DISSECTING ROLE OF TYPE III POLYKETIDE SYNTHASE, METHYLTRANSFERASE AND SULFOTRANSFERASE IN THE BIOLOGY OF CORYNEBACTERINEAE

Dissertation submitted to South Asian University in partial fulfilment of the requirements for the award of the degree of

Doctor of Philosophy

By

Gorkha Raj Giri

Faculty of Life sciences and Biotechnology

South Asian University

New Delhi

2020

Declaration

I hereby declare that the dissertation entitled "**Dissecting Role of type III polyketide synthase, methyltransferase and sulfotransferase in the biology of corynebacterineae"** being submitted to the South Asian University, New Delhi, in partial fulfilment of the requirements for the award of the degree of **Doctor of Philosophy** is my original work carried out under the guidance of my supervisors at the Faculty of Life Sciences & Biotechnology, South Asian University. The research work reported in this dissertation has not been submitted either in part or full to any other university or institution for the award of any degree or diploma.

Gorkha Raj Giri

Enrollment No.: SAU/BIO(P)/2014/08

SOUTH ASIAN UNIVERSITY

(A University established by SAARC Nations)
Akbar Bhawan, Chanakyapuri, New Delhi – 110 021, India
Tel.: 91-11-2412 2512-14 Fax: 91-11-2412 2511
Website: www.southasianuniversity.org

Certificate

I/We have gone through the PhD dissertation entitled "**Dissecting role of type III polyketide synthase, methyltransferase and sulfotransferase in the biology of corynebacterineae**" of **Mr. Gorkha Raj Giri**, and I/We recommend that this dissertation be placed before the examiners for evaluation.

21/05/2020

Dr. PRITI SAXENA
Assistant Professor
Faculty of Life Sciences & Biotechnology
South Asian University
New Delhi-110 021

Dean
Dr. Senthil K. Venugopal
Professor and Dean
Faculty of Life Sciences
& Biotechnology
South Asian University
New Delhi

Supervisor
Dr. Priti Saxena
Assistant Professor (SG)
Faculty of Life Sciences
& Biotechnology
South Asian University
New Delhi

Acknowledgements

I would like to thank my supervisors, Dr Priti Saxena, for being so generous with her time, for her invaluable guidance and support throughout the duration of my PhD. Despite some stumbles and falls her continued positivity and trust made this project run so efficiently. I am very grateful to our project collaborator Associate Professor Dr Vengadesan Krishnan, from Regional Centre for Biotechnology, NCR Biotech Science Cluster, Hariyana (NCR Delhi) for his valuable help in crystallography and related work. I also like to thank Professor Y. Singh, IGIB, India for providing *Mycobacterium marinum* (strain ATCC BAA-535/M) and Dr Rajesh S. Gokhale, NII for *M. smegamtis* Δpks10 knockout strain.

Similarly, I am obliged to Dr Archana Singh, IGIB for Transmission Electron Microscopy sample processing and imaging, Mrs Hemlata Gautam for technical support and imaging of Scanning Electron Microscopy sample. I am delightful to Professor Lakshmaiah Sreerama, Qatar University, for his guidance in sample analysis and Professor Kambadur Muralidhar for giving me research insight and encouragement. I also want to thank, Ankit Srivastava from IIT Delhi, Shivendra Pratap and Abhiruchi Kant, from RCB for technical help respectively in Atomic Force Microscopy and crystallography related work. I am thankful all of my labmates from Chemical Biology Group Monica Vardhan, Sameer Giri, Dr Amreesh Parvez, Miss Renu Bisht, Umesh Hatti for their generous help and support. I am also thankful to our lab Msc student Anjali, Apoorva, Satish, Dipu, Pankhuri, Deepak, Minhazur (Minaz), Gaurav and Upasana for making lab with an excellent atmosphere to work together during my PhD tenure.

I express my sincere gratitude to my parents Mrs Saraswoti Giri and Mr Brish Lal Giri, for their care and love. I also like to thank my late grandparents for providing me love, wisdom for life. I am also incredibly grateful to my siblings Mr Mahesh Kumar Giri, Mr Manoj Kumar Giri, Miss Usha Giri for their support and help and encouragement in the endeavour of my life. I also like to thank my sister-in-law Mrs Sunita Tiwari, Mrs Shanti Giri for their love and good wishes.

I am thankful to all of my batchmates from 2014 batch Roshan Kumar Dutta, Rakesh Sharma, Faiz Ahmad, Ravi Dubey, Dr Parul Gupta, Dr Muskan Jain and Deepika Bhardwaj all of them for their scientific aptitude, humour, laughs and encouragement making PhD with such a great experience of life. My incredible gratitude also goes to all of my seniors of 2013 batch and specially to elder brother Dr Uddhav Timilsina, Dr Kishor Pant and Dr Nawneet Mishra and Dr Shivendra Lal Karn for their suggestion and help in shaping my research. I am also very much thankful to my roommate Ajay Kumar Yadav, Rabindra Mahato, Srimal Farnendo, Mohiuddin Al-Amin for making hostel room as home and caring me. Besides I like to thank remarkable people and upcoming scientist including Dibya Ghimire, Preeti Nagar, Bilal Habib and Muhammad Hussain, Nerina Shahi and Amrendra Sah for their friendliness, help and concern to me. I also like to give heartful thanks to all other PhD scholars and masters students of CIF who make lab as a homely environment during my PhD. Also, I would like to acknowledge all the faculty members and specially Dean Prof. Dr Senthil K. Venugopal of the Faculty of Life Science and Biotechnology for their unswerving support and help during my PhD work. I thank technical staff members including Mr Nitin Sharma, Mr Milind Dongardrive for their technical support in fixing instruments and Mr Jagdish Yadav, Mr. Tarun, Satyaram Jee and Mohd. Izrail for their coordination and lending their hand throughout my PhD. I also want to thank faculty assistant Mrs Sree Devi and Mrs Simple Kumari for their assistance in office work related to me. Lastly, I would also like to acknowledge the South Asian University, New Delhi, for providing me PhD scholarship and to my supervisor with Start-up grant for pursuing research. I am also thankful to the Department of Science and Technology and Department of Biotechnology, Government of India for SERB grant, IYBA grant respectively to my supervisor.

Gorkha Raj Giri

Table of contents

Abbreviations

2PS	2-pyrone synthase
Å	angstrom
ACN	Acetonitrile
ACP	acyl carrier protein
AdoMet	*S*-adenosyl *L*-methionine dependent
ASST	arylsulfate sulfotransferase, or aryl sulfotransferase
ATP	Adenosine 5'-triphosphate
APS	ammonium per sulphate
BAS	benzalacetone synthase
bp	base pair
BSA	bovine serum albumin
*Bc*OMT	*Bacillus cereus O*-methyltransferase
CCoAOMT	caffeoyl coenzyme A 3-*O*-methyltransferase
CHS	chalcone synthase
CMTs/C-MTs	C-methyltransferases
CoA	Coenzyme A
COMT	catechol *O*-methyltransferase
CTAS	*p*-coumaroyltriacetic acid synthase
CUS	curcuminoid synthase
dNTP	deoxy nucleotide triphosphate
DTT	dithiothreitol
ER	enoyl reductase
EDTA	ethylene diamine tetraacetic acid
ESI-MS	electrospray ionization mass spectrometry
FASs	fatty acid synthases
HPLC	high performance liquid chromatography
hr	hour
HRMS	High-Resolution Mass Spectrometer

ICMTs	isoprenylcysteine carboxyl methyltransferases
IDA	information-dependent acquisition
IPTG	isopropyl-1-thio-*β*-D-thiogalactopyranoside
Kb	kilobase
kDa	kilo Dalton
KR	ketoreductase
KS	ketosynthase
L	litre
LB	Luria Bertini medium
LPS	Lipopolysaccharide
M	Molar
M. sativa	*Medicago sativa*
mAGP	mycolylarabonogalactan peptidoglycan
MCoA	malonyl Coenzyme A
µg	microgram
mg	milligram
min	minute
ml	milliliter
µl	microliter
µM	micromolar
mM	millimolar
Mmar/MMAR	*Mycobacterium(M.) marinum*
MCoA	malonyl Coenzyme A
MMCoA	methylmalonyl Coenzyme A
MRM	multiple-reaction-monitoring
msCHS	chalcone synthase from *Medicago sativa*
Msmeg/MSMEG	*Mycobacterium(M.) smegmatis*
MTase	methyltransferase
Mtb	*Mycobacterium(M.) tuberculosis*
MWCO	molecular weight cut off
Ni^{2+}-NTA	Nickel Nitrilo-Triacetic Acid

NMTs/N-MTs	N-methyltransferases
NRPSs	non-ribosomal peptide synthases (NRPs)
OKS	octaketide synthase
OD	optical density
OMTs	*O*-methyltransferases
PAP	3'-Phosphoadenosine-5'-phosphate
PAPS	3'-Phosphoadenosine-5'-phosphosulfate
PCR	Polymerase Chain Reaction
PCS	pentaketide chromone synthase
PEG	polyethylene glycol
PKSs	polyketide synthases
pmol	picomole
PSB	phosphosulfate binding
RMSD	root mean square deviation
RNase A	ribonuclease A
rpm	revolution per minute
SAH	S-adenosyl *L*-homocysteine
SAM	*S*-adenosyl *L*-methionine
SAM-MT	*S*-adenosylmethionine-dependent methyltransferase
SDS-PAGE electrophoresis	sodium dodecyl sulfate-polyacrylamide gel
sec	seconds
SMTs	S-methyltransferases
SPS	styrylpyrone synthase
STS	stilbene synthase
STS	stilbene synthase
SULTs	sulfotransferases
TAL	triacetic acid lactone
TB	Tuberculosis
TCEP	tris (2-carboxyethyl) phosphine
TDM	treehalose dimycolate

TEMED	N, N,N’,N’-tetra methyl ethylene diamine
THN	1,3,6,8- tetrahydroxynaphthalene
THNS	1,3,6,8- tetrahydroxynaphthalene synthase
TMM	Treehalose monomycolate
TOF-MS	Time of flight mass spectrometry
WHO	World Health Organization

ABSTRACT

The present thesis provides a thorough insight into type III polyketide synthases (PKSs) from microorganisms belonging to sub-order corynebacterineae. Type III PKSs are versatile enzymes making variable products with different bioactivities. *M. marinum* (Mmar) has four kinds of type III polyketide synthases (PKSs) belonging to three types of clusters. The PKSs produce a spectrum of polyketides and polyphenolics. MMAR_2470 and MMAR_2474 belong to the first cluster. The orthologous genes are found to be present in pathogenic bacteria from 27 species of mycobacteria. Whereas, MMAR_2190, from the second cluster, belongs to nine genomes of pathogenic corynebacterineae. Our biochemical characterization of MMAR_2190 revealed unprecedented cyclization potential of an enzyme to form four classes of non-methylated products using a common substrate pool through a single catalytic site. The structural study of protein depicts the presence of three active site residues cysteine (Cys), Histidine (His) and Asparagine (Asn) similar to other type III PKSs but with unique features, i.e. malonate bound to active site Cys and constriction in the active site cavity. It provides a continuous channel for the access of substrate, which is further regulated through the orientation of gatekeeper residues. Leu266, Leu348, Ile319 and Leu223 regulate the binding of substrate to the substrate-binding tunnel, and different catalytic pockets orient type III PKS by the substrate. We further attempted to identify the physiological roles through over-expression of different type III PKSs from Mmar and NFIII, an ortholog of MMAR_2190 from *N. farcinica* in Δpks10 *M.* smegmatis strain. This type III PKS knockout strain shows defect in biofilm formation. An acetamide inducible pMyNT vector was utilized to over-express different type III PKSs and generate different Δpks10 Msmeg/MyNT-type III PKS strains. These over-expressing strains displayed interesting

physiological features. The Δpks10 Msmeg/*pMyNT_MMAR_2190* strain showed complete biofilm recovery, similar to wildtype *M. smegmatis* and *M. marium* biofilms. Whereas, Δpks10 Msmeg/*pMyNT_MMAR_2474* showed partial recovery of biofilm. Δpks10 Msmeg/*pMyNT_MMAR_2470* and Δpks10 Msmeg/*pMyNT_NFIII* strains could not revive biofilm growth. Scanning Electron Microscopy (SEM) also showed thick interconnecting pigmented cells in the biofilm of Δpks10Msmeg/*pMyNT_MMAR_2190* strain. This study showed metabolic diversity of polyketide molecules in over-expressed strains along with the presence of a series of quinone like molecules in over-expressed pMyNT strains. Possibly, the polyketide molecules help in remodelling the cell wall as inferred from the consistent cell envelope observed through the Transmission Electron Microscopy (TEM) of planktonic cultures of pMyNT_*MMAR_2190*. We found acyl phloroglucinols in MMAR_2190 overexpressed mycobacterial strain. We observed polyketide modification by two groups of enzymes, methyltransferase and sulfotransferase from MMAR_2190 cluster. *O*-methylation seemed essential in biofilm formation and was also observed in a biofilm of wildtype *M. marinum*. MMAR_2190 genomic cluster contains a methyltransferase (MMAR_2193) and two sulfotransferases (MMAR_2191 and MMAR_2192). Our biochemical analysis of these Mmar proteins revealed unprecedented potential of the enzymes to modify all available hydroxyl groups in polyphenolic compounds. In an interesting experiment, type III polyketide products produced by cognate type III PKSs could be directly methylated by MMAR_2193 to generate multi-methylated polyketides. Collectively, these studies implicate the genomic cluster of type III PKSs to aid mycobacterial growth and survival. Additionally, this thesis unravels novel biosynthetic potentials of the Mmar type III PKSs and modifying enzymes that could be utilized for production of divergent chemical scaffolds with unique bioactive potentials.

Introduction

INTRODUCTION

Corynebacterial disease diagnosis has been challenging in many pathogenic bacteria by thick cell envelope consisting of macromolecules and small molecules as constituents. Type III polyketides and polyphenolic compounds are small secondary metabolites, synthesized through an acyl-coenzyme A (CoA) as building blocks with elongation through extender-like malonyl-CoA using type III polyketide synthases (PKSs) enzyme-based machinery. The machinery includes multiple-rounds of condensation (Claisen and aldol) or lactonization[1-4] to form diverse cyclic compounds[1,5,6] or polyketide natural products[7] with varied functions[8-13]. Our study deals with the mechanistic potential of type III polyketide synthases of 40 to 45 kDa homodimeric enzymes from *Mycobacterium marinum* (Mmar) and *Nocardia farcinica*. Both organisms belong to sub-order corynebacterineae of phylum actinomycetes to which pathogenic bacteria including *M. tuberculosis*, *M. leprae*, and *Corynebacterium diphththeriae* belong. Thesis reviews the structural aspects of different type III PKSs from plants, fungi, algae, slime molds and bacteria with respect to the generation of polyketide molecules like chalcones, stilbene, α-pyrones, resorcinols and phloroglucinols. The thesis describes the biochemical and structural characterization of one of a type III PKS, MMAR_2190 from Mmar to form diverse polyketide compounds along with roles of two classes of modifying enzymes, a methyltransferase and two sulfotransferases to give *O*-methylated and *O*-sulfated derivatives. It also dissects the physiological relevance of different type III PKSs coded by the mycobacterial genes- *MMAR_2190*, *MMAR_2470*, *MMAR_2474* and nocardial *NFIII* by overexpression studies in Δpks10 knockout strain of *M. smegmatis*. Planktonic culture

and biofilm culture-based studies of overexpressed and control strains are used to find the roles of different genes with respect to wildtype mycobacterial strains (*M. marinum and M. smegmatis*). Metabolomics and advanced microscopy techniques investigated possible roles of genes in the regeneration of biofilm, change cell envelope, and production of bioactive molecules of unexplored functions.

Chapter 1

Review of Literature

CHAPTER 1: REVIEW OF LITERATURE

Mycobacterial cell envelope:

Mycobacterium tuberculosis is the deadliest pathogen of humans and cattle, causing public health problems with the disease tuberculosis (TB) (World Health Organization [WHO], 2015). According to WHO, annual report 2019, approximately 10 million people fell ill, and 1.2 million died of TB in the year 2018. Bacteria ability to survive in the adverse condition is due to a highly complex and well-organized cell envelope consisting of peptidoglycan covalently linked to mycolylarabinogalactan with phosphodiester bond. Most mycobacterial cell envelope consists of three covalently linked macromolecules viz. peptidoglycan, arabinogalactan, and mycolic acids forming mycolylarabinogalactan peptidoglycan (mAGP) and liporabinomannnan embedded to the framework of the mAGP. Peptidoglycan consists of alternating units of N-acetylglucosamine and N-glyolylmuramic acid. Peptidoglycan is crosslinked with arabinogalactan through the tetrapeptide side chain (L-alaninyl-isoglutaminyl-meso-diaminopimelyl-D-alanyl, with diaminopimelic acid further amidated) attached to muramic acid. Arabinogalactan comprises D-galactofuranoses and D-arabinofuranoses exclusively. Mycolic acids are long-chain (C_{70} and C_{90}), α-alkyl branched or β-hydroxylated and located terminal to pentarabinofuranosyl unit. Mycolic acid attached to cell wall arabinogalactan and organized with other lipids. It forms an outer permeability barrier with low fluidity and acts as a screen to resist the antibiotics and trehalose mycolate function in mycobacterial physiology and virulence[14,15]. Slow growing mycobacteria (*M. tuberculosis*), mycolic acid undergoes cyclopropanation, whereas rapid growing (*M. smegmaits*) do not. Mycolates also undergo epoxidation and

methylation to form epoxymycolate and methoxymycolate and also occur in the form of free trehalose dimycolate (TDM, cord factor) and trehalose monomycolate (TMM) both occur on the fluid matrix of the cell surface of bacteria. The enzyme involved in the modification of mycolate such as Hma (MmaA4), an S-adenosylmethionine-dependent methyltransferase (SAM-MT) leads to the formation of keto- or methoxymycolates in *M. tuberculosis*[16]. Different species of mycobacteria including *M. tuberculosis*, *M. smegmatis* and *M. marinum* shared conserved domain predicting for cell wall-associated proteins, regulatory proteins, conserved hypothetical proteins and enzymes involved in intermediatory metabolism and lipid metabolism. The 85 % amino acid identity of *M. marinum* with *M. tuberculosis* and 76 % with *M. smegmatis* make *M. marinum* as the best model understand the fish tuberculosis and better link with human disease. The overexpression of genes from the pathogenic cluster into *M. smegmatis* growing in BSL-2 and with less generation time usually 4 h compared with more than 20 h in *M. tuberculosis* make the model convenient to study the deadliest pathogen. It is interesting to consider surface-specific glycolipids, lipooligosaccharides and phenolic glycolipids. Understanding these molecules that vary with species of pathogens can be useful to better deal with toxin and understand better the pathogenesis. The illustrative architecture of the mycobacterial cell envelope is shown in figure 1.

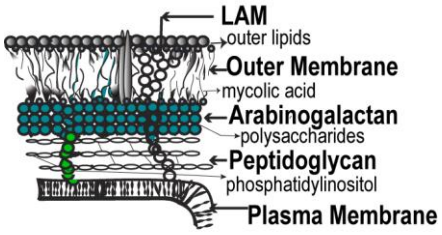


Figure 1.1: Mycobacterial cell envelope model with principal components mycolylarabinogalactan peptidoglycan (mAGP) and lipoarabinomannan. Mycobacterial cell wall with mAGP and LAM are the two principal constituents surrounded by capsular-like substances rich in extracellular polysaccharides usually consisting glycogen (glucan), arabinomannan and

mannan[17]. The surface glycolipids include a variety of species and strain-specific glycopeptidolipids, lipooligosaccharides, and aromatic polyphenolic derivatives (termed "polyphenolics"), the chemical identity and amount of which varies from one species to another.

1.1. Polyketides and polyphenolics

Polyketides are the secondary metabolites with multiple β-hydroxy-ketone or in some reduced to β-hydroxy-aldehyde functional group. Polyketides have diverse structures and functions[2,18]. Polyketides are found commonly in bacteria[19], fungi[20], lichen[21], and plants[22]. Protists[23], insects[24], molluscs[25], and sponges[26] produce polyketides with different bioactivities either through the symbiotic association with bacteria or fungi. Polyketides have diverse structures and share a typical pattern of biosynthesis similar to that of fatty acid synthase as hypothesized by Birch and Denovan in 1953[27]. In 1806 John Norman Collie, reported that orcinol could be directly synthesized using dehydroacetic acid and the term polyketides as the oligomer of C2 acetate and polyphenols biosynthesis occur using C2 precursors[28] using multiple keten groups came to the chemical literature in 1907. The remaining of the alternate keto group during the repetitive condensation gives rise to the name "polyketide", for the family of secondary metabolites. Polyketides are synthesized by polyketide synthase. The synthesis of diverse compounds occurs through (C2)n linear intermediates, a highly reactive poly β-ketoacyl intermediate, H-(CH2C=O)n-OH by the successive decarboxylative Claisen condensation of coenzyme A-derived units to the multi-carbon compound containing keto or hydroxyl groups[29-31]. The group comprises of polyphenols, macrolides, polyenes, enediynes, and polyethers.

Although the exact role of polyketides in their original biological context is unknown, they are believed to function in imparting pigments to flowers, acting as virulence factors in microorganisms, playing a role as info chemicals or signals and have a role in defences. The group of secondary metabolites are also widely used as therapeutic agents. Polyketides are used as antibiotics, immunosuppressants, antiparasitics, cholesterol-lowering, and

antitumoral agents. The biomedical significance of polyketide is increased with the selection of polyketide or polyketide based synthetic products as a "List of Essential Medicines by the World Health Organization" (For WHO Model List of Essential Medicines, on October 22, 2015).

http://www.who.int/medicines/publications/essentialmedicines/en/

Galvanolide (potent against wheat stem rust fungus *Puccinia graminis, Botrytis cinerea* and other fungal phytopathogens and also inhibit the growth of human opportunistic fungal pathogen *Cryptococcus neoformans*)[32], erythromycin, daunorubicin, and lovastatin are some of the clinically important drugs synthesized using polyketide synthase pathway.

Some of the polyketides, like afaltoxins[33] and maitotoxin[34], have a detrimental effect on the health of human beings and fish upon chronic exposure. Mycolactone, a macrolide toxin has a polyketide side chain is a hallmark of pathogenesis in *M. ulcerans i*nfection-causing Buruli ulcer[35].

1.2. Polyketide synthase

Polyketide synthases (PKSs) are the multifunctional enzymes catalyze the synthesis of polyketide. Biosynthesis of polyketide is similar to fatty acid synthase (FAS) that catalyzes an attachment to the substrate through acyltransferase (acetyl or malonyl) to the acyl carrier protein (ACP), ketosynthase (KS) and undergoes the condensation of substrate attached in ACP. The polyketide intermediate is subsequently processed by ketoreductase (KR), dehydratase (DH) and enoyl reductase (ER). The latter three steps are optional in polyketide synthase[12,36,37]. PKSs and FASs protein structure and mechanism of biosynthesis of products are similar. Both undergo decarboxylative condensation of simple acetyl-CoA thioesters to form various products. Based on the protein architecture and

function, PKSs are categorised into three types: type I, II and III. Type I and II are multifunctional proteins, whereas type III are small homodimeric protein[38-41].

1.3. Type I PKSs and type II PKSs

Type I polyketide synthases involve multifunctional catalytic domains organized with modular with many modules. Each module undergoes single decarboxylation reactions through ketosynthase domain, add many extender substrates to form polyketide[42-45]. Iterative with single ketosynthase domain as a functional domain is clustered in a single module that repeatedly tethered to an acyl carrier protein (ACP) as an anchor for several times of extension to form product[44-46]. Whereas, type II PKSs consist of mono-functional polypeptides a "minimal PKS" and with the auxiliary subunit bearing a single set of active sites used repeatedly with an anchor acyl carrier protein to join to nascent polypeptide[47]. It has two ketosynthases units: KSα and KSβ, catalyze the condensation of precursors and act as chain determining factors respectively[43]. Further, the poly β-keto chain is transformed into the aromatic core by modifying enzymes-like ketoreductases, cyclases, and aromatases. The products from different reactions are acted upon by post- tailoring enzyme-like oxygenases, glycosyl, and methyltransferases to give rise to functional aromatic polyketide[48-50].

1.4. Type III PKSs

Type III PKSs are the homodimeric ketosynthases[1] which catalyze repeated chain elongation using the acyl-CoAs linked starter unit (usually an aromatic CoA) and acetyl units (derived from malonyl-CoA) to generate a vast number of polyketide products through repetitive priming, extension, and cyclization reactions using same active site cavity. Type III PKSs synthesize aromatic compounds with wide function and bioactivities in both plants and microorganisms[1]. Type III PKSs were exclusively studied in the plant

kingdom for several decades and are known as Chalcone/stilbene synthases/*p*-coumaroyltriacetic acid synthase (CHS/STS/CTAS). The first type III PKS from the plant was discovered in 1970s[1]. In the plant, the enzyme uses phenylpropanoid substrates (such as *p*-coumaroyl-CoA) assembling the chain with extender (e.g. malonyl-CoA) to synthesize flavonoid precursors[1] (e.g. naringenin chalcone, resveratrol and *p*-coumaroyltriacetic acid lactone from thioester linked tetraketide intermediate by regiospecific C1 to C6 Claisen condensation, C2 to C7 Aldol condensation and intramolecular C5 oxygen to C1 lactonization respectively). Wide ranges of functionally divergent CHS like type III PKSs are present in plants. The various polyketide products play roles as signalling molecules between plant and microbes, antimicrobials agents and feeding deterrents, male fertility of some species, protecting from UV damage[51]. In-plant, CHS cooperates with CHS like proteins such as benzalacetone synthases (BAS), styrylpyrone synthase (SPS), 4-coumaroyl triacetic acid synthase (CTAS), stilbene synthase (STS), acridone synthase, and valerophenone synthase and other synthetic enzymes such as chalcone flavanone isomerase giving a flavonoid.

The products from CHS family of proteins undergo a tailoring reaction at multiple positions. Some of the modifications are hydroxylation, *O*-methylation, glycosylation, acylation, prenylation and conjugation. The tailoring modifications are important for the biosynthesis of more than 6000 naturally occurring flavonoids with diverse functions[1]. Similarly, octaketide synthase (OCS) from *Aloe arborescens* found to synthesize a wide range of products including SEK4 and SEK4b, unnatural C21 heptaketide chalcone and C19 hexaketide stilbene, C18 heptaketide phloroglucinol and C16 hexaketide resorcinol due to broad substrate specificity[52]. Further, pentaketide chromone synthase, a type III PKS from the same plant produces 5,7-dihydroxy-2-methylchromone, a pentaketide chromone using five molecules of malonyl-CoA as starter and extender[53].

In bacteria, type III PKS were discovered with the study of RppA, a chalcone synthase-like protein exclusively present in the Streptomyces species (*S. griseus*) and other bacteria[2] and the gene conferred red-brown pigment production. Diversity of the polyketide results from programmed events like a condensation of a pool of starter and extender unit to the PKS forming the backbone[29]. The variety of the polyketide products depends on the choice of the chain length, a number of biosynthetic building blocks used[40] and the cavity volume of the enzyme.[54] The backbone is an intermediate commonly in the form of triketide, tetraketide, i.e. dictated by β-keto processing reactions, the stereochemistry of intermediate, alkylations or branching reactions as well as other *in situ* reactions. The intermediate undergoes lactonization, hydrolysis, and other nucleophilic attack or reductive release forming primary cyclized core structure. Which further undergoes subsequent cyclization, C-C cleavage, and condensation or rearrangement to form secondary core structure.

1.5. Structural insights on type III PKSs

The structural study of type III PKSs from bacteria, fungi, and plants revealed a common mechanism[55] for assembly of the polyketide scaffold with the loading of the acyl-CoA derived substrate onto the active site cysteine of the conserved catalytic triad, Cys-His-Asn. Cysteine acts as a nucleophile and attachment site for polyketide intermediate. Histidine acts as a general base during the nucleophilic thiolate anion formation. It is followed by the repetitive decarboxylation condensation reaction of the extender substrate (usually malonyl-CoA) assisted by Phenylalanine and Asparagine with elongation of the polyketide chain giving an intermediate product. The intermediate product undergoes final cyclization assisted by the number of amino acids of the catalytic site and active site, release of the polyketide product from the PKS[1]. The diversity of the product from type III PKSs depends on the use of different starter units (acetyl, propionyl, or other aliphatic and

aromatic acyl-CoA), and extender units (malonyl-CoA or modified malonyl-CoAs), the number of condensation steps and final cyclization mechanism (Claisen condensation, Aldol condensation, or lactonization). The type III PKSs has $\alpha\beta\alpha\beta\alpha$-fold architecture and belongs to the thiolase-fold group of the enzyme. Different crystallographic and architecture based studies made on type III PKSs from plants and microorganisms show the active site amino acid residues of type III PKS dictate the starter molecule preferences, polyketide chain-length, and regio-specificity of polyketides during cyclization[56]. Further, the volume of the initiation/ elongation cavity also guide the starter molecule selectivity and polyketide length[57,58]. It paves the path to the rapid diversification of polyketide products through molecular clarity and understanding biochemistry for structure-assisted engineering experiments[59,60].

1.5.1. Structure of plant type III PKSs

Plant type III PKSs have a high level of sequence identity (70-95 %) compared to bacterial type III PKSs (with 20-40 %). Polyketides from plants are used for different purposes such as protection, flower colour, pollen development, root nodulation, astringent, anti-microbial phytoalexins and chemical defence. The schematic representation for the formation of various polyketide products from the plant type III PKSs is shown in figure 1.2.

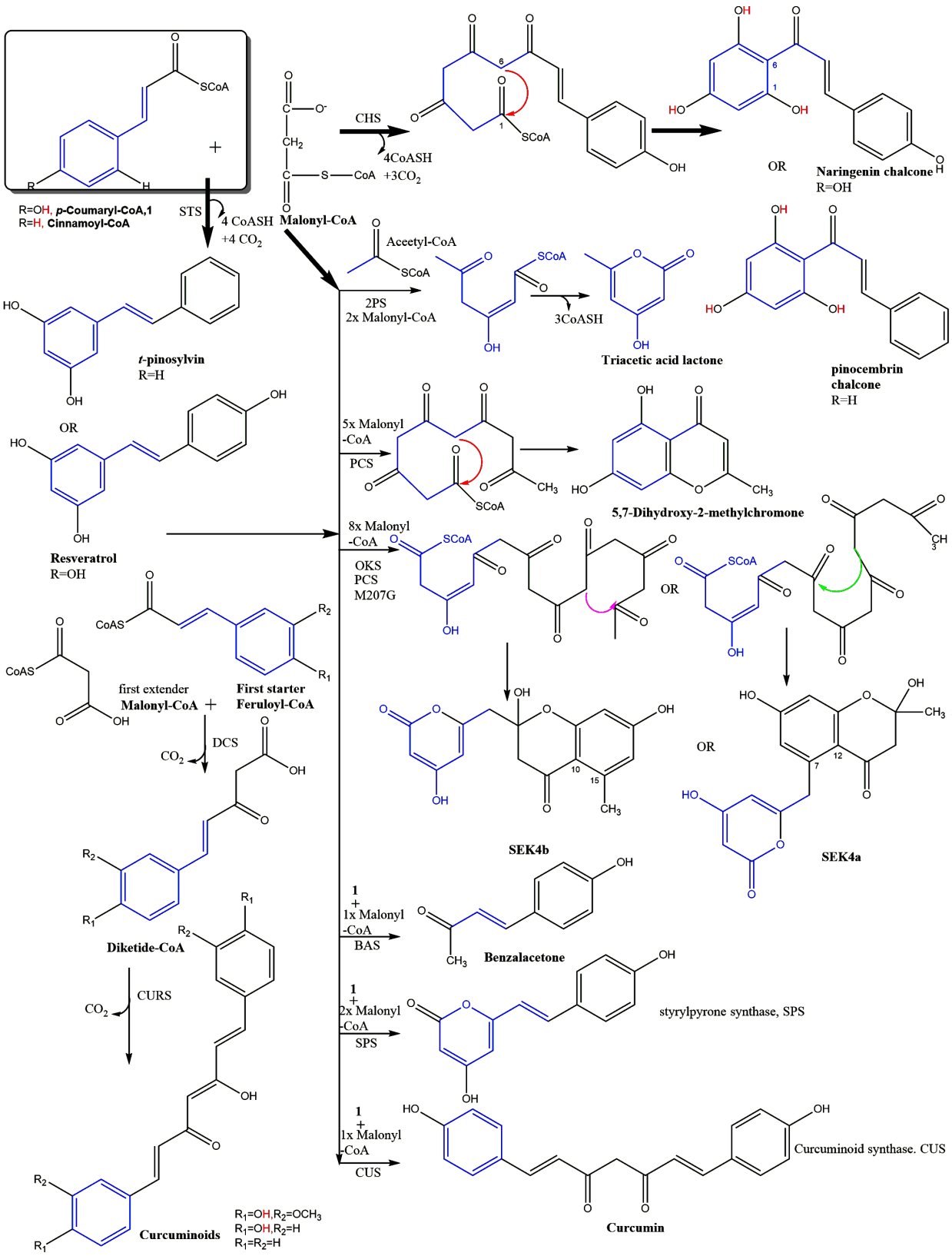


Figure 1.2: Diversity of polyketides generated using different plant type III polyketide synthases. The schematic representation shows the formation of different products using different acyl-CoA as starter and malonyl-CoA as an extender by chalcone synthase (CHS), stilbene synthase (STS), 2-pyrone synthase (2PS), pentaketide chromone synthase (PCS), octaketide synthase (OKS), the mutant of PCS

(M207G), benzalacetone synthase (BAS), styrylpyrone synthase (SPS), curcuminoid synthase (CUS), diketide-CoA synthase (DCS) and curcumin synthase (CURS)

1.5.1.1.Chalcone synthase

Plant chalcone synthase is the first type III PKS described functionally and structurally. It has three cavities viz. the CoA-binding tunnel, a coumaroyl-binding pocket, and a cyclization pocket. The CoA-binding tunnel is 15˚A long, which provides access *p*-coumaroyl-CoA as a starter to load onto the active site residues (Cys164, His303, and Asn336). At the entrance of the CoA-binding tunnel, there are Lys55, Arg58, and Lys62 residues making hydrogen bonds with two-phosphates of CoA. The additional hydrogen bond between the nitrogen of Ala308 and the first carbonyl of pantotheine moiety and van der Walls contacts maintain the interaction between CHS and CoA. The set of conserved residues of catalytic triad are situated in a bi-lobed internal cavity (initiation/elongation cavity). An imidazolium ion stabilizes the transfer of acyl moiety to the thiolate anion of the Cys164 on the His303. It is followed by the decarboxylative condensation of malonyl-CoA facilitate by protonated nitrogen's of the side chain of His303 and Asn336, the key catalytic residues forming an "oxyanion hole". Oxyanion hole provides an electron sink and facilitates decarboxylation.

Further, the iterative process of a recapture of extender and condensation onto the starter-acetyl-diketide-CoA and additional rounds of elongation results in triketide and tetraketide intermediate. It proceeds to intramolecular cyclization reaction and release of polyketide. The gatekeeper residues Phe215 and Phe265, at the lower end are thought to block the lower portion opening between the "Co-A binding tunnel" and the active site cavity and help in folding and internal orientation of the intermediate during cyclization. The gatekeeper blocks the bulky substrate like N-methylanthraniloyl-CoA to binds onto the Cys164. The structure-based engineering approach can be a metabolic approach to increase

the chemical diversity of CHS products. Besides there are chemically inert residues lining the active site cavity includes Thr132, Ser133, Thr194, Thr197, Gly256, Phe265, and Ser 338 which are conserved in CHS and thought to control substrate and product specificity. The mutant CHS S338V from *Scutellaria baicalensis* was found to produce octaketide SEK4/ SEK4b from eight molecules of malonyl-CoA. The activity was found dramatically increased using CHS triple mutant (T197G/G256L/S338T). This functional conversion is based on the simple steric modulation conferred by the inert amino acid residues lining the active site cavity[61].

1.5.1.2. Stilbene synthase

Like CHS stilbene synthase (STS) use most preferably *p*-coumaroyl-CoA as a starter with iterative condensation of three acetyl units derived from malonyl-CoA to form identical tetraketide intermediates. The phylogenetic analysis also shows the CHS and STS have a 75-90% amino acid sequence identity over there 400 amino acid residues. The crystallography study of stilbene synthase crystal structure solved for *Pinus sylvestris*[62] and *Arachis hypogaea*[63] at 2.1˚A and 2.4 ˚A respectively form pinosylvin and resveratrol. However, the cyclization of the intermediate in case of STS is C2/C7 aldol condensation with the product having stilbene backbone unlike that of CHS is C6/C1 Claisen condensation forming chalcone scaffold. The comparative structural study of *Medicago sativa* CHS2 with STS shows a minor topological difference in the active site cavity. However, the loop region between residues 132 and 136 in the active site of STS is reported to have significant backbone change. Displacement of Thr132 is found to have an effect on hydrogen bond network around the catalytic residue Cys164 also called: aldol-switch." Aldol switch is a thioesterase-like electron hydrogen bond network of Ser338-H2O-Thr132-Glu192. The cyclization mechanism of STS most likely activated by nucleophilic water-based "aldol-switch", which cleaves the C1 thioester bond of enzyme-bound

tetraketide intermediate leading to C2/C7 aldol cyclization. The aldol cyclization in STS is favoured when the C1-position of a linear polyketide as a part of free acid[64] rather than ester (thioester) as reported in CHS. It is followed by the additional decarboxylative elimination of the resulting acidic intermediate (carboxylate) to give stilbene backbone.

1.5.1.3. 2-Pyrone synthase (2-PS)

The X-ray crystal structure of 2-pyrone synthase (2-PS) from *Gerbera hybrida* complexed with its natural substrate acetoacetyl-CoA at 2.05˚A revealed the fact that 2-PS has smaller initiation/elongation cavity[4]. It has dimerization surface (with residues Met142 and Pro143) allow methionine from one monomer to protrude to adjoining monomer's active site form structurally complete 2-PS active site. The active site dimerization surface has one-third of its area of dimerization 1805 $Å^2$ /monomer compared to CHS[56,57]. Cavity volume dictates polyketide length and influences the choice of starter molecule. It does not use bulky *p*-coumaroyl-CoA as a starter to give product chalcone but uses acetoacetyl-CoA with two units of malonyl-CoA to give product 6-methyl-4-hydroxy-2-prone (methylpyrone) or triacetic acid lactone (TAL)[65]. It shares 74% amino acid sequence identity with CHS. It has a similar three-dimensional fold with a set of conserved catalytic residues 2-PS (Cys169, His308, and Asn 341) position as that in CHS (Cys164, His303, Asn336), and KSAII (Cys163, His303, and His340). The superimposition of the structure of 2-PS and CHS show nearly identical structure with RMSD 0.64˚A for Cα-atoms. Unlike the chemically inert residues of 2-PS: Leu202, Leu261, and Ile343 (*M. sativa* CHS2: Thr197, Gly256, and Ser338) lining the active site cavity aid steric contraction and orient the acetoacetyl moiety at the end of 15 ˚A long tunnel decide the fate of initiation and elongation steps of polyketide formation. The proof on the structural basis for volume and shape of the active site cavity dictate functional diversity was made by the four amino acid substitutions made on CHS2 with T197L, I254M, G256L, and S338I. The cavity

volume of CHS reduce from 923 $Å^3$ in CHS to 274 $Å^3$ in 2-PS and the triple mutant T197L/G256L/S338I were found functionally identical to 2PS, producing TAL from acetyl-CA as a starter and two molecules of malonyl-CoA as an extender. In summary, the functional diversity of 2-PS is controlled by the selectivity of the starter substrate and malonyl-CoA condensation by steric modulation of the active site cavity[57].

1.5.1.4. Pentaketide chromone synthase and octaketide synthase (PCS and OKS)

Pentaketide chromone synthase (PCS) and octaketide synthase (OKS) are the type III PKS characterized by aloe (*Aloe arborescens*) with conserved Cys174-His316-Asn349 as well as most of the CHS active site residues. They are 91 % identical and share a 50-60 % amino acid identity with CHS. PCS and OKS undergo iterative condensation of five and eight molecules of malonyl-CoA respectively to form respective pentaketide (5,7-dihydroxy-2-methylchromone), and octaketides (SEK4 and SEK4b). The X-ray crystallographic study with 1.6 ˚A and site-directed mutagenesis shows the single inert amino acid (Met207 in PCS, and Gly207 in OKS) replace Thr197of CHS2 lining the active site cavity, dictate the polyketide length and product specificity depending on the bulkiness of side chain[66]. The M207G substitution opens the gate to the pockets expanding the putative polyketide chain elongation tunnel forming octaketide, SEK4a, and SEK4b rather than pentaketide in PCS mutant[53] with same overall folding except for residues 211-220 like the PCS. Further, the structural comparison of PCS, CHS2 and 2PS reveal a significant change in the backbone residues 296-309 of PCS (*M. sativa* CHS: 289-296, G. hybrid 2PS: 291-301). Although the residues allow the formation of an expanded surface-exposed loop (residue: 296-309), they increase the dimerization surface area. But are not thought to be crucial for catalytic activity. The comparison of amino acid from the active site architecture depicts the PCS and CHS with nearly identical superimposable active site residues but with only exceptions in Cys143, Thr204, Met207, Leu266, and Val351(*M. sativa* CHS: Ser133, Thr194,

Thr197, Gly256, and Ser338 respectively). The latter three amino acid Met207, Leu266, and Val351 are altered in many PKSs and are critical for substrate and product specificities.[53,57,67] Octaketide molecule gets accommodate to M207G mutant of PCS pocket A as like OKS because the mutant contains two additional novels, buried pockets (pocket A and B) and the cavity volume is 2.6 times (649 $Å^3$) larger than wild-type PCS (247 $Å^3$). The position and orientation of pocket A of the PCS mutant overlap with an acyl-binding tunnel of type III PKSs (THNS[68] and PKS18[54]) share similar architectural strategies to accept long-chain (C12-C20) fatty acyl starter to produce long-chain polyketides, including pyrones, phloroglucinols, and resorcinols.[69,70] PCS and M207G mutant (as well as OKS) undergo different kinds of cyclization such as C-1/C-6 Claisen-type cyclization of pentaketide intermediate, and C-10/C-15 and C-12/C-7-aldol type cyclization of octaketide form the product 5, 7-dihydroxy-2-methylchromone and SEK4b: SEK4a(4:1) respectively. Furthermore, the structural analysis of PCS depicts Met207 more flexible than the corresponding Thr197 of *M. sativa* CHS and Leu202 of G. hybrid 2-PS[56,57].

1.5.1.5. Benzalacetone synthases (BAS) and Curcuminoid synthase

Some of the plant-specific type III PKSs like benzalacetone synthase (BAS) and curcuminoid synthase (CUS) undergo condensation of extender onto the starter acyl-CoA without cyclization[71].

1.5.1.5.1. Benzalacetone synthase

BAS shares approximately 70 % amino acid sequence identity with CHS and STS. It catalyzes one-step decarboxylative condensation of starter 4-coumaroyl-CoA with extender malonyl-CoA to produce diketide benzalacetone, 4-(4-hydroxyphenyl) but-3-en-2-yne. The crystal structure of Rheum palmatum wild-type BAS with 1.6 Å resolution and

I207L/L208F mutant producing chalcone with 1.8 Å resolution is the first direct evidence for monoketide intermediate (4-coumaroyl thioester) covalently bound to SH group of Cys157 residue[72]. Superimposition of BAS structure with *M. sativa* CHS and *Pinus sylvestris* STS show homology with RMSD of 0.66 Å and 0.73 Å respectively. The structural study depicts the presence of a CoA-binding tunnel and large internal cavity along with alternative, novel active site pocket for binding the aromatic moiety of the coumarate. Due to steric contraction, the chain elongation is interrupted at diketide stage. Phe208 in BAS (corresponds to gatekeeper residue Phe 215 in CHS of *M. sativa*) accounts for the interruption of the polyketide. I207L/L208F mutant show restored function of coumaroyl binding pocket with the formation of chalcone. BAS has hydrophobic Leu in place of Thr132 in STS. The enzyme uses a hydrogen bond network between catalytic triad, Cys157-His296-Asn329, and a putative nucleophilic water molecule. The general mechanism of catalysis in BAS is similar to CHS- thioester bond cleavage and decarboxylative Claisen-type condensation of malonyl-CoA with the production of C6-C4 benzalacetone. In summary, BAS use unique thioester bond cleavage of enzyme-bound intermediate, that initiate reaction by the "aldol-switch" like a rearrangement of hydrogen bond networks of the *β*-keto acid produced by the nucleophilic attack of water on the catalytic triad, and proton abstraction by the His296, reactivated by Cys157 thiolate forming a stabilized His296-Asn329 oxyanion hole. It subsequently undergoes decarboxylation, and further tautomerized to produce benzalacetone with the restoration of the Cys157-His296 thiolate-imidazolium ion pair.

1.5.1.5.2. Curcuminoid synthase

Curcuminoid synthase (CUS) share 40-51 % identity with another plant type III PKSs and synthesize secondary metabolites curcuminoids. Curcuminoids are the polyphenolic compounds imparting yellow colour. Curcuminoids include curcumin,

demethoxycurcumin, and bisdemethoxycurcumin (widely used in food additives and traditional medicine). The rhizome of turmeric (*Curcuma longa*) and root of *Oryza sativa* are sources for curcuminoids. Based on the 2.0 Å resolution crystal structure of CUS from *O. sativa*, the enzyme form homodimer with root means square deviation (RMSD) 0.5 Å and the buried surface area 4995.8 $Å^2$ between the protomers[73]. It has a functionally independent active site pocket with catalytic triad Cys174, His316, and Asn349. The reaction begins with the attachment of *p*-coumaroyl-CoA, the first starter substrate with thiol moiety of Cys-174. CUS structure shows a starter substrate-binding pocket similar to CHS and STS and occupied by the residues Glu202 and Tyr207. It is followed by the decarboxylative condensation of malonyl-CoA (the first extender substrate) onto the thioester of *p*-coumarate with the formation of diketide- CoA intermediate. The hydrolysis of the intermediate result in the formation of *β*-keto acid (second extender substrate). The structural study of enzyme depicts that the CUS has large characteristic active site pocket to accommodate *β*-keto acid probably due to the replacement of gatekeeper phenylalanine (Phe265) of *M.sativa* CHS (that act as a mobile steric gate during successive rounds of polyketide elongation) with glycine residue increasing the cavity volume. The *β*-keto acid condenses onto the thioester of *p*-coumarate give bisdemethoxycurcumin. The approximate 15 Å distance from the deepest point in the catalytic pocket to the cysteine and the orientation of Arg217 that divides the pocket from solvent probably help in the unique head-to-head assembly of the polyketide in CUS. Unlike *O. sativa* CUS, biosynthesis of curcuminoids in turmeric (*C. longa*) is catalyzed by two enzymes diketide-CoA synthase (DCS) and curcumin synthase (CURS)[74]. The former synthesize feruloyl-diketide-CoA using feruloyl-CoA as starter and malonyl-CoA as extender which on coupling with CURS gives curcumin.

1.5.1.6. Alkylresorcinol synthase, alkylpyrone synthase, and phloroglucinol synthase

The preliminary X-ray diffraction study from polyketide synthase-1 from *Cannabis sativa* shows the formation of hexanoyl triaceticacid lactone from the iterative condensation of hexanoyl-CoA with three molecules of malonyl-CoA[75] probably by using three amino acid in catalytic triad Cys-His-Asn. Similarly, recombinant olivetol synthase (OLS) made from cDNA encoding type III PKS from the same plant synthesizes olivetol, the decarboxylative from of olivetolic acid as the main product[76]. Alkylresorcinol moiety of cannabinoids is derived from olivetol and olivetolic acid. The genome sequencing projects of the model plant *Arabidopsis thaliana* revealed two chalcone synthase (CHS) homologs PKS-A and PKS-B with 20-40 % amino acid identity to plant type III PKSs[77]. The two PKSs were reported to accept long-chain fatty acyl-CoAs (C_6-C_{20}) as a starter and carry out iterative condensations reaction with two or three molecules of malonyl-CoA to produce long-chain triketide or tetraketide *α*-pyrones as a major product with phloroglucinol as a minor product. It shows similarity to bacterial type III PKS from *Mycobacterium tuberculosis*, PKS18 that produces long-chain lipid metabolites triketide *α*-pyrone and tetraketide *α*-pyrones[78].

Nevertheless, the structural study of phenolic compounds producing type III PKSs is not done in a plant to do mechanistic research and unravel the cyclization mechanism. Plant, algal and bacterial type III PKSs diversity can be studied based on the common cyclization mechanism to form similar kind of polyphenolic compounds as illustrated below in figure 1.3.

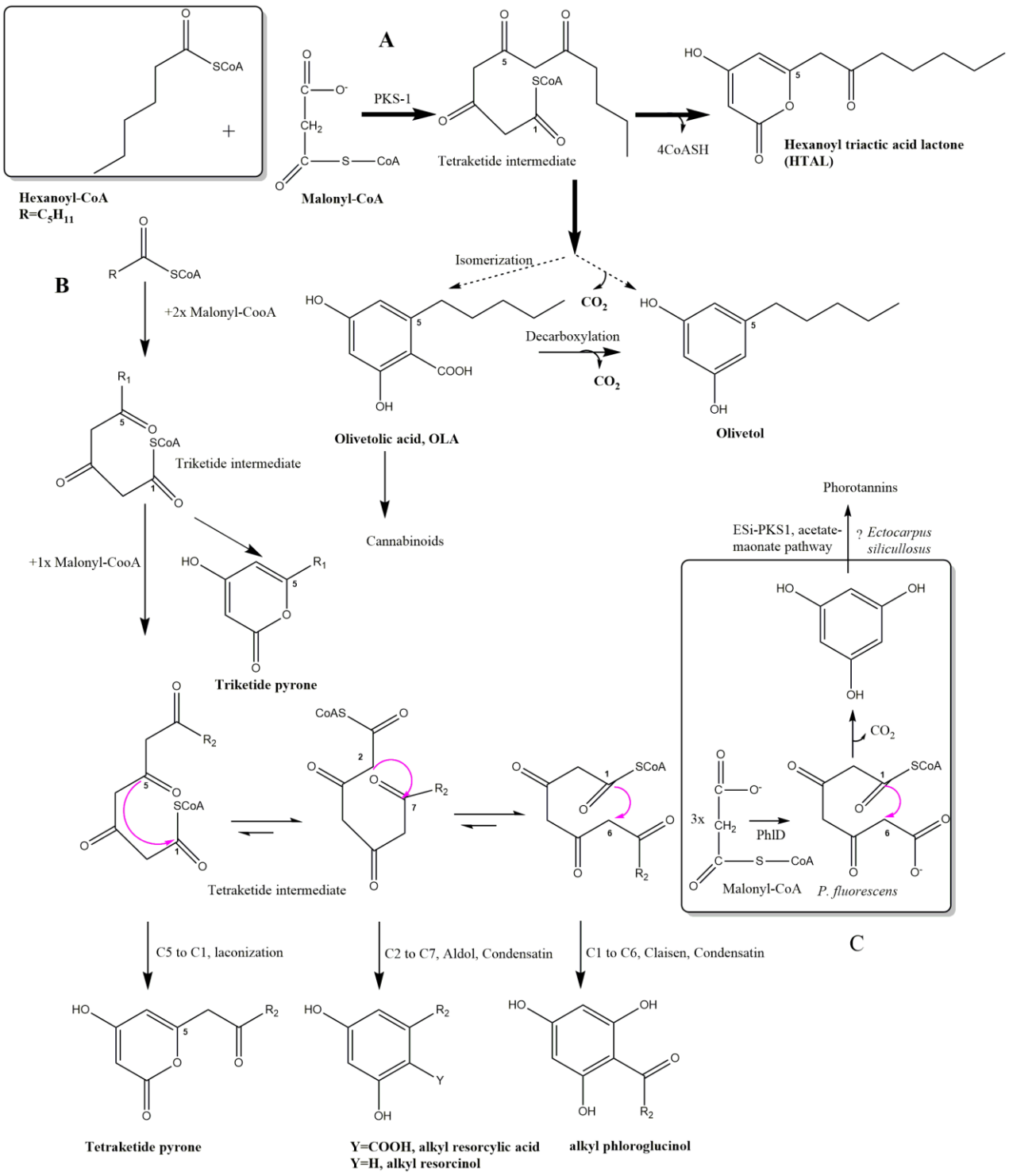


Figure 1.3: Purposed mechanisms for the formation of various polyketide products in plants, algae and bacteria based on the different studies of type III PKSs. The diagram illustrates the possible mechanism for the formation of alkyl pyrones like molecules- hexanoyl triacetic acid lactone, triketide and tetraketide *α*-pyrone in plants (A) and bacteria (B) using different C-chain acyl-CoA as starter and malonyl-CoA as an extender drawn in a schematic representation. For plant PKS-1 and bacterial PKS18 was purposed with a similar mechanism. The formation of olivetol by plant olivetol synthase (OLS, A) is identical to the one for the formation of alkyl resorcinol through C2 to C7 Aldol condensation (B). The formation of phloroglucinol in a bacterium, *P. fluorescens* through type III PKS, PhlD (C) is similar to the

algal *Esi*- PKS1 which is hypothesized to follow acetate malonate pathway to give condensed polymerized products. The mechanism of polymerization to form phlorotannins is still unknown.

1.5.2. Structural insights on type III PKSs from bacteria

1.5.2.1. RppA/THNS

The study of type III PKSs from bacteria started with THNS/RppA from gram-positive, filamentous bacterium *Streptomyces coelicolor* A3 with 2.2 ˚A resolution protein crystal structure[68]. The electron density of THNS crystal structure revealed conserved *αβαβα*-fold core domain with a homodimeric interface and Ramchandran plot show 86.6 % of amino acid residue conformations in the most favoured region, 11.2 % in additional allowed, 2.2 % in generously allowed, and no residues in the disallowed region. Further, the C-termini of THNS has unusual extensions (approximately 25 residues) not found in other plants or bacterial type III PKSs. The enzyme catalyzes the synthesis of 1,3,6,8-tetrahydroxynaphthalene (THN), a precursor for hexahydroxyperylenequinone (HPQ) melanin, using five molecules of malonyl-CoA as starter and extender. It uses a novel cavity extending into the "floor" of the traditional active site cavity for extra polyketide extension beyond triketide when primed with a long chain starter. It catalyzed polyketide extension using the conserved residue of catalytic triad Cys138-His270-Asn303 in an internal cavity which further interacts with the surrounding aqueous phase using narrow CoA-binding tunnel. THNS has four cysteines at positions 106, 138, 168, and 171 (CHS: 132,194, 164 and 197)[1]. Cys138 (CHS: Cys164) act as a universal catalytic site for the malonyl-primed polyketide extension beyond triketide stage. The additionally exposed cysteine (replaced by threonine in CHS responsible for an increase in cavity volume) might contribute the enzyme unusual physiological reaction, and sequence divergence from bacterial CHS-like proteins.

Further, presence of the buried tunnel extending out from the active site cavity of RppA, similar to PKS18[54] probably assist in binding with long-chain acyl-CoA and help in polyketide extensions. Pairwise superimposition of THNS with plant type III PKSs (CHS, STS, and 2-PS) structures shows the root mean square deviation closer to 2 Å, and the THNS has obscure reaction pathway. Still, the mechanism is described in terms of initial C6/C1 Claisen condensation and C10/C5 aldol condensation with the formation of aromatic phloroglucinol-like intermediate and resorcinol-like intermediate, respectively. Further intermediate undergoes subsequent condensation and aromatization to give THN. Site-directed mutagenesis and structure approach show Tyr224 is important in THNS for starter specificity contribute steric constriction limiting the conformational freedom of Cys138-loaded malonyl starter (small starter molecule) while adequate cysteine probably provides sufficient volume for additional polyketide chain extension steps a dual cyclization necessary for THN production. Cys106 (CHS: Thr132) form hydrogen bonding and shuttle electron from adjacent Glu192 (conserved in type III PKS) activate hydrolytic water next to catalytic cysteine[62]. The substitution of C106S employs derailment of the triketide intermediate to form triacetic acid lactone (TAL), but the mechanism of partitioning between THN and TAL derailment is unknown.

1.5.2.2. PKS18 and PKS11

PKS18 and PKS 11 are type III PKS from *M. tuberculosis*. The former gives non-methylated tri- and tetraketide *α*-pyrone using long-chain acyl-coenzyme A (acyl-CoA) as starter units (C_6-C_{20}) and malonyl-CoA as an extender unit by the iterative condensation of extender into the starter using amino acid of the catalytic triad. The latter gives methyl-branched triketide and tetraketide pyrones using long-chain starter (C_{12}-C_{20}), and methylmalonyl-CoA as the first extender and malonyl-CoA as the second extender followed by lactonization. The two enzymes have 26 % amino acid identity. The X-ray

crystal structure reveals the presence of long 20 Å[54] and 20×5 Å substrate-binding tunnel, respectively, lined primarily with hydrophobic residues in PKS18 and PKS11. In contrast to CHS, which has a small pocket volume to accommodate shorter fatty acids substrate and cyclized chalcone product, PKS18 and PKS11 both can bind with long-chain fatty acid substrates and alkyl pyrones products. Both the proteins are homodimeric enzymes with each monomer bury 15 % of total surface area (PKS18: 2360 $Å^2$ and PKS11:2323 $Å^{2)}$ with backbone C*α* RMSD between the two structures is 2.1 Å over 329 residues. The catalytic triad of PKS18 consists of Cys175, His313, and Asn346 (PKS11: Cys138, His277, and Asn310), and the hydrophobic tunnel is proximal to the active site cysteine to load the substrate.

1.5.2.2.1. PKS18

The crystal structure of PKS18 protein at a resolution of 2.25 Å co-purified with C14 fatty acid (myristic acid) depicts the fact that the ligand did not occupy the tunnel completely, and few water molecules are occupying the remaining space. The observation implies that PKS18 has an open tunnel to accommodate substrate longer than C_{14} fatty acids. Although PKS18 and msCHS have highly superimposable overall fold with RMSD 1.5 Å in C*α*-position for 326 atoms and 42 % homology in protein sequences. However, the major difference is found in the region corresponding to residues 204-209 of PKS18 (193-198 in msCHS). Comparative study of msCHS and PKS18 on the region showed Thr197 blocks the entrance of the substrate-binding tunnel in msCHS, but the corresponding PKS18 residue Asn208 does not protrude in the channel. Whereas, Cys205 corresponds to Thr194, protrude onto the region analogous to msCHS substrate-binding pocket. Besides Cys175 that acts as a catalytic nucleophile, the structure revealed two catalytic cysteine-Cys205 and Cys275. The detailed study showed there is a subtle change in φ and ψ angle result change in the conformation of the residues and Asn208 moves away from the tunnel region.

Mutation by large residue in PKS18 constricts the cavity and analogous to msCHS which is mainly responsible for opening a new tunnel. PKS18 has Leu266 instead of Gly256 and Ala148 in place of Met137, which results in loss of cross subunit interaction. Substitution of msCHS residue M137A, Gly256L, S338L (msCHS: M137, Gly256, S338 to respective PKS18:Ala148, Leu266, and Leu348) result loss of cross-subunit interaction in msCHS . The structural analysis further reveals amino acids Thr144, Cys205 and Ala209 might be crucial for the cavity volume in PKS. It is validated by the mutagenesis experiment of T144F mutant show complete loss of activity using C_{14} and C_{16}- acyl-CoA, C205F mutant preferably use C_6- starter to form tetraketide α-pyrone and A209F mutant switch the specificity of PKS18 toward small acyl-CoA. The unique location of the catalytic site and subtle change in a restricted region of a protein molecule, PKS18 use unusual substrate molecules.

1.5.2.2.2. PKS11

The crystal structure of PKS11 co-purified with ligand palmitoyl-CoA (C_{16}-CoA) at a resolution of 2.05 Å reveal the backbone with 1.56 Å Cα RMSD to plant CHS (over 343 residues. However, the backbone deviates significantly in the four regions of substrate binding tunnel lining the active site cavity corresponding to residues 18-49, 52-71, 172-181, and 228-241 contributing the difference in substrate specificity. The typical chain-elongation occurs by loading of a substrate to catalytic nucleophile Cys-138 and chain elongation by binding of the extender (malonyl-CoA, MCoA or methylmalonyl-CoA, MMCoA) with subsequent decarboxylation.

The complexes with both covalent and noncovalent attached long-chain fatty acids, along with the products formed by reaction with (M)MCoA, are consistent with the typical chain-elongation cycle. It is carried out by other types of type III PKSs, including a substrate

loading phase (acyl transfer from palmitoyl-CoA to Cys-138), binding of an extender unit (MCoA or MMCoA). The catalytic nucleophile facilitates transesterification of acyl-substrate between CoA thioesters and cysteine thiolate and Asn-310 stabilize hydrogen bond to the thioester oxygen. It is followed by decarboxylation and enolate formation, nucleophilic attack on the thioester to form fatty acyl-CoA extended by two carbons, and finally reloading (acyl transfer back to Cys-138). Additional Pro-341 which is conserved in all other types III PKSs (including CHS and STS) postulate to promote cyclization of products[79].

1.5.3. Structural insights on type III PKSs from fungi

PKSIII*Nc* is the first fungal type III PKS reported in 2007 from *Neurospora crassa*[80]. Crystal structure of PKSIII*Nc* with resolution 2.58 Å without ligand[81] and 1.75 Å co-crystallized with ligand[82] reveal the presence of catalytic triad (Cys-152, His305, and Asn-338) situated at a location and orientation with active site flanked by the deep hydrophobic substrate-binding tunnel for binding long-chain fatty acid substrate (C_{14}-C_{22}) similar to PKS18 and THNS[54,68]. According to the structure and mutagenesis experiment of PKSIII*Nc*, the protein has broad substrate specificity to synthesize n-alkylresorcinol[81] with long-chain acyl-CoA as substrate and pyrones from small and medium-chain precursors. Whereas, Horinouchi's group showed the synthesis of pentaketide alkylresorcylic[80] acid with long-chain thioesters. Unlike the PKS18, 2.0 Å resolution crystal structure of trORAS, a separate study from ORAS co-crystallized with eicosanoid acid reveals α7i helical insertion that gives the difference to novel fatty acid-binding tunnel[82].

The active site residues, second cysteine Cys-120, Phe-252 (msCHS: Gly-256) and Met-189 (msCHS: Thr 197) determine the specificity of PKSIII*Nc* to different substrates. Cys-120 (mCHS: Thr-132, *S. coelicolor*-THNS: Cys106) precludes the formation of a

thioesterase-like hydrogen bonding network for partitioning intermediates. Phe-252 residue acts as a steric regulator as conferred by mutagenesis and crystal structure of F252G mutant with 2.1 Å resolution. Similarly, the truncated ORAS-eicosanoic acid co-crystal structure reveals the role of Met-189 to orient the substrate onto the active site.

1.5.4. Structural insights on type III PKSs from algae

The type III PKS from brown algae, *Ectocarpus siliculosus*, PKS1 (*ESi*-PKS1), catalyzes the synthesis of phloratannins[83]. pIn brown algae, phlorotannins act as antifouling substance[84] and provide chemical defences against herbivory[85] and act as UV sunscreens for the seaweeds[86]. The condensation of acetate units as starter and malonyl-CoA as extender catalyzed by *ESi*-PKS1 with cyclization and tautomerization of intermediates form phloroglucinol which further condenses into phlorotannins[71,83]. The crystal structure of *Esi*-PKS1 with 2.85 Å resolution superimpose with PKS18 shows identical residues corresponding to the initiation and elongation pocket of the active site (*Esi*-PKS1/PKS18: Ser162/ Ser143, Thr163/ Thr144, Phe243/ Phe224, Ile282/ Ile264, Leu284/ Leu266, Ile291/ Ile273, Cys293/ Cys275)[87]. The molecule in the catalytic triad is conserved with Cys-194, His-331, and Asn-364. Like PKS18, Esi-PKS1 uses a tunnel created by thiolase fold of protein for substrate binding. Unlike there is an additional cavity filled with bulky aromatic residues and covered by two adjacent and asymmetric loops containing Gly-Phe stretch. The cavity allows the insertion of a small molecule, malonyl-CoA, and the loop might enhance the release of phloroglucinol units out of the molecule. Pocket with malonic acid ligand formed by Phe-128, Val-159, Phe-165, Leu-166, Leu-170, Leu-188, and Phe-191 has a volume of 210 $Å^3$. Phe-191 acts as a gatekeeper residue for the ligand exit and entry making interface between two pockets. Further, the position of residues is well conserved with the equivalent cavity volume. Two amino acids Ser224 and His227 that differ in position from residues in PKS18 (Cys205 and Asn 208) probably determine the

cyclization of two enzymes. Further, His-227 occupying more space might be crucial to produce acylphloroglucinols.

1.5.5. Structural insights on type III PKSs from slime molds

"Steely" protein from *Dictyostelium discoideum* consists of N-terminal type I FAS like the domain to produce starter hexanoyl-CoA. The six catalytic domains in N-terminal transfer the hexanoyl-CoA to the Steely C-terminal domain with conserve homodimeric type III PKS fold. Malonyl-CoA condenses with starter hexanoyl-CoA to synthesize polyketide derived phlorocaprophenone (PCP) acylphloroglucinol skeleton of DIF-1. The 2.9 Å resolution crystal structure depicts the conservation of overall fold with homodimeric assembly, and internal active site cavity along with conserved residues in catalytic triad[88].

1.6. Type III PKSs and modifying enzyme

Polyketides undergo tailoring reactions by modifying enzymes. Some of the tailoring reactions are glycosylations, alkylations, acylation, hydroxylation, epoxidation, halogenation, transamination, nitrile formation, desaturation, methylation and sulfation[29]. Type III PKSs either alone or using modifying enzymes are essential to generate compound with anti-bacterial and anti-influenza A virus activity by overexpression of cryptic secondary biosynthetic gene cluster. Methylated Violapyrones are produced by overexpression of type III PKS gene *vioA* under the control of constitutive promoter PgapDH in three Streptomyces strain[89]. The availability of gene clusters along with precursors for the biosynthesis based on the selection of expression host, enzymatic machinery, and regulatory system guide the production of varied secondary metabolites[90,91] The precursors or intermediates moreover facilitate functional secondary metabolite biosynthesis from the primary metabolism[92]. The crosstalk between primary

and secondary metabolism perhaps balances the excess nutrient and substance and helps the organism to survive in adverse conditions[93,94].

