Characterization of the Genes Involved in Santalene Biosynthetic Pathway in Indian Sandalwood *Santalum album* Linn.

Thesis submitted to the University of Pune
For the degree of
Doctor of Philosophy
in
Biotechnology
by

Prabhakar Lal Srivastava

Research Supervisor
Dr. H. V. Thulasiram

Division of Organic Chemistry
CSIR-National Chemical Laboratory
Pune 411008
India

June 2014

Problems cannot be solved

by the same level of

thinking that created them.

-Albert Einstein

CERTIFICATE

This is to certify that the work presented in this thesis entitled, "**Characterization of the Genes Involved in Santalene Biosynthetic Pathway in Indian Sandalwood *Santalum album* Linn.**" by **Mr. Prabhakar Lal Srivastava**, for the degree of **Doctor of Philosophy**, was carried out by the candidate under my supervision in the Division of Organic Chemistry, CSIR-National Chemical Laboratory, Pune-411008, India. Any material that has been obtained from other sources has been duly acknowledged in the thesis.

Date:

Place: Pune

Dr. H. V. Thulasiram

(Research Guide)

Organic Chemistry Division

CSIR-National Chemical Laboratory

Pune-411008, India

DECLARATION

I, Prabhakar Lal Srivastava, hereby declare that the work incorporated in the thesis entitled "**Characterization of the Genes Involved in Santalene Biosynthetic Pathway in Indian Sandalwood *Santalum album* Linn**" submitted by me to **University of Pune** for the degree of **Doctor of Philosophy** is original and has not been submitted to this or any other University or Institution for the award of Degree or Diploma. Such material, as has been obtained from other sources, has been duly acknowledged.

Date:

Place: Pune

Prabhakar Lal Srivastava

Dedicated to my beloved father…….

Acknowledgement

It gives me immense pleasure to look over my Ph.D. journey and remember all who have helped and supported me enroute. First and foremost, I would like to express my sincere and heartfelt gratitude towards my research supervisor, Dr. H. V. Thulasiram, for his invaluable advice, constructive support and his extensive discussions around my work, without which it would have been improbable for me to have completed my thesis. I am extremely grateful to him to have introduced me to the area of terpene biosynthesis and to have given me the opportunity to explore various dimensions.

I take this opportunity to sincerely acknowledge the Council of Scientific and Industrial Research (CSIR), New Delhi, for providing with the Junior and Senior Research Fellowships which facilitated me to carry out my work. I would like to thank The Director, National Chemical Laboratory (NCL), Pune and HOD, Division of Organic Chemistry, for infrastructure and support to accomplish my research work.

It is probably the best time to express my obligations towards my M.Sc. professors, Prof. J. K. Pal, Prof. W. N. Gade and Retd. Prof. V. Sitaraman, for kindling my scientific interest and encouraging me to explore various dimensions.

I am highly indebted to Late Dr. M. I. Khan for sowing scientific seeds during my M.Sc. dissertation under him at NCL and to Dr. Sanjay Nene for providing me the opportunity to start my career in research. I would also like to thank Dr. R. V. Gadre and Mr. V. V. Jogdand for their continuous support and motivation throughout my tenure under Dr. Nene.

I am extremely obliged to Dr. Jomon Joseph, Dr. Ashok P. Giri, Dr. Mahesh J. Kulkarni, Dr. Vidya Gupta, Dr. Dhiman Sarkar and Dr. Srinivas Hotha for their continuous evaluation of my Ph. D. work and their valuable advice for the same.

I would like to extend my gratitude to Dr. S. Dhanashekaran, Dr. M. N. Jagdeesha, Dr. Suresh Kumar, Dr. Nilanjana Sengupta and Dr. Karthikeyan for their support and suggestions.

There is a reason we don't thank friends, because it is too small a word in return for the unconditional love, support and encouragement that they share with you during your

best and worst times. Nevertheless, I would like to acknowledge my special friends, Gyan Prakash, Yashwant Kumar, Atul Anand, Jitendra Gupta, Harekrishna Punjal, Pradeep Soni, Sandeep Golegaonkar, Ravi Kalunke, Shamim Akhtar, Abhishek Mishra, Devdutta Dandekar and Shankar Suman.

"A journey is easier when you travel together." Words fail me when I express my gratitude to Krithika Ramakrishnan, Soniya Mantri, Pankaj Daramwar, Rincy Sanjay, Sharanbasappa Gangashetty, Nilofer Jahan Attar, Shiva Shankar S., Avinash Pandreka Rakesh Joshi and Krunal Patel for their efforts in completing critical parts of my work, which would have been impossible without their help.

I would like to extend a huge thanks to my labmates, Pankaj, Swati, Atul, Saikat, Devdutta, Dipesh, Harshal, Namdev, Sudhir, Deepak, Rincy, Deepti, Nilofer, Trushna, Ashwini, Krithika, Sreekant, Prasad, Sharan, Imran, Rohil, Shiva, Soniya, Priya, Avinash, Vijayshree, Uttara, Rani, Anurag, Montu, Ramesha, Aarthy, Supriya, Anirban, Manoj, Balaji, Fayaz, Pruthviraj, Nilesh, Ajay, Vijay, Yashpal, Rahul and Khandekar Uncle for providing a healthy and pleasant environment for work.

My warm thanks to my room-mate, Chaitanya Kiran, for bearing with me for 5 years and extending brotherly love towards me. I would like to give a special mention to some of my friends at NCL, namely, Bhushan Sir, Suresh, Arvind, Kapil, Yogesh, Rishi, Mudit, Sanjay, Kaushal, Anand, Hemender, Anuj, Brijesh, Manoj, Prathit, Achintya, Manzoor, Mrityunjoy, Wahid, Prakash, Datta, Rupesh, Shweta, Arati, Rubina, Reema, Yugendra, Nishant, Jagdeesha, Tejal, Rupa and Vishwanath.

I am indebted to my friends since graduation, Neeraj, Nitesh, Manish, Santosh, Kishore, Saif, Amol, Vivek, Kumar, Luxman, Prashant, Pankaj Verma, Ranjeet, Manbodh and Diwakar Mishra for their constant support.

"The love of a family is life's greatest blessing." I would like to pay my highest regards to my family for their love, care, support, understanding and sacrifices: my late father, Dr. Dinesh Chandra Srivastava, my pillar of strength, my mother, Mrs. Saraswati Srivastava, my elder brother, Mr. Diwakar, who has always stood by me, my beloved sisters, Suman and Neelam, for never losing their faith in me, my younger brother, Sudhakar, who has been immensely supportive of me, my sister-in-law, Mrs. Anita, for

her care and affection and my little nephew, Arush, for inspiring me to go on. I would also like to thank my Uncles, Mr. Suresh Chandra Srivastava and Mr. Ramesh Chandra Srivastava, my aunties Mrs. Kalawati Srivastava and Mrs. Aparna Srivastava and my cousins Mr. Shat Rudra Srivastava and Mr. Shatbir Srivastava, Brajesh, Abhishek and Abhijeet for their support and encouragement.

Last, but not least, I would like to thank God for the strength and courage He has bestowed upon me in every phase of my life.

--Prabhakar Lal Srivastava

Contents

List of tables

List of figures

Abbreviations

Å	Angstrom
Amp	Ampicillin
AMV-RT	Avian Myeloblastosis Virus-Reverse Transcriptase
AA	Amino acid
BLAST	Basic Local Alignment Search Tool
bp	Base pair
Cam	Chloramphenicol
cDNA	Complementary DNA
CTAB	Cetyl Trimethyl Ammonium Bromide
C-terminal	Carboxy terminal
DEPC	Diethylpyrocarbonate
DTT	Dithiothritol
EDTA	Ethylene diamine tetra acetic acid disodium salt
FAD	Flavin adenine dinucleotide
IPTG	Isopropyl β-D-1-thiogalactopyranoside
Kan	Kanamycin
kDa	Kilo dalton
KEGG	Kyoto Encyclopedia of Genes and Genomes
LA	Luria agar
LB	Luria broth
mg	Milligram
mL	Millilitre
µg	Microgram
µL	Microlitre
µM	Micromolar
NADP	Nicotinamide adenine dinucleotide phosphate
NADPH	Nicotinamide adenine dinucleotide phosphate reduced
NCBI	National Center for Biotechnology Information
ng	Nanogram
N-terminal	Amino terminal

OD	Optical density
ORF	Open Reading Frame
PCR	Polymerase Chain Reaction
PDB	Protein Data Bank
pI	Isoelectric point
PMSF	Phenyl methyl sulphonyl fluoride
Pfam	Protein families
PVPP	Polyvinylpolypyrrolidone
RACE	Rapid amplification of cDNA ends
RNA	Ribonucleic acid
Rt	Retention time
Sa	*Santalum album*
Sec	Second
SDS-PAGE	Sodium dodecyl sulfate polyacrylamide gel electrophoresis
TB	Terrific broth
UTR	Untranslated Region
X-gal	5-bromo-4-chloro-3-indolyl β-D-galactoside

Thesis abstract

Sandalwood oil is one of the oldest, most valuable and highly prized fragrances. The essential oil of Indian Sandalwood is dominated by a blend of sesquiterpenoids such as: α, β-Santalenes, α, β-Santalols, α-Bergamotene, α-Bergamotol, Bisabolol and Farnesol, of which, α- and β-santalols comprise >80 %. This thesis describes the characterization of genes involved in santalene biosynthetic pathway in Indian Sandalwood *Santalum album* L. We cloned and characterized a prenyl transferase (***SaFDS***) and four terpene synthases (***SaSS***, ***Saβ-BS***, ***SaTPS1*** and ***SaTPS2***) from the interface of heartwood and sapwood. Enzymatic incubation of *SaSS* with (*E,E*)-FPP resulted in the formation of a complex mixture of six sesquiterpenes, whereas *Saβ-BS*, *SaTPS1* and *SaTPS2* gave exclusively β-bisabolene and sesquisabinene B, respectively, on incubation with (*E,E*)-FPP. Evolutionary origin of santalene biogenesis and structural basis for the product specificity of santalene synthase was determined using homology based structural modeling and site directed mutagenesis. We have also cloned one CYP450 system and one CYP450 reductase, which may be involved in the hydroxylation of the santalene mixture for the formation of santalols.

Chapter 1: Introduction

Isoprenoid are largest group of secondary metabolites and present in all living systems. Over 55,000 individual structures, containing a truly incredible array of carbon skeletons and functional groups have been reported[1,2]. All isoprenoid compounds are constructed from two simple five-carbon building blocks,

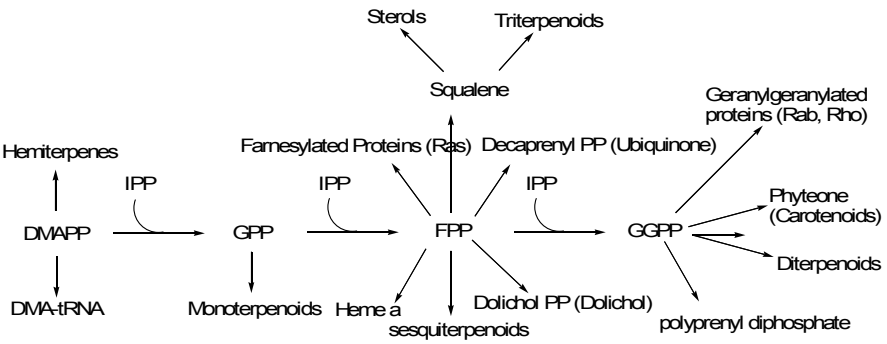


Figure1.1: Basic pathway for biosynthesis of various isoprenoids

isopentenyl diphosphate (IPP) and dimethylallyl diphosphate (DMAPP) and are classified by the number of five carbon units present in the core structure (Figure 1.1).

Plants continue to be the exemplary source of isoprenoid compounds with a vast array of chemical structures ranging from universal primary metabolites, such as sterols, carotenoids, ubiquinones, and hormones, to more unique and species-specific secondary metabolites which may be involved in plant defence and communication[3]. India has a vast and inexhaustible resource of medicinal and aromatic plants. Over 45,000 plant species are known in India, of which about six thousand are known to possess medicinal and organoleptic properties and very few plants have been explored for the natural products which possess medicinal properties[4]. About 1000 medicinal plant species face a threat to their existence, such as the most endangered Indian sandalwood (*Santalum album*, Linn).

Sandalwood or Chandan (*Santalum*. *album*) belongs to the *Santalaceae* family, a medium-sized evergreen hemi root parasitic tree, highly valued for its fragrant heartwood, which contains the essential oil known as sandalwood oil that is used in perfumes, cosmetics, aromatherapy, incense sticks industries[5] and has been reported to have various biological properties such as antiviral[5], anticarcinogenic[6], and antitumor[5,7,8] effects. Among the reported constituents of sandalwood oil are sesquiterpenoids[9-11], triterpenoids[11], and phenylpropanoids[11]. Over 90 % of the world's sandalwood (*S. album*) production is from India. α- and β-santalols (hydroxylated products of α- and β-santalenes), which are the major components (~80 %) of the essential oil from a well-matured tree (≥80 years old), are responsible for most of the biological activity. The prime-most enzyme of the terpene biosynthesis in sandalwood, leading to the formation of santalene derivatives, is santalene synthase. This enzyme catalyzes a complex, multi-step reaction capable of generating a mixture of six different sesquiterpenes (α-santalene, α-bergamotene, epi β-santalene, β-santalene, β-bergamotene and (*E*)-β-farnesene), with varied carbon skeletons from a single diphosphate substrate, farnesyl diphosphate (*E,E*-FPP)[12].

Chapter 2: RNA isolation and transcriptome sequencing for screening of genes involved in santalene biosynthesis

Isolation of intact and functional RNA from interface of heartwood and sapwood from Sandalwood is very challenging because it contains high percentage of phenolics and secondary metabolites. Moreover, phenolic compounds are found in a wide range of polymerization states, including lignans, stilbenes, flavonoids and quinones, with the latter two associated with the colour change occurring during heartwood formation. There are several protocols[13-15] available for the isolation of RNA from tissues rich in phenolic and polysaccharide compounds but most of them are tissue specific. The presence of phenolics and polysaccharides complicate RNA isolation from the interface of heartwood and sapwood because they co-precipitate with nucleic acids giving the pellet a brownish colour. Also, the solubility of the pellet decreases immensely. In this study, a protocol is developed for the isolation of high quality of total RNA from the interface of heartwood and sapwood from *S. album*.

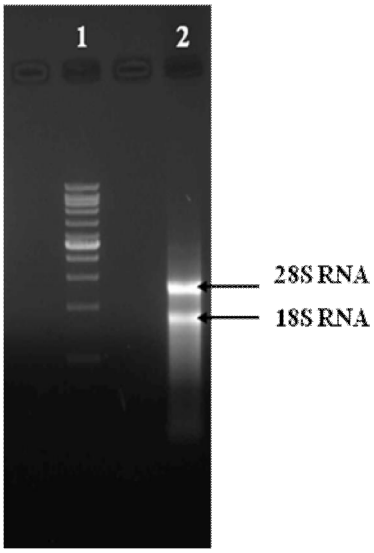


Figure 2.1: Agarose gel electrophoresis of RNA isolated from modified protocol, **Lane 1:** 1Kb DNA ladder, **Lane 2:** Total RNA, $A_{260/280}$ ratio = 2.14, $A_{260/230}$ ratio = 2.35

This high quality RNA was further used for transcriptome sequencing, a quick and cost effective method to profile complete coding sequence of a genome for the screening of genes involved in the biosynthesis of santalenes and their derivatives. From

transcriptome sequencing and annotation, we screened 18 transcripts having the terpene synthase domain, 72 transcripts having CYP450 domain whereas 3 transcripts for CYP450 reductase. Some of the selected transcripts were screened for the identification of genes involved in santalene biosynthesis.

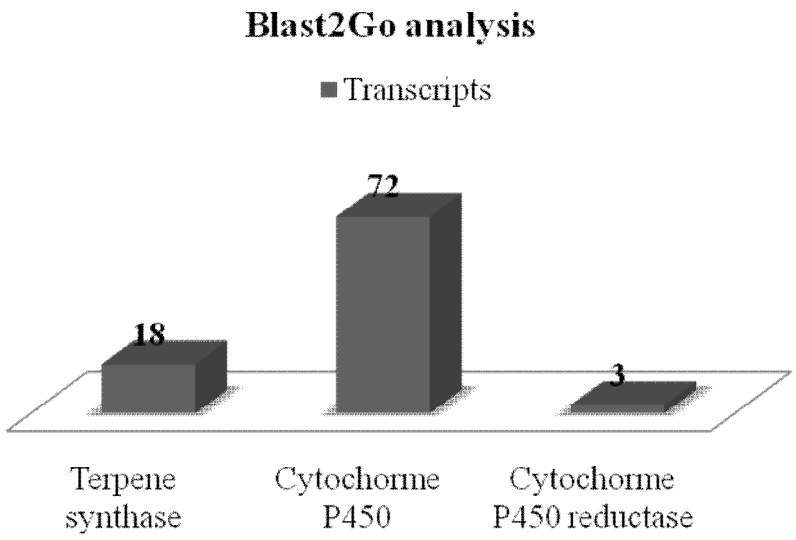


Figure 2.2: BLAST2GO analysis of unigenes assigned with Pfam ID for the screening of terpenoid biosynthetic pathway genes

Chapter 3: Screening, cloning, expression and characterization of terpene synthases involved in biosynthesis of santalene derivatives

---

To understand the molecular and biochemical mechanism controlling terpene diversity in Indian Sandalwood, we have isolated and characterized five terpene synthases **(*SaFDS, SaSS*, *SaBBS*, *SaTPS1* and *SaTPS2*)** involved in the biosynthesis of farnesyl diphosphate (FPP), santalenes, bergamotenes, β-bisabolene and sesquisabinene B, respectively. There was not much information available on how many enzymes are involved in the formation of the sesquiterpene mixture in Indian Sandalwood, *S album.* Therefore, to characterize the genes involved in santalene biosynthesis, we selected five transcripts on the basis of sequence comparison to existing sesquiterpene synthase sequences in NCBI database. A RACE was performed to generate the full-length cDNA

sequences of all these five genes. ORFs were isolated and cloned in expression vector for functional characterization. Protein expression was carried out in BL21DE3 (***SaFDS***)/Rosetta 2 DE3 (***SaSS, SaBBS, SaTPS1*** **and** ***SaTPS2***) cells and recombinant protein was purified to the homogeneity. Purified proteins were incubated with prenyl-diphosphate substrate for functional characterization.

When santalene synthase (*SaSS*) was incubated with (*E,E*)-FPP in presence of Mg^{2+}, it resulted in the formation of six sesquiterpenes of which three major sesquiterpenes (α-, β-santalenes and exo-α-bergamotene) and three minor sesquiterpenes (*epi*- β-santalene, (*E*)-β-farnesene and exo-β-bergamotene) indicating that *SaSS* is moderately promiscuous in nature. In Sandalwood, these Santalene derivatives will be further converted into their corresponding alcohols catalyzed by CYP450 mono-oxygenase systems[12,16].

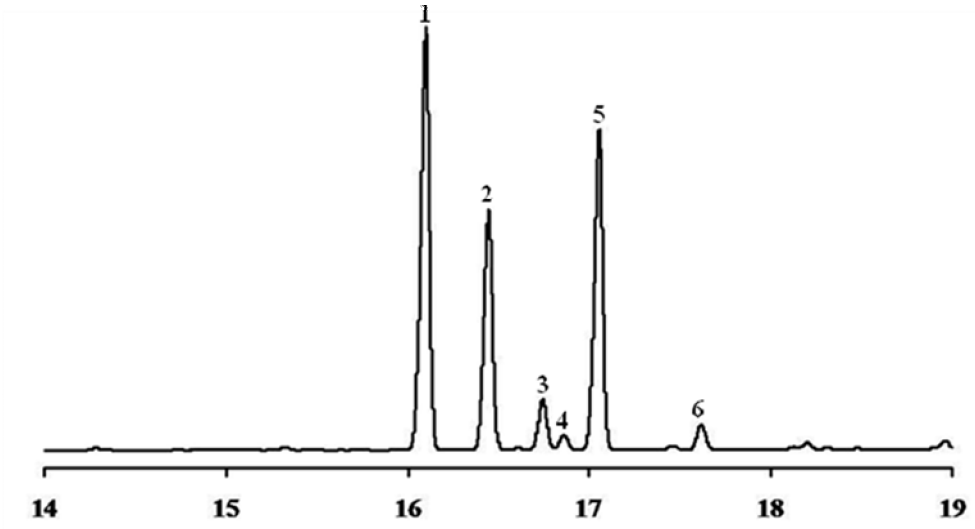


Figure 3.1: GC profile of *SaSS* assay with (*E,E*)-FPP, α- santalene (**1**), exo-α-bergamotene (**2**), *epi*- β-santalene (**3**), (*E*)-β-farnesene (**4**), β-santalene (**5**) and exo-β-bergamotene (**6**)

Incubation of β-bisabolene synthase (*SaβBS*) with *(E,E)*-FPP in the presence of Mg^{2+} resulted in the exclusive formation of β-bisabolene (1) and a small proportion of bisabolol (2) as shown in figure 3.2. In recent years, sesquiterpenes of farnesene and bisabolene skeletons have been recognized as replacements for petroleum-derived jet-engine fuel[17,18].

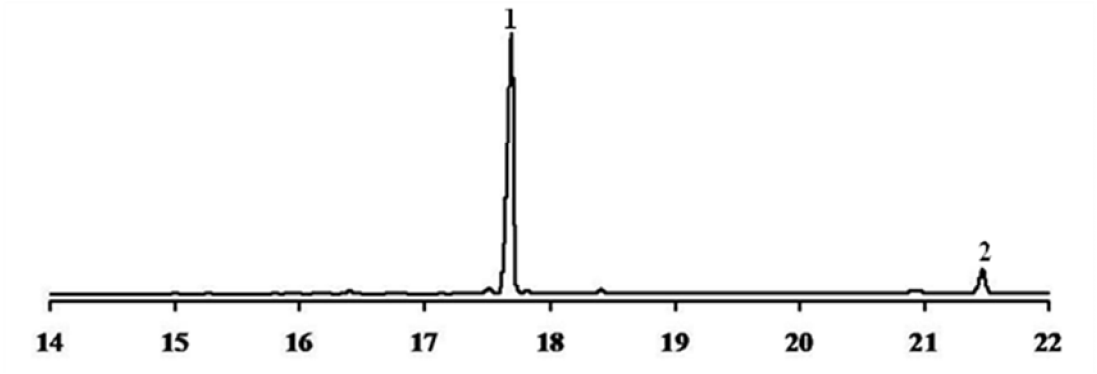


Figure 3.2: *SaβBS* assay profile with (*E,E*)-FPP, β-bisabolene (**1**), bisabolol (**2**)

Two isoforms of sesquiterpene synthases (*SaTPS1* and *SaTPS2*) involved in the biosynthesis of sesquisabinene B in Indian Sandalwood, *S. album* have been isolated and functionally characterized. Sesquisabinene B and its hydroxyl derivative sesquisabinene hydrate are reported to be present in several plant species[19,20], but till today no report is available on the isolation and characterization of sesquisabinene B synthase. This is the first report on the isolation and functional characterization of sesquisabinene B synthase. Functional characterization of *SaTPS1* and *SaTPS2* revealed that both the enzymes catalyze the exclusive formation of sesquisabinene B from (*E,E*)-FPP as substrate with varying kinetic parameters (Figure 3.3). We also tested the *in vivo* production of *SaTPS1* and *SaTPS2* metabolites by incorporating these genes in pET-Duet vector along with *SaFDS* and expressing the same in C41DE3 cells containing pRARE plasmid for rare codons. We observed that sesquisabinene B could be produced in 1-2 mg/L of non-optimized bacterial culture and demonstrated the feasibility of metabolic engineering of *SaTPS1* in a heterologous system for the large-scale production of sesquisabinene B.

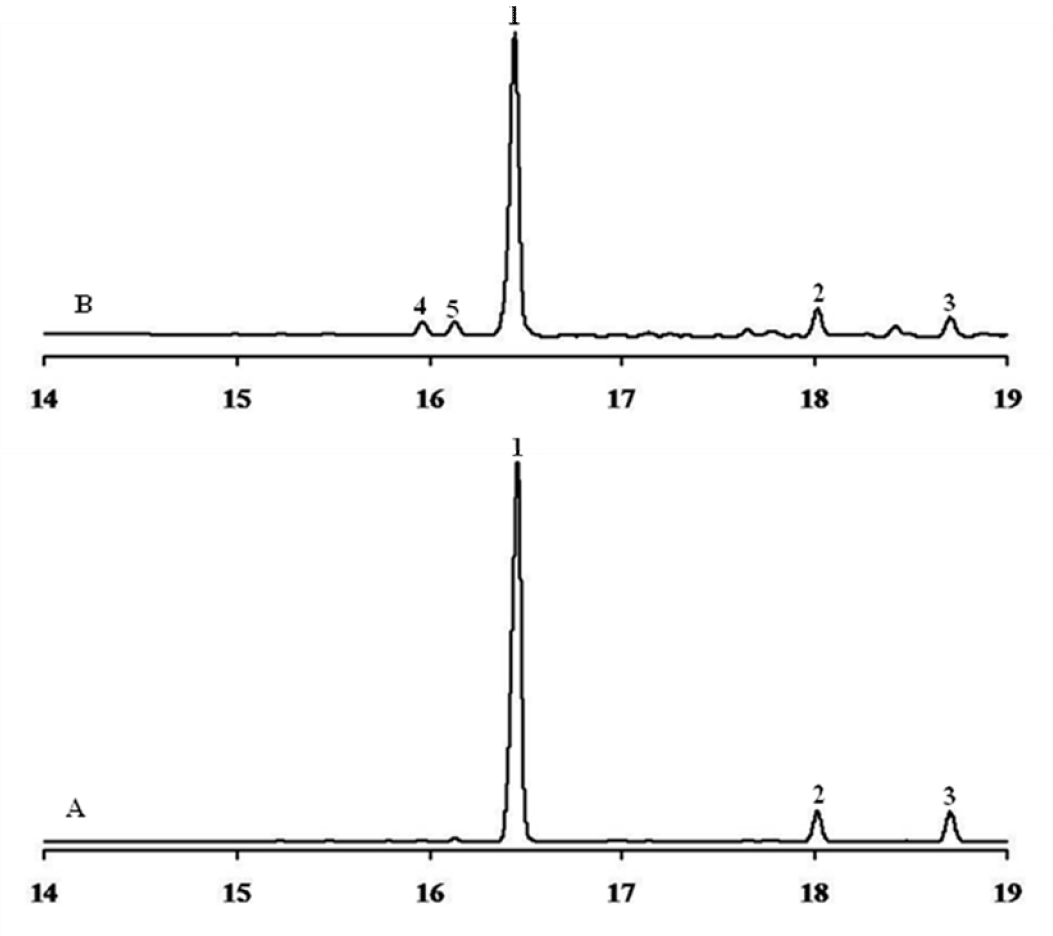


Figure 3.3: A) *SaTPS1* assay profile with (*E,E*)-FPP, Sesquisabinene B (**1**), **B)** *SaTPS2* assay profile with (*E,E*)-FPP, Sesquisabinene B (**1**)

Chapter 4: Santalene synthase: A comprehensive site directed mutagenesis study to understand the dynamic nature of the sesquiterpene synthase active site

---

In this chapter we describe the homology structural model based engineering of santalene synthase (*SaSS*), which was isolated and cloned from the transition zone of the sapwood-heartwood of the *Santalum album* Linn (sandalwood tree). Santalene synthase catalyzes the cyclization of acyclic farnesyl diphosphate (FPP) into α-, β-santalenes, and α-exo-bergamotene as major products along with a small amount of epi-β-santalene, (*E*)-β-

farnesene and exo-β-bergamotene. In order to engineer *SaSS* for the selective production of these products or few very different sesquiterpenes, the residues involved in stabilizing the intermediates in carbocation cascade for the formation of sesquiterepens from FPP were identified using homology structural model built using the crystal structure of 5-*epi*-aristolochene synthase. Homology model based directed mutagenesis was used to modify the selected amino acids in the active site of the santalene synthase (*SaSS*) to better understand sesquiterpene biosynthesis and molecular origins of chemical diversity in *S. album*. This study also provides insights into the evolutionary aspects of various terpene synthases even though there is a very low level of sequence identity among them. In this chapter, 22 amino acids were selected including all common residues present in 5-*epi*-aristolochene synthase and 35 mutants were constructed to understand the biosynthesis of santalenes and bergamotenes. These surrounding amino acids are thought to be involved in maintaining electronic density and correct orientation of carbocation for cyclization in the active site pocket. The aspartate and arginine rich motifs which are highly conserved in all sesquiterpene synthases, were not considered because studies have already proved that such alterations would result in significant loss of activity[21,22]. The product ratio analyses using (*E,E*)-FPP as substrate indicated that mutants, S290G, T318A, R474M, R474L, T467A, W293F, Y37F, Y396F and Y539F produces exclusively α-exo-bergamotene where as mutant Y539W produces endo-bergamotene and exo-β-bergamotene in significant amounts. Mutant L427A increases the active site volume which inhibits the cyclization of highly unstable carbocation which leads to the formation of (*E*)-β-farnesene in a significant amount. Mutants I422A produced a very different sesquiterpene β-curcumene as a major product by stabilizing the bisabolyl carbocation. We successfully constructed the sesquiterpene synthases such as: α-santalene and exo-α-bergamotene synthase (N463D), exo-α-bergamotene synthase (S290G, W293F, T318A, Y539F, R474M/L, and Y539S), endo- α-bergamotene synthase (Y539W), β-curcumene synthase (I422A), (*E*)-β-farnesene synthase (L427A) producing a single or multiple product derived from the predominant reaction pathway of *SaSS* by maintaining the kinetic parameters (Figure 4.4).

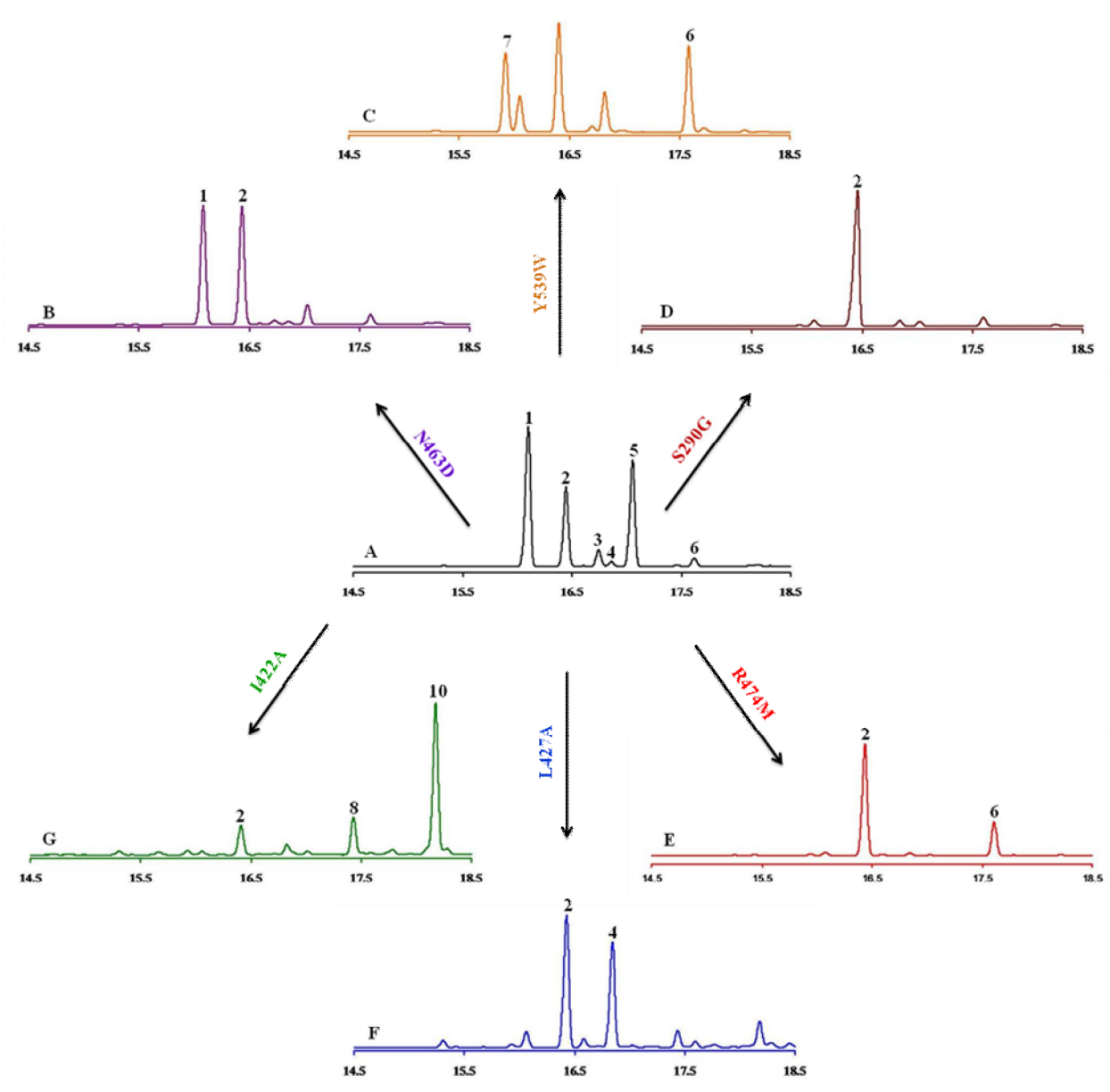


Figure 4.4: GC profile of novel sesquiterpene synthases with (*E,E*)-FPP derived from *S. album* santalene synthase by directed mutagenesis, **A)** Santalene synthase assay profile with (*E,E*)-FPP, **B)** Mutant N463D (α-santalene and exo- α-bergamotene synthase), **C)** Mutant Y539W (endo- α- bergamotene synthase), **D)** Mutant S290G (exo-α-bergamotene synthase), **E)** Mutant R474M (exo-α-bergamotene and exo-β-bergamotene synthase), **F)** Mutant L427A (*E*)-β-farnesene synthase), **G)** Mutant I422A (β-curcumene synthase), α-santalene (**1**), exo-α-bergamotene (**2**), epi-β-santalene (**3**), (*E*)-β-farnesene (**4**), β-santalene (**5**), exo-β-bergamotene (**6**), endo-α-bergamotene (**7**), γ-curcumene (**8**), β-curcumene (**10**)

Chapter 5: Screening, isolation and cloning of Cytochrome P450 systems from Indian Sandalwood *S. album*

α- and β-santalols (hydroxy derivative of α- and β-santalenes), which are the major components (~80 %) of sandalwood oil, are responsible for most of the biological activity related to sandalwood oil. Sandalwood oil is commercially produced by steam distillation of heartwood chip with varying yield from 4-7 % from a well-matured tree. Since the demand for this essential oil is increasing, overharvesting and several diseases have led to a serious decline in the population of Sandalwood tree. Alternate approach for the production of these essential components of sandalwood oil is engineering a microbial system by incorporating the genes involved in the biosynthesis of santalols. From the 10 transcripts showing similarity to the CYP domains, we selected one transcript matching with Premnaspirodiene oxygenase (>Locus_41531_Transcript_1/1_Confidence_1.000_Length_749 bp), which may be involved in the hydroxylation of santalenes. Transcript matching with Premnaspirodiene oxygenase was used for extending the partial transcript sequence to generate full-length cDNA sequence by 5′ RACE. Full length ORF sequence of SaCYP450 mono-oxygenase was generated and cloned in pESC yeast duet vector MCS2 (containing CYP450 reductase from *Azadirachta indica*). Yeast expression and microsome preparation was carried out and this microsome was used in *in vitro* assay with santalenes' mixture, but could not result in hydroxylation of the substrate.

References

(1) Thulasiram, H. V.; Erickson, H. K.; Poulter, C. D. *Science* **2007**, *316*, 73-76.

(2) Reiling, K. K.; Yoshikuni, Y.; Martin, V. J. J.; Newman, J.; Bohlmann, J.; Keasling, J. D. *Biotechnol. Bioeng.* **2004**, *87*, 200-212.

(3) Lange, B. M.; Rujan, T.; Martin, W.; Croteau, R. *P. Natl. Acad. Sci. USA* **2000**, *97*, 13172-13177.

(4) Grover, J. K.; Yadav, S.; Vats, V. *J. Ethnopharmacol.* **2002**, *81*, 81-100.

(5) Kaur, M.; Agarwal, C.; Singh, R. P.; Guan, X.; Dwivedi, C.; Agarwal, R. *Carcinogenesis* **2005**, *26*, 369-380.

(6) Kim, T. H.; Ito, H.; Hayashi, K.; Hasegawa, T.; Machiguchi, T.; Yoshida, T. *Chem. Pharm. Bull.* **2005**, *53*, 641-644.

(7) Okugawa, H.; Ueda, R.; Matsumoto, K.; Kawanishi, K.; Kato, A. *Phytomedicine* **1995**, *2*, 119-126.

(8) Hongratanaworakit, T.; Heuberger, E.; Buchbauer, G. *Planta Medica* **2004**, *70*, 3-7.

(9) Ochi, T.; Shibata, H.; Higuti, T.; Kodama, K.-h.; Kusumi, T.; Takaishi, Y. *J. Nat. Prod.* **2005**, *68*, 819-824.

(10) Prema, B. R.; Bhattacharyya, P. K. *Appl. Microbiol.* **1962**, *10*, 529-531.

(11) Ishida, T. *Chem. Biodivers.* **2005**, *2*, 569-590.

(12) Jones, C. G.; Moniodis, J.; Zulak, K. G.; Scaffidi, A.; Plummer, J. A.; Ghisalberti, E. L.; Barbour, E. L.; Bohlmann, J. r. *J. Biol. Chem.* **2011**, *286*, 17445-17454.

(13) Pandit, S.; Mitra, S.; Giri, A.; Gupta, V. *J. Plant Biol.* **2007**, *50*, 60-64.

(14) Kolosova, N.; Miller, B.; Ralph, S.; Ellis, B. E.; Douglas, C.; Ritland, K.; Bohlmann, J. *Biotechniques* **2004**, *36*, 821-824.

(15) Salzman, R. A.; Fujita, T.; Zhu-Salzman, K.; Hasegawa, P. M.; Bressan, R. A. *Plant Mol. Biol. Rep.* **1999**, *17*, 11-17.

(16) Daramwar, P. P.; Srivastava, P. L.; Priyadarshini, B.; Thulasiram, H. V. *Analyst* **2012**, *137*, 4564-4570.

(17) Rude, M. A.; Schirmer, A. *Curr. Opin. Microbiol.* **2009**, *12*, 274-281.

(18) Peralta-Yahya, P. P.; Keasling, J. D. *Biotechnol. J.* **2011**, *5*, 147-162.

(19) Radulovic, N. S.; DordeviC, N. D.; Zlatkovic, B. K.; Palic, R. M. *Chem. Pap.* **2008**, *62*, 603-607.

(20) Sonboli, A.; Bahadori, M. B.; Dehghan, H.; Aarabi, L.; Savehdroudi, P.; Nekuei, M.; Pournaghi, N.; Mirzania, F. *Chem. Biodivers.* **2013**, *10*, 687-694.

(21) Starks, C. M.; Back, K.; Chappell, J.; Noel, J. P. *Science* **1997**, *277*, 1815-1820.

(22) Greenhagen, B. T.; Oâ€™Maille, P. E.; Noel, J. P.; Chappell, J. *P. Natl. Acad. Sci. USA* **2006**, *103*, 9826-9831.

Chapter 1

Introduction

1.1 Plant secondary metabolites and their biosynthesis

Plants have evolved specialized secondary biosynthetic pathways for the synthesis of structurally and functionally complex small molecules, which aid in the growth and development of the plants but are not required for the survival of the plants. In plants, the pattern of secondary metabolites is complex as it changes in a tissue- and/or organ-specific manner. Secondary metabolites can be present in an active state or as a prodrug which are activated by wounding and infection. Secondary metabolites play a key role in the protection of plants against microbial infections (fungal, bacterial), viral infections, herbivory, UV radiation, attraction of pollinators and frugivores, allelopathy, and signalling[1]. The abundance of secondary metabolites in plants is often less than 1 % of the total storage in a dedicated tissue. The ability to synthesize secondary metabolites has been selected throughout the course of evolution in different plant lineages. Plant genomes are estimated to contain 20,000-60,000 genes out of which 15-25 % of genes encode for the enzymes involved in secondary metabolism[2,3]. Since ancient times, secondary metabolites have been used for several purposes, such as flavours, fragrances, dyes, stimulants, insecticides, hallucinogens as well as therapeutic agents.

Secondary metabolites encompass three major classes of products including flavonoids, isoprenoids, and alkaloids derived from primary metabolic pathway (fig 1.1). These metabolites are unique to plants and are essential for adaptation to diverse and inconstant surroundings. Flavonoids are derived from phenylalanine and acetyl-CoA and play important roles in defence, signalling, pathogenesis and symbiosis[4]. They are also known for providing pigments to flowers, fruits, leaves and seeds. Alkaloids are nitrogen containing compounds found in about 20 % of plant species and play an important role in the defence of plants against herbivores and pathogens[5]. Isoprenoids represent the largest class of secondary metabolites and play diverse functional roles in plants as hormones, photosynthetic pigments, electron carriers, structural components of membrane, as well as an important role in communication and defence. The inter-relationship of biosynthetic pathway of secondary metabolites derived from primary pathway is shown in figure 1.1.

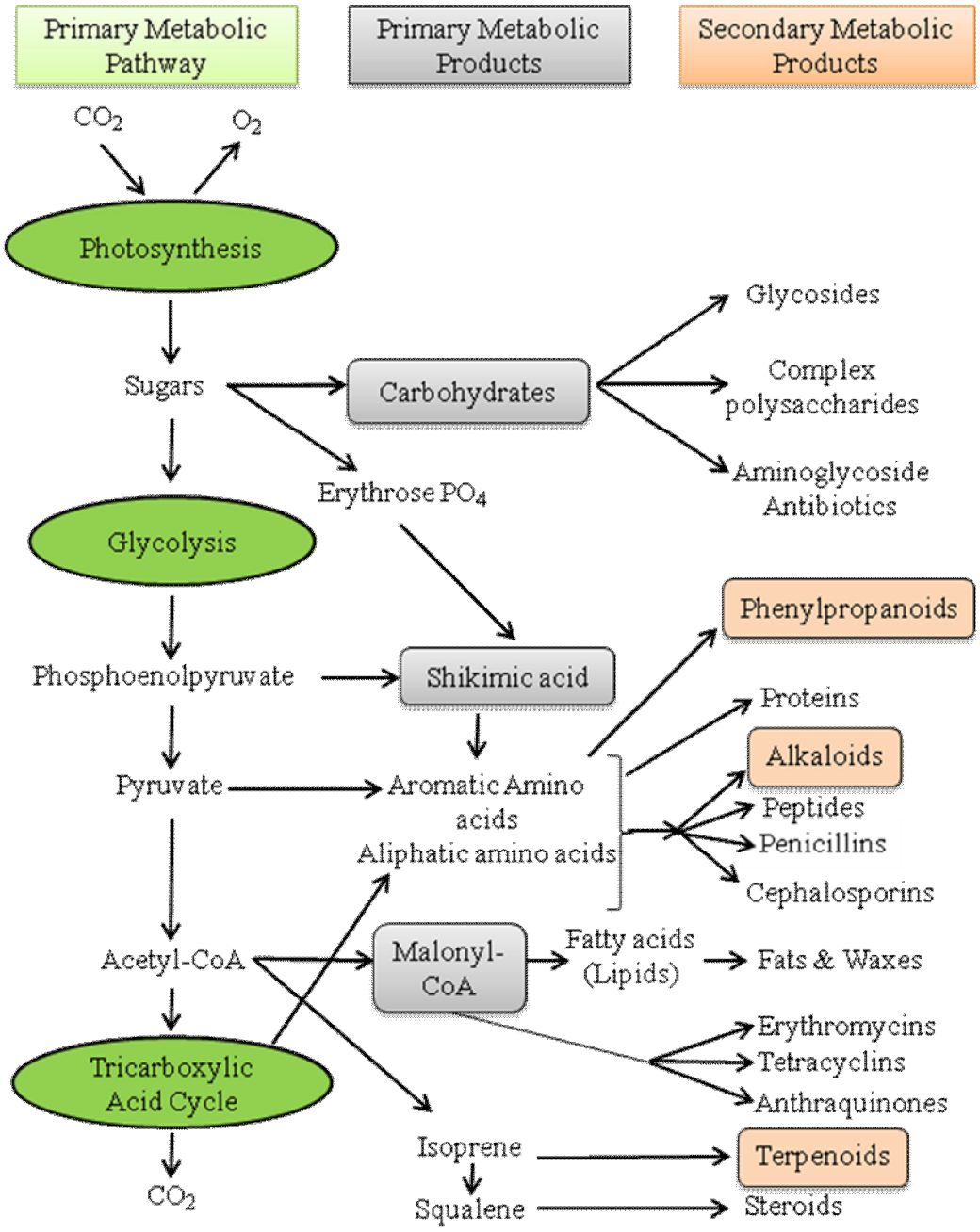


Figure 1.1: Inter-relationship of biosynthetic pathway leading to secondary metabolites in plants

The genes responsible for generating great diversity of secondary metabolites arise from gene duplication, origin of new genes for secondary metabolism, gain and loss of function for specific compounds, etc. The biosynthesis of plant secondary metabolites is usually restricted to a particular tissue and starts at specific developmental stages.

1.2 Isoprenoids

Isoprenoids are a large and diverse class of naturally occurring organic compounds present in all forms of living systems. All isoprenoid compounds are constructed from two simple five-carbon building blocks, isopentenyl diphosphate (IPP) and dimethylallyl diphosphate (DMAPP). Over 70,000 individual structures, containing a truly incredible array of carbon skeletons and functional groups have been reported[6,7]. Plants continue to be the exemplary source of isoprenoid compounds with a vast array of chemical structures ranging from universal primary metabolites such as sterols (as essential component of bio-membrane)[8,9], carotenoids and chlorophyll (as photosynthetic pigments)[10,11], ubiquinones (acting in electron transport chain)[12,13], vitamins and hormones[14,15], and to more unique and species-specific secondary metabolites which may be involved in plant defence and communication[16]. Terpene derivatives are also used for the production of numerous pharmaceuticals including taxol (anti-cancer drug)[17], artemisinin (anti-malerial drug)[18], etc. Biosynthesis of isoprenoids can be divided into four basic steps[19]: (i) production of the C5 monomers isopentenyl diphosphate (IPP) and dimethylallyl diphosphate (DMAPP), (ii) the head to tail condensation of C_5-unit isopentyl diphosphate with its isomer, dimethylallyl diphosphate into geranyl diphosphate (GDP), farnesyl diphosphate (FDP) and geranylgeranyl diphosphate (GGDP), (iii) cyclization and/or rearrangement to form the diverse class of terpene carbon backbone and (iv) tailoring and downstream functional group modifications of the terpene skeleton. Biosynthesis of isoprenoids and its compartmentalisation in plant cells is represented in figures 1.2A and 1.2B.

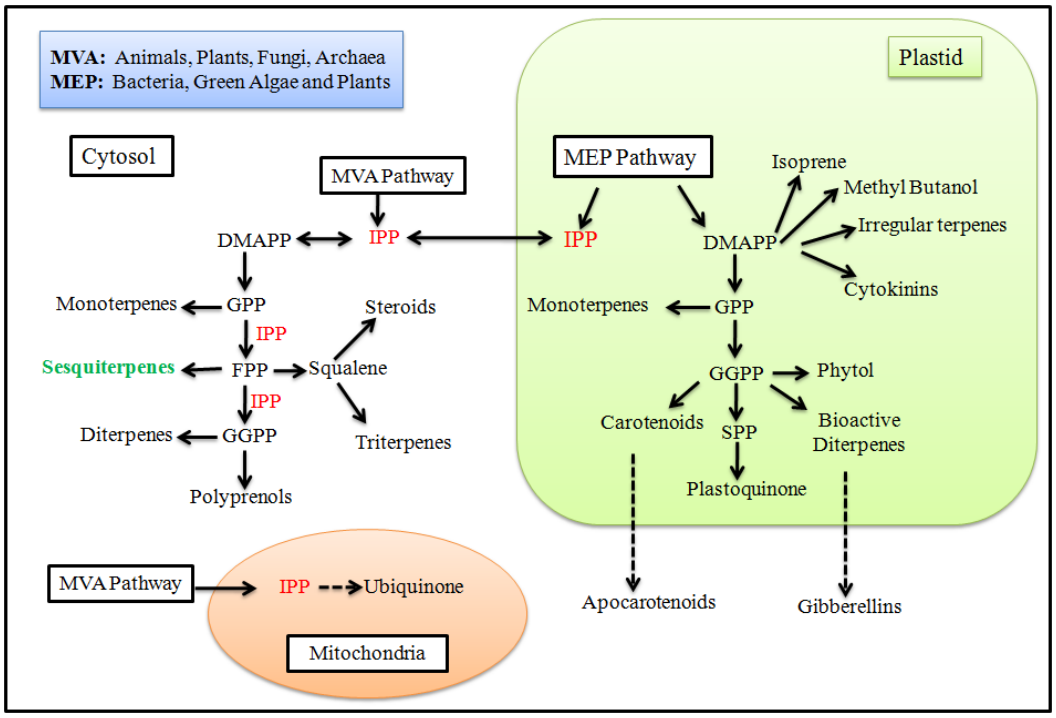


Figure 1.2A: Biosynthesis of various isoprenoids and its compartmentalisation in plant

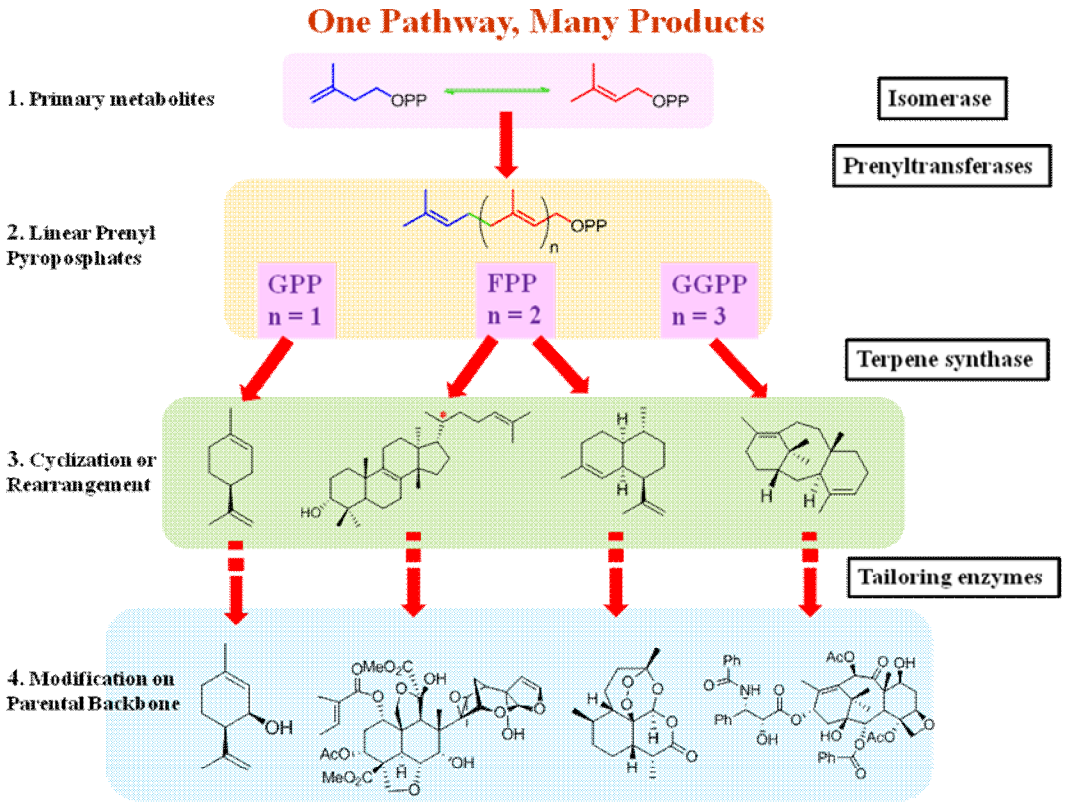


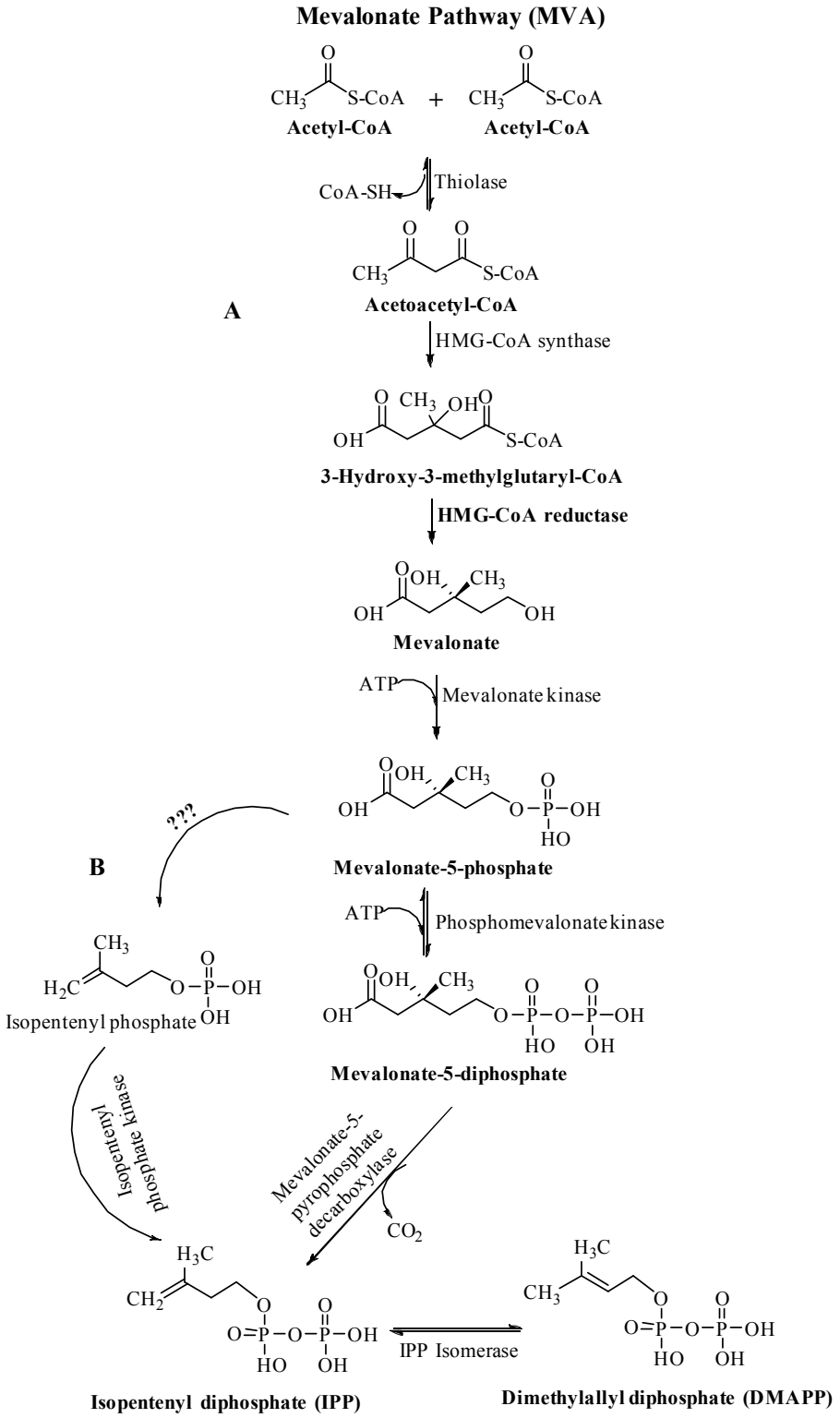
Figure 1.2B: Basic steps involved in isoprenoids biosynthesis

1.3 Biogenesis of IPP and DMAPP

1.3.1 Mevalonate (MVA) pathway

Mevalonate pathway (MVA) or HMG-CoA reductase pathway was first discovered in yeasts and animals in 1950, and it was thought to be responsible for the biosynthesis of IPP and DMAPP in all the organisms in six enzymatic steps starting from Acetyl-CoA[20,21], followed by IPP isomerase (IDI) which maintains the balance between IPP and DMAPP. This pathway occurs in cytosol and is present in animals, plants, fungi, archaea and some bacteria.

The first step of MVA pathway is the synthesis of acetoacetyl-CoA which occurs by condensation of two molecules of acetyl-Co-A, the reaction being catalyzed by acetoacetyl-CoA transferase. In the second step, acetoacetyl-CoA is converted into 3-hydroxy-3-methylglutaryl-CoA (HMG-CoA) catalyzed by HMG synthase. The subsequent step involves conversion of HMG-CoA to mevalonic acid (MVA) catalyzed by 3-hydroxy-3-methylglutaryl-CoA reductase in the presence of NADPH. Further, mevalonic acid (MVA) undergoes phosphorylation in two successive steps, catalyzed by MVA kinase and phospho-MVA kinase to synthesize MVA 5-diphosphate. In the last step of IPP biosynthesis MVA 5-diphosphate undergoes ATP-dependent decarboxylation catalyzed by diphospho-MVA decarboxylase. IPP undergoes isomerisation to form DMAPP in a reaction catalyzed by IPP isomerase (IDI). Two types of isopentenyl diphosphate isomerase (IDI) have been characterized based on their requirement of the cofactor. Type 1 IDI belongs to zinc metallo-proteins, catalyzing the inter-conversion of IPP and DMAPP, and is widely distributed in fungi, mammals and plants[22]. Type 2 IDI is restricted to archaea and bacteria which requires flavin mononucleotide (FMN) and NADH for its activity[23,24]. Recently an enzyme from *Methanocaldococcus jannaschii* was shown to catalyze phosphorylation of isopentenyl phosphate to isopentenyl pyrophosphate[25-27]. Based on this observation, a modified MVA pathway was proposed in which mevalonate-5-phosphate is decarboxylated into isopentenyl phosphate which further undergoes phosphorylation to form isopentenyl diphosphate[28].

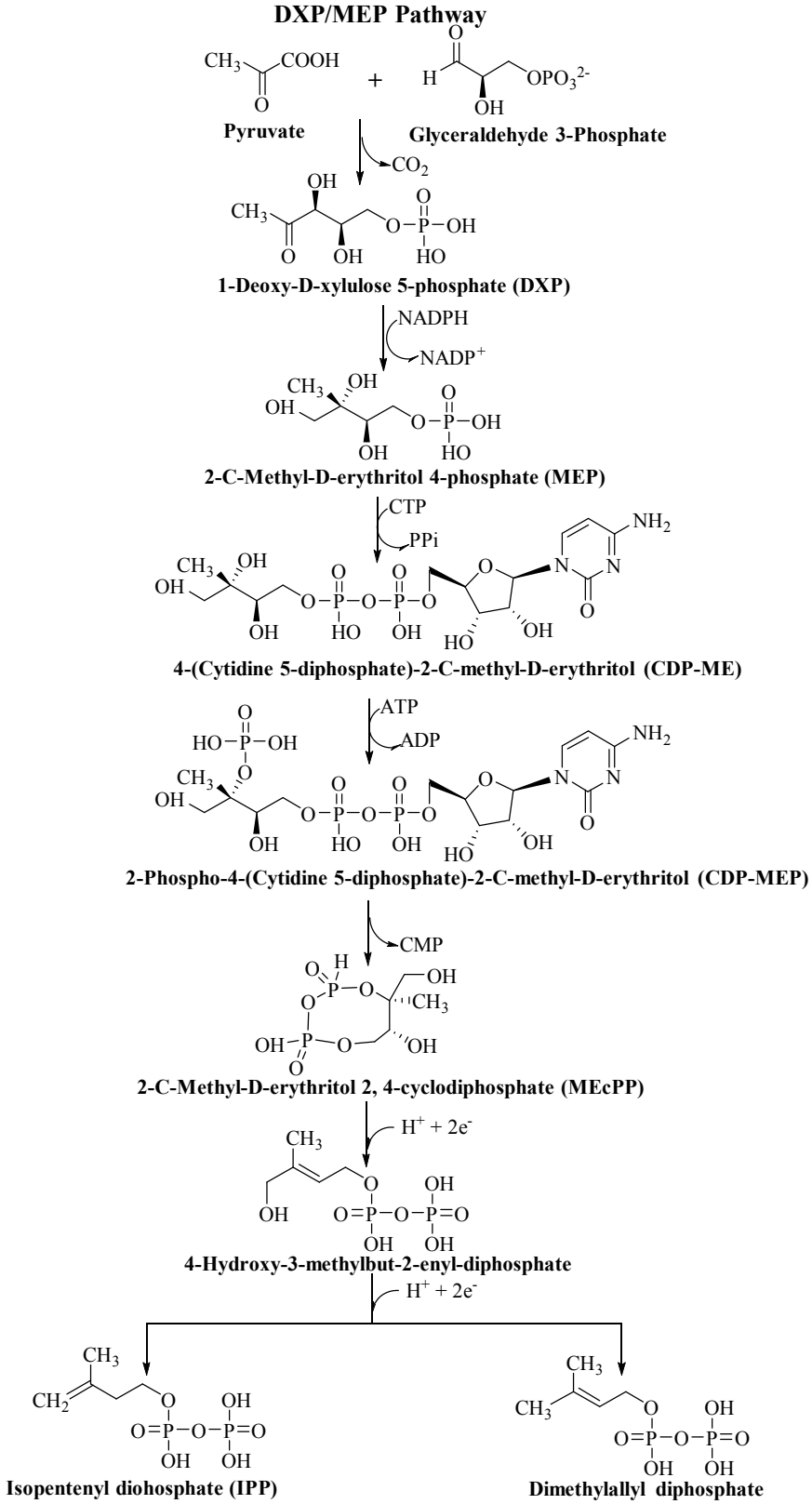


Scheme 1.3.1: A) Schematic representation of mevalonate pathway (MVA), **B)** Modified mevalonate pathway

1.3.2 Methyl Erythritol Pathway (MEP) / 1-deoxy-D-xylulose 5-phosphate (DXP) pathway

Earlier it was believed that MVA pathway is the sole pathway for the biosynthesis of IPP and DMAPP. In the early 1990s, Rohmer and co-workers proved the existence of an alternative pathway (the MEP or DXP pathway)[29-31] which is present in many eubacteria and plant plastids, but absent in humans[16]. Therefore MEP pathway could be an excellent target for developing broad spectrum antibiotics[32-34], as many human pathogenic bacteria rely on MEP pathway for the biosynthesis of their essential isoprenoids. In plants, carotenoids, chlorophylls and prenylquinones are synthesized through the MEP pathway operated in plastids. MEP pathway consists of seven enzymatic steps starting from glyceraldehyde-3-phosphate and pyruvic acid via 1-deoxy-D-xylulose-5-phosphate for the synthesis of IPP and DMAPP in a ratio of 5:1.

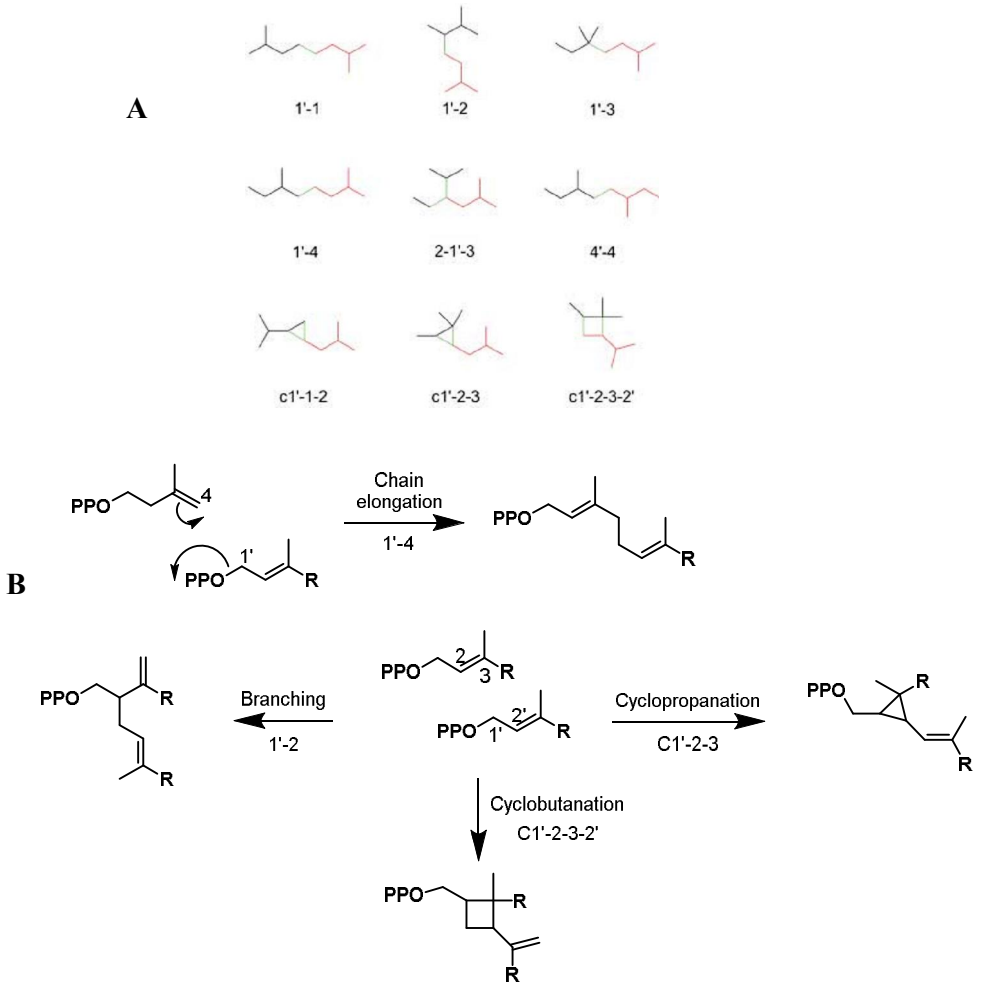
The first step of MEP pathway is the synthesis of DOXP (1-deoxy-D-xylulose-5-phosphate), which is catalyzed by DOXP synthase (DXS) in a transketolase-type decarboxylation from glyceraldehyde-3-phosphate and pyruvate. In the second step, DOXP is transformed into 2-C-methyl-D-erythritol-4-phosphate (MEP) catalyzed by the NADPH-dependent DOXP reductoisomerase. In the next step, MEP is converted into 4-(diphosphocytidyl)-2-C-methyl-D-erythritol (CDP-ME) in a CTP (cytidine triphosphate) dependent reaction catalyzed by 2-C-methyl-D-erythritol 4-phosphate cytidylyltransferase. The hydroxyl group in C2 position of CDP-ME in the next step undergoes phosphorylation and gets converted into 4-(diphosphocitidyl)-2-C-methyl-D-erythritol-2-phosphate (CDP-ME2P), which is catalyzed by CDP-ME kinase. CDP-ME2P is subsequently converted into a cyclic 2-C-methyl-D-erythritol 2,4 cyclodiphosphate (MEcPP) catalyzed by MECP synthase. In the last step of this pathway, MEcPP is converted into IPP and DMAPP via an intermediate step 1-hydroxy-2-methyl-2-(E)-butenyl-4-diphosphate (HMBPP) catalyzed by HMBPP synthase and HMBPP reductase, both of these steps require NADPH as a reducing agent. Unlike MVA pathway, IPP isomerase (IDI) is present in only minority of bacteria and seems to be non-essential in *E.coli*.



Scheme 1.3.2: Schematic representation of Methylerythritol phosphate (MEP) pathway

1.3.3 Terpene synthase chemistry

Terpene synthases catalyze biosynthesis of terpenoids with high regio- and stereo-chemical precision in a cascade of complex reactions involving highly reactive carbocationic intermediates that undergo a sequence of reactions such as cyclizations, alkylations, rearrangements, deprotonations and hydride shifts.[35-42] The hydrocarbon skeletons thus obtained are further added with functional groups, such as hydroxy group, by the action of upstream enzymes belonging to monooxygenase class.[43] The installed functional groups may subsequently undergo a series of functional group transformations to ultimately result in final terpenoid molecules. The carbon skeletons of isoprenoid compounds are constructed from 3-methyl-1-butyl units (isoprene), which are joined in one of the nine different patterns (Scheme 1.3.3).



Scheme 1.3.3 A) Patterns found in nature for connecting isoprene units, Colours indicate isoprene units (red and black) and bonds joining the isoprene units (green), **B)** Four basic coupling reactions in terpene biosynthesis[44].

Of the only four biosynthetic reactions comprising of chain elongation, cyclopropanation, branching and cyclobutanation are involved in joining of simpler units, whereas others involve rearrangements of 1'-2-3 structures (Scheme 1.3.3). Of these, chain elongation is the most common reaction, which involves head to tail condensation of IPP with allylic diphosphates such as DMAPP, GPP or FPP to form acyclic diphosphates[6] and are catalyzed by prenyl transferases such as geranyl diphosphate synthase (GDS), farnesyl diphosphate synthase (FDS) or geranylgeranyl diphoshate synthase (GGDS) respectively. Early structural investigations made by Wallach (1914) led to formulate a principle that, most terpenoids could be hypothetically constructed by a repetitive joining of isoprene units. This was the first integrated concept for a common structural relationship among the terpenoid natural products. Subsequently, Ruzicka refined the original concept to establish the 'biogenetic isoprene rule[45,46] which proposed that all terpenoids arose from an "active" or biogenic isoprene. This hypothesis ignores the specific character of the biological precursor and considers only that their structure is "isoprenoid" in nature. Active isoprene, which was identified as isopentenyl pyrophosphate (IPP), and its allylic ester isomer, dimethylallyl pyrophosphate (DMAPP) are now known to form the central intermediates in all isoprenoid biosynthetic pathways.

1.3.4 Major branch points and downstream products

In plants, isopentenyl diphosphate (IPP) and dimethylallyl diphosphate (DMAPP) synthesized from MVA pathway are used for the synthesis of isoprenoids in cytosol and mitochondria, whereas IPP and DMAPP synthesized from MEP/DOXP pathway are used in the synthesis of isoprenoids in the plastid. DMAPP acts as the primary active substrate for the synthesis of prenyl diphosphates: geranyl diphosphate (GPP), farnesyl diphosphate (FPP), and geranylgeranyl diphosphate (GGPP) by head to tail condensation with its isomer IPP. DMAPP also acts as a direct source for synthesis of cytokines and hemiterpenes whereas GPP, FPP and GGPP are the branch points for synthesis of all isoprenoids. GPP synthase, localized in plastids, produces GPP as a precursor of monoterpenes biosynthesis and functions as homomeric and heteromeric enzymes[47-49], FPP is synthesized mainly in cytosol and mitochondria by homodimeric enzyme (FPP synthase) and act as a main branching point for the biosynthesis of sterols, brassinosteroids, dolichols, polyprenols and for protein prenylation[50]. GGPP produced by GGPP synthase act as a precursor for the synthesis of diterpenoids,

gibberellins, chlorophylls, phylloquinone, plastoquinone, tocopherols, carotenoids, abscisic acid and oligoprenols. GGPP synthase functions as homodimeric enzyme localized in plastids, mitochondria, and endoplasmic reticulum and utilizes all three allylic prenyl diphosphates (DMAPP, GPP, and FPP)[51]. The ability of terpene synthases to convert single prenyl diphosphate substrate to diverse products of varying carbon skeleton is one of the most unique features of this class of enzymes.

1.3.5 Cross talk between cytosolic and plastidial pathway

In higher plants, both the pathways operate simultaneously with a compartmental segregation. Various feeding experiments with stable isotopes indicate that compartmentalisation of two biosynthetic pathways in plants which are leading to the formation of DMAPP/IPP is not absolute, because in several cases one metabolite can exchange between these two pathways[48,52]. Biosynthetic studies of diterpene ginkgolide using ^{13}C glucose indicated that three C5 units (IPP) are formed through MVA pathway and one C5 unit (IPP) from DXP pathway is being utilized to form ginkgolide. Similar studies in liverwort (*H. planus*) and hornwort (*A. punctatus*) shows FPP portion (end C15 unit) of phytol is derived from MVA pathway whereas, terminal IPP portion (C5 unit) is derived from the DXP pathway[53-55]. These results clearly indicate the potential exchange of isoprenoids between both MVA and DXP pathways which may play a crucial role in regulation of isoprenoid biosynthesis[56-58]. The presence of MVA and MEP pathway enzymes in different compartments of plants suggest a mixed biosynthetic origin. In *Arabidopsis thaliana*, GGPP synthase which synthesizes GGPP as a precursor for diterpenoid, carotenoids and chlorophylls was found to be present in endoplasmic reticulum and mitochondria in addition to plastids[51]. Sesquiterpenoid biosynthesis usually takes place in cytosol of higher plants, however, (*Z,Z*)-FDS and sesquiterpene synthases have been found in plastids of the tomato trichome[59].

1.3.6 Regulation of MVA and MEP pathway

Isoprenoid biosynthesis in all organisms is regulated at multiple steps. In yeast and mammalian cells, regulation occurs at the levels of transcription, translation, posttranslational and protein degradation[60], whereas in bacteria, regulation occurs mainly at transcriptional level[61]. The regulation of both the pathways (MVA and MEP pathway) in plants is more complex as compared to other organisms. In plants,

presence of both the pathways in different compartments of one cell, existence of isozymes for several enzymatic steps and wide range of environmental stimuli add to the complexity of regulation of isoprenoid biosynthesis.

Regulation of MVA and MEP pathways in plants occurs mainly at transcriptional level; transcription of genes encoding enzymes of MVA and MEP pathway are not tightly co-ordinated. Transcription of individual enzymes or isozymes can vary significantly in different tissues and developmental stages. MVA pathway genes are mainly expressed in roots, flowers and seeds, whereas MEP pathway genes are predominantly expressed in photosynthetic tissues[62]. Light plays a very crucial role in regulation of isoprenoid biosynthesis from MEP pathway. Most of the pigment protein complexes responsible for photosynthesis (chlorophylls, carotenoids, xanthophylls, side chains of phylloquinone and plastoquinone) are derived from the MEP pathway. Exposure of dark-grown *Arabidopsis* seedlings to low frequency red light (which is known to induce photomorphogenesis) resulted in photomorphogenesis and etioplast development in chloroplast because of accumulation of chlorophylls, xanthophyll, β-carotene and phylloquinone. On the other hand, plants exposed to far red light (which does not induce photomorphogenesis) do not synthesize chlorophylls, β-carotene, and phylloquinone, but in turn accumulate lutein and α-tocopherol without changes in transcript abundance. These results suggests that MEP pathway is regulated at transcriptional and post transcriptional levels during photomorphogenesis[63].

The main rate determining enzyme of MVA pathway in fungi, mammals, insects and plants is hydroxymethylglutaryl CoA reductase (HMGR), which catalyzes the formation of Mevalonate from 3-hydroxy-3-methylglutaryl-CoA[64-68] which is post translationally regulated by feedback-inhibition mechanism. HMGR is anchored to endoplasmic reticulum (ER) whereas all other enzymes of MVA pathway have been found in cytoplasm and peroxisome[69,70].

In vivo feeding experiments suggested that DXS in MEP pathway is a rate limiting step in the biosynthesis of IPP and DMAPP[71,72]. Analysis of *Arabidopsis* transgenic lines[73], tomato[74,75], potato[76]and *Ginkoba biloba*[77] indicates that changes in the level of DXS changes the formation of final product of isoprenoid including chlorophyll, carotenoids, tocopherols and abscisic acid (ABA). In addition to DXS, DXR (DXR reductoisomerase) and HDR (4-hydroxy-3-methylbut-2-enyl diphosphate

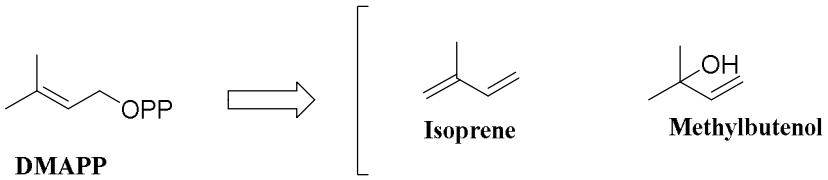
reductase) also have rate limiting roles in MEP pathway for the biosynthesis of IPP and DMAPP.

1.4 Structural classification of Isoprenoids

Isoprenoids can be classified into different groups according to the number of five carbon units (isoprene unit) present in the core structure.

1.4.1 Hemiterpene (C5)

Hemiterpenes consist of a single isoprene unit (C5), but their oxygenated derivatives are called as hemiterpenoids eg: prenol, isovaleric acid. Several plant species from mosses, fern, and trees emit isoprenes in very large amounts. Isoprenes are synthesized by chloroplastic enzymes directly from DMAPP by diphosphate elimination, which is produced from MEP pathway [78-80].



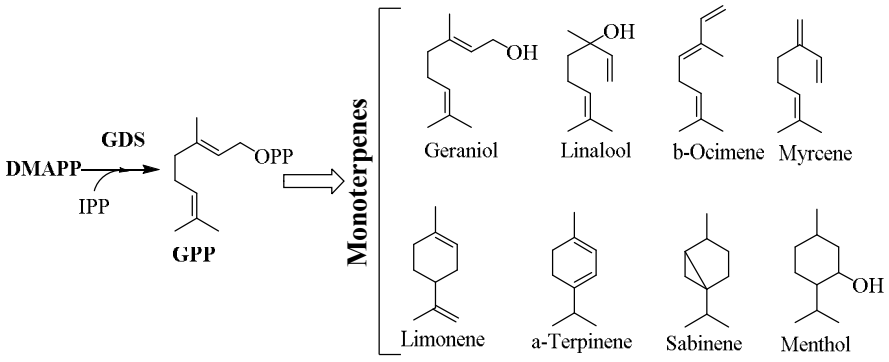
Scheme 1.4.1: Biosynthesis of hemiterpenes

Isoprene emission plays an important role in protecting leaves from abiotic stress (short high temperature condition). The evolution of isoprene emission in plants may have been important in allowing plants to survive during rapid temperature changes. Emission of isoprenes from plants account for a significant proportion of atmospheric hydrocarbon[81].

1.4.2 Monoterpene (C10)

Monoterpenes consist of two isoprene units with the molecular formula $C_{10}H_{16}$ and are present in secretory tissues such as oil glands of higher plants, insects, fungi and marine organisms. These molecules are volatile in nature (boiling point in the range of 150-185 °C), and are less dense than water. These compounds are widely used in flavour and fragrance industry, due to their characteristic odour. Oxygenated derivatives of monoterpenes are more widespread in nature with greater importance. Monoterpenes can be divided into three subgroups: acyclic (myrcene, geraniol, linalool), monocyclic (limonene, α-terpineol and terpinolene) and bicyclic (α-pinene,

sabinene and camphor). Monoterpenes are biosynthesized from geranyl pyrophosphate (GPP) catalyzed by monoterpene synthase.

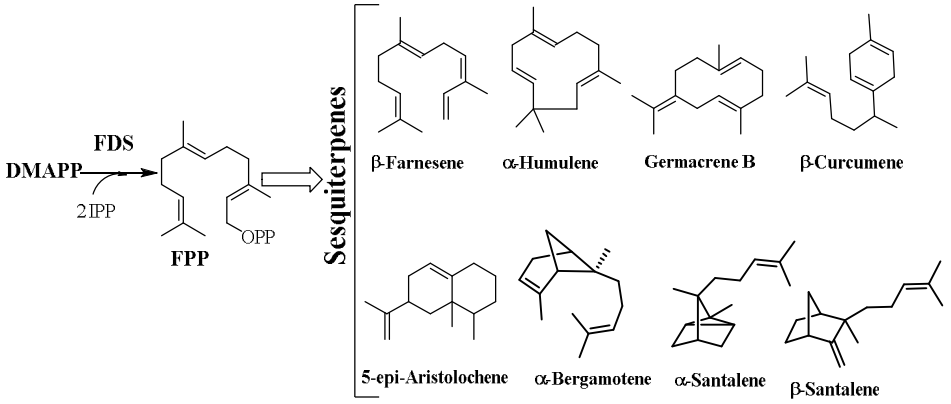


Scheme 1.4.2: Biosynthesis of monoterpenes

Several monoterpenes possess various pharmacological properties including antibacterial antifungal, antioxidant, and anticancerous[82-84]. In plants, monoterpenes are synthesized in plastids through the DXP pathway, whereas in other higher organisms and in yeast, they are synthesized through the MVA pathway[85].

1.4.3 Sesquiterpene (C15)

Sesquiterpenes are the most diverse group of isoprenoids consisting of three isoprene units with the molecular formula $C_{15}H_{24}$. This abundant group of isoprenoids consists of over 7000 molecules with more than 300 stereochemically distinct hydrocarbon skeletons. Sesquiterpenes are derived from the linear substrate, farnesyl pyrophosphate (C15) which is biosynthesized by FDS by head to tail condensation of IPP and GPP or IPP and DMAPP. They represent one of the most important components of plant essential oils along with monoterpenes. The large carbon skeleton of farnesyl diphosphate (FPP) and the presence of three double bonds greatly increase the structural diversity of the resulting products. Sesquiterpenes can be divided into four subgroups: acyclic (farnesene, farnesol, nerolidol), monocyclic (bisabolene, curcumene, zingiberene), bicyclic (bergamotene, β-santalene, aristolochene, cadinene, caryophyllene) and tricyclic (α-santalene, longifolene). Sesquiterpenes are biosynthesized from MVA pathway and targeted in the cytosol[86].

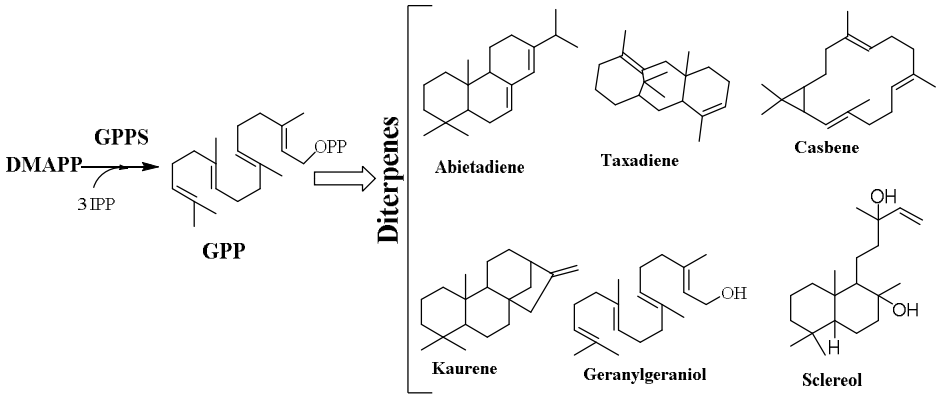


Scheme 1.4.3: Biosynthesis of sesquiterpenes

Sesquiterpenes are traditionally used as flavours and fragrances but they have potential to serve as anticancer[87] and antimalarial[88]. In recent years, sesquiterpenes of farnesene and bisabolene skeletons have been recognized as replacements for petroleum-derived fuels[89].

1.4.4 Diterpene (C20)

Diterpenes consist of four isoprene units with the molecular formula $C_{20}H_{32}$ and are widely distributed in nature. The diterpene compounds are derived from geranylgeranyl diphosphate (GGPP) and divided into the following subgroups: linear (geranylgeraniol, geranyl linalool), bicyclic (casbene, sclareol), tricyclic (abietadiene, taxadiene), tetracyclic (kaurene aphidicolanes, gibberellenes) and macrocyclic (cembranes, daphnanes, tiglianes). They are mainly present in polyoxygenated form with keto and hydroxyl group.

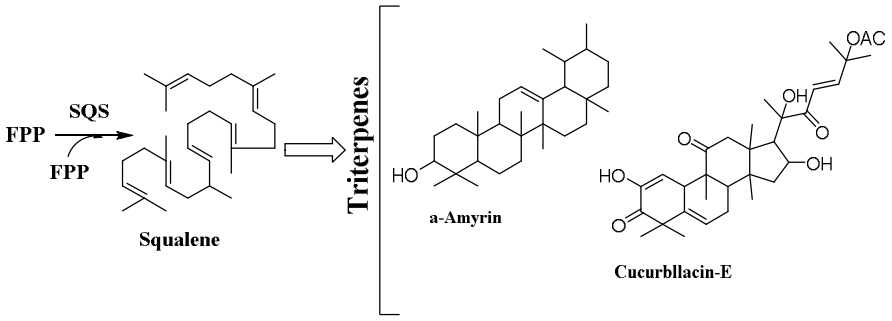


Scheme 1.4.4: Biosynthesis of diterpenes

Diterpenes have attracted attention because of their biological and pharmacological activities such as anti-bacterial, anti-fungal, anti-inflammatory and anti-leishmanial activity. Taxol is one of the most effective diterpenes which possesses anticancer properties against a wide range of cancers.

1.4.5 Triterpene (C30)

Triterpenes consist of six isoprene units and are derived from squalene which is synthesized by head to head condensation of two units of farnesyl diphosphate. Triterpenes constitute the most important class of terpenoids and exhibit a wide range of structural diversity and regulate membrane fluidity and permeability with lipids and proteins[90]. Sterols are precursors for a vast array of compounds involved in important cellular processes in animals whereas plant sterols are linked to brassinosteroids synthesis[91,92]. Brassinosteroids are polyhydroxy plant steroids essential for normal growth and development and present in all parts of plants[93].

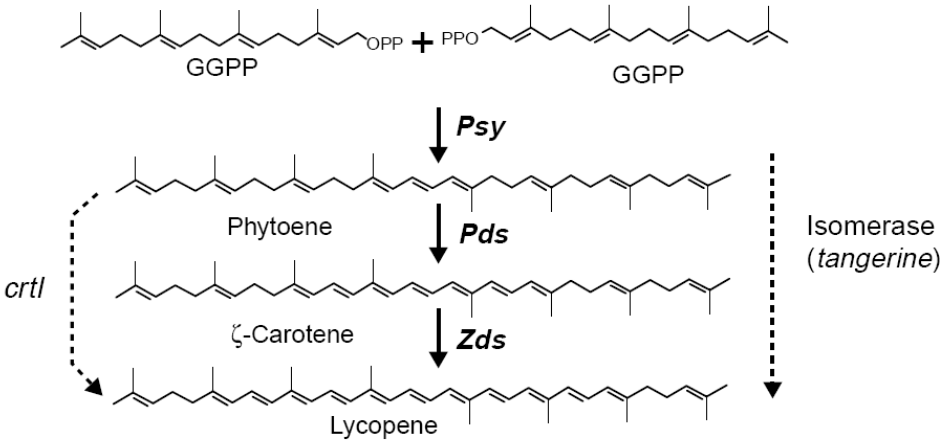


Scheme 1.4.5: Biosynthesis of triterpenes

1.4.6 Tetraterpene (C40)

Carotenoids belong to tetraterpenoids which consist of eight isoprene units, having the molecular formula $C_{40}H_{56}$ instead of $C_{40}H_{64}$ and are found in all plants, bacteria and fungi. Carotenoids are synthesized from the DOXP pathway in plants and some bacteria whereas in fungi, they are synthesized by the MVA pathway[94,95]. The first committed step in carotenoid biosynthesis is condensation of two molecules of GGPP to form phytoene catalyzed by phytoene synthase. Oxygenated derivatives of carotenoids such as lutein and zeaxanthin are known as xanthophylls, whereas mono-oxygenated carotenoids such as α-carotene, β-carotene and lycopene are known as carotenes. The presence of conjugation system in carotenoids enables them to possess

light absorbing properties in the range of visible spectrum. This property of carotenoids is exploited by photosynthetic organisms. Xanthophylls are necessary pigments for light harvesting antennae of chloroplast which prevents chlorophyll bleaching in intense light[96]. Carotenoids play another important role in providing distinct colours to flowers and fruits of plants.



Scheme 1.4.6: Biosynthesis of tetraterpenes

Commercially available carotenoids such as β-carotene, lycopene and astaxanthin are widely used as food colorants, animal feed supplements (since they cannot synthesize carotenoids *de novo*), cosmetic purposes, etc. There is a considerable interest in carotenoids because of their antioxidant properties and ability to alleviate chronic diseases[97].

1.4.7 Polyterpene

Polyterpenes consist of many isoprene units with the molecular formula $(C_5H_8)_n$. In eukaryotic system cis-prenyltransferases synthesize long chain polyprenyl diphosphates (C_{55}-C_{100}) by successive condensation of IPP on FPP. Polyprenyl diphosphate is converted to dolichol by dephosphorylation and saturation of α-isoprene unit[98-100], which serve as glycosyl carriers in protein glycosylation.

Rubber (*cis*-1,4-polyisoprene) is a polyisoprenoid without any known physiological function to plants, produced by more than 2000 plant species[101,102] and used in the preparation of heavy duty tyres and various industrial applications. *Hevea brasiliensis* is the only commercial source for natural rubber[102], which occurs on the surface of

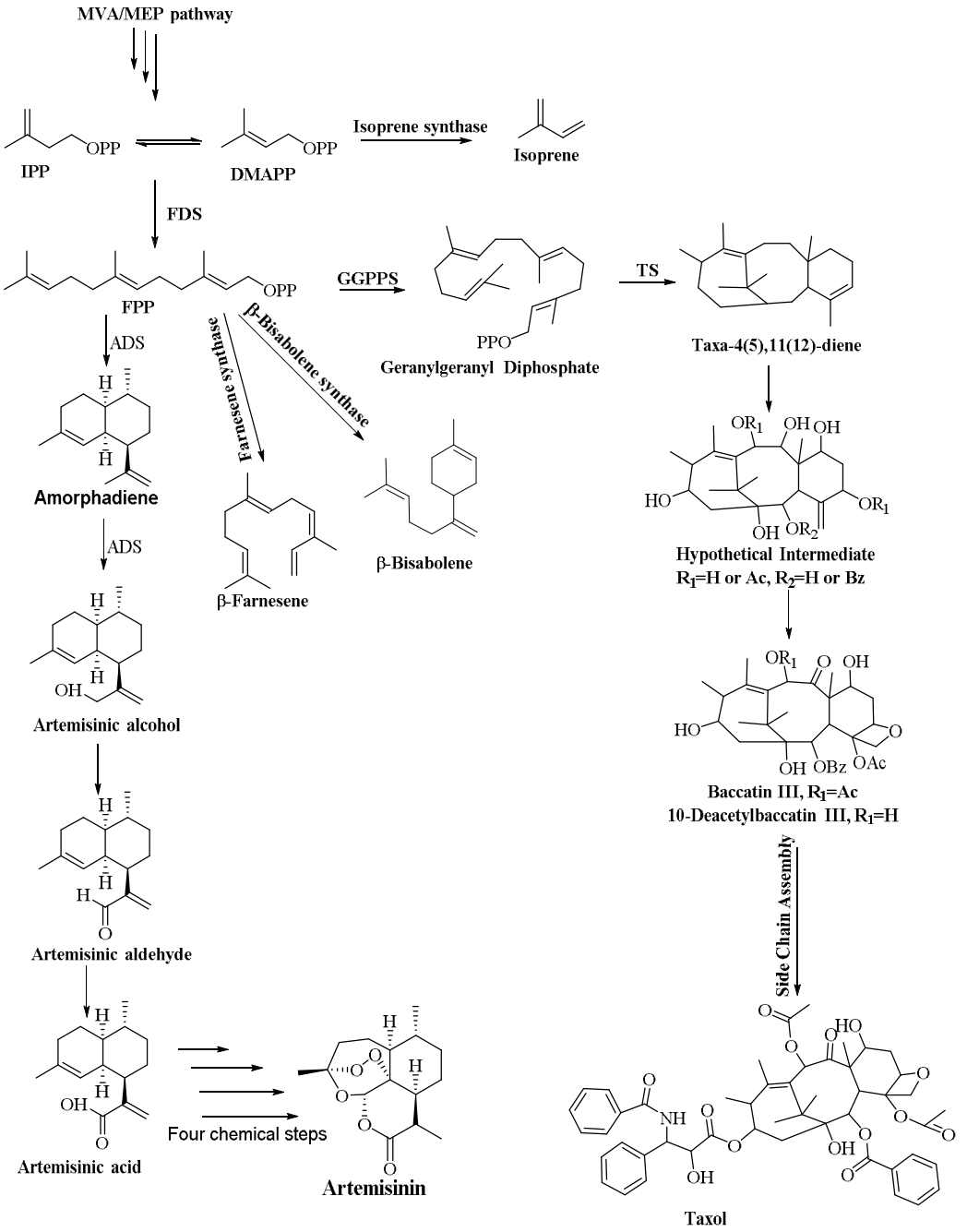
rubber particles suspended in latex[103]. Rubber prenyl-transferases synthesize polymers by condensation of thousands of IPP units.

1.5 Metabolic engineering for the production of isoprenoids

Isoprenoids are the most structurally diverse natural products and have found various applications in medicine (artemisinin as antimalarial, taxol as anticancer), nutrition (carotenoids), flavours and fragrances (essential oils). Plants are one of the major sources for the production of these isoprenoids. However the supply of these products suffers from low yields, impurities and consumption of large amounts of natural resources. Chemical synthesis of these natural isoprenoids is difficult and costly because of their structural complexity and low yields. Engineering of metabolic pathway in microbial systems for the production of terpenes in large scale presents a cost-effective alternative tool for isoprenoid biosynthesis. Recent advances in molecular techniques have enabled the engineering of microorganisms for the production of isoprenoids. Metabolic engineering for the production of isoprenoids requires two main steps: (i) the production of common intermediate prenyl diphosphate (farnesyl diphosphate) by FDS from basic building blocks (IPP and DMAPP) synthesized by either MVA or MEP pathway, and (ii) conversion of prenyl diphosphate into products of interest by incorporating heterologous genes into *E.coli* or yeast through synthetic biology tools.

