Engineering Selina-4(15),7(11)-diene synthase for the Production of Novel Products

Emily Turri

Submitted in accordance with the requirements for the degree of Doctor of Philosophy

The University of Leeds

Astbury Centre for Structural Molecular Biology

September 2020

Acknowledgements

Throughout my PhD I have been extremely lucky to receive support and encouragement from family, friends, colleagues and the university. Without these people and their support, this thesis would not be possible.

First, I would like to thank my supervisors Alan Berry and Adam Nelson for their encouragement, guidance and patience over these four years. I would also like to thank Glyn Hemsworth and Nasir Khan for their continued advice and support.

I am very grateful to Glyn Hemsworth, Sheena Radford, David Brockwell and John Blacker for letting me use their equipment while offering me direction to give me the best results. In addition, I would like to thank James Ault, Rachel George, Mary Bayana, Mark Howard and Ricardo Labes for their technical assistance and biochemical analysis.

My PhD experience would have been nothing without the present and past members of the Hemrry's and Radford's with specific thanks to Jess, Alex, Jess and Dan. Thanks to everyone willing to distract me from the science with Old Bar pitchers and baby Guinness'.

Finally, I would like to thank my Mum, Dad, Grandparents and Megan, without your support I would never have been able to reach this stage. Thanks to the best quiz team, James and the Giant Peaches, to Kirsty, Wayne and Hannah for keeping in the real world.

Abstract

Terpenes are secondary natural products consisting of carbon five units produced by terpene synthase. They have been found to have functional importance ranging from medical uses (taxadiene), biofuel potential and taste and smell for cinnamon and mint. One terpene precursor substrate can produce up to 400 different products through various terpene synthase with some single enzymes producing over 50 different compounds. This broad chemical diversity is a critical focus for the engineering effort to produce novel terpenes.

Terpene synthase reactions are all initiated by the removal of a pyrophosphate to create a cation that is shaped in a hydrophobic section of the active site controlling the product structure. Little is known about how this hydrophobic region controls production, so this site is the focus of engineering work.

The aim of this project was to engineer *Streptomyces pristinaespiralis* selina-4(15),7(11)-diene synthase to produce novel compounds. Through this engineering, we could gain mechanistic insight into *Streptomyces pristinaespiralis* selina-4(15),7(11)-diene synthase and terpene synthases. *Streptomyces pristinaespiralis* selina-4(15),7(11)-diene synthase is a specific terpene synthase with selina-4(15),7(11)-diene comprising over 90% of the product profile and the second by-product germacrene B.

Using ligand binding software, a series of site directed mutants was produced with the aim of altering product binding and product distribution. An “in culture” GC/MS screen was developed to screen these targeted mutagenesis variants. Initially alanine scanning was performed at specific points in the active sites to identify potential hotspots for further targeted mutagenesis. While no novel products were found in the current screening, variant V187F showed a switch in specificity between the products and a Y152F showed a 50:50 ratio of wild-type products. The kinetic parameters of V187F and Y152F were determined and showed an increased K_m suggesting the variant reduces the substrate affinity favouring specificity of germacrene B. It was concluded that, V187 and Y152 have a key roles in the conversion of germacrene B to selina-4(15),7(11)-diene and the formation of the two rings. This mechanistic insight may be used to engineer futher terpene synthase to produce novel ring terpenes.

Table of Contents

List of Tables

List of Figures

Abbreviations

ADP	Adenosine diphosphate
AT-AS	*Aspergillus terreus* aristolochene synthase
ATP	Adenosine triphosphate
B. subtilis	Bacillus subtilis
BcBOT2	*Botrytis cinerea* terpene synthase producing presilphiperfolan-8β-ol
CoA	Coenzyme A
CTP	Cytidine 5'-triphosphate
DMAPP	Dimethylallyl pyrophosphate
DNA	Deoxyribonucleic Acid
dNTPs	Deoxyribonucleotide triphosphate
DXP	1-dexoy-D-xylulose-5-phosphate
DXR	1-dexoy-D-xylulose-5-phosphate reductoisomerase
DXS	1-dexoy-D-xylulose-5-phosphate synthase
E. coli	Escherichia coli
EI-MS	Electron impact mass spectrometry
G3P	Glyceraldehyde-3-phosphate
GC	Gas chromatography
GC/MS	Gas chromatography mass spectrometry
HMG-CoA	3-hydroxy-3-methylglutaryl-CoA
HSQC	Heteronuclear single quantum coherence spectroscopy
IDI	Isopentyl pyrophosphate isomerase (or isopentyl diphosphate isomerase)
IPP	Isopentyl pyrophosphate
IPTG	Isopropyl β-D-1-thiogalactopyranoside
IspC	1-dexoy-D-xylulose-5-phosphate reductoisomerase
IspG	4-hydroxy-3-methylbut-2-enyl-diphosphate synthase
IspH	4-hydroxy-3-methylbut-2-enyl-diphosphate reductase

LB	Lysogeny Broth
MEP	Methylerythritol phosphate pathway
MS	Mass spectrometry
MVA	Mevalonic acid pathway
NADP	Nicotinamide adenine dinucleotide phosphate
NEB	New England Biolabs
NMR	Nuclear magnetic resonance
NOESY	Nuclear overhauser effect spectroscopy
PCR	Polymerase chain reaction
PMSF	Phenylmethylsulfonyl fluoride
PPi	Pyrophosphate
PR-AS	*Penicillium roqueforti* aristolochene synthase
QM/MM	Quantum mechanics / molecular mechanics
rpm	Revolutions per minute
S. aureus	Staphylococcus aureus
S. cerevisiae	*Streptomyces cerevisiae*
S. epidermidis	Staphylococcus epidermidis
SCOPE	Structure-based combinatorial protein engineering
SDS-PAGE	Sodium dodecayl sulfate polyacrylamide gel electrophoresis
SeDS	Selina-4(15),7(11)-diene synthase
SfCinS1	*Salvia fruticose* 1,8-cineole synthase
TEMED	Tetramethylethylenediamine
VAR3-PinS	α-pinene synthase variant

Chapter 1 – Introduction

1.1 Natural Products

Natural products are defined as compounds of limited molecular weight (less than 3000 Da) produced by living organisms and can be classified into primary and secondary metabolites. Primary metabolites are directly involved in normal growth and development of the organism whereas secondary metabolites are not necessarily produced under all conditions (Dewick, 2011). Therefore, secondary metabolites can be described as non-essential molecules, not required for the organism's normal metabolism (Figure 1-1) (Dar *et al.*, 2017). A wide variety of organisms including animal, bacteria and fungi produce a wide variety of natural products with a range of functions. These functions range from defensive responses to toxic products. The chemical diversity of natural products comes from the diversity of the organisms and their biological activities (Bernardini *et al.*, 2018). There is a large potential number of uncharacterised natural products in unexplored species and organisms' silent biosynthetic gene clusters (Boufridi and Quinn, 2018).

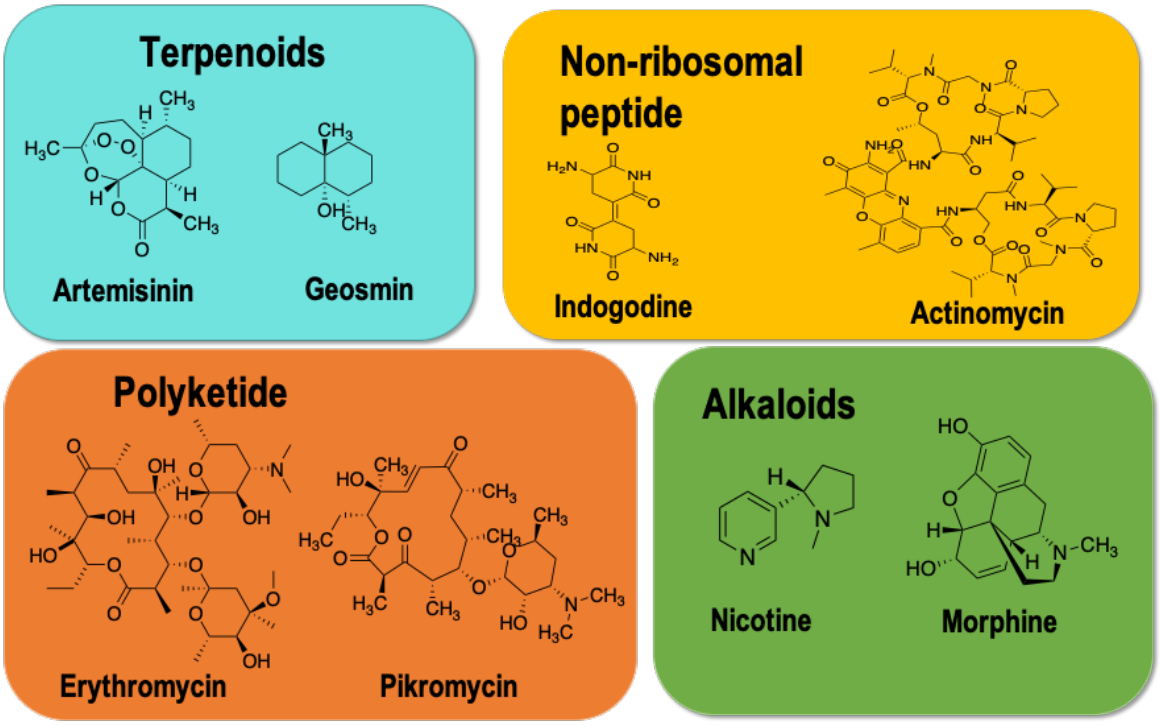


Figure 1-1 The four main structural classes of natural products. *Structural examples for each of the four main classes of natural products or secondary metabolites*.

Natural products have been a large source for pharmaceuticals with approximately half of all drugs in current use being natural products or natural product derivatives (Dar *et al.*, 2017). Of the 175 anticancer drugs that were FDA approved between 1940 and 2014, 63% were

natural products or natural product derivatives (Boufridi and Quinn, 2018). On average natural products have higher oxygen content and a lower nitrogen content when compared to the potential chemical space, while favouring aliphatic over aromatic structures with 38% of characterised natural products containing arenes (Rodrigues *et al.*, 2016).

Natural products have been used as medicinal sources throughout history, with a significant shift towards natural products following the isolation of penicillin in 1928 by Alexander Fleming and its subsequent development for clinical use in the 1940s. Following this the focus was on the isolation of products from microbial sources and the discovery of antibiotics including; tetracycline and streptomycin (Bernardini *et al.*, 2018) (Carlson, 2010). Preceding the discovery of penicillin, Friedrich Sertürner discovered alkaloids by isolating morphine from the poppy plant in 1817 and the consequent isolation of other alkaloids, caffeine and nicotine from plant sources (Bernardini *et al.*, 2018). In addition, the first natural compound produced by chemical synthesis was salicylic acid in 1853 by Charles Gerhardt leading to the development of the drug aspirin (Jeffreys, 2004). Following these advances a large focus was put on the isolation of natural products until the second half of the 20th century when pharmaceutical efforts shifted towards the use of combinatorial chemistry and the production of libraries by chemical synthesis (Carlson, 2010). However, these libraries did not have the hoped-for impact when compared to previous natural product libraries results with a several fold higher hit rate (Newman and Cragg, 2016). Recent efforts show 43% of published clinal candidates in 2016 and 2017 were derived from previously known compounds compared to 29% identified from random high throughput screening (Brown and Boström, 2018). So, in recent years the focus has shifted back towards natural products and their derivatives with the advancement of emerging genetic technologies.

1.2 Terpenes

Terpenes are a range of chemically diverse compounds consisting of isoprene units with over 80,000 chemically distinct characterised products (Figure 1-2) (Byres *et al.*, 2007) (Harms, Kirschning and Dickschat, 2020) (Leferink *et al.*, 2019). The number of characterised terpenes is increasing due to emerging genome analysis, genome mapping and biosynthetic tools (Citron *et al.*, 2012). Terpene diversity is derived from cation quenching and the range of hydrocarbon backbones (Zhou and Peters, 2011). This assembly mechanism leads to a diverse range of linear and chiral skeletons from a limited number of substrates, with one substrate producing over 400 characterised products (Ajikumar *et al.*, 2008).

Terpene substrates are produced from sequential isomerisation of isoprene units. As terpenes are produced from carbon five isoprene units, they can be categorised by constituent number. A single isoprene unit is found in the hemiterpenes and two isoprene units are monoterpenes (Figure 1-2). Three isoprene units are sesquiterpenes and four isoprene units are diterpenes (Figure 1-2). Both five and seven isoprene units, sesterpenes and sesquaterpenes respectively, are rare in nature, however both six and eight isoprene units, triterpenes and tetraterpenes respectively, are common in nature (Figure 1-2). Larger terpenes are produced in nature for example natural rubber consists of cis polyisoprene units (Gao, Honzatko and Peters, 2012).

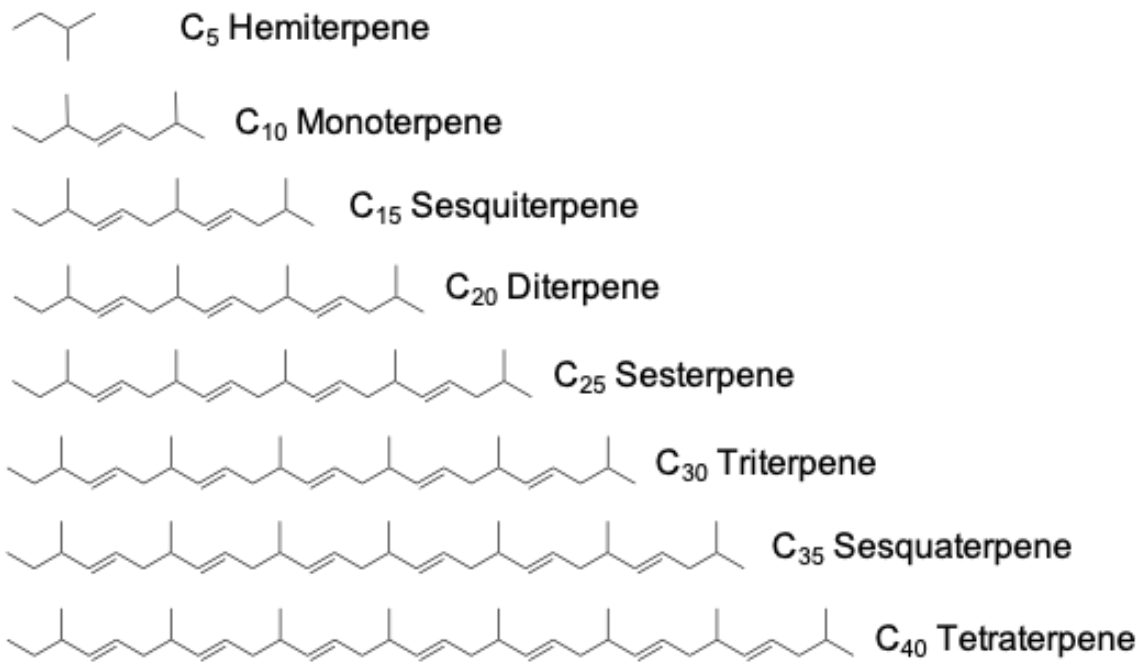


Figure 1-2 Schematic of terpene nomenclature. *Schematic of increasing numbers of C5 isoprene units and the associated nomenclature.*

Terpenes are produced by a variety of organisms including bacteria, plants, yeast and fungi. In addition, recently terpenes have been found in *Dictyostelium discoideum,* a social amoeba with various terpenes dependent on their developmental stage (Rabe *et al.*, 2016).

Terpenoids are functionally and chemically modified terpene skeletons, modified by tailoring enzymes including; P450 and acyl transferases (Ajikumar *et al.*, 2008) (Rinkel *et al.*, 2016). Terpenes are volatile compounds and have a high vapour pressure at room temperature (Ajikumar *et al.*, 2008).

1.2.1 Functional importance of terpene's

Characterised terpenes have a wide variety of functional roles within the producing organism and to the benefit humans. These roles include; antimicrobial, antifungal, antiviral, antiparasitic and potential biofuels (Ajikumar *et al.*, 2008). One of the most notable terpenoid's is paclitaxel, an anticancer agent that acts by binding tubulin heterodimers stopping cell division (Figure 1-3a). Paclitaxel is a diterpenoid extracted from bark of yew trees *Taxus chinenis* that can be used to treat breast and ovarian cancers (Edgar *et al.*, 2017) (Ajikumar *et al.*, 2008). Other terpenes with anticancer activity include D-limonene and perillyl alcohol that inhibit post-translational isoprenylation of proteins regulating cell growth and can be used to treat a range of cancers (Figure 1-3b and c) (Ajikumar *et al.*, 2008).

Another medically important terpenoid's is artemisinin extracted from *Artemisia annua* to treat malaria using the epoxide ring in Figure 1-1 (Withers and Keasling, 2007) (Tu, 2017). Various antibacterial diterpenes have been extracted from *Salvia* species acting against bacterial species which include; *S. aureus, S. epidermis, B. subtilis, E. coli.* In addition to antibacterial activity, terpenoids can have antifungal activity for example carvone is antifungal agent which is active against *Candida albicans,* a human pathogenic fungi (Figure 1-3d). Terpenes and terpenoids have natural advantages as drugs with low toxicity and low skin irritation along with good penetration (Gao, Honzatko and Peters, 2012). In addition there are a variety of terpene functions that are not medically relevant such as a hydrogenated bisabolene, bisabolane is a potential biofuel with similar properties to diesel fuel D2 (Figure 1-3e) (McAndrew *et al.*, 2011). Terpenoids are also the molecules responsible for the taste and scents in cinnamon and mint (Fujisawa *et al.*, 2010).

Terpenes also have vital roles within the organism that produce them including; defence, attraction, ecological communication between organism and cell regulation (Miller and Allemann, 2012). For example (E)-α-bisabolene from *Abies grandis* is produced following wound inducing signals as shown from stem wounding experiments (Bohlmann *et al.*, 1998). Terpenoids constitute a large range of plant hormones with a variety of key functions. Stringolactones are essential plant terpenoid hormones that regulate roots structures and branching (Figure 1-3f) (Chen, Zhang and Lindley, 2019). Carotenoids are the light harvesting terpenoids, involved in regulation of photosynthesis and protect plants from the effect of UV radiation (Figure 1-3g) (Chen, Zhang and Lindley, 2019). Volatile terpenes can be released to attract pollinators for example (E)-β-ocimene and myrcene are released from snapdragon flowers in floral development (Figure 1-3h and i) (Tholl, 2006). Other plant hormones are used for defence and therefore can be used as environmentally friendly insectides (Ajikumar *et al.*, 2008).

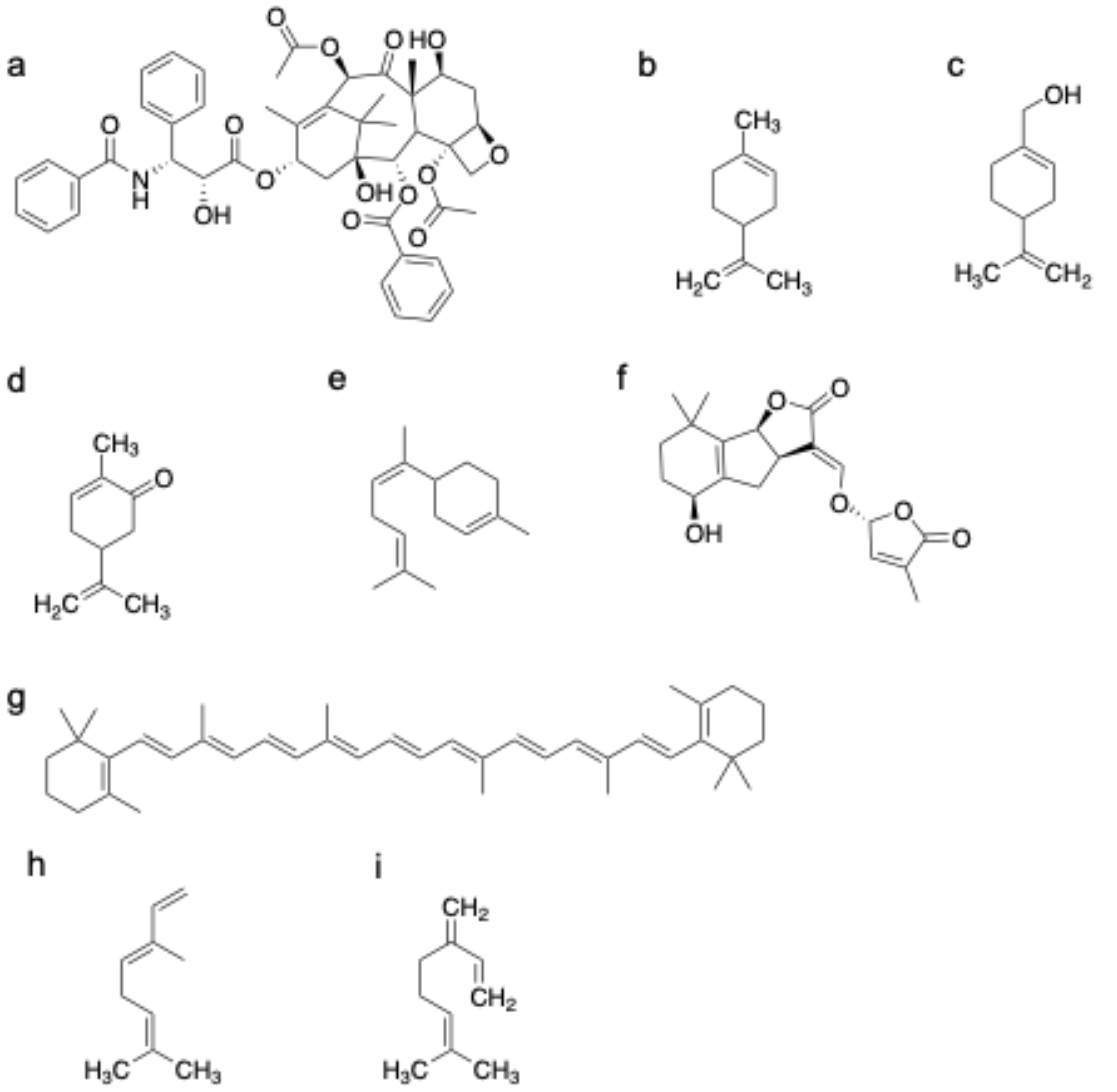


Figure 1-3 Schematic of functionally important terpenes. *(a) Structure of paclitaxel. (b) Structure of d-limonene. (c) Structure of perilly alcohol. (d) Structure of carvone. (e) Structure of bisabolene. (f) Structure of (+)-strigol a stringolactone. (g) Structure of β-carotene a carotenoid. (h) Structure of (E)- β-ocimene. (i) Structure of myrcene.*

1.3 Isoprene unit production

Isoprenes are five carbon units with the molecular formula C_5H_8 and are produced by combinations of isopentyl pyrophosphate and dimethylallyl pyrophosphate (Tholl, 2006). The isoprene precursor compounds are produced by two separate pathways; the mevalonic acid (MVA) pathway (Figure 1-4) and the methylerythritol phosphate (MEP) pathway (Figure 1-5). The MVA pathway is active in the cytosol of bacteria, plants, fungi and other eukaryotes mainly for the production of triterpenes and sesquiterpenes. The MVA pathway consist of 7 enzymatic reactions. Whereas the MEP pathway is active in plants and bacteria to supply precursors mainly for monoterpene and diterpene production (Roberts, 2007) (Ajikumar *et*

al., 2010). This second pathway consists of 8 reaction steps. Mechanisms for both pathways were determined using isotopic labelling experiments and are described below (Figure 1-4 and 1-5) (Dickschat, 2017).

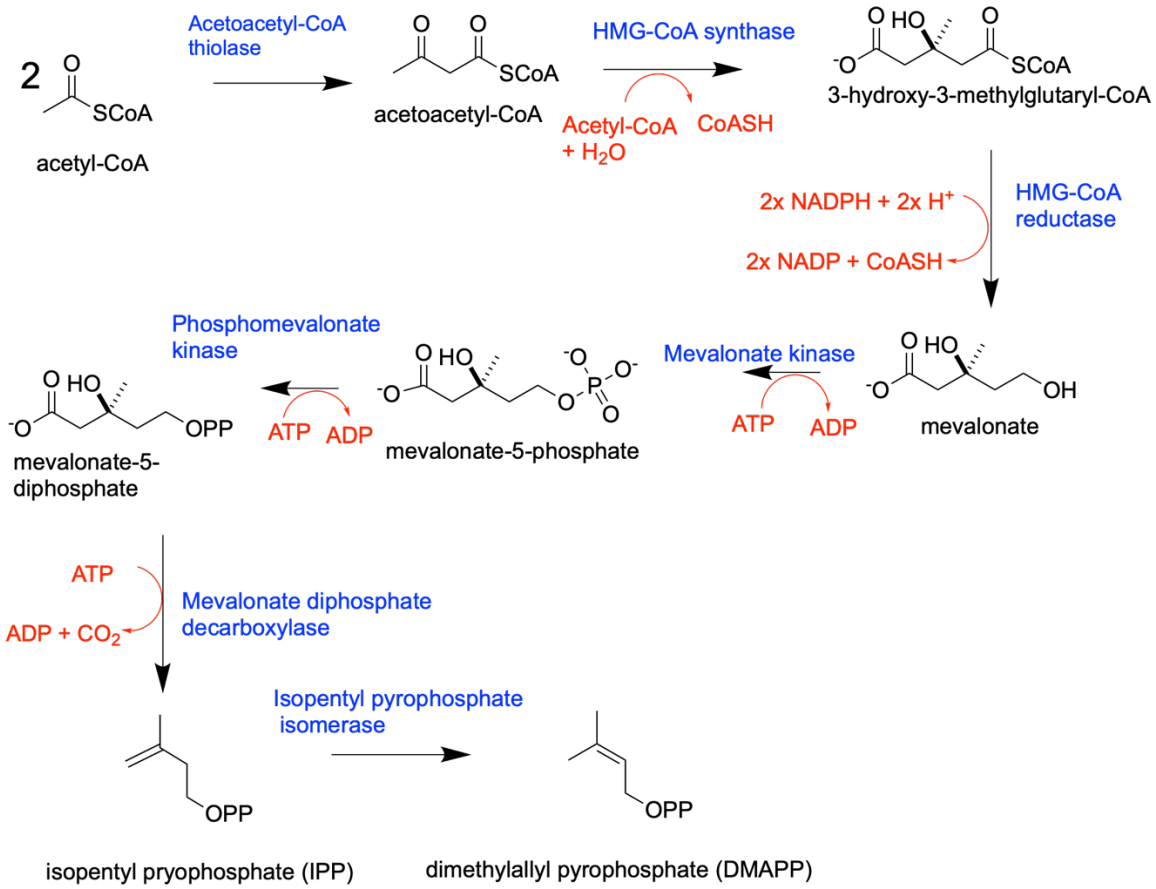


Figure 1-4 Schematic of the mevalonate pathway producing isopentyl pyrophosphate (IPP) and dimethylallyl pyrophosphate (DMAPP). *The pathway begins with a thioester Claisen condensation reaction joining two molecules of acetyl-CoA forming acetoacetyl-CoA. Next is a thioester Claisen condensation reaction by 3-hydroxy-3-methylglutaryl-CoA synthase (HMG-CoA synthase) joining acetoacetyl-CoA to an acetyl-CoA forming 3-hydroxy-3-methylglutaryl-CoA (HMG-CoA)* (Ajikumar *et al.*, 2008). *HMG-CoA synthase contains a nucleophilic cysteine residue used to acetylate the enzyme releasing the reduced co-enzyme A followed by a nucleophilic attack of acetyl-CoA* (Theisen *et al.*, 2004). *Next HMG-CoA is reduced and the co-enzyme A is released forming mevalonate using HMG-CoA reductase and NADPH* (Haines, Wiest and Stauffacher, 2013) (Ajikumar *et al.*, 2008). *Mevalonate and phosphomevalonate kinases perform sequential phosphorylation of mevalonate using two molecules of ATP forming mevalonate-5-phosphate and then mevalonate-5-diphosphate. Mevalonate-5-diphosphate is converted to IPP through an ATP dependent decarboxylation in a two-step reaction. The first step of reaction is phosphorylation to produce a reactive intermediate which undergoes dephosphorylation and decarboxylation as the second step* (Byres *et al.*, 2007). *To convert IPP to DMAPP isopentyl pyrophosphate isomerase utilises a protonation / deprotonation mechanism to isomerise the double bond. OPP is used on drawing IPP and DMAPP to signify pyrophosphate.*

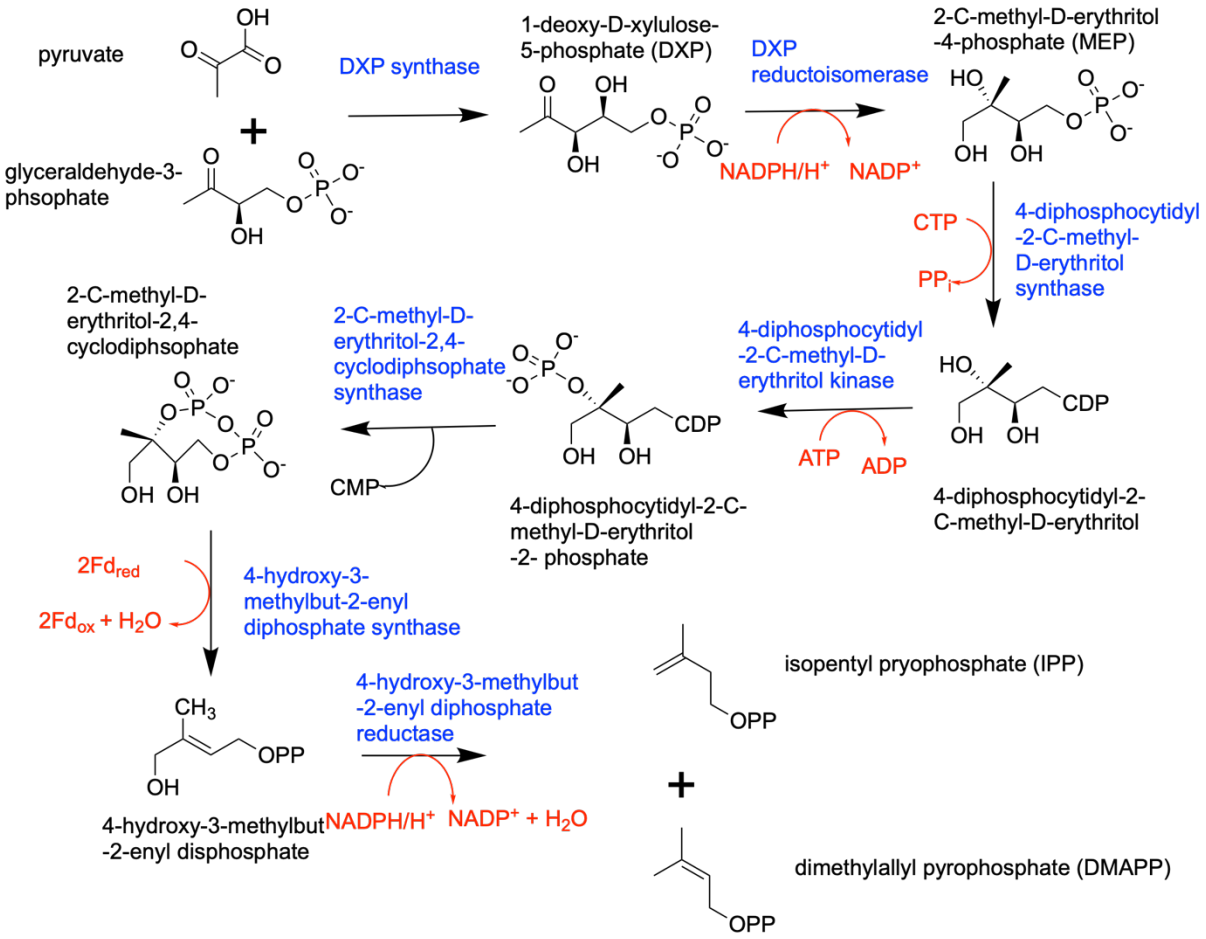


Figure 1-5 Schematic of the methylerythritol phosphate pathway producing isopentyl pyrophosphate and dimethylallyl pyrophosphate. *Pyruvate and glyceraldehyde-3-phosphate (G3P) undergo acyloin condensation to produce 1-deoxy-D-xylulose-5-phosphate (DXP) by DXP synthase and a thiamine pyrophosphate cofactor (UniProt, 2017)* (Ajikumar *et al.*, 2008). *DXP undergoes and NADP dependent reduction and rearrangement under the control of DXP reductoisomerase (IspC) forming 2-C-methyl-D-erythritol-4-phosphate* (Ajikumar *et al.*, 2008). *After 4-diphosphocytidyl-2-C-methyl-d-erythritol synthase utilises a nucleophilic attack mechanism to join cytidine diphosphate to 2-C-methyl-D-erythritol-4-phosphate forming 4-diphosphocytidyl-2-C-methyl-d-erythritol* (Jin *et al.*, 2016). *4-diphosphocytidyl-2-C-methyl-D-erythritol kinase utilises ATP to phosphorylate 4-diphosphocytidyl-2-C-methyl-d-erythritol forming 4-diphosphocytidyl-2-C-methyl-d-erythritol-2-phosphate. Following this 2-C-methyl-D-erythritol-2,4-cyclodiphsophate synthase catalyses the removal of the CMP forming 2-C-methyl-D-erythritol-2,4-cyclodiphsophate* (Richard *et al.*, 2002) (Kemp, Bond and Hunter, 2002). *2-C-methyl-D-erythritol-2,4-cyclodiphsophate was then oxidised 4-hydroxy-3-methylbut-2-enyl diphosphate by 4-hydroxy-3-methylbut-2-enyl-diphosphate synthase (IspG). 4-hydroxy-3-methylbut-2-enyl diphosphate reductase (IspH) then produces IPP and DMAPP which are used further in terpene production.*

Both pathways utilise common cell metabolism products to produce the isomers IPP and DMAPP. Isopentyl-diphosphate isomerase (or isopentyl pyrophosphate isomerase) (IDI) stereospecificially converts IPP to DMAPP by isomerising the carbon-carbon double bond (Ajikumar *et al.*, 2008). There are two types of IDI enzymes; IDI-I found in eukaryotes and some bacteria (including *E. coli*) and IDI-II found in some *Streptomyces* strains, archaea and bacteria. Both types are monomeric, divalent cation-dependent enzymes with alternating α-helices and β-sheets creating a buried active site (Zheng *et al.*, 2007). Metal ion coordination and key glutamate residues stabilise the carbocation formed by the initial protonation in the reaction mechanism (Zheng *et al.*, 2007).

1.3.1 Prenyltransferases

Prenyltransferases catalyse the synthesis of linear prenyl pyrophosphates by the head to tail condensation of DMAPP and IPP in a sequential manner to produce products of various terpene precursors lengths (Tholl, 2015). The general mechanism of action is the elimination of a pyrophosphate ion to form an allylic cation which is subsequently attacked by an IPP molecule and the stereospecific removal of a proton from the newly formed carbon-carbon bond (Figure 1-6) (Tholl, 2015) (Ajikumar *et al.*, 2008). Cis- or trans- stereochemistry of the double bonds in the prenyl pyrophosphate product determines whether the enzyme operates as a cis or trans prenyltransferase. The majority of work has focused on trans-short chain prenyltransferases producing monoterpene, sesquiterpene and diterpenes but there is evidence of short chain cis prenyltransferases (Tholl, 2015).

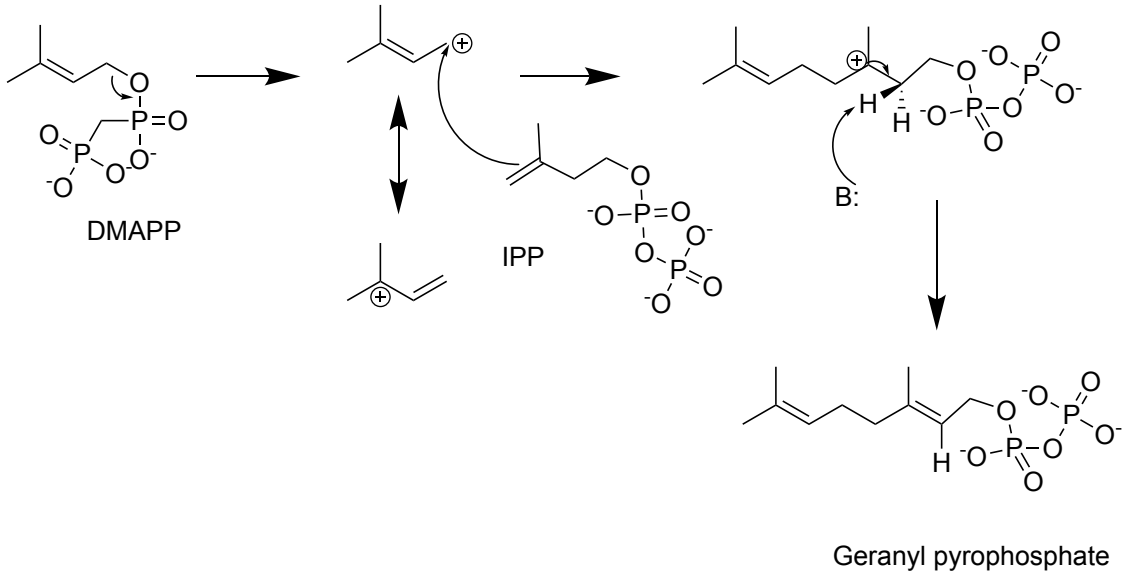


Figure 1- 6 Schematic shows the sequential addition of IPP units in a head-to-tail condensation. *The schematic shows head-to-tail condensation of IPP and DMAPP. The DMAPP is ionised as the pyrophosphate leaves creating an electrophile. This electrophile is then attacked by the double bond in IPP creating a new carbon bond. Consequently a proton is abstracted to remove the positive charge on the cation using a base provided in the enzyme creating a double bond (Kennepohl, 2016).*

Short-chain prenyltransferases share a common 13 α-helices protein fold with 10 α-helices forming the active site. Substrates are bound within the active site that contains 2 highly conserved aspartate-rich regions; DDx2-4D (FARM or first aspartate-rich-motif) and DDxxD (SARM or second aspartate-rich-motif) binding divalent metal ions. The product chain length is in part regulated by the amino acids upstream of the FARM in the hydrophobic substrate binding region: the bulkier the residues the shorter the chain. This is shown when comparing farnesyl pyrophosphate synthase's bulkier aromatic residues and geranylgeranyl pyrophosphate synthase's smaller residues in the same locations. In addition, the second layer of active site residues can provide additional active site flexibility and can influence the chain lengths (Tholl, 2015) (Christianson, 2017).

1.4 Terpene synthase

Terpene synthases catalyse the cyclisation reactions of isoprene-based substrates. The ionisation of a pyrophosphate ester generates an allylic carbocation. Once the highly reactive cation is formed it can undergo a reaction cascade and structural rearrangement before final reaction termination by proton elimination or solvent addition (Christianson, 2017). The reaction cascades include; hydride transfers, methyl migrations and

rearrangements (Christianson, 2017). Folding of the substrate carbocation within the active site may act as a template for the product (Zhou and Peters, 2011). This folding brings the carbocation and the C-C cyclisation bond in close proximity for initiating ionisation and cyclisation due to anchimeric assistance or interaction of an electron pair and its adjacent reaction centre (Zhou and Peters, 2011). Terpene synthases catalyse reactions with high regio- and stereospecificity and produce enantiomerically pure products (Rinkel *et al.*, 2016). Some terpene synthases are highly specific producing one specific product and some produce product mixtures (Tholl, 2006). Due to their reaction mechanism and the stereospecificity of products terpene synthases are mechanistically and chemically interesting enzymes (Dickschat, 2017).

All terpene synthases have tertiary homology containing catalytically critical aspartate residues and a hydrophobic pocket but have low primary sequence homology (Wendt and Schulz, 1998) (Shishova *et al.*, 2007). The primary sequence homology is conserved to the aspartate rich motifs that coordinate divalent cations to stabilise the pyrophosphate moiety, predominantly magnesium ions (Rynkiewicz, Cane and Christianson, 2001) (Little and Croteau, 2002) (McAndrew *et al.*, 2011) (Gao, Honzatko and Peters, 2012). The binding of divalent cations is postulated to induce the conformational changes needed to close and sequester the active site (Vedula *et al.*, 2005). The hydrophobic pocket is the lower region of the active site and is predominantly lined with hydrophobic and aromatic residues that create a template for shaping and orienting the substrate. The aromatic residues create an electrostatic effect to stabilise the cation through cation-pi interactions. (Dougherty, 1996). However the secondary layer of residues for the active site can generate further diversity with an effect on the active site contour (Lodeiro *et al.*, 2004). All terpene synthases have these key features but otherwise can be classified in two classes; I and II (see section 1.4.1 and 1.4.2). Terpene synthases can also be classed as specific or promiscuous enzymes depending on the number of products produced.

1.4.1 Class I terpene synthases

Class I terpene synthases have a classic α-helical terpene synthase fold called the α-domain. The active site contains two conserved aspartate rich motifs that bind divalent cations; DDxxD motif and NSE or DTE motif (Figure 1-7). The DDxxD motif is typically found on helix D. The NSE or DTE motif is found on helix H with the sequence (N,D)D(L,I,V)X(S,T)XXXE (Christianson, 2017). These motifs resemble the motif in prenylelongases, leading to a similar catalytic mechanism with ionisation of the allylic diphosphate ester bond and pyrophosphate removal forming a cation (Figure 1-7) (Gao,

Honzatko and Peters, 2012). The quenching of the cation, enzyme activity and active site contour produces the diversity in the products (Tholl, 2015). Class I terpene synthases are traditionally monoterpene synthase, sesquiterpene synthase, sesterterpene synthase and some diterpene synthase (Christianson, 2017) (Tholl, 2006).

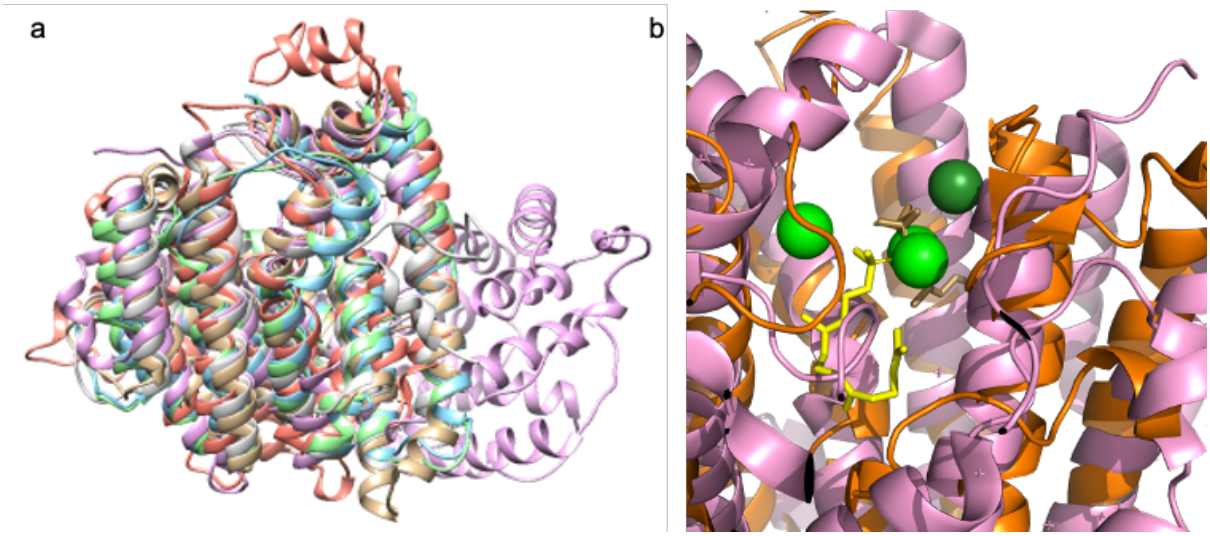


Figure 1-7 Alignment of class I active sites with farnesyl pyrophosphate synthase. *(a) Alignment of class I active sites of Selina-4(15),7(11)-diene synthase in grey (PDB 4OKM), pentalene synthase in gold (PDB 1PS1),* Aspergillus terreus *aristolochene synthase in blue (PDB 2OA6*), Penicillium roqueforti *aristolochene synthase in green (PDB 1DI1), 5-epi-arisotolchene synthase in pink (PDB 51K9) and farnesyl pyrophosphate synthase in orange (PDB 1UBX) using SALIGN. (b) Active site alignment in pymol of farnesyl pyrophosphate synthase (PDB 1UBX) with magnesium ions and farnesyl pyrophosphate and* Nicotiana tabacum *5-epi-aristolochene synthase (PDB 5IK9) with farnesyl monophosphate. Overall the active sites helices align well in shape and the magnesium ions can coordinate both bound substrates.*

Monoterpene synthases catalyse the cyclisation of the C_{10} isoprenoid substrate geranyl pyrophosphate. All cyclisation reactions proceed through the formation of the α-terpinyl carbocation resulting from the C1-C6 bond formation (Christianson, 2017). The diverse products are then controlled by the enzyme active site contour acting as a product template, with the more product-like the template the more high-fidelity the terpene synthase (Christianson, 2017). Previously a number of monoterpene synthases have been structurally and functionally characterised producing various crystal structures. These crystal structures suggest some monoterpene synthases have a coordinated solvent molecule for further diversity in the termination products (Christianson, 2017). One example of a monoterpene synthase is bornyl diphosphate synthase, a homodimer with multiple crystal structures

containing an active site water. However this active site water is not directly involved in the termination of the product but in the coordination of the cation (Christianson, 2017).

Specific sesquiterpene synthases cyclise the C_{15} farnesyl pyrophosphate into over 400 products through six initial cyclisation routes compared to monoterpenes one (Yamada *et al.*, 2015). These cyclisation events will either form; 1,10-(*E*,*E*)-germacradienyl cation, 1,11-(*E*,*E*)-humulyl cation or nerolidyl cation leading to a 1,6-bisabolyl cation, 1,7-cycloheptenyl cation and (*E*,*Z*)- germacradienyl and humulyl cations (Figure 1-8) (Harms, Kirschning and Dickschat, 2020). Many sesquiterpene synthases have been structurally and functionally characterised and show a diverse range of products and features all with the core class I components. For example, pentalene synthase is a monomeric sesquiterpene synthase whereas trichodiene synthase is an antiparallel homodimer (Lesburg *et al.*, 1997) (Rynkiewicz, Cane and Christianson, 2001). These differences are even seen between the same enzyme but from different species as aristolochene synthase from *Penicillium roqueforti* is a monomer and produces side products germacrene A and valencene whereas that from *Aspergillus terreus* is a tetramer with tight specificity (Caruthers *et al.*, 2000) (Shishova *et al.*, 2008) (Yu, Miller and Allemann, 2007). When comparing the active site of each enzyme the active site contour shape is similar suggesting a similar level of control on their final product with differences seen at the secondary layer of the active site contour (Shishova *et al.*, 2007).

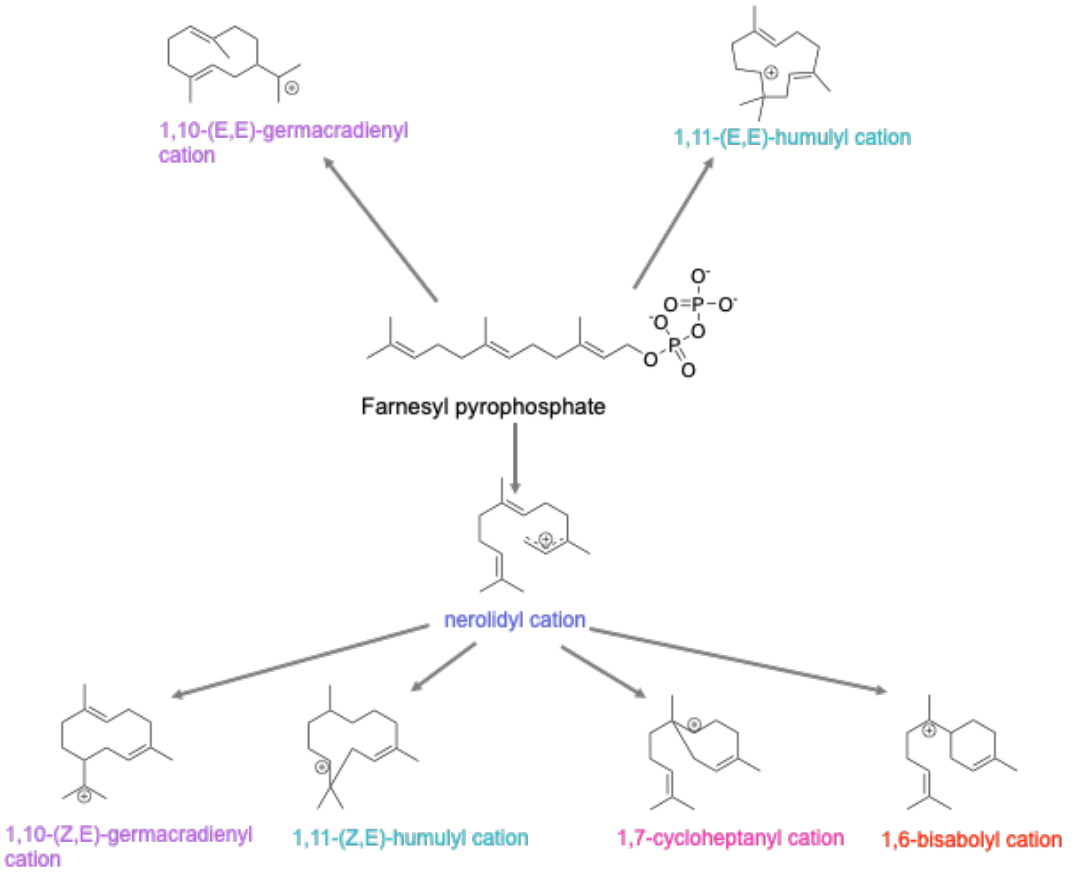


Figure 1-8 Schematic of the sesquiterpene cyclisation cations. *Schematics of the six cyclisation mechanisms from farnesyl pyrophosphate and the six distinct cations which are the basis of the over 400 sesquiterpene products.*

1.4.2 Class II terpene synthases

Class II terpene synthases all contain a hydrophobic pocket and an aspartate motif within their enzyme active sites similar to class I. However only one aspartate motif, DXDD, is found in class II active sites, unrelated to the class I aspartate motif (Christianson, 2017). Structurally, class II enzymes have a minimum of two domains, β and γ. Both domains are made up of α helices and the class II active site is at the interface of the two domains (Christianson, 2017). Class II terpene synthases initiate reactions differently from class I by protonating to C-C or π bond in the epoxide moiety forming a carbocation that moves towards the hydrophobic pocket of the active site (Christianson, 2017). Class II terpene synthases include triterpene synthase and some diterpene synthases.

Triterpene synthases act on C_{30} isoprenoid substrates and due to the substrate size the enzyme active site contour is critical to the product (Christianson, 2017). Triterpene synthases initiate reactions by protonating the terminal π bond of squalene or the terminal epoxide moiety of squalene oxide creating a cation for the cascade of carbon-carbon bond

forming and rearranging reactions (Christianson, 2017). One example of a triterpene synthase is squalene-hopene cyclase which has been extensively studied and crystallised in a variety of mechanistic conformations. In addition, squalene-hopene cyclase can also be active (Bronsted acid activation) with C_{15}, C_{20}, C_{30} and C_{35} substrates. An additional example is oxidosqualene cyclase with the expected $\beta\gamma$ domain structure homologous to squalene-hopene cyclase. Through computational and crystallographic studies, a mechanism has been postulated for oxidosqualene cyclase initiated by protonation (Christianson, 2017).

Diterpene synthases act on geranylgeranyl pyrophosphate and initiate reactions by protonating the terminal π bond of geranylgeranyl pyrophosphate. Generally this initiation goes along with the formation of both the C10-C15 and C6-C11 bonds forming the bicyclic intermediate before the cascade of further reactions (Christianson, 2017). One diterpene synthase has been extensively studied in two organisms ent-Copalyl diphosphate synthase from *Arabidopis thaliana* and *Streptomyces platensis* CB00739 (Sun Tai ping and Kamiya, 1994) (Rudolf *et al.*, 2016). The bacterial enzyme has a $\beta\gamma$ domain architecture whereas, the plant enzyme has a $\alpha\beta\gamma$ domain with a non-functional α domain. However using crystal structures with a variety of substrates and sequence comparison the same mechanism of action has been postulated for both enzymes (Christianson, 2017).

1.4.3 Bifunctional terpene synthases

Bifunctional terpene synthases catalyse tandem or coupled cyclisation reactions in two distinct active sites on the same protein. Previously it was thought these reactions are only tandem diterpene cyclisation reactions but other examples have been found more recently (Christianson, 2017). In these bifunctional enzymes there is not always a channel to direct the product of one reaction to the second active site. One example geosmin synthase, is an $\alpha\alpha$ domain bifunctional enzyme that first catalyses the cyclisation of farnesyl pyrophosphate to germacradienol and minor amounts of germacrene D. Following this a protonation dependent cyclisation-fragmentation produces C_{12} geosmin and acetone (Cane *et al.*, 2006). Geosmin synthase has two class I terpene synthase α domains both with bound magnesium ions with the *N*-terminal domain catalysing the first reaction and *C*-terminal the second reaction (Christianson, 2017). Another example is abietadiene synthase which contains a class I and a class II active site cyclising geranylgeranyl pyrophosphate (Christianson, 2017). The first reaction is the initial cyclisation of geranylgeranyl pyrophosphate copalyl pyrophosphate in a class II active site by protonation. The second reaction occurs in the class I active site with bound magnesium ions to form a tricyclic carbocation and consequently abietadiene (Christianson, 2017).

1.5 Terpene synthase structures

As discussed terpene synthases have low primary sequence homology but all have catalytically critical aspartate residues and a hydrophobic active site (Degenhardt, Köllner and Gershenzon, 2009). The catalytic importance of aspartate rich motifs was established due to the loss of activity and loss of efficiency through the decrease in k_{cat} and increase in K_m (Yoshikuni, Ferrin and Keasling, 2006) (Felicetti and Cane, 2004). The hydrophobic active sites are electron rich and unreactive but have a key role in shaping and orientating the cation into productive conformations (Lodeiro *et al.*, 2004) (Li *et al.*, 2014) (Vedula *et al.*, 2008). Generally terpene synthase reactions are successful as a more product like isoprenoid analogue is more thermodynamically favourable (Christianson, 2008).

1.5.1 Class I terpene synthase structures

Class I terpene synthases have a characteristic α-helical domain called the classic terpene synthase fold by Oldfield (Harris *et al.*, 2015). Generally the key DDxx(D,E) motif is located on helix D and coordinates magnesium ions A and C whereas the DTE/NSE motif is on helix H and binds magnesium ion B (Figure 1-9) (Miller and Allemann, 2012) (Degenhardt, Köllner and Gershenzon, 2009). From structural analysis it is postulated that the closing of the active site cap is dependent on metal ion binding and not substrate binding (Rynkiewicz, Cane and Christianson, 2002). In addition even though terpene synthases have low primary sequence homology there is a conserved arginine residue 46 residue upstream of the NSE motif which is postulated to be a pyrophosphate binding sensor (Harms, Kirschning and Dickschat, 2020).

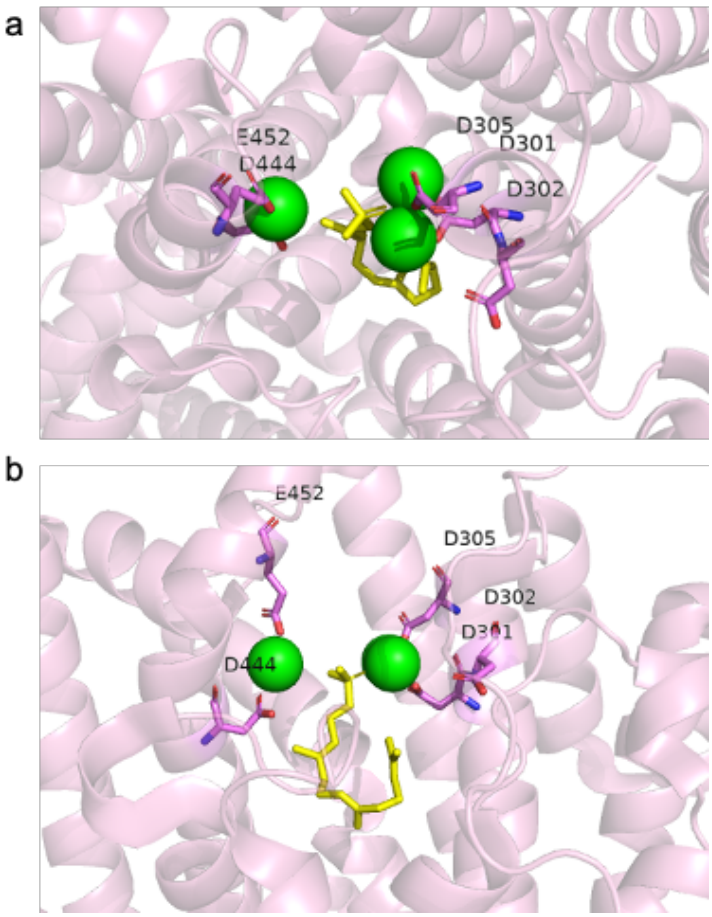


Figure 1-9 Class I active site with key aspartate motif residues highlighted*. Active site of* Nicotiana tabacum *5-epi-aristolochene synthase (PDB 5IK9) with farnesyl monophosphate and three magnesium ions bound. Highlighted are the key aspartate residues D301, D304, D305, D444 and E452 in the critical aspartate motifs. These residues are shown to be involved in the binding and control of three magnesium ions and consequently the phosphate moiety on farnesyl monophosphate. D301, D304 and D305 are within the DDxxD and D444 and E452 are part of the NSE motif.*

Limonene synthases are monoterpene synthase and have been found and studied in various organisms. Despite all isotypes having a sequence identity below 50%, they all have similar tertiary structural features (Figure 1-10). Two structural examples are (+)-limonene synthase from *Citrus sinensis* and (4S)(-)-limonene synthase from *Mentha spicata* (Christianson, 2017). Both limonene synthases have a 2 α-helical domain architecture with both domains adopting similar overall folds (Figure 1-10). The *C*-terminal domain contains the monoterpene synthase active site and has all the characteristic features of a class I terpene synthase. (+)-Limonene synthase *C*-terminal domain terpene fold consists of 12 helices with the enzyme active site consisting of mostly nonpolar, hydrophobic and aromatic amino acids but is flanked on either side with aspartate motifs (Figure 1-10b) (Morehouse *et*

al., 2017). The enzyme contour of both enzymes is similar to orient the carbocation in the same manner. Both *N*-terminal domains have no known function except being involved in the closing of the *C*-terminal active site. The (+)-limonene synthase has a pair of arginine residues involved in sensing the closure of the active site (Morehouse *et al.*, 2017). Whereas in (4S)(-)-limonene synthase the *N*-terminal domain folds back to form a component of the active site cap with the J and K helices (Figure 1-10c) (Hyatt *et al.*, 2007). Both limonene synthases have been crystallised with a variety of substrate and intermediate analogues including; 2-flurogeranyl pyrophosphate and 2-fluorolinalyl pyrophosphate in which the electron withdrawing fluorine atoms make the analogues unreactive (Hyatt *et al.*, 2007).

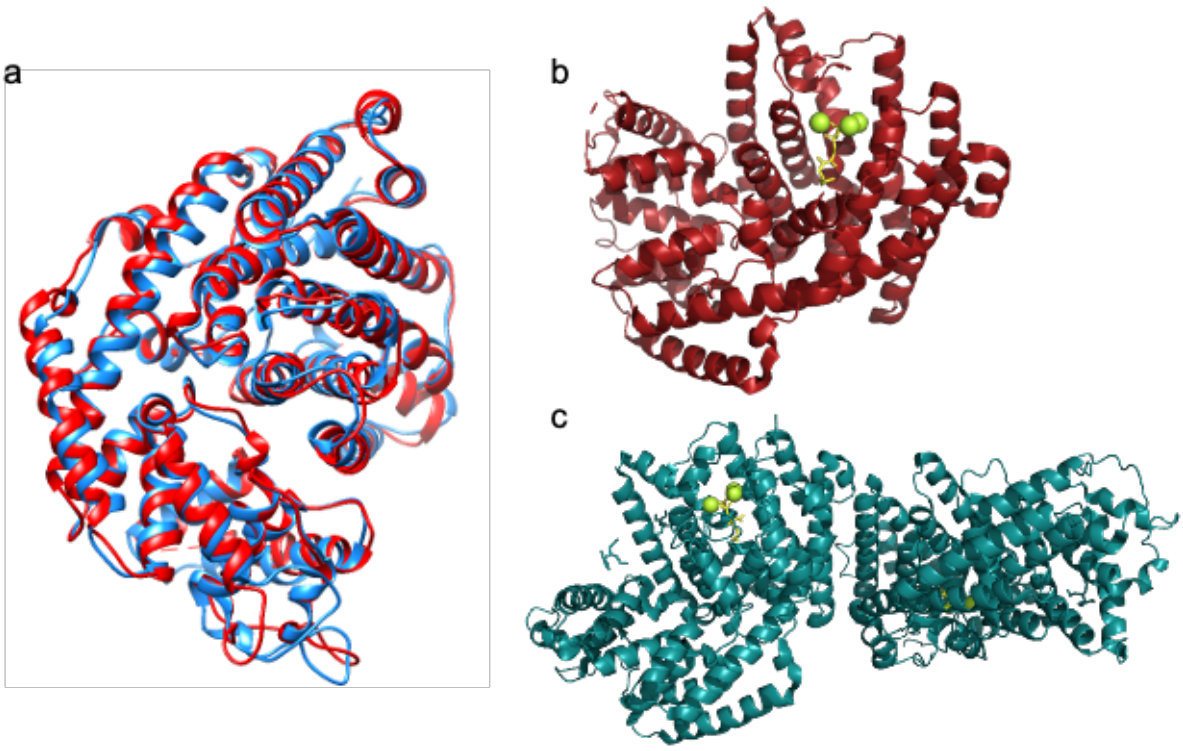


Figure 1-10 Comparison of limonene synthase structures. *(a) SALIGN alignment of + limonene synthase from* Citrus sinensis *in red (PDB 5UV1) and chain A of – limonene synthase from* Mentha spicata *in blue (PDB 2ONG). The alignment has a score of 96.5% showing how similar the two structures and active site are. (b) Structure of +limonene synthase from* Citrus sinensis *(PDB 5UV1) with the manganese ions used for crystallisation displayed in green and 2-flurogeranyl pyrophosphate bound in the active site displayed in yellow. (c) Structure of (4S)(–)-limonene synthase from* Mentha spicata *(PDB 2ONG) with the magnesium ions used for crystallisation displayed in green and 2-flurogeranyl pyrophosphate bound in the active site displayed in yellow.*

Sesquiterpene synthases have been extensively studied. They have a large structural variation especially at a quaternary level, but all contain the key class I terpene synthase features. Pentalene synthase is the smallest sesquiterpene synthase as an α-terpene domain monomer with 11 α-helices (Figure 1-11d). Five of these 11 α-helices with longer loops surround the active site with a cavity 15 Å deep and 9 Å wide (Lesburg *et al.*, 1997). As with other terpene synthases the bottom of the active site cavity is predominantly hydrophobic with aromatic residues to shape the cation, with F77 and N219 said to direct the cation into a U shaped conformation (Lesburg *et al.*, 1997). The most closely related terpene synthase to pentalene synthase is epi-isozizianene from *Streptomyces coelicolor,* with 10 α helices forming the α-helical terpene fold (Figure 1-11c) (Aaron *et al.*, 2010). One significant difference is the epi-isozizianene synthase active site is deeper and defined by 5 helices (Aaron *et al.*, 2010). There are a number of structures of epi-isozizanene with a variety of substrate analogues and a number of mutagenesis studies including in the aspartate motifs and aromatic residues in the hydrophobic cavity of the active site (Blank, Barrow and Christianson, 2019). These mutations create changes in the active site contour and the product profile of the enzyme. Hedycaryol synthase from *Kitasatospora setae KM-6054*, is another monomeric sesquiterpene synthase with 11 antiparallel α-helices all connected through short loops and contains all the key class I terpene synthase structural features (Figure 1-11b) (Baer, Rabe, Citron, *et al.*, 2014). Despite producing chemically distinct terpenes and despite having low sequence homology these three single domain terpene synthases align with a score of 91 % (Braberg *et al.*, 2012). Therefore, diversity is based on the specific protein sequence and the active site contour the sequence forms (Figure 1-11a) (Braberg *et al.*, 2012).

