Investigation and utilisation of the promiscuity of the sesquiterpene synthase amorpha-4,11-diene synthase

A thesis submitted to Cardiff University

for the degree of Doctor of Philosophy by

Heulwen Gwawr Lloyd Davies

Supervisor: Rudolf K. Allemann

Cardiff University
School of Chemistry
December 2021

Abstract

Terpenoids are a vast family of natural products utilised by humans as insect repellents, medicines and fragrances. The sesquiterpenoid (15-carbon terpenoid) artemisinin, a secondary metabolite of *Artemisia annua*, is the key antimalarial used to treat the devastating illness of malaria which kills around three million people a year. It is naturally generated by oxidative metabolism of the sesquiterpene amorpha-4,11-diene, produced from farnesyl diphosphate (FDP) by amorpha-4,11-diene synthase (ADS). Currently artemisinin is predominantly sourced by extraction from *A. annua*, which is converted from FDP in six steps, but more stable and cost-effective alternative sources are still desired.

The first part of this thesis details an investigation into which amino acid residues are key in facilitating conversion of achiral FDP into bicyclic amorpha-4,11 diene by ADS, probing 12 residues. Variants of ADS were created using site-directed mutagenesis and changes to the product profile analysed by GC-MS. The variants G401L and G401Y, in the G helix of ADS, generated two compounds which were identified as selina-4(15),7(11)-diene and germacrene B; never before seen products of ADS or its variants. This work shows that active site selectivity of ADS is finely tuned and furthers our understanding for its application in biocatalytic production of artemisinin.

12-Hydroxy farnesyl diphosphate (12-OH FDP), an analogue of FDP, can be converted to dihydroartemisinic aldehyde (DHAAl) in one step by ADS, compared to the four from FDP to DHAAl occurring in *A. annua*. The investigation of new chemoenzymatic routes to these key intermediates is described. A novel synthesis of DHAAl was developed using prenol and isoprenol, a chemical oxidation, two kinases, a prenyltransferase and ADS. Although less laborious than previous chemical synthesis, both kinases used were inefficient, and work to improve their activity was unsuccessful. Improving upon this, a simpler synthesis was developed from the diphosphorylation of 12-hydroxyfarnesol using two kinases to make 12-OH-FDP with ~99% enzymatic conversion achieved compared to the maximum 52% reported for the previous chemical method. This was subsequently converted by ADS to DHAAl with 1:1 epimeric selectivity for the desired 11-*R* epimer, compared to 7:2 reported previously.

Acknowledgments

First and foremost, I would like to thank my supervisor, Prof. Rudolf K. Allemann, for giving me the opportunity to do a PhD in his group, for providing the excellent research facilities that made this work possible and for his continued supervision. I would also like to thank Dr. David Miller who has helped to supervise my project, with guidance and advice during my PhD, and for proofreading this thesis so thoroughly. I thank the EPSRC for funding my PhD.

I would also like to thank several current and past members of the Allemann group. Dr. Donya Valikani, Dr. Enas Behiry, Dr. Antonio Angelastro, Dr. Irene Castellan, Dr. Victor Gonzalez Requena, Gareth Smith, Dr. Martin Ubler, and Dr. Zulfa Yoosuf Aly. Each of you made the Allemann group a warmer, funnier place to work. It was a pleasure to work with you and thank you for all the help you gave throughout my PhD. I am incredibly grateful for all the help provided by Dr. Alan Scott, Dr. Chris Jones, and Dr. Alice Dunbabin. You were always happy to help and went above and beyond when supporting me in my work, which I truly appreciated.

I must acknowledge the endless support, advice and guidance provided by Dr. Adura Adesina, Dr. Prabhakar Srivastava, and Dr. Rob Mart. You were a tremendous help throughout my PhD, always willing to share your expertise and take time from your many responsibilities. You also provided entertaining and thoughtful discussion outside of our academic work, and I'm so happy to have met you.

Your enjoyment of the PhD experience is greatly influenced by the people around you. I couldn't have asked for better friends to support me during my PhD. Ed, Raquel, Jenny, Florence, and Sara filled the past four years with laughter, joy, and countless memorable moments. Your friendships will be ones I treasure for years to come. Thank you for the support, advice, and kindness, I am so grateful to have met you all.

Outside of the lab I have to thank the many wonderful friends I met in Cardiff, Silvia, Chantelle, Muireann and Laura. Thank you for reminding me there is a life outside of the lab, you made that life so wonderful, and kept me sane throughout the endless lockdowns of 2020 and 2021.

Ac yn olaf, gallwn i byth fod wedi llwyddo yn fy mywyd heb anogaeth fy rhieni. Diolch am ddioddef y cwyno pan oedd pethau ddim yn mynd yn dda, ac am godi fy nghalon. Mae cefnogaeth ddiamod o bobl sydd yn eich caru werth y byd, a gallai byth diolch digon.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

°C	Degrees Celsius
12-OH farnesol	12-hydroxy farnesol
12-OH FDP	12-hydroxy farnesyl diphosphate
A	Adenosine
Abs	Absorbance
ACT	Artemisinin based combination therapy
ADP	Adenosine diphosphate
ADS	Amorpha-4,11-diene synthase
AMP	ampicillin
APS	Ammonium persulfate
Aq	Aqueous
ATP	Adenosine triphosphate
B	Unspecified base
β-ME	β-Mercaptoethanol
bp	Base pair
BSA	Bovine serum albumin
C	Cytosine
CDC	Centers for Disease Control and Prevention
CDP-ME	4-(cytidine-5-diphospho)-2-*C*-methyl-D-erythritol
CDP-MEP	4-(cytidine-5-diphospho)-2-*C*-methyl-D-erythritol-2-phosphate
CHO	Chinese hamster ovary
CMP	Cytidyl monophosphate
CoA	Coenzyme A
COSY	Correlated spectroscopy
cpm	Counts per minute
CTP	Cytidine triphosphate
CYP	Cytochrome P450
d	Doublet
Da	Daltons
DCM	Dichloromethane
dH_2O	Deionised water
DHAAl	Dihydroartemisinic aldehyde
DHFR	Dihydrofolate reductase
DHP	Dihydropyran
DHPS	Dihydropteroate synthase
DMADP	Dimethylallyl diphosphate
DNA	Deoxyribonucleic acid
DMF	dimethylformamide
DMSO	Dimethyl sulfoxide
dpm	Disintegrations per minute
DTBM-SEGPHOS	5,5′-Bis[di(3,5-di-*tert*-butyl-4-methoxyphenyl)phosphino]-4,4′-bi-1,3-benzodioxole, [(4*R*)-(4,4′-bi-1,3-benzodioxole)-5,5′-diyl]bis[bis(3,5-di-*tert*-butyl-4-methoxyphenyl)phosphine]
DTT	Dithiothreitol
DXP	1-Deoxy-D-xylulose-5-phosphate
e^-	Electron
E. coli	*Escherichia coli*
EDTA	Ethylenediaminetetraacetic acid

EI	Electron ionisation
EPR	Electron paramagnetic resonance
ER	Endoplasmic reticulum
ETC	Electron transport chain
FAD	Flavin adenine dinucleotide
Fd	Ferredoxin
FDP	(*E*,*E*) - Farnesyl diphosphate
FDPS	Farnesyl diphosphate synthase
FeS	Iron sulphur cluster
FHP	1-Hydroxy-3,7,11-trimethyldodeca-2,6,10-triene phosphonic acid
FLAG	N terminus DYKDDDDK C-terminus
FMN	Flavin mononucleotide
G	Guanine
GB1	Protein G B1 domain
GBS	Germacrene B synthase
GC	Gas chromatography
GC-FID	Gas chromatography flame ionisation detection
GC-MS	Gas chromatography-mass spectrometry
GDP	Geranyl diphosphate
GGDP	Geranylgeranyl diphosphate
h	Hour
H	Hydrogen
HEPES	2[4-(2-Hydroxyethyl)-1-piperazin-1yl]ethanesulfonic acid
His tag	Histidine tag
HMB-PP	1-hydroxy-2-methyl-2-(*E*)-butenyl 4-diphosphate
HMG-CoA	Hydroxymethylglutaryl coenzyme A
Hz	Hertz
IC_{50}	Half maximal inhibitory concentration
IDP	Isopentenyl diphosphate
IPK	Isopentenyl phosphate kinase
IPTG	Isopropyl-β-D-1-thiogalactopyranoside
J	Coupling constant (in Hz)
k_{cat}	Turnover number
K_D	Dissociation constant
kDa	Kilo daltons
K_M	Michaelis constant
l	Pathlength
LB	Luria-Bertani growth media
M	Molar
MeCN	Acetonitrile
MEcPP	2-*C*-methyl-D-erythritol-2,4-cyclodiphosphate
MEP	2-*C*-Methyl-D-erythritol-4-phosphate
min	Minute(s)
MS	Mass spectrometry
$MsCl_2$	Methane sulfonyl chloride
MVA	Mevalonic acid
MW	Molecular weight
Myc	N terminus EQKLISEEDL C terminus
NAD^+	Oxidised nicotinamide adenine dinucleotide
NADH	Reduced nicotinamide adenine dinucleotide

$NADP^+$	Oxidised nicotinamide adenine dinucleotide phosphate
NADPH	Reduced nicotinamide adenine dinucleotide phosphate
NCS	*N*-Chlorosuccinimide
NDP	Nerolidyl diphosphate
NIST	National Institute of Standards and Technology
NMR	Nuclear magnetic resonance spectroscopy
NTA	Nitriolotriacetic acid
O	Oxygen
OD_{600}	Optical density (600 nm)
ODP	Diphosphate
ox	Oxidised
P450	Cytochrome P450
PAGE	Polyacrylamide gel electrophoresis
PCR	Polymerase chain reaction
PEG	Polyethylene glycol
PI3P	phosphatidylinositol-3-phosphate
PMSF	Phenylmethyl sulfonyl fluoride
ppm	Parts per million
PVDF	Polyvinylidene difluoride
q	Quartet
R	Unspecified substituent
red	Reduced
RNA	Ribose nucleic acid
rpm	Revolutions per minute
RT	Room temperature
s	Singlet
S. aureus	*Staphylococcus aureus*
S. cerevisiae	*Saccharomyces cerevisiae*
SDM	Site directed mutagenesis
SDS-PAGE	Sodium dodecyl sulphate-polyacrylamide gel electrophoresis
SE	Standard error
SERCA	Sacro/endoplasmic reticulum calcium ATPase
SUMO	Small Ubiquitin-like Modifier
t	Triplet
TAE	Tris-acetate-EDTA
TB	Terrific broth
TBST	Tris buffered saline tween
TCEP	Tris(2-carboxyethyl)phosphine
TCTP	Translationally controlled tumour protein
TEMED	N,N,N’,N’-tetramethylenediamine
THF	Tetrahydrofuran
ThiM	Hydroxyethylthiazole kinase
TIC	Total ion current chromatogram
TLC	Thin layer chromatography
TRIS	Trisaminomethane
tRNA	Transfer ribonucleic acid
U	Enzyme unit
UPK	Undecaprenol kinase
UV	Ultraviolet
UV-Vis	Ultra-violet visible

WHO	World Health Organisation
WT	Wild type
YPD	Yeast peptone dextrose

Amino acid abbreviations

Amino acids are expressed as a triplet or single letter code

Amino acid	3-letter code	1-letter code
L-Alanine	Ala	A
L-Arginine	Arg	R
L-Asparagine	Asn	N
L-Aspartic acid	Asp	D
L-Cysteine	Cys	C
L-Glutamic acid	Glu	E
L-Glutamine	Gln	Q
L-Glycine	Gly	G
L-Histidine	His	H
L-Isoleucine	Ile	I
L-Leucine	Leu	L
L-Lysine	Lys	K
L-Methionine	Met	M
L-Phenylalanine	Phe	F
L-Proline	Pro	P
L-Serine	Ser	S
L-Threonine	Thr	T
L-Tryptophan	Trp	W
L-Tyrosine	Tyr	Y
L-Valine	Val	V

Introduction

1.1. Terpenes

Terpenes are the largest class of natural products comprising upwards of 80,000 compounds with the broadest range of structural types.[1,2] Terpenes are used in a variety of ways including spices such as 1,8-cineole (**1**), a component of eucalyptus oil,[3] flavours such as menthol (**2**) from mint,[4] and as a component of fragrances such as limonene (**3**) found in citrus fruit.[5] They are also used for medicinal purposes, such as the anti-malarial drug artemisinin (**4**),[3] the antibiotic pentalenolactone (**5**),[6] and the anticancer drug taxol (**6**).[7] Terpenes are widespread in plants, and can act as insect attractants *via* pleasant scents released from flowers such as hedycaryol (**7**) from *Camellia hiemalis*,[8] and act as insect repellents such as 7-*epi*-zingiberene (**8**) released from wild tomatoes.[9] There is even interest in using terpenes as biofuels, such as a dimer of pinene (**9**),[10] showcasing the sheer diversity of this family of compounds. Terpenes are universally present in all kingdoms of life, and although are generally referred to as secondary metabolites, some examples such as squalene (**10**), ubiquinone (**11**) and carotenoids (**12**) (Figure 1) are vital for life.[3]

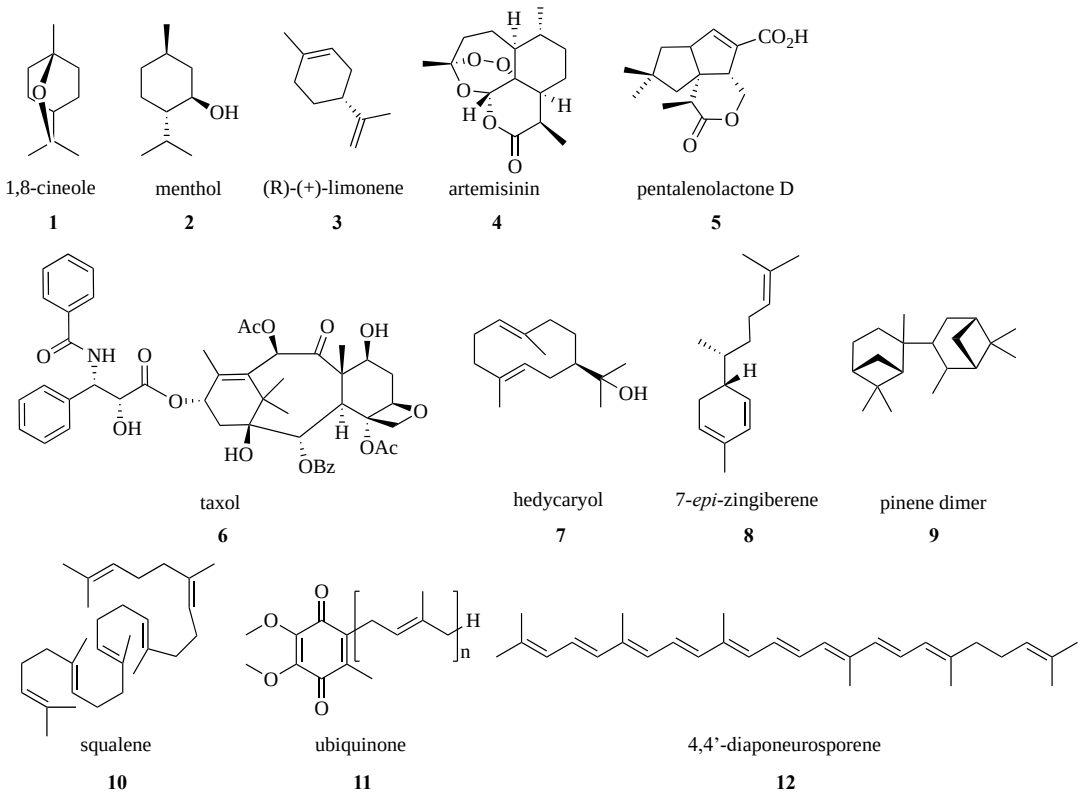
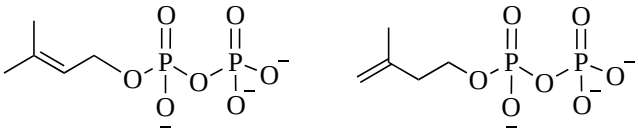


*Figure 1: Chemical structures of 1,8-cineole (**1**),[3] menthol (**2**),[4] limonene (**3**),[5] artemisinin (**4**),[3] pentalenolactone D (**5**),[6] taxol (**6**),[7] hedycaryol (**7**),[8] 7-*epi*-zingiberene (**8**),[9] pinene dimer (**9**),[10] squalene (**10**),[3] ubiquinone (**11**)[11] and the carotenoid 4,4'-diaponeurosporene (**12**).[3]*

In 1953 Ruzicka proposed the biogenetic isoprene rule, suggesting that all terpenes' carbon skeletons are composed of isoprene (**32**) units, joined in a head-to-tail or tail-to-head fashion. The number of isoprene (**32**) units in a terpene determines how they are classified e.g. two isoprene (**32**) unit compounds are monoterpenes, see Table 1.[12] It was later discovered that all terpenes are made from one of two five carbon scaffolds, iso-pentenyl diphosphate (IDP) (**14**) or dimethylallyl diphosphate (DMADP) (**13**), referred to as prenyl diphosphate precursors (Figure 2). These are produced *via* the mevalonate or methylerythritol pathway.[13]



dimethylallyl diphosphate **13**

isopentenyl diphosphate **14**

*Figure 2: DMADP (**13**), IDP **(14)***

The mevalonate pathway

The mevalonate pathway is used to produce isoprenoids in eukaryotes, archaea and the cytosols of higher plants. This pathway begins with the condensation of two molecules of acetyl-CoA (**15**) catalysed by the enzyme acetyl-CoA acetyltransferase to give the thioester acetoacetyl CoA (**16**). An aldol reaction occurs between **7** and another molecule of acetyl-CoA (**15**) to give 3-hydroxy-3-methylglutaryl-CoA (**17**), catalysed by HMG-CoA synthase. Two equivalents of NADPH are used to reduce **17** to produce mevalonate (**18**), catalysed by HMG-CoA reductase. At this point, various species may differ in subsequent steps, and the mevalonate pathway follows three different routes to yield DMADP (**13**) and IDP (**14**). In eukaryotic organisms mevalonate (**18**) is phosphorylated twice using ATP to give diphosphomevalonate (**20**). A final ATP is consumed with the release of carbon dioxide and water, and DMADP (**13**) and IDP (**14**) are produced. This is catalysed by diphosphomevalonate decarboxylase. DMADP (**13**) and IDP (**14**) can be reversibly isomerised *via* a proton tautomerisation catalysed by IDP isomerase.[14]

In some archaea, bacteria and eukarya such as the bacteria *Roseiflexus castenholzii*, the Archaeal Mevalonate Pathway I is present which branches from the conventional pathway at phosphomevalonate (**19**). In contrast decarboxylation occurs first, catalysed by mevalonate decarboxylase, to produce isopentenyl phosphate (**23**). This is then phosphorylated to IDP (**14**) catalysed by isopentenyl kinase.[15]

The third pathway is termed Archaeal Mevalonate Pathway II and occurs in archaea such as *Thermoplasma acidophilum*. Its branch point occurs higher up in the pathway at mevalonate (**18**) which is phosphorylated to mevalonate-3-phosphate (**21**) rather than **19.** This is catalysed by mevalonate-3-kinase consuming ATP. It is believed that the next steps involve a second phosphorylation catalysed by a mevalonate-3-phosphate-5-kinase producing mevalonate-3,5-bis-phosphate (**22**). Decarboxylation of **22** yields isopentenyl phosphate (**23**), and the enzyme catalysing this reaction has been termed mevalonate-3,5-bisphosphate decarboxylase. Here the pathway joins the Archaeal Mevalonate Pathway I (Figure 3).[16]

This pathway can be manipulated in medicine to reduce cholesterol levels, as it is the beginning of the cholesterol biosynthesis pathway, *via* the administration of statins which are HMG-CoA reductase inhibitors. This has long term benefits in reducing cardiovascular disease and morbidity.

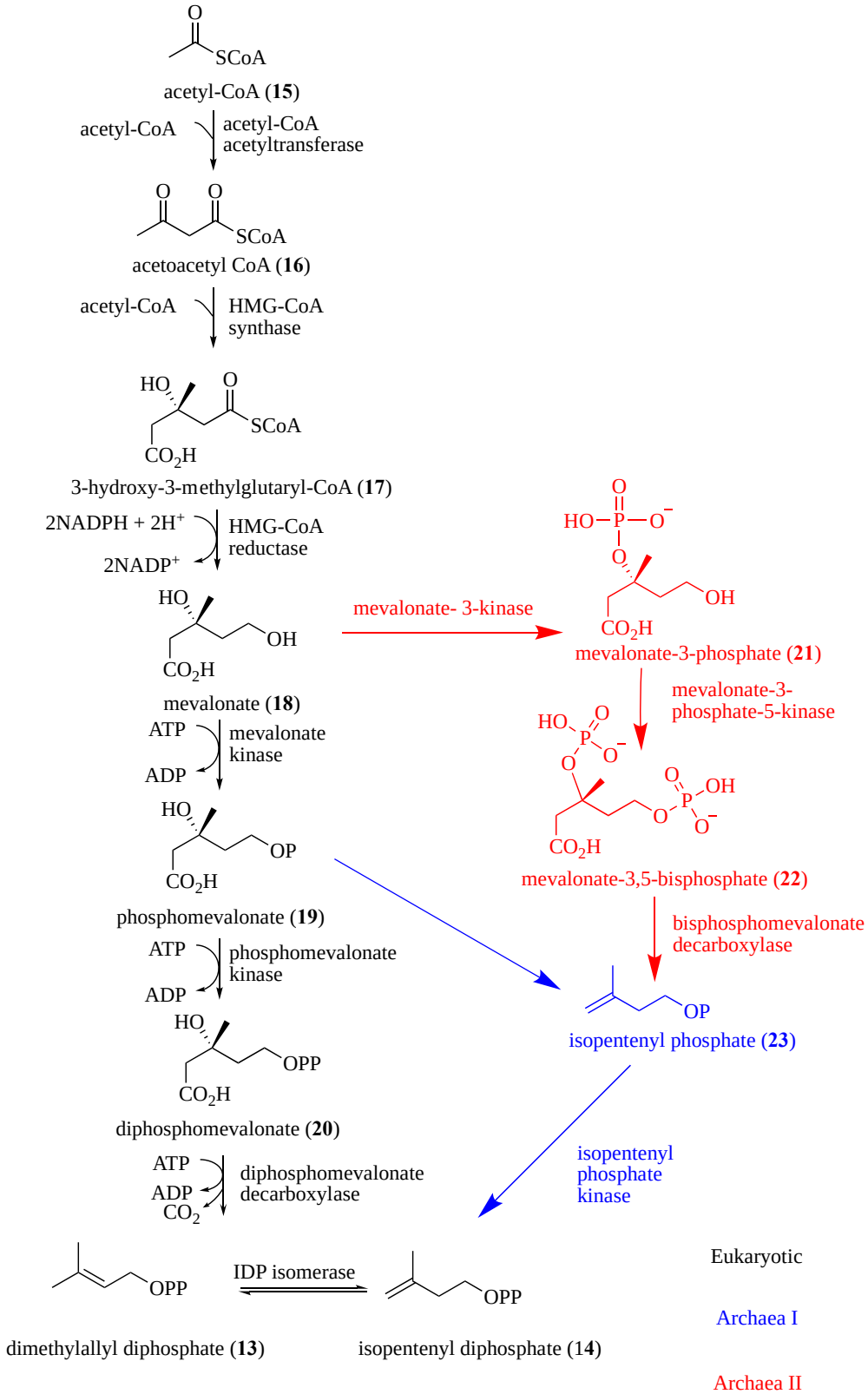


Figure 3: The mevalonate pathways

The non-mevalonate pathway

The methylerythritol pathway (non-mevalonate pathway) is an alternative metabolic pathway for producing isoprenoids. It occurs in the plastids of higher plants, algae, apicomplexan parasites and cyanobacteria. As this pathway is not used by animals, it could be a target for antimicrobials.[17] This pathway begins with the condensation of pyruvic acid (**24**), and glyceraldehyde-3-phosphate (**25**); two metabolites of the glycolysis pathway. This reaction uses the cofactor thiamine diphosphate and is catalysed by DXP synthase (DXPS) releasing carbon dioxide and DXP (**26**). DXP reductoisomerase (DXR) isomerises DXP (**26**) to the aldehyde 2-*C*-methyl-D-erythose-4-phosphate followed by a reduction consuming NADPH to give MEP (**27**).

CDP-MEP synthase (CMS) adds cytidine monophosphate to MEP (**27**) releasing diphosphate and CDP-ME (**28**). CDP-ME kinase (CMK) consumes ATP releasing ADP and CDP-MEP (**29**). Cytidine monophosphate is released to produce MEcPP (**30**) catalysed by MEcPP synthase (MCS).

The final two steps of the non-mevalonate pathway involve iron-sulfur cluster proteins that have yet to be successfully expressed in eukaryotic organisms[14] and the catalytic mechanisms of these enzymes are unclear. A ring opening of MEcPP (**30**) occurs, catalysed by HMB-PP synthase (HDS) oxidizing two ferredoxin (Fd) proteins producing HMB-PP (**31**)**.** A molecule of NADH is consumed by HMB-PP reductase (HDR) producing water, IDP (**14**) and DMADP (**13**) (Figure 4).[17] For every one DMADP (**13**) made, five to six IDP (**14**) are made, which is maintained at the ratio 3:7 by isopentenyl diphosphate isomerase.[18] HDS is believed to have three crucial cysteine residues for activity and a radical mechanism has been postulated. HDR is also hypothesised to complete the reaction *via* a radical mechanism, and both enzymes are suggested to have a $[4Fe\text{-}4S]^{2+}$ cluster.[17]

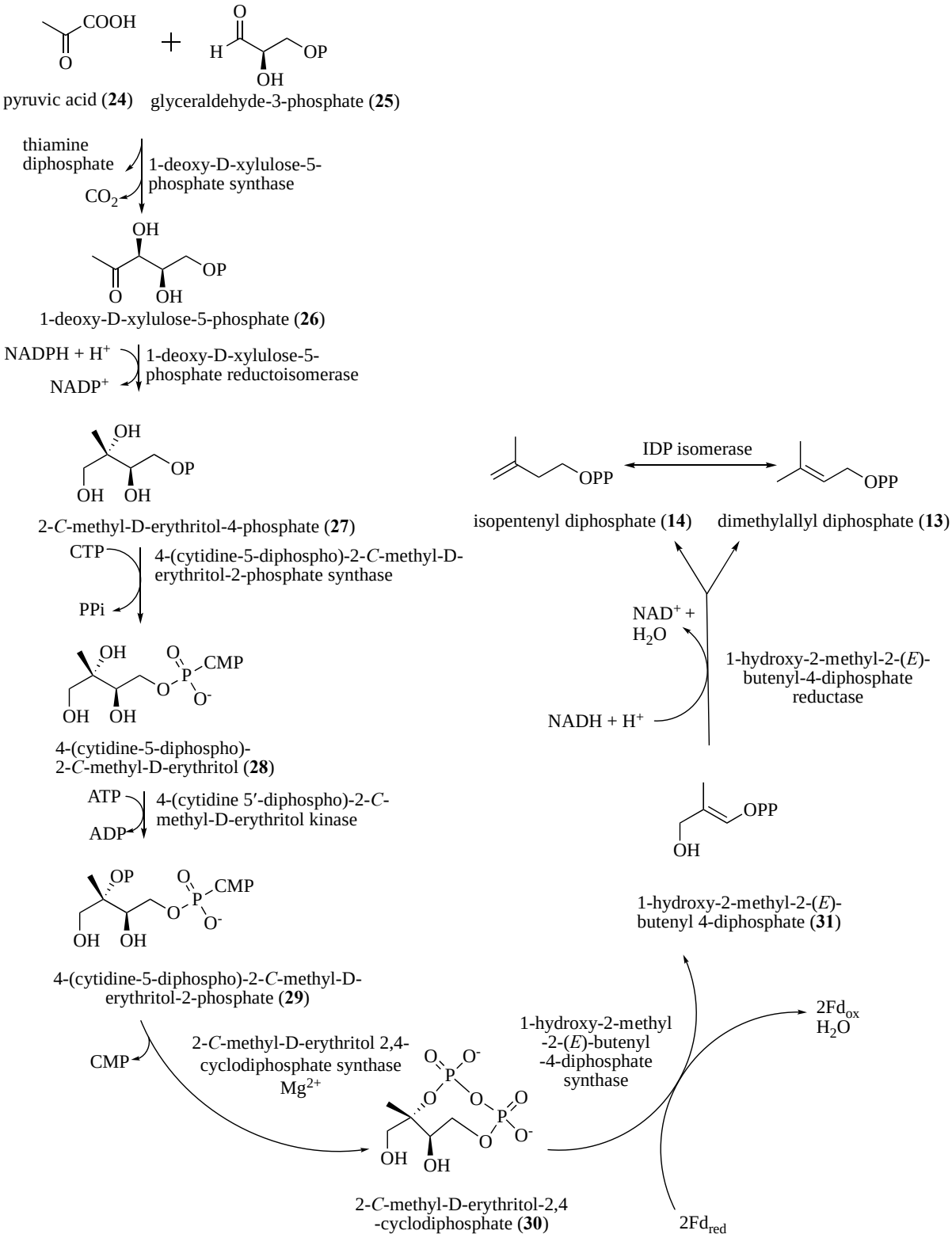


Figure 4: The non-mevalonate pathway

Prenyltransferases

IDP (**14**) and DMADP (**13**) are substrates for prenyltransferases, also called isoprenyl diphosphate synthases, which utilise terpene precursors. There are four possible reactions to build prenyl chains.[19] The most common attachment reaction involves an isoprenoid diphosphate chain adding to IDP (**14**) *via* a one-four linkage, referred to as a head-to-tail condensation reaction (catalytic mechanism in Figure 5). An example of an enzyme that catalyses such a reaction is avian FDPase.[20] In a head-to-tail condensation reaction, the diphosphate group of the substrate is removed, forming a carbocation at carbon one, aided by a divalent metal ion, most commonly, magnesium. A nucleophilic attack occurs from the double bond, and a stable tertiary carbocation forms in the elongated allylic chain. Deprotonation occurs producing the final isoprenyl diphosphate. This could be a substrate for further elongation reactions, or for a terpene cyclase.[21]

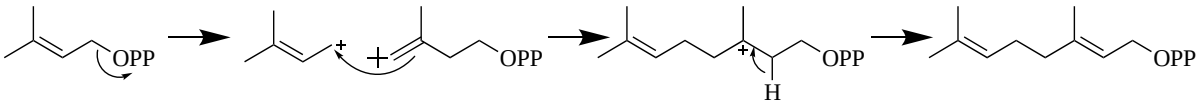


Figure 5: Head-to-tail prenyltransferase reaction.[21]

Although this is the most common elongation reaction, others can occur. Metabolites derived from cyclopropanation reactions are also widely found in nature and one such example is chrysanthemyl diphosphate synthase, which combines two molecules of DMADP (**13**) to form chrysanthemyl diphosphate (Figure 6).[19]

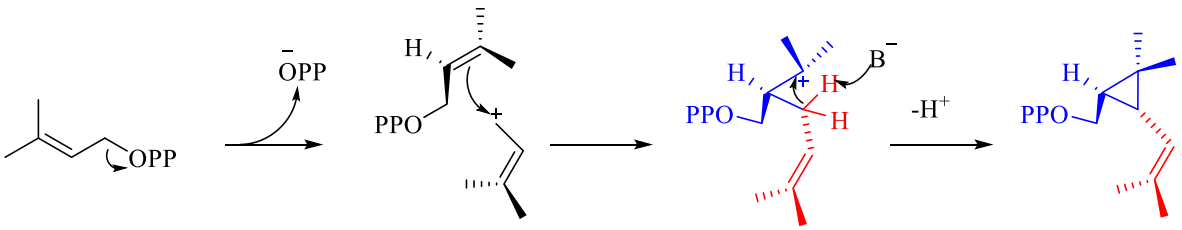


Figure 6: Cyclopropanation prenyltransferase reaction.[19]

Less commonly found are metabolites derived from head-to-middle condensation reactions, (Figure 7), which can lead to branched products such as lavandulyl diphosphate.[19]

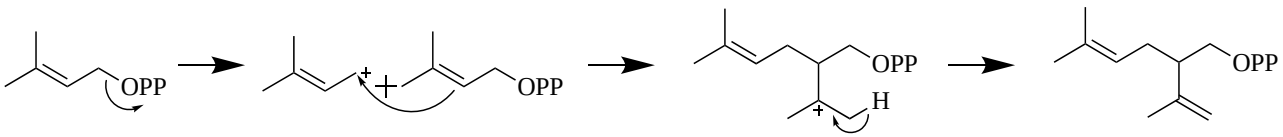


Figure 7: Head-to-middle prenyltransferase reaction.[21]

Products of cyclobutanation prenyltransferase reactions are also known, such as maconelliyl diphosphate (Figure 8). Enzymes which complete branching and cyclobutanation reactions have yet to be identified.[19]

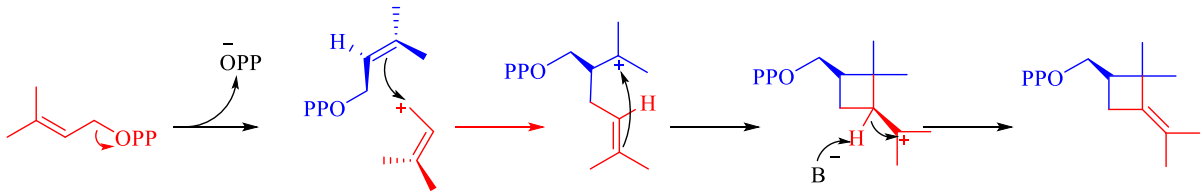


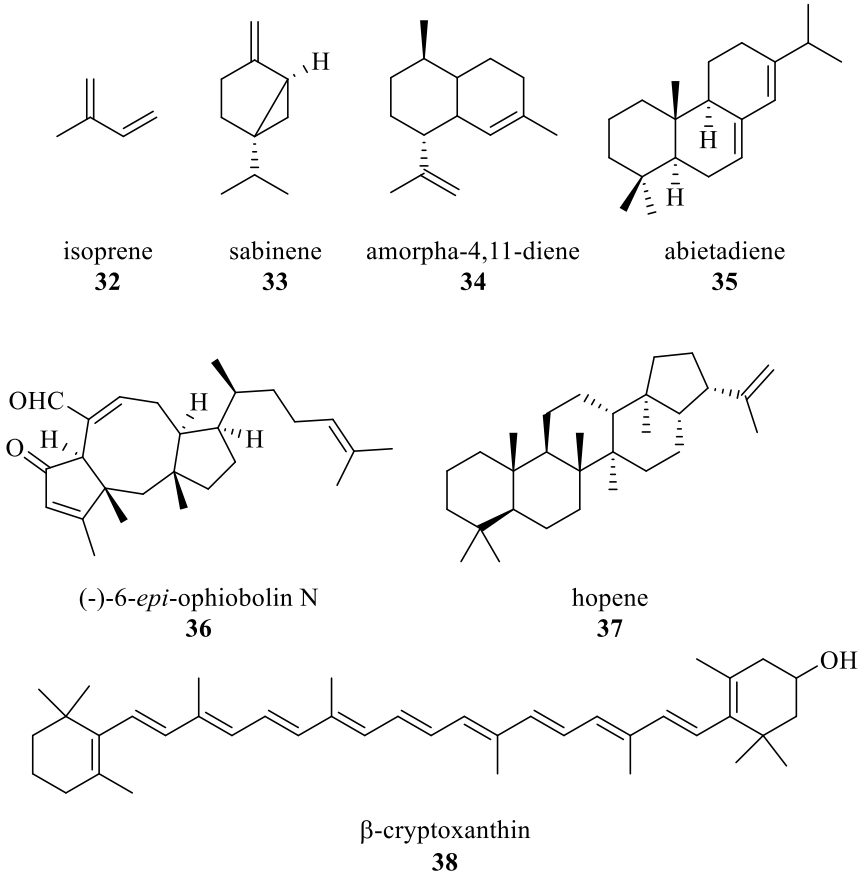
Figure 8: Cyclobutanation prenyltransferase reaction.[19]

GDP synthase catalyses a head-to-tail condensation of IDP (**14**) and DMADP (**13**), yielding geranyl diphosphate (a 10-carbon unit with *E*-stereochemistry). C_{10} terpene units are referred to as monoterpenes.[15,22] A 15-carbon unit, farnesyl diphosphate (FDP) is then produced from GDP and IDP (**14**). This can then be used to produce sesquiterpenes such as artemisinin (**4**).[23] Four 5-carbon units can be used to build geranylgeranyl diphosphate (GGDP (**42**)), a 20-carbon unit. FDP and GGDP (**42**) can be used in elongation reactions to produce compounds containing upwards of 50 carbons. Some examples are presented in Table 1 and figure 9.

Terpenes can be divided into classes based on the number of 5-carbon units they contain and the cyclases can be categorised *via* their preferred substrate: those which accept GDP are called monoterpene cyclases, those which accept FDP are called sesquiterpene cyclases, and so on.[24]

Table 1: Terpene classifications and examples

Class	No. C's	Precursor	Example terpene
Hemiterpenes	5	DMADP (**13**)	isoprene[3] (**32**)
Monoterpenes	10	GDP	sabinene[25] (**33**)
Sesquiterpenes	15	FDP	amorpha-4,11-diene[26] (**34**)
Diterpenes	20	GGDP (**42**)	abietadiene[27] (**35**)
Sesterterpenes	25	Geranylfarnesyl diphosphate (GFDP)	(-)-6-*epi*-ophiobolin N[28] (**36**)
Triterpenes	30	2x FDP	hopene[29] (**37**)
Tetraterpenes	40	2x GGDP (**42**)	β-cryptoxanthin[30] (**38**)



isoprene **32** | sabinene **33** | amorpha-4,11-diene **34** | abietadiene **35**

(-)-6-*epi*-ophiobolin N **36** | hopene **37**

β-cryptoxanthin **38**

*Figure 9: Structures of isoprene (**32**),[3] sabinene (**33**),[25] amorpha-4,11-diene (**34**),[26] abietadiene (**35**),[27] (-)-6-epi-ophiobolin N (**36**),[28] hopene (**37**),[29] and β-cryptoxanthin (**38**)[30]*

The family of prenyltransferases have very broad activities with substrates ranging from 10-carbon compounds to those containing upwards of 5000 carbons. Koyama split this family of enzymes into four classes according to chain length of the product and stereochemistry,[24] and later added a fifth class to distinguish *cis*-prenyltransferases of varying product lengths.[31] Class I prenyltransferases synthesise short chains from geranyl diphosphate (C_{10}) to geranylgeranyl diphosphate (GGDP (**42**)) (C_{20}) with *E* selectivity.[24] Class II are those which catalyse the synthesis of medium chain prenyl diphosphates, (C_{30} and C_{35}), also with *E* selectivity. Long chain prenyl diphosphates (C_{40}, C_{45}, and C_{50}) are synthesised by class III prenyltransferases with *E* selectivity. These enzymes share characteristic DDXXD motifs for binding magnesium ions to coordinate diphosphate groups and are *trans*-prenyltransferases.[32] Residues close to the first DDXXD motif affects the length of the final prenyl chain[24], such as the residues at the fifth and sixth position before the motif.[32]

The products formed by class IV enzymes differ in the stereochemistry of the elongation reactions and include *cis*-farnesyl diphosphate synthase. Class V enzymes are also *cis*-

prenyltrasnferases and synthesise prenyl chains varying in length from 50 carbons to over 5000.[31] Class IV and V prenyltransferases bare no sequence similarity to the *trans*-prenyltransferases and do not contain the conserved DDXXD motifs.[24] *Cis*-prenyltrasnferases have a conserved RX_5RX_nE sequence which recognises the diphosphate group of the substrate. The positively charged residues interact directly with the diphosphate while the negative residues interact *via* magnesium ions. These enzymes do not contain the characteristic 'α fold' like *trans*-prenyltrasnferases, the active site is surrounded by a mix of α helices and β-sheets.[31] Residues at the bottom of the substrate pocket are believed to determine chain length in *cis*-prenyltrasnferases.[24]

As more prenyltransferases were discovered it was found that the stereochemistry of elongation reactions varies in an *E* or *Z* manner. Although *E,E*-FDP (**40**) is the most commonly reported substrate for sesquiterpene synthases, exceptions do exist. In the insect species *Phyllotreta striolata*, a gene was found which encoded an enzyme which catalysed the production of *Z,E*-FDP (**41**), which was the substrate used to produce (6*R*,7*S*)-himachala-9,11-diene, an insect hormone.[33] *Z,Z*-FDP (**39**), (Figure 10), is also an alternative substrate for sesquiterpene synthases. Examples of terpenes produced *via* this alternative substrate include santalene and bergamotene. These discoveries show that the vast library of terpenes of which we are aware is likely to only be a snapshot of the terpenes that exist in nature.[34]

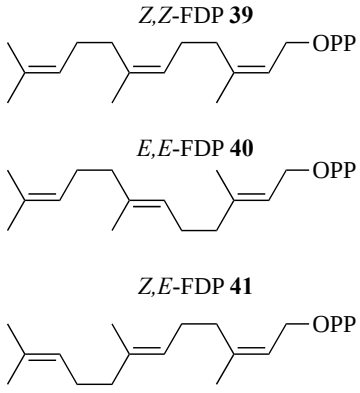


Figure 10: Isomers of FDP[33,34]

Terpene cyclases

Terpene cyclases catalyse the conversion of linear prenyl diphosphates to cyclic compounds *via* a complex cascade of stereospecific reactions. This process can involve hydride shifts, carbocation rearrangements, deprotonations and the use of surrounding solvents, namely water,

for carbocation quenching. The carbocation formed is stabilised by the diphosphate group lost as well as *via* various basic and aromatic residues in the active site.[3] Terpene cyclases tend to have active sites in deep hydrophobic pockets, protecting the cationic intermediates of the reaction from quenching by surrounding solvent. The residues lining these pockets dictate the conformations taken by the substrate and the resulting product profile.[35]

As the first X-ray crystal structures of terpene cyclases were solved it was seen that they generally fit into two classes, class I and class II and all contain one or more of three domains: α, β and γ. The class I active site is in the α domain, while the class II active site sits between the β and γ domains. All cyclisations begin with the formation of a carbocation and the mode by which it forms also splits cyclases into two classes. Class I terpene cyclases are grouped together as the initial formation of the carbocation occurs *via* the same catalytic mechanism. It involves a triad of metal ions in a cluster coordinating the diphosphate and enabling the cleavage of the diphosphate group from the prenyl.[3] Class II terpene cyclases rely on protonation of a double bond to form the carbocation, or of an epoxide leading to ring opening and the movement of the carbocation to an internal double bond. [36]

Class I terpene cyclases

The active site of class I terpene cyclases are located in the 'α fold'; an α helical bundle with key sequences of amino acid residues.[3] These active sites, also termed the class I terpenoid synthase fold, generally constitute of five α helices surrounding the active site.[35] It is believed that class I terpene cyclases evolved from class I prenyltransferases as they have a very similar conserved structure in the α fold.[37] One example of this class is aristolochene synthase which produces the sesquiterpene aristolochene.[3] The monomer of aristolochene synthase is shown in Figure 11. When the enzyme originating from *Aspergillus terreus* was crystallised it was discovered to exist as a tetramer with each monomer containing the α - helical fold. It is believed that the enzyme exists as a dimer in solution but at appropriately high concentrations can oligomerise. [38]

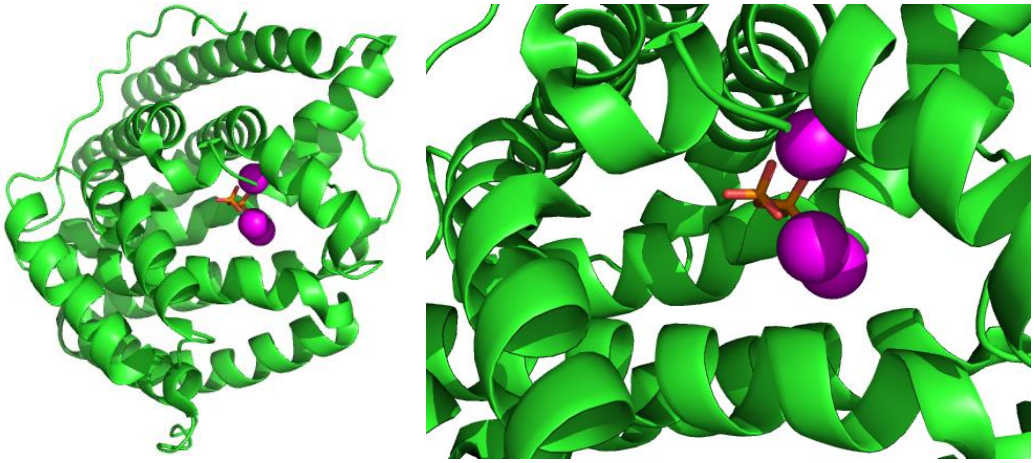


Figure 11: Cartoon representations of a monomer of aristolochene synthase (PDB:2OA6) with three magnesium ions (pink) and a diphosphate group (orange)[38]

Two key sequences of amino acids are vital for terpene cyclase activity, one on the D helix and one on the H helix. The first, termed the aspartate rich motif, generally has the sequence DDXXD/E, and helps coordinate two metal ions in the active site. These magnesium ions are designated A and C.[39] These coordinate to the diphosphate group of the substrate, aiding in its removal along with a third ion. The metal ions are most commonly magnesium ions but manganese ions will facilitate turn over, not always with the same product profile. The second sequence on helix H is called the NSE/DTE motif, these are the residues that bind the third magnesium ion,[4] termed the B magnesium ion,[39] and has the sequence (N,D)D(L,I,V)X(S,T)XXXE.[4] There are some exceptions to this such as δ-cadinene synthase from *Gossypium arboreum*, which contains a second DDXXD motif instead of the NSE/DTE

motif,[40] and germacrene D synthase from *Solidago canadensis*, in which the aspartate rich motif is **N**DXXD.[41] In Figure 12, the three aspartate residues of the DDXXD/E motif and the DDXXTXXXE residues of the NSE/DTE motif of *epi*-aristolochene synthase are coloured in red, surrounding the three magnesium ions in magenta, and the substrate analogue 1-hydroxy-3,7,11-trimethyldodeca-2,6,10-triene phosphonic acid (FHP), which allows prediction of where FDP would sit in the active site.[42]

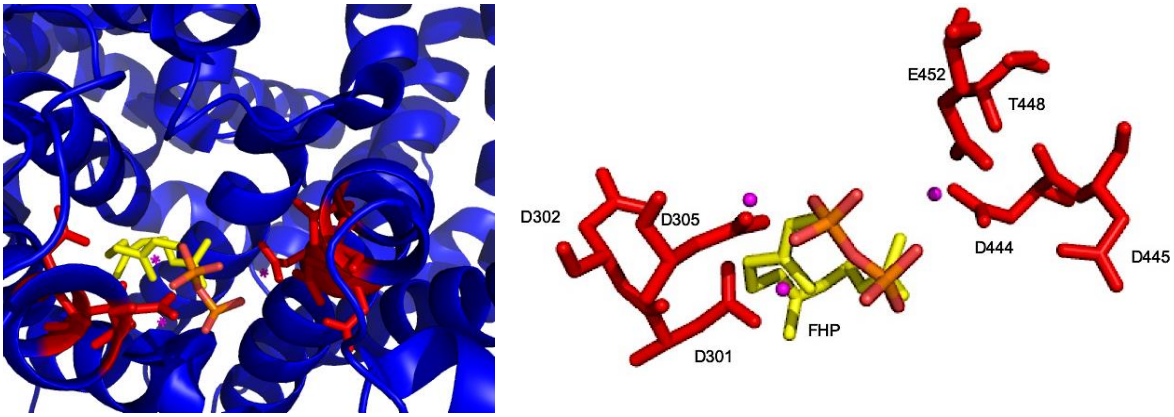


Figure 12: Left: cartoon representation of epi-*aristolochene synthase (PDB:5EAT) with the DDXXD/E and NSE/DTE motif of shown in red, FHP in yellow and magnesium ions in magenta. Right: the DDXXD/E and NSE/DTE motif residues only, with FHP and magnesium ions* [42]

Although the α domain contains the active site, class I cyclases can contain β or γ domains which are generally inactive.[3] One such example is tobacco *epi*-aristolochene synthase.[4] Cyclases have been crystallised with water molecules in the active site that aid in deprotonation reactions, but these molecules can also be used to hydroxylate the final product. How the cyclases control this differentiation is not yet understood.[3]

Bacterial and fungal sesquiterpene cyclases generally have only the α domain and are active as monomers but homodimers have been characterised. Plants on the other hand can have an additional domain at their N terminus that has some weak resemblance to the structure of a class II fold. It is suggested that the N terminus of the β domain helps stabilise the active site of the C terminal class I domain, by forming a cap over the entrance.[4] Tobacco *epi*-aristolochene synthase is a sesquiterpene synthase that has the α and β domain, where the α domain catalyses the production of *epi*-aristolochene. The function of the β domain is so far unknown[42], but is an evolutionary relic of a terpene cyclase ancestor which had all three domains, which were all active.[3] Taxadiene synthase is a diterpene synthase involved in the pathway producing taxol (**6**), a chemotherapeutic. This enzyme has class I activity but

contains an α, β and γ domain; (Figure 13), the α domain exhibits class I activity, while the β and γ domains have no known catalytic activity, but show some similarity to a class II terpene cyclase.[43]

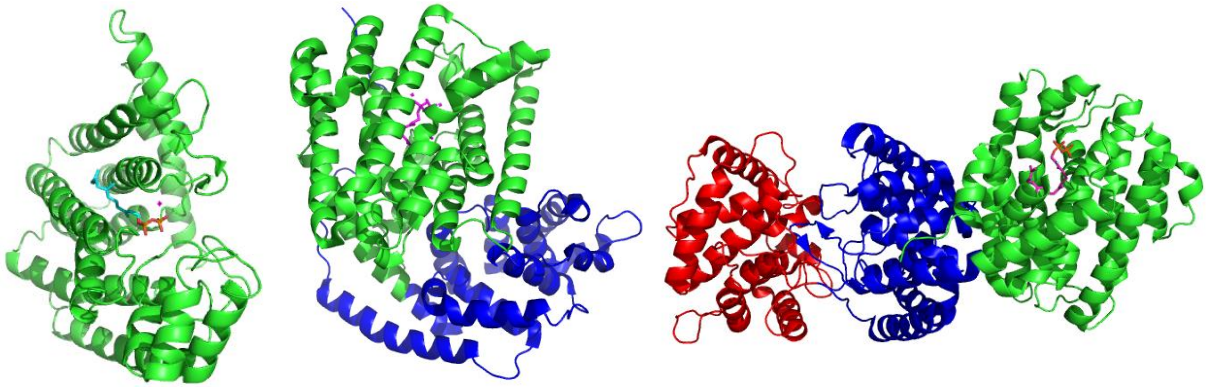


Figure 13:Cartoon representations of each terpene cyclase domain: first, the avian farnesyl diphosphate synthase monomer (PDB:1UBX) which represents the α helical bundle of class I synthases;[44] *second,* epi-*aristolochene synthase (PDB:5EAT), a class I sesquiterpene synthase containing the α domain in green and the β domain in blue;*[42] *and third. taxadiene synthase (PDB:3P5R), a class I diterpene synthase containing the α domain in green, the β domain in blue and γ domain in red.*[43]

Class II terpene cyclases

Class II cyclases have two α barrel fold structures, the β and γ domains, termed the class II fold and, unlike class I cyclases, the structure is formed before the substrate binds.[4] Unlike class I terpene cyclases there is no clear relation of class II cyclases to prenyltransferases and they seem to have evolved independently,[37] but it is hypothesised that the two domains arose from a gene duplication as they are structurally very similar.[45]

The substrates of class II cyclases are commonly squalene (**10**), or oxido-squalene which are the precursors of triterpenes, or GGDP (**42**), the precursor of diterpenes.[36] Products of class II cyclases are often multicyclic structures with complex cyclisation mechanisms.[36] The key sequence of class II cyclases is the DXDD motif, which is believed to be the proton donor, but there is another more poorly characterised domain, which has an EDXXD motif, that aids in class II diterpene activity.[46] In diterpene cyclases which use GGDP (**42**) as a substrate, the central aspartate in the DXDD motif aids in positioning a magnesium ion that coordinates the diphosphate group.[36] Examples of class II domains in sesquiterpene synthases have been found but they are not always active, such as α-bisabolene synthase, which has the β and γ domains of a class II active site which is non-functional.[47]

Squalene-hopene cyclase, (Figure 14) is a terpene synthase that accepts the triterpene squalene (**10**) and cyclises it to hopene (**37**) or diplopterol. This is an example of a typical class II terpene cyclase that contains only the β and γ domains where the active site sits in between both these domains. The γ domain is an α_6-α_6 barrel with two rings of parallel α helixes. The β domain has a mobile α-α barrel likely evolved from the α_6-α_6 barrel.[29]

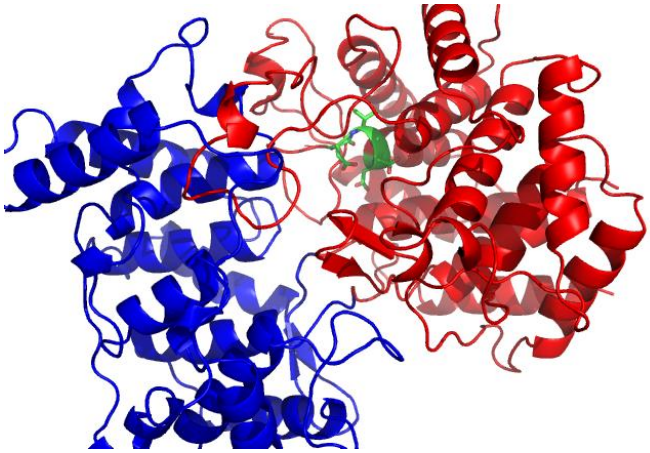


Figure 14: Cartoon representation of squalene-hopene cyclase (PDB:3SQC). The image shows the β domain in red and the γ domain in blue. The DXDD motif of the class II active site is in green, which sits between the β and γ domain.

An example of a class II terpene synthase containing the three domains is abietadiene synthase, (Figure 15) which is involved in grand fir tree resin synthesis. The class II active site sits in between the β and γ domains at the N-terminus of the protein. [27]

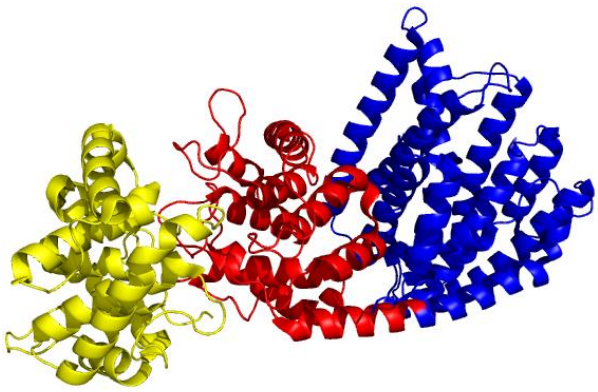


Figure 15: Cartoon representation of abietadiene synthase (PDB:3S9V), the α domain is in blue, the β domain in red and the γ domain in yellow[27]

Bifunctional terpene cyclases

Although terpene cyclases can be divided into class I and class II, those with both active sites exist, although both active sites are not functional in all. An example of a bifunctional terpene

cyclases which has a functional class I and II active sites, where the reactions occur sequentially is abietadiene synthase from *Abies grandis*.[45] Another example is the diterpene *ent*-kaurene cyclase from the fungus *Phaeosphaeria* sp. L487, whose catalytic mechanism is shown in Figure 16. This bifunctional terpene cyclase initially uses the β and γ domains, which corresponds to the class II cyclases, to cyclise GGDP (**42**) to *ent*-copalyl diphosphate (**43**). The class I α domain then cyclises *ent*-copalyl diphosphate (**43**) to form *ent*-kaurene (**44**).[48]

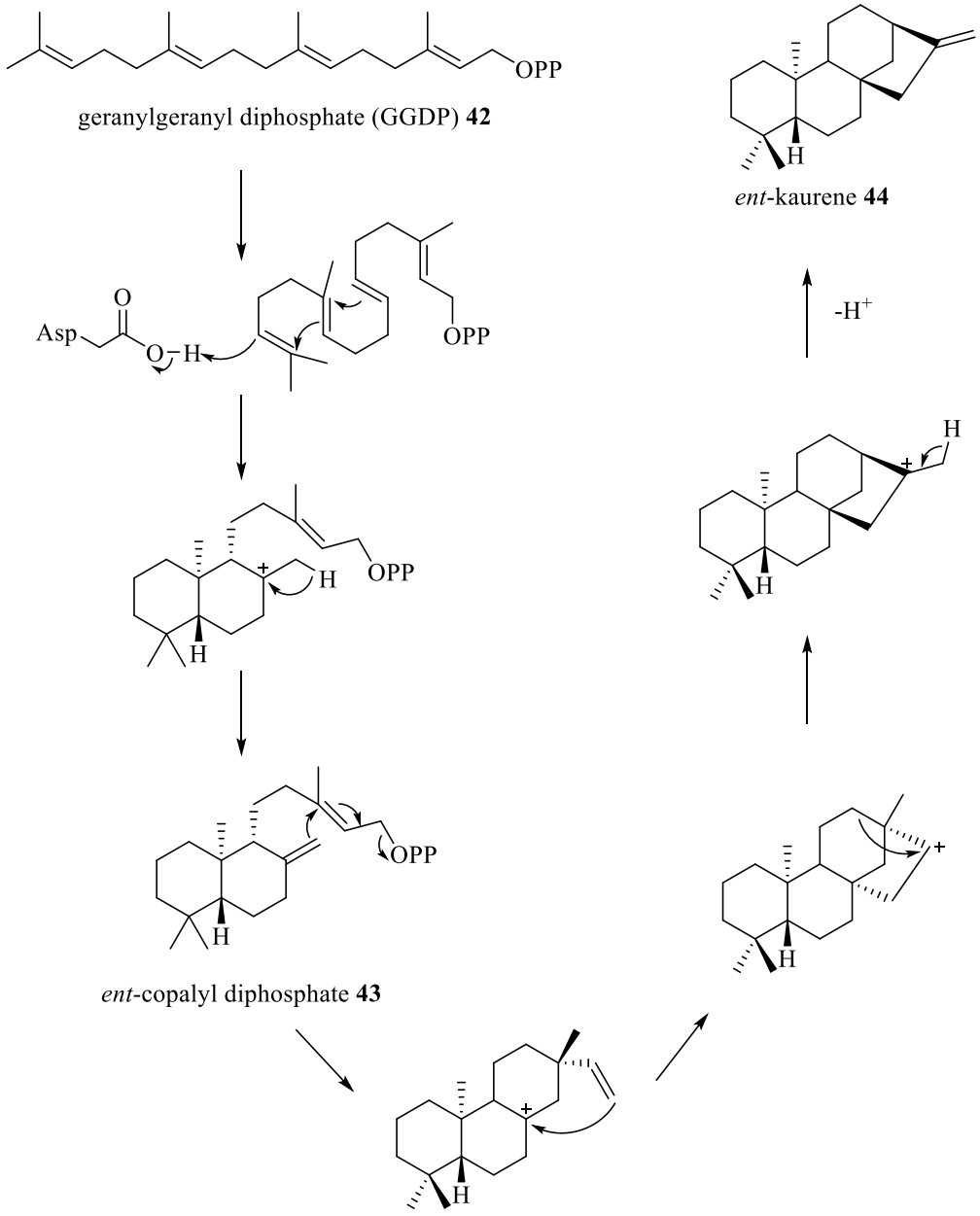
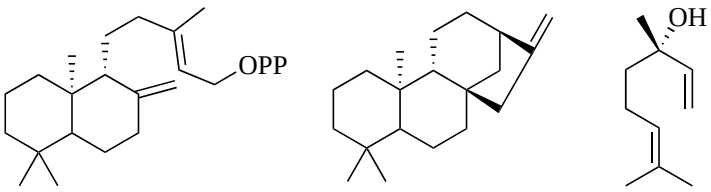


Figure 16: ent-kaurene (**44**) synthesis[45]

Alternative classifications

Within the plant kingdom, an alternative classification system of terpene synthases exists. All plant terpene synthases can be divided into six subfamilies based on their similarity and to be within the same subfamily, members must have 40% identity. These subfamilies are Tps-a through to Tps-f (Tps meaning terpene synthases).[49] It is believed that all plant terpene synthases share a common ancestor which was a terpene synthase involved in primary metabolism.[50] It is hypothesised that this ancestor gene duplicated, where one copy was modified slightly with time and is responsible for primary metabolite synthesis. The other copy mutated more varyingly, allowing the creation of a diverse panel of secondary metabolites, which were selected for if they aided in competition and survival. Tps-a, b, d and f make secondary metabolites while Tps-c and e make primary metabolites. Members of the Tps-a subfamily are mostly sesquiterpene synthases from angiosperms, while members of the Tps-b subfamily are monoterpene synthases from angiosperms. Tps-c subfamily members are (-)-copalyldiphosphate (CDP) synthases, a diterpene intermediate which can be derivatised *e.g.* into plant hormones gibberellins. The Tps-d subfamily includes monoterpene synthases from gymnosperms,[49] and the Tps-e subfamily encompasses kaurene synthases, also a precursor of gibberellins. The final subfamily, Tps-f, is a subfamily of linalool synthases.[50] Exemplary structures of products of these enzymes are shown in Figure 17.



(-)-CDP **45** (-)-kaurene **46** (+)-linalool **47**

*Figure 17: Examples of products from Tps-c: (-)-CDP (**45**), Tps-e: (-)-kaurene (**46**)[49], and Tps-f: (+)-linalool (**47**).*

1.2. Sesquiterpene synthases

Sesquiterpenes are a class of natural products with online libraries noting almost 900 compounds in the family.[51] They are all derived from FDP but, depending on the sesquiterpene synthase that acts on FDP, they may cyclise and various functional groups can be added.[4] Sesquiterpene synthases generally lack high amino acid identity but share some common structural features[4] such as the α-helical bundle and both metal binding motifs.[36] Sesquiterpene

synthases generally consist of a bundle of α helices, with loops connecting either end of the helices. On the side furthest from the active site these loops tend to be shorter, and for the site closest to the active site these are longer.[52] Class I sesquiterpene cyclases share several commonalities such as a flexible G helix-break motif which moves as the substrate binds changing the space of the active site. They also have a triad of residues termed the effector triad, an example of which from selina-4,7-diene synthase is shown in Figure 18. These residues are termed the diphosphate sensor which forms hydrogen bonds with the diphosphate group of the substrate; the linker which forms hydrogen bonds with the sensor; and the effector that moves into the active site as the substrate binds allowing formation of the carbocation. In bacteria and fungi these residues are found on the G1 helix, while in plants they are found on helix H.[53]

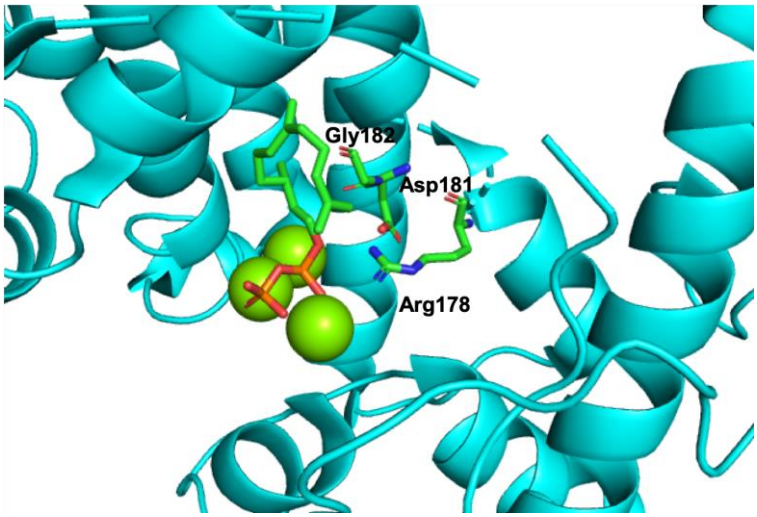


Figure 18:The effector triad of selina-4,7-diene synthase, with arginine 178 as the diphosphate sensor, the aspartate 181 as the linker and the glycine 182 as the effector.

These enzymes generally have high stereochemical precision, achieved *via* an active site surface with many aromatic and aliphatic residues that stabilise carbocations formed within the cyclisation cascade. These residues also determine product specificity as they aid in the folding of the carbon backbone. Site directed mutagenesis of various sesquiterpene synthases, where bulky aromatic residues are mutated to smaller residues, often show a change in product distribution.[39] One such example is *epi*-isozizaene synthase, whose reactions scheme to *epi*-isozizaene and sesquisabinine A is shown in Figure 19, and active site phenylalanine 95: when mutated to smaller polar residues, cyclisation from a homobisabolyl cation (**50**) to the next cation, acorenyl (**51**), is stunted. Mutations of the adjacent phenylalanine 96 in the active site to smaller polar residues leads to a different product, sesquisabinine A (**52**), with very high selectivity.[54]

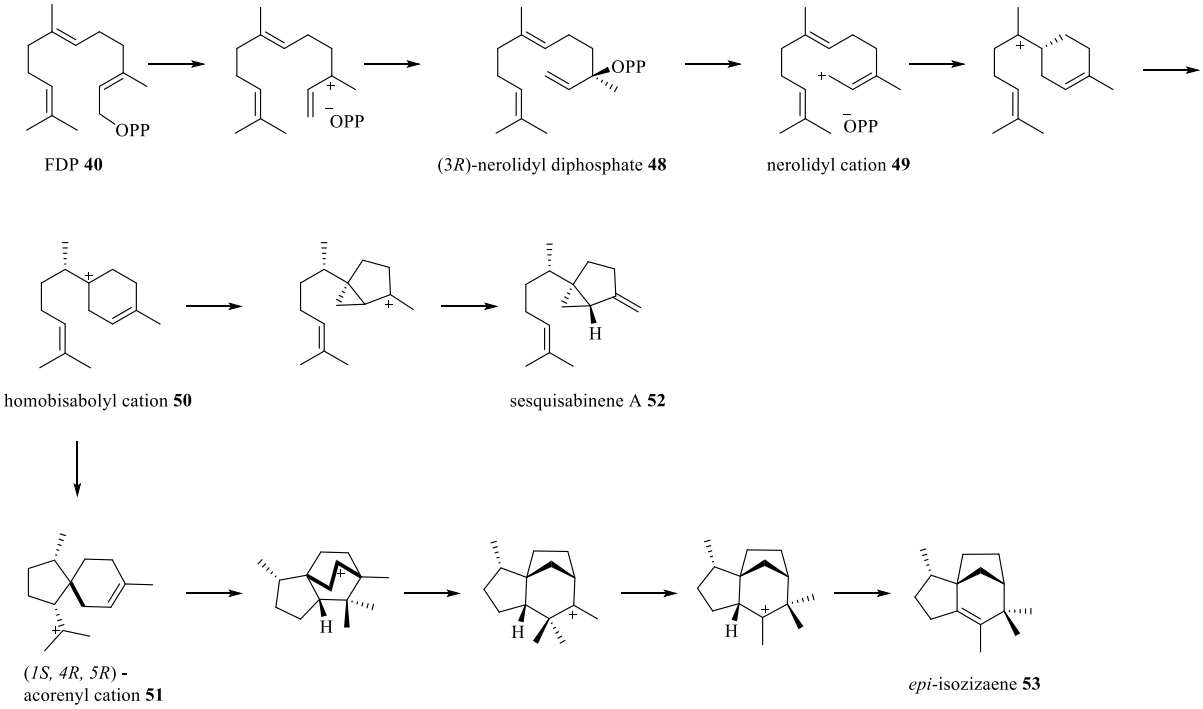


*Figure 19:Structures of intermediates of epi-isozizaene (**53**) synthesis, and alternate product sesquisabinene A (**52**)*[54]

In plants, sesquiterpene cyclases tend to be localised to the cytosol and mostly use *E,E*-FDP (**40**) as their substrates. It is hypothesised that plant sesquiterpene synthases may have formed as a diterpene synthase lost its transit peptide for entrance into the plastid and was therefore in the cytosol where FDP is the main substrate rather than GDP or GGDP (**42**) as in the plastid.[46] All sesquiterpene synthases have a class I terpene synthase α helical fold in which catalysis occurs, and most contain the characteristic DDXXD/E and DTE/NSE motifs. Some sesquiterpene synthases, such as in plants, have a class II terpene synthase fold that does not catalyse sesquiterpene synthesis.[39]

Sesquiterpene synthases activity

Sesquiterpene synthases are generally regarded as 'slow enzymes' with an average k_{cat} value of 2.5 s^{-1} compared to vital metabolic enzymes with an average k_{cat} of 79 s^{-1}. Sesquiterpene synthases generally catalyse complex multistep reactions to form cyclic products, therefore there are many possible steps in the synthesis to cause this lower efficiency.[55] Using trichodiene synthase it was determined that the rate limiting step of the overall reaction is product release, as the single-turnover rate measured is much faster than the steady-state rate.