1.5.1 Microbial engineering for the production of Artemisinin

Artemisinin, a sesquiterpene lactone extracted from *Artemisia annua* is an antimalarial drug and highly effective against multidrug resistant strains of *P. falciparum*[104]. Artemisinin is accumulated in glandular trichome to the level of 0.01-1 % of dry weight[105]. Supply from natural resource is limiting because of low yield. Although, chemical synthesis is possible, structural complexity and low yield makes it economically nonviable for drug production[106]. Therefore a semisynthetic route for the production of artemisinin could be utilized by starting from artemisinic acid[107]. 25 g/litre of artemisinic acid was produced in an engineered yeast strain (*S. cerevisiae*) starting from a simple inexpensive carbon source by fermentation[108] (Scheme 1.5). This result would constitute a superior route for the development of economically viable processes for the production of semi synthetic artemisinin.



Scheme 1.5: Schematic representation of metabolic engineering of various isoprenoids in microorganism

1.5.2 Microbial engineering for the production of Taxol

Taxol is a polyoxygenated cyclic diterpene and is well-established as an anti cancer molecule[109] and as a potent antineoplastic drug against a wide range of cancer[110]. Naturally, taxol is isolated from the bark of taxus *sp.* with a yield of 0.001-0.05 % of dry weight[111]. Total synthesis of taxol is reported but the approach is commercially

not viable[112], therefore increased demand for drug shifted from large scale destruction of low yielding *Taxus sp*. to semisynthetic approach[113]. Taxol can be produced semi-synthetically from more abundant taxoid like baccatin III and 10-deacetyl baccatin III isolated from needles or cell cultures of various *Taxus* species as renewable resources[114-117]. However, the extraction of semisynthetic precursors from *Taxus* is also very difficult and expensive. Therefore, the production of taxol or its precursor in a genetically engineered microorganism is a desirable alternative. The most promising approach for the production of taxol or suitable precursors at industrial level is through plant cell culture.

1.6 Sandalwood (*Santalum album* Linn)

Sandalwood is a slow growing medium-sized evergreen hemi root parasitic tree belonging to the Santalaceae family. Sandalwood is one of the most valuable trees of the plant kingdom in the world. There are 18 species of Sandalwood trees belonging to genus *Santalum* which are distributed in the tropical and temperate regions of India, Indonesia, Australia and Pacific Island[118-120]. Sandalwood trees grow between 4 to 9 meter in length and girth of up to 1 meter which lives for more than 100 years. Six out of 18 species can be found in the Hawaiian Island which represents the highest level of Sandalwood diversity. This species is highly valued for its fragrant heartwood, which contains the essential oil known as sandalwood oil. The oil content varies widely between species and even within the species. Indian Sandalwood (*Santalum album*) is the most valuable Sandalwood species because of the high oil content in its heartwood and high quality of its oil. Sandalwood oil is commercially produced by steam distillation of heartwood chip with varying yields from 4-7 % from a well matured tree. Sandalwood species are renowned for their fragrant heartwood and oil to the extent that they have been exploited in many parts of their natural distribution. To produce commercially valuable high quality Sandalwood oil, harvested Sandalwood trees have to be at least 40 year of age, but 80 or above are preferred. The annual production of sandalwood is estimated between 200-300 tonnes worldwide, out of which 90 % of total production is from India. In India 80 % of the production is from the states of Tamil Nadu, Kerala and Karnataka. Due to high demand for Sandalwood and its oil, activities such as illegal cutting, over harvesting and illegal poaching of sandalwood tree have increased drastically, ultimately resulting in depletion of its natural resources. This leads to several disadvantages such

as: adulteration of Sandalwood oil, dramatic increases in the cost of Sandalwood oil, and scarcity of resource. The rank of highly valued Sandalwood oil in the world Sandalwood market is shown in table 1.1[120].

Table 1.1: Commercially valued Sandalwood species and their natural distribution

Rank	Scientific name	Vernacular name	Natural distribution
1	*S. album*	Indian Sandalwood	India, Indonesia, Sri Lanka
2	*S. yasi*		Fiji, Niue, Tonga
3	*S. astrocaledonicum*	Vanuata Sandalwood	New Caledonia, Vanuatu
4	*S. macgregorii*		Papua New Guinea, Indonesia
5	*S. spicatum*	Australian Sandalwood	Australia
6	*S. lanceolatum*	Northern Sandalwood	Australia

1.6.1 Classification

Kingdom:	Plantae
Phylum	Steptophyta
Class	Eudicots
Sub-class	Core eudicots
Order:	Santalales
Family:	*Santalaceae*
Genus:	*Santalum*
Species:	*S. album*

India has a vast and inexhaustible resource of medicinal and aromatic plants. Over 45,000 plant species are known in India, of which about six thousand are known to possess medicinal and organoleptic properties and very few plants have been explored for the natural products which possess medicinal properties[121]. About 1000 medicinal plant species face a threat to their existence. Indian sandalwood (*Santalum album*, Linn) is a threatened species and grows in the Western Ghats and other mountain regions such as Kalrayan and Shevaroy Hills.

1.6.2 Heartwood formation and oil content

The hallmark of heartwood formation is the death of both axial and ray parenchyma cells within the narrow transition zone between heartwood and sapwood. Events

immediately receding cell death are responsible for the best known characteristics of heartwood and are themselves suggestive of an active cell metabolism before death. These include rapid synthesis of secondary metabolites, formation of blockages (e.g. tyloses and gels) within conducting elements and changes in the structure of pits connecting conducting elements with parenchyma and with each other.

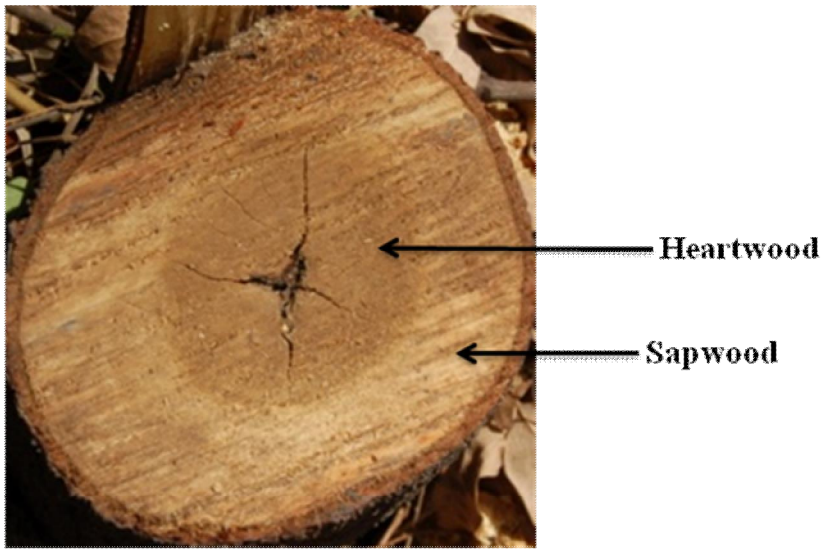


Figure 1.6.2: Cross-section of Sandalwood trunk, inner dark coloured region is known as heartwood and outer light coloured region is known as sapwood.

In many conifers and angiosperms, parenchyma in the transition zone, synthesize secondary metabolites (sometimes termed extractives). These compounds are deposited in the walls of neighbouring cells and often fill parenchymal lumen, imparting properties of decay resistance, colour and reduced permeability to the heartwood. They are predominantly phenolic compounds found in a range of polymerization states, including lignans, stilbenes, flavonoids and quinones, with the latter two associated with colour change occurring during heartwood formation.

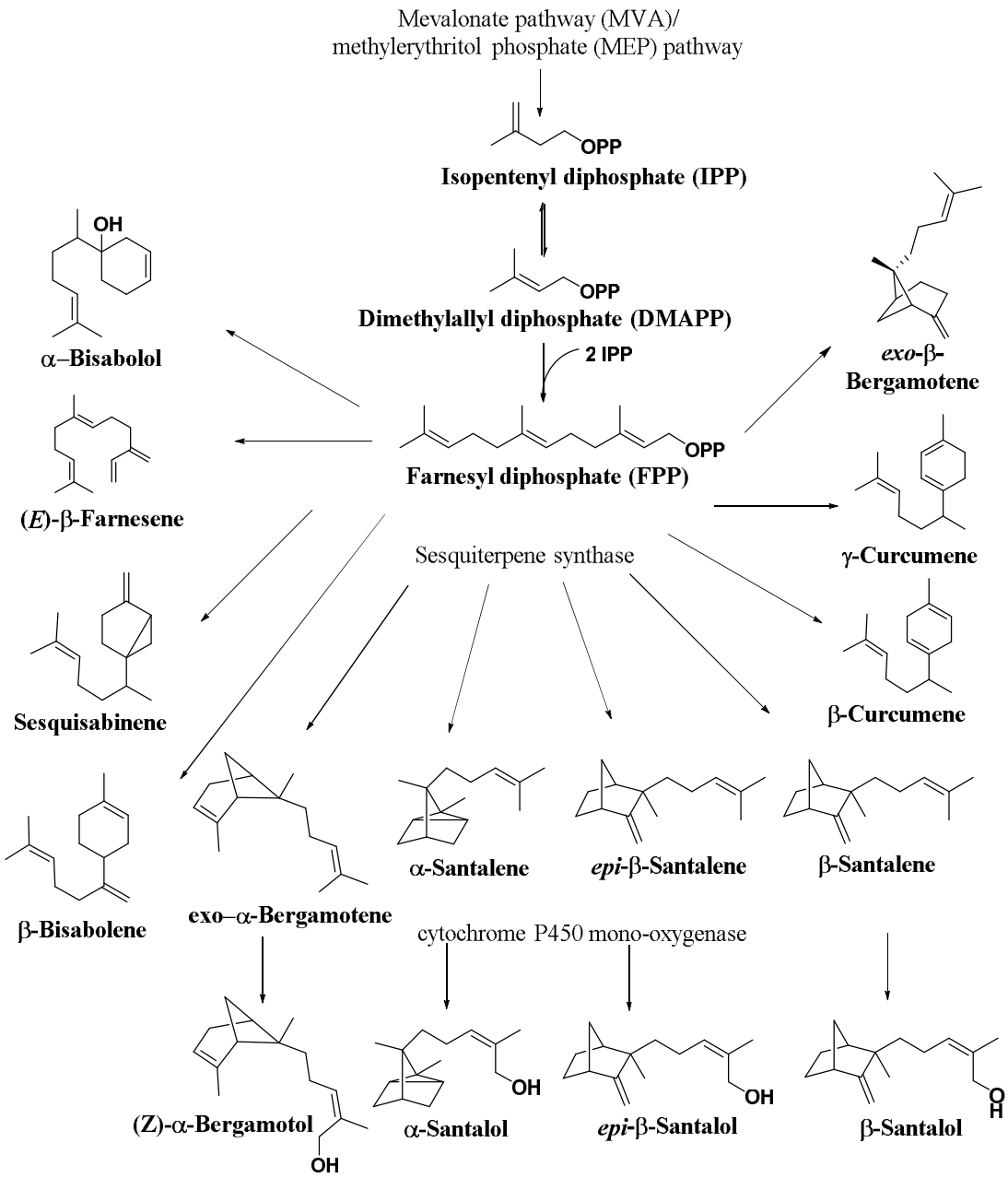
Heartwood formation in sandalwood trees generally starts around 10 – 13 years of age but what triggers this process has not been very well understood. Certain factors, generally relating to stress, such as gravel-dry soil, insulation and range of elevation (500 – 700 m), seem to provide the right environment for the formation of heartwood. According to an estimate, about 65 % of the heartwood is used to produce essential oil, 10 % of heartwood is used in carving, and 25 % of heartwood is used in incense sticks.

1.6.3 Essential oil composition of Sandalwood

Among the reported constituents of sandalwood oil, major components are sesquiterpenene alcohols identified as (*Z*)-α-santalol, (*Z*)-β-santalol, (*Z*)-α-bergamotol, (*Z*)-*epi*-β-santalol along with minor quantities of parental hydrocarbons, α-santalene, β-santalene, α-bergamotene, *epi*-β-santalene, sesquisabinene hydrates, α-curcumene, β-curcumene, γ-curcumene, β-bisabolene, α-bisabolol, cedrene, cedrol, sesquisabinene, triterpenoids and phenylpropanoids[122-124]. (*Z*)-α- and (*Z*)-β-santalols are the major constituents (~80 %) of the essential oil from a well-matured tree (≥80 years old) and are responsible for most of the biological activity of the oil. Specific aroma of sandalwood oil is produced by mainly two compounds (*Z*)-α- and (*Z*)-β-santalols. Therefore, in order to maintain the quality of Sandalwood oil, ISO standards have been formulated for *Santalum album* (*Z*)-α-santalol: 41-55 %, (*Z*)-β-santalol: 16-24 % [120].

1.6.4 Biosynthesis of santalene derivatives in Indian Sandalwood

Isopentenyl diphosphate (IPP) and Dimethylallyl diphosphate (DMAPP) which are synthesized from either MVA or MEP pathway, condense in a head to tail fashion to generate farnesyl diphosphate (*E,E*-FPP) with the help of farnesyl diphosphate synthase. Further (*E,E*-FPP) undergoes cyclization with the help of sesquiterpene synthases to form sesquiterpenes (santalene derivatives). These santalenes derivatives undergo hydorxylation at their *cis*-methyl position to form santalols with the help of CYP450 systems, which are the major constituents of sandalwood oil. The prime-most enzymes of the terpene biosynthesis in sandalwood, leading to the formation of santalene derivatives, are santalene synthase, β-bisabolene synthase, and sesquisabinene synthase. These enzymes catalyze a complex, multi-step reaction capable of generating a mixture of sesquiterpenes such as α-santalene, β-santalene, *epi*-β-santalene, exo-α-bergamotene, (*E*)-β-farnesene, β-bisabolene, sesquisabinene, with varied carbon skeletons from a single diphosphate substrate (*E,E*-FPP)[125]. Further the CYP450 mono-oxygenase system will hydroxylate at the cis methyl group of the side chain to form (*Z*)-α-santalol, (*Z*)-α-exo-bergomotol, (*Z*)-epi-β-santalol, (*Z*)-β-santalol, and sesquisabinene hydrates representing >90 % of oil component. Biosynthesis of sesquiterpenoids present in Indian Sandalwood is represented in scheme 1.6.



Scheme 1.6: Schematic representation of biosynthesis of Sesquiterpenes in Indian Sandalwood *Santalum album*

1.6.5 Application of Sandalwood oil

Indian Sandalwood is known worldwide for its fragrant heartwood. Essential oil of Sandalwood is a pale yellow to yellow colour viscous liquid with a distinctive soft, smooth, fragrant, persistent, warm and woody scent. It has been extensively used in many purposes like perfumes, cosmetics, aromatherapy, incense sticks industries, religious ceremonies and pharmaceutical industries. Sandalwood oil is used extensively in perfume industry as a harmonizing agent, base note and fixative. The

oil is used in flavouring food substances such as dairy desserts, baked food, gelatine, etc. Sandalwood has been part of religious and spiritual ceremonies in India. In Ayurveda, Chinese and Tibetan medicinal systems, Sandalwood oil is used in a wide variety of medicinal applications such as antipyretic, antiseptic, for nervous disorders, depression, anxiety, anti-inflammatory, treatment for urinary infection, dysentery and against several skin diseases for thousands of years. (*Z*)-α-santalol has particularly attracted attention for its neuroleptic[126] properties and chemopreventive[127,128] effects in *in vitro* and *in vivo* assay systems. A recent study suggests that (*Z*)-β-santalol possesses anti-influenza A/HK virus activity of 86 % with no cytotoxic activity at the concentration of 100 µg/mL[129].

1.7 Scope of thesis

There were no reports available at the start of this work towards understanding the biosynthesis of santalenes and its derivatives, and mechanism of their storage in heartwood of Indian Sandalwood, *S. album*. Since the supply of sandalwood oil mainly depends on harvesting from Sandalwood trees, this leads to destruction of natural resources. Excessive harvesting of natural resources without replenishment has substantially reduced the sandalwood industries leading to global shortage and increase in market price. Indian Sandalwood, *S. album* has been categorized as vulnerable by IUCN (International Union for Conservation of nature). The primary aim of this work was to characterize the genes involved in the biosynthesis of santalene derivative in Indian Sandalwood S*antalum album* and establish the mechanism utilized by these genes/enzymes for the catalysis. The characterized enzymes can be utilized in heterologous host cells to produce santalenes and its derivatives in large quantities.

The thesis is divided into five chapters

Chapter 1: Introduction

Chapter 2: RNA isolation and transcriptome screening for genes involved in santalene biosynthesis

Chapter 3: Screening, cloning, expression and characterization of terpene synthases involved in biosynthesis of santalene derivatives

Chapter 4: Santalene synthase: A comprehensive site directed mutagenesis study to understand the dynamic nature of santalene synthase active site pocket

Chapter 5: Screening, isolation and cloning of Cytochrome P450 systems from Indian Sandalwood *S. album*

1.8 References

(1) Dixon, R. A. *Curr. Opin. Biotech.* **1999**, *10*, 192-197.

(2) Bevan, M.; Bancroft, I.; Bent, E.; Love, K.; Goodman, H.; Dean, C.; Bergkamp, R.; Dirkse, W.; Van Staveren, M.; Stiekema, W.; Drost, L.; Ridley, P.; Hudson, S. A.; Patel, K.; Murphy, G.; Piffanelli, P.; Wedler, H.; Wedler, E.; Wambutt, R.; Weitzenegger, T.; Pohl, T. M.; Terryn, N.; Gielen, J.; Villarroel, R.; De Clerck, R.; Van Montagu, M.; Lecharny, A.; Auborg, S.; Gy, I.; Kreis, M.; Lao, N.; Kavanagh, T.; Hempel, S.; Kotter, P.; Entian, K. D.; Rieger, M.; Schaeffer, M.; Funk, B.; Mueller-Auer, S.; Silvey, M.; James, R.; Montfort, A.; Pons, A.; Puigdomenech, P.; Douka, A.; Voukelatou, E.; Milioni, D.; Hatzopoulos, P.; Piravandi, E.; Obermaier, B.; Hilbert, H.; Dusterhoft, A.; Moores, T.; Jones, J. D. G.; Eneva, T.; Palme, K.; Benes, V.; Rechman, S.; Ansorge, W.; Cooke, R.; Berger, C.; Delseny, M.; Voet, M.; Volckaert, G.; Mewes, H. W.; Klosterman, S.; Schueller, C.; Chalwatzis, N. *Nature* **1998**, *391*, 485-488.

(3) Somerville, C.; Somerville, S. *Science* **1999**, *285*, 380-383.

(4) Koes, R. E.; Quattrocchio, F.; Mol, J. N. M. *BioEssays* **1994**, *16*, 123-132.

(5) Caporale, L. H. *P. Natl. Acad. Sci. USA* **1995**, *92*, 75-82.

(6) Thulasiram, H. V.; Erickson, H. K.; Poulter, C. D. *Science* **2007**, *316*, 73-76.

(7) Reiling, K. K.; Yoshikuni, Y.; Martin, V. J. J.; Newman, J.; Bohlmann, J.; Keasling, J. D. *Biotechnol. Bioeng.* **2004**, *87*, 200-212.

(8) Ding, V. D. H.; Sheares, B. T.; Bergstrom, J. D.; Ponpipom, M. M.; Perez, L. B.; Poulter, C. D. *Biochem. J.* **1991**, *275*, 61-65.

(9) Anderson, M. S.; Yarger, J. G.; Burck, C. L.; Poulter, C. D. *J. Biol. Chem.* **1989**, *264*, 19176-19184.

(10) Beyer, P.; Mayer, M.; Kleinig, H. *Eur. J. Biochem.* **1989**, *184*, 141-150.

(11) Krinsky, N.; Mathews-Roth, M.; Taylor, R.; Britton, G. In *Carotenoids*; Springer US: 1989, p 167-184.

(12) Momose, K.; Rudney, H. *J. Biol. Chem.* **1972**, *247*, 3930-3940.

(13) Trumpower, B. L.; Houser, R. M.; Olson, R. E. *J. Biol. Chem.* **1974**, *249*, 3041-3048.

(14) Rock, C. D.; Zeevaart, J. A. *P. Natl. Acad. Sci. USA* **1991**, *88*, 7496-7499.

(15) Parry, A. D.; Horgan, R. *Phytochemistry* **1991**, *30*, 815-821.

(16) Lange, B. M.; Rujan, T.; Martin, W.; Croteau, R. *P. Natl. Acad. Sci. USA* **2000**, *97*, 13172-13177.

(17) Jennewein, S.; Croteau, R. *Appl. Microbiol. Biot.* **2001**, *57*, 13-19.

(18) Rodriguez-Concepion, M. *Curr. Pharm. Design* **2004**, *10*, 2391-2400.

(19) McCaskill, D.; Croteau, R. *Adv. Biochem. Eng. Biot.* **1997**, *55*, 107-46.

(20) McGarvey, D. J.; Croteau, R. *Plant Cell* **1995**, *7*, 1015-26.

(21) Bach, T. J. *Lipids* **1995**, *30*, 191-202.

(22) Wouters, J.; Oudjama, Y.; Barkley, S. J.; Tricot, C.; Stalon, V.; Droogmans, L.; Poulter, C. D. *J. Biol. Chem.* **2003**, *278*, 11903-11908.

(23) Kaneda, K.; Kuzuyama, T.; Takagi, M.; Hayakawa, Y.; Seto, H. *P. Natl. Acad.Sci. USA* **2001**, *98*, 932-937.

(24) Hemmi, H.; Ikeda, Y.; Yamashita, S.; Nakayama, T.; Nishino, T. *Biochem. Bioph. Res. Co.* **2004**, *322*, 905-910.

(25) Chen, M.; Poulter, C. D. *Biochemistry* **2009**, *49*, 207-217.

(26) Mabanglo, M. F.; Schubert, H. L.; Chen, M.; Hill, C. P.; Poulter, C. D. *ACS Chem. Biol.* **2010**, *5*, 517-527.

(27) Dellas, N.; Noel, J. P. *ACS Chem. Biol.* **2010**, *5*, 589-601.

(28) Grochowski, L. L.; Xu, H.; White, R. H. *J. Bacteriol* **2006**, *188*, 3192-3198.

(29) Rohmer, M.; Knani, M.; Simonin, P.; Sutter, B.; Sahm, H. *Biochem. J.* **1993**, *295*, 517-524.

(30) Sprenger, G. A.; Schorken, U.; Wiegert, T.; Grolle, S.; de Graaf, A. A.; Taylor, S. V.; Begley, T. P.; Bringer-Meyer, S.; Sahm, H. *P. Natl. Acad. Sci. USA* **1997**, *94*, 12857-12862.

(31) Eisenreich, W.; Schwarz, M.; Cartayrade, A.; Arigoni, D.; Zenk, M. H.; Bacher, A. *Chem. Biol.* **1998**, *5*, R221-R233.

(32) Zeidler, J.; Schwender, J.; Mueller, C.; Lichtenthaler, H. K. *Biochem. Soc. T.* **2000**, *28*, 796-798.

(33) RodrÃ-guez-ConcepciÃ³n, M.; ForÃ©s, O.; MartÃ-nez-GarcÃ-a, J. F.; GonzÃ¡lez, V. c.; Phillips, M. A.; Ferrer, A.; Boronat, A. *Plant Cell* **2004**, *16*, 144-156.

(34) Rohdich, F.; Bacher, A.; Eisenreich, W. *Biochem. Soc. T.* **2005**, *33*, 785-791.

(35) Cane, D. E. *Accounts Chem. Res.* **1985**, *18*, 220-226.

(36) Cane, D. E. *Chem. Rev.* **1990**, *90*, 1089-1103.

(37) Croteau, R. *Chem. Rev.* **1987**, *87*, 929-954.

(38) Davis, E. M., Croteau, R. ; Springer: 2000; Vol. 209, p 53-95.

(39) Lesburg, C. A.; Caruthers, J. M.; Paschall, C. M.; Christianson, D. W. *Curr. Opin. Struc. Biol.* **1998**, *8*, 695-703.

(40) Poulter, C. D. *Phytochem. Rev.* **2006**, *5*, 17-26.

(41) Tantillo, D. J. *Nat. Prod. Rep.* **2011**, *28*, 1035-1053.

(42) Thulasiram, H. V.; Poulter, C. D. *J. Am. Chem. Soc.* **2006**, *128*, 15819-15823.

(43) Ohnishi, T.; Yokota, T.; Mizutani, M. *Phytochemistry* **2009**, *70*, 1918-1929.

(44) Thulasiram, H. V.; Erickson, H. K.; Poulter, C. D. *J. Am. Chem. Soc.* **2008**, *130*, 1966-1971.

(45) Ruzicka, L.; Stoll, M. *Helv. Chim. Acta* **1922**, *5*, 923-936.

(46) Ruzicka, L. *Experientia* **1953**, *9*, 357-67.

(47) Burke, C. C.; Wildung, M. R.; Croteau, R. *P. Natl. Acad. Sci. USA* **1999**, *96*, 13062-13067.

(48) Burke, C.; Croteau, R. *Journal of Biological Chemistry* **2002**, *277*, 3141-3149.

(49) Burke, C.; Croteau, R. *J. Biol. Chem.* **2002**, *277*, 3141-3149.

(50) Delourme, D.; Lacroute, F. o.; Karst, F. *Plant Mol. Biol.* **1994**, *26*, 1867-1873.

(51) Okada, K.; Saito, T.; Nakagawa, T.; Kawamukai, M.; Kamiya, Y. *Plant Physiol.* **2000**, *122*, 1045-1056.

(52) Thiel, R.; Adam, K. P. *Phytochemistry* **2002**, *59*, 269-274.

(53) Itoh, D.; Karunagoda, R. P.; Fushie, T.; Katoh, K.; Nabeta, K. *J. Nat. Prod.* **2000**, *63*, 1090-1093.

(54) Itoh, D.; Kawano, K.; Nabeta, K. *J. Nat. Prod.* **2003**, *66*, 332-336.

(55) Nabeta, K.; Komuro, K. *Chem. Commun.* **1998**, 671-672.

(56) Lichtenthaler, H. K. *Annu. Rev. Plant Phys.* **1999**, *50*, 47-65.

(57) Rohmer, M. *Nat. Prod. Rep.* **1999**, *16*, 565-574.

(58) Eisenreich, W.; Rohdich, F.; Bacher, A. *Trends Plant Sci.* **2001**, *6*, 78-84.

(59) Sallaud, C.; Rontein, D.; Onillon, S.; Jabes, F.; Duffe, P.; Giacalone, C.; Thoraval, S.; Escoffier, C.; Herbette, G.; Leonhardt, N.; Causse, M.; Tissier, A. *Plant Cell* **2009**, *21*, 301-317.

(60) Burg, J. S.; Espenshade, P. J. *Prog. Lipid Res.* **2011**, *50*, 403-410.

(61) Bach, T. J.; Rohmer, M.; Rodriguez-Concepcion, M.; Boronat, A. In *Isoprenoid Synthesis in Plants and Microorganisms*; Springer New York **2013**, 1-16.

(62) Schmid, M.; Davison, T. S.; Henz, S. R.; Pape, U. J.; Demar, M.; Vingron, M.; Scholkopf, B.; Weigel, D.; Lohmann, J. U. *Nat. Genet.* **2005**, *37*, 501-506.

(63) Ghassemian, M.; Lutes, J.; Tepperman, J. M.; Chang, H.-S.; Zhu, T.; Wang, X.; Quail, P. H.; Markus Lange, B. *Arch. Biochem. Biophys.* **2006**, *448*, 45-59.

(64) Basson, M. E.; Thorsness, M.; Rine, J. *P. Natl. Acad. Sci. USA* **1986**, *83*, 5563-5567.

(65) Goldstein, J. L.; Brown, M. S. *Nature* **1990**, *343*, 425-430.

(66) Monger, D. J.; Lim, W. A.; Kozdy, F. J.; Law, J. H. *Biochem. Bioph. Res. Co.* **1982**, *105*, 1374-1380.

(67) Chappell, J.; Wolf, F.; Proulx, J.; Cuellar, R.; Saunders, C. *Plant Physiol.* **1995**, *109*, 1337-1343.

(68) Bach, T. J. *Plant Physiol. Bioch.* **1987**, *25*, 163-178.

(69) Sapir-Mir, M.; Mett, A.; Belausov, E.; Tal-Meshulam, S.; Frydman, A.; Gidoni, D.; Eyal, Y. *Plant Physiol.* **2008**, *148*, 1219-1228.

(70) Simkin, A.; Guirimand, G.; Papon, N.; Courdavault, V.; Thabet, I.; Ginis, O.; Bouzid, S.; Giglioli-Guivarc'h, N.; Clastre, M. *Planta* **2011**, *234*, 903-914.

(71) Guevara-Garcia, A.; San Roman, C.; Arroyo, A.; Cortes, M. E.; de la Luz Gutierrez-Nava, M.; Leon, P. *Plant Cell* **2005**, *17*, 628-643.

(72) Wolfertz, M.; Sharkey, T. D.; Boland, W.; Kuhnemann, F. *Plant Physiol.* **2004**, *135*, 1939-1945.

(73) Estevez, J. M.; Cantero, A.; Reindl, A.; Reichler, S.; Leon, P. *J. Biol. Chem.* **2001**, *276*, 22901-22909.

(74) Lois, L. M.; Campos, N.; Putra, S. R.; Danielsen, K.; Rohmer, M.; Boronat, A. *P. Natl. Acad. Sci. USA* **1998**, *95*, 2105-2110.

(75) Enfissi, E. M. A.; Fraser, P. D.; Lois, L.-M.; Boronat, A.; Schuch, W.; Bramley, P. M. *Plant Biotechnol. J.* **2005**, *3*, 17-27.

(76) Morris, W. L.; Ducreux, L. J. M.; Hedden, P.; Millam, S.; Taylor, M. A. *J. Exp. Bot.* **2006**, *57*, 3007-3018.

(77) Gong, Y.-f.; Liao, Z.-h.; Guo, B.-h.; Sun, X.-f.; Tang, K.-x. *Planta Med.* **2006**, *72*, 329-335.

(78) Sharkey, T. D.; Yeh, S. *Annu. Rev. Plant Phys.* **2001**, *52*, 407-436.

(79) Silver, G. M.; Fall, R. *Plant Physiol.* **1991**, *97*, 1588-1591.

(80) Kuzma, J.; Fall, R. *Plant Physiol.* **1993**, *101*, 435-440.

(81) Zimmerman, P. R.; Chatfield, R. B.; Fishman, J.; Crutzen, P. J.; Hanst, P. L. *Geophys. Res. Lett.* **1978**, *5*, 679-682.

(82) Garcia, R.; Alves, E. S. S.; Santos, M. P.; Aquije, G. r. M. F. V. g.; Fernandes, A. A. R.; Santos, R. B. d.; Ventura, J. A.; Fernandes, P. M. B. *Braz. J. Microbiol.* **2008**, *39*, 163-168.

(83) Singh, P.; Shukla, R.; Prakash, B.; Kumar, A.; Singh, S.; Mishra, P. K.; Dubey, N. K. *Food Chem. Toxicol.* **2010**, *48*, 1734-1740.

(84) Karkabounas, S.; Kostoula, O. K.; Daskalou, T.; Veltsistas, P.; Karamouzis, M.; Zelovitis, I.; Metsios, A.; Lekkas, P.; Evangelou, A. M.; Kotsis, N.; Skoufos, I. *Exp. Oncol.* **2006**, *28*, 121-5.

(85) Hampel, D.; Mosandl, A.; Wust, M. *Phytochemistry* **2005**, *66*, 305-311.

(86) Yu, F.; Utsumi, R. *Cell. Mol. Life Sci.* **2009**, *66*, 3043-3052.

(87) Bommareddy, A.; Rule, B.; VanWert, A. L.; Santha, S.; Dwivedi, C. *Phytomedicine*, *19*, 804-811.

(88) Klayman, D. L. *Science* **1985**, *228*, 1049-1055.

(89) Tippmann, S.; Chen, Y.; Siewers, V.; Nielsen, J. *Biotechnol. J.* **2013**, *8*, 1435-1444.

(90) Ohvo-Rekila, H.; Ramstedt, B.; Leppimaki, P.; Slotte, J. P. *Prog. Lipid Res.* **2002**, *41*, 66-97.

(91) Sakurai, A. *Plant Physiol. Bioch.* **1999**, *37*, 351-361.

(92) Bishop, G. J.; Yokota, T. *Plant Cell Physiol.* **2001**, *42*, 114-120.

(93) Clouse, S. D.; Langford, M.; McMorris, T. C. *Plant Physiol.* **1996**, *111*, 671-678.

(94) Lichtenthaler, H. K.; Rohmer, M.; Schwender, J. *Physiol. Plantarum* **1997**, *101*, 643-652.

(95) Hirschberg, J. *Curr. Opin. Plant Biol.* **2001**, *4*, 210-218.

(96) Demmig-Adams, B.; Gilmore, A. M.; Adams, W. W. *FASEB J.* **1996**, *10*, 403-12.

(97) Rao, A. V.; Rao, L. G. *Pharmacol. Res.* **2007**, *55*, 207-216.

(98) Sagami, H.; Igarashi, Y.; Tateyama, S.; Ogura, K.; Roos, J.; Lennarz, W. J. *J. Biol. Chem.* **1996**, *271*, 9560-9566.

(99) Sagami, H.; Kurisaki, A.; Ogura, K. *J. Biol. Chem.* **1993**, *268*, 10109-10113.

(100) Chojnacki, T.; Dallner, G. *Biochem. J.* **1988**, *251*, 1-9.

(101) Pootakham, W.; Chanprasert, J.; Jomchai, N.; Sangsrakru, D.; Yoocha, T.; Therawattanasuk, K.; Tangphatsornruang, S. *Am. J. Bot.* **2011**, *98*, e337-e338.

(102) Backhaus, R. A. *Israel J. Bot.* **1985**, *34*, 283-293.

(103) Oh, S. K.; Kang, H.; Shin, D. H.; Yang, J.; Chow, K.-S.; Yeang, H. Y.; Wagner, B.; Breiteneder, H.; Han, K.-H. *J. Biol. Chem.* **1999**, *274*, 17132-17138.

(104) Lin, A. J.; Klayman, D. L.; Milhous, W. K. *J. Med. Chem.* **1987**, *30*, 2147-2150.

(105) Duke, M. V.; Paul, R. N.; Elsohly, H. N.; Sturtz, G.; Duke, S. O. *Int J. Plant Sci.* **1994**, *155*, 365-372.

(106) Avery, M. A.; Chong, W. K. M.; Jenningswhite, C. *J. Am. Chem. Soc.* **1992**, *114*, 974-979.

(107) Roth, R. J.; Acton, N. *J. Nat. Prod.* **1989**, *52*, 1183-1185.

(108) Paddon, C. J.; Westfall, P. J.; Pitera, D. J.; Benjamin, K.; Fisher, K.; McPhee, D.; Leavell, M. D.; Tai, A.; Main, A.; Eng, D.; Polichuk, D. R.; Teoh, K. H.; Reed, D. W.; Treynor, T.; Lenihan, J.; Fleck, M.; Bajad, S.; Dang, G.; Dengrove, D.; Diola, D.; Dorin, G.; Ellens, K. W.; Fickes, S.; Galazzo, J.; Gaucher, S. P.; Geistlinger, T.; Henry, R.; Hepp, M.; Horning, T.; Iqbal, T.; Jiang, H.; Kizer, L.; Lieu, B.; Melis, D.; Moss, N.; Regentin, R.; Secrest, S.; Tsuruta, H.; Vazquez, R.; Westblade, L. F.; Xu, L.; Yu, M.; Zhang, Y.; Zhao, L.; Lievense, J.; Covello, P. S.; Keasling, J. D.; Reiling, K. K.; Renninger, N. S.; Newman, J. D. *Nature* **2013**, *496*, 528-+.

(109) Wall, M. E.; Wani, M. C. In *Taxane Anticancer Agents: Basic Science and Current Status* 1995; Vol. 583, p 18-30.

(110) Goldspiel, B. R. *Pharmacotherapy* **1997**, *17*, S110-S125.

(111) Malik, S.; Cusido, R. M.; Mirjalili, M. H.; Moyano, E.; Palazon, J.; Bonfill, M. *Process Biochem.* **2011**, *46*, 23-34.

(112) Borman, S. T. U. *Chem. Eng. News* **1994**, *72*, 32-34.

(113) Wuts, P. G. *Curr. Opin. Drug Disc.* **1998**, *1*, 329-37.

(114) Kingston, D. G. I. *J. Nat. Prod.* **2000**, *63*, 726-734.

(115) Cragg, G. M.; Schepartz, S. A.; Suffness, M.; Grever, M. R. *J. Nat. Prod.* **1993**, *56*, 1657-1668.

(116) Ketchum, R. E. B.; Gibson, D. M. *Plant Cell Tiss. Org.* **1996**, *46*, 9-16.

(117) Ketchum, R. E. B.; Gibson, D. M.; Croteau, R. B.; Shuler, M. L. *Biotechnol. Bioeng.* **1999**, *62*, 97-105.

(118) Harbaugh, D. T.; Baldwin, B. G. *Am. J. Bot.* **2007**, *94*, 1028-1040.

(119) Page, T.; Southwell, I.; Russell, M.; Tate, H.; Tungon, J.; Sam, C.; Dickinson, G.; Robson, K.; Leakey, R. R. B. *Chem. Biodivers.* **2010**, *7*, 1990-2006.

(120) Subasinghe, S. M. C. U. P. *J. Trop For. Env.* **2013**, *3*, 1-8.

(121) Grover, J. K.; Yadav, S.; Vats, V. *J. Ethnopharmacol.* **2002**, *81*, 81-100.

(122) Ochi, T.; Shibata, H.; Higuti, T.; Kodama, K.-h.; Kusumi, T.; Takaishi, Y. *J. Nat. Prod.* **2005**, *68*, 819-824.

(123) Prema, B. R.; Bhattacharyya, P. K. *Appl. Microbiol.* **1962**, *10*, 529-531.

(124) Ishida, T. *Chem. Biodivers.* **2005**, *2*, 569-590.

(125) Jones, C. G.; Moniodis, J.; Zulak, K. G.; Scaffidi, A.; Plummer, J. A.; Ghisalberti, E. L.; Barbour, E. L.; Bohlmann, J. r. *J. Biol. Chem.* **2011**, *286*, 17445-17454.

(126) Okugawa, H.; Ueda, R.; Matsumoto, K.; Kawanishi, K.; Kato, A. *Phytomedicine* **1995**, *2*, 119-126.

(127) Zhang, X.; Dwivedi, C. *Front. biosci.* **2011**, *3*, 777-87.

(128) Santha, S.; Dwivedi, C. *Photochem. Photobiol.* **2013**, *89*, 919-926.

(129) Paulpandi, M.; Kannan, S.; Thangam, R.; Kaveri, K.; Gunasekaran, P.; Rejeeth, C. *Phytomedicine* **2012**, *19*, 231-235.

Chapter 2

RNA isolation and transcriptome sequencing for screening of genes involved in santalene biosynthesis

Chapter 2

RNA isolation and transcriptome sequencing for screening of genes involved in santalene biosynthesis

RNA sequencing is a quick and cost effective method to profile complete coding sequence of a genome due to its high throughput accuracy and reproducibility. Isolation of intact and functional RNA from interface of heartwood and sapwood from Indian Sandalwood *Santalum album* Linn is extremely difficult as it contains high percentage of polysaccharides, phenolics and other secondary metabolites. The presence of phenolics and polysaccharide compounds complicates the RNA isolation as they affect the purity as well the yield of the RNA. In this chapter, a protocol is developed for the isolation of high quality of total RNA from the interface of heartwood and sapwood from *S. album*. High quality total RNA obtained using modified protocol was used for RNA sequencing using Illumina GAII Analyzer and screening for genes involved in terpene biosynthesis in Indian Sandalwood.

2.1 Introduction

The endangered**,** Sandalwood or Chandan (*Santalum album* L.) belongs to the *Santalaceae* family, a medium-sized evergreen hemi-root parasitic tree, and is highly valued for its oil[1,2], which is synthesized at the interface of heartwood and sapwood and subsequently stored in heartwood. It is believed that expression of genes involved in the biosynthesis of santalenes and their hydroxy derivatives santalols, occurs at interface of heartwood and sapwood. Isolation of total RNA from interface of heartwood and sapwood from Sandalwood is very challenging as it contains high percentage of polysaccharides, phenolics and other secondary metabolites[3,4]. Moreover, phenolic compounds are found in a wide range of polymerization states, including lignans, stilbenes, flavonoids and quinones, with the latter two associated with the colour change occurring during heartwood formation. There are several protocols[5-7] available for the isolation of total RNA from tissues rich in phenolic and polysaccharide compounds but most of them are tissue specific. However, the extraction of total RNA from plant tissues often requires modification in existing procedure or development of new procedure. The presence of phenolics and polysaccharides complicate RNA isolation from tissue rich with these molecules, as they tend to co-precipitate with nucleic acids giving a brownish colour to the RNA pellet, decreasing its solubility immensely and further leading to the degradation of RNA. The quality of RNA is very important for downstream processing because these phenolics and polysaccharides act as the inhibitors for reverse transcriptase, as well as they interfere in PCR reaction. In this study, a protocol was developed for the isolation of high quality total RNA from the interface of heartwood and sapwood from *S. album*, with extensive modifications of the protocol reported for isolation of high quality of RNA from gymnosperm and angiosperm tissue[7].

RNA sequencing has become a powerful tool for the complete profiling of coding sequences of the genome[8]. Here we present the generation of massive amount of data from transcriptome sequencing using Next-generation sequencing (NGS) technology. Assembled data set was compared with NCBI database\Protein Data Bank (PDB) for screening for unigenes involved in terpenoid biosynthesis in Sandalwood, specifically prenyl transferase, terpene synthases and CYP450 systems.

2.2 Materials and Methods

2.2.1 Plant material

Sandalwood tissue samples were collected using wood-borer (5.15mm diameter) at a height of about 1 meter from the ground from a 15 year old *Santalum album* tree and the tissues were immediately flash-frozen in liquid nitrogen and stored at -80 °C till further use.

2.2.2 Reagents

All the reagents and chemicals were procured from Sigma-Aldrich or Invitrogen, All the solutions were prepared in 0.1% diethylpyrocarbonate (DEPC) and sterilezed in an autoclave, except Tris buffer, which was prepared in sterile DEPC treated water. All the plastic wares were soaked in freshly prepared DEPC water, dried, and sterilized in autoclave before use.

Extraction buffer

Lithium dodecyl sulphate (1.5 %), Lithium Chloride (300 mM), EDTA disodium salt (20 mM), Sodium Deoxycholate (1 % w/v), Tergitol nonidet NP-40 (1 % v/v). The buffer containing these constituents was autoclaved and the Tris-HCl (200 mM) pH 8.5 was added thereafter that and following constituents were added just before use: Thiourea (5 mM), Aurintricarboxylic acid (1 mM), and DTT (10 mM).

PVPP (Polyvinylpolypyrrolidone)

Acetone (HPLC grade)

Chloroform/Isoamylalcohol (24:1; v/v)

3.3 M Sodium Acetate pH 6.1 (prepared in DEPC treated water and autoclaved)

5 M NaCl (prepared in DEPC treated water and autoclaved)

10 % CTAB (prepared in DEPC treated water and autoclaved)

Isopropanol

TE: 10 mM Tris-HCl; 1 mM EDTA; pH 8.0

10 M LiCl (prepared in DEPC treated water and autoclaved)

70 % Ethanol (prepared with DEPC treated autoclaved water)

DEPC treated water (DEPC from Sigma-Aldrich)

Centrifuges

Waterbath

2.2.3 Modified protocol for RNA isolation

- 1 g of plant tissue was ground in liquid nitrogen with 0.3 g of PVPP per gram of tissue in bead-beater at a frequency of 30 vibrations/second for 5 minutes.

Critical step

Important for minimizing the oxidation of phenolic compounds

- Finely crushed powder was transferred into tube and washed with acetone to remove phenolic contamination.

Critical step

This step is very important to remove the phenolics (we can see the yellow colour of supernatant).

- This was centrifuged at 3000 × *g* for 10 minutes at 4 °C.
- Pellet was dried and 20 mL extraction buffer/g of tissue was added.
- Vortexed at room temp for 10 minute.
- This was centrifuged at 3000 × *g* for 20 minutes at 4 °C.
- The supernatant was transferred into a fresh tube and kept on ice.
- To this, 2 mL 10 % CTAB solution was added at room temperature and incubated for 5 minutes at 60 °C, to remove the residual polysaccharides.

Critical step

This step is very important to remove the polysaccharide to avoid the co-precipitation, which later leads to the degradation of RNA.

- This mixture was extracted with Chloroform: Isoamyl alcohol (24:1) till the interface became clear and the supernatant was retrieved each time.

Critical step

Extraction should be performed till the clear interface is observed to avoid the protein contamination.

- The supernatant was transferred into another tube and to this, $1/9^{th}$ volume of 3.3 M Sodium acetate and 0.6 volume of ice-cold isopropanol were added.
- This was incubated at -20 °C for 3 hrs and then centrifuged at 14000 × *g* for 20 minutes at 4 °C and the supernatant was discarded.
- To the pellet, 1 mL of TE Buffer (10 mM) and 1 mL of 5 M NaCl were added and incubated on ice for 30 minutes, with periodic vortexing.
- This mixture was extracted with chloroform: isoamyl alcohol (24:1) and the supernatant was retrieved each time. This was repeated till the interface became clear.

Critical step

Extraction should be performed till the clear interface is observed to avoid protein contamination

- To the supernatant, 2.5 M final concentration of LiCl was added and incubated for overnight at -20 ºC.
- The RNA was pelleted down by centrifugation at 14000 × *g* for 30 minutes at 4 °C.
- The resultant pellet was washed with 70 % ethanol by centrifugation at 14,000 × *g* for 10 minutes at 4 °C.
- The pellet was air dried at room temperature and re-suspended in 50 µl DEPC treated water.

2.2.4 Quantification of total RNA

Total RNA was quantified using a spectrophotometer (NanoDrop, Thermo Scientific) by measuring optical density of isolated RNA in 10 mM TE buffer at a wavelengths of 230, 260, and 280 nm and purity was checked by comparing the ratio of 230/260 and 260/280. The integrity of total RNA was assessed by sharpness of rRNA (28S and 18S rRNA) on 1.5 % agarose gels and visualized by GelRed™ (Biotium).

2.2.5 RT-PCR

To test RNA quality, 1µg of total RNA was used for cDNA synthesis using RT kit from Promega according to manufacturer's instructions. PCR Primers were designed for amplification of 18S rRNA gene (Sa18S-F 5'TGACGGAGAATTAGGGTTCG3', and Sa18S-R 5'GTGCCAGCGGAGTCCTATAA3') from 18S rRNA gene sequence (Accession no L24416) reported from Sandalwood. The PCR mixture was initially denatured at 94 °C for 5 minutes and then subjected to PCR condition at 20 seconds at 94 °C, 20 seconds at 55 °C and 1 minute at 72 °C for 32 cycles, and a final extension for 10 minutes at 72 °C. Reaction product was analyzed by electrophoresis on a 1 % agarose gel and visualized by GelRed™ (Biotium).

2.2.6 Transcriptome sequencing

Transcriptome library for RNA sequencing was constructed according to the Illumina TruSeq RNA library protocol outlined in "TruSeq RNA Sample Preparation Guide" (Part # 15008136; Rev. A; Nov 2010) from Genotypic Pvt. Ltd. In briefly, mRNA was purified from 1 µg of intact total RNA using oligodT beads (TruSeq RNA

Sample Preparation Kit, Illumina). The purified mRNA was fragmented for 2 minute at elevated temperature (94 °C) in the presence of divalent cations and reverse transcribed with Superscript II Reverse transcriptase by priming with Random Hexamers. Second strand cDNA was synthesized in the presence of DNA polymerase I and RnaseH. The cDNA was cleaned up using Agencourt Ampure XP SPRI beads (Beckman Coulter). Illumina adapters were ligated to the cDNA molecules after end repair and addition of A base and SPRI cleanup was performed after ligation. The library was amplified using 8 cycles of PCR for enrichment of adapter-ligated fragments. The prepared library was quantified using Nanodrop and validated for quality by running an aliquot on High Sensitivity Bioanalyzer Chip (Agilent).

2.2.7 De novo transcriptome assembly

Total 10.08 GB paired end raw data was generated with read length of 150 bp and primary QC check of the raw data was performed using the inbuilt tool SeqQC-V2.1. To obtain high quality clean read data for *De novo* assembly, the raw reads were filtered by discarding the reads containing adaptor sequence and poor quality raw reads (Phred score <20). The clean reads were first assembled into contigs using the Velvet_1.1.05 with an optimized hash length of 59. Assembled contigs were given as input for Oasis_0.2.01 to generate transcripts. The redundancy in the output transcripts of Oasis_0.2.01 was removed by using CD-HIT to generate unique unigenes.

2.2.8 Transcriptome annotation

To assign molecular function, biological processes and cellular components of unigenes, functional annotations were performed. ORFs were predicted in all six frames by Virtual Ribosome online program. The longest ORFs were selected for each unigenes and submitted to Pfam-A database to identify protein domain and architecture. Unigenes assigned with Pfam ID were used to perform BLAST2GO search against NCBI Nr database, SwissProt/Uniprot database, Protein Data Bank (PDB) with an E-value$\leq 10^{-5}$. The FASTA format of all the unigenes were submitted to KEGG database to assign KO (KEGG Orthology) number and generate KEGG pathways.

2.3 Results and Discussion

2.3.1 Optimization of RNA isolation from the interface of heartwood and sapwood

Several standard protocols have been reported for isolation of total RNA from various plant tissue specialized for high phenolic and polysaccharide rich tissue, including CTAB-NaCl method[9], modified SDS/ Phenol method[10]. However, these protocols failed to yield high quality of total RNA from the interface of heartwood and sapwood from Sandalwood. Isolation of total RNA using CTAB-NaCl buffer, from the interface of heartwood and sapwood from Sandalwood was not efficient as the tissue is rich in polysaccharides and phenolics, SDS/Phenol method was also not effective to remove these molecules, which led to the degradation of RNA due to co-precipitation of polysaccharides and phenolics. High level of polyphenolic compounds interact irreversibly with DNA, RNA and proteins[11], leading to RNA degradation[12].

The RNA isolation protocol reported for RNA isolation from gymnosperm and angiosperm tissue[7] was not efficient for isolation of high quality of RNA. This method resulted in very low yield with partially degraded RNA (indicated by strong smear) and also was found to be contaminated with genomic DNA. Moreover, the 28s rRNA and 18S rRNA band are not clear (Figure 2.3.1). The isolated pellet was also brown in colour and difficult to dissolve. This may be due to browning effect[11], where phenolic compound, upon oxidation, develops a brown colour supernatant. The absorbance ratio $A_{260/280}$ and $A_{260/230}$ were found to be 1.12-1.52 and 0.42-0.65 respectively indicating contamination of protein, polyphenolics and polysaccharide. This quality of RNA was not suitable for the downstream processing for the discovery of genes involved in santalene biosynthesis.

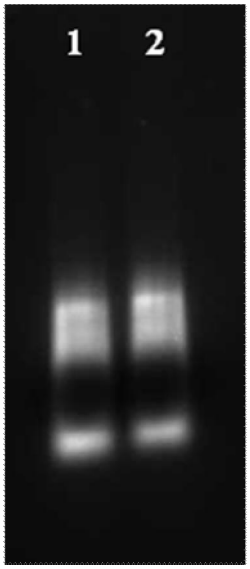


Figure 2.3.1: Agarose gel electrophoresis of total RNA isolated using protocol reported for RNA isolation from xylem tissue from gymnosperm, **Lane 1 and 2:** Total RNA $A_{260/280}$ ratio = 1.52, $A_{260/230}$ ratio = 0.65

To overcome these problems, the above protocol was extensively modified to isolate better quality of RNA. PVPP which is an inhibitor of polyphenol oxidase and can prevent browning effect[13,14] was mixed with tissue before crushing, so that it could bind to phenolic compounds and prevent their oxidation, which in turn could not bind to nucleic acids. The ground tissue mixture was washed with acetone/methanol to remove soluble phenolic contaminants rendering a significant improvement in RNA quality. CTAB treatment before chloroform:isoamyl alcohol extraction was extremely helpful in removing the polysaccharide contamination and avoiding its co-precipitation with nucleic acid.

Total RNA extracted from interface of heartwood and sapwood of Indian Sandalwood produced two distinct rRNA bands (28S rRNA and 18S rRNA) on a 1.5 % agarose gel electrophoresis in the ratio of 2:1 without degradation and also showed no genomic DNA contamination (Figure 2.3.2).

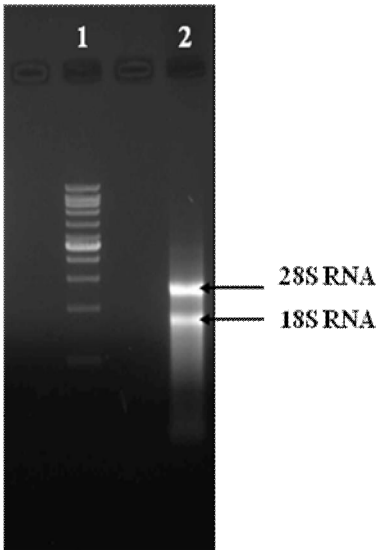


Figure 2.3.2: Agarose gel electrophoresis of RNA isolated from modified protocol, **Lane 1:** 1 Kb DNA ladder, **Lane 2:** Total RNA, $A_{260/280}$ ratio = 2.14, $A_{260/230}$ ratio = 2.35

The ratio of $A_{260/280}$ and $A_{260/230}$ were found to be 2.14 and 2.35 respectively, indicating no contamination of protein, polysaccharides, polyphenolics, salts, and solvents. The yield of total RNA ranges from 50-70 µg/g of tissue. To check the integrity of isolated total RNA, it was reverse transcribed as discussed earlier, and PCR amplification with 18 rRNA primer resulted in amplification of the expected 750 bp amplicon (Figure 2.3.3).

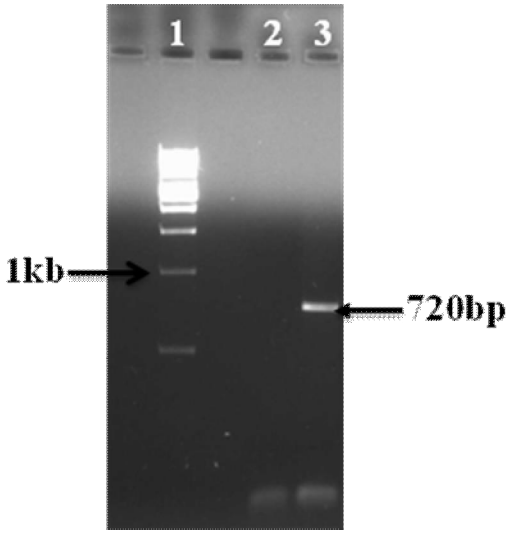


Figure 2.3.3: PCR amplification of 18S rRNA for cDNA verification, **Lane1:** 1 Kb DNA ladder, **Lane2:** no insert control, **Lane3:** 750 bp of 18S rRNA amplicon

In summary, our modified protocol is simple and highly effective for extraction of high quality total RNA from the wood tissue which is rich in polysaccharides,

polyphenolics and other secondary metabolites. The use of PVPP along with tissue crushing and washing with solvent renders a significant improvement in the quality of total RNA as compared to the other protocols. The optimized protocol enabled us to isolate high quality total RNA without degradation and secondary metabolite contamination, which was used for downstream processing such as reverse transcription, EST library preparation and gene amplification.

2.3.2 Transcriptome sequencing and screening of genes involved in terpenoid biosynthesis in Indian Sandalwood

2.3.2.1 cDNA library preparation and sequencing

cDNA library was constructed from the mRNA purified from total RNA isolated from the interface of heartwood and sapwood of Indian Sandalwood, and was subjected to RNA sequencing using Illumina GAII Analyzer. cDNA library was constructed as discussed earlier and amplified by PCR to enrich the adaptor ligated fragments. The cDNA library was sequenced on one lane of the flow cell using paired end sequencing. A total of 33323756 raw reads were generated with a length of 150 bp corresponding to 10.08 GB. Primary QC check for raw reads was performed using the inbuilt tool SeqQC-V2.1. From the observation of QC report, adapter trimming and low quality trimming was performed throughout the sequence to get better quality of reads.

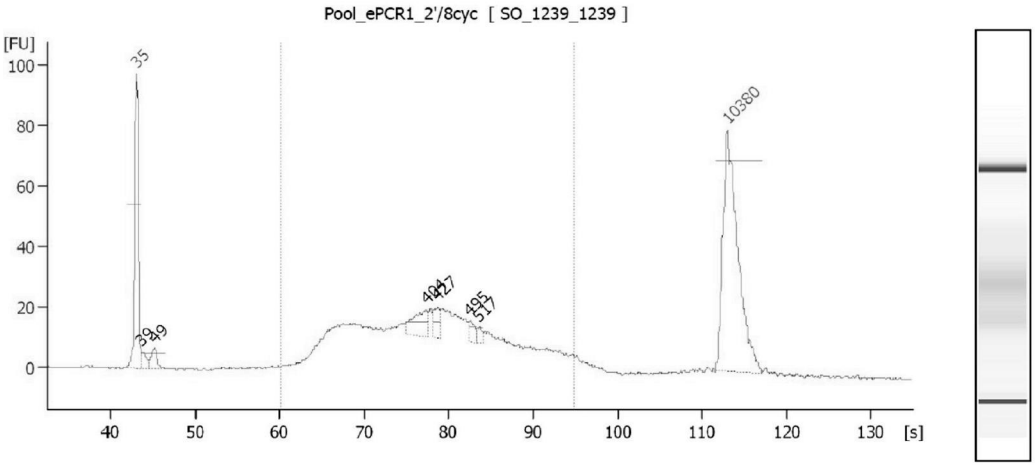


Figure 2.3.2.1: Bioanalyzer profile of amplified adaptor ligated fragments

2.3.2.2 *De novo* transcriptome assembly

Total 33323756 high quality paired end reads were generated with read length of 150bp. By discarding reads containing adaptor sequence, poor quality reads (Phred score <20), clean reads were preliminarily assembled into contigs by Velvet_1.1.05 with various hash lengths from 51 to 113. Total 58221 contigs were generated with optimised hash length of 59 with average contig length 571.6 bp and N50 value 863. These contigs were given as input for Oasis_0.2.01 to generate 90,478 transcripts having N50 value of 695 and average transcript length 474.18 bp. The total transcripts

were further subjected to cluster and assembly analysis using CD-HIT to remove the redundancy. Finally, it resulted in a total of 84,094 unique transcripts with an average size of 494.17 bp and N50 length 717, which contains 9136 transcripts (10.86 %) with length greater than 1 kb and 24912 transcripts (29.62 %) with length greater than 500 bp. The length distribution of transcripts is shown in fig 2.3.2.2. These results showed that the high throughput sequencing and assembly quality is good, which could be used for functional annotation.

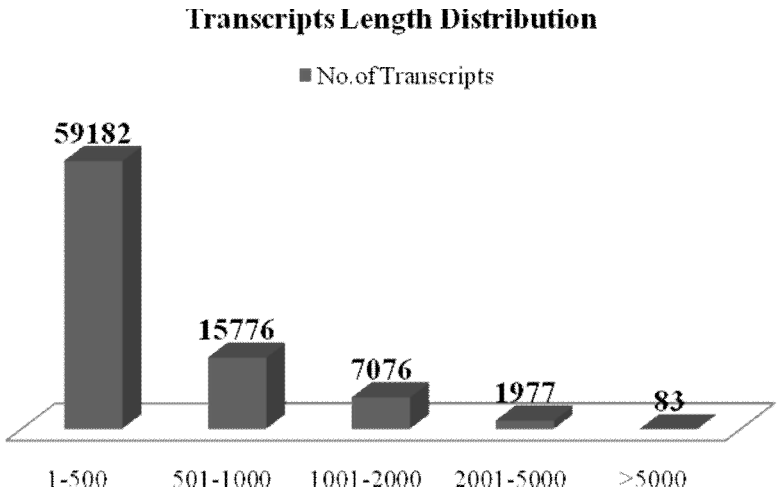


Figure 2.3.2.2: Transcript length distribution graph

Table 2.1: Transcriptome assembly statistics

	Velvet_1.1.05 (Oases 0.2.01) Transcripts Statistics		
	Contigs	**Total transcript**	**Unigenes**
Hash_length	59	59	55
Transcripts Generated :	58221	90478	84094
Maximum Transcript Length :	11726	12279	12279
Minimum Transcript Length :	117	100	100
Average Transcript Length	571.671	474.18	494.17
Total Transcripts Length :	33283265	42902769	41557227
Total Number of Non-ATGC Characters :	20	120683	112096
Percentage of Non-ATGC Characters :	6.00903e-07	0.00281294	0.00269739
Transcripts > 100 bp :	58221	90476	84092

Transcripts > 500 bp :	21371	25145	24912
Transcripts > 1 Kb :	8779	9187	9136
Transcripts > 10 Kb :	1	3	3
Transcripts > 1 Mb :	0	0	0
N50 value :	863	695	717

2.3.2.3 Functional annotation of unigenes

Functional annotation of unigenes allows for insight into particular molecular function and biological processes in which the putative proteins are involved. We applied various approaches to functionally annotate the assembled transcripts.

2.3.2.3.1 KASS analysis

To identify the terpenoid biosynthetic pathways present at the interface of heartwood and sapwood of Indian Sandalwood, all the 84,094 unigenes were submitted to KEGG database for functional annotation of genes by bidirectional BLAST comparison against manually curated KEGG (Kyoto Encyclopedia of Genes and Genomes) database[15].

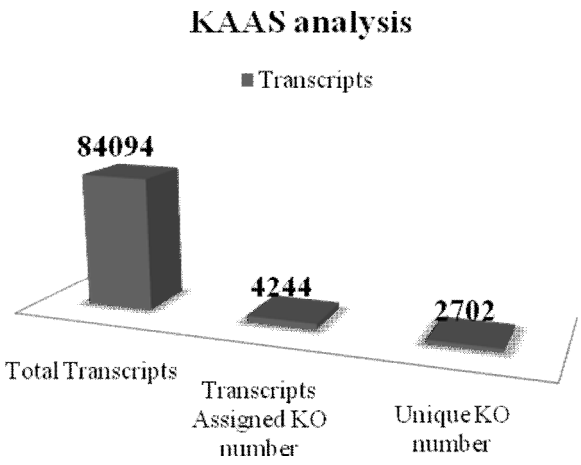


Figure 2.3.2.3: KAAS analysis of unigenes for KEGG pathway mapping

The results contains 4244 unigenes are assigned with KO number representing 298 KEGG pathways, from which 2702 unigenes represent unique KO number. The predicted pathways represented the majority of plant biochemical pathways including metabolism, cellular processes and genetic information processing. According to KEGG pathway mapping of our sequence dataset, 30 unigenes were found to be involved in terpenoid backbone biosynthesis and assigned with KO number.

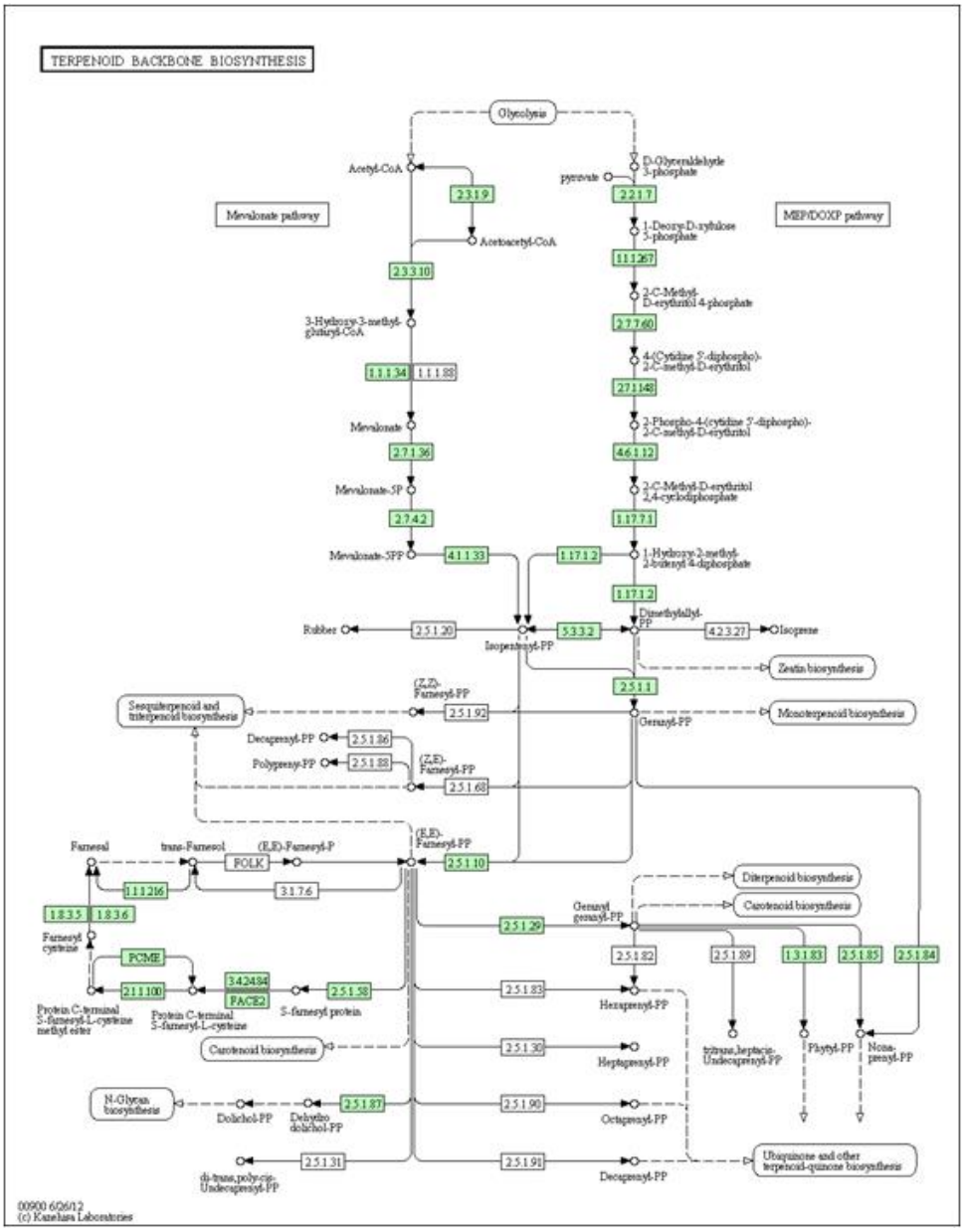


Figure 2.3.2.4: KEGG pathways map for terpenoid biosynthesis, green colour highlighted IDs represent the presence of these pathway enzymes in our unigenes

2.3.2.3.2 Pfam analysis

Pfam is a database of manually curated protein families which contains information about protein domain and protein families represented as multiple sequence alignments and as profile hidden Markov models[16]. For the functional annotation of unigenes, we made an Open Reading Frame (ORF) prediction analysis using online software Virtual Ribosome-V1.1. 84,094 unigenes were submitted to Virtual Ribosome to predict ORF of maximum length for each unigene in all six frames. Total 83,823 unigenes (99.66%) were identified as having ORF starting at an ATG

codon, out of which 17,119 unigenes (20.35 %) were found to have the ORF of ≥100 amino acid length without redundency, whereas 66697 (79.31 %) unigenes contains ORF ≤ 100 amino acid length and 278 (0.33 %) unigenes were found without having ORF in any of the frame (Figure 2.3.2.5).

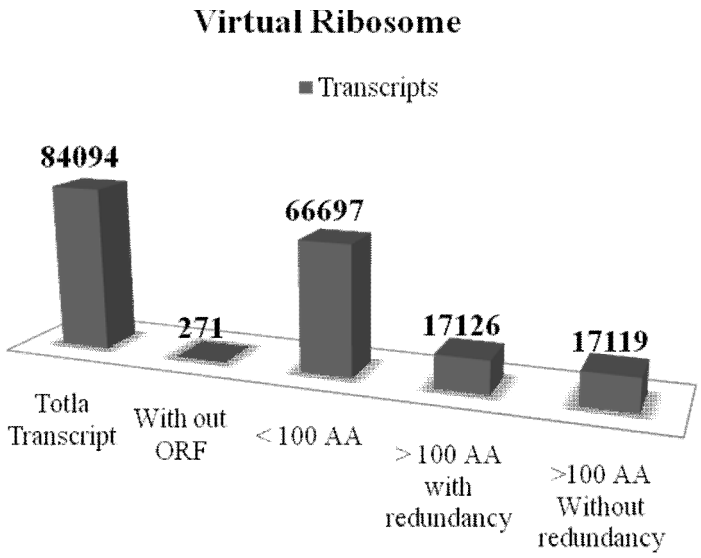


Figure 2.3.2.5: ORF prediction using Virtual Ribosome-V-1.1

17,119 unigenes having ORF of ≥ 100 amino acid were submitted to Pfam-A database, which allows efficient and sustainable manual curation of alignments and annotation (Figure 2.3.2.6). From the Pfam analysis only 10,668 unigenes were assigned with Pfam ID out of 17,119 unigenes. Further analysis of 10,668 unigenes having Pfam ID revealed that 18 unigenes belonged to terpene biosynthesis representing Pfam ID: PF01397 (terpene synthase N-terminal domain), PF03936 (terpene synthase family metal binding domain), PF00432 (prenyl transferase and squalene oxidase repeat) and PF13243 (prenyl transferase-like) were selected for screening of terpene synthases.

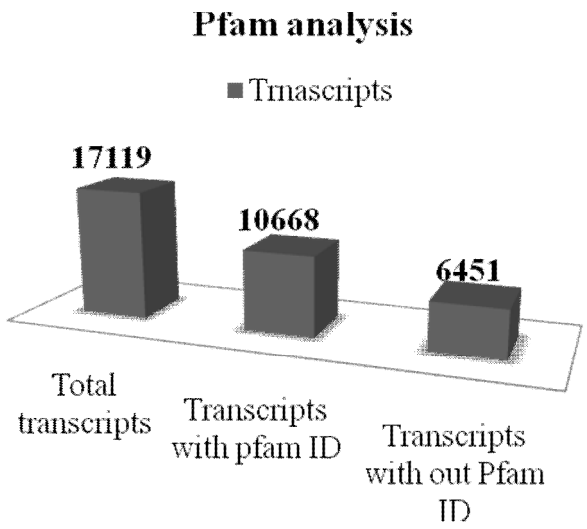


Figure 2.3.2.6: Pfam analysis of transcript having ORF ≥100 amino acid

10,668 unigenes assigned with Pfam ID were subjected to Blast2Go analysis against NCBI Nr-database, Swissprot/ Uniprot database to screen the genes involved in terpenoid biosynthesis in Indian Sandalwood.

2.3.2.3.3 BLAST2GO analysis

BLAST2GO is a tool often used in functional genomics research. It works by blasting assembled transcript sequences against protein database in NCBI and assign GO terms for functional characterization of the unigenes.

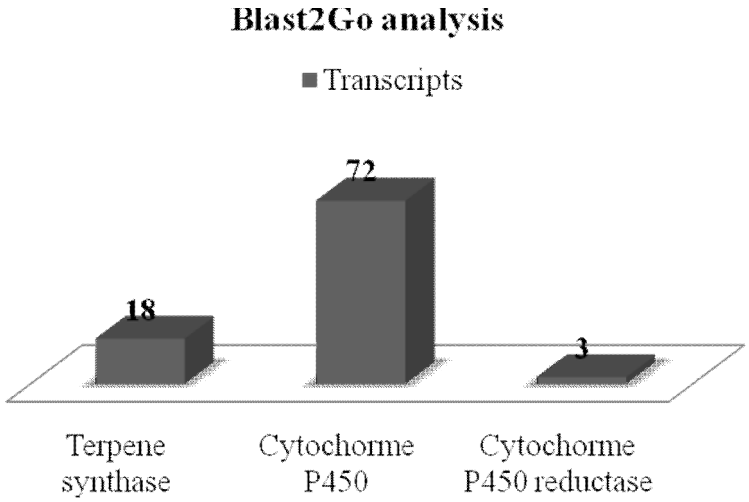


Figure 2.3.2.7: BLAST2GO analysis of unigenes assigned with Pfam ID for the screening of terpenoid biosynthetic pathway genes

BLAST2GO analysis resulted in the screening of 18 unigenes (Figure 2.3.2.7), which have terpene synthase domain and may be involved in biosynthesis of sesquiterpenes present in Indian Sandalwood. Further screening for CYP450 system resulted in 72 unigenes containing the domain of CYP450, from which 10 unigenes were matching with terpene hydroxylase. These CYP450 systems require a CYP450 reductase, which helps in the hydroxylation. Transcriptome screening resulted in identification of 3 transcripts shown to have the domain of CYP450 reductase were selected. These transcripts will be used for establishing the biosynthesis of santalols in Indian Sandalwood *S. album.*

2.4 Conclusion

In plant, high-throughput RNA sequencing accelerated the discovery of new genes, transcription pattern, and functional analysis. In the present work, we have optimized the protocol for the isolation of high quality of total RNA from the interface of heartwood and sapwood of Indian Sandalwood. RNA sequencing was performed on Illumina GAII Analyzer and 10.08 GB of sequence data file was generated. The raw data was assembled into 84,094 unique unigenes and screened against KEGG database, Pfam database, NCBI Nr-database, and Swissprot / Uniprot database for unigenes containing terpene synthase domain. These analyses resulted in identification of 18 unigenes related to terpene synthases, 72 unigenes for CYP450 mono-oxygenase and 3 unigenes representing CYP450 reductases. These unigenes were used for cloning and functional characterization of the genes involved in biosynthesis of santalene derivatives.

2.5 References

(1) Hongratanaworakit, T.; Heuberger, E.; Buchbauer, G. *Planta Med.* **2004**, *70*, 3-7.

(2) Kaur, M.; Agarwal, C.; Singh, R. P.; Guan, X.; Dwivedi, C.; Agarwal, R. *Carcinogenesis* **2005**, *26*, 369-380.

(3) Ochi, T.; Shibata, H.; Higuti, T.; Kodama, K.; Kusumi, T.; Takaishi, Y. *J. Nat. Prod.* **2005**, *68*, 819-824.

(4) Prema, B. R.; Bhattacharyya, P. K. *Appl Microbiol.* **1962**, *10*, 529-531.

(5) Zeng, Y.; Yang, T. *Plant Mol. Biol. Rep* **2002**, *20*, 417-417.

(6) Pandit, S.; Mitra, S.; Giri, A.; Gupta, V. *J. Plant Biol.* **2007**, *50*, 60-64.

(7) Kolosova, N.; Miller, B.; Ralph, S.; Ellis, B. E.; Douglas, C.; Ritland, K.; Bohlmann, J. *Biotechniques* **2004**, *36*, 821-824.

(8) Vijay, N.; Poelstra, J. W.; Künstner, A.; Wolf, J. B. W. *Mol. Ecol.* **2012**, *22*, 620-634.

(9) Jorge, A. R.-P. a.; Omar, Z.-P. r. *Isolation of total RNA from tissues rich in polyphenols and polysaccharides of mangrove plants*, 2011.

(10) Gao, J.; Liu, J.; Li, B.; Li, Z. *Plant Mol. Biol. Rep.* **2001**, *19*, 185-186.

(11) Loomis, W. D. *Method enzymol.* **1974**, *31*, 528-44.

(12) Dabo, S. M.; Mitchell, E. D.; Melcher, U. *Anal. Biochem.* **1993**, *210*, 34-38.

(13) Figueiredo, S.; Albarello, N.; Campos Viana, V. *In Vitro Cell Dev. Pl* **2001**, *37*, 471-475.

(14) Chen, G. Y. J.; Jin, S.; Goodwin, P. H. *J. Phytopathol.* **2000**, *148*, 57-60.

(15) Moriya, Y.; Itoh, M.; Okuda, S.; Yoshizawa, A. C.; Kanehisa, M. *Nucleic Acids Res.* **2007**, *35*, W182-W185.

(16) Finn, R. D.; Tate, J.; Mistry, J.; Coggill, P. C.; Sammut, S. J.; Hotz, H.-R.; Ceric, G.; Forslund, K.; Eddy, S. R.; Sonnhammer, E. L. L.; Bateman, A. *Nucleic Acids Res.* **2008**, *36*, D281-D288.

Chapter 3

Screening, cloning, expression and characterization of terpene synthases involved in biosynthesis of santalene derivatives

Chapter 3

Screening, cloning, expression and characterization of terpene synthases involved in biosynthesis of santalene derivatives

This chapter deals with the isolation and characterization of five terpene synthases ***(SaFDS*, *SaSS*, *SaβBS*, *SaTPS1* and *SaTPS2*)** involved in the biosynthesis of Farnesyl diphosphate (FPP), Santalenes, β-Bisabolene and Sesquisabinene B, respectively, in Indian Sandalwood, *S album.* There was not much information available on how many enzymes are involved in the formation of the sesquiterpene mixture in Indian Sandalwood, *S album.* Therefore, to characterize the genes involved in santalene biosynthesis, we selected five transcripts on the basis of sequence comparison to existing sesquiterpene synthase sequences in NCBI database and ORFs were isolated and cloned in expression vector for functional characterization. Protein expression was carried out in BL21DE3 (***SaFDS***) / Rosetta 2 DE3 (***SaSS*, *SaβBS*, *SaTPS1* and *SaTPS2***) cells and recombinant protein was purified to the homogeneity. Purified proteins were incubated with prenyl-diphosphate substrate for their functional characterization. When *SaSS,* was incubated with (*E,E*)-FPP resulted in the formation of mixture of six sesquiterpenes of which three major sesquiterpenes, identified as α-santalene, β-santalenes and exo-α-bergamotene, and three minor sesquiterpenes such as: *epi*- β-santalene, (*E*)-β-farnesene and exo-β-bergamotene. *SaβBS*, *SaTPS1* and *SaTPS2* upon incubation with (*E,E*)-FPP resulted in exclusive formation of β-bisabolene and sesquisabinene B. We also tested the bacterial production of sesquisabinene B from *SaTPS1* and *SaTPS2* by incorporating these genes in pET-Duet vector along with *SaFDS* and expressing the same in C41DE3 cells containing pRARE plasmid for rare codons. We observed that these sesquiterpenes could be produced in large quantity in genetically tractable heterologous systems.

3.1 Introduction

Isoprenoids constitute the largest class of naturally occurring secondary metabolites, which play diverse functional roles in all living systems. The basic units of isoprenoid biosynthesis in plants are IPP and DMAPP, which in turn are synthesized from either MVA (cytosol) or MEP (Plastid)[1,2] pathway, and undergo prenyl transferase catalyzed chain elongation reaction in a head to tail fashion to give linear prenyl diphosphates[3]. FDS is a key chain elongation enzyme in isoprenoid biosynthesis which catalyzes the condensation of one or two molecules of IPP with the allylic carbocation generated from allylic diphosphates, GPP or DMAPP respectively to form C_{15} molecules (FPP). This C_{15} derivative of prenyl diphosphate (*E,E*-FPP) is the precursor for most of the physiologically significant isoprenoid compounds such as: sesquiterpenoids, sterols[3-5], dolichols[6,7], heme a[8], prenylated proteins[8] and ubiquinone[9,10]. Till date, over 7000 sesquiterpene molecules with more than 300 stereochemically distinct hydrocarbon skeletons are reported. Sesquiterpenes, along with monoterpenes, have been traditionally used as flavor and fragrance ingredients, but they have the potential to serve as anticancer[11], antimalarial[12] and antiseptic agent. In recent years, sesquiterpenes of farnesene and bisabolene skeletons have been recognized as replacements for petroleum-derived jet-engine fuel[13,14].

The endangered, Sandalwood or Chandan (*Santalum album* L.) is highly valued for its fragrant heartwood oil. The essential oil of Indian Sandalwood is dominated by a blend of sesquiterpenes such as α- and β-santalenes, α- and β-santalols, α-bergamotene, α-bergamotol, sesquisabinene hydrate, curcumene, cedrene, sesquisabinene, bisabolol, β-bisabolene, farnesol etc. The sesquiterpene alcohols, i.e., (Z)-α- and (Z)-β-santalols are the major components (~80 %)[15-17] of the essential oil from a well matured tree (≥80 years old) and are responsible for most of the biological activity of the Sandalwood oil[18-22]. The first committed step in santalol biosynthesis is the cyclization of (*E,E*)-FPP catalyzed by santalene synthase to yield α-and β-santalenes along with other minor sesquiterpenes. Santalene synthase is the prime most enzyme present at the interface of heartwood and sapwood in *S. album* which catalyzes the divalent metal-ion dependent cyclization of (*E,E*)-FPP into (+)-α-santalene, (-)-α-exo-bergomotene, (+)-*epi*-β-santalene, and (-)-β-santalene as the major sesquiterpene hydrocarbons along with (*E*)-β-farnesene as a minor product[23]. Further, the cytochrome P450 mono-oxygenase system carries out hydroxylation at

the *cis* methyl group of the side chain to form (*Z*)-α-santalol, (*Z*)-α-exo-bergomotol, (*Z*)-epi-β-santalol, and (*Z*)-β-santalol.

In this chapter, we present the isolation, cloning and characterization of one prenyl transferase (*SaFDS*) and four sesquiterpene synthases (*SaSS*, *SaβBS*, *SaTPS1*, and *SaTPS2*) from the interface of heartwood and sapwood in Indian Sandalwood *S. album* to establish the biosynthesis of santalenes and their derivatives. In this study, we have also demonstrated the feasibility of metabolic engineering of *SaTPS1* and *SaTPS2* for the production of sesquisabinene in large quantity in heterologous systems.

3.2 Materials and Methods

3.2.1 Materials used in the study

3.2.1.1 Bacterial Strains and plasmids used in the study

Escherichia coli TOP10 (Invitrogen)

Escherichia coli BL21DE3 (Invitrogen)

Escherichia coli Rosetta2DE3 (Novagen)

TA cloning vector (Invitrogen)

pRSET expression vector (Invitrogen)

pET32 expression vector (Novagen)

pET28 expression vector (Novagen)

pET-Duet expression vector (Novagen)

Restriction enzymes (New England Biolabs)

3.2.1.2 Kits and reagent used in the study

cDNA synthesis kit (Promega)

GeneRacer Kit (Invitrogen)

JumpStart taq DNA polymerase (Sigma-Aldrich)

Long-Accurate taq DNA polymerase (Sigma-Aldrich)

Accuprime taq DNA polymerase (Invitrogen)

GenElute™ PCR cleanup kit (Sigma-Aldrich)

GenElute™ Gel extraction kit (Sigma-Aldrich)

GenElute™ Plasmid Miniprep Kit (Sigma-Aldrich)

GelRed™ (Biotium)

3.2.1.3 Buffer composition

Lysis buffer

50 mM NaH_2PO_4, 300 mM NaCl, pH 7.4, 10 % glycerol, containing 1 mg/mL lysozyme

Wash buffer

50 mM NaH_2PO_4, 300 mM NaCl, 30 mM Imidazole, pH 7.4, 10 % glycerol

Elution buffer

50 mM NaH_2PO_4, 300 mM NaCl, 250 mM Imidazole, pH 7.4, 10 % glycerol

Desalting buffer

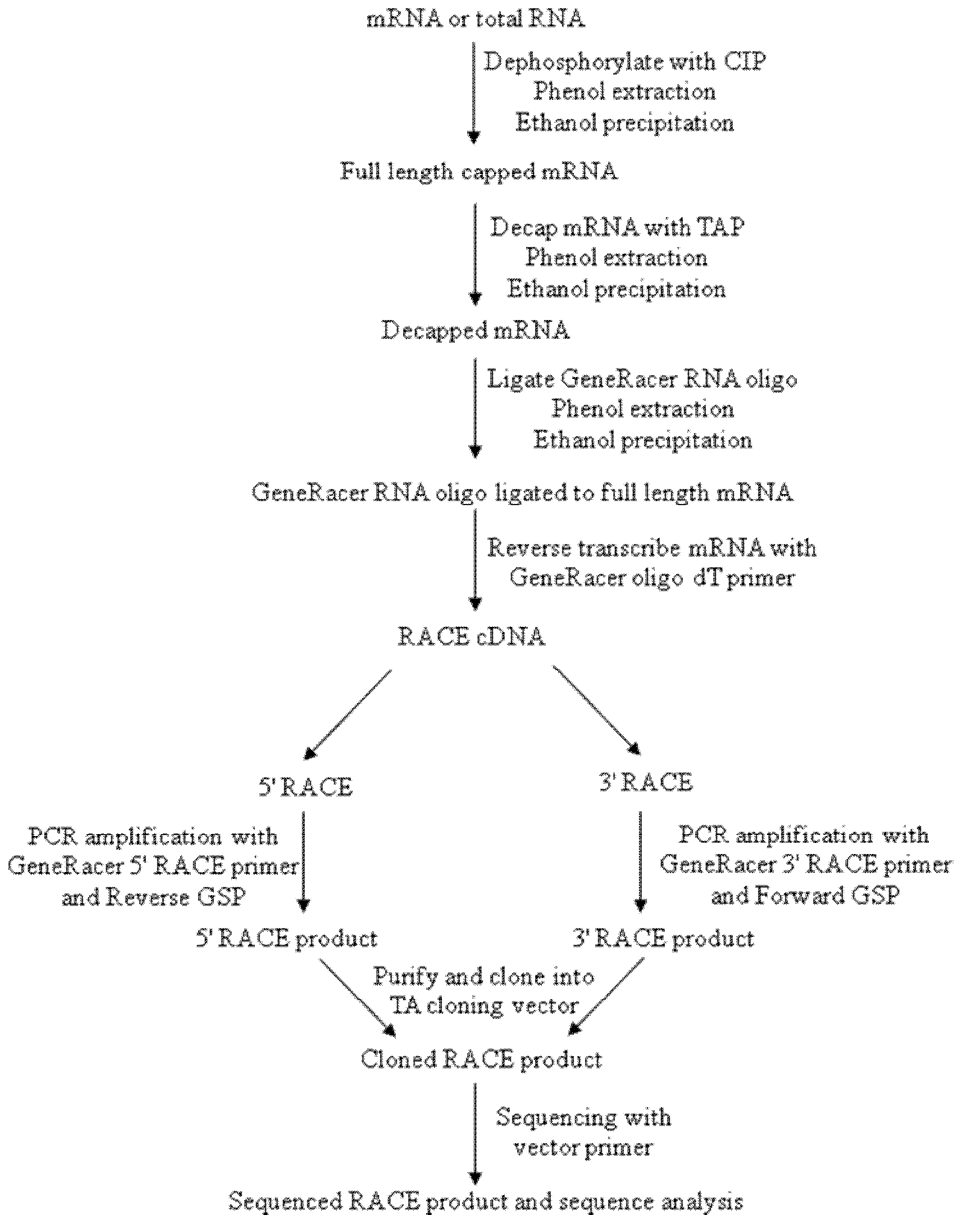
25 mM HEPES, 100 mM KCl, 10 % glycerol, pH 7.4

3.2.2 RNA isolation and cDNA preparation

As discussed in Chapter 2, total RNA was isolated from the interface of heartwood and sapwood. 1 µg of total RNA was used for first strand cDNA synthesis using oligo-$(dT)_{18}$ primer and AMV-RT kit (Promega). The cDNA thus constructed, was stored at -20 °C till further use.

3.2.3 Rapid amplification of cDNA ends (RACE)

RACE provides a method to obtain full length sequences (5' and 3' ends) of cDNA using a known small fragment of expressed sequence tags (ESTs). This technique is based on RNA ligase mediated oligo-capping rapid amplification of cDNA ends methods. This results in the selective ligation of RNA oligo to the 5' ends of de-capped mRNA using T4 RNA ligase. RACE protocol is described below in the flowchart.



Scheme 3.2.1: Schematic representation of RACE procedure

3.2.3.1 Sequence of RNA Oligo and Oligo dT primer

3.2.3.1.1 RACE RNA Oligo sequence

5′-CGACUGGAGCACGAGGACACUGACAUGGACUGAAGGAGUAGAAA-3′
(44 bases)

3.2.3.1.2 Oligo dT primer sequence

5′-GCTGTCAACGATACGCTACGTAACGGCATGACAGTG$\mathbf{(T)_{18}}$-3′ (54 bases)

Table 3.2.1 RACE universal primer

RACE primer	5'-3' sequence
5′ RACE Universal Primer	CGACTGGAGCACGAGGACACTGA
5′ RACE Nested Universal Primer	GGACACTGACATGGACTGAAGGAGTA
3′ RACE Universal Primer	GCTGTCAACGATACGCTACGTAACG
3′ RACE Nested Universal Primer	CGCTACGTAACGGCATGACAGTG

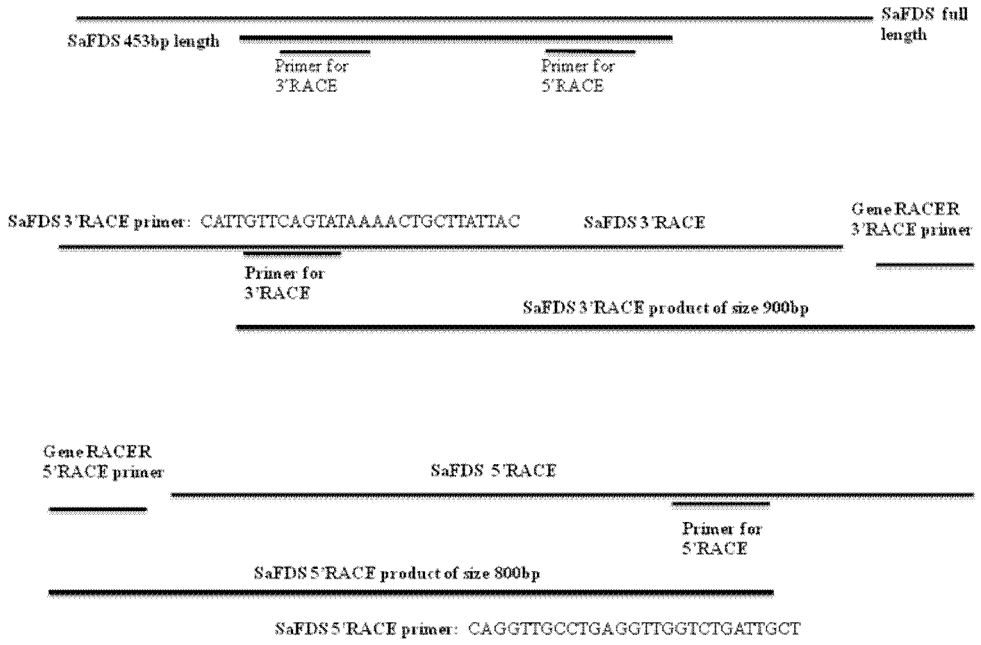
Table 3.2.2 Gene specific primers for RACE amplification

RACE primer	5'-3' sequence
5′ RACE *SaFDS*	TGCGATGGTGAATAGGCATTGAAT
3′ RACE *SaFDS*	TTCTTGTTCTCGATGATATTATGGA
3′ RACE *SaSS*	AGCGAATATGAACCCAACACTACTCA
5′ RACE *SaSS*	TCAGCAGCCCTTTTACGTTGTCACAAA
3′ RACE *SaβBS*	GTTCCAGAACAAGGAAACTGGCAAAT
3′ RACE *SaTPS1*	GGTTCCAGAACAAGGAAACTGGCAAAT
3′ RACE *SaTPS2*	CACATATGAAGAAGAAGAAGACGTGAACC

3.2.3.2 RACE PCR conditions and cloning

3.2.3.2.1 RACE PCR for *SaFDS*

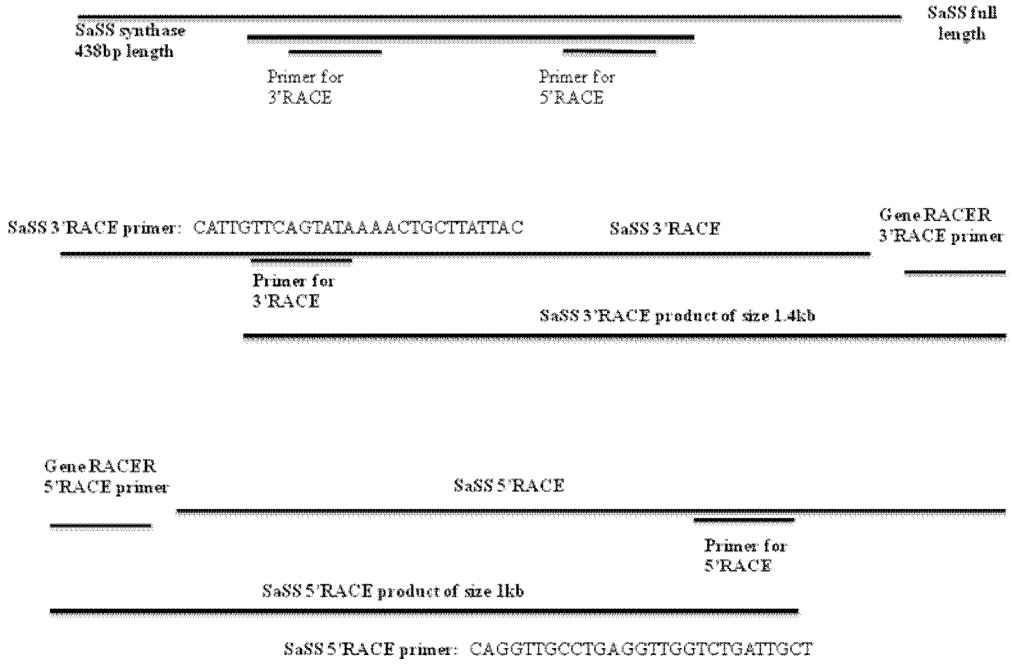
In order to extend partial cDNA fragments of farnesyl diphosphate synthase to full-length cDNA sequence, RACE PCR was performed using GeneRacer kit (Invitrogen) as per manufacturer's instruction. 3' RACE PCR for transcript 19031 was performed using 3' RACE *SaFDS* primer and GeneRacer 3' RACE primer using the following PCR program: 94 °C for 5 min, followed by 35 cycles of 94 °C for 20 sec, 58 °C for 20 sec, 72 °C for 1 min, and final extension at 72 °C for 10 min, whereas 5′ RACE PCR was performed using 5′ *SaFDS*-RACE primer and GeneRacer 5′ RACE primer using the PCR program: 94 °C for 5 min, followed by 35 cycles of 94 °C for 20 sec, 62 °C for 20 sec, 72 °C for 1 min, and final extension at 72 °C for 10 min. Both the 3′and 5′ RACE amplicon of expected size were purified from agarose gel and ligated into TA cloning vector to generate full-length sequence.