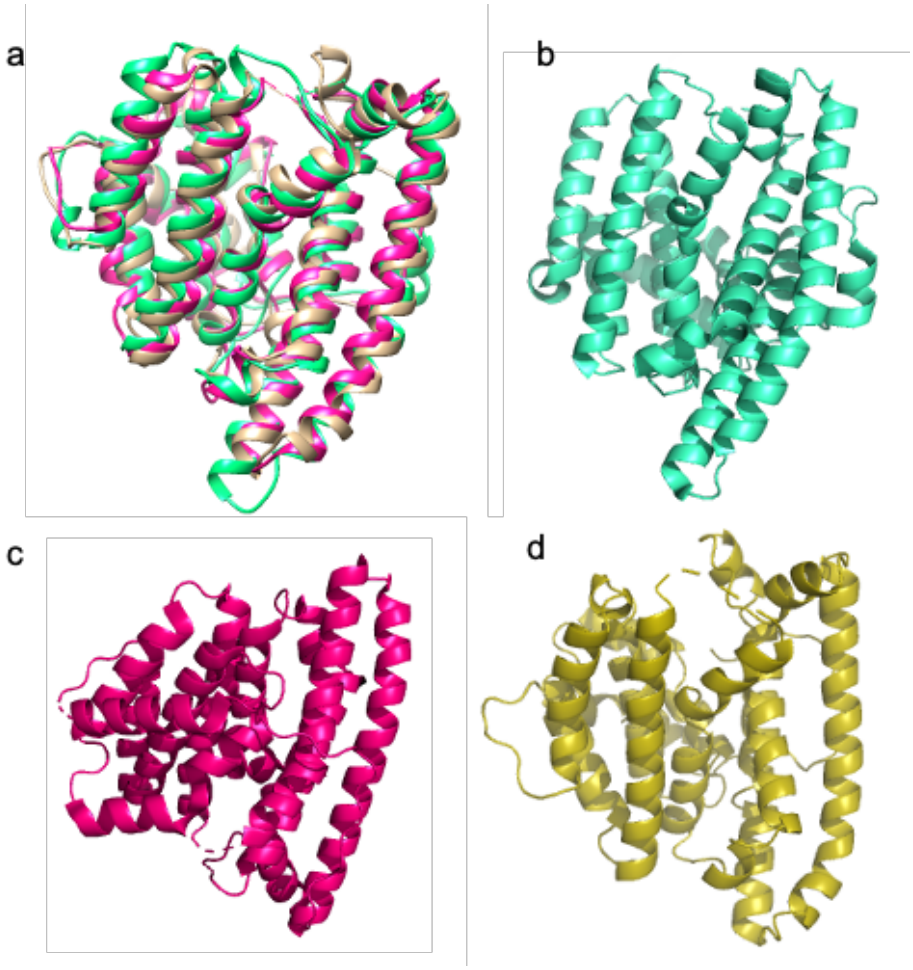


Figure 1-11 Structures and structural alignment of single domain sesquiterpene synthases. *(a) SALIGN alignment of epi-isozizaene synthase from* Streptomyces coelicolor *in green (PBD 4LTV), hedycaryol synthase from* Kitasatospora setae KM-6054 *in pink (PDB 4MC0) and pentalene synthase from* Streptomyces exfolitaus *in gold (PDB 1PS1). Overall, the structures align well with a score of 91.0 % with the main differences seen in the loops and ends of helices with the majority of the helices giving the overall terpene fold. (b) Structure of epi-isozizaene synthase from* Streptomyces coelicolor *in the apo form (PDB 4LTV) showing the classical terpene synthase fold. (c) Structure of hedycaryol synthase from* Kitasatospora setae KM-6054 *in apo from (PDB 4MC0) showing the classical terpene synthase fold. (d) Structure of pentalene synthase from* Streptomyces exfolitaus *in apo from (PDB 1PS1) showing the classical terpene synthase fold. Comparison of the structures shows the key differences are on the left-hand side of the structures.*

Other sesquiterpene synthases are dimeric or heteromeric and have additional domains with unknown functions like limonene synthase. Trichodiene synthase is an antiparallel homodimeric terpene synthase from *Fusarium sporotrichiodes* (Rynkiewicz, Cane and Christianson, 2001). A monomer of trichodiene synthase has 17 α-helices with 6 helices

creating the active site and containing the DDxx(D,E) and NSE motifs for binding 3 magnesium ions. As with other sesquiterpene synthases, trichodiene synthase has multiple structures with various ligands as well as having undergone mutagenesis studies, specifically of the critical aspartate residues. This includes D100E which was crystallised in apo conformation and with pyrophosphate. These structures show the active site is over 10 % larger by attenuating the closure of the active site. As a result, only two magnesium ions bind to the active site, weakening the electrostatic environment around the pyrophosphate moiety. This mutation produces 4 premature products and a 20 fold decrease in k_{cat}/K_m compared to wild type (Rynkiewicz, Cane and Christianson, 2002). Another dimer-forming sesquiterpene synthase is α-eudesmol synthase from *Streptomyces chartreusis* NRRL 3882 with each monomer consisting of 14 α-helices connected by short loops (Kracht *et al.*, 2019). However other sesquiterpene synthases have a 2 domain, $\alpha\beta$ architecture. The α domain is a classic class I terpene synthase whereas the β domain has no known function, but as with limonene synthase forms a component of the active site cap (Blank, Shinsky and Christianson, 2019).

Structural differences can be seen for the same enzyme from different organisms. Both aristolochene synthase from *Penicillium roqueforti* (PR-AS) and *Aspergillus terreus* (AT-AS) were characterised biochemically and structurally with a variety of substrate and intermediate analogues (Figure 1-12) (Caruthers *et al.*, 2000) (Shishova *et al.*, 2008). The two enzymes share 61 % sequence similarity and have major structural differences with minor differences in the product profile as AT-AS is highly specific and PR-AS produces minor amounts of germacrene A and valencene (Caruthers *et al.*, 2000). PR-AS is 38 kDa monomeric consisting of 11 α-helices with 6 helices creating a 15 Å deep and 20 Å wide active site cavity (Figure 1-12b). Whereas AT-AS is a tetramer, a dimer of dimers (Figure 1-12c). Each AT-AS monomer is made up of 13 α-helices but is only 36 kDa with PR-AS 6 α-helices making up the active site. Both active sites are complementary to aristolochene structure and contain all the key class I active site features. The DDXX(D, E) motif is on helix D and the NSE motif is on helix H. Both active sites display induced fit closure with helices G1, H and the F-G1 loop showing shift between the apo and bound structures forming the active site cap (Shishova *et al.*, 2007) (Shishova *et al.*, 2008) (Yu, Miller and Allemann, 2007). Both active sites have three key aromatic residues (two phenylalanine's and one tryptophan) that stabilise the eudesmane cation through aromatic π interactions. In PR-AS these residues are F112, F178 and W334 whereas in AT-AS they are F87, F153 and W308 (Caruthers *et al.*, 2000) (Shishova *et al.*, 2007).

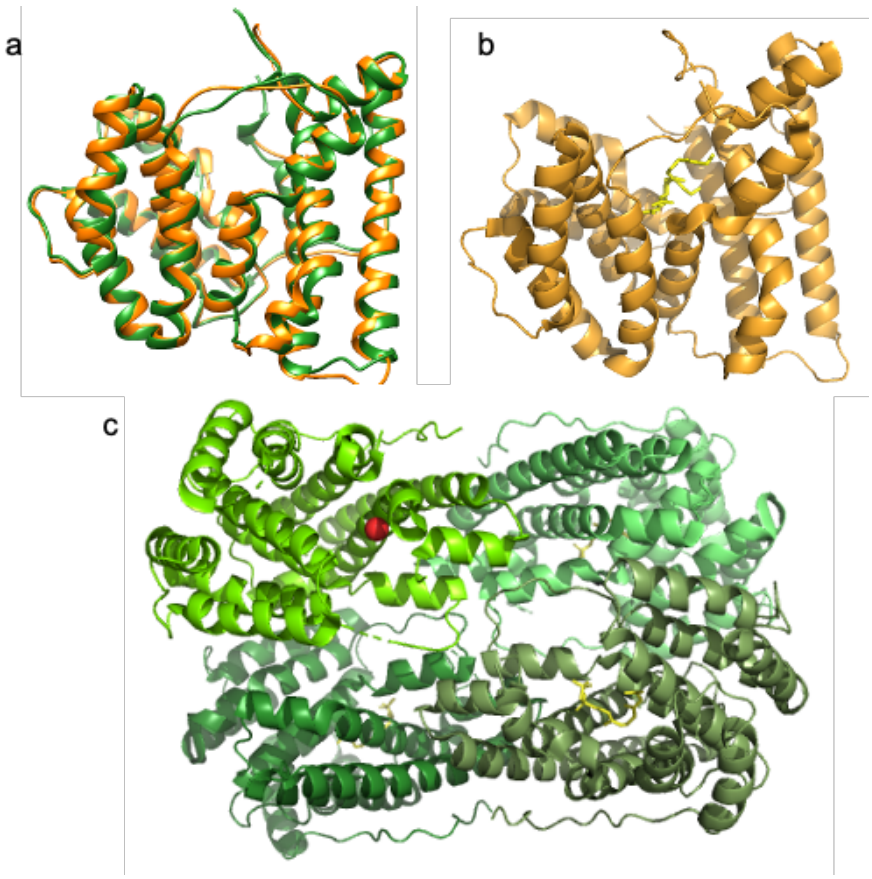


Figure 1-12 Structures of aristolochene synthase*. (a) SALIGN of* Penicillium roqueforti *aristolochene synthase in orange (PDB 1DGP) and chain A of* Aspergillus terreus *aristolochene synthase in green (PDB 3BNX). Alignment displays only subtle differences away from the active site with an alignment score of 97.9%. (b) Structure of monomeric* Penicillium roqueforti *aristolochene synthase (PDB 1DGP) with farnesol bound in yellow. (c) Structure of the tetramer of* Aspergillus terreus *aristolochene synthase (PDB 3BNX) with each chain coloured a different shade of green. Chains A, B and C have farnesyl pyrophosphate coloured yellow bound and chain D has magnesium ions bound, coloured red.*

Diterpene synthases can also be classified as class I terpene synthases, for example cyclooctat-9-en-7-ol synthase or CotB2 which produces cyclooctat-9-en-7-ol which is further transformed to terpenoid cyclooctatin (Driller *et al.*, 2018). CotB2 has the classical α-helical fold with both the DDxx(D, E) and NSE motifs with the active site. When comparing apo and substrate analogue-bound crystal structures, an induced fit closing mechanism can be postulated as with other class I terpene synthase.

The identification of a large terpene synthase ($C_{25}/C_{30}/C_{35}$) from *Bacillus subtilis*, β-curcumene synthase (BsuTS) led to the identification of an alternative subclass of terpene

synthase, class IB (Fujihashi *et al.*, 2018). One homologous β-(C25/C30/C35)-prene synthase class IB terpene synthase from *Bacillus alcalophilus*, BalTS, was crystallised showing an α-terpene synthase fold. BalTS forms a dimer in solution where cavities in each subunit fuse to form a large active site capable of holding long chain prenyl pyrophosphates (Fujihashi *et al.*, 2018). As with class I terpene synthases three magnesium ions are bound in the active site however the two aspartate motifs differ in sequence; DYLDNLxD and DY(F,L,W)IDxxED (Fujihashi *et al.*, 2018). BsuTS forms products with a single ring while BalTS produces linear products, however the active site residues are conserved between the two enzymes. This implies that the control over product determination is in the secondary level (Fujihashi *et al.*, 2018).

1.5.2 Class II terpene synthase structure

Class II terpene synthases have a characteristic 2 domain architecture (β and γ) with the interface creating the active site (McAndrew *et al.*, 2011) (Tholl, 2015). The structures consist of α-helices and one aspartate motif DXDD. The interface interaction contains repetitive QW sequence motifs on the outer side of helices (Wendt and Schulz, 1998). Class II terpene synthases ionise the substrate by protonation of the epoxide ring at a double bond (Tholl, 2015) (Gao, Honzatko and Peters, 2012). Along with bifunctional enzymes, some class II enzymes have an additional 250 amino terminal insert for correct folding (Tholl, 2006). Class II enzymes include some diterpene synthases and triterpene synthases (Christianson, 2017).

One example of a classical class II terpene synthase structure is *ent*-copalyl pyrophosphate synthase from *Streptomyces platensis* CB00739 with its $\beta\gamma$ domain architecture (Figure 1-13c) (Rudolf *et al.*, 2016). *ent*-Copalyl pyrophosphate synthase produces platensimycin and platencin (Rudolf *et al.*, 2016). As with aristolochene synthase another *ent*-copalyl pyrophosphate synthase from a different species (*Arabidopsis thaliana*) has different structural features with an $\alpha\beta\gamma$ domain architecture (Figure 1-13b) (Potter *et al.*, 2014). The α-domain in *Arabidopsis thaliana ent*-copalyl pyrophosphate synthase has no known function but structurally resembles a class I active site without the characteristic aspartate rich motifs (Christianson, 2017). Despite the structural differences between the two *ent*-copalyl pyrophosphate synthases the proposed mechanism is the same with corresponding key residues in the same position (Figure 1-13) (Christianson, 2017).

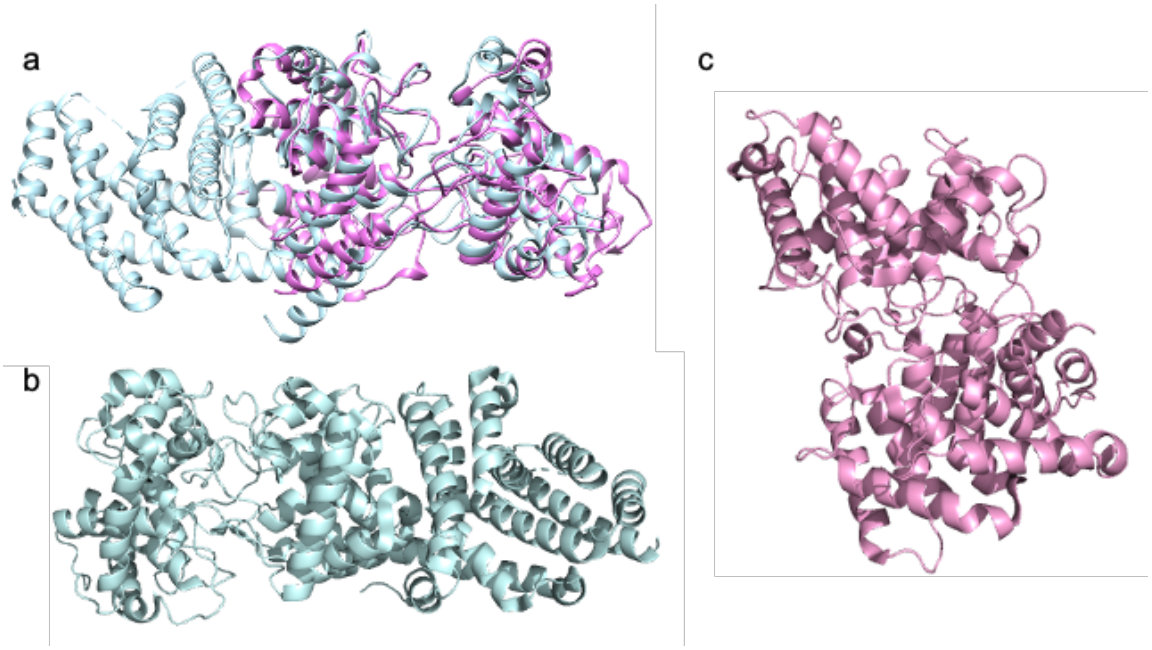


Figure 1-13 Structures of ent-copalyl pyrophosphate synthases*. (a) SALIGN diagram of ent-copalyl pyrophosphate synthases from* Streptomyces platensis CB00739 *in pink (PDB 5BP8) and* Arabidopsis thaliana *in grey (PDB 4LIX). The alignment shows a well aligned $\beta\gamma$ domain architecture and Arabidopsis thaliana ent-copalyl pyrophosphate synthase distinct separate α domain. (b) Structure of* Arabidopsis thaliana *ent-copalyl pyrophosphate synthase (PDB 4LIX) with its three domain $\alpha\beta\gamma$ structure. (c) Structure of* Streptomyces platensis CB00739 *ent-copalyl pyrophosphate synthase with its two-domain class II terpene domain structure, $\beta\gamma$.*

As previously mentioned, squalene-hopene cyclase from *Abicyclobacillus acidocaldarius* is a well-studied triterpene synthase with a $\beta\gamma$ domain architecture. Squalene-hopene cyclase is a membrane bound enzyme with the interface active site accessible through a membrane passageway (Christianson, 2017). This structure is similar to membrane bound oxidosqualene cyclase also with a $\beta\gamma$ domain architecture and a membrane passage to the interface active site (Christianson, 2017). Both of these structures, along with other experiments have been used to postulate the mechanism of action with the traditional protonation of terminal π bond or epoxide moiety initiation.

1.5.3 Bifunctional terpene synthase structures

Bifunctional terpene synthase catalyses the tandem or coupled cyclisation reactions in two distinct active sites. Bifunctional terpene synthase active sites have the key terpene synthase structural features and can be either class I, class II or a combination. In bifunctional enzymes both active sites must be functional; however a channel directing

products between the two sites is not always necessary (Christianson, 2017). The most studied example of a bifunctional enzyme is geosmin synthase which has an $\alpha\alpha$ domain structure with two class I active sites (Harris *et al.*, 2015). There is no direct channel between the two active sites that cyclise farnesyl pyrophosphate to germacradienol before fragmentation to C_{12} geosmin (Harris *et al.*, 2015). Both α domains have a characteristic class I terpene synthase fold binding three magnesium ions (Christianson, 2017).

However other bifunctional enzymes have both class I and class II active sites together. One example is (E)-biformene synthase a labene-related diterpene. (E)-biformene synthase catalyse two consecutive reactions; cyclisation of GGPP by protonation in a class II active site followed by cyclisation of a bicyclic to pentacyclic product triggered by metal ion dependent dephosphorylation, in a class I active site (Centeno-Leija *et al.*, 2019). Both active sites have the key structural features of class I and class II active sites with the α-helical fold, aspartate motifs and magnesium ions (Centeno-Leija *et al.*, 2019).

1.5.4 Alternative terpene synthase structures

Some terpene synthases do not contain the key identified features of terpene synthases. One well characterised example is the sesquiterpene synthase, δ-cadinene synthase which does not contain the NSE/DTE motif but instead contains a DVAE sequence similar to farnesyl pyrophosphate synthase aspartate motifs (Gennadios *et al.*, 2009) (Gao, Honzatko and Peters, 2012). In addition, δ-cadinene synthase has a $\beta\alpha$ didomain structure with a flexible region of the *N*-terminal β domain being involved in the capping of the α domain active site. This was shown by progressive truncation of the N-terminal leading to increased production of sesquiterpene alcohol due to inefficient sealing of the active site and solvent exposure (Gonzalez *et al.*, 2016).

Two other class I terpene synthases with unusual features are α-bisabolene synthase from *Abies grandis* and taxadiene synthase from *Taxus brevifolia* that both have a $\alpha\beta\gamma$ domain architecture (Figure 1-14) (McAndrew *et al.*, 2011) (Köksal *et al.*, 2011). Both enzymes have a functional α-domain class I active site at the *C*-terminal of the protein. The *N*-terminal of each protein is a class II active site fold missing with mutated aspartate rich motifs and like δ-cadinene synthase, forms part of the active site cap. The fact these additional domains can form part of the active site cap differs from other terpene synthase that have non-functional domains.

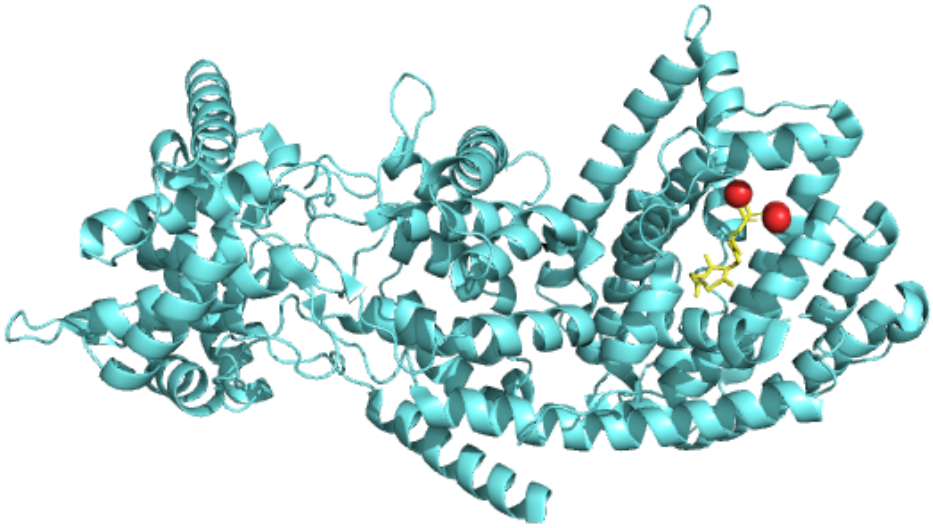


Figure 1-14 Structure of taxadiene synthase from* Taxus brevifolia *(PDB 3P5R). *Taxadiene synthase has a three-domain architecture with a $\alpha\beta\gamma$ structure. The α domain is a functional class I active site shown in the structure with magnesium ions in red and 2-fluorogeranylgeranyl pyrophosphate in yellow bound in the active site. The $\beta\gamma$ domain is a non-functional class II active site that forms part of the cap of the class I active site.*

1.5.5 Prenylelongase structures

Prenylelongases catalyse the formation of terpene precursors by catalysing the sequential condensation of DMAPP and IPP and are related structurally to class I terpene synthases. Both enzymes hace α-helical terpene folds with an active site containing aspartate rich motifs, however the sequence of these motifs differs between prenylelongases and class I terpene synthases (Wendt and Schulz, 1998). This is also similar to the glutamate residues that coordinate divalent metal ions that can stabilise the carbocation in IDI (Zheng *et al.*, 2007). The depth of the active site determines the length of the chain product as the active site acts as a well-defined template. This has been confirmed by mutagenesis increasing the depth of active sites which results in longer polyisoprenoid products (Christianson, 2008). Trans-isoprenyl pyrophosphate synthase generates a trans C-C double bond in geranyl, farnesyl or geranylgeranyl pyrophosphates produced by a corresponding pyrophosphate synthase (Gao, Honzatko and Peters, 2012).

The most studied prenylelongase is the homodimer avian farnesyl pyrophosphate synthase consisting of 10 α-helices forming a large central cavity active site (Tarshis *et al.*, 1994). The active site cavity consists of hydrophobic interactions with two conserved aspartate rich motifs FARM and SARM coordinating the magnesium ions (Tarshis *et al.*, 1994). The helices at the lower end of the active site have short loops connecting helices whereas other helices

have large connecting loops of up 60 residues (Tarshis *et al.*, 1994). The dimer interface is formed of a large number of hydrophobic interactions (Tarshis *et al.*, 1994).

The structural homology and similarities between successive enzymes suggest enzymes that catalyse successive steps in biosynthetic pathways may evolve with similar structure features through divergence regardless of sequence. This suggests a sequence and function evolves more rapidly than tertiary structures (Caruthers *et al.*, 2000). The terpene synthase α fold has evolved to accommodate a variety of different reactions with subtle changes to the active site contour (Rynkiewicz, Cane and Christianson, 2001). Other suggestions of this structural homology come from terpene synthases that contain $\alpha\beta\gamma$ domain architecture of both class I and class II active sites including; α-bisabolene synthase and taxadiene synthase. This suggests these enzymes could have evolved from a common ancestor. This is further shown by the bifunctional terpene synthase with the fusions of two different active sites. Overall the low sequence homology but high tertiary homology suggest enzymes could have evolved from each other time maintaining the classic fold but having large differences in sequence (Wendt and Schulz, 1998).

1.6 Reaction mechanisms of terpene synthases

Terpene diversity arises from the enzyme activity and subtle differences in the electrophilic mechanism within the active site. Reactions are either initiated by magnesium dependent removal of the pyrophosphate moiety or the protonation of a π or epoxide moiety, both creating a cation (Gao, Honzatko and Peters, 2012). Reactions then proceed through a multistep reaction cascade with cationic intermediates through; protonation, deprotonation, hydride shifts, intramolecular proton transfers and electrophilic attack (Dickschat, 2017) (Tantillo, 2017) (Hong and Tantillo, 2015). These reactions combine several bond forming and rearranging steps and are classified as cationic S_N1-like alkylations (Salmon *et al.*, 2015) (Wendt and Schulz, 1998). Terpene synthase reactions are chemically important as despite being promiscuous or high fidelity the reactions are chemo selective, regioselective, diastereoselective and enantioselective (Harms, Kirschning and Dickschat, 2020). As a result the hydrophobic active site acts as a template, controlling the reaction cascade by aligning the cation into productive spatial orientations and stabilising transition states (Miller and Allemann, 2012). It has also been postulated that the pyrophosphate anion contributes to the electrostatic effect created by aromatic and hydrophobic residues or could act as an acid/base within the reaction cascade, potentially re-joining the cation (Zhou and Peters, 2011). Terpene synthases control the reaction cascade by raising the energy of initial intermediates to bypass the formation of unwanted side products (Major, 2017). The

reactions proceed with a delicate balance of kinetic, thermodynamic and dynamic control on the cationic intermediates (Dixit *et al.*, 2017). A variety of chemical and computational methods have been used to study these chemically important reactions.

1.6.1 Substrate analogues

Substrate or intermediate analogues can be used in combination with x-ray crystallography and product determination methods. Usually substrates or intermediates are modified with fluorine, methyl groups, nitrogen, hydroxyl groups, deuterium and ^{13}C labelling (Figure 1-15) (Prisic *et al.*, 2007) (Cascon *et al.*, 2012). Due to the volatile nature of terpenes, products can be detected with GC/MS and further characterised by NMR (Forcat and Allemann, 2006). Isotopic labelling or isotopic sensitive branching of substrate analogues in combination with *in vitro* incubation of purified terpene synthase produces a list of intermediates determining the precise reaction pathway.

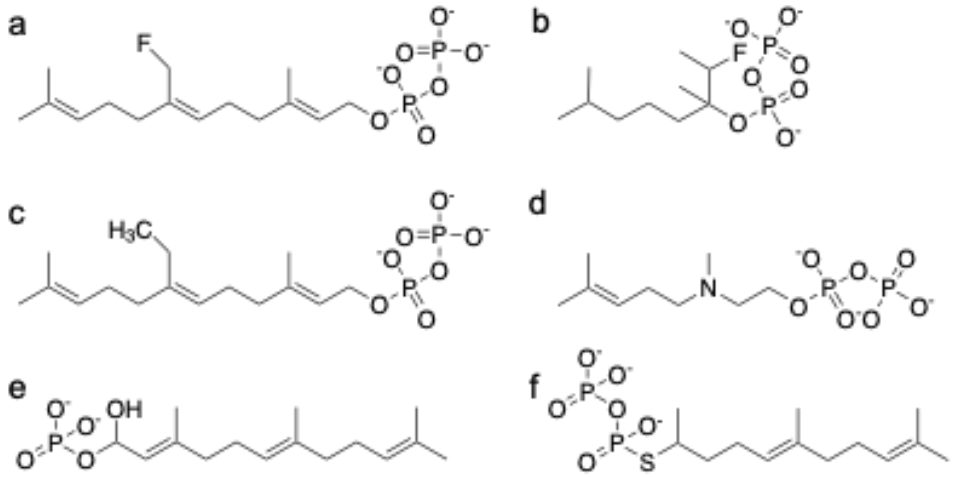


Figure 1-15 Structures of substrate analogues used in determining terpene synthase reaction mechanisms. *(a) Structure of 6-fluorofarnesyl pyrophosphate. (b) Structure of 2-flurorlinalyl pyrophosphate. (c) Structure of 6-methylfarnesyl pyrophosphate. (d) Structure of 3-aza-2,3-dihydrogeranyl pyrophosphate. (e) Structure of farnesyl hydroxy pyrophosphate. (f) Structure of farnesyl thiopyrophosphate.*

Isotopic sensitive branching involves using a substrate analogue with a single atom swapped for an isotope that is fed to the enzyme. The ratio of labelled products is compared to an unlabelled substrate (Tantillo, 2017). One key atom used in substrate analogues is fluorine, specifically fluoro-farnesyl pyrophosphate isomers, that have been used with a variety of sesquiterpenes synthases. Fluorine is a popular choice due to its small size not affecting binding but its electronic properties have strong inhibitory effects (Harms, Kirschning and Dickschat, 2020). Fluorine can be set in specific positions for targeted inhibitions at key sites

(Harms, Kirschning and Dickschat, 2020). PR-AS was tested with variety of fluoro-farnesyl pyrophosphate isomers and the results suggested a 1,10 cyclisation and germacryl cation which it can be blocked by the destabilisation of the product cation by the fluorine (Yu, Miller and Allemann, 2007) (Miller *et al.*, 2009). Studies into (E)-β-farnesene and isoprene synthase used the electron withdrawing effect of fluorine to find competitive inhibitors blocking the formation of trans-farnesyl cation and the proceeding reaction steps (Faraldos *et al.*, 2012). The inhibitors reflect cationic intermediates with fluorine and help determine the reaction pathway. Both reactions go through regiospecific 1,4 conjugate elimination of either farnesyl pyrophosphate or dimethylallyl pyrophosphate. The strongest inhibitor for (E)-β-farnesene are inhibitors with fluorine at position 15 including; 15-flurorfarnesyl pyrophosphate and 15-triflurorofarnesyl pyrophosphate (Faraldos *et al.*, 2012).

In addition to fluorine, sulfur has been used make farnesyl-S-thiolopyrophosphate which has be used in crystallisation to mimic farnesyl pyrophosphate binding conformations to induce the reaction pathway. This has also been used on AT-AS with ammonium and iminium analogues of carbocation proposed intermediates for reaction pathway (Chen *et al.*, 2013). Binding of these analogues in crystal structures leads to a proposed mechanism of action including identifying key residues involved in cation stability specifically the aromatic triad of Y61 F81 and F147 stabilising cation π interactions (Chen *et al.*, 2013). Some cation intermediates are stabilised or directed by interactions with the pyrophosphate anion as suggested by the aza analogue's interactions with the pyrophosphate anion (Chen *et al.*, 2013). This is consistent with the orientation of the final carbocation being proximal to the pyrophosphate anion and its potential key role in quenching the final carbocation intermediate (Chen *et al.*, 2013). Analogues closer in structure to the final product are more thermodynamically favourable than less product-like analogues, however it is possible a high-fidelity terpene synthase can bind analogues with "incorrect" structures or stereochemistry (Chen *et al.*, 2013). As a result, these studies should be used in combination with additional methods to determine the reaction pathway.

Deuterium site-specific labelling, also known as following the proton, can test the feasibilityof potential mechanisms (Tantillo, 2017). One example was to gain insight into the cyclisation mechanism of 1,8 cineol synthase from *Streptomyces clavuligerus* by specific labelling of the protons at the C1 of geranyl pyrophosphate (Dickschat, 2017). Labelled substrates were incubated with purified 1,8 cineol synthase and products are analysed by heteronuclear single quantum coherence (HSQC) spectroscopy to propose a mechanism. In addition, amorpha-4,11-diene synthase from *Artemisia annua* was studied with *R*- and *S*-deuterated

farnesyl pyrophosphate to determine cation intermediates to determine a mechanism of action (Miller and Allemann, 2012).

Site-specific labelling can also use ^{13}C-labelled substrate analogues to determine reaction pathways and identify potential intermediates. One example is the study of three sesquiterpene synthases from *Streptomyces* species using a variety of ^{13}C farnesyl pyrophosphate isomers (Rabe, Rinkel, Klapschinski, *et al.*, 2016). This was a fast-easy method to determine products and the reactions' stereochemical courses. These incubations identified the products of the three enzymes; (-)-neomeranol B, isodauc-8-en-11-ol and intermedeol (Rabe, Rinkel, Klapschinski, *et al.*, 2016). The stereochemical course was determined using HSQC spectroscopy and nuclear overhauser effect spectroscopy (NOESY) of the unlabelled and labelled products along with stereochemical farnesyl pyrophosphate isomers (Rabe, Rinkel, Klapschinski, *et al.*, 2016). An additional example is the study of epi-isozizaene synthase using ^{13}C farnesyl pyrophosphate isomers with EI-MS fragmentation and position-specific mass shift analysis (Rabe, Klapschinski and Dickschat, 2016). These can be used in combination with other mass spectroscopy techniques including; high resolution GC/MS quadrupole time-of-flight and MS-MS for position specific mass shift analysis. Using this labelling every carbon atom within a fragment can be identified through the corresponding ^{13}C farnesyl pyrophosphate mass shift and from this a mechanism can be postulated (Rabe, Klapschinski and Dickschat, 2016).

1.6.2 Site directed mutagenesis

Structural analysis of terpene synthase bound with substrate analogues has identified key residues involved in directing the product formation and stabilising the cation intermediates. These residues can then be studied by site directed mutagenesis to determine the effect on the product profile and the residues specific role. One of the most studied active sites is aristolochene synthase from both PR-AS and AT-AS. The main target of mutagenesis for PR-AS was the aromatic residues comparable to the aromatic triad identified in AT-AS targeting Y92, F112 F178 and W334 (Figure 1-16b) (Miller and Allemann, 2012) (Forcat and Allemann, 2006). Mutagenesis of W334 to a variety of hydrophobic residues, specifically valine, led to the production of intermediate germacrene A showing its role in stabilising the eudesmane cation which is formed following the protonation of germacrene A (Miller and Allemann, 2012). Mutagenesis of Y92 displays a similar eudesmane cation stabilisation role as Y92V gives a similar product profile to W334V. In addition the size of the residue at position 92 has an effect on the folding of the initial cation, with increasing residue size the production of aristolochene increased and as the volume decreases amounts of linear products increase (Miller and Allemann, 2012). F178 was mutated to cysteine, isoleucine,

tyrosine and tryptophan to investigate the effect of size and aromaticity at position 178 (Forcat and Allemann, 2006). When mutated to other aromatic residues the product profile is similar to wild-type PR-AS with aristolochene as the main product with minor amounts of germacrene A and valencene (Forcat and Allemann, 2006). Mutations removing the aromatic residues, like cysteine and valine, reduced the amount of aristolochene and increased the amount of germacrene A along with additional by-products (Forcat and Allemann, 2006). The results suggest F178 has a critical role in the 1,2 hydride shift from C2 to C3, that initiates the conversion of eudesmane to aristolochene and a bulky side chain is key for this role (Forcat and Allemann, 2006). F112 is critically placed between W334 and F178 so to define its role it was mutated to alanine, producing only a small amount of germacrene A and some linear farnesene products (Figure 1-16c). As a result it was determined F112 is critical for the roles of W334 and F178 to allow the conversion of eudesmane to aristolochene (Forcat and Allemann, 2006). K206 is predicted to be the acid/base in the reaction mechanism due its proximity to Y92. When K206 was mutated to glutamine, production was reduced by 3 orders of magnitude and 4 fold when mutated to arginine (Miller and Allemann, 2012). Site directed mutagenesis was used in combination with labelled substrates (see section 1.6.1) to postulate a mechanism of action with specific stereochemistry (Miller *et al.*, 2009).

AT-AS mutagenesis focused more on the active site cleft compared to the aromatic triad on PR-AS. One focus was probing the trapped water molecule in the active site as a component of the active site contour but not actively involved in the cation quenching (Chen *et al.*, 2016). Site directed mutagenesis looked into the relationship between the polar amino acid side chains and the trapped water molecule (Chen *et al.*, 2016). S303 and N299 were targeted and mutated to non-polar alanine and the water molecules remained (Figure 1-16a). However, the structure of the S303A/N299A double mutant has no solvent molecule resolved showing the loss of both interactions results in the loss of the bound solvent molecules (Chen *et al.*, 2016). The structures of N299A and S303A/N299A show a rotation of Y61 hindering the re-protonation of germacrene A to form the eudesmane cation (Chen *et al.*, 2016). Other mutagenesis efforts targeted residues within the active site contour including V88 and T89. When mutated to the bulkier aromatic phenylalanine at both positions, germacrene A production was increased (Faraldos, Gonzalez, *et al.*, 2011).

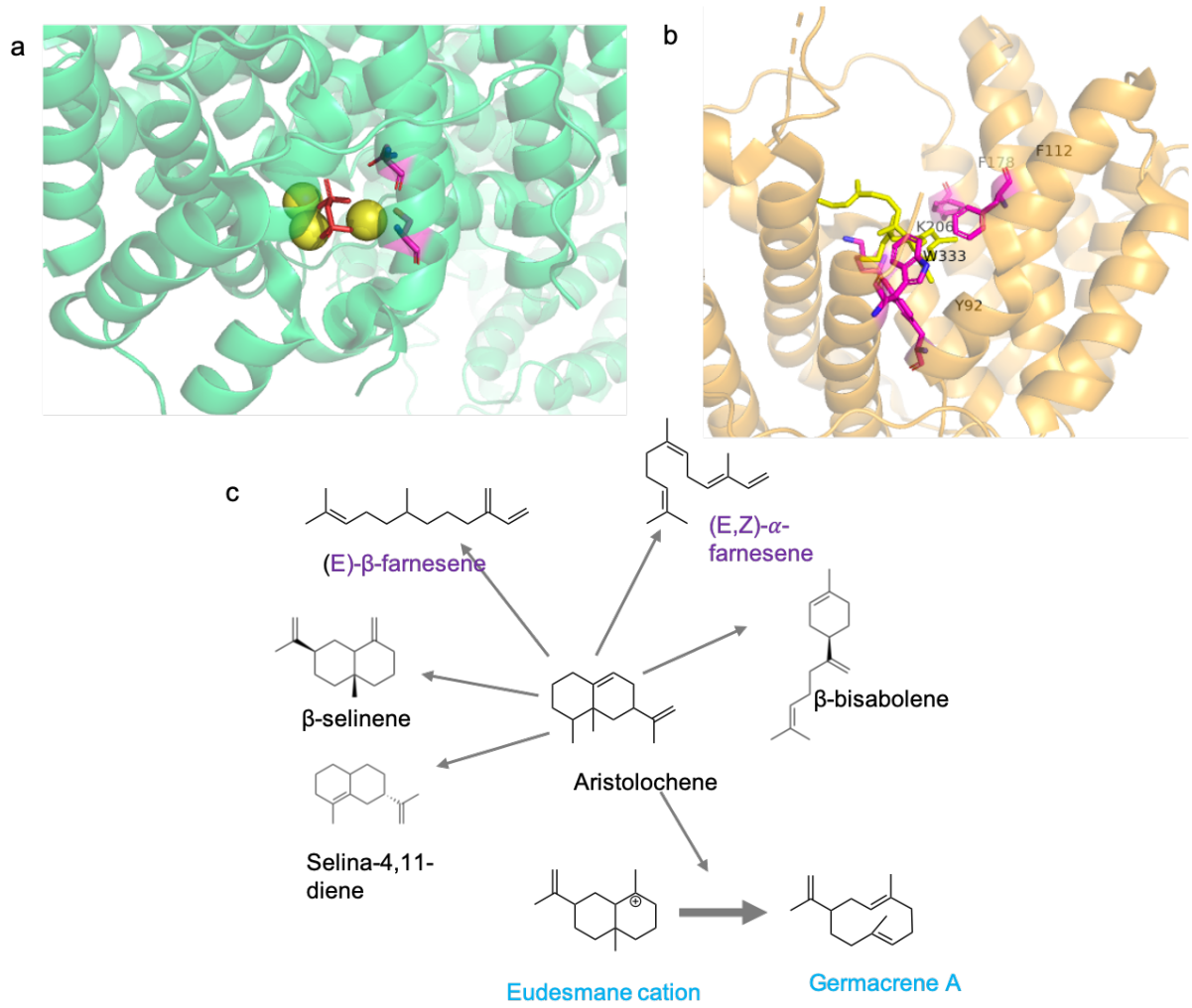


Figure 1-16 Active site images of aristolochene synthase positions targeted and a schematic of some of the products of mutagenesis. *(A) Active site structural image of* Aspergillus terreus *aristolochene synthase (PDB 3BNX) bound with pyrophosphate in pink and yellow magnesium ion. Residues N299 and S303 that were targeted for mutagenesis are highlighted in pink. (B) Active site structural image of* Penicillium roqueforti *aristolochene synthase (PDB 1DGP) with farnesol bound and displayed in yellow. Residues Y92, F112, F178, K206 and W334 were the targets of mutagenesis and surround the bottom of the active site are shown in pink. (C) Structures of some of the alternative products produced by variant enzymes. The products labelled in purple are produced from mutagenesis at position Y92. The products labelled in blue are key products from the mutagenesis of W334 and specifically mutant W334V.*

In pentalene synthase, aromatic residues were also selected as a target for site directed mutagenesis. F76A and F77A showed significant reduction in activity along with derailment of cyclisation leading to the production of germacrene A (Miller and Allemann, 2012). In addition H309 was proposed to be the active site base in the mechanism of pentalene

synthase (Seemann *et al.*, 1999). However unexpectedly all H309 mutants retained some activity with no significant differences in the product profile. Therefore H309 has no base role in catalysis and has only a minimal effect on the active site contour from the minor product profile changes (Seemann *et al.*, 1999).

As with all techniques for determining reaction mechanisms site directed mutagenesis can be combined with other techniques to give further mechanistic insight. One example is germacardien-4-ol synthase, a sesquiterpene synthase with a bound water molecule, however unlike AT-AS the active site does not contain any polar residues (Grundy *et al.*, 2016). As with other studies, mutations of D80, D81 and D84 (key aspartate residues) to glutamate showed significantly less activity and with minor functional changes producing germacrene A and D (Grundy *et al.*, 2016). Whereas when mutated to asparagine the enzymes are inactive showing the critical role in coordinating magnesium ions (Grundy *et al.*, 2016). Other key residues were targeted including N218 within the NSE motif. When N218 was mutated to glutamine the catalytic efficiency was reduced and the product profile was altered, similar to mutagenesis with aspartate residues, by increasing production of germacrene A and germacrene D. This shows that the longer side chain is perturbing the magnesium ion coordination (Grundy *et al.*, 2016). Due to the lack of obvious residues, the enzyme was incubated with $H_2{}^{18}O$ to confirm the hydroxyl group came from the active site trapped water, however it is unclear how the water is captured. In addition, germacardien-4-ol was incubated with a variety of substrate and intermediate analogues using the activity and inhibition to determine a mechanism. The postulated mechanism involves the farnesyl cation undergoing a 1,10 cyclisation to the germacrenyl cation before a rapid 1,3 hydride shift where the cation is quenched by the water molecule (Figure 1-17) (Grundy *et al.*, 2016).

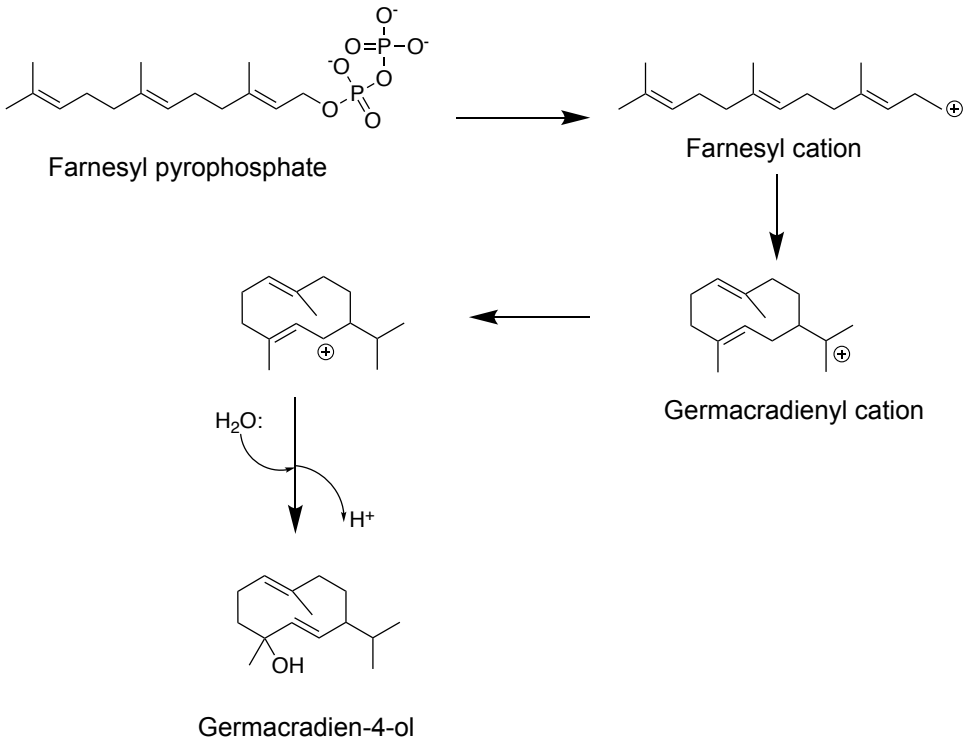


Figure 1-17 Schematic of the postulated mechanism of germacradien-4-ol. *As with all class I terpene synthase the magnesium dependent pyrophosphate removal produces the farnesyl cation. The farnesyl cation was cyclised through a 1,10-cyclisation forming a germacradienyl cation which was consequently attacked by a hydroxyl group of a water molecule forming germacradien-4-ol.*

A way to expand the current repertoire of amino acid is to use non-canonical amino acids. This was tested with PR-AS to further investigate W334 and its effect on eudesmane cation. The aim was to prove if residues with lower π electron densities than tryptophan would decrease the rate of reaction and a change the product profile. W334Y had a 3-fold decrease in efficiency with an increase in germacrene A as the phenolic hydroxyl prevents the optimal orientation for cation stabilisation (Faraldos, Antonczak, *et al.*, 2011). W334H had a 70 fold decrease in efficiency with aristolochene activity significantly reduced but there was an 11 fold increase in germacrene A whereas W334L showed no activity (Faraldos, Antonczak, *et al.*, 2011). Building on these results, non-canonical aromatic amino acids with a diverse range of properties were substituted into the enzyme using the amber suppression method. The amber suppression method utilises the amber stop codon, the least used stop codon in *E. coli,* and a specifically evolved amino acyl t-RNA synthetase to insert the desired non-canonical amino acid (Goodman *et al.*, 1968). If germacrene A was a consequence of the reduced steric size in the active site, then any bulky aromatic residue at 334 would re-establish aristolochene production due to the π stabilisation of eudesmane cation. The

results show decreasing π densities increase the production of germacrene A (Faraldos, Antonczak, *et al.*, 2011).

1.6.3 Computer modelling and simulations

Increasing technological advances and availability of computer modelling tools have produced the most recent advances in mechanistic understanding of terpene synthases. There are a variety of methods that can be used to study a variety of parameters relating to the reaction mechanism. One of the most common modelling methods is QM/MM, a hybrid of quantum mechanics and molecular mechanics combining accuracy and speed to model environments (Warshel and Levitt, 1976). Molecular mechanics uses a force field to calculate the forces acting on each particle. This involves finding a molecular solution or a structure with the minimum energy (Warshel and Levitt, 1976). One common use in terpene synthase is to predict intramolecular proton transfers that can be affected by various barriers including; ring strain, the degree of substitution, the level of delocalization and whether the environment is endocyclic or exocyclic (Hong and Tantillo, 2015).

One example involves using modelling techniques MD, QM/MM and QM/(DFT)/MM on AT-AS, *Nictotina tabacum* 5-epi-aristolochene synthase and farnesyl pyrophosphate synthase (Zhang *et al.*, 2016). Initially the analysis focused on the pyrophosphate anion cleavage which is universal in all three reactions and is usually the rate limiting step in the reaction pathway. From the calculations the pyrophosphate anion cleavage and the C-C bond formation is a nucleophilic substitution reaction where the pyrophosphate acts as the leaving group and the C10-C11 double bond acts as the nucleophile (Zhang *et al.*, 2016). Throughout the remaining reaction pathway, the pyrophosphate can adopt a variety of protonation states to assist the final product formation. The protonation state of the pyrophosphate can depend on the cleavage mechanism, the native active site environment, the magnesium ion coordination and the enzyme class (Zhang *et al.*, 2016). These findings can be applied to other terpene synthases to determine the effect of the pyrophosphate on the remaining reaction pathway. Further, these results can be combined with additional QM/MM and substrate docking to identify the single possible orientation of the farnesyl cation in 5-epi-aristolochene synthase for productive reaction (O'Brien *et al.*, 2016).

Another example is the use of QM/MM and multidimensional free energy simulations on bornyl pyrophosphate synthase (Major and Weitman, 2012). This analysis showed that the pyrophosphate moiety has a level of control directing the reaction towards the product (Major and Weitman, 2012). It identifies that the bornyl cation is the branching point of the reaction

pathway where the enzyme pushes the reaction towards the formation of the terpinyl cation (Major and Weitman, 2012). The analysis proposed a multilevel catalytic pathway where transition states are separated by low free-energy barriers due to the thermodynamic properties of the enzyme (Figure 1-18). This is supplemented by the initial product like binding configuration of geranyl pyrophosphate (Major and Weitman, 2012).

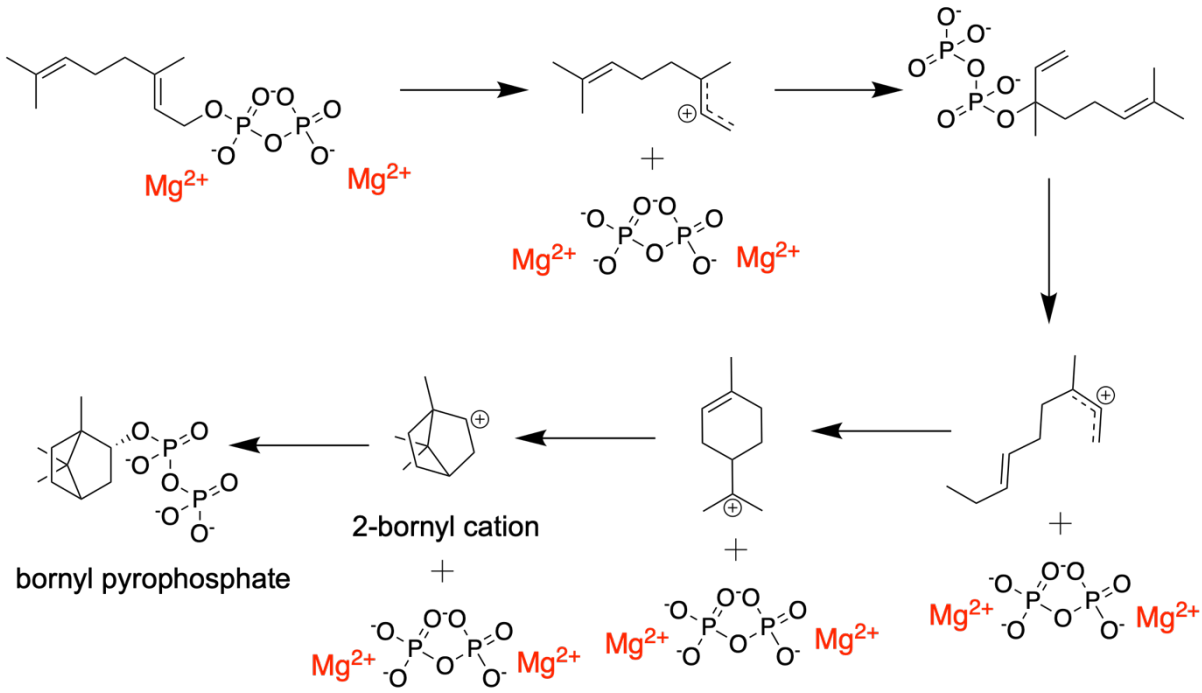


Figure 1-18 Schematic of the proposed reaction mechanism of bornyl pyrophosphate synthase. *The proposed mechanism is initiated by the magnesium ion-dependent pyrophosphate removal to generate the geranyl cation. The pyrophosphate moiety re-joins the farnesyl cation at a different position and is subsequent removed forming a different cation that can be cyclised to the 2 bornyl cation. The pyrophosphate moiety cation quenches the cation forming bornyl pyrophosphate.*

Bornyl pyrophosphate synthase was used to test a computer modelling software (TerDockin) specifically designed to predict orientation of carbon skeletons of substrates and intermediate carbocations with a pyrophosphate moiety in terpene synthase active sites (O'Brien *et al.*, 2018). As with QM/MM software, the binding orientation appears to be relevant to the reaction and shows the stereochemical preference for the enzyme. The reaction pathway predicted binding modes from TerDockin are consistent with previous experimental observations and there are no significant movements between each reaction step (O'Brien *et al.*, 2018).

A popular technique is *in silico* modelling which combines both *in vivo* and *in vitro* experiments and can control an unlimited number of parameters (Colquitt, Colquhoun and

Thiele, 2011). This technique was used on taxadiene synthase using the open and inactive crystal structure with substrate analogues like 2-fluorogernaylgernayl pyrophosphate. Modelling created a closed active site conformation enabling insight into the structure function relationship and productive geranylgeranyl pyrophosphate folding patterns. Consequently, this leads to the localisation of the intermediate carbocations and the key residue interactions. These results enable *in silico* protein engineering of the identified residues to alter the predicted cascade.

Overall the best way to establish the complex reaction pathways of terpene synthases is to use a sequential combination of techniques. A useful protocol is to use substrate analogues to create crystal structures and find potential binding orientations. These structures can then be used for modelling and simulation studies. From the modelling studies, residues can be identified for site directed mutagenesis and potential future engineering. The observations from these mechanistic studies can develop chemical methods to mimic terpene synthase. The majority of synthetic terpene synthases are based on class II reactions while class I reactions are rare. One example is the biomimetic penta-cyclisation of polyenes which are derivative oleananes. This can be achieved under acidic conditions without enzymes (Fish and Johnson, 1994). The first enantioselective cyclisation of a polyprenoid used a Lewis acid-assisted chiral Bronsted acid, however there was a mix diastereoisomers (Harms, Kirschning and Dickschat, 2020).

1.7 Engineering terpene synthase

Terpene synthases chemically interesting due to their chemical variation from limited substrates while being highly specific and enantioselective. As a result, terpene synthases are a great target for engineering and potential novel products production. There are two main approaches for engineering enzymes; rational engineering and non-rational engineering or directed evolution. Rational engineering uses structural or modelling information to propose mutations to be created using site-directed mutagenesis. There is an increased probability of mutations being beneficial or productive reducing the library size especially if there is no high-throughput screen available (Acebes *et al.*, 2016). On the other hand directed evolution needs no structural information and small changes, that are sometimes random, are made over several rounds of evolution (Steiner and Schwab, 2012). A mix between the two approaches is semi-rational engineering that uses smaller focused libraries that are designed using structural or biochemical knowledge (Steiner and Schwab, 2012).

1.7.1 Rational engineering

One of the first examples of rational engineering was linked to the determination of the trichodiene synthase reaction mechanism. The first step in engineering targets the aspartate motif, specifically D100. The mutation D100E leads to a 22-fold loss in activity as the active site closure is impaired and only two of the three magnesium ions are bound (Rynkiewicz, Cane and Christianson, 2002). D101E had a similar effect but mutation of D104E has no significant change in kinetic parameters (Rynkiewicz, Cane and Christianson, 2002). All three mutants produced a variety of additional products including a number of linear products (Figure 1-19) (Cane and Xue, 1996) (Cane, Xue and Fitzsimons, 1996). Following this the diphosphate interaction was probed, specifically residue Y305 (Figure 1-19). The Y305F structure shows a significant active site deviation, losing the hydrogen bond with the pyrophosphate without impacting the whole structure.

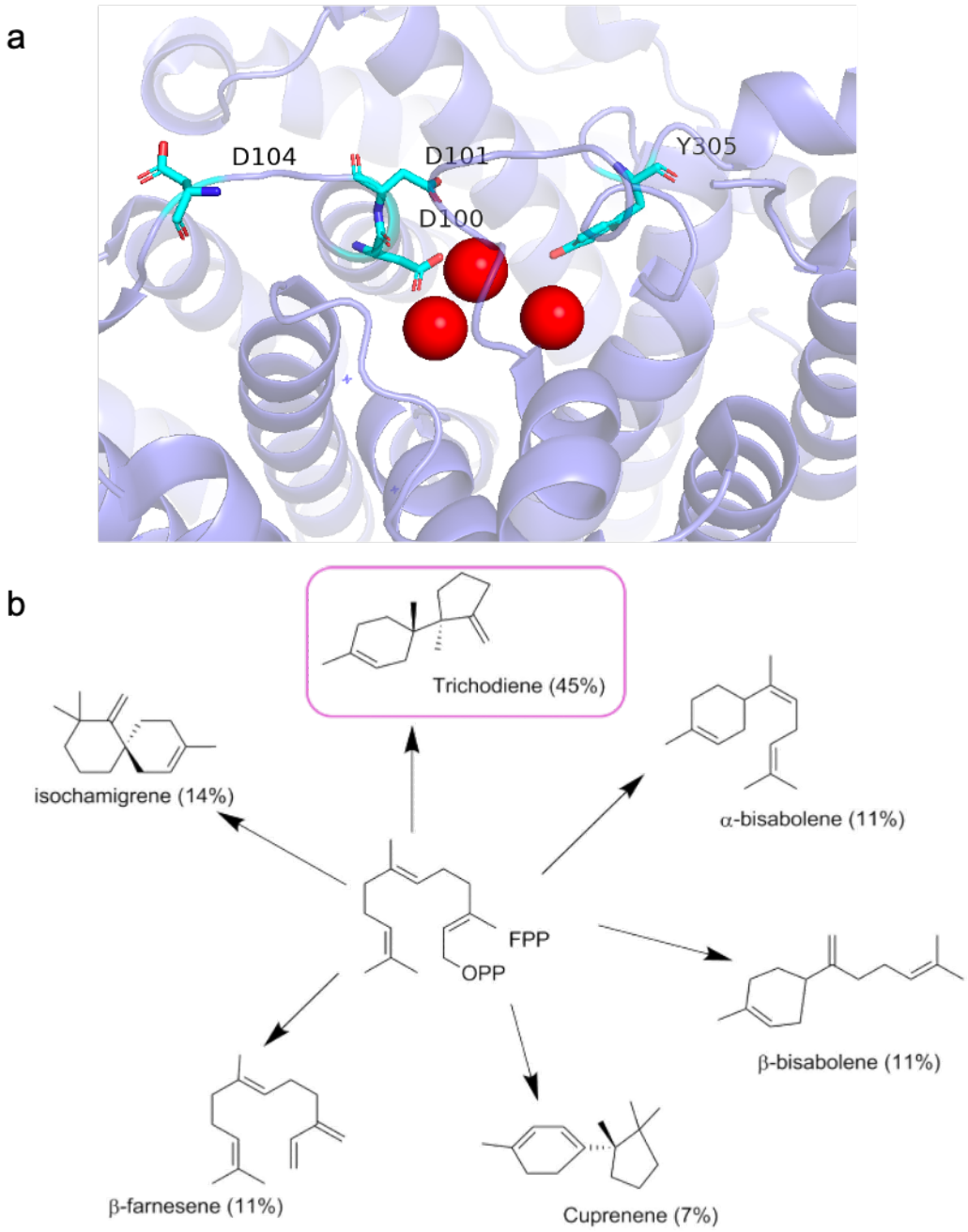


Figure 1-19 Structural image of trichodiene synthase active site (2Q9Z) and Schematic of some of the alternative products for trichodiene synthase mutants. *(a) Pymol image of structural image of the active site of trichodiene synthase highlighting the residues targeted in mutagenesis. (b) All mutants produced trichodiene as the predominant product, but mutants also produced a range of linear and cyclic products with examples of some of the key products and their approximate product profile ratio for mutant D100E .*

This initial engineering was an inspiration for further engineering that moved towards targeting the hydrophobic active site. One example is the selection of single point mutations in epi-isozizaene synthase (Li *et al.*, 2014). Initial work focused on F96 in the hydrophobic

cleft and was targeted with a range of mutations creating a variety of different products. In addition 26 single-site mutations targeting three areas; aromatic triad F95, F96 and F198 that encircle the active site, aromatic residues in the active site contour W203, W325 and F332 and aliphatic active contour residues L72 and V329 (Li *et al.*, 2014). Overall significant changes were seen with decreases in activity between 11-730 fold lower however, F332A had the most significant change of over 800-fold (Li *et al.*, 2014). Mutagenesis produced 6 separate sesquiterpene synthases with alternative predominant products (Figure 1-20). In general, the aromatic residues within the active site are a hot spot for reprogramming or engineering terpene synthase.