This may be as the hydrophobic product interacts favourably with hydrophobic residues within the active site and unfavourably with the hydrophilic solution surrounding it, or that a slow conformational change of the enzyme is needed before product can be released. [56]

Sesquiterpene synthases cyclisation mechanism

The cyclisation of sesquiterpenes can occur *via* six initial cyclisation events that can be split into two categories, as shown in Figure 20. FDP can directly cycle *via* a 1,10 or 1,11-cyclisation, such as the (*E,E)*-germacradienyl (**54**) or (*E,E)*-humulenyl cation (**56**), where the *E* geometry of the C2-C3 double bond does not prohibit 1,10 or 1,11-cyclisation. The alternative category involves the isomerisation of the C2-C3 double bond to the *cis* nerolidyl cation. This can be followed by a 1,6-cyclisation to the bisaboloyl cation.[2] The nerolidyl cation can alternatively cycle *via* a 1,7-cyclisation to the cycloheptyl cation, 1,10-cyclisation to (*Z,E*)-germacradienyl cation (**60**), or *via* a 1,11-cyclisation to (*Z,E*)- humulenyl cation.[57]

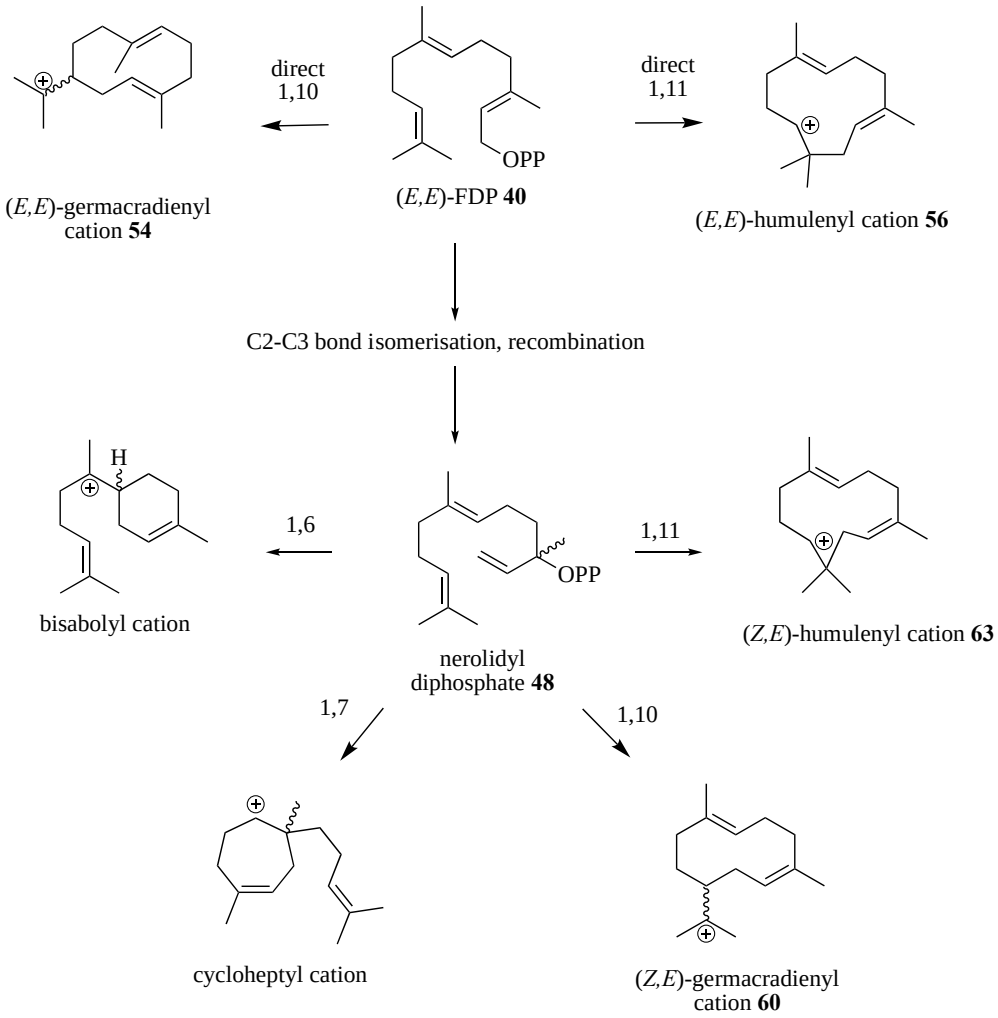


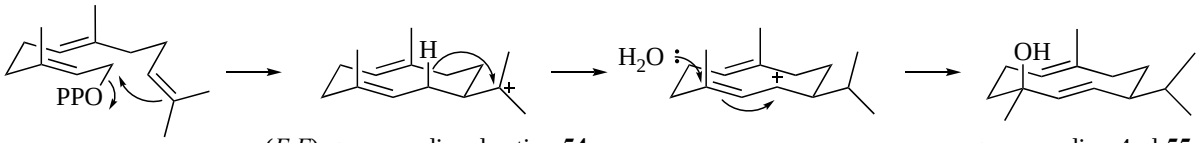
*Figure 20: Six initial cyclisation events of FDP (**40**)*

The mechanism of sesquiterpene synthases can be probed *via* various techniques, such as site directed mutagenesis, where various residues within the active site believed to impact the enzymes product profile and catalytic activity are mutated to alternative residues. One such example of this technique is the mutation of *epi*-isozizaene synthase from *Streptomyces coelicolor* A3, which produces *epi*-isozizaene with up to 99% fidelity, to produce sesquisabinene A with 97% fidelity *via* the mutation F96Q.[54] This method is discussed in greater detail in chapter 2.

Alternatively, analogues of the enzymes natural substrate can be used and the alternate product, if any is produced, can be used to deduce information reading the mechanism. Aza-analogues are nitrogen containing compounds that mimic the carbocation intermediates produced in sesquiterpene synthases but cannot be turned over by the enzyme. They can be used to provide evidence for the formation of carbocations at specific carbons in the skeleton such as the indication of the eudesmane cation as an intermediate in aristolochene synthesis.[58] Fluorinated analogues are also useful for identification of intermediate carbocations. Fluorine atoms are similar enough in size to hydrogen atoms, but are highly electronegative. Fluorinated analogues were used with germacradien-4-ol to indicate the initial steps of formation of the germacryl cation.[59] Isotopologues, such as deuterated analogues of the natural substrate, can be analysed by mass spectrometry and NMR to show hydride movement, aiding in the determination of mechanisms.[60] See pages 40-42 for an example.

Germacradien-4-ol synthase – 1,10-cyclisation via *(*E,E)*-germacradienyl cation (**54**)*

Germacradien-4-ol-synthase (Gd-4-ol) is found in *Streptomyces citricolor* and converts FDP (**40**) to germacradien-4-ol (**55**), using water to hydroxylate a carbocation. The catalytic mechanism begins with the ionisation of FDP (**40**) to a farnesyl cation, before 1,10-ring closure occurs to form the (*E*,*E*)-germacradienyl cation (**54**). A 1,3-hydride shift occurs, forming a stabilised carbocation, before the cation is quenched with a water molecule forming germacradien-4-ol (**55**), (Figure 21).[59]



(*E,E*)-germacradienyl cation **54** germacradien-4-ol **55**

Figure 21: Catalytic mechanism of germacradien-4-ol synthase

Pentalenene synthase – 1,11-cyclisation via *(*E,E*)-humulenyl cation (**56**)*

Pentalenene synthase is an enzyme which converts FDP (**40**) to pentalenene (**58**), which in *Streptomyces* is converted to a variety of antibiotic compounds in the pentalenolactone (**5**) family. Figure 22 shows the catalytic mechanism of this enzyme, which uses the farnesyl cation for cyclisation. FDP (**40**) is ionised and the diphosphate group is removed, the terminal double bond adds to the carbocation on C1 to form a humulenyl cation (**56**). A 1,2-hydride shift occurs, followed by a double bond rearrangement to form a tricyclic intermediate, the 7-protoilludyl cation (**57**).[61] A dyotropic rearrangement occurs between a methyl group and a hydrogen, in which the cation is not involved. Finally, proton abstraction occurs forming the product, pentalenene (**58**).[62]

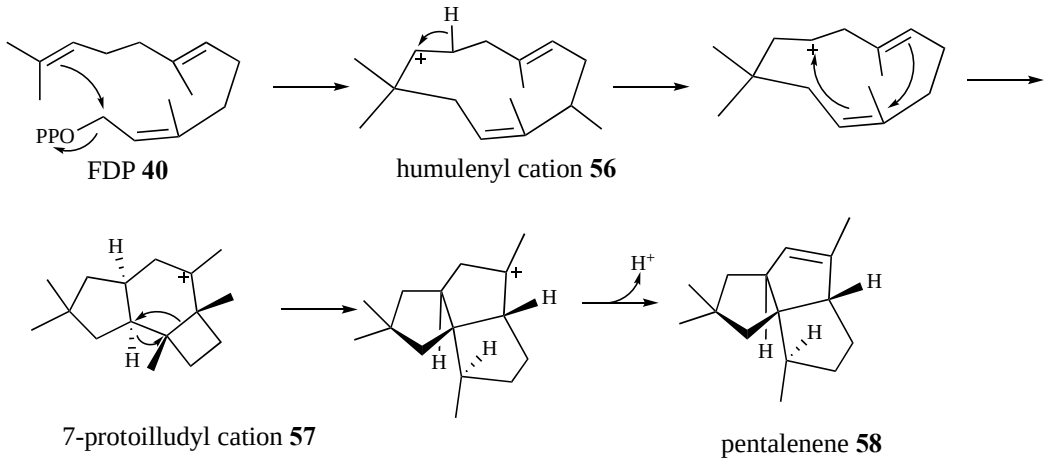


Figure 22: Catalytic mechanism of pentalenene synthase

Amorpha-4,11-diene synthase (ADS) – 1,6-cyclisation via *bisabolyl cation*

Amorpha-4,11-diene synthase (ADS) is a sesquiterpene synthase that converts FDP (**40**) to amorpha-4,11-diene (**34**), *via* the alternative cyclisation mechanism, (Figure 23), whereby the farnesyl cation isomerises to the *cis* nerolidyl cation. This begins with isomerisation to form nerolidyl diphosphate (**48**).[26] The formation of NDP (**48**) allows the C2-C3 double bond to rotate forming the *cis* conformation. NDP (**48**) then undergoes ionisation to form a 2,3-*cis*-farnesyl cation, and a 1,6-ring closure occurs to form a bisabolyl carbocation intermediate.[26,63] Following this, a 1,3-hydride shift occurs. Next, a ring closure occurs in a 1,10-cyclisation, and a cation forms at C11. Finally, a proton is lost from C11 to form a double bond between C11 and C13, to form amorpha-4,11-diene (**34**). This enzyme converts FDP (**40**) to amorpha-4,11-diene (**34**) in one step, creating two six-membered rings, four stereo-centres and two double bonds, showcasing precise stereo and chemoselectivity.[26]

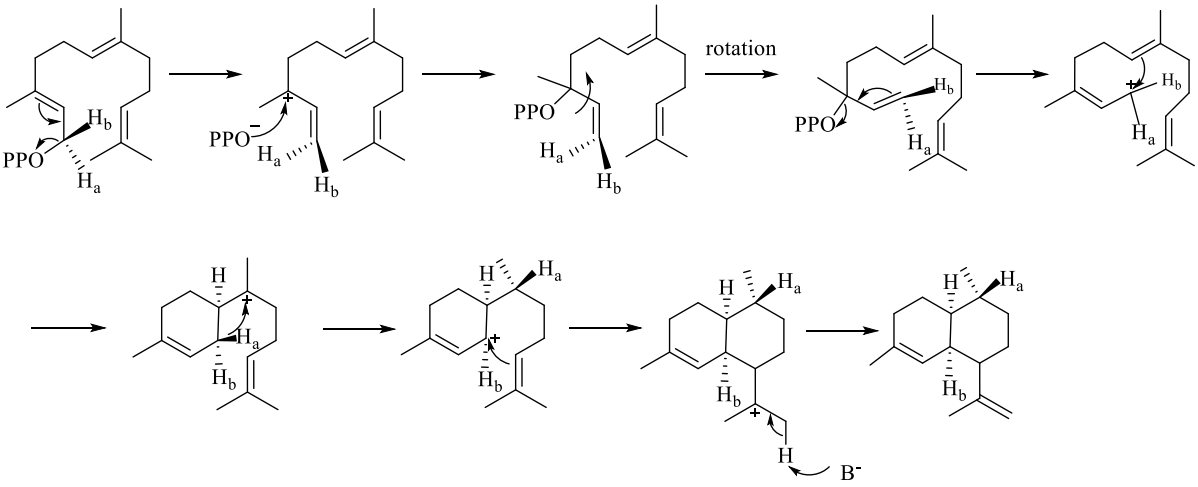


*Figure 23: Catalytic mechanism of FDP (**40**) to amorpha-4,11-diene (**34**)*[26]

*Daucene (**59**) - 1,7-cyclisation* via *cycloheptenyl cation*

Daucanes are a class of sesquiterpenes derived from plant, fungal and marine organisms, also named carotanes in the literature. FDP (**40**) is initially ionised and isomerised to NDP (**48**), allowing a rotation of the C2-C3 bond into the *cis* conformation. The 2,3-*cis*-farnesyl cation can then undergo a 1,7-ring closure to form a cycloheptenyl ion. A cyclopentane ring then forms, and 1,2-hydride shifts with proton loss allows the formation of daucene (**59**), (Figure 24), one example of the family of daucanes. Daucanes show a range of activities, such as fulvoferruginin which has anti-fungal activity, and lasidiol angelate has repellent activity towards leafcutter ants.[64]

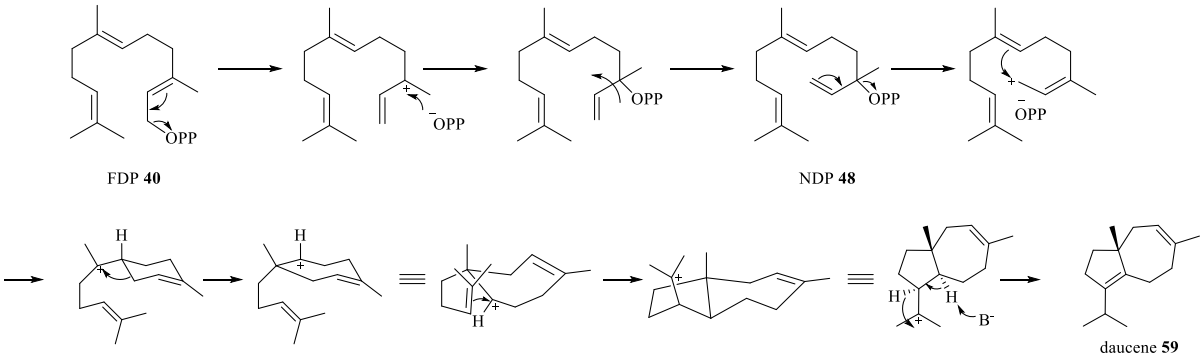


*Figure 24: Catalytic mechanism of daucene (**59**) formation from FDP (**40**)* [64,65]

*δ-Cadinene (**62**) - 1,10-cyclisation* via *(*Z,E*)-germacradienyl cation (**60**)*

δ-Cadinene synthase is a sesquiterpene synthase found in cotton and is a precursor of compounds such as hemigossypol which the plant uses for defence against pathogenic fungal species. The putative catalytic mechanism of δ-cadinene synthase is shown in Figure 25, beginning with the isomerisation of FDP (**40**) to NDP (**48**). The formation of NDP (**48**) allows the C2-C3 double bond to rotate to the *cis* conformation. NDP (**48**) then undergoes ionisation to form a 2,3-*cis*-farnesyl cation, before a 1,10-cyclisation occurs to form the (*Z*,*E*)-germacradienyl cation (**60**). This is followed by a 1,3-hydride shift, before a second cyclisation to form the cadinanyl cation (**61**). The cation is finally deprotonated to form δ-cadinene (**62**).[66]

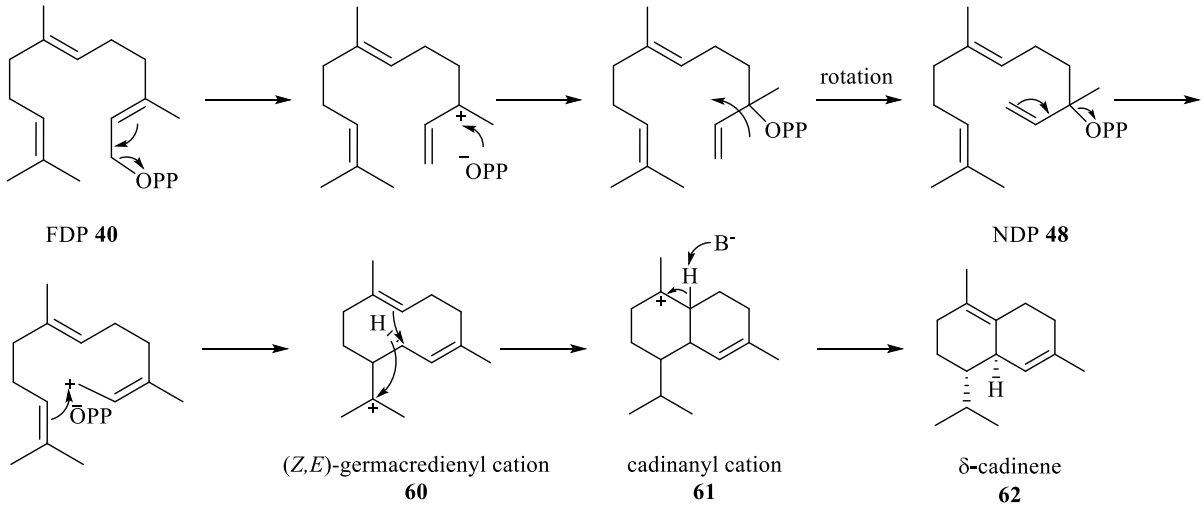


Figure 25: Putative catalytic mechanism of δ-cadinene synthase[66]

(+)-Longifolene synthase - 1,11-cyclisation via *(*Z,E*)-humulenyl cation (**63**)*

(+)-Longifolene synthase was first identified in a Norway spruce tree, and its major product, 60.6%, is longifolene (**64**). It is believed longifolene (**64**) is one of many terpene released by the tree to protect against pathogens.[67] The catalytic mechanism, (Figure 26), begins with the isomerisation of FDP (**40**) to the 2,3-*cis*-farnesyl cation, before a 1,11-cyclisation occurs to form the (*Z*,*E*)-humulenyl cation (**63**). A 1,3-hydride shift occurs,[68] followed by a 3,7-ring closure. A deprotonation occurs at the methyl group to form longifolene (**64**).[69]

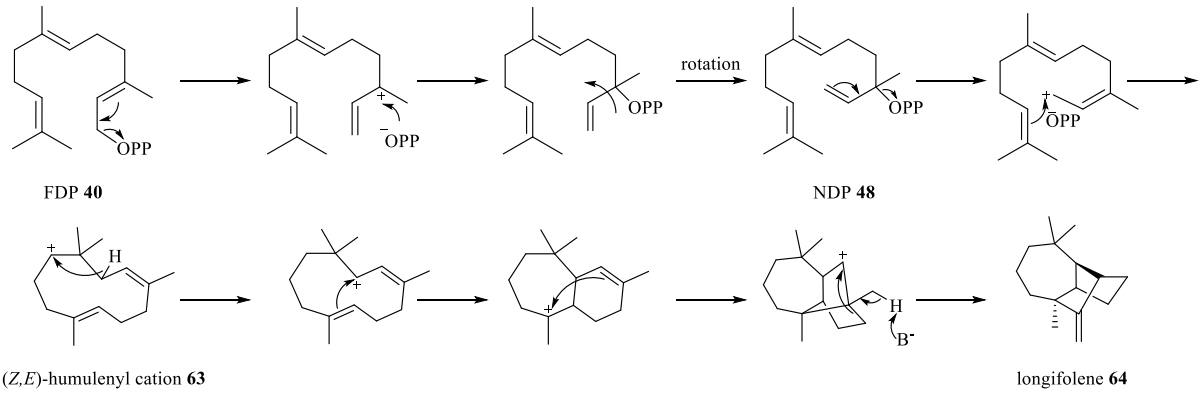


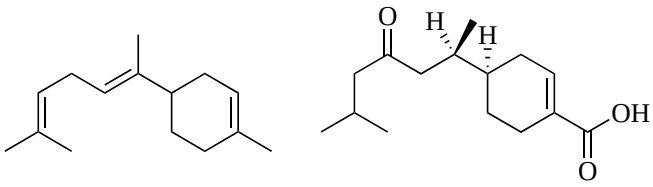
Figure 26: Catalytic mechanism of longifolene synthase[68]

1.3. Activities of sesquiterpenes

Sesquiterpenes are known to have a broad range of useful activities, including insect repellents,[9] aromatics,[70] and medications.[71]

Semiochemicals

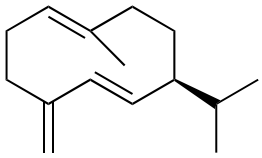
(*E*)-α-bisabolene synthases produces (*E*)-α-bisabolene (**65**) as a sole product in the grand fir tree. (*E*)-α-bisabolene (**65**) is released only when the tree is wounded, for example when an insect feeds on the tree. Downstream, this compound is believed to be used to produce todomatuic acid (**66**), (Figure 27), whose activity mimics juvenile hormone preventing larvae from maturing. This compound has also exhibited antifungal properties.[72]



(*E*)-a-bisabolene **65** todomatuic acid **66**

*Figure 27:(E)- α -bisabolene (**65**) and todomatuic acid (**66**).*

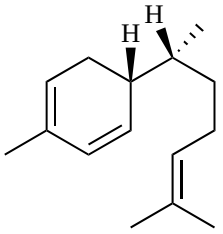
(-)-Germacrene D synthase produces (-)-germacrene D (**67**), (Figure 28), in many plant species, such as *Hemizygia petiolata*,[73] and *Solidago canadensis*.[41] When (-)-germacrene D was tested with grain aphids at naturally occurring concentrations, it was found to be a repellent.[73] Contrastingly, in experiments with the females of the moth *Heliothis virescens*, (-)-germacrene D (**67**) was found to be an attractant.[74]



(-)-germacrene D **67**

*Figure 28:(-)-Germacrene D (**67**)*

7-*Epi*-zingiberene synthase is an unusual sesquiterpene synthase obtained from the wild tomato plant, *Solanum habrochaites*. Unlike most sesquiterpene synthases, its substrate is (*Z,Z*)-FDP (**39**) instead of the (*E,E)* (**40**) form. The sequence of this enzyme is also unusual, as it seems more similar to a diterpene synthase than a sesquiterpene synthase. However, the product 7-*epi*-zingiberene (**8**), (Figure 29) is a sesquiterpene which displays insect repellent activity. Use of 7-*epi*-zingiberene (**8**) reduces new growth of whitefly and spider mites, two major agricultural pests. In future the genes for the production of 7-*epi*-zingiberene (**8**) may be inserted into commercial tomato species as a natural insect repellent.[9]



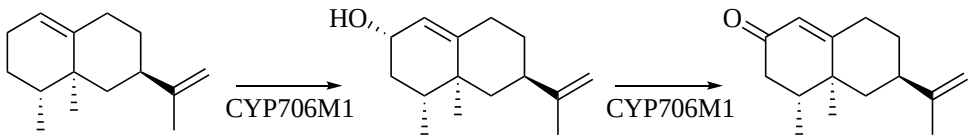
7-*epi*-zingiberene **8**

*Figure 29: 7-*epi*-zingiberene (**8**)*

Fragrances

Many components of the terpene family contain fragrant compounds long used by humans in perfumes and essential oils. Valencene (**68**) is the only sesquiterpene synthesised by valencene synthase and is obtained in low quantities from citrus fruit. Valencene (**68**) accumulates in oranges as they mature and is responsible for the typical 'orange' scent.[70] It is an inexpensive compound used in flavouring as well as a fragrance.

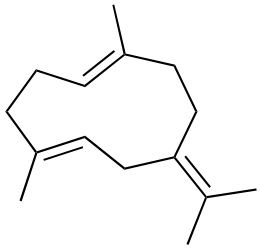
(+)-Nootkatone (**70**) is a sesquiterpene initially isolated from the cypress tree but is also present in grapefruit. (+)-Nootkatone (**70**) is obtained from the oxidation of valencene (**68**) catalysed by CYP706M1, *via* nootkatol (**69**). (+)-Nootkatone (**70**), (Figure 30), demonstrates insect repellent activity against termites,[75] but is primarily used in the food and cosmetic industry.[76] There is much interest in biologically engineering microbes to produce (+)-nootkatone (**70**).[76]



(+)-valencene **68** trans-nootkatol **69** (+)-nootkatone **70**

Figure 30: Sesquiterpene aromatics

Germacrene B (**71**), as shown in figure 31is a sesquiterpene with a sweet-geranium like odour and contributes to the scent of lime peel.[77] Germacrene B (**71**) is synthesised by germacrene B synthase, whose cDNA has been obtained from *Curcuma zedoaria* of the Zingiberaceae family.[78] Generally research regarding germacrene B (**71**) has analysed this compound as a component of essential oils, claiming it demonstrates antimicrobial activity. One such example is the essential oil of the flowers of *Aloysia gratissima*, which showed antimicrobial activity against *P.aeruginosa* and *C. albicans.*[79] Isolation of germacrene B (**71**) as a pure compound would allow a definitive answer on its antimicrobial properties.



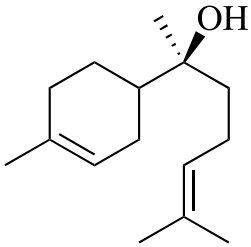
germacrene B **71**

*Figure 31: Structure of germacrene B (**71**)*

Medications

α-Bisabolol (**72**) (Figure 32) is the main product arising from α-bisabolol synthase catalysis and is found in a variety of plants including *Artemisia annua, Lippia dulcis,*[3] and *Matricaria chamomilla*, the latter more commonly known as chamomile.[80] It shows antimicrobial properties against the bacterial species *Solobacterium moorei*, one cause of halitosis, and is present in some oral healthcare products.[80] It has also been shown to inhibit pancreatic cancer cell growth, induce apoptosis,[81] and reduce invasiveness.[82,80] It has been investigated as a pain reliever in rodents, reducing nervous excitability. The suggested mechanism is blocking voltage-dependent sodium channels which would normally open to cause a rise in membrane potential in nerve impulses.[83] It has been demonstrated that the use of α-Bisabolol (**72**) can

reduce inflammatory cytokine release and increase anti-inflammatory cytokine release in mice with sciatic nerve lesions.[84]



α-bisabolol **72**

*Figure 32: Structure of α-bisabolol (**72**)*

Epi-isozizaene (**53**) is a sesquiterpene precursor of the antibiotic albaflavenone (**75**), produced in the bacteria *Streptomyces coelicolor*. *Epi*-isozizaene synthase converts FDP (**40**) to *epi*-isozizaene (**53**), which is then oxidised to albaflavenone (**75**), by CYP170A1 *via* (4*R*/*S*)-albaflavenol (**73-74**) (Figure 33)[85] Albaflavenone (**75**) demonstrates antibiotic activity against *Bacillus subtilis*.[86]

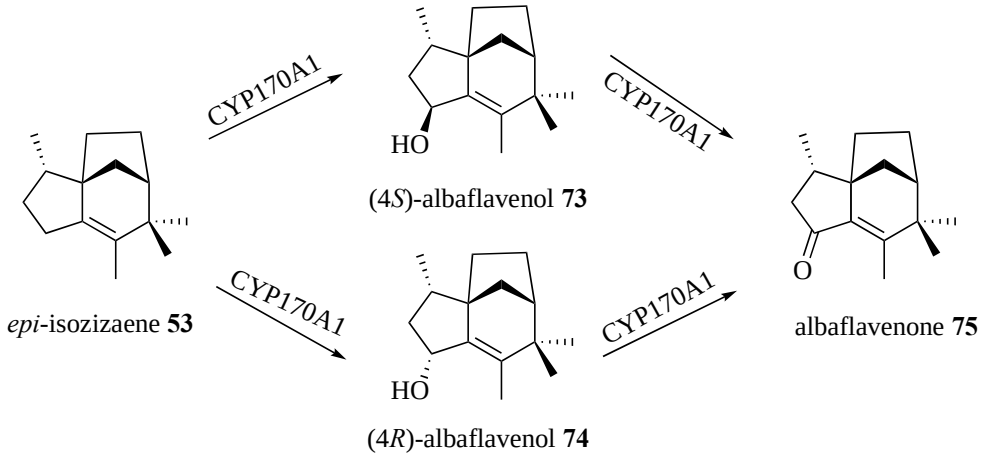
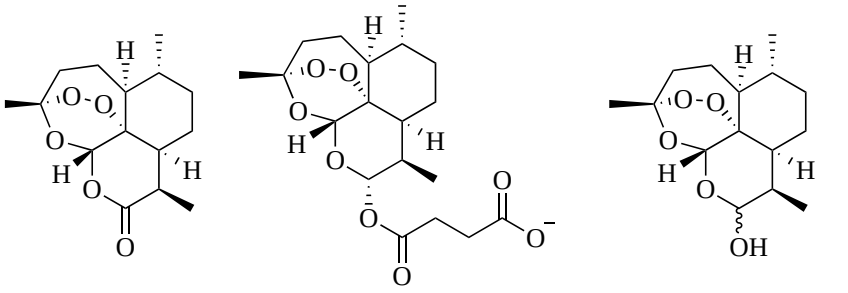


*Figure 33:Biosynthetic pathway of albaflavenone (**75**)*[85]

Amorpha-4,11-diene (**34**) is a sesquiterpene synthesised from ADS, found in the plant *Artemisia annua L.* Amorpha-4,11-diene (**34**) is a precursor to the antimalarial drug artemisinin (**4**). ADS is of great importance as artemisinin (**4**) is such a vital medication and extraction from the plant is very low yielding (around 0.1% dry weight of the plant)[87] Artemisinin (**4**) behaves as an antimalarial due to the breakage of its endoperoxide bridge forming free radicals. There is little agreement on how exactly these radicals kill malarial parasites but the bridge is necessary for antimalarial activity.[88]

Artemisinin (**4**) is often modified to more soluble derivatives such as dihydroartemisinin (**77**) and artesunate (**76**) (Figure 34). Although artemisinin (**4**) and its derivatives are known for

antimalarial activity, there is interest in its anticancer properties. Artemisinin (**4**) has been shown to induce apoptosis in human hepatoma cells, and disrupt the cell cycle, but cytotoxicity to normal liver cells was much lower. [89] Dihydroartemisinin (**77**) has been used in *in vitro* experiments with human metastatic melanoma A375 cells, showing that apoptosis was induced *via* upregulation of selective apoptotic genes. Simultaneously the study showed that dihydroartemisinin (**77**) did not induce apoptosis in healthy human melanocytes.[90] Dihydroartemisinin (**77**) has also been used to reduce cell viability of human ovarian carcinoma OVCAR-3 cells, while non-tumourigenic ovarian epithelial cells were far less affected. [91] Artesunate (**76**) has been used in *in vitro* studies against human epidermoid carcinoma A431 cells and was demonstrated to inhibit cell growth. It was demonstrated that artesunates (**76**) activity comes from inhibiting the cell cycle and inducing apoptosis in an iron dependent manner. [92]



artemisinin **4** artesunate **76** dihydroartemisinin **77**

*Figure 34: Structure of artemisinin (**4**) and derivatives*[89]

As is clear from a few examples of useful sesquiterpenes these compounds are often modified from the basic carbon backbone structure obtained from FDP. These modifications are carried out by a variety of enzymes but often by a class of decorating enzymes called cytochromes P450.[93]

1.4. Cytochromes P450

Cytochromes P450 are a family of enzymes that have a key heme ligand aiding complex chemical reactions and are referred to as CYPs. All P450s contain type-*b* heme (Figure 35): a protoporphyrin IX ring structure which is a planar molecule with an iron ion in the centre. This iron coordinates to four nitrogen atoms and a thiol group from a cysteine residue. The sixth ligand changes depending on the environment and the step of the P450 catalytic cycle.[94] CYPs are found in all types of life, from bacteria to mammals, but not in every species. CYPs are

involved in drug metabolism, production of sterols, various hormones and many natural products.[95] Likely one of the first descriptions of P450 proteins occurred in experiments by Garfinkel in 1958, as he isolated microsomes from pig liver and looked at their composition and the haemoglobin 'contaminant'. Here the 450 nm absorbance peak that P450 is named for was described.[96]

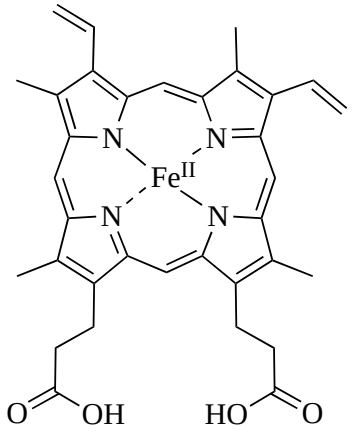


Figure 35: Structure of type b heme [97]

P450s have a distinctive maximum absorbance at 450 nm when bound to carbon monoxide in the presence of a reducing agent, which is called a Soret peak. This is due to the thiolate ligand of the cysteine interacting with the carbon monoxide.[98] P450 proteins produce a variety of difference spectra depending on the environment and the stage of the P450 cycle. A type I spectral change can be seen when the substrate displaces the water ligand, and the spectrum has a maximum at 385 nm and a minimum at 420 nm. A reverse of the type I with a maximum at 420 nm and the minimum at 385 nm is caused by the substrate leaving and the water ligand returning. A type II spectrum has a maximum at 425 – 435 nm and a broad minimum between 390 – 410 nm. The maxima and minima of spectra can shift depending on the environment and the type of ligand, and can be due to the heme dropping from the high spin state to the low spin state, where the heme bonding is perturbed, or due to the low spin state becoming perturbed.[98] The heme group of the P450 normally interacts with the protein *via* a coordinate bond between the iron ion and the thiolate group of the conserved cysteine residue. However, there are some exceptions to this, such as a few members of the CYP4A, CYP4B and CYP4F family which bind the heme in a covalent fashion *via* an base-labile ester linkage to a glutamic acid residue.[99]

P450 enzymes commonly perform hydroxylation reactions, general formula in Figure 36, but they can also add other functional groups such as methyl or nitrogen containing groups. P450s have fascinated scientists for decades as they are able to insert oxygen into a carbon hydrogen

bond, a process considered difficult using synthetic chemistry.[100] The reaction should not be possible as it is spin forbidden, but by complexing the oxygen molecule to a metal this issue is overcome allowing it to react with hydrocarbons.[101] CYPs also generally possess high regioselectivity, allowing a more controlled reaction than *via* chemical synthesis.[100]

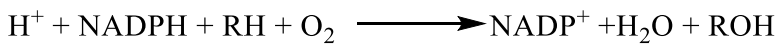


Figure 36: General equation of P450 activity

Nomenclature of P450s

P450s are named CYP followed by a number, then a letter and then another number. The numbering system is based on amino acid sequence similarity and was first introduced in 1987 by Nebert *et al.*[102] The first number signifies the family in which the P450 belongs and members of these families must have greater than 40% identity. The letter signifies the sub-family and must share greater than 55% identity. The final number is arbitrarily assigned and signifies the individual protein, which is therefore unique.[103] There are over 200 known P450 families, a number which is continually growing.[104]

CYPs generally share low sequence identity, comparative to many other enzymatic families. The general structure of these enzymes involves 14 α-helices designated A-L, and five β-sheets, some of which are shown in Figure 37 using CYP124A1 as an example.[105] P450s share a common fold containing four helical bundles: D, L and I run parallel, and the fourth, E, is antiparallel. This fold binds the heme group that is common to all CYPs, hence the similarity. The heme group is in between the I and L helix and a loop containing the key cysteine reside is also bound to the heme. This cysteine residue acts as a ligand to the iron ion in the middle of the heme,[95] while nitrogens of the porphyrin ring act as four more ligands and water is the last.[106] This loop containing the cysteine has a signature amino acid sequence, FxxGx(H/R)xCxG. This key cysteine residue is also part of producing the characteristic absorbance band. The I helix contains a sequence of conserved amino acids, (A/G)-Gx(E/D)T, that is believed to be involved in catalysis and sits in the active site, and this conserved sequence causes a kink in the middle of the helix. The J and C terminal of the K helix are also conserved, and the C terminus of the K helix contains the conserved sequence ExxR. Two β-sheets, β1 and β2, are also conserved. [105] CYPs accept an incredibly broad range of substrates, and it is believed that key 'substrate recognition sites' determine what the enzyme turns over.[95]

The most diverse areas of P450s are normally the A, B', B, F, G, H, and K' helix, the β3 and β4 β sheets, and the unstructured loops. Six sites in the protein are believed to be involved in substrate recognition. The first, SRS1, is in the B'-helix region; SRS2 involves the F helix, SRS3 involves the G helix; SRS4 involves part of the I-helix; SRS5 involves the β4 hairpin and SRS6 is the K-helix β2 connecting region. These areas have been identified *via* point mutation studies and they also move when substrates bind to CYPs implying they are flexible regions accommodating the substrate.[105] Some of these structures are labelled in Figure 37.

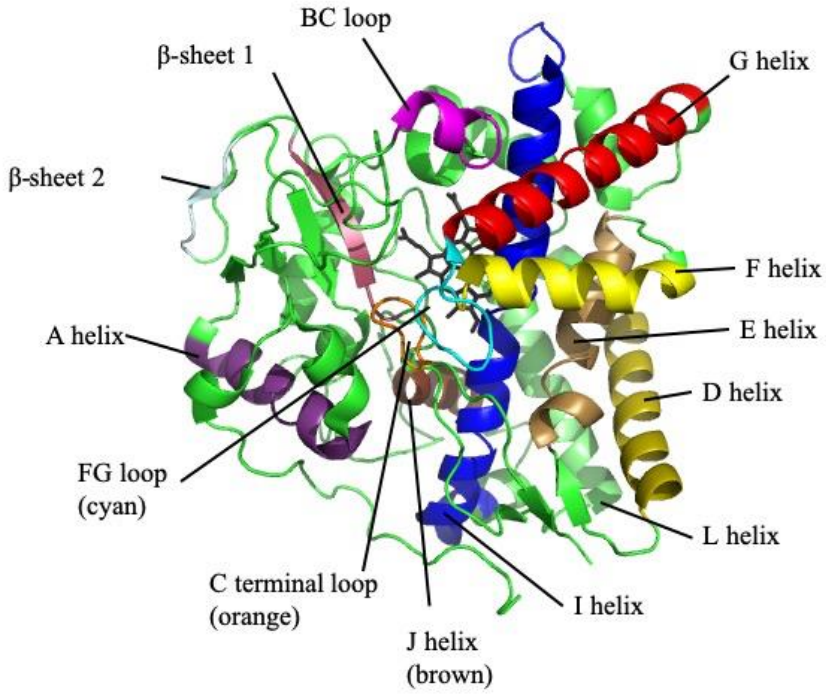


Figure 37: Cartoon representation of CYP124A1 (PDB:2WM4) with heme shown in black[107]

Classification of Cytochromes P450

P450s are divided into four classes determined by how electrons are transported from NADPH to the heme, as summarised in Figure 38. Class I are a three component system which use a reductase protein containing FAD and an iron sulfur redoxin.[104] These are most commonly bacterial enzymes and include CYP101A1, a well-studied P450.[94] Class II are a two-component system and only require the reductase proteins which contains FAD or FMN. CYP102A1 is an example of this class, which are most commonly found in eukaryotes and are therefore usually membrane bound.[104] Class III do not need an electron donor or oxygen and are self-contained systems as they react on compounds produced by dioxygenases.[104] These

have the diflavin reductase fused to the P450.[108] Class IV obtain their electrons directly from NADH and an example is CYP55A1.[104] These contain a reductase with FMN, an iron sulfur redoxin and the P450 in one protein.[108] As more P450s have been discovered additional classes have been named and therefore alternative designations of classes exist in the literature.[94]

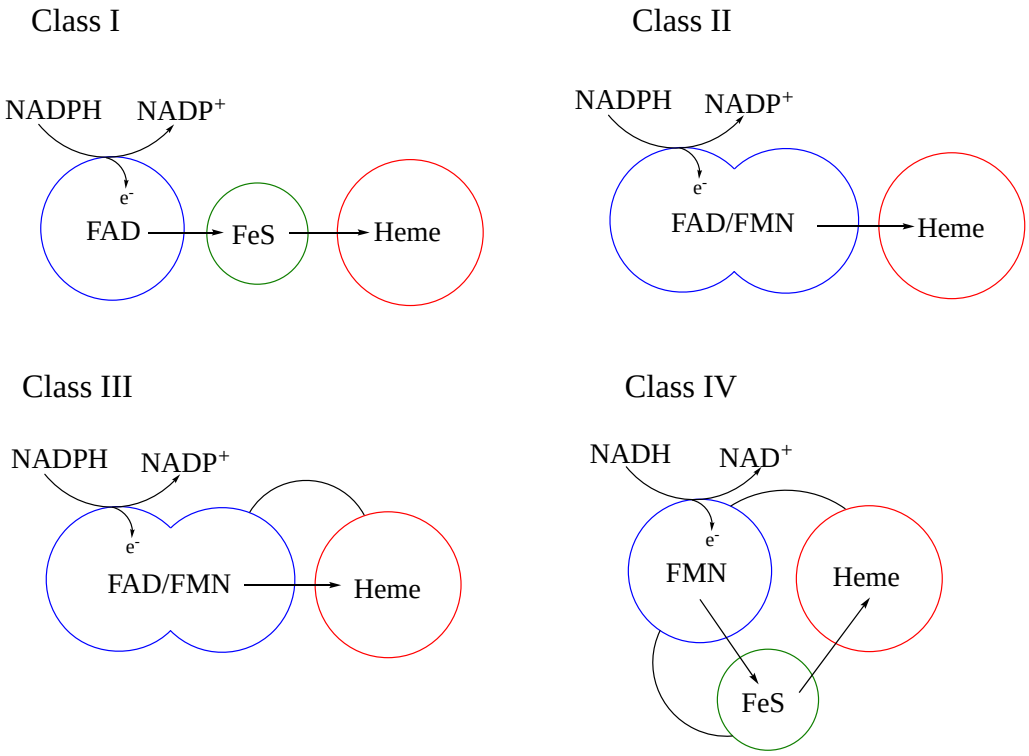


Figure 38: Schematic of the components of each class of P450[104] with the flavin containing components outlined in blue, the heme containing components outlined in red, and the iron-sulfur components outlined in green.

The cytochrome P450 cycle

P450s complete their reaction *via* the P450 catalytic cycle in which the iron ion is key (Figure 39). Normally, the iron ion is bound to four nitrogen atoms, a thiol group of a cysteine residue and the sixth ligand is water which is displaced by the substrate. In this initial resting state, the iron ion is in a low spin state. As the substrate binds and displaces the water ligand, there is often a change (or partial change) to a high spin state, and the ion shifts from a six coordinated state to a five coordinated state. The iron (II) moves out of the plane of the ring and is destabilised by the loss of the sixth ligand leading to the shift in redox potential. Various substrates cause different shifts in this redox potential.[95]

The shift in redox potential means the reduction can occur more readily as an electron originating from NADPH is donated. This stops NADPH being wasted in the system as a substrate must be present.[95] A molecule of oxygen then joins and binds to iron (III) forming an oxy-P450 complex which is hexa-coordinated. This is a diamagnetic complex and is a relatively stable intermediate in the cycle. An X-ray crystal structure of this intermediate in CYP101 shows that the dioxygen does not interact sterically with the substrate.[95]
Another electron is donated *via* a reduction forming a peroxo-ferric intermediate which is far less stable. In this intermediate, the iron ion is displaced out of the plane of the porphyrin ring of heme. One of the oxygen atoms is bound to a proton forming a hydroperoxo-ferric intermediate, which is a fast reaction. This intermediate was visualised by cryoradiolytic reduction and EPR. [95]

A second proton joins, directed by a highly conserved threonine residue near the dioxygen binding site,[109] triggering the release of H_2O *via* heterolysis of the O-O bond where an electron is donated from the porphyrin and the iron (III) oxidises to iron (IV).[95] This donation leads to the formation of a cation radical on the porphyrin ring. This intermediate is commonly referred to as compound I and is very reactive. It is a high-valent iron-oxo intermediate and has been visualised by absorption spectra[95] *via* comparisons with chloroperoxidase[109] and cryo-X-ray crystallography but never under native conditions.[95] Compound I removes a proton from the substrate to form a hydroxy species called compound II which then binds with the substrate radical and forms the hydroxylated product.[109] Finally, the hydroxylated substrate is released, the iron (IV) is reduced to iron (III) and water joins as the sixth ligand regenerating the enzyme.[95]

Where exactly the charge is located on the complex throughout the cycle is ambiguous and is represented in different ways in the literature: it can be represented on the iron ion, on the iron oxygen bond or delocalised across the porphyrin ring. [101]

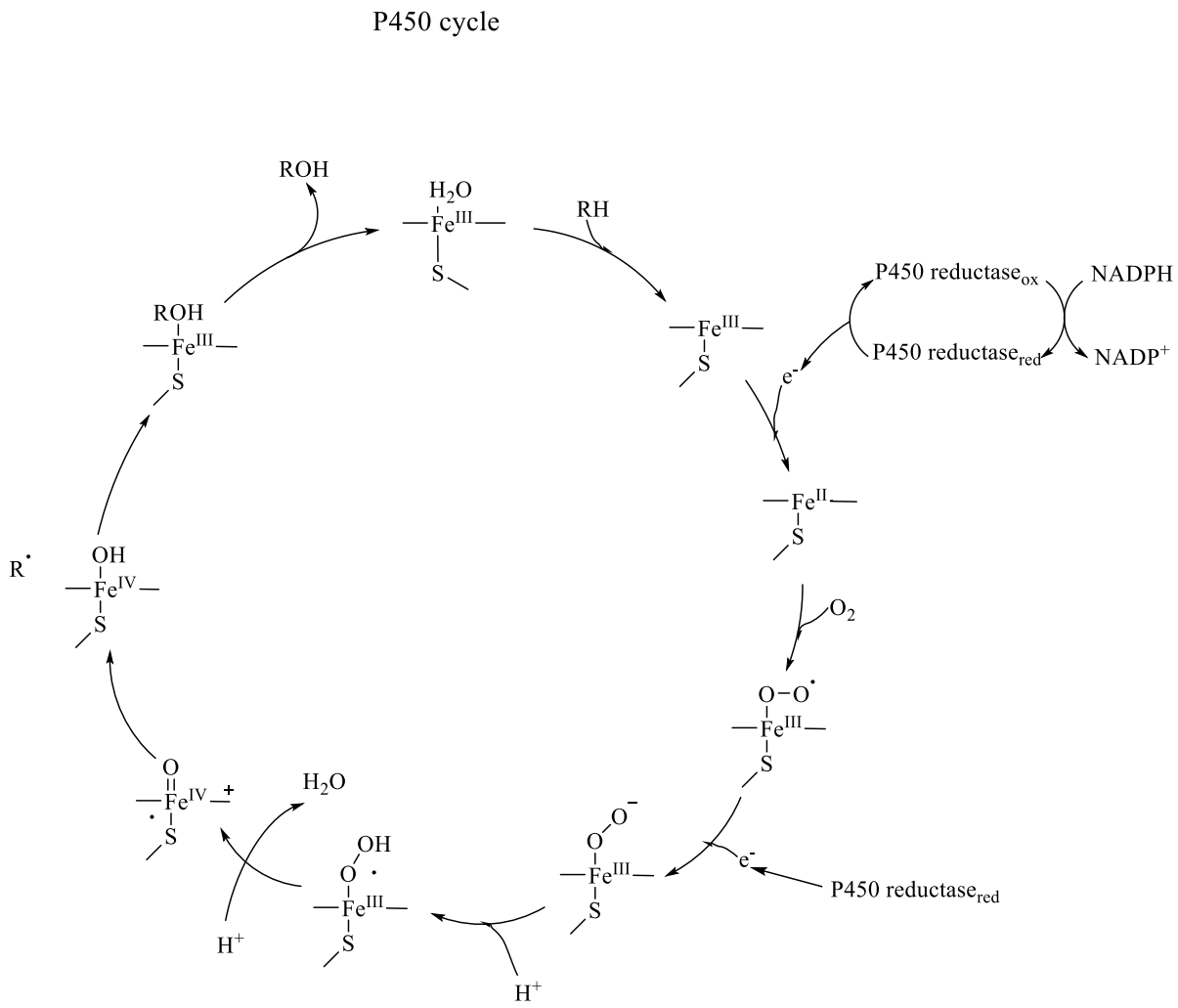


Figure 39: P450 cycle[110]

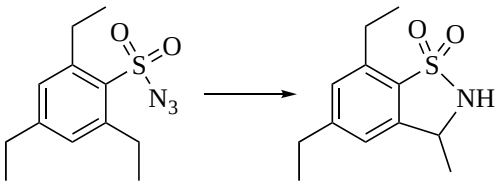
Cytochrome P450 reductases

Electrons used in the reduction steps of the cycle often come from NADPH and can be delivered to the active site by accessory proteins. These are split into class I and class II proteins.[106] Class I encompasses ferredoxin reductase, ferredoxins, flavodoxins, adrenodoxin reductase or adrenodoxin, all of which can be involved in delivering electrons.[111] Class I are more common with soluble P450 enzymes or with mitochondrial membrane reductases. Class II redox proteins are membrane bound and contain two domains in the same polypeptide, both of which contain flavin, in the form of FMN in one and FAD in the other. Occasionally these accessory proteins will be in a single structure as separate domains with its respective P450. The most famous of these examples is CYP102A1.[106] Not all P450s require a reductase partner. Some use peroxides to jump straight to the hydroperoxo-ferric intermediate, allowing oxidation. Examples include CYP152B1 and CYP152A1 which oxidise long chain fatty acids.[111]

Diversifying and mutating P450s for novel uses

Much of P450 research involves investigating their natural behaviour e.g. identifying which P450 in an organism hydroxylates a drug to increase its solubility for excretion.[112] But there is research into using P450s as biocatalysts for alternative reactions. The traditional reaction associated with CYPs is the hydroxylation of organic molecules, although there are exceptions such as CYP170A1 which acts as a pyrophosphatase[113] and CYP1A2 and CYP2E1 which have phospholipase D activity.[114] More unusual activities can be engineered in CYPs such as the formation of C-N bonds from C-H bonds. Arnold *et al.* attempted mutations of P450BM3 to catalyse the conversion of 2,4,6-triethylbenzene-1-sulfonylazide (**78**) to benzosultam (**79**) (Figure 40). After several rounds of screening, a mutant with 73% yield of the desired product was obtained. It was shown that hemin alone catalysed the reaction, but gave a racemic mixture of benzosultam (**79**), whereas one of the mutant CYPs gave 73% enantiomeric excess. The enzymes structure clearly plays a significant role in the specificity of the reaction. These mutants are not technically P450s as they lack the conserved cysteine residue, which is mutated to serine. Arnold therefore named these mutants P411 as the Soret peak of the reduced CO spectrum is shifted to 411 nm, but they are derived from a P450. This work showed that CYPs already present in nature can be modified to complete reactions for which very few enzymes

are known to catalyse.[115] In the same vein, a mutant of P450BM3 converted carbonazidates to oxazolidinones while the wild type and hemin showed little activity.[116]



2,4,6-triethylbenzene -1-sulfonylazide **78** → benzosultam **79**

Figure 40: Reaction catalysed by mutant P450BM3[115]

1.5. Synthesis of Artemisinin (**4**)

Natural synthesis of Artemisinin (**4**)

Artemisinin (**4**) is an example of a sesquiterpene which is derived from the hydrocarbon amorpha-4,11-diene (**34**) and is decorated *via* various enzymes including a P450 to produce a medicinally active compound. In nature, the plant *Artemisia annua* or sweet wormwood produces the anti-malarial drug artemisinin (**4**) in its glandular secretory trichomes.[117] Artemisinin (**4**) was discovered in China in the 1970s by Prof. You-You Tu, a chemist. This discovery led to her winning the Nobel Prize for Medicine in 2015. She was looking at ancient remedies from Chinese history and discovered the anti-malarial activity of artemisinin (**4**) when extracted at low temperatures. Tu went on to identify the structure of the active compound extracted from *Artemisia annua* in 1975. Artemisinin (**4**) is a sesquiterpene lactone, and the structure was published in 1977.[118]

The first specific step of artemisinin (**4**) synthesis is catalysed by the sesquiterpene synthase ADS that converts FDP (**40**) into amorpha-4,11-diene (**34**) (Figure 41).[117]After amorpha-4,11-diene (**34**) is produced, it is hydroxylated by a member of the cytochrome P450 family, CYP71AV1, and concomitantly a molecule of O_2 is converted to a molecule of water. This produces artemisinic alcohol (**80**). The alcohol is oxidised to artemisinic aldehyde **(81)** by alcohol dehydrogenase 1 using NAD^+. The alkene is then reduced to dihydroartemisinic aldehyde (**82**) by artemisinic aldehyde Δ11(13) reductase using NADPH. Dihydroartemisinic aldehyde (**82**) is oxidised to dihydroartemisinic acid (**83**) catalysed by aldehyde dehydrogenase 1 using $NADP^+$ and water. Finally, dihydroartemisinic acid (**83**) is converted to artemisinin (**4**) using oxygen, catalysed by UV light.[117]

The initial work regarding artemisinin (**4**) biosynthesis involved some confusion. Amorpha-4,11-diene (**34**) was first discovered from the plant *Viguiera oblongifolia* and named cadina-4,11-diene.[119] Amorpha-4,11-diene (**34**) was identified as a likely precursor to artemisinin (**4**) in 1999 as a compound obtained from leaf extracts of *A. annua.* Its structure was confirmed as amorpha-4,11-diene (**34**) when compared to an authentic standard synthesised from artemisinic acid (**92**) for gas chromatography-mass spectrometry (GC-MS) identification. The sesquiterpene synthase which produces it, ADS, was identified at the same time. It was suggested that due to the high activity of ADS, but the low quantities of amorpha-4,11-diene (**34**) obtained from the plant, that it is likely modified further. Also, the conversion of FDP (**40**) to amorpha-4,11-diene (**34**) is the rate limiting step of artemisinin (**4**) synthesis as very little of the intermediates between FDP (**40**) and artemisinin (**4**) are obtained. All this made amorpha-4,11-diene (**34**) a viable candidate for a precursor of artemisinin (**4**). [120] Additionally, the biosynthetic pathway described above, (Figure 41), was not the first pathway hypothesised. Initially it was believed that artemisinic acid (**92**), which can also be extracted from plants such as *Artemisia annua,* was a precursor to artemisinin (**4**).[121]

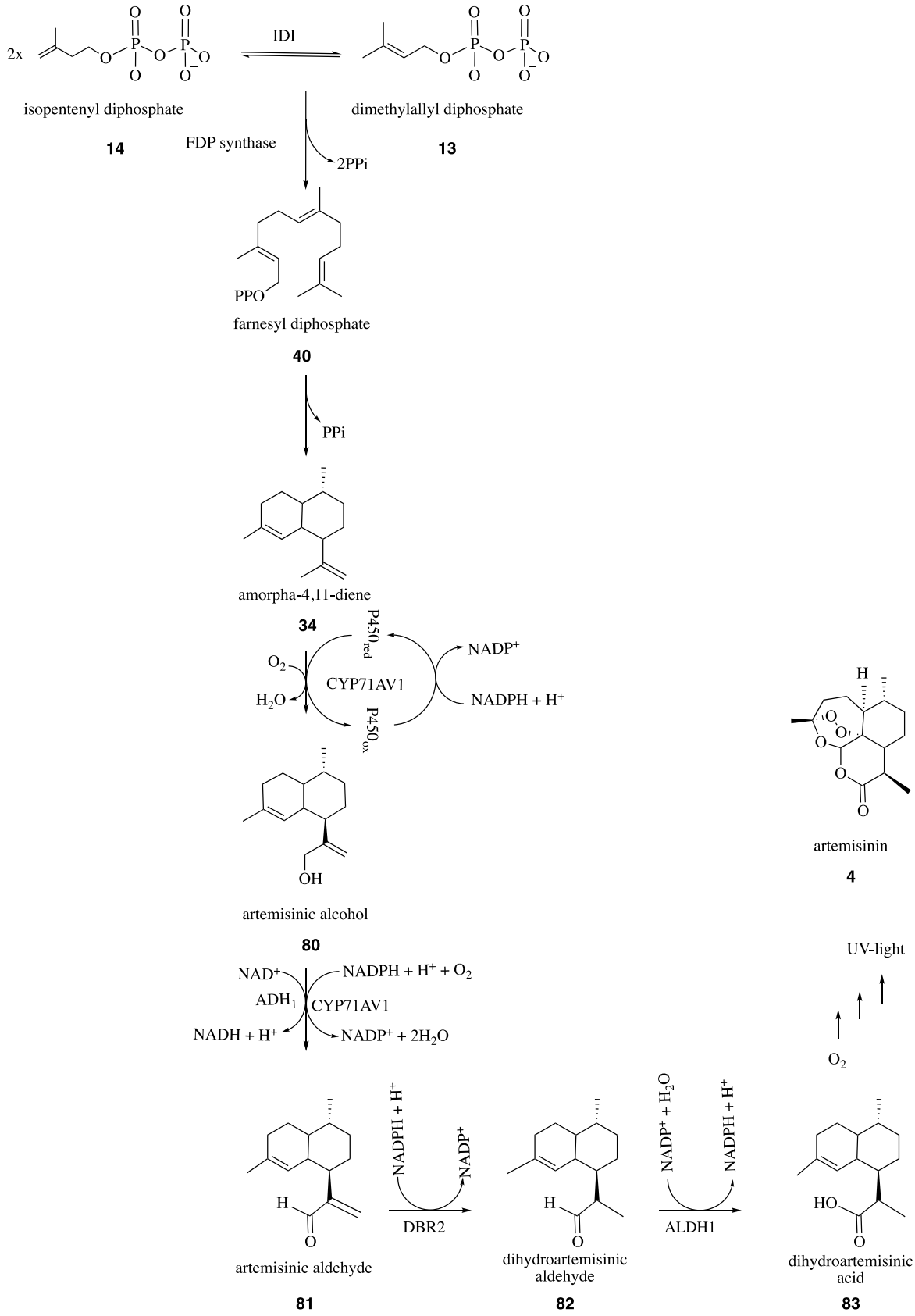


*Figure 41: Synthesis of artemisinin (**4**) from dimethyl allyl diphosphate (**13**) and isopentenyl diphosphate (**14**) in Artemisia anuua*[117]

The main product of ADS, amorpha-4,11-diene (**34**), makes up 91.2% of the total with the rest made up of hydrocarbons and alcohols such as γ-humelene (**84**) (1%) and amorpha-4-en-7-ol (**85)** (2.1%) whose structures are shown in Figure 42. [87] ADS, although recognised as a sesquiterpene synthase, does also accept GDP producing limonene (**3**), α-terpineol and terpinen-4-ol.[60]

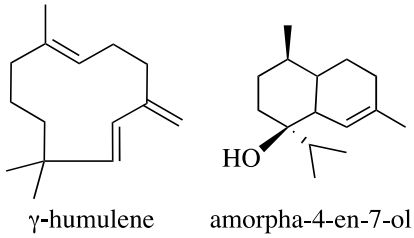


*Figure 42: Structure of γ-humelene (**84**), amorpha-4-en-7-ol (**85**)*

Discovery of the catalytic mechanism of amorpha-4,11-diene synthase

The enzyme amorpha-4,11-diene synthase was isolated and recombinantly expressed in *E. coli* in the early 2000s[87,122] and a catalytic mechanism was proposed.[87] It was suggested that the enzyme completes a 1,6-ring closure rather than a 1,10-ring closure as β-sesquiphellandrene was identified as a minor product. No intermediates suggesting a 1,10-ring closure such as germacrenes were produced. Using this information, along with the intermediates produced and the catalytic mechanism of *epi*-aristolochene synthase, a tentative catalytic mechanism was proposed, (Figure 43).[87]

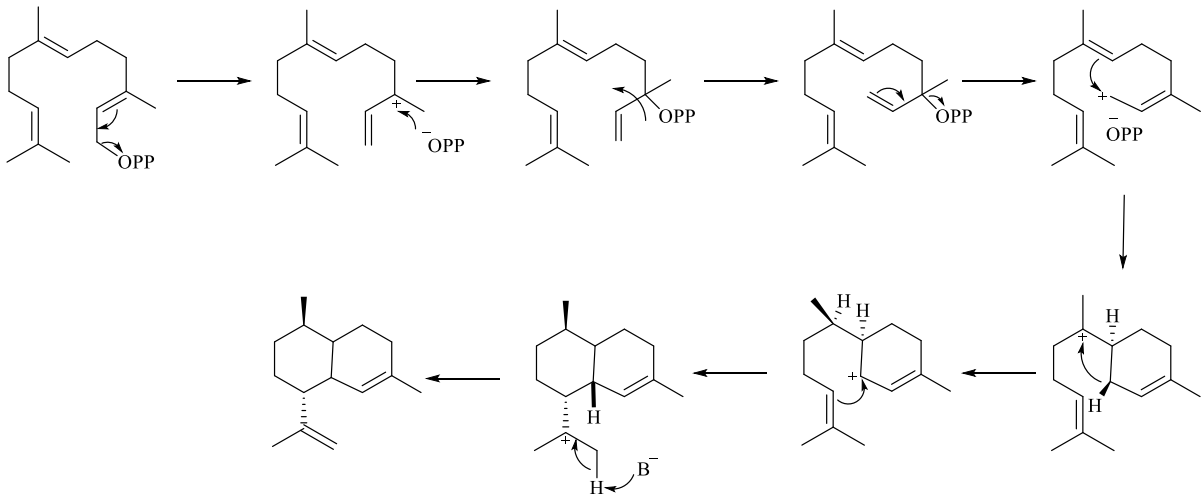


Figure 43: An early version of the proposed ADS catalytic mechanism[87]

By isoptopically labelling FDP (**40**), the substrate of sesquiterpene synthases, it is possible to identify the catalytic mechanism the enzyme follows. Deuterium labelled FDP (**40**) can be used to determine whether the hydrogens at certain positions are retained due to the mass ion being heavier, and if they are retained in their original position depending on which of the fragments are heavier. The position of the deuterium may also be identified by various NMR experiments.

