et al., 1994). Many times some of the tailoring region genes are required to be expressed together to bring about an effect in the polyketide structure like those for attachment of sugars or those involved in oxidation reactions. Different tailoring enzymes were coexpressed in heterologous *E. coli* host to achieve diversification of glycosylation in order to get better erythromycin analogues (Jiang et al., 2013, Zhang et al., 2015). Similarly, biosynthesis of

19-*O*-perosaminyl-amphoteronolide B was engineered in *Streptomyces nodosus* by co-expressing *Streptomyces aminophilus perDII* perosamine synthase and a hybrid of *perDI* perosaminyltransferase and *amphDI* genes (Hutchinson et al., 2010). In such cases, concerned tailoring functions were required to be expressed together to bring about the desired change/modification in polyketide structure. Substrate specificities of tailoring genes have also been one of the major concerns in producing altered polyketide moieties for example in erythromycin biosynthesis deletion of *eryF* gene blocks the hydroxylation of 6-deoxyerythronolide-B at C-6. On the other hand, this change does not affect subsequent attachment of the sugars, mycarose and desoamine, by glycosyltransferases (Weber et al., 1991).

1.10 Resistance in polyketide producers

The organism that produces antibiotics also has mechanisms/strategies to protect itself from them. Thus it has been found that genes for particular antibiotic production and those for resistance from it are not only found together within the natural PKS gene clusters but also co-regulated together to protect them from the effect of produced polyketide. Therefore, it has been suggested that mechanisms for self-protection have co-evolved with the biosynthetic pathways of antibiotics (Fischbach and Walsh, 2006, Walsh, 2006). The gene for the resistance from erythromycin, that is *ermE*, is found in erythromycin gene cluster of *S. erythraea* (Dhillon et al., 1989). Similarly, genes for the resistance from actinorhodin are found with the genes responsible for its biosynthesis in *Streptomyces coelicolor* (Caballero et al., 1991). Genes for resistance to carbomycin are located on either side of the gene cluster responsible for the biosynthesis of this antibiotic (Epp et al., 1989).

1.11 The *trans*-AT polyketide synthases

There is a class of type I modular PKSs that do not follow the colinearity principle of biosynthesis between the genetic organization of biosynthetic cluster or enzyme structure and the structure of the metabolite that they produce, which is because some of the domains in their enzyme architecture are not at all used or are redundant during biosynthesis. To the contrary, some are used more than once resulting in metabolites whose structure does not clearly reflect structure/organization of their enzyme/genes in the cluster. This class of type I modular PKSs is characterized by the lack of cognate (*cis*) AT domain in their structure which is present separately in *trans* as one or more *trans*-ATs. These PKS systems utilize discrete AT functions encoded by different genes to load ACPs and are therefore called *trans*-AT type I modular systems as against to the classical which are *cis*-AT PKS (Piel, 2010, Musiol and Weber, 2012, Helfrich and Piel, 2016).

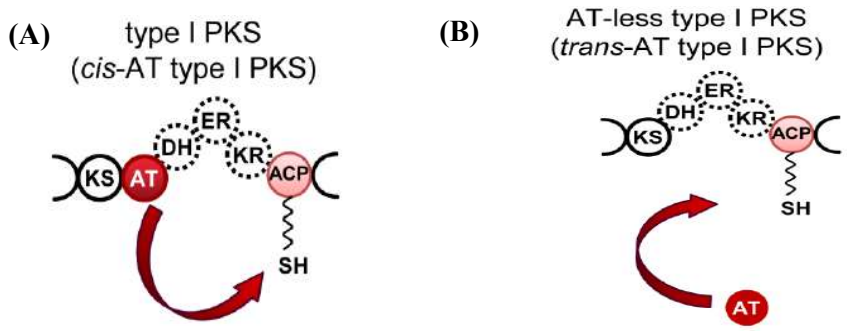


Figure 1.12 Domain architectures of type I PKS.(A) a *cis*-AT type I PKS and (B) *trans*-AT type-I PKS (Reproduced from Musiol 2012).

The *trans*-AT PKSs do not obey the colinearity rule about PKS structure and the metabolic product or any other colinearity between the order of genes encoding a PKS and the steps involved in its biosynthesis which is observed about erythromycin gene cluster. Examples of such peculiarities are the presence of unusual domain orders, repeated domains or the presence of domains in two parts with each part in a different protein etc. An analysis

of these PKSs has revealed that those KS modules share more similarities with each other which process similar substrates (Nguyen et al., 2008). Similarly, apart from the single type of dehydrating modules found in *cis*-AT PKS systems, two types of dehydrating bimodules are commonly found in *trans*-AT PKS systems. These are: type A which has domain structure KS KR ACP KS^0 DH ACP and type B that has KS KR ACP KS DH ACP KR; each of which introduces characteristic double bonds (Piel, 2010). While KS^0 is the KS domain that lacks the specific histidine residue at its active site which is required for elongation of the carbon chain. Examples of *trans*-AT PKS are mupirocin, thiomarinol, kalimantacin/batumin, myxovirescin, kirramycin etc. It has been shown that among fused tandem free-standing/trans AT's the function of other AT (AT2) is not to load the substrate units but is exclusively dedicated to proofreading activity (Jensen et al., 2012). The *trans* acting oxidoreductase domain is also reported to be found fused with tandem or single ATs, as *trans*-acting ER (Musiol and Weber, 2012) among *trans*-AT polyketide systems. The *trans*-AT polyketide systems also show both types of carbon branching in their structures. Methyl, ethyl or another carbon branching at α-position in their structures results either from the incorporation of corresponding branched derivative substrate or by S-adenosylmethionine (SAM)-dependent α-methylation which involves a methyltransferase (MT) domain in their corresponding module. Instead, β-branching is the result of functions acting in *trans* that include enzymes of the HMG-CoA pathway (Piel, 2010).

Table 1.2 Domain architecture of discrete ATs (Musiol and Weber, 2012).

Domain architecture	AT type	Examples (domains of clusters of)
	Single AT	BaeC (PksC) and BaeD (PksD) of Bacillaene, FenF of Mycosubtilin, KirCII of Kirramycin, OzmC of Oxazolomycins, VirI of Virginiamycin, ZmaF of Zwittermicin LkcD of Lankacidin.
	Tandem AT	BryP of Bryostatins, RhiG of Rhizoxin, Tav of Myxovirescin
	Single AT with N-terminal ER	BaeE, (PksE) of Bacillaene, DisD (DszD) of Disorazol, LnmG of Leinamycin
	Tandem AT with N-terminal ER	KirCl of Kirramycin, MmpC of Mupirocin, OzmM of Oxazolomycins

1.12 Structural model for type I polyketide synthases

Till now no experimental data was available about structures of polyketide synthases but now there is some structural insight available to test models that were proposed about their structures. Now new models have been suggested based on the structural data made available by using techniques of SAXS and cryo-EM (Weissman, 2015). Earlier, two models were proposed for the structure of erythromycin polyketide synthase based on the data of structural studies and other findings (Figures 1.13 and 1.14).

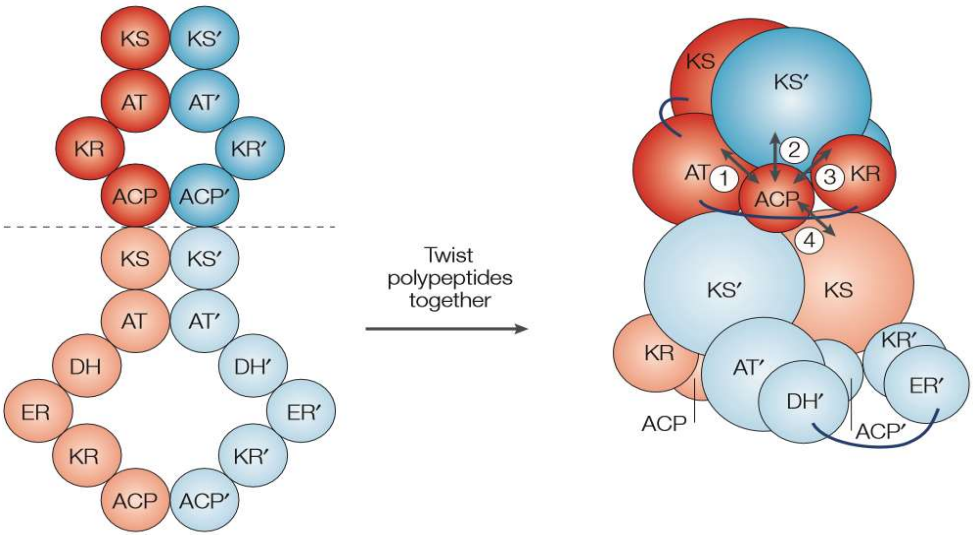


Figure 1.13 Cambridge model for type I PKS of erythromycin (DEBS). This model proposed head-to-head and tail-to-tail arrangement of homodimeric PKS modules. According to this model DEBS PKS subunits were twisted together to form double helical architecture. This model gave most acceptable arrangement of domains and convincingly explained their interaction during functioning of the PKS. Recent data about structure of PKS generated by using techniques of cryo-EM and SAXS also supported these features proposed in this model. All the new models that are proposed about PKS architecture are based on these features of this model (Reproduced from Weissman and Leadlay, 2005).

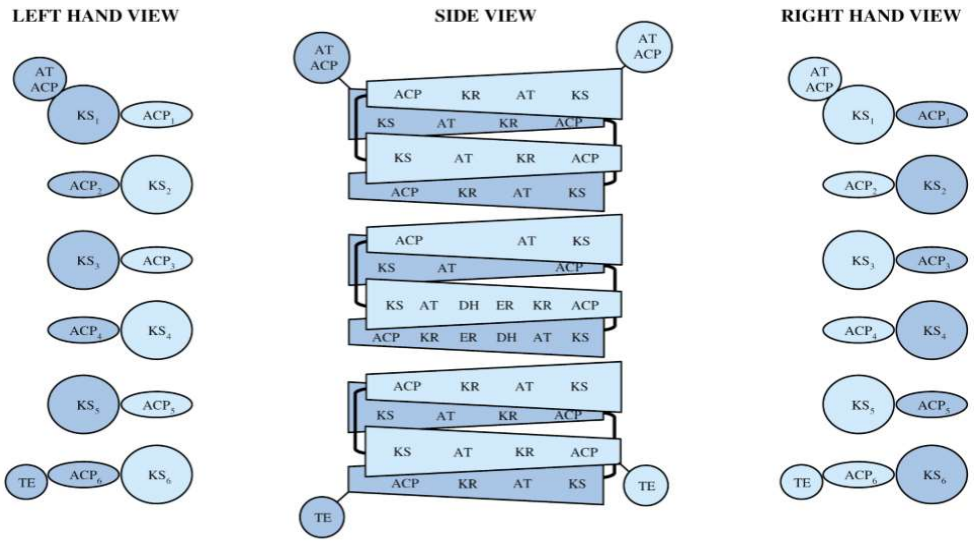


Figure 1.14 Head-to-tail model for the structure of type I PKS of erythromycin (DEBS).This was the first model that was proposed about the structure of type I PKS. It failed to support experimental data of proteolytic digestions of DEBS PKS (Reproduced from Staunton and Weissman, 2001).

1.13 Protein-protein interactions in polyketide biosynthesis

Biosynthesis of polyketides by PKS requires communication between different components present in its architecture like successive modules, domains, etc. Whenever modules are split into two polypeptides they need to interact with each other to permit transfer of intermediate metabolites. For example, in the case of DEBS, which is present as three polypeptides, the interaction among them should occur in a certain specific manner so that the reaction intermediates encounter different active sites present in various modules/domains in the correct order. For successful biosynthesis of erythromycin aglycon, C-terminus of DEBS1 module should interact in a proper way with the N-terminus of DEBS2, not of DEBS3. Short stretches of amino acids called as linkers that join various domains and modules covalently in the PKS architecture play a crucial role in the passage of acyl intermediates during polyketide biosynthesis (Donadio and Katz, 1992). Short stretches of amino acids which join domains within modules are known as inter-domain linkers while longer linkers called inter-modular linkers join modules covalently. Still larger linkers which connect C-terminal of a module with N-terminal of another module which is present on different polypeptide (PKS subunit) are called docking domains (Gokhale et al., 1999). It has been reported using genetic engineering that mismatched linker pairs in a PKS significantly impair protein-protein interactions (Tsuji et al., 2001a, Tsuji et al., 2001b, Wu et al., 2001, Weissman et al., 2006, Weissman, 2006). Apart from mismatched linkers, the correct pairing of donor ACP in the biosynthetic machinery with its natural downstream acceptor KS domain is very crucial for inter-modular transfer of intermediates (Wu et al., 2002).

1.14 Combinatorial biosynthesis of polyketides/synthases

The principle of colinearity about PKSs combined with the flexibility of various options/flexibilities regarding PKS domains and levels of modification allows them to be a

very promising tool in the drug discovery using combinatorial biosynthesis and approaches of combinatorial genetics (Cane et al., 1998, Weissman and Leadlay, 2005, Wong and Khosla, 2012, Ladner and Williams, 2016, Till and Race, 2016). One-to-one correspondence between complement of enzymes/enzyme activities in PKSs with the biosynthetic steps for biosynthesis of metabolites and their modular structure in which such activities are unambiguously represented in the form of specific discrete domains gives the platform for combinatorial biosynthesis of polyketides. On account of linear nature of correlation between PKS structure and product, particularly for type I PKS, these are amenable to genetic manipulation (Kittendorf and Sherman, 2006). Methods of genetic engineering are used to modify them for addition or deletion of domains or modules or just few amino acids to affect docking so as to alter the activity of enzymes or to alter their substrate specificity etc. by adding or deleting gene(s) responsible for the same. Thus, when the specific change(s) are made in PKS structure by targeting particular genes(s) it is often possible to predict correctly, assuming the various PKS domains act independently, which biosynthetic step(s) will be affected and what product it will result in. This way of considering the colinear relationship between the PKS structure and enzymatic steps in the biosynthetic pathway of polyketide antibiotics theoretically means that it is possible to create such novel/modified molecules by manipulating biosynthetic pathways by recombining various components (modules or domains) of relevant biosynthetic genes from two similar systems (Kumar et al., 2004, Reeves et al., 2004, Weissman, 2004, Weissman and Leadlay, 2005). Using such approaches, it was possible to produce analogues of avermectins with the desired changes (Figure 1.15) or to engineer biosynthesis of triketide by DEBS 1 polypeptide alone (Figure 1.16). There are many polyketide antibiotic biosynthesizing systems similar to the mupirocin biosynthetic cluster like kalimantacin/batumin, myxovirescin etc. whose components could be engineered into the mupirocin system using the knowledge of genetic engineering and approaches of

combinatorial biosynthesis (Kumar et al., 2004, Menzella et al., 2005, Zhou et al., 2008) in order to produce mupirocin like novel structures/derivatives which may have novel useful activities. In practicality, there are several limitations to this approach as this not only requires complete knowledge of relevant biosynthetic steps but also deeper knowledge on specific issues like substrate preference/tolerance of specific domains, interaction between various catalytic domains and proper folding of hybrid proteins etc. (Khosla et al., 1999, Rowe et al., 2001). In practice it has been observed that despite various specific enzymatic activities clearly and exclusively assigned to specific domains, the efficiency of combinatorial biosynthesis is marred by the fact that PKS behaves like assembly lines, as a result the processing of metabolic intermediates by engineered/hybrid PKSs is not efficient (Figures 1.15 and 1.17). Success in combinatorial biosynthesis relies not only upon recognition of specific enzymes of their substrates but also upon successful passage of metabolic intermediates during the course of synthesis through its assembly line architecture right from its N-terminal until their end at C-terminal. There the TE domain finally releases the product for further modification cyclization and/or for the action of tailoring enzymes, if any. It has been dissected that various non-covalent (docking etc.) and covalent protein-protein interactions between elements of the PKS architecture are responsible for the successful passage of metabolites through it (Tsuji et al., 2001a, Tsuji et al., 2001b, Weissman and Muller, 2008). In multimodular PKSs, a module is responsible for one round of elongation after which the elongated intermediate is transferred to the next module. Various recognition and protein-protein interactions ensure that the intermediate is transferred from ACP to KS of next module and not back to the earlier module (Kapur et al., 2012). The ACP, which is present in each module, on one hand coordinates with the KS of its own module for chain elongation at the catalytic site while on the other hand coordinates with the KS of the proceeding module for chain translocation. The KS acts as the gate-keeper of

the assembly line system where it accepts the intermediate chain only if it recognizes it (Kapur et al., 2010). Thus, a large number of issues like substrate specificity/tolerance of various domains, communication between different modules (inter-modular as well as intra-modular) and even linkers joining the various domains/modules have been found to play critical role in the biosynthesis of polyketides by modular PKSs (Wu et al., 2002).

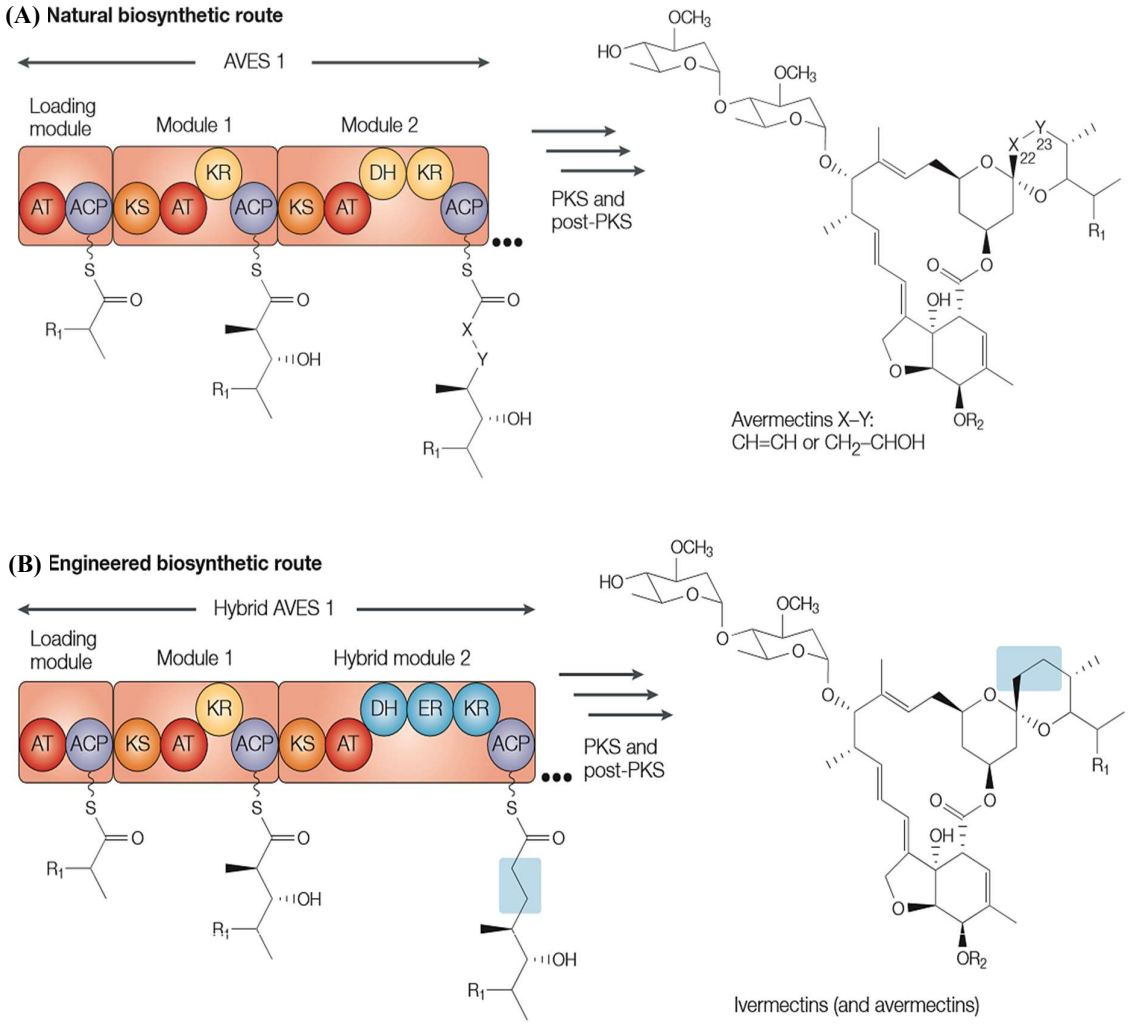


Figure 1.15 Production of ivermectins and related drugs-analogues of avermectins by genetic engineering of the natural avermectin biosynthetic cluster.(A) Avermectins that are produced by natural biosynthetic pathway have unsaturation at C22-C23 or hydroxyl group at C23. (B) Replacement of the cognate module responsible for these modifications by a module from rapamycin PKS that had complete set of reducing domains (all the three DH, ER and KR) needed to achieve complete reduction resulted in the production of various analogues of avermectins (ivermectins) in which these carbons were fully reduced. Hybrid PKS also produced unreduced avermectins that they naturally produce, which appeared to be the result of incomplete processing of biosynthetic intermediates. KS, ketosynthase; AT, acyl transferase; DH, dehydratase; ER, enoyl reductase; KR, ketoreductase; ACP, acyl carrier protein and AVES, avermectin PKS (Reproduced from Weissman and Leadlay, 2005).

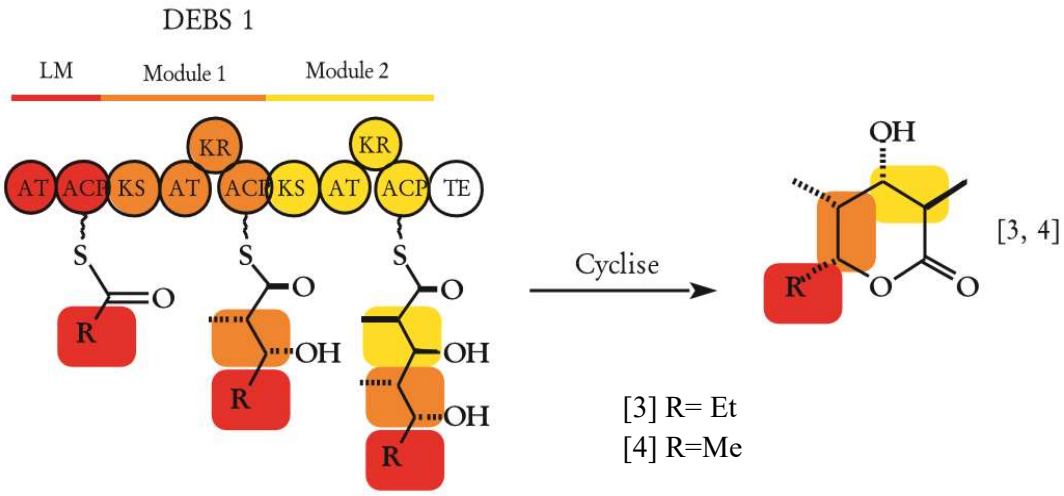


Figure 1.16 Domain organization of an engineered PKS, a bimodular synthase, comprising DEBS-1 and TE. The ketide produced by this bimodular synthase *in vivo* was the triketide lactone which was expected (Reproduced from Rowe et al., 2001)**.**

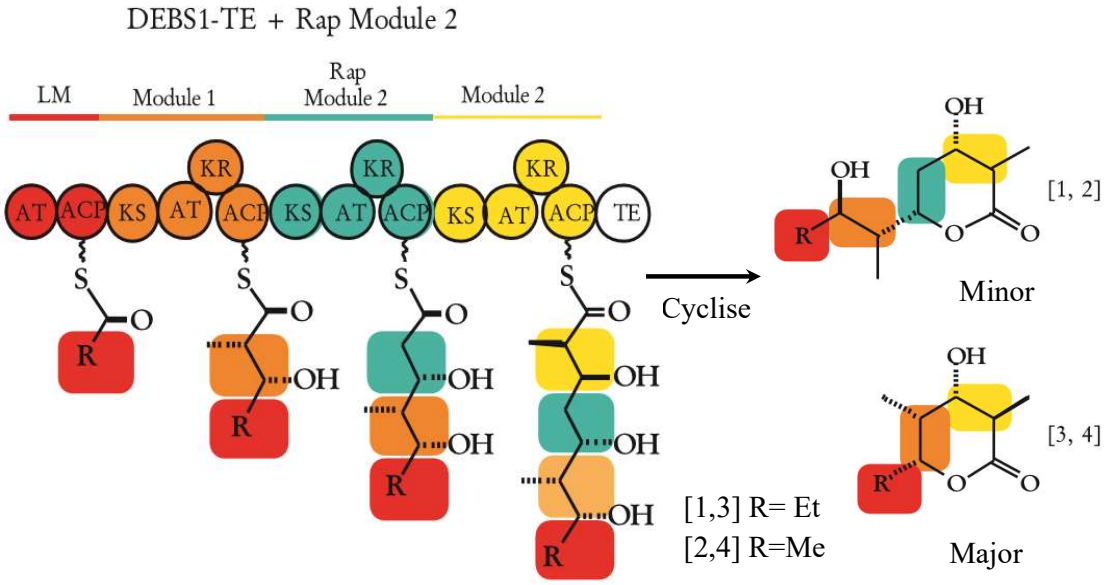


Figure 1.17 Domain organization of an engineered hybrid PKS: A trimodular synthase, obtained by insertion of module-2 of rapamycin PKS between C-terminus of module 1 and N-terminus of module 2 of DEBS-1 polypeptide. Two types of ketides were produced by this trimodular synthase *in vivo*. The expected product was minor while triketide lactone 1, which was produced by bimodular synthase shown above, was the major product. This was explained on the basis of phenomenon called 'skipping'(Reproduced from Rowe et al., 2001).

Several hybrid polyketides have been reported to be made using these approaches (Staunton and Wilkinson, 2001, Menzella et al., 2005). Identification of various compatible domains particularly KS, AT motifs and ACP domains can be exploited for engineering novel PKS molecules with novel bioactivities (Weissman, 2004, Weissman and Leadlay, 2005).

1.15 Nonribosomal peptide synthetases

Nonribosomal Peptide examples include various antibiotics tyrocidine, bactracin, vancomycin, daptomycin, cyclosporin A, penicillin anticancer drugs like bleomycin etc. NRPS has modular organization analogous to type I PKS systems or FAS in which catalytic units are folded into distinct domains that are arranged as an assembly line (Figure 1.18 and 1.19). These multimodular synthetases have well-defined initiation, elongation and termination modules. Elongation modules comprise distinct domains: a condensation (C) domain, an adenylation (A) domain and a peptidyl carrier protein (PCP or ArCP) or thiolation (T) domain which is equivalent to ACP in polyketide synthases (Figure 1.18). As in PKS or FAS systems, the condensation domain is the point of attachment of activated monomers. Similarly, other domains for methylation (M) or reduction (R domain) or epimerization (E) or cyclization (Cy), that are found in PKS, can also be present in NRPS which modify the functionality of amino acid molecules by methylation/ reduction or epimerization or cyclization respectively during the cycles of chain elongation (Ehmann et al., 1999). Domains for oxidation or hydroxylation (Silakowski et al., 1999) are also found. Owing to their modular organization as in PKS, the number and order of modules in an NRPS protein directly correspond to the sequence and number of amino acids in the resulting peptide (Figures 1.18 and 1.19) (Fischbach and Walsh, 2006).

As against to the PKS multimodular assembly lines that are dimeric, NRPS assembly lines are monomeric (Sieber et al., 2002, Hur et al., 2012).

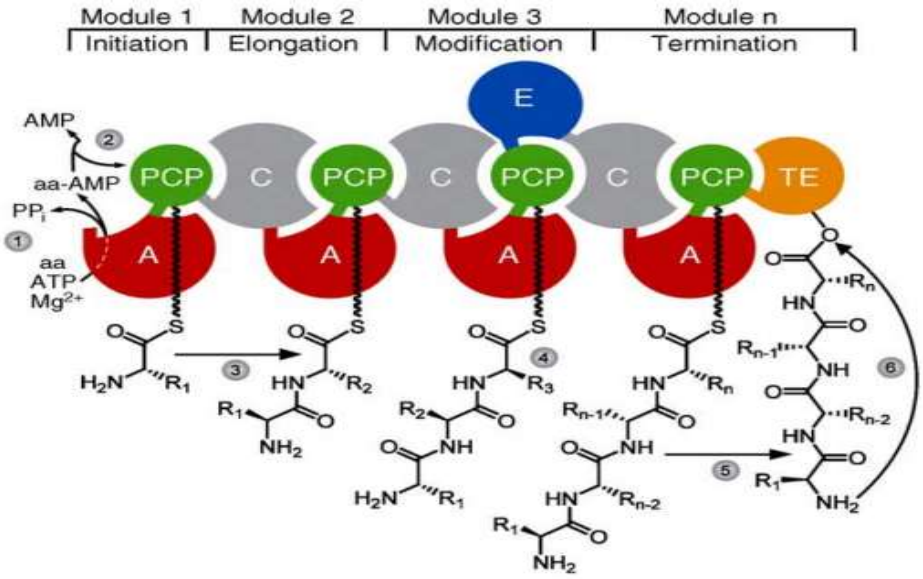


Figure 1.18 Simplified representation of modular organization and mechanism of biosynthesis by NRPS. Various NRPS domains are arranged into modules that play defined roles during the biosynthesis (Reproduced from Strieker et al., 2010).

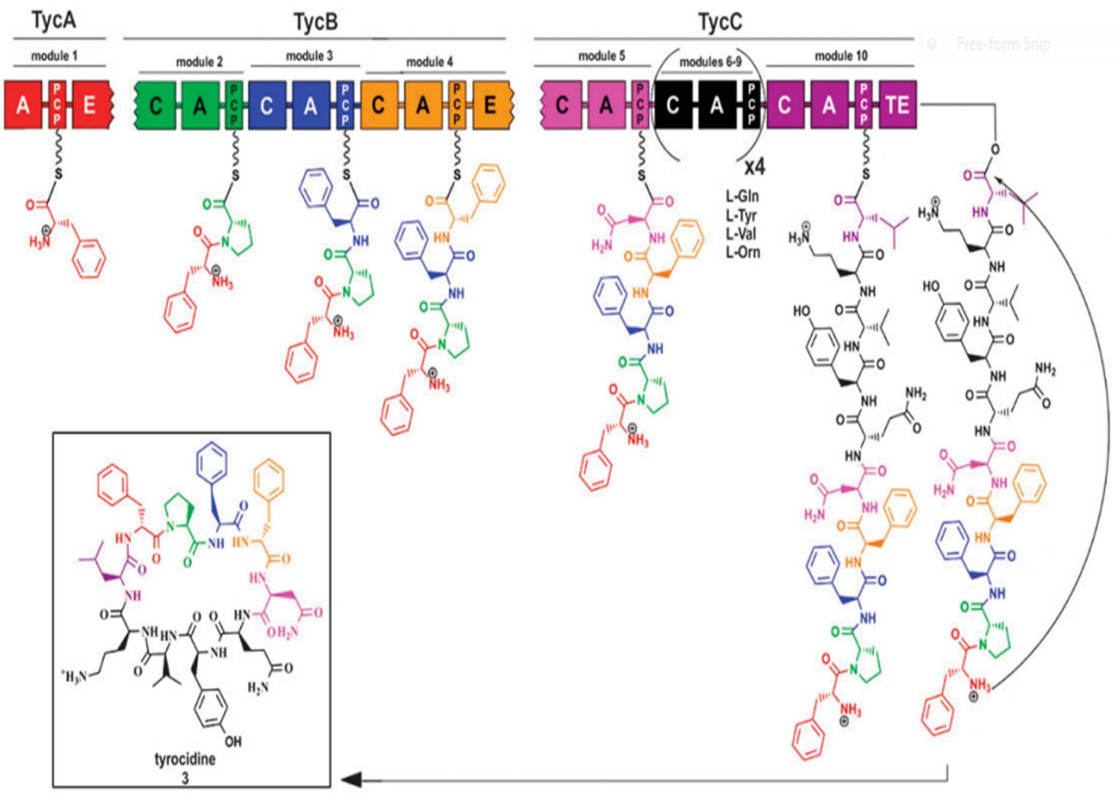
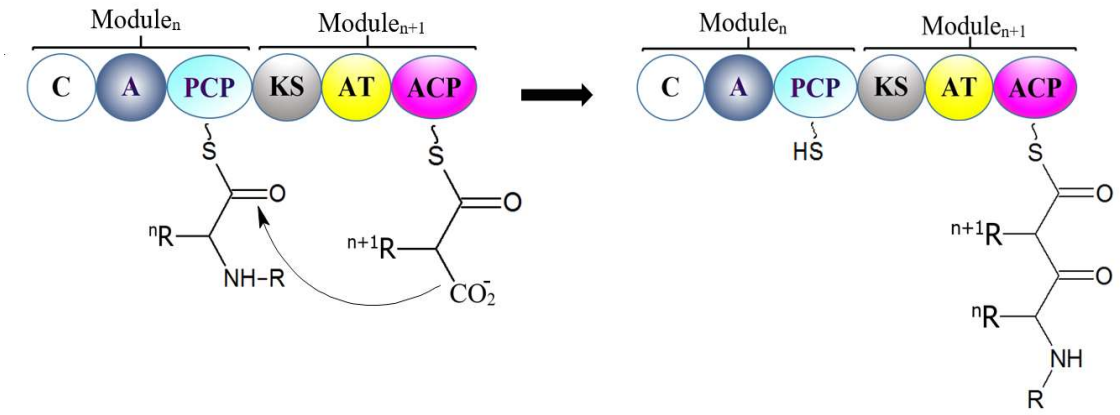


Figure 1.19 Organization of domains in type I NRPS for the biosynthesis of tyrocidine (Reproduced from Meier and Burkart, 2009).

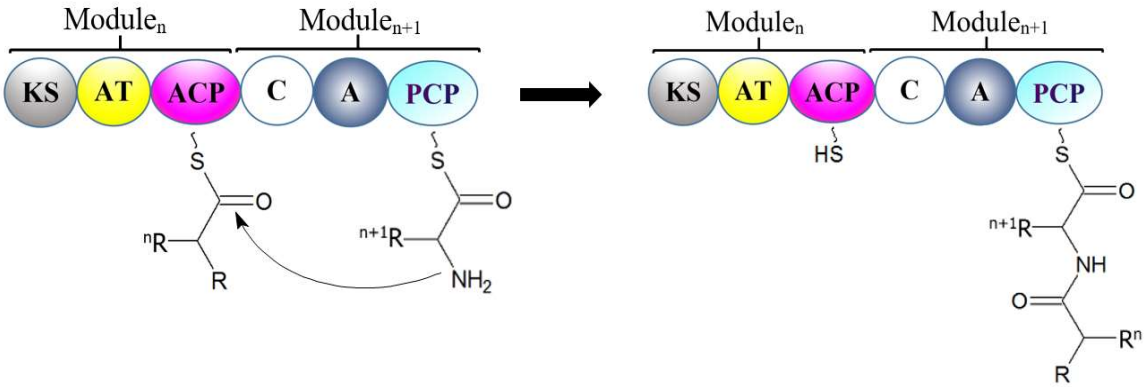
1.16 Biosynthesis of combined peptide-polyketide natural products and hybrid multimodular synthetases

The multimodular synthetases that are described above are also found in nature as hybrid fused synthetases either as PKS-NRPS or as NRPS-PKS synthetases or in various combinations, in which the polyketide synthetase part can be type I or II, and it may be iterative or non-iterative, etc (Fischbach and Walsh, 2006). Some natural products are produced by the action of both NRPS and PKS whereby a moiety biosynthesized by one synthetase is taken as substrate by the other synthetase to deliver the final product molecule. Numerous such biosynthetic clusters are also suspected to be associated with fatty acid synthetase activity like the product of *mmpB* of the *mup* cluster (Thomas et al., 2010). In hybrid multimodular synthetase systems, the PKS and NRPS components fuse together, and there is functional hybridization between modules of both types such that the metabolic intermediate synthesized by one is transferred directly onto the other synthetase for processing in assembly line without any previous conversion **(**Figure 1.20**).**

NRPS-



PKS-



PPTas

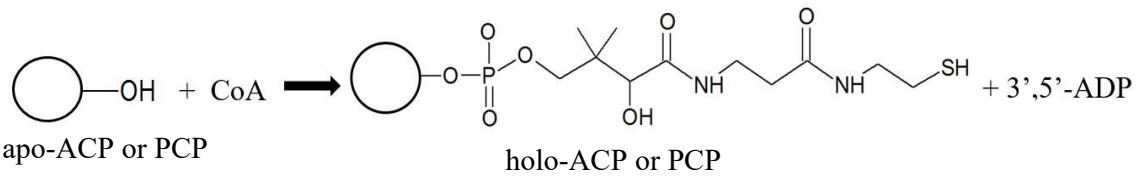


Figure 1.20 Schematic representation of modular organization of hybrid NRPS-PKS and PKS-NRPS synthases and of posttranslational modification of apo ACP or apo PCP into holo ACP or holo PCP by PPTase(Adapted from Du et al., 2001).

In the coronatine biosynthesis polyketide component, coronafacic acid and amino acid component coronamic acid are individually biosynthesized by PKS and NRPS respectively and final product is made by joining two components by a ligase. In case of

cyclosporin biosynthesis a polyketide intermediate is first converted into an amino acid and then it is incorporated into the natural product by an NRPS (Figure 1.21)

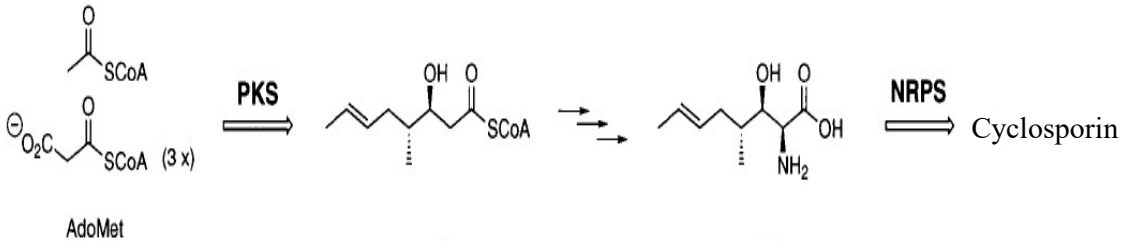


Figure 1.21 Biosynthetic pathway for cyclosporin in *T. niveum* (Reproduced from Du et al., 2001).

There exist numerous combinations of NRPS and PKS hybrid multimodular systems, for example, biosynthesis of pristinamycin by *S. pristinaespiralis* involves PKS/NRPS/PKS/NRPS hybrid system while biosynthesis of bleomycin by *S. verticillus* represents PKS/NRPS/PKS hybrid system (Figure 1.22). Rapamycin is biosynthesized by *S. hygroscopicus* on a PKS/NRPS multimodular system.

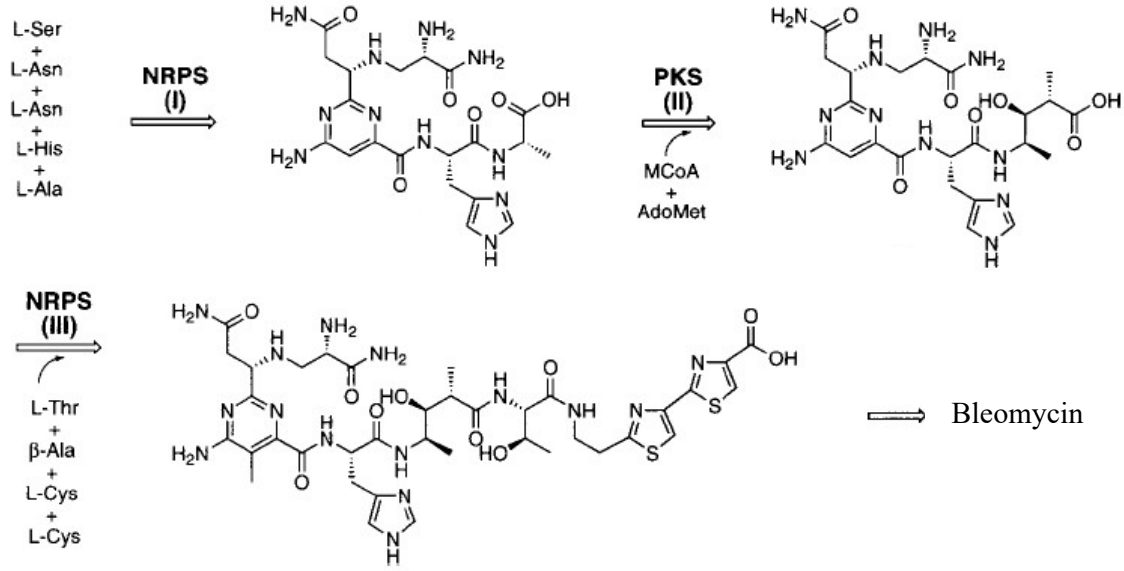


Figure 1.22 Biosynthetic pathway of bleomycin in *S. verticillus* involving a NRPS/PKS/NRPS system (Reproduced from Du et al., 2001).

Among *trans*-AT systems, thiomarinol, mupirocin and kalimantacin/batumin related gene clusters (*kal/bat)* are other mixed hybrid multimodular systems. Out of these, the thiomarinol and *kal/bat* clusters which produce antibiotics thiomarinol and kalimantacin/batumin in *Pseudoalteromonas sp.* SANK 73990 (Fukuda et al., 2011), and *Pseudomonas fluorescens* BCCM_ID9359 (Mattheus et al., 2010), respectively are PKS-NRPS hybrid multimodular systems. Mupirocin producing (*mup*) cluster in *Pseudomonas fluorescens* NCIMB10586 consists of a mix of type I and II PKSs and is suspected of including an iterative FAS (*mmpB*) (Thomas et al., 2010).

1.17 Pseudomonic acids (mupirocin)

Mupirocin is a polyketide antibiotic which was first isolated as pseudomonic acid by Fuller et al. in 1971 by the fermentation of soil bacteria *Pseudomonas fluorescens* NCIMB 10586. These bacteria are mainly found associated with plant rhizosphere and on the surface of plant leaves along with closely related other strains of *Pseudomonas sp.* (Couillerot et al., 2009). Pseudomonads belong to Gamma subclass of Proteobacteria; order Pseudomonadales, and family Pseudomonadaceae (Palleroni, 1984, Palleroni, 2008). They show remarkable ecological and metabolic diversity. Their existence is widespread in nature where they are found in soil, water, plant surfaces, organic matter, animals, insects, and humans (Gross and Loper, 2009, Palleroni, 2010). *Pseudomonas spp.* are Gram-negative rod shaped which have diameter of 1 μm and length 1.5 to 5 μm. They are motile as they have one or more polar flagella and are non-sporulating (Couillerot et al., 2009, Palleroni, 2010). These are catalase positive aerobic bacteria though some species can utilize nitrate as electron acceptor (Couillerot et al., 2009). The GC content of different pseudomonads ranges from 58-69 % (Palleroni, 2008, Silby et al., 2011) falling within this range (averaging 61%) is that of *Pseudomonas fluorescens* NCIMB 10586 (El-Sayed et al., 2003). Pseudomonads can utilize

an array of substrates as source of energy and produce a range of secondary metabolites (Palleroni, 2010). This includes a number of bioactive and antimicrobial compounds that are known to positively impact plant growth by controlling phytopathogens (Haas and Defago, 2005, Couillerot et al., 2009). All fluorescent pseudomonads including *Pseudomonas fluorescens* produce fluorescent compounds particularly pyoverdins (Palleroni, 2010). Mupirocin produced by *Pseudomonas fluorescens* NCIMB 10586 is produced as a mixture of four types of pseudomonic acids (A, B, C and D) (Fuller et al., 1971). At least two other strains of *pseudomonas* isolated from sediment ground water have also been reported to produce pseudomonic acid A (Fritz et al., 2009).

1.17.1 Structure and composition

All the pseudomonic acids share a common structure in which a C-17 unsaturated polyketide moiety called monic Acid, that contains a pyran ring, is esterified to a C-9 saturated fatty acid, 9-hydroxynonanoic acid (9-HN) (Chain and Mellows, 1977b, Chain and Mellows, 1977a). The difference among different pseudomonic acid types is because of the way various carbons of this basic structure are modified in each of them **(**Figure 1.23**).** Pseudomonic Acid A is the primary natural product (90%) in the mupirocin mixture while pseudomonic acid B which has an extra hydroxyl group at C-8 makes up 8% of the mupirocin mixture. PA-C that has a double bond instead of epoxide group at C 10-11 accounts only 2%. The minor component is (<2%) is PA-D that has double bond at C'4 and 5' (Thomas et al., 2010).

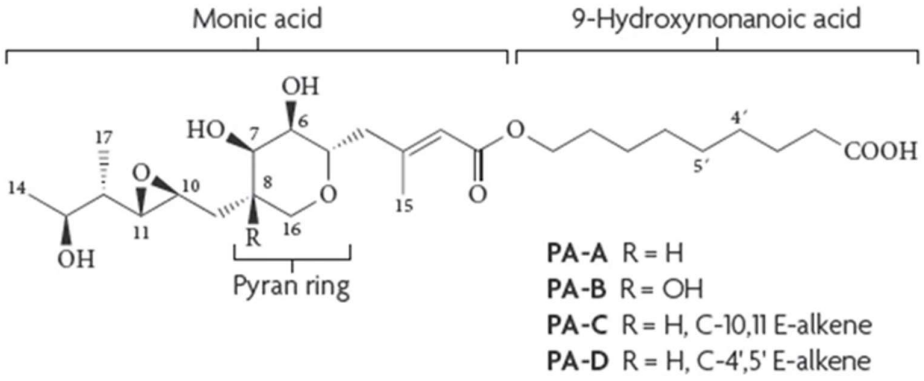


Figure 1.23 Structures of pseudomonic acid (PA) A, B, C and D(Reproduced from Thomas et al., 2010)**.**

1.17.2 Antimicrobial activity

Mupirocin has activity against Gram-positive bacteria as well as against a few Gram-negative bacteria. It is primarily used against *Staphylococcus aureus* and specifically against methicillin-resistant *S. aureus* (MRSA) (El-Sayed et al., 2003, McConeghy et al., 2009, Thomas et al., 2010). Because of the side effects it induces in the tissues in which it is introduced, mupirocin is used topically only against bacterial skin infections and to clear nasal passage and as a pre-operative scrub for medical staff (Thomas et al., 2010). It is effective under the conditions when other antibiotics like vancomycin and ciprofloxacin fail to bring any effect (Ha et al., 2008). Several factors limit the use of mupirocin apart from its toxicity to the tissues in which it is introduced. Firstly, it gets quickly deactivated once it is in the body fluids because of the hydrolysis of the ester bond between its Monic acid and 9-HN moiety. Secondly, it strongly binds to serum proteins thereby reducing its bioavailability. Finally, its biological activity is lower at higher pH (Thomas et al., 2010). It is for these factors and despite the fact that it is proved to be many-many times more potent on bacterial IleRS than on the corresponding eukaryotic enzymes - a property that makes it a drug of

choice, mupirocin cannot be used systemically (Nakama et al., 2001). Additionally, there is the problem of increasing mupirocin resistance and of which high level mupirocin resistance is a cause for concern as the same can be acquired through horizontal gene transfer, may be through plasmid carriage (Thomas et al., 2010). Because of the limitations in the use of mupirocin, there is need to find ways to produce alternative compounds that may be having the novel or improved activities. These compounds may be derived from mupirocin, thus having similar or altered structures but without the limitations put on their use like mupirocin.

1.17.3 Mechanism of action

Mupirocin targets bacterial isoleucyl-tRNA synthetase enzyme (IleRs) to which it competitively and reversibly binds to inhibit protein synthesis (Hughes and Mellows, 1978). IleRS catalyses the transfer of amino acid isoleucine by charging it so that it can be joined to its cognate tRNA prior to protein synthesis. Structural examination of IleRS with and without mupirocin bound to it, revealed that the 14 methyl end of the pseudomonic acid mimics side-chain of Isoleucine (Nakama et al., 2001), while the region of C1-to C-3 of mupirocin molecule and its pyran ring occupy the ATP binding groove of IleRS as it mimics the structure of adenine and ribose (Yanagisawa et al., 1994). The 9-hydroxylnonanoic acid portion of the mupirocin molecule stabilizes this complex by fitting into the hydrophobic groove of the complex of IleRS with bound mupirocin (Brown et al., 2000).

1.17.4 Self-Resistance to mupirocin

Two types of self-resistance have been reported against mupirocin by *Pseudomonas fluorescens* NCIMB 10586, low and high (resistance to more than 512 $\mu g/ml^{-1}$ of mupirocin). These are related to the presence of two types of IleRS (IleRS-R1 and IleRS-R2) encoded by two separate *ileS* genes (Yanagisawa and Kawakami, 2003). Low level of resistance is due to

the presence of a typical bacterial IleRS. The high-level resistance is attributed to the presence of a eukaryotic-like IleRS which is the product of *mupM* present in the tailoring region of the mupirocin biosynthesis cluster of *P. fluorescens* (El-Sayed et al., 2003). The product of *mupM* gene, the eukaroyotic like IleRS, shares similarity with the other *ileS* reported from other bacterial species that are naturally resistant to mupirocin, for example the product of *mupA* of *S. aureus* strains which is present on its plasmids that are acquired by horizontal gene transfer.

1.17.5 Mupirocin biosynthetic cluster

The mupirocin gene cluster, which is a *trans*-AT PKS cluster of about 75 kb, was first identified by transposon mutagenesis (Whatling et al., 1995, Thomas et al., 2010). Later it was analysed by DNA sequencing, specific gene knockouts and complementation studies (El-Sayed et al., 2003). It is found to encode 35 open reading frames (ORFs) which could be classified into two parts. The first part comprises of 3 larger multifunctional proteins while the second group includes another three smaller multifunctional proteins along with 29 individual genes called 'tailoring genes'. The multifunctional genes of the cluster are designated as mupirocin multifunctional polypeptides (*mmp*A *to mmp*F) (El-Sayed et al., 2003) five of which are PKS while *mmpB* is a fatty acid synthase (El-Sayed et al., 2003, Gao et al., 2014). Some of these genes show similarity with type II PKS systems indicating that the mupirocin PKS system is a mix of both type I and type II. While individual genes are designated as *mup*A-X and *macp*A-E (El-Sayed et al., 2003). The putative functions of various tailoring region genes as well as different domains present in multifunctional proteins are listed below.

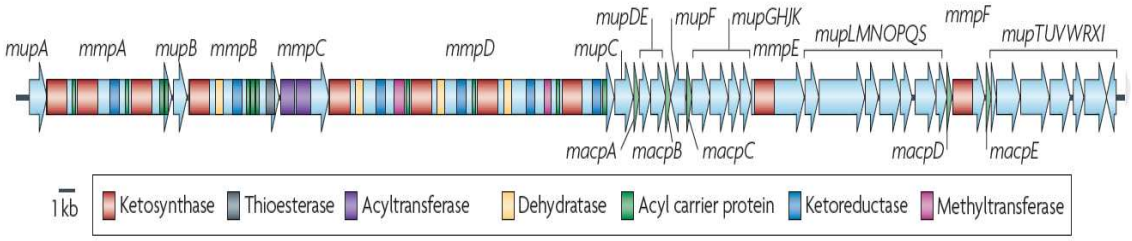


Figure 1.24 The organization of mupirocin biosynthetic cluster
(Adapted from Thomas, et al., 2010).

Table 1.3 List of putative gene functions of mupirocin biosynthetic cluster
(El-Sayed et al., 2003, Hothersall et al., 2007, Thomas et al., 2010)

ORF	Proposed function
mupA	Reduced flavin mononucleotide (FMNH2) oxygenase
mmpA	PKS: load/transfer KS, ACP; module 5 KS, KR, ACP; module 6 KS, ACP, ACP
mupB	3-Oxo-ACP synthase
mmpB	PKS: KS, DH, KR, ACP, ACP, ACP, TE
mmpC	Dual AT and putative ER
mmpD	PKS: module 1 KS, DH, KR, MT, ACP; module 2 KS, DH, KR, ACP; module 3 KS, DH, KR, MT, ACP; module 4 KS, KR, ACP
mupC	Dienoyl-CoA reductase
macpA	ACP
mupD	3-Oxo-ACP synthase
mupE	ER
macpB	ACP
mupF	KR
macpC	ACP
mupG	3-Oxo-ACP synthase
mupH	β-hydroxyl-β-methyl glutarate (HMG) CoA synthase

mupJ	Enoyl CoA hydratase
mupK	Enoyl CoA hydratase
mmpE	PKS: KS, hydroxylase
mupL	Hydrolase
mupM	Isoleucyl-tRNA synthetase
mupN	Phosphopantetheinyl transferase
mupO	Cytochrome P450
mupP	Unknown
mupQ	Acyl CoA synthase
mupS	3-Oxo-ACP reductase
macpD	ACP
mmpF	PKS: KS
macpE	ACP
mupT	Ferrodoxin dioxygenase
mupU	Acyl CoA synthase
mupV	Oxidoreductase
mupW	Dioxygenase
mupR	Transcriptional activator
mupX	Amidase
mupI	N-Acyl homoserine lactone synthase

Abbreviations: *mmp*, mupirocin multifunctional polypeptide gene; PKS, polyketide synthase; KS, ketosynthase; DH, dehydratase; KR, ketoreductase; ACP, acyl carrier protein; TE, thioesterase; AT, acyltransferase; ER, enoyl reductase; MT, methyltransferase; *macp,* mupirocin acyl carrier protein gene; ORF, open reading frame.

1.17.6

1.17.6 Mupirocin biosynthetic pathway

It has been pointed out that the mupirocin biosynthetic pathway does not follow the principle of colinearity which is that the order of genes in the cluster does not match with the biosynthetic steps as reported in many other PKS type I systems like erythromycin (Donadio and Katz, 1992). Thus, it has neither separate loading domain for loading of activated starter unit covalently on the first KS domain, nor a cognate AT domain for sole use of any of its modules. Apart from these peculiarities, it also has some other unique features like tandemly repeated ACPs in its modules and the presence of ER domains/activity in *trans.* The biosynthesis of both the components of pseudomonic acid i.e. monic acid and 9-hydroxynonanoic is now thought to take place together, which explains why transposon mutation or targeted mutagenesis to abolish production of either of the two components failed (El-Sayed et al., 2003, Cooper et al., 2005b, Hothersall et al., 2007). Mutational and re-feeding studies done by Gao et al. (2014) showed that the biosynthesis of mupirocin by the *mup* cluster involves two biosynthetic pathways running in parallel at the same time. Both the pathways are proposed to start from the same precursor, but branches at the point of action of MmpEOR whose role in epoxidation has been established. Out of these two pathways, the major pathway, all of the products of which are epoxide derived and in which pseudomonic acid B is an intermediate, finally results in the production of pseudomonic acid A as the main product while the minor pathway results in the biosynthesis of pseudomonic acid C-an alkene derivative which is produced in a minor amount. It has been shown that epoxidation, tetrahydropyran (THP) ring formation and 6-hydroxylation are essential for downstream processing by the biosynthetic assembly line (Gao et al., 2014). The gene function responsible for 6-hydroxylation is not known yet. It has also been shown that once the fatty acid side chain biosynthesis is underway, the formation of THP ring is not possible and that epoxidation can occur before THP ring formation is catalysed by *mupT* and *mupW*

(Cooper et al., 2005a, Gao et al., 2014). Also, the modification of the pyran ring by tailoring genes occurs after biosynthesis and esterification of 9-HN on monic acid (Gao et al., 2014).

The biosynthesis of mupirocin by *mup* cluster represents an assembly line process as it involves orderly and coordinated action of various enzymes represented in the assembly line that are coded by the genes in this cluster. Mupirocin is produced by the action of different enzymatic units represented in the *mup* assembly line on the same building blocks (starter unit acetyl-CoA and extender unit malonyl-CoA) in a set pattern as described in section (1.17.7-1.17.9). During the biosynthesis some of these enzymatic units act iteratively and some non-iteratively to complete a cycle for the production of mupirocin molecule. MmpD and MmpA multifunctional proteins act non-iteratively to produce monic acid C-14 precursor while others (MmpB) act iteratively to elaborate fatty acid component of the mupirocin molecule. During biosynthesis some of the enzymatic functions (AT and ER) are provided, in *trans* by *mmpC*, that are common to all the steps in the biosynthetic pathway during chain elongation. Role of MupE (alone or together with MupD) is also predicted in the biosynthesis of mupirocin as *trans* acting enoyl reductase that also commonly affects various steps during biosynthesis (Hothersall et al, 2007, Gao et al, 2014). None of the modules of *mup* cluster has domains with enoyl reductase activity. By mutational and knock out studies it has been shown that the biosynthesis of mupirocin is an integrated process because mutation in most of the several components of the *mup* cluster either resulted in complete loss of production of pseudomonic acids or drastically reduced their production (El-Sayed et al., 2003, Hothersall et al., 2007, Thomas et al., 2010). This mutational analysis of *mup* cluster also revealed "leaky hosepipe" behaviour of the biosynthetic pathway further supporting the integrated nature of the *mup* assembly line (Wu et al., 2008). This name has been adopted to describe the behaviour of *mup* biosynthetic pathway whereby mutation in many of the several gene functions of the pathway (like any of the genes of HCS cassette or any ACP of *mmpA,* that of

mupB, mupL, mupQ and *mupS,* as well as point mutations in *mmpB* (S1390A) and *mmpF* (C183A), resulted in essentially identical phenotype which was identified by complete loss of mupirocin production and accumulation of same intermediates (mupiric acid and/or mupirocin H) (Wu et al., 2008, Cooper et al., 2005a, Hothersall et al., 2007). This has been described due to the existence of labile points in the *mup* biosynthetic pathway. It has been hypothesized that these labile points, from which preferential and spontaneous release of biosynthetic intermediates occurs, come into play whenever any mutation blocked the passage of biosynthetic intermediates through the *mup* assembly line. These labile points are predicted to be independent of the site of blockage in the *mup* biosynthetic pathway which may be close to them or located distantly (Wu et al., 2008).

1.17.7 Monic acid backbone biosynthesis

It has been proposed that C-17 ketide backbone of monic acid is synthesised by six condensation reactions of acetate derived units which are catalysed by MmpD and MmpA that are predicted to have three and four modules respectively (Figure 1.25) (El-Sayed et al., 2003). The synthesis begins at MmpD as was evidenced by the isolation of mupiric acid as a result of active site mutation in ketoreductase domain of its putative module four (Thomas et al., 2010). For the biosynthesis to begin, the *trans* acting AT domains, which are part of the MmpC module, load an activated starter unit (acetyl-CoA intermediate) onto the 4'-phosphopantethein arm of ACP of the first module of MmpD. An activated extender unit like malonyl-CoA is then transferred by these acyl transferases to the ACP of first module of MmpD. The active site of which becomes vacant in the meantime by transferring starter unit onto thiol group of cysteine present at the active site of KS domain of the same module of MmpD, which catalyses first decarboxylative (Claisen) condensation leading to chain formation. After that, the action of KR and MT domains causes ketoreduction and α-

methylation, respectively. The same process of extension and modifications of carbon backbone is repeated by all the remaining three modules of MmpD after which the elongated chain is transferred to MmpA module. The first module of MmpA is non-elongating because of the presence of KS^{o} domain which does not have an active site *his* residue that is critically required by all the elongating KSs to perform their function. This non-elongating domain plays an active role in biosynthesis which is evidenced by complete abolition of mupirocin production on mutating its active site or upon its complete deletion (El-Sayed et al., 2003). It has been speculated that this module may have a role in transferring elongated intermediate from MmpD to MmpA (Thomas, et al. 2010). Two further rounds of chain elongation by the remaining two elongating modules of MmpA on the intermediate that is transferred to it from MmpD finally result in the heptaketide precursor of monic acid. Figure 1.25 below shows this scheme that is predicted for the biosynthesis of mupirocin by *mup* cluster.

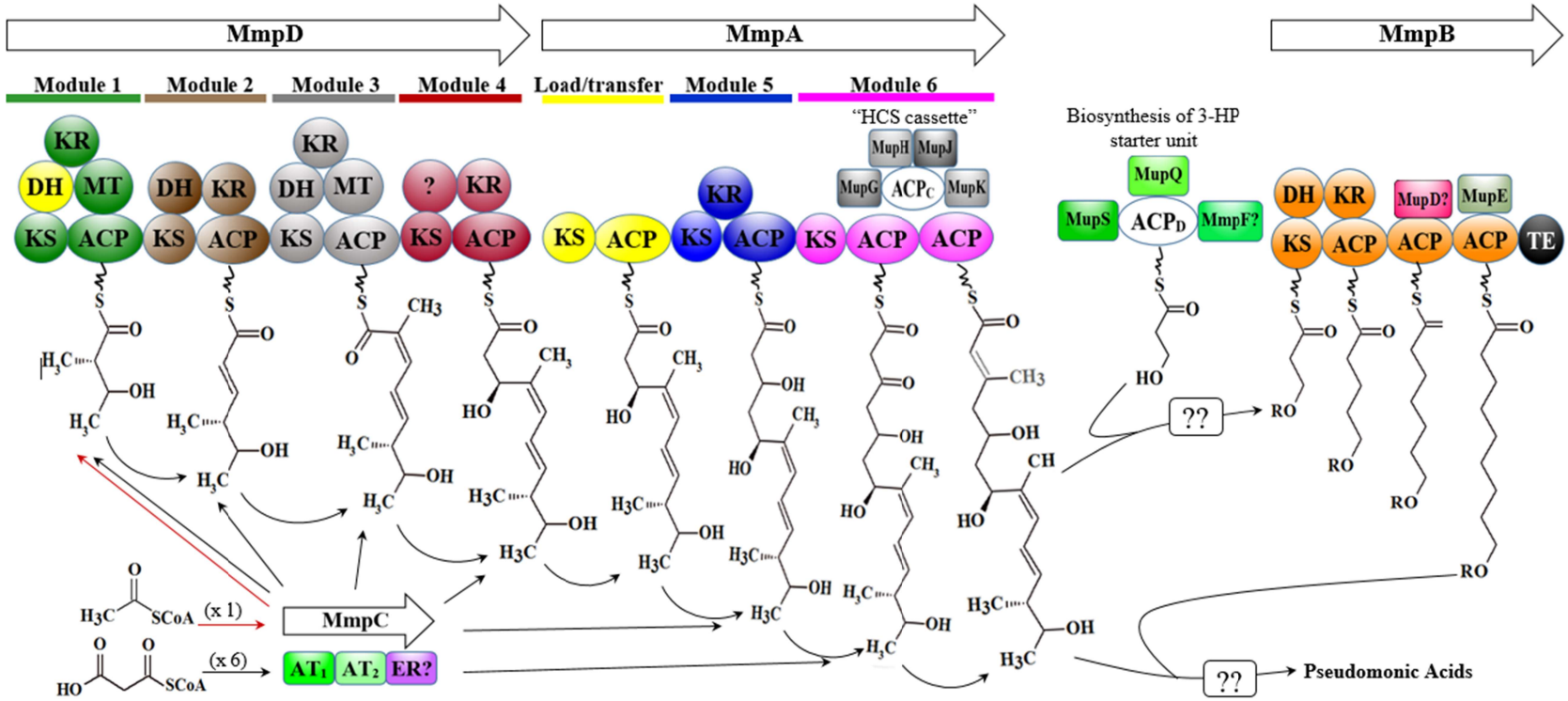


Figure 1.25 Scheme predicted for pseudomonic acid biosynthetic pathway. Biosynthesis of monic acid, the polyketide backbone of pseudomonic acids is proposed to occur by the polypeptides (comprising six modules) encoded by the *mup* cluster in the order shown using acetate as starter unit and malonate as extender units. Biosynthesis of 9-HN-the fatty acid component-is proposed to be from malonate by three condensations catalysed by MmpB module, iteratively, using 3-hydroxy propionate (3-HP) as starter unit during which MupE (probably along with MupD) is proposed to be involved in the reduction of the extended moiety. The MacpC and MacpD proteins are shown where they are predicted to act during the biosynthesis. It is still not clear whether monic acid and 9-HN moieties are elaborated separately and then they are joined together by esterification to give pseudomonic acids (scheme in which R=H) or 9-HN moiety is elaborated while already esterified to monic acid (scheme in which R=Monic Acid). Domains shown in yellow are not involved in chain extension or modification while those shown in boxes involve activities that act *in trans* during the biosynthesis. KS, ketosynthase; AT, acyl transferase; DH, dehydratase; ER, enoyl reductase; KR, ketoreductase; ACP, acyl carrier protein and TE, thioesterase domain, Mmp, mupirocin multifunctional polypeptide; MT, methyltransferase, MA, Monic Acid (Adapted from Gurney and Thomas, 2011) .

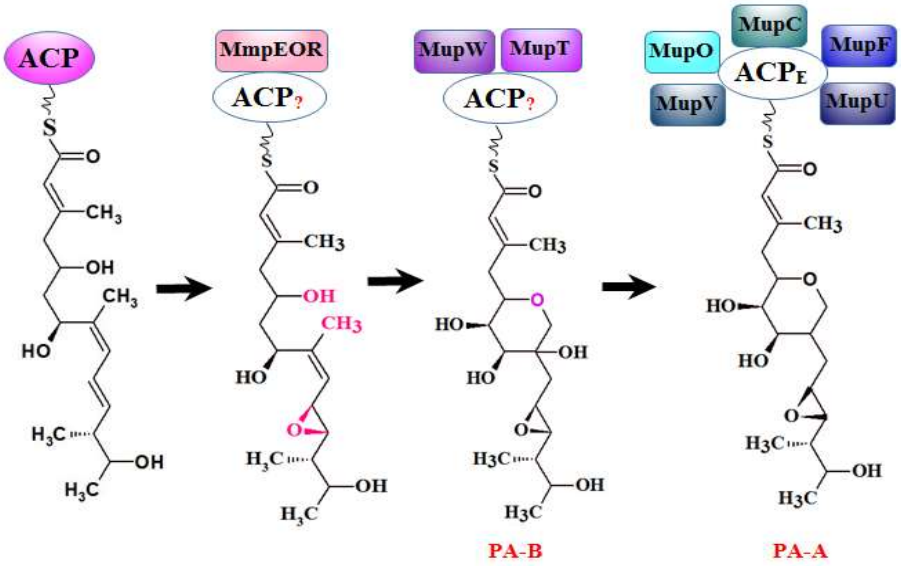


Figure 1.26 Modifications of monic acid moiety during the biosynthesis of pseudomonic acids. The Monic acid structure is modified in the order shown by the action of several enzymes of the *mup* cluster that act in *trans* during the biosynthesis. After the addition of C-15 methyl group by HCS cassette, epoxide group is added at C-10,11 by the action of oxidoreductase domain of MmpE in the intermediates that are processed to become PA-B and finally PA-A. Pyran ring formation is catalysed by MupW and MupT on epoxidised and unepoxidised intermediates. The site of these reactions (concerned ACPs) is not known yet. Thereafter, pyran ring is modified by the action of MupO, U, V, C and F when the moiety is tethered to ACP_E. Unepoxidised intermediates yield PA-C. MmpEOR, mupirocin multifunctional protein oxidoreductase domain; ACP, acyl carrier protein (Adapted from Gurney and Thomas, 2011).