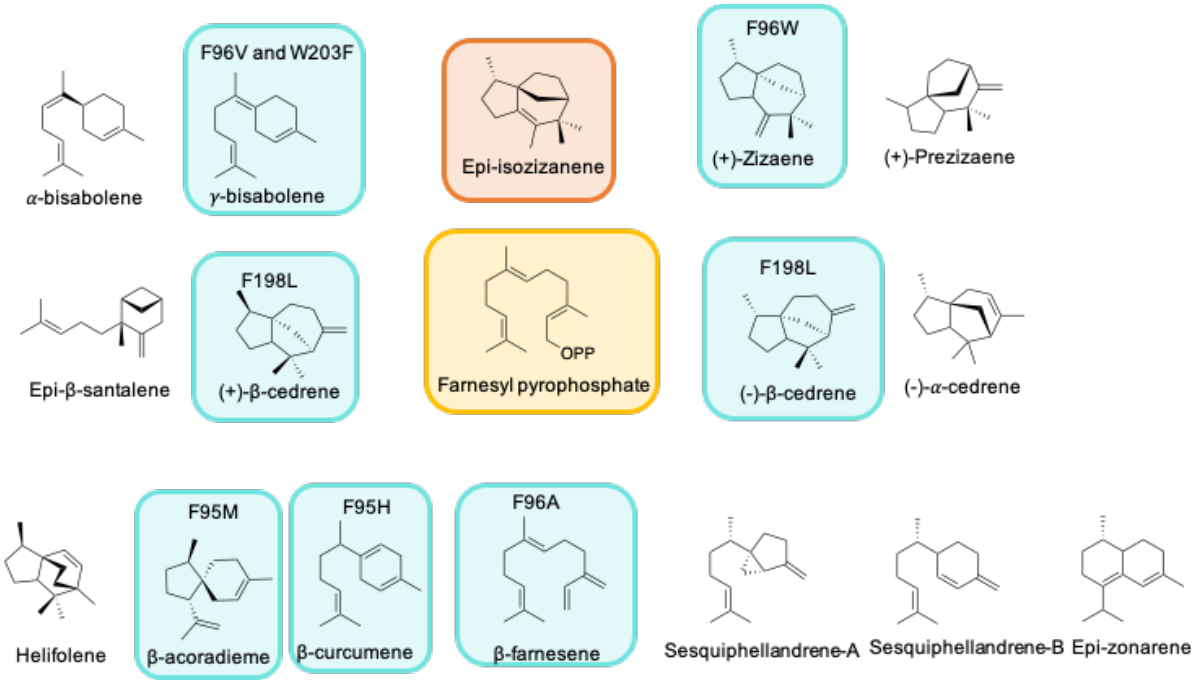


Figure 1-20 Schematic of the products of wild type and mutant epi-isozizanene synthase. *The single substrate farnesyl pyrophosphate is shown in yellow with the wild-type product epi-isozizanene shown in orange. The products highlighted in blue are the predominant products of the mutants above showing switches of specificity. The other products surrounding the key products are minor products produced by the mutants.*

Site-directed mutagenesis can also have a specific purpose; for example, a switch in cyclisation from 1,10 to 1,11 (Figure 1-21). Structural alignments of multiple 1,10 and 1,11 synthase structures or homology models were analysed and identified a triad of residues in germacrene A synthase to target; T410, G402 and G403 (Gonzalez *et al.*, 2014). Generally increasing the residue size at G402 progressively leads to an increase in 1,11 cyclisation (Gonzalez *et al.*, 2014). G402A and G402S both produce α-humulene with the 1,11 cyclisation however 1,10 cyclisation products are still produced (Figure 1-21). G402C

sterically hinders the active site and 1,11 cyclisation activity was increased to 68 % of the final activity. However when G402 is mutated to leucine, phenylalanine or valine 1,11 cyclisation activity is reduced or lost (Gonzalez *et al.*, 2014). This was repeated with germacrene D and G404, G404V produces 20% bioclogermacrene which is a cyclopropane ring underlining dual 1,10 and 1,11 cyclisation (Gonzalez *et al.*, 2014).

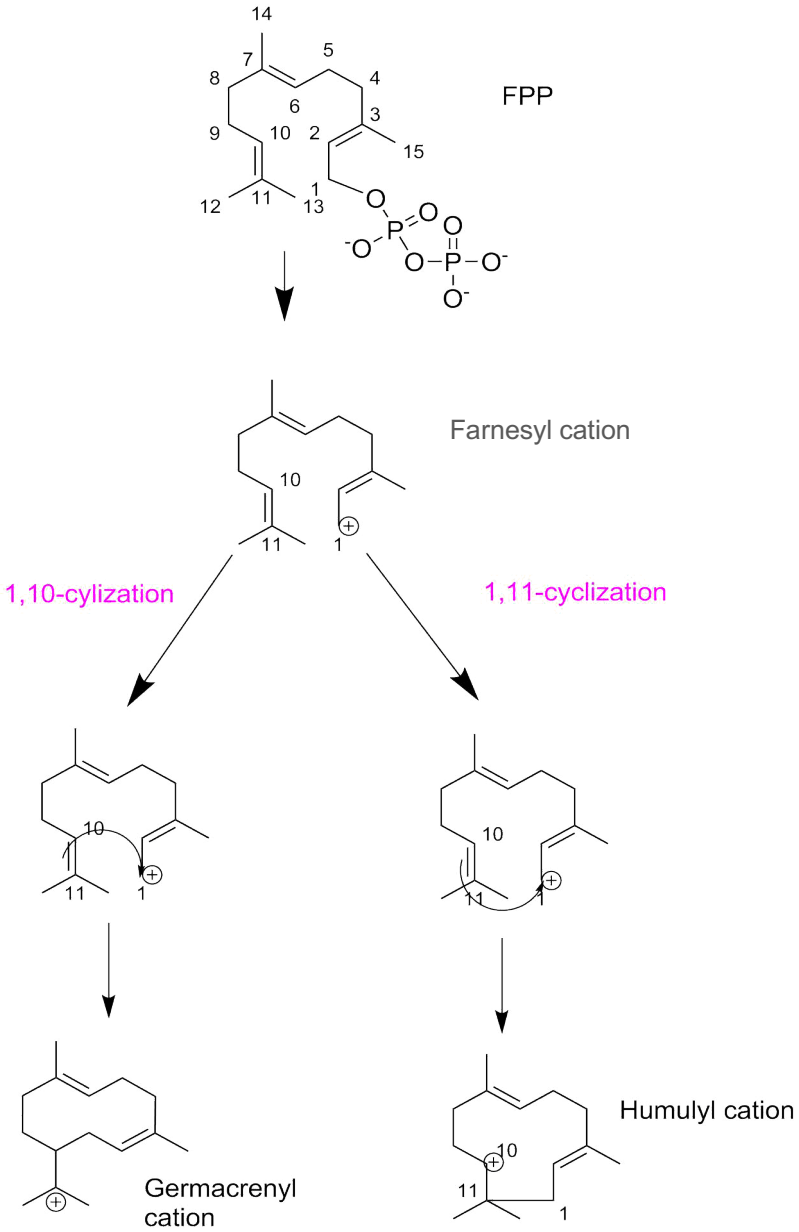


Figure 1-21 Schematic of 1,10 and 1,11 cyclisation of farnesyl cation. *The schematic depicts the 1,10-cyclisation producing the germacrenyl cation, a two ringed structure. The schematic also depicts a 1,11-cyclisation forming a large extended ring, humulyl cation.*

An additional site-directed mutagenesis focus is to select between two products. Taxadiene synthase naturally produces 2 taxadiene isomers; taxa-4(5),11(12)-diene and taxa-

4(20),11(12)-diene (92:8). The following Taxol production pathway can act on both isomers but acting on taxa-4(20),11(12)-diene has higher specificity improving the production (Edgar *et al.*, 2017). The first step was alanine scanning to identify residues directly involved in the final step in the reaction pathway, the proton abstraction. No single residue was identified as responsible as there was no complete product profile switch (Edgar *et al.*, 2017). The number of residues targeted was increased and saturation mutagenesis was completed in the predicted active site cap. Some mutants showed a specificity switch, however this was achieved by reducing the production of the undesired isomer (Edgar *et al.*, 2017). The most successful mutant was Y688L (a residue predicted to be close to the pyrophosphate moiety) with a 2.4-fold increase in taxa-4(20),11(12)-diene production (Edgar *et al.*, 2017). Y688L was co-expressed with the remaining pathway and was combined with further upstream engineering, increasing Taxol production (Edgar *et al.*, 2017).

So far, most site directed mutagenesis has focused on maintaining the hydrophobicity of the hydrophobic active site contour, however mutagenesis strategies progressed to including mutagenesis with polar residues. Epi-isozizaene synthase has been a previous mutagenesis target and from these results residues Y69, F95, F96 and W203 were selected to create 16 additional polar mutations (Blank *et al.*, 2017). This mutagenesis yielded functional enzymes and expanded the chemo diversity with 8 new products, however activity is between 21 and 44% of wild-type (Figure 1-22) (Blank *et al.*, 2017). These products are acyclic and cyclic hydroxylated products due to the binding of a water molecule. Y69 mutations largely maintain the wild-type product profile. F95 mutations have a small effect of the active site and all mutants generate epi-isozizaene along with side products (Blank *et al.*, 2017). W203H and W302Y both produce γ-bisabolene and sesquisabiene, W203H also generates α-bisabolol (Blank *et al.*, 2017). As with nonpolar substitution, polar substitutions at F96 generate significant quantities of sesquisabinene and minor pathway termination products.

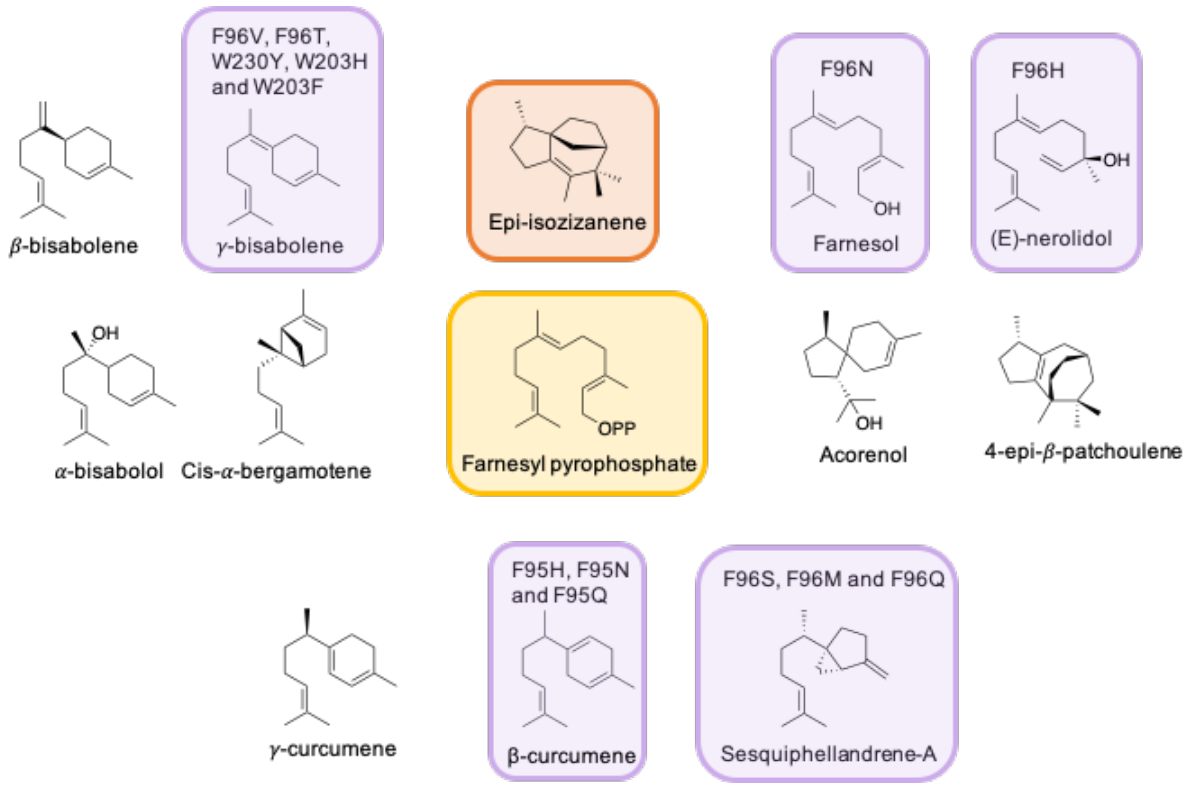


Figure 1-22 Schematic of the products of wild-type and polar mutant epi-isozizanene synthases. *The single substrate of farnesyl pyrophosphate is shown in yellow with the wild-type product epi-isozizanene shown in orange. The products highlighted in purple are the predominant products of polar epi-isozizanene synthase mutants. Some products are consistent with previous mutagenesis attempts, but some are novel products. The surrounding products are minor products produced by a number of the polar mutants.*

A novel way to expand the terpenoid repertories is to use terpene synthase to act on non C5 isoprenoid substrates. The first example of this was to evolve a C11 terpene synthase from a monoterpene synthase. The first step was to use pathway engineering to increase the production of 2-methyl-gernayl pyrophosphate and an overall combination of engineering methods increased production by 760-fold (Ignea *et al.*, 2018). Following this 7 monoterpene synthases were tested with 2-methyl-gernayl pyrophosphate, producing 36 C11 compounds but each monoterpene synthase still favoured geranyl pyrophosphate (Ignea *et al.*, 2018). To engineer a C11-specific terpene synthase SfCinS1, a 1,8-cineole synthase from *Salvia fruticose,* was selected as the start point due to its corresponding crystal structure (Ignea *et al.*, 2018). The first target was to expand the narrow fit active site to allow extended conformations of 2-methyl-gernayl pyrophosphate at N338 and I451 (Ignea *et al.*, 2018). The best mutant is N388S-I451A with 88% of its profile being C11 products, specifically 2-methylmycrene (Ignea *et al.*, 2018). Following further investigation of the SfCinS1 a residue

involved in the closing of the active site could be a C11 single residue switch, F571. In homologous monoterpene synthase this residue is typically an aromatic amino acid F or H (Ignea *et al.*, 2018). F571H improved specificity for 2-methylenebornane and produced 84.7% C11 products. F571Y produces 2-methyl-α-terpineol as its major product. Mutating analogous residues in other monoterpene synthase showed a similar switch to C11 specificity (Ignea *et al.*, 2018). This approach could now progress to other larger isoprenoid substrates and terpene synthase.

1.7.2 Directed evolution

Directed evolution is best suited to situation where little structural or biochemical information is available. This is the case for Cop2 a multi-product sesquiterpene synthase from *Coprinus cinereus* which was engineered with the aim of improving the product selectivity (Lauchli *et al.*, 2014). Error prone PCR was the mutagenesis strategy was used to produce random mutants which were screened through a cyclization activity assay using a non-natural substrate, vinyl methyl ether producing methanol. Methanol was detected with alcohol oxidase and purpald reagent which was analysed by measuring the absorbance at 500 nm (Lauchli *et al.*, 2014). The most successful mutant found through screening was 17H2 (L59H, T65A and S310Y) with a 77% specificity toward germacrene-D-4-ol and produced no α-cardinol. When the mutations were made separately small changes were seen showing the mutations are synergistic for significant changes (Lauchli *et al.*, 2014). All of the mutations are predicted to be near the active site but do not line the active site or directly interact with the substrates (Lauchli *et al.*, 2014).

Another directed evolution method is to target specific positions with degenerate codons meaning a selection of amino acids could be inserted at the position. Using an automated pipeline and liquid handling coupled to GC/MS, libraries can be screened at high throughput. This method was used on monoterpene α-pinene synthase variant (VAR3-PinS) which produces an equal mixture of linear, monocyclic and bicyclic monoterpenes (Leferink *et al.*, 2019). 16 positions were targeted within the three areas of plasticity, previously identified in α-pinene synthase, in the hydrophobic active site. To target these positions the NBT degenerate codon (a codon that could introduce more than one amino acid randomly) was selected covering polar and hydrophobic residues including; G, A, V, I, L, F, P, C, T, S and R (Leferink *et al.*, 2019). 1000 colonies were screened with 65 unique sequences and results were validated with using larger culture volumes and traditional expression shake flasks. The positions were scored based on the spread of amino acids found at these positions (Leferink *et al.*, 2019). The highest scoring positions produce the greatest product diversity,

specifically residue 335, which corresponds to other key product determination residues in monoterpene synthases (Leferink *et al.*, 2019). The results also identify the key residues in bicyclic products (557-563) and in linear products (443,447 and 448) determination (Leferink *et al.*, 2019).

Directed evolution can also be used to evolve other biochemical properties including thermostability. This was completed on BcBOT2 (presilphiperforlan-8 β-ol synthase) and utilised the high throughput screen with a vinyl methyl ether modified substrate, alcohol oxidase and purpald reagent (Lauchli *et al.*, 2014). 2800 BcBOT2 mutants were produced by error-prone PCR with an average of 3 base pairs mutated per mutant. Following library generation, it is treated for 10 minutes at 45 ºC. The most improved mutant was mutant 19B7, which is missing 3 *C*-terminal amino acids and K58R mutation. The T_{50} was increased from 42 to 47 ºC with no change in expression, activity or specificity. 19B7 was then used as the template to further increase the T_{50} to 54 ºC with the additional mutation H383R.

1.7.3 Semi-rational engineering

Semi-rational engineering combines the use of structural information to create targeted libraries and the use of screening or mutagenesis methods in directed evolution. An additional directed evolution method is structure-based combinatorial protein engineering (SCOPE) uses structural information coupled with DNA manipulation techniques creating multiple crossover libraries from non-homologous genes (Figure 1-23) (O'Maille, Bakhtina and Tsai, 2002). One example used amorpha-4,11-diene synthase from *Artemisia annua* that produces cyclic terpenes and (E)- β-farnesene synthase that produces linear terpenes including (E)- β-farnesene but with similar mechanisms (Salmon *et al.*, 2015). Structural analysis identified 24 positions to target within a 6 Å radius of the active site. SCOPE was used to create diversity in 9 libraries with varying number of mutations with each library screened 3 times (Salmon *et al.*, 2015). 32 mutants from each library were sampled with 94 % inactive, 5 % wildtype activity, 1% cyclic activity. Y420L was the most common cyclic mutation producing 15 cyclic terpenes and this activity was abolished when the mutation is removed (Figure 1-24) (Salmon *et al.*, 2015). This mutation can be combined with Y430A to enhance cyclic catalytic efficiency (Salmon *et al.*, 2015). V467G reverts any cyclase activity to farnesene activity while reducing efficiency (Salmon *et al.*, 2015). Also with accumulating number of mutations there was an exponential decline in efficiency therefore, a select number of mutations are more efficient (Salmon *et al.*, 2015).

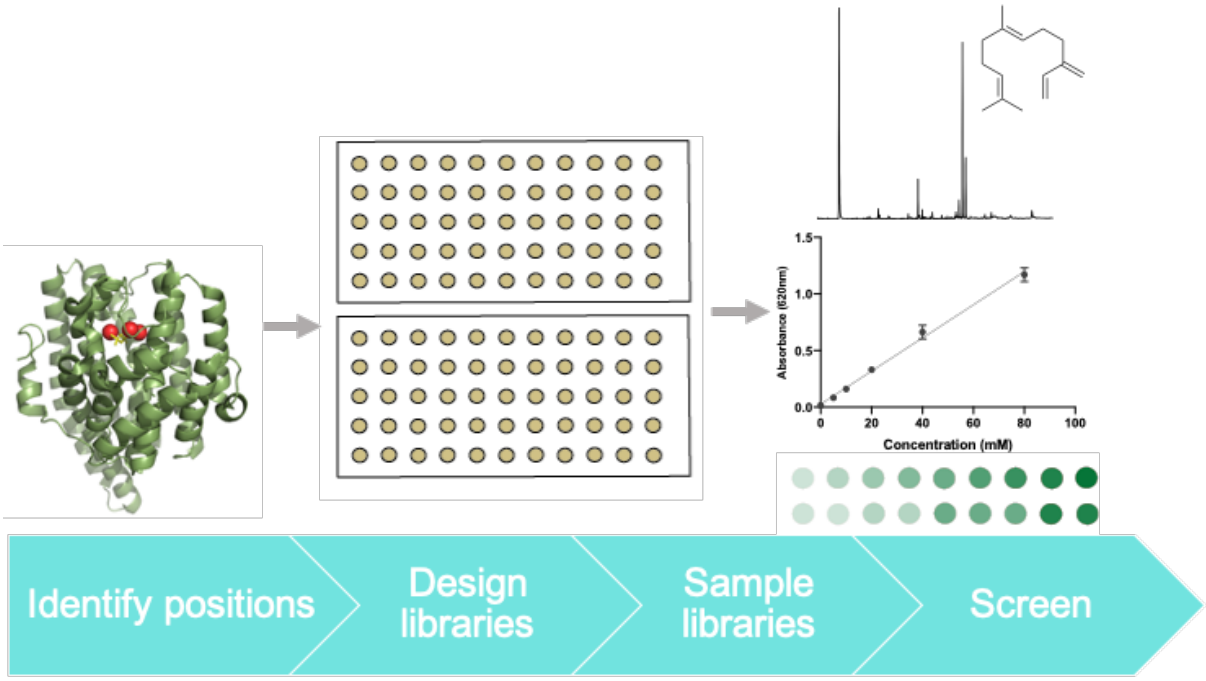


Figure 1-23 Schematic of the structure-based combinatorial protein engineering (SCOPE) process. *The schematic identifies positions to target using a crystal structure or a homology model. The structure included is epi-isozizaene synthase that was used to create the homology model of amorpha-4,11-diene synthase. Libraries for each position are designed and produced using DNA manipulation techniques. A selected number of mutants from each library were screened. In this example mutants were screened with the malachite green activity assay and GC/MS product profile analysis. Mutants identified could then be taken forward or used templates in further rounds of SCOPE to improve properties.*

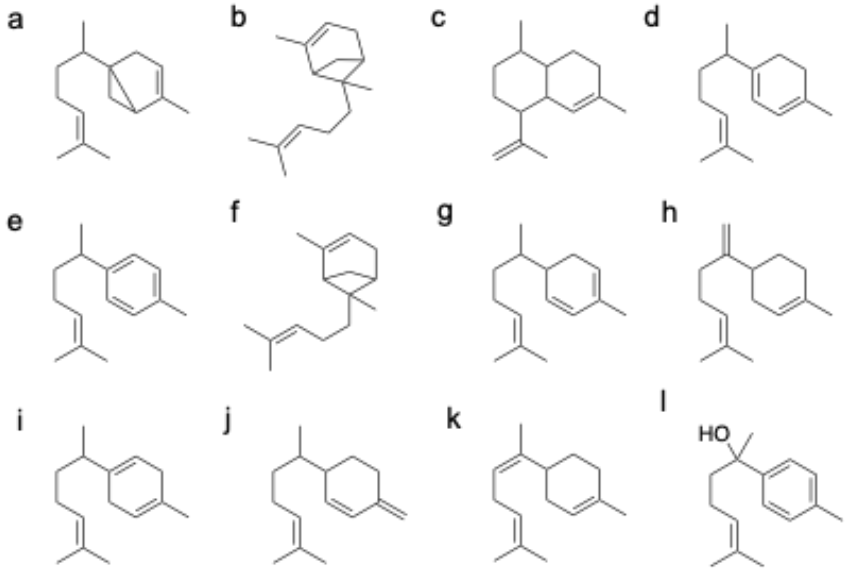


Figure 1-24 Schematic of Y402L SCOPE products. *(a) Structure of sesquithujene. (b) α-exo-bergamotene. (c) Structure of an amorphadiene isomer. (d) Structure of γ-cumene. (e) Structure of arcurcumene. (f) Structure of cis- α-bergamotene. (g) Structure of zingiberene. (h) Structure of β-bisabolene. (i) Structure of β-curcumene. (j) Structure of β-sesquiphelladrene. (k) Structure of α-bisabolene. (l) Structure of α-bisabolol.*

Another example of semi-rational design is the probing of the plasticity residues in γ-humulene synthase. γ-humulene synthase is a promiscuous terpene synthase that can produce over 52 products with a variety of regio-chemistries and stereo-chemistries (Yoshikuni, Ferrin and Keasling, 2006). To determine the best areas to target a homology model was produced using the structure of 5-epi-aristolochene synthase. From the structural homology model 19 residues in active site contour were selected for saturation libraries (Yoshikuni, Ferrin and Keasling, 2006). The product profiles of each mutant were normalised against the wild-type product profile to spot altered specificity. This analysis identified 4 key residues that significantly shifted selectivity of γ-humulene synthase; W315, M447, S484 and Y566. The residues were further investigated using an algorithm to predict a product specificity change. The algorithm was successful in predicting a β-bisabolene synthase with both predictions successfully switching specificity and maintaining activity (M447H/A336V/I562T and M447H/A336V/I562V) (Yoshikuni, Ferrin and Keasling, 2006). Plasticity residues behaved as if they were independent and can be the driving force behind molecular evolution, while the mutations were additive (Yoshikuni, Ferrin and Keasling, 2006).

Overall the mutagenesis effort has led to a variety of product specificity and enzyme property changes while giving insight into how terpene synthases control product diversity and reaction mechanisms. This knowledge was used to create an aromatic enzyme-mimicking supramolecular capsule to catalyse the cyclisation of a sesquiterpene (Zhang *et al.*, 2018). Isolongifolene was produced by an artificial enzyme mimic that combined the terpene synthase features of encapsulating the substrate in a hydrophobic environment and controlling the orientation of the substrate (Zhang *et al.*, 2018).

1.8 Metabolic engineering for the production of terpenes

As terpenes have such a wide range of uses, they are targets for pathway optimisation to increase their production. Pathway optimisation can target substrates and products and can be done in native and non-native species. Metabolic engineering uses genetic engineering tools to modulate existing or heterologous pathways in cells to overproduce target chemicals with high yields, rates and with reduced stereochemistry issues (Chen, Zhang and Lindley, 2019). Metabolic engineering involves understanding metabolism to keep the overall metabolic flux harmonized and utilise the molecular biology and high throughput analytical tools and techniques to create a systematic approach (Ajikumar *et al.*, 2008).

Native expression and engineering of terpenes especially in plant cell culture has a number of associated challenges. One of the largest challenges is the fact that large number of terpenoid pathway elements are undefined including regulation and rate limiting steps (Roberts, 2007). Few plant genomes have been fully sequenced, and this can pose problems moving between whole plant and plant cell culture studies and using traditional genetic methods. Without sequencing information there are no known plant inducible promoters unlike in heterologous platforms (Roberts, 2007). In addition plant cells can have a large variability in terpene productions and are generally 2-3 % of the plant dry weight (Roberts, 2007).

There has only been one notable example of engineering endogenous plant terpenoid pathways - the increase in mint oil production in *Mentha x piperita* (Withers and Keasling, 2007). Over expression of DXP reductoisomerase by 50 % increased essential oil production. Essential oils are a complex mix of terpenoids and the ratio of peppermint in essential oils can be modulated by the production of menthofuran. The production of methofuran is targeted through transcriptional regulation of menthofuran synthase as it down regulates pulegone reductase (Mahmoud and Croteau, 2003).

There are two main heterologous industrial production platforms; *E. coli* and *S. cerevisiae* (Chen, Zhang and Lindley, 2019). These production platforms are generally safer, higher yielding and sustainable as extraction from plant material is expensive due to the low yields and complicated extraction methods (Chen, Zhang and Lindley, 2019). Heterologous expression produces higher yields as there is no feedback regulation as a native control but high yields can lead to toxicity (Chang and Keasling, 2006). Heterologous hosts may produce the precursor and substrates but it is hard to predict the levels and how to balance the flux of metabolites (Chang and Keasling, 2006).

1.8.1 Balancing pathways and building blocks

A large focus of engineering terpenes involves increasing the supply of precursors through overexpression and heterologous expression. There are three key stages which can be targeted for engineering; transcription, translation and post-translational modifications. To target at the level of transcription involves the balancing of promoters to control the levels of mRNA. The use of model organisms as industrial platforms is successful as promoters have been well characterised to fine tune the balance of mRNA. The use of promoters can be used to upregulate the desired genes while also using weak promoters to downregulate genes that could divert resources (Chen, Zhang and Lindley, 2019). For example the temperature sensitive promoter GalM9 was used to increase the production of astaxanthin synthase in *S. cerevisiae* when the temperature of the culture is lowered post induction (Chen, Zhang and Lindley, 2019). The second stage to target is translation specifically targeting the ribosomal binding site and the 5'-UTRs. Mutations of 3 or more nucleotides in the ribosomal binding site region can alter translation initiation rate by 50-fold. Post translation involves the co-localisation of enzyme to increase local substrate concentration and facilitate substrate channelling. One way to do this is to fuse two sequential enzymes in the pathway, in close proximity the diffusional limitations are reduced increasing the reaction rate. Overall all three stages need to be engineered in a push and pull manner to increase precursors and co-factor availability without reaching a toxic level. This can generally improve pathway flux when the balance is reached and is functioning (Chen, Zhang and Lindley, 2019).

The first stage in increasing production is to identify the rate limiting steps of the pathways to increase the overall flux of precursor production. The rate limiting steps in the methylerythritol phosphate pathway are either the reactions of DXS, DXR or IspC and IPP-DMAPP isomerase (Figure 1-4) (Chang and Keasling, 2006). To overcome rate limiting steps, combinations of precursor feeding, modulation of gene expression and gene deletion experiments were used. Increased expression of DXS and IPP-DMAPP isomerase

increased the precursor pool and balanced the ratio G3P and pyruvate were combined with carotenoid synthase expression increasing the yield 10-fold (Ajikumar *et al.*, 2008). This level was further increased when combined with overexpression of geranylgeranyl pyrophosphate synthase to a 50-fold improvement (Ajikumar *et al.*, 2008). The flux of the pathway remains limited by the intracellular regulatory mechanisms of the organism.

Heterologous expression of a pathway can overcome the intracellular regulation limits. The methylerythritol phosphate pathway is unique to eubacteria and plants but it is not suitable for heterologous expression in *S. cerevisiae* due to the lack of iron clusters for IspG and IspH (Chen, Zhang and Lindley, 2019). Whereas the mevalonate pathway has been successfully heterologously expressed in 2 synthetic operons (Figure 1-4) (Chang and Keasling, 2006) (Withers and Keasling, 2007). MevT is a plasmid that encode the enzymes that convert acetyl-coA to mevalonate. MBis is a plasmid that converts mevalonate to farnesyl pyrophosphate (Chang and Keasling, 2006) (Withers and Keasling, 2007). Expression of just the pathway leads to precursor accumulation toxicity, however co-expression with codon optimised terpene synthase relieves the toxicity (Chang and Keasling, 2006). An example of this is the co-expression of codon optimised amorphadiene synthase to reduce toxicity and the testing of the mevalonate pathway in an *E. coli* strain deficient in isoprenoid synthesis (Martin *et al.*, 2003).

This mevalonate pathway spilt has been further studied in *E. coli* for the optimisation of protoilludene production. Protoilludene derivatives have medically relevant properties being used as antitumor and antimicrobial activities but purification of protoilludene has low yields (Yang *et al.*, 2016). MvU is a plasmid containing the sequences for the enzymes; acetyl-coA acetyltransferase, HMG-CoA reductase, HMG-CoA synthase. MvL is a plasmid containing the sequences for the enzymes; mevalonate kinase, phosphomevalonate kinase, pyrophosmevalonate decarboxylase, IPP isomerase. The third section is the farnesyl pyrophosphate and codon optimised protoilludene synthase (Yang *et al.*, 2016). The balance between the expression of the MvU and MvL was altered to increase the overall production of protoilludene. For MvL the gene order was arranged in order of activity level from low to high. In addition, MvL was further optimised by random sequential permutations throughout the plasmid (Yang *et al.*, 2016). Whereas MvU expression was modulated by changes to the promoter causing varied transcription levels. Also a homolog of mevalonate kinase was substituted in which further increased protoilludene (Yang *et al.*, 2016).

Increasing the supply of acetyl-CoA for the native mevalonate pathway in *S. cerevisiae* would therefore increase the production of amorphadiene. To increase acetyl-CoA the

method focused on primary metabolism using a pyruvate dehydrogenase bypass (Shiba *et al.*, 2007). The pyruvate dehydrogenase bypass converts pyruvate to acetyl-CoA using a combination of pyruvate decarboxylase, cytosolic acetaldehyde dehydrogenase and acetyl-CoA synthase (Shiba *et al.*, 2007). The combination of enzymes was heterologously overexpressed and tested with a variety of promoters to find the optimal balance (Shiba *et al.*, 2007). Engineered *S. cerevisiae* strains have been tested with heterologous expression of novel terpene synthase found through mining bacterial genome data and characterised 13 novel cyclic terpenes (Yamada *et al.*, 2015). This shows engineered platforms can be used to facilitate the discovery or characterisation of terpene synthase and can be used alongside computer predictive frameworks predicting function or activity (Chen, Zhang and Lindley, 2019).

1.8.2 Terpene synthase pathway engineering examples

The most significant engineering effort is for the production of anticancer Taxol by targeting taxadiene synthase and the following enzymes. Taxol is natively produced by *Taxus chinenis* and is extracted from bark of yew trees. However, this extraction is only 0.001 - 0.002 % efficient and therefore cannot meet the demand (Engels, Dahm and Jennewein, 2008). Taxadiene synthase and Taxol production have been metabolically engineered in both *S. cerevisiae* and *E. coli. S. cerevisiae* engineering initially involved the heterologous genes for the biosynthetic enzymes, that constitute the early Taxol pathway, isoprenoid pathway and regulatory factors inhibiting competitive pathways (Figure 1-25) (DeJong *et al.*, 2006). *S. cerevisiae* natively uses the mevalonate pathway and its main feedback regulation is HMG-CoA reductase. To target HMG-CoA regulation the *N*-terminal regulation domain is removed to stop the negative feedback regulation. Expression of this truncated version increased taxadiene production by 50 % (Figure 1-25) (Engels, Dahm and Jennewein, 2008). An additional method to increase flux is to reduce competing pathways that include sterol production using geranylgeranyl pyrophosphate synthase from *S. acidocaldarius* to compete for resources with farnesyl pyrophosphate synthase. *S. acidocaldarius* geranylgeranyl pyrophosphate synthase can utilise DMAPP and IPP building blocks to produce geranylgeranyl pyrophosphate. This was combined with a previously characterised transcription factor UPC2 allele mutant that allows steroid uptake in aerobic conditions to utilise steroid uptake from the culture medium (Figure 1-25) (Engels, Dahm and Jennewein, 2008) (DeJong *et al.*, 2006). Also taxadiene synthase was expressed without the *N*-terminal plastid target sequence to improve expression (DeJong *et al.*, 2006). Despite this there was insufficient geranylgeranyl pyrophosphate so the geranylgeranyl pyrophosphate synthase gene was put under the control of the strong constitutive phosphate glycerol kinase promoter using glucose as an activator in yeast fermentation (DeJong *et al.*, 2006). A combination of

all of these engineering steps led to a 40-fold increase in taxadiene levels to 8.7 mg/L (Figure 1-25) (DeJong *et al.*, 2006) (Engels, Dahm and Jennewein, 2008).

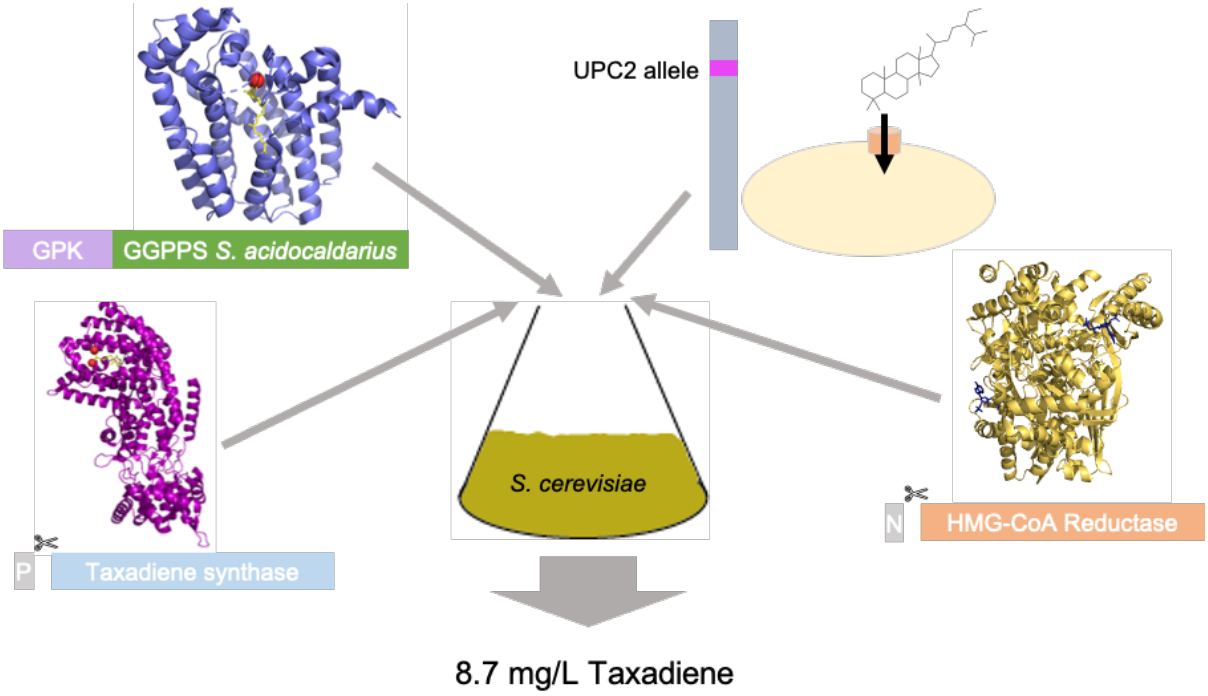


Figure 1-25 Schematic of the engineering attempts for taxadiene production in S. cerevisiae*. The engineering efforts included expression of* S. acidocadarius *geranylgeranyl pyrophosphate synthase under the strong glycerol kinase promoter and the removal of the plastid targeting sequence from* Taxus chinenis *taxadiene synthase. In addition, the UPC2 allele mutant allows steroid uptake from culture medium in aerobic conditions. The was combined with the expression of the truncated HMG-CoA reductase. Overall the engineering efforts produce 8.7 mg/L of taxadiene.*

Taxol production was also engineered in *E. coli* utilising the native methylerythritol phosphate pathway. The methylerythritol phosphate pathway consists of 8 genes with 4 rate determining gene expressions modulating the production of IPP. Downstream of this pathway is a two gene heterologous pathway to produce taxadiene (Ajikumar *et al.*, 2010). The balance of both pathways needs to be achieved so the downstream steps can accommodate the upstream production levels. The balance is achieved by using a variety of promoters and gene copy numbers to modulate each step. The strongest promoter was used for the downstream steps but alongside a low copy number to increase the taxadiene heterologous pathway (Figure 1-26) (Ajikumar *et al.*, 2010). The upstream enzymes were expressed at lower levels and were chromosomally integrated for stability. High levels of specific intermediates can lead to inhibition, so the lower expression levels increase the overall production without overwhelming the culture (Ajikumar *et al.*, 2010). In addition, a

variety of chimeric P450 enzymes were co-expressed in engineered strains for the oxidation of taxadiene to taxadiene-5 α-ol (Figure 1-26). This combination of methods allows *E. coli* to produce 1 g/L in fed-batch bioreactor fermentations (Ajikumar *et al.*, 2010). Other Taxol production engineering strategies in *E. coli* include removing the signal peptide of taxadiene synthase as in *S. cerevisiae* to increase soluble expression, enzyme fusions and overexpression of rate limiting steps (Huang *et al.*, 2001). In addition, fusions of IPP isomerase and geranylgeranyl pyrophosphate synthase with a 21 amino acids linker were overexpressed to increase taxadiene synthase substrate production (Figure 1-26) (Huang *et al.*, 2001). Along with previous methods rate limiting enzymes in the methylerythritol phosphate pathway including DXP synthase genes were collected on one plasmid to increase soluble protein to overcome the rate limiting steps (Huang *et al.*, 2001).

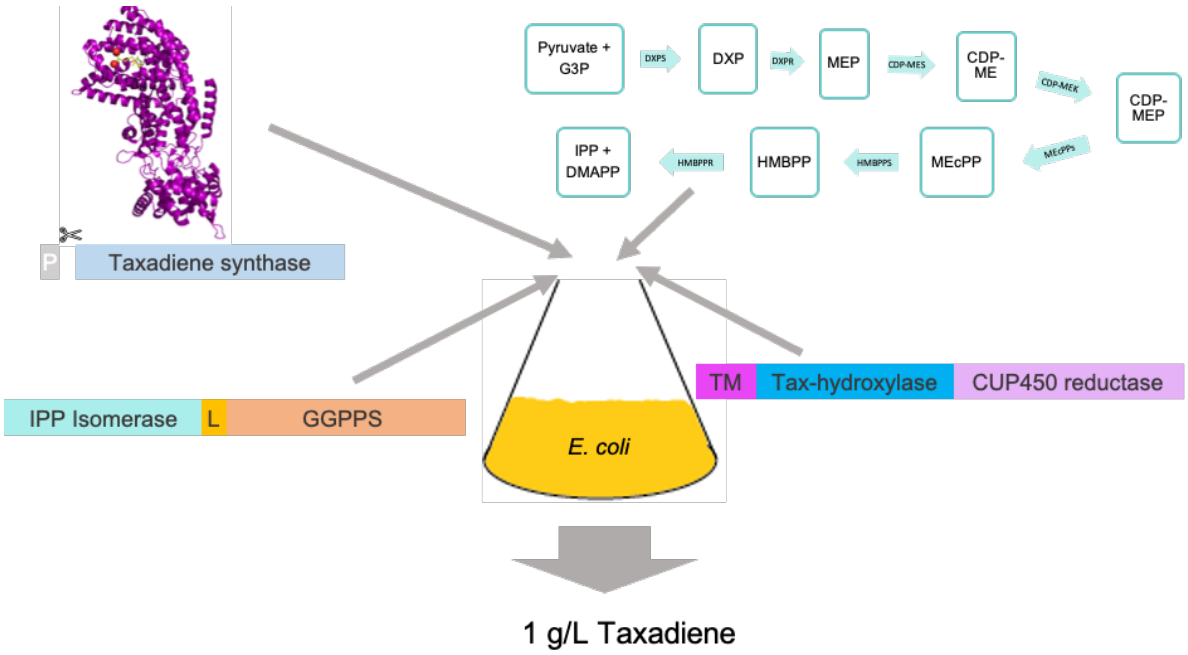


Figure 1-26 Schematic of engineering taxadiene production in E. coli. *The engineering efforts included the removal of the plastid targeting sequence from* Taxus chinenis *taxadiene synthase and engineering the methylerythritol phosphate pathway. In addition, a fusion protein of IPP isomerase and geranylgeranyl pyrophosphate synthase reducing the substrate diffusion rate. Also, the chimeric P450 enzyme oxidise taxadiene to taxadiene-5 α-ol the next step in the Taxol pathway. Overall the engineering efforts yielded 1 g/L taxadiene or taxadiene-5 α-ol.*

Amorphadiene is another medically relevant terpene that is chemically functionalised to the terpenoid artemisinin which has antimalarial activity. Production of artemisinic acid is engineered in *S. cerevisiae* through engineering the mevalonate pathway, amorphadiene synthase, and novel cytochrome P450 monooxygenase. The first target was to upregulate

farnesyl pyrophosphate synthesis with the upregulation of truncated soluble HMG-CoA reductase (Ro *et al.*, 2006). This is combined with reducing the steroid production by increasing steroid uptake with UPC2-1 mutant allele, with the down regulation of squalene synthase (ERG9) (Chang and Keasling, 2006) (Ro *et al.*, 2006). This was combined with a novel cytochrome P450 to convert amorphadiene to artemisinic acid. Cytochrome P450 fragments from *Artemisia annua* were isolated and a P450 of close linage was found that could catalyse hydroxylation of terpenoids. This is active on amorphadiene and was put under control of a galactose-inducible promoter with its native redox partner NADPH (Ro *et al.*, 2006). This engineered strain of *S. cerevisiae* produces artemisinic acid two orders of magnitude higher than the native host. The extraction of artemisinic acid was simple as it is released from the cells and can be separated with alkaline buffer washing (Ro *et al.*, 2006) (Chang and Keasling, 2006).

Bisabolane fuel is an alternative to D2 diesel fuel and is produced by hydrogenating bisabolene (McAndrew *et al.*, 2011). Bisabolene production was engineered in both *E. coli* and *S. cerevisiae* and from both of these sources bisabolene can be extracted with a solvent layer and hydrogenated to bisabolane (McAndrew *et al.*, 2011). *Abies grandis* bisabolene synthase is a specific terpene synthase and following the screening of a selection of bisabolene synthases it was taken forward due to it high level of bisabolene production (Peralta-Yahya *et al.*, 2011). Further optimisation was achieved by codon optimising the sequences of the methylerythritol phosphate pathway enzymes and changing the promoters to increase production. The overall yield achieved 912 mg/L which is only 4 % of the theoretical yield (Peralta-Yahya *et al.*, 2011). Following this production optimisation in *S. cerevisiae*, engineering in *E. coli* used previously engineered strains as the starting point as well. The engineered strain involved a truncated HMG-CoA reductase, an overexpressed farnesyl pyrophosphate synthase and overexpression of a global transcription regulator of the sterol pathway. Also squalene synthase, an enzyme involved in a competitive pathway, was downregulated increasing available farnesyl pyrophosphate (Peralta-Yahya *et al.*, 2011). The *Abies grandis* bisabolene synthase was put under the galactose promoter on a high copy number plasmid. Overall *E. coli* engineering achieved bisabolene production levels of 994 mg/L.

Pathway engineering can be combined with terpene synthase engineering to further improve the production of a specific terpene in the case of diterpene levopimaradiene. The first step was to target diterpenoid and its precursor biosynthesis through the methylerythritol phosphate pathway. This engineering involved overexpression of the methylerythritol phosphate pathway 10-fold to a point where the downstream pathway is insufficient to cope

with the precursor level (Leonard *et al.*, 2010). To increase the capacity of the downstream pathway both geranylgeranyl pyrophosphate synthase and levopimaradiene synthase were codon optimised and functionally expressed however, there was still significant precursor build up. To further increase the capacity both geranylgeranyl pyrophosphate synthase and levopimaradiene synthase were engineered. First levopimaradiene synthase was engineered using a homology structure to identify 15 potential key residues within the active site. At these positions a saturation or targeted library was produced. The key mutations were M593L increasing 2-fold, M593I increasing 3.7-fold and Y700F, M or W increasing 5-fold. A combination of M593I/Y700F increased production 10-fold (Leonard *et al.*, 2010). There is no structural information on geranylgeranyl pyrophosphate synthase so directed evolution was used and screened with a lycopene biosynthetic pathway and a colorimetric assay with an increasing terpene production increasing the intensity of the red produced (Leonard *et al.*, 2010). The 15 best mutants were screened with the M593I/Y700F levopimaradiene synthase. The best mutation combination with an increase of 17.7-fold was S239C and G295D (Leonard *et al.*, 2010). Overall with 10-fold overexpression of the methylerythritol phosphate pathway, S239C/G295D geranylgeranyl pyrophosphate synthase and M593I/Y700F levopimaradiene synthase expression of levopimaradiene is increased 2600-fold over wildtype (Leonard *et al.*, 2010).

1.9 Project background

This thesis will focus on selecting and characterising a sesquiterpene synthase (Chapter 3) with the goal of engineering this sesquiterpene synthase for the production of novel products (Chapter 4). Through this engineering we hope to gain further insight into the mechanism of action and cation control within the sesquiterpene synthase active site.

Sesquiterpene synthase selection was based on the following criteria: ability to heterologously express in *E. coli*, availability of a crystal structure to allow for rational mutagenesis and no previous extensive engineering. After shortlisting four sesquiterpene synthases, two were selected based on the chemical structure and chemical potential of the primary product. The selected sesquiterpene synthase will be rationally engineered with the aim of producing novel sesquiterpene. The aims of this project were as follows.

1.9.1 To select a suitable sesquiterpene synthase as target for engineering.
Sesquiterpene synthase were selected based on their ability to heterologously express in *E. coli* to a sufficient level along with a suitable purification method (Section 3.1). In addition, the sesquiterpene synthase must have a solved crystal structure without having undergone

extensive engineering to allow for a broad rational mutagenesis strategy. Following the shortlisting of the sesquiterpene synthase the chemical structure and chemical potential of the product was taken into account for narrowing the choice. From this two sesquiterpene synthases were taken forward for expression and purification tests to select the sesquiterpene synthase (Section 3.3).

1.9.2 To kinetically characterise the selected sesquiterpene synthase

Using the malachite green assay and purified enzymes pyrophosphate release from terpene synthases can be detected when coupled with a pyrophosphatase, creating inorganic phosphate ions. These inorganic phosphate ions from a complex with malachite green and ammonium molybdate that forms a green complex that can be measured at 620 nm. The selected sesquiterpene synthase was be tested with a range of farnesyl pyrophosphate concentrations with a set enzyme concentration of a set amount of time and this data could be used to calculate the kinetic parameters. These wild type parameters can then be compared to any generated variants (Section 3.4).

1.9.3 To complete alanine scanning and screen development

The first step of rational engineering uses alanine scanning at select residues across the hydrophobic cleft of the active site. These alanine mutants can then be used to asses screening methods for further mutagenesis. One screening option is to use the malachite green activity assay to test whether the mutants are active but does not confirm the products. An additional screening method is using GC/MS on either purified protein or cell culture samples with a solvent overlay. This method can detect the product profile of the enzymes to detect any changes. Following the screening test the alanine mutants can be evaluated and be used in further rational engineering (Section 4.2).

1.9.4 To complete further rational engineering of sesquiterpene synthase

Using the results from the alanine mutants, further study of the crystal structure and the throughput of the selected screening method a set of mutants was compiled and created using site directed mutagenesis. The selected mutants were screened, and any interesting results were flagged, the mutants were characterised, and these positions were further explored. The aim of further mutagenesis is to produce a sesquiterpene synthase that produce novel sesquiterpene (Section 4.3).

Chapter 2 - Materials and Methods

2.1 Materials

2.1.1 Chemicals

Table 2-1 List of chemicals used in methods described.

Chemical	Supplier
1,2-Dithiothreitol DTT	Fluorochem, Derybshire, UK
Acetone	Fisher Scientfic, Loughborough, UK
Acrylamide 30%	Sigma-Aldrich, Dorset, UK
Agar	Melford Laboratories, Suffolk, UK
Agarose	Melford Laboratories, Suffolk, UK
Ammonium Acetate	Fisher Scientific, Loughborough, UK
Ammonium molybdate	Sigma-Aldrich, Dorset, UK
Ammonium Persulfate	Acros Organics (Fisher Scientific)
Ammonium Sulphate	Sigma-Aldrich, Dorset, UK
Ampicillin	Sigma-Aldrich, Dorset, UK
Bisabolene	Sigma-Aldrich, Dorset, UK
Bovine serum albumin	Bio-Rad, Hertfordshire, UK
Bradford Reagent	Bio-Rad, Hertfordshire, UK
Bromophenol Blue	Sigma-Aldrich, Dorset, UK
CAPS	Calbiochem (Merk Millipore)
Chloramphenicol	Sigma-Aldrich, Dorset, UK
Chloroform-d	Sigma-Aldrich, Dorset, UK
Coomassive Brilliant Blue	Bio-Rad, Hertfordshire, UK
Cumene	Sigma-Aldrich, Dorset, UK
Dimethyl sulphoxide, DMSO	Sigma-Aldrich, Dorset, UK
Dodecane	Sigma-Aldrich, Dorset, UK
Ethanol	Fisher Scientific, Loughborough, UK
Ethylenediamine tera acetic acid, EDTA	Sigma-Aldrich, Dorset, UK
Farensyl pyrophosphate	Isoprenoids, Tampa, USA
Glacial acetic acid	Sigma-Aldrich, Dorset, UK
Glycerol	Alfa Aesar, Lancashire, UK
Glycine	Sigma-Aldrich, Dorset, UK
HEPES	Fluorochem, Derybshire, UK
Imidazole	Acros Organics (Fisher Scientific)

InstantBlue	Expendeon, Cambridge, UK
Isopropyl β-D-1-thiogalactopyranoside, IPTG	Generon, Berkshire, UK
Kanamycin	Sigma-Aldrich, Dorset, UK
Lysozyme (Mammalian)	Sigma-Aldrich, Dorset, UK
Malachite green oxalate salt	Sigma-Aldrich, Dorset, UK
MES Monohydrate	Alfa Aesar, Lancashire, UK
Phenylmethylsulfonyl fluoride, PMSF	Sigma-Aldrich, Dorset, UK
Pierce™ Protease Inhibitor Tablets, EDTA-free	Thermo Fisher Scientific, Loughborugh, UK
Propan-2-ol	Fisher Scientific, Loughborough, UK
Sodium Chloride, NaCl	Across Organics (Fisher Scientific)
Sodium dodecyl sulphate, SDS	Sigma-Aldrich, Dorset, UK
Sodium hydroxide, NaOH	Fisher Scientific, Loughborough, UK
Sodium phosphate	Sigma-Aldrich, Dorset, UK
Sodium pyrophosphate	Sigma-Aldrich, Dorset, UK
Sodium Sulphate	Sigma-Aldrich, Dorset, UK
SYBRsafe DNA stain	Bio-Rad, Hertfordshire, UK
Tetracycline	Sigma-Aldrich, Dorset, UK
Tetramethylethylenediamine, TEMED	Sigma-Aldrich, Dorset, UK
Tris Base	Fisher Scientific, Loughborogh, UK
Tryptone	Melford Laboratories, Suffolk, UK
Yeast Extract	Sigma-Aldrich, Dorest, UK
β-mercaptoethanol	Sigma-Aldrich, Dorset, UK

All restriction enzymes and Q5® High-Fidelity Polymerase, Quick-Load® 1 kb Extend DNA Ladder, yeast inorganic pyrophosphatase and broad range Colour Prestained Protein Standard were purchased from NEB. *Pfu* DNA polymerase was purchased from Agilent. Deyoxynucleotide triphosphate (dNTPs) and 100 bp and 1 kb DNA ladders were purchased from Promega.

2.1.2 Consumables and chromatography reagents and kits

Miniprep kits were purchased from Promega or NEB. Qiaquick® PCR purification Kit and Qiaquick® Gel extraction kit was purchased from Qiagen. Viviaspin® centrifugal concentrators were purchased from Generon.

Unless otherwise stated all chromatography resins and prepacked columns were purchased from GE Healthcare.

2.1.3 Media and antibiotics

All growth media was prepared by weighing out the appropriate ingredients (Table 2-2) before dissolving them in deionsed water. Media was then sterilised by autoclaving at 121°C for 20 minutes at 15 psi. The media was the stored at room temperature until use.

Table 2-2 Growth media components

Solution	Ingredients
Lysogeny Broth (LB)	10 g/L Tryptone 5 g/L Yeast Extract 10 g/L Sodium Chloride
2 TY Media	16 g/L Tryptone 10 g/L Yeast Extract 5 g/L Sodium Chloride
SOC Media	5 g/LYeast Extract 20 g/L Tryptone 0.5 g/L Sodium Chloride 2.5mL/L 1M Potassium Chloride 10mL/L 1M Magnesium Chloride 20mL 1M Glucose
LB Agar Plates	10 g/L Tryptone 5 g/L Yeast Extract 10 g/L Sodium Chloride 20 g/L Agar Bacteriological

Antibiotic stocks were prepared by weighing out the appropriate amount as described in Table 2-3. The antibiotic was dissolved in deionised water or ethanol and filtered using 0.22μM. The antibiotic solution was aliquoted into aliquots between 0.15 and 1ml and stored at -20°C.

Table 2-3 Antibiotic solutions and concentrations

Antibiotic	Stock Concentration and Preparation	Working Concentration
Ampicillin	100 mg/ml in dH_2O	100 µg/ml
Chloramphenicol	25 mg/ml in ethanol	25 µg/ml
Kanamycin	50 mg/ml in dH_2O	50 µg/ml
Tetracycline	10 mg/ml in dH_2O	10 µg/ml

2.1.4 Bacterial strains

E. coli strains used in this work are described below in Table 2-4. All transformations were performed using the manufactures guidelines.

Table 2-4 E. coli strains and genotypes.

Strain (Supplier)	Purpose	Genotype
XL 10-Gold® Ultracompetent (Stratagene)	DNA Manipulation	Tet^r Δ(*mcrA*)183 Δ(*mcrCB-hsdSMR-mrr*)*173 end*A1 *sup*E44 *thi*-1 *rec*A1 *gyr*A96 *rel*A1 *lac* Hte [F´ *pro*AB *lacI*q ZΔM15 *Tn*10(Tet^r) Amy Cam^r]
BL21-Gold(DE3) (Agilent)	Protein Expression	(B F^- *ompT hsdS*(r_B^- m_B^-) *dcm*$^+$ *Tet*^{r} *gal* λ(DE3) *end*A Hte
BL21-CodonPlus RIPL	Protein Expression	E. coli B F– ompT hsdS(rB – mB –) dcm+ Tetr gal endA Hte [argU ileY leuW Camr]
BL21-Gold(DE3)-pLysS	Autoinduction Protein Expression	B F- dcm+ Hte ompT hsdS(r_B- m_B-) gal λ (DE3) [pLysS $Cam_{r]}^a$ endA Tet^r

2.1.5 Vectors, genes and primers

Primers were synthesised by Intergrated DNA Technologies, Inc. Codon optimised selina-4(15),7(11)-diene was synthesized by GenScript into a pET-28a vector. pET His6 Sumo TEV LIC cloning vector (1S) was a gift from Scott Gradia (Addgene plasmid # 29659; http://n2t.net/addgene:29659; RRID:Addgene 29659). pET-TEV was a kind gift from Professor Radford (Faculty of Biological Science, University of Leeds, UK).

2.1.6 Production of competent cells

A 10mL culture of the desired strain was grown at 37°C shaking at 200 rpm overnight. 100 mL of LB was inoculated with a 1:20 ratio using the 10 mL overnight culture. Cultures were grown at 37°C shaking at 200 rpm until the OD_{600nm} is between 0.40 and 0.45. At this point

the cells were centrifuged at 4000 rpm for 10 minutes at 4°C with cold rotors. The supernatant was discarded, and the pellet resuspended in 10 mL of sterile chilled 100 mM calcium chloride. This was incubated on ice for 10 minutes before being centrifuged as before. The supernatant was discarded, and the pellet was reuspended in 2 mL of 100 mM calcium chloride, 30% (v/v) glycerol. The cells were aliquoted into 100 µL aliquots on eppendorfs on dry ice. The cells were stored at -80°C.

2.2 DNA methods

2.2.1 Primers

Primers were designed for a variety of purposes using the NEB T_m calculator and sometimes the QuikChange primer design online tool to design primers for mutagenesis. Primers are shown in the Table 2-5.

Table 2-5 The primers used in DNA cloning and mutagenesis.