[1,1-$^2H_2$]FDP (**86**), (1*S*)-[1-^{2}H]FDP (**87**) and (1*R*)-[1-^{2}H]FDP (**88**), were used as substrates for recombinantly produced ADS to produce amorpha-4,11-diene (**34**). The amorpha-4,11-diene produced from [1,1-$^2H_2$]FDP (**89**) was two atomic mass units heavier than that produced with non-labelled FDP (**40**) confirming that both deuterium atoms were retained in the final product. The mass fragments produced due to a loss of the methyl group (at position 12, 14 or 15 of FDP, numbering corresponds to Figure 44) were also two atomic mass units heavier than that produced with non-labelled FDP, suggesting no deuterium is located on any of the methyl groups.[60]

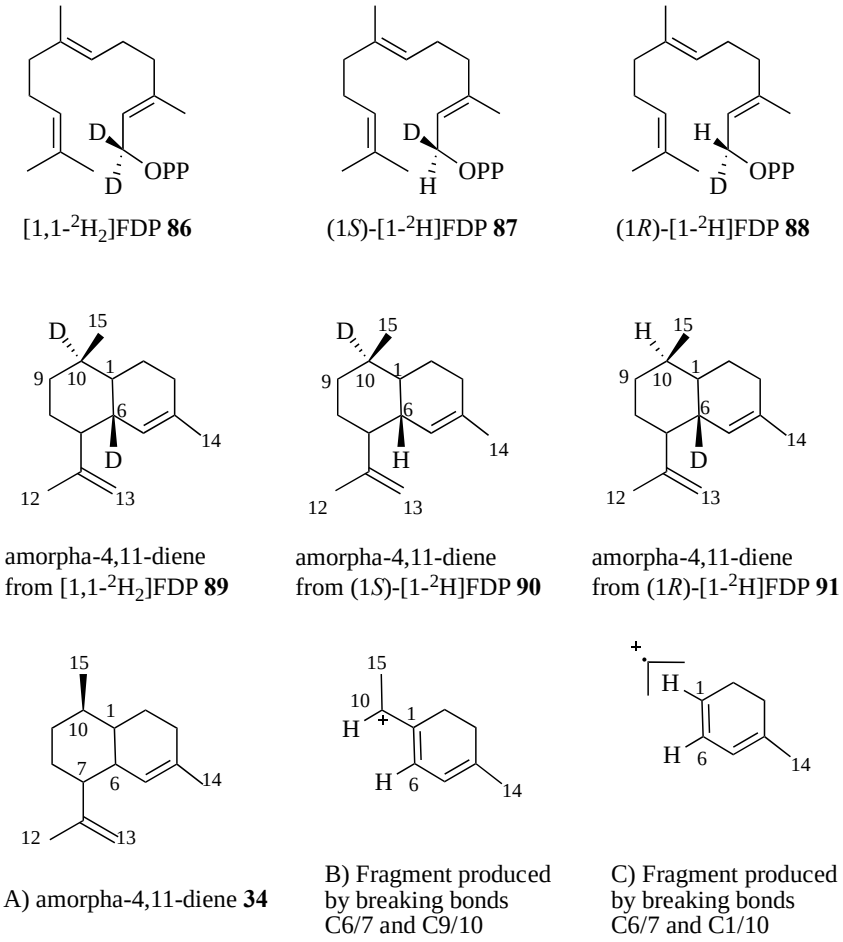


Figure 44: Structures of deuterium labelled FDP, the deuterium labelled amorpha-4,11-diene produced when the FDP was incubated with ADS, A) a numbered structure of amorpha-4,11-diene and B) and C) fragments mentioned that were produced via *mass spectrometry*

From mass fragmentation patterns, regardless of which labelled substrate is used, it was possible to detect deuteriums in fragment B, produced by the breaking of the bonds between C6/7 and C9/10. Additionally, the ion C which is formed when the bonds between C6/7 and C1/10 are cleaved, retains the deuterium when (1*R*)-[1-^{2}H]FDP (**88**) and [1,1-$^2H_2$]FDP (**86**) are used as substrates but not when (1*S*)-[1-^{2}H]FDP (**87**) is the substrate. This implies that the H-1*si*-proton from FDP moves to carbon 10 of amorpha-4,11-diene (**34**) during the catalytic mechanism.[60] These results were corroborated with proton NMR and COSY experiments. The proton NMR of amorpha-4,11-diene obtained from (1*R*)-[1-^{2}H]FDP (**91**) showed no peak for H6 and the product of (1*S*)-[1-^{2}H]FDP (**90**) incubated with ADS showed no peak for H10 as the respective carbons have a deuterium resulting in a loss of signal compared to the native substrate. The amorpha-4,11-diene analogue obtained from incubating ADS with [1,1-$^2H_2$]FDP (**89**) showed no peak for H6 and in the COSY no cross-peak between H10 and H15. This shows that the deuterium atoms in the product were located at C6 and C10 as the H6 proton peak had been lost due to the deuterium present, and there was no coupling between protons at positions C10 and C15. [60] The results obtained by Picaud *et al.* were confirmed shortly after by Kim *et al*.[26]

Commercial production of artemisinin (**4**)

Extraction from Artemisia annua

Initially artemisinin (**4**) was extracted from the plant in which it was discovered, *Artemisia annua.* Various solvents were used for the extraction of artemisinin (**4**) such as ethanol, water, dichloromethane, chloroform and petroleum ether. A study conducted in the USA on air-dried *A. annua* showed that petroleum ether was the most selective solvent of choice for extraction of artemisinin (**4**). Various parts of the plant were extracted in solvent between 30 – 60 °C for 48 hours. This extract was then further extracted with acetonitrile and chloroform to remove waxes and then purified using silica chromatography. Using this method, the authors obtained a 0.06% yield from the leaves and flowers of the plant.[123] This yield was improved upon by using a slightly different method at a much larger scale, beginning with 400 kg of leaves. The leaves were extracted with hexane for 48 hours, the hexane evaporated, and residue dissolved in fresh hexane before filtering. The hexane containing soluble compounds was then washed with aqueous acetonitrile. Water was removed and artemisinic acid (**92**) partially removed *via*

crystallisation. The rest of the extract was purified by silica chromatography and the artemisinin (**4**) recrystalised to obtain pure compound with 0.12% yield. This method was much more industrially feasible than previous methods used.[124]

As the yield *via* extraction from the plant is relatively low, individuals have been attempting to increase the content of artemisinin (**4**) in the plant. One method attempted involved fusing FDP synthase with ADS (the rate limiting enzyme) in an attempt to channel the product of the first enzyme directly to the next. These were cloned into vectors and transformed into two high yielding varieties of *A. annua.* The genetically engineered plants showed higher levels of the ADS transcript and higher levels of artemisinin (**4**) production: 2 and 2.5 times the wild type strains.[117] Using a similar hypothesis that accumulation of FDP (**40**) will increase overall artemisinin (**4**) production, Chen *et al.* introduced an extra copy of the gene FDP synthase. When the gene was introduced, the best producing plant had 3 times the production of artemisinin (**4**) compared to the control plants.[125]

The issue with extracting artemisinin (**4**) from plants is that the production depends on the harvest, which can be very unreliable. Prices can vary, ranging from $350 – 1700 US dollars per kilogram. As the number of companies harvesting artemisinin (**4**) has increased the prices have dropped, but the medication is still too expensive for many who need it. Alternative methods of creating artemisinin (**4**) have been investigated due to the inconsistency and expense of extracting it from plants.[88]

Total synthesis

There are various routes for total synthesis of artemisinin (**4**) starting from chiral substrates, several of which are summarised in Table 2, all of which involve 10 or more steps and have under 5% overall yield.[126–131] Although producing artemisinin (**4**) synthetically would improve the reliability of the supply compared to extraction from the natural source, none of the current published syntheses are economically viable. All have numerous steps, overall low yields and starting materials are not inexpensive. Chemical synthesis therefore offers little advantage compared to extracting the compound from its natural source.[132]

*Table 2: Summary of total synthesis routes to artemisinin (**4**)*

Author	Year	Starting material	Number of steps	Overall yield
Xu, X. X., Zhu, J., Huang, D. Z., and Zhou, W. S.[126]	1986	(*R*)-(+)-citronellal	20	0.25%
Schmid, G., and Hofheinz, W.[127]	1983	(-)-isopulegol	13	2.96%
Ravindranathan, T., Kumar, M. A., Menon, R. B., and Hiremath, S. V.[128]	1990	(+)-isolimonene	12	1.26%
Liu, H. J., Yeh, W. L., and Chew, S. Y.[129]	1993	(-)-β-pinene	18	3.47%
Yadav, J. S., Babu, R. S., and Sabitha, G.[130]	2003	(+)-isolimonene	11	0.35%
Avery, M. A., Chong, W. K. M., and Jennings-White, C.[131]	1992	(*R*)-(+)- pulegone	10	3.42%

Microbial biosynthesis

As chemical synthesis has proved inefficient, an alternative possibility is biosynthetic production. The progression of microbial biosynthesis of artemisinin (**4**) precursors is summarised in Figure 45. As parts of the enzymatic pathway for artemisinin (**4**) synthesis were discovered this became a possibility. Initial work on the production of artemisinin (**4**) using a mixture of microbes and chemistry began with large-scale production of amorpha-4,11-diene (**34**), a precursor of artemisinin (**4**), in *E. coli*. (Figure 41) This was achieved by inserting parts of the mevalonate pathway from *S. cerevisiae*, *S. aureus*, and a synthetic copy of the ADS gene into the organism. The organism was grown in a fermenter and sugar and nitrogen were monitored for optimal production. Using the optimised strain of *E. coli* and the optimum conditions for fermentation, with a two-phase partitioning bioreactor to capture the amorpha-4,11-diene (**34**) in the organic phase, a yield of 27.4 g/L was obtained.[133] Amorpha-4,11-diene (**34**) synthesis was then moved to *S. cerevisiae* CEN.PK2. In this strain, the mevalonate pathway up to FDP synthase was upregulated, a synthetic copy of ADS was introduced and the need for galactose for metabolism was removed leading to a yield of 41 g/L.[134]

Once production of amorpha-4,11-diene (**34**) was successful, the next step was to produce artemisinic acid (**92**), which is easier to convert chemically to artemisinin (**4**). Initial attempts at producing artemisinic acid (**92**) in yeast gave very low yields compared to amorpha-4,11-

diene (**34**) production. Producing artemisinic acid (**92**) involves introduction of CYP71AV1 and its partnering reductase into the synthetic pathway, but when this was done, the viability of the organisms were reduced. When the ratio of expression of the CYP and its reductase were optimised, cell viability was increased but artemisinic acid (**92**) production decreased. Therefore, a further protein from *A. annua*, a cytochrome b_5, was introduced into the yeast strain. This led to not only an increase in artemisinic acid (**92**) production, but also an increase in artemisinic aldehyde (**81**) production, which was presumed to be toxic. A dehydrogenase, ALDH1, that converts artemisinic aldehyde (**81**) to the acid (**92**) was introduced into the yeast, increasing artemisinic acid (**92**) yield to 7.7 g/L. Unfortunately, this was still much lower than the yield obtained for amorpha-4,11-diene (**34**) production. A further dehydrogenase, ADH1, was identified as part of the pathway and with its introduction the yield increased once more. The final issue was the crystallisation of artemisinic acid (**92**) in the cultures, leading to inaccurate measurements of concentrations. This was solved by overlaying the cultures with isopropyl myristate oil which solubilised the precipitate. This, along with an ethanol feed fermentation, allowed yields of 25 g/L.[135] This value is key as it is the value estimated to be required for economical production of artemisinin (**4**) derivatives.[71]

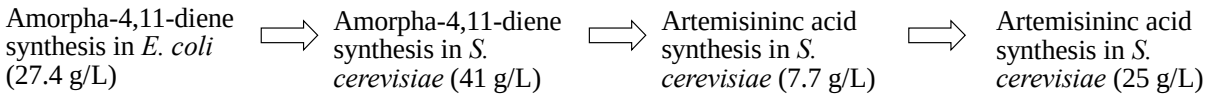


*Figure 45: Summary of microbial synthesis of artemisinin (**4**)*

Finally, artemisinic acid (**92**) must be converted into artemisinin (**4**). In *A. annua*, dihydroartemisinic acid (**83**) is converted to artemisinin (**4**) involving oxygen and UV light.[117] For the chemical synthesis, the first step is converting artemisinic acid (**92**) to dihydroartemisinic acid (**83**) in a diastereoselective manner. Hydrogenation using $RuCl_2$[(*R*)-DTBM-Segphos](DMF)$_n$, as the catalyst afforded the desired product with 95% selectivity and quantitative yield. The final step is the conversion of dihydroartemisinic acid (**83**) to artemisinin (**4**), which is a highly complex reaction involving oxidation and rearrangement. Dihydroartemisinic acid (**83**) is converted to the anhydride before undergoing a Schenck ene reaction to produce a linear peroxide. The reaction is initiated by light and addition of acid. A Hock cleavage, and oxidation cyclisation occurs to give artemisinin (**4**), which was obtained in 55% isolated yield. This synthesis is shown in Figure 46. [136]

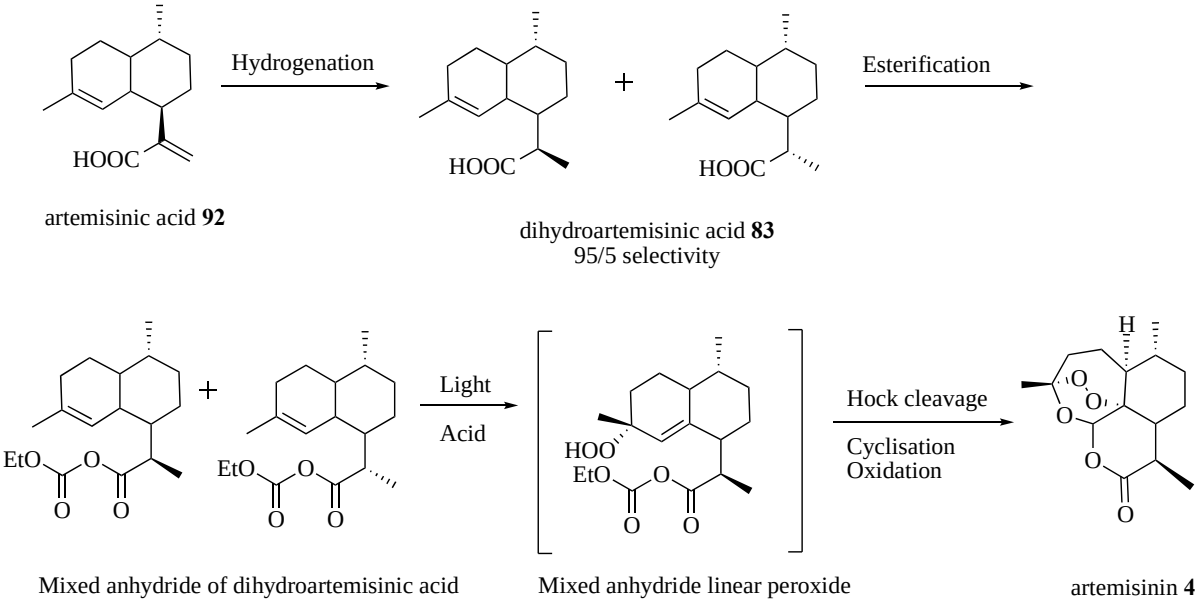


*Figure 46: Chemical synthesis of artemisinin (**4**) from artemisinic acid (**92**)* [136]

This synthesis was taken to an industrial scale by Sanofi,[136] but after only a few years manufacturing artemisinin (**4**) the company decided to stop production, as they were unable to compete with the price of artemisinin (**4**) extracted from *A. annua*.[137]

1.6. Artemisinin based combination therapy

Malaria and the *Plasmodium* life cycle

Malaria is the largest global parasitic disease that kills around three million people every year, a large proportion of which are children under five years of age. Close to half of the worlds' population live in a country where malaria is a major health concern. The disease is most common in developing countries and tropical countries, but outbreaks have occurred in Europe. It is estimated that malaria costs the world around $12 billion US dollars every year from direct issues such as the cost of treatment. The cost including the loss of economic growth due to the money spent on malaria is likely to be much higher.[138]

Malaria is caused by the single celled protozoan *Plasmodium*. There are five known species of *Plasmodium* which infect humans. Four are very commonly known, *P. falciparum, P. vivax, P. ovale* and *P. malaria. P. falciparum* is responsible for the majority of infections.[139] *P. vivax* and *P. ovale* can cause relapses once the blood is clear of parasites, as they can become a dormant form in the liver, called a hypnozoite. These can initiate the liver life cycle again,

weeks or months after the initial symptoms have cleared.[140] The final species is *P. knowlesi* which is normally found in macaque monkeys, but many cases in humans have been identified, begging the question as to whether this parasite has switched host and can now infect humans.[141]

Malaria is spread *via* the bite of the female of the *Anopheles* genus of mosquitoes. The *Plasmodium* sit in the salivary glands of the mosquitoes, and when the mosquito bites the mammalian host to draw blood, the parasite is injected into the wound. The *Plasmodium* is in the form of a sporozoite at this stage of the life cycle and can travel through the host skin surface to a blood vessel. They then continue on to the host's liver and within 48 hours are in a hepatocyte where they form a parasitophorous vacuole in which the parasite can replicate.[142] Some *Plasmodium* species will form hypnozoites in the liver, which can lead to relapses. Otherwise the parasite forms schizonts with numerous nuclear divisions occurring. This leads to the hepatocyte enlarging as the cytoplasm of the parasite increases also. Each nucleus will form a merozoite, as they are encircled by a little of the cytoplasm.[140] These merozoites will be released into the blood stream as membrane bound merosomes, which protect them until ready to infect erythrocytes, where the merosome will rupture releasing the merozoites.[143] This cycle will occur only once, for each infection.

These merozoites enter red blood cells where they form a ring structure. In the red blood cells the parasite feeds on the cytoplasm breaking down the haemoglobin. This leads to pigment formation which is visible in the parasite, which is now in the trophozoite form. The parasite will then return to the schizont form, this time in the red blood cell, and create many new daughter cells. The merozoites are released when the red blood cells rupture allowing further infection.[140] Rather than differentiating into schizonts a small number of the parasites will complete the sexual reproductive cycle forming gametocytes.[143] The female forms are called macrogametocytes and the male microgametocytes. A mosquito will bite an infected host containing the gametocytes, these will fuse, and a zygote is formed. This will mature to an ookinete one to two days after fertilization. Ookinetes will undergo mitotic division to form oocyst, which will divide thousands of times. These daughter cells are sporozoites, which travel through the mosquito and congregate in the salivary gland ready to infect the next host.[140] The life cycle is summarised in Figure 47.

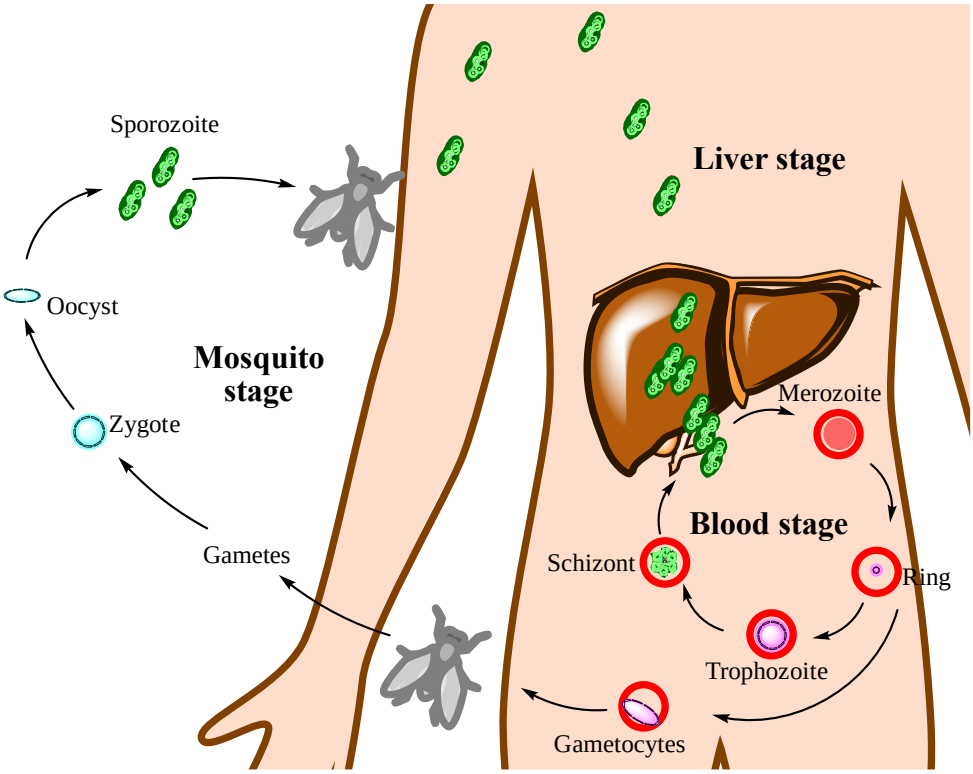


Figure 47: Cartoon diagram of the stages of the malarial life cycle [143]

Malarial symptoms and past treatments

Perhaps the most well-known symptom of malaria is the cyclical fevers caused by the red blood cells rupturing, releasing the parasite into the blood stream. The fever occurs once a certain concentration of the parasite is reached. This fever normally lasts only a few hours and occurs every 48 hours for *P. falciparum, viva* and *ovale*, and every 72 hours in the case of *P. malariae*. Some sufferers will not follow this pattern *e.g.* if they are infected with more than one type of *Plasmodium*. It has been demonstrated that a fever can reduce the number of *Plasmodium* in the host, but the parasite has learnt to use the fever to their advantage. In some clinical studies the parasites seem to use the fever as a signal allowing them to synchronise the stage of the life cycle they are in. Malaria can also cause more severe symptoms such as anaemia due to the cyclical rupturing of red blood cells and cerebral malaria.[144]

Cerebral malaria involves seizures and the host having impaired consciousness or entering a coma due to their infection. It is hypothesised that this is due to infected red blood cells in the vasculature of the brain adhering to the lining of vessels blocking blood supply. Others hypothesise it is due to inflammation caused by the parasite. The consequences of cerebral malaria are severe, and without treatment is fatal. Even if patients receive treatment and

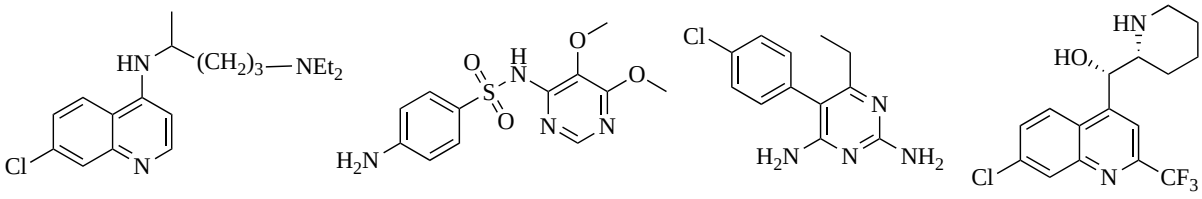
recover, they often have long-lasting effects. These can include impaired cognition and motor-function, blindness and epilepsy.[145] As in the brain, infected blood cells can move to the lining of arteries and veins in the body which can lead to blocking of blood flow and oxygen supply. This, along with the inflammation triggered by the infection, can lead to multi-organ failure.[146]

Chloroquine (**93**) (Figure 48) is likely one of the most successful drugs used for malaria. This drug was developed in 1934 but was not recognised as an anti-malarial until 1942. Unfortunately, in the 1950s, resistance began to develop and by the 1980s resistant *P. falciparum* was the norm in most of the affected countries. Chloroquine (**93**) inhibits the parasites polymerisation of hematin, a breakdown product of the heme in red blood cells, to hemozoin. Hematin is poisonous to *Plasmodium*, hence chloroquine's (**93**) activity.[147]

Due to the increase in resistance to chloroquine (**93**), new drugs were needed to treat malaria. The family of anti-folate drugs, including pyrimethamine (**95**) and sulfadoxine (**94**) (Figure 48) became the next choice during the 1990s. Pyrimethamine (**95**) is a competitive inhibitor of dihydrofolate reductase (DHFR), an enzyme vital for synthesis of proteins and DNA, which is also a target of many cancer therapies. Sulfadoxine (**94**) inhibits dihydropteroate synthase (DHPS) which is involved in folate synthesis. These drugs work very effectively together but less so when working alone. Unfortunately, the life-span of this treatment was rather short lived with resistance emerging in many regions within five years. The resistance is due to point mutations in DHFR and DHPS, with the point mutation S108N in DHFR reducing pyrimethamine's (**95**) efficiency 100-fold. It is hypothesised that resistance developed much quicker to antifolates versus chloroquine (**93**) as antifolates were already used against bacterial infections, therefore the *Plasmoidum* may have already encountered the drugs before the malarial treatment.[148]

Mefloquine (**96**) (Figure 48) was introduced in 1984 in Thailand[149] and is a member of the quinoline containing antimalarials like chloroquine (**93**). How this drug works is subject to many hypotheses, one of which is the same as chloroquine (**93**); the inhibition of hemozoin formation. An alternative is the observation that it inhibits the endocytosis of haemoglobin into the food vacuole.[150] As with previous malarial treatments, resistance to mefloquine (**96**) was seen within six years of use in some regions. One known resistance mechanism is the amplification and overexpression of the pfmdr1 gene, which produces an (ATP)-Binding Cassette transporter. It seems resistant *Plasmodium* have an increased rate of efflux of

mefloquine (**96**). To combat this resistance, mefloquine (**96**) is now being used in combination therapy with artemisinin (**4**) derivatives.[149]

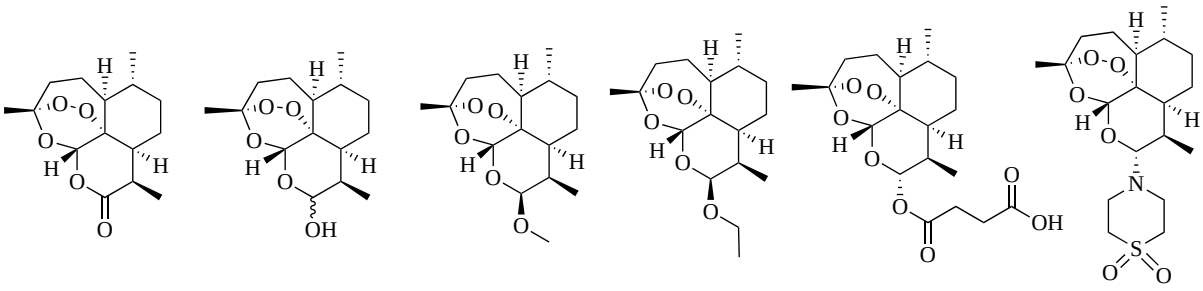


chloroquine **93** sulfadoxine **94** pyrimethamine **95** mefloquine **96**

*Figure 48: Structures of Chloroquine (**93**)[150], Sulfadoxine (**94**), Pyrimethamine (**95**)[151] and Mefloquine (**96**)[150]*

Artemisinin (**4**) and its derivatives

The current recommended treatment for malaria is artemisinin-based combination therapies (ACT). In the 1980s artemisinin (**4**) caught the world's attention as it was helping thousands of patients in China and by 2005 WHO announced that the first line treatment for malaria should be artemisinin combination therapies.[118] Although artemisinin (**4**) showed good antimalarial activity, as a viable drug, it has its downfalls. It shows poor solubility in both water and oil, leading scientists to create more soluble derivatives (Figure 49).[152] The most common derivatisation occurs on the C10 position of artemisinin (**4**). The lactone group was first converted to a lactol group to obtain dihydroartemisinin (**77**) which was much more active than artemisinin (**4**). This lactol group was further derivatised with ethers and esters to produce oil soluble compounds, such as artemether (**97**) and arteether (**98**). An alternative water-soluble compound called artesunate (**76**) was produced by reacting the ether with succinic acid. Artesunate (**76**) is commonly used for severe malarial infections and is often used in combination therapy with other antimalarials for which resistance is known such as sulfadoxine (**94**)-pyrimethamine (**95**), and mefloquine (**96**).[153]



artemisinin **4** dihydroartemisinin **77** artemether **97** arteether **98** artesunate **76** artemisone **99**

*Figure 49: The structure of artemisinin (**4**) and many of its derivatives* [153]

Throughout the history of malarial treatment, resistance to medication has emerged time and time again. When artemisinin (**4**) was established as a treatment, it was a golden opportunity to try and stop resistance developing to this new class of anti-malarials.[154] Therefore the recommended treatment for malaria from WHO is now artemisinin combination therapy, where an artemisinin (**4**) derivative is administered along with another longer lasting antimalarial.[88] This combination therapy is also useful as many artemisinin (**4**) derivatives such as artemether (**97**) and artesunate (**76**) are eliminated very rapidly from the body, both with a half-life of approximately one hour. When artemisinin (**4**) derivatives are administered alone a much longer course is needed, therefore compliance is much lower.[88]

Artemisinin (**4**) derivatives generally clear the majority of the parasites, but the partner drug administered in the ACT mops up any remaining parasites. This means resistant parasites are much less likely to develop, as, if one treatment does not kill them the other will. Artesunate (**76**)-mefloquine (**96**) was introduced in 1994 in Thailand and after 14 years the treatment was still effective. Artesunate (**76**) has also been administered in partnership with sulfadoxine (**94**)-pyrimethamine (**95**).[88] It is advantageous to have multiple ACT as they may show differing clearance of the parasite depending on the species, or on the patients' environment.[154] There are concerns regarding artemisinin's safety in the first trimester of pregnancy. It has yet to be established whether artemisinin (**4**) is teratogenic. It is clear the effort to discover better anti-malarials must continue.[88]

Artemisinin (**4**) - mechanism of action

The endoperoxide bridge found in artemisinin (**4**) is key to its activity, although its presence does not guarantee antimalarial activity in derivatives, and the cyclic ether aids in activity.[152] It is generally agreed that the iron in haem is needed to activate artemisinin (**4**) to produce free radicals, and as the majority of mammalian haem is incorporated into haemoglobin, it does not interact with the iron. *Plasmodium* on the other hand, break down haemoglobin for nutrients and do interact closely with haem, making them vulnerable to artesmisinin's effects.[155] What happens next, and how exactly artemisinin (**4**) works is not clear and there are many hypotheses with evidence for and against these ideas.

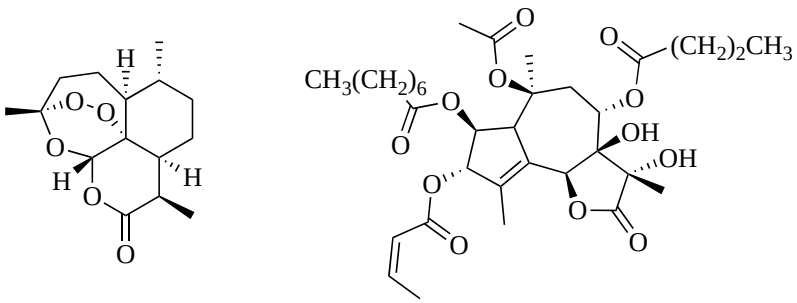
The first of these is that artemisinin (**4**) affects the haem detoxification pathway. *Plasmodium* break down the haemoglobin in the host's red blood cells releasing haem, which is toxic as it can produce hydroxyl free radicals. These parasites get around this issue by detoxifying haem into hemozoin *via* a polymerisation reaction. The hypothesis is that artemisinin (**4**) and its derivative alkylate haem inhibiting this polymerisation, and this has been shown *in vitro* and *in vivo* in mice models. Therefore more free haem is available to produce free radicals, causing oxidative damage.[97] In contrast to this hypothesis, other research has demonstrated that compounds with antimalarial activity and containing an endoperoxide bridge do not interact with haem or free iron. This would suggest that an alternative mechanism for anti-malarial activity is occurring, but possibly alongside the haem alkylation.[156]

The second hypothesis is that artemisinin (**4**) damages various proteins such as translationally controlled tumour protein. Studies used radiolabelled artemisinin (**4**) derivatives to treat *Plasmodium* infected and uninfected red blood cells. When comparing infected an uninfected red blood cells, only the infected red blood cells showed radioactively labelled proteins, and this only occurred with the antimalarial artemisinin (**4**) derivatives, not the inactive derivative used as a control.[157] One of these radiolabelled proteins was later discovered to be the *Plasmodium* translationally controlled tumour protein (TCTP) homolog, and it was found that TCTP binds haem, possibly linking this hypothesis with the first.[158] A study looking at a mouse infecting *Plasmodium* showed that artemisinin (**4**) resistant parasites express much more TCTP compared to the sensitive parasites, indicating this protein does play a part in the mechanism of antimalarials, but it may be a knock on effect from the original target. Without the understanding of the role of TCTP in *Plasmodium* it is difficult to be certain that it is a target. Perhaps in future a knock-out strain of *Plasmodium* lacking TCTP could be generated to see if this is a true target or a downstream effect.[159]

The third hypothesis is that artemisinin (**4**) inhibits the sacro/endoplasmic reticulum calcium ATPase (SERCA), PfATP6. Artemisinins have a similar central structure to a known SERCA inhibitor, thapsigargin (**100**), which is also a sesquiterpene lactone (Figure 50) leading to the hypothesis that artemisinins may work in a similar way. *P. falciparum* have two calcium ATPase transporters, only one of which is a SERCA. When PfATP6 was expressed in frog eggs, inhibition of this transporter was induced by thapsigargin (**100**) and artemisinin (**4**) but not by other antimalarials such as quinine. It was also confirmed that artemisinins only inhibit this transporter and not other membrane proteins such as the other calcium transporter PfATP4,

even at much higher concentrations than necessary to inhibit PfATP6. When an artemisinin (**4**) derivative that lacks the endoperoxide bridge (and the anti-malarial activity) was tested no inhibition of PfATP6 was seen.[160] A point mutation found in PfATP6, S769N, in samples from populations in French Guiana seemed to confer resistance to ACT as six of seven samples that had this mutation had raised IC_{50} values for artemether (**97**). How this mutation allows resistance has not been proven, but it is hypothesised that it stops artemisinin (**4**) affecting conformational changes required of the protein during the ATPase cycle and for calcium interaction.[161]

Another residue in PfATP6 suggested to be related to artemisinin (**4**) activity is L263. Artemisinin (**4**) does not inhibit mammalian SERCA, and in the mammalian equivalent of PfATP6 L263 is a glutamic acid residue. When a mutant version of PfATP6, L263E, was expressed in frog eggs, artemisinin's (**4**) inhibitory activity was lost.[162] When the mutant PfATP6, L263E, was expressed in *P. falciparum*, very little change was seen in artemisinin's (**4**) activity, showing more *in vivo* studies of this protein are vital to validate the *in vitro* results.[163] When enantiomers of analogues of artemisinin (**4**) are tested for activity against *P. falciparum*, no clear difference is seen between each enantiomer's activity, leading to the argument that PfATP6 cannot be the direct target of artemisinin (**4**) or derivatives as it would surely involve some stereoselectivity.[164] A study looking at changes in organelle structure of *Plasmodium* when exposed to various drugs showed that thapsigargin (**100**) causes a change in the morphology of the endoplasmic reticulum, its target. However, artemisinin (**4**) and other bicyclic endoperoxides did not cause the same effect.[165]



artemisinin **4** thapsigargin **100**

*Figure 50: Structure of artemisinin (**4**) and the SERCA inhibitor thapsigargin(**100**)* [164]

The final hypothesis is that artemisinin (**4**) interferes with the plasmodium mitochondria. A study using *S. cerevisiae* as a model showed that artemisinin (**4**) does not affect yeast at medically relevant conditions when it is allowed to produce energy *via* fermentation. When the

media is switched, forcing the yeast to use the electron transport chain to produce energy, artemisinin (**4**) has a marked effect on growth suggesting the mitochondria is affected. When the membrane potential of the mitochondria was investigated using a cationic dye, treatment with artemisinin (**4**) depolarised the inner mitochondrial membrane. Numerous genes were individually knocked out to see if there was a change in reactivity to artemisinins. Two genes, *NDE1* and *NDI1*, both NADH dehydrogenases, were found to be involved in the effect of artemisinin (**4**), as the knock out strains became resistant to artemisinin (**4**). In agreement to this, when the genes were overexpressed, the strains showed increased sensitivity to artemisinin (**4**). To show relevance to *P. falciparum* the homologue to *NDI1*, the only NADH dehydrogenase in the parasite, PfNDI1 was expressed in the knockout of *NDI1* restoring partial sensitivity to artemisinin (**4**). There is no known homologue of this gene in humans. This suggests that components of the *Plasmodium* mitochondria are targets of artemisinin (**4**), whether it is a direct or secondary effect.[166]

When this research was continued by treating a mouse infected with a strain of *Plasmodium* with artemisinin (**4**), membrane depolarisation of the parasites mitochondrial membrane was seen. Further experiments implied this was not an artefact of cell death as was mostly reversible *via* the removal of artemisinin (**4**). When deoxyartemisinin, a derivative of artemisinin (**4**) lacking anti-malarial ability, was used as a control these effects were not seen, and neither were they seen when using mitochondria obtained from CHO cells. When artemisinin (**4**) was used concomitantly with DPI, an inhibitor of a component of the ETC, there was a clear antagonistic effect, whereas with chloroquine (**93**), which is not known to affect the mitochondria, there was an additive affect. Treatment of yeast and *Plasmodium* mitochondria with artemisinin (**4**) causes an increase in reactive oxygen species, but this is not seen in mammalian mitochondria. This suggests that artemisinin's (**4**) effect stems from an increase in production of reactive oxygen species and the damages these cause down the line.[167]

A study conducted with novel bicyclic endoperoxides using artemisinin (**4**) as a control had contrasting results. When very high doses of artemisinin (**4**) and bicyclic endoperoxides were added to *Plasmodium* at two varying points of the life cycle and left for four hours, no change in membrane potential was detected. By twenty-four hours, only some change to membrane potential was seen, but this may be an artefact of cell death rather than a direct result of the drug. When using a dye that labels the digestive vacuole of the *Plasmodium*, experiments using physiologically relevant concentrations of artemisinin (**4**) caused changes in the dye

distribution within a few hours, but no affects were seen on the mitochondria or endoplasmic reticulum. After four hours, with much higher concentrations of artemisinin (**4**), changes to the digestive vacuole structure were seen, but only after longer incubations were changes to other organelles seen. These results suggest that effects on the mitochondria may be secondary, and are not the first target of artemisinin (**4**).[165]

There is clearly much evidence for and against each hypothesis, that may suggest that there are multiple mechanisms at play depending on the derivative, the strain of parasite and the health of the host. Of course experimental design of all the research discussed will have an effect on the result *e.g. in vitro* vs *in vivo*, the stage of life cycle the parasite is in when tested; and it may be these are all off target effects and we have yet to discover the mode of action of artemisinin (**4**).

Artemisinin (**4**) resistance

Resistance is a continual problem with anti-malarials and although ACT was used as an attempt to curb resistance, and succeeded for much longer than previous generations of antimalarials, some resistance has emerged in the Thailand-Cambodia and Thailand-Myanmar regions.[154] In a study conducted between 2006 and 2007 in Cambodia, two patients of the 94 enrolled in the study had resistant *Plasmodium* infections, with IC_{50}s up to four times that of a 'non-resistant' infection. These resistant strains did not seem to correlate with any known point mutations in the *PfATPase6* gene.[168] Another marker of resistance seen in areas of Cambodia is a longer parasite clearance time, which does not show itself with the standard *in vitro* testing as it is believed to be due to resistance at the young ring stage parasites.[169]

Although many targets have been discussed in the hypothesised mechanisms for artemisinin (**4**), a gene determined to be involved in artemisinin (**4**) resistance was not discovered. The PfKelch13 protein is a propeller protein found on chromosome 13 of the *Plasmodium*. When resistant parasites isolated from Cambodia were analysed, four mutations in this gene were associated with longer parasite survival. These included R539T, Y493H, I543T and C580Y. Interestingly none of these mutations showed an increase in IC_{50} values, showing that the experiments conducted may bias the results obtained.[170]

How this protein is involved in resistance to artemisinins (**4**) cytotoxicity is not clear. One proposal is that it interacts with phosphatidylinositol-3-kinase (PI3K) which produces phosphatidylinositol-3-phosphate (PI3P). PI3P is involved in transporting proteins from the parasite's ER to the red blood cells of the host. Nanomolar concentrations of dihydroartemisinin (**77**) inhibited PI3P production, whereas deoxyartemisinin did not, and dihydroartemisinin (**77**) does not inhibit the human homologue of PfPI3K. Supporting this proposal, some resistant strains found in Cambodia expressed 1.5-2 times the levels of PfPI3PK compared to sensitive strains. A mutation of PfKelch13, C580Y, was introduced into a wild type strain of *Plasmodium* and expression of PI3K increased 2-3-fold with no change in PfKelch13 expression levels. The mutant PfKelch13 did not bind PI3K as well as the wild type providing a possible explanation for this.[171] Many mutations in the PfKelch13 gene have been found in numerous samples in the Myanmar region also. Some of these mutations have been associated with resistance to artemisinins (**4**) cytotoxicity, and slower parasite clearance.[172]

As previously mentioned, strains of *Plasmodium* in French Guiana were also found to be resistant to ACT.[161] ACTs ability to ensure resistance will not develop is dependent on the parasites' susceptibility to both drugs used, if the parasite previously developed resistance to the partner drug and then develops resistance to artemisinins (**4**) cytotoxicity then the possibility of resistance to ACT occurs.[173] This may be how resistance has developed in Southeast Asia.

1.7. Aims

This project had two aims, which were designed to try and further our understanding of the flexibility and selectivity of ADS for its application in biocatalytic production of artemisinin. The first part involved the investigation of the residues involved in the flexibility and efficiency of ADS by site directed mutagenesis. This work was an attempt to probe which residues are involved in shaping the achiral substrate FDP into the bicyclic product amorpha-4,11 diene, and which residues are key for the efficiency of this cyclisation.

Secondly this work set out to improve the synthesis of 12-OH FDP, which can be converted using ADS to DHAAl. This exploits the flexibility of the ADS active site for accepting analogues of FDP. The previous method used for the synthesis of 12-OH FDP is inefficient, time consuming and laborious.[174] Therefore, various kinases and a P450 were investigated to

try an create a mostly or fully enzymatic pathway for the synthesis of 12-OH FDP. Site directed mutagenesis was utilised throughout to try and improve various enzymes affinity for unnatural substrates. An improved route to 12-OH FDP would generate a more efficient and less labour-intensive synthesis of DHAAl which can be converted to the key antimalarial artemisinin.

1 J. Buckingham, *Dictionary of Natural Products, Vol 7*, CRC Press, London, 1st edn., 1994.
2 J. Degenhardt, T. G. Köllner and J. Gershenzon, *Phytochemistry*, 2009, **70**, 1621–1637.
3 D. W. Christianson, *Chem. Rev.*, 2017, **117**, 11570–11648.
4 D. W. Christianson, *Chem. Rev.*, 2006, **106**, 3412–3442.
5 J. Isac-Garcia, J. A. Dobado, F. G. Calvo-Flores and H. Martinez-Garcia, in *Experimental Organic Chemistry: Laboratory Manual*, Academic Press, London, 1st edn., 2016, pp. 207–238.
6 C. N. Tetzlaff, Z. You, D. E. Cane, S. Takamatsu, S. Omura and H. Ikeda, *Biochemistry*, 2006, **45**, 6179–6186.
7 A. L. Wheeler, R. M. Long, R. E. B. Ketchum, C. D. Rithner, R. M. Williams and R. Croteau, *Arch. Biochem. Biophys.*, 2001, **390**, 265–278.
8 J. ichiro Hattan, K. Shindo, T. Sasaki, F. Ohno, H. Tokuda, K. Ishikawa and N. Misawa, *Sci. Rep.*, 2018, **8**, 1–12.
9 P. M. Bleeker, R. Mirabella, P. J. Diergaarde, A. VanDoorn, A. Tissier, M. R. Kant, M. Prins, M. de Vos, M. A. Haring and R. C. Schuurink, *Proc. Natl. Acad. Sci.*, 2012, **109**, 20124–20129.
10 S. Sarria, B. Wong, H. G. Martín, J. D. Keasling and P. Peralta-Yahya, *ACS Synth. Biol.*, 2014, **3**, 466–475.
11 A. Boronat and M. Rodriguez-Concepcion, in *Biotechnology of Isoprenoids*, eds. J. Schrader and J. Bohlmann, Springer International Publishing AG Switzerland, 1st edn., 2014, vol. 148, pp. 3–18.
12 L. Ruzicka, *Experientia*, 1994, **50**, 395–405.
13 T. Endo and T. Suga, *Phytochemistry*, 1992, **31**, 1565–1568.
14 T. Kuzuyama, *Biosci. Biotechnol. Biochem.*, 2002, **66**, 1619–1627.
15 N. Dellas, S. T. Thomas, G. Manning and J. P. Noel, *Elife*, , DOI:10.7554/elife.00672.
16 Y. Azami, A. Hattori, H. Nishimura, H. Kawaide, T. Yoshimura and H. Hemmi, *J. Biol. Chem.*, 2014, **289**, 15957–15967.
17 W. Eisenreich, A. Bacher, D. Arigoni and F. Rohdich, *Cell. Mol. Life Sci.*, 2004, **61**, 1401–1426.
18 A. Hemmerlin, J. L. Harwood and T. J. Bach, *Prog. Lipid Res.*, 2012, **51**, 95–148.
19 H. V Thulasiram, H. K. Erickson and C. Dale Poulter, *Science (80-.).*, 2007, **316**, 73–76.
20 H. V. Thulasiram and C. D. Poulter, *J. Am. Chem. Soc.*, 2006, **128**, 15819–15823.
21 J. Gao, T. P. Ko, L. Chen, S. R. Malwal, J. Zhang, X. Hu, F. Qu, W. Liu, J. W. Huang, Y. S. Cheng, C. C. Chen, Y. Yang, Y. Zhang, E. Oldfield and R. T. Guo, *Angew. Chemie - Int. Ed.*, , DOI:10.1002/anie.201710185.
22 T.-H. Chang, F.-L. Hsieh, T.-P. Ko, K.-H. Teng, P.-H. Liang and A. H.-J. Wang, *Plant Cell*, 2010, **22**, 454–467.
23 B. Singh and R. A. Sharma, *3 Biotech*, 2015, **5**, 129–151.
24 T. Koyama, *Biosci. Biotechnol. Biochem.*, 1999, **63**, 1671–1676.
25 S. C. Kampranis, D. Ioannidis, A. Purvis, W. Mahrez, E. Ninga, N. A. Katerelos, S. Anssour, J. M. Dunwell, J. Degenhardt, A. M. Makris, P. W. Goodenough and C. B. Johnsona, *Plant Cell*, 2007, **19**, 1994–2005.
26 S. H. Kim, K. Heo, Y. J. Chang, S. H. Park, S. K. Rhee and S. U. Kim, *J. Nat. Prod.*, 2006, **69**, 758–762.
27 K. Zhou, Y. Gao, J. A. Hoy, F. M. Mann, R. B. Honzatko and R. J. Peters, *J. Biol. Chem.*, 2012, **287**, 6840–6850.
28 Z. G. Brill, H. K. Grover and T. J. Maimone, *Science (80-.).*, 2016, **352**, 1078–1082.

29 K. U. Wendt, K. Poralla and G. E. Schulz, *Science (80-.).*, 1997, **277**, 1811–1815.
30 B. J. Burri, M. R. La Frano and C. Zhu, *Nutr. Rev.*, 2016, **74**, 69–82.
31 Y. Kharel and T. Koyama, *Nat. Prod. Rep.*, 2003, 20, 111–118.
32 P. H. Liang, T. P. Ko and A. H. J. Wang, *Eur. J. Biochem.*, 2002, **269**, 3339–3354.
33 F. Beran, P. Rahfeld, K. Luck, R. Nagel, H. Vogel, N. Wielsch, S. Irmisch, S. Ramasamy, J. Gershenzon, D. G. Heckel and T. G. Köllner, *Proc. Natl. Acad. Sci.*, 2016, **113**, 2922–2927.
34 C. Sallaud, D. Rontein, S. Onillon, F. Jabès, P. Duffé, C. Giacalone, S. Thoraval, C. Escoffier, G. Herbette, N. Leonhardt, M. Causse and A. Tissier, *Plant Cell*, 2009, **21**, 301–317.
35 K. U. Wendt and G. E. Schulz, *Structure*, 1998, **6**, 127–133.
36 Y. Gao, R. B. Honzatko and R. J. Peters, *Nat. Prod. Rep.*, 2012, 29, 1153–1175.
37 M. Baunach, J. Franke and C. Hertweck, *Angew. Chemie - Int. Ed.*, 2015.
38 E. Y. Shishova, F. Yu, D. J. Miller, J. A. Faraldos, Y. Zhao, R. M. Coates, R. K. Allemann, D. E. Cane and D. W. Christianson, *J. Biol. Chem.*, 2008, **283**, 15431–15439.
39 D. J. Miller and R. K. Allemann, *Nat. Prod. Rep.*, 2012, **29**, 60–71.
40 M. Loizzi, V. González, D. J. Miller and R. K. Allemann, *ChemBioChem*, 2018, **19**, 100–105.
41 I. Prosser, I. G. Altug, A. L. Phillips, W. A. König, H. J. Bouwmeester and M. H. Beale, *Arch. Biochem. Biophys.*, 2004, **432**, 136–144.
42 C. M. Starks, K. Back, J. Chappell and J. P. Noel, *Science (80-.).*, 1997, **277**, 1815–1820.
43 M. Köksal, Y. Jin, R. M. Coates, R. Croteau and D. W. Christianson, *Nature*, 2011, **469**, 116–122.
44 L. C. Tarshis, J. C. Sacchettini, M. Yan and C. D. Poulter, *Biochemistry*, 1994, **33**, 10871–10877.
45 M. Chen, G. G. Harris, T. A. Pemberton and D. W. Christianson, *Curr. Opin. Struct. Biol.*, 2016, **41**, 27–37.
46 F. Chen, D. Tholl, J. Bohlmann and E. Pichersky, *Plant J.*, 2011, **66**, 212–229.
47 R. P. McAndrew, P. P. Peralta-Yahya, A. Degiovanni, J. H. Pereira, M. Z. Hadi, J. D. Keasling and P. D. Adams, *Structure*, 2011, **19**, 1876–1884.
48 H. Kawaide, T. Sassa and Y. Kamiya, *J. Biol. Chem.*, 2000, **275**, 2276–2280.
49 J. Bohlmann, G. Meyer-Gauen and R. Croteau, *Proc. Natl. Acad. Sci. U. S. A.*, 1998, **95**, 4126–4133.
50 S. C. Trapp and R. B. Croteau, *Genomic Organization of Plant Terpene Synthases and Molecular Evolutionary Implications*, 2001.
51 W. A. Konig, D. Joulain and D. H. Hochmuth, Terpenoids Library, https://massfinder.com/wiki/Terpenoids_Library, (accessed 20 May 2020).
52 C. A. Lesburg, G. Zhai, D. E. Cane and D. W. Christianson, *Science (80-.).*, 1997, **277**, 1820–1824.
53 P. Baer, P. Rabe, K. Fischer, C. A. Citron, T. A. Klapschinski, M. Groll and J. S. Dickschat, *Angew. Chemie - Int. Ed.*, 2014, **53**, 7652–7656.
54 P. N. Blank, G. H. Barrow, W. K. W. Chou, L. Duan, D. E. Cane and D. W. Christianson, *Biochemistry*, 2017, **56**, 5798–5811.
55 J.-X. Li, X. Fang, Q. Zhao, J.-X. Ruan, C.-Q. Yang, L.-J. Wang, D. J. Miller, J. A. Faraldos, R. K. Allemann, X.-Y. Chen and P. Zhang, *Biochem. J.*, 2013, **451**, 417–426.
56 D. E. Cane, H.-T. Chiu, P.-H. Liang and K. S. Anderson, *Biochem.*, 1997, **36**, 8332-8339.
57 V. Harms, A. Kirschning and J. S. Dickschat, *Nat. Prod. Rep.*, 2020, **37**, 1080-1097.

58 J. A. Faraldos, B. Kariuki and R. K. Allemann, *J. Org. Chem.*, 2010, **75**, 1119–1125.
59 D. J. Grundy, M. Chen, V. González, S. Leoni, D. J. Miller, D. W. Christianson and R. K. Allemann, *Biochemistry*, 2016, **55**, 2112–2121.
60 S. Picaud, P. Mercke, X. He, O. Sterner, M. Brodelius, D. E. Cane and P. E. Brodelius, *Arch. Biochem. Biophys.*, 2006, **448**, 150-155.
61 L. Zu, M. Xu, M. W. Lodewyk, D. E. Cane, R. J. Peters and D. J. Tantillo, *J. Am. Chem. Soc.*, 2012, **134**, 11369–11371.
62 M. W. Lodewyk, D. Willenbring and D. J. Tantillo, *Org. Biomol. Chem.*, 2014, **12**, 887–894.
63 I. I. Abdallah, M. Czepnik, R. Van Merkerk and W. J. Quax, *J. Nat. Prod.*, 2016, **79**, 2455–2463.
64 E. L. Ghisalberti, 1994, **37**, 597–623.
65 L. H. Zalkow and M. G. Clower, *Tetrahedron Lett.*, 1975, **16**, 75–78.
66 C. R. Benedict, J.-L. Lu, D. W. Pettigrew, J. Liu, R. D. Stipanovic and H. J. Williams, *Plant Physiol.*, 2001, **125**, 1754–1765.
67 D. M. Martin, J. Fäldt and J. Bohlmann, *Plant Physiol.*, 2004, **135**, 1908–1927.
68 A. McDonald and International Union of Biochemistry and Molecular Biology, Longifolene synthase EC 4.2.3.58, https://www.enzyme-database.org/reaction/terp/longi.html, (accessed 13 May 2020).
69 J. Rinkel and J. S. Dickschat, *Beilstein J. Org. Chem.*, 2019, **15**, 1008–1019.
70 L. Sharon-Asa, M. Shalit, A. Frydman, E. Bar, D. Holland, E. Or, U. Lavi, E. Lewinsohn and Y. Eyal, *Plant J.*, 2003, **36**, 664–674.
71 C. J. Paddon and J. D. Keasling, *Nat. Rev. Microbiol.*, 2014, **12**, 355–367.
72 J. Bohlmann, J. Crock, R. Jetter and R. Croteau, *Proc. Natl. Acad. Sci.*, 2002, **95**, 6756–6761.
73 T. J. A. Bruce, M. A. Birkett, J. Blande, A. M. Hooper, J. L. Martin, B. Khambay, I. Prosser, L. E. Smart and L. J. Wadhams, *Pest Manag. Sci.*, 2005, **61**, 1115–1121.
74 R. Mozuraitis, M. Stranden, M. I. Ramirez, A. K. Borg-Karlson and H. Mustaparta, *Chem. Senses*, 2002, **27**, 505–509.
75 K. Cankar, A. Van Houwelingen, M. Goedbloed, R. Renirie, R. M. De Jong, H. Bouwmeester, D. Bosch, T. Sonke and J. Beekwilder, *FEBS Lett.*, 2014, **588**, 1001–1007.
76 T. Wriessnegger, P. Augustin, M. Engleder, E. Leitner, M. Müller, I. Kaluzna, M. Schürmann, D. Mink, G. Zellnig, H. Schwab and H. Pichler, *Metab. Eng.*, 2014, **24**, 18–29.
77 B. C. Clark, T. S. Chamblee and G. A. Iacobucci, *J. Agric. Food Chem*, 1987, **35**, 514–518.
78 J. I. Hattan, K. Shindo, T. Sasaki and N. Misawa, *J. Oleo Sci.*, 2018, **67**, 1235–1246.
79 F. M. Santos, J. E. B. P. Pinto, S. K. V. Bertolucci, A. A. Alvarenga, M. N. Alves, M. C. T. Duarte and A. Sartoratto, *Rev. Bras. Plantas Med.*, 2013, **15**, 583–588.
80 M. Forrer, E. M. Kulik, A. Filippi and T. Waltimo, *Arch. Oral Biol.*, 2013, **58**, 10–16.
81 T. Seki, T. Kokuryo, Y. Yokoyama, H. Suzuki, K. Itatsu, A. Nakagawa, T. Mizutani, T. Miyake, M. Uno, K. Yamauchi and M. Nagino, *Cancer Sci.*, 2011, **102**, 2199–2205.
82 M. Uno, T. Kokuryo, Y. Yokoyama, T. Senga and M. Nagino, *Anticancer Res.*, 2016, **36**, 583–9.
83 A. d M. H. Alves, J. C. R. Gonçalves, J. S. Cruz and D. A. M. Araújo, *Neurosci. Lett.*, 2010, **472**, 11–15.
84 L. L. Fontinele, L. Heimfarth, E. W. M. Pereira, M. M. Rezende, N. T. Lima, Y. M. Barbosa Gomes de Carvalho, E. Afonso de Moura Pires, A. G. Guimarães, M. T. Bezerra Carvalho, R. de Souza Siqueira Barreto, A. R. Campos, A. R. Antoniolli, A.

Antunes de Souza Araújo, L. J. Quintans-Júnior and J. de Souza Siqueira Quintans, *Neurochem. Int.*, 2019, **131**, 104530.
85 B. Zhao, X. Lin, L. Lei, D. C. Lamb, S. L. Kelly, M. R. Waterman and D. E. Cane, *J. Biol. Chem.*, 2008, **283**, 8183–8189.
86 H. Gūrtler, R. Pedersen, U. Anthoni, C. Christophersen, P. H. Nielsen, E. M. H. Wellington, C. Pedersen and K. Bock, *J. Antibiot. (Tokyo).*, 1994, **47**, 434–439.
87 P. Mercke, M. Bengtsson, H. J. Bouwmeester, M. A. Posthumus and P. E. Brodelius, *Arch. Biochem. Biophys.*, 2000, **381**, 173-180.
88 N. J. White, *Science (80-.).*, 2008, **320**, 330–334.
89 J. Hou, D. Wang, R. Zhang and H. Wang, *Clin. Cancer Res.*, 2008, **14**, 5519–5530.
90 C. M. Cabello, S. D. Lamore, W. B. Bair, S. Qiao, S. Azimian, J. L. Lesson and G. T. Wondrak, *Invest. New Drugs*, 2012, **30**, 1289–1301.
91 T. Chen, M. Li, R. Zhang and H. Wang, *J. Cell. Mol. Med.*, 2009, **13**, 1358–1370.
92 Z. Jiang, J. Chai, H. H. F. Chuang, S. Li, T. Wang, Y. Cheng, W. Chen and D. Zhou, *Anticancer. Drugs*, 2012, **23**, 606–613.
93 C. Weitzel and H. T. Simonsen, *Phytochem. Rev.*, 2015, 14, 7–24.
94 D. J. Cook, J. D. Finnigan, K. Cook, G. W. Black and S. J. Charnock, in *Advances in Protein Chemistry and Structural Biology*, Academic Press Inc., 2016, vol. 105, pp. 105–126.
95 I. G. Denisov, T. M. Makris, S. G. Sligar and I. Schlichting, *Chem. Rev.*, 2005, 105, 2253–2277.
96 D. Garfinkel, *Arch. Biochem. Biophys.*, 1958, **77**, 493–509.
97 B. Meunier and A. Robert, *Acc. Chem. Res.*, 2010, **43**, 1444–1451.
98 C. R. Jefcoate, *Methods Enzymol.*, 1978, **52**, 258–279.
99 Y. M. Zheng, B. R. Baer, M. B. Kneller, K. R. Henne, K. L. Kunze and A. E. Rettie, *Biochemistry*, 2003, **42**, 4601–4606.
100 R. Fasan, *Nat. Chem.*, 2017, **9**, 609–611.
101 F. P. Guengerich, *ACS Catal.*, 2018, 8, 10964–10976.
102 D. W. Nebert, M. Adesnik, M. J. Coon, R. W. Estabrook, F. J. Gonzalez, F. P. Guengerich, I. C. Gunsalus, E. F. Johnson, B. Kemper, W. Levin, I. R. Phillips, R. Sato and M. R. Waterman, *DNA*, 1987, **6**, 1–11.
103 M. J. Coon, X. Ding, S. J. Pernecky and A. D. N. Vaz, *FASEB J.*, 1992, **6**, 669–673.
104 D. Werck-Reichhart and R. Feyereisen, *Genome Biol.*
105 O. Pylypenko and I. Schlichting, *Annu. Rev. Biochem.*, 2004, **73**, 991–1018.
106 H. Ouellet, J. B. Johnston and P. R. Ortiz de Montellano, *Arch. Biochem. Biophys.*, 2010, **493**, 82–95.
107 J. B. Johnston, P. M. Kells, L. M. Podust and P. R. Ortiz de Montellano, *Proc. Natl. Acad. Sci.*, 2009, **106**, 20687–20692.
108 G. A. Roberts, G. Grogan, A. Greter, S. L. Flitsch and N. J. Turner, *J. Bacteriol.*, 2002, **184**, 3898–3908.
109 J. Rittle, J. M. Younker and M. T. Green, *Inorg. Chem.*, 2010, **49**, 3610–3617.
110 C. M. Krest, E. L. Onderko, T. H. Yosca, J. C. Calixto, R. F. Karp, J. Livada, J. Rittle and M. T. Green, *J. Biol. Chem.*, 2013, **288**, 17074–17081.
111 F. P. Guengerich and A. W. Munro, *J. Biol. Chem.*, 2013, **288**, 17065–17073.
112 A. G. Warrilow, A. T. Nishimoto, J. E. Parker, C. L. Price, S. A. Flowers, D. E. Kelly, P. D. Rogers and S. L. Kelly, 2019, **63**, e02586-18.
113 B. Zhao, L. Lei, D. G. Vassylyev, X. Lin, D. E. Cane, S. L. Kelly, H. Yuan, D. C. Lamb and M. R. Waterman, *J. Biol. Chem.*, 2009, **284**, 36711–36719.
114 C. H. Yun, T. Ahn, F. P. Guengerich, H. Yamazaki and T. Shimada, *Arch. Biochem. Biophys.*, 1999, **367**, 81–88.

115 J. A. McIntosh, P. S. Coelho, C. C. Farwell, Z. J. Wang, J. C. Lewis, T. R. Brown and F. H. Arnold, *Angew. Chemie - Int. Ed.*, 2013, **52**, 9309–9312.
116 R. Singh, J. N. Kolev, P. A. Sutera and R. Fasan, *ACS Catal.*, 2015, **5**, 1685–1691.
117 J. Han, H. Wang, S. Kanagarajan, M. Hao, A. Lundgren and P. E. E. Brodelius, *Mol. Plant*, 2016, **9**, 946–948.
118 F. Liao, *Molecules*, 2009, **14**, 5362-5366.
119 F. Bohlmann, T. Gerke, J. Jakupovic, R. M. King and H. Robinson, *Phytochemistry*, 1984, **23**, 1183–1184.
120 H. J. Bouwmeester, T. E. Wallaart, M. H. A. Janssen, B. Van Loo, B. J. M. Jansen, M. A. Posthumus, C. O. Schmidt, J. W. De Kraker, W. A. König and M. C. R. Franssen, *Phytochemistry*, 1999, **52**, 843–854.
121 A. Akhila, K. Rani and R. S. Thakur, *Phytochemistry*, 1990, **29**, 2129-2132.
122 T. E. Wallaart, H. J. Bouwmeester, J. Hille, L. Poppinga and N. C. A. Maijers, *Planta*, 2001, **212**, 460–465.
123 D. L. Klayman, A. J. Lin, N. Acton, J. P. Scovill, J. M. Hoch, W. K. Milhous, A. D. Theoharides and A. S. Dobek, *J. Nat. Prod.*, 1984, **47**, 715–717.
124 H. N. Elsohly, E. M. Groom, F. S. El-Feraly and M. M. El-Sherei, *J. Nat. Prod.*, 1990, **53**, 19.
125 D. H. Chen, H. C. Ye and G. F. Li, *Plant Sci.*, 2000, **155**, 179–185.
126 X.-X. Xu, J. Zhu, D.-Z. Huang and W.-S. Zhou, *Tetrahedron*, 1986, **42**, 819–828.
127 G. Schmid and W. Hofheinz, *J. Am. Chem. Soc.*, 1983, 105, 624–625.
128 T. Ravindranathan, M. Anil Kumar, R. B. Menon and S. V. Hiremath, *Tetrahedron Lett.*, 1990, **31**, 755–758.
129 L. Hsing-Jang, Y. Wen-Lung and Y. C. Sew, *Tetrahedron Lett.*, 1993, **34**, 4435–4438.
130 J. S. Yadav, R. Satheesh Babu and G. Sabitha, *Tetrahedron Lett.*, 2003, **44**, 387–389.
131 M. A. Avery, W. K. M. Chong and C. Jennings-White, *J. Am. Chem. SOC*, 1992, **114**, 974–979.
132 P. S. Covello, *Phytochemistry*, 2008, **69**, 2881–2885.
133 H. Tsuruta, C. J. Paddon, D. Eng, J. R. Lenihan, T. Horning, L. C. Anthony, R. Regentin, J. D. Keasling, N. S. Renninger and J. D. Newman, *PLoS One*, 2009, **4**, e4489.
134 P. J. Westfall, D. J. Pitera, J. R. Lenihan, D. Eng, F. X. Woolard, R. Regentin, T. Horning, H. Tsuruta, D. J. Melis, A. Owens, S. Fickes, D. Diola, K. R. Benjamin, J. D. Keasling, M. D. Leavell, D. J. McPhee, N. S. Renninger, J. D. Newman and C. J. Paddon, *Proc. Natl. Acad. Sci.*, 2012, **109**, E111-E118.
135 C. J. Paddon, P. J. Westfall, D. J. Pitera, K. Benjamin, K. Fisher, D. McPhee, M. D. Leavell, A. Tai, A. Main, D. Eng, D. R. Polichuk, K. H. Teoh, D. W. Reed, T. Treynor, J. Lenihan, H. Jiang, M. Fleck, S. Bajad, G. Dang, D. Dengrove, D. Diola, G. Dorin, K. W. Ellens, S. Fickes, J. Galazzo, S. P. Gaucher, T. Geistlinger, R. Henry, M. Hepp, T. Horning, T. Iqbal, L. Kizer, B. Lieu, D. Melis, N. Moss, R. Regentin, S. Secrest, H. Tsuruta, R. Vazquez, L. F. Westblade, L. Xu, M. Yu, Y. Zhang, L. Zhao, J. Lievense, P. S. Covello, J. D. Keasling, K. K. Reiling, N. S. Renninger and J. D. Newman, *Nature*, 2013, **496**, 528–532.
136 J. Turconi, F. Griolet, R. Guevel, G. Oddon, R. Villa, A. Geatti, M. Hvala, K. Rossen, R. Göller and A. Burgard, *Org. Process Res. Dev.*, 2014, **18**, 417–422.
137 M. Peplow, *Nature*, 2016, 530, 389–390.
138 CDC, Centers for Disease Control and Prevention, https://www.cdc.gov/malaria/malaria_worldwide/impact.html, (accessed 20 July 2019).
139 A. Persidis, *Nat. Biotechnol.*, 2000, **18**, 111–112.

140 R. S. Bray and P. C. C. Garnham, *Br. Med. Bull.*, , DOI:10.1093/oxfordjournals.bmb.a071746.
141 B. Singh, L. K. Sung, A. Matusop, A. Radhakrishnan, S. S. G. Shamsul, J. Cox-Singh, A. Thomas and D. J. Conway, *Lancet*, 2004, **363**, 1017–1024.
142 M. Prudêncio, M. M. Mota and A. M. Mendes, *Trends Parasitol.*, 2011.
143 A. F. Cowman, D. Berry and J. Baum, *J. Cell Biol.*, 2012, **198**, 961–971.
144 M. S. Oakley, N. Gerald, T. F. McCutchan, L. Aravind and S. Kumar, *Trends Parasitol.*, 2011.
145 R. Idro, K. Marsh, C. C. John and C. R. J. Newton, *Pediatr. Res.*, 2010, 68, 267–274.
146 G. M. de Jong, M. B. B. McCall, W. A. Dik, R. T. Urbanus, L. J. Wammes, R. Koelewijn, R. W. Sauerwein, A. Verbon, J. J. van Hellemond and P. J. J. van Genderen, *Cytokine*, 2020, **125**, 154838.
147 T. E. Wellems and C. V Plowe, *J. Infect. Dis.*, 2001, **184**, 770–776.
148 C. Hopkins Sibley, J. E. Hyde, P. F. G. Sims, C. V. Plowe, J. G. Kublin, E. K. Mberu, A. F. Cowman, P. A. Winstanley, W. M. Watkins and A. M. Nzila, *Trends Parasitol.*, 2001, **17**, 582–588.
149 O. Reamtong, K. Srimuang, N. Saralamba, P. Sangvanich, N. P. J. Day, N. J. White and M. Imwong, *Int. J. Mass Spectrom.*, 2015, **391**, 82–92.
150 M. Ghavami, C. H. Dapper, S. Dalal, K. Holzschneider, M. Klemba and P. R. Carlier, *Bioorganic Med. Chem. Lett.*, 2016, **26**, 4846–4850.
151 A. Acheampong, A. Gyebi, G. Darko, J. Apau, W. Owusu Gyasi and S. Addai-Arhin, *Cogent Chem.*, 2018, **4**, 1472198.
152 R. K. Haynes, *Curr. Top. Med. Chem.*, 2006, **6**, 509–537.
153 E. Martino, M. Tarantino, M. Bergamini, V. Castelluccio, A. Coricello, M. Falcicchio, E. Lorusso and S. Collina, *Future Med. Chem.*, 2019, **11**, 1443–1459.
154 M. P. Anthony, J. N. Burrows, S. Duparc, J. Jmoehrle and T. N. C. Wells, *Malar. J.*, 2012, **11**, 316.
155 J. Wang, C. Xu, Z. R. Lun and S. R. Meshnick, *Trends Pharmacol. Sci.*, 2017.
156 R. K. Haynes, W. Y. Ho, H. W. Chan, B. Fugmann, J. Stetter, S. L. Croft, L. Vivas, W. Peters and B. L. Robinson, *Angew. Chemie - Int. Ed.*, 2004, **43**, 1381–1385.
157 W. Asawamahasakda, Y.-M. Pu, H. Ziffer and A. R. Meshnick', *Antimicrob. Agents Chemother.*, 1994, **38,** 1854-1858.
158 J. Bhisutthibhan, X.-Q. Pan, P. A. Hossler, D. J. Walker, C. A. Yowell, J. Carlton, J. B. Dame and S. R. Meshnick, *J. Biol. Chem.*, 1998, **273**, 16192-16198.
159 D. J. Walker, J. L. Pitsch, M. M. Peng, B. L. Robinson, W. Peters, J. Bhisutthibhan and S. R. Meshnick, *Antimcirob. Agents Chemother.*, 2000, **44**, 344-347.
160 U. Eckstein-Ludwig, R. J. Webb, I. D. A. Van Goethem, J. M. East, A. G. Lee, M. Kimura, P. M. O'neill, P. G. Bray, S. A. Ward and & S. Krishna, Nature, 2003, **424**, 957-961.
161 R. Jambou, E. Legrand, M. Niang, N. Khim, P. Lim, B. Volney, M. T. Ekala, C. Bouchier, P. Esterre, T. Fandeur and O. Mercereau-Puijalon, *Lancet*, 2005, **366**, 1960–1963.
162 A. C. Uhlemann, A. Cameron, U. Eckstein-Ludwig, J. Fischbarg, P. Iserovich, F. A. Zuniga, M. East, A. Lee, L. Brady, R. K. Haynes and S. Krishna, *Nat. Struct. Mol. Biol.*, 2005, **12**, 628–629.
163 S. G. Valderramos, D. Scanfeld, A. C. Uhlemann, D. A. Fidock and S. Krishna, *Antimicrob. Agents Chemother.*, 2010, **54**, 3842–3852.
164 P. M. O'Neill, S. L. Rawe, K. Borstnik, A. Miller, S. A. Ward, P. G. Bray, J. Davies, C. H. Oh and G. H. Posner, *ChemBioChem*, 2005, **6**, 2048–2054.
165 M. D. P. Crespo, T. D. Avery, E. Hanssen, E. Fox, T. V. Robinson, P. Valente, D. K.

Taylor and L. Tilley, *Antimicrob. Agents Chemother.*, 2008, **52**, 98–109.
166 W. Li, W. Mo, D. Shen, L. Sun, J. Wang, S. Lu, J. M. Gitschier and B. Zhou, *PLoS Genet.*, 2005, **1**, 329–334.
167 J. Wang, L. Huang, J. Li, Q. Fan, Y. Long, Y. Li and B. Zhou, *PLoS One*, 2010, **5**, e9582.
168 H. Noedl, Y. Se, K. Schaecher, B. L. Smith, D. Socheat and M. Fukuda, *N. Engl. J. Med.*, 2008, **359**, 2619–2620.
169 A. M. Dondorp, F. Nosten, P. Yi, D. Das, A. P. Phyo, J. Tarning, K. M. Lwin, F. Ariey, W. Hanpithakpong, S. J. Lee, P. Ringwald, K. Silamut, M. Imwong, K. Chotivanich, P. Lim, T. Herdman, S. S. An, S. Yeung, P. Singhasivanon, N. P. J. Day, N. Lindegardh, D. Socheat and N. J. White, *N. Engl. J. Med.*, 2009, **361**, 455–467.
170 J. Straimer, N. F. Gnädig, B. Witkowski, C. Amaratunga, V. Duru, A. P. Ramadani, M. Dacheux, N. Khim, L. Zhang, S. Lam, P. D. Gregory, F. D. Urnov, O. Mercereau-Puijalon, F. Benoit-Vical, R. M. Fairhurst, D. Ménard and D. A. Fidock, *Science.*, 2015, **347**, 428-431.
171 A. Mbengue, S. Bhattacharjee, T. Pandharkar, H. Liu, G. Estiu, R. V. Stahelin, S. S. Rizk, D. L. Njimoh, Y. Ryan, K. Chotivanich, C. Nguon, M. Ghorbal, J. J. Lopez-Rubio, M. Pfrender, S. Emrich, N. Mohandas, A. M. Dondorp, O. Wiest and K. Haldar, *Nature*, 2015, **520**, 683–687.
172 K. M. Tun, M. Imwong, K. M. Lwin, A. A. Win, T. M. Hlaing, T. Hlaing, K. Lin, M. P. Kyaw, K. Plewes, M. A. Faiz, M. Dhorda, P. Y. Cheah, S. Pukrittayakamee, E. A. Ashley, T. J. C. Anderson, S. Nair, M. McDew-White, J. A. Flegg, E. P. M. Grist, P. Guerin, R. J. Maude, F. Smithuis, A. M. Dondorp, N. P. J. Day, F. Nosten, N. J. White and C. J. Woodrow, *Lancet Infect. Dis.*, 2015, **15**, 415–421.
173 S. Krishna, C. J. Woodrow, H. M. Staines, R. K. Haynes and O. Mercereau-Puijalon, *Trends Mol. Med.*, 2006, **12**, 200–205.
174 X. Tang, M. Demiray, T. Wirth and R. K. Allemann, *Bioorganic Med. Chem.*, 2018, **26**, 1314–1319.

2 Chapter 2 Investigation into the promiscuity of the ADS active site

2.1 Introduction

Amorpha-4,11-diene (**34**) is a sesquiterpene precursor of the anti-malarial artemisinin (**4**). Although the importance of understanding the mechanism of action of amorpha-4,11-diene synthase (ADS) is widely acknowledged, attempts to obtain an X-ray crystal structure have yet to succeed. ADS is a 62.2 kDa class I terpene cyclase which is predicted to adopt the mainly α-helical fold characteristic of this family.[12] Residues corresponding to the conserved DDxxD/E and NSE/DTE magnesium binding motifs of this fold have been identified[3] as have some additional key residues and their roles (discussed below), but there is little more detail in the literature regarding how the structure of this enzyme influences its mechanism. A homology model was therefore required, which can be used to predict residues involved in guiding the enzymes mechanism and how they might interact with the substrate or any cofactors in the active site. Mutational studies then allow us to test these hypotheses by selecting specific changes to illustrate how residues are involved in shaping the product, positioning the substrate and/or cofactors. Mutations can also change enzymatic behaviour, including altering the product profile,[4] the catalytic efficiency[5] or enzyme stability.[6]

Previous investigations into sesquiterpene synthases *via* site directed mutagenesis have allowed interconversion of related enzymes,[7] or narrowing of product distributions.[8] For example, 5-*epi*-aristolochene synthase from *Nicotiana tabacum* and premnaspirodiene synthase from *Hyoscyamus muticus* share a high amino acid identity and the same catalytic mechanism for the initial steps of the conversions they catalyse. When 9 amino acids of 5-*epi*-aristolochene synthase (EAS) were swapped for the corresponding amino acids in premnaspirodiene synthase (PS), the activity of the mutant EAS resembled that of wild type PS. When the corresponding 9 residues of EAS were introduced to PS, the activity of the mutant PS resembled that of wild type EAS.[7] Mutagenesis of γ-humulene synthase also illustrate how the product distribution of a promiscuous enzyme can be modified to produce a narrower range of substrates. This sesquiterpene synthase naturally produces 52 products. An algorithm that looks at the effect of changing residues that are predicted to be in the active site, guided the production of multiple highly-specific sesquiterpene synthases with novel products;

e.g. just three mutated residues converted γ-humulene synthase into a highly specific sibirene synthase, an enzyme activity not yet identified in nature.[8]

Certain residues in class I terpene synthases are proposed to have a large impact on the product profile of the enzyme. The metal binding motifs of terpene synthase are key for positioning cations to assist in the removal of the diphosphate group to initiate the reaction. When these motifs are mutated, a reduction or outright loss of activity is commonly seen, but changes in the product profile are also common.[9,10] One such example is the mutation of the metal binding domains of aristolochene synthase, where the mutation D115E caused an increase in the proportion of (*S*)-(-)-germacrene A and (-)-valencene (**68**) side products. The N244D mutant showed an even more significant change in product profile, with (*S*)-(-)-germacrene A becoming the major product.[9] Similar results were seen with pentalenene synthase, when different aspartate residues of the metal binding motifs were mutated to glutamate varying proportions of the side product (+)-germacrene A were produced.[10]

The G helix, logically divided into G1 and G2 helices by a kink in the middle, has been reported to play a key role in product determination in a variety of sesquiterpene synthases and is a focus of particular interest. For example, mutations to residues in the G helix of selina-4(15),7(11)-diene synthase increased production of the side product germacrene B (**71**).[11] When the G helix of δ-cadinene synthase was subjected to site directed mutagenesis an increase in the alternate hydroxylated product, germacradien-4-ol, was seen.[12,13] Mutations in the G helix of germacradien-11-ol synthase were shown to significantly affect the hydroxylation step of the mechanism, and showed to subtly affect the shape of the active site.[14] A similar motif is present in monoterpene synthases and mutations in this region lead to an altered product profile in 1,8-cineole synthase, converting it to a sabinene synthase.[15] This helix breaking kink is also present in diterpene synthase structures, and when mutated in a kaurene synthase from rice, the major product was changed from *ent*-isokaur-15-ene to *ent*-primara-8(14),15-diene.[16]

There are many literature reports of investigating the activity and mechanism of ADS *via* site-directed mutagenesis. Previous work includes the creation of mutants with greater activity, ADS T399S, postulated to be due to a faster release of product from the active site.[17] Another single residue change in ADS, T296V, impacted on the ability of ADS to isomerise FDP (**40**) to the normal (3*R*,6*E*)-nerolidyl intermediate, with the main product changing from amorpha-

4,11-diene (**34**) to (*E*)-*β*-farnesene.[18] It is hypothesised the hydroxy group of the threonine helps stabilise the intermediate carbocation as FDP (**40**) is isomerised to nerolidyl diphosphate (**48**).[19] The Allemann group also truncated the *N*-terminus of the protein, based on the hypothesis that this could allow more water into the active site, increasing the percentage of hydroxylated products as this seemed effective with δ-cadinene synthase. Unfortunately, this was not successful, and the enzyme activity was reduced[20].