Scheme 3.2.2: Schematic representation of 5' and 3' RACE of *SaFDS* (>Locus_19031)

3.2.3.2.2 RACE PCR for *SaSS*

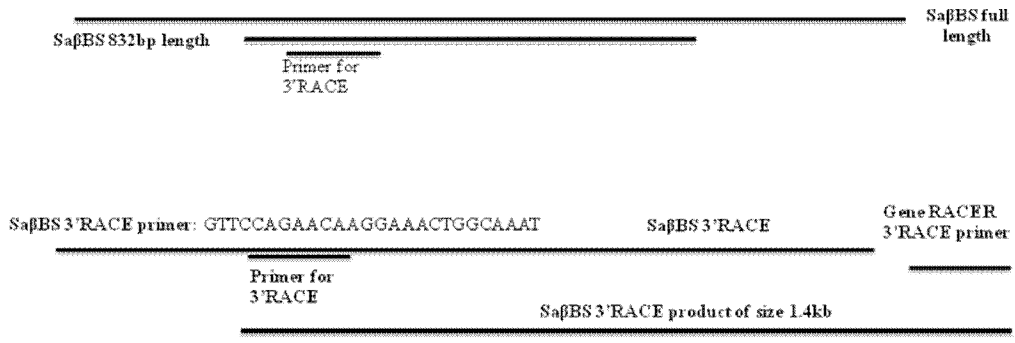
In order to extend partial cDNA fragments of Locus_1838_Transcript_1/1_Confidence_1.000 matching with terpene synthase in NCBI database to full-length cDNA sequence, RACE PCR was performed. 3' RACE PCR was performed using 3' RACE *SaSS* primer and GeneRacer 3' RACE primer using the following PCR program: 94 ℃ for 4 min, followed by 35 cycles of 94 ℃ for 20 sec, 60 ℃ for 20 sec, 72 ℃ for 1 min 30 sec, and final extension at 72 ℃ for 10 min. The expected size (1.5 kb) of RACE amplicon was purified from 1 % agarose gel and ligated in TA cloning vector (Invitrogen) according to manufacturer's instruction.



Scheme 3.2.3: Schematic representation of 5' and 3' RACE of *SaSS* (Locus_1838)

3.2.3.2.3 RACE PCR for *SaβBS*

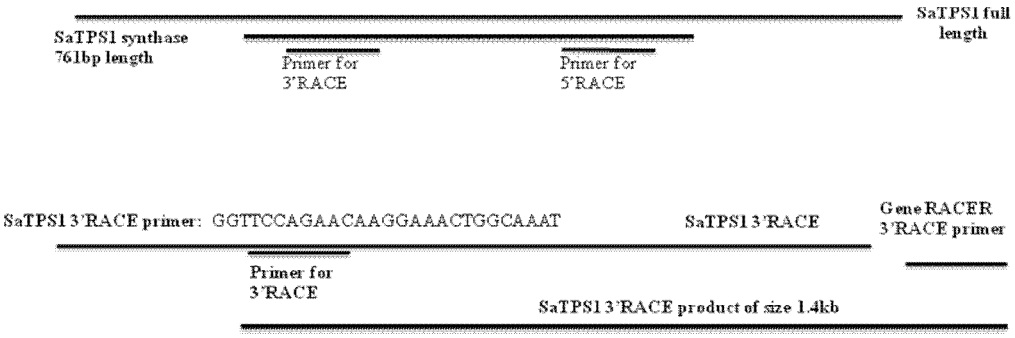
In order to extend partial cDNA fragments of β-bisabolene synthase (>Locus_5558_Transcript_1/1_Confidence_1.000_Length_832) to full-length cDNA sequence, RACE PCR was performed using GeneRacer 3' universal primer and *SaβBS* 3' RACE primer with RACE cDNA. PCR was performed at varying annealing temperatures to amplify 3′ RACE amplicon using the PCR program: 94 °C for 5 min, followed by 35 cycles of 94 °C for 20 sec, 58-64 °C for 20 sec, 72 °C for 1 min 30 sec, and final extension at 72 °C for 10 min. The expected size of RACE amplicon was purified from agarose gel and ligated in TA cloning vector to generate 3' RACE sequence.



Scheme 3.2.4: Schematic representation of 3' RACE scheme for *SaβBS* (>Locus_5558)

3.2.3.2.4 RACE PCR for *SaTPS1*

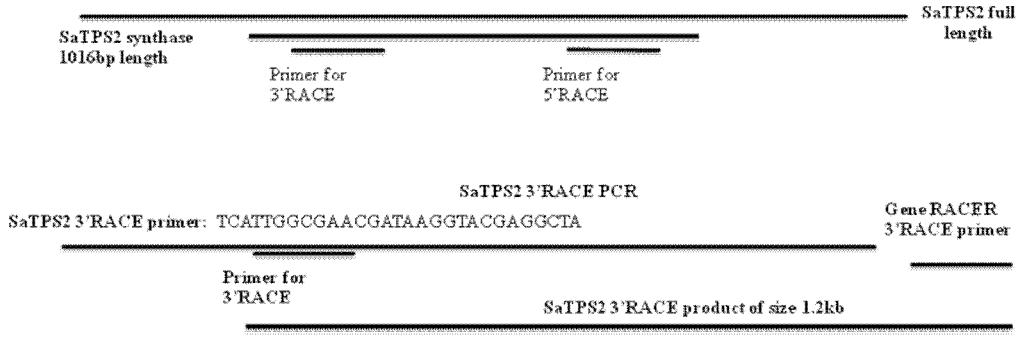
In order to extend partial cDNA fragments of sesquiterpene synthase 1 (Locus_33105_Transcript_1/1_Confidence_1.000_Length_761) to full-length cDNA sequence, 3' RACE PCR was performed using 3' RACE *SaTPS1* primer and GeneRacer 3' RACE primer. A gradient PCR program was used to amplify the 3' RACE amplicon by varying annealing temperature in the PCR program: 94 °C for 5 min, followed by 35 cycles of 94 °C for 20 sec, 60-64 °C for 20 sec, 72 °C for 1 min 30 sec, and final extension at 72 °C for 10 min. The expected size of RACE amplicon was purified from 1 % agarose gel and ligated in TA cloning vector (Invitrogen) according to manufacturer's instruction.



Scheme 3.2.5: Schematic representation of 3' RACE of *SaTPS1* (Locus_33105)

3.2.3.2.5 RACE PCR for *SaTPS2*

In order to extend partial cDNA fragments of sesquiterpene synthase 2 (Locus_8408_Transcript_1/1_Confidence_1.000_Length_1016) to full-length cDNA sequence, 3' RACE PCR was performed using 3' RACE *SaTPS2* primer and GeneRacer 3' RACE primer. A gradient PCR program was used to amplify the 3' RACE amplicon by varying annealing temperature at the PCR program: 94 °C for 5 min, followed by 35 cycles of 94 °C for 20 sec, 58-64 °C for 20 sec, 72 °C for 1 min 30 sec, and final extension at 72 °C for 10 min. The expected size of RACE amplicon was purified from 1 % agarose gel and ligated in TA cloning vector (Invitrogen) according to manufacturer's instruction.



Scheme 3.2.6: Schematic representation of 3' RACE of *SaTPS2* (Locus_8408)

3.2.4 Transformation and selection of positive clones

Ligation mixtures were transformed in TOP10 chemically competent cells and the transformants were selected on a plate containing the antibiotic ampicillin and X-gal for blue white screening. Colony PCR of white colonies were performed using M13 forward primer (5'-GTAAAACGACGGCCAG-3') and M13 reverse primer (5'-CAGGAAACAGCTATGAC-3'). Plasmids were isolated from the positive clones and sequenced with M13 forward and reverse primers.

3.2.5 Sequence analysis and ORF construction

After generating the missing sequence of RACE products, sequences were analyzed by NCBI GenBank database and overlapped with the EST fragments to generate full-length sequences. ORFs were selected using online ORF finder software for all the genes for their cloning and functional characterization. Full length ORF primers were designed containing RE sites at both the ends for all the genes for cloning in expression vector.

Table 3.2.3 Primer sequence for isolation of full length ORF of terpene synthases

Primer name	5'-3' Primer sequence
*SaFDS*_F	**GGATCC**ATGGGCGATCGGAAAACCAAAT
*SaFDS*_R	**CTGCAG**CTACTTCTGCCGCTTGTATATCT
*SaSS*_F	**GATATC**ATGGATTCTTCCACCGCCACC
*SaSS*_R	**GAGCTC**CTACTCCTCGCCGAGAGGAATAG
*Saβ-BS*_F	**GGATCC**ATGGATGCCTTTGCCACTTCTCCGACCT
*Saβ-BS*_R	**GCGGCCGC**TCAATCCTCCTCGTTCAGTGGAATAGGG
*SaTPS1*_F	**GATATG**ATGGATTTGTGTCAGATCCCGCCCACCTCT

*SaTPS1*_R	**GGATCC**TTACTCCTCATCTAGCGTAATTGGGTGAAT
*SaTPS2*_F	**GGATCC**ATGGCCTCTGTGATTGTTGAACCCATTCGT
*SaTPS2*_R	**CTCGAG**CTACTCTTCATTGAGTGGAATTGGATGGATC

Bold sequences are the Restriction site sequences used in cloning strategy

3.2.6 Full-length gene isolation and cloning into expression vector

3.2.6.1 Isolation and cloning of ORF of *SaFDS* in pRSETB expression vector

ORF of *SaFDS* was isolated from cDNA using full length ORF primers having RE site at both the ends (BamHI at 5' and PstI at 3') with Proof reading taq DNA polymerase (Invitrogen) using the PCR program: 94 ℃ for 5 min, followed by 35 cycles at 94 °C for 20 sec, 62 °C for 20 sec, 68 °C for 1 min 10 sec followed by final extension at 68 °C for 10 min. PCR product was purified from agarose gel using gel extraction kit and digested with BamHI and PstI restriction enzymes, pRSETB was also digested with same set of restriction enzymes. Digested fragments were purified from agarose gel and ligated using T4 DNA ligase (Invitrogen) by incubating overnight at 14 °C. Ligation mixture was transformed in TOP10 chemically competent cells, plated on LA containing 100 µg/mL of ampicillin and incubated overnight at 37 °C. Colony PCR was performed with T7 promoter primer and SaFDS reverse primer for the screening of positive clones. Plasmids were isolated from 5 positive clones of colony PCR and sequenced with T7 promoter and T7 reverse primers to check for the sequence in the correct vector frame.

3.2.6.2 Isolation and cloning of ORF of *SaSS* in pET32b expression vector

ORF of *SaSS* was isolated from cDNA using full length ORF primers having RE site at both the ends (EcoRV at 5' and SacI at 3') with LA (long accurate) taq DNA polymerase (Sigma-Aldrich) using the PCR program: 94 °C for 5 min, followed by 35 cycles at 94 °C for 20 sec, 60 °C for 20 sec, 68 °C for 1 min 50 sec followed by final extension at 68 ℃ for 10 min. The PCR product was purified from agarose gel using gel extraction kit and digested with EcoRV and SacI restriction enzyme along with pET32b expression vector. Both the digested fragments were purified from agarose gel and ligated using T4 DNA ligase (Invitrogen) by incubating overnight at 14 °C. The ligation mixture was transformed in TOP10 chemically competent cells, plated on LA containing 100 µg/mL of ampicillin and incubated overnight at 37 °C. Colony PCR was performed with *SaSS* forward primer and T7 reverse primer for the

screening of positive clones. Plasmids were isolated from 5 positive clones of colony PCR and sequenced with T7 promoter and T7 reverse primers to check for the sequence in the correct vector frame.

3.2.6.3 Isolation and cloning of ORF of *SaβBS* in pET28a expression vector

ORF of SaβBS was isolated from cDNA using full length ORF primers having RE site at both the ends (BamHI at 5' and NotI at 3') using Proof reading taq DNA polymerase (Invitrogen) using the PCR program: 94 ºC for 5 min, followed by 35 cycles at 94 ºC for 20 sec, 62 ºC for 20 sec, 68 ºC for 1 min 50 sec followed by final extension at 68 ºC for 10 min. PCR product was purified from agarose gel using gel extraction kit and digested with BamHI and NotI restriction enzymes along with pET28a expression vector. Digested fragments were purified from agarose gel and ligated using T4 DNA ligase (Invitrogen) by incubating overnight at 14 ºC. Ligation mixture was transformed in TOP10 chemically competent cells and plated on LA containing 50 µg/mL kanamycin and incubated overnight at 37 ºC. Colony PCR was performed with T7 promoter primer and *SaβBS* reverse primer for the screening of positive clones. Plasmids were isolated from 5 positive clones of colony PCR and sequenced with T7 promoter and T7 reverse primers to check for the sequence in the correct vector frame.

3.2.6.4 Isolation and cloning of ORF of *SaTPS1* and *SaTPS2* in pET28a expression vector

ORF of *SaTPS1* and *SaTPS2* were isolated from cDNA using full length ORF primers having RE site at both the ends (NdeI at 5' and BamHI at 3' for SaTPS1, whereas BamHI at 5' and XhoI at 3' for SaTPS2) using Proof reading taq DNA polymerase (Invitrogen) using the PCR program: 94 ºC for 5 min, followed by 35 cycles at 94 ºC for 20 sec, 60-62 ºC for 20 sec, 68 ºC for 1 min 50 sec followed by final extension at 68 ºC for 10 min. PCR products were purified from agarose gel using gel extraction kit and digested with respective restriction enzymes incorporated at their ends. pET28a was also digested with same set of restriction enzymes. Digested fragments were purified from agarose gel and ligated using T4 DNA ligase (Invitrogen) by incubating overnight at 14 ºC. The ligation mixture was transformed in TOP10 chemically competent cells, plated on LA containing 50 µg/mL kanamycin and incubated overnight at 37 ºC. Colony PCR was performed with T7 promoter primer

and *SaTPS1*/*SaTPS2* reverse primers. Plasmid was isolated from 5 positive clones of colony PCR and sequenced with T7 promoter and T7 reverse primers to check for the sequence in the correct vector frame.

3.2.7 Heterologous expression and protein purification

3.2.7.1 Heterologous expression and protein purification of *SaFDS*

pRSETB harboring *SaFDS* was transformed in *E. coli* BL21 (DE3) chemically competent cells and plated on LA containing 100 µg/mL ampicillin and incubated overnight at 37 °C . A single colony was inoculated in 5 mL LB+Amp and incubated overnight at 37 °C at 200 rpm. 1 mL of overnight grown culture was transferred in 100 mL of LB+Amp media and incubated at 37 °C at 200 rpm till OD reached 0.6 at 600 nm. After that, the culture was induced with IPTG to a final concentration of 1 mM and further incubated at 30 °C for 6 hr. The culture was harvested by centrifugation at 4500 × *g* for 10 minutes at 4 °C and cell pellet was stored at -80 °C till further use. The cell pellet was suspended in lysis buffer and incubated on ice for 30 minutes. After that, it was sonicated using probe sonicator for 5 cycles (pulse on 30 sec, pulse off 30 sec, amplitude 70 %), lysed suspension was centrifuged at 10,000 × *g* for 20 minutes. Crude lysate was separated, mixed with Ni-NTA slurry and incubated at 4 °C for 1 hr on rocking. The mixture was loaded onto a protein purification column of 1 cm diameter and unbound fraction was collected. Column was washed with wash buffer till O.D. at 280 nm reached zero. Finally, it was eluted into fractions of 1 mL of elution buffer. Purified protein fractions were pooled together and desalted on HiPrep™ 26/10 desalting column in HEPES buffer using ACTA (GE Healthcare). All the fractions were loaded on 10 % SDS-PAGE to check for purity and integrity of purified *SaFDS*.

3.2.7.2 Heterologous expression and protein purification of *SaSS*

pET32b harboring *SaSS* was introduced in Rosetta 2 DE3 chemically competent cells and colonies were grown on LA containing 100 µg/mL ampicillin and 34 µg/mL chloramphenicol and incubated overnight at 37 °C. A single colony was inoculated in 5 mL LB+Amp+Cam and incubated overnight at 37 °C at 200 rpm. 1 mL of overnight grown culture was transferred to 100 mL of TB+Amp+Cam and incubated at 37 °C at 200 rpm till OD reached to 0.8 at 600 nm. After that, the culture was induced with 0.5 mM IPTG and incubated overnight at 16 °C. The culture was harvested by

centrifugation at 4500 × *g* for 10 minutes at 4 °C, frozen and stored at -80 °C till further use. The cell pellet was resuspended in lysis buffer (10 mL/g of cell pellet) containing 1 mg/mL of lysozyme, protease inhibitor cocktail and incubated on ice for 30 minutes. Sonication was carried out using probe sonicator for 5 cycles (pulse on 30 sec, pulse off 30 sec, and amplitude 70 %). Lysed suspension was centrifuged at 10,000 × *g* for 20 minutes at 4 °C. Crude lysate was mixed with Ni-NTA slurry and incubated at 4 °C for 1 hr on rocking. It was loaded onto a protein purification column of 1 cm diameter. After collecting the unbound fraction, the column was washed with wash buffer till O.D. at 280 nm reached zero (4 mL each time). Finally, it was eluted into six fractions of 0.5 mL elution buffer each. Purified protein fractions were pooled together and desalted on HiPrep™ 26/10 desalting column in HEPES buffer using ACTA (GE Healthcare). All the fractions were analyzed on 10 % SDS-PAGE to check for purity and integrity of purified *SaSS*.

3.2.7.3 Heterologous expression and protein purification of *SaβBS*, *SaTPS1* and *SaTPS2*

pET28a harboring *SaβBS*, *SaTPS1* and *SaTPS2* were introduced in Rosetta 2 DE3 chemically competent cells and colonies were grown on LA plate containing 50 µg/mL of kanamycin and 34 µg/mL of chloramphenicol and incubated overnight at 37 °C. A single colony was inoculated in 5 mL LB+Kan+Cam and incubated overnight at 37 °C at 200 rpm. 1 mL of overnight grown culture was transferred in 100 mL of TB+kan+cam and incubated at 37 °C at 200 rpm till O.D. reached 0.8 at 600 nm. After that, the culture was induced with 0.5 mM IPTG and incubated overnight at 16 °C. The culture was harvested by centrifugation at 4500 × *g* for 10 minutes at 4°C, frozen and stored at -80 °C till further use. Cell pellet was suspended in lysis buffer and incubated on ice for 30 minutes. After that it was sonicated using probe sonicator for 5 cycles (pulse on 30 sec, pulse off 30 sec, amplitude 70 %). The lysed suspension was centrifuged at 10,000 × *g* for 20 minutes at 4 °C. Clear lyaste was mixed with Ni-NTA slurry and incubated at 4 °C for 1 hr with rocking. It was loaded onto a protein purification column of 1 cm diameter. After collecting the unbound fraction, the column was washed with wash buffer (4 mL each time) till O.D. at 280 nm reached zero. Finally, it was eluted into eight fractions of 0.5 mL elution buffer. Purified protein fractions were pooled together and desalted on HiPrep™ 26/10 desalting

column in HEPES buffer using ACTA (GE Healthcare). All the fractions were loaded on 10 % SDS-PAGE to check for purity and integrity of the purified proteins.

3.2.8 Enzymatic characterization

3.2.8.1 Enzyme assay of prenyl transferase (*SaFDS*)

50 µg of purified and desalted *SaFDS* protein was incubated with 100 µM IPP and 100 µM of allylic diphosphate (GPP/DMAPP) in HEPES Buffer, pH 7.4 (25 mM HEPES, 10 % v/v Glycerol, 5 mM Dithiothreitol, 10 mM Magnesium chloride) in a final reaction volume of 500 µL, and incubated for 2 hr at 30 °C. After the incubation, 10 units of alkaline phosphatase was added and further incubated at 37 °C for 1 hr. The reaction mixture was extracted twice with 500 µL n-Hexane. The extracted volume was reduced to ~50 µL with a stream of dry nitrogen. Products were characterized by injecting 1 µL of concentrated sample in GC/GC-MS and comparing with the authentic standards.

3.2.8.2 Enzyme assay of terpene synthases (*SaSS*, *SaβBS*, *SaTPS1*, and *SaTPS2*)

50 µg of purified and desalted recombinant proteins of terpene synthases (*SaSS*, *SaβBS*, *SaTPS1*, and *SaTPS2*) were incubated with 100 µM FPP in HEPES Buffer, pH 7.4 (25 mM HEPES, 10 % v/v Glycerol, 5 mM Dithiothreitol, 10 mM Magnesium chloride) in a final reaction volume of 500 µL, and incubated for 2 hr at 30 °C. The reaction mixtures were extracted twice with 500 µL n-hexane. The extracted volume was reduced to ~50 µL with a stream of dry nitrogen. Products were characterized by injecting 1 µL of concentrated sample in GC/GC-MS.

3.2.9 GC/GC-MS analysis

1 µL of extracts of all the assay products were injected in GC equipped with a 30 m × 0.32 mm × 0.30 µm capillary column (HP-5, J & W Scientific) and a FID. The column was equilibrated at 70 °C followed by a temperature gradient from 70 °C to 170 °C at 5 °C/min, followed by a second temperature gradient of 15 °C/min from 170 °C to 180 °C and a final hold at 180 ºC for 5 min, when nitrogen was the carrier gas with a flow rate of 1 mL/min. Analysis by GCMS was performed under the similar conditions using a 30 m × 0.25 mm × 0.25 µm capillary column (HP-5 MS, J & W Scientific) with helium as the carrier gas at a flow rate of 1 mL. All the products formed were characterized by comparing the retention time and also by co-injection with purified standard samples in GC and GC-MS.

3.2.10 Kinetic characterization of terpene synthases (*SaSS*, *SaβBS*, *SaTPS1*, and *SaTPS2*)

Steady state kinetics of *SaSS*, *SaβBS*, *SaTPS1* and *SaTPS2* were performed using 5 µM of protein with varying substrate concentrations ranging from 0.5 µM to 40 µM (0.5, 1, 2.5, 5, 7.5, 10, 12.5, 15, 20 and 40), in HEPES buffer (25 mM HEPES, 10 % v/v Glycerol, 5 mM DTT, 10 mM Magnesium chloride) in a reaction volume of 500 µL for 5 min at 30 °C and 50 rpm. Reactions were quenched by adding 250 µL of saturated Barium hydroxide and 250 µL of 0.1 M Zinc sulphate, followed by vigorous vortexing. Reaction mixtures were extracted twice with 1 mL of n-hexane. The extracts were supplemented with 100 ng of Dodecane as an internal standard concentrated to ~50 µL with a stream of dry nitrogen. The product ratios were analyzed using Gas chromatography. The calculated amounts of products formed were used for the determination of K_m, V_{max}, K_{cat} and K_{cat}/K_m values using graph pad prism software.

3.2.11 Real time analysis of *SaTPS1*, *SaTPS2* and *SaβBS*

In order to carry out the semi-quantitative real time analysis for studying the transcription level of *SaTPS1*, *SaTPS2* and *SaβBS*, 3 µg of total RNA was used for the synthesis of cDNA using Superscript cDNA synthesis kit (Invitrogen) according to manufacturer's instruction. PCR was carried out in duplicates in a total volume of 10 µL reaction scale using Jump-start taq DNA polymerase (Sigma). PCR was setup for 18S rRNA using 10 µM of forward and reverse primers with 150 ng, 30 ng, and 15 ng of cDNA using the PCR program: 94 °C for 5 min, followed by 30 cycles at 94 °C for 20 sec, 55 °C for 20 sec, 72 °C for 1 min followed by final extension at 72 °C for 10 min. PCR was setup for *SaTPS1*, *SaTPS2* and *SaβBS* ORF using 10 µM of forward and reverse primer with 15 ng of cDNA using the PCR program: 94 °C for 5 min, followed by 35 cycles at 94 °C for 20 sec, 60 °C for 20 sec, 72 °C for 1 min 40 sec followed by final extension at 72 °C for 10 min.

Table 3.2.4: Primer sequence used for semi-quantitative real time PCR

Primer name	5'-3' Primer sequence
***Sa18S*_F**	TGACGGAGAATTAGGGTTCG
***Sa18S*_R**	GTGCCAGCGGAGTCCTATAA
***SaTPS1*_F**	ATGGATTTGTGTCAGATCCCGCCCACCTCT

***SaTPS1*_R**	TTACTCCTCATCTAGCGTAATTGGGTGAAT
***SaTPS2*_F**	ATGGCCTCTGTGATTGTTGAACCCATTCGT
***SaTPS2*_R**	CTACTCTTCATTGAGTGGAATTGGATGGATC
***Saβ-BS*_F**	ATGGATGCCTTTGCCACTTCTCCGACCT
***Saβ-BS*_R**	**C**TCAATCCTCCTCGTTCAGTGGAATAGGG

3.2.12 *In-vivo* production of *SaTPS1* and *SaTPS2*

To test the microbial production of *SaTPS1* and *SaTPS2* metabolites, *SaFDS* was sub-cloned in pETDuet vector frame using BamHI site in multiple cloning site 1. Positive clones were screened by colony PCR using pEtDuet upstream primer 1 and *SaFDS* reverse primer and sequenced to confirm the frame of *SaFDS*. Similarly *SaTPS1* was sub-cloned in multiple cloning site 2 of pETDuet-1:SaFDS using NdeI and KpnI, whereas *SaTPS2* was sub-cloned in multiple cloning site 2 of pETDuet-1:*SaFDS* using KpnI restriction site. Positive clones were screened for both the terpene synthases using pETDuet upstream primer 2 and *SaTPS1*/*SaTPS2* reverse primers, respectively and sequenced to confirm the frame. Dual expression vector construct pETDuet-*SaFDS*:*SaTPS1*/*SaTPS2* as well as empty vector control were transformed in C41 DE3 cells containing pRARE plasmid isolated from Rosetta 2 DE3 cells for RARE codons. Transformants were selected on selection media containing 100 µg/mL ampicillin and 34 µg/mL chloramphenicol by incubating overnight at 37 °C. Individual colonies were inoculated in 5 mL of LB+Amp+Cam and incubated overnight at 37 °C at 200 rpm. Overnight grown cultures were transferred to 100 mL of Terrific Broth containing same antibiotics and incubated at 37 °C till absorbance at 600 nm reached to 0.8. Cultures were then equilibrated at 20 °C for 1 hr and then induced with 0.5 mM IPTG and further incubated at same temperature for 14-16 hr. Thereafter, the cultures were harvested by centrifugation at 4500 × *g* and 4 °C for 10 minutes. Supernatant was supplemented with 10 µg of dodecane and extracted twice with 100 mL of n-hexane each time, whereas pellet was lysed by alkaline lysis (0.2 M NaOH) and extracted with n-hexane. Extracted samples were concentrated and analyzed by GC/GC-MS using the same conditions as described earlier.

3.2.13 Phylogenetic analysis

Phylogenetic tree was constructed for all the terpene synthases isolated from Sandalwood. Multiple sequence alignments were performed using software ClustalX 2.1 and phylogenetic tree was generated using the nearest neighbour joining method though ClustalX 2.1 and visualized using MEGA 6 software[24].

3.2.14 Large scale enzyme assay

For the characterization of *SaTPS1* metabolite, large scale enzyme assay was setup using 80 mg of purified and desalted protein and 1.6 mM FPP in HEPES buffer, pH 7.4 in a reaction volume of 12 mL of 5 tubes.

Protein (2 mg/mL)	= 8 mL
50 mM FPP	= 0.450 mL (1.6 mM)
HEPES Buffer (pH 7.4)	= 3.550 mL
Total	**= 12 mL**

Reaction mixtures were overlaid with 4 mL of n-hexane and incubated at 30 °C for 24 hr at 50 rpm. After incubation, reaction mixtures were pooled together and extracted twice with n-hexane and analyzed by GC/GCMS.

3.2.15 $AgNO_3$ silica column for purification of metabolites

$AgNO_3$ coated silica gel (230-400 mesh size) was prepared by adding silica gel into methanolic solution of silver nitrate (5 % w/v). Methanolic slurry was dried over rotary evaporator under reduced pressure and then in a hot air oven at 80 °C for 1 hr to get dry $AgNO_3$ coated silica gel. A column was packed (10 cm height and 1 cm internal diameter) using 5 cm of uncoated silica gel (230-400 mesh) and then 5 cm of silver nitrate coated silica gel. The column was washed with three column volumes of pet ether. Crude assay sample was loaded on top of the column and fractionation was started using pet ether as the elution solvent. Slowly ethyl acetate was added drop by drop and fractions were collected. All the fractions were analyzed using GC/GC-MS and the fractions containing pure product were pooled together. H^1, C^{13} and DEPT NMR spectral data was recorded in $CDCl_3$ at 500 MHz for the characterization of metabolite 1.

3.3 Results and Discussion

3.3.1 Isolation, cloning, expression and characterization of FPP synthase (*SaFDS*)

3.3.1.1 Screening, isolation and cloning of farnesyl diphosphate synthase (*SaFDS)*

From transcriptome sequencing and screening as discussed in previous chapter, one transcript (Locus_19031_Transcript_1/1_Confidence_1.000) with the length of 453 bp was identified using BLAST analysis with NCBI database, which matched with FPP synthase reported from *Panax quinquefolius* with 88 % identity at the amino acid level, lacking its 5' and 3' sequences. To obtain the full length cDNA sequence of *SaFDS*, 5' and 3' RACE were performed (Section 3.2.3.8.1). Both the RACE fragments with expected size of 800 bp and 900 bp were obtained and cloned in TA vector. Sequencing of both the RACE fragments with M13 forward and reverse primer revealed the homology with the reported FPP synthase from *Panax quinquefolius* with 88 % identity containing 70 bp 5' UTR and 296 bp of 3' UTR.

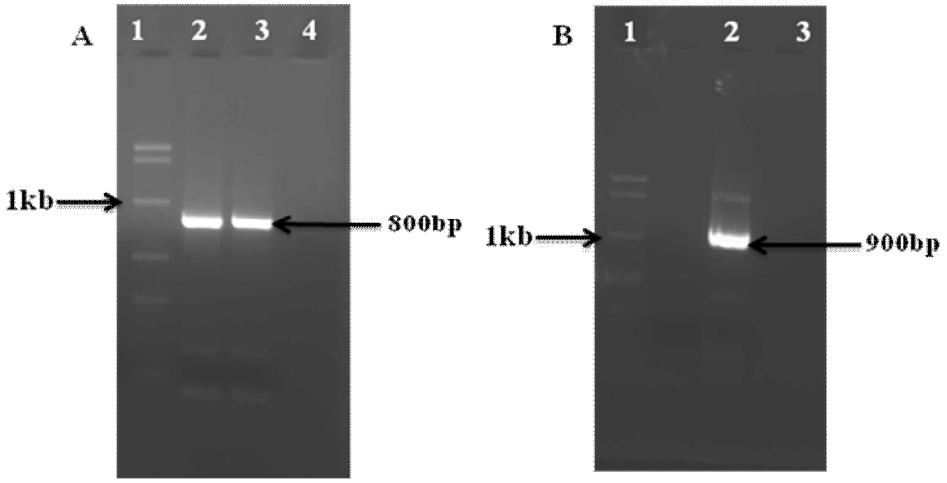


Figure 3.3.1.1: RACE amplification of *SaFDS* (>Locus 19031), **A)** *SaFDS* 3′ RACE PCR gel image, **Lane 1:** 1 Kb DNA ladder, **Lanes 2-3:** FPP 3′ RACE product, **Lane 4:** no insert control, **B)** *SaFDS* 5' RACE PCR gel image, **Lane 1:** 1 Kb DNA ladder, **Lane 2:** FDS 5′ RACE product, **Lane 3:** no insert control

The 5' and 3' RACE sequence of FPP synthase (*SaFDS*) were overlapped and full length sequences were generated. The full length ORF of *SaFDS* is composed of 1029 bp encoding a polypeptide of 342 amino acids with a calculated molecular weight of 39.4 kDa and pI (isoelectric point) of 5.26. The deduced amino acid

sequence of *SaFDS* was compared with prenyl transferases reported from other plant system and observed that it was matching with farnesyl diphosphate synthate from *Panax quinquefolius* (84 % identity), and farnesyl diphosphate synthase from *Humulus lupulus* (82 % identity). The analysis *of SaFDS* protein sequence revealed the presence of seven highly conserved regions (region I-VII) from which Region II (**L**, X_4LDDxxDxxxxRRG) and VI (GxxFQxxDDxxD....GK) are the most remarkable regions due to aspartate rich motifs and are involved in binding of both homoallylic (IPP) and allylic diphosphate substrates (GPP and DMAPP).

ORF of *SaFDS* was cloned in pRSETB vector frame with N terminal His_6 tag for affinity purification under the control of T7-RNA polymerase promoter for expression of active protein in *E.coli* BL21 (DE3). Positive clones were screened by colony PCR, followed by sequencing the positive clones with T7 promoter and T7 reverse primer.

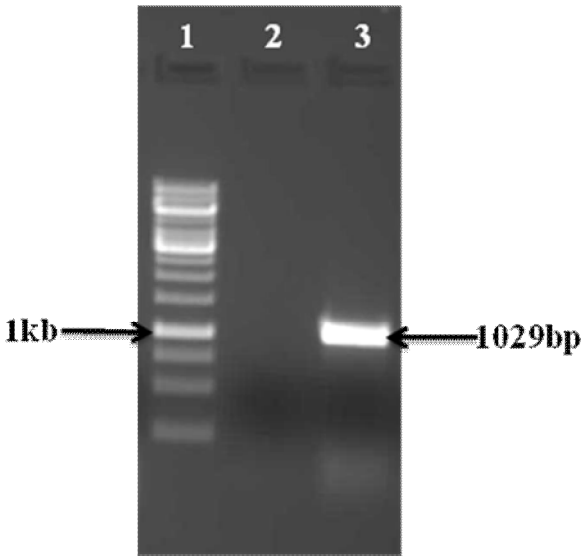


Figure 3.3.1.2: FPP Synthase full length ORF amplification, **Lane 1:** 1 Kb DNA ladder, **Lane 2:** no insert control, **Lane 3:** FPP synthase full length ORF

3.3.1.2 Bacterial expression and protein purification

pRSETB harboring the ORF of *SaFDS* was transformed in BL21(DE3) for active protein expression. Recombinant protein was over-expressed as discussed earlier and purified to the homogeneity using Ni-NTA resin. The recombinant protein was eluted at 250 mM imidazole, which gave a sharp band with ≥95 % purity at 40 kDa on a 10 % SDS-PAGE with the yield of 10-12 mg/L of bacterial culture. The result for

purification of *SaFDS* is shown in Figure 3.3.1.3. Pure fractions were desalted, flash-frozen and stored at -80 °C till further use.

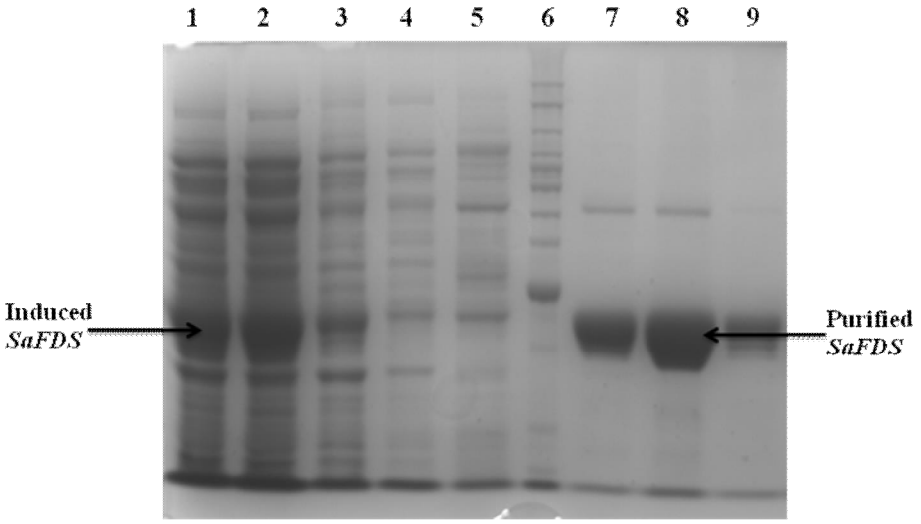
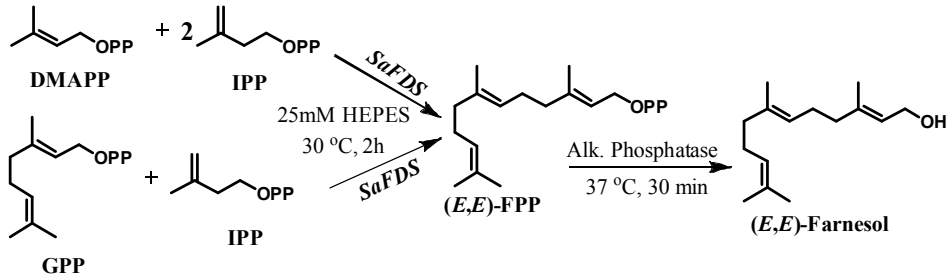


Figure 3.3.1.3: SDS-PAGE for FPP synthase (*SaFDS*) protein purification, **Lane 1:** Crude lysate pellet fraction, **Lane 2:** Crude lysate supernatant, **Lane 3:** Unbound fraction, **Lane 4-5:** Wash fraction, **Lane 6:** Protein marker, **Lanes 7-9:** Elution fractions

3.3.1.3 Enzymatic characterization of *SaFDS*

Farnesyl diphosphate (FPP) is the main branching point in terpene biosynthesis, where it acts as a precursor for biosynthesis of sesquiterpenes, sterols, dolichols, ubiquinones, and prenylated proteins. Functional characterization of *SaFDS* was carried out by incubation of purified and desalted *SaFDS* with allylic and homoallalic diphosphate substrate. Incubation of *SaFDS* with equimolar ratio of IPP and GPP or/ 1:2 molar ratio of DMAPP and IPP resulted in the formation of (*E,E*)-FPP which on subsequent treatment with alkaline phosphatase yielded the hydrolyzed product (*E,E*)-farnesol. The product from both the reactions was characterized by GC-MS fragmentation and co-injection of standard (*E,E*)-farnesol (Fig 3.3.1.4). A reaction mechanism for the formation of farnesol is proposed in scheme 3.3.1.



Scheme 3.3.1: Schematic representation of *SaFDS* assay

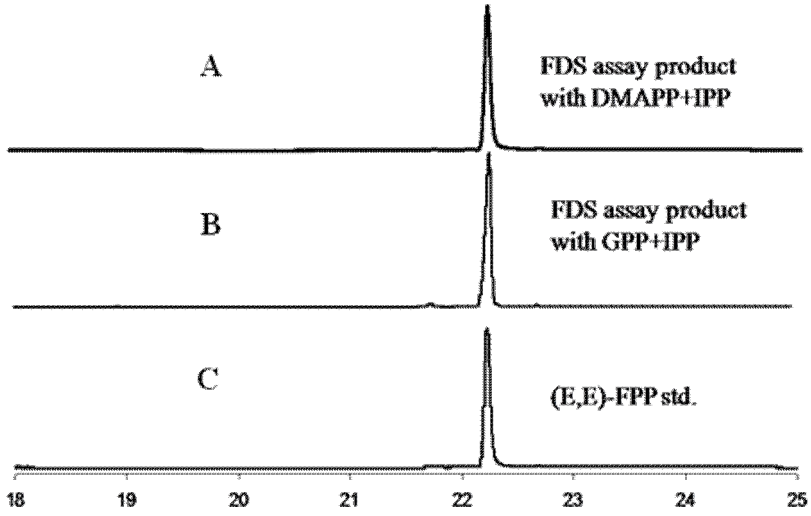


Figure 3.3.1.4: GC Profile of *SaFDS* assay, **A)** *SaFDS* assay product with DMAPP and IPP, **B)** *SaFDS* assay with GPP and IPP, **C)** (*E,E*)-farnesol standard

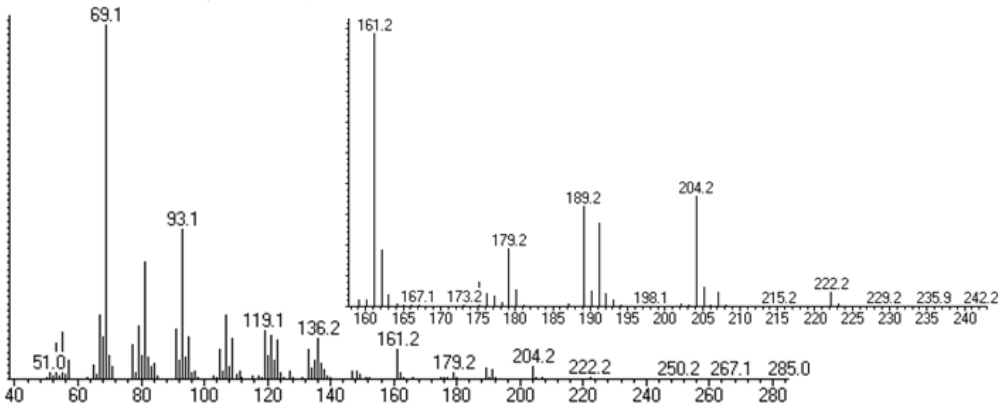


Figure 3.3.1.5: EI Mass Spectrum of (*E,E*)-Farnesol

3.3.2 Isolation, cloning, expression and characterization of santalene synthase (*SaSS*)

3.3.2.1 Screening, isolation and cloning of santalene synthase (*SaSS*)

From transcriptome screening, one transcript with the length of 438 bp was identified (Locus_1838_Transcript_1/1_Confidence_1.000), using BLAST analysis against NCBI database, which matched with limonene synthase reported from *Ricinus communis* with 55 % identity and lacking its 5' and 3' sequences. To obtain the full length cDNA sequence of *SaSS*, 5' and 3' RACE were performed as discussed earlier. 3' RACE fragments with expected size of 1.4 kb was obtained, cloned in TA vector and sequenced with M13 forward and reverse primer. Sequence analysis of 3' RACE fragment revealed the homology to santalene synthase reported from *Santalum album* with 100 % identity.

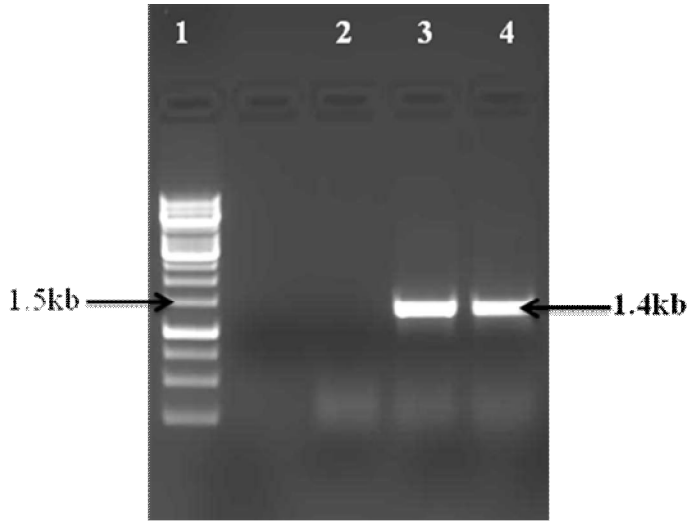


Figure 3.3.2.1: *SaSS* (Locus_1838) 3' RACE gel image, **Lane 1:** 1 Kb DNA ladder, **Lane 2:** no insert control, **Lanes 3-4:** Santalene Synthase 3' RACE amplicon

The full length *SaSS* is composed of 2223 bp having the ORF of 1710 bp encoding a polypeptide of 569 amino acids with a calculated molecular weight of 65.16 kDa and pI 5.63. The deduced amino acid sequence of *SaSS* was compared with terpene synthases reported from other plant systems and observed that it is matched with terpene synthases such as: limonene synthase reported from *Ricinus communis* (49 % identity at amino acid level), myrcene synthase from *Vitis vinifera* (42 % identity at amino acid level), and TPS10 from *Ricinus communis* (44 % identity at amino acid level). The *SaSS* protein shares several highly conserved residues with known sesquiterpene synthase from other plant species including DDxxD motif

(Asp^{321}, Asp^{322}, and Asp^{324}) involved in substrate binding, (D/**N**)Dxx(S/**T**)xxx**E** motif for metal binding, RRx_8W motif (Arg^{32}, Arg^{33}, and Trp^{42}) conserved at N-terminal. It lacks RxR which is conserved in terpene synthase and present at 35 amino acids upstream of DDxxD motif, but a similar RNN sequence is present. Biosynthesis of monoterpenes and diterpenes are compartmentalized in plastids, whereas sesquiterpenes are biosynthesized in cytosol. The predicted polypeptide sequence of SaSS lacks N terminal organelle targeting sequence, suggesting that the enzyme is directed to cytoplasm.

ORF of SaSS was amplified from cDNA as discussed in section 3.2.7.2 and cloned in pET32b vector frame between EcoRV and SacI restriction sites with N-terminal thioredoxin tag (Trx-tag) as solubilizing domain, His_6 tag for affinity purification, and S-tag for detection under the control of T7 RNA polymerase.

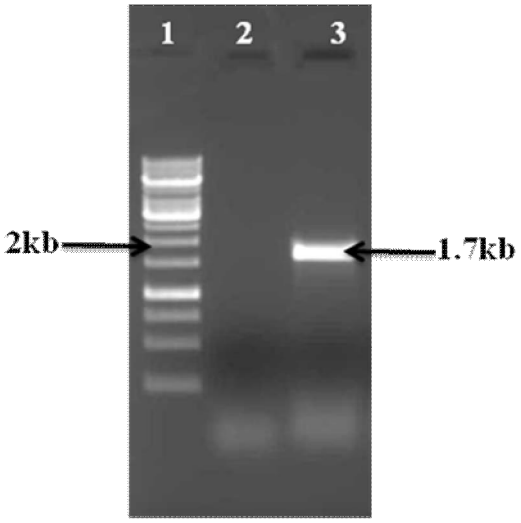


Figure 3.3.2.2: *SaSS* full length ORF amplification, **Lane 1:** 1 Kb DNA ladder, **Lane 2:** no insert control, **Lane 3:** *SaSS* ORF of 1.7 kb amplicon

3.3.2.2 Bacterial expression and protein purification

pET32b containing ORF of *SaSS* was introduced in Rosetta 2 DE3 cells for the expression of active protein. Rosetta 2 DE3 cells contain pRARE-plasmid for the expression of tRNA codons for arginine that are rare in *E. coli* and present more commonly in eukaryotes. Protein was over-expressed as discussed in materials and methods and purified to the homogeneity using Ni-NTA with the yield of 25-30 mg/L of bacterial culture. The recombinant protein was eluted at 250 mM imidazole, which gave a sharp band at 83 kDa (containing 18 kDa of N-terminal thioredoxin domain) of ≥90 % purity on a 10 % SDS-PAGE (Fig 3.3.2.3).

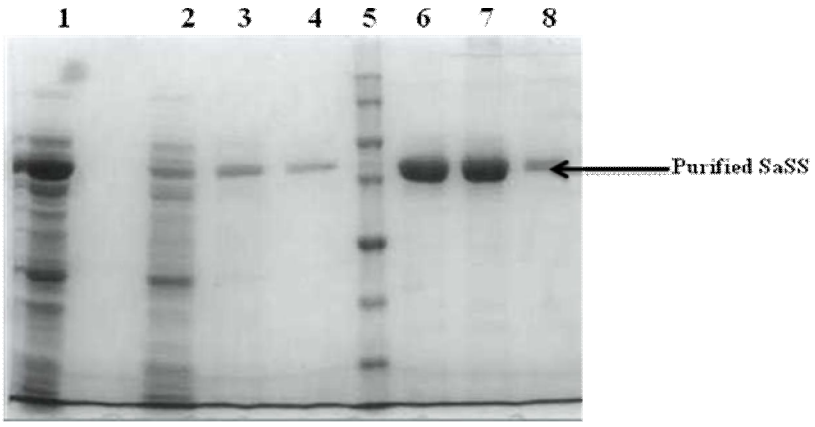


Figure 3.3.2.3: SDS-PAGE gel image of *SaSS* protein purification, **Lane 1:** Crude lysate of induced santalene synthase, **Lane 2:** Unbound proteins, **Lanes 3-4:** Wash fractions, **Lane 5:** Protein marker, **Lanes 6 -8:** Elution fraction 1-3

3.3.2.3 Enzymatic characterization of *SaSS*

For the functional characterization of recombinant *SaSS*, enzyme assay was performed using purified and desalted protein with (*E,E*) FPP in presence of 10 mM Mg^{2+}. GC and /GC-MS analyses of assay product mixture resulted in the formation of six sesquiterpenes (Figure 3.3.2.4): 1) α-santalene (41.22 %), 2) exo-α- bergamotene (21.62 %), 3) epi-β-santalene (4.36 %), 4) (*E*)-β-farnesene (1.13 %), 5) β-santalene (29.48 %), and 6) exo-β-bergamotene (2.19 %). All these products were confirmed by GC-MS fragmentation analyses (Figure 3.3.2.5 - 3.3.2.10) and co-injecting with purified terpene standards (Figure 3.3.2.15). Initial identification of **6** (exo-β-bergamotene) by comparison of its GC-MS fragmentation pattern with the terpenes from NIST mass spectral library was misleading as the EIMS fragmentation pattern showed similarity with that of farnesene. However, formation of exo-β-bergamotene as enzymatic product was confirmed by comparing the co-injection studies carried out with all farnesenes (Figure 3.3.2.15 J-K) as well as with a purified terpene obtained by large-scale enzymatic assay using mutant Y539W (Figure 3.3.2.15 L-M). This mutant produced metabolite **6** as one of the major products as discussed in chapter 4 (Section 4.3.1.1). All these six products were formed from single diphosphate substrate by generating carbocations which undergoes a sequence of intramolecular cyclizations involving a number of Wagner-Meerwein rearrangements to form single or multiple rings as proposed in scheme 3.3.2.1. The ability of santalene synthase to form multiple products from a single diphosphate substrate makes this enzyme moderately promiscuous in nature.

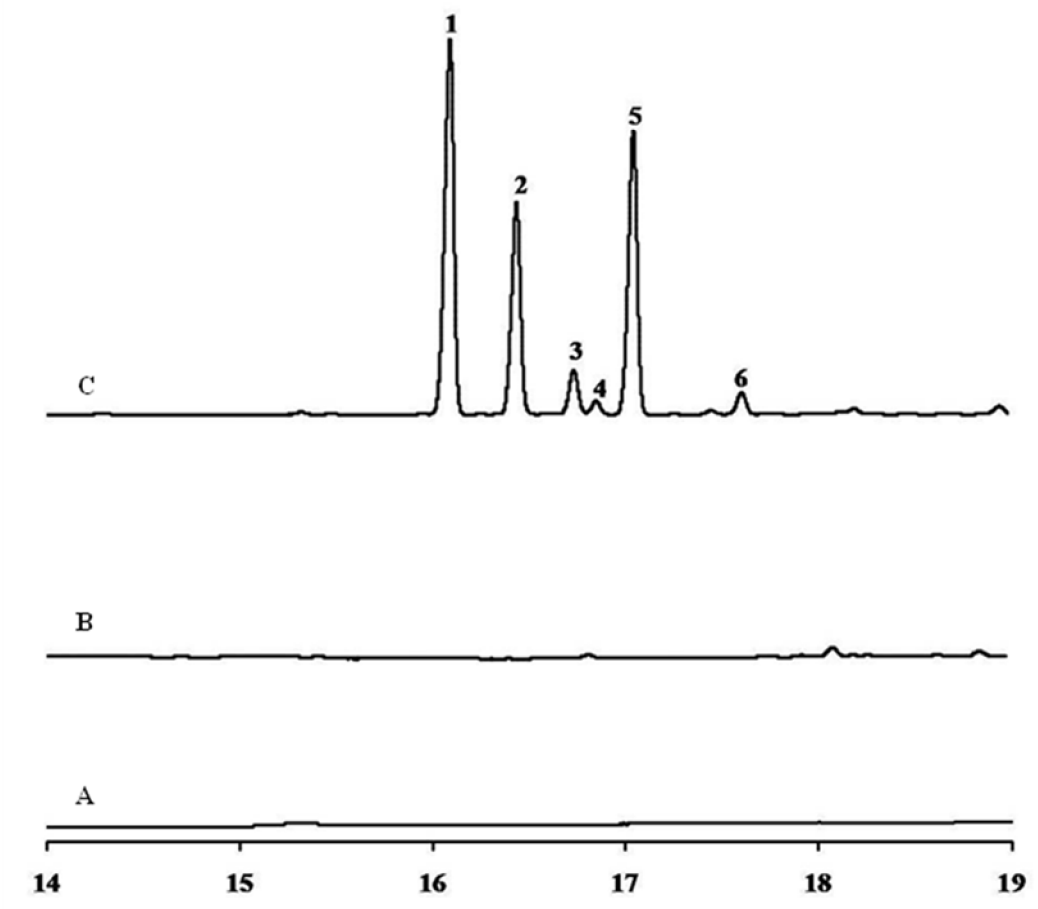


Figure 3.3.2.4: GC profile of *SaSS* enzyme assay with (*E,E*)-FPP, **A)** Empty vector control, **B)** Substrate control without enzyme, **C)** Assay of SaSS with (*E,E*)-FPP: α-santalene (**1**), exo-α-bergamotene (**2**), *epi*-β-santalene (**3**), (*E*)-β-farnesene (**4**), β-santalene (**5**), exo-β- bergamotene (**6**).

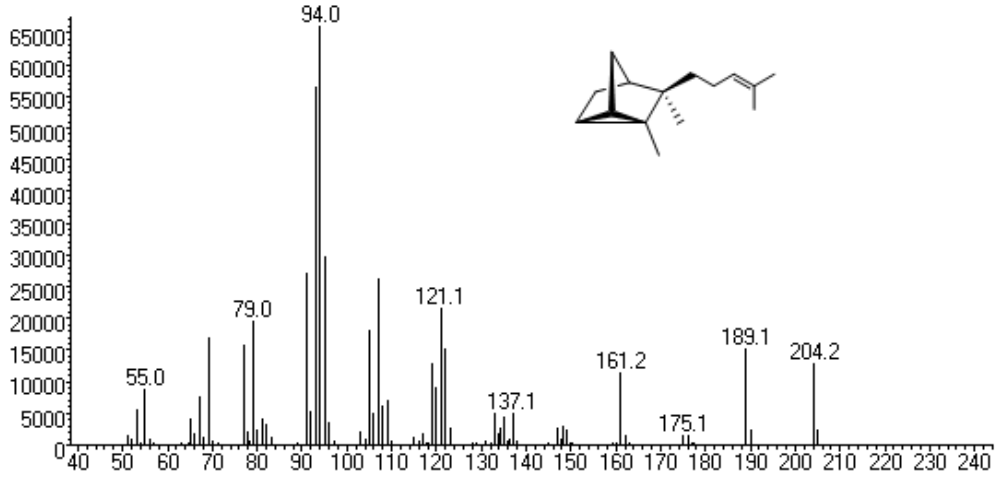


Figure 3.3.2.5: EI Mass spectrum of α-Santalene (**1**)

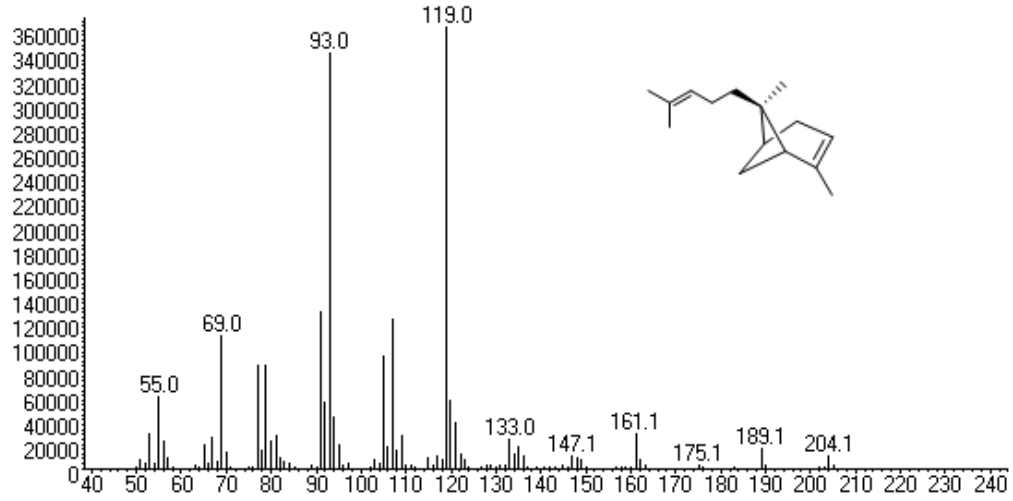


Figure 3.3.2.6: EI Mass spectrum of *exo*-α-Bergamotene (**2**)

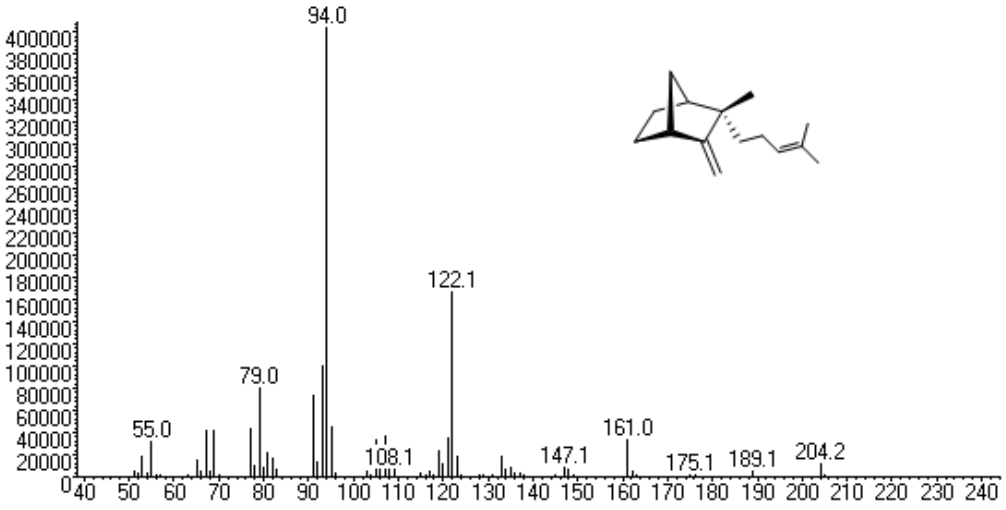


Figure 3.3.2.7: EI Mass spectrum of *epi*-β-Santalene (**3**)

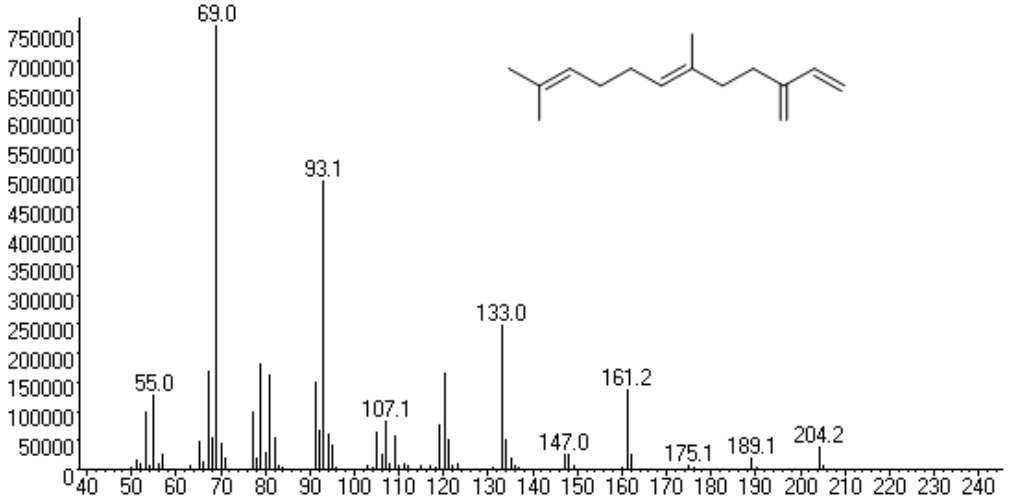


Figure 3.3.2.8: EI Mass spectrum of (*E*)-β-Farnesene (**4**)

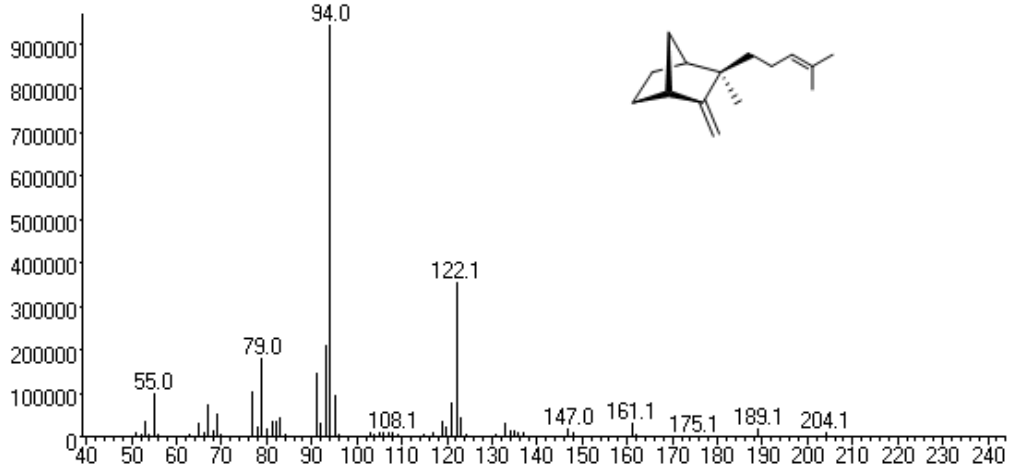


Figure 3.3.2.9: EI Mass spectrum of β-Santalene (**5**)

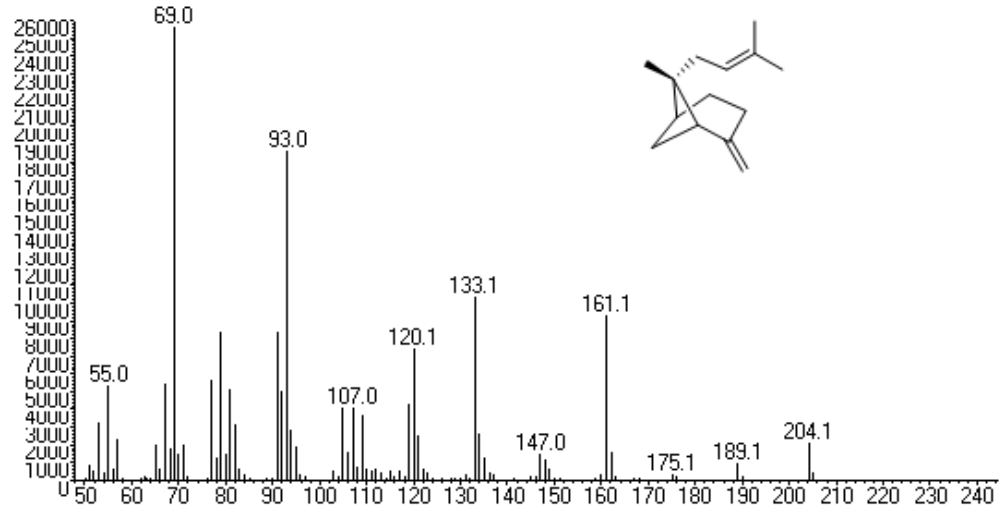
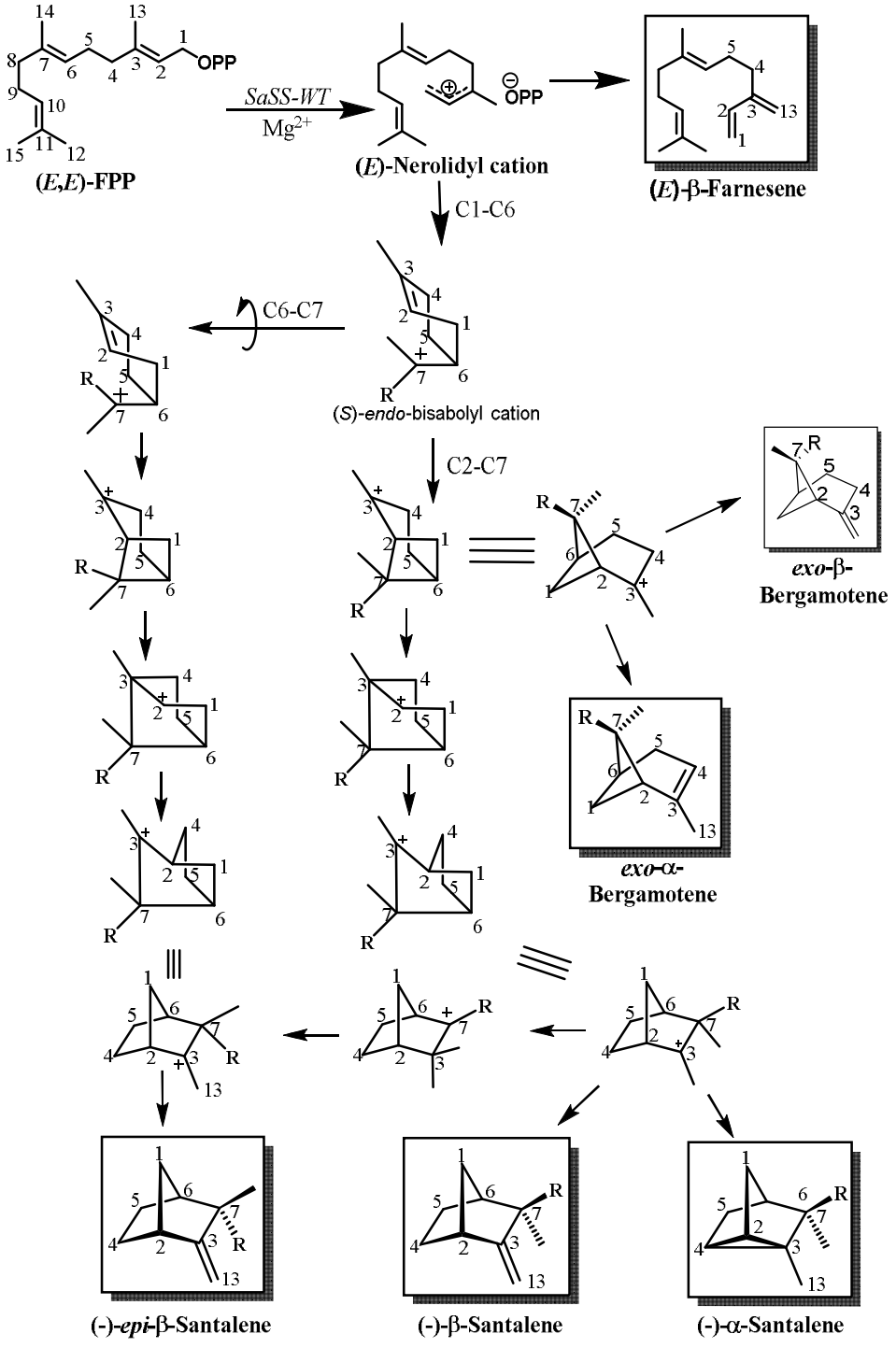


Figure 3.3.2.10: EI Mass spectrum of *exo*-β-Bergamotene (**6**)



Scheme 3.3.2.1: Schematic representation of proposed mechanism for the biosynthesis of santalene and bergamotene derivatives from (*E,E*)-FPP catalyzed *SaSS* in presence of Mg^{2+}.

3.3.2.4 Removal of thioredoxin tag

Thioredoxin tag (18 kDa) was removed by treating the purified and desalted protein of *SaSS* with thrombin. When the treated sample was analyzed by 10 % SDS-PAGE, a single sharp band of 66 kDa was obtained (Fig 3.3.2.11). Incubation of thrombin treated protein (without TRX tag) with (*E,E*)-FPP in presence of $MgCl_2$ resulted in a product profile similar to that obtained with *SaSS* with tag (3.3.2.12).

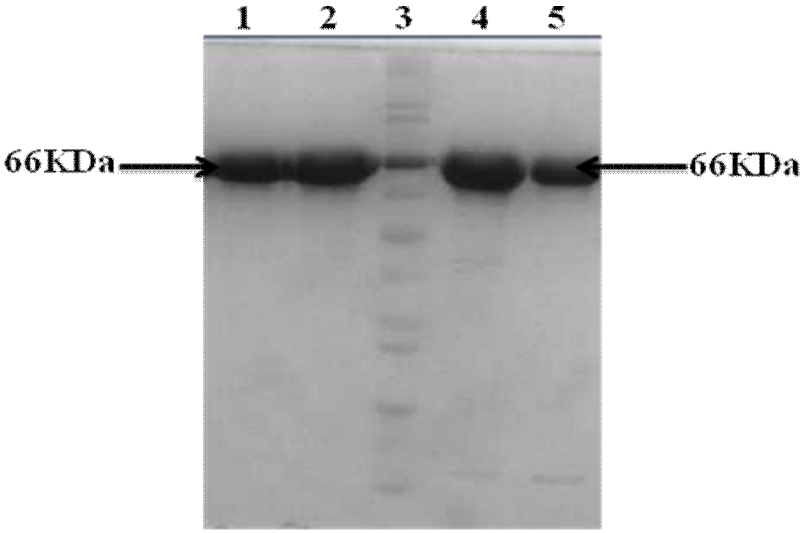


Figure 3.3.2.11: SDS-PAGE gel image of *SaSS* purified and thrombin treated sample, **Lanes 1, 2, 4 & 5:** Thrombin treated protein samples, **Lane 3:** Protein marker

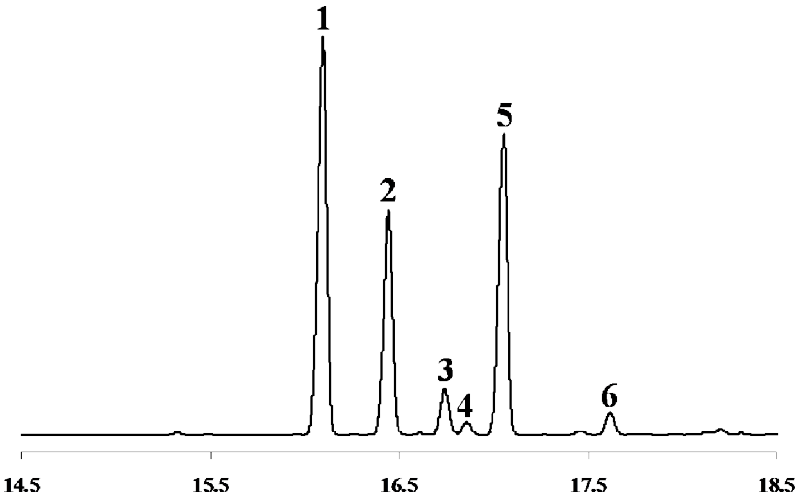


Figure 3.3.2.12: GC profile of *SaSS* enzyme assay using thrombin treated protein (without TRX taq) with (*E,E*)-FPP, α-santalene (**1**), exo-α-bergamotene (**2**), *epi*-β-santalene (**3**), (*E*)-β-farnesene (**4**), β-santalene (**5**), exo-β- bergamotene (**6**)

These results clearly indicate that 18 kDa of thioredoxin tag expressed along with the protein of interest does not affect the secondary structure of *SaSS* (Figure 3.3.2.10).

3.3.2.5 Product ratio studies of combined enzyme assay of *SaSS* and *SaFDS* with IPP and DMAPP/GPP as substrate

SaSS and *SaFDS* were incubated along with allylic diphosphate substrate GPP/DMAPP and isopentenyl diphosphate (IPP) compare the product ratios with that of *SaSS* incubation with (*E,E*)-FPP as explained in section 3.3.2.3. Separate experiments were performed, wherein *SaFDS* was incubated with IPP + DMAPP and/ IPP + GPP for 1 hr at 30 °C, followed by addition of sesquiterpene synthase *SaSS* to both the mixtures further incubated for 1 hr at similar conditions. GC and GC-MS analyses of both the assay mixtures resulted in the formation of sesquiterpenes **1** to **6** in the same ratio as was obtained from incubation of *SaSS* with (*E,E*)-FPP (Figure 3.3.2.13).

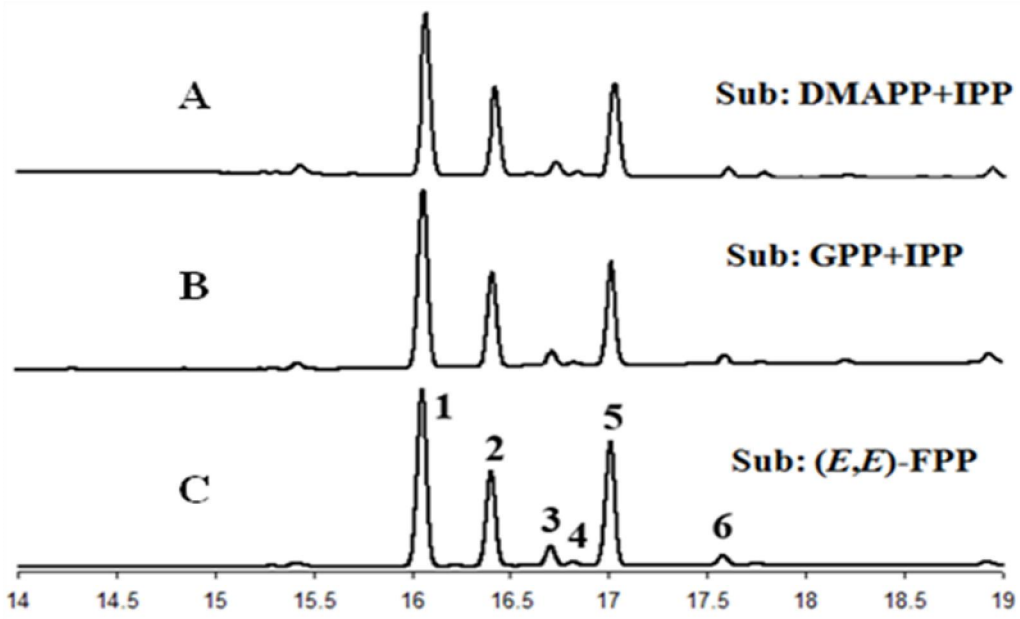
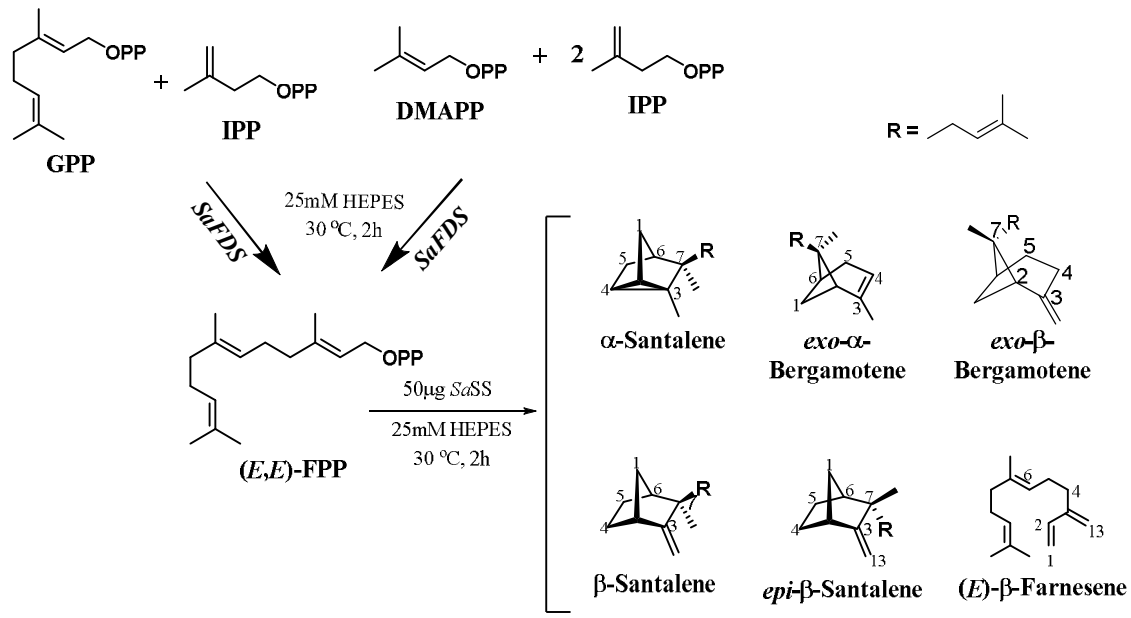


Figure 3.3.2.13: GC profile of combined enzyme assay *SaSS* and *SaFDS* **A)** Assay profile of *SaSS* when incubated with DMAPP + IPP along with *SaFDS*, **B)** Assay profile of *SaSS* when incubated with GPP + IPP along with *SaFDS*, **C)** Assay profile of *SaSS* when incubated with (*E,E*)-FPP

These observations clearly indicate that one can use this combined assay strategy utilizing sequential catalysis by *SaFDS* and *SaSS* along with appropriately deuterium labeled IPP and/or DMAPP/GPP to generate the desired deuterium labeled

(*E,E*)-FPP analogues which in turn are cyclized to produce deuterated sesquiterpene products. GC and GC-MS analyses of these deuterated sesquiterpene products by studying the kinetic isotope effects as well as the mass fragmentation pattern would establish the mechanism of sesquiterpene biosynthesis through carbocation cascade reactions. These results have been discussed in detail by Mr. Pankaj in his thesis chapter 3[25].



Scheme 3.3.2.2: Schematic representation of biosynthesis of santalene derivatives in combined enzyme assay of *SaSS* and *SaFDS* starting from GPP/DMAPP and IPP as substrate

3.3.2.6 Kinetic characterization of *SaSS*

Kinetic constants for *SaSS* were determined using GC for (*E,E*)-FPP with varying substrate concentrations from 0.5 µM to 20 µM. Amount of product formed was calculated using dodecane as an internal standard in the assay sample as discussed in materials and methods section (Section 3.2.11) and used to determine kinetic parameters using Graph pad Prism software. Kinetic parameters for SaSS were estimated to be K_m (substrate specificity) = 0.586 µM, V_{max} (rate of reaction) = 0.502 µM/min, K_{cat} (turn over number) = 0.100 and Kcat/Km (enzyme efficiency) = 0.170.

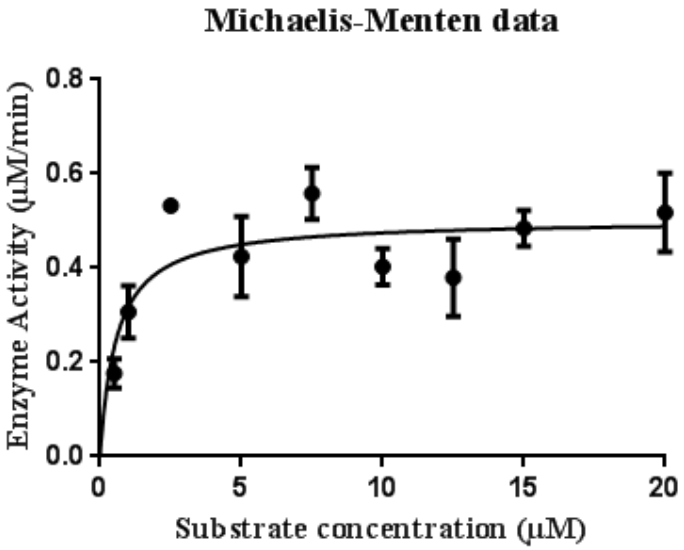
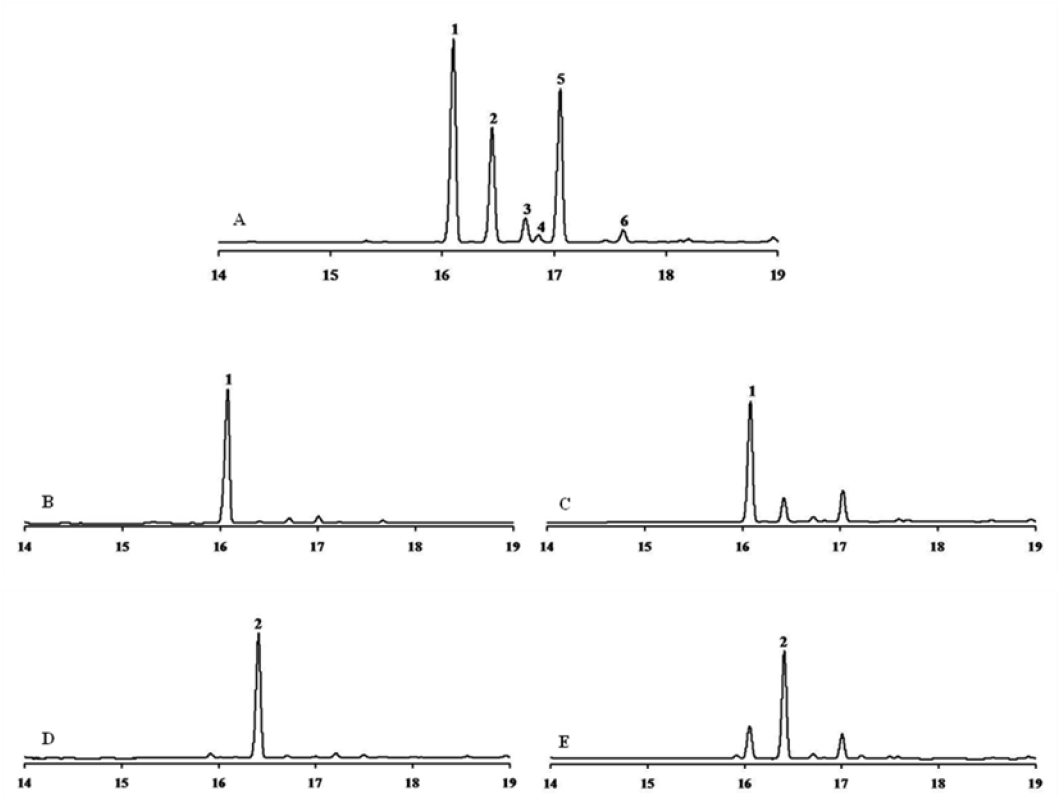


Figure 3.3.2.14: Michaelis Menten plot for *SaSS* kinetics with varying substrate concentration

3.3.2.7 Co-injection of standard terpenes with assay mixture



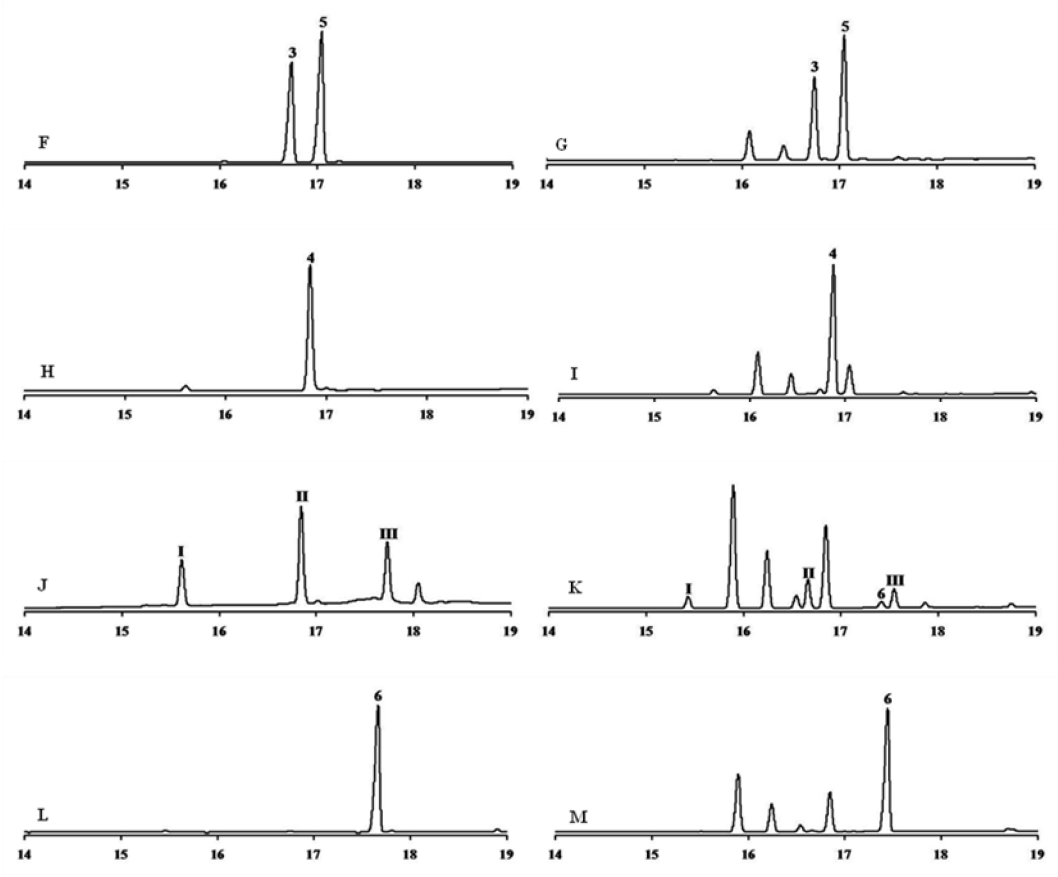


Figure 3.3.2.15: Co-injection of Standard terpene with *SaSS* assay mixture for the characterization of metabolites, **A)** Product profile of *SaSS* assay with (*E*,*E*)-FPP, **B)** α-Santalene (**1**) standard, **C)** Co-injection of *SaSS* assay mixture with α-Santalene, **D)** exo-α-Bergamotene (**2**) standard, **E)** Co-injection of *SaSS* assay mixture with *exo*-α-Bergamotene, **F)** *epi*-β-Santalene (**3**) and β-Santalene (**5**), **G)** Co-injection of *SaSS* assay mixture with *epi*-β-Santalene and β-Santalene, **H)** (*E*)-β-Farnesene (**4**), **I)** Co-injection of *SaSS* assay mixture with (*E*)-β-Farnesene, **J)** Synthesized mixture of (6*E*)-Farnesenes, **K)** Co-injection of SaSS assay mixture with mixture of (E)-farnesene for the characterization of product 6, **L)** exo-β-Bergamotene (**6**), **M)** Co-injection of *SaSS* assay mixture with exo-β-Bergamotene, α-santalene (**1**), exo-α-bergamotene (**2**), *epi*-β-santalene (**3**), (*E*)-β-farnesene (**4**), β-santalene (**5**), exo-β-bergamotene (**6**), mixture of (6*E*)-farnesene (**I**, **II**, **III**)

3.3.3 Isolation, cloning, expression and characterization of β-bisabolene synthase (*SaβBS*)

3.3.3.1 Screening, isolation and cloning of β-bisabolene synthase (*SaβBS*)

From transcriptome screening, one of the transcripts when analyzed using BLAST against NCBI database, was identified as (Locus_5558_Transcript_1/1_Confidence_1.000) with 832 bp length, and matched with β-bisabolene synthase reported from *S. album* with 99 % identity at amino acid level, and lacked its 3' sequences. To obtain the full length cDNA sequence of *SaβBS*, 3' RACE was performed as discussed earlier in methodology section. 3' RACE fragments with expected size of 1.25 kb was obtained and cloned in TA vector. Sequencing of 3' RACE fragment with M13 forward and reverse primer revealed the homology to reported β-bisabolene synthase from *S. album* with 99 % identity at amino acid level containing the 235 bp of 3' UTR sequence.

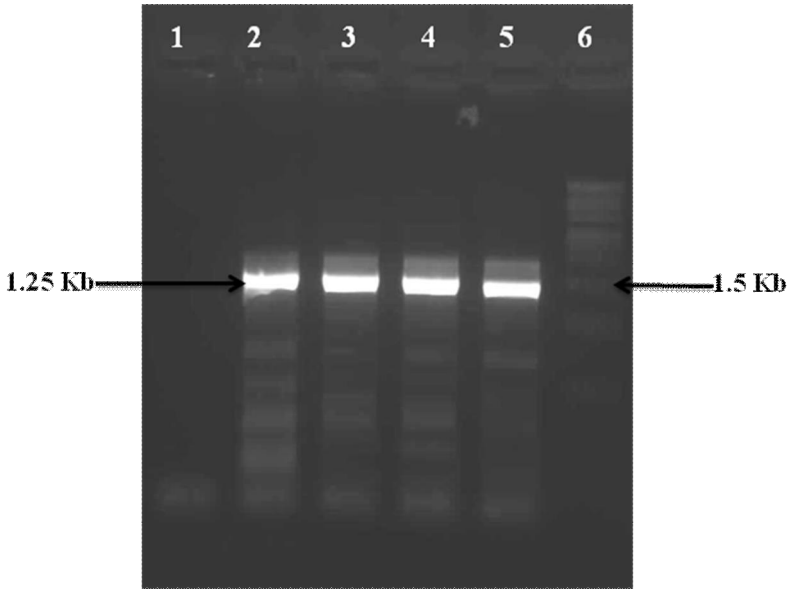


Figure 3.3.3.1: *Saβ-BS* 3′ RACE amplification, **Lane 1:** no insert control, **Lanes 2-5:** 3'RACE amplicon of β-bisabolene synthase, **Lane 6:** 1 Kb DNA ladder

The 3' RACE sequence of terpene synthases (*SaβBS*) was overlapped with EST fragment of Locus_5558_Transcript_1/1_Confidence_1.000 and the full length sequence was generated. The full length ORF of *SaβBS* was found to be composed of 1731 bp encoding a polypeptide of 577 amino acids with a calculated molecular weight of 65.90 kDa and pI of 5.48. Amino acid sequence analysis of *SaβBS* with

terpene synthases reported from another plant showed resemblance to *S. album* β-bisabolene synthase (99 % identity).

ORF of *SaβBS* was amplified from cDNA (Figure 3.3.3.2) as discussed in methods section and cloned in pET28a vector frame between BamHI and NotI restriction site with N terminal His_6 tag and S-tag under the control of T7 RNA polymerase.

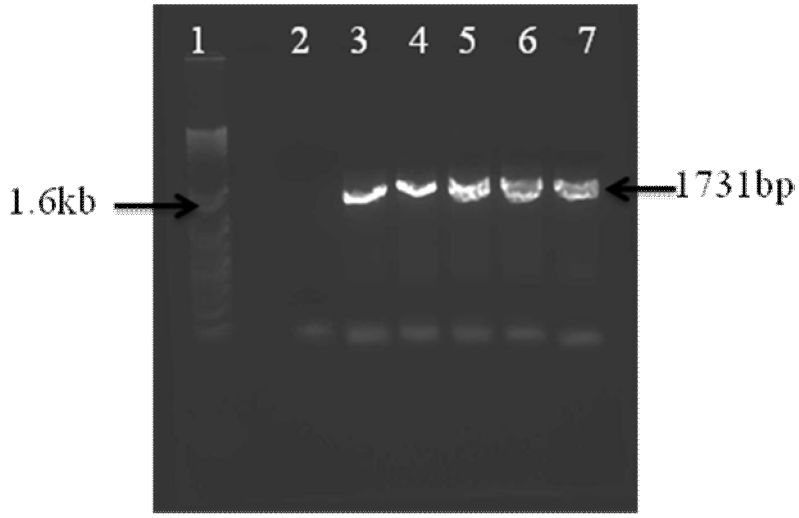


Figure 3.3.3.2: *Saβ-BS* full-length ORF amplification, **Lane 1:** DNA ladder, **Lane 2:** no insert control, **Lanes 3-6:** ORF of β-bisabolene synthase

3.3.3.2 Bacterial expression and protein purification

pET28a containing ORF of *SaβBS* was introduced into Rosetta 2 DE3 chemically competent cells for the expression of active protein. Active protein was over-expressed, as discussed in methodology section, and purified to the homogeneity using Ni-NTA resin. The recombinant protein was eluted with 250 mM imidazole, with the yield of 25-30 mg/L of bacterial culture. Purity and size of the expressed proteins analyzed by SDS-PAGE revealed ≥95 % purity of 66 kDa sharp band (Figure 3.3.3.2).

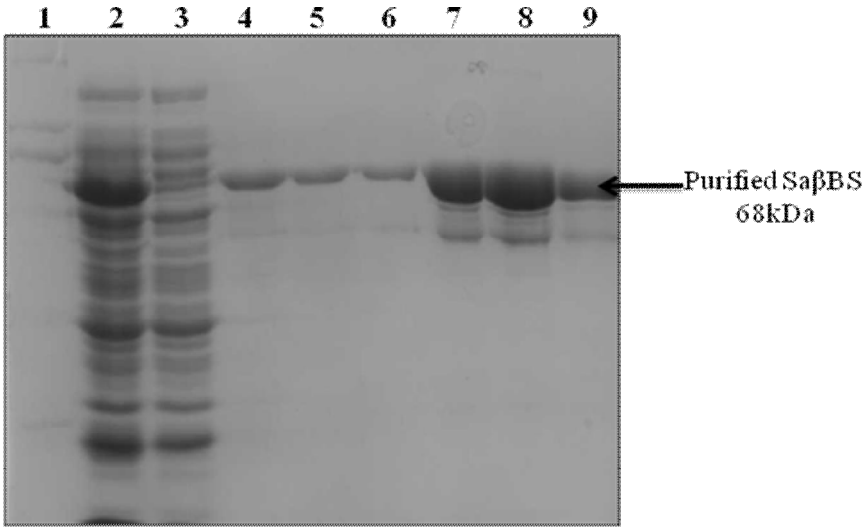


Figure 3.3.3.3: SDS-PAGE of *SaβBS* protein purification, **Lane 1:** Protein Marker, **Lane 2:** Crude Lysate, **Lane 3:** Unbound fraction, **Lane 4:** Wash 1, **Lane 5:** Wash 2, **Lane 6:** Wash 3, **Lane 7-9:** Elution fractions 1-3

3.3.3.3 Enzymatic characterization of *SaβBS*

For functional characterization, incubation of *SaβBS* with *(E,E)*-FPP in the presence of Mg^{2+} resulted in the exclusive formation of β-bisabolene and small proportion of bisabolol (Figure 3.3.3.4.C). Products were confirmed by GC-MS fragmentation analysis and co-injecting with purified standard (Figure 3.3.3.4.D-E). Analysis of assay sample on Astec CHIRAL DEX™ B-DA Capillary Column (Supelco™) showed the presence of a single isomer (Figure 3.3.3.5). A reaction mechanism for the formation of β-bisabolene is proposed in scheme 3.3.3. In recent years, sesquiterpenes of farnesene and bisabolene skeletons have been recognized as replacements for petroleum-derived jet-engine fuel[13,14].