Primer Function	Primer Name	Direction	Primer Sequence (5' to 3')
Addition of stop codon to pET His6 Sumo TEV LIC cloning vector (1S)	SeDS_1	Fwd	ggtggtggtggtggtggatgattgcggccgcaagcttg
	SeDS_2	Rev	caagcttgcggccgcaatcatccaccaccaccaccacc
Mutate F55 to E	SeDS_F55E_Fwd	Fwd	ggcagaatacgcgcgctctcggtaccgatgtcttggg
	SeDS_F55E_Rev	Rev	cccaagacatcggtaccgagagcgcgcgtattctgcc
Mutate F55 to H	SeDS_F55H_Fwd	Fwd	gcagaatacgcgcgctatgggtaccgatgtcttggg
	SeDS_F55H_Rev	Rev	cccaagacatcggtacccatagcgcgcgtattctgc
Mutate F55 to I	SeDS_F55I_Fwd	Fwd	agaatacgcgcgctaatggtaccgatgtcttgg
	SeDS_F55I_Rev	Rev	ccaagacatcggtaccattagcgcgcgtattct
Mutate F55 to L	F55L Fwd	Fwd	cccaagacatcggtaccttaagcgcgcgta
	F55L Rev	Rev	tacgcgcgcttaaggtaccgatgtcttggg
Mutate F55 to M	SeDS_F55M_Fwd	Fwd	gcagaatacgcgcgctcatggtaccgatgtcttgg
	SeDS_F55M_Rev	Rev	ccaagacatcggtaccatgagcgcgcgtattctgc
Mutate F55 to Q	SeDS_F55Q_Fwd	Fwd	ggcagaatacgcgcgctctgggtaccgatgtcttggg
	SeDS_F55Q_Rev	Rev	cccaagacatcggtacccagagcgcgcgtattctgcc
Mutate F55 to Y	F55Y Fwd	Fwd	caagacatcggtacctatagcgcgcgtattctg
	F55Y Rev	Rev	cagaatacgcgcgctataggtaccgatgtcttg
Mutate L78 to C	SeDS_L78C_Fwd	Fwd	gccatcgtccacaccaaagcaccacaggatgaaatccgc
	SeDS_L78C_Rev	Rev	gcggatttcatcctgtggtgctttggtgtggacgatggc
Mutate L78 to E	SeDS_L78E_Fwd	Fwd	catcgtccacaccaaactcccacaggatgaaatccg
	SeDS_L78E_Rev	Rev	cggatttcatcctgtgggagtttggtgtggacgatg

Mutate L78 to G	SeDS_L78G_Fwd	Fwd	catcgtccacaccaaaccccacaggatgaaatccg
	SeDS_L78G_Rev	Rev	cggatttcatcctgtgggggtttggtgtggacgatg
Mutate L78 to I	SeDS_L78I_Fwd	Fwd	gccatcgtccacaccaaatatccacaggatgaaatccgc
	SeDS_L78I_Rev	Rev	gcggatttcatcctgtggatatttggtgtggacgatggc
Mutate L78 to Q	SeDS_L78Q_Fwd	Fwd	catcgtccacaccaaactgccacaggatgaaatcc
	SeDS_L78Q_Rev	Rev	ggatttcatcctgtggcagtttggtgtggacgatg
Mutate L78 to T	SeDS_L78T_Fwd	Fwd	ccatcgtccacaccaaacgtccacaggatgaaatccgc
	SeDS_L78T_Rev	Rev	gcggatttcatcctgtggacgtttggtgtggacgatgg
Mutate L78 to W	SeDS_L78W_Fwd	Fwd	ccatcgtccacaccaaaccaccacaggatgaaatccgc
	SeDS_L78W_Rev	Rev	gcggatttcatcctgtggtggtttggtgtggacgatgg
Mutate F79 to E	SeDS_F79E_Fwd	Fwd	gtggccatcgtccacaccctccagccacaggatgaaatc
	SeDS_F79E_Rev	Rev	gatttcatcctgtggctggagggtgtggacgatggccac
Mutate F79 to H	SeDS_F79H_Fwd	Fwd	ggccatcgtccacaccatgcagccacaggatgaaat
	SeDS_F79H_Rev	Rev	atttcatcctgtggctgcatggtgtggacgatggcc
Mutate F79 to I	SeDS_F79I_Fwd	Fwd	ccatcgtccacaccaatcagccacaggatgaaa
	SeDS_F79I_Rev	Rev	tttcatcctgtggctgattggtgtggacgatgg
Mutate F79 to M	SeDS_F79M_Fwd	Fwd	ggccatcgtccacacccatcagccacaggatgaaa
	SeDS_F79M_Rev	Rev	tttcatcctgtggctgatgggtgtggacgatggcc
Mutate F79 to Q	SeDS_F79Q_Fwd	Fwd	gtggccatcgtccacaccctgcagccacaggatgaaatc
	SeDS_F79Q_Rev	Rev	gatttcatcctgtggctgcagggtgtggacgatggccac
Mutate Y152 to F	SeDS_Y152F_Fwd	Fwd	aaacaacgctaaagaagaactcacgcagcgcatc
	SeDS_Y152F_Rev	Rev	gatgcgctgcgtgagttcttctttagcgttgttt
Mutate Y152 to G	SeDS_Y152G_Fwd	Fwd	ggttgatgcgctgcgtgagggcttctttagcgttgtttgg

	SeDS_Y152G_Rev	Rev	ccaaacaacgctaaagaagccctcacgcagcgcatcaacc
Mutate Y152 to H	SeDS_Y152H_Fwd	Fwd	aacaacgctaaagaagtgctcacgcagcgcatcaa
	SeDS_Y152H_Rev	Rev	ttgatgcgctgcgtgagcacttctttagcgttgtt
Mutate Y152 to I	SeDS_Y152I_Fwd	Fwd	ccaaacaacgctaaagaagatctcacgcagcgcatcaacc
	SeDS_Y152I_Rev	Rev	ggttgatgcgctgcgtgagatcttctttagcgttgtttgg
Mutate Y152 to L	SeDS_Y152L_Fwd	Fwd	gttgatgcgctgcgtgagttattctttagcgttgtttggg
	SeDS_Y152L_Rev	Rev	cccaaacaacgctaaagaataactcacgcagcgcatcaac
Mutate Y152 to M	SeDS_Y152M_Fwd	Fwd	ggttgatgcgctgcgtgagatgttctttagcgttgtttggg
	SeDS_Y152M_Rev	Rev	cccaaacaacgctaaagaacatctcacgcagcgcatcaacc
Mutate Y152 to S	SeDS_Y152S_Fwd	Fwd	ccaaacaacgctaaagaagctctcacgcagcgcatcaacc
	SeDS_Y152S_Rev	Rev	ggttgatgcgctgcgtgagagcttctttagcgttgtttgg
Mutate Y152 to T	SeDS_Y125T_Fwd	Fwd	ccaaacaacgctaaagaaggtctcacgcagcgcatcaacc
	SeDS_Y152T_Rev	Rev	ggttgatgcgctgcgtgagaccttctttagcgttgtttgg
Mutate Y152 to W	SeDS_Y152W_Fwd	Fwd	cccaaacaacgctaaagaaccactcacgcagcgcatcaac
	SeDS_Y152W_Rev	Rev	gttgatgcgctgcgtgagtggttctttagcgttgtttggg
Mutate Y152 to V	SeDS_Y152V_Fwd	Fwd	ggttgatgcgctgcgtgaggtcttctttagcgttgtttgg
	SeDS_Y152V_Rev	Rev	ccaaacaacgctaaagaagacctcacgcagcgcatcaacc
Mutate T184 to C	SeDS_T184C_Fwd	Fwd	gcagaaccacgctgcacgcgccgtcataca
	SeDS_T184C_Rev	Rev	tgtatgacggcgcgtgcagcgtggttctgc
Mutate T184 to D	SeDS_T184D_Fwd	Fwd	ggcagaaccacgctgtccgcgccgtcatacag
	SeDS_T184D_Rev	Rev	ctgtatgacggcgcggacagcgtggttctgcc
Mutate T184 to L	SeDS_T184L_Fwd	Fwd	atcggcagaaccacgcttagcgcgccgtcatacagac
	SeDS_T184L_Rev	Rev	gtctgtatgacggcgcgctaagcgtggttctgccgat

Mutate T184 to N	SeDS_T184N_Fwd	Fwd	gcagaaccacgctgttcgcgccgtcatac
	SeDS_T184N_Rev	Rev	gtatgacggcgcgaacagcgtggttctgc
Mutate T184 to S	SeDS_T184S_Fwd	Fwd	cagaaccacgctgctcgcgccgtcata
	SeDS_T184S_Rev	Rev	tatgacggcgcgagcagcgtggttctg
Mutate T184 to V	SeDS_T184V_Fwd	Fwd	ggcagaaccacgctgaccgcgccgtcatacag
	SeDS_T184V_Rev	Rev	ctgtatgacggcgcggtcagcgtggttctgcc
Mutate T184 to Y	SeDS_T184Y_Fwd	Fwd	atcggcagaaccacgctatacgcgccgtcatacagac
	SeDS_T184Y_Rev	Rev	gtctgtatgacggcgcgtatagcgtggttctgccgat
Mutate V187 to D	SeDS_V187D_Fwd	Fwd	cagcatcggcagatccacgctggtcgc
	SeDS_V187D_Rev	Rev	gcgaccagcgtggatctgccgatgctg
Mutate V187 to F	SeDS_V187F_Fwd	Fwd	agcatcggcagaaacacgctggtcgcg
	SeDS_V187F_Rev	Rev	cgcgaccagcgtgtttctgccgatgct
Mutate V187 to G	SeDS_V187G_Fwd	Fwd	cagcatcggcagacccacgctggtcgc
	SeDS_V187G_Rev	Rev	gcgaccagcgtgggtctgccgatgctg
Mutate V187 to I	SeDS_V187I_Fwd	Fwd	cgcgaccagcgtgattctgccgatgct
	SeDS_V187I_Rev	Rev	agcatcggcagaatcacgctggtcgcg
Mutate V187 to L	SeDS_V187L_Fwd	Fwd	gcatcggcagaagcacgctggtcgc
	SeDS_V187L_Rev	Rev	gcgaccagcgtgcttctgccgatgc
Mutate V187 to M	SeDS_V187M_Fwd	Fwd	cgcgaccagcgtgatgctgccgatgctgg
	SeDS_V187M_Rev	Rev	ccagcatcggcagcatcacgctggtcgcg
Mutate V187 to N	SeDS_V187N_Fwd	Fwd	ccagcatcggcagattcacgctggtcgcgc
	SeDS_V187N_Rev	Rev	gcgcgaccagcgtgaatctgccgatgctgg
Mutate V187 to S	SeDS_V187S_Fwd	Fwd	gcgcgaccagcgtgagtctgccgatgctgg

	SeDS_V187S_Rev	Rev	ccagcatcggcagactcacgctggtcgcgc
Mutate V187 to T	SeDS_V187T_Fwd	Fwd	ccagcatcggcagagtcacgctggtcgcgc
	SeDS_V187T_Rev	Rev	gcgcgaccagcgtgactctgccgatgctgg
Mutate V187 to T repeat	SeDS_V187T2_Fwd	Fwd	gcgcgaccagcgtgactctgccgatgctgg
	SeDS_V187T2_Rev	Rev	ccagcatcggcagagtcacgctggtcgcgc
Mutate V187 to Y	SeDS_V187Y_Fwd	Fwd	gcgcgaccagcgtgtatctgccgatgctgg
	SeDS_V187Y_Rev	Rev	ccagcatcggcagatacacgctggtcgcgc
Mutate S217 to C	SeDS_S217C_Fwd	Fwd	caggtaatgatgaagcacgccatctccgcaac
	SeDS_S217C_Rev	Rev	gttgcggagatggcgtgcttcatcattacctg
Mutate S217 to D	SeDS_S217D_Fwd	Fwd	cccaggtaatgatgaagtccgccatctccgcaaccg
	SeDS_S217D_Rev	Rev	cggttgcggagatggcggacttcatcattacctggg
Mutate S217 to L	SeDS_S217L_Fwd	Fwd	gtcccaggtaatgatgaatagcgccatctccgcaaccgc
	SeDS_S217L_Rev	Rev	gcggttgcggagatggcgctattcatcattacctgggac
Mutate S217 to N	SeDS_S217N_Fwd	Fwd	ccaggtaatgatgaagttcgccatctccgcaac
	SeDS_S217N_Rev	Rev	gttgcggagatggcgaacttcatcattacctgg
Mutate S217 to T	SeDS_S217T_Fwd	Fwd	ccaggtaatgatgaaggtcgccatctccgcaa
	SeDS_S217T_Rev	Rev	ttgcggagatggcgaccttcatcattacctgg
Mutate S217 to V	SeDS_S127V_Fwd	Fwd	cccaggtaatgatgaagaccgccatctccgcaaccg
	SeDS_S217V_Rev	Rev	cggttgcggagatggcggtcttcatcattacctggg
Mutate S217 to Y	SeDS_S217Y_Fwd	Fwd	gtcccaggtaatgatgaaatacgccatctccgcaaccgc
	SeDS_S217Y_Rev	Rev	gcggttgcggagatggcgtatttcatcattacctgggac
Mutate I220 to E	SeDS_I220E_Fwd	Fwd	aagatatcgttgtcccaggtctcgatgaagctcgccatctccg
	SeDS_I220E_Rev	Rev	cggagatggcgagcttcatcgagacctgggacaacgatatctt

Mutate I220 to F	SeDS_I220F_Fwd	Fwd	tcgttgtcccaggtaaagatgaagctcgccatc
	SeDS_I220F_Rev	Rev	gatggcgagcttcatctttacctgggacaacga
Mutate I220 to L	SeDS_I220L_Fwd	Fwd	tcgttgtcccaggtaaggatgaagctcgccatc
	SeDS_I220L_Rev	Rev	gatggcgagcttcatccttacctgggacaacga
Mutate I220 to M	SeDS_I220M_Fwd	Fwd	ttgtcccaggtcatgatgaagctcgccatctcc
	SeDS_I220M_Rev	Rev	ggagatggcgagcttcatcatgacctgggacaa
Mutate I220 to Q	SeDS_I220Q_Fwd	Fwd	aagatatcgttgtcccaggtctggatgaagctcgccatctccg
	SeDS_I220Q_Rev	Rev	cggagatggcgagcttcatccagacctgggacaacgatatctt
Mutate I220 to T	SeDS_I220T_Fwd	Fwd	tatcgttgtcccaggtagtgatgaagctcgccatc
	SeDS_I220T_Rev	Rev	gatggcgagcttcatcactacctgggacaacgata
Mutate I220 to V	SeDS_I220V_Fwd	Fwd	tcgttgtcccaggtaacgatgaagctcgccatc
	SeDS_I220V_Rev	Rev	gatggcgagcttcatcgttacctgggacaacga
Mutate T221 to C	SeDS_T221C_Fwd	Fwd	tcgttgtcccagcaaatgatgaagctcgccatctcc
	SeDS_T221C_Rev	Rev	ggagatggcgagcttcatcatttgctgggacaacga
Mutate T211 to D	SeDS_T221D_Fwd	Fwd	atcgttgtcccagtcaatgatgaagctcgccatctccgc
	SeDS_T211D_Rev	Rev	gcggagatggcgagcttcatcattgactgggacaacgat
Mutate T221 to L	SeDS_T221L_Fwd	Fwd	gatatcgttgtcccatagaatgatgaagctcgccatctccgcaacc
	SeDS_T221L_Rev	Rev	ggttgcggagatggcgagcttcatcattctatgggacaacgatatc
Mutate T221 to N	SeDS_T221N_Fwd	Fwd	tcgttgtcccagttaatgatgaagctcgccatctc
	SeDS_T221N_Rev	Rev	gagatggcgagcttcatcattaactgggacaacga
Mutate T221 to S	SeDS_T221S_Fwd	Fwd	cgttgtcccagctaatgatgaagctcgccatc
	SeDS_T211S_Rev	Rev	gatggcgagcttcatcattagctgggacaacg
Mutate T221 to V	SeDS_T221V_Fwd	Fwd	atcgttgtcccagacaatgatgaagctcgccatctccgc

	SeDS_T221V_Rev	Rev	gcggagatggcgagcttcatcattgtctgggacaacgat
Mutate T221 to Y	SeDS_T221Y_Fwd	Fwd	gatatcgttgtcccaataaatgatgaagctcgccatctccgcaacc
	SeDS_T221Y_Rev	Rev	gatatcgttgtcccaataaatgatgaagctcgccatctccgcaacc

2.2.2 Polymerase chain reaction (PCR)

Both Q5® High fidelity polymerase and *Pfu* DNA polymerase were used throughout this work. Both were used following the manufactures guidelines.

2.2.3 PCR Purification

Purification of PCR products was completed using QIAquick PCR Purification kit as described by the supplier's guidelines.

2.2.4 Agarose gel electrophoresis

0.6-1% agarose gel were produced by weighing out accurately the agarose and dissolving the agarose in 1X TAE buffer (40 mM Tris-acetate, 1 mM EDTA pH 8.0) using the microwave. After the solution had cooled 10,000X SYBRsafe DNA stain was added and the solution was poured into a cast. When the gel had set the gel was placed in that tank and covered in 1X TAE buffer for sample insertion.

Samples were prepared using 6x Gel loading dye purple (NEB). Either 1kb Quick load extend (NEB) or 1kb ladder (Promega). After the gel was ran at 90 volts to 100 volts for 30 to 1 hour and 30 minutes. Following running the gel the gel was imaged under UV light.

2.2.5 DNA gel extraction

Bands were excised from the gel under UV light by a scalpel. QIAquick Gel Extraction Kit (Qiagen) was used a described in the suppliers' guidelines.

2.2.6 Restriction digestion

Restriction digests were carried out as described in the supplier's guidelines. 3-10 µL of DNA was added and 10-20 µL of PCR products were added to 30-50 µL restriction digests for an approximate concentration between 10-20 ng/µL. The enzymes used were; Hind III (1 µL – 20 units of enzyme), Eco RI (3 µL – 60 units of enzyme) and XhoI (2 µL - 40 units of enzyme). The volume was made up to 30-50 µL (depending on the reaction) with nuclease free water. If the digestion products were to be used further the digestion was terminated by incubating at 80 °C for 20 minutes.

2.2.7 Transformation and DNA plasmid purification

Cloning and mutagenesis products were transformed into *Escherichia coli (E. coli)* XL-10 Gold UltraCompetent cells (Aligent Technologies) or *E. coli* BL21 DE3 (-) using the supplier's guidelines.

DNA was purified using either the Wizard® Plus SV Minipreps DNA Purification Kit (Promega) or the Monarch® Plasmid Miniprep Kit (NEB) from cultures. Using a 200 µL tip a colony was picked from transformation products and used to inoculate 5 mL of LB media with the corresponding antibiotic which was incubated at 37°C shaking at 250 rpm overnight. Cells were harvested by centrifugation at 13,000 x g for 5 minutes. Following this the kits protocol was followed and the product DNA was stored at -20°C.

2.2.8 Measuring DNA concentration and sequencing

DNA concentration was measured using a Nanodrop spectrophotometer (Thermo-Scientific). 1 µL of DNA was used after cleaning with deionised water and blanking the spectrophotometer. The concentration is measured using A_{260nm} reading and using Beers law to calculate given in ng/µl.

DNA was sequenced by, Beckman Coulter Genomics with the primers in Table 5 or universal primers. DNA (15 µl at 100 ng/µl) and primers (10 µM) were sent for the express single sequence service and the results was analysed using BioEdit Sequence Alignment Editor.

2.3 Protein expression and purification

2.3.1 Protein expression of selina-4(15),7(11)-diene synthase for purification

One colony from a transformation into BL21-DE3 cells or a glycerol stock was used to inoculate 5 mL of 2x TY media which was incubated for 6-8 hours shaking at 200 rpm at 37°C. 1 mL of this culture was used to inoculate 50 mL of 2x TY media which was incubated overnight shaking at 200 rpm at 37°C. 10 mL of this overnight culture was used to inoculate 1L of 2x TY media. 500 mL cultures were incubated at 37°C until reaching OD_{600nm} 0.4 when the temperature was lowered to 20°C. When the OD_{600nm} reached 0.6-0.8 the cultures were induced with 1 mM IPTG and the temperature was lowered further to 18°C and incubated overnight. 1 mL samples were taken before induction and before harvesting and stored at -20°C. Cells were harvested at 10°C by centrifuging at 8,000 x g for 20 minutes, the pellet was stored at -80°C.

2.3.2 Protein expression of selina-4(15),7(11)-diene synthase for GC/MS

One colony from a transformation into BL21-DE3 cells or a glycerol stock was used to inoculate 5 mL of 2x TY media which is incubated overnight shaking at 200 rpm at 37°C. 1 mL of this culture was used to inoculate 50 mL of 2x TY media which was incubated shaking at 200 rpm at 37°C. 50 mL cultures were incubated until OD_{600nm} reached 0.6-0.8. At this point the temperature was lowered to 25°C, the culture induced with 1 mM IPTG and the

culture layer with 2.5 mL of dodecane. The induced culture was then incubated shaking at 200 rpm at 25°C for over 72 hours.

2.3.3 Protein expression of TEV protease

One colony from a transformation into BL21-CodonPlus RIPL cells or a glycerol stock was used to inoculate 50 mL of LB media which is incubated overnight shaking at 200rpm at 37°C. 10 mL of overnight culture was used to inoculate 1L of LB media which was then incubated shaking at 200rpm at 37°C. When OD_{600nm} reached 0.6 when the cultures were induced with 1 mM IPTG. Expression was allowed for 4 hours at 30°C before cells were pelleted at 8,000 x g for 20 minutes and the pellet was stored at -80°C.

2.3.4 Purification of selina-4(15),7(11)-diene synthase

The cell pellet was thawed and resuspended in lysis buffer (20 mM Tris-HCl pH 8.0, 500 mM NaCl, 10 mM $MgCl_2$, 25 mM imidazole, 10% glycerol, 0.1 mg/ml lysozyme, 0.05 mg/ml DNase, EDTA-free protease inhibitors). The pellet was homogenised using the homogeniser before disrupting the cells using the cell disruptor at 30 kpsi. After cell disruption cell lystate centrifuged at 50,000 *g* for 30 minutes at 10°C. Supernatant was applied to HisTrap FF crude (GE Healthcare, USA) that was pre-equilibrated with wash buffer (20 mM Tris-HCl pH 8.0, 500 mM NaCl, 10 mM $MgCl_2$, 25 mM imidazole, 10% glycerol). After the supernatant was applied the column was washed with 10 column volumes of wash buffer before elution of selina-4(15),7(11)-diene synthase with 2 volumes of elution buffer (20 mM Tris-HCl pH 8.0, 500 mM NaCl, 10 mM $MgCl_2$, 400 mM imidazole, 10% glycerol). Samples of each stage were taken for analysis on sodium-dodecayl sulphate polyacrylamide gel electrophoresis (SDS-PAGE gel). This His column elution was pooled and cleaved by 0.1 mg/ml TEV protease in 5L dialysis of TEV cleveage buffer (20 mM Tris-HCl pH 7.5, 100 mM NaCl, 10 mM $MgCl_2$, 10% glycerol, 14mM β-mercaptoethanol) overnight at 4°C. Following TEV Cleavage overnight sample concentrated for application onto size exclusion column. Hiload 16/600 Superdex 75pg (GE Healthcare, USA) was equilibrated with size exclusion (20 mM Tris-HCl pH 7.5, 100 mM NaCl, 10 mM $MgCl_2$, 10% glycerol, 5mM DTT) before the sample is applied. After a void volume of 40 mls, fraction collecting began collecting 1.8 ml fractions. Fractions were analysed on an SDS-PAGE. The fractions containing SEDS were pooled and dialysed three times into 5L of SEDS storage buffer (50 mM Tris-HCl pH 7.5, 100 mM NaCl, 10 mM $MgCl_2$, 10% glycerol) before snap freezing in liquid nitrogen to be stored at -80°C.

2.3.5 Protein purification of selina-4(15),7(11)-diene synthase mutants for activity screening

The cell pellet was thawed and resuspended in lysis buffer (20 mM Tris-HCl pH 8.0, 500 mM NaCl, 10 mM $MgCl_2$, 25 mM imidazole, 10% glycerol, 0.1 mg/ml lysozyme, 0.05 mg/ml DNase, EDTA-free protease inhibitors). The pellet was homogenised using the homogeniser before disrupting the cells using the cell disruptor at 30 kpsi. After cell disruption cell lystate centrifuged at 50,000 *g* for 30 minutes at 10°C. Supernatant was applied to nickel resin that was pre-equilibrated with wash buffer (20 mM Tris-HCl pH 8.0, 500 mM NaCl, 10 mM $MgCl_2$, 25 mM imidazole, 10% glycerol). After the supernatant was applied the resin was washed with 50 ml of wash buffer. The SEDS mutants were eluted with 5 ml of elution buffer (20 mM Tris-HCl pH 8.0, 500 mM NaCl, 10 mM $MgCl_2$, 400 mM imidazole, 10% glycerol). Before testing the elution was cleaved with 0.1 mg/ml TEV protease in 5L dialysis of TEV cleveage buffer (20 mM Tris-HCl pH 7.5, 100 mM NaCl, 10 mM $MgCl_2$, 10% glycerol, 14mM β-mercaptoethanol) overnight at 4°C.

2.3.6 Protein purification of TEV protease

Cell pellet thawed and resuspeneded in TEV lysis buffer (25 mM sodium phosphate pH 8.0, 200 mM NaCl, 10% glycerol, 25 mM imidazole, PMSF). Pellet was homogenised using the homogeniser before cell disrupting at 30 kpsi. After cell disruption cell lystate centrifuged at 20,000 *g* for 30 minutes at 10°C. Supernatant was filtered with 22 µM filter before application to HisTrap FF crude (GE Healthcare, USA) that was pre-equilibrated with TEV wash buffer (20 mM sodium phosphate pH 8.0, 200 mM NaCl, 25 mM imidazole, 10% glycerol). After the supernatant was applied the column was washed with 10 column volumes of wash buffer before elution with 2 column volumes of TEV elution buffer (20 mM sodium phosphate pH 8.0, 200 mM NaCl, 500 mM imidazole, 10% glycerol). The elution sample was concentrated before application on to a pre-equlibrated SEC column with TEV SEC buffer (25 mM sodium phosphate pH 8.0, 200 mM NaCl, 13 mM β-mercaptoethanol, 10% glycerol). The SEC column used is a Hiload Superdex 75pg 26/60. Column ran at 1 ml/min and 1.8ml fraction began collecting after the column void volume. Samples were collected at all stages of the purification for analysis by SDS-PAGE. Fractions containing TEV protease were pooled before snap freezing in liquid nitrogen to be stored at -80°C.

2.3.7 SDS-PAGE

All samples are diluted appropriately with cell culture samples diluted using their OD_{600nm} values for dilution factors. The insoluble fraction resuspended in 500 µL using the sonicator and was diluted by a factor of 10 along with the soluble fraction. The cell lysate is diluted by

a factor of 20. All diluted samples were diluted with 1x PBS (137 mM NaCl, 2.7 mM KCl, 10 mM Na_2HPO_4 and KH_2PO_4). All wash and elution samples were not diluted. 10 µL of each sample is mixed with 10 µL of 2x sample loading buffer (200 mM Tris-HCl, pH 6.8, 20% (v/v) glycerol, 8% (w/v) SDS, 0.4% (w/v) bromophenol blue and 20% (v/v) β-mercaptoethanol). The sample were incubated at 95°C for 10 minutes and then spun down 13,000 x g for 3 minutes. SDS PAGE gels were produced as described in Table 2-6. Samples alongside a molecular weight ladder were run on gel in SDS-PAGE running buffer (25 mM Tris, 250 mM glycine, 0.1% (w/v) SDS, pH 8.3). Gels were run at 100 volts for 15 minutes until a clear line of dye is seen across the gel and then the voltage was increase to 200 volts and run for 45-90 minutes. Gels were stained for 15 minutes using InstantBlue (Expedeon) rocking before being de-stained with deionised water. The gel was imaged and after could be sent for mass spectrometry.

Table 2-6 SDS-PAGE gel components

Component	12% Separation Gel	4% Stacking Gel
Deionised Water	3.2 mL	3.6 mL
1.5 M Tris-HCl pH 8.8	2.6 mL	-
1 M Tris-HCl pH 6.8	-	625 µL
10% SDS	100 µL	50 µL
Acrylamide Bisphosphate 30%	4 mL	650 µL
10% (w/v) Ammonium persulfate*	100 µL	50 µL
TEMED*	12.5 µL	12.5 µL

2.3.8 Protein concentration determination

Protein concentration was determined using a Nanodrop spectrophotometer (Thermo-Scientific. 2 µL of protein was used after cleaning with deionised water and blanking the spectrophotometer. The concentration was calculated using the other protein setting with the molar extinction coefficient and protein molecular weight.

2.3.9 Mass spectrometry

All mass spectrometry (MS) analysis was performed by Dr James Ault and Rachel George at the Astbury Centre for Structural and Molecular Biology, University of Leeds, UK.
Accurate molecular mass was determined by LC/MS. A 70 µl sample of greater than 20 µM was submitted for analysis. For protein identification through peptide fingerprinting a protein

sample was ran on a 12% SDS-PAGE gel and stained for 15 minutes with InstantBlue (Expendeon). The protein band was excised and digested with trypsin for 18 hours. LC-MS/MS was used to identify the peptides that were then screened against the UniProtKB/SwissProt databases and sequences of codon SeDS provided to the MS department.

2.4 Malachite green assay

2.4.1 Initial activity testing

The initial activity assays were completed with purified protein as described in 2.3.4. The components of the reaction are described in Table 2-7 and were all prepared before the assay. The malachite green components, highlighted in green in Table 2-7, were combined to from the malachite green colour reagent before initiating the assay. The farnesyl pyrophosphate and the assay buffer were added to a 96 well plate on ice in a determined order. The plate was incubated at 30 °C for up to 20 minutes with each time point initiated by the addition of the specific enzyme concentration. For each enzyme concentration the assay was completed in triplicate with a negative control without enzyme. Following the 20-minute incubation the plate was developed with the addition of 12 µL malachite green colour reagent. Before the absorption was measured at 620 nm the plate was incubated for 15 minutes at room temperature. All reactions were performed in clear round bottom, 96 well plates.

Table 2-7 Components of initial activity malachite green assays

Component	Stock Concentration	Final Concentration	Volume for reaction (µl)
Farnesyl Pyrophosphate	10 mM	100 µM and 200 µM	5 (of dilution)
Enzyme	2.5 µM	Concentrations ranged between 25 and 10000 nM	20
Assay Buffer	50 mM MES, 50 mM CAPS, 100 mM Tris, 10 mM $MgCl_2$, 60 mU Inorganic Pyrophosphatase	25 mM MES, 25 mM CAPS, 50 mM Tris, 5 mM $MgCl_2$ 30 mU Inorganic Pyrophosphatase	25
Malachite Green	1.83 mM (in 3M sulphuric acid)	1.44 mM	-
Ammonium Molybdate	60.87 mM	11.98 mM	-
Tween 20	11%	0.17%	-
Malachite Green Colour Reagent	1.44 mM Malachite green, 11.98 mM Ammonium molybdate, 0.17% Tween 20		12

2.4.2 Kinetic characterization

Kinetic characterisation of SeDS enzymes was completed with at set concentration of enzyme and farnesyl pyrophosphate concentrations ranging from 5 µM and 160 µM. Each kinetic assay was performed alongside a standard curve of sodium pyrophosphate concentrations to convert absorbance to pyrophosphate concentration. Reaction components are described in Table 2-8 and were set up on ice. The malachite green colour reagent was produced by mixing the appropriate amount of the malachite green and ammonium molybdate. For calibration curve reactions sodium pyrophosphate was aliquoted into a 96 well plate on ice and reactions were initiated at set time points by the addition of assay buffer. For assay reactions farnesyl pyrophosphate concentrations and assay buffer were aliquoted and the reactions were initiated with the addition of the SeDS enzyme and incubated at 30 °C. The reactions were initiated at the set points and incubated for the stated amount of time. Following the incubation, the reaction was quenched by the addition of 100 µL of the malachite green colour reagent. The plate was then incubated at room temperature for 30 minutes before the absorbance was measured at 620 nm.

Table 2-8 Components of kinetic assay reactions

Component	Stock Concentration	Assay Concentration	Volume for reaction (µl)	Method
Farnesyl Pyrophosphate	10mM	5µM and 160µM	5	Added to plate before assay initiated
SeDS	20µM	500nM	20	Added to each reaction at the time point.
Sodium pyrophosphate	10mM	1.25µM-15µM	25	Positive control and creates standard curve.
Assay Buffer	500mM Tris, 10mM $MgCl_2$, 60mU Inorganic Pyrophosphatase	250mM Tris, 5mM $MgCl_2$, 30mU Inorganic Pyrophosphatase	25	Added to reaction wells, before assay initiated but added at each time point for negative and positive control wells.
Malachite green	0.12% (in 1M HCl)	0.09%	-	Component of malachite green colour reagent
Ammonium molybdate	4.2% (in 4.5M HCl)	1.05%	-	Component of malachite green colour reagent. Solution filtered before use.
Malachite green colour reagent	0.09% Malachite green (in 1M HCl), 1.05% (in 4.5M HCl)	-	100	Components mixed fresh and filtered before use.

2.5 Product Analysis

2.5.1 GC/MS

GC/MS samples are diluted in ethyl acetate spiked with 15 µg/ml trans-carophyllene. 250 ml of this diluted sample is added to a glass vial for analysis.

The dodecane layer (2.5.2) was sampled after 72 hours using a glass syringe. This sample was centrifuged at 13,000 x g for 1 minutes. 100 µL of the dodecane layer was added to 100 µL cumene-spiked ethyl acetate (80 µg/mL cumene to make final concentration 40 µg/mL).

A 10 mg/mL stock solution of bisabolene was produced using 10 µL of bisabolene (density 0.89 g/mL) with 899 µL of ethyl acetate. The stock was diluted by a factor of 100 with ethyl acetate making the first concentration 100 µg/mL. This solution was used to create successive dilutions by a factor of 2. Samples were analysed using the program EMI. GC/MS was carried using a HP-5MS column (30 m length x 0.32 mm diameter x 0.25 μm film thickness) (ThermoTrace Ultra). Parameters were inlet pressure, 77.1 kPa, He 23.3 mL/min, injection volume, 2 µL, transfer line, 300 °C, and electron energy 70 eV. The GC was programmed as follows: 5 min at 50 °C increasing at 10 °C/min to 320 °C and operated in splitless mode. The carrier gas was He at 1 mL/min and a HP-5MS column fused silica capillary column was used. Instrumental parameters were (1) inlet pressure, 77.1 kPa, He 23.3 mL/min, (2) injection volume, 2 µL, (3) transfer line, 300 °C, and (4) electron energy 70 eV. The GC was programmed as follows: 5 min at 50 °C increasing at 10 °C/min to 250 °C and operated in splitless mode. The carrier gas was He at 1 mL/min (Baer, Rabe, Citron, *et al.*, 2014).

Chapter 3 – Selection, Expression and Characterisation of a Sesquiterpene Synthase

The aim of this work was to select a sesquiterpene synthase for expression and characterisation. The sesquiterpene synthase will then be used in engineering or mutagenesis experiments with the aim of producing a novel, previously uncharacterised terpene (See section 4.1). In order to determine whether our engineering attempts have been successful, kinetic characterisation of the wild-type and variant enzymes will be carried out, but this will require sufficient quantities of purified enzyme to be generated and so, the chosen sesquiterpene synthase must be able to be expressed recombinantly in *E. coli* and be purified using affinity chromatography to produce enough enzyme for kinetic studies. We will also need to quickly screen for active variants, and particularly those with altered product profiles. Since it is possible to quickly extract the terpene products from bacterial cultures by growing samples overlaid with organic solvents, it should be possible to us GC/MS of the solvent extract to quickly and easily assess the level and identity of the enzyme products. This chapter will describe the choice of system and the initial expression and characterisation.

3.1 Terpene synthase selection

Terpenes consist of a broad range of compounds composed of C5 units (Gao, Honzatko and Peters, 2012). For this work, sesquiterpenes, also known as C15 terpenes, were selected as the target length terpene for engineering. Sesquiterpenes are produced from farnesyl pyrophosphate with over 400 characterised sesquiterpene products (Yamada *et al.*, 2015). Monoterpenes were not selected due to their small size, limiting the potential chemistry available and the number of scaffolds available. Larger lengths were not selected because there would be a large potential chemistry available which could be overwhelming when using targeted mutagenesis (Figure 1-2).

Criteria for selecting the synthase to engineer included ability to express heterologously in *E. coli*, availability of a crystal structure to aid with targeted engineering and an enzyme that has not been extensively engineered previously. Using an enzyme that has not previously been extensively engineered will expand the field of terpene synthase engineering and produce novel insights. Four sesquiterpene synthases were shortlisted; α-bisabolene synthase, hedycaryol synthase, germacradienol synthase and selina-4(15),7(11)-diene synthase (McAndrew *et al.*, 2011) (Baer, Rabe, Citron, *et al.*, 2014) (Jiang, He and Cane,

2006) (Baer et al., 2014) (Figure 3-1). This list was further shortened by comparing the structures of the major products synthesised by each enzyme (Figure 3-1).

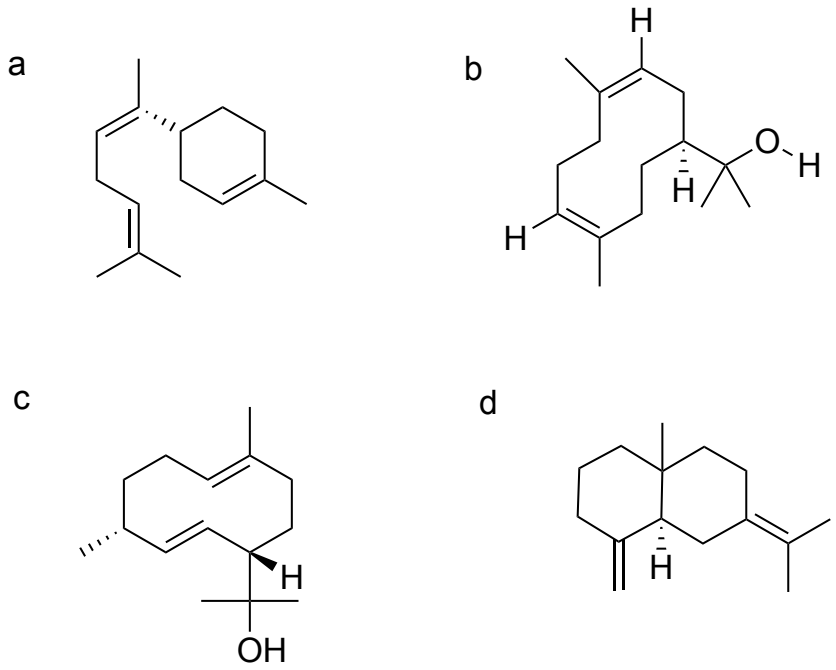


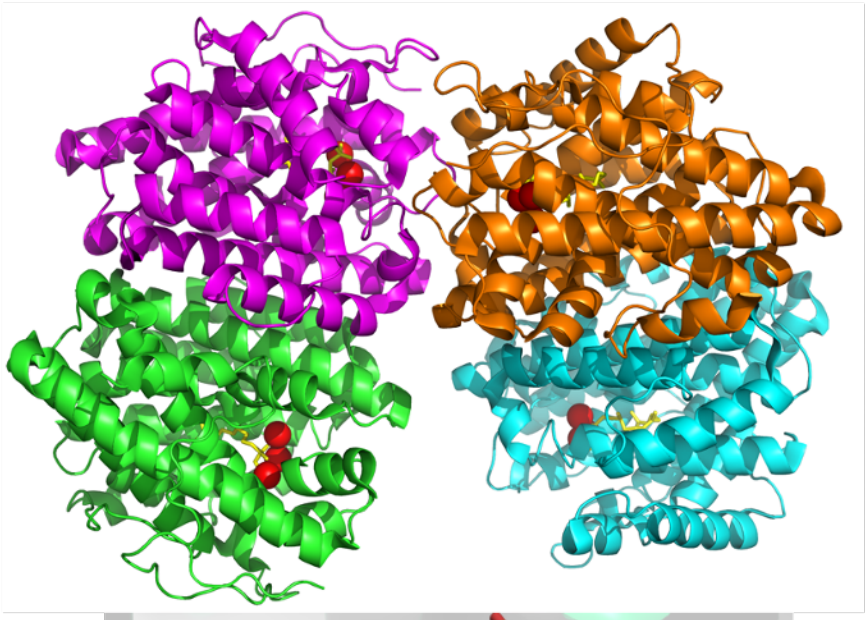
Figure 3-1 Structures of the major products of the four shortlisted sesquiterpene synthases. *(a) Bisabolene produced by α-bisabolene synthase. (b) Hedycaryol produced by hedycaryol synthase. (c) Germacradienol produced by germacradienol synthase. (d) Selina-4(15),7(11)-diene produced by selinadiene synthase*

Both hedycaryol and germacradienol (Figure 3-1 B and C) have one medium-sized ring structure compared to bisabolene and selina-4(15),7(11)-diene which both have two distinct component scaffolds either two rings or linear chains. Due to the extended chemistry potential with a double ring structure, α-bisabolene synthase and selina-4(15),7(11)-diene synthase were both considered further. Both these products lack oxygen and as part of our engineering we considered whether introduction of more space in the active site might allow quenching of intermediate biosynthetic carbocation by water. This could lead to substituted products that might be easier to detect if the natural product was non-hydroxylated and produced by elimination to an alkene. This potential has been shown with δ-cadiene synthase where a L405S variant produces germacrene D-4-ol. The L405S mutant targeted the G helix allowing a water molecule in the active site (Yoshikuni *et al.*, 2006).

3.1.1 Selina-4(15),7(11)-diene synthase

Selina-4(15),7(11)-diene synthase (SeDS) from *Streptomyces pristinaespiralis* ATCC 25486 is a tetrameric enzyme with each monomer consisting of one class I active site (Figure 3-2 to 3-4) (see section 1.4.1). SeDS produces selina-4(15),7(11)-diene as over 90% of its product profile (major product) and a germacrene B as a minor product (Figure 3-5) (Baer *et al.*, 2014). Selina-4(15),7(11)-diene synthase has been crystalized in the apo form with $(Mg^{2+})_3$ and diphosphate or with 2,3-dihydrofarnesyl pyrophosphate and $(Mg^{2+})_3$ (Figure 3-2) (Baer et al., 2014). Each monomer is an α domain with 11 anti-parallel α-helices connected by short linker sequences. The active site consists of helices B, C, H and K with linker helix G. The DDXXD motif is on helix C with the NSE triad on helix H (see section 1.4.1). The NSE triad coordinates the magnesium ion at the top of the active site and hydrophobic residues structure the active site cleft (Figure 3-3). Helices C and K are the helices rearranged on substrate binding in an induced fit reaction mechanism (Baer et al., 2014). The G linker helix has two distinct states shown in the crystal structure between the apo and bound conformations and closes into the active site when substrate is bound (Figure 3-4) (Baer et al., 2014). Other selinadiene synthases have been identified across many species producing variations in the arrangement of the two double bonds on the key selinane scaffold including, selina-4(14),7(11)-diene synthase from *Humulus lupulus* (Hartley and Fawcett, 1969). Both selina-4(14),7(11)-diene and selina-3,7(11)-diene were isolated from hop varieties in micro-quantities. From these extractions the compounds structures were resolved using gas-liquid chromatography and proton magnetic resonance, but no further studies were carried out.

a



b

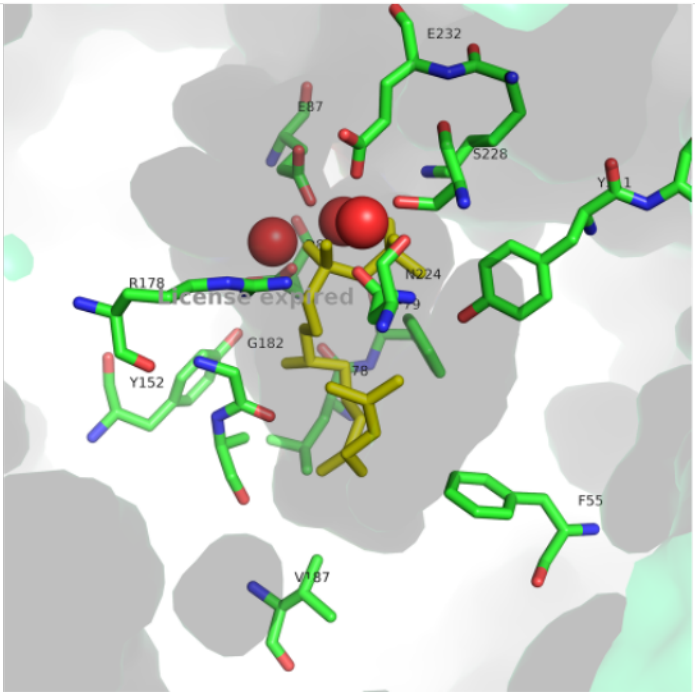


Figure 3-2 Structure of tetrameric selina-4(15),7(11)-diene synthase from Streptomyces pristinaespiralis ATCC 25486 (PDB 4OKZ). *(a) Selina-4(15),7(11)-diene synthase is naturally tetrameric (chain A shown in cyan, chain B shown in orange, chain C shown in magenta and chain D in green). In the crystal structure, each monomer has an enclosed active site containing three magnesium ions, depicted as red spheres, and a non-hydrolysable farnesyl pyrophosphate mimic, dihydrofarnesyl pyrophosphate (yellow). It is postulated that all four active sites are catalytically competent. (b) Active site on chain A of figure A with the key residues coordinating the magnesium ions and the cations produced, depicted as sticks coloured by elements as identified by Ligplot (Wallace, Laskowski and Thornton, 1995).*

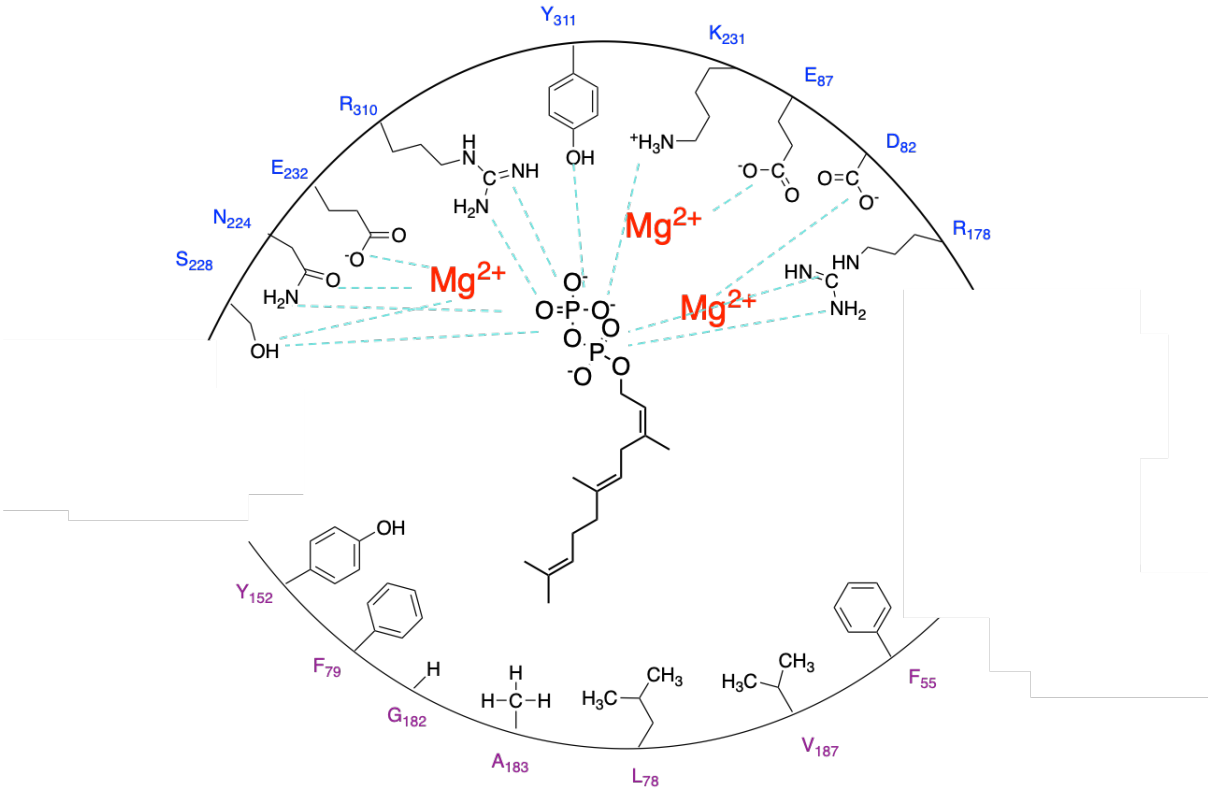


Figure 3-3 Schematic of the active site of selina-4(15),7(11)-diene with farnesyl pyrophosphate. *2D schematic of the active site of SeDS highlighting the residues that interact with the substrate. Residues labelled in blue interact with the pyrophosphate and magnesium ions with the predicted hydrogen bonds depicted a cyan lines. The residues labelled with purple form the hydrophobic cleft that shapes the cation formed by the removal of the pyrophosphate moiety.*

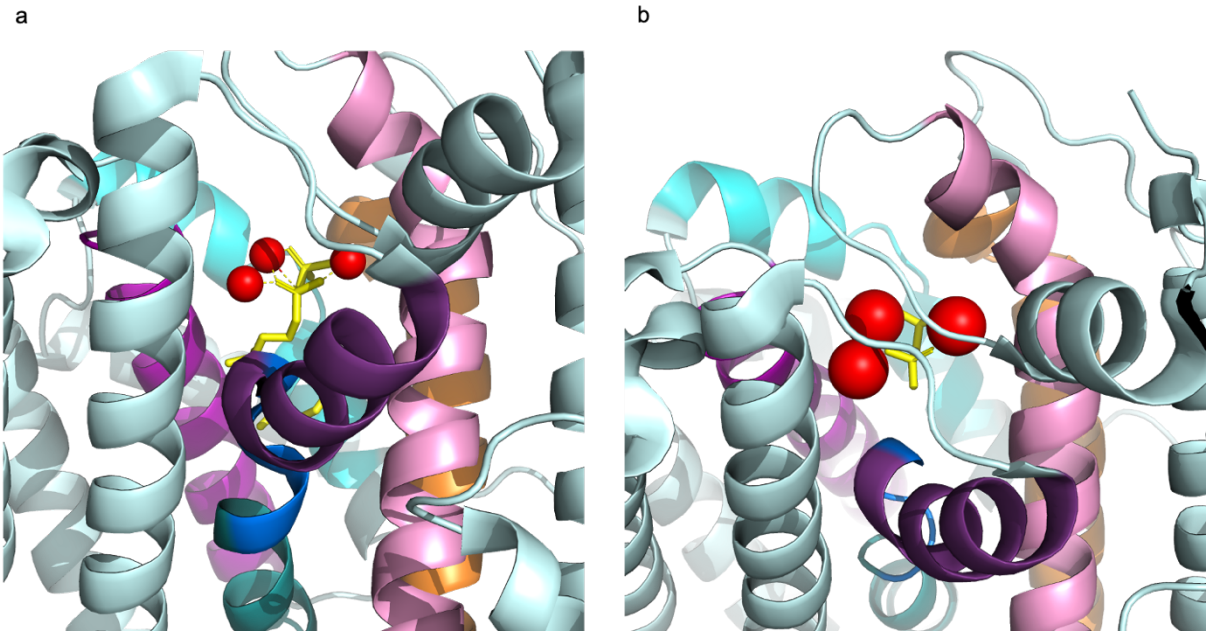


Figure 3-4 Structural comparison of the apo and bound conformations of selina-4(15),7(11)-diene synthase from Streptomyces pristinaespiralis ATCC 25486 (PDB 4OKM and 4OKZ). *Both figures show an active site view of chain A with the key active site helices coloured. Helix B is coloured cyan. Helix C is violet, helix G1 is purple, the G helix break is blue, helix G2 is teal, helix H is pink and helix K is orange. Magnesium ions are coloured in red. (A) Structure of the bound conformation (PDB 4OKZ) has a non-hydrolysable farnesyl pyrophosphate mimic, dihydrofarnesyl pyrophosphate in yellow. (B) Structure of the apo conformation (PDB 4OKM) has diphosphate bound in yellow. The G1 helix depicted in purple has the most significant change between the two structures.*

A catalytic mechanism has been proposed for SeDS in light of the reported crystal structures, computer modelling and mutagenesis studies (Figure 3-5) (Baer et al., 2014) (Das, Dixit and Major, 2016). It is not known which residues involved in the specific restructuring of the farnesyl cation. Baer and co-workers designed SeDS mutants based on three groups of catalytic residues mutating a total of 8 positions, producing 28-point mutants in a bid to gain further mechanistic insight into the enzyme's action. The three groups were formed from; hydrophobic residues that line the active site, residues that coordinate the catalytic magnesium ions, and residues involved in induced fit changes that occurred upon substrate binding. Mutations in the first group of hydrophobic residues resulted mainly in changes to the distribution of the major and minor products but some variants also showed alternative product profiles producing linear products including; (E)-β-farnesene and (2Z,6E)-α-farnesene (Baer et al., 2014). Mutations to the magnesium ion coordination residues and induced fit residues, however, showed reduced activity with some induced fit residue mutants being completely inactive including: D82N, D82E, E159Q E159D. (Baer et al.,

2014). Consequently this proved the induced fit mechanism showing that the coordination of magnesium ions alone is not sufficient to initiate the reaction (Baer et al., 2014). In summary, the hydrophobic residues contouring the active site are the residues that determine the product profile and mutations in the region can lead to alternative products. So far, engineering has only led to linear products with no novel terpenes, however further residues are available for targeting, making SeDS a good target for further engineering.

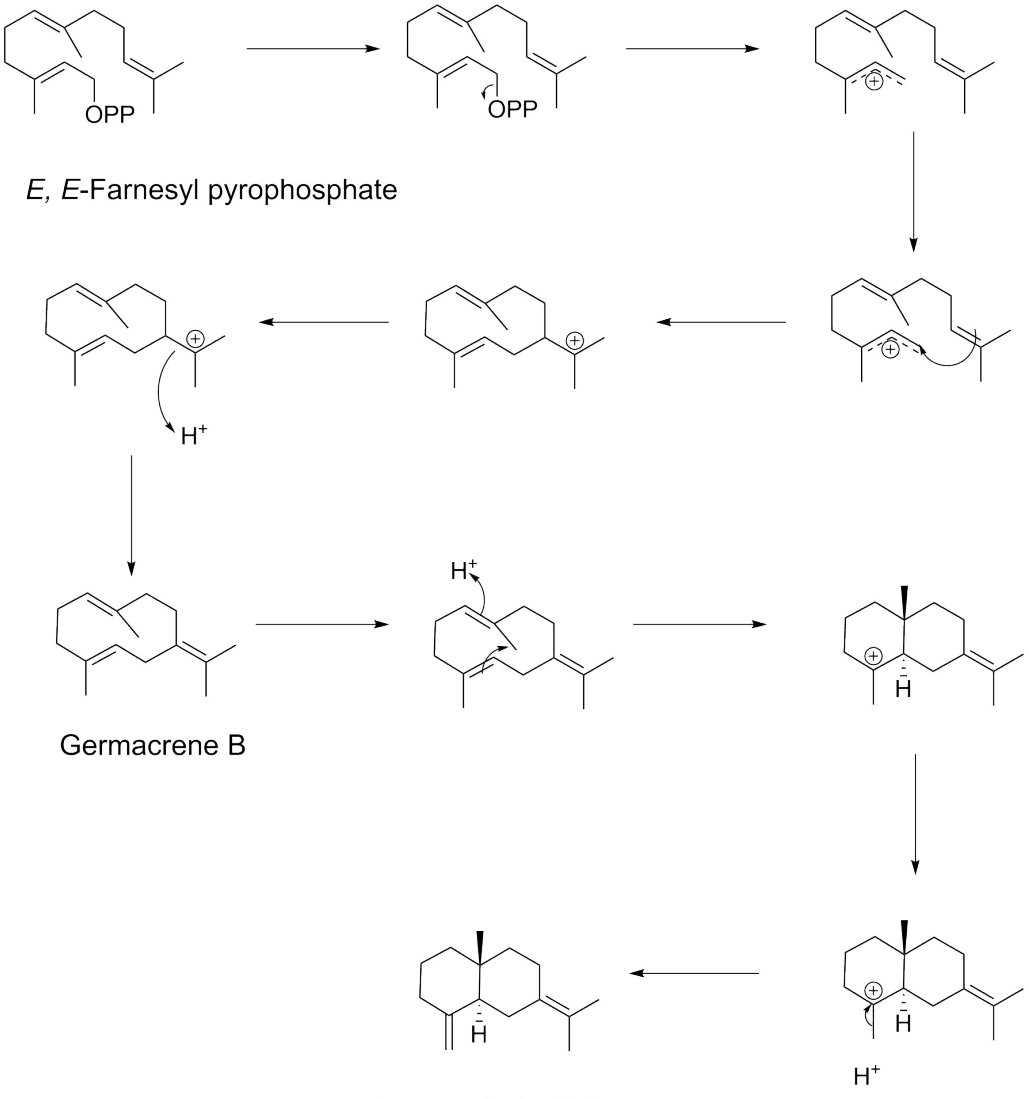


Figure 3-5 Reaction schematic of selina-4(15),7(11)-diene synthase from Streptomyces pristinaespiralis. *The reaction is initiated with the removal of the pyrophosphate moiety coordinated by the magnesium ions producing an allyl carbocation. The allyl carbocation is cyclised to create a double ring moiety before deprotonation and re-protonation of the ring. The final product is produced by the removal of a proton forming a double bond (Baer et al., 2014).*

3.2.2 α-Bisabolene synthase

α-Bisabolene synthase (αBS) produces bisabolene for 90% of its product profile with its postulated mechanism shown below (Figure 3-6) (Christopher L Steele *et al.*, 1998). Bisabolene is produced by a variety of species with a variety of species-specific roles. *Abies grandis* α bisabolene synthase (AgαBS) is the most studied bisabolene synthase to date and is induced as a stress response to wounds in plants (Christopher L Steele *et al.*, 1998). Other uses include as a food additive and as a precursor to hernandulcin, a natural sweetener from plant *Phyla dulcis* (PubChem, 2019). In biotechnology, hydrogenated bisabolene (bisabolane) has desirable properties similar to D2 diesel fuel and so thanks to its potential as a biofuel, bisabolene biosynthetic engineering for increased production has been investigated in *Streptomyces venezuelae* (Phelan *et al.*, 2015). Protein engineering work included; optimising the gene for expression, removing the genes that compete for intracellular resources (including acetyl-coA, pyruvate and glyceraldehyde-3-phosphate) and optimisation of native isoprenoid producing pathway (Phelan *et al.*, 2015). Bisabolene synthase has been mined and characterised from variety of species but only AgαBS has been expressed in *E. coli* (Fujisawa *et al.*, 2010) (Bohlmann *et al.*, 1998).

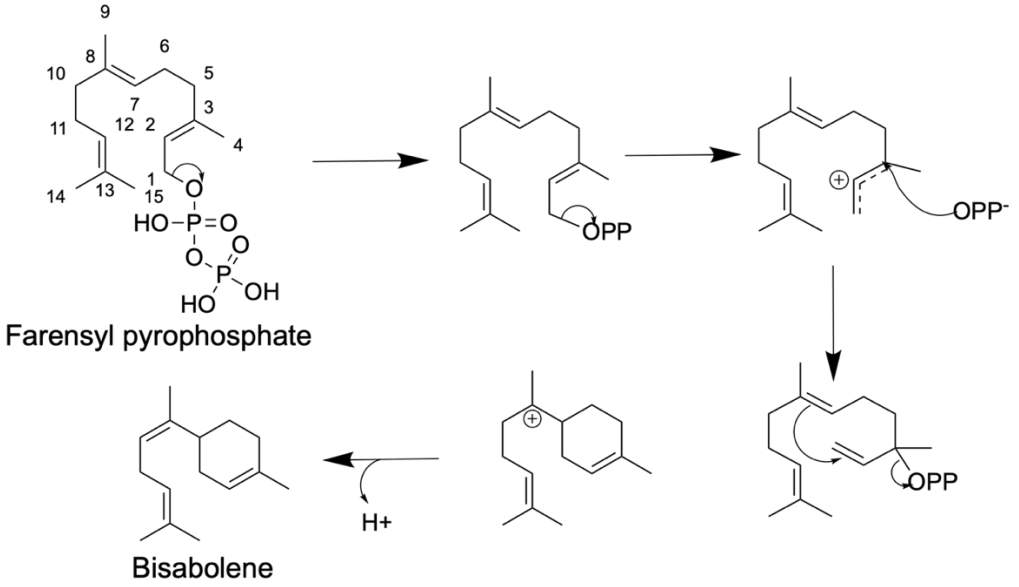


Figure 3-6 Schematic of the postulated reaction mechanism for bisabolene synthase. *The pyrophosphate undergoes metal dependent ionisation and is removed away from the farnesyl carbocation which is delocalized over C1, 2, and 3. The pyrophosphate anion electrophilically attacks at position C3 to assist the cyclisation (McAndrew et al., 2011). The bisabolene synthase active site rearranges the cation so C1 is positioned near the C7-C8 double bond for the nucleophilic attack leaving C8 a cation. This C8 cation leads to acidic hydrogens surrounding this leading to deprotonation at position C10. An alternative mechanism has been suggested where the pyrophosphate anion does not electrophilically attack, but the enzyme can rearrange the cation for cyclization.*

α-Bisabolene synthase from *Abies grandis* is a monomeric enzyme with a unique structure which can be divided into three α-helical domains; α, β and γ (Figure 3-7). The α domain contains the class I active site with the key catalytic motifs including the DDXXD motif (McAndrew *et al.*, 2011). At the β and γ domain interface, is a non-functional class II active site lacking the key conserved catalytic DXDD motif (McAndrew *et al.*, 2011). This shows close phylogenetic relationship to a bi-functional diterpene synthase; for example *Abies grandis* abietadiene synthase compared to monofunctional terpene synthase (McAndrew *et al.*, 2011) (see section 1.5.4). AgαBS has been crystalized in the apo form and in complex with magnesium ions and a range of compounds including, geranlyl thiopyrophosphate and farnesyl thiopyrophosphate (McAndrew *et al.*, 2011). The farnesyl carbocation was directed into the hydrophobic cleft. Hydrophobic residues may direct C1 to the π orbital of the C7-C8 double bond to undergo nucleophilic attack producing the C1-C7 bond. This produces a bisabolyl cation with cationic C8 leading to more acidic C7, C9 and C10 which can subsequently lead to the loss of a hydrogen at position C10 (McAndrew *et al.*, 2011).

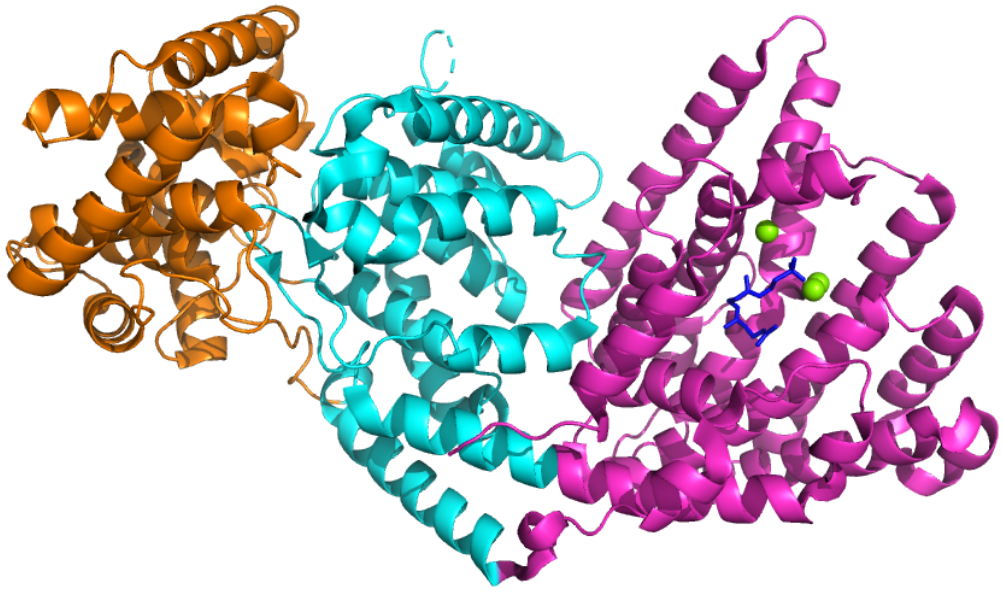


Figure 3-7 Structure of bisabolene synthase from Abies grandis (PDB 3SAE). *The α domain is coloured magenta with the class I active site highlighted with the three magnesium ions displayed as green spheres and the farnesyl pyrophosphate mimic, farnesyl thiopyrophosphate in blue. The β domain is coloured cyan and the γ domain is coloured orange with the inactive class II active site in the interface between the domains.*

3.2 Construction of selina-4(15),7(11)-diene synthase plasmid expression vector

Here, in preliminary work, both AgαBS and SeDS genes were codon-optimised, using the Genscript codon optimisation programme, for expression in *E. coli and* made synthetically based on previous work. This construct was taken forward for expression testing (McAndrew *et al.*, 2011) (Baer et al., 2014). Following expression trials of both constructs, selina-4(15),7(11)-diene synthase was identified as the best candidate with a good level of soluble expression as described in previous work (Baer et al., 2014). Whereas the expression of bisabolene synthase was problematic as despite using a codon optimised sequencing and optimising expression, 90% of expressed protein still remained in the insoluble fraction (Figure 3-8). This would hinder the throughput of screening options since larger expression cultures would be needed and/or expression levels might be too low to detect products in culture media (section 4.2.5).