Extensive mutagenesis by Abdallah *et al.* identified several key residues for catalysis, including N443 which binds a magnesium ion coordinating FDP (**40**) and F525 whose aromatic ring helps stabilise a carbocation intermediate.[3] Further research from this group created a T399S/H448A double mutant which had a k_{cat} 5-fold greater than that of the wild type. They also identified other important residues such as R262 which binds the diphosphate group, W271 which acts as a base in the active site aiding in the removal of a proton, and G400, G439, and L515 which were all indicated to aid 1,10 ring closure.[5] Studies by Fang *et al.* noted other residues that play a part in 1,10 ring closure, such as G401 which when mutated to alanine impaired 1,10 cyclization, as did L374Y, and F370L. Whereas, T399 and T447 are proposed to bind water *via* their hydroxy groups and aid deprotonation at H12/H13 because mutation of these residues caused an increase in the variety of side products, including amorpha-4,7-diene.[19]

2.2 Aims

The aim of this work was to investigate the flexibility of the active site of ADS and to determine key residues that aid in substrate cyclisation. It has been shown that single point mutations can have a dramatic impact on product ratio, indicating key residues determine an enzymes product profile.[21] A 3D homology model was needed to guide further mutagenesis to allow residues hypothesised to determine product profile to be chosen. Once the model was created various mutations could be chosen to clarify if certain residues are involved in the ADS cyclisation mechanism. As the metal binding regions and G helix of sesquiterpene synthases, and the corresponding regions of other class I terpene synthases, have been shown to play a part in product profile determination, these were chosen as the primary regions of focus.[22]

2.3 Results

Characterisation of Wild Type Amorpha-4,11-diene Synthase

Characterisation of GC-MS spectrum

The ADS gene from *Artemisia annua* was previously cloned into the *pET-21d E. coli* expression vector in the correct reading frame to give a C-terminal His_6 tag fusion.[20] It was expressed in BL21(DE3) cells and purified by nickel column chromatography. See materials and methods 6.1.24, and supplementary information 7.1.1.1.

Assays were performed to confirm the product distribution of the wild type enzyme and to obtain an authentic retention time and mass spectrum for amorpha-4,11-diene (**34**) allowing comparison with any products obtained from ADS variants. Briefly, ADS was incubated with FDP (**40**), in a magnesium containing buffer. Reactions were overlaid with pentane and left to incubate overnight, see materials and methods 6.1.26 6. The pentane extract was then separated and injected into a GC-MS, and the retention time and fragmentation pattern, Figure 51, compared to that of known standards of amorpha-4,11-diene (**34**), see materials and methods 6.1.36.4. The retention time and fragmentation pattern matched that of a known standard of amorpha-4,11-diene, confirming the enzyme was active and produced the correct compound.

ADS produces amorpha-4,11-diene as its major product, reported as 91.2%, but it produces numerous other compounds as very minor side products. These include amorpha-4,7-diene (3.7%), amorpha-4-en-7-ol (2.1%), γ-humulene (1.0%), (*E*)-β-farnesene (0.8%), β-sesquiphellandrene (0.5%), (*E*)-α-bisabolol (0.3%) and amorpha-4-en-11-ol (0.2%). Some of these minor products can be seen in the TIC of assays with WT ADS. (*E*)-β-farnesene was available as an authentic standard and could therefore be confirmed as the peak with the retention time 11.83 minutes, but none of the other minor products mentioned were available.[23]

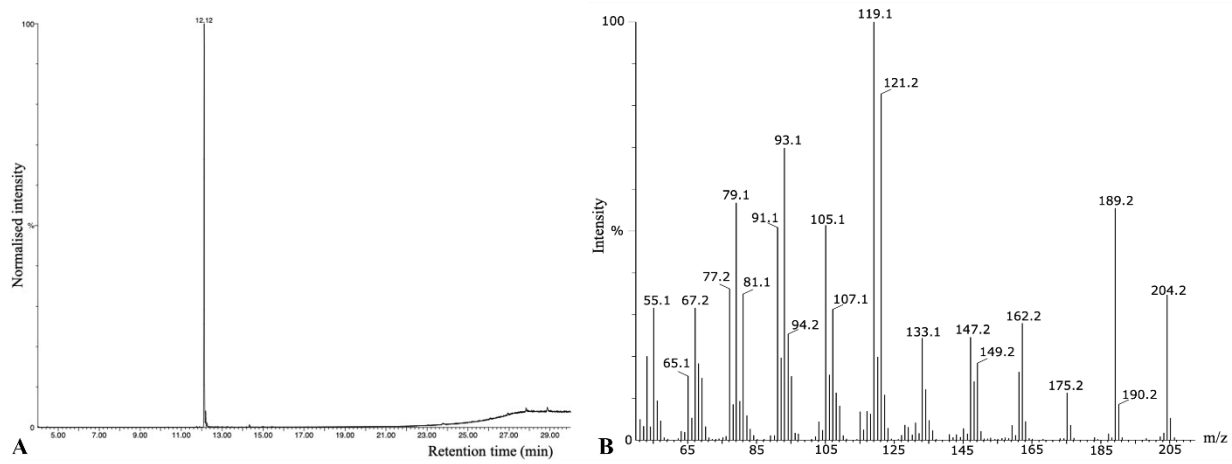


*Figure 51: A) Total ion current chromatogram (TIC) of the pentane extracts arising from the incubations of wild type amorpha-4,11-diene synthase (1 μM) in a buffer containing FDP (**40**) (0.4 mM) and magnesium chloride (5 mM). Amorpha-4,11-diene (**34**) is the major product, eluting at 12.12 minutes. B) EI mass spectrum of retention time 12.12 minutes in A (amorpha-4,11-diene (**34**))*

Characterisation of wild type ADS Steady State Kinetics

Radioactivity assays to determine the kinetic parameters of wild type ADS were carried out as a reference for subsequent mutants (see materials and methods 6.1.26). Briefly tritiated FDP across a range of concentrations was used in assays with ADS for a set time, to determine steady state kinetics. The hydrophobic product, amorpha-4,11-diene, was extracted in an organic solvent and the radioactivity quantified using a scintillation counter. The substrate is not soluble in the organic solvent and is not present in the organic extract. The results are used to plot a Michaelis-Menten curve, Figure 52, allowing the calculation of k_{cat} and K_M as an average of three separate experiments. The values obtained were in good agreement with the values obtained previously.[20]

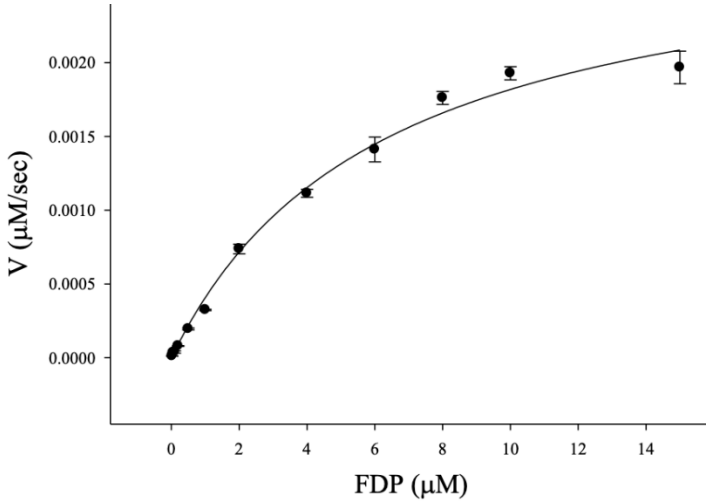


*Figure 52: Michaelis-Menten plot for the conversion of FDP (**40**) by ADS to amorpha-4,11-diene (**34**)*

Mutagenesis of Amorpha-4,11-diene Synthase

Building a Homology Model

An ADS homology model was built using the I-TASSER software[24–26] to thread the amino acid sequence of *Artemisia annua* ADS onto the crystal structure of 5-*epi*-aristolochene synthase (PDB: 5EAT), as shown in figure 53. 5-*Epi*-aristolochene synthase was chosen as it shares 56% sequence similarity with ADS (as calculated by pairwise sequence alignment) at amino acid level. The crystal which generated this structure was co crystallised with magnesium chloride and FHP,[27] an FDP (**40**) analogue that allows a prediction of where FDP (**40**) would sit in the active site of ADS. This model was used to choose residues that seemed to interact with FDP (**40**), and particular attention was paid to residues in the G helix, and those in and around the metal binding motif.

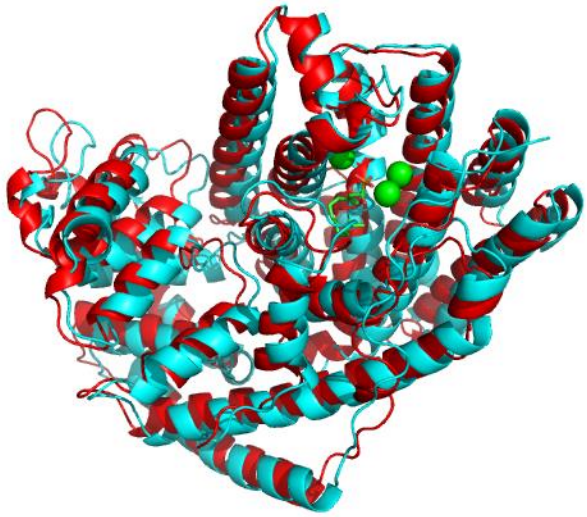


Figure 53: Cartoon representation of an overlay of homology model built of ADS (cyan), and the crystal structure of epi-aristolochene synthase (PDB 5EAT)[27] (red). FHP and magneisum ions are shown in green.

The active site with FHP is shown in Figure 54. Residues that lay within a reasonable distance (<10 Å) of the substrate, either by coordinating the diphosphate group or its attendant magnesium ions or guiding its folding, and residues in the G helix that could play a part in shaping the active site were carefully examined. The existing literature regarding ADS mutants was also used to guide decisions.

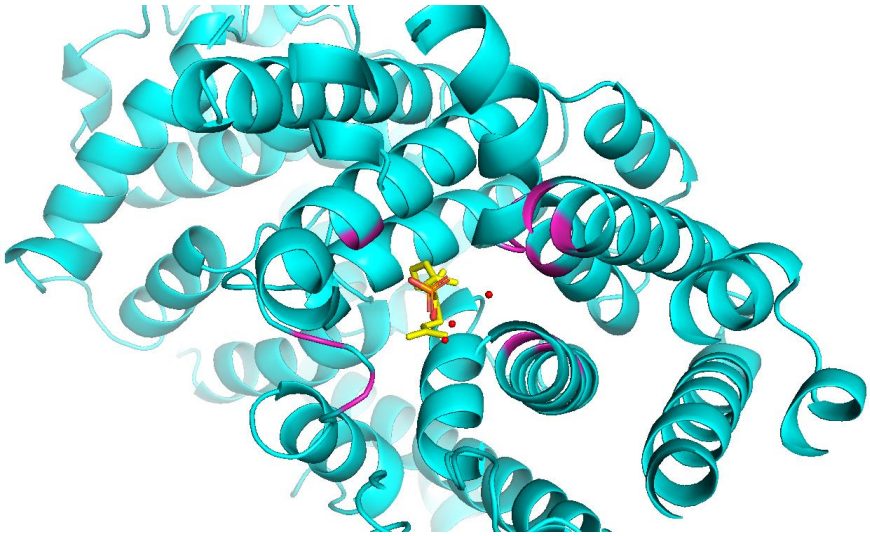


*Figure 54: Cartoon representations of the active site of the ADS homology model (cyan) with residues of interest coloured magenta within the backbone, FDP (**40**) mimic FHP (yellow) and magnesium ions (red).*

Choosing Residues to Change

Many of the positions chosen for mutation were in the G helix of the homology model of ADS (I397, G401, A402, N403, L405), with other residues that were predicted to interact with FDP (**40**) based on FHP in the active site (I295, D299, L374, E377, N443, N524, and T526) see Figure 55.

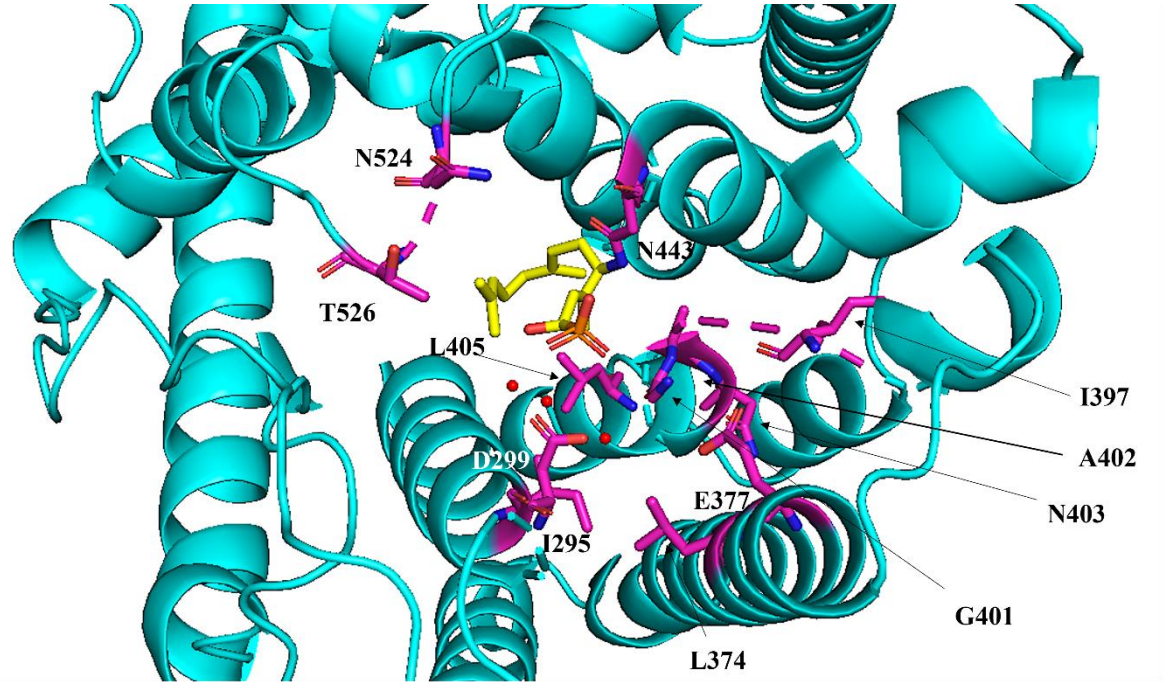


*Figure 55: Cartoon representations of the ADS homology model (cyan) indicating residues chosen for site-directed mutagenesis (magenta), FDP (**40**) mimic FHP (yellow), and magnesium ions (red).*

Mutations were carried out by the method described by Liu *et al.*[28] where primers are designed to incompletely overlap with the changed codon located within the overlapping section (see

materials and methods 6.1.15 for further detail). Once the desired DNA sequences were confirmed by sequencing by Eurofins Genomics, they were transformed into expression cell lines. Mutant ADS were expressed and purified according to the same protocols as for the wild type enzyme and kinetic analyses were carried out as previously described and the results compared to those obtained for wild type ADS.

Assay Results

Non-Perturbing Mutations

Many of the mutations yielded enzymes that produced very similar product profiles to wild type ADS, see supplementary information 7.1.2. Some of the non-perturbing mutations showed very minor variations in the quantity of side products produced, compared to WT ADS, that are visible in the TIC. Nevertheless, none showed peaks with different retention times to the TIC of WT ADS, suggesting no novel products were produced. None of the variants TICs showed any significant variation in the size of the peaks of the side products produced compared to the WT TIC, suggesting the variants did not impact the cyclisation mechanism in a significant way. These mutants included: I295F, I295A, L374A, E377A, I397Y, A402Y, A402L, N403S, L405F, N524D, T526L and T526D. These residues are shown in Figure 56.

I295 was chosen as it sits close to the DDXXD motif, in the expectation that mutations at this position may lead to distortions in this important region. The corresponding residue at this position is commonly alanine in other sesquiterpenes. When I295 was mutated to the conserved alanine little change was seen in the K_M but the k_{cat} decreased dramatically, see Table 3. Mutating I295 to phenylalanine had a similar consequence to the k_{cat} but also greatly decreased the K_M compared to wild type. Phenylalanine is larger than isoleucine and alanine and may create a tighter fit for FDP (**40**). k_{cat} is affected by product release, the product is released along with the magnesium ions in the active site. Often a network of water molecules aid in coordinating these magnesium ions. Changes in the amino acid residues may affect these networks, such as cation-π interactions between phenylalanine and magnesium ions, affecting the release of these ions from the active site.

L374 is highly conserved, as shown in Figure 57, as a tyrosine residue in many other sesquiterpene synthases, unlike ADS, and in the homology model, see Figure 56, sits near the

magnesium ions that coordinate diphosphate. When mutated to alanine, a much smaller residue the enzymes turnover was too slow to measure. It may be that a larger residue such as leucine is needed to aid in forming the structure of the active site, and with an alanine residue the magnesium ions or substrate may be able to escape the active site. Too large a residue, such as the generally conserved tyrosine, disrupts 1,10-ring closure.[19]

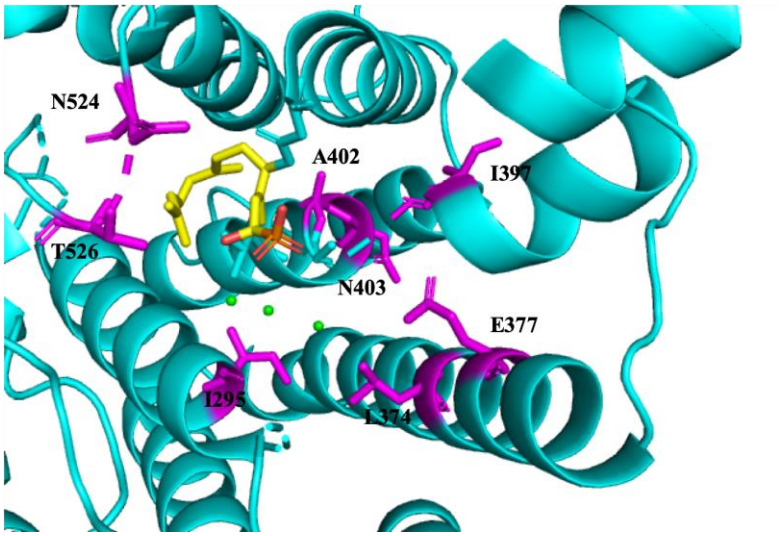


Figure 56: Cartoon representation of homology model of ADS (cyan) focused on the active site, FHP (yellow), and magnesium ions (green) showing the residues mutated which caused no change to the active site and their position in comparison to the active site and substrate.

E377 and T526 are highly conserved residues throughout a large number of sesquiterpene synthases including germacrene D synthase, *epi*-cedrol synthase and δ-cadinene synthase, as shown in Figure 57. Mutating E377 to alanine resulted in slight change in K_M with large reduction in $k_{cat.}$ Changing glutamic acid to alanine removes a potential site for hydrogen bonding, this may change the interaction with magnesium ions or water molecules. It was proposed by Dickschat *et al.* that a collapse in hydrogen bonding network in the active site of selinadiene synthase may be important for product release. If ADS has a similar network within the active site, changes may hinder product release slowing the overall reaction.[29] Both mutations of T526, to aspartate and leucine respectively, decreased the $k_{cat,}$ but the mutation to aspartate also produced a significant increase to the K_M. T526 sits near the diphosphate group of FDP (**40**) in the homology model of ADS. It may be that the negative charge of aspartate repulses the diphosphate group leading to the increase in K_M. A similar issue to E377A may have arose by mutating T526 to leucine, as threonine may hydrogen bond to water, but leucine would be unable to. Mutating T526, which is near the diphosphate group of FDP, to aspartate,

while possibly repulsing the diphosphate group may also introduce new interactions with the magnesium ions, which may hinder their release from the active site.

A402, N403 and L405 sit in the G helix and distortion of this structure is known to affect product profile. A402 is highly conserved as a tyrosine residue at this position, and L405 is highly conserved as leucine, with only *epi*-cedrol synthase having alanine in this position in the alignments obtained, (Figure 57). When examining the homology model A402 and N403 do not branch into the active site, Figure 56, the side chains point outwards from the protein backbone. This may explain why changes in these residues do not change the product profile, while changes to D299, G401 and L405, which do point into the active site, can. Although no change in product profile was seen with mutations to A402 and N403, the structure of the G helix clearly plays a part in the enzymes catalysis as rates of these mutants were too slow to measure or did not follow Michaelis-Menten kinetics. Both mutants of A402 exhibited sigmoidal kinetics, likely due to a change in binding of the substrate, meaning a larger concentration of substrate was needed before the active site was inhabited. While L405A showed a change in product profile, L405F did not. L405F decreased the k_{cat} suggesting that position 405 needs a moderately sized residue to aid in shaping the product as alanine was too small to retain the product profile, but too large a residue impinges on the rate of catalysis. Introducing a phenylalanine residue with the potential for cation-π interactions with magnesium ions may affect the overall network of water molecules and magnesium ions, and may also slow the release of the magnesium from the active site.

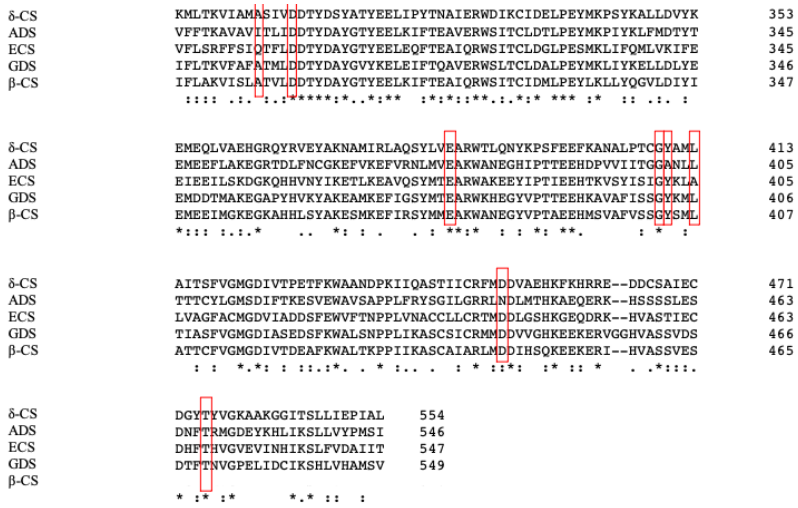


Figure 57:Sequence alignment of δ-cadinene synthase (δ-CS) (O49853), ADS (Q9AR04), epi-cedrol synthase (ECS) (Q9LLR9), germacrene D synthase (GDS) (I6QPS5) and β-caryophyllene synthase (β-CS) (Q8SA63). Sequence alignment were done using the Clustal Omega software. Codes in brackets signify the accession numbers.

N524 may be part of the catalytic triad responsible for later steps of the mechanism *e.g.* in 5-*epi*-aristolochene synthase it is needed for a re-protonation step.[30] When mutated to aspartate the K_M and k_{cat} decreased compared to wild type. If this residue is truly part of the re-protonation step, it may be that the negative charge of aspartate compared to asparagine helps stabilise the carbocation intermediate. Mutating asparagine 524 to aspartate, which is near the diphosphate group within the active site, introduces a residue with a negative charge. It is possible the residue had novel interactions with the magnesium ions which coordinate the diphosphate group, slowing their, and subsequently the products, release. Of all the mutants which had similar product profiles to ADS, I397Y is the only mutant which retained similar kinetic activity, with both K_M and k_{cat} measured at close values to wild type.

Table 3: Steady state kinetic parameters of non-perturbing mutations

Mutant	K_M ± Error (SE) (μM)	k_{cat} ± Error (SE) (s^{-1}) (x 10^{-3})	Catalytic efficiency ± Error (propagation of error) ($\mu M^{-1} s^{-1}$) (x 10^{-3})	Type of variant
WT	6.24 ± 0.87	30.00 ± 2.00	4.81 ± 0.74	Non-perturbing
I295A	7.57 ± 2.44	1.00 ± 0.20	0.13 ± 0.05	
I295F	0.67 ± 0.11	3.00 ± 0.11	4.51 ± 0.78	
D299E	2.94 ± 0.85	1.00 ± 0.12	0.34 ± 0.11	Perturbing
L374A	Not determined*			Non-perturbing
E377A	7.65 ± 1.67	7.00 ± 0.77	0.92 ± 0.22	
I397Y	5.27 ± 1.32	18.00 ± 2.00	3.41 ± 0.93	
G401L	5.58 ± 1.44	9.00 ± 1.00	1.60 ± 0.46	Perturbing
G401Y	1.09 ± 0.21	7.00 ± 0.31	6.36 ± 1.30	
A402L	Not determined*			Non-perturbing
A402Y	Not determined*			
N403S	Not determined*			
L405A	1.19 ± 0.16	1.00 ± 0.04	0.84 ± 0.12	Perturbing
L405F	2.77 ± 0.69	0.30 ± 0.02	0.11 ± 0.03	Non-perturbing
N443I	Not determined**			Inactive
N443L	Not determined**			
N524D	1.87 ± 0.15	7.00 ± 0.15	3.75 ± 0.31	Non-perturbing
T526D	28.01 ± 17.24	4.00 ± 0.20	0.14 ± 0.09	
T526L	6.69 ± 1.18	3.00 ± 0.28	0.45 ± 0.09	

*Variant was too slow for kinetics to be measurable or did not follow Michaelis-Menten curve.

**Variant was inactive and therefore kinetics could not be measured

Although all residues mentioned above are within a reasonable distance to interact with FDP (**40**) in the homology model they do not seem to have an impact on product profile, see Figure 58 for comparison with wild type chromatogram and mass spectrum, Figure 51. It is clear that the majority of the active site of ADS is very robust to changes in product profile. Some of the residues chosen to mutate were highly conserved *e.g.* T526 which had no effect on the enzymes product. It suggests conservation of amino acids may be as they work in coordination with other residues to aid catalysis or to bind the substrate rather than shaping it. Also, when residues were mutated to ones conserved in other sesquiterpene synthases *e.g.* A402Y no change in product profile was seen suggesting the aberration in ADS is not significant to shaping its substrate to product and is likely to do with catalysis of substrate to product.

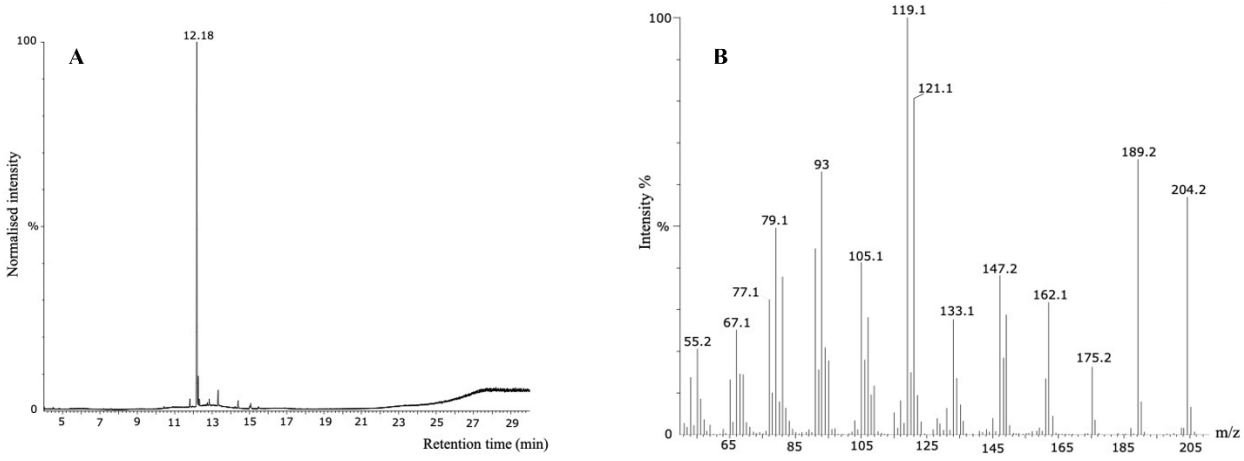


*Figure 58: A) TIC of the pentane extracts arising from the incubations of ADS A402Y (1 μM) which showed the same activity as wild type ADS, in this example, with FDP (**40**) (0.4 mM) and magnesium chloride (5 mM). B) EI mass spectrum of retention time 12.18 minutes in A (amorpha-4,11-diene (**34**))*

Inactive Mutants

Both N443L and N443I showed no product when incubated with FDP (**40**). This residue was proposed to complex magnesium due to its polar side chain. These magnesium ions aid in the removal of the diphosphate group from the FDP (**40**),[31] a key step in the mechanism. When this asparagine residue is mutated to a hydrophobic amino acid enzyme activity is abolished supporting this hypothesis, see supplementary information 7.1.2.2 and 7.1.3.

Mutants With Novel Activities

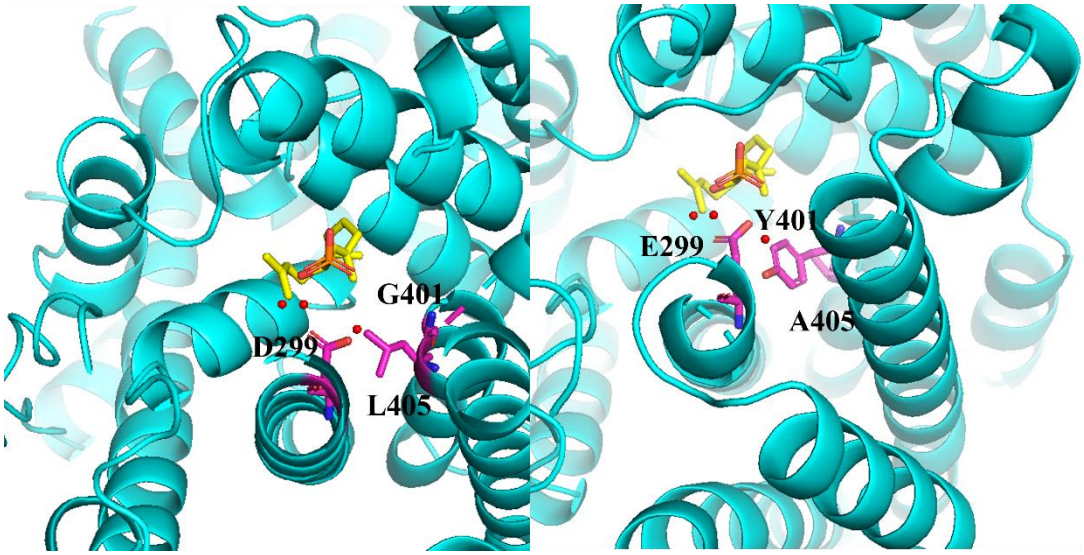


Figure 59: Cartoon representation of homology model of ADS (cyan) focused on the G helix, FHP (yellow), and magnesium ions (red) showing the three residues, D299, G401 and L405 which when mutated to E299, Y401 and A405 respectively cause change in product profile and how they interact with the active site and substrate.

All kinetic experiments were performed as three separate experiments run in triplicate and k_{cat} and K_M were calculated as a mean of the three. Error of k_{cat} and K_M were calculated as the standard error, and the error of the catalytic efficiency was calculated with the formula of the propagation of error. See 6.1.27.4 of materials and methods.

G401Y and G401L

In the homology model residue G401 sits in the kink of helix G1 and G2 which is known to affect product shaping, see Figure 60. Previous work has shown that ADS G401A shows bisabolyl-cation-derived products characteristic of impaired 1,10 cyclisation.[19] When this same residue is mutated to threonine, any activity is reported to be abolished.[3]

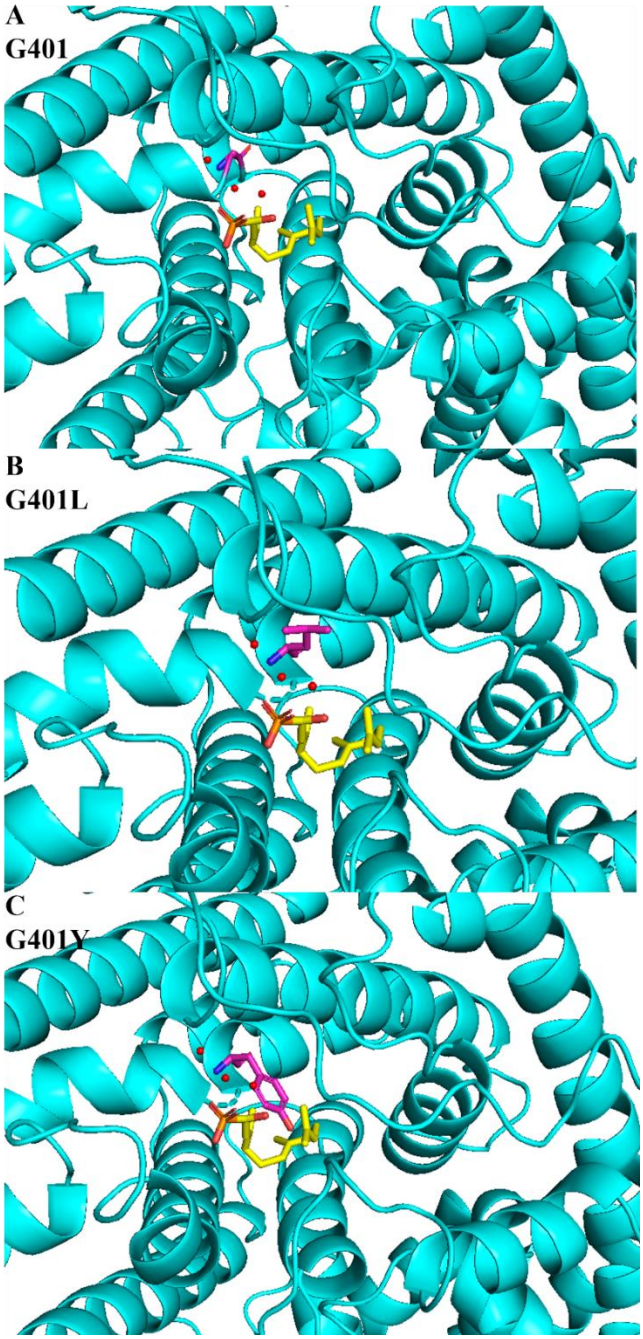


Figure 60: Cartoon representation of homology model of ADS (cyan) focused on the G helix, FHP (yellow), and magnesium ions (green) showing A) the G401 residue (magenta). B) residue 401 mutated to leucine (magenta). C) residue 401 mutated to tyrosine (magenta).

In this work, when G401 was mutated to tyrosine and leucine the product profile showed production of mixtures containing a 1:1 ratio of selina-4(15),7(11)-diene (**101**) and germacrene B (**71**). Both of these amino acids are larger than those previously investigated and much larger than the wild type residue glycine. They are likely to take up more space in the active site and affect the cyclisation mechanism. This result suggests that incorporation of bulkier residues at G401 position changes the initial cyclisation of farnesyl cation from 1-6 ring closure to 1-10

ring closure, to give rise to germacrene carbocation which subsequently undergoes rearrangements to form selina-4(15),7(11)-diene (**101**) (see Figure 61).

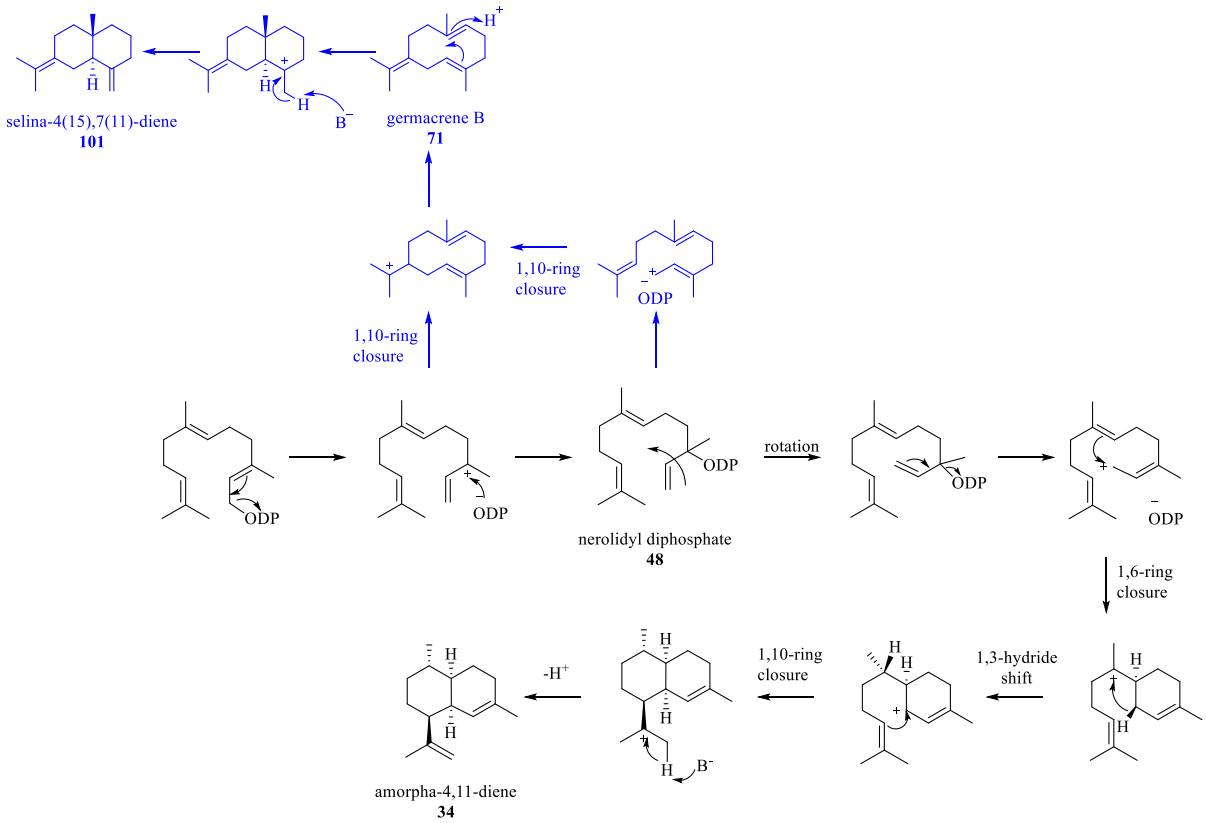


Figure 61: Schematic of the chemical steps catalysed by ADS and SDS showing common intermediates

To ensure that the results obtained were accurate, authentic standards of selina-4(15),7(11)-diene (**101**) and germacrene B (**71**) were obtained from assays performed with selina-4(15),7(11)-diene synthase (SDS) by Dr. P. Srivastava. GC-MS spectra of the compounds obtained from assays with ADS G401L and ADS G401Y were recorded and the fragmentation patterns were compared to those of authentic selina-4(15),7(11)-diene (**101**) and germacrene B (**71**) samples to ensure they corresponded, see Figure 62 to Figure 65, as well as injected together to ensure the peaks co-eluted. These standards were injected onto a GC-FID to compare retention times and were added to pentane extracts from assays with ADS G401L and ADS G401Y to check the peaks co-eluted with the authentic standards. See GC-FID traces in section 7.1.3.2 of supplementary information.

During this work numerous attempts were made to isolate both selina-4(15),7(11)-diene (**101**) and germacrene B (**71**) as pure compounds by column chromatography but this proved incredibly difficult. Large scale assays using G401L were completed, and the pentane extract used for purification. Neither compound could be visualised on ordinary silica TLC, or on

silver nitrate coated silica TLC. Silica columns, with and without silver nitrate, were performed to try and separate the compounds without knowing where they were on the column, with the fractions analysed by GC-MS immediately afterwards. Unfortunately, the compounds were never seen in any of the fractions analysed. Possibly, as they are volatile, they were lost or degraded on the silica.

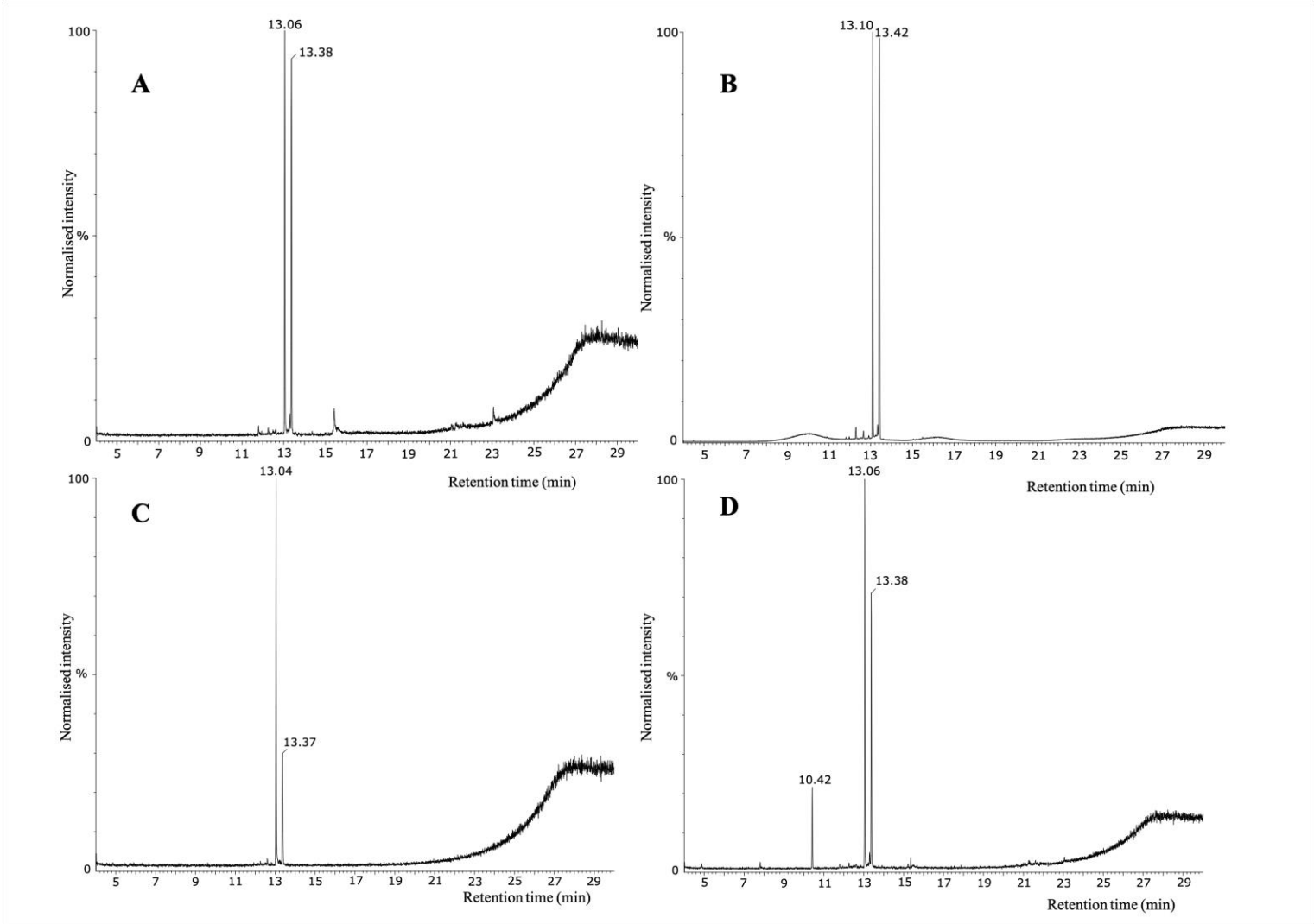


*Figure 62: TIC of the pentane extracts arising from the incubations of enzyme (1 μM) with FDP (**40**) (0.4 mM) and magnesium chloride (5 mM). A) ADS G401L. B) ADS G401Y. C) SDS wt. D) coelution of ADS G401L and SDS extracts*

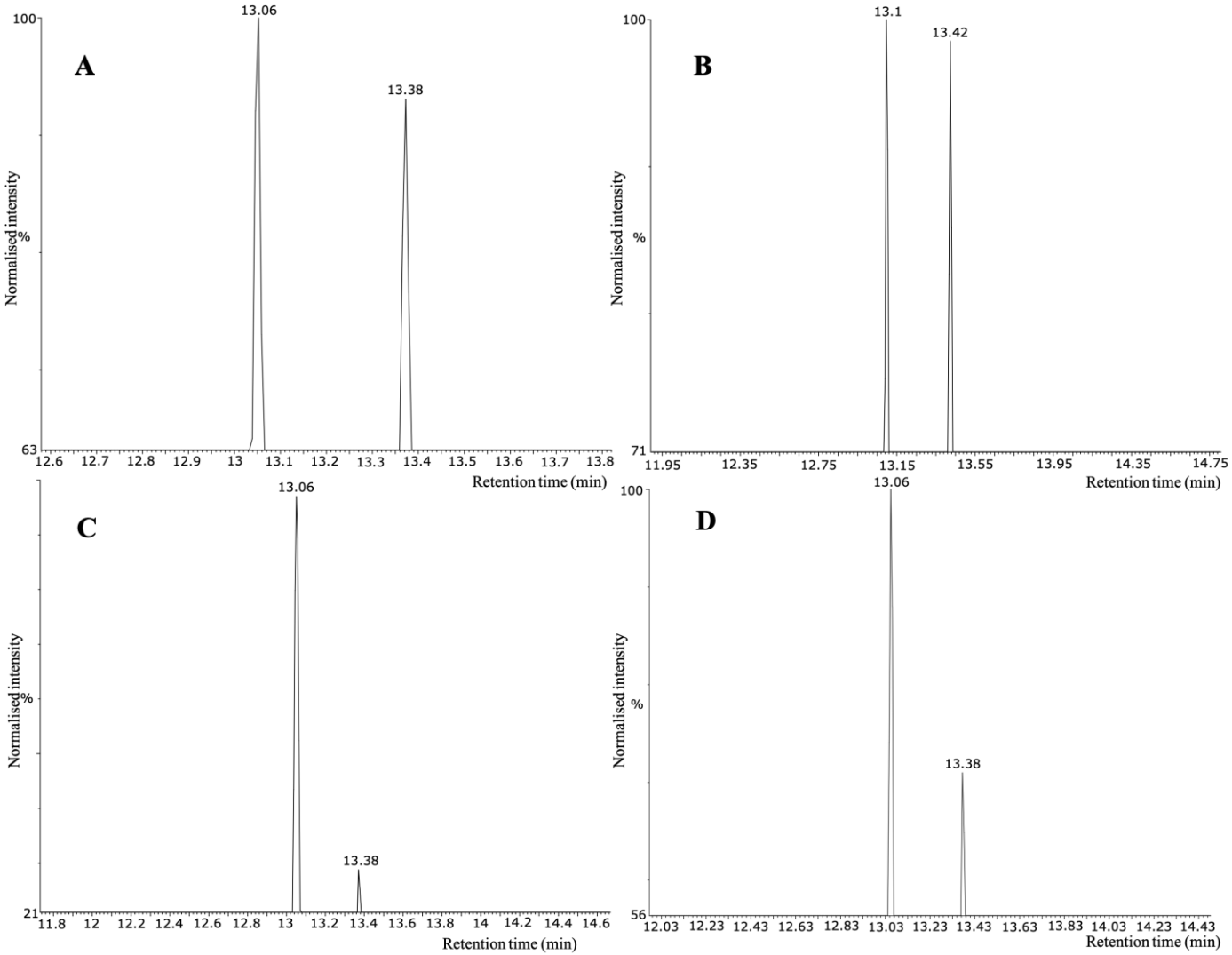


*Figure 63: Magnified images of TIC of the pentane extracts arising from the incubations of enzyme (1 μM) with FDP (**40**) (0.4 mM) and magnesium chloride (5mM). A) ADS G401L. B) ADS G401Y. C) SDS wt. D) coelution of ADS G401L and SDS extracts*

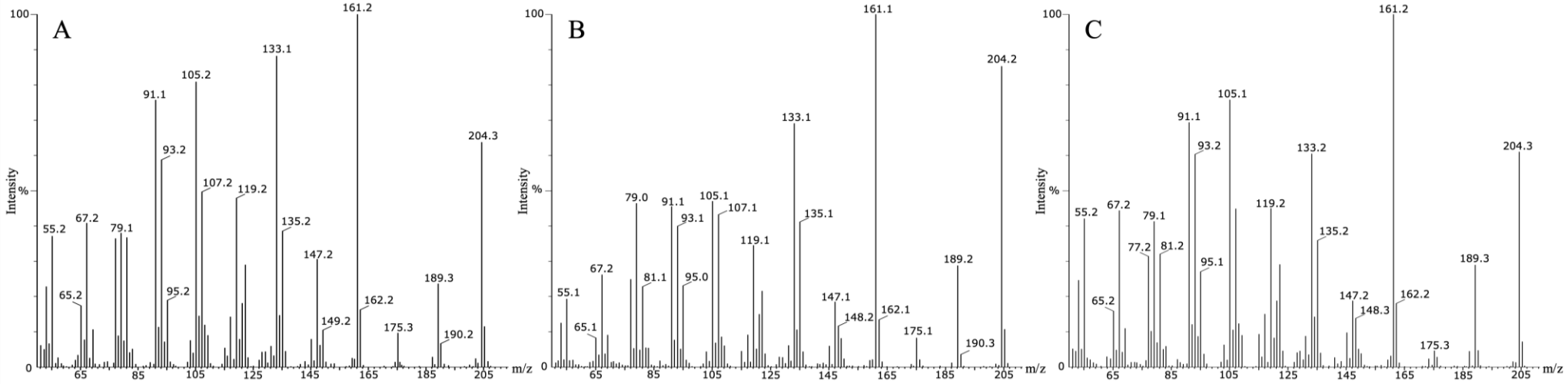


*Figure 64: EI mass spectrum of retention time: A)13.06 minutes of Figure 62A. B)13.10 minutes of Figure 62B. C)13.06 minutes of Figure 62C (selina-4(15),7(11)-diene **101**)*

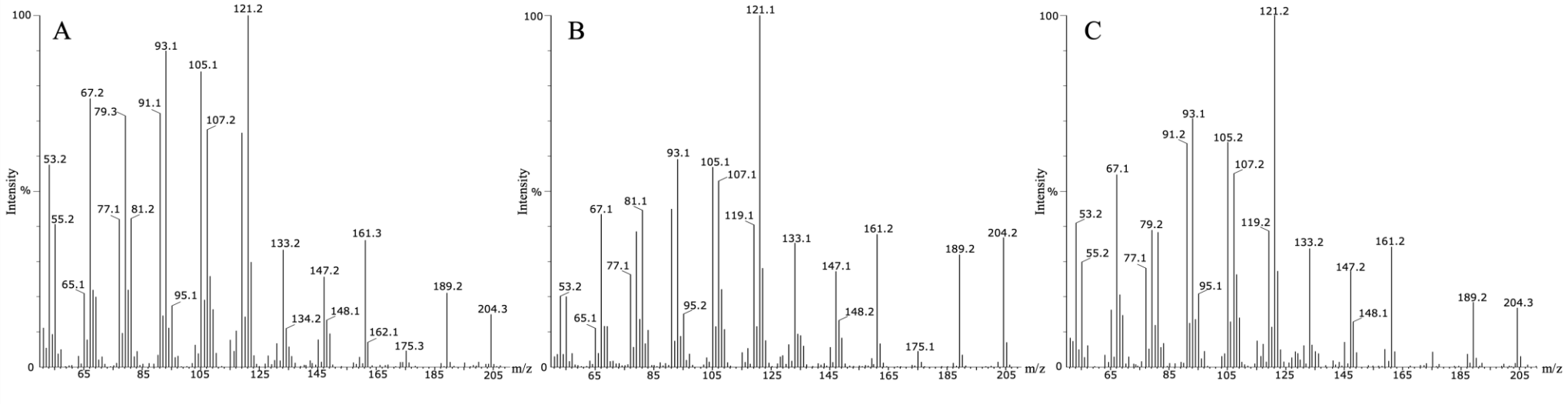


*Figure 65: EI mass spectrum of retention time: A)13.38 minutes of Figure 62A, B) 13.42 minutes of Figure 62B, C)13.38 minutes of Figure 62C (Germacrene B **71**)*

Kinetic parameters were measured for G401Y was found to have comparable k_{cat} and K_M values to wild type ADS, see Figure 66, indicating the promiscuity of the new variant is derived by small changes to the active site pocket without altering the catalytic efficiency. The catalytic efficiency for G401L is three-fold slower than the value obtained for the wild type ADS, which is not a dramatic change. It is likely that the small changes to the active site pocket did not dramatically alter the catalytic efficiency.

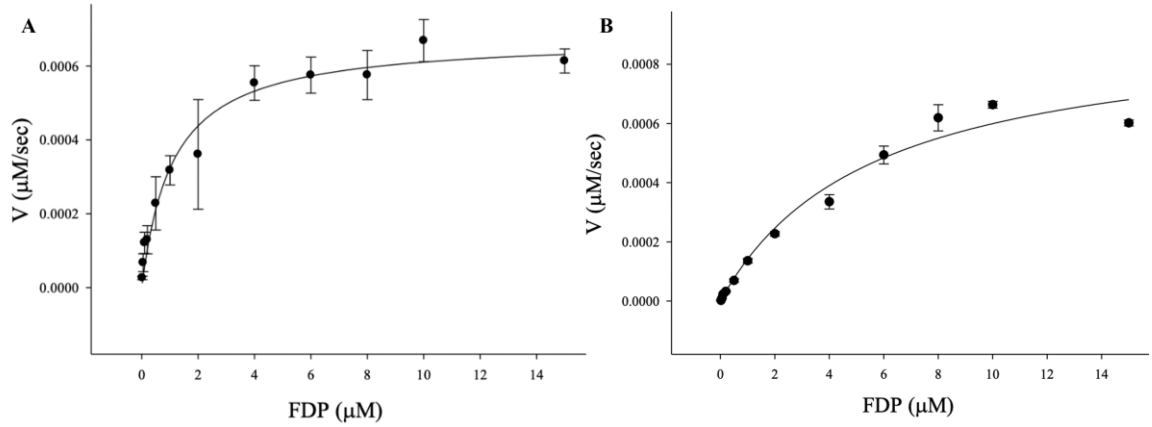


*Figure 66: Michaelis-Menten plot for the conversion of FDP (**40**) by A) G401Y B) G401L ADS to sesquiterpene products*

L405A

The L405 residue sits in the G helix just after the kink between G1 and G2 helices as seen in the homology model in Figure 59. In previous work this residue has been proposed to aid the 1,10 ring closure as an increase in bisabolyl cation derived products was noted when this leucine was mutated to isoleucine.[19] When this residue was mutated to alanine a similar result was seen. There was an increase in the proportion of bisabolene (**65**) and bisabolol (**72**) production, along with linear product β-farnesene, as shown in Figure 67. This mutant has a more than seven-fold reduction in the catalytic efficiency compared to wild type ADS. As shown in Figure 68, as the catalytic efficiency is slower there is large error in the data.

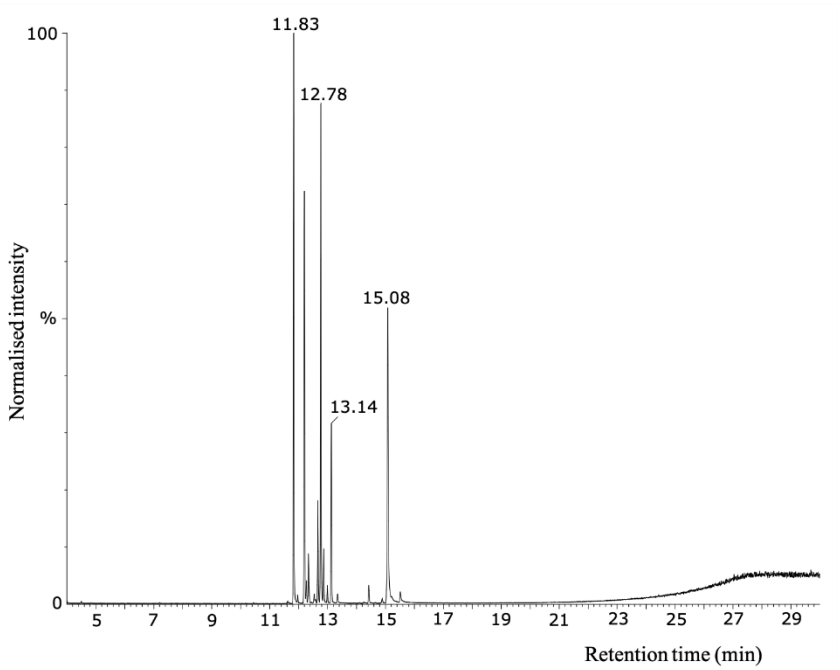


*Figure 67: TIC of the pentane extracts arising from the incubations of L405A (1 μM) with FDP (**40**) (0.4 mM) and magnesium chloride (5 mM). The peak at 11.83 minutes corresponds to β-farnesene, the one at 12.20 minutes corresponds to amorpha-4,11-diene (**34**) and 13.14 minutes corresponds to α-bisabolene (**65**) as compared with authentic standards. The NIST library identifies the peak at 12.78 minutes as γ-bisabolene, and 15.08 minutes as α-bisabolol (**72**).*

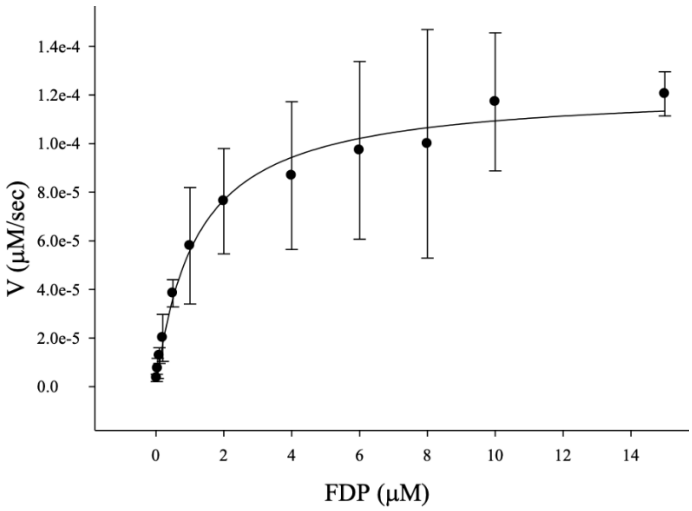


*Figure 68: Michaelis-Menten plot for the conversion of FDP (**40**) by L405A ADS to sesquiterpene products*

Alanine and leucine are both hydrophobic amino acids, but alanine has much smaller side chain, as shown in Figure 69. Reduction in the hydrophobic side chain of L405 affects the initial cyclisation as well the rearrangement of the bisabolyl carbocation which undergoes a 1-10 ring closure to form the carbocation required for amorpha-4,11-diene (**34**) formation. Ultimately this resulted in the majority of products deriving from either the farnesyl or bisabolyl skeleton.

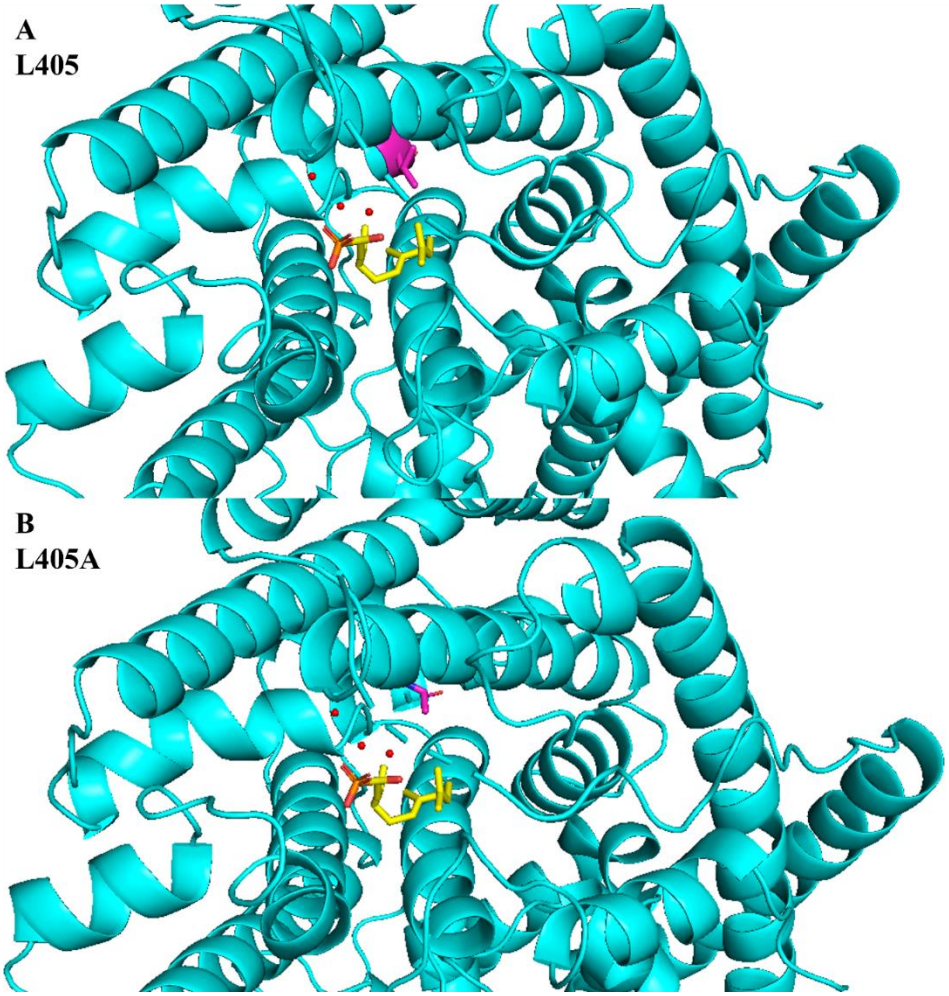


Figure 69: Cartoon representation of homology model of ADS (cyan), FHP (yellow), and magnesium ions (green): A) the L405 residue chosen for mutagenesis (magenta). B) residue 405 mutated to alanine (magenta).

D299E

Residue 299 is the first aspartic acid in the **D**DXXD motif that contributes to binding of magnesium ions. This mutant has a significant reduction in the catalytic efficiency (14-fold slower) compared to the value measured for wild type ADS, see Figure 70.

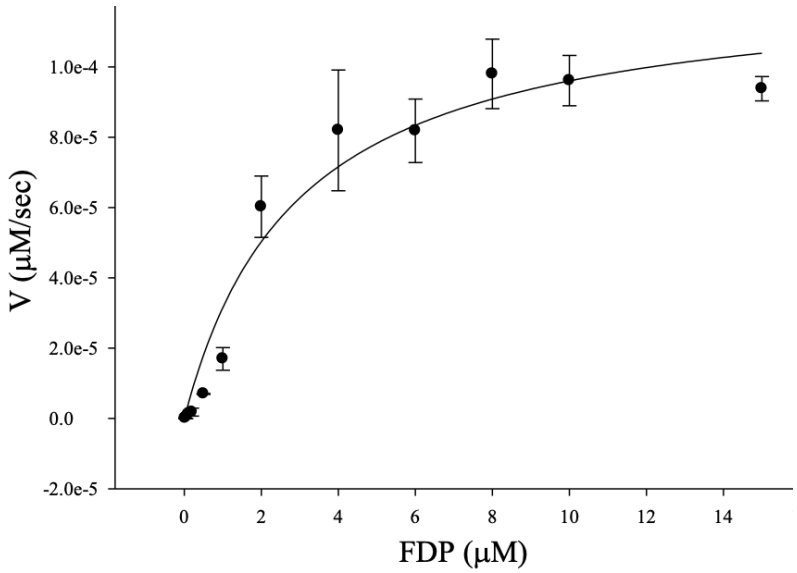


*Figure 70: Michaelis-Menten plot for the conversion of FDP (**40**) by D299E ADS to sesquiterpene product*

While both glutamic acid and aspartic acid are negatively charged, both have different orientations within a protein backbone, as shown in Figure 71. This may be why a different cyclisation mechanism is possible to produce germacrene B (**71**), and the variation in orientation may affect magnesium binding leading to the drop in catalytic activity. When a D299E mutant was tested, as with mutations in residue G401, selina-4(15),7(11)-diene (**101**) and germacrene B (**71**) were determined to be produced. Unlike the mutations at G401, there remained some production of amorpha-4,11-diene (**34**) as well as traces of β-farnesene. The ratio of products was around 52% selina-4(15),7(11)-diene (**101**), 37% germacrene B (**71**) and 11% amorpha-4,11-diene (**34**), as shown in Figure 72. This result suggests that changes in the active site pocket of ADS affects the initial cyclisation pattern (1-6 ring closure) and favour 1-10 cyclisation representing 90% of total products.

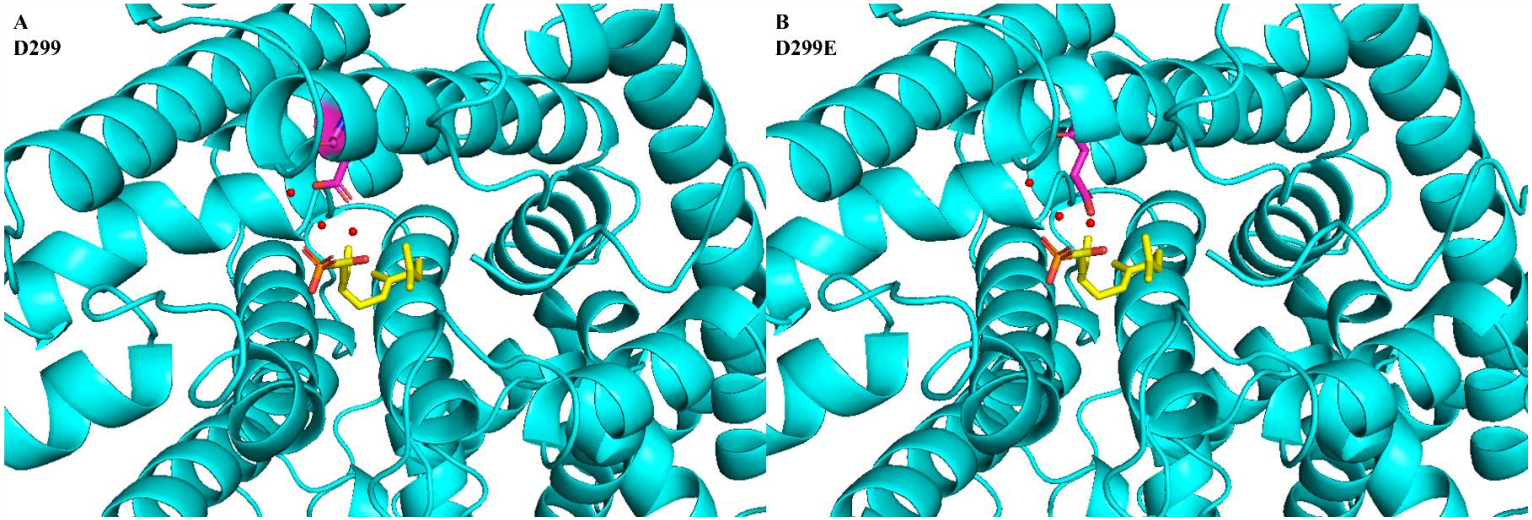


Figure 71: Cartoon representation of homology model of ADS (cyan), FHP (yellow), and magnesium ions (green): A) the D299 residue chosen to mutate (magenta). B) residue 299 mutated to glutamic acid (magenta).