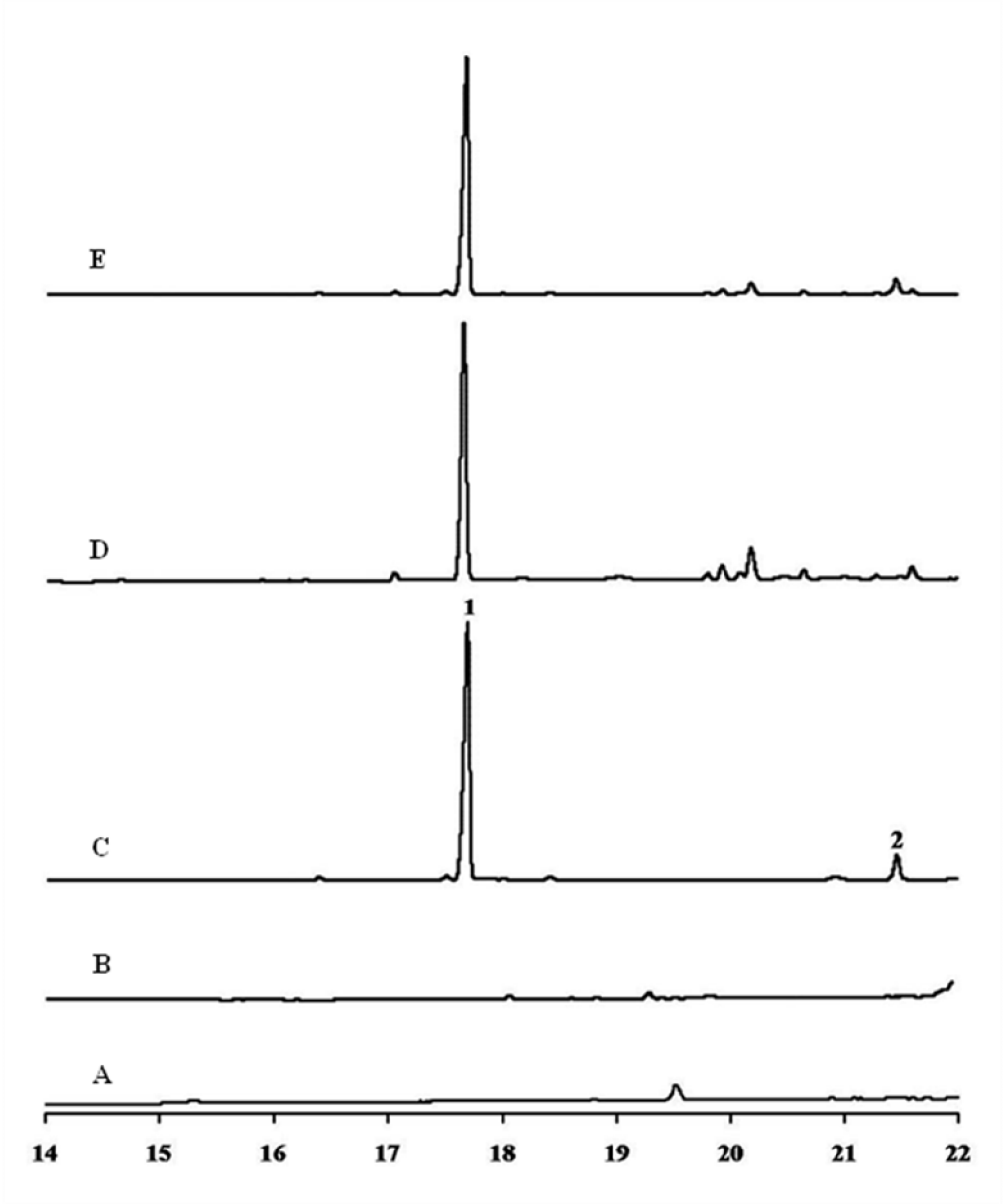


Figure 3.3.3.4: GC chromatogram of *Saβ-BS* assay profile with (*E,E*)-FPP, **A)** Empty vector control, **B)** Substrate control without enzyme, **C)** *SaβBS* assay profile with (*E,E*)- FPP, β-bisabolene (**1**), bisabolol (**2**), **D)** β-bisabolene standard, **E)** Co-injection of *SaβBS* assay sample with β-bisabolene standard

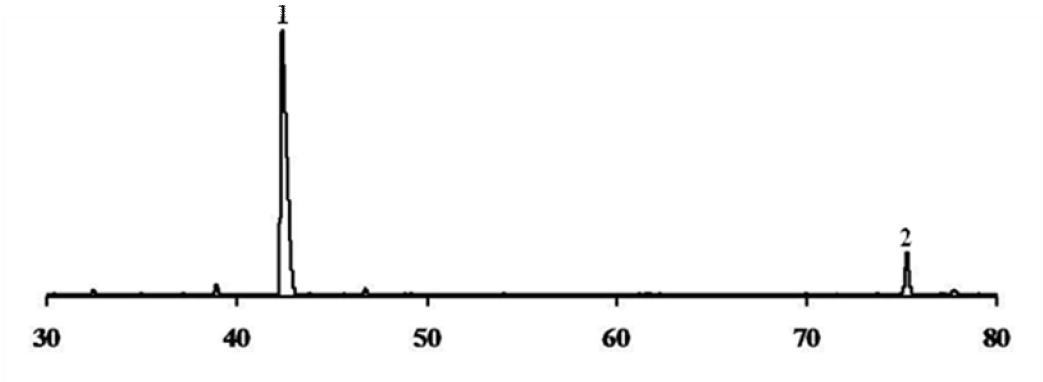


Figure 3.3.3.5: Chiral resolution of *Saβ-BS* assay sample on Astec CHIRAL DEX™ B-DA Capillary Column, β-bisabolene (**1**), α-bisabolol (**2**)

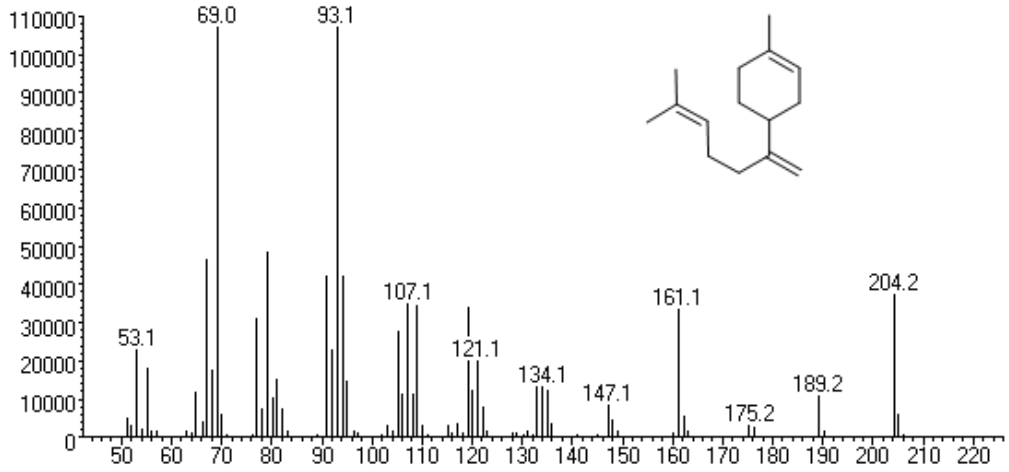


Figure 3.3.3.6: EI Mass spectrum of β-Bisabolene (**1**) produced in assay mixture of *SaβBS* with (*E,E*)-FPP

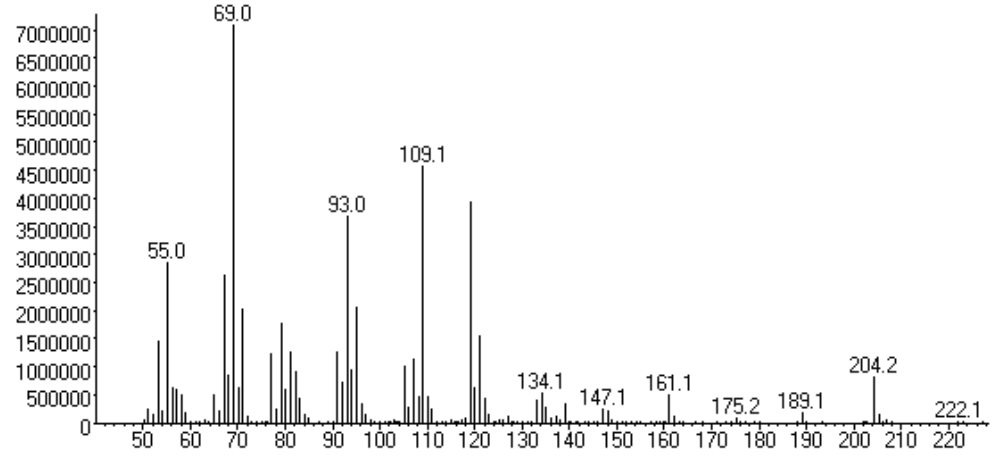
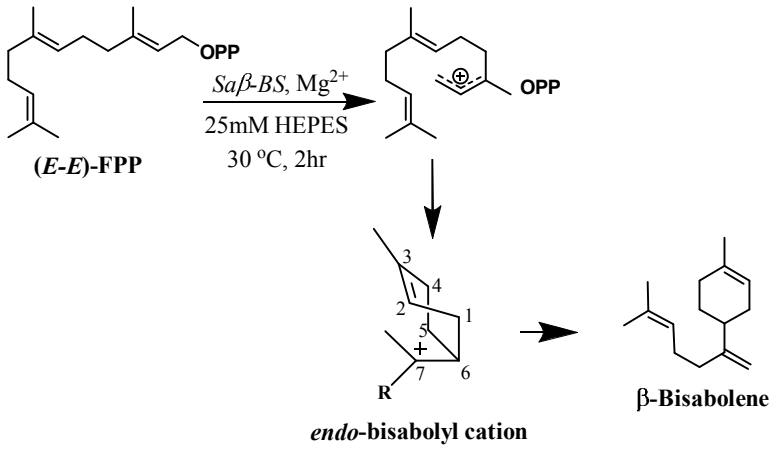


Figure 3.3.3.7: EI Mass spectrum of α-Bisabolol (**2**) produced in assay mixture of *SaβBS* with (*E,E*)-FPP



Scheme 3.3.3: Schematic representation of proposed mechanism for biosynthesis of β-bisabolene in assay mixture of *SaβBS* with (*E,E*)-FPP

3.3.3.4 Kinetic characterization of *SaβBS*

Kinetic constants for *SaβBS* were determined using GC for (*E,E*)-FPP with varying substrate concentration from 0.5 µM to 40 µM. The calculated amounts of product formed were used to determine the kinetic parameters using Graph pad Prism software. Kinetic parameters for SaβBS were estimated to be K_m (substrate specificity) = 12.59 µM, V_{max} (rate of reaction) = 0.2076 µM/min, K_{cat} (turn over number) = 0.0415 and K_{cat}/K_m (enzyme efficiency) =0.003.

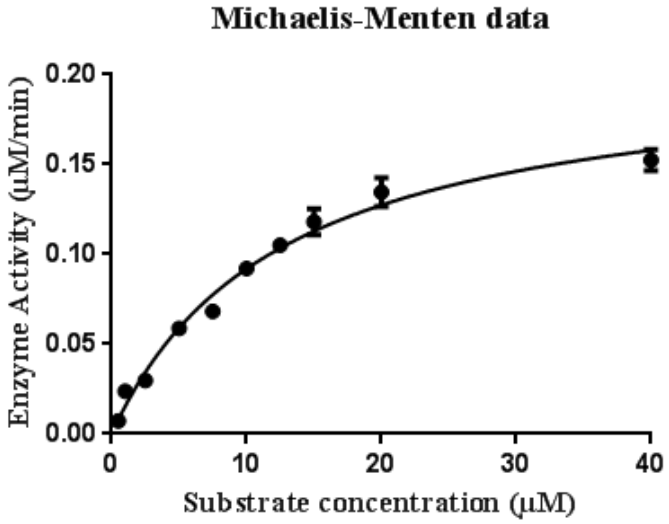


Figure 3.3.3.8: Michaelis Menten plot for *SaβBS* kinetics with varying substrate concentration

3.3.4 Isolation, cloning, expression and characterization of *SaTPS1* from Indian Sandalwood, *S. album*

3.3.4.1 Screening, isolation and cloning of *SaTPS1*

From transcriptome screening as discussed in previous chapter (Chapter 2), one of the transcripts of length 761 bp (Locus_33105_Transcript_1/1_Confidence_1.000), when analysed usng BLAST against NCBI database, was selected based on its homology with other reported terpene synthases in the database. EST fragment was found to match with sesquisabinene B synthase from *S. album* with 76 % identity at amino acid level, lacking its 3' sequence. To obtain the full length cDNA sequence of this EST fragment (Locus_33105), 3' RACE was performed as discussed earlier. 3' RACE fragments of expected size of 1.4 kb was obtained and cloned in TA cloning vector to generate 3' sequence. Sequencing of 3' RACE fragment revealed the homology with sesquisabinene B synthase reported from *S. album* with 92 % identity at amino acid level having the 3' UTR sequence of 260 bp.

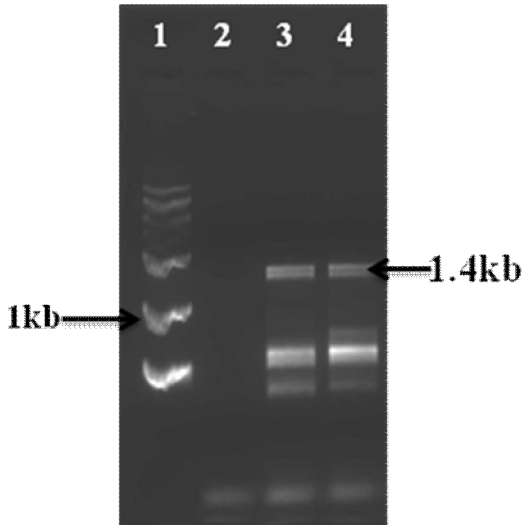


Figure 3.3.4.1: 3' RACE amplification of *SaTPS1*, **Lane 1:** 1 Kb DNA Ladder, **Lane 2:** no insert control, **Lane 3&4:** 3′ RACE amplicon of *SaTPS1*

The 3' RACE sequence of sesquiterpene synthase (*SaTPS1*) was overlapped with EST fragment of Locus_33105_Transcript_1/1_Confidence_1.000 and the full length sequence was generated. The full length ORF of SaTPS1 was found to be composed of 1701 bp encoding a polypeptide of 566 amino acids with a calculated molecular weight of 65.22 kDa and pI of 5.01. Predicted polypeptide sequence of *SaTPS1* lacks N-terminal organelle targeting sequence, suggesting that this enzyme is

directed to the cytoplasm. The analysis of deduced amino acid sequence of *SaTPS1* with other terpene synthases reported in NCBI database, showed resemblance to *S. album* sesquisabinene B synthase (88 % identity), and β-bisabolene synthase (63 % identity). The proteins sequence of *SaTPS1* shares several highly conserved residues with known sesquiterpene synthase from other plant species including DDxxD motif (Asp^{316}, Asp^{317}, and Asp^{320}) involved in substrate binding, (D/**N**)Dxx(S/**T**)xxx**E** motif for metal binding, RRx_8W motif (Arg^{27}, Arg^{28}, and Trp^{37}) conserved at N-terminal. It lacks the RxR which is conserved in terpene synthases and located 34 amino acids upstream of DDxxD motif, but a similar RNG sequence is present.

ORF of *SaTPS1* was amplified from cDNA and cloned in pET28a vector frame between NdeI and BamHI restriction site with N terminal His_6 tag under the control of T7 RNA polymerase for the expression of recombinant active protein.

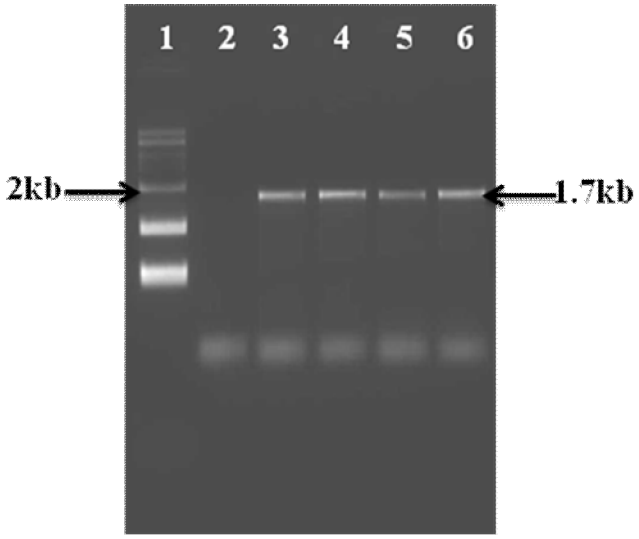


Figure 3.3.4.2: Amplification of ORF of *SaTPS1*, **Lane 1:** 1 Kb DNA ladder, **Lane 2:** no insert control, **Lanes 3-6:** ORF of *SaTPS1*

3.3.4.2 Bacterial expression and protein purification

pET28a containing ORF of *SaTPS1* was introduced into Rosetta 2 DE3 chemically competent cells for the expression of active protein. Active protein was over-expressed as discussed in methodology section and purified to the homogeneity using Ni-NTA, with the yield of 25-30 mg/L of bacterial culture. Purity and size of the expressed protein analyzed on a 10 % SDS-PAGE revealed a sharp band at 66 kDa and of >95 % purity (Fig 3.3.4.3). The purified protein was desalted using ACTA, flash frozen and stored at -80 ºC till further use.

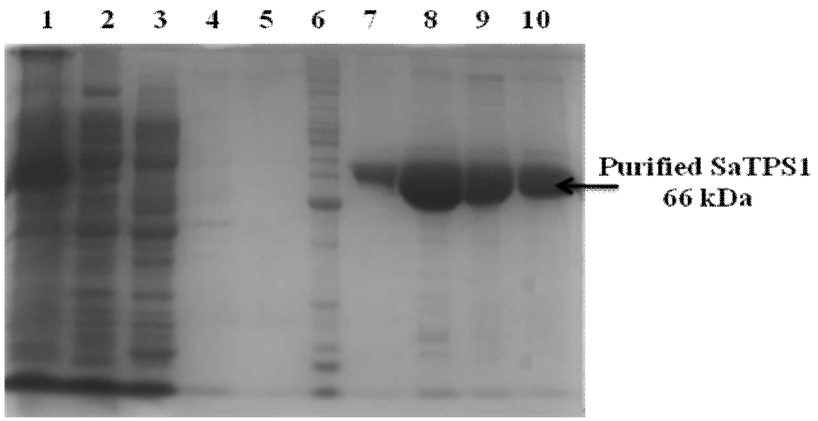


Figure 3.3.4.3: SDS-PAGE gel image of *SaTPS1* protein purification, **Lane 1:** Crude lysate, **Lane 2:** Unbound fraction, **Lane 3:** Wash 1, **Lane 4:** Wash 2 **Lane 5:** Wash 3, **Lane 6:** Protein marker, **Lanes 7-9:** Elution fraction 1-4

3.3.4.3 Enzymatic characterization of *SaTPS1*

SaTPS1 was functionally characterized by incubation of 50 µg purified and desalted recombinant protein with 100 µM (*E,E*)-FPP in the presence of 10 mM Mg^{2+}. GC and GCMS analyses resulted in exclusive formation of Sesquisabinene B (**1**, 86.92 %), and minor metabolites: β-Sesquiphellandrene (**2**, 6.58 %), and α-Himachalene (**3**, 6.50 %) as shown in figure 3.3.4.4.

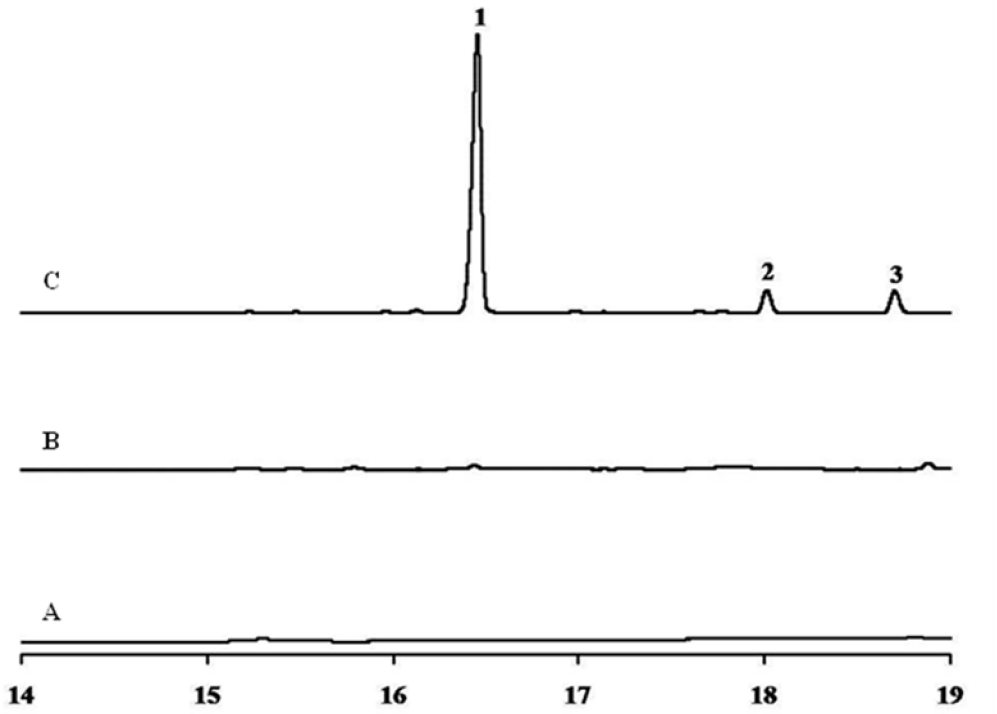


Figure 3.3.4.4: GC chromatogram of *SaTPS1* assay profile with (*E,E*)-FPP, **A)** Empty vector control, **B)** Substrate control without enzyme, **C)** Assay profile of *SaTPS1* with (*E,E*)-FPP, Sesquisabinene B (**1**), β-Sesquiphellandrene (**2**), α-Himachalene (**3**)

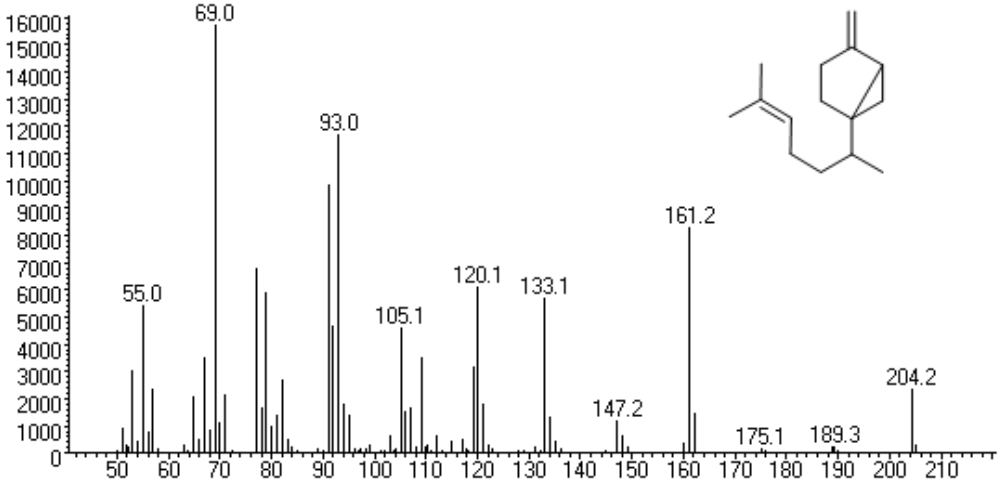


Figure 3.3.4.5: EI Mass Spectrum of Sesquisabinene B (**1**) produced in the assay mixture of *SaTPS1*

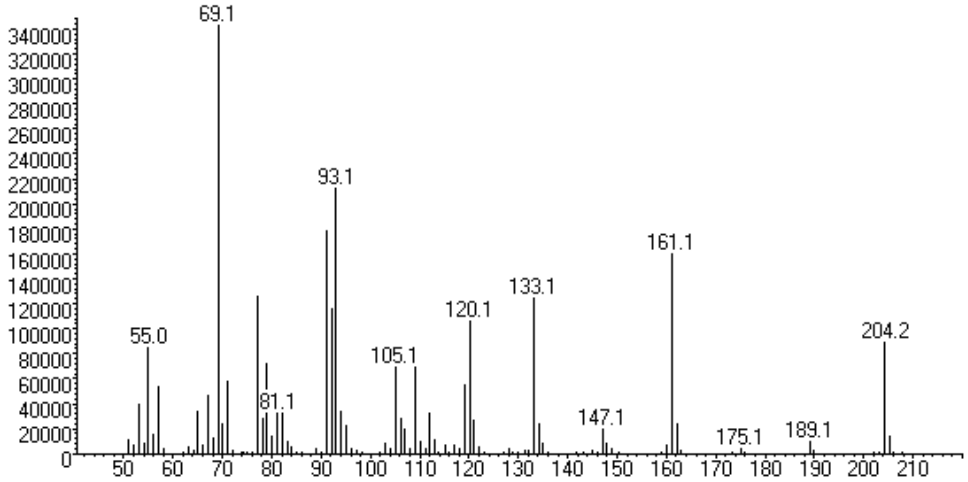


Figure 3.3.4.6: EI Mass spectrum of β-Sesquiphellandrene (**2**) produced in the assay mixture of *SaTPS1*

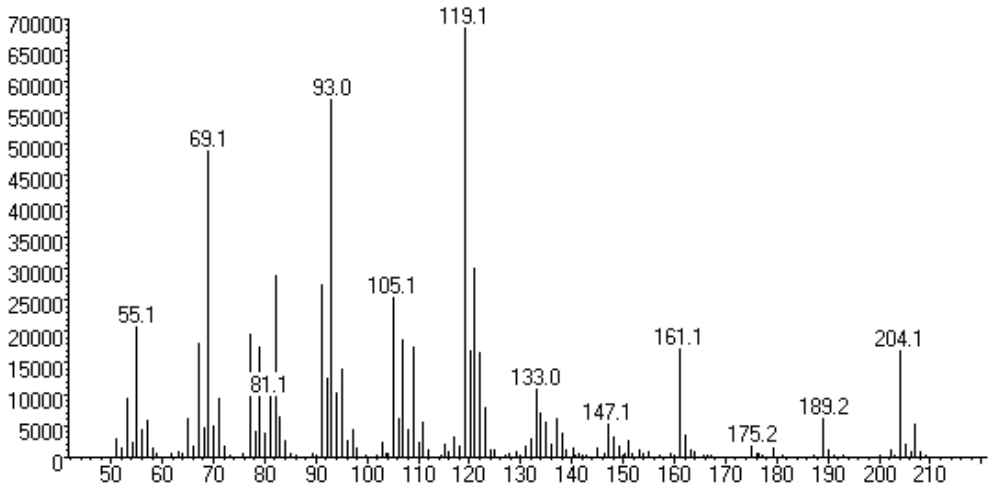


Figure 3.3.4.7: EI Mass spectrum of α-Himachalene (**3**) produced in the assay mixture of *SaTPS1*

3.3.4.4 Characterization of *SaTPS1* metabolite

For the characterization of the major metabolite of *SaTPS1*, large scale assay was set up, as discussed earlier, and extracted using n-hexane. 17.1 mg of crude extract was obtained which was further loaded on $AgNO_3$ coated silica column for purification of major metabolites. Metabolite 1 was purified to ≥95% purity (Figure 3.3.4.8) and analyzed by H^1, C^{13}, and DEPT NMR for structure elucidation. On analysis of the spectral data, the sesquiterpene was assigned as sesquisabinene B and the data matched with the literature reports [26]. Minor metabolites were characterized on the basis of GC-MS fragmentation analysis and NIST library match. The proposed reaction mechanism for the formation of all the metabolites from *SaTPS1* with (*E,E*)-FPP is represented in scheme 3.3.4.

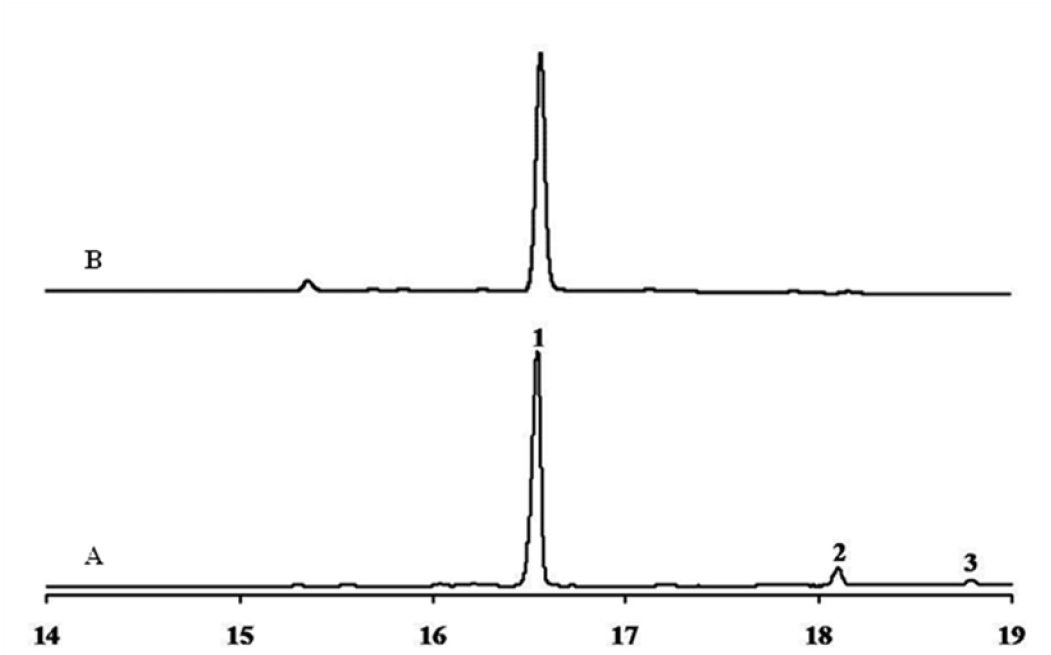
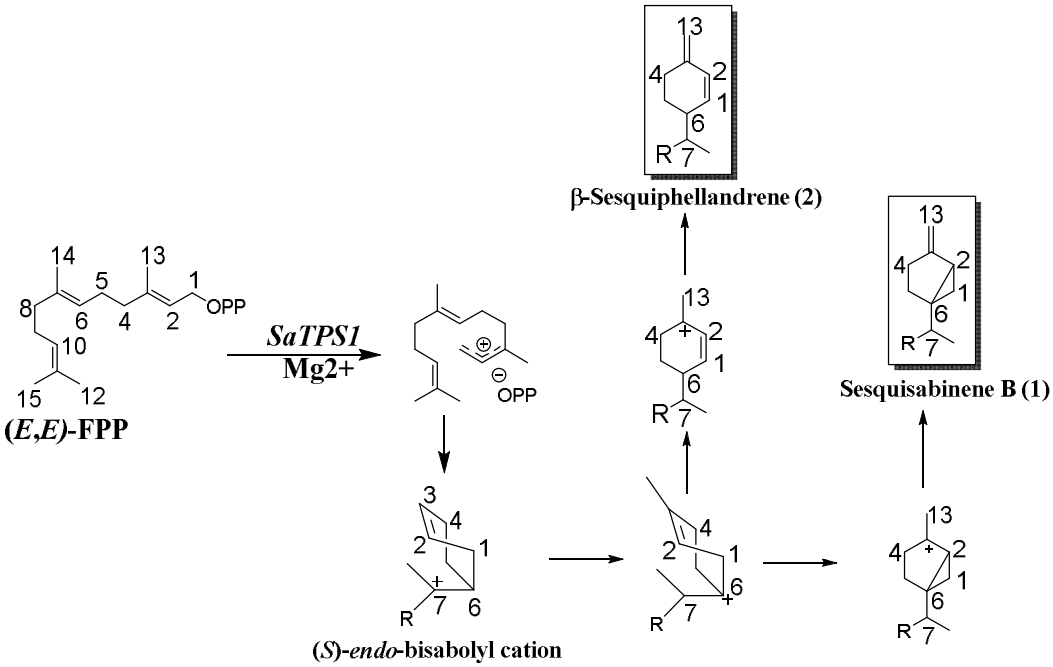


Figure 3.3.4.8: GC chromatogram of purified sesquisabinene from large scale enzyme assay product of *SaTPS1*, **A)** Crude extract of large scale assay mixture, **B)** Purified sesquisabinene B (**1**)



Scheme 3.3.4: Proposed mechanism for the formation of sesquiterpene mixture from *SaTPS1* upon incubation with (*E,E*)-FPP

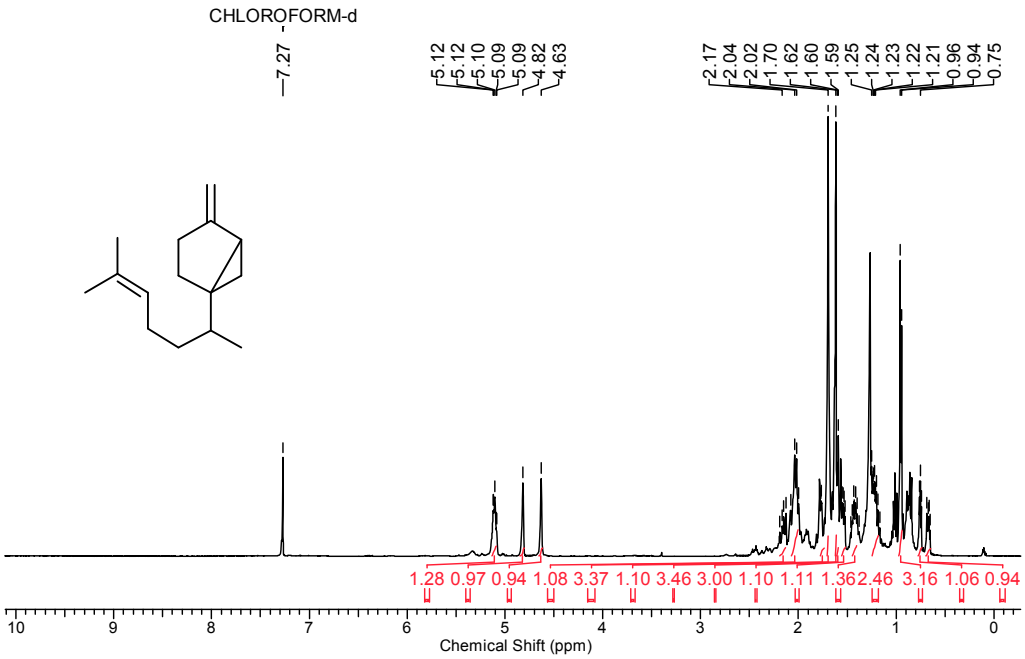


Figure 3.3.4.9: ^{1}H NMR of Sesquisabinene B in $CDCl_3$ at 400 MHz.

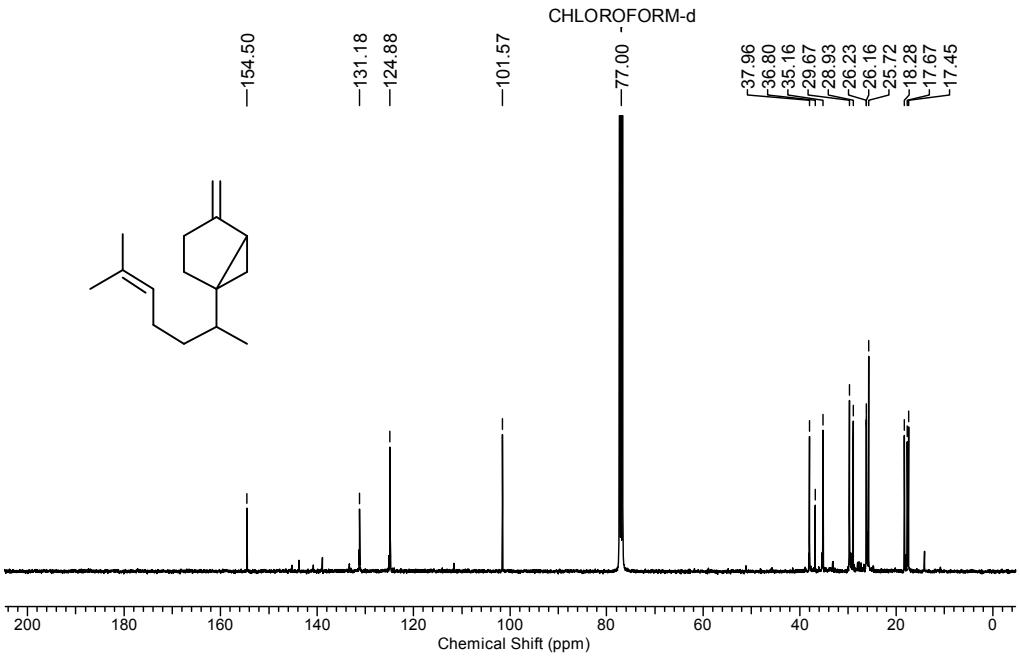


Figure 3.3.4.10: ^{13}C NMR of Sesquisabinene B in $CDCl_3$ at 400 MHz.

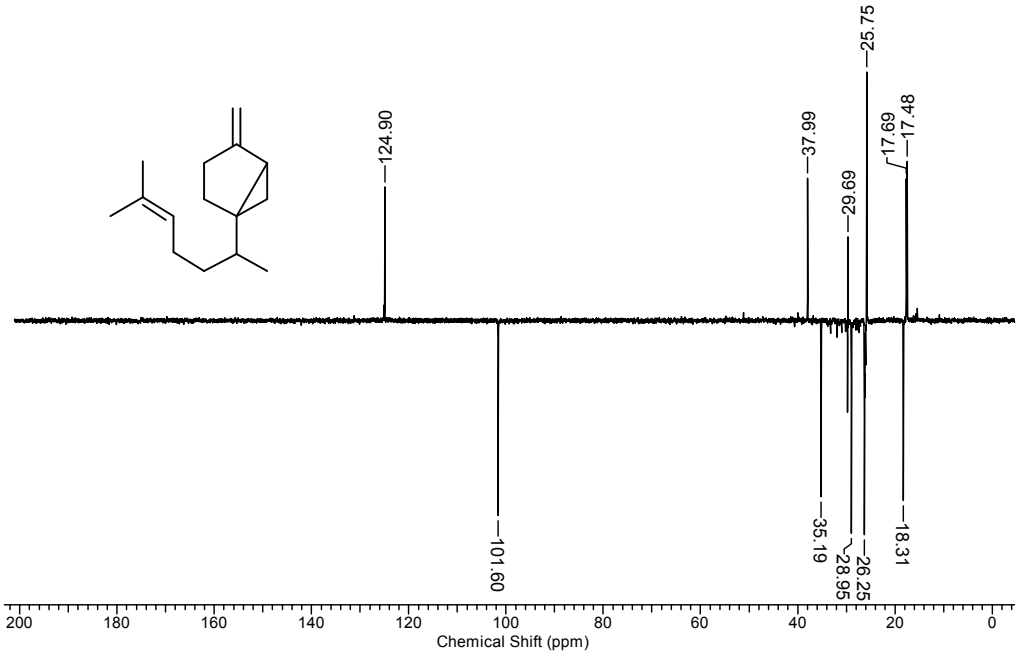


Figure 3.3.4.11: DEPT NMR of Sesquisabinene B in $CDCl_3$ at 400 MHz.

3.3.4.5 Kinetic characterization of *SaTPS1*

Kinetic constants for *SaTPS1* were determined using GC for (*E,E*)-FPP with varying substrate concentrations from 0.5 µM to 60 µM. Amount of product formed was calculated using dodecane as an internal standard in the assay sample. The calculated amount of products formed was used to determine kinetic parameters using Graph pad Prism software. The kinetic parameters were estimated to be K_m (substrate specificity) = 11.12µM, V_{max} (rate of reaction) = 1.76 µM/min, K_{cat} (turn over number) = 0.3524/min and K_{cat}/K_m (enzyme efficiency) = 0.031.

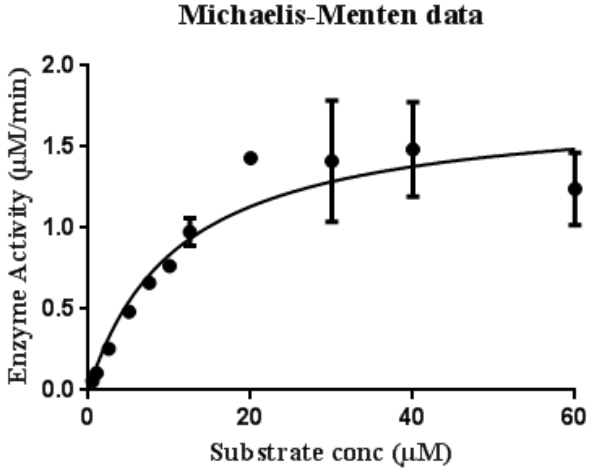


Figure 3.3.4.12: Michaelis Menten plot for *SaTPS1* kinetics with varying substrate concentration

Section 3.3.5 Isolation, cloning, expression and characterization of *SaTPS2* from Indian Sandalwood *S. album*

3.3.5.1 Screening, isolation and cloning of *SaTPS2*

From transcriptome screening as discussed in previous chapter (Chapter 2), one of the transcripts of length 1016 bp (Locus_8408_Transcript_1/1_Confidence_1.000), was identified using BLAST analysis based on its homology with other terpene synthases reported in NCBI database. EST fragment matched with sesquisabinene B synthase from *S. album* with 99 % identity at amino acid level lacking its 3′ sequence. To obtain the full length of *SaTPS2*, 3' RACE was performed as discussed earlier. 3' RACE fragment of expected size of 1.2 kb was obtained and cloned in TA cloning vector. Sequencing of 3' RACE fragment revealed the homology with sesquisabinene B synthase with 98 % identity at amino acid level having the 3' UTR sequence of 173 bp.

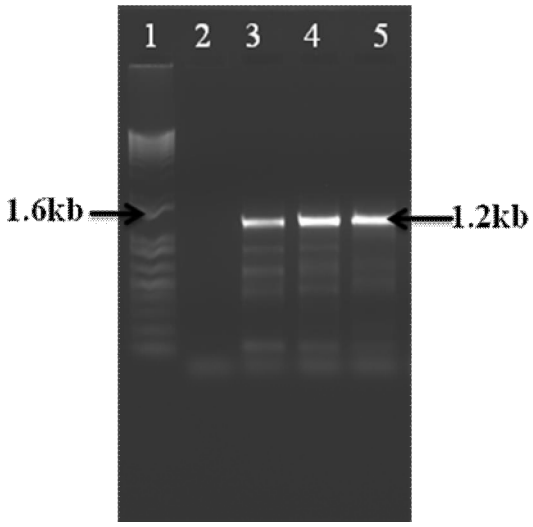


Figure 3.3.5.1: 3' RACE amplification of *SaTPS2*, **Lane 1:** 1 Kb DNA Ladder, **Lane 2:** no insert control, **Lane 3-5:** 3′ RACE amplicon of *SaTPS2*

The 3' RACE sequence of terpene synthases (*SaTPS2*) was overlapped with EST fragments and the full length sequence was generated. The full length ORF of *SaTPS2* is composed of 1701 bp encoding a polypeptide of 566 amino acids with a calculated molecular weight of 65.44 kDa and pI of 5.10. The analysis of amino acid sequence of *SaTPS2* with other terpene synthases showed resemblance to *S. album* sesquisabinene B synthase (98 % identity), and β-bisabolene synthase (63 % identity). *SaTPS1* and *SaTPS2* were found to be very similar to each other with 81.7 % identity

at DNA sequence level and 82.8 % identity and 89.6 % similarity at protein sequence level (Figure 3.3.5.2).

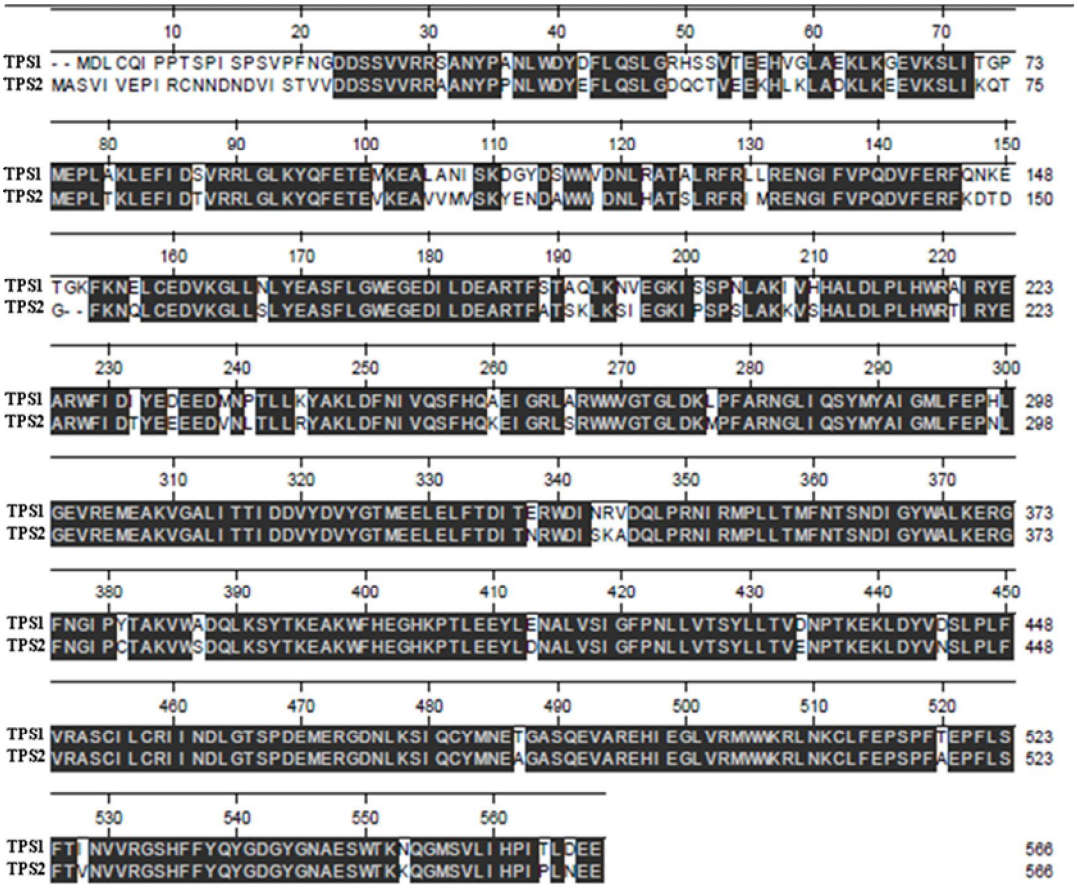


Figure 3.3.5.2 Protein sequence alignment of *SaTPS1* and *SaTPS2*

Predicted polypeptide sequence of *SaTPS2* lacked N terminal organelle targeting sequence, suggesting that this enzyme is directed to the cytoplasm. The protein sequence of *SaTPS2* shared several highly conserved residues with known sesquiterpene synthases from other plant species including DDxxD motif (Asp^{316}, Asp^{317}, and Asp^{320}) involved in substrate binding, (D/**N**)Dxx(S/**T**)xxx**E** motif for metal binding, RRx_8W motif (Arg^{29}, Arg^{30}, and Trp^{38}) conserved at N-terminal. It lacked RxR which is conserved in terpene synthases and is located 34 amino acids upstream of DDxxD motif, but a similar RNG sequence was present.

ORF of *SaTPS2* was amplified from cDNA and cloned in pET28a vector frame between BamHI and XhoI restriction site with N terminal His_6 for affinity purification, and S-tag for detection under the control of T7 RNA polymerase.

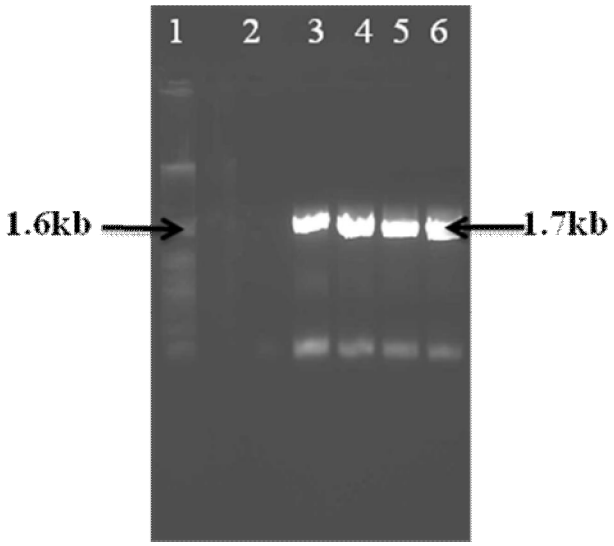


Figure 3.3.5.3: Amplification of ORF of *SaTPS2*, **Lane 1:** 1 Kb DNA ladder, **Lane 2:** no insert control, **Lanes 3-6:** Amplicon of *SaTPS2*

3.3.5.2 Bacterial expression and protein purification

pET28a containing ORF of *SaTPS2* was introduced into Rosetta 2 DE3 chemically competent cells for the expression of active protein. Active protein was over-expressed as discussed earlier in methodology section and purified to the homogeneity using Ni-NTA resin with the yield of 70-75 mg/L of bacterial culture. Purity and size of the expressed protein analyzed by SDS-PAGE revealed >95 % purity of a 66 kDa sharp band (Figure 3.3.5.4). The purified protein was desalted, flash frozen and stored at -80 °C till further use.

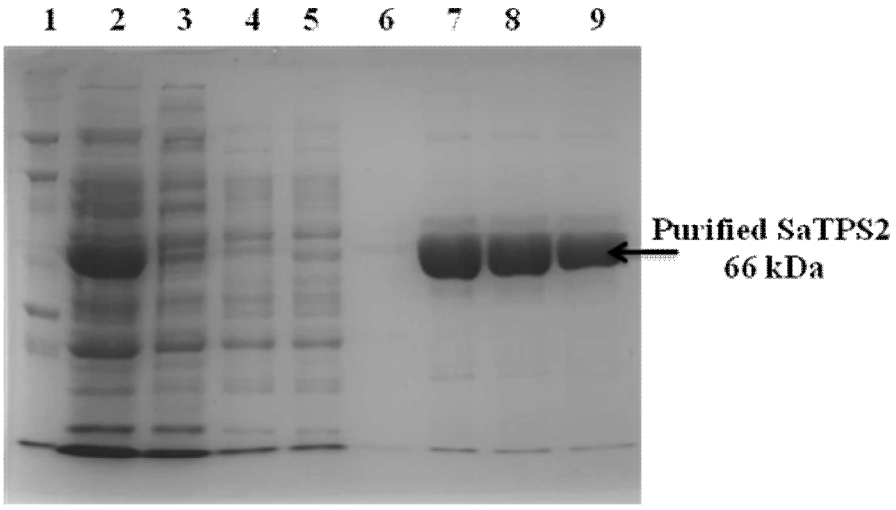


Figure 3.3.5.4: SDS-PAGE gel image of *SaTPS2* protein purification, **Lane 1:** Protein marker, **Lane 2:** Crude lysate, **Lane 3:** Unbound fraction, **Lane 4:** Wash 1, **Lane 5:** Wash 2, **Lane 6:** Wash 3, **Lanes 7-9:** Elution fraction 1-3

3.3.5.3 Enzymatic characterization of *SaTPS2*

Incubation of *SaTPS2* with (*E,E*)-FPP in presence of Mg^{+2} for functional characterization, resulted in exclusive formation of Sesquisabinene B (**1**, 80 %) and smaller amount of β-Sesquiphellandrene (**2**, 7.1 %), α-Himachalene (**3**, 4.9 %), exo-α-Bergamotene (**4**, 3.4 %) and (*E*)-α-Farnesene (**5**, 3.4 %) as shown in figure 3.3.5.5. Major metabolite (**1**) was confirmed by co-injection of *SaTPS2* assay sample with purified sesquisabinene B from *SaTPS1* assay mixture.

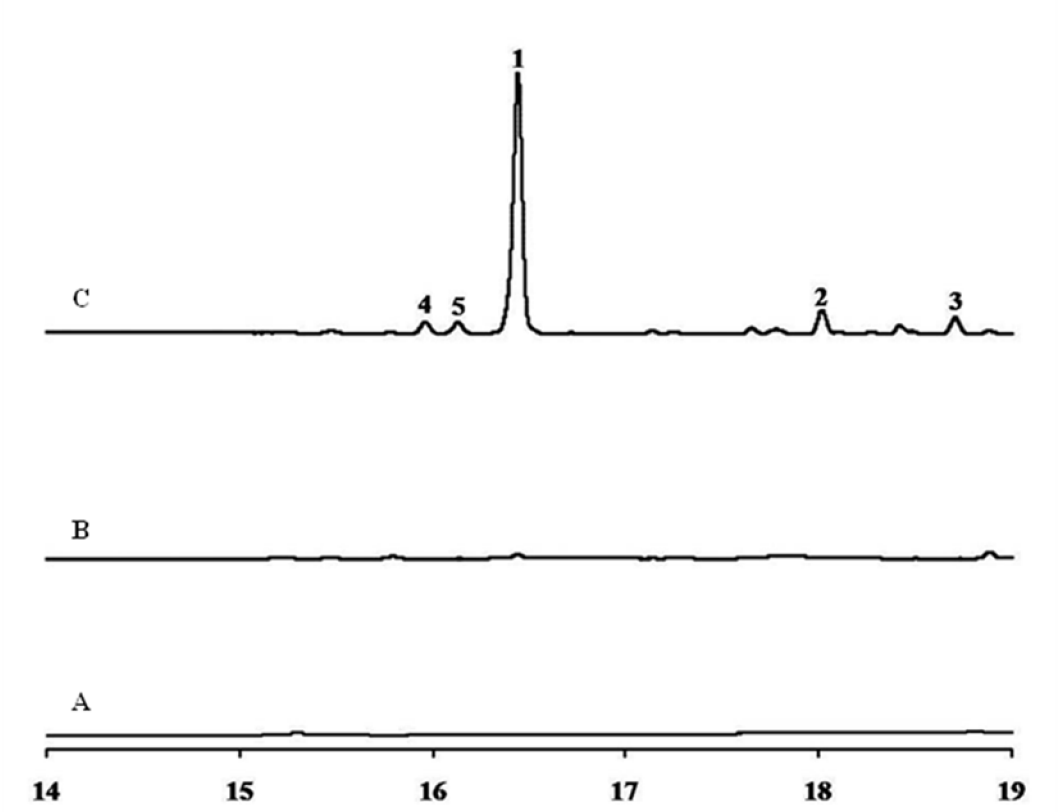


Figure 3.3.5.5: GC chromatogram of *SaTPS2* assay profile with (*E,E*)-FPP, **A)** Empty vector control, **B)** Substrate control without enzyme, **C)** Assay profile of *SaTPS2* with (*E,E*)-FPP, Sesquisabinene B (**1**), β-Sesquiphellandrene (**2**), α-Himachalene (**3**), exo-α-Bergamotene (**4**), (*E*)-α-Farnesene (**5**)

Analysis of assay sample of *SaTPS1*, *SaTPS2* and their co-injection on Astec CHIRAL DEX™ B-DA Capillary Column (Supelco™) showed the presence of a single isomer of major metabolite 1 (Figure 3.3.5.6) and confirmed that both the sesquierpene synthases are isoform of sesquisabinene synthase. Minor metabolites are characterized on the basis of GC-MS fragmentation analysis and NIST library match.

The proposed reaction mechanism for the formation of all the metabolites from *SaTPS2* with (*E,E*)-FPP is represented in scheme 3.3.5.

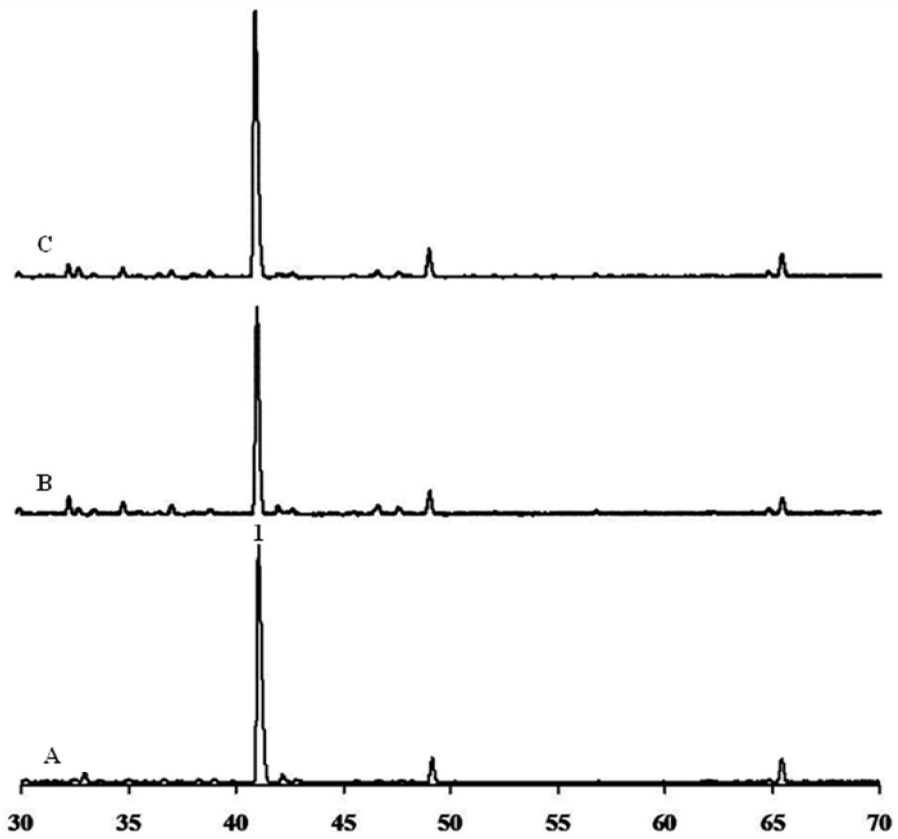


Figure 3.3.5.6: Chiral GCMS chromatogram of *SaTPS1* and *SaTPS2* assay sample on Astec CHIRAL DEX™ B-DA Capillary Column, **A)** *SaTPS1* assay sample, **B)** *SaTPS2* assay sample, **C)** Co-injection of *SaTPS1* and *SaTPS2* assay sample

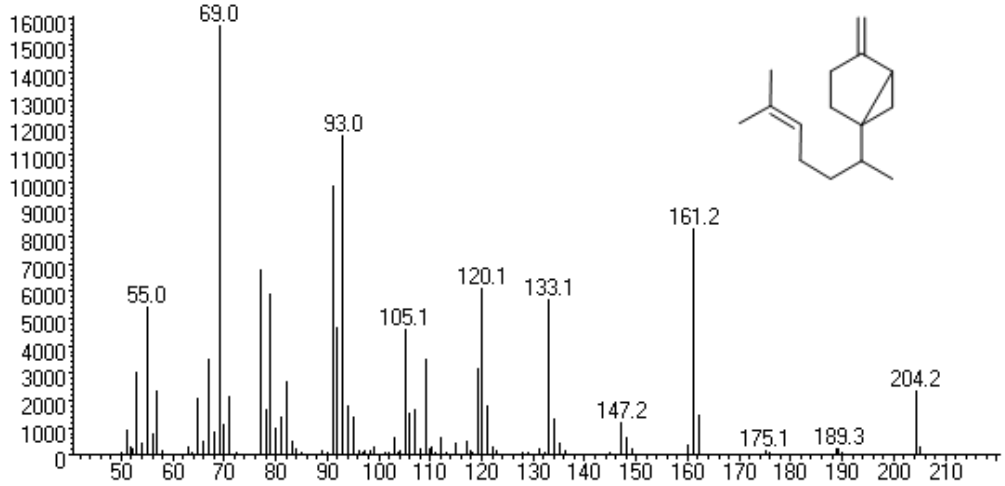


Figure 3.3.5.7: EI Mass spectrum of Sesquisabinene B (**1**) produced in assay mixture of *SaTPS2*

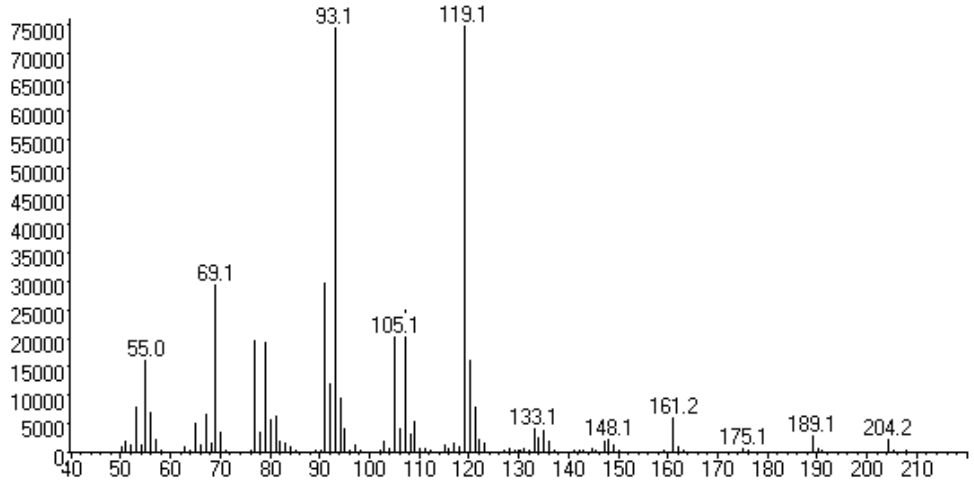


Figure 3.3.5.8: EI Mass spectrum of exo-α-Bergamotene (**4**) produced in assay mixture of *SaTPS2*

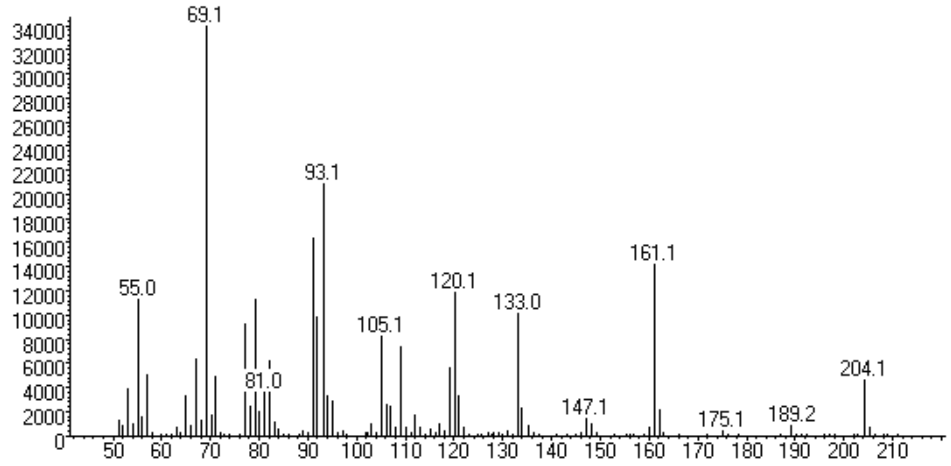
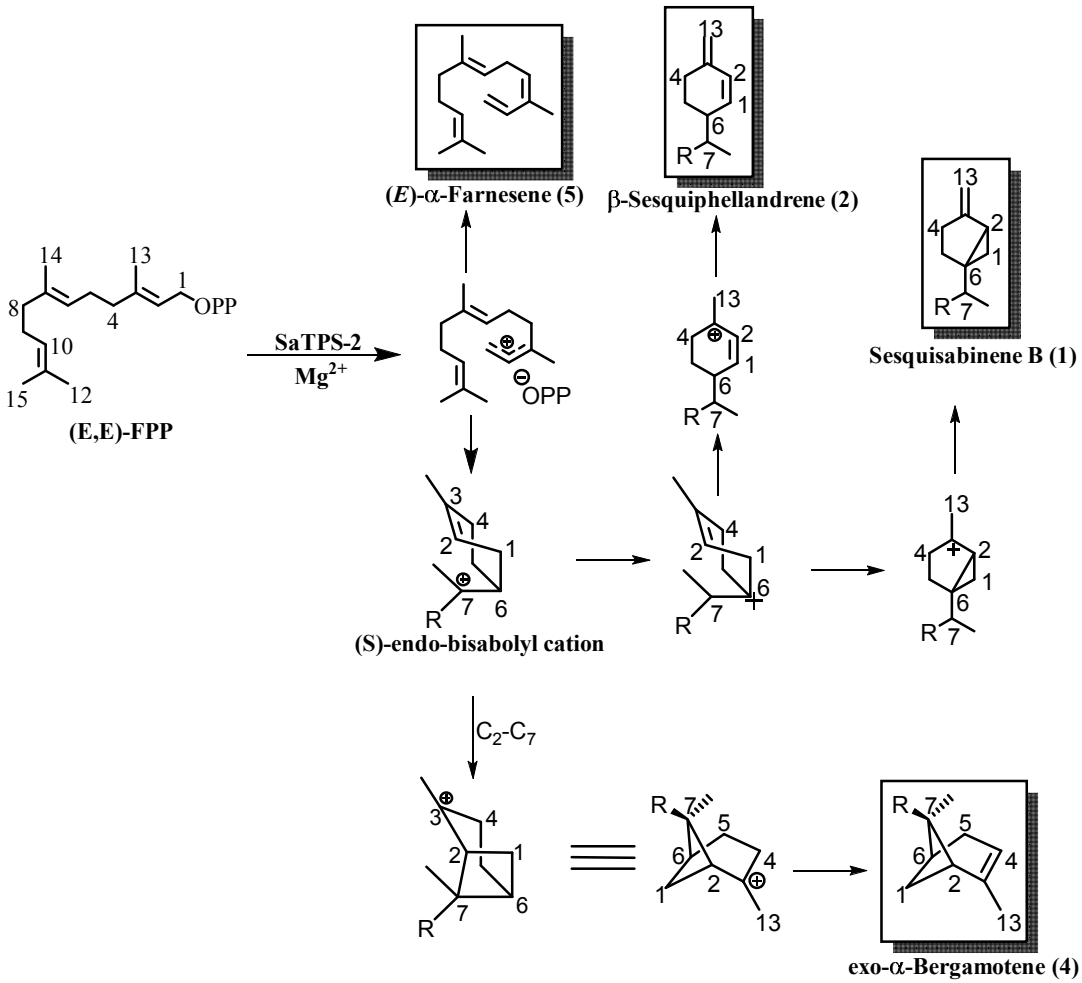


Figure 3.3.5.9: EI Mass spectrum of (*E*)-α-Farnesene (**5**) produced in assay mixture of *SaTPS2*



Scheme 3.3.5: Proposed mechanism for the formation of sesquiterpene mixture from *SaTPS2* upon incubation with (*E,E*)-FPP

3.3.5.4 Kinetic characterization of *SaTPS2*

Kinetic constants for *SaTPS2* were determined for (*E,E*)-FPP using GC with varying substrate concentrations from 0.5 µM to 60 µM. Amount of product formed was calculated with the help of known amount of dodecane added as an internal standard in the reaction samples. The kinetic parameters were determined using Graph pad Prism software 6.04. Kinetic parameters for *SaTPS2* were estimated to be K_m = 15.30µM, V_{max} = 0.71µM/min, K_{cat} = 0.1418/min and K_{cat}/K_m = 0.009. These results clearly indicate that *SaTPS1* is naturally more favored over *SaTPS2* owing to its better kinetic parameters.

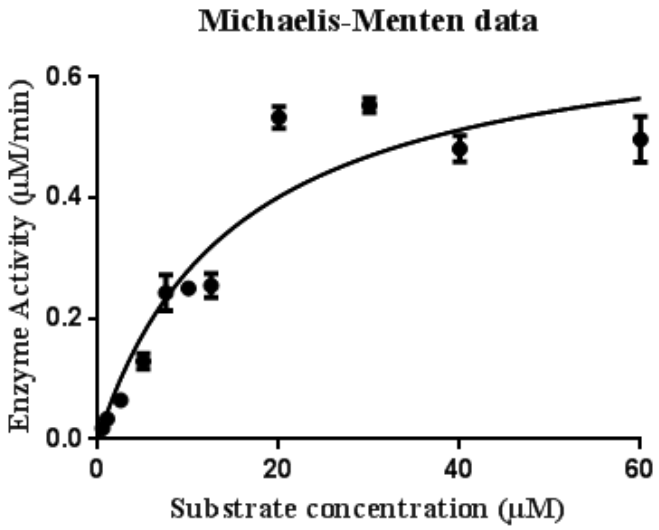


Figure 3.3.5.10: Michaelis Menten plot for *SaTPS2* kinetics with varying substrate concentration

Sesquisabinene B and its hydroxy derivative sesquisabinene hydrate are reported to be present in several plant species[27,28], but till today no report is available on the characterization of sesquisabinene B synthase. Functional characterization of *SaTPS1* and *SaTPS2* revealed that both the enzyme catalyzes the exclusive formation of sesquisabinene B from (*E,E*)-FPP as substrate with varying kinetic parameter.

Section 3.3.6: Bacterial production of *SaTPS1* and *SaTPS2* metabolites, real time analysis and phylogenetic analysis

3.3.6.1 Cloning of *SaFDS* in pETDuet-1 MCS 1

ORF of *SaFDS*, as discussed earlier, was cloned in pETDuet-1 in multiple cloning site 1 using BamHI restriction site with N-terminal His_6 under the control of T7 RNA polymerase for active protein expression. Colony PCR was performed for screening of positive clones and sequenced to confirm the frame of *SaFDS*.

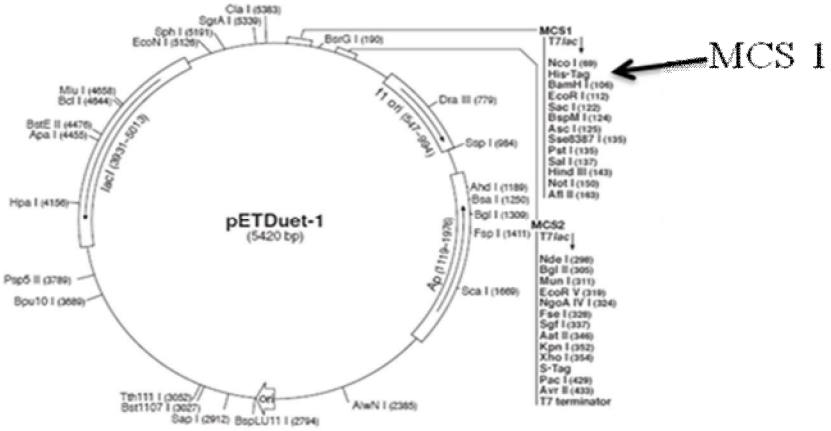


Figure 3.3.6.1: Vector map of cloning of *SaFDS* in pETDuet-1 in MCS1 with BamH1 cloning site

3.3.6.2 Cloning of *SaTPS1* and *SaTPS2* in MCS 2 of *SaFDS*-pETDuet-1

pETDuet-1:*SaFDS* was used to clone *SaTPS1* and *SaTPS2* in second multiple cloning site using different sets of restriction enzyme, as discussed in methodology section, under the control of T7 RNA polymerase for expression of active protein without having N terminal His_6 tag. Positive clones were screened by colony PCR as discussed in methodology section and sequenced to confirm the presence of insert and their orientation. Two independent clones for both the synthase were selected and expressed to check the bacterial production of their metabolites.

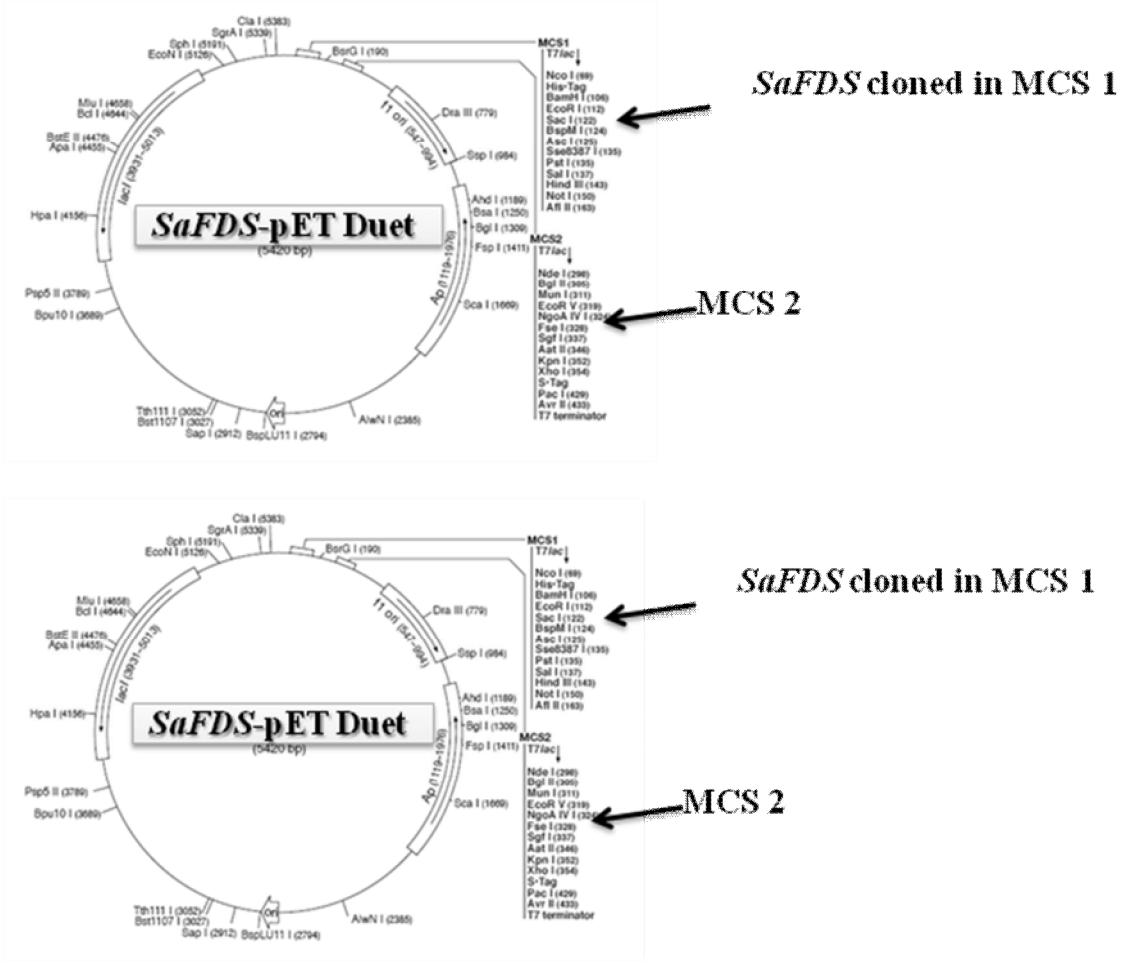


Figure 3.3.6.2: Cloning strategy for cloning of *SaTPS1* and *SaTPS2*, **A)** Vector map of cloning of *SaTPS1* in pETDuet-1-*SaFDS* in MCS 2, **B)** Vector map of cloning of *SaTPS2* in pETDuet-1:*SaFDS* in MCS 2

3.3.6.3 Bacterial production of *SaTPS1* and *SaTPS2* metabolites in C41 DE3 cells

Microbial production of sesquisabinene B from both the sesquiterpene synthases using *in vivo* expression system: pETDuet-1:*SaFDS*:*SaTPS1*/*SaTPS2* were performed in C41DE3 cells containing pRARE plasmid. No detectable level of sesquisabinene was found in culture containing empty vector, whereas it was detected in hexane extract of both the overnight grown culture media of pETDuet containing *SaFDS* and *SaTPS1*/*SaTPS2*. Small fraction of sesquisabinene B was detected in pellet fraction of both the terpene synthases.

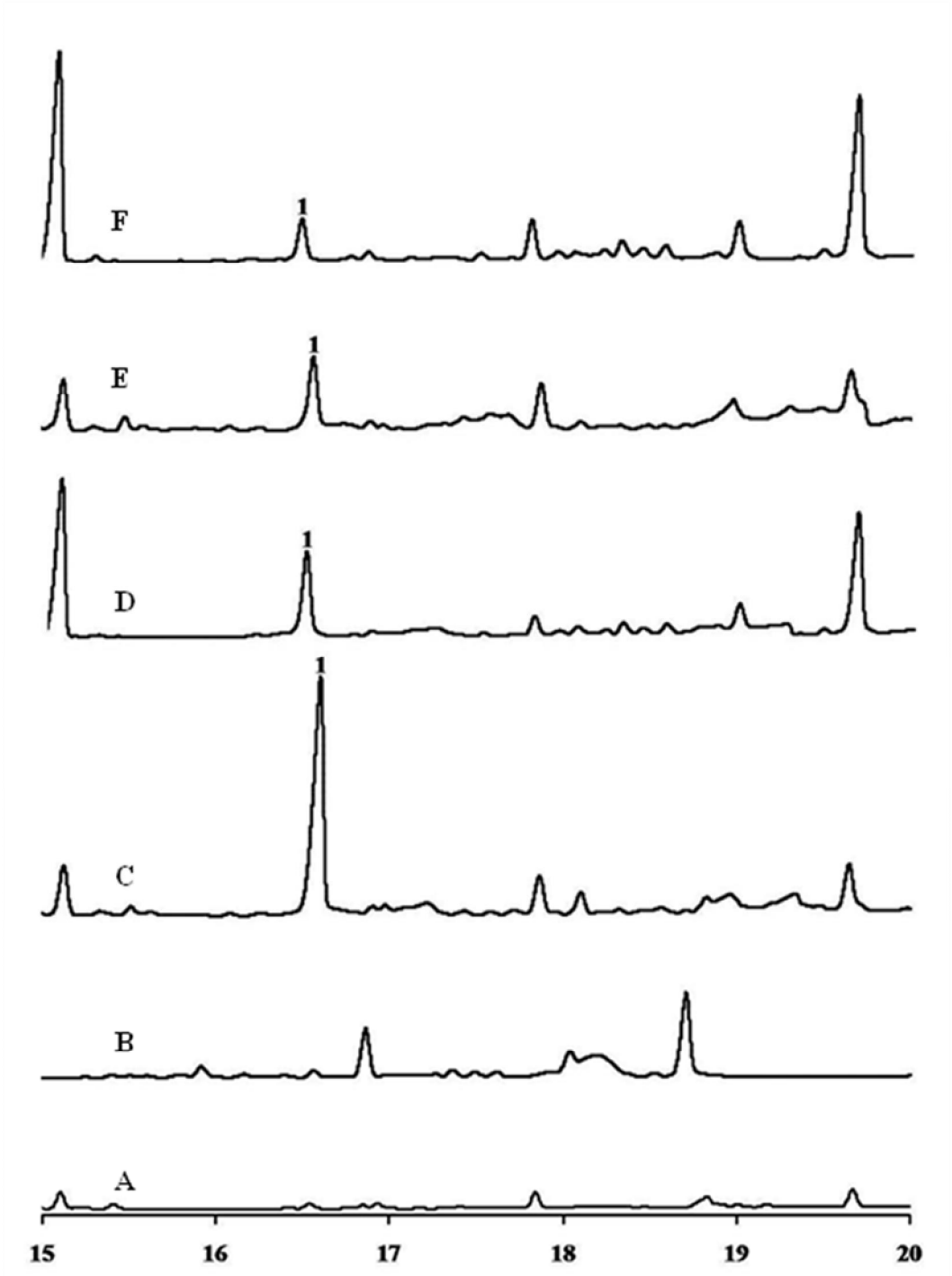


Figure 3.3.6.3: GC chromatogram of *SaTPS1* and *SaTPS2 in vivo* production, **A)** Hexane extract of supernatant of empty vector control**, B)** hexane extract of pellet of empty vector, **C)** GC profile of hexane extract of supernatant of *SaTPS1*, **D)** hexane extract of pellet of *SaTPS1I*, **E)** GC profile of hexane extract of supernatant of *SaTPS2*, **F)** hexane extract of pellet of *SaTPS2*

The yield of sesquisabinene from both the sesquiterpene synthase corresponds to their kinetic parameters, from which it was observed that *SaTPS1* is a more

efficient enzyme than *SaTPS2*. Sesquisabinene production by *SaTPS1* was 1-2 mg/L of bacterial culture, whereas the yield of sesquisabinene was found to be as low as 0.5 mg/L by *SaTPS2*. These results validated the *in vivo* activity of *SaTPS1* and *SaTPS2* in non-optimized bacterial strain and demonstrated the feasibility of metabolic engineering of *SaTPS1* in a heterologous system for the large scale production of sesquisabinene B.

3.3.6.4 Semi-quantitative real time PCR for *SaTPS1*, *SaTPS2* and *SaβBS*

To check the transcription level of *SaTPS1*, *SaTPS2* and *SaβBS*, semi-quantitative real time PCR was performed (as discussed in methodology). The cDNA dilution was normalized by amplifying 750 bp of 18S rRNA fragment, and 1:10 dilution (15ng) was selected (Figure 3.3.6.4.1) for semi-quantitative PCR of *SaTPS1*, *SaTPS2* and *SaβBS*. PCR amplification of *SaTPS1*, *SaTPS2* and *SaβBS* using same amount of cDNA and at same PCR conditions resulted in a significant difference at their transcript copy numbers (Figure 3.3.6.4.2). Transcription level of *SaTPS1* is 2-3 times more than *SaTPS2* indicating that *SaTPS1* is naturally more preferred over *SaTPS2*. These results correlate with the kinetic parameters. However, transcription level of β-bisabolene synthase is comparable to *SaTPS1* expression level.

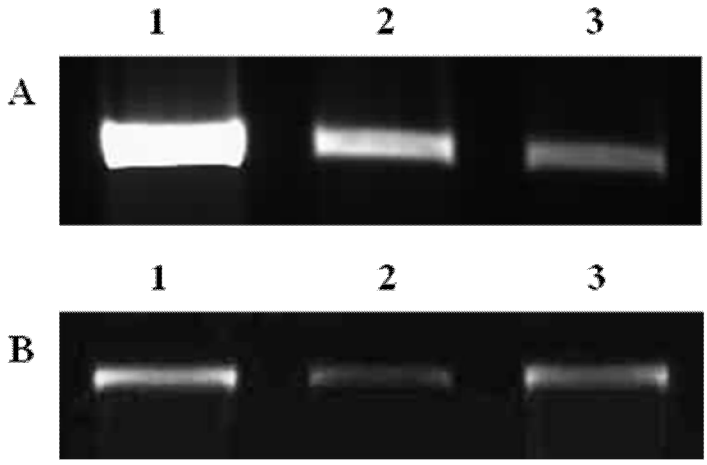


Figure 3.3.6.4: Semi-quantitative real time PCR gel image, **A)** 18S rRNA semi-qPCR, **Lane 1:** undiluted cDNA, **Lane 2:** 1:5 diluted cDNA (30ng), **Lane 3:** 1:10 diluted cDNA (15ng) , **B)** *SaTPS1*, *SaTPS2*, and *SaβBS* semi- qPCR with 1:10 diluted cDNA (15ng), **Lane 1:** semi-qPCR of *SaTPS1*, **Lane 2:** semi-qPCR of *SaTPS2*, **Lane 3:** semi-qPCR of *SaβBS*

3.3.6.5 Phylogenetic analysis of terpene synthases isolated from *S. album*

A neighbor joining phylogenic tree placed all the full-length sequence of terpene synthase isolated from *S. album* in a separate clades. *SaSS* forms a separate branch with other terpene synthase (*SaβBS*, *SaTPS1* and *SaTPS2*). The other branch is diverse into two nodes, one representing beta bisabolene synthase (*SaβBS*) and the other getting separated in two clades representing sesquisabinene B synthase (*SaTPS1* and *SaTPS2*), respectively (Figure 3.3.6.5). The isoform of sesquisabinene B synthase (*SaTPS1* and *SaTPS2*) as shown in phylogenetic tree are evolved from a common ancestor indicating *SaTPS1* to be of recent origin.

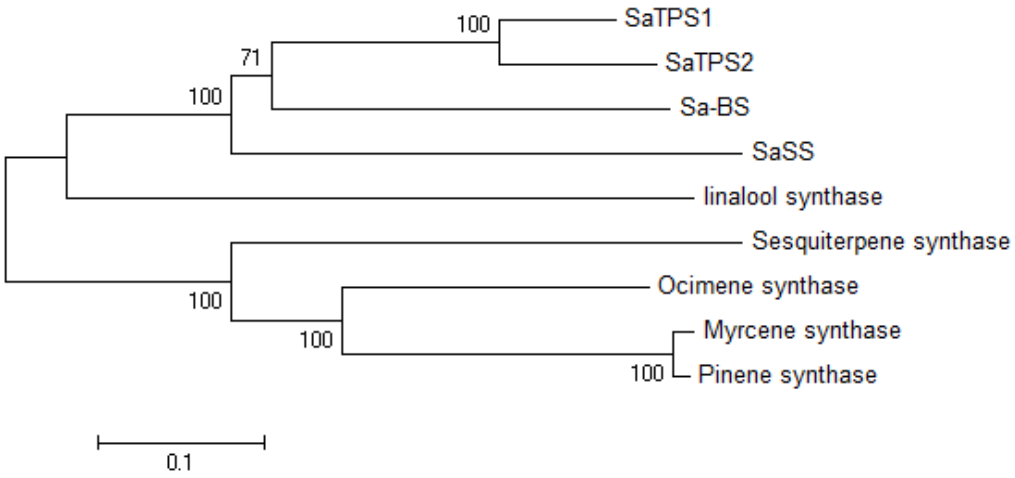
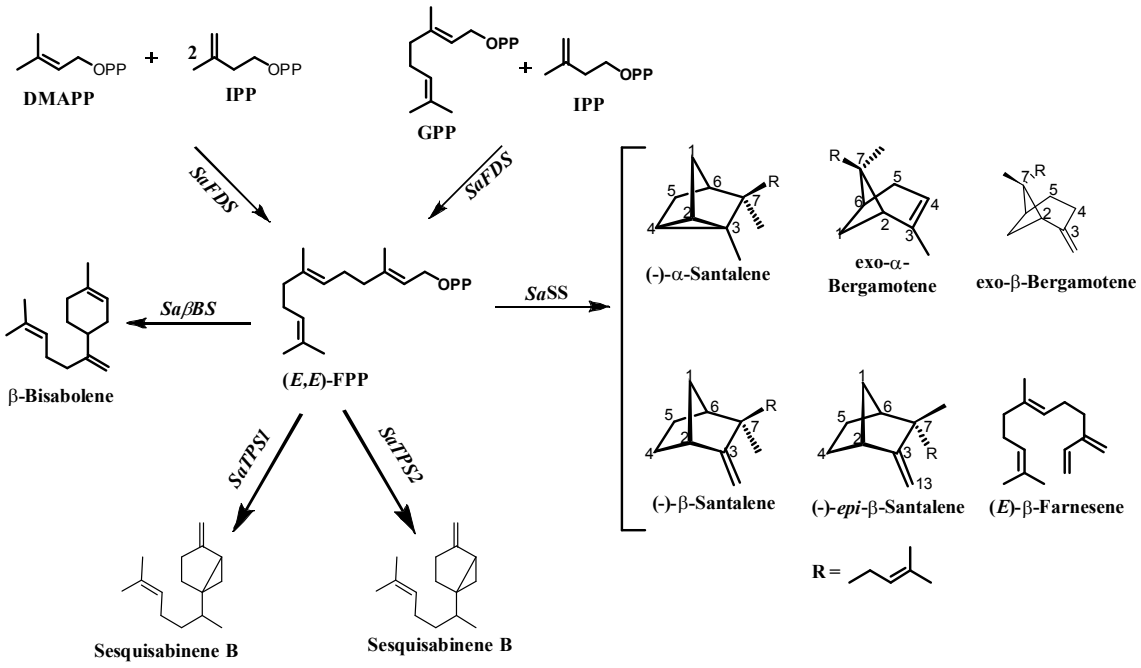


Figure 3.3.6.5: Phylogenetic analysis of terpene synthases isolated from *S. album*, Sequence used for phylogenetic tree construction are: *SaTPS1* (KJ665776), *SaTPS2* (KJ665777), *SaβBS* (KJ665778), *SaSS* (KF011938), Linalool synthase from *Backhousia citriodora* (BAG82825), Sesquiterpene synthase 6 from *Valeriana officinalis* (AGB05615), (*E*)-β-ocimene/myrcene synthase from *Vitis vinifera* (ADR74206), Myrcene synthase from *Quercus ilex* (Q93X23), pinene synthase from *Quercus ilex* (CAK55186)

3.4 Conclusion

To establish the sesquiterpene biosynthesis in Sandalwood, we have isolated and characterized one prenyl transferase (*SaFDS*) and two classes of sesquiterpene synthases from the interface of heartwood and sapwood of Indian Sandalwood. One class represents santalene synthase (SaSS), a multiproduct enzyme, which forms mixture of six products from a single prenyl diphosphate substrate with varying carbon skeletons: α-santalene (41.22 %), exo-α- bergamotene (21.62 %), *epi*-β-santalene (4.36 %), (*E*)-β-farnesene (1.13 %), β-santalene (29.48 %), and exo-β-bergamotene (2.19 %). Whereas, the second class of sesquiterpenes represents, β-bisabolene synthase (*SaβBS*), *SaTPS1* and *SaTPS2*, which form the dominant single products, β-bisabolene, and sesquisabinene B, respectively. A schematic representation is shown in scheme 3.4 for the biosynthesis of sesquiterpenes in Indian Sandalwood.



Scheme 3.4: Schematic representation of sesquiterpenes biosynthesis in Indian Sandalwood *Santalum album*

We also tested the bacterial production by incorporating *SaTPS1* and *SaTPS2* genes in pET-Duet vector along with *SaFDS* and expression was performed in C41DE3 cells containing pRARE plasmid. We observed that sesquisabinene B could be produced in 1-2 mg/L of non-optimized bacterial culture. These results validated the *in vivo* activity of *SaTPS1* and *SaTPS2* and demonstrated the feasibility

of metabolic engineering of *SaTPS1* in a heterologous system for large-scale production of sesquisabinene B in large quantity due to better kinetic parameter. *SaTPS1* and *SaβBS* producing exclusively sesquisabinene B and beta bisabolene, respectively, could be utilized for the production of jet-engine fuel. In recent years, sesquiterpenes of farnesene and bisabolene skeletons have been recognized as replacements for petroleum-derived jet-engine fuel.

Appendix 3: Colony PCR screening for gene cloning

3.1 Cloning of *SaFDS* in pRSETB vector frame

Colony PCR screening of farnesyl diphosphate synthase cloned in pRSETB vector frame with T7 promoter primer and *SaFDS* reverse primer.

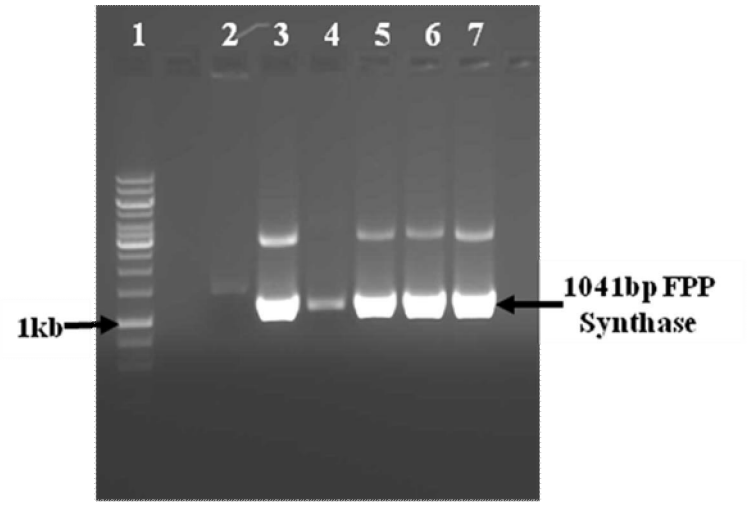


Figure 3.1: FPP Synthase full length ORF cloned in pRSETB, **Lane 1:** 1 Kb DNA ladder, **Lane 2:** no insert control, **Lanes 3-7:** PCR with T7 promoter and *SaFDS* reverse primer

3.2 Cloning of 3′ RACE amplicon of *SaSS* in TA vector

Colony PCR screening of santalene synthase cloned in pET32b vector frame with *SaSS* forward primer and T7 reverse primer.