1.17.8 9-Hydroxynonanoic biosynthesis (9-HN)

To date it has been very difficult to predict the source of biosynthesis of 9-HN which is fully saturated nine carbon fatty acid chain (Thomas et al., 2010). Since direct evidence has implicated *mmpD* and *mmpA* in monic acid backbone synthesis, it must be one of the remaining modules that is involved in 9-HN biosynthesis. The MmpB module which could not be characterized either as PKS or FAS and which has unusual structure in that it has just one module and three ACPs, is proposed to be involved in 9-HN biosynthesis by an iterative

mechanism although there is no direct published evidence to support this (however, unpublished data about involvement of *tmlU*-induced truncation of the 9-HN moiety and mutations of the TE domain do support this). Feeding experiments of labelled precursors are consistent with biosynthesis of 9-HN coming from 3-hydroxypropionate (3-HP) which primes its synthesis by *mmpB*. Circumstantial evidence indicates that products of *macpD*, *mupS* and *mupQ* are involved in the biosynthesis of 3-hydroxypropionate (Thomas et al., 2010). With 3-HP as starter unit three rounds of condensation of with malonate extender unit should result in the synthesis of 9-HN backbone. However, since this module does not have an ER domain to account for saturated fatty acid chain it has been proposed that MupE, which shows sequence similarity with ER domains, is involved in enoyl reduction (Hothersall et al., 2007). Mutation in MupE resulted in the incorporation of a double bond in the fatty acid chain of 9-HN between $C_{6'}$-$C_{7'}$ in pseudomonic acid A (Hothersall et al., 2007). However, the fact that this did not result in complete unsaturation of 9-HN indicated that this enoyl reductase along with other enzymes is responsible for the reduction (Gao et al., 2014). The third domain in MmpC is also predicted to have ER activity which may also be involved in the reduction of 9-HN. The terminal TE domain is proposed to catalyse release of the saturated 9-HN moiety (El-Sayed et al., 2003, Hothersall et al., 2007).

1.17.9 Tailoring of monic acid backbone (biosynthesis of active mupirocin)

Monic acid in the pseudomonic acid backbone (i.e. the monic acid esterified with 9-HN) needs to be worked upon by products of several genes of the mup cluster, each having different individual activities in order to make the biologically active mupirocin molecule. Mutagenesis has showed that all 26 tailoring ORFs are required for normal mupirocin production (Hothersall et al., 2007). These enzymes modify the PKS-bound intermediate in many ways, hence, they are called tailoring enzymes. Various modifications like

hydroxylation at C-6, epoxidation at C-10,11, incorporation of a methyl group at C-15, oxidation of C-16 to form tetrahydropyran ring (THP) are carried out by tailoring enzymes (Cooper et al., 2005a, Cooper et al., 2005b, Hothersall et al., 2007). Bioinformatic and mutational analyses have indicated that in the biosynthesis of pseudomonic acids, several tailoring genes work together in groups. Product analysis of various gene knock-outs has shown that sets of genes are involved in the particular type of modification of mupirocin structure (Hothersall et al., 2007, Gao et al., 2014). One such group hydroxymethylglutaryl-CoA synthase (HCS) cassette that is revealed, which includes genes *macpC, mupG, mupH mupJ, mupK* that are located in *mup* cluster next to each other, is implicated in β-methyl branching in the molecule (Figure 1.25). The other group of genes *macpD, mupS, mupQ, mmpF* that are proposed to act together in a block, which are also located next to each other, is implicated 9-HN biosynthesis (Figure 1.25). Similarly, another group of genes, *mupD* and *mupE,* that are located next to each other in *mup* cluster, is implicated in providing enoyl reductase activity that is predicted to result in the reduced 9-HN (Gao et al., 2014). Mutation in *mupW* resulted in the biosynthesis of mupirocin derivative without THP ring named as mupirocin W, which was also produced when *mupT* gene was knocked out indicating the two genes work together in the oxidation of C-16 to form pyran ring (Figure 1.26) (Cooper et al., 2005a). Knockouts of either of *macpE/ mupO/ mupU/ mupV* resulted in the biosynthesis of pseudomonic acid B only and at the same time entirely abolished the production of pseudomonic acid A, indicating that products of these genes work together in reducing C8-C9 double bond (Hothersall et al., 2007). Similarly, deletion of *mupC* and *mupF* resulted in novel metabolites that were named mupirocin C and mupirocin F, respectively (Hothersall et al., 2007, Gao et al., 2014) (Figure 1.28). From the structures of these compounds it was proposed that *mup*F encodes a ketoreductase and is involved in the reduction of 7 keto group while *mup*C, which is predicted to be enone reductase, is involved in the reduction of 8'-9'

bond. By double knockout studies it has been established that MupW acts early before the action of MupO, MupU, MupV and MupC and that the 7-hydroxyl group is required for the action of mupW/T and MmpEOR (Figure 1.28). Double knock-out study has also confirmed that MupF acts only after MupC in the biosynthetic pathway (Gao et al., 2014). Deletion of *mup*H resulted in the isolation of a novel metabolite mupirocin H, the structure of which pointed towards its action early in the mupirocin biosynthetic pathway (Figure 1.27)(Wu et al., 2007). Apart from these, there is no information about the point of action of products of the other tailoring genes in the mupirocin biosynthetic pathway. All mutant strains of either *mupG/mupJ* or *mupK* produced mupirocin H as a major product (Wu et al., 2007). Products of all these enzymes together with MupH are implicated in the introduction of the C-15 methyl group in pseudomonic-A biosynthesis as pointed out above.

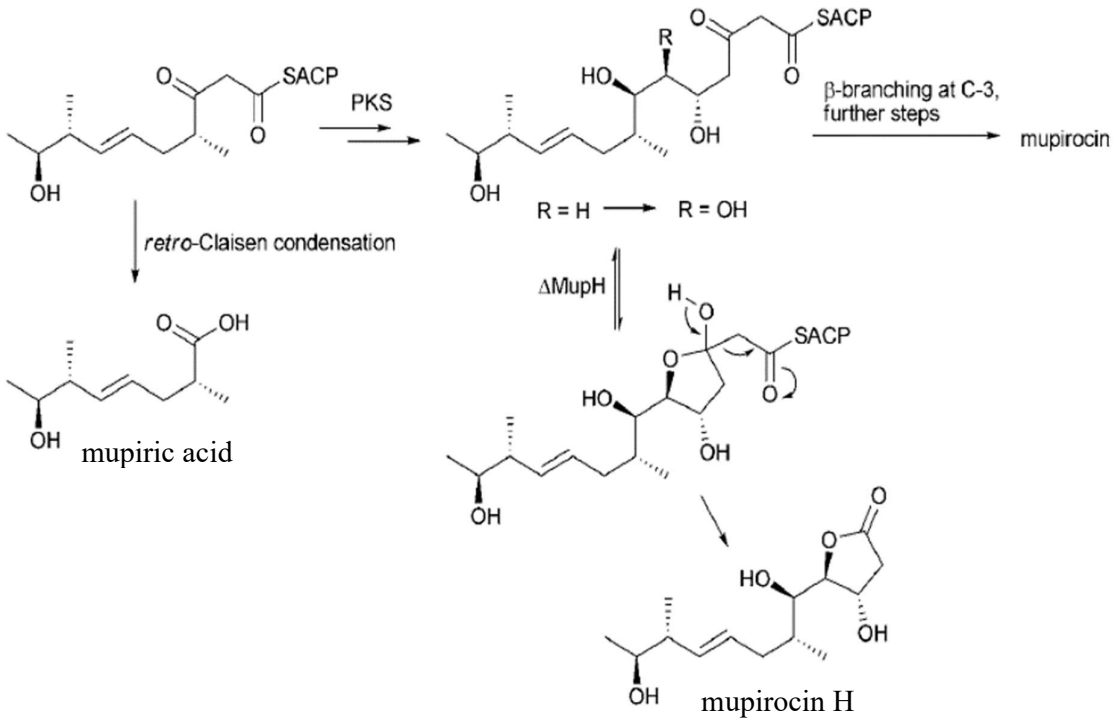


Figure 1.27 Mechanisms proposed for the biosynthesis of mupirocin H and mupiric acid(Reproduced from Piel, 2010).

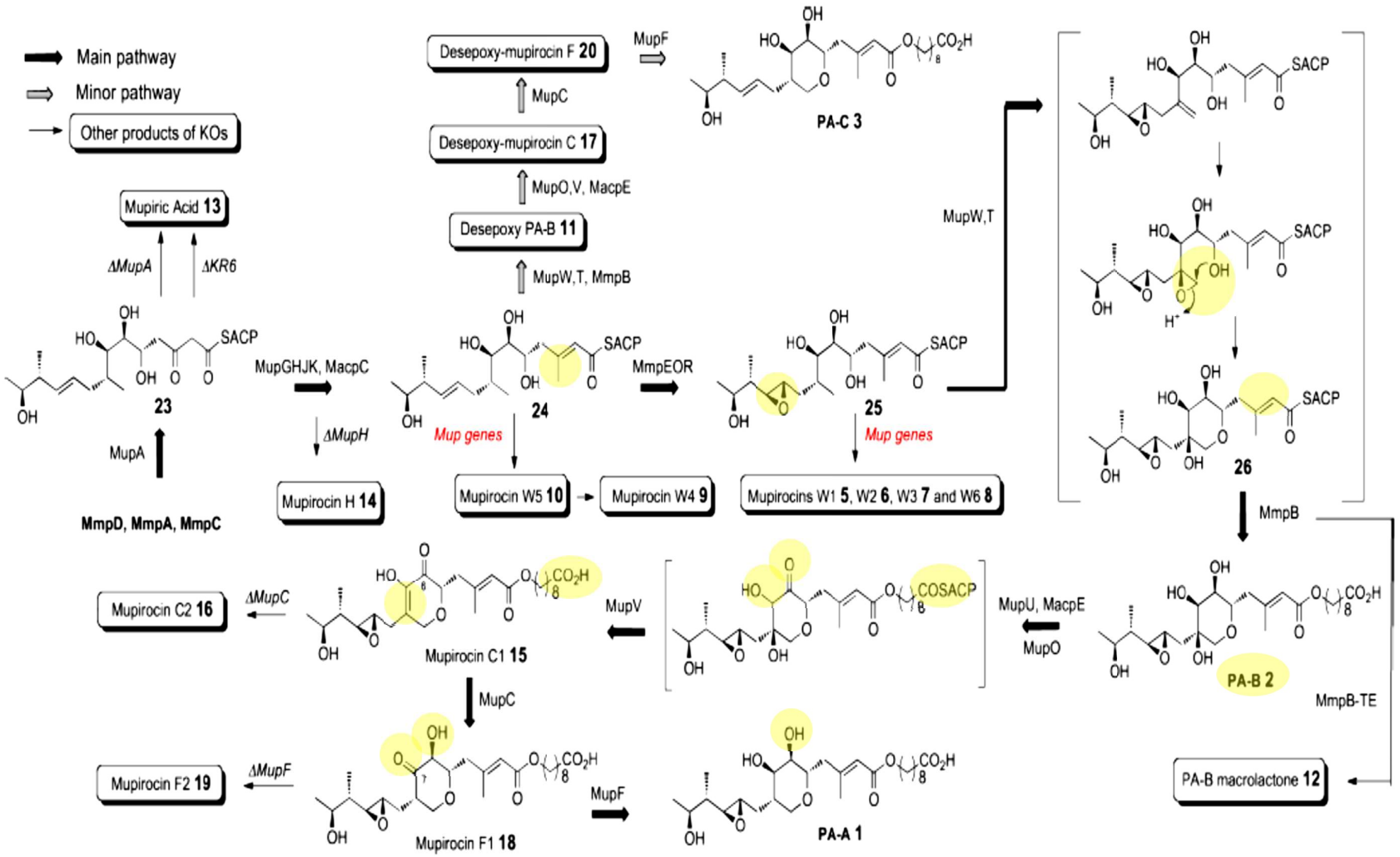


Figure 1.28 Overall scheme for the biosynthesis of mupirocin showing orderly action of some of the key tailoring enzymes on various biosynthetic intermediates that lead to the production of different pseudomonic acids.(Reproduced from Gao et al., 2014).

1.17.10 Regulation of the *mup* cluster

Quorum sensing is used to control the expression of genes of the *mup* cluster (El-Sayed et al., 2001). In quorum sensing bacterial populations communicate with each other through a diffusible signal molecule which is called an auto-inducer. There is constitutive expression of autoinducer by bacteria and with the increase in bacterial population; there is a consequent increase in the concentration/level of auto-inducer. At a certain population density of bacteria which is called a quorum, the concentration of autoinducer becomes such that it induces the target gene expression resulting in consequent target action. In the mupirocin cluster, *mup*R and *mup*I are found to be involved in quorum regulation of cluster. MupR is the transcriptional activator which activates *mup* promoter while MupI is an inducer that responds to it (El-Sayed et al., 2003). This way transcription of *mup* operon is activated upon binding to *lux* box promoter regions of *mup*F, *macp*C and *mup*A. *mup*I is reported not to have *lux* box region (El-Sayed et al., 2001).

1.17.11 Characteristic features of *mup* cluster

Mupirocin producing (*mup*) cluster has some of the unique characteristic features which distinguishes it from PKS system that encodes erythromycin antibiotic-a prototype of PKS systems. The following sections list the distinguishing features of the *mup* system.

1.17.11.1 Mix type hybrid PKS

As discussed above *mup* cluster does not comprise just type I PKS like DEBS PKS system. Instead, it is a combination of type I non-iterative and type II iterative PKS which is also suspected to include iterative FAS system (Thomas et al., 2010).

1.17.11.2 The *trans*-AT system

Unlike DEBS PKS system whereby each module of the PKS has got its own cognate AT domains, in *mup* cluster function of AT for each module is done by standalone ATs that are part of MmpC. A similar arrangement is found in kirramycin and Oxazolomycins antibiotic clusters (Piel, 2010, Musiol and Weber, 2012). As a consequence of this in the *mup* system, each module does not have a choice of selecting particular extender unit for incorporation in the growing polyketide chain, a liberty that DEBS system may have.

1.17.11.3 Lack of integrated ER domains

There are no well-defined integrated ER domains as found in DEBS PKS system. ER activity that is known, is provided *in trans* by one of the tailoring region genes *mupE* and may be by *mupD* as well (El-Sayed et al., 2003, Hothersall et al., 2007). Like other *trans* AT systems ER activity is also suspected to be provided by other proteins (Weber et al., 2008), for which MmpC of the *mup* system is under investigation.

1.17.11.4 Methyl group incorporation

During biosynthesis methyl groups are incorporated in the mupirocin molecules in both the positions α as well as β. While α groups are incorporated owing to the presence of *cis* acting MT domains in module one and three, methyl group in β position is inserted by a group of *trans* acting auxiliary genes described above which functions together as a cassette known as HCS cassette (Thomas et al., 2010). The α methyl groups are inserted in the growing polyketide chain as C-16 and 17 in the final mupirocin molecule while methyl group at position C-15 appears as β-branch. Methyl groups at α position are derived from S-adenosyl methionine (SAM) (Feline et al., 1977, El-Sayed et al., 2003, Wu et al., 2008).

1.17.11.5 Multiple acyl carrier proteins

Tandem repeats of ACPs are present in *mup* cluster. Two ACPs are present in the MmpA module while three ACPs are found in its MmpB (El-Sayed et al., 2003). Thus it belongs to the group of other *trans* AT PKS systems having such unusual architecture of domains like bacillaene, lankacidin, virginamycin and difficidin etc. (Chen et al., 2006, Mochizuki et al., 2003, Pulsawat et al., 2007, Schneider et al., 2007). It has been shown that just one ACP among these double/triplets is essential and sufficient for the production of mupirocin, while extra number of ACPs only contributed to increase mupirocin production significantly, indicating that extra ACPs are redundant (Rahman et al., 2005). It has been shown that each of the ACPs at the end of *mmpA* has unique recognition motif specific for their interaction with the HCS cassette proteins that are involved in introduction of the 15 methyl group (β-branching) (Haines et al., 2013a).

1.18 Thiomarinols

Thiomarinols represent another class of antibiotics which is in fact a hybrid of two known antibiotics (Figure 1.30) for which independent biosynthetic pathways already exist in nature (Fukuda et al., 2011). One of these is monic acid which is the polyketide part of mupirocin antibiotic and the other one is pyrrothine moiety which is part of antibiotic holomycin and related compounds produced by *Streptomyces clavuligerus* (Kenig and Reading, 1979). Pyrrothine is a yellow chromophore that imparts yellow colour to the molecule (Fukuda et al., 2011). Isolation of thiomarinol (Thiomarinol A) was first reported by the fermentation of marine bacterium *Alteromonas rava sp. nov.* SANK73390 collected from the sea water of Japan (Shiozawa et al., 1993). Later, different forms of thiomarinol named as B-to-G were also reported to be produced by these bacteria (Shiozawa et al., 1995, Shiozawa et al., 1997). All the different forms of thiomarinol have related structures therefore they are collectively

called thiomarinols. This producer of thiomarinols was later classified to belong to a new genus named as *Pseudoalteromonas* different from *Alteromonas* and therefore was renamed as *Pseudoalteromonas rava sp. nov.* SANK 73390 (Gauthier et al., 1995). The *Pseudoalteromonas* genus belongs to the *gamma* subclass of *Proteobacteria* of the order *Alteromonadales* (Bowman, 2007). Pseudoalteromonads are found in marine water at different levels from surface to deeper sea. These are often found associated with higher eukaryotes (Holmstrom and Kjelleberg, 1999). They can be pigment producing or non-pigmented type. Pigmented *Pseudoalteromonas sp.* are involved in the production of bioactive substances (Bowman, 2007). Several strains of *Pseudoalteromonas rava sp. nov.* SANK have been isolated from marine water at different places in Japan which either produced black pigment or yellow pigment (Fukuda et al., unpublished data). Thiomarinol producer *Pseudoalteromonas rava sp. nov.* SANK73390 is yellow pigmented (Fukuda et al., 2011). Pseudoalteromonads are Gram-negative bacteria that are rod shaped. They have one or more polar or lateral flagella making them motile (Bowman, 2007). These are non-spore forming strict aerobes which have chemoautotrophic metabolism. These are oxidase positive and are able to grow well at 23 ^{o}C (Gauthier et al., 1995). The GC content of genomic DNA for thiomarinol producer strain *Pseudoalteromonas rava sp. nov.* SANK 73390 is 40.6 % (Fukuda et al., 2011).

1.18.1 Structure and activity

In thiomarinols, the pyrrothine moiety is attached to the 8-hydroxyoctanoic acid side chain of a mupirocin-like derivative known as marinolic acid via amide bond. Marinolic acid derives its name because of its structural analogy with the mupirocin molecule in which monic acid is found attached to a fatty acid (Figure 1.30) (Fukuda et al., 2011, Murphy et al., 2014). Thus, marinolic acid essentially represents thiomarinols without their thiol component. Though the

fatty acid component that is present in pseudomonic acids has a nine carbon chain which is 9-HN. In both these molecules, thiomarinols and pseudomonic acids, the fatty acid chain is attached to the polyketide moiety by an ester linkage. The polyketide moiety in both the molecules is very similar with slight differences in the side groups present. Various types of thiomarinols differ from each other only in the presence or absence of different/ additional side groups around their one primary structure (Figure 1.29) (Shiozawa et al., 1993, Shiozawa et al., 1995, Shiozawa et al., 1997). Among all pseudomonic acids, pseudomonic acid C is structurally closest to thiomarinols and lacks the epoxide ring like them (Shiozawa and Takahashi, 1994).

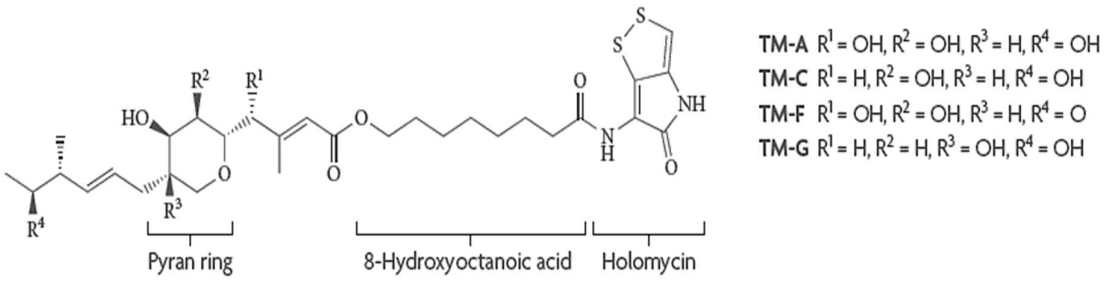


Figure 1.29 Structures of some thiomarinol compounds(TMs) (Reproduced from Thomas et al., 2010).

While mupirocin is effective against Gram-positive bacteria, thiomarinols are also effective against many Gram-negative bacteria as well. The antimicrobial activity of thiomarinols is stronger than the combined activity of both the known antibiotics the components of which thiomarinols have. Thiomarinol A has more potency than any of the pseudomonic acids which is about twenty fold more against organisms like MRSA (Shiozawa et al., 1993). The increased strength of thiomarinols over mupirocin could be due to many reasons: increased uptake; targeting of more than one cellular function; or increased activity against the usual mupirocin target-isoleucyl t-RNA synthetase (Hughes and Mellows, 1978, Thomas et al.,

2010). Despite their strong antimicrobial properties, thiomarinols are not currently used as an antibiotic because of their toxicity.

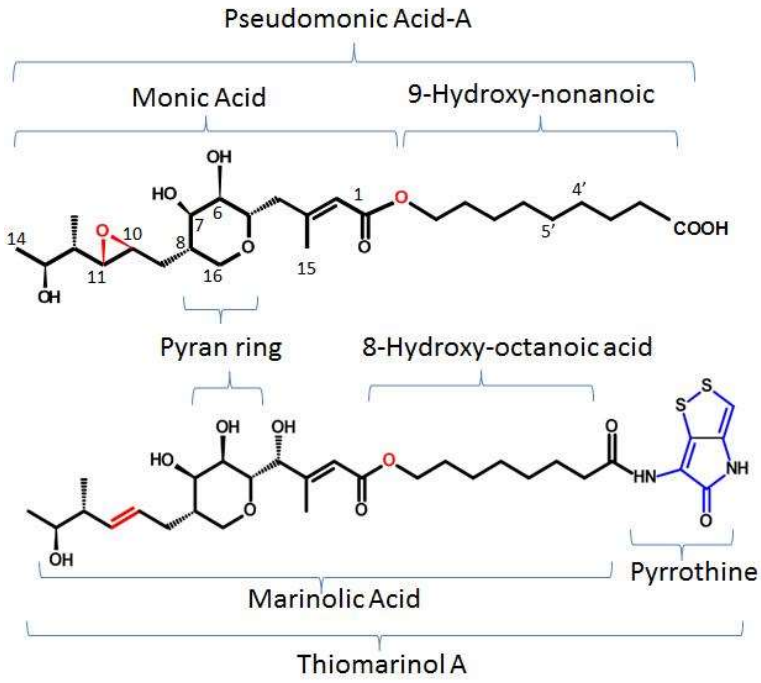


Figure 1.30 A comparative account of the structures of pseudomonic acid A and thiomarinol A(Adapted from Fukuda et al., 2011).

1.18.2 The thiomarinol gene cluster

Efforts by Fukuda et al. (2011) led to the identification of thiomarinol gene cluster (*tml)* from marine bacterium *Pseudoalteromonas sp*. SANK 73390, which was found to be entirely located on a 97 kb circular plasmid named pTML1 (Figure 1.31). Out of the 97 kb of this plasmid, 7.6 kb is occupied by five ORFs that are commonly present in mobile DNA like replication, partitioning, integration and transposition. Apart from this mobile DNA and 7 ORFs coding for NRPS component of thiomarinols, the other 27 ORFs of this cluster are very similar to mupirocin biosynthetic cluster. Even the gene order for type I PKS from *tmpA* to *tmpD* was the same as in mup cluster and both are *trans*-AT type clusters. Some notable differences in the two biosynthetic clusters, in parts, which otherwise have significant

similarities are (1) presence of an extra ACP in the third module of TmpD, (2) an extra ACP in the last module of TmpA and (3) two extra ACPs in TmpB, one of which is present as extra module together with a KS^0. Also, the GC content of thiomarinol encoding genes (43.2%) is significantly less than that of mupirocin encoding gene cluster (56%) found in *P. fluorescens* NCIMB 15860. Contrary to the notable similarities in PKS genes of two clusters, the tailoring region genes are significantly rearranged in the *tml* cluster compared to in *mup* cluster. No putative regulatory gene/protein was initially detected on plasmid pTML1. The entire gene set was found to split into five transcriptional units (Fukuda et al., 2011). Bioinformatics analysis revealed that there has been conservation of PKS and associated genes in the biosynthetic clusters of these two antibiotics (Figure 1.32). In fact, amino acid sequence alignments of products predicted of *tml* cluster showed significant identity (40% to 60%) with the corresponding proteins of *mup* cluster (Fukuda et al., 2011).

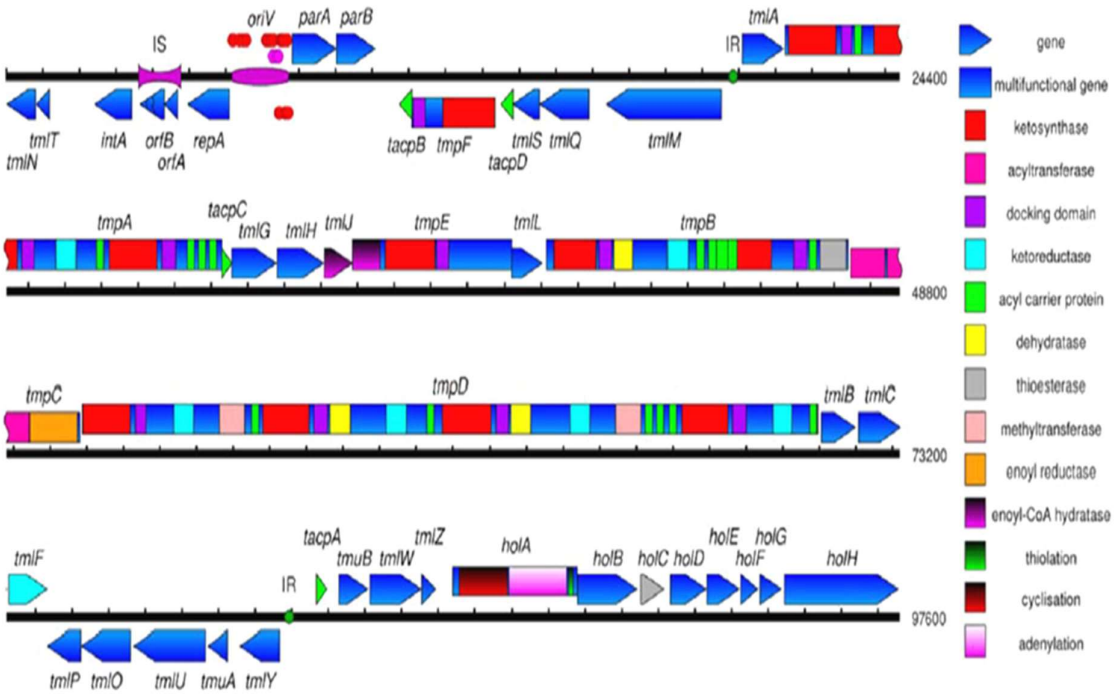


Figure 1.31 The genetic organization of thiomarinol biosynthetic cluster (*tml*) found on plasmid pTML1(Reproduced from Fukuda et al., 2011).

1.18.3 Thiomarinol biosynthetic pathway

The biosynthetic pathway for thiomarinols is predicted to be same as for mupirocin biosynthetic pathway as far as biosynthesis of marinolic acid- a monic acid analogue is concerned (Figure 1.32) (Fukuda et al., 2011). Differences from the biosynthetic scheme of mupirocin, that are evident on account of the presence or absence of particular gene functions or were evident by mutasynthesis experiments, were accounted for in the predicted scheme (Figure 1.32) (Fukuda et al., 2011). Suicide mutagenesis in *tmpD* (PKS) gene led the cells to produce only pyrrothine while suicide mutagenesis in *holA* (NRPS) gene produced only marinolic acid. Similarly, suicide mutagenesis of *tmlU* led to the production of both marinolic acid as well as pyrrothine but not of thiomarinols that indicated the role of *tmlU* in joining the two components to give active thiomarinol (Figure 1.33) (Fukuda et al., 2011). The absence

of *macpE* in the *tml* cluster, product of which is proposed to be involved in the last step of pseudomonic acid production while the presence of an extra module in TmpB indicated that all tailoring region steps take place at TmpB (Figure 1.32) (Fukuda et al., 2011). Similarly, absence of function of the second hydratase, that is encoded by MupK of *mup* cluster and which is implicated in the insertion of C-15 methyl group along with other genes of HCS cassette in mupirocin biosynthesis, is accounted for in thiomarinol cluster by the fact that N-terminal end of TmpE aligns with the MupK protein. Under this context, work was undertaken to test those tailoring region gene functions of thiomarinols, products of which share significant amino acid sequence identity with the corresponding *mup* tailoring genes to find whether they can complement functions of their *mup* counterparts**.**

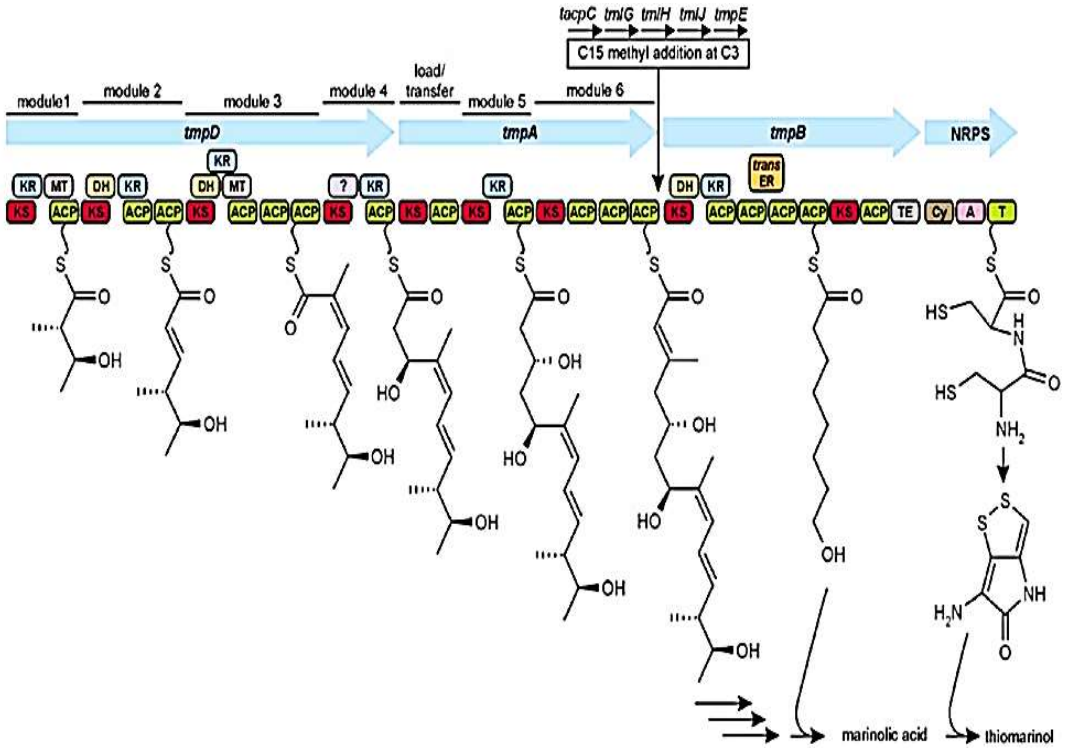


Figure 1.32 Scheme predicted for thiomarinol biosynthesis. Biosynthesis of thiomarinols is hypothesized to occur in the same way as that of mupirocin is done by *mup* cluster using same substrates by equivalent *tml* gene functions. TmpD module 1-4 and module 1 and 2 of TmpA are involved in the biosynthesis of polyketide component while TmpB elaborates 8-ON acid. Pyrrothine part of thiomarinols is elaborated separately by NRPS and then it is joined to the marinolic acid component to give thiomarinols. KS, ketosynthase; ACP, acyl carrier protein; DH, dehydratase; ER, enoyl reductase; KR, ketoreductase; MT, methyl transferase; TE, thioesterase; C, condensation; T, thiolation domain or peptidyl carrier protein; A, aminoacyl adenylation domain (Reproduced from Fukuda et al., 2011).

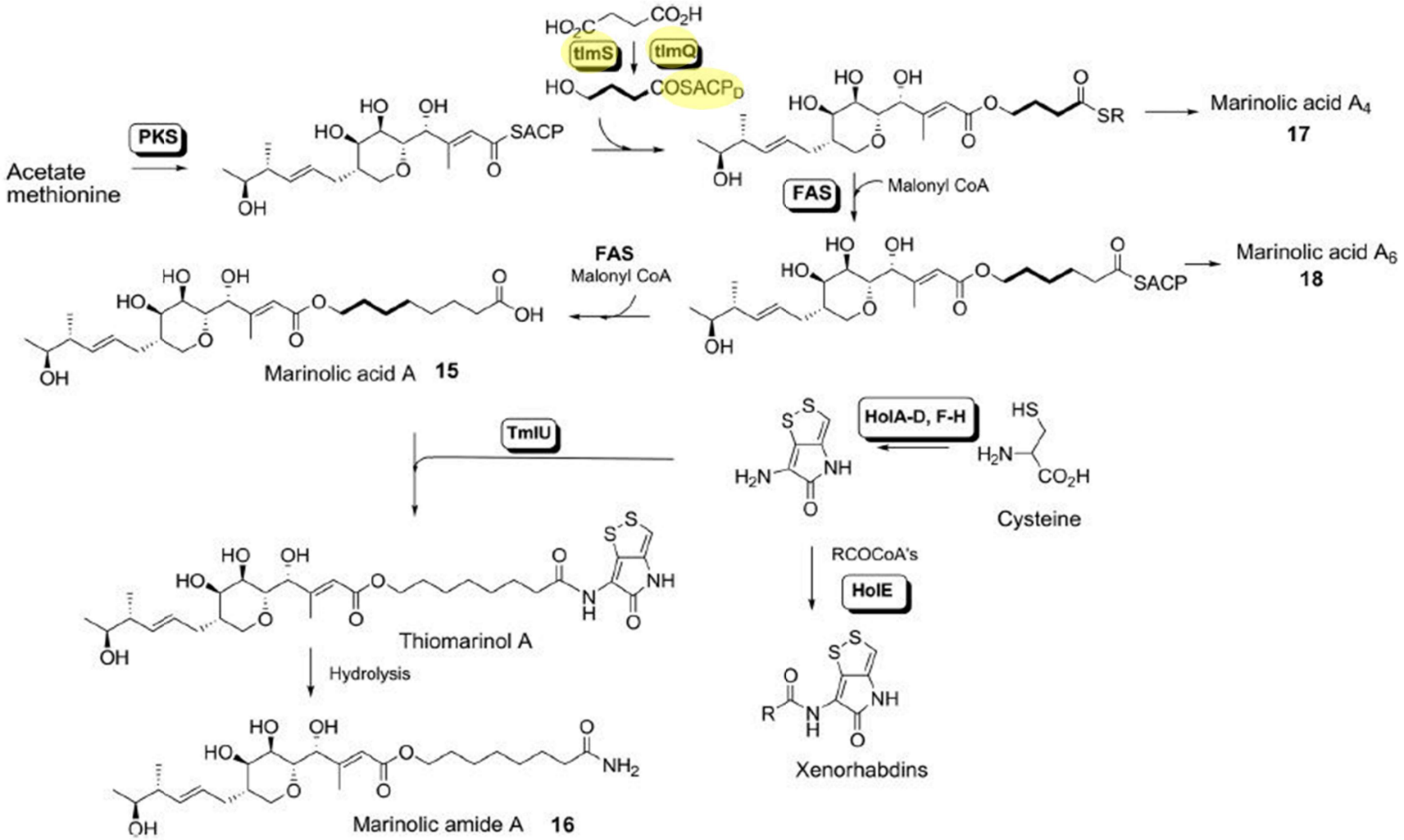


Figure 1.33 Some key biosynthetic steps in the biosynthesis of thiomarinols(Reproduced from Murphy et al., 2014).

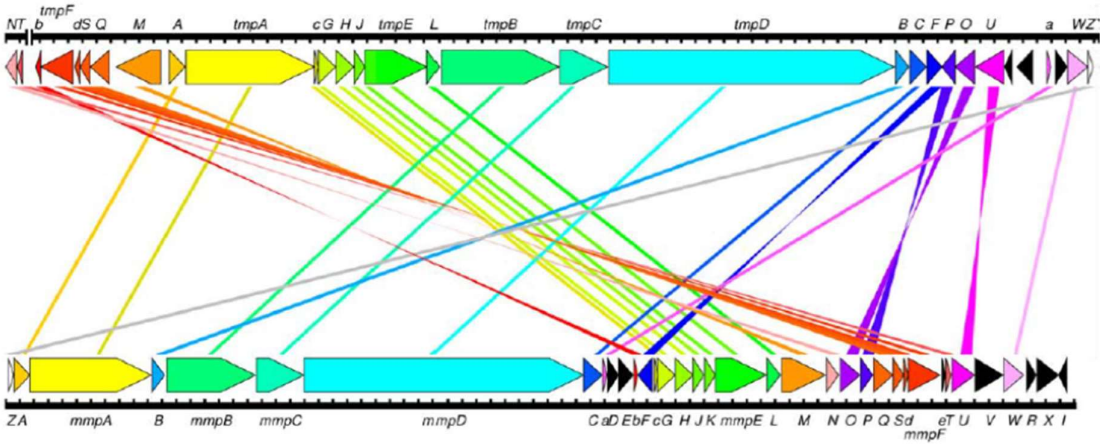


Figure 1.34 Comparison of the organization of the part of thiomarinol gene cluster that codes for its polyketide and fatty acid components with the mupirocin biosynthetic gene cluster. Upper line represents *tml* cluster while lower line represents *mup* cluster. Only equivalent gene functions (ORFs) are shown through lines joining the two clusters (Adapted from Fukuda et al., 2011).

1.19 Aims and objectives

The general aim of the doctoral work described in this thesis was to increase our understanding of the mupirocin and thiomarinol biosynthetic pathways to underpin their manipulation to create novel molecules, particularly production of hybrid pathways. Production of hybrid pathways involves the elements of different pathways working together and so understanding the extent to which related proteins from different pathways can be moved between pathways should help us to identify factors such as substrate specificity and protein-protein interactions that limit the creation of hybrid pathways. A major part of my project, therefore, aimed to investigate the extent to which genes could be swapped between the related mupirocin and thiomarinol pathways. The specific project follows from a study published by Fukuda et al., 2011 and aimed, in the first place, at testing functional cross-complementation by *in trans* expression of singular genes out of 22 homologous auxiliary/tailoring genes of the mupirocin and thiomarinol biosynthetic clusters whose products share significant identity in amino acid sequences**.** Thus, Chapter 4 of thesis is about testing cross-complementation by these homologous thiomarinol genes, singularly, in the corresponding *mup* knockouts of *Pseudomonas fluorescens* NCIMB10586. With an aim of testing reverse complementation of these genes to know more about protein-protein interactions between the two biosynthetic pathways and, for deciphering individual roles played by each of the tailoring genes in the thiomarinol biosynthetic pathway, this study also involved making of in-frame knockouts of single genes in the *tml* gene cluster. This is described in Chapter 3 of the thesis. Chapter 5 of this thesis is an extension of approach of testing functional cross-complementation used in Chapter 4 to investigate cross-complementation by a group of genes whose products are predicted to work together in the biosynthesis of fatty acid components of these two molecules.

CHAPTER 2

Material and Methods

2.1 Bacterial strains and plasmids

Details of various strains and plasmids that were commonly used throughout this work are given in Table 2.1 and Table 2.4 respectively. Strains and plasmids, that are specific to the study of individual chapters, are given in those chapters. Wild type thiomarinol producing *Pseudoalteromonas sp.* SANK 73390 strain was used as the source of thiomarinol coding genes. *Escherichia coli* DH5*α* strain was used for transformation and propagation of plasmids. *E. coli* S17-1 strain was used for mating/conjugation purpose for the mobilisation of expression vectors. Bioassay was used to test the production/activity of pseudomonic acids by the test organisms - various mutants of *Pseudomonas fluorescens* NCIMB10586 where wild type was used as one of the controls. *Bacillus subtilis* 1064 was used as mupirocin sensitive organism in the bioassays. Plasmid pGEM-T Easy vector supplied by Promega and plasmid pJH10 were used to clone genes for sequencing/amplification and to achieve expression of cloned genes, respectively. Plasmid pAKE604 was used for suicide mutagenesis.

Table 2.1 Bacterial strains used in this study.

Bacterial strain	Genotype/ phenotype	Source/ reference
Pseudoalteromonas sp. SANK 73390	Wild type (WT) strain producing thiomarinol	Fukuda et al., 2011
Escherichia coli DH5α	F$^-$*Φ80lacZΔM15*, *recA1*, *endA1*, *gyrA86*, *thi-1*, *hsdR17* (r_K^-, m_K^+) *supE44, relA1, deoR, Δ (lacZYA-ArgF)U169*	GIBCO BRL
Escherichia coli S17-1	*rec*A *pro hsd*R RP*4-2-Tc::Mu*-km::Tn7	(Simon et al., 1983)
Escherichia coli C2110	K-12 *pol*A1 *his rha*	(Stachel et al., 1985)
Escherichia coli ER2925	*dcm*$^-$ *and dam*$^-$	New England Bio labs
Bacillus subtilis 1064	*trp*C2	(Moir et al., 1979)
Pseudomonas fluorescens NCIMB 10586	Wild type (WT) mupirocin producing strain	G.T Banks

Table 2.2 Bacterial strains of *Pseudomonas fluorescens* used or constructed in Chapter 4.

Bacterial strain	Genotype	Source/ reference
NCIMB 10586Δ*mupO*	*mupO* gene deleted	Hothersall et al.,2007
NCIMB 10586Δ*mupS*	*mup*S gene deleted	Hothersall et al.,2007
NCIMB 10586Δ*mupQ*	*mup*Q gene deleted	Hothersall et al.,2007
NCIMB 10586Δ*macpD*	*macpD* gene deleted	Hothersall et al.,2007
NCIMB 10586Δ*mmpF*	*mmpF* gene with point mutation (C183A) in its active site	Hothersall et al.,2007
NCIMB 10586Δ*mupF*	*mupF* gene deleted	Hothersall et al.,2007
NCIMB 10586Δ *mupF*(pJH10-*mupF*)	*mupF* gene deleted and carrying pJH10-*mupF* in *trans*	Hothersall et al.,2007
NCIMB 10586Δ*mupC*	*mupC* gene deleted	Hothersall et al.,2007
NCIMB 10586Δ*mupK*	*mupK* gene deleted	Hothersall et al.,2007
NCIMB 10586Δ*mupJ*	*mupJ* gene deleted	Hothersall et al., 2007
NCIMB 10586Δ*mupO* (pJH10)	*mupO* gene deleted and carrying pJH10 in *trans*	This study
NCIMB 10586Δ*mupO* (pMY2)	*mupO* gene deleted and carrying pMY2 in *trans*	This study

NCIMB 10586Δ*mupS* (pJH10)	*mup*S gene deleted and carrying pJH10 in *trans*	This study
NCIMB 10586Δ *mupS*(pMY4)	*mup*S gene deleted and carrying pMY4 in *trans*	This study
NCIMB 10586Δ*mupQ* (pJH10)	*mup*Q gene deleted and carrying pJH10 in *trans*	This study
NCIMB 10586 Δ*mupQ*(pMY6)	*mup*S gene deleted and carrying pMY6 in *trans.*	This study
NCIMB 10586Δ*macpD* (pJH10)	*macpD* gene deleted and carrying pJH10 in *trans*	This study
NCIMB 10586Δ*macpD* (pMY8)	*macpD* gene deleted and carrying pMY8 in *trans*	This study
NCIMB 10586Δ*mmpF* (pJH10)	*mmpF* gene deleted and carrying pJH10 in *trans*	This study
NCIMB 10586Δ *mmpF*(pMY10)	*mmpF* gene deleted and carrying pMY10 in *trans*	This study
NCIMB 10586Δ*mupF* (pJH10)	*mupF* gene deleted and carrying pJH10 in *trans*	This study
NCIMB 10586Δ*mupF* (pMY12)	*mupF* gene deleted and carrying pMY12 in *trans*	This study
NCIMB 10586Δ*mupC* (pJH10)	*mupC* gene deleted and carrying pJH10 in *trans*	This study
NCIMB 10586Δ*mupC* (pMY14)	*mupC* gene deleted and carrying pMY14 in *trans*	This study

NCIMB 10586Δ*mupK* (pJH10)	*mupK* gene deleted and carrying pJH10 in *trans*	This study
NCIMB 10586 Δ*mupK*(pMY16)	*mupK* gene deleted and carrying pMY16 in *trans*	This study
NCIMB 10586Δ*mupJ* (pJH10)	*mupJ* gene deleted and carrying pJH10 in *trans*	This study
NCIMB 10586 Δ*mupJ*(pMY18)	*mupJ* gene deleted and carrying pMY18 in *trans*	This study

Table 2.3 Bacterial strains constructed in Chapter 5.

Bacterial strain	Genotype	Reference
P. fluorescens NCIMB 10586Δ*mmpF*Δ*mupS* Δ*mup*QΔ*macpD*	*P. fluorescens* mutant with *mmpF, mupS, mupQ* and *macpD* genes deleted	This study
NCIMB 10586Δ*mmpF* Δ*mupS*Δ*mupQ*Δ*macpD* (pJH10)	*P. fluorescens* mutant with deleted *mmpF, mupS, mupQ* and *macpD* genes carrying pJH10 in *trans*	This study
NCIMB 10586 Δ*mmpF*Δ*mupS*Δ*mupQ* Δ*macpD*(pMY23)	*P. fluorescens* with deleted *mmpF, mupS, mupQ* and *macpD* genes carrying pMY23 in *trans.*	This study

Table 2.4 Plasmids used in this study.

Plasmid	Size (kb)	Description	Reference
pGEM-T Easy	3.0	Ap^R, *lacZ*α (PCR cloning vector)	Promega
pJH10	13.74	Tc^R; *lac*Iq, *tacp*, IncQ replicon, multiple cloning site (*Eco*RI-*Sac*I)	El Sayed et al., 2003
pAKE604	7.2	pMB1 replicon, Ap^R, Km^R, *oriT, lacZα, sacB*, multiple cloning sites (*Bam*HI-*Eco*RI)	El Sayed et al., 2001

Table 2.5 Plasmids constructed in Chapter 3.

Plasmid	Size (kb)	Description	Objective	Reference
pMY24	3.54	530 bp PCR product with *Bam*HI and *Xba*I sticky ends (arm1) cloned into pGEM-T Easy vector for the deletion of *tmlW*	For sequencing of the PCR product	This study
pMY25	3.54	527 bp PCR product with *Eco*RI and *Xba*I sticky ends (arm2) cloned into pGEM-T Easy vector for the deletion of *tmlW.*	For sequencing of the PCR product	This study
pMY26	8.24	1063 bp fragment (produced by ligating arm1 of 530 bp with arm2 of 527 bp using *Xba*I sticky ends) cloned into pAKE604 vector at *Bam*HI and *Eco*RI sites.	For the deletion of *tmlW* gene (precisely for deleting a fragment of Δ1377 bp).	This study
pMY27	3.6	583 bp PCR product with *Bam*HI and *Xba*I sticky ends (arm1) cloned into pGEM-T Easy vector for the deletion of *tmlT*	For sequencing of the PCR product	This study
pMY28	3.6	591 bp PCR product with *Eco*RI and *Xba*I sticky ends (arm2) cloned into pGEM-T Easy vector for the deletion of *tmlT*	For sequencing of the PCR product	This study
pMY29	8.3	1180 bp fragment (produced by ligating arm 1 of 583 bp with arm2 of 591 bp using *Xba*I sticky ends) cloned into pAKE604 vector at *Bam*HI and *Eco*RI sites.	For the deletion of *tmlT* gene (precisely for deleting a fragment of Δ303 bp).	This study

pMY30	3.5	500 bp PCR product with *Bam*HI and *Xma*I sticky ends (arm1) cloned into pGEM-T Easy vector for the deletion of *tmlO*	For sequencing of the PCR product	This study
pMY31	3.5	489 bp PCR product with *Eco*RI and *Xma*I sticky ends (arm2) cloned into pGEM-T Easy vector for the deletion of *tmlO*	For sequencing of the PCR product	This study
pMY32	8.18	995 bp fragment (produced by ligating arm1 of 500 bp with arm2 of 489 bp using *Xma*I sticky ends) cloned into pAKE604 vector at *Bam*HI and *Eco*RI sites.	For the deletion of *tmlO* gene (precisely for deleting a fragment of Δ1305 bp).	This study
pMY33	3.5	490 bp PCR product with *Bam*HI and *Xba*I sticky ends (arm1) cloned into pGEM-T Easy vector for the deletion of *tmlS.*	For sequencing of the PCR product	This study
pMY34	3.5	492 bp PCR product with *Xba*I and *Eco*RI sticky ends (arm2) cloned into pGEM-T Easy vector for the deletion of *tmlS.*	For sequencing of the PCR product	This study
pMY35	8.17	988 bp fragment (produced by ligating arm1 of 490 bp with arm2 of 492 bp using *Xba*I sticky ends) cloned into pAKE604 vector at *Bam*HI and *Eco*RI sites.	For the deletion of *tmlS* gene (precisely for deleting a fragment of Δ675 bp).	This study

pMY36	3.50	488 bp PCR product with *Xma*I and *Xba*I sticky ends (arm1) cloned into pGEM-T Easy vector for the deletion of *tacpD*	For sequencing of the PCR product	This study
pMY37	3.53	518 bp PCR product with *Eco*RI and *Xba*I sticky ends (arm2) cloned into pGEM-T Easy vector for the deletion of *tacpD*	For sequencing of the PCR product	This study
pMY38	3.55	536 bp PCR product with *Xma*I and *Xba*I sticky ends (arm1) cloned into pGEM-T Easy vector for the deletion of *tmlQ*	For sequencing of the PCR product	This study
pMY39	3.54	528 bp PCR product with *Eco*RI and *Xba*I sticky ends (arm2) cloned into pGEM-T Easy vector for the deletion of *tmlQ*	For sequencing of the PCR product	This study
pMY40	3.5	530 bp PCR product with *BamHI* and XbaI sticky ends (arm1) cloned into pGEM-T Easy vector for the deletion of *tmpF*	For sequencing of the PCR product	This study
pMY41	3.5	519 bp PCR product with *XmaI* and XbaI sticky ends (arm2) cloned into pGEM-T Easy vector for the deletion of *tmpF*	For sequencing of the PCR product	This study

Table 2.6 Plasmids constructed in Chapter 4.

Plasmid	Size (kb)	Description	Reference
pMY1	4.35	*tmlO* ORF of 1352 bp cloned in pGEM-T Easy vector after A-tailing	This study
pMY2	15	*tmlO* ORF of 1352 bp cloned in pJH10 using *Kpn*I and *Xba*I	This study
pMY3	3.75	*tmlS* ORF of 752 bp cloned in pGEM-T Easy vector after A-tailing	This study
pMY4	14.46	*tmlS* ORF of 725 bp cloned in pJH10 using *Kpn*I and *Xba*I	This study
pMY5	4.33	*tmlQ* ORF of 1337 bp cloned in pGEM-T Easy vector after A-tailing	This study
pMY6	15	*tmlQ* ORF of 1337 bp cloned in pJH10 using *Eco*RI and *Xba*I	This study
pMY7	3.33	*tacpD* ORF of 335 bp cloned in pGEM-T Easy vector after A-tailing	This study
pMY8	14.07	*tacpD* ORF of 335 bp cloned in pJH10 using *Eco*RI and *Xba*I	This study
pMY9	5.25	*tmpF* ORF of 2252 bp cloned in pGEM-T Easy vector after A-tailing	This study
pMY10	15.99	*tmpF* ORF of 2252 bp cloned in pJH10 using *EcoR*I and *Xba*I	This study
pMY11	4.05	*tmlF* ORF of 1053 bp cloned pGEM-T Easy vector after A-tailing	This study

pMY12	14.79	*tmlF* ORF of 1053 bp cloned in pJH10 using *Kpn*I and *Sac*I	This study
pMY13	4.11	*tmlC* ORF of 1140 bp cloned in pGEM-T Easy vector after A-tailing	This study
pMY14	14.88	*tmlC* ORF of 1140 bp cloned in pJH10 using *Kpn*I and *Sac*I	This study
pMY15	3.74	*tmlK*[1] ORF of 749 bp cloned in pGEM-T Easy vector after A-tailing	This study
pMY16	14.48	*tmlK*[1] ORF of 749 bp cloned in pJH10 using *Kpn*I and *Xba*I	This study
pMY17	3.76	*tmlJ* ORF of 767 bp cloned in pGEM-T Easy vector after A-tailing	This study
pMY18	14.50	*tmlJ* ORF of 767 bp cloned in pJH10 using *Kpn*I and *Xba*I	This study

[1] *tmlK* gene is non-existent in *tml* cluster. It is the ORF at the N-terminal end of TmpE that aligns with MupK because it shares significant amino acid sequence identity with it, which was supplied with external stop codon.

Table 2.7 Plasmids constructed in Chapter 5.

Plasmid	Size (kb)	Genotype description	Purpose	Reference
pMY19	3.55	548 bp *Eco*RI-*Xba*I PCR product containing arm1 cloned into pGEM-T Easy vector	Sequencing of PCR product (arm1) for creating quadruple mutation in wild type *P. fluorescens*	This study
pMY20	3.51	505 bp *Xba*I-*Bam*HI PCR product containing arm2 cloned into pGEM-T Easy vector	Sequencing of PCR product (arm2) for creating quadruple mutation in wild type *P. fluorescens*	This study
pMY21	8.27	pAKE604 with fragment of 1053 kb (obtained by joining arm1 and arm2 using *Xba*I sticky ends) cloned using *Eco*RI-*Bam*HI sticky ends.	For deletion of block of four genes (Δ*mmpF,* Δ*mupS*, Δ*mupQ,* Δ*macpD*) from wild type *Pseudomonas fluorescens* NCIMB 10586.	This study
pMY22	5.21	pGEM-T Easy vector with cloned PCR amplified fragment *Xba*I-*Sac*I carrying ORFs for *tmlQ, tmlS* and *tacpD.*	Sequencing of PCR amplified fragment for expression of *tmlQ, tmlS* and *tacpD* ORFs.	This study
pMY23	21.45	pJH10-tmpF::tmlQ-tmlS tacpD-pGEM-T Easy co-integrate obtained by ligating plasmids pMY10 (carrying *tmpF* as *Eco*RI and *Xba*I fragment) and pMY22 (carrying fragment with ORFs of genes *tmlQ, tmlS* and *tacpD*), using *Xba*I restriction site.	Expression of thiomarinol proteins TmpF, TmlQ, TmlS and TacpD.	This study

2.2 Growth, media and culture conditions

All *Pseudomonas* strains were grown at 30 ^{o}C either in L broth or on L agar. *Pseudoalteromonas sp.* SANK 73390 strain was grown at 23 ^{o}C on M agar or M broth. L broth had composition as yeast extract (10 g/l), NaCl (10 g/l), tryptone (5 g/l) and glucose (1 g/l) while L agar additionally included 15 g/l of agar. Marine broth comprised of peptone (5.0 g/l), yeast extract (10 g/l), NaCl (19.45 g/l), ferric citrate (0.1 g/l), magnesium chloride (5.9 g/l), magnesium sulphate (3.24 g/l), $CaCl_2$ (1.80 g/l), KCl (0.55 g/l), $NaHCO_3$ (0.16 g/l), strontium chloride (0.034 g/l), boric acid (0.022 g/l), sodium silicate (0.004 g/l), sodium fluoride (0.0024 g/l), ammonium nitrate (0.0016 g/l) and disodium phosphate (0.008 g/l). In M agar all components were same as in M broth except magnesium sulphate, $MgCl_2$ (8.8 g/l), sodium sulphate (3.24 g/l) and agar (15.0 g/l) which were additionally added. These media were supplemented with antibiotics in required concentrations as and when needed (as mentioned in Table 2.8 below). *E. coli* strains were grown at 37 ^{o}C on L broth or L agar supplemented with tetracycline 15 µg/ ml or IPTG 0.5 µl/l (in L agar) as and when required. *Bacillus subtilis* 1064 strain was grown at 37 ^{o}C on L agar or L broth without any antibiotic selection. All media were sterilized by autoclaving at a pressure of 15 pounds per square inch at 121 ^{o}C for 15 min and whenever any temperature sensitive supplements like antibiotics or IPTG were added to it, these were added after filter sterilization through 0.2 µ membrane filters. For HPLC analysis all the strains of *Pseudomonas fluorescens* NCIMB 105806 were grown in secondary stage medium (SSM) that contained soya flour 25 g/l, spray dried corn liquor 2.5 g/l, $(NH_4)_2SO_4$ 5.0 g/l, $MgSO_4.7H_2O$ 0.5 g/l, Na_2HPO_4 1.0 g/l, KH_2PO_4 1.5 g/l, KCl 1.0 g/l, $CaCO_3$ 6.25 g/l. This medium was made in distilled water and pH was adjusted to 7.5. After sterilizing it by autoclaving, separately autoclaved 40% w/v glucose was added to it before any use. For plasmid DNA isolation, cultures were incubated overnight (14-16 hours). For HPLC, seed culture was also incubated for overnight while cultures for the

production of mupirocin were incubated up to 40 hours (till secondary/stationary stage). Liquid cultures were always set up in sterilized containers having volumes four times of the amount of culture that was set up. For the preparation of competent cells of different strains of *E. coli* L broth was used. All the stock solutions of antibiotics were prepared in suitable solvent. Table 2.8 gives the concentrations of all the antibiotics and the other supplements used in the cultures.

Table 2.8 Antibiotics and other substrates used in this study.

Antibiotic/substrate	**Solvent used**	**Working concentration used (µg/ml)**
Ampicillin	H_2O	50
Tetracycline hydrochloride	70% C_2H_5OH	15
Kanamycin sulphate	H_2O	50
Carbenicillin	H_2O	50
IPTG	H_2O	0.5 mM
X-gal	N, N-dimethylformamide	50
Sucrose	H_2O	5.5 % w/v

2.3 Polymerase Chain Reaction

Polymerase Chain reaction was used to amplify segments of DNA (Saiki et al., 1985, Saiki et al., 1988) allowing amplification of short DNA sequences with the help of two primers designed to complement the ends of the DNA sequence to be amplified. DNA template for the reaction was prepared by adding a loop-full of O/N grown bacteria in 50 µl of sterilized distilled water which was suspended by vortexing. Cell suspension was then heated in boiling water for 10 min and then centrifuged at 14000×g for 5 min at 4 °C. Thereafter supernatant was collected. This supernatant had bacterial DNA which was used as template for PCR and was kept at ice. This method of preparing template DNA is known as boil prep method. The enzyme used for amplification depended upon the size of the DNA fragment to be amplified and the desired fidelity. For most of the amplifications Velocity PCR kit (Bioline), that makes the use of velocity DNA polymerase, was used. For amplification of larger fragments of DNA, KOD polymerase (Novagen) was used. Wherever proof-reading activity was not required, for example when checking for cloned inserts, *Taq* polymerase (Invitrogen) was used. Primers were obtained from Alta Biosciences, University of Birmingham upon ordering and supplying their designs. Primers were dissolved in sterile distilled water and stored at -20 °C. These were used after diluting to 10 pmol/µl concentration. Primers used in the study are mentioned in respective chapters. Primers were designed to have 18-24 bases approximately and to have melting temperatures (T_m) between 55 and 60 °C. Designed primers were checked for melting temperature and for secondary structure formation, if any, by on-line OligoCal software of Northwestern University, USA. For thorough analysis of primers NetPrimer (www.premierbiosoft.com/netprimer) was used. Annealing temperature was roughly set at 4 °C less than melting temperature (T_m), which was calculated using equation given below.

$$T_m = 4\,(G+C) + 2\,(A+T)\ ^oC$$

Extension time was calculated after considering the rate of addition of base per second specified by the manufacturer for the polymerase enzyme used and the size of fragment to be amplified. All the PCR reactions were carried out (incubated) using laboratory thermal cycler (Gene flow SensoQuest Lab-cycler).

Table 2.9 Primers used in Chapter 3.

Name of Primer[1]	Objective	5' Co-ordinate[2]	Primer sequence, 5'→3'	Number of bases	Restriction sites[3]
SA1F	Deletion of *tmlS* gene	15013	GGTGGATCCGCCAAATAT AAGGTTCAAGATGGAGTC C	37	*Bam*HI
SA1R		14524	TGGTCTAGACACTGTACC ATTTTCACTTTCCATCGG	36	*Xba*I
SA2F		13848	GGTTCTAGAGTAATAGAC GGAG GTTTTACATGT TAAACG	39	*Xba*I
SA2R		13357	TGGGAATTCGTTTATACG CTTTATTAATAAGCGTGA GCAC	40	*Eco*RI
OA1F	Deletion of *tmlO* gene	77074	GGTGGATCCCATTAACAT CAGGTGCAACGCTTAGC	35	*Bam*HI
OA1R		76575	TGGCCCGGGCTGATTCTG CGAATTAGTCGTCACAAC	36	*Xma*I
OA2F		75269	GGTCCCGGGCGCGCAACC GATCAAATCGTATAG	33	*Xma*I
OA2R		74781	GGGAATTCCCTAAGAATC GGCTTTGTCAGATCC	33	*Eco*RI
WA1F	Deletion of *tmlW* gene	82631	GGTGGATCCCCAGGAGAA AGCAGGCAGGCTATAC	34	*Bam*HI
WA1R		83160	TGGTCTAGACGGTGCCAA TTTATTCATGTTTCAACCC TG	39	*Xba*I

WA2F		84538	TGGTCTAGAGAAAGCTAG GAGGCTGAACGATGTTGG	36	*Xba*I
WA2R		85064	TGGGAATTCGCAATATAT CATTTAACACTAAGCCCA CACTGAG	43	*Eco*RI
TA1F	Deletion of *tmlT* gene	1745	GGTGGATCCGATACCGCA CCTGTCGAAGAATTTACT C	37	*Bam*HI
TA1R		1155	TGGTCTAGAGTTCCCTTGT GCTAGTGTCATAGGTAC	36	*Xba*I
TA2F		851	GGTTCTAGAGAAATAGAG TCCGTGTCTTGTTAAGTTG AG	39	*Xba*I
TA2R		269	TGGGAATTCGGTATAGAC ATC CCCTGTCCACAT	33	*Eco*RI
TMA1F	Deletion of *tmpF* gene	13793	GGTGGATCCGGATAGTAT TATCCAAGAAGTAGCCAC	36	*Bam*HI
TMA1R		13264	TGGTCTAGAGAATTGACT CGGAGAGCTTGCTTC	33	*Xba*I
TMA2F		11100	GGTTCTAGAGGTGACGAA GAC TAACATTGCCAC	33	*Xba*I
TMA2R		10582	TGGCCCGGGGGTGACAGA AACTCCTTTTAGCTTTTC	36	*Xma*I

QA1F	Deletion of *tmlQ* gene	16431	GGTCCCGGGCCAATACGATGTGGATGGCCATACCTTG	37	*Xma*I
QA1R		15896	TGGTCTAGACACCAGGTTGGCTTCCTTTTTAATGGTTCATATAAC	45	*Xba*I
QA2F		14590	GTTCTAGAGAAAGATTCAAAAAAATCAATAAGATATAAGGAATAAAAACCGATGGAAAG	60	*Xba*I
QA2R		14063	TGGGAATTCCTCGTTATACCATCATAGCAGCCTTTGTAG	39	*Eco*RI
TcpA1F	Deletion of *tacpD* gene	14307	GGTCCCGGGGGTTTAGTCAACAATGCAGCCATAGGAG	37	*Xma*I
TcpA1R		13820	TGGTCTAGAGTTTAACATGTAAAACCTCCGTCTATTACGATG	42	*Xba*I
TcpA2F		13504	GGTTCTAGACAAGTAAGGTGATCTACCACGAGTATAAGC	39	*Xba*I
TcpA2R		12987	TGGGAATTCCACCTGTTTCGTTTG TCTGTGCTGTTG	36	*Eco*RI

[1]For all the amplifications using these primers, plasmid pTML1 from WT *Pseudoalteromonas sp.* SANK 73390 was used as template DNA.
[2]Coordinates given in column 3 indicate the base numbers of pTML1plasmid. These represent the arms that were PCR amplified.
[3]Restriction sites, start and stop codons in primers are shaded in turquoise, green and red colours respectively.

Table 2.10 Primers used in Chapter 4.

Gene (Size in kb)	Name of Primer[1]	Primer sequence, 5'→3'	Number of bases	Objective	Restriction sites[2]
*tml*O (1.352)	RtmlO	TGGTCTAGACTATACGATT TGATCGGTTGCG	31	Amplificat-ion of *tml O* gene	*Kpn*I and *Xba*I
	FtmlO	GGTGGTACCAAGGAAATT GTTGTGACGACTAATTC	35		
*tml*S (0.725)	RtmlS	TGGTCTAGATTAACATGA AAACCTCCGTCTATTAC	35	Amplificat-ion of *tml S* gene	*Kpn*I and *Xba*I
	FtmlS	GGTGGTACCAAGGAATAA AAACCGATGGAAAGTG	34		
*tacp*D (0.335)	RtacpD	TGGTCTAGATCACCTTACT TGGTTTTTAATGGCCTTAA G	39	Amplificat-ion of *tacpD* gene	*Eco*RI and *Xba*I
	FtacpD	GGTGAATTCAGAAGGAGG TTTTACATGTTAAACGTAC AG	39		
*tml*Q (1.337)	RtmlQ	TGGTCTAGATTATATCTTA TTGATTTTTTTGAATCTTTC GCTGTGG	46	Amplificat-ion of *tml Q* gene	*Eco*RI and *Xba*I
	FtmlQ	GGTGAATTCAAGGAAGCC AACCTGGTGAATAGAAC	35		
*tmp*F (2.252)	RtmpF	TGGTCTAGATTAGTCTTCG TCACCCAACCAATAGC	35	Amplificat-ion of *tmpF* gene	*Eco*RI and *Xba*I
	FtmpF	GGTGAATTCAAGTTAGGA AATGTATGATAGCTACTG ATGATATAG	45		

tmpF (internal primers)	Forward	GCGATCGCCGCAGCCAGTTG	20	For verifying sequence of *tmpF* gene	None
	Reverse	GCTGCCAATGCTTGTTGACACCAC	24		None
tmlF (1.053)	RtmlF	TGGGAGCTCCTAGCTTACATTTAACGCGACGG	32	Amplificat-ion of *tmlF* gene	*Kpn*I and *Sac*I
	FtmlF	GGTGGTACCAGGAGTATTATCGTGCGTGAAACAG	34		
tmlC (1.140)	RtmlC	TGGGAGCTCTTATAACACACGAGCTGCTTGACG	33	Amplificat-ion of *tmlC* gene	*Kpn*I and *Sac*I
	FtmlC	GGTGGTACCGGAGAAATTGATGAAAAAAAACGCAGATAATG	41		
tmlK[3] (0.749)	RtmlK[3]	TGGTCTAGAGCCACTCTATTGCCCAAACAGCTG	33	Amplificat-ion of *tml*[3] gene	*Kpn*I and *Xba*I
	FtmlK[3]	GGTGGTACCGGGAAGGTTAATGAATAAGACAATTATCAACTC	42		
tmlJ (0.767)	RtmlJ	TGGTCTAGATTAACCTTCCCATGGAAGCTTACC	33	Amplificat-ion of *tmlJ* gene	*Kpn*I and *Xba*I
	FtmlJ	GGTGGTACCGAGTTAGGACAGTGATGTATGAGATG	35		

[1]For all amplifications using these primers, plasmid pTML1 from WT *Pseudoalteromonas sp.* SANK 73390 was used as template DNA.
[2]Restriction sites, start and stop codons in primers are shaded in turquoise, green and red colours respectively.
[3]*tmlK* gene is non-existent in *tml* cluster. It is the ORF for N-terminal end of TmpE that share significant amino acid sequence identity with MupK, which was supplied with external stop codon for which primers were designed accordingly to amplify the same.