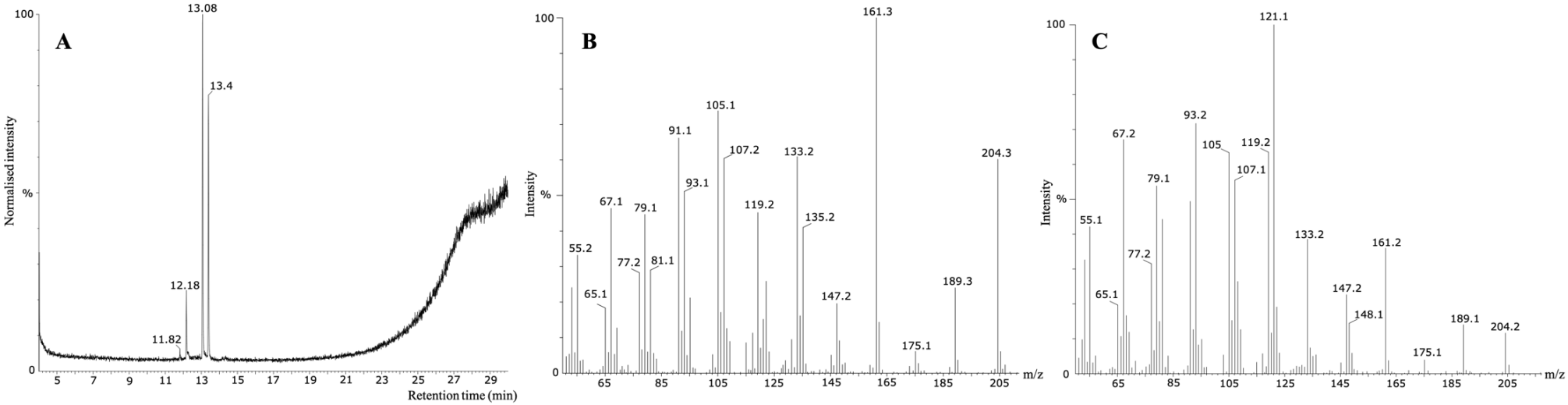


*Figure 72: A) TIC of the pentane extracts arising from the incubations of D299E (1 μM) with FDP (**40**) (0.4 mM) and magnesium chloride (5 mM). 12.18 minutes corresponds to amorpha-4,11-diene (**34**), 13.08 minutes corresponds to selina-4(15),7(11)-diene (**101**) and 13.40 minutes corresponds to Germacrene B (**71**). EI mass spectrum of retention time B) 13.08 minutes. C) 13.40 minutes*

2.4 Conclusion

Described in this chapter are site directed mutagenesis experiments which were used to investigate the influence of amino acid residues surrounding the active site of ADS. The majority of mutations resulted in no change in the product profile, however, they showed significant change to the activity indicating that many of the active sites residues work together in substrate binding and carbocation stabilisation to catalyse the reaction. Several residues in the G helix showed significantly decreased or abnormal activity, reinforcing the proposal that this region is particularly important for enzyme activity.

Some key residues in the G helix and metal binding region showed clear differences in product formation. The metal binding region of sesquiterpene synthases have previously been shown to affect product profile. Mutations of aspartate residues to glutamate in this region are well known to lead to an increase in side products.[9,10] In this work the mutant D299E was made with the expectation of a change in product profile. A larger amount of the linear side product, β-farnesene, was seen, along with low level of the wild type product amorpha-4,11-diene (**34**), but there was also production of two new compounds not normally seen in the ADS product profile; selina-4(15),7(11)-diene (**101**) and germacrene B (**71**). As observed previously, mutating the metal binding motif of sesquiterpene synthases, the D299E mutant showed a very large drop in catalytic efficiency.[9] This may be because the position of the magnesium ions in the active site were altered weakening the binding.

Mutations in the G helix are also known to change enzyme product profiles.[11–13] The L405 residue lies in the G helix and likely aids in folding the substrate. When this residue was mutated to phenylalanine, a larger amino acid, no change in product profile was seen, although leucine is generally conserved in this position, see Figure 57. But when mutated to alanine, a smaller residue, a clear increase in the proportion of bisabolyl cation derived products was seen. Clearly there is some flexibility in the residue that can be at this position, and perhaps the planar phenylalanine can rotate into a position where it is less hindering. Although this mutant showed a decrease in catalytic efficiency, it was not as large as that observed for D299E.

The most interesting results were obtained from the mutation of G401. This residue is highly conserved across sesquiterpene synthases (Figure 57), and previous work showed a link to 1,10 ring closure.[19] Previous work within the Allemann group has showed that when the analogue

of FDP, 8-methoxy FDP, is given to ADS 8-methoxy-γ-humulene is formed. This compound forms from a 1,11-ring closure, showing that it is possible to manipulate the catalytic mechanism of ADS.[32] When mutated to larger residues, leucine and tyrosine, selina-4(15),7(11)-diene (**101**) and germacrene B (**71**) were obtained as products. Neither of these sesquiterpenes have been identified as side products of ADS, therefore these mutants show novel activity. Furthermore, although they cause very dramatic shifts to the product profile, G401Y shows comparable catalytic efficiency to the wild type enzyme and G401L does not dramatically differ.

While the active site of ADS proved robust to mutation in general, single mutations in particular motifs created dramatic changes in product profile and produced novel activity suggesting potential promiscuity of the enzyme.

1 P. Alam, U. Kiran, M. Mobeen Ahmad, M. Ali Khan, S. Jhanwar and M. Z. Abdin, *Bioinformation*, 2010, **4**, 421–429.
2 H. J. Bouwmeester, T. E. Wallaart, M. H. A. Janssen, B. van Loo, B. J. M. Jansen, M. A. Posthumus, C. O. Schmidt, J. W. de Kraker, W. A. König and M. C. R. Franssen, *Phytochemistry*, 1999, **52**, 843–854.
3 I. I. Abdallah, M. Czepnik, R. van Merkerk and W. J. Quax, *Journal of Natural Products*, 2016, **79**, 2455–2463.
4 R. Li, W. K. W. Chou, J. A. Himmelberger, K. M. Litwin, G. G. Harris, D. E. Cane and D. W. Christianson, *Biochemistry*, 2014, **53**, 1155–1168.
5 I. I. Abdallah, R. van Merkerk, E. Klumpenaar and W. J. Quax, *Scientific Reports*, 2018, **8**, 1–11.
6 J. E. Diaz, C. S. Lin, K. Kunishiro, B. K. Feld, S. K. Avrantinis, J. Bronson, J. Greaves, J. G. Saven and G. A. Weiss, *Protein Science*, 2011, **20**, 1597–1606.
7 B. T. Greenhagen, P. E. O'Maille, J. P. Noel and J. Chappell, *Proc Natl Acad Sci U S A*, 2006, **103**, 9826–9831.
8 Y. Yoshikuni, T. E. Ferrin and J. D. Keasling, *Nature*, 2006, **440**, 1078–1082.
9 B. Felicetti and D. E. Cane, *J Am Chem Soc*, 2004, **126**, 7212–7221.
10 M. Seemann, G. Zhai, J. W. de Kraker, C. M. Paschall, D. W. Christianson and D. E. Cane, *J Am Chem Soc*, 2002, **124**, 7681–7689.
11 P. Baer, P. Rabe, K. Fischer, C. A. Citron, T. A. Klapschinski, M. Groll and J. S. Dickschat, *Angewandte Chemie - International Edition*, 2014, **53**, 7652–7656.
12 M. Loizzi, V. González, D. J. Miller and R. K. Allemann, *ChemBioChem*, 2018, **19**, 100–105.
13 Y. Yoshikuni, V. J. J. Martin, T. E. Ferrin and J. D. Keasling, *Chemistry and Biology*, 2006, **13**, 91–98.
14 P. L. Srivastava, A. M. Escorcia, F. Huynh, D. J. Miller, R. K. Allemann and M. W. van der Kamp, *ACS Catalysis*, 2021, **11**, 1033–1041.
15 S. C. Kampranis, D. Ioannidis, A. Purvis, W. Mahrez, E. Ninga, N. A. Katerelos, S. Anssour, J. M. Dunwell, J. Degenhardt, A. M. Makris, P. W. Goodenough and C. B. Johnsona, *Plant Cell*, 2007, **19**, 1994–2005.
16 M. Xu, P. R. Wilderman and R. J. Peters, *Proc Natl Acad Sci U S A*, 2007, **104**, 7387–7401.
17 J.-X. Li, X. Fang, Q. Zhao, J.-X. Ruan, C.-Q. Yang, L.-J. Wang, D. J. Miller, J. A. Faraldos, R. K. Allemann, X.-Y. Chen and P. Zhang, *Biochemical Journal*, 2013, **451**, 417–426.
18 Z. Li, R. Gao, Q. Hao, H. Zhao, L. Cheng, F. He, L. Liu, X. Liu, W. K. W. Chou, H. Zhu and D. E. Cane, *Biochemistry*, 2016, **55**, 6599–6604.
19 X. Fang, J.-X. Li, J.-Q. Huang, Y.-L. Xiao, P. Zhang and X.-Y. Chen, *Biochemical Journal*, 2017, **474**, 2191–2202.
20 M. Demiray, PhD, Cardiff University , 2016.
21 A. di Girolamo, J. Durairaj, A. van Houwelingen, F. Verstappen, D. Bosch, K. Cankar, H. Bouwmeester, D. de Ridder, A. D. J. van Dijk and J. Beekwilder, *Archives of Biochemistry and Biophysics*, 2020, **695**, 108647.
22 P. S. Karunanithi and P. Zerbe, *Frontiers in Plant Science*, 2019, **10**, 1–23.
23 P. Mercke, M. Bengtsson, H. J. Bouwmeester, M. A. Posthumus and P. E. Brodelius, *Archives of Biochemistry and Biophysics*, , DOI:10.1006/abbi.2000.1962.
24 J. Yang, R. Yan, A. Roy, D. Xu, J. Poisson and Y. Zhang, *Nature Methods*, 2014, **12**, 7–8.
25 A. Roy, A. Kucukural and Y. Zhang, *Nature Protocols*, 2010, **5**, 725–738.
26 Y. Zhang, *BMC Bioinformatics*, 2008, **9**, 1–8.
27 C. M. Starks, K. Back, J. Chappell and J. P. Noel, *Science (1979)*, 1997, **277**, 1815–1820.
28 H. Liu and J. H. Naismith, *BMC Biotechnol*, 2008, **8**, 1–10.

29 Y. H. Wang, H. Xu, J. Zou, X. B. Chen, Y. Q. Zhuang, W. L. Liu, E. Celik, G. D. Chen, D. Hu, H. Gao, R. Wu, P. H. Sun and J. S. Dickschat, *Nature Catalysis*, 2022, **5**, 128–135.
30 A. Vattekkatte, S. Garms, W. Brandt and W. Boland, *Organic and Biomolecular Chemistry*, 2018, **16**, 348–362.
31 D. W. Christianson, *Chemical Reviews*, 2017, **117**, 11570–11648.
32 M. Demiray, D. J. Miller and R. K. Allemann, *Beilstein Journal of Organic Chemistry*, 2019, **15**, 2184–2190.

3 Chapter 3 - Utilising 4-hydroxy Prenol (106) to synthesise 12-hydroxy FDP (105)

3.1 Introduction

An economically viable route for artemisinin (**4**) synthesis is needed for the treatment of the global epidemic, malaria, a parasitic disease which kills around three million people a year.[1] Previous work shows that the hydroxylated form of FDP (**40**) can be used in a shortened enzymatic synthesis pathway, comparative to the naturally occurring pathway in *Artemisia annua*, to obtain the medically relevant compound dihydroartemisinic aldehyde (DHAAl, **82**), a precursor in the synthesis of artemisinin (**4**). Although this chemoenzymatic synthesis is optimised comparative to the pathway that occurs in nature, the chemical synthesis of 12-hydroxy FDP (12-OH FDP, **105**) reported is still time consuming and inefficient.[2] It also suffers many practical issues such as isolation of the diphosphate reagent in an anhydrous condition and keeping this in the usable dry form. As the diphosphate reagent is generally used in excess, it can also co-elute with the product in purification.[3] In the synthesis of 12-OH FDP (**105**) there are two chemical synthesis steps that dramatically reduce the yield: the hydroxylation which has a reported yield of 49% and the diphosphorylation of 12-hydroxy farnesyl chloride (**104**), which has a reported yield of 52%, giving an overall yield of ~25%, as shown in Figure 73.[2] This chapter focusses on the investigation of ways to improve these steps using enzymatic alternatives.

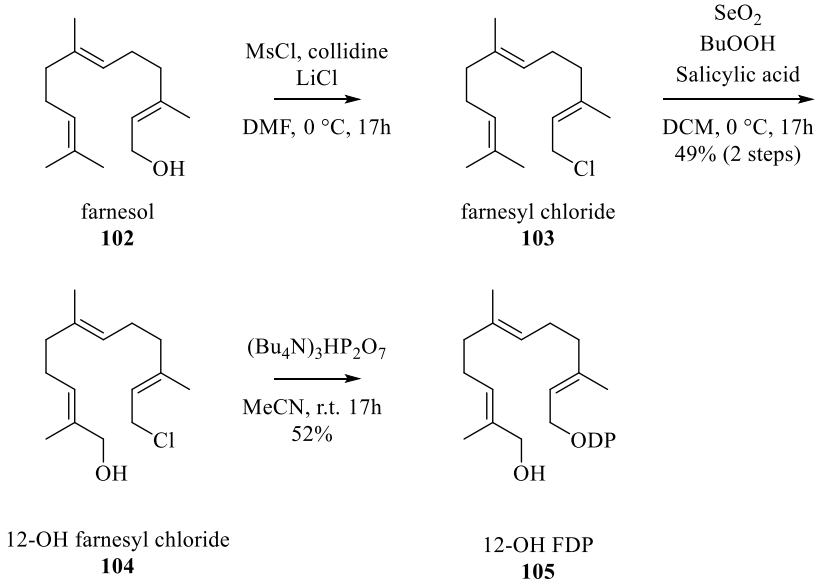


Figure 73: Summary of chemical synthesis of 12-OH FDP[2]

3.2 Aims

Previous work in the group had used two kinases, hydroxyethylthiazole kinase (ThiM) from *Escherichia coli*, and isopentenyl phosphate kinase (IPK) from *Methanocaldococcus jannaschii*, with prenol and isoprenol, to produce IDP (**14**) and DMADP (**13**). IDP (**14**) and DMADP (**13**) were combined by FDP synthase (FDPS) to produce FDP (**40**), which was turned over by amorpha-4,11-diene synthase (ADS) to produce amorpha-4,11-diene (**34**).[4] As these kinases showed some substrate diversity, they were tested with 4-hydroxyprenol (**106**) (a 5-carbon terpene precursor of DHAAl (**82**)). The resulting 4-hydroxy dimethylallyl diphosphate (4-OH DMADP) (**108**), along with IDP and FDPS, could be used to produce 12-OH FDP (**105**), which is known to be converted to DHAAl (**82**) by ADS,[4] this pathway is shown in Figure 74 along with the natural synthesis in *A. annua* for comparison. If this synthesis proved successful, it was hypothesised it would be more efficient, begin from less expensive starting materials and be less laborious than the chemical synthesis of 12-OH FDP (**105**) to generate DHAAl (**82**)

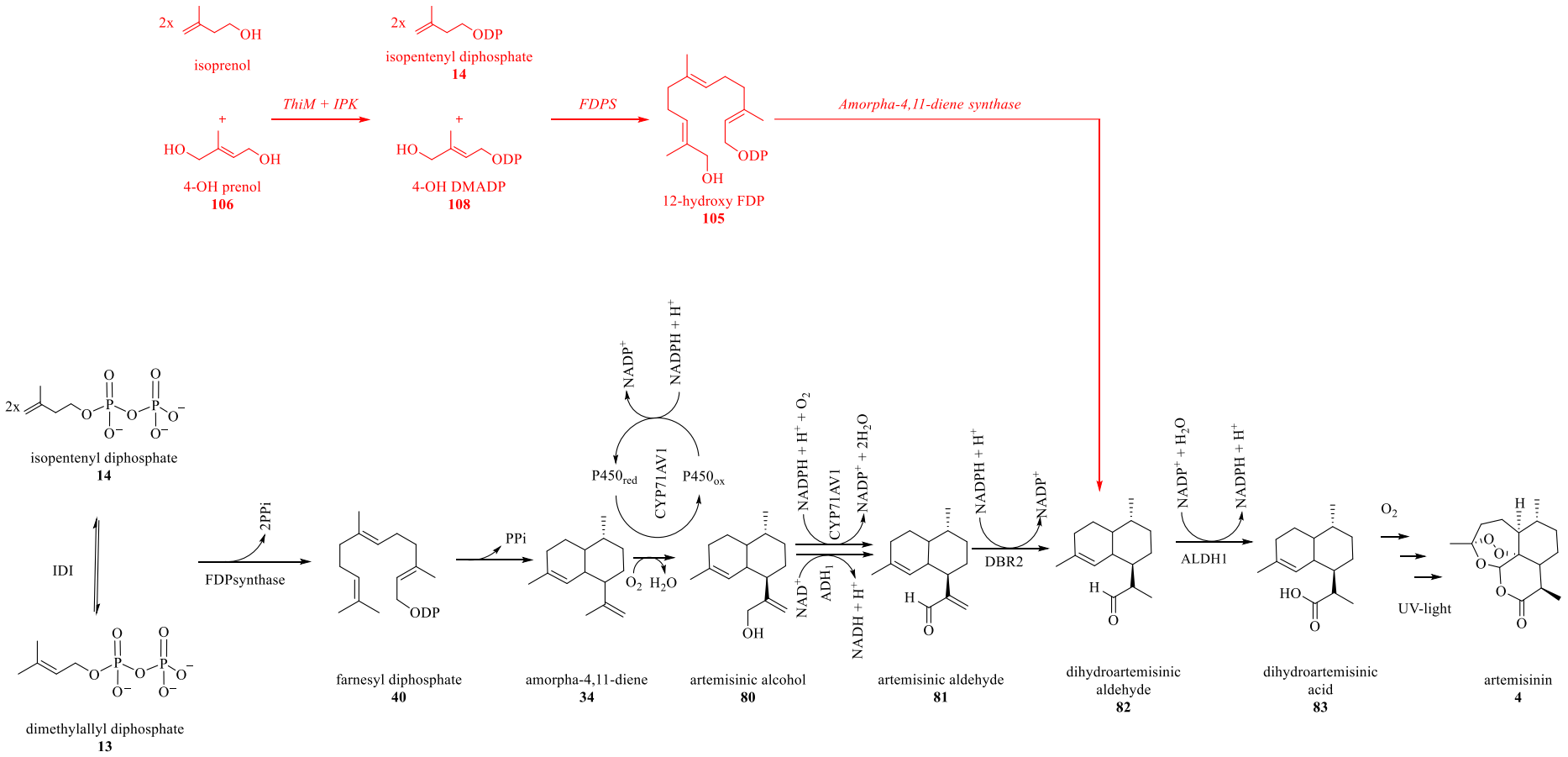
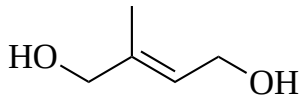


Figure 74: Natural synthesis pathway of artemisinin in Artemisia annua *in black, and the artificial shortened pathway to DHAAl (**82**) shown in red.* [5]

3.3 Results

3.3.1 Chemical synthesis of 4-hydroxyprenol

Briefly, commercially sourced prenol was hydroxylated *via* a selenium dioxide oxidation before purification *via* column chromatography, to obtain 4-hydroxy prenol (**106**), Figure 75. See materials and methods 6.2.5 for further detail. The product was confirmed by ^{1}H and ^{13}C NMR spectrum; and mass spectrometry, see 7.2.1 of supplementary information.



4-hydroxy prenol
106

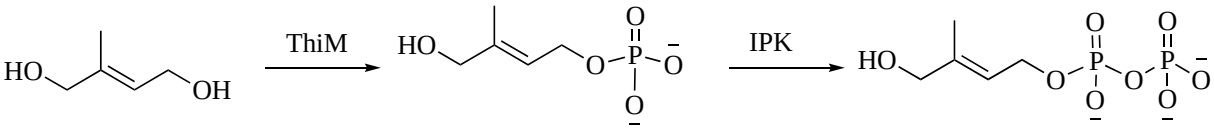
*Figure 75: Chemical structure of 4-hydroxy prenol (**106**)*

3.3.2 Diphosphorylation of 4- hydroxyprenol (**106**)

Previous work carried out in the group showed that the kinases ThiM and IPK could convert alcohols into diphosphates,[4] these strategies were used in experiments to see if 4-OH DMADP (**108**) could be produced from 4-hydroxyprenol (**106**). The chemically synthesised 4-hydroxyprenol (**106**) was added to a solution of D_2O, KCl (2.5 M), $MgCl_2$ (500 mM), ATP (100 mM), pyruvate kinase, phosphoenolpyruvate (500 mM) and ThiM (68 μM) in a buffer of 100 mM TRIS pH 8.0. ATP was added as a source of phosphate; pyruvate kinase and phosphoenolpyruvate were added as an ATP regenerating system. See 6.1.36 of materials and methods for further detail. Initially, the reaction was left incubating overnight and a ^{31}P NMR spectrum of the reaction was performed to ensure the monophosphorylation reaction had occurred. It was observed that 48-hour incubation was necessary to obtain significant monophosphorylation. IPK was then added to the reaction, and the same process repeated, leaving the reaction overnight only. Once the system was confirmed to work, both ThiM and IPK were added in one pot and left for 48h, producing 4-hydroxy dimethylallyl diphosphate, Figure 76.

Mono and diphosphorylation of 4-hydroxyprenol (**106**) was monitored *via* $^{31}PNMR$ spectroscopy. The monophosphorylated product was identified by comparison with a reaction with no enzyme, and the diphosphorylated product by comparison with the monophosphorylated product. See 7.2.3 of supplementary information for NMR spectra. Both

steps were also compared to control reactions completed with prenol as a substrate. ThiM accepted 4-hydroxyprenol (**106**) as a substrate, and IPK accepted 4-hydroxy dimethylallyl monophosphate (4-OH DMAMP) (**107**). It was clear from comparison with the NMR spectra of prenol subjected to the same experiments that 4-hydroxyprenol (**106**) is a poorer substrate for these kinases. The increase in the peaks generated by the mono (2.5 ppm) and di-phosphorylated species (-6.3, -9.5 ppm) were far smaller in the 4-hydroxyprenol (**106**) spectrum than in the prenol spectrum. See 7.2.3 of supplementary information for NMR spectra. Work to optimise the monophosphorylation step was ongoing within the group therefore the diphosphorylation step was chosen for optimisation. IPK has previously been mutated to improve its substrate scope for longer substrates.[6,7] It is possible that ThiM and IPK could be mutated to take a hydroxylated analogue, such as 4-hydroxyprenol (**106**), over more favourable substrates such as prenol.



4-hydroxy prenol
106

4-hydroxy dimethyl allyl monophosphate
107

4-hydroxy dimethyl allyl diphosphate
108

Figure 76: Enzymatic synthesis of hydroxylated DMADP from hydroxy prenol

3.3.3 Mutations of IPK

While carrying out phosphorylation reactions with ThiM and IPK, it was seen in the ^{31}P NMR spectroscopic analysis that IPK was far slower at diphosphorylating the 4-OH DMAMP (**107**) than DMAMP when compared after four hour incubation. The reduction in the peak representing the monophosphorylated starting material, at 3.5 ppm, see 7.2.3 of supplementary information for NMR spectra, was far slower for 4-OH DMAMP (**107**) than for DMAMP. The peaks for ATP (-3.5, -6 and -10.5 ppm) and PEP (-1 ppm), both needed for the diphosphorylation reaction were also still present in the NMR spectrum of 4-OH DMAMP (**107**) suggesting the reaction had not reached completion. Once left overnight both reactions had similar spectra. Therefore, changes in the amino acid sequence were designed to try and improve IPKs affinity for 4-OH DMAMP (**107**).

Some work has been described in the literature on investigating the substrate scope of IPK, generally involving the use of wild type IPK with analogues of DMADP (**13**) and IDP (**14**).[8,4] Work has been done on generating variants of IPK, both in the Allemann group,[6] and described

in the literature,[7] but both were designed for IPK to take a larger substrate. A crystal structure of IPK (PDB: 3K4Y)[7] was assessed, and residues believed to be near the hydroxyl group of 4-OH DMAMP (**107**) were chosen for mutations. G59 was mutated to threonine allowing hydrogen bonding with the hydroxyl group of 4-OH DMAMP (**107**) as was I156. M90, G144, and D145 were all mutated to serine for the same reason. The wild type residues and the mutated residues in the active site are shown in Figure 77, with individual changes shown in 7.2.4 of supplementary information.

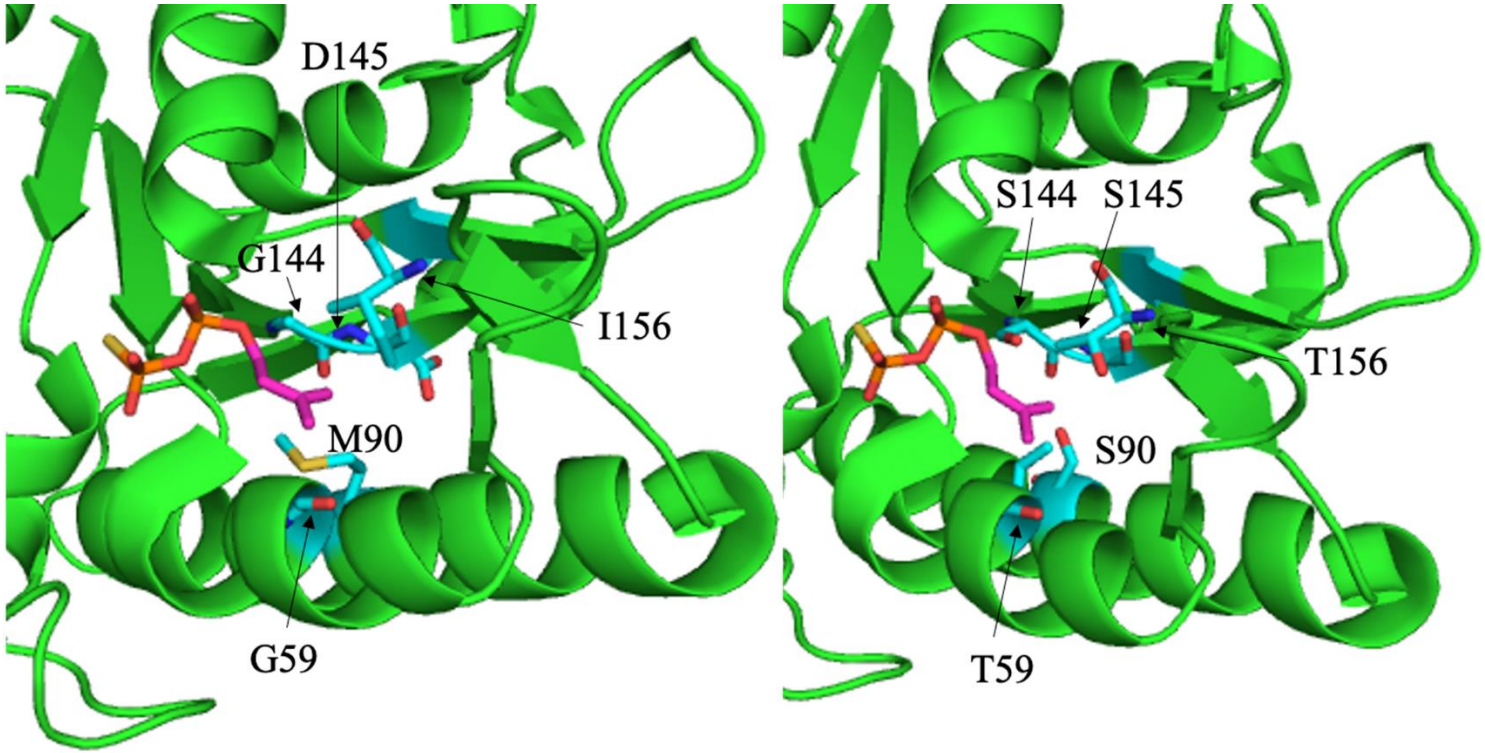


Figure 77: Cartoon representations of a crystal structure of IPK in green (PDB:3K4Y)[7], with the key mutated residues highlighted in cyan, IPP is shown in magenta with the diphosphate group in red and orange

The variant enzymes were obtained, the enzyme was tested with 4-OH DMAMP (**107**) and monitored via ^{31}P NMR spectrum analysis for the presence of diphosphate peaks with reaction conditions as described previously. 4-OH DMAMP (**107**) was obtained as described above using ThiM. The reactions mixtures were then filtered to remove the ThiM enzyme, and the remaining components of the reaction were freeze dried. The powder obtained, which contained 4-OH DMAMP (**107**) was then used in the next step of the reaction, using IPK or a variant. The reaction also contained the internal standard methylphosphonic acid.

Product formation was monitored over 48 hours by ^{31}P NMR spectrum analysis every 4 hours, Figure 78, for comparison with wild type IPK, as shown in Table 4. The results are represented as a comparison to wild type. Monitoring the reaction at regular time intervals ensured that the reaction had reached completion; and that the product did not begin to precipitate once it had reached a certain concentration, leading to a drop in perceived product formation. This also allowed monitoring of the stability of the variants. If variants were less stable than WT IPK they could have denatured within a short space of time, but the initial rate may have been faster than WT IPK. This was not seen, product formation for all variants along with the WT IPK proceeded in a linear fashion, and was complete by 36 hours. Therefore, 36 hours was the time point chosen for comparison.

Table 4:Each variant was analysed by ^{31}P NMR spectroscopy after 36 hours and the speed of the reaction was compared to the standard, wild type IPK. Speed is reported as a fold change compared to wild type IPK.

IPK variant	Speed of reaction compared to WT IPK
WT	1
G59T	0.53
M90S	0.56
G144S	0.53
D145S	0.77
I156T	0.67

No variant of IPK showed improved activity compared to the wild type IPK. It may be that the residues changed play a key part in the enzymes catalytic activity and therefore changing the residues to accommodate the hydroxyl group had a negative effect on the rate of catalysis.

It was anticipated that ThiM and IPK could phosphorylate either hydroxyl group of 4-OH prenol (**106**), and therefore not all phosphorylated product would be usable. Although this was possible, if 4-OH prenol (**106**) was diphosphorylated at the 1 position, it would not be converted by FDPS and ADS downstream to produce DHAAl (**82**). Successful turnover of the product formed from incubation of 4-OH prenol (**106**), ThiM and IPK; with FDPS and ADS indicated that diphosphorylation had occurred, to at least some extent, at the 4 position. The product could have been identified by purification and NMR, but HPLC purification of the product proved extremely difficult and time consuming, therefore turnover by FDPS and ADS to DHAAl (**82**) was deemed sufficient confirmation.

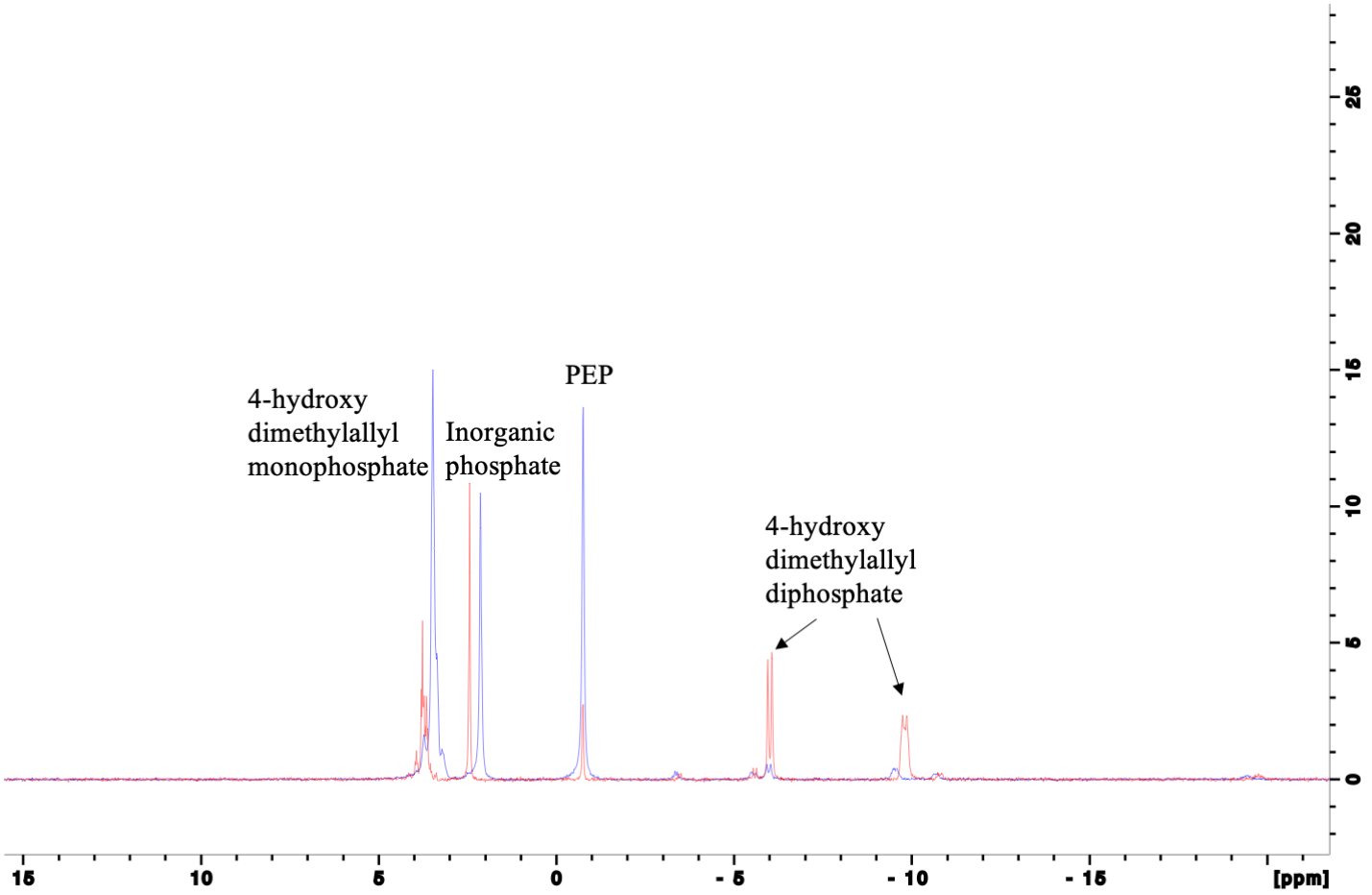


*Figure 78: ^{31}P NMR spectra (243 MHz, D_2O, 298K) of the diphosphorylation reaction of 4-OH DMAMP (**107**) (peak at 3.5 ppm) catalysed by IPK, showing the decrease in 4-OH DMAMP (**107**) and increase in 4-OH DMADP (**108**) (peaks at -6, -10 ppm) over time. The peak at 2.5 ppm is inorganic phosphate, and at -0.5 ppm is PEP. The blue trace shows the time point where IPK was added, and the red trace shows the reaction at 36 hours.*

3.3.4 Enzymatic synthesis of 12-OH FDP (**105**) from 4-OH DMADP (**108**) and IDP (**14**)

Once 4-OH DMADP (**108**) was obtained it was coupled in a reaction with enzymatically produced IDP (**14**) and FDPS. Both were added to the reaction described in 3.3.2 and incubated for 24 hours. 12-OH FDP (**105**), like FDP (**40**) precipitates from solution in the reaction and therefore progression of the reaction could be monitored by eye. The (*E,E*)-FDPS used in these reactions is the wild type form from *Bacillus stearothermophilus*. Previous work has shown that this FDPS accepts 8-hydroxy GDP and couples it with IDP (**14**) to produce 12-OH FDP (**105**), but with an almost tenfold reduction in product comparative to the wild type reaction to produce (*E,E*)-FDP.[9,10] Therefore it was very likely that 4-OH DMADP (**108**) would be turned over with IDP (**14**) to produce 12-OH FDP (**105**).

3.3.5 Activity experiments

To confirm 12-OH FDP (**105**) was produced, the final step using ADS to convert it to DHAAl (**82**) was completed, Figure 79. ADS was added to the reaction containing biologically synthesised 12-OH FDP (**105**), the reaction was overlaid with pentane and left to incubate overnight, see 6.1.36 of materials and methods. The pentane extraction was analysed by GC-MS, and the retention time and fragmentation pattern compared to that of a known standard of DHAAl (**82**) and amorpha-4,11-diene (**34**). Comparable amounts of DHAAl (**82**) and amorpha-4,11-diene (**34**) were seen on the GC-MS, Figure 80. Hence, the chemical diphosphorylation step of 12-OH FDP (**105**) synthesis can be replaced by biological phosphorylation using kinases beginning from a 5-carbon precursor of DHAAl (**82**). The only synthetic chemical reaction required still for this pathway is the initial oxidation.

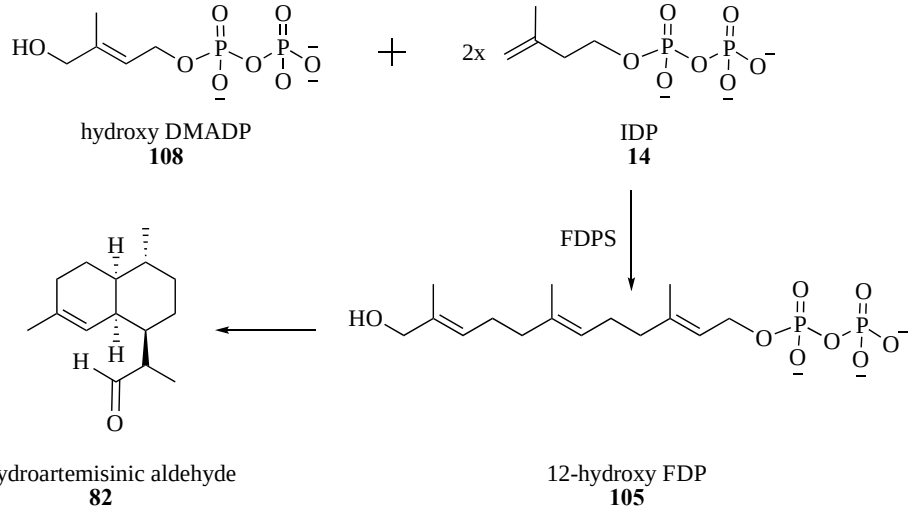


*Figure 79: Enzymatic synthesis of DHAAl (**82**) from 4-OH DMADP (**108**) and IDP (**14**)*

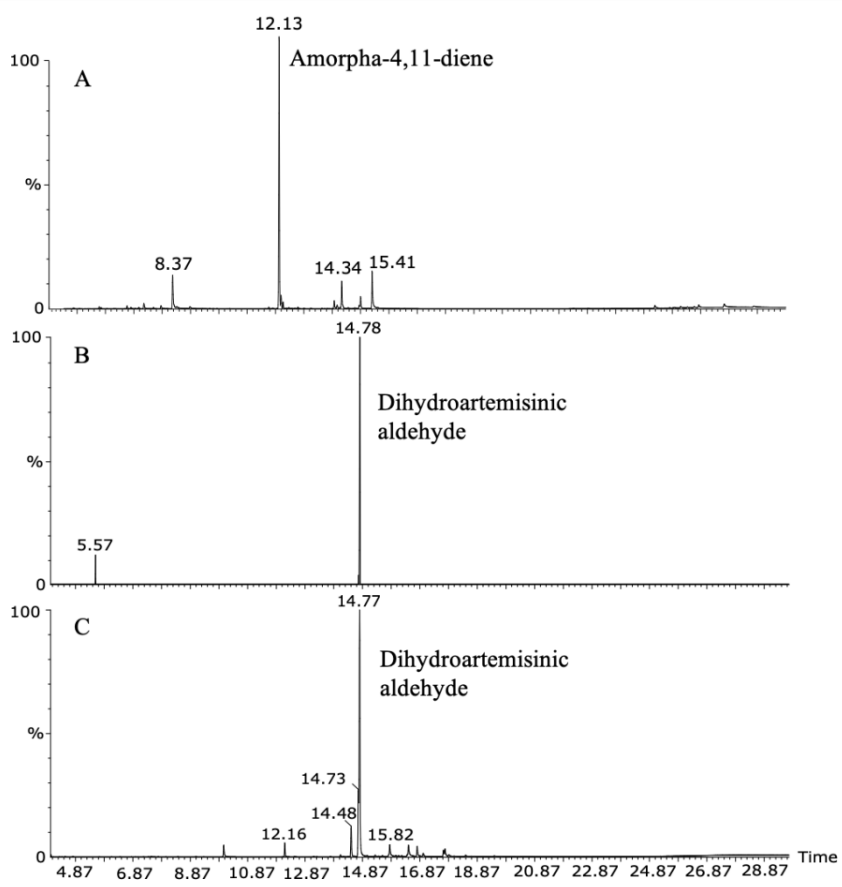


*Figure 80: TIC of the pentane extracts of reactions with A: prenol (2 mg/mL), IDP (**14**) (2.9 mg/mL), ThiM (68 μM), IPK (0.15 μM), FDPS (10 μM), and ADS (100 nM); C: 4-hydroxyprenol (**106**) (2 mg/mL), IDP (**14**) (2.9 mg/mL), ThiM (68 μM), IPK (0.15 μM), FDPS (10 μM), and ADS (100 nM). Reactants were incubated at room temperature for 48h. In panel B 14.78 minutes is a standard of S-DHAAl (**82**), with a small amount of* R-*DHAAl (**82**) present as a shoulder on the left-hand side (obtained from Dr. F. Huynh). In panel C amorpha-4,11-diene (**34**) (12.16 minutes),* S-*DHAAl (**82**) (14.77 minutes),* R-*DHAAl (**82**) (14.73 minutes), and artemisinic-11S,12-epoxide (14.48 minutes) are also present. Amorpha-4,11-diene (**34**) is present as IDP (**14**) can isomerise to DMADP (**13**) over time. No 4-hydroxprenol (**106**) (4.79 minutes) is present.*

Although 12-OH FDP (**105**) was produced as seen before in the literature, and as confirmed by the production of DHAAl (**82**), this was not a favourable substrate. FDPS has been mutated to allow the synthesis of alternative products, e.g. hydroxy-GGDP.[9] The crystal structure of FDPS has been deposited (PDB: 5AYP),[11] which was used to design mutations of this enzyme to increase its affinity for 4-OH DMADP (**108**).

3.3.6 Variants of FDPS

The crystal structure of FDPS was assessed to determine which residues would be near the 12-C end of FDP (**40**) and could therefore be changed to allow better affinity for the

analogue, 12-OH FDP (**105**), and to introduce more space. Methionine 156 was hypothesised to be near where the hydroxyl group of 12-OH FDP (**105**) would sit within the active site. This residue was therefore mutated to serine (M156S), threonine (M156T) and aspartate (M156D) to aid in hydrogen bonding. These residues vary in size, including a mutation that increases the space in the active site for the additional hydroxyl group, to a mutation that is similar in size to the original residue to not dramatically change the enzymes structure, see Figure 81.

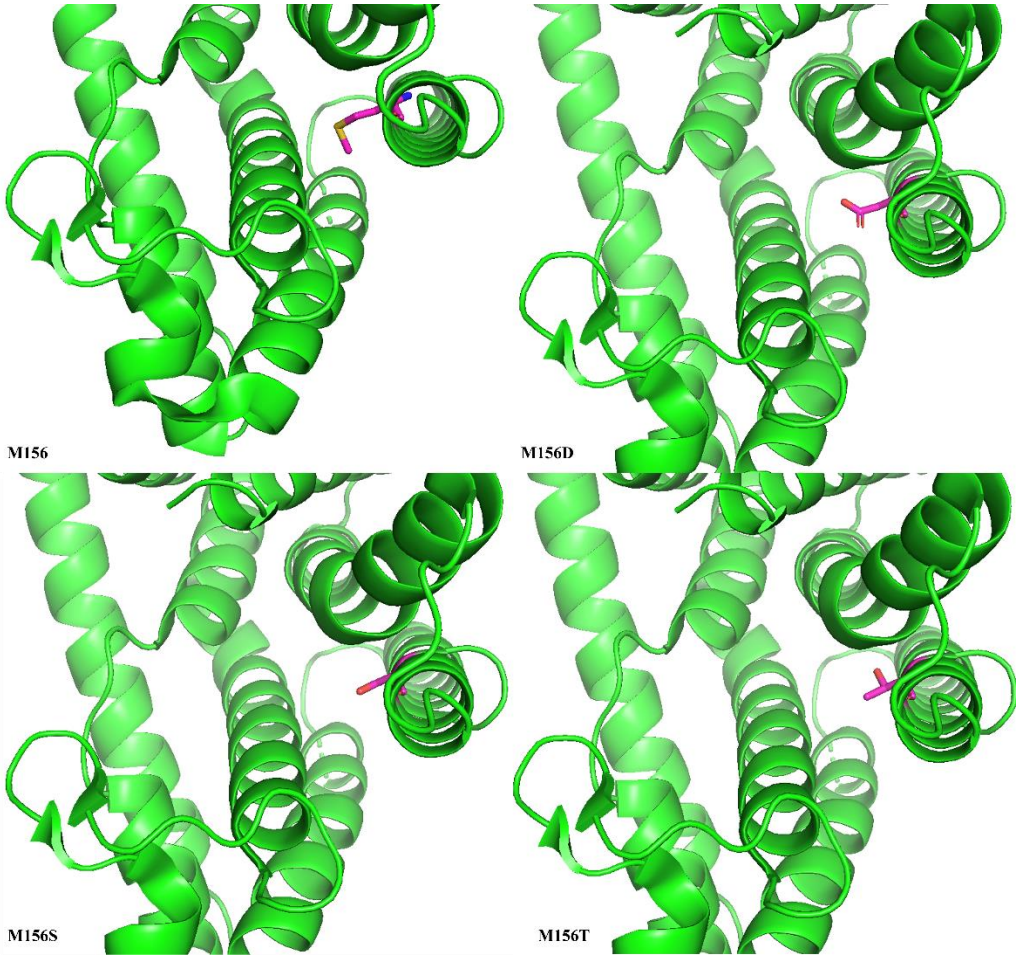


Figure 81: Cartoon representations of a crystal structure of FDPS (PDB: 5AYP)[11] in green, with the key mutated residue highlighted in magenta.

4-OH DMADP (**108**) and IDP (**14)** were added to a buffer containing DTT, HEPES, and $MgCl_2$. FDPS or a variant were added (1 μM) and ADS added (20 μM) to ensure ADS was present in an excess and FDPS would be the limiting factor. The reactions were incubated for four hours, before alkaline phosphatase was added. The reaction was then extracted with pentane which contained geraniol as an internal standard. Any 4-OH DMADP (**108**) or 12-OH FDP (**105**) not converted to DHAAl (**82**) would then be present as an alcohol and would be

extracted into the organic layer along with the DHAAl (**82**). Reactions with FDPS, ADS, IDP (**14**) or 4-OH DMADP (**108**) omitted were performed to ensure no DHAAl (**82**) was produced from another source. See materials and methods 6.1.36.3 for further details.

*Table 5: DHAAl (**82**) produced by each variant of FDPS in a reaction with ADS, was compared to the DHAAl (**82**) produced in a reaction containing WT FDPS and ADS, and reported as a fold increase of product as normalised to the geraniol internal standard. .*

FDPS variant	Product generated compared to WT (fold increase)
M156D	13.8
M156S	3.2
M156T	1.0

The variant of FDPS, M156D, produced almost 14 times more DHAAl (**82**) compared to WT FDPS in a four-hour incubation, as shown in Table 5. Aspartate is only a little smaller than methionine, with a volume of 111.1 $Å^3$ compared to 162.9 $Å^3$ respectivley,[12] and contains an oxygen available for hydrogen bonding with the hydroxyl group at the 12^{th} position of 12-OH FDP (**105**). Both the M156T and M156S mutants did not show such a dramatic increase in product production, but M156S did shows some improvement in activity compared to the WT FDPS, as seen in Figure 82. It may be that the hydroxyl group of serine is available for hydrogen bonding, but as it is smaller it may not be in the appropriate position. Threonine is smaller than methionine, and thought this residue also contains a hydroxyl group, it may not point into the correct position to allow hydrogen bonding with the hydroxyl group of 12-OH FDP (**105**). In the literature, FDPS has previously been mutated to generate a GDPS[9] showing it is possible to modify this enzymes activity by site directed mutagenesis.

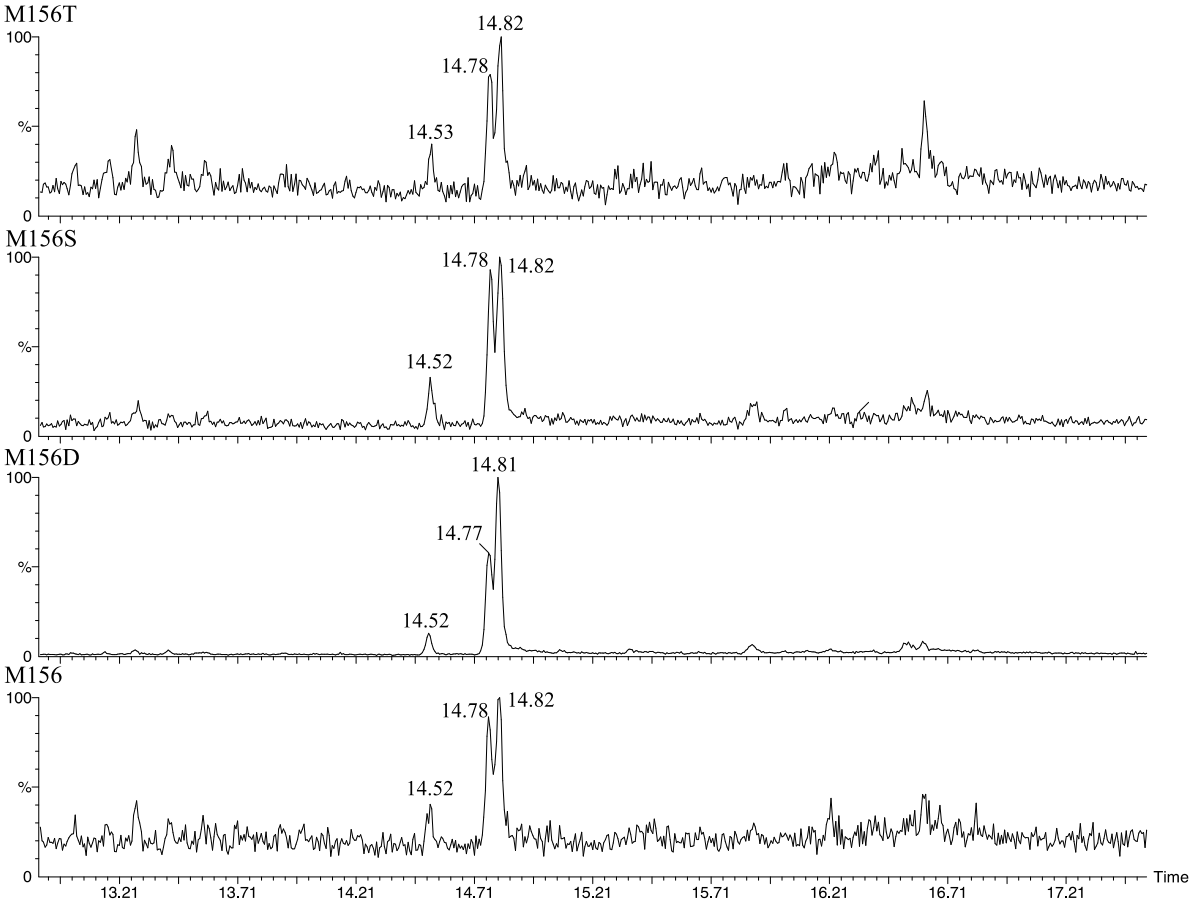


Figure 82: TIC from GC-MS analysis of pentane extractions of reactions with FDPS or a variant of, 4-OH DMADP ***(108)*** *2.9 mg/mL, IDP (**14**) 2.9 mg/mL, ADS (20 μM), FDPS (1 μM), and a buffer containing DTT, HEPES and* $MgCl_2$*. Reactions were left incubating at 37 °C for four hours. The chromatograms show:* S-DHAAl ***(82)*** *(14.82 minutes),* R-DHAAl ***(82)*** *(14.78 minutes), and artemisinic-11S,12-epoxide (14.52 minutes).*

3.4 Conclusion

Two kinases, ThiM and IPK were investigated for improving the diphosphorylation step of 4-hydroxy prenol (**106**), a chemically hydroxylated analogue of prenol. Previous work in the group had shown that prenol and isoprenol could be diphosphorylated by ThiM and IPK and converted to FDP (**40**) using FDPS.[4] Synthesis of 4-hydroxy prenol (**106**) is quick and has similar yield to the oxidation of farnesyl chloride (**103**) but involves only one step of synthesis rather than two. ThiM, IPK and FDPS have been shown to accept the 4-hydroxy analogue, albeit less favourably. In reactions of 4-OH DMADP (**108**), IDP (**14**), FDPS and ADS comparable yields of DHAAl (**82**) were seen as determined by GC-MS with respect to an analogous reaction with DMADP (**13**) leading to the production of amorpha-4,11-diene (**34**) when the reaction was incubated for 48 hours.

When conducting these experiments, it was observed that 4-OH DMAMP (**107**) was a poor substrate for IPK. Therefore, mutations were designed that exchanged wild type residues for those that allowed hydrogen bonding, in the region were the hydroxyl group was postulated to sit. Previous work has shown that IPK is amenable to mutation,[6,7] so it was hypothesised that the changes would not result in loss of activity. Although none of the mutants showed loss of activity, none showed an improved activity with the unnatural substrate, 4-OH DMAMP (**107**).

The FDP synthase used originally was seen to take 4-OH DMADP (**108**) less favourably than DMADP (**13**) and IDP (**14**). It was therefore mutated with wild type residues exchanged for those which allow hydrogen bonding where the hydroxyl group was hypothesised to sit. Three mutants were generated, M156D, M156S and M156T. M156D produced almost 14 times more product than wild type FDPS in a 4-hour incubation period, suggesting this mutation greatly improves the enzymes affinity for 4-OH DMADP (**108**). This change keeps a similarly sized residue but introduces a hydroxyl group for hydrogen bonding with the hydroxyl group of 4-hydroxy DMADP (**108**). M156S showed a little improvement compared to the wild type, but not a dramatic improvement, and M156T showed little change.

This chapter describes the design of a more efficient synthesis of DHAAl (**82**), that involves a far less laborious method for diphosphorylation than previously reported. [2,3] Although attempts to optimise the diphosphorylation step were unsuccessful, a more efficient FDPS for the

condensation of 4-OH DMADP (**108**) and IDP (**14**) was designed. Future work could involve attempting to optimise the monophosphorylation of 4-OH DMADP (**108**) and further attempts to improve the diphosphorylation reaction.

1 CDC, Centers for Disease Control and Prevention, https://www.cdc.gov/malaria/malaria_worldwide/impact.html, (accessed July 20, 2019).
2 X. Tang, M. Demiray, T. Wirth and R. K. Allemann, *Bioorganic and Medicinal Chemistry*, 2018, **26**, 1314–1319.
3 H. J. Korhonen, H. L. Bolt and D. R. W. Hodgson, *Beilstein Journal of Organic Chemistry*, 2015, **11**, 469–472.
4 R. K. Allemann, L. Johnson, A. Dunbabin, J. Benton and R. Mart, *Angewandte Chemie*, 2020, **59**, 8486–8490.
5 J. Han, H. Wang, S. Kanagarajan, M. Hao, A. Lundgren and P. E. E. Brodelius, *Molecular Plant*, 2016, **9**, 946–948.
6 F. Huynh, PhD, Cardiff University, 2020.
7 N. Dellas and J. P. Noel, *ACS Chemical Biology*, 2010, **5**, 589–601.
8 S. Lund, C. Taylor and G. J. Williams, *Chem Bio Chem*, 2019, **20**, 2217–2221.
9 M. Nagaki, M. Nakada, T. Musashi, J. Kawakami, N. Ohya, M. Kurihara, Y. Maki, T. Nishino and T. Koyama, *Bioscience, Biotechnology, and Biochemistry*, 2007, **71**, 1657–1662.
10 M. Nagaki, T. Ichijo, R. Kobashi, Y. Yagihashi, T. Musashi, J. Kawakami, N. Ohya, T. Gotoh and H. Sagami, *Journal of Molecular Catalysis B: Enzymatic*, 2012, **80**, 1–6.
11 K. Makabe and T. Kijima, 5AYP, to be published, https://www.rcsb.org/structure/5AYP, (accessed May 11, 2021).
12 A. A. Zamyatnin, *Progress in Biophysics and Molecular Biology*, 1972, **24**, 107–123.

4 Chapter 4 – Utilising 12-Hydroxy Farnesol to Synthesise 12-Hydroxy FDP (**105**) and DHAAl (**82**)

4.1 Introduction

In chapter 3 a synthetic pathway to DHAAl (**82**) beginning from isoprenol, and the chemically synthesised 4-hydroxy prenol (**106**) was described, as well as modifications made to improve this pathway. Nevertheless, this pathway still involves an initial chemical synthesis with a 40% yield. The next step, utilising ThiM for monophosphorylation is time consuming, involving 48-hour incubation, and although attempts were made to improve the diphosphorylation step they proved unsuccessful. This synthesis involves 4 enzymes and one chemical reaction to go from two relatively low-cost starting materials to DHAAl (**82**). Although this pathway is less laborious than the previous chemical phosphorylation route described in chapter 3, there was still scope for improvement.

An improved route to DHAAl (**82**), began with the possibility of the enzymatic hydroxylation of a 15-carbon terpene precursor. A cytochromes P450 (CYP), CYP124A1, has been reported to hydroxylate some terpenes at the 12 position including FDP (**40**) and farnesol (**102**).[3] CYP124A1 is a CYP obtained from *Mycobacterium tuberculosis* that is believed to have evolved to hydroxylate methyl branched lipids at the ϖ-position. A crystal structure, as shown in Figure 83, of this enzyme is available both free and bound to a substrate. This enzyme was shown to hydroxylate geraniol, farnesol (**102**), FDP (**40**) and geranylgeraniol. The K_D value of FDP (**40**) (90 ± 13) was 90 times higher than of farnesol (**102**) (1.04 ± 0.05), implying the diphosphate group greatly hinders the binding. A K_M value for FDP (**40**) was not measurable.[3]

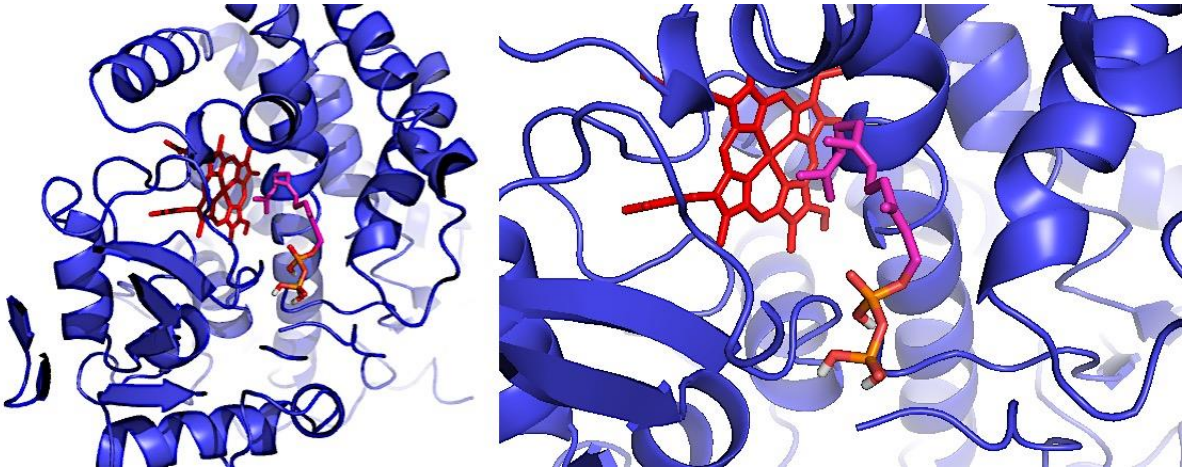


*Figure 83: Cartoon representations of a crystal structure of CYP124A1 in blue, with the heme cofactor in red, and a molecule of FDP (**40**) in magenta modelled into the active site, with the image on the right focusing more on the FDP (**40**) and heme cofactor. (PDB: 2WM4)*[3]

Some of the most studied CYPs include CYP101 and CYP102A1. CYP102A1 has been studied by many and has been mutated to accept alternative substrates. The natural substrates of CYP102A1 are fatty acids. Carmichael and Wong successfully modified CYP102A1 to accept hydrophobic polycyclic aromatic hydrocarbons (PAH). Residues at the entrance to the access channel which interact with the acid region of the natural substrate were changed to more hydrophobic residues to increase the likelihood of interaction with more hydrophobic substrates, residues are highlighted in Figure 84. Residues within the active site were also changed to create a larger cavity, and the unnatural substrates were larger than the fatty acids generally accept by the enzyme. One of the mutant CYP102A1s created showed activity with PAHs that was almost 50% of that shown with the wild type enzyme and its natural substrate.[8] CYP102A1 has also been mutated to accept unnatural terpene substrates. By altering residue I401 to proline (+)-α-pinene (**9**), a substrate not accepted by the wild type, was oxidised to verbenol. The rate of metabolism for many other non-natural substrates was also increased.[9]

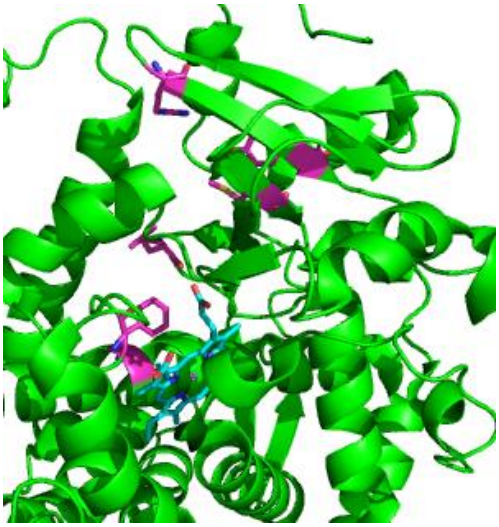


Figure 84: Cartoon representations of a crystal structure of CYP102A1 in green, with the heme cofactor in blue. (PBD: 1FAH)[10] *Key residues within the work done by Carmichael and Wong are shown in magenta.*

If CYP124A1 could be used to hydroxylate farnesol (**102**), (if the hydroxylation of FDP (**40**) proved unsuccessful) 12-OH farnesol (**109**) would then have to be diphosphorylated. The method previously used for diphosphorylation of 12-OH farnesol (**109**) is laborious and low yielding.[11] It involves two chemical synthesis steps, two cation exchange columns, a final purification step and multiple lyophilisations.[11] See materials and methods 6.2.3 for further detail. Thus, an alternative would be required for a less laborious synthesis. Undecaprenol kinase (UPK) (also called diacylglycerol kinase A) from *Streptococcus mutans* has been shown to phosphorylate farnesol to farnesyl phosphate.[12] A mutant form of IPK from *Methanocaldococcus jannaschii*, IPK F83A I86A I146A, has been reported to phosphorylate farnesyl phosphate to FDP (**40**) with almost 30% turnover.[13] Utilising UPK, a screen of various IPK mutants was investigated for optimising the diphosphorylation of farnesol (**102**) to FDP (**40**) with one mutant showing approximately 99% conversion.[14]

4.2 Aims

As it had been reported by Johnston *et al.* that FDP (**40**) could be hydroxylated by CYP124A1, a shortened synthesis to 12-OH FDP (**105**) would be possible, with a one step hydroxylation, and then conversion by ADS to DHAAl (**82**), as shown in red in Figure 85. This could potentially occur *in vivo* as organisms synthesise FDP (**40**) for other uses vital to life. Despite numerous attempts to repeat the work described by Johnston *et al.* successful oxidation of FDP (**40**) was never seen, only of farnesol (**102**), and this result proved difficult to reproduce.

Therefore, mutations in CYP124A1 were attempted to increase the affinity of the enzyme for the large, charged diphosphate group.

As the substrate for hydroxylation by the WT CYP124A1 could only be farnesol, an alternative route to 12-OH FDP (**105**) was also investigated, involving the enzymatic diphosphorylation of 12-OH farnesol (**109**). As screening IPK mutants had proved so successful in identifying an enzyme that diphosphorylated farnesol (**102**), they were repurposed in this work, along with UPK, in a screen for the diphosphorylation of 12-OH farnesol (**109**), see the route shown in black in Figure 85.

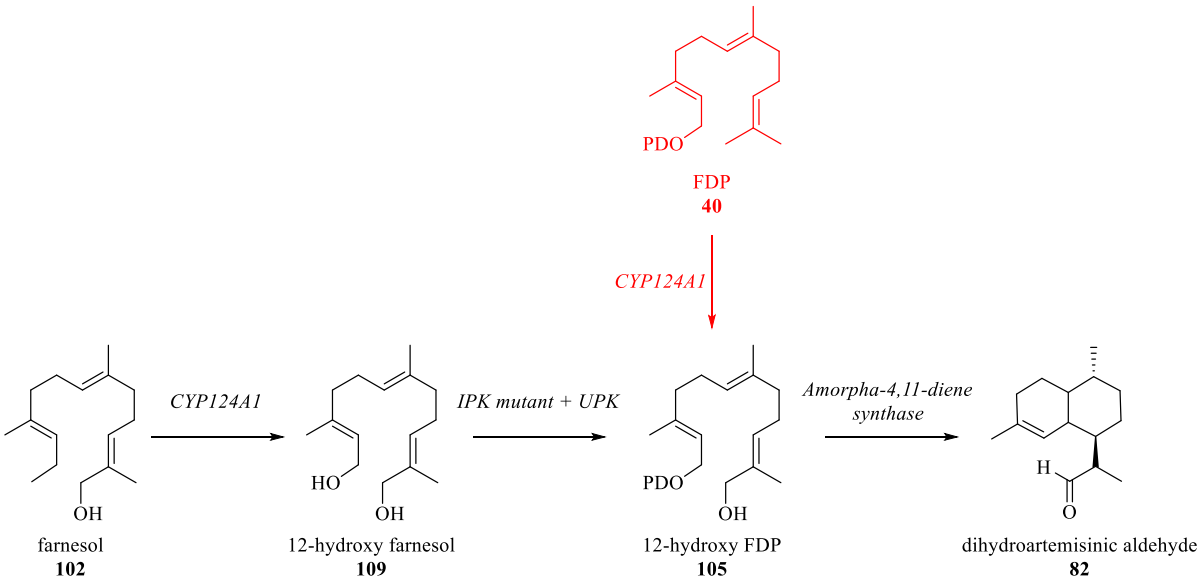


*Figure 85:Two possible enzymatic routes for the synthesis of 12-OH FDP (**105**) to be coupled with ADS for the synthesis of DHAAl (**82**)*

4.3 Results

4.3.1 Expression of CYP124A1

CYP124A1, a cartoon representation of the crystal structure is shown in Figure 86, was obtained in a pCWori vector with an N terminal His_6-tag and expressed in *Escherichia coli* (*E. coli*) BL21(DE3) cells as described in materials and methods 6.1.21. The plasmid was a gift from Professor Ortiz de Montellano, Washington State Univeristy.[3]

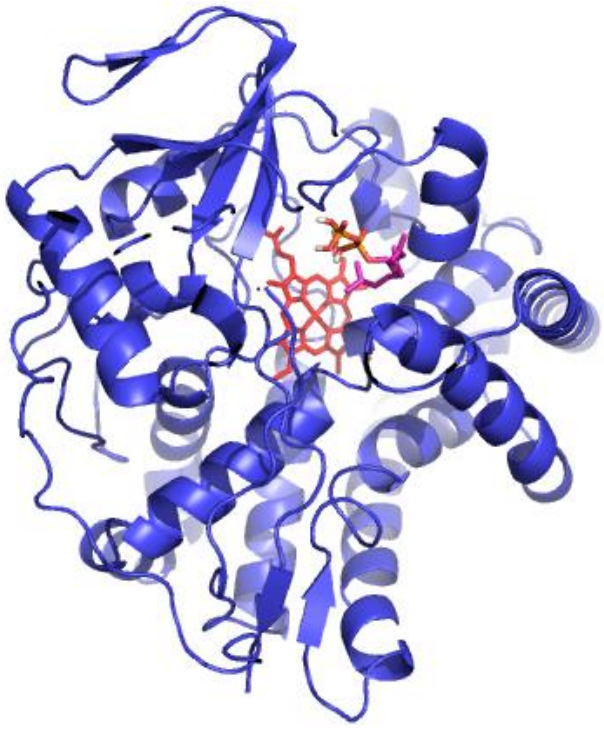


*Figure 86:Cartoon representations of a crystal structure of CYP124A1 in blue, with the heme cofactor in red, and a molecule of FDP (**40**) in magenta modelled into the active site. (PDB: 2WM4)*[3]

Optimisation of expression

Initial attempts to express CYP124A1 in *E. coli* proved unsuccessful with no expression bands seen on the protein gel, although the method reported in literature was followed. Briefly, CYP124A1 in the expression vector pCWori was expressed using DH5α cells, as described in the literature.[3] Cells were grown at 37 °C in TB media. Cells were induced with IPTG and supplemented with δ-aminolevulinic acid and $FeCl_3$.[15] Protein expression was investigated at various induction temperatures, 16 °C, 20 °C, and 25 °C overnight, and 37 °C for four hours. Purification of expressions showed very little to no protein in any of the fractions, see example polyacrylamide gel (7.3.4 in supplementary information). Various expression cell lines were

also tested. BL21(DE3) were tested as they are the most common expression cell line used. Rosetta (DE3) cells were tested to ensure the issue was not lack of tRNAs for various rare codons. ArcticExpress (DE3) cells were used to ensure the issue was not insolubility of the protein. C41(DE3) and C43(DE3) cell lines were used as they are designed to express toxic proteins. In expression tests all cells lines (but ArticExpress (DE3)) were tested at the varying induction temperatures as lower temperatures allow slower folding (ArcticExpress (DE3) cells were tested only at 16 °C) but none showed clear protein expression. See materials and methods 6.1.4 for further information regarding cell lines.