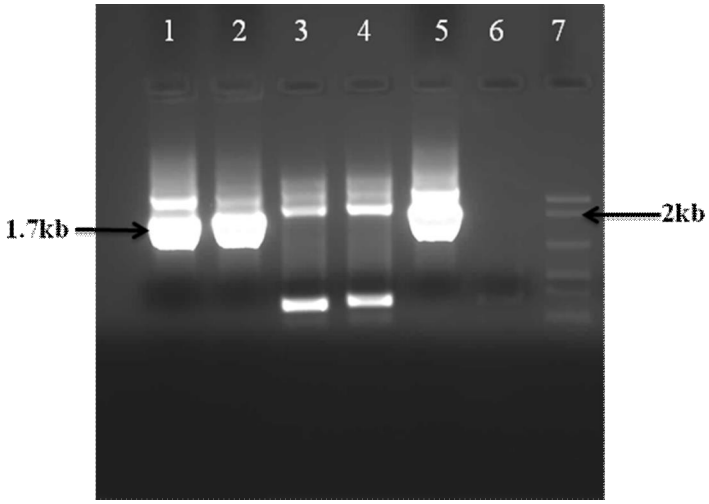


Figure 3.2: Colony PCR screening of 3′ RACE amplicon of *SaSS* cloned in TA vector, **Lanes 1-5:** PCR with M13 forward and reverse primer, **Lane 6:** no insert control, **Lane 7:** 1 Kb DNA ladder

3.3 Cloning of *SaSS* in pET32b vector frame

Colony PCR screening of santalene synthase cloned in pET32b vector frame with T7 promoter primer primer and T7 reverse primer

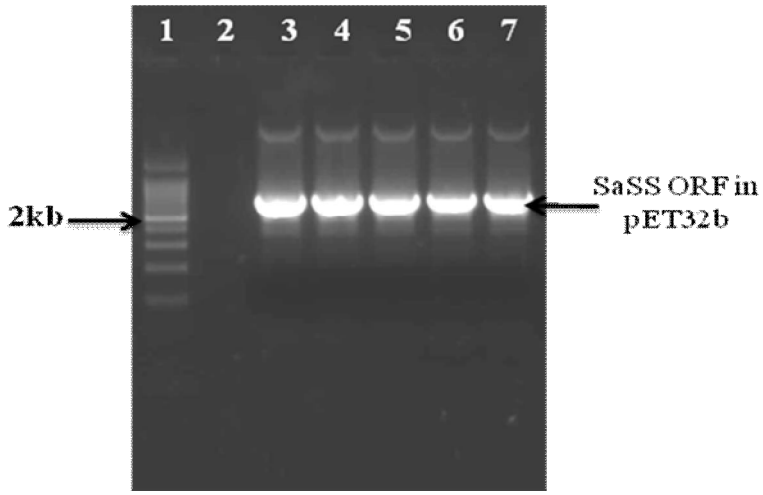


Figure 3.3: Santalene synthase full length ORF cloned in pET32b, **Lane 1:** 1 Kb DNA ladder, **Lane 2:** no insert control, **Lanes 3-7:** PCR with T7 promoter primer and T7 reverse primer

3.4 Cloning of 3′ RACE amplicon of *SaβBS* in TA cloning vector

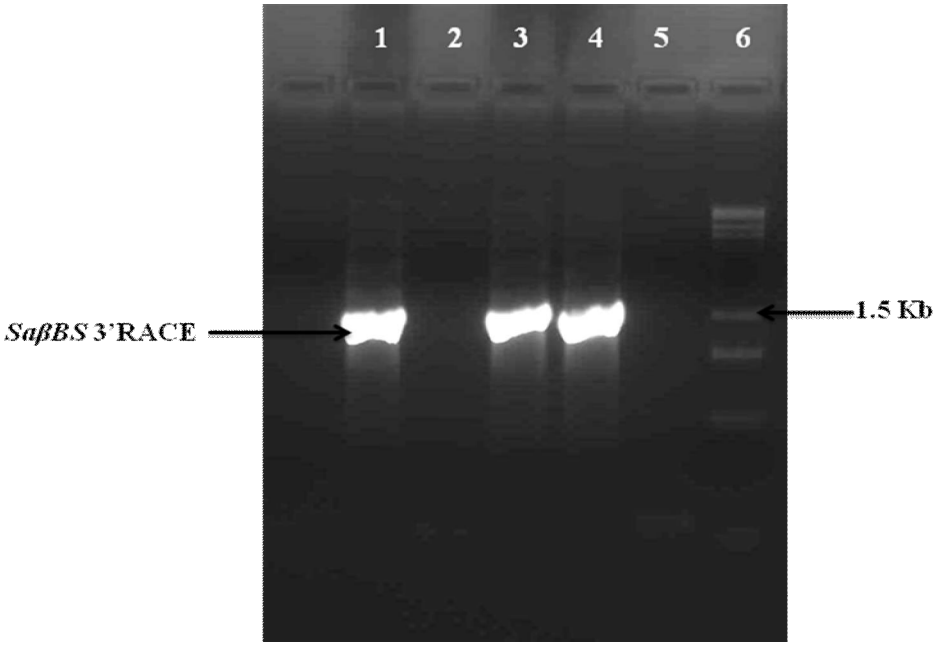


Figure 3.4: Colony PCR screening of 3′ RACE amplicon of *SaβBS* cloned in TA vector, **Lanes 1-4:** PCR with M13 forward and reverse primer, **Lane 5:** no insert control, **Lane 6:** 1 Kb DNA ladder

3.5 Cloning of ORF of *SaβBS* in pET28a

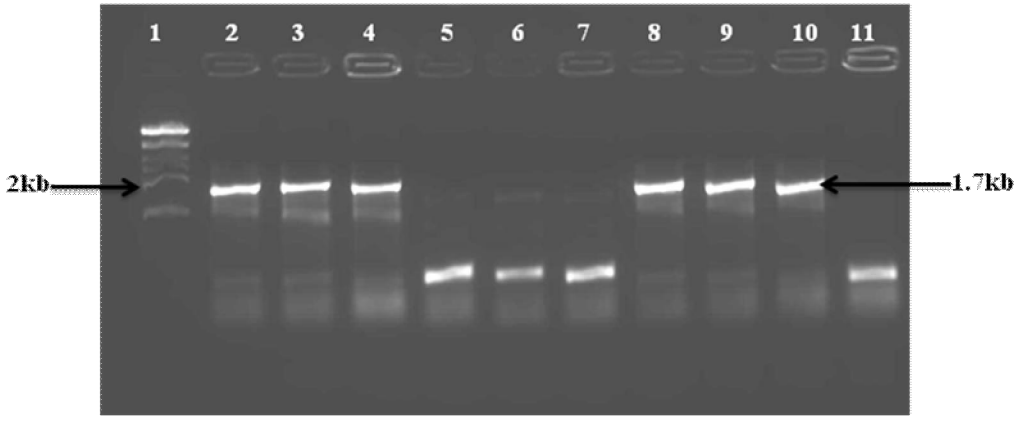


Figure 3.5: Colony PCR screening for ORF of *SaβBS* cloned in pET28a vector, **Lane 1:** 1 Kb DNA ladder, **Lanes 2-10:** PCR amplicon with T7 promoter and *SaβBS* reverse primer, **Lane 11:** no insert control

3.6 Cloning of 3' RACE amplicon of *SaTPS1* in TA cloning vector

Colony PCR screening of 3' RACE amplicon of *SaTPS1* cloned in TA vector with M13 forward and reverse primer.

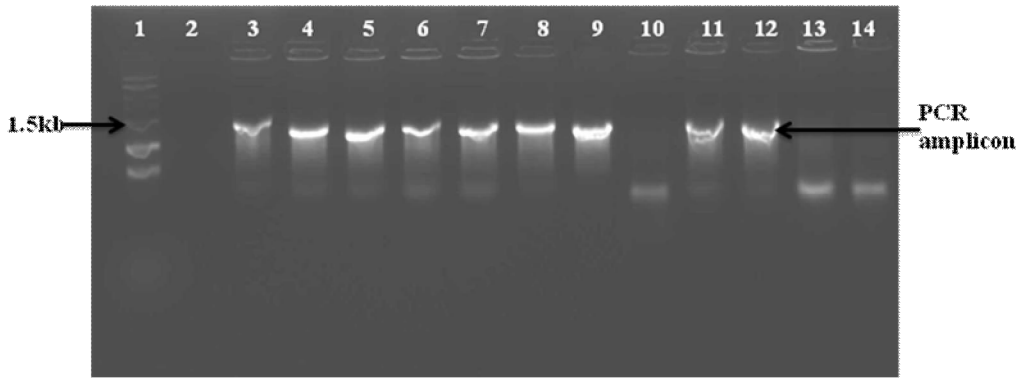


Figure 3.6: Colony PCR screening of 3′ RACE amplicon of *SaTPS1* cloned in TA vector, **Lane 1:** 1 Kb DNA ladder, **Lane 2:** no insert control, **Lane 3-14:** PCR with M13 forward and reverse primer

3.7 Cloning of ORF of *SaTPS1* in pET28a

Colony PCR screening of ORF of *SaTPS1* cloned in pET28a vector frame with T7 promoter and *SaTPS1* reverse primer.

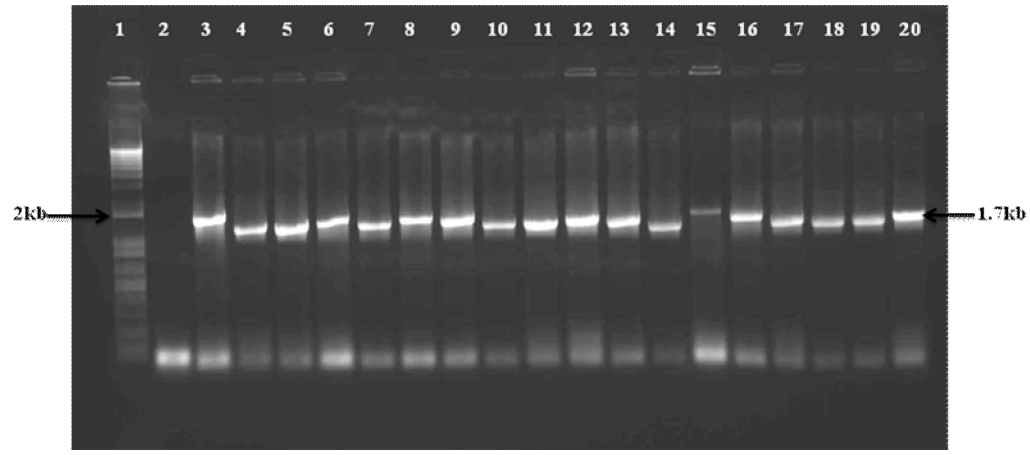


Figure 3.7: Colony PCR screening for ORF of *SaTPS1* cloned in pET28a vector, **Lane 1:** 1 Kb DNA ladder, **Lane 2:** no insert control, **Lanes 3-20:** PCR amplicon with T7 promoter and *SaTPS1* reverse primer

3.8 Cloning of 3' RACE amplicon of *SaTPS2* in TA cloning vector

Colony PCR screening of 3' RACE amplicon of *SaTPS2* cloned in TA vector with M13 forward and reverse primers.

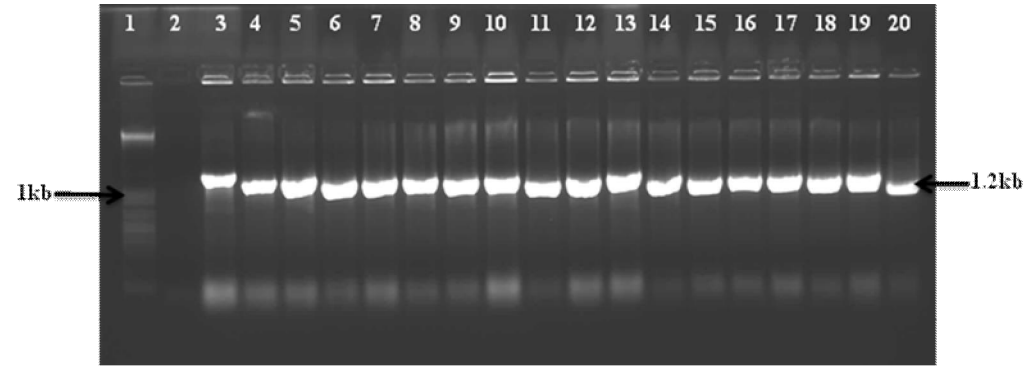


Figure 3.8: Colony PCR screening of 3' RACE amplicon of *SaTPS1* cloned in TA vector, **Lane 1:** 1 Kb DNA ladder, **Lane 2:** no insert control, **Lanes 3-20:** PCR with M13 forward and reverse primer

3.9 Cloning of ORF of *SaTPS2* in pET28a

Colony PCR screening for ORF of *SaTPS2* cloned in pET28a vector frame with T7 promoter and *SaTPS2* reverse primer

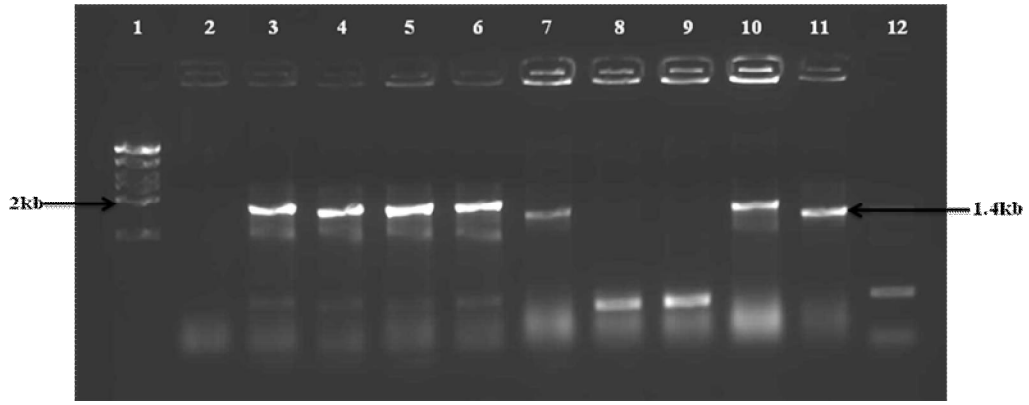


Figure 3.9: Colony PCR screening for ORF of *SaTPS2* cloned in pET28a vector, **Lane 1:** 1 Kb DNA ladder, **Lane 2:** no insert control, **Lanes 3-12:** PCR amplicon with T7 promoter and *SaTPS2* reverse primer

3.10 Cloning of *SaFDS* in pETDuet1:

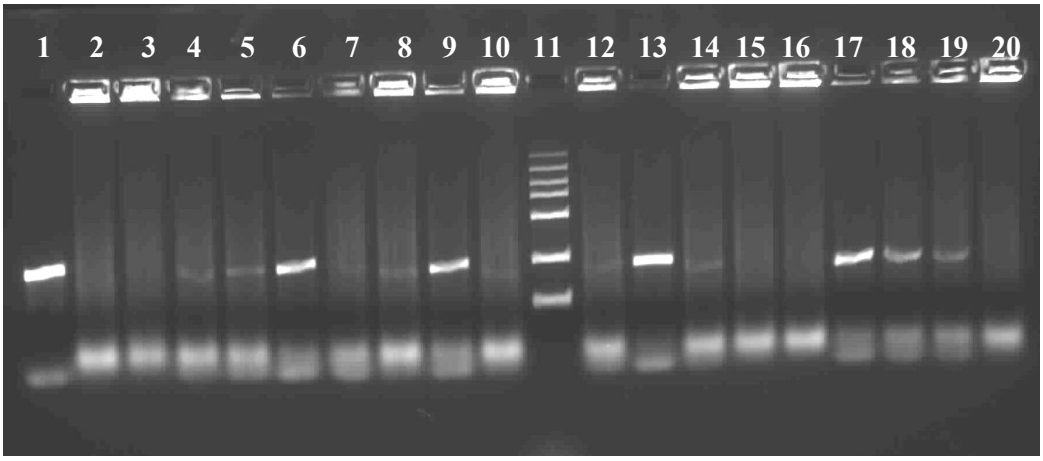


Figure 3.10: Colony PCR screening for screening of positive clones of *SaFDS* in pETDuet-1, **Lanes 1-10 and 12-19:** Colony PCR with pETDuet upstream primer 1 and *SaFDS* reverse primer, **Lane 11:** 1 Kb DNA ladder, **Lane 20:** No insert control

3.11 Cloning of *SaTPS1* in MCS 2 of *SaFDS*-pETDuet-1

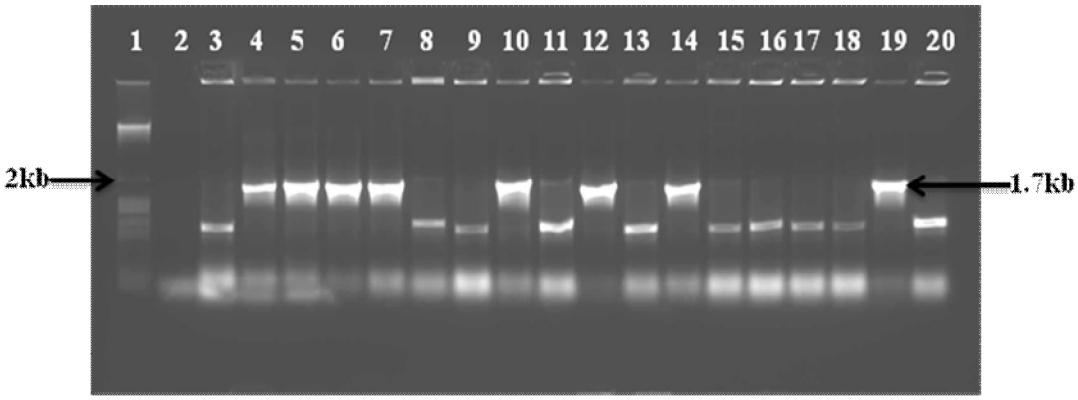


Figure 3.11: Colony PCR screening for screening of positive clones of *SaTPS1* in *SaFDS*-pETDuet-1, **Lane 1:** 1 Kb DNA ladder, **Lane 2:** No insert control, **Lanes 3-20:** Colony PCR with pETDuet upstream primer 2 and *SaTPS1* reverse primer

3.12 Cloning of *SaTPS2* in MCS 2 of *SaFDS*-pETDuet-1

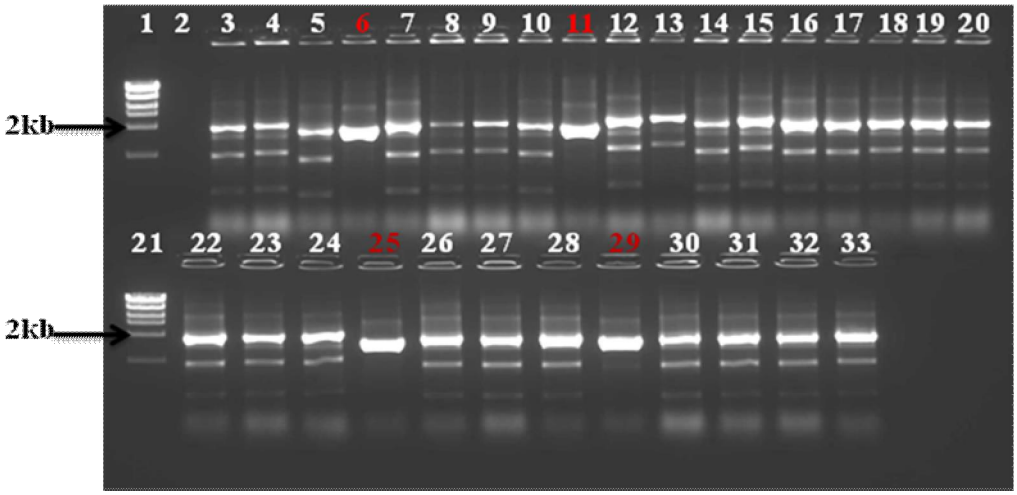


Figure 3.12: Colony PCR screening for screening of positive clones of *SaTPS2* in *SaFDS*-pETDuet-1, **Lanes 1 and 21:** 1 Kb DNA ladder, **Lane 2:** No insert control, **Lanes 3-33:** Colony PCR with pETDuet upstream primer 2 and *SaTPS2* reverse primer

3.5 References

(1) Bouvier, F.; Rahier, A.; Camara, B. *Prog. Lipid Res.* **2005**, *44*, 357-429.

(2) Pichersky, E.; Noel, J. P.; Dudareva, N. *Science* **2006**, *311*, 808-811.

(3) Ding, V. D. H.; Sheares, B. T.; Bergstrom, J. D.; Ponpipom, M. M.; Perez, L. B.; Poulter, C. D. *Biochem. J.* **1991**, *275*, 61-65.

(4) Anderson, M. S.; Yarger, J. G.; Burck, C. L.; Poulter, C. D. *J. Biol. Chem.* **1989**, *264*, 19176-19184.

(5) Hugueney, P.; Camara, B. *FEBS Letters* **1990**, *273*, 235-238.

(6) Wong, T. K.; Decker, G. L.; Lennarz, W. J. *J. Biol. Chem.* **1982**, *257*, 6614-8.

(7) Lee Adair Jr, W.; Cafmeyer, N. *Arch. Biochem. Biophys* **1987**, *259*, 589-596.

(8) Rilling, H. C.; Bruenger, E.; Epstein, W. W.; Kandutsch, A. A. *Biochem. Bioph. Res. Co.* **1989**, *163*, 143-148.

(9) Momose, K.; Rudney, H. *J. Biol. Chem.* **1972**, *247*, 3930-3940.

(10) Trumpower, B. L.; Houser, R. M.; Olson, R. E. *J. Biol. Chem.* **1974**, *249*, 3041-3048.

(11) Bommareddy, A.; Rule, B.; VanWert, A. L.; Santha, S.; Dwivedi, C. *Phytomedicine* **2012**, *19*, 804-811.

(12) Klayman, D. L. *Science* **1985**, *228*, 1049-1055.

(13) Rude, M. A.; Schirmer, A. *Curr. Opin. Microbiol.* **2009**, *12*, 274-281.

(14) Peralta-Yahya, P. P.; Keasling, J. D. *Biotechnol. J.* **2011**, *5*, 147-162.

(15) Jones, C. G.; Ghisalberti, E. L.; Plummer, J. A.; Barbour, E. L. *Phytochemistry* **2006**, *67*, 2463-2468.

(16) Daramwar, P. P.; Srivastava, P. L.; Priyadarshini, B.; Thulasiram, H. V. *Analyst*, *137*, 4564-4570.

(17) Baldovini, N.; Delasalle, C.; Joulain, D. *Flavour Frag. J.*, *26*, 7-26.

(18) Kim, T. H.; Ito, H.; Hatano, T.; Hasegawa, T.; Akiba, A.; Machiguchi, T.; Yoshida, T. *J. Nat. Prod.* **2005**, *68*, 1805-1808.

(19) Hongratanaworakit, T.; Heuberger, E.; Buchbauer, G. *Planta Med* **2004**, *70*, 3-7.

(20) Kaur, M.; Agarwal, C.; Singh, R. P.; Guan, X. M.; Dwivedi, C.; Agarwal, R. *Carcinogenesis* **2005**, *26*, 369-380.

(21) Dwivedi, C.; Guan, X.; Harmsen, W. L.; Voss, A. L.; Goetz-Parten, D. E.; Koopman, E. M.; Johnson, K. M.; Valluri, H. B.; Matthees, D. P. *Cancer Epidem. Biomar.* **2003**, *12*, 151-156.

(22) Kim, T. H.; Ito, H.; Hayashi, K.; Hasegawa, T.; Machiguchi, T.; Yoshida, T. *Chem. Pharm. Bull.* **2005**, *53*, 641-644.

(23) Jones, C. G.; Moniodis, J.; Zulak, K. G.; Scaffidi, A.; Plummer, J. A.; Ghisalberti, E. L.; Barbour, E. L.; Bohlmann, J. *J. Biol. Chem.* **2011**, *286*, 17445-17454.

(24) Tamura, K.; Stecher, G.; Peterson, D.; Filipski, A.; Kumar, S. *Mol. Biol. Evol.* **2013**.

(25) Daramwar, P. P. *University of Pune* **2014**, *Ph. D. thesis*.

(26) Adams, R. P.; Zanoni, T. A.; van Beek, T. A.; Posthumus, M. A.; van de Haar, C. *J. Essent. Oil Res.* **1998**, *10*, 175-178.

(27) Radulovic, N. S.; Dordevic, N. D.; Zlatkovic, B. K.; Palic, R. M. *Chem. Pap.* **2008**, *62*, 603-607.

(28) Sonboli, A.; Bahadori, M. B.; Dehghan, H.; Aarabi, L.; Savehdroudi, P.; Nekuei, M.; Pournaghi, N.; Mirzania, F. *Chem. Biodivers.* **2013**, *10*, 687-694.

Chapter 4

Santalene synthase: A comprehensive site directed mutagenesis study to understand the dynamic nature of santalene synthase active site pocket

Directed engineering of enzyme function to produce predefined product remains a major challenge in protein engineering. Structural and biochemical analyses indicated that the active or binding site residues are important for the functional plasticity of the terpene synthases which carry out the cyclization of acyclic prenyl diphosphate into diversified carbon skeletons. Here we describe the homology structural model based engineering of santalene synthase (*SaSS*), which was isolated and cloned from the sapwood-heartwood transition zone of Indian Sandalwood *Santalum album* Linn (Chapter 3). Santalene synthase catalyses the cyclization of farnesyl diphosphate (FPP) into a mixture of α-, β-santalenes, and exo-α-bergamotene as major products and *epi*-β-santalene, (*E*)-β-farnesene, and exo-β-bergamotene as minor products. In order to understand the structural basis for cyclization of santalenes and bergamotene, the residues involved in stabilizing the intermediates in carbocation cascade were identified using homology structural model, which was built using co-ordinates of 5-*epi*-aristolochene synthase. Twenty-two residues surrounding the active site pocket were selected and 34 mutants were constructed (29 single mutants, four double mutants and one triple mutant). The product ratio analyses using (*E,E*)-FPP as the substrate indicated that mutants, S290G, T318A, R474M, R474L, T467A, W293F, Y37F, Y396F and Y539F produces exclusively exo-α-bergamotene whereas mutant Y539W produces endo-α-bergamotene and exo-β-bergamotene to significant level. Mutants I422A produced a very different monocyclic sesquiterpene, β-curcumene, as a major product by stabilizing the bisabolyl carbocation, whereas mutant L427A resulted in the formation of an acyclic sesquiterpene (*E*)-β-farnesene in a significant amount by minimizing the hydrophobic interaction required for 1-6 closure to form the bisabolyl carbocation.

4.1 Introduction

Terpenoids, also referred to as isoprenoids, are the largest group of natural products present in all living systems with immense diversity in their structure and function. Isoprenoid compounds play crucial metabolic and structural roles in cells[1,2]. Although terpenes traditionally have been used in fragrances and flavors, they have the potential to serve as advanced biofuels and chemical precursors[3,4]. Over 70,000 naturally occurring isoprenoids with a truly incredible array of carbon skeletons and functional groups have been reported[5]. Sesquiterpenoids constitute an abundant group of isoprenoids with over 7000 molecules and over 300 stereochemically distinct hydrocarbon skeletons. The basic building reaction in the isoprenoid biosynthetic pathway is divided into two steps: the first step is chain elongation by head to tail condensation of C5 unit isopentyl diphosphate and its isomer dimethylallyl diphosphate into geranyl diphosphate (GPP), farnesyl diphosphate (FPP) and geranylgeranyl diphosphate (GGPP), whereas, the second step deals with the cyclization of these linear substrates catalyzed by terpene synthases to form a diverse class of terpenes from respective prenyl diphosphates. Sesquiterpene synthases catalyze the metal-dependent cyclization of the universal substrate, farnesyl diphosphate (FPP) for the formation of skeletons containing fifteen carbons, by generating carbocations which undergo a sequence of intra-molecular cyclizations involving a number of Wagner-Meerwein rearrangements to form single or multiple rings[6]. Despite the low level of sequence similarity, all the terpene synthases share a similar active site pocket[7-10]. The molecular recognition and initial ionization of the substrate diphosphate group is triggered by metal coordinated interactions and hydrogen bond interactions by amino acids located in the upper part of the active site pocket. Complexity in terpene skeleton arises from the double bond isomerization, proton elimination, stereo-specific hydride, methyl, and methylene migrations, and solvation of carbocation by water to form terpene alcohols[11]. The diversity in the product structure and stereochemistry are precisely determined by folding of the acyclic diphosphate substrate in the substrate binding pocket of the terpene synthase[12,13]. Several biochemical studies have suggested that, proteins have the ability to generate novel or altered function with a small number of amino acid (plasticity residues) substitutions in the active site pocket[13-16]. Amino acid changes in the active site of the terpene cyclases changes the electronic distribution as well as

geometry of active site cleft which in turn, can modify the bond forming reactions including cyclization, cation rearrangements, or deprotonation[17]. In this direction, directed engineering of the terpene synthase is emerging as a prolific tool for fine tuning the terpene synthases for selective formation of desired product[18,19]. Recently, structures of several sesquiterpene synthases have been determined by X-ray diffraction studies, which greatly facilitated the study on role of catalytically important amino acids as well as engineering of the sesquiterpene synthase activities[8-10,20-22].

To investigate the evolutionary origins of santalene biogenesis, we have isolated the gene, which encodes for santalene synthase (*SaSS*) from the transition zone of heartwood and sapwood of *S. album* using transcriptome screening and comparing sequence homology (Chapter 3). Santalene synthase from *S. album* catalyzes the divalent metal-ion dependent cyclization of (*E,E*)-FPP into a mixture of (+)-α-santalene, (-)-exo-α-bergomotene, and (-)-β-santalene as major sesquiterpene hydrocarbons along with (+)-epi-β-santalene, (*E*)-β-farnesene and exo-β-bergamotene as minor products[23]. Further the cytochrome P450 mono-oxygenase system will hydroxylate at the cis methyl group of the side chain to form (*Z*)-α-santalol, (*Z*)-α-exo-bergomotol, (*Z*)-epi-β-santalol, and (*Z*)-β-Santalol. The sesquiterpene alcohols (*Z*)-α-santalol and (*Z*)-β-santalol together constitute over 80 % of the heartwood oil extracted from the well matured sandalwood tree[23,24].

In order to understand the structural basis for cyclization of santalenes and bergamotene, the amino acid residues present in and around the active site were identified using homology structural model, which was built using co-ordinates of 5-*epi*-aristolochene synthase[8].

4.2 Materials and Methods

4.2.1 Computational modeling and docking

Santalene synthase homology model was built using the co-ordinates of 5-*epi*-aristolochene synthase as template with Schrodinger-Maestro 1.6. The five lowest energy structures were validated using Ramachandran plot. Reaction intermediates were constructed by ChemBioDraw ultra 13 for docking study. To determine the amino acid interactions involved in the product formation, docking was performed using all the intermediates involved in santalene biosynthesis and images were generated by PYMOL (v.1.6).

4.2.2 Site–Directed mutagenesis

All the mutation reactions were carried out with a pET-32b expression vector (Novagen) harboring the coding sequence for santalene synthase as a template. The mutants were prepared using the QuikChange Lightning site directed mutagenesis kit from Stratagene according to the manufacturer's instructions. The mutagenic primers (with mutagenic bases shown in boldface and underlined) which were designed on the basis of homology based model of *SaSS* are listed in the table 4.2.2.1.

Table 4.2.2.1: Primer sequences for mutation

Primer name	Primer sequence (5'-3')
G323V&Y324F Forward	TGTTGACGAT**GTATTT**GACGTCT
G323V&Y324F Reverse	AGACGTC**AAATAC**ATCGTCAACA'
Y324F Forward	GACGATGGA**TTT**GACGTCTAT
Y324F Reverse	ATAGACGTC**AAA**TCCATCGTC'
T318A Forward	CGGAAGTGTACTC**GCA**GTTGTTGACG
T318A Reverse	CGTCAACAAC**TGC**GAGTACACTTCCG
Y539F Forward	CTCATTTCTTC**TTT**GAATTTGGGGA
Y539F Reverse	TCCCCAAATTC**AAA**GAAGAAATGAG
Y539W Forward	CTCATTTCTTC**TGG**GAATTTGGGGA
Y539W Reverse	TCCCCAAATTC**CCA**GAAGAAATGAG'
D549A Forward	TTTGGGGTGACG**GCT**AGCTGGACAAAG

D549A Reverse	CTTTGTCCAGCT**AGC**CGTCACCCCAAA
R474M Forward	GATGAGATGGCA**ATG**GGCGATAATCTG
R474M Reverse	CAGATTATCGCC**CAT**TGCCATCTCATC
R474L Forward	GATGAGATGGCA**CTA**GGCGATAATCTG
R474L Reverse	CAGATTATCGCC**TAG**TGCCATCTCATC
W293F Forward	TGCAGAGCTATATG**TTT**AGCTGCGCGATTGC
W293F Reverse	GCAATCGCGCAGCT**AAA**CATATAGCTCTGCA
I422A Forward	GGACTTGTTTCC**GCA**GGATTCCCTCTC
I422A Reverse	GAGAGGGAATCC**TGC**GGAAACAAGTCC
Q289A Forward	AATAATTTACTG**GCG**AGCTATATGTG
Q289A Reverse	CACATATAGCT**CGC**CAGTAAATTATT
S290A Forward	AATTTACTGCAG**GCC**TATATGTGGAG
S290A Reverse	CTCCACATATA**GGC**CTGCAGTAAATT
L427A Forward	GGATTCCCTCTC**GCG**TTAATCACGGG
L427A Reverse	CCCGTGATTAA**CGC**GAGAGGGAATCC
T467A Forward	AATGATATAGGA**GCG**TCTCCGGATGA
T467A Reverse	TCATCCGGAGA**CGC**TCCTATATCATT
W551F Forward	TGACGGATAGC**TTT**ACAAAGGTTGAT
W551 Reverse	ATCAACCTTTGT**AAA**GCTATCCGTCA
Y37F Forward	AGGATGGGAAAT**TTT**AAACCCAGCATT
Y37F Reverse	AATGCTGGGTTT**AAA**ATTTCCCATCCT
Y396F Forward	ACAGTGTAAATCA**TTC**CAGAAAGAAGCAA
Y396F Reverse	TTGCTTCTTTCTG**GAA**TGATTTACACTGT
N463I Forward	AGTCGCCTCATC**ATT**GATATAGGAACG
N463I Reverse	CGTTCCTATATC**AAT**GATGAGGCGACT
D470A Forward	GAACGTCTCCG**GCT**GAGATGGCAAG
D470A Reverse	CTTGCCATCTC**AGC**CGGAGACGTTC
F538Y Forward	AGGGTCTCATTTC**TAC**TATGAATTTGGGG
F538Y Reverse	CCCCAAATTCATA**GTA**GAAATGAGACCCT
S459F & L461F Forward	CTCCTCCCTCCTT**TTT**CGC**TTC**ATCAATGATATAG
S459F & L461F	CTATATCATTGAT**GAA**GCG**AAA**AAGGAGGGAGGAG

Reverse	
S290L Forward	ATAATTTACTGCAG**CTC**TATATGTGGAGCTG
S290L Reverse	CAGCTCCACATATA**GAG**CTGCAGTAAATTAT
S290G Forward	ATAATTTACTGCAG**GGC**TATATGTGGAGCTG
S290G Reverse	CAGCTCCACATATA**GCC**CTGCAGTAAATTAT
L427I Forward	TAGGATTCCCTCTC**ATC**TTAATCACGGGCTAC
L427I Reverse	GTAGCCCGTGATTAA**GAT**GAGAGGGAATCCTA
F538M Forward	GAGGGTCTCATTTC**ATG**TATGAATTTGGGGAT
F538MReverse	ATCCCCAAATTCATA**CAT**GAAATGAGACCCTC
E471Q Forward	ACGTCTCCGGAT**CAG**ATGGCAAGAGGC
E471Q Reverse	GCCTCTTGCCAT**CTG**ATCCGGAGACGT
D543N Forward	TATGAATTTGGG**AAT**GGCTTTGGGGTG
D543N Reverse	CACCCCAAAGCC**ATT**CCCAAATTCATA
D543L Forward	TATGAATTTGGG**CTT**GGCTTTGGGGTG
D543L Reverse	CACCCCAAAGCC**AAG**CCCAAATTCATA
N463D Forward	AGTCGCCTCATC**GAT**GATATAGGAACG
N463D Reverse	CGTTCCTATATC**ATC**GATGAGGCGACT
Y539L Forward	GGTCTCATTTCTTC**TTA**GAATTTGGGGATGGC
Y539L Reverse	GCCATCCCCAAATTC**TAA**GAAGAAATGAGACC
Y539S Forward	GGTCTCATTTCTTC**TCT**GAATTTGGGGATGGC
Y539S Reverse	GCCATCCCCAAATTC**AGA**GAAGAAATGAGACC

Mutant Y539F & T318A was prepared using the *SaSS* Y539F DNA as template and the primers of *SaSS* T318A to introduce the necessary additional mutation, while the Mutant Y539F & I422A was prepared using the *SaSS* Y539F DNA as template and primers of *SaSS* I422A. Mutant Y37F, S459F & L461F was constructed using *SaSS* Y37F as template and primer of *SaSS* S459F & L461F. Mutation reactions were transformed in XL10 gold ultra-competent cells and the mutations were confirmed by automated nucleotide sequencing. Plasmid DNA constructs were further transformed into Rosetta 2 (DE3) cells for protein expression.

4.2.3 Heterologous expression and purification of recombinant enzymes

pET32b harboring *SaSS* mutants were introduced into Rosetta 2 (DE3) chemically competent cells and colonies were grown on LA containing 100 µg/ml ampicillin and 34 µg/ml chloramphenicol and incubated overnight at 37 ºC. A single colony was inoculated in 5 mL LB+Amp+Cam and incubated overnight at 37 ºC at 200 rpm. 1 mL of overnight grown cultures were transferred to 100 mL of TB-amp-cam and incubated at 37 °C at 200 rpm till OD at 600 nm reached 0.8, after which, the cultures were induced with 0.5 mM IPTG and incubated overnight at 16 °C. Cultures were harvested by centrifugation at 4500 × g for 10 minutes at 4 °C, frozen and stored at -80 °C till further use. Cell pellet was re-suspended in lysis buffer (10 mL/g of cell pellet) containing 1 mg/mL of lysozyme, protease inhibitor cocktail and incubated on ice for 30 minutes. Cells were lysed by sonication was carried out using probe sonicator for 5 cycles (pulse on 30 sec, pulse off 30 sec, and amplitude 70 %), and the lysed suspension was centrifuged at 10,000 × g for 20 minutes. Crude lysate was mixed with Ni-NTA slurry and incubated at 4 °C for 1 hr on rocking. It was loaded onto a protein purification column of 1 cm diameter. After collecting the unbound fraction, the column was washed with wash buffer till O.D. at 280 nm achieved zero (4 mL each time). Finally it was eluted into six fractions of 0.5 mL elution buffer each. Purified protein fractions were pooled, and desalted on HiPrep™ 26/10 desalting column in HEPES buffer using ACTA (GE Healthcare). All the fractions were analyzed on a 10 % SDS-PAGE to check for purity.

4.2.4 Enzymatic characterization and product analysis

Enzymatic assays were carried out for all the santalene synthase mutants in HEPES buffer pH 7.4 (25 mM HEPES, 10 % glycerol, 5 mM DTT and 10 mM $MgCl_2$) with100 µM (*E,E*)-FPP in a reaction volume of 500 µL, and incubated for 2 hr at 30 °C. The reaction mixtures were extracted twice with 500 µL n-hexane and the extracts were passed through anhydrous Na_2SO_4. The extracted volume was reduced to ~50 µL with a stream of dry nitrogen. Products were characterized by injecting 1 µL of concentrated sample in GC and GC-MS equipped with a HP-5 capillary column (30 m × 0.32 mm × 0.30 µm, HP-5, J & W Scientific). The column was equilibrated at 70 °C followed by a temperature gradient from 70 °C to 170 °C at 5 °C/min, followed by a second temperature gradient of 15 °C/min from 170 °C to 180 °C and a final hold at 180 ºC for 5 minutes using nitrogen as the carrier gas with a flow rate of 1 mL/min.

Analysis by GCMS was performed under the similar conditions on HP-5MS capillary column (30 m × 0.25 mm × 0.25 μm,HP-5 MS, J & W Scientific) using helium as the carrier gas with a flow rate of 1 mL He/min. All the products formed were verified by co-injection with purified standard samples and comparing with NIST Mass Spectral Library.

4.2.5 Kinetic characterization of santalene synthase mutants

Steady state kinetics of santalene synthase mutants were performed using 4.8 μM (200 μg) of protein with varying substrate concentrations ranging from 0.5 μM to 40 μM (0.5, 1.0, 2.5, 5.0, 7.5, 10.0, 12.5, 15.0, 20.0 and 40.0μM), in the 25 mM HEPES buffer, in a 500 μL scale for 5 minutes at 30 °C and 50 rpm in a water bath shaker. Reactions were quenched by adding 250 μL of saturated Barium hydroxide and 250 μL of 0.1 M Zinc sulphate and vortexing. Further, 1.5 mL of n-hexane was added, vortexed, and these mixtures frozen at -80 °C till further use. The extracts were supplemented with 100 ng of dodecane as an internal standard and the product ratios were analyzed using GC. The calculated amounts of products formed were used for the determination of K_m, V_{max}, K_{cat} and K_{cat}/K_m values.

4.2.6 Large scale enzyme assay

For the characterization of major metabolites from mutant Y539W (products **6** and **7**) and from mutant I422A (products **8** and **10**), large scale enzyme assays were setup using 260 mg (Y539W) and 300 mg (I422A) of purified proteins with 1.6 mM FPP in HEPES buffer, pH 7.4 in a total reaction volume of 12 mL distributed in 16 test tubes.

4.2.6.1 Large scale assay setup of Y539W

260 mg of protein was purified from16 L expression of SaSS mutant Y539W same as discussed in section 4.2.3.

Protein (2 mg/mL) = 8 mL
50mM FPP (1.6mM) = 0.450 mL
Buffer (pH 7.4) = 3.550 mL
Total = **12 mL** (overlaid with 3 mL n-Hexane)

4.2.6.2 Large scale assay setup of I422A

Protein (2.3mg/mL) = 8 mL
50mM FPP (1.6mM) = 0.450 mL

Buffer (pH 7.4) = 3.550 mL

Total = **12 mL** (overlaid with 3 mL n-Hexane)

Reaction mixtures were overlaid with 4 mL of n-hexane and incubated at 30 °C for 24 hr at 50 rpm. After incubation, reaction mixtures were pooled together and extracted twice with n-hexane and analyzed by GC and GCMS.

4.2.7 $AgNO_3$ silica column for purification of metabolites

$AgNO_3$ coated silica gel (230-400 mesh size) was prepared by adding silica gel into methanolic solution of silver nitrate[24] (5 % w/v). Methanolic slurry was dried over rotary evaporator under reduced pressure and then in a hot air oven at 80 °C for 1 hr to get dry $AgNO_3$ coated silica gel. Column was prepared (10 cm height and 1 cm internal diameter) using 5 cm of uncoated silica gel (230-400 mesh) and then 5 cm of silver nitrate coated silica gel. The column was washed with three column volumes of pet ether. Crude assay sample was loaded on top of the column and fractionation was started using pet ether as the solvent system. Slowly ethyl acetate was added drop by drop and fractions were collected. All the fractions were analyzed by GC and GC-MS. All the pure fractions were pooled together and analyzed by GC. The sesquiterpene NMR sample was prepared using $CDCl_3$ solution and H^1, C^{13} and DEPT spectral data was acquired at 500 MHz for the characterization of the purified metabolite.

4.3 Results and Discussion

In the present study, homology based directed mutagenesis was used to replace the selected amino acid residues in the active site of santalene synthase (*SaSS*) to understand sesquiterpene biosynthesis and molecular origins of chemical diversity in *Santalum album*. This study also provided valuable insights into the evolutionary aspects of various terpene synthases even though there is very low level of sequence identity among them.

4.3.1 Homology based structural analysis of *SaSS*

Terpene synthases are a diverse enzyme family responsible for the biosynthesis of complex secondary metabolites. The specificity and activity of terpene synthases are highly dependent on the small number of amino acids present inside or near the active site cleft. Santalene synthase is a moderately promiscuous enzyme which produces a mixture of various carbon skeletons from linear (*E*-β farnesene) to tricyclic (α-santalene) starting from (*E,E*)-FPP as the substrate (as proposed in scheme 4.3.1). Even though terpene synthases vary in sequence similarity, their active sites are highly conserved. To understand the basis of cyclization of various carbon skeletons in the product profile of *SaSS*, homology based model was built using co-ordinates of 5-*epi*-aristolochene synthase mutant (3M00) as a template, based on an overall sequence similarity of 32 % at amino acid level, using Schrodinger.

3D homology model of SaSS was validated using Ramachandran validation tool with default parameters. Ramachandran PROCHEK validation tools suggested that 406 residues (84.2 %) of the modeled *SaSS* were present in the most favored region, whereas 67 residues (13.9 %) were present in the additional allowed region. Only 2 residues (0.4 %) were present in the disallowed region. This result clearly suggests that the model built is valid and can be used for docking studies.

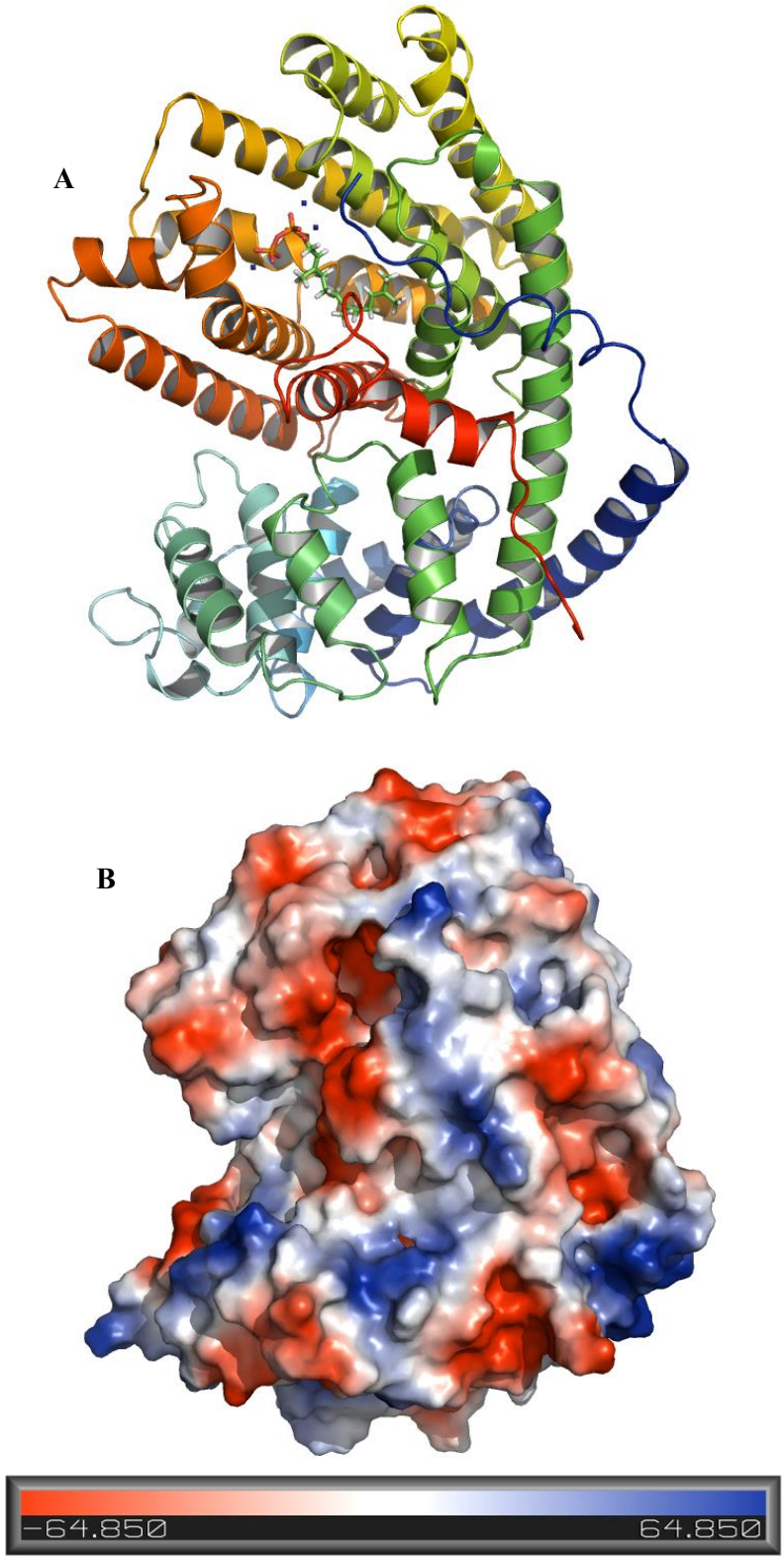


Figure 4.3.1: A) Ribbon diagram of overall structure of predicted santalene synthase 3D structure using 5-*epi*-aristolochene synthase as template, **B)** Surface representation of modeled santalene synthase colour coded by surface electrostatic potential

The *SaSS* modeled structure entirely consists of α-helices connected by small loops and is organized into two domains (NH_2 terminal domain and COOH terminal domain). Active site of santalene synthase was identified by presence of highly conserved DDxxD motif located in the upper edge of the C-terminal helical barrel domain[7,8,22]. This is involved in substrate binding whereas, opposite wall of the active site is composed of highly conserved residues (D/**N**)Dxx(S/**T**)xxx**E** which are involved in divalent metal ion, Mg^{2+} binding[10,21,22]. Active site cavity of santalene synthase is flanked by six α-helices connected with small loops, making the active site cavity 14Å wide and 20Å deep. Bottom of the active site cavity is mainly composed of aromatic and aliphatic hydrophobic residues (such as: W293, F538, F545, Y539, I422, L427) which help in giving shape to the active site pocket for the binding of FPP in correct orientation for cyclization, whereas top of the pocket is mainly composed of polar amino acids (such as: N463, T467, E471, R460, R474, K479, D321, D322) and thought to be involved in binding the phosphate end of the substrate. Despite the low level of sequence similarity, majority of the active site residues are common in *SaSS* model and 3M00 structure whose functions are assigned[8].

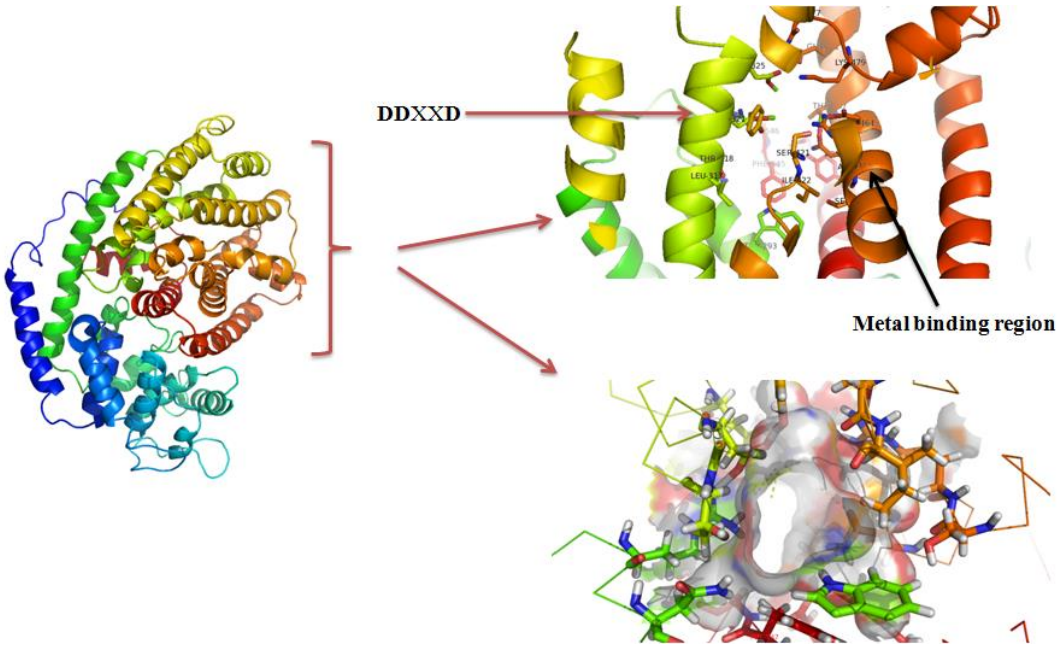


Figure 4.3.2: Active site pocket analysis of homology structure of *SaSS*, **A)** *SaSS* model structure, **B)** Six helices surrounding the active site pocket, **C)** representing active site volume and shape

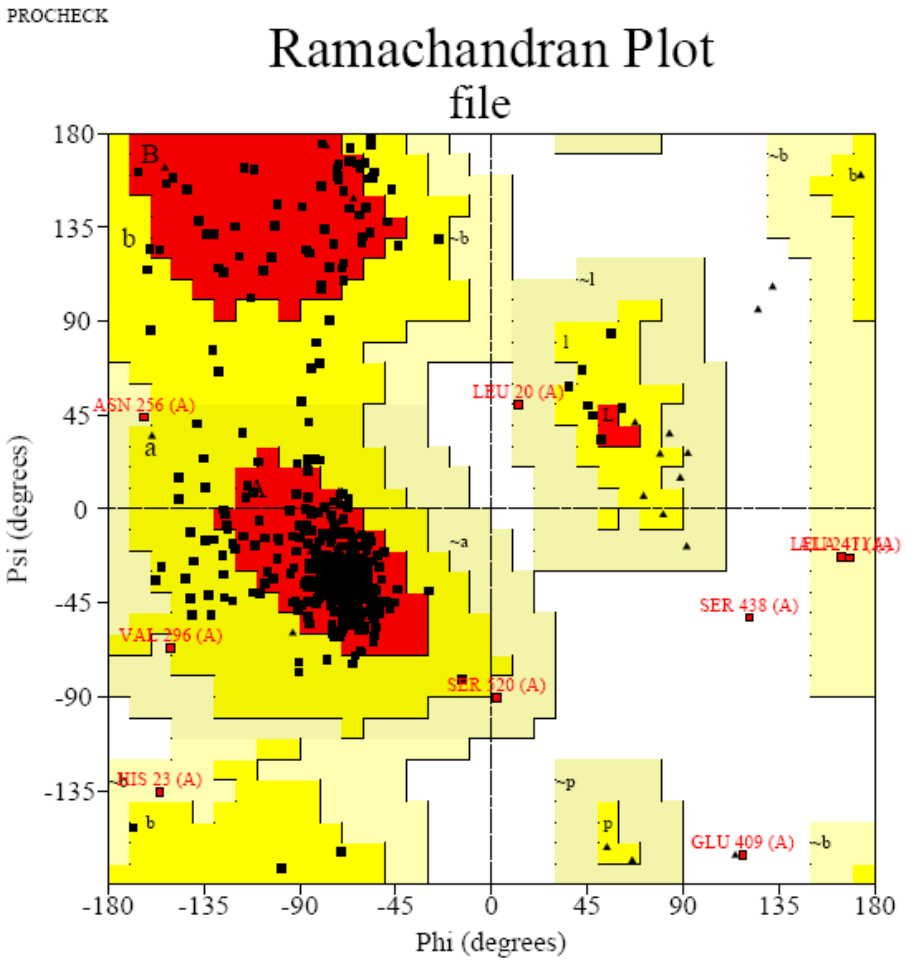
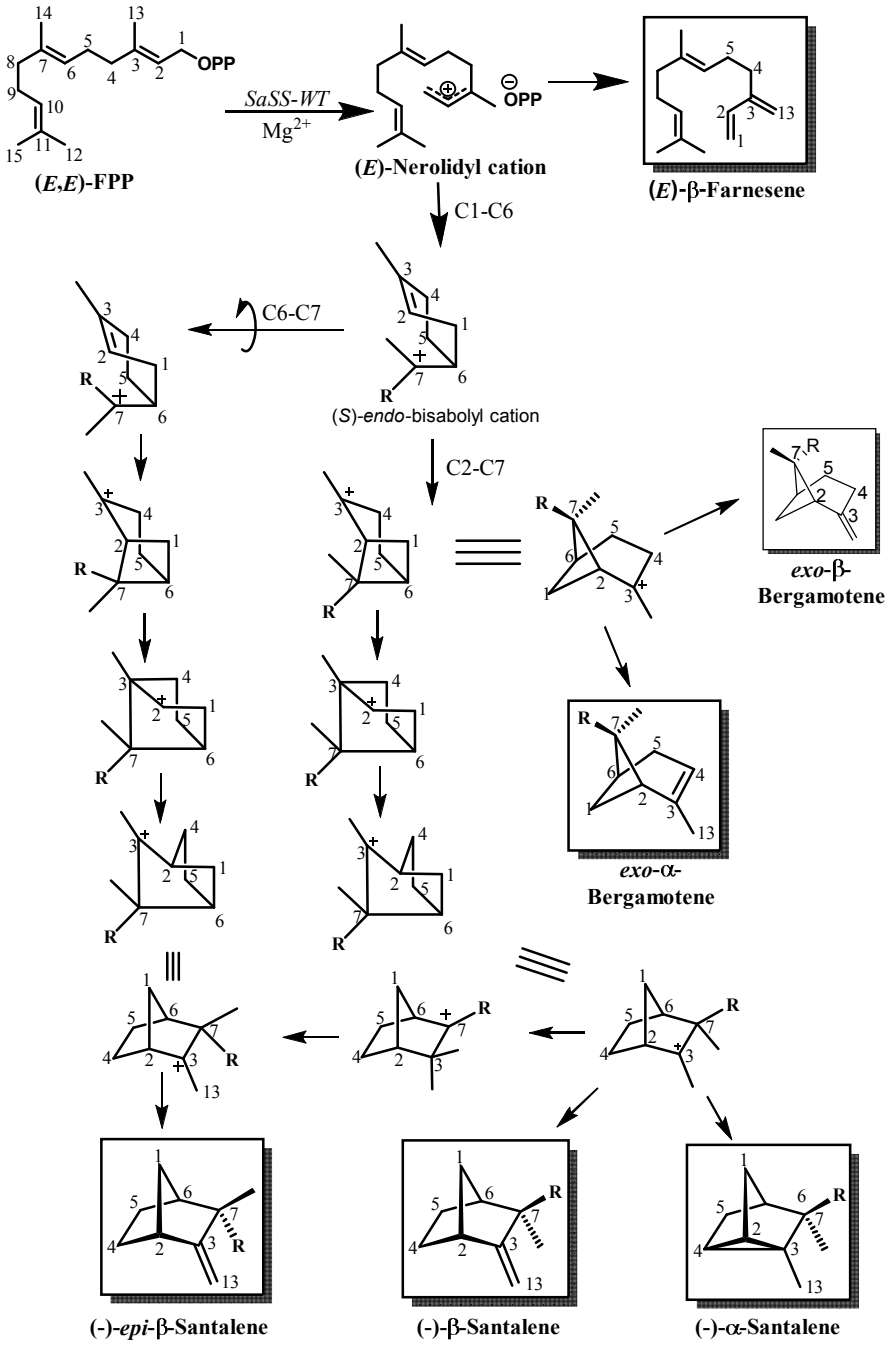


Figure 4.3.3: Ramachandran PROCHECK validation for homology model of *SaSS* built using 5-*epi*-aristolochene synthase as template

4.3.2 Proposed mechanism of santalene biosynthesis

Santalene synthase, upon incubation with (*E,E*)-FPP, in the presence of divalent metal ion (Mg^{2+}), produced a mixture of six sesquiterpenes with varied carbon skeletons (Scheme 4.3.1). An initial step in the cyclization reactions resulting in biosynthesis of santalenes involves Mg^{2+} assisted ionization of the allylic diphosphate substrate (*E,E*)-FPP at C1 to produce the charged diphosphate anion and the farnesyl cation that subsequently isomerizes to its tertiary counterpart. (*E*)-β-farnesene is produced from the (*E*)-nerolidyl carbocation by direct deprotonation. The formed carbocation undergo C1-C6 cyclization reaction to form (*S*)-endo-bisabolyl cation. The endo bisabolyl carbocation undergoes C2-C7 ring closure to produce bicyclic bergamotyl carbocation, which loses the proton from C4 and forms exo-α-bergamotene. Majority of bergamotyl carbocation formed undergoes several Wagner-

Meerwein rearrangements to generate santalyl carbocation. From the santalyl carbocation α- and β-santalenes are produced by deprotonation from C4 and C13, respectively.



Scheme 4.3.1: Proposed reaction mechanism for the biosynthesis of santalenes and bergamotenes from (*E,E*)-FPP upon incubation with *SaSS* in presence of Mg^{+2}.

4.3.3 Docking of intermediates involved in santalene biosynthesis

In order to determine the contribution of particular residues involved directly or indirectly in binding and catalysis of santalene synthase, docking studies were performed using flexible ligand conformations. All the intermediates involved in santalene biosynthesis (Scheme 4.3.1) were docked and structures were minimized using Schrodinger. 22 amino acid residues surrounding the active site cavity were selected including all common residues present in 5-*epi* aristolochene synthase, for site-directed mutagenesis to understand the biosynthesis of santalenes and bergamotenes. These surrounding amino acids were thought to be involved in maintaining electronic electrostatic gradient, conserving the active site frame work, correct folding and providing complementarity to the carbocation for cyclization in active site pocket of *SaSS*. The aspartate and arginine rich motifs highly conserved in all sesquiterpene synthases were not considered, because studies have already proved that such alterations would result in significant loss of activity[21].

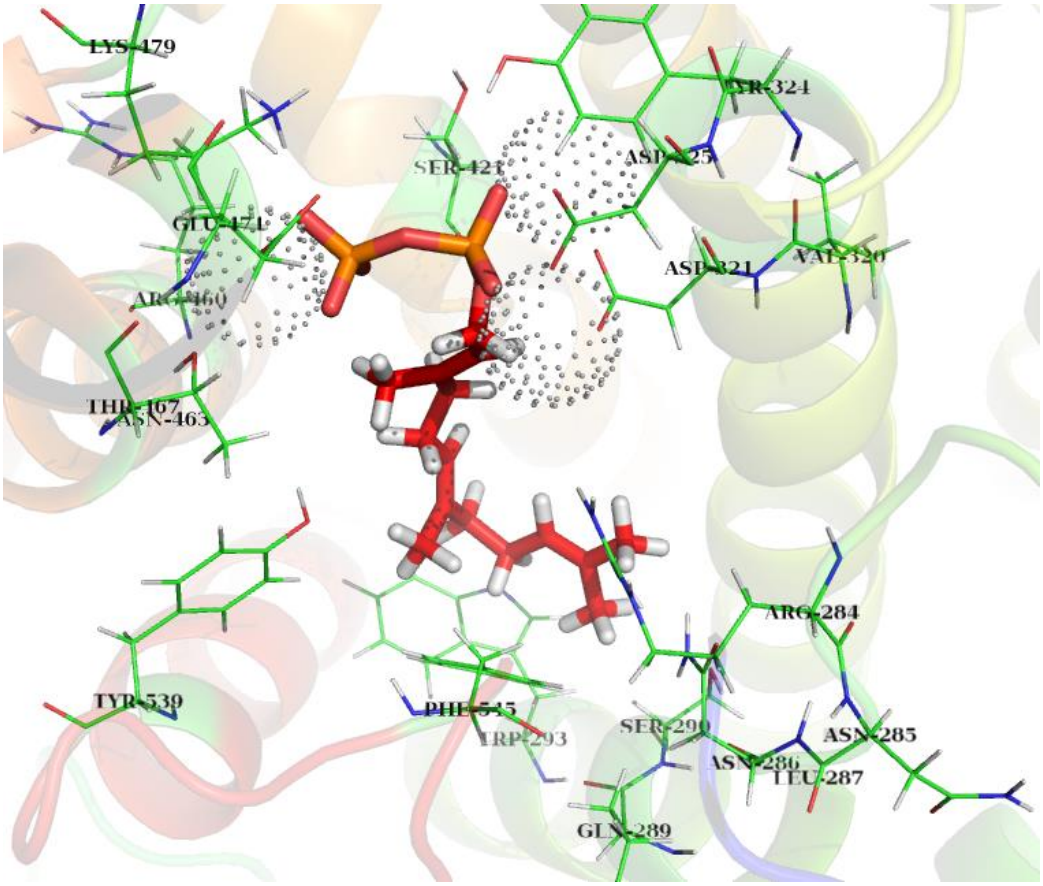


Figure 4.3.4: Amino acid interaction in active site pocket of *SaSS* homology model with (*E,E*)-FPP

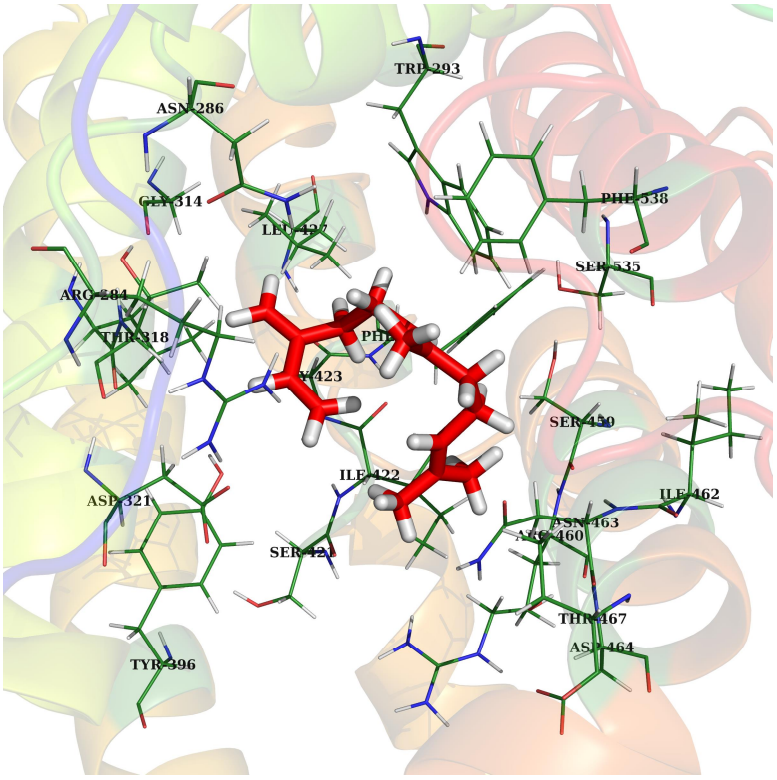


Figure 4.3.5: Amino acid interaction with nerolidyl carbocation docked with *SaSS* homology model

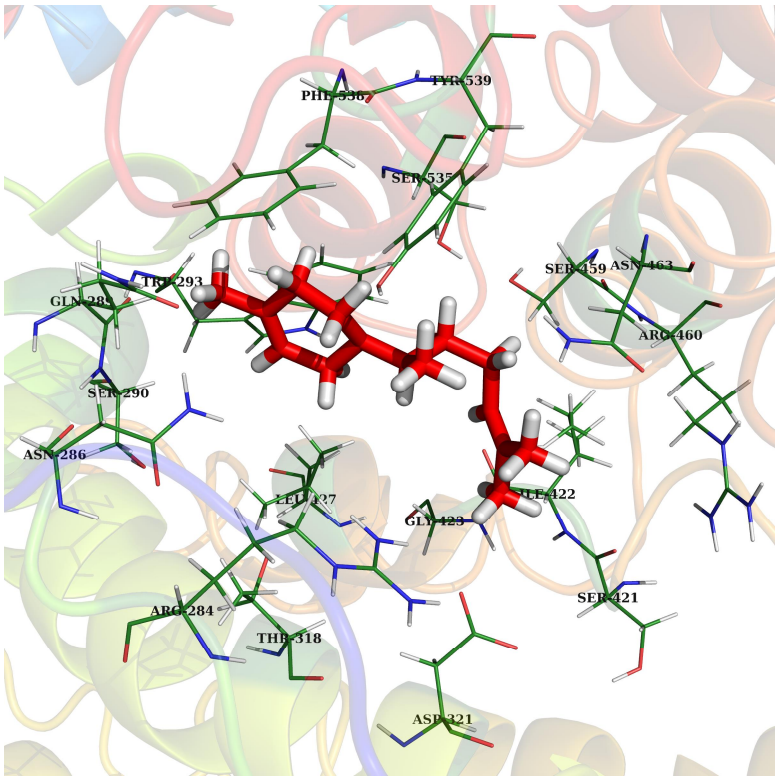


Figure 4.3.6: Amino acid interaction with bisabolyl carbocation docked with *SaSS* homology model

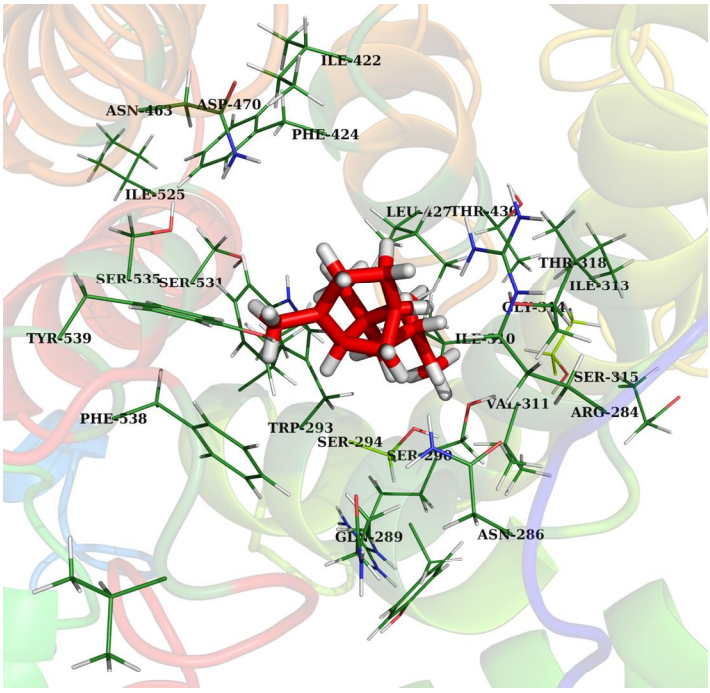


Figure 4.3.7: Amino acid interaction with bergamotyl carbocation docked with *SaSS* homology model

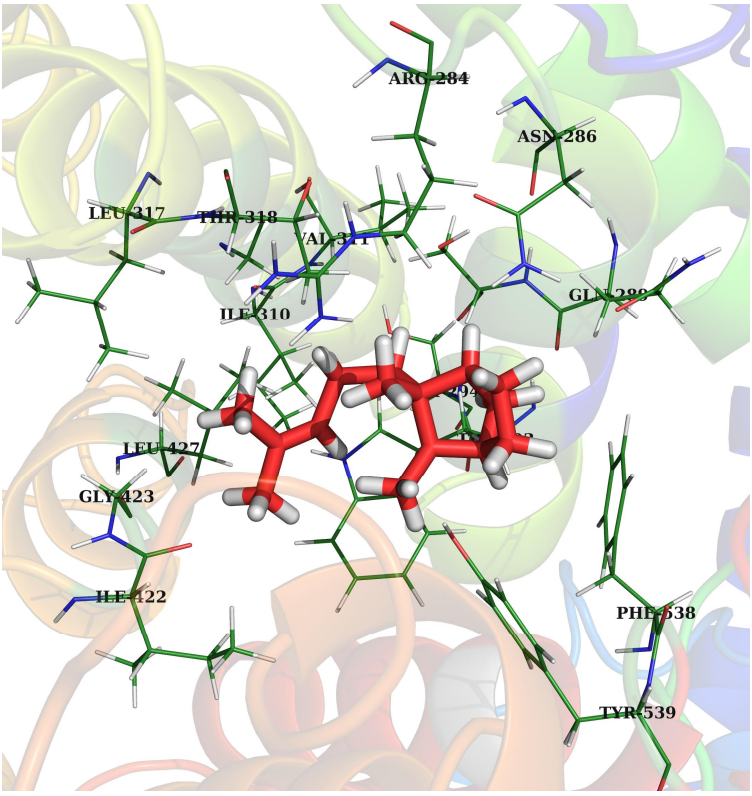


Figure 4.3.8: Amino acid interaction with santalyl carbocation docked with *SaSS* homology model

4.3.4 Residues involved in catalytic triad/dyad

Stark *et al.*[8] hypothesized that Asp^{444}, Tyr^{520}, and Asp^{525} form a catalytic triad in the active site of 5-epi-aristolochene synthase and help in the isomerization of germacrene A intermediate. Sequence comparison of *SaSS* and 5-EAT showed that one of these two aspartate residues is conserved (D543), whereas second aspartate is replaced by asparagine (N463). To verify the essential role of these residues in santalene synthase active site pocket, site directed mutagenesis of the corresponding residues (Asn^{463}, Tyr^{539}, and D^{543}) in *SaSS* was performed as discussed in the material and methods section.

4.3.4.1 Tyrosine 539 (Y539)

Tyrosine 539 (corresponding to Y520 in 5-*epi*-aristolochene synthase and Y528 in maize TPS4) is a highly conserved residue among terpene synthases and found to be involved in protonation / isomerization of carbocation. To investigate the role of Y539 in santalene synthase (*SaSS*) active site pocket, site directed mutagenesis was carried out. Removal of hydroxyl group from tyrosine (Y539F) resulted in exclusive formation of exo-α-bergamotene (84 %) and exo-β-bergamotene (11 %), along with small percentage of santalenes. Similar profile was observed with the mutant Y539S (Table 4.3.2). Incorporation of larger size side chain by replacing tyrosine 539 with tryptophan (Y539W) led to the formation of endo-α-bergamotene due to the alteration of the geometry of active site pocket. This change in secondary structure resulted in the formation of isomeric mixture of (*R*)- and (*S*)-bisabolyl carbocations by rotation of the C6-C7 bond before closure of C2-C7 (Scheme 4.3.3). (*R*)-bisabolyl carbocation undergoes C2-C7 closure, and thereafter, loses proton from C4 to form endo-α-bergamotene (21 %) (which is absent in the products formed by native form of the enzyme), along with α-santalene (6 %), exo-α-bergamotene (28 %), *epi*-β-santalene (1 %), β-farnesene (8 %) and exo-β-bergamotene (20 %) from the parallel pathway through counterpart (*S*)-bisabolyl carbocation. Mutant Y539L also resulted in the formation of exo-α-bergamotene (68 %) as the major product (Figure 4.3.9). These results clearly indicate that both the hydroxy group and aromatic ring of tyrosine 539 are required for the biosynthesis of α- and β-santalenes. The selective production of exo-α-bergamotene provides evidence for an exo-α-bergamotene intermediate in the cyclization reaction for santalene formation.

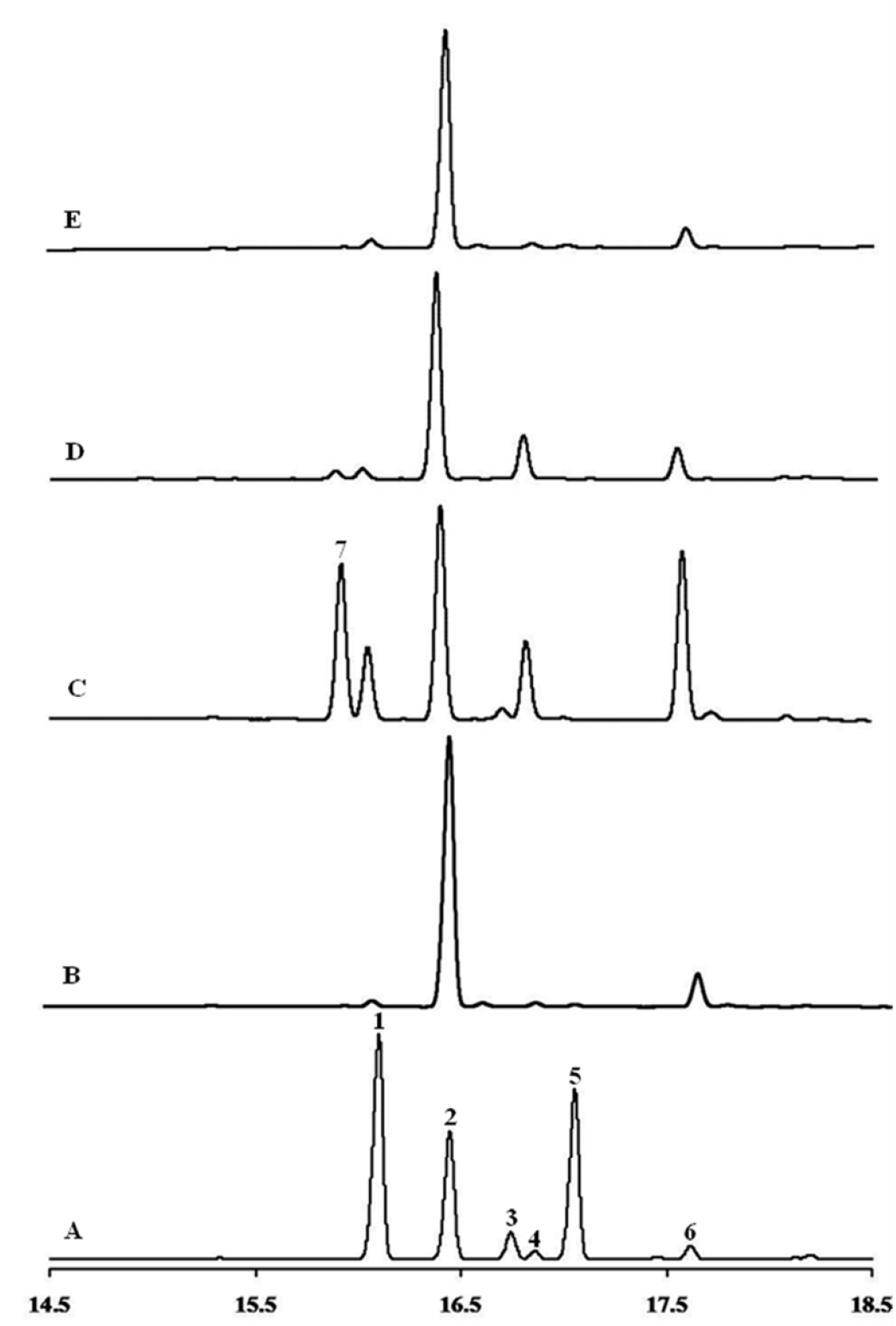


Figure 4.3.9: GC profile of products from incubation of (*E,E*)-FPP with mutants of Y539, **A)** *SaSS* wild type, α-santalene (**1**), exo-α-bergamotene (**2**), *epi*-β-santalene (**3**), (*E*)-β-farnesene (**4**), β-santalene (**5**), exo-β-bergamotene (**6**), endo-α-bergamotene, (**7**), **B)** *SaSS* mutant **Y539F**, **C)** *SaSS* mutant **Y539W**, **D)** *SaSS* mutant **Y539L**, **E)** *SaSS* mutant **Y539S**

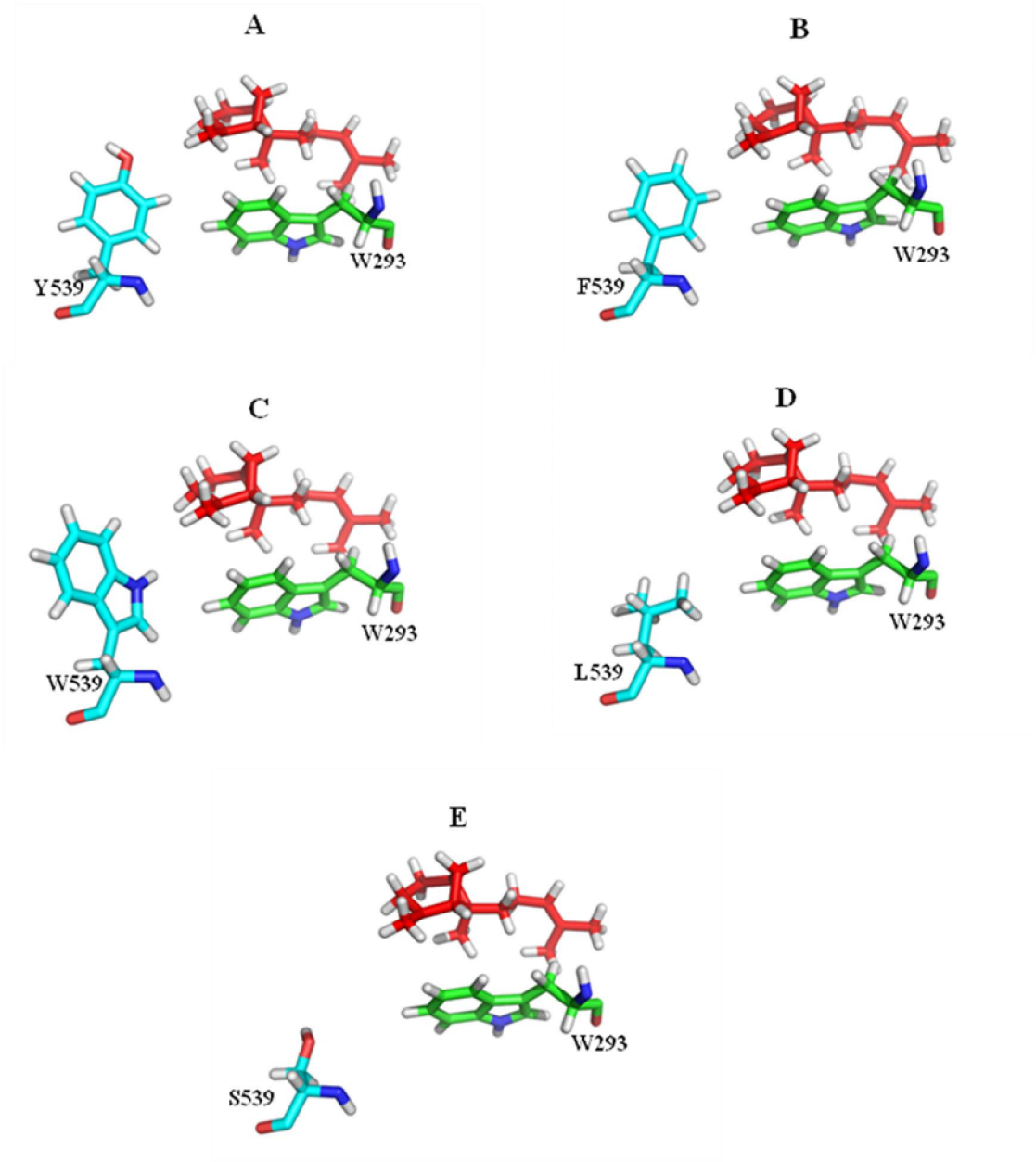


Figure 4.3.10: Docking of bergamotyl carbocation with homology model of *SaSS* and all the Y539 mutants, **A)** Interaction with Y539, **B)** Interaction with *SaSS* mutant **F539**, **C)** Interaction with *SaSS* mutant **W539**, **D)** Interaction with *SaSS* mutant **L539**, **E)** Interaction with *SaSS* mutant **S539**, Residue shown in cyan colour is site of mutation and residue shown in green colour is involved in stabilizing the carbocation

4.3.4.2 Asparagine 463 (N463)

N463 is the first residue of the highly conserved metal binding domain with the consensus of (D/**N**)Dxx(S/T)xxxE and involved in Mg^{2+} binding. The introduction of mutation N463D exhibits the selectivity for producing α-santalene and exo-α-bergamotene in equal proportions by minimizing the formation of β-santalene (Figure

4.3.11). Thus, the mutant N463D seems incapable of deprotonation from C13 required for the formation of β-santalene. Whereas, the alteration of asparagine 463 to larger side chain isoleucine resulted in complete loss of activity. This loss of activity may be attributed to the perturbation of the active site pocket by incorporating a large hydrophobic side chain, as well as the disturbance caused in the metal ion interaction which is required for diphosphate removal from FPP for cyclization (Figure 4.3.12).

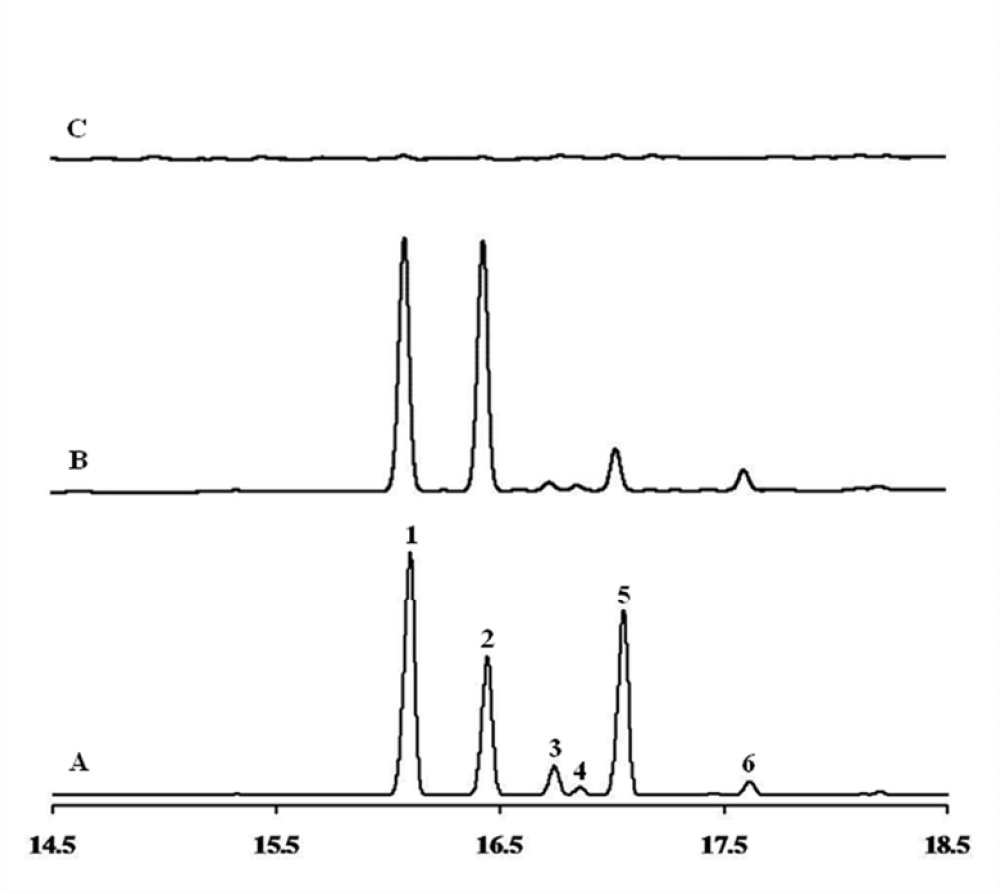


Figure 4.3.11: GC profile of products from incubation of (*E,E*)-FPP with mutants of N463, **A)** *SaSS* wild type, α-santalene (**1**), exo-α-bergamotene (**2**), *epi*-β-santalene (**3**), (*E*)-β-farnesene (**4**), β-santalene (**5**), exo-β-bergamotene (**6**), **B)** *SaSS* mutant **N463D**, **C)** *SaSS* mutant **N463I**

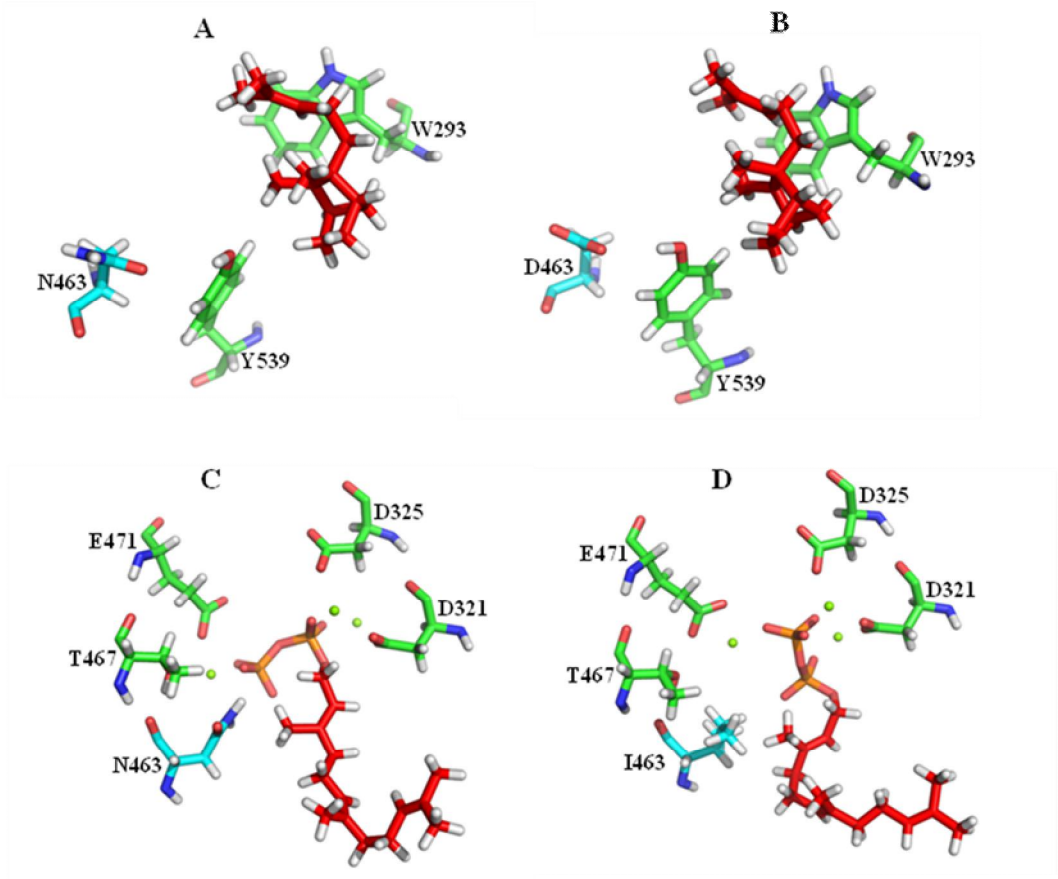


Figure 4.3.12: Interaction of carbocation with N463 and its mutant, **A)** Interaction of **N463** with santalyl carbocation, **B)** Interaction of **D463** with santalyl carbocation (mutant N463D), **C)** Interaction of **N463** with (*E,E*)-FPP, **D)** Interaction of **I463** (mutant N463I) with (*E,E*)-FPP, Residue shown in cyan colour is site of mutation and residue shown in green colour is for stabilizing the carbocation

4.3.4.3 Aspartate 543 (D543)

Aspartate 543 is also a part of the catalytic triad reported in several terpene synthases[8,25]. The mutation D543N resulted in the formation of exo-α-bergamotene (83 %) and exo-β-bergamotene (14 %) as a major enzymatic product by minimizing the santalene formation. This result clearly suggests that D543 plays a crucial role in rearrangement of bergamotyl carbocation to santalyl carbocation (Scheme 4.3.1). However, the replacement of aspartate 543 to leucine leads to complete loss of activity. This loss of activity may be due to disturbance of active site pocket resulting from incorporation of a large hydrophobic side chain.

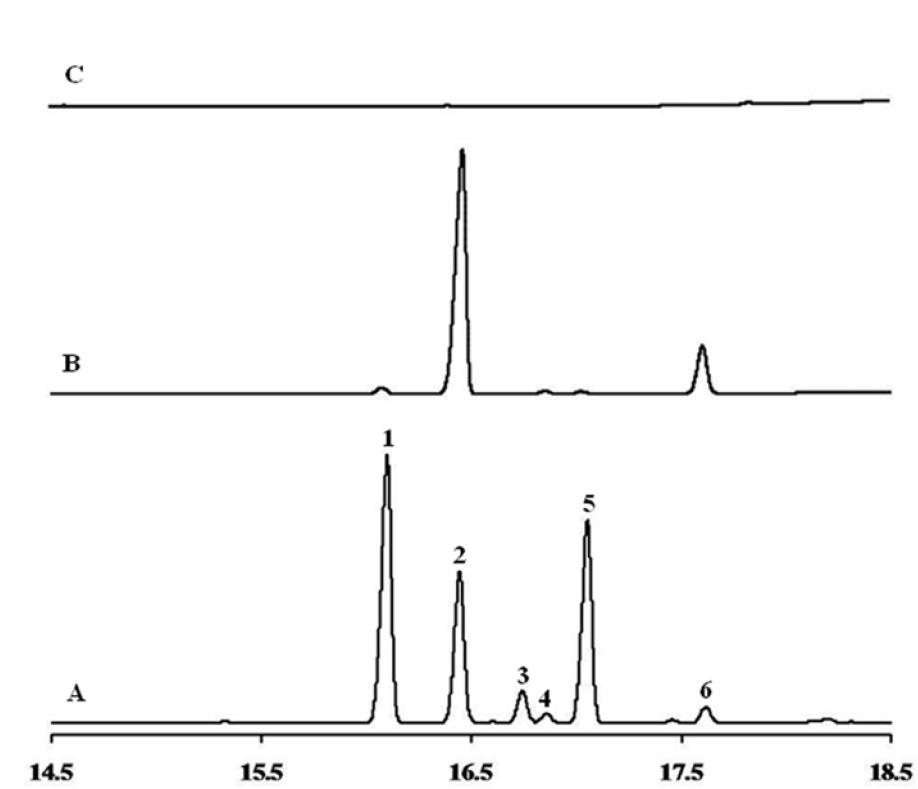


Figure 4.3.13: GC profile of products from incubation of (*E,E*)-FPP with mutants of D543, **A)** SaSS wild type, α-santalene (**1**), exo-α-bergamotene (**2**), *epi*-β-santalene (**3**), (*E*)-β-farnesene (**4**), β-santalene (**5**), exo-β-bergamotene (**6**), **B)** *SaSS* mutant **D543N**, **C)** *SaSS* mutant **D543L**

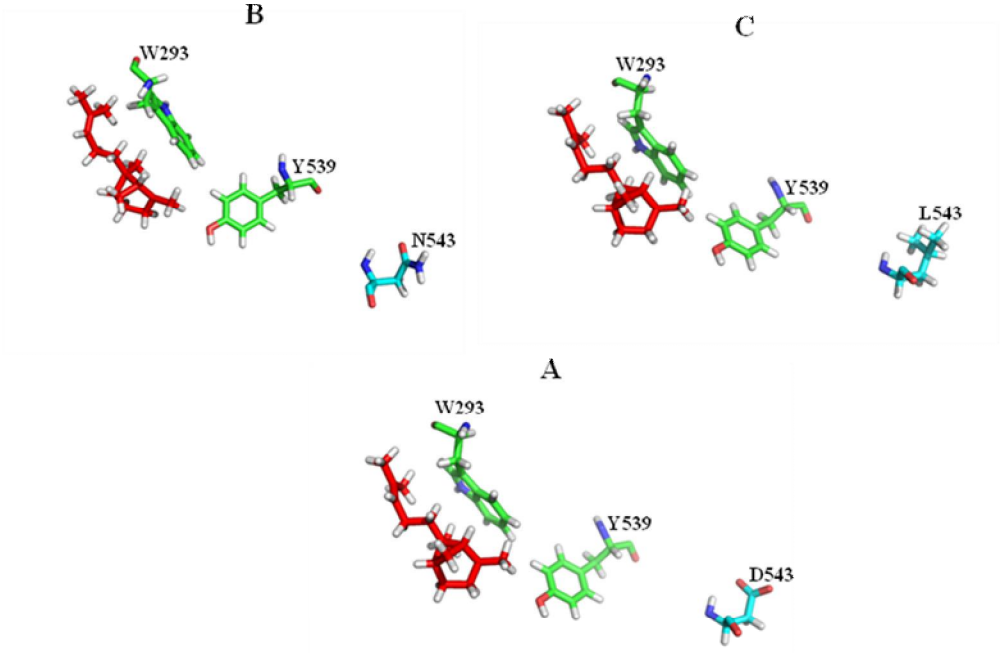


Figure 4.3.14: Interaction of bergamotyl carbocation with D543 and its mutant, **A)** Interaction of **D543**, **B)** Interaction of **N543** (mutant D543N), **C)** Interaction of **L543**

(mutant D543L), Residue shown in cyan colour is site of mutation and residue shown in green colour is for stabilizing the carbocation

4.3.4.4 Large scale expression and purification of *SaSS* mutant Y539W for the characterization of metabolites 6 and 7

The major metabolites of mutant Y539W were characterized by setting up a large-scale enzyme assay as discussed in materials and method section and the products were extracted using n-hexane. 51 mg of crude extract was obtained which was further subjected to $AgNO_3$ coated silica gel column chromatography for the purification of major metabolites. Metabolites **6** and **7** were purified from the column with a purity of ≥95 % (Figure 4.3.15) and spectral data (^{1}H, ^{13}C, and DEPT NMR) were acquired. Products were confirmed by comparing NMR data with the reported NMR values of endo-α-bergamotene[26] and exo-β-bergamotene[27].