Table 2.11 Primers used in Chapter 5.

Name of primer[1]	Objective	5' Co-ordinate[2]	Primer sequence, 5'→3'	Number of bases	Restriction sites[3]
QDA1F	Deletion of block of four genes *mmpF, mupQ, macpD* and *mupS.*	60394	GGTGAATTCGAAG ATCTGGAGGACGC CGTG	30	*Eco*RI
QDA1R		60941	TGGTCTAGAGTCG CTGATCCAATTAC GTTCCTC	33	*Xba*I
QDA2F		65463	GGTTCTAGAGACG TTGTGCAGGCCGG AATC	30	*XbaI*
QDA2R		65967	TGGGGATCCGCTC ACGTTGGGCGAAC CAC	29	*Bam*HI
RQE3	Amplificat-ion of plasmid DNA fragment (2.4 kb) comprising *tmlS, tmlQ* and t*acpD*	13491	TGGGAGCTCGATC ACCTTACTTGGTT TTTAATGGCCTTA AG	41	*Sac*I
FQE3		15922	GGTTCTAGACCAT TAAAAAGGAAGC CAACCTGGTGAAT AG	40	*Xba*I

QE Int Seq R	Sequencing of DNA fragment (2.4 kb) comprising *tmlS, tmlQ* and t*acpD*	14160	CATCGCGGTTCTG AAGTAG	19	None
QE Int Seq F		15177	GTCAGTGAAGATA GATCGCAAC	22	None

[1]For amplifications using primers for deletion, genomic DNA from *Pseudomonas fluorescens* NCIMB 10586 was used. For the amplification of plasmid DNA fragments and for their sequencing, pTML1 plasmid DNA from WT *Pseudoalteromonas sp.* SANK 73390 was used as template DNA.
[2]Coordinates given in column 3 indicate for primers for the deletion of genes, the base numbers of *Pseudomonas fluorescens* NCIMB 10586 genome and those for amplification/sequencing of plasmid DNA, the base numbers of pTML1 plasmid of WT *Pseudoalteromonas sp.* SANK 73390.
[3]Restriction sites, start and stop codons in primers are shaded in turquoise, green and red colours respectively.

2.3.1

2.3.1 PCR using KOD polymerase (Novagen)

KOD polymerase is a recombinant DNA polymerase with high fidelity and elongation rate compared to other DNA polymerases like *Pfu* or *Taq*. It is derived from *Thermococcus kodakarensis* KOD DNA polymerase. This enzyme also has proof-reading activity and can amplify up to 2 kb of target DNA fragment from genomic template and up to 6 kb from plasmid DNA template. It produces blunt-ended DNA products. Following mix (Table 2.12) was prepared for each PCR reaction which was incubated as per programme given in Table 2.13**.**

Table 2.12 Reaction conditions for KOD polymerase.

Reagent	Volume added (µl) per reaction	Final concentration in the reaction
Template DNA	1	50 pg-10 ng
10x Buffer 1	5	1x
KOD DNA polymerase (2.5 U/ µl)	0.4	1 U
dNTPs (2.5 mM each)	5	250 µM (each)
Forward primer (15 µM stock)	1.3	0.4 µM
Reverse primer (15 µM stock)	1.3	0.4 µM
50% Glycerol	5	5**%**
$MgSO_4$ (25 mM)	2	1 mM
Distilled water	29	
Total reaction volume	**50**	

Table 2.13 KOD polymerase PCR program details.

Stage	Description	Duration	Temperature (°C)	Cycles
1	Initial denaturation	2 min	98	1
2	Denaturation	20 s	98	25-30
	Annealing	*s	T_m-5	
	Elongation	*s/ kb	72	
3	Final elongation	7 min	72	1

*For 1-2 kb target DNA, annealing time was 2 s and elongation time was 20 s; for 3-4 kb target annealing time was 5 s and elongation time was 40 s.

2.3.2 PCR using Velocity™ DNA Polymerase (Bioline)

For amplification of most of the DNA fragments, Velocity™ DNA polymerase was used which was supplied with Velocity PCR kit (Bioline). This polymerase amplifies DNA fragment at a rate of 15-30 s/ kb and generates products with blunt ends. For each reaction following mix (Table 2.14) was prepared which was incubated as per programme given in Table 2.15.

Table 2.14 Velocity ™ PCR kit (Bioline) reaction conditions.

Reagent	**Volume added (µl) per reaction**	**Final concentration in the reaction**
Template DNA	2	50 pg-10 ng
5x HIFI Buffer	10	1x
DMSO	1.5	3%
dNTPs (2.5 mM each)	5	250 µM (each)
Forward primer (15 pmol/µl)	2	0.6 pmol
Reverse primer (15 pmol/µl)	2	0.6 pmol
Velocity *Taq* polymerase (2 U/ µl)	0.5	1 U
Distilled water	27	
Total reaction volume	**50**	

Table 2.15 Velocity™ DNA polymerase PCR program details.

Stage	Description	Duration	Temperature (°C)	Cycles
1	Initial denaturation	2 min	98	1
2	Denaturation	30 s	98	30
	Annealing	30 s	T_m-5	
	Elongation	*s	72	
3	Final elongation	7 min	72	1

*Length of the DNA fragment to be amplified in terms of base pair was taken to calculate the elongation time in view of capacity of Velocity™ DNA polymerase to catalyse the reaction.

2.3.3 PCR using Invitrogen *Taq* polymerase

This was used for checking any mutations (insertions or deletions) introduced in the DNA plasmid/bacterial genome. The PCR reaction used *Taq* DNA polymerase isolated from *Thermus aquaticus.* A typical PCR reaction was set up in the total volume of 50 µl with the following composition (Table 2.16) which was incubated as per programme given in Table 2.17.

Table 2.16 *Taq* polymerase PCR reaction conditions.

Reagent	Volume added (µl) per reaction	Final concentration in the reaction
Template DNA	2	1-500 ng
10x PCR Buffer minus Mg^{++}	5	1x
dNTPs (2.5 mM each)	4	250 µM (each)
Taq DNA polymerase (5 U/ µl)	0.6	3 U
50% Glycerol	10	10%
Forward primer (100 µM)	0.3	0.6 µM
Reverse primer (100 µM)	0.3	0.6 µM
$MgCl_2$ (50 mM)	1.0	1 mM
Sterile distilled water (SDW)	26.8	
Total reaction volume	**50**	

Depending on the number of clones to be checked for mutation, a master mix with above composition was prepared (without adding template DNA) and 45 µl of which was aliquoted for each reaction. Template DNA obtained by boil-prep method (as described above in section 2.3) from the concerned bacterial colony was added to each reaction. For negative

control reaction, instead of DNA, 5 µl of sterile distilled water was added. Other controls were set up by using DNA obtained from wild type strain and the plasmid construct that was used to create the desired effect (insertion or deletion).

Table 2.17 *Taq* polymerase PCR program details.

Stage	Description	Duration	Temperature (°C)	Cycles
1	Initial denaturation	2 min	94	1
2	Denaturation	15 s	94	30
	Annealing	30 s	T_m-5	
	Elongation	*s	72	
3	Final elongation	7 min	72	1

*Length of the DNA fragment to be amplified in terms of base pair was taken to calculate the elongation time in view of capacity of *Taq* polymerase to catalyse the reaction which was 90 s/ kb.

2.4 Manipulation of DNA

2.4.1 Plasmid extraction using Alkaline SDS method

As per this method of Birnboim and Dolly (1979) overnight cultures of plasmid carrying bacteria were set up in 5 ml of L broth supplemented with appropriate antibiotic. Cells in the overnight grown culture were pelleted by centrifugation at 14000×g for 5 min in a microfuge tube. Supernatant was discarded and cells were resuspended in 100 µl of ice cold lysis buffer (25 mM tris pH 8.0, 10 mM EDTA, 50 mM glucose) by vortexing. Thereafter, 200 µl of NaOH-SDS solution (2% SDS and 0.4 M NaOH) that was kept on ice, was added to the resuspended suspension of cells which was then mixed by gently inverting the tube

several times before incubation on ice for 5 min. Afterwards, 150µl of neutralizing solution (3 M solution of sodium acetate pH 5.0) was added and mixed by gentle inversion several times. Incubation was done on ice again for 5 min. The entire mix was then centrifuged at 14000×g at 4ºC for 5 min to pellet the cell debris. The supernatant containing plasmid DNA was transferred to another microfuge tube in which it was precipitated by adding 400 µl of Isopropanol by mixing and incubating at 15 ºC at room temperature. Plasmid DNA was recovered from it by centrifugation of the mix again at 14000×g at 4 ºC this time for 10 min. The precipitated DNA pellet was resuspended in 100 µl TNE buffer (10 mM *tris* pH 7.5, 50 mM NaCl and 5 mM EDTA).

2.4.2 Large scale high quality plasmid DNA extraction (maxi-prep)

Large scale plasmid DNA was isolated using a method that was basically a modified alkaline-SDS method of Birnboim and Dolly (1979). In this method usually 400ml of overnight grown cultures of bacteria carrying the plasmid of interest was set up in L broth using appropriate antibiotic. The culture was spun in centrifuge at 10,000 x g for 10 min at 4 ºC in two pots in equal amounts of 200 ml. The bacterial pellet collected from both the pots after discarding supernatant was resuspended in a total of 25 mls of lysis buffer (solution 1). Thereafter 50 mls of freshly prepared NaOH-SDS solution mix (1:1 mix of 0.4 M NaOH and 2 % SDS; solution 2) was added to the resuspended bacterial pellet in the pot. The pot was inverted several times to mix the solutions thoroughly so as to get complete lysis of the bacterial cells. It was incubated on ice for 5 min and then 37.5 ml of 3 M sodium acetate solution with pH 5 (solution 3) was added to neutralise the mixture and precipitate cell debris and chromosomal DNA, the contents mixed by inverting the pot several times and after another 5 min incubation on ice the pot was centrifuged for 15 min at 10,000 x g at 4 ºC to pellet the cell

debris. The collected supernatant was further filtered through Whatman 1 mm paper to remove any of the remaining cell debris. 100 ml of isopropanol was added to the filtrate in a fresh pot to precipitate the plasmid DNA, mixed and again centrifuged for 15 min at 10,000 x g at 4 ^{o}C. After centrifugation the supernatant was discarded and the DNA and RNA pellet allowed to dry for 30 min to get rid of any remaining isopropanol. Then the DNA pellet, free from any isopropanol, was dissolved in 3 ml of 1 X TNE buffer (pH 8). The resuspended pellet was transferred to a bottle containing 4.62 gms of CsCl and salt was mixed by gentle shaking. The total volume of the mix was measured and raised to 4.5 ml with 1 X TNE and 0.5 ml of ethidium bromide (10 mg/ml stock) was added to this. After mixing, the solution was transferred to two Beckman ultracentrifugation tubes which were balanced using 1 X TNE and filled up to the frosted line. The tubes were then sealed using the cap heating machine and centrifuged in Beckman TL 100 ultracentrifuge for overnight at 100,000 rpm at 20 ^{o}C.

Next day after overnight centrifugation, the DNA bands in tubes were visualized under UV light source. If two DNA bands were observed there, both of them were collected because in that case the upper one might be open circular DNA. Bands were collected separately in clean tubes using a wide bore syringe after puncturing the tube with needle and carefully extracting the desired layer of DNA. To the separated layer collected in tube, an equal volume of CsCl/H_2O saturated solution of isopropanol was gently added and mixed thoroughly. This separated ethidium bromide as a separate pink coloured top layer which was carefully removed until no more pink colour was seen indicating that all ethidium bromide was removed. The remaining DNA solution was transferred to another fresh tube. To 400 µl of this DNA solution, 500 µl of distilled water, 100 µl of 3 M Sodium acetate and 530 µl of isopropanol were added in a 1.5 ml microfuge tube. Contents were mixed well and centrifuged at 10,000 rpm for 10 min at the room temperature. Supernatant was discarded and

pellet was re-dissolved in 200 µl of 1X TNE. 25 µl of 3 M Sodium acetate and 500 µl of ethanol were added to it and mixed. The mix was again centrifuged at 10,000 rpm for 10 min at the room temperature that resulted in pellet. This pellet was dried well and resuspended in 100 µl of 1:10 diluted TNE and stored at -20 ºC for further use.

2.4.3 Plasmid extraction using *AccuPrep*® kit (Bioneer)

Plasmid DNA isolation was done using *AccuPrep*® plasmid extraction kit (Bioneer) as per the protocol supplied with it by the manufacturer. Cells in the 5 ml of overnight cultures were pelleted by centrifugation at 14000×g for 5 min at room temperature. Supernatant was discarded and supernatant free cells were resuspended by vortexing in 250 µl of pre-chilled resuspension buffer (buffer1) supplied with the kit. 250 µl of lysis buffer (buffer 2) was added to this and the tube was gently inverted 3-4 times to mix the contents. Thereafter, 350 µl of neutralising buffer (buffer 3) was added immediately and mixed gently by 3-4 inversions before centrifugation at 14000×g for 10 min and 4 ºC to pellet the cell debris. The cleared lysate was transferred to the binding column supplied by the manufacturer and centrifuged at 14000×g for 1 min. The flow-through was discarded and DNA binding column was reassembled with the collection tube. Thereafter, 500 µl of denaturation buffer (buffer D) supplied by the manufacturer was added to this and incubation was done for 5 min before again centrifuging it for 1 min at 14000×g. The flow-through was again discarded and column reassembled with the collection tube. 700 µl of wash buffer supplied with the kit was added to this column and column was re-centrifuged for 1 min at 14000×g. Again flow through was discarded and column was refitted with the tube and centrifuged for 1 min at 14000×g. It was then left for 5 min to dry completely and refitted into fresh microfuge tube. Finally, 50-100 µl of elution buffer (buffer 5) was added to the column which was incubated for at least 1 min and thereafter centrifuged at 14000×g for 1 min at the room temperature to

collect the plasmid DNA suspension in a microfuge tube which was stored in freezer at -20 ^{o}C.

2.4.4 Plasmid extraction using Bioline kit

As for the Bioneer kit, plasmid DNA isolation was done following the protocol supplied with the Bioline kit by the manufacturer. 5 ml of overnight grown cell cultures were pelleted by centrifugation at 11000×g for 30 s at the room temperature. After discarding supernatant, bacteria were resuspended in 500 µl of pre-chilled resuspension buffer (buffer 1) supplied with the kit. Cell lysis was then done by adding 500 µl of denaturation buffer (buffer 2) and mixing gently by inverting the tube 6-8 times. Finally, 600 µl of neutralization buffer (buffer 3) was added and mixed gently by slowly inverting the tube 6-8 times before centrifugation at 11000×g for 10 min at room temperature. The cleared lysate was then filtered through the binding column supplied by the manufacturer, in batches of 750 µl followed by centrifugation at 11000×g for 1 min**.** Each time the flow through was discarded and the DNA binding column was reassembled with the collection tube. Once all the lysate was filtered through DNA binding column, 500 ml of wash buffer PW1 (pre-heated to 50 ^{o}C) was added to it and the column was centrifuged again at 11000×g for 1 min. After discarding flow through, column was refitted back in the tube. 600 µl of wash buffer PW2 (supplemented with ethanol) was added to the filter in the column which was centrifuged again at 11000×g for 1 min. The flow-through was discarded and filter was refitted into the column. The filter was then dried by centrifuging for 2 min at 11000×g to remove the residual ethanol. After drying the filter was shifted to fresh 1.5 ml micro-centrifuge tube. 50 µl of (pre-heated to 70 ^{o}C) elution buffer (buffer P) was then added to the filter and incubated for 2 min at 70 ^{o}C. Finally, plasmid DNA was collected by centrifugation for 1 min at 11000×g at room temperature and stored in freezer at -20 ^{o}C.

2.4.5 Restriction enzyme digests

Restriction enzymes from the laboratory stock were used that were either supplied by New England Biolabs (NEB), Fermentas or by Invitrogen. Digestions were made in two different volumes depending on whether the digests were to be recovered after gel-electrophoresis for further use or not. For recovery, every time 20 µl of plasmid DNA was used in a total volume of 40 µl of reaction mix while for other purposes like checking of insert 5-10 µl of plasmid DNA stock was used in a total volume of 20 µl of reaction mix. The volume of enzyme stock solution used in any case never exceeded 10% of the reaction mix. Appropriate concentrated buffer was used along with bovine serum albumin wherever it was required. The reaction mix was finally diluted to compensate for concentrated buffer by sterilized distilled water. All digestions were typically carried out for 2 hours at 37 ^{o}C.

2.4.6 Agarose gel electrophoresis

All DNA whether obtained by PCR or restriction digests were analysed by agarose gel electrophoresis. 1% agarose gels were made by dissolving 1 gm of agarose in 100 ml of TAE buffer (40 mM *Tris*-acetate, 1 mM EDTA pH 8.0). The agarose solution was prepared by heating in a microwave oven for 1 min, stirring and then heating again for 30 s. Before pouring it on the casting gel, it was allowed to cool and then 2 µl of ethidium bromide was added to it that allowed DNA visualization under UV light. The gel-casting tray was prepared by sealing its ends with tapes and inserting combs of different sizes as per the requirement of number and size of wells. After pouring, the gel was allowed to set for about 20 min at room temperature. DNA samples were prepared for loading by adding loading-dye in the ratio of 2 µl dye to 10 µl of DNA. Two types of loading dyes were used depending upon the requirement i.e. size of the DNA to be analysed. Top loading dye was composed of 0.25% w/v xylene cyanol and 15% w/v ficoll while bottom loading dye was composed of 0.25% w/v

bromophenol blue and 15% w/v ficoll. Samples were always loaded alongside 1 kb DNA ladder (Fermentas). Power for electrophoresis was supplied from power pack at about 100 V and 500 mA which took about 45 min to complete the process. DNA was finally visualized by UV transilluminator.

2.4.7 Extraction of DNA from agarose gel

DNA from electrophoresis gels was extracted by using Illustra GFX PCR DNA purification kit (GE Healthcare). The kit comprised of four solutions. The first one was the Capture buffer which was added to the gel pieces of interest excised from the electrophoresis gel under UV of longer wavelength in the ratio of 10 µl or more of capture buffer per 10 mg of gel slice. After adding Capture buffer, the mix was incubated for 15-30 min at 60 ^{o}C with intermittent mixing every 3 min by inverting the microfuge tube. This resulted in dissolution of the gel into capture buffer to give a yellow/pale coloured uniform solution that indicated correct pH of this solution which was effective for binding of DNA to the DNA binding column to be used in the next step. If the colour was not yellow or pale, that required an addition of 3M sodium acetate to adjust the pH to the level which promotes effective binding of DNA to the column. This solution was transferred to the DNA Binding column fixed in collection tube and incubation was done for 1 minute at the room temperature before centrifuging it for 30 s at 16000 rpm in a microfuge. Flow through was discarded and 500 µl of wash buffer was again added to the column and it was re-centrifuged for 30 s at 16000 rpm. Flow through was discarded again and the empty column was re-centrifuged for about a minute to dry the membrane in column on which DNA was bound. This DNA was finally eluted from the column by adding 15 µl of elution buffer and incubating the column for at least 1 min at the room temperature before centrifuging it for 1 min at 16000 rpm. Eluted DNA was stored at -20 ^{o}C for future use.

2.4.8 Addition of 'A'-overhangs to PCR products

Amplified DNA fragments obtained by using Velocity^{TM} or KOD DNA polymerase lacked 'A' tails –a stretch of 'A's at the 3' end of the amplified fragment which were required in order to facilitate their cloning into pGEM-T Easy vector for sequencing. Therefore 'A' tails were added to amplified DNA fragments after subjecting them to agarose electrophoresis and eluting it from gel as explained earlier. Thereafter, purified PCR product was treated with 1 µl of *Taq* polymerase, 10 µl of *Taq* polymerase buffer, 2 µl of $MgCl_2$ (all the three of Invitrogen) and 2 mM dATPs. The volume of mix was made up to 100 µl with sterilized distilled water and heated at 70 ^{o}C for 30 min. The 'A'-tailed PCR product was again run on agarose gel and extracted from it for ligation into pGEM-T Easy vector.

2.4.9 DNA ligations

These were performed using T4 DNA ligase (Invitrogen) and incubating the reaction mix at 4 ^{o}C for overnight. The concentration (ng/µl) of insert and vector digested with the same restriction enzyme(s) was estimated by comparing with the respective bands of the 1 kb marker ladder (Fermentas) run along with them and the digested DNA to be ligated was eluted from the gel. The ratio of insert to vector concentration in the ligation reaction was maintained approximately at 3:1 in the standard reaction volume of 10 µl. Using the following formula, quantities of vector and insert DNA to be used in ligation reaction were estimated.

$$\frac{\text{ng of vector X size of insert(kb)}}{\text{size of vector (kb)}} \text{ X } \frac{3 \text{ insert}}{1 \text{ vector}} = \text{ng of insert DNA}$$

2.4.10 Preparation of competent cells of *E. coli*

Competent cells were prepared as per the calcium chloride method of Cohen et al (1972), by which a culture of *E. coli* was set up at 37 ^{o}C in 100 ml of L broth without antibiotic selection

using an overnight grown (at 37 ºC) seed culture. Inoculation was done in the ratio of 1-part seed culture in 100 part of L broth. The culture was incubated for about 2 hr in a shaker at 200 rpm until OD_{600} reached at 0.4-0.6. The culture was centrifuged at 5000 rpm at 4 ºC for 7 min. Supernatant was discarded and pellet of cells was resuspended by vortex in pre-chilled 100 mM $CaCl_2$ added in the ratio of 2 ml of it per 5 ml of starting culture. After incubation on ice for 20 min, the cell suspension was re-centrifuged at 5000 rpm at 4 ºC for 7 min. Supernatant was again discarded and cells were gently resuspended in pre-chilled 100 mM $CaCl_2$ containing 15% v/v glycerol, this time in the ratio of 0.5 ml for each 5 ml of starting culture. The entire suspension was divided into aliquots of 100 µl in sterile microfuge tubes which were frozen in liquid nitrogen before being stored at -80 ºC.

2.4.11 Bacterial transformation by DNA and selection of transformants

Plasmid DNA (20 ng) was added to 100 µl of ice cold competent cells and incubated on ice for 30 min followed by a 2 min heat shock in a water bath maintained at 42 ºC. Thereafter, cells were again kept on ice for 5 min before 1 ml of L broth medium was added and incubation was carried out at 37 ºC for 1 hour. After 1 hour the cell suspension was plated in two volumes of 50-200 µl on separate agar plates with suitable selection medium i.e. appropriate antibiotic and/or indicator like X-gal/IPTG (for X-gal assay). The X-gal assay was used to detect transformed bacterial cells in which the pGEM-T Easy vector has an insert. Detection relied upon disrupting expression of the gene encoding the lacZα fragment because of the insertion of PCR product. An undisrupted *lacZα* gene that encoded the fragment in combination with the chromosomally-encoded β-fragment metabolizes the X-gal substrate into blue coloured metabolite allowing one to screen colonies on the basis of their colour –dark blue, light blue or white.

2.4.12 DNA sequencing

The sequence of PCR amplified genes/DNA fragments cloned into pGEM-T Easy vector was confirmed by DNA sequencing using universal primers in combination with the Sanger dideoxy method (Sanger et al., 1977) carried out by the Functional Genomics Laboratory of the University of Birmingham on ABI 3700 analyser. 5 ml overnight cultures were used for plasmid DNA extraction using Bioneer *AccuPrep*® plasmid extraction kit. The concentration of the plasmid DNA preparation which was submitted for sequencing, was determined by nanodrop spectrophotometer (NanoDrop ND-100). Typically, 200-500 ng of plasmid DNA was used for the reaction along with 3-4 picomoles of primer in a final volume of 10 µl made up with sterile distilled water. Sequences were analysed using Chromas Lite (http://www.technelysium.com.au/chromas_lite.html).

2.4.13 Sequence analysis

Pairwise alignments of sequences of proteins and DNA were done using EMBOSS Needle (http://www.ebi.ac.uk/Tools/psa/emboss_needle/) (Rice et al., 2000), and BLAST programmes (https://blast.ncbi.nlm.nih.gov/Blast.cgi) (Altschul et al., 1990). Local alignment of proteins was also performed using EMBOSS Water programme (http://www.ebi.ac.uk/Tools/psa/emboss_water/) (Rice et al., 2000). Default parameters were used for all the alignments that were performed online. These programmes were also used to verify sequences of cloned fragments obtained by DNA sequencing. Databases of National Centre for Biotechnology (www.ncbi.nlm.nih.gov) and European Bioinformatics Institute (http://www.ebi.ac.uk/) were used for obtaining reference (wild type) nucleotide and protein sequences of interest. Artemis software (http://www.sanger.ac.uk/resources/software/artemis/) was also used for the analysis of sequences (Rutherford et al., 2000). Primers were designed and analysed using either

NetPrimer (www.premierbiosoft.com/netprimer) or OligoCal software programme of Northwestern University of USA (http://www.basic.northwestern.edu/biotools/oligocalc.html). Programme NEB cutter2 and other tools of New England Biolabs were used for mapping of restriction sites (http://tools.neb.com/NEBcutter2/).

2.4.14 Bi-parental mating (conjugation) and suicide vector excision

This was used for transferring/mobilization of plasmid DNA carrying cloned inserts into *Pseudomonas fluorescens* strains or *Pseudoalteromonas sp.* SANK 73390. These plasmids were either expression vector pJH10 (Figure 2.4) or the suicide vector pAKE604 (Figure 2.2) that carried cloned inserts of interest. For this purpose, *E. coli* S17-1 strain was used which was transformed with these plasmids. 0.5 ml of late exponential phase cultures of each of the mating strains were mixed in a sterile bottle. Alternatively, the two cultures to be mated were filtered through 0.45 µM Millipore filter. For mating with *Pseudomonas fluorescens* strains, the mix of two cultures was plated on L agar plate without any antibiotic (or the filter was transferred to it) that was incubated at 30 ^{o}C for overnight for conjugation to happen. Thereafter, cells were scratched from the lawn of bacteria on plate or washed from the filter and suspended in 1 ml saline (0.85% NaCl). Mixture was vortexed and serial dilutions were prepared. 100 µl of each dilution was plated on M9 minimal media plates supplemented with tetracycline or kanamycin. Minimal media was made by mixing 200 ml of salt solution consisting of Na_2HP0_4 (6 g/l), KH_2P0_4 (3 g/l), NH_4Cl (1 g/l), $MgSO_4$ (1 mM), thiamine HCl (1 mM), $CaCl_2$ (0.1 mM) and glucose (0.2%) with 200 ml of 50% H_2O agar. Plates were incubated at 30 ^{o}C for 2 days. Thereafter, single colonies that appeared were picked and restreaked on M9 minimal media plates with antibiotics (tetracycline or kanamycin) and

again incubated at 30 ºC for 3-4 days. Finally, single colonies were purified from these plates by restreaking onto L agar plates with tetracycline and ampicillin as supplement.

Suicide vectors designed and constructed as described below were mobilized into the target organism by conjugation and transconjugants were selected and purified as described above for the expression vector. Integrant colonies among transconjugants were selected on minimal media plates supplemented with antibiotic kanamycin which were purified by restreaking on L agar plates. Sucrose sensitivity was checked by streaking kanamycin resistant colonies onto L agar plates supplemented with 5.5% w/v sucrose. For excision to take place, overnight cultures of integrants were set in 5 ml L broth without any antibiotic selection at 30 ºC in a shaker at 200 rpm. Serial dilutions of this overnight culture were made in L broth to give 10^{-1}-10^{-5} dilutions which were plated on L agar plates containing 5.5% w/v sucrose and incubated for 3-4 days at 30 ºC, for the selection of strains in which integrants had excised. Single colonies from these plates were initially purified by streaking on L agar with ampicillin and by incubating for 3-4 days at 30 ºC. From these, isolated single colonies were picked up and were simultaneously restreaked onto L agar plates supplemented with ampicillin and on L agar plates supplemented with kanamycin, which were again incubated for 3-4 days at 30 ºC. Finally, ampicillin resistant and kanamycin sensitive colonies were selected which were subjected to the confirmation of desired mutant genotype by PCR.

In case of *Pseudoalteromonas sp.* SANK 73390, the methods for conjugation and suicide vector excision were same except as explained in Chapter 3 (under section 3.2).

2.4.15 Strategy for constructing in-frame deletions

The strategy for creating in-frame deletions for removing large proportions of the target genes to render them non-functional involved amplifying two arms of about 500 bp,

overlapping with the ends of the target gene to be deleted (Figure 2.1). This strategy was used by Rahman et al., 2005 to create defined mutants of *Pseudomonas fluorescens* NCIMB 10586. Deletions need to be made in-frame to ensure that the implications of polar effects, if any, are minimal. This in turn ensured that the resulting phenotype was just the effect of lacking the specific gene which was targeted for deletion. Primers that were used to amplify these arms using polymerase chain reaction, were designed to have a unique restriction site at their 5' ends and common restriction sites at the other ends. These fragments were then cloned into pGEM-T Easy vector and sequenced. Thereafter, these arms were cut out from clones that had got the correct sequence, purified and joined together using common restriction site to get a fragment of 1 kb which was then cloned into vector pAKE604 (El-Sayed, 2001), and this resulting plasmid was transformed into *E. coli* S17-1 cells for transfer by conjugation into wild type target organism. This suicide vector was integrated into the host chromosome (or the targeted plasmid pTML1 in case of *Pseudoalteromonas sp.* SANK 73390) once after inside the host cell, via homologous recombination between two 500 bp arms that it carried and their cellular (chromosomal or plasmid) homologues (Figure 2.3). The suicide plasmid that got integrated, carried the s*acB gene* which encoded enzyme *levansucrase* that catalysed polymerisation of sucrose resulting in *levan* which was lethal to the cell. For this reason, those cells that had this plasmid integrated into their chromosome (or the targeted plasmid pTML1), became sensitive to media supplemented with sucrose. A second recombination event could occur between homologous regions resulting into the excision of suicide plasmid from the site of its integration in the host cell, thereby causing cells to become insensitive to sucrose. The probability of the second recombination event occurring in either arm should be equal so that half of the recombination events will leave the target gene intact and the other half produce mutants with in-frame deletion which could be checked by PCR.

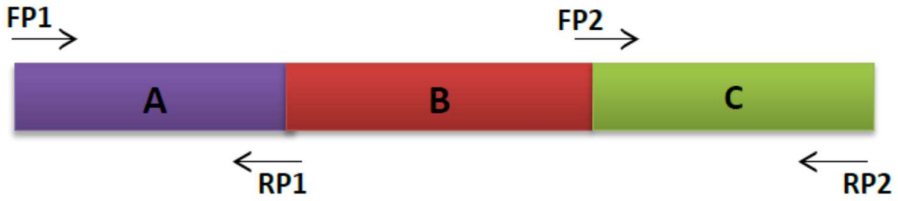


Figure 2.1 Strategy for PCR amplification of 500 base pair arms flanking the gene targeted for deletion. The arms A and C flank the gene B that was targeted for deletion. (Arrows indicate position of forward or reverse primers indicated as FP and RP respectively for the amplification of arms).

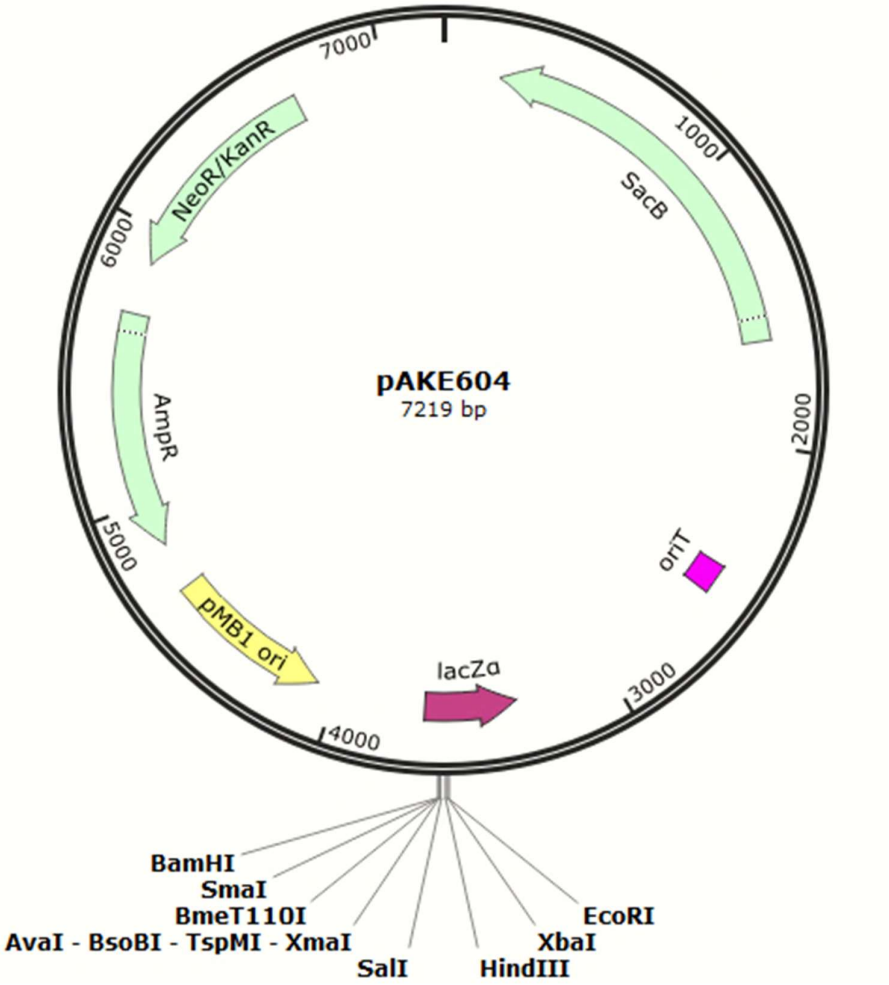


Figure 2.2 Map of suicide vector pAKE604 used for creating specific vectors that were used for making defined deletions(El Sayeed et al., 2001). Created using SnapGene®.

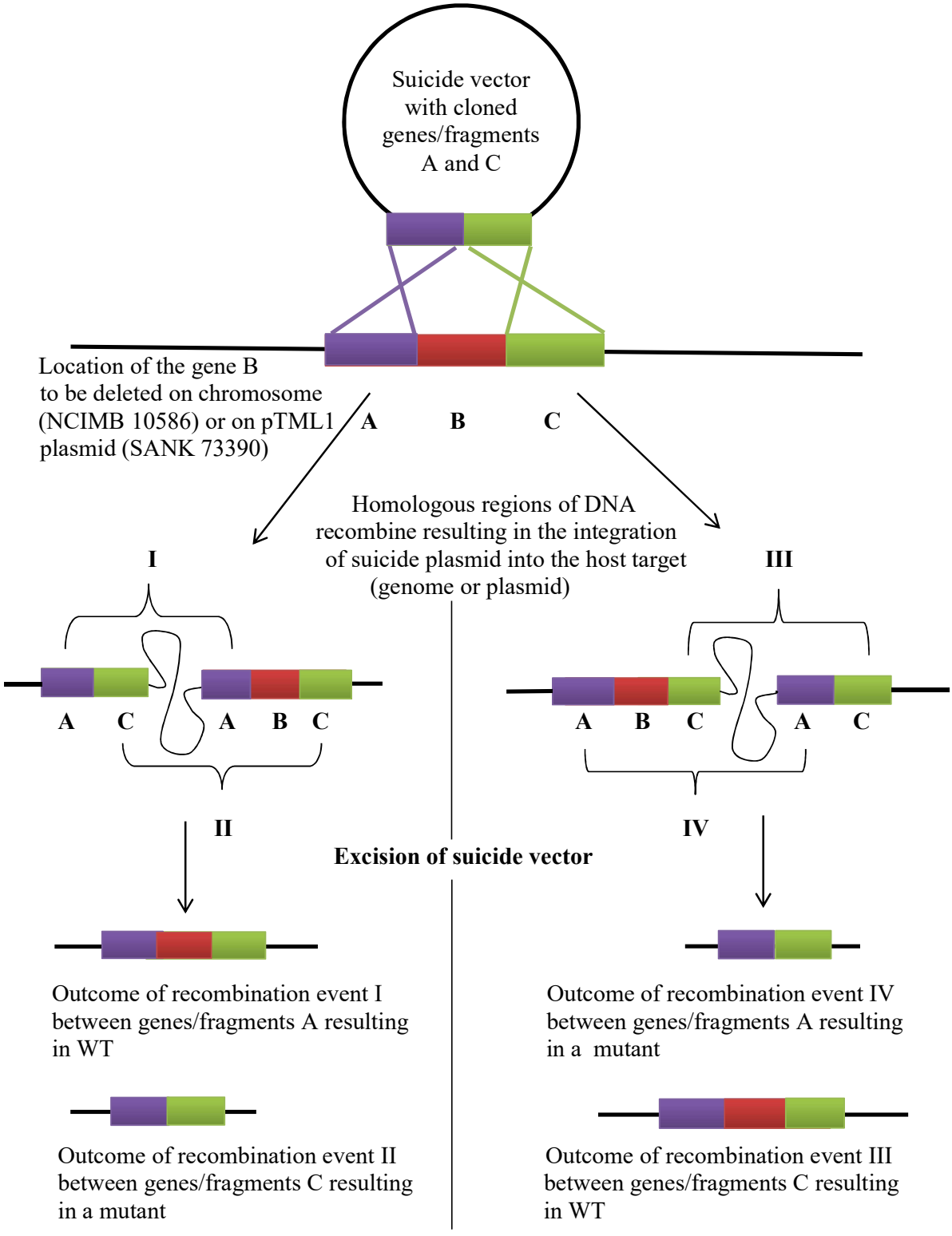


Figure 2.3 Scheme depicting outcomes of suicide vector integration and exclusion in the host target DNA used for the creation of in-frame deletions. Suicide vector carried arms flanking the gene targeted for deletion.

2.5 Construction of expression plasmids

The gene of interest was PCR-amplified and cloned into pGEM-T Easy vector (promega) after 'A'-tailing. The cloned gene with pGEM-T Easy vector (the ligation) mix was transformed into *E.coli* DH5*α* and transformants were selected on X-gal (80 µg/ml) with IPTG (0.5 mM) and ampicillin (50 µg/ml) plates for blue-white screening. Plasmid DNA isolated from selected white colonies was digested with those restriction enzymes sites for which were present on both sides of the multiple cloning site of pGEM-T Easy vector for example *EcoR*I. Thereafter, the cloned gene was sequenced using universal primers. The insert from a clone carrying the correct gene sequence was excised from pGEM-T Easy vector by digestion and ligated into expression vector (pJH10) (Figure 2.4) using restriction enzymes and T4 DNA ligase respectively. The ligation mix (vector and gene) was transformed into *E. coli* DH5α cells and was selected on L agar antibiotic plates. Plasmid DNA from colonies selected in this way was checked for inserted genes by restriction digestion or by colony PCR.

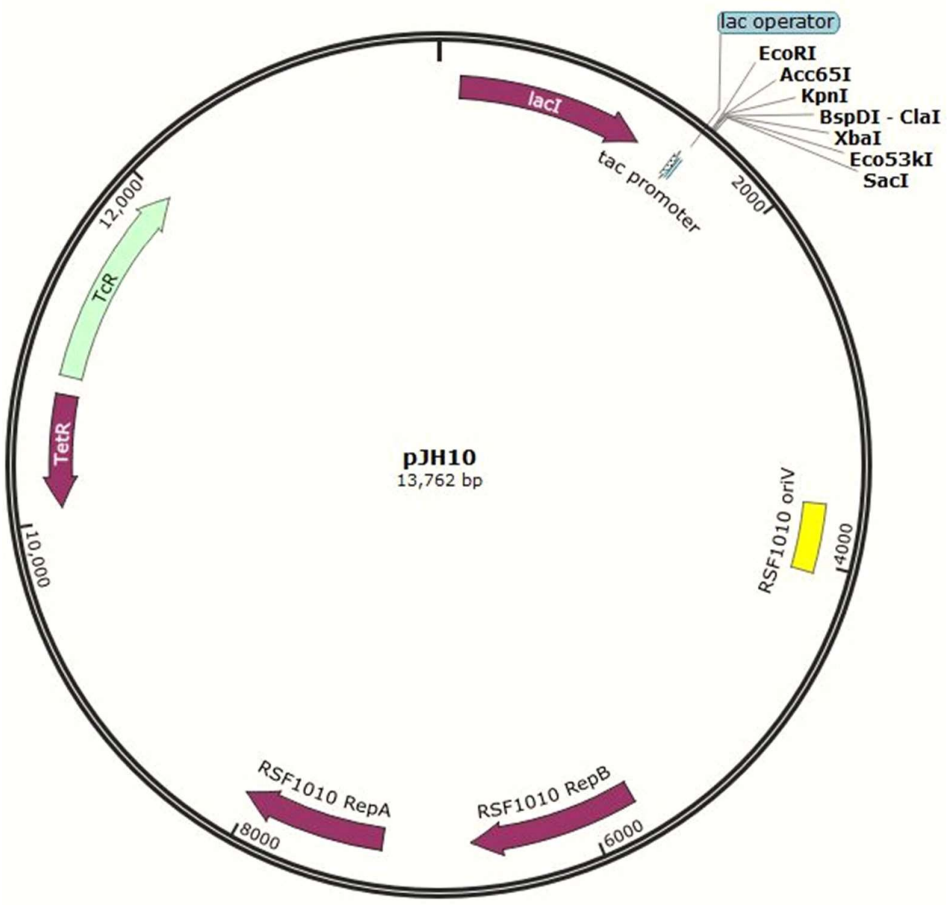


Figure 2.4 Map of expression vector pJH10 used to derive *in trans* expression of thiomarinol proteins(El Sayeed et al., 2003). Created using SnapGene®.

2.6 Bioassay for mupirocin production

For each bioassay, single colonies of *Pseudomonas fluorescens* (constructed test strains, wild type NCIMB 10586, control strains i.e. mutant strains of genes being tested transformed with pJH10 plasmid) were inoculated, each in 5 ml L broth with tetracycline for test strains and controls with pJH10 plasmid, and ampicillin for all the other control strains and for wild type. Cultures were incubated for overnight at 30 ^{o}C. Cell density was measured by UV spectrophotometer at 600 nm. After normalising cell density of all cultures, spot cultures were laid on L agar plates (poured with measured volume) which were left on the bench for overnight incubation. Overnight culture of *Bacillus subtilis* 1064 grown in L broth at 37 ^{o}C to late log phase was used to test mupirocin production. Overnight grown spot cultures were overlaid with equal volumes of molten L agar (100 ml) mixed with 4 ml of *Bacillus subtilis* 1064 culture and 0.5 ml of 5% w/v solution of 2,3,5 triphenyl tetrazolium chloride (TTC). These overlaid plates were incubated for overnight at 37 ^{o}C. The area of zone of clearance around spot culture was measured and was taken as an estimate of mupirocin production by concerned strains.

2.7 High performance liquid chromatography (HPLC)

2.7.1 Culture preparation for HPLC

Overnight (16 hours) seed cultures were set up in 5 ml of L broth at 25 ^{o}C, 200 rpm. Thereafter, 25 ml of culture was set up at 22 ^{o}C, 200 rpm for 40 hours in fresh secondary stage medium (SSM) using 1.25 ml of this seed culture that was used for inoculation. SSM broth comprised of 25 gm Soya flour, 2.5 gm spray dried grown flour, 5.0 gm $(NH_4)_2SO_4$, 0.5 gm $MgSO_4.7H_2O$, 1.0 gm Na_2HPO_4, 1.5 gm K_2HPO_4, 1.09 gm KCl, 6.25 gm $CaCO_3$ and 50

ml glucose, all made up to 1 litre with distilled water. After 40 hours of incubation, cultures were centrifuged at 13000 rpm for 7 min. Supernatant was collected and its pH was adjusted to 4.5 using HCl before being stored at -20 ºC. Supernatant was filtered using 0.2 µm PTFE syringe filters.

2.7.2 Sample analysis by HPLC

HPLC was performed using Unipoint LC system software reverse phase C18 column (15 cm × 4.6 mm) with UV detection (Gilson) at 233 nm, and mobile phase water/ acetonitrile gradient (5-70%) using acetonitrile with 0.01% triflouroacetic acid, over 60 min at 1 ml/ min rate of flow.

2.8 Liquid chromatography mass spectroscopy (LCMS)

This was performed by the collaborator of our group (Dr. Zhongshu Song) at the University of Bristol, UK. As per the procedure followed by him, L agar plates were inoculated with test strains and incubated at 30 °C for 30 hours. A single colony from each mutant and wild type strain was picked from the agar plates and inoculated into 12 ml of L broth with carbenicillin (50 µg/ ml culture) in a 30 ml universal tube. The seed cultures were incubated at 25 °C and 200 rpm overnight. 5 ml of the seed culture was inoculated into each 500 ml Erlenmeyer flask containing 100 ml of L broth medium enriched with 4% glucose. Two flasks for each strain were used in the experiment. These fermentation flasks were incubated at 22 °C and 220 rpm for 48 hours. After the fermentation finished all the flasks were examined for pH values, cultural outlook and smell for any sign of contamination. Cultures from the 2 flasks for each strain were combined and subjected to chemical extraction. 500 ml to 1 litre of cultures were incubated for target product purification.

The cells were removed by centrifugation at 22,000 ×g for 20 min. The supernatant was then extracted by ethyl acetate (1:1) once, followed by an extra ethyl acetate extraction after the aqueous was acidified to pH 5.0. The two extracts were combined and ethyl acetate was evaporated *in vacuo*. The residue was collected by MeOH for LCMS analysis. All the samples were dissolved in 4.0 ml MeOH. Analytical samples were prepared by 10-fold dilution with MeOH and analysed by LCMS using a Waters HPLC system (Waters 2545 Binary Gradient Module and Waters SFO System Fluidics Organizer). Detection was achieved by UV between 200 and 400 nm using a Waters 2998 diode array detector, and by simultaneous electrospray (ES) mass spectrometry using a Waters Quattro Micro™ API spectrometer detecting between 150 and 600 *m/z* units. Chromatography (flow rate 1 ml·min^{-1}) was achieved using Phenomenex Kinetex column (5 μ, C_{18}, 100 Å, 4.6 × 250 mm). Solvents were: (A) HPLC grade H_2O containing 0.05% formic acid; (B) HPLC grade CH_3CN containing 0.045% formic acid. Analytical gradients were as follows: 0 min, 5% B; 22 min, 60% B; 24 min, 95% B; 26 min, 95% B; 27 min, 5% B; 30 min, 5% B. LCMS purification of target compounds was performed on prep Phenomenex Kinetex column (5 μ, C_{18}, 100 Å, AXIA 21.2 × 250 mm). Collection of target peaks was triggered by masses. Both positive (PI) and negative (NI) mode were employed for the characterization of the target compounds together with UV absorption pattern.

CHAPTER 3

Investigations on the role of tailoring/auxiliary genes of the *tml* cluster in the biosynthesis of thiomarinol

3.1 Introduction

As explained in Chapter 1, thiomarinols are more effective than mupirocin in antibacterial properties and therefore can be a potent antibiotic or can be used to develop a more efficient antibiotic. pTML1 is the 97 kb plasmid that is naturally present in marine bacterium *Pseudoalteromonas rava* SANK 73990 that produces thiomarinols. Fukuda et al., (2011) had proposed that all the genes that are needed to code for thiomarinols are entirely found on this plasmid. Genetic annotation of thiomarinol cluster on pTML1 has showed that it consists of 45 ORFs (Fukuda et al., 2011).

Very little is known about biosynthetic steps in the thiomarinol biosynthetic pathway. But by sequence homology it has been shown that 27 out of total 45 ORFs of the gene functions of thiomarinol cluster have significant amino acid identity with those of mupirocin biosynthetic cluster of *Pseudomonas fluorescens* NCIMB 10586. Bioinformatics analysis has shown that there has been clear conservation of PKS and associated gene functions between the thiomarinol and mupirocin gene clusters, which is also reflected as striking similarity in the structure of these two molecules (Fukuda et al., 2011). Sequence homology indicates that the thiomarinol cluster, like mupirocin biosynthetic cluster, also has an unusually large number of accessory/tailoring region genes (Thomas et al., 2010, Fukuda et al., 2011).

By comprehensive in-frame deletion studies in the accessory/tailoring open reading frames of mupirocin biosynthetic cluster, functions of most of these ORFs have been deduced (Hothersall et al., 2007). By *in trans* complementation studies of defined knockouts of mupirocin biosynthetic cluster, it has been shown that all of these are required for mupirocin production with role of most of the specific genes in the biosynthetic pathway determined (Hothersall et al., 2007). While in the *tml* cluster, gene knockout studies have not been so extensive. Yet, by gene knockout studies by Fukuda et al. (2011) in *tml* cluster, carried out

using the pAKE604 suicide vector, it has been shown that *tmpD* (PKS) mutant produced only the acyl pyrrothine while mutation in *holA* (NRPS) resulted in the production of marinolic acid only. Mutation in *tmlU* leads to the production of both marinolic acid and acyl pyrrothines but no thiomarinol was produced (Fukuda et al., 2011). Using the same strategy knockout of *tmlF* accessory/tailoring gene function of thiomarinol gene cluster could also be made recently by Thomas laboratory at the University of Birmingham which is under study for any changes in the phenotype that this specific knockout might cause.

By applying similar approach of creating logically defined in-frame mutations in the various accessory/tailoring functions of thiomarinol gene cluster and studying the resulting phenotype of mutants it should be possible to deduce steps in the biosynthetic pathway of thiomarinols. Each of the defined deletion of a gene function may result in non-production or deficient production of thiomarinols. And it may also not affect production of thiomarinols at all. It is also possible that a new metabolite or a novel derivative of thiomarinol is produced by the mutant. This way, by characterizing the product it should be possible to pin point the physiological role for which a particular specific gene function is responsible if the same effect is not caused as a result of actions of two or more gene functions that act simultaneously on the metabolic intermediate in the pathway. Any of these many situations depend on the specific contribution made by each gene of the tailoring/accessory region in the biosynthetic pathway of thiomarinols. It will also depend upon the relative role like specific catalytic function played by the specific gene function in the biosynthetic pathway that is mutated. At the same time, by *in trans* expression of the knocked out function in the mutant strain, it can be ensured that the resulting phenotype of the mutant strain is not caused by polar effect of gene expression.

The importance of accessory/tailoring genes has been shown in the biosynthesis of active secondary metabolites. Their role has been shown as accessory genes in the intermediate biosynthetic steps in the biosynthesis of daptomycin, lovastatin and mupirocin (Kennedy et al., 1999, Nguyen et al., 2006, Hothersall et al., 2007). While their role as post PKS tailoring- genes has been shown in the biosynthesis of erythromycin and anthracyclines (Katz, 1997, Weymouth-Wilson, 1997, Weissman and Leadlay, 2005). The importance of manipulating tailoring enzymes has been shown in the development of molecules with novel activities. It has been shown that minor modifications in the tailoring region genes have resulted in novel activities (Katz and Donadio, 1993, Rix et al., 2002, Olano et al., 2010).

This work was aimed at investigating gene functions of specific tailoring region genes of the thiomarinol cluster by defined in-frame deletion of singular genes. The initial knockouts attempted were of *tmlT, tmlS, tmlO* and *tmlW* with an aim of extending such approach towards understanding roles of *tacpD, tmlQ* and *tmpF* as well.

By *in silico* studies it has been predicted that *tmlT* and *tmlW* are ferredoxin and dioxygenases, respectively. By knockout and complementation studies it has been shown that *mupT* and *mupW,* with which *tmlT* and *tmlW* share significant amino acid sequence identity of 39% and 56% respectively, are involved in the closure of pyran ring found in the polyketide component of mupirocin (monic acid). It has been shown that formation of pyran ring involved attack of the 5-OH onto an activated 16-hydroxy group to finalise (close) the pyran ring (Cooper et al., 2005b, Hothersall et al., 2007, Thomas et al., 2010). Mutants of *mupW* of *Pseudomonas fluorescens* NCIMB 10586 resulted in the isolation of PA-W which is essentially pseudomonic acid but with open pyran ring (Cooper et al., 2005a) while mutants of *mupT* resulted in significant loss in the production of pseudomonic acids (Cooper et al., 2005b). On this basis and in view of bioinformatics analysis it could be predicted that *tmlW* knock outs of thiomarinol producer may result into thiomarinol molecules without the

completed pyran ring that is also present in it as part of polyketide component of the molecule that has similarity with monic acid component of Pseudomonic acids (Fukuda et al., 2011). Knockouts of *tmlT* should also drastically affect the production and/or activity of thiomarinols.

Similarly, knockout of *tmlO* which is predicted to be a cytochrome P450 enzyme, is expected to result in changes in the oxidation state around the pyran ring found in polyketide moiety of thiomarinols. Which is because this enzyme shared significant amino acid sequence identity with *mupO* (44%), a cytochrome P450 enzyme and the knockout yielded the same result in the mupirocin biosynthesis (Cooper et al., 2005b).

tmlS, tmlQ, tacpD and *tmpF* were the other accessory/tailoring genes, knockouts of which were planned to be made. On the basis of bioinformatics studies *tmlS, tmlQ, tacpD* and *tmpF* are predicted, like equivalent Mup proteins, to be 3-oxoacyl-ACP reductase, acyl-CoA synthase, acyl carrier protein and ketosynthase domain, respectively. MupS, MupQ, MacpD and MmpF have been found to share amino acid identity of 62%, 46%, 46%, and 43% with their corresponding Tml protein respectively. *In silico* studies indicated that proteins like MupS are often found to act together with other proteins in the furnishing of starter unit/primer unit for the biosynthesis of fatty acids (Wu et al., 2008). Gene functions of such proteins that act together are also found located together like such genes of defficidin polyketide system of *Bacillus amyloliquifaciens* (Chen et al., 2006) in which these are also found located together as in *mup* cluster (Thomas et al., 2010). Knockouts of *mupS, mupQ, macpD* and *mmpF* have been reported to result in abolition of production of pseudomonic acids (or in steep decrease in their production) as confirmed by plate bioassay and HPLC (Cooper et al., 2005b, Hothersall et al., 2007). Therefore, based on bioinformatics evidence and structural similarities in these two molecules, mupirocin and thiomarinol, knocking out *tmlS* may result in abolition of the production of thiomarinols. Similarly, mutation in any of

the genes *tmlQ, tacpD* and *tmpF* might result in the abolition or drastic decrease in the production of thiomarinols by the corresponding mutants.

To confirm that the engineered mutation was the only cause of change in the phenotype of the mutant strain and was not because of polar effects, it may be complemented by *in trans* expression of the knocked out gene. If there was no polar effect, then this should have restored the wild type phenotype in the mutant strain.

3.2 Results

Objective of this study was to create defined in-frame deletions of specific genes in the thiomarinol cluster of wild type *Pseudoalteromonas sp.* SANK 73390 present on the pTML1 plasmid and to study the effect of each deletion on the phenotype. Defined knockouts of *tmlT, tmlW, tmlO, tmlS, tacpD, tmpF* and *tmlQ* were planned to be made.

As described in Chapter 2 (section 2.4.14) suicide vector insertion and excision technique that depended on homologues recombination taking place, was used in order to delete each gene. First of all, two arms of DNA of about 500 base pairs on either side of each of these genes to be deleted were identified/ designed. Unique restriction sites were identified to facilitate ligation of these arms with each other and to incorporate joined arms into the suicide vector pAKE604 (El-Sayed et al., 2001). Scheme for the deletion of each of these seven genes is outlined in Figures 3.1 to 3.7. Primers were designed accordingly to include selected restriction sites in the arms. Details of all the primers designed to amplify arms to delete above said seven genes are given in Table 2.9 giving name of each primer, size, 5'-coordinate, incorporated restrictions sites etc. Arms were then amplified by PCR using specific pair of primers for each of the arm designed for the deletion of a specific gene.

PCR amplified arms were A-tailed as described in Chapter 2, to facilitate their cloning into pGEM-T Easy vector, in which these were cloned to confirm their sequence by sequencing using universal primers. All the pGEM-T Easy derivative vectors that were constructed for the deletion of *tmlT, tmlW, tmlO, tmlS, tacpD, tmpF* and *tmlQ* genes are listed in Table 2.5 which gives name and size of each plasmid constructed this way with the size of each arm that was cloned into it.

Following sequencing and cloning into pGEM-T Easy vector of each of the two arms of a pair designed to delete a specific gene, these were cloned into suicide vector pAKE604

by three-way ligation as a fragment of size about 1 kb using restriction sites that were incorporated during the PCR amplification of respective pair of arms. Table 2.5 also details suicide vectors made for the deletion of genes *tmlW, tmlT, tmlS and tmlO* giving names of suicide vectors, size of specific pair of arms cloned, the restriction sites into which it is cloned and the size of vector.

Suicide vectors so constructed were transformed into *E. coli* S 17-1 in order to transfer the constructed specific suicide vector into wild type *Pseudoalteromonas sp.* SANK 73390 by way of mobilization through conjugation as explained in Chapter 2. Briefly, for mating 5 ml cultures of *E. coli* S 17-1 strain carrying constructed suicide vector of interest were grown overnight at 37 ^{o}C in L broth medium supplemented with kanamycin (50 µg/ ml). Similarly, 5 ml culture of wild type *Pseudoalteromonas sp.* SANK 73390 was set up in 5 ml of M broth and allowed to grow at 23 ^{o}C for 1-2 days until there was dense growth. Thereafter, 1 ml of both the cultures was mixed and vortexed for 10 sec. 100 µl of the mix was plated onto M agar plate with no antibiotic selection and was incubated for 2-3 days at 23 ^{o}C to allow conjugation to takes place. After incubation, serial dilutions were prepared in M broth for up to 10^{-5} fold and 100 µl of the serial dilutions were plated onto M agar plates with kanamycin, for the selection of wild type *Pseudoalteromonas sp.* SANK 73390 with suicide vector integrated in pTML1 plasmid (first homologous recombination event). That is because in the first place suicide vector cannot replicate by itself and secondly, it has got kanamycin resistance in its genotype. This ensured that the selected transconjugants had suicide vector integrated into the wild type pTML1 at sites dictated by homology of specific two arms incorporated in the suicide vector. Plates were incubated for up to 2-3 days at 23 ^{o}C for selection to take place. Single colonies that were typical of wild type *Pseudoalteromonas sp.* SANK 73390 were then re-streaked on M agar plates supplemented with kanamycin (50 µg/ml) and incubated at 23 ^{o}C for another 2-3 days. Using these purified single colonies,

cultures were inoculated in 5 ml M broth without antibiotic to allow excision of the suicide vector integrated into plasmid pTML1 to take place (second homologues recombination event). Cultures were incubated for 1-2 days at 23 ^{o}C in the shaker. Thereafter, serial dilutions were prepared in M broth for up to 10^{-5} fold and 100 µl of serial dilutions were plated onto M agar plates containing sucrose (5.5% w/v) to allow counter selection of those transconjugants in which suicide vector had excised from pTML1. In order to allow this to happen, plates were incubated for 2-3 days at 23 ^{o}C. Colonies having integrated suicide vector did not grow because of the presence of sucrose in the medium which was polymerised by the product of *sacB* gene present in the integrated suicide vector, into solid *levan-sucrose* accumulation of which in the periplasm of the cell caused cell death. This ensured that colonies that grew had no suicide vector in their genotype. Single colonies were then selected from these plates and checked for their sensitivity to kanamycin. Sensitivity to kanamycin confirmed excision of the suicide plasmid from the pTML1. Such colonies that grew on sucrose but did not grow on kanamycin were then checked for the desired mutation by PCR using specific pair of outer primers that were used to amplify arms designed for deletion of a particular gene. Wild type *Pseudoalteromonas sp.* SANK 73390 and specific suicide vector were used as negative and positive controls, respectively, to confirm the desired mutation. Results for the knock-outs of *tmlO* and *tmlW* turned out to be negative as confirmed by of PCR because band of desired size could not be seen in any case. In case of knockouts of *tmlS* and *tmlT,* no colony that was resistant to the presence of sucrose in the culture media (and sensitive to the kanamycin) could be obtained indicating failure of second recombination event (excision) to take place.

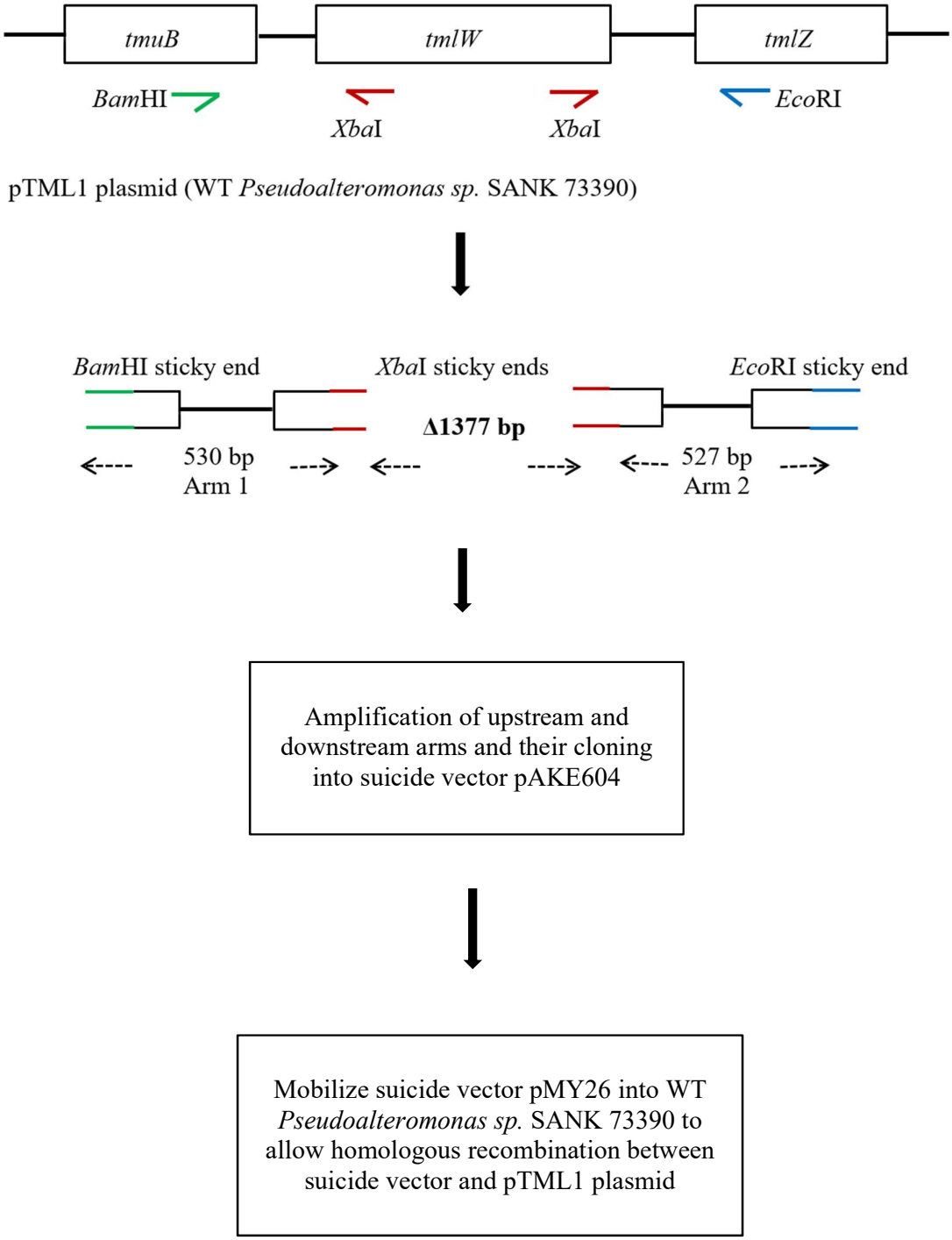


Figure 3.1 Construction for the deletion of *tmlW* gene from pTML1 plasmid of *Pseudoalteromonas sp.* SANK 73390.