1.6.1. Methyltransferase

Methylation is an important phenomenon in all three domains of life, and transfer of methyl moiety occurs from methyl donor to acceptor nucleophile atom usually O, N, C or S and sometimes halide ion[95,96]. Methyltransferase is the enzyme responsible for the methylation of a wide array of substrates including hormones, neurotransmitters, proteins, nucleic acids, lipids, carbohydrates, small molecule, and natural products modulating the bioavailability and bioactivity of acceptor molecules. The biological fate of methylation includes biosynthesis of diverse metabolites, host defence compounds, pigments, prosthetic groups, cofactors, cell membrane, and cell wall components, xenobiotics, metabolism, signal transduction, protein sorting, nucleic acid processing. Methylation is vital in epigenetics regulation as studied in bacteria with DNA methyltransferase that transfer methyl groups from S-adenosyl-*L*-methionine to adenine or cytosine bases.

1.6.1.1. Types of methyltransferase:

Methyltransferase can be classified as *O*-methyltransferases (OMTs), N-methyltransferases (NMTs), C-methyltransferases (CMTs) and S-methyltransferases (SMTs) based on the nucleophiles targeted for methylation. OMTs are the most abundant methyltransferase*s.* Based on the methyl donor methyltransferases can be classified as SAM-dependent (AdoMet) methyltransferase and non-SAM dependent methyltransferase. *S*-adenosylmethionine (SAM, AdoMet) acts as a universal methyl donor in most of the organisms giving a methylated acceptor with S-adenosyl *L*-homocysteine (SAH) as a by-product[97]. Following ATP, AdoMet is the second most widely used enzyme substrate[98].

Based on the unique sequence and structural features, currently, eight classes (class I-VIII) of AdoMet methyltransferases are known[99-102].

The structural elucidation of first AdoMet-dependent methyltransferase was done for DNA C5-cytosine methyltransferase MHhaI in 1993[103]. It belongs to Class I AdoMet methyltransferase, which has the majority among the known AdoMet methyltransferase[99]. Class I methyltransferase has highly conserved seven stranded *β*-sheets flanked by *α*-helices to form *αβα* sandwich. Most of the Class I methyltransferase are monomers, some are homodimer or tetramer with variant structural features[104]. Smallest methyltransferase like catechol *O*-MTase, histamine N-methyltransferase, DNA (cytosine) methyltransferase, Precorrin-6y C-Methyltransferase, isoflavone O4'-methyltransferase, O6-alkylguanine DNA methyltransferase, protein-arginine N-methyltransferase, isoliquiritigenin (chalcone) O2'-methyltransferase, FtsJ RNA methyltransferase belongs to Class I. Class II methyltransferase has conserved RxxGY motif with dominating long central antiparallel *β*-sheets flanked by groups of helices at either end[105]. Class I, Class IV derived from Rossmann-fold proteins with conserved GxG motif, i.e. also conserved in Class III CbiF for nucleotide-binding showing common evolutionary history. Class II and Class V, Class VI, Class VII, and Class VIII methyltransferase do not have structural analogs known, so the evolutionary history is unknown[99]. Transmembrane protein, isoprenylcysteine carboxyl MTase (ICMT) belong to the VI class[106]. Cfr is a 23S ribosomal methyltransferase methylate the C-8 position of rRNA and has conserved the CX3CX2C motif that confers antibiotic resistance to bacteria belongs to Class VII SAM-dependent methyltransferase[101]. The family also includes BchQ and BchR methylate C-8^2 and C-12^1 atoms of bacteriochlorophyllide[107]. Based on the recent findings and records Protein Data Bank of >100 structures for 50 distinct AdoMet-dependent methyltransferases from 31 different classes, enzymes are defined by the Enzyme Classification (EC) system[104].

1.6.1.1.1. N-methylation

The AdoMet-dependent methyltransferase undergoing N-methylation has a conserved motif, DPPY [D/N/S]PP[Y/F] selective for nitrogens, extend from the C-terminus of fourth *β*-strand of Class I structure. Phospholipid N-methyltransferase, PmtA from plant pathogen *Agrobacterium* tumefaciens catalyze the formation of phosphatidylcholine, via phosphatidyl-N-monomethylcholine and phosphatidyl-N-dimethylcholine help in crown gall/tumour formation by playing the role of the second messenger belongs to this family[108]. Similarly, DNA N6-adenine[109] and DNA N4-cytosine[110] MTases, N5-glutamine MTase, PrmC/HemK[111] undergo N-methylation.

1.6.1.1.2. C-methylation

The *in silico* study on methyltransferase domain showed N-MT domain-containing protein has significant homology to C-MT proteins, in the study made in NRPS/PKS methyltransferase during the myriad of secondary metabolite synthesis[112]. The exact mechanism of C-methylation is unknown. However, the study of much substrate-bound structures from DNA C5-cytosine methyltransferase showed that C-methylation is favoured by the conserved ProCys dipeptide in the active site, resemble as PY portion of the DPPY motif[113,114]. The aspartic acid functionally replaces the [D/N/S] residue of the DPPY motif positioning the N4 nitrogen of the cytosine base further away. Covalent Cys-thiol and C6 carbon generate a negative charge on C5, facilitate the methyl transfer from AdoMet[115]. The crystal structure of CouU, a C-methyltransferase from *Streptococcus rishirensis* and a study on NovO from *Streptococcus spheroids* show C-C bond formation catalyzed by SAM-dependent methyltransferase. The presence of novel His-Arg motif facilitates the deprotonation of the phenolic proton from C-7 position, followed by methylation at C-8 giving methylated coumermycin A1 and novobiocin respectively[116].

Both the enzymes showed activity towards various naphthalene derivatives and also undergo Friedel-Crafts alkylation by using non-natural SAM analogue to alkylate regiospecifically.

1.6.1.1.3. *O*-methylation

O-methyltransferases (OMTs) are a group of enzymes methylate the electron-rich *O*-position of the acceptor molecule. They are present in diverse organisms, including plants, animals, bacteria, and fungi conferring a wide range of biological functions[117]. OMTs can be classified into two classes Class I and Class II based on the sequence and phylogenetic analysis[118]. Class I protein include caffeoyl coenzyme A 3-OMT (CCoAOMT) from alfalfa (*Medicago sativa*)[119], catechol OMT (COMT) from rat (*Rattus norvegicus*)[120], repair enzyme protein L-isoaspartyl (D-aspartyl) OMTs from human[121], *Pyrococcus furiosus*[122], and *Vibrio cholerae*[123], Cmcl from *Streptomyces clavuligerus*[124]. It exists as a homodimer with molecular weights of 23-28 kDa and typically requiring divalent cation as Mg^{2+}[125]. Mg^{2+} acts primarily organize the substrate-binding site and is not a general base[126]. Among the OMTs catechol *O*-methyltransferases (COMT) is the most widely studied[120] because the enzyme has a role in methylating the hydroxyl group of catechol, a neurotransmitter and used to inactivate the catechol-type compounds such as L-Dopamine[127]. The structural superimposition of CmcI and COMT revealed both the protein bind to Mg^{2+} at the same position and in CmcI coordinated by Asp160 (Asp141 in COMT), Glu186 and Asp187[124]. Plants OMTs biosynthesize specialized metabolites like lignin from methylated caffeoyl-CoA, flavonoids, and phenylpropanoid conjugate through the methylation of phenolic compounds[128]. There is another class of methyltransferase protein with larger subunit sizes (38-43 kDa) and use catalytic histidine residue rather than Mg^{2+} dependent methylation. It belongs to Class II and includes alfalfa chalcone OMT[129] that methylate 2'-hydroxyl position of isoliquiritigenin (4, 2',4'-trihydroxychalcone) giving 4, 4'-dihydroxy-2'-

methoxychalcone, a signalling molecule and an inducer of nodulation in soil rhizobia[130,131]. Besides Class II OMTs include alfalfa isoflavone OMT[129], alfalfa caffeic acid OMT[132], and isoflavonoid OMT from *Medicago truncatula*[133]. Domain analysis of Class II methyltransferase shows C-terminal catalytic domain with both SAM and substrate binding site and N-terminal end has dimerization domain[134].

Class I AdoMet-dependent OMTs are also present in bacterial species like *Leptospira interrogans* (LiOMTs)[135] and *Bacillus cereus* (BcOMT2)[134]. Based on sequence and structural analysis, the former uses the metal-dependent catalytic site with a possibly phenolic substrate and the latter uses flavonoids with regiospecific methylation following dimerization mechanism. However, the endogenous substrate of both the enzyme is unknown. The substrate specificity of the BcOMT2 was found to be determined by the SAH binding and disordered orientation of Glu171 to Gly186.

1.6.2. Sulfated metabolites

Sulfated metabolites are widespread in a variety of organisms from prokaryotes to multicellular eukaryotes. In human, sulfated carbohydrate molecules mediate cell to cell communication and signaling[136,137]. Phenolic compounds get detoxified by sulfoconjugation in animals[138]. Sulfate conjugation seems to be important in a plant for its growth, development, and adaptation to stress[139]. Gallic acid glucoside sulfate is a sulfatide from plants responsible for a biological activity like seismonastic and gravitropic movements. Several sulfated secondary metabolites of unknown functions are known to accumulate in a variety of plant species. Flavonoids and sulfate ester play an essential role in plant growth by binding the naphthyl-thalamic acid receptor, thus blocking the quercetin-stimulated accumulation of auxin phytohormone. *Sinorhizobium meliloti*, a gram-negative bacterium, use sulfated compounds as a mediator for plant host interaction.

M. tuberculosis has the abundant sulfolipids, and sulfatides determine the virulence of pathogenic species[140].

History of the existence of sulfated molecules was reported back to 1876. Nevertheless, after 80 years, the mechanism of sulfonation revealed by the discovery of a sulfate donor 3'-phosphoadenosine 5'-phosphosulfate (PAPS)[141]. Sulfotransferase is the transferase enzymes that catalyze the transfer of sulfate group from a donor molecule onto the acceptor alcohol or amine. PAPS act as a universal sulfate donor. Sulfotransferase can be classified as PAPS dependent sulfotransferase and PAPS independent sulfotransferase (arylsulfate sulfotransferase, or aryl sulfotransferase or ASST)[142].

1.6.2.1.Types of Sulfotransferase

PAPS-dependent sulfotransferases use PAPS to transfer the sulfuryl groups from the activated anhydride donor to hydroxyl groups of acceptor molecule sugar, alcohol, phenol, or amine. Based on the site of sulfation, sulfotransferases are cytosolic- and membrane-associated sulfotransferase. Cytosolic sulfotransferase is commonly present in the cytosol of prokaryotes (bacteria) or cytosol of eukaryotes. In contrast, membrane-associated is present in the Golgi apparatus membrane in eukaryotes and cytoplasmic membrane of both prokaryotes and eukaryotes. Cytosolic sulfotransferase is mainly responsible for the clearance of endogenous and exogenous compounds such as hormones, drugs, bio-amines, and xenobiotics compounds. Unlike the membrane-associated sulfotransferase sulfate larger biomolecules like carbohydrates and proteins. But the mechanism behind the sulfation of PAPS-dependent sulfotransferase in both the cytosolic and membrane-associated is believed to occur by the attack of a nucleophile to the sulfate group.[143,144]

In bacteria, like *M. tuberculosis*, the cell to cell interaction of bacterium and host like a human is modulated by S881, a sulfated menaquinone, localized in the outer envelope of

the *M. tuberculosis*, negatively controlling the virulence of the pathogen. The biosynthesis of sulfated compounds depends on the PAPS, and sulfotransferase stf3[145].

PAPS-independent sulfotransferases are also known as aryl sulfotransferases (ASSTs). ASSTs are mainly present in pathogens[146]. Periplasmic ASSTs are commonly present in several commensal intestinal bacterial organisms like Klebsiella K-36[147], *Eubacterium* A-44[148], *Enterobacter amnigenus*[149] which use *p*-Nitrophenyl sulfate. *Haemophilus* K-12[150] preferably use 4-methyl-umbelliferyl sulfate followed by *β*-naphthyl sulfate, *p*-nitrophenyl sulfate and *α*-naphthyl sulfate as sulfate donor for the sulfation of acceptor *α*-naphthol followed by phenol and resorcinol. Although the physiological role of many ASSTs is unknown, the sulfurylation is also reported in gram-negative opportunistic pathogens like *Edwardsiella tarda*[151], *Citrobacter freundii*[152], *Salmonella typhimurium*[153], and uropathogenic *E. coli* CFT073[154]. *Eubacterium* A-44 accepts a variety of aromatic sulfuric acid esters and phenols as sulfuryl donor and acceptors, respectively. The bacterium is reported to sulfurylase phenolic antibiotics like amoxicillin, cefadroxil, and cefoperazone[155] using ASST. The ASST also reported with the activity of transfer of the sulfuryl groups from the laxative sodium picosulfate onto other phenols[156]. Other sulfuryl donors for the enzyme are 4-acetyl phenyl sulfate, 4-methyl umbelliferyl sulfate, and 4-nitrophenyl sulfate with an acceptor naphthol, estradiol, phenol, tyrosine methyl ester, tyramine, and epinephrine, and a variety of other naturally occurring compounds[157]. ASST from *Klebsiella* K-36 sulfurylate small phenolic compounds with up to three fused aromatic rings[147,158].

1.6.2.2.Eukaryotic Sulfotransferase

Besides a wide range of sulfotransferase in prokaryotes, eukaryotes have two kinds of sulfotransferase: cytosolic sulfotransferase and membrane-associated sulfotransferase. The

former also termed as SULTs that sulfurylate small molecules such as hormones, amines, and xenobiotics, including drugs. It has a role in detoxification, hormone regulation, and drug metabolism. Whereas, the latter sulfurylate larger substrates such as carbohydrates and tyrosine residues of peptide chain during the secretory pathway are also called Golgi-resident sulfotransferase and are not studied extensively. The fate of sulfation of xenobiotics by cytosolic sulfotransferases majorly include detoxification of phenolic compounds in the liver[157,159], enhanced excretion of metabolites, increase solubility and bioavailability of drugs with more significant pharmacological activity or in conversion xenobiotics to carcinogen[160].

1.6.2.3. Sulfotransferase, polyketide sulfation

Besides sulfation of proteins, sugars, antibiotics, and a variety of other low molecular weight metabolites, the transfer of sulfur to phenolic compounds is assisted by PAPS, i.e. catalyzed by sulfotransferase enzyme[142]. Type III PKS are found to form polyphenolic compounds similar to the substrate used by ASST[161]. The polyphenolic compounds are aromatic resorcinol, triketide pyrones, tetraketide pyrones, and phloroglucinol mainly. *O*-sulfation of the metabolites is possible and assisted by the conserved sulfotransferase gene of the cluster. *Streptomyces sp*. MK730-62F2 consists of such a cluster (*cpz9-cpz31*) for the caprazamycin, a liponeucleoside antibiotics biosynthesis[162]. Recently, one of the studies described the generation of genuine sulfate donor for the ASST. It shows the type III PKS Cpz6 form a tetraketide α-pyrone, a polyketide molecule. The pyrone is sulfated by the PAPS-dependent sulfotransferase Cpz8 to generate a physiological sulfate donor for ASST Cpz4 for the synthesis of sulfated Caprazamycins[161].

Chapter 2

Biochemical Characterization of Type III Polyketide Synthase

CHAPTER 2: BIOCHEMICAL CHARACTERIZATION OF TYPE III PKS

2.1. INTRODUCTION

Type III polyketide synthases are proteins differing in the specificities in the generation of diverse secondary metabolites called polyketides by the condensation of multiple extender units to thioester linked starter substrate. Crystallographic analysis of type III PKSs reveals a dimeric nature of the protein with two catalytically independent active sites including the signature catalytic triad of Cys-His-Asn and common three-dimensional overall fold[54,56,57,62,68,163]. The PKS has coenzyme A binding tunnel, substrate-binding pocket, and cyclization pocket and undergoes successive decarboxylative condensations of the extender to form products. The diversity in metabolites is due to functionally divergent nature of enzymes with an ability to use a wide range of starter substrates. Type III PKSs use a varying number of extender units such as malonyl-CoA, methylmalonyl-CoA or ethylmalonyl-CoA and starter units ranging from small acetyl-CoA or acetate unit derived from malonyl-CoA or bulky *p*-coumaroyl-CoA or aliphatic-CoA (C_6-C_{22}) or aromatic-CoA (benzoyl-CoA) or isovaleryl-CoA. Further, the type of condensation, cyclizations and aromatization reactions determine the products.

In plants, flavonoids and naringenin chalcone- the precursor of a large number of flavonoids, 4,2',4',6'-tetrahydroxychalcone- a precursor for anthocyanin pigments, anti-microbial phytoalexins, and inducers or Nod-factor for *Rhizobium* based root nodulation, antifungal stilbene phytoalexins, and plant polyphenols are the products of malonyl-CoA condensed with CoA-linked starter molecule by CHS/STS proteins. Although the plant

prototypes of CHS are not identified in the microbial world, there are several type III PKSs or CHS like proteins from microorganisms that are biochemically and structurally characterized. Bacterial type III PKS is used to biosynthesize a large number of secondary metabolites due to easy genetic manipulations, the simple structure of protein under study and feasibility in creating a multitude of bioactive products.

Bacterial type III PKS study for the first time reported with the characterization of RppA from *Streptomyces griseus*. RppaA undergoes condensation of malonyl-CoA to acetyl units derived from malonyl-CoA to synthesize majorly 1,3,6,8-tetrahydroxynaphthalene (THN), the precursor of pigment melanin through intramolecular condensation of pentaketide intermediate. RppA like enzymes are found in other bacterial genera including *Sachharopolyspora erythrea, Sterptomyces coelicolor* and *Sorangium cellulosum*. PhlD, a protein from fluorescent *Pseudomonas spp* synthesizes 2,4-diacetylphloroglucinol, a biocontrol agent against fungal pathogen by the condensation of three molecules of malonyl-CoA. It utilizes from CHS like C6-C1 intramolecular cyclization of a tetraketide intermediate. DpgA from *A. oreintalis*, synthesizes an additional amino acid 3, 5-dihydroxyphenylglycine (DHPG, a precursor of glycopeptide antibiotics) by the decarboxylative condensation of four molecules of malonyl-CoA. *Azotobacter* spp. use type III PKS, ArsB and ArsC, undergoes the condensation of malonyl-CoA with acetyl-CoA to synthesize phenolic lipids. ArsB synthesize 5-alkylresorcinol by C2-C7 aldol condensation, and ArsC catalyzed C5 oxygen-to-C1 lactonization to yield tri- and tetra-ketide pyrones. PKSA and PKSB from *A. thaliana*, catalyzed the condensation of ethyl- or methyl-malonyl-CoA units to form tri-and tetraketide pyrones, the precursor for sporopollenin, the principal constituent of exine in outer pollen wall[13]. Germicidin synthase (GCS) from *S. coelicolor* A3(2) condensed ethyl- or methyl-malonyl-CoA units to *β*-ketoacyl thioester to form germicidin pyrones from the cyclization of the triketide *β,δ*-

diketothioester. Type III PKSs, Steely1/2 from *Dictyostelium discoideum* catalyze the synthesis of signalling molecules like differentiation-inducing factors[88]. Intramolecular C6-to C1 Claisen condensation biosynthesizes the acylphloroglucinol skeleton of DIF-1.

PKS18 is a type III PKS characterized from *M. tuberculosis*, condense malonyl-CoA with aliphatic acyl-CoA (C_6 to C_{22}) to produce tri- and tetraketide pyrones[54,78]. Whole-genome sequencing of *M. tuberculosis*[164] and *M. marinum* give an insight that large numbers of polyketide biosynthetic genes cluster are conserved in pathogenic bacteria to synthesize novel metabolites with distinct functions[4,78,165-169] or complex lipids, which might be essential for the survival of organism[170]. This study is based on the cluster study of conserved type III polyketide synthase and associated genes in pathogenic bacteria belonging to the suborder of Corynebacterineae of the order Actinomycetales. We come to our hypothesis in mycobacteria, the similarity between two orthologous type III polyketide synthases MtbPKS18 and MMAR_2190 and then tried to study the role of this specific cluster from *M. marinum*.

2.2. MATERIALS AND METHODS

2.2.1 Materials

Gel-extraction kit, Plasmid extraction kit, and Ni^{2+}-NTA agarose resin were bought from Qiagen. Pierce™ BCA Protein Assay Kit was purchased from Thermo Scientific. Similarly, cloned *Pfu* polymerase, Q5-polymerase, DNA restrictions enzymes, and dNTPs were purchased from New England Biolabs (NEB). Fatty acyl-CoA substrates (starter and extender), were purchased from SIGMA. The cloning and expression vector pBluescript SK+, pET21c and pET28c were procured from Adgene and Novagen. The cloning host- *Escherichia coli* XL-1 blue and expression host- BL21(DE3) strains were obtained from American Type Culture Collection (ATCC). *Mycobacterium marinum* (strain ATCC BAA-535/M) provided by Y. Singh (IGIB, India). Luria Bertani broth (LB), 7H9 broth, 7H11 Agar base, Sauton's media, glucose, glycerol, sodium chloride, sodium dodecyl sulfate (SDS), Tris base, hydrochloric acid), chloroform, acetic acid, ethyl acetate, acetonitrile, methanol, formaldehyde, paraformaldehyde and antibiotics (kanamycin, ampicillin) were bought from Hi-Media, SIGMA, Merck-millipore, Thermo-Fisher Scientific.

2.2.2 Methods

2.2.2.1 . *In silico* study of mycobacterial clusters

The whole-genome sequences of *M. marinum* and other Corynebacterineae members were retrieved from the National Centre for Biotechnology Information (NCBI). Cluster analysis of nucleotide sequences was done using AntiSmash metabolite Shell to find the conserved type III PKSs clusters. The homology search of potentially related proteins of mycobacterial type III clusters was done using NCBI BLASTp program[171].

2.2.2.2 . Genomic DNA isolation

Genomic DNA was isolated from streaked colonies of a pure culture of *M. marinum*, harvested from sterile disposable toothpick on Middlebrook 7H11 agar plate suspended in 200 µl chloroform within 1.5 ml microcentrifuge tube containing tiny pores on a cap by using the boiling method. The tube was heated at 95 ˚C followed by intermittent vortex and addition of 100 µl chloroform again. Further, de-lipidation was enhanced by heating and repeating the process. To the dried pellet 100 µl of molecular grade water was added, and the tube was centrifuged at 11200 rpm for 5 min in a table-top centrifuge with the aerosol-tight rotor. The supernatant containing nucleic acid was transferred to the new microcentrifuge tube for a polymerase chain reaction. DNA was quantified using Nanodrop (Eppendorf, Germany).

2.2.2.3 . Polymerase Chain Reaction

Type III *pks* gene, *MMAR_2190* was amplified using the following set of forward and reverse primers: PS51 (fp) 5'TTCATATGAGCACCGCAGCCGAGGGC3' and PS52 (rp) 5'AAGAATTCCCGCGGCGGATGATGTCGAACAG3' respectively. The 50 µl reaction mixture contains 10 µl (5x) Reaction buffer, 10 µl (5x) GC buffer, 1 µl (17.3 ng/µl) genomic DNA, 1 µl (12.5 pmol/µl) of each primer, 4 µl dNTP mixture (2.5 mM), 0.6 µl (2U/µl) Q5 high-fidelity DNA polymerase, and 22.4 µl double distilled autoclaved water. Incubation of reaction mixture was done in a thermal cycler (Eppendorf, Germany) with an initial hot start at 98ºC for 2 min and then underwent 30 cycles of 30 sec at 98ºC, 45 sec at 60ºC and 50 sec at 72ºC, followed by final extension of 10 min at 72ºC.

2.2.2.4 . Cloning of amplified PCR product

The sample *(MMAR_2190* amplicon) containing 1× gel loading dye was loaded into the well with bromophenol blue as a tracker dye. It was resolved on 1 % agarose gel containing 0.5 µg/ml of ethidium bromide at 50 V in 1× TAE buffer. The resolved bands were visualized using UV transilluminator at 312 nm.

The amplified products with correct size were excised using a sterile scalpel blade after the preparative agarose gel. The excision was further processed through the protocol of the QIAquick gel extraction kit (Qiagen, Germany), and the product was eluted using the binding columns. The blunt-ended amplicon was phosphorylated using T4 polynucleotide kinase by incubation at 37 ˚C for 45 min and purified further using binding columns. pBluescript-Sk^+ plasmid DNA (1 µg) was digested using one unit of restriction endonuclease (*Eco*RV). The reaction was incubated in a final volume of 20 µl at 37 ˚C for 1 hr. After the completion of restriction digestion, 1µl of calf intestinal phosphatase (CIAP) was added to the reaction mixture and further incubated for 1hr. Also, it was purified using Qiagen DNA binding columns based on the manufacturer's protocol.

MMAR_2190 amplicon was ligated using 1 µl T4 DNA ligase into the linearized vector DNA with insert in 15 µl reaction containing insert vector ration 3:1 and 1× T4 DNA ligase buffer (containing ATP) at 16 ˚C for overnight and transformed into cloning host *Escherichia coli*, (*E.coli*) XL-1 Blue.

2.2.2.4.1 Competent cell preparation

The competent cells were prepared for various *E. coli* strains: cloning host (XL-1 Blue) and expression host-BL21 (DE3).

Competent cell preparation: *E. coli* strains

For *E. coli* based cloning and expression study, chemical competent cells were prepared for various *E. coli* strains: XL-1 and BL21(DE3) used in this study. Pure culture of respective strains was streak onto the LB agar plate. The primary culture was grown for an overnight in a 37 ˚C incubator shaker from the single pure colony. The inoculum (1 %) was transferred into 100 ml of media from the primary culture and incubated at 37 ˚C incubator-shaker until the optical density (OD) 600 nm reached 0.5 units. The cell growth was slowed by incubating the flask on ice for 10 min. The culture was centrifuged at 3000 rpm for 15 min at 4 ˚C. The cell pellet was gently resuspended in 20 ml chilled TF1 buffer and further incubated on ice for 10 min. Cells were pelleted again at 3000 rpm for 15 min at 4 ˚C, and the cell pellet was resuspended in 4 ml of TF2 buffer. It is aliquoted in 100 µl in the micro-centrifuge tube, snap-frozen in liquid nitrogen and stored at -80 ˚C.

2.2.2.4.2 Plasmid isolation

The plasmid was isolated from overnight grown culture by using alkaline lysis method[172]. An overnight grown culture from bacterial colonies was harvested by centrifugation to pellet down the cells. The cells pellet was resuspended in P1 buffer (50 mM Tris-HCl pH 8.0 and 10 mM EDTA and 100 µg/ml RNase A), lysed by the addition of an equal volume of P2 buffer (200 mM sodium hydroxide and 1 % SDS) and incubated at room temperature for 3-5 min. The plasmid DNA was renatured by the addition of neutralizing P3 buffer (3 M Potassium Acetate, pH 5.5) and was centrifuged at 13000 rpm for 10 min. The supernatant was transferred into a new microcentrifuge tube and was precipitated by the addition of 0.7 volume of isopropanol. The precipitated nucleic acid was pelleted down by centrifugation and further washed by 70 % ethanol and was air-dried. For further use, the air-dried plasmid was resuspended in 10 mM Tris-HCl (pH 8.0).

The construct containing pBluescript-Sk^+ and *MMAR_2190* was digested using NdeI and *Eco*RI-Hf restriction enzymes. The digested product was electrophoresed as described in section 2.2.2.4, and the *MMAR_2190* gene was excised. The correct size fragment of *MMAR_2190* was purified by Qiagen DNA purification columns and ligated onto the pET21C linearized vector with sticky end created by double digestion using the compatible enzyme.

2.2.2.4.3 Expression and purification of MMAR_2190 in *E.coli* BL21 (DE3)

The sticky end of *MMAR_2190* was sub-cloned into the pET21c (Novagen) expression vector and transformed into *E. coli* strain XL-1 Blue. The plasmid was isolated as described above, and the clone was confirmed by restriction digestion and Sanger nucleotide sequencing. The correct *MMAR_2190* gene with expression vector pET 21c plasmid was further transformed into the expression host of *E. coli* strain BL21 (DE3).

The pET21c plasmid carrying the gene (*MMAR_2190*) for protein expression was transformed in an expression strain of *E.coli* BL21(DE3) for the over-expression of protein. The recombinant *E.coli* cells with expression plasmid construct were grown in LB broth at 30 ˚C until an O.D of 0.6. A log phase the culture was induced using 0.5 mM IPTG and further incubated in a shaker at 22 ˚C for 8 hrs. The cells were harvested by centrifugation. The pellet was resuspended in lysis buffer (50mM Tris, pH 8.0, 10 % glycerol and 0.15 % NaCl) followed by 10 cycles of sonication (on and off cycles: 20-sec cycle and 10-sec rest) using 30 % amplitude. Next, the centrifugation at 17000 rpm for 30 min was done at 4 ˚C to remove the cellular debris. The crude recombinant protein from the supernatant was purified by using Ni^{2+} -NTA based affinity chromatography using 1 ml Ni^{2+} -NTA slurry/litre culture. The slurry was allowed to bind to the protein for 1hr at 4 ˚C and passed to the column using gravity flow followed by a wash of the resin with

wash buffer (50 mM Tris, pH 8.0, 10 % glycerol). After the complete removal of unbound protein, elution buffer with an increasing concentration gradient of imidazole (5, 10, 25, 50, 100, 75, 125, 150 mM imidazole, 50 mM Tris pH 8.0, 10 % glycerol, 0.15 % NaCl) was used to elute the pure protein. SDS-PAGE was run, and the fraction containing protein of interest was pooled, and 1 mM TCEP was added. The pooled protein was further purified by Resource Q (anion exchange, 6 ml) column using Buffer A (equilibration buffer 50 mM Tris pH 8.0, 10 % glycerol, 1 mM EDTA) with linear gradient of 0-1M NaCl in Buffer B (50 mM Tris pH 8.0, 10 % glycerol, 1 mM EDTA, 1M NaCl). The fraction from purified protein was pooled and concentrated using an Amicon Centriprep filter (25 kDa MWCo) followed by the addition of TCEP and stored at -70 ˚C.

2.2.2.4.4 Protein Estimation using BCA

Protein estimation was done using BCA method[173]. Triplicate of protein unknown concentration samples were taken and diluted to 5X and 10X. Similarly, the standard sample was prepared (0 to 2 mg/ml) diluting the BSA stock of 2 mg/ml. To the 25 µl unknown and standard sample 0.2 ml BCA working reagent, WR (WR prepared by mixing 50 parts of BCA Reagent A with 1 part of BCA Reagent B) was added according to the protocol of Pierce™ BCA Protein Assay Kit (Thermo). The mixture was incubated at 37 ˚C for 30 min. The absorbance was taken at 562 nm, and the calibration curve plotted was used to estimate the protein concentration.

2.2.2.4.5 Reconstitution of biochemical activity and product characterization

The enzymatic assay of protein was reconstituted using 100 µM each starter molecule (C_{10}-, C_{14}-, C_{16}-, C_{18}-, C_{22}-CoA) and 100 µM malonyl-CoA as extender molecule. The enzymatic reactions were carried out using 50 µg of purified MMAR_2190 protein incubated at 30 ˚C for 120 min. The reaction mixture was quenched with 5 % acetic acid,

and the products were extracted with 2×300 µl of ethyl acetate. The extracts were dried under vacuum. The reaction products were further resolved on analytical C5 reverse-phase ultra-fast liquid chromatography (UFLC) column using a binary gradient of solvent B (ACN) and solvent A (water) each containing 1 % acetic acid. The flow rate of 1 ml/min was maintained with a linear gradient of 5-30 % ACN for 5 min, 30-60 % for 10 min, 60-80 % for 10 min, 80-90 % for 10 min, 90-100 % for 10 min, 100 % for 10 min and 5 % for 5 min. The absorbance was detected at 280 nm using a Photo Diode Array (PDA) detector, and the fractions of resolved polyketides products were collected and concentrated suing speed-vac. The resolved polyketide products were characterized using SCIEX Triple-ToF 5600 HRMS.

2.3. RESULTS

Mycobacterium marinum is a close relative of *M. tuberculosis* sharing 3000 orthologs with an identical amino acid. It has a large genome size and shows a broad host range with several genes encoding polyketide synthases (PKSs) and non-ribosomal peptide synthetases (NRPS). Genome downsizing and lateral gene transfer with significant data referring evolution of other species of mycobacteria as *M. tuberculosis*, and *M.ulcerans*, made *M. marinum* as a model to understand determinants of pathogenesis[174]. It has 33 clusters of genes those coding for type I PKS and type III PKS, NRPS and terpene and amglyccycl biosynthesis, lans peptide (putative class II), bacteriocin and others suggested by using AntiSMASH secondary metabolites Shell[175]. There are diverse metabolites that are generated by bacterial type III PKSs, which are important in secondary metabolite biosynthetic pathways.

MMAR_2190 is predicted to belong to type III PKS cluster with two sulfotransferases- *MMAR_2191* and *MMAR_2192*, and one methyltransferase- *MMAR_2193* downstream to it. It has a single desaturase, *MMAR_2189*. MtbPKS18 from *M. tuberculosis* has been previously characterized to form α-pyrones with long and very long-chain acyl-CoA substrate[78]. MMAR_2190 interestingly shows 80 % amino acid identity with MtbPKS18, and from *M. tuberculosis* H37Rv. The clusters of both mycobacterial strains have a single desaturase and a single methyltransferase. *M. marinum* has two sulfotransferases, whereas *M. tuberculosis* has Rv1373. Rv1373 has been previously characterized as glycolipid sulfotransferase[176] and is orthologous to MMAR_2192. This type III pks cluster is conserved only in pathogenic corynebacterial genomes, as illustrated in Figure 2.1.

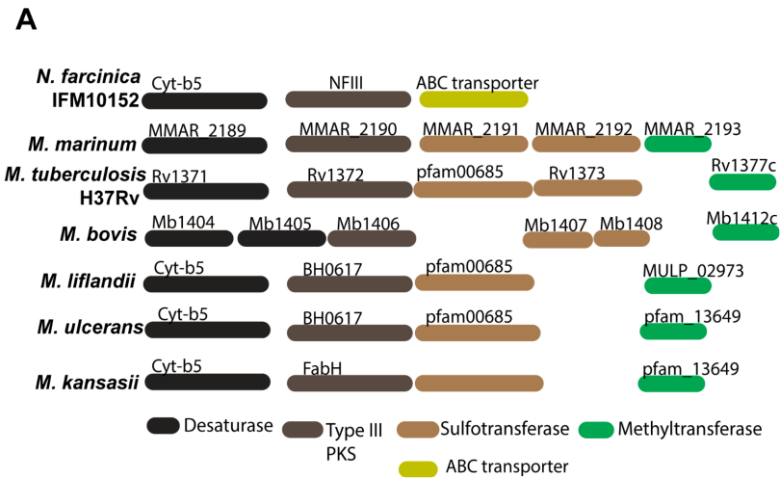
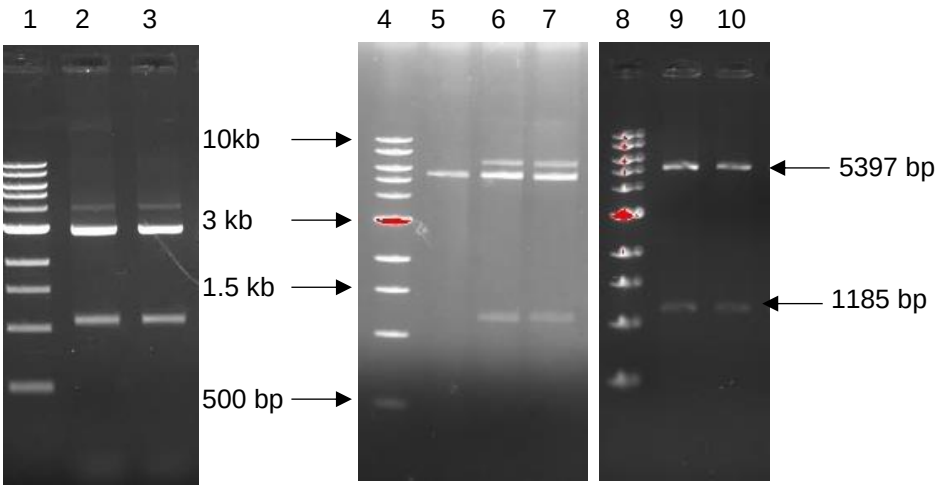


Figure 2.1: The cluster of genes orthologous in *M. marinum* and different species of suborder corynebacterineae. In silico cluster study of type III pks cluster using AntiSMASH Metabolites Shell showing conserved desaturase, type III PKS, sulfotransferase, and methyltransferase in pathogenic bacteria *N. farcinica IFM10152*, *M. marinum*, *M. tuberculosis H37Rv*, *M. bovis*, *M. liflandii*, *M. ulcerans* and *M. kansasii***.**

2.3.1 Cloning of type III PKS gene

Type III PKS gene, *MMAR_2190* was amplified using gene-specific primers using genomic DNA isolated from *M. marinum.* It was cloned into cloning host *E.coli* (Xl-1 blue) using blunt-end cloning vector pBluescript SK+ and the clones were double digested using restriction endonucleases NdeI-*Eco*RI generating fragments of 1182 bp and 2954 bp. The clone was further conformed by Plasmid DNA sequencing. The expression constructs KA4 (MMAR_2190 encoding plasmid) was generated by cloning digested 1182 bp fragments in pET21c by using NdeI *and Eco*RI sites. The restriction analysis of KA4 with NdeI-HindIII generated approximately 1185 bp and 5397 bp, as shown in figure 2.2**.**



1, 4, 8, 13.	1kb ladder (NEB)	6, 9, 10.	Ka4 (NdeI-HindIII)
2, 3.	MMAR_2190 clone (NdeI-*Eco*RI)		
5.	pET21c ((NdeI-HindII)		

Figure 2.2: Restriction digestion of MMAR_2190, pET21c, and Ka4 plasmids

2.3.2 Expression and purification of MMAR_2190 in *E.coli* BL21 (DE3)

Type III PKS circular plasmid Ka4 (plasmid containing *MMAR_2190* in pET21c), and pET21c (an empty vector) were independently transformed into BL21/DE3 strain of *E.coli* to express MMAR_2190. Small scale cultures were grown at 30 ˚C with 0.5mM IPTG induction at logarithmic (log) phase followed by protein overexpression at 22 ˚C for 8 hrs. The overexpressed proteins of the size of approximately 45 kDa were purified by Ni^{2+} - NTA affinity chromatography, as shown in figure 2.3 (lane 1-10). Large scales cultures were grown at the same condition, followed by IPTG induction and overexpression at 22 ˚C for 10 hrs.

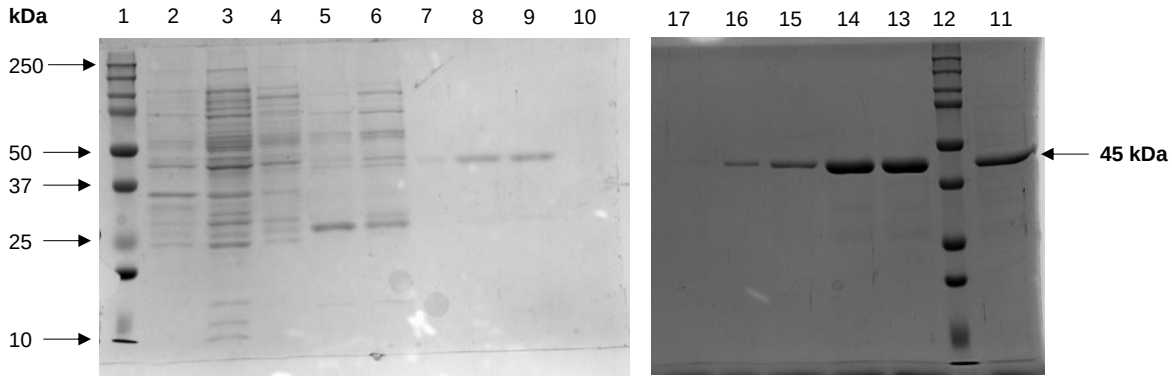


Figure 2.3: SDS-PAGE profile of Ni^{2+}-NTA purified MMAR_2190 protein

1, 12.	Protein ladder		
2.	Crude supernatant	6, 11, 13	10mM
3.	Pellet	7, 8, 14, 15.	25mM
4.	Flow through (FT)	16, 17.	50mM
5	5mM Imidazole eluent		

Most of the proteins was obtained as a soluble form. The protein was further purified by Ni^{2+} -NTA affinity chromatography as shown in figure (lane 1-17)

MMAR_2190 was further resolved on anion exchange column (Resource Q) and eluted using increasing the step gradient of NaCl buffer (Buffer B). MMAR_2190 mostly eluted with 250 mM NaCl to 300 mM NaCl, as shown in figure A2.2 in appendix 1**.**

2.3.3 Protein quantification using BCA

The purified protein was estimated using the BCA method. The calibration curve was drawn, as shown in figure 2.4, using standard BSA. From the estimated protein stock concentration, 50 µg was used for catalytic activity with different starters of acyl-CoA with extender malonyl-CoA.

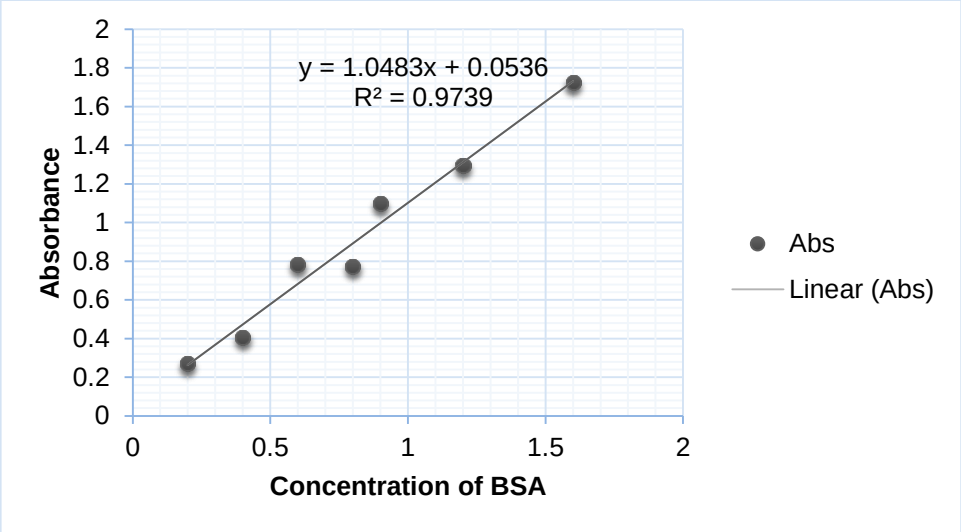
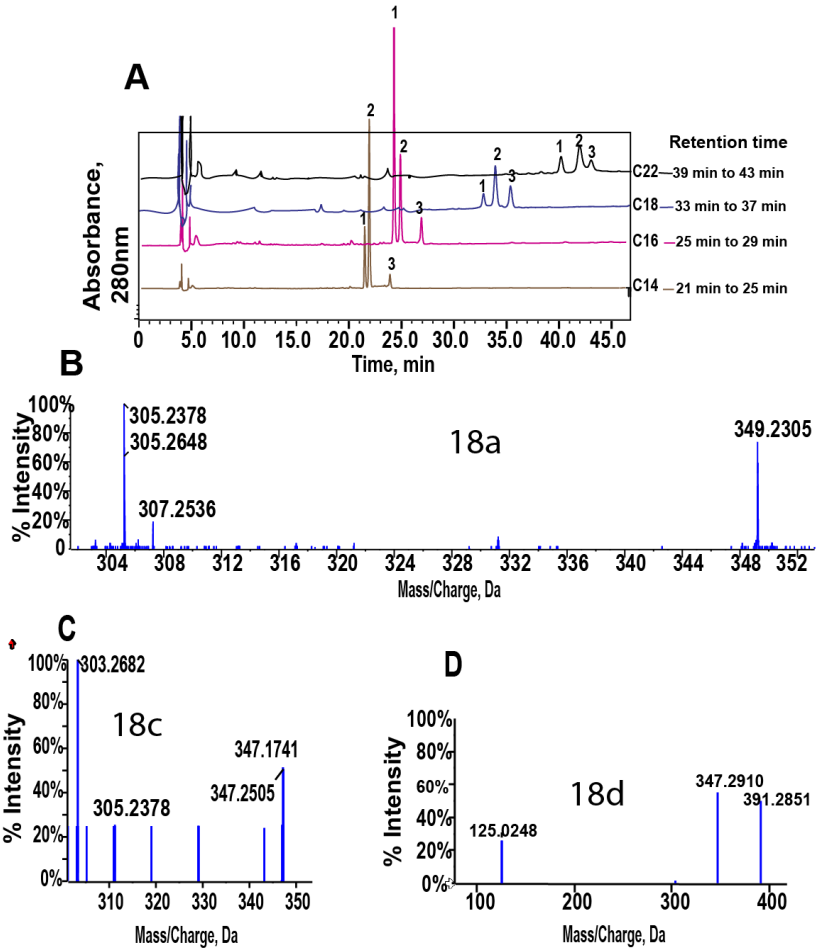


Figure 2.4: Standard calibration curve for protein estimation

2.3.4 Reconstitution of biochemical activity and product characterization

The *in vitro* reaction of MMAR_2190 with different carbon aliphatic chain starters and malonyl-CoA as extender showed the ability of the enzyme to accept various medium and long-chain acyl-CoAs (C_{14} to C_{22}) as shown in figure 2.5. The purified product peaks on the reverse phase C5 column using UFLC with PDA detector revealed multiple polyketide product peaks, particularly for long-chain aliphatic acyl-CoA. The retention time of the products was found to increase with the increasing length of fatty acyl-CoAs. The medium and long-chain acyl-CoAs were found to produce three significant product peaks 1, 2, 3, as shown in the chromatogram of figure 2.5A with multiple combinations of products, shown in the tabulated form. ESI-MS analysis from the different peak was found to give cyclized triketide *α*-pyrone (a), tetraketide *α*-pyrone (b), alkyl resorcinol (c), and acyl phloroglucinol (d). The representative profile with fragmentation stearoyl triketide *α*-pyrone (18a), palmitoyl tetraketide *α*-pyrone (16b), stearoyl resorcinol (18c) and stearoyl phloroglucinol (18d) are shown in the figure. The cyclized triketide *α*-pyrone, 18a with molecular ion peak of $[M-H]^-$ at m/z 349.2305 fragmented into $[M-CO_2H]^-$, the peak of

m/z 305.2378, and 307.2536 as shown in figure 2.5B. Whereas, the tetraketide *α*-pyrone (18b) of $[M-H]^-$ at m/z 391 formed from a starter- Stearoyl-CoA (C18) and extender-malonyl-CoA was not observed in our *in vitro* assay. The pattern of products profile from peak 1 from starter acyl-CoA, C-14 to C-22 with extender malonyl-CoA were found giving majorly triketide *α*-pyrone, tetraketide *α*-pyrone, and alkyl resorcinol. Peak 2 yield majorly triketide *α*-pyrone and tetraketide *α*-pyrone. Peak 3 yields cyclized triketide *α*-pyrone and acyl phloroglucinol. Stearoyl phloroglucinol (18d) gave a molecular ion peak of m/z $[M-H]^-$ at m/z 391.2851 which further fragmented giving fragments of m/z 347.2910 and 125.0248 similar to the pattern in tetraketide *α*-pyrone fragmentation except for $[M-CO_2H]^-$ peak is with higher intensity than $[C_6H_5O_3]^-$ peak as shown in figure 2.5D. We used the fragmentation pattern reported for non-methylated acyl phloroglucinol products from RppA, a protein of *S. griseus* in our study[177]. Stearoyl resorcinol at $[M-H]^-$ at 347.1741 which fragmented into $[M-CO_2H]^-$ peak at m/z 305.2378 and 303.2682 correspond to M-42 and M-44 fragmentation pattern as shown in figure 2.5C.



16a 18a	C16 C18	Triketide pyrone	321.2434 350.2305
16b	C16	Tetraketide pyrone	364.2328
16c 18c	C16 C18	Alkyl resorcinol	320.1241 348.1741
16d 18d	C16 C18	Acyl phlorogluci-nol	364.2551 392.2851

Protein	Acyl-CoA	UFLC Products		
		1	2	3
MMAR_2190	C14	a, b, c	a, b	a, d
	C16	a, b ,c	a, b	a,, d
	C18	a, c	a	a, d
	C22	a, b	a, b	a,d
	Products profile from UFLCpeaks 1,2,3			

Figure 2.5: Figure showing the biochemical activity of type III PKSs (MMAR_2190) to form various products (a- triketide pyrone, b- tetraketide pyrone, c- acyl resorcinol, d- acyl phloroglucinol). A. shows fractionation of products from *in vitro* assay with different C-chain starters (C14-CoA to C22-CoA) and malonyl-CoA as an extender to form products with Mmar type III polyketide synthase fractionated in UFLC in the form of peak 1, 2, 3 at 280 nm. B, C, and D show ESI-MS/MS analysis of reaction catalyzed with malonyl-CoA to form diverse products shown in the table

A. UFLC chromatogram with C_{14} to C_{22}
B. Triketide α-pyrone (18a)
D. Alkyl resorcinol (18c)
E. Acyl phloroglucinol (18d)

MMAR_2190 was found to have promiscuous substrate specificity with cyclization ability to form C_{16} triketide α-pyrone with molecular ions of $[M-H]^-$ at m/z 321.25 (a) and C_{16}-tetraketide α-pyrone $[M-H]^-$ at m/z 363.26 (b) similar to the close homologs protein of PKS18 from *M. tuberculosis*. Besides it also catalyze the condensation of starter with an extender to form C_{16}-phloroglucinol $[M-H]^-$ at m/z 363.26 (c) and C_{16}-resorcinol $[M-H]^-$ at m/z 319.27 (d) as shown in figure 2.6. Moreover, the C_{16}-triketide-α-pyrone, and C_{16}-tetraketide α-pyrone formation occur by lactonization. The C_{16}-phloroglucinol formed by using Claisen condensation and C_{16}-resorcinol products are formed by aldol condensation, as summarized in figure 2.7.

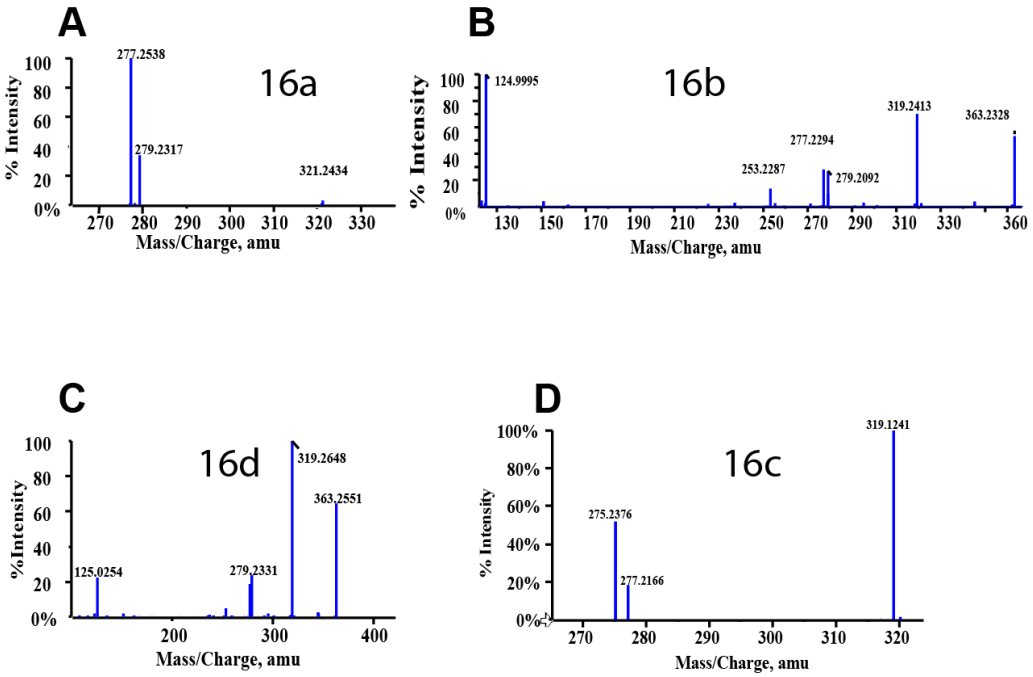


Figure 2.6: Figure showing the biochemical activity of type III PKSs (MMAR_2190) to form different products (A, triketide pyrone- 16a; B, tetraketide pyrone- 16b; D, acyl resorcinol- 16c; C, acyl phloroglucinol- 16d) observed in mass spectrometry.

A. C_{16}-triketide α-pyrone (16a)
B. C_{16}-tetraketide α-pyrone (16b)
C. C_{16}-acyl phloroglucinol (16d)
E. C_{16}-alkyl resorcinol (16c)

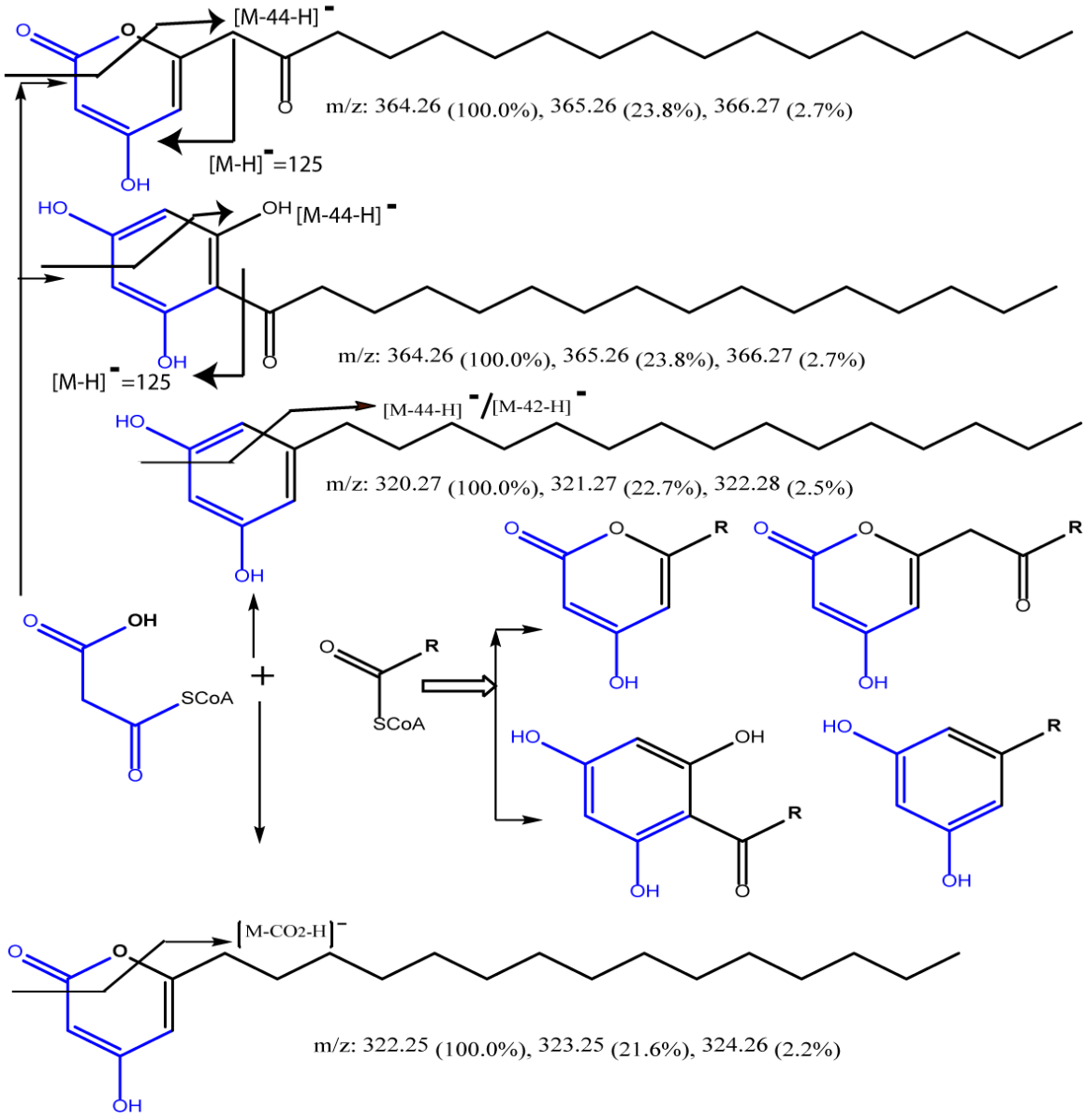


Figure 2.7: Schematic representation of the reaction module of the formation of polyketide products using separate starter and malonyl-CoA as an extender by MMAR_2190. Reactions summarized with C_{16}-starter (palmitoyl-CoA shown in blue) and extender (malonyl-CoA drawn in black) to form a-C_{16}-triketide α-pyrone $[M-H]^-$=321.25 b-C_{16} tetraketide α-pyrone $[M-H]^-$=363.26, c-C_{16} phloroglucinol $[M-H]^-$=363.26 and d-C_{16} resorcinol $[M-H]^-$=319.27 in negative ion mode. The actual m/z of the precursor of molecule in neutral mode is shown in the figure with respect to relative abundance as depicted by Chemdraw. The figure shows the formation of four kinds of phenolic compounds using starter acyl-CoA as a starter (R=alkyl group attached to the acyl-CoA) and malonyl-CoA as an extender. C_{16}-triketide α-pyrone in negative ion mode gives a -44 peak with the precursor mass. C_{16} tetraketide α-pyrone gives a low intensity -44 peak and an additional peak of $[M-H]^-$=125. C_{16} phloroglucinol gives a high intensity -44 peak compared to the additional peak of $[M-H]^-$=125. C_{16} resorcinol gives -44 and -42 peaks. MMAR_2190 in general use this reaction module to form various polyketide products. The backbone derived from starter molecule is shown in black and extender, malonyl-CoA is shown in blue.

2.4. DISCUSSION

M. marinum causes tuberculosis like symptoms in aquatic animals like fish and amphibians. It is a close relative of human pathogen *M. tuberculosis H37Rv*. The genome sequencing of *M. marinum* reveals it as an appropriate pathogen to understand tuberculosis disease. The genome downsizing and lateral gene transfer from 5424 genes in *M. marinum*[174] to 3999 genes in *M. tuberculosis H37Rv* make a bacterium a specialized pathogen to understand the biology of a member of Corynebacterineae family. The functional study of different type III PKSs proteins from microbes contribute to a better understanding of new dimensions for the product formation and chemical diversity of polyketides[4,70,88,165,178-181]. Diverse products are results of successive decarboxylative condensation of malonyl-CoA, or coenzyme A (CoA) derived into a keto or hydroxyl functional group of muti- carbon compound[30].

Our *in silico* analysis of MMAR_2190 is a type III PKS from *M. marinum* share 55 % protein similarity to CHS-like proteins of plants, fungi, and bacteria. It shows 79.9 % with Rv1372 protein from *M. tuberculosis H37Rv*. The biochemical study of MMAR_2190 protein showed remarkable specificity of the proteins for long-chain aliphatic-CoA analogues to give long-chain *α*-pyrones similar to that observed for Rv1372 from *M. tuberculosis*[78]. Several bacterial and plant homologs are reported to utilize the long-chain precursor molecules[67-70,182]. MMAR_2190 was found to have promiscuous substrate specificity and unprecedented cyclization. Similar to its homolog Rv1372 in *M. tuberculosis*, MMAR_2190 was found to accept only malonyl-CoA as an extender when primed *in vitro* with acyl-CoA from C_{12} to C_{22} to form non-methylated aliphatic triketide and tetraketide pyrones, alkyl phloroglucinol and alkylresorcinols as products. Enzymes

from *Azotobacter*[70] and *Neurospora*[80,81] *are rep*orted to form resorcinolic lipids. PKS-A and PKS-B are the two type III polyketide synthases from *Arabidopsis thaliana* involved in the pollen development and sporopollenin using precursor type III polyketide molecules triketide and tetraketide α-pyrone[77].

Similarly, resorcinolic metabolites, α- and β-leprosols are reported from pathogenic mycobacteria such as *M. leprae*[183]. Phenolic type III polyketides *A. vinelandii* play a crucial role during exine formation in dormant cells[70]. It can be speculated that lipidic type III polyketides in mycobacteria play roles in cell envelope modulation and growth. In this chapter, we have functionally characterized MMAR_2190, a novel type III PKS from Mmar that catalyzes the formation of diverse non-methylated metabolites. The protein displays unprecedented cyclization capability to biosynthesize variably cyclized chemically distinct and functionally diverse metabolites from a single protein core.

Chapter 3

Structural Determination of Type III Polyketide Synthase

CHAPTER 3: STRUCTURAL DETERMINATION OF TYPE III PKS

3.1. INTRODUCTION

Type III polyketide synthase is a homodimeric protein that catalyzes the biosynthesis of polyketide products using the single active site of each monomer. The architecture of the protein shows "thiolase fold" having five layered *αβαβα* topologies in secondary structural elements[1,184]. Understanding of the mechanism behind a broad range of biological activities of type III PKSs is based on the crystallographic studies of chalcone synthase msCHS from *Medicago sativa*[56]. Type III PKSs belong to plant chalcone synthase (CHS) superfamily of condensing enzymes each having mechanistic uniqueness to generate chemically distinct polyketide cores that get variously modified into bio-active natural products. The product formation is based on catalysis by type III PKS proteins with the unique substrate-binding pocket with the presence of coenzyme A (CoA)- binding tunnel links the surrounding solvent with an active site and catalytic residues[56]. The active site of protein has conserved amino acid forming catalytic triad consisting of Cys-His-Asn. Type III PKS, in general, uses an acyl-CoA as a substrate and undergoes condensation with extender malonyl-CoA or methyl malonyl-CoA or derivatives, with a difference in the mode of cyclization resulting in divergent products.

The functional diversity of products from type III PKS in plant and microorganisms depends on the selection of different starter molecules, the number of malonyl-CoA condensation, the mechanism of cyclization along with the specificity of individual members of the family controlling cyclization and termination of polyketide chain

length[1,54,56,68,177]. Our biochemical characterization of MMAR_2190 protein from *M. marinum* revealed medium and long-chain starter specificity of the protein with metabolite divergence (Chapter2). From homology-based comparisons of MMAR_2190 with the crystal structure of MtbPKS18, it was not known how single enzyme active site accommodates substrate and orchestrates series of products in MMAR_2190. To gain insight into the function of MMAR_2190, we determined the crystal structure of the protein at a resolution of 2.3 Å, in collaboration with Dr Krishnan Vengadesan group at Regional Centre for Biotechnology (RCB), India.

3.2. MATERIALS AND METHODS

3.2.1. Materials

HiPrep 16/60 Sephacryl S-100 size exclusion column (GE Healthcare), Polyethylene glycol 3000 and HEPES buffer, PEG400 were bought from SIGMA. Other materials used in the study are listed in chapter 2.

3.2.2. Methods

3.2.2.1.1. Protein purification

The purified MMAR_2190 protein from chapter 2 was pooled, concentrated and loaded onto HiPrep 16/60 Sephacryl S-100 size exclusion column (GE Healthcare). The column was pre-equilibrated with elution buffer buffer C (50 mM Tris pH;8.0, 150 mM NaCl, 1 mM DTT, 1 mM EDTA). Single eluted MMAR_2190 peak was collected and compared with low molecular weight standards. Further, the fraction was concentrated up to 10 mg/ml using an Amicon ultra-centrifugal device (Millipore) with 10 kDa molecular cut-off.

3.2.2.1.2. Crystallization

Purified MMAR_2190 protein was used for initial crystallization trials using available crystal screens in 96-well plates using an automatic liquid handling system (TTB labtech) to obtain defined shaped crystals in multiple conditions. The initial conditions were further optimized using 24 well plates by using the hanging-drop vapour diffusion method at temperature 293 K varying pH, salt and precipitant concentration. A supersaturated solution of protein and mother liquor was made by mixing in a 1:1 ratio making 2 µl drop suspended over 1ml reservoir solution. The optimized conditions were obtained for

diffraction quality crystal using 20 % (w/v) Polyethylene glycol 3000 and 100 mM HEPES at pH 7.5 with protein concentration 10 mg/ml. The crystal was soaked in a cryoprotectant solution containing reservoir solution supplemented with 30 % PEG400 followed by flash freezing in liquid nitrogen prior to data collection.

3.2.2.1.3. Data collection and refinement statistics

The crystals were diffracted onto the X-ray Diffractometer. The data from the protein crystals were collected in total 360 frames with oscillation step 0.5˚, 3 second exposure time and crystal-to detector distance of 170 mm on synchrotron radiation beamline BM14 at ESRF, Grenoble equipped with MAR 225 CCD detector. Further, the diffraction data were indexed and integrated with XDS[185] and the scaling and merging program SCALA using the CCP4 suite[186]. Macromolecular phasing of MMAR_2190 protein was done through Phaser to solve the structure with the monomer of *M. tuberculosis* PKS III (PDB: 1TED) as a search template using the molecular replacement method. The initial round of iterative refinement was done using REFMAC[187] and model building, and more real and space refinement was done using Coot[188]. Anisotropy due to diffraction limit and attenuation of diffraction was corrected using STARANISO software[189] followed by the final cycle of refinement.

3.3. RESULTS

3.3.1. Insights to the crystal structure of MMAR_2190

Single eluted MMAR_2190 peak from HiPrep 16/60 Sephacryl S-100 size exclusion column was eluted at 41 ml with a molecular weight of approximately 45 kDa (Figure: 3.1: A), and the crystal from the protein was found to have hexagonal symmetry (Figure: 3.1: B)

M. marinum, MMAR_2190 protein was crystallized in the absence of any substrate, products or intermediate with a resolution of 2.3 Å. The data collection and refinement statistics are given in table 3.1. The structure of the protein was solved by a molecular replacement method using MtbPKS18, type III PKS (PDB ID: 1TED) from Mtb[54]. The close homologs protein from two mycobacterium species share 79 % sequence identity. The asymmetric unit in crystal contains tightly packed two molecules forming homodimer related by two-fold rotational symmetry (figure 3.1: C). Each monomer has 393 amino acid residues arranged in a typical $\alpha\beta\alpha\beta\alpha$-thiolase fold with 11$\alpha$-helices and 13$\beta$-strands (figure 3.1:D) and the catalytic triad deeply buried inside the catalytic pocket consisting of Cys175, His313 and Asn346 (figure 3.1:E) position and conformation are similar to the known structure of type III PKS, PKS18 from Mtb (1TED)[54], *Dictyostelium*-Steely1 (2H84)[88], *Neurospora crassa*-PKSIII (3E1H)[81] and *Streptomyces coelicolor*-THNS (1UoM)[68] with RMSD of 0.8, 2.1, 2.4 and 2.0 Å respectively for the Cα-atoms. A ribbon diagram of a dimer of MMAR_2190 from *M. marinum* has approximate dimensions of 58×56×29 Å with each monomer enclosed 14.2 % (2158.9 Å^2) total surface area buried. Like 1TED, a crystal structure of PKS18 conserved amino acid of a catalytic triad with a CoA-binding tunnel of 16 Å length connect active site cavity with an outer surface of the

protein. Superimposed monomers of MMAR_2190 has an RMSD of 0.1 Å for Cα-atoms, and dimeric form has apparent electron density localized at the active site Cys175 and its tunnel. The interface between dimer is stabilized by polar, non-polar and salt bridge interactions. The data collection and refinement statistics of crystal are given below.