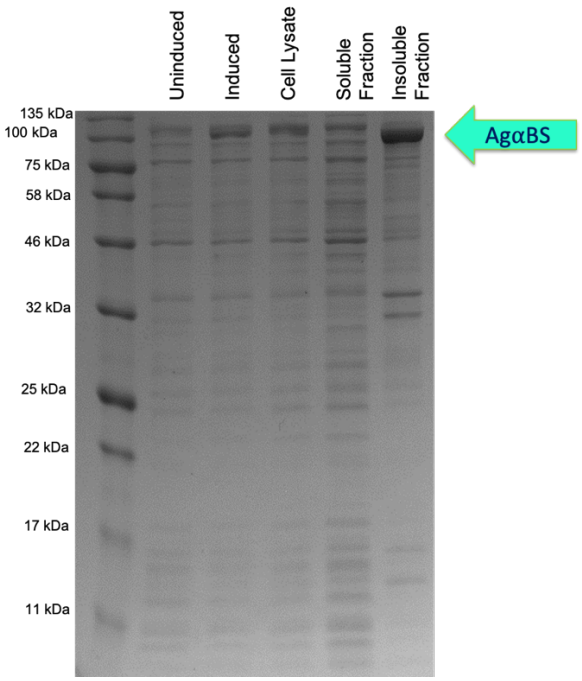
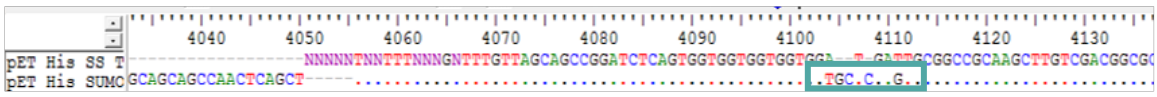


Figure 3-8 SDS PAGE expression gel of* Abies grandis *bisabolene synthase*. Expression test of* Abies grandsis *bisabolene synthase in 50 ml culture, in BL21 DE3 cells, induced with 1 mM IPTG and expressed at 20 ºC. A pre induction and post induction sample were taken along with samples of the cell lysate produced after cell disruption. Following cell disruption, the cell lysate was separated using centrifugation creating a soluble and insoluble fraction. The large band in insoluble fraction correspond to the expected band of* Abies grandis *bisabolene synthase showing the majority of the enzyme is insoluble.*

As previously discussed, the SeDS gene from *Streptomyces pristinaespiralis* ATCC 25486 (WP 005317515) was optimised for expression in *E. coli* using the Genscript programme, synthetically made (Genscript) and cloned into a pET 28a vector digested with the restriction enzymes BamHI and Hind III (Baer et al., 2014). The synthetic gene was sequenced before further work was carried out .The previous *E. coli* expression work was used as a basis for the expression plasmid design (Rabe, Citron and Dickschat, 2013). This used a His_6-SUMO-Ser-Gly tag to purify the protein and a protocol that includes a tag cleavage step (Baer et al., 2014). As a result the expression vector pET His6 SUMO TEV (Addgene) was selected as the suitable expression vector with the added benefit of a TEV cleavage site to remove the purification tag (Baer et al., 2014). The vector includes the option of both an N- and C-terminal His_6 tag. The N-terminal His_6 tag is followed by a SUMO tag and a TEV protease site with a small serine glycine loop for effective cleavage. Primers were designed to add a double stop codon to stop the expression of the C-terminal His_6 tag (Figure 3-9a) (Table 2-5). Mutagenesis was confirmed through sequencing before cloning with synthetic SeDS gene (Figure 3-9a). Cloning of the expression vector and gene was completed using ligation dependent cloning, with the specific restriction sites added to the synthetic gene based on the expression vector; BamHI and HindIII corresponding to the selected expression vector. Cloning was successfully confirmed by sequencing. This produced the SeDS expression vector, named pET_SeDS_exp (Figure 3-10).

a



b

Figure 3-9 Sequencing results for cloning of selina-4(15),7(11)-diene synthase and its cloning vector. *(A) Sequencing results of the double stop codon (highlighted in the box) in expression vector pET His6 SUMO TEV to prevent the expression of the C-terminal His tag before cloning in SeDS. (B) Sequencing results of SeDS gene in expression vector pET His6 SUMO TEV using T7 primer creating SeDS expression vector, pET_SeDS_exp.*

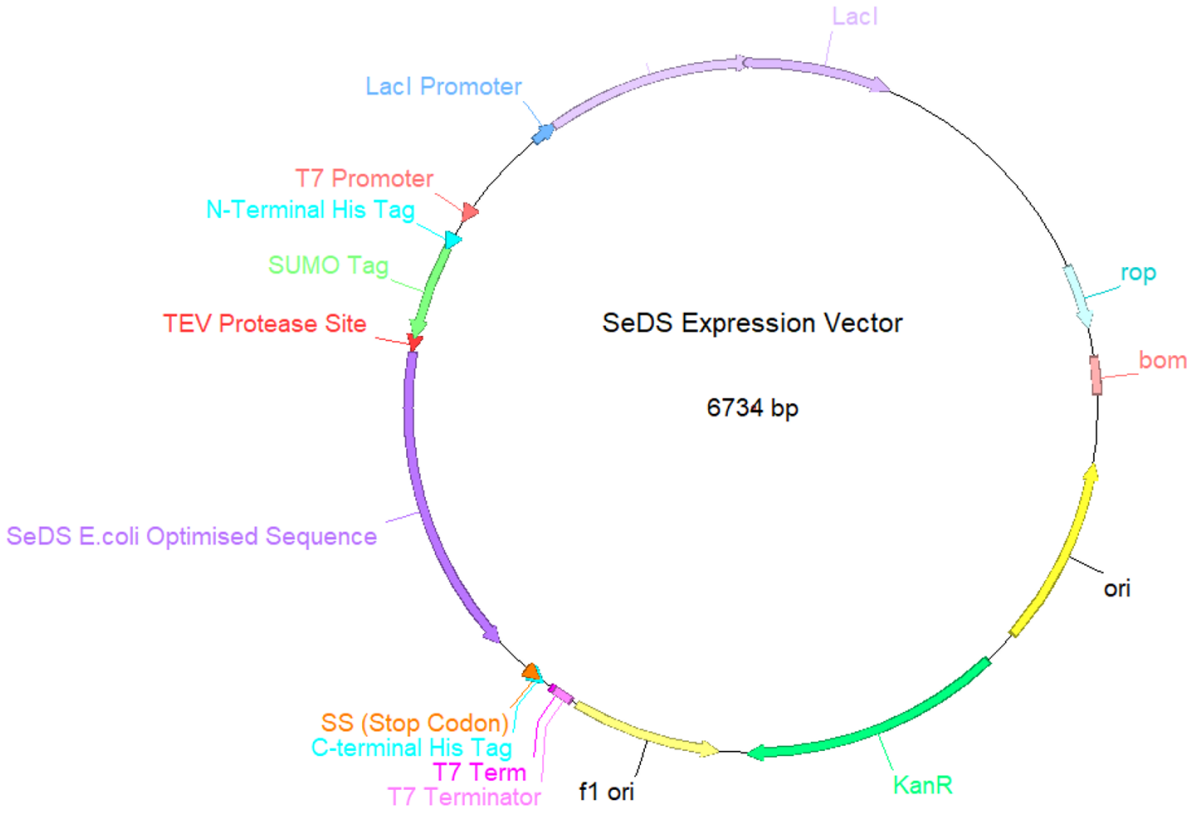


Figure 3-10 Schematic of the selina-4(15),7(11)-diene synthase expression vector (pET_SeDS_exp). *The SeDS expression vector uses the lac operon to provide a promoter for the expression and has a kanamycin resistance gene for selective expression. The SUMO tag can help solubilise SeDS while the* His_6 *will be used as an affinity purification tag.*

3.3 Recombinant selina-4(15),7(11)-diene synthase expression and purification

The pET_SeDS_exp vector was transformed into *E. coli* BL21 (DE3) competent cells for overexpression. Expression was conducted as described in 2.3.1 based on previous *E. coli* expression (Baer et al., 2014). 2xTY media was inoculated with a starter culture and incubated at 37°C, shaking at 200rpm until it reached an OD_{600nm} of 0.4. At this OD_{600nm} the temperature of the incubator was lowered to 18°C. Following further incubation of approximately 1 hour, when the OD_{600nm} reached 0.6, cultures were induced with 1mM IPTG and incubated further overnight. Cells were pelleted at 9,000 g for 20 minutes before being stored at -80°C. Cells are thawed with lysis buffer containing lysozyme, DNase and protease inhibitor tablets. Cells were homogenised before being disrupted at 30 kpi in the cell disruptor (Constant systems). Following cell lysis and pelleting of the cell debris, the His_6 tag in the expression vector was utilised in affinity chromatography purification (Figure 3-11). Following affinity chromatography, the His_6 and SUMO tag were removed by cleavage with

TEV protease (produced in-house) to prevent the tag affecting the folding and organisation of the tetrameric structure.

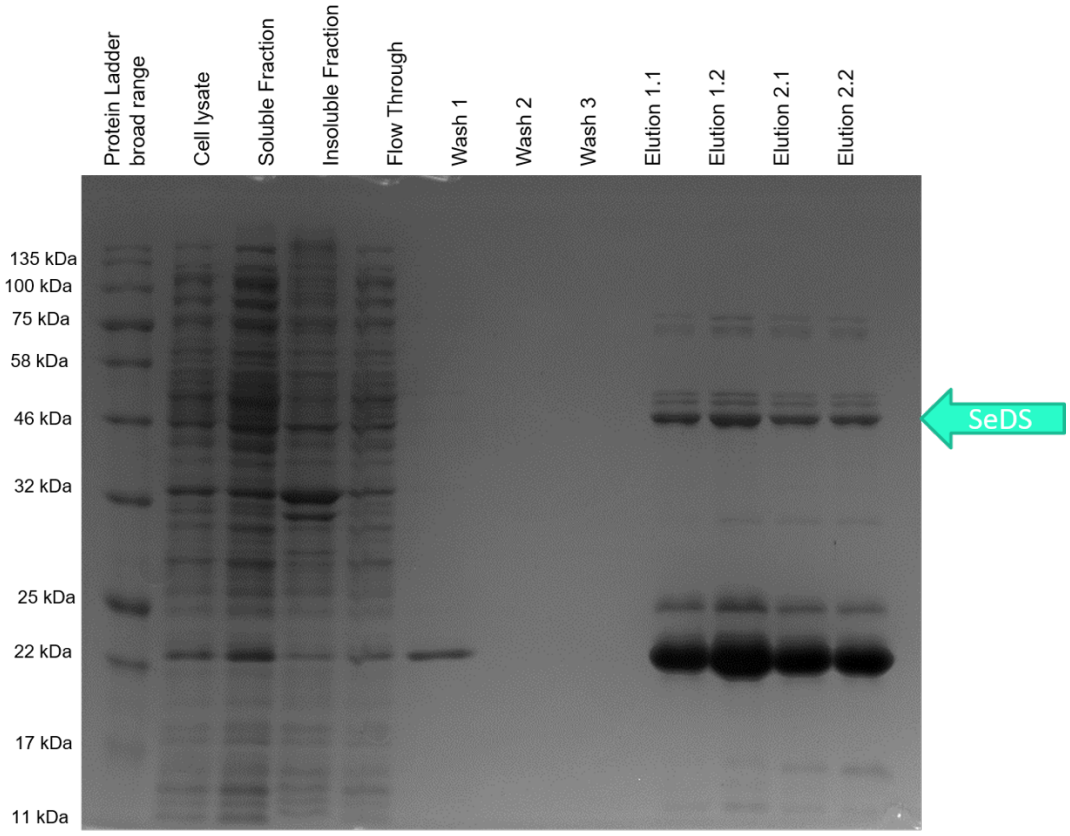


Figure 3-11 Purification of selina-4(15),7(11)-diene synthase. *Purification of SeDS using nickel affinity chromatography. SeDS was eluted in four 2.5ml fractions using 400mM imidazole. Band confirmed with protein identification mass spectrometry using a tryptic digest of the band. All elution fractions were pooled for tag cleavage with TEV protease. Other bands seen in elution fractions are E. coli contaminant proteins with the largest band at approximately 20 kDa predicted to be the E. coli contaminant protein SlyD (Robichon et al., 2011).*

Following tag cleavage, size exclusion chromatography was performed to improve the purity of SeDS and remove the TEV protease (as TEV protease also contains a His tag). According to the calibration curve of the HiLoad 26/600 Superdex S75 an SeDS monomer would elute at approximately 50 ml, so size exclusion chromatography was performed in reduced conditions to keep the samples oligomeric state consistent (Figure 3-12). Following size-exclusion chromatography, fractions containing SeDS were pooled and dialysed to remove DTT. Purity of approximately 75% was achieved which is suitable for kinetic studies but not for structural studies. Positive ESI mass spectrometry was used to confirm the size of a monomer 426045 Da (Figure 3-13).

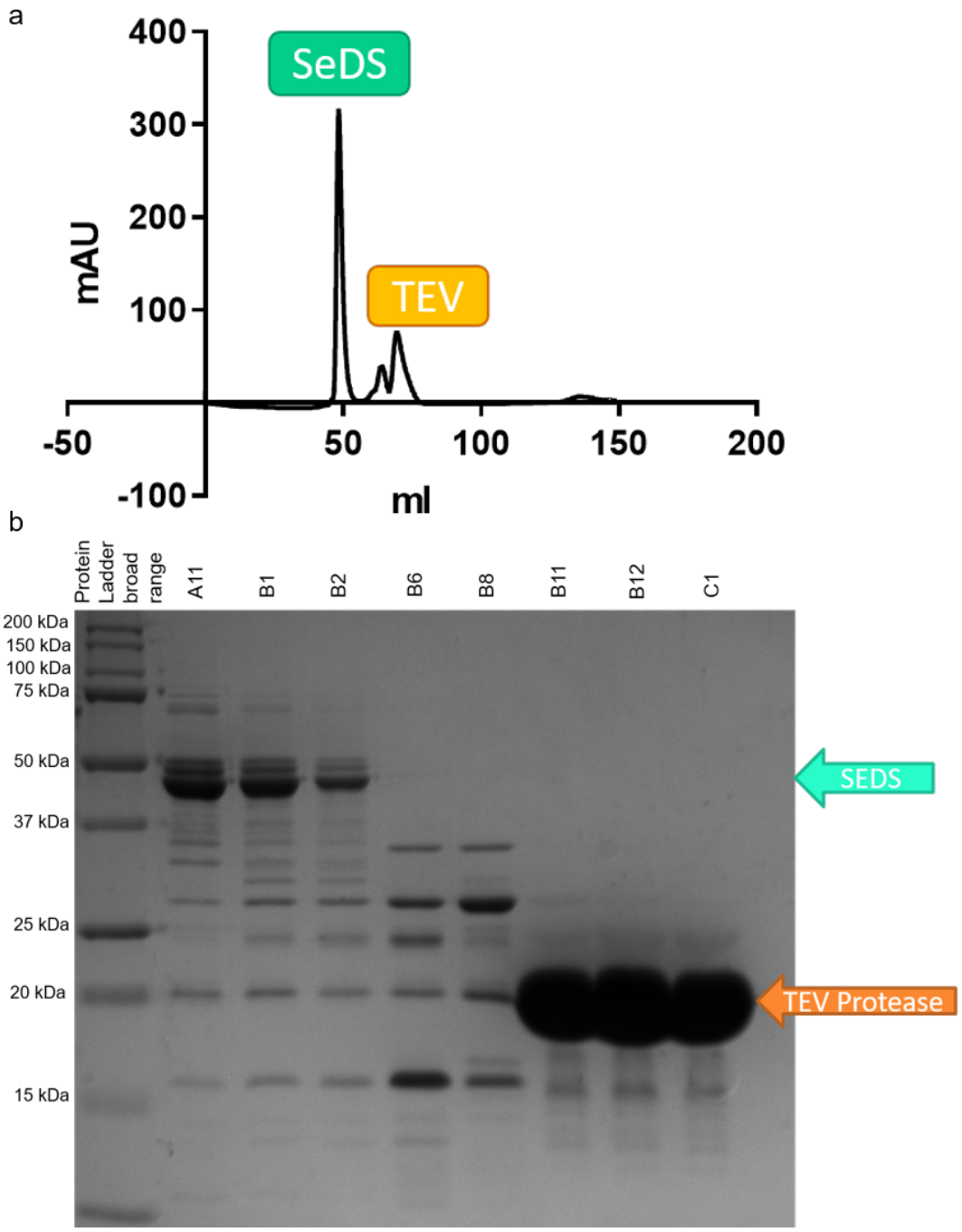


Figure 3-12 Size-exclusion chromatography of selina-4(15),7(11)-diene synthase. *(a) Elution profile of chromatography performed with HiLoad 26/600 Superdex 75 column with a flow rate of 1 ml/min in 20 mM Tris-HCL pH 7.5, 100 mM NaCl, 10 mM $MgCl_2$ and 10 mM DTT. One large peak was observed (SeDS) with two smaller connected peaks (TEV and other E. coli contaminant proteins). (b) SDS PAGE gel analysis of size exclusion chromatography fractions showing SeDS was purified to over 75% purity which is suitable for kinetic studies.*

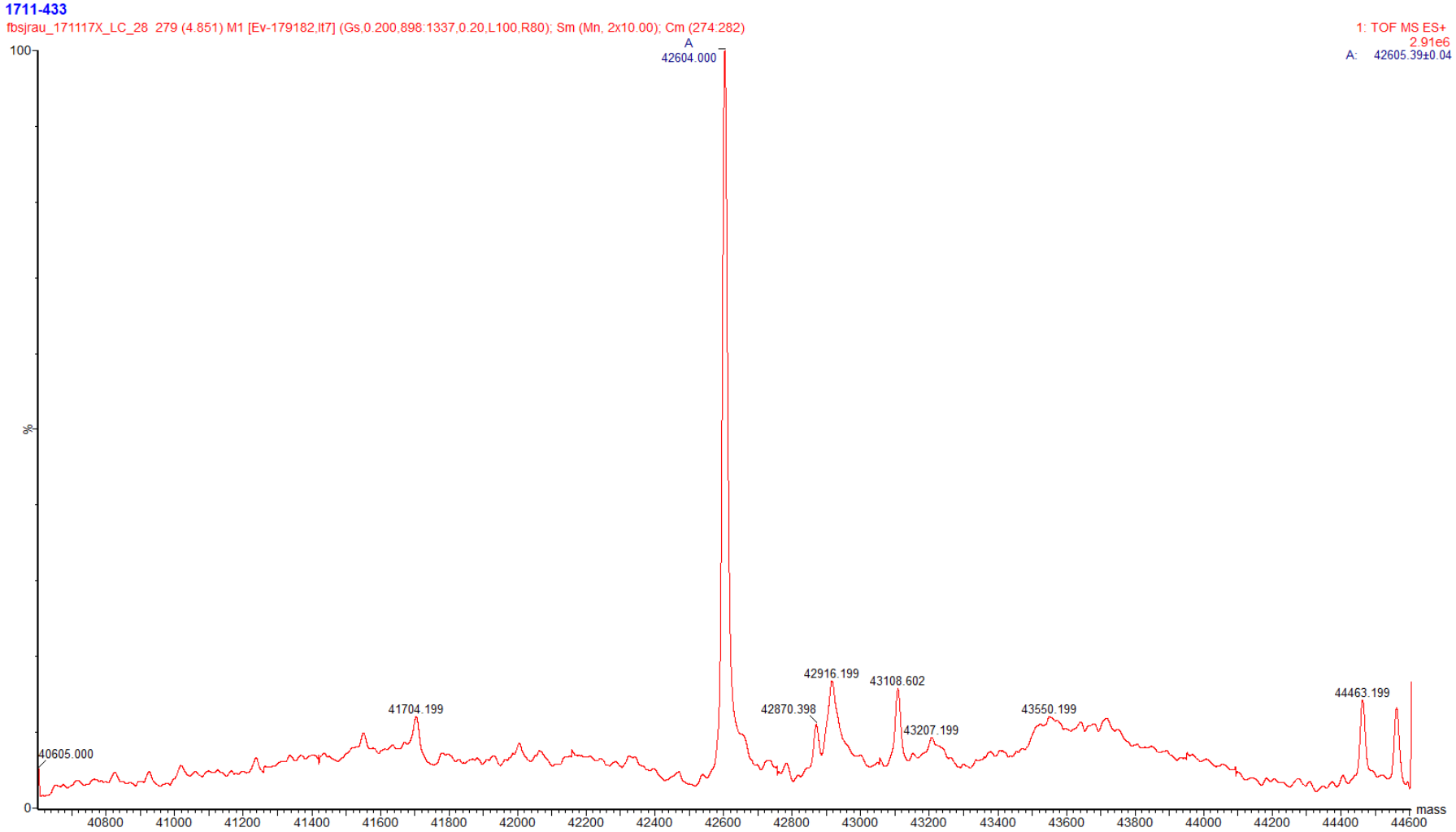


Figure 3-13 Positive ESI mass spectrometry of selina-4(15),7(11)-diene synthase. *The largest peak at 42604 Da corresponds to wild type SeDS with additional contaminant or adducts seen at a significantly lower abundance.*

3.4 Development of an assay for kinetic characterisation

To further confirm that SeDS was successfully expressed and to kinetically characterise the enzyme, the malachite green assay was selected to quantify the enzymes activity (Scott *et al.*, 2010). The malachite green assay is used to detect phosphate production and has been used to measure the activity for a wide range of enzymes including; ATPases, GTPases, terpene synthases and other phosphatases (Scott *et al.*, 2010) (Shatton *et al.*, 1983). In this assay, phosphate is detected as it forms a complex with malachite green and molybdate forming the malachite phosphomolybdate complex (Baginski, Epstein and Zak, 1975). The colorimetric response can be quantified as the concentration of phosphate is in a linear relationship with the absorbance between 600-650 nm (Scott *et al.*, 2010). The reaction is an end point assay as both malachite green and ammonium molybdate are stored in acidic conditions and inactivate the enzyme on addition. The inactivation of the enzyme limits the use of the assay and increase the risk of error. The malachite green assay has been extensively used in activity studies of ATPases or GTPases (Niwa *et al.*, 2012). For terpene synthase activity the enzyme is coupled to an inorganic pyrophosphatase to break down the pyrophosphate released in all terpene synthase mechanisms, to two phosphate ions (Figure 3-14) (Vardakou *et al.*, 2014).

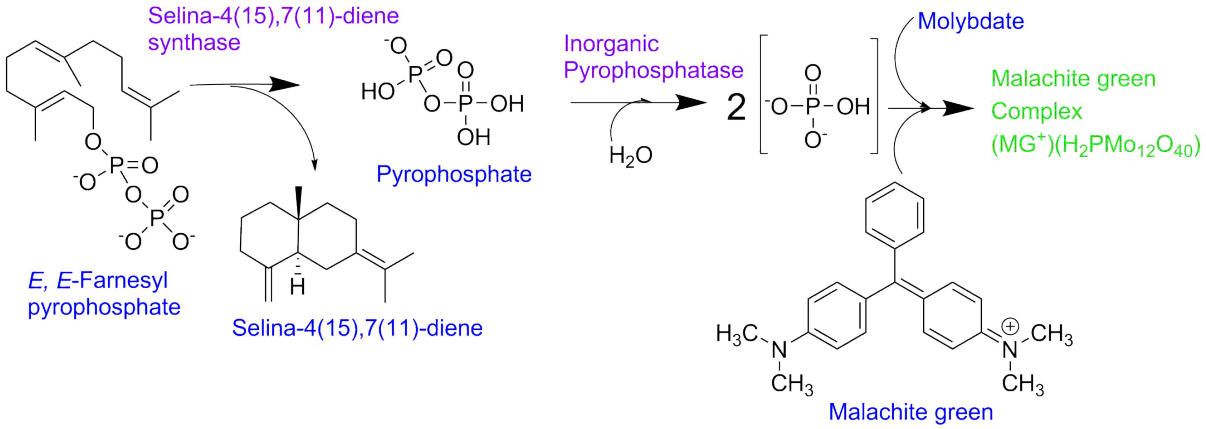


Figure 3-14 Schematic of malachite green with selina-4(15),7(11)-diene synthase. *Selina-4(15),7(11)-diene synthase converts farnesyl pyrophosphate to selina-4(15),7(11)-diene and pyrophosphate. Consequently, the pyrophosphate is broken down into component phosphate ions. These phosphate ions react with molybdate and malachite green to form the malachite green phosphate complex. This green complex can be detected by measuring the absorbance with a filter between 600-650 nm.*

3.4.1 Initial selina-4(15),7(11)-diene synthase activity test

Initial activity tests were performed before final optimisation of the assay to confirm purification of SeDS. Assay conditions and specifically the malachite colour dye solution

have been adapted from previously described methods (Vardakou *et al.*, 2014). The assay was carried out in 25mM MES, 25mM CAPS, 50mM Tris, 5mM $MgCl_2$ pH 7.5 with the reaction components in Table 3-1.

***Table 3-1 Table of malachite green assay components used for initial activity tests**. Components were added and used in three stages (coloured in table)*

Component	Stock Concentration	Final Concentration	Volume for reaction (µl)
Farnesyl Pyrophosphate	10 mM	100 µM and 200 µM	5 (of dilution)
SeDS	2.5 µM	Concentrations ranged between 25 and 10000 nM	20
Assay Buffer	50 mM MES, 50 mM CAPS, 100 mM Tris, 10 mM $MgCl_2$, 60 mU Inorganic Pyrophosphatase	25 mM MES, 25 mM CAPS, 50 mM Tris, 5 mM $MgCl_2$ 30 mU Inorganic Pyrophosphatase	25
Malachite Green	1.83 mM (in 3M sulphuric acid)	1.44 mM	-
Ammonium Molybdate	60.87 mM	11.98 mM	-
Tween 20	11%	0.17%	-
Malachite Green Colour Reagent	1.44 mM Malachite green, 11.98 mM Ammonium molybdate, 0.17% Tween 20		12

Each assay component was prepared individually before the assay into 6 different solutions highlighted in yellow and green in Table 3-1. The reaction was initiated by combining the farnesyl pyrophosphate, with the assay buffer before the addition of SeDS and incubation of these components for a total of 20 minutes at 30°C. The malachite green components (highlighted in green) were combined to create the malachite green colour reagent. After the 20-minute incubation the plate was moved to room temperature for the addition of 12 µl malachite green colour reagent. Subsequently the reactions were incubated for 15 minutes before the absorption was measured at 620 nm. Each SeDS concentration assay was completed in triplicate with a negative control of farnesyl pyrophosphate in the absence of SeDS. All reactions were performed in clear round bottom, 96 well plates.

The reaction results showed that with an increase in SeDS concentration the absorbance at 620 nm increases (Figure 3-15). This proved that the SeDS was active but further optimisation of the malachite green assay and the kinetic conditions was completed before calculation of kinetic parameters.

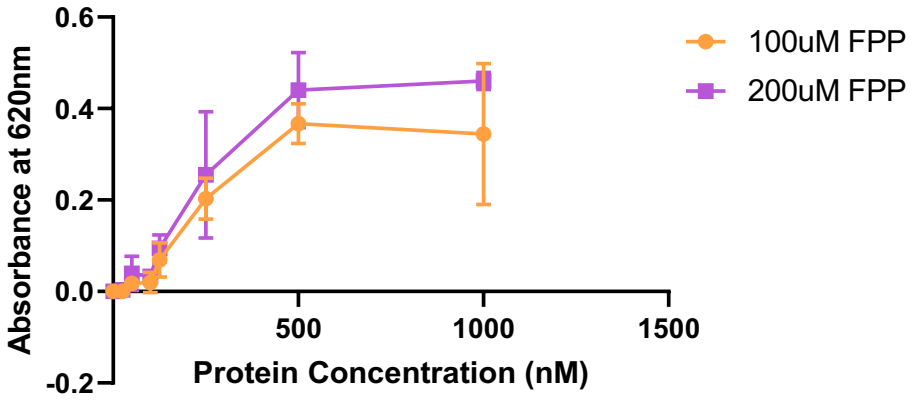


Figure 3-15 Initial activity test of selina-4(15),7(11)-diene synthase. *The concentration of SeDS was plotted against the baselined absorbance of each reaction after 20 minutes. Each SeDS concentration was baselined against the control with no farnesyl pyrophosphate. As the concentration of SeDS increases the absorbance at 620 nm increases with the absorbance until the absorbance plateaus after 500 nM SeDS for both farnesyl pyrophosphate concentrations.*

Figure 3-15 shows that at higher SeDS concentrations the absorbance plateaus showing the maximum rate has been achieved. To achieve conditions for kinetic characterisation either the incubation length needs to be shortened for the initial rate of reaction due to the plateau or that the concentration of farnesyl pyrophosphate is too high for these reaction conditions. For the assay to detect sufficient pyrophosphate release the enzyme concentration should be 250 nM or above. This concentration produces a sufficient absorbance to be above the level of error and variation below 0.1 absorbance units, while still being within the linear range of activity at 200 µM farnesyl pyrophosphate.

3.4.2 Optimisation of malachite green assay

The malachite green assay was developed from previously described methods using terpene synthases (Vardakou *et al.*, 2014). Initial optimisation of the malachite green assay used a thermostable pyrophosphatase with an optimal temperature of 75°C due to the availability of the enzyme. Both sodium pyrophosphate and sodium phosphate dilutions were used as positive controls and tests to optimise the assay conditions before using terpene synthase. The parameter to optimise was the concentration of the thermostable pyrophosphatase to ensure the reaction hydrolysing pyrophosphate is not the rate limiting step. A range of thermostable pyrophosphatase concentrations were tested to find the minimum concentration needed (Figure 3-16). The tests were performed in the assay buffer previously described, with a 50µl reaction consisting of 25µl of 2x assay buffer and 25µl of

sodium phosphate or pyrophosphate (2 x final concentration) (Vardakou *et al.*, 2014). Three repeats were conducted for each concentration of sodium pyrophosphate. The reactions were set up on ice in a 96 well plate and incubated at 75°C for 20 minutes before the addition of the malachite green colour reagent. Following a further 15-minute incubation at room temperature the plate was scanned with a 620 nm filter. (Table 3-1).

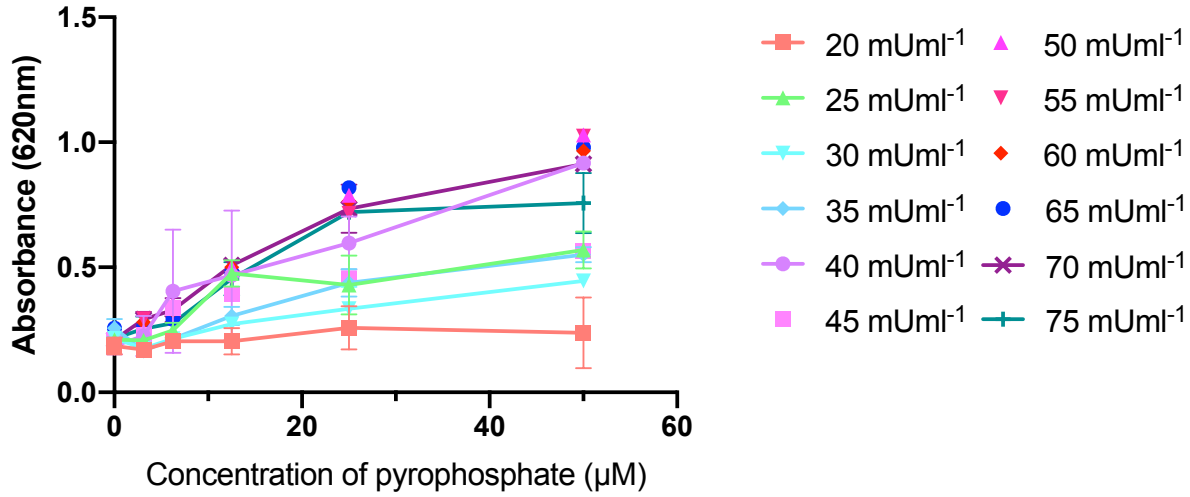


Figure 3-16 Testing 20 mUml^{-1} to 75 mUml^{-1} thermostable pyrophosphatase.
Thermostable pyrophosphatase tested with 20 mUml^{-1} to 75 mUml^{-1} with 50 to 0 μM sodium pyrophosphate, incubated at 75 °C for 20 minutes. Concentrations of 50 mUml^{-} and above produce the same curve giving the optimal concentration as 50 mUml^{-1}.

Figure 3-16 shows that concentrations above 50 mUml^{-1} would be suitable to make sure the hydrolysis of pyrophosphate is not the rate limiting step. To test whether these conditions would work in a coupled reaction the thermostable pyrophosphatase was coupled to the action of Q5 polymerase. This polymerase releases pyrophosphate as a by-product of the reaction, similar to terpene synthase. A 25μl example PCR reaction was set up (10μl Q5 reaction buffer, 5μl 2mM dNTPs, 2μl 200ng/μl DNA, 0.5μl Q5 polymerase) and incubated with 25μl assay buffer (Table 3-1) with 50mU pyrophosphatase. The reaction was incubated for 1 hour before developing the plate as described above (Figure 3-17).

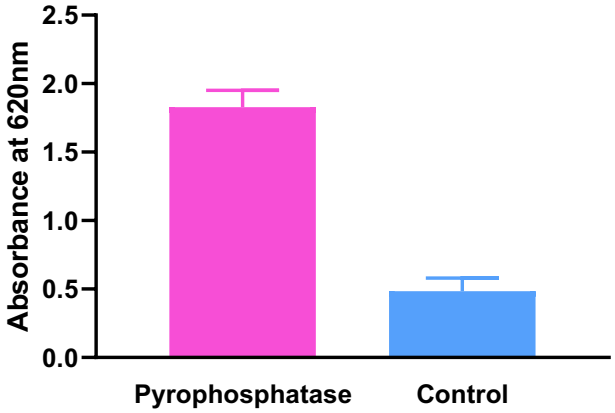


Figure 3-17 Polymerase malachite green coupling test. *Using a basic PCR reaction (10µl Q5 reaction buffer, 5µl 2mM DNTPs, 2µl 200ng/µl DNA, 0.5µl Q5 polymerase) releasing pyrophosphate from the nucleotides. This pyrophosphate could be detected by the action of the thermostable pyrophosphatase and malachite green reagent. Control reaction is a replicate of the pyrophosphatase reaction without a polymerase.*

Following a successful coupled reaction with inorganic pyrophosphatase and Q5 polymerase, initial activity assay testing was completed with both terpene synthases; SeDS and AgαBS. However further optimisation was needed before kinetic characterisation including; incubation at 30°C for terpene synthase reaction, the volume of malachite dye solution and ways to reduced variation seen on results. Throughout this optimisation issues arose including; high backgrounds and evaporation of reactions due to the high reaction temperature of the thermostable pyrophosphatase and the two incubation steps. Therefore, an alternative non-thermostable pyrophosphatase as sourced with an optimum temperature equal or similar to SeDS to prevent the two separate incubation steps and the problems this causes.

Inorganic pyrophosphatase (yeast) (NEB) has an optimal temperature of 30°C equal to SeDS, therefore only one incubation period is needed before the addition of the malachite green colour reagent. As with the thermostable pyrophosphatase initial optimisation began by developing the previously described method (Vardakou *et al.*, 2014). Initial reaction conditions were the same as above except the incubation was at 30°C with a range of concentrations between 10 $mUml^{-1}$ and 60 $mUml^{-1}$ inorganic pyrophosphatase (Figure 3-18).

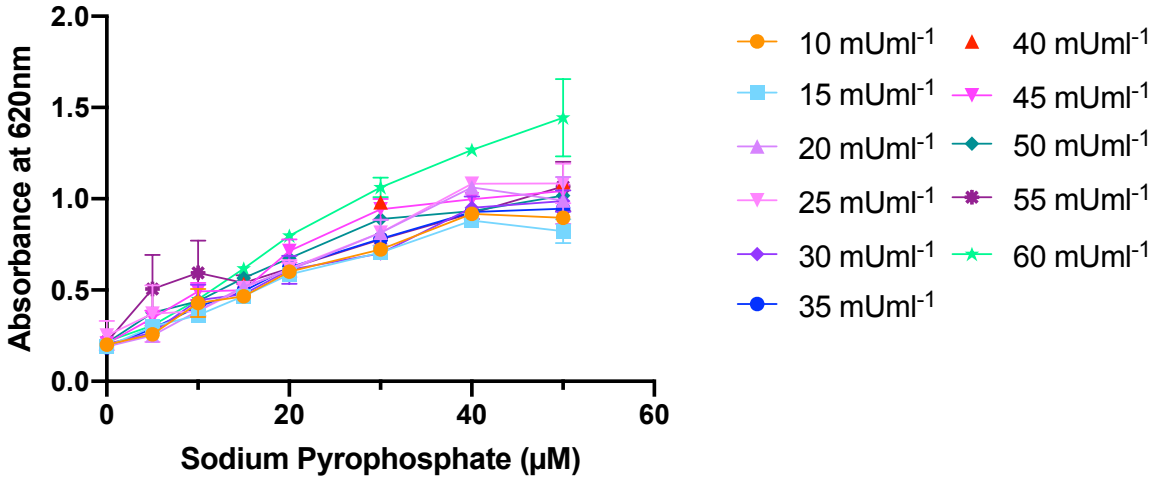


Figure 3-18 Inorganic pyrophosphatase concentration test with 20-minute incubation. *Inorganic pyrophosphatase between 10 mUml^{-1} and 60 mUml^{-1} was tested with a range of sodium pyrophosphate concentration between 0-50 µM. Data used was collected with 20-minute incubation at 30 °C.*

Optimisation of inorganic pyrophosphatase concentration suggests 60 mUml^{-1} for 20 minutes produces a sodium pyrophosphate standard curve with a linear response showing the enzyme is not rate limiting (Figure 3-18). This concentration of inorganic pyrophosphatase was then taken forward for further optimisation of this assay to couple with the SeDS reaction.

3.4.3 Development of a method for kinetic characterization of selina-4(15),7(11)-diene synthase

Following the initial activity test of SeDS (Section 3.4.1) both farnesyl pyrophosphate and SeDS concentrations needed to be optimised to measure the initial rate of reaction to consequently calculate kinetic parameters. This concentration of enzyme should be for a substrate concentration that covers the range above and below the K_m. Following the initial activity test (Figure 3-15) further concentrations of both SeDS and farnesyl pyrophosphate were tested along with incubation period at 30°C. Optimisation included testing a range of SeDS concentrations between 500 nM and 4 µM, a range of farnesyl pyrophosphate concentrations between 5 µM and 200 µM. This was alongside varying the reaction incubation length between 10 and 60 minutes and taking time points within.
However, with each optimisation overlapping error bars were seen, suggesting further optimisation of the assay method was needed and not the reaction concentrations. From a literature search malachite green can form precipitates which could alter absorbance readings. The literature suggested two methods to overcome the precipitation issue; filtering

the malachite green colour reagent before quenching the reaction and dissolving the malachite green in more concentrated, 6 M sulphuric acid (Shatton *et al.*, 1983) (Baykov, Evtushenko and Avaeva, 1988). Both methods were tested with a 25 μM sodium pyrophosphate standard and 60 $mUml^{-1}$ inorganic pyrophosphatase. However, neither method made any difference to the accuracy of the absorbance values. Therefore, other assay components were investigated including pH of the malachite green assay buffer. Malachite green is pH sensitive and has previously been used as a pH indicator along with a phosphate indicator (The Merck Index Online, 2020). Below pH 2.0 malachite green is yellow, between pH 2.0 and 11.5 malachite green is green. Above 11.6 it is a blue green before going colourless at pH 14.0. The malachite green for this assay is stored in acid so below pH 2.0 and its addition can affect the pH so increasing the buffer capacity, to 500 mM Tris, can prevent this pH effect (The Merck Index Online, 2020).

Nonetheless, variability was still seen affecting initial rate calculations thought to be due to small bubbles seen in some wells containing SeDS and not in sodium pyrophosphate control wells. Therefore, alternative malachite green colour reagent compositions were investigated and tested. This could mean different ratio of components, different volumes added to quench the reaction and additional stabilising elements. 11 solutions were selected from the literature and tested in standard conditions with two sodium pyrophosphate concentrations and inorganic pyrophosphatase (Appendix) (Carter and Karl, 1982) (Cogan, Birrell and Griffith, 1999) (Feng *et al.*, 2011) (Geladopoulos, Sotiroudis and Evangelopoulos, 1991) (Guo *et al.*, 2018) (Hua *et al.*, 2018) (Itaya and Ui, 1966) (Lanzetta *et al.*, 1979) (Maehama *et al.*, 2000) (Niwa *et al.*, 2012) (Shatton *et al.*, 1983).

Following this test, a new malachite green colour reagent was taken forward for kinetic parameter optimisation. This new reagent has reduced ammonium molybdate and hydrochloric acid while increasing the concentration of malachite green and removing the detergent, Tween-20 (Table 2-8). The optimisation continued as described above to find the optimal conditions to calculate the initial rate of reactions. The Michealis-Menten principles suggest the ideal condition for calculating kinetic parameters is to test substrate concentrations either side of the K_m (Voet, D., Voet, J.G. & Pratt, 2008). Following further optimisation, the final experimental farnesyl pyrophosphate concentration range was 5 μM to 160 μM with 500 nM SeDS and an incubation of 10 minutes at 30°C (Figure 3-19).

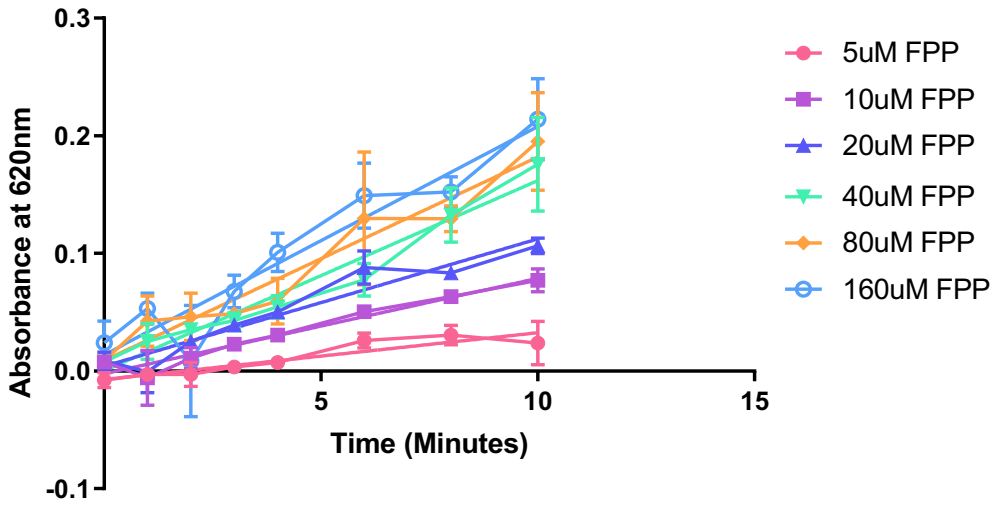


Figure 3-19 500nM SeDS farnesyl pyrophosphate broad range test. *A broad range of farnesyl pyrophosphate concentrations between 5 µM and 160 µM. The reactions were tested over 10 minutes with 500 mM Tris-HCl (pH 7.5) and new malachite green solution, which is a mix 0.12 % malachite green in 1M HCl and 4.2 % ammonium molybdate in 4.5 M HCl in the ratio 3:1. Each absorbance was baselined with the average of the corresponding FPP Blank.*

The results and farnesyl pyrophosphate concentrations tested in Figure 3-18 looked to cover a kinetic concentration range and potentially cover a range of concentrations above and below the predicted K_m so were taken forward for kinetic analysis with Michealis-Menten equation (Johnson and Goody, 2011). The predicted K_m was calculated using GraphPad Michaelis-Menten analysis using the data in Figure 3-19.

3.4.4 Kinetic characterisation of selina-4(15),7(11)-diene synthase

Subsequently the optimised malachite green assay conditions were used to kinetically characterise wild type SeDS. To complete kinetic analysis each of the 6 farnesyl pyrophosphate concentrations were tested in triplicate for each time point. Alongside this each time point has a negative control of substrate with no enzyme and positive controls of sodium pyrophosphate concentrations that make up a standard curve for each assay. The final optimised conditions are described below in Table 3-2.

Table 3-2 – Components for kinetic characterisation of selina-4(15)-7(11)-diene synthase

Component	Stock Concentration	Assay Concentration	Volume for reaction (µl)	Method
Farnesyl Pyrophosphate	10mM	5µM and 160µM	5	Added to plate before assay initiated
SeDS	20µM	500nM	20	Added to each reaction at the time point.
Sodium pyrophosphate	10mM	1.25µM-15µM	25	Positive control and creates standard curve.
Assay Buffer	500mM Tris, 10mM $MgCl_2$, 60mU Inorganic Pyrophosphatase	250mM Tris, 5mM $MgCl_2$, 30mU Inorganic Pyrophosphatase	25	Added to reaction wells, before assay initiated but added at each time point for negative and positive control wells.
Malachite green	0.12% (in 1M HCl)	0.09%	-	Component of malachite green colour reagent
Ammonium molybdate	4.2% (in 4.5M HCl)	1.05%	-	Component of malachite green colour reagent. Solution filtered before use.
Malachite green colour reagent	0.09% Malachite green (in 1M HCl), 1.05% (in 4.5M HCl)	-	100	Components mixed fresh and filtered before use.

Both reactions and controls were set up on ice in a 96 well plate before being incubated at 30°C for up 10 minutes with reactions initiated at each time point. The reactions were stopped with 100 µl of the malachite green colour reagent described in Table 3-2 before being incubated for an additional 30 minutes at room temperature. The plate was then analysed by measuring the absorbance at 620 nm. For accurate kinetic characterisation the conditions were repeated three times with three separate substrate and enzyme stocks and the kinetic parameters were calculated (Figure 3-20, Table 3-3).

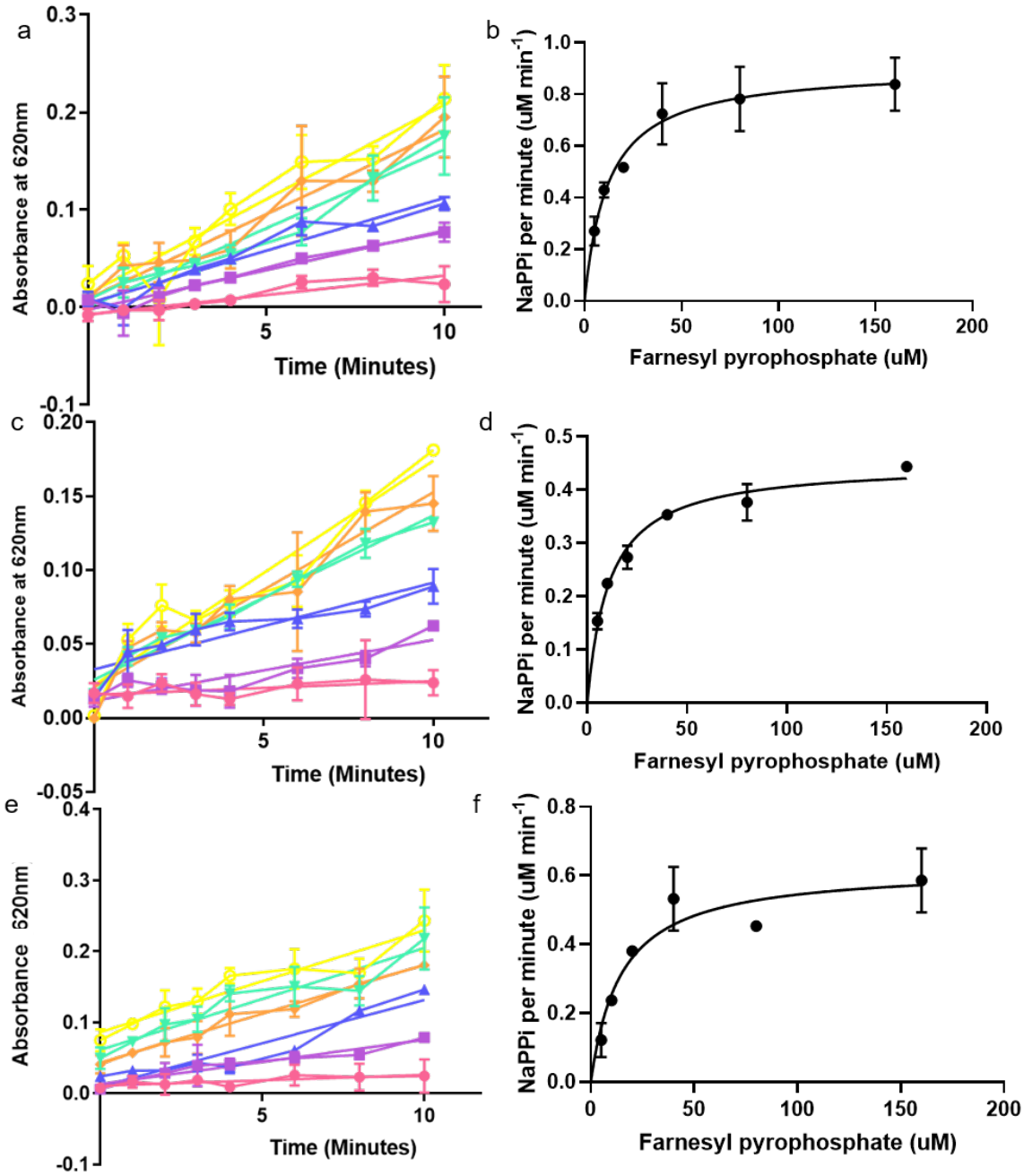


Figure 3-20 Kinetic characterisation of selina-4(15),7(11)-diene synthase. *(a, c, e) The three repeats time vs absorbance at 620nm for 6 farnesyl pyrophosphate concentration between 5 μM and 160 μM with 500 nM of SeDS. For each concentration the initial rate of reaction was calculated using GraphPad linear regression. (B, D, F) Michaelis-Menten curve using the farnesyl pyrophosphate concentration against the initial rate of reaction in pyrophosphate produced per minute.*

Table 3-3 Kinetic parameters of wild-type selina-4(15),7(11)-diene synthase

	V_{max} (μMmin^{-1})	K_m (μM)	k_{cat} (min^{-1})
Repeat 1	0.90 ± 0.04	12.07 ± 2.2	1.81
Repeat 2	0.45 ± 0.01	10.93 ± 1.1	0.99
Repeat 3	0.62 ± 0.04	14.64 ± 3.5	1.25

Kinetic analysis gives a range of K_m between 10 μM and 15 μM and V_{max} between 0.4 and 0.9 μMmin^{-1} and a k_{cat} between 0.9 and 1.8 min^{-1} (Table 3-3). The range between the different repeats is small and is due to kinetic parameters being estimated from experimental data (Vetter, Burcham and Doherty, 2014). The malachite green assay is a discontinuous assay which increase the potential error due to the increased number of reactions needed for a kinetic test. The kinetic values were calculated using an appropriate standard curve taken alongside the kinetic experiments to convert absorbance values to concentration of pyrophosphate. GraphPad Michaelis-Menten analysis was used to calculate the kinetic parameters. As a result, kinetic parameters can be compared from different experiments removing the potential experimental absorbance shifts, occurring from external conditions. Therefore, these reaction conditions and kinetic parameters were taken forward and used to compare the effect of engineering efforts.

3.5 GC/MS characterisation of product profile

Terpenes are volatile compounds that can be analysed using gas chromatography mass spectrometry in organic solvent layers. This method has been used throughout previous work with terpene synthase product profile analysis. No standard of selina-4(15),7(11)-diene or germacrene B was available (Figure 3-5). Therefore, a standard of bisabolene from bisabolene synthase and α-humulene was used to test GC/MS methods and methods of extraction (Figure 3-21 and Figure 3-22). The final GC/MS method used is described in methods 2.5.1 and uses electron impact mass spectrometry.

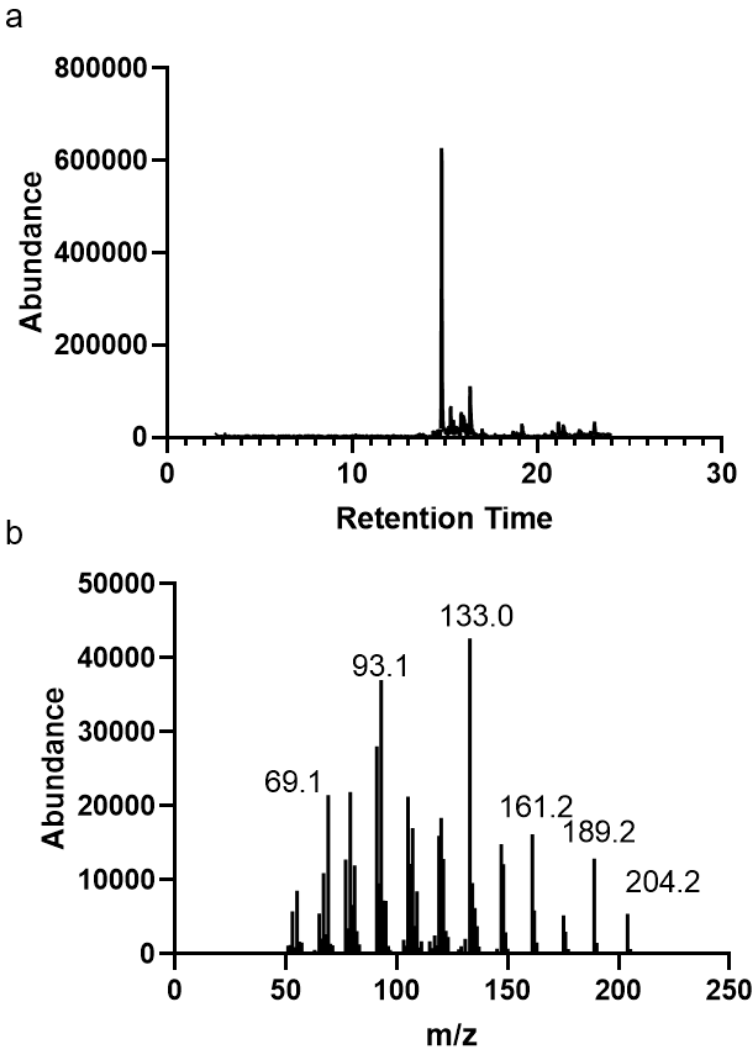


Figure 3-21 GC/MS output for 50 ug bisabolene (mixed isomers) *(a) The GC/MS output of bisabolene standards with retention time against abundance. A large peak is seen at 15.0 minute which corresponds to the mass spectrometry fragmentation pattern in b. The largest mass is 204.2 Da which corresponds to the mass of a sesquiterpene or farnesyl pyrophosphate without the pyrophosphate moiety. Other peaks correspond to other bisabolene isomers and contaminants present in bisabolene and in ethyl acetate. (b) Mass fragmentation pattern of the peak at 15 minutes showing fragments produced when bisabolene isomers break up in the detector.*

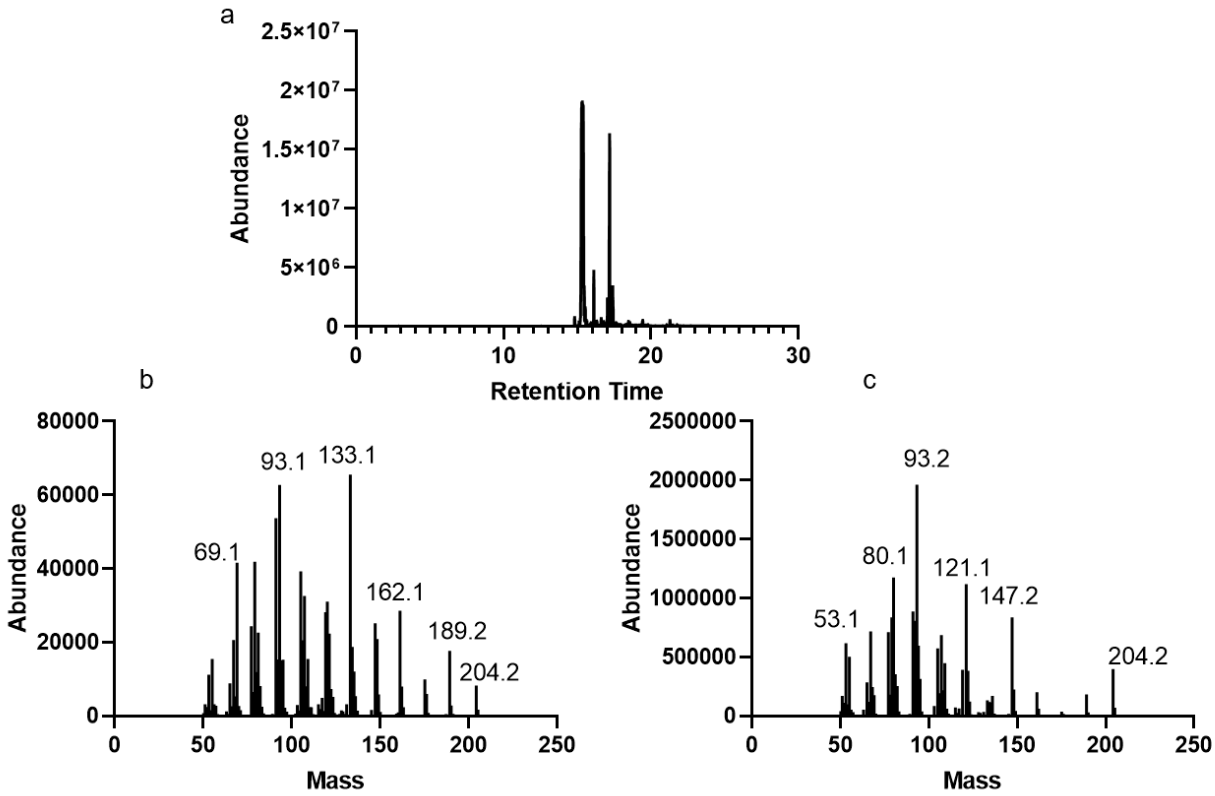


Figure 3-22 GC/MS output for 50 ug α-humulene. *(a) The GC/MS output of α-humulene standards with retention time against abundance. Two peaks with masses of 204 Da were seen at 14.8 and 15.3 minutes. The mass of 204.2 Da corresponds to the mass of a sesquiterpene or farnesyl pyrophosphate without the pyrophosphate moiety. (b and c) The mass fragmentation patterns for the peaks at 14.8 minutes and 15.3 minutes respectively. Mass fragmentation patterns represent the molecular ions produced when α-humlene is broken up in the detector.*

Following confirming a GC/MS method and testing it with sesquiterpene standards, purified SeDS was tested. The method of sample collection for SeDS, optimised from previous methods, is to quench the 200 μl reaction of enzyme and farnesyl pyrophosphate using snap freezing in liquid nitrogen (Styles *et al.*, 2017). Following this, 150 μl ethyl acetate was added and the sample vortex until thawed. To separate the organic layer, it was spun down at 2,100g. The organic layer was sampled and analysed by GC/MS (Figure 3-23) (Styles *et al.*, 2017).