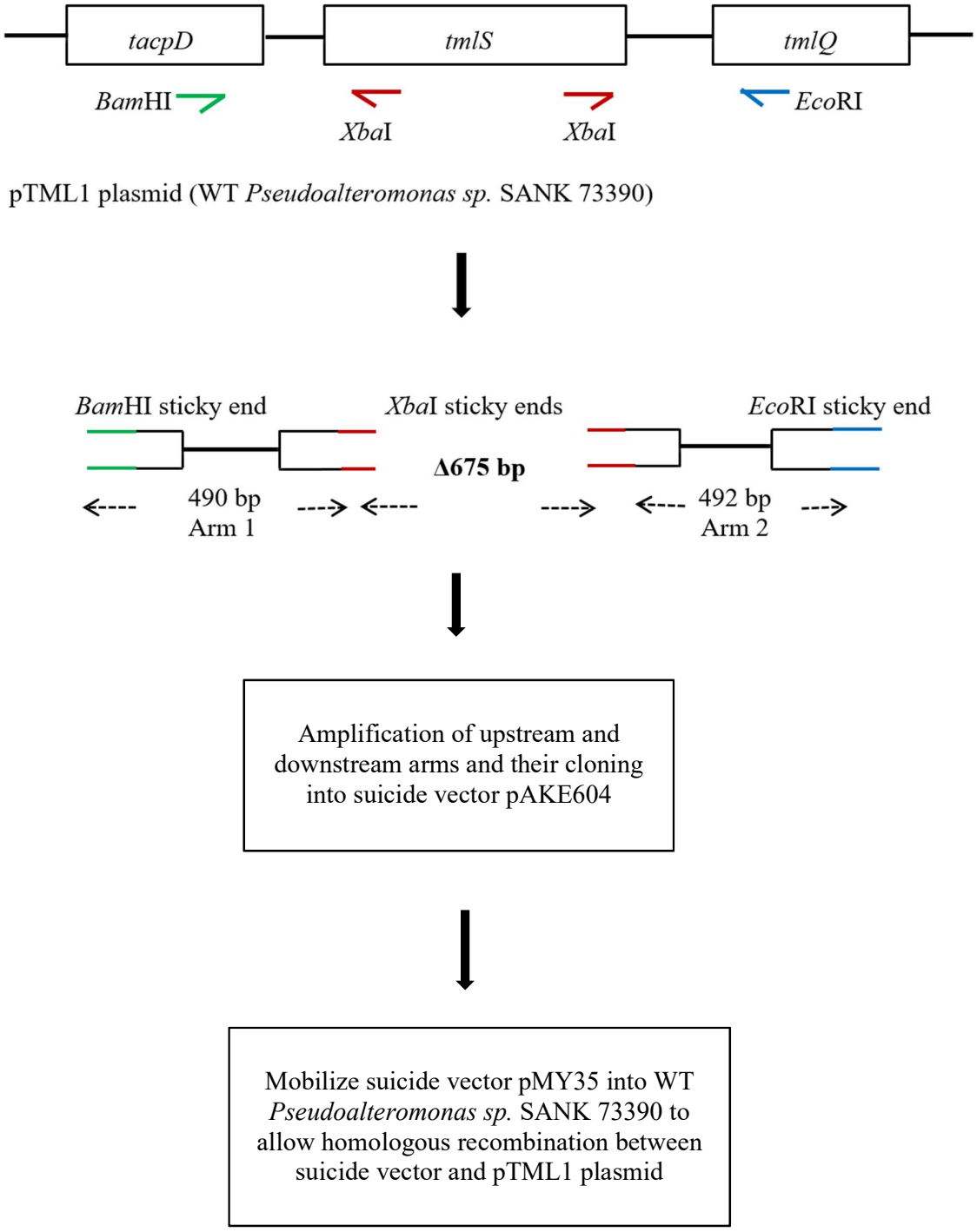
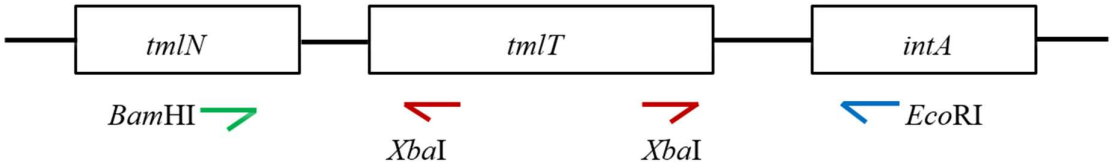
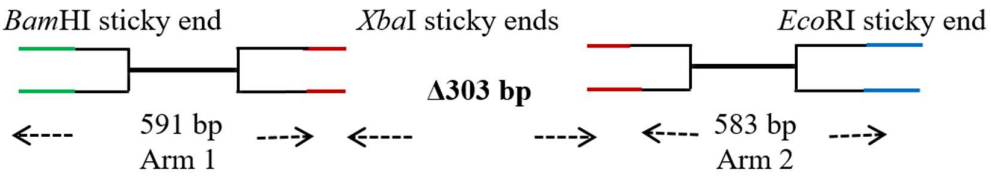


Figure 3.2 Construction for the deletion of *tmlS* gene from pTML1 plasmid of *Pseudoalteromonas sp.* SANK 73390.



pTML1 plasmid (WT *Pseudoalteromonas sp.* SANK 73390)



Amplification of upstream and downstream arms and their cloning into suicide vector pAKE604

Mobilize suicide vector pMY29 into WT *Pseudoalteromonas sp.* SANK 73390 to allow homologous recombination between suicide vector and pTML1 plasmid

Figure 3.3 Construction for the deletion of *tmlT* gene from pTML1 plasmid of *Pseudoalteromonas sp.* SANK 73390.

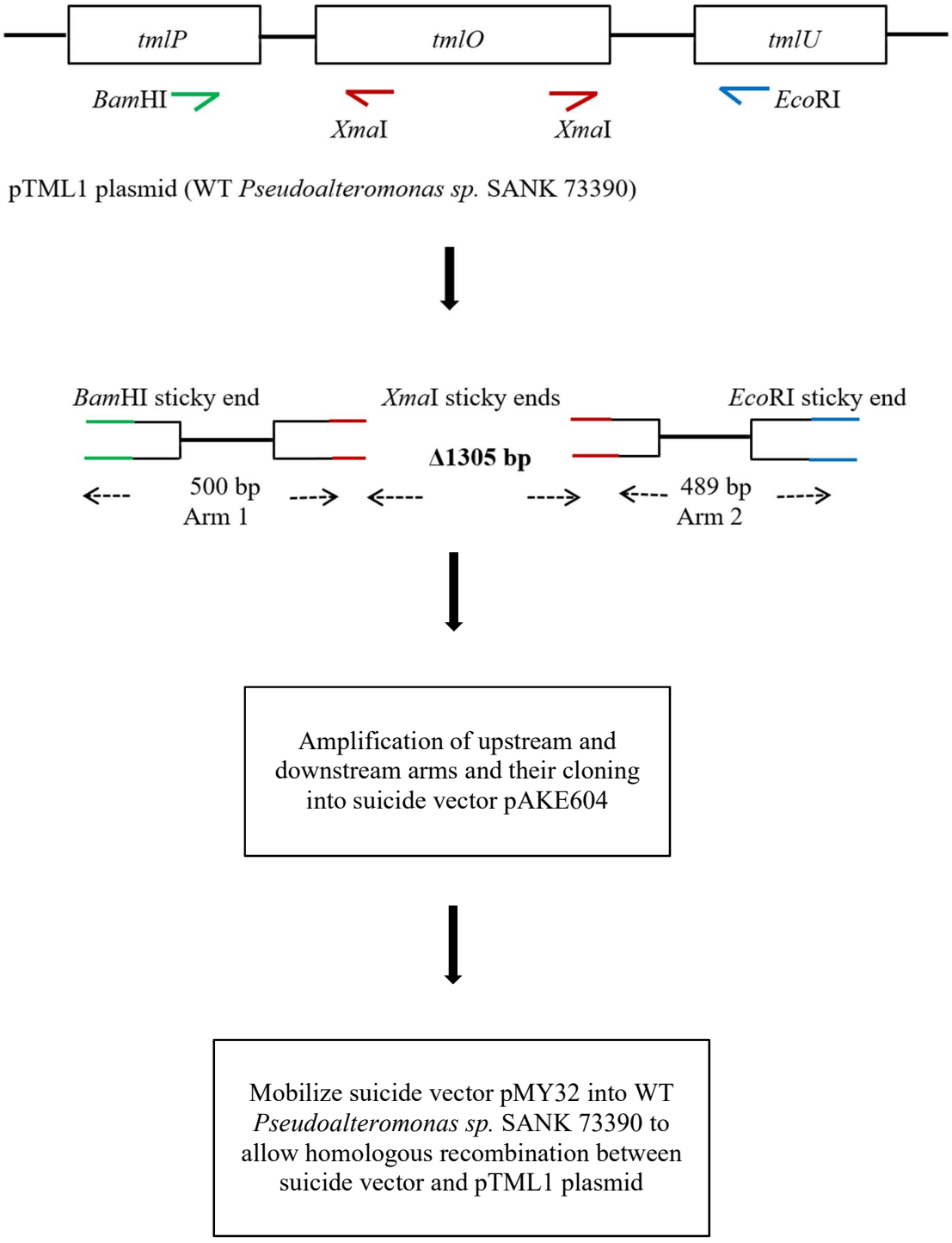


Figure 3.4 Construction for the deletion of *tmlO* gene from pTML1 plasmid of *Pseudoalteromonas sp.* SANK 73390.

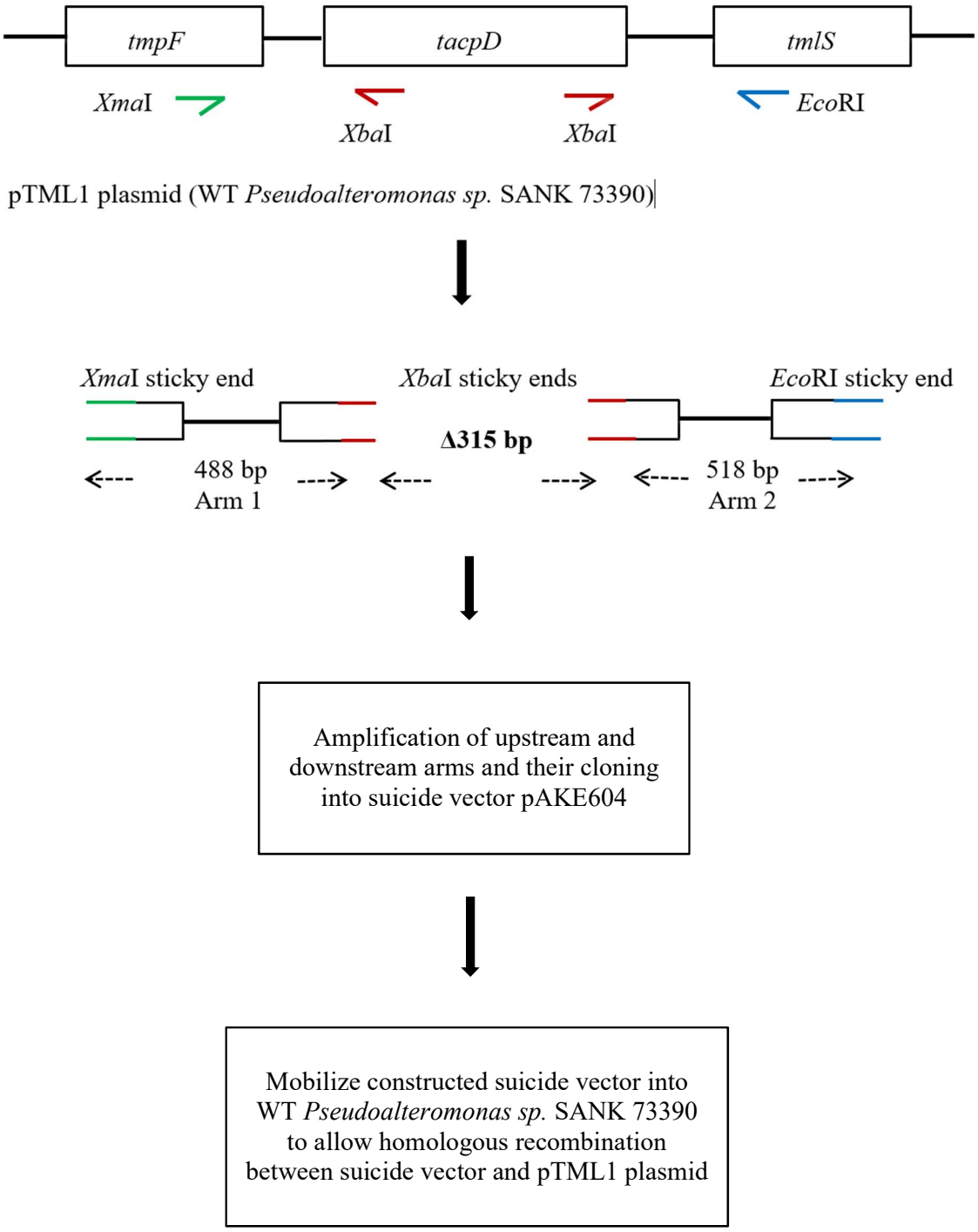


Figure 3.5 Construction for the deletion of *tacpD* gene from pTML1 plasmid of *Pseudoalteromonas sp.* SANK 73390.

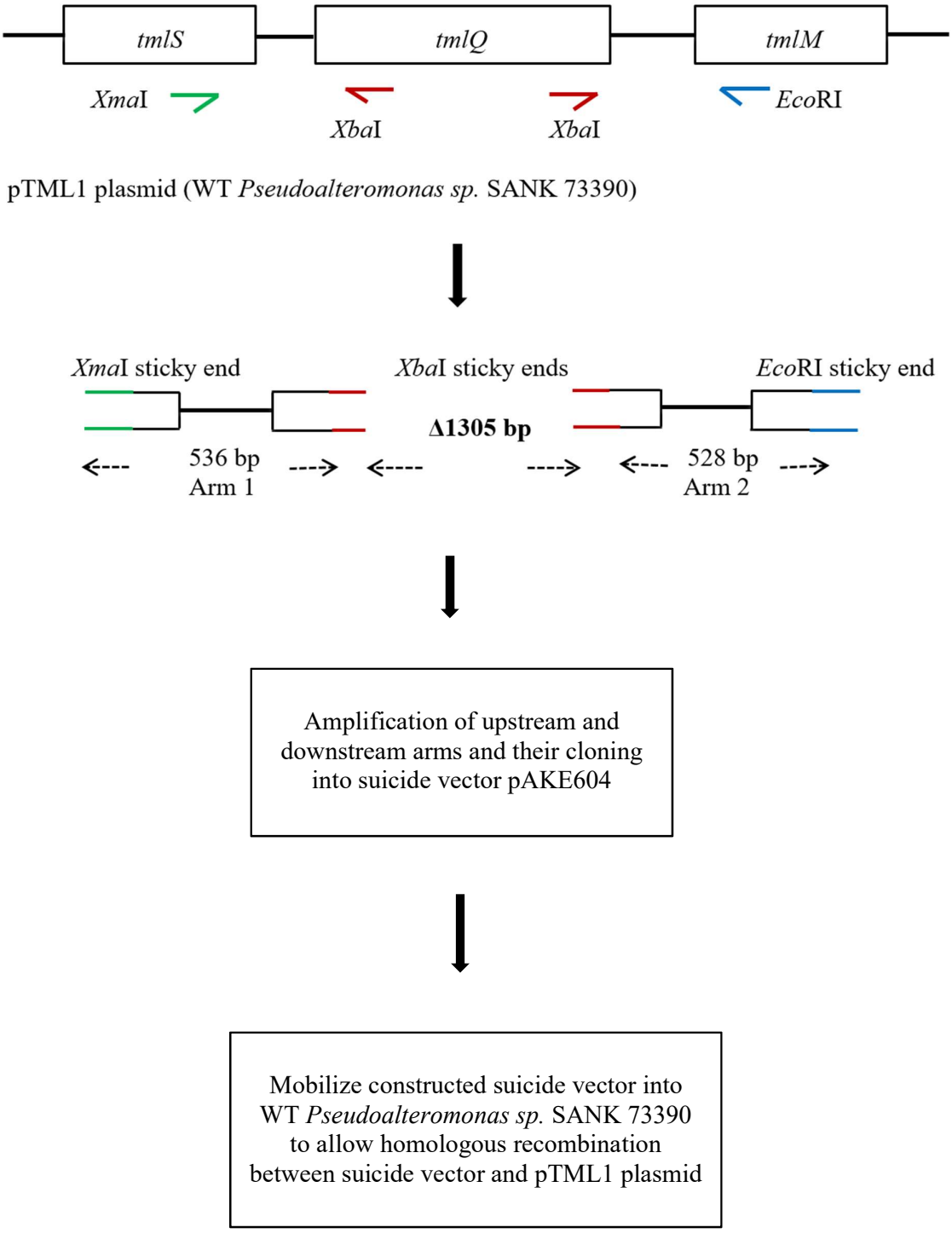


Figure 3.6 Construction for the deletion of *tmlQ* gene from pTML1 plasmid of *Pseudoalteromonas sp.* SANK 73390.

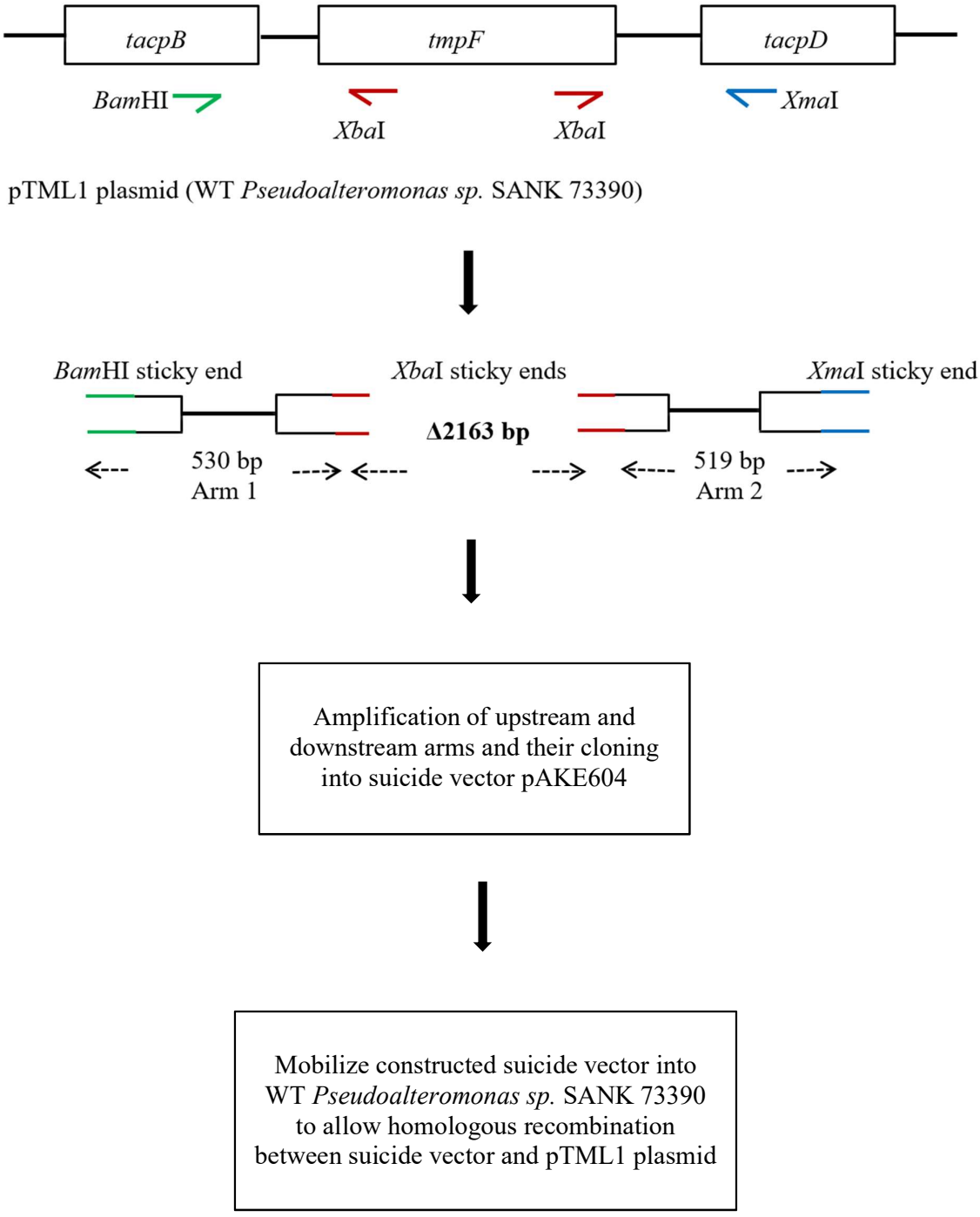
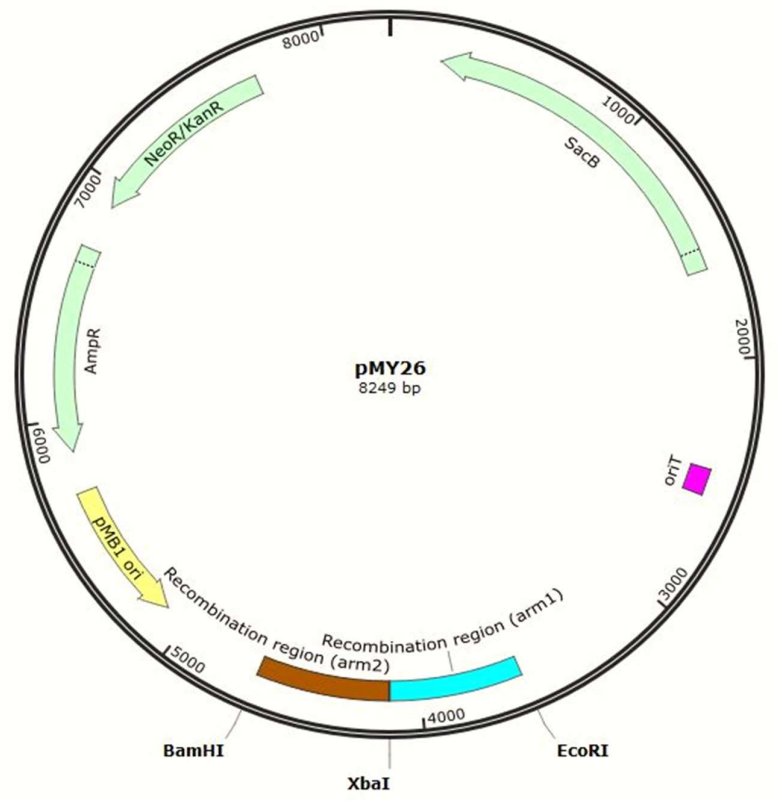


Figure 3.7 Construction for the deletion of *tmpF* gene from pTML1 plasmid of *Pseudoalteromonas sp.* SANK 73390.

(A)



(B)

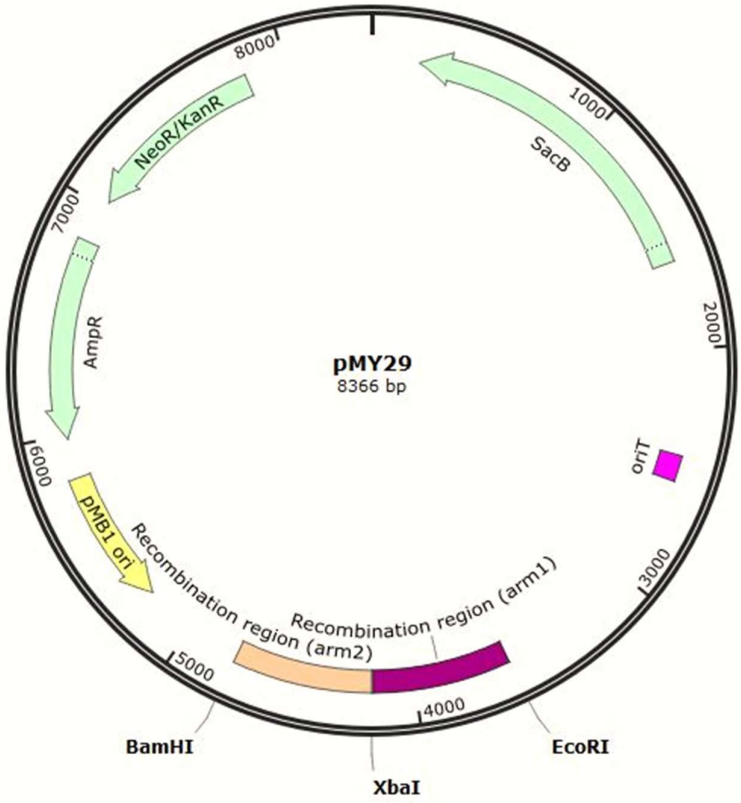
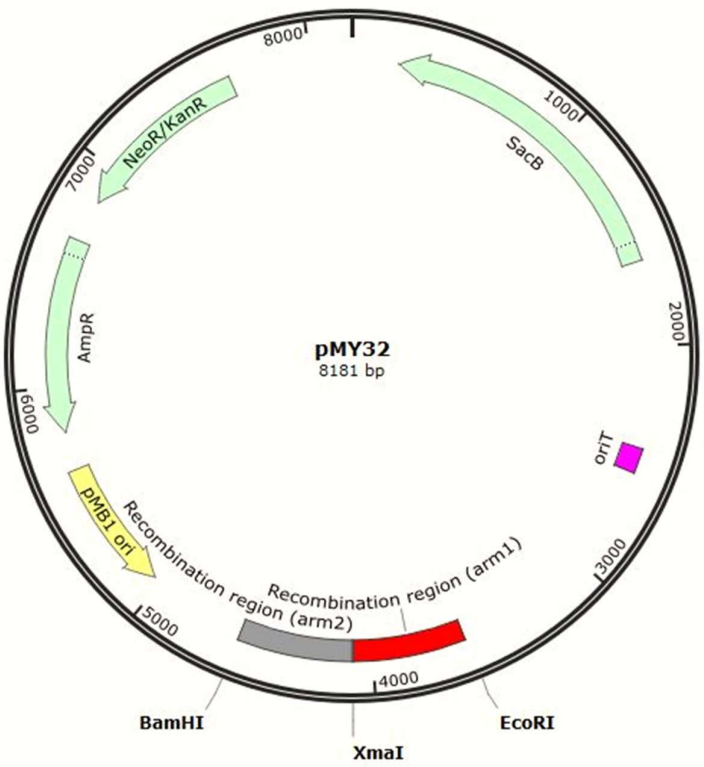


Figure 3.8 Map of suicide vectors pMY26 (A) and pMY29 (B) constructed for the deletion of *tmlW* and *tmlT* genes, respectivelyCreated using SnapGene®.

(A)



(B)

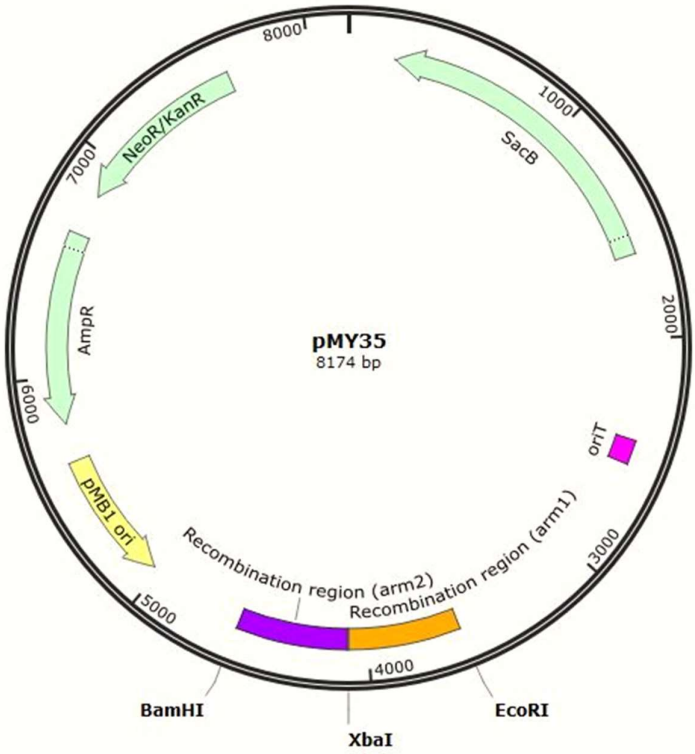


Figure 3.9 Map of suicide vectors pMY32 (A) and pMY35 (B) constructed for the deletion of *tmlO* and *tmlS* genes, respectivelyCreated using SnapGene®.

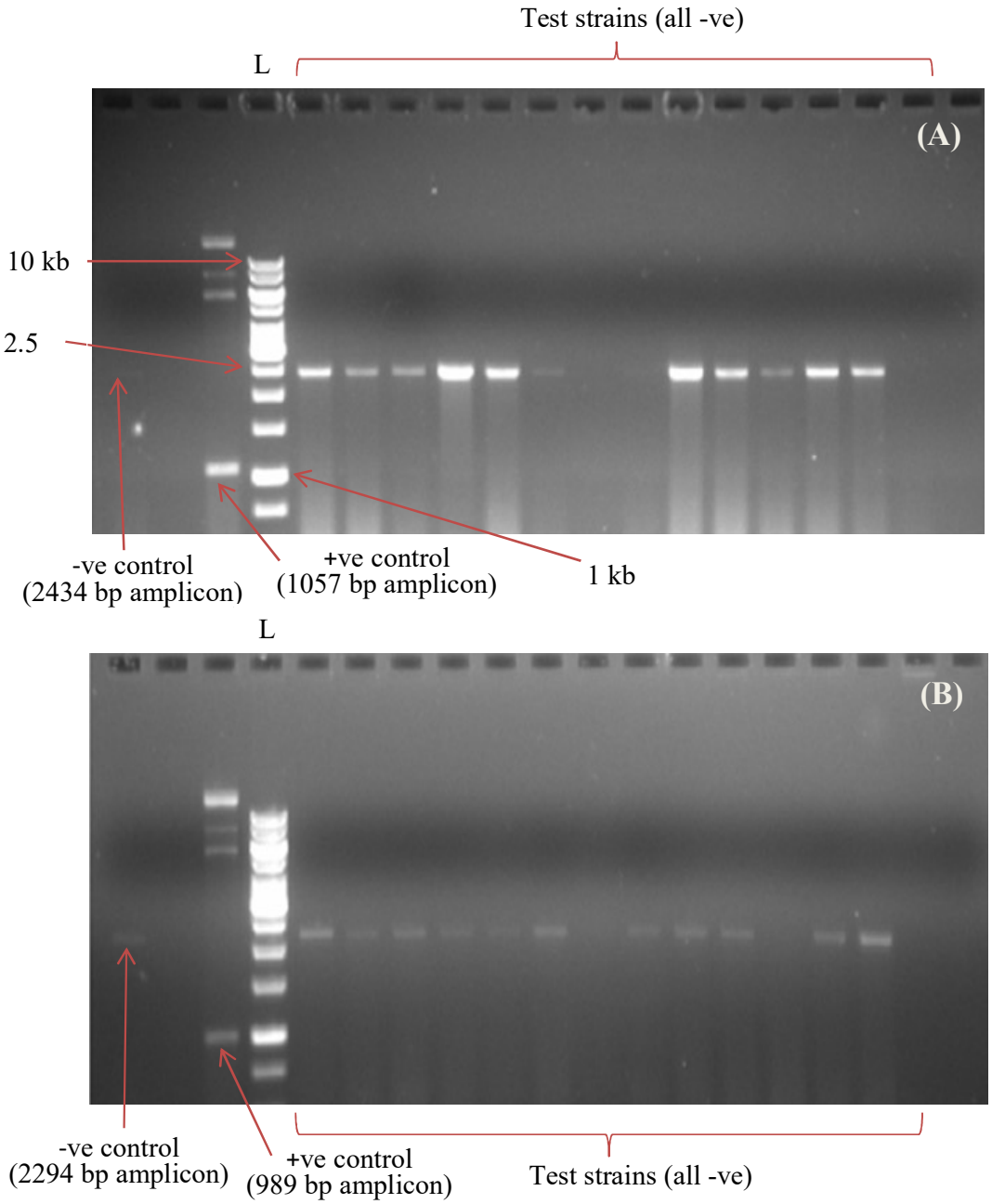


Figure 3.10 Screening for mutants of *tmlW* and *tmlO* using PCR. Excisants- the cells that had lost suicide vector from pTML1 plasmid which were identified by their resistance to sucrose and sensitivity to kanamycin were screened by PCR using (A) vector pMY26 as template for the deletion of *tmlW* (positive control) and (B) vector pMY32 as template for the deletion of *tmlO* (positive control). For negative control, wild type *Pseudoalteromonas sp.* SANK 73390 was used as template. In both the cases none of the colonies screened turned out to be a mutant. L= 1 kb ladder.

3.3 Discussion

Failure to get expected result (mutants) using strategy given in Chapter 2 (section 2.4.15) (Rahman et al., 2005), might be because of similar unknown reasons for which several other researchers of Thomas laboratory who used this strategy failed to get knockouts of many other genes of *tml* cluster. Defined in-frame knockouts of *tmlC, holE, holA, tmuA, tmuB,* module 2 of *tmpB* genes were attempted by several researchers of this laboratory using this strategy but expected mutants of none of these could be obtained. On the other hand, this strategy worked well for creating some other mutants like disruptive point mutations in *holA, tmuB, tmpD* and for creating defined in-frame knockouts of *tmlU* (Fukuda et al., 2011), and *tmlF* (Thomas, 2015), which was supportive of the fact that all the processes on which this strategy relied upon, were effective in these cases. Though, the *tmlF* mutants could be obtained but only after several attempts while repeated attempts made about other genes by various researchers of Thomas laboratory did not yield any mutants. *Pseudoalteromonas* species despite having multidrug resistance genes and drugs efflux pumps that may render it to survive antibiotic pressures and develop resistance, was sensitive to 50 µl/ ml concentration of kanamycin which was used to select its transconjugants (Webber and Piddock, 2003, Qin et al., 2011, Yu et al., 2013). Use of 5.5% w/v concentration of sucrose also worked to induce expression of *Bacillus subtilis sacB* gene to promote cell death thus permitted counter selection. Growth conditions that were used to perform conjugation also worked with *E. coli* S-17 strain grown O/N at 37 ºC and wild type *Pseudoalteromonas sp.* SANK 73390 grown at 23 ºC for 1-2 days until there was good growth. Selection of transconjugants also worked well at 23 ºC whereby colonies, which were selected by plating on M agar containing 50 µl/ ml concentration of kanamycin, indicated that mating had been successful. Single colonies of transconjugants were purified by restreaking before setting up

excision step. None of the transconjugants that were counter-selected because of being sensitive to kanamycin and insensitive to the presence of sucrose in the growth medium, turned out to be expected knockouts but wild type as confirmed by colony PCR for *tmlW* and *tmlO*. In case of *tmlS* and *tmlT* knockouts, none of the transconjugants turned out to be insensitive to the presence of sucrose in the growth medium when these were selected after the step that promoted second homologous recombination event.

Factors that might be responsible for the failure to get knockouts could be on account of the fact that the bacteria that are found in sea habitat have some features significantly different from those found in other habitats. Marine bacteria grow under extreme conditions of high salinity, lower temperature and higher pressure. As a result, they have different adaptations to survive in such environment. They have higher solute concentrations and therefore, integrity of cell membrane becomes very important to maintain it (Wood, 2015). Cell membrane is home to various processes like electron transport chain, various membrane transport systems/channels etc. Correct physical state of cell membrane is required for proper functioning of gene regulation, for intracellular signal transduction and for protein-protein interactions that takes place within it. The ability to survive under adverse conditions is achieved by changes in various factors including protein contents and mainly the lipid content of the cell. Various protein factors/enzymes of marine bacteria require high salt environment to become functional. Membrane lipids are modified to maintain the membrane fluidity according to the extreme habitat (Allen et al., 1999, Allen and Bartlett, 2000). It is well known that with the increase in temperature, percentage of monounsaturated fatty acid (MUFA) increases in the membrane phospholipids of bacteria. In marine bacterium *P. profundum* SS9 it has been shown that MUFA is essential for growth at low temperature and/or high pressure (Allen et al., 1999).

To support various adaptations to the extreme marine habitat, bacteria are often found to have many special genetic features. It has been reported that recombination and repair gene *recD* found in deep sea bacteria *Photobacterium profundum* ss9 is critical for growth at high pressure and, therefore, is different from *recD* found in *E. coli* (Bidle and Bartlett, 1999). Similarly, the *fabF* gene found in this bacteria codes for a beta-ketoacyl-acyl carrier protein synthase II which can adapt to changes in pressure compared to the one coded by *fabF* gene found in *E. coli* (Allen and Bartlett, 2000). In *P. profundum* ss9, this enzyme is involved in the production of unsaturated fatty acid, cis-vaccenic acid, which is required for normal growth at higher hydrostatic pressure (Allen and Bartlett, 2000). *E. coli* did not regulate its fatty acid composition in an adaptive response to high pressure. Several genetic loci have been reported from *P. profundum* ss9 that are associated with the growth at lower temperature and/or at high temperature. Mutations in any of these loci affected adaptations of these bacteria to grow in their natural environment as they affected various regulatory and other related processes for example those involved in ribosome assembly and function, cell envelope or chromosome structure etc. (Lauro et al., 2008). As part of adaptations for survival in its natural environment marine bacteria also develop resistance mechanisms to counter various adverse conditions and substances, and produce various secondary metabolites in order to have competitive edge over others (de Carvalho and Fernandes, 2010). All these adaptations are genetically represented. There are gene clusters in marine bacteria responsible for the production of secondary metabolites and genes that confers it resistance to survive in the multidrug environment (de Carvalho and Fernandes, 2010, Cordero et al., 2012, Yu et al., 2013, Giordano et al., 2015). It has been shown in *Pseudoalteromonas flavipulchra* JG1 strain that its genome represented several advantages over others as it encoded for adaptations to marine environment (Yu et al., 2013). This included expression of proteins like PfaP which is not just antibacterial but also contributed towards physiology by

its involvement in the biosynthesis of small molecules, which may be part of several of antagonizing and biocontrol mechanisms of this bacterial strain (Yu et al., 2013). Such genetic variations in *Pseudoalteromonas sp.* SANK 73390 species that are due to its marine habitat but currently unknown, might have affected the fidelity of various cellular and biochemical processes on which gene knockout protocol depended (that included DNA-DNA and DNA-protein interactions) resulting in the failure to get any mutants. Culturing of marine bacteria under laboratory conditions on artificial media (M agar/ M broth) might have also contributed towards this because these organisms normally grow under stressful conditions prevailing in their natural habitat where they may depend on the associated ecology for their various requirements like nutrition etc. (de Carvalho and Fernandes, 2010, Giordano et al., 2015).

In *Pseudoalteromonas sp.* SANK 73390, there could also be a different and varied set of restriction-modification systems that this species harboured that might have caused degradation of foreign DNA. This could have been the reason behind failure of recombination events to take place in the first place. It may also be the case of failure in the maintenance of acquired sequence even if it got integrated by recombination or by any other genetic method because of not being supported by host's replicative machinery for various adaptive reasons discussed above. Various protein factors/enzymes involved in the replication or regulation of replication of plasmid pTML1 (into which integration and excision events are supposed to occur) either did not function properly for the changed homeostatic conditions inside the cell or were expressed in sub-optimal concentrations because of bacteria being grown out of its natural marine habitat but under laboratory conditions and on artificial medium. Improper folding of concerned enzymes/proteins could also be one of the reasons behind failure of recombination events to take place. The fact that this strategy was successful for creating some mutants, though only after repetitive attempts,

was supportive of the fact that these restriction-modification systems and/or the host's replicative machinery / processes were not very rigid for foreign DNA which it received from distantly related species *E. coli* S17-1 strain. These systems/processes may also be dependent on the sequence (GC content or nature of sequence) that was introduced in the host strain as suicide vector. Ability to get some knockouts by others using this method also indicated that the various cellular processes of the bacteria were not drastically affected by its growth in laboratory on artificial medium. This may explain why same strategy failed in case of creating knockouts of other genes of same bacteria (this study and by others). *Pseudoalteromonas* species are thus not amenable to ordinary methods of genetic manipulations. New protocols are being developed to knockout genes effectively in these species (Yu et al., 2014, Wang et al., 2015).

One of the other reasons for failure to get expected mutants might be that knocking out of any of these particular gene(s) resulted in the loss of cell viability. This could have been the case because the product of that specific gene was critical for the survival of bacterium or knocking it out induced polar effects that rendered inactivation/lowered (than required) optimal expression of a gene, product of which was critical for survival of bacteria.

CHAPTER 4

Investigations on functional cross-complementation between products of homologous genes of *mup* and *tml* clusters

4.1 Introduction

As described in Chapter 1, in recent years there has been a growing problem of rising resistance to antibiotics among pathogens. As a result, there have been attempts to find newer antibiotic molecules with higher potency and broad spectrum. Newer molecules having novel activities have been reported from organisms from diverse environments including extreme habitats ranging from deep seas to deserts (Bredholt et al., 2008, Cragg and Newman, 2013, Orlova et al., 2015, Tiwari et al., 2015, Kozuma et al., 2016, Newman and Cragg, 2016). On the other hand, there have also been attempts to develop new antibiotic molecules from existing antibiotic biosynthetic clusters by using the approaches of combinatorial biosynthesis and applying the techniques of genetic engineering (Walsh and Fischbach, 2010, Wong and Khosla, 2012, Sun et al., 2015). Fully rational application of combinatorial biosynthesis to develop newer molecules with enhanced properties depends on the complete understanding of the antibiotic biosynthetic cluster, its gene functions, complete biosynthetic pathway including all the biosynthetic steps and intermediates involved their timings, the gene functions involved and protein-protein interactions taking place in those steps. It has been shown that tailoring functions apart from the core polyketide functions play an important role in deciding the final bioactive molecule (Rix et al., 2002, Olano et al, 2010). In recent years, manipulation of tailoring region genes has been a great source of novel or improved bioactivities of molecules as well as a way for making them less/non-toxic with enhanced stability (Rix et al., 2002, Olano et al., 2010). In efforts to get novel molecules or to get improved activities/properties of existing antibiotics, focus has also been on the group of antibiotics belonging to the *trans*-AT class (Till and Race, 2014, Helfrich and Piel, 2016). This work was an attempt to understand more about biosynthetic steps in such *trans*-AT

antibiotics and was focussed on finding common biosynthetic steps/intermediates in the two biosynthetic systems by studying protein- protein interactions.

Fukuda et al. 2011 has shown that *trans*-AT antibiotics; mupirocin and thiomarinol, have very similar structures which is explained in Chapter 1 (section 1.18). It raised the hypothesis that this structural similarity between the two antibiotics might be reflected at the genetic level. The biosynthetic systems of these antibiotics must have things in common (Figures 1.30 and 1.34) including individual biosynthetic steps, enzymes and/or biosynthetic intermediates. It has been pointed out that out of the total 45 ORFs of the thiomarinol cluster, 27 ORFs encode products that one can expect for a mupirocin like biosynthetic cluster producing marinolic acid (Figure 1.32) (Fukuda et al., 2011). Marinolic acid is the name derived by analogy with pseudomonic acid on account of the similarities between the structures of thiomarinol without pyrrothine (thiol group) and pseudomonic acid (Figure 1.30) (Fukuda et al., 2011). Although in thiomarinols, 9-HN is replaced with 8-HO, the only major difference from pseudomonic acids is the presence of pyrrothine moiety in thiomarinols that is attached to fatty acid via an amide bond. In both the molecules, mupirocin and thiomarinol, fatty acid moiety is in ester linkage with the monic acid component.

As stated above, the genetic setup for the type I PKS of the thiomarinol cluster is similar to *mup* cluster. Though there are differences in the number of ACP's, tailoring regions and regulatory aspects and there is an extensive rearrangement of tailoring/auxiliary genes compared to *mup* cluster which is detailed in Chapter 1 (Figure 1.34) (El-Sayed et al., 2003, Fukuda et al., 2011). By gene knockout, complementation and refeeding studies, much knowledge has been gathered about the mupirocin biosynthetic pathway of *Pseudomonas fluorescens* NCIMB 10586. Still a lot remains to be discovered about exact timings of

various biosynthetic steps in the biosynthetic process and about the role of some gene functions and interaction of various gene products resulting in specific modifications in the final molecule (Thomas et al., 2010, Gao et al., 2014). Relatively very little is known by experimental evidence (rather most is predicted) about gene functions of the thiomarinol cluster and biosynthetic steps involved in its biosynthesis (Fukuda et al., 2011). This situation warrants studying the homologous gene functions, products of which share a significant amino acid sequence identity, between the two biosynthetic clusters.

Gene knockout and complementation studies have been routinely used to study functional cross-complementation between homologues of different genetic systems. *In vivo* complementation studies have helped to elucidate functional organization of genes in the genome and to confirm whether specific phenotype relates to specific gene function. By *in vivo* cross-complementation studies using various regulatory genes of the nystatin biosynthetic gene cluster of *Streptomyces noursei* ATCC11455, hierarchy of differential control exerted by them in nystatin biosynthesis could be established (Sekurova et al., 2004). By cross-complementation studies in chlorothricin and tetrocarcin gene clusters it has been shown that the homologues of AbyA3 and AbyA4 that encode an ACP and dehydrogenase, respectively, are interchangeable (Gottardi et al., 2011). Interspecies cross-complementation in *Sacchropolyspora erythraea* has been used to elucidate the functions of *oleP1, oleG1* and *oleG2* genes of oleandomycin biosynthetic cluster of *Streptomyces antibioticus* which also led to the isolation of novel erythromycin derivatives (Doumith et al., 1999). Zhang et al. (2008) used cross-complementation studies between SgcE and NcsE PKS biosynthetic clusters to show that there is common mechanism for initiation of enediyne biosynthesis in nine-membered enediynes (Zhang et al., 2008). By complementation, it has been demonstrated that regulatory elements *lanI*, *lndI and jadR1* from different biosynthetic clusters could restore landomycins E and A production in *S. globisporus* I2-1 and *S.*

cyanogenus lanI7 regulatory mutants, respectively (Rebets et al., 2008). Complementation studies have also been used to shed more light on enzyme promiscuity. A survey of enzyme promiscuity and ambiguity in *E. coli* was conducted using self-complementation by which it was shown that 20% of the 104 auxotrophs, that were tested, survived as a result of overexpression of at least one of its noncognate gene (Patrick et al., 2007, Khersonsky and Tawfik, 2010). Using *in vivo* complementation studies together with other techniques it was shown that PTPS-I proteins **(**COG0720 subfamily involved in queuosine biosynthesis) in prokaryotes were inherently promiscuous as they catalysed same reaction on different substrates as well as different reactions on the same substrate (Phillips et al., 2012). Using complementation studies it has been shown that *Chlamydia* species have evolved a variant of tetrahydrofolate biosynthesis pathway that involve promiscuous and adaptable enzymes (Adams et al., 2014).

Studies for functional cross-complementation across the two systems; mupirocin and thiomarinol, for such specific homologous gene functions will shed more light on the similarity and differences in the two biosynthetic systems/pathways particularly about protein-protein interactions. It will also tell us about the extent of conservation and divergence of homologous gene functions during evolution in the two distantly related systems involved in the biosynthesis of similar structures.

Gene knock out and *in trans* complementation studies in the thiomarinol system could provide useful information to identify and assign the putative function of various genes in the *tml* cluster and help to elucidate biosynthetic steps much in the same way, as was done for the *mup* biosynthetic pathway and for other biosynthetic systems (Hothersall et al., 2007, Thomas et al., 2010, Gao et al., 2014). The cross-complementation studies between the two systems; *mup* and *tml*, may also reveal additional information about both systems as these should help us to identify common biosynthetic intermediates, biosynthetic steps and related

protein interactions taking place etc. in the two clusters. Complementation of a *mup* knockout by its corresponding homologous Tml protein will indicate whether the concerned metabolic intermediate in the two biosynthetic pathways is similar or not. This will also reveal about similarity or differences in the sequence of biosynthetic steps up to that stage i.e. actions of preceding enzymes and structure of metabolic intermediates between two systems. Lack of complementation by specific gene/product will indicate differences in the biosynthetic pathways in these two systems and/or the stringent specificity of respective enzymes for their specific substrate- the metabolic intermediates in the biosynthetic pathway. Such studies will also help in identifying the differences in these two biosynthetic pathways concerning different biosynthetic intermediates/steps employed for otherwise structurally very similar molecules. The outcome of the study should also provide useful data to shed more light on some of the questions that are still unanswered about mupirocin biosynthetic pathway. Complementation was expected in view of the fact that both these systems produce polyketide molecules with similar structures which is also reflected as homology of proteins of two clusters that are proposed to utilize the same building blocks (the starter units as acetate and the extender units as malonate).

One of the important unanswered questions is how fatty acid components in both the antibiotics are synthesized and what gene functions are responsible for their biosynthesis? One such question is why, despite the two biosynthetic systems coding for structurally similar polyketide molecules there are some enzyme functions present in one biosynthetic pathway but are missing in the other? And how do these two biosynthetic pathways differ from each other which otherwise produce structurally similar molecules that differ in their fatty acids by just one carbon? In addition, in neither system the gene functions responsible for 6-hydroxylation are known yet. By such cross-complementation studies between the two systems it has been shown that TmuB protein, a putative dioxygenase, of *tml* cluster is

responsible for the hydroxylation of the fourth carbon of the polyketide moiety in thiomarinols (Mohammad, 2016). It has been demonstrated by *in trans* expression of *tmuB* in *Pseudomonas fluorescens* NCIMB10586 that it can hydroxylate the fourth carbon of the polyketide moiety of pseudomonic acids which is similar to the one that is found in thiomarinols indicating that there is some similarity in the biosynthetic steps in the two systems.

It was proposed to study all of the 22 tailoring/auxiliary genes of the *tml* cluster, products of which shared significant amino acid sequence identity with their *mup* equivalents. The genes were tested in the order based on information available about their putative functioning as groups whereby either they worked together or processed biosynthetic intermediates one after the other in the predicted *mup* biosynthetic pathway. In this regard their order of occurrence in the gene cluster was also taken into account. Gene knockout, complementation and refeeding studies in *mup* cluster by Hothersall et al. (2007) and Gao et al. (2014) have established the way in which products of various genes collaborated/acted, to process biosynthetic intermediates in the *mup* biosynthetic pathway. Based on this information, biochemical nature of gene products and significant homology shared by them in these two systems, complementation by *tmlS, tmlQ, tmpF,* and *tacpD* was studied because their corresponding *mup* equivalents have been predicted to be involved in furnishing the starter unit for the biosynthesis of the fatty acid component of pseudomonic acids.

Similarly, *tmlO, tmlC* and *tmlF* were studied, as all of their *mup* counterparts have been shown to be involved in determining oxidation state acting one after the other around the pyran ring common to both the molecules (Figure 1.28). *tmlJ* and *tmlK* gene functions were studied together because their *mup* counterparts have been implicated along with a group of several other genes, equivalents of which are also present in the *tml* cluster, in the

insertion of a methyl group (C-15) at β- position at C-3 and which is present in both the molecules.

Knowledge gained as a result of this study about these and other biosynthetic pathways may be used to reengineer them using approaches of combinatorial biosynthesis to generate novel molecules; derivatives of pseudomonic acids and thiomarinols, with improved bioactivities and desired properties such as lack of toxicity that could be commercially produced.

4.2 Result

Functional cross-complementation studies of nine tailoring region genes out of the initial plan to test 22 were completed. Nine genes of the tailoring region of thiomarinol cluster namely *tmlJ, tmlK,* tml*S, tmlQ, tacpD, tmpF, tmlO, tmlF* and *tmlC* were tested by *in trans* expression of their ORFs in the corresponding mutants of *Pseudomonas fluorescens* NCIMB 10586. Primers were designed to amplify each of these nine ORFs incorporating unique restriction sites at their either end, details of which are given in Table 2.10.

PCR was carried out as described in Chapter 2 to amplify each of the ORFs. Amplification of most of the ORFs could be completed in the first attempt by using Velocity™ DNA polymerase. For amplifying *tmlO* ORF, initially 49 ºC was used as the annealing temperature but that yielded no result. While product of expected size was obtained on repeating the PCR with the annealing temperature raised to 53 ºC**.** ORFs *tmlC, tmlF, tmlJ, tmlK, tmlQ* and *tacpD* could be amplified in the first attempt. ORFs of *tmlC, tmlF* and *tmlJ* could be amplified at an annealing temperature of 51ºC, while good yield of product of expected size for *tacpD* and *tmlQ* ORFs could be obtained at annealing temperature of 55 ºC. The *tmlK* ORF could be amplified at annealing temperature of 53 ºC. For amplification of *tmlS* ORF, initially 49 ºC was used as the annealing temperature but after optimising temperature to 53 ºC, a good yield of product of expected size was obtained. It took several attempts to get PCR product of *tmpF* ORF for which KOD DNA polymerase was also used for amplification (see section 2.3). Initially, 53 ºC was used as the annealing temperature for amplification but it is only after optimising annealing temperature to 55 ºC and extension period of 2 min 30 s, a good yield of product of expected size could finally be obtained using Velocity™ DNA polymerase.

PCR products were purified, A-tailed and cloned into pGEM-T Easy vector in each case in order to sequence them. For every ORF after confirming their sequence to be correct, these fragments were cloned into inducible pJH10 expression vector (induced by IPTG) by digestion/ligation using incorporated restriction sites. All the plasmids thus developed are listed in Table 2.6. Before proceeding to the next stage presence of PCR amplified ORFs was confirmed in the respective expression vectors by restriction/digestion (Figures 4.2, 4.7, 4.11, 4.15, 4.21, 4.23, 4.28 and 4.36).

Each of the expression plasmids so constructed was then introduced into *E. coli* S17-1 by transformation and the purified strain was used to transfer this plasmid into the corresponding *Pseudomonas fluorescens* NCIMB 10586 mutant strain by conjugative transfer as described in Chapter 2 (section 2.4.14). Complementation was assessed by bioassay which was performed as described in Chapter 2 (section 2.6). In each case, visual inspection of the plate bioassay was followed by quantitative analysis displayed as a bar-diagram.

4.2.1 Complementation by TmlO

It could be seen from figures 4.3A, 4.3B, 4.4 and 4.5, that TmlO did not complement functions of its *mup* equivalent when expressed in the corresponding mutant strain *Pseudomonas fluorescens* NCIMB10586Δ*mupO.* Which was despite TmlO shared significant amino acid sequence identity with MupO (Fukuda et al., 2011). Figure 4.1 gives global alignment of these two proteins. In TmlO complementation, it was observed in bioassay plates that the zone of clearance was further reduced when IPTG was included in growth medium indicating that enhanced expression of TmlO interfered with the biosynthesis of active moiety that was responsible for the development of clearing zone (Figure 4.3). HPLC analysis of the supernatant from constructed strain also showed chromatogram different from wild type strain of mupirocin producer in agreement with the results of bioassay (Figure 4.4).

Presence of new peaks in the chromatogram compared to that of wild type strain indicated existence of new moieties. Results of LC-MS also confirmed non-complementation of *mupO* by *tmlO* (Figure 4.5). HPLC and LC-MS chromatograms of *tmlO* complementation showed peaks confirming the production of PA-B which is produced by Δ*mupO* strain (Figures 4.4 and 4.5) (Cooper et al., 2005b).

```
TmlO          1 MTTNSQNQSMG----QPPIAKNITDNGQILTWLDNPLTYLRKCYDQYGNI    46
                ||:..:..|.|    |.|:.|.....|..|..:.|||.:|:.....||::
MupO          1 MTSWEREVSSGAGNRQLPVVKGWPLLGSALALIRNPLGFLQTTRSTYGDV    50

TmlO         47 FRIGSEKNSVTVLAGLEANQLMSTESKELLSSNAFWSDTLSEMDCPHSFI    96
                :|:.:...:..||||:|||:.::.:.|:...|:.||..||.||.||||||
MupO         51 YRVKAAHMNFVVLAGMEANRFVADKGKDCFVSSGFWEATLQEMQCPHSFI   100

TmlO         97 GVDGDVHRYQRRLMAPMFCKASLKECVGEFGKTFRSILN-HHCTQPKTLL   145
                |||||.||:||.||.|:|.|::..|.:....:.|...|. .:....|  :
MupO        101 GVDGDAHRFQRNLMKPLFSKSAFNERIPMLAQIFTDTLQARYGVDQK--V   148

TmlO        146 SPFLRHTLSTQIGYALQGYAPSPEEVEALIDYQATVMNACSFKRISKRVL   195
                |...||.||.|||.:||||.|:|:|||||:.||.|.||.|:.|:..:..|
MupO        149 SALFRHVLSQQIGGSLQGYQPTPDEVEALMRYQNTAMNVCALKKWPRLAL   198

TmlO        196 MSPEYLKTKQTVHDLADRIIHNEQSEHAKERYIDVLISKGVKNNPQWFTP   245
                ..|.|...|:.|..||||||.:|:::..:.|...|..||....||||||
MupO        199 RLPGYRAAKKQVQALADRIIESERTQEQTQGYFQTLKEKGQMVQPQWFTP   248

TmlO        246 GDIRNHAIIPFLAGIDTVTSALSFALRELLLRPDIRQELQQEV-AQLNLY   294
                ||:|||||.:|||||||.:.|||.|.||..:|.:.|.|:.|| |..:..
MupO        249 GDMRNHAIISYLAGIDTVGATLSFMLLELFKQPHLHQALRDEVDACFSQG   298

TmlO        295 DPDLSVINDMSKVNLFTKEIMRLYPPAFAIYRVATNDFHFDGYQINKGET   344
                .||...:.:|..:..|.:|:|||||.|:|:.|.|..||.|.||.|:||:.
MupO        299 LPDADGLENMETLKNFIREVMRLYPTAYAVRRTAAKDFEFQGYSIDKGQD   348

TmlO        345 LVFFLTASHTDERYFSDPYEFKLDRFADPDLD---SEAYAPFGKGPHTCI   391
                ::.|.||:|||..:|.:|..|.:.|:.:|.|:   |.|:||||:||||||
MupO        349 IILFTTANHTDPAWFKNPQVFDITRYEEPRLEHRASGAWAPFGRGPHTCI   398

TmlO        392 GAGLSNIILPLNLAIILRFANFEYEGDITNLAFDYTKPSLSIDPNFTLSF   441
                ||||:||:|.||||:.|...:.....::::..|::.|:..:...|.:||
MupO        399 GAGLANILLSLNLALFLYHTDLRPACKLSDIKMDFSNPAAGLSERFAISF   448

TmlO        442 TPRATDQIV    450
                |||...
MupO        449 TPRNRP---    454
```

Figure 4.1 Global alignment of TmlO and MupO proteins using EMBOSS Needle.(Rice et al., 2000). Default parameters for Needleman-Wunsch algorithm used (Needlema.Sb and Wunsch, 1970). Percent similarity between protein sequences 60.1% which is identified, depending on the level of similarity, as ':' (more similar) and as '.' less similar, while identity is indicated as '|'.

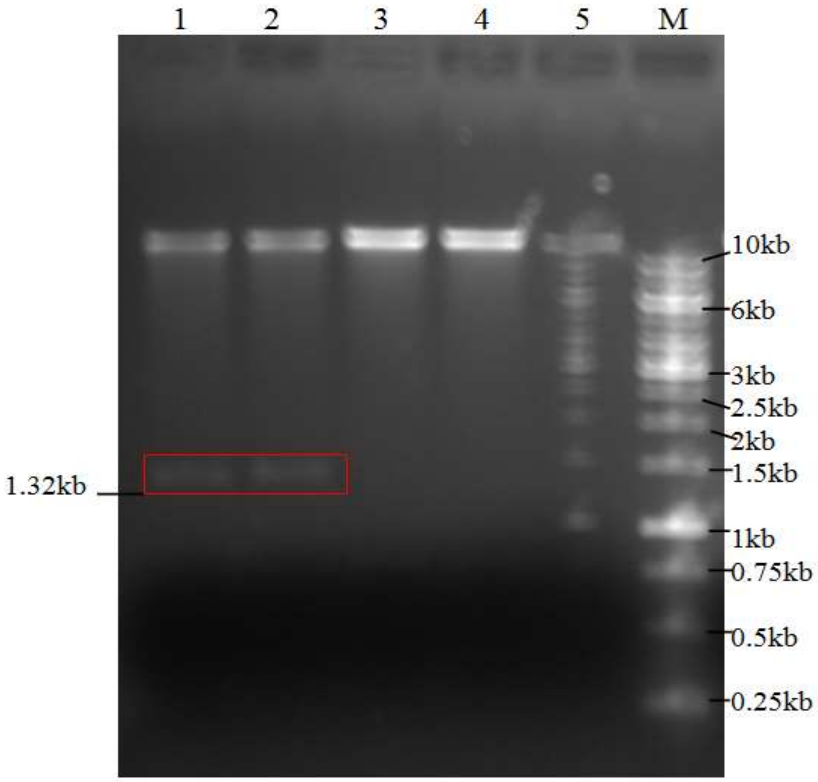
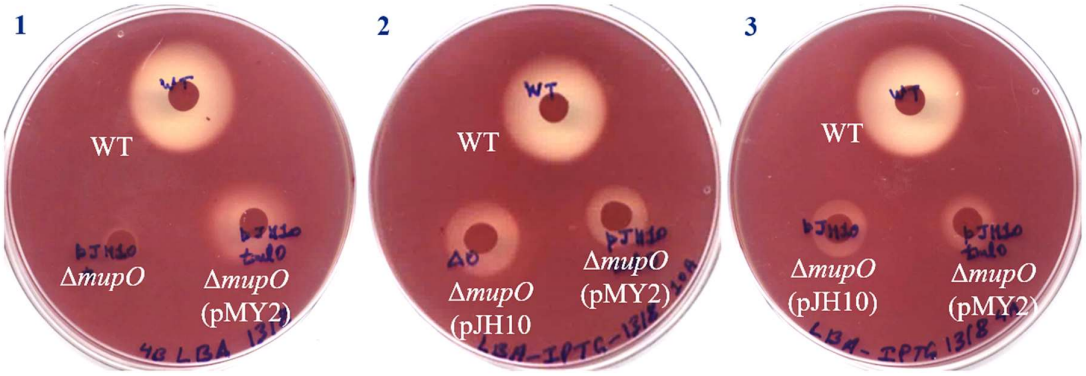


Figure 4.2 Confirmation of the *tmlO* insert in pMY2 expression vector. Double digest with *Kpn*I and *Xba*I yielded a fragment size of 1.35 kb that confirmed the presence of cloned *tmlO* gene in these clones (lane 1 and 2). Other fragment of size larger than 10 kb in these clones corresponds to linearized pJH10 in which it was cloned. M, 1 kb ladder.

(A)



(B)

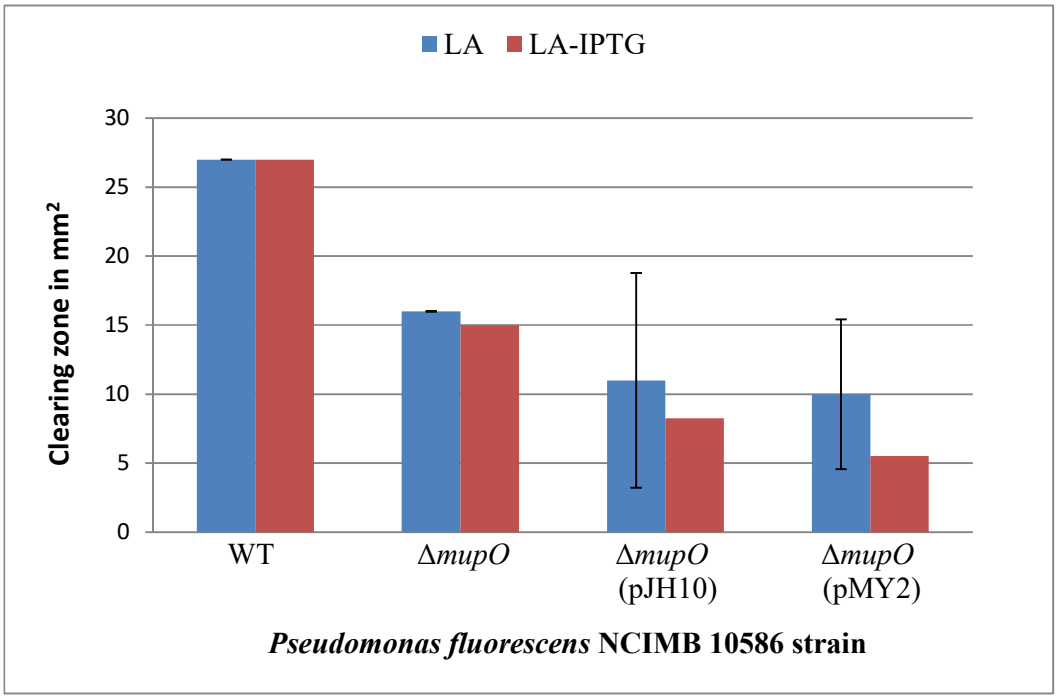


Figure 4.3 Bioassay to determine complementation by *tmlO* of the *mupO* single mutant of *P. fluorescens* NCIMB10586. (A) Bioassay plates showing clearing zones with (2 and 3) and without (1) induction by 0.5 mM of IPTG (B) Chart of quantitative bioassay (clearing zone measurements). WT, Δ*mupO* and Δ*mupO*(pJH10) strains of *P. fluorescens* NCIMB 10586 used as controls while Δ*mupO*(pMY2) was the test strain. Three replicates were used to collect data. Each bar represents the mean ± standard error of the mean.

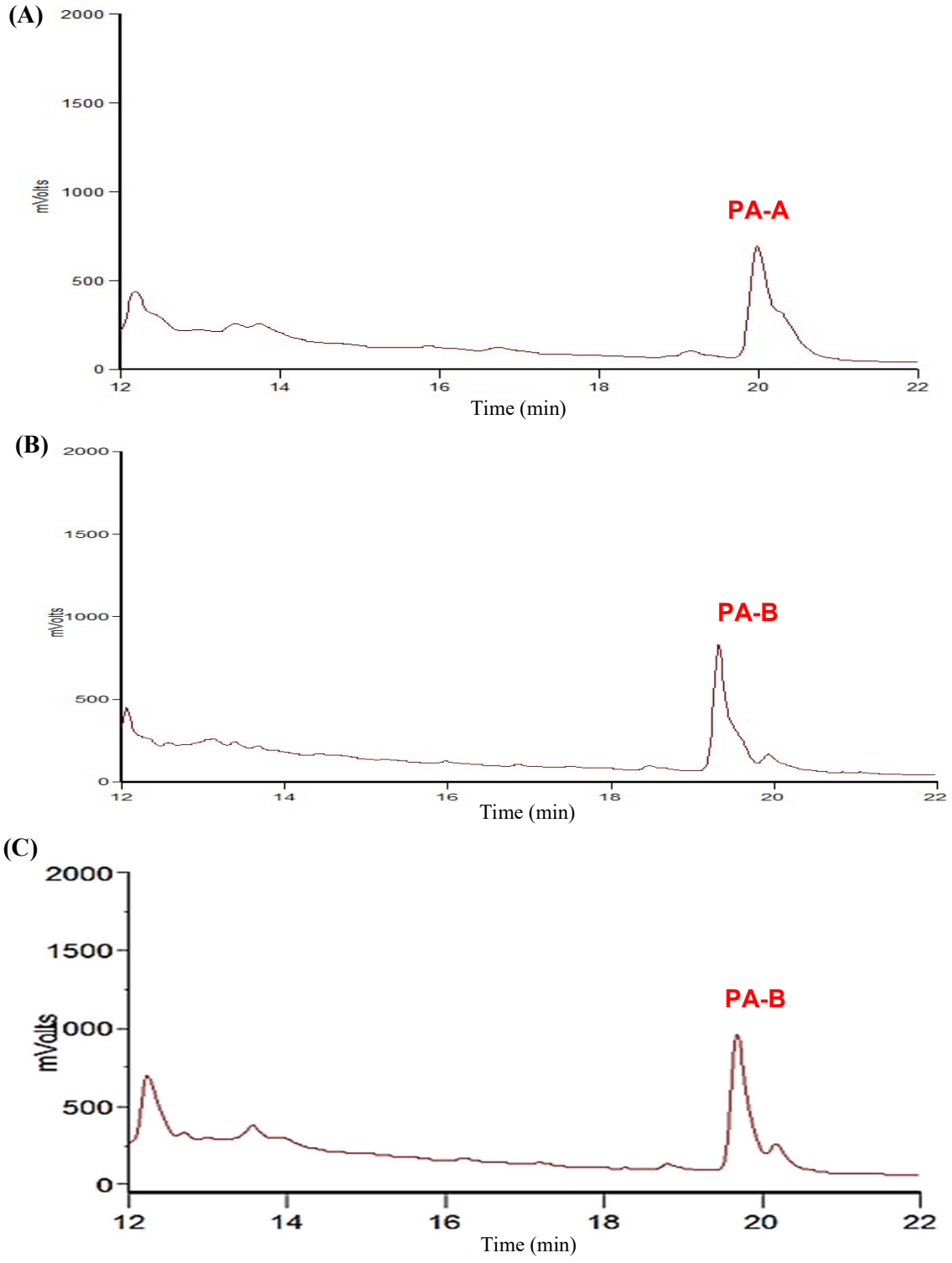


Figure 4.4 HPLC analysis of extracts from strains of *P. fluorescens* NCIMB 10586 for the complementation by *tmlO* of Δ*mupO* single mutant. Test strain Δ*mupO*(pMY2) did not produce PA-A both in the presence (C) or absence (B) of 0.5 mM IPTG and instead produced PA-B only (retention time 18.4 min). Positive control WT strain produced peak characteristic of PA-A (retention time 20.4 min) (A).