Table 3.1: Data collection and refinement statistics of MMAR_2190[#*]

Parameters	**Values**
Wilson B-factor	49.4
R-merge	0.058 (0.560)
CC1/2	0.999 (0.576)
Refinement	
No. reflections (working / test set)	26660 / 1552
R_{work} / R_{free} (%)	19.79 / 21.74
No. atoms	5403
Protein	5263
Ligand	64
Water	76
RMSD	
Bond lengths (Å)	0.012
Bond angles (°)	1.57
Ramachandran plot (%)	
Favoured	97.07
Allowed	2.93
Outliers	0.00
Average B-factor (Å^2)	58.64

Protein	58.69
Ligands	72.39
Water	43.62
PDB code	5ZQ9

[#] Diffraction data was generated using a single crystal, and the values in the parentheses are for the highest resolution shell.

*Refinement statistics and ellipsoidal completeness of crystal calculated after anisotropy correction using STARANISO.

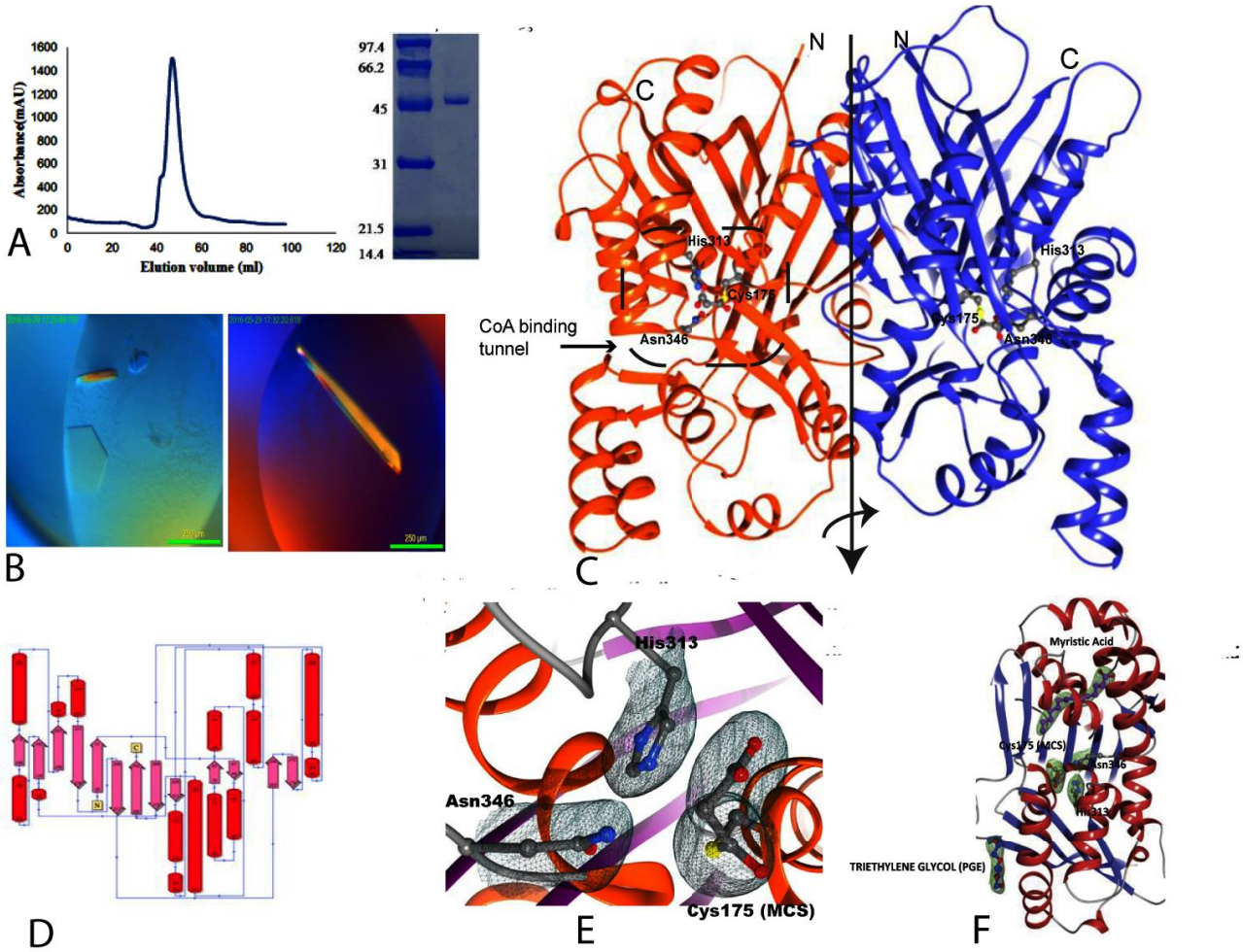


Figure 3.1: Insights to the crystal of MMAR_2190 with a study on the tertiary structure of the dimeric protein. A: Elution profile in gel-filtration chromatography and SDS-PAGE of purified protein B: Crystal (hexagonal shaped) from100 mM HEPES buffer pH 7.5 containing 20 % (w/v) PEG 3000), C: A ribbon diagram of MMAR_2190 with two-fold symmetry showing two monomers (orange and blue respectively) labelled N- and C-terminal with residues of the active site and CoA-binding tunnel, D: Thiolase fold of MMAR_2190 with 11 α-helices and 13 β-strands, red and pink cylinders respectively, E: A close-up view of residues of a catalytic triad consisting of Cys 175 (malonyl-Cysteine, MCS), His313 and Asn346. Electron density Map (2Fo-Fc; with Cutoff level 1sigma) is shown in black mesh, and F:

Structure of monomeric MMAR_2190 was shown with an electron density of active side residues, myristic acid (MYR, violet) and triethylene glycol (PEG, blue) (2Fo-Fc; with Cutoff level 1sigma).

To understand the mechanism of uptake of a substrate to give varied products, the dimeric crystal structure of MMAR_2190 (PDB: 5ZQ9, unreleased) was analyzed for protein-ligand interaction using Biovia Discovery Studio 4.5 (Figure 3.2) by docking individual ligands myristate (MYR), triethylene glycol (PGE1) and acetate (ACT1) separately to the protein.

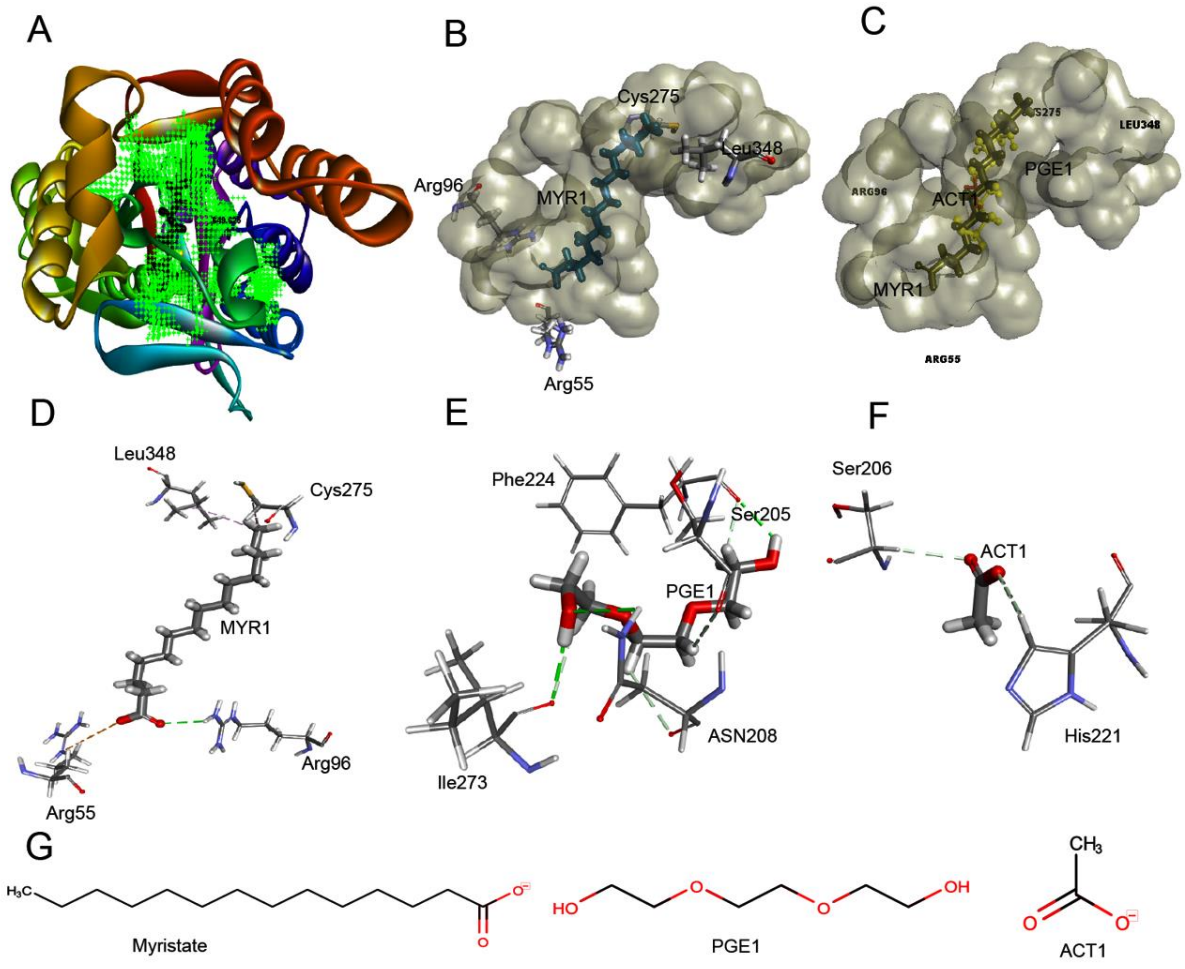


Figure 3.2: Protein-ligand interactions in the crystal structure of MMAR_2190 using docking studies. A: Electron density of binding cavity (green region) in protein monomer showing the fitting of ligand, MYR1 deep inside the cavity, B: Fitting of ligand, MYR1 to the binding cavity C: Binding cavity showing separate fitting region for MYR1, PGE1 and ACT1, D: Interaction of MYR1 with protein using amino acid residues Cys275, Leu348, arg55 and Arg96, E: A close-up view of the interaction of PGE1 to the protein. Phe224, Ser205, Ile273, and Asn208 are the interacting amino acid residues, and F: Acetate (ACT1) is shown interacting with the protein using His221 and Ser206 G: Structure formula of myristate, PGE1, and Act1 with functional groups are shown in red.

The native crystal of MMAR_2190 portrayed different electron density maps close to the active site, as shown in figure 3.2. Our docking study of MMAR_2190 with different ligands shows various residues of active sites, catalytic sites and gating residues get conserved similar to that of FabH and KSIII family of proteins suggesting the functional importance of amino acid residues. Different ligand MYR, PGE1, and ACT1 sequestered within the protein crystal interacting with some of the amino acid residues at the active site (Cys175, His313, and Asn346), catalytic site (Arg55, Arg96, Ser205, Ser206, Asn208, His221, Phe224, Ile273, Cys275) and near to the cavity volume (the gatekeeper amino acid- Leu266, Leu348, Ile319 and Leu223) are predicted to play an important role in functional diversity of products. Further, the importance of amino acid can be deduced based on the comparison of the sequence alignment of MMAR_2190 with MtbPKS18 as shown in figure 3.3. It shows most of the amino acid are conserved in two orthologous protein, and few are replaced with similar and weakly similar amino acid residues.

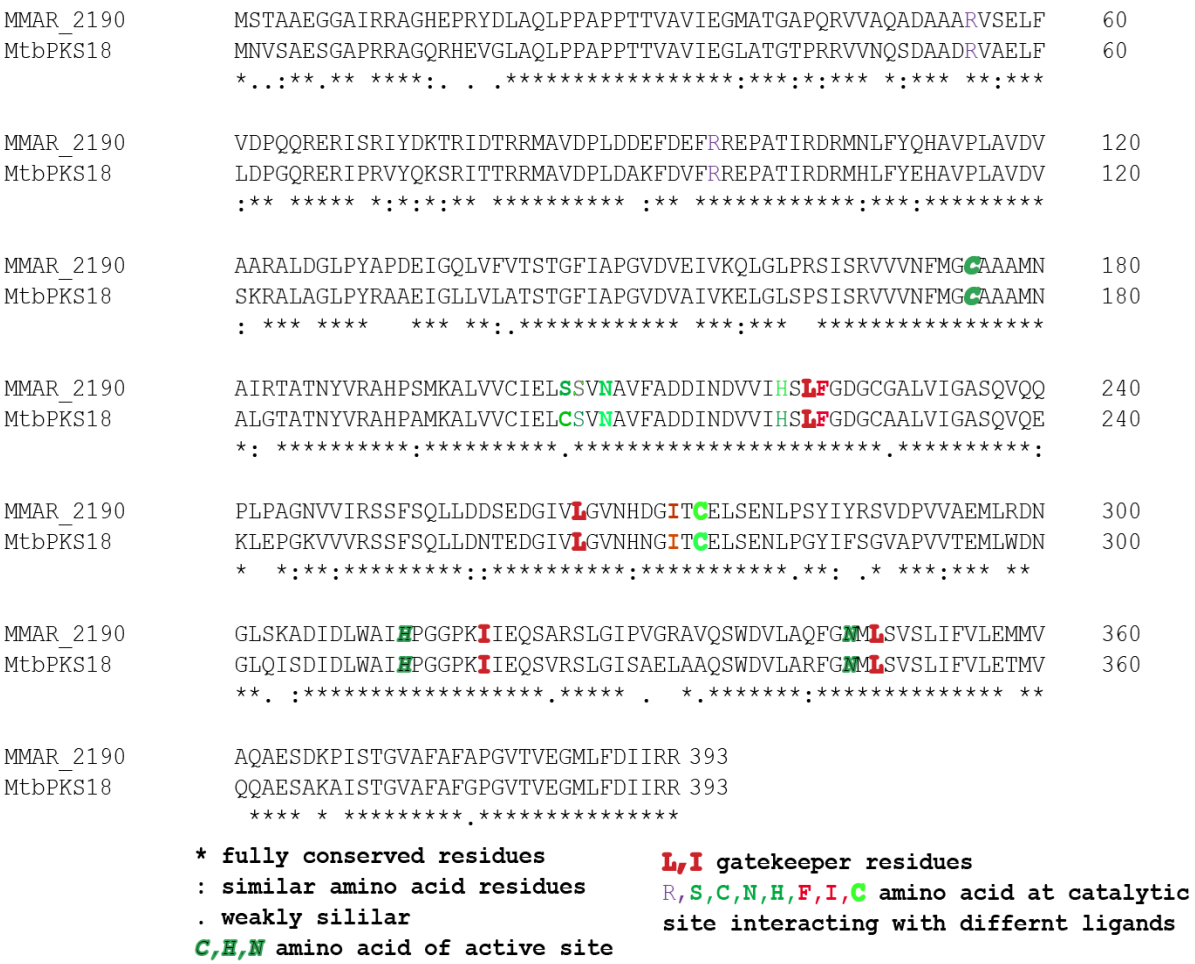


Figure 3.3: Sequence alignment of MMAR_2190 and MtbPKS18 protein. Alignment depicts fully conserved amino acid, similar amino acid, and weakly similar amino acids in two orthologous proteins. Different amino acids in the active site are shown in dark green colour (Cys175 (*C*), His313 (*H*), and Asn346 (*N*). Leu266 (L266), Leu348 (L348), Ile319(I319) and Leu223(L223) were found conserved in both orthologs shown in dark red colour act differently in MMAR_2190. Arg55(R55), Arg96(R96), Ser205(S205), Ser206(S206), Asn208(N208),His221(H221), Phe224(F224), Ile273(I223) Cys275(C275) are the amino acid predicted to interacts with different ligands. All coloured amino acids in their respective positions are shown fully conserved in both proteins except Ser205 is found replaced by Cys205 in MtbPKS18.

Our crystal structure depicts the addition of the malonyl unit to the catalytic Cys175 with no apparent conformational change observed at the active site (Figure 3.1: E). Similar active site cysteine modification was also observed in CerJ, a ketosynthase from *Streptomyces tendae*[190]. It acts as the first stage of reaction where malonyl moiety is trapped to the catalytic pocket in anticipation of the fatty acyl receptor.

3.3.2. Structural overview of the gating mechanism in regulating active site volume

The comparative sequence and structural analysis have provided information to dissect the functional relevance of the active site and crucial structural residues of the MMAR_2190 protein. Regulation of the gating of active site volume in MMAR_2190 can be described in terms of the presence of two tunnels- tunnel A and tunnel B, connecting the catalytic cysteine to the active site cleft and the rotamer study. The delineating residues of both tunnels are largely hydrophobic and accommodate long-chain acyl substrates. Most of the type III PKS, CoA substrates was reported to bind to the first tunnel, and another tunnel was found to bound myristic acid or long-chain intermediate by accommodating or release of the product. The latter tunnel extends toward the surface of protein through another but a narrow hydrophobic tunnel. The presence of hydrophobic residues and the quality of electron density at a position similar to the substrate-binding tunnel reported for MtbPKS18[54]. The approximate length and volume of the tunnel cavity also seem identical in both MMAR_2190 and MtbPKS18. Based on the fact myristic acid was modelled in both monomers (figure 3.4), and the proteins show a remarkable difference in surface openings with 239 $Å^2$ and 122 $Å^2$ respectively as calculated by CasTp[191]. The difference in orientation of carboxyl-terminal of myristic acid seems to dictate side-chain amino acid orientation. Few of the amino acid residues; Leu59, Leu88, Phe211, Pro87, and Arg96 opens the tunnel in MMAR_2190 whereas, the tunnel path is blocked in case of MtbPKS18 (figure 3.4). The change in relative positions of Phe211 (6.8 Å), Leu88 (4.8 Å), Arg96 (4.26 Å), Leu59 (3.69 Å), Pro87 (1.96 Å) away from the tunnel cavity widens the tunnel opening for the substrate binding in MMAR_2190 whereas, in MtbPKS18 tunnel is partially occluded by Phe211. Arg55 another amino acid at the tunnel cavity, the opening of the tunnel is disordered. Phe211 and Arg55 likely act as a regulator for closing and opening of the cavity in MMAR_2190. Although MMAR_2190 and Steely1 structures

resemble each other with fold and structure, the latter has a substitution of bulkier Phe191 instead of Ala209 at the tunnel cavity. It results in clashes with myristic acid, the ligand of MMAR_2190 and MtbPKS18. Similarly, the shifts in the Cα backbone is observed in the superimposed amino acid Try2842 at the position of Pro87 of MMAR_2190 and MtbPKS18.

Contrary to these facts, the crystal structure of PKSIIINc from *N. crassa*, and THNS from *S. coelicolor* A3(2) subtle difference in the similar tunnel is found as MMAR_2190, but there is a difference in the relative position of amino acid. Similarly, PEG molecule and triethylene glycol (PGE) were modelled at the equivalent position of MMAR_2190. The former has significant electron density and adjacent to the PEG; a narrow cleft connects the surface with the active site cavity was observed. Although the significance is not known equivalent tunnel are found in other types of type III PKSs. The narrow cleft probably acts as an acyl chain substrate reservoir that needs experimental validation.

In MMAR_2190, an interesting observation was made as both the tunnels were closed towards active site cysteine (Figure: 3.4 A). The lining residues toward the opening gate of the tunnel are hydrophobic. Leu266 and Leu348 are found to present on tunnel A gate whereas tunnel B with Ileu319 and Leu223. Structural comparison with MtbPKS18 (Figure: 3.4D) shows both tunnels are arranged at perpendicular to each other and merge at active site with no hindrance observed in the opening at both ends. Although most of the residues of the lining the two-cavity volume are almost identical to both of the structures. The opening and closing of the tunnel are decided by the gatekeeper residues such as Phe224 and Ile319 (figure: 3.4 C). Compared to the MtbPKS18, the relative orientation of the side chain differs in MMAR_2190. The channel is blocked in case of rotation of Phe224 through 90˚ from that of MtbPKS18, altering the side-chain conformation of gatekeeper residue Leu348.

Similarly, the rotation of the side chain of Leu348 through 90˚ towards tunnel shows steric clash with a rotated side chain of Phe224 resulted close to the channel A opening. The minor change in the side-chain orientation of Ile319 in tunnel B of MMAR_2190 results in the closure of the active site end of the CoA-binding tunnel because of the hydrophobic interaction with Leu223. Substitution of Ile319 with small alanine may widen the CoA tunnel. Change in rotamer of Leu348 and Ile319, two important gatekeeper residues both the channel get merged near to the active site cysteine (figure: 3.4C) similar to that observed of merging channel in case of MtPKS118 two tunnel as reported by Sankaranarayanan *et al*.

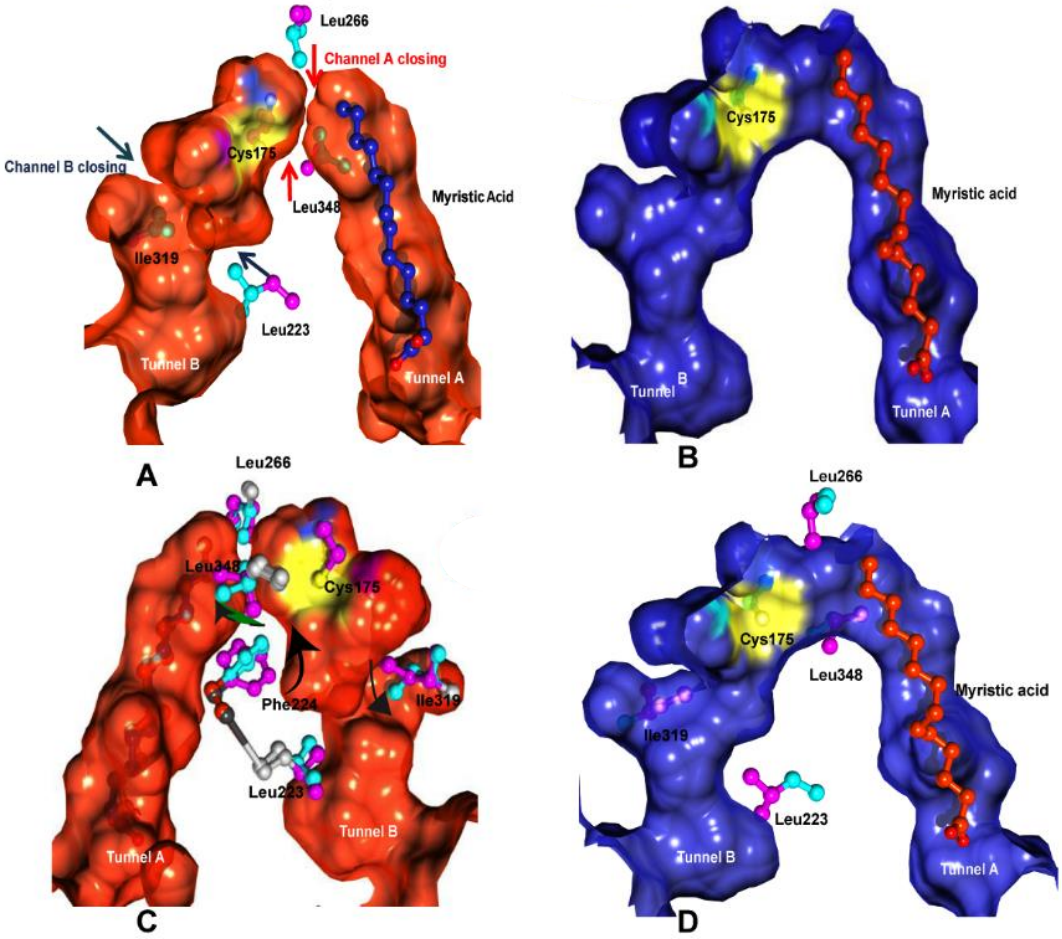


Figure 3.4: Comparison of the residues lining the active site cysteine through two-cavity volume. **A**: The surface diagram is sliced to show the active site with two tunnel A and B meeting at the catalytically active Cys175, and the opening is marked with an arrow (blue and red respectively for tunnel B and A). Leu266 and Leu348 are present in tunnel A with myristic acid bound to the cavity (shown in blue) whereas, Ileu319 and Leu223 are the gatekeeper residue shown. **B**: Reconstruction of MMAR_2190 tunnel (PDB: 5EZQ9, unpublished) after orientation change of the gatekeeper residues as per 1TED structure, both tunnels were found arranged perpendicular to each other and merge at active site Cys175. **C**: The rotamer study based on the orientation of gatekeeper residues in MMAR_2190 (5ZQ9) with corresponding amino acid MtbPKS18 (1TED) is shown. Residues of MMAR_2190 are shown in cyan and MtbPKS18 in magenta. Rotation of Phe224 (shown with black arrow) results in a change in side-chain conformation of Leu348. Whereas, Leu348 rotated through 90˚ (green, black arrow) shows the steric clash with Phe224 closing of the channel A opening. In channel B rotamer of Ile319 (shown in black arrow) change the conformation of Leu223 by hydrophobic interaction close the channel A opening. **D**: Change in rotamer of Ile319 and Leu348 both the channel get merged to the Cys175**.**

3.4. DISCUSSION

MMAR_2190 protein from *M. marinum* displays a remarkable ability to tethered acyl-CoA thioesters as a starter molecule to condense with extender malonyl-CoA undergoing elongation steps with the synthesis of linear triketide and tetraketide as polyketide intermediates. Further, Aldol condensation or Claisen condensation along with intramolecular cyclization result in the biosynthesis of functionally important phenolic compounds (described in chapter 2). There are dozens of studies made on type III PKS from plant and bacteria to understand the specific selection of starter molecule followed by chain extension, but the cyclization specificity is still a mystery. To understand the unusual specificity of MMAR_2190 protein to catalyze diverse class of functionally important molecules, we generated three-dimensional protein crystal structure. We probed the ligand myristic acid (MYR1), and malonated cysteine onto the structure based on the previous study on homologous protein MtbPKS18 from *M. tuberculosis.* Our structural characterization of the MMAR_2190 protein elucidates the characteristics features of the type III PKS enzyme. It has a presence of three-dimensional thiolase fold consisting of internal active site residues, Cys175-His313-Asn346 making catalytic triad[62,57,201] similar

to other types of type III PKSs homologs *M. tuberculosis* MtbPKS18 (1TED)[54], *Dictyostelium* STEELY1 (2H84)[88] *N. crassa* PKSIII (3E1H)[81] and *S. coelicolor* THNS (1U0M)[68]. The iterative condensation of acetyl units to CoA linked starter unit followed by linear cyclization occurs on the same active site cavity. The acetyl unit is derived from the malonyl-CoA (MCS), binds to the Cys derived thiolate nucleophile, and left to the catalytic pocket. The ligand myristic acid (MYR1) binds to the long acyl-CoA binding tunnel near to the floor of the PKS active site with a carboxyl group close to both the active site Cys residue. A similar modification was also reported in the crystal structure of 3-Hydroxy-3-methylglutaroyl CoA synthase *from B. juncea*[192]. Near to the active site of MMAR_2190, there is the presence of 20 Å long CoA-binding tunnel extends from the active site to the surface of the protein is similar to the MtbPKS18[54]. MMAR_2190 protein works in the dimeric form. The CoA-linked starter/ ligand (MYR1) covalently gets loaded onto the catalytic cysteine by the opening of channel A into another narrow tunnel extending to the surface of the protein. Leu266 and Leu348 are the gatekeeper residues for tunnel A whereas, Ileu319 and Leu223 are responsible for the opening of tunnel B. It is presumed that both myristic acid and malonated cysteine bind to the cavity of protein but the change in orientation of Leu348 and Phe224 decide opening and closing of channel A. The hydrophobic interaction and change in rotamer of Ile319 and Leu223 decide the channel B opening. Lu348 rotated through 90˚toward the cavity cause a steric clash with Phe224 and close the channel A and the product formation only occurs on cavity B. Similarly, the minor change in orientation of Ile319 change the hydrophobic interaction between Leu223 closing the channel B.

In this chapter, MMAR_2190 uniqueness is described in terms of catalytic malonated Cys175 and based on confirmation of gatekeeper residues. The decarboxylation of malonyl-CoA as an extender produces activated acetyl unit that goes repetitive

condensation with cysteine bound starter unit. The formation of triketide and tetraketide intermediates and cyclization to form four type of products depend on the conformation change associated with change in rotamer or side chain along with the gating mechanism[193]. For alkyl triketide and alkyl tetraketide *α*-pyrone formation, the triketidie and tetraketide intermediate undergo intramolecular lactonization from C5-oxygen to C1 as described for *p*-coumaroyl triacetid acid synthase[194], and MtbPKS18[78]. Similarly, the Claisen condensation from C6 to C1 like the one reported in CHS[62] and STEELY2[88] pattern is followed in the biosynthesis of acyl-phloroglucinol. Alkyl resorcinol formation from tetraketide intermediate follow the pattern C2 to C7 Aldol condensation of pine stilbene synthase (STS)[62] or *N. crassa*, PKSIII*Nc*[81] to give alkyl resorcinol.

Our crystallization studies thus reveal a novel type III PKS structure and provide clues to possible modes used by this family of homologous proteins to generate diverse bioactive metabolites from limited substrate pools.

Chapter 4

Investigating the physiological roles of type III Polyketide synthases (PKSs)

CHAPTER 4: INVESTIGATING PHYSIOLOGICAL ROLES OF TYPE III PKSs

4.1. INTRODUCTION

Mycobacterium marinum has four type III PKSs, as shown in figure 4.1.

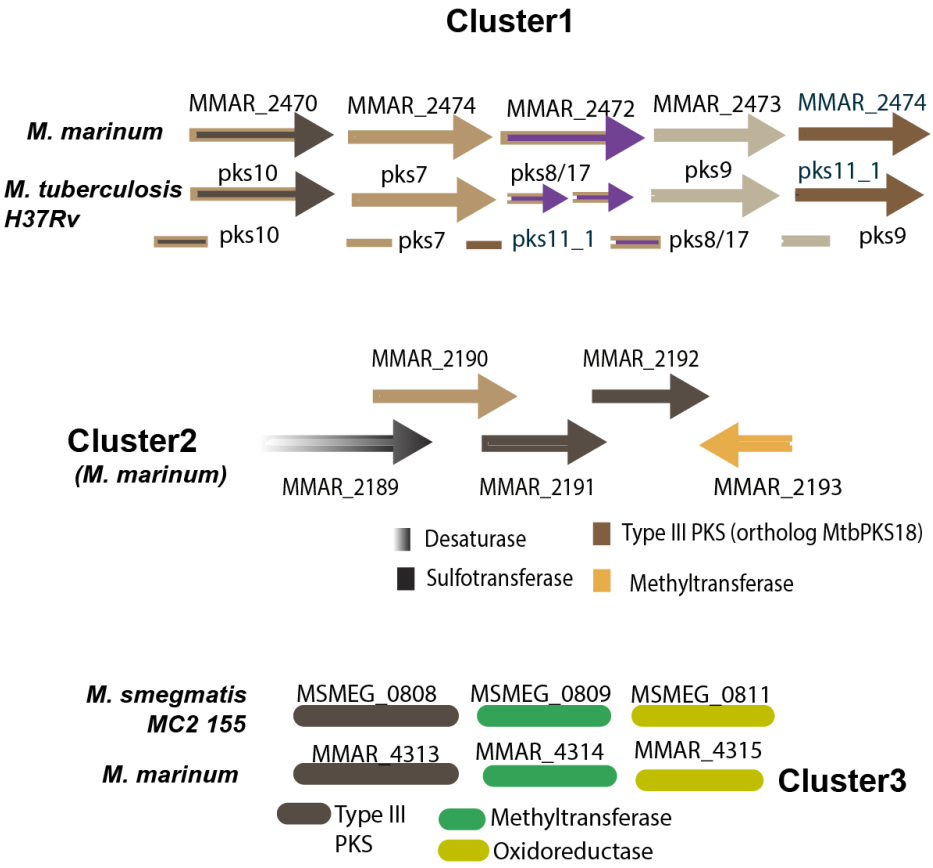


Figure 4.1: Four type III polyketide synthases (PKSs) are conserved in *M. mrinum* and other mycobacteria in the form of three pks cluster- cluster1, cluster2 and cluster3. *M. marinum*,Cluster1 consists of two type III PKSs,- MMAR_2470 and MMAR_2474 with respective orthologs pks10 and pks11 conserved in *M. tuberculosis*. Cluster2 has single type III PKS, MMAR_2190 with two orthologous protein MtbPKS18 and NFIII conserved in *M. tuberculosis* and *N. farcinica* respectively. Cluster3 consists of MMAR_4313 with orthologous protein MSMEG_0808 present in *M. smegmatis*.

The cluster study depicted in chapter II, and figure 4.1 show *M. marinum* has thee types of polyketide synthase cluster- cluster1, cluster2 and cluster3. Cluster1 includes MMAR_2470 and MMAR_2474 with respective orthologous protein pks10 and pks11 conserved in *M. tuberculosis*. Cluster2 consists of MMAR_2190 with two orthologous protein MtbPKS18 and NFIII conserved respectively in *M. tuberculosis* and *N. farcinica*. Different genes coding for cluster1 are exclusively present in pathogenic mycobacteria. While those coding for cluster2 are present only in the genomes of the pathogenic member of corynebacterineae. Cluster3 consists of type III PKS, MMAR_4313 from *M. marinum*. MMAR_4313 belongs to the same family of orthologous protein, recently characterized from polyketide quinone biosynthetic gene cluster of *M. smegmatis* MC^2 155[11]. *M. smegmatis* MC^2 15 has single type III PKS, MSMEG_0808. *MMAR_2190*, a type III *pks* gene produces non-methylated polyketide products. Wheres, two of the type III *pks* genes- *MMAR_2470* and *MMAR_2474*, have been recently characterized by our group to produce C-methylated polyketides[195]. NFIII makes products similar to MMAR_2190, as characterized in chapter three and appendix A2 using *in vitro* assay, followed by overexpression in Δpks10 knockout of *M. smegmatis* devoid of *MSMEG_0808* as shown in figure 4.2. In this study, we used the Δpks10 knockout strain of *M. smegmatis* to study the role of different type III PKSs from corynebacterineae through overexpression.

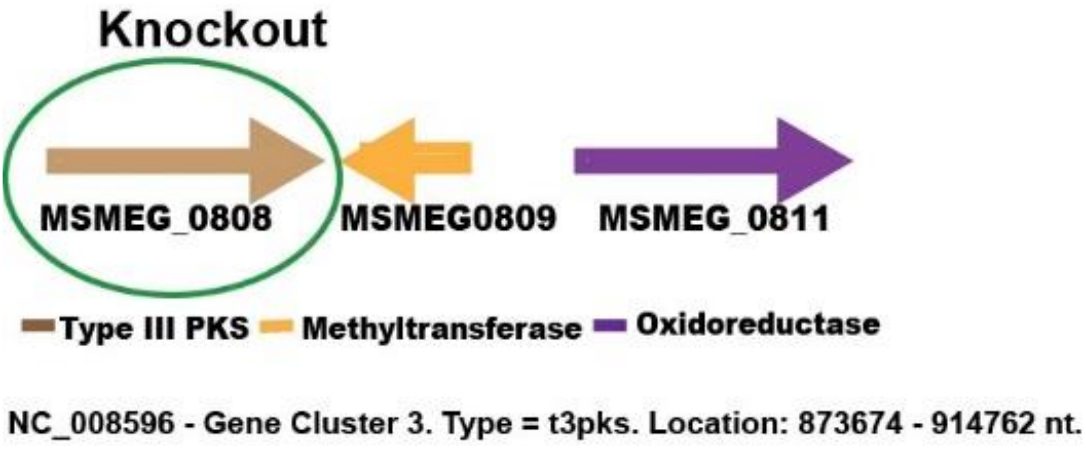


Figure 4.2: Figure showing Δpks10 knockout strain of *M. smegmatis* devoid of *MSMEG_0808* gene (circled in green). *M. smegmatis* Mc^2155 (with accession number NC_008596) type III PKS cluster, cluster3 starting from gene location 873674 to 914762 nt is illustrated in the figure. The figure is drawn based on gene prediction made using AntiSMASH[175].

There are already many studies done on microorganisms and microbial communities which is a rich source of secondary metabolites[196]. Many Small molecules or natural products isolated from bacteria are becoming lead compounds against many human diseases[197]. Almost 80 % of antibiotics in nature are produced by *Streptomyces* spp, a bacteria belonging to the Corynebacterineae. The other members of the family, *Mycobacterium* spp. and *Nocardia spp.* are less known for secondary metabolite production. However, the physiological relevance of genes can be understood in terms of a pathogenic factor of microorganisms, also known as a biofilm. Credit for microbial biofilm discovery goes to van Leeuwenhoek who first observed tiny organisms attached to the surface of the teeth as a biofilm. Biofilm is seen in 80 % of bacterial infections on the surface, causing more severe gingivitis, infectious kidney stone, bacterial endocarditis, infections in the inner ear, and nosocomial infections from catheters and ports[198,199]. Biofilm is a survival strategy in which microorganisms stay in a state of dynamic equilibrium with the host cell or tissue until the detachment from the surface to form planktonic cells[200].

The cells in a biofilm are attached to the surface but not permanently sessile, with an extracellular matrix consisting of water, polysaccharides and constantly changing macromolecules generating nutrients and osmotic gradient making the population heterogeneous to each other and act like single planktonic cells[201,202]. Various molecules like glycopeptidolipids[203], shorter chain mycolic acids, GroEL1, and chaperone[204], polyketide quinone[205] are also reported from the biofilm. Nutrients, carbon source and metal ions influence biofilm formation. Biofilm conferred features of the multi-layered

type of defense[206]. It has increased resistance to antimicrobials[205,207-209] and UV light[210] with increase rates of genetic exchange, alteration in bio-degradative capabilities and increased secondary metabolites[211] production. The uniqueness and functionality of polyketides as secondary metabolites have an impact on cell wall remodelling of mycobacteria and predicted to have diverse biological activities[7]. Many studies showed that these unique features of biofilm lead to the formation of persister cells[212] through nutrient limitation[213] and slow growth[214] compared to planktonic cells. Management and understanding of biofilm is quite necessary to combat the infection from non-tuberculous mycobacterial diseases and different kinds of *M. tuberculosis* diseases[215].

4.2. MATERIALS AND METHODS

4.2.1. Materials

The mycobacterial expression vector, pMyNT kindly donated by Annabel Parret and Matthias Wilmanns (Addgene plasmid # 42191) was used for overexpression of type III PKS gene. *M. smegmatis* knockout of type III PKS, *Δpks10* strain and wildtype *M. smegamtis* (Mc^2155) were kindly provided by Dr. Rajesh S. Gokhale (NII, India) and were utilized as host for overexpression of type III PKSs. Restriction enzymes were procured from New England Biolabs (NEB). Luria Bertani Broth (LB), LB agar, Middlebrook 7H9, Middlebrook 7H11 media, Sauton's Fluid Media Base, Glycerol were purchased from Himedia. Acetamide was obtained from Merck. All other analytical grades and HPLC grade solvents were obtained from Merck. Hygromycin B (50 mg/ml) was obtained from Thermo Fisher Scientific.

4.2.2. Methods

4.2.2.1. Generation of pMyNT clones

Construct G was prepared by ligating *MMAR_4313* gene excised from pET28c using restriction enzyme NcoI and HindIII, with mycobacterial expression vector *pMyNT* using NcoI site. The clone generated was digested with NdeI *a*nd HindIII site, and the *MMAR_4313* was excised out. The backbone generated was used to ligate independently *MMAR_2190,* and *NfIII, MMAR_2470 and MMAR_2474* genes using NdeI site and HindIII site. The constructs so made were transformed to *Escherichia coli* XL-1 blue. The clones were screened by restriction digestion using *NdeI* and *HindIII* and electrophoresed, as described in chapter two. The plasmid from positive clones was used separately for different type III PKSs encoding genes electroporation into *Δpks10* knockout strain of *Mycobacterium smegmatis* Mc^2155 with pMyNT plasmid as a control.

4.2.2.2. Mycobacterial electrocompetent cell preparation and electroporation*:*

The *Δpks10* knockout strain of *Mycobacterium smegmatis* Mc^2155 electrocompetent cells were prepared like chemical competent cell preparation for *E. coli* strains as described in chapter two till the OD600 nm reached 0.8 units from 1 % inoculum inoculated from overnight grown primary culture incubated at 37 ˚C in an incubator shaker. The cell growth was stopped by incubating the flask on ice for 1.5 hr. The culture was centrifuged at 3000 rpm for 15 min at 4 ˚C. The cell pellet was gently washed four times with sterile 10 % glycerol and aliquoted further using 4 ml 10 % glycerol in a microcentrifuge tube, snap-frozen in liquid nitrogen and stored at -80 ˚C for further uses.

Electroporation

Plasmid DNA (about 100 ng in 1-2 µL distilled water) was added to 50 µL competent cells in a 1.5-mL microfuge tube and incubated on ice for about 5 min. The mixture was then carefully transferred to an ice-cold electroporation cuvette (0.2 cm), avoiding the formation of air bubbles, and electroporation was carried out using a BioRad electroporator set to 2.5 kV. The optimal time constant is 4.5 – 5.0 ms. Middlebrook 7H9 (1 mL 7H9) medium was immediately pipetted into the cuvette, and the suspension was transferred to a microfuge tube and incubated on a water bath or a shaker (170 rpm) for 4 h at 37 °C. The mixture was spread on 7H11 agar plates containing the appropriate antibiotic(s) (no more than 200 µL per plate), and the plates were incubated at 37 °C for 4 to 7 days.

4.2.2.3. Screening positive transformants

The ligation mixture was transformed into 100 µl of an aliquot of competent cells at 4 ˚C. After the addition of ligated DNA mixture, the aliquots of the cell was incubated in ice for 30 min followed by heat shocks for precisely 90 seconds and back to the ice for 5 min. To the mixture, 500 µl of LB medium was added and incubated at 37 ˚C in an incubator shaker at 225 rpm. The content was plated on an LB Agar plate containing 100 µg/ml ampicillin and spread uniformly using a spreader and dried. The dried plate was incubated at 37 ˚C for 16 hrs.

The transformed colonies after 16 -hours (and 4-7 days) were grown to isolate plasmid from 3 ml culture with ampicillin (100 µg/ml) (and 25 µg/ml kanamycin and 50 µg/ml hygromycin B solution) incubated overnight (or for 4 days in 7H9 media) at 37 ˚C incubator-shaker followed by restriction digestion of the plasmid, and further confirmation of the transformants was done by Sanger sequencing.

4.2.2.4. Profiling different metabolites from Δpks10 overexpressed pMyNT- strains

The transformed colonies were screened on Middlebrook 7H11 agar plate (7H11) with kanamycin and hygromycin as selectable markers. For overexpression studies in a mycobacterial host, 10 ml primary culture was grown from the single screened colonies of *pMyNT-MMAR_2190* and *pMyNT-NfIII* and *pMyNT* vector control. The cultures were incubated in 7H9 media (having composition of supplemented ADC, 0.18 % glycerol and 0.05 % tween 80) with 50 µg/ml kanamycin, and 100 µg/ml hygromycin for four days. One per cent of the primary culture was used as inoculum for planktonic culture and biofilm generation. The planktonic culture was grown in triplicate in a 500 ml capacity flask with 250 ml each. The culture was induced with 1 mM acetamide when O. D at 600 nm reached 0.8. The biofilm was grown in triplicate in 140 mm polystyrene tissue culture grade sterile petri dish and sterile 6-well plate with 70 ml and 4 ml Sauton's fluid media respectively, supplemented with 2 % glucose and 2 % glycerol, 50 µg/ ml kanamycin and 100 µg/ ml hygromycin. Acetamide (1 mM) was used at the start of biofilm culture for induction. Cells of *pMyNT-MMAR_2190*, *pMyNT-NfIII* and *pMyNT* control planktonic culture were harvested after ten days, and biofilm after 14 days respectively.

Similarly, for comparison of extracts profile in mass spectrometry *pMyNT-MMAR_2470* and *pMyNT-MMAR_2474* were also grown as described above. Both the state cells were pelleted down by centrifugation and supernatant, and the pellet was separated. The pellet was washed with 100 mM Tris, pH 8.0 and was further re-suspended using a fresh buffer. Cells in this stage were used for transmission electron microscopy (TEM). For Scanning Electron Microscopy (SEM), the biofilm of *pMyNT-MMAR_2190*, *pMyNT-NfIII* and *pMyNT* control were grown on 6-well plate with sterile coverslip inside the media before seeding inoculum to the Sauton's media. The biofilm was harvested after 14 days. The

flakes of biofilm were transferred to the coverslip from Petri dish or already grown biofilm inside 6-well plate were washed twice using 0.1 M sodium cacodylate buffer.

Cells from grown planktonic culture and biofilm pellicles were harvested and resuspended in 100 mM Tris-Cl, pH 8.0. After resuspension, the culture supernatants and pellicles were acidified with acetic acid to pH 4, and polyketide derived lipid molecules were extracted by adding a double volume of ethyl acetate. The upper organic layer was decanted and filtered using Whatman filter paper (#54). A pinch of anhydrous sodium sulfate was added to the filtrate to remove traces of water molecules. The extract was decanted, and extracted products in solution were concentrated using rotary multivapor (Buchi) under reduced pressure. Methanol was used to solubilise the extracts. Products were further characterized using SCIEX Triple-TOF 6600 and 5600 High-Resolution Mass Spectrometer (HRMS) using multiple-reaction-monitoring (MRM) and information-dependent acquisition (IDA) methods with heated capillary temperature 450 ˚C; sheath gas nitrogen, collision gas argon). The auto-sampler was used at 4 ˚C to inject the sample volume 7 µl and 20 µl respectively for Triple TOF 6600 and 5600. Direct infusion of the sample was done without separation in LC. 0.1 % of formic acid was used to acidify the solvent system to give a flow rate of 400 µl/ min and 200 µl/ min. Linear-gradient of 70 % B and 90 % B was used for 3 min and 2 min to maintain the flow in Hybrid Triple-TOF Mass Spectrometer. IDA samples were run for 4 min. The declustering potential was optimized from 50 to 90, and collision energy (CE) was set to 35 (+/-20).

4.2.2.5. Transmission Electron Microscopy (TEM)

The harvested and washed cells from the planktonic culture in section 4.2.2.1 were further washed using 0.1 M sodium cacodylate buffer. The cells were pelleted down and were fixed overnight at 4 ˚C using 2.5 % glutaraldehyde and 4 % paraformaldehyde made on

cacodylate buffer. Fixed cells were further washed with 0.1 M cacodylate buffer to remove the excess of fixative and were embedded in 2 % agar blocks on the next day**.** The embedded cells were then post-fixed using 2 % osmium tetraoxide for 1-hour (hr). Different samples of *pMyNT-MMAR_2190*, *pMyNT-NfIII* and *pMyNT* control cells were dehydrated using a graded series of ethanol from 30 % to 100 % (30 %, 50 %, 70 % and 100 %). Epson 812 resin was used to infiltrate the samples and polymerized at 60 ˚C for 72-hours (hrs). Ultramicrotome was used for microtomy on a copper grid to make ultrathin sections (70 nm). The fine sections were stained using 5 % uranyl acetate and 0.2 % lead citrate, followed by the observation of sections in the transmission electron microscope (Tecnai twin 20, FEI).

4.2.2.6. Scanning Electron Microscopy (SEM)

The morphological difference of surface properties of harvested and washed cells from biofilm cultures, as mentioned in section 4.2.2.4 was studied using scanning electron microscopy. The primary fixation of samples was done using a mixture of 2.5 % (vol/vol) glutaraldehyde and 4 % paraformaldehyde for overnight. The samples were stored at 4 ˚C freezers. On the next day, samples were washed twice at room temperature with 0.1 M sodium cacodylate buffer. Secondary fixation of samples was done by using 1 % (weight/ volume) Osmium tetraoxide in 0.1 M Cacodylate buffer incubated at room temperature for 40 min in the dark. The fixed samples were washed three times with MilliQ water. The dehydration of the samples was done increasing the concentration of ethanol from 30 % to 100 % (30 % for 5 min, 40 % for 5 min, 50 % for 5 min, and 70 % for 5 min, 90 % for 5 min and 100 % for 5 min). Each dehydration was followed by the removal of ethanol from previous dehydration, and the use of absolute ethanol is repeated twice during drying of samples. Critical point drying was achieved for the samples using a chemical drying agent 100 % Hexamethyldisilazane (HMDS) solution for 15 min in the dark at room temperature.

The samples prepared are directly mounted on adhesive carbon film stick on a stub, and then gold coating was achieved by using Quorum SC7620 mini Sputter coater with a thickness of 30 Å (Angstroms). The topological difference between the cells was studied using Scanning Electron Microscope ZEISS-EVO LS 15 using an accelerating voltage of 20 kV.

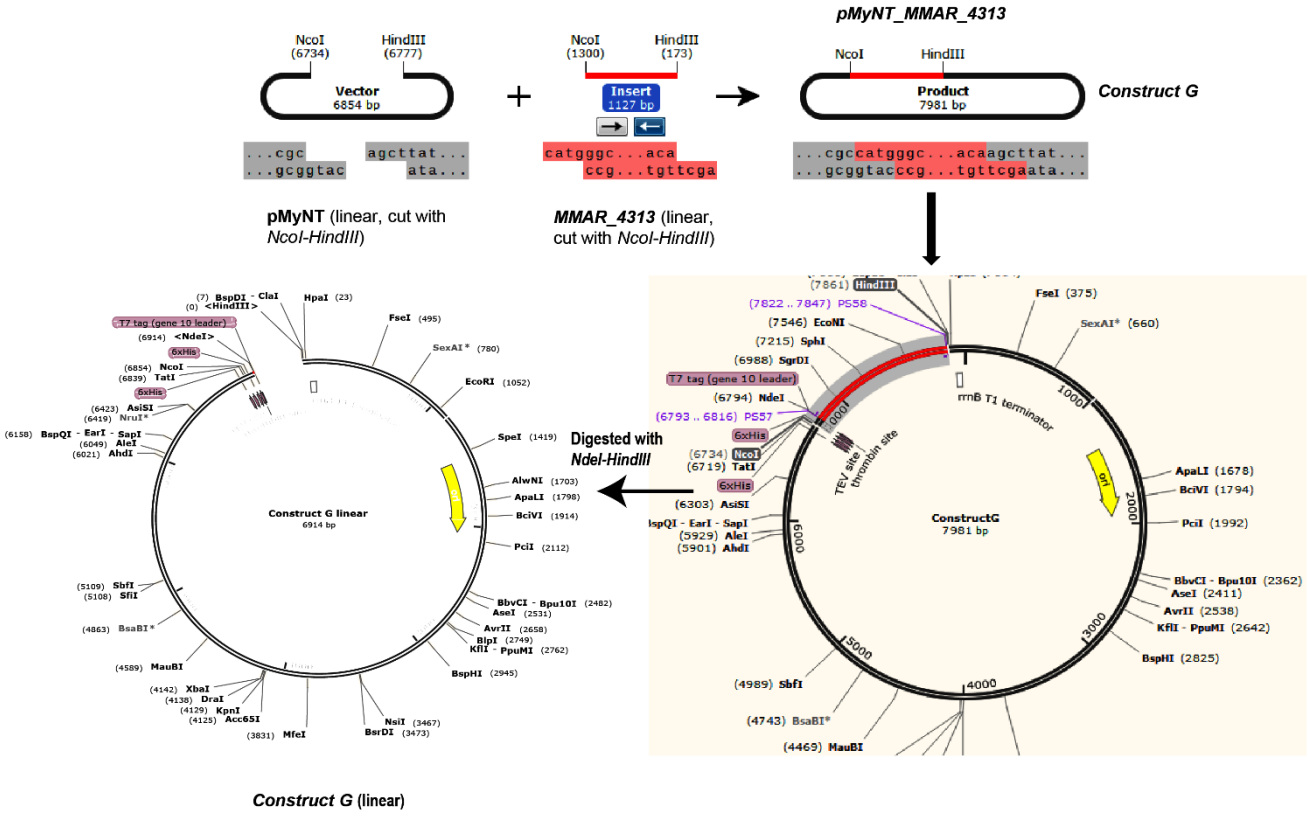
4.2.2.7. Atomic Force Microscopy (AFM)

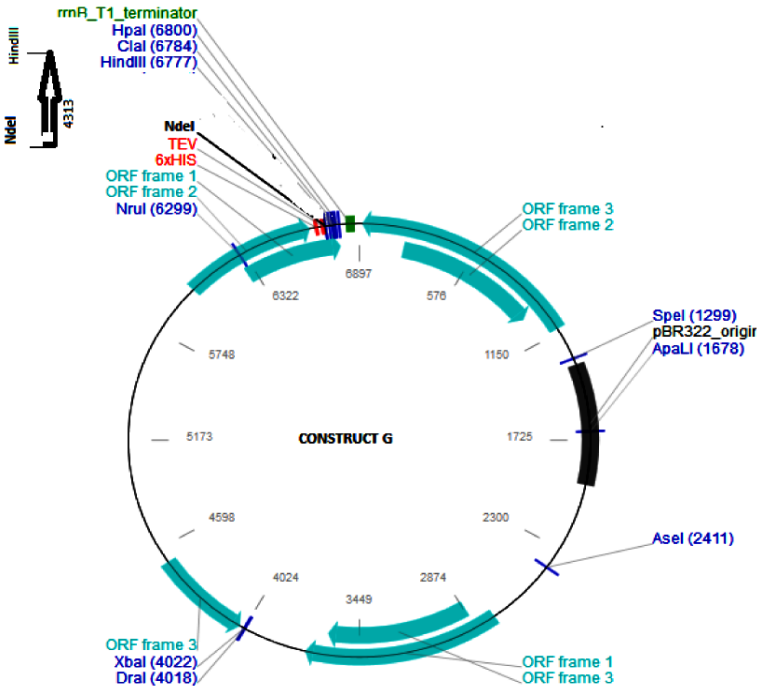
The biofilm cells from *pMyNT-MMAR_2190*, *pMyNT-NfIII* and *pMyNT* control and knockout (Δpks10) strain of *M. smegmatis* were washed using phosphate-buffered saline (1X PBS, pH 7.4). The bacterial cells from each biofilm cultures were separately diluted ten times and then applied to a freshly cleaved muscovite mica substrate. The substrate was rinsed using 50 µl deionized ultrapure water after the 60s. The samples were transferred into AFM liquid cell. The images were captured in a tapping mode at 20 ˚C through a moving probe on a BioScope Catalyst Atomic Force Microscope (Digital Instrument, Santa Barbara, CA) using an optical lever (NanoscopeIV controller) and type E scanner equipped to the microscope. The images were captured using single-beam silicon cantilever probes.

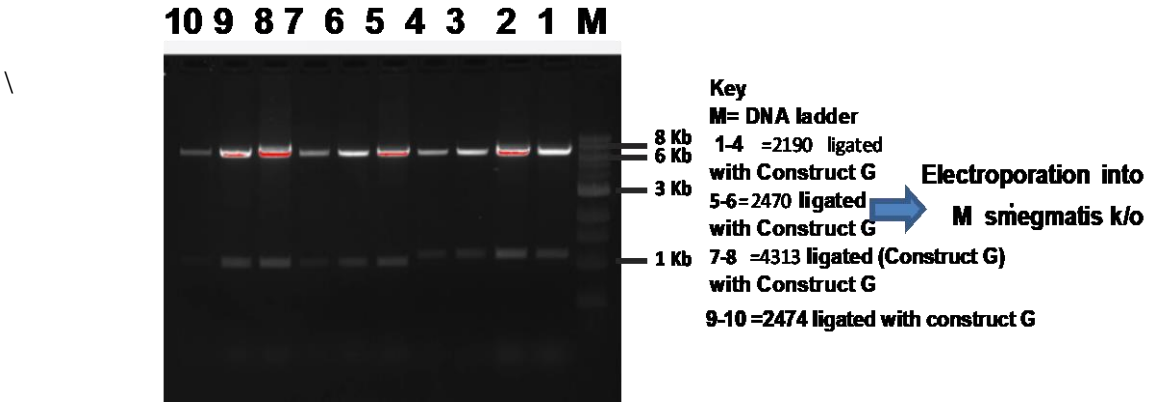
4.3. RESULTS

4.3.1. Generation of different type III PKS derived *pMyNT* clones

Master construct G of 7981bp was generated using the NcoI site (Figure 4.3). NdeI and HindIII-Hf site was used to excise out the *MMAR_4313* gene from the master clone, and the backbone was used to ligate *MMAR_2190*, *MMAR_2470* and *MMAR_2474*, and *Nf III* (Figure 4.4). The clones were screened and were confirmed by restriction digestion using NdeI and HindIII-*Hf* (Figure 4.3) followed by sequencing. All the pMyNT clones were electroporated into *M. smegmatis* k/o of Δpks10. The electroporated product was plated using a 7H11 agar plate with 50 µg/ml kanamycin and 100 µg/ml hygromycin as a selectable marker.







M	DNA ladder (1kb)	7,8.	*MMAR_4313* ligated with construct G
1-4.	*MMAR_2190* ligated with construct G	9-10.	*MMAR_2474* ligated with construct G
5,6.	*MMAR_2470* ligated with construct G		

Figure 4.3: Vector map of construct G (top) and agarose gel electrophoresis (bottom) showing restriction digestion using NdeI and HindIII site of *MMAR_2190* (lane 1 to lane 4) *MMAR_2470* (lane 5 and lane 6), Construct G (lane 7 and lane 8) and *MMAR_2474* (lane 9 and lane 10)

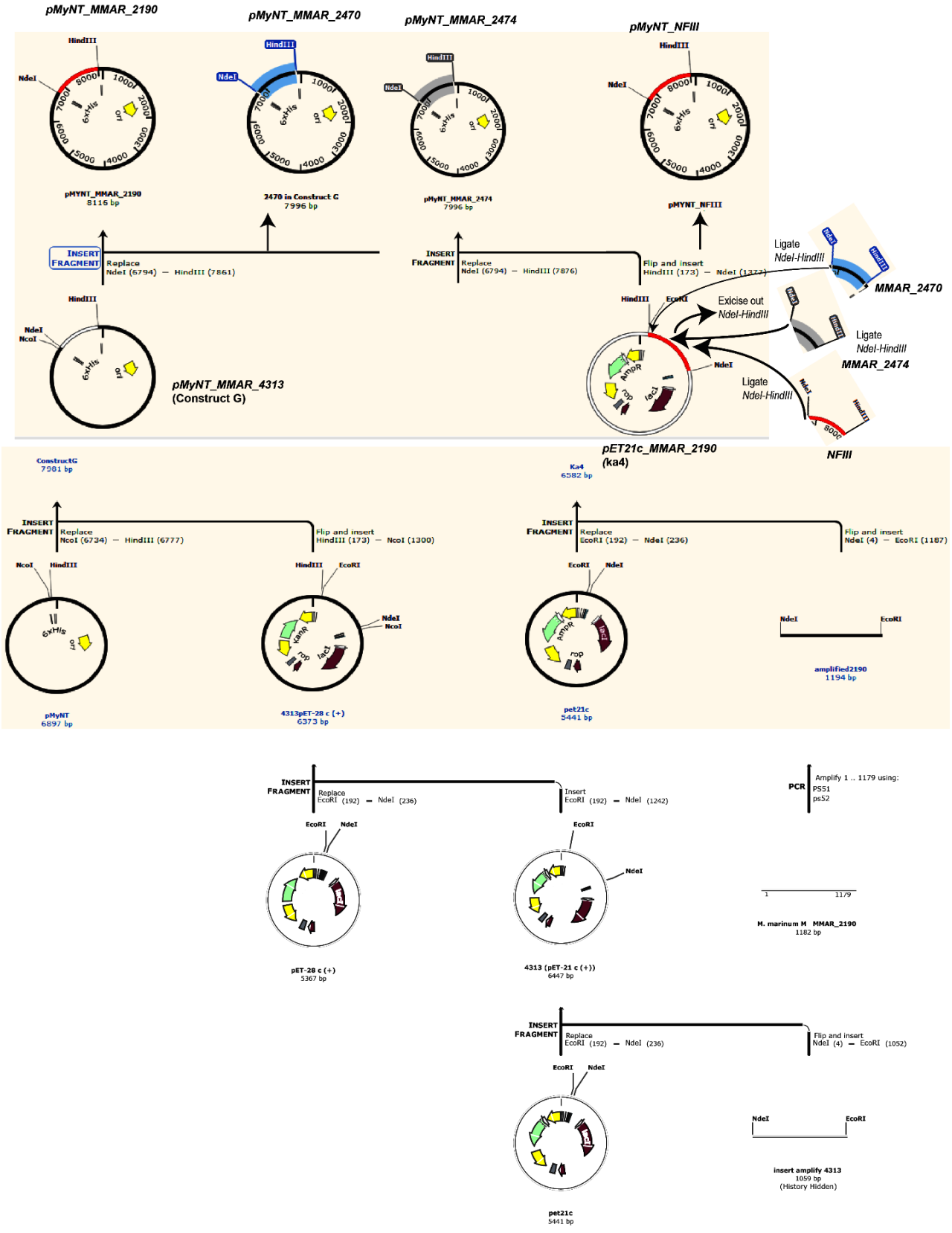


Figure 4.4: Cloning strategy used to generate different type III PKS pMyNT clones using construct G. *Mmar_4313* gene was excised using NcoI and HindIII-*Hf* from the pET28c-4313 clone. The product was purified using a gel extraction kit (Qiagen) and ligated into pMyNT vector with a

compatible restriction site. The construct is named as construct G. Further, restriction digestion of construct G using the same sets of restriction enzymes followed by ligation using prep digested excised gene of *MMAR_2190*, *MMAR_2470* and *MMAR_2474* and *NfIII* were ligated to generate *pMyNT*- a clone of the type III PKSs for overexpression studies.

4.3.2. Probing the physiological relevance of type III PKSs using planktonic cultures and biofilms pellicle assay

Type III PKSs derived overexpressed mycobacterial strains *pMyNT-MMAR_2190* and *pMyNT-NFIII*, *pMyNT-MMAR_2470* and *pMyNT-MMAR_2474* and *pMyNT-NfIII* were observed using pellicle assay of biofilm. Two pMyNT strains- *pMyNT-MMAR_2190* and *pMyNT-NFIII* found to show distinct colouration in both planktonic and biofilm cultures (figure 4.5). The biofilm of *pMyNT-MMAR_2190*, *pMyNT-MMAR_2474*, *pMyNT-MMAR_2470* inferred the formation of pellicle on the air-liquid interface with dying dead cells below the bottom surface of the culture. *pMyNT-MMAR_2190* showed brownish dark red and brownish light red colouration respectively in planktonic and biofilm cultures. *pMyNT-NFIII* was found to form transparent light brown with planktonic culture taking the colour of media and colourless with no biofilm with dead cells lying below the surface. Wildtype *M. smegmatis* Mc^2155 shows the presence of biofilm in the form of thick pellicles with brownish-yellow pigmentation. With overexpression of type III PKS, MMAR_2190 and MMAR_2474 to in knockout strain of *M. smegmatis* Δpks10, there was found better restoration of biofilm formation in *pMyNT-MMAR_2190* strains similar to the wild-type strains of mycobacteria, *M. marinum* and *M. smegmatis*. Whereas, *pMyNT-MMAR*_2474 and *pMyNT-MMAR*_2470 were observed with partial recovery and as very thin aggregated cells, respectively. The *pMyNT-NfIII* strain there was no observation of biofilm restoration. Instead of pellicle formation, dead bacteria were found below the surface of the stationary culture of *pMyNT-NfIII*.

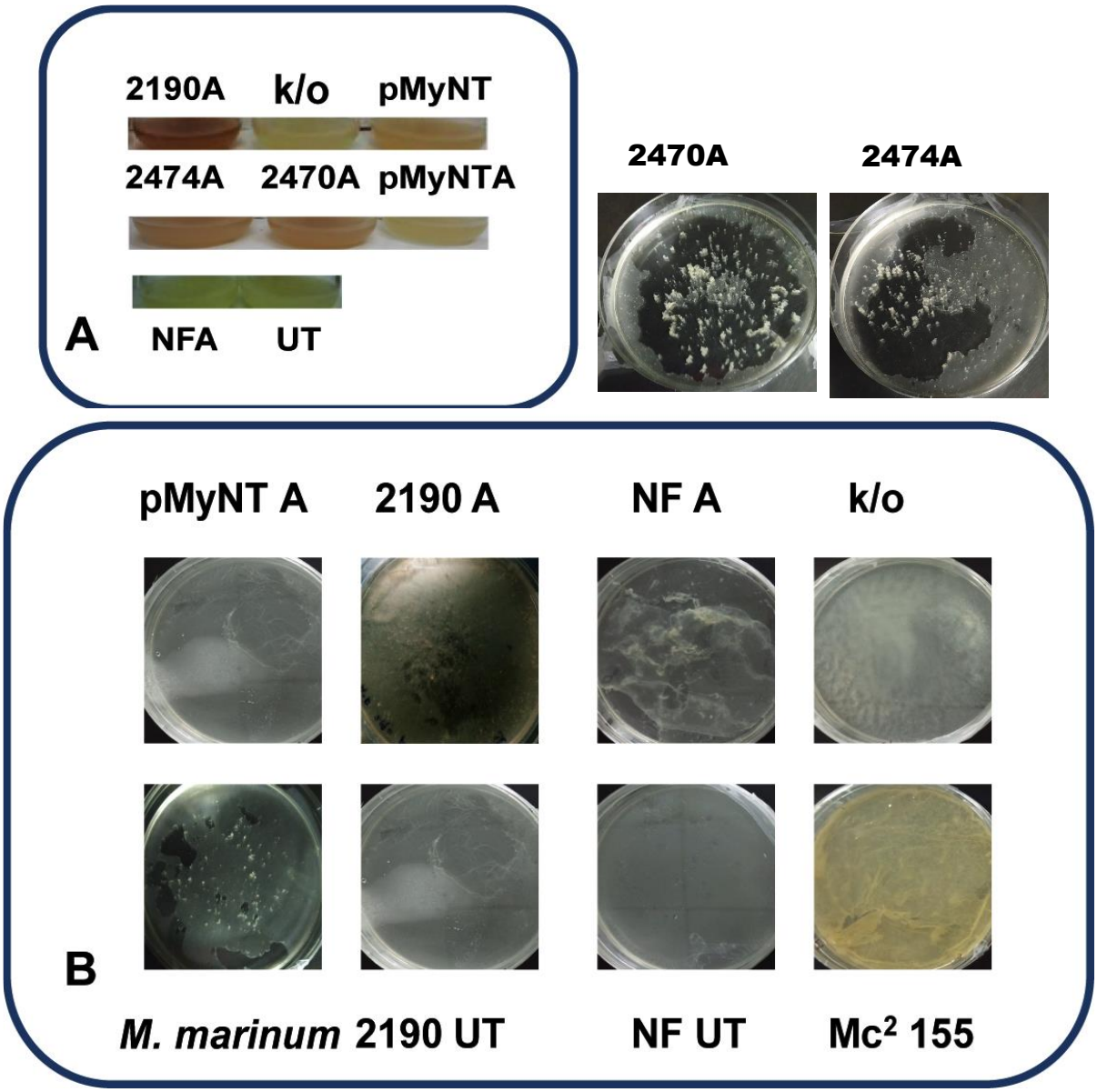


Figure 4.5: Physiological relevance of type III PKSs gene overexpressed in *M. smegmatis* knockout (k/o) of Δpks10. Type III PKSs gene *MMAR_2190* (2190), *MMAR_2470* and *MMAR_2474* from *M. marinum* are shown to be overexpressed in two culture conditions. A and B show the respective planktonic cultures and biofilm cultures of acetamide induced *pMyNT-MMAR_2190* (2190A), *pMyNT-MMAR_2470* (2470A), *pMyNT- MMAR_2474* (2474A) and *pMyNT-NFIII* (NFA) and *M. smegmatis* knockout strain of Δpks10 as control. Similarly, to compare the relevance of gene wild-type *M. marinum* and *M. smegmatis* biofilm are shown in figure B. The figure depicts type III PKSs (MMAR_2190 and MMAR_2474) revive the biofilm with the overexpression of PKSs gene in *M. smegmatis* knockout (k/o) of Δpks10.