All expression tests with various media utilised BL21(DE3) cells. LB, LBE, TYE and TB 1% glucose media were used in expression trials. Purifications from both TYE and the TB 1% glucose media run on an SDS-PAGE gel showed clear bands of overexpression of the correct molecular weight in the elution fractions (see 7.3.4 in supplementary information) whilst cell grown in the other two medias did not. It seemed that the commonality was the presence of glucose. Although C41(DE3) and C43(DE3) were tested to determine if the issue was protein toxicity, it seemed that leaky expression of all cell lines when not grown in 1% glucose meant that no protein was expressed. Possibly this was due to the toxicity of CYP124A1. All temperatures and times tested for various cell lines are summarised in Table 6.

Table 6: Summary of cell lines and conditions tested for optimisation of CYP124A1 expression

CELL LINE	Bl21(DE3)	Rosetta (DE3)	C41(DE3)	C43(DE3)	Artic Express (DE3)
Induction temperature (°C)	**Induction time (h)**				
16	16	16	16	16	16
20	16	16	16	16	
25	16	16	16	16	
37	4	4	4	4	

CYP124A1 activity assays

CYPs require an electron donor which is commonly NADPH. Glucose-6-phosphate and glucose-6-phosphate dehydrogenase regenerate this NADPH from $NADP^+$ in the assay as shown in Figure 87. NADPH is a two-electron donor and donates its electrons to ferredoxin catalysed by the ferredoxin reductase. Ferredoxin then passes these electrons one at a time from NADPH to the CYP124A1.[16] See page 34-35 of introduction for further information.

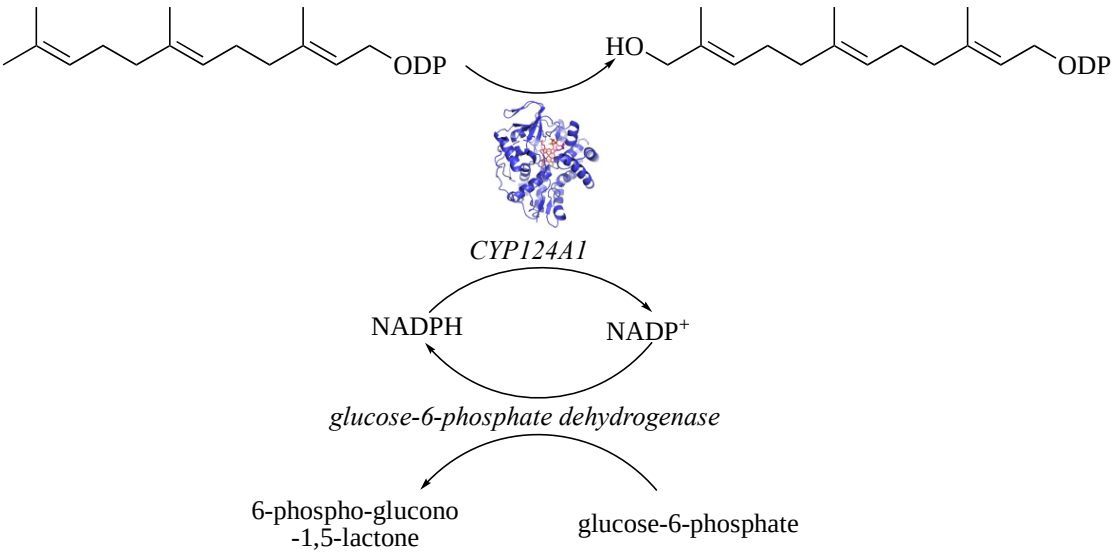


Figure 87: Schematic of the CYP124A1 oxidation and regeneration system.

Once CYP124A1 was successfully expressed and purified, it could be used in assays with FDP (**40**) to determine its activity. Assays were performed concurrently, with farnesol (**102**) as a positive control and negative control assays with either the enzyme or NADPH omitted. CYP124A1 and the appropriate substrate were incubated in buffer with ferredoxin and ferredoxin reductase, catalase, glucose-6-phosphate and glucose-6-phosphate dehydrogenase for 10 minutes at room temperature. NADPH in excess was then added to initiate the reaction. Assays were left overnight shaking at room temperature. See 6.1.23 of materials and methods for further information.

After overnight incubation, the pentane extracts from the assay samples containing farnesol (**102**) were analysed using a GC-MS. Whereas assay samples containing FDP (**40**) were treated with alkaline phosphatase to remove the diphosphate group allowing its extraction with organic solvent. As the expected product for both assays on the GC-MS would be 12-OH farnesol (**109**), it was synthesised chemically in the laboratory and used as a standard for GC-MS analysis. Assays containing farnesol (**102**) as a substrate showed a product whose retention time corresponded to the chemically synthesised 12-OH farnesol (**109**) standard, see figure 88, however, assays containing FDP (**40**) as a substrate showed a product whose retention time corresponded to a standard of farnesol (**102**). Therefore, FDP (**40**) was not turned over by wild type CYP124A1, while farnesol was (**102**).

Chemical synthesis of 12-OH farnesol (**109**) as a standard

A chemical standard was needed for gas chromatography mass spectrum analysis to identify if CYP124A1 successfully hydroxylated farnesol. The standard was used as a comparison for any products extracted from CYP124A1 assays, see Figure 88. Briefly, commercially sourced farnesol (**102**) was hydroxylated *via* a selenium dioxide oxidation before purification *via* column chromatography. See 6.2.4 of materials and methods for further information.

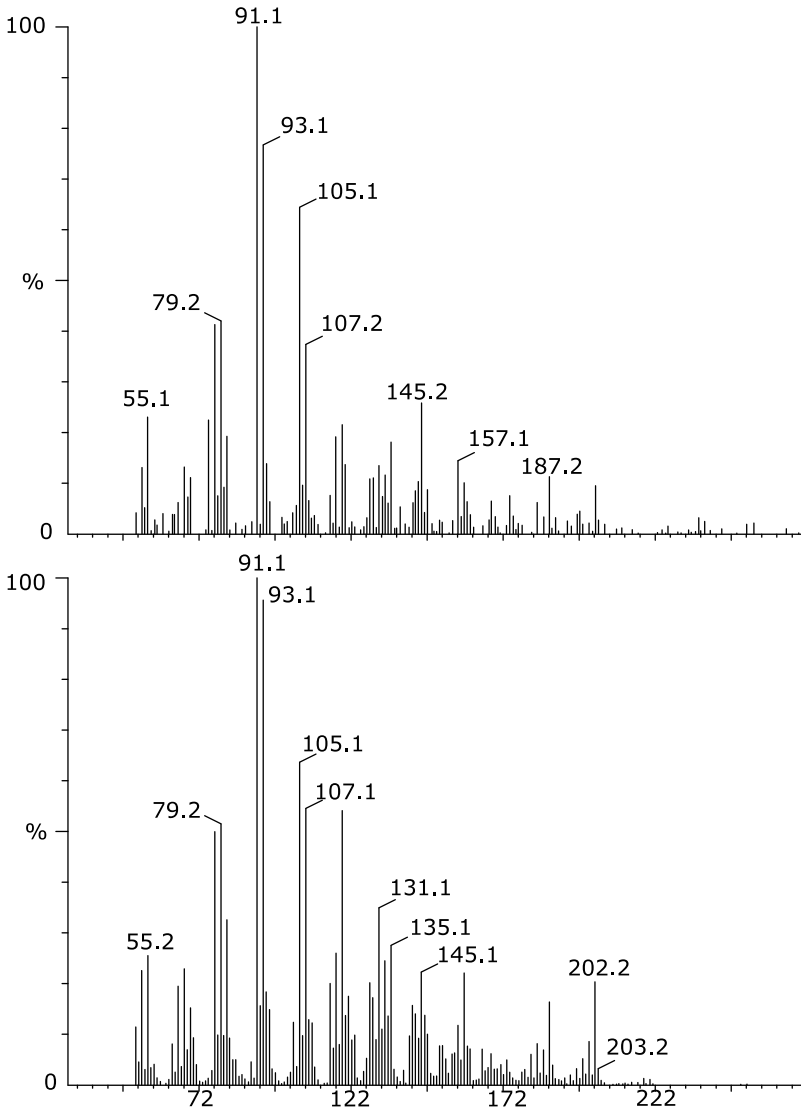


*Figure 88: EI mass spectrum from GC-MS analysis of: top, the chemical standard of 12OH-Farnesol; bottom, the biological assay containing farnesol (**102**) and CYP124A1*

In vivo work in Saccharomyces cerevisiae

Whilst the previous work was being done, simultaneous attempts were made to express CYP124A1 in *Saccharomyces cerevisiae* AM109 along with ADS. The strain was designed by

Ignea *et al.* for the optimisation of caryophyllene production. It contains a knock-out of one copy of squalene synthase ERG9 which uses FDP (**40**) for ergosterol production, allowing more FDP (**40**) to funnel into sesquiterpene synthesis. It also has deletion of 3 genes that degrade HMG-CoA reductase, a key gene in the mevalonate pathway, see page 4 of the introduction for further information. The strain also has a gene in the tryptophan pathway, trp1, knocked out to allow for selection of colonies successfully transformed with a plasmid containing trp1.[17]AM109 was transformed with a plasmid, pESC-Rv2266-ADS, containing trp1, ADS and CYP124A1. It was hypothesised that the natural FDP synthesis of *S. cerevisiae* could be hydroxylated by CYP124A1 and ADS could convert it to DHAAl (**82**). DHAAl (**82**) would be released as a volatile into the cultures organic overlay. To ensure transformations were successful colony PCR was performed.

Successfully transformed *S. cerevisiae* were cultured for two days before induction with galactose and the cultures were overlaid with dodecane. Both ADS and CYP124A1 were under the control of a galactose promoter. Samples of the dodecane overlay were taken every day for five days once the cultures were induced and analysed on a GC-MS to identify any volatiles produced. Any peaks were then compared to standards of amorpha-4,11-diene (**34**) and DHAAl (**82**). Despite numerous attempts only amorpha-4,11-diene (**34**) was seen in the dodecane overlay, as shown in Figure 89.

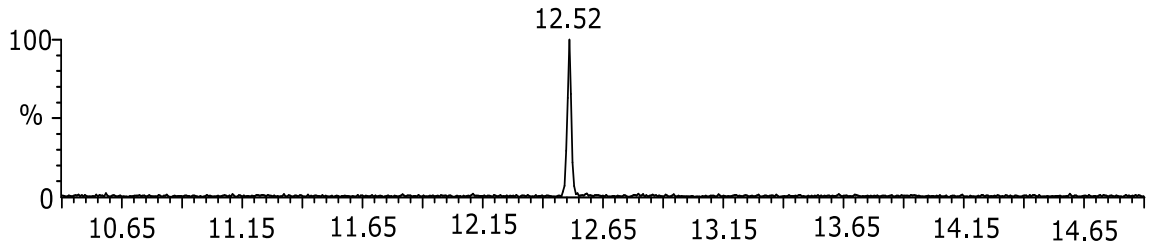


Figure 89: TIC from GC-MS analysis of an exemplary trace of S. cerevisiae *transformed with the plasmid containing ADS and CYP124A1 after induction, the peak at 12.52 minutes corresponds to the chemical standard of amorpha-4,11-diene* ***(34)***

It was suspected that CYP124A1 was not expressed in *S. cerevisiae* due to the same issues as expressing in *E. coli*. Therefore, western blots for both ADS and CYP124A1 expression was performed. A Myc tag had been placed before the ADS gene for western blotting with an anti-Myc antibody, and a FLAG tag placed before the CYP124A1 gene. When using an antibody for the Myc tag which was fused to the ADS gene, bands were seen on the western blot as shown in Figure 90. ADS is 62 kDa. When using an anti-FLAG tag antibody no bands or very

faint bands across the entire blot were seen indicating no expression, or degradation of CYP124A1 as shown in Figure 90. A positive control was also run on the western blot to ensure the issue was not the antibody, or the technique. *In vivo* cell work involving CYP124A1 was therefore unlikely to succeed.

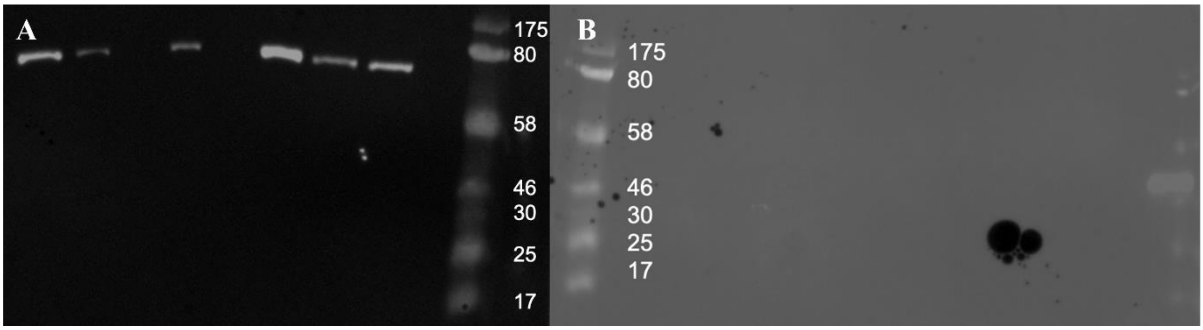


Figure 90:Western blot of S. cerevisiae *cultures transformed with ADS and CYP124A1. A) Top bright bands of chemiluminescence show antibody binding to a Myc tag which was fused to ADS. The last lane is a ladder. B) The last lane is a positive control with a known protein containing a FLAG tag. The first lane is a ladder. No expression of CYP124A1 was seen, as there should be bands in all lanes but the first and last at 48.6kDa.*

4.3.2 Mutants of CYP124A1

Since the wild type CYP124A1 showed no activity with the desired substrate FDP (**40**), but was shown to be active with farnesol (**102**), it was concluded that optimisation was necessary for the enzyme to allow hydroxylation of FDP (**40**). The crystal structure of CYP124A1 (PDB 2WM4) with a known substrate, phytanic acid (**110**) was obtained.[3] The natural substrate was modified, *in silico* using PyMol, to FDP (**40**). The YASARA force field energy minimiser was used to determine the most appropriate conformation of the FDP (**40**).[18]

Residues that seemed to be able to coordinate the diphosphate group when examining the model were chosen to be changed to residues with polar side chains to hopefully facilitate hydrogen bonding to the oxygens of the diphosphate group. Residues chosen are summarised in Table 7 and the structures shown in Figure 91.

Table 7: Table of CYP124A1 M. tuberculosis *variants created and expressed within this work* via *SDM*

Amino acid position	WT residue	Changed residue
101	T	S
103	N	S
104	D	S
421	Q	L
422	F	R
423	I	T

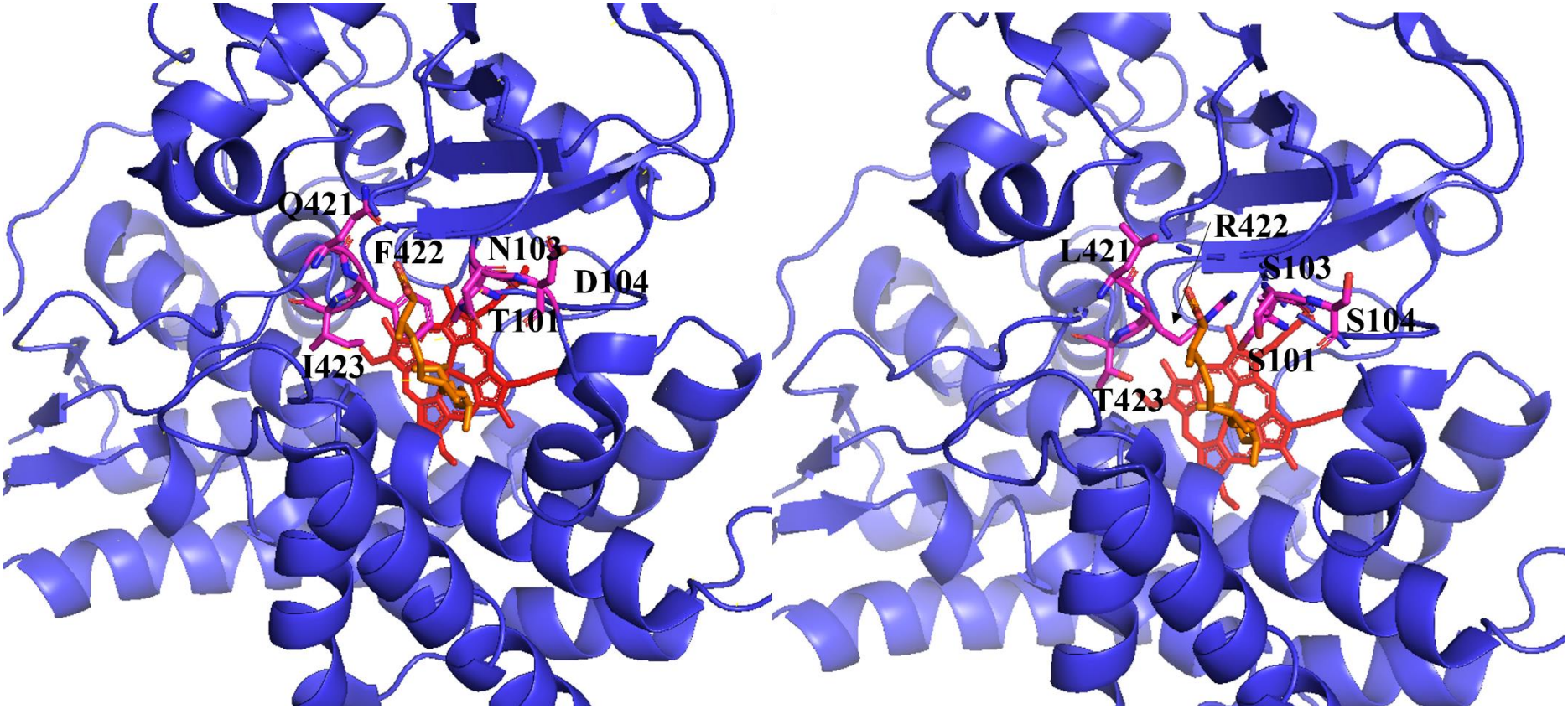
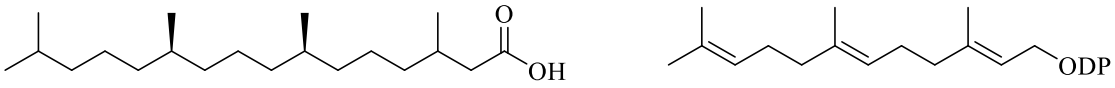


Figure 91: Cartoon representation of an X-ray crystal structure of M. tuberculosis *CYP124A1. On the left are the wild type residues, and on the right are the changed residues, shown in magenta. Heme is shown in red, and the substrate is shown in orange, (PDB: 2WM4)*[3]

Threonine 101 was changed to serine as it still contains a hydroxyl group for hydrogen bonding but it is a smaller residue. Asparagine 103 was changed to serine, with the hope that it could hydrogen bond to an oxygen in the diphosphate group making FDP (**40**) binding more favourable. Glutamine 421 was chosen due to its close proximity of 3.4 Å to the diphosphate group of FDP (**40**) modelled into CYP124A1's active site. Changing glutamine to leucine creates a little more space for the bulky diphosphate group, but the change is not too dramatic as to try and avoid disrupting the structure of the helix around the residue, *e.g.* if it had been changed to alanine. Isoleucine 423 was chosen to change to threonine to increase the space in the active site and to introduce a hydroxyl group to interact with the diphosphate group. Phenylalanine 422 sits near where the diphosphate group is predicted to be. By changing this to arginine a positively charged residue it was hypothesised this would aid in the substrate sitting in the active site via positive electrostatic interactions. Aspartate 104 sits in the active site near where the diphosphate group of FDP (**40**) is predicted to be. As this is a negatively charged residue it may repulse the diphosphate group of FDP (**40**) hindering it ability to sit in the active site. Therefore, this residue was changed to serine, as it is a similar size, but has a hydroxyl group promoting hydrogen bonding with the diphosphate group. See 7.3.5 of supplementary information for cartoon representations of a crystal structure of CYP124A1 with the individual various changes described.

Unfortunately, none of the variants showed any activity with FDP (**40**) in the assays described above. It may be that such a large change in substrate requires a combination of changes to accommodate the large pyrophosphate group and that individual point mutations are not sufficient. Also, as the crystal structure of the enzyme did not have FDP (**40**) in the active site it is impossible to be certain that the residues chosen are close to the diphosphate group. It may also be that a mutation further away from the active site is needed to create a larger structural change to allow room for the diphosphate group. Phytanic acid (**110**), as shown in Figure 92, the substrate which was crystallised in CYP124A1, has a longer carbon chain, but a smaller tail group so may not curve, in the active site in the same way. It is incredibly difficult to predict activity in engineered proteins and often a large panel of mutants are made before one with the desired activity is identified.[8] This work highlights how it is still not fully understood how proteins interact with ligands within the active site during the enzymatic reaction. Before resuming this work, a crystal structure containing an FDP (**40**) analogue; *e.g.* 1-hydroxy-3,7,11-trimethyldodeca-2,6,10-triene phosphonic acid[19] which can be used in sesquiterpene synthase crystallisation; would be greatly beneficial. This could help clarify how substrates

with shorter carbon chains and larger polar tails interact with the residues within the active site. Throughout this work it was found that the hydroxylation of farnesol (**102**) was unreliable and difficult to replicate, therefore the remaining work was focused on the diphosphorylation step.



phytanic acid
110

FDP
40

*Figure 92: Structure of the ligand bound within the crystal structure of CYP124A1, phytanic acid (**110**), and FDP (**40**) the substrate modelled into the CYP124A1 active site*

4.3.3 Enzymatic diphosphorylation of hydroxylated prenyl analogues

Enzymatic diphosphorylation of 12-OH Farnesol (**109**)

As CYP124A1 had been shown to hydroxylate farnesol (**102**) at the 12 position, kinases could be used to phosphorylate the 1 position, synthesising 12-OH FDP (**105**). Previous work has shown that UPK and a mutant of IPK can diphosphorylate farnesol (**102**)[12,13], therefore UPK and a panel of IPK mutants, shown in Table 8 (cloned by Dr. F. Huynh)[14] were screened in this work for turnover with 12-OH farnesol (**109**).

Table 8: Table of M. jannaschii *variants expressed and tested in this work, generated by Dr. F. Huynh* via *SDM*

Variant	Residue of interest				
	F76	**F83**	**I86**	**I146**	**I156**
WT	F	F	I	I	I
IPK mutants					
No. 57	F	A	A	A	I
No. 47	A	A	A	A	A
No. 45	F	A	A	A	A
No. 43	A	A	A	I	A
No. 41	F	A	I	A	A
No. 39	F	A	A	I	A
No. 35	A	A	A	I	I
No. 24	A	A	I	A	A
No. 23	A	A	I	A	I
No. 16	A	A	I	I	A

Assays containing UPK, ATP and the regenerating system along with 12-OH farnesol (**109**) and a separate assay with farnesol (**102**) as a control were incubated for two hours. See 6.1.37 of materials and methods for further details. A ^{31}P NMR spectrum of the assay mixture was performed to ensure the monophosphorylation reaction had occurred and compared to the NMR spectrum of the same reaction performed on farnesol. See 7.3.6 of supplementary information for NMR spectra. A white precipitate formed which was isolated and analysed on a TLC (6:3:1 isopropanol:ammonia:water). It was compared to an authentic standard of farnesyl monophosphate and the starting alcohol substrate to confirm it was the monophosphorylated product. Throughout this work, it was anticipated that both hydroxyl groups could be individually monophosphorylated by UPK, summarised in Figure 93, and that both hydroxyl groups may be monophosphorylated on the same compound, but only the correct compound, 12-OH FDP (**105**) would be converted to DHAAl (**82**).

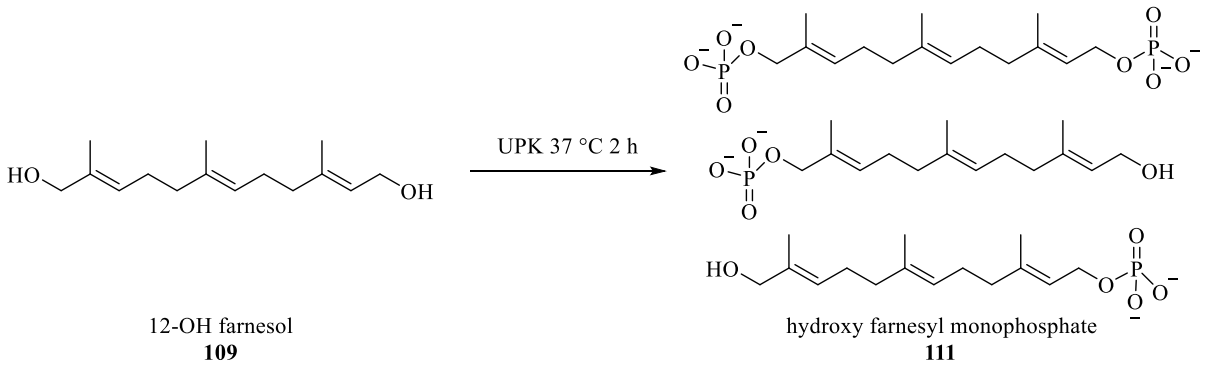


*Figure 93: Monophosphorylation of 12-OH farnesol (**109**) by UPK and the possible products*

Once the white precipitate was confirmed to be the monophosphorylated product, it was then used for further assays with a panel of IPK mutants. As both the monophosphorylated and the diphosphorylated products precipitate in high concentrations, it was difficult to identify the diphosphorylated product using NMR spectroscopy. Therefore, the diphosphorylation reaction was coupled to a sesquiterpene synthase, ADS, to allow quantification on a GC-MS. If the IPK mutant converted the monophosphate (**111**) to the diphosphate (**105**), the monophosphate (**111**) would not be present, while the product of incubating ADS and 12-OH FDP (**105**), DHAAl (**82**) would be present in the organic extract, as shown in Figure 94.

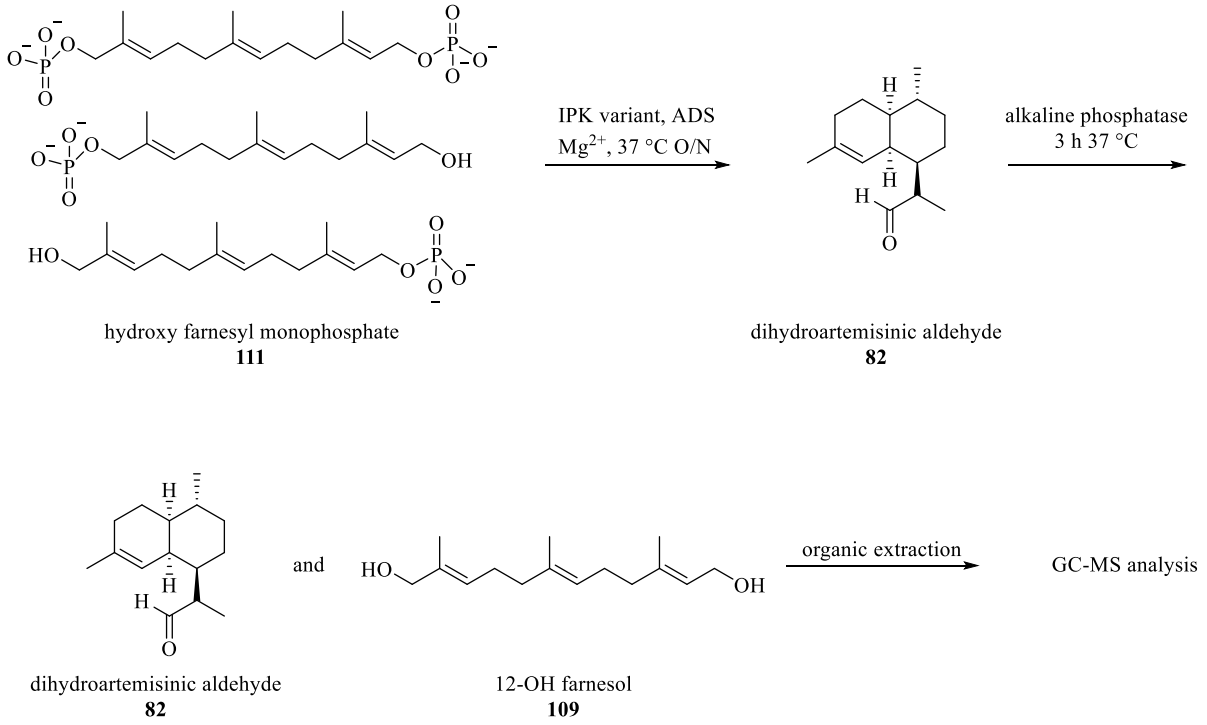


Figure 94: IPK-ADS coupled reaction treated with alkaline phosphatase to test IPK variants activity.

ADS was added to the reaction at 20 times the concentration of IPK to ensure ADS was present in a large excess. See 6.1.37 of materials and methods for further details. After overnight

incubation, alkaline phosphatase was added to the assay to convert any remaining monophosphate (**111**) back to the alcohol (**109**) to allow extraction into the organic solvent. The organic layer was then analysed by GC-MS as exemplified in Figure 95. The activity of each IPK variant could then be compared *via* the ratio of 12-OH farnesol (**109**) (from treating the monophosphate (**111**) with alkaline phosphatase) and DHAAl (**82**). Controls without IPK, ADS or either were also performed, and no DHAAl (**82**) was observed.

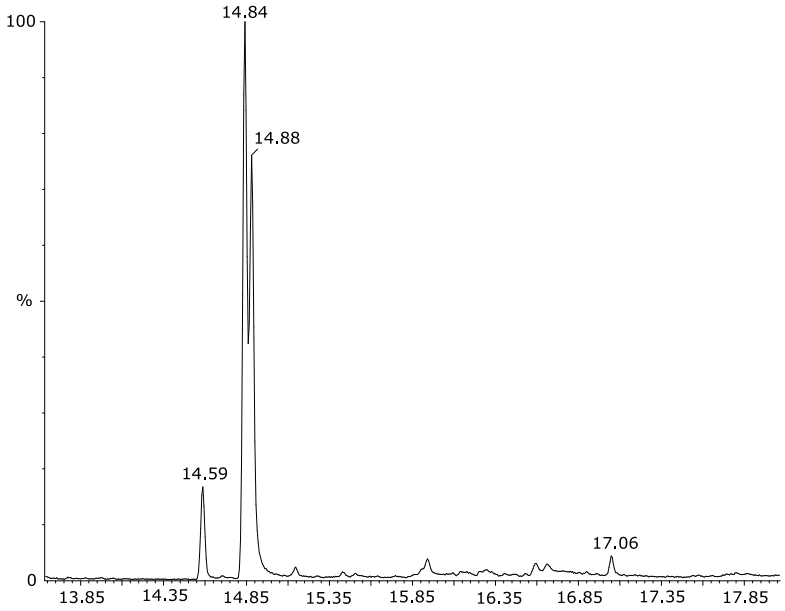


*Figure 95: Example TIC from GC-MS analysis of IPK16-ADS coupled assay treated with alkaline phosphatase. The peak at 14.59 minutes is artemisinic-11S,12-epoxide, 14.84 minutes is R- dihydroartemisinic aldehyde (**82**), 14.88 minutes is S-dihydroartemisinic aldehyde (**82**), and at 17.06 minutes is 12-OH farnesol (**109**).*

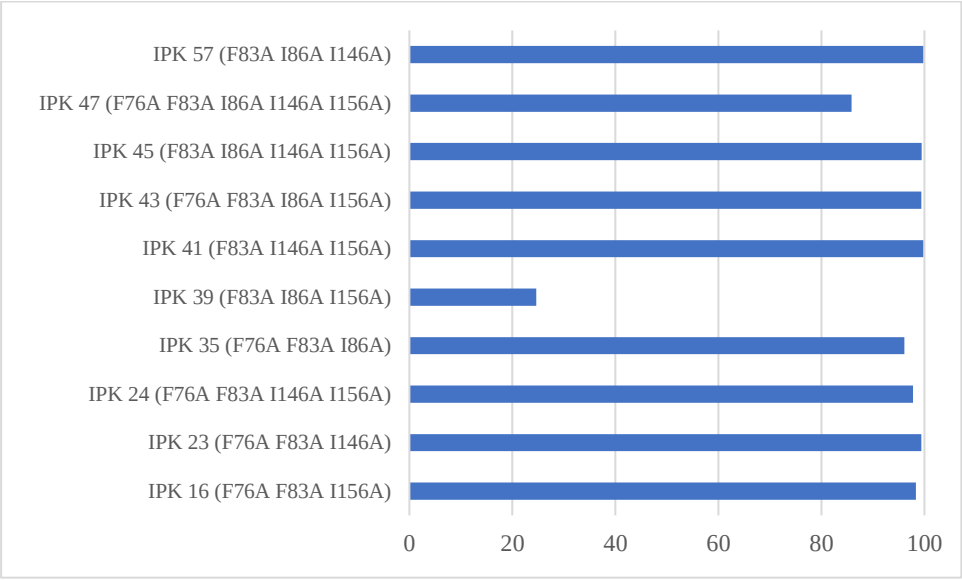


Figure 96: Comparison of the activity of 10 IPK variants in the IPK-ADS coupled assay

Although there is no definitive answer for which mutations are best for improving yield, there are certain trends within the results shown in Figure 96. F83A is present in all variants and therefore its contribution cannot be distinguished. Generally, diphosphorylation decreases with the presence of the change F76A which sits in the back of the pocket, as shown in Figure 97. This is shown in the yield of variants IPK 45 and IPK 24, which have the same changes except I86A instead of F76A respectively. Also, variants IPK 24 and 41 have the same changes, apart from the addition of F76A to IPK 24, which has the lower yield of the two.

As illustrated in Figure 97, I86A sits in the front of the binding pocket and it may be that increasing space in the front of the pocket is more important than in the back, possibly as the carbon chain may bend within the active site. Corroborating this, variants containing the change I146A, which sits in the front of the active site, is always associated with higher yield, with the only exception being IPK 47. This variant additionally contains 4 other changes, and it may be that such a drastic change in the enzymes structure inhibits general activity. The change I156A is associated with higher yields when the change I146A is not present, but when variants with the same changes except I146A vs I156A are compared, such as IPK 23 and IPK 16, it is seen that I146A leads to higher yields. Therefore, it seems that of the residues that can be assessed, those which have the largest impact are in the front of the binding pocket, I86A and I146A, which can be seen from the data in Figure 96.

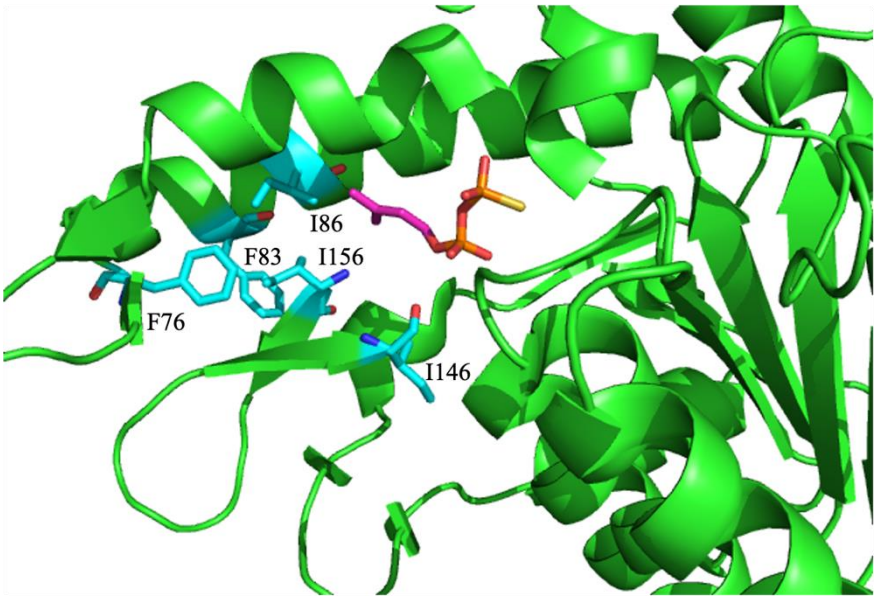


Figure 97: Cartoon representation of an X-ray crystal structure of M. jannaschii *IPK with residues changed labelled (PDB: 3K56)*[13]

ADS yields 3 products in a reaction with chemically synthesised 12-OH FDP (**105**), which are *R* and *S*-DHAAl (**82**) and artemisinic-11-*S*,12-epoxide,[14] with around 10% of the product being the epoxide, and the aldehyde epimers in a 2:7 ratio respectively, with the 11-*R* epimer being the desired product.[20] Using UPK and IPK to diphosphorylate 12-OH farnesol (**109**) to make 12-OH FDP (**105**) gave different ratios of aldehydes as shown in Figure 98, with the ratios being approximately 1:1. This is in contrast to the DHAAl (**82**) produced by 4-hydroxy prenol (**106**), ThiM, WT IPK, FDPS and ADS which matches the ratios reported by Demiray *et al.*,[20] as shown in the previous chapter.

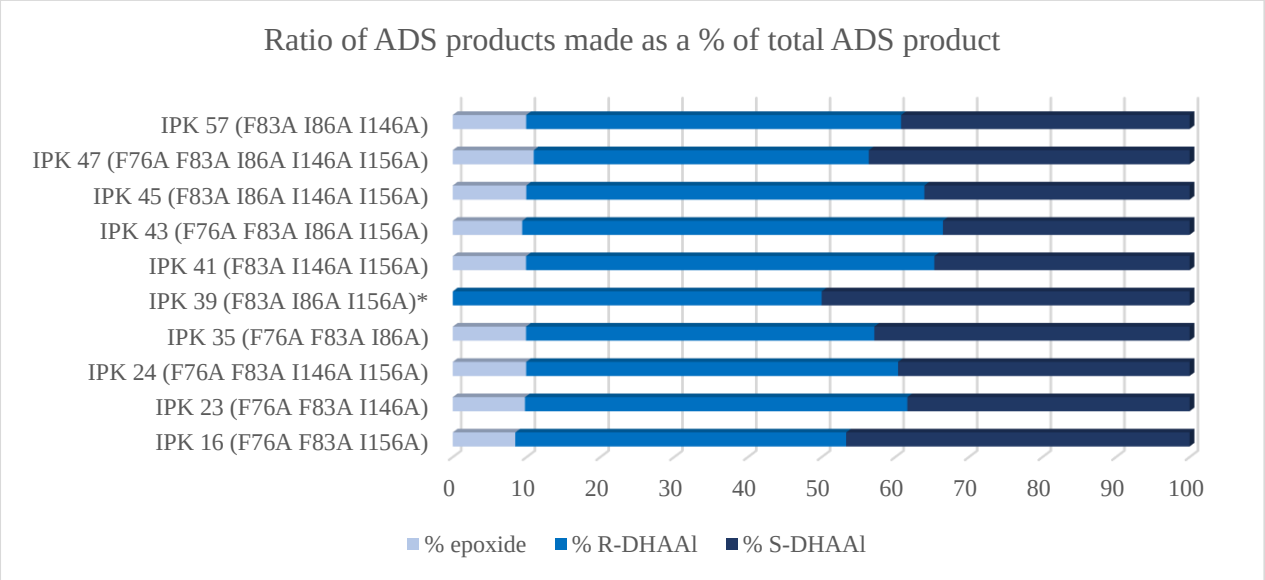


*Figure 98: Comparison of the ratio of products produced from incubations of 12-OH farnesol (**109**), UPK, ADS and various IPK variants from total ADS products. *IPK 39 ratios are unreliable as so little product was made.*

This change in ratio was unexpected, and it was proposed that the variation was due to a change in substrate availability. In the case of the chemical synthesis of 12-OH FDP (**105**), all of the substrate is available at the beginning of the assay, saturating the enzyme. In contrast, the enzymatically synthesised 12-OH FDP (**105**) is made progressively, as the IPK turns over the monophosphate (**111**). Therefore, assays varying the concentration of IPK:ADS from 1:20, 1:10, 1:5, 1:1, 4:1, 8:1, 16:1 to 20:1 µM were completed. If the change in ratio of *R*:*S* were related to substrate availability at the beginning of the assay, it would be expected that the ratio of *R*:*S* would change as the ratio of IPK increases respective to ADS. This was not clearly seen, with all assays showing close to a 1:1 ratio of *R*:*S*. Some assays showed some variation from the 1:1 ratio, but unfortunately, this change did not occur in a linear way as the IPK concentration increased. Therefore, it could reasonably be argued that this was due to variation

between assays, rather than an increase in IPK concentration, see figure 7.3.6.3 and 7.3.6.4 in supplementary information. The other hypothesis was that the presence of the ammonium counter ion in the chemically synthesised 12-OH FDP (**105**) affected the ratio of *R*:*S*, but assays performed in the presence of ammonia also showed an ~1:1 ratio of *R*:*S*, see figure 7.3.6.5 in supplementary information.

Future work to determine if this aberration in *R:S* ratio is caused by substrate availability could involve large scale synthesis of 12-OH FDP (**105**) made by UPK and an IPK variant. Enzymatically synthesised 12-OH FDP (**105**) could then be purified and lyophilised, so it is in the same state as the chemically synthesised 12-OH FDP (**105**), before being subjected to the same assay as chemically synthesised 12-OH FDP (**105**) with ADS. Nevertheless, the consistently low yields obtained in UPK expression could turn out to be a severe limitation in the large scale set up required for an efficient diphosphate purification. For this reason and time constraints, this avenue was not explored further.

4.4 Conclusion

A shortened synthetic pathway for the production of DHAA1 (**82**), the precursor of antimalarial artemisinin (**4**), has previously been reported,[20] where several steps are inefficient and laborious. This chapter focusses on a route investigated to improve the synthesis of 12-OH FDP (**105**), a precursor of DHAAl (**82**).[21] Both the hydroxylation at the 12 position of FDP (**40**) / farnesol (**102**) and the diphosphorylation at the 1 position of 12-OH farnesol (**109**) were investigated.

Initial work investigated a P450, CYP124A1, reported to hydroxylate FDP (**40**) at the 12 position. These results were never replicated, neither *in vitro* nor *in vivo* with both routes hindered by CYP124A1's toxicity to the host organism. Once a method for obtaining quantifiable amounts of CYP124A1 as a pure protein was discovered, *in vitro* assays could be used. As the natural substrate of CYP124A1 does not have a large polar group like the FDPs (**40**) diphosphate group, variants of CYP124A1 were designed with residues containing polar side chains to favour hydrogen bonding, or smaller side chains to aid in accommodating the large diphosphate group. Unfortunately, none of these variants yielded hydroxylated FDP (**105**). CYP124A1 was shown to hydroxylate farnesol (**102**) in the correct position, but over time it became clear the result was unreliable and difficult to repeat.

As CYP124A1 can hydroxylate farnesol (**102**) to produce 12-OH farnesol (**109**), the diphosphorylation of this compound was investigated. Previous work within the group had utilised a pair of kinases, UPK and mutant IPK,[14] diphosphorylate farnesol (**102**). This work was repurposed for the diphosphorylation of 12-OH farnesol (**109**), and a variant of IPK was identified (IPK F83A I146A I156A) that gave very high turnover from the monophosphate (**111**), producing DHAAl (**82**) in a coupled reaction with ADS. It was also seen that UPK, which was used for the monophosphorylation step, was far more efficient than ThiM. This was the enzyme used for the monophosphorylation of the 5 carbon compounds in the synthetic pathway described in chapter 3. This enzymatic synthesis also showed improved epimeric selectivity for the desired 11-*R* epimer of DHAAl (**82**).

Enzymatic hydroxylation of farnesol (**102**) has been shown, in the literature,[22] and within this work. Although this proved difficult to reproduce, it is likely this would be a viable route to producing 12-OH farnesol (**109**) in a regiospecific reaction after further optimisation.

Enzymatic diphosphorylation of 12-OH farnesol (**109**) is a less laborious route, and 12-OH FDP (**105**) can be isolated if desired as it can precipitate in solution. The diphosphorylation and subsequent cyclisation by ADS can all be done in one pot, rather than involving multiple lyophilisation steps and ion exchange columns[11] before an assay can be set up. This is therefore an improved route to synthesised 12-OH FDP (**105**), a precursor of the medically relevant compound dihydroartemisinic aldehyde (**82**).

1 CDC, Centers for Disease Control and Prevention, https://www.cdc.gov/malaria/malaria_worldwide/impact.html, (accessed July 20, 2019).
2 N. J. White, *Science (1979)*, 2008, **320**, 330–334.
3 J. B. Johnston, P. M. Kells, L. M. Podust and P. R. Ortiz de Montellano, *Proceedings of the National Academy of Sciences*, 2009, **106**, 20687–20692.
4 D. J. Cook, J. D. Finnigan, K. Cook, G. W. Black and S. J. Charnock, in *Advances in Protein Chemistry and Structural Biology*, Academic Press Inc., 2016, vol. 105, pp. 105–126.
5 I. G. Denisov, T. M. Makris, S. G. Sligar and I. Schlichting, *Chemical Reviews*, 2005, 105, 2253–2277.
6 R. Fasan, *Nature Chemistry*, 2017, **9**, 609–611.
7 J. Han, H. Wang, S. Kanagarajan, M. Hao, A. Lundgren and P. E. E. Brodelius, *Molecular Plant*, 2016, **9**, 946–948.
8 A. B. Carmichael and L. L. Wong, *European Journal of Biochemistry*, 2001, **268**, 3117–3125.
9 C. J. C. Whitehouse, S. G. Bell, W. Yang, J. A. Yorke, C. F. Blanford, A. J. F. Strong, E. J. Morse, M. Bartlam, Z. Rao and L. L. Wong, *ChemBioChem*, 2009, **10**, 1654–1656.
10 H. Yeom, S. G. Sligar, H. Li, T. L. Poulos and A. J. Fulco, *The Role of Thr268 in Oxygen Activation of Cytochrome 450ßμ-3*, 1995, vol. 34.
11 V. J. Davisson, A. B. Woodside, T. R. Neal, K. E. Stremler, M. Muehlbacher and C. D. Poulter, *Journal of Organic Chemistry*, 1986, **51**, 4768–4779.
12 L. Y. Huang, S. H. Huang, Y. C. Chang, W. C. Cheng, T. J. R. Cheng and C. H. Wong, *Angewandte Chemie - International Edition*, 2014, **53**, 8060–8065.
13 N. Dellas and J. P. Noel, *ACS Chemical Biology*, 2010, **5**, 589–601.
14 F. Huynh, PhD, Cardiff University, 2020.
15 H. Ouellet, L. M. Podust and P. R. Ortiz De Montellano, *Journal of Biological Chemistry*, 2008, **283**, 5069–5080.
16 D. Hamdane, C. Xia, S. C. Im, H. Zhang, J. J. P. Kim and L. Waskell, *Journal of Biological Chemistry*, 2009, **284**, 11374–11384.
17 C. Ignea, F. A. Trikka, I. Kourtzelis, A. Argiriou, A. K. Kanellis, S. C. Kampranis and A. M. Makris, *Microbial Cell Factories*, 2012, **11**, 1–16.
18 E. Krieger, K. Joo, J. Lee, J. Lee, S. Raman, J. Thompson, M. Tyka, D. Baker and K. Karplus, *Proteins: Structure, Function and Bioinformatics*, 2009, 77, 114–122.
19 C. M. Starks, K. Back, J. Chappell and J. P. Noel, *Science (1979)*, 1997, **277**, 1815–1820.
20 M. Demiray, X. Tang, T. Wirth, J. A. Faraldos and R. K. Allemann, *Angewandte Chemie - International Edition*, 2017, **56**, 4347–4350.
21 X. Tang, M. Demiray, T. Wirth and R. K. Allemann, *Bioorganic and Medicinal Chemistry*, 2018, **26**, 1314–1319.
22 J. B. Johnston, A. A. Singh, A. A. Clary, C. K. Chen, P. Y. Hayes, S. Chow, J. J. De Voss and P. R. Ortiz De Montellano, *Bioorganic and Medicinal Chemistry*, 2012, **20**, 4064–4081.

5 Chapter 5: Discussion and Future Work

The work described in this thesis was aimed at furthering our understanding of the flexibility and selectivity of amorpha-4,11-diene synthase (ADS) for its application in biocatalytic production of artemisinin (**4**). This was done *via* site directed mutagenesis of key residues within the active site thought to affect shaping the achiral substrate FDP (**40**) into the bicyclic product amorpha-4,11 diene (**34**). The product profile generated, and kinetic parameters of these variants were determined. The other part of this project involved utilising this flexibility to produce an efficient whole or partially enzymatic synthesis of DHAAl (**82**), which involves ADS turning over the FDP (**40**) analogue 12-OH FDP (**105**). Two key chemical reactions in the synthesis of DHAAl (**82**) were identified as being inefficient, and throughout this work enzymatic approaches were attempted to replace these chemical steps with more selective and efficient enzyme catalysed reactions. The first route identified used inexpensive 5 carbon starting materials, employing just one chemical (non-enzymatic) step and 4 enzymatic steps to produce DHAAl (**82**). The second route employed the starting material farnesol (**102**) and a route using 4 enzymatic steps was investigated to produce DHAAl (**82**). Both routes proved far less labour intensive, and the second showed significant improvement in yield for the diphosphorylation step.

5.1 Chapter 2

Work has been done in the Allemann group and in others to investigate the catalytic role played by various residues of ADS on its mechanism and kinetic activity.[1–3] Steady state kinetic measurements of the wild type ADS were measured, with good agreement to literature values, and used as a comparison for steady state kinetics of variants throughout this work. Investigations into the impact of changing residues in the active site of ADS showed a general robustness to change in product with the majority of variants showing no change in product profile but clear changes in kinetic parameters.

Extensive previous work by the Allemann group and others has shown that changes in terpene synthase G helices often lead to changes in product profile.[4–8] In this work, varying key residues in the G helix of ADS such as G401 and L405 effected significant changes in the product profile, producing selina-4(15),7(11)-diene (**101**) and germacrene B (**71**). These compounds are produced traditionally by a 1,10-cyclisation rather than the 1,6-cyclisation path ADS is known to take and have never been reported as products of ADS variants before.

Although changes to the G helix produced dramatic changes in product, G401 variants showed little difference in activity to ADS.

These results show minor changes in sesquiterpene synthases' structures can produce radical changes in the final product. They also support the growing body of literature suggesting the G helix of terpene synthases play a major role in shaping their product. [4–8] This work may also suggest an ancestral link between these enzymes, possibly arising from a gene duplication that occurred in a common ancestor.[9]

Attempts were made throughout this work to generate diffraction-quality crystals of ADS and the variant G401L using commercially sourced screens. Obtaining a high-quality crystal structure of ADS would aid in any future site directed mutagenesis work and allow better informed rationales when designing potential FDP (**40**) analogues for downstream use in artemisinin (**4**) synthesis. A crystal structure of G401L would allow comparison with the ADS structure, which may aid in explaining why a 1,10-ring closure product is formed rather than 1,6 as in the wild type. Clearly there is a precarious balance in which route FDP (**40**) cyclises *via* and minor changes to the wild type enzyme tips this balance in favour of a 1,10-ring closure. Also, although in this work, germacrene B (**71**) and selina-4(15), 7(11)-diene (**101**) are confidently identified as products of the ADS variants G401L and G401Y, obtaining NMR spectra would make this result conclusive. As described in Chapter 2, obtaining NMR spectra was attempted, but ultimately proved unsuccessful.

Variations of germacrene B synthase could be designed to try and generate products of a 1,6-ring closure. Comparison of the crystal structure of any successful variants, wt ADS and ADS G401L/Y may aid in the understanding of how one pathway is favoured over another. The variants generated in this work contribute to our understanding of biocatalytic compound production and may aid in future modification of sesquiterpene cyclases to manufacture valuable terpene analogues.

5.2 Chapter 3

Previous work in the group showed that ADS can accept FDP (**40**) with a hydroxyl group at the C-12 position. This yielded an intermediate present in the natural biosynthesis of artemisinin (**4**), DHAAl (**82**), in a shortened synthesis compared to that which occurs in *Artemisia annua*.[10] An inefficient step in the published chemical synthesis of 12-OH FDP (**105**) is the diphosphorylation of 12-hydroxy farnesol.16 A more recent approach developed within the group has employed two promiscuous kinases, ThiM and IPK, and FDP synthase (FDPS)

to produce FDP (**40**) and analogues.[12] This pathway was therefore utilised with a chemically hydroxylated form of prenol, 4-hydroxy prenol (**106**), to produce 12-hydroxy FDP (**105**) without the inefficient and cumbersome chemical diphosphorylation step. This was then converted to DHAAl (**82**) by ADS.

Although the kinase pathway initially developed within the group was a significant improvement on the chemical diphosphorylation,[13] the kinases were far slower with 4-hydroxy prenol (**106**) than with prenol. Various mutations of IPK were designed to try and improve the affinity of this enzyme for 4-hydroxy prenol (**106**). None of the variants showed improved activity compared to the wild type IPK with the unnatural substrate 4-hydroxy dimethylallyl monophosphate (4-OH DMAMP, **107**). Further work in this project could try screening other residues in the active site, which sit near the hydroxyl group, and combinations of mutations could be designed to improve the affinity for 4-OH DMAMP (**107**). IPK is clearly robust to multiple mutations, as shown by Dr. F. Huynh,[14] with variants of IPK with four changes allowing efficient diphosphorylation of a larger analogue.

Further investigation into the active site of IPK to allow diphosphorylation of unnatural analogues, as was attempted in this work, would be beneficial for the synthesis of unnatural analogues of FDP (**40**). The route described in Chapter 4 begins from a 15-carbon starting material; therefore modular building of analogues from 5 carbon precursors is not possible. Attempts at optimising the slow monophosphorylation step could also be investigated as this would make the synthesis less time consuming.

Three variants of FDPS were designed, and the variant M156D showed dramatic improvement in product formation compared to wild type FDPS, improving the condensation step of 4-hydroxy dimethylallyl diphosphate (4-OH DMADP, **107**) and IDP (**14**). Future analogues of DMADP could be tested with this vastly improved mutant of FDPS to generate other FDP analogues. It is known that FDPS can be changed to produce larger products such as HO-GGDP.13 FDPS has been shown in this work to be amenable to changes in residues to improve turnover of an unnatural analogue of DMADP to generate a desirable FDP analogue. Many FDP analogues can be used to generate valuable compounds such as (*S*)-14,15-dimethylgermacrene D,14 and 9-methyl germacrene B.15 Alternative variants of FDPS may allow improved turnover of five carbon precursors to produce these useful analogues, similar to variants described in this work.

5.3 Chapter 4

The synthesis reported in Chapter 3 was far less laborious and technically difficult than the chemical diphosphorylation previously reported which involved two chemical reactions, multiple lyophilisation steps and ion exchange columns and a product that must be kept anhydrous.16 But the synthesis described in Chapter 3 still involved a chemical hydroxylation and time consuming diphosphorylation tranformations of the 5 carbon precursors.

This chapter detailed attempts to use a P450, CYP124A1, and variants of this enzyme to complete the hydroxylation of FDP (**40**). Although these variants were not able to hydroxylate FDP (**40**), they are believed to be the first reported variants of this enzyme, and the wild type enzyme was able to hydroxylate farnesol (**102**). Concurrent to this work, CYP124A1 and ADS were inserted into a strain of *S. cerevisiae* to try and synthesise DHAAl (**82**) *in vivo*. This work proved unsuccessful due to the lack of expression of CYP124A1 in the organism.

A shorter diphosphorylation pathway than that reported in Chapter 3 was also investigated. Follow up work in the group involved investigating a different kinase, UPK, for the monophosphorylation of farnesol (**102**), and mutating the IPK previously reported to complete the diphosphorylation. These were tested with 12-OH farnesol as a substrate and showed good yield. Surprisingly, the 12-OH FDP (**105**) produced from the enzymatic diphosphorylation gave a different ratio of 11-*R:*11-*S* epimer of DHAAl (**82**) compared to similar assays using the chemically synthesised 12-OH FDP (**105**), 1:1 and 2:7 *R:S* respectively.[12] The 11-*R* epimer is the desired isomer for synthesis to artemisinin (**4**).12 Therefore, a fully enzymatic synthesis, with improved epimeric selectivity from farnesol (**102**) to DHAAl (**82**) was designed.

Previously, Sanofi used a 'semi-synthetic route' to generate artemisinin (**4**) in yeast based largely on work devised by Keasling *et al.*,[18] but this route was economically unviable and discontinued. A shortened route, similar to the one described in this work may be optimised for artemisinin (**4**) precursor synthesis and could be a step towards affordable production of synthetic artemisinin (**4**).

Although none of the variants of CYP124A1 generated were able to hydroxylate FDP (**40**), it is possible that multiple mutations around the active site would allow more space for the diphosphate group as well as positive residues to negate the large concentration of negative charge. If this were successful, a reliable fully enzymatic synthesis starting from the 15 carbon compound farnesol (**102**) to DHAAl (**82**) would be possible.

As enzymatic assays using UPK and a variant of IPK yielded a far higher ratio of the 11-*R* epimer of DHAAl (**82**). Future work could determine if this aberration in *R:S* (**82**) ratio is due to substrate availability. This could involve large scale synthesis of 12-OH FDP (**105**) made by UPK and an IPK variant. Enzymatically synthesised 12-OH FDP (**105**) could then be purified

in the same buffer as the chemically synthesised 12-OH FDP (**105**) (see materials and methods 6.2.3 for chemical synthesis) before lyophilisation. This would yield pure enzymatically synthesised 12-OH FDP (**105**), in the same state as the chemically synthesised 12-OH FDP (**105**). This could then be subjected to the same assay as chemically synthesised 12-OH FDP (**105**) with ADS. This would clarify if any of the components of the kinase assay, see materials and methods 6.1.37 for assay components, or if substrate availability affects the ratio of *R:S* DHAAl (**82**). Unfortunately, this would be difficult to do on a large scale, as UPK expression gave very low yields (see supplementary information 7.3.4.3 for SDS-PAGE image) and purification involves lengthy refolding steps. Also, total purification of the diphosphate from the components of the kinase assay is difficult and time consuming.

5.4 Conclusion

This work aimed to improve our understanding of ADS for its application in biocatalytic production of artemisinin (**4**) and to develop an improved synthetic route to artemisinin (**4**) *via* 12-OH FDP (**105**). Novel products from ADS variants were generated which suggest point mutations in ADS allows it to follow a 1,10-ring closure mechanism. This insight shows there is much more to discover about how ADS shapes it product. This work involved generating 18 variants of ADS by site directed mutagenesis. Although effective in identifying novel products, a more efficient screening system of randomly generated variants would likely have higher throughput. One such example is utilising directed evolution coupled to a colourimetric assay to identify improved activity,[19] or generating a large number of variants of small regions of interest and analysing the product profile by automated GC-MS.[20]

An improved synthetic route to DHAAl (**82**) was designed and optimised giving good yield and selectivity. Although IPK variants were designed they did not have improved affinity. Research and development of chemoenzymatic methods, such as generating protein variants, is continually needed. As this work shows, when the ideal variant is identified improvement can be exceptional, as with the FDPS, but producing these variants is still challenging. In future it is likely computational chemistry will be more regularly employed to allow the design of variants for desired activity such as improved affinity for substrate analogues, or the generation of desired products.[21,22]

Although a one-pot enzymatic synthesis of artemisinin (**4**) was not ultimately successful this work has explored novel chemoenzymatic routes to DHAAl (**82**) and demonstrated the flexibility of several enzymes to be used in synthesis of unnatural terpenoids and their

precursors. Along with bioinformatic searches for novel terpene cyclases, engineering enzymes for modified activity will play a key part in the future of novel terpenoid synthesis.

1 J.-X. Li, X. Fang, Q. Zhao, J.-X. Ruan, C.-Q. Yang, L.-J. Wang, D. J. Miller, J. A. Faraldos, R. K. Allemann, X.-Y. Chen and P. Zhang, *Biochemical Journal*, 2013, **451**, 417–426.

2 I. I. Abdallah, R. van Merkerk, E. Klumpenaar and W. J. Quax, *Scientific Reports*, 2018, **8**, 1–11.

3 I. I. Abdallah, M. Czepnik, R. van Merkerk and W. J. Quax, *Journal of Natural Products*, 2016, **79**, 2455–2463.

4 M. Loizzi, V. González, D. J. Miller and R. K. Allemann, *ChemBioChem*, 2018, **19**, 100–105.

5 Y. Yoshikuni, V. J. J. Martin, T. E. Ferrin and J. D. Keasling, *Chemistry and Biology*, 2006, **13**, 91–98.

6 S. C. Kampranis, D. Ioannidis, A. Purvis, W. Mahrez, E. Ninga, N. A. Katerelos, S. Anssour, J. M. Dunwell, J. Degenhardt, A. M. Makris, P. W. Goodenough and C. B. Johnsona, *Plant Cell*, 2007, **19**, 1994–2005.

7 M. Xu, P. R. Wilderman and R. J. Peters, *Proc Natl Acad Sci U S A*, 2007, **104**, 7387–7401.

8 P. Baer, P. Rabe, K. Fischer, C. A. Citron, T. A. Klapschinski, M. Groll and J. S. Dickschat, *Angewandte Chemie - International Edition*, 2014, **53**, 7652–7656.

9 S. C. Trapp and R. B. Croteau, *Genetics*, 2001, **158**, 811–832.

10 X. Tang, M. Demiray, T. Wirth and R. K. Allemann, *Bioorganic and Medicinal Chemistry*, 2018, **26**, 1314–1319.

11 V. J. Davisson, A. B. Woodside, T. R. Neal, K. E. Stremler, M. Muehlbacher and C. D. Poulter, *Journal of Organic Chemistry*, 1986, **51**, 4768–4779.

12 R. K. Allemann, L. Johnson, A. Dunbabin, J. Benton and R. Mart, *Angewandte Chemie*, 2020, **59**, 8486–8490.

13 M. Demiray, X. Tang, T. Wirth, J. A. Faraldos and R. K. Allemann, *Angewandte Chemie - International Edition*, 2017, **56**, 4347–4350.

14 F. Huynh, PhD, Cardiff University, 2020.

15 M. Nagaki, M. Nakada, T. Musashi, J. Kawakami, N. Ohya, M. Kurihara, Y. Maki, T. Nishino and T. Koyama, *Bioscience, Biotechnology, and Biochemistry*, 2007, **71**, 1657–1662.

16 S. Touchet, K. Chamberlain, C. M. Woodcock, D. J. Miller, M. A. Birkett, J. A. Pickett and R. K. Allemann, *Chemical Communications*, 2015, **51**, 7550–7553.

17 R. P. Brazil and J. G. C. Hamilton, *Memorias do Instituto Oswaldo Cruz*, 2002, **97**, 435–436.

18 C. J. Paddon and J. D. Keasling, *Nature Reviews Microbiology*, 2014, **12**, 355–367.

19 R. Lauchli, K. S. Rabe, K. Z. Kalbarczyk, A. Tata, T. Heel, R. Z. Kitto and F. H. Arnold, *Angewandte Chemie - International Edition*, 2013, **52**, 5571–5574.

20 N. G. H. Leferink, M. S. Dunstan, K. A. Hollywood, N. Swainston, A. Currin, A. J. Jervis, E. Takano and N. S. Scrutton, *Scientific Reports*, 2019, **9**, 11936.

21 P. L. Srivastava, A. M. Escorcia, F. Huynh, D. J. Miller, R. K. Allemann and M. W. van der Kamp, *ACS Catalysis*, 2021, **11**, 1033–1041.

22 S. C. Dodani, G. Kiss, J. K. B. Cahn, Y. Su, V. S. Pande and F. H. Arnold, *Nature Chemistry*, 2016, **8**, 419–425.

6 Chapter 6: Materials and Methods

6.1 Biological

6.1.1 General

Restriction enzymes, unstained protein marker (14.4 kDa to 116 kDa), DNA dye, SYBR safe gel dye, and DNA ladder were obtained from Fisher Scientific. PrimeSTAR HS DNA polymerase (premix) was obtained from TaKaRa. Western blot equipment was obtained from BioRad. Easy Pure® Plasmid MiniPrep Kit was obtained from TransGen Biotech and the dialysis membrane obtained from Medicell membranes Ltd, 12-14,000 Da. The DNA sequencing service used was Eurofins Genomics. PCR was run on either Techne Touchgene Gradient Thermal Cycler or Biometra Thermocycler T-Gradient Thermoblock. Nickel beads used for nickel affinity columns were obtained from Expedion. Oligonucleotide primers and all other biological reagents were obtained from Sigma Aldrich. [1-^{3}H]-FDP was purchased from American Radiolabelled Chemicals, 'Ecoscint^{TM} O' scintillation fluid was obtained from National Diagnostics.

All other chemicals were obtained from Sigma-Aldrich, Fisher Scientific, or Melford. All gel images were visualised using the ChemiDoc MP imaging systems from BioRad. DNA concentration was measured via a nanodrop (Nanodropo 3300 Fluorospectrometer, Thermo scientific). UV/vis spectroscopy was completed on a Shimadzu UV-2600 UV-Vis spectrophotometer. GC-MS analysis was run on a Perkin Elmer Clarus 680 GC fitted with a Perkin Elmer Elite-1 column (30 m x 0.25 mm internal diameter) and a Perkin Elmer Clarus SQ 8 C mass spectrometer. GC-FID analysis was run on an Agilent 7890A fitted with a Restex Rt-betaDEXsm fused silica capillary column (30 m x 0.32 mm ID x 0.25 µm film) and a flame ionisation detector (FID).

6.1.2 Media

6.1.2.1 Luria-Bertani (LB) media

NaCl (10 g L^{-1}), tryptone (10 g L^{-1}) and yeast extract (5 g L^{-1}) were dissolved in deionised water (900 mL), the solution made up to a litre, and autoclaved (121 °C, 20 minutes). Solutions

were allowed to cool to room temperature before use, and antibiotics required added at time of use.

6.1.2.2 LBE media

Yeast extract (5 g L^{-1}), tryptone (10 g L^{-1}), glucose (5 g L^{-1}), NaCl (10 g L^{-1}) were dissolved in deionised water (900 mL), the solution made up to a litre, and autoclaved (121 °C, 20 minutes). Once the solution cooled to room temperature any antibiotics required were added, as was phosphate buffer (100 mL), magnesium sulfate, (2 M, 1 mL), and metals mix (1 mL).

6.1.2.3 TYE media

Yeast extract (10 g L^{-1}), tryptone (16 g L^{-1}), glucose (5 g L^{-1}), NaCl (5 g L^{-1}) were dissolved in deionised water (900 mL), the solution made up to a litre, and autoclaved (121 °C, 20 minutes). Once the solution cooled to room temperature any antibiotics required were added, as was phosphate buffer (100 mL), magnesium sulfate, (2 M, 1 mL), and metals mix (1 mL).

Metals mix (1000X)

50 mM HCl (0.5 mL) and 0.1 M $FeCl_3$-$6H_2O$ (1.35 mg) were dissolved in deionised water (50 mL). 1 M $MnCl_2$-$4H_2O$ (9.9 mg), 1 M $ZnSO_4$-$7H_2O$ (14.4 mg), 0.2 M $CoCl_2$-$6H_2O$ (2.4 mg), 0.2 M $NiCl_2$-$6H_2O$ (2.4 mg) were dissolved in deionised water (50 mL). both solutions were mixed to make the 1000X metals mix solution.

6.1.2.4 Terrific broth (TB) media with 1% glucose

Yeast extract (26.7 g L^{-1}), tryptone (13.3 g L^{-1}), glycerol (4.44 mL L^{-1}), and glucose (9 g, 1% w/v) were dissolved in deionised water (900 mL), the solution made up to a litre, and autoclaved (121 °C, 20 minutes). Once the solution cooled to room temperature any antibiotics required were added, as was phosphate buffer (100 mL).

Phosphate buffer

Potassium phosphate monobasic (0.17 M, 23.14 g L^{-1}), and potassium phosphate dibasic (0.72 M, 125.4 g L^{-1}) were dissolved in deionised water (1 L) and the solution autoclaved (121 °C, 20 minutes).

6.1.2.5 LB agar plates

NaCl (10 g L^{-1}), tryptone (10 g L^{-1}), yeast extract (5 g L^{-1}) and agar (20 g L^{-1}) were dissolved in deionised water and the solution autoclaved (121 °C, 20 minutes). The solution was allowed to cool to body temperature before the appropriate antibiotics were added, and the plates poured under aseptic conditions. Plates were stored at room temperature.

6.1.2.6 TB 1% glucose agar plates

Yeast extract (26.7 g L^{-1}), tryptone (13.3 g L^{-1}), glycerol (4.44 mL L^{-1}), glucose (9 g, 1% w/v), and agar (20 g L^{-1}) were dissolved in deionised water (900 mL), the solution made up to a litre, and autoclaved (121 °C, 20 minutes). The solution was allowed to cool to body temperature before the appropriate antibiotics were added, and the plates poured under aseptic conditions. Plates were stored at room temperature.

6.1.2.7 YPD

Yeast extract (10 g L^{-1}), tryptone (20 g L^{-1}), and glucose (20 g L^{-1}) were dissolved in deionised water (900 mL), the solution made up to a litre, and autoclaved (121 °C, 20 minutes). Solutions were allowed to cool to room temperature before use.

6.1.2.8 YPD agar

Yeast extract (10 g L^{-1}), tryptone (20 g L^{-1}), glucose (20 g L^{-1}) and agar (20 g L^{-1}) were dissolved in deionised water (900 mL), the solution made up to a litre, and autoclaved (121 °C, 20 minutes). The solution was allowed to cool to body temperature, and the plates poured under aseptic conditions. Plates were stored at room temperature.

6.1.2.9 Yeast minimal media for growth

Yeast nitrogen base without amino acids (6.9 g L^{-1}), complete supplement mixture minus tryptophan (0.77 g L^{-1}), and glucose (20 g L^{-1}) were dissolved in deionised water (900 mL), the solution made up to a litre, and autoclaved (121 °C, 20 minutes). Solutions were allowed to cool to room temperature before use.