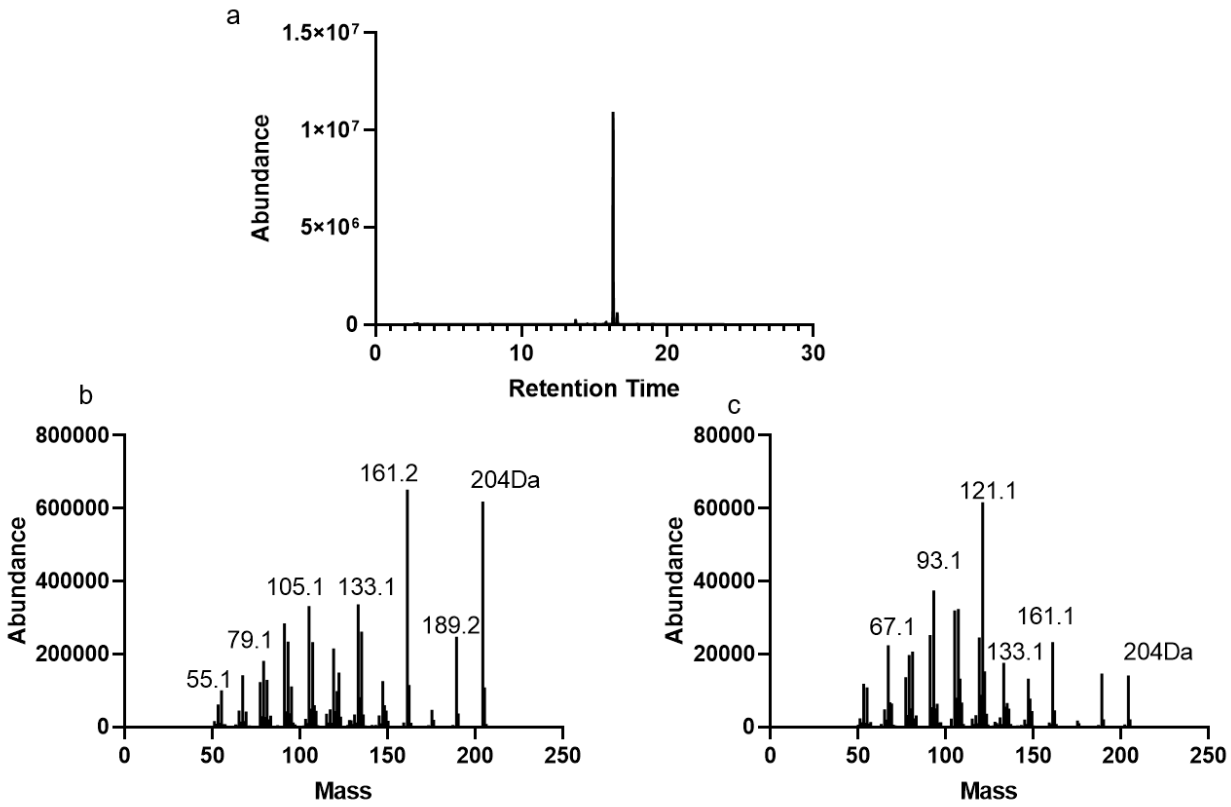


Figure 3-23 GC/MS output of product profile of selina-4(15),7(11)-diene synthase. *(a) The GC/MS output of the products formed by selina-4(15),7(11)-diene synthase, with retention time against abundance. 2 mM farnesyl pyrophosphate was incubated with 5 μM SeDS, incubated overnight at 30℃ (b) The mass spectrometry fragmentation pattern of the largest peak at 16.3 minutes is predicted to be selina-4(15),7(11)-diene. (c) The mass spectrometry fragmentation pattern of the smaller peak at 16.5 minutes and is predicted to be germacrene B.*

The product profile in Figure 3-23 show two products with a mass of 204 Da that corresponds to a sesquiterpene product with no additional atoms. The two different products correspond to the mechanism suggesting the larger peak is selina-4(15),7(11)-diene (Figure 3-23 b) and the smaller peak is germacrene B (Figure 3-23 c) due to previous work (Baer et al., 2014). After confirming the product profile using purified protein the method was adapted to extraction from expressing cultures. This culture expression method was adapted from work with *Saccharomyces cerevisiae* where post induction with IPTG a culture is layered with organic solvent with a density lower than water. The terpenes produced are released from the cells and following a period of expression the organic solvent containing the terpenes can be sampled from the culture (Figure 3-24) (Rodriguez *et al.*, 2014).

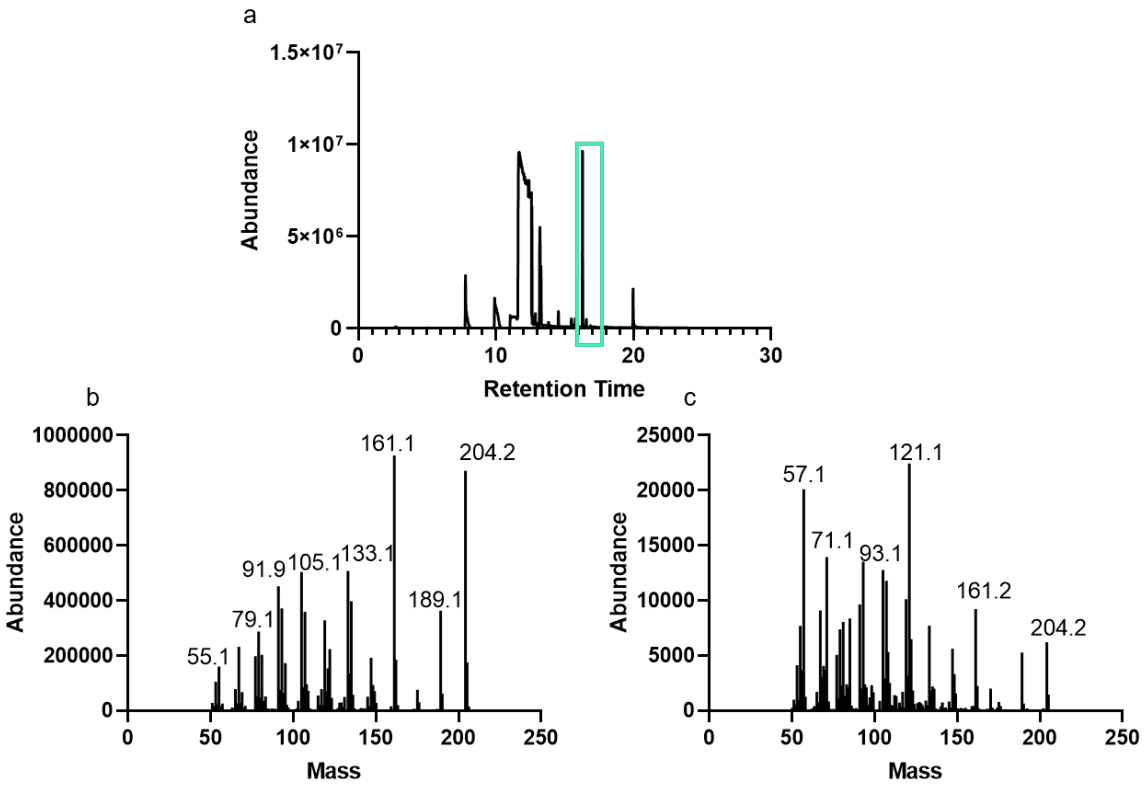


Figure 3-24 GC/MS output of product profile of selina-4(15),7(11)-diene synthase from culture overlay. *(A) The GC/MS output of the products produced by selina-4(15),7(11)-diene synthase, with retention time against abundance. The sesquiterpene peaks are highlighted by the cyan box. (b) The mass spectrometry fragmentation pattern of the largest peak at 16.3 minutes predicted to be selina-4(15),7(11)-diene. (c) The mass spectrometry fragmentation pattern of the smaller peak at 16.5 minutes and is predicted to be germacrene B.*

As both products have the same mass and no standards are available for either product, the only way is to confirm the products by NMR.

3.6 Summary

This chapter described the steps taken to choose a sesquiterpene synthase to engineer. Subsequently the chosen sesquiterpene was expressed, purified, kinetically characterised and identification of the products was attempted

The SeDS gene from *Streptomyces prostinaespiralis* ATCC 25486 was optimised for expression in *E. coli*, synthetically made and cloned into a pET His6 SUMO TEV vector with a double stop codon (Figure 3-10). SeDS was expressed and purified using nickel affinity

chromatography and size exclusion chromatography (Figure 3-11 and Figure 3-12). The size was confirmed by mass spectrometry.

Size exclusion was performed in reducing conditions but these reducing conditions were removed to allow SeDS to achieve tetrameric formation for activity testing. Initial activity testing used a partially optimised version of the malachite green enzyme assay to validate the purification of active SeDS (Figure 3-15). The initial activity test confirmed purification produced an active enzyme and estimated a starting concentration of 250 μM for further malachite green assay optimisation.

The malachite green assay was optimised, testing a range of assay conditions and components. Initial testing used a thermostable pyrophosphatase, however large variances and errors were seen. These errors were due to the necessary two incubation periods for the terpene synthase reaction. The thermostable pyrophosphatase was used to validate the coupling of the assay to another reaction to Q5 polymerase (Figure 3-17). Going forward, an inorganic pyrophosphatase from yeast was used for the remaining optimisation where the minimum excess concentration is 60 $mUml^{-1}$ (Figure 3-18). Subsequent optimisation before kinetic conditions could be determined included; pH of reactions with increasing the buffer capacity of the assay buffer and the components of the malachite green colour reagent.

Following the optimisation of the malachite green assay for the phosphate range that SeDS was active in the kinetic conditions were optimised. The kinetic parameters were calculated using the initial rates of reaction that cover a range of substrate concentrations above and below the K_m. Both the SeDS concentration and the farnesyl pyrophosphate concentration range were optimised along with the reaction incubation length. This optimisation led to the following conditions; 500 nM SeDS, 5 μM to 160 μM farnesyl pyrophosphate with a ten-minute reaction incubation (Figure 3-19). The kinetic parameters were calculated; K_m between 10μM and 15μM and V_{max} between 0.4 and 0.9μM min^{-1}.

To confirm the product profile of SeDS gas chromatography mass spectrometry was used. Using methods adapted from previous work, two peaks with corresponding sesquiterpene masses were seen in the GC/MS output using both purified protein and SeDS expressing cultures (Figure 3-23 and Figure 3-24) (Baer, Rabe, Citron, *et al.*, 2014) (Styles *et al.*, 2017). In summary this chapter shows the expression, purification, kinetic characterisation and product profile analysis for selina-4(15),7(11)-diene synthase.

Chapter 4 - Investigating the specificity of selina-4(15),7(11)-diene synthase variants

The aim of this work was to probe the function of a variety of residues within the active site of selina-4(15),7(11)-diene synthase (SeDS) with the goal of producing novel products. Due to the low to medium throughput of potential screening methods a targeted engineering approach was taken. It was envisaged that a series of targeted mutations would be created and screened using either GC/MS or the malachite green assay (Section 4.2 and 4.3). Any mutation with interesting or altered product profiles would then be further characterised.

4.1 Aims of mutagenesis

The aim of this work was to produce novel compounds by engineering SeDS however, due to the low to medium throughput of potential screening methods, a targeted approach using site directed mutagenesis was taken and the workflow is overviewed in Figure 4-1. Alanine scanning was selected as an initial screening technique to probe the function of a variety of active site residues. Alanine scanning involves selecting a number of residues to target and mutate to alanine, a neutral residue with no extended side chain, and screening the effect (Weiss *et al.*, 2000). These select alanine variants were also used to optimise the GC/MS and malachite green assay screening and for further mutagenesis. As a result of alanine scanning, a set of residues were selected for further mutagenesis. Following completion of further mutagenesis any interesting variants were taken forward for further characterisation and study.

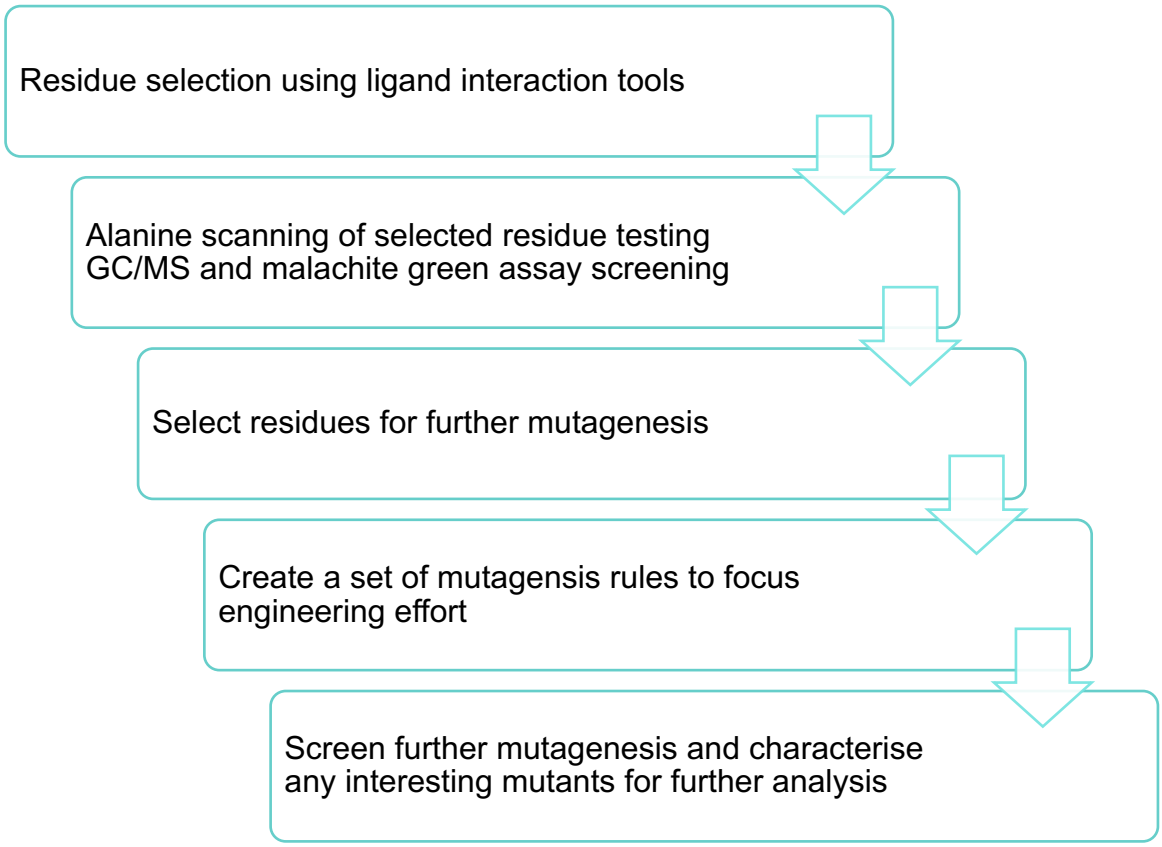


Figure 4-1 Flow chart of the engineering process for selina-4(15),7(11)-diene synthase. *Flow chart depicts the process taken in engineering SeDS in stages. Initial mutagenesis used alanine scanning and these variants were used to determine the best initial screening method. Following alanine scanning, residue positions will be narrowed down and a set of mutagenesis rules determined for further engineering.*

4.2 Alanine mutagenesis

4.2.1 Residue selection

The selection of residues for mutagenesis was aided by SeDS crystal structures including a structure with substrate mimic, dihydrofarnesyl pyrophosphate (PDB 4OKZ) (Baer, Rabe, Fischer, *et al.*, 2014). This crystal structure was analysed using pymol software along with using a ligand interaction analysis tool Ligplot (Wallace, Laskowski and Thornton, 1995) (Laskowski and Swindells, 2011) (Figure 4-2 and 4-3). This online tool generates a 2D schematic of the protein and ligand interactions from a PDB file. Ligplot figures can display a variety of intermolecular interactions to show the different potential interactions with the ligand. The algorithm involves the flattening of rings and the unrolling of the 3D structure around rotatable bonds meaning the atoms connected to this bond are in the same plane, without disrupting the bond lengths in the process. In addition, the algorithm uses the 3D coordinates of the PDB file to process the hydrogen bond information for the ligand and

protein. Following these stages, clean up stages improve the results to minimize overlap giving a clear 2D image (Laskowski and Swindells, 2011). Ligplot software has been widely used including in the aid of designing peptide locks and interactions of lysozyme with monocationic and dicationic liquids (Ernst *et al.*, 2020) (Islam *et al.*, 2020).

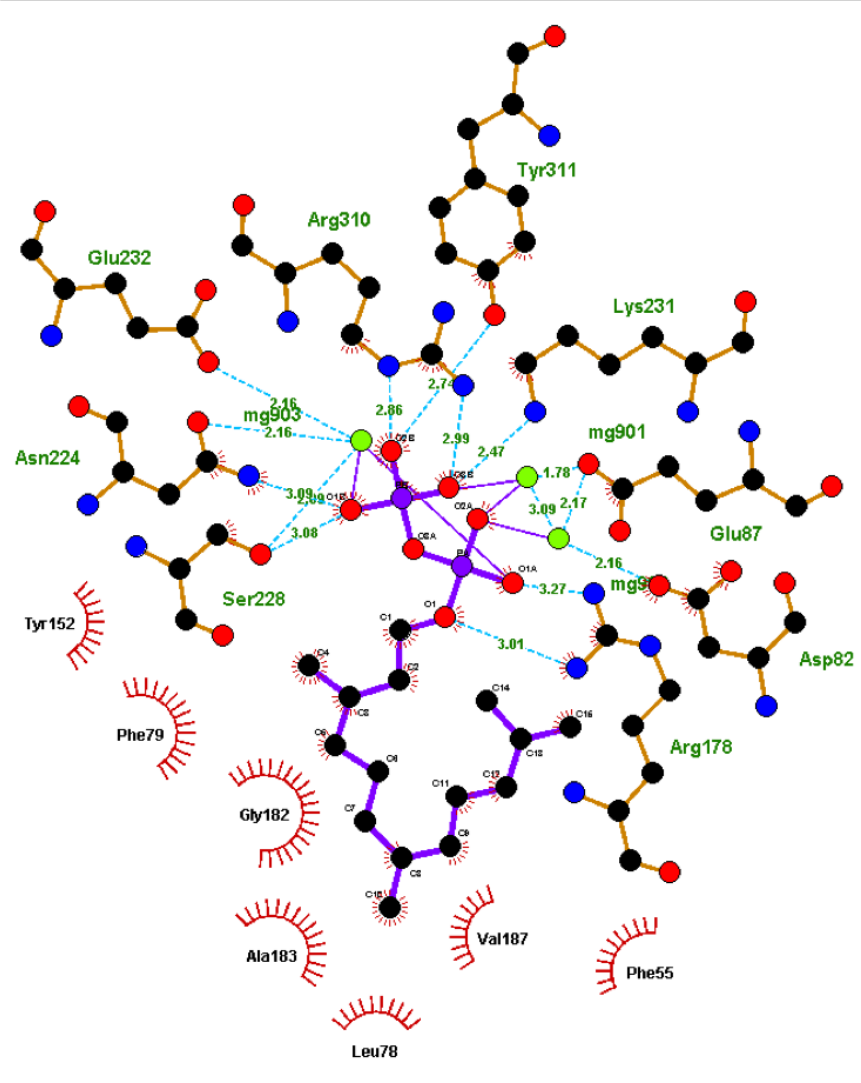


Figure 4-2 Ligplot+ analysis diagram of selina-4(15),7(11)-diene synthase active site from Streptomyces pristinaespiralis ATCC 25486 (PDB 4OKZ). *Selina-4(15),7(11)-diene synthase had been crystallised with an inactive substrate mimic dihydrofarnesyl pyrophosphate (shown in purple). The selina-4(15),7(11)-diene synthase active site has three magnesium ions coordinated by the NSE/DTE motif. Magnesium ions coordinate the pyrophosphate moiety of the substrate whereas the hydrophobic region of the active site coordinates the cation and hydrophobic substrate end with a variety of hydrophobic interactions. Hydrophobic interactions are depicted by red lines and the seven residues highlighted with red circles were selected for alanine mutagenesis.*

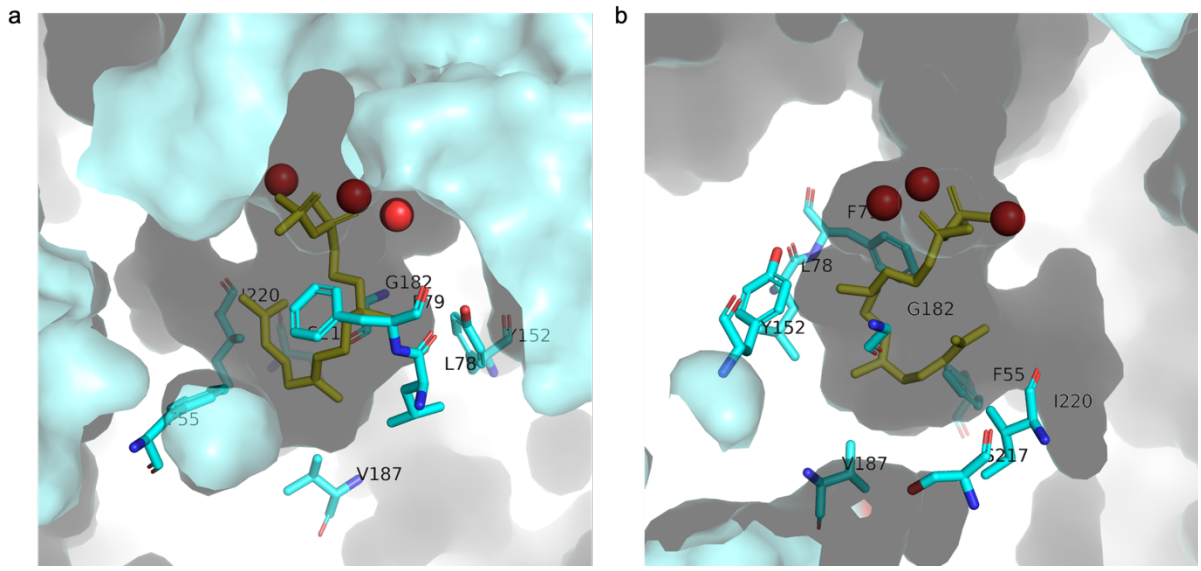


Figure 4-3 Selina-4(15),7(11)-diene synthase active site from Streptomyces pristinaespiralis ATCC 25486 (PDB 4OKZ) with residues for alanine mutagenesis. *The active site of chain 1 of SeDS with magnesium ions depicted as red spheres and an un-hydrolysable farnesyl pyrophosphate mimic, dihydrofarnesyl pyrophosphate in yellow. The residues depicted as sticks coloured by elements are the key residues to coordinate the cation produced in the reaction mechanism as identified by Ligplot and by Pymol structural analysis (Wallace, Laskowski and Thornton, 1995) (PDB 4OKZ). The residues identified are all hydrophobic residues found in the bottom of the active site that control the quenching of the cation and the structure of the final products. A and B show different views of the active site.*

Figure 4-2 shows the Ligplot analysis identifying 7 key residues that interact hydrophobically with the ligand. All residues, excluding the residue A183, were taken forward, in addition to S217 and I220 selected from pymol analysis of the active site.

4.2.2 Mutant creation

Alanine mutants were created using site directed mutagenesis with primers designed using the Agilent online quick change tool (Table 2-5) (Agilent Technologies, 2013) (Novoradovsky *et al.*, 2005). Mutagenesis was completed using PfU Ultra High-fidelity DNA polymerase (Agilent) and used the Agilent quick change 2 protocol (Section 2.2.2). Mutagenesis was confirmed by sequencing before transformation for expression. For each mutant a 50 ml expression test was completed in wild-type expression conditions to confirm expression before large scale expression (Section 2.3). The expression test samples were lysed and the lysate was processed through a mini His spin column to bind any expressed His tagged or

binding proteins. The results of the expression test were analysed by SDS PAGE gel (Figure 4-4).

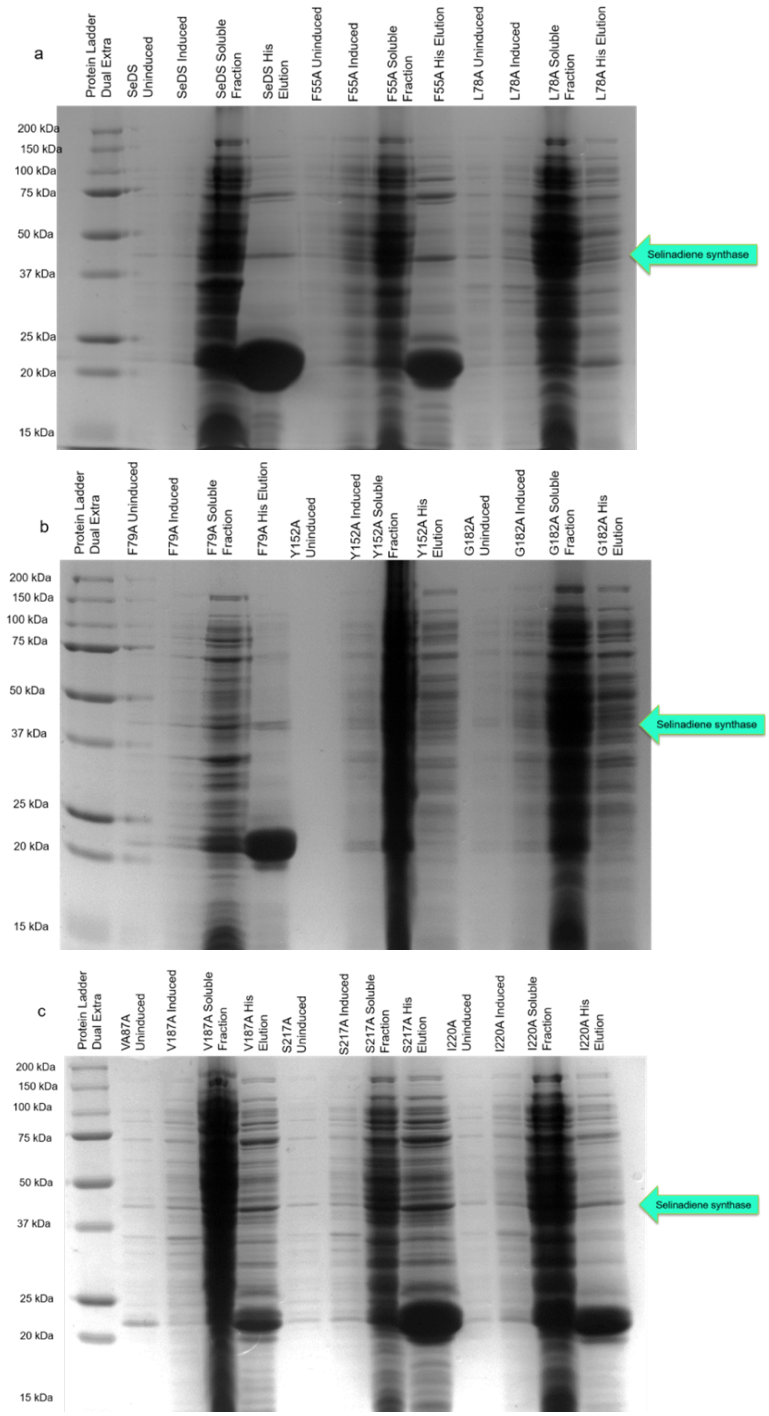


Figure 4-4 Expression testing of selina-4(15),7(11)-diene synthase alanine mutants. *Each alanine mutant was expressed in a 50ml culture and samples were analysed by SDS PAGE gel image. Each mutant shows a band at the expected position in elution samples; however, each elution sample is heavily contaminated. The largest band at around 20 kDa is E. coli contaminant protein SlyD. The expression levels vary between the mutants with Y152A and G182A showing the lowest expression levels but sufficient levels for future studies.*

Figure 4-4 shows the expression level of each mutant varied, with mutants Y152A and G182A having the lowest expression. From the expression level it was estimated that 4 L of large-scale expression would produce a sufficient amount of enzyme for an initial malachite green assay test (Section 2.4). Each alanine mutant was purified using the gravity flow purification protocol (Section 2.3.5).

4.2.3 Production, purification and confirmation of alanine mutants

After producing and confirming the alanine mutants, two screening methods were used to assess the activity and the product profile of the variants. The malachite green assay can be used to test a variant's activity levels based on the consumption of substrate FPP and the consequent pyrophosphate production. The assay is performed at a concentration of FPP above the wild-type K_m over the course of 10 minutes. Using these standard conditions and concentration, an estimate of the mutant's initial activity in absorbance units per minute can be calculated and compare the expected activity of the alanine mutants. This assay will show whether the mutant is active, but the screening method has some pitfalls. This assay needs purified protein lowering the throughput of the assay and its benefit as a screening method. To screen for the product profile, GC/MS in culture screening can be used. This method has a medium throughput and can be done in culture. However, there is limitations as the culture need to be incubated for 72-hour post induction for the optimum concentration of terpenes for analysis. Both screening methods were tested and evaluated for reliability and the value of the results to determine the screening protocol for further mutagenesis.

4.2.4 Malachite green assay screening

The first stage in developing the screening assay was to determine whether full purification or whether the His elution could be used in the assay (Section 3.4). A single farnesyl pyrophosphate concentration was be tested giving comparable initial activity. This would reduce the accuracy of the protein concentration but would significantly reduce the time and increase the throughput of this screening method. Therefore, going forward alanine variants were purified alongside wild-type SeDS using the gravity flow purification method (Section 2.3.5). Following this, wild-type SeDS was tested in the assay conditions with and without TEV cleavage (Figure 4-5).

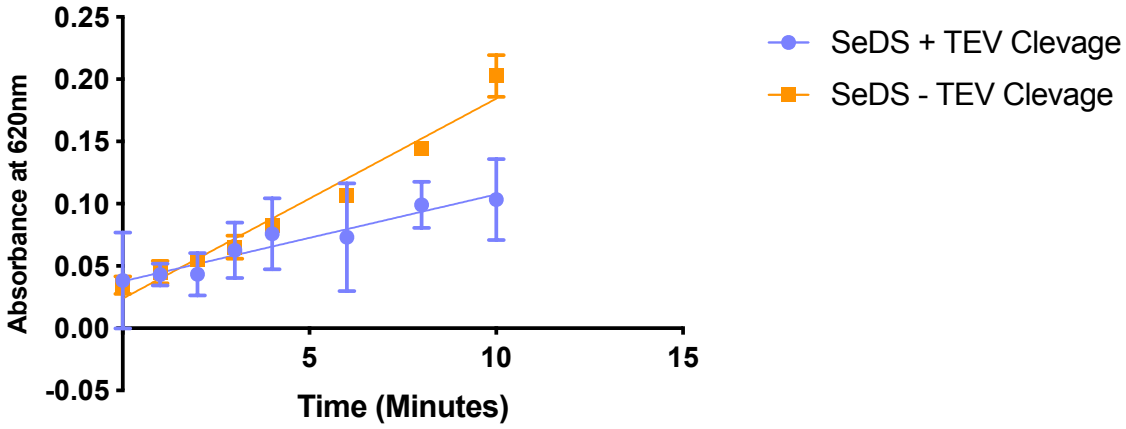


Figure 4-5 Testing the activity of cleaved and uncleaved selina-4(150,7(11)-diene synthase samples. *SeDS was purified using a gravity flow purification method. The His elution was spilt in two fractions, one fraction was cleaved with TEV protease and one fraction was not cleaved. Following cleavage both fractions were buffer exchanged into SeDS storage buffer (50 mM Tris, 100 mM NaCl, 10 mM $MgCl_2$, 10 % glycerol). Both samples were tested with the malachite green assay with 40 µM FPP and 500 nM enzyme over the course of 10 minutes.*

Figure 4-5 shows that un-cleaved SeDS (orange) is still active and therefore the cleavage of a SUMO tag is not necessary consequently improving the throughput of the malachite green assay. The rates of both enzymes were expected to be same, however the addition of TEV reduced the activity rate. This could be due to TEV protease interacting within a step of the assay or the acidic termination of the reaction. In the full purification the TEV protease is separated in the size exclusion chromatography step removing the potential for any assay interactions. Consequently, each mutant was tested in these standard conditions without cleaving the tag and an initial rate was calculated for each mutant (Figure 4-6).

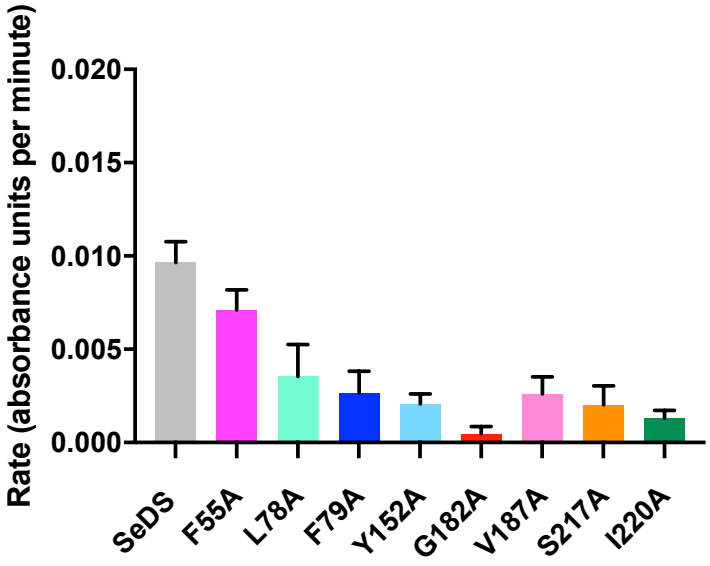
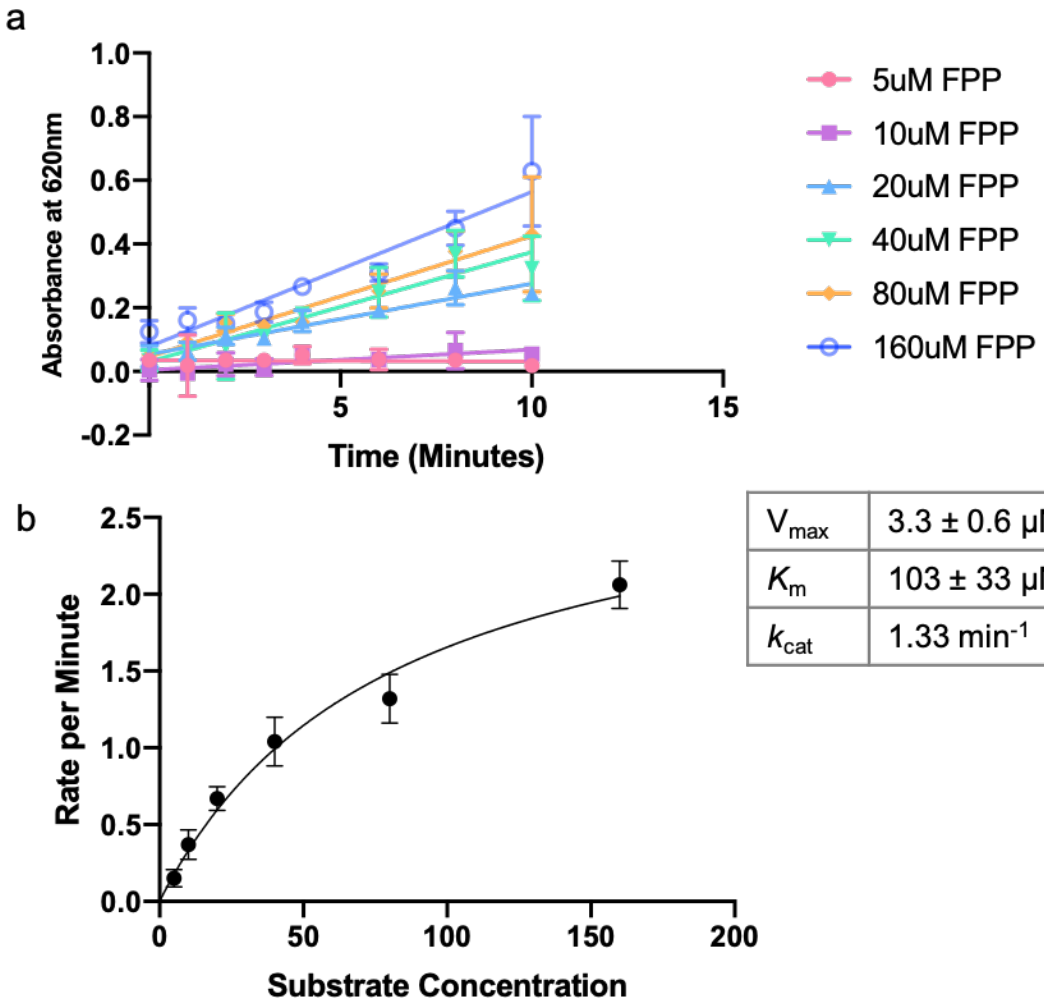


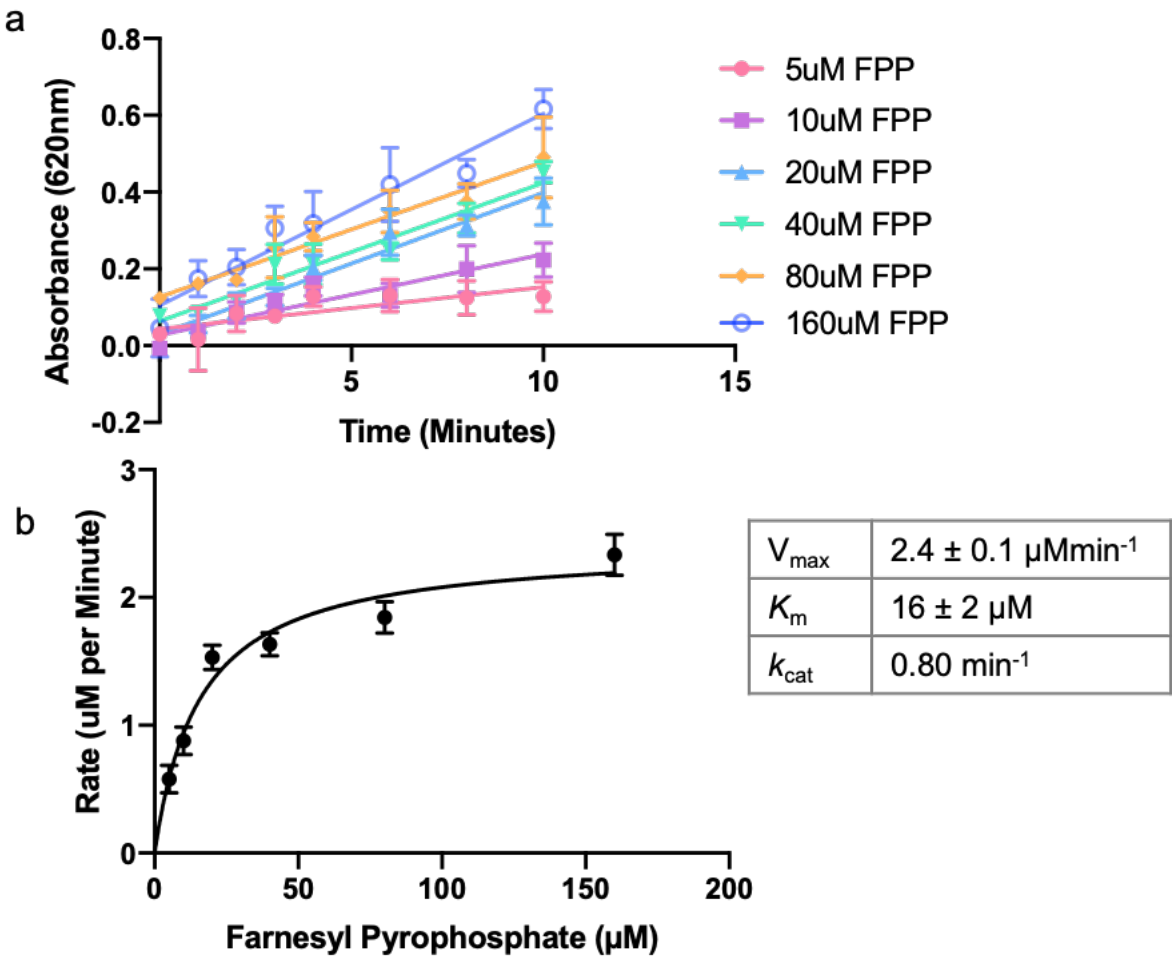
Figure 4-6 Alanine variant testing with the malachite green assay. *Each alanine variant was tested at 40 μM FPP with 500 nM enzyme over the course of 10 minutes along with a wild-type control. From this data an initial rate in absorbance units per minute was calculated and this rate is plotted above. The mutants with the fastest rate were F55A, L78A and F79A.*

Figure 4-6 displays the results of the initial activity tests for each alanine variant showing varying activity levels. The error bars show the error associated with the purity of the protein and the limitations of the discontinuous assay. However, the error was not statistically significant, and the results were used to select variants to take forward. The three most active variants, F55A, L78A and F79A were taken forward for further activity testing and to validate the assay method (Figure 4-6). For further testing a larger scale expression and wild-type purification was completed for each variant (Section 2.3.4). Variant purification was confirmed by mass spectrometry. Initial testing of these purified variants using the same conditions reported in Section 3.4 for the wild-type enzyme confirmed the variants had a lower activity. In order to obtain accurate values for the kinetic parameters for these variants it was therefore necessary to increase the enzyme concentration or the length of the assay. After such optimisation, the kinetic parameters for the variants could be obtained (Figure 4-8 to 4-10).



V_{max}	3.3 ± 0.6 μMmin^{-1}
K_m	103 ± 33 μM
k_{cat}	1.33 min^{-1}

***Figure 4-7 Kinetic characterisation of F55A SeDS**. (A) The F55A variant at 2.5 μM was assayed at six farnesyl pyrophosphate concentrations between 5 μM and 160 μM. For each concentration the initial rate of reaction was calculated using GraphPad linear regression. (B) Michealis-Menten curve using the initial rate of reaction (in μmoles of pyrophosphate produced per minute) plotted against the farnesyl pyrophosphate concentration with the calculated kinetic parameters displayed in the table.*



V_{max}	2.4 ± 0.1 μMmin^{-1}
K_m	16 ± 2 μM
k_{cat}	0.80 min^{-1}

Figure 4-8 Kinetic characterisation of L78A SeDS*. (A) The L78A variant at 3 μM was assayed at six farnesyl pyrophosphate concentrations between 5 μM and 160 μM. For each concentration the initial rate of reaction was calculated using GraphPad linear regression. (B) Michealis-Menten curve using the initial rate of reaction (in μmoles of pyrophosphate produced per minute) plotted against the farnesyl pyrophosphate concentration with the calculated kinetic parameters displayed in the table.*

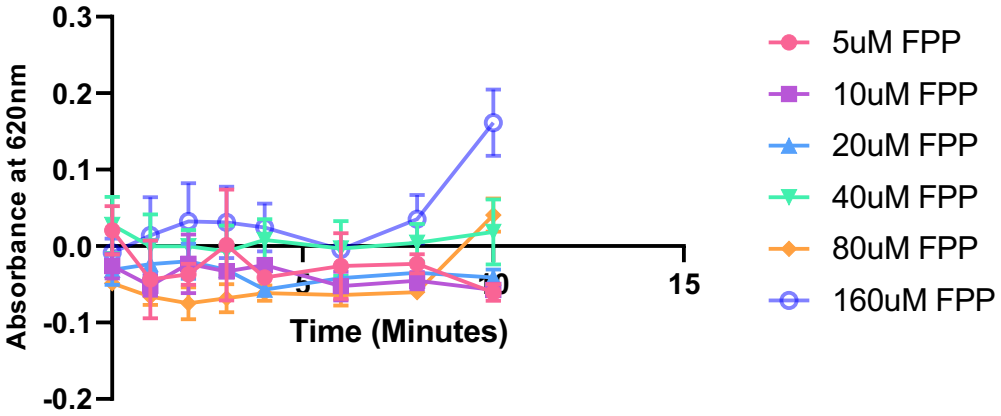


Figure 4-9 Kinetic testing of F79A. *F79A was tested at 2.5 µM enzyme with various FPP concentrations between 5 µM and 160 µM. F79A displayed relative inactivity in these higher enzyme conditions.*

For both F55A and L78A kinetic parameters were calculated by using the same substrate concentrations but increasing the enzyme concentration to five- or six-times the wild-type concentration (Figure 4-7 and 8) (Table 3-2). F55A shows an increase in K_m of approximately 10-fold compared to wild-type SeDS, whereas L78A has a K_m similar to wild-type SeDS (Table 3-3). However, F79A was too inactive to calculate kinetic parameters in any of these conditions (Figure 4-9). To calculate these parameter the reaction time could be increased. However, the further activity testing of F79A reduces the validity of this screening method. This suggests that future mutants should be further purified to be tested in the malachite green assay.

4.2.5 GC/MS screening

GC/MS screening was initially completed using an *E. coli* culture overlay method. A 100 ml culture of each alanine variant in *E. coli* BL21-DE3 cells was expressed and induced with 1 mM IPTG at an OD_{600nm} 0.6. After induction 2 mls of dodecane is layered on the culture and the culture is expressed for 72 hours. During the expression period, terpenes are produced and released by the cells and extracted into the organic solvent layer. This means that amount produced is related to both the activity of the enzyme and its expression level. Following the expression period, the organic layer is sampled and diluted in ethyl acetate, spiked with 40 µg/ml cumene as an internal standard, and analysed by GC/MS (Section 2.5). The GC/MS output provides relative abundance with respect to the sample provided. Therefore, to quantify and compare GC/MS results an internal standard at a set concentration was added to al GC/MS samples. All GC/MS results are standardised against the cumene internal standard relevant abundance. This was calculated by setting the

relevant abundance of cumene to 1,000,000 and creating a dilution factor to standardise the remaining data against. For easiest presentation of the results of the GC/MS analysis the section of resulting chromatogram between retention times of 13.3 and 19.9 minutes is presented (after standardisation to the internal cumene standard). This allows the peaks of the enzymically produced terpenes to be viewed best. (Figure 4-10 and 11). All peaks within this range with a mass between 200 and 300 Da are then flagged for further investigation.

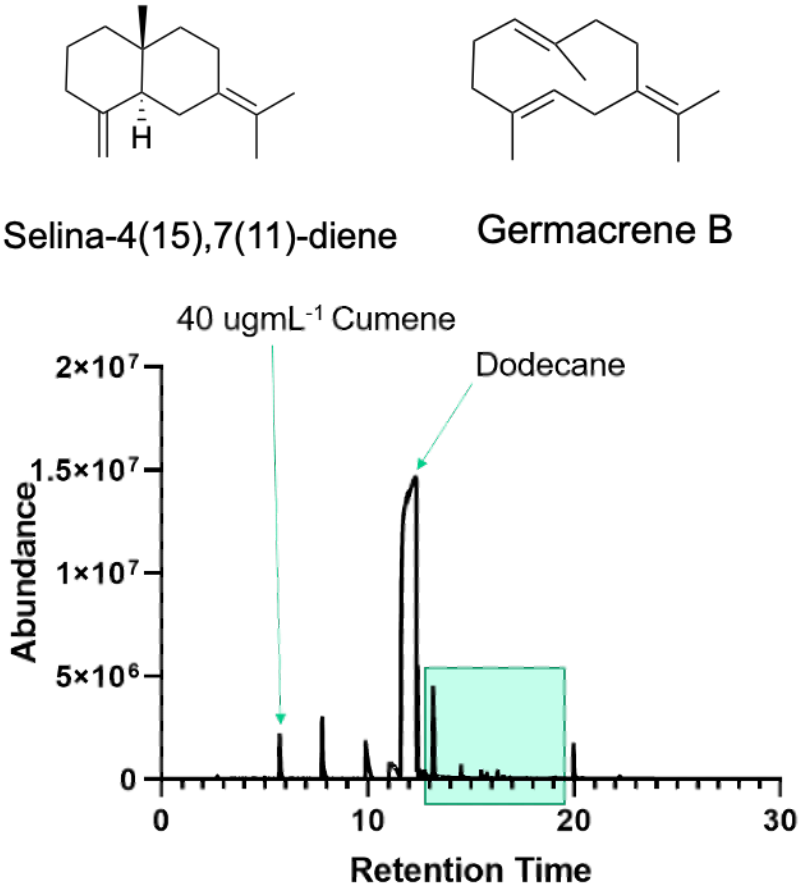


Figure 4-10 Example GC/MS output of terpene products of wild-type selina-4(15),7(11)-diene synthase. *Highlighted in the output are the peaks that correspond to cumene and dodecane. The highlighted green box shows the peaks between 13.3 and 19.9 minutes that will be displayed in further GC/MS analysis to easily compare results. Also displayed are the structures of the two products of selin-4(15),7(11)-diene synthase; selina-4(15),7(11)-diene and germacrene B.*

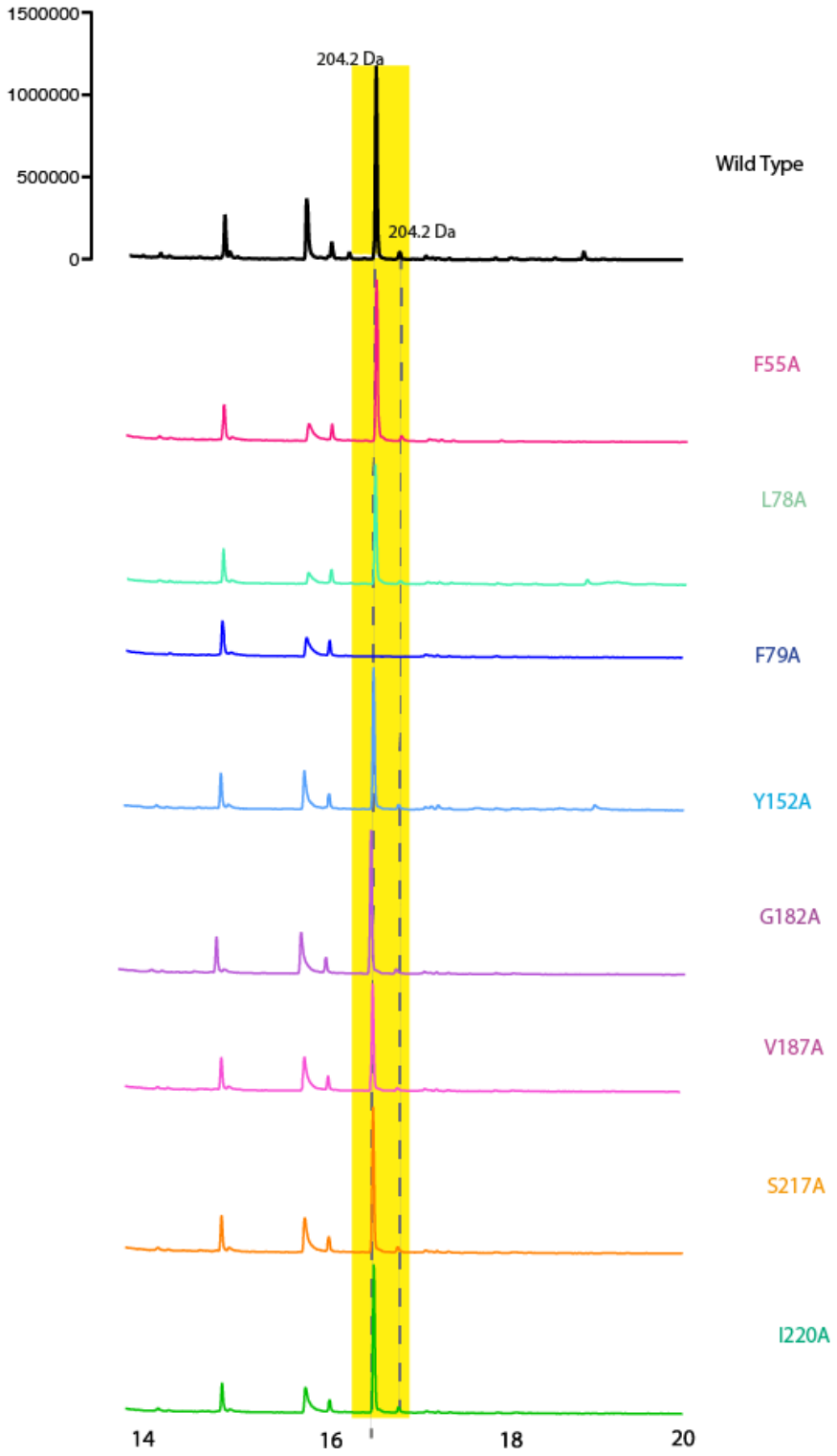


Figure 4-11 GC/MS output of terpene products of for alanine variants of SeDS. *GC/MS output for each alanine variant between 13.3 and 19.9 minutes. All abundances have been standardised against the internal standard 40 µg.mL^{-1} cumene peak and replotted. The yellow box highlights the selina-4(15),7(11)-diene and germacrene B peaks and the dotted lines highlight the two products. The samples have varying abundances of each product with F79A showing no abundance at this level for either product or any additional products.*

From Figure 4-11 no alanine variant showed any novel terpene peaks in these conditions, but all showed production of the wild-type products at varying levels. Therefore, the mutation to alanine does not significantly affect the product profile of the variants excluding F79A. F79A appears inactive in these conditions which correlates with kinetic studies (Figure 4-9). Using the standardised data these results can be quantified by calculating approximate percentage total activity compared to wild-type SeDS and the ratio of the two wild-type products, by calculating the area under the peak (Figure 4-12).

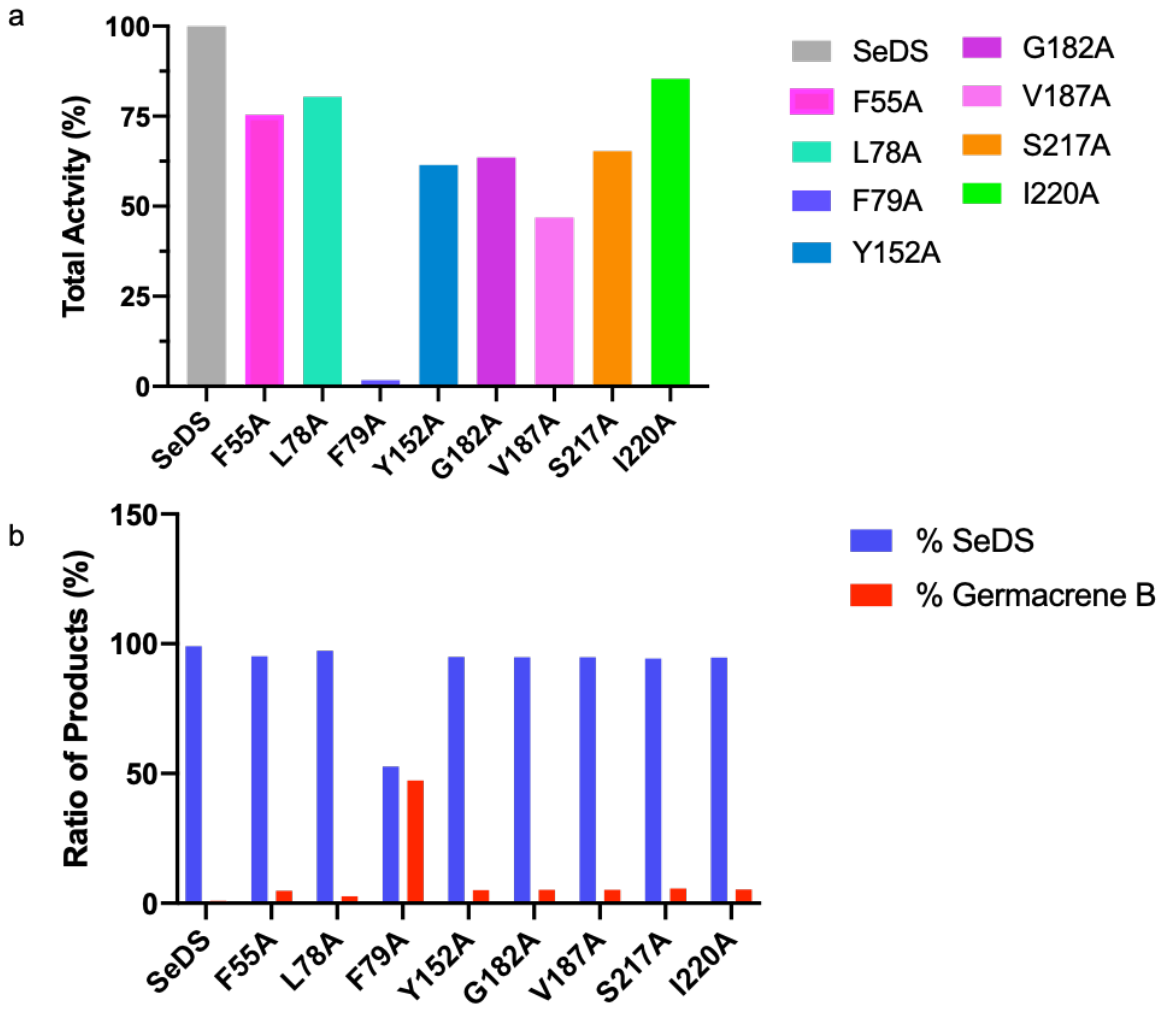


Figure 4-12 GC/MS analysis of alanine variants of SeDS. *(a) Bar chart showing the total activity as a percentage of the total amount of both products in (µg) when compared to wild-type selina-4(15),7(11)-diene synthase. Activity is based on the enzyme's activity level and its expression level. All variants show lower activity than wild-type with F79A showing no, or low, activity. (b) Bar chart showing the ratio of two wild-type products for each alanine variant. The ratios do not take into account the total amount of production in relation to wild-type. Both figures used cumene standardised GC/MS data in analysis.*

As previously mentioned, GC/MS activity levels are affected by both the enzyme activity level and the enzyme expression level. All variants show a lower total production of both products when compared to wild-type SeDS. F79A has the most significant reduction in activity compared to any other position. However, F79A has the most significant change in product profile distribution with a near 50:50 ratio.

As F55A, L78A and F79A were kinetically characterised, GC/MS analysis was conducted on the purified protein to verify the results of the GC/MS screening method from the overlay method (Figure 4-13).

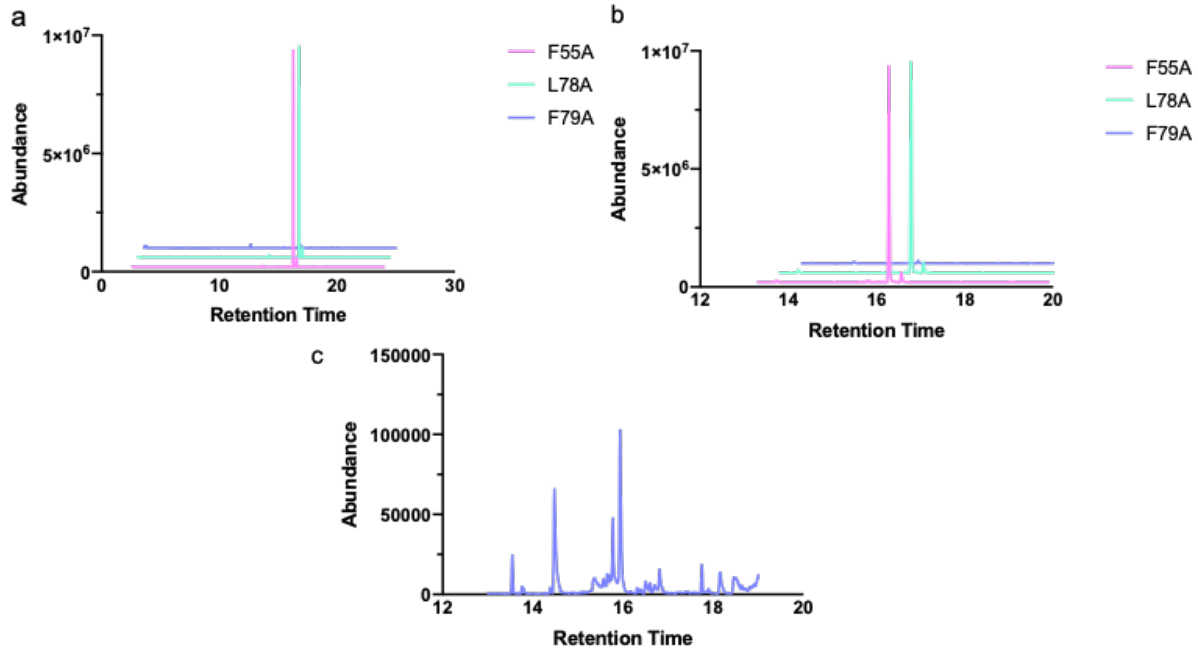


Figure 4-13 GC/MS output of terpene products of F55A, L78A and F79A variants of SeDS. *(a) The GC/MS output of product profiles of the F55A, L78A and F79A purified variants . For presentation, the data of F55A has been shifted by 200000 abundance units and the following values were shifted by 0.5 minutes and 400000 abundance units. (b) The GC/MS output of terpene products of F55A, L78A and F79A, with retention time against abundance cut between 13.3 and 19.9 minutes for clarity. F55A was shifted by 200000 abundance units and the following values were shifted by 0.5 minutes and 400000 abundance units. (A and B) 2 mM farnesyl pyrophosphate was incubated with 5 μM enzyme incubated overnight at 30℃ for F55A and L78A. Both peaks correspond to mass of 204 Da and have the same retention time as wild type. (a, b and c) The GC/MS output of F79A product profile separated, with retention time against abundance. 2 mM farnesyl pyrophosphate was incubated with 10 μM enzyme incubated overnight at 30℃. (c) F79A data cut between 13.3 and 19.9 minutes and the abundance y-axis was rescaled to see the peaks clearly.*

Figure 4-13 shows the GC/MS results confirming the culture overlay GC/MS results showing the same product profiles. Both F55A and L78A produce wild-type products in a similar ratio to the wild-type SeDS, however the level of activity is undermined as the reactions go to completion overnight unlike the kinetic analysis showing lower activity. Whereas, F79A is inactive in all three studies suggesting the mutation has a drastic effect on enzyme activity and potentially change in product ratio (Figure 4-12 B). There are additional peaks in the product profile of F79A that have masses between 200 and 300 Da but there were insufficient amounts to take this forward.

4.2.6 Alanine mutagenesis summary

Overall eight alanine variants were created at residues within the hydrophobic cleft of the active site as identified by Ligplot and pymol analysis. These variants were screened with both the malachite green assay and GC/MS methods to evaluate the protocol for further mutagenesis screening with the result displayed in Table 4-1 (Figure 4-1).

The results show that F79 is the most interesting residue tested with the alanine scanning. It has the most significant change in both activity level and products ratio (Table 4-1). This change in product profile to 50:50 (selina-4(15),7(11)-diene : germacrene B) is the most appealing change with the aim of producing novel products, nonetheless the amino acid substitution does lead to a significant loss in activity. F79A is near the magnesium ions close to the mouth of the active site (Figure 4-14) and therefore could control how the cation moves deeper within the hydrophobic cleft of the active site. Residues F55 and L78 are also close to F79 at the mouth of the active site but alanine scanning showed they do not have the same role. Mutation of phenylalanine to alanine significantly reduces the size of the amino acid, potentially increasing the volume of the active site cleft, in turn giving more room for the flexible substrate to be orientated differently in the active site. This might cause the substrate cation to be misplaced in the hydrophobic cleft resulting in reduced activity and changes in product profile as the new orientation might not favour selina-4(15),7(11)-diene over germacrene B as in wild-type SeDS. Also, this rationale suggests why there was a reduction in activity as nothing is controlling the cation to be correctly positioned alongside any catalytic residues in the hydrophobic cleft. In addition, when aligning SeDS to 5-epi-aristolochene synthase with germacrene A and pyrophosphate bound (PDB 5IK6), F79A in SeDS is found in the same position as Y520 and Y376 in 5-epi-aristolochene synthase. This suggests a bulky residue containing a benzene ring may be crucially involved in correct positioning of the cation in the hydrophobic enzyme active site (Figure 4-14).