4.3.3. Metabolomics profiling of type III PKSs in *M. smegmatis* knockout strain of Δpks10

The mycobacterial type III PKS, MMAR_2190 and nocardial, NFIII proteins are orthologous protein belong to a unique cluster of pathogenic bacterial species in which MtbPKS18 belongs. Type III PKSs, MMAR_2474 from *M. marinum* was recently reported to catalyze C-methylated products (methylated alkyl-resorcinol and methylated phloroglucinol) by the use of acyl-CoA as starter and malonyl-CoA and methyl-malonyl-CoA as extender whereas, MMAR_2470 with restricted product profile[195]. Our study from MMAR_2190 and NFIII proteins both show the catalysis of non-methylated products,long-chain-triketide *α*-pyrone, long-chain-tetraketide *α*-pyrone, alkyl-phloroglucinol, alkyl-resorcinol and some quinones like molecules. The biofilm of *M. marinum* also shows the presence of some of the long-chain- triketide and tetraketide *α*-pyrone, phloroglucinol, alkyl resorcinol and quinones from both planktonic and biofilm cultures of type III PKSs overexpressed genes, in *M. smegmatis* knockout strain of Δpks10.

4.3.3.1. Study of type III PKSs derived metabolites from overexpressed strains

4.3.3.1.1. Alkyl triketide *α*-pyrones (A)

The T7 promoter based histidine tag overexpressed stain of MMAR_2190 in pET21c *E.coli* expression vector shows the presence of metabolites reported in chapter 2**.** The overexpressed type III PKS strains show the presence of even chain, odd-chain, and methylated tetraketide *α*-pyrone from C_6 to C_{30}.The triketide *α*-pyrones in MMAR_2190 from *E. coli* shows presence of 6A, 8A, 9A, 10A, 11A, 12A, 13A, 14A, 16A, 18A, 20A, 22A, 24A, 26A, 28A and 30A.. The small chain tetraketide *α*-pyrone derived from C_6-CoA and malonyl-CoA like a molecule with m/z at negative ion mode 181.01 was also identified from the biofilm culture of *M. marium*. The relevance of such a molecule needs to be

further investigated. The summary of different type III PKS derived metabolites- alkyl triketide *α*-pyrones are shown below under different headings :

Alkyl triketide *α*- pyrone	**Overexpressed BL21/DE3 strain (MMAR_2190)**	**Overexpressed *pMyNT_MMAR_2190*)**	**Overexpress *pMyNT_NFIII***	**Wildtype *M. marinum***
Even Chain	6A to 30A	6A to 22A	6A to 22A	12A, 16A
Odd-chain	9A, 11A, 13A	9A, 13A, 15A, 17A, 21A, 23A	9A, 13A, 15A, 17A, 21A, 23A	Not observed
Methylated	Not observed	6A' to 18A', 20A', 24A'	6A' to 18A'	8A', 12A', 16A, 18A', 20A', 22A'

MMAR_2190 and *NFIII* genes overexpressed in *M. smegmatis* k/o of Δpks10 strain shows the presence of 6A, 6A', 8A, 9A, 9A', 10A, 10A', 12A, 12A', 13A, 14A, 14A', 15A, 16A, 17A, 18A, 18A', 20A, 21A, 22A and 23A and the formation of triketide *α*-pyrone was found more in biofilm culture rather than planktonic culture. Lower chain C_6 (6A') to C_{10} (10A') and higher chain methylation 20A', 24A' are observed in extracts from MMAR_2190.

We observed the presence of both mix odd chain (9A, 13A, and 21A) and even chain (14A, 16A, 18A', 20A, 20A' and 22A triketide *α*-pyrone) from extracts of *pMyNT-MMAR _4313*. On the other hand the metabolome of *pMyNT-MMAR_2474* (2474A) and *pMyNT-MMAR_2470* (2470A) show presence of 10A, 12A, 12A', 14A, 14A', 16A, 16A', 18A, 18A', 20A, 22A majorly in the biofilm extracts. Our results on MMAR_2474 and MMAR _2470 show besides the formation of C-methylated products, using acyl-CoA as starter and methylmalonyl-CoA as an extender, the biofilm has pool or metabolites specially triketide *α*-pyrone similar to that formed by MMAR_2190, and NFIII using acyl-CoA as starter and malonyl-CoA as an extender. The significance of the molecule was checked by a

growing biofilm of wildtype *M. marinum*. We observed the presence of triketide *α*-pyrone 8A, 10A, 10A', 12A, 12A', 14A, 14A', 16A (16A-1.5 % NaCl)), 16A', 18A, 18A', 20A, 20A', 22A.

To cross-check the difference in triketide *α*-pyrones in our overexpressed strain, we isolated metabolites from close homologs of MMAR_2190, i.e. MtbPKS18-pET21c based *E.coli* extracts which show the presence of long-chain triketide *α*-pyrone from C_{10} to C_{24}-C-chain, with products 10A, 14A, 14A', 16A, 18A, 24A and 24A'. The presence of similar molecules in cultures from *M. smegmatis* Mc^2155 with tetraketide *α*-pyrone in the products, 16A (in the planktonic culture with 1.5 % NaCl stress), 15A, 24A (in biofilm culture). The mass spectrometry profile with the fragmentation of different alkyl triketide *α*-pyrones is illustrated in figure 4.6 to figure 4.10, and appendix A4.1 to appendix A4.2.

This result also gives a hint that the molecules are important in biofilm formation. Moreover, the results show the production of metabolites in adverse condition. It is tempting to speculate that these molecules could play roles in pathogen survival but further validation is required.

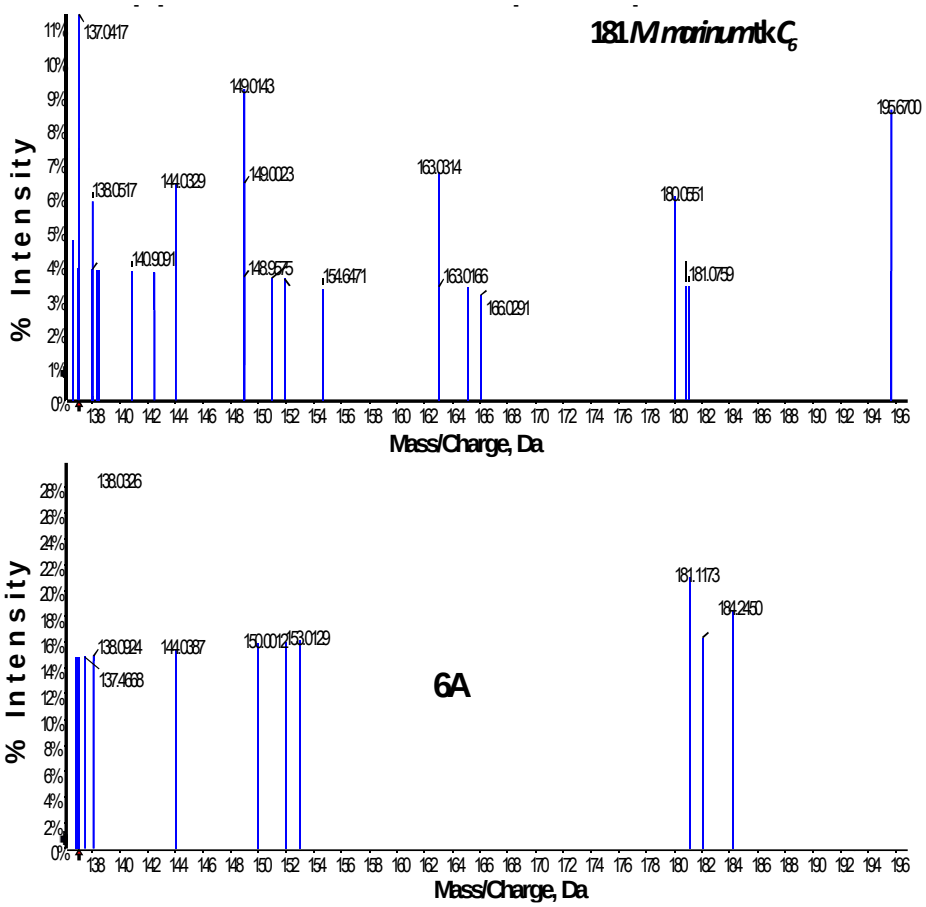


Figure 4.6: Metabolomics analysis of triketide *α*-pyrones (6A from wildtype *M. marinum* with m/z and MS2 fragments similar to type III PKS homologs of MMAR_2190 overexpressed in pMyNT based overexpressed knockout (strain devoid type IIII PKS, Δ pks10)

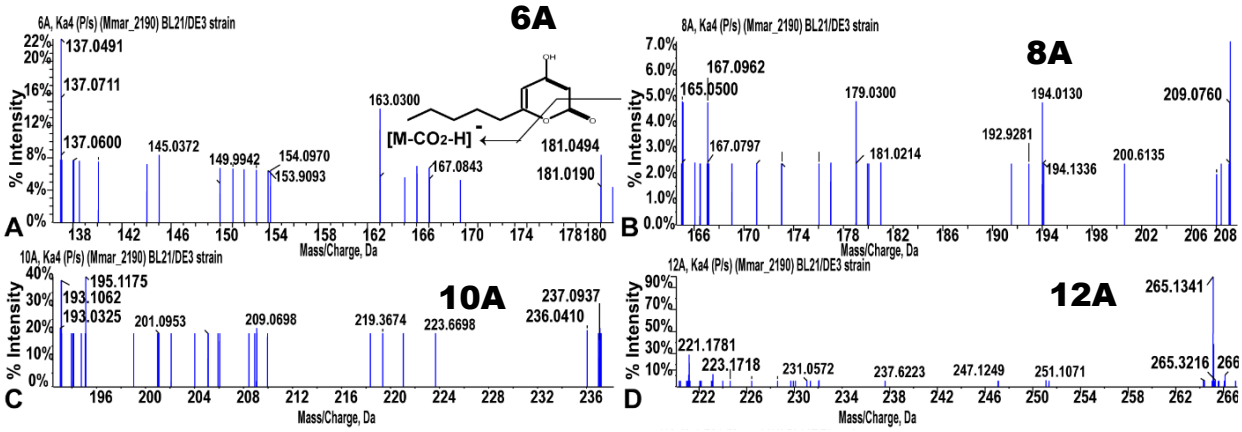


Figure 4.7: Metabolomics analysis of triketide *α*-pyrone from the *MMAR_2190* gene overexpressed in BL21(DE3) *E.coli* strain. A to D show the fragmentation of precursor ion derived from C_6 to C_{12} carbon chain with a mass of precursor $[M-H]^-$ to give fragments of $[M-CO_2-H]^-$ with loss of -44 mass unit from $[M-H]^-$.

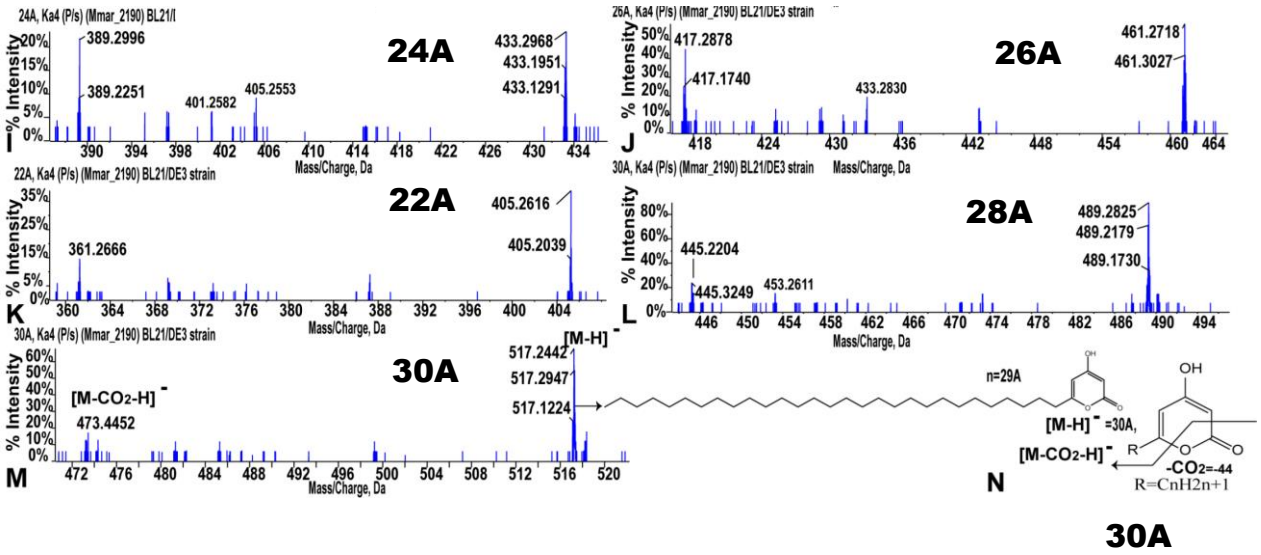


Figure 4.8: Metabolomics analysis of triketide α-pyrone from the *MMAR_2190* gene overexpressed in BL21(DE3) *E.coli* strain. I to M show the fragmentation of precursor ion derived from C_{24} to C_{30} carbon chain with a mass of precursor $[M-H]^-$ to give fragments of $[M-CO_2-H]^-$ with loss of -44 mass unit from $[M-H]^{-}$. The fragmentation pattern is given in N, showing a general strategy for all the C-chain α-pyrone. An arrow is shown from 30A derived triketide α-pyrone of the mass of precursor m/z 517.2442 to give sub-fragments of $[M-CO_2-H]^-$ with m/z 473.4452.

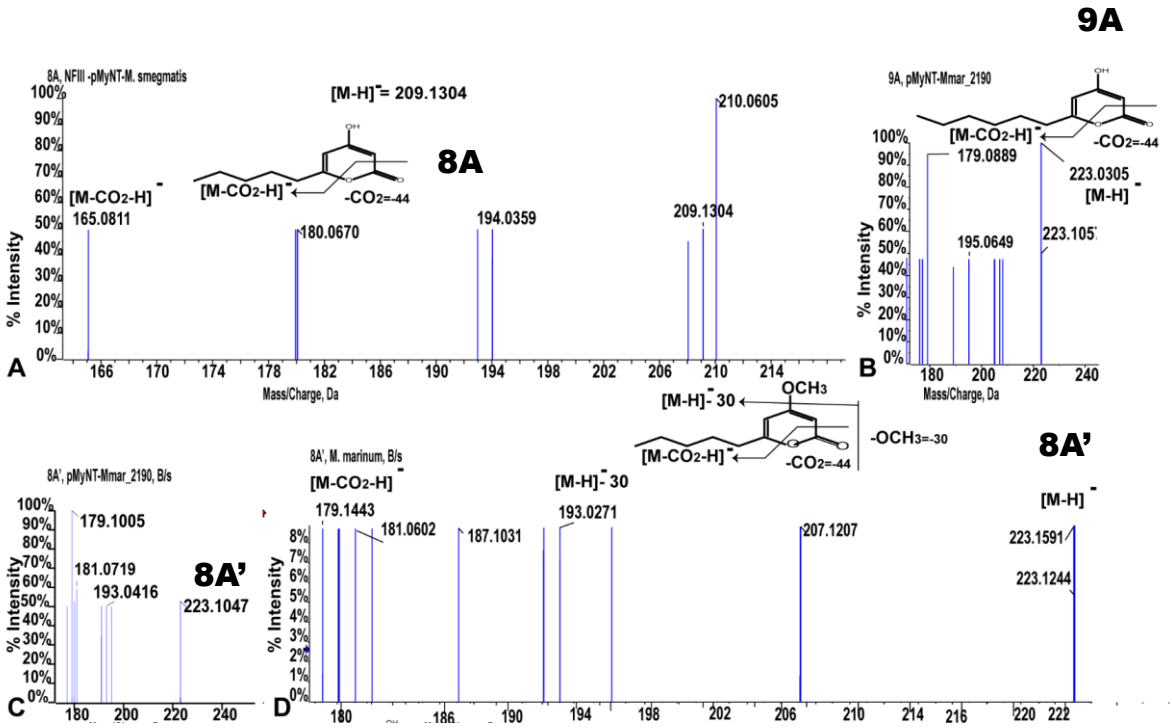


Figure 4.8: Metabolomics analysis of triketide α-pyrones (8A, 9A, 8A') from mycobacterial strains- wildtype *M. marinum* and *M. smegamtis* and different type III PKS homologs of MMAR_2190 overexpressed in pMyNT based overexpressed knockout (strain devoid type IIII PKS, Δ pks10). A and B show the fragmentation of 8A and 9A triketide α-pyrone from *pMyNT-NFIII* strain, *pMyNT-MMAR_2190* strains with the fragmentation of

precursor ion derived from C_8 to C_9 carbon chain with the mass of precursor $[M-H]^-$ to give fragments of $[M-CO_2-H]^-$ with loss of -44 mass unit from $[M-H]^{-\cdot}$. The strategy of fragmentation pattern is shown above the mass spectrometry data. C and D show the fragmentation from biofilm culture of overexpressed pMyNT_MMAR_2190 and *M. marinum* with *O*-methylated precursor ion of 8A' (m/z 223) to give fragments of $[M-CO_2-H]^-$ with loss of -44 mass unit from $[M-H]^-$ and $[M-H]^-$-30 as unique fragments. The presence of odd chain triketide α-pyrone like the one shown in B with the mass of precursor with $[M-H]^-$ at m/z 223 gives fragments $[M-CO_2-H]^-$ with loss of -44 mass unit similar to that observed in even chain triketide α-pyrone.

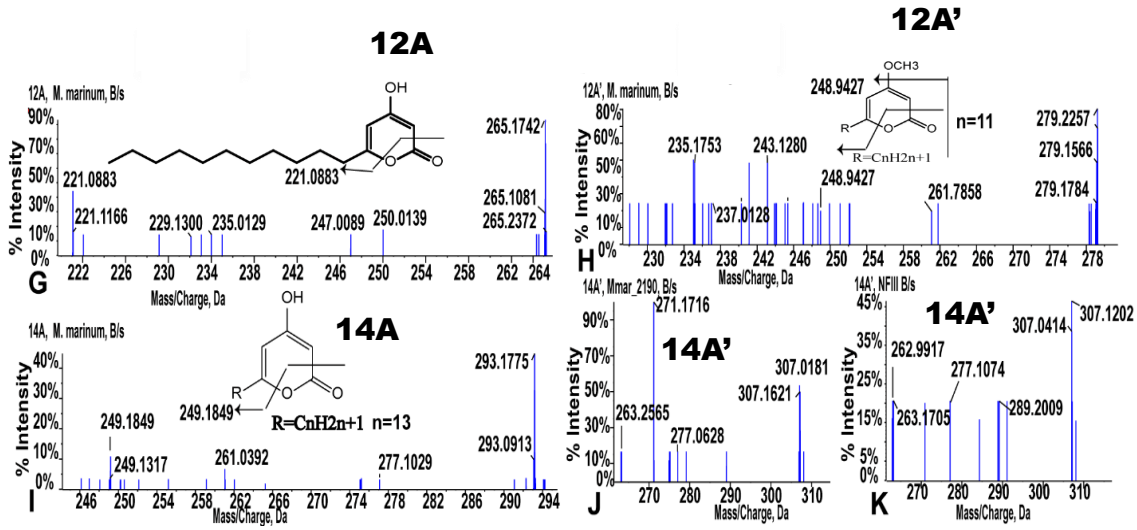


Figure 4.9: Metabolomics analysis of triketide α-pyrones (12A, 12A', 14A and 14A') from wildtype biofilm of *M. marinum* and MMAR_2190 overexpressed in pMyNT based overexpressed knockout (strain devoid type IIII PKS, Δ pks10). G and I show the fragmentation of 12A, and 14A triketide α-pyrone from *M. marinum* biofilm culture with the fragmentation of precursor ion derived from C_{12} to C_{14} carbon chain with the mass of precursor $[M-H]^-$ to give fragments of $[M-CO_2-H]^-$ with loss of -44 mass unit from $[M-H]^{-\cdot}$. The strategy of fragmentation pattern is shown above the mass spectrometry data. H and J and K show the fragmentation of *O*-methylated molecules from wildtype *M. marinum* and *pMyNT_NFIII* strain with precursor ion of 12A' (m/z at 279) and 14A' (m/z, 307) to give fragments of $[M-CO_2-H]^-$ with loss of -44 mass unit from $[M-H]^-$ and $[M-H]^-$-30 as unique fragments.

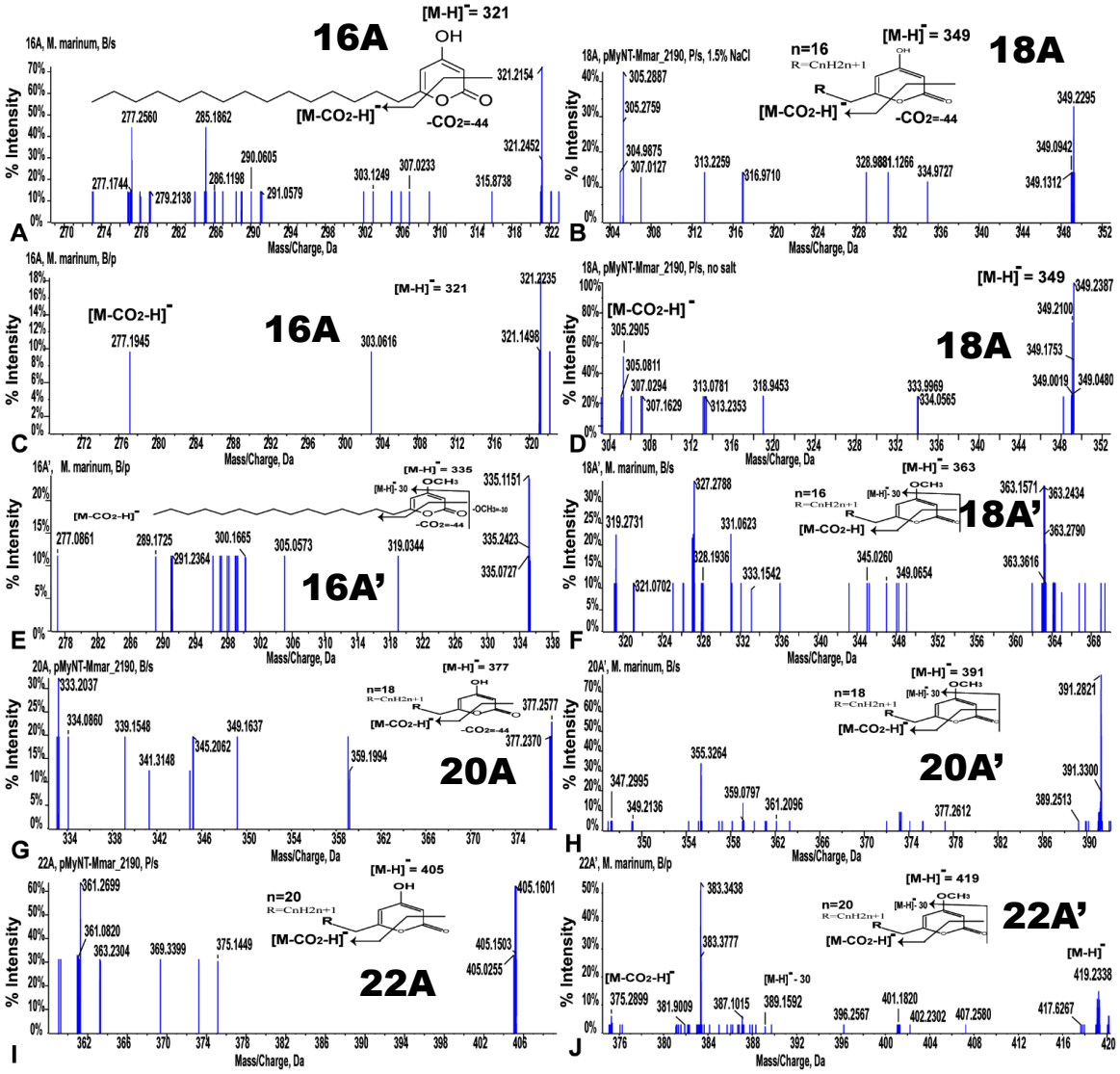


Figure 4.10: Metabolomics analysis of triketide *α*-pyrones (16A to 22A) from wildtype *M. marinum* and *pMyNT-MMAR_2190 with O*-methylated pyrone products. A, B, C, D, G and I show the fragmentation of 16A, 18A, 20A and 22A tetraketide *α*-pyrone with the fragmentation of precursor ion derived from C_6 to C_{22} carbon chain with mass of precursor $[M-H]^-$ to give fragments of $[M-CO_2-H]^-$ with loss of -44 mass unit from $[M-H]^{-\cdot}$. The strategy of fragmentation pattern is shown above the mass spectrometry data. E, F, G, H and J show the fragmentation of O-methylated precursor ion of 16A' (m/z 335), 18A' (m/z at 363), 20A' (m/z at 391) and 22A (m/z, 419) to give fragments of $[M-CO_2-H]^-$ with loss of -44 mass unit from $[M-H]^-$ and $[M-H]^-$-30. Although long-chain triketide *α*-pyrone fragmentation pattern was found in biofilm culture of *pMyNT-MMAR_2190* overexpressed strain and *M. marinum* wild-type strain with exception to the 18A present in the planktonic culture of MMAR_2190 without salt and with 1.5 % salt, the *O*-methylated products were better observed in *M. marinum* wildtype as shown in biofilm culture study in the above figure.

4.3.3.1.2. Alkyl tetraketide *α*-pyrone and alkyl-phloroglucinol from *pMyNT*- strains

Even chain alkyl tetraketide *α*-pyrone (6B, 8B, 12B, and 14B) and alkyl phloroglucinol (6D, 8D, 12D, 14D, and 16D, 24D and 28D) were found on both the metabolite extracts from planktonic and biofilm cultures from overexpressed strains of *MMAR_2190* in Δpks10 knockout and *M. marinum*. However, we could identify a limited number of phloroglucinol (8D, 10D, 12D, 14D, and 16D) products in *NFIII* and tetraketide *α*-pyrone (16B) from biofilm (supernatant) extracts of *M. marinum* wildtype. We studied the role of salt stress on the metabolome through the introduction of 0 % to 2 % NaCl. It was found that 1 % salt shift the production of tetraketide *α*-pyrone (8B) in the planktonic culture of *MMAR_2190* compared to the one with no salt with phloroglucinol (8D). Unlike 1.5 % salt favour the production of phloroglucinol (8D) in planktonic culture similar to the one identified phloroglucinol (8C) from biofilm extracts (figure 4.11). The tabulated summary of alkyl tetraketide *α*-pyrone and alkyl phloroglucinol from two overexpressed pMyNT strains with MMAR_2190 and NFIII is shown below. Representative MS/MS fragmentation profiles the compounds are given in figure 4.11, figure 4.12 and appendix A4.3 to appendix A4.5.

Bacteria	**Overexpressed *pMyNT_MMAR_2190*)**	**Overexpress *pMyNT_NFIII***	**Wildtype *M.* marinum**
Tetraketide pyrone (B)	6B, 8B, 12B, 14B	16B	6B, 8B, 12B, 14B
Phloroglucinol	6D, 8D, 12D, 14D, 16D, 24D, 28C	8D, 10D, 12D, 14D, 16D	6D, 8D, 12D, 14D, 16D, 24D, 28D

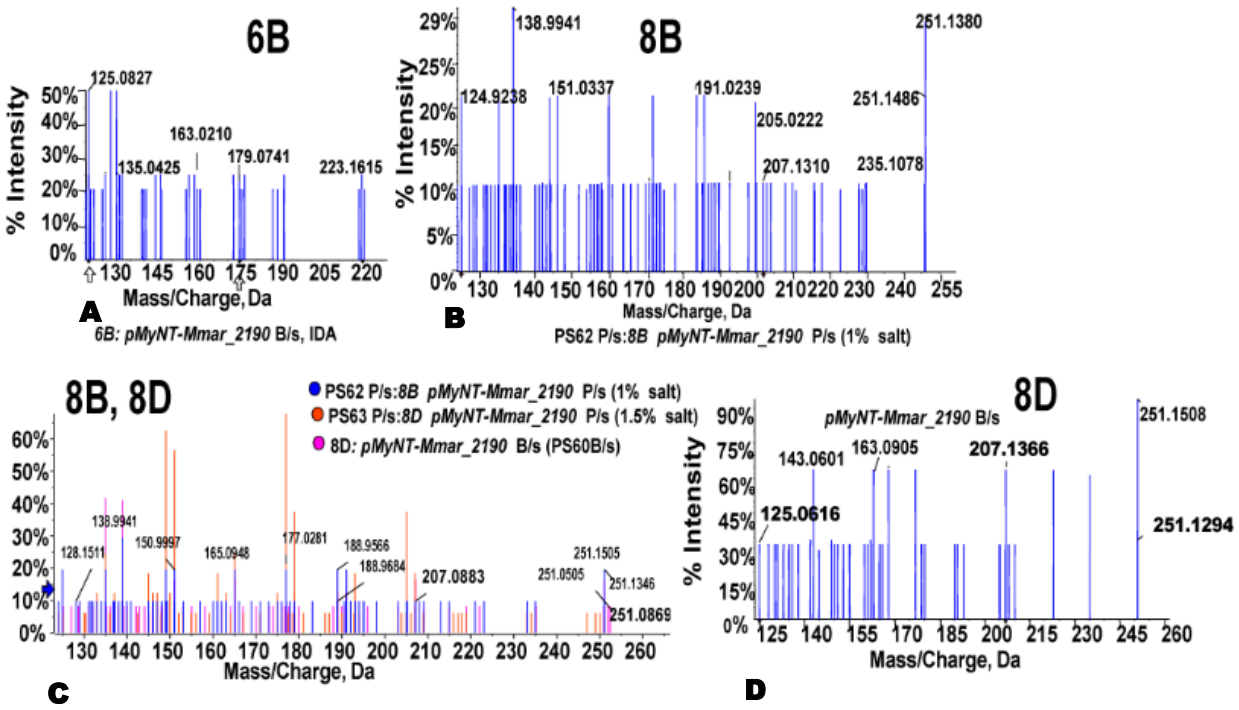


Figure 4.11: Fragmentation of precursor's ions with m/z 223.1615 (6B), 251.1367(8B and 8D) in negative ion mode with comparative study of bacterial culture using saline condition with (NaCl) from 1 to 1.5 %. Type III PKSs extracts from *pMyNT_MMAR_2190* (2190), give molecular ion peak $(M-H)^-$ at m/z 223 and 251 in biofilm culture and planktonic culture. The precursor ion $(M-H)^-$ at m/z 223 fragments into m/z 179 and 125 observed in methanol extract from the supernatant of biofilm culture of *pMyNT_MMAR_2190* shown in **A**). It shows the characteristics peak pattern for phloroglucinol 6B as described in the schematic representation (Chapter 2, figure 2.7**).** The biofilm culture of the same overexpressed strain showed characteristics peak pattern for alkyl phloroglucinol (8D) without salt (shown in D) and alkyl tetraketide *α*-pyrone (8B) with 1 % salt stress (as shown in figure B). The fragmentation is described in the schematic representation (Chapter 2, figure 2.7**).** With an increase in salt stress to 1.5 %, the planktonic culture of *pMyNT_MMAR_2190* strain was found producing alkyl phloroglucinol 8D (shown in figure C) similar to biofilm culture.

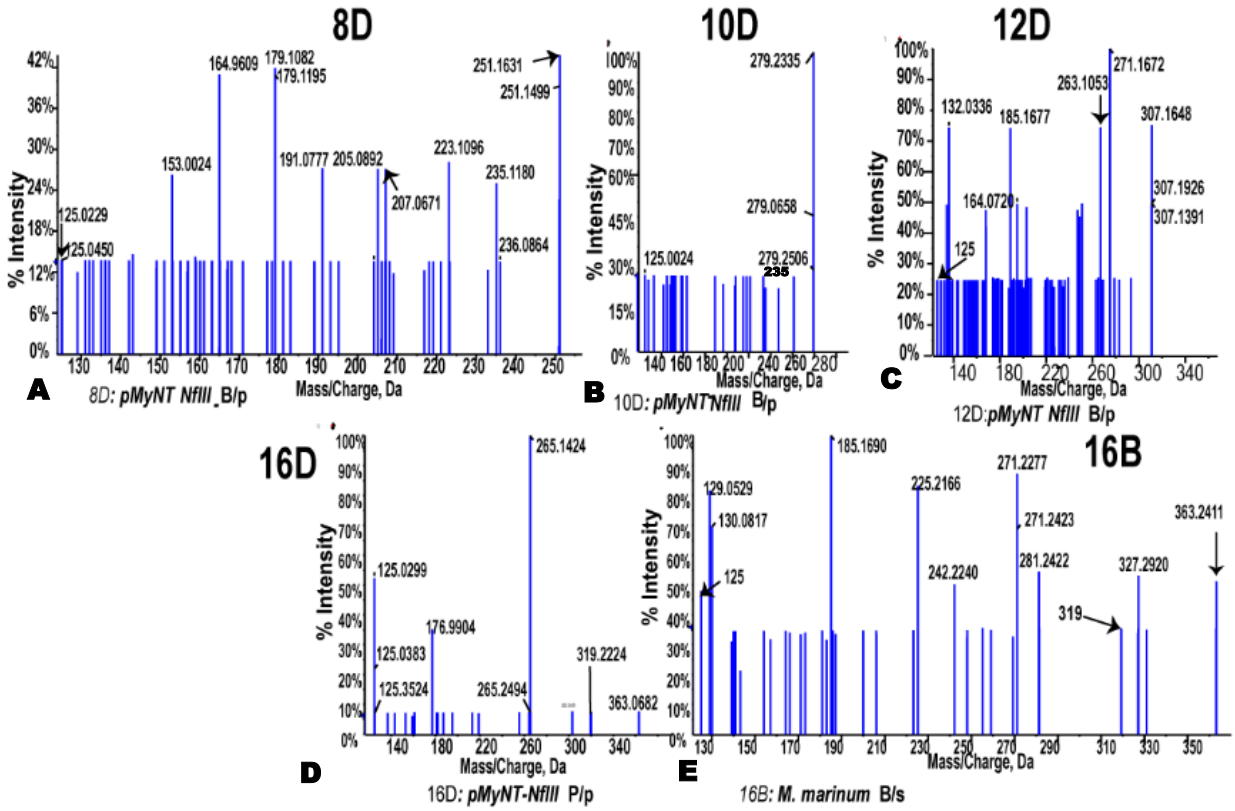


Figure 4.12: Fragmentation of alkyl phloroglucinol (8D, 10D, 12D and 16D) and alkyl tetraketide *α*-pyrone (16B) from the biofilm and planktonic culture of *pMyNT_NFIII*, and biofilm culture of *M. marinum*. Alky phloroglucinol from *pMyNT_NFIII* strain (shown in figure A to D) have precursor ions at m/z 251.1631 (8D), 279.2335 (10D), 307.1391 (12D), and 363.0682 (16D). Alkyl tetraketide *α*-pyrone (16B) shown in figure E show precursor ion at m/z 363.2411.

4.3.3.1.3. Alkyl resorcinol from *pMyNT*- strains

Metabolomics study of type III PKSs, mainly from biofilm extracts of the overexpressed strains of different genes show the presence of both odd and even chain alkyl resorcinol. We observed alkyl resorcinol- 7C, 8C, 9C, 10C, 11C, 12C, 13C, 14C, 15C, 16C, 17C, 18C, 19C, 20C, 21C, 22C, 23C, and 24C from the biofilm and planktonic culture extracts of *pMyNT_MMAR_2190*. *pMyNT_NFIII* extract shows the presence of alkyl-resorcinol of 7C, 8C, 9C, 10C, 11C, 15C, 16C, and 17C, 19C. Few of the alkyl resorcinol 7C, 8C, 9C, 11C, 14C, 15C, 16C, 17C, and 19C are observed in MMAR_2190 and NFIII. MMAR_4313 shows C18- carbon chain in biofilm extract, whereas we observed 19C in

MMAR_2470 and MMAR_2474. Higher chain resorcinol 22C was observed in both MMAR_2474 planktonic supernatant derived extracts possibly because of the binding of starter to the cavity volume. In general, we observed the increase in stress result in planktonic culture with higher even and odd chain alkyl resorcinol, but the metabolites are not present in biofilm culture. We observed resorcinol 10C, 13C, 16C, 19C in 1 % salt, 11C, 13C in 1.5 % salt. Thus, it shows with the increase in salinity; there is a gradual decrease in alkyl-resorcinol of the higher chain, and biofilm culture has better resorcinol similar to that observed in triketide α-pyrone. The tabulated form of alkyl resorcinol from different pMyNT- strains is given below. MS/MS fragmentation of a few representative alkyl resorcinols is given from figure 4.13 to figure 4.15.

Bacteria	**Overexpressed *pMyNT_MMAR_2190*)**	**Overexpress *pMyNT_NFIII***	**Wildtype *M.* marinum**
Alkyl resorcinol (C)	7C, 8C, 9C, 10C, 11C, 12C, 13C, 14C, 15C, 16C, 17C, 18C, 19C, 20C, 21C, 22C, 23C, 24C, 28C	7C, 8C, 9C, 10C, 11C, 15C, 16C, 17C, 19C	12C, 16C, 18C
Other type III PKS: *M. marinum* ***pMyNT_MMAR_4313*** 18C ***pMyNT_MMAR_2470*** 19C ***pMyNT_MMAR_2474*** 19C, 22C ***pMyNT_MMAR_2190*** 10C, 13C, 16C, 19C in 1% salt and 11C, 13C in 1.5% salt			

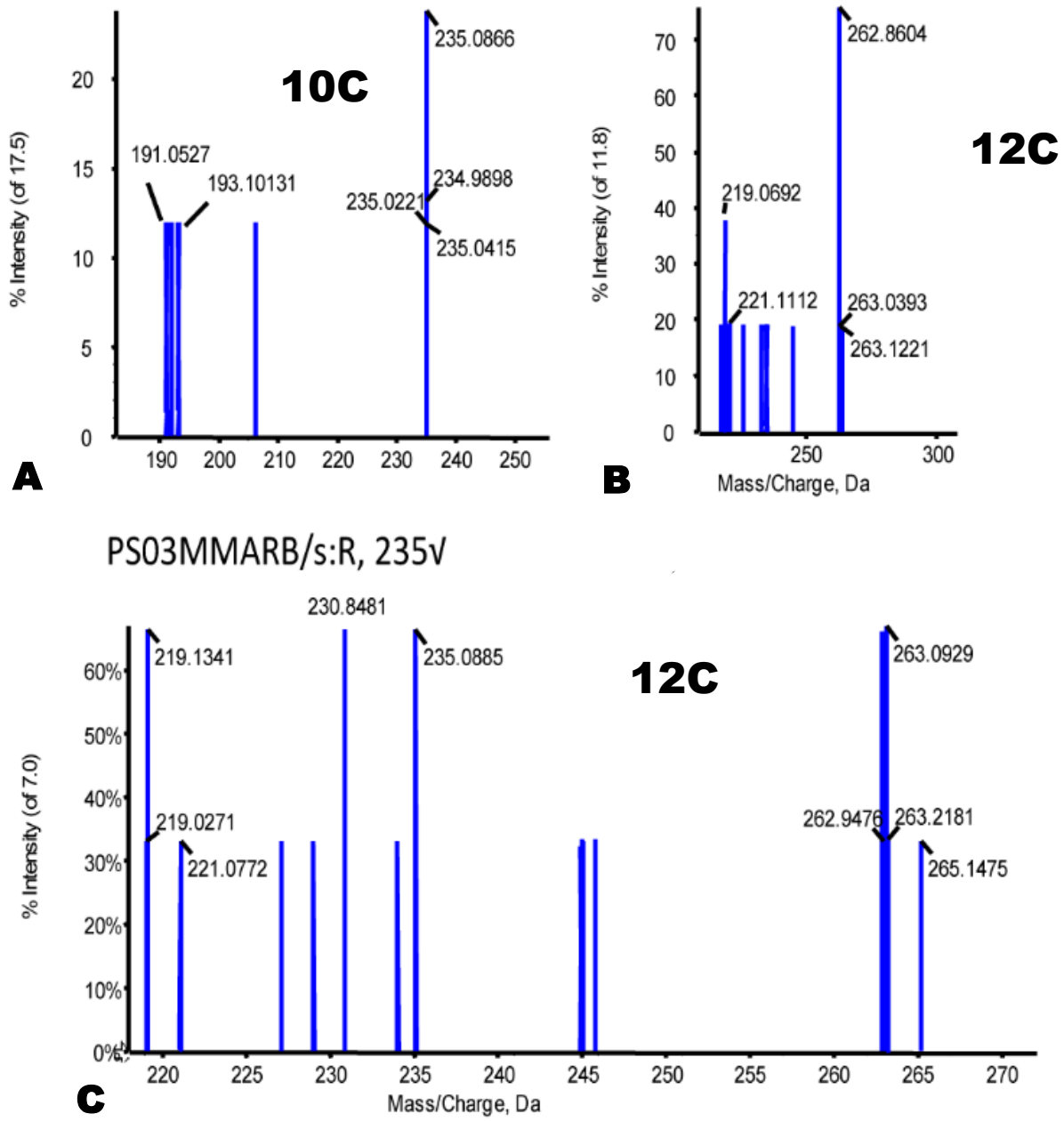


Figure 4.13: MS/MS fragmentation pattern observed for alkyl resorcinol (10C and 12C) Biofilm culture of *M. marinum* showed the presence of 10C(figure A) with a precursor ion peak at m/z 265 fragments into 191.0527 and 193.1031. B and C show the presence of 12C in *pMyNT_MMAR_2190* and *pMyNT_NFIII* strain in the both in planktonic and biofilm culture with precursor ion 263 showing fragments of 219 and 221.

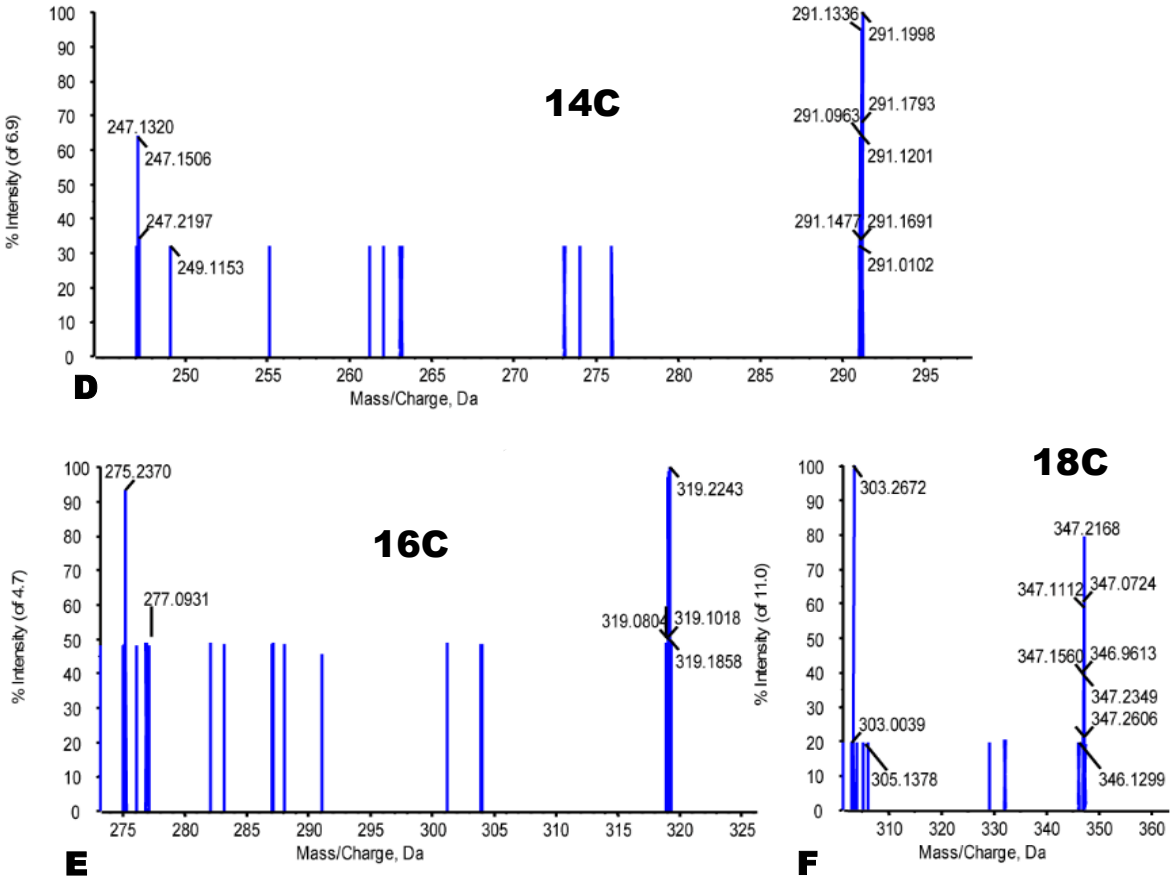


Figure 4.14: Representative MS/MS fragmentation pattern of alkyl resorcinol (14C, 16C and 18C) observed from biofilm culture of type III PKS n (*pMyNT-MMAR_2190* and *pMyNT-NFIII*). Biofilm culture of *M. marinum* also showed the presence of 16C and 18C. Figure D depicts the presence of 14C with precursor ion 291 showing fragments of 247 and 249. Alkyl resorcinol, 16C (E) with precursors ion peak $(M-H)^-$ at m/z 319 fragments into 275 and 277 peaks. The precursor ion from alkyl resorcinol 18C (F), 347 gets fragmented giving fragment ions of 303 and 305.

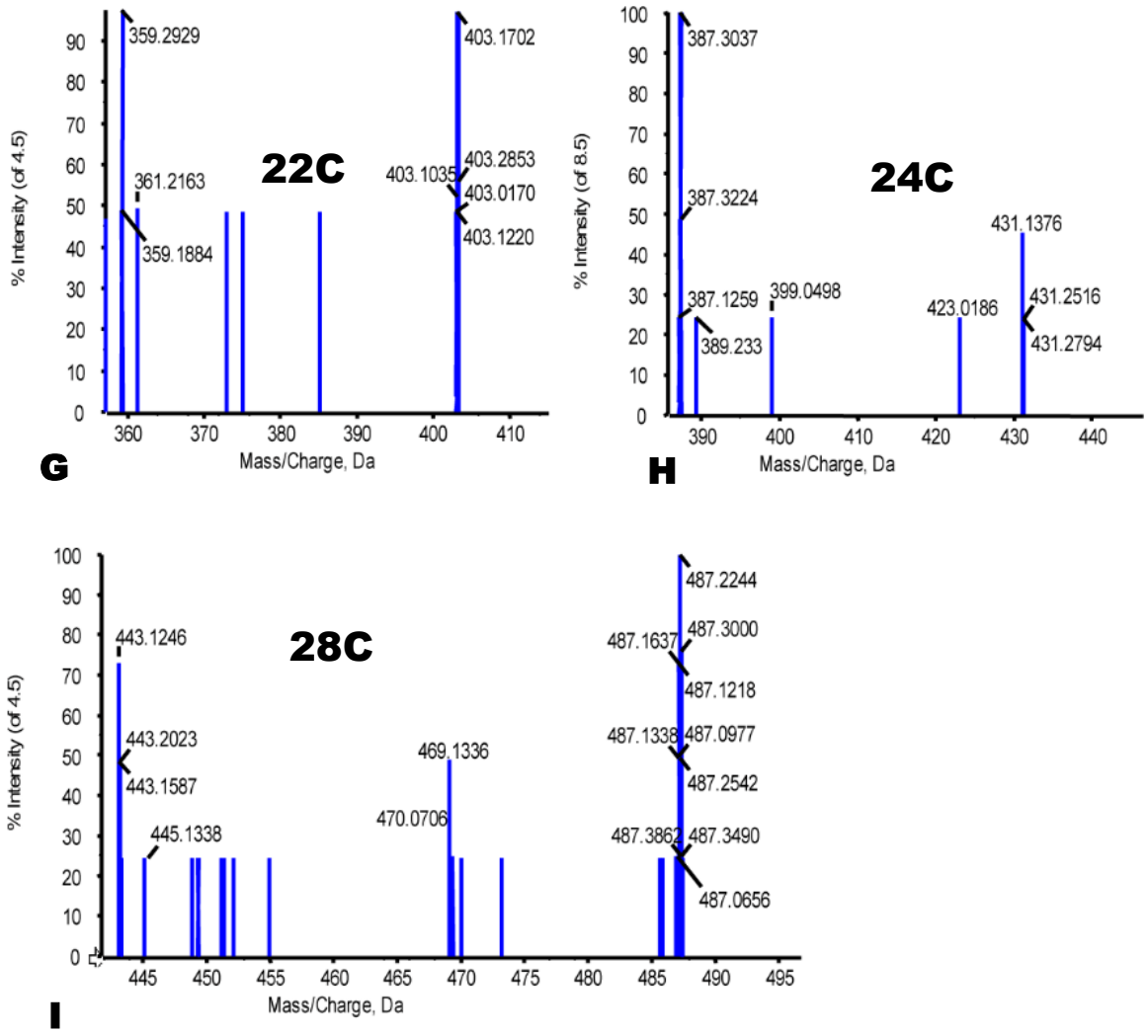


Figure 4.15: Representative MS/MS fragmentation pattern of higher chain alkyl resorcinol (22C, 24C and 28C). Figure G depicts the presence of 22C with precursor ion 403 showing fragments of 359 and 361. Alkyl resorcinol, 24C (H) with precursors ion peak $(M-H)^-$ at m/z 431 fragments into 387 and 389 peaks. The precursor ion from alkyl resorcinol 28C (I), 487 gets fragmented giving fragment ions of 443 and 445.

4.3.3.1.4. Quinones from pMyNT- strains

The physiological relevance of our metabolome products from *pMyNT-MMAR_2190* overexpressed strain shows various quinones with fragmentation similar to that reported by Anand *et al*. Besides C-methylated quinones reported, this study revealed the presence of odd and even chain, C-methylated and non-methylated quinones in MMAR_2190.

Similar type quinone like molecules were also found in the biofilm culture of wildtype *M. marinum* but not in pMyNT control and knockout control of *M. smegmatis* strain. The summary for the formation of quinones like molecules is tabulated below.

Bacteria	**Overexpressed *pMyNT_MMAR_2190*)**	**Overexpress *pMyNT_NFIII***	**Wildtype *M. marinum***
Quinone (E)	5E, 6E, 7E, 10E, 11E, 12E, 14E, 15E 16E, 18E, 20E,	5E, 11E	5E, 6E, 7E, 11E, 12E, 14E, 18E, 20E
Other type III PKS: *M. marinum* ***pMyNT_MMAR_4313*** Not observed ***pMyNT_MMAR_2470*** Not observed ***pMyNT_MMAR_2474*** Not observed ***pMyNT_MMAR_2190*** **P/s** 12E, 14E in 1% salt and 12E, 14E in 1.5% salt			

This study reveals the presence of quinones 5E, 6E, 7E, 10E, 11E, 12E, 14E, 15E, 16E, 18E, and 20E. The representative figure of few odd and even chain quinones like molecules are shown in figure 4.16 and 4.17.

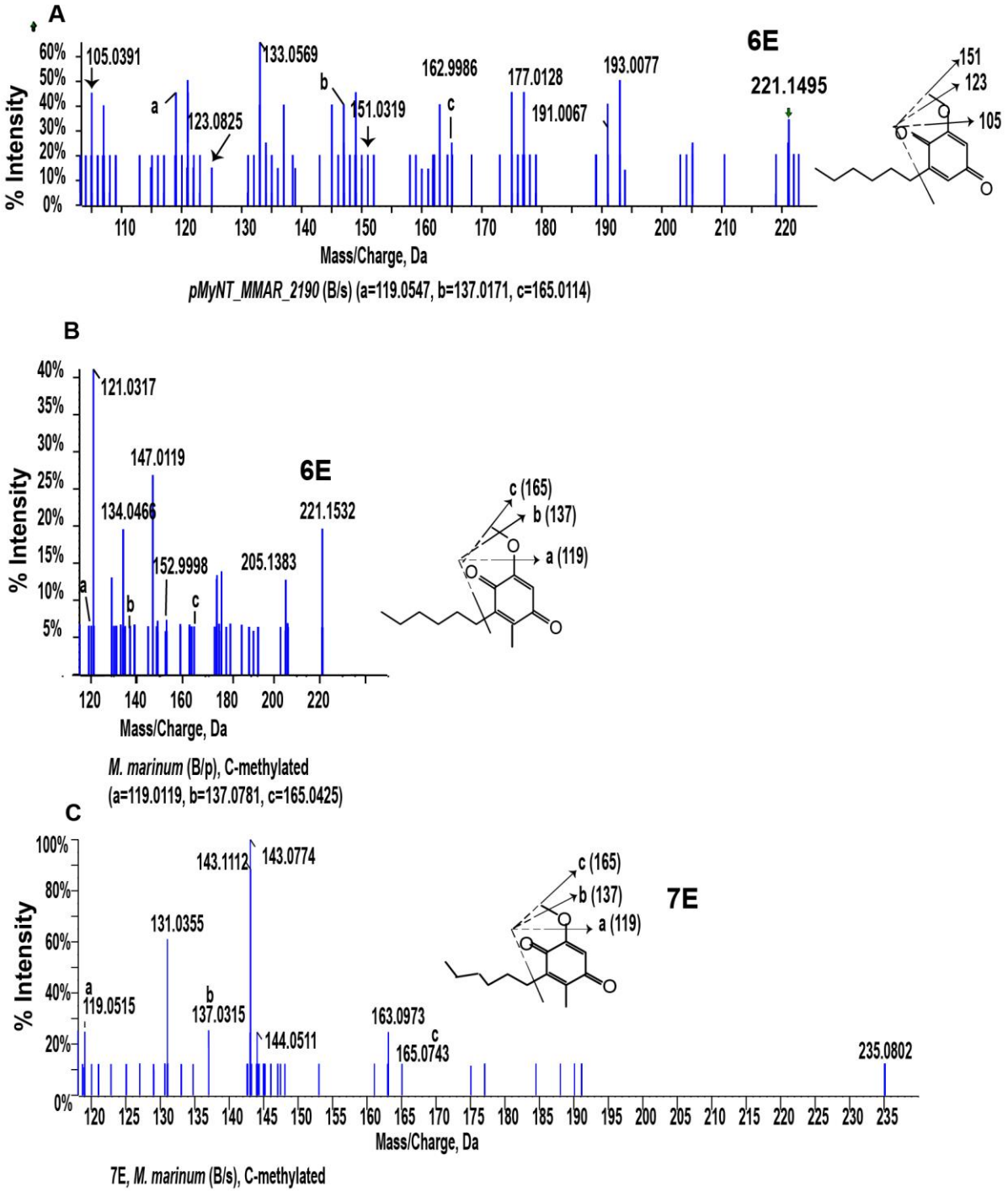


Figure 4.16: Fragmentation of quinones like molecule 6E and 7E observed from pMyNT based overexpressed strain (*pMyNT-MMAR_2190*) and wildtype *M. marinum*. The figure shows the fragmentation of C-methylated even chain quinone 6E (A and B) with precursor ion peak $(M-H)^-$ at m/z 221 (calculated from the structure) get fragmented into sub-fragments with m/z 119, 137 and 165 in *pMyNT_MMAR_2190* and *M. marinum* biofilm culture. The three fragments are denoted by a, b, and c. Besides *pMyNT_MMAR_2190* also showed fragmentation of precursor ion 221 with sub-fragments 105,123 and 151 describing the presence odd chain quinones as illustrated in figure A. Odd chain quinone, 7E present in both *pMyNT_MMAR_2190* and *M.marinum* biofilm culture with precursor ion of m/z 235 giving fragmentation pattern of C-methylated quinone. Figure C shows the fragmentation of C-methylated quinone 7E from *M. marinum* biofilm culture extracted from the supernatant (B/s).

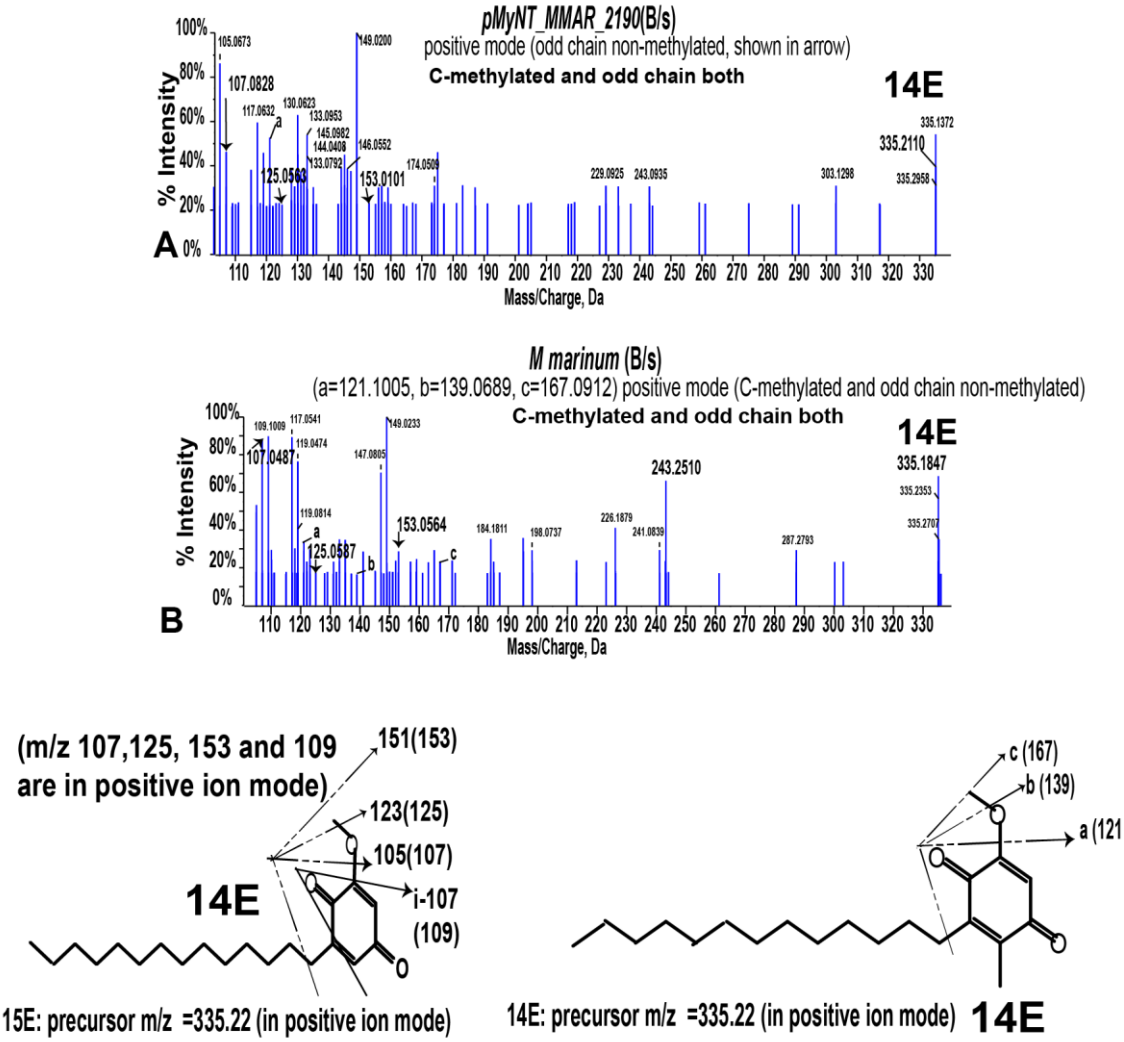


Figure 4.17: Fragmentation of 14E and 15E observed from biofilm cultures of pMyNT based overexpressed strain (*pMyNT-MMAR_2190*) and wildtype *M. marinum*. The figure depicts the presence of C-methylated quinones 14E (A: 14E) and odd chain non-C-methylated quinone (B: 14E) in *pMyNT-MMAR_2190* and *M. marinum* biofilm extracts. The fragmentation of precursor ion at m/z 335.22 (14E) in positive ion mode with sub-fragments with m/z 121, 139 and 167 for even chain C-methylated quinone is shown in figure A and is denoted by a, b, and c respectively. The fragmentation at m/z 335 into sub-fragments with m/z 107,125 and 153 is also observed in figure A and B.

The fragmentation patterns observed for other quinones are shown in Appendix A4.8 to Appendix A4.10.

Atomic Force Microscopy (AFM)

Atomic force microscopy was done to find surface topology of *pMyNT-MMAR_2190* and *pMyNT-MMMAR_NFIII* strains. AFM of control strain was found to have a smooth surface with almost perfect packing, whereas nocardial type III *pks* gene (*pMyNT-NfIII*) overexpressed in *M. smegmatis* knockout of Δpks10 was found to have a fairly rough surface. The *pMyNT-MMAR_2190* overexpressed in control strain has an interesting surface similar to knockout control strain except with bubbly expression of surface molecules (proteins). It was found that possibly the expression of different type III *pks* genes, *MMAR_2190* and *NfIII* affect the cell surface topology and thereby affecting interactions of protein deciding the communication of bacteria overall during the growth. The probe that was used to get the image was a ScanAsyst air probe that gives only the topological map, as shown in figure 4.18. For further details of molecules in overexpressed type III pks strains, we need to use a molecular probe which is out of the scope of our study and further experiments would be required to confirm these observations.

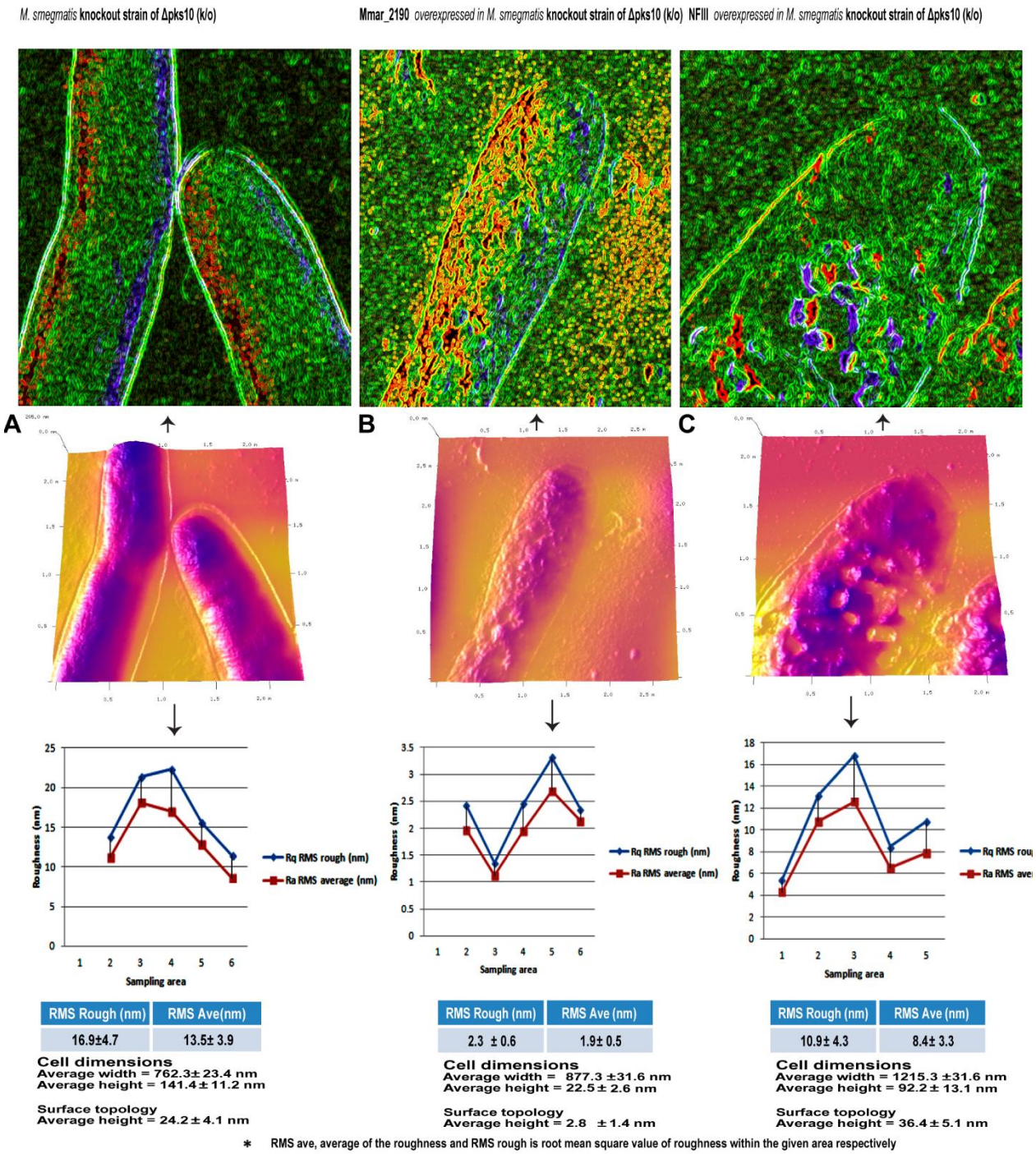


Figure 4.18: Atomic Force Microscopy Study of overexpressed (*pMyNT-MMAR_2190* and *pMyNT-NFIII)* gene on *M. smegmatis* knockout strain of Δpks10 (k/o). Figure A to C show the surface topology with cell dimensions observed for k/o control strain, *pMyNT-MMAR_2190* and *pMyNT-MMAR_NFIII* respectively.