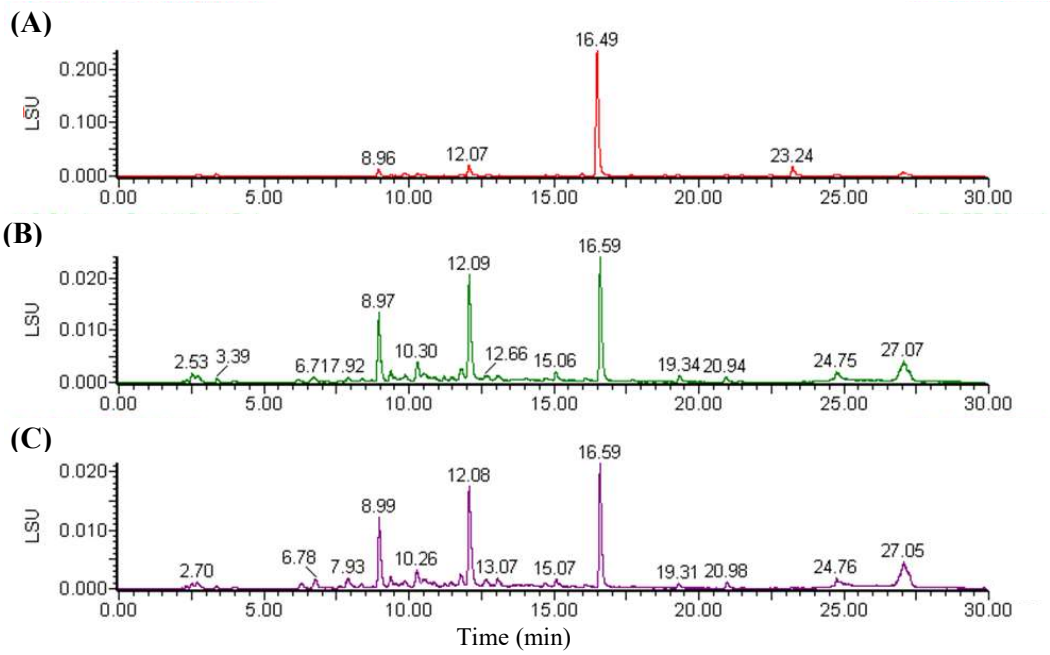


Figure 4.5 LC-MS chromatograms of extracts from strains of *P. fluorescens* NCIMB 10586 for the complementation by *tmlO* of Δ*mupO* single mutant. Neither of the three strains (A), (B) or (C) produced PA-A (retention time 20.4 min) despite induction by 0.5 mM IPTG but only PA-B (retention time 16.59 min) (done by Dr Song of University of Bristol, UK). (A) and (B), strains 10586Δ*mupO*(pMY2); (C), 10586Δ*mupO*(pJH10); LSU, light scattering units.

4.2.2 Complementation by TmlC

It can be observed from the results of plate bioassay (Figures 4.8A and 4.8B) that TmlC did not complement functions of its *mup* equivalent, MupC, when *tmlC* ORF was expressed *in trans* in the corresponding mutant strain *Pseudomonas fluorescens* NCIMB 10586Δ*mupC*. This was despite TmlC shared significant amino acid sequence identity with MupC protein (Fukuda et al., 2011). A global alignment of these two proteins is shown in Figure 4.6. In TmlC complementation study the presence of *tmlC* ORF, *in trans*, decreased the clearing zones in bioassay plates almost to the same level when no insert was present in the expression vector used, which was suggestive of an active effect (Figure 4.8). As a control, it could be observed that *mupC* showed complementation in single mutant NCIMB 10586Δ*mupC* when

expressed *in trans,* in the absence of IPTG in the growth medium which was in agreement with previous findings (Hothersall et al., 2007). LC-MS chromatograms of *tmlC* complementation showed peaks for the production of mupirocin C instead of that for PA-A (Figure 4.9) (Hothersall et al., 2007).

Figure 4.6 Global alignment of TmlC and MupC proteins using EMBOSS Needle(Rice et al., 2000). Default parameters for Needleman-Wunsch algorithm used (Needlema.Sb and Wunsch, 1970). Percent similarity between protein sequences 61.4%, which is identified, depending on the level of similarity as ':' (more similar) and as '.' less similar, while identity is indicated as '|'.

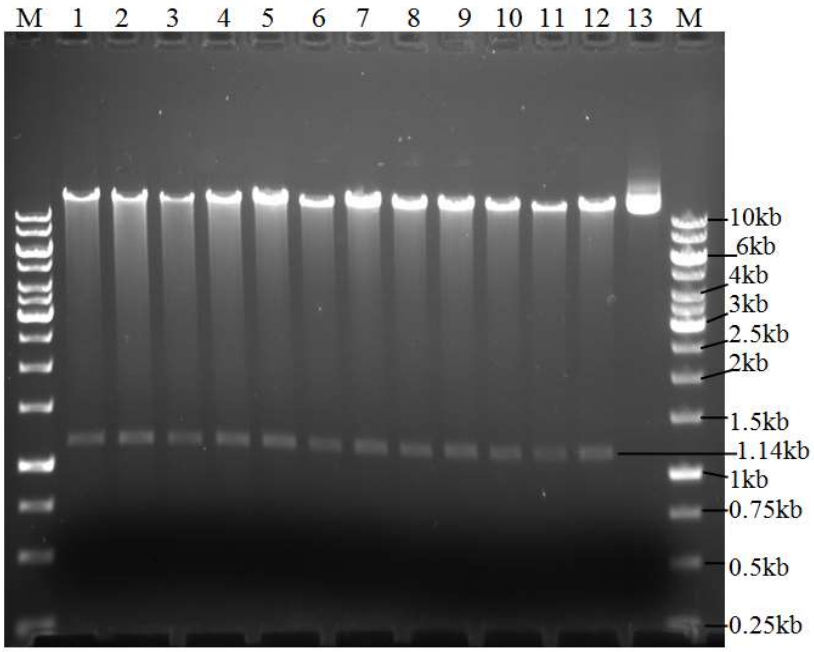
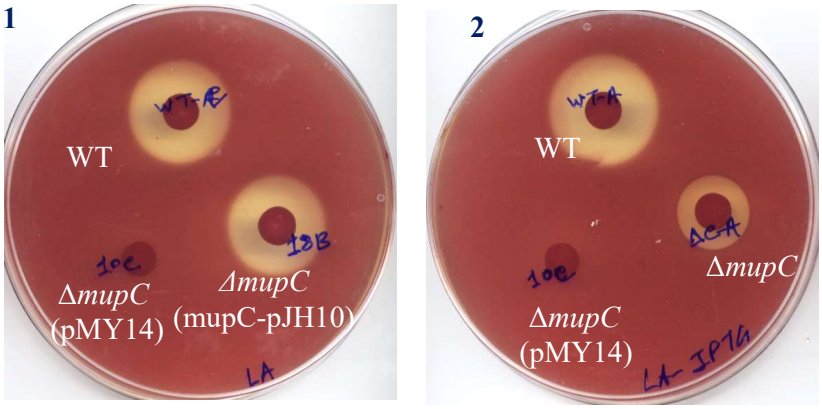


Figure 4.7 Confirmation of the *tmlC* insert in pMY14 expression vector. Double digest with *Kpn*I-*Sac*I yielded a fragment size of 1.14 kb that confirmed the presence of cloned *tmlC* gene all the 12 clones analysed (lane 1 to 12). Other fragment of size larger than 10 kb in these lanes corresponds to linearized pJH10 in which it was cloned. M, 1 kb ladder.

(A)



(B)

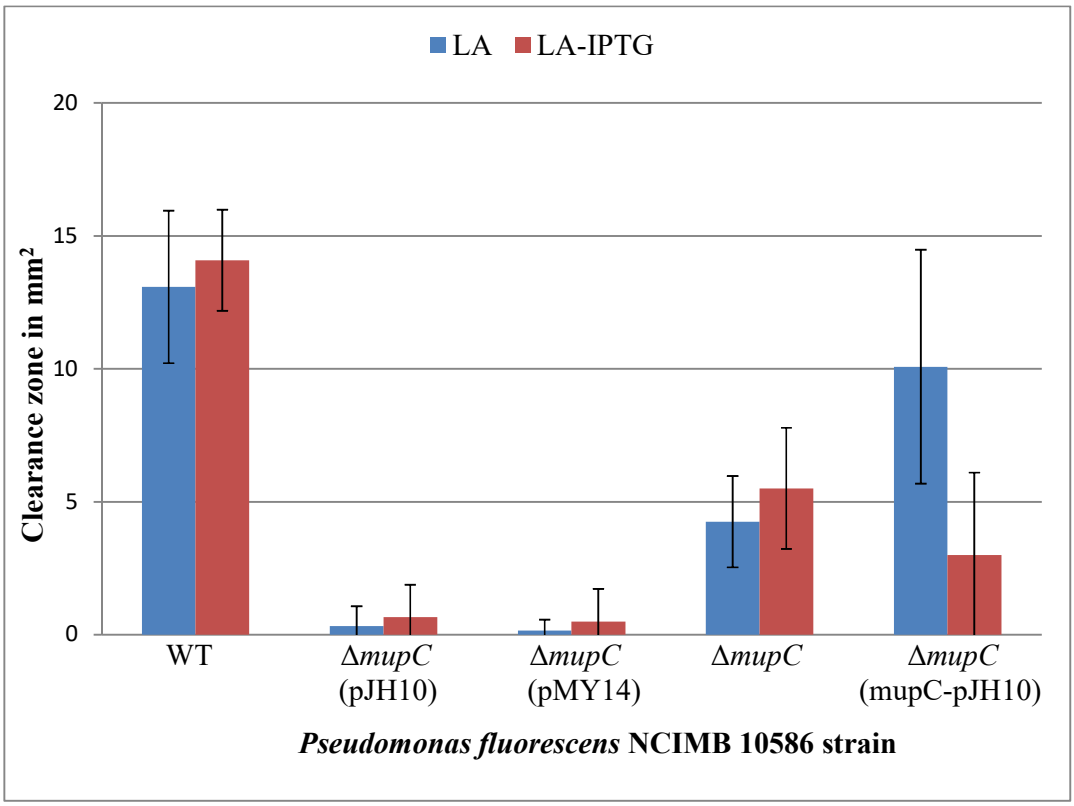


Figure 4.8 Bioassay to determine complementation by *tmlC* of the *mupC* single mutant of *P. fluorescens* NCIMB10586. (A) Bioassay plates showing clearing zones with (2) and without (1) induction by 0.5 mM of IPTG (B) Chart of quantitative bioassay (clearing zone measurements). WT, Δ*mupC,* Δ*mupC*(pJH10) and Δ*mupC*(mupC-pJH10) strains of *P. fluorescens* NCIMB 10586 used as controls while Δ*mupC*(pMY14) was the test strain. Three replicates were used to collect data. Each bar represents the mean ± standard error of the mean.

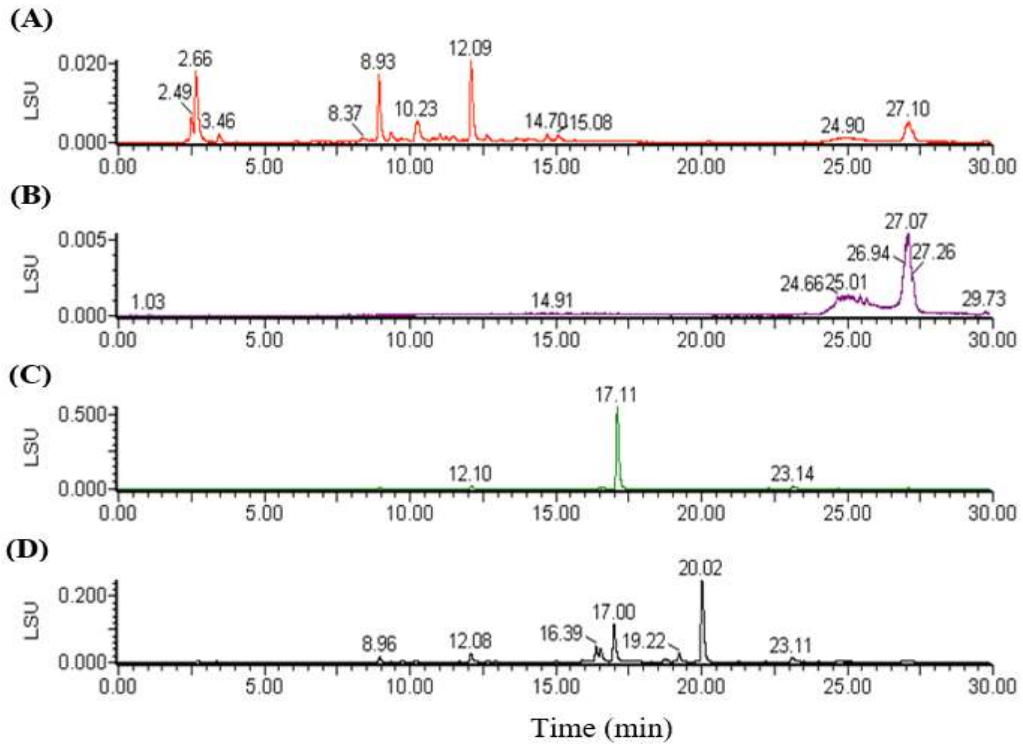


Figure 4.9 LC-MS chromatograms of extracts from strains of *P. fluorescens* NCIMB 10586 for the complementation by *tmlC* of Δ*mupC* single mutant. Test strains (A) and (B) did not produce and restore production of PA-A (retention time 20.4 min), (C) and (D) were the two control strains (done by Dr Song of University of Bristol, UK). (A) and (B), strain 10586Δ*mupC*(pMY14); (C), 10586Δ*mupC*(pJH10); (D), 10586Δ*mupC*(mupC-pJH10); LSU, light scattering units.

4.2.3 Complementation by TmlF

From figures 4.12A and 4.12B that show results of plate bioassay, it can be observed that TmlF, similarly like TmlC, also did not complement function of its *mup* equivalent, MupF, when *tmlF* ORF was expressed *in trans,* in the corresponding single mutant strain *Pseudomonas fluorescens* NCIMB 10586Δ*mupF*. This was despite TmlF shared significant amino acid sequence identity with MupF protein (Fukuda et al., 2011). Figure 4.10 gives global alignment of these two proteins. LC-MS analysis of the supernatant from both the test strains confirmed non-complementation by TmlF as the chromatogram did not show peak characteristic of PA-A production, instead it showed peaks for the production of mupirocin F

as was shown by control strain NCIMB 10586Δ*mupF* strain (Figure 4.13) (Hothersall et al., 2007).

```
TmlF              1 MRETVCVTGASGYLGTHVVAELLAQGYKVIASVRDINSEGVHAMTKAYEE     50
                    |...:|:|||||:||:|:|.:||..|..|.|.:|   ....|....|...
MupF              1 MPRHICITGASGFLGSHIVEKLLHHGCTVDALLR---RPTAHLQGLAARH     47

TmlF             51 GMFKVMAADFTVAHSLDEALAQSDFLIYCAATTALGFTNDVNVPQAHQSA    100
                    ...::...|.....:.:...::.|.:|:|||:..........:.|.....
MupF             48 AQLRLHVVDLAQPGACNALFSECDTVIHCAASVTNNARPGSRLWQQTLEM     97

TmlF            101 NIIGTGNMIQSIKRCKNLKKVVYSSSMASVFRSDVADDHLYDESDWNNDD    150
                    |:.||.|:|..|::..:::..||:||||::...|:...|.:.|:|||:..
MupF             98 NVNGTQNVIDGIRQAASVRTFVYTSSMAAILSPDMPAGHAFSEADWNHAS    147

TmlF            151 KANSEPYFYSKTQAELMLSALYDQDDPNTLPALVRFNPSVLMGPVKVASH    200
                    ...|:||:||||.||.:::..: |..| ..|.:|..||||::|||....|
MupF            148 LEASDPYWYSKTAAERLVTDAF-QPWP-CEPRVVCINPSVIIGPVLDPRH    195

TmlF            201 QNSSISILRNLLNARSGGCPRLYYSIADVRDVARIYVEALHNPSAKGRYL    250
                    ..:||:||.:|...::..||.|.:...||||:|.::|.|..:...:||::
MupF            196 AATSIAILHDLHTGKTPACPDLNFHFVDVRDLAELHVRAALDDVIEGRFI    245

TmlF            251 LPGKSISLLAIAQIVAQHFPEYKMPTRELKDTFVYALASHNTGLTTAYLD    300
                    :||...|:|.:|.::.|.||:.|.|.|...|..:|..|..|..|:..||.
MupF            246 VPGHEASMLELATMIKQRFPDSKAPQRRAPDWLMYLSAMLNPRLSRRYLR    295

TmlF            301 KFLGIEHRFNDSKLRRDFIEPYCIDLVQTIKDSVA-SIQAIEQQAVALNV    349
                    :.||:...|:|.:.||.|...|     :.::.||| ::|::...|.
MupF            296 QNLGVRRVFDDQRARRAFQLSY-----RPLEHSVADTLQSLSPTAP----    336

TmlF            350 S    350

MupF            337 -    336
```

Figure 4.10 Global alignment of TmlF and MupF proteins using EMBOSS. Needle(Rice et al., 2000). Default parameters for Needleman-Wunsch algorithm used (Needlema.Sb and Wunsch, 1970). Percent similarity between protein sequences 52.7%, which is identified, depending on the level of similarity, as ':' (more similar) and as '.' less similar, while identity is indicated as '|'.

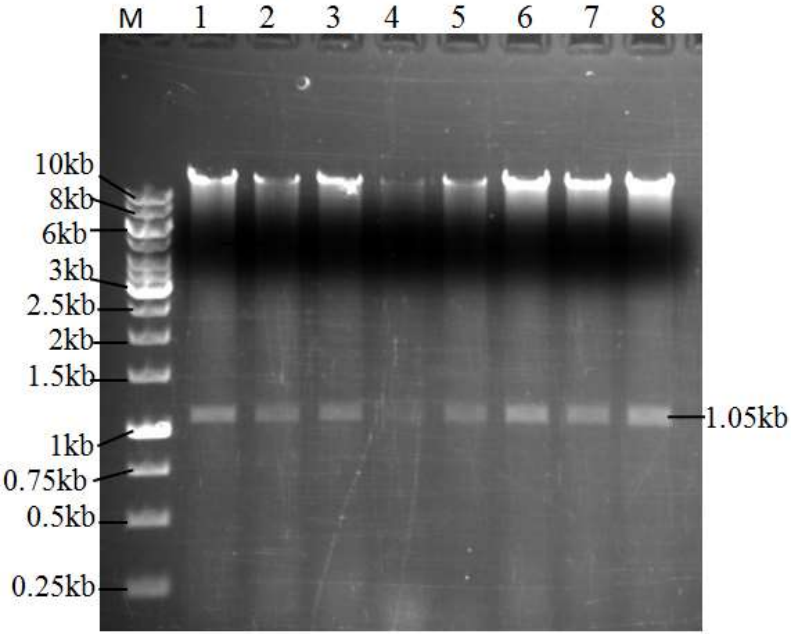
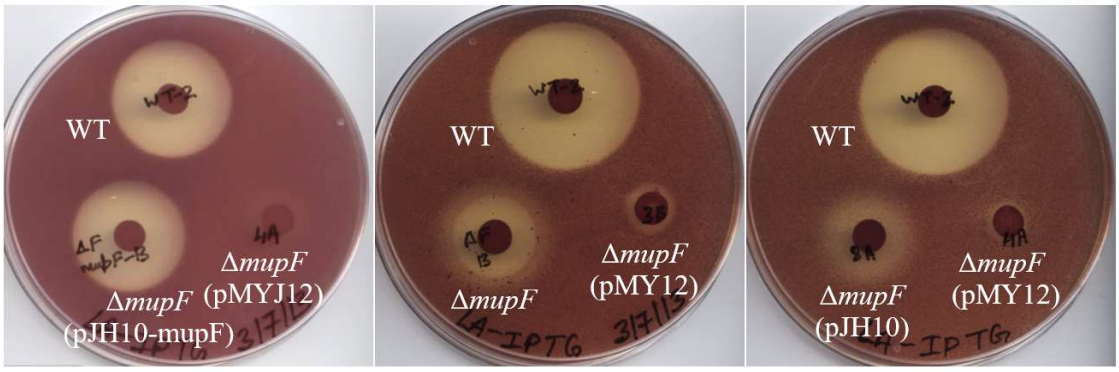


Figure 4.11 Confirmation of the *tmlF* insert in pMY12 expression vector. Double digest with *Kpn*I-*Sac*I yielded a fragment size of 1.05 kb that confirmed the presence of cloned *tmlF* gene in all the 8 clones analysed (lane 1 to 8). Other fragment of size larger than 10 kb in these lanes corresponds to linearized pJH10 in which it was cloned. M, 1 kb ladder.

(A)



(B)

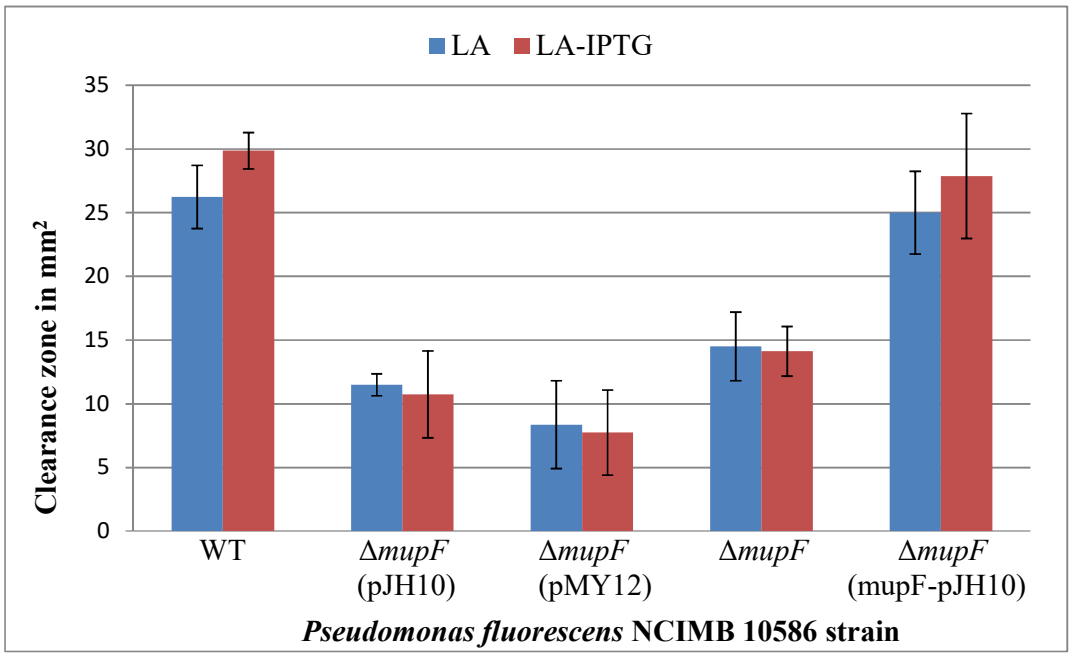


Figure 4.12 Bioassay to determine complementation by *tmlF* of the *mupF* single mutant of *P. fluorescens* NCIMB10586.(A) Bioassay plates induced by 0.5 mM of IPTG showing clearing zones (B) Chart of quantitative bioassay (clearing zone measurements). WT, Δ*mupF* and Δ*mupF*(mupF-pJH10), Δ*mupF*(pJH10) strains of *P. fluorescens* NCIMB10586 used as controls while Δ*mupF*(pMY12) was the test strain. Three replicates were used to collect data. Each bar represents the mean ± standard error of the mean.

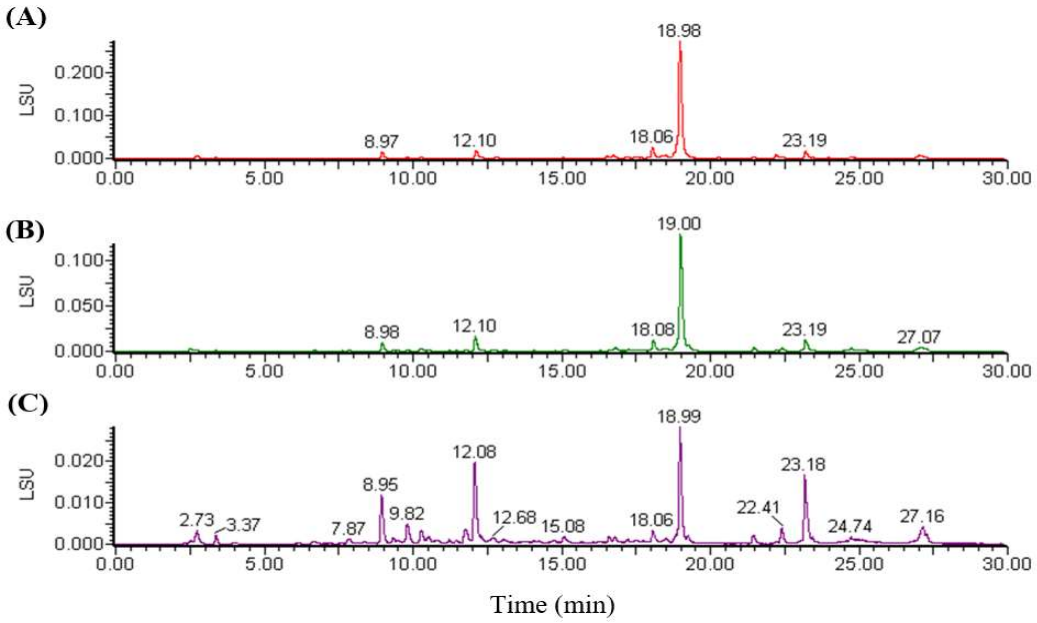


Figure 4.13 LC-MS chromatograms of extracts from strains of *P. fluorescens* NCIMB 10586 for the complementation by *tmlF* of Δ*mupF* single mutant. Under similar conditions neither of the test strains (A) or (B) and control strain (C) produced PA-A (retention time 20.4 min) but test strains still produced mupirocin F (mw498) like NCIMB 10586Δ*mupF* strain (done by Dr Song of University of Bristol, UK). (A) and (B), strain 10586Δ*mupF*(pMY12); (C), 10586Δ*mupF*(pJH10); LSU, light scattering units.

4.2.4 Complementation by TmlJ

In case of complementation by TmlJ, observation of bioassay plates and the quantitative analysis that was displayed as bar diagram (Figures 4.16A and 4.16B) indicated that TmlJ complemented functions of MupJ only when *in trans* expression of *tmlF* ORF in the corresponding single knockout strain *Pseudomonas fluorescens* NCIMB 10586Δ*mupF* was induced by IPTG, supporting the *in silico* predictions which indicated that these proteins shared significant amino acid sequence identity (Fukuda et al., 2011). A global alignment of these two proteins is given in Figure 4.14. In the absence of IPTG, there was no complementation (Figure 4.16). Findings of bioassay were also supported by the results of HPLC analysis of culture supernatants from the constructed *tmlJ* strain, which in the absence

of IPTG, displayed no peak in the chromatogram that was characteristic of pseudomonic acid-A (Figure 4.17). Instead, additional peaks could be seen at the retention time characteristic of pseudomonic acid-A in the presence of IPTG (Figure 4.17). Result was further confirmed by LC-MS (Figure 4.18).

```
TmlJ              1 -MYEMIKVNHNADICTIQFDRPKHRNTLNKQMIEECLHVIAECEERSKIV     49
                    .::..:.....:..::|.||:..||:|:||::|||.|:..|:..|.::
MupJ              1 MNFQATEFRTEGSLAFLRFARPESNNTINRQMVDECLWVLEHCQAHSTVL     50

TmlJ             50 IIKGLPDVFCFGADFSWMNEQEYTNNQQGSEHDPSMLYTLWRRIALGPFI     99
                    :::|.|:||||||||..|::.....:  .|.|..||.:|:|:..|.||
MupJ             51 VLEGGPEVFCFGADFQGMHDDAVRGASE--RHQPERLYEIWQRMVTGSFI     98

TmlJ            100 SIAAVEGRVNAGGMGFVAACDIVLSGSRAQYSLSELIFGILPACVMPFLI    149
                    :::.|.|:.||||||||||.|||||.:.||:|||||||.::||||:||||
MupJ             99 TVSHVRGKANAGGMGFVAASDIVLSHTDAQFSLSELIFDVMPACVLPFLI    148

TmlJ            150 KKVGEQKAHYMTLMTSPFSANEITQWGLADAVADDLDELLRRHLLRLRRL    199
                    ::||.|:|||:||.|.|.||.::.:||||||.||.|.|.||||||||||
MupJ            149 RRVGFQRAHYLTLSTQPVSAAQASQWGLADDVASDSDALLRRHLLRLRRL    198

TmlJ            200 PRDGIVRHKNYMNTLNPQLLEHEQVAIQENRATFTLPSVVANIGRYVQSG    249
                    .:.|:.::|.|||.|:....:..|:||::|...|:.|....||.|||..|
MupJ            199 SKHGVEQYKGYMNRLSTLPFDARQMAIEQNHQVFSNPRTRENIHRYVSHG    248

TmlJ            250 KLPWEG-    255
                    ..|||.
MupJ            249 LFPWEPD    255
```

Figure 4.14 Global alignment of TmlJ and MupJ proteins using EMBOSS Needle(Rice et al., 2000). Default parameters for Needleman-Wunsch algorithm used (Needlema.Sb and Wunsch, 1970). Percent similarity between protein sequences 67.3%, which is identified, depending on the level of similarity, as ':' (more similar) and as '.' less similar, while identity is indicated as '|'.

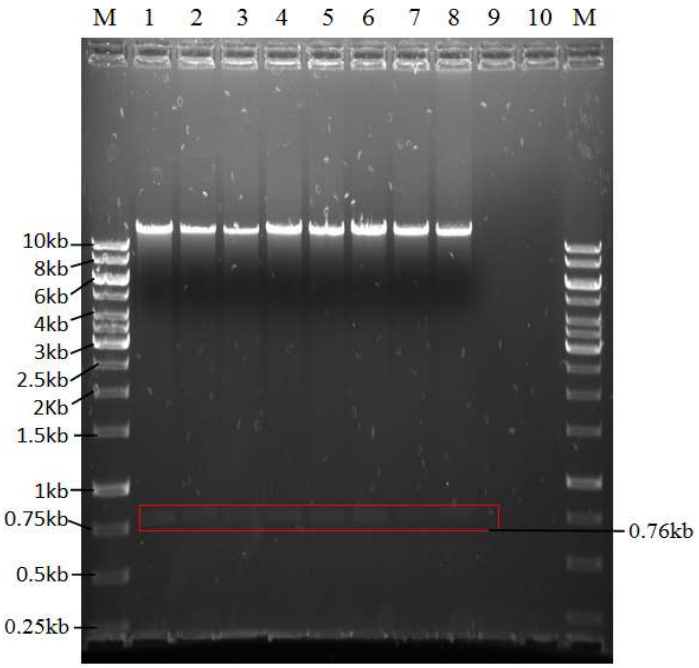
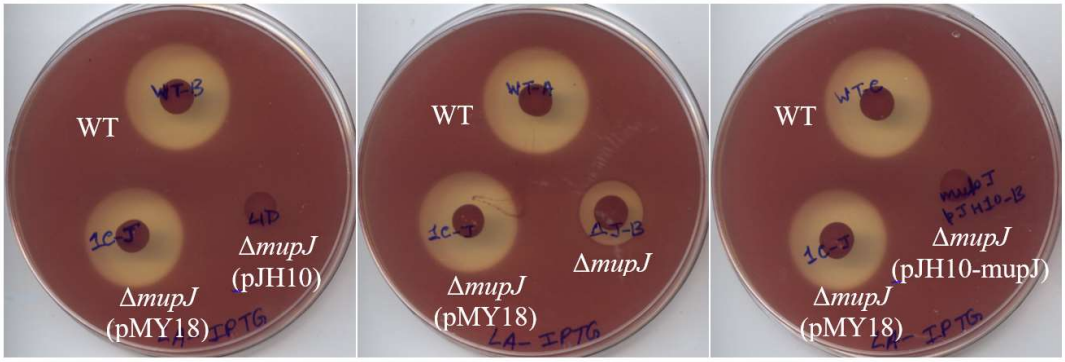


Figure 4.15 Confirmation of the *tmlJ* insert in pMY18 expression vector. Double digest with *Kpn*I and *Xba*I yielded a fragment size of 0.76 kb that confirmed the presence of cloned *tmlJ* gene in these clones. Other fragment of size larger than 10 kb corresponds to linearized pJH10 in which it was cloned. M, 1 kb ladder.

(A)



(B)

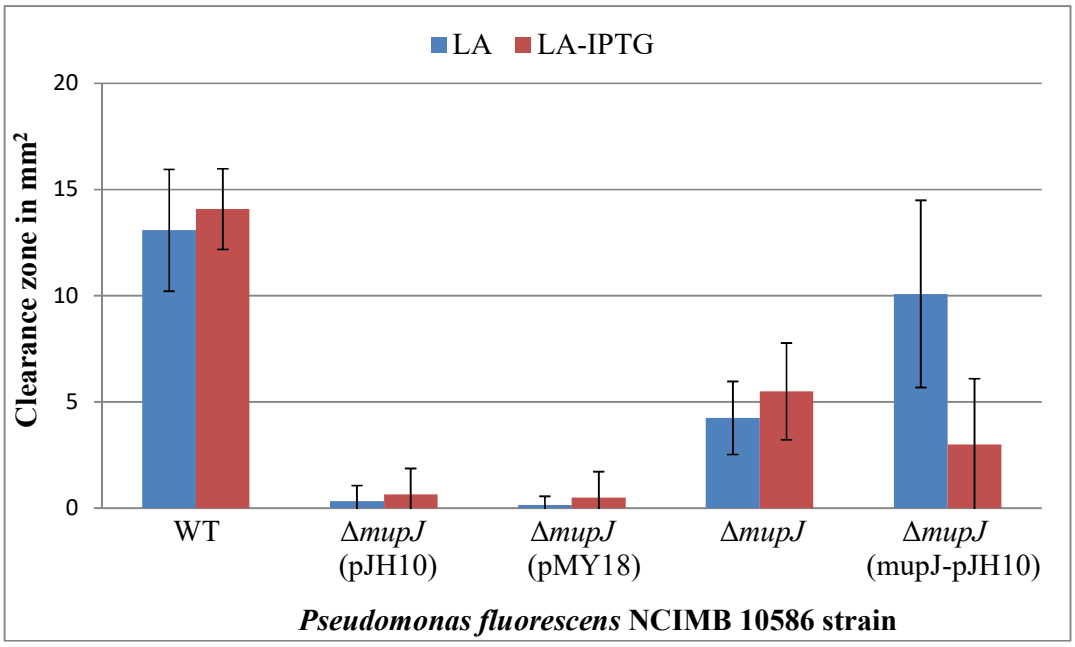


Figure 4.16 Bioassay to determine complementation by *tmlJ* of the *mupJ* single mutant of *P. fluorescens* NCIMB 10586 (A) Bioassay plates induced by 0.5 mM of IPTG showing clearing zones (B) Chart of quantitative bioassay (clearing zone measurements). WT, Δ*mupJ* and Δ*mupJ*(mupJ-pJH10), Δ*mupJ*(pJH10) strains of *P. fluorescens* NCIMB 10586 used as controls while Δ*mupJ*(pMY18) was the test strain. Three replicates were used to collect data. Each bar represents the mean ± standard error of the mean.

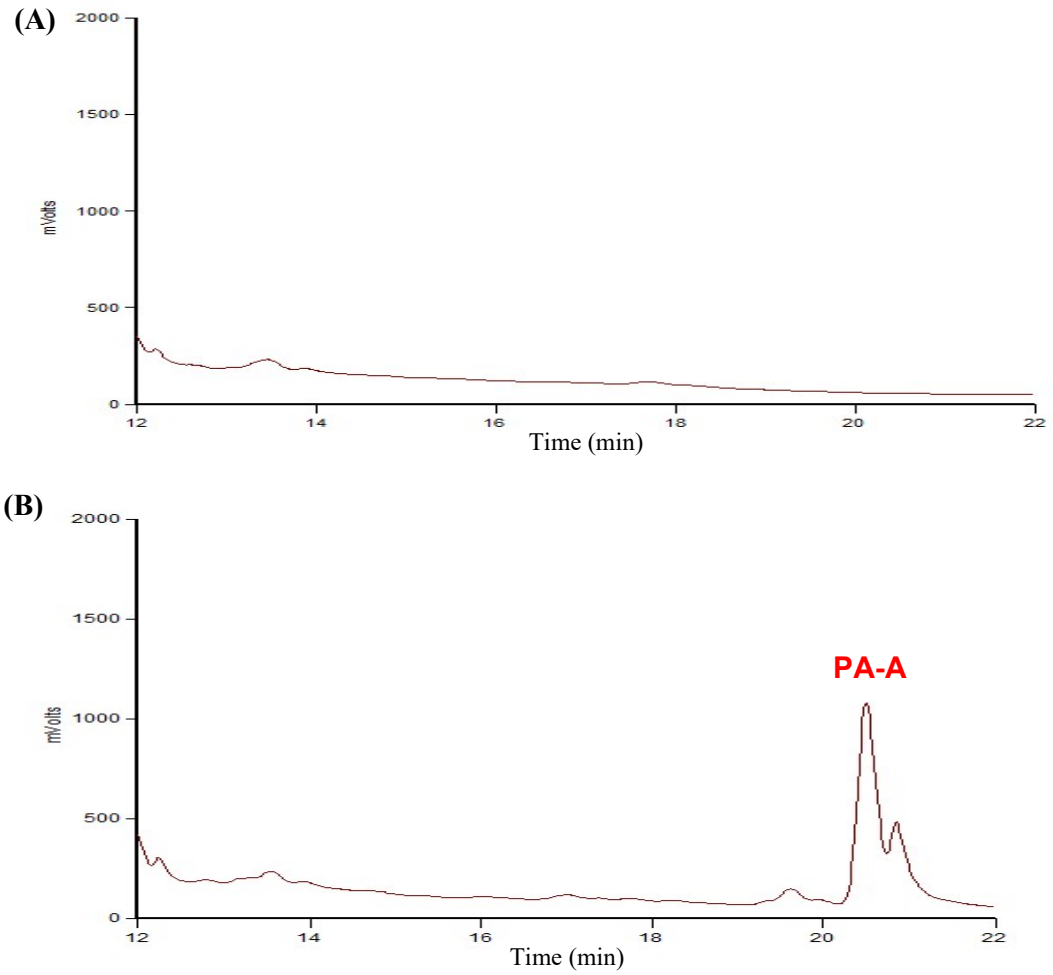


Figure 4.17 HPLC analysis of extracts from strains of *P. fluorescens* NCIMB 10586 for the complementation by *tmlJ* of Δ*mupJ* single mutant.Test strain NCIMB10586Δ*mupJ*(pMY18) complemented in the presence of IPTG and restored the peak characteristic of PA-A at retention time of 20.4 min (B), In the absence of IPTG there was no complementation (A). WT strain used as control (not shown).

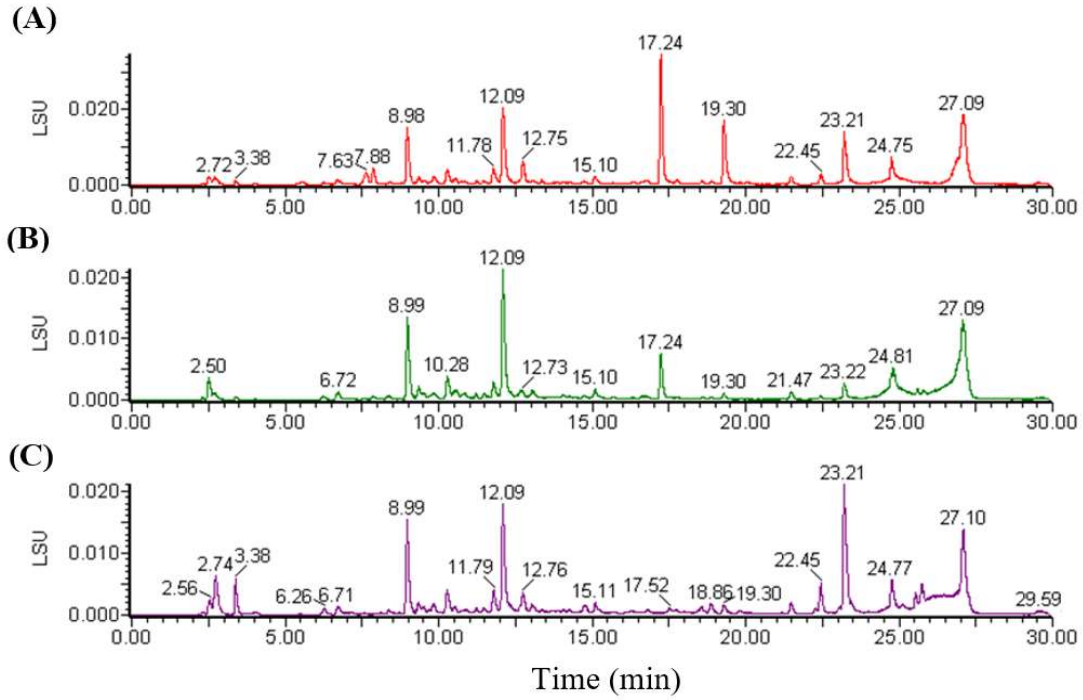


Figure 4.18 LC-MS chromatograms of extracts from strains of *P. fluorescens* NCIMB 10586 for the complementation by *tmlJ* of Δ*mupJ* single mutant. Test strains (A) and (B) restored the production of PA-A in the presence of IPTG (0.5 mM) while control strain (C) did not show any complementation (data by Dr Song of University of Bristol, UK). (A) and (B), 10586Δ*mupJ*(pMY18); (C), 10586Δ*mupJ*(pJH10); LSU, light scattering units.

4.2.5 Complementation by TmlK

Observation of bioassay plates and accompanying bar diagram indicated that there was no complementation of functions of MupK protein by TmlK upon *in trans* expression of *tmlK* ORF in the Δ*mupK* single mutant strain *Pseudomonas fluorescens* NCIMB 10586Δ*mupK* (Figure 4.22). This was despite MupK shared significant amino acid sequence identity with the N-terminal end of its TmpE protein (Figure 4.20) which represented putative enoyl-CoA hydratase similar to MupK (Fukuda et al., 2011). Figure 4.19 gives global alignment of MupK protein with TmpE. The *mupK* ORF cloned in pJH10 vector which was used as negative control also did not show any complementation as was reported earlier by Hothersall et al., 2007 (Figure 4.22).

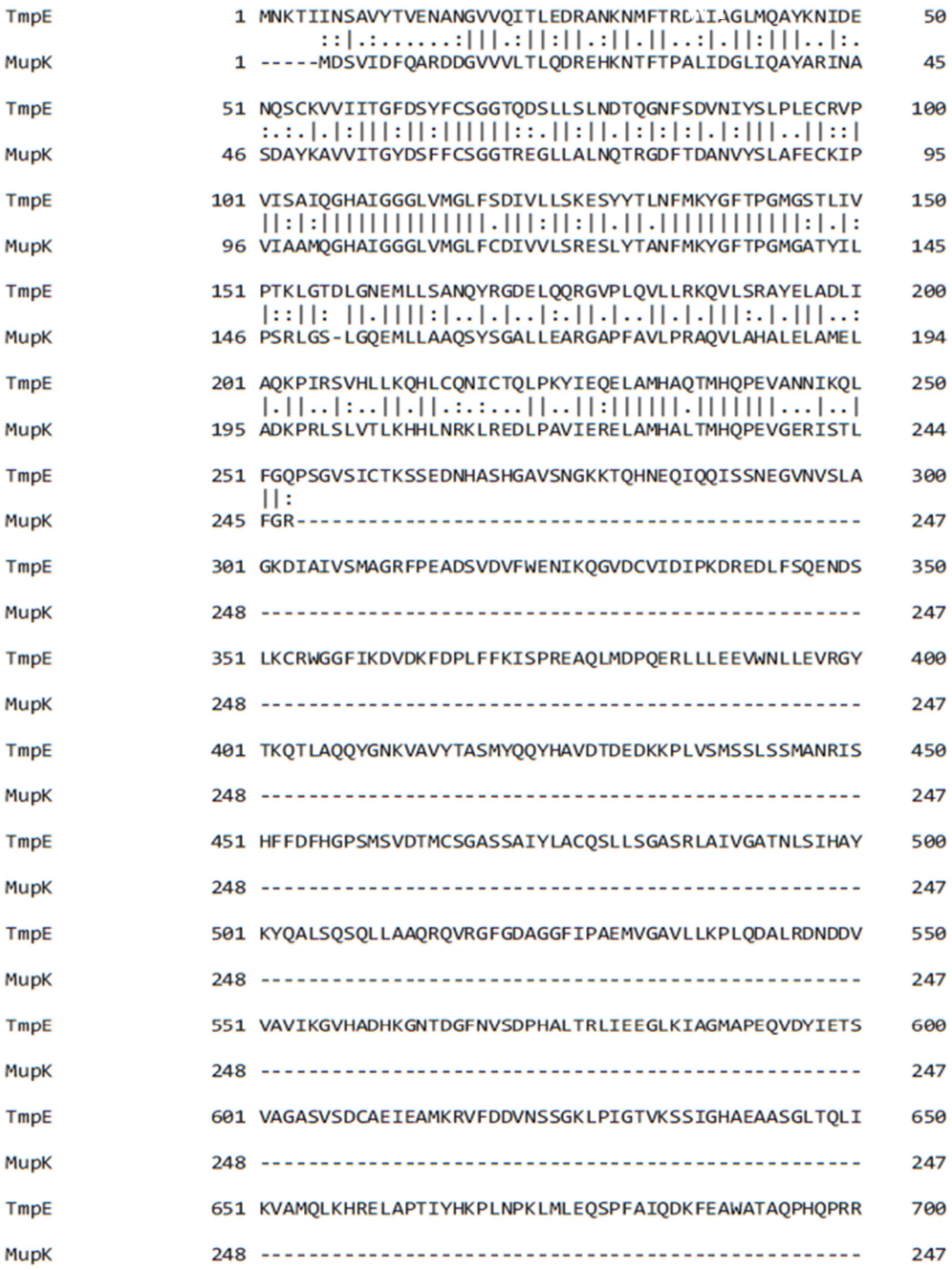


Figure 4.19 Global alignment of MupK with TmpE protein using EMBOSS Needle.(Rice et al., 2000). Default parameters for Needleman-Wunsch algorithm used (Needlema.Sb and Wunsch, 1970). MupK protein has 247 aa that aligns with the N-terminal of TmpE protein of size 1444 aa, therefore alignment for only 1-700 aa from its N-terminal end is shown. Percent similarity between protein sequences 12.7%, which is identified, depending on the level of similarity, as ':' (more similar) and as '.' less similar, while identity is indicated as '|'.

```
TmpE             17 NGVVQITLEDRANKNMFTRDIIAGLMQAYKNIDENQSCKVVIITGFDSYF     66
                    :|||.:||:||.:||.||..:|.||:|||..|:.:.:.|.|:|||:||:|
MupK             12 DGVVVLTLQDREHKNTFTPALIDGLIQAYARINASDAYKAVVITGYDSFF     61

TmpE             67 CSGGTQDSLLSLNDTQGNFSDVNIYSLPLECRVPVISAIQGHAIGGGLVM    116
                    |||||::.||:||.|:|:|:|.|:|||..||::|||:|:||||||||||
MupK             62 CSGGTREGLLALNQTRGDFTDANVYSLAFECKIPVIAAMQGHAIGGGLVM    111

TmpE            117 GLFSDIVLLSKESYYTLNFMKYGFTPGMGSTLIVPTKLGTDLGNEMLLSA    166
                    |||.|||:||:||.||.||||||||||||:|.|:|::||: ||.||||:|
MupK            112 GLFCDIVVLSRESLYTANFMKYGFTPGMGATYILPSRLGS-LGQEMLLAA    160

TmpE            167 NQYRGDELQQRGVPLQVLLRKQVLSRAYELADLIAQKPIRSVHLLKQHLC    216
                    ..|.|..|:.||.|..||.|.|||:.|.|||..:|.||..|:..||.||.
MupK            161 QSYSGALLEARGAPFAVLPRAQVLAHALELAMELADKPRLSLVTLKHHLN    210

TmpE            217 QNICTQLPKYIEQELAMHAQTMHQPEVANNIKQLFGQ    253
                    :.:...||..||:||||||.|||||||...|..|||:
MupK            211 RKLREDLPAVIERELAMHALTMHQPEVGERISTLFGR    247
```

Figure 4.20 Local alignment of TmpE and MupK proteins using EMBOSS Water. (Rice, et al., 2000). Default parameters for Smith-Waterman algorithm used (Smith and Waterman, 1981). Percent similarity between protein sequences 75.9%, which is identified, depending on the level of similarity, as ':' (more similar) and as '.' less similar, while identity is indicated as '|'.

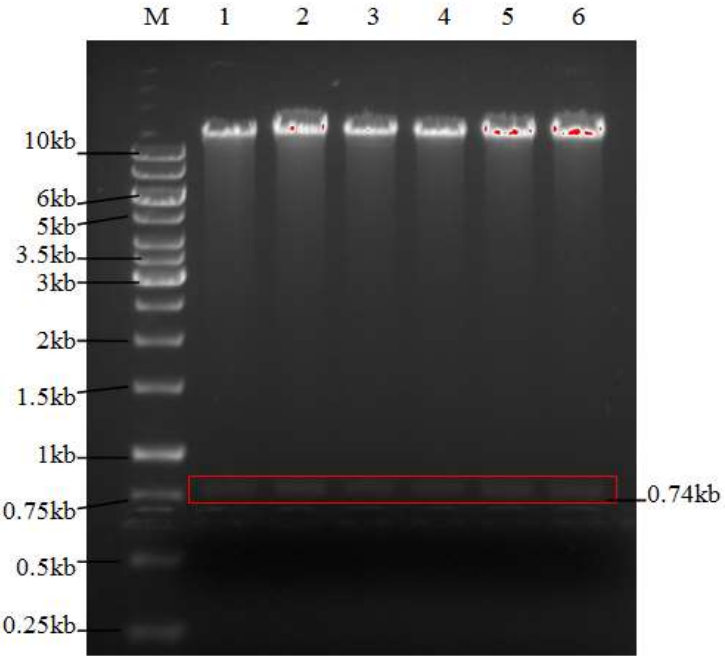
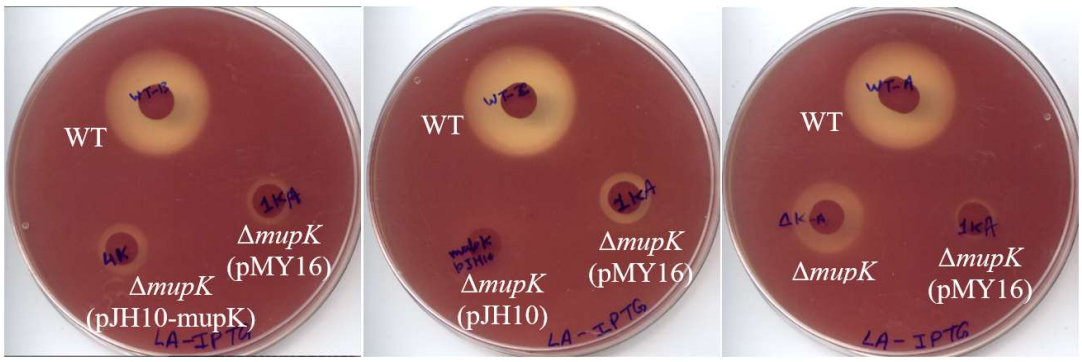


Figure 4.21 Confirmation of the *tmlK* insert in pMY16 expression vector. Double digest with *Kpn*I and *Xba*I yielded a fragment size of 0.749 kb that confirmed the presence of cloned *tmlK* gene in these clones. Other fragment larger than 10 kb corresponds to linearized pJH10 in which it was cloned. M, 1 kb ladder.

(A)



(B)

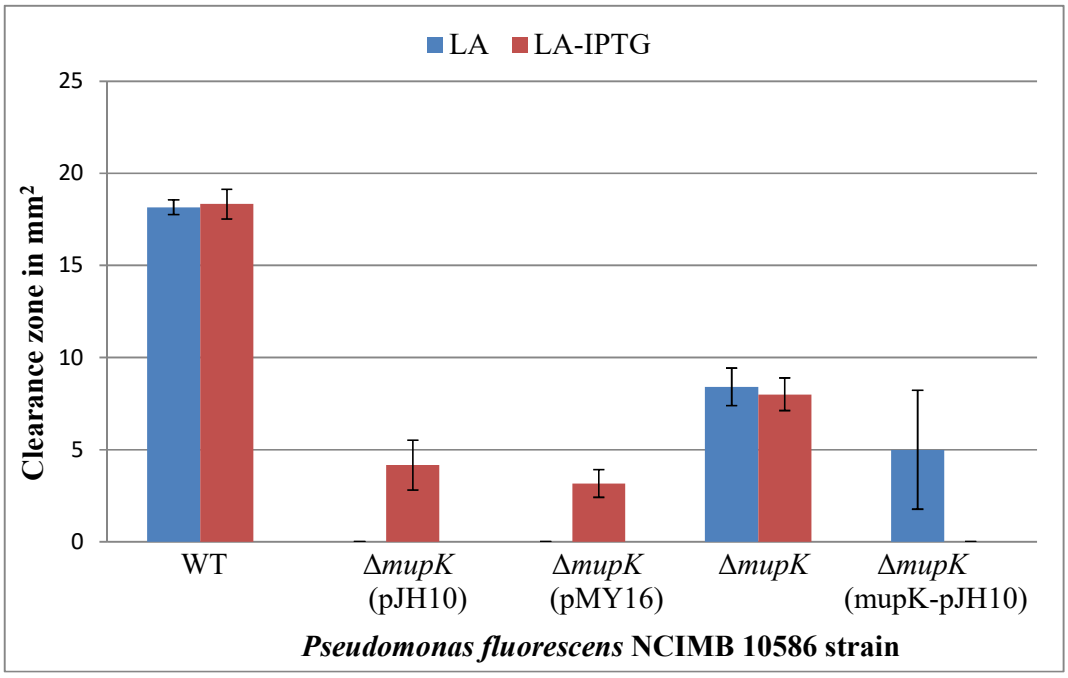


Figure 4.22 Bioassay to determine complementation by *tmlK* of the *mupK* single mutant of *P. fluorescens* NCIMB 10586. (A) Bioassay plates induced by 0.5 mM of IPTG showing clearing zones (B) Chart of quantitative bioassay (clearing zone measurements). WT, Δ*mupK* and Δ*mupK*(mupK-pJH10), Δ*mupK*(pJH10) strains of *P. fluorescens* NCIMB 10586 used as controls while Δ*mupK*(pMY16) was the test strain. Three replicates were used to collect data. Each bar represents the mean ± standard error of the mean.

4.2.6 Complementation by TmlQ

No complementation was shown by TmlQ as can be observed from bioassay plates and the accompanying quantitative analysis displayed as bar diagram (Figures 4.25A and 4.25B), when *tmlQ* ORF was expressed, *in trans,* in the corresponding single knockout strain *Pseudomonas fluorescens* NCIMB 10586Δ*mupQ*. This was despite the fact that the two proteins, TmlQ and MupQ, shared significant amino acid sequence identity (Fukuda et al., 2011). Figure 4.24 gives global alignment of these two proteins. LC-MS chromatogram also displayed non-complemntation of Δ*mupQ* strain by *tmlQ* (Figure 4.26).

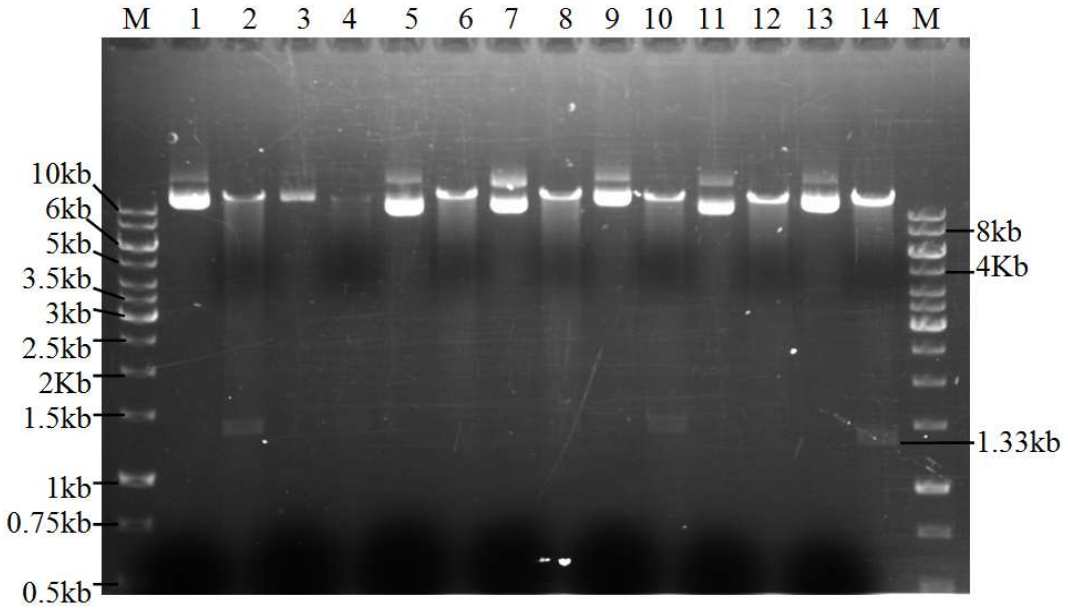
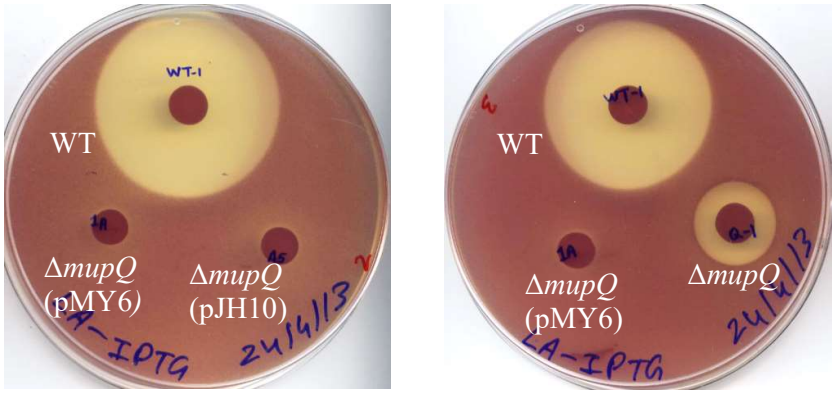


Figure 4.23 Confirmation of the *tmlQ* insert in pMY6 expression vector. Double digest with *Eco*RI and *Xba*I yielded a fragment size of 1.33 kb (lane 2, 10 and 14) that confirmed the presence of cloned *tmlQ* gene in these clones. Other fragment of size larger than 10 kb in these lanes corresponds to linearized pJH10 in which it was cloned. M, 1 kb ladder**.**

Figure 4.24 Global alignment of TmlQ and MupQ proteins using EMBOSS Needle(Rice et al., 2000). Default parameters for Needleman-Wunsch algorithm used (Needlema.Sb and Wunsch, 1970). Percent similarity between protein sequences 64.3%, which is identified, depending on the level of similarity, as ':' (more similar) and as '.' less similar, while identity is indicated as '|'.

(A)



(B)

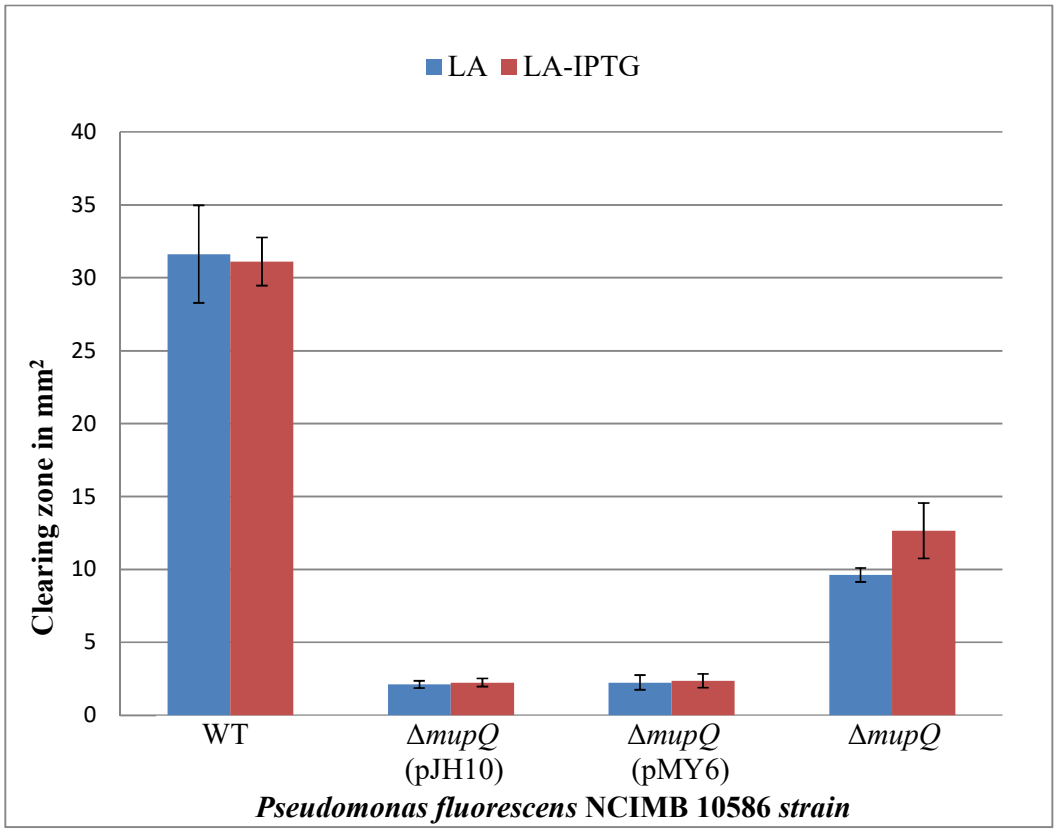


Figure 4.25 Bioassay to determine complementation by *tmlQ* of *mupQ* single mutant of *P. fluorescens* NCIMB 10586 (A) Bioassay plates induced by 0.5 mM of IPTG showing clearing zones (B) Chart of quantitative bioassay (clearing zone measurements). WT, Δ*mupQ* and Δ*mupQ*(pJH10) strains of *P. fluorescens* NCIMB 10586 used as controls while *ΔmupQ*(pMY6) was the test strain. Three replicates were used to collect data. Each bar represents the mean ± standard error of the mean.

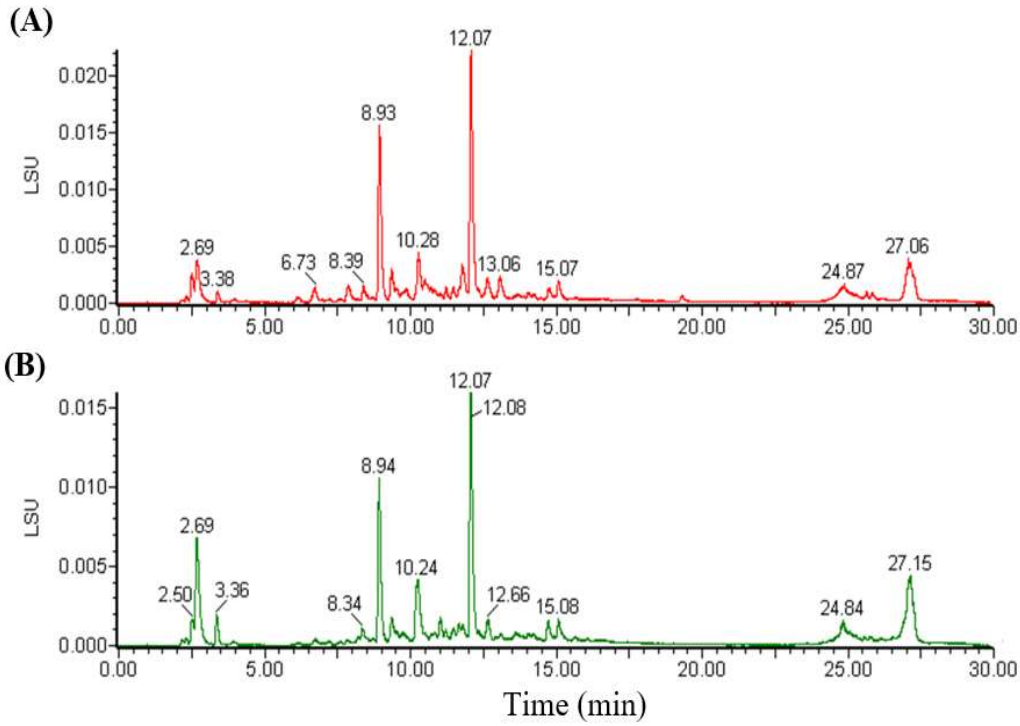


Figure 4.26 LC-MS chromatograms of extracts from strains of *P. fluorescens* NCIMB 10586 for the complementation by *tmlQ* of Δ*mupQ* single mutant. Neither of the two test strains (A) or (B) restored the production of PA-A (data by Dr Song of the University of Bristol, UK). (A) and (B) are test strains 10586Δ*mupQ*(pMY6). LSU, light scattering units.