6.1.2.10 Yeast minimal media for induction

Yeast nitrogen base without amino acids (6.9 g L^{-1}), complete supplement mixture minus tryptophan (0.77 g L^{-1}), galactose (18 g L^{-1}) and glucose (2 g L^{-1}) were dissolved in deionised

water (900 mL), the solution made up to a litre, and autoclaved (121 °C, 20 minutes). Solutions were allowed to cool to room temperature before use.

6.1.2.11 Yeast minimal media agar plates

Yeast nitrogen base without amino acids (6.9 g L^{-1}), complete supplement mixture minus tryptophan (0.77 g L^{-1}), glucose (20 g L^{-1}) and agar (20 g L^{-1}) were dissolved in deionised water (900 mL), the solution made up to a litre, and autoclaved (121 °C, 20 minutes). The solution was allowed to cool to body temperature, and the plates poured under aseptic conditions. Plates were stored at room temperature.

6.1.3 Antibiotics

6.1.3.1 Ampicillin

Ampicillin (100 mg mL^{-1}) was dissolved in deionised water, filtered through a 0.2 μm syringe filter, aliquoted (0.5 mL), and stored at -20 °C.

6.1.3.2 Kanamycin

Kanamycin (50 mg mL^{-1}) was dissolved in deionised water, filtered through a 0.2 μm syringe filter, aliquoted (0.5 mL), and stored at -20 °C.

6.1.3.3 Tetracycline

Tetracycline (10 mg mL^{-1}) was dissolved in deionised water, filtered through a 0.2 μm syringe filter, aliquoted (0.5 mL), and stored at -20 °C.

6.1.3.4 Chloramphenicol

Chloramphenicol (25 mg mL^{-1}) was dissolved in deionised water, filtered through a 0.2 μm syringe filter, aliquoted (0.5 mL), and stored at -20 °C.

6.1.4 Strains of bacteria used

Various *E. coli* cells lines were used during this research. All cloning work and plasmid preparation were completed in XL1-Blue and DH5α cloning strains. Expression cell lines tested included: Rosetta (DE3), ArcticExpress (DE3), C41(DE3), C43(DE3), Bl21-AI and BL21(DE3). BL21(DE3) are an expression cell line with two proteases knocked out decreasing the degradation of the protein expressed. Rosetta (DE3) cells are derived from BL21 expression

cells, but contain additional tRNAs for rarer codons.[1] CYP124A1 was used in Rosetta cells as the gene contains the CGA and CGG codon for Arginine, and the CCC codon for proline. ArcticExpress (DE3) cells express chaperonins from the psychrophilic bacterium *Oleispira antartica* to aid in protein folding. It is believed that expression at lower temperatures aids in the recovery of soluble protein.[2]

C41(DE3) and C43(DE3) are expression cell lines derived from BL21(DE3) cells which allow expression of toxic proteins even when they are unable to be expressed in BL21(DE3).[3]

All assays of CYP124A1, amorphadiene synthase and FDP synthase came from expressions in BL21(DE3) cells. All assays with isopentenyl kinase, hydroxyethylthiazole kinase, and undecaprenol kinase came from expressions in BL21-AI cells.

6.1.5 *E. coli* Competent cell preparation

Competent cell buffer I

Sodium acetate (0.04 g, 10 mM), and NaCl (0.01 g, 5 mM) was dissolved in 45 mL of deionised water and adjusted with acetic acid to a pH of 5.6. Then $MnCl_2$ (0.2 g, 20 mM) was added and the solution made up to 50 mL. The solution was filter sterilised through a 0.2 μm syringe filter before use.

Competent cell buffer II

Sodium acetate (0.04 g, 10 mM), $CaCl_2$ (0.11 g, 20 mM), $MnCl_2$ (0.05 g, 5 mM), glycerol (2.5 g, 5% v/v) was dissolved in deionised water (50 mL). The solution was filter sterilised through a 0.2 μm syringe filter before use.

Method of competent cell preparation

All competent cells were prepared via the same method.

50 μL of the cell line was added to LB media (20 mL), antibiotics were added when required, and left to grow overnight at 37 °C while shaking. The culture (5 mL) was added to LB media (50 mL), antibiotics were added when required, and grown for two hours at 37 °C while shaking. The cells were centrifuged (4000 rpm/ 10 minutes / 4 °C) and the supernatant discarded. The cell pellet was resuspended in competent cell buffer I (20 mL) and left on ice (20 minutes). The cells were centrifuged (4000 rpm/ 5 minutes / 4 °C) and the supernatant

discarded. The cell pellet was gently resuspended in competent cell buffer II (5 mL) and aliquoted (50 μL) before being flash frozen in liquid nitrogen and stored at -80 °C.

6.1.6 Transformation of bacterial competent cells

The desired plasmid and aliquot of competent cells were thawed on ice. Plasmid (1 μL [or 5μL for SDM products]) was added to the competent cells. Cells were incubated on ice (1 hour), then heat shocked in a water bath (45 seconds, 42 °C). Cells were immediately placed back on ice (5 minutes), then appropriate media added (950 μL). Cells were incubated (90 minutes, 37 °C) with shaking, then centrifuged (1 minute, 6000 rpm). The majority of the supernatant was discarded, and the pellet resuspended in ~200 μL. Cells (50 μL) were plated on the appropriate agar plates and incubated overnight (37 °C). A negative control of an aliquot of competent cells which contained no plasmid was subjected to the same protocol and plated on plates with and without the appropriate antibiotic to ensure no microbial contamination alongside each transformation. A positive control of a known plasmid was added to an aliquot of competent cells was subjected to the same protocol and plated on the appropriate antibiotic plate to ensure cell competency.

6.1.7 *Saccharomyces cerevisiae* competent cell preparation

6.1.7.1 Buffer

Lithium acetate (0.33 g, 0.1 M) was dissolved in 45 mL of deionised water and adjusted with acetic acid to a pH of 7.5, before being made up to 50 mL. The solution was filter sterilised through a 0.2 μm syringe filter before use.

6.1.7.2 Protocol

A single colony of the desired strain was added to YPD media (10 mL) and left to grow overnight at 30 °C while shaking. The culture (10 mL) was added to YPD media (40 mL) in a 250 mL flask, and grown for five hours at 30 °C while shaking. The cells were centrifuged (3000 rpm/ 5 minutes / 4 °C) and the supernatant discarded. The cell pellet was resuspended in sterile deionised water (25 mL) and centrifugation repeated. The cell pellet was resuspended in sterile lithium acetate buffer (1 mL) and transferred to a sterile Eppendorf. The cells were

centrifuged (14,500 rpm/ 15 seconds) and the supernatant discarded. The cell pellet was resuspended in sterile lithium acetate buffer (400 μL) and aliquoted (50 μL). The cells were centrifuged (14,500 rpm/ 15 seconds) and the supernatant discarded.[4]

6.1.8 Transformation of *Saccharomyces cerevisiae* competent cells

6.1.8.1 Buffers

PEG 3350 (25 g, 50% w/v) was dissolved in 50 mL of deionised water.

Lithium acetate (3.3 g, 1 M) was dissolved in 45 mL of deionised water and adjusted with acetic acid to a pH of 7.5 before being made up to 50 mL

TE buffer - TRIS base (6.1 g, 1 M) and EDTA (7.3 g, 0.5 M) was dissolved in 45 mL of deionised water and adjusted with acetic acid to a pH of 7.5 before being made up to 50 mL.

Solutions were filter sterilised through a 0.2 μm syringe filter before use.

Salmon sperm DNA (100 mg, 2 mg mL^{-1}) was dissolved in TE buffer (50 mL) and aliquoted (200 μL) before being flash frozen in liquid nitrogen and stored at -80 °C.

6.1.8.2 Protocol

The desired plasmid was thawed on ice. An aliquot of competent cells was placed on ice, and an aliquot of salmon sperm DNA denatured (99 °C, 5 minutes). PEG 3350 (240 μL), sterile lithium acetate buffer (36 μL), salmon sperm DNA (50 μL) and DNA (made up to 34 μL with deionised water) were added to the competent cells in this order. Cells were vortexed briefly before being incubated (30 °C, 30 minutes then 42 °C, 30 minutes). Cells were then centrifuged (15 seconds, 6000 rpm) and resuspended in deionised water (1 mL).

The suspension (200 μL) was spread on yeast minimal media plates and incubated for 3-5 days (30 °C). A negative control of an aliquot of competent cells which contained no plasmid was subjected to the same protocol and plated on YPD and yeast minimal media plates to ensure no microbial contamination alongside each transformation. A positive control of a known plasmid was added to an aliquot of competent cells and was subjected to the same protocol and plated on yeast minimal media plates to ensure cell competency.[4]

6.1.9 *Saccharomyces cerevisiae* culture conditions

All *S. cerevisiae* were grown at 30 °C, plates were grown in a static incubator while cultures were grown shaken. All wild type *S. cerevisiae* were grown in YPD and transformed *S. cerevisiae* in yeast minimal media. All cultures were grown in sterile 250 mL conical flasks. Cultures were grown for 48 hours in non-inducing media 6.1.2.9. Cells were centrifuged (3000 rpm/ 5 minutes / 4 °C) and with aseptic technique the supernatant discarded and replaced with inducing media 0. Cultures grown for compound extraction were overlaid with 5 mL of dodecane once exchanged into inducing media.

6.1.10 Saccharomyces cerevisiae colony PCR

6.1.10.1 Buffers

Lithium acetate (0.66 g, 0.2 M) and SDS (0.5 g, 1% w/v) was dissolved in 50 mL of deionised water.

Ethanol (96 mL) was mixed with deionised water (4 mL) for a 96% solution.

Ethanol (70 mL) was mixed with deionised water (30 mL) for a 70% solution.

6.1.10.2Protocol

A single colony of the desired strain was added to the appropriate media (10 mL) and left to grow to OD 0.4 at 30 °C while shaking. Cells (200 µL) were centrifuged (13,000 rpm/ 3 minutes). Cells were resuspended in lithium acetate SDS buffer and incubated in a heat block (70 °C, 5 minutes). 96% ethanol (300 µL) was added and the solution was vortexed. The solution was centrifuged (15,000 rpm/ 3 minutes). The DNA pellet was washed with 70% ethanol (300 µL), and centrifuged (13,000 rpm/ 3 minutes). The DNA pellet was dissolved in deionised water (100 µL) and centrifuged (15,000 rpm/ 15 seconds). 1 µL of this supernatant was used for PCR.

6.1.10.3PCR protocol

All PCR reactions contained 25 µL of BioMix^{TM} Red, 23.4 µL of de-ionised water, 1 µL of the target DNA, and 0.3 µL of each primer (10 µM concentration).

Table 1: Colony PCR protocol

Step	**Temperature °C)**	**Repeats**	**Length (seconds)**
Initial denaturation	98	1	120
Denaturation	98	30	20
Annealing	55	30	10
Elongation	72	30	60 seconds per kbp of DNA
Final elongation	72	1	600

Table 2: Primers used for colony PCR

Forward 5’ to 3’	Reverse 5’ to 3’
TTACAAGGATGACGACGATAAGGGCCT	ACTAGTATCGATGGATTACAAGGATGAC

6.1.11 Glycerol stock preparation

0.6 mL of an overnight culture was added to a sterile Eppendorf and 0.6 mL of sterile 50% glycerol was added. These were then stored at -80 °C. When stocks were needed for culture the stock were placed on ice, and overnights set up aseptically from the stocks.

6.1.12 Bacterial Plasmid DNA isolation

Method of plasmid miniprep

Appropriate media (15 mL) was inoculated with a single colony of the desired organism, appropriate antibiotic added and left to grow overnight at 37 °C while shaking. Cells were centrifuged (4000 rpm/ 10 minutes / 4 °C) and the supernatant discarded. The DNA was purified using the TransGen Biotech ‘EasyPure ® HiPure Plasmid MiniPrep Kit’ (TransGen, Beijing) following the protocol provided by the manufacturers.

6.1.13 1% Agarose gel preparation

6.1.13.1 TAE buffer (50X)

Tris-Base (242 g, 2 M), glacial acetic acid (57.1 mL, 1 M), and disodium EDTA (100 mL, 5 mM) were added to deionised water (900 mL), and the pH adjusted to 8 before the total volume (1 L) was made up with deionised water.

6.1.13.2Agarose gel

50X TAE buffer (1 mL) was dissolved in deionised water (49 mL), agarose powder (0.5 g) and SYBR safe dye (7 μL) added. The solution was warmed until the agarose powder had dissolved and then the gel was cast. DNA samples (5 μL) were mixed with DNA dye (2 μL) and placed into the gel wells. DNA ladder (5 μL) was run on each gel to identify the size of DNA samples. Electrophoresis was run at a constant amplitude of 85 mA for 40 minutes.

6.1.14 Gel extraction of DNA

DNA samples were extracted from the agarose gel and 3 volumes of QG buffer (QIAGEN) added to 1 volume of gel (100 mg gel ~ 100 μL). samples were incubated at 50 °C for 10 minutes, vortexing every 3 minutes. Isopropanol (1 gel volume) was added and the sample inverted 5 times. The solution was added to a QIAquick column and centrifuged (1 minute, 13,000 rpm). QG buffer (500 μL) was added to the column and the sample was centrifuged (1 minute, 13,000 rpm). PE buffer (750 μL) was added to the column, left to stand for 5 minutes and the sample was centrifuged (1 minute, 13,000 rpm). Centrifugation was repeated. The column was placed into a clean eppendorf, sterile water (30 μL) added, the column was left for 1 minute and centrifuged (1 minute, 13,000 rpm).

6.1.15 Site directed mutagenesis

Mutations of all enzymes were completed via primers containing the desired mutation and a site directed mutagenesis PCR protocol. Primers are designed to have 15-20 base pairs which overlap at the 5’ end, and then a further 20-25 base pairs of overhang at the 3’ end which are not complimentary to try and avoid primer dimerization. The desired codon to mutate is positioned within the complimentary region of the primers to increase mutation efficiency.[5]

All PCR reactions contained 25 μL of PrimeSTAR pre-mix, 20.9 μL of de-ionised water, 0.5 μL of the target DNA, 3 μL of DMSO, and 0.3 μL of each primer (10 μM concentration).

The PCR protocol used was standardised apart from the annealing temperature:

Table 3: SDM PCR protocol

Step	**Temperature (°C)**	**Repeats**	**Length (seconds)**
Initial denaturation	98	1	120
Denaturation	98	30	20
Annealing	Variable 50-70	30	30
Elongation	72	30	60 seconds per kbp of DNA
Final elongation	72	1	600

Table 4: Primers for site directed mutagenesis with the changed bases in lower case

Plasmid	Mutation	Primer		Annealing temperature (°C)
		Forward 5'-3'	**Reverse 5'-3'**	
ADS	I295A	GCT GTT gcg ACC CTG ATC GAT GAC ACT TAC GAT GCT TAC G	CAG GGT cgc AAC AGC TAC AGC TTT AGT GAA GAA CAC ACG	65
ADS	I295F	GCT GTT tTt ACC CTG ATC GAT GAC ACT TAC GAT GCT TAC G	CAG GGT aAa AAC AGC TAC AGC TTT AGT GAA GAA CAC ACG	65
ADS	D299E	C CTG ATC GAa GAC ACT TAC GAT GCT TAC GGC ACC TAC G	GT GTC tTC GAT CAG GGT GAT AAC AGC TAC AGC TTT AGT G	65
ADS	L374A	CGT AAC gcG ATG GTT GAA GCT AAA TGG GCT AAC GAA GGC	C AAC CAT Cgc GTT ACG TAC GAA TTC TTT AAC AAA CTC	60
ADS	E377A	G ATG GTT Gcg GCT AAA TGG GCT AAC GAA GGC CAT ATC CCG	CCA TTT AGC cgC AAC CAT CAG GTT ACG TAC GAA TTC TTT AAC	65
ADS	I397Y	GCC GGT GAT ata AAC AAC CGG GTC ATG TTC TTC GGT AGT CGG	G GTT GTT tat ATC ACC GGC GGT GCA AAC CTG CTG ACC ACC	65
ADS	G401L	C ACC GGC ctg GCA AAC CTG CTG ACC ACC ACT TGC TAT C	G GTT TGC cag GCC GGT GAT GAT AAC AAC CGG GTC ATG TTC	65
ADS	G401Y	C ACC GGC taT GCA AAC CTG CTG ACC ACC ACT TGC TAT CTG G	GTT TGC Ata GCC GGT GAT GAT AAC AAC CGG GTC ATG TTC	65
ADS	A402L	GGC GGT ctg AAC CTG CTG ACC ACC ACT TGC TAT CTG GG	CAG GTT cag ACC GCC GGT GAT GAT AAC AAC CGG GT	65
ADS	A402Y	GGC GGT tat AAC CTG CTG ACC ACC ACT TGC TAT CTG GG	CAG GTT ata ACC GCC GGT GAT GAT AAC AAC CGG GTC	70
ADS	N403S	GGT GCA AgC CTG CTG ACC ACC ACT TGC TAT CTG GGT ATG TCC	GGT CAG CAG GcT TGC ACC GCC GGT GAT GAT ACC AAC CGG GTC	65
ADS	L405A	GCA AAC CTG gcG ACC ACC ACT TGC TAT CTG GGT ATG TCC GAC	GGT GGT Cgc CAG GTT TGC ACC GCC GGT GAT GAT AAC AAC C	65
ADS	L405F	CA AAC CTG tTt ACC ACC ACT TGC TAT CTG GGT ATG TCC GAC	GGT GGT aAa CAG GTT TGC ACC GCC GGT GAT GAT AAC AAC C	60

ADS	N443I	CGT CTG Att GAC CTG ATG ACC CAC AAA GCA GAG CAG G	CAG GTC aaT CAG ACG ACG ACC CAG AAT ACC GGA GTA ACG	65
ADS	N443L	CGT CTG ctg GAC CTG ATG ACC CAC AAA GCA GAG CAG G	CAG GTC cag CAG ACG ACG ACC CAG AAT ACC GGA GTA ACG	65
ADS	N524D	GT AAA GAT gAt TTC ACT CGC ATG GGC GAC GAA TAC AAA CAC C	CG AGT GAA aTc ATC TTT ACC AGC GTA CTG TAC TTC CAG GAA CTG	65
ADS	T526D	GAT AAC TTC gaT CGC ATG GGC GAC GAA TAC AAA CAC CTG ATC	CAT GCG Atc GAA GTT ATC TTT ACC AGC GTA CTG TAC TTC C	65
ADS	T526L	GAT AAC TTC ctg CGC ATG GGC GAC GAA TAC AAA CAC CTG ATC	CAT GCG cag GAA GTT ATC TTT ACC AGC GTA CTG TAC TTC CAG	65
CYP124A1	T101S	C ATC tct ATC AAC GAC CAG ACA CCA GAG TTA GCC G	GTC GTT GAT aga GAT GTT GGG GTA CGA ACT GAA AAT G	65
CYP124A1	N103S	ATC tcg GAC CAG ACA CCA GAG TTA GCC GAA TAC TTC GGC	GG TGT CTG GTC cga GAT CGT GAT GTT GGG GTA CGA ACT G	60
CYP124A1	D104S	CG ATC AAC agC CAG ACA CCA GAG TTA GCC GAA TAC TTC G	GT CTG Gct GTT GAT CGT GAT GTT GGG GTA CGA ACT G	70
CYP124A1	F416R	TCG CAG cgC ATT CAC GGA ATC AAG ACG CTG CCA GTT	GTG AAT Gcg CTG CGA CAA CAG CCG TGC GGG	70
CYP124A1	Q421L	G TTG TCG CtG TTC ATT CAC GGA ATC AAG ACG CTG CC	G AAT GAA CaG CGA CAA CAG CCG TGC GGG CTC CTC G	50
CYP124A1	I423T	G CAG TTC Acc CAC GGA ATC AAG ACG CTG CCA G	CC GTG ggT GAA CTG CGA CAA CAG CCG TGC G	65
FDPS	M156D	GAA GGG gat GTC GCC GGT CAG GCA GCC GAT ATG GAA GG	C GGC GAC atc CCC TTC CGG ACC GGC CGC TTT CG	60
FDPS	M156S	GAA GGG Agc GTC GCC GGT CAG GCA GCC GAT ATG GAA GG	GC GAC gcT CCC TTC CGG ACC GGC GCG TTT C	60
FDPS	M156T	GAA GGG Acc GTC GCC GGT CAG GCA GCC GAT ATG GAA GG	C GGC GAC ggT CCC TTC CGG ACC GGC CGC TTT CG	65
IPK	G59T	GCA TTT acT CAT CCG GTT GCC AAA AAG TAT CTG	CGG ATG Agt AAA TGC ACC ACC ACC ATG AAC CAG AAT C	65
IPK	M90S	CGT GCA tcG CGT CGT TTT AAC AAC ATT ATT ATC GAT ACC CTG C	CG ACG Cga TGC ACG CTG AAT TTC CCA AAA GCC TTT TTC C	65

IPK	G144S	G ATT CAT tcT GAT ATT GTG ATC GAT GAT AAA AAC GGC	C AAT ATC Aga ATG AAT CAC CGG AAC CAG ATT ACG	70
IPK	D145S	CAT GGT tcT ATT GTG ATC GAT GAT AAA AAC GGC	CAC AAT Aga ACC ATG AAT CAC CGG AAC CAG	55
IPK	I156T	C TAT CGC AcT ATT AGC GGT GAT GAT ATT GTT C	CT AAT AgT GCG ATA GCC GTT TTT ATC ATC G	70

Once the DNA was replicated via PCR it was digested with the restriction enzyme *DpnI* (1 μL, 2 hours, 37 °C). DpnI is a restriction enzyme which recognises N6-methyladenine (found in almost all DNA obtained from *E. coli*) in the recognition sequence GmATC.[6] The DNA was then transformed into cloning cell lines 6.1.6, miniprepped 6.1.12 and confirmed by sequencing.

6.1.16 Golden gate assembly

Golden gate assembly is a one pot cloning method that utilises a type IIs restriction enzyme that with proper design of restriction sites can remove the recognition site once ligation is complete.[7]

Insert DNA (1 μL, 100 ng / μL), plasmid DNA (1 μL, 50 ng / μL), BSA (1.5 μL), BsaI-HF®v2 (0.5 μL), T4 DNA Ligase Reaction Buffer (1.5 μL) ,T4 DNA ligase (0.5 μL) (all four from NEB®), and DTT (0.1 μL, 0.1 M) were combined and the deionised water added (8.9 μL).

Solutions were incubated at 37 °C for an hour and then subjected to the following protocol:

Temperature (°C)	Time (minutes)	Repeats
37	3	26
16	4	
50	5	
80	5	

6.1.17 Primers

Table 5: Primers for golden gate cloning, the BsaI sites are underlined, the sequence corresponding to the plasmid are in lower case and the sequence in capitals are the overhangs

	Primer	
Plasmid	Forward	Reverse
GBS	GCT GAA GGT CTC ggt atg gct gct tct tct gc	G CTG AAG GTC TCG gag cag ccg gat ctc agt gg
pET-32 Xa LIC	G CTG AAG GTC TCC gct cta act ctc ctc tgg cc	GCT GAA GGT CTC CAT acc ctc aat acc gga gcc

6.1.18 Sodium dodecyl sulfate – polyacrylamide gel electrophoresis

6.1.18.1Resolving buffer

Tris-Base (27.3 g, 1.5 M) was dissolved in deionised water (80 mL), the pH adjusted to 8.8 with HCl, and the final volume made to 150 mL.

6.1.18.2Stacking buffer

Tris-Base (6.1 g, 0.5 M) was dissolved in deionised water (60 mL), the pH adjusted to 6.8 with HCl, and the final volume made to 100 mL.

6.1.18.3Running buffer

Tris-Base (30.3 g, 0.25 M), glycine (144 g, 1.92 M) and SDS (10 g, 10% w/v) was dissolved in deionised water (1 L). Running buffer was diluted ten fold with deionised water before use.

6.1.18.4 Protein sample buffer

β-ME (500 μL, 7.1 mM), glycerol (2.5 mL, 34 mM), stacking buffer (1.25 mL), SDS (2 mL of 10% w/v solution), and bromophenol blue (0.2 mL of 0.6% w/v solution) was dissolved in deionised water (3.55 mL) and stored at 4 °C.

6.1.18.5 Resolving gel

Resolving buffer (2 mL), deionised water (4 mL), 10% w/v SDS (0.2 mL), and 30% bis acrylamide (3.7 mL) were added together. Prior to pouring the gel 10% APS (30 μL, 100 mg mL^{-1} in dH_2O) and TEMED (30 μL) were added to the solution.

6.1.18.6 Stacking gel

Stacking buffer (0.6 mL), deionised water (2 mL), 10% w/v SDS (60 μL), and 30% bis acrylamide (0.5 mL) were added together. Prior to pouring the gel 10% APS (30 μL, 100 mg mL^{-1} in dH_2O) and TEMED (30 μL) were added to the solution.

6.1.18.7 SDS-PAGE gel stain

Coomasie brilliant blue R-250 (80 mg) was dissolved in ethanol (10 mL). Concentrated HCl (2 mL) was added and made up to 500 mL with deionised water.

6.1.18.8 Method of SDS-PAGE gel preparation

Initially the resolving gel was poured into the gel stand, overlaid with isopropanol (1 mL) and left to polymerise. The isopropanol was drained off and the stacking gel poured on top of the resolving gel, with the combs to create wells added immediately. The stacking gel was left to polymerise. All proteins samples were prepared in a 1:1 ratio with protein sample buffer, and incubated for 10 minutes at 80 °C. 7 μL of sample was loaded onto the gel, running buffer added, and the gel run for 45 minutes at 180 V.

6.1.18.9 Method of SDS-PAGE gel visualisation

Gels were removed from the plates into deionised water and microwaved for 2 minutes. This was repeated twice more. The water was then replaced with SDS-PAGE gel stain, microwaved for 2 minutes and left to cool. The stain was replaced with water and bands were visible.

6.1.19 *Saccharomyces cerevisiae* protein extraction

Appropriate media (15 mL) was inoculated with a single colony of the desired organism, left to grow overnight at 30 °C while shaking. Cells (1.5 mL) were centrifuged (13,000 rpm/ 3 minutes) and the supernatant discarded. The cells were resuspended in lithium acetate buffer (500 μL, 2 M, pH 7.5) and left on ice for five minutes. The solution was centrifuged (13,000 rpm/ 3 minutes) and the supernatant discarded. The cells were resuspended in NaOH buffer (500 μL, 0.4 M) and left on ice for five minutes. The solution was centrifuged (13,000 rpm/ 3 minutes) and the supernatant discarded. The cells were resuspended in protein sample buffer 6.1.18.4 (100 μL) and incubated in a heat block (100 °C, 5 minutes). The solution was centrifuged (13,000 rpm/ 3 minutes) and the supernatant loaded onto the protein gel, 6.1.18.

6.1.20 Western blot preparation

6.1.20.1 Buffers

Trans-Blot Turbo Buffer – prepared according to manufacturers instructions

TBST buffer – TRIS base (3 g, 25 mM), NaCl (8.76 g, 150 mM) and Tween 20 (0.05% v/v) were dissolved in deionised water (900 mL) and adjusted with hydrochloric acid to a pH of 7.6 before being made up to 1 L.

Blocking buffer – BSA (1.25 g) dissolved in 50 mL of TBST

Probe buffer – the appropriate volume of probe according to the manufacture was placed in 10 ml of TBST
Buffers were stored at 4 °C.

An SDS-PAGE gel was run with Prestained NEB ladder (P77085), and if needed a sample of protein with the tag of interest, in this work either the FLAG or MYC tag. The MYC tag antibody used was obtained from Proteintech and was the monoclonal mouse antibody. The FLAG tag antibody used was obtained from Biorad and was a monoclonal rat antibody. The kit used was Biorad Trans-Blot® Turbo™ RTA Transfer kit, LF PVDF mini size. Substrates used for imaging came from the Clarity™Western ECL substrate kit. Transfers were done using the Trans-Blot Turbo Transfer system using the mixed molecular weight setting.

Once an SDS-PAGE gel was run it was placed in water rather than stained.
One PVDF membrane per SDS-PAGE gel was placed in 100% methanol until translucent.
The PVDF membrane was transferred into Trans-Blot Turbo Buffer (30 mL) and left for 3 minutes.
Two transfer stacks per SDS-PAGE gel were placed in Trans-Blot Turbo Buffer (20 mL) and left for 3 minutes.
Components were assembled according to the manufactures instructions in the systems tray. Excess buffer was removed before the transfer began.
Once the transfer was complete the PVDF membrane was washed in TBST buffer, before incubation in blocking buffer (1 hour/ room temperature/ with shaking).
The membrane was washed with TBST (10 minutes/ 15 mL room temperature/ with shaking) twice.
The membrane was incubated in probe buffer (1 hour/ room temperature/ with shaking). The membrane was washed with TBST (10 minutes/ 15 mL room temperature/ with shaking) four times.
The membrane was imaged according to the manufactures instructions, incubating with the substrates provided (5 minutes) before imaging.

6.1.21 CYP124A1 expression

The CYP124A1 gene was in the pCWori plasmid, which contains a tac promoter. A single colony of BL21(DE3) containing the CYP124A1 plasmid was inoculated into 20 mL of media (TB 1% glucose) along with ampicillin stock solution (20 μL) and grown overnight with shaking (37 °C). The overnight was added to flasks of media (500 mL, TB 1% glucose), and ampicillin stock solution (500 μL) added. Flasks were grown at 37 °C until optical density reached 1.2 (OD at 600 nm) and then IPTG (120 mg), $FeCl_3$ (250 μM, 41 mg) and 5-aminolevulinic acid (ALA, 70 mg) were added to induce protein expression and supplement heme production respectively. The flasks were incubated for 48 hours at 25 °C. Cells were harvested by centrifugation (5000 rpm, 20 minutes, 4 °C). Cell pellets were used or stored at -20 °C.

6.1.22 CYP124A1 purification

Cell lysis buffer I

Tris-Base (6.1 g, 50 mM), NaCl (29.2 g, 0.5 M), imidazole (1.36 g, 20 mM), PMSF (0.17 g, 1 mM), and glycerol (10% w/v) were dissolved in deionised water (900 mL), the pH adjusted to 7.5 with HCl, and made up to a litre.

6.1.22.1 Sonication

Pellets were defrosted on ice and cell lysis buffer I (~100 mL), and lysozyme (10 mg) were added. Cells were resuspended with stirring at 4 °C. Cells were placed in an ice bath and sonicated for 5 minutes, with 5 second pulses and 20 second rests. Lysed cells were centrifuged (18,000 rpm/ 40 minutes/ 4 °C) and supernatant kept. Supernatant was passed through a 0.2 μm syringe filter before use.

6.1.22.2 Ni^{2+}-NTA affinity column protocol I

Ni^{2+} affinity resin was equilibrated with cell lysis buffer I. Supernatant was added and the resin was left to bind while mixing (4 °C, 1 hour). The protein was eluted with an imidazole gradient (20 mM, 50 mM, 100 mM, 250 mM, 500 mM; 50 mL) dissolved in cell lysis buffer I. The protein was obtained in the 100 mM fraction, (MW 48647 Da).

6.1.22.3 Dialysis of protein

The protein obtained was dialysed in dialysis tubing with a 14 kDa membrane, in 3 L of dialysis buffer overnight. Dialysis buffer I contained Tris base (7.3 g, 20 mM), glycerol (300 g, 10% w/v) NaCl (1.8 g, 10 mM), pH 7.7.

6.1.22.4 Concentration of protein

Once the protein was dialysed it was placed in a spin column with a molecular cut off of 10 kDa and concentrated to 2 mL by centrifugation.

6.1.23 CYP124A1 assays

Assays were completed in a final volume of 4 mL, with a concentration of 0.5 μM of the appropriate CYP and 0.1 mM of the appropriate substrate. The final volume was made up with buffer (40 mM [3-(N-morpholino) propanesulfonic acid], 50 mM NaCl, 5 mM $MgCl_2$ pH 7.2). Control assays did not contain CYP124A1. All other components were kept constant.

All components of the assay but NADPH were added together and left to incubate at room temperature for 10 minutes. Then NADPH was added to initiate the reaction, which was left shaking at room temperature overnight.

If the assay involved FDP, the assay was treated with alkaline phosphatase for four hours at 37 °C to remove the diphosphate group, before extraction with pentane allowing mass spectrometry analysis for any hydroxylated product. If the assay contained farnesol, the assay was extracted immediately with pentane for mass spectrometry analysis. 12-OH farnesol was synthesised as a standard for the results of the assay see 6.2.4.

Table 6: CYP124A1 assay components

Component	**Concentration**
NADPH	0.5 mM
Ferredoxin	0.005 μM
Ferredoxin reductase	0.1 U
Catalase	0.042 nM
Glucose-6-phosphate	1 mM
Glucose-6-phosphate dehydrogenase	2.4 U

6.1.24 Amorpha-4,11-diene synthase (ADS) expression

A single colony of BL21(DE3) containing the pET-21d ADS plasmid was inoculated into 20 mL of media (LB) along with ampicillin stock solution (20 μL) and grown overnight with shaking (37 °C). The overnight was added to flasks of media (500 mL, LB) and ampicillin stock solution (500 μL) added. Flasks were grown at 37 °C until optical density reached 0.6 (OD at 600 nm) and then IPTG (120 mg) was added to induce protein expression. The flasks were incubated for 6 hours at 20 °C. Cells were harvested by centrifugation (5000 rpm, 20 minutes, 4 °C). Cell pellets were used or stored at -20 °C.

6.1.24 ADS purification

6.1.25.1 Cell lysis buffer II

Tris-Base (6.1 g, 50 mM), NaCl (29.2 g, 0.5 M), β-ME (1.4 mL, 20 mM), $MgCl_2$ (0.5 g, 5 mM), and glycerol (10% w/v) were dissolved in deionised water (900 mL), the pH adjusted to 8 with HCl, and made up to a litre.

6.1.25.2 Sonication

Using cell lysis buffer II, and the same protocol as CYP124A1

6.1.25.3 Ni^{2+}-NTA affinity column protocol II

Ni^{2+} affinity resin was equilibrated with cell lysis buffer II. Supernatant was added and allowed to flow through under gravity. The protein was eluted with an imidazole gradient (20 mM, 50 mM, 100 mM, 250 mM, 500 mM; 50 mL) dissolved in cell lysis buffer II. The protein was obtained in the 100 mM fraction (MW 66496 Da).

6.1.25.4 Dialysis of protein

The protein obtained was dialysed in dialysis tubing with a 14 kDa membrane, in 4 L of dialysis buffer overnight. Dialysis buffer II contained HEPES (26 g, 25 mM), NaCl (23.4 g, 100 mM), and DTT (618 mg, 1 mM), pH 7.5.

6.1.25.5 Concentration of protein

Once the protein was dialysed it was placed in a spin column with a molecular cut off of 10 kDa and concentrated to 2 mL by centrifugation.

6.1.26 ADS activity assays

6.1.26.1 Assay buffer

HEPES (20 mM), $MgCl_2$ (5 mM), and DTT (1 mM) were dissolved in deionised water and the pH adjusted to 7.5. Amorphadiene synthase (1 μM) and FDP (0.4 mM) were added and the solution overlaid with 1 mL of pentane. The incubations were left to gently shake overnight at room temperature. The next day, the assays were vortexed, the organic layer was extracted and analysed via GC-MS.

6.1.27 Steady state kinetics

Optimal conditions for measuring wild type amorphadiene synthase kinetics have already been determined by Dr. M. Demiray. ADS was used at a concentration of 100 nM with an incubation time of 10 minutes. The concentration of [1-^{3}H]-FDP (240000 dpm $nmol^{-1}$) was varied. Reaction mixtures contained buffer A and B, ADS at the appropriate concentration and the appropriate concentrations of FDP.

6.1.27.1 Incubation buffer A

HEPES (20 mM), and DTT (1 mM) were dissolved in deionised water and the pH adjusted to 7.5.

6.1.27.2 Incubation buffer B

HEPES (20 mM), DTT (1 mM) and $MgCl_2$ (10 mM) were dissolved in deionised water and the pH adjusted to 7.5.

6.1.27.3 Protocol

Incubation buffer A was used to dilute the enzyme to an appropriate concentration, final concentration 100 nM. The volume of enzyme and FDP needed was calculated, and incubation buffer A used to make up the total volume to 125 μL. A range of FDP concentrations, from

0.02 to 15 μM were used. 125 μL of incubation buffer B made up the total volume of 250 μL for the assays.

The assay solution without enzyme were placed on ice, ADS was added to initiate the reaction and overlaid with hexane (1 mL). The assays were incubated at 30 °C for 10 minutes before quenching with EDTA (50 μL, 0.5 M), vortexing and cooling on ice. The hexane overlay, along with two extractions (1 mL) of hexane:diethylether (11:1) were passed through a small silica column in a Pasteur pipette (~1 g). A final addition of hexane:diethylether solution (1 mL) was passed through the column. All extracts were added to Ecoscint™ O fluid (15 mL) with vortexing. Radioactivity was detected via a scintillation counter (Packard 2500 TR™) in ^{3}H mode for 4 min per sample. Background radioactivity was obtained by measuring samples of the range of FDP concentrations without the addition of enzyme. These were measured and subtracted from the data obtained for enzymatic assays.

Data obtained was fitted to the Michaelis-Menten equation, Equation 1, using Systat Sigmaplot.

$$Equation\ 1: V = \frac{V_{max}[S]}{K_M + [S]}$$

The turnover, k_{cat} was derived from the calculated V_{max} using Equation 2

$$Equation\ 2: \frac{V_{max}}{t.c.A}$$

Where V_{max} is counts per minute (cpm), t is the incubation time in seconds, c is the enzyme concentration in μM and A is the activity of FDP in cpm μM^{-1}.

Catalytic efficiency was calculated by dividing the value obtained for the k_{cat} over the value obtained for the K_M.

6.1.27.4

Error bars in this work are calculated as the standard deviation (sd) of a sample, using

$$sd = \sqrt{\frac{\Sigma(X - M)^2}{(n - 1)}}$$

Where X = each value measured in the sample, M = mean of the sample, and n = sample size.

Error of the K_M and k_{cat} is calculated as the standard error of the mean using

$$SE = sd/\sqrt{n}$$

Where SE = Standard error sd = standard deviation n = sample size

The propagation of the errors for catalytic efficiency was calculated as:

$$EC = C\sqrt{\left(\frac{EX}{X}\right)^2 + \left(\frac{EY}{Y}\right)^2}$$

Where EC = error of catalytic efficiency, C = the catalytic efficiency, X = k_{cat}, EX = error of k_{cat}, Y = K_M EY = error of K_M.

6.1.28 IPK and IPK variants expression

A single colony of BL21-AI containing the isopentenyl phosphate kinase (IPK) gene from *Methanocaldococcus jannaschii* inserted into the pET-28a vector previously made by Dr. L. A. Johnson was inoculated into 20 mL of media (LB) along with kanamycin stock solution (20 μL) and grown overnight with shaking (37 °C). The overnight was added to flasks of media (500 mL, LB) and kanamycin stock solution (500 μL) added. Flasks were grown at 37 °C until optical density reached 0.8 (OD at 600 nm) and then arabinose (2 g) was added to induce protein expression. The flasks were incubated for 4 hours at 37 °C. Cells were harvested by centrifugation (5000 rpm, 20 minutes, 4 °C). Cell pellets were used or stored at -20 °C.

6.1.29 IPK purification

6.1.29.1 Cell lysis buffer III

Tris-Base (6.1 g, 50 mM), NaCl (17.5 g, 300 mM), and TCEP (143 mg, 0.5 mM) were dissolved in deionised water (900 mL), the pH adjusted to 8 with HCl, and made up to a litre.

6.1.29.2 Sonication

Using cell lysis buffer III, and the same protocol as CYP124A1

6.1.29.3 Ni^{2+}-NTA affinity column protocol III

Ni^{2+} affinity resin was equilibrated with cell lysis buffer III. Supernatant was added and the protein was eluted with an imidazole gradient (20 mM, 50 mM, 100 mM, 250 mM, 500 mM; 50 mL) dissolved in cell lysis buffer III. The protein was obtained in the 250 mM fraction, (MW 32648 Da).

6.1.29.4 Dialysis of protein

The protein obtained was dialysed in dialysis tubing with a 14 kDa membrane, in 4 L of dialysis buffer overnight. Dialysis buffer III contained TRIS (48.5 g, 100 mM), pH 8.

6.1.29.5 Concentration of protein

Once the protein was dialysed it was placed in a spin column with a molecular cut off of 10 kDa and concentrated to 10 mL by centrifugation. The protein had a tendency to precipitate.

6.1.30 ThiM expression

A single colony of BL21-AI containing the hydroxyethylthiazole kinase (ThiM) gene from *E. coli* inserted into the pET28a vector previously made by Dr. L. A. Johnson was inoculated into 20 mL of media (LB) along with kanamycin stock solution (20 μL) and grown overnight with shaking (37 °C). The overnight was added to flasks of media (500 mL, LB) and kanamycin stock solution (500 μL) added. Flasks were grown at 37 °C until optical density reached 0.8 (OD at 600 nm) and then arabinose (2 g) was added to induce protein expression. The flasks were incubated for 4 hours at 37 °C. Cells were harvested by centrifugation (5000 rpm, 20 minutes, 4 °C). Cell pellets were used or stored at -20 °C.

6.1.31 Thim purification

Via the same protocol as IPK expression (MW 27339 Da), this protein did not have the same tendency to precipitate.

6.1.32 FDP Synthase expression

A single colony of BL21(DE3) containing the FDP synthase gene from *Bacillus stearothermophilus*, cloned into a pET-28a(+) plasmid, with a GB1 solubility tag was inoculated into 20 mL of media (LB) along with kanamycin stock solution (20 μL) and grown overnight with shaking (37 °C). The overnight was added to flasks of media (500 mL, LB) and kanamycin stock solution (500 μL) added. Flasks were grown at 30 °C until optical density reached 0.8 (OD at 600 nm) and then IPTG (120 mg) was added to induce protein expression. The flasks were incubated for 3 hours at 30 °C. Cells were harvested by centrifugation (5000 rpm, 20 minutes, 4 °C). Cell pellets were used or stored at -20 °C.

6.1.33 FDP Synthase purification

6.1.33.1 Cell lysis buffer IV

Tris-Base (3.0 g, 25 mM), disodium EDTA (0.34 g, 1 mM), β-ME (0.7 mL, 10 mM), PMSF (0.02 g, 0.1 mM), were dissolved in deionised water (900 mL), the pH adjusted to 7.7 with HCl, and made up to a litre.

6.1.33.2 Sonication

Using cell lysis buffer IV, and the same protocol as CYP124A1

6.1.33.3 Ni^{2+}-NTA affinity column protocol IV

Ni^{2+} affinity resin was equilibrated with cell lysis buffer IV. Supernatant was added and the protein was eluted with an imidazole gradient (20 mM, 50 mM, 100 mM, 250 mM, 500 mM; 50 mL) dissolved in cell lysis buffer IV. The protein was obtained in the 100 and 250 mM fraction (MW 39197 Da).

6.1.33.4 Dialysis of protein

The protein obtained was dialysed in dialysis tubing with a 14 kDa membrane, in 4 L of dialysis buffer overnight. Dialysis buffer IV contained TRIS (48.5 g, 100 mM), glycerol (300 g, 10% w/v) NaCl (1.8 g, 10 mM), pH 7.7.

6.1.33.5 Concentration of protein

Once the protein was dialysed it was placed in a spin column with a molecular cut off of 10 kDa and concentrated to 10 mL by centrifugation.

6.1.34 UPK expression

A single colony of BL21-AI containing the undecaprenol kinase (UPK) gene from *Streptococcus mutans* inserted into the pET32 vector previously made by Dr. F. Huynh, was inoculated into 20 mL of media (LB) along with ampicillin stock solution (20 μL) and grown overnight with shaking (37 °C). The overnight was added to flasks of media (500 mL, LB) and ampicillin stock solution (500 μL) added. Flasks were grown at 37 °C until optical density reached 0.6 (OD at 600 nm) and then IPTG (120 mg) was added to induce protein expression. The flasks were incubated for 20 hours at 16 °C. Cells were harvested by centrifugation (5000 rpm, 20 minutes, 4 °C). Cell pellets were used or stored at -20 °C.

6.1.35 UPK purification

6.1.35.1 Cell lysis buffer V

Tris-Base (6.1 g, 50 mM), NaCl (29.2 g, 500 mM), Tween (10 g, 1%), TCEP (0.03 g, 0.1 mM) and glycerol (100 g, 10% w/v) were dissolved in deionised water (900 mL), the pH adjusted to 8 with HCl, and made up to a litre.

6.1.35.2 Sonication

Using cell lysis buffer V, and the same protocol as CYP124A1

6.1.35.3 Basic extraction

The insoluble fraction obtained from centrifugation was resuspended in cell lysis buffer V (60 mL) and stirred at 4 °C for 30 minutes. NaOH (5 M) was added dropwise to the stirring solution until pH 12 was reached. The solution was left stirring for 30 minutes at 4 °C. Hydrochloric

acid (pH 6) was added dropwise until the solution reached pH 8. The solution was left stirring for 30 minutes at 4 °C. The solution was centrifuged (18,000 rpm/ 40 minutes/ 4 °C) and supernatant kept. Supernatant was passed through a 0.2 μm syringe filter before use.

6.1.35.4 Ni^{2+}-NTA affinity column protocol V

Ni^{2+} affinity resin was equilibrated with cell lysis buffer V. Supernatant was added and the protein was eluted with an imidazole gradient (20 mM, 50 mM, 100 mM, 250 mM, 500 mM; 50 mL) dissolved in cell lysis buffer V. The protein was obtained in the 250 mM fraction but with very low concentrations (MW 31,819 Da).

6.1.35.6 Dialysis of protein

The protein obtained was dialysed in dialysis tubing with a 14 kDa membrane, in 4 L of dialysis buffer overnight. Dialysis buffer V contained TRIS (9.7 g, 20 mM), glycerol (400 g, 10% w/v) NaCl (23.4 g, 100 mM), pH 8.

6.1.35.7 Concentration of protein

Once the protein was dialysed it was placed in a spin column with a molecular cut off of 10 kDa and concentrated to 10 mL by centrifugation.

6.1.36 Assays

6.1.36.1 Assay I

Pyruvate kinase (1 mg), D_2O (55 μL), ATP (5.5 μL, 100 mM), $MgCl_2$ (5.5 μL, 500 mM), KCl (44 μL, 2.5 M), desired substrate (2 mg mL^{-1}), phosphoenolpyruvate (PEP) (82.5 μL, 500 mM) were made in a solution of 100 mM TRIS pH 8. 68 μM ThiM, 1.5 μM IPK, 10 μM of FDPS, 100 nM ADS final concentration in solution were added as needed and the final volume made up to 550 μL.

The assays were incubated at room temperature with shaking for 48 hours, before extraction with pentane three times. The pentane extraction was run on the GC-MS for analysis.

Assays that did not include ADS were treated with alkaline phosphatase for four hours at 37 °C to remove the diphosphate group, before extraction with pentane allowing mass spectrometry analysis for the alcohol product.

Assays that did not include ADS and FDPS were subjected to ^{31}P NMR spectroscopy to show consecutive phosphorylation reactions had occurred. An assay without substrate was used as a comparison for the reaction with ThiM producing the monophosphorylated product, and both were used for identifying the diphosphorylated product formed from IPK. See supplementary information 7.2.3 for NMR spectra.

6.1.36.2 Assay II

Assay I (6.1.36.1) was repeated but with IPK, FDPS and ADS omitted.

The assays were incubated at room temperature with shaking for 48 hours, before centrifugation at 4000 rpm in a 10,000 Da Vivaspin centrifugal concentrator. The flow through was subjected ^{31}P NMR analysis to ensure the monophosphorylated product was present, before lyophilisation overnight. This powder was then resuspended in a solution of 100 mM TRIS pH 8. Pyruvate kinase (1 mg), D_2O (55 μL), ATP (5.5 μL, 100 mM), phosphoenolpyruvate (PEP) (82.5 μL, 500 mM) and methyl phosphonic acid (5.5 μL, 1 M) were added along with IPK or a variant of IPK to reach a final concentration of 5 μM. The reaction was monitored by ^{31}P NMR analysis.

6.1.36.3 Assay III

Assay I (6.1.36.1) was repeated but with FDPS and ADS omitted. The assays were centrifuged at 4000 rpm in a 10,000 Da Vivaspin centrifugal concentrator. The flow through was subjected ^{31}P NMR analysis to ensure the diphosphorylated product was present, before lyophilisation overnight. This powder was then resuspended in a solution of Incubation buffer B (6.1.27.2) along with IDP (synthesised as described in assay I without FDPS and ADS). WT FDPS or one of the mutations generated (M156D, M156S, M156T) (1 μM final concentration) were added along with ADS (20 μM final concentration). The assays were incubated at 37 °C for four hours at 350 rpm before the addition of 300 μL of $MgCl_2$ (20 mM) and alkaline phosphatase (10 mg mL^{-1}). The solution was incubated at 37 °C for three hours before pentane (500 μL) was added to extract the organic compounds. The pentane extraction also contained geraniol (35 μM) as an internal standard, and was analysed by GC-MS. Assays with FDPS, ADS, IDP or 4-hydroxy dimethylallyl diphosphate omitted were performed to ensure no DHAAl was produced from another source.

6.1.36.4 GC-MS method

The program used an injection port temperature of 100 °C with a split ratio of 19:1. The initial temperature of 80 °C was held for 2 minutes, and a ramp of 8 °C a minute completed until a temperature of 280 °C was reached, where it was held for 3 minutes. 5 μL of sample was injected per run.

6.1.36.5 GC-FID method

The program used an injection port temperature of 200 °C with a split ratio of 19:1. The oven temperature was held at 80 °C for 1 min, then raised 8 °C a minute until a temperature of 200 °C was reached, where it was held for 2 minutes. 2 μL of sample was injected per run.

6.1.37 UPK IPK FDPS ADS assays

6.1.37.1 Assay I

Pyruvate kinase (1 mg), D_2O (100 μL), 1 μL of desired substrate (5 mM), ATP (7.5 μL, 100 mM), PEP (15 μL, 500 mM), and UPK (10 μM) were made in a 750 μL solution of 20 mM TRIS, 20 mM $MgCl_2$.The solution was incubated at 37 °C for two hours at 350 rpm before quenching with methanol. The solution was subjected to ^{31}P NMR every 15 minutes to try and follow the reaction, and by 2 hours, no PEP was visible, suggesting the reaction was over. The NMR showed in the supplementary information 7.3.6 is after 30 minutes, as from this time point on the product peak dropped due to precipitation. The solution was centrifuged to pellet the white precipitate, which was used in further assays with IPK or dissolved in butanol and 1% phosphoric acid (1:1, 200 μL) for analysis by TLC.

6.1.37.2 Assay II

Hydroxyfarnesyl monophosphate (2 mM) was dissolved in a buffer containing β-cyclodextrin (2 mM), HP-β-cyclodextrin (2 mM), TRIS (20 mM) and $MgCl_2$ (20 mM). ATP (25 μL, 5 mM), IPK (0.5 μM), ADS (10 μM), PEP (10 μL, 500 mM), pyruvate kinase (1 mg), D_2O (100 μL) were added for a final volume of 500 μL. The solution was incubated at 37 °C overnight at 350 rpm before the addition of 300 μL of $MgCl_2$ (20 mM) and alkaline phosphatase (10 mg mL^{-1}). The solution was incubated at 37 °C for three hours before heptane (500 μL) was added to extract the organic compounds. The heptane extraction was analysed by GC-MS.

6.1.37.3 GC-MS method

As described in 6.1.36.4.

6.1.38 Germacrene B synthase purification

6.1.38.1 Cell lysis buffer VI

Tris-Base (6.1 g, 50 mM), NaCl (2.9 g, 50 mM), Tween (2 g, 0.2% w/w), and β-ME (350 µL, 5 mM) were dissolved in deionised water (900 mL), the pH adjusted to 8 with HCl, and made up to a litre.

6.1.38.2 Sonication

Using cell lysis buffer VI, and the same protocol as CYP124A1

6.1.38.3 Basic extraction

The insoluble fraction obtained from centrifugation was resuspended in cell lysis buffer VI (60 mL) and stirred at 4 °C for 30 minutes. NaOH (5 M) was added dropwise to the stirring solution until pH 12 was reached. The solution was left stirring for 30 minutes at 4 °C. Hydrochloric acid (pH 6) was added dropwise until the solution reached pH 8. The solution was left stirring for 30 minutes at 4 °C. The solution was centrifuged (18,000 rpm/ 40 minutes/ 4 °C) and supernatant kept. Supernatant was passed through a 0.2 µm syringe filter before use. Both the cell pellet sample and the supernatant sample were analysed *via* SDS-PAGE.

6.1.39 Bradford assay

6.1.39.1 Bradford reagent

Brilliant blue G250 (10 mg) was dissolved in ethanol (2 mL) and phosphoric acid (80%, 10 mL) added. Deionised water was added to bring the volume to 100 mL, and the solution was stored in the dark at 4 °C.

6.1.39.2 Method

BSA (1 mg mL^{-1}) was dissolved in deionised water and a range of concentrations (0- 20 μg mL^{-1}, 200 μL) prepared. Bradford reagent (800 μL) was added and the solutions vortexed. Each concentration (200 μL) was placed in a 96 well plate with three repeats and the absorbance measured at 590 nm and 450 nm on a FLUOstar Omega plate reader, from BMS LABTECH. The ratio of absorbance was calculated and used to plot a standard curve. Purified protein was prepared in the same way and the ratio of absorbance measured. The standard curved obtained from BSA concentrations was then used to determine the concentration of the protein sample.

6.2 Organic chemistry

6.2.1 General

All chemicals were obtained from Sigma-Aldrich, Alfa-Aesar, Fluorochem or Fischer Scientific and used with no further purification. Anhydrous dichloromethane, acetonitrile, diethylether, and tetrahydrofuran were obtained from a MBraun MB SPS-800 solvent purification system. ^{1}H, ^{13}C, and ^{31}P NMR spectra were measured on a Bruker Avance III 600, Bruker Avance 500, Bruker Avance III HD 400 and a Bruker Fourier 300 NMR spectrometer. All spectra are reported as chemical shifts in parts per million, downfield from tetramethylsilane (^{1}H and ^{13}C) and phosphoric acid (^{31}P). Multiplicity is reported as: s-singlet, d-doublet, t-triplet, q-quartet, quin-quintet, m-multiplet, dd-doublet of doublets, dq-doublet of quartet. Coupling is rounded to the nearest 0.5 Hz, and integrals to the nearest whole number. All spectra are assigned to the appropriate image where each peak is assigned a letter of the roman alphabet. Unless specified otherwise all compounds were purified on Biotage® SNAP Ultra HP-Sphere™ 25 μM columns 10 – 100 grams. All flash chromatography was performed on a Biotage Isolera™ Four.

GC-MS analysis was run as above 6.1.1. Thin layer chromatography (TLC) was completed using pre-coated aluminium plates (20 x 20 cm) of silica 60 F_{254}. TLC visualisations were performed with UV light or the stains:

I. Ceric ammonium molybdate: 4.2 % ammonium molybate, 0.2 % ceric sulfate in 5 % H_2SO_4 made up with deionised water
II. Potassium permanganate: 0.75 % potassium permanganate, 5 % potassium carbonate, 0.625 mL of 10% NaOH solution in deionised water

III. Berberine: 0.1% berberine hydrochloride in ethanol

6.2.2 Bromination of farnesol

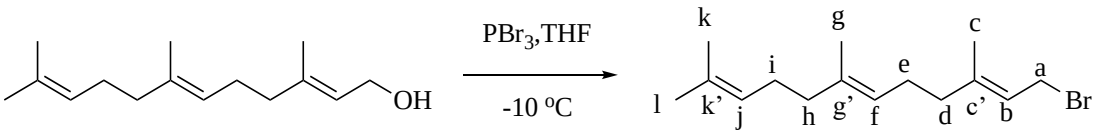


Figure 1: Farnesol, phosphorus tribromide, THF, -10 °C, under argon, 100% yield.

Phosphorus tribromide (90 μL, 1.14 mmol, 1 eq.) was added dropwise to a stirred solution of (2*E*, 6*E*)-farnesol (0.56 mL, 2.25 mmol, 2 eq.) in anhydrous tetrahydrofuran (30 mL) at -10 °C. The solution was stirred for 15 minutes under inert atmosphere and quenched with ammonium chloride. The product was extracted with diethylether (3 x 20 mL), washed with brine (20 mL), dried (Na_2SO_4) and the solvent removed under reduced pressure. The product was obtained as a yellow oily solution which was used without further purification.

^{1}H-NMR ($CDCl_3$; 400 MHz) 1.52 (6H, s, Hk, Hl), 1.60 (3H, s, Hg) 1.65 (3H, s, Hc), 1.87-1.93, (2H, m, Hd) 1.96-2.06 (6H, m, He, Hh, Hi), 3.93 (2H, d *J* 5.95, Ha), 5.01 (2H, t *J* 6.94, Hj, Hf), 5.45 (1H, t *J* 8.21, Hb).
^{13}C-NMR ($CDCl_3$; 100 MHz) C 143.5 (c'), 135.6 (g'), 131.2 (k'), CH 124.4 (C-b), 123.4 (C-f), 120.7 (C-j), CH_2 39.7 (C-h), 39.5 (C-a), 29.5 (C-d), 26.7 (C-e) 26.1 (C-i), CH_3 25.7 (C-l), 17.7 (C-k), 16.0 (C-g), 16.0 (C-c).
HRMS (EI) m/z calculated for $C_{15}H_{25}Br$: 284.1134 Found: 284.1
See supplementary information 7.3.2 for spectra.

6.2.3 Diphosphorylation of farnesylbromide

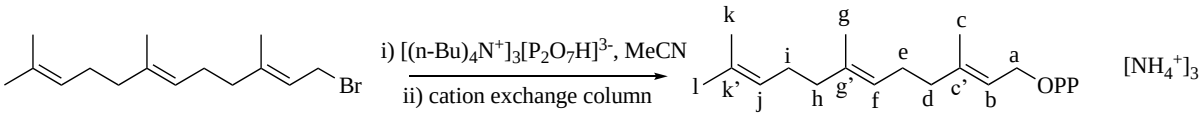


Figure 2:Farnesyl bromide, tris(tetrabutyl ammonium) hydrogen pyrophosphate, acetonitrile, RT, under argon, 26% yield.

The bromide was taken up in anhydrous acetonitrile (30 mL) and tris(tetrabutyl ammonium) hydrogen pyrophosphate (3.15 g, 3.5 mmol, 2 eq.) added. The reaction was left stirring overnight at room temperature under inert atmosphere. The solvent was removed under reduced pressure, and the oily residue dissolved in ion exchange buffer (25 mM NH_4HCO_3,

2% v/v isopropanol) (5 mL). This cloudy solution was passed slowly through a Amberlyst 131 (wet H^+ form) mesh cation exchange resin column which was pre-equilibrated in the ion exchange buffer. The tris(tetrabutyl ammonium) hydrogen pyrophosphate salt counter ions were exchanged for ammonium ions during the process. Fractions were eluted with ion exchange buffer and analysed by TLC (3:1:1, isopropanol:NH_4OH:ion exchange buffer). The appropriate fractions were pooled and lyophilized to give a white powder. This white powder was dissolved in ion exchange buffer (5 mL) before purification by flash column chromatography using a Biotage® SNAP 10 g HP-Biosphere C18 column (water:acetonitrile) and the product eluted at 60% acetonitrile.

The fractions containing the product were collected and lyophilized to give (2*E*, 6*E*)-farnesyl diphosphate as a white powder (0.22 g, 26%).

^{1}H-NMR (D_2O; 400 MHz) 1.50 (6H, s, Hk, Hl), 1.57 (3H, s, Hg) 1.60 (3H, s, Hc), 1.88-1.94, (2H, m, Hd) 1.97-2.07 (6H, m, He, Hh, Hi), 4.36 (2H, t *J* 6.21, Ha), 5.09 (2H, t *J* 13.4, Hj, Hf), 5.35 (1H, t *J* 7.11, Hb).

^{13}C-NMR (D_2O; 100 MHz) C 142.9 (c'), 136.6 (g'), 133.5 (k'), CH 124.4 (C-f), 124.2 (C-j), 119.6 (C-b), CH_2 62.6 (C-a), 38.8 (C-d), 38.7 (C-h), 25.7 (C-e) 25.5 (C-i), CH_3 24.8 (C-l), 16.9 (C-k), 15.5 (C-g), 15.2 (C-c).

^{31}P (D_2O, 162 MHz) -8.4 (d *J* 21.94, CH_2-O-P-O-*P*), -10.6 (dt *J* 20.27, 6.01, CH_2-O-*P*).

HRMS (ESI-) $[M-H]^-$ m/z calculated for $C_{15}H_{27}P_2O_7$: 381.1232 Found: 381.1247

See supplementary information 7.3.3 for spectra.

6.2.4 Synthesis of 12- hydroxy farnesol

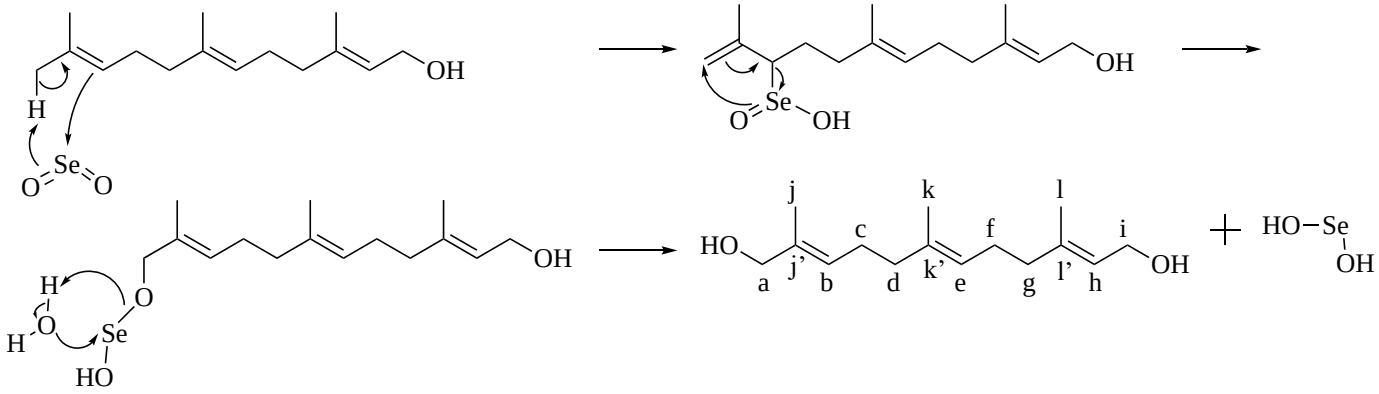


Figure 3: Farnesol, SeO_2, salicylic acid, tert-butyl hydroperoxide, DCM, 0 °C, 68% yield.

SeO_2 (109 mg, 1 mmol), salicylic acid (135 mg, 1 mmol), and *tert*-butyl hydroperoxide (1.6 mL, 16.3 mmol) were stirred in dichloromethane (20 mL) for 30 min at room temperature. The temperature was reduced to 0 °C before the addition of farnesol (0.58 mL, 3.3 mmol). The reaction was left to stir for 24 hours at 0-4 °C. Sat. Na_2SO_3 (20 mL) was added, and the product was extracted with DCM (3 x 20 mL), washed with brine (20 mL), dried (Na_2SO_4) and the solvent removed under reduced pressure.

The product was purified by flash column chromatography (ethyl acetate:hexane 20:80), to give 12-hydroxy farnesol as a yellow oil (0.32 g, 40%).

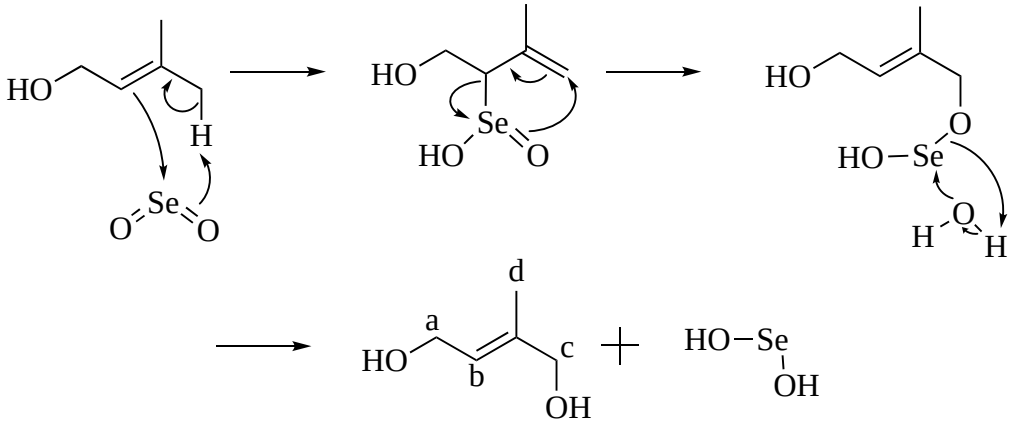
^{1}H-NMR ($CDCl_3$; 400 MHz) 1.59, 1.65, 1.66(3x 3H, s, Hj, Hk, Hl) 1.99-2.07, 2.09-2.16, (2x 4H, m, Hc, Hd, Hf, Hg), 3.98 (2H, s, Ha), 4.14, (2H, d *J* 6.68 Hz, Hi) 5.10 (1H, t *J* 6.51, He), 5.38 (2H, q *J* 7.24, Hb, Hh).

^{13}C-NMR ($CDCl_3$; 100 MHz) C 139.4 (j'), 134.9 (l'), 134.8 (k'), CH 125.7 (C-h), 124.1 (C-b), 123.5 (C-e), CH_2 68.9 and 59.4 (C-i, a), 39.4 and 39.2 (C-c,f) 26.1 and 25.9 (C-d,g), CH_3 16.2 and 16.0 (C-j, l), 13.7 (C-k).

LR-MS (EI) $[M-H]^+$ m/z calculated for $C_{15}H_{26}O_2$: 238.37 Found: 202.2 (loss of two OH groups)

See supplementary information 7.3.1 for spectra.

6.2.5 Synthesis of 4-hydroxyprenol



hydroxy prenol

Figure 4: Prenol, SeO_2, ethanol, 60 °C, 40% yield

Prenol (3.3 mL, 33 mmol) was dissolved in ethanol (50 mL) and SeO_2 (3.66 g, 33 mmol) was added and heated to 60 °C. The reaction was stirred for 40 min. Sat. Na_2SO_3 (20 mL) was added and the product was extracted with diethyl ether (3 x 20 mL), washed with brine (20 mL), dried (Na_2SO_4) and the solvent removed under reduced pressure. The product was

purified by flash column chromatography (ethyl acetate:hexane 20:80), to give hydroxyprenol as a yellow oil (1.33 g, 40%).

^{1}H-NMR ($CDCl_3$; 400 MHz) 1.60 (3H, s, Hd), 3.91, (2H, s, Hc), 4.10 (2H, d *J* 6.5, Ha), 5.55 (1H, t *J* 12.99, Hb). ^{13}C-NMR ($CDCl_3$; 100MHz) C 137.9 (alkene), CH 123.3 (C-b), CH_2 67.4 (C-a), 58.7 (C-c), CH_3 13.7 (C-d).
LR-MS (EI) $[M-H]^+$ m/z calculated for $C_5H_{10}O_2$: 102.13 Found: 102.2
See supplementary information 7.2.1 for spectra

7 Appendix I: Supplementary information

7.1 Chapter 2 Investigation into the promiscuity of the ADS active site

7.1.1 Protein purification

7.1.1.1 SDS-PAGE gel of ADS

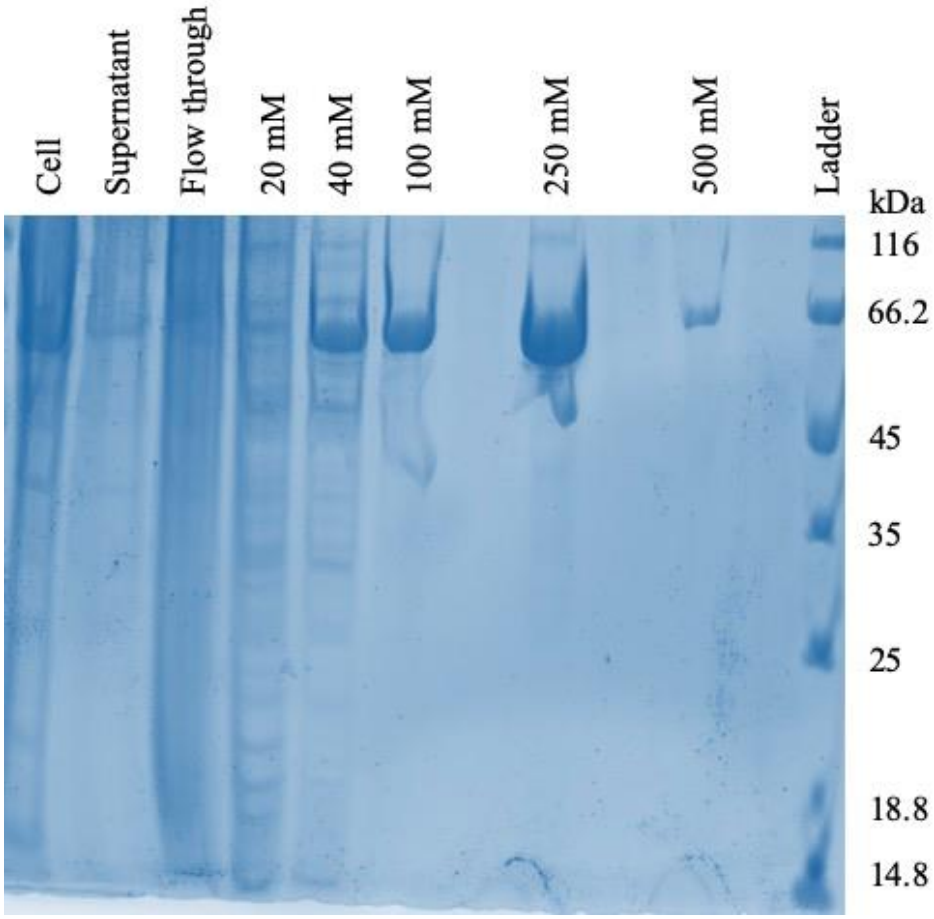


Figure a. Example SDS-PAGE gel of nickel column purification of ADS

7.1.2 Images of mutations of ADS

7.1.2.1 Mutants which behaved like wild type amorphadiene synthase

I295A and I295F

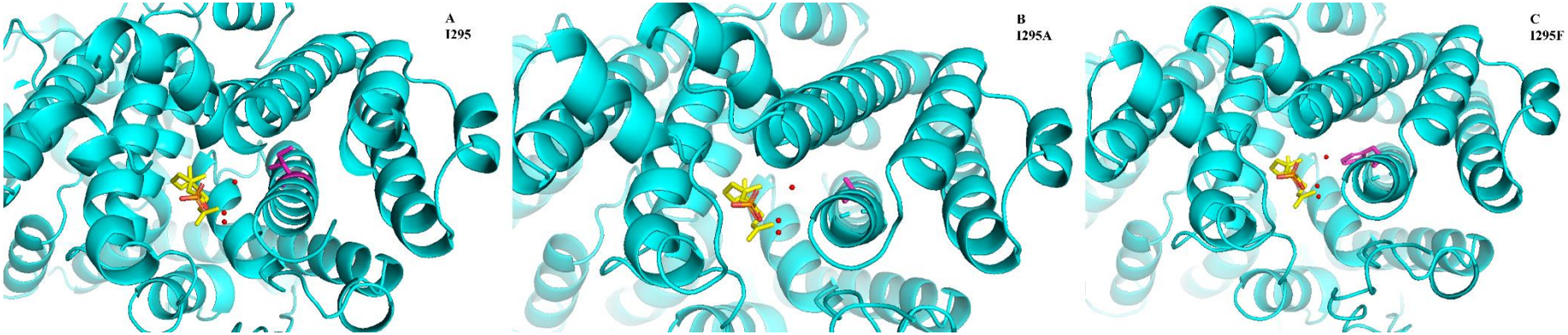


Figure b. Cartoon representation of homology model of amorphadiene synthase in cyan, FHP in yellow, and magnesium ions in red: A) I295 the residue chosen to mutate in magenta. B) residue 295 mutated to alanine in magenta. C) residue 295 mutated to phenylalanine in magenta.