Transmission Electron Microscopy (TEM)

Transmission electron microscopy (TEM) of *pMyNT-MMAR_2190* and *pMyNT-NFIII* strain both overexpressed in *M. smegmatis show* distinct mycobacterial outer membrane, electron transparent layer (ETL) and mycobacterial cell membrane (CM). The diameter measured for the overexpressed strains was found to have 0.4 to 0.6 µm. The difference between two strains was very less except for the ETL was found wavy in *pMyNT-NFIII* ultrathin sections compared to smooth one from *pMyNT-MMAR_2190*. Both the overexpressed strain shows ETL measurement carried out using ImageJ with 11.53±0.5 to 18.37±0.04 nm (unlike *M. tuberculosis* patient-reported ranging from 10.7 nm to 40 nm[216]). The cell diameter measured from the mycobacterial outer membrane (MOM) through ultrathin sections was found ranging from 0.4-0.6 µm similar to the value reported by Yamada *et al*[217](0.51-0.66). The graph for the grey value with distance shows the value for cell membrane (CM) and MOM for the type III *pks* genes from *pMyNT_MMAR_2190* and *pMyNT_NFIII* strains, as shown in figure 4.19.

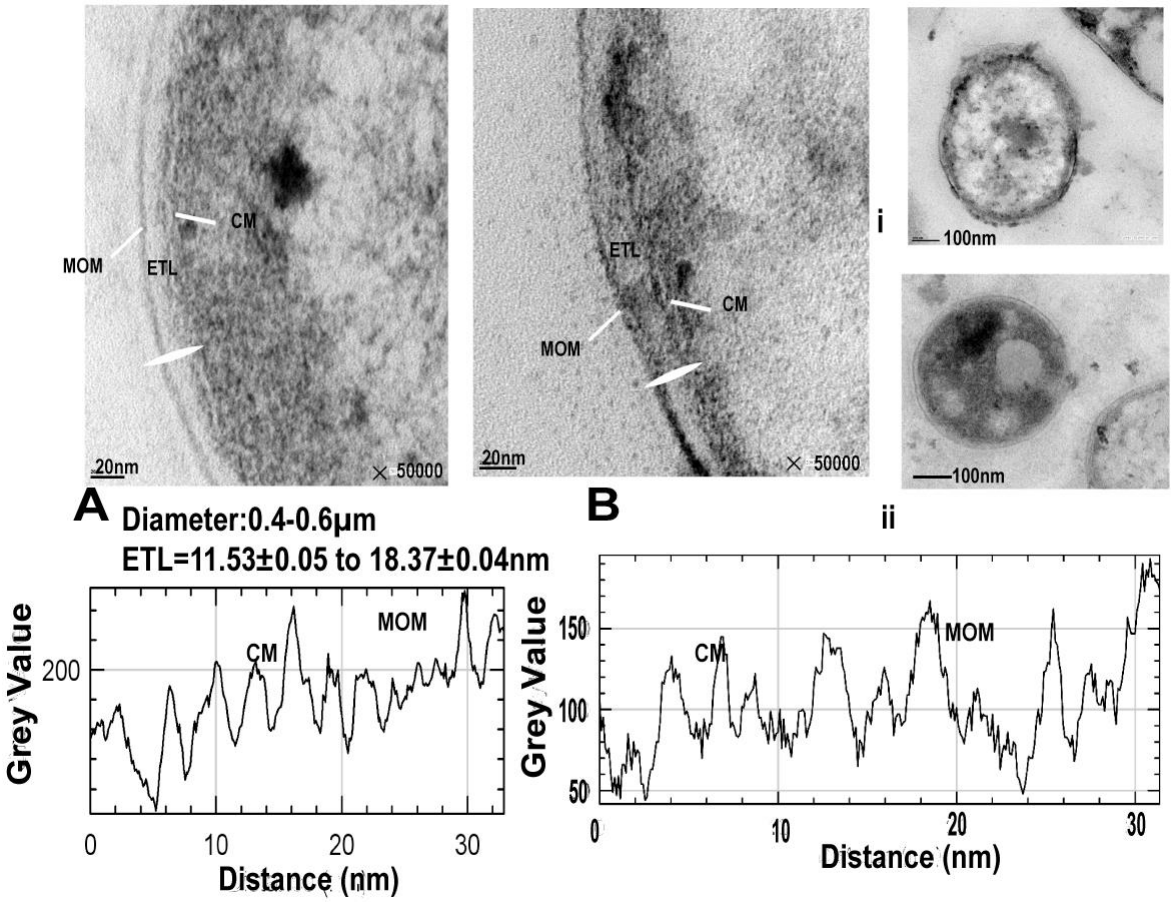


Figure 4.19: Images of ultrathin sections and transmission electron microscopy of *pMyNT-MMAR_2190* (A, ii) and *pMyNT-NFIII* (B, i) expressed in *M. smegmatis* knockout of Δpks10 cells. The graph for distance vs grey value was measured using ImageJ for A and B showing the difference in cell envelope layer between the over-expressed strains.

Scanning Electron Microscopy (SEM)

The morphology of biofilm was observed using Scanning Electron Microscopy (SEM) of *pMyNT* control, *pMyNT-MMAR_2190* and *pMyNT-NFIII* strain and *M. smegmatis* wildtype as shown in below photograph. The overexpressed *pMyNT-MMAR_2190* and *pMyNT-NFIII* displayed a reticulate network of biofilm with interconnected cells similar to the wildtype *M. smegmatis* strain except for *pMyNT-MMAR_2190* biofilm with pigmented pellicle that showed thick extrapolysaccharide covering in SEM, of unknown function. Whereas *pMyNT* control exhibited separate cells with no formations of biofilm and pigment production.

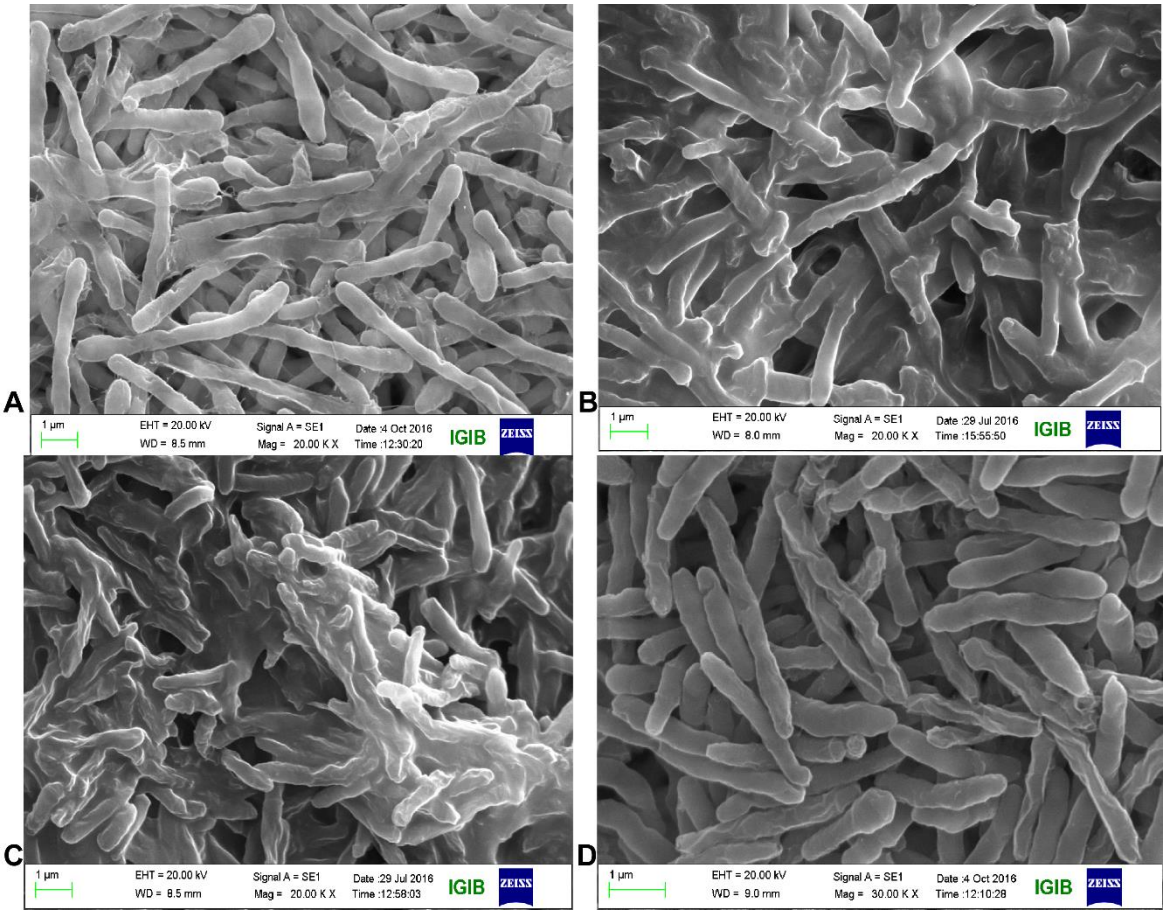


Figure 4.20: Scanning electron microscopy of biofilm cultures from overexpressed *pMyNT* control (A), *pMyNT-MMAR_2190* (B), *M. smegmatis* wildtype (C) and *pMyNT-NFIII* (D) expressed in *M. smegmatis* knockout of Δpks10 cells.

4.4. DISCUSSION

Biofilm formation is a common pathogenic factor of most of the mycobacterial infections[200]. There is a good correlation between biofilm formation and pellicle formation in bacteria. Our study revealed that the Δpks10 knockout strain of *M. smegmatis* (k/o) restores growth in biofilm similar to wildtype *M. smegmatis* (Mc^2155 strain) by overexpression of MMAR_2190. Unlike MMAR_2190, a type III PKS from *M. marinum*, its ortholog NFIII from *N. farcinica* shows no biofilm. In other type III PKS overexpressed strains, *pMyNT-MMAR_2470* was found to have severely impaired biofilm as reported by our group compared to *pMyNT-MMAR_2474* that shows partially recovered biofilm less defective as compared to Δpks10 k/o Msmeg control strain[195]. The factors influencing the difference in biofilm formation include the availability of nutrients[218], carbon source, ions, cofactors like Fe^{2+}, glycopeptidolipids, and gene[215]. In this study, we used non-tuberculous bacteria *M. smegmatis* which grow as a planktonic culture with a tendency to aggregate after three days to form the transition stage between planktonic and biofilm culture. The aggregation might be possibly due to the change in C: N ratio[219].

It is reported that low availability of carbon compared to available nitrogen, bacteria grow as a planktonic culture. We tried to probe the physiological relevance of the overexpressed gene from *M. marinum* and *N. farcinica* in the Δpks10 knockout strain of *M. smegmatis* using metabolomics study from both planktonic culture and biofilm cultures. Type III PKS genes from pathogenic *M. marinum* strain were studied using the most common model organism, *M. smegmatis*. The comparative study of type III PKSs metabolites shows MMAR_2190 and its ortholog protein NFIII undergo catalysis to form even-chain, odd-

chain, non-methylated and triketide *α*-pyrone from 6A to 30A. The metabolomics analysis of pMyNT-strains revealed fragmentation pattern of *O*-methylated triketide *α*-pyrone with unknown functions in both MMAR_2190 and *M. marinum* cultures. The production of triketide *α*-pyrone and its *O*-methylated derivatives is similar to the catalyzed product form in *B. subtilis* by the action of two genes *bpsA* and *bpsB* in an operon[220]. Similarly, *srsA* and srsB from *S. griseus* also form phenolic lipids (alkylpyrones and alkylresorcinols)[12]. The possible mechanism behind the formation of *O*-methylated derivatives is given in chapter 5. The generation of diverse type triketide *α*-pyrone was found more in biofilm cultures of different type III PKS overexpressed strains and some extent in *M. marinum* culture.

We identified different C-chain tetraketide *α*-pyrone from 6B, 8B, 12B, and 14B in *pMyNT-MMAR_2190* and 16B from *M. marinum* wildtype. Among phloroglucinol, a wide range of alkyl phloroglucinol from 6D to 28D were found on different metabolite extracts from planktonic, and biofilm cultures extract from overexpressed strains of *MMAR_2190* and wildtype Mmar. However, we could identify the limited number of phloroglucinol 8D, 10D, 12D, 14D, and 16D products in *pMyNT_NFIII*. We studied the role of salt stress on the metabolome through the introduction of 0 % to 2 % NaCl; it was found that 1 % salt shift the production of tetraketide *α*-pyrone (8B) in the planktonic culture of *MMAR_2190* compared to the one with no salt with phloroglucinol (8D). Unlike 1.5 % salt favour the production of phloroglucinol (8D) in planktonic culture similar to the one identified phloroglucinol (8D) from biofilm extracts. The possible explanation for no/very less colour in *pMyNT-NFIII* compared to red-coloured biofilm in its ortholog *pMyNT-MMAR_2190* is due to varied C-chain alkyl phloroglucinol. However, the identification of only tetraketide *α*-pyrone from *M. marinum* culture might be due to mild salt stress similar to the 1 % salt favouring tetraketide *α*-pyrone (8B) in *pMyNT-MMAR_2190* illustrated in our results. Thin biofilm observed in overexpressed mycobacterial strain compared to the wildtype *M.*

smegmatis and wildtype *M. marinum* possibly due to the requirements of carbon-based precursors and co-production of toxic metabolites inhibiting bacterial growth[221,222].

We found diverse odd- and even-chain alkyl resorcinol from 7C to 19C in both planktonic and biofilm cultures extracts of *pMyNT-MMAR_2190* and *pMyNT-NFIII*. Both odd and even higher chain alkyl resorcinol from 20C to 24C were identified only from the *M. marinum* type III PKSs rather than nocardial ortholog. This might be due to the better compatibility of Δpks10 knockout *M. smegmatis* strain for mycobacterial origin gene expression compared to the nocardial origin. The effect of salt stress on alkyl resorcinol production was opposite to that of phloroglucinol, i.e. less salt stress favour the production of a particular group of alkyl resorcinol better than limited alkyl resorcinol production observed in high salinity. Unlike the stationary phase, biofilm had shown better production of alkyl resorcinol with an increase in mild stress. The benzoquinone ring that gives quinone like molecules with the downstream enzymes, derived from resorcinol was reported by Anand *et al.* act as an electron carrier to survive bacteria under hypoxia. Type III PKS cluster with conserved *MMAR_2190* and cluster of other genes in pathogenic bacteria, possibly play a similar role for the survival of bacteria in stressful conditions of biofilm.

We observed majorly C-methylated even chain quinones like molecule as reported by Anand *et al.* from the biofilm extracts from *pMyNT-MMAR_2190*. The basis of fragmentation pattern in our quinone analysis for C-methylated even chain polyketide quinone is the fragmentation pattern reported by Anand *et al*. Besides we used fragmentation prediction tool of PeakView software from Sciex to devise the pattern in odd and non-methylated quinones of unknown functions as revealed from the biofilm culture study. The study of quinones with salinity shows the production of odd chain C-methylated quinone (14E). Possibly, odd chain C-methylated quinones work as an electron

carrier similar even to chain C-methylated ones to overcome the salinity stress for the survival of bacteria. The presence of even chain C-methylated quinone from biofilm extracts of *M. marinum* biofilm shows the importance of quinone like molecules for the survival of bacteria during stress. Further experiments are needed to confirm these findings.

Chapter 5

Study on type III PKS Modifying Enzymes: Methyltransferase and Sulfotransferase

CHAPTER 5: STUDY ON TYPE III PKS MODIFYING ENZYMES: METHYLTRANSFERASE AND SULFOTRANSFERASE

5.1. INTRODUCTION

Several polyketide cores from type III PKSs undergo modifications to give functional molecules. *O*-methylation and *O*-sulfation of bioactive metabolites on the hydroxyl groups are common modifications widely present in most of the pathogenic microorganisms. Besides these two modifications, bacteria undergo *O*-acetylation, *O*-glycosylation *O*-mannosylation to complete the polyketide core structure. Bacteria undergo *O*-methylation during quinone formation to survive in anaerobic conditions (biofilm) and stress conditions. Methyltransferases are enzymes that methylate the molecule improve the stability of a molecule and determine the efficacy of transport through a biological membrane. *O*-methylated flavonoids used as anticancerous compounds are reported for resistance property to liver enzymes and found to have high absorption potential in the intestine[223,224]. The biological activity of the compound increased by methylation. *O*-methylated compounds are observed in stress conditions[225]. A wide range of the biological property of final products from plants, animals, bacteria, and fungi depends on the tailoring reaction at the electron-rich *O*-position of the acceptor molecules. Two classes of *O*-methyltransferases are commonly present in a wide range of organisms, Class I proteins with a size of 23-28 kDa and class II with size 38-43 kDa. The former uses divalent cation Mg^{2+} as a cofactor[125] to organize the substrate-binding site[126], and the latter uses catalytic histidine residue in place of Mg^{2+} for methylation.

Sulfotransferases are the enzymes widely studied in eukarya for producing sulfated molecules. The sulfated molecules have roles in numerous physiological processes from embryogenesis, blood clotting, inflammation, and cancer to sulfate conjugation of endogenous and exogenous molecules. In both prokaryotes and eukaryotes, sulfurylation of biomolecules occurs. The sulfate group is added from an activated universal donor. PAPS-dependent sulfotransferases catalyze 3'-phosphoadenosine-5'-phosphosulfate (PAPS) to the hydroxyl or amine functional group of a small polar molecule, a carbohydrate, antibiotics, hormones or a tyrosine residue within the protein. Unlike PAPS-dependent sulfotransferases, PAPS-independent sulfurylation there is the use variety of aromatic sulfuric acid esters as sulfate donors and phenolic compounds as an acceptor. PAPS-dependent sulfotransferases were reported to be present in both prokaryotes and eukaryotes cytosol, and Golgi apparatus. Unlike the PAPS-independent were studied from the prokaryotic periplasm[142,160]. Sulfated molecules act as mediators and determinant for the host specificity for the cell to cell interactions, hepatic detoxification and extracellular signaling[154]. NodH and NoeE are the two sulfotransferases from nitrogen-fixing bacteria that catalyze the formation of sulfated glycolipid, a root nodulation factor. In essence, root nodulation factors mediate the communication between the legumes (host) to the bacteria[226]. Sulfated molecules are often found in pathogens, upregulated during infection [227]. The abundance of numerous sulfated glycolipids and their significance in making bacteria strain virulent and acquire drug resistance was identified from the study on human pathogens *M. tuberculosis* and *M. avium*[228-230]. It has been suggested the early stage of infection of M. tuberculosis starts with the adherence of bacteria to the extracellular matrix and virulence of several mycobacterial strains is mediated by sulfated glycoconjugates[231] and sulfolipids[232], respectively.

Our initial *in silico* analysis of a cluster of genes downstream to *MMAR_2190*, a type IIII polyketide synthase revealed two putative sulfotransferases- *MMAR_2191* and *MMAR_2192,* and one methyltransferase- *MMAR_2193*. The identity of genes and the polyketides' tailoring reactions were studied by heterologous expression of each gene in the *E. coli* strain BL21(DE3) and activity against various polyketide substrates.

5.2 MATERIALS AND METHODS

5.2.1 Materials

Bacterial strains *Escherichia coli* XL-1 blue and BL21(DE3) were used for cloning and expression, respectively. Isolated DNA from section 2.2.2.2 was used to amplify the *MMAR_2191*, *MMAR_2192*, and *MMAR_2193* genes using their respective genes specific primers. Wildtype *M. marinum* (strain ATCC BAA-535/M) kindly provided by Prof. Y. Singh (IGIB, India) was grown on Middlebrook 7H9, Middlebrook 7H11 media. The strain was further used for biofilm growth on Sauton's Fluid Media Base. The MtbPKS18 expression clone was kindly gifted by R. S. Gokhale (NII, India) and was used for substrate biosynthesis. S-adenosyl *L*-methionine (SAM) was used as a methyl donor, and 3'-Phosphoadenosine-5'-phosphosulfate (PAPS) used as sulfate donor, acyl-CoA starter and extender substrates were procured from SIGMA. *Escherichia coli* strains were grown on LB medium. Restriction endonucleases and PCR master mix were obtained from New England Biolabs. PCR cleanup kit and Ni^{2+}-NTA agarose were purchased from Qiagen. HPLC and MS grade solvents were purchased from Merck and SIGMA.

5.2.2. Methods

5.2.2.1. *In silico study of PAPS-dependent sulfotransferases and SAM-dependent methyltransferase*

Cluster analysis of *M. marinum*, M complete genome (accession number: CP000854) for type III PKS cluster was done using AntiSMASH[175]. Multiple Sequence alignments of different PAPS-dependent sulfotransferases and SAM-dependent methyltransferase were done by using JalView Software[233]. The NCBI (RefSeq or GenBank) accession number of protein sequences used for sequence comparison. The FASTA format of *MMAR_2193*, *MMAR_2191*, and *MMAR_2192* gene sequences were used as the target sequence to search for the template based on the sequence alignment with Protein Databank (PDB) database. Homology modelling of MMAR_2193, MMAR_2191 and MMAR_2192 models were performed based on the crystal structure of 4NEC_A, 1FMJ and 2GWH downloaded from RCSB protein databank using Biovia Discovery Studio. The models were superimposed with their templates. The structure of the ligand library was generated using ChemDraw Professional 17.0. Protein and ligand were prepared for *in silico* docking. Different docked poses of protein with different ligands were generated using the Biovia Discovery Studio CDOCKER tool. CHARMm-based energy forcefields were used to generate the docked score and predict the correct poses[234].

5.2.2.2 Cloning of gene and sequence analysis

Bacterial genomic DNA was used as a template to amplify methyltransferase gene *MMAR_2193* using a set of the gene-specific forward primer

(5’ **TTCAT**ATGGATTTTGATGCG CTG 3’ containing NdeI restriction enzyme site and reverse primer 5’**TTAAGCTT**GTTTTGCCGCCGCGC 3’) containing a HindIII

restriction site. The *MMAR_2193* gene was amplified using 50 µl reaction mixture. The mixture contains 10 µl (5x) reaction buffer, 10 µl (5x) GC buffer, 1 µl (17.3 ng/µl) genomic DNA, 1 µl (12.5 pmol/µl) of each primer,4 µl dNTP mixture(2.5 mM), 0.6 µl (2 U/µl) Q5 high-fidelity DNA polymerase, and 22.4 µl double distilled autoclaved water. The reaction mixture was incubated in a thermal cycler (Eppendorf, Germany) with an initial hot start at 98ºC for 2 min and then underwent 30 cycles of 30 sec at 98ºC, 45 sec at 60ºC and 50 sec at 72ºC, followed by final extension of 10 min at 72ºC. Similarly, the two sulfotransferase genes *MMAR_2191* and *MMAR_2192* were amplified using a set of respective gene-specific forward primers (5' **AACAT**ATGGCCGCTGCCCAGCCG 3' and 5'**AACAT**ATGGCCACCCGAGTCGGG 3' containing NdeI restriction enzyme site and respective reverse primers 5' **AAAAGCTT**GGACTCCCCCCGCCCGA 3' and 5' **TTAAGCTT**GCCCACCGG CCCGGCG 3') with HindIII restriction site. All the amplicons *MMAR_2193, MMAR_2191* and *MMAR_2192* genes were cloned into the pBLuescript SK+ cloning vector (Stratagene). Restriction digestion and automated nucleotide sequencing were used to find the identify and confirm the clones. Genes encoding methyltransferase and sulfotransferases were further sub-cloned into pET21c (Novagen) expression vector, for protein purification.

5.2.2.3 *Expression and purification of proteins*

BL21/ (DE-3) strain of *E.coli* was used to overexpress MMAR_2193, MMAR_2191, and MMAR_2192 proteins as C-terminal hexahistidine tagged protein. The recombinant *E.coli* BL21(DE3) cells harbouring expression plasmid was grown in Luria Bertani broth incubated at 30 ˚C until the optical density at 600 nm reached 0.5 units. It is followed by the induction of cultures using the 1 mM final concentration of isopropyl-1-thio-*β*-D-thiogalactopyranoside (IPTG) and incubated at 22 ˚C. After 16 hrs, the culture was harvested to pellet down the cells, and the bacterial pellet was re-suspended in lysis buffer

(50 mM Tris, pH 8.0 with 10 % glycerol, 0.15 M NaCl). Each of the recombinant proteins was released by sonication (30 s cycle, 10 s rest, 15 cycles at 30 % amplitude) and the lysate formed was centrifuged at 17000 rpm for 40 min at 4 ˚C to remove cell debris. The supernatant from each protein was collected and used for binding with 0.5 ml Ni^{2+} -NTA slurry. Unbound protein was washed with wash buffer (50 mM Tris pH 8.0, 10 % glycerol). The Ni^{2+} -NTA bound protein was eluted through the imidazole gradient using gravity flow from 5 mM to 150 mM. SDS-PAGE was used to check the purified protein elution. For the co-enzyme assay, MMAR_2190, a type III PKSs clone was used to purify protein similar to MMAR_2193, MMAR_2191, and MMAR_2192, and purified using Ni^{2+} -NTA based affinity purification using a modified protocol of Saxena *et al*[78].

5.2.2.4. *In vitro* enzymatic assays for methyltransferase and sulfotransferases

Enzyme activity assay for methyltransferase, MMAR_2193 and two sulfotransferases (MMAR_2191 and MMAR_2192) was done by using a reaction mixture consisting of each of standard resorcinol, phloroglucinol, and commercially available 5-pentyl resorcinol as a substrate. The *in vitro* reaction was performed for MMAR_2190, a type III PKS products together with MMAR_2193, a methyltransferase. Similarly, the sequential reaction was also carried out with products from type III PKS with MMAR_2193. C-chain starters: C_{16}-CoA with an extender malonyl-CoA was used for the first reaction with MMAR_2190 to get type III polyketide products. The enzymatic reaction was carried out using 100 µM each starter molecule (C_{16}-CoA) and 100 µM malonyl-CoA as an extender molecule. The enzymatic reactions were carried using 50 µg of purified MMAR_2190 at 30 ˚C for 120 min. The reaction was quenched using 5 % acetic acid, and the products were extracted using 2×300 µl of ethyl acetate. The extracts were dried using a vacuum and were resuspended and dissolved using 10 % ethanol in 50 mM HEPES buffer. The products obtained from the first reaction were used as a starter substrate for the second sequential

reaction. *O*-methylated and sulfated products were extracted using an acid-based hydrolysis method.

Methyltransferase and sulfotransferases assays were done using the master mixture containing 50 µM HEPES buffer with extracted products from the first reaction (with type III PKS) or commercially available standard resorcinol, olivetol and phloroglucinol as a starter, 10 mM $MgSO_4.7H_2O$, 0.1 % BSA, pH 7.4). For methyltransferase 400 µM *S*-adenosyl *L*-methionine (SAM) and 0.9495 mg/ml of the MMAR_2193 protein was added in a master mixture containing 50 mM HEPES. For sulfotransferases reaction, 50 µM 3'-phosphoadenosine-5'-phosphosulfate (PAPS) was added as sulfate donor in a reaction buffer (25 mM HEPES), and 428 µg/ml MMAR_2191 and MMAR_2192 proteins were also separately added. Control reactions were kept without enzymes. Reactions were incubated at 30 ˚C for 8h. Products were quenched by adding 5 % of acetic acid. Products were extracted with the 2×equal volume of ethyl acetate was added to the reaction mixture and dried under vacuum. The extracts from the assay were resuspended using 50 µl methanol. The methylated and sulfated polyketide products from standard commercially available resorcinol, olivetol and phloroglucinol was further fractionated in UFLC, and that from the sequential reaction was directly characterized using SCIEX Triple TOF 5600 high-resolution mass spectrometry (HRMS)

UFLC analysis: Reaction products were resolved in reverse phase column (ES Industries, Sonama C5, 5 µ 100 ˚A; 25 cm×4.6 mm). A shallow gradient of 5 % CH3CN in water (each containing 1 % acetic acid) to 30 % CH3CN over 5 min, 60 % CH3CN in 15 min, 80 % CH3CN in 25 min, 90 % CH3CN in 35 min, and 100 % CH3CN in 45 min was used for separation of the products from the assay. Methylated and sulfated products were characterized using SCIEX Triple TOF 5600 HRM.

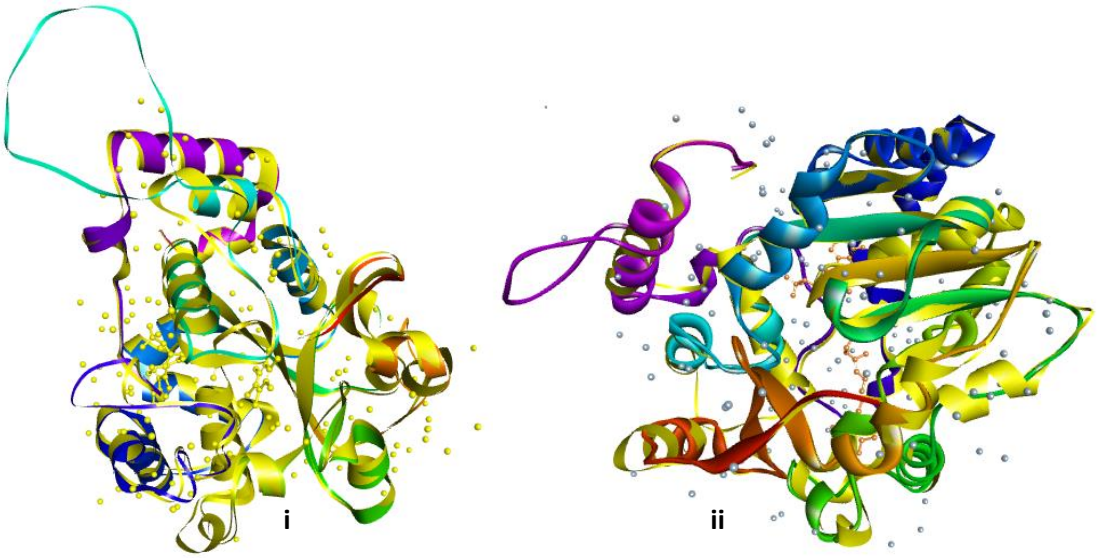
5.2.2.5. Metabolomic analysis from biofilm culture

M. marinum wildtype biofilm pellicles were grown from 1 % inoculum of primary culture in triplicate using (150 mm ×25 mm) sterile polystyrene coated cell culture dish (SIGMA). Biofilm was grown using 70 ml Sauton's fluid medium base supplemented with glucose (2 %) and glycerol (2 %) and was incubated for 14 days as described in chapter 4. The pellicle from biofilm was scraped out and resuspended in 100 mM Tris, pH 8.0 and further acidified to pH 4.0. The metabolites were extracted with the 2×equal volume of ethyl acetate from the acid-based hydrolyzed mixture and dried under vacuum. The extracts from the assay were resuspended using 200 µl methanol. Products from biofilm were characterized using SCIEX Triple TOF 5600 HRM.

5.3 RESULTS

5.3.1 *In silico* study predicting the substrate specificity of the methyltransferase and sulfotransferases

The homology models of MMAR_2191 and MMAR_2192 show RMSD of 1.303 and 2.01 respectively with the template structure 1FMJ and 2GWH. The basis for the template selection for two models of protein shown in figure 5.1, is the amino acid sequence identity of 25 % and 27 % with 1FMJ and 2GWH. Both of the proteins belong to PAPS-dependent sulfotransferases. MMAR_2191 and MMAR_2192 have 5'-phosphosulfate binding (5'PSB) motif and 3'-phosphosulfate binding (3'PSB) motif. The *in silico* studies of two sulfotransferases along with motif analysis of PAPS-dependent sulfotransferases is illustrated in figure 5.1



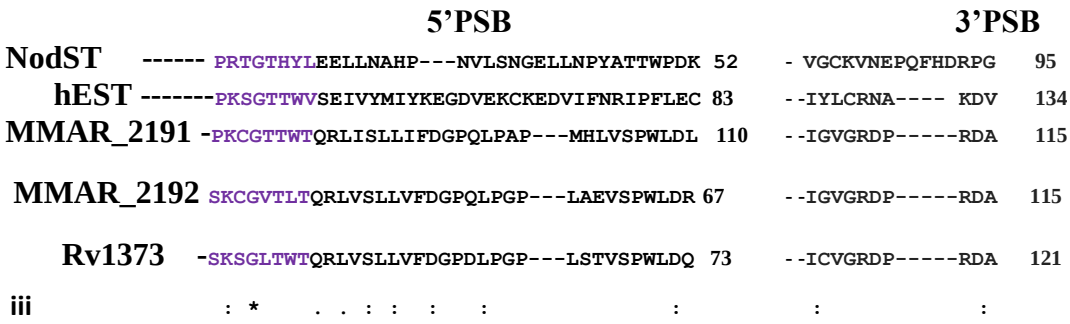


Figure 5.1: *In silico* studies of sulfotransferases. Figure showing superimposed homology models of (i) MMAR_2191 and (ii) MMAR_2192 with two template protein structures 1FMJ and 2GWH respectively and (iii) multiple sequence alignment (MSA) with other PAPS-dependent sulfotransferases. The MSA was generated to find 5'-phosphosulfate binding (5'PSB) motif and 3'-phosphosulfate binding (3'PSB) motif conserved rhizobial NodST (accession no: CDX57624.1), human estrogen sulfotransferase, hEST (accession no: AAC50286.1), two putative sulfotransferase from *M. marinum* cluster (MMAR_2191, MMAR_2192), and the orthologs of MMAR_2192 in *M. tuberculosis* (Rv1373).

Besides two sulfotransferases cluster2 has MMAR_2193, a methyltransferase. Based on multiple sequence alignment, the protein was identified as SAM-dependent methyltransferase with four motifs- motif I with "IGCGLGD", motif II with "DI", motif III with "ADA" and motif IV with "S" as a conserved amino acid. Based on the motif analysis MMAR_2193 has two domains on protein: a SAM binding domain to bind the SAM or SAH mainly at the N-terminal with GxGxG motif, and the substrate-binding domain toward the C-terminal and small segment at N-terminal site as shown in figure 5.2. The homology model of protein with RMSD of 1.234 with template structure of 4NEC_A from Protein Data Bank (PDB) shows a model protein binding to SAH and SAM using the residues as shown in figure 5.3**.** The cavity volume comparison of the ligands/products binding cavities of MMAR_2193 shows the multi-methylated molecules fitted well to the cavity near to the S-Adenosyl *L* homocysteine (SAH) binding cavity, as shown in figure 5.4.

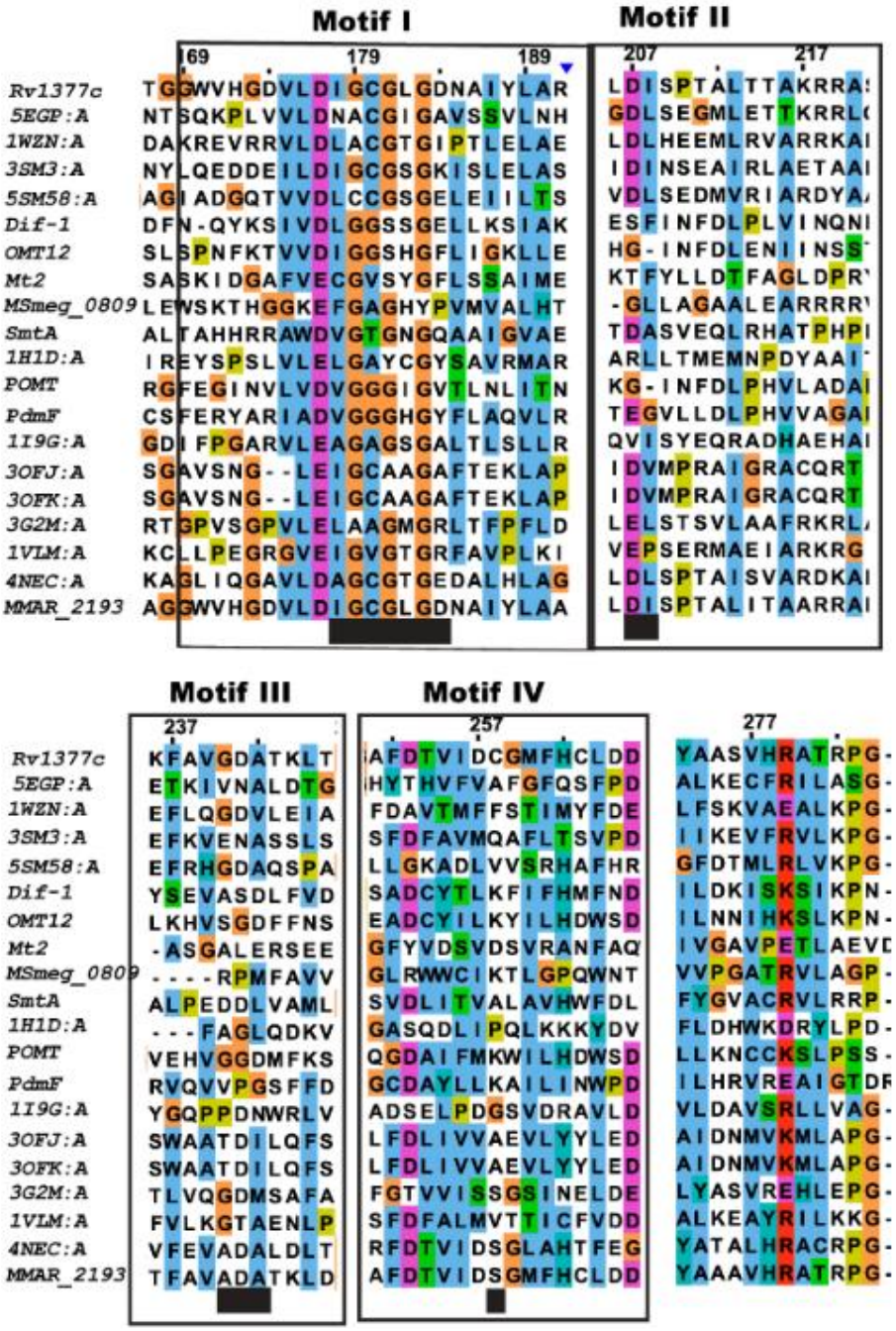


Figure 5.2: Multiple sequence alignment (MSA) of MMAR_2193. The MSA was generated to find N-terminal and C-terminal conserved amino acid for all bacterial and a few *O*-methylated representative plant *O*-methyltransferases (OMTs). Four conserved motifs were analyzed for MMAR_2193 generally conserved in OMTs.

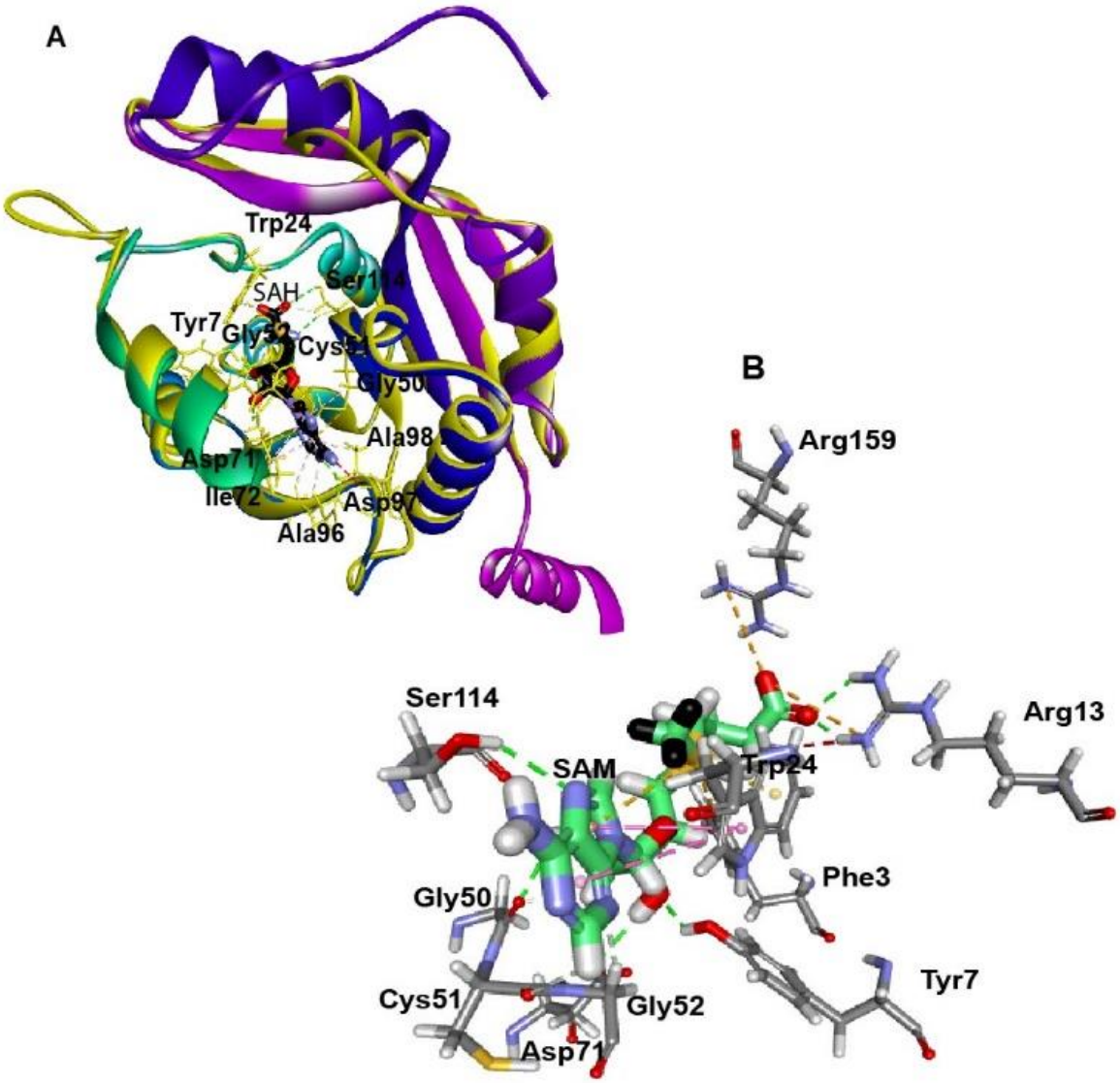


Figure 5.3: Homology Model of MMAR_2193 with interacting *S*-adenosyl *L*-homocysteine (SAH) and *S*-adenosyl *L*-methionine (SAM). The template structure of 4NEC_A from Protein Data Bank (PDB) was used to generate a homology model using the Biovia Discovery study version 4.5. The interaction between protein and methyl donor, SAM (B) and reduced SAH (A) was derived using CHARM based force field. The methyl group of SAM is shown in black colour.

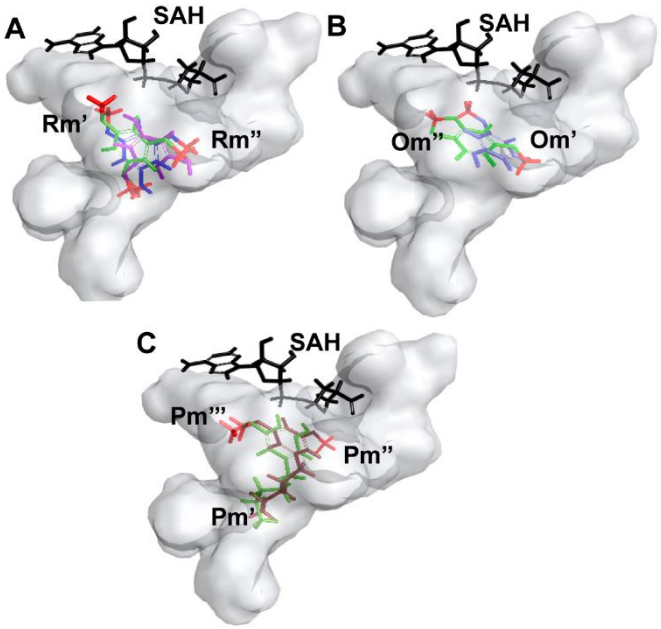


Figure 5.4: Cavity volume comparison of ligand/ products binding cavities of MMAR_2193. Homology model of Mamr_2193 was used to study cavity volume near to the SAH binding cavity for the fitting of methylated resorcinol (A- Rm', and Rm"), methylated olivetol (B- Om', and Om"), and methylated phloroglucinol (C- Pm', Pm" and Pm"')

5.3.2 Cloning of the PKS modifying genes

Modifying genes, *MMAR_2191*, *MMAR_2192,* and *MMAR_2193* were amplified using gene-specific primers using genomic DNA isolated from *M. marinum.* It was cloned into cloning host *E.coli* (Xl-10 blue) using blunt-end cloning vector pBluescript SK+, and the clone was confirmed using restriction digestion with NdeI-HindIII generating fragments of 1095 bp, 981 bp and 636 bp. The clone was further conformed by Plasmid DNA sequencing. The expression constructs *MMAR_2191*, *MMAR_2192* and *MMAR_2193* were made by cloning digested 1095 bp, 981 bp, and 636 bp, fragments in pET21c by using NdeI *and* HindIII sites, figure 5.5.

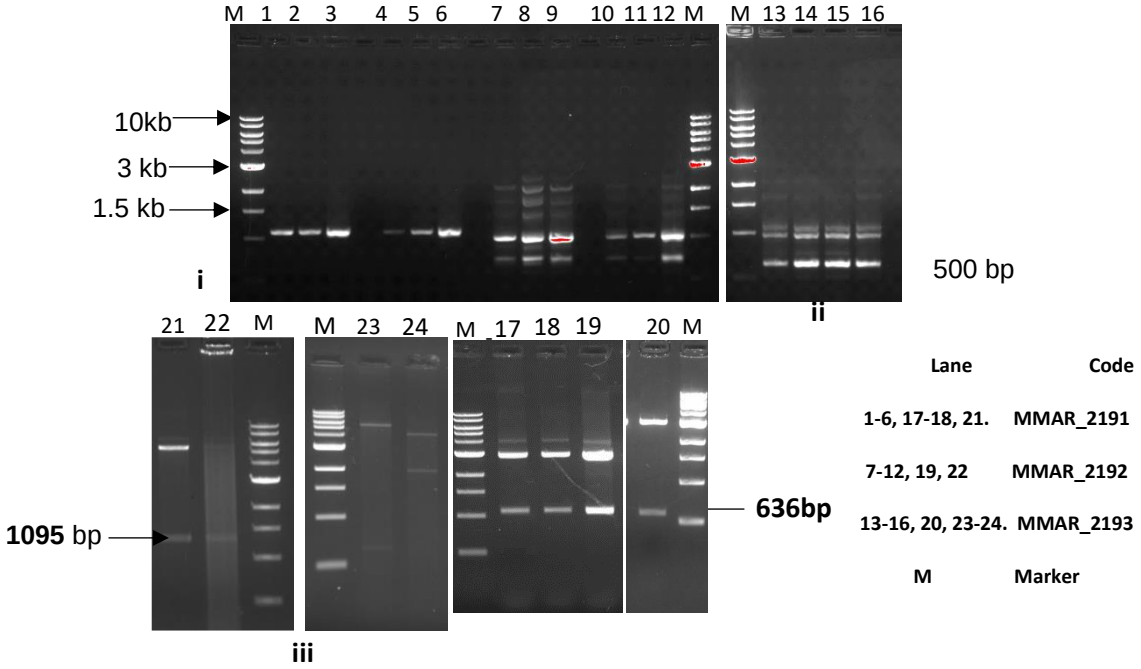


Figure 5.5: Cloning of MMAR_2191, MMAR_2192 and MMAR_2193 and clone confirmation by restriction digestion using NdeI and HindIII-Hf. Amplified modifying genes are shown in (i) from lane 1 to 16, (ii) clone digested from the recombinant plasmid PBS SK+ are shown from lane 17-18 and (iii) the one ligated with expression vector peT21c from lane 21 to 24. Recombinant plasmids from pET21c confirmed were used for overexpression studies.

5.3.3 Expression and purification of polyketide modifying proteins

Type III polyketide modifying proteins MMAR_2191, and MMAR_2192, MMAR_2193 (cloned in the pET system) and pET21c (an empty vector) were independently transformed into BL21/DE3 strain of *E.coli*. Small scale cultures were grown at 30 ˚C with 0.5 mM IPTG induction at logarithmic (log) phase followed by protein overexpression at 22 ˚C for 8 hrs. The overexpressed proteins of size approximately 25 kDa and 42 kDa were purified by Ni^{2+} -NTA affinity chromatography, as shown in figure 5.6 (lane 1-24). Large scales cultures were grown at the same condition, followed by IPTG induction and overexpression at 22 ˚C for 10 hrs.

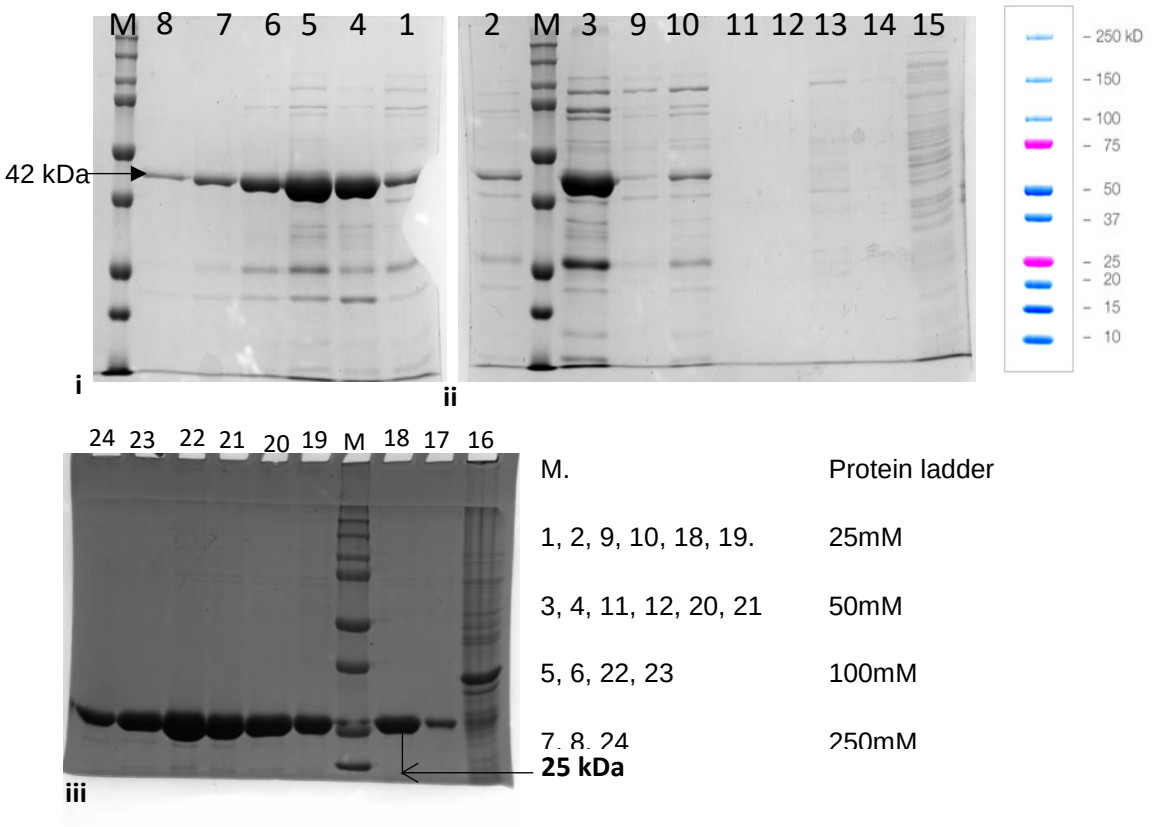


Figure 5.6: SDS-PAGE profile of Ni^{2+} -NTA purified (i) MMAR_2191, (ii) Mmar_2192, and (iii) Mmar_2193 proteins

5.3.4 Biochemical activity of polyketide modifying methyltransferase

It was inferred that MMAR_2193 could catalyze *O*-methylation of varied substrates including triketide *α*-pyrone, resorcinol, 5-pentyl resorcinol and phloroglucinol from our *in silico* docking studies. We used a pET based recombinant plasmid system to conjugate *M. marinum MMAR_2193* gene and heterologous over-express it in Bl21/DE3, an *E.coli* strain. MMAR_2193 was expressed as a hexahistidine tagged protein and purified using Ni^{2+}-nitrilotriacetic acid affinity chromatography as a single protein band of ~25 kDa was determined on SDS-PAGE (Figure 5.7). Recombinant MMAR_2193 was purified. *In vitro*, enzymatic assays were done using SAM as a methyl donor and resorcinol/phloroglucinol/olivetol as the acceptor substrates. Extracted reaction products resolved using ultra-fast liquid chromatography (UFLC) and subjected to high-resolution

mass spectrometry (HRMS) corroborated in silico predictions. The UFLC peaks in figure 5.7 revealed methylated products from resorcinol, phloroglucinol and olivetol, respectively, in our HRMS analyses. Fragmentation pattern in Figure 5.8 revealed ions with [M-H]- at m/z 122.9031 and m/z 136.9603 from resorcinol primed reaction. Reaction with phloroglucinol led to product ions with [M-H]- at m/z 138.9629, m/z 152.9535 and m/z 166.9535. Olivetol primed reactions formed product ions with [M-H]- at m/z 191.1410 and m/z 207.0473 in profile. Tandem MS/MS analyses as shown in figure 5.8, confirmed these product ions as mono-methylated resorcinol (Rm', [M-H]- at m/z 122.9031: fragments at m/z 95.0132, 93.0363, 68.9971 and 53.0019); di-methylated resorcinol (Rm", [M-H]- at m/z 136.9603: fragments at m/z 120.9648, 93.0545 and 76.9827); mono-methylated phloroglucinol (Pm', [M-H]- at m/z 138.9629: fragments at m/z 122.9633, 109.0423, 96.9766, 94.9446 and 80.9300); di-methylated phloroglucinol (Pm", [M-H]- at m/z 152.9535: fragments at m/z 120.9789, 108.9605 and 96.9756); tri-methylated phloroglucinol (Pm"', [M-H]- at m/z 166.9535: fragments at m/z 151.0087, 134.9532); mono-methylated olivetol (Om', [M-H]- at m/z 191.1410: fragments at m/z 177.0893, 165.1183, 149.0827 and 135.0674) and di-methylated olivetol (Om", [M-H]- at m/z 207.0473: fragments at m/z 191.1410, 144.9777, 143.0722 and 117.0314). Appendix figure 5.2 and appendix table 3 provide details of the mass spectrometric characterization of the methylated molecules. Our functional characterization of MMAR_2193 provided evidence for the *O*-methylation potential of the mycobacterial protein. Notably, MMAR_2193 displayed the capability to perform multiple *O*-methylations on a single substrate molecule.

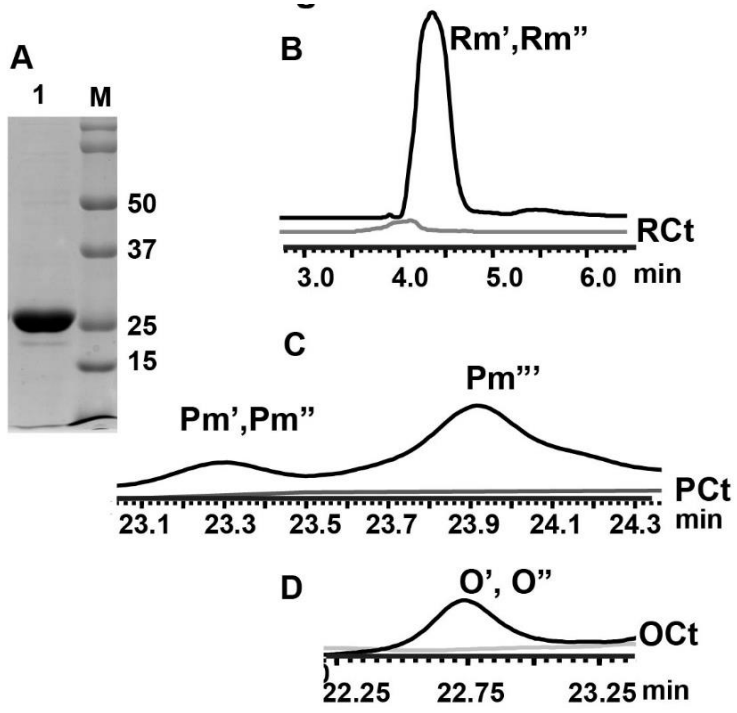


Figure 5.7: Purified methyltransferase and fractionation of methylated products using Ultra-Fast Liquid Chromatography (UFLC). A shows the purified MMAR_2193 with approximate size 25 kDa, B, C, and D display the fractionation of methylated polyketide products concerning their control RCt, PCt, and Oct. B shows the fractionation of mono (Rm')- and di (Rm")-methylated resorcinol together in 4 min. C shows the fractionation of mono (Pm')- and di (Pm")-methylated phloroglucinol from 23.1 to 23.4 min whereas, tri-methylated phloroglucinol (Pm"') fractionated at 24.1 min respectively. D shows the fractionation of mono-methylated (Om') and dimethylated olivetol (Om") from 22.25 to 23.25 min.

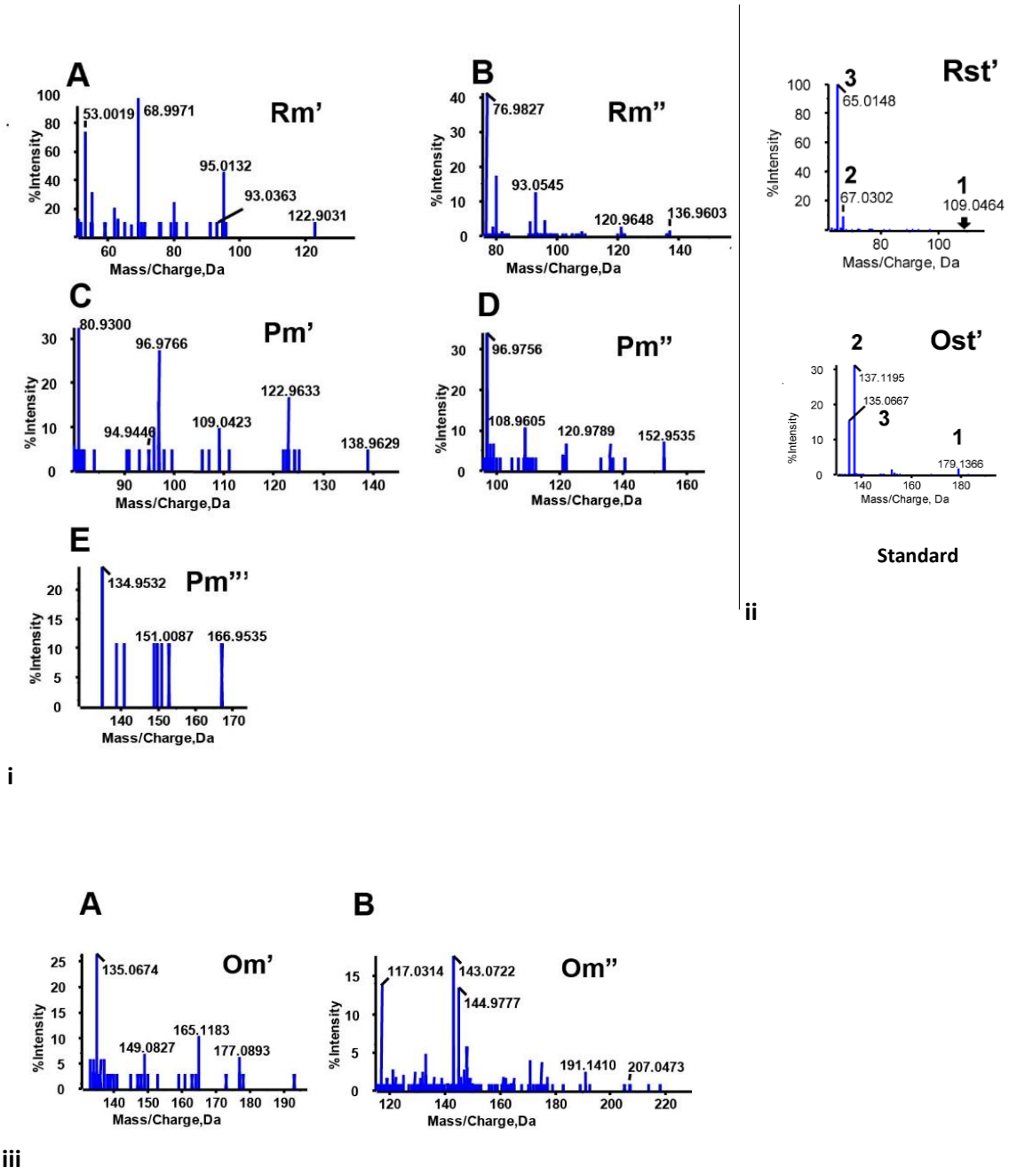


Figure 5.8: Tandem mass spectrometry for different methylated products formed from (i) resorcinol, phloroglucinol, and (iii) olivetol. Standard resorcinol Rst' and commercially available standard olivetol, Ost' fragmentation pattern is shown in (ii).

5.3.4.1 *O*-methylated alkyl *α*-pyrones identified from sequential *in vitro* reaction and *M. marinum* biofilm

MMAR_2190, a type III polyketide synthases biochemically characterized in chapter 2 results in the catalysis for the formation of wide arrays of products with diverse scaffolds.

We tried to perform the sequential reaction of two enzymes MMAR_2190 and MMAR_2193. MtbPKS18, a type III PKS from *Mycobacterium tuberculosis* has been previously characterized to catalyze the biosynthesis of long-chain triketide- and tetraketide alkyl *α*-pyrones in cell-free assays. In a similar *in vitro* reaction, we incubated MMAR_2190 protein with palmitoyl-CoA (C_{16}-CoA) and malonyl-CoA to biosynthesize triketide- and tetraketide palmitoyl *α*-pyrones, 16A and 16B, respectively. These metabolites were extracted from the quenched assays and used as substrates for a sequential reaction with SAM and MMAR_2193 protein. An HRMS analysis of the methyltransferase reaction products identified ions of [M-H]- at m/z 335.1987 and m/z 377.2351. Tandem MS/MS of these precursor ions confirmed the identity of these molecules as methylated alkyl *α*-pyrones, 16A' and 16B', respectively (Figure 5.9).

M. marinum harbours several genomic clusters with polyketide biosynthetic genes, including the MMAR_2193 cluster that is exclusively identified in pathogenic species. Biofilms have lately been associated with several pathogenic diseases[198,199,235,236] and are a natural form of existence in mycobacteria. We probed the possibility of the presence of *O*-methylated polyketides in wild-type Mmar biofilms. Biofilm pellicles were developed for Mmar cells (appendix A5.1) and extracted for HRMS metabolomics analyses. A multiple-reaction-monitoring (MRM) based metabolomics approach identified two ions of $[M-H]^-$ at m/z 279.2257 and m/z 335.1987. Tandem MS/MS confirmed identities of *O*-methylated triketide lauroyl *α*-pyrone (12A') and triketide palmitoyl *α*-pyrone (16A') corresponding to the two identified ions, respectively. Our results revealed the presence of *O*-methylated polyketides in wild-type Mmar biofilms.

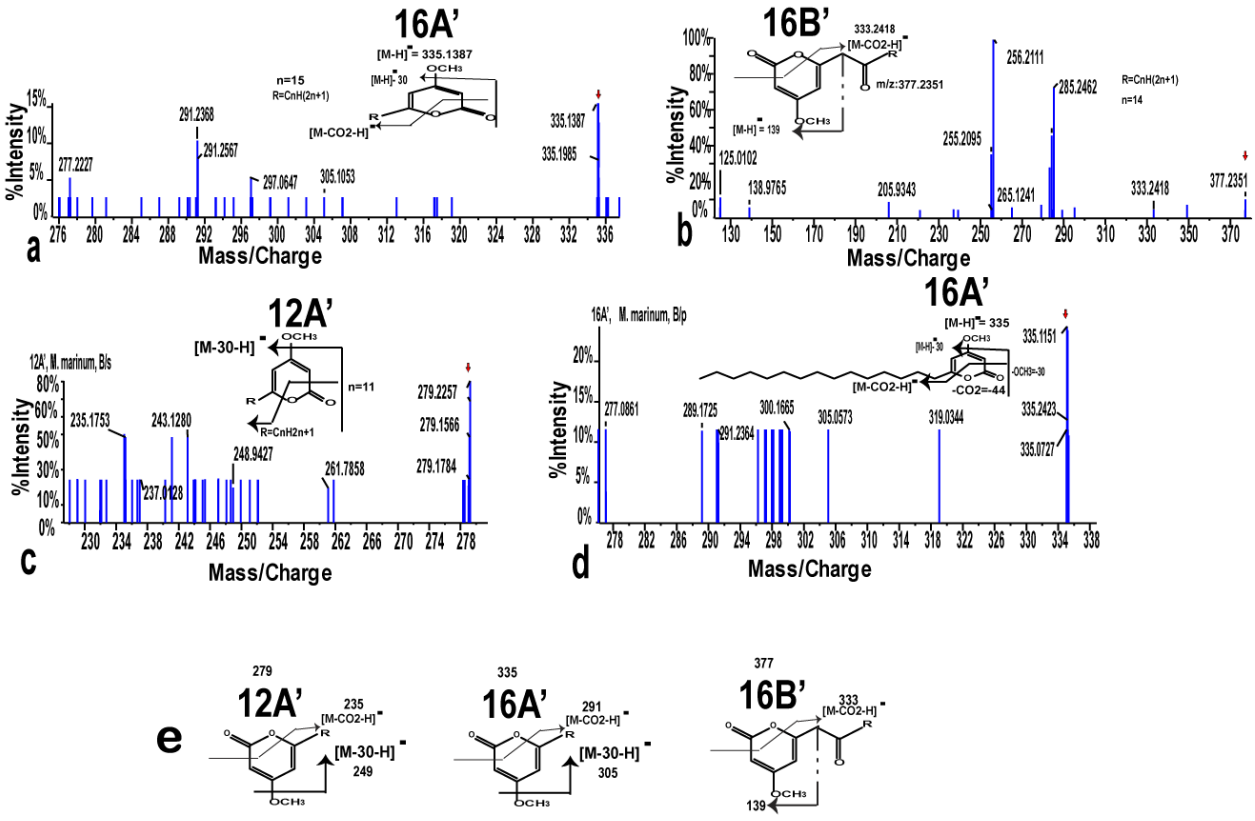


Figure 5.9: Tandem mass spectrometry of methylated *α*-triketide pyrones (12A' and 16A'), tetraketide *α*-pyrone (16B') from sequential assay and wildtype *M. marinum* biofilm extracts. The subsequent assay from products of MMAR_2190 with methyltransferase shows the formation of methylated triketide *α*-pyrone (a: 16A') and tetraketide *α*-pyrone (b: 16B') formed from 16C-acyl CoA as starter and malonyl-CoA as an extender. The metabolomics profiling of extracts from wild-type *M. marinum* also shows the presence of triketide *α*-pyrone (12A' and 16A') as shown in c and d. Figure e shows the fragmentation pattern for the respective methylated triketide and tetraketide *α*-pyrones.