4.2.7 Complementation by TmlS

In case of TmlS complementation, inspection of bioassay plates followed by quantitative analysis displayed a bar-diagram (Figures 4.29A and 4.29B) as a result of *in trans* expression of *tmlS* ORF in the corresponding single knockout strain *Pseudomonas fluorescens* NCIMB 10586Δ*mupS*, which indicated that TmlS partially complemented functions of MupS when IPTG was included in the culture medium. The result was in agreement, to some extent, with the *in silico* predictions made on the basis of the fact that these proteins shared significant amino acid sequence identity (Fukuda et al., 2011). A global alignemnt of these two proteins is shown in Figure 4.27. Complementation was further confirmed by HPLC and also by LC-

MS that displayed partial restoration of the peak characteristic of pseudomonic acid-A in these chromatograms (Figures 4.30 and 4.31).

```
TmlS              1 ----------------MESENGTVIISGGSRGLGFEIASQFLAQGYAVAT     34
                                    ..:..||:|:||||:|||......||..||.|||
MupS              1 MTDAVSDALHTRHDSAPSAVKGTIIVSGGSQGLGLTTVRCFLEAGYNVAT     50

TmlS             35 FSRGSSEQVVALAEEPRFFWKSVDGSDYQALTEFLKEAQKKLGNIVGLVN     84
                    |||..|..|..|:|...|.|:::|.:||.|||.|:::.:|:.|.:.||||
MupS             51 FSRRESPAVTELSERADFHWQALDCTDYSALTAFVQQVEKRFGGLDGLVN    100

TmlS             85 NAAIGADGVLSTMRTSDIDRAIDVNLKAQIYLSKLVSKKLLQNRDGFIIN    134
                    |||.|.:|:|||||.:|||.|:|:|||.|:||:|||:.|||:...|.::|
MupS            101 NAATGVEGILSTMRVADIDSALDINLKGQLYLTKLVTAKLLKRGAGSVVN    150

TmlS            135 ISSIMGVRGLPGVSIYSATKAAMDGITRSLAKELGRKGIRVNSVSPGYFS    184
                    :|||..:||..|:::||||||||||:|||||||||.:|||||||||||||
MupS            151 VSSINALRGHSGLTVYSATKAAMDGLTRSLAKELGPRGIRVNSVSPGYFS    200

TmlS            185 SDMVKDLSDDILRKIERRTPLGRLGTQDEIAKLVLFLATEGKFITGQNIV    234
                    ||||||||...|.:|||||||||||||.|:|.|:|:|...|.|:|||||.
MupS            201 SDMVKDLSPQTLSRIERRTPLGRLGTQQEVADLILYLVDRGTFVTGQNIA    250

TmlS            235 IDGGFTC    241
                    :||||||
MupS            251 VDGGFTC    257
```

Figure 4.27 Global alignment of TmlS and MupS proteins using EMBOSS Needle(Rice et al., 2000). Default parameters for Needleman-Wunsch algorithm used (Needlema.Sb and Wunsch, 1970). Percent similarity between protein sequences 72%, which is identified, depending on the level of similarity, as ':' (more similar) and as '.' less similar, while identity is indicated as '|'.

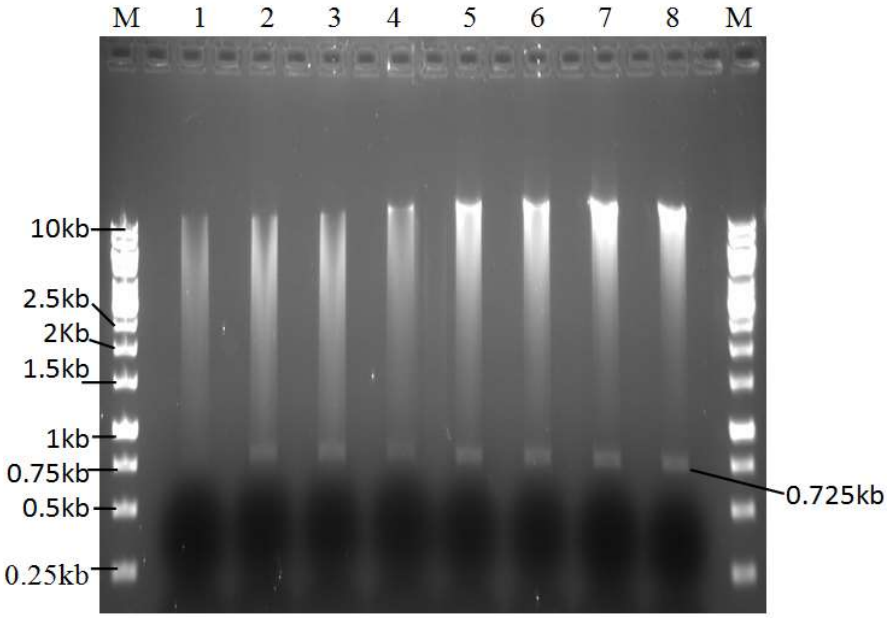
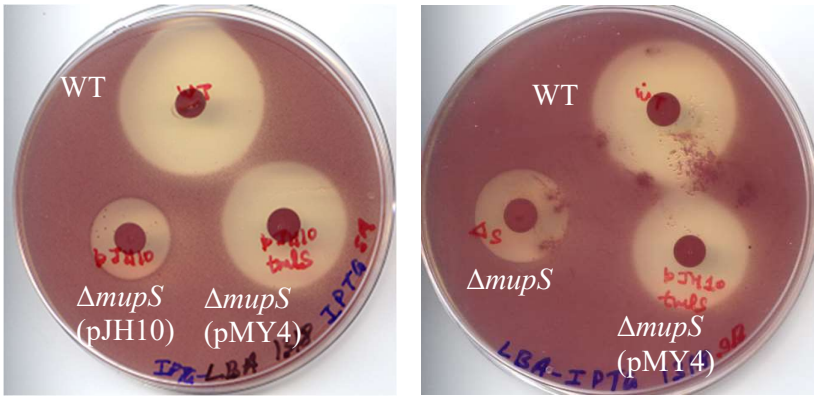


Figure 4.28 Confirmation of the *tmlS* insert in pMY4 expression vector. Double digest with *Kpn*I and *Xba*I yielded a fragment size of 0.725 kb that confirmed presence of cloned *tmlS* gene. Other fragment of size larger than 10 kb corresponds to linearized pJH10 in which it was cloned. M, 1 kb ladder.

(A)



(B)

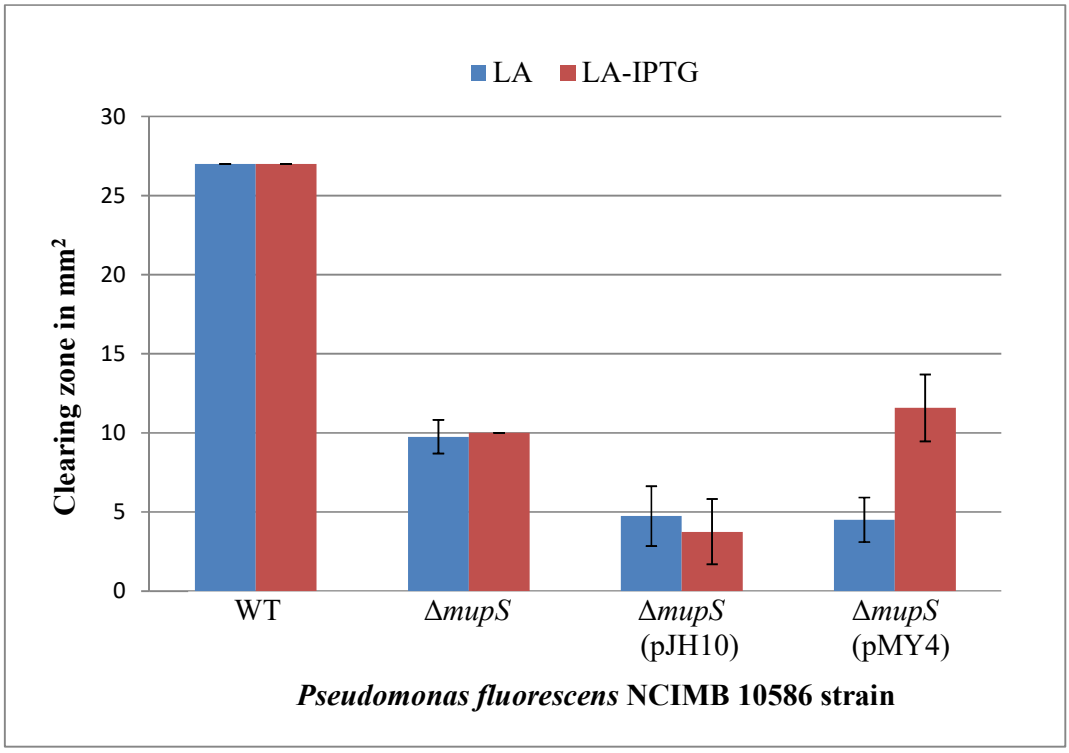


Figure 4.29 Bioassay to determine complementation by *tmlS* of *mupS* single mutant of *P. fluorescens* NCIMB 10586 (A) Bioassay plates induced by 0.5 mM of IPTG showing clearing zones (B) Chart of quantitative bioassay (clearing zone measurements). WT, *ΔmupS* and *ΔmupS*(pJH10) strains of *P. fluorescens* NCIMB 10586 used as controls while *ΔmupS*(pMY4) was the test strain. Three replicates were used to collect data. Each bar represents the mean ± standard error of the mean.

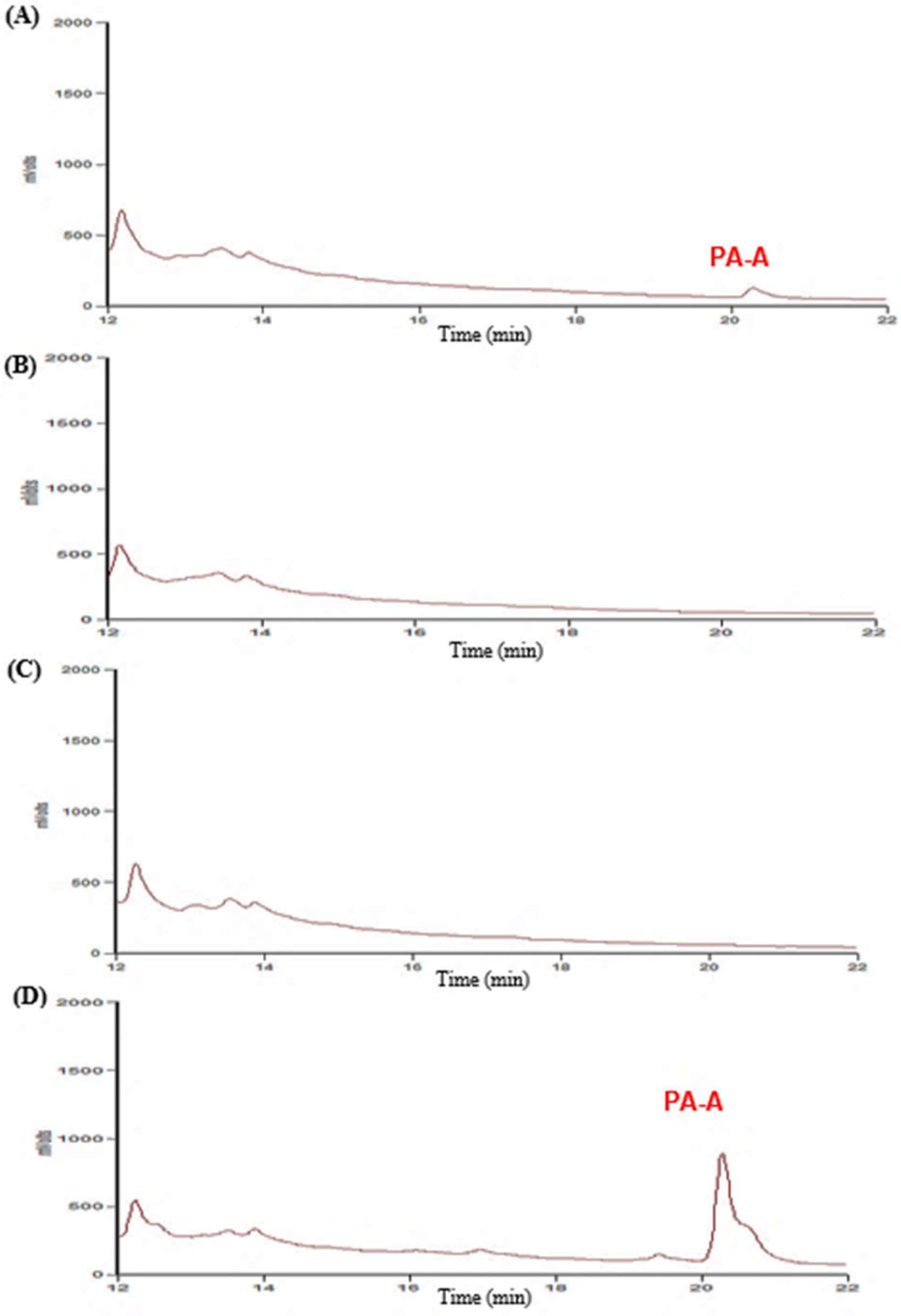


Figure 4.30 HPLC analysis of extracts from strains of *P. fluorescens* NCIMB 10586 for the complementation by *tmlS* of Δ*mupS* single mutant. (A) Test strain NCIMB 10586Δ*mupS*(pMY4) partially restored the production of PA-A in the presence IPTG (of 0.5 mM) (retention time 20.4 min), (B) Non-production of PA-A by control strain 10586Δ*mup*S (pJH10), (C) Non-production of PA-A by control strain 10586Δ*mupS,* (D) Positive control WT strain.

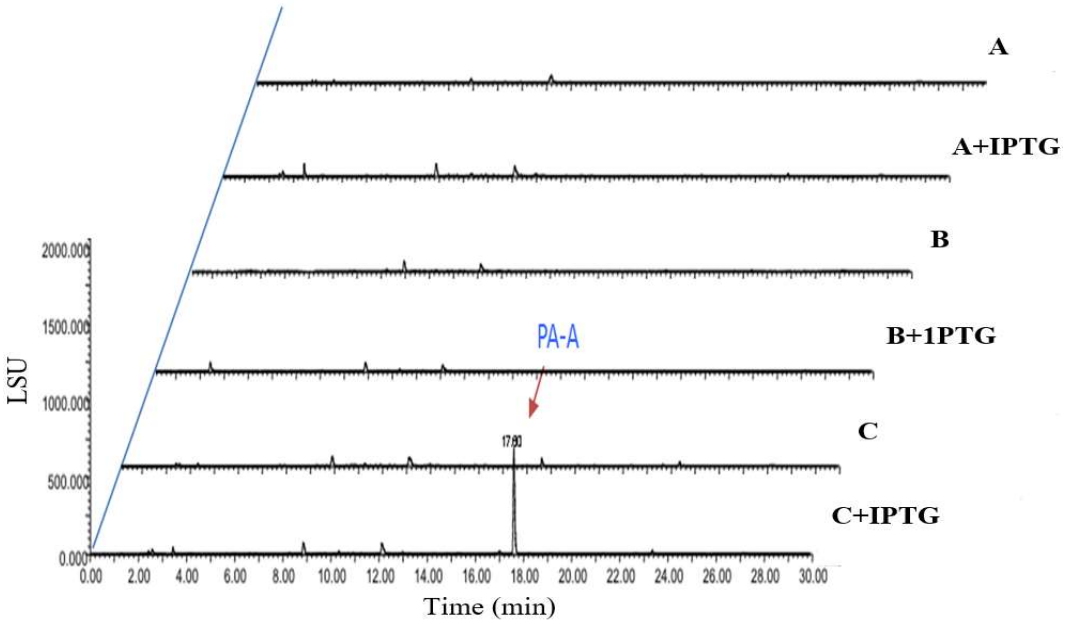


Figure 4.31 LC-MS chromatograms of extracts from strains of *P. fluorescens* NCIMB 10586 for the complementation by *tmlS* of Δ*mupS* single mutant. Test strain C restored the production of PA-A only in the presence IPTG (of 0.5 mM) while both of the control strains A and B did not produce any PA-A (data by Dr Song of the University of Bristol, UK). A, B, and C are strains 10586Δ*mupS,* 10586Δ*mupS*(pJH10) and 10586Δ*mupS*(pMY4), respectively. LSU, light scattering units.

4.2.8 Complementation by TacpD

TacpD did not complement functions of MacpD protein which can be observed from bioassay plates and the accompanying quantitative analysis displayed as bar diagram (Figures 4.33A and 4.33B) obtained as a result of *in trans* expression of *tacpD* ORF in the corresponding single knockout strain *Pseudomonas fluorescens* NCIMB 10586*ΔtacpD*. This was despite the fact that both TacpD and MacpD proteins shared significant amino acid sequence identity (Fukuda et al., 2011). Figure 4.32 gives a global alignment of these two proteins. LC-MS chromatogram also displayed non-complemntation by TacpD of functions of MacpD (Figure 4.34).

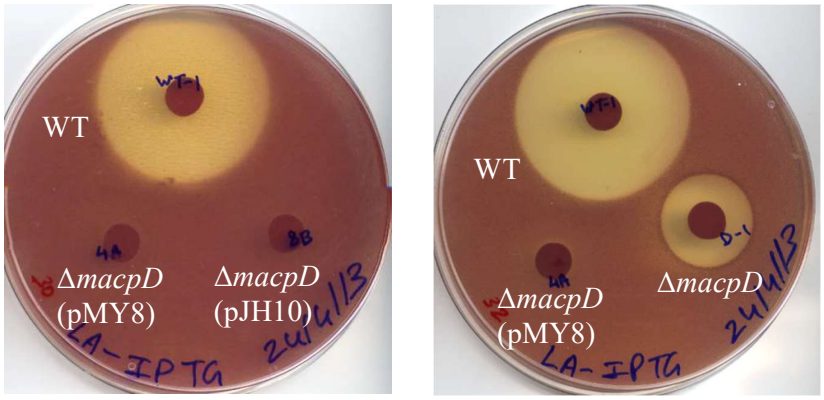
```
TacpD             1 MLNVQQIKNEVLDSIIQEVATILSEKLLPITHIENSTAFSELSLTSLDLA     50
                    ||| .|:.::|.|.:..::|.:|..|..|:..:|..:.||:|.|:|||||
MacpD             1 MLN-HQVMDQVFDQVEHQIAQVLGAKGGPLVAVEIDSRFSDLGLSSLDLA     49

TacpD            51 ELISNLEARYEVDPFEELVAITSIVTIEDLASAYA-LSLSGNTEDSHDLL     99
                    .|||||||.|..|||.:.|||||||:.||..||| ..:.|.:.|..|..
MacpD            50 TLISNLEAVYGTDPFADAVAITSIVTVADLCRAYAQQGVPGPSPDPLDAQ     99

TacpD           100 SDELKAIKNQVR    111
                    ..:|:.:
MacpD           100 LRDLRQL-----    106
```

Figure 4.32 Global alignment of TacpD and MacpD proteins using EMBOSS Needle(Rice et al., 2000). Default parameters for Needleman-Wunsch algorithm used (Needlema.Sb and Wunsch, 1970). Percent similarity between protein sequences 59.8%, which is identified, depending on the level of similarity, as ':' (more similar) and as '.' less similar, while identity is indicated as '|'.

(A)



(B)

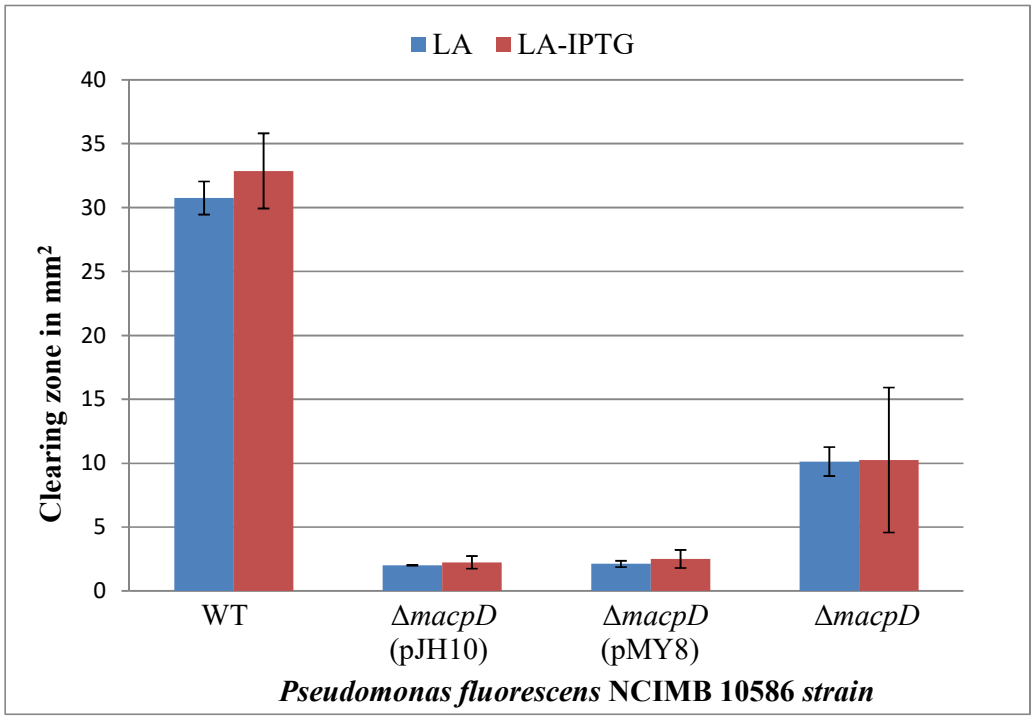


Figure 4.33 Bioassay to determine complementation by *tacpD* of *macpD* single mutant of *P. fluorescens* NCIMB 10586 (A) Bioassay plates induced by 0.5 mM of IPTG showing clearing zones (B) Chart of quantitative bioassay (clearing zone measurements). WT, Δ*macpD* and Δ*macpD*(pJH10) strains of *P. fluorescens* NCIMB 10586 used as controls while Δ*macpD*(pMY8) was the test strain. Three replicates were used to collect data. Each bar represents the mean ± standard error of the mean.

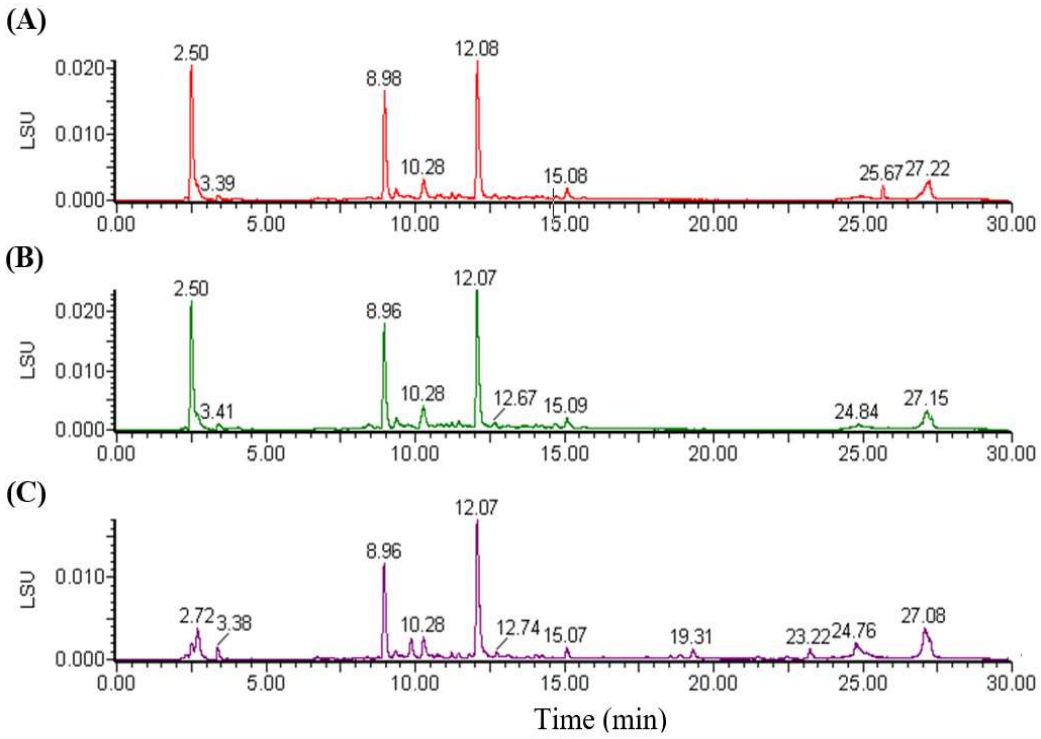


Figure 4.34 LC-MS chromatograms of extracts from strains of *P. fluorescens* NCIMB 10586 for the complementation by *tacpD* of Δ*macpD* single mutant. None of the test strains (A) or (B) or control strain (C) restored the production of PA-A despite induction by IPTG (0.5 mM) (data by Dr Song of the University of Bristol, UK). (A) and (B), 10586Δ*macpD*(pMY8); (C), 10586Δ*macpD*(pJH10); LSU, light scattering units.

4.2.9 Complementation by TmpF

Observation of bioassay plates and quantitative analysis that resulted into accompanying bar diagram (Figures 4.38A and 4.38B) obtained as a result of *in trans* expression of *tmpF* ORF in the corresponding single mutant strain *Pseudomonas fluorescens* NCIMB 10586Δ*mmpF* indicated that there was no complementation of functions of MmpF protein by TmpF. This was in disagreement with the *in silico* prediction made on the basis of the fact that these two proteins shared significant amino acid sequence identity (Fukuda et al., 2011). A global alignment of TmpF and MmpF proteins is given in Figure 4.35. Results of LC-MS also confirmed non-complementation of functions of MmpF protein by TmpF (Figure 4.37).

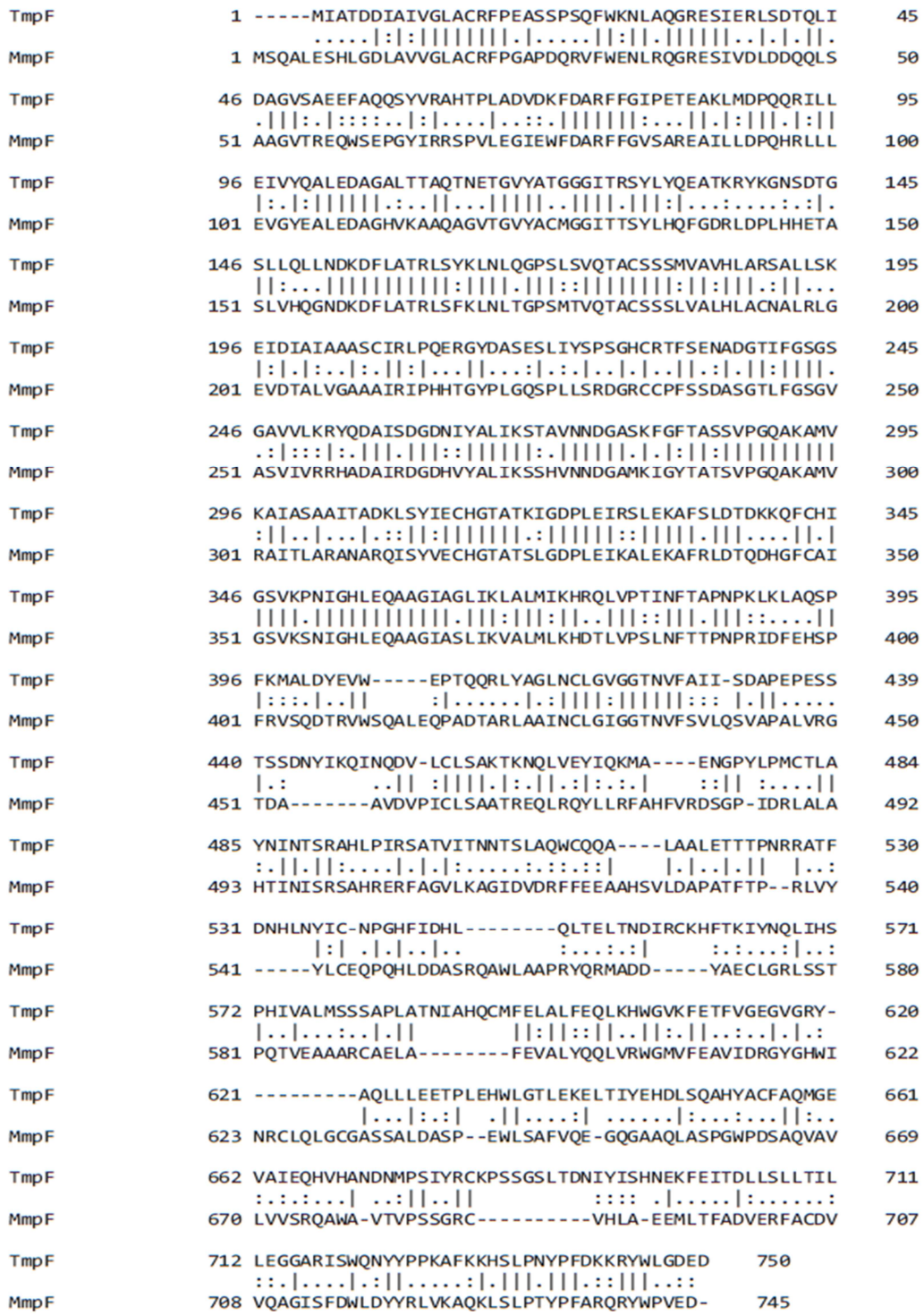


Figure 4.35 Global alignment of TmpF and MmpF proteins using EMBOSS Needle(Rice et al., 2000). Default parameters for Needleman-Wunsch algorithm used (Needlema.Sb and Wunsch, 1970). Percent similarity between protein sequences 57.8% which is identified, depending on the level of similarity, as ':' (more similar) and as '.' less similar while identity is indicated as '|'.

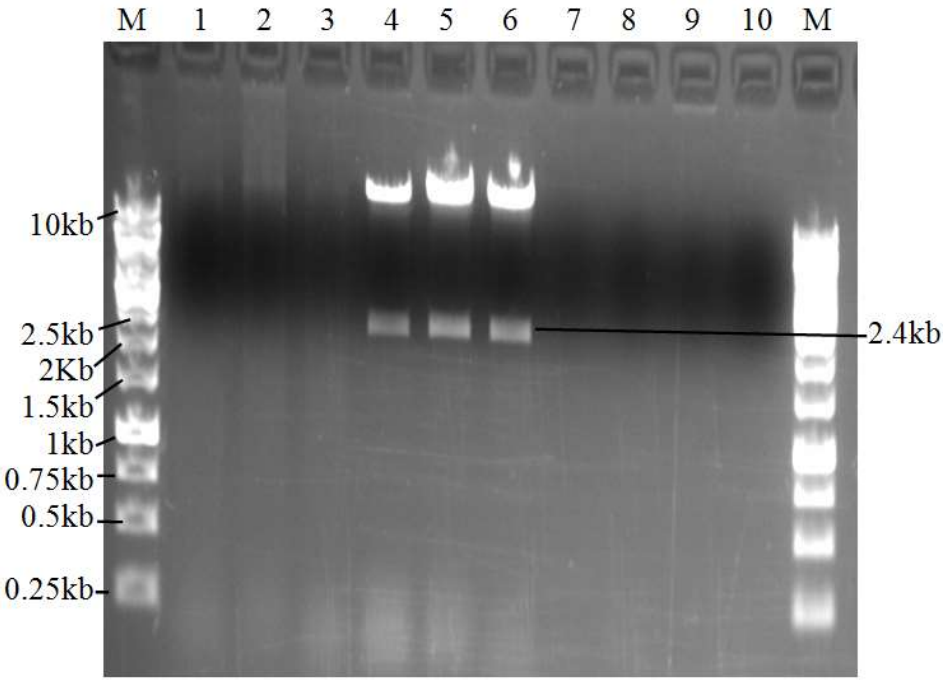


Figure 4.36 Confirmation of the *tmpF* insert in pMY10 expression vector.Double digest with *Eco*RI and *Xba*I yielded a fragment size of 2.4 kb that confirmed the presence of cloned *tmpF* gene in these clones. Other fragment larger of size than 10 kb corresponds to linearized pJH10 in which it was cloned. M, 1 kb ladder.

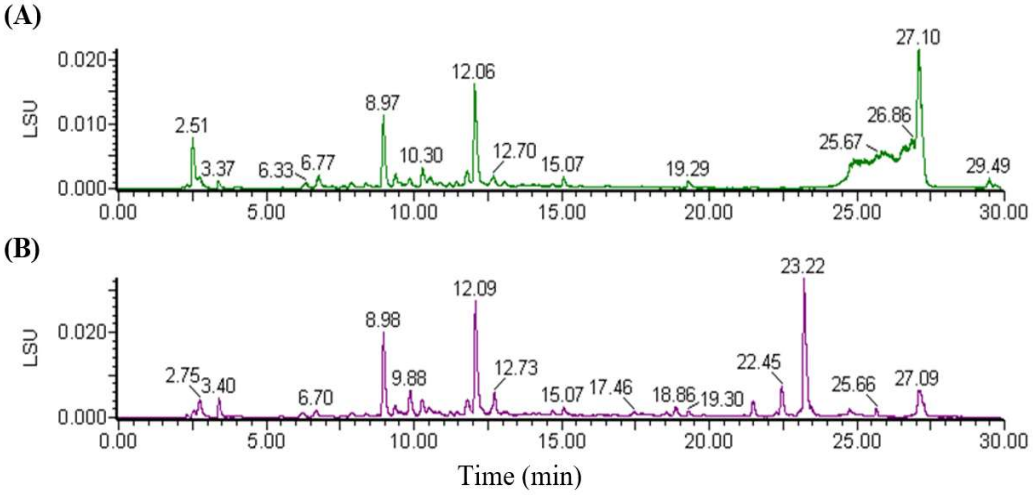
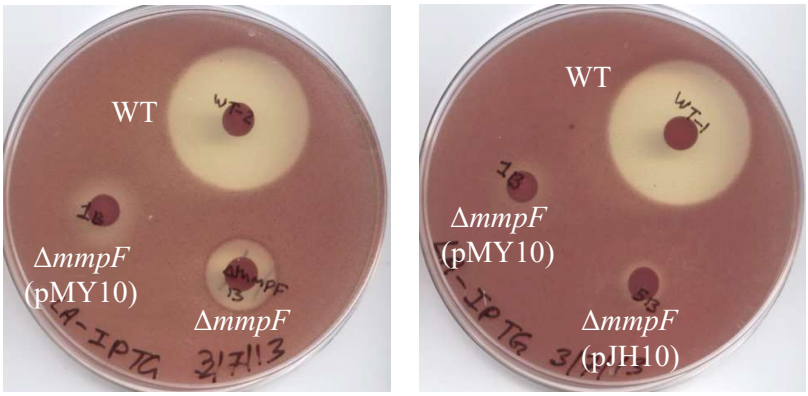


Figure 4.37 LC-MS chromatograms of extracts from strains of *P. fluorescens* NCIMB 10586 for the complementation by *tmpF* of Δ*mmpF* single mutant. Test strain (A) and control strain (B) did not show any complementation as there was no restoration of peak characteristic of PA-A (done by Dr Song of University of Bristol, UK). (A) and (B) are 10586Δ*mmpF*(pMY10) and 10586Δ*mmpF*(pJH10), respectively. LSU, Light scattering units.

(A)



(B)

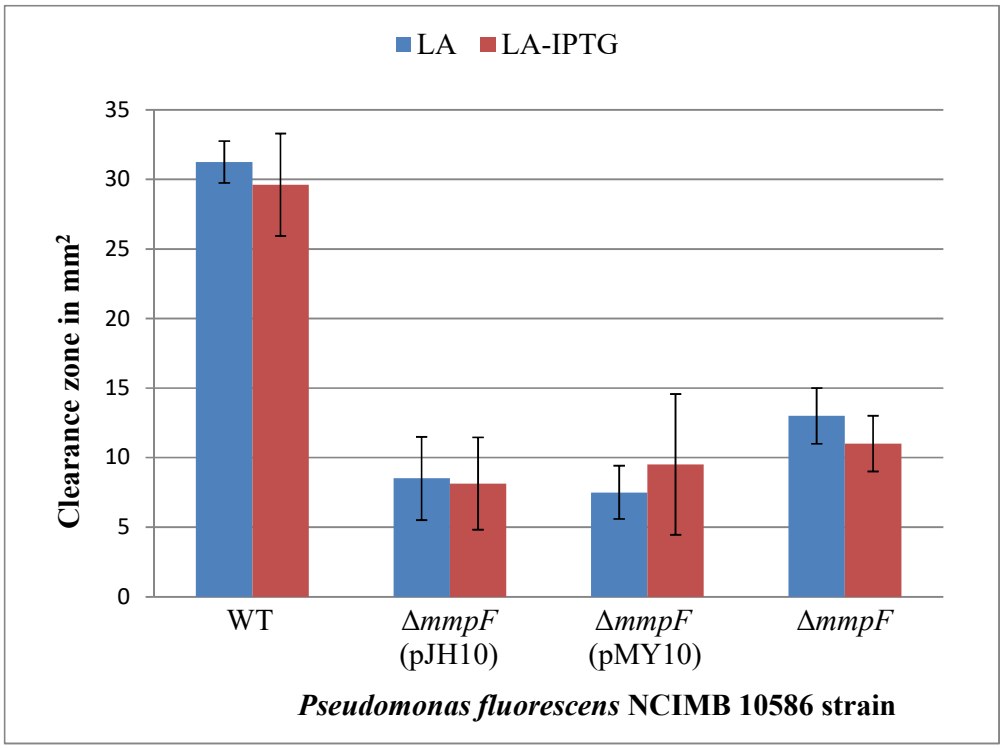


Figure 4.38 Bioassay to determine complementation by *tmpF* of *mmpF* single mutant of *P. fluorescens* NCIMB 10586. (A) Bioassay plates induced by 0.5 mM of IPTG showing clearing zones (B) Chart of quantitative bioassay (clearing zone measurements). WT, Δ*mmpF* and Δ*mmpF*(pJH10) strains of *P. fluorescens* NCIMB 10586 used as controls while Δ*mmpF*(pMY10) was the test strain. Three replicates were used to collect data. Each bar represents the mean ± standard error of the mean.

4.3 Discussion

Prior work by Fukuda et al. (2011) has shown that the two antibiotics mupirocin and thiomarinol share striking structural similarities, which were also reflected at the genetic level as the products of 27 ORFs of the thiomarinol cluster shared significant identity of amino acid sequence with their *mup* equivalents. The 22 ORFs out of these 27 belonged to the tailoring region genes that are extensively rearranged in these two clusters. By gene knockout and complementation studies, functions of most of the tailoring genes of *mup* cluster have been deduced (Hothersall et al., 2007). On the other hand, very little is known about tailoring functions of *tml* biosynthetic cluster. In view of this, in the present study it was planned to test whether these 22 ORFs from *tml* cluster can complement the functions of their *mup* equivalents in the *mup* system *in vivo*. The idea behind was not just to test convergence and divergence of related proteins in these two distantly related systems but also to gather information about common biosynthetic steps/ intermediates, by way of protein–protein interactions, that might exist in these two systems which produce structures with striking similarities. It has been reported that there is high probability of interaction between products of gene pairs that are conserved (Dandekar et al., 1998). The common biosynthetic steps in these two systems can be used to create hybrid molecules with novel/ enhanced properties by reengineering these two biosynthetic pathways by using genetic components from either (combinatorial biosynthesis) and using the techniques of genetic engineering (Weissman and Leadlay, 2005). Genetic manipulation of tailoring region genes in the biosynthetic clusters has been a great source of getting novel molecules with improved activities (Rix et al., 2002, Olano et al., 2010).

Complementation by nine tailoring region genes could be tested. Results indicated that out of these nine, only two thiomarinol genes *tmlS* and *tmlJ* complemented functions of their *mup* equivalents in the corresponding mutant of *P. fluorescens* NCIMB 10586. While all the

remaining seven *tml* genes namely *tmlO, tmlC, tmlF, tmlK, tmlQ, tacpD* and *tmpF* did not show any complementation when these were expressed, *in trans,* in the corresponding mutant of *mup* system (NCIMB 10586). There could be several reasons behind non-complementation by *tml* genes that may range from common to specific for each of the concerned gene functions. Specific reasons for complementation/non-complementation are discussed separately.

4.3.1 Common reasons behind complementation/ non-complementation by *tml* proteins in *mup* system for each complementation test

In the first place, *tml* genes might not have been efficiently expressed by host machinery because these two organisms have significant differences in their GC content (12.8%). This could have caused poor transcription by transcriptional machinery of mupirocin producer (Gustafsson et al., 2004, Thomas and Nielsen, 2005, Fukuda et al., 2011). Even if these were expressed, the resulting proteins might be non-functional as these might have failed to fold properly in a different milieu in host cell that may have different physiological pH, salt concentration or temperature (Nilsson and Anderson, 1991). This is because the natural habitats of both the organisms are very different (Shiozawa et al., 1993, Thomas et al., 2010). The two organisms have different optimum temperature for growth and nutritional requirements. Moreover, there could be problems with the expression of cloned gene itself as the control of *tac* promoter of pJH10 vector used in this study has been reported to be leaky (de Boer et al., 1983, El-Sayed et al., 2003, Cooper et al., 2005b). As a result, there could be basal level of expression of cloned genes in the absence of any induction, which in itself could have been varied for different genes. Therefore, in case of genes that required induction by IPTG for optimal expression of genes to show any complementation, these could have been induced to different levels with respect to different cloned genes which might have resulted in their varied expressions. This might have affected the results of

complementation where a protein could have been expressed at sub-optimal level than required for complementation (thus showed false non-complementation) in the presence of an amount of IPTG, which was sufficient enough for others to be expressed at levels that resulted in complementation. It may also be possible that the concerned Tml protein modified metabolic intermediates altogether into a different intermediate that was not at all processed further in the metabolic pathway by downstream processing assembly line to result into bioactive molecule. Non-complementation might also be the result of minor differences in the structure of polyketide component of two molecules like 4-OH group is unique to thiomarinols while none of them have epoxide group which is present in the major forms of pseudomonic acids. Also compared to pseudomonic acids, thiomarinols not only have an extra moiety present in them as pyrrothine but also the fatty acid component in these two systems differs by one carbon. Owing to these apparent differences, concerned *tml* proteins might not have been able to process metabolic intermediates without 4-OH group or those with epoxide group, which is because the major pathway of pseudomonic acid production involves metabolic intermediates with epoxide group (Gao et al., 2014). Non-complementation might also be the result of differences in the genetic systems like presence of extra module in *tmpB* gene of *tml* cluster, presence of extra ACPs in the other modules of PKSs and complete absence of equivalent of *macpE* in the thiomarinol system which indicated that molecules might be differently processed in *tml* system than in *mup* system (Thomas et al., 2010, Fukuda et al., 2011, Gao et al., 2014, Murphy et al., 2014). Non-complementation might also be due to the lack of essential specific recognition motifs in *tml* proteins that are required by the components of *mup* system to process biosynthetic intermediates down through *mup* assembly line. It has been shown that each of the two ACPs at the end of module 6 of *mmpA* in *mup* system has unique specific recognition motifs necessary for association with the "β-hydroxymethylglutaryl-CoA synthase" (HCS) cassette

involved in 15 methyl β-branching in pseudomonic acids (Haines et al., 2013b). Non-complementation indicated that the molecules are processed differently; as a result, two pathways have different metabolic intermediates in the biosynthetic pathways, which require specific enzymes to process them in different stepwise order. It is likely because of the presence of pyrrothine moiety in thiomarinols, concerned *tml* protein is involved in the processing of a completely different metabolic intermediate from what its *mup* equivalent processes in the *mup* system. As a result, concerned *tml* protein completely failed to recognize the substrate of its *mup* equivalent so as to process the same. And the similar situation might have caused because of the difference of one carbon in the fatty acid component of these two molecules. It may also be because the fatty acid component in thiomarinols is also in amide linkage with the pyrrothine moiety. Therefore, it is highly likely the *tml* protein encountered a completely different substrate compared to what **its** *mup* equivalent processes in the *mup* system owing to the differences in the biosynthetic steps (Gao et al., 2014, Murphy et al., 2014).

One of the main reasons for non-complementation could be what is indicated from the extensive rearrangement of tailoring genes in the two systems (Fukuda et al., 2011). The rearrangement of genes indicates that the equivalent gene functions might be arranged differently in these two biosynthetic clusters under different operon organizations (Dandekar et al., 1998, Hothersall et al., 2007). Consequently, the equivalent gene functions of these two systems might be under completely different regulatory controls whereby these are coregulated in a different manner along with the other genes of the concerned operon (Dandekar et al., 1998). It has been reported from the study of *trp* operons in seven different prokaryotic genomes that the only gene pair that was conserved among them, was *trpA-trpB* gene pair, while other genes of this operon were present in them differently. For example, in some of these genomes, many member genes of *trp* operon were interrupted by the presence

of stretches of genome sequences or were oriented in opposite direction, while in others the two genes were present as fused with each other and which two genes were fused, were different in different genomes (Dandekar et al., 1998). The differences in the regulation of these two systems become apparent on account of location of their regulatory functions whereby *mup* cluster is found on chromosome while *tml* cluster is entirely located on a plasmid and very little is known about its regulatory aspects (El-Sayed et al., 2001, Fukuda et al., 2011). On the other hand, the singular expression of a gene function out of its natural context might not have yielded a functional product, perhaps because it needed to be transcribed together with its neighbouring genes as part of the organizational requirements of concerned operon (Dandekar et al., 1998, Hothersall et al., 2007). This also holds true for the expression of genes of the *mup* biosynthetic cluster of the mutants, in which *tml* genes were expressed *in trans* (Dandekar et al., 1998, Hothersall et al., 2007). It may also be the situation that in the new context, in a heterologous system, a gene is expressed at a different level (Nishizaki et al., 2007, Lim et al., 2011). As a result, it might be possible that the expressed protein was produced in concentrations and at a rate that was either more or less than the optimum required, at which it could interact with other partner proteins in the biosynthetic pathway to exert its effect in a biochemical reaction, or may be for the same reason it failed to diffuse in proximity to its substrate (Hiroe et al., 2012). This may be the case because it has been found that mupirocin biosynthesis involves several protein complexes and the concerned *tml* protein was not produced at the required level (Cooper et al., 2005b, Hothersall et al., 2007). Similarly, some genes need to be transcribed together as a pair from the same m-RNA to yield a functional product which may be the case because the expressed polypeptides need to fold with each other to get into a conformation which only was functionally active (Dandekar et al., 1998, Hothersall et al., 2007). It has been shown that ACPs are very important in protein-protein interactions in mupirocin biosynthetic

pathway as two of its free standing type II ACPs; MacpB and E could be converted to holo form only by mupirocin PPTase that is MupN (Shields et al., 2010). Further studies with chimeric ACPs indicated their C-terminal end played critical role in protein-protein interactions (Shields et al., 2010, Haines et al., 2013b).

The common reasons behind complementation shown by two of the thiomarinol genes; *tmlS* and *tmlJ* of corresponding *mup* functions might be because of promiscuity of these enzymes (Khersonsky and Tawfik, 2010).

4.3.2 Complementation by TmlO

There could be many specific reasons behind failure of complementation by TmlO. In the first place, failure to complement could be because of different and stringent substrate specificity of this enzyme. MupO has been predicted to be putative cytochrome P450 monoxygenase with which TmlO shared 44% identity of amino acids (Thomas et al., 2010, Fukuda et al., 2011). A number of cytochrome P450 monoxygenases have been implicated in polyketide biosynthetic clusters with their roles in tailoring functions. They are involved in electron transfer reactions. Knocking out these genes in biosynthetic clusters often resulted in the isolation of final biosynthetic product without a hydroxyl or epoxide group. For example, inactivation of cytochrome P450 monooxygenase, in amphotericin B biosynthetic pathway of *S. nodosus,* that inserts 8-OH group in the final molecule resulted in isolation of 8-deoxy amphotericin B molecule. While knocking out of AmphN another cytochrome P_{450} monooxygenase in this pathway resulted in the isolation of amphotericin B derivative having unoxidized methyl group at C-16 position (Carmody et al., 2005). Cytochrome P450 *eryF* of *Saccharopolyspora erythreae* is implicated in the C-6 hydroxylation of 6-deoxyerythronolide B in the biosynthesis of erythromycin (Weber et al., 1991). Similar cytochrome P450 monoxygenase function is also reported from pimaricin biosynthetic cluster of *S. natalensis*

(Mendes et al., 2001). A common feature of cytochrome P450 monoxygenases is that they require proteins to transfer electrons and therefore function as part of a multicomponent system (Urlacher et al., 2004). In *mup* system, it is MupV which is predicted to possess NAD(P)H dependent reductase activity that works in association with MupO (Cooper et al., 2005b). Some atypical reactions are also catalysed by cytochrome P450 enzymes including dehydrations, reductions, isomerizations and dehydrogenations (Mansuy, 1998).

In mupirocin and thiomarinol biosynthesis, the role that is predicted for MupO/TmlO, is hydroxylation of metabolic intermediates of the biosynthetic pathway. MupO is particularly involved in determining oxidation state around pyran ring in pseudomonic acids (Cooper et al., 2005b). It has been shown by Gao et al., 2014, that modifications of the pyran ring occur after the elaboration of 9-hydroxynonanoic acid moiety on monic acid. Acyclic monic acid thioester that is first epoxidised and cyclised, on which elaboration of fatty acid takes place, is the putative substrate that is encountered for oxidation by MupO through MupU and MacpE, which is then dehydrated by MupV and thereafter reduced by the action of MupC and MupF, respectively, to yield PA-A. MupO is predicted to be specifically involved in the reduction at 7^{th} carbon of monic acid, to reduce hydroxyl group to keto and is found to be essential for the production of pseudomonic acid A but not of pseudomonic acid B (Thomas et al., 2010).

It is highly likely that TmlO has a specific requirement for an extra hydroxyl group to be present at the fourth carbon of the metabolic intermediates to process the same, or it could not process an intermediate with 10-11 epoxide group, or it imparted its effect on a metabolic intermediate with altogether different number of carbons compared to MupO etc. Moreover, it has been proposed in the biosynthesis of mupirocin that several tailoring proteins work together as part of auxiliary complex to make multiprotein complex which processes the biosynthesis of mupirocin as an integrated structure (Hothersall et al., 2007). There is

another probability that TmlO might have failed to become part of this multiprotein complex. It has been proposed that these modifying reactions on biosynthetic intermediates take place when these are still tethered on mupirocin acyl carrier proteins (MacpE) (Thomas et al., 2010). Since TmlO could not become part of such biosynthetic complex of auxiliary proteins perhaps because of topographical reasons or owing to the fact that *tml* cluster lacks *macpE,* TmlO lacked certain key residues (that were present in MupO) that were necessary for it to collaborate with the biosynthetic complex of *mup* proteins (Fukuda et al., 2011). This way it could not access the substrate held by the ACP in the multiprotein complex to impart its effect**.** Another probability is that TmlO might have failed to collaborate with MupV, with which it was required to work because of being an oxygenase it has the requirement to collaborate with proteins that can transfer electrons.

Reason behind reduced zones of clearance when IPTG was added to the medium could be interference in pseudomonic acids production by the TmlO protein, which might be inhibiting its production, probably by binding to any of the proteins involved in the biosynthetic pathway. Otherwise, in *mup* system MupO has been reported to complement when it was expressed *in trans* using the same vector and complementation was better when IPTG was included in growth medium (Cooper et al., 2005b).

4.3.3 Complementation by TmlC

MupC enzyme of mupirocin biosynthetic cluster is predicted to be a dienoyl-CoA reductase (an oxido-reducatse) (Hothersall et al., 2007, Gao et al., 2014). The TmlC, product of *tmlC* ORF in the tailoring region of thimarinol cluster is also predicted to be an oxidoreductase with which MupC shared 56% amino acid sequence identity (Fukuda et al., 2011, Thomas et al., 2010). *In vivo* complementation of Δ*mupC* mutant by *in trans* expression of *mupC* ORF resulted into the restoration of pseudomonic acid A production from more than 10-fold

reduction that occurred in its production by the Δ*mupC* mutant (Cooper et al., 2005b). As a result of gene knockout and complementation studies of several other ORFs of the tailoring region of *mup* cluster, it has been predicted that MupC acts as enone reductase and is involved in the reduction of enol ketones metabolic intermediates in the biosynthetic pathway that resulted in the production of pseudomonic acid-A (Hothersall et al., 2007). Functional activity like that of *mupC* is also reported from *Clostridium sp.* which showed high homology with dienoyl-CoA reductases (Rohdich et al., 2001).

As per the putative biosynthetic pathway that has been suggested, MupC acts after the action of MupV- an oxidoreductase, that reduces the hydroxyl group present at C-8 of the metabolic intermediate in the pathway into a double bond, and this metabolic intermediate becomes substrate for the action of MupC (Thomas et al., 2010). All the tailoring modifications by *mupO, V, C* and *F* are done while the metabolic intermediate is tethered at *macpE*/15 (Thomas et al., 2010). It has been predicted that all these enzymes function together as part of complex of auxillary enzymes that function as multiprotein complex to modify the substrate attached to it (Hothersall et al., 2007). TmlC, like TmlO might have failed to become part of this multiprotein complex of several enzymes acting together. Perhaps this is because of some key residues which other *mup* proteins might be having, by virtue of which these could become part of such complex to interact with MacpE, were missing from TmlC. Moreover, for TmlC which is a putative NADH: flavin oxidoreductase, either it failed to recognize the substrate-metabolic intermediate- of MupC which is putative dienoyl-CoA reductase in the *mup* pathway or modified the substrate in a way such that it blocked the pathway i.e. created metabolite that is no more substrate for the action of subsequent enzyme(s) in *mup* assembly line and this way diverted the pathway towards a

product that no longer had biological activity. While other general reasons explained under section 4.3.1 might also be responsible for non-complementation by TmlC.

4.3.4 Complementation by TmlF

It has been shown that MupF complemented the Δ*mupF* mutant strain when expressed *in trans,* cloned in the same IncQ vector which was used for complementation study by *in trans* expression of *tmlF* (Hothersall et al., 2007). Production of PA-A was restored to wild type level on *in trans* expression of *mupF* from less than 2% production by Δ*mupF* mutant (Cooper et al., 2005b, Hothersall et al., 2007). By gene knockout and complementation studies MupF is predicted to be specifically involved in the reduction of keto group present at the 7th carbon of the metabolic intermediate in the biosynthesis of pseudomonic acid-A. In the biosynthetic pathway that is proposed, MupF is predicted to act immediately after the action of MupC (Thomas et al., 2010, Gao et al., 2014). By *in silico* studies MupF is predicted to be ketoreductase (Hothersall et al., 2007). TmlF, the product of *tmlF* ORF in the tailoring region of thiomarinol cluster shared 34% of identity of amino acids with MupF and is also predicted to be a ketoreductase (Hothersall et al., 2007, Fukuda et al., 2011). Possible reasons for non-complementation by TmlF might be similar to those for TmlO and TmlC, because in *mup* system all these genes have been proposed to function together as part of multiprotein complex (Hothersall et al., 2007). As was expected in case of TmlO and TmlC, TmlF also lacked certain residues which MupF possessed and which were critical for TmlF to become part of this multiprotein complex of tailoring genes tethered on Macp**E** of *mup* system (Figure 1.26). It might be possible that owing to its high specificity, TmlF completely failed to recognize substrate, the metabolic intermediate in *mup* biosynthetic pathway, that did not have 4-OH group as pointed out in section 4.3.1. Complete absence of an *macpE*

homologue in thiomarinols clearly indicated that TmlF processed a completely different biosynthetic intermediate than what MupF processed.

4.3.5 Complementation by TmlJ and TmlK

TmlJ complemented the functions in Δ*mupJ* mutant strain and restored the production of PA-A while TmlK did not show complementation in Δ*mupK* mutant and did not restore pseudomonic acid production. Both *mupJ* and *mupK* mutants show drastically reduced production of pseudomonic acids (Hothersall et al., 2007). These findings **are** contrary to the fact that either *mupJ* or *mupK* completely failed to complement when these were expressed singularly, *in trans,* cloned in the same expression vector, which was used for *tmlJ* or *tmlK* expression under the same growth medium and other conditions in the corresponding *mup* single mutants. Thas was inspite of induction of their expression by IPTG (Hothersall et al., 2007).

Complementation by *mupJ or mupK* could only be achieved in Δ*mupJ* mutant when both these ORFs were expressed together *in trans* under a single promoter and transcribed as single *m*-RNA. While insignificant complementation was reported to be achieved when same construct was expressed *in trans* in Δ*mupK* mutant**.** Complementation, close to wild type levels, could only be achieved when this construct was expressed *in trans* in the double mutant Δ*mupJK* (Hothersall et al., 2007). The requirement of expression of both these ORFs together to achieve complementation was explained on account of the organization of the concerned operon and as a requirement for the correct folding of these two proteins so as to make them functional (Hothersall et al., 2007). Decrease in complementation in Δ*mupJ* mutant by *mupJ* and no complementation in Δ*mupK* mutant by *mupK* had been explained, because the remaining ORF of either *mupJ* or *mupK,* depending on the case, was expressed as

non-functional polypeptide by the *mup* cluster which interfered with the MupJ-MupK protein that was expressed *in trans* from the vector (Hothersall et al., 2007).

In silico studies have shown that *mupJ* and *mupK* are two discrete ORFs in the tailoring region of the mupirocin gene cluster with other several genes. Biochemically, MupJ and MupK are predicted to be enoyl-CoA hydratases of the type belonging to crotonase superfamilies that are involved in β-oxidation of fatty acids (Hothersall et al., 2007). Bioinformatics search have revealed many genes related to *mupJ* and *mupK,* the products of which have similar or related biochemical function as enoyl-CoA hydratases. One such pair of homologous proteins is PksH and PksI of *B. subtilis* subsp. *subtilis* str.168 which is present within the polyketide biosynthesis cluster and this pair shows highest homology with these *mup* proteins (Albertini et al., 1995). Though exact roles of these two enoyl-CoA hydratases in mupirocin biosynthetic pathway are not deduced till date but bioinformatics analysis revealed that MupK homologues, PksH (60.4% identity), Tay (57.8%) and CurF ECH_2 (31.6%) in bacillaene, myxovirescin and curacin biosynthetic clusters, respectively, all catalyse decarboxylation *in vitro* (Hothersall et al., 2007). While MupJ homologues in the same clusters, PksI (49.6% identity), Tax (51.1%) and CurE ECH_1 (37.6%) have been shown to catalyse dehydration (Gu et al., 2006, Calderone et al., 2006, Calderone et al., 2007). In mupirocin biosynthetic pathway both these enzymes are predicted to act along with the products of other tailoring region genes *mupG, mupH* and *macpC* in the incorporation of C-15 methyl group (Hothersall et al., 2007) at C-3 of the mupirocin polyketide moiety (Hothersall et al., 2007, Thomas et al., 2010, Haines et al., 2013a). In the putative biosynthetic pathway that is proposed for β-branching, MupJ acts before MupK as dehydratase and subsequently MupK decarboxylates the product (Figure 4.39) (Hothersall et al., 2007).

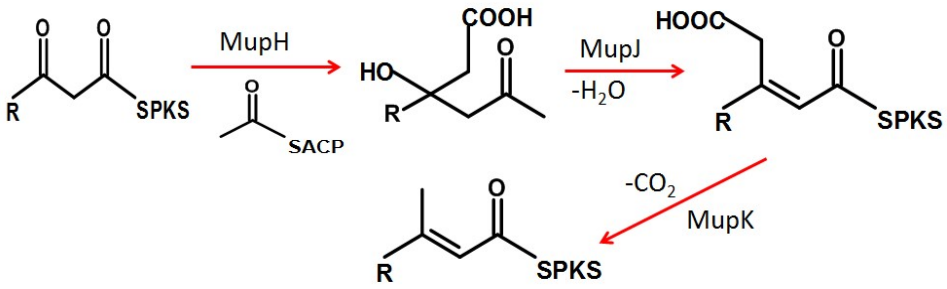


Figure 4.39 Scheme predicted for incorporation of C-15 methyl group by MupJ and MupK acting along with other proteins(Adapted from Hothersall et al., 2007).

TmlJ that showed 49% amino acid sequence identity with MupJ, is also predicted to be enoyl-CoA hydratase while homologue of MupK is not present in the *tml* cluster. Instead, the N-terminal end of TmpE, which is a multifunctional protein, aligned with MupK sequence and is predicted to have enoyl-CoA hydrates activity. These two proteins share 40% identity of amino acid sequence between them (Fukuda et al., 2011).

Restoration of PA-A production by *tmlJ* in the single mutant Δ*mupJ* clearly indicated that its product could actively fold itself on its own so as to be functional in the *mup* system. As a result, it could process the concerned metabolic intermediate in the *mup* biosynthetic pathway, acting along with all the member proteins encoded by HCS cassette, such that it was processed by the downstream functions of the assembly line. TmlJ dehydrated the concerned metabolic intermediate in the proposed pathway before decarboxylation by MupK. It has been shown in the *mup* system that each of the ACPs at the end of sixth module of *mmpA* has unique recognition motif specific for their interaction with the HCS cassette proteins that are involved in β-branching (Haines et al., 2013a). It clearly appears that TmlJ could collaborate with these unique motifs of the *mup* assembly line while TmlK might have failed to collaborate. Though the complementation by *tmlJ* was dependent on induction by IPTG indicating that higher than basal level of enzymatic protein was required to bring about concerned reaction. Since the complementation was achieved to wild type levels, it indicated

non-interference in its action by MupK which in case of complementation by *mupJ*, interfered with the expression/action of MupJ thus lowered the degree of complementation.

Primary reasons behind non-complementation of the single Δ*mupK* mutant by TmlK protein could be because of its failure to fold so as to become functional enzyme. Which is because in *tml* system this hydratase activity exists only as part of a multifunctional protein TmpE that includes ketosynthase activity as well. Therefore, in the first place, expression of a part of this multifunctional protein, that is predicted to have enoyl-CoA hydratase activity might not have resulted into a functional protein having activity when expressed alone cloned in a vector. Which is because it might have required its cognate partner sequence of the *tml* multifunctional protein, to which it originally belonged, in order to fold properly so as to retain its enzymatic activity. While there could be any other reasons behind non-complementation that are discussed above under section 4.3.1. Presence of this second hydratase activity in the *tml* cluster fused with multifunctional protein and the requirement of *mup* system to express both the hydratases (MupJ and MupK) together to achieve complementation, clearly indicated differences in the genetic organization of these two systems, which might result into differences in the biosynthetic steps leading to β-branching in these two systems. *tml* system also has HCS cassette similar to the one found in *mup* system with each of its member proteins which shares significant amino acid sequence identity with the corresponding members of *mup* system.

4.3.6 Complementation by TmlS

Results of TmlS complementation indicated that it complemented the functions of MupS upon expression of *tmlS* ORF, *in trans*, in the Δ*mupS* mutant and partially restored the production of PA-A. *In silico* studies have predicted that proteins like TmlS and MupS, which share amino acid sequence identity of 62%, act together with the products of other

genes as a group and are implicated in furnishing starter/primer unit for the biosynthesis of fatty acids in those biosynthetic systems. This is supported by the genetic evidence because genes for all these proteins that act in a group are found located together in the biosynthetic gene cluster. In mupirocin cluster, *mupS* (3-oxo acyl carrier protein reductase) is located together with *mmpF* (decarboxylase and ketosynthase), *macpD* (acyl carrier protein) and *mupQ* (acyl-CoA synthase). On the basis of bioinformatics studies, product of genes like *mupS*, *mupQ* and *macpD* are predicted to act together and are implicated in furnishing the starter/primer unit for the biosynthesis of the fatty acid component of pseudomonic acids (9-HN) by *mmpB* with the help of associated ACPs (Thomas et al., 2010). These aspects of concerned *mup* proteins are also supported by the existence of genes for a similar group of proteins in other similar *trans* AT systems like difficidin polyketide system of *Bacillus amyloliquifaciens* in which these are found located together as in *mup* cluster, indicating that these groups of proteins act together. In difficidin biosynthetic cluster *diffE, diffD and diffC* genes are located together which code for reductase, acyl-CoA synthase and acyl carrier protein, respectively. All of these proteins are predicted to act together with the product of another gene *diffB* (a kinase) located with them to dehydrate 3-hydroxypropionate to derive acryl thiol ester which acts as priming unit for the chain elongation in difficidin biosynthesis (Chen et al., 2006). Similarly, proteins encoded by *tml* cluster namely TmlS, TmlQ, TmpF and TacpD each of which share significant amino acid sequence identity with the corresponding mup proteins, are also found located together in *tml* cluster (Fukuda et al., 2011). As discussed in section 4.3.1, it has been reported that conservation of gene pairs/clustering across different systems suggested that the concerned proteins are interacting (Dandekar et al., 1998). Based on these evidences, in *tml* biosynthetic system this group of *tml* proteins is also predicted to act together in furnishing starter/primer unit for the biosynthesis of the fatty acid component of thiomarinols- the 8-hydroxyoctanoic acid by

TmpB and associated ACPs. Apart from other evidences, the mutagenesis studies in which knockout or inactivation of any of these genes namely MupS, MupQ, MmpF and MacpD proteins resulted into similar phenotype of producing mupiric acid and mupirocin H with complete loss in the production of pseudomonic acids, also support that this group of proteins functions together (Hothersall et al., 2007, Wu et al., 2008).

It has been reported that knockout of *mupS* could be complemented by *in trans* expression of *mupS* ORF cloned in an IncQ expression vector (Cooper et al., 2005b). Therefore, the partial complementation of *mupS* function by TmlS in the Δ*mupS* mutant under expression from same IncQ vector under same growth conditions indicated that these two enzymes shared substrate specificities of certain substrates found in both these biosynthetic pathways. Complementation indicated that up to the step where substrates could be processed by both these enzymes, either two biosynthetic pathways being similar or that the similar substrate existed in these pathways which might have other totally different biosynthetic steps. Since TmlS, also a 3-oxo acyl carrier protein reductase, could only partially restore production of pseudomonic acid A, this indicated that it could only partially catalyse the reaction in the *mup* biosynthetic pathway to modify the substrate of MupS such that it could be loaded onto concerned acyl carrier protein (MacpD) (Figure 1.25).

While there could be other reasons behind the functional complementation shown by TmlS particularly in view of the fact that complementation was only partial indicating that it might have been because of the substrate promiscuity shown by TmlS (Khersonsky and Tawfik, 2010) as pointed out in section 4.7.1.

4.3.7 Complementation by TmlQ

In view of the reported finding that *mupQ* could complement when expressed *in trans,* cloned in the same vector under same growth conditions which were used for the expression of *tmlQ*

in the Δ*mupQ* mutant of NCIMB 10586, it could be concluded that TmlQ did not complement. Despite sharing significant (46%) amino acid sequence identity with MupQ, TmlQ completely failed to act upon the substrate of MupQ in the *mup* system to produce its CoA (Cooper et al., 2005b, Thomas et al., 2010). Non-complementation, therefore, indicated that there has been divergence in the function of these two proteins during evolution. Non-complementation specifically indicated that each of these proteins have high specificity for their respective substrates (concerned metabolic intermediates) in the respective biosynthetic pathways, which probably used different starter units for priming the biosynthesis of their fatty acid moieties on account of what is discussed in section 4.3.7 above. While non-complementation might also be due to other reasons explained above under section 4.3.1.