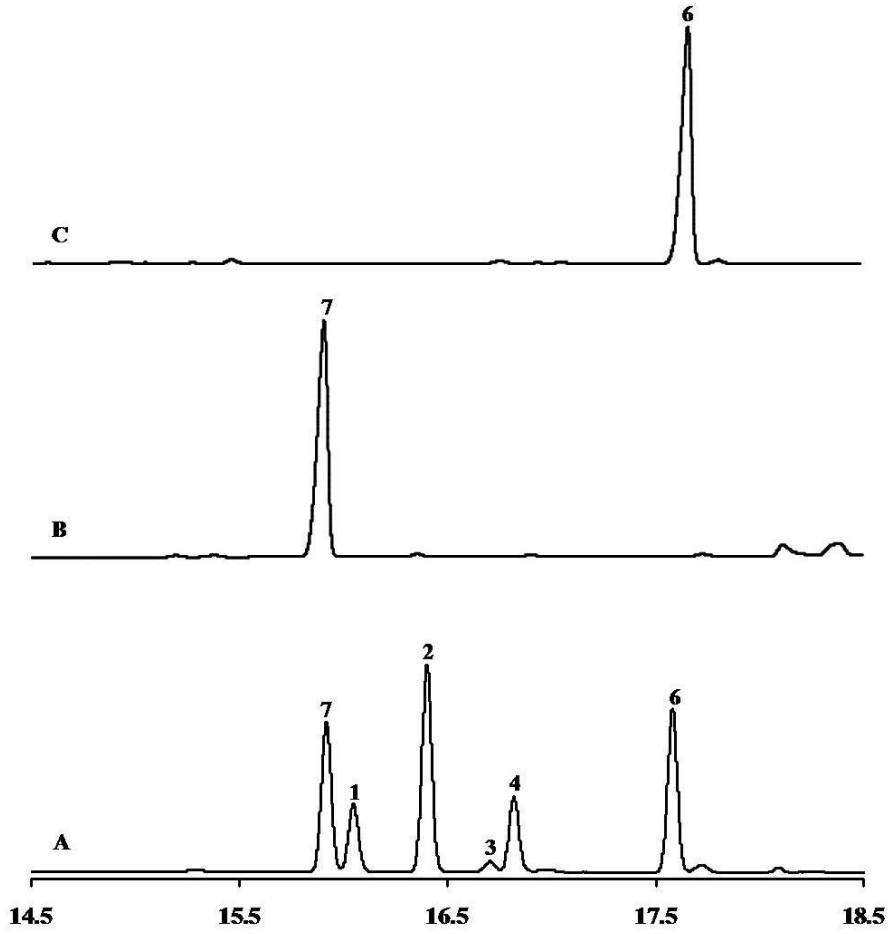


Figure 4.3.15: Purification of metabolite **6** and **7** from large scale assay mixture of *SaSS* mutant Y539W, **A)** GC profile of *SaSS* mutant **Y539W** with (*E,E*)-FPP, **B)** Purified endo-α-bergamotene (**7**), **C)** Purified exo-β-bergamotene (**6**)

These results clearly indicate that *SaSS* requires a dyad of Y539 and D543 for the rearrangement of bergamotyl carbocation into santalyl carbocation for the formation of santalenes. Although the precise function of these amino acid residues in active site is not known, they clearly play a crucial role in stabilizing the carbocation in the later step of santalene synthase catalysis.

4.3.4.5 Threonine 467 and Glutamic acid 471 (T467 and E471)

Threonine 467 (T467) and glutamic acid 471 (E471) are two important residues, along with asparagine (N463), present in the metal binding domain with a consensus of (D/**N**)Dxx(S/**T**)xxx**E** and involved in chelating Mg^{2+} ion (Figure 4.3.17). These residues are highly conserved in terpene synthases and their role has been demonstrated by alanine scanning. Alanine substitution of these metal binding residues reduces catalytic efficiency[28,29]. The mutant T467A resulted in exclusive formation of exo-α-bergamotene (70-75 %), exo-β-bergamotene (19-25 %) and a very small percentage of α-santalene (1-2 %). Mutant E471Q also resulted in a similar profile at the basal level (Figure 4.3.16).

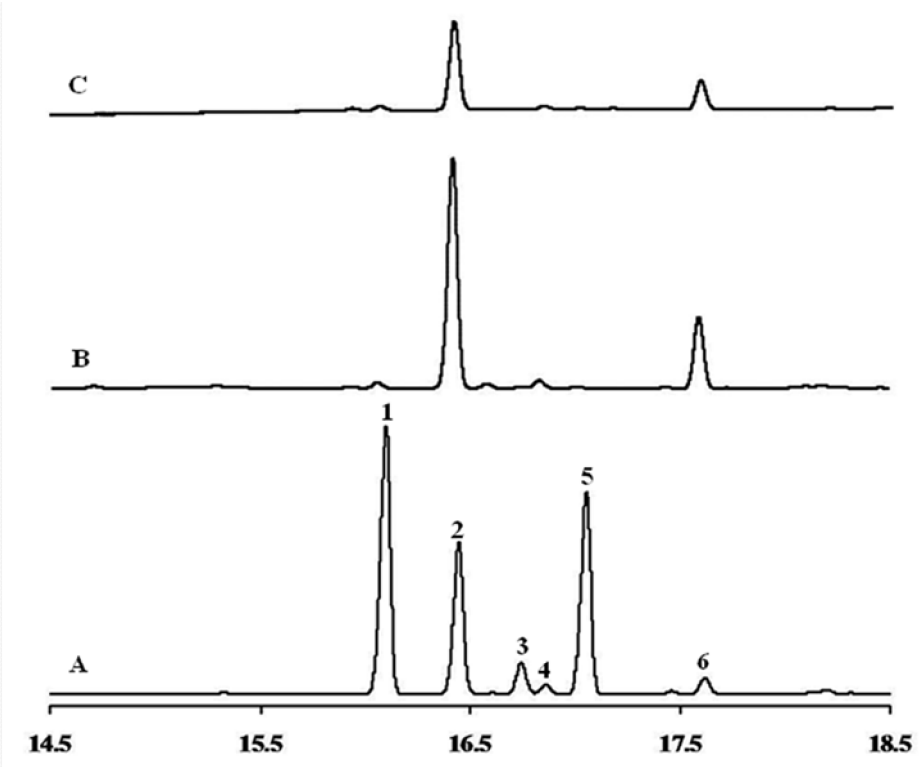


Figure 4.3.16: GC profile of mutants of T467 and E471 with (*E,E*)-FPP, **A)** *SaSS* wild type, α-santalene (**1**), exo-α-bergamotene (**2**), epi-β-santalene (**3**), (*E*)-β-

farnesene (**4**), β-santalene (**5**), exo-β-bergamotene (**6**), **B)** *SaSS* mutant **T467A**, **C)** *SaSS* mutant **E471Q**

As it is reported that these residues involve in metal ion binding and helps in removal of diphosphate moiety for cyclization reaction, mutation to these residues does not completely block removal of the diphosphate moiety, but it minimizes the santalene formation.

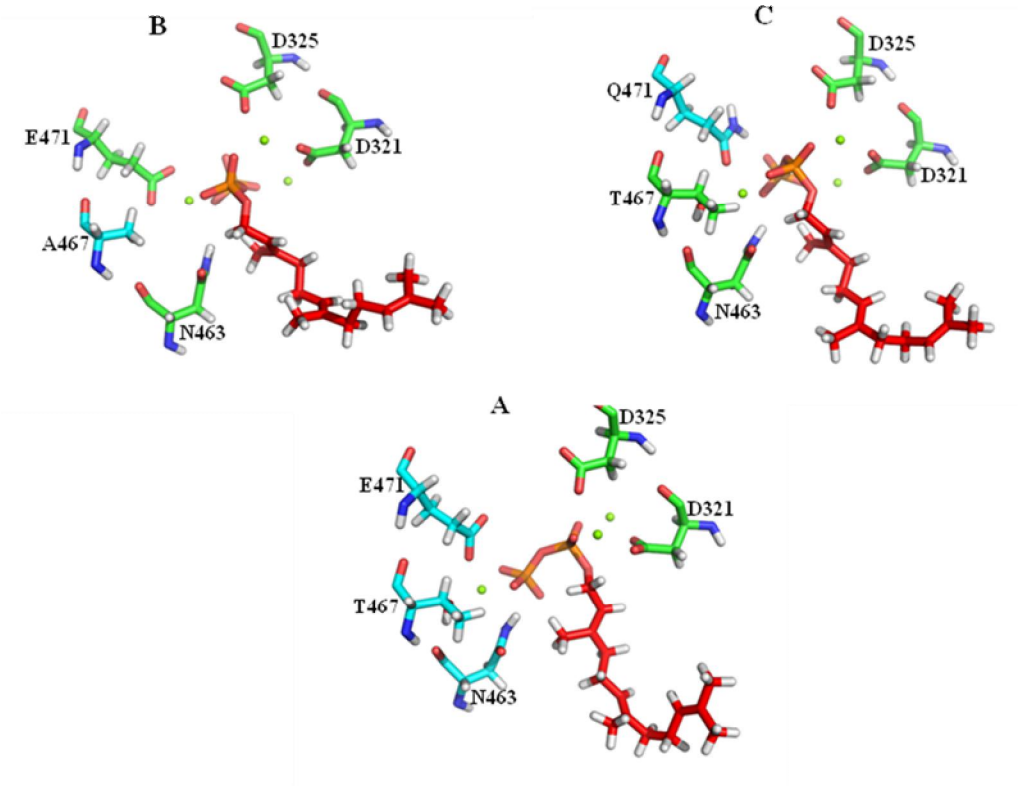


Figure 4.3.17: Interaction of FPP with residues involved in metal binding, **A)** Interaction of **N463**, **T467**, **E471**, **B)** Interaction of **A467** (mutant T467A), **C)** Interaction of **Q471** (mutant E471Q), Residue shown in cyan colour are site of mutation

4.3.5 Residues involved in santalene and bergamotene carbocation stabilization

The mechanism proposed for santalenes and bergamotene biosynthesis in scheme 4.3.1 is highly dependent on the hydrophobic interactions and ionic distribution inside the active site pocket aiding in the formation of santalyl carbocation through a complex rearrangement of nerolidyl carbocation (Scheme 4.3.1), which then undergoes deprotonation from C4 and C13 to convert into α- and β-santalenes, respectively. These reactions are governed by co-ordinated interactions of amino acids present in carbocation vicinity. Little changes in the active site pocket may alter

the free space and electronic distribution which changes the orientation and rearrangement of carbocation cascades.

4.3.5.1 Tryptophan 293 (W293)

Several experimental studies have provided support that the presence of aromatic amino acids in active site pocket appears well placed to provide shape to the carbocation cascade and help in stabilizing the carbocation by π-π interactions. Stark *et al.* postulated the role of W273 (corresponding residue in 5-*epi*-aristolochene synthase) to be involved in stabilizing the carbocation and also performing final deprotonation from the germacryl carbocation for the formation of the final product. Tryptophan 293 is present in the active site pocket and is highly conserved in terpene synthases. Replacement of tryptophan to phenylalanine (W293F) resulted in the formation of α-exo-bergamotene (62 %) as a major product (Figure 4.3.18). This result clearly indicates that W293 is not involved in the final deprotonation for formation of the final products, but it may be involved in stabilizing the bergamotyl carbocation for the formation of α- and β-santalenes. Docking results also suggested that W293 makes π-stacking and helps in stabilizing the bergamotyl carbocation (Figure 4.3.19).

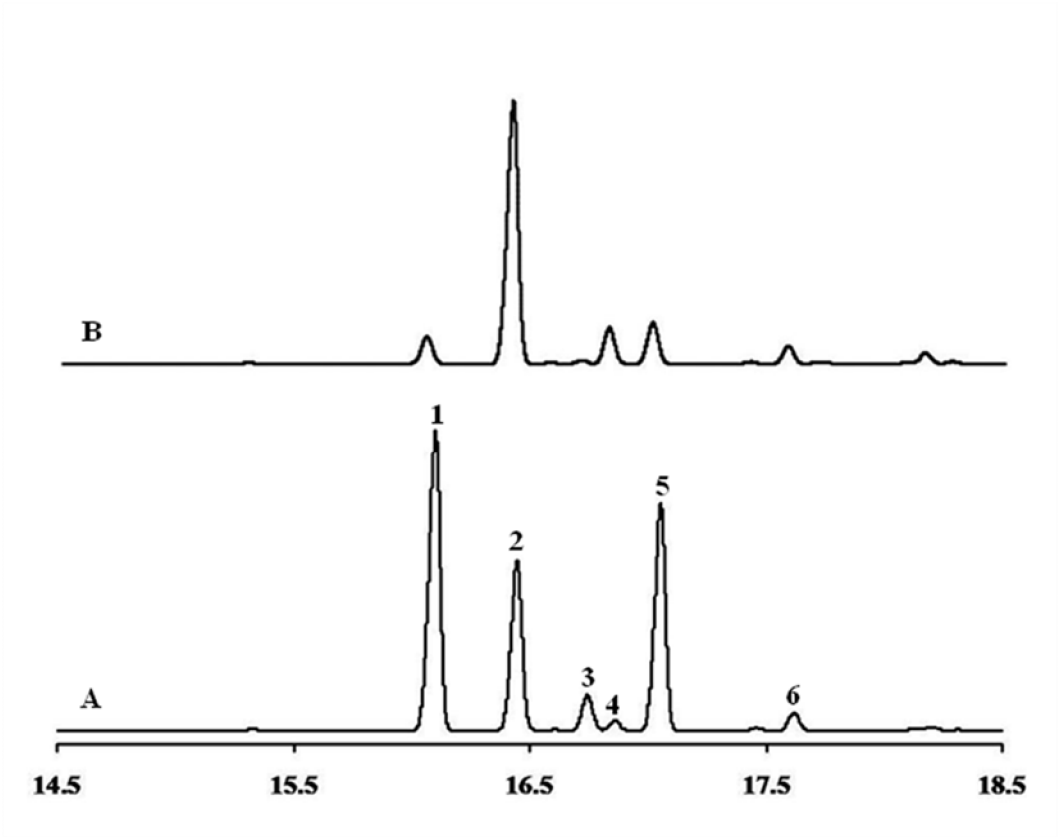


Figure 4.3.18: GC profile of mutant of W293 with (*E,E*)-FPP, **A)** *SaSS* wild type, α-santalene (**1**), exo-α-bergamotene (**2**), *epi*-β-santalene (**3**), (*E*)-β-farnesene (**4**), β-santalene (**5**), exo-β-bergamotene (**6**), **B)** *SaSS* mutant **W293F**

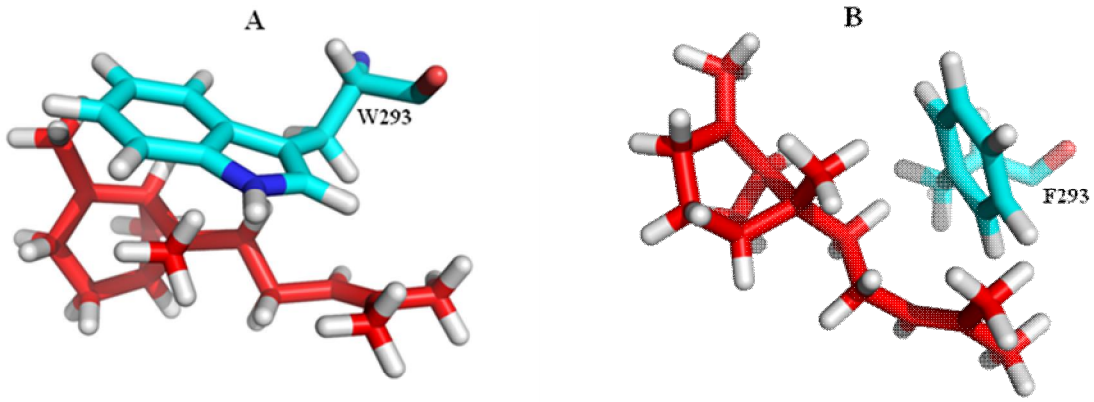


Figure 4.3.19: Interaction of bergamotyl carbocation with W293 and mutant W293F **A)** Tryptophan **(W293)** interaction with bergamotyl carbocation, **B)** Mutant **F293** interaction with bergamotyl carbocation, Residue shown in cyan colour is site of mutation

4.3.5.2 Threonine 318 (T318)

T318 is present in the active site and found to interact with bergamotyl carbocation as well as santalyl carbocation in docking studies (Figure 4.3.21). Mutational analysis of T318 (T318A) resulted in exclusive formation of exo-α-bergamotene (80 %) by minimizing the formation of both α- and β-santalenes (Figure 4.3.20). This result suggests that the well packed active site with carbocation stabilizing functionalities are mere requirements for the conversion of bergamotyl carbocation to santalyl carbocation, leading to the formation of α- and β-santalenes.

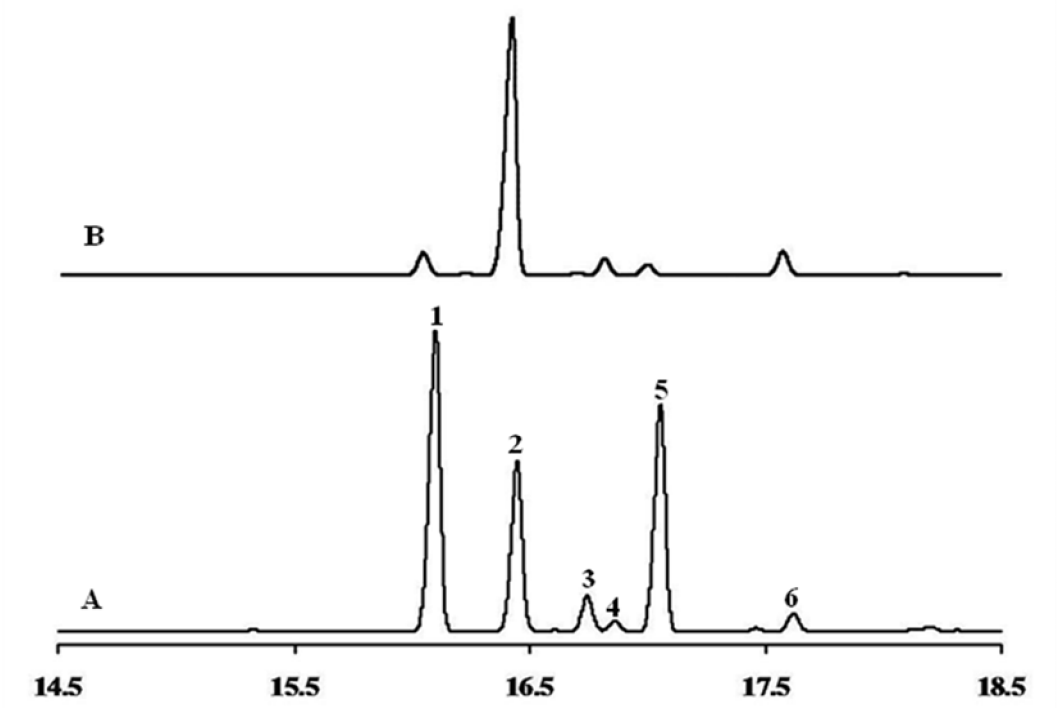


Figure 4.3.20: GC profile of mutant of T318 with (*E,E*)-FPP, **A)** *SaSS* wild type, α-santalene (**1**), exo-α-bergamotene (**2**), *epi*-β-santalene (**3**), (*E*)-β-farnesene (**4**), β-santalene (**5**), exo-β-bergamotene (**6**), **B)** *SaSS* mutant **A318**

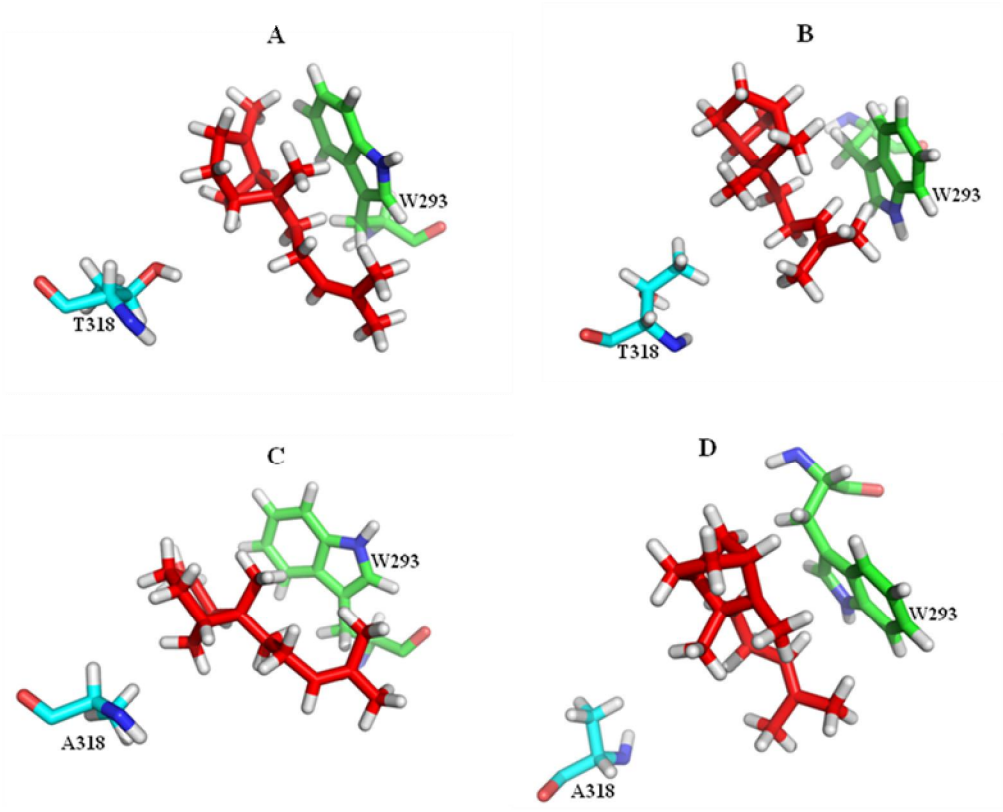


Figure 4.3.5.21: Interaction of carbocation with residue T318 (Cyan color) and W293 (Green colour) for stabilizing the carbocation, **A)** T318 interaction with bergamotyl carbocation, **B)** T318 interaction with santalyl carbocation, **C)** Mutant A318 interaction with bergamotyl carbocation, **D)** Mutant A318 interaction with santalyl carbocation

4.3.5.3 Arginine 474 (R474)

Active site analysis revealed that Arginine 474 is positioned at the opening of active site pocket and might be stabilizing the carbocation as well as may offset the diphosphate released after dephosphorylation of farnesyl diphosphate for the cyclization. The mutation of R474M/L resulted in exclusive formation of exo-α-bergamotene (69-71 %) and exo-β-bergamotene (24-26 %) as shown in figure 4.3.22. These results clearly indicate that R474 also plays a crucial role in stabilizing the carbocation cascade leading to the formation of α- and β-santalenes.

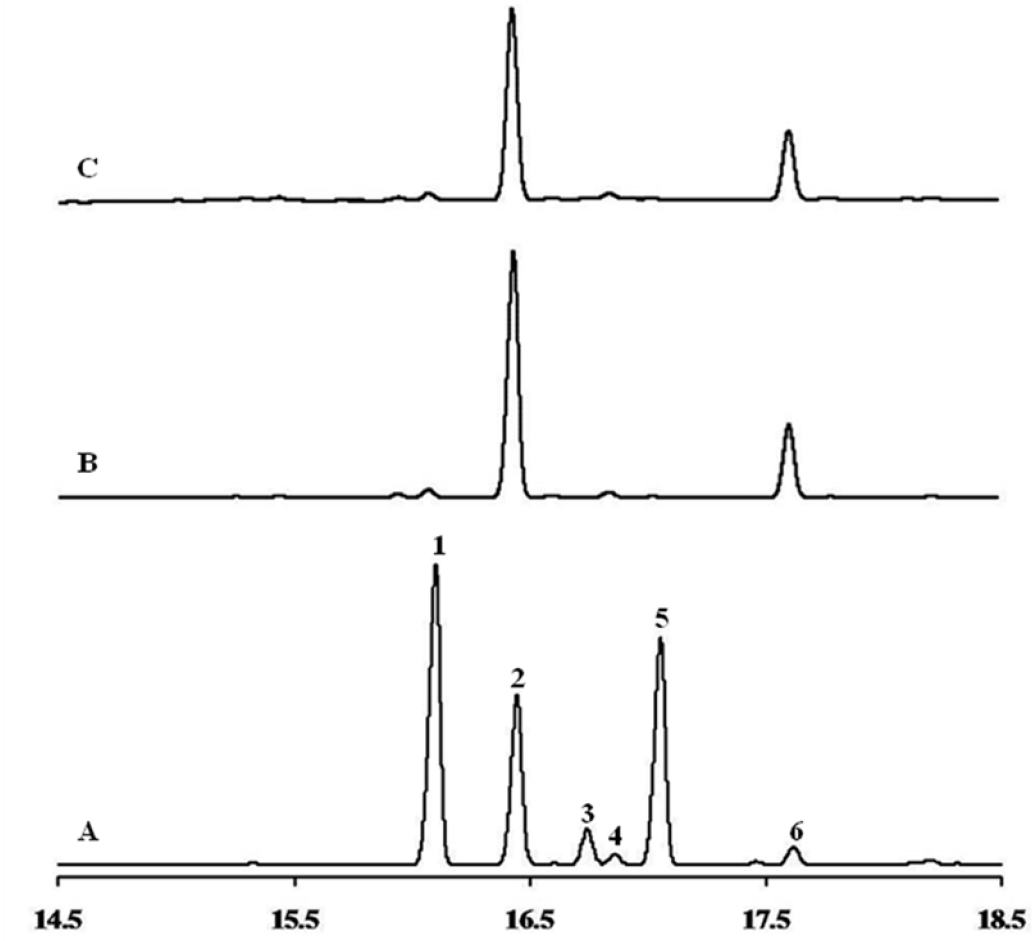


Figure 4.3.22: GC profile of mutants of R474 with (*E,E*)-FPP, **A)** *SaSS* wild type, α-santalene (**1**), exo-α-bergamotene (**2**), *epi*-β-santalene (**3**), (*E*)-β-farnesene (**4**), β-santalene (**5**), exo-β-bergamotene (**6**), **B)** *SaSS* mutant **R474M**, **C)** *SaSS* mutant **R474L**

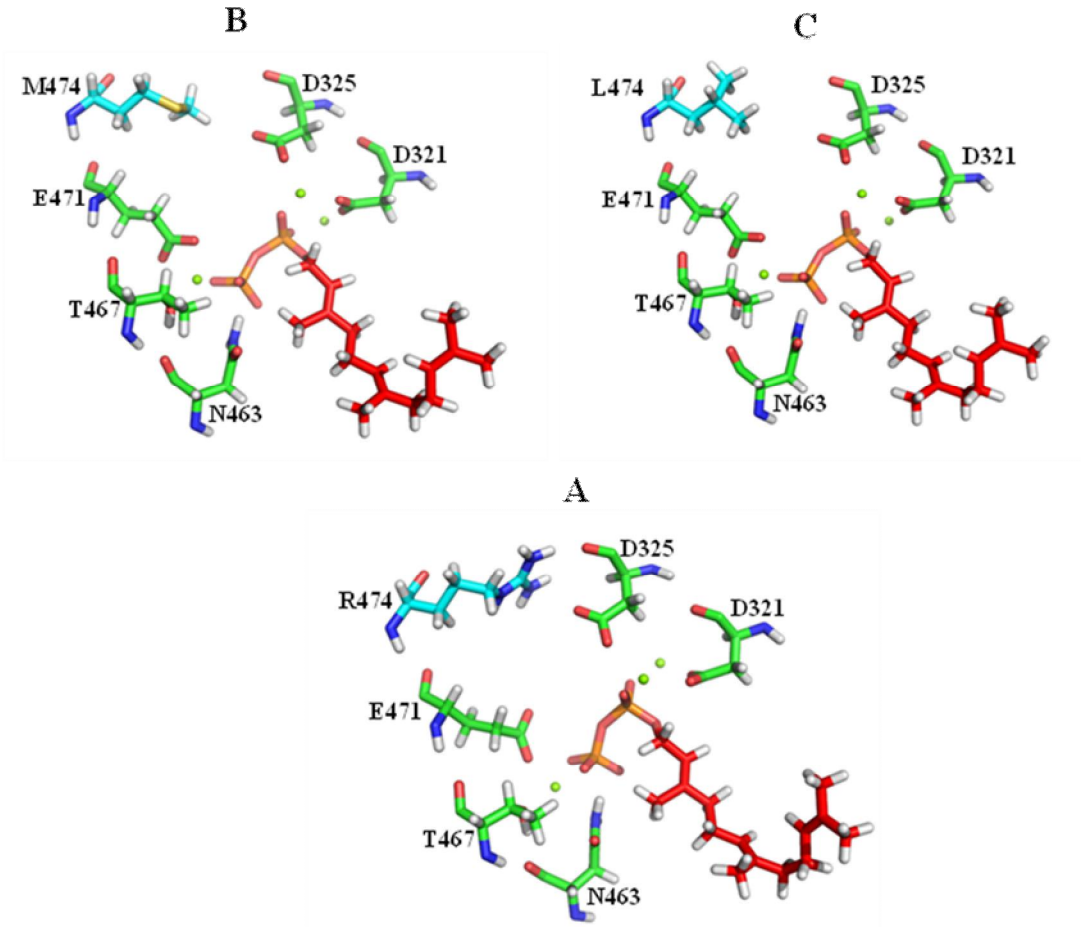


Figure 4.3.5.23: Interaction of (*E,E*)-FPP with residue R474 (Cyan color) and metal binding domain (Green colour) for stabilizing the carbocation, **A)** R474 interaction with (*E,E*)-FPP, **B)** M474 (mutant R474M) interaction with (*E,E*)-FPP, **C)** L474 (mutant R474L) interacting with (*E,E*)-FPP

4.3.5.4 Glutamine 289 and Serine 290 (Q289 and S290)

Glutamine 289 and Serine 290 are present at the edge of one of the helices surrounding the active site pocket. Mutant Q289A and S290A resulted in the formation of *exo*-α-bergamotene (53-57 %) as a major product along with small amount of α- and β-santalenes, but the mutant S290G resulted in exclusive formation of exo-α-bergamotene (86 %), whereas S290L leads to complete loss of activity (Figure 4.3.24). This loss may be due to incorporation of long bulky side chain which may lead to perturbation in the active site pocket.

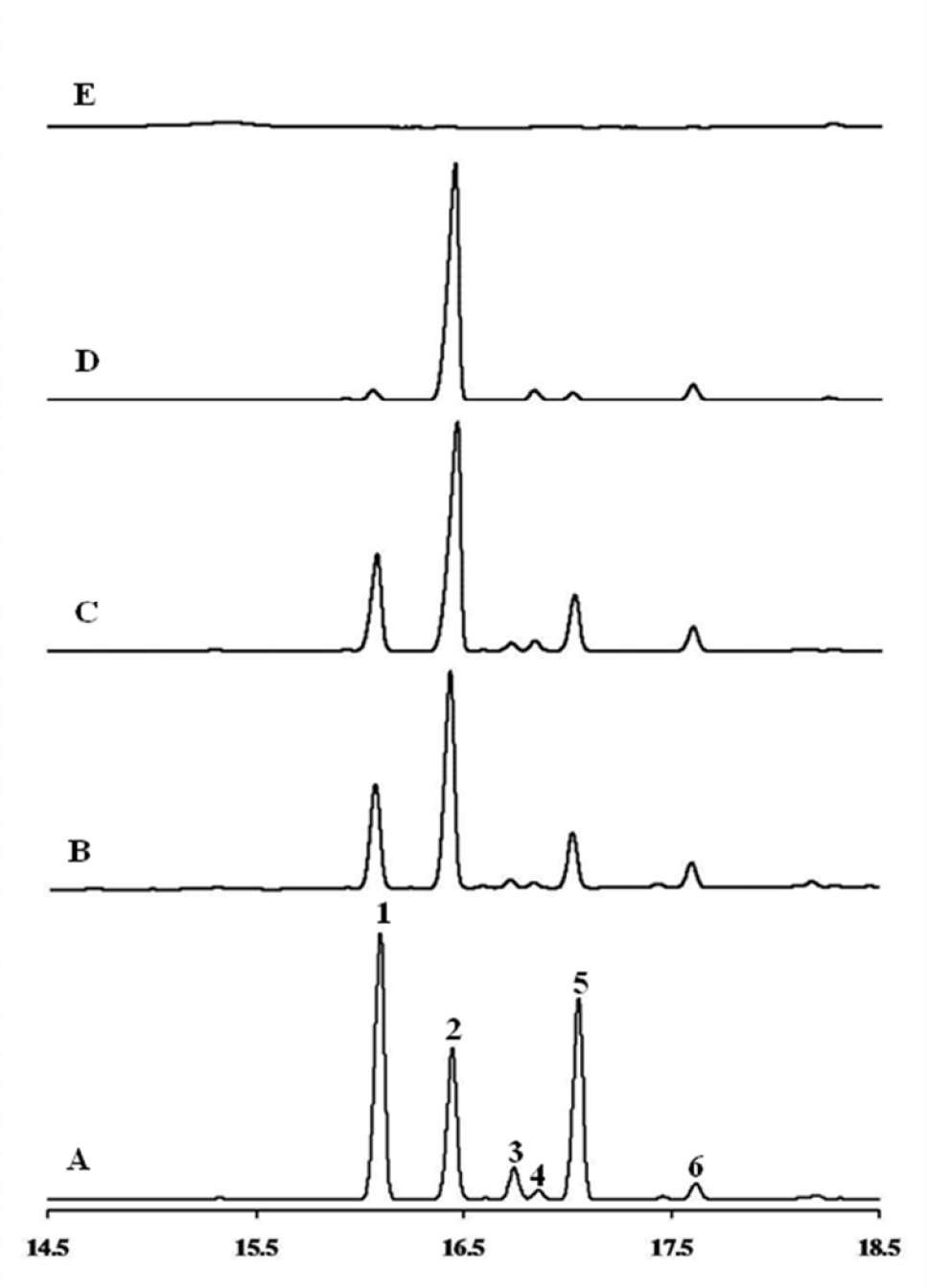


Figure 4.3.24: GC profile of mutants of Q289 and S290 with (*E,E*)-FPP, **A)** *SaSS* wild type, α-santalene (**1**), exo-α-bergamotene (**2**), *epi*-β-santalene (**3**), (*E*)-β-farnesene (**4**), β-santalene (**5**), exo-β-bergamotene (**6**), **B)** *SaSS* mutant **Q289A**, **C)** *SaSS* mutant **S290A**, **D)** *SaSS* mutant **S290G**, **E)** *SaSS* mutant **S290L**

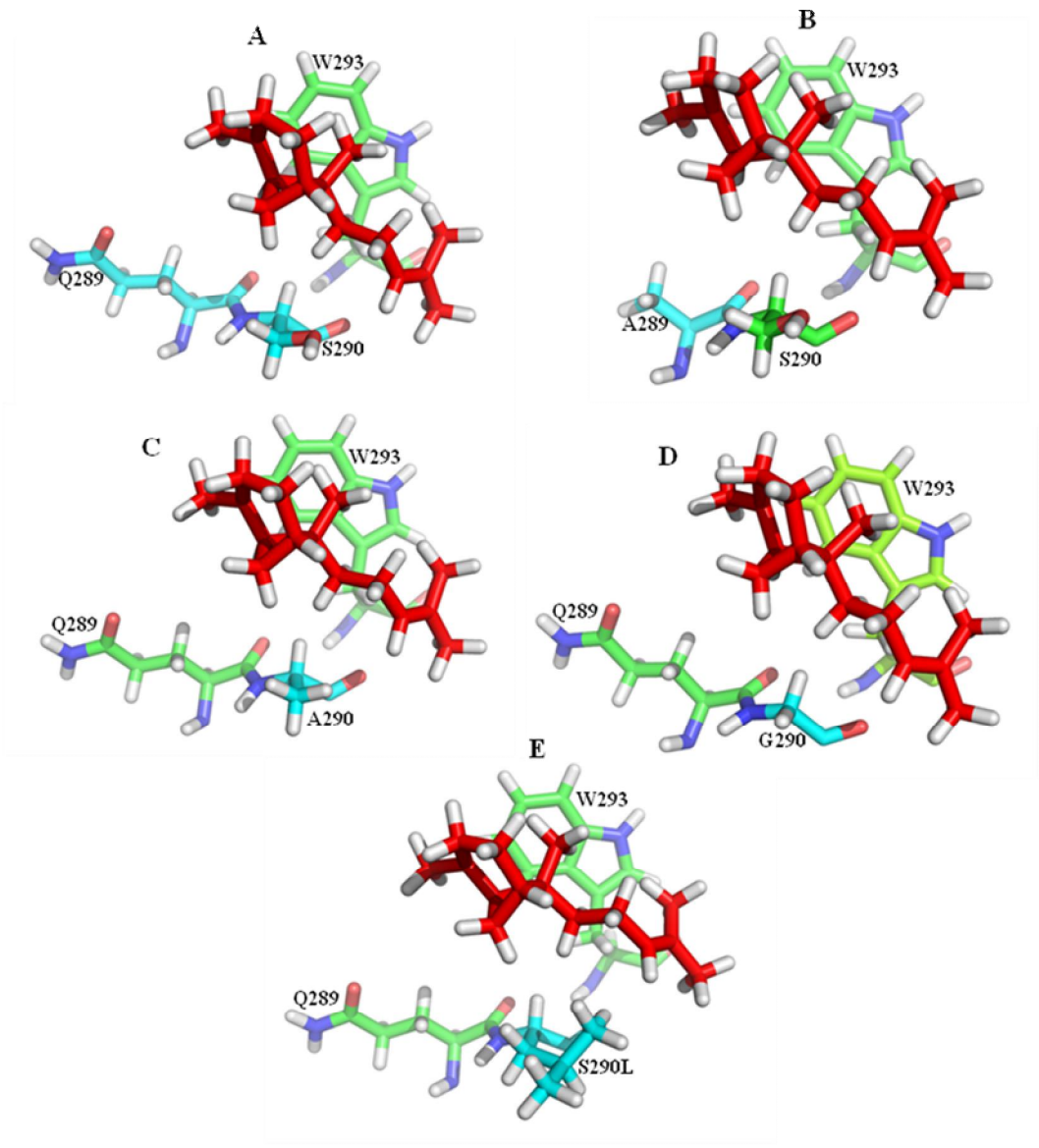


Figure 4.3.25: Interaction of residues Q289 and S290 with bergamotyl carbocation, **A)** Q289, S290 (Cyan colour) interaction with bergamotyl carbocation, **B)** *SaSS* mutant **Q289A** (Cyan colour), **C)** *SaSS* mutant **S290A** (Cyan colour), **D)** *SaSS* mutant **S290G** (Cyan colour), **E)** *SaSS* mutant **S290L** (Cyan colour)

These results provide an insight that W293, Q289, S290, T318, T467, Y37, Y396 and R474 amino acid residues play an important role in maintaining the electronic distribution in the active site pocket for stabilizing the complex carbocation cascade in the formation of santalenes.

4.3.5.5 Tyrosine 37 and Tyrosine 396 (Y37 and Y396)

Analysis of docking studies suggests that tyrosine 37 and tyrosine 396 are present on opposite sides of the active site, positioning their hydroxyl group inside the pocket. It

was believed that presence of multiple hydroxyl derivatives of amino acids is responsible for the formation of multiple products. Incorporation of phenylalanine at tyrosine 37 (Y37F) resulted in the formation of exo-α-bergamotene as the major product (64 %) along with small percentages of α-santalene (17 %) and β-santalene (9 %). A similar profile was also observed with mutant Y396F (Figure 4.3.26). The activity of Y37F and Y396F demonstrated the crucial role of hydroxyl group in stabilizing the carbocation cascade for the formation of α- and β-santalenes (Figure 4.3.27).

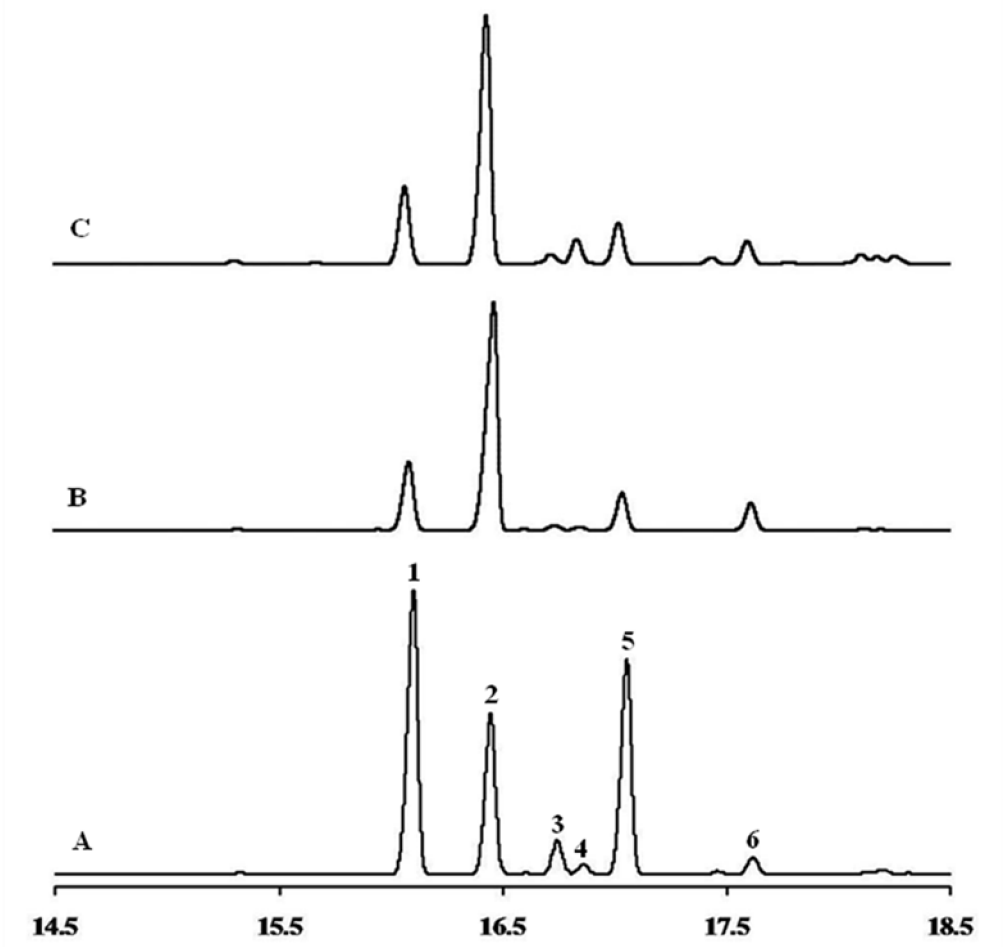


Figure 4.3.26: GC profile of mutants of Y37 and Y396 with (*E,E*)-FPP, **A)** *SaSS* wild type, α-santalene (**1**), exo-α-bergamotene (**2**), *epi*-β-santalene (**3**), (*E*)-β-farnesene (**4**), β-santalene (**5**), exo-β-bergamotene (**6**), **B)** *SaSS* mutant **Y37F**, **C)** *SaSS* mutant **Y396F**

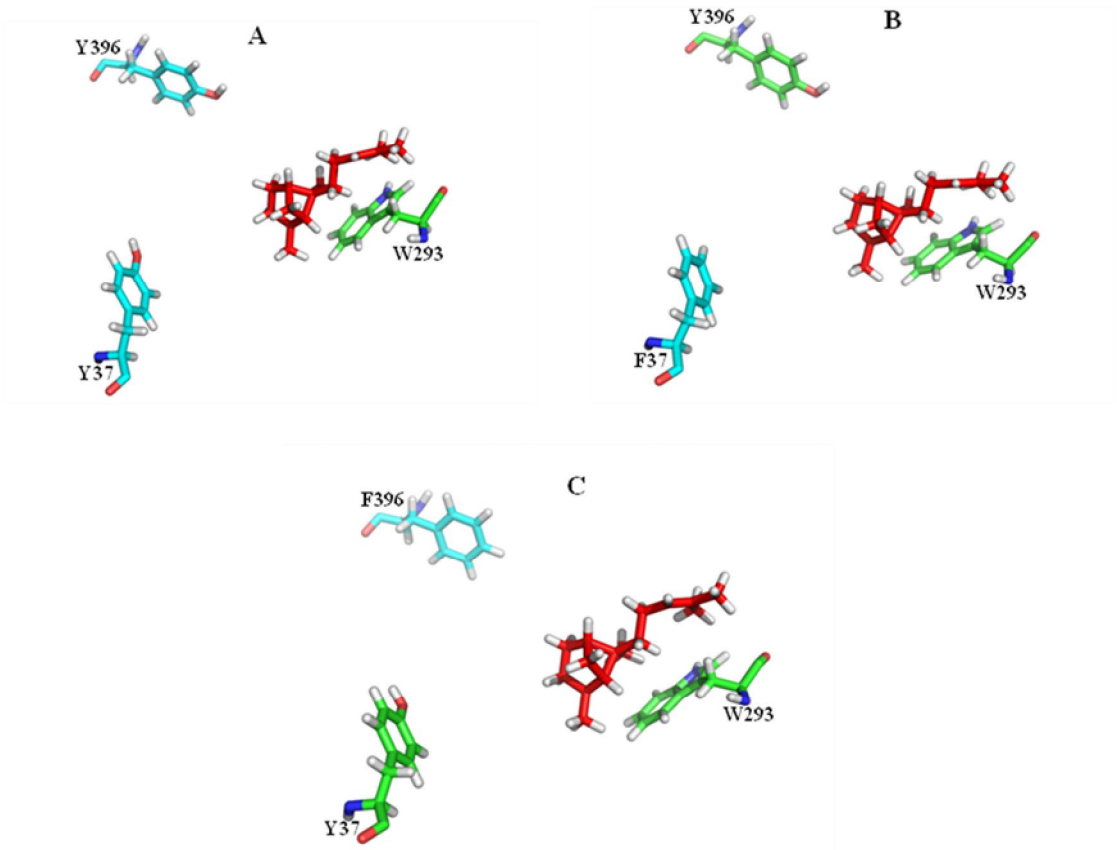


Figure 4.3.27: Interaction of bergamotyl carbocation with Y37 and Y396 and their mutants, **A)** Bergamotyl carbocation exposing to Y37, Y396 (Cyan colour), **B)** Bergamotyl carbocation exposing to Y37F (Cyan colour) **C)** Bergamotyl carbocation exposing to Y396F (Cyan colour)

4.3.6 Residues involved in giving shape to active site pocket

As proposed in Scheme 4.3.1 of santalene biosynthesis, acyclic FPP, after dephosphorylation, undergoes 1-6 ring closure to form bisabolyl carbocation followed by C2-C7 closure for the formation of bergamotene and santalenes. Hydrophobic amino acids (I422, L427) present at the bottom of the active site cleft are thought to be involved in maintaining the well packed environment surrounding the carbocation and to be helping in pushing the allylic end of linear FPP to form sesquiterpenes of multiple ring. The reaction co-ordinates of C2 and C7 were found to be 4.5Å apart in the bisabolyl carbocation, which decreases to 1.8Å in the bergamotyl carbocation (Scheme 4.3.1). This decrease in distance between the reaction co-ordinates may be governed by hydrophobic interactions, facilitating the C2-C7 closure and ultimately resulting in the formation of the bergamotyl cation in closed form of the enzyme.

4.3.6.1 Isoleucine 422 (I422)

Mutation of amino acids present in active site cleft may not only affect the side chain but can also lead to altering the secondary structure and product specificity. The reaction co-ordinates, C2 and C7 in the bisabolyl carbocation were found to be 4.5Å apart in the bisabolyl carbocation, which decreases to 1.8Å in the bergamotyl carbocation (Scheme 4.3.1). This decrease in distance between the reaction co-ordinates may be governed by hydrophobic interactions, facilitating the C2-C7 closure and ultimately resulting in the formation of the bergamotyl cation. Docking studies suggests that I422 is positioned at the bottom of active site pocket and was found to be involved in providing support to hydrophobic tail of bisabolyl carbocation for positioning the co-ordinates of C2-C7 in the closure of bergamotyl carbocation. Removal of hydrophobic side chain in mutant I422A led to a change in the active site pocket by increasing the free space, which in turn, minimizes the hydrophobic interactions responsible for C2-C7 closure. Because of this, majority of the products are formed by deprotonating the bisabolyl carbocation or by C7-C6 hydride shift, followed by deprotonation from C5, resulting in β-curcumene (59 %) and C1, resulting in γ-curcumene (13 %), depending on the positioning of the surrounding amino acids (Figure 4.3.28). I422 is present in the active site pocket and influences the cyclization of bisabolyl carbocation to bergamotyl carbocation by enforcing the specific folding on substrate with the help of side chain (Figure 4.3.29). This mutant has resulted in remarkably different product profile providing the structural basis of rapid evolution by a small number of mutations in the active site pocket.

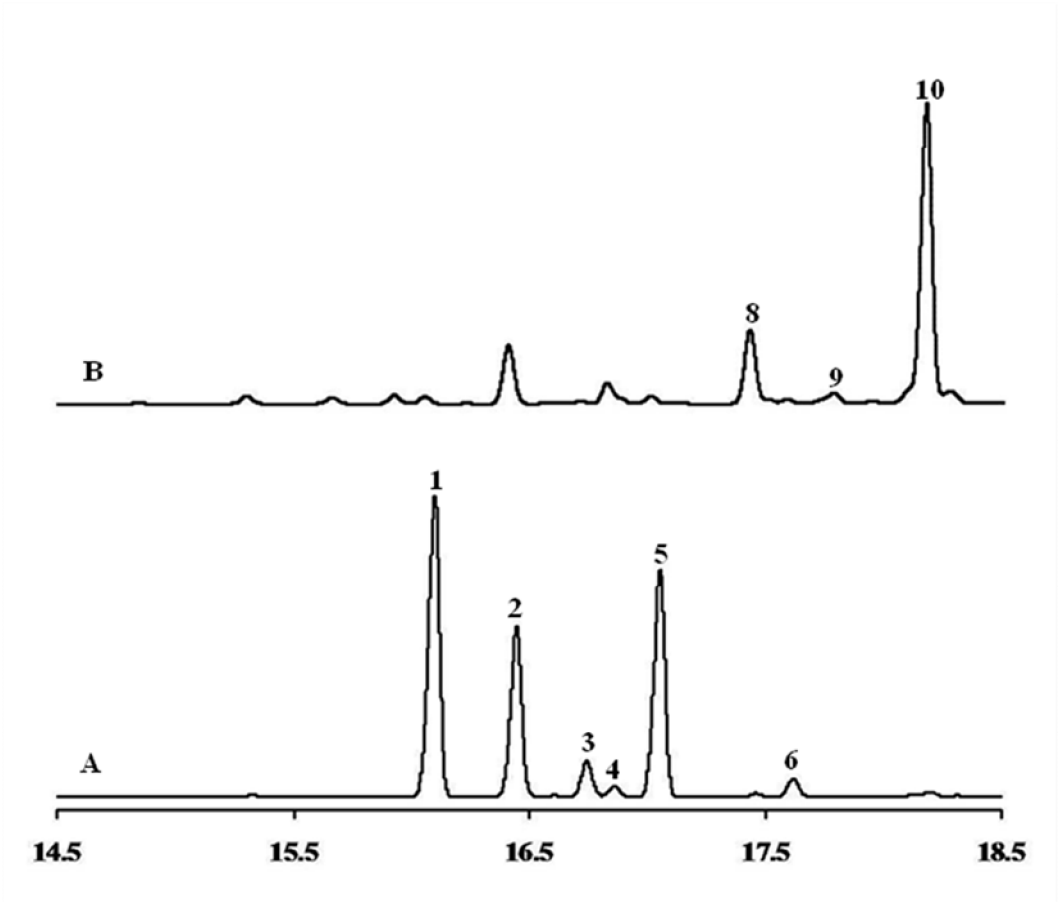


Figure 4.3.28: GC profile of *SaSS* mutant of I422A with (*E,E*)-FPP, **A)** *SaSS* wild type, α-santalene (**1**), exo-α-bergamotene (**2**), *epi*-β-santalene (**3**), (*E*)-β-farnesene (**4**), β-santalene (**5**), exo-β-bergamotene (**6**), **B)** *SaSS* mutant **I422A**, γ-curcumene (**8**), α-zingiberene (**9**), β-curcumene (**10**)

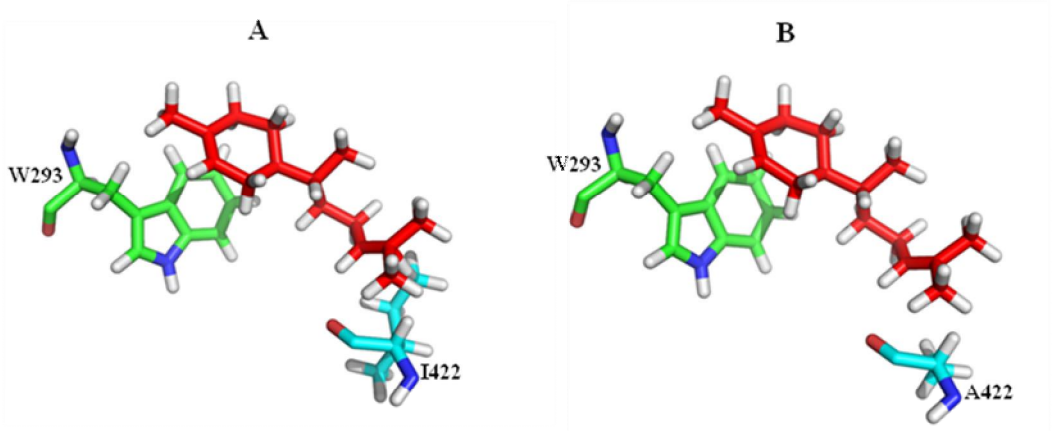


Figure 4.3.29: Interaction of isoleucine 422 (I422) and its mutant with bisabolyl carbocation, **A)** Isoleucine 422 (Cyan colour) positioning bisabolyl carbocation, **B)** Mutant I422A (Cyan colour) positioning bisabolyl carbocation

4.3.6.2 Leucine 427 (L427)

Leucine 427 is a part of the conserved leucine rich domain and is present at the bottom of the active site cleft to provide shape and hydrophobic interactions for cyclization of linear carbocation. Introduction of mutation L427A facilitated the formation of (*E*)-β-farnesene with 40-50 folds increase, as compared to the wild type and decreased the bicyclic and tricyclic product formation (Figure 4.3.30) presumably by alteration in the active site pocket.

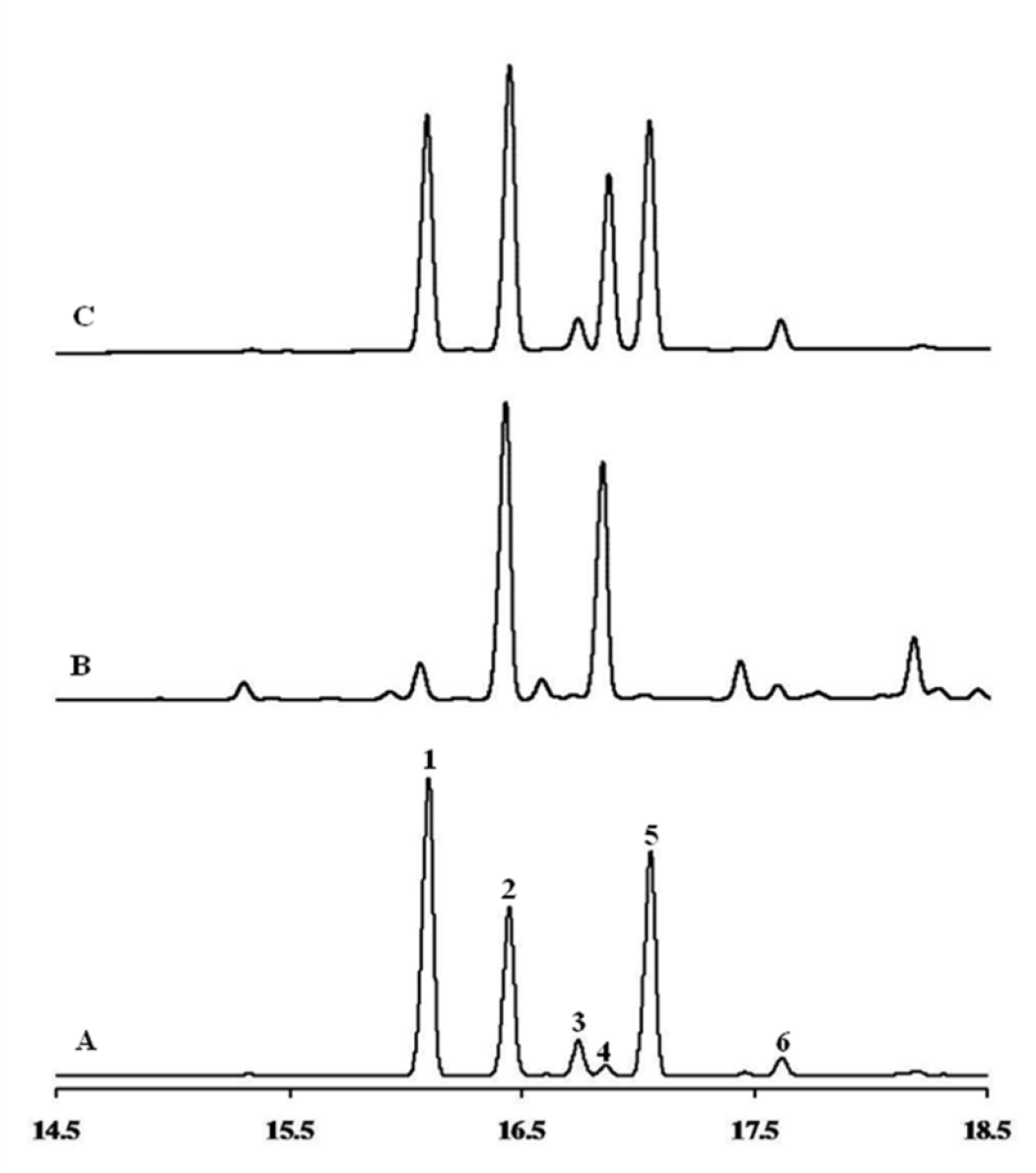


Figure 4.3.30: GC profile of mutants of L427 with (*E,E*)-FPP, **A)** *SaSS* wild type, , α-santalene (**1**), exo-α-bergamotene (**2**), *epi*-β-santalene (**3**), (*E*)-β-farnesene (**4**), β-santalene (**5**), exo-β-bergamotene (**6**), **B)** *SaSS* mutant **L427A**, **C)** *SaSS* mutant **L427I**

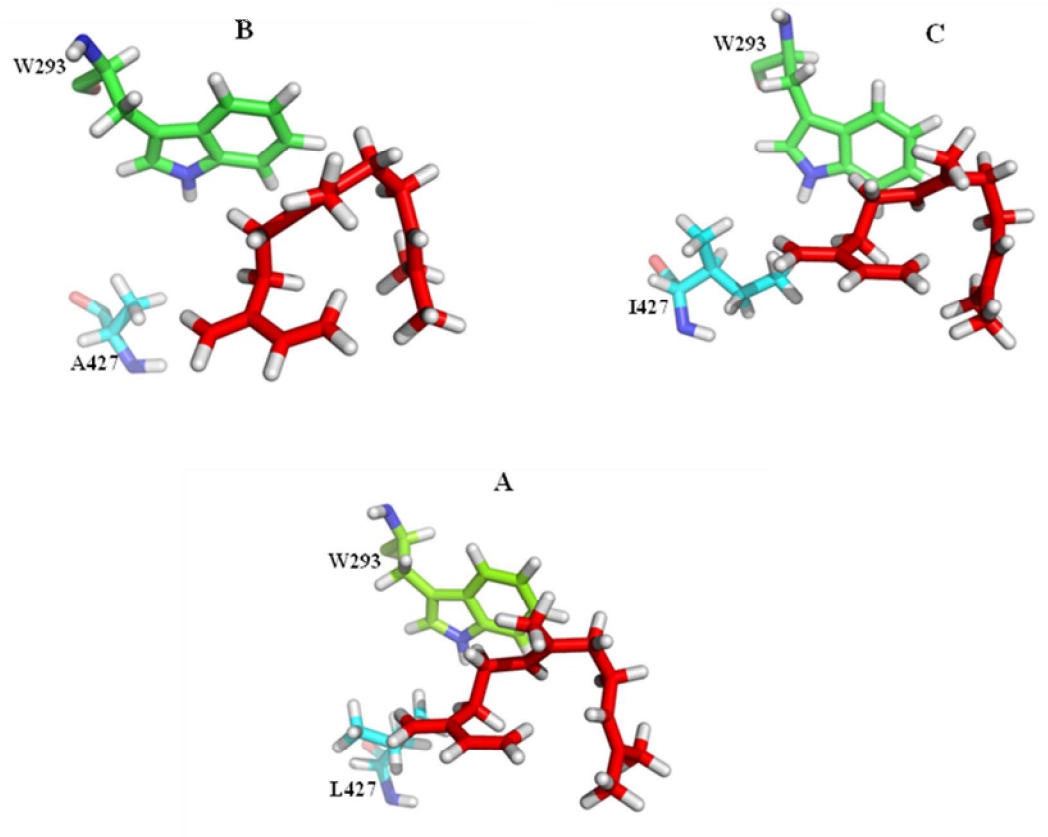


Figure 4.3.31: Docking posses for interaction of nerolidyl carbocation with leucine 427 and its mutants (Cyan colour), **A) L427** interaction with nerolidyl carbocation, **B) A427** interaction with nerolidyl carbocation, **C) I427** interaction with nerolidyl carbocation

Removal of long side chain from leucine to alanine in mutant L427A has an enlarged active site pocket which minimizes the hydrophobic interactions. This result clearly indicates that L427 is involved in providing the support for early cyclization step (C1-C6 ring closure, scheme 4.3.1) for the formation of bisabolyl carbocation (Figure 4.3.31).

4.3.6.3 Phenylalanine (F538)

Docking of intermediates involved in santalene biosynthesis and analysis of active site residues revealed that F538 is positioned to provide shape to the active site cleft for appropriate folding of carbocation cascade. Incorporation of hydroxyl group to phenylalanine (F538Y) resulted in similar profile as *SaSS* wild type. However, when phenylalanine was replaced with methionine (F538M), the mutant yielded one major uncharacterized solvated product (Figure 4.3.32) along with other sesquiterpenes as observed with wild type *SaSS*. These results clearly indicate that F538 is positioned to provide shape to active site cleft by positioning terminal part outside the active site

pocket (Figure 4.3.33) because of which F538Y did not result in any change in the profile. Incorporation of linear long side chain distorted the shape of the active site and the carbocation became accessible to water molecules leading to hydrolysis of carbocation intermediate to form a solvated product.

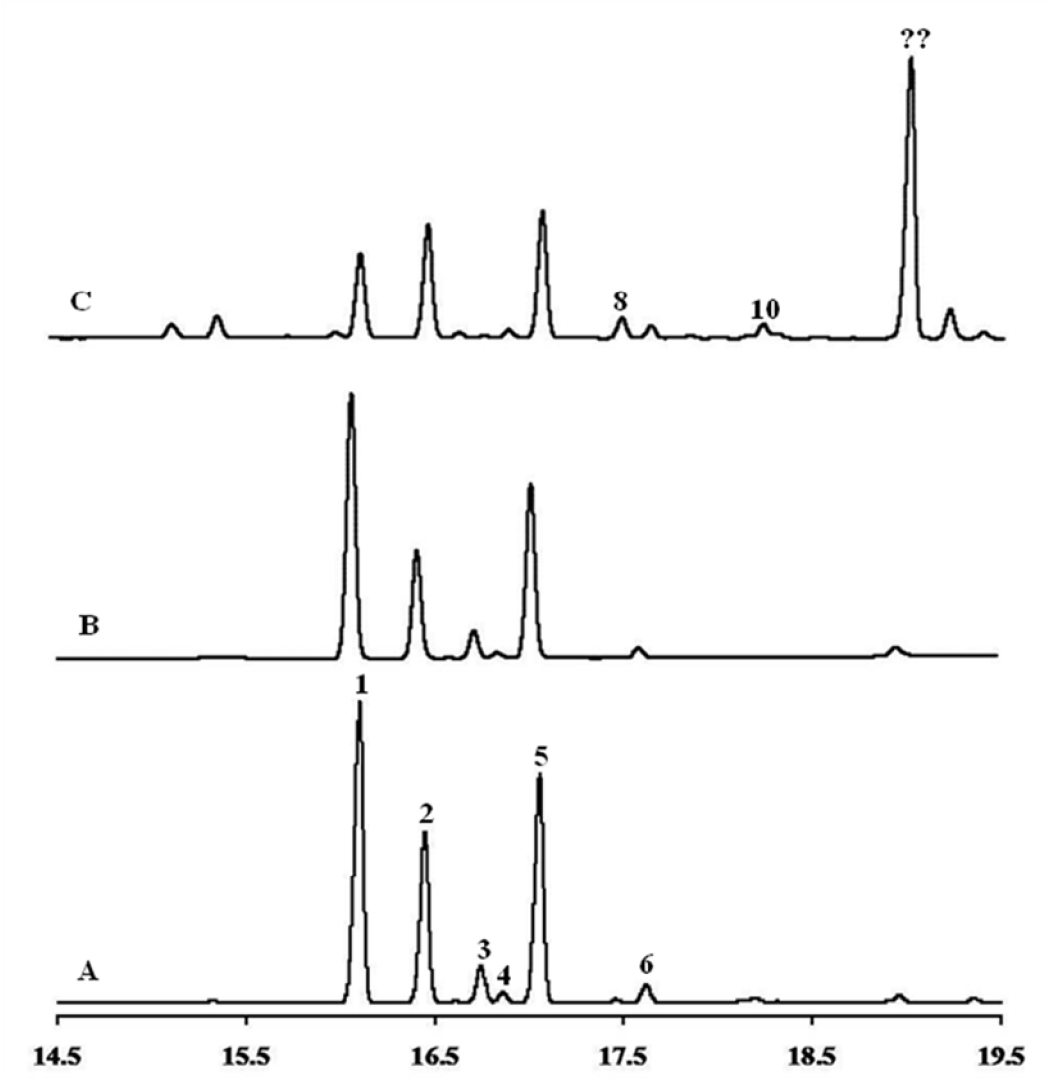


Figure 4.3.32: GC profile of mutants of F538 with (*E,E*)-FPP, **A)** *SaSS* wild type, α-santalene (**1**), exo-α-bergamotene (**2**), *epi*-β-santalene (**3**), (*E*)-β-farnesene (**4**), β-santalene (**5**), exo-β-bergamotene (**6**), **B)** *SaSS* mutant **F538Y**, **C)** *SaSS* mutant **F538M**

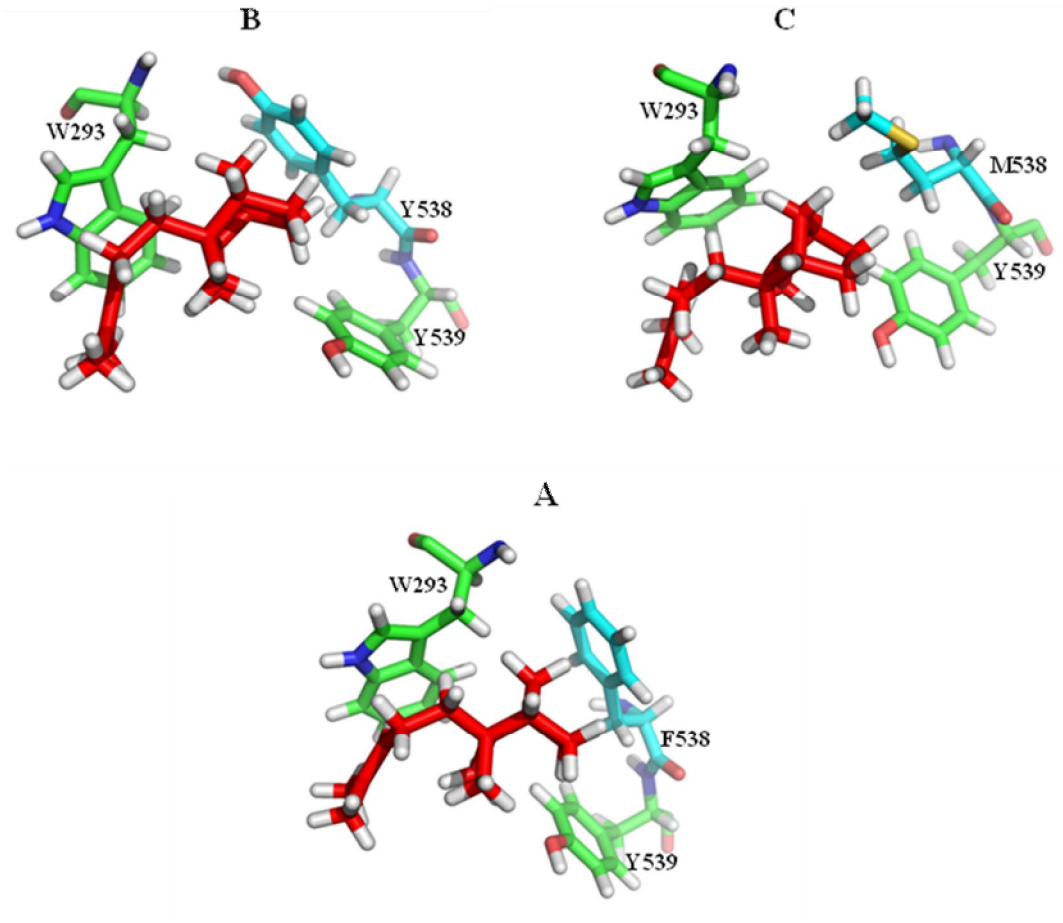


Figure 4.3.33: Docking posses of F538 (Cyan colour) and its mutants with santalyl carbocation, **A)** F538 interaction with santalyl carbocation, **B)** Y538 interaction with santalyl carbocation showed no change in the active site pocket, **C)** M538 interaction with santalyl carbocation showed the disturbance of active site

4.3.6.4 Large scale expression and purification of *SaSS* mutant I422A for the characterization of metabolite 8 and 10

The major metabolite of mutant I422A was characterized by setting up a large-scale enzyme assay as discussed in materials and method section and the product was extracted using n-hexane. Crude extract (69 mg) obtained was further loaded on $AgNO_3$ coated silica column for purification of major metabolites. Metabolites **8** and **10** were purified from the column with the purity of ≥95 % (Figure 4.3.34) and spectral data was acquired from H^1, C^{13}, and DEPT NMR. Products were confirmed by comparing NMR data with reported NMR values of γ-curcumene and β-curcumene[30]. The reaction mechanism for the formation of these metabolites is proposed in scheme 4.3.2.

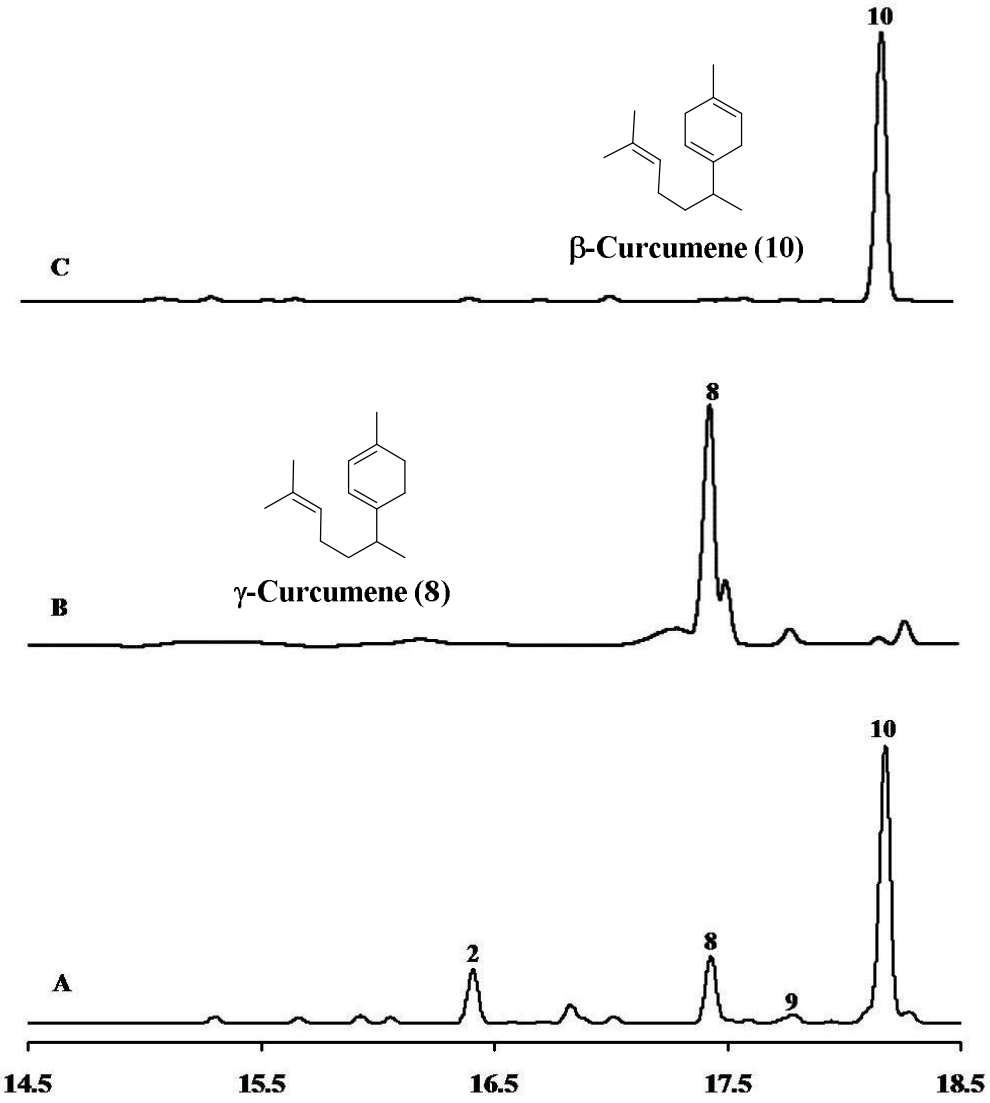
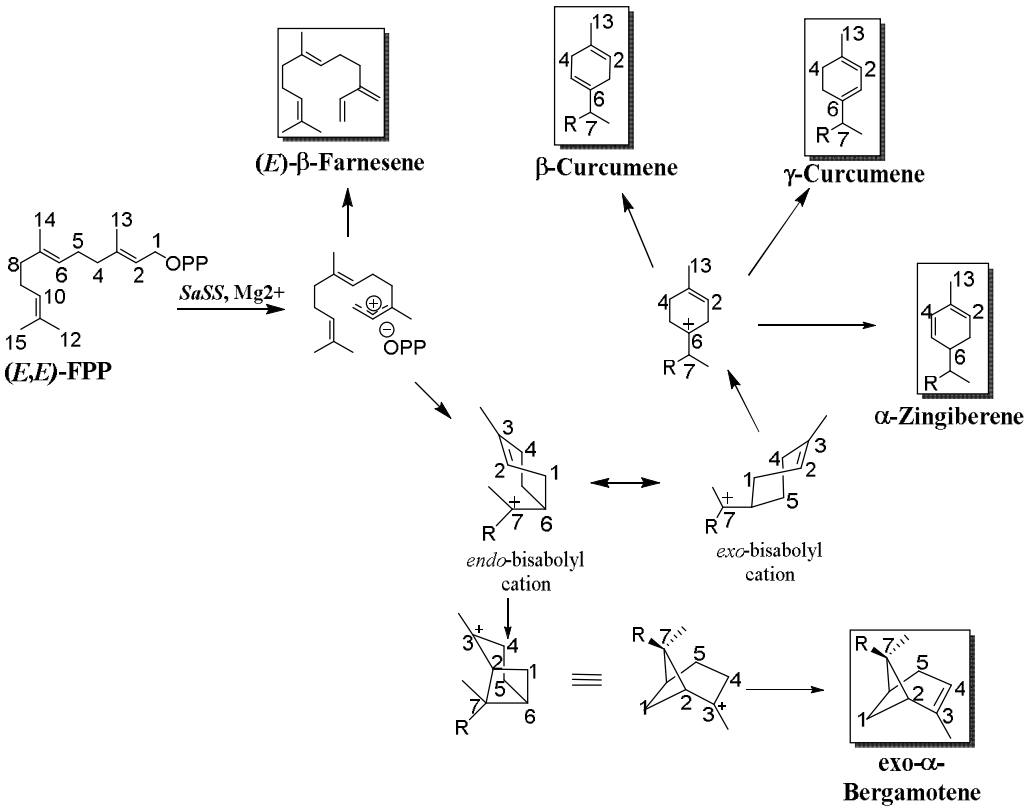


Figure 4.3.34: Purification of metabolite **8** and **10** from large scale assay mixture of **I422A**, **A)** GC profile of *SaSS* mutant **I422A** with (*E,E*)-FPP, **B)** Purified γ-curcumene (**8**), **C)** Purified β-curcumene (**10**)



Scheme 4.3.2: Schematic representation for the proposed mechanism of metabolites formed by I422A and L427A

4.3.6.5 Double mutant (T318A/Y539F and I422A/Y539F)

In order to find out the exact roles of tyrosine 539, threonine 318 and isoleucine 422, double mutants (T318A/Y539F and I422/Y539F) were constructed. Analysis of double mutant T318A/Y539F resulted in exclusive formation of bergamotene similar to the independent profile of T318A and Y539F mutants, whereas double mutant I422A/Y539F resulted in profile similar to that of I422A without santalenes formation (Figure 4.3.35). These results provide insight that isoleucine 422 plays a crucial role in positioning C2-C7 for the second cyclization step responsible for the formation of bergamotyl carbocation. However, no effect on the first cyclization step responsible for the formation of bisabolene carbocation was observed, as major metabolites are formed from bisabolene skeleton. Tyrosine 539 plays a crucial role for the arrangement of bergamotyl carbocation into santalyl carbocation but has no effect on the steps leading to the formation of bergamotyl carbocation. Threonine 318 also helps in arrangement of bergamotyl carbocation (Figure 4.3.36).

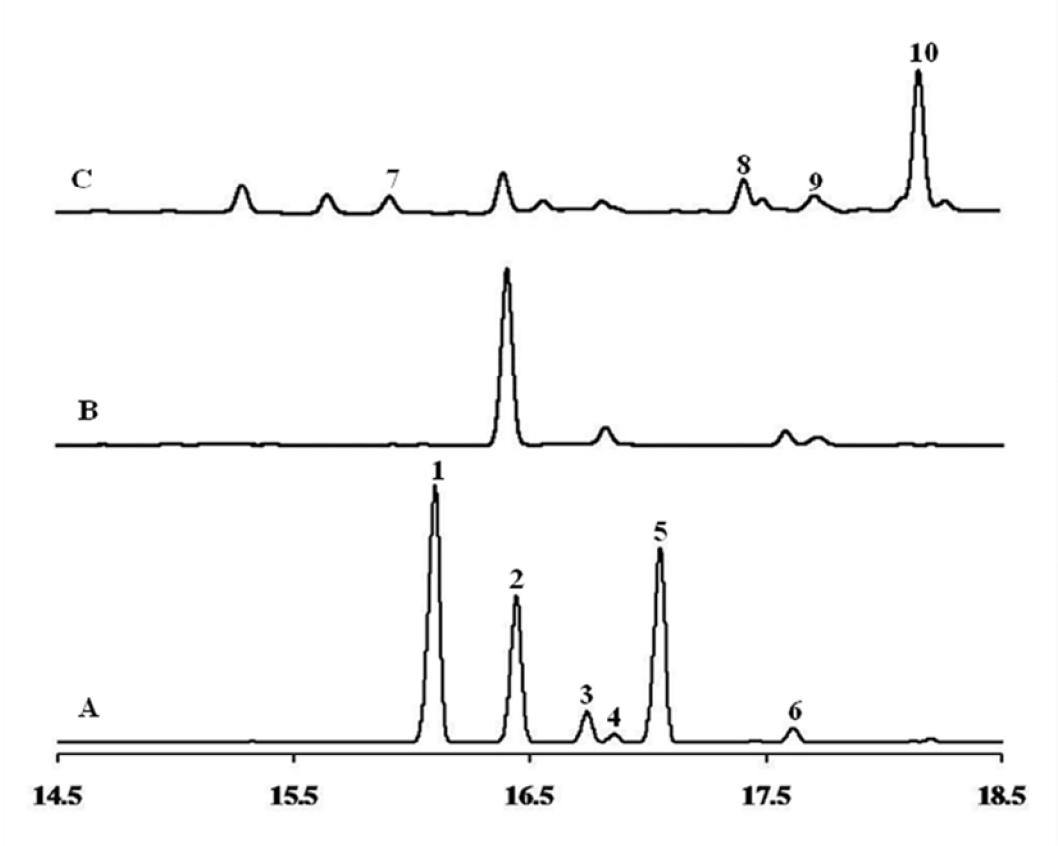


Figure 4.3.35: GC profile of double mutants with (*E,E*)-FPP, **A)** *SaSS* wild type, α-santalene (**1**), exo-α-bergamotene (**2**), *epi*-β-santalene (**3**), (*E*)-β-farnesene (**4**), β-santalene (**5**), exo-β-bergamotene (**6**), **B)** *SaSS* mutant **T318A & Y539F**, **C)** *SaSS* mutant **I422A & Y539F**

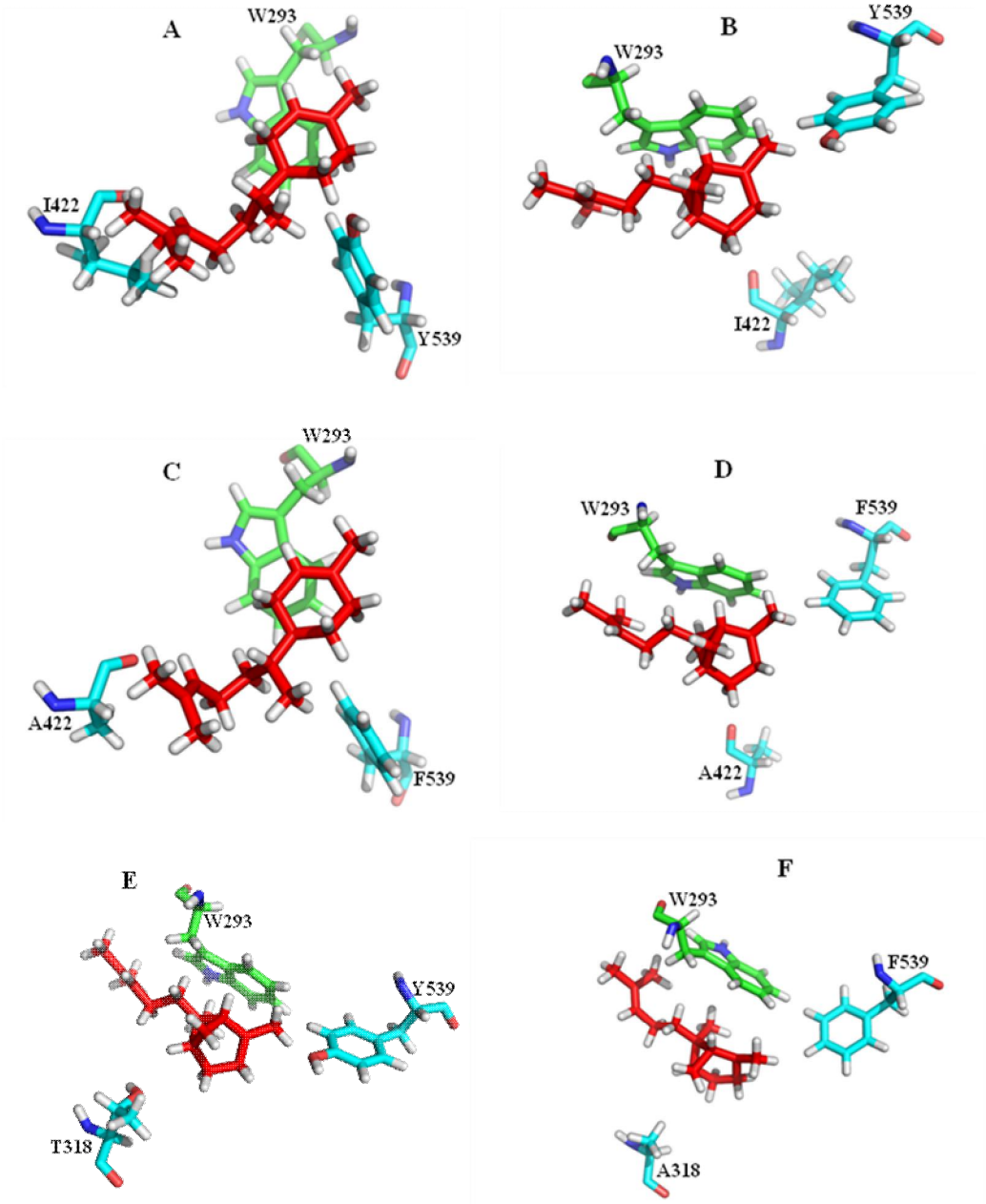


Figure 4.3.36: Interaction of I422, Y539 and T318 (Mutated residues are shown in cyan colour) with carbocation cascade**, A)** Interaction of bisabolyl carbocation with I422 and Y539, **B)** Interaction of bergamotyl carbocation with I422 and Y539, **C)** Interaction of bisabolyl carbocation with I422A and Y539F, **D)** Interaction of bergamotyl carbocation with I422A and Y539F, **E)** Interaction of bergamotyl carbocation with T318 and Y539, **F)** Interaction of bergamotyl carbocation with T318A and Y539F

4.3.7 Mutation without significant change in the product ratio

Mutants G323V/Y324F, Y324F, D470A, D549A, and W551F did not cause much alteration in the product ratios of wild type santalene synthase, even though these residues were present in the active site pocket (Figure 4.3.37).

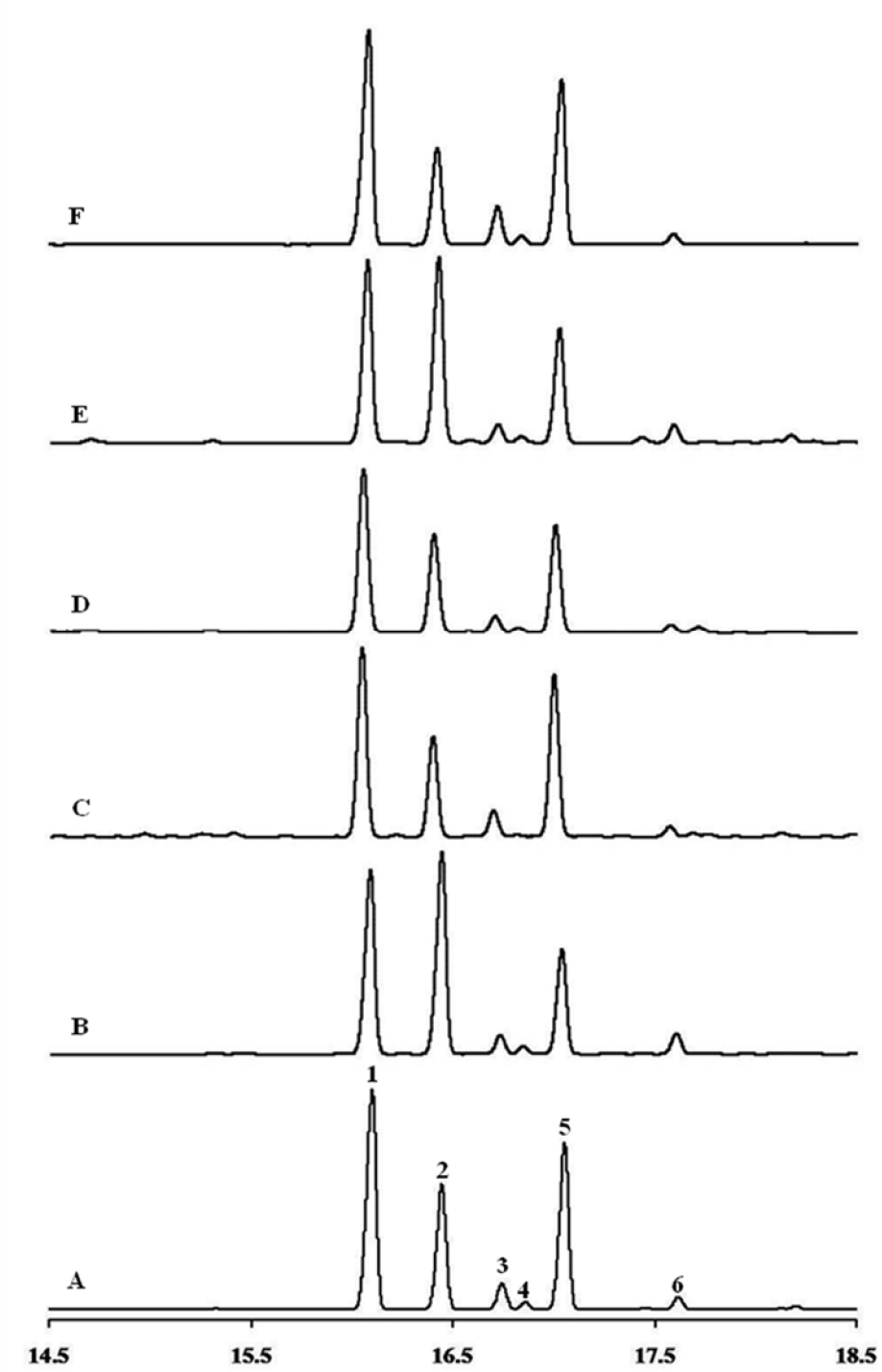


Figure 4.3.37: GC profile of mutants with (*E,E*)-FPP without significant change in the profile of wild type, **A)** *SaSS* wild type, α-santalene (**1**), exo-α-bergamotene (**2**), *epi*-β-santalene (**3**), (*E*)-β-farnesene (**4**), β-santalene (**5**), exo-β-bergamotene (**6**), **B)** *SaSS* mutant G323V/Y324F, **C)** *SaSS* mutant Y324F, **D)** *SaSS* mutant D549A, **E)** *SaSS* mutant W551F, **F)** *SaSS* mutant D470A

4.3.8 Lethal Mutation

Amino acids at the active site of SaSS were replaced with bulkier side chain amino acids with similar functionality as in the case of S459F/L461F and Y37F/S459F/L461F. Both the mutants S459F/L461F and Y37F/S459F/L461F resulted in complete loss of activity (Figure 4.3.38). This loss of activity may be due to loss of secondary structure of the mutants. This factor was supported further by solubility of these proteins which reduced drastically.