5.3.5. Reconstitution of biochemical activity of Sulfotransferases

The two putative sulfotransferase- MMAR_2191 and MMAR_2192 show PAPS-dependent sulfation of type III polyketides: resorcinol, phloroglucinol and olivetol. *In vitro* assay of polyketide products as substrate and 3'-phosphoadenosine-5'-phosphosulfate (PAPS) as sulfate donor with respective sulfotransferases give mono- and di-sulfated resorcinol and olivetol, and mono-, di- and tri-sulfated phloroglucinol. The products were fractionated, as shown in the chromatogram (figure 5.10).

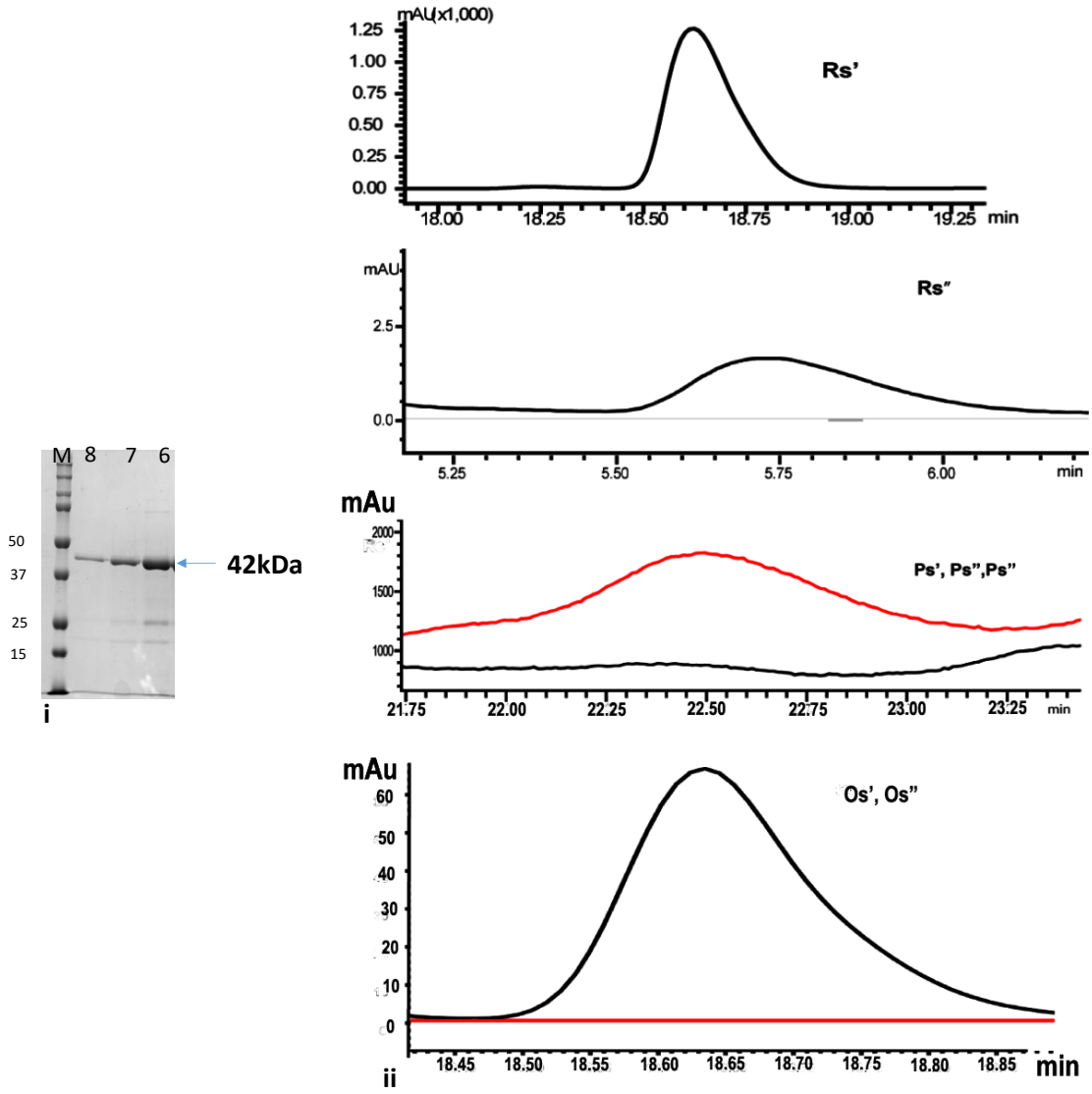
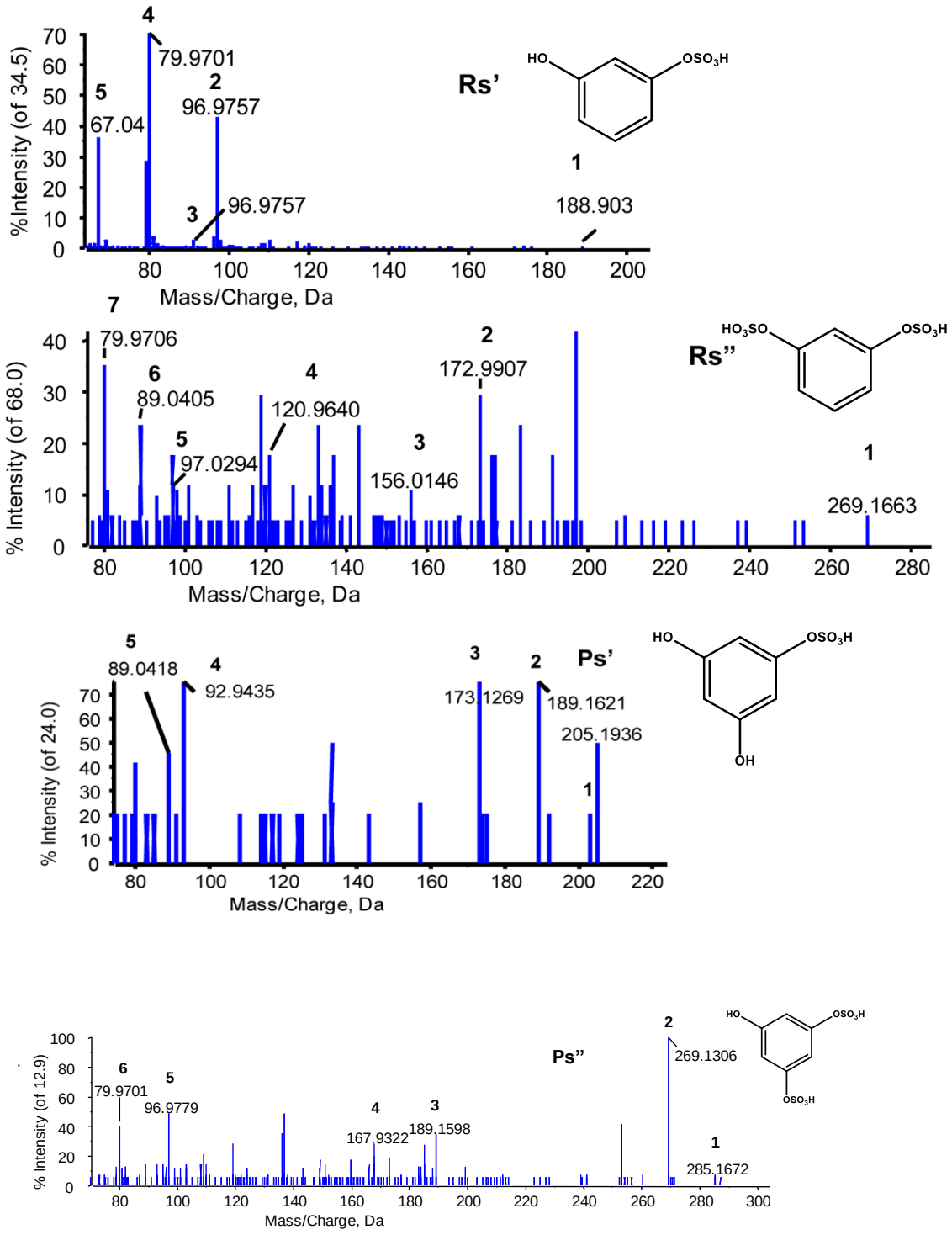


Figure 5.10: Purified sulfotransferase and fractionation of sulfated products using Ultra-Fast Liquid Chromatography (UFLC). SDS-PAGE gel picture (i) shows the purified sulfotransferase with approximately size 42 kDa. Chromatograms (ii) shows the fractionation of sulfated polyketide products concerning their control with fractionation of mono (Rs')- sulfated resorcinol in 18 to 20 min and di(Rs")-sulfated resorcinol in 5 to 6 min. Fractionation of mono(Ps')-, di(Ps")- and tri(Ps"')-sulfated phloroglucinol from 21.75 to 23.25 min whereas, mono(Os')- and di(Os")-sulfated olivetol fractionated at 18.45 min to 18.85 min.

The identity of the products was further confirmed using tandem mass spectrometry, as shown in the below figure (figure 5.11). The basis of the fragmentation pattern for the methylated and standard compounds is described in Appendix table 3.


Rs'
HO
OSO3H
%Intensity (of 34.5)
Mass/Charge, Da
79.9701
96.9757
67.04
96.9757
188.903
Rs"
HO3SO
OSO3H
% Intensity (of 68.0)
79.9706
89.0405
120.9640
97.0294
156.0146
172.9907
269.1663
Ps'
OH
% Intensity (of 24.0)
89.0418
92.9435
173.1269
189.1621
205.1936
Ps"
% Intensity (of 12.9)
79.9701
96.9779
167.9322
189.1598
269.1306
285.1672

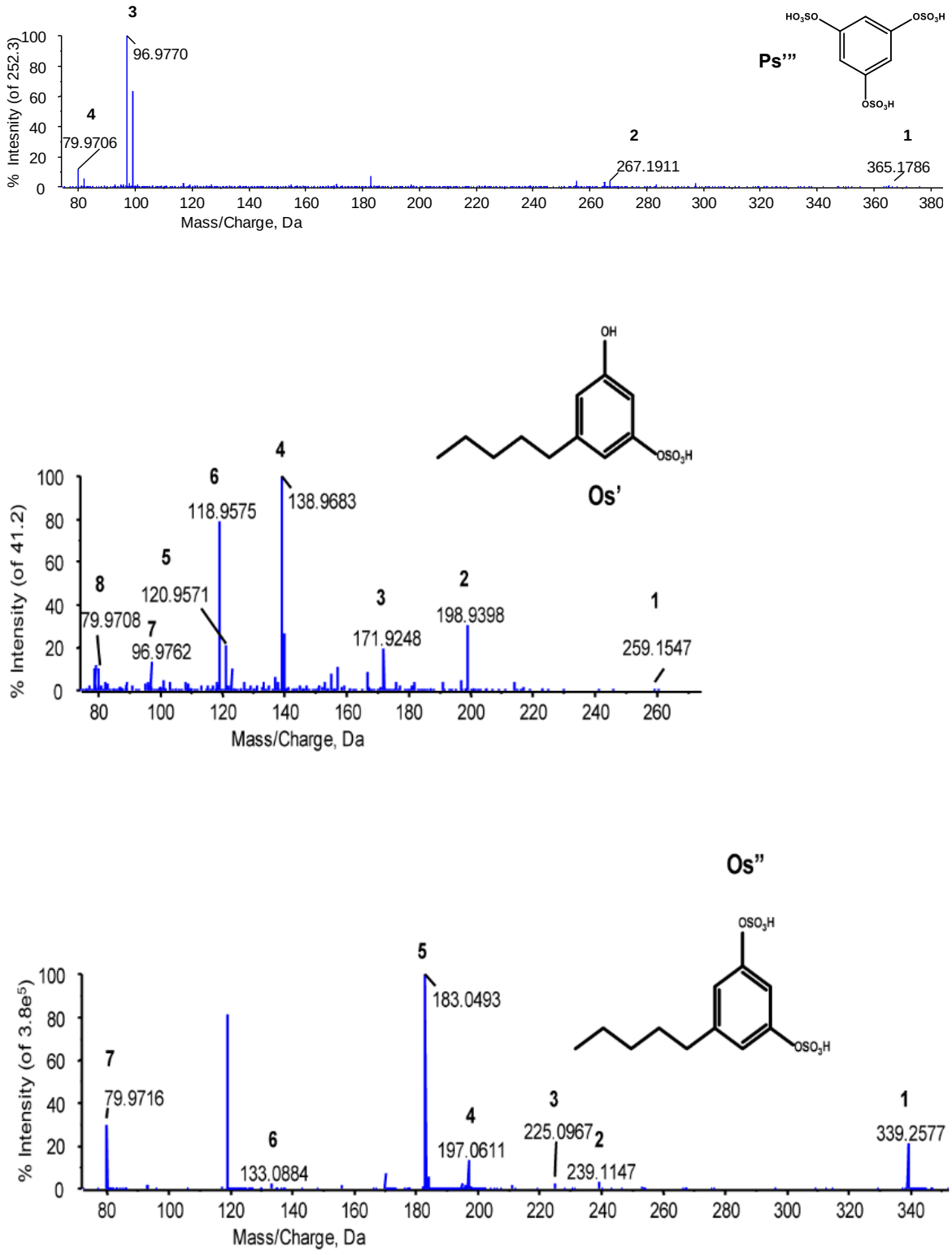


Figure 5.11: Tandem mass spectrometry for different sulfated products formed from resorcinol (Rs', Rs"), phloroglucinol (Ps', Ps" and Ps"') and olivetol (Os', and Os"). The precursor mass for parent compound is dented by 1 and MS/MS fragments denoted by 2, 3 and so on…

5.4. DISCUSSION

Members of the corynebacterial sub-order are reported to have a profound presence of large numbers of methylated and sulfated secondary metabolites. During the metabolism of drugs and xenobiotics, mono-conjugated metabolites produced by catechol *O*-methyltransferase get sulfated by the second group of enzymes called sulfotransferases (SULTs). SULTs are present both in eukaryotes and prokaryotes[237-240]. Our study on methyltransferase and sulfotransferases is based on the conserved signature sequences of SAM/SAH-binding motif and PAPS-binding motif, respectively. Their substrates binding sites predicted through *in silico* studies are used to devise possible substrates for *in vitro* assay of respective enzymes. We tried to predict the high accuracy of substrates for methyltransferase taking into consideration substrate-bound with and without SAH, the docking studies and the cavity volume analysis of the homology model to dictate the screening candidate substrate compounds. A similar, strategy is used to screen high accurate substrate compounds [241] for sulfotransferases using bound with and without PAP(S). *M. tuberculosis* has sulfated and methylated bioactive compounds including trehalose-2-sulfate esters, with a rich pool of various sulfolipids. The bioactive molecules act as a biomarker of virulence usually present in pathogenic mycobacterial strains[229]. We studied methyltransferase and sulfotransferases by heterologous expression in *E. coli*. The in vitro assays show the multiple methylation and sulfation of resorcinol, phloroglucinol and 5-pentyl resorcinol. MMAR_2191 and MMAR_2192 found downstream to the MMAR_2190, a type III PKS, were found to have sulfation of polyketide metabolites. *MMAR_2192* has an orthologous gene- *Rv1373* from *M. tuberculosis*. Both the genes undergo PAPS-binding and transfer of sulfate. The real *in vivo* substrates for both the

enzymes are not characterized yet. Rv1373 catalyzed transfer of sulfate to eukaryotic sulfatide (galactose, glucose and lactose ceramide)[176].

Whereas, *M. marinum* sulfotransferases catalyze polyketide sulfation to form multi-sulfated molecules or derivatives thereof. This property of sulfotransferases, the enzyme can be further utilize to synthesize sulfated (*O*- sulfonated) small phenolic compounds, with a broad range of functions. *O*-sulfated molecules derived from different PAPS-dependent *O*- sulfotransferases are reported controlling cell signalling, regulate development processes, maintain communications between plant and bacteria, transport of molecules, play a role in stress responses, act during detoxification processes and used in antibiotics biosynthesis in different organisms[161,242-248]. Sulfotransferases from *M. marinum* can have applications in different sulfation reactions using different templates in understanding the life process, host-pathogen interactions and synthesize new compounds in chemical and pharmaceutical industry.

Although there are few studies on SAM-dependent *O*-methyltransferases from bacteria, the complete role of the enzyme is still unknown. *O*-methylation act as common modification in the biosynthesis of many natural products. Different bacterial methyltransferases are reported with *O*-methylation of varied plant-based polyphenolic compounds including flavonoids, isoflavones, flavonols, flavones, flavanones, naringenin and Cirsimaritin[249,250]. A plant-based *O*-methyltransferases characterized recently from the Amaryllidaceae family were found to have multiple *O*-methylation reactions at alkaloids (norbelladine)[251]. One of the methyltransferases (ypbQ) studied from gram-positive microorganisms *Bacillus subtilis* with bcsA-ypbQ operon was found to methylate a polyketide triketide α-pyrones to alkyl pyrone methyl ethers[220]. MMAR_2193 displays novel functionality to catalyze multiple *O*-methylation on all available hydroxyl positions

in a single turns substrate molecule. The protein can be utilized to generate a palette of a chemically distinct and functionally diverse family of methylated bioactive molecules. Whereas, PMT2 (pinosylvin *O*-methyltransferase), and CCoAOMT, the similar plant proteins to MMAR_2193 methylate a wide variety of substrates such as flavonoids, alkaloids and stilbenes (only plant products) with single methylation[252].

Summary and Conclusions

SUMMARY AND CONCLUSIONS

The genome sequencing of different corynebacterial strain led to the identification of a cluster of genes coding for type III polyketide synthases (PKSs) and modifying enzymes. Type III PKSs from different organisms are reported to generate functionally diverse chemical compounds with multiple keten groups[29,39,40,42,78,170,177]. We used comparative genome study of open reading frames of some of the polyketide synthase (PKSs) and modifying enzymes with their orthologues in different corynebacterineae were predicted to form hundreds of polyketide or modified molecules to predict their roles in mycobacterial pathogenesis.

In this thesis, we have performed biochemical characterization of type III PKS, sulfotransferases and methyltransferase. We used *in vitro* assay with different starters acyl-CoA (C_{14} to C_{22}) and extender malonyl-CoA to confirm the identity of products based on Mmar targeted metabolomics approach, using mass spectrometry. Product formation of Mmar type III PKS, MMAR_2190 is explained using structural analysis. Overexpression studies of mycobacterial MMAR_2190, MMAR_2470, MMAR_2474 and nocardial NFIII protein in Δpks10 knockout strain of *M. smegamtis* were used to study the physiological relevance of type III PKSs from *M. marinum* and *N. farcinica*.

The first chapter describes type III PKSs and modifying enzymes (methyltransferase and sulfotransferases) in details. The chapter talks on the structural and biochemical perspectives of different type III PKSs from plants, algae, bacteria and slime molds with diversity in the formation of products, and possible cyclization mechanisms. It also

describes different types of methyltransferases and sulfotransferases and importance of product modifications.

The second chapter describes biochemical characterization of a MMAR_2190 with unusual substrate specificity and cyclization potential to form four types of polyketide molecules- triketide pyrone (a), tetraketide pyrone (b), alkyl resorcinol (c) and acyl phloroglucinol (d). An orthologous type III PKS protein MtbPKS18, from *M. tuberculosis* was earlier reported to produce only triketide pyrones and tetraketide pyrones[78]. Pyrones formed from long-chain fatty-acyl chains are speculated to modulate cell envelope of mycobacteria. The third chapter describes the three-dimensional structural characterization of *M. marinum*, MMAR_2190 by using an *E. coli* strain in collaboration with Dr Krishnan Vengadesan group at Regional Centre for Biotechnology (RCB), India. The 2.3 Å crystal structure of MMAR_2190 reveals substrate-binding tunnel similar to that observed in MtbPKS18[54]. Crystal structure of MMAR_2190 shows malonated active site cysteine, and the difference in orientation of gating residues compared to MtbPKS18. These changes could aid the divergent product profile of MMAR_2190.

In chapter four, we investigated possible roles of corynebacterial type III PKSs growth, survival and cell envelope modulation. Type III PKSs, MMAR_2190 and an orthologous protein NFIII from *N. farcinica* were overexpressed in Δpks10 knockout strain of Msmeg using pMyNT inducible system. Both genes belong with MtbPKS18, belong to an orthologous genomic cluster, named cluster2. MMAR_2470 and MMAR_2474 belong to cluster1 to which MtbPKS10 and MtbPKS11 belong. Type IIII PKSs from Cluster1 and cluster2 are exclusively present in pathogenic corynebacterial strains. We studied the physiological relevance by overexpressed pMyNT-strains containing type III PKSs belonging to cluster1 and cluster2 using biofilm and planktonic culture-based studies. The *pMyNT_MMAR_2190* strain was found to have a better recovery of biofilm with

pigmentation, and *pMyNT_MMAR_2474* strain was found to have a partial recovery of biofilm. The restoration of biofilm is similar to the wildtype *M. marinum* and thick biofilm of wildtype *M. smegmatis* Mc^2 155. *pMyNT_MMAR_2470* strain, in contrast, was found to form fragile and layers of cells. Similarly, *pMyNT_NFIII* and knockout strain failed to form a biofilm. Metabolomics analysis of *pMyNT_MMAR_2190* strain revealed the formation of non-methylated a, b, c, d, C-methylated even chain quinone like molecules and odd chain non-C-methylated quinones. The metabolome of *pMyNT_NFIII* showed limited C-chain non-methylated a, b, c and d similar to *pMyNT_MMAR_2190* with no C-methylated even chain quinone. Generation of C-methylated bioactive molecules a, b, c and d by *pMyNT_MMAR_2474*, and *pMyNT_MMAR_2470* has been characterized by our group recently[195]. *pMyNT_MMAR_2190* planktonic cells revealed an alteration in the cell envelope, linking towards possible roles of these metabolites in cell wall remodellation. Further, studies should confirm these findings.

Chapter five describes the biochemical characterization of MMAR_2191 and MMAR_2192 and MMAR_2193 proteins that are present downstream to MMAR_2190 in the genomic cluster2. The two sulfotransferases MMAR_2191 and MMAR_2192 were found to form *O*-sulfated polyketide products. Whereas, MMAR_2193 was characterized to perform multiple *O*-methylation of standard resorcinol, 5-pentyl resorcinol and phloroglucinol. The protein was also found to *O*-methylate C_{16}-CoA derived polyketide molecule, alkyl triketide α-pyrone. Metabolomics analysis of biofilm of wildtype *M. marinum* strain revealed *O*-methylated triketide α-pyrone showing the significance of tailoring reaction in mycobacterial physiology.

Thus, in conclusion, the thesis is a piece of work to dissect the role of different type III polyketide synthases, methyltransferase and sulfotransferase from corynebacterineae using biochemical and structural characterization of proteins and metabolomics study of the

overexpressed genes in mycobacterial host. It further explains on the utilization of enzymes, specially methyltransferase and sulfotransferase to form functionally bioactive *O*-methylated and *O*-sulfated molecules.

Bibliography

BIBLIOGRAPHY

1 Austin, M. B. & Noel, J. P. The chalcone synthase superfamily of type III polyketide synthases. *Natural product reports* **20**, 79-110 (2003).

2 Moore, B. S. & Hopke, J. N. Discovery of a new bacterial polyketide biosynthetic pathway. *Chembiochem : a European journal of chemical biology* **2**, 35-38 (2001).

3 Li, Y. & Muller, R. Non-modular polyketide synthases in myxobacteria. *Phytochemistry* **70**, 1850-1857, doi:10.1016/j.phytochem.2009.05.003 (2009).

4 Funa, N. *et al.* A new pathway for polyketide synthesis in microorganisms. *Nature* **400**, 897-899, doi:10.1038/23748 (1999).

5 Gruschow, S., Buchholz, T. J., Seufert, W., Dordick, J. S. & Sherman, D. H. Substrate profile analysis and ACP-mediated acyl transfer in Streptomyces coelicolor Type III polyketide synthases. *Chembiochem : a European journal of chemical biology* **8**, 863-868, doi:10.1002/cbic.200700026 (2007).

6 Satou, R. *et al.* Structural basis for cyclization specificity of two Azotobacter type III polyketide synthases: a single amino acid substitution reverses their cyclization specificity. *The Journal of biological chemistry* **288**, 34146-34157, doi:10.1074/jbc.M113.487272 (2013).

7 Quadri, L. E. Biosynthesis of mycobacterial lipids by polyketide synthases and beyond. *Critical reviews in biochemistry and molecular biology* **49**, 179-211, doi:10.3109/10409238.2014.896859 (2014).

8 Chen, H., Tseng, C. C., Hubbard, B. K. & Walsh, C. T. Glycopeptide antibiotic biosynthesis: enzymatic assembly of the dedicated amino acid monomer (S)-3,5-dihydroxyphenylglycine. *Proceedings of the National Academy of Sciences of the United States of America* **98**, 14901-14906, doi:10.1073/pnas.221582098 (2001).

9 Pfeifer, V. *et al.* A polyketide synthase in glycopeptide biosynthesis: the biosynthesis of the non-proteinogenic amino acid (S)-3,5-dihydroxyphenylglycine. *The Journal of biological chemistry* **276**, 38370-38377, doi:10.1074/jbc.M106580200 (2001).

10 Risdian, C., Mozef, T. & Wink, J. Biosynthesis of Polyketides in Streptomyces. *Microorganisms* **7**, doi:10.3390/microorganisms7050124 (2019).

11 Anand, A. *et al.* Polyketide Quinones Are Alternate Intermediate Electron Carriers during Mycobacterial Respiration in Oxygen-Deficient Niches. *Molecular cell* **60**, 637-650, doi:10.1016/j.molcel.2015.10.016 (2015).

12 Funabashi, M., Funa, N. & Horinouchi, S. Phenolic lipids synthesized by type III polyketide synthase confer penicillin resistance on Streptomyces griseus. *The Journal of biological chemistry* **283**, 13983-13991, doi:10.1074/jbc.M710461200 (2008).

13 Kim, S. S. *et al.* LAP6/POLYKETIDE SYNTHASE A and LAP5/POLYKETIDE SYNTHASE B encode hydroxyalkyl alpha-pyrone synthases required for pollen development and sporopollenin biosynthesis in Arabidopsis thaliana. *The Plant cell* **22**, 4045-4066, doi:10.1105/tpc.110.080028 (2010).

14 Daffe, M. & Draper, P. The envelope layers of mycobacteria with reference to their pathogenicity. *Advances in microbial physiology* **39**, 131-203 (1998).

15 Brennan, P. J. & Nikaido, H. The envelope of mycobacteria. *Annual review of biochemistry* **64**, 29-63, doi:10.1146/annurev.bi.64.070195.000333 (1995).

16 Boissier, F. *et al.* Further insight into S-adenosylmethionine-dependent methyltransferases: structural characterization of Hma, an enzyme essential for the biosynthesis of oxygenated mycolic acids in Mycobacterium tuberculosis. *The Journal of biological chemistry* **281**, 4434-4445, doi:10.1074/jbc.M510250200 (2006).

17 Ortalo-Magne, A., Andersen, A. B. & Daffe, M. The outermost capsular arabinomannans and other mannoconjugates of virulent and avirulent tubercle bacilli. *Microbiology* **142 (Pt 4)**, 927-935, doi:10.1099/00221287-142-4-927 (1996).

18 Rokem, J. S., Lantz, A. E. & Nielsen, J. Systems biology of antibiotic production by microorganisms. *Natural product reports* **24**, 1262-1287, doi:10.1039/b617765b (2007).

19 Das, A. & Khosla, C. Biosynthesis of aromatic polyketides in bacteria. *Accounts of chemical research* **42**, 631-639, doi:10.1021/ar8002249 (2009).

20 Crawford, J. M. & Townsend, C. A. New insights into the formation of fungal aromatic polyketides. *Nat Rev Microbiol* **8**, 879-889, doi:10.1038/nrmicro2465 (2010).

21 Muggia, L. & Grube, M. Type III polyketide synthases in lichen mycobionts. *Fungal biology* **114**, 379-385, doi:10.1016/j.funbio.2010.03.001 (2010).

22 Noel, J. P., Austin, M. B. & Bomati, E. K. Structure-function relationships in plant phenylpropanoid biosynthesis. *Current opinion in plant biology* **8**, 249-253, doi:10.1016/j.pbi.2005.03.013 (2005).

23 Ghosh, R. *et al.* Dissecting the functional role of polyketide synthases in Dictyostelium discoideum: biosynthesis of the differentiation regulating factor 4-methyl-5-pentylbenzene-1,3-diol. *The Journal of biological chemistry* **283**, 11348-11354, doi:10.1074/jbc.M709588200 (2008).

24 Amnuaykanjanasin, A., Phonghanpot, S., Sengpanich, N., Cheevadhanarak, S. & Tanticharoen, M. Insect-specific polyketide synthases (PKSs), potential PKS-nonribosomal peptide synthetase hybrids, and novel PKS clades in tropical fungi. *Applied and environmental microbiology* **75**, 3721-3732, doi:10.1128/AEM.02744-08 (2009).

25 Lin, Z. *et al.* A bacterial source for mollusk pyrone polyketides. *Chemistry & biology* **20**, 73-81, doi:10.1016/j.chembiol.2012.10.019 (2013).

26 Blunt, J. W. *et al.* Marine natural products. *Natural product reports* **24**, 31-86, doi:10.1039/b603047p (2007).

27 Birch, A. & Donovan, F. Studies in relation to Biosynthesis. I. Some possible routes to derivatives of Orcinol and Phloroglucinol. *Australian Journal of Chemistry* **6**, 360-368, doi:https://doi.org/10.1071/CH9530360 (1953).

28 Collie, J. N. CLXXI.—Derivatives of the multiple keten group. *Journal of the Chemical Society, Transactions* **91**, 1806-1813, doi:10.1039/CT9079101806 (1907).

29 Hertweck, C. The biosynthetic logic of polyketide diversity. *Angewandte Chemie* **48**, 4688-4716, doi:10.1002/anie.200806121 (2009).

30 Dunn, B. J. & Khosla, C. Engineering the acyltransferase substrate specificity of assembly line polyketide synthases. *Journal of the Royal Society, Interface* **10**, 20130297, doi:10.1098/rsif.2013.0297 (2013).

31 Fischbach, M. A. & Walsh, C. T. Assembly-line enzymology for polyketide and nonribosomal Peptide antibiotics: logic, machinery, and mechanisms. *Chemical reviews* **106**, 3468-3496, doi:10.1021/cr0503097 (2006).

32 Kim, H. J. *et al.* A single module type I polyketide synthase directs de novo macrolactone biogenesis during galbonolide biosynthesis in Streptomyces galbus. *The Journal of biological chemistry* **289**, 34557-34568, doi:10.1074/jbc.M114.602334 (2014).

33 Abrar, M. *et al.* Aflatoxins: biosynthesis, occurrence, toxicity, and remedies. *Critical reviews in food science and nutrition* **53**, 862-874, doi:10.1080/10408398.2011.563154 (2013).

34 Murata, M. & Yasumoto, T. The structure elucidation and biological activities of high molecular weight algal toxins: maitotoxin, prymnesins and zooxanthellatoxins. *Natural product reports* **17**, 293-314, doi:10.1039/A901979K (2000).
35 Stinear, T. P. *et al.* Giant plasmid-encoded polyketide synthases produce the macrolide toxin of Mycobacterium ulcerans. *Proceedings of the National Academy of Sciences of the United States of America* **101**, 1345-1349, doi:10.1073/pnas.0305877101 (2004).
36 Lal, R. *et al.* Regulation and manipulation of the gene clusters encoding type-I PKSs. *Trends in biotechnology* **18**, 264-274 (2000).
37 Okamoto, S., Taguchi, T., Ochi, K. & Ichinose, K. Biosynthesis of actinorhodin and related antibiotics: discovery of alternative routes for quinone formation encoded in the act gene cluster. *Chemistry & biology* **16**, 226-236, doi:10.1016/j.chembiol.2009.01.015 (2009).
38 Hopwood, D. A. Genetic Contributions to Understanding Polyketide Synthases. *Chemical reviews* **97**, 2465-2498 (1997).
39 Gokhale, R. S. & Tuteja, D. Biochemistry of polyketide synthases. *Biotechnology: Special Processes* **10**, 341-372 (2001).
40 Khosla, C., Gokhale, R. S., Jacobsen, J. R. & Cane, D. E. Tolerance and specificity of polyketide synthases. *Annual review of biochemistry* **68**, 219-253, doi:10.1146/annurev.biochem.68.1.219 (1999).
41 Hutchinson, C. R. Microbial polyketide synthases: more and more prolific. *Proceedings of the National Academy of Sciences of the United States of America* **96**, 3336-3338 (1999).
42 Staunton, J. & Weissman, K. J. Polyketide biosynthesis: a millennium review. *Natural product reports* **18**, 380-416 (2001).
43 Shen, B. Polyketide biosynthesis beyond the type I, II and III polyketide synthase paradigms. *Current opinion in chemical biology* **7**, 285-295 (2003).
44 Rawlings, B. J. Type I polyketide biosynthesis in bacteria (part B). *Natural product reports* **18**, 231-281 (2001).
45 Rawlings, B. J. Type I polyketide biosynthesis in bacteria (Part A--erythromycin biosynthesis). *Natural product reports* **18**, 190-227 (2001).
46 Cox, R. J. Polyketides, proteins and genes in fungi: programmed nano-machines begin to reveal their secrets. *Organic & biomolecular chemistry* **5**, 2010-2026, doi:10.1039/b704420h (2007).
47 Hutchinson, C. R. *et al.* Aspects of the biosynthesis of non-aromatic fungal polyketides by iterative polyketide synthases. *Antonie van Leeuwenhoek* **78**, 287-295 (2000).
48 Hertweck, C., Luzhetskyy, A., Rebets, Y. & Bechthold, A. Type II polyketide synthases: gaining a deeper insight into enzymatic teamwork. *Natural product reports* **24**, 162-190, doi:10.1039/b507395m (2007).
49 Zhang, Z., Pan, H. X. & Tang, G. L. New insights into bacterial type II polyketide biosynthesis. *F1000Research* **6**, 172, doi:10.12688/f1000research.10466.1 (2017).
50 Chan, Y. A., Podevels, A. M., Kevany, B. M. & Thomas, M. G. Biosynthesis of polyketide synthase extender units. *Natural product reports* **26**, 90-114 (2009).
51 Winkel-Shirley, B. Flavonoid biosynthesis. A colorful model for genetics, biochemistry, cell biology, and biotechnology. *Plant physiology* **126**, 485-493 (2001).
52 Shi, S. P. *et al.* Enzymatic formation of unnatural novel chalcone, stilbene, and benzophenone scaffolds by plant type III polyketide synthase. *Organic letters* **11**, 551-554, doi:10.1021/ol802606w (2009).
53 Abe, I. *et al.* A Plant Type III Polyketide Synthase that Produces Pentaketide Chromone. *Journal of the American Chemical Society* **127**, 1362-1363, doi:10.1021/ja0431206 (2005).

54 Sankaranarayanan, R. *et al.* A novel tunnel in mycobacterial type III polyketide synthase reveals the structural basis for generating diverse metabolites. *Nature structural & molecular biology* **11**, 894-900, doi:10.1038/nsmb809 (2004).
55 Keating, T. A. & Walsh, C. T. Initiation, elongation, and termination strategies in polyketide and polypeptide antibiotic biosynthesis. *Current opinion in chemical biology* **3**, 598-606 (1999).
56 Ferrer, J. L., Jez, J. M., Bowman, M. E., Dixon, R. A. & Noel, J. P. Structure of chalcone synthase and the molecular basis of plant polyketide biosynthesis. *Nature structural biology* **6**, 775-784, doi:10.1038/11553 (1999).
57 Jez, J. M. *et al.* Structural control of polyketide formation in plant-specific polyketide synthases. *Chemistry & biology* **7**, 919-930 (2000).
58 Jez, J. M., Bowman, M. E. & Noel, J. P. Structure-guided programming of polyketide chain-length determination in chalcone synthase. *Biochemistry* **40**, 14829-14838 (2001).
59 Steele, C. L., Gijzen, M., Qutob, D. & Dixon, R. A. Molecular characterization of the enzyme catalyzing the aryl migration reaction of isoflavonoid biosynthesis in soybean. *Archives of biochemistry and biophysics* **367**, 146-150, doi:10.1006/abbi.1999.1238 (1999).
60 Dixon, R. A. Natural products and plant disease resistance. *Nature* **411**, 843-847, doi:10.1038/35081178 (2001).
61 Abe, I., Watanabe, T., Morita, H., Kohno, T. & Noguchi, H. Engineered biosynthesis of plant polyketides: manipulation of chalcone synthase. *Organic letters* **8**, 499-502, doi:10.1021/ol052912h (2006).
62 Austin, M. B., Bowman, M. E., Ferrer, J. L., Schroder, J. & Noel, J. P. An aldol switch discovered in stilbene synthases mediates cyclization specificity of type III polyketide synthases. *Chemistry & biology* **11**, 1179-1194, doi:10.1016/j.chembiol.2004.05.024 (2004).
63 Shomura, Y. *et al.* Crystal structure of stilbene synthase from Arachis hypogaea. *Proteins* **60**, 803-806, doi:10.1002/prot.20584 (2005).
64 Harris, T. M. & Carney, R. L. Biogenetically Modeled Synthesis of β-Resorcylic Acids[1]. *Journal of the American Chemical Society* **88**, 2053-2054, doi:10.1021/ja00961a042 (1966).
65 Eckermann, S. *et al.* New pathway to polyketides in plants. *Nature* **396**, 387, doi:10.1038/24652 (1998).
66 Morita, H. *et al.* Structural insight into chain-length control and product specificity of pentaketide chromone synthase from Aloe arborescens. *Chemistry & biology* **14**, 359-369, doi:10.1016/j.chembiol.2007.02.003 (2007).
67 Abe, I., Oguro, S., Utsumi, Y., Sano, Y. & Noguchi, H. Engineered biosynthesis of plant polyketides: chain length control in an octaketide-producing plant type III polyketide synthase. *J Am Chem Soc* **127**, 12709-12716, doi:10.1021/ja053945v (2005).
68 Austin, M. B. *et al.* Crystal structure of a bacterial type III polyketide synthase and enzymatic control of reactive polyketide intermediates. *The Journal of biological chemistry* **279**, 45162-45174, doi:10.1074/jbc.M406567200 (2004).
69 Abe, I., Watanabe, T. & Noguchi, H. Enzymatic formation of long-chain polyketide pyrones by plant type III polyketide synthases. *Phytochemistry* **65**, 2447-2453, doi:10.1016/j.phytochem.2004.08.005 (2004).
70 Funa, N., Ozawa, H., Hirata, A. & Horinouchi, S. Phenolic lipid synthesis by type III polyketide synthases is essential for cyst formation in Azotobacter vinelandii. *Proceedings of the National Academy of Sciences of the United States of America* **103**, 6356-6361, doi:10.1073/pnas.0511227103 (2006).
71 Abe, I. & Morita, H. Structure and function of the chalcone synthase superfamily of plant type III polyketide synthases. *Natural product reports* **27**, 809-838, doi:10.1039/b909988n (2010).

72 Morita, H. *et al.* A structure-based mechanism for benzalacetone synthase from Rheum palmatum. *Proceedings of the National Academy of Sciences of the United States of America* **107**, 669-673, doi:10.1073/pnas.0909982107 (2010).
73 Miyazono, K. *et al.* Crystal structure of curcuminoid synthase CUS from Oryza sativa. *Proteins* **79**, 669-673, doi:10.1002/prot.22888 (2011).
74 Katsuyama, Y., Kita, T., Funa, N. & Horinouchi, S. Curcuminoid biosynthesis by two type III polyketide synthases in the herb Curcuma longa. *The Journal of biological chemistry* **284**, 11160-11170, doi:10.1074/jbc.M900070200 (2009).
75 Taguchi, C. *et al.* Crystallization and preliminary X-ray diffraction studies of polyketide synthase-1 (PKS-1) from Cannabis sativa. *Acta crystallographica. Section F, Structural biology and crystallization communications* **64**, 217-220, doi:10.1107/S1744309108003795 (2008).
76 Taura, F. *et al.* Characterization of olivetol synthase, a polyketide synthase putatively involved in cannabinoid biosynthetic pathway. *FEBS letters* **583**, 2061-2066, doi:10.1016/j.febslet.2009.05.024 (2009).
77 Mizuuchi, Y., Shimokawa, Y., Wanibuchi, K., Noguchi, H. & Abe, I. Structure function analysis of novel type III polyketide synthases from Arabidopsis thaliana. *Biological & pharmaceutical bulletin* **31**, 2205-2210 (2008).
78 Saxena, P., Yadav, G., Mohanty, D. & Gokhale, R. S. A new family of type III polyketide synthases in Mycobacterium tuberculosis. *The Journal of biological chemistry* **278**, 44780-44790, doi:10.1074/jbc.M306714200 (2003).
79 Suh, D. Y. *et al.* Identification of amino acid residues important in the cyclization reactions of chalcone and stilbene synthases. *Biochem J* **350 Pt 1**, 229-235 (2000).
80 Funa, N., Awakawa, T. & Horinouchi, S. Pentaketide resorcylic acid synthesis by type III polyketide synthase from Neurospora crassa. *The Journal of biological chemistry* **282**, 14476-14481, doi:10.1074/jbc.M701239200 (2007).
81 Goyal, A. *et al.* Structural insights into biosynthesis of resorcinolic lipids by a type III polyketide synthase in Neurospora crassa. *J Struct Biol* **162**, 411-421, doi:10.1016/j.jsb.2008.02.009 (2008).
82 Gokulan, K. *et al.* Crystal structure of Mycobacterium tuberculosis polyketide synthase 11 (PKS11) reveals intermediates in the synthesis of methyl-branched alkylpyrones. *The Journal of biological chemistry* **288**, 16484-16494, doi:10.1074/jbc.M113.468892 (2013).
83 Stengel, D. B., Connan, S. & Popper, Z. A. Algal chemodiversity and bioactivity: sources of natural variability and implications for commercial application. *Biotechnology advances* **29**, 483-501, doi:10.1016/j.biotechadv.2011.05.016 (2011).
84 Sieburth, J. M. & Conover, J. T. Sargassum Tannin, an Antibiotic which Retards Fouling. *Nature* **208**, 52, doi:10.1038/208052a0 (1965).
85 Toth, G. B. & Pavia, H. Water-borne cues induce chemical defense in a marine alga (Ascophyllum nodosum). *Proceedings of the National Academy of Sciences of the United States of America* **97**, 14418-14420, doi:10.1073/pnas.250226997 (2000).
86 Amsler, C. D. & Fairhead, V. A. in *Advances in Botanical Research* Vol. 43 1-91 (Academic Press, 2005).
87 Meslet-Cladiere, L. *et al.* Structure/function analysis of a type iii polyketide synthase in the brown alga Ectocarpus siliculosus reveals a biochemical pathway in phlorotannin monomer biosynthesis. *The Plant cell* **25**, 3089-3103, doi:10.1105/tpc.113.111336 (2013).
88 Austin, M. B. *et al.* Biosynthesis of Dictyostelium discoideum differentiation-inducing factor by a hybrid type I fatty acid-type III polyketide synthase. *Nature chemical biology* **2**, 494-502, doi:10.1038/nchembio811 (2006).
89 Hou, L. *et al.* Generation of methylated violapyrones with improved anti-influenza A virus activity by heterologous expression of a type III PKS gene in a marine Streptomyces strain.

Bioorganic & medicinal chemistry letters **28**, 2865-2868, doi:10.1016/j.bmcl.2018.07.029 (2018).

90 Gomez-Escribano, J. P. & Bibb, M. J. Heterologous expression of natural product biosynthetic gene clusters in Streptomyces coelicolor: from genome mining to manipulation of biosynthetic pathways. *Journal of industrial microbiology & biotechnology* **41**, 425-431, doi:10.1007/s10295-013-1348-5 (2014).

91 Zhang, M. M., Wang, Y., Ang, E. L. & Zhao, H. Engineering microbial hosts for production of bacterial natural products. *Natural product reports* **33**, 963-987, doi:10.1039/c6np00017g (2016).

92 Bilyk, O. & Luzhetskyy, A. Metabolic engineering of natural product biosynthesis in actinobacteria. *Current opinion in biotechnology* **42**, 98-107, doi:10.1016/j.copbio.2016.03.008 (2016).

93 Knoten, C. A., Hudson, L. L., Coleman, J. P., Farrow, J. M., 3rd & Pesci, E. C. KynR, a Lrp/AsnC-type transcriptional regulator, directly controls the kynurenine pathway in Pseudomonas aeruginosa. *Journal of bacteriology* **193**, 6567-6575, doi:10.1128/JB.05803-11 (2011).

94 Zummo, F. P. *et al.* Tryptophan catabolism via kynurenine production in Streptomyces coelicolor: identification of three genes coding for the enzymes of tryptophan to anthranilate pathway. *Applied microbiology and biotechnology* **94**, 719-728, doi:10.1007/s00253-011-3833-y (2012).

95 Ohsawa, N., Tsujita, M., Morikawa, S. & Itoh, N. Purification and characterization of a monohalomethane-producing enzyme S-adenosyl-L-methionine: halide ion methyltransferase from a marine microalga, Pavlova pinguis. *Bioscience, biotechnology, and biochemistry* **65**, 2397-2404, doi:10.1271/bbb.65.2397 (2001).

96 Attieh, J. M., Hanson, A. D. & Saini, H. S. Purification and characterization of a novel methyltransferase responsible for biosynthesis of halomethanes and methanethiol in Brassica oleracea. *The Journal of biological chemistry* **270**, 9250-9257 (1995).

97 Fontecave, M., Atta, M. & Mulliez, E. S-adenosylmethionine: nothing goes to waste. *Trends in biochemical sciences* **29**, 243-249, doi:10.1016/j.tibs.2004.03.007 (2004).

98 Cantoni, G. L. Biological methylation: selected aspects. *Annual review of biochemistry* **44**, 435-451, doi:10.1146/annurev.bi.44.070175.002251 (1975).

99 Schubert, H. L., Blumenthal, R. M. & Cheng, X. Many paths to methyltransfer: a chronicle of convergence. *Trends in biochemical sciences* **28**, 329-335, doi:https://doi.org/10.1016/S0968-0004(03)00090-2 (2003).

100 Kozbial, P. Z. & Mushegian, A. R. Natural history of S-adenosylmethionine-binding proteins. *BMC structural biology* **5**, 19, doi:10.1186/1472-6807-5-19 (2005).

101 Kaminska, K. H. *et al.* Insights into the structure, function and evolution of the radical-SAM 23S rRNA methyltransferase Cfr that confers antibiotic resistance in bacteria. *Nucleic Acids Research* **38**, 1652-1663, doi:10.1093/nar/gkp1142 (2010).

102 Kimura, S. *et al.* Discovery of the beta-barrel-type RNA methyltransferase responsible for N6-methylation of N6-threonylcarbamoyladenosine in tRNAs. *Nucleic Acids Res* **42**, 9350-9365, doi:10.1093/nar/gku618 (2014).

103 Cheng, X., Kumar, S., Posfai, J., Pflugrath, J. W. & Roberts, R. J. Crystal structure of the HhaI DNA methyltransferase complexed with S-adenosyl-L-methionine. *Cell* **74**, 299-307 (1993).

104 Martin, J. L. & McMillan, F. M. SAM (dependent) I AM: the S-adenosylmethionine-dependent methyltransferase fold. *Current opinion in structural biology* **12**, 783-793 (2002).

105 Dixon, M. M., Huang, S., Matthews, R. G. & Ludwig, M. The structure of the C-terminal domain of methionine synthase: presenting S-adenosylmethionine for reductive methylation of B12. *Structure* **4**, 1263-1275 (1996).

106 Stephenson, R. C. & Clarke, S. Characterization of a rat liver protein carboxyl methyltransferase involved in the maturation of proteins with the -CXXX C-terminal sequence motif. *The Journal of biological chemistry* **267**, 13314-13319 (1992).
107 Gomez Maqueo Chew, A., Frigaard, N. U. & Bryant, D. A. Bacteriochlorophyllide c C-8(2) and C-12(1) methyltransferases are essential for adaptation to low light in Chlorobaculum tepidum. *Journal of bacteriology* **189**, 6176-6184, doi:10.1128/JB.00519-07 (2007).
108 Aktas, M., Gleichenhagen, J., Stoll, R. & Narberhaus, F. S-adenosylmethionine-binding properties of a bacterial phospholipid N-methyltransferase. *Journal of bacteriology* **193**, 3473-3481, doi:10.1128/JB.01539-10 (2011).
109 Goedecke, K., Pignot, M., Goody, R. S., Scheidig, A. J. & Weinhold, E. Structure of the N6-adenine DNA methyltransferase M.TaqI in complex with DNA and a cofactor analog. *Nature structural biology* **8**, 121-125, doi:10.1038/84104 (2001).
110 Gong, W., O'Gara, M., Blumenthal, R. M. & Cheng, X. Structure of pvu II DNA-(cytosine N4) methyltransferase, an example of domain permutation and protein fold assignment. *Nucleic Acids Res* **25**, 2702-2715 (1997).
111 Schubert, H. L., Phillips, J. D. & Hill, C. P. Structures along the catalytic pathway of PrmC/HemK, an N5-glutamine AdoMet-dependent methyltransferase. *Biochemistry* **42**, 5592-5599, doi:10.1021/bi034026p (2003).
112 Ansari, M. Z., Sharma, J., Gokhale, R. S. & Mohanty, D. In silico analysis of methyltransferase domains involved in biosynthesis of secondary metabolites. *BMC bioinformatics* **9**, 454, doi:10.1186/1471-2105-9-454 (2008).
113 Klimasauskas, S., Kumar, S., Roberts, R. J. & Cheng, X. HhaI methyltransferase flips its target base out of the DNA helix. *Cell* **76**, 357-369 (1994).
114 Kumar, S. *et al.* DNA containing 4'-thio-2'-deoxycytidine inhibits methylation by HhaI methyltransferase. *Nucleic Acids Res* **25**, 2773-2783 (1997).
115 Wu, J. C. & Santi, D. V. Kinetic and catalytic mechanism of HhaI methyltransferase. *The Journal of biological chemistry* **262**, 4778-4786 (1987).
116 Pavkov-Keller, T. *et al.* Crystal Structure and Catalytic Mechanism of CouO, a Versatile C-Methyltransferase from Streptomyces rishiriensis. *PloS one* **12**, e0171056, doi:10.1371/journal.pone.0171056 (2017).
117 Ibrahim, R. K., Bruneau, A. & Bantignies, B. Plant O-methyltransferases: molecular analysis, common signature and classification. *Plant Molecular Biology* **36**, 1-10, doi:10.1023/a:1005939803300 (1998).
118 Joshi, C. P. & Chiang, V. L. Conserved sequence motifs in plant S-adenosyl-L-methionine-dependent methyltransferases. *Plant Molecular Biology* **37**, 663-674, doi:10.1023/a:1006035210889 (1998).
119 Ferrer, J. L., Zubieta, C., Dixon, R. A. & Noel, J. P. Crystal structures of alfalfa caffeoyl coenzyme A 3-O-methyltransferase. *Plant physiology* **137**, 1009-1017, doi:10.1104/pp.104.048751 (2005).
120 Vidgren, J., Svensson, L. A. & Liljas, A. Crystal structure of catechol O-methyltransferase. *Nature* **368**, 354-358, doi:10.1038/368354a0 (1994).
121 Smith, C. D. *et al.* Crystal structure of human L-isoaspartyl-O-methyl-transferase with S-adenosyl homocysteine at 1.6-Å resolution and modeling of an isoaspartyl-containing peptide at the active site. *Protein Science* **11**, 625-635, doi:doi:10.1110/ps.37802 (2002).
122 Griffith, S. C. *et al.* Crystal structure of a protein repair methyltransferase from Pyrococcus furiosus with its l-isoaspartyl peptide substrate11Edited by I. A. Wilson. *Journal of Molecular Biology* **313**, 1103-1116, doi:https://doi.org/10.1006/jmbi.2001.5095 (2001).

123 Chatterjee, T., Mukherjee, D., Banerjee, M., Chatterjee, B. K. & Chakrabarti, P. Crystal structure and activity of protein L-isoaspartyl-O-methyltransferase from Vibrio cholerae, and the effect of AdoHcy binding. *Archives of biochemistry and biophysics* **583**, 140-149, doi:https://doi.org/10.1016/j.abb.2015.08.001 (2015).

124 Öster, L. M. *et al.* Insights into Cephamycin Biosynthesis: the Crystal Structure of CmcI from Streptomyces clavuligerus. *Journal of Molecular Biology* **358**, 546-558, doi:https://doi.org/10.1016/j.jmb.2006.02.004 (2006).

125 Ibdah, M., Zhang, X.-H., Schmidt, J. & Vogt, T. A Novel Mg2+-dependent O-Methyltransferase in the Phenylpropanoid Metabolism of Mesembryanthemum crystallinum. *Journal of Biological Chemistry* **278**, 43961-43972, doi:10.1074/jbc.M304932200 (2003).

126 Zheng, Y.-J. & Bruice, T. C. A Theoretical Examination of the Factors Controlling the Catalytic Efficiency of a Transmethylation Enzyme: Catechol O-Methyltransferase. *Journal of the American Chemical Society* **119**, 8137-8145, doi:10.1021/ja971019d (1997).

127 Bonifacio, M. J. *et al.* Kinetics and crystal structure of catechol-o-methyltransferase complex with co-substrate and a novel inhibitor with potential therapeutic application. *Molecular pharmacology* **62**, 795-805 (2002).

128 Noel, J. P., Dixon, R. A., Pichersky, E., Zubieta, C. & Ferrer, J.-L. in *Recent Advances in Phytochemistry* Vol. 37 (ed John T. Romeo) 37-58 (Elsevier, 2003).

129 Zubieta, C., He, X. Z., Dixon, R. A. & Noel, J. P. Structures of two natural product methyltransferases reveal the basis for substrate specificity in plant O-methyltransferases. *Nature structural biology* **8**, 271-279, doi:10.1038/85029 (2001).

130 Maxwell, C. A., Harrison, M. J. & Dixon, R. A. Molecular characterization and expression of alfalfa isoliquiritigenin 2'-O-methyltransferase, an enzyme specifically involved in the biosynthesis of an inducer of Rhizobium meliloti nodulation genes. *The Plant journal : for cell and molecular biology* **4**, 971-981 (1993).

131 Maxwell, C. A., Edwards, R. & Dixon, R. A. Identification, purification, and characterization of S-adenosyl-L-methionine: isoliquiritigenin 2'-O-methyltransferase from alfalfa (Medicago sativa L.). *Archives of biochemistry and biophysics* **293**, 158-166 (1992).

132 Zubieta, C., Kota, P., Ferrer, J. L., Dixon, R. A. & Noel, J. P. Structural basis for the modulation of lignin monomer methylation by caffeic acid/5-hydroxyferulic acid 3/5-O-methyltransferase. *The Plant cell* **14**, 1265-1277 (2002).

133 Liu, C. J. *et al.* Structural basis for dual functionality of isoflavonoid O-methyltransferases in the evolution of plant defense responses. *The Plant cell* **18**, 3656-3669, doi:10.1105/tpc.106.041376 (2006).

134 Cho, J. H., Park, Y., Ahn, J. H., Lim, Y. & Rhee, S. Structural and functional insights into O-methyltransferase from Bacillus cereus. *J Mol Biol* **382**, 987-997, doi:10.1016/j.jmb.2008.07.080 (2008).

135 Hou, X. *et al.* Crystal structure of SAM-dependent O-methyltransferase from pathogenic bacterium Leptospira interrogans. *Journal of Structural Biology* **159**, 523-528, doi:https://doi.org/10.1016/j.jsb.2007.04.007 (2007).

136 Varki, A. Biological roles of oligosaccharides: all of the theories are correct. *Glycobiology* **3**, 97-130, doi:10.1093/glycob/3.2.97 (1993).

137 Hooper, L. V., Manzella, S. M. & Baenziger, J. U. From legumes to leukocytes: biological roles for sulfated carbohydrates. *FASEB journal : official publication of the Federation of American Societies for Experimental Biology* **10**, 1137-1146 (1996).

138 Dodgson, K. S. & Tudball, N. The metabolic fate of the ester sulphate group of potassium p-nitrophenyl [35S]sulphate. *The Biochemical journal* **74**, 154-159 (1960).

139 Varin, L., Marsolais, F., Richard, M. & Rouleau, M. Sulfation and sulfotransferases 6: Biochemistry and molecular biology of plant sulfotransferases. *FASEB journal : official publication of the Federation of American Societies for Experimental Biology* **11**, 517-525 (1997).

140 Goren, M. B. Sulfolipid I of Mycobacterium tuberculosis, strain H37Rv. II. Structural studies. *Biochimica et biophysica acta* **210**, 127-138 (1970).

141 Robbins, P. W. & Lipmann, F. Identification of enzymatically active sulfate as adenosine-3'-phosphate-5'-phospho-sulfate1. *Journal of the American Chemical Society* **78**, 2652-2653, doi:10.1021/ja01592a097 (1956).

142 Chapman, E., Best, M. D., Hanson, S. R. & Wong, C. H. Sulfotransferases: structure, mechanism, biological activity, inhibition, and synthetic utility. *Angewandte Chemie* **43**, 3526-3548, doi:10.1002/anie.200300631 (2004).

143 Negishi, M. *et al.* Structure and function of sulfotransferases. *Archives of biochemistry and biophysics* **390**, 149-157, doi:10.1006/abbi.2001.2368 (2001).

144 Rath, V. L., Verdugo, D. & Hemmerich, S. Sulfotransferase structural biology and inhibitor discovery. *Drug discovery today* **9**, 1003-1011, doi:10.1016/S1359-6446(04)03273-8 (2004).

145 Mougous, J. D. *et al.* A sulfated metabolite produced by stf3 negatively regulates the virulence of Mycobacterium tuberculosis. *Proceedings of the National Academy of Sciences of the United States of America* **103**, 4258-4263, doi:10.1073/pnas.0510861103 (2006).

146 Malojcic, G. & Glockshuber, R. The PAPS-independent aryl sulfotransferase and the alternative disulfide bond formation system in pathogenic bacteria. *Antioxidants & redox signaling* **13**, 1247-1259, doi:10.1089/ars.2010.3119 (2010).

147 Kim, D.-H., Kim, H.-S. & Kobashi, K. Purification and Characterization of Novel Sulfotransferase Obtained from Klebsiella K-36, an Intestinal Bacterium of Rat1. *The Journal of Biochemistry* **112**, 456-460, doi:10.1093/oxfordjournals.jbchem.a123921 (1992).

148 Kim, D. H. & Kobashi, K. Kinetic studies on a novel sulfotransferase from Eubacterium A-44, a human intestinal bacterium. *Journal of biochemistry* **109**, 45-48 (1991).

149 Kwon, A. R., Yun, H. J. & Choi, E. C. Kinetic mechanism and identification of the active site tyrosine residue in Enterobacter amnigenus arylsulfate sulfotransferase. *Biochemical and biophysical research communications* **285**, 526-529, doi:10.1006/bbrc.2001.5185 (2001).

150 Lee, N. S., Kim, B. T., Kim, D. H. & Kobashi, K. Purification and reaction mechanism of arylsulfate sulfotransferase from Haemophilus K-12, a mouse intestinal bacterium. *Journal of biochemistry* **118**, 796-801 (1995).

151 Mathew, J. A., Tan, Y. P., Srinivasa Rao, P. S., Lim, T. M. & Leung, K. Y. Edwardsiella tarda mutants defective in siderophore production, motility, serum resistance and catalase activity. *Microbiology* **147**, 449-457, doi:10.1099/00221287-147-2-449 (2001).

152 Kang, J. W. *et al.* Cloning, sequence analysis, and characterization of the astA gene encoding an arylsulfate sulfotransferase from Citrobacter freundii. *Archives of pharmacal research* **24**, 316-322 (2001).

153 Kang, J.-W., Kwon, A.-R., Kim, D.-H. & CHOI, E.-C. Cloning and sequencing of the astA gene encoding arylsulfate sulfotransferase from Salmonella typhimurium. *Biological and Pharmaceutical Bulletin* **24**, 570-574 (2001).

154 Malojcic, G. *et al.* A structural and biochemical basis for PAPS-independent sulfuryl transfer by aryl sulfotransferase from uropathogenic Escherichia coli. *Proceedings of the National Academy of Sciences of the United States of America* **105**, 19217-19222, doi:10.1073/pnas.0806997105 (2008).

155 KIM, D.-H., YOON, H.-K., KOIZUMI, M. & KOBASHI, K. Sulfation of phenolic antibiotics by sulfotransferase obtained from a human intestinal bacterium. *Chemical and pharmaceutical bulletin* **40**, 1056-1057 (1992).
156 Kim, D. H., Hyun, S. H., Shim, S. B. & Kobashi, K. The role of intestinal bacteria in the transformation of sodium picosulfate. *Japanese journal of pharmacology* **59**, 1-5 (1992).
157 Kim, D. H., Konishi, L. & Kobashi, K. Purification, characterization and reaction mechanism of novel arylsulfotransferase obtained from an anaerobic bacterium of human intestine. *Biochimica et biophysica acta* **872**, 33-41 (1986).
158 Baek, M. C., Kim, S. K., Kim, D. H., Kim, B. K. & Choi, E. C. Cloning and sequencing of the Klebsiella K-36 astA gene, encoding an arylsulfate sulfotransferase. *Microbiology and immunology* **40**, 531-537 (1996).
159 Kim, D. H. & Kobashi, K. The role of intestinal flora in metabolism of phenolic sulfate esters. *Biochemical pharmacology* **35**, 3507-3510 (1986).
160 Bojarova, P. & Williams, S. J. Sulfotransferases, sulfatases and formylglycine-generating enzymes: a sulfation fascination. *Current opinion in chemical biology* **12**, 573-581, doi:10.1016/j.cbpa.2008.06.018 (2008).
161 Tang, X. *et al.* A two-step sulfation in antibiotic biosynthesis requires a type III polyketide synthase. *Nature chemical biology* **9**, 610-615, doi:10.1038/nchembio.1310 (2013).
162 Kaysser, L. *et al.* Identification and Manipulation of the Caprazamycin Gene Cluster Lead to New Simplified Liponucleoside Antibiotics and Give Insights into the Biosynthetic Pathway. *Journal of Biological Chemistry* **284**, 14987-14996, doi:10.1074/jbc.M901258200 (2009).
163 Jez, J. M., Ferrer, J. L., Bowman, M. E., Dixon, R. A. & Noel, J. P. Dissection of malonyl-coenzyme A decarboxylation from polyketide formation in the reaction mechanism of a plant polyketide synthase. *Biochemistry* **39**, 890-902 (2000).
164 Cole, S. T. *et al.* Deciphering the biology of Mycobacterium tuberculosis from the complete genome sequence. *Nature* **393**, 537-544, doi:10.1038/31159 (1998).
165 Bangera, M. G. & Thomashow, L. S. Identification and characterization of a gene cluster for synthesis of the polyketide antibiotic 2,4-diacetylphloroglucinol from Pseudomonas fluorescens Q2-87. *Journal of bacteriology* **181**, 3155-3163 (1999).
166 Cortes, J. *et al.* Identification and cloning of a type III polyketide synthase required for diffusible pigment biosynthesis in Saccharopolyspora erythraea. *Molecular microbiology* **44**, 1213-1224 (2002).
167 Gross, F. *et al.* Bacterial type III polyketide synthases: phylogenetic analysis and potential for the production of novel secondary metabolites by heterologous expression in pseudomonads. *Archives of microbiology* **185**, 28-38, doi:10.1007/s00203-005-0059-3 (2006).
168 Izumikawa, M. *et al.* Expression and characterization of the type III polyketide synthase 1,3,6,8-tetrahydroxynaphthalene synthase from Streptomyces coelicolor A3(2). *Journal of industrial microbiology & biotechnology* **30**, 510-515, doi:10.1007/s10295-003-0075-8 (2003).
169 Moore, B. S. *et al.* Plant-like biosynthetic pathways in bacteria: from benzoic acid to chalcone. *Journal of natural products* **65**, 1956-1962, doi:10.1021/np020230m (2002).
170 Gokhale, R. S., Saxena, P., Chopra, T. & Mohanty, D. Versatile polyketide enzymatic machinery for the biosynthesis of complex mycobacterial lipids. *Natural product reports* **24**, 267-277, doi:10.1039/b616817p (2007).
171 Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. Basic local alignment search tool. *Journal of molecular biology* **215**, 403-410 (1990).
172 Sambrook, J. & Russell, D. W. Preparation of Plasmid DNA by Alkaline Lysis with SDS: Minipreparation. *CSH protocols* **2006**, doi:10.1101/pdb.prot4084 (2006).