4.3.8 Complementation by TacpD

Results indicated that TacpD did not complement the functions of MacpD in Δ*macpD* mutants. *In trans* expression of *macpD* cloned in the same expression vector (IncQ), as was used for *tacpD* expression, is already reported to have complemented Δ*macpD* mutants (Hothersall et al., 2007). In this view, non-complementation by TacpD clearly indicated that it failed to recognize substrates that were processed by MacpD protein of the *mup* biosynthetic pathway and that TacpD/MacpD had very high specificity for their specific substrates. This was despite both these proteins are putative acyl carrier proteins that shared 46% amino acid sequence identity between them which is significant. Non-complementation, therefore, indicated that there is divergence in the functions of these proteins during evolution and in view of what is discussed above under section 4.3.7 that different priming unit is utilized by *mup* system for the biosynthesis of its fatty acid component (9-HN moiety) from that of *tml* system that has 8-HO (Thomas et al., 2010, Fukuda et al., 2011). Other reasons behind non-complementation could be same as discussed above under section 4.3.1.

4.3.9 Complementation by TmpF

TmpF did not complement mutant of *mmpF* and did not restore the production of PA-A. It has been shown by mutating active site of *mmpF* (C183A) that this gene is necessary for the production of mupirocin (Hothersall et al., 2007). This mutation in *mmpF* could be complemented by the *in trans* expression of the *mmpF* ORF cloned in IncQ plasmid pJH10 that was used in this study (Hothersall et al., 2007). Since deletion of any of the genes of the *mup* cluster so far has completely failed to produce a mutant that either produced monic acid or 9-hydroxynonanoic, this has indicated that biosynthesis of pseudomonic acids is a single assembly line integrated process (Thomas et al., 2010). This is also supported by the findings of refeeding experiments done by Gao et al., 2014. Thus, MmpF has a definite role to play in this integrated biosynthesis of pseudomonic acids. TmpF and MmpF both are predicted to be ketosynthases which share amino acid sequence identity of 43% which is significant. Though specific role could still not be proposed for MmpF in the *mup* biosynthetic pathway yet based on its proximity to *mupS, mupQ* and *macpD* in the *mup* cluster that are predicted to play role in furnishing primer unit for the biosynthesis of fatty acid moiety (9-HN) and, on the basis of *in silico* studies pointed out above under section 4.3.6., MmpF is also predicted to be involved in furnishing this starter unit in the *mup* system. The *mmpF* might be present with these group of genes like as *diffB* is present in difficidin gene cluster to play a role by collaborating with the products of other genes of this cluster in furnishing the starter unit (see section 4.3.6). Non-complementation of MmpF functions by TmpF, in the first place clearly indicated that these two proteins despite sharing significant amino acid sequence identity have functionally diverged during the course of evolution.

CHAPTER 5
Investigations on cross-complementation by a group of homologous accessory genes in the biosynthesis of antibiotics mupirocin and thiomarinol

5.1 Introduction

As described in Chapter 4, some *tml* genes were found to complement to their corresponding *mup* knockout mutants but this tended to be the exception rather than the rule. One reason for lack of complementation could be the need to interact with partner proteins and it may be that at least some of the protein-protein interactions have diverged sufficiently between the *mup* and *tml* systems that function complexes are no longer able to form. The first example of likely protein-protein interaction in the *mup* system is the *mupD-E* pair where it has been shown that these two genes need to be expressed together to complement a mutation in either gene. Also an active site point mutation in *mupD* had no effect on PA-A production while an in-frame deletion abolished production, suggesting that MupD now acted as chaperone for MupE (Hothersall et al., 2007).

Similarly, as described in Chapter 4 (section 4.3.5), complementation of MupJ and MupK proteins could only be achieved to wild type levels when these were expressed together as part of single m-RNA cloned in the pJH10 expression vector in the double mutant Δ*mupJ* and Δ*mupK of Pseudomonas fluorecsens* NCIMB10586. These two proteins are implicated in 15 methyl β-branching in mupirocin biosynthesis acting together with the products of several other genes as part of "β-hydroxymethylglutarylCo-A synthase" known as HCS cassette (Hothersall et al., 2007).

In both the above examples, expression of single genes of the pair when expressed separately from each other resulted in the production of non-functional elements which in turn interfered with the expression of other partners. As a result, complementation by either member of the pair in both the cases was either reduced or blocked (Hothersall et al., 2007).

The other aspect of complementing group of related genes whose products are predicted to work together becomes clear by the cross-complementation studies done about group of genes that are proposed to be involved in pyran ring formation in these two biosynthetic systems (Hothersall, unpublished studies, 2014). These studies gives broader understanding of protein-protein interactions taking place between groups of proteins that are predicted to work together in the mupirocin and thiomarinol biosynthetic system. It was shown that cross-complementation by *tml* genes; TmlW and TmlT was close to wild type levels when both these ORFs were expressed together, *in trans,* as a single mRNA cloned in an expression vector in the double mutant Δ*mupT* and Δ*mupW* of *Pseudomonas fluorecsens* NCIMB10586 compared to when these were expressed singularly in the corresponding single mutant. The *mupW* and *mupT* encode a putative dioxygenase and related ferredoxin dioxygenase, respectively, and both have been implicated in the closure of tetrahydropyran ring (THP) in pseudomonic acid molecules. Mutation of either of these resulted in the isolation of similar compounds named as pseudomonic acid W, which was characterized by having an open pyran ring (Cooper et al., 2005a, Gao et al., 2014). This indicated that products of both these genes were required for oxidative closure of THP. TmlW and TmlT are also predicted to be putative dioxygenase and associated ferredoxin dioxygenase like their *mup* counter parts MupW and MupT, with which they share significant amino acid sequence identity of 56% and 39%, respectively (Fukuda et al., 2011). Similarly, tetrahydropyran ring (THP) as it is found in pseudomonic acids, is also present in thiomarinols with minor differences in side groups (Thomas et al., 2010).

Considering other groups of tailoring genes, whose products are predicted to work together and whose analysis would be useful, led to the group of genes proposed to be involved in production of the starter unit for 9-HN namely *mupQ, macpD, mupS* and *mmpF* in mupirocin biosynthetic cluster along with those, that are either involved or predicted to be

involved in furnishing priming unit for 8-HO biosynthesis namely *tmlS, tmlQ, tacpD* and *tmpF* in thiomarinol biosynthetic cluster (Feline et al., 1977, Piel, 2010, Murphy et al., 2014, Helfrich and Piel, 2016). *In silico* studies described in Chapter 4 (section 4.3.6), have predicted that proteins from such group of genes which existed together in biosynthetic clusters, were also found to work together. Studies about conserved gene-pairs/clustering, that are conserved as part of operon architecture across different prokaryotic systems, also suggested that protein products of such gene pairs interact physically (Dandekar et al., 1998). Considering these facts, cross-complementation study of this group of genes between both the biosynthetic clusters was planned to study interaction of proteins as a group between the two systems.

The results of cross-complementation studies for singular genes of this group showed that TmlS could interact with the other *mup* proteins of this group such that it partially restored the production of pseudomonic acid-A in the corresponding mutant strain *Pseudomonas fluorecsens* NCIMB 10586Δ*mup*S (see section 4.3.6). TmlS and MupS are both predicted to be reductases that share 62% amino acid sequence identity which is very significant (Figure 4.27). Partial complementation of functions of MupS by TmlS indicated that to some extent there has been functional conservation of concerned protein-protein interactions in the two proteins. Though the results of singular cross-complementation of the remaining three genes of this group namely *tmpF, tmlQ* and *tacpD* have been negative despite product of each of them also shared significant amino acid sequence identity with the corresponding *mup* protein (Chapter 4 section 4.3.7-10) (Figures 4.24, 4.32 and 4.35). In this study, above explained approach of complementation of group of genes was extended to find whether extra complementation could be achieved by expressing the four thiomarinol genes (*tacpD, tmlS, tmlQ* and *tmpF*) together in the *Pseudomonas fluorecsens* NCIMB10586 with its all the four corresponding gene functions knocked out (*macpD, mupS, mupQ* and *mmpF).*

5.2 Results

5.2.1 Construction of *P. fluorescens* NCIMB 10586 mutant strain with defined chromosomal deletion of *macpD, mupS, mupQ* and *mmpF* (quadruple mutant)

To create in-frame chromosomal deletion of four genes, strategy mentioned in Chapter 2 (section 2.4.14) was followed (Figure 5.2). Accordingly, sequences of around 500 base pairs were identified on the chromosome in the *mup* cluster that overlapped with the ends of the region that contained *mupQ* and *mmpF*. Sets of primers given in Table 2.11 were designed to amplify these arms. An upstream arm of 547 bp was amplified using primers QDA1F and QDA1R that incorporated *EcoR*I and *Xba*I sites, respectively, to facilitate its cloning into the suicide vector pAKE604 and ligation with the downstream arm. Similarly, downstream arm of size 504 bp was amplified using primers QDA2F and QDA2R that incorporated restriction sites *Xba*I and *Bam*HI, respectively. QD in the name of primers indicated that they were developed for the deletion of a block of four concerned genes from the chromosome of *Pseudomonas fluorescens* NCIMB 10586 while numbers 1 and 2 indicated that these were meant to amplify upstream and downstream arms, respectively**.** The letter F and R in primer names represented forward and reverse primers, respectively. The amplified arms were cloned into pGEM-T Easy vector after A-tailing as per method mentioned in Chapter 2 (section 2.4.8) to yield vectors pMY19 and pMY20 in which upstream and downstream arms were cloned, respectively (Table 2.7). After sequencing using universal primers, to confirm that the sequence of these arms was correct, these fragments were cloned into suicide vector pAKE604 using restriction sites *EcoR*I*, Xba*I and *Bam*HI by three way ligation as a fragment of 1051 bp to yield suicide vector pMY21 (Table 2.7, Figure 5.1).

The constructed suicide vector pMY21 was transformed into *E. coli* S17-1 cells and purified transformants were then used to mobilize it into wild type *Pseudomonas fluorescens* NCIMB 10586 strain as per methods described in Chapter 2 (section 2.4.13). In this way homologous recombination was promoted to take place that resulted in the insertion of suicide vector into the chromosome of wild type *P. fluorescens* NCIMB 10586. The resulting putative integrants were purified by restreaking to single colonies to remove any bacteria that had not received the suicide plasmid. Purified strains were then grown without selection and then spread on L agar sucrose plates to isolate sucrose resistant bacteria and colonies were tested for the loss of kanamycin resistance. Kanamycin sensitive clones were screened for deletion of the target region by PCR reaction in which plasmid pMY21 was used as positive control while wild type *Pseudomonas fluorescens* NCIMB 10586 strain was used as negative control (Figure 5.3). Out of 22 colonies tested this way for the desired mutation, twelve turned out to be mutants.

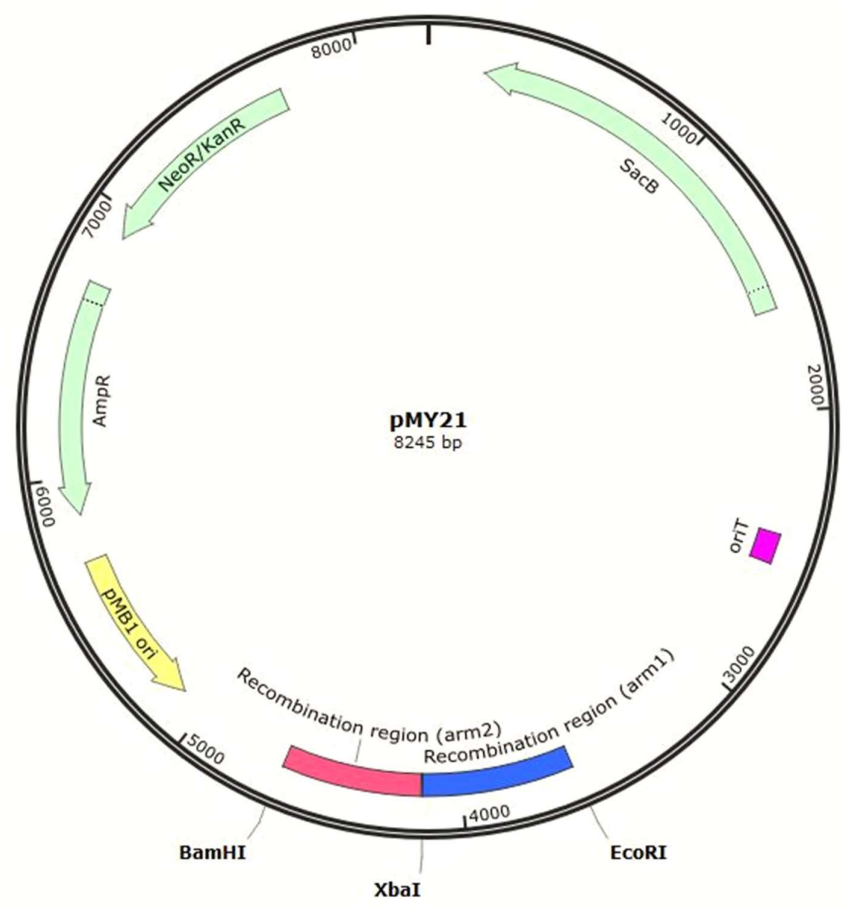


Figure 5.1 Map of suicide vector pMY21 constructed from pAKE604 for the deletion of a block of four *mup* genes; *mupS, mupQ, macpD* and *mmpF* in the chromosome of NCIMB 10586.Created using SnapGene®.

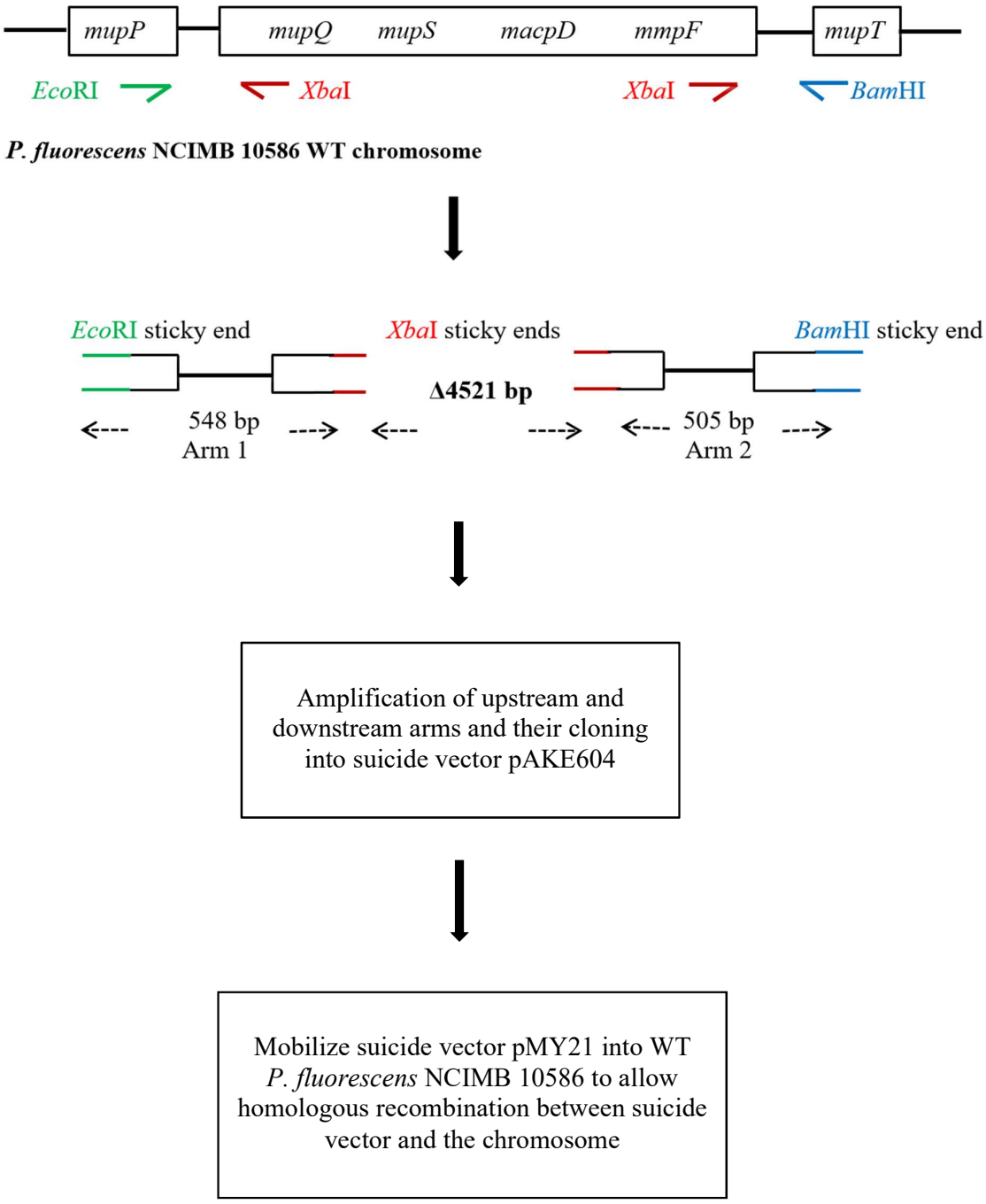


Figure 5.2 Construction for the deletion of quadruple genes *mupQ, mupS, macpD* and *mmpF* in the chromosome of *P. fluorescens* NCIMB 10586.

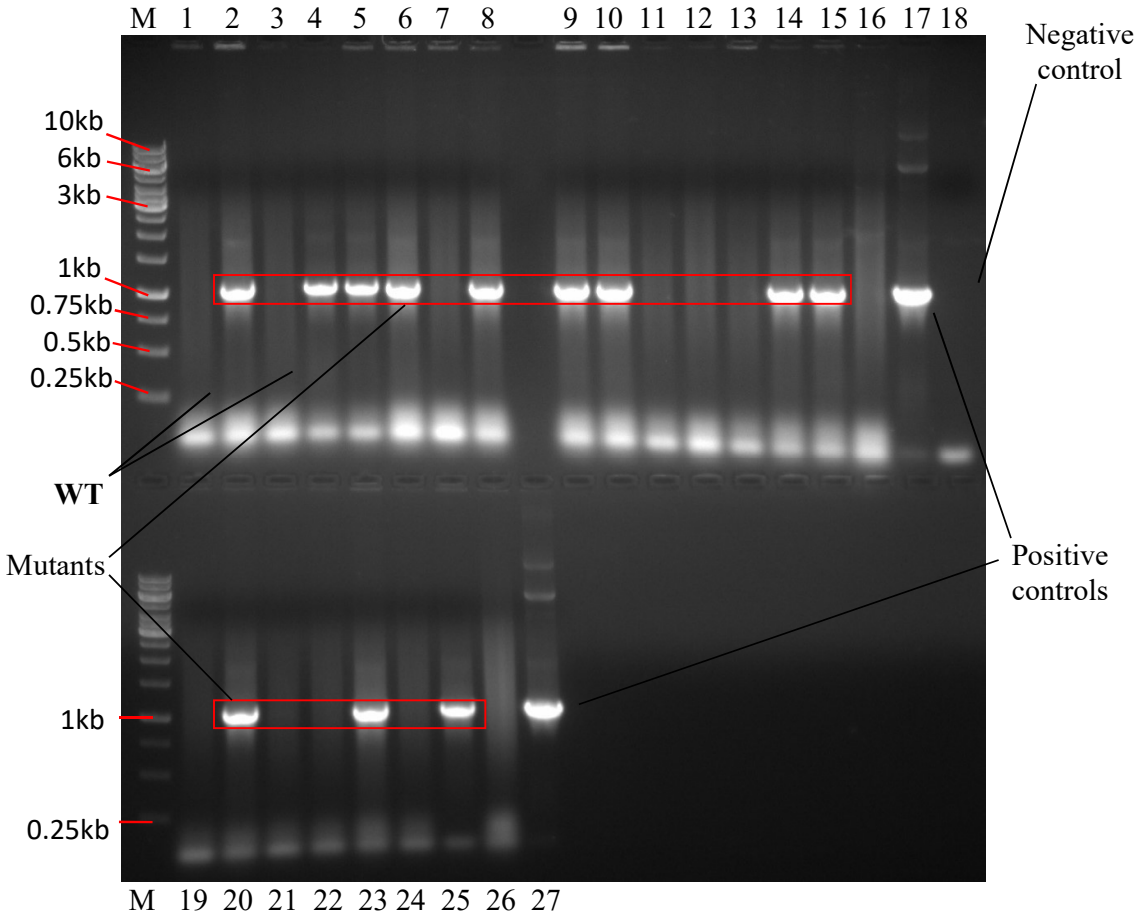


Figure 5.3 The result of PCR to check potential chromosomal mutants of NCIMB 10586 for the deletion of fragment containing *mupQ, mupS, macpD* and *mmpF*. Top and bottom lanes from left: M, 1 kb marker ladder; lane 1 and 3, WT NCIMB 10586; lane 17 and 27, positive control; lane 18, negative control. Out of total 22 clones analysed 12 confirmed to be mutants; lane 2, 4, 5, 6, 8, 9, 10, 14, 15, 20, 23 and 25.

5.2.2 Construction of expression vector carrying *tmlQ, tmlS, tacpD* and *tmpF* (quadruple expression vector)

For cross-complementation analysis of group of thiomarinol genes in the constructed mutant strain of *Pseudomonas fluorescens* NCIMB 10586Δ*mmpF*Δ*mupS*Δ*mupQ*Δ*macpD* (quadruple mutant), an expression plasmid was constructed to express all the four corresponding genes of thiomarinol cluster (*tacpD, tmlQ, tmlS and tmpF)* cloned under *tac* promoter in broad host range vector pJH10 (Figure 2.4)(El-Sayed et al., 2003). To create plasmid expressing four thiomarinol genes, three of these namely *tacpD, tmlQ* and *tmlS* were PCR amplified as a fragment of 2.2 kb. Primers were designed to amplify this fragment that incorporated restriction sites *Xba*I and *Sac*I (Table 2.11). The fragment was amplified using Velocity^{TM} DNA polymerase as described in Chapter 2 (section 2.3.2). A good yield of product of expected size was obtained at annealing temperature of 54 ^{o}C. The ORF for the fourth gene of this group, *tmpF* was used from the vector pMY10 that was constructed for the previous study on singular complementation (Chapter 4, Table 2.6).

PCR amplified fragment was A-tailed and cloned into pGEM-T Easy vector to give 5.21 kb plasmid, pMY22, which was subjected for sequencing of the cloned fragment. It is worth mentioning that in view of the large size of the cloned fragment (2.2 kb), internal primers have to be designed and used to confirm integrity of its sequence (Table 2.11). Finally, cloning of PCR fragment from pMY22 clones that showed no mismatch, was attempted in pMY10 vector at sites *Xba*I and *Sac*I using restriction digestion/ligation as described in Chapter 2 (sections 2.4.5 and 2.4.9). Plasmid pMY10 of 15.99 kb which was derived from pJH10 vector, had the *tmpF* ORF cloned into it as an *EcoR*I-*Xba*I fragment under the control of *tac* promoter (Table 2.6). After the failure of repeated attempts made to clone fragment from pMY22 carrying three *tml* genes into pMY10 by restriction digestion/ligation as double digest of *Xba*I and *Sac*I, an alternative strategy that depended on the use of single restriction

site for cloning, was used. As per this strategy, plasmids pMY22 and pMY10 were digested with *Xba*I and ligated to give cointegrate pMY23 that had these two plasmids ligated through *Xba*I site. This cointegrate was transformed into *E. coli* C2110 (*polA1*-) competent cells. Transformants that showed resistance to ampicillin were selected confirming success of ligation and existence of cointegrate because resistance to ampicillin could only be conferred by an intact cointegrate, which was because pMY22 itself could not replicate in this strain of *E. coli*, and expression of its ORF that conferred resistance to ampicillin was driven by the *tac* promoter of pMY10, which was possible only if both these plasmids were present as cointegrate. Clones of transformant colonies were analysed to get the clone of cointegrate in which the two plasmids were joined together in the desired orientation. That was the clone in which the fragment carrying ORFs of three *tml* genes in pMY22 was ligated at the end of the *tmpF* ORF of pMY10 vector through the *Xba*I site. Out of eight clones analysed by *Eco*RI restriction digestion, six turned out to have the desired orientation (Figures 5.5 and 5.6).

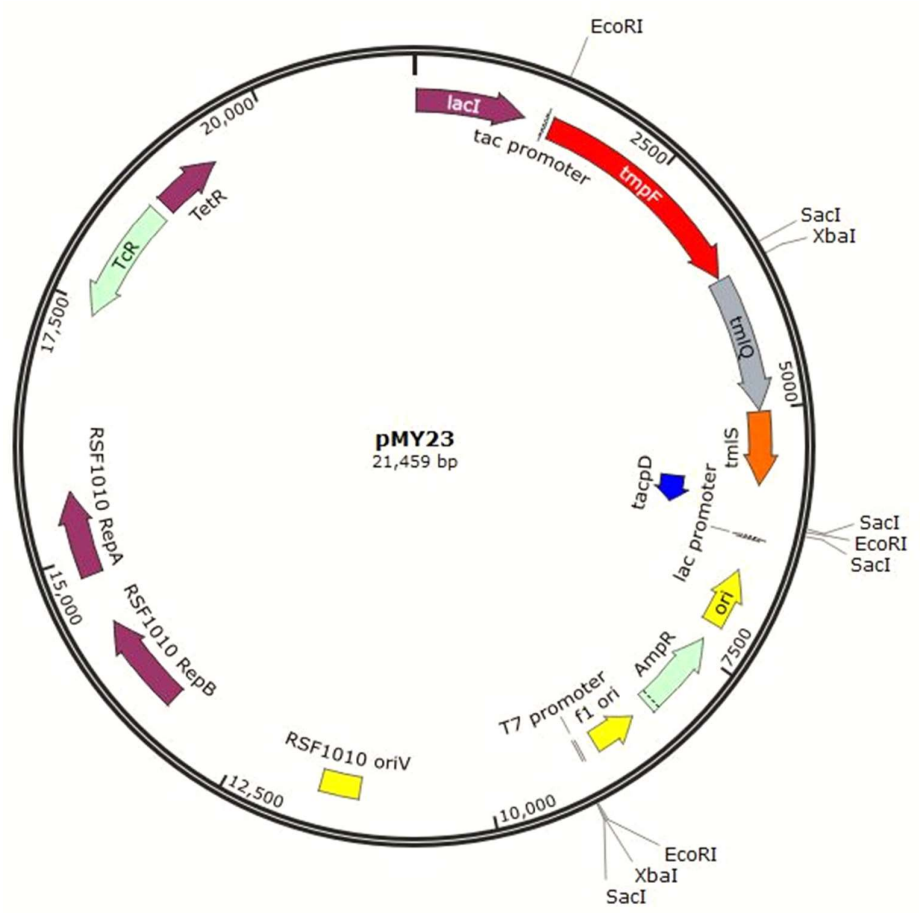


Figure 5.4 Map of expression vector pMY23 used for *in trans* expression of Tml proteins; TmlS, TmlQ, TacpD and TmpF in NCIMB 10586.Plasmid pMY23 constructed as a cointegrate by ligating pMY10 (another expression vector) with pMY22 vector through ***Xba***I restriction sites. C**reated** using SnapGene®.

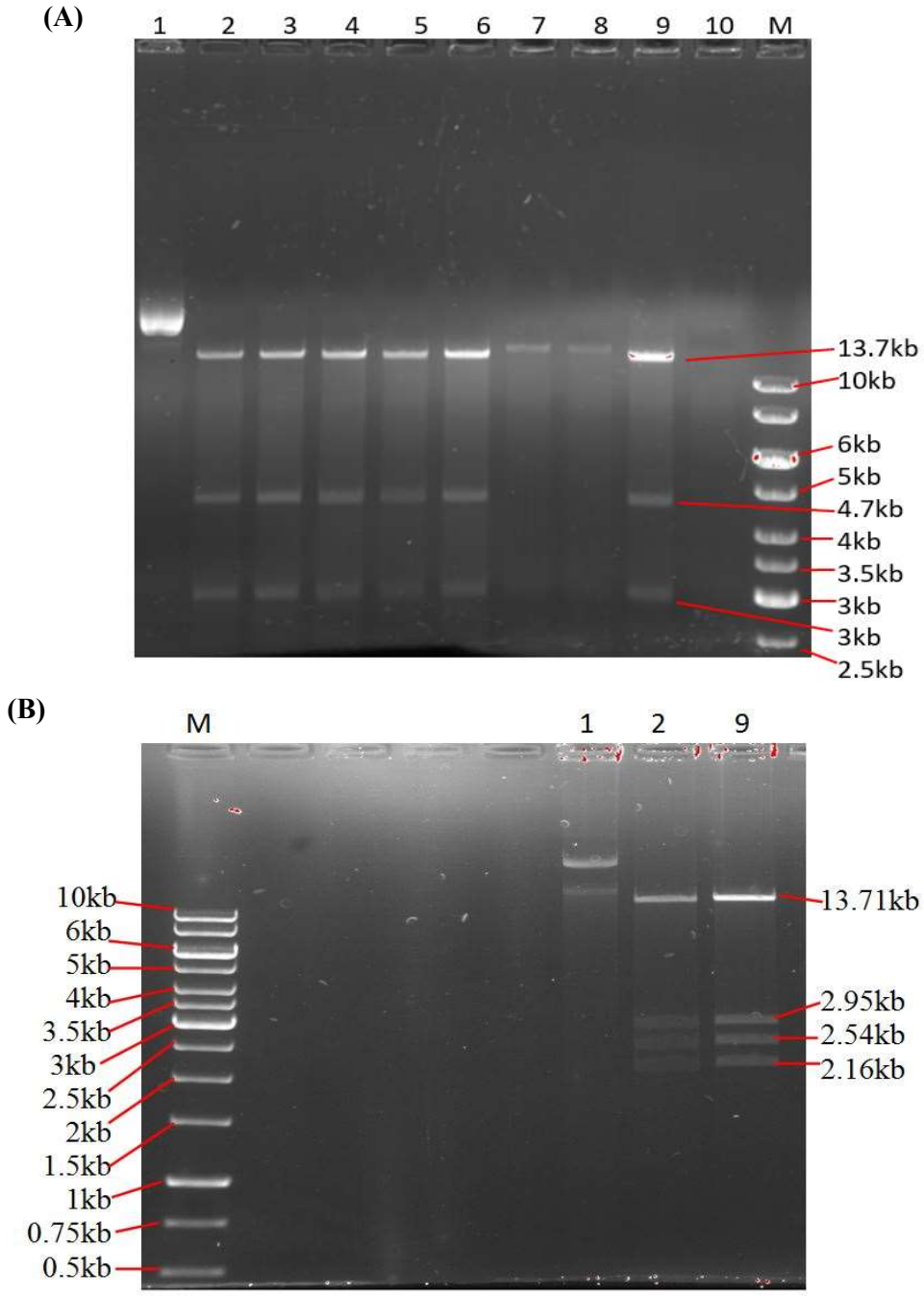


Figure 5.5 Confirmation of the correct orientation of genes in the constructed expression vector pMY23. Created by ligating vectors pMY10 and pMY22 through their *Xba*I sites (A) Top gel *Eco*RI digestion. No band of 2.4 kb could be seen in any of the clones and presence of a band of 4.7 kb in all of the six clones out of 10 analysed confirmed that they had genes in the correct orientation in the cointegrate. (B) Bottom gel *Eco*RI and *Sac*I double digestion. Correct pMY10 clones used were again confirmed by double digestion. Clones in lane 2 and 9 of upper and lower gel are same. Lane M in both had the 1 kb ladder. Lane 1 in upper and lower gel had uncut plasmid.

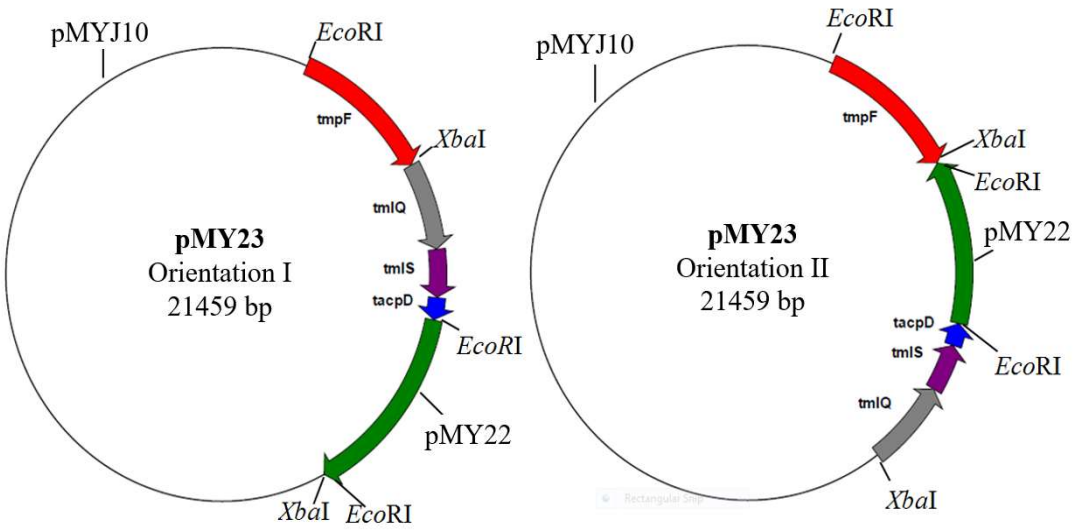


Figure 5.6 Schematic presentation of two of the possible orientations of *tml* genes in the expression vector pMY23, a cointegrate created by ligating vectors pMY10 and pMY22 through their *Xba*I restriction sites. The desirable orientation was orientation I (figures not to scale).

5.2.3 Complementation analysis of constructed *P. fluorescens* NCIMB10586 Δ*mmpF*Δ*mupS*Δ*mupQ*Δ*macpD* strain

To transfer the constructed vector pMY23, that should express the four chosen thiomarinol genes, into the mutant strain to test complementation it was transformed into *E. coli* S17-1 strain. Purified single colonies were used to mobilize this vector into the mutant of *P. fluorescens* NCIMB 10586Δ*mmpF*Δ*mupS*Δ*mupQ*Δ*macpD* (quadruple mutant), by conjugating the two strains together as per methodology described in Chapter 2 (section 2.4.14). The resulting strain *Pseudomonas fluorescens* NCIMB10586 Δ*mmpF*Δ*mupS*Δ*mupQ*Δ*macpD*(pMY23) was purified to single colonies before checking for complementation (Table 2.3).

A B C

Figure 5.7 Bioassay to show influence of deletion of *mmpF, mupS, mupQ* and *macpD* in *P. fluorescens* NCIMB10586 and, to show cross-complementation of this mutant by corresponding quadruple *tml* genes (*tmpF, tmlQ, tmlS* and *tacpD*) as a group. *P. fluorescens* NCIMB10586Δ*mmpF*Δ*mupS*Δ*mupQ*Δ*macpD* strain (A) and WT strain (B) taken as controls while NCIMB 10586 Δ*mmpF*Δ*mupS*Δ*mupQ* Δ*macpD* (pMY23) was the test strain (C). The zone of clearance around the central spot culture was taken as a measure of antibacterial activity.

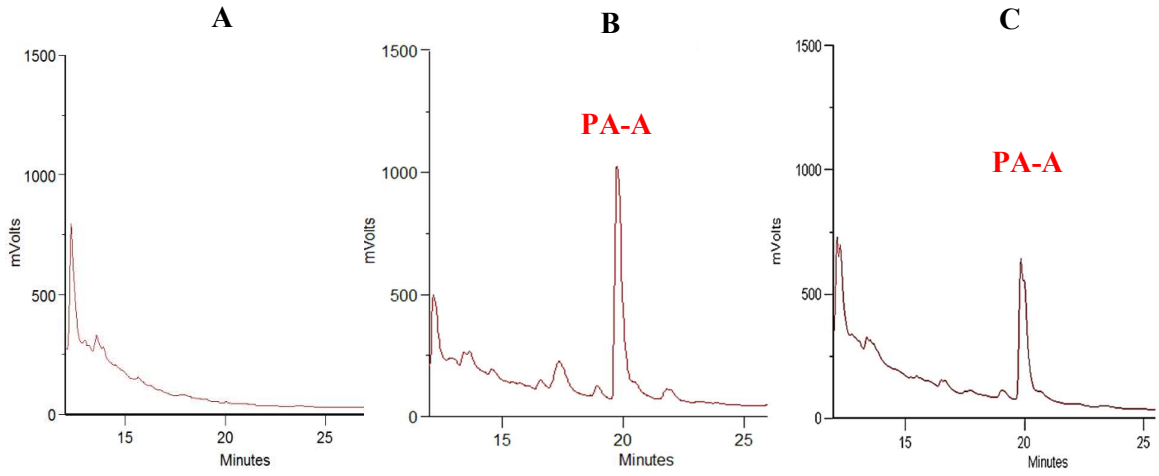


Figure 5.8 HPLC analysis of extracts from strains of *P. fluorescens* NCIMB 10586 for the complementation by quadruple *tml* genes (*tmpF, tmlQ, tmlS* and *tacpD*) of *P. fluorescens* NCIMB 10586Δ*mmp*FΔ*mup*SΔ*mup*QΔ*macp*D quadruple mutant strain. Test strain 10586Δ*mmpF*Δ*mupS*Δ*mupQ*Δ*macpD*(pMY23) (C) restored the production of PA-A (retention time 20.4 min). Quadruple mutant strain (A) did not produce PA-A while WT (B) produced PA-A both of which were used as controls.

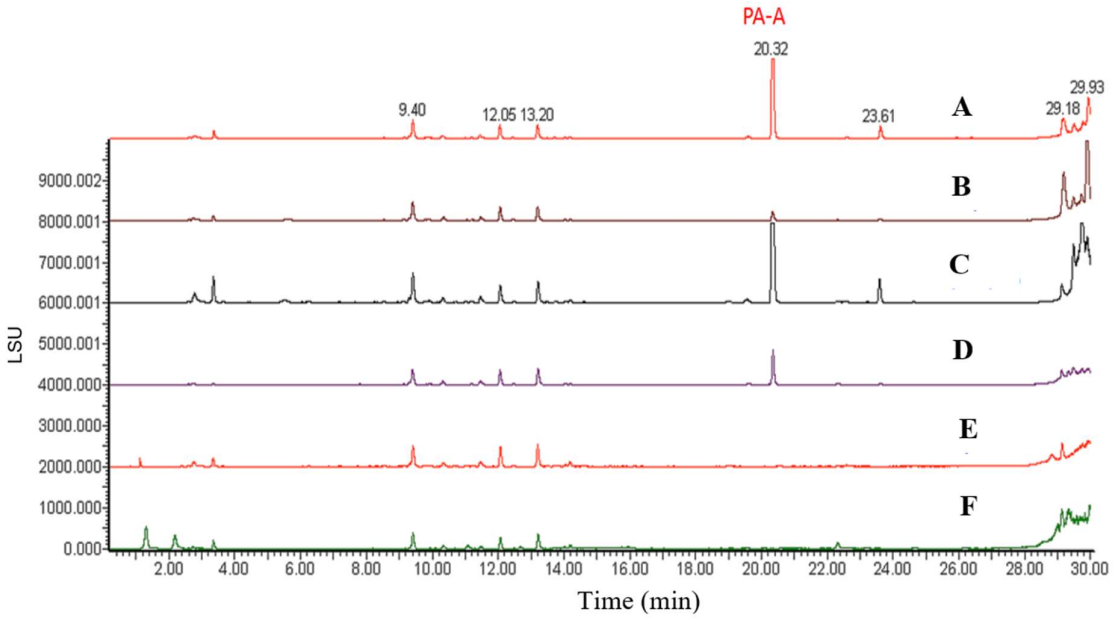


Figure 5.9 LC-MS chromatograms of extracts from strains of *P. fluorescens* NCIMB 10586 for the complementation by the group of four *tml* genes *tmlS, tmlQ, tacpD* and *tmpF* of Δ*mupS*Δ*mupQ*Δ*macpD*Δ*mmpF* quadruple mutant. Test strain C restored the production of PA-A in the presence of IPTG (0.5 mM). None of the two types of negative controls B, E or F produced PA-A. The control D is same as E or F but it turned out to be false mutant as it produced PA-A (data by Dr Song of the University of Bristol, UK). A, WT; B, NCIMB10586Δ*mmpF*Δ*mupS*Δ*mupQ* Δ*macpD*(pJH10); C, NCIMB 10586Δ*mmpF*Δ*mupS*Δ*mupQΔmacpD*(pMY23); D-F, NCIMB 10586 Δ*mmpF*Δ*mupS*Δ*mupQ*Δ*macpD*; LSU, light scattering units.

Results of *in trans* expression of the group of thiomarinol genes *tmpF, tmlQ, tmlS* and *tacpD* cloned under *tac* promoter in *P. fluorescens* NCIMB 10586 *ΔmmpF ΔmupS ΔmupQ ΔmacpD* mutant strain clearly indicated functional complementation by this group of thiomarinol genes expressed together. All the three test strains restored the production of PA-A as confirmed by HPLC. The quadruple mutant strain of *P. fluorescens* NCIMB 10586 did not produce PA-A (Figures 5.7and 5.8). The results were confirmed by LC-MS study for one test strain and the mutants (Figure 5.9).

5.3 Discussion

The idea to achieve complementation by *in trans* expression of a group of genes that are predicted to work together, in an heterologous system, when singular complementation by all or any gene(s) of this group in this other system had failed, was based on the fact that it might be possible that the product of the single gene did not fit in the host's biosynthetic pathway, but the group of genes might process a metabolic intermediate and turn it into a product that could be accepted and processed by the downstream biosynthetic machinery of the other biosynthetic system. It is a widely observed fact that the product of a group of genes that are located together, also collaborates together to process the metabolic intermediates. It has also been shown that gene pairs that were conserved as part of operon architecture across different prokaryotic systems, also suggested that protein products of such gene pairs physically interact (Dandekar et al., 1998).

It has already been shown that complementation can be achieved by expressing related genes together as a group, when either member of the group completely failed to show any complementation when it was expressed individually in the respective single mutant. This could be because of restrictions that were imposed either genetically or because product of action of single gene failed to get processed by the downstream biosynthetic machinery without the action of product of another related gene. It had been shown by the complementation studies of *mupK/mupJ* and *mupD/mupE* genes in *P. fluorescens* NCIMB 10586 wherein these gene pairs, that are located next to each other, could only be complemented when these were expressed as pairs. Otherwise there was no complementation shown by either member of gene pairs when expressed singularly.

Results for the cross-complementation of group of *tml* genes namely *tmlS, tacpD, tmpF* and *tmlQ* expressed together *in trans* cloned under the *tac* promoter in the

corresponding quadruple knockout of *P. fluorescens* NCIMB 10586, have been positive. All the other members of this group of *tml* genes except *tmlS*, failed to show any complementation when these were expressed singularly *in trans* in the corresponding single knockouts of *P. fluorescens* NCIMB10586.

Results of HPLC for the three test mutants that were constructed for this group complementation study, clearly showed restoration of pseudomonic acid-A peak while at the same time concerned quadruple mutants did not show peak characteristic of pseudomonic acid-A (retention time 20.3 min). The result was also confirmed with LC-MS study which was consistent with the earlier results where deletion of *mupS*, a member of the gene group tested, completely abolished the production of pseudomonic acids by *P. fluorescens* NCIMB 10586 (Cooper et al., 2005b). The result of group complementation was positive despite the fact that only one member of this group *tmlS* showed complementation and restored the production of pseudomonic acid-A only partially, when it was expressed singularly *in trans* under the same *tac* promoter in the corresponding single knockout strain (Chapter 4, section 4.3.6).

On the basis of bioinformatics studies it has been pointed out that groups of genes like as those of *tml* cluster or of *mup* cluster, are not only closely located with each other but are also involved in the furnishing of starter unit for the biosynthesis of fatty acids (Chen et al., 2006, Piel, 2010, Gao et al., 2014, Helfrich and Piel, 2016). In the biosynthesis of mupirocin, it has been proposed that enzymes like MupS (reductase), MacpD (acyl carrier protein) and MupQ (acyl-CoA synthase) are involved in furnishing of the starter unit from 3-hydroxypropionate for the biosynthesis of its 9-hydroxynonanoic acid component (Feline et al., 1977, Thomas et al., 2010, Murphy et al., 2014). Though there is no direct evidence to support this but it is indirectly supported by studies done by feeding of radioactive substrates

(Feline et al., 1977, Murphy et al., 2014). In the metabolism, 3-hydroxypropionate acts as a starter unit for the biosynthesis of fatty acids with odd number of carbon atoms. It has been suggested that 9-HN is produced by the three rounds of condensation reactions catalysed by MmpB using malonate as extender unit and probably for each cycle of condensation it uses its three ACPs.

Contrary to the results of singular complementation shown by all the members of this group of *tml* genes, result of group complementation indicated that all these enzymes of this group in mupirocin biosynthesis were in the same role as they played in thiomarinol biosynthesis. This suggested that some other gene function in thiomarinol biosynthetic system causes the difference of one carbon from the fatty acid component of pseudomonic acids, or it is the result of some other gene function in *mup* biosynthetic system that causes its fatty acid to have one more extra carbon compared to that of thiomarinols.

Results of group complementation compared to that for single complementation clearly indicated that the biosynthetic intermediates processed by the products of genes *tmlS, tacpD, tmlQ* and *tmpF,* which have same chemical nature as their corresponding *mup* counterparts, were acceptable to the downstream biosynthetic machinery of mupirocin biosynthetic system. While products of *tacpD, tmlQ* and *tmpF* when expressed singularly in the corresponding single *mup* mutant completely failed to accept or process biosynthetic intermediates of the mupirocin biosynthetic assembly line (see sections 4.3.8, 4.3.7, and 4.3.9). This was despite products of each of these gene functions shared significant amino acid sequence identity with the corresponding gene functions of the *mup* system, which indicated a divergence in the evolution of these enzymes from those of *mup* system.

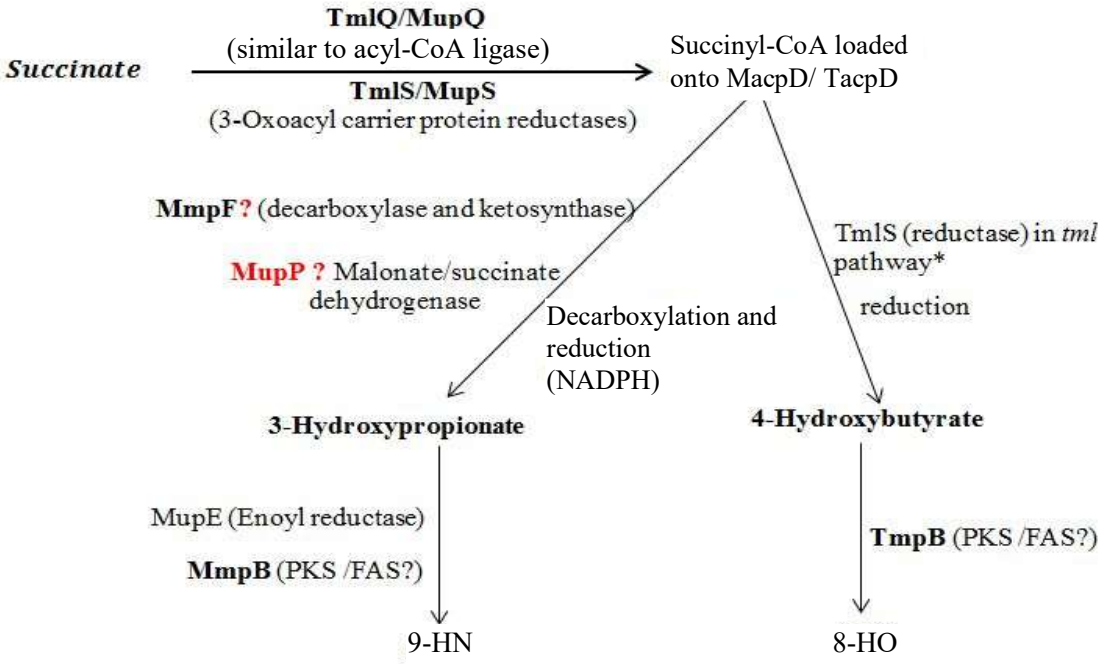


Figure 5.10 Predicted scheme for 9-HN and 8-HO acid biosynthesis in mupirocin and thiomarinol biosynthesis.(*) It is proposed that TmlS protein is in dual role in *tml* biosynthetic pathway where it is also involved in the reduction of succinyl-CoA apart from playing role in its loading onto TacpD protein similarly like MupS.

Positive complementation by group of *tml* genes, individual members of which either showed no complementation or complemented partially (*tmlS*), indicated that concerned proteins of the group shared enough homology so that working together these could complement the task of similar group of proteins in the other system (*mup* system). This supported the prediction that groups of these proteins work together, which was made on the basis of close location of their genes in the respective biosynthetic cluster and on the basis of *in silico* predictions about them. Since it is already established that 4-hydroxybutyrate is the starter unit for the biosynthesis of 8-HO in thiomarinols (Murphy et al., 2014), this indicated that the biosynthetic pathway so far as action of these enzymes as a group is concerned, is same in both the systems, mupirocin and thiomarinol. Otherwise, there could not have been

restoration of pseudomonic acid production by the quadruple mutant strain upon *in trans* expression of corresponding *tml* genes. This tells us that the biosynthetic steps in *mup* biosynthetic pathway deviate from *tml* cluster in the furnishing of starter units for their respective fatty acid moieties. Based on this it is proposed that there is some other gene function, product of which is involved in the furnishing of 3-HP in *mup* system from succinate, which is the substrate, from which in *tml* biosynthetic pathway 4-hydroxybutyrate is derived by three members of this group of genes for priming the biosynthesis of 8-HO.

The gene function *tmpF,* which is located next to this group of genes in the *tml* cluster, complemented the functions of MmpF but only when complemented together with this group of genes, otherwise singularly it failed to complement. This indicated that TmpF worked along with the product of this group of genes. It is established that Δ*mmpF* mutant of *P. fluorescens* NCIMB10586 does not produce PA-As. Based on the role of other similar genes that are often found located near group of genes with which they function in furnishing of starter units like role played by *diffB* in difficidin biosynthetic cluster of *Bacillus amyloliquifaciens* (Chen et al., 2006), it is proposed that *mupP* which is located next to this group of genes in *mup* cluster, is a putative malonate dehydrogenase. Probably it is MupP, that reduces the malonate derived from the succinyl-CoA, decarboxylation of which to give malonate is catalysed may be by MmpF or by some other protein, to yield 3-HP substrate. This scheme is shown in Figure 5-10.

CHAPTER 6

General Discussion and Future Work

6.1 General comments

In the past few decades a large number of polyketides and their derivatives have been discovered. Out of these, many have been found to be effective in the treatment of various diseases working as immuno-suppressants, anticancer agents, antifungals, anti-parasitic agents and antibacterial agents and, therefore, are already in clinical use. But during the same period and with the discovery of new antibiotics, resistance to them has also been growing (Davies and Davies, 2010, Ventola, 2015). In fact, during last few decades there has been rapid increase in the rate by which resistance is developing in pathogens against available antibiotics because of their unregulated use (Ventola, 2015). Unfortunately, the rate of development of resistance against existing antibiotics has outpaced that of discovery of new antibiotic molecules that can tackle the situation. In fact, for certain human pathogens we are left with no effective antibiotic and there are very few available for some other pathogens. The scene on growing antibiotic resistance, has become so alarming that the organizations like World Health Organization (WHO) and World Economic Forum have taken several initiatives to control it (Schaberle and Hack, 2014, Nathan and Cars, 2014). Now there is a lot of pressure to look out for new molecules with novel activities and improved properties that can qualify as drugs for clinical use (Doshi, 2015). As part of new drug discovery initiatives, several other compounds like non-ribosomal proteins that have been successfully used as drugs in medicine are also being considered (Mootz et al., 2000, Williams, 2013). Many compounds from diverse natural sources, some of which are found as polyketide-NRPS hybrid molecules of different types, are also being tested for clinical use and commercial exploitation. Apart from chemical methods that are being applied to synthesise new molecules or modify existing molecules to get new bioactive molecules, approaches of combinatorial biosynthesis are also being explored. As part of this drive, techniques of

genetic engineering are being applied to exploit and integrate biosynthetic steps from various biosynthetic clusters to achieve the production of hybrid or modified bioactive molecules with improved properties. Success of combinatorial biosynthesis completely depends on in-depth knowledge about the biosynthetic steps and metabolic intermediates of the biosynthetic clusters involved. The existence of all the genes at one place in a genome as shown in different biosynthetic clusters and the modular architecture of PKS assembly lines along with colinearity in their biosynthesis have strongly supported the approaches of combinatorial biosynthesis (Weissman and Leadlay, 2005, Cummings et al., 2014). That is also being extended to type II and III PKSs after understanding of their pattern of biosynthesis and to other modular assembly lines like NRPS and FAS involved in the production of diverse secondary metabolites (Weissman and Leadlay, 2005). Despite the small number of successful examples, combinatorial biosynthesis is still under development because its success depends upon the molecular compatibility of genes or proteins from different sources. Attempts are being made to bypass regularly-encountered bottlenecks by swapping gene functions (domains, modules, tailoring functions etc.) from different systems by developing standard sets of linkers (inter-domain and inter-modular linkers) that can be used to reprogramme components from different PKSs so that they can act as functional assembly lines (Weissman, 2004). Similarly, attempts are being made to develop pairs of docking domains that can be used in reprogrammed biosynthetic systems to facilitate interaction of proteins in the correct order, thus ensuring their functionality. All these efforts are aimed at getting engineered modular assembly lines to deliver new molecules with high potency and desired properties like non-toxicity as well as in large amounts.

Mupirocin and thiomarinol are two antimicrobials of importance, biosynthetic clusters of which pose new challenges to the approaches of combinatorial biosynthesis, thus adding value to study them. It is because these systems follow the principle of colinearity of

biosynthesis only partially that is shown by type I PKSs like Erythromycin. As described in Chapter 1, mupirocin and thiomarinol are synthesised by mixed clusters which belonged to the *trans*-AT group of polyketides. Although mupirocin is still an effective drug against MRSA, it suffers from structural weakness and, therefore, it cannot be used parenterally. While thiomarinols are broad spectrum antimicrobials effective even against some Gram negative bacteria, still these cannot be used because of their toxicity to eukaryotic cells. As a result, these molecules are the subject of ongoing research to improve or introduce them as drugs (Thomas et al., 2010, Fukuda et al., 2011, Murphy et al., 2011, Hothersall et al., 2011, Gao et al., 2014, Dunn et al., 2015)(Fukuda et al., 2011, Murphy et al., 2011, Murphy et al., 2014, Hothersall et al., 2011, Gao et al., 2014). Remarkable structural similarities between these two molecules that are also reflected at a genetic level are indicative of the fact that these two systems might be having certain biosynthetic steps and intermediates in common. This further provides an opportunity to exploit the biosynthetic pathways of the two systems to engineer hybrid molecules using the approaches of combinatorial biosynthesis. Thiomarinols also provide the opportunity to develop new molecules by mutasynthesis because it is naturally a hybrid molecule of two known antibiotics. As a result, the research is ongoing in different aspects of these two biosynthetic pathways. Although most of the biosynthetic steps in the mupirocin biosynthetic pathway have been deduced, gene functions behind certain modifications and timing of various events still remained to be known. On the other hand, very little is known about thiomarinol biosynthetic pathway. This study, in general, contributes to the field of antibiotics particularly to the *trans*-AT group of antibiotic and more specifically to the biosynthesis of mupirocin and thiomarinol antibiotics.

6.2 Main findings and key conclusions

6.2.1 *There is divergence in the evolution of pair of proteins that are tested belonging to the tailoring region of the biosynthetic clusters of these two antibiotics*

Each of the thiomarinol tailoring region proteins namely TmlO, TmlC, TmlF, TmlK, TmlQ, TmpF and TacpD shared significant amino acid sequence identity with its equivalent *mup* protein encoded by the tailoring region functions of the mupirocin biosynthetic cluster. But despite this, all of these failed to show functional cross-complementation when expressed in the mupirocin system indicating divergence in the function of pair of these enzymatic proteins of the two clusters during evolution.

6.2.2 *There has been convergence in the evolution of reductases and hydratases belonging to the tailoring region of the thiomarinol/mupirocin biosynthetic cluster.*

Both of these tested enzymatic proteins belonging to the tailoring regions of *tml* cluster namely TmlS and TmlJ that shared significant identity of amino acid sequence with the respective Mup proteins, showed positive complementation. TmlS and MupS both are putative reductases while TmlJ and MupJ are putative enoyl-CoA hydratases. This indicated that there has been convergence during the evolution of these protein pairs as a result of which there is conservation of function between them.

6.2.3 *The biosynthetic steps in β-methyl branching (C-15) in the thiomarinol biosynthesis are different from those involved in this branching in mupirocin biosynthesis.*

Based on analysis of the products of defined in-frame knockouts of genes like *mupH, mupJ, mupG, mupK* and *macpC* which yielded same product (mupirocin H) for all these mutations, products of these genes have been implicated in the biosynthetic steps involved in

β-methylation (C-15) of C-3 of the pseudomonic acids. This is also supported by analogy of existence of similar gene functions in other similar biosynthetic clusters. Biochemically, *mupJ* and *mupK* are predicted to be enoyl-CoA hydratases. The *mupJ* and *mupK* are found associated in *mup* cluster with the group of other genes that are HMG-CoA synthase, ACP and KS. Genes of this kind, which are found as a group called as "HMG cassette", are also found together in other biosynthetic clusters. This group of genes has been proposed to introduce a methyl group onto the polyketide backbone from acetate. In the mupirocin biosynthetic pathway, two enoyl-CoA hydratases MupJ and MupK have been proposed to act one after the other on the metabolic biosynthetic intermediate to dehydrate and then to decarboxylate it, respectively. It has also been observed that these enzymes act as protein complexes in a group. This was evident from the gene knockout and complementation studies done by Hothersall et al. (2007) by which it was shown that these genes could only be complemented to wild type levels when they were expressed together such that they were transcribed from a single mRNA in the double mutant of *mupJ* and *mupK*. Therefore, positive complementation shown by TmlJ in the corresponding single mutant of NCIMB 10586, which is also an enoyl-CoA like MupJ and with which it shares significant amino acid sequence identity, clearly indicated that the biosynthetic steps for β-methylation (C-15) of C-3 in thiomarinol biosynthetic pathway are completely different from that of mupirocin. Though the second enoyl-CoA hydratase is present in thiomarinols as fused protein TmpE, the N-terminal end of which shares significant amino acid sequence identity with MupK. But because TmlJ alone could complement functions of MupJ to the wild type levels in the single mutant, NCIMB10586 *ΔmupJ,* and also because MupJ failed to complement the same mutant alone without MupK, this indicated that TmlJ could act in the concerned biosynthetic pathway alone without the requirement of forming any protein complex with MupK of the kind that MupJ was required to form for proper functioning. Different biosynthetic steps for

β-branching pathway in two biosynthetic clusters also become evident from the fact that MupK in the Δ*mupJ* mutant did not interfere with the expression of TmlJ while it interfered when MupJ was expressed in it (Hothersall et al., 2007). Since complementation could be achieved only in the presence of IPTG, this indicated that higher amount of enzyme was required to restore the production of pseudomonic acids.

6.2.4 *The gene functions involved in furnishing starter unit for the biosynthesis of fatty acid component in both the system are homologous and share same substrate for deriving primer unit biosynthesis. With respect to this, both the biosynthetic pathways overlap to some extent but divulges later as far as functioning of three enzymes MupS, MupQ, and MacpD and their tml equivalents is concerned. The TmpF and MmpF are also proposed to have some role with them in the respective pathways. But because mup cluster has some other gene functions probably a dehydrogenase and/or a decarboxylase which together furnish a three carbon primer unit from the same substrate from which tml system derives its four carbon primer unit for the biosynthesis.*

Mupirocin antibiotic has fatty acid 9-hydroxynonanoic acid attached to the monic acid moiety by ester linkage while in thiomarinols this fatty acid is 8-hydroxyoctanoic acid. The biosynthetic steps involved in the biosynthesis of fatty acid components in these antibiotics are not known. In *mup* cluster, *mmpB* gene function is predicted to be involved in its biosynthesis by iterative mechanism only on the basis of indirect evidence (Thomas et al., 2010). Equivalent *tml* protein TmpB shares 36% identity of its amino acid sequence with MmpB. In mupirocin, biosynthesis of 9-HN is predicted to be from 3-hydroxypropionate. Based on *in silico* studies, it has been found that genes like *mupS, macpD, mupQ* and their thiomarinol equivalents are found to be involved in furnishing starter unit for furnishing of 3-hydroxypropionate. From which biosynthesis of 9-HN is proposed to takes place by three

condensations of malonate units. Since extender unit in thiomarinol system is also malonate, therefore, the difference of one carbon in these two fatty acids, 9-HN and 8-HO acid, is proposed to occur at the level of starter unit. The starter unit in thiomarinols is a four carbon unit and 8-HO is biosynthesised by the condensation of two malonates. Restoration of pseudomonic acid A production in NCIMB10586Δ*mupsΔmupQΔmmpFΔmacpD* quadruple mutant on expression of group of *tml* genes *tmlQ, tmlS, tmpF* and *tacpD* indicated that some other gene functions were responsible for this difference of one carbon in the fatty acid components of these two antibiotics.

6.2.5 *Complementation by a group of genes that are predicted to work together can be used as a tool to extend complementation between two systems, even when, complementation by any or all of the members of that group failed to show complementation when expressed singularly.*

As described in **C**hapter **4, gene** knockout and complementation studies in the mupirocin biosynthetic pathway have established that the products of genes *mupQ, macpD, mupS* and *mmpF* are necessary for the production of PA-A (Cooper et al., 2005b, Hothersall et al., 2007). *In silico* studies in other similar biosynthetic clusters have indicated that the products of genes of this type that are found located close to each other, also work together to furnish the starter unit for the biosynthesis of fatty acids (Chen et al., 2006, Thomas et al., 2010). Both pseudomonic acids and thiomarinols have a fatty acid component. In thiomarinol system such genes are *tmlQ, tmpF*, *tacpD* and *tmlS,* products of which share significant amino acid sequence identity with the products of corresponding gene functions of the *mup* system (Fukuda et al., 2011). Results in the chapter 4 (section 4.3) concluded that *in trans* complementation by each of the *tmlQ, tmpF* and *tacpD,* singularly, did not restore the production of PA-A in the corresponding *mup* single mutant strain of NCIMB10586,

however, it was only *tmlS* that restored the production of PA-A in the Δ*mupS* mutant but only partially. Contrary to this, complementation fully restored the production of PA-A when all these thiomarinol genes were expressed together, *in trans,* as a group in the corresponding *mup* (quadruple) mutant strain in which all the corresponding four genes were knocked out. This extended complementation supported *in silico* prediction that the products of group of corresponding four genes in both these clusters function together in the respective biosynthetic pathway. Products of these *tml* genes shared enough homology with their *mup* equivalents so that these were able to extend complementation while working together, compared to what was shown by individual members of *tml* group.

6.2.6 *The method for making defined gene knockouts that relied on suicide mutagenesis using* pAKE604 *(El Sayeed et al., 2001), which was used by Rahman et al. (2005) for making defined in-frame chromosomal mutations in Pseudomonas fluorescens* NCIMB 10586*, is not very efficient in Pseudoalteromonas sp.* SANK 73390 *and, therefore, does not guarantee mutations in pTML1.*

The method used for making mutants relied upon suicide mutagenesis and homologous recombination between the derivatives of suicide vector pAKE604 and the pTML1 plasmid of *Pseudoalteromonas sp.* SANK 73390. The site of homologous recombination was directed by cloning of specific pTML1 sequences in the suicide vector. Apart from this, method depended on conjugative transfer of suicide vector into SANK 73390 strain and effective selection of transconjugants using kanamycin (at the concentration of 50 µl/ml) and on negative selection of clones at a concentration of 5.5% w/v sucrose in M Agar. Negative selection of clones depended on effective expression of *sacB* gene in the SANK73390 strain. Failure to get knockouts of pTML1 in this study and of some different genes by others using this method while at the same time success of others to make knockouts of some genes

(Fukuda et al., 2011), raise doubts about the utility of this method for making SANK 73390 knockouts. This is strengthened by the fact that knockouts of some genes could only be made after several attempts by several workers while some other genes could not be deleted at all despite repeated attempts. Failure to get knockouts could be because of failure of cellular machinery of *Pseudoalteromonas sp.* SANK 73390 to support the suicide vector in the cell or to support the processes on which this method depended. This could be because of specialized adaptations of its cellular components, that are currently unknown, owing to its marine habitat. A variety of adaptations for survival in marine environment has been reported in the cellular components of marine bacteria (Bidle and Bartlett, 1999, Allen et al., 1999, Allen and Bartlett, 2000, Qin et al., 2011, Yu et al., 2013). As a result, the cellular processes on which success of this method depended were not supported in the laboratory environment where it was grown on artificial media out of its ecological habitat. New methods are being developed with improved protocols for achieving gene knockouts in *Pseudoalteromonas sp.* (Yu et al., 2014, Wang et al., 2015).

6.3 Future work

Creating defined knockouts of specific genes has been one of the most widely used method to elucidate the basic information about the role played by different gene functions in a biosynthetic system/cluster that involves examining changes in the phenotype by characterizing the biosynthetic intermediates with respect to every mutation that is made. Complementation studies using such defined knockouts have also been widely used as the next step to shed light on the organization of a gene cluster and to question the results of knockout studies made about specific genes to enquire their specific roles. A comprehensive analysis of various defined knockouts and their complementation about a biosynthetic cluster can give an overall idea about the protein-protein interactions taking place in that

biosynthetic system. The specific protein-protein interactions can then be identified using various methods available to study protein-protein interactions which depend on the detection or estimation of specific proteins, for example by co-immunoprecipitation, by quantitative immunoprecipitation combined with knock-down (QUICK) or by bimolecular fluorescence complementation (BiFC) to visualize the protein interactions etc. As mentioned earlier that though most of the steps in the biosynthesis of *mup* biosynthetic pathway have been determined mainly by using comprehensive gene knockout and complementation studies, timing for various biosynthetic steps as well as gene functions responsible for certain modifications still remained unknown. For example, it is still not known which gene function is responsible for 6-hydroxylation and what are the biosynthetic steps that lead to the formation of ester linkage between polyketide moiety and the fatty acid component. On the other hand, very little is known about the thiomarinol biosynthetic pathway. Knowledge of biosynthetic steps in these two systems mupirocin and thiomarinol is needed in order to genetically manipulate them to develop strains that produce molecules with better properties like high stability and no toxicity. Knowledge about common biosynthetic steps and intermediates in these two biosynthetic pathways is needed to develop novel molecules with improved activities and properties by applying the knowledge of combinatorial biosynthesis. This knowledge can be used to develop novel hybrid molecules out of these two molecules like pseudomonic acid derivatives with 8-HO acid or thiomarinol derivatives with 9-HN etc. which can be tested for novel activities and other beneficial properties.