Table 4-1 Summary table of alanine variant testing

Mutation	Rate of activity (Absorbance units min^{-1})	Percentage of activity (%)	Ratio of products (selina-4(15),7(11)-diene: germacrene B)
Wild-type	0.0097	100	99:1
F55A	0.0071	75.3	95:5
L78A	0.0035	80.3	97:3
F79A	0.0027	1.7	53:47
Y152A	0.0021	61.4	95:5
G182A	0.00043	63.6	95:5
V187A	0.0026	46.8	95:5
S217A	0.0020	65.3	94:6
I220A	0.0013	85.4	95:5

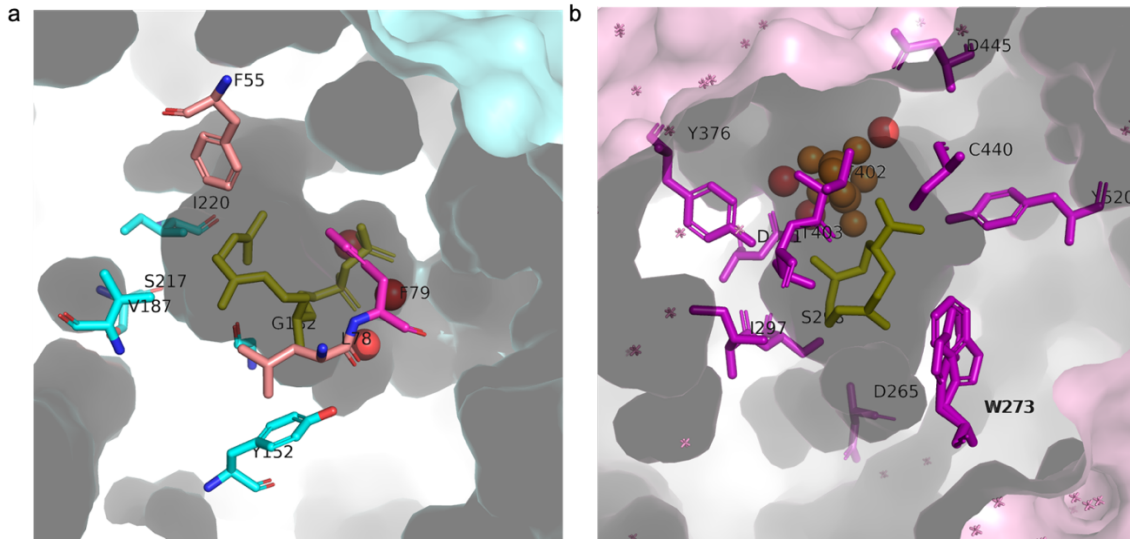


Figure 4-14 Selina-4(15),7(11)-diene synthase active site from Streptomyces pristinaespiralis ATCC 25486 (PDB 4OKZ) with F79A highlighted compared to the tobacco 5-epi-aristolochene synthase active site (PDB 5IK6). (*A) The active site of chain 1 of SeDS with magnesium ions depicted as red spheres and an un-hydrolysable farnesyl pyrophosphate mimic, dihydrofarnesyl pyrophosphate in yellow. The residues depicted as sticks coloured by elements are the key residues to coordinate the cation produced in the reaction mechanism as identified by Ligplot and by Pymol structural analysis (Wallace, Laskowski and Thornton, 1995) (PDB 4OKZ). F79 (shown in pink) appears to be closer to the magnesium ions in the active site along with F55 and L78 (shown in orange). (B) Active site of 5 epi-aristolochene synthase with key residues highlighted as sticks (from Leview analysis) (PDB 5IK6). Magnesium ions are depicted as red spheres with diphosphate depicted as orange spheres. Germacrene A is depicted as yellow sticks in the hydrophobic cleft of the active site.*

Along with identifying key residues to target, alanine scanning was used to compare two screening methods and optimise a screening protocol for further mutagenesis. The malachite green assay needs purified protein reducing the throughput of the assay. Also, the results for variant F79A were not accurate with two other methods showing the variant is near inactive. Low accuracy in this assay could be due to potential interactions with E.coli contaminant proteins not separated in the purification method used. GC/MS screening can be done with culture samples and an organic layer increasing the throughput of the screening method. The GC/MS produced reliable results with the in culture and purified protein GC/MS corroborating the kinetic characterisation. The only drawback of GC/MS screening is the 72-hour incubation required. In conclusion, the experiments demonstrate that the best protocol for initial screening is to use GC/MS to provide information on product

profile (to identify those variants with altered profiles) and to give some indication of activity level which can be further confirmed and quantified using the malachite green assay and GC/MS on purified protein.

4.3 Further mutagenesis

4.3.1 Mutant selection

Based on the results of alanine scanning and Leview analysis, positions were selected for further mutagenesis (Figure 4-15). Leview uses a similar algorithm to Ligplot to analyse residue ligand interactions (Caboche, 2013). Analysis was performed to validate the Ligplot analysis or find any additional positions to target.

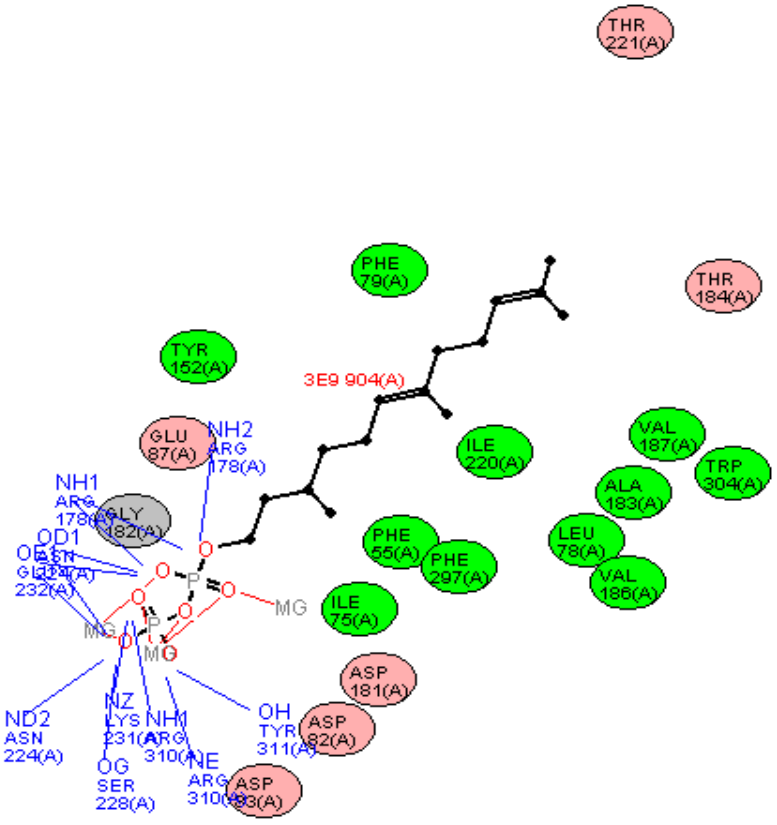


Figure 4-15 Leview analysis diagram of the active site of selina-4(15),7(11)-diene synthase from Streptomyces pristinaespiralis ATCC 25486 (PDB 4OKZ). *Selina-4(15),7(11)-diene synthase was crystallised with inactive substrate mimic dihydrofarnesyl pyrophosphate. The selina-4(15),7(11)-diene synthase active site has three magnesium ions coordinated by the NSE/DTE motif. Magnesium ions coordinate the pyrophosphate moiety of the substrate whereas the hydrophobic region of the active site binds the cation and hydrophobic substrate end with a variety of hydrophobic interactions. Residues are coloured based on their hydrophobic properties with green residues non-polar and pink residues polar.*

Using a combination of results described above and active site analysis tools, residues were selected for further mutagenesis. Leview analysis identified all residues mutated in alanine scanning, excluding G182 which was highlighted by Ligplot (Figure 4-15). G182 was tested in alanine mutagenesis and had the lowest expression level and activity level of all the variants and no significant effect on the product profile of the enzyme (Figure 4-12). Therefore, residue G182 was not taken any further in mutagenesis studies. In addition to the residues identified with Ligplot and previously mutated, additional analysis was completed with Leview software and identified two polar residues in the secondary layer of the active site, T184 and T221. Consequently, the remaining residues tested in alanine scanning were taken forward for further mutagenesis along with two polar residues further away from the active site, T184 and T221.

Both ligand protein interaction analysis software identified residue A183 which was not included in previous alanine scanning. Due to its close proximity to many other residues being screened it was not taken forward for screening but was noted for potential further screening.

Consequently, nine residues were taken forward for further mutagenesis, but with this number of residues and the screening capabilities, full saturation libraries at each residue were not practical. As a result, a set of mutation rules or criteria was created to direct the mutagenesis effort. Therefore, if any residues produce significant results further mutations can be made at this position. The criteria were designed to have the maximum effect by making the most significant changes in size, polarity and charge with the aim of changing the product profile. The general approach followed was;

- Every selected residue will be mutated to aliphatic residues (L, I or V)
- Small amino acids changed to larger bulkier amino acids of similar charge and properties.
- Bulky amino acids changed to smaller amino acids of similar charge or properties
- Non-polar amino acid changed to polar amino acid of similar size
- Polar amino acid change to non-polar amino acid of similar size
- Reverse charge of charged residues
- Remove bulky residues to smaller polar residues to allow for potential water molecules

Both F55 and F79 have previously been mutated to tryptophan, tyrosine and leucine with only leucine changing the product profile to produce small amounts of linear products including; (E)-β-farnesene and (2Z,6E)- α-farnesene (Baer, Rabe, Fischer, *et al.*, 2014). These results corroborate the alanine scanning results showing the key role of the ringed amino acids at the top of the hydrophobic cleft; however, these mutations were not repeated. Therefore, both F55 and F79 were mutated to isoleucine to see if the rearrangement of the atoms of a leucine also has an effect on the product profile. Also, mutations to glutamine and methionine maintain a similar size or space but the addition of chemical groups changes the polarity of the residue. To change the charge of the residue in the position of the natural phenylalanine while retaining a similar size, the residues were mutated to histidine and glutamic acid.

To change the polarity of the residue at L78 it was substituted by cysteine, threonine and glutamine with the additions of different chemical groups and by glutamic acid to add a negative charge. To affect the size or orientation of the residue at position 78 while maintaining the same polarity and charge it was mutated to tryptophan, glycine and isoleucine. For Y152 it was substituted by serine and threonine to change the size while still containing a hydroxyl group. To just change the polarity the ringed residue was mutated to tryptophan and phenylalanine, both benzene ring containing residues. In addition, Y152 was mutated to isoleucine, which changes both size and polarity and histidine to change the charge at this position.

For both threonine residues, T184 and T221, changes were made to hydroxyl containing residues serine and tyrosine, to change the size without affecting the polarity and charge. To affect the charge of these residues it was changed to aspartic acid, whereas to change the polarity it was substituted by valine and cysteine, removing or substituting the hydroxyl group of the side chain. To change the size while maintaining other properties, V187 was mutated to phenylalanine with a benzene ring, small residue glycine and leucine. To change the polarity V187 was changed to threonine and asparagine while it was changed to aspartic acid to alter the charge. For S217 it was replaced by valine, cysteine, asparagine and leucine to change the polarity of the residue by substituting or removing the hydroxyl group. Also, to change the size while maintaining other properties it was changed to threonine and to change the charge, to aspartic acid. Finally, I220 was changed to leucine, phenylalanine and valine to maintain properties but affect the size of the residue. To change the polarity at I220 it was changed to residues with additional chemical groups specifically threonine, glutamine and methionine. To change the charge of I220 it was mutated to glutamic acid. The mutagenesis strategy is summarised in Table 4-2.

Table 4-2 Summary table of further mutagenesis

Position	Size Change	Polarity Change	Charge Change
F55		Glutamine	Histidine
		Methionine	Glutamic acid
	Isoleucine		
L78	Tryptophan	Cysteine	Glutamic acid
	Glycine	Threonine	
	Isoleucine	Glutamine	
F79		Glutamine	Histidine
		Methionine	Glutamic acid
	Isoleucine		
Y152	Serine	Tryptophan	Histidine
	Threonine	Phenylalanine	
	Isoleucine		
T184	Serine	Valine	Aspartic acid
	Tyrosine	Cysteine	
		Asparagine	
		Leucine	
V187	Phenylalanine	Threonine	Aspartic acid
	Glycine	Asparagine	
	Leucine		
S217	Threonine	Valine	Aspartic acid
		Cysteine	
		Asparagine	
		Leucine	
I220	Leucine	Threonine	Glutamic acid
	Phenylalanine	Glutamine	
	Valine	Methionine	
T221	Serine	Valine	Aspartic acid
	Tyrosine	Cysteine	

4.3.2 GC/MS screening of variants

Variants were produced using site directed mutagenesis using the same protocol as used with the alanine variants (4.2.2). After confirming the mutations with sequencing, variants were screened using *E. coli* in culture GC/MS screening. As with previous GC/MS

screening, the data has been cut to improve visibility of terpenes and the abundance relates to both the activity of the enzyme and the expression level of the enzyme (Figure 4-10, 16-18).

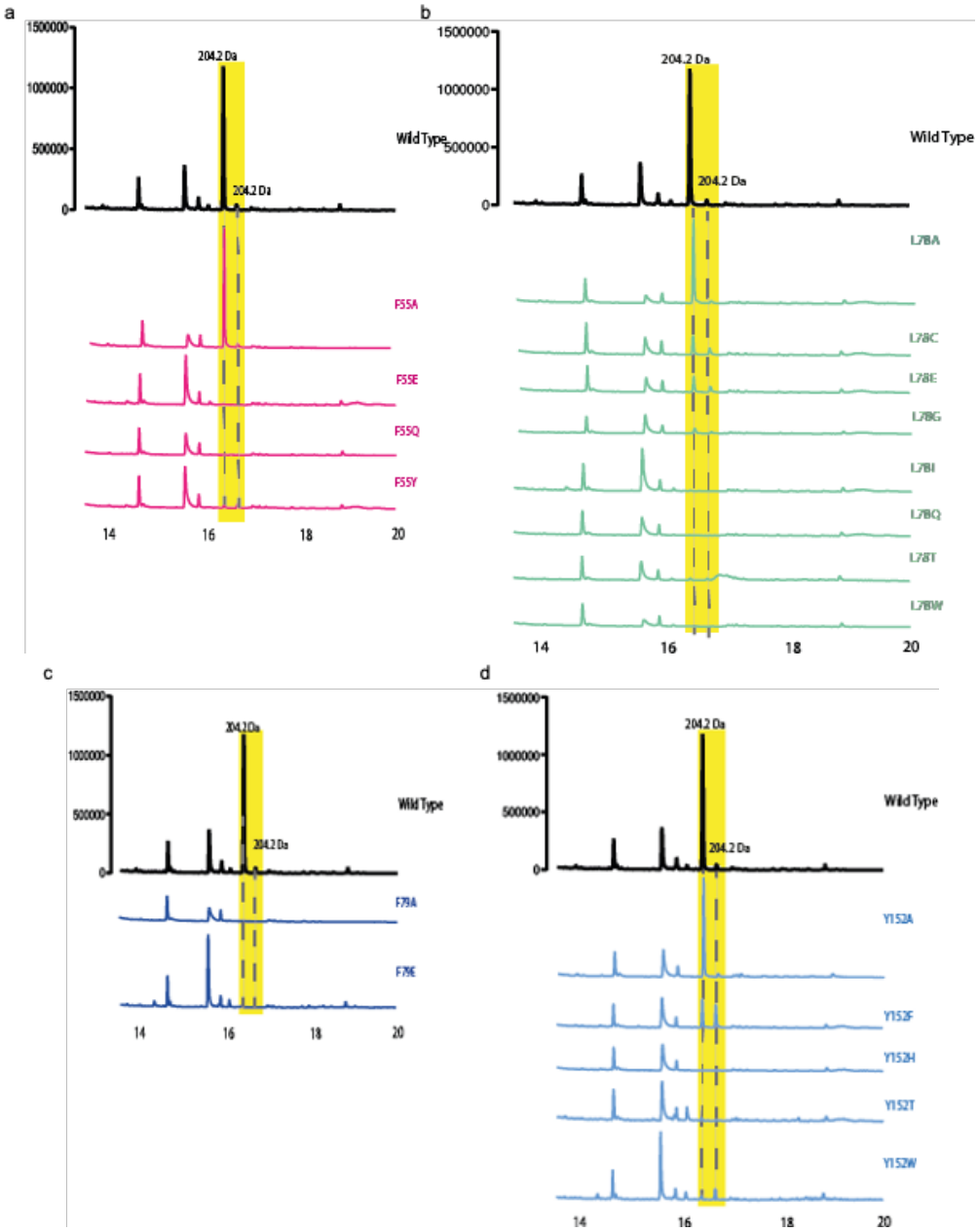


Figure 4-16 GC/MS analysis of terpene products for variant SeDSs at positions F55, L78, F79 and Y152. *GC/MS output of terpene products between 13.3 and 19.9 minutes standardised against the relative abundance of the internal standard cumene (not shown in figures). For each variant peaks between 200 and 300 Da are annotated. The yellow box signifies the selina-4(15),7(11)-diene and germacrene B peaks and the dotted lines highlight the two products. (a) The GC/MS output for the mutations at position F55. (b) The GC/MS output for mutations at position L78. (c) The GC/MS output for mutations at position F79. (d) The GC/MS output for the mutations at position Y152. No mutation shows any novel products, but some mutations show differences in product specificity (Y152W and Y152F).*

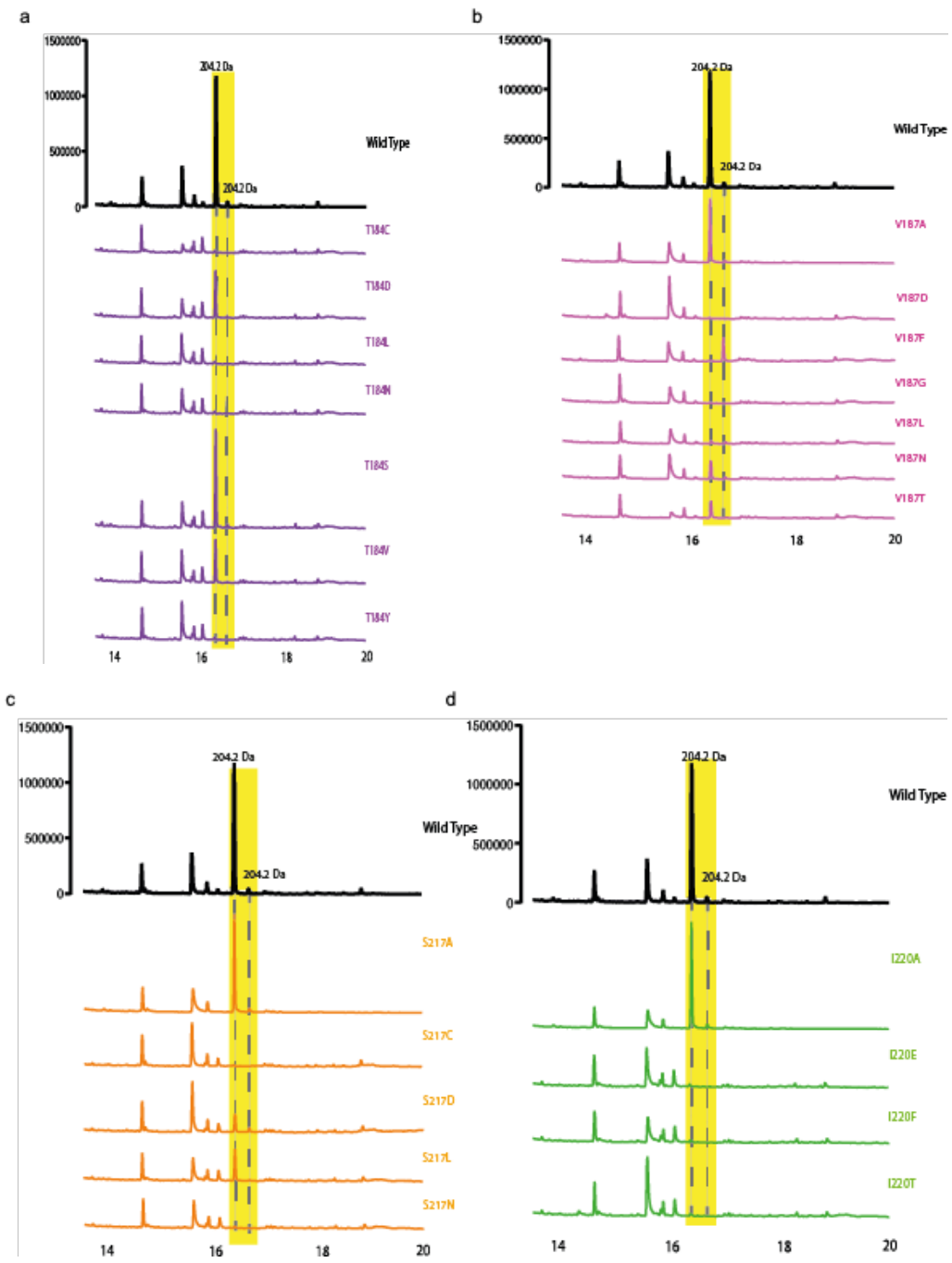


Figure 4-17 GC/MS analysis of terpene products for variant SeDSs at positions T184, V187, S217 and I220. *GC/MS output between 13.4 and 19.9 minutes standardised against the relative abundance of the internal standard cumene (not shown in figures). For each mutation and peaks between 200 and 300 Da are annotated. The yellow box signifies the selina-4(15),7(11)-diene and germacrene B peaks and the dotted lines highlight the two products. (a) The GC/MS output for the mutations at position T184. (b) The GC/MS output for mutations at position V187. (c) The GC/MS output for mutations at position S217. (d) The GC/MS output for the mutations at position I220. No mutation shows any novel products, but some mutations show differences is product specificity (V187F).*

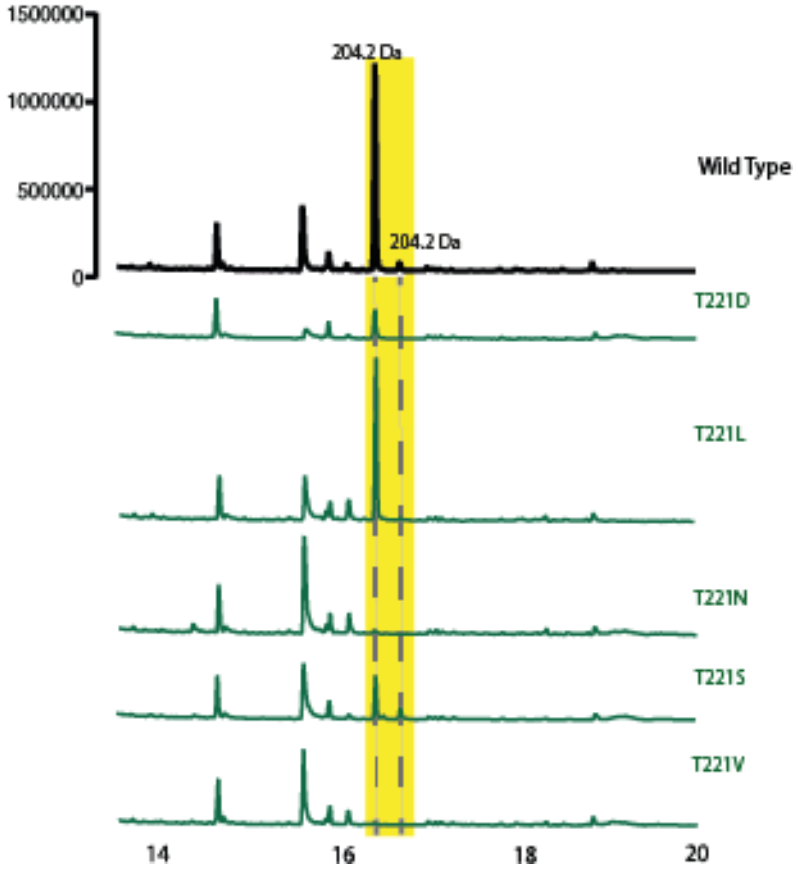


Figure 4-18 GC/MS analysis of terpene products for variant SeDSs at position T221. *GC/MS output between 13.4 and 19.9 minutes standardised against the relative abundance of the internal standard cumene (not shown in figures). For each mutation and peaks between 200 and 300 Da are annotated. The yellow box signifies the selina-4(15),7(11)-diene and germacrene B peaks and the dotted lines highlight the two products. No mutation shows any novel products.*

As shown in Figures 4-16 to 4-18, no variant produced any alternative products to a detectable level. Some variants produced altered amounts of each of the wild-type products so an alternative way to display these results is to calculate the amount of each product and plot these against each other. The area under each peak in the GC/MS profile was calculated using the data standardised with the cumene internal standard (Figure 4-19).

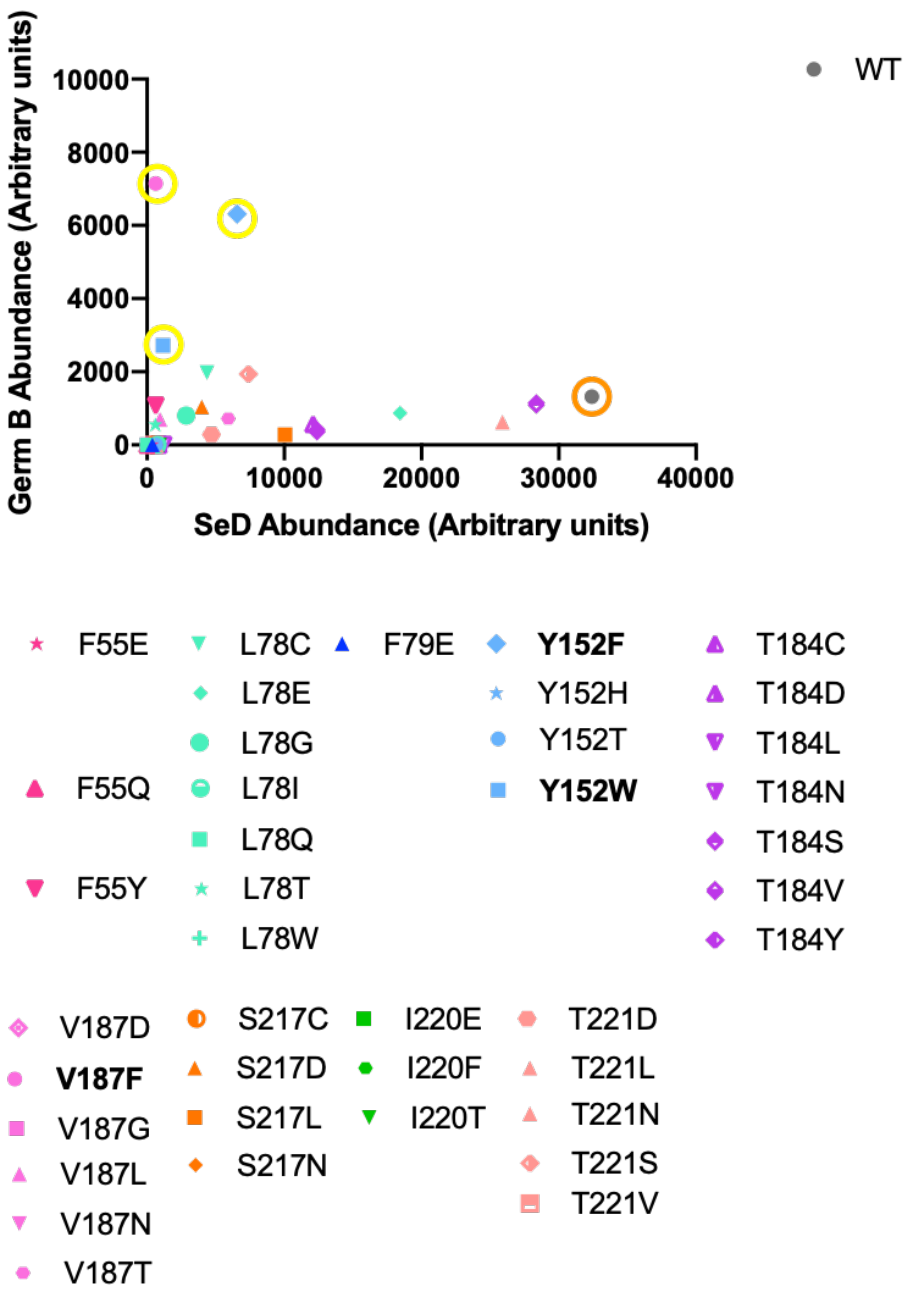


Figure 4-19 Selina-4(15),7(11)-diene and germacrene B abundance for each variant. *Graph plotting the abundance of each of the wild-type products of selina-4(15),7(11)-diene synthase for both wild-type and variant enzymes. Scatter diagram of the abundance of selina-4(15),7(11)-diene against germacrene B. The y-axis for germacrene abundance was reduced by a factor of 4 to improve the clarity. Wild-type SeDS was also included and highlighted in orange. The values circled and in bold are residues selected for the biggest change in product specificity.*

From both analysis V187F and Y152F have the most significant change in product specificity and profile. V187F shows a complete switch in specificity from selina-4(15),7(11)-diene to germacrene B (Figure 4-17 and 4-19). Y152F shows a switch to a near 50:50 ratio between

the two products (Figure 4-16 and 4-19). These variants were taken forward for further characterisation including activity assays.

In addition, other variants were selected for further characterisation, to investigate these positions and to validate the screening results. Y152W was selected for further analysis as its product profile is similar to V187F with a switch in specificity, but either has a lower activity or lower expression compared to V187F. Both alanine variants of V187 and Y152 were taken forward for further analysis of these positions. Also, T221D has a common product profile similar to most variants with either low activity or expression with some production of both products, as a way to test the screening methods with GC/MS on purified protein. Likewise, L78E was taken forward as one mutation with a similar product profile as wild-type SeDS to validate screening.

4.3.3 Variant characterisation

Variants V187F, Y152F, Y152W, Y152A, V187A, T221D and L78E were selected from GC/MS screening and taken forward for further characterisation. Each variant was expressed on a large scale and purified with wild-type methods for kinetic studies and GC/MS verification (Section 2.3.4). Initial activity tests were performed over an extended period of time due to potential activity of some variants being significantly lower based on GC/MS results. Variants were tested over 5 hours with the same standard substrate and enzyme concentration as in alanine variant testing, a concentration above the wild-type K_m (40 μM FPP) and at 500 nM enzyme. An initial rate against time was calculated for each variant in μM min^{-1}, using the standard curve, to compare all experimental tests completed on different days (Figure 4-20).

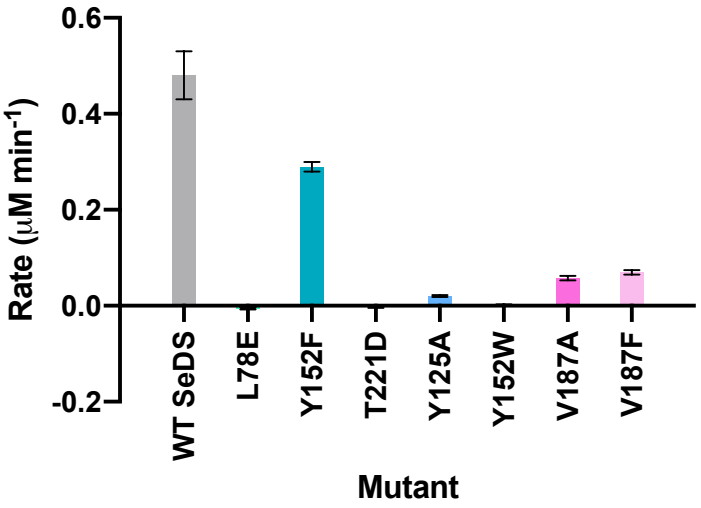


Figure 4-20 Initial activity test for variants selected for further testing. *Initial activity of each variant at 40 μM farnesyl pyrophosphate over the course of 5 hours. Each absorbance value was converted to pyrophosphate produced in μM using the standard curves performed alongside each experiment. From this an initial rate was calculated per minute to easily compare rates.*

The initial activity test showed that Y152F and V187F have the highest activity out of the variants characterised. Also, it validates the GC/MS screening results showing V187F having a higher activity to Y152W. Following this initial kinetic test, variants were analysed using GC/MS on purified enzyme. Enzyme and substrate were incubated overnight at 30 ºC before terpenes were sampled for GC/MS (Section 2.5) (Figure 4-21).

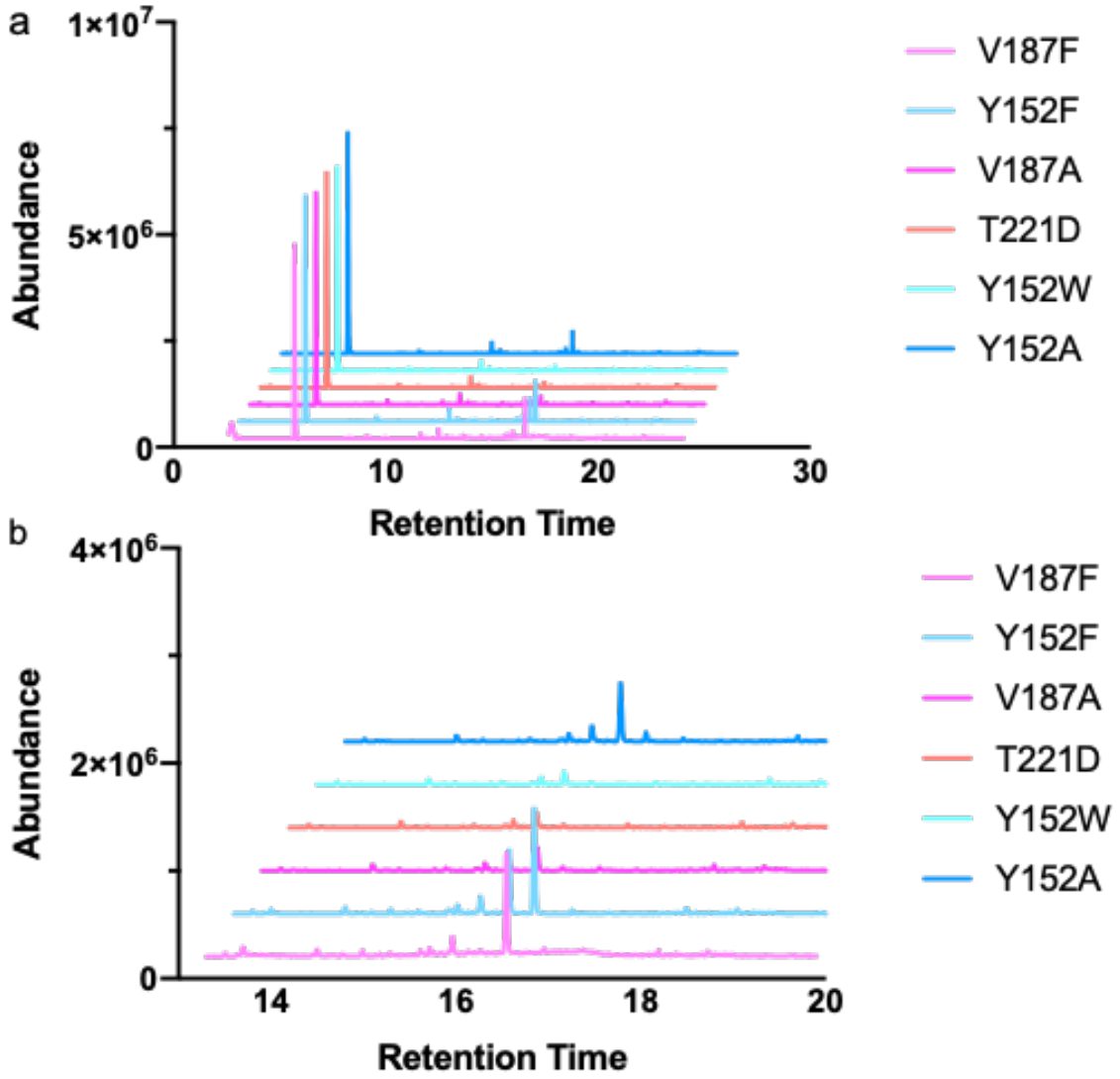


Figure 4-21 GC/MS output for V187F, Y152F V187A, T221D, Y152W and Y152A using purified protein. *(A) GC/MS output of 10 µM of enzyme and 2 mM farnesyl pyrophosphate incubated overnight at 30 ºC before terpenes were sampled by snap freezing and extraction in ethyl acetate. Each result is shifted by 200000 abundance units and all results after V187F were shifted 0.5 minutes. (B) The GC/MS output between 13.3 and 19.9 minutes. Each result is shifted by 200000 abundance units and all result after V187F were shifted 0.3 minutes.*

Figure 4-21 shows all results confirm with the culture overlay method except Y152W which shows none of the expected terpene peaks for selina-4(15),7(11)-diene and germacrene B. Y152W enzyme appears inactive in both Figure 4-20 and 4-20, suggesting the enzyme has significantly lower activity than expected. As the in-culture GC/MS screen is over 72 hours there was no way to estimate the enzyme activity level. To confirm that the peaks at the same retention time correspond to the expected products; selina-4(15),7(11)-diene and

germacrene B, GC/MS samples of V187F were spiked with wild-type SeDS samples. If the peaks do correspond to the same products, the relative abundance will increase (Figure 4-22).

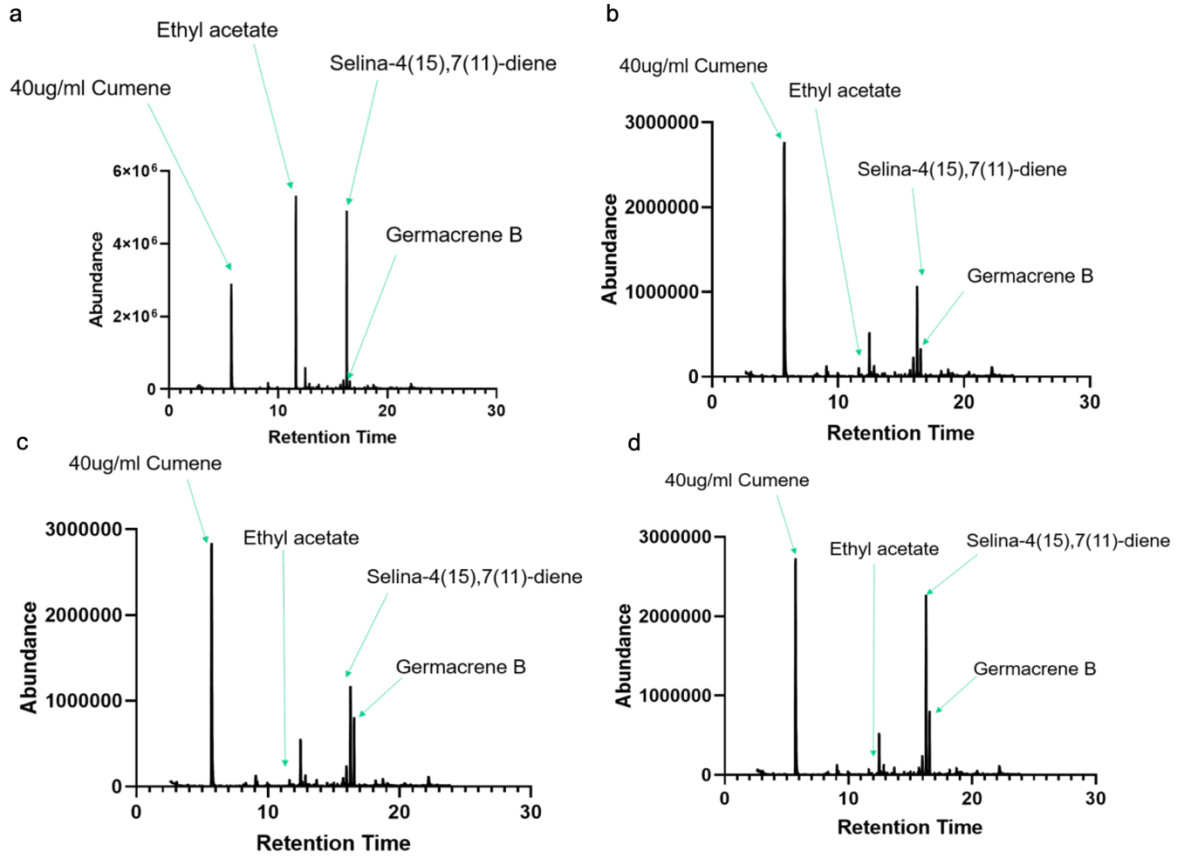


Figure 4-22 GC/MS outputs of spiking V187F samples with wild-type selina-4(15),7(11)-diene synthase products. *(a) GC/MS output of 5 µM of wild-type selina-4(15),7(11)-diene synthase and 2 mM farnesyl pyrophosphate incubated overnight at 30 ºC before terpenes were sampled by snap freezing and extraction in ethyl acetate. (b) GC/MS output of 10 µM of V187F and 2 mM farnesyl pyrophosphate incubated overnight at 30 ºC before terpenes were sampled by snap freezing and extraction in ethyl acetate. (c) GC/MS output of V187F and wild type selina-4(15),7(11)-diene synthase in a ratio of 4:1. (d) GC/MS output of V187F and wild type selina-4(15),7(11)-diene synthase in a ratio of 3:2.*

GC/MS spiking confirmed that the predicted peaks are the wild-type products; selina-4(15),7(11)-diene and germacrene B. As Figure 4-22 shows the peak increases with increased amount of wild-type GC/MS sample present further confirming the peaks to be for the same products. Following confirmation of the product profiles, Y152F and V187F were taken forward for kinetic characterisation. As with previous characterisation of alanine variants F55A and L78A, optimisation of the kinetic conditions was necessary. Using the initial kinetic tests as a starting point the first stage of optimisation used an extended time

course and increased enzyme concentration compared to wild-type SeDS kinetic conditions (Figure 4-23).

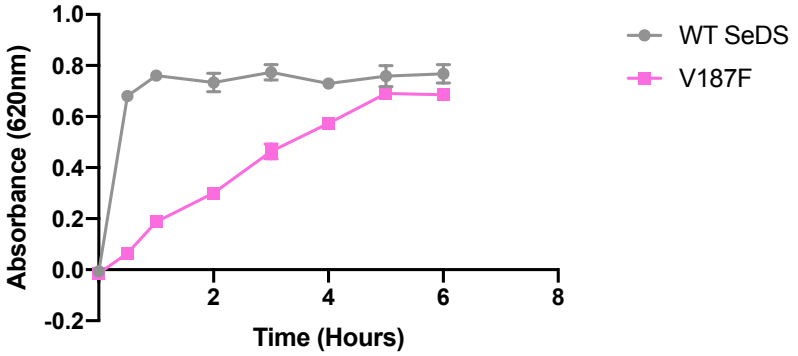


Figure 4-23 – V187F kinetic optimisation. *V187F was tested at 40 μM FPP over the course of 6 hours with 2.5 μM enzyme. The time range tested is an extended time range compared to wild-type kinetics and shows linear activity between 1 and 5 hours.*

Following initial kinetic condition tests (Figure 4-23) a 2-hour time range was selected with a range of concentrations between 5 and 80 μM FPP, which is similar to the wild-type 5 and 160 μM FPP. These optimised conditions provided the opportunity to calculate kinetic parameters (Figure 4-24). In addition, Y152F's kinetic conditions were optimised to calculate kinetic parameters by altering the substrate concentration range and the assay length.

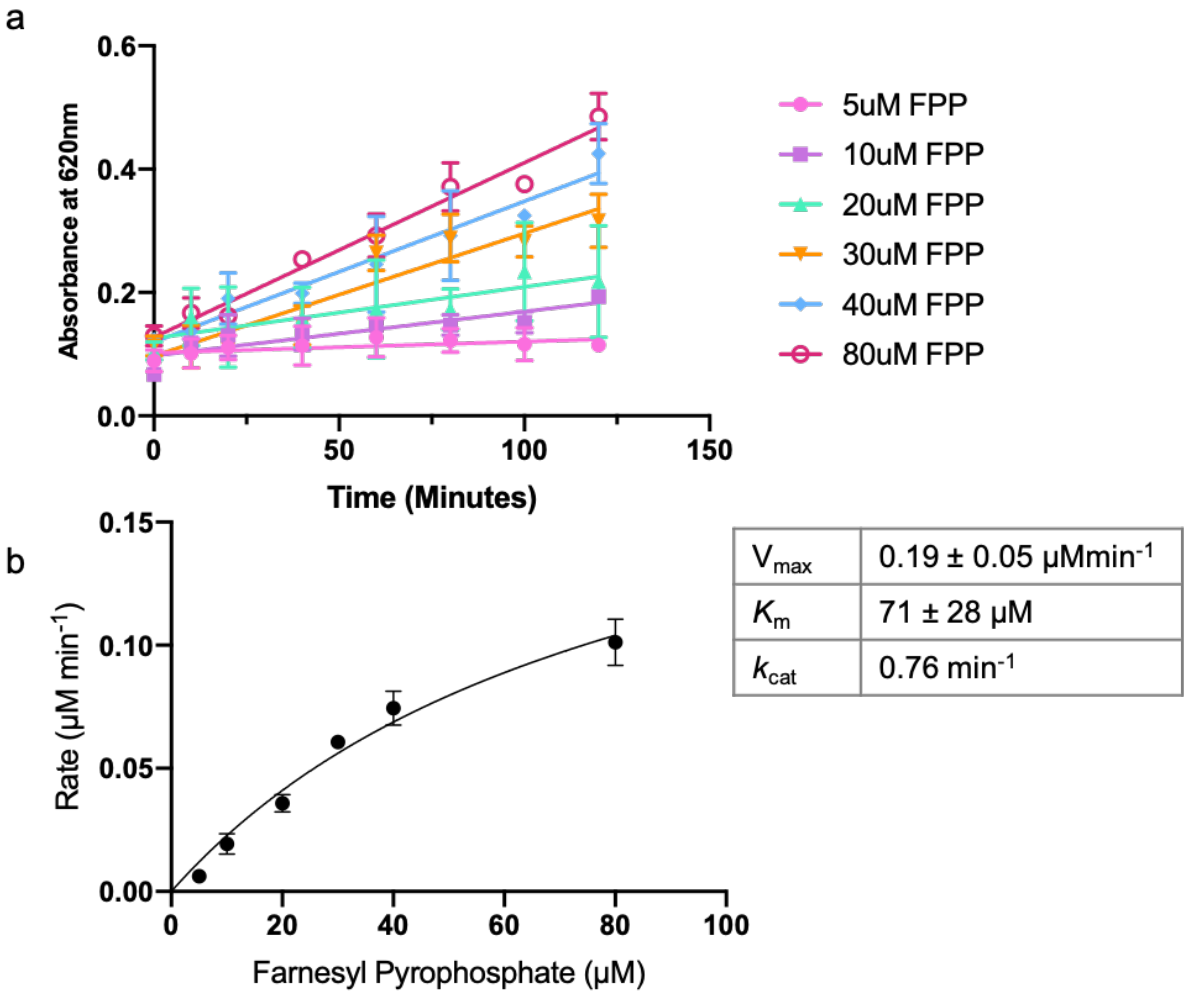
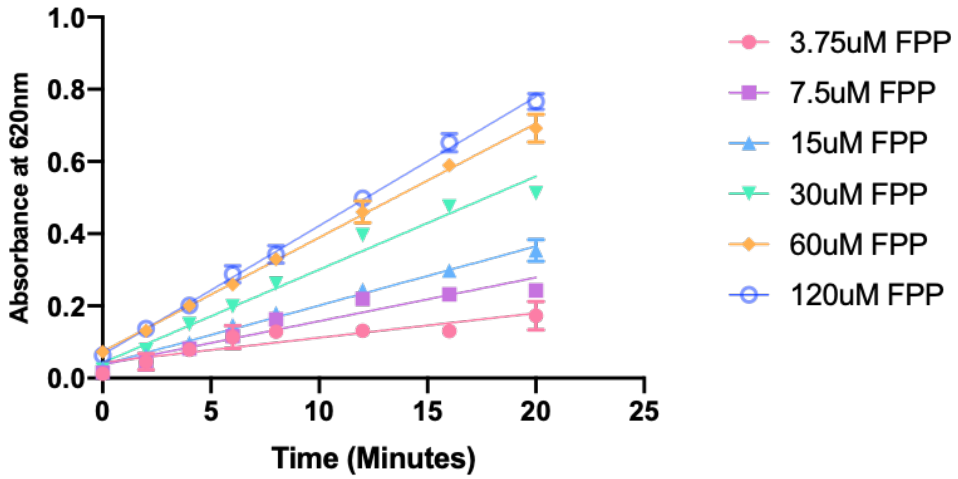
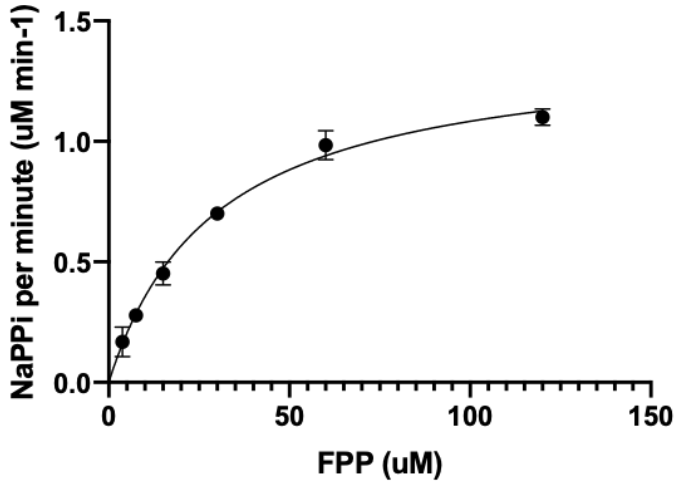


Figure 4-24 V187F kinetic characterisation. *(a) V187F at 2.5 μM was tested at 6 farnesyl pyrophosphate concentration between 5 μM and 80 μM. For each concentration the initial rate of reaction was calculated using GraphPad linear regression. (b) Michealis Menten curve using the farnesyl pyrophosphate concentration against the initial rate of reaction in pyrophosphate produced per minute with the kinetic parameters displayed in the table.*





V_{max}	1.403 ± 0.095 μMmin^{-1}
K_m	29.55 ± 5 μM
k_{cat}	2.806 min^{-1}

Figure 4-25 Y152F kinetic characterisation*. (a) Y152F at 0.5 μM was tested at 6 farnesyl pyrophosphate concentration between 3.75 μM and 120 μM. For each concentration the initial rate of reaction was calculated using GraphPad linear regression. (b) Michealis Menten curve using the farnesyl pyrophosphate concentration against the initial rate of reaction in pyrophosphate produced per minute with the kinetic parameters displayed in the table.*

The kinetic parameters of the V187F variant show a significant increase in K_m, from approximately 15 μM to 70 μM a factor of 4. (Figure 4-24). An increase in K_m suggests a reduction in affinity for the substrate which could explain the switch in specificity and relate to the stabilisation of the cation. In comparison Y152F has a K_m of approximately 2-fold higher than wild-type while also slightly increasing the V_{max} (Figure 4-25). A less significant increase in K_m could explain the change in product profile while not seeing the full switch in specificity as with V187F.

Both Y152 and V187 were detected in both ligand protein interacting softwares as residues interacting with SeDS's substrate mimic (Figure 4-2 and 4-15). Both of these residues are found within the hydrophobic cleft of the enzyme's active site and predicted to be where the cation is moved to be quenched and form the final products. Y152 is in the middle of the hydrophobic cleft opposite residue F79, however the residue is deeper into the active site with the ring angled in the same direction (Figure 4-26). Mutagenesis of F79 has previously shown product profile changes that resemble Y152 but there is a significant effect on activity, suggesting these residues have a direct effect on the formation of the final products. V187F is a distance away from the magnesium ions, deep within the active site cleft and sticks into the hydrophobic cleft (Figure 4-26).

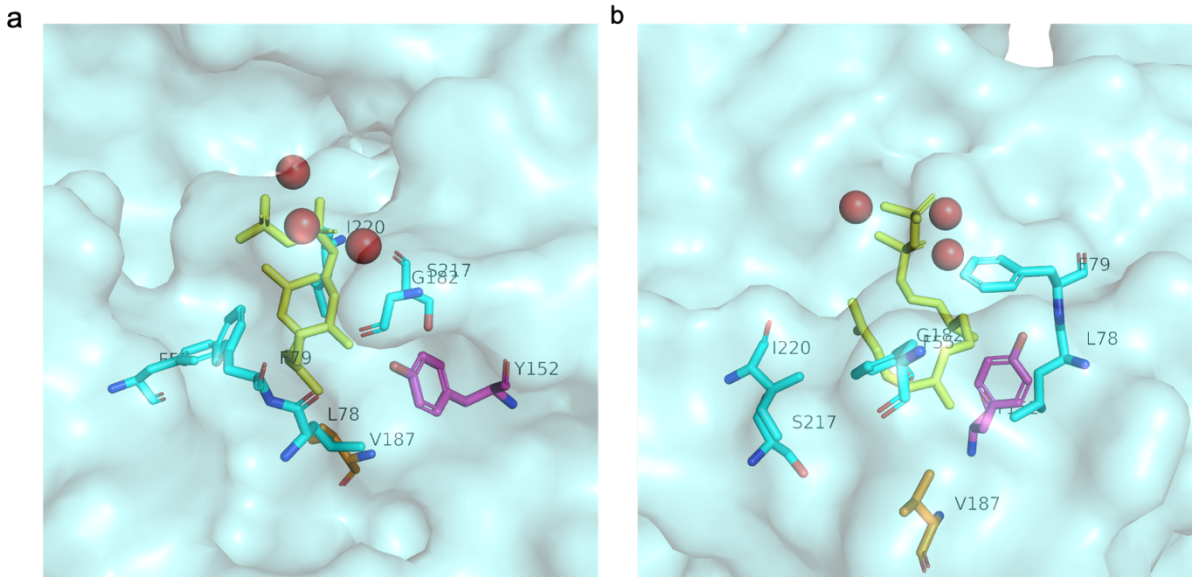


Figure 4-26 Active site image of selina-4(15),7(11)-diene synthase (4OKZ) with residues Y152 and V187 highlighted. *Surface representation of the active site of SeDS with the magnesium ions depicted as red spheres and substrate mimic dihydrofarnesyl pyrophosphate in yellow sticks. The key residues mutated in this engineering attempt are depicted as blue sticks with residue Y152 coloured pink and V187 coloured orange (A) Image shows the position and orientation of Y152 in the active site and its position in relation to F79. (B) Image shows the position and orientation of V187 in the active site and the effect a large change in size would have on the active site surface.*

Y152F is a small change in polarity with the removal of a hydroxyl group but maintains the volume with a benzene ring. This smaller change within the active site could explain why there is only a change in product distribution and not a significant change in activity. When residue F79 was mutated to benzene-containing amino acids there was a shift in product profile with the increased production of side product of germacrene B. However, when residue F79 mutagenesis removes the benzene ring, for example F79A and F79E, the

enzyme activity is lost or significantly impacted but, there is still the product shift. Therefore, the loss of this hydroxyl group shifts the product ratio to an approximate 50:50 ratio suggesting the hydroxyl group could push the cation into favouring selina-4(15),7(11)-diene production following the formation of germacrene B (Figure 3-5).

V187F is a significant change in size and volume with the addition of a benzene ring changing the contour of the hydrophobic cleft. This significant change deep within the hydrophobic cleft could potentially explain why there is a full specificity switch to germacrene B. The addition of a ring deep within the hydrophobic cleft affects the proton abstraction from germacrene B and the formation of two distinct rings in the selina-4(15),7(11)-diene pathway (Figure 3-5 and 4-26). Following the identification of these two positions further mutagenesis was suggested at these two sites including other benzene ring-containing amino acids and additional chemical groups.

4.3.4 Mutagenesis analysis

Other selinadiene synthases have been characterised from biosynthetic gene cluster analysis but there are no other crystal structures for structural comparison. However, the product profile of these enzymes has been characterised with one selinadiene synthase, δ-selinadiene synthase from *Abies alba,* which produces; δ -selinadiene, germacrene B, guaia-6,9-diene, germacrene A and δ-cadinene (Christopher L. Steele *et al.*, 1998). Germacrene products and intermediates in other selinadiene synthase supports the mechanism, however due to the low primary sequence homology between terpene synthase, no structural comparison can be made.

Other terpene synthases' also produce germacrene products that have been structurally characterised. As a result, these structures and enzyme mechanisms could be compared to SeDS and potentially explain engineering effort and specificity switch. One significant example is 5-epi-aritolochene synthase which has been crystallised with both FPP and PPi, with germacrene A, in addition to an apo structure (Figure 4-27). A structural alignment of all selina-4(15),7(11)-diene synthases and a selection of the many 5-epi-aritsolchene synthase structures (variety of ligands and apo forms) (Figure 4-28) was performed to find a structural basis for the specificity switches.

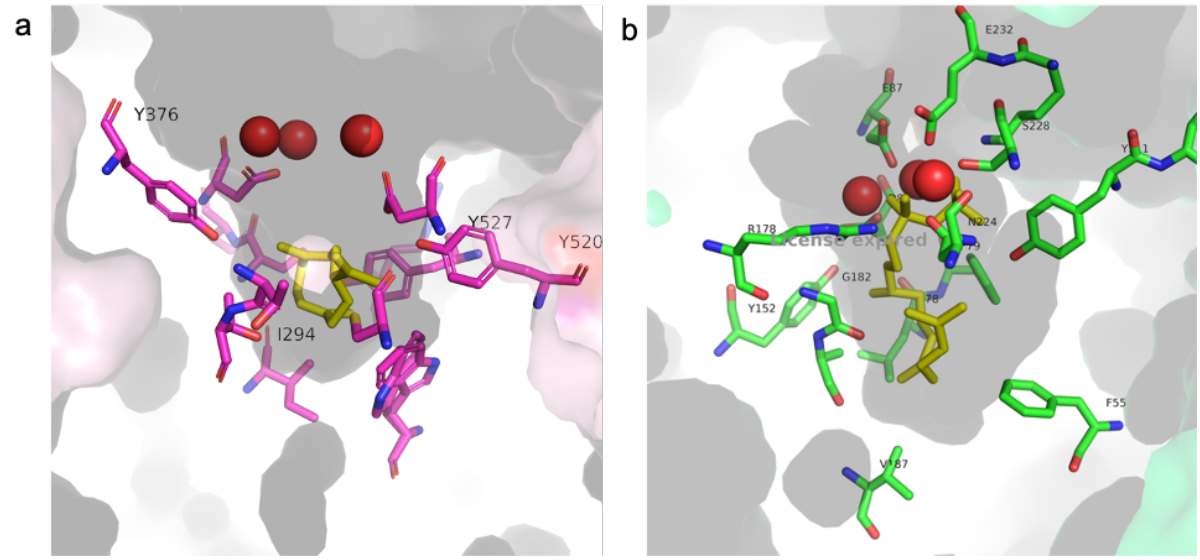


Figure 4-27 Active site image of tobacco 5-epi-aristolochene synthase (PDB 5IK6) and selina-4(15),7(11)-diene synthase (40KZ). *(a) Active site image of tobacoo-5-epi-aristolochene synthase showing the residues interacting with germacrene A as analysed by Leview. Germacrene A is displayed as yellow sticks with the three magnesium ions displayed as red dots. Resides Y527, Y367, Y520 and I294 are labelled and displayed as sticks due to their location in relation to similar residues in the SeDS active site as shown by alignment (Figure 4-28). (b) Active site image of selina-4(15),7(11)-diene synthase showing the residues interacting with farnesyl pyrophosphate substrate mimic.*

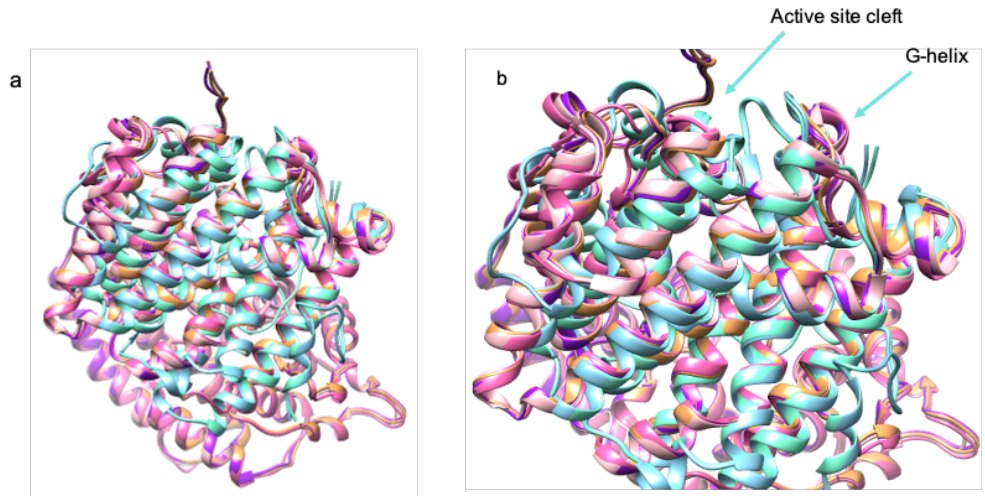


Figure 4-28 Structural alignment of selina-4(15),7(11)-diene synthase with tobacco 5-epi-aristolochene synthase (PDB 4OKZ, 4OKM, 5EAS, 5IK0, 5IK6 and 5IKA). *Structural alignment using the SALIGN web tool gave a score of 81.8 %. 4OKZ is coloured aquamarine, 4OKM is coloured blue, 5EAS is coloured pink, 5IK0 is coloured in pale pink, 5IK6 is coloured purple and 5IKA is coloured orange. (A) Structural alignment of all structures using only chain A of both 4OKM and 4OKZ. From the image the active site of all enzymes aligns well with a score of 81.8%. 5-epi-aristlochene is a significantly larger enzyme than one chain of SeDS. (B) Enlargement of the active site. Subtle differences between the two active sites specifically at the top of the enzyme active site can be seen.*

The SeDS active site (Figure 4-27 and 4-28) has three ring containing residues in the top half of the active site, F55, F79 and Y152. Engineering at both F79 and Y152 have led to changes in product profile and in activity levels of the enzyme and have been postulated to push the cation to favour the major product. Residues in similar positions are seen in 5-epi-aristolochene with three tyrosines at positions 367, 520 and 527. Potentially these bulky residues direct the cation into a germacrene like shape and the remaining residues lead to the differences in product profile in between the two enzymes. In addition, the enzyme has an isoleucine at position 294 which is similar to the valine at position 187 in SeDS. This residue allows more space in the active site and could be a key residue in allowing the formation of the two rings of epi-aritsolochene, as with the two rings of selina-4(15),7(11)-diene.