173 Smith, P. K. *et al.* Measurement of protein using bicinchoninic acid. *Analytical Biochemistry* **150**, 76-85, doi:https://doi.org/10.1016/0003-2697(85)90442-7 (1985).
174 Stinear, T. P. *et al.* Insights from the complete genome sequence of Mycobacterium marinum on the evolution of Mycobacterium tuberculosis. *Genome research* **18**, 729-741, doi:10.1101/gr.075069.107 (2008).
175 Weber, T. *et al.* antiSMASH 3.0-a comprehensive resource for the genome mining of biosynthetic gene clusters. *Nucleic acids research* **43**, W237-243, doi:10.1093/nar/gkv437 (2015).
176 Rivera-Marrero, C. A., Ritzenthaler, J. D., Newburn, S. A., Roman, J. & Cummings, R. D. Molecular cloning and expression of a novel glycolipid sulfotransferase in Mycobacterium tuberculosis. *Microbiology* **148**, 783-792, doi:10.1099/00221287-148-3-783 (2002).
177 Funa, N., Ohnishi, Y., Ebizuka, Y. & Horinouchi, S. Properties and substrate specificity of RppA, a chalcone synthase-related polyketide synthase in Streptomyces griseus. *The Journal of biological chemistry* **277**, 4628-4635, doi:10.1074/jbc.M110357200 (2002).
178 van Wageningen, A. M. *et al.* Sequencing and analysis of genes involved in the biosynthesis of a vancomycin group antibiotic. *Chemistry & biology* **5**, 155-162 (1998).
179 Pelzer, S. *et al.* Identification and analysis of the balhimycin biosynthetic gene cluster and its use for manipulating glycopeptide biosynthesis in Amycolatopsis mediterranei DSM5908. *Antimicrobial agents and chemotherapy* **43**, 1565-1573 (1999).
180 Pootoolal, J. *et al.* Assembling the glycopeptide antibiotic scaffold: The biosynthesis of A47934 from Streptomyces toyocaensis NRRL15009. *Proceedings of the National Academy of Sciences of the United States of America* **99**, 8962-8967, doi:10.1073/pnas.102285099 (2002).
181 Song, L. *et al.* Type III polyketide synthase beta-ketoacyl-ACP starter unit and ethylmalonyl-CoA extender unit selectivity discovered by Streptomyces coelicolor genome mining. *J Am Chem Soc* **128**, 14754-14755, doi:10.1021/ja065247w (2006).
182 Zha, W., Rubin-Pitel, S. B. & Zhao, H. Characterization of the substrate specificity of PhlD, a type III polyketide synthase from Pseudomonas fluorescens. *The Journal of biological chemistry* **281**, 32036-32047, doi:10.1074/jbc.M606500200 (2006).
183 Bu'Lock, J. D. & Hudson, A. T. beta-Leprosol: the identification of a trialkylresorcinol from bacterial lipids. *Journal of the Chemical Society. Perkin transactions 1* **1**, 61-63 (1969).
184 Mathieu, M. *et al.* The 2.8 A crystal structure of peroxisomal 3-ketoacyl-CoA thiolase of Saccharomyces cerevisiae: a five-layered alpha beta alpha beta alpha structure constructed from two core domains of identical topology. *Structure* **2**, 797-808 (1994).
185 Kabsch, W. Xds. *Acta crystallographica. Section D, Biological crystallography* **66**, 125-132, doi:10.1107/S0907444909047337 (2010).
186 Winn, M. D. *et al.* Overview of the CCP4 suite and current developments. *Acta crystallographica. Section D, Biological crystallography* **67**, 235-242, doi:10.1107/S0907444910045749 (2011).
187 Murshudov, G. N. *et al.* REFMAC5 for the refinement of macromolecular crystal structures. *Acta crystallographica. Section D, Biological crystallography* **67**, 355-367, doi:10.1107/S0907444911001314 (2011).
188 Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta crystallographica. Section D, Biological crystallography* **60**, 2126-2132, doi:10.1107/S0907444904019158 (2004).
189 Tickle, I. *et al.* (2016). (http://staraniso.globalphasing.org/cgi-bin/staraniso.cgi)
190 Bretschneider, T. *et al.* A ketosynthase homolog uses malonyl units to form esters in cervimycin biosynthesis. *Nature chemical biology* **8**, 154-161, doi:10.1038/nchembio.746 (2011).
191 Dundas, J. *et al.* CASTp: computed atlas of surface topography of proteins with structural and topographical mapping of functionally annotated residues. *Nucleic Acids Res* **34**, W116-118, doi:10.1093/nar/gkl282 (2006).

192 Pojer, F. *et al.* Structural basis for the design of potent and species-specific inhibitors of 3-hydroxy-3-methylglutaryl CoA synthases. *Proceedings of the National Academy of Sciences of the United States of America* **103**, 11491-11496, doi:10.1073/pnas.0604935103 (2006).
193 Shimizu, Y., Ogata, H. & Goto, S. Type III Polyketide Synthases: Functional Classification and Phylogenomics. *Chembiochem : a European journal of chemical biology* **18**, 50-65, doi:10.1002/cbic.201600522 (2017).
194 Akiyama, T., Shibuya, M., Liu, H. M. & Ebizuka, Y. p-Coumaroyltriacetic acid synthase, a new homologue of chalcone synthase, from Hydrangea macrophylla var. thunbergii. *European journal of biochemistry* **263**, 834-839 (1999).
195 Parvez, A. *et al.* Novel Type III Polyketide Synthases Biosynthesize Methylated Polyketides in Mycobacterium marinum. *Scientific reports* **8**, 6529, doi:10.1038/s41598-018-24980-1 (2018).
196 Xiong, Z. Q., Wang, J. F., Hao, Y. Y. & Wang, Y. Recent advances in the discovery and development of marine microbial natural products. *Marine drugs* **11**, 700-717, doi:10.3390/md11030700 (2013).
197 Shen, B. A New Golden Age of Natural Products Drug Discovery. *Cell* **163**, 1297-1300, doi:10.1016/j.cell.2015.11.031 (2015).
198 Davies, D. Understanding biofilm resistance to antibacterial agents. *Nature reviews. Drug discovery* **2**, 114-122, doi:10.1038/nrd1008 (2003).
199 Parsek, M. R. & Singh, P. K. Bacterial biofilms: an emerging link to disease pathogenesis. *Annual review of microbiology* **57**, 677-701, doi:10.1146/annurev.micro.57.030502.090720 (2003).
200 Hall-Stoodley, L. & Stoodley, P. Biofilm formation and dispersal and the transmission of human pathogens. *Trends in microbiology* **13**, 7-10, doi:10.1016/j.tim.2004.11.004 (2005).
201 Zhang, X., Bishop, P. L. & Kupferle, M. J. Measurement of polysaccharides and proteins in biofilm extracellular polymers. *Water Science and Technology* **37**, 345-348, doi:https://doi.org/10.1016/S0273-1223(98)00127-9 (1998).
202 Sutherland, I. W. The biofilm matrix - an immobilized but dynamic microbial environment. *Trends in microbiology* **9**, 222-227, doi:10.1016/S0966-842X(01)02012-1 (2001).
203 Recht, J. & Kolter, R. Glycopeptidolipid acetylation affects sliding motility and biofilm formation in Mycobacterium smegmatis. *Journal of bacteriology* **183**, 5718-5724, doi:10.1128/JB.183.19.5718-5724.2001 (2001).
204 Ojha, A. K. *et al.* Growth of Mycobacterium tuberculosis biofilms containing free mycolic acids and harbouring drug-tolerant bacteria. *Molecular microbiology* **69**, 164-174, doi:10.1111/j.1365-2958.2008.06274.x (2008).
205 Anand, A. *et al.* Polyketide quinones are alternate intermediate electron carriers during mycobacterial respiration in oxygen-deficient niches. *Molecular cell* **60**, 637-650 (2015).
206 O'Toole, G., Kaplan, H. B. & Kolter, R. Biofilm Formation as Microbial Development. *Annual review of microbiology* **54**, 49-79, doi:10.1146/annurev.micro.54.1.49 (2000).
207 Marsollier, L. *et al.* Impact of Mycobacterium ulcerans biofilm on transmissibility to ecological niches and Buruli ulcer pathogenesis. *PLoS pathogens* **3**, e62 (2007).
208 Ojha, A. K. *et al.* Growth of Mycobacterium tuberculosis biofilms containing free mycolic acids and harbouring drug-tolerant bacteria. *Molecular microbiology* **69**, 164-174 (2008).
209 Islam, M. S., Richards, J. P. & Ojha, A. K. Targeting drug tolerance in mycobacteria: a perspective from mycobacterial biofilms. *Expert review of anti-infective therapy* **10**, 1055-1066 (2012).
210 Elasri, M. O. & Miller, R. V. Study of the response of a biofilm bacterial community to UV radiation. *Applied and environmental microbiology* **65**, 2025-2031 (1999).

211 Brazil, G. M. *et al.* Construction of a rhizosphere pseudomonad with potential to degrade polychlorinated biphenyls and detection of bph gene expression in the rhizosphere. *Applied and environmental microbiology* **61**, 1946-1952 (1995).
212 Lewis, K. Persister cells, dormancy and infectious disease. *Nature reviews. Microbiology* **5**, 48-56, doi:10.1038/nrmicro1557 (2007).
213 Goodman, A. E., Marshall, K. C. & Hermansson, M. Gene transfer among bacteria under conditions of nutrient depletion in simulated and natural aquatic environments. *FEMS Microbiology Ecology* **15**, 55-60 (1994).
214 Annachhatre, A. P. Anaerobic treatment of industrial wastewaters. *Resources, Conservation and Recycling* **16**, 161-166, doi:https://doi.org/10.1016/0921-3449(95)00053-4 (1996).
215 Esteban, J. & Garcia-Coca, M. Mycobacterium Biofilms. *Frontiers in microbiology* **8**, 2651, doi:10.3389/fmicb.2017.02651 (2017).
216 Vijay, S. *et al.* Ultrastructural Analysis of Cell Envelope and Accumulation of Lipid Inclusions in Clinical Mycobacterium tuberculosis Isolates from Sputum, Oxidative Stress, and Iron Deficiency. *Frontiers in microbiology* **8**, 2681, doi:10.3389/fmicb.2017.02681 (2017).
217 Yamada, H. *et al.* Mycolicibacterium smegmatis, Basonym Mycobacterium smegmatis, Expresses Morphological Phenotypes Much More Similar to Escherichia coli Than Mycobacterium tuberculosis in Quantitative Structome Analysis and CryoTEM Examination. *Frontiers in microbiology* **9**, 1992, doi:10.3389/fmicb.2018.01992 (2018).
218 Williams, M. M. *et al.* Structural analysis of biofilm formation by rapidly and slowly growing nontuberculous mycobacteria. *Applied and environmental microbiology* **75**, 2091-2098, doi:10.1128/AEM.00166-09 (2009).
219 DePas, W. H., Bergkessel, M. & Newman, D. K. Aggregation of Nontuberculous Mycobacteria Is Regulated by Carbon-Nitrogen Balance. *mBio* **10**, e01715-01719, doi:10.1128/mBio.01715-19 (2019).
220 Nakano, C., Ozawa, H., Akanuma, G., Funa, N. & Horinouchi, S. Biosynthesis of aliphatic polyketides by type III polyketide synthase and methyltransferase in Bacillus subtilis. *Journal of bacteriology* **191**, 4916-4923, doi:10.1128/JB.00407-09 (2009).
221 Abdel-Ghany, S. E., Day, I., Heuberger, A. L., Broeckling, C. D. & Reddy, A. S. N. Production of Phloroglucinol, a Platform Chemical, in Arabidopsis using a Bacterial Gene. *Scientific reports* **6**, 38483-38483, doi:10.1038/srep38483 (2016).
222 Zhang, R., Cao, Y., Liu, W., Xian, M. & Liu, H. Improving phloroglucinol tolerance and production in Escherichia coli by GroESL overexpression. *Microbial cell factories* **16**, 227, doi:10.1186/s12934-017-0839-x (2017).
223 Das, S. & Rosazza, J. P. Microbial and enzymatic transformations of flavonoids. *Journal of natural products* **69**, 499-508, doi:10.1021/np0504659 (2006).
224 Bernini, R., Crisante, F. & Ginnasi, M. C. A convenient and safe O-methylation of flavonoids with dimethyl carbonate (DMC). *Molecules* **16**, 1418-1425, doi:10.3390/molecules16021418 (2011).
225 Pezet, R. *et al.* Delta-viniferin, a resveratrol dehydrodimer: one of the major stilbenes synthesized by stressed grapevine leaves. *Journal of agricultural and food chemistry* **51**, 5488-5492, doi:10.1021/jf030227o (2003).
226 Gibson, K. E., Kobayashi, H. & Walker, G. C. Molecular determinants of a symbiotic chronic infection. *Annual review of genetics* **42**, 413-441, doi:10.1146/annurev.genet.42.110807.091427 (2008).
227 Lloyd, A. L., Rasko, D. A. & Mobley, H. L. Defining genomic islands and uropathogen-specific genes in uropathogenic Escherichia coli. *Journal of bacteriology* **189**, 3532-3546, doi:10.1128/JB.01744-06 (2007).

228 Tsukamura, M. & Mizuno, S. [Subgrouping of strains of Mycobacterium avium-Mycobacterium intracellulare complex by thin-layer chromatography of acetone-soluble fraction after incubation with 35S-sulfate]. *Kekkaku : [Tuberculosis]* **55**, 481-484 (1980).

229 Goren, M. B., Brokl, O. & Schaefer, W. B. Lipids of putative relevance to virulence in Mycobacterium tuberculosis: correlation of virulence with elaboration of sulfatides and strongly acidic lipids. *Infection and immunity* **9**, 142-149 (1974).

230 Khoo, K. H. *et al.* Altered expression profile of the surface glycopeptidolipids in drug-resistant clinical isolates of Mycobacterium avium complex. *The Journal of biological chemistry* **274**, 9778-9785, doi:10.1074/jbc.274.14.9778 (1999).

231 Pethe, K. *et al.* The heparin-binding haemagglutinin of M. tuberculosis is required for extrapulmonary dissemination. *Nature* **412**, 190-194, doi:10.1038/35084083 (2001).

232 Hossain, M. M., Moriizumi, Y., Tanaka, S., Kimura, M. & Kakuta, Y. Crystal structure of sulfotransferase STF9 from Mycobacterium avium. *Molecular and cellular biochemistry* **361**, 97-104, doi:10.1007/s11010-011-1093-x (2012).

233 Waterhouse, A. M., Procter, J. B., Martin, D. M., Clamp, M. & Barton, G. J. Jalview Version 2--a multiple sequence alignment editor and analysis workbench. *Bioinformatics* **25**, 1189-1191, doi:10.1093/bioinformatics/btp033 (2009).

234 Wu, G., Robertson, D. H., Brooks, C. L., 3rd & Vieth, M. Detailed analysis of grid-based molecular docking: A case study of CDOCKER-A CHARMm-based MD docking algorithm. *Journal of computational chemistry* **24**, 1549-1562, doi:10.1002/jcc.10306 (2003).

235 Hall-Stoodley, L., Costerton, J. W. & Stoodley, P. Bacterial biofilms: from the natural environment to infectious diseases. *Nature reviews. Microbiology* **2**, 95-108, doi:10.1038/nrmicro821 (2004).

236 Chakraborty, P. & Kumar, A. The extracellular matrix of mycobacterial biofilms: could we shorten the treatment of mycobacterial infections? *Microbial cell* **6**, 105-122, doi:10.15698/mic2019.02.667 (2019).

237 Hui, Y., Yasuda, S., Liu, M. Y., Wu, Y. Y. & Liu, M. C. On the sulfation and methylation of catecholestrogens in human mammary epithelial cells and breast cancer cells. *Biological & pharmaceutical bulletin* **31**, 769-773, doi:10.1248/bpb.31.769 (2008).

238 Yasuda, S. *et al.* Concerted action of the cytosolic sulfotransferase, SULT1A3, and catechol-O-methyltransferase in the metabolism of dopamine in SK-N-MC human neuroblastoma cells. *Neuroscience research* **64**, 273-279, doi:10.1016/j.neures.2009.03.011 (2009).

239 Donovan, J. L., Kasim-Karakas, S., German, J. B. & Waterhouse, A. L. Urinary excretion of catechin metabolites by human subjects after red wine consumption. *The British journal of nutrition* **87**, 31-37, doi:10.1079/bjn2001482 (2002).

240 Del Rio, D. *et al.* Bioavailability and catabolism of green tea flavan-3-ols in humans. *Nutrition* **26**, 1110-1116, doi:10.1016/j.nut.2009.09.021 (2010).

241 Cook, I., Wang, T., Falany, C. N. & Leyh, T. S. High accuracy in silico sulfotransferase models. *The Journal of biological chemistry* **288**, 34494-34501, doi:10.1074/jbc.M113.510974 (2013).

242 Gidda, S. K. *et al.* Biochemical and molecular characterization of a hydroxyjasmonate sulfotransferase from Arabidopsis thaliana. *The Journal of biological chemistry* **278**, 17895-17900, doi:10.1074/jbc.M211943200 (2003).

243 Lerouge, P. *et al.* Symbiotic host-specificity of Rhizobium meliloti is determined by a sulphated and acylated glucosamine oligosaccharide signal. *Nature* **344**, 781-784, doi:10.1038/344781a0 (1990).

244 Baek, D. *et al.* A stress-inducible sulphotransferase sulphonates salicylic acid and confers pathogen resistance in Arabidopsis. *Plant, cell & environment* **33**, 1383-1392, doi:10.1111/j.1365-3040.2010.02156.x (2010).

245 Mueller, J. W., Gilligan, L. C., Idkowiak, J., Arlt, W. & Foster, P. A. The Regulation of Steroid Action by Sulfation and Desulfation. *Endocrine reviews* **36**, 526-563, doi:10.1210/er.2015-1036 (2015).

246 Velderrain-Rodríguez, G. R. *et al.* Phenolic compounds: their journey after intake. *Food & Function* **5**, 189-197, doi:10.1039/C3FO60361J (2014).

247 Ehrhardt, D. W. *et al.* In vitro sulfotransferase activity of NodH, a nodulation protein of Rhizobium meliloti required for host-specific nodulation. *Journal of bacteriology* **177**, 6237-6245, doi:10.1128/jb.177.21.6237-6245.1995 (1995).

248 Jendresen, C. B. & Nielsen, A. T. Production of zosteric acid and other sulfated phenolic biochemicals in microbial cell factories. *Nature communications* **10**, 4071, doi:10.1038/s41467-019-12022-x (2019).

249 Kim, B. G. *et al.* Regiospecific flavonoid 7-O-methylation with Streptomyces avermitilis O-methyltransferase expressed in Escherichia coli. *Journal of agricultural and food chemistry* **54**, 823-828, doi:10.1021/jf0522715 (2006).

250 Yoon, Y. *et al.* Characterization of O-methyltransferase ScOMT1 cloned from Streptomyces coelicolor A3(2). *Biochimica et biophysica acta* **1730**, 85-95, doi:10.1016/j.bbaexp.2005.06.005 (2005).

251 Li, W., Qiao, C., Pang, J., Zhang, G. & Luo, Y. The versatile O-methyltransferase LrOMT catalyzes multiple O-methylation reactions in amaryllidaceae alkaloids biosynthesis. *International journal of biological macromolecules* **141**, 680-692, doi:10.1016/j.ijbiomac.2019.09.011 (2019).

252 Paasela, T., Lim, K. J., Pietiainen, M. & Teeri, T. H. The O-methyltransferase PMT2 mediates methylation of pinosylvin in Scots pine. *The New phytologist* **214**, 1537-1550, doi:10.1111/nph.14480 (2017).

253 Shevchenko, A., Wilm, M., Vorm, O. & Mann, M. Mass spectrometric sequencing of proteins silver-stained polyacrylamide gels. *Analytical chemistry* **68**, 850-858 (1996).

Appendices

APPENDICES

Peptide analysis using mass spectrometry

MMAR_2190 and NFIII proteins were loaded on a 10 % SDS-PAGE and visualized by staining the gel with Coomassie Brilliant Blue R250 followed by destaining. In gel-digestion was done from the excised protein bands from polyacrylamide gels using trypsin[253].

Trypsin digestion involves repeated washing of the gel pieces using freshly prepared 100 µl of 50 % acetonitrile (ACN)/50 % 50 mM ammonium bicarbonate. It was incubated for 20 min at 37 ˚C till the removal of the stain resulting shrunken gel. The gel pieces were dried after the aspiration of acetonitrile using a speed vacuum. It is followed by swelling and undergoing reduction using freshly prepared 10 mM dithiotreitol/50 mM ammonium bicarbonate. The incubation was continued for 56 min at 45 ˚C and chilled to room temperature. Excess liquid was removed from the gel. The liquid was replaced by the same volume of freshly prepared 50 mM iodoacetamide (in 50mM ammonium bicarbonate) to undergo alkylation in the dark (for 30 min) at room temperature. Iodoactetamide solution was removed from the gel. The gel was rewashed using 100 µl of 50 % acetonitrile (ACN)/50 % 50 mM ammonium bicarbonate. The gel pieces were covered using 100 µl acetonitrile for 10 min, and acetonitrile was aspirated and dried again to remove water molecules. To the gel pieces, freshly prepared (diluted extraction buffer, 25mM ammonium bicarbonate solution, pH 8.5 to trypsin 1:20 ratio) lyophilized Promega trypsin (sequencing grade) was added to cover the gel in ice-cold condition. It was incubated at 37 ˚C for 30 min. The excess enzyme solution was removed, and the 2-3 µl extraction buffer

was added to keep the gel wet at 37 ˚C overnight. The peptides were extracted by centrifugation at 11200 rpm for 1 min, and the peptides in extraction buffer were transferred to fresh 0.5 ml PCR tube, cleaned with 75 µl of 50 % ACN/ 5 % formic acid solution. The gel pieces were covered with 75 µl of 50 % ACN/ 5 % formic acid solution for 15 min and centrifuged again, and the supernatant was combined with peptides in earlier extraction. The step was repeated, and the final extracted peptide sample was passed through µ-C18 Zip tip equilibrated 3X with acetonitrile and 0.1 % formic acid for cleaning up peptide sample. 4 µl aliquots of elution buffer (60 % acetonitrile/0.1 % formic acid) were passed through the tip loaded with the sample. The peptide was eluted down in the elution buffer and was dried completely using speed-vac and stored at -80 ˚C till further analysis. The sample was reconstituted in 10 µl of 5 % ACN/ 5 % formic acid solution. The resulting peptides from the sample were spotted on the MALDI target plate to obtain peptide spectra, and the identification of protein was carried out by b- and y ions using ABI SCIEX MALDI-TOF/TOF 5800 and further analyzed by using SCIEX PeakView 2.1 software. The peptides were blasted against proteins using the National Centre for Biotechnology Information (NCBI) to identify the protein.

Peptide analysis of type III polyketide synthase by mass spectrometry

In-gel trypsin digestion of purified type III PKS, MMAR_2190 and NFIII protein and peptides spotted on MALDI-TOF-MS with tryptic peptide spectra were analyzed using database searching of the b- and y- ions through the use of Mascot search. Furthermore, any modification in the active site residue cysteine was analyzed using SCIEX PeakView 2.1 software. The peptide analysis showed protein belong to type III PKS, and it is homologs of PKS18, an *M. tuberculosis* type III PKS. The bold peptide of the active site of MMAR_2190 was found to have m/z of peptide 1666.8226.

```
  1  MSTAAEGGAI RRAGHEPRYD LAQLPPAPPT TVAVIEGMAT GAPQRVVAQA DAAARVSELF
 61  VDPQQRERIS RIYDKTRIDT RRMAVDPLDD EFDEFRREPA TIRDRMNLFY QHAVPLAVDV
121 AARALDGLPY APDEIGQLVF VTSTGFIAPG VDVEIVKQLG LPRSISRVVV NFMGCAAAMN
181 AIRTATNYVR AHPSMKALVV CIELSSVNAV FADDINDVVI HSLFGDGCGA LVIGASQVQQ
241 PLPAGNVVIR SSFSQLLDDS EDGIVLGVNH DGITCELSEN LPSYIYRSVD PVVAEMLRDN
301 GLSKADIDLW AIHPGGPKII EQSARSLGIP VGRAVQSWDV LAQFGNMLSV SLIFVLEMMV
AQAESDKPIS TGVAFAFAPG VTVEGMLFDIIRR
```

The peptide masses from your sequence are:

[Theoretical pI: 4.85 / Mw (average mass): 42335.47 / Mw (monoisotopic mass): 42308.71]

mass	position	#MC	modifications	peptide sequence
6346.2349	334-392	0		AVQSWDVLAQFGNMLSVSLI FVLEMMVAQAESDKPISTGV AFAFAPGVTVEGMLFDIIR
5502.8762	197-250	0		ALVVCIELSSVNAVFADDIN DVVIHSLFGDGCGALVIGAS QVQQPLPAGNVVIR
4085.9284	251-287	0		SSFSQLLDDSEDGIVLGVNH DGITCELSENLPSYIYR
3530.8668	124-157	0		ALDGLPYAPDEIGQLVFVTS TGFIAPGVDVEIVK
2764.4338	19-45	0		YDLAQLPPAPPTTVAVIEGM ATGAPQR
2015.0531	106-123	0		MNLFYQHAVPLAVDVAAR
1698.7315	83-96	0		MAVDPLDDEFDEFR
1666.8226	168-183	0		VVVNFMGCAAAMNAIR
1489.7797	305-318	0		ADIDLWAIHPGGPK
1317.6797	56-66	0		VSELFVDPQQR
1215.6401	288-298	0		SVDPVVAEMLR
1063.5200	1-11	0		MSTAAEGGAIR
971.5268	46-55	0		VVAQADAAAR
824.4261	184-190	0		TATNYVR
816.4574	319-325	0		IIEQSAR
798.4832	326-333	0		SLGIPVGR
686.3831	98-103	0		EPATIR
683.4199	158-163	0		QLGLPR
670.3341	191-196	0		AHPSMK
666.3318	13-18	0		AGHEPR
633.3202	299-304	0		DNGLSK

Appendix A2.1.1: Predicted peptide mass from *in-gel* trypinization of MMAR_2190 using PeptideMass from ExPASy

The Nf protein from *Nocardia farcinica* 10152 showed 68 % identity and 98 % query cover to MMAR_2190. The signature peptide from the active site cysteine as shown below. The peptide was found to have m/z of 1706.7404 conformed by the b- and y-ions analysis, as shown in **appendix figure 2.**

```
  1  MSITVDEGGA RPATEPRQRI HPDLGHAHTP MPPAPPVTIG VVEGIATGSP AQIVDQAEAA
 61  ERVAALFTDP AQRARIARVY EKTRIETRRM AVDPTAPEFR SFSRQPGTLR ERMNLFYRHA
121  APLAVDVAGR ALADSGAAAA DIGLLVFVTS TGFIAPGVDV AVLRELGLVP TVGRVVVNFM
181  GCAAAMNGIR TGVDYVRAHP DKKALVVCLE LSSVNAVFAD DVNDVIIHSL FGDGCGAVVL
241  GASQARQPLA PGRVVVRDSF SHLFDAAEDG IVLGVDHNGI TCELSENLPR YIYRGVDPVV
301  REVLARNGLR KSDIDLWAIH PGGPKIIEES VRSLGLSPEQ AAPSWDVLAR HGNMLSVSLI
361 FVLEQMIAQS ATAEPISTGV AFSFAPGVTL EGLIFDIIRR
```

The peptide masses from your sequence are:

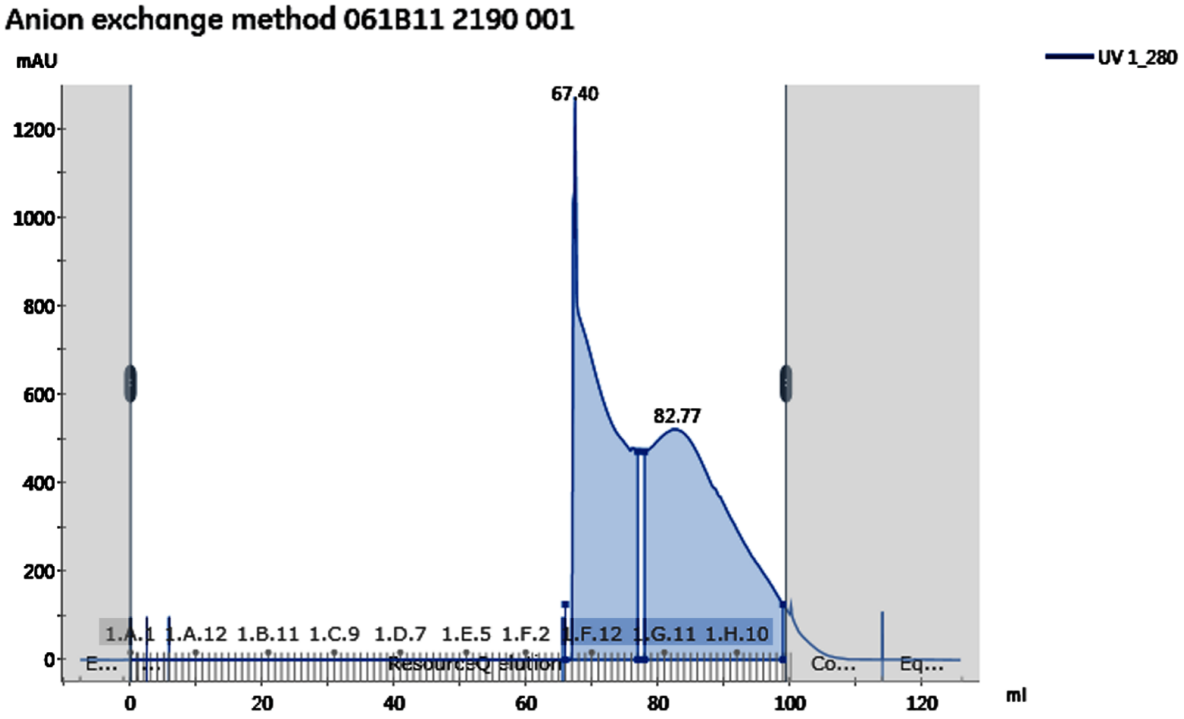
[Theoretical pI: 5.72 / Mw (average mass): 42588.82 / Mw (monoisotopic mass): 42562.20]

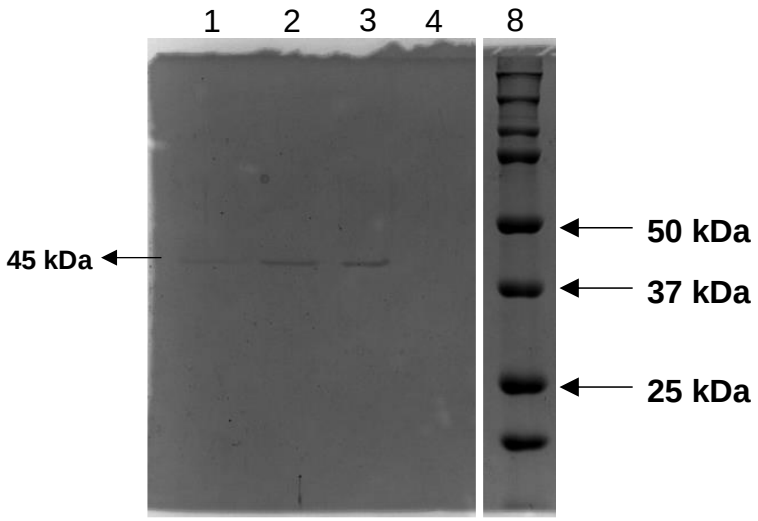
mass	position	#MC	modifications	peptide sequence
5190.7257	351-399	0		HGNMLSVSLIFVLEQMIAQS ATAEPISTGVAFSFAPGVTL EGLIFDIIR
4347.2189	20-62	0		IHPDLGHAHTPMPPAPPVTI GVVEGIATGSPAQIVDQAEA AER
4344.2002	204-246	0		ALVVCLELSSVNAVFADDVN DVIIHSLFGDGCGAVVLGAS QAR
3570.6805	258-290	0		DSFSHLFDAAEDGIVLGVDH NGITCELSENLPR
3257.7779	131-164	0		ALADSGAAAADIGLLVFVTS TGFIAPGVDVAVLR
1896.9813	333-350	0		SLGLSPEQAAPSWDVLAR
1786.8752	1-17	0		MSITVDEGGARPATEPR
1652.8069	175-190	0		VVVNFMGCAAAMNGIR
1505.7746	312-325	0		SDIDLWAIHPGGPK
1233.5932	90-100	0		MAVDPTAPEFR
1188.6371	63-73	0		VAALFTDPAQR
1176.6483	119-130	0		HAAPLAVDVAGR
1040.6098	165-174	0		ELGLVPTVGR
845.4727	326-332	0		IIEESVR
843.4182	113-118	0		MNLFYR
809.4152	191-197	0		TGVDYVR
741.4253	295-301	0		GVDPVVR
738.4257	247-253	0		QPLAPGR
671.3835	105-110	0		QPGTLR
614.3297	291-294	0		YIYR
587.3511	302-306	0		EVLAR
567.2885	198-202	0		AHPDK
538.2871	79-82	0		VYEK
518.2933	85-88	0		IETR

```
Query  4    TVDEGGA--RPATEPRQRIHPDLGHAHTPMPPAPPVTIGVVEGIATGSPAQIVDQAEAAE
61

            T  EGGA  R   EPR     DL      +PPAPP T+ V+EG+ATG+P ++V QA+AA

Sbjct  3    TAAEGGAIRRAGHEPRY----DLAQ----LPPAPPTTVAVIEGMATGAPQRVVAQADAAA
54


Query  62   RVAALFTDPAQRARIARVYEKTRIETRRMAVDPTAPEFRSFSRQPGTLRERMNLFYRHAA
121

            RV+ LF DP QR RI+R+Y+KTRI+TRRMAVDP   EF  F R+P T+R+RMNLFY+HA

Sbjct  55   RVSELFVDPQQRERISRIYDKTRIDTRRMAVDPLDDEFDEFRREPATIRDRMNLFYQHAV
114


Query  122  PLAVDVAGRALADSGAAAADIGLLVFVTSTGFIAPGVDVAVLRELGLVPTVGRVVVNFMG
181

            PLAVDVA RAL     A  +IG LVFVTSTGFIAPGVDV ++++LGL  ++ RVVVNFMG

Sbjct  115  PLAVDVAARALDGLPYAPDEIGQLVFVTSTGFIAPGVDVEIVKQLGLPRSISR VVVNFMG
174


Query  182  CAAAMNGIRTGVDYVRAHPDKKALVVCLELSSVNAVFADDVNDVIIHSLFGDGCGAVVLG
241

            CAAAMN IRT  +YVRAHP  KALVVC+ELSSVNAVFADD+NDV+IHSLFGDGCGA+V+G

Sbjct  175  CAAAMNAIRTATNYVRAHPSMKALVVCIELSSVNAVFADDINDVVIHSLFGDGCGALVIG
234


Query  242  ASQARQPLAPGRVVVRDSFSHLFDAAEDGIVLGVDHNGITCELSENLPRYIYRGVDPVVR
301

            ASQ +QPL  G VV+R SFS L D +EDGIVLGV+H+GITCELSENLP YIYR VDPVV
```

```
Sbjct  235  ASQVQQPLPAGNVVIRSSFSQLLDDSEDGIVLGVNHDGITCELSENLPSYIYRSVDPVVA
294

Query  302  EVLARNGLRKSDIDLWAIHPGGPKIIEESVRSLGLSPEQAAPSWDVLARHGNMLSVSLIF
361
            E+L  NGL K+DIDLWAIHPGGPKIIE+S RSLG+   +A  SWDVLA+ GNMLSVSLIF
Sbjct  295  EMLRDNGLSKADIDLWAIHPGGPKIIEQSARSLGIPVGRAVQSWDVLAQFGNMLSVSLIF
354

Query  362  VLEQMIAQSATAEPISTGVAFSFAPGVTLEGLIFDIIRR  400
            VLE M+AQ+ + +PISTGVAF+FAPGVT+EG++FDIIRR
Sbjct  355  VLEMMVAQAESDKPISTGVAFAFAPGVTVEGMLFDIIRR  393
```

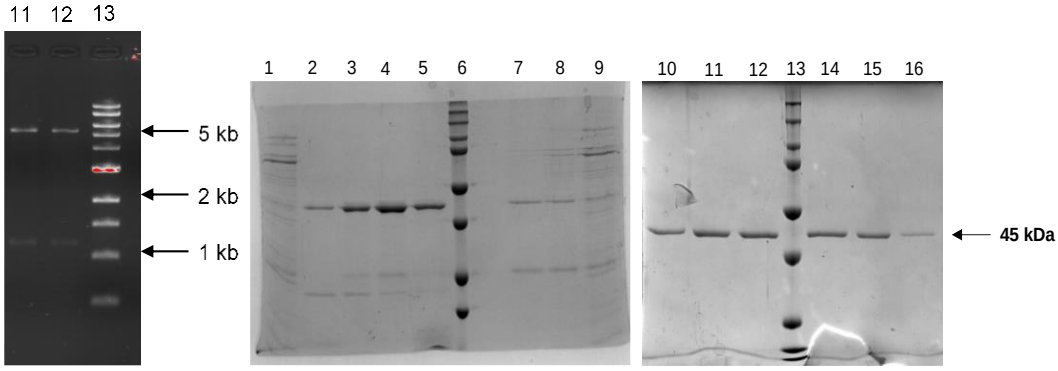
Appendix A2.1.2: Predicted peptide masses from *in-gel* trypinization of Nf using PeptideMass from ExPASy and BlastP of Nf (Query) with MMAR_2190 (Subject)





1, 2, 3, 4	MMAR_2190
8.	Protein ladder

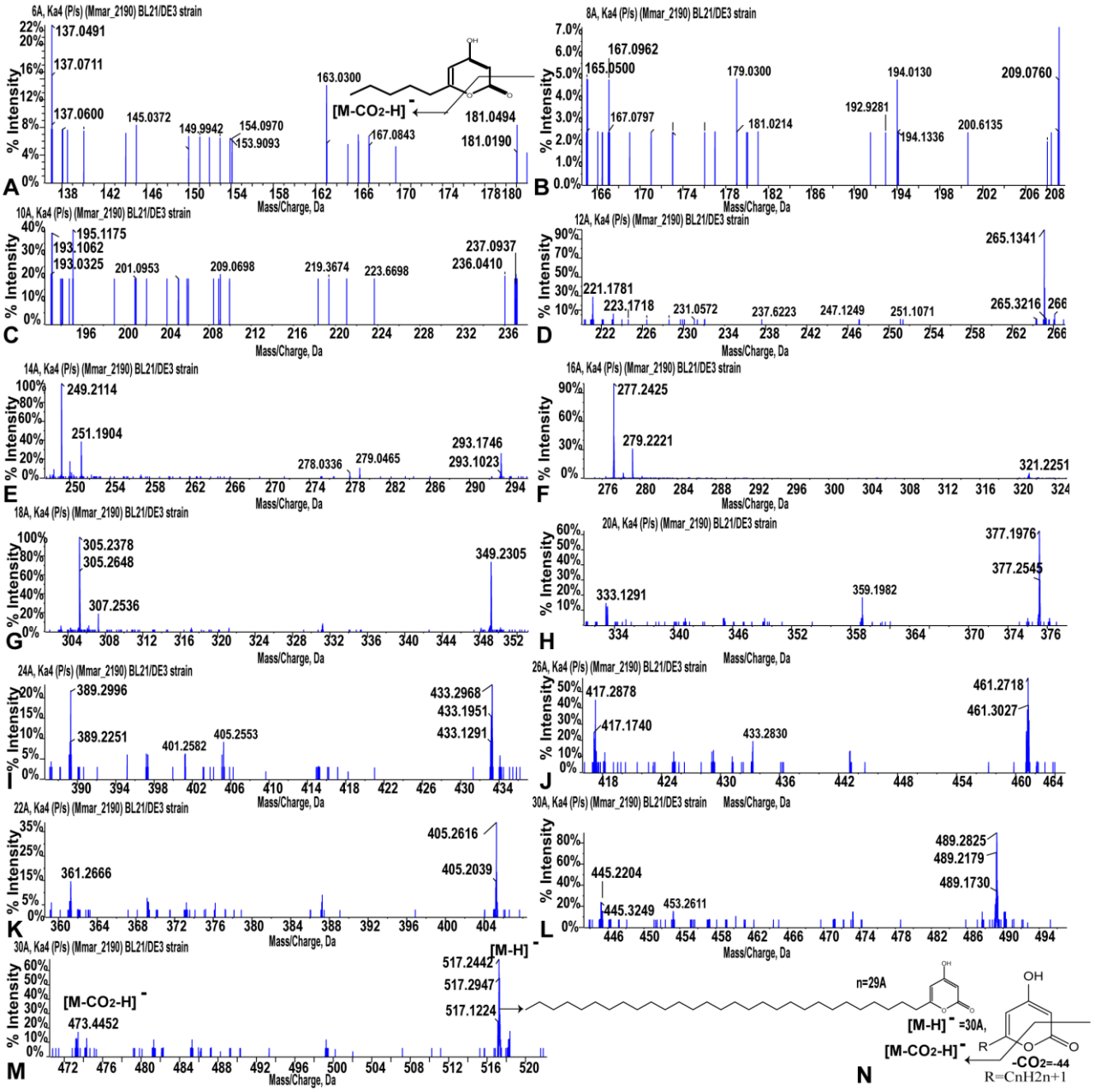
Appendix A2.2.1: SDS-PAGE profile of overexpressed MMAR_2190 by Anion exchange chromatography



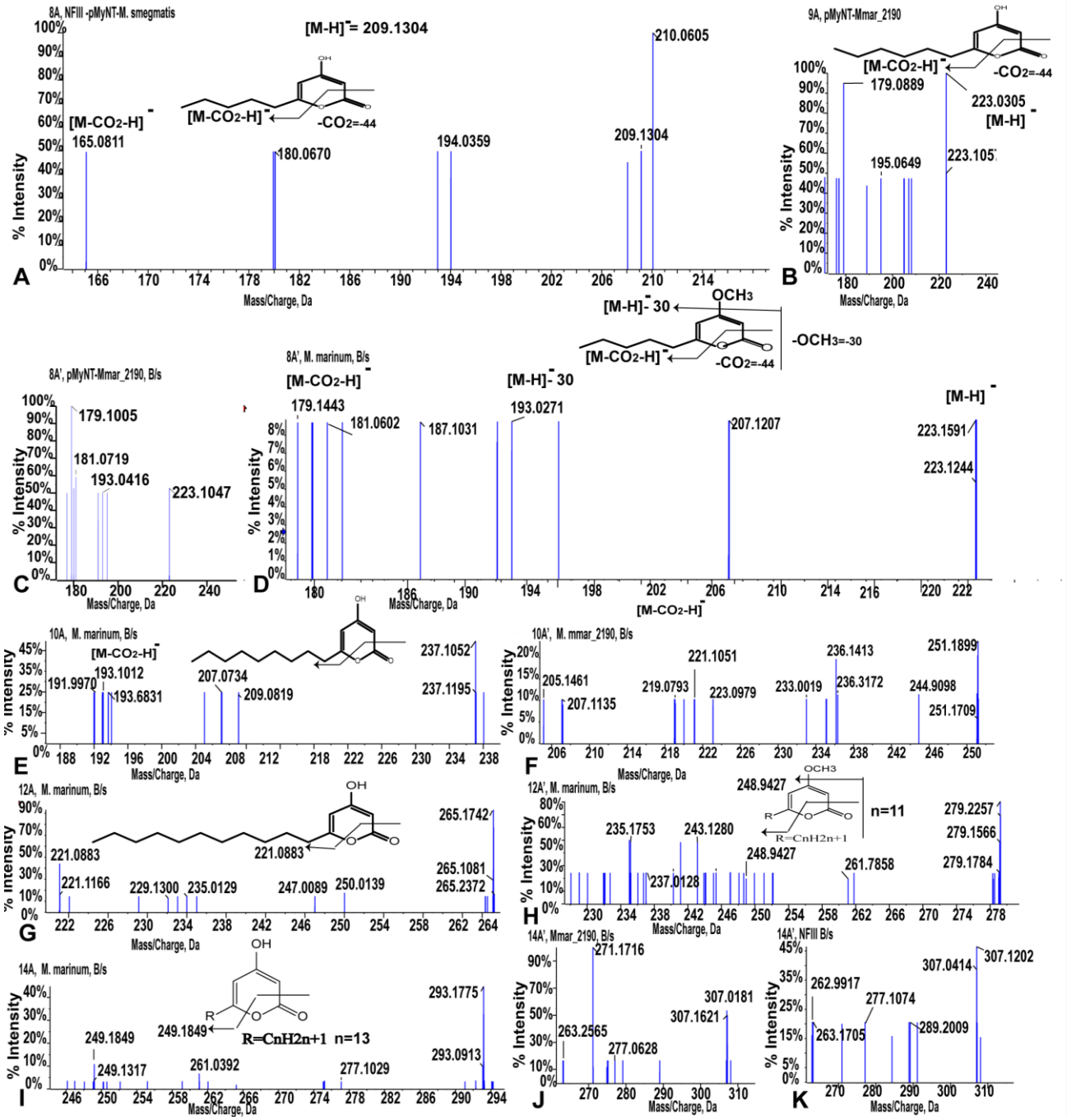
Appendix A2.2.2: Screening NFIII plasmid using (i) restriction digestion and (ii) protein SDS-PAGE profile for overexpressed NFIII protein purified through Ni^{2+}-NTA column chromatography

6, 13.	Protein ladder
9.	Crude supernatant
1.	Pellet
7, 8.	5mM elution
2, 16.	10mM elution
3, 4, 14, 15	10mM elution
5, 11, 12.	25mM elution
10.	50mM elution

Appendix A2.2.2: Screening NFIII plasmid for overexpression

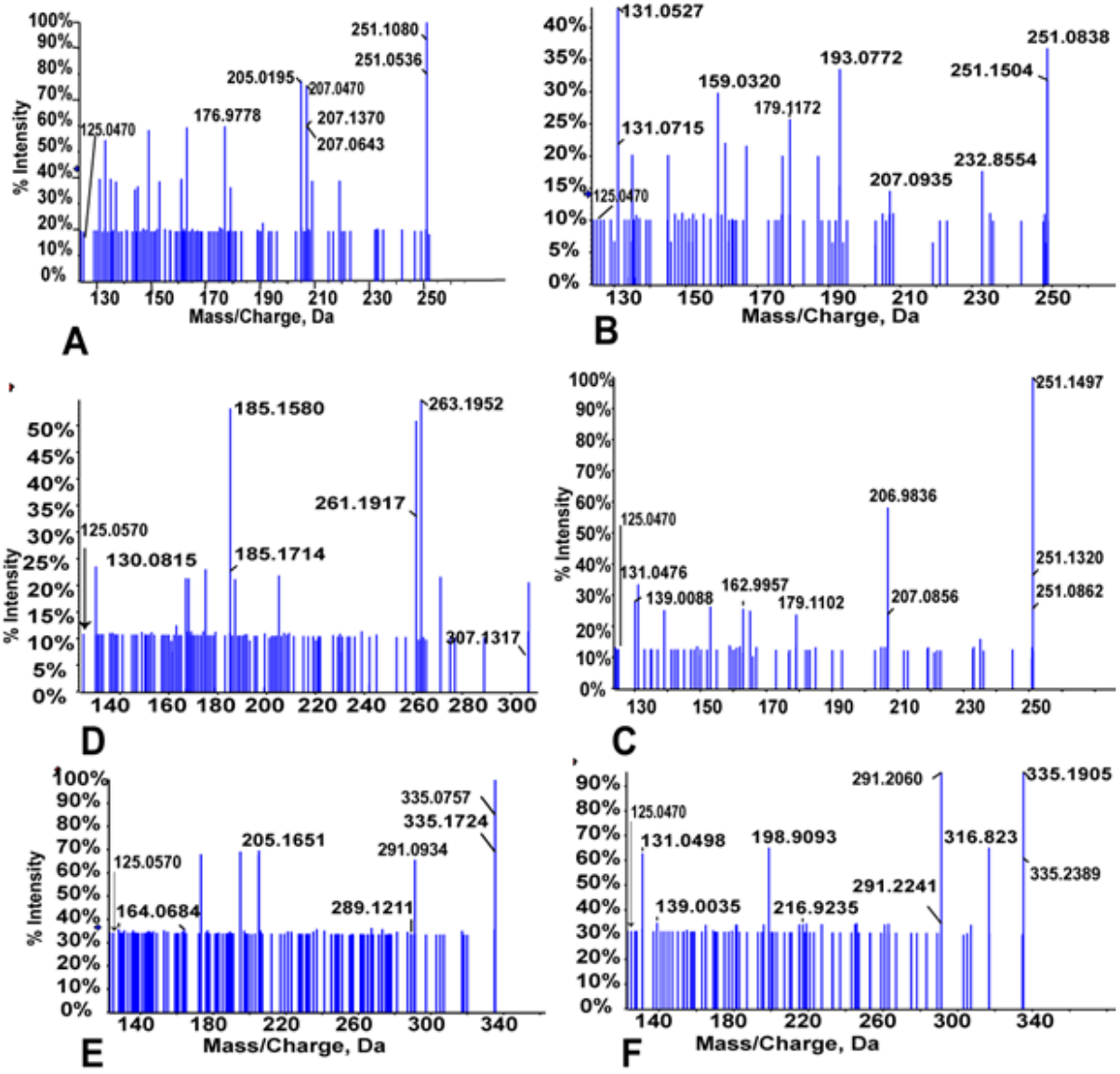


Appendix A4.1: Metabolomics analysis of triketide α-pyrone from the *MMAR_2190* gene overexpressed in BL21(DE3)*E.coli* strain. A to M show the fragmentation of precursor ion derived from C_6 to C_{30} carbon chain with a mass of precursor $[M-H]^-$ to give fragments of $[M-CO_2-H]^-$ with loss of -44 mass unit from $[M-H]^{-\cdot}$. The fragmentation pattern is given in N, showing a general strategy for all the C-chain α-pyrone. An arrow is shown from 30A derived triketide α-pyrone of the mass of precursor m/z 517.2442 to give sub-fragments of $[M-CO_2-H]^-$ with m/z 473.4452

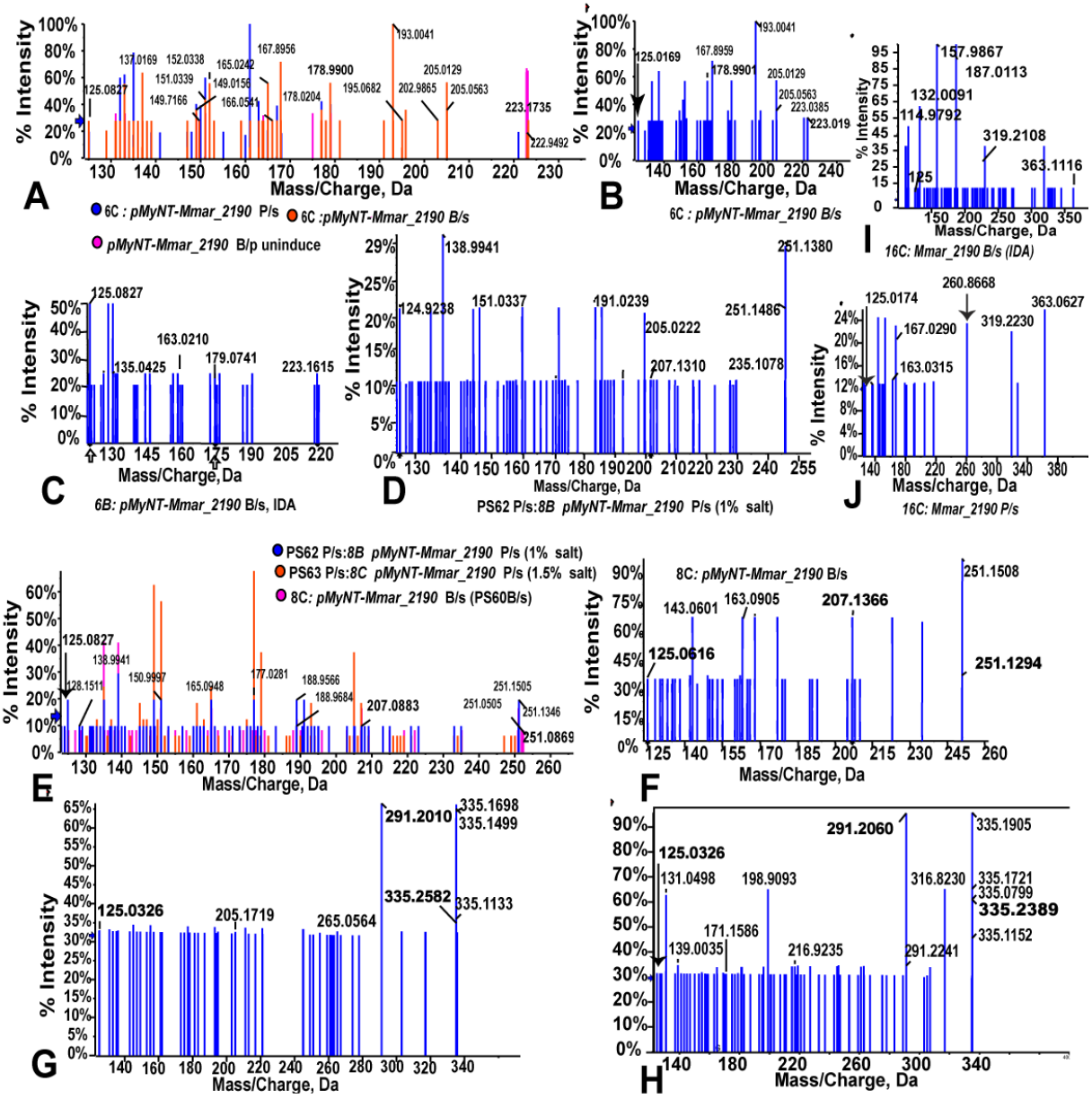


Appendix A4.2: Metabolomics analysis of triketide *α*-pyrones (8A to 14A) from wildtype *M. marinum* and different type III PKS homologs of MMAR_2190 overexpressed in pMyNT based overexpressed knockout (strain devoid type IIII PKS, Δ pks10). A, B, E, G and I show the fragmentation of 8A, 9A, 10A, 12A, and 14A triketide *α*-pyrone from *pMyNT-NFIII* strain, *pMyNT-MMAR_2190* and *M. marinum* strains with the fragmentation of precursor ion derived from C_8 to C_{14} carbon chain with mass of precursor $[M-H]^-$ to give fragments of $[M-CO_2-H]^-$ with loss of -44 mass unit from $[M-H]^{-\cdot}$. The strategy of fragmentation pattern is shown above the mass spectrometry data. C, D, F, H, J and K show the fragmentation of O-methylated precursor ion of 8A' (m/z 223), 10A' (m/z at 251), 12A' (m/z at 279) and 14A' (m/z, 307) to give fragments of $[M-CO_2-H]^-$ with loss of -44 mass unit from $[M-H]^-$ and $[M-H]^-$-30 as unique fragments. Although short and medium-chain triketide *α*-pyrone fragmentation pattern was found in biofilm culture of *pMyNT-MMAR_2190* overexpressed strain and *M. marinum* wild-type strain but a close homologs *pMyNT-NFIII* also show the *O*-methylated products in biofilm culture study. The presence of odd chain triketide *α*-pyrone like the one shown in B with

the mass of precursor with $[M-H]^-$ at m/z 223 gives fragments $[M-CO_2-H]^-$ with loss of -44 mass unit similar to that observed in even chain triketide α-pyrone.

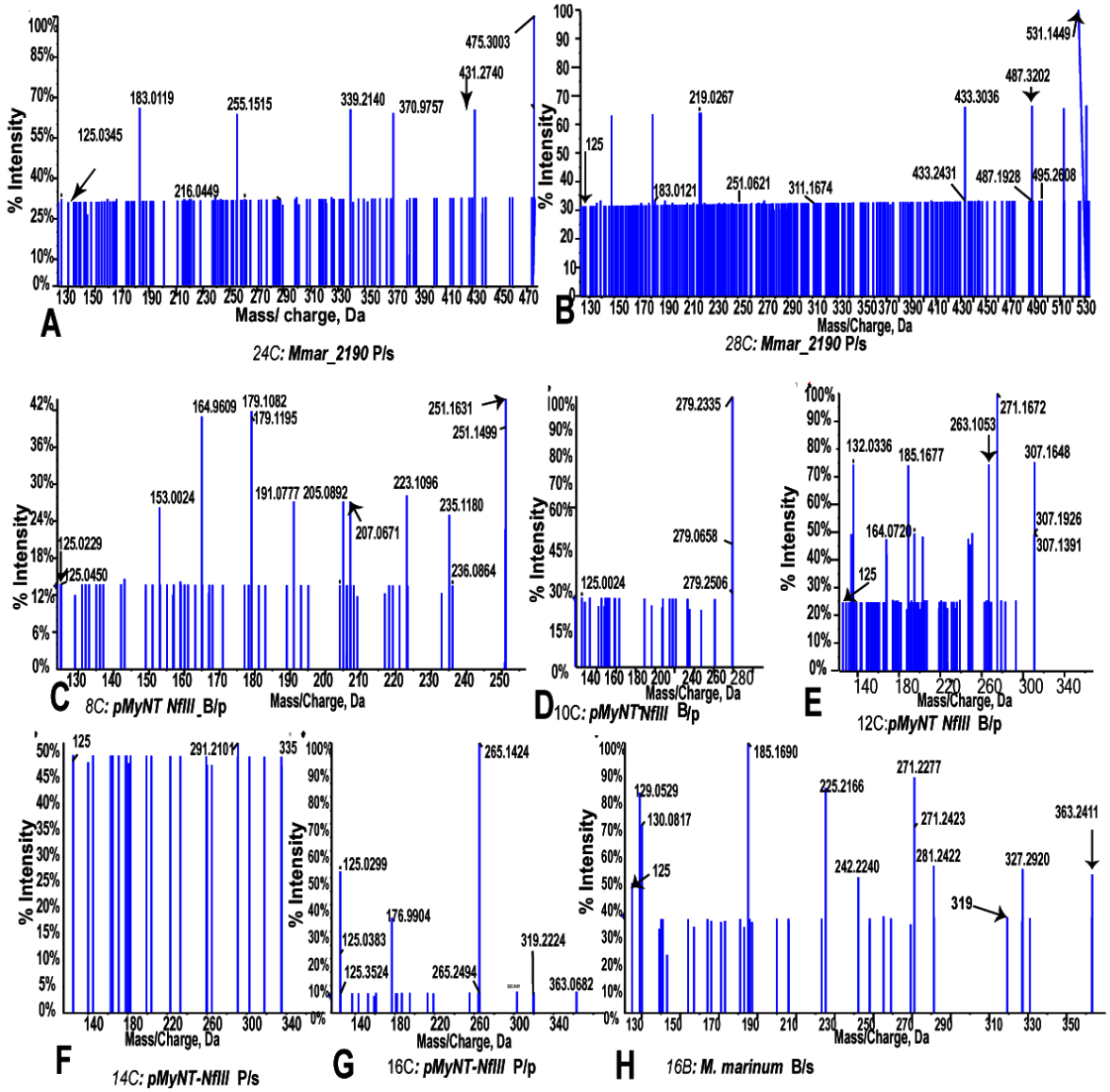


Appendix A4.3: Fragmentation of precursor's ions with m/z 251.1367(8B and 8D), 307.1993 (12B and 12D) and 335(14B and 14D) in negative ion mode. Type III PKSs extracts from MMAR_2190 (2190), and *M. marinum* wildtype give molecular ion peak $(M-H)^-$ at m/z 251. The ion fragments into m/z 207 and 125 observed in methanol extracts from the supernatant of planktonic culture of MMAR_2190 shown in **A** and biofilm culture of *M. marinum* (B from biofilm supernatant, B/s, and C from biofilm pellet B/p) shows characteristics peak pattern for phloroglucinol 8D as described in the schematic representation (Chapter 3, figure 3.5**).** Another precursor molecular ion peak at $(M-H)^-$ at m/z 307 and obtained from extracts of supernatants from planktonic culture (P/s) of MMAR_2190 representing to phloroglucinol 12D show fragments of 263 and 125 (D). The precursor ion peak $(M-H)^-$ at m/z 335 fragments into m/z 125 and 291. The peak pattern is found both in extracts from MMAR_2190 P/s (E) and *M. marinum* B/s (**F**).



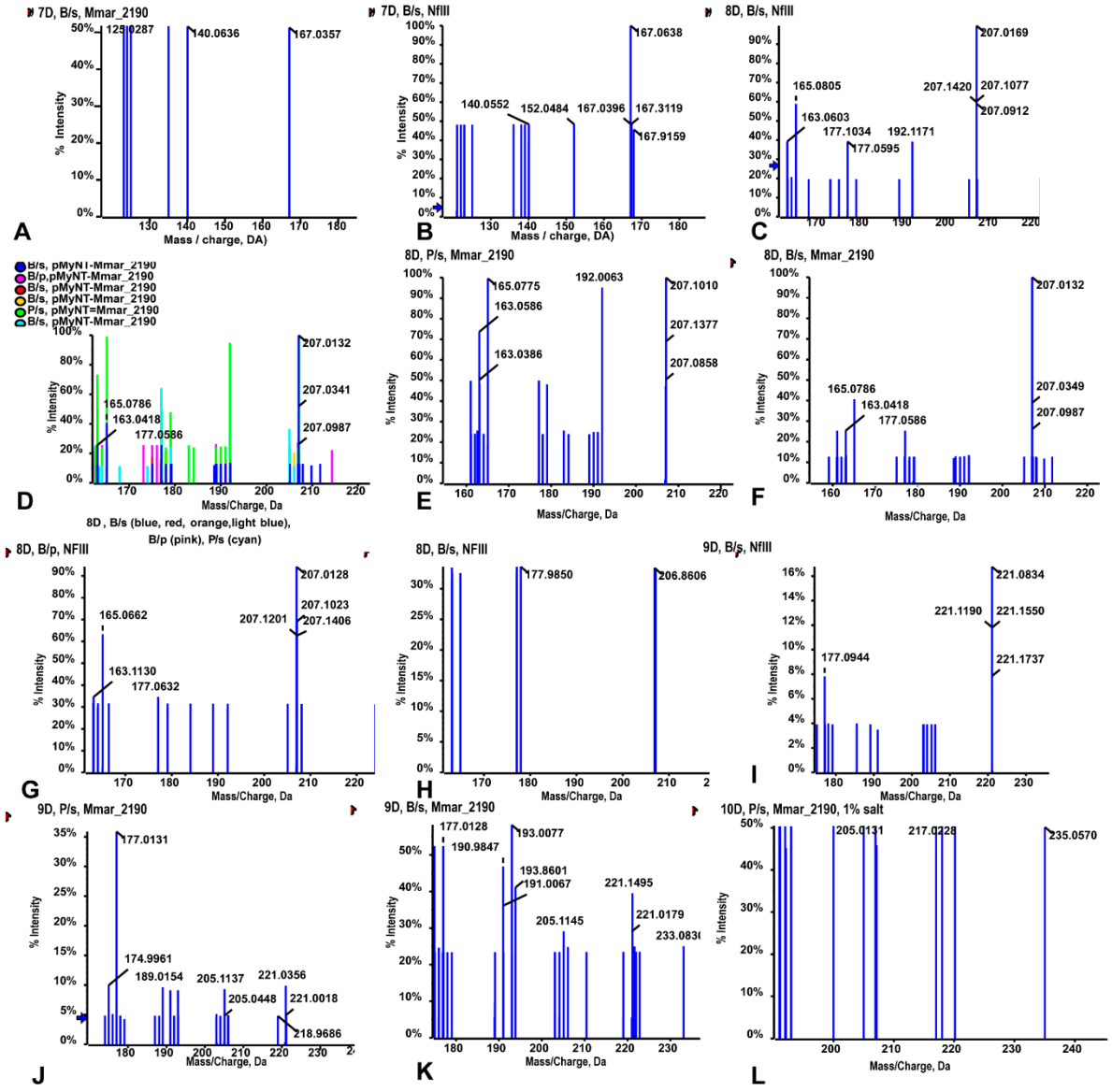
Appendix A4.4: Fragmentation of low molecular weight precursor's ions with m/z 223.1367(6C and 6B), 251.1367 (8B and 8D), and 307.1993 (12B and 12D) and 335(14B and 14D) and 363(16B and 16D) in negative ion mode with comparative study of bacterial culture using saline condition with (NaCl) from 1 to 1.5 %. A shows the presence of molecular ion peak $(M-H)^-$ at m/z 223 fragments into 125 and 179 peaks both in planktonic culture (P/s) and biofilm culture (B/s) along with very less observed in acetamide un-induced overexpressed biofilm culture (obtained from the pellet, B/p)culture of MMAR_2190 (Figure A, and B). C shows the formation of α-tetraketide α-pyrone 6B ($(M-H)^-$ at m/z 223 from biofilm culture (B/s) with fragments, m/z corresponding to fragments m/z 125 with high intensity and m/z 179 peaks with low intensity as shown in figure C. Figure D and figure F shows the presence of molecular ion peak of $(M-H)^-$ at m/z 251.1367 fragments into m/z 125 and 207 ion both in stress condition due to salt (1 % NaCl) in planktonic culture and biofilm culture but in 1 % salt stress C8 tetraketide *α*-pyrone (8B) was observed as shown in figure D and the biofilm culture shows presence of alkyl-phloroglucinol, 8D. The comparative study of bacterial culture shows the better production of phloroglucinol in 1.5 % NaCl, planktonic culture compared to the biofilm culture is shown in figure E. Figure G, and figure H depicts the precursor ion peak at $(M-H)^-$ at 335 with fragmentation pattern similar to the alkyl phloroglucinol bor both planktonic culture from MMAR_2190 (P/s) and *M. marinum* biofilm

culture (B/s). I and J respectively show the fragmentation of *pMyNT-MMAR_2190* C_{16}-phloroglucinol (16D) from both from biofilm supernatant (B/s) and planktonic supernatant (P/s). The presence of precursor ion peak of $(M-H)^-$ 363.1116 and 363.0627 with fragments of 125 and 319 as described in the pattern Chapter 2 confirm the presence of 16D in both biofilm and planktonic culture using MRM and IDA based methods.