There is need to develop a new method for making defined in-frame knockouts of the gene functions in the thiomarinol biosynthetic cluster because the method used did not always yield a result (El Sayeed et al., 2001, Fukuda et al., 2011). Qualitative analysis and quantitative estimation of products from defined mutants and elucidating their structures will shed light on the biosynthetic steps in the thiomarinol biosynthetic pathway.

Functional complementation of remaining ORFs of the mupirocin biosynthetic cluster, products of which show significant similarity of amino acid sequence identity with those of thiomarinol cluster, should also be completed. This will give a broader view of the common biosynthetic steps by way of protein-protein interaction taking place and gene functions between the two systems. On the other hand, results of negative complementation could be verified by confirming that there was actual expression of the concerned protein that was being tested. This could be done by various methods for example by mass-spectrometry or by using antibodies raised specifically against selected epitopes on purified proteins of interest.

Though the negative result of complementation of singular proteins that has been tested tells about their divergence, a better answer can be obtained by testing them by *in vitro* assay. Because of different cellular environment apart from genetic rearrangement of tailoring functions in the *mup* system, the *tml* enzymes might not have been active or did not pair up with the protein complex formed by other *mup* proteins to bring out the desirable phenotype. It has been shown that several *mup* proteins work as part of protein complexes. In this regard, interaction of TmlJ protein with the *mup* proteins should be investigated because of the positive result of its complementation. It is interesting because its equivalent partner protein MupJ itself failed to show complementation in the *mup* system when *mupJ* was expressed singularly without the gene function *mupK* which encodes another enoyl-CoA hydratase (MupK) proposed in the *mup* pathway.

In view of the failure of complementation of *tmlF* and *tmlC* genes to complement corresponding singular *mup* mutants of NCIMB 10586, group complementation by expressing them together in the concerned double mutant might help to achieve complementation. It is because both these ORFs are located next to each other in the *tml*

cluster and may have a requirement for being transcribed together as a single unit for efficient transcription. This will also shed light on question whether their polypeptides need to be folded together before they can become functionally active.

Cross-complementation by different gene combinations (sub-groups) among the group of genes that were tested along with those that are predicted, will reveal much more information about the key biosynthetic steps involved in furnishing primer unit for the biosynthesis of fatty acid component in pseudomonic acids.

The approach of reverse complementation for testing functional complementation by all the *mup* proteins that show significant similarity of amino acid sequence identity with the thiomarinol functions in the corresponding thiomarinol knockouts, should also be pursued. This could be done not only for singular gene functions but also to test complementation by group of genes, products of which are predicted to act together. This will help to get complete picture about the functional organization of different gene functions in these two biosynthetic clusters and will reveal similarities and differences in the biosynthetic steps and intermediates between the two when compared with the results of forward singular and group complementation studies. Complementation of various gene groups in view of their proposed functionality on account of *in silico* predictions or experimental results or their location in the respective biosynthetic cluster, should also reveal their functional organization in the cluster like regulatory requirements, operon organization and transcriptional requirements to become functional proteins.

REFERENCES

ADAMS, N. E., THIAVILLE, J. J., PROESTOS, J., JUAREZ-VAZQUEZ, A. L., MCCOY, A. J., BARONA-GOMEZ, F., IWATA-REUYL, D., DE CRECY-LAGARD, V. & MAURELLI, A. T. 2014. Promiscuous and adaptable enzymes fill "holes" in the tetrahydrofolate pathway in *Chlamydia* species. *MBio,* 5**,** e01378-14.

ALBERTINI, A. M., CARAMORI, T., SCOFFONE, F., SCOTTI, C. & GALIZZI, A. 1995. Sequence around the 159 degree region of the *Bacillus subtilis* genome: the *pksX* locus spans 33.6 kb. *Microbiology,* 141 (Pt 2)**,** 299-309.

ALLEN, E. E. & BARTLETT, D. H. 2000. FabF is required for piezoregulation of cis-vaccenic acid levels and piezophilic growth of the deep-Sea bacterium *Photobacterium profundum* strain SS9. *J Bacteriol,* 182**,** 1264-71.

ALLEN, E. E., FACCIOTTI, D. & BARTLETT, D. H. 1999. Monounsaturated but not polyunsaturated fatty acids are required for growth of the deep-sea bacterium *Photobacterium profundum* SS9 at high pressure and low temperature. *Appl Environ Microbiol,* 65**,** 1710-20.

ALTSCHUL, S. F., GISH, W., MILLER, W., MYERS, E. W. & LIPMAN, D. J. 1990. Basic Local Alignment Search Tool. *Journal of Molecular Biology,* 215**,** 403-410.

ANTONIO, M., MCFERRAN, N. & PALLEN, M. J. 2002. Mutations affecting the Rossman fold of isoleucyl-tRNA synthetase are correlated with low-level mupirocin resistance in *Staphylococcus aureus*. *Antimicrob Agents Chemother,* 46**,** 438-42.

ARISAWA, A., KAWAMURA, N., TAKEDA, K., TSUNEKAWA, H., OKAMURA, K. & OKAMOTO, R. 1994. Cloning of the macrolide antibiotic biosynthesis gene acyA, which encodes 3-O-acyltransferase, from *Streptomyces thermotolerans* and its use for direct fermentative production of a hybrid macrolide antibiotic. *Appl Environ Microbiol,* 60**,** 2657-60.

BIDLE, K. A. & BARTLETT, D. H. 1999. RecD function is required for high-pressure growth of a deep-sea bacterium. *J Bacteriol,* 181**,** 2330-7.

BISANG, C., LONG, P. F., CORTES, J., WESTCOTT, J., CROSBY, J., MATHARU, A. L., COX, R. J., SIMPSON, T. J., STAUNTON, J. & LEADLAY, P. F. 1999. A chain initiation factor common to both modular and aromatic polyketide synthases. *Nature,* 401**,** 502-5.

BORISOVA, S. A. & LIU, H. W. 2010. Characterization of glycosyltransferase DesVII and its auxiliary partner protein DesVIII in the methymycin/picromycin biosynthetic pathway. *Biochemistry,* 49**,** 8071-84.

BOWMAN, J. P. 2007. Bioactive compound synthetic capacity and ecological significance of marine bacterial genus *pseudoalteromonas*. *Mar Drugs,* 5**,** 220-41.

BREDHOLT, H., FJAERVIK, E., JOHNSEN, G. & ZOTCHEV, S. B. 2008. Actinomycetes from sediments in the Trondheim fjord, Norway: diversity and biological activity. *Mar Drugs,* 6**,** 12-24.

BROWN, M. J., MENSAH, L. M., DOYLE, M. L., BROOM, N. J., OSBOURNE, N., FORREST, A. K., RICHARDSON, C. M., O'HANLON, P. J. & POPE, A. J. 2000. Rational design of femtomolar inhibitors of isoleucyl tRNA synthetase from a binding model for pseudomonic acid-A. *Biochemistry,* 39**,** 6003-11.

CABALLERO, J. L., MARTINEZ, E., MALPARTIDA, F. & HOPWOOD, D. A. 1991. Organisation and functions of the actVA region of the actinorhodin biosynthetic gene cluster of *Streptomyces coelicolor*. *Mol Gen Genet,* 230**,** 401-12.

CALDERONE, C. T., IWIG, D. F., DORRESTEIN, P. C., KELLEHER, N. L. & WALSH, C. T. 2007. Incorporation of nonmethyl branches by isoprenoid-like logic: multiple beta-alkylation events in the biosynthesis of myxovirescin A1. *Chem Biol,* 14**,** 835-46.

CALDERONE, C. T., KOWTONIUK, W. E., KELLEHER, N. L., WALSH, C. T. & DORRESTEIN, P. C. 2006. Convergence of isoprene and polyketide biosynthetic machinery: isoprenyl-S-carrier proteins in the *pksX* pathway of *Bacillus subtilis*. *Proc Natl Acad Sci U S A,* 103**,** 8977-82.

CANE, D. E., WALSH, C. T. & KHOSLA, C. 1998. Harnessing the biosynthetic code: combinations, permutations, and mutations. *Science,* 282**,** 63-8.

CARMODY, M., MURPHY, B., BYRNE, B., POWER, P., RAI, D., RAWLINGS, B. & CAFFREY, P. 2005. Biosynthesis of amphotericin derivatives lacking exocyclic carboxyl groups. *Journal of Biological Chemistry,* 280**,** 34420-34426.

CHAIN, E. B. & MELLOWS, G. 1977a. Pseudomonic acid. Part 1. The structure of pseudomonic acid A, a novel antibiotic produced by *Pseudomonas fluorescens*. *J Chem Soc Perkin 1***,** 294-309.

CHAIN, E. B. & MELLOWS, G. 1977b. Pseudomonic acid. Part 3. Structure of pseudomonic acid B. *J Chem Soc Perkin 1***,** 318-24.

CHEN, X. H., VATER, J., PIEL, J., FRANKE, P., SCHOLZ, R., SCHNEIDER, K., KOUMOUTSI, A., HITZEROTH, G., GRAMMEL, N., STRITTMATTER, A. W., GOTTSCHALK, G., SUSSMUTH, R. D. & BORRISS, R. 2006. Structural and functional characterization of three polyketide synthase gene clusters in *Bacillus amyloliquefaciens* FZB 42. *J Bacteriol,* 188**,** 4024-36.

COOPER, S. M., COX, R. J., CROSBY, J., CRUMP, M. P., HOTHERSALL, J., LAOSRIPAIBOON, W., SIMPSON, T. J. & THOMAS, C. M. 2005a. Mupirocin W, a novel pseudomonic acid produced by targeted mutation of the mupirocin biosynthetic gene cluster. *Chem Commun (Camb)***,** 1179-81.

COOPER, S. M., LAOSRIPAIBOON, W., RAHMAN, A. S., HOTHERSALL, J., EL-SAYED, A. K., WINFIELD, C., CROSBY, J., COX, R. J., SIMPSON, T. J. & THOMAS, C. M. 2005b. Shift to Pseudomonic acid B production in *P. fluorescens* NCIMB10586 by mutation of mupirocin tailoring genes *mupO, mupU, mupV,* and *macpE. Chem Biol,* 12**,** 825-33.

CORDERO, O. X., WILDSCHUTTE, H., KIRKUP, B., PROEHL, S., NGO, L., HUSSAIN, F., LE ROUX, F., MINCER, T. & POLZ, M. F. 2012. Ecological populations of bacteria act as socially cohesive units of antibiotic production and resistance. *Science,* 337**,** 1228-31.

COUILLEROT, O., PRIGENT-COMBARET, C., CABALLERO-MELLADO, J. & MOENNE-LOCCOZ, Y. 2009. *Pseudomonas fluorescens* and closely-related fluorescent pseudomonads as biocontrol agents of soil-borne phytopathogens. *Lett Appl Microbiol,* 48**,** 505-12.

CRAGG, G. M. & NEWMAN, D. J. 2013. Natural products: a continuing source of novel drug leads. *Biochim Biophys Acta,* 1830**,** 3670-95.

CUMMINGS, M., BREITLING, R. & TAKANO, E. 2014. Steps towards the synthetic biology of polyketide biosynthesis. *Fems Microbiology Letters,* 351**,** 116-125.

DANDEKAR, T., SNEL, B., HUYNEN, M. & BORK, P. 1998. Conservation of gene order: a fingerprint of proteins that physically interact. *Trends Biochem Sci,* 23**,** 324-8.

DAVIES, J. & DAVIES, D. 2010. Origins and evolution of antibiotic resistance. *Microbiol Mol Biol Rev,* 74**,** 417-33.

DE BOER, H. A., COMSTOCK, L. J. & VASSER, M. 1983. The *tac* promoter: a functional hybrid derived from the *trp* and *lac* promoters. *Proc Natl Acad Sci U S A,* 80**,** 21-5.

DE CARVALHO, C. C. & FERNANDES, P. 2010. Production of metabolites as bacterial responses to the marine environment. *Mar Drugs,* 8**,** 705-27.

DE LIMA PROCÓPIO, R. E., DA SILVA, I. R., MARTINS, M. K., DE AZEVEDO, J. L. & DE ARAÚJO, J. M. 2012. Antibiotics produced by *Streptomyces. The Brazilian Journal of Infectious Diseases,* 16**,** 466-471.

DHILLON, N., HALE, R. S., CORTES, J. & LEADLAY, P. F. 1989. Molecular characterization of a gene from *Saccharopolyspora erythraea* (*Streptomyces erythraeus*) which is involved in erythromycin biosynthesis. *Mol Microbiol,* 3**,** 1405-14.

DONADIO, S. & KATZ, L. 1992. Organization of the enzymatic domains in the multifunctional polyketide synthase involved in erythromycin formation in Saccharopolyspora erythraea. *Gene,* 111**,** 51-60.

DONADIO, S., STAVER, M. J., MCALPINE, J. B., SWANSON, S. J. & KATZ, L. 1991. Modular organization of genes required for complex polyketide biosynthesis. *Science,* 252**,** 675-9.

DOSHI, P. 2015. Speeding new antibiotics to market: a fake fix? *BMJ,* 350**,** h1453.

DOUMITH, M., LEGRAND, R., LANG, C., SALAS, J. A. & RAYNAL, M. C. 1999. Interspecies complementation in *Saccharopolyspora erythraea*: elucidation of the function of *oleP1, oleG1* and *oleG2* from the oleandomycin biosynthetic gene cluster of *Streptomyces antibioticus* and generation of new erythromycin derivatives. *Molecular Microbiology,* 34**,** 1039-1048.

DU, L., SANCHEZ, C. & SHEN, B. 2001. Hybrid peptide-polyketide natural products: biosynthesis and prospects toward engineering novel molecules. *Metab Eng,* 3**,** 78-95.

DUNN, Z. D., WEVER, W. J., ECONOMOU, N. J., BOWERS, A. A. & LI, B. 2015. Enzymatic Basis of "Hybridity" in Thiomarinol Biosynthesis. *Angewandte Chemie-International Edition,* 54**,** 5137-5141.

EHMANN, D. E., GEHRING, A. M. & WALSH, C. T. 1999. Lysine biosynthesis in Saccharomyces cerevisiae: mechanism of alpha-aminoadipate reductase (Lys2) involves posttranslational phosphopantetheinylation by Lys5. *Biochemistry,* 38**,** 6171-7.

EL-SAYED, A. K. 2001. *Molecular biological studies on the genes of Pseudomonas fluorescens NCIMB10586 which encode the biosynthesis of polyketide antibiotic mupirocin.* PhD, The University of Birmingham, UK.

EL-SAYED, A. K., HOTHERSALL, J., COOPER, S. M., STEPHENS, E., SIMPSON, T. J. & THOMAS, C. M. 2003. Characterization of the Mupirocin Biosynthesis Gene Cluster from *Pseudomonas fluorescens* NCIMB 10586. *Chemistry & Biology,* 10**,** 419-430.

EL-SAYED, A. K., HOTHERSALL, J. & THOMAS, C. M. 2001. Quorum-sensing-dependent regulation of biosynthesis of the polyketide antibiotic mupirocin in *Pseudomonas fluorescens* NCIMB 10586. *Microbiology,* 147**,** 2127-39.

EMMERT, E. A., KLIMOWICZ, A. K., THOMAS, M. G. & HANDELSMAN, J. 2004. Genetics of zwittermicin a production by *Bacillus cereus*. *Appl Environ Microbiol,* 70**,** 104-13.

EPP, J. K., HUBER, M. L., TURNER, J. R., GOODSON, T. & SCHONER, B. E. 1989. Production of a hybrid macrolide antibiotic in *Streptomyces ambofaciens* and *Streptomyces lividans* by introduction of a cloned carbomycin biosynthetic gene from *Streptomyces thermotolerans. Gene,* 85**,** 293-301.

FELINE, T. C., JONES, R. B., MELLOWS, G. & PHILLIPS, L. 1977. Pseudomonic acid. Part 2. Biosynthesis of pseudomonic acid A. *J Chem Soc Perkin 1***,** 309-18.

FISCHBACH, M. A. & WALSH, C. T. 2006. Assembly-line enzymology for polyketide and nonribosomal Peptide antibiotics: logic, machinery, and mechanisms. *Chem Rev,* 106**,** 3468-96.

FRITZ, E., FEKETE, A., LINTELMANN, J., SCHMITT-KOPPLIN, P. & MECKENSTOCK, R. U. 2009. Isolation of two *Pseudomonas* strains producing pseudomonic acid A. *Syst Appl Microbiol,* 32**,** 56-64.

FUJII, I., WATANABE, A., SANKAWA, U. & EBIZUKA, Y. 2001. Identification of Claisen cyclase domain in fungal polyketide synthase WA, a naphthopyrone synthase of *Aspergillus nidulans*. *Chem Biol,* 8**,** 189-97.

FUKUDA, D., HAINES, A., BRAGGINTON, E., NZOKURUM, N., KODAMA, K., SHIOZAWA, H., HOTHERSALL, J., STEPHENS, E. & THOMAS, C. M. unpublished data. Genomics of marine bacteria producing the antibiotic thiomarinol. The University of Birmingham, UK.

FUKUDA, D., HAINES, A. S., SONG, Z., MURPHY, A. C., HOTHERSALL, J., STEPHENS, E. R., GURNEY, R., COX, R. J., CROSBY, J., WILLIS, C. L., SIMPSON, T. J. & THOMAS, C. M. 2011. A natural plasmid uniquely encodes two biosynthetic pathways creating a potent anti-MRSA antibiotic. *PLoS One,* 6**,** e18031.

FULLER, A. T., MELLOWS, G., WOOLFORD, M., BANKS, G. T., BARROW, K. D. & CHAIN, E. B. 1971. Pseudomonic acid: an antibiotic produced by *Pseudomonas fluorescens*. *Nature,* 234**,** 416-7.

GAISSER, S., BOHM, G. A., CORTES, J. & LEADLAY, P. F. 1997a. Analysis of seven genes from the *eryAI-eryK* region of the erythromycin biosynthetic gene cluster in *Saccharopolyspora erythraea*. *Mol Gen Genet,* 256**,** 239-51.

GAISSER, S., BOHM, G. A., DOUMITH, M., RAYNAL, M. C., DHILLON, N., CORTES, J. & LEADLAY, P. F. 1998. Analysis of *eryBI, eryBIII* and *eryBVII* from the erythromycin biosynthetic gene cluster in *Saccharopolyspora erythraea*. *Mol Gen Genet,* 258**,** 78-88.

GAISSER, S., TREFZER, A., STOCKERT, S., KIRSCHNING, A. & BECHTHOLD, A. 1997b. Cloning of an avilamycin biosynthetic gene cluster from *Streptomyces viridochromogenes* Tu57. *J Bacteriol,* 179**,** 6271-8.

GAO, S. S., HOTHERSALL, J., WU, J. E., MURPHY, A. C., SONG, Z. S., STEPHENS, E. R., THOMAS, C. M., CRUMP, M. P., COX, R. J., SIMPSON, T. J. & WILLIS, C. L. 2014. Biosynthesis of Mupirocin by *Pseudomonas fluorescens* NCIMB 10586 Involves Parallel Pathways. *Journal of the American Chemical Society,* 136**,** 5501-5507.

GAUTHIER, G., GAUTHIER, M. & CHRISTEN, R. 1995. Phylogenetic analysis of the genera *Alteromonas, Shewanella*, and *Moritella* using genes coding for small-subunit rRNA sequences and division of the genus *Alteromonas* into two genera, *Alteromonas* (emended) and *Pseudoalteromonas* gen. nov., and proposal of twelve new species combinations. *Int J Syst Bacteriol,* 45**,** 755-61.

GIORDANO, D., COPPOLA, D., RUSSO, R., DENARO, R., GIULIANO, L., LAURO, F. M., DI PRISCO, G. & VERDE, C. 2015. Marine Microbial Secondary Metabolites: Pathways, Evolution and Physiological Roles. *Adv Microb Physiol,* 66**,** 357-428.

GOKHALE, R. S., TSUJI, S. Y., CANE, D. E. & KHOSLA, C. 1999. Dissecting and exploiting intermodular communication in polyketide synthases. *Science,* 284**,** 482-5.

GOKHALE, R. S. & TUTEJA, D. 2008. Biochemistry of Polyketide Synthases. *Biotechnology Set.* Wiley-VCH Verlag GmbH.

GOTTARDI, E. M., KRAWCZYK, J. M., VON SUCHODOLETZ, H., SCHADT, S., MUHLENWEG, A., UGURU, G. C., PELZER, S., FIEDLER, H. P., BIBB, M. J.,

STACH, J. E. M. & SUSSMUTH, R. D. 2011. Abyssomicin Biosynthesis: Formation of an Unusual Polyketide, Antibiotic-Feeding Studies and Genetic Analysis. *Chembiochem,* 12**,** 1401-1410.

GREGORY, M. A., PETKOVIC, H., LILL, R. E., MOSS, S. J., WILKINSON, B., GAISSER, S., LEADLAY, P. F. & SHERIDAN, R. M. 2005. Mutasynthesis of rapamycin analogues through the manipulation of a gene governing starter unit biosynthesis. *Angew Chem Int Ed Engl,* 44**,** 4757-60.

GROSS, H. & LOPER, J. E. 2009. Genomics of secondary metabolite production by *Pseudomonas spp. Nat Prod Rep,* 26**,** 1408-46.

GU, L., JIA, J., LIU, H., HAKANSSON, K., GERWICK, W. H. & SHERMAN, D. H. 2006. Metabolic coupling of dehydration and decarboxylation in the curacin A pathway: functional identification of a mechanistically diverse enzyme pair. *J Am Chem Soc,* 128**,** 9014-5.

GU, L. C., GEDERS, T. W., WANG, B., GERWICK, W. H., HAKANSSON, K., SMITH, J. L. & SHERMAN, D. H. 2007. GNAT-like strategy for polyketide chain initiation. *Science,* 318**,** 970-974.

GURNEY, R. & THOMAS, C. M. 2011. Mupirocin: biosynthesis, special features and applications of an antibiotic from a gram-negative bacterium. *Appl Microbiol Biotechnol,* 90**,** 11-21.

GUSTAFSSON, C., GOVINDARAJAN, S. & MINSHULL, J. 2004. Codon bias and heterologous protein expression. *Trends Biotechnol,* 22**,** 346-53.

HA, K. R., PSALTIS, A. J., BUTCHER, A. R., WORMALD, P. J. & TAN, L. W. 2008. In vitro activity of mupirocin on clinical isolates of *Staphylococcus aureus* and its potential implications in chronic rhinosinusitis. *Laryngoscope,* 118**,** 535-40.

HAAS, D. & DEFAGO, G. 2005. Biological control of soil-borne pathogens by fluorescent pseudomonads. *Nat Rev Microbiol,* 3**,** 307-19.

HAINES, A. S., DONG, X., SONG, Z., FARMER, R., WILLIAMS, C., HOTHERSALL, J., PLOSKON, E., WATTANA-AMORN, P., STEPHENS, E. R., YAMADA, E., GURNEY, R., TAKEBAYASHI, Y., MASSCHELEIN, J., COX, R. J., LAVIGNE, R., WILLIS, C. L., SIMPSON, T. J., CROSBY, J., WINN, P. J., THOMAS, C. M. & CRUMP, M. P. 2013a. A conserved motif flags acyl carrier proteins for beta-branching in polyketide synthesis. *Nat Chem Biol,* 9**,** 685-92.

HAINES, A. S., DONG, X., SONG, Z. S., FARMER, R., WILLIAMS, C., HOTHERSALL, J., PLOSKON, E., WATTANA-AMORN, P., STEPHENS, E. R., YAMADA, E., GURNEY, R., TAKEBAYASHI, Y., MASSCHELEIN, J., COX, R. J., LAVIGNE, R., WILLIS, C. L., SIMPSON, T. J., CROSBY, J., WINN, P. J., THOMAS, C. M. & CRUMP, M. P. 2013b. A conserved motif flags acyl carrier proteins for beta-branching in polyketide synthesis. *Nature Chemical Biology,* 9**,** 685-+.

HAYDOCK, S. F., DOWSON, J. A., DHILLON, N., ROBERTS, G. A., CORTES, J. & LEADLAY, P. F. 1991. Cloning and sequence analysis of genes involved in erythromycin biosynthesis in *Saccharopolyspora erythraea*: sequence similarities between EryG and a family of S-adenosylmethionine-dependent methyltransferases. *Mol Gen Genet,* 230**,** 120-8.

HELFRICH, E. J. & PIEL, J. 2016. Biosynthesis of polyketides by *trans*-AT polyketide synthases. *Nat Prod Rep,* 33**,** 231-316.

HERTWECK, C. 2009. The biosynthetic logic of polyketide diversity. *Angew Chem Int Ed Engl,* 48**,** 4688-716.

HIRAMATSU, K., CUI, L., KURODA, M. & ITO, T. 2001. The emergence and evolution of methicillin-resistant *Staphylococcus aureus*. *Trends Microbiol,* 9**,** 486-93.

HIROE, A., TSUGE, K., NOMURA, C. T., ITAYA, M. & TSUGE, T. 2012. Rearrangement of Gene Order in the *phaCAB* Operon Leads to Effective Production of Ultrahigh-Molecular-Weight Poly[(R)-3-Hydroxybutyrate] in Genetically Engineered *Escherichia coli. Applied and Environmental Microbiology,* 78**,** 3177-3184.

HODGSON, J. E., CURNOCK, S. P., DYKE, K. G., MORRIS, R., SYLVESTER, D. R. & GROSS, M. S. 1994. Molecular characterization of the gene encoding high-level mupirocin resistance in *Staphylococcus aureus* J2870. *Antimicrob Agents Chemother,* 38**,** 1205-8.

HOLMSTROM, C. & KJELLEBERG, S. 1999. Marine *Pseudoalteromonas* species are associated with higher organisms and produce biologically active extracellular agents. *FEMS Microbiol Ecol,* 30**,** 285-293.

HONG, J. S., PARK, S. J., PARAJULI, N., PARK, S. R., KOH, H. S., JUNG, W. S., CHOI, C. Y. & YOON, Y. J. 2007. Functional analysis of desVIII homologues involved in glycosylation of macrolide antibiotics by interspecies complementation. *Gene,* 386**,** 123-30.

HOPWOOD, D. A. 2007. *Streptomyces in Nature and Medicine:The Antibiotic Makers: The Antibiotic Makers*, Oxford University Press, USA.

HOPWOOD, D. A. & SHERMAN, D. H. 1990. Molecular genetics of polyketides and its comparison to fatty acid biosynthesis. *Annu Rev Genet,* 24**,** 37-66.

HOTHERSALL, J. 2014. Unpublished data. The University of Birmingham, UK.

HOTHERSALL, J., MURPHY, A. C., IQBAL, Z., CAMPBELL, G., STEPHENS, E. R., WU, J., COOPER, H., ATKINSON, S., WILLIAMS, P., CROSBY, J., WILLIS, C. L., COX, R. J., SIMPSON, T. J. & THOMAS, C. M. 2011. Manipulation of quorum sensing regulation in *Pseudomonas fluorescens* NCIMB 10586 to increase mupirocin production. *Appl Microbiol Biotechnol,* 90**,** 1017-26.

HOTHERSALL, J., WU, J., RAHMAN, A. S., SHIELDS, J. A., HADDOCK, J., JOHNSON, N., COOPER, S. M., STEPHENS, E. R., COX, R. J., CROSBY, J., WILLIS, C. L., SIMPSON, T. J. & THOMAS, C. M. 2007. Mutational analysis reveals that all tailoring region genes are required for production of polyketide antibiotic mupirocin by *pseudomonas fluorescens*: pseudomonic acid B biosynthesis precedes pseudomonic acid A. *J Biol Chem,* 282**,** 15451-61.

HUGHES, J. & MELLOWS, G. 1978. Inhibition of isoleucyl-transfer ribonucleic acid synthetase in *Escherichia coli* by pseudomonic acid. *Biochem J,* 176**,** 305-18.

HUGLER, M., MENENDEZ, C., SCHAGGER, H. & FUCHS, G. 2002. Malonyl-coenzyme A reductase from *Chloroflexus aurantiacus*, a key enzyme of the 3-hydroxypropionate cycle for autotrophic CO(2) fixation. *J Bacteriol,* 184**,** 2404-10.

HUR, G. H., VICKERY, C. R. & BURKART, M. D. 2012. Explorations of catalytic domains in non-ribosomal peptide synthetase enzymology. *Nat Prod Rep,* 29**,** 1074-98.

HUTCHINSON, E., MURPHY, B., DUNNE, T., BREEN, C., RAWLINGS, B. & CAFFREY, P. 2010. Redesign of polyene macrolide glycosylation: engineered biosynthesis of 19-(O)-perosaminyl-amphoteronolide B. *Chem Biol,* 17**,** 174-82.

IKEDA, H., NONOMIYA, T., USAMI, M., OHTA, T. & OMURA, S. 1999. Organization of the biosynthetic gene cluster for the polyketide anthelmintic macrolide avermectin in *Streptomyces avermitilis*. *Proc Natl Acad Sci U S A,* 96**,** 9509-14.

JENSEN, K., NIEDERKRUGER, H., ZIMMERMANN, K., VAGSTAD, A. L., MOLDENHAUER, J., BRENDEL, N., FRANK, S., POPLAU, P., KOHLHAAS, C., TOWNSEND, C. A., OLDIGES, M., HERTWECK, C. & PIEL, J. 2012. Polyketide proofreading by an acyltransferase-like enzyme. *Chem Biol,* 19**,** 329-39.

JEZ, J. M., FERRER, J. L., BOWMAN, M. E., AUSTIN, M. B., SCHRODER, J., DIXON, R. A. & NOEL, J. P. 2001. Structure and mechanism of chalcone synthase-like polyketide synthases. *J Ind Microbiol Biotechnol,* 27**,** 393-8.

JIANG, M., ZHANG, H., PARK, S. H., LI, Y. & PFEIFER, B. A. 2013. Deoxysugar pathway interchange for erythromycin analogues heterologously produced through *Escherichia coli*. *Metab Eng,* 20**,** 92-100.

KAPUR, S., CHEN, A. Y., CANE, D. E. & KHOSLA, C. 2010. Molecular recognition between ketosynthase and acyl carrier protein domains of the 6-deoxyerythronolide B synthase. *Proc Natl Acad Sci U S A,* 107**,** 22066-71.

KAPUR, S., LOWRY, B., YUZAWA, S., KENTHIRAPALAN, S., CHEN, A. Y., CANE, D. E. & KHOSLA, C. 2012. Reprogramming a module of the 6-deoxyerythronolide B synthase for iterative chain elongation. *Proc Natl Acad Sci U S A,* 109**,** 4110-5.

KATZ, L. 1997. Manipulation of Modular Polyketide Synthases. *Chem Rev,* 97**,** 2557-2576.

KATZ, L. & DONADIO, S. 1993. Polyketide synthesis: prospects for hybrid antibiotics. *Annu Rev Microbiol,* 47**,** 875-912.

KEATINGE-CLAY, A. T. 2012. The structures of type I polyketide synthases. *Nat Prod Rep,* 29**,** 1050-73.

KENIG, M. & READING, C. 1979. Holomycin and an antibiotic (MM 19290) related to tunicamycin, metabolites of *Streptomyces clavuligerus*. *J Antibiot (Tokyo),* 32**,** 549-54.

KENNEDY, J., AUCLAIR, K., KENDREW, S. G., PARK, C., VEDERAS, J. C. & HUTCHINSON, C. R. 1999. Modulation of polyketide synthase activity by accessory proteins during lovastatin biosynthesis. *Science,* 284**,** 1368-72.

KHERSONSKY, O. & TAWFIK, D. S. 2010. Enzyme Promiscuity: A Mechanistic and Evolutionary Perspective. *Annual Review of Biochemistry, Vol 79,* 79**,** 471-505.

KHOSLA, C., GOKHALE, R. S., JACOBSEN, J. R. & CANE, D. E. 1999. Tolerance and specificity of polyketide synthases. *Annu Rev Biochem,* 68**,** 219-53.

KITTENDORF, J. D. & SHERMAN, D. H. 2006. Developing tools for engineering hybrid polyketide synthetic pathways. *Curr Opin Biotechnol,* 17**,** 597-605.

KOZUMA, S., HIROTA-TAKAHATA, Y., FUKUDA, D., KURAYA, N., NAKAJIMA, M. & ANDO, O. 2016. Identification and biological activity of ogipeptins, novel LPS inhibitors produced by marine bacterium. *J Antibiot (Tokyo)*.

KUMAR, P., KHOSLA, C. & TANG, Y. 2004. Manipulation and analysis of polyketide synthases. *Methods Enzymol,* 388**,** 269-93.

KURODA, M., OHTA, T., UCHIYAMA, I., BABA, T., YUZAWA, H., KOBAYASHI, I., CUI, L., OGUCHI, A., AOKI, K., NAGAI, Y., LIAN, J., ITO, T., KANAMORI, M., MATSUMARU, H., MARUYAMA, A., MURAKAMI, H., HOSOYAMA, A., MIZUTANI-UI, Y., TAKAHASHI, N. K., SAWANO, T., INOUE, R., KAITO, C., SEKIMIZU, K., HIRAKAWA, H., KUHARA, S., GOTO, S., YABUZAKI, J., KANEHISA, M., YAMASHITA, A., OSHIMA, K., FURUYA, K., YOSHINO, C., SHIBA, T., HATTORI, M., OGASAWARA, N., HAYASHI, H. & HIRAMATSU, K. 2001. Whole genome sequencing of meticillin-resistant *Staphylococcus aureus*. *Lancet,* 357**,** 1225-40.

LADNER, C. C. & WILLIAMS, G. J. 2016. Harnessing natural product assembly lines: structure, promiscuity, and engineering. *J Ind Microbiol Biotechnol,* 43**,** 371-87.

LAMBALOT, R. H., CANE, D. E., APARICIO, J. J. & KATZ, L. 1995. Overproduction and characterization of the erythromycin C-12 hydroxylase, EryK. *Biochemistry,* 34**,** 1858-66.

LAURO, F. M., TRAN, K., VEZZI, A., VITULO, N., VALLE, G. & BARTLETT, D. H. 2008. Large-scale transposon mutagenesis of *Photobacterium profundum* SS9 reveals new genetic loci important for growth at low temperature and high pressure. *J Bacteriol,* 190**,** 1699-709.

LEE, T. S., KHOSLA, C. & TANG, Y. 2005. Engineered biosynthesis of aklanonic acid analogues. *J Am Chem Soc,* 127**,** 12254-62.

LEVY, S. B. & MARSHALL, B. 2004. Antibacterial resistance worldwide: causes, challenges and responses. *Nat Med*.

LIM, H. N., LEE, Y. & HUSSEIN, R. 2011. Fundamental relationship between operon organization and gene expression. *Proceedings of the National Academy of Sciences of the United States of America,* 108**,** 10626-10631.

LIOU, G. F. & KHOSLA, C. 2003. Building-block selectivity of polyketide synthases. *Curr Opin Chem Biol,* 7**,** 279-84.

LIU, T., CHIANG, Y. M., SOMOZA, A. D., OAKLEY, B. R. & WANG, C. C. 2011. Engineering of an "unnatural" natural product by swapping polyketide synthase domains in *Aspergillus nidulans*. *J Am Chem Soc,* 133**,** 13314-6.

LIU, T., SANCHEZ, J. F., CHIANG, Y. M., OAKLEY, B. R. & WANG, C. C. 2014. Rational domain swaps reveal insights about chain length control by ketosynthase domains in fungal nonreducing polyketide synthases. *Org Lett,* 16**,** 1676-9.

LIU, W., CHRISTENSON, S. D., STANDAGE, S. & SHEN, B. 2002. Biosynthesis of the enediyne antitumor antibiotic C-1027. *Science,* 297**,** 1170-3.

LU, W., LEIMKUHLER, C., GATTO, G. J., JR., KRUGER, R. G., OBERTHUR, M., KAHNE, D. & WALSH, C. T. 2005. AknT is an activating protein for the glycosyltransferase AknS in L-aminodeoxysugar transfer to the aglycone of aclacinomycin A. *Chem Biol,* 12**,** 527-34.

MAHAJAN, G. B. & BALACHANDRAN, L. 2012. Antibacterial agents from actinomycetes - a review. *Front Biosci (Elite Ed),* 4**,** 240-53.

MANSUY, D. 1998. The great diversity of reactions catalyzed by cytochromes P450. *Comparative Biochemistry and Physiology C-Pharmacology Toxicology & Endocrinology,* 121**,** 5-14.

MARSDEN, A. F., CAFFREY, P., APARICIO, J. F., LOUGHRAN, M. S., STAUNTON, J. & LEADLAY, P. F. 1994. Stereospecific acyl transfers on the erythromycin-producing polyketide synthase. *Science,* 263**,** 378-80.

MATTHEUS, W., GAO, L. J., HERDEWIJN, P., LANDUYT, B., VERHAEGEN, J., MASSCHELEIN, J., VOLCKAERT, G. & LAVIGNE, R. 2010. Isolation and purification of a new kalimantacin/batumin-related polyketide antibiotic and elucidation of its biosynthesis gene cluster. *Chem Biol,* 17**,** 149-59.

MCCONEGHY, K. W., MIKOLICH, D. J. & LAPLANTE, K. L. 2009. Agents for the decolonization of methicillin-resistant *Staphylococcus aureus. Pharmacotherapy,* 29**,** 263-80.

MCDANIEL, R., EBERT-KHOSLA, S., HOPWOOD, D. A. & KHOSLA, C. 1995. Rational design of aromatic polyketide natural products by recombinant assembly of enzymatic subunits. *Nature,* 375**,** 549-54.

MEIER, J. L. & BURKART, M. D. 2009. The chemical biology of modular biosynthetic enzymes. *Chem Soc Rev,* 38**,** 2012-45.

MELANCON, C. E., 3RD, TAKAHASHI, H. & LIU, H. W. 2004. Characterization of *tylM3/tylM2* and *mydC/mycB* pairs required for efficient glycosyltransfer in macrolide antibiotic biosynthesis. *J Am Chem Soc,* 126**,** 16726-7.

MENDES, M. V., RECIO, E., FOUCES, R., LUITEN, R., MARTIN, J. F. & APARICIO, J. F. 2001. Engineered biosynthesis of novel polyenes: a pimaricin derivative produced by targeted gene disruption in *Streptomyces natalensis*. *Chemistry & Biology,* 8**,** 635-644.

MENZELLA, H. G., REID, R., CARNEY, J. R., CHANDRAN, S. S., REISINGER, S. J., PATEL, K. G., HOPWOOD, D. A. & SANTI, D. V. 2005. Combinatorial polyketide biosynthesis by de novo design and rearrangement of modular polyketide synthase genes. *Nat Biotechnol,* 23**,** 1171-6.

MOCHIZUKI, S., HIRATSU, K., SUWA, M., ISHII, T., SUGINO, F., YAMADA, K. & KINASHI, H. 2003. The large linear plasmid pSLA2-L of *Streptomyces rochei* has an unusually condensed gene organization for secondary metabolism. *Mol Microbiol,* 48**,** 1501-10.

MOHAMMAD, H. H. 2016. Unpublished data. University of Birmingham, Birmingham, United Kingdom.

MOIR, A., LAFFERTY, E. & SMITH, D. A. 1979. Genetics analysis of spore germination mutants of *Bacillus subtilis* 168: the correlation of phenotype with map location. *J Gen Microbiol,* 111**,** 165-80.

MONCRIEFFE, M. C., FERNANDEZ, M. J., SPITELLER, D., MATSUMURA, H., GAY, N. J., LUISI, B. F. & LEADLAY, P. F. 2012. Structure of the glycosyltransferase EryCIII in complex with its activating P450 homologue EryCII. *J Mol Biol,* 415**,** 92-101.

MOOTZ, H. D., SCHWARZER, D. & MARAHIEL, M. A. 2000. Construction of hybrid peptide synthetases by module and domain fusions. *Proceedings of the National Academy of Sciences of the United States of America,* 97**,** 5848-5853.

MORRIS, A., KELLNER, J. D. & LOW, D. E. 1998. The superbugs: evolution, dissemination and fitness. *Current Opinion in Microbiology,* 1**,** 524-529.

MURPHY, A. C., FUKUDA, D., SONG, Z., HOTHERSALL, J., COX, R. J., WILLIS, C. L., THOMAS, C. M. & SIMPSON, T. J. 2011. Engineered thiomarinol antibiotics active against MRSA are generated by mutagenesis and mutasynthesis of *Pseudoalteromonas* SANK73390. *Angew Chem Int Ed Engl,* 50**,** 3271-4.

MURPHY, A. C., GAO, S. S., HAN, L. C., CAROBENE, S., FUKUDA, D., SONG, Z. S., HOTHERSALL, J., COX, R. J., CROSBY, J., CRUMP, M. P., THOMAS, C. M., WILLIS, C. L. & SIMPSON, T. J. 2014. Biosynthesis of thiomarinol A and related metabolites of *Pseudoalteromonas sp* SANK 73390. *Chemical Science,* 5**,** 397-402.

MUSIOL, E. M. & WEBER, T. 2012. Discrete acyltransferases involved in polyketide biosynthesis. *MedChemComm,* 3**,** 871.

NAKAMA, T., NUREKI, O. & YOKOYAMA, S. 2001. Structural basis for the recognition of isoleucyl-adenylate and an antibiotic, mupirocin, by isoleucyl-tRNA synthetase. *J Biol Chem,* 276**,** 47387-93.

NATHAN, C. & CARS, O. 2014. Antibiotic Resistance - Problems, Progress, and Prospects. *New England Journal of Medicine,* 371**,** 1761-1763.

NEEDLEMA.SB & WUNSCH, C. D. 1970. A General Method Applicable to Search for Similarities in Amino Acid Sequence of 2 Proteins. *Journal of Molecular Biology,* 48**,** 443-&.

NEWMAN, D. J. & CRAGG, G. M. 2016. Natural Products as Sources of New Drugs from 1981 to 2014. *J Nat Prod,* 79**,** 629-61.

NGUYEN, H. C., KARRAY, F., LAUTRU, S., GAGNAT, J., LEBRIHI, A., HUYNH, T. D. & PERNODET, J. L. 2010. Glycosylation steps during spiramycin biosynthesis in

Streptomyces ambofaciens: involvement of three glycosyltransferases and their interplay with two auxiliary proteins. *Antimicrob Agents Chemother,* 54**,** 2830-9.

NGUYEN, K. T., KAU, D., GU, J. Q., BRIAN, P., WRIGLEY, S. K., BALTZ, R. H. & MIAO, V. 2006. A glutamic acid 3-methyltransferase encoded by an accessory gene locus important for daptomycin biosynthesis in *Streptomyces roseosporus*. *Mol Microbiol,* 61**,** 1294-307.

NGUYEN, T., ISHIDA, K., JENKE-KODAMA, H., DITTMANN, E., GURGUI, C., HOCHMUTH, T., TAUDIEN, S., PLATZER, M., HERTWECK, C. & PIEL, J. 2008. Exploiting the mosaic structure of trans-acyltransferase polyketide synthases for natural product discovery and pathway dissection. *Nat Biotechnol,* 26**,** 225-33.

NII. 2010. *The World of Polyketides* [Online]. New Delhi. [Accessed 30 May 2012].

NILSSON, B. & ANDERSON, S. 1991. Proper and Improper Folding of Proteins in the Cellular Environment. *Annual Review of Microbiology,* 45**,** 607-635.

NISHIZAKI, T., TSUGE, K., ITAYA, M., DOI, N. & YANAGAWA, H. 2007. Metabolic engineering of carotenoid biosynthesis in *Escherichia coli* by ordered gene assembly in *Bacillus subtilis*. *Applied and Environmental Microbiology,* 73**,** 1355-1361.

OLANO, C., MENDEZ, C. & SALAS, J. A. 2010. Post-PKS tailoring steps in natural product-producing actinomycetes from the perspective of combinatorial biosynthesis. *Natural Product Reports,* 27**,** 571-616.

ORLOVA, T. I., BULGAKOVA, V. G. & POLIN, A. N. 2015. [Secondary Metabolites from Marine Microorganisms. I. Secondary Metabolites from Marine Actinomycetes]. *Antibiot Khimioter,* 60**,** 47-59.

PALLERONI, N. J. 1984. Genus *Pseudomonas In Bergey's Manual of Systematic Bacteriology,* Vol. 1. Krieg, N.R., and Holt, J.G. (eds) Baltimore, MD, USA: Williams & Wilkins pp. 141–199.

PALLERONI, N. J. 2008. The road to the taxonomy of *Pseudomonas*. *In Pseudomonas. Genomics and Molecular Biology.,* Cornelis, P. (ed.). Norfolk, UK: Caister Academic Press**,** pp. 1–18.

PALLERONI, N. J. 2010. The Pseudomonas story. *Environ Microbiol,* 12**,** 1377-83.

PATRICK, W. M., QUANDT, E. M., SWARTZLANDER, D. B. & MATSUMURA, I. 2007. Multicopy suppression underpins metabolic evolvability. *Mol Biol Evol,* 24**,** 2716-22.

PAULUS, T. J., TUAN, J. S., LUEBKE, V. E., MAINE, G. T., DEWITT, J. P. & KATZ, L. 1990. Mutation and cloning of *eryG*, the structural gene for erythromycin O-methyltransferase from *Saccharopolyspora erythraea*, and expression of *eryG* in *Escherichia coli*. *J Bacteriol,* 172**,** 2541-6.

PHILLIPS, G., GROCHOWSKI, L. L., BONNETT, S., XU, H., BAILLY, M., BLABY-HAAS, C., EL YACOUBI, B., IWATA-REUYL, D., WHITE, R. H. & DE CRECY-LAGARD, V. 2012. Functional promiscuity of the COG0720 family. *ACS Chem Biol,* 7**,** 197-209.

PIEL, J. 2010. Biosynthesis of polyketides by *trans*-AT polyketide synthases. *Nat Prod Rep,* 27**,** 996-1047.

PIEL, J., WEN, G. P., PLATZER, M. & HUI, D. Q. 2004. Unprecedented diversity of catalytic domains in the first four modules of the putative pederin polyketide synthase. *Chembiochem,* 5**,** 93-98.

PULSAWAT, N., KITANI, S. & NIHIRA, T. 2007. Characterization of biosynthetic gene cluster for the production of virginiamycin M, a streptogramin type A antibiotic, in *Streptomyces virginiae*. *Gene,* 393**,** 31-42.

QIN, Q. L., LI, Y., ZHANG, Y. J., ZHOU, Z. M., ZHANG, W. X., CHEN, X. L., ZHANG, X. Y., ZHOU, B. C., WANG, L. & ZHANG, Y. Z. 2011. Comparative genomics

reveals a deep-sea sediment-adapted life style of *Pseudoalteromonas sp.* SM9913. *ISME J,* 5**,** 274-84.

RAHMAN, A. S., HOTHERSALL, J., CROSBY, J., SIMPSON, T. J. & THOMAS, C. M. 2005. Tandemly duplicated acyl carrier proteins, which increase polyketide antibiotic production, can apparently function either in parallel or in series. *J Biol Chem,* 280**,** 6399-408.

REBETS, Y., DUTKO, L., OSTASH, B., LUZHETSKYY, A., KULACHKOVSKYY, O., YAMAGUCHI, T., NAKAMURA, T., BECHTHOLD, A. & FEDORENKO, V. 2008. Function of lanI in regulation of landomycin A biosynthesis in *Streptomyces cyanogenus* S136 and cross-complementation studies with *Streptomyces* antibiotic regulatory proteins encoding genes. *Arch Microbiol,* 189**,** 111-20.

REEVES, C. D., WARD, S. L., REVILL, W. P., SUZUKI, H., MARCUS, M., PETRAKOVSKY, O. V., MARQUEZ, S., FU, H., DONG, S. D. & KATZ, L. 2004. Production of hybrid 16-membered macrolides by expressing combinations of polyketide synthase genes in engineered *Streptomyces fradiae* hosts. *Chem Biol,* 11**,** 1465-72.

RICE, P., LONGDEN, I. & BLEASBY, A. 2000. EMBOSS: the European Molecular Biology Open Software Suite. *Trends Genet,* 16**,** 276-7.

RIX, U., FISCHER, C., REMSING, L. L. & ROHR, J. 2002. Modification of post-PKS tailoring steps through combinatorial biosynthesis. *Nat Prod Rep,* 19**,** 542-80.

ROHDICH, F., WIESE, A., FEICHT, R., SIMON, H. & BACHER, A. 2001. Enoate reductases of *Clostridia* - Cloning, sequencing, and expression. *Journal of Biological Chemistry,* 276**,** 5779-5787.

ROWE, C. J., BOHM, I. U., THOMAS, I. P., WILKINSON, B., RUDD, B. A., FOSTER, G., BLACKABY, A. P., SIDEBOTTOM, P. J., RODDIS, Y., BUSS, A. D., STAUNTON, J. & LEADLAY, P. F. 2001. Engineering a polyketide with a longer chain by insertion of an extra module into the erythromycin-producing polyketide synthase. *Chem Biol,* 8**,** 475-85.

RUTHERFORD, K., PARKHILL, J., CROOK, J., HORSNELL, T., RICE, P., RAJANDREAM, M. A. & BARRELL, B. 2000. Artemis: sequence visualization and annotation. *Bioinformatics,* 16**,** 944-5.

SAIKI, R. K., GELFAND, D. H., STOFFEL, S., SCHARF, S. J., HIGUCHI, R., HORN, G. T., MULLIS, K. B. & ERLICH, H. A. 1988. Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. *Science,* 239**,** 487-91.

SAIKI, R. K., SCHARF, S., FALOONA, F., MULLIS, K. B., HORN, G. T., ERLICH, H. A. & ARNHEIM, N. 1985. Enzymatic amplification of beta-globin genomic sequences and restriction site analysis for diagnosis of sickle cell anemia. *Science,* 230**,** 1350-4.

SALAH-BEY, K., DOUMITH, M., MICHEL, J. M., HAYDOCK, S., CORTES, J., LEADLAY, P. F. & RAYNAL, M. C. 1998. Targeted gene inactivation for the elucidation of deoxysugar biosynthesis in the erythromycin producer *Saccharopolyspora erythraea*. *Mol Gen Genet,* 257**,** 542-53.

SANGER, F., NICKLEN, S. & COULSON, A. R. 1977. DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci U S A,* 74**,** 5463-7.

SCHABERLE, T. F. & HACK, I. M. 2014. Overcoming the current deadlock in antibiotic research. *Trends in Microbiology,* 22**,** 165-167.

SCHNEIDER, G. 2005. Enzymes in the biosynthesis of aromatic polyketide antibiotics. *Curr Opin Struct Biol,* 15**,** 629-36.

SCHNEIDER, K., CHEN, X. H., VATER, J., FRANKE, P., NICHOLSON, G., BORRISS, R. & SUSSMUTH, R. D. 2007. Macrolactin is the polyketide biosynthesis product of the *pks2* cluster of *Bacillus amyloliquefaciens* FZB42. *J Nat Prod,* 70**,** 1417-23.

SCHRODER, J. 1999. Probing plant polyketide biosynthesis. *Nat Struct Biol,* 6**,** 714-6.

SEKUROVA, O. N., BRAUTASET, T., SLETTA, H., BORGOS, S. E., JAKOBSEN, M. O., ELLINGSEN, T. E., STROM, A. R., VALLA, S. & ZOTCHEV, S. B. 2004. In vivo analysis of the regulatory genes in the nystatin biosynthetic gene cluster of *Streptomyces noursei* ATCC 11455 reveals their differential control over antibiotic biosynthesis. *J Bacteriol,* 186**,** 1345-54.

SHIELDS, J. A., RAHMAN, A. S., ARTHUR, C. J., CROSBY, J., HOTHERSALL, J., SIMPSON, T. J. & THOMAS, C. M. 2010. Phosphopantetheinylation and specificity of acyl carrier proteins in the mupirocin biosynthetic cluster. *Chembiochem,* 11**,** 248-55.

SHIOZAWA, H., KAGASAKI, T., KINOSHITA, T., HARUYAMA, H., DOMON, H., UTSUI, Y., KODAMA, K. & TAKAHASHI, S. 1993. Thiomarinol, a new hybrid antimicrobial antibiotic produced by a marine bacterium. Fermentation, isolation, structure, and antimicrobial activity. *J Antibiot (Tokyo),* 46**,** 1834-42.

SHIOZAWA, H., KAGASAKI, T., TORIKATA, A., TANAKA, N., FUJIMOTO, K., HATA, T., FURUKAWA, Y. & TAKAHASHI, S. 1995. Thiomarinol-B and Thiomarinol-C, New Antimicrobial Antibiotics Produced by a Marine Bacterium. *Journal of Antibiotics,* 48**,** 907-909.

SHIOZAWA, H., SHIMADA, A. & TAKAHASHI, S. 1997. Thiomarinols D, E, F and G, new hybrid antimicrobial antibiotics produced by a marine bacterium; isolation, structure, and antimicrobial activity. *J Antibiot (Tokyo),* 50**,** 449-52.

SHIOZAWA, H. & TAKAHASHI, S. 1994. Configurational Studies on Thiomarinol. *Journal of Antibiotics,* 47**,** 851-853.

SIEBER, S. A., LINNE, U., HILLSON, N. J., ROCHE, E., WALSH, C. T. & MARAHIEL, M. A. 2002. Evidence for a monomeric structure of nonribosomal Peptide synthetases. *Chem Biol,* 9**,** 997-1008.

SILAKOWSKI, B., SCHAIRER, H. U., EHRET, H., KUNZE, B., WEINIG, S., NORDSIEK, G., BRANDT, P., BLOCKER, H., HOFLE, G., BEYER, S. & MULLER, R. 1999. New lessons for combinatorial biosynthesis from myxobacteria. The myxothiazol biosynthetic gene cluster of *Stigmatella aurantiaca* DW4/3-1. *J Biol Chem,* 274**,** 37391-9.

SILBY, M. W., WINSTANLEY, C., GODFREY, S. A., LEVY, S. B. & JACKSON, R. W. 2011. *Pseudomonas* genomes: diverse and adaptable. *FEMS Microbiol Rev,* 35**,** 652-80.

SIMON, R., PRIEFER, U. & PUHLER, A. 1983. A Broad Host Range Mobilization System for In Vivo Genetic Engineering: Transposon Mutagenesis in Gram Negative Bacteria. *Nat Biotech,* 1**,** 784-791.

SMITH, S. & TSAI, S. C. 2007. The type I fatty acid and polyketide synthases: a tale of two megasynthases. *Nat Prod Rep,* 24**,** 1041-72.

SMITH, T. F. & WATERMAN, M. S. 1981. Identification of common molecular subsequences. *J Mol Biol,* 147**,** 195-7.

STACHEL, S. E., AN, G., FLORES, C. & NESTER, E. W. 1985. A Tn3 *lacZ* transposon for the random generation of beta-galactosidase gene fusions: application to the analysis of gene expression in *Agrobacterium. EMBO J,* 4**,** 891-8.

STASSI, D., DONADIO, S., STAVER, M. J. & KATZ, L. 1993. Identification of a *Saccharopolyspora erythraea* gene required for the final hydroxylation step in erythromycin biosynthesis. *J Bacteriol,* 175**,** 182-9.

STASSI, D. L., KAKAVAS, S. J., REYNOLDS, K. A., GUNAWARDANA, G., SWANSON, S., ZEIDNER, D., JACKSON, M., LIU, H., BUKO, A. & KATZ, L. 1998. Ethyl-substituted erythromycin derivatives produced by directed metabolic engineering. *Proc Natl Acad Sci U S A,* 95**,** 7305-9.

STAUNTON, J. & WEISSMAN, K. J. 2001. Polyketide biosynthesis: a millennium review. *Nat Prod Rep,* 18**,** 380-416.

STAUNTON, J. & WILKINSON, B. 2001. Combinatorial biosynthesis of polyketides and nonribosomal peptides. *Curr Opin Chem Biol,* 5**,** 159-64.

STRIEKER, M., TANOVIC, A. & MARAHIEL, M. A. 2010. Nonribosomal peptide synthetases: structures and dynamics. *Curr Opin Struct Biol,* 20**,** 234-40.

SUMMERS, R. G., DONADIO, S., STAVER, M. J., WENDT-PIENKOWSKI, E., HUTCHINSON, C. R. & KATZ, L. 1997. Sequencing and mutagenesis of genes from the erythromycin biosynthetic gene cluster of *Saccharopolyspora erythraea* that are involved in L-mycarose and D-desosamine production. *Microbiology,* 143 (Pt 10)**,** 3251-62.

SUN, H., LIU, Z., ZHAO, H. & ANG, E. L. 2015. Recent advances in combinatorial biosynthesis for drug discovery. *Drug Des Devel Ther,* 9**,** 823-33.

SWAN, D. G., RODRIGUEZ, A. M., VILCHES, C., MENDEZ, C. & SALAS, J. A. 1994. Characterisation of a *Streptomyces antibioticus* gene encoding a type I polyketide synthase which has an unusual coding sequence. *Mol Gen Genet,* 242**,** 358-62.

THOMAS, C. M. 2015. Unpublished data. The University of Birmingham, UK.

THOMAS, C. M., HOTHERSALL, J., WILLIS, C. L. & SIMPSON, T. J. 2010. Resistance to and synthesis of the antibiotic mupirocin. *Nat Rev Microbiol,* 8**,** 281-9.

THOMAS, C. M. & NIELSEN, K. M. 2005. Mechanisms of, and barriers to, horizontal gene transfer between bacteria. *Nature Reviews Microbiology,* 3**,** 711-721.

TILL, M. & RACE, P. R. 2014. Progress challenges and opportunities for the re-engineering of *trans*-AT polyketide synthases. *Biotechnol Lett,* 36**,** 877-88.

TILL, M. & RACE, P. R. 2016. The Assembly Line Enzymology of Polyketide Biosynthesis. *In:* EVANS, B. S. (ed.) *Nonribosomal Peptide and Polyketide Biosynthesis: Methods and Protocols.* New York, NY: Springer New York.

TIWARI, K., RAJ, V. S., UPADHYAY, D. J. & GUPTA, R. K. 2015. In vitro activity of bioactive extracts from rare actinomycetes against multi-drug resistant *Streptococcus pneumoniae. J Appl Microbiol,* 118**,** 1306-14.

TSUJI, S. Y., CANE, D. E. & KHOSLA, C. 2001a. Selective protein-protein interactions direct channeling of intermediates between polyketide synthase modules. *Biochemistry,* 40**,** 2326-31.

TSUJI, S. Y., WU, N. & KHOSLA, C. 2001b. Intermodular communication in polyketide synthases: comparing the role of protein-protein interactions to those in other multidomain proteins. *Biochemistry,* 40**,** 2317-25.

UCHIYAMA, H. & WEISBLUM, B. 1985. N-Methyl transferase of *Streptomyces erythraeus* that confers resistance to the macrolide-lincosamide-streptogramin B antibiotics: amino acid sequence and its homology to cognate R-factor enzymes from pathogenic bacilli and cocci. *Gene,* 38**,** 103-10.

URLACHER, V. B., LUTZ-WAHL, S. & SCHMID, R. D. 2004. Microbial P450 enzymes in biotechnology. *Applied Microbiology and Biotechnology,* 64**,** 317-325.

VENTOLA, C. L. 2015. The antibiotic resistance crisis: part 1: causes and threats. *P T,* 40**,** 277-83.
WALSH, C. 2003. Where will new antibiotics come from? *Nat Rev Microbiol,* 1**,** 65-70.
WALSH, C. T. & FISCHBACH, M. A. 2010. Natural products version 2.0: connecting genes to molecules. *J Am Chem Soc,* 132**,** 2469-93.
WANG, P., YU, Z., LI, B., CAI, X., ZENG, Z., CHEN, X. & WANG, X. 2015. Development of an efficient conjugation-based genetic manipulation system for *Pseudoalteromonas. Microb Cell Fact,* 14**,** 11.
WEBBER, M. A. & PIDDOCK, L. J. 2003. The importance of efflux pumps in bacterial antibiotic resistance. *J Antimicrob Chemother,* 51**,** 9-11.
WEBER, J. M., LEUNG, J. O., SWANSON, S. J., IDLER, K. B. & MCALPINE, J. B. 1991. An erythromycin derivative produced by targeted gene disruption in *Saccharopolyspora erythraea. Science,* 252**,** 114-7.
WEBER, J. M., SCHONER, B. & LOSICK, R. 1989. Identification of a gene required for the terminal step in erythromycin A biosynthesis in *Saccharopolyspora erythraea* (*Streptomyces erythreus*). *Gene,* 75**,** 235-41.
WEBER, T., LAIPLE, K. J., PROSS, E. K., TEXTOR, A., GROND, S., WELZEL, K., PELZER, S., VENTE, A. & WOHLLEBEN, W. 2008. Molecular analysis of the kirromycin biosynthetic gene cluster revealed beta-alanine as precursor of the pyridone moiety. *Chem Biol,* 15**,** 175-88.
WEISSMAN, K. J. 2004. Polyketide biosynthesis: understanding and exploiting modularity. *Philos Transact A Math Phys Eng Sci,* 362**,** 2671-90.
WEISSMAN, K. J. 2006. The structural basis for docking in modular polyketide biosynthesis. *Chembiochem,* 7**,** 485-94.
WEISSMAN, K. J. 2015. Uncovering the structures of modular polyketide synthases. *Natural Product Reports,* 32**,** 436-453.
WEISSMAN, K. J., HONG, H., POPOVIC, B. & MEERSMAN, F. 2006. Evidence for a protein-protein interaction motif on an acyl carrier protein domain from a modular polyketide synthase. *Chem Biol,* 13**,** 625-36.
WEISSMAN, K. J. & LEADLAY, P. F. 2005. Combinatorial biosynthesis of reduced polyketides. *Nat Rev Microbiol,* 3**,** 925-36.
WEISSMAN, K. J. & MULLER, R. 2008. Protein-protein interactions in multienzyme megasynthetases. *Chembiochem,* 9**,** 826-48.
WEYMOUTH-WILSON, A. C. 1997. The role of carbohydrates in biologically active natural products. *Nat Prod Rep,* 14**,** 99-110.
WHATLING, C. A., HODGSON, J. E., BURNHAM, M. K. R., CLARKE, N. J., FRANKLIN, F. C. H. & THOMAS, C. M. 1995. Identification of a 60 kb region of the chromosome of *Pseudomonas fluorescens* NCIB 10586 required for the biosynthesis of pseudomonic acid (mupirocin). *Microbiology,* 141**,** 973-982.
WILLIAMS, G. J. 2013. Engineering polyketide synthases and nonribosomal peptide synthetases. *Current Opinion in Structural Biology,* 23**,** 603-612.
WONG, F. T. & KHOSLA, C. 2012. Combinatorial biosynthesis of polyketides--a perspective. *Curr Opin Chem Biol,* 16**,** 117-23.
WOOD, J. M. 2015. Bacterial responses to osmotic challenges. *J Gen Physiol,* 145**,** 381-8.
WU, J., COOPER, S. M., COX, R. J., CROSBY, J., CRUMP, M. P., HOTHERSALL, J., SIMPSON, T. J., THOMAS, C. M. & WILLIS, C. L. 2007. Mupirocin H, a novel metabolite resulting from mutation of the HMG-CoA synthase analogue, mupH in *Pseudomonas fluorescens*. *Chem Commun (Camb)***,** 2040-2.

WU, J., HOTHERSALL, J., MAZZETTI, C., O'CONNELL, Y., SHIELDS, J. A., RAHMAN, A. S., COX, R. J., CROSBY, J., SIMPSON, T. J., THOMAS, C. M. & WILLIS, C. L. 2008. In vivo mutational analysis of the mupirocin gene cluster reveals labile points in the biosynthetic pathway: the "leaky hosepipe" mechanism. *Chembiochem,* 9**,** 1500-8.

WU, K., CHUNG, L., REVILL, W. P., KATZ, L. & REEVES, C. D. 2000. The FK520 gene cluster of *Streptomyces hygroscopicus* var. ascomyceticus (ATCC 14891) contains genes for biosynthesis of unusual polyketide extender units. *Gene,* 251**,** 81-90.

WU, N., CANE, D. E. & KHOSLA, C. 2002. Quantitative analysis of the relative contributions of donor acyl carrier proteins, acceptor ketosynthases, and linker regions to intermodular transfer of intermediates in hybrid polyketide synthases. *Biochemistry,* 41**,** 5056-66.

WU, N., TSUJI, S. Y., CANE, D. E. & KHOSLA, C. 2001. Assessing the balance between protein-protein interactions and enzyme-substrate interactions in the channeling of intermediates between polyketide synthase modules. *J Am Chem Soc,* 123**,** 6465-74.

YANAGISAWA, T. & KAWAKAMI, M. 2003. How does *Pseudomonas fluorescens* avoid suicide from its antibiotic pseudomonic acid?: Evidence for two evolutionarily distinct isoleucyl-tRNA synthetases conferring self-defense. *J Biol Chem,* 278**,** 25887-94.

YANAGISAWA, T., LEE, J. T., WU, H. C. & KAWAKAMI, M. 1994. Relationship of protein structure of isoleucyl-tRNA synthetase with pseudomonic acid resistance of *Escherichia coli*. A proposed mode of action of pseudomonic acid as an inhibitor of isoleucyl-tRNA synthetase. *J Biol Chem,* 269**,** 24304-9.

YU, M., TANG, K., LIU, J., SHI, X., GULDER, T. A. & ZHANG, X. H. 2013. Genome analysis of *Pseudoalteromonas flavipulchra* JG1 reveals various survival advantages in marine environment. *BMC Genomics,* 14**,** 707.

YU, Z. C., ZHAO, D. L., RAN, L. Y., MI, Z. H., WU, Z. Y., PANG, X., ZHANG, X. Y., SU, H. N., SHI, M., SONG, X. Y., XIE, B. B., QIN, Q. L., ZHOU, B. C., CHEN, X. L. & ZHANG, Y. Z. 2014. Development of a genetic system for the deep-sea psychrophilic bacterium *Pseudoalteromonas sp.* SM9913. *Microb Cell Fact,* 13**,** 13.

ZHANG, G., LI, Y., FANG, L. & PFEIFER, B. A. 2015. Tailoring pathway modularity in the biosynthesis of erythromycin analogs heterologously engineered in *E. coli*. *Science Advances,* 1.

ZHANG, J., VAN LANEN, S. G., JU, J., LIU, W., DORRESTEIN, P. C., LI, W., KELLEHER, N. L. & SHEN, B. 2008. A phosphopantetheinylating polyketide synthase producing a linear polyene to initiate enediyne antitumor antibiotic biosynthesis. *Proceedings of the National Academy of Sciences,* 105**,** 1460-1465.

ZHOU, H., XIE, X. & TANG, Y. 2008. Engineering natural products using combinatorial biosynthesis and biocatalysis. *Curr Opin Biotechnol,* 19**,** 590-6.