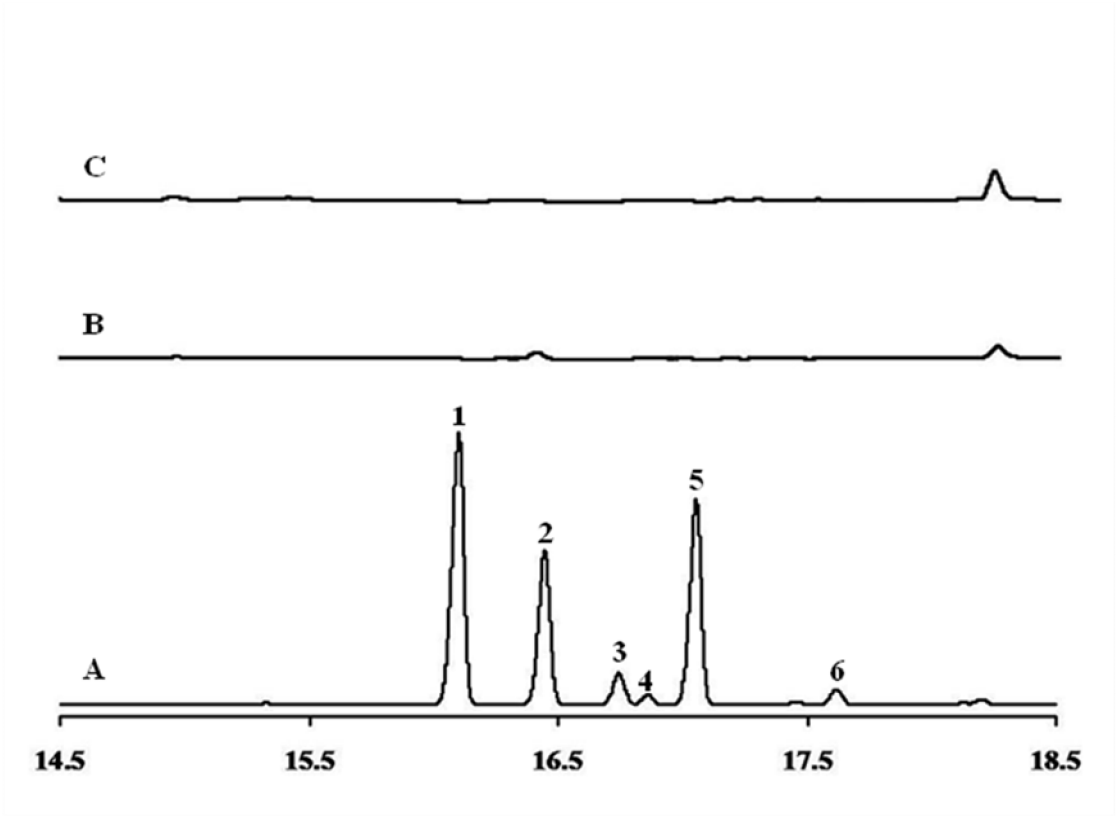


Figure 4.3.38: GC profile of mutants causing loss of secondary structure with (*E,E*)-FPP, **A)** *SaSS* wild type, α-santalene (**1**), exo-α-bergamotene (**2**), *epi*-β-santalene (**3**), (*E*)-β-farnesene (**4**), β-santalene (**5**), exo-β-bergamotene (**6**), **B)** *SaSS* mutant **S459F/L461F**, **C)** *SaSS* mutant **Y37F/S459F/L461F**

Table 4.3.1: Santalene synthase mutant product profile with (*E,E*)- FPP

	α-endo Bergamotene Rt (15.92)	α-Santalene Rt (16.05)	α-Bergamotene Rt (16.40)	*epi*-β Santalene Rt (16.70)	β-Farnesene Rt (16.82)	β-Santalene Rt (17.00)	γ-Curcumene Rt (17.43)	β-Bergamotene Rt (17.57)	α-Zingiberene Rt (17.82)	β-Curcumene Rt (18.22)	Uncharacerized molecules
SaSS Wild Type	---	41.22 ± 1.03	21.62 ± 0.55	4.36 ± 0.01	1.13 ± 0.04	29.48 ± 0.37	---	2.19 ± 0.06	---	---	---
SaSS Mutant G323V& Y324F	---	33.77 ± 1.23	37.23 ± 0.45	3.35 ± 0.25	1.17 ± 0.07	20.24 ± 1.00	---	4.24 ± 0.50	---	---	---
SaSS Mutant Y324F	---	39.33 ± 0.61	21.43 ± 0.57	4.87 ± 0.25	0.74 ± 0.48	31.61 ± 1.45	---	2.02 ± 0.05	---	---	---
SS Mutant T318A	---	5.66 ± 0.90	79.68 ± 0.58	---	4.82 ± 0.43	2.54 ± 0.43	---	7.30 ± 0.32	---	---	---
SaSS Mutant Y539F	---	1.98 ± 0.02	83.96 ± 1.58	---	1.23 ± 0.01	1.14 ± 0.14	---	11.23 ± 1.40	---	---	---
SaSS Mutant Y539W	23.46 ± 0.72	10.15 ± 0.22	30.54 ± 0.30	1.44 ± 0.01	10.51 ± 0.50	---	---	23.90 ± 0.31	---	---	---
SaSS Mutant Y539L	2.63 ± 0.00	3.5 ± 0.00	68.70 ± 0.84	---	14.74 ± 0.14	---	---	10.34 ± 0.50	---	---	---
SaSS Mutant Y539S	---	3.16	86.32	---	1.67	1.23	---	7.622	---	---	---
SaSS Mutant D549A	---	41.14 ± 0.99	24.10 ± 0.12	4.15 ± 0.19	0.92 ± 0.06	27.74 ± 0.70	---	1.95 ± 0.07	---	---	---
SaSS Mutant R474M	1.00 ± 0.11	2.4 ± 0.01	70.75 ± 2.43	---	1.49 ± 0.01	---	---	24.36 ± 2.56	---	---	---
SaSS Mutant R474L	0.97 ± 0.08	2.49 ± 0.01	69.13 ± 0.41	---	1.47 ± 0.12	---	---	25.94 ± 0.62	---	---	---
SaSS Mutant W293F	---	7.27 ± 0.54	65.89 ± 1.32	0.70 ± 0.06	8.61 ± 0.07	10.92 ± 0.66	---	4.48 ± 0.02	---	2.13 ± 0.15	---
SaSS Mutant I422A	1.85 ± 0.05	1.34 ± 0.14	11.83 ± 0.17	---	5.10 ± 0.34	1.26 ± 0.04	14.82 ± 0.59	1.01 ± 0.09	2.04 ± 0.03	60.75 ± 0.01	---
SaSS Mutant T318A& Y539F	---	---	83.02 ± 2.75	---	10.50 ± 2.43	---	---	6.48 ± 0.32	---	---	---
SaSS Mutant I422A& Y539F	5.99 ± 0.18	---	14.56 ± 0.36	---	1.99 ± 0.08	---	9.72 ± 0.89	1.34 ± 0.46	---	45.88 ± 0.49	20.52
SaSS Mutant Q289A	---	25.47 ± 0.39	52.99 ± 0.68	1.65 ± 0.01	1.11 ± 0.02	12.87 ± 0.34	---	5.91 ± 0.04	---	---	---
SaSS Mutant S290A	---	22.46 ± 0.43	57.02 ± 0.40	1.57 ± 0.02	2.01 ± 0.01	12.20 ± 0.17	---	4.74 ± 0.19	---	---	---
SaSS Mutant S290L	**No Activity**										

SaSS Mutant S290G	0.47 ± 0.09	3.36 ± 0.12	85.90 ± 0.39	---	2.92 ± 0.19	2.34 ± 0.01	---	5.01 ± 0.00	---	---	---
SaSS Mutant L427I	---	24.06 ± 0.41	28.40 ± 0.40	3.61 ± 0.10	17.21 ± 0.07	23.53 ± 0.42	---	3.19 ± 0.26	---	---	---
SaSS Mutant L427A	2.17 ± 0.98	5.86 ± 0.63	43.64 ± 0.76	0.37 ± 0.37	32.91 ± 0.40	---	4.28 ± 0.89	1.88 ± 0.04	---	6.08 ± 1.18	2.81
SaSS Mutant T467A	---	1.62 ± 0.06	73.76 ± 0.01	---	2.03 ± 0.06	---	---	22.59 ± 0.02	---	---	---
SaSS Mutant W551F	---	33.70 ± 0.54	33.95 ± 0.81	3.49 ± 0.09	1.27 ± 0.08	21.30 ± 0.03	1.41 ± 0.48	3.29 ± 0.03	---	1.59 ± 0.54	---
SaSS Mutant Y37F	---	17.88 ± 0.29	64.20 ± 0.76	1.26 ± 0.11	0.81 ± 0.08	9.2 ± 0.20	---	6.65 ± 0.07	---	---	---
SaSS Mutant Y396F	---	18.19 ± 0.18	58.53 ± 0.23	2.21 ± 0.05	5.48 ± 0.11	9.29 ± 0.05	1.43 ± 0.01	4.87 ± 0.19	---	---	---
SaSS Mutant N463I	**No Activity**										
SaSS Mutant N463D	---	42.73	43.02	1.82	1.29	7.34	---	3.80	---	---	---
SaSS Mutant E471Q	---	3.53 ±.12	72.90 ± 1.2	---	1.80 ± 0.11	---	---	22.80 ± 1.0	---	---	---
SaSS Mutant D470A	---	41.48 ± 0.17	18.00 ± 0.16	6.62 ± 0.10	1.43 ± 0.02	30.59 ± 0.23	---	1.89 ± 0.01	---	---	---
SaSS Mutant F538Y	---	44.98 ± 0.57	19.35 ± 1.04	4.32 ± 0.09	1.12 ± 0.23	28.42 ± 0.74	---	1.81 ± 0.12	---	---	---
SaSS Mutant F538M	---	11.48	14.97	---	0.78	16.54	2.735	1.71	---	1.19	50.58
SaSS Mutant S459F& L461F	**No Activity**										
SaSS Mutant Y37F, S459F& L461F	**No Activity**										
SaSS Mutant D543N	---	2.20 ± 0.03	83.15 ± 0.03	---	---	---	---	14.72 ± 0.00	---	---	---
SaSS Mutant D543L	**No Activity**										

4.3.9 Kinetic characterization of santalene synthase mutants

Kinetic characterization of all the mutants in comparison with wild type revealed that enzyme efficiency for mutant G323V& Y324F,Y539F,Y539W, R474M & R474L has slightly increased while in all other cases it is decreased. However, mutant Y539F and D549A have very high K_m, suggesting the role of these amino acids in facilitating the binding of substrate in active site pocket. These results clearly suggest that a promiscuity of the sesquiterpene enzymes can be utilized for directed engineering to produce new enzymes with desired activity without affecting the kinetic parameters much, suggesting the spontaneous theory of evolutionary origin where specific enzymes arise from surprisingly few mutations.

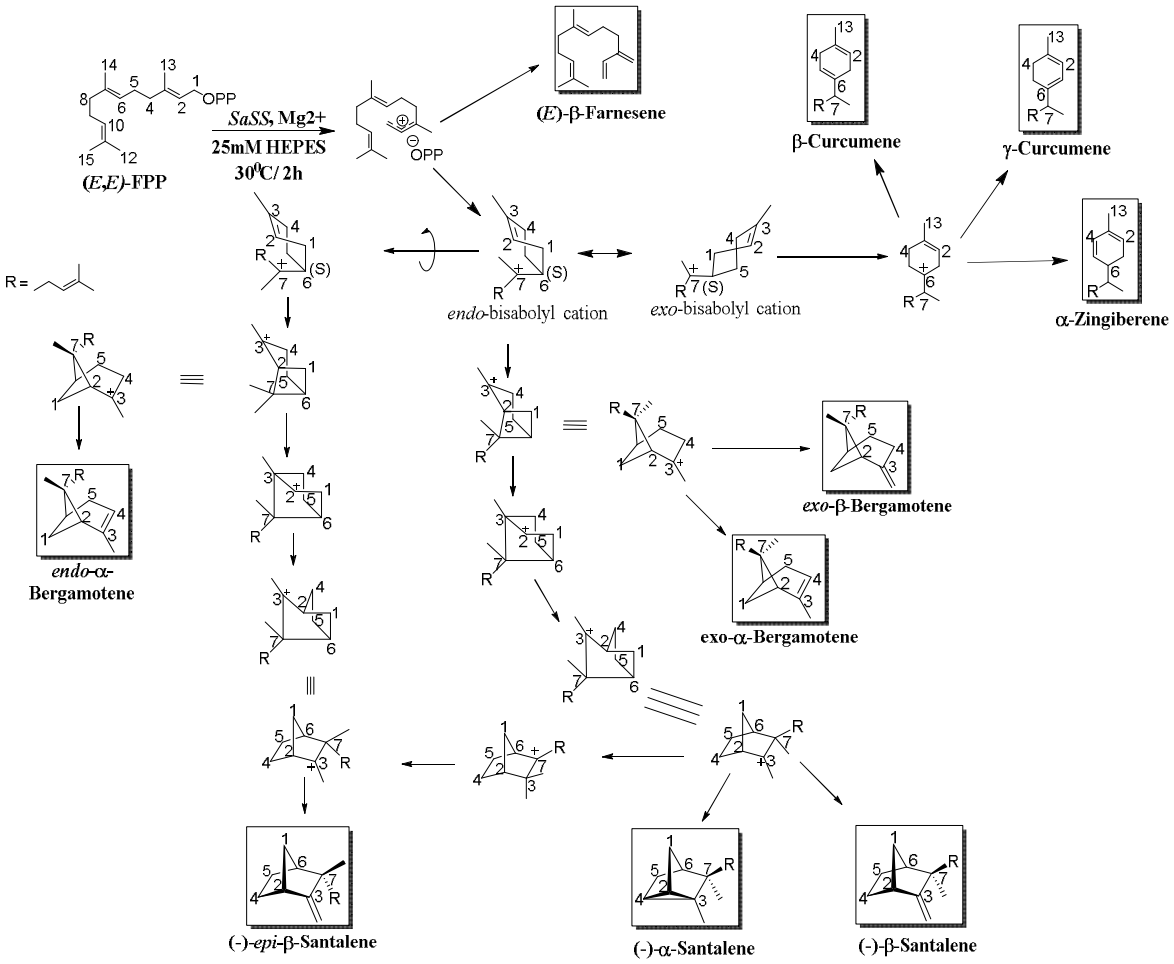
Table 4.3.2: Kinetic characterization table of santalene synthase mutant

Mutants	Km (µM)	Vmax (µM/min)	Kcat /min	Kcat/Km
SaSS wild type	0.586± 1.2	0.502 ± 0.01	0.100 ± 0.00	0.170
Mutant 1 (G323V &Y324F)	3.01 ± 1.2	0.42 ± 0.07	0.09 ± 0.01	0.031
Mutant 2 (Y324F)	23.41 ± 5.6	0.64 ± 0.07	0.13 ± 0.01	0.006
Mutant 3 (T318A)	6.17 ± 1.9	0.25 ± 0.03	0.05± 0.01	0.009
Mutant 4 (Y539F)	5.52 ± 0.8	0.61± 0.03	0.13 ± 0.01	0.023
Mutant 5 (Y539W)	3.11 ± 1.4	0.35 ± 0.03	0.07± 0.01	0.026
Mutant 6 (D549A)	17.3	0.360	0.23± 0.05	0.004
Mutant 7 (R474M)	2.69 ± 0.3	0.40 ± 0.03	0.08 ± 0.01	0.031
Mutant 8 (R474L)	ND	ND	ND	ND
Mutant 10 (W293F)	6.58 ± 0.3	0.37 ± 0.01	0.08 ± 0.00	0.012
Mutant 11 (I422A)	5.21 ± 0.7	0.38 ± 0.01	0.08 ± 0.00	0.015
Mutant 12 (T318A & Y539F)	2.77± 0.9	0.08 ± 0.01	0.02 ± 0.00	0.006
Mutant 13 (I422A & Y539F)	3.28 ± 0.2	0.09 ± 0.00	0.02 ± 0.00	0.006

4.4 Conclusion

The complex catalytic mechanism and formation of a wide variety of products catalyzed by terpene synthases from a single diphosphate has been a subject of great biochemical analysis. Several studies have indicated that a small number of amino acid substitutions are sufficient to alter the product profile dramatically. Site directed mutagenesis studies provides a platform to understand the functional role of a particular amino acid present in the active site pocket.

Our directed approach was to identify plasticity residues of the active site pocket, which determines the product profile of santalene synthase and to engineer them for the production of novel sesquiterpenes. 10 plasticity residues were identified (S290, W293, T318, Y396, I422, L427, N463, T467, R474, and Y539) which are responsible for the promiscuous nature of santalene synthase, and for drastic changes in the product outcome. As discussed in chapter 3, *SaSS* was able to cyclize the acyclic diphosphate, (*E,E*)-FPP into six sesquiterpenes of which three were major sesquiterpenes (α-, β-santalenes and exo-α-Bergamotene) and three were minor sesquiterpenes (*epi*-β-santalene, (*E*)-β-farnesene and exo-β-bergamotene), indicating that *SaSS* is moderately promiscuous in nature. Terpene cyclases function as a template and chaperone can be measured based on product diversity. Active site pocket analysis reveals that 7 hydroxy amino acids (Y37, S290, T318, Y324, Y396, T467, Y539,) are present in the pocket, which help in stabilizing the carbocation and final deprotonation, this being the main reason for the formation of multiple products. Our results suggest that these amino acids, surrounding the pocket, help in maintaining the geometry and electronic distribution in the active site for the formation of santalenes and bergamotenes. Identification of such a high plasticity residue in the active site pocket of *SaSS* provides the structural basis for rapid evolution, wherein small changes in the active site pocket can drastically modify the product outcome.



Scheme 4.4.1: Proposed mechanism for all the metabolites formed from all the mutants of *SaSS*

We successfully constructed the sesquiterpene synthases such as: α-santalene and exo-α-bergamotene synthase (N463D), exo-α-bergamotene synthase (S290G, W293F, T318A, Y539F, R474M/L, and Y539S), endo- α-bergamotene synthase (Y539W), β-curcumene synthase (I422A), (*E*)-β-farnesene synthase (L427A) producing a single or multiple product derived from the predominant reaction pathway of *SaSS* by maintaining the kinetic parameters (Scheme 4.4.1). These results provide the evidence that a single amino acid substitution in the active site pocket is sufficient to dramatically alter the product profile of *SaSS*. We believe that the results described here will aid in the production of novel and specific isoprenoids in pharmaceutical, flavor and fragrance as well as in biofuel industry.

Appendix 4.A GC-MS fragmentation of indentified terpenes

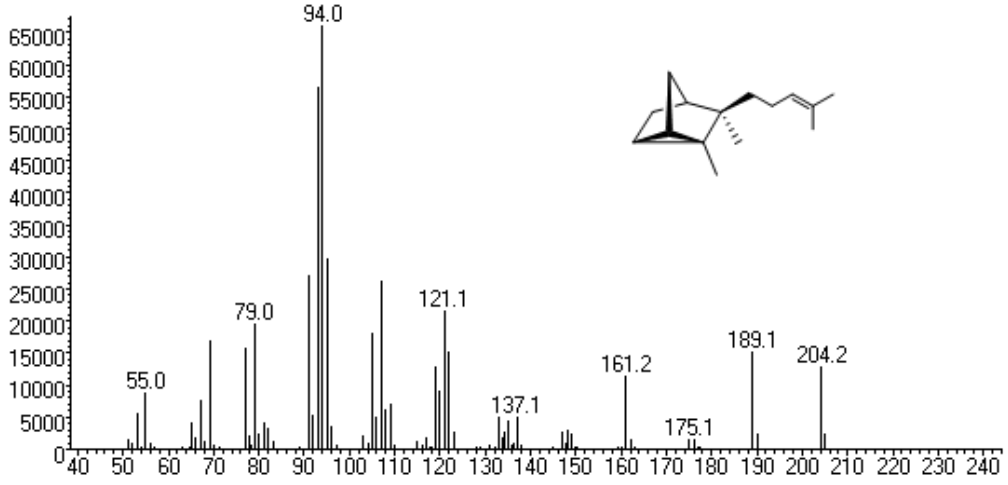


Figure 4.A.1: EI Mass Spectrum of α-Santalene (**1**)

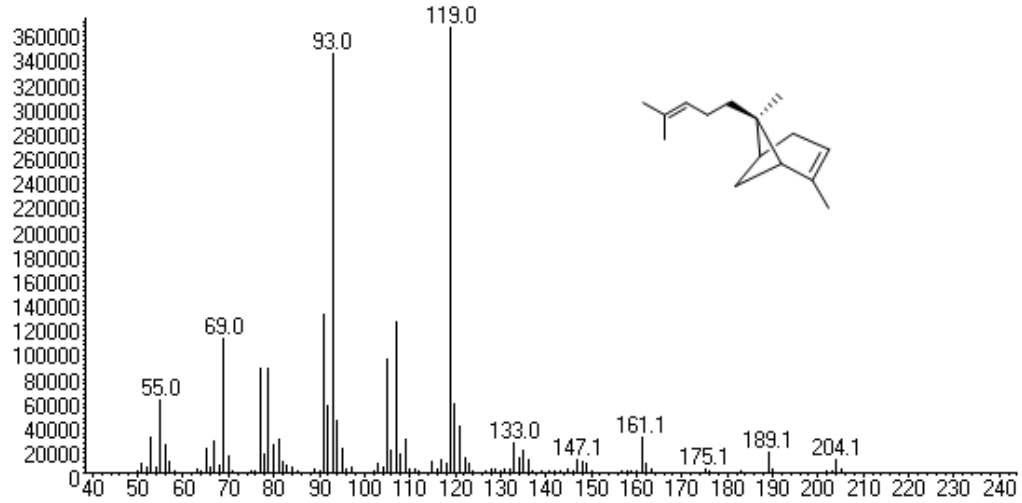


Figure 4.A.2: EI Mass Spectrum of *exo*-α-Bergamotene (2)

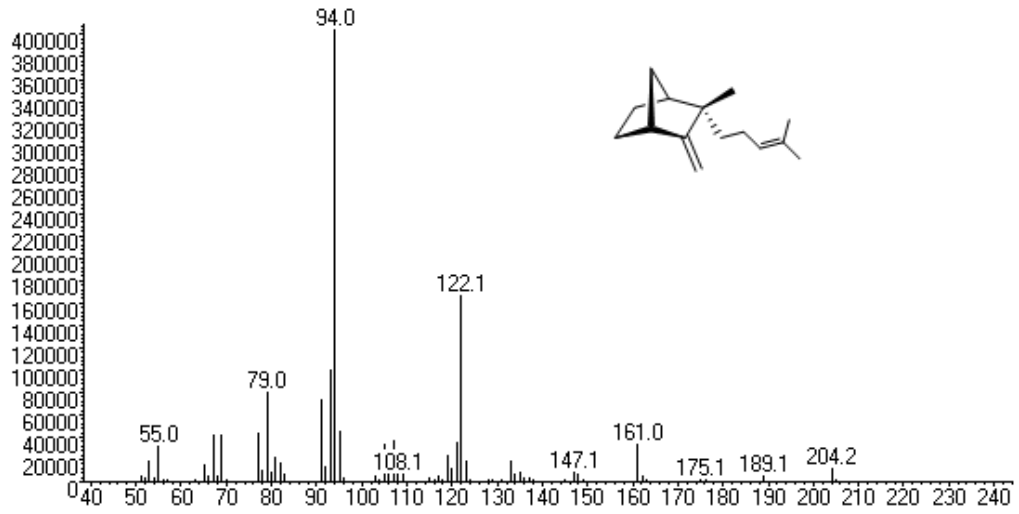


Figure 4.A.3: EI Mass Spectrum of *epi*-β-Santalene (3)

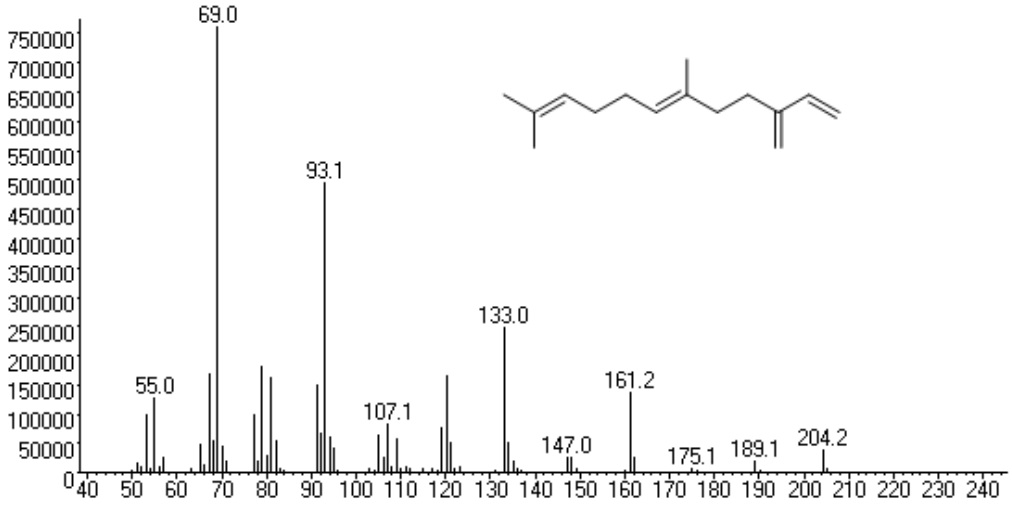


Figure 4.A.4: EI Mass Spectrum of (*E*)-β-Farnesene (**4**)

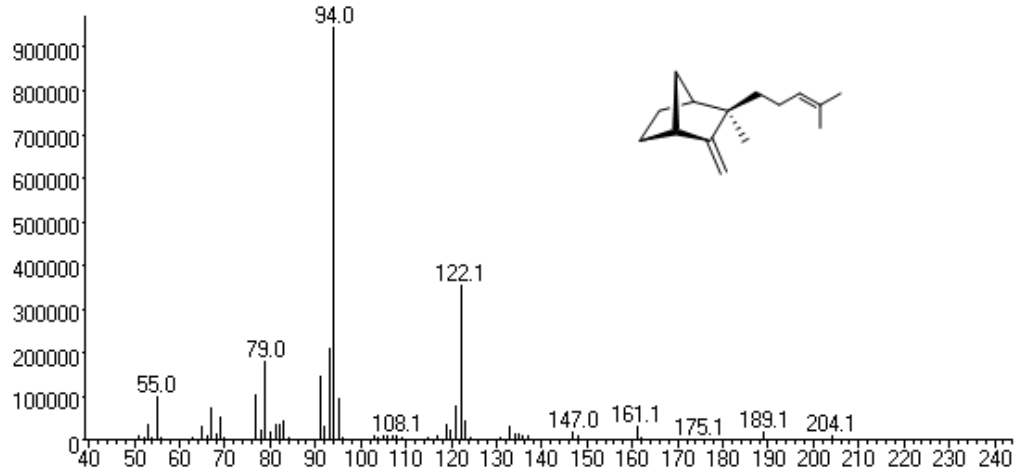


Figure 4.A.5: EI Mass Spectrum of β-Santalene (5)

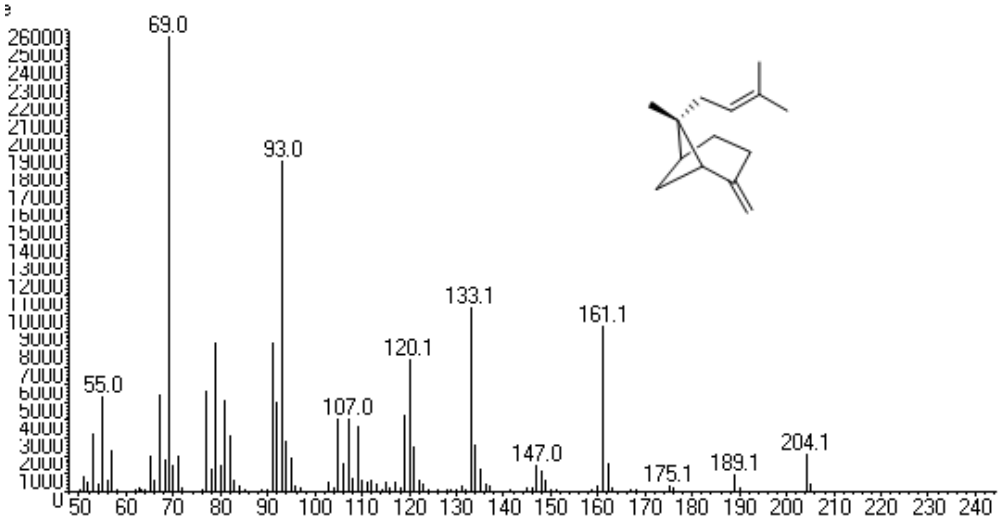


Figure 4.A.6: EI Mass Spectrum of *exo*-β-Bergamotene (**6**)

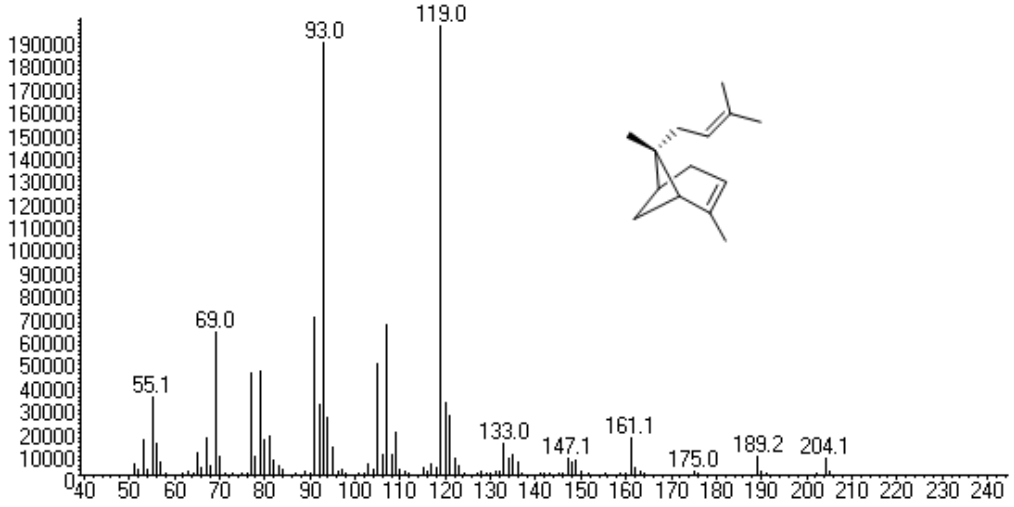


Figure 4.A.7: EI Mass Spectrum of *endo*-Bergamotene (**7**)

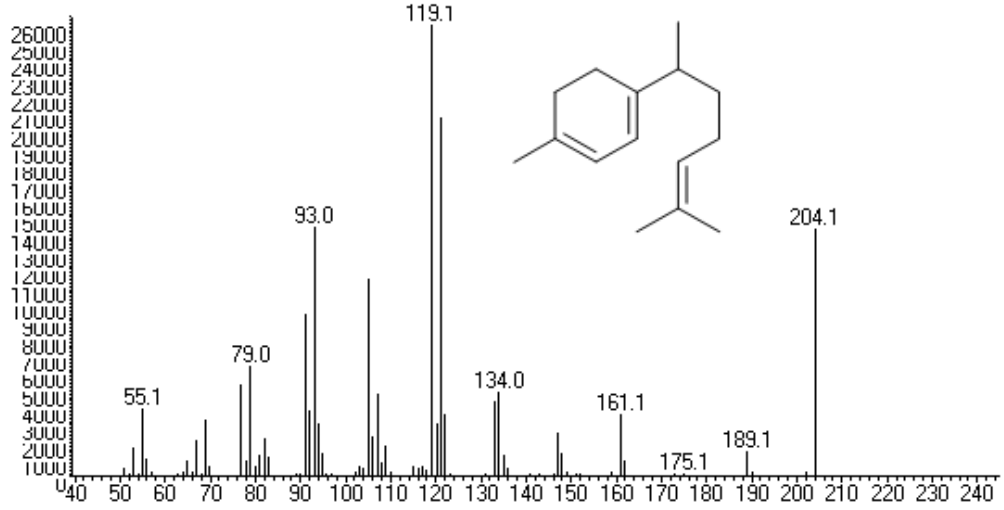


Figure 4.A.8: EI Mass Spectrum of γ-Curcumene (**8**)

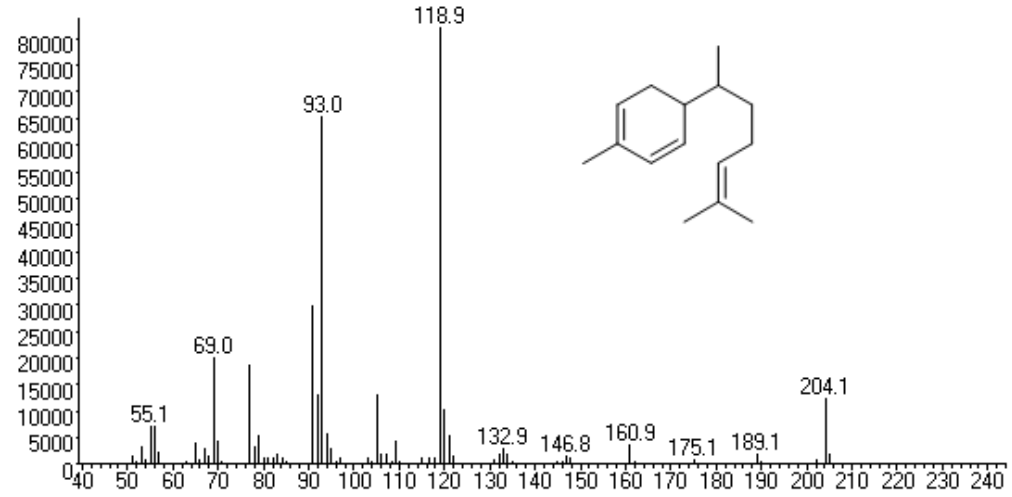


Figure 4.A.9: EI Mass Spectrum of α-Zingiberene (**9**)

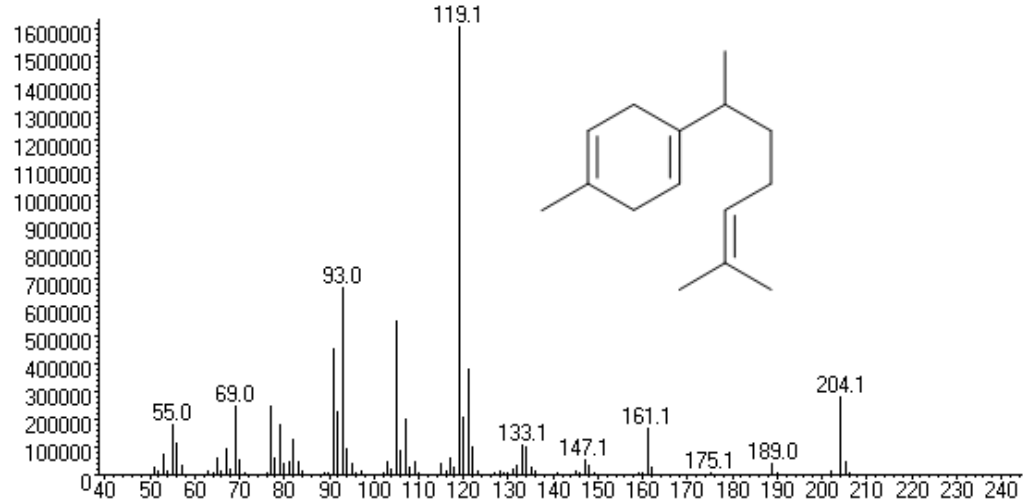


Figure 4.A.10: EI Mass Spectrum of β-Curcumene (**10**)

Appendix 4.B SDS-PAGE gel image for purified SaSS mutants

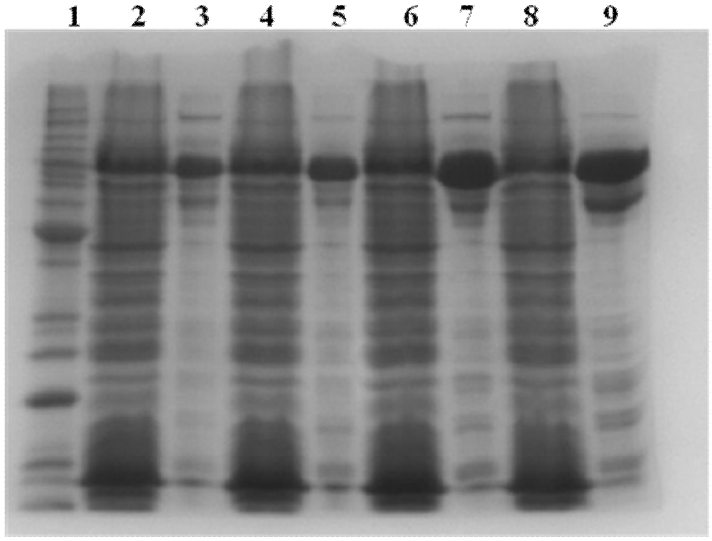


Figure 4.B.1: Protein gel image of santalene synthase mutants 1, 2, 3 and 4, **Lane 1:** Protein marker, **Lane 2:** *SaSS* mutant G323V&Y324F crude lysate, **Lane 3:** *SaSS* mutant G323V&Y324F purified, **Lane 4:** *SaSS* mutant Y324F crude lysate, **Lane 5:** *SaSS* mutant Y324F purified, **Lane 6:** *SaSS* mutant T318A crude lysate, **Lane 7:** *SaSS* mutant T318A purified, **Lane 8:** *SaSS* mutant Y539F crude lysate, **Lane 9:** *SaSS* mutant Y539F purified

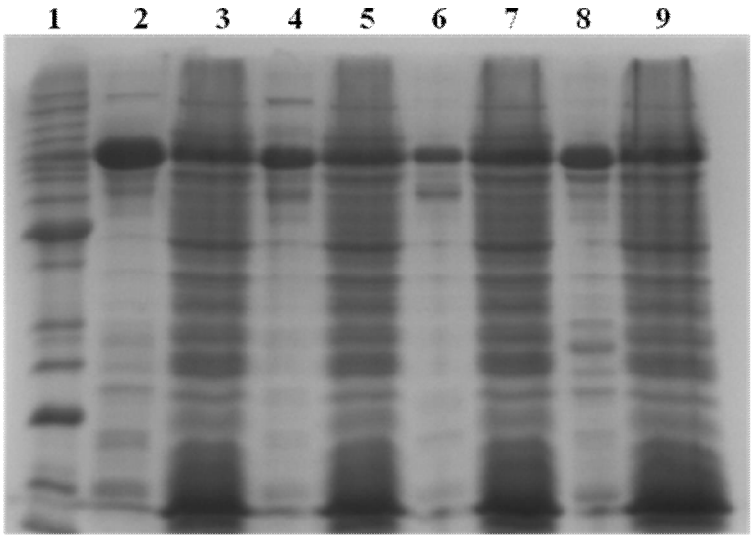


Figure 4.B.2: Protein gel image of santalene synthase mutants 5, 6, 7 and 8, **Lane 1:** Protein marker, **Lane 2:** *SaSS* mutant Y539W purified, **Lane 3:** *SaSS* mutant Y539W crude lysate **Lane 4:** *SaSS* mutant D549A purified, **Lane 5:** *SaSS mutant* D549A crude lysate, **Lane 6:** *SaSS* mutant R474M purified, **Lane 7:** *SaSS* mutant R474M crude lysate, **Lane 8:** *SaSS* mutant R474L purified, **Lane 9:** *SaSS mutant* R474L crude lysate

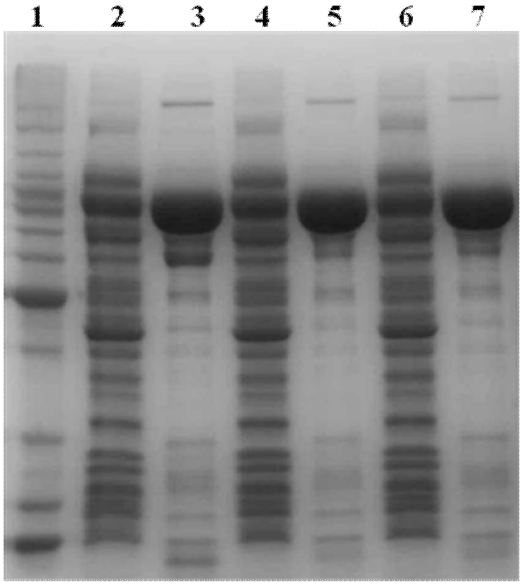


Figure 4.B.3: Protein gel image of santalene synthase mutants 10, 11 and 12, **Lane 1:** Protein marker, **Lane 2:** *SaSS* mutant W293F crude lysate, **Lane 3:** *SaSS mutant* W293F purified, **Lane 4:** *SaSS* mutant I422Acrude lysate, **Lane 5:** *SaSS mutant* I422A purified, **Lane 6:** *SaSS* mutant Y539F & T318A crude lysate, **Lane 7:** *SaSS* mutant Y539F & T318A purified

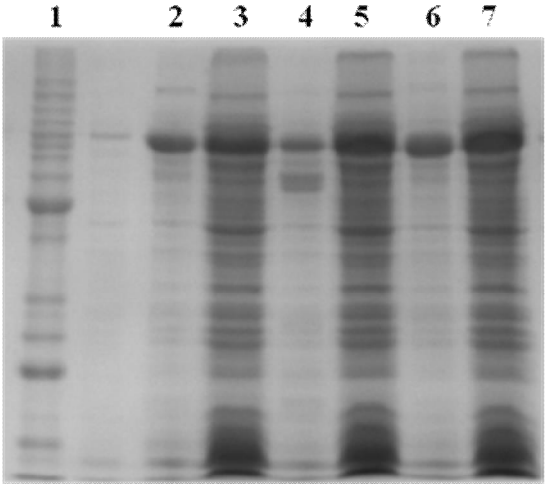


Figure 4.B.4: Protein gel image of santalene synthase mutants, **Lane 1:** Protein marker, **Lane 2:** *SaSS* mutant Y539F & I422A purified, **Lane 3:** *SaSS mutant* Y539F & I422A crude lysate, **Lane 4:** *SaSS* mutant Q289A purified, **Lane 5:** *SaSS mutant* Q289A crude lysate, **Lane 6:** *SaSS* mutant S290A purified, **Lane 7:** *SaSS* mutant S290A crude lysate

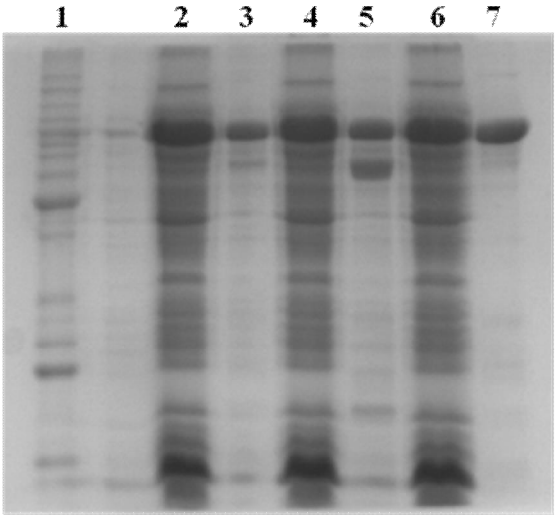


Figure 4.B.5: Protein gel image of santalene synthase mutants, **Lane 1:** Protein marker, **Lane 2:** *SaSS* mutant L427A crude lysate, **Lane 3:** *SaSS mutant* L427A purified, **Lane 4:** *SaSS* mutant T467A crude lysate, **Lane 5:** *SaSS mutant* T467A purified, **Lane 6:** *SaSS* mutant W551F crude lysate, **Lane 7:** *SaSS* mutant W551F purified

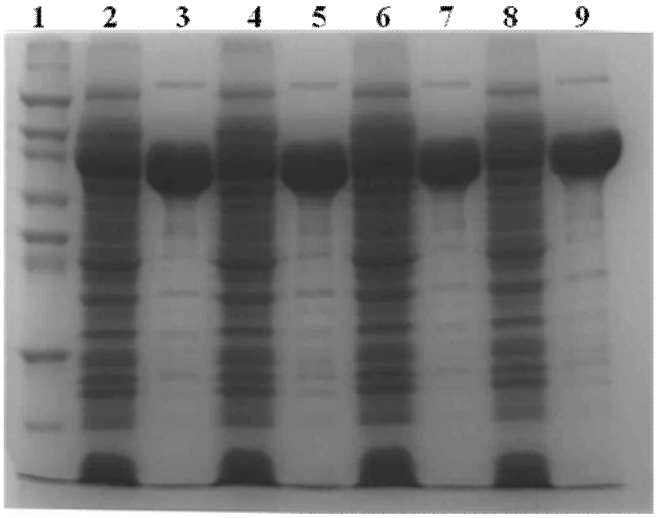


Figure 4.B.6: Protein gel image of santalene synthase mutants, **Lane 1:** Protein marker, **Lane 2:** *SaSS* mutant Y37F crude lysate, **Lane 3:** *SaSS mutant* Y37F purified, **Lane 4:** *SaSS* mutant Y396F crude lysate, **Lane 5:** *SaSS mutant* Y396F purified, **Lane 6:** *SaSS* mutant D470A crude lysate, **Lane 7:** *SaSS* mutant D470A purified, **Lane 8:** *SaSS* mutant F538Y crude lysate, **Lane 9:** *SaSS mutant* F538Y purified

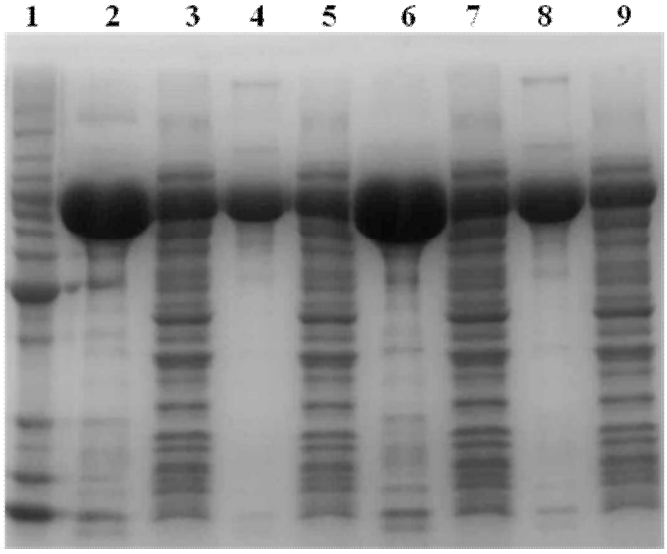


Figure 4.B.7: Protein gel image of santalene synthase mutants, **Lane 1:** Protein marker, **Lane 2:** *SaSS* mutant S290L purified, **Lane 3:** *SaSS mutant* S290L crude lysate, **Lane 4:** *SaSS* mutant S290G purified, **Lane 5:** *SaSS mutant* S290G crude lysate, **Lane 6:** *SaSS* mutant L427I purified, **Lane 7:** *SaSS* mutant L427I crude lysate, **Lane 8:** *SaSS* mutant F538M purified, **Lane 9:** *SaSS mutant* F538M crude lysate

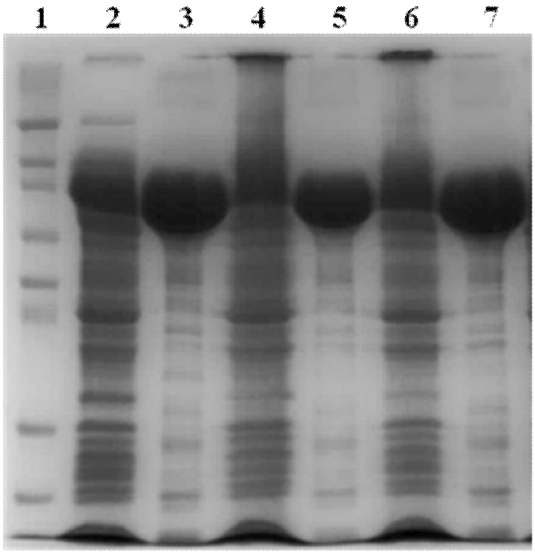


Figure 4.B.8: Protein gel image of santalene synthase mutants, **Lane 1:** Protein marker, **Lane 2:** *SaSS* mutant E471Q crude lysate, **Lane 3:** *SaSS mutant* E471Q purified, **Lane 4:** *SaSS* mutant D543N crude lysate, **Lane 5:** *SaSS mutant* D543N purified, **Lane 6:** *SaSS* mutant D543L crude lysate, **Lane 7:** *SaSS* mutant D543L purified

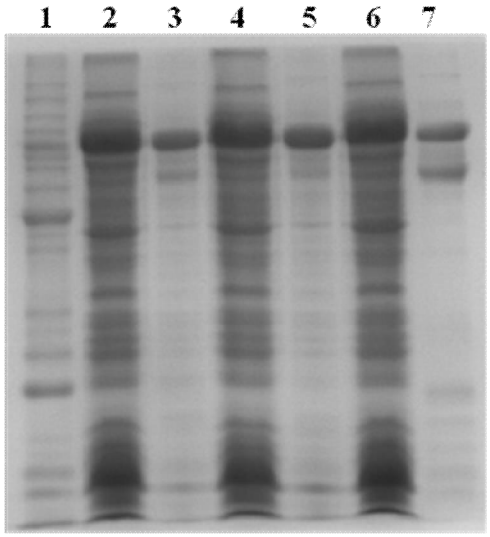


Figure 4.B.9: Protein gel of santalene synthase mutants, **Lane 1:** Protein marker, **Lane 2:** *SaSS* mutant N463D crude lysate, **Lane 3:** *SaSS mutant* N463D purified, **Lane 4:** *SaSS* mutant Y539L crude lysate, **Lane 5:** *SaSS mutant* Y539L purified, **Lane 6:** *SaSS* mutant Y539S crude lysate, **Lane 7:** *SaSS* mutant Y539S purified

Appendix 4.C: NMR Data of isolated and synthesized terpenes

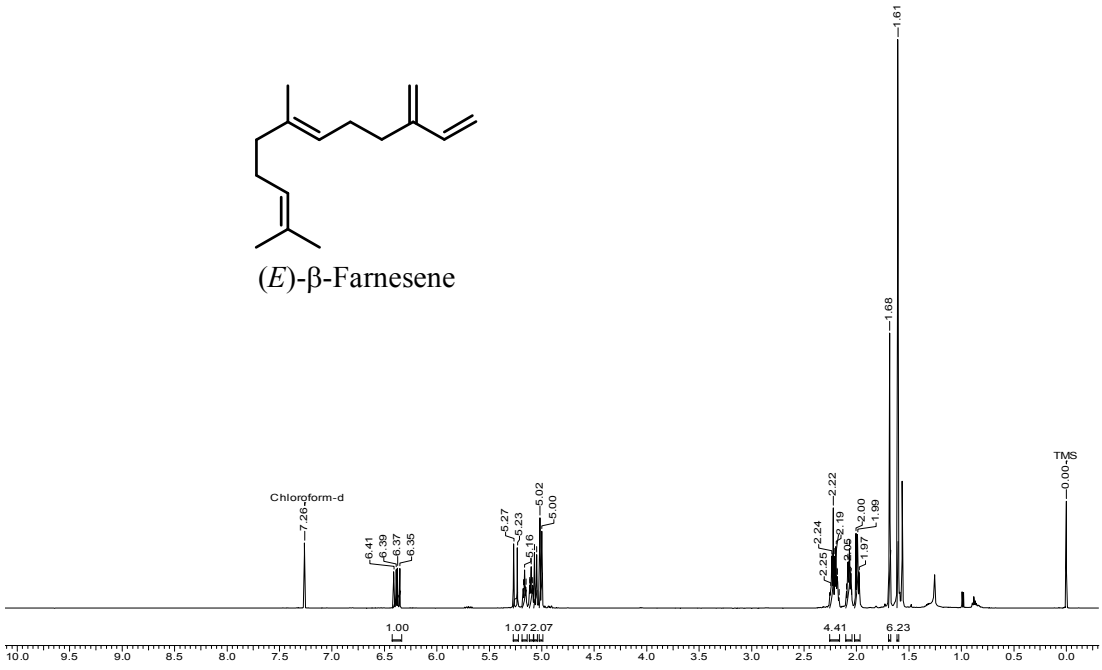


Figure 4.C.1.1: ^{1}H NMR of (*E*)-β-Farnesene (4) at 500 MHz.

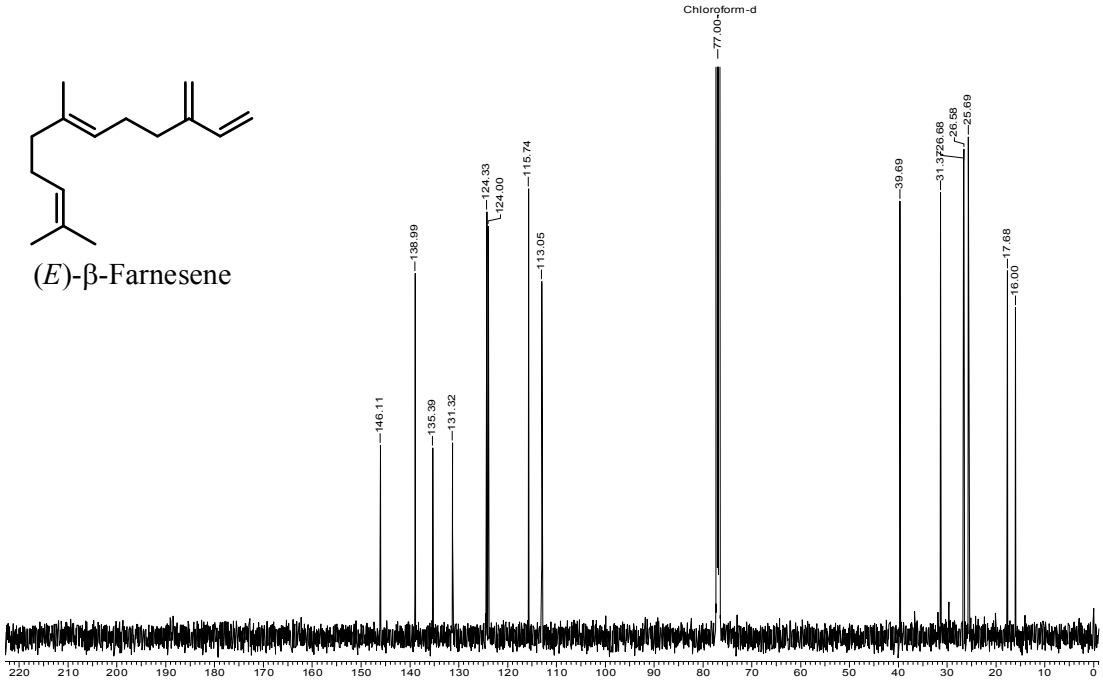


Figure 4.C.1.2: ^{13}C NMR of (*E*)-β-Farnesene (4) at 125 MHz.

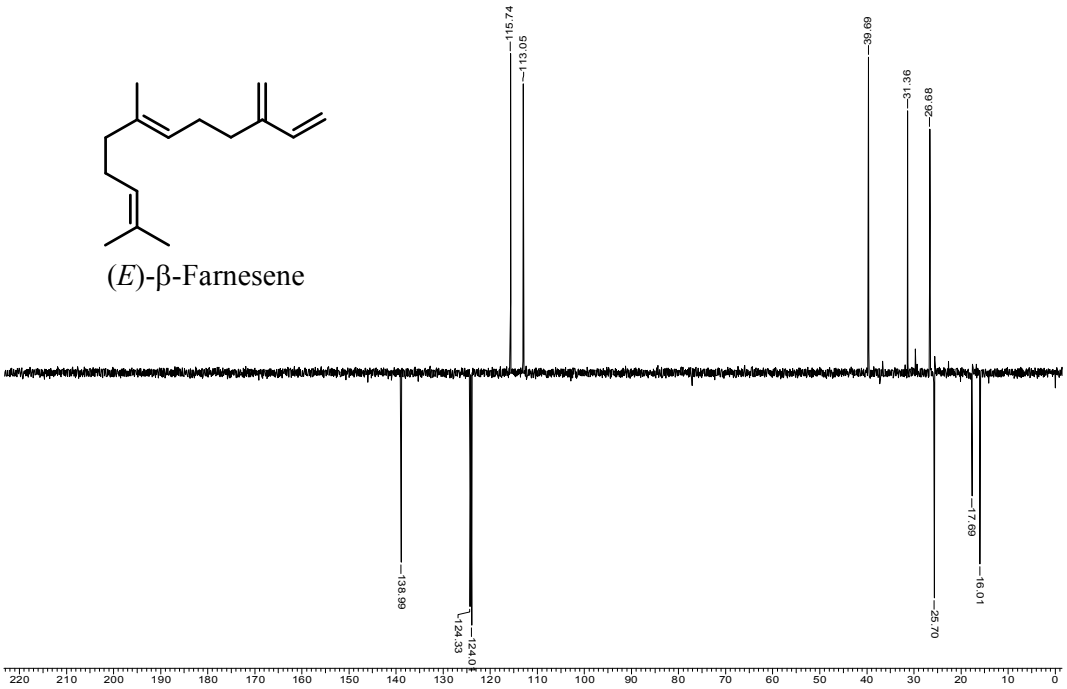


Figure 4.C.1.3: DEPT NMR of (*E*)-β-Farnesene (4) at 125 MHz.

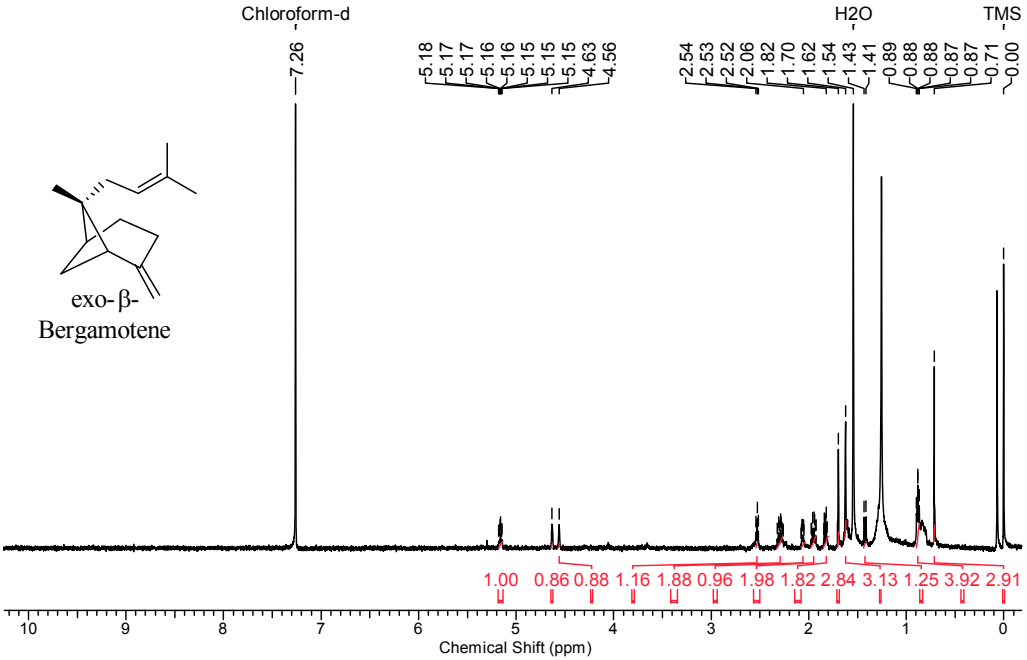


Figure 4.C.2.1a: [1]H NMR of exo-β-Bergamotene[6] (6) at 500 MHz.

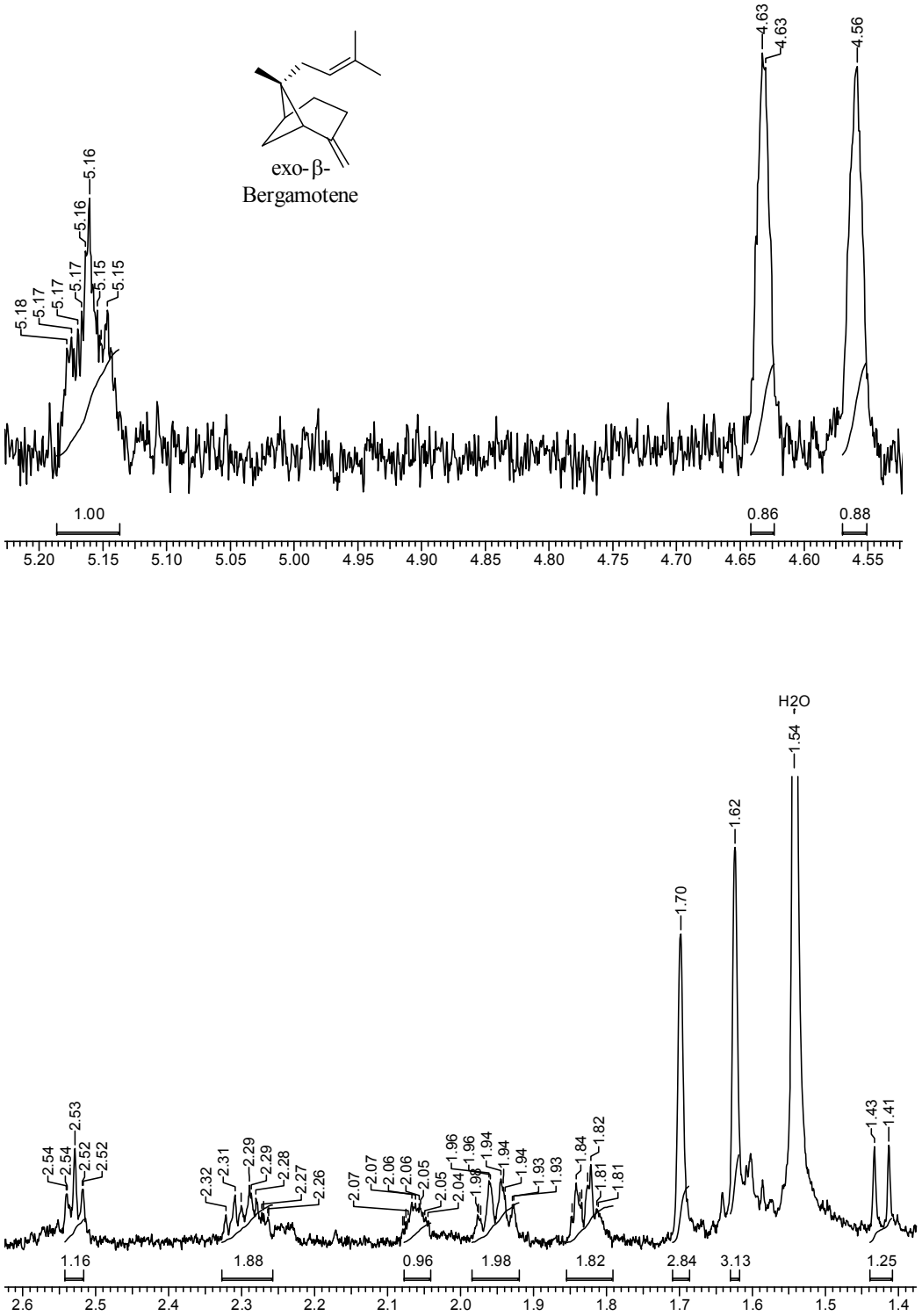


Figure 4.C.2.1b: Partial ^{1}H NMR of exo-β-Bergamotene[6] (6) at 500 MHz.

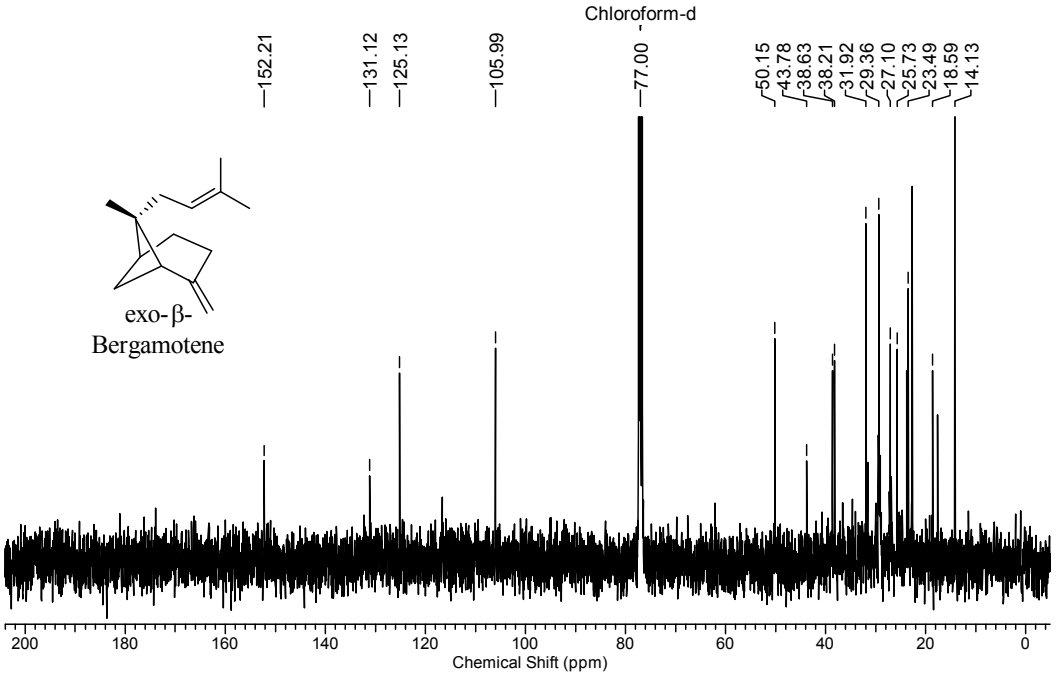


Figure 4.C.2.: ^{13}C NMR of *exo*-β-Bergamotene (6) at 125 MHz.

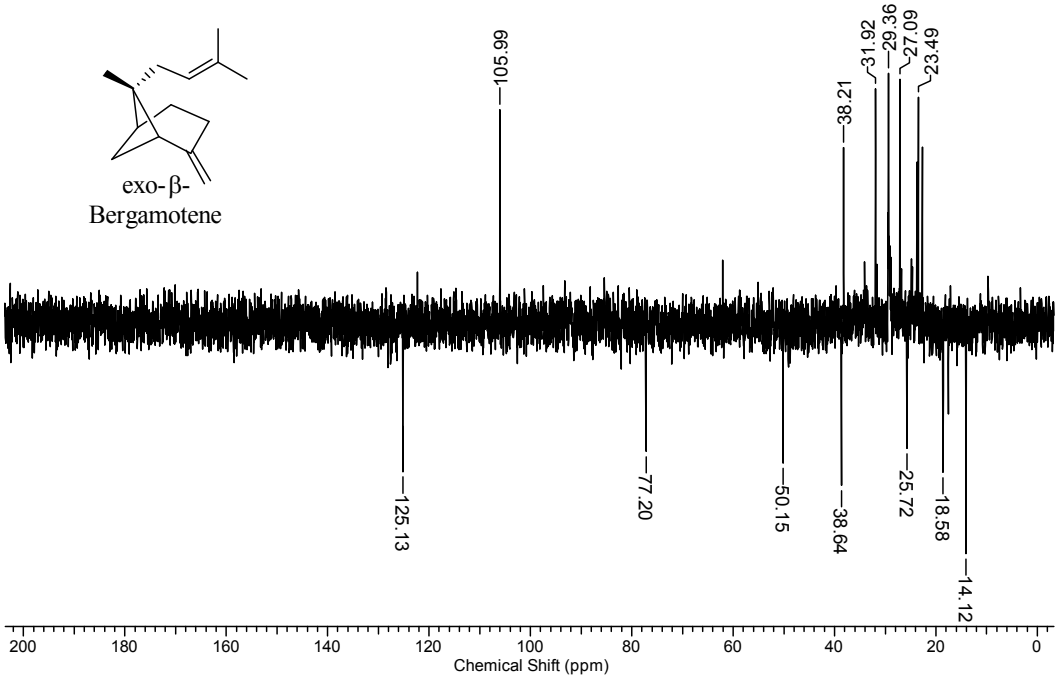


Figure 4.C.2.3: DEPT NMR of *exo*-β-Bergamotene (6) at 125 MHz.

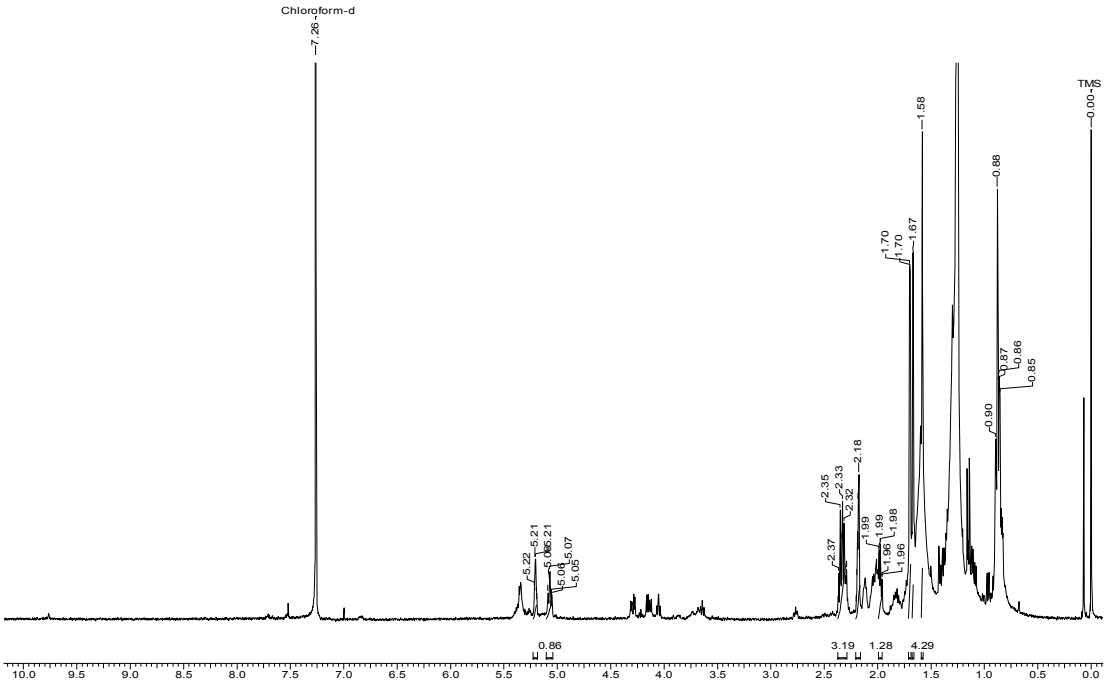


Figure 4.C.3.1a: ^{1}H NMR of *endo*-α-Bergamotene (7) at 400 MHz.[5]

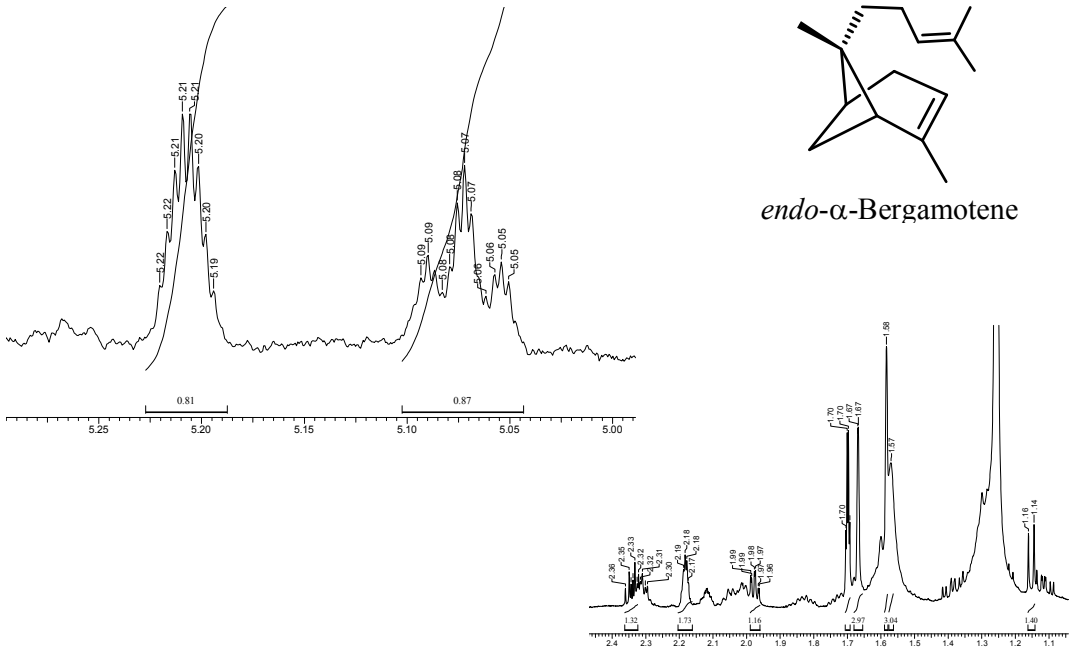
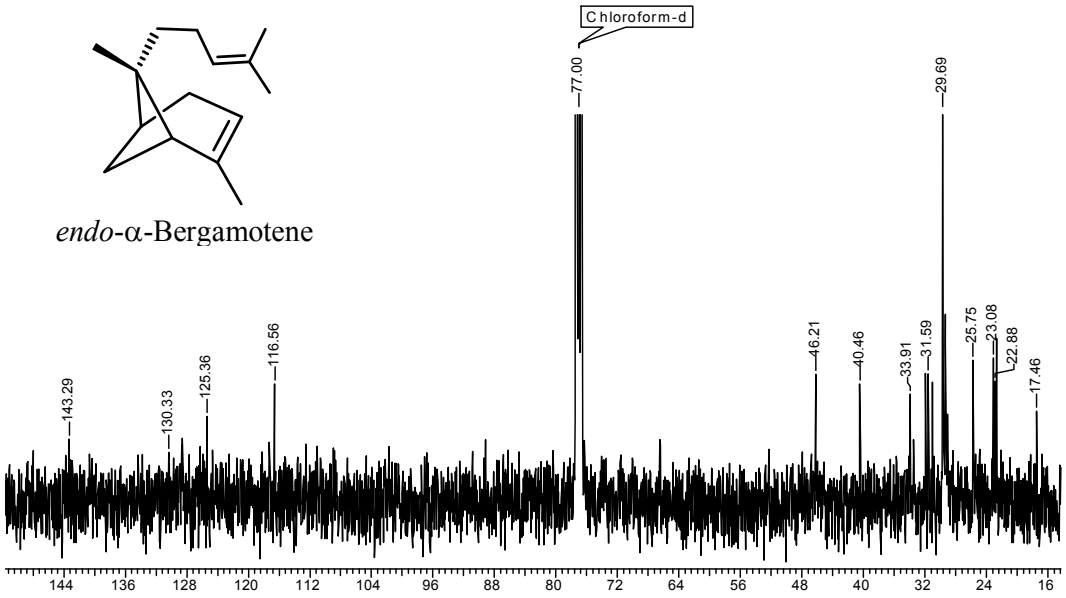
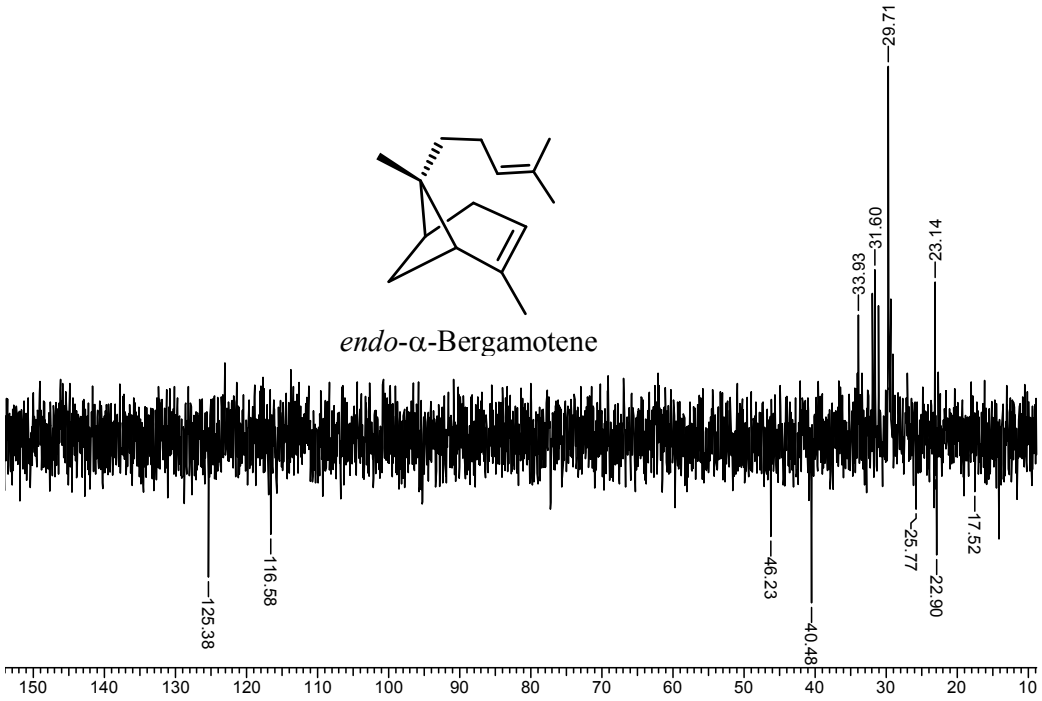


Figure 4.C.3.1b: Partial ^{1}H NMR of *endo*-α-Bergamotene (7) at 400 MHz.[5]



endo-α-Bergamotene

Figure 4.C.3.2: ^{13}C NMR of *endo*-α-Bergamotene (7) at 100 MHz.[5]



endo-α-Bergamotene

Figure 4.C.3.3: DEPT NMR of *endo*-α-Bergamotene (7) at 100 MHz.[5]

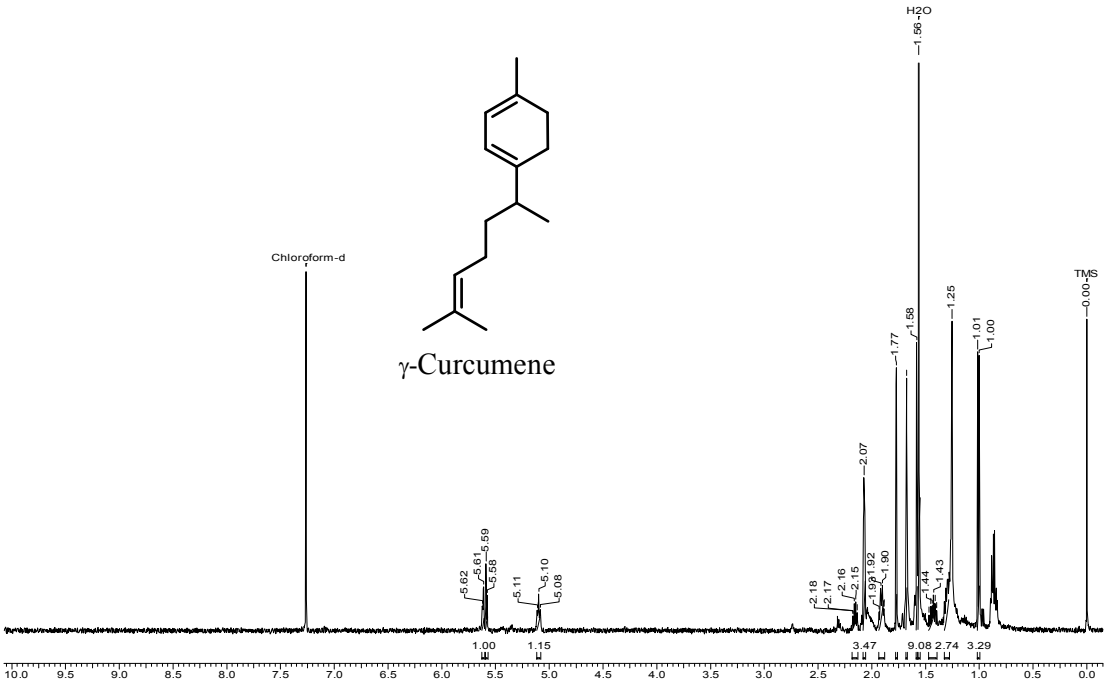


Figure 4.C.4.1: ^{1}H NMR of γ-Curcumene (8) at 500 MHz.

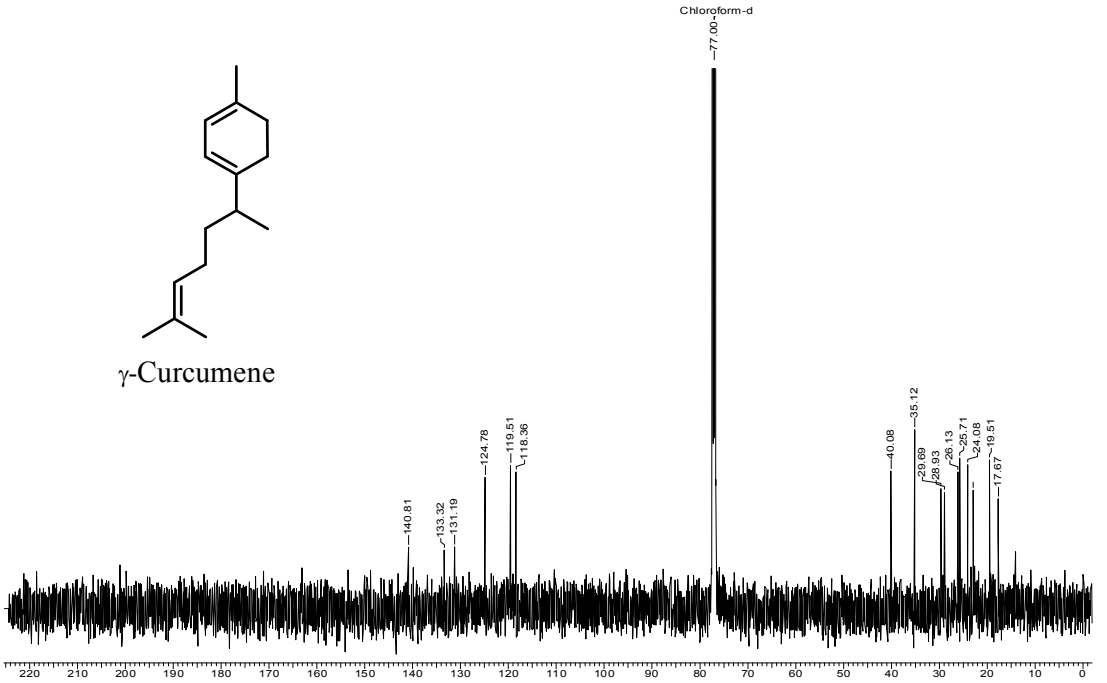


Figure 4.C.4.2: ^{13}C NMR of γ-Curcumene (8) at 125 MHz

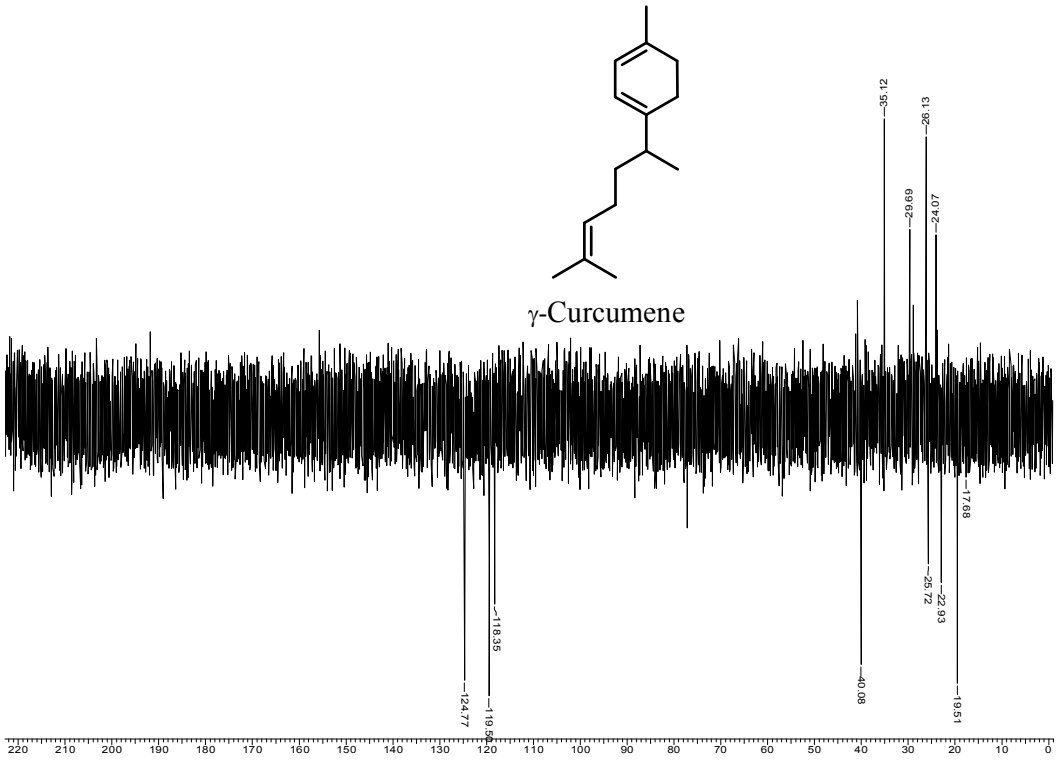


Figure 4.C.4.3: DEPT NMR of γ-Curcumene (8) at 125 MHz.

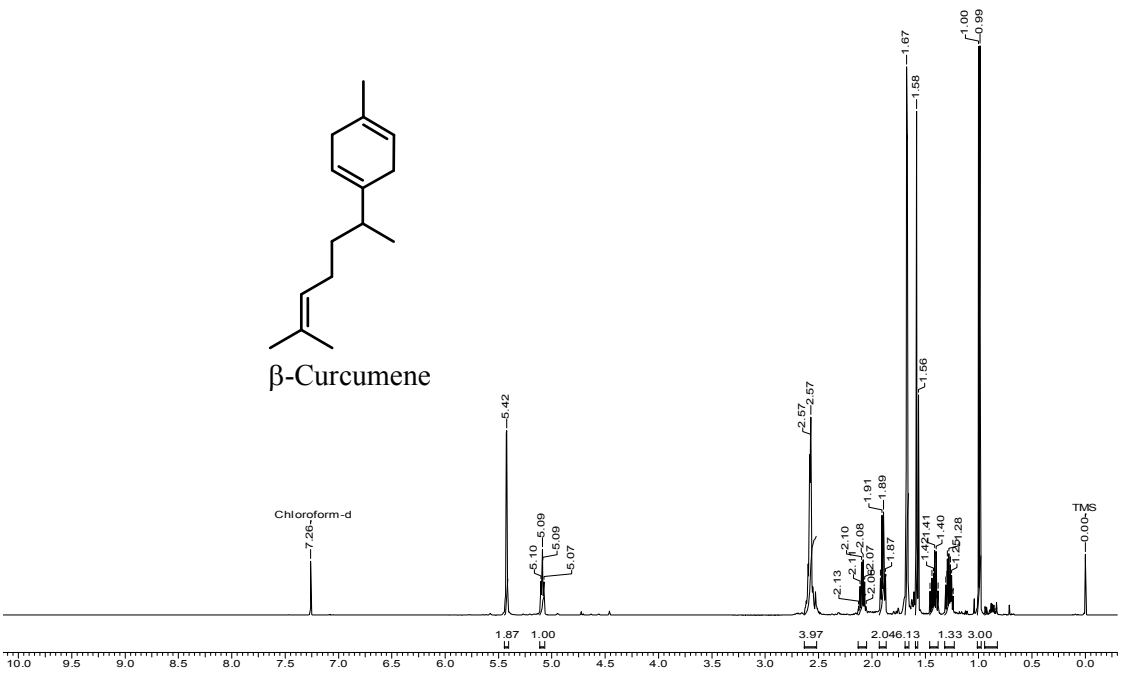


Figure 4.C.5.1: ^{1}H NMR of β-Curcumene (10) at 500 MHz.[4]

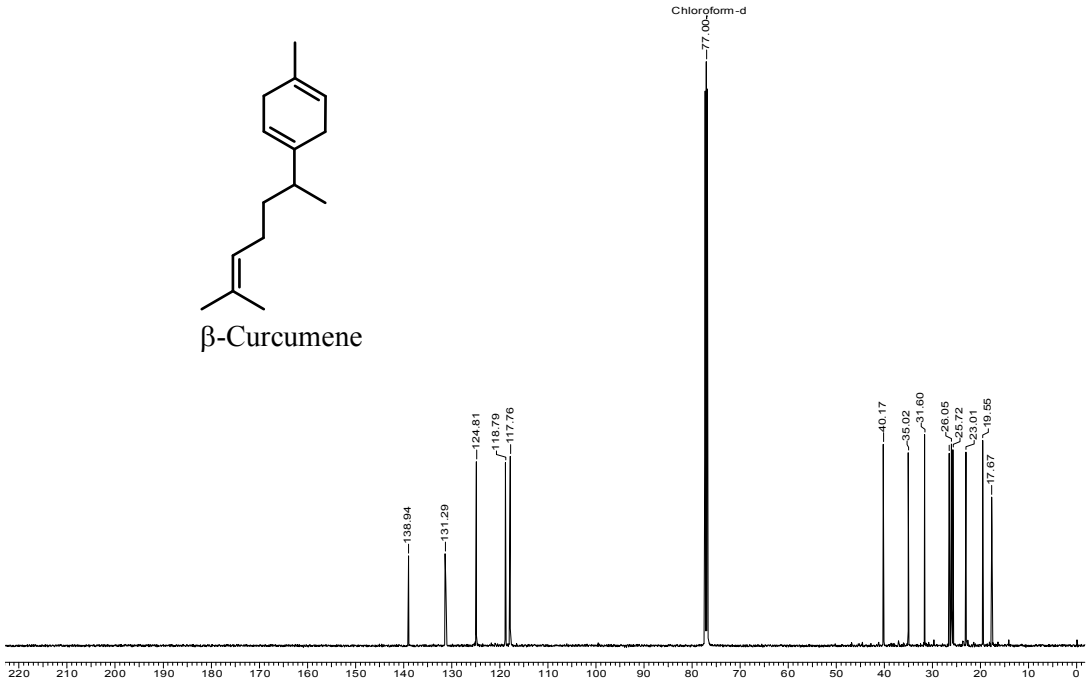


Figure 4.C.5.2: ^{13}C NMR of β-Curcumene (10) at 125 MHz.

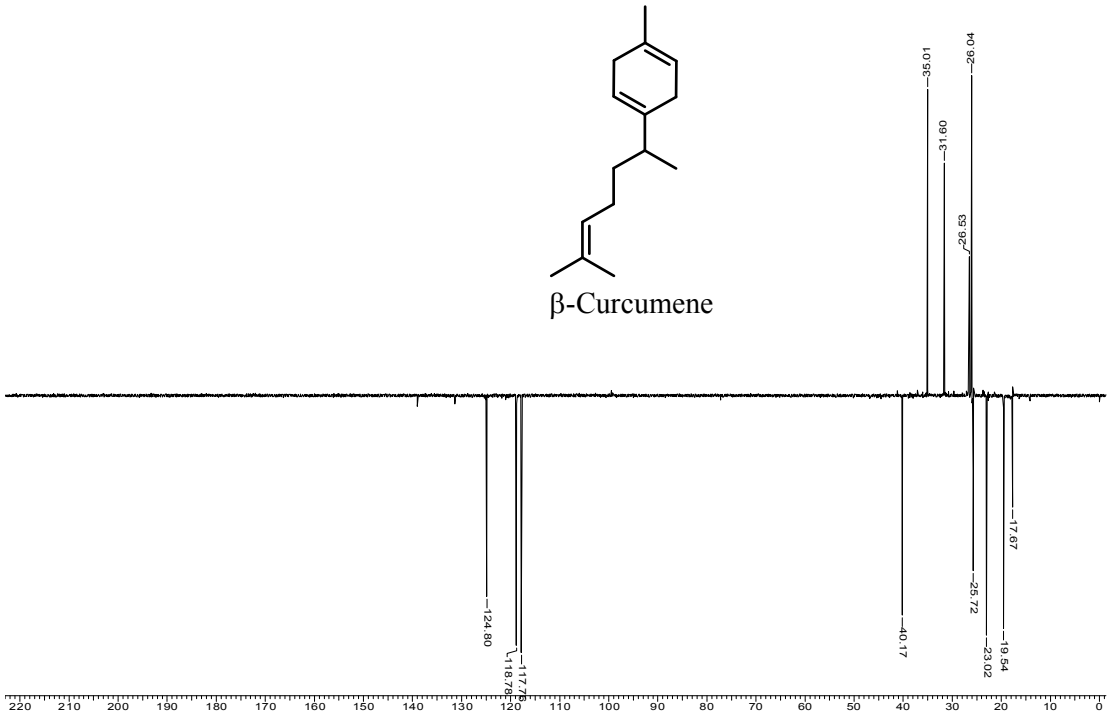


Figure 4.C.5.3: DEPT NMR of β-Curcumene (10) at 125 MHz.

4.5 References

(1) Ding, V. D. H.; Sheares, B. T.; Bergstrom, J. D.; Ponpipom, M. M.; Perez, L. B.; Poulter, C. D. *Biochem. J.* **1991**, *275*, 61-65.

(2) Anderson, M. S.; Yarger, J. G.; Burck, C. L.; Poulter, C. D. *J. Biol. Chem.* **1989**, *264*, 19176-19184.

(3) Peralta-Yahya, P. P.; Ouellet, M.; Chan, R.; Mukhopadhyay, A.; Keasling, J. D.; Lee, T. S. *Nat. Commun.* **2011**, *2*.

(4) Rude, M. A.; Schirmer, A. *Curr. Opin. Microbiol.* **2009**, *12*, 274-281.

(5) Thulasiram, H. V.; Erickson, H. K.; Poulter, C. D. *Science* **2007**, *316*, 73-76.

(6) Croteau, R.; Satterwhite, D. M.; Wheeler, C. J.; Felton, N. M. *J. Biol. Chem.* **1989**, *264*, 2075-80.

(7) Lesburg, C. A.; Zhai, G. Z.; Cane, D. E.; Christianson, D. W. *Science* **1997**, *277*, 1820-1824.

(8) Starks, C. M.; Back, K. W.; Chappell, J.; Noel, J. P. *Science* **1997**, *277*, 1815-1820.

(9) Caruthers, J. M.; Kang, I.; Rynkiewicz, M. J.; Cane, D. E.; Christianson, D. W. *J Biol. Chem.* **2000**, *275*, 25533-25539.

(10) Rynkiewicz, M. J.; Cane, D. E.; Christianson, D. W. *P. Natl. Acad. Sci. USA* **2001**, *98*, 13543-13548.

(11) Cane, D. E. *Chem. Rev.* **1990**, *90*, 1089-1103.

(12) Cane, D. E. *Accounts Chem. Res.* **1985**, *18*, 220-226.

(13) Bone, R.; Silen, J. L.; Agard, D. A. *Nature* **1989**, *339*, 191-195.

(14) Aharoni, A.; Gaidukov, L.; Khersonsky, O.; Gould, S. M.; Roodveldt, C.; Tawfik, D. S. *Nat. Genet.* **2005**, *37*, 73-76.

(15) Khersonsky, O.; Tawfik, D. S. In *Annu. Rev. Biochem.* 2010; Vol. 79, p 471-505.

(16) van den Heuvel, R. H. H.; Fraaije, M. W.; Ferrer, M.; Mattevi, A.; van Berkel, W. J. H. *P. Natl. Acad. Sci. USA* **2000**, *97*, 9455-9460.

(17) Rising, K. A.; Starks, C. M.; Noel, J. P.; Chappell, J. *J. Am. Chem. Soc.* **2000**, *122*, 1861-1866.

(18) Yoshikuni, Y.; Martin, V. J. J.; Ferrin, T. E.; Keasling, J. D. *Chem. Biol.* **2006**, *13*, 91-98.