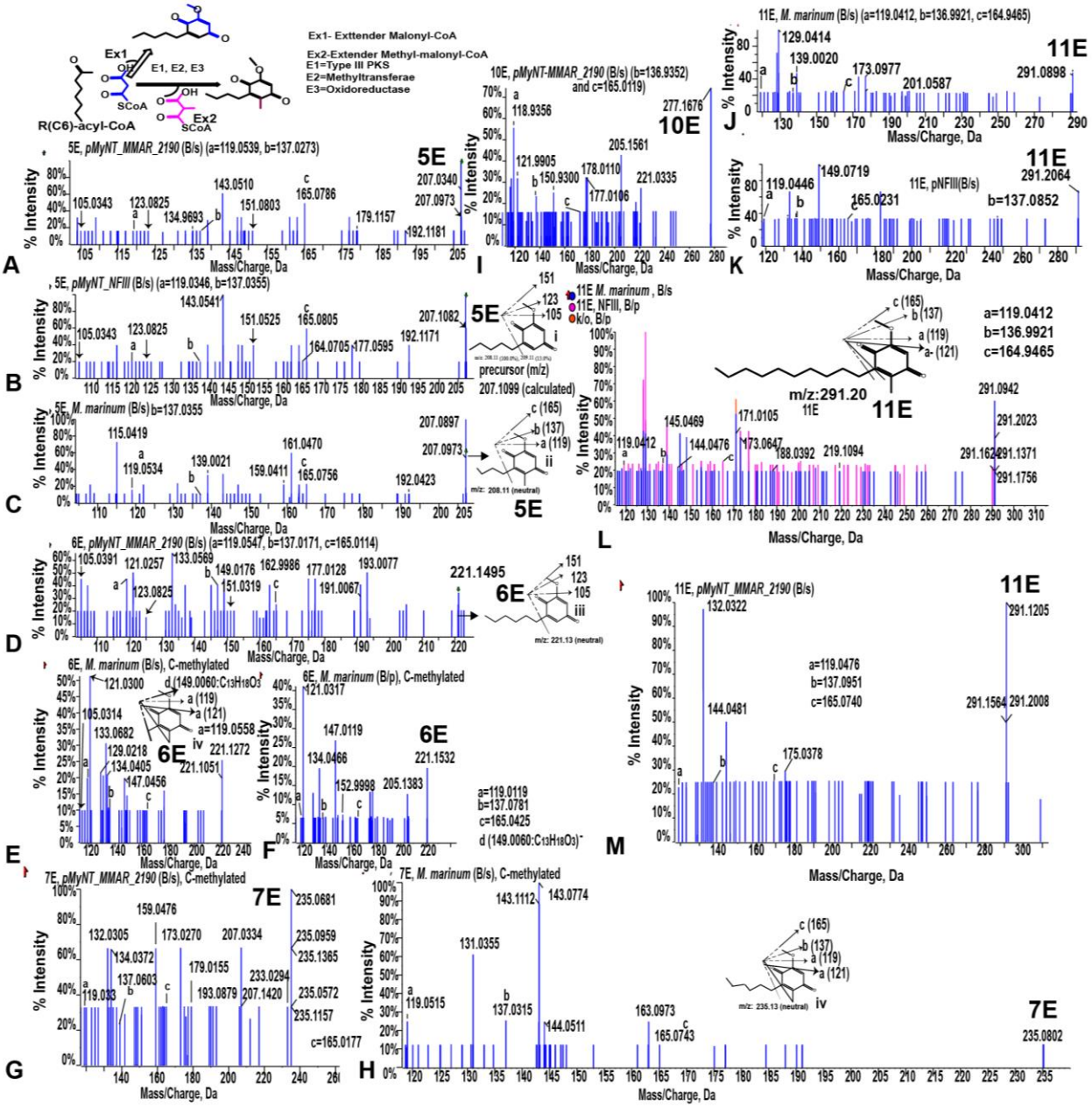


Appendix A4.5: Fragmentation of high molecular weight precursor's ions with m/z 431.2740(24C, figure A)), 531.1449 (28D, figure B) from *pMyNT-MMAR_2190* metabolites extracted from planktonic supernatant culture (P/s) and low and medium chain acyl-CoA derived alkyl phloroglucinol with precursors ion of m/z 251.1631 (8D, Figure C), 279.2335 (10D, figure D), 307.1648 (12D, figure E), 335.1533 (14D, figure F), 363.0682 (16c, Figure G) from *pMyNT-NFIII* strain and 363.2411(16B, Figure H) obtained from *M. marinum* wildtype in negative ion mode. Figure A and Figure B from P/s culture show the fragmentation of precursors ion peak $(M-H)^-$ at m/z 431.2740 fragments into 125 and 387 peaks, 531.1449 give the fragment ions into 125 and 487. The biofilm pellet culture (B/p) from NFIII gives alkyl phloroglucinol 8C, 10C and 12C which fragments to give low intensity 125 as characteristic peak along 207, 235, 263 respectively. The planktonic culture-derived both pellets) (*pMyNT_NFIII* and supernatant (P/s) show the presence of 14D and 16D with 125, 291 fragments, and 125 and 319 fragments

respectively Wildtype M. *marinum also* shows 16B fragment with characteristic peak representing alkyl tetraketide α-pyrone which was found in *E. coli* extracts from MMAR_2190 as described in chapter 2.

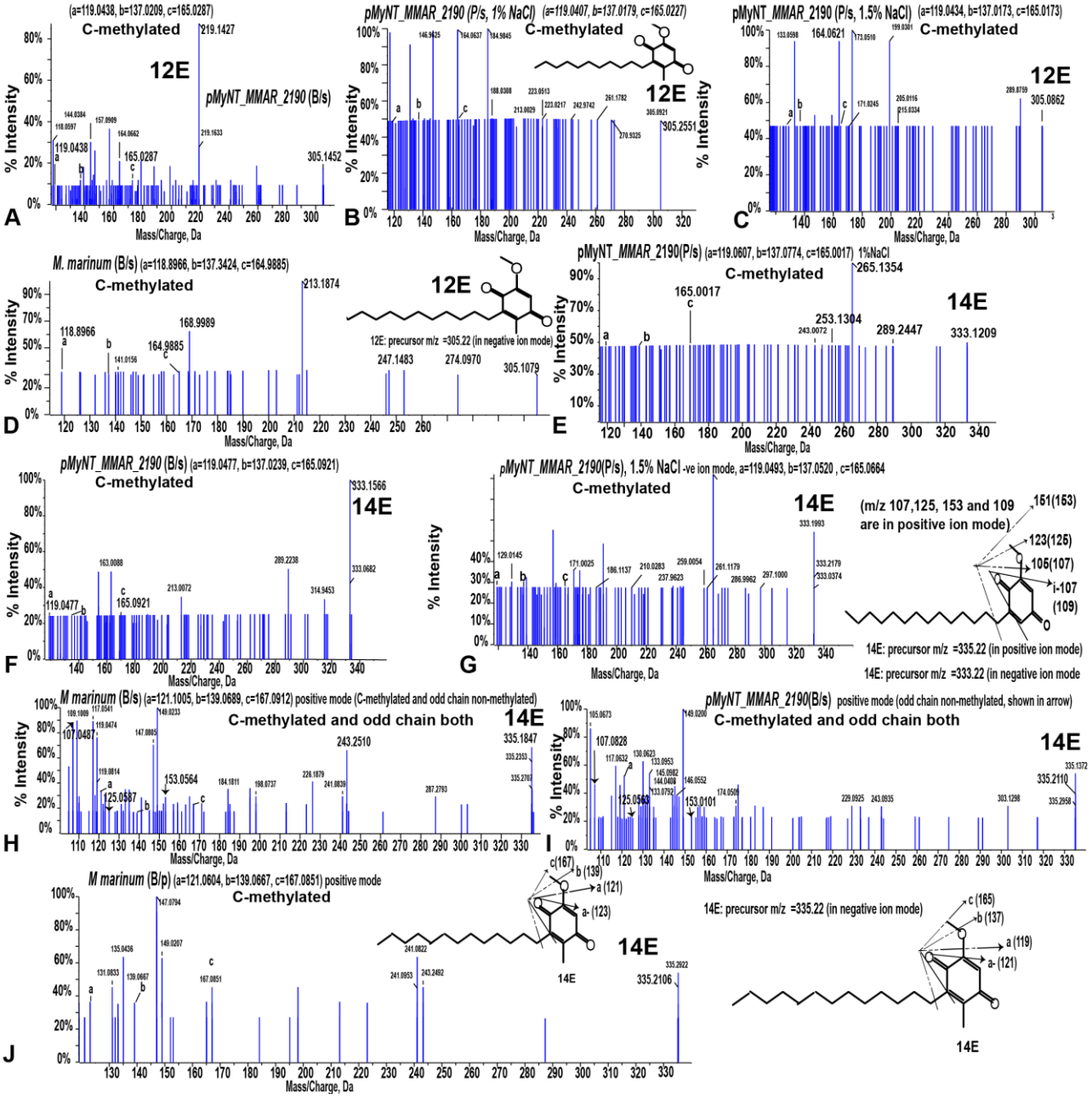


Appendix A4.6: Comparative study of alkyl resorcinol from 7C to 10C of *pMyNT-MMAR_2190* and *pMyNT-NFIII*.Figure A and Figure B shows the presence of odd carbon chain resorcinol 7Cfrom the extracts from biofilm culture supernatant (B/s) of both *pMyNT-MMAR_2190* and *pMyNT-NFIII* overexpressed strain with the fragmentation of precursors ion peak $(M-H)^-$ at m/z 167 fragments into 123 and 125 peaks. C shows the presence of even chain resorcinol 8C with precursor ion of m/z 207 gives the fragment ions into 163.0603 and 165.0805 in B/s culture. D shows the presence of 8C more in biofilm culture (biofilm supernatant, B/s and biofilm pellet, B/p) than that present in planktonic supernatant (P/s) in MMAR_2190. The fragmentation of overlaid P/s and B/s culture, 8C MMAR_2190 is shown in E and F and the compound 8C detected in B/p and B/s culture of NFIII. I, J and K show the presence of odd chain resorcinol 9C in NFIII B/s and P/s and B/s of MMAR_2190, respectively. L shows the presence of 10C, and even chain resorcinol in saline condition (1 % NaCl) of MMAR_2190 with precursor ion 235.0570 showing fragments of 191 and 193.



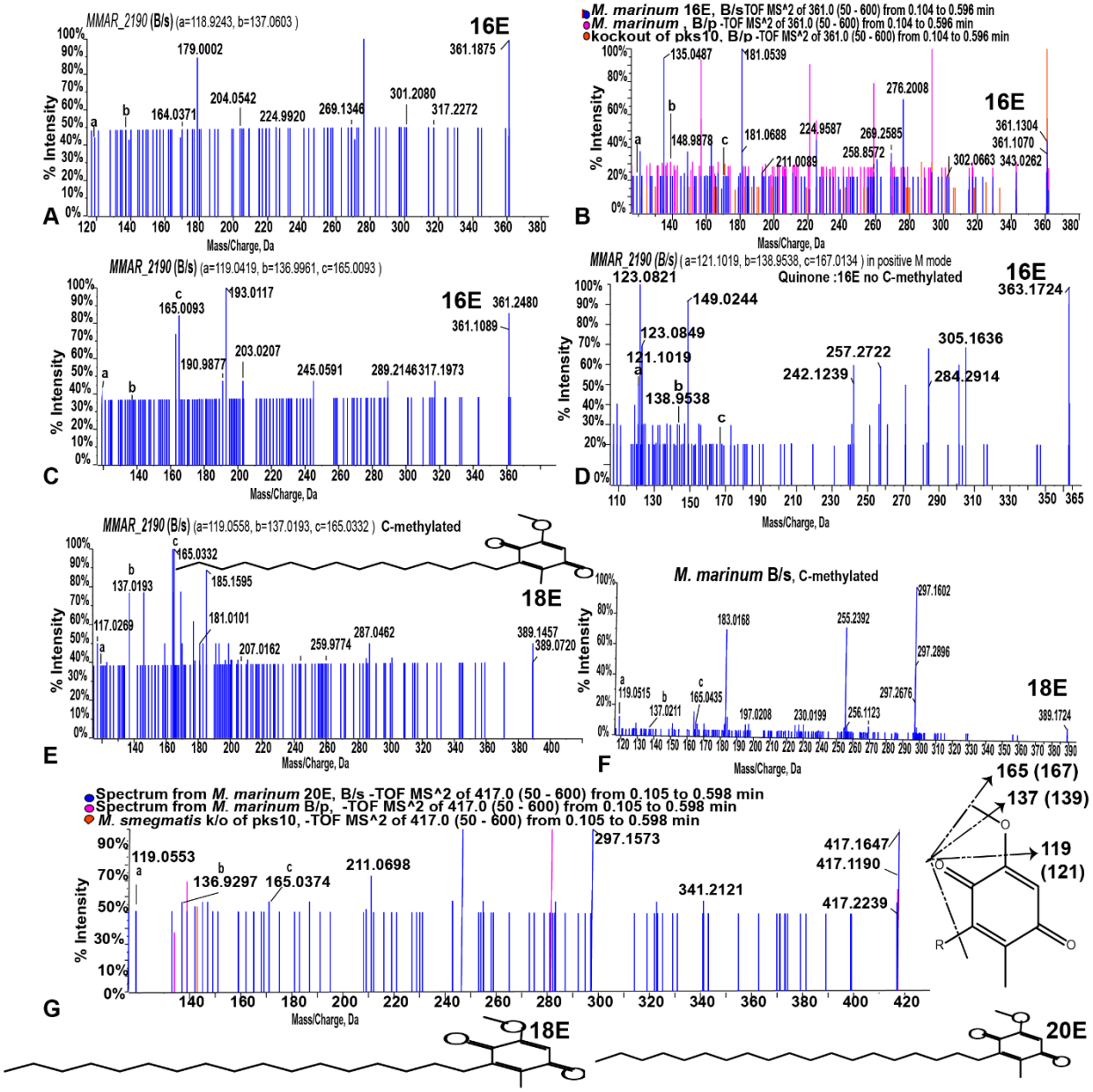
Appendix A4.7: Comparative study of quinones from 5E to 11E from pMyNT based overexpressed strain (*pMyNT-MMAR_2190* and *pMyNT-NFIII* with wildtype *M. marinum*. The figure shows the fragmentation of odd chain quinone 5A (A to C). A also shows the precursors ion peak $(M-H)^-$ at with m/z 207.1099 (calculated from structure) get fragmented into sub-fragments with m/z 105,123 and 151 describing the presence odd chain quinones. C- Methylated quinone give the fragments with m/z at negative ion mode showing sub-fragments of m/z 119, 137, and 165. Both odd chain (non-methylated) and even chain (C-methylated) quinones (5E) were observed in MMAR_2190 (B/s: A), and NFIII (B/s) except for *M. marinum* biofilm culture (B/s) shows presence of C-methylated quinone (C) only in the extracts. 6E was observed in the extract from biofilm pellet of MMAR_2190 (D: B/s) and *M. marinum* (E: B/s and F: B/p). The biofilm culture shows the unique pattern of fragmentation observed with the presence of precursor ion of m/z 221 with sub-fragments of m/z 119, 137, and 165 for C-methylated quinone for both MMAR_2190 (D) and *M. marinum* (E and F) biofilm culture whereas non-C-methylated pattern was found with 105, 123 and 151 in MMAR_2190 only (D). G and H show the odd chain quinone, 7E present in both MMAR_2190 and *M.marinum* biofilm culture with precursor ion of m/z 235 giving fragmentation pattern of C-methylated quinone. Even chain quinone 10E (with precursor ion of m/z, 277) was found only in B/s culture (I) with fragmentation pattern for both C-methylated and non-C-

methylated quinone. Odd chain quinone 11E was found for biofilm culture of MMAR_2190 (M), NFIII (K) and *M. marinum* (J) with precursor ion of m/z 291 giving MS^2 fragments for C-methylated pattern only. The comparative study of odd chain quinone 11E between NFIII (B/p), *M. marinum* (B/s) and Δpks10 knockout of type III PKS in *M. smegmatis* (B/p, control) in L show 11E present better in NFIII strain followed by *M. marinum* but not in knockout strain.

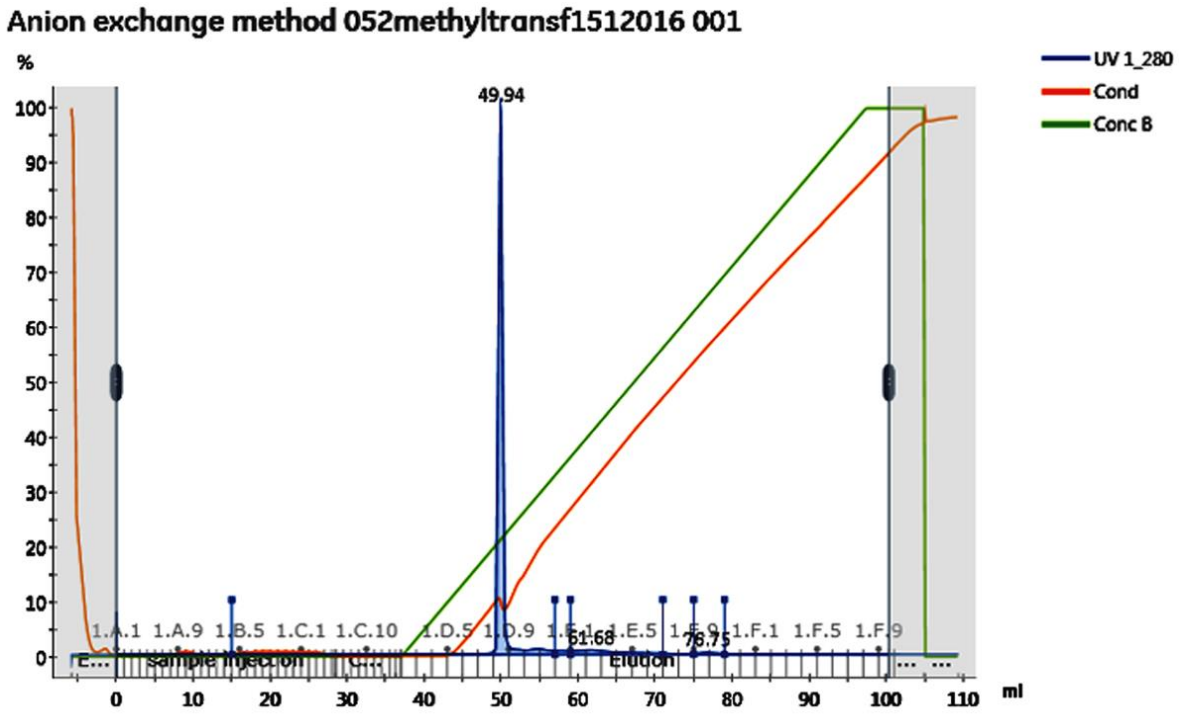


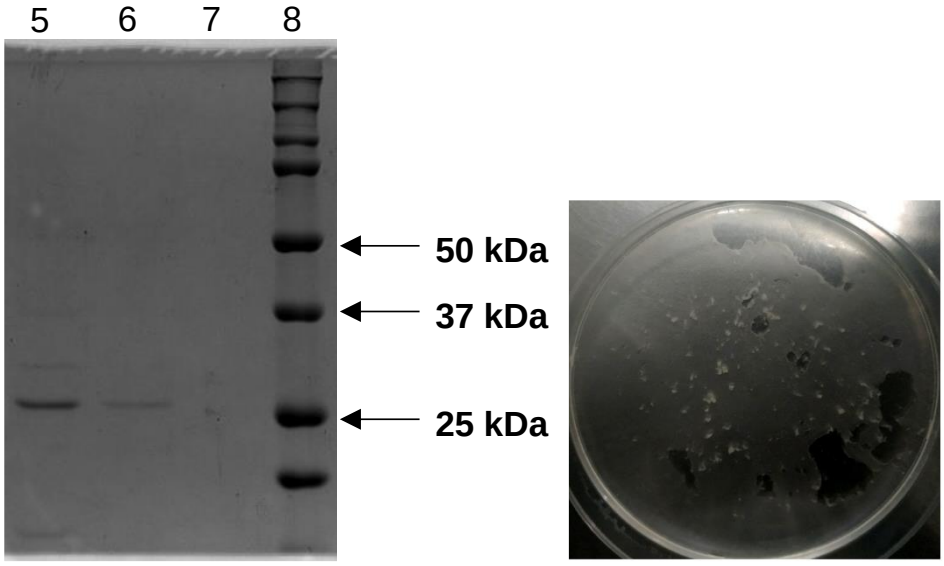
Appendix A4.8: Study of quinones from 12E to 14E from pMyNT based overexpressed strain (*pMyNT-MMAR_2190* and *pMyNT-NFIII* with respect to salt, NaCl stress) and wildtype *M. marinum* biofilm culture. Figure A to D shows the presence of C-methylated quinone, 12E in biofilm culture of MMAR_2190 (A: 12E, B/s) and *M. marinum* wildtype (D: 12E, B/s). No, C-methylated quinone was observed in planktonic culture, but the 1 % (B: 12E, P/s) and 1.5 % NaCl (C: 12E, P/s) show the precursor molecule with $(M-H)^-$ at with m/z 305 fragments into three ions 119, 137 and 165 similar to the overexpressed type III PKS gene and biofilm-based culture study. Figure E to G shows the presence of C-methylated, even chain and non-C-methylated odd chain quinones. E, F, G, H, I and J shows C-methylated quinone present in MMAR_2190 biofilm culture (F: 14E, B/s) majorly. Planktonic supernatant culture (1 % and 1.5 % NaCl treated, P/s) culture also shows the presence of odd chain C-methylated 14E similar to 12E. F and G show the presence of C-methylated and odd chain non-C-

methylated quinone 14E are present in biofilm culture (B/s) of *pMyNT-MMAR_2190* (I) and *M. marinum* wildtype (H) strain. J shows even chain C-methylated quinone is present in pellet extract of *M. marinum* biofilm.



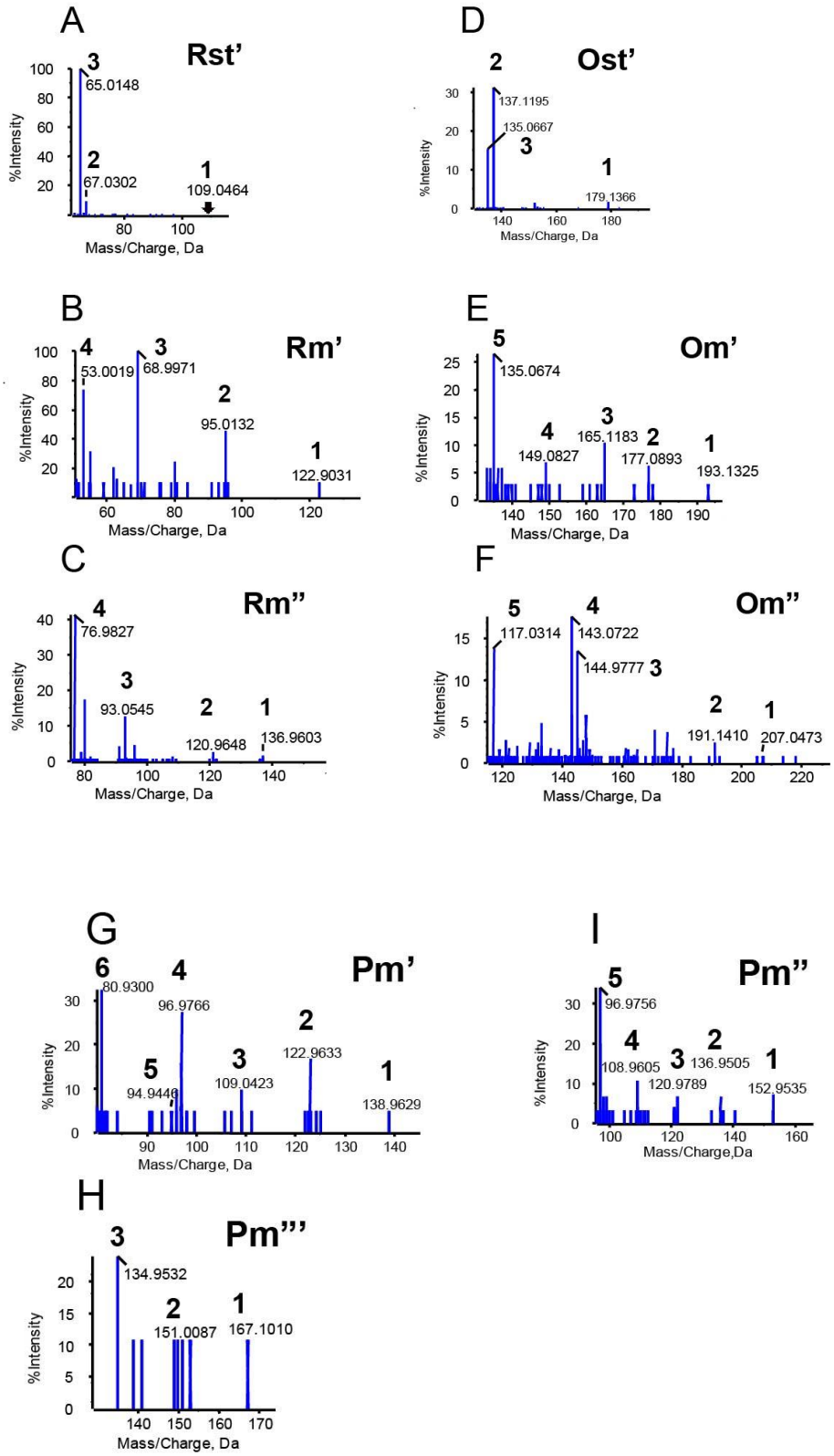
Appendix A4.9: Study of quinones from 16E to 20E from pMyNT based overexpressed (*pMyNT-MMAR_2190* and *pMyNT-NFIII)* and wildtype *M. marinum* biofilm culture. Figure A to D show presence of C-methylated quinone, 16E in biofilm culture of *pMyNT-MMAR_2190* (A, C and D) and *M. marinum* (B: 16E, B/s).The precursor ion $(M-H)^-$ at with m/z 361 fragmented into sub-fragments of m/z 119, 137 and 165 (A, B, and C)and $(M-H)^+$ get fragmented into sub-fragments of m/z 119, 137 121, 139 and 167 (D). Type III PKS, MMAR_2190 overexpressed strain and *M. marinum* wildtype showed a similar, pattern of fragmentation, for 18E, 20E quinone (E, F, and G). No, C-methylated quinone was observed in planktonic culture and Δpks10 knockout of type III PKS. The alky chain R is the denotation for the alkyl group of 16E and 18E and 20E (with CnH2(n+1), where n denotes the length of carbon chain generated using different carbon chain starter and malonyl-CoA as an extender.





5,6,7,8	MMAR_2193	*M. marinum* biofilm culture (10
8.	Protein ladder	days)

Appendix A5.1.: SDS-PAGE profile of overexpressed MMAR_2193 by Anion exchange chromatography and *M. marinum* wildtype biofilm culture

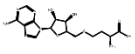
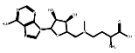
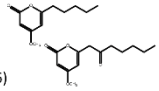
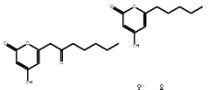
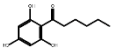
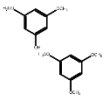


Appendix A5.2: MS-2 fragmentation pattern of different methylated products derived from *in vitro* assay of methyltransferase with different standard starters, namely resorcinol (R), olivetol (O) and phloroglucinol (P). Rst' and Ost' are the standard resorcinol and standard olivetol (5-pentyl resorcinol). Rm', Rm'', Om', Om'', Pm', Pm'', Pm''' denotes methylation of resorcinol, olivetol and phloroglucinol. The symbol with **'**indicates single *O*- methylation, **"**denotes double *O*-methylation and **"'** denotes triple *O*-methylation at the free hydroxyl group.

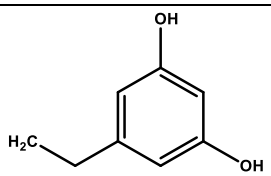
Appendix table 1: Guide table for targeted precursor masses to be used by a targeted metabolomics approach in negative and positive mode of mass spectrometry.

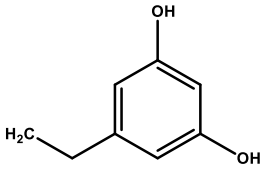
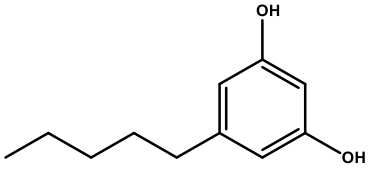
S.N	Code	Starter	Compound	Exact Masses(M)	$[M-H]^-$	$[M-H]^+$
1	6A	C6(Hexanoyl CoA)	Triketide pyrone	182.0948	181.0948	183.0948
2	6B	C6	Tetraketide pyrone	224.1054	223.1054	225.1054
3	6D	C6	Phloroglucinol	224.1054	223.1054	225.1054
4	6C	C6	Resorcinol	180.1156	179.1156	181.1156
5	6E	C6	Quinone	222.1261	221.1261	223.1261
6	8A	C8	Triketide pyrone	210.1261	209.1261	211.1261
7	8B	C8	Tetraketide pyrone	252.1367	251.1367	253.1367
8	8D	C8	Phloroglucinol	252.1367	251.1367	253.1367
9	8C	C8	Resorcinol	208.1469	207.1469	209.1469
10	8E	C8	Quinone	250.1574	249.1574	251.1574
11	10A	C10	Triketide pyrone	238.1574	237.1574	239.1574
12	10B	C10	Tetraketide pyrone	280.168	279.168	281.168
13	10D	C10	Phloroglucinol	280.168	279.168	281.168
14	10C	C10	Resorcinol	236.1782	235.1782	237.1782
15	10E	C10	Quinone	278.1887	277.1887	279.1887
16	12A	C12	Triketide pyrone	266.1887	265.1887	267.1887
17	12B	C12	Tetraketide pyrone	308.1993	307.1993	309.1993
18	12D	C12	Phloroglucinol	308.1993	307.1993	309.1993
19	12C	C12	Resorcinol	264.2095	263.2095	265.2095
20	12E	C12	Quinone	306.22	305.22	307.22
21	14A	C14	Triketide pyrone	294.22	293.22	295.22
22	14B	C14	Tetraketide pyrone	336.2306	335.2306	337.2306
23	14D	C14	Phloroglucinol	336.2306	335.2306	337.2306
24	14C	C14	Resorcinol	292.2408	291.2408	293.2408
25	14E	C14	Quinone	334.2513	333.2513	335.2513
26	16A	C16	Triketide pyrone	322.2513	321.2513	323.2513
27	16B	C16	Tetraketide pyrone	364.2619	363.2619	365.2619
28	16D	C16	Phloroglucinol	364.2619	363.2619	365.2619
29	16C	C16	Resorcinol	320.2721	319.2721	321.2721
30	16E	C16	Quinone	362.2826	361.2826	363.2826
31	18A	C18	Triketide pyrone	350.2826	349.2826	351.2826
32	18B	C18	Tetraketide pyrone	392.2932	391.2932	393.2932
33	18D	C18	Phloroglucinol	392.2932	391.2932	393.2932
34	18C	C18	Resorcinol	348.3034	347.3034	349.3034
35	18E	C18	Quinone	390.3139	389.3139	391.3139
36	20A	C20	Triketide pyrone	378.3139	377.3139	379.3139
37	20B	C20	Tetraketide pyrone	420.3245	419.3245	421.3245
38	20D	C20	Phloroglucinol	420.3245	419.3245	421.3245
39	20C	C20	Resorcinol	376.3347	375.3347	377.3347
40	20E	C20	Quinone	418.3452	417.3452	419.3452
41	22A	C22	Triketide pyrone	406.3452	405.3452	407.3452
42	22B	C22	Tetraketide pyrone	448.3558	447.3558	449.3558
43	22D	C22	Phloroglucinol	448.3558	447.3558	449.3558
44	22C	C22	Resorcinol	404.366	403.366	405.366
45	22E	C22	Quinone	446.3765	445.3765	447.3765
46	24A	C24	Triketide pyrone	434.3765	433.3765	435.3765
47	24B	C24	Tetraketide pyrone	476.3871	475.3871	477.3871
48	24D	C24	Phloroglucinol	476.3871	475.3871	477.3871
49	24C	C24	Resorcinol	432.3973	431.3973	433.3973
50	24E	C24	Quinone	474.4078	473.4078	475.4078
51	26A	C26	Triketide pyrone	462.4078	461.4078	463.4078
52	26B	C26	Tetraketide pyrone	504.4184	503.4184	505.4184
53	26D	C26	Phloroglucinol	504.4184	503.4184	505.4184
54	26C	C26	Resorcinol	460.4286	459.4286	461.4286
55	26E	C26	Quinone	502.4391	501.4391	503.4391
56	28A	C28	Triketide pyrone	490.4391	489.4391	491.4391
57	28B	C28	Tetraketide pyrone	532.4497	531.4497	533.4497
58	28D	C28	Phloroglucinol	532.4497	531.4497	533.4497
59	28C	C28	Resorcinol	488.4599	487.4599	489.4599
60	30A	C30	Triketide pyrone	518.4704	517.4704	519.4704
61	30B	C30	Tetraketide pyrone	560.481	559.481	561.481
62	30D	C30	Phloroglucinol	560.481	559.481	561.481
63	30C	C30	Resorcinol	516.4912	515.4912	517.4912

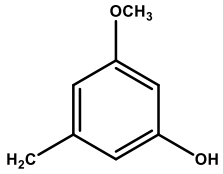
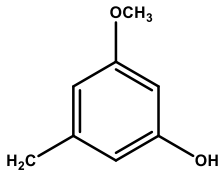
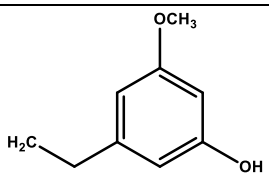
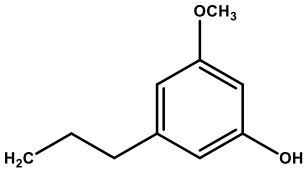
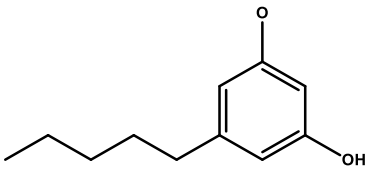
Appendix table 2: Docking of different ligands to the MMAR_2193, a homology model, SAM-dependent methyltransferase generated using Biovia Discovery Studio 4.5. The table shows the Docked structures with their –cDocker energy and –cDocker Interaction Energy with the probale interacting residues.

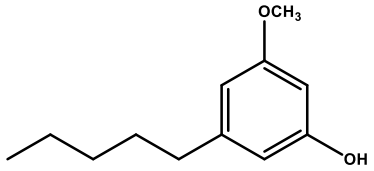
Docked Structure	-cDocker Energy	-cDocker Interaction Energy	Interacting residues
S-Adenosyl Homocysteine (SAH) 	38.044	59.9029	Cys51, Gly50, Arg13, Trp24, His118, Ser114,Gly115, Gly155, Gln56,Cys51
S-Adenosyl Methionine (SAM) 	33.1272	57.3199	Gly50,Gly115, Ser114, His118,Trp24,Aeg188,Gln189,Asp71,Cys119
4-methoxy-6-pentyl-2H-pyran-2-one(mtkC6) 	15.154	30.637	His118,Met116,Ile72,Gln189,Trp24,Phe3
4-methoxy-6-(2-oxoheptyl)-2H-pyran-2-one(mttkC6)	21.147	36.0611	His118,Trp24,Phe147,Ser114,Gly52
4-hydroxy-6-pentyl-2H-pyran-2-one(tkC6) 	33.9799	42.09	Arg13, Trp24, Cys119
4-hydroxy-6-(2-oxoheptyl)-2H-pyran-2-one(ttkC6)	39.5764	49.2252	His118,Arg13,Phe3
1-(2,4,6-trihydroxyphenyl)hexan-1-one(PhlC6) 	32.8321	44.8317	Gln156,Arg13,Arg159
benzene-1,3,5-triol(Phloroglucinol) 		24.633	His118,Ser114,Trp24
5-methoxybenzene-1,3-diol (Pm') 	23.71	27.02	Ser114,Trp24,His118
3,5-dimethoxyphenol(Pm") 	17.734	25.5793	Trp24,Gly50
1,3,5-trimethoxybenzene(Pm"')	14.9319	25.8614	Gly50,Ser114,Asp25,Asn187
5-pentylbenzene-1,3-diol(Olivetol, ResorcinolC6) 	42.6294	41.5929	Trp24,Arg13,Met116
3-methoxy-5-pentylphenol(Olivetol,Om') 	33.3851	37.1157	Trp24,Asp25,Asp55
1,3-dimethoxy-5-pentylbenzene(olivetol,Om") 	31.936	37.3736	Phe3,Ser114,Trp24,Asp25,Arg188
Resorcinol 	17.749	20.6411	Cys119,Trp24,Asp71,Gly115,Trp24
3-methoxyphenol(Resorcinol,1m) 	15.7515	20.7804	Cys119,Trp24,Asp71,Gly115
1,3-dimethoxybenzene(Resorcinol,2m) 	15.3348	22.8399	Asp71,Gly115,Trp24

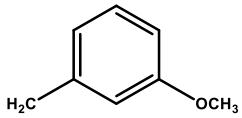
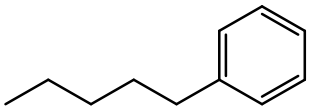
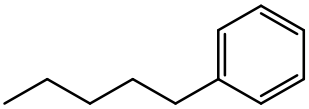
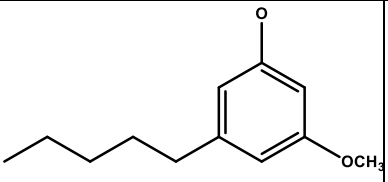
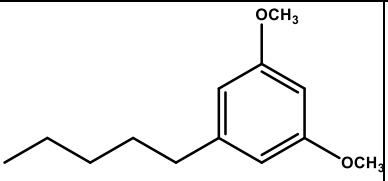
Appendix table 3: Fragmentation of standard resorcinol (Rst), standard olivetol (Ost) and *O*-methylated derivatives with their m/z of fragments derived from precursor ion (1). The statistical analysis of the fragments were shown to confirm the identity of the structure.

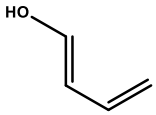
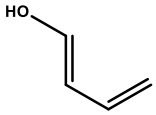
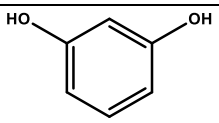
m/z(Ost)	Structure	Composition	Intensity %	Error Da
135.0667 **(3)**		$C_8H_7O_2^-$	15.24	0.0222

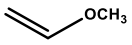
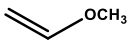
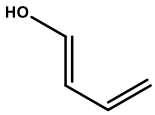
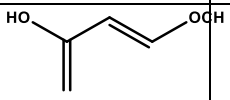
137.1195 **(2)**		$C_8H_9O_2^-$	31.19	0.059
179.1366 (**Ost**) **(1)**		$C_{11}H_{15}O_2^-$	1.57	0.029

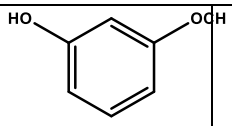
m/z(Om')	**Structure**	**Composition**	**Intensity %**	**Error Da**
135.0674 **(6)**		$C_8H_7O_2^-$	26.44	0.022
137.0469 **(5)**		$C_8H_9O_2^-$	5.75	0.014
149.0827 **(4)**		$C_8H_9O_2^-$	6.90	0.022
165.1183 **(3)**		$C_{10}H_{13}O_2^-$	10.34	0.026
177.0893 **(2)**		$C_{11}H_{13}O_2^-$	6.32	0.003

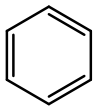
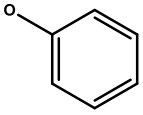
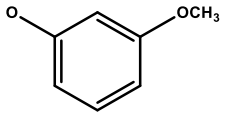
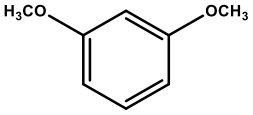
193.1325 **(Om')** **(1)**		$C_{12}H_{17}O_2^-$	2.87	0.001

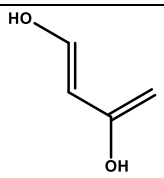
m/z(Om")	Structure	Composition	Intensity %	Error Da
117.0314 **(5)**		$C_8H_5O^-$	13.84	0.003
143.0722 **(4)**		$C_{11}H_{11}^-$	17.68	0.014
144.9777 **(3)**		$C_{11}H_{13}^-$	13.51	0.125
191.1410 **(2)**		$C_{13}H_{19}O_2^-$	2.50	0.033
207.0473 **(1)Om"**		$C_{11}H_{19}O_2^-$	0.83	0.092

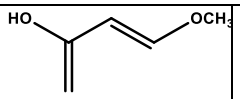
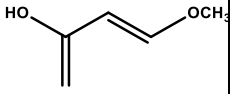
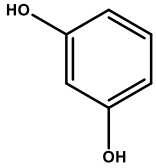
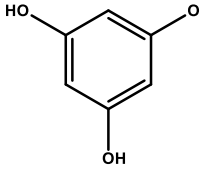
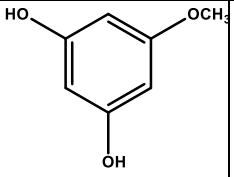
m/z(Rst)	Structure	Composition	Intensity %	Error Da
65.0148 **(3)**		C_4HO^-	100	0.012
67.0302 **(2)**		$C_4H_3O^-$	9.25	0.011
109.0464 (**Rst**) **(1)**		$C_6H_5O_2^-$	0.06	0.017

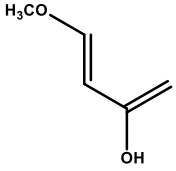
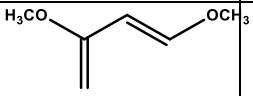
m/z(Rm')	Structure	Composition	Intensity %	Error Da
53.0019 **(5)**		C_3HO^-	73.33	0.001
55.1087 **(4)**		$C_3H_3O^-$	31.19	0.000
68.9971 **(3)**		$C_4H_5O^-$	100	0.038
95.0132 **(2)**		$C_5H_3O_2^-$	45.61	0.001

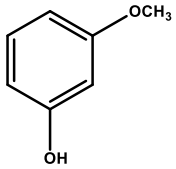
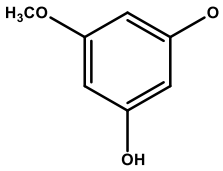
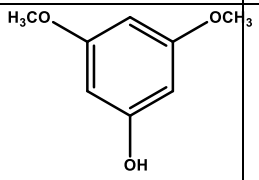
122.9031 **(Rm')** **(1)**		$C_7H_7O_2^-$	10.25	0.142

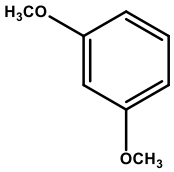
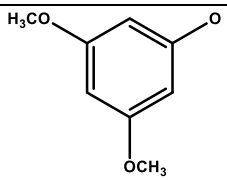
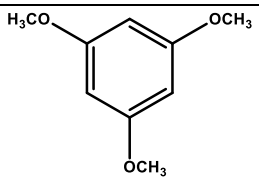
m/z(Rm")	Structure	Composition	Intensity %	Error Da
76.9827 **(4)**		$C_6H_5^-$	41.41	0.057
93.0545 **(3)**		$C_6H_5O^-$	12.61	0.020
120.9648 **(2)**		$C_7H_5O_2^-$	2.65	0.065
136.9603 **(Rm")** **(1)**		$C_8H_9O_2^-$	1.46	0.100

m/z(Pm')	Structure	Composition	Intensity%	Error Da
80.9300 **(6)**		$C_4HO_2^-$	32.35	0.068

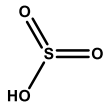
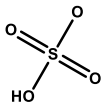
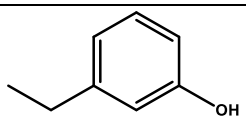
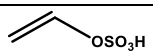
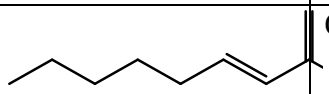
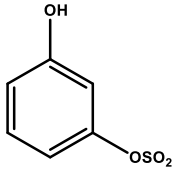
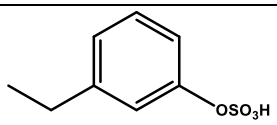
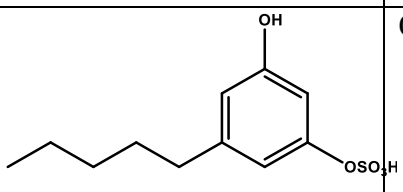
94.9446 **(5)**		$C_5H_3O_2^-$	4.90	0.069
96.9766 **(4)**		$C_5H_5O_2^-$	27.45	0.053
109.0423 **(3)**		$C_6H_5O_2^-$	9.80	0.013
122.9633 **(2)**		$C_6H_3O_3^-$	16.67	0.045
138.9629 **(Pm')** **(1)**		$C_7H_7O_3^-$	4.90	0.077

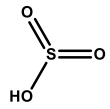
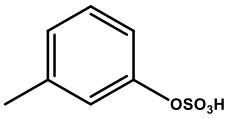
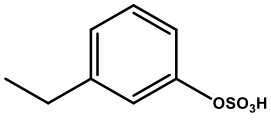
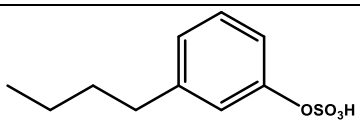
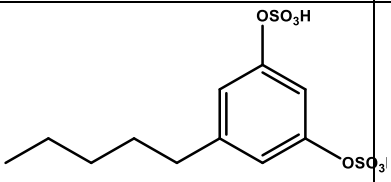
m/z(Pm")	**Structure**	**Composition**	**Intensity%**	**Error Da**
96.9756 **(5)**		$C_5H_5O_2^-$	34.00	0.054
108.9605 **(4)**		$C_6H_5O_2^-$	10.67	0.062

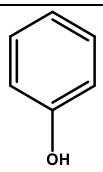
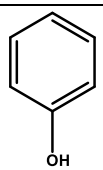
120.9789 **(3)**		$C_7H_5O_2^-$	4.00	0.051
136.9505 **(2)**		$C_7H_5O_3^-$	3.33	0.074
152.9535 **(Pm")** **(1)**		$C_8H_9O_3^-$	7.33	0.102

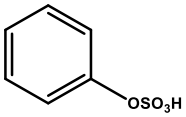
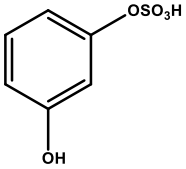
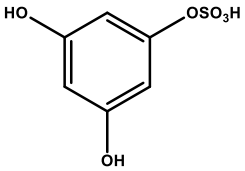
m/z(Pm"')	**Structure**	**Composition**	**Intensity %**	**Error Da**
134.532 **(3)**		$C_8H_7O_2^-$	23.91	0.092
151.0087 **(2)**		$C_8H_7O_3^-$	10.73	0.013
167.1010 **(Pm"')** **(1)**		$C_9H_{12}O_3^-$	10.73	0.013

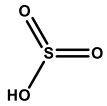
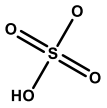
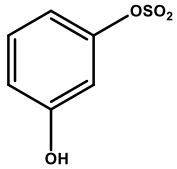
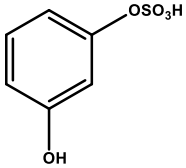
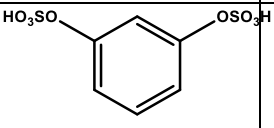
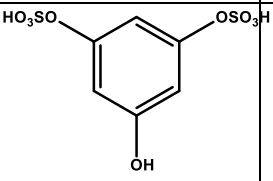
Appendix table 3: Fragmentation of *O*-sulfated derivatives with their m/z of fragments derived from precursor ion (1) and the statistical analysis of the fragments to confirm the identity of the structure.

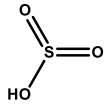
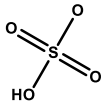
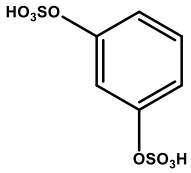
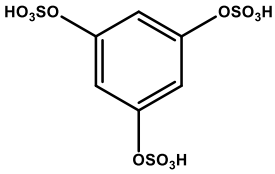
m/z(Os')	Structure	Composition	Intensity %	Error Da
79.9708 **(8)**		O_3S^-	10.36	0.013
96.9762 **(7)**		HO_4S^-	13.23	0.016
118.9575 **(6)**		$C_8H_7O^-$	79.16	0.093
120.9571 **(5)**		$C_2HO_4S^-$	21.39	0.003
138.9683 **(4)**		$C_9H_{15}O^-$	100.00	0.144
171.9248 **(3)**		$C_6H_4O_4S^-$	19.85	0.059
198.9398 **(2)**		$C_8H_7O_4S^-$	30.43	0.067
259.1547 **(Os') (1)**		$C_{11}H_{15}O_5S^-$	1.1	0.031

m/z(Os”)	Structure	Composition	Intensity%	Error Da
79.9716 **(5)**		O_3S^-	30.10	0.014
183.0493 **(4)**		$C_7H_3O_4S^-$	100.00	0.074
197.0611 **(3)**		$C_8H_5O_4S^-$	13.29	0.070
225.0967 **(2)**		$C_{10}H_9O_4S^-$	2.40	0.074
339.2577 **(1)**		$C_{11}H_{15}O_8S_2^-$	21.45	0.236

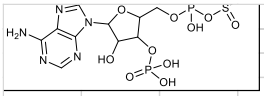
m/z(Ps’)	Structure	Composition	Intensity%	Error Da
89.0418 **(5)**		C_6HO^-	45.83	0.038
92.9435 **(4)**		$C_6H_5O^-$	75.00	0.091

173.1269 **(3)**		$C_6H_5O_4S^-$	75.00	0.136
189.1621 **(2)**		$C_6H_5O_5S^-$	75.00	0.176
205.1936 **(1)Ps'**		$C_6H_5O_6S^-$	50.00	0.212

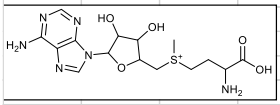
m/z(Ps")	**Structure**	**Composition**	**Intensity%**	**Error Da**
79.9701 **(6)**		O_3S^-	40.48	0.013
96.9779 **(5)**		HO_4S^-	48.81	0.018
167.9322 **(4)**		$C_6O_4S^-$	28.57	0.020
189.1598 **(3)**		$C_6H_5O_5S^-$	34.52	0.173
269.1306 **(2)**		$C_6H_5O_8S_2^-$	100.00	0.187
285.1672**(1)Ps"**		$C_6H_5O_9S_2^-$	7.14	0.229

m/z(Ps”’)	Structure	Composition	Intensity %	Error Da
79.9706 **(4)**		O_3S^-	11.58	0.013
96.9770 **(3)**		HO_4S^-	100.00	0.017
267.1911 **(2)**		$C_6H_3O_8S_2^-$	4.16	0.264
365.1786 **(1)Ps”’**		$C_6H_5O_{12}S_3^-$	0.15	0.284

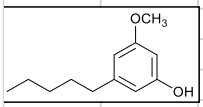
Appendix table 4: Calculated exact masses of different molecules described in *in vitro* assays of methyltransferase and sulfotransferase assay. The m/z of the molecules are calculated in negative and positive ion mode.



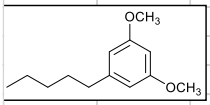
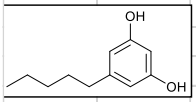
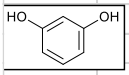
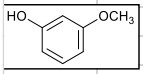
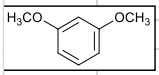
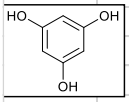
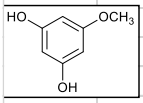
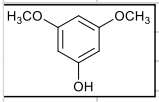
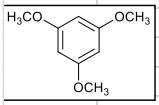
Chemical Formula: C10H15N5O13P2S	PAPS			
Exact Mass: 506.99		C10H14N5O13P2S-		505.979
Molecular Weight: 507.26				
m/z: 506.99 (100.0%), 507.99 (10.8%), 508.98 (4.5%), 508.99 (2.7%), 507.98 (1.8%)		C10H16N5O13P2S+		507.9935
Elemental Analysis: C, 23.68; H, 2.98; N, 13.81; O, 41.00; P, 12.21; S, 6.32				



Chemical Formula: C15H23N6O5S+	SAM	C15H24N6O5S+		400.1523
Exact Mass: 399.14				
Molecular Weight: 399.45				
m/z: 399.14 (100.0%), 400.15 (16.2%), 401.14 (4.5%), 400.14 (2.2%), 401.15 (1.2%), 401.15 (1.0%)		C15H22N6O5S-		398.1378
Elemental Analysis: C, 45.10; H, 5.80; N, 21.04; O, 20.03; S, 8.03				



Chemical Formula: C12H18O2		C12H17O2-	193.1234
Exact Mass: 194.13			
Molecular Weight: 194.27		C12H19O2+	195.138
m/z: 194.13 (100.0%), 195.13 (13.0%)			
Elemental Analysis: C, 74.19; H, 9.34; O, 16.47			

		Chemical Formula: C13H20O2		C13H19O2-	207.1391
		Exact Mass: 208.15			
		Molecular Weight: 208.30		C13H21O2+	209.1536
		m/z: 208.15 (100.0%), 209.15 (14.1%)			
		Elemental Analysis: C, 74.96; H, 9.68; O, 15.36			
	Olivitol	Chemical Formula: C11H16O2		C11H15O2-	179.1078
		Exact Mass: 180.12			
		Molecular Weight: 180.25		C11H17O2+	181.1223
		m/z: 180.12 (100.0%), 181.12 (11.9%)			
		Elemental Analysis: C, 73.30; H, 8.95; O, 17.75			
	Resorcinol	Chemical Formula: C6H6O2		C6H5O2-	109.0295
		Exact Mass: 110.04			
		Molecular Weight: 110.11		C6H7O2+	111.0441
		m/z: 110.04 (100.0%), 111.04 (6.5%)			
		Elemental Analysis: C, 65.45; H, 5.49; O, 29.06			
	o-methylated resorcinol	Chemical Formula: C7H8O2		C7H7O2-	123.0452
		Exact Mass: 124.05			
		Molecular Weight: 124.14		C7H9O2+	125.0597
		m/z: 124.05 (100.0%), 125.06 (7.6%)			
		Elemental Analysis: C, 67.73; H, 6.50; O, 25.78			
	1,3-dimethoxybenzene	Chemical Formula: C8H10O2		C8H9O2-	137.0608
		Exact Mass: 138.07			
		Molecular Weight: 138.17		C8H9O2+	138.0675
		m/z: 138.07 (100.0%), 139.07 (8.7%)			
		Elemental Analysis: C, 69.55; H, 7.30; O, 23.16			
		Chemical Formula: C6H6O3		C6H5O3-	125.0244
	Phloroglucinol	Exact Mass: 126.03			
		Molecular Weight: 126.11		C6H7O3+	127.039
		m/z: 126.03 (100.0%), 127.04 (6.5%)			
		Elemental Analysis: C, 57.14; H, 4.80; O, 38.06			
	Methylated phloroglucinol	Chemical Formula: C7H8O3		C7H7O3-	139.0401
		Exact Mass: 140.05			
		Molecular Weight: 140.14		C7H9O3+	141.0546
		m/z: 140.05 (100.0%), 141.05 (7.6%)			
		Elemental Analysis: C, 60.00; H, 5.75; O, 34.25			
	3,5-dimethoxy phenol	Chemical Formula: C8H10O3		C8H9O3-	153.0557
		Exact Mass: 154.06			
		Molecular Weight: 154.17		C8H11O3+	155.0703
		m/z: 154.06 (100.0%), 155.07 (8.7%)			
		Elemental Analysis: C, 62.33; H, 6.54; O, 31.13			
	1,3-5-trimethoxybenzene	Chemical Formula: C9H12O3		C9H11O3-	167.0714
		Exact Mass: 168.08			
		Molecular Weight: 168.19		C9H13O3+	169.0859
		m/z: 168.08 (100.0%), 169.08 (9.7%)			
		Elemental Analysis: C, 64.27; H, 7.19; O, 28.54			

APPENDICES

LB Broth:

Component	g/100ml
NaCl	1.0g
Yeast Extract	0.5g
Tryptone	1.0g

All the components were dissolved in distilled water and the volume was made up to 100 ml. The media pH was adjusted to 7 using 10N NaOH. The media was autoclaved for 15 minutes at 15 psi pressure.

LB Agar:

Component	g/100ml
NaCl	1.0g
Yeast Extract	0.5g
Tryptone	1.0g
Agar	1.5g

All components were dissolved in distilled water before adding agar and the volume ws made up to 100 ml. The media pH was adjusted to 7 using 10N NaOH. The media was autoclaved for 15 minutes at 15 psi pressure.

Middlebrook 7H9 Broth Base:

2.35 g of 7H9 Broth Base was suspended in 450 ml of distilled water. 2 ml of glycerol was, and 0.05 % Tween-80 was added to the media. The medium was dissolved completely and was autoclaved at 121 ˚C for 10 minutes. The final mixture was cooled to 45 ˚C, and aseptically 1 vial of Middlebrook ADC growth supplement was added. The whole mixture was mixed thoroughly and aliquoted in different vials.

Middlebrook 7H11 Agar Base:

10.25 g of 7H11 Agar Base was suspended in 450ml of distilled water. 2.5 ml of glycerol was added to the media. The medium was heated and dissolved completely. It was sterilized at 121 ˚C for 15 minutes and was cooled down to 50 ˚C. Aseptically 1 vial of Middlebrook OADC growth supplement was added. The whole mixture was mixed thoroughly before dispensing.

Sauton's Fluid Medium Base

3.19g of mineral media was suspended in 1000ml distilled water containing 20ml glycerol. It was dissolved completely using a hot plate with a magnetic stirrer and was sterilized by autoclaving at 121 ˚C for 15 minutes. The final mixture was cooled to 45 ˚C and aseptically and aseptically 2 % glucose was added and mixed thoroughly.

Antibiotic Stocks:

Ampicillin:

100 mg/ml stock solution was made in Millipore water. It was filter-sterilized using a 0.22 µm filter. The aliquots were made and stored at –20 °C.

Kanamycin:

500 mg/ml stock solution was made in Millipore water. It was filter-sterilized using a 0.22 µm filter. The aliquots were made and stored at –20 °C.

Acetamide solution:

5.643 M acetamide (Molecular weight =59.07) stock solution was prepared in millipore water. The solution was filter sterilized and stored at 4 °C. 1 mM acetamide was used for the induction of pMyNT- strains.

Solutions for chemical competent cells:

TF1:

Component	**/100ml**
1M Potassium Acetate	3 ml (0.29442 gm)
1M KCl	10 ml (0.74550 gm)
0.5M $CaCl_2$	2 ml(0.147 gm)
1M $MnCl_2$	5 ml (0.9895 gm)
50 % Glycerol v/v	30 ml
Millipore water	~50 ml

The pH was adjusted to 5.8 with 0.2M acetic acid and the volume was made100ml with Millipore water. The solution was filter sterilized and stored at 4°C.

TF2:

Component	**/100 ml**
1M MOPS	1 ml (0.2093g)
1M KCl	1 ml (0.07455g)
0.5M $CaCl_2$	15 ml (1.1025g)
50 % Glycerol v/v	30 ml
Millipore water	~50 ml

The pH was adjusted to 6.5 with 1 M KOH and the volume was made up to 100ml. It was filter sterilized and was stored at 4°C.

Buffers for plasmid isolation of DNA:

Resuspension Buffer (P1):

Component	**/1000ml**
Tris.Cl	50 mM/6.06g
EDTA	10 mM/3.75g
Rnase A	100 µg/ml

6.06g Tris base was dissolved in 800ml of Millipore water and 3.75g EDTA.2H2O added to dissolve in it. The pH was adjusted to 8.0 with HCl and the volume was made up to 1 L with Millipore water. It was autoclaved and cooled. 10 ml RNaseA (10 mg/ml stock) added and whole solution (P1) was stored at 4°C.

Lysis Buffer (P2):

Component	**/1000 ml**
NaOH	200 mM
1 % SDS	50 ml of 20%SDS

8.0g NaOH pellets was dissolved in 950 ml autoclaved millipore water. 50 ml 20% SDS solution made in autoclaved millipore water was added. The solution was filter sterilized and stored at room temperature.

Neutralizing Buffer (P3):

Component	**/1000ml**
Potassium Acetate	3M/294.5 g

294.5g Potassium Acetate was added in 500 ml millipore water. The pH was adjusted to 5.5 with glacial acetic acid (~110 ml) and the volume was made up to 1L with Millipore water. It was autoclaved and stored at room temperature.

RNase A (10 mg/ml):

1 ml, 0.5 M Tris pH 7.5 was added and 1.5 ml 0.5M NaCl to 25 ml millipore water. The pH was adjusted to 7.5 with 0.1 M HCl. The volume was made up to 50ml. It was autoclaved and cooled. 0.02 gm RNase A was dissolved in 2ml of 10mM Tris pH 7.5 – 15mM NaCl buffer prepared above. The solution was aliquoted into 2 microfuge tubes 1ml in each and was heated at 100 °C for exactly 15 min. It was cooled at room temperature and stored at –20 °C.

Solutions for agarose gel electrophoresis:

Ethidium Bromide (10 mg/ml):

Ethidium Bromide: 1g

It was dissolved in 100 ml autoclaved Millipore water and was stored in an amber colored bottle at 4°C.

50X TAE:

Component	**/1000ml**
Tris base	242g
$Na_2EDTA.2H_2O$	37.2g
glacial acetic acid	57.1 ml

Dissolved in 900 ml Millipore water. Added 57.1 ml glacial acetic acid. Volume made up to 1 L with Millipore water. It was autoclaved and stored at room temperature.

1Kb DNA Ladder Stock:

1Kb DNA Ladder: 10 µl (NEB)

Agarose Gel Sample Buffer (6X): 10 µl (NEB, ready to use)

Autoclaved Millipore water: 30 µl

Store at -20°C.

Requirements for PCR:

dNTP solution:

Nucleotide	**Required concentration**	**Stock concentration**	**Amount of nucleotide per 100 µl**
dATP	2.5 mM	100 mM	2.5 µl

dCTP	2.5 mM	100 mM	2.5 µl
dTTP	2.5 mM	100 mM	2.5 µl
dGTP	1.25 mM	100 mM	1.25 µl
7-deaza GTP	1.25 mM	5 mM	25 µl
Total			33.75 µl

Autoclaved Millipore water 66.25 µl was added to make volume to 100 µl. Stored at –20°C.

IPTG:

IPTG: 0.4g

0.4g was dissolved in 1 ml Millipore water. Volume was adjusted to 2 ml. It was filter sterilized and was dispensed into 1ml aliquots. The aliquots were stored at –20°C.

Solutions for SDS-PAGE:

30 % Acrylamide Mix:

Component	**/100 ml**
Acrylamide	29g
N, N-methylene bis acrylamide	1g
Millipore water	100 ml

1g N, N-methylene bis acrylamide was dissolved in 50 ml Millipore water and 29g acrylamide was added and dissolved. The volume was made up to 100 ml and stored at 4°C in the amber colored bottle.

5X Tris Glycine Electrophoresis Buffer (1L):

Component	**/100ml**
Tris base	15.1g
Glycine	94g
SDS	50 ml 10 % (w/v) stock solution of SDS

Dissolved in 900 ml Millipore water. Added 50 ml 10 % (w/v) stock solution of SDS. Made up volume to 1000 ml and stored at room temperature.

1M Tris (pH 6.8):

Tris base: 12.11g

Dissolved in 70ml Millipore water. Adjusted pH to 6.8 with conc. HCl (few mls). Made up volume to 100ml and stored at room temperature.

1.5M Tris (pH 8.8):

Tris base: 45.417g

Dissolved in 200 ml Millipore water. Set pH to 8.8 with conc HCl (few µl). Made up volume to 250 ml and stored at room temperature.

10 % SDS Stock Solution:

Dissolved 5g SDS in 50ml Millipore water and stored at room temperature.

10 % Ammonium Persulfate (APS):

Dissolved 0.1g APS in 1ml Millipore water and stored at room temperature.

4X SDS Gel Loading Buffer (10 ml):

Component	**/10 ml**
1M Tris (pH 6.8)	2.5 ml
β-mercaptoethanol	2 ml
SDS	0.8g
Bromophenol Blue	1.7mg
Glycerol	4 ml

Aliquots were made and stored in microfuge tubes at –20°C.

Coomassie Brilliant Blue Stain:

Component	**/1000ml**
Coomassie Brilliant Blue R250/G250	1.25g
Methanol	450ml
Millipore water	450ml
Glacial acetic acid	100ml

1.25g Coomassie Brilliant Blue R250 was dissolved in 900ml of methanol: H2O (1:1 v/v). 100ml glacial acetic acid was added to the mixture. The solution was filtered through Whatman No.1 filter to remove any particulate matter and was stored at room temperature.

Destain:

Component	**/1000ml**
Methanol	450ml
Millipore water	450ml
Glacial acetic acid	100ml

Above mentioned individual components were mixed all together and stored at room temperature.

Solutions for Ni^{2+}-NTA affinity chromatography:

1M Imidazole:

Imidazole: 6.808g

Imidazole (6.808g) was dissolved in 100 ml Millipore water and stored at 4°C.

1M Tris (pH 8.0):

Tris base: 12.11g

12.11g tris base was dissolved 80ml Millipore water and the pH was adjusted to 8.0 with 0.1M HCl. The volume was made up volume to 100ml. It was stored at room temperature.

Loading buffer:

Component	/100 ml
Tris-HCl	50mM
Glycerol	10 %
NaCl	150mM

50mM Tris pH8, 10 % glycerol, 150mM NaCl

Wash buffer:

Component	/100 ml
Tris-HCl	50mM
Glycerol	10 %

50mM Tris pH8, 10 %glycerol

Elution buffers:

Component	/100 ml
Tris-HCl	50 mM
Glycerol	10 %
Imidazole	5-250 mM

50mM Tris pH8, 10 % glycerol, 10-100 mM imidazole

FPLC buffers:

Buffer A:

Component	/1000ml
Tris-HCl (pH 8)	50 mM
Glycerol	10 %
EDTA	2 mM

Buffer B:

Component	/100ml
Tris-HCl (pH8)	50mM
Glycerol	10 %
EDTA	2 mM
NaCl	1 M

Publications, Patents and Presentations

Publications, Patents, and Presentations

Publications

Pratap, S., Kant A, Giri, S., **Giri, G.R,** Saxena, P* and Krishan, V*. (2018). Crystal structure of type III polyketide synthases from *Mycobacterium marinum*, PDB ID: 5ZQ9, RCSB Protein Data Bank (2018)

Parvez, A.[1], Giri, S.[1], **Giri, G.R**, Kumari, M, Bisht, R. & Saxena, P. (2018). Novel Type III Polyketide Synthases Biosynthesize Methylated Polyketides in *Mycobacterium marinum. Scientific Reports* 8, 6529.

Patents

Methodology for Production of Acylphloroglucinol or Analogues using Microorganisms, Saxena, P., Giri, S., Parvez, A., **Giri, G.R**. and Bisht, R., IPA No. 201811013921.

Methodology for Production of Alkylresorcinol or Analogues using Microorganisms, Saxena, P., Giri, S., Parvez, A., **Giri, G.R**. and Bisht, R., IPA No. 201811013918.

Presentations

Poster Presentation at **South Asian Biotechnology Conference-2017**, (**March 16-18**) Kathmandu, Nepal on "Study of the mycobacterial type III polyketide synthases and modifying enzyme." (Best poster award from American Society of Microbiology)

Poster Presentation at **International Conference on Agriculture, Allied & Applied Sciences** (ICAAAS-2018, April 28-29**),** New Delhi, India on "In-silico study of methyltransferase and sulfotransferase modifying mycobacterial type III PKSs products."

Oral talk on **SABC conference 2018 (March 28-30, 2018), Colombo, Srilanka** on "Role of type III PKS and modifying enzyme from *M. marinum* in the regeneration of biofilm."

Poster Presentation at **1st Annual MANA conference from November 15 – 17**, 2019 at Georgia Tech, Atlanta, Georgia, USA on "Multiple *O*-methylated phenolic scaffolds characterized from methyltransferase of *Mycobacterium marinum*."