Figure 4-28 shows the structural alignment of chain A of SeDS and 5-epi-aristolochene showing that the active sites are similar with a SALIGN alignment score of 81.8% (Braberg et al., 2012). Chain A is significantly smaller that 5-epi-aristolochene potentially explaining the product differences between the two enzymes and their two ringed products, as can be

seen at the bottom of Figure 4-28a. Figure 4-28b shows an enlargement of the active site region and shows the subtle differences between the two enzymes, specifically at what is described as the G-helix in SeDS (Figure 3-3). The subtle differences between the active sites could explain the product profile differences but show how the proposed reaction mechanism goes through the same initial stage with the formation of a germacrene ion.

4.4 Summary

The aim of this project was to engineer a sesquiterpene synthase to produce novel products, while gaining insight into the mechanism of action. Initial engineering used the technique of alanine scanning to identify a set of residues to target for further engineering. From the alanine scanning results the most interesting residue was F79A with the most significant loss in activity but also the most significant change in product profile. Despite the product profile change, a 50:50 ratio of wild-type products with additional peaks, the activity is too low to be useful. In addition to the further study of F79A, both F55A and L78A were kinetically characterised due to their initial activity test results (Figure 4-6). F55A shows an increase in K_m whereas L78A shows no significant difference. Increasing the K_m shows a reduction in affinity for the substrate (Figure 4-7 and 4-8). Therefore, F55A reduces the affinity for the substrate whereas L78A affects the reaction rate. Increasing the K_m and therefore decreasing affinity for the substrate would be useful method to potentially relax product specificity as terpene synthase products are determined by the active site contour shaping the cation. With reduced substrate affinity the cation may take up new conformations consequently diversifying the product profile. However, this must be balanced with the substrate having sufficient binding affinity for a productive conformation.

Along with testing positions to target in further mutagenesis, but also tested screening methods for reliable and efficient results, both GC/MS in culture and malachite green assay were tested and compared. Overall, GC/MS screening was selected as the first initial screen as it was the most reliable, the easiest and of the two methods had the highest throughput. Following screening, a further two variants (Y152F and V187F) had the most significant change in product profile. No novel products were detected, however a selection of variants showed changes in the product distribution of the wild-type products (Figure 4-16 to 4-18). These mutations, along with other mutations with the most common result of screening, were taken forward for basic characterisation. The variants were purified and tested with the malachite green assay and with GC/MS screening (Figure 4-20 and 4-21). GC/MS product analysis from purified proteins confirmed the results of the ‘overlay screen’ (Figure 4-21 and 4-22). Y15F and V187F were taken forward for kinetic characterisation to provide further

information on how the mutation affects the reaction (Figure 4-23 and 4-24). V187F shows an increase of 4-fold in K_m showing a reduction in affinity for the substrate similar to F55A. V187F is a significant change in the active site shape reducing the available space in the enzyme active site pushing the product profile in favour of germacrene B. No novel products have been characterised from our engineering efforts, however a change in specificity has been characterised with the mutation V187F. In addition, Y152F has shown a change in product profile ratio with 50:50 ratio of the two products. Y152F shows a 2-fold increase in K_m a smaller reduction in affinity than V187F and a less significant change in active site contour and product profile.

This engineering has shown key residues in determining the product profile and also key residues for activity levels. When comparing SeDS to other enzymes that also have a mechanism that progress through a germacrene cation, or which produce germacrenes, we find that these have similar size residues in the same position in the active site. This suggests that they could be important in steps leading to the formation of the germacrene cation. These residues include three containing ring structures at the top of the active site and a smaller non-polar residue at the bottom of the hydrophobic cleft. In conclusion the engineering of SeDS was successful in changing the product profile and giving mechanistic insight, but the changes made here did not produce any novel products.

Chapter 5 – Conclusion

This thesis has described the kinetic characterisation of engineering efforts of selina-4(15),7(11)-diene synthase with the aim of producing novel products.

Chapter 1 introduces natural product terpenes, which are a broad range of chemical products from simple C5 skeletons and the enzymes that produce this chemical diversity (Harms, Kirschning and Dickschat, 2020) (Zhou and Peters, 2011) (Ajikumar *et al.*, 2008). Due to the broad range of uses for terpenes the recent focus has switched to engineering terpene synthase with the goal of increasing or decreasing specificity or expanding the range of characterised terpenes (Steiner and Schwab, 2012). The engineering of terpene synthase can be rational, semi-rational engineering and directed evolution (Rynkiewicz, Cane and Christianson, 2002) (Li *et al.*, 2014) (Lauchli *et al.*, 2014) (Salmon *et al.*, 2015). In addition to engineering for novel terpenes, terpene production pathways are engineering to increase the production of terpenes particularly in heterologous hosts (Chen, Zhang and Lindley, 2019) (Chang and Keasling, 2006). SeDS has previously been engineered with the intention of gaining mechanistic insight into its tight specificty, however it has never been engineered with the aim of producing novel products. As a result, further mutagenesis and further residue were targeted with the specific goal of producing novel terpenes form a highly specific terpene synthase.

The first results chapter describes the selection of a sesquiterpene synthase to study and take forward. The criteria were the sesquiterpene synthase must be expressed heterologously expressed in *E. coli*, have a crystal structure available and limited previous engineering. From this criteria four sesquiterpene synthase were narrowed down to four sesquiterpene synthase which were further narrowed down to *Abies grandis* bisabolene synthase and *Streptomyces pristinaespiralis* selina-4(15),7(11)-diene synthase. Following initial expression trials, selina-4(15),7(11)-diene synthase was selected as the target enzyme due to its higher expression levels. Subsequently SeDS was used to optimise the malachite green enzyme assay to kinetically characterise the enzyme and calculate the kinetic parameters. The V_{max} ranged from 0.9 to 0.4 μMmin^{-1}, the K_m ranges from 10.9 to 14.6 μM and the k_{cat} 1 to 1.8 min^{-1}. In addition, the product profile of SeDS was analysed using gas chromatography mass spectrometry on purified protein and “in culture” using and organic extraction method.

The second results chapter describes the engineering efforts of SeDS through site directed mutagenesis. Initial mutagenesis used alanine scanning at positions identified using ligand protein analysis software Ligplot+. These alanine mutants were screened using two methods to evaluate the screening throughput; malachite green assay and in culture GC/MS. Following the evaluation of the screening methods in-culture GC/MS screening was selected as the initial screening method for further mutagenesis, due to the accuracy between purified protein and expression culture samples therefore removing the need to purify samples and increasing the throughput. In addition, the alanine mutagenesis screening highlighted three positions which were further explored; F55A, L78A, F79A. All three variants were further tested in the malachite green assay to calculate kinetic parameters and GC/MS analysis. F55A's and L78A's product profile is similar to the wild-type product profile but with lower activity. Both variants' kinetic parameters were calculated with F55A having a significant (6-fold) increase in K_m whereas the K_m for L78A is within the wild-type range as with the variant's other kinetic parameters. However, F79A was inactive in both screening method conditions. As a result, no kinetic parameters were calculated for F79A and an accurate product profile could not be determined. These results show F79 has a key role within the enzyme active site and due to its position near the mouth of active site potentially directing the cation down into the depth of the hydrophobic cleft and provided a starting point for further mutagenesis.

Subsequently further mutagenesis was completed using the positions in alanine scanning and additional positions identified by the protein ligand interaction software Leview. A series of variants were produced for each position by site directed mutagenesis, based on a set of defined rules, and the variants were screened using "in culture" GC/MS. The majority of variants showed low activity or expression with a product profile resembling the wild-type profile. Two variants had interesting product profiles with V187F showing a switch specificity and Y152F's profile changing to a 50:50 product ratio of selina-4(15),7(110-diene and germacrene B. Along with additional related variants, V187F and Y152F were tested with the malachite green enzyme assay and GC/MS with purified protein. In addition, the kinetic parameters were calculated. Both variants showed an increased K_m and a reduced affinity for the substrate. V187F K_m increased by a factor of 4 while Y152F K_m increased by a factor of 2. This could explain the difference in the product profile changes with Y152F having higher binding affinity for farnesyl pyrophosphate reducing the potential orientations therefore there is only a partial switch in specificity. The differences might also be show that residue size at this position plays a more significant role in specificity at this position.

The immediate future research area would be to confirm the structures of the products by NMR spectroscopy as GC/MS analysis can only give a mass and all sesquiterpene products have the same mass. As a result, although a switch in specificity in the product distribution is observed, the identity of the products had not been elucidated. Initial work had begun on finding the concentration limit for farnesyl pyrophosphate with the spectrometer. To complete this concentration test, a sesquiterpene bisabolene standard was used and the optimal conditions, with the minimum farnesyl pyrophosphate concentration while maintaining a low background, was 100 μM with 1024 scans. Following the concentration optimisation, the sample preparation method was optimised as the product is separated from an incubated reaction in an aqueous solutions. The optimisation was to separate the products using snap freezing with an organic solvent based on methods of purification for GC-MS samples along with filtration with sodium sulfate. However, this work was unable to be completed using a reaction with purified SeDS. This confirmation would be undisputable proof the two peaks are selina-4(15),7(11)-diene and germacrene B.

In addition, further investigation of positions Y152 and V187 would benefit the engineering effort. This further engineering could improve the selectivity of the altered product distribution or alternatively lead to the further changes in activity and product profiles. This further investigation would involve rational engineering, varying additional properties, at both positions including the addition of a sulfide in methionine to alter the active site environment. These changes can change the hydrophobicity within the active site altering the orientations of the cation. Also, a number of small hydrophobic amino acids were selected to see the how the loss of bulk affects the product ratio. This work could have given further insight into the effect or importance of these positions in the control of the reaction mechanism.

In conclusion, this research has provided insights into the active site cleft of SeDS and potential key residues involved in the determination of the products. Variants of F79, Y152 and V187 have shown changes in the product profile or in activity. These positions vary within the active site, with F79A near the active site mouth while Y152 and V187 are deeper within the active site. This work can be a grounding point for further engineering in SeDS with the potential to switch specificity or produce novel products.

Chapter 6 - Bibliography

Aaron, J. A., Lin, X., Cane, D. E. and Christianson, D. W. (2010) 'Structure of Epi-Isozizaene Synthase from Streptomyces coelicolor A3(2), a Platform for New Terpenoid Cyclization Templates', *Biochemistry*, 49(8), pp. 1787–1797. doi: 10.1021/bi902088z.

Acebes, S., Fernandez-Fueyo, E., Monza, E., Lucas, M. F., Almendral, D., Ruiz-Dueñas, F. J., Lund, H., Martinez, A. T. and Guallar, V. (2016) 'Rational Enzyme Engineering Through Biophysical and Biochemical Modeling', *ACS Catalysis*, 6(3), pp. 1624–1629. doi: 10.1021/acscatal.6b00028.

Agilent Technologies (2013) *QuickChange Primer Design*. Available at: http://www.genomics.agilent.com/primerDesignProgram.jsp.

Ajikumar, P. K., Tyo, K., Carlsen, S., Mucha, O., Phon, T. H. and Stephanopoulos, G. (2008) 'Terpenoids: Opportunities for Biosynthesis of Natural Product Drugs Using Engineered Microorganisms', *Mol. Pharm*, 5(2), pp. 167–190. doi: 10.1021/mp700151b.

Ajikumar, P. K., Xiao, W.-H., Tyo, K. E. J., Wang, Y., Simeon, F., Leonard, E., Mucha, O., Phon, T. H., Pfeifer, B. and Stephanopoulos, G. (2010) 'Isoprenoid Pathway Optimization for Taxol Precursor Overproduction in Escherichia coli', *Science (New York, N.Y.)*, 330(6000), pp. 70–74. doi: 10.1126/science.1191652.

Baer, P., Rabe, P., Citron, C. A., de Oliveira Mann, C. C., Kaufmann, N., Groll, M. and Dickschat, J. S. (2014) 'Hedycaryol Synthase in Complex with Nerolidol Reveals Terpene Cyclase Mechanism', *ChemBioChem*, 15(2), pp. 213–216. doi: 10.1002/cbic.201300708.

Baer, P., Rabe, P., Fischer, K., Citron, C. A., Klapschinski, T. A., Groll, M. and Dickschat, J. S. (2014) 'Induced-fit mechanism in class i terpene cyclases', *Angewandte Chemie - International Edition*, 53(29), pp. 7652–7656. doi: 10.1002/anie.201403648.

Baginski, E. S., Epstein, E. and Zak, B. (1975) 'Review of phosphate methodologies', *Annals of Clinical & Laboratory Science*, 5(5), pp. 399–416. Available at: http://www.annclinlabsci.org/content/5/5/399.abstract.

Baykov, A. A., Evtushenko, O. A. and Avaeva, S. M. (1988) 'A malachite green procedure for orthophosphate determination and its use in alkaline phosphatase-based enzyme immunoassay', *Analytical Biochemistry*, 171(2), pp. 266–270. doi: 10.1016/0003-2697(88)90484-8.

Bernardini, S., Tiezzi, A., Laghezza Masci, V. and Ovidi, E. (2018) 'Natural products for human health: an historical overview of the drug discovery approaches', *Natural Product Research*. Taylor and Francis Ltd., pp. 1926–1950. doi: 10.1080/14786419.2017.1356838.

Blank, P. N., Barrow, G. H. and Christianson, D. W. (2019) 'Crystal structure of F95Q epi-isozizaene synthase, an engineered sesquiterpene cyclase that generates biofuel precursors β- and γ-curcumene', *Journal of Structural Biology*, 207(2), pp. 218–224. doi:

10.1016/j.jsb.2019.05.011.
Blank, P. N., Barrow, G. H., W Chou, W. K., Duan, L., Cane, D. E. and Christianson, D. W. (2017) 'Substitution of Aromatic Residues with Polar Residues in the Active Site Pocket of epi-Isozizaene Synthase Leads to the Generation of New Cyclic Sesquiterpenes', *Biochemistry*, 56, pp. 5798–5811. doi: 10.1021/acs.biochem.7b00895.
Blank, P. N., Shinsky, S. A. and Christianson, D. W. (2019) 'Structure of Sesquisabinene Synthase 1, a Terpenoid Cyclase That Generates a Strained [3.1.0] Bridged-Bicyclic Product', *ACS Chemical Biology*, 14(5), pp. 1011–1019. doi: 10.1021/acschembio.9b00218.
Bohlmann, J., Crock, J., Jetter, R. and Croteau, R. (1998) 'Terpenoid-based defenses in conifers: cDNA cloning, characterization, and functional expression of wound-inducible (E)-α-bisabolene synthase from grand fir (Abies grandis)', *Proceedings of the National Academy of Sciences of the United States of America*, 95(12), pp. 6756–6761. Available at: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC22624/.
Boufridi, A. and Quinn, R. J. (2018) 'Harnessing the Properties of Natural Products', *Annual Review of Pharmacology and Toxicology Annu. Rev. Pharmacol. Toxicol*, 58, pp. 451–70. doi: 10.1146/annurev-pharmtox.
Braberg, H., Webb, B. M., Tjioe, E., Pieper, U., Sali, A. and Madhusudhan, M. S. (2012) 'SALIGN: a web server for alignment of multiple protein sequences and structures', *Bioinformatics*, 28(15), pp. 2072–2073. doi: 10.1093/bioinformatics/bts302.
Brown, D. G. and Boström, J. (2018) 'Where Do Recent Small Molecule Clinical Development Candidates Come From?', *Journal of Medicinal Chemistry*, 61(21), pp. 9442–9468. doi: 10.1021/acs.jmedchem.8b00675.
Byres, E., Alphey, M. S., Smith, T. K. and Hunter, W. N. (2007) 'Crystal Structures of Trypanosoma brucei and Staphylococcus aureus Mevalonate Diphosphate Decarboxylase Inform on the Determinants of Specificity and Reactivity', *Journal of Molecular Biology*, 371(2), pp. 540–553. doi: https://doi.org/10.1016/j.jmb.2007.05.094.
Caboche, S. (2013) 'LeView: Automatic and interactive generation of 2D diagrams for biomacromolecule/ligand interactions', *Journal of Cheminformatics*, 5(8), p. 40. doi: 10.1186/1758-2946-5-40.
Cane, D. E., He, X., Kobayashi, S., Omura, S. and Ikeda, H. (2006) 'Geosmin Biosynthesis in Streptomyces avermitilis. Molecular Cloning, Expression, and Mechanistic Study of the Germacradienol/Geosmin Synthase', *J Antibiot*, 59(8), pp. 471–479. Available at: http://dx.doi.org/10.1038/ja.2006.66.
Cane, D. E. and Xue, Q. (1996) 'Trichodiene Synthase. Enzymatic Formation of Multiple Sesquiterpenes by Alteration of the Cyclase Active Site', *Journal of the American Chemical Society*, 118(6), pp. 1563–1564. doi: 10.1021/ja953751n.
Cane, D. E., Xue, Q. and Fitzsimons, B. C. (1996) 'Trichodiene Synthase. Probing the Role

of the Highly Conserved Aspartate-Rich Region by Site-Directed Mutagenesis', *Biochemistry*, 35(38), pp. 12369–12376. doi: 10.1021/bi961344y.
Carlson, E. E. (2010) 'Natural Products as Chemical Probes', *ACS Chemical Biology*, 5(7), p. 639,653. doi: 10.1021/cb100105c.
Carter, S. G. and Karl, D. W. (1982) 'Inorganic phosphate assay with malachite green: an improvement and evaluation.', *Journal of biochemical and biophysical methods*, 7(1), pp. 7–13. Available at: http://www.ncbi.nlm.nih.gov/pubmed/7153458 (Accessed: 11 July 2018).
Caruthers, J. M., Kang, I., Rynkiewicz, M. J., Cane, D. E. and Christianson, D. W. (2000) 'Crystal Structure Determination of Aristolochene Synthase from the Blue Cheese Mold, Penicillium roqueforti', *Journal of Biological Chemistry*, 275(33), pp. 25533–25539. doi: 10.1074/jbc.M000433200.
Cascon, O., Touchet, S., Miller, D. J., Gonzalez, V., Faraldos, J. A. and Allemann, R. K. (2012) 'Chemoenzymatic preparation of germacrene analogues', *Chemical Communications*, 48(78), pp. 9702–9704. doi: 10.1039/c2cc35542f.
Centeno-Leija, S., Tapia-Cabrera, S., Guzmán-Trampe, S., Esquivel, B., Esturau-Escofet, N., Tierrafría, V. H., Rodríguez-Sanoja, R., Zárate-Romero, A., Stojanoff, V., Rudiño-Piñera, E., Sánchez, S. and Serrano-Posada, H. (2019) 'The structure of (E)-biformene synthase provides insights into the biosynthesis of bacterial bicyclic labdane-related diterpenoids', *Journal of Structural Biology*, 207(1), pp. 29–39. doi: 10.1016/j.jsb.2019.04.010.
Chang, M. C. Y. and Keasling, J. D. (2006) 'Production of isoprenoid pharmaceuticals by engineered microbes', *Nat Chem Biol*, 2(12), pp. 674–681. Available at: http://dx.doi.org/10.1038/nchembio836.
Chen, M., Al-lami, N., Janvier, M., D'Antonio, E. L., Faraldos, J. A., Cane, D. E., Allemann, R. K. and Christianson, D. W. (2013) 'Mechanistic Insights from the Binding of Substrate and Carbocation Intermediate Analogues to Aristolochene Synthase', *Biochemistry*, 52(32), pp. 5441–5453. doi: 10.1021/bi400691v.
Chen, M., Chou, W. K. W., Al-Lami, N., Faraldos, J. A., Allemann, R. K., Cane, D. E. and Christianson, D. W. (2016) 'Probing the Role of Active Site Water in the Sesquiterpene Cyclization Reaction Catalyzed by Aristolochene Synthase', *Biochemistry*, 55(20), pp. 2864–2874. doi: 10.1021/acs.biochem.6b00343.
Chen, X., Zhang, C. and Lindley, N. D. (2019) 'Metabolic engineering strategies for sustainable terpenoid flavour and fragrance synthesis', *Journal of Agricultural and Food Chemistry*, p. acs.jafc.9b06203. doi: 10.1021/acs.jafc.9b06203.
Christianson, D. W. (2008) 'Unearthing the roots of the terpenome', *Current Opinion in Chemical Biology*, 12(2), pp. 141–150. doi: http://dx.doi.org/10.1016/j.cbpa.2007.12.008.
Christianson, D. W. (2017) 'Structural and Chemical Biology of Terpenoid Cyclases'. doi: 10.1021/acs.chemrev.7b00287.

Citron, C. A., Gleitzmann, J., Laurenzano, G., Pukall, R. and Dickschat, J. S. (2012) 'Terpenoids are Widespread in Actinomycetes: A Correlation of Secondary Metabolism and Genome Data', *ChemBioChem*, 13(2), pp. 202–214. doi: 10.1002/cbic.201100641.
Cogan, E. B., Birrell, G. B. and Griffith, O. H. (1999) *A Robotics-Based Automated Assay for Inorganic and Organic Phosphates*. Available at: http://www.idealibrary.com (Accessed: 19 September 2018).
Colquitt, R. B., Colquhoun, D. A. and Thiele, R. H. (2011) 'In silico modelling of physiologic systems', *Best Practice and Research: Clinical Anaesthesiology*. Bailliere Tindall Ltd, pp. 499–510. doi: 10.1016/j.bpa.2011.08.006.
Dar, R. A., Shahnawaz, M., Rasool, S., Qazi, P. H. and Qazi, H. (2017) *Natural product medicines: A literature update*, *The Journal of Phytopharmacology*. Available at: www.phytopharmajournal.com (Accessed: 23 March 2020).
Das, S., Dixit, M. and Major, D. T. (2016) 'First principles model calculations of the biosynthetic pathway in selinadiene synthase', *Bioorganic and Medicinal Chemistry*, 24(20), pp. 4867–4870. doi: 10.1016/j.bmc.2016.07.002.
Degenhardt, J., Köllner, T. G. and Gershenzon, J. (2009) 'Monoterpene and sesquiterpene synthases and the origin of terpene skeletal diversity in plants', *Phytochemistry*, 70(15–16), pp. 1621–1637. doi: http://dx.doi.org/10.1016/j.phytochem.2009.07.030.
DeJong, J. M., Liu, Y., Bollon, A. P., Long, R. M., Jennewein, S., Williams, D. and Croteau, R. B. (2006) 'Genetic engineering of taxol biosynthetic genes in Saccharomyces cerevisiae', *Biotechnology and Bioengineering*, 93(2), pp. 212–224. doi: 10.1002/bit.20694.
Dewick, P. M. (2011) *Medicinal Natural Products: A Biosynthetic Approach*. 3rd edn. John Wiley & Sons, Ltd. Available at: https://books.google.co.uk/books?id=A4zptjOJfKQC&printsec=frontcover&dq=natural+products&hl=en&sa=X&ved=0ahUKEwjLms3v2o3oAhU1Q0EAHQLNAicQ6AEIKTAA#v=onepage&q=natural products&f=false (Accessed: 9 March 2020).
Dickschat, J. S. (2017) 'Modern Aspects of Isotopic Labellings in Terpene Biosynthesis', *European Journal of Organic Chemistry*, 2017(33), pp. 4872–4882. doi: 10.1002/ejoc.201700482.
Dixit, M., Weitman, M., Gao, J. and Major, D. T. (2017) 'Chemical Control in the Battle against Fidelity in Promiscuous Natural Product Biosynthesis: The Case of Trichodiene Synthase', *ACS Catalysis*, 7(1), pp. 812–818. doi: 10.1021/acscatal.6b02584.
Dougherty, D. A. (1996) 'Cation-π Interactions in Chemistry and Biology: A New View of Benzene, Phe, Tyr, and Trp', *Science*, 271(5246), pp. 163–168. doi: 10.1126/science.271.5246.163.
Driller, R., Janke, S., Fuchs, M., Warner, E., Mhashal, A. R., Major, D. T., Christmann, M., Brück, T. and Loll, B. (2018) 'Towards a comprehensive understanding of the structural

dynamics of a bacterial diterpene synthase during catalysis', *Nature Communications*, 9(1), p. 3971. doi: 10.1038/s41467-018-06325-8.
Edgar, S., Li, F.-S., Qiao, K., Weng, J.-K. and Stephanopoulos, G. (2017) 'Engineering of Taxadiene Synthase for Improved Selectivity and Yield of a Key Taxol Biosynthetic Intermediate', *ACS Synthetic Biology*, 6(2), pp. 201–205. doi: 10.1021/acssynbio.6b00206.
Engels, B., Dahm, P. and Jennewein, S. (2008) 'Metabolic engineering of taxadiene biosynthesis in yeast as a first step towards Taxol (Paclitaxel) production', *Metabolic Engineering*, 10(3–4), pp. 201–206. doi: 10.1016/j.ymben.2008.03.001.
Ernst, P., Zosel, F., Reichen, C., Nettels, D., Schuler, B. and Plückthun, A. (2020) 'Structure-guided design of a peptide lock for modular peptide binders', *ACS Chemical Biology*, 15(2), pp. 457–468. doi: 10.1021/acschembio.9b00928.
Faraldos, J. A., Antonczak, A. K., González, V., Fullerton, R., Tippmann, E. M. and Allemann, R. K. (2011) 'Probing Eudesmane Cation−π Interactions in Catalysis by Aristolochene Synthase with Non-canonical Amino Acids', *Journal of the American Chemical Society*, 133(35), pp. 13906–13909. doi: 10.1021/ja205927u.
Faraldos, J. A., Gonzalez, V., Li, A., Yu, F., Köksal, M., Christianson, D. W. and Allemann, R. K. (2012) 'Probing the Mechanism of 1,4-Conjugate Elimination Reactions Catalyzed by Terpene Synthases', *Journal of the American Chemical Society*, 134(51), pp. 20844–20848. doi: 10.1021/ja311022s.
Faraldos, J. A., Gonzalez, V., Senske, M. and Allemann, R. K. (2011) 'Templating effects in aristolochene synthase catalysis: elimination versus cyclisation', *Organic & Biomolecular Chemistry*, 9(20), pp. 6920–6923. doi: 10.1039/c1ob06184d.
Felicetti, B. and Cane, D. E. (2004) 'Aristolochene Synthase: Mechanistic Analysis of Active Site Residues by Site-Directed Mutagenesis', *Journal of the American Chemical Society*, 126(23), pp. 7212–7221. doi: 10.1021/ja0499593.
Feng, J., Chen, Y., Pu, J., Yang, X., Zhang, C., Zhu, S., Zhao, Y., Yuan, Y., Yuan, H. and Liao, F. (2011) 'An improved malachite green assay of phosphate: Mechanism and application', *Analytical Biochemistry*, 409(1), pp. 144–149. doi: 10.1016/J.AB.2010.10.025.
Fish, P. V and Johnson, W. S. (1994) *The First Examples of Nonenzymic, Biomimetic Polyene Pentacyclizations. Total Synthesis of the Pentacyclic Triterpenoid*, *Tetrahedron Lett*. UTC. Available at: https://pubs.acs.org/sharingguidelines (Accessed: 12 May 2020).
Forcat, S. and Allemann, R. K. (2006) 'Stabilisation of transition states prior to and following eudesmane cation in aristolochene synthase', *Organic & Biomolecular Chemistry*, 4(13), pp. 2563–2567. doi: 10.1039/b604147g.
Fujihashi, M., Sato, T., Tanaka, Y., Yamamoto, D., Nishi, T., Ueda, D., Murakami, M., Yasuno, Y., Sekihara, A., Fuku, K., Shinada, T. and Miki, K. (2018) 'Crystal structure and functional analysis of large-terpene synthases belonging to a newly found subclass',

Chemical Science, 9(15), pp. 3754–3758. doi: 10.1039/c8sc00289d.
Fujisawa, M., Harada, H., Kenmoku, H., Mizutani, S. and Misawa, N. (2010) 'Cloning and characterization of a novel gene that encodes (S)-β-bisabolene synthase from ginger, Zingiber officinale', *Planta*, 232(1), pp. 121–130. doi: 10.1007/s00425-010-1137-6.
Gao, Y., Honzatko, R. B. and Peters, R. J. (2012) 'Terpenoid synthase structures: a so far incomplete view of complex catalysis', *Natural Product Reports*, 29(10), pp. 1153–1175. doi: 10.1039/c2np20059g.
Geladopoulos, T. P., Sotiroudis, T. G. and Evangelopoulos, A. E. (1991) 'A malachite green colorimetric assay for protein phosphatase activity', *Analytical Biochemistry*, 192(1), pp. 112–116. doi: http://dx.doi.org/10.1016/0003-2697(91)90194-X.
Gennadios, H. A., Gonzalez, V., Di Costanzo, L., Li, A., Yu, F., Miller, D. J., Allemann, R. K. and Christianson, D. W. (2009) 'Crystal Structure of (+)-δ-Cadinene Synthase from Gossypium arboreum and Evolutionary Divergence of Metal Binding Motifs for Catalysis', *Biochemistry*, 48(26), pp. 6175–6183. doi: 10.1021/bi900483b.
Gonzalez, V., Grundy, D. J., Faraldos, J. A. and Allemann, R. K. (2016) 'The amino-terminal segment in the [small beta]-domain of [small delta]-cadinene synthase is essential for catalysis', *Organic & Biomolecular Chemistry*, 14(31), pp. 7451–7454. doi: 10.1039/c6ob01398h.
Gonzalez, V., Touchet, S., Grundy, D. J., Faraldos, J. A. and Allemann, R. K. (2014) 'Evolutionary and Mechanistic Insights from the Reconstruction of α-Humulene Synthases from a Modern (+)-Germacrene A Synthase', *Journal of the American Chemical Society*, 136(41), pp. 14505–14512. doi: 10.1021/ja5066366.
Goodman, H. M., Abelson, J., Landy, A., Brenner, S. and Smith, J. D. (1968) 'Amber suppression: A Nucleotide change in the anticodon of a tyrosine transfer RNA', *Nature*, 217(5133), pp. 1019–1024. doi: 10.1038/2171019a0.
Grundy, D. J., Chen, M., González, V., Leoni, S., Miller, D. J., Christianson, D. W. and Allemann, R. K. (2016) 'Mechanism of Germacradien-4-ol Synthase-Controlled Water Capture', *Biochemistry*, 55(14), pp. 2112–2121. doi: 10.1021/acs.biochem.6b00115.
Guo, C., Hu, Y., Yang, C., Nanjaraj Urs, A. N. and Zhang, Y. (2018) 'Developing a colorimetric assay for Fe(II)/2-oxoglutarate-dependent dioxygenase', *Analytical Biochemistry*, 548, pp. 109–114. doi: 10.1016/J.AB.2018.02.013.
Haines, B. E., Wiest, O. and Stauffacher, C. V (2013) 'The Increasingly Complex Mechanism of HMG-CoA Reductase', *Accounts of Chemical Research*, 46(11), pp. 2416–2426. doi: 10.1021/ar3003267.
Harms, V., Kirschning, A. and Dickschat, J. S. (2020) 'Nature-driven approaches to non-natural terpene analogues.', *Natural product reports*. doi: 10.1039/c9np00055k.
Harris, G. G., Lombardi, P. M., Pemberton, T. A., Matsui, T., Weiss, T. M., Cole, K. E.,

Köksal, M., Murphy, F. V, Vedula, L. S., Chou, W. K. W., Cane, D. E. and Christianson, D. W. (2015) 'Structural Studies of Geosmin Synthase, a Bifunctional Sesquiterpene Synthase with αα Domain Architecture That Catalyzes a Unique Cyclization–Fragmentation Reaction Sequence', *Biochemistry*, 54(48), pp. 7142–7155. doi: 10.1021/acs.biochem.5b01143.
Hartley, R. D. and Fawcett, C. H. (1969) 'The separation and identification of selina-4(14),7(11)-diene, a new sesquiterpene from hops (humulus lupulus)', *Phytochemistry*, 8(3), pp. 637–643. doi: 10.1016/S0031-9422(00)85414-X.
Hong, Y. J. and Tantillo, D. J. (2015) 'Feasibility of Intramolecular Proton Transfers in Terpene Biosynthesis – Guiding Principles', *Journal of the American Chemical Society*, 137(12), pp. 4134–4140. doi: 10.1021/ja512685x.
Hua, G., Hu, Y., Yang, C., Liu, D., Mao, Z., Zhang, L. and Zhang, Y. (2018) 'Characterization of santalene synthases using an inorganic pyrophosphatase coupled colorimetric assay', *Analytical Biochemistry*, 547, pp. 26–36. doi: 10.1016/J.AB.2018.02.002.
Huang, Q., Roessner, C. A., Croteau, R. and Scott, A. I. (2001) 'Engineering Escherichia coli for the synthesis of taxadiene, a key intermediate in the biosynthesis of taxol', *Bioorganic & Medicinal Chemistry*, 9(9), pp. 2237–2242. doi: http://dx.doi.org/10.1016/S0968-0896(01)00072-4.
Hyatt, D. C., Youn, B., Zhao, Y., Santhamma, B., Coates, R. M., Croteau, R. B. and Kang, C. (2007) 'Structure of limonene synthase, a simple model for terpenoid cyclase catalysis', *Proceedings of the National Academy of Sciences of the United States of America*, 104(13), pp. 5360–5365. doi: 10.1073/pnas.0700915104.
Ignea, C., Pontini, M., Motawia, M. S., Maffei, M. E., Makris, A. M. and Kampranis, S. C. (2018) 'Synthesis of 11-carbon terpenoids in yeast using protein and metabolic engineering', *Nature Chemical Biology*, 14(12), pp. 1090–1098. doi: 10.1038/s41589-018-0166-5.
Islam, M. M., Barik, S., Preeyanka, N. and Sarkar, M. (2020) 'Interaction of Lysozyme with Monocationic and Dicationic Ionic Liquids: Toward Finding a Suitable Medium for Biomacromolecules', *Journal of Physical Chemistry B*, 124(6), pp. 961–973. doi: 10.1021/acs.jpcb.9b10270.
Itaya, K. and Ui, M. (1966) 'A new micromethod for the colorimetric determination of inorganic phosphate', *Clinica Chimica Acta*, 14(3), pp. 361–366. doi: http://dx.doi.org/10.1016/0009-8981(66)90114-8.
Jeffreys, D. (2004) *Aspirin : the remarkable story of a wonder drug*. 1st edn. Bloomsbury.
Jiang, J., He, X. and Cane, D. E. (2006) 'Geosmin Biosynthesis. Streptomyces coelicolor Germacradienol/Germacrene D Synthase Converts Farnesyl Diphosphate to Geosmin', *Journal of the American Chemical Society*, 128(25), pp. 8128–8129. doi: 10.1021/ja062669x.
Jin, Y., Liu, Z., Li, Y., Liu, W., Tao, Y. and Wang, G. (2016) 'A structural and functional study

on the 2-C-methyl-D-erythritol-4-phosphate cytidyltransferase (IspD) from Bacillus subtilis', *Scientific Reports*, 6, p. 36379. doi: 10.1038/srep36379 https://www.nature.com/articles/srep36379#supplementary-information.
Johnson, K. A. and Goody, R. S. (2011) 'The original Michaelis constant: Translation of the 1913 Michaelis-Menten Paper', *Biochemistry*, 50(39), pp. 8264–8269. doi: 10.1021/bi201284u.
Kemp, L. E., Bond, C. S. and Hunter, W. N. (2002) 'Structure of 2C-methyl-D-erythritol 2,4-cyclodiphosphate synthase: an essential enzyme for isoprenoid biosynthesis and target for antimicrobial drug development', *Proc Natl Acad Sci U S A*, 99(10), pp. 6591–6596.
Kennepohl, D. (2016) *27.5 Terpenoids*. LibreTexts: LibreTexts. Available at: https://chem.libretexts.org/LibreTexts/Athabasca_University/Chemistry_360%3A_Organic_Chemistry_II/Chapter_27%3A_Biomolecules%3A_Lipids/27.05_Terpenoids.
Köksal, M., Jin, Y., Coates, R. M., Croteau, R. and Christianson, D. W. (2011) 'Taxadiene synthase structure and evolution of modular architecture in terpene biosynthesis', *Nature*, 469(7328), pp. 116–122. doi: 10.1038/nature09628.
Kracht, O. N., Correia Cordeiro, R. S., Håkansson, M., Stockmann, J., Sander, D., Bandow, J., Senges, C. H. R., Logan, D. T. and Kourist, R. (2019) 'Discovery of three novel sesquiterpene synthases from Streptomyces chartreusis NRRL 3882 and crystal structure of an α-eudesmol synthase', *Journal of Biotechnology*, 297, pp. 71–77. doi: 10.1016/j.jbiotec.2019.03.006.
Lanzetta, P. A., Alvarez, L. J., Reinach, P. S. and Candia, O. A. (1979) 'An improved assay for nanomole amounts of inorganic phosphate', *Analytical Biochemistry*, 100(1), pp. 95–97. doi: 10.1016/0003-2697(79)90115-5.
Laskowski, R. A. and Swindells, M. B. (2011) 'LigPlot+: Multiple ligand-protein interaction diagrams for drug discovery', *Journal of Chemical Information and Modeling*, 51(10), pp. 2778–2786. doi: 10.1021/ci200227u.
Lauchli, R., Pitzer, J., Kitto, R. Z., Kalbarczyk, K. Z. and Rabe, K. S. (2014) 'Improved selectivity of an engineered multi-product terpene synthase', *Organic & Biomolecular Chemistry*, 12(23), pp. 4013–4020. doi: 10.1039/c4ob00479e.
Leferink, N. G. H., Dunstan, M. S., Hollywood, K. A., Swainston, N., Currin, A., Jervis, A. J., Takano, E. and Scrutton, N. S. (2019) 'An automated pipeline for the screening of diverse monoterpene synthase libraries', *Scientific Reports*, 9(1). doi: 10.1038/s41598-019-48452-2.
Leonard, E., Ajikumar, P. K., Thayer, K., Xiao, W.-H., Mo, J. D., Tidor, B., Stephanopoulos, G. and Prather, K. L. J. (2010) 'Combining metabolic and protein engineering of a terpenoid biosynthetic pathway for overproduction and selectivity control', *Proceedings of the National Academy of Sciences of the United States of America*, 107(31), pp. 13654–13659. doi: 10.1073/pnas.1006138107.

Lesburg, C. A., Zhai, G., Cane, D. E. and Christianson, D. W. (1997) 'Crystal Structure of Pentalenene Synthase: Mechanistic Insights on Terpenoid Cyclization Reactions in Biology', *Science*, 277(5333), pp. 1820–1824. doi: 10.1126/science.277.5333.1820.
Li, R., Chou, W. K. W., Himmelberger, J. A., Litwin, K. M., Harris, G. G., Cane, D. E. and Christianson, D. W. (2014) 'Reprogramming the Chemodiversity of Terpenoid Cyclization by Remolding the Active Site Contour of epi-Isozizaene Synthase', *Biochemistry*, 53(7), pp. 1155–1168. doi: 10.1021/bi401643u.
Little, D. B. and Croteau, R. B. (2002) 'Alteration of product formation by directed mutagenesis and truncation of the multiple-product sesquiterpene synthases δ-selinene synthase and γ-humulene synthase', *Archives of Biochemistry and Biophysics*, 402(1), pp. 120–135. doi: http://dx.doi.org/10.1016/S0003-9861(02)00068-1.
Lodeiro, S., Segura, M. J. R., Stahl, M., Schulz-Gasch, T. and Matsuda, S. P. T. (2004) 'Oxidosqualene Cyclase Second-Sphere Residues Profoundly Influence the Product Profile', *ChemBioChem*, 5(11), pp. 1581–1585. doi: 10.1002/cbic.200400086.
Maehama, T., Taylor, G. S., Slama, J. T. and Dixon, J. E. (2000) 'A Sensitive Assay for Phosphoinositide Phosphatases', *Analytical Biochemistry*, 279(2), pp. 248–250. doi: 10.1006/ABIO.2000.4497.
Mahmoud, S. S. and Croteau, R. B. (2003) 'Menthofuran regulates essential oil biosynthesis in peppermint by controlling a downstream monoterpene reductase', *Proceedings of the National Academy of Sciences of the United States of America*, 100(SUPPL. 2), pp. 14481–14486. doi: 10.1073/pnas.2436325100.
Major, D. T. (2017) 'Electrostatic Control of Chemistry in Terpene Cyclases', *ACS Catalysis*, 7(8), pp. 5461–5465. doi: 10.1021/acscatal.7b01328.
Major, D. T. and Weitman, M. (2012) 'Electrostatically Guided Dynamics—The Root of Fidelity in a Promiscuous Terpene Synthase?', *Journal of the American Chemical Society*, 134(47), pp. 19454–19462. doi: 10.1021/ja308295p.
Martin, V. J. J., Pitera, D. J., Withers, S. T., Newman, J. D. and Keasling, J. D. (2003) 'Engineering a mevalonate pathway in Escherichia coli for production of terpenoids', *Nat Biotech*, 21(7), pp. 796–802. doi: http://www.nature.com/nbt/journal/v21/n7/suppinfo/nbt833_S1.html.
McAndrew, R. P., Peralta-Yahya, P. P., DeGiovanni, A., Pereira, J. H., Hadi, M. Z., Keasling, J. D. and Adams, P. D. (2011) 'Structure of a Three-Domain Sesquiterpene Synthase: A Prospective Target for Advanced Biofuels Production', *Structure*, 19(12), pp. 1876–1884. doi: http://dx.doi.org/10.1016/j.str.2011.09.013.
Miller, D. J. and Allemann, R. K. (2012) 'Sesquiterpene synthases: Passive catalysts or active players?', *Nat. Prod. Rep.*, 29(1), pp. 60–71. doi: 10.1039/C1NP00060H.
Miller, D. J., Yu, F., Knight, D. W. and Allemann, R. K. (2009) '6- and 14-Fluoro farnesyl

diphosphate: mechanistic probes for the reaction catalysed by aristolochene synthase', *Organic & Biomolecular Chemistry*, 7(5), pp. 962–975. doi: 10.1039/b817194g.
Morehouse, B. R., Kumar, R. P., Matos, J. O., Olsen, S. N., Entova, S. and Oprian, D. D. (2017) 'Functional and Structural Characterization of a (+)-Limonene Synthase from Citrus sinensis', *Biochemistry*, 56(12), pp. 1706–1715. doi: 10.1021/acs.biochem.7b00143.
Newman, D. J. and Cragg, G. M. (2016) 'Natural Products as Sources of New Drugs from 1981 to 2014', *Journal of Natural Products*. American Chemical Society, pp. 629–661. doi: 10.1021/acs.jnatprod.5b01055.
Niwa, H., Ewens, C. A., Tsang, C., Yeung, H. O., Zhang, X. and Freemont, P. S. (2012) 'The role of the N-domain in the ATPase activity of the mammalian AAA ATPase p97/VCP.', *The Journal of biological chemistry*, 287(11), pp. 8561–70. doi: 10.1074/jbc.M111.302778.
Novoradovsky, A., Zhang, V., Ghosh, M., Hogrefe, H., Sorge, J. A. and Gaasterland, T. (2005) *Computational Principles of Primer Design for Site Directed Mutagenesis*. Anaheim. Available at: www.nsti.org, (Accessed: 15 April 2020).
O'Brien, T. E., Bertolani, S. J., Tantillo, D. J. and Siegel, J. B. (2016) 'Mechanistically informed predictions of binding modes for carbocation intermediates of a sesquiterpene synthase reaction', *Chemical Science*, 7(7), pp. 4009–4015. doi: 10.1039/C6SC00635C.
O'Brien, T. E., Bertolani, S. J., Zhang, Y., Siegel, J. B. and Tantillo, D. J. (2018) 'Predicting Productive Binding Modes for Substrates and Carbocation Intermediates in Terpene Synthases—Bornyl Diphosphate Synthase As a Representative Case', *ACS Catalysis*, pp. 3322–3330. doi: 10.1021/acscatal.8b00342.
O'Maille, P. E., Bakhtina, M. and Tsai, M. D. (2002) 'Structure-based combinatorial protein engineering (SCOPE)', *Journal of Molecular Biology*, 321(4), pp. 677–691. doi: 10.1016/S0022-2836(02)00675-7.
Peralta-Yahya, P. P., Ouellet, M., Chan, R., Mukhopadhyay, A., Keasling, J. D. and Lee, T. S. (2011) 'Identification and microbial production of a terpene-based advanced biofuel', *Nature Communications*, 2(1), p. 483. doi: 10.1038/ncomms1494.
Phelan, R. M., Sekurova, O. N., Keasling, J. D. and Zotchev, S. B. (2015) 'Engineering Terpene Biosynthesis in Streptomyces for Production of the Advanced Biofuel Precursor Bisabolene', *ACS Synthetic Biology*, 4(4), pp. 393–399. doi: 10.1021/sb5002517.
Potter, K., Criswell, J., Zi, J., Stubbs, A. and Peters, R. J. (2014) 'Novel product chemistry from mechanistic analysis of ent-copalyl diphosphate synthases from plant hormone biosynthesis', *Angewandte Chemie - International Edition*, 53(28), pp. 7198–7202. doi: 10.1002/anie.201402911.
Prisic, S., Xu, J., Coates, R. M. and Peters, R. J. (2007) 'Probing the Role of the DXDD Motif in Class II Diterpene Cyclases', *ChemBioChem*, 8(8), pp. 869–874. doi: 10.1002/cbic.200700045.

PubChem (2019) *(Z)-alpha-Bisabolene | C15H24 - PubChem*, *PubChem*.
Rabe, P., Citron, C. A. and Dickschat, J. S. (2013) 'Volatile Terpenes from Actinomycetes: A Biosynthetic Study Correlating Chemical Analyses to Genome Data', *ChemBioChem*, 14(17), pp. 2345–2354. doi: 10.1002/cbic.201300329.
Rabe, P., Klapschinski, T. A. and Dickschat, J. S. (2016) 'Position-Specific Mass Shift Analysis: A Systematic Method for Investigating the EI-MS Fragmentation Mechanism of epi -Isozizaene', *ChemBioChem*, 17(14), pp. 1333–1337. doi: 10.1002/cbic.201600237.
Rabe, P., Rinkel, J., Klapschinski, T. A., Barra, L. and Dickschat, J. S. (2016) 'A method for investigating the stereochemical course of terpene cyclisations', *Organic & Biomolecular Chemistry*, 14(1), pp. 158–164. doi: 10.1039/C5OB01998B.
Rabe, P., Rinkel, J., Nubbemeyer, B., Köllner, T. G., Chen, F. and Dickschat, J. S. (2016) 'Terpene Cyclases from Social Amoebae', *Angewandte Chemie International Edition*, 55(49), pp. 15420–15423. doi: 10.1002/anie.201608971.
Richard, S. B., Ferrer, J. L., Bowman, M. E., Lillo, A. M., Tetzlaff, C. N., Cane, D. E. and Noel, J. P. (2002) 'Structure and mechanism of 2-C-methyl-D-erythritol 2,4-cyclodiphosphate synthase. An enzyme in the mevalonate-independent isoprenoid biosynthetic pathway', *Journal of Biological Chemistry*, 277(10), pp. 8667–8672.
Rinkel, J., Rabe, P., zur Horst, L. and Dickschat, J. S. (2016) 'A detailed view on 1,8-cineol biosynthesis by Streptomyces clavuligerus', *Beilstein Journal of Organic Chemistry*, 12(1), pp. 2317–2324. doi: 10.3762/bjoc.12.225.
Ro, D.-K., Paradise, E. M., Ouellet, M., Fisher, K. J., Newman, K. L., Ndungu, J. M., Ho, K. A., Eachus, R. A., Ham, T. S., Kirby, J., Chang, M. C. Y., Withers, S. T., Shiba, Y., Sarpong, R. and Keasling, J. D. (2006) 'Production of the antimalarial drug precursor artemisinic acid in engineered yeast', *Nature*, 440(7086), pp. 940–943. doi: http://www.nature.com/nature/journal/v440/n7086/suppinfo/nature04640_S1.html.
Roberts, S. C. (2007) 'Production and engineering of terpenoids in plant cell culture', *Nature Chemical Biology*, 3(7), pp. 387–395. Available at: http://dx.doi.org/10.1038/nchembio.2007.8.
Robichon, C., Luo, J., Causey, T. B., Benner, J. S. and Samuelson, J. C. (2011) 'Engineering Escherichia coli BL21(DE3) derivative strains to minimize E. coli protein contamination after purification by immobilized metal affinity chromatography.', *Applied and environmental microbiology*, 77(13), pp. 4634–46. doi: 10.1128/AEM.00119-11.
Rodrigues, T., Reker, D., Schneider, P. and Schneider, G. (2016) 'Counting on natural products for drug design', *Nature Chemistry*. Nature Publishing Group, pp. 531–541. doi: 10.1038/nchem.2479.
Rodriguez, S., Kirby, J., Denby, C. M. and Keasling, J. D. (2014) 'Production and quantification of sesquiterpenes in Saccharomyces cerevisiae, including extraction,

detection and quantification of terpene products and key related metabolites', *Nature Protocols*, 9(8), pp. 1980–1996. doi: 10.1038/nprot.2014.132.
Rudolf, J. D., Dong, L. Bin, Cao, H., Hatzos-Skintges, C., Osipiuk, J., Endres, M., Chang, C. Y., Ma, M., Babnigg, G., Joachimiak, A., Phillips, G. N. and Shen, B. (2016) 'Structure of the ent-Copalyl Diphosphate Synthase PtmT2 from Streptomyces platensis CB00739, a Bacterial Type II Diterpene Synthase', *Journal of the American Chemical Society*, 138(34), pp. 10905–10915. doi: 10.1021/jacs.6b04317.
Rynkiewicz, M. J., Cane, D. E. and Christianson, D. W. (2001) 'Structure of trichodiene synthase from Fusarium sporotrichioides provides mechanistic inferences on the terpene cyclization cascade', *Proceedings of the National Academy of Sciences of the United States of America*. 2001/11/08, 98(24), pp. 13543–13548. doi: 10.1073/pnas.231313098.
Rynkiewicz, M. J., Cane, D. E. and Christianson, D. W. (2002) 'X-ray Crystal Structures of D100E Trichodiene Synthase and Its Pyrophosphate Complex Reveal the Basis for Terpene Product Diversity', *Biochemistry*, 41(6), pp. 1732–1741. doi: 10.1021/bi011960g.
Salmon, M., Laurendon, C., Vardakou, M., Cheema, J., Defernez, M., Green, S., Faraldos, J. A. and O'Maille, P. E. (2015) 'Emergence of terpene cyclization in Artemisia annua', *Nature Communications*, 6, p. 6143. doi: 10.1038/ncomms7143 http://www.nature.com/articles/ncomms7143#supplementary-information.
Scott, D. P., Yang, T., Valerie, S. and Andrew, D. M. (2010) 'A Universal, Fully Automated High Throughput Screening Assay for Pyrophosphate and Phosphate Release from Enzymatic Reactions', *Combinatorial Chemistry & High Throughput Screening*, 13(1), pp. 27–38. doi: http://dx.doi.org/10.2174/138620710790218203.
Seemann, M., Zhai, G., Umezawa, K. and Cane, D. (1999) 'Pentalenene Synthase. Histidine-309 Is Not Required for Catalytic Activity', *Journal of the American Chemical Society*, 121(3), pp. 591–592. doi: 10.1021/ja983657h.
Shatton, J. B., Ward, C., Williams, A. and Weinhouse, S. (1983) 'A microcolorimetric assay of inorganic pyrophosphatase', *Analytical Biochemistry*, 130(1), pp. 114–119. doi: 10.1016/0003-2697(83)90657-7.
Shiba, Y., Paradise, E. M., Kirby, J., Ro, D.-K. and Keasling, J. D. (2007) 'Engineering of the pyruvate dehydrogenase bypass in Saccharomyces cerevisiae for high-level production of isoprenoids', *Metabolic Engineering*, 9(2), pp. 160–168. doi: http://dx.doi.org/10.1016/j.ymben.2006.10.005.
Shishova, E. Y., Di Costanzo, L., Cane, D. E. and Christianson, D. W. (2007) 'X-ray Crystal Structure of Aristolochene Synthase from Aspergillus terreus and Evolution of Templates for the Cyclization of Farnesyl Diphosphate', *Biochemistry*, 46(7), pp. 1941–1951. doi: 10.1021/bi0622524.
Shishova, E. Y., Yu, F., Miller, D. J., Faraldos, J. A., Zhao, Y., Coates, R. M., Allemann, R.

K., Cane, D. E. and Christianson, D. W. (2008) ‘X-ray Crystallographic Studies of Substrate Binding to Aristolochene Synthase Suggest a Metal Ion Binding Sequence for Catalysis’, *Journal of Biological Chemistry*, 283(22), pp. 15431–15439. doi: 10.1074/jbc.M800659200.
Steele, Christopher L, Crock, J., Bohlmann, J. and Croteau, R. (1998) ‘Sesquiterpene Synthase from Grand Fir (Abies grandis)’, *The Journal of Biological Chemistry*, 273(4), pp. 2078–2089. Available at: http://www.jbc.org/content/273/4/2078.short.
Steele, Christopher L., Crock, J., Bohlmann, J. and Croteau, R. (1998) ‘Sesquiterpene synthases from grand fir (Abies grandis): Comparison of constitutive and wound-induced activities, and cDNA isolation, characterization, and bacterial expression of δ-selinene synthase and γ- humulene synthase’, *Journal of Biological Chemistry*, 273(4), pp. 2078–2089. doi: 10.1074/jbc.273.4.2078.
Steiner, K. and Schwab, H. (2012) ‘Recent advances in rational approaches for enzyme engineering’, *Computational and Structural Biotechnology Journal*, 2, p. e201209010. doi: 10.5936/csbj.201209010.
Styles, M. Q., Nesbitt, E. A., Marr, S., Hutchby, M. and Leak, D. J. (2017) ‘Characterization of the first naturally thermostable terpene synthases and development of strategies to improve thermostability in this family of enzymes’, *FEBS Journal*, 284(11), pp. 1700–1711. doi: 10.1111/febs.14072.
Sun Tai ping and Kamiya, Y. (1994) ‘The Arabidopsis GA1 locus encodes the cyclase ent-kaurene synthetase A of gibberellin biosynthesis’, *Plant Cell*, 6(10), pp. 1509–1518. doi: 10.1105/tpc.6.10.1509.
Tantillo, D. J. (2017) ‘Importance of Inherent Substrate Reactivity in Enzyme-Promoted Carbocation Cyclization/Rearrangements’, *Angewandte Chemie International Edition*, 56(34), pp. 10040–10045. doi: 10.1002/anie.201702363.
Tarshis, L. C., Yan, M., Poulter, C. D. and Sacchettini, J. C. (1994) ‘Crystal Structure of Recombinant Farnesyl Diphosphate Synthase at 2.6-ang Resolution’, *Biochemistry*, 33(36), pp. 10871–10877. doi: 10.1021/bi00202a004.
The Merck Index Online (2020) *Malachite Green*. Available at: https://www.rsc.org/Merck-Index/monograph/m7033/malachite green?q=unauthorize (Accessed: 31 March 2020).
Theisen, M. J., Misra, I., Saadat, D., Campobasso, N., Miziorko, H. M. and Harrison, D. H. T. (2004) ‘3-hydroxy-3-methylglutaryl–CoA synthase intermediate complex observed in “real-time”’, *Proceedings of the National Academy of Sciences of the United States of America*, 101(47), pp. 16442–16447. doi: 10.1073/pnas.0405809101.
Tholl, D. (2006) ‘Terpene synthases and the regulation, diversity and biological roles of terpene metabolism’, *Current Opinion in Plant Biology*, 9(3), pp. 297–304. doi: 10.1016/j.pbi.2006.03.014.
Tholl, D. (2015) ‘Biosynthesis and Biological Functions of Terpenoids in Plants’, in

Biotechnology of Isoprenoids. Cham: Springer International Publishing, pp. 63–106. doi: 10.1007/10_2014_295.
Tu, Y. (2017) *From Artemisia annua L. to Artemisinins: the discovery and development of artemisinins and antimalarial agents*. London, UK: Academic Press.
UniProt (2017) *UniProtKB - P77488 (DXS_ECOLI)*. 153rd edn. UniProt. Available at: http://www.uniprot.org/uniprot/P77488.
Vardakou, M., Salmon, M., Faraldos, J. A. and O'Maille, P. E. (2014) 'Comparative analysis and validation of the malachite green assay for the high throughput biochemical characterization of terpene synthases', *MethodsX*, 1, pp. 187–196. doi: 10.1016/j.mex.2014.08.007.
Vedula, L. S., Jiang, J., Zakharian, T., Cane, D. E. and Christianson, D. W. (2008) 'Structural and Mechanistic Analysis of Trichodiene Synthase using Site-Directed Mutagenesis: Probing the Catalytic Function of Tyrosine-295 and the Asparagine-225/Serine-229/Glutamate-233–Mg(2+)(B) Motif', *Archives of Biochemistry and Biophysics*, 469(2), pp. 184–194. doi: 10.1016/j.abb.2007.10.015.
Vedula, L. S., Rynkiewicz, M. J., Pyun, H.-J., Coates, R. M., Cane, D. E. and Christianson, D. W. (2005) 'Molecular Recognition of the Substrate Diphosphate Group Governs Product Diversity in Trichodiene Synthase Mutants', *Biochemistry*, 44(16), pp. 6153–6163. doi: 10.1021/bi050059o.
Vetter, T., Burcham, C. L. and Doherty, M. F. (2014) 'Attainable Regions in Crystallization Processes: Their Construction and the Influence of Parameter Uncertainty', *Computer Aided Chemical Engineering*, 34, pp. 465–470. doi: 10.1016/B978-0-444-63433-7.50062-6.
Voet, D., Voet, J.G. & Pratt, C. W. (2008) *Fundamentals of Biochemistry - Life at the Molecular Level*. Third. Edited by S. Dumas. United States of America: John Wiley & Sons, Inc.
Wallace, A. C., Laskowski, R. A. and Thornton, J. M. (1995) 'LIGPLOT: a program to generate schematic diagrams of protein-ligand interactions.', *Protein engineering*, 8(2), pp. 127–34. Available at: http://www.ncbi.nlm.nih.gov/pubmed/7630882 (Accessed: 27 June 2019).
Warshel, A. and Levitt, M. (1976) 'Theoretical studies of enzymic reactions: Dielectric, electrostatic and steric stabilization of the carbonium ion in the reaction of lysozyme', *Journal of Molecular Biology*, 103(2), pp. 227–249. doi: 10.1016/0022-2836(76)90311-9.
Weiss, G. A., Watanabe, C. K., Zhong, A., Goddard, A. and Sidhu, S. S. (2000) 'Rapid mapping of protein functional epitopes by combinatorial alanine scanning', *Proceedings of the National Academy of Sciences of the United States of America*, 97(16), pp. 8950–8954. doi: 10.1073/pnas.160252097.
Wendt, K. U. and Schulz, G. E. (1998) 'Isoprenoid biosynthesis: manifold chemistry

catalyzed by similar enzymes', *Structure*, 6(2), pp. 127–133. doi: http://dx.doi.org/10.1016/S0969-2126(98)00015-X.
Withers, S. T. and Keasling, J. D. (2007) 'Biosynthesis and engineering of isoprenoid small molecules', *Applied Microbiology and Biotechnology*, 73(5), pp. 980–990. doi: 10.1007/s00253-006-0593-1.
Yamada, Y., Arima, S., Nagamitsu, T., Johmoto, K., Uekusa, H., Eguchi, T., Shin-ya, K., Cane, D. E. and Ikeda, H. (2015) 'Novel terpenes generated by heterologous expression of bacterial terpene synthase genes in an engineered Streptomyces host', *J Antibiot*, 68(6), pp. 385–394. doi: 10.1038/ja.2014.171.
Yang, L., Wang, C., Zhou, J. and Kim, S.-W. (2016) 'Combinatorial engineering of hybrid mevalonate pathways in Escherichiacoli for protoilludene production', *Microbial Cell Factories*, 15, p. 14. doi: 10.1186/s12934-016-0409-7.
Yoshikuni, Y., Ferrin, T. E. and Keasling, J. D. (2006) 'Designed divergent evolution of enzyme function', *Nature*, 440(7087), pp. 1078–1082. doi: http://www.nature.com/nature/journal/v440/n7087/suppinfo/nature04607_S1.html.
Yoshikuni, Y., Martin, V. J. J., Ferrin, T. E. and Keasling, J. D. (2006) 'Engineering Cotton (+)-δ-Cadinene Synthase to an Altered Function: Germacrene D-4-ol Synthase', *Chemistry & Biology*, 13(1), pp. 91–98. doi: http://dx.doi.org/10.1016/j.chembiol.2005.10.016.
Yu, F., Miller, D. J. and Allemann, R. K. (2007) 'Probing the reaction mechanism of aristolochene synthase with 12,13-difluorofarnesyl diphosphate', *Chemical Communications*, (40), pp. 4155–4157. doi: 10.1039/b709562g.
Zhang, F., Chen, N., Zhou, J. and Wu, R. (2016) 'Protonation-Dependent Diphosphate Cleavage in FPP Cyclases and Synthases', *ACS Catalysis*, 6(10), pp. 6918–6929. doi: 10.1021/acscatal.6b02096.
Zhang, Q., Rinkel, J., Goldfuss, B., Dickschat, J. S. and Tiefenbacher, K. (2018) 'Sesquiterpene cyclizations catalysed inside the resorcinarene capsule and application in the short synthesis of isolongifolene and isolongifolenone', *Nature Catalysis*, 1(8), pp. 609–615. doi: 10.1038/s41929-018-0115-4.
Zheng, W., Sun, F., Bartlam, M., Li, X., Li, R. and Rao, Z. (2007) 'The Crystal Structure of Human Isopentenyl Diphosphate Isomerase at 1.7 Å Resolution Reveals its Catalytic Mechanism in Isoprenoid Biosynthesis', *Journal of Molecular Biology*, 366(5), pp. 1447–1458. doi: 10.1016/j.jmb.2006.12.055.
Zhou, K. and Peters, R. J. (2011) 'Electrostatic effects on (di)terpene synthase product outcome', *Chemical Communications*, 47(14), p. 4074. doi: 10.1039/c0cc02960b.