(19) Yoshikuni, Y.; Ferrin, T. E.; Keasling, J. D. *Nature* **2006**, *440*, 1078-1082.

(20) Christianson, D. W. *Chem. Rev.* **2006**, *106*, 3412-3442.

(21) Cane, D. E.; Xue, Q.; Fitzsimons, B. C. *Biochemistry* **1996**, *35*, 12369-12376.

(22) Gennadios, H. A.; Gonzalez, V.; Di Costanzo, L.; Li, A.; Yu, F.; Miller, D. J.; Allemann, R. K.; Christianson, D. W. *Biochemistry* **2009**, *48*, 6175-6183.

(23) Jones, C. G.; Moniodis, J.; Zulak, K. G.; Scaffidi, A.; Plummer, J. A.; Ghisalberti, E. L.; Barbour, E. L.; Bohlmann, J. *J. Biol. Chem.* **2011**, *286*, 17445-17454.

(24) Daramwar, P. P.; Srivastava, P. L.; Priyadarshini, B.; Thulasiram, H. V. *Analyst* **2012**, *137*, 4564-4570.

(25) Kollner, T. G.; O'Maille, P. E.; Gatto, N.; Boland, W.; Gershenzon, J.; Degenhardt, J. *Arch. Biochem. Biophys.* **2006**, *448*, 83-92.

(26) Snider, B. B.; Beal, R. B. *J. Org. Chem.* **1988**, *53*, 4508-4515.

(27) Sallaud, C.; Rontein, D.; Onillon, S.; Jabes, F.; Duffe, P.; Giacalone, C.; Thoraval, S.; Escoffier, C.; Herbette, G.; Leonhardt, N.; Causse, M.; Tissier, A. *Plant Cell* **2009**, *21*, 301-317.

(28) Peters, R. J.; Croteau, R. B. *P. Natl. Acad. Sci. USA* **2002**, *99*, 580-584.

(29) Little, D. B.; Croteau, R. B. *Arch. Biochem. Biophys.* **2002**, *402*, 120-135.

(30) Zilenovski, J. S. R.; Hall, S. S. *Synthesis-Stuttgart* **1979**, 698-699.

Chapter 5

Screening, isolation and cloning of Cytochrome P450 system from Indian Sandalwood *S. album*

Chapter 5

Screening, isolation and cloning of Cytochrome P450 systems from Indian Sandalwood *S. album*

α- and β-santalols are the most important components present in the essential oil of Sandalwood and highly valued because of their desirable fragrance and biological properties. These compounds are formed through hydroxylation at *cis* methyl group of the corresponding sesquiterpene hydrocarbons α-and β-santalenes catalyzed by monooxygenase system, cytochrome P450 (CYP450). Here, we present the screening and isolation of CYP450 system belonging to CYP71D subfamily.

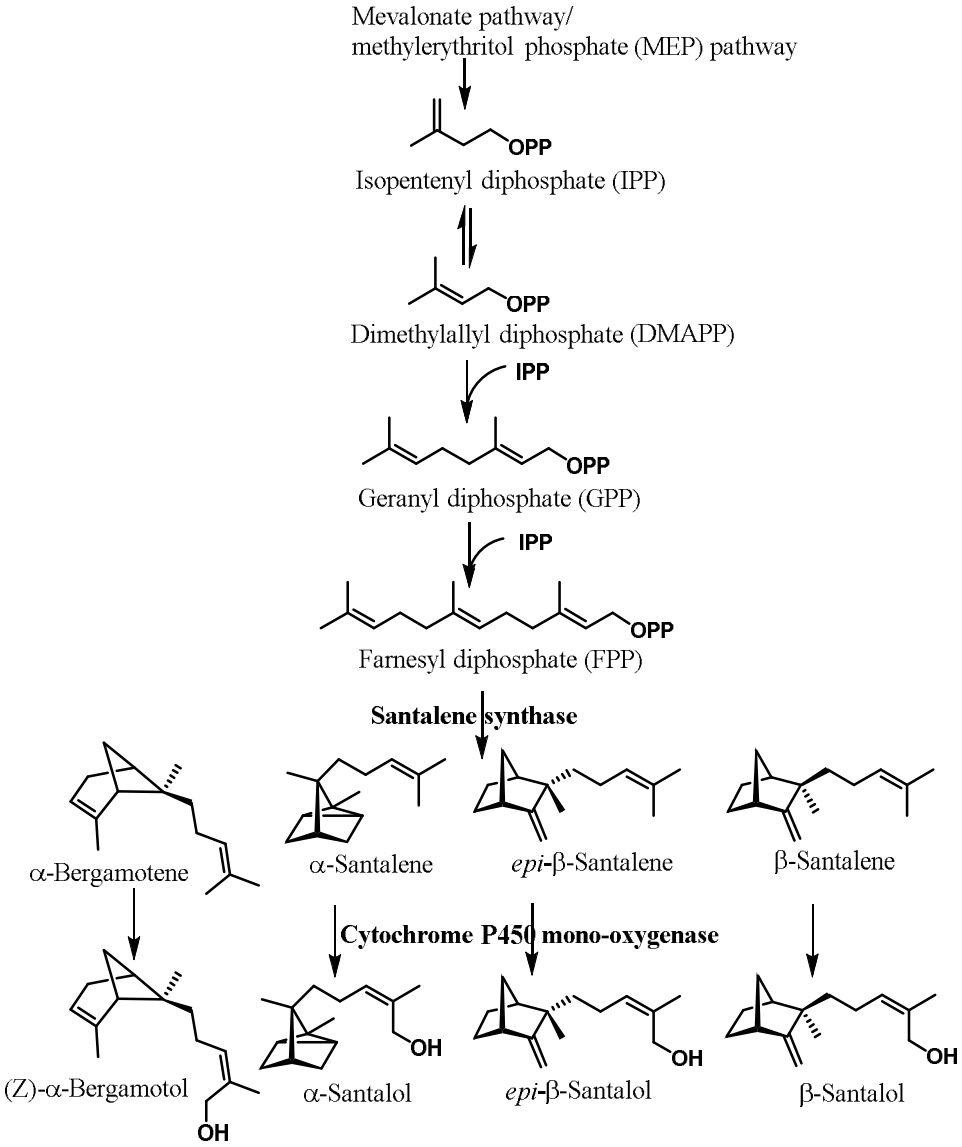
5.1 Introduction

Sandalwood oil is the world's most prized essential oil, commercially produced by steam distillation of heartwood chip with varying yields from 4-7% from a well matured tree (>80 years old tree). α, β, *epi*-β-santalols and exo-α-bergamotol represent the major components (~90%) of sandalwood essential oil, from which α- and β santalols are the most important components responsible for most of the biological activity[1-3]. Since the demand for this essential oil is increasing, overharvesting and several plant diseases have led to a serious decline in population of Sandalwood tree. Although total synthesis of α- and β santalols has been attempted[4,5] but due to structure complexity and low yield, the conventional organic synthesis of these molecules becomes uneconomical. An alternate approach for the production of these essential components of sandalwood oil is engineering a microbial system by incorporating the biosynthetic pathway involved in biosynthesis of santalols. The first committed step in santalol biosynthesis is generation of FPP by head to tail condensation of IPP and DMAPP catalyzed by FPP synthase. The linear FPP undergoes cyclization by santalene synthase into a mixture of α-, β-santalenes, exo-α-bergamotene, which is further hydroxylated at *cis* methyl position by CYP450 mono-oxygenase system to generate α-, β-santalols and exo-α-bergamotol[6-8]. Cytochrome P450 mono-oxygenases are a large group of enzymes which contain a characteristic heme binding motif and play an important role in secondary metabolism pathway such as: terpenoid pathway[9,10], phenylpropanoid pathway[11,12], hormone biosynthesis[13] and detoxification of xenobiotics[14,15] and are ubiquitously distributed in all biological kingdoms. The general reaction catalyzed by CYP450 systems are represented as below[16]:

$$\mathbf{RH} + O_2 + NAD(P)H + H^+ \quad = \quad \mathbf{ROH} + H_2O + NAD(P)^+$$

The terminology of CYPP450 is described as: CYP for cytochrome, P stand for pigments and 450 reflects the spectral properties having a strong peak at the 450 nm wavelength in reduced and complexed state with carbon mono-oxide[17]. Depending on the electron supporting systems, CYP450s are divided into ten subclasses, of which bacterial/mitochondrial type and microsomal type are two main classes. Microsomal CYP450 are membrane bound and accept electrons from microsomal NADPH-CYP450 reductase system. The CYP450 reductase belongs to diflavin protein family and contains two conserved domains for binding FAD (C-terminal region) and FMN

(N-terminal region), which transfer the required electrons to CYP450 systems. Depending on the species, each plant contains 1-3 paralogs of CYP450 reductase. CYP reductase from different plant species or from different kingdoms can complement each other functionally. The reaction mechanism for santalol/bergamotol biosynthesis in Sandalwood is proposed in scheme 5.1.



Scheme 5.1: Proposed pathway for the biosynthesis of santalols and bergamotol in Indian Sandalwood, *S. album*

In this chapter, we present the screening, isolation and cloning of CYP450 of CYP71D55 subfamily, which might be involved in hydroxylation of santalenes and bergamotene at the *cis* methyl group of the side chain.

5.2 Materials and methods

5.2.1 Bacterial strains and plasmids used in the study

Escherichia coli TOP10F (Invitrogen)

TA cloning vector (Invitrogen)

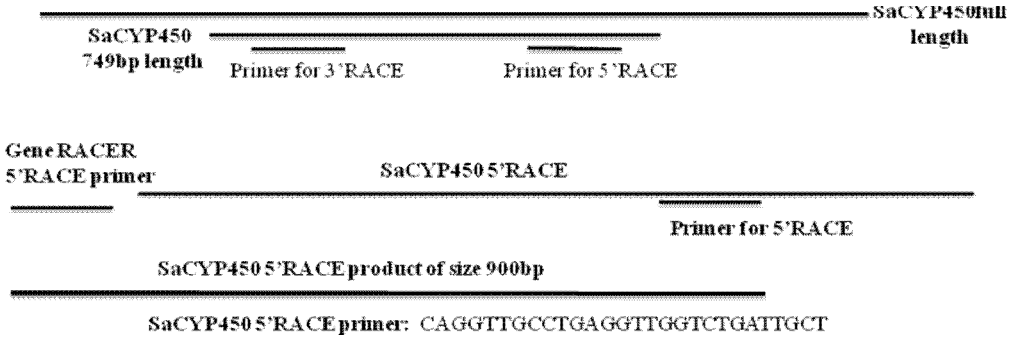
pESC-Duet yeast expression vector

5.2.2 RNA isolation and cDNA preparation

As discussed in chapter 2, total RNA was isolated from the interface of heartwood and sapwood. 1µg of total RNA was used for first strand cDNA synthesis using oligo-$(dT)_{18}$ primer and AMV-RT (Promega). The prepared cDNA was stored in -20 °C till further use.

5.2.3 RACE PCR for CYP450 mono-oxygenase (>Locus_41531)

In order to extend EST fragment (>Locus_41531) of CYP450 mono-oxygenase to full length cDNA sequence, 5′ RACE was performed using GeneRacer 5′ universal primer (CGACTGGAGCACGAGGACACTGA) and 5′ RACE primer of locus_41531 (GGAGGGGGTGCCCGGGTTCTTCCTTCG). RACE PCR was performed using the PCR program: 94 °C for 5 minutes, followed by 35 cycles of 94 °C for 20 seconds, 60 °C for 20 seconds, 72 °C for 1 minute, and final extension at 72 °C for 10 minutes. The expected size of 5′ RACE amplicon was purified from 1 % agarose gel and ligated in TA cloning vector. Positive clones were sequenced using M13 forward and reverse primers to generate RACE sequence.



Scheme 5.2.1: Schematic representation of 5′ RACE for transcript (>Locus_41531) matching with CYP450 mono-oxygenase

5.2.4 Primer sequence for isolation of full length ORF sequence of SaCYP450 mono-oxygenase system

Primer name	5'-3' Primer sequence
SaCYP450_F	**GTCGAC**ATGGACGAATTTACCCTTCTAGCCG
SaCYP450_R	**GGTACC**TTACAGGTGCTTTGGGACAAGTAT

5.2.5 Isolation and cloning of ORF of SaCYP450 in pESC duet vector MCS 2

ORF of SaCYP450 mono-oxygenase was isolated from cDNA using full length ORF primers having RE site at both the ends (SalI at 5' and KpnI at 3') with Proof reading taq DNA polymerase (Invitrogen) using the PCR program: 94 °C for 5 min, followed by 35 cycles of 94 °C for 20 sec, 62-64 °C for 20 sec, 68 °C for 1 min 30 sec followed by a final extension at 68 °C for 10 min. PCR amplicon was purified from 1 % agarose gel using gel extraction kit and subjected to restriction digestion with SalI and KpnI along with MCS 2 of pESC yeast-Duet vector (containing CYP450 reductase from *Azadirachta indica* in MCS 1)[18]. Digested fragments were purified from agarose gel and ligated using T_4 DNA ligase (Invitrogen) at 14 ºC for 24 hours. The ligation mixture was transformed in TOP10 chemically competent cells, plated on LA containing 100 µg/mL of ampicillin and incubated overnight at 37 °C. Colony PCR was performed with Gal 1 promoter primer and SaCYP450 reverse primer to screen for positive clones. Plasmids were isolated from 5 positive clones of colony PCR and sequenced with Gal 1 forward and reverse primers to check for the sequence in the vector frame.

5.2.6 Yeast expression and microsome preparation (SaCYP450)

Expression of active protein was carried out in INVSc1 yeast competent cells. Cells were grown overnight in synthetic complete medium without Uracil (SC-U), containing 2 % glucose at 30 °C, then transferred to induction medium (SC-U, containing 20 % galactose) and further incubated at 30 °C for 12 hours. The cells were centrifuged at 3000 × *g* for 10 minutes at 4 °C. The cell pellet obtained was washed with TEK buffer (50 mM Tris-HCl, 1 mM EDTA, pH 7.4, 100mM KCl) (1 mL/g of cell pellet X 3) and centrifuged. The cell pellet (1 g/5 mL) was re-suspended in 50 mM Tris-HCl buffer (containing 1 mM EDTA, 600 mM Sorbitol, 5 mM DTT, 0.25 mM PMSF and pH 7.4) and cells were lysed using a bead-beater (with acid-

washed glass beads, 425-600 µm) for 6 cycles (pulse on 30 sec, pulse off 30 sec, manual rocking for 3 × 30 sec). The lysed cells were centrifuged at 1000 × *g* for 5 min at 4 °C to remove the glass beads. Further, the supernatant was subjected to centrifugation at 10,000 × *g* for 30 min at 4 °C. The 10,000 × *g* supernatant was centrifuged at 1,00,000 × *g* for 1 hr 30 min at 4 °C. The microsomal pellet, thus obtained was suspended in TEG buffer (50 mM Tris-HCl, 1 mM EDTA, 30% glycerol, pH 7.5) and homogenized. The homogenized microsome fraction was aliquoted (0.2 mL), flash-frozen in liquid nitrogen and stored at –80 °C.

5.2.7 Enzymatic characterization of SaCYP450 mono-oxygenase system

Microsomal fraction (1 mg protein) was incubated with santalenes' mixture in 0.5 ml Tris-HCl buffer (20 mM Tris-HCl, pH 7.6, 1 mM EDTA, 1 mM DTT), containing 1 mM NADPH, 2.5 µM Glucose 6-phosphate, 1U Glucose 6-phosphate dehydrogenase, 10 µM FAD, 10 µM FMN and incubated at 30 °C for 1 hour. The mixture was then extracted twice with dichloromethane (CH_2Cl_2). The extracted volume was reduced to ~50 µL with a stream of dry nitrogen. Products were characterized by injecting 1 µL of concentrated sample in GC/GC-MS equipped with a 30 m × 0.32 mm × 0.30 µm capillary column (HP-5, J & W Scientific) and a FID. The column was equilibrated at 70 °C followed by a temperature gradient from 70 °C to 170 °C at 5 °C/min, followed by a second temperature gradient of 15 °C/min from 170 °C to 180 °C and a final hold at 180 ºC for 5 min, Nitrogen was the carrier gas with a flow rate of 1 mL/min.

5.3 Results and discussion

5.3.1 Screening, isolation and cloning of SaCYP450 system from *S. album*

5.3.1.1 Transcriptome screening for SaCYP450 system (CYP450_1)

From the transcriptome screening as discussed in chapter 2, ten unigenes were identified in the pool of unigenes (as shown in table 5.3.1.1) belonging to CYP450 71D, CYP450 76B and CYP450 93E subfamilies which may involved in hydroxylation of santalene and its derivative for the biosynthesis of α-, β-santalols and exo-α-bergamotol.

Table 5.3.1: Unigenes selected from transcriptome analysis having CYP450 domain.

Sr. No.	Transcript ID	Status	Blast Result	Maximum Identity
1	>Locus_41531_Transcript_1/1_Confidence_1.000_Length_749	5′ sequence is missing	RecName: Full=**Premnaspirodiene oxygenase**; Short=HPO; AltName: Full=Cytochrome P450 71D55	53 %
2	>Locus_7513_Transcript_1/1_Confidence_1.000_Length_694bp	5′ sequence is missing	**Premnaspirodiene oxygenase**; Short=HPO; AltName: Full=Cytochrome P450 71D55	46 %
3	>Locus_24201_Transcript_1/1_Confidence_1.000_Length_708	Both the ends are missing	RecName: Full=**Premnaspirodiene oxygenase**; Short=HPO; AltName: Full=Cytochrome P450 71D55	49 %
4	>Locus_28896_Transcript_1/1_Confidence_1.000_Length_814	5′ sequence is missing	RecName: Full=**Premnaspirodiene oxygenase**; Short=HPO; AltName: Full=Cytochrome P450 71D55	64 %
5	>Locus_22546_Transcript_1/1_Confidence_1.000_Length_1184	Both the ends are missing	RecName: Full=**Geraniol 8-hydroxylas**e; AltName: Full=Cytochrome P450 76B10; AltName: Full=Geraniol 10-hydroxylase; Short=SmG10H	50 %
6	>Locus_3853_Transcript_1/1_Confidence_1.000_Length_790	3′ sequence is missing	RecName: Full=**Premnaspirodiene oxygenase**; Short=HPO; AltName: Full=Cytochrome	41 %

			P450 71D55	
7	>Locus_9117_Transcript_1/1_Confidence_1.000_Length_782	5′ sequence is missing	RecName: Full=**Geraniol 8-hydroxylase**; AltName: Full=Cytochrome P450 76B10; AltName: Full=Geraniol 10-hydroxylase; Short=SmG10H	44 %
8	>Locus_7653_Transcript_1/1_Confidence_1.000_Length_610	Both the ends are missing	RecName: Full=**Geraniol 8-hydroxylase**; AltName: Full=Cytochrome P450 76B10; AltName: Full=Geraniol 10-hydroxylase; Short=SmG10H	56 %
9	>Locus_36009_Transcript_1/1_Confidence_1.000_Length_611	Both the ends are missing	RecName: Full=**Taxane 13-alpha-hydroxylase**; AltName: Full=Cytochrome P450 725A2	35 %
10	>Locus_55002_Transcript_1/1_Confidence_1.000_Length_881	Both the ends are missing	RecName: Full=**Beta-amyrin 24-hydroxylase**; AltName: Full=Cytochrome P450 93E1; AltName: Full=Sophoradiol 24-hydroxylase	43 %

5.3.2 RACE amplification and full length sequence generation

From transcriptome screening as shown in table 5.3.1, 1 unigene (>Locus_41531) out of 10, with the length of 749 bp was selected using BLAST analysis based on its homology with other plant terpene mono-oxygenase in NCBI database. EST fragment analysis revealed the 59 % identity at amino acid level with cytochrome P450 71A1 from *Persea americana*, 53 % identity with Premnaspirodiene oxygenase (cytochrome P450 71D55) from *Hyoscyamus muticus* and lacking 300 amino acids at its 5′ end. To obtain the full-length sequence of this EST fragment (Locus_41531), 5' RACE was performed as discussed in methodology. 5' RACE fragment of expected size of 900 bp was obtained and cloned in TA cloning vector to generate 5′ missing sequence. The 5' RACE sequence of SaCYP450_1 was overlapped with EST fragment of >Locus_41531_Transcript_1/1_Confidence_1.000 and full length sequence was generated.

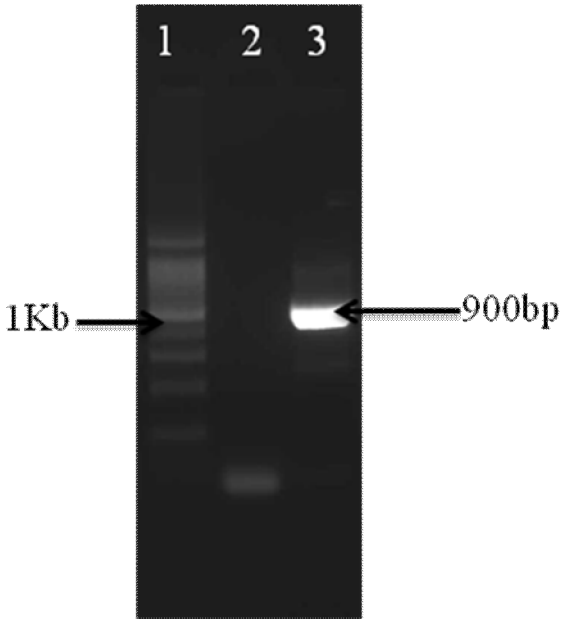


Figure 5.3.1: 5′ RACE amplification of >Locus_41531_Transcript matching with Premnaspirodiene oxygenase, **Lane 1:** 1 Kb DNA ladder, **Lane 2:** no insert control, **Lane 3:** 5′ RACE amplicon

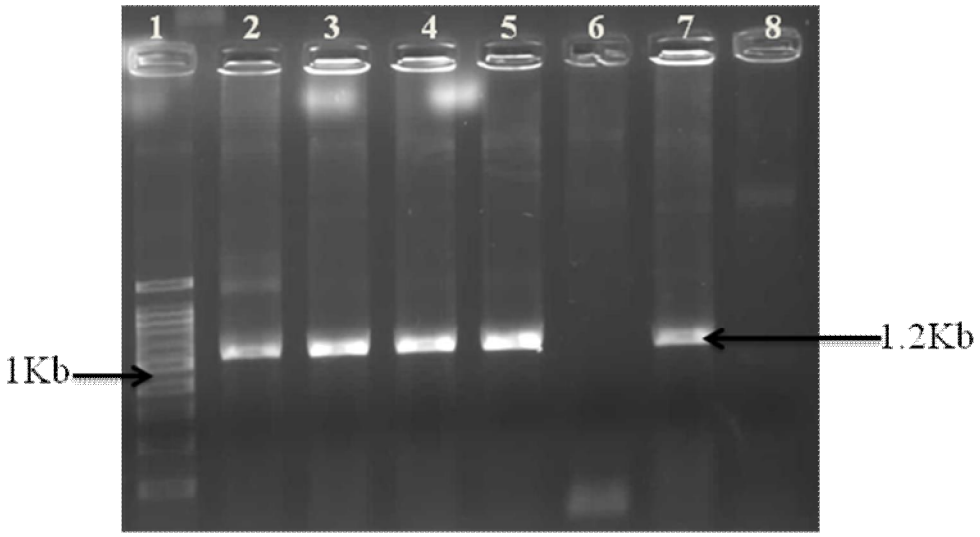


Figure 5.3.2: Colony PCR screening for 5′ RACE amplicon of locus_41531, **Lane 1:** 1 Kb DNA ladder, **Lanes 2-7:** Colony PCR for 5′ RACE amplicon with M13 forward and reverse primer, **Lane 8:** no insert control

The full-length sequence of SaCYP450_1 consists of 1704 bp having 5′ UTR of 36 bp and 3′ UTR of 162 bp. The ORF of SaCYP450_1 is composed of 1506 bp encoding a polypeptide of 501 amino acid residues with a calculated molecular weight of 56.06 kDa and pI of 6.51. The analysis of deduced amino acid sequence of SaCYP450_1 with other CYP450 system reported in NCBI database shows

resemblance to 47 % identity at amino acid level with cytochrome P450 71A1 from *Persea americana*, and 43 % identity with Premnaspirodiene oxygenase (cytochrome P450 71D55) from *Hyoscyamus muticus*. The protein sequence analysis of SaCYP450_1 revealed several highly conserved structural motifs characteristic of CYP450 system including heme-binding motif (**F**GG**G**R**R**G**C**P**G**) present at 65 residues upstream from C-terminal, proline rich region present at N-terminal. It also contains threonine rich motif for binding of an oxygen molecule present at amino acid 305-310 as AGTDTT with a consensus sequence of (**A**/G)**G**x(**D**/E)**T**(**T**/S).

5.3.3 SaCYP450 _1 full-length sequence generation and alignment

5' UTR sequence:

GATCACTTCACTCCGCCTCTCTTTCGGTCAAAATAA

ATGGACGAATTTACCCTTCTAGCCGTCTCCCTCTCAACCCTACTACTATTAGCCGCCGCCT
ACGTCCTCCATGCAAAGCCACCGCCCCGCTCAGCCGCCGGAGCGGCACCGCCATCTCCCC
GGGGCCTCCCTATCATCGGTCACCTCCACCTCGTCTCCGACATGCCCCACCACGCCTTCGC
CGACCTAGCGCGGCGGCTCGGCCCCATCTTCAGCCTCGAGCTCGGCCGAGTCCCCGCCGT
GGTCGTCTCCTCCGCTGACCTCGCCCGCAACGTCCTCAAGACTCACGACCACGTGTTCTCG
AACCGCCCGCAGCTCCTGTCGGCCCAGTACCTCTCCTTCGGCTGCTCCGACGTCACCTTTT
CCCCACACGGCCCCTACTGGCGCCAGGCCCGGAGGATATGCGTGACGGAGCTCCTCAGCC
CGCGGCGCGTGGCCTTGTTCGACCGCGTGCGGGCCGAGGAGGTCGGCCGGCTGATAACAA
GCCTCTCGGATCGGACCGGGACGGCGGTGGACGTGAGCAGGACGCTGTTCAATTTGGCGA
ACGACATTCTGTGTCGGGTCGCTTTCGGGCGGCGATTCGGCGGAGAGGGGGAGGAGAAGA
AGGGGAGGTTGGTGGAGGTGATGAAGGAGACGCAGGAGCTGTTTGCTGGGTTCTGCGTGG
GGGATTTTTATCCGGAGTGGGGATGGGTGAACTGTGTGGGTGGGTATAAGAGGAGGCTAG
AGAAGAACCTGGAGGATTTGAGAGAGGTTTGTGATGAGATCATCAAGGAACATCTCGTGA
GACAAGGGGATGAGGAGGAGAGAGGGGAAGATTTTGTGGACGTGTTACTTCGGGTGCAA
AAGAGAGAAGACTTTGAGGTTCCCATTACAGAAGATAATCTCAAAGCCCTTGTCCTGGAC
ATGTTTGTTGCTGGAACTGATACGACCTCATCAACCCTCGAATGGGCAATGACCGAGCTG
GCCAAGCACCCGAAAGTGCTGAAGAAGGCGCAAGAAGAAGTTCGAAAAATCGCATTCAA
GGAAGGAAAAGTGGATGAAACTCATCTTCAACACCTGCATTACATGAAAGCCGTGATAAA
GGAAACGCTCCGCCTGCACCCACCAGCGCCTCTGCTGGTGCCTCGCGAGTCCATGGAGAA
GTGCGTTCTTGCTGGCTACGAGATACCCGCCAAAACTCGGGTTCTGATCAACGCCTACGCA
ATCGGGAGGGACCCGCAGTCGTGGGAGAACCCTCTCGAGTACAATCCCGACAGATTTTGC
GAGGAGGGATCTGATGCGAAGGGTCAAGACTTCAGGTTTTTACCATTTGGGGGAGGAAGG
AGGGGGTGCCCGGGTTCTTCCTTCGGGCTGGCGACTGTGGAGCTCGCGCTTGCTCGACTTC
TCCATCATTTTGATTGGGCATTGCCTCCGGGGGTGAAGACGGATGATGTTGACCTGAGCGA
GATTTTCGGCCTAGCCAGTAGGAAGAAAAATGCTTTCATACTTGTCCCAAAGCACCTGTAA

3'UTR sequence:

GCCCAAAGGATGCATGATGGTGCCCAGTTTATCAGCTCCTAGAATAAGAGATATAATCCC
AATGGACATATAAGATAGAGGAAGGAGTTTGTGCTCTCGGATATCAAGCCTACGTCGTTT
TGATATGATAAGAGCAGAAAAAAAAAAAAAAAAAAAAAAAA

Protein sequence:

MDEFTLLAVSLSTLLLLAAAYVLHAKPPPRSAAGAAPPSPRGLPIIGHLHLVSDMPHHAFADLA
RRLGPIFSLELGRVPAVVVSSADLARNVLKTHDHVFSNRPQLLSAQYLSFGCSDVTFSPHGPY
WRQARRICVTELLSPRRVALFDRVRAEEVGRLITSLSDRTGTAVDVSRTLFNLANDILCRVAFG
RRFGGEGEEKKGRLVEVMKETQELFAGFCVGDFYPEWGWVNCVGGYKRRLEKNLEDLREV
CDEIIKEHLVRQGDEEERGEDFVDVLLRVQKREDFEVPITEDNLKALVLDMFVAGTDTTSSTLE
WAMTELAKHPKVLKKAQEEVRKIAFKEGKVDETHLQHLHYMKAVIKETLRLHPPAPLLVPRE

SMEKCVLAGYEIPAKTRVLINAYAIGRDPQSWENPLEYNPDRFCEEGSDAKGQDFRFLPFGGG
RRGCPGSSFGLATVELALARLLHHFDWALPPGVKTDDVDLSEIFGLASRKKNALILVPKHL-

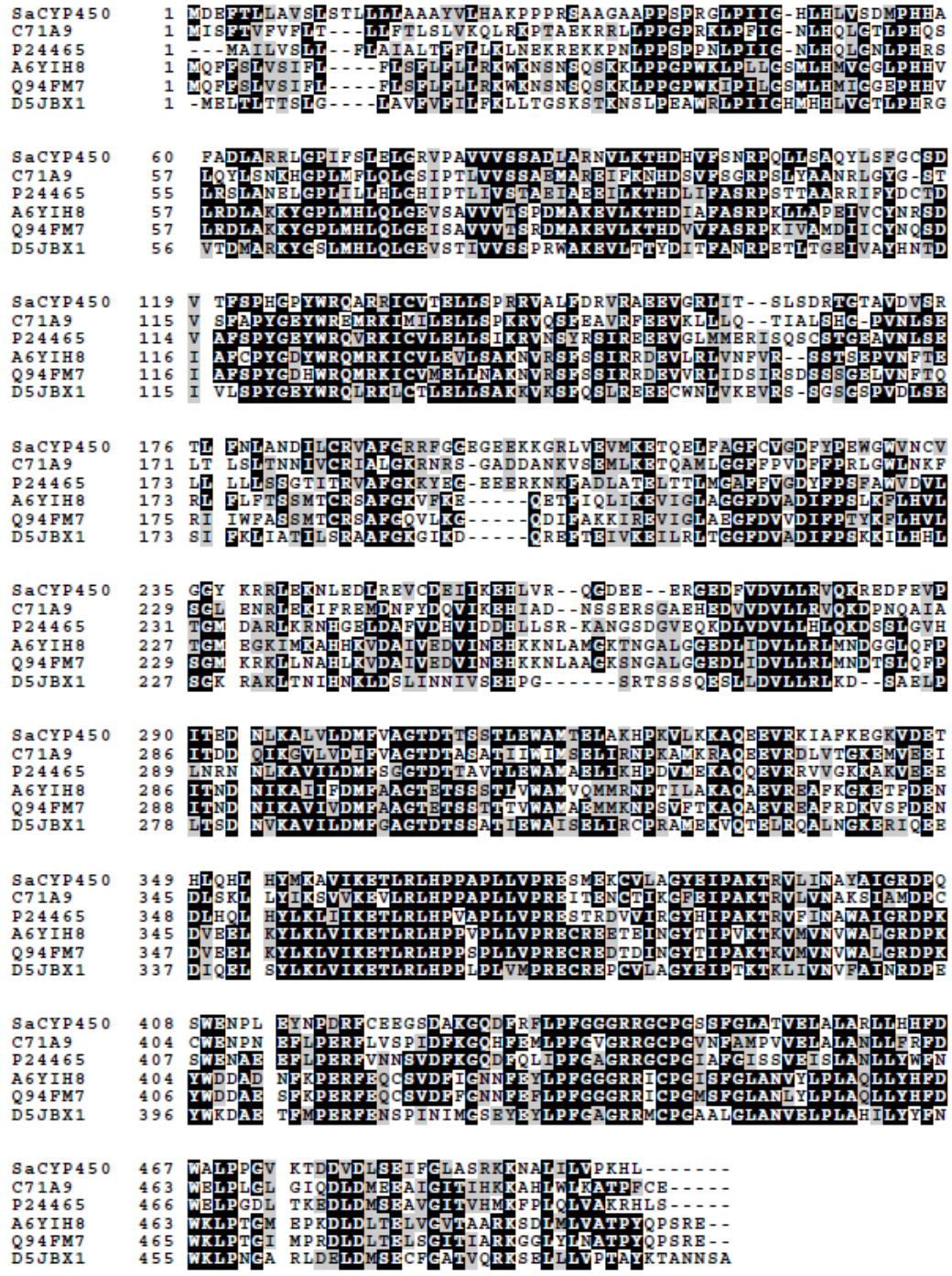


Figure 5.3.3: Multiple sequence alignment of SaCYP450_ mono-oxygenase 1 with CYP450 mono-oxygenase reported in NCBI database from other plant species, C71A9: Cytochrome P450 CP1 from *Glycine max*, P24465: C71A1_PERAE from *Persea americana*, A6YIH8: Premnaspirodiene oxygenase from *Hyoscyamus muticus* Q94FM7: 5-epiaristolochene 1,3-dihydroxylase from *Nicotiana tabacum*, D5jBX1: Germacrene A oxidase from *Barnadesia spinosa*

5.3.4 Full length PCR amplification of SaCYP450 and cloning in pESC- Duet-MCS2

The ORF of CYP450_1 was amplified from cDNA using proof reading taq DNA polymerase. The expected 1506 bp of amplicon was purified from agarose gel and cloned in multiple cloning site 2 (MCS2) of pESC-Duet vector frame (containing CYP450 reductase from *A. indica* in MCS 1) between SalI and KpnI restriction sites under the control of Gal 10 promoter for the expression of active protein. Positive clones were sequenced to confirm the frame of CYP450_1 in pESC vector.

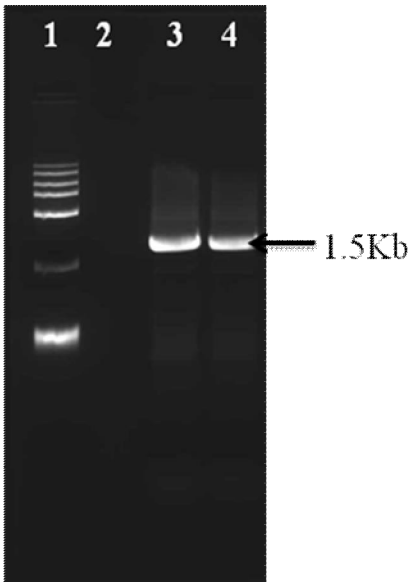


Figure 5.3.4: Full length ORF amplification of >Locus_41531, **Lane 1:** 1 Kb DNA ladder, **Lane 2:** no insert control, **Lanes 3-4:** ORF amplicon of transcript 41531

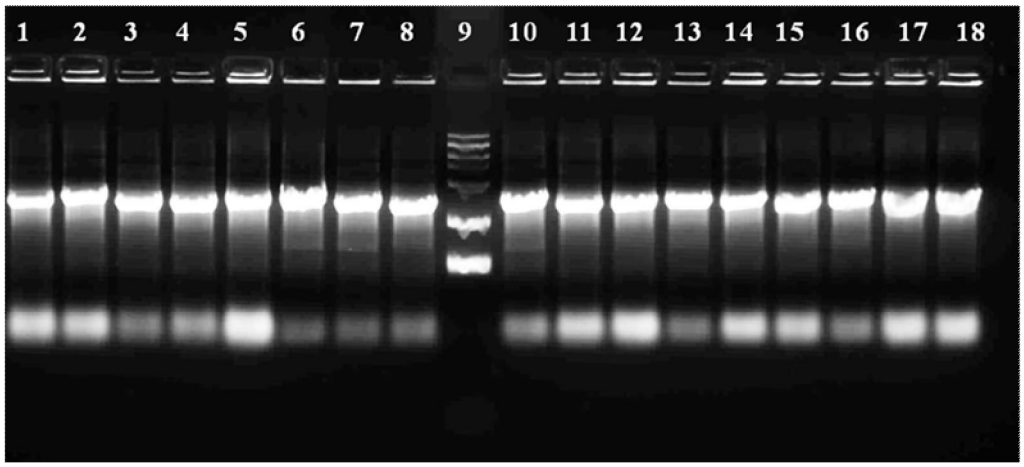


Figure 5.3.5: Colony PCR screening of SaCYP450_1 cloned in MCS 1 of pESC vector, **Lanes 1-8 and 10-18:** Colony PCR amplicon for ORF of CYP450_1 with Gal 1 promoter primer and CYP450_1 reverse primer, **Lane 9:** 1 Kb DNA ladder

5.3.5 Expression of active protein and enzymatic characterization of SaCYP450 mono-oxygenase system

pESC yeast-Duet vector harboring SaCYP450 mono-oxygenase system in MCS 2 and AiCYP450 reductase in MCS1 was introduced in INVSc1 yeast competent cells for the expression of active protein. Microsomes were prepared as discussed in materials and methods and enzyme assay was performed using microsomal fraction for the functional characterization of SaCYP450 mono-oxygenase system, with mixture of santalenes. GC and /GC-MS analyses of assay product mixture could not resulted in the hydroxylation of any of the santalene derivatives (Figure 5.3.1.6).

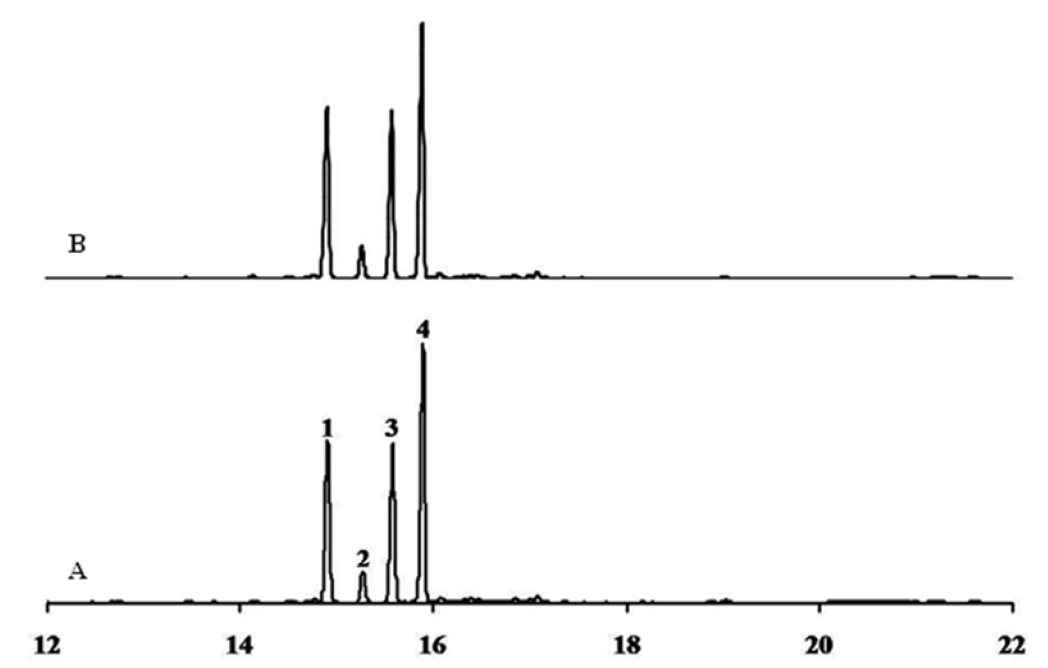


Figure 5.3.6: Enzyme assay of SaCYP450 mono-oxygenase system, **A)** Mixture of santalenes used for enzyme assay: α-santalene (**1**), exo-α-bergamotene (**2**), *epi*-β-santalene (**3**), β-santalene (**4**), **B)** Enzyme assay mixture of SaCYP450 mono-oxygenase system with santalene mixture

5.4 Conclusion

In order to establish the santalol/bergamotal biosynthesis in Indian Sandalwood, *S. album*, we have identified ten unigenes from the pool of transcriptome data set having the domain of CYP450-monooxygenase and 3 CYP450 reductases. 1 unigene having the domain of CYP450 71D family was extended to generate full-length cDNA sequence. The full length ORF of SaCYP450_1 was amplified and cloned in multiple cloning site 2 of pESC yeast expression vector (containing CYP450 reductase from *A. indica*). This was expressed in yeast and prepared microsomes were used for its ability to carry out hydroxylation at *cis* methyl group of santalenes and bergamotenes. However, the GC and GC-MS analyses indicated that, the expressed SaCYP450 mono-oxygenase system failed to carry out the hydroxylation on these sesquiterpene hydrocarbons under the experimental conditions used.

5.5 References

(1) Adams, D. R.; Bhatnagar, S. P.; Cookson, R. C. *Phytochemistry* **1975**, *14*, 1459-1460.

(2) Christenson, P. A.; Secord, N.; Willis, B. J. *Phytochemistry* **1981**, *20*, 1139-1141.

(3) Dwivedi, C.; Valluri, H. B.; Guan, X.; Agarwal, R. *Carcinogenesis* **2006**, *27*, 1917-1922.

(4) Muratore, A.; Clinet, J.-C.; Dunach, E. *Chem. Biodivers.* **2010**, *7*, 623-638.

(5) Brocke, C.; Eh, M.; Finke, A. *Chem. Biodivers.* **2008**, *5*, 1000-1010.

(6) Jones, C. G.; Moniodis, J.; Zulak, K. G.; Scaffidi, A.; Plummer, J. A.; Ghisalberti, E. L.; Barbour, E. L.; Bohlmann, J. *J. Biol. Chem.* **2011**, *286*, 17445-17454.

(7) Daramwar, P. P.; Srivastava, P. L.; Priyadarshini, B.; Thulasiram, H. V. *Analyst* **2012**, *137*, 4564-4570.

(8) Jones, C. G.; Moniodis, J.; Zulak, K. G.; Scaffidi, A.; Plummer, J. A.; Ghisalberti, E. L.; Barbour, E. L.; Bohlmann, J. r. *J. Biol. Chem.* **2011**, *286*, 17445-17454.

(9) Hefner, J.; Rubenstein, S. M.; Ketchum, R. E. B.; Gibson, D. M.; Williams, R. M.; Croteau, R. *Chem. Biol.* **1996**, *3*, 479-489.

(10) Hedden, P.; Kamiya, Y. *Annu. Rev. Plant Phys.* **1997**, *48*, 431-460.

(11) Akashi, T.; Aoki, T.; Ayabe, S. *Plant Physiol.* **1999**, *121*, 821-828.

(12) Jung, W.; Yu, O.; Lau, S. M. C.; O'Keefe, D. P.; Odell, J.; Fader, G.; McGonigle, B. *Nat. Biotechnol.* **2000**, *18*, 208-212.

(13) Chapple, C. *Annu. Rev. Plant Phys.* **1998**, *49*, 311-343.

(14) Nielsen, J. S.; Moller, B. L. *Arch. Biochem. Biophys.* **1999**, *368*, 121-130.

(15) Kraus, P. F.; Kutchan, T. M. *P. Natl. Acad. Sci. USA* **1995**, *92*, 2071-2075.

(16) Bernhardt, R. *J. Biotechnol.* **2006**, *124*, 128-145.

(17) Omura, T.; Sato, R. *J. Biol. Chem.* **1964**, *239*, 2370-8.

(18) Pandreka, A. *Ph. D. Work*, *Unpublished data*.

Sequences of terpene synthases isolated from Indian Sandalwood *Santalum album*

1) *SaFDS*

>Locus_19031_Transcript_1/1_Confidence_1.000_Length_453bp

GTGCATTGAATGGCTTCAAGCATATTTTCTTGTTCTCGATGATATTATGGATGGTTCTCATA
CACGCCGAGGTCAGCCTTGTTGGTTCAGGTTGCCTGAGGTTGGTCTGATTGCTGTAAATGA
TGGCATAATGCTTCGCAACCACATCCCAAGAATTCTCAAGAAGCACTTCAAAAATAAGCC
TTATTATGTGGAACTGTTGGATTTATTTAATGAGGTCGAGTTCCAAACAACTTCAGGACAG
ATGATAGATTTGATAACCACGCTTGAAGGGCAGAAAGATCTTTCAAAGTATTCAATGCCT
ATTCACCATCGCATTGTTCAGTATAAAACTGCTTATTACTCCTTTTACCTTCCGGTTGCTTG
TGCACTGCTTATGTCAGGTGAGAATCTGGACAGCCACACTGAAGTGGAGAAAATCCTTGT
TGAAATGGGAACCTATTTTCAAGTACA

Protein Sequence

CIEWLQAYFLVLDDIMDGSHTRRGQPCWFRLPEVGLIAVNDGIMLRNHIPRILKKHFKNKPYY
VELLDLFNEVEFQTTSGQMIDLITTLEGQKDLSKYSMPIHHRIVQYKTAYYSFYLPVACALLMS
GENLDSHTEVEKILVEMGTYFQV

***SaFDS* 5' RACE Sequence**

Nucleotide sequence

5' UTR

GAAAATTCTTCATCCCTCAAATCTCTCAAATCTTTCAAATCTCTCAAAATCTCAAACCTTTC
TTCACACA

ATGGGCGATCGGAAAACCAAATTTCTCGAGGCCTACTCTGTCTTGAAATCGGAGCTCCTCC
GGGACCCTGCTTTCAATTTTACAGACGCTTCCCGTCAATGGGTCGACCGGATGCTGGACTA
CAATGTGCCTGGAGGGAAACTGAATCGAGGGCTCTCAGTGATTGACAGCTATGAGTTGCT
GAAAGAAGGAAAAGAGCTAACTGATGATGAAATATTTCTTGCATCTGCACTCGGTTGGTG
CATTGAATGGCTTCAAGCATATTTTCTTGTTCTCGATGATATTATGGATGGCTCTCATACAC
GCCGAGGTCAGCCTTGTTGGTTCAGGTTGCCTGAGGTTGGTCTGATTGCTGTAAATGATGG
CATAATGCTTCGCAACCACATCCCAAGAATTCTCAAGAAGCACTTCAAAAATAAGCCTTA
TTATGTGGAACTGTTGGATTTATTTAATGAGGTCGAGTTCCAAACAACTTCAGGACAGATG
ATAGATTTGATAACCACGCTTGAAGGGCAGAAAGATCTTTCAAAGTATTCAATGCCTATTC
ACCATCGCATTGTTCAGTATAAAACTGCTTATTACTCCTTTTACCTTCCCGGTTGCTTGTGC
ACTGCTTATGTCAGTGAGAATCTGGACAGCCACACTGAAGTGGAGAAAATCCTTGTTGAA
ATGGGAACCTATTTTCAAGTACA

***SaFDS* 3' RACE Sequence**

Nucleotide Sequence

TGTAAATGATGGCATAATGCTTCGCAACCACATCCCAAGAATTCTCAAGAAGCACTTCAA
AAATAAGCCTTATTATGTGGAACTGTTGGATTTATTTAATGAGGTCGAGTTCCAAACAACT
TCAGGACAGATGATAGATTTGATAACCACGCTTGAAGGGCAGAAAGATCTTTCAAAGTAT

TCAATGCCTATTCACCATCGCATTGTTCAGTATAAAACTGCTTATTACTCCTTTTACCTTCC
GGTTGCTTGTGCACTGCTTATGTCAGGTGAGAATCTGGACAGCCACACTGAAGTGGAGAA
AATCCTTGTTGAAATGGGAACCTATTTTCAAGTACAGGATGATTACCTGGACTGCTTTGGT
CATCCTGATGTCATTGGAAAGATTGGAACAGATATTGAAGATTTTAAGTGTTCTTGGTTGG
TTGTAAAAGCGTTGGAACTTTCCAACGAGGAACAGAAGAAATTATTATATGAGAACTATG
GGAAAGCCGATGAAGCCAGCGTTGCAAAAGTAAAGGCACTTTATAAGGAACTTGACCTTG
AGGGTGCATTTGTGGAGTACGAGAATGCTAGTTATGAGAAGATAATCAGCTCAATTGAGG
TGCAGCCAAGCAAAGCAGTACAAGCAGTGCTGAAATCCTTTTTGGCGAAGATATACAAGC
GGCAGAAGTAG

3' UTR sequence

GAATTTGAGGCGTTAACTGGTGATGGACTAAGTAGATTTTTCGTCATTGTGCCATTGGAGG
AATAAGAAATGTATTTGTATTTTGTTCTCCTGAAGCCTTTTATTATTTCGGTTTTGTATTGG
TTTCGAAGCTTCCCCTTAAGACTAAAATAACTACCGAGTTTCTCCCCTGTTATGCTGCCCC
AAATAGCAATTTGAGTTGCAGTGAACCTTTTATCATGACCTGTTTGTTGTGACTTTGAGCA
CCTACCGTTGAATGAGTTATTTTTCATTAAAAAAAAAAAAAAAAAAAAAAA

***SaFDS*_full length ORF sequence (Accession number: KF011939)**

Nucleotide sequence

ATGGGCGATCGGAAAACCAAATTTCTCGAGGCCTACTCTGTCTTGAAATCGGAGCTCCTCC
GGGACCCTGCTTTCAATTTTACAGACGCTTCCCGTCAATGGGTCGACCGGATGCTGGACTA
CAATGTGCCTGGAGGGAAACTGAATCGAGGGCTCTCAGTGATTGACAGCTATGAGTTGCT
GAAAGAAGGAAAAGAGCTAACTGATGATGAAATATTTCTTGCATCTGCACTCGGTTGGTG
CATTGAATGGCTTCAAGCATATTTTCTTGTTCTCGATGATATTATGGATGGCTCTCATACAC
GCCGAGGTCAGCCTTGTTGGTTCAGGTTGCCTGAGGTTGGTCTGATTGCTGTAAATGATGG
CATAATGCTTCGCAACCACATCCCAAGAATTCTCAAGAAGCACTTCAAAAATAAGCCTTA
TTATGTGGAACTGTTGGATTTATTTAATGAGGTCGAGTTCCAAACAACTTCAGGACAGATG
ATAGATTTGATAACCACGCTTGAAGGGCAGAAAGATCTTTCAAAGTATTCAATGCCTATTC
ACCATCGCATTGTTCAGTATAAAACTGCTTATTACTCCTTTTACCTTCCGGTTGCTTGTGCA
CTGCTTATGTCAGGTGAGAATCTGGACAGCCACACTGAAGTGGAGAAAATCCTTGTTGAA
ATGGGAACCTATTTTCAAGTACAGGATGATTACCTGGACTGCTTTGGTCATCCTGATGTCA
TTGGAAAGATTGGAACAGATATTGAAGATTTTAAGTGTTCTTGGTTGGTTGTAAAAGCGTT
GGAACTTTCCAACGAGGAACAGAAGAAATTATTATATGAGAACTATGGGAAAGCCGATG
AAGCCAGCGTTGCAAAAGTAAAGGCACTTTATAAGGAACTTGACCTTGAGGGTGCATTTG
TGGAGTACGAGAATGCTAGTTATGAGAAGATAATCAGCTCAATTGAGGTGCAGCCAAGCA
AAGCAGTACAAGCAGTGCTGAAATCCTTTTTGGCGAAGATATACAAGCGGCAGAAGTAG

Protein Sequence

MGDRKTKFLEAYSVLKSELLRDPAFNFTDASRQWVDRMLDYNVPGGKLNRGLSVIDSYELLK
EGKELTDDEIFLASALGWCIEWLQAYFLVLDDIMDGSHTRRGQPCWFRLPEVGLIAVNDGIML
RNHIPRILKKHFKNKPYYVELLDLFNEVEFQTTSGQMIDLITTLEGQKDLSKYSMPIHHRIVQY
KTAYYSFYLPVACALLMSGENLDSHTEVEKILVEMGTYFQVQDDYLDCFGHPDVIGKIGTDIE

DFKCSWLVVKALELSNEEQKKLLYENYGKADEASVAKVKALYKELDLEGAFVEYENASYEK
IISSIEVQPSKAVQAVLKSFLAKIYKRQK

2) Santalene Synthase (*SaSS*)

>Locus_1838_Transcript_1/1_Confidence_1.000_Length_ 438bp

CATGCGACATCTCTCCGATTTAGGCTGCTACGACAGTGTGGGCTTTTTATTCCCCAAGATG
TGTTTAAAACGTTCCAAAACAAGACTGGGGAATTTGATATGAAACTTTGTGACAACGTAA
AAGGGCTGCTGAGCTTATATGAAGCTTCATACTTGGGATGGAAGGGTGAAAACATCCTAG
ATGAAGCCAAGGCCTTCACCACCAAGTGCTTGAAAAGTGCATGGGAAAATATATCCGAAA
AGTGGTTAGCCAAAAGAGTGAAGCATGCATTGGCTTTGCCTTTGCATTGGAGAGTCCCTCG
AATCGAAGCTAGATGGTTCATTGAGGCATATGAGCAAGAAGCGAATATGAACCCAACACT
ACTCAAACTCGCAAAATTAGACTTTAATATGGTGCAATCAATTCATTAGAAAGAGATTTG
GGAATTAGCAAGGT

Protein Sequence

HATSLRFRLLRQCGLFIPQDVFKTFQNKTGEFDMKLCDNVKGLLSLYEASYLGWKGENILDEA
KAFTTKCLKSAWENISEKWLAKRVKHALALPLHWRVPRIEARWFIEAYEQEANMNPTLLKLA
KLDFNMVQSIH

***SaSS* 3' RACE Sequence**

Nucleotide Sequence

AGCGAATATGAACCCAACACTACTCAAACTCGCAAAATTAGACTTTAATATGGTGCAATC
AATTCATCAGAAAGAGATTGGGGAATTAGCAAGGTGGTGGGTGACTACTGGCTTGGATAA
GTTAGCCTTTGCCAGGAATAATTTACTGCAGAGCTATATGTGGAGCTGCGCGATTGCTTCC
GACCCGAAGTTCAAACTTGCTAGAGAAACTATTGTCGAAATCGGAAGTGTACTCACAGTT
GTTGACGATGGATATGACGTCTATGGTTCAATCGACGAACTTGATCTCTACACAAGCTCCG
TTGAAAGGTGGAGCTGTGTGGAAATTGACAAGTTGCCAAACACGTTAAAATTAATTTTTAT
GTCTATGTTCAACAAGACCAATGAGGTTGGCCTTCGAGTCCAGCATGAGCGAGGCTACAA
TAGCATCCCTACTTTTATCAAAGCGTGGGTTGAACAGTGTAAATCATACCAGAAAGAAGC
AAGATGGTTCCACGGGGGACACACGCCTCCATTGGAAGAATATAGCTTGAATGGACTTGT
TTCCATAGGATTCCCTCTCTTGTTAATCACGGGCTACGTGGCAATCGCTGAGAACGAGGCT
GCACTGGATAAAGTGCACCCCCTTCCTGATCTTCTGCACTACTCCTCCCTCCTTAGTCGCCT
CATCAATGATATAGGAACGTCTCCGGATGAGATGGCAAGAGGCGATAATCTGAAGTCAAT
CCATTGTTACATGAACGAAACTGGGGCTTCCGAGGAAGTTGCTCGTGAGCACATAAAGGG
AGTAATCGAGGAGAATTGGAAAATACTGAATCAGTGCTGCTTTGATCAATCTCAGTTTCA
GGAGCCTTTTATAACCTTCAATTTGAACTCTGTTCGAGGGTCTCATTTCTTCTATGAATTTG
GGGATGGCTTTGGGGTGACGGATAGCTGGACAAAGGTTGATATGAAGTCCGTTTTGATCG
ACCCTATTCCTCTCGGCGAGGAGTAG

3′ UTR sequence

TAAGCTCGAAGCTTATGTTTGGTTTATGGTGATAGTGGTTCGACAAAAATATAACATCGGA
CAGGAGTATCTATTTTTAATGATCGTGAGTTAAATAGGTAAGAAATATCTATTTCGCTCAT
GATTCATAGAAATATTATTCCTTACACGTTATATTTCTATCAAATAACTATTTTATTCCCGT
AAGCCAACTTTAACTTTACTCTTAAGGGTCACTAGTGGTTTTACTATATAATAAGTCAACT
AGCTTGAAGTAATAGAATAAAGGTTTGTTTTTAAGAGGGTTATGCACTACGGTTTCGATTG
TCTTTTACTAATAAGTTTCTTAACGAAACTCAGATGCTTTTTGGGATCATTCCCAAGGGATT
TGTGCTTGTATTATCATCGTCTCTCTTATTAATATATTCCCCTTTATGGCTGGTGTGGAAAA
AAAAAAAAAAAAAAAAAAAA

***SaSS*_full length ORF sequence (Accession number: KF011938)**

Nucleotide Sequence

ATGGATTCTTCCACCGCCACCGCCATGACAGCTCCATTCATTGATCCTACTGATCATGTGA
ATCTCAAAACTGATACGGATGCCTCAGAGAATCGAAGGATGGGAAATTATAAACCCAGCA
TTTGGAATTATGATTTTTTACAATCACTTGCAACTCATCACAATATTGTGGAAGAGAGGCA
TCTAAAGCTAGCTGAGAAGCTGAAGGGCCAAGTGAAGTTTATGTTTGGGGCACCAATGGA
GCCGTTAGCAAAGCTGGAGCTTGTGGATGTGGTTCAAAGGCTTGGGCTAAACCACCTATTT
GAGACAGAGATCAAGGAAGCGCTGTTTAGTATTTACAAGGATGGGAGCAATGGATGGTG
GTTTGGCCACCTTCATGCGACATCTCTCCGATTTAGGCTGCTACGACAGTGTGGGCTTTTT
ATTCCCCAAGATGTGTTTAAAACGTTCCAAAACAAGACTGGGGAATTTGATATGAAACTTT
GTGACAACGTAAAAGGGCTGCTGAGCTTATATGAAGCTTCATACTTGGGATGGAAGGGTG
AAAACATCCTAGATGAAGCCAAGGCCTTCACCACCAAGTGCTTGAAAAGTGCATGGGAAA
ATATATCCGAAAAGTGGTTAGCCAAAAGAGTGAAGCATGCATTGGCTTTGCCTTTGCATTG
GAGAGTCCCTCGAATCGAAGCTAGATGGTTCATTGAGGCATATGAGCAAGAAGCGAATAT
GAACCCAACACTACTCAAACTCGCAAAATTAGACTTTAATATGGTGCAATCAATTCATCA
GAAAGAGATTGGGGAATTAGCAAGGTGGTGGGTGACTACTGGCTTGGATAAGTTAGCCTT
TGCCAGGAATAATTTACTGCAGAGCTATATGTGGAGCTGCGCGATTGCTTCCGACCCGAA
GTTCAAACTTGCTAGAGAAACTATTGTCGAAATCGGAAGTGTACTCACAGTTGTTGACGAT
GGATATGACGTCTATGGTTCAATCGACGAACTTGATCTCTACACAAGCTCCGTTGAAAGGT
GGAGCTGTGTGGAAATTGACAAGTTGCCAAACACGTTAAAATTAATTTTTATGTCTATGTT
CAACAAGACCAATGAGGTTGGCCTTCGAGTCCAGCATGAGCGAGGCTACAATAGCATCCC
TACTTTTATCAAAGCGTGGGTTGAACAGTGTAAATCATACCAGAAAGAAGCAAGATGGTT
CCACGGGGGACACACGCCTCCATTGGAAGAATATAGCTTGAATGGACTTGTTTCCATAGG
ATTCCCTCTCTTGTTAATCACGGGCTACGTGGCAATCGCTGAGAACGAGGCTGCACTGGAT
AAAGTGCACCCCCTTCCTGATCTTCTGCACTACTCCTCCCTCCTTAGTCGCCTCATCAATGA
TATAGGAACGTCTCCGGATGAGATGGCAAGAGGCGATAATCTGAAGTCAATCCATTGTTA
CATGAACGAAACTGGGGCTTCCGAGGAAGTTGCTCGTGAGCACATAAAGGGAGTAATCGA
GGAGAATTGGAAAATACTGAATCAGTGCTGCTTTGATCAATCTCAGTTTCAGGAGCCTTTT
ATAACCTTCAATTTGAACTCTGTTCGAGGGTCTCATTTCTTCTATGAATTTGGGGATGGCTT

TGGGGTGACGGATAGCTGGACAAAGGTTGATATGAAGTCCGTTTTGATCGACCCTATTCCT
CTCGGCGAGGAGTAG

Protein Sequence

MDSSTATAMTAPFIDPTDHVNLKTDTDASENRRMGNYKPSIWNYDFLQSLATHHNIVEERHL
KLAEKLKGQVKFMFGAPMEPLAKLELVDVVQRLGLNHLFETEIKEALFSIYKDGSNGWWFGH
LHATSLRFRLLRQCGLFIPQDVFKTFQNKTGEFDMKLCDNVKGLLSLYEASYLGWKGENILDE
AKAFTTKCLKSAWENISEKWLAKRVKHALALPLHWRVPRIEARWFIEAYEQEANMNPTLLKL
AKLDFNMVQSIHQKEIGELARWWVTTGLDKLAFARNNLLQSYMWSCAIASDPKFKLARETIV
EIGSVLTVVDDGYDVYGSIDELDLYTSSVERWSCVEIDKLPNTLKLIFMSMFNKTNEVGLRVQ
HERGYNSIPTFIKAWVEQCKSYQKEARWFHGGHTPPLEEYSLNGLVSIGFPLLLITGYVAIAEN
EAALDKVHPLPDLLHYSSLLSRLINDIGTSPDEMARGDNLKSIHCYMNETGASEEVAREHIKGV
IEENWKILNQCCFDQSQFQEPFITFNLNSVRGSHFFYEFGDGFGVTDSWTKVDMKSVLIDPIPL
GEE-

3) β-Bisabolene Synthase (*SaβBS*)

>Locus_5558_Transcript_1/1_Confidence_1.000_Length_832

Nucleotide Sequence

TACAAGTCGAATAGGTAAGAAATAATTACTCGATGGACGAAATGAGCTCCCCACAAGAAT
AGGGTTGAAACATAGTAGCTACTATCATGAACCATGACAGAACGTAA
ATGTCTAACGAATCTTGTCTTTTTATGTTTAGTGATTGTAGCTTGTTGTACGTTTGATATTT
ATGCTATGTGGTTGGAAATGCAGCAGGAGAAGCATATGAAGATGGCTGAGAAGCTGAAG
GAGGAAGTGAAGTCTATGATCAAGGGGCAAATGGAACCCGTGGCGAAGCTTGAGTTGATC
AATATACTTCAGCGGCTCGGCCTGAAATATAGGTTTGAGTCTGAGATCAAGGAAGAATTG
TTTAGTCTTTACAAAGACGGTACCGATGCATGGTGGGTCGACAATTTGCATGCTACGGCCC
TCCGATTTCGGCTTCTACGTGAAAATGGGATATTTGTTCCCCAAGATGTGTTCGAGACTTT
GAAGGACAAAAGTGGTAAATTTAAGAGTCAATTGTGCAAGGATGTGAGGGGACTCCTGA
GCTTATATGAAGCCTCTTACTTGGGTTGGGAAGGTGAGGATTTACTTGATGAGGCAAAGA
AGTTTAGTACTACAAACTTGAACAATGTGAAGGAAAGCATATCTTCTAACACCTTGGGAA
GATTGGTGAAGCATGCTTTGAATCTGCCATTGCATTGGTCAGCAGCAAGATATGAGGCCA
GGTGGTTTATTGATGAGTATGAGAAAGAGGAAAATGTGAACCCTAATTTACTAAAATATG
CCAAGTTGGACTTCAATATTGTGCAATCAATTCATCAGCAGGAGCTTGGCAACCTAGCAA
GGT

Protein Sequence

MTERKCLTNLVFLCLVIVACCTFDIYAMWLEMQQEKHMKMAEKLKEEVKSMIKGQMEPVA
KLELINILQRLGLKYRFESEIKEELFSLYKDGTDAWWVDNLHATALRFRLLRENGIFVPQDVFE
TLKDKSGKFKSQLCKDVRGLLSLYEASYLGWEGEDLLDEAKKFSTTNLNNVKESISSNTLGRL
VKHALNLPLHWSAARYEARWFI DEYEKEENVNPNLLKYAKLDFNIVQSIHQQELGNLAR

SaβBS 3′ RACE sequence

GAAAGAGGAAAATGTGAACCCTAATTTACTAAAGTATGCCAAGTTTGACTTCAATATTGT
GCAATCAATTCATCAGCGGGAGCTTGGCAACCTAGCAAGGTGGTGGGTGGAGACGGGGTT
GGATAAGCTAAGCTTTGTGAGGAACACTTTAATGCAAAACTTCATGTGGGGCTGTGCCAT
GGTGTTTGAGCCACAGTACGGCAAAGTCAGAGATGCGGCCGTCAAACAGGCCAGTCTCAT
TGCTATGGTTGATGATGTTTATGATGTCTATGGCTCCTTGGAGGAATTGGAGATCTTCACA
GATATTGTTGACAGGTGGGACATCACTGGAATCGACAAACTACCAAGAAATATAAGTATG
ATTCTGCTTACGATGTTTAACACTGCAAATCAGATAGGCTATGACCTCCTGAGAGACCGAG
GCTTCAACGGCATTCCTCATATAGCCCAAGCTTGGGCAACTCTGTGCAAGAAGTACCTCAA
GGAAGCAAAATGGTACCACAGCGGATACAAGCCCACGCTCGAGGAGTACTTGGAGAACG
GATTAGTTTCCATTTCCTTTGTTCTTAGTCTCGTCACAGCATACTTGCAAACGGAAACCCTG
GAGAATTTAACCTATGAGAGCGCTGCGTACGTTAATTCTGTACCGCCTCTTGTCCGGTACT
CAGGTCTTCTCAACCGCCTCTACAATGATCTCGGAACGTCTTCAGCCGAGATAGCGAGAG
GAGACACGCTCAAGTCGATCCAGTGTTACATGACCCAAACCGGTGCAACTGAGGAGGCTG
CACGCGAGCACATCAAAGGGTTGGTGCATGAGGCTTGGAAAGGCATGAACAAGTGCTTGT
TTGAGCAAACCCCATTTGCCGAGCCTTTTGTGGGCTTCAATGTGAACACTGTTCGTGGGTC
TCAATTCTTCTACCAGCATGGAGATGGGTATGCTGTTACAGAGAGTTGGACTAAGGACCTT
TCCCTCTCAGTTCTCATCCACCCTATTCCACTGAACGAGGAGGATTGA

3′ UTR sequence

GTAAAGTTTCTCAAGGGAAGGGGGTGTTAGTAAGAAATAATTAGGGGAAGGATGCATGA
AAAACTATGTTTGTATGGTGCCCTGCGCAAGGAAAATGTCTCTTTAGCTATATGTGCTCGG
CACAGTGTAATGGATGTTTTAAGTATATGGTTCAGTATAATACAACTGATGCTGGCTCCGA
TCCAATAATAAAATTAATGTGGTTGGATCTAAAAAAAAAAAAAAAAAAAAAAAA

*SaβBS*_full length ORF sequence (Accession number: KJ665778)

Nucleotide sequence

ATGGATGCCTTTGCCACTTCTCCGACCTCAGCCCTCATCAAAGCAGTTAACTGTATTGCGCATGTCAC
TCCTATGGCTGGGGAAGATTCGTCTGAGAATCGACGGGCCAGCAATTACAAGCCAAGCAGTTGGGAC
TATGAGTTCCTGCAGTCACTCGCAACCTCTCACAACACTGTGCAGGAGAAGCATATGAAGATGGCTG
AGAAGCTGAAGGAGGAAGTGAAGTCTATGATCAAGGGGCAAATGGAACCCGTGGCGAAGCTTGAGT
TGATCAATATACTTCAGCGGCTCGGCCTGAAATATAGGTTTGAGTCTGAGATCAAGGAAGAATTGTT
TAGTCTTTACAAAGACGGTACCGATGCATGGTGGGTCGACAATTTGCATGCTACGGCCCTCCGATTTC
GGCTTCTACGTGAAAATGGGATATTTGTTCCCCAAGATGTGTTCGAGACTTTGAAGGACAAAAGTGG
TAAATTTAAGAGTCAATTGTGCAAGGATGTGAGGGGACTCCTGAGCTTATATGAAGCCTCTTACTTG
GGTTGGGAAGGTGAGGATTTACTTGATGAGGCAAAGAAGTTTAGTACTACAAACTTGAACAATGTGA
AGGAAAGCATATCTTCTAACACCTTGGGAAGATTGGTGAAGCATGCTTTGAATCTGCCATTGCATTG
GTCAGCAGCAAGATATGAGGCCAGGTGGTTTATTGATGAGTATGAGAAAGAGGAAAATGTGAACCC
TAATTTACTAAAGTATGCCAAGTTTGACTTCAATATTGTGCAATCAATTCATCAGCGGGAGCTTGGCA
ACCTAGCAAGGTGGTGGGTGGAGACGGGGTTGGATAAGCTAAGCTTTGTGAGGAACACTTTAATGCA
AAACTTCATGTGGGGCTGTGCCATGGTGTTTGAGCCACAGTACGGCAAAGTCAGAGATGCGGCCGTC
AAACAGGCCAGTCTCATTGCTATGGTTGATGATGTTTATGATGTCTATGGCTCCTTGGAGGAATTGGA

GATCTTCACAGATATTGTTGACAGGTGGGACATCACTGGAATCGACAAACTACCAAGAAATATAAGT
ATGATTCTGCTTACGATGTTTAACACTGCAAATCAGATAGGCTATGACCTCCTGAGAGACCGAGGCT
TCAACGGCATTCCTCATATAGCCCAAGCTTGGGCAACTCTGTGCAAGAAGTACCTCAAGGAAGCAAA
ATGGTACCACAGCGGATACAAGCCCACGCTCGAGGAGTACTTGGAGAACGGATTAGTTTCCATTTCC
TTTGTTCTTAGTCTCGTCACAGCATACTTGCAAACGGAAACCCTGGAGAATTTAACCTATGAGAGCGC
TGCGTACGTTAATTCTGTACCGCCTCTTGTCCGGTACTCAGGTCTTCTCAACCGCCTCTACAATGATCT
CGGAACGTCTTCAGCCGAGATAGCGAGAGGAGACACGCTCAAGTCGATCCAGTGTTACATGACCCA
AACCGGTGCAACTGAGGAGGCTGCACGCGAGCACATCAAAGGGTTGGTGCATGAGGCTTGGAAAGG
CATGAACAAGTGCTTGTTTGAGCAAACCCCATTTGCCGAGCCTTTTGTGGGCTTCAATGTGAACACTG
TTCGTGGGTCTCAATTCTTCTACCAGCATGGAGATGGGTATGCTGTTACAGAGAGTTGGACTAAGGA
CCTTTCCCTCTCAGTTCTCATCCACCCTATTCCACTGAACGAGGAGGATTGA

Protein sequence

MDAFATSPTSALIKAVNCIAHVTPMAGEDSSENRRASNYKPSSWDYEFLQSLATSHNTVQEKHMKMAEK
LKEEVKSMIKGQMEPVAKLELINILQRLGLKYRFESEIKEELFSLYKDGTDAWWVDNLHATALRFRLLRE
NGIFVPQDVFETLKDKSGKFKSQLCKDVRGLLSLYEASYLGWEGEDLLDEAKKFSTTNLNNVKESISSNT
LGRLVKHALNLPLHWSAARYEARWFIDEYEKEENVNPNLLKYAKFDFNIVQSIHQRELGNLARWWVET
GLDKLSFVRNTLMQNFMWGCAMVFEPQYGKVRDAAVKQASLIAMVDDVYDVYGSLEELEIFTDIVDR
WDITGIDKLPRNISMILLTMFNTANQIGYDLLRDRGFNGIPHIAQAWATLCKKYLKEAKWYHSGYKPTLE
EYLENGLVSISFVLSLVTAYLQTETLENLTYESAAYVNSVPPLVRYSGLLNRLYNDLGTSSAEIARGDTLK
SIQCYMTQTGATEEAAREHIKGLVHEAWKGMNKCLFEQTPFAEPFVGFNVNTVRGSQFFYQHGDGYAV
TESWTKDLSLSVLIHPIPLNEED*

4) *SaTPS1*

Locus_33105_Transcript_1/1_Confidence_1.000_Length_761

Nucleotide Sequence

CCCCCACCCTCTCTCTGTACCTCTTTCCGATCTAGGATTACTATAAATACATGAAGATGTG
AACCTTACGAGGCACCAACACAAGCTACGTTGCACTCTCATTCTCTCATTGTTGCAACCCT
AATTAGCCCTCCCTTCCATTCTACGTTCTCTTCATTTTCCTCACTTTCTCCCCA

ATGGATTTGTGTCAGATCCCGCCCACCTCTCCTATCTCCCCGTCCGTTCCATTTAATGGAGA
CGATTCGTCGGTGGTGCGGCGTTCCGCAAATTACCCTGCAAACTTGTGGGACTATGATTTC
CTCCAGTCCCTTGGACGCCATTCCAGCGTGACGGAGGAGCACGTAGGGCTGGCTGAGAAG
CTGAAGGGCGAAGTGAAGTCTCTGATTACGGGACCGATGGAGCCATTGGCCAAGTTGGAG
TTTATCGACAGTGTTCGGAGGCTCGGCCTAAAATACCAATTTGAGACTGAAATGAAGGAA
GCACTGGCTAACATTAGTAAGGATGGTTATGATTCATGGTGGGTCGACAATTTGCGTGCCA
CTGCTCTCCGATTTAGGCTCCTCCGCGAAAATGGGATATTTGTTCCTCAAGATGTGTTTGA
AAGGTTCCAGAACAAGGAAACTGGCAAATTCAAGAATGAACTATGTGAGGATGTGAAGG
GGCTGCTAAATTTGTATGAAGCATCTTTCTTGGGATGGGAAGGTGAGGACATCCTTGATGA
GGCAAGGACCTTTTCAACTGCACAACTAAAGAATGTTGAG

Protein Sequence

MDLCQIPPTSPISPSVPFNGDDSSVVRRSANYPANLWDYDFLQSLGRHSSVTEEHVGLAEKLK
GEVKSLITGPMEPLAKLEFIDSVRRLGLKYQFETEMKEALANISKDGYDSWWVDNLRATALRF
RLLRENGIFVPQDVFERFQNKETGKFKNELCEDVKGLLNLYEASFLGWEGEDILDEARTFSTA
QLKNVE

***SaTPS1* 3′ RACE sequence**

TTCAACTGCACAACTAAAGAATGTTGAGGGGAAAATTTCCTCCCCAAACTTGGCCAAAAT
AGTGCACCATGCTTTGGATTTGCCACTACATTGGAGAGCGATACGGTATGAGGCCAGATG
GTTCATTGACATTTATGAGGACGAAGAAGACATGAATCCTACTTTACTCAAATATGCAAA
ATTAGATTTCAACATCGTGCAATCATTTCATCAAGCGGAGATAGGTAGATTAGCGAGATG
GTGGGTCGGTACGGGCTTGGACAAGCTGCCGTTTGCGAGGAATGGTTTGATACAGAGCTA
CATGTATGCCATAGGTATGTTGTTTGAGCCACACCTGGGAGAAGTTAGAGAAATGGAAGC
AAAAGTTGGAGCTTTAATAACGACAATCGATGATGTCTATGATGTGTACGGAACAATGGA
AGAACTAGAACTCTTCACTGATATCACCGAGAGGTGGGATATCAATAGAGTCGATCAATT
ACCACGGAATATTAGAATGCCTCTCCTCACCATGTTCAACACTTCTAACGATATCGGCTAC
TGGGCATTAAAAGAGAGAGGGTTTAATGGAATTCCTTACACAGCTAAAGTGTGGGCTGAT
CAACTAAAAAGTTACACCAAAGAAGCAAAATGGTTTCACGAGGGGCACAAGCCAACCTT
GGAGGAGTACTTGGAAAATGCTCTGGTCTCCATAGGATTCCCAAACCTTCTTGTCACCAGC
TACCTACTTACCGTAGACAATCCAACTAAGGAGAAGCTTGACTACGTCGACAGCCTCCCTC
TCTTCGTGCGTGCATCTTGCATTCTTTGTCGGATCATCAACGACTTGGGAACTTCTCCGGAT
GAGATGGAAAGAGGAGATAATCTCAAATCGATTCAATGCTACATGAATGAAACCGGTGCT
TCTCAAGAAGTCGCTCGCGAGCACATTGAAGGTCTGGTTCGTATGTGGTGGAAGAGGTTG
AACAAATGCTTGTTTGAGCCGTCGCCTTTCACCGAGCCTTTCCTAAGTTTCACCATAAATG
TGGTGCGAGGTTCACATTTCTTCTACCAATATGGAGATGGGTACGGCAATGCAGAGAGTT
GGACTAAGAATCAGGGCATGTCTGTGCTAATTCACCCAATTACGCTAGATGAGGAGTAA

3′ UTR sequence

ACAATACAGAACATGTGATTTATAATAAAGTTGTTGCATATAACTAAAGCAAGATCATAA
ATAATTAAAGGAAATCCAGCTATAAATCTATACGCTTGTAACGATTTGCTGTTCTACTTCT
CCTTTTTGTCATGCTCTCCATGTATTGATTCCAATTTTTTTCTATACTTTTTAGATATATATT
GGATTCAAGGACCATAATTGTACTACTAAATAATTTTTGAAATGTTCCTTGAAAAAAAAA
AAAAAAAAAAAAAAAA

***SaTPS1*_Full length ORF sequence (Accession number: KJ665776)**

Nucleotide Sequence

ATGGATTTGTGTCAGATCCCGCCCACCTCTCCTATCTCCCCGTCCGTTCCATTTAATGGAGA
CGATTCGTCGGTGGTGCGGCGTTCCGCAAATTACCCTGCAAACTTGTGGGACTATGATTTC
CTCCAGTCCCTTGGACGCCATTCCAGCGTGACGGAGGAGCACGTAGGGCTGGCTGAGAAG
CTGAAGGGCGAAGTGAAGTCTCTGATTACGGGACCGATGGAGCCATTGGCCAAGTTGGAG

TTTATCGACAGTGTTCGGAGGCTCGGCCTAAAATACCAATTTGAGACTGAAATGAAGGAA
GCACTGGCTAACATTAGTAAGGATGGTTATGATTCATGGTGGGTCGACAATTTGCGTGCCA
CTGCTCTCCGATTTAGGCTCCTCCGCGAAAATGGGATATTTGTTCCTCAAGATGTGTTTGA
AAGGTTCCAGAACAAGGAAACTGGCAAATTCAAGAATGAACTATGTGAGGATGTGAAGG
GGCTGCTAAATTTGTATGAAGCATCTTTCTTGGGATGGGAAGGTGAGGACATCCTTGATGA
GGCAAGGACCTTTTCAACTGCACAACTAAAGAATGTTGAGGGGAAAATTTCCTCCCCAAA
CTTGGCCAAAATAGTGCACCATGCTTTGGATTTGCCACTACATTGGAGAGCGATACGGTAT
GAGGCCAGATGGTTCATTGACATTTATGAGGACGAAGAAGACATGAATCCTACTTTACTC
AAATATGCAAAATTAGATTTCAACATCGTGCAATCATTTCATCAAGCGGAGATAGGTAGA
TTAGCGAGATGGTGGGTCGGTACGGGCTTGGACAAGCTGCCGTTTGCGAGGAATGGTTTG
ATACAGAGCTACATGTATGCCATAGGTATGTTGTTTGAGCCACACCTGGGAGAAGTTAGA
GAAATGGAAGCAAAAGTTGGAGCTTTAATAACGACAATCGATGATGTCTATGATGTGTAC
GGAACAATGGAAGAACTAGAACTCTTCACTGATATCACCGAGAGGTGGGATATCAATAGA
GTCGATCAATTACCACGGAATATTAGAATGCCTCTCCTCACCATGTTCAACACTTCTAACG
ATATCGGCTACTGGGCATTAAAAGAGAGAGGGTTTAATGGAATTCCTTACACAGCTAAAG
TGTGGGCTGATCAACTAAAAAGTTACACCAAAGAAGCAAAATGGTTTCACGAGGGGCACA
AGCCAACCTTGGAGGAGTACTTGGAAAATGCTCTGGTCTCCATAGGATTCCCAAACCTTCT
TGTCACCAGCTACCTACTTACCGTAGACAATCCAACTAAGGAGAAGCTTGACTACGTCGA
CAGCCTCCCTCTCTTCGTGCGTGCATCTTGCATTCTTTGTCGGATCATCAACGACTTGGGAA
CTTCTCCGGATGAGATGGAAAGAGGAGATAATCTCAAATCGATTCAATGCTACATGAATG
AAACCGGTGCTTCTCAAGAAGTCGCTCGCGAGCACATTGAAGGTCTGGTTCGTATGTGGT
GGAAGAGGTTGAACAAATGCTTGTTTGAGCCGTCGCCTTTCACCGAGCCTTTCCTAAGTTT
CACCATAAATGTGGTGCGAGGTTCACATTTCTTCTACCAATATGGAGATGGGTACGGCAAT
GCAGAGAGTTGGACTAAGAATCAGGGCATGTCTGTGCTAATTCACCCAATTACGCTAGAT
GAGGAGTAA

Protein Sequence

MDLCQIPPTSPISPSVPFNGDDSSVVRRSANYPANLWDYDFLQSLGRHSSVTEEHVGLAEKLK
GEVKSLITGPMEPLAKLEFIDSVRRLGLKYQFETEMKEALANISKDGYDSWWVDNLRATALRF
RLLRENGIFVPQDVFERFQNKETGKFKNELCEDVKGLLNLYEASFLGWEGEDILDEARTFSTA
QLKNVEGKISSPNLAKIVHHALDLPLHWRAIRYEARWFIDIYEDEEDMNPTLLKYAKLDFNIV
QSFHQAEIGRLARWWVGTGLDKLPFARNGLIQSYMYAIGMLFEPHLGEVREMEAKVGALITTI
DDVYDVYGTMEELELFTDITERWDINRVDQLPRNIRMPLLTMFNTSNDIGYWALKERGFNGIP
YTAKVWADQLKSYTKEAKWFHEGHKPTLEEYLENALVSIGFPNLLVTSYLLTVDNPTKEKLD
YVDSLPLFVRASCILCRIINDLGTSPDEMERGDNLKSIQCYMNETGASQEVAREHIEGLVRMW
WKRLNKCLFEPSPFTEPFLSFTINVVRGSHFFYQYGDGYGNAESWTKNQGMSVLIHPITLDEE-

5) *SaTPS2*

>Locus_8408_Transcript_1/1_Confidence_1.000_Length_1016

Nucleotide Sequence

ATGGCCTCTGTGATTGTTGAACCCATTCGTTGCAACAATGACAATGATGTCATTTCGACTG
TGGTCGACGATTCGTCCGTGGTTCGGCGGTCCGCCAATTACCCGCCAAACCTGTGGGACTA
TGAATTTCTTCAGTCACTTGGCGACCAGTGCACCGTGGAGGAGAAGCATTTAAAGTTGGCT
GACAAACTGAAGGAGGAAGTGAAGTCCCTGATTAAGCAAACCATGGAGCCATTAGCAAA
GTTGGAGTTCATCGATACCGTTCGGAGGCTGGGCTTAAAATACCAGTTTGAGACTGAGGT
GAAGGAAGCAGTCGTTATGGTTAGCAAATATGAGAACGATGCATGGTGGATCGACAATTT
ACATGCCACGTCTCTCCGGTTCAGGATCATGCGAGAAAACGGGATATTTGTTCCCCAAGAT
GTTTTTGAAAGGTTCAAGGACACTGATGGATTCAAGAATCAACTGTGTGAGGATGTGAAG
GGGTTGCTCAGTTTGTATGAGGCTTCATTCTTAGGCTGGGAAGGTGAGGACATCCTTGATG
AGGCAAGGACCTTTGCAACCTCCAAACTAAAGAGCATCGAGGGGAAAATTCCCTCCCCAA
GCTTGGCCAAAAAAGTGAGCCATGCTTTGGATTTGCCGCTTCATTGGCGAACGATAAGGT
ACGAGGCTAGATGGTTCATCGACACATATGAAGAAGAAGAAGACGTGAACCTTACTTTGC
TTAGATACGCTAAACTGGACTTCAACATCGTGCAATCATTTCATCAAAAGGAGATTGGTA
GATTATCAAGGT

Protein Sequence

MASVIVEPIRCNNDNDVISTVVDDSSVVRRSANYPPNLWDYEFLQSLGDQCTVEEKHLKLAD
KLKEEVKSLIKQTMEPLAKLEFIDTVRRLGLKYQFETEVKEAVVMVSKYENDAWWIDNLHAT
SLRFRIMRENGIFVPQDVFERFKDTDGFKNQLCEDVKGLLSLYEASFLGWEGEDILDEARTFAT
SKLKSIEGKIPSPSLAKKVSHALDLPLHWRTIRYEARWFIDTYEEEEDVNLTLLRYAKLDFNIV
QSFHQKEIGRLSR

***SaTPS2* 3' RACE Sequence**

CACATATGAAGAAGAAGAAGACGTGAACCTTACTTTGCTTAGATACGCTAAACTGGACTT
CAACATCGTGCAATCATTTCATCAAAAGGAGATTGGTAGATTATCAAGGTGGTGGGTGGG
TACGGGCCTGGACAAGATGCCATTTGCAAGAAATGGCTTGATACAGAGCTACATGTATGC
CATAGGCATGCTGTTTGAACCGAACTTGGGAGAAGTTAGAGAAATGGAAGCCAAAGTTGG
AGCTTTGATAACGACAATCGATGATGTTTATGATGTGTACGGAACAATGGAAGAACTAGA
ACTCTTCACAGATATCACCAACAGATGGGATATTAGTAAAGCTGATCAGTTACCACGAAA
TATTAGAATGCCTCTCCTTACAATGTTCAACACTTCGAATGATATTGGTTATTGGGCACTA
AAGGAAAGAGGGTTTAACGGCATTCCTTGCACAGCCAAAGTGTGGTCCGACCAACTAAAA
AGCTACACCAAAGAAGCAAAATGGTTCCATGAGGGACACAAGCCAACCTTGGAGGAGTA
CTTGGACAATGCTTTGGTTTCCATAGGATTCCCAAATCTTCTCGTCACTAGCTACCTTCTCA
CAGTAGAAAATCCGACTAAGGAGAAACTCGATTACGTCAATAGCCTTCCCCTTTTTGTGCG
CGCGTCTTGCATTCTATGTCGGATCATCAATGACTTGGGAACTTCTCCAGATGAGATGGAG
AGAGGAGATAATCTCAAATCAATCCAATGCTACATGAACGAAGCTGGAGCTTCTCAAGAG
GTTGCTCGCGAGCACATAGAAGGTTTGGTTCGCATGTGGTGGAAGAGGCTAAACAAATGT

TTGTTTGAGCCGTCTCCATTCGCGGAACCATTCCTGAGTTTCACCGTGAACGTGGTGCGAG
GGTCGCACTTCTTCTACCAATACGGAGATGGGTACGGCAATGCGGAGAGCTGGACCAAAA
AGCAGGGAATGTCTGTGTTGATCCATCCAATTCCACTCAATGAAGAGTAG

3′ UTR sequence

GAGGGGATGTCATTACTTTGGTCCCAAAATAAGAGCCTGGCCATAATACTTGTAGAGGAA
GGCTTGGTGGAAACCAATTGTTCCTCAACTGAATTTTCTTTAAATTTTACCTTCTATTGTAA
TTCTTATGATTTATTGATCGAAAAATAAAAAAAAAAAAAAAAAAAAAAAA

***SaTPS2*_Full length ORF sequence (Accession number: KJ665777)**

Nucleotide Sequence

ATGGCCTCTGTGATTGTTGAACCCATTCGTTGCAACAATGACAATGATGTCATCTCGACTG
TGGTCGACGATTCGTCCGTGGTTCGGCGGGCCGCCAATTACCCGCCAAACCTGTGGGACT
ATGAATTTCTTCAGTCACTTGGCGACCAGTGCACCGTGGAGGAGAAGCATTTAAAGTTGG
CTGACAAACTGAAGGAGGAAGTGAAGTCCCTGATTAAGCAAACCATGGAGCCATTAACA
AAGTTGGAGTTCATCGATACCGTTCGGAGGCTGGGCTTAAAATACCAGTTTGAGACTGAG
GTGAAGGAAGCAGTCGTTATGGTTAGCAAATATGAGAACGATGCATGGTGGATCGACAAT
TTACATGCCACGTCTCTCCGGTTCAGGATCATGCGAGAAAACGGGATATTTGTTCCCCAAG
ATGTTTTTGAAAGGTTCAAGGACACTGATGGATTCAAGAATCAACTGTGTGAGGATGTGA
AGGGGTTGCTCAGTTTGTATGAGGCTTCATTCTTAGGCTGGGAAGGTGAGGACATCCTTGA
TGAGGCAAGGACCTTTGCAACCTCCAAACTAAAGAGCATCGAGGGGAAAATTCCCTCCCC
AAGCTTGGCCAAAAAAGTGAGCCATGCTTTGGATTTGCCGCTTCATTGGCGAACGATAAG
GTACGAGGCTAGATGGTTCATCGACACATATGAAGAAGAAGAAGACGTGAACCTTACTTT
GCTTAGATACGCTAAACTGGACTTCAACATCGTGCAATCATTTCATCAAAAGGAGATTGGT
AGATTATCAAGGTGGTGGGTGGGTACGGGCCTGGACAAGATGCCATTTGCAAGAAATGGC
TTGATACAGAGCTACATGTATGCCATAGGCATGCTGTTTGAACCGAACTTGGGAGAAGTT
AGAGAAATGGAAGCCAAAGTTGGAGCTTTGATAACGACAATCGATGATGTTTATGATGTG
TACGGAACAATGGAAGAACTAGAACTCTTCACAGATATCACCAACAGATGGGATATTAGT
AAAGCTGATCAGTTACCACGAAATATTAGAATGCCTCTCCTTACAATGTTCAACACTTCGA
ATGATATTGGTTATTGGGCACTAAAGGAAAGAGGGTTTAACGGCATTCCTTGCACAGCCA
AAGTGTGGTCCGACCAACTAAAAAGCTACACCAAAGAAGCAAAATGGTTCCATGAGGGA
CACAAGCCAACCTTGGAGGAGTACTTGGACAATGCTTTGGTTTCCATAGGATTCCCAAATC
TTCTCGTCACTAGCTACCTTCTCACAGTAGAAAATCCGACTAAGGAGAAACTCGATTACGT
CAATAGCCTTCCCCTTTTTGTGCGCGCGTCTTGCATTCTATGTCGGATCATCAATGACTTGG
GAACTTCTCCAGATGAGATGGAGAGAGGAGATAATCTCAAATCAATCCAATGCTACATGA
ACGAAGCTGGAGCTTCTCAAGAGGTTGCTCGCGAGCACATAGAAGGTTTGGTTCGCATGT
GGTGGAAGAGGCTAAACAAATGTTTGTTTGAGCCGTCTCCATTCGCGGAACCATTCCTGA
GTTTCACCGTGAACGTGGTGCGAGGGTCGCACTTCTTCTACCAATACGGAGATGGGTACG
GCAATGCGGAGAGCTGGACCAAAAAGCAGGGAATGTCTGTGTTGATCCATCCAATTCCAC
TCAATGAAGAGTAG

Protein Sequence

MASVIVEPIRCNNDNDVISTVVDDSSVVRRAANYPPNLWDYEFLQSLGDQCTVEEKHLKLAD
KLKEEVKSLIKQTMEPLTKLEFIDTVRRLGLKYQFETEVKEAVVMVSKYENDAWWIDNLHAT
SLRFRIMRENGIFVPQDVFERFKDTDGFKNQLCEDVKGLLSLYEASFLGWEGEDILDEARTFAT
SKLKSIEGKIPSPSLAKKVSHALDLPLHWRTIRYEARWFIDTYEEEEDVNLTLLRYAKLDFNIV
QSFHQKEIGRLSRWWVGTGLDKMPFARNGLIQSYMYAIGMLFEPNLGEVREMEAKVGALITT
IDDVYDVYGTMEELELFTDITNRWDISKADQLPRNIRMPLLTMFNTSNDIGYWALKERGFNGI
PCTAKVWSDQLKSYTKEAKWFHEGHKPTLEEYLDNALVSIGFPNLLVTSYLLTVENPTKEKLD
YVNSLPLFVRASCILCRIINDLGTSPDEMERGDNLKSIQCYMNEAGASQEVAREHIEGLVRMW
WKRLNKCLFEPSPFAEPFLSFTVNVVRGSHFFYQYGDGYGNAESWTKKQGMSVLIHPIPLNEE-

List of publications from thesis

1 **Prabhakar Lal Srivastava**, Ramakrishnan Krithika, Mantri T. Soniya, Pankaj Daramwar, Sharanabasappa Gangashetty, Raju Rincy, and Hirekodathakallu V. Thulasiram

Santalene synthase: A comprehensive site directed mutagenesis study to understand the dynamic nature of santalene synthase active site pocket (**Manuscript under preparation**)

2 **Prabhakar Lal Srivastava**, Pankaj P. Daramwar , Ramakrishnan Krithika, Nilofer J. Siddiqui, Avinash Pandreka, Shiva S. Shankar and Hirekodathakallu V. Thulasiram

Isolation and characterization of two isoforms of sesquisabinene B synthase from Indian Sandalwood *S. album* using transcriptome sequencing

(**Manuscript under preparation**)