L374A

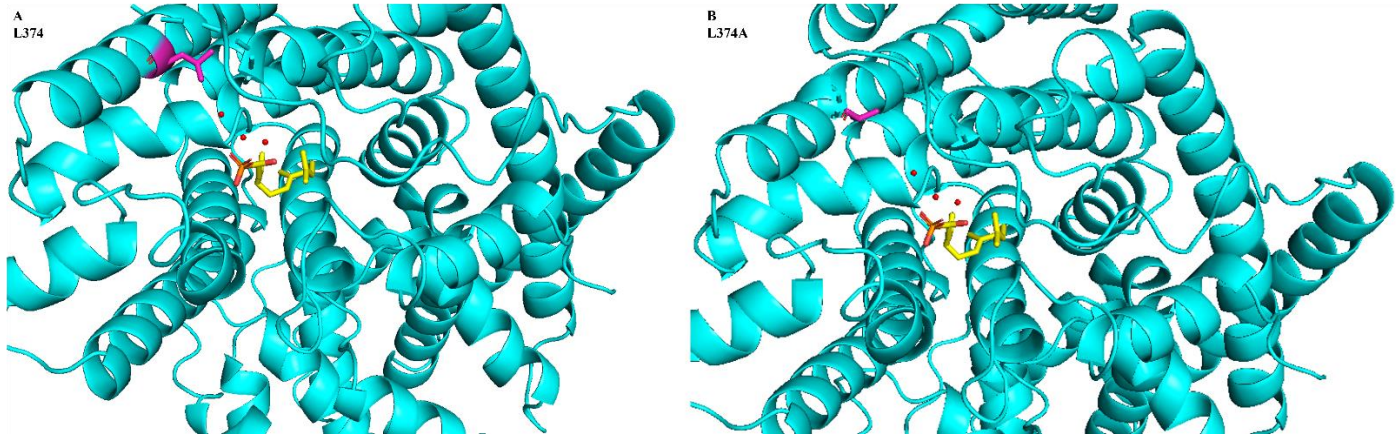


Figure c.Cartoon representation of homology model of amorphadiene synthase in cyan, FHP in yellow, and magnesium ions in red: A) L374 the residue chosen to mutate in magenta. B) residue 374 mutated to alanine in magenta.

E377A

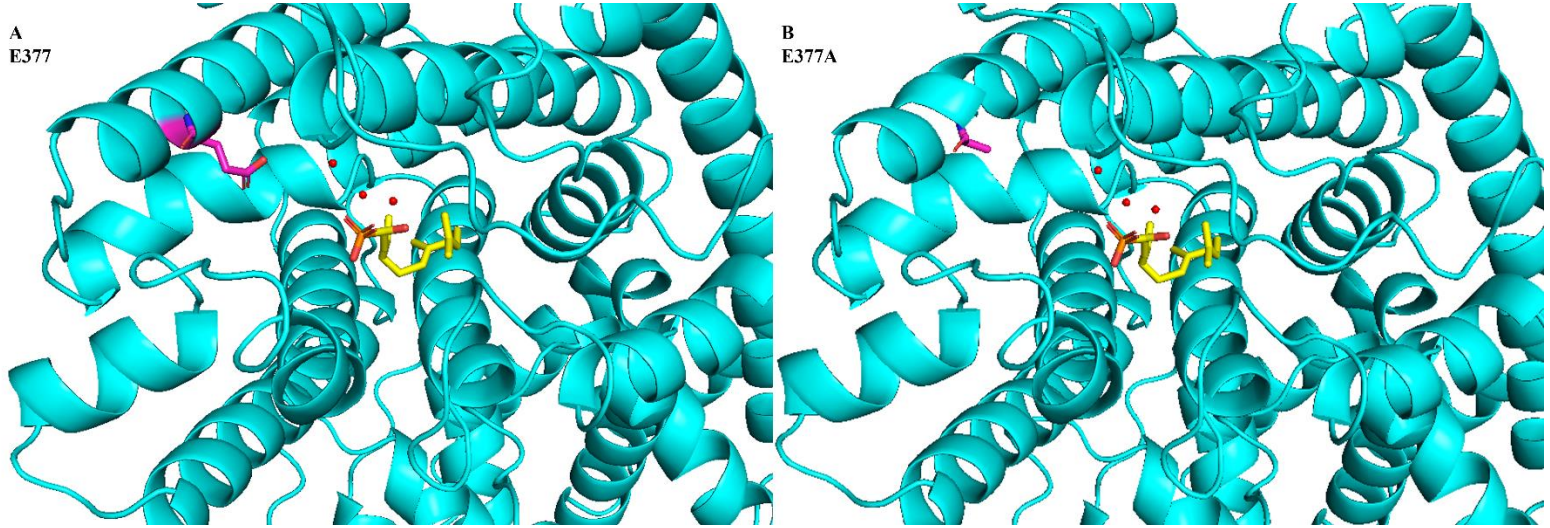


Figure d.Cartoon representation of homology model of amorphadiene synthase in cyan, FHP in yellow, and magnesium ions in red: A) E377 the residue chosen to mutate in magenta. B) residue 377 mutated to alanine in magenta.

A402Y, A402L

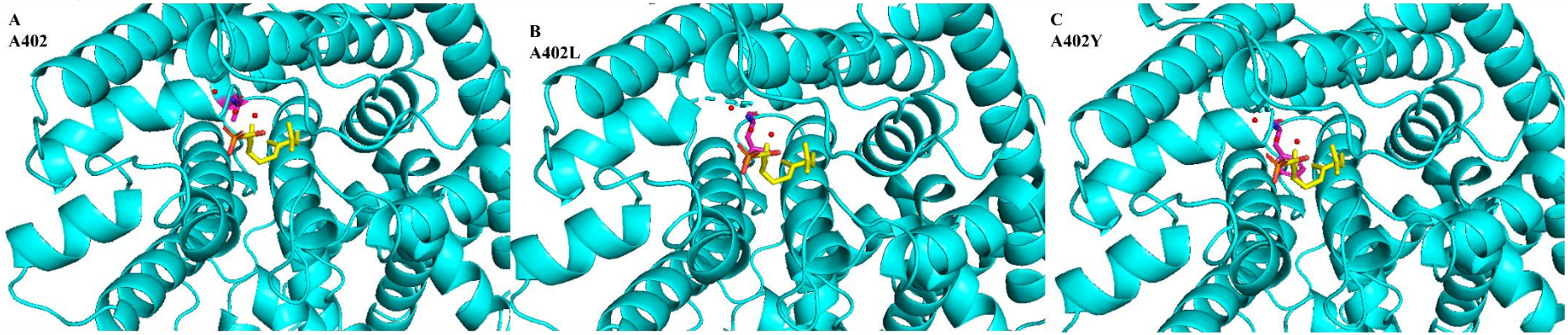


Figure e.Cartoon representation of homology model of amorphadiene synthase in cyan, FHP in yellow, and magnesium ions in red: A) A402 the residue chosen to mutate in magenta. B) residue 402 mutated to leucine in magenta. C) residue 402 mutated to tyrosine in magenta

N524D

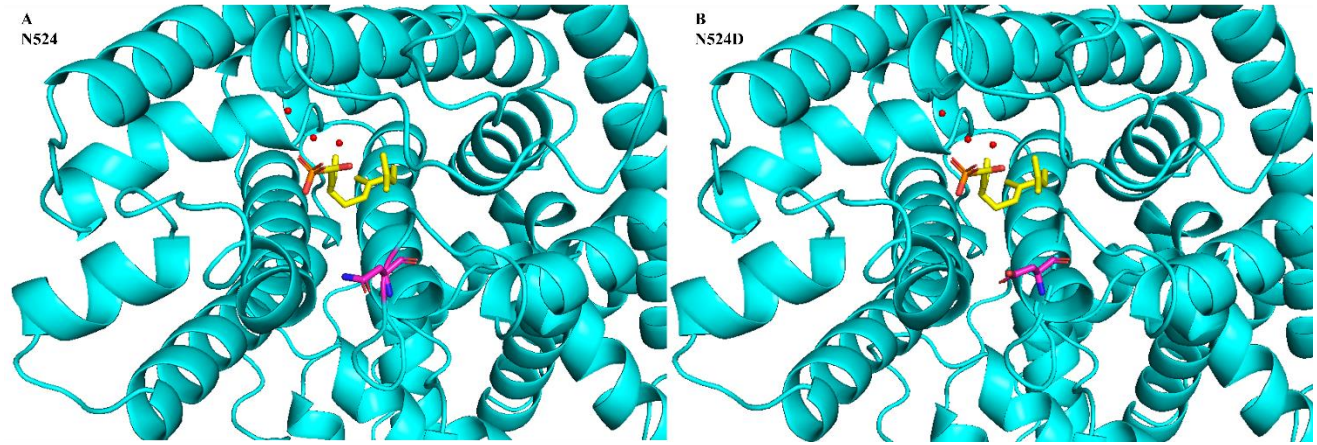


Figure f. Cartoon representation of homology model of amorphadiene synthase in cyan, FHP in yellow, and magnesium ions in red: A) N524 the residue chosen to mutate in magenta. B) residue 524 mutated to aspartate in magenta.

T526L and T526D

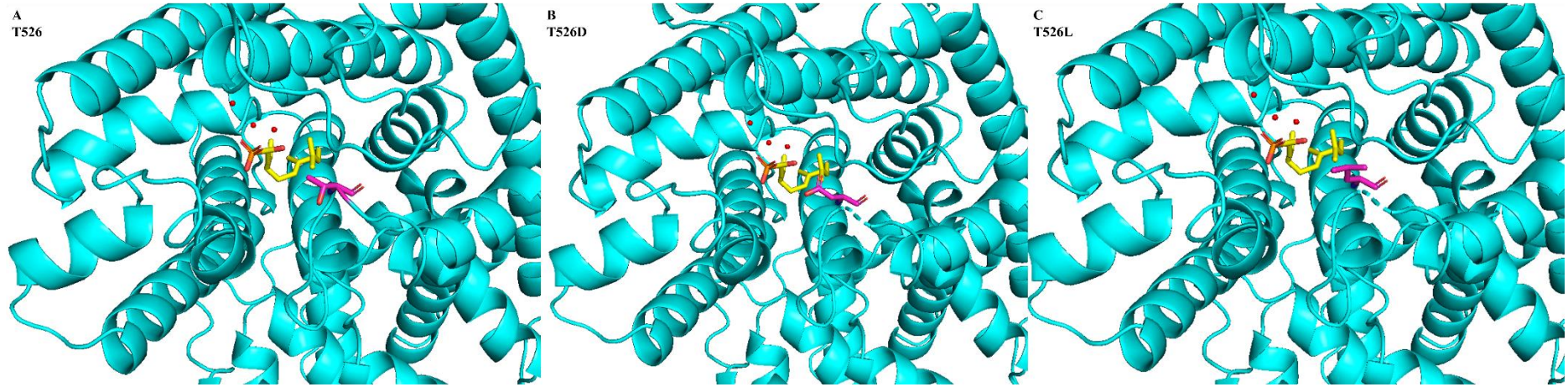


Figure g.Cartoon representation of homology model of amorphadiene synthase in cyan, FHP in yellow, and magnesium ions in red: A) T526 the residue chosen to mutate in magenta. B) residue 526 mutated to aspartate in magenta. C) residue 526 mutated to leucine in magenta

7.1.2.2 Mutants of amorphadiene synthase which were inactive

N443L and N443I

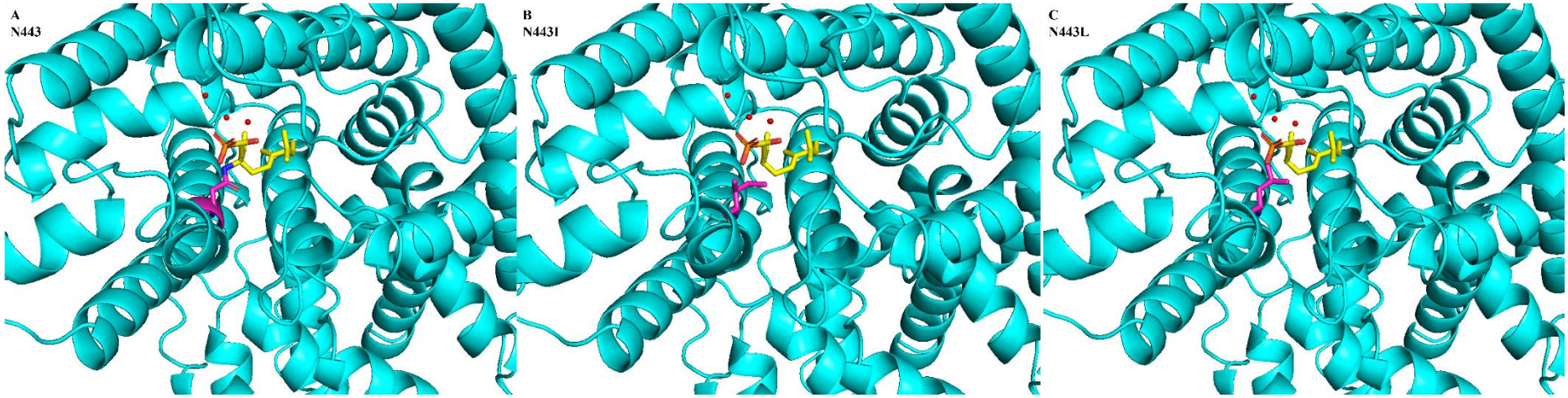


Figure h.Cartoon representation of homology model of amorphadiene synthase in cyan, FHP in yellow, and magnesium ions in red: A) N443 the residue chosen to mutate in magenta. B) residue 443 mutated to isoleucine in magenta. C) residue 443 mutated to leucine in magenta

7.1.3 Gas-chromatography analysis of ADS mutants

7.1.3.1 Gas chromatogram of an assay with N443L

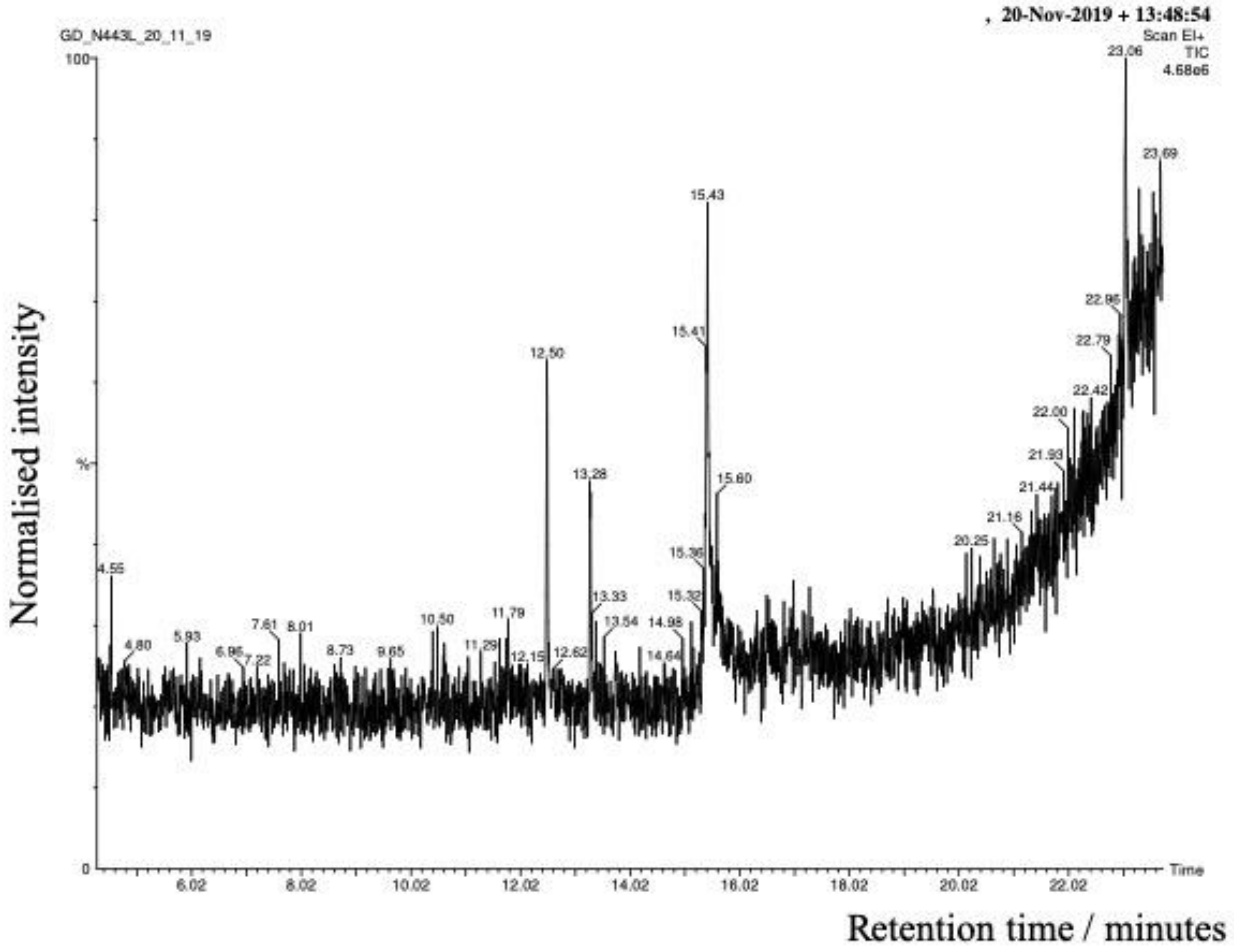


Figure i.TIC from GC-MS analysis of pentane extractions of an assays with 1 mM ADS N443L with 0.4mM FDP and 5mM $MgCl_2$.

7.1.3.2 GC-FID traces of assays comparing products of ADS mutant G401L and wtSDS

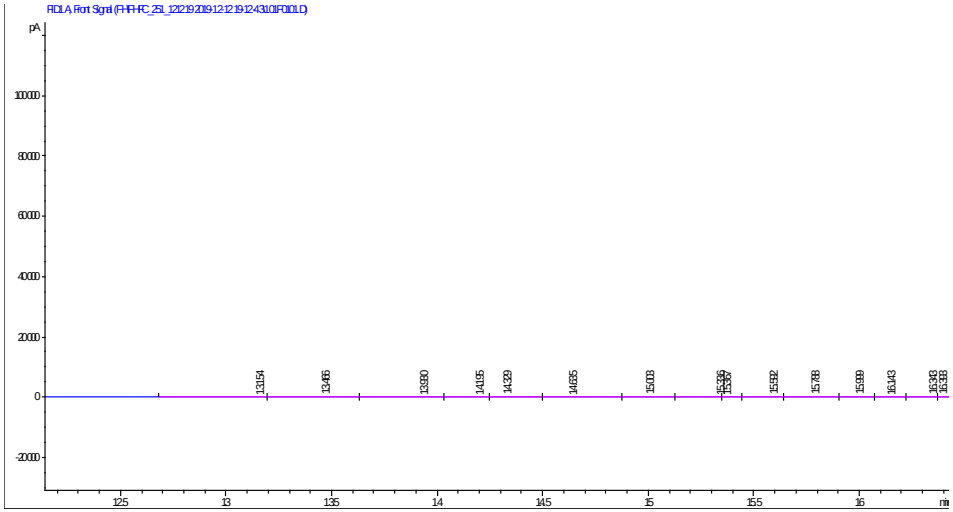


Figure j.GC-FID chromatograms of a pentane blank

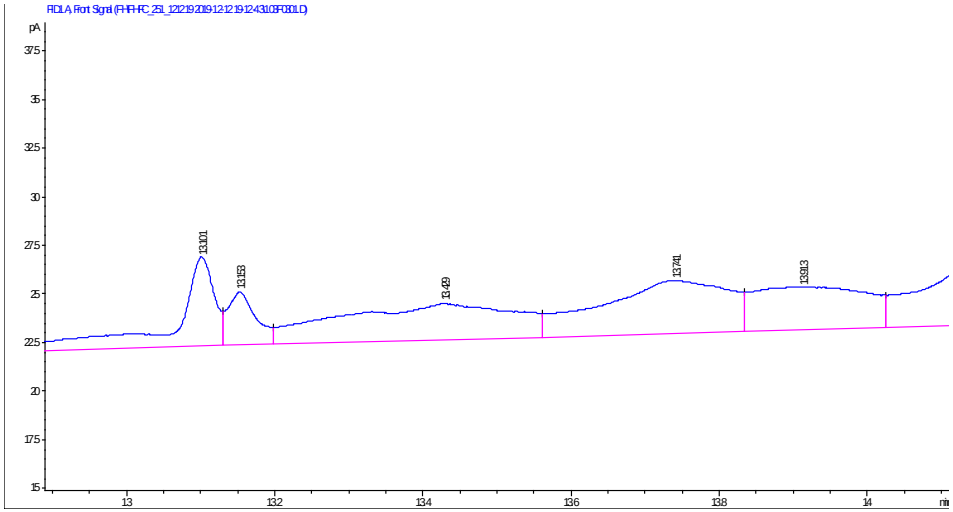


Figure k. GC-FID chromatograms of pentane extractions of assays with ADS mutant G401L

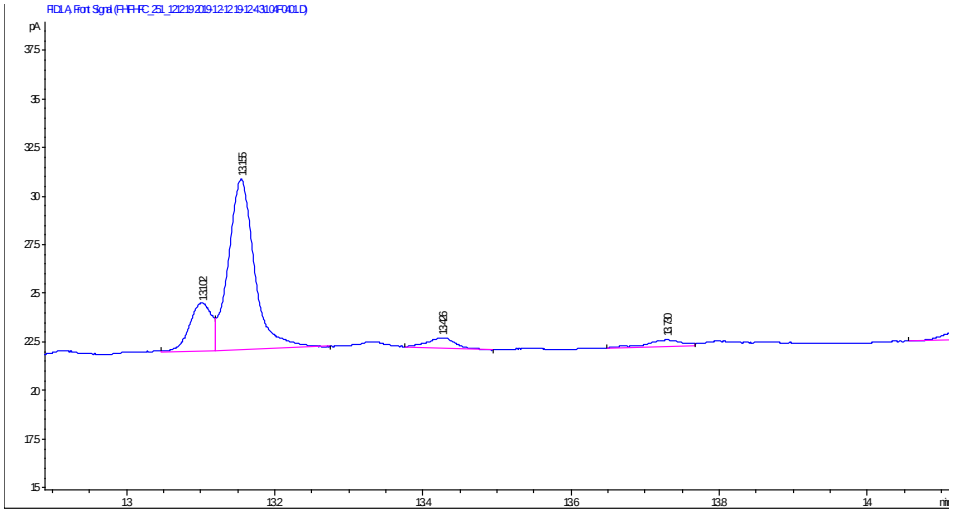


Figure l. GC-FID chromatograms of pentane extractions of assays with wt SDS

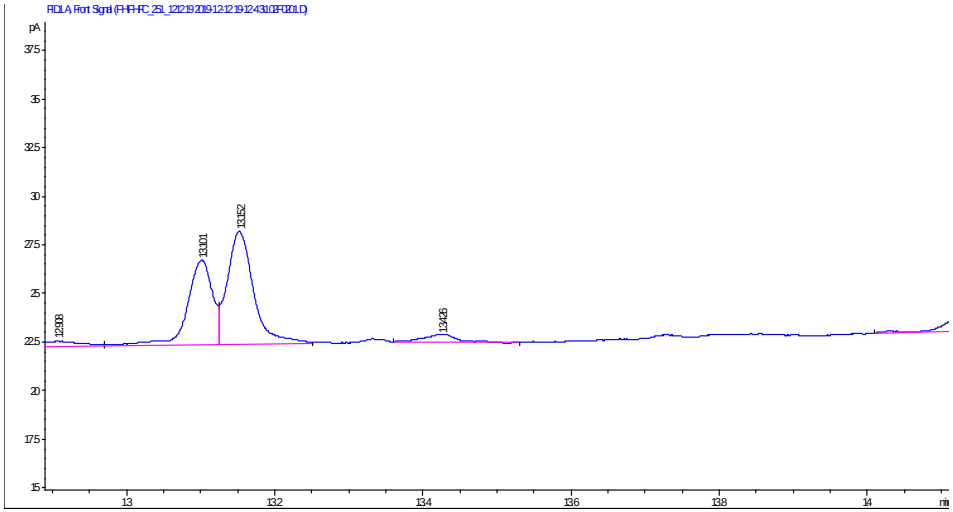


Figure m.GC-FID chromatograms of pentane extractions of assays with wt SDS and ADS mutant G401L as a co-elution

7.2 Chapter 3

7.2.1 4-hydroxy prenol

7.2.1.1 Hydrogen NMR of 4-hydroxyprenol in $CDCl_3$

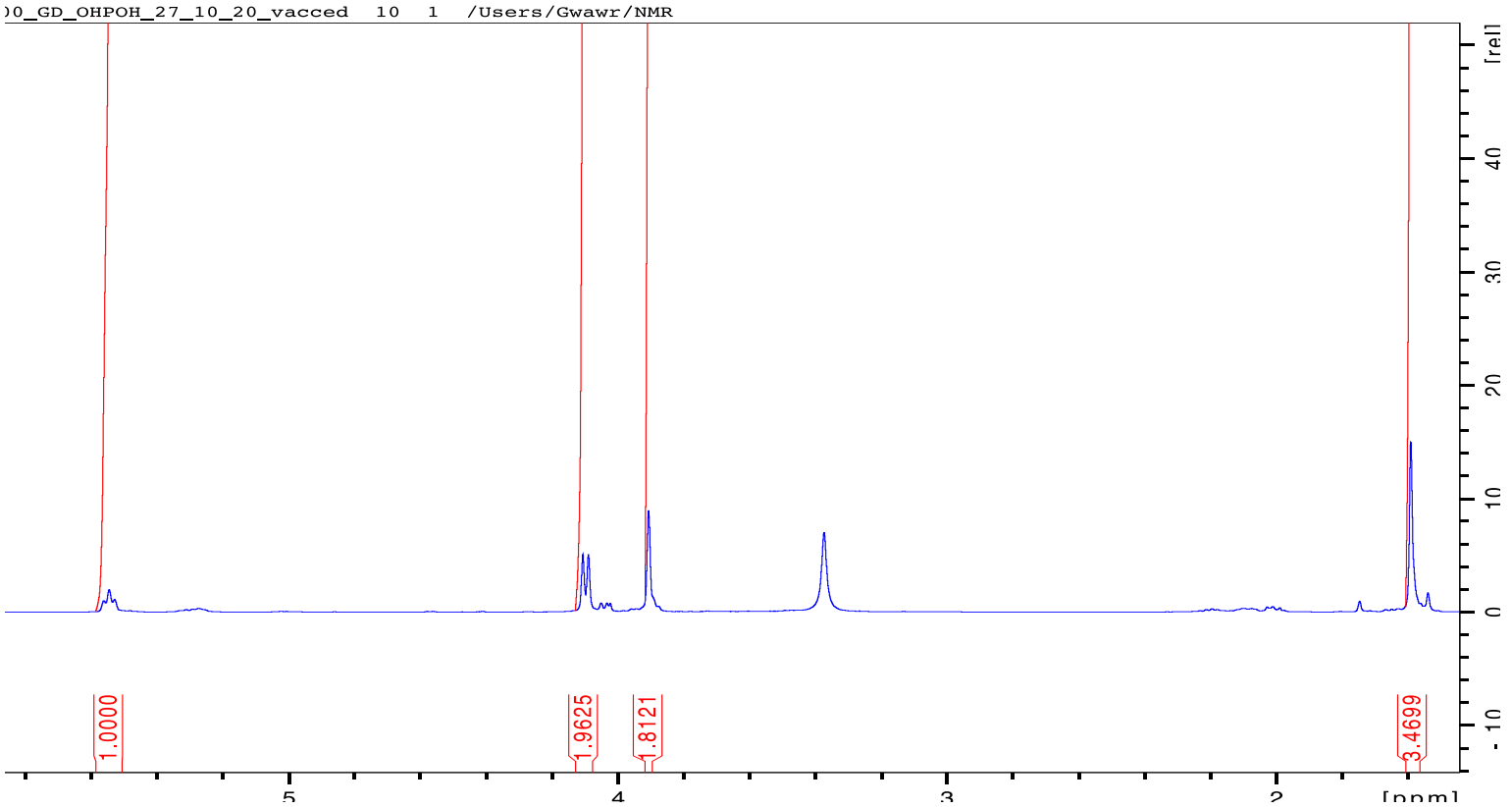


Figure n. ^{1}H NMR spectrum (400 MHz, $CDCl_3$, 298K) of 4-hydroxy prenol

7.2.1.2 Carbon NMR of 4-hydroxyprenol in $CDCl_3$

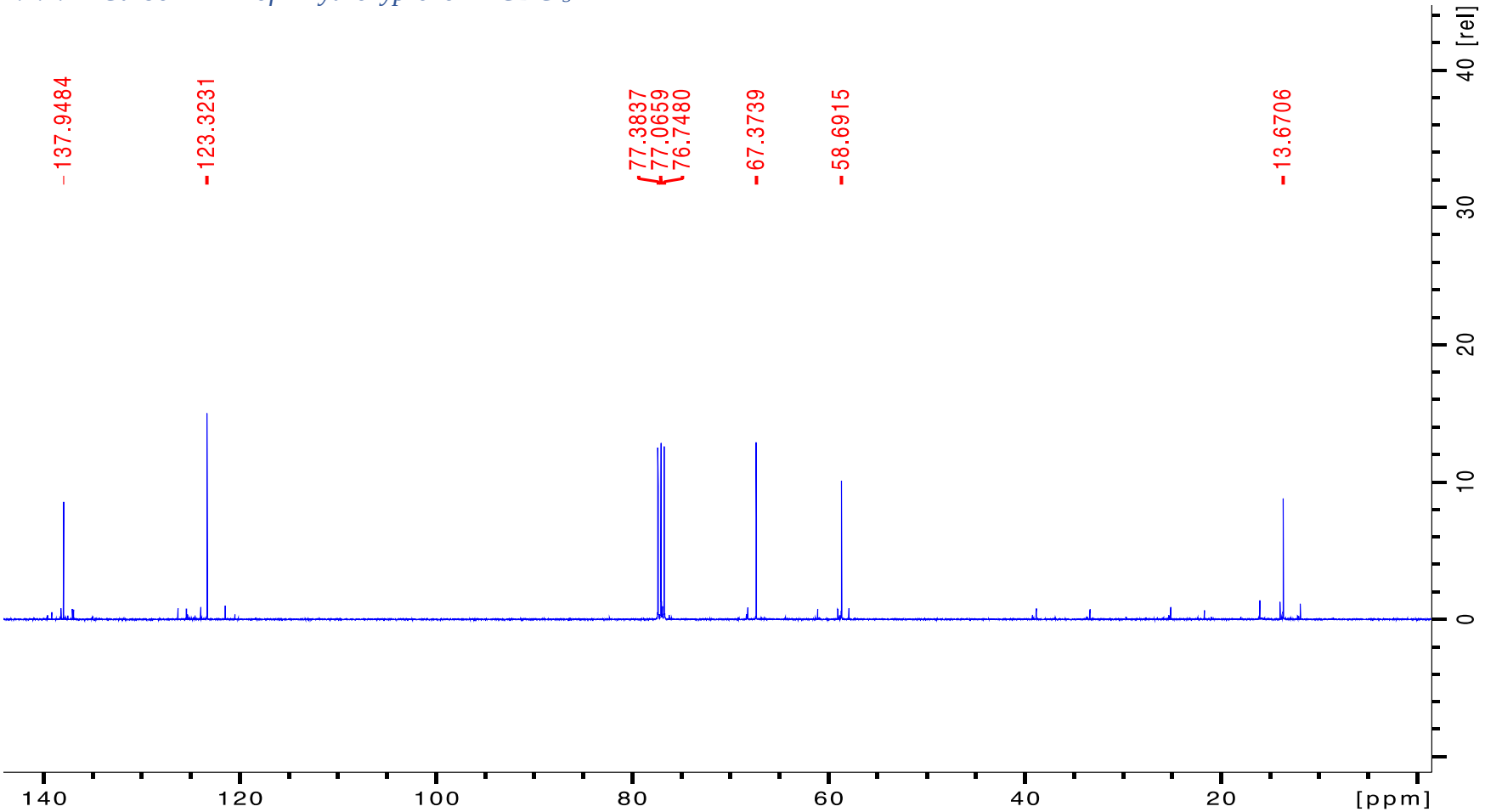


Figure o. ^{13}C NMR spectrum (100 MHz, $CDCl_3$, 298K) of 4-hydroxy prenol

7.2.1.3 Mass spectrum of 4-hydroxy prenol

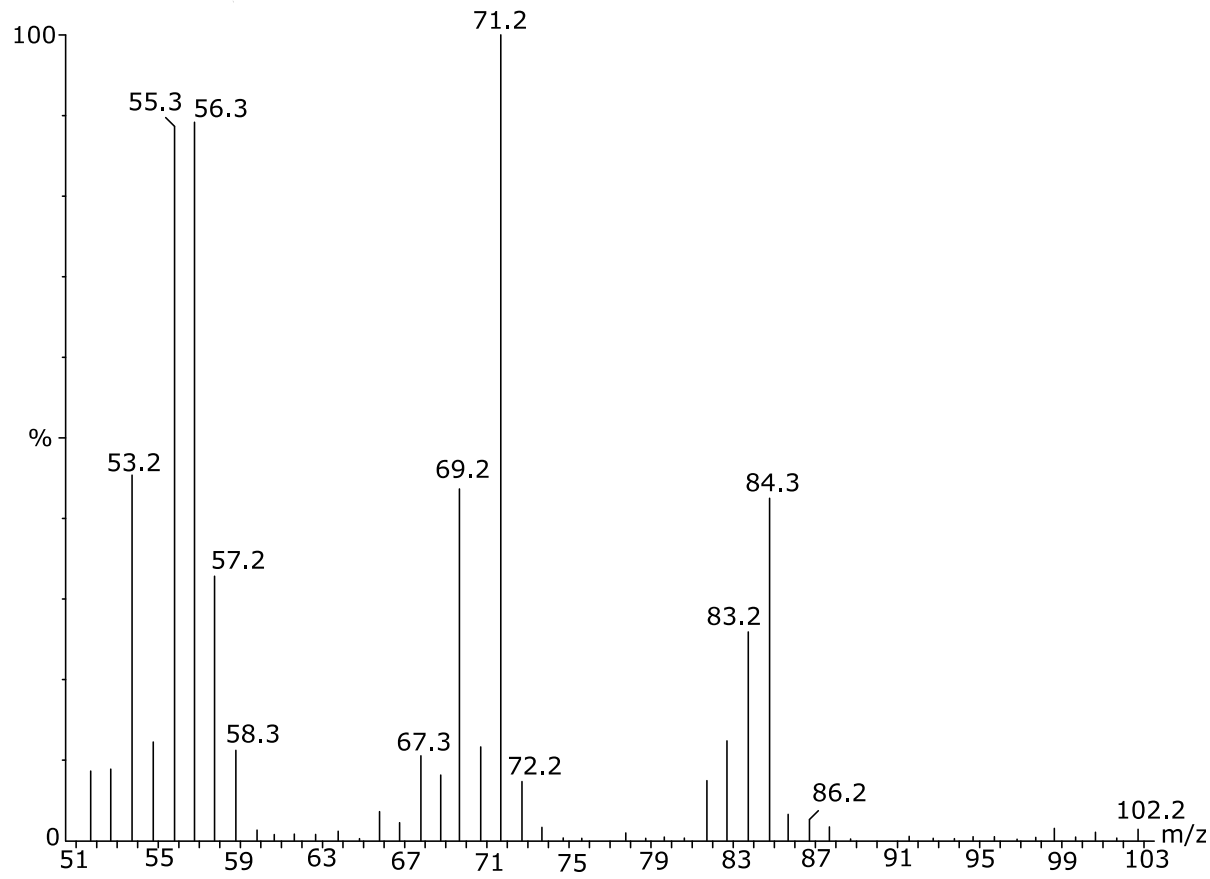


Figure p.Mass spectrum of peak at 4.79 min of total ion chromatogram of the synthesised compound 4-hydroxy prenol.

7.2.2 Protein purification

7.2.2.1 SDS-PAGE Gel of ThiM purification

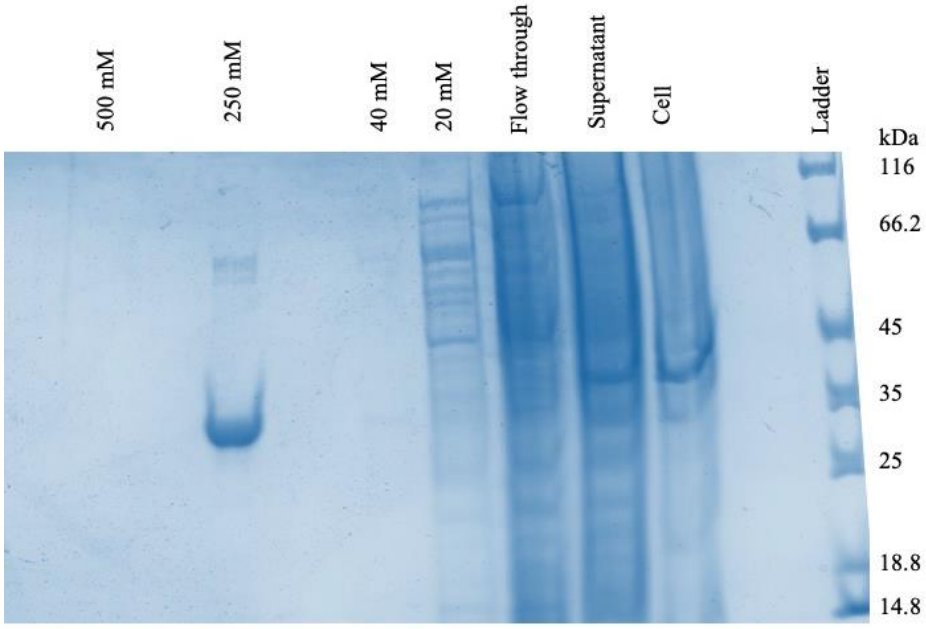


Figure q. Example SDS-PAGE gel of nickel column purification of ThiM

7.2.2.2 SDS-PAGE Gel of IPK purification

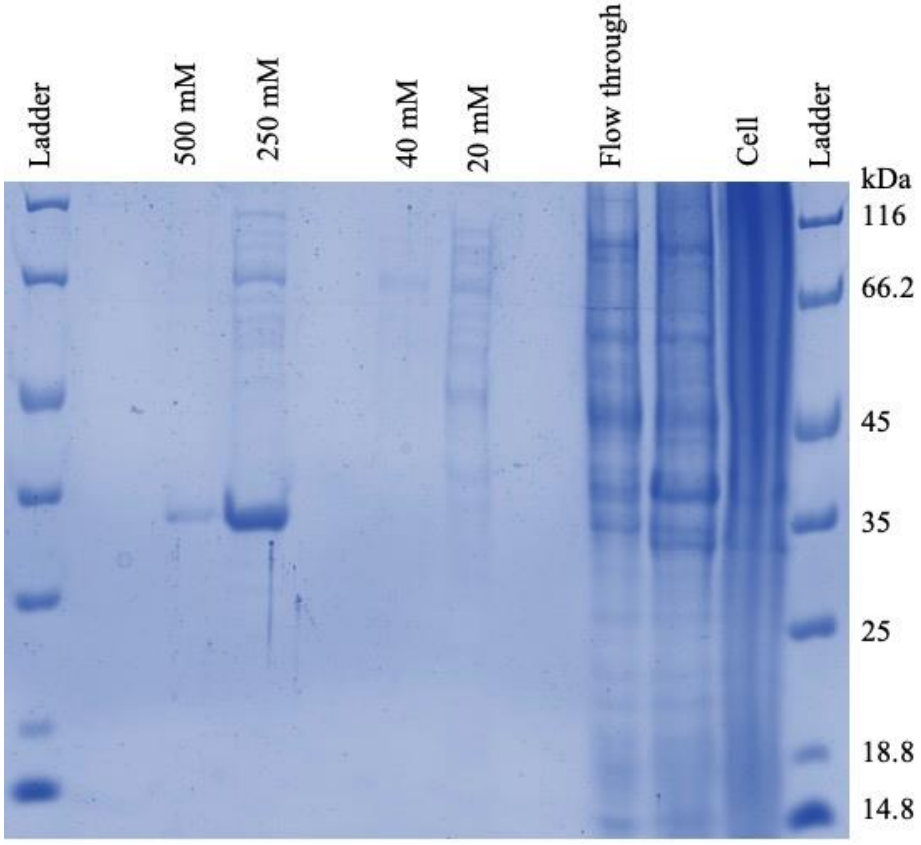


Figure r.Example SDS-PAGE gel of nickel column purification of IPK

7.2.2.3 SDS-PAGE Gel of FDPS purification

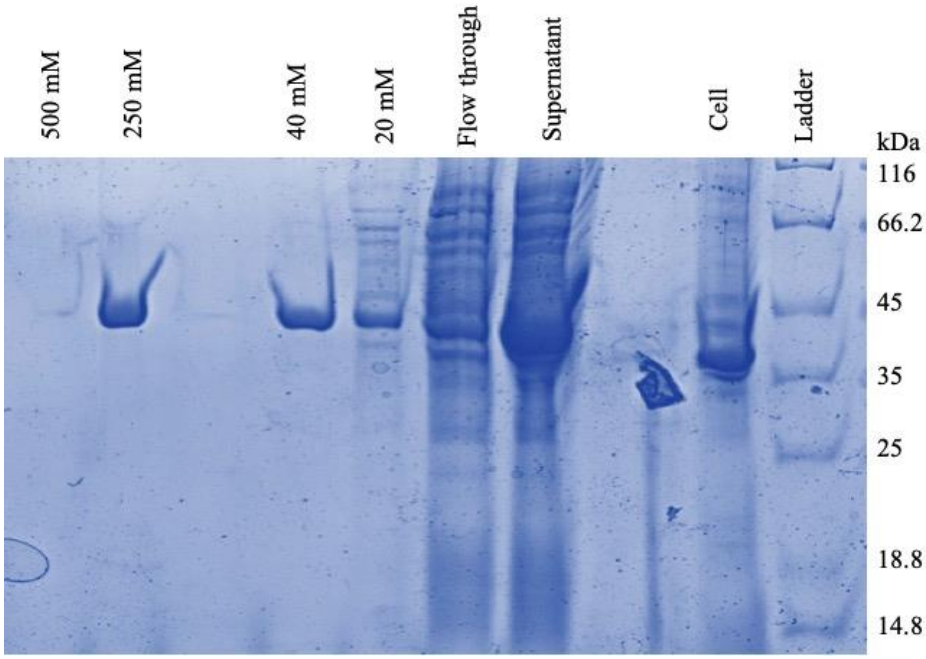


Figure s. Example SDS-PAGE gel of nickel column purification of FDPS

7.2.3 Phosphorylation assays

7.2.3.1 Phosphorus NMR of isoprenol and prenol monophosphorylated by ThiM, with a negative control in D_2O

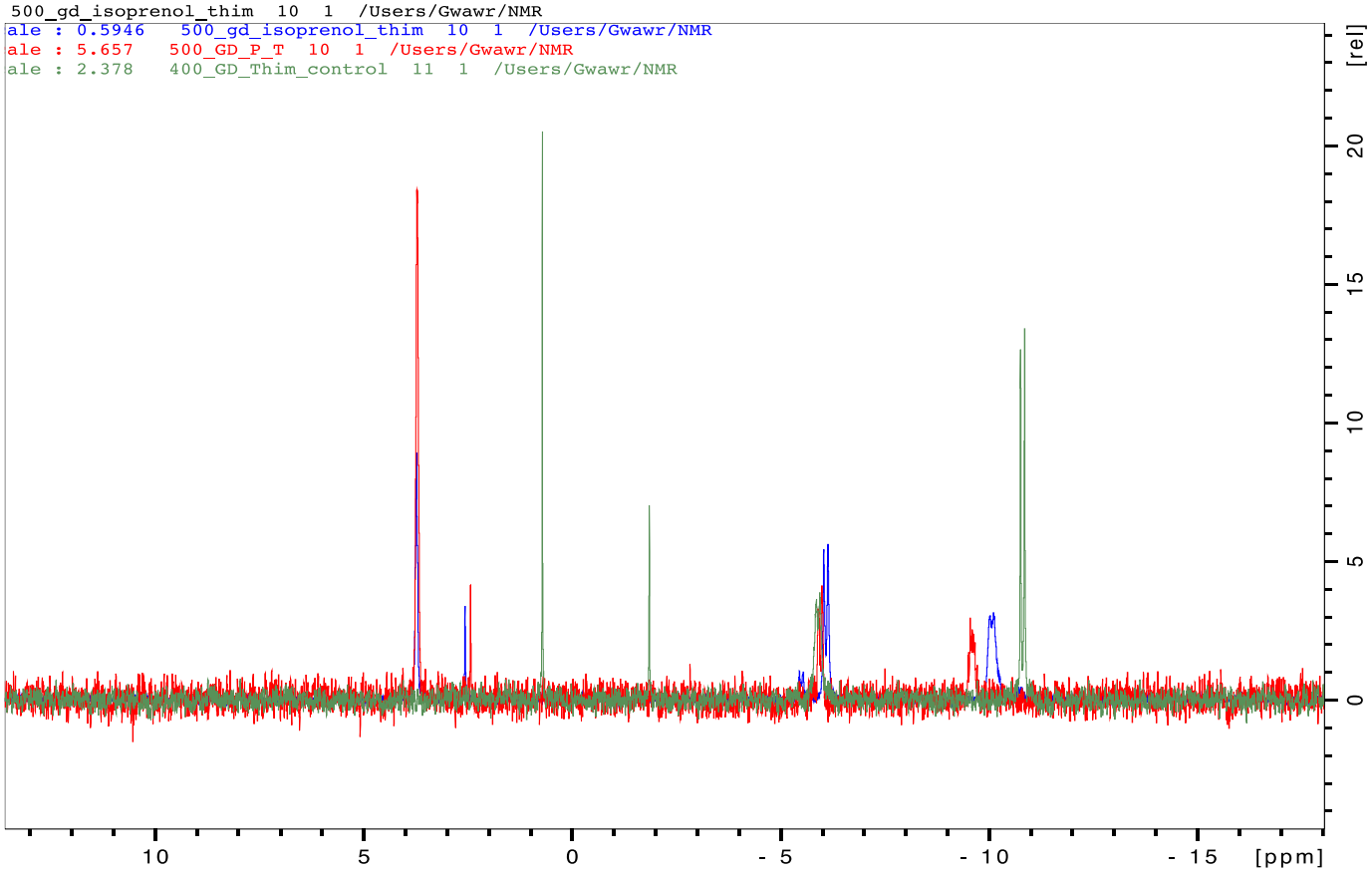


Figure t. ^{31}P NMR spectrum (243 MHz, D_2O, 298K) of ThiM with: blue – isoprenol, red – prenol, green – no substrate (this is a little shifted due to a variation in pH as no reaction occurred). The peak at: 4 ppm is the monophosphorylated product, 2.5 ppm [1 ppm in the green spectrum] is free phosphate, -1.9 ppm is PEP and the peaks at -5.9 and -10.5 ppm are ADP.

7.2.3.2 Phosphorus NMR of isoprenol and prenol diphosphorylated by Thim and IPK

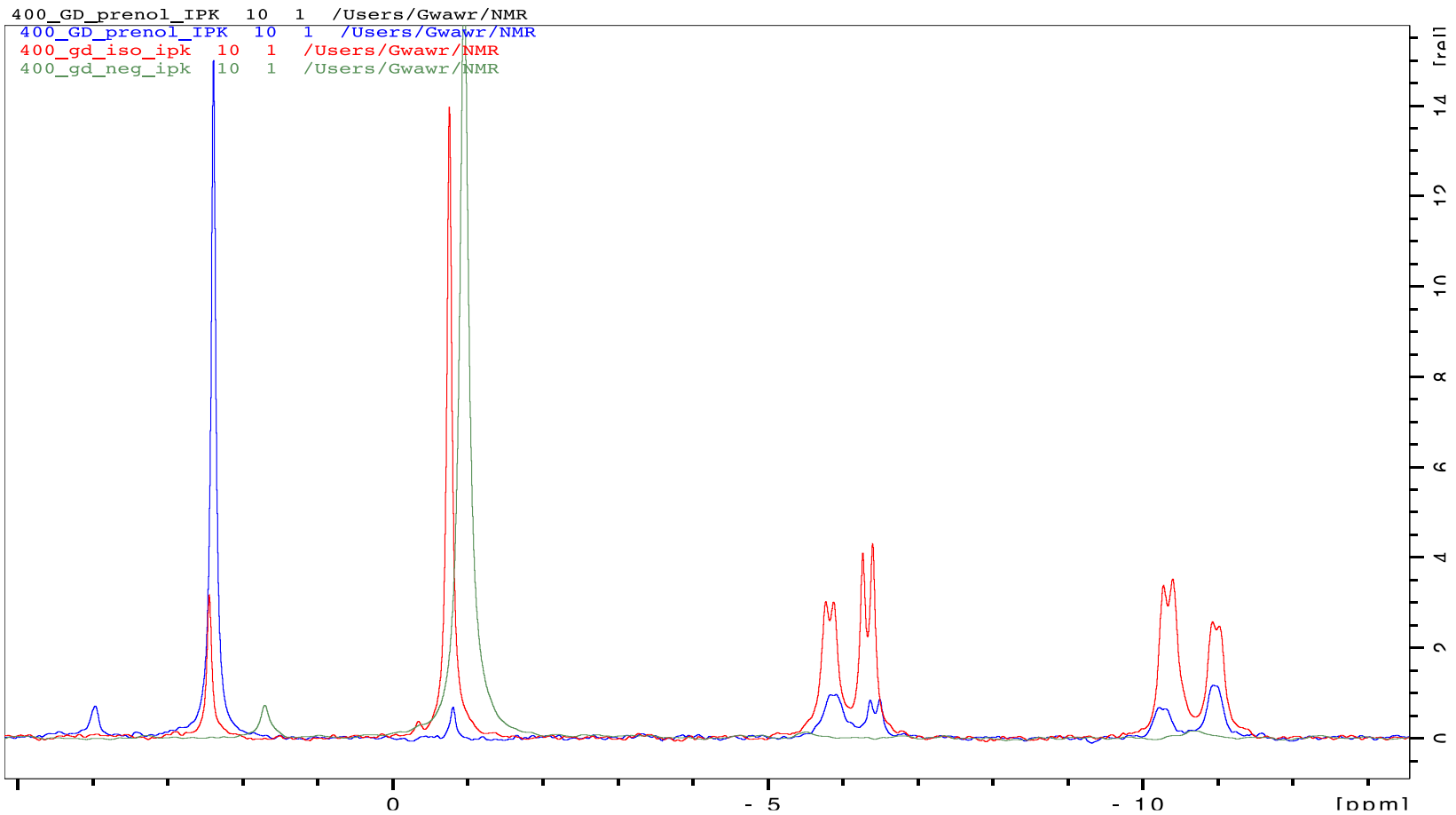


Figure u. ^{31}P NMR spectrum (243 MHz, D_2O, 298K) of ThiM and IPK with: blue – prenol, red – isoprenol, green – no substrate. The peak at: 2.5 ppm is the monophosphorylated product, -0.6 ppm is PEP and the peaks at -5.9 and -11 ppm are ADP; and -6.3 and -10.1 ppm are the diphosphorylated products.

7.2.3.3 Phosphorus NMR of 4-hydroxyprenol monophosphorylated by ThiM, and of prenol monophosphorylated by ThiM for comparison.

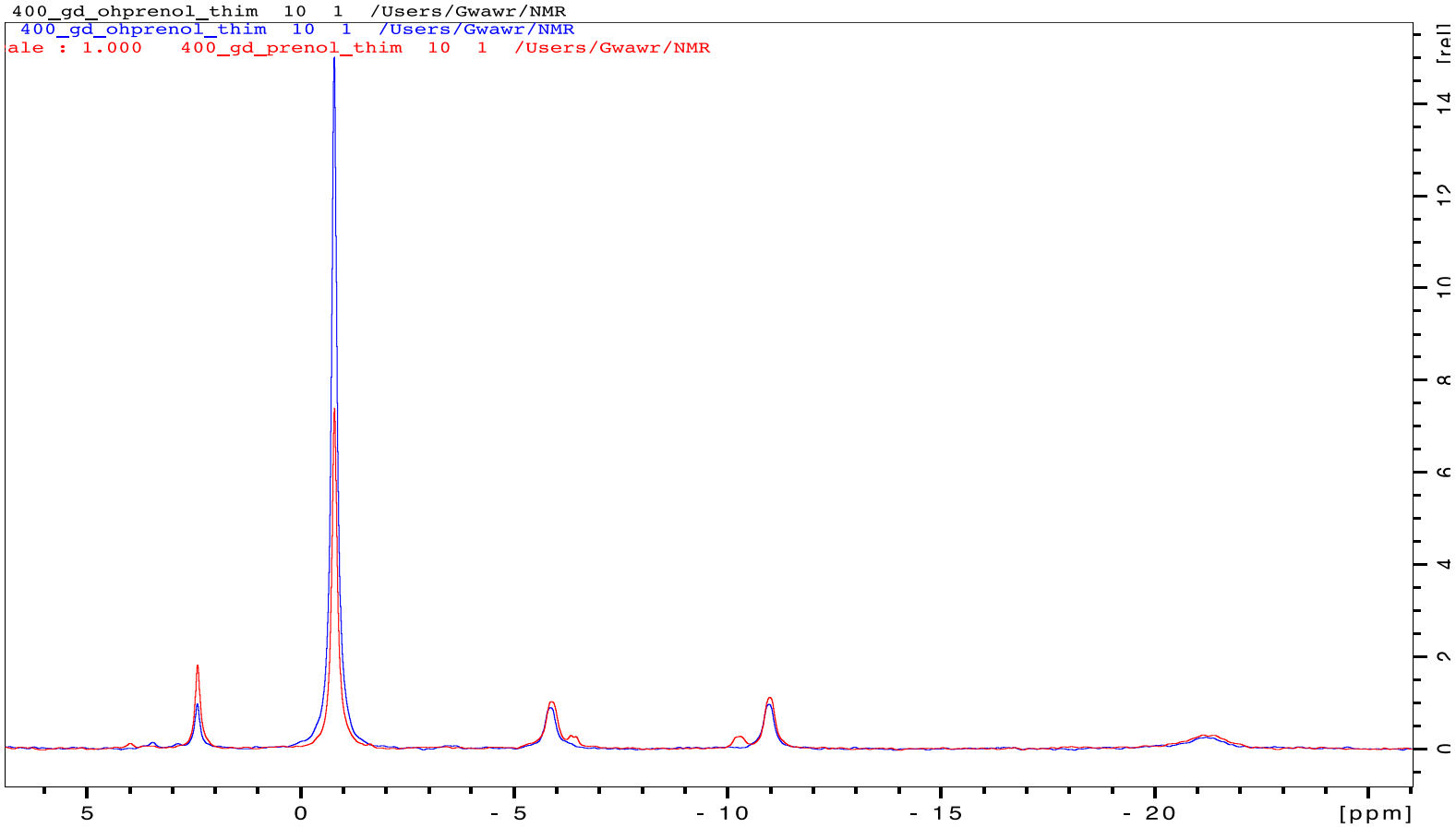


Figure v. ^{31}P NMR spectrum (243 MHz, D_2O, 298K) of ThiM with: red– prenol, blue – 4-hydroxyprenol. The peak at: 2.5 ppm is the monophosphorylated product, -0.6 ppm is PEP, and the peaks at -5.9 and -11 ppm are ADP.

7.2.3.4 Phosphorus NMR of 4-hydroxyprenol diphosphorylated by IPK and of prenol diphosphorylated by IPK for comparison

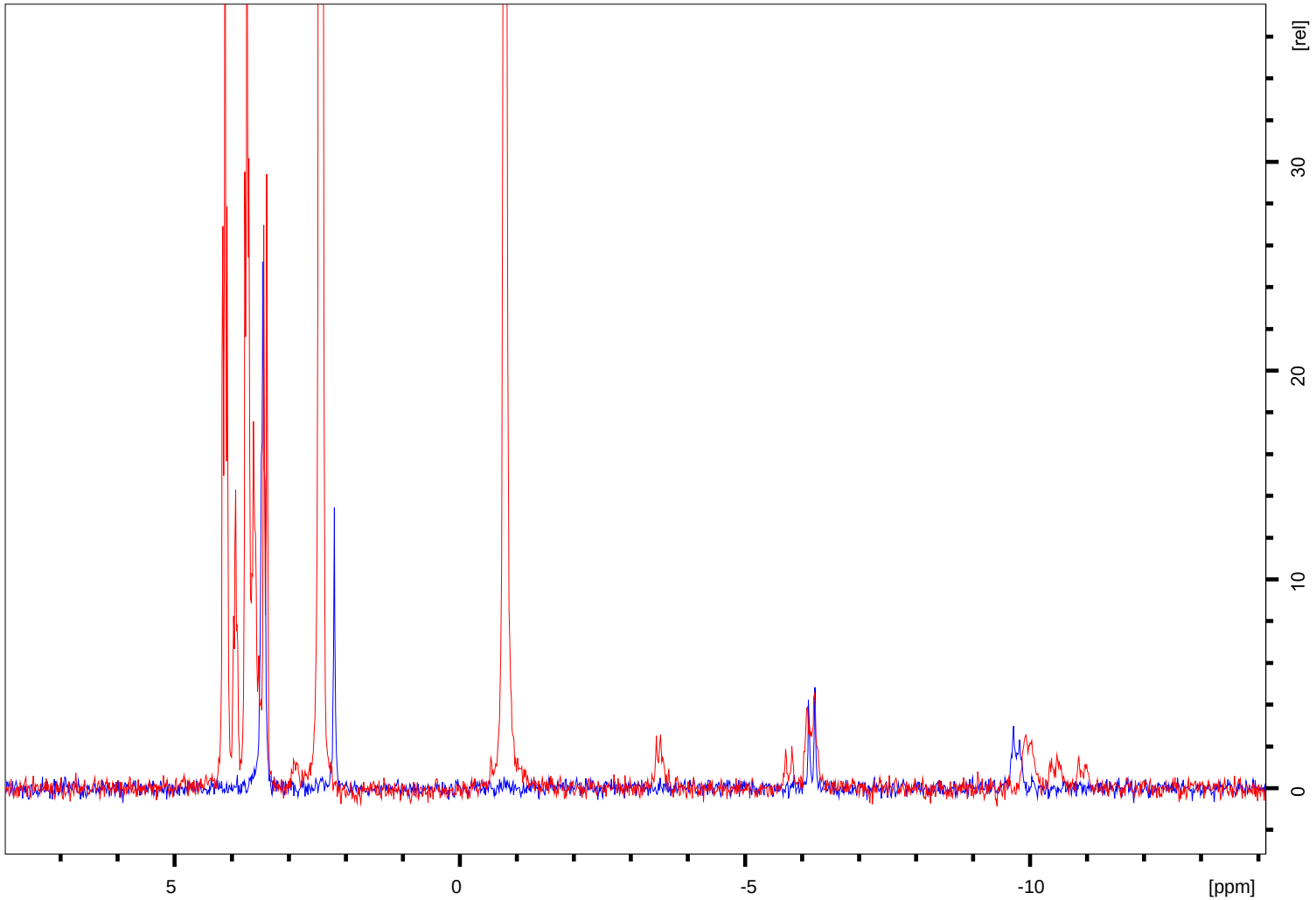


Figure w. ^{31}P NMR spectrum (243 MHz, D_2O, 298K) of ThiM and IPK with: red – 4-hydroxyprenol blue – prenol after 4 hour incubation. The peak at: 3.5 ppm is the monophosphorylated product, 2.3 ppm is free phosphate, -1 ppm is PEP, and the peaks at -6.3 and -9.5 ppm are the diphosphorylated products. In the red spectrum there are peaks at 4.5 ppm which is monophosphorylation of components of the buffer, as 4-hydroxy prenol is an unfavourable substrate , and at -3.5, -6 and -10.5 ppm which are ATP.

7.2.4 Images of mutations of IPK

G59T

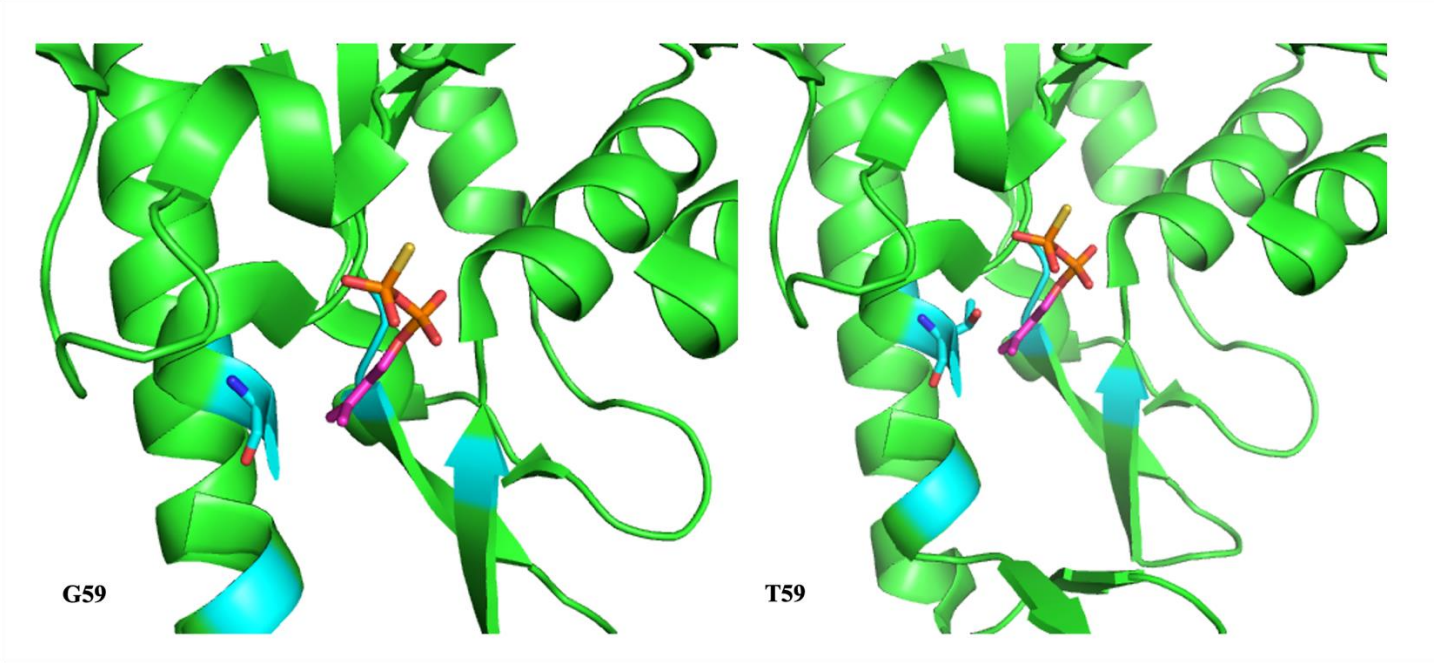


Figure x. Cartoon representation of homology model IPK in green, 3-methylbut-3enyl trihydrogen diphosphate in magenta, and the residue of interest G59 muatetd to threonine in cyan.

M90S

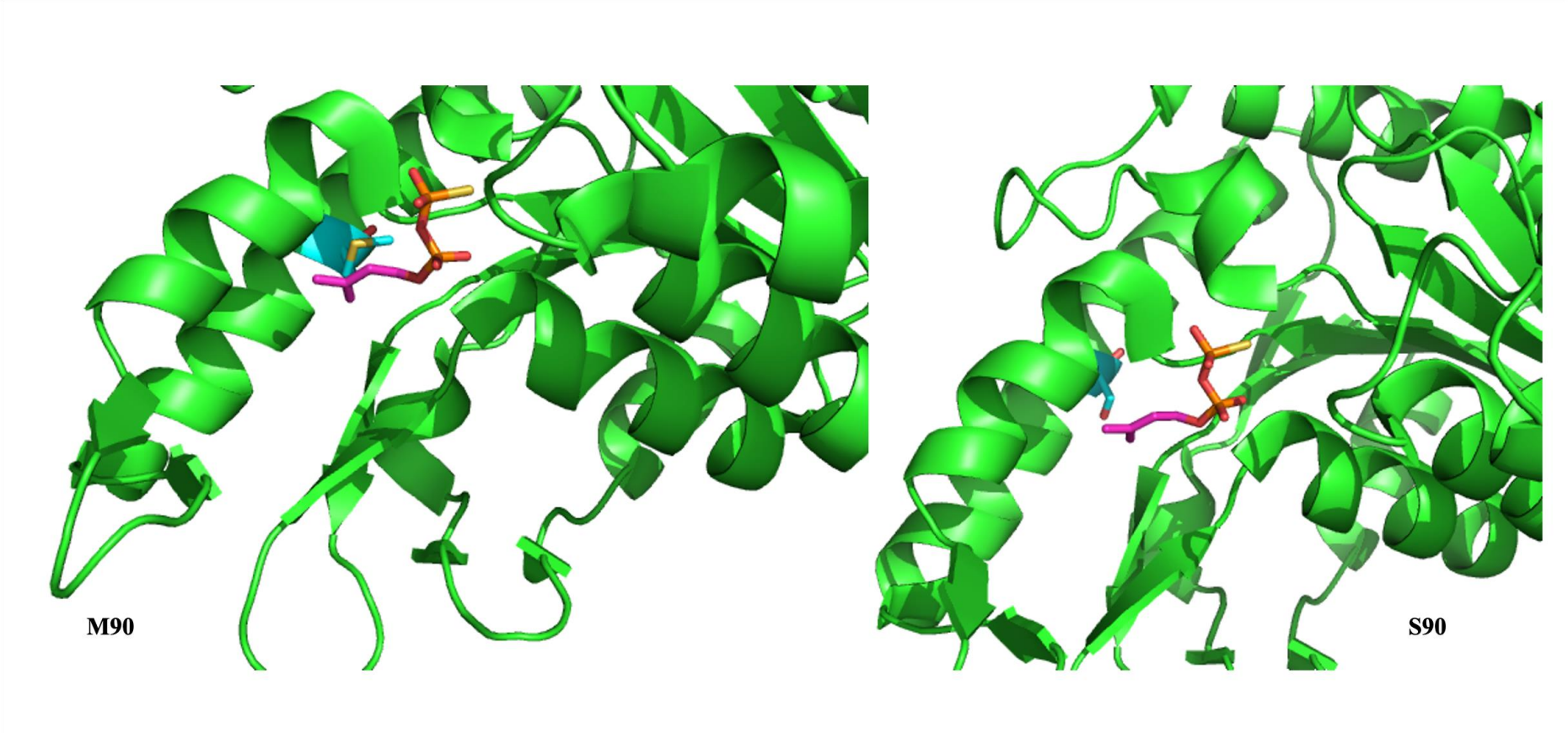


Figure y. Cartoon representation of homology model IPK in green, 3-methylbut-3enyl trihydrogen diphosphate in magenta, and the residue of interest, M90, mutated to serine, in cyan.

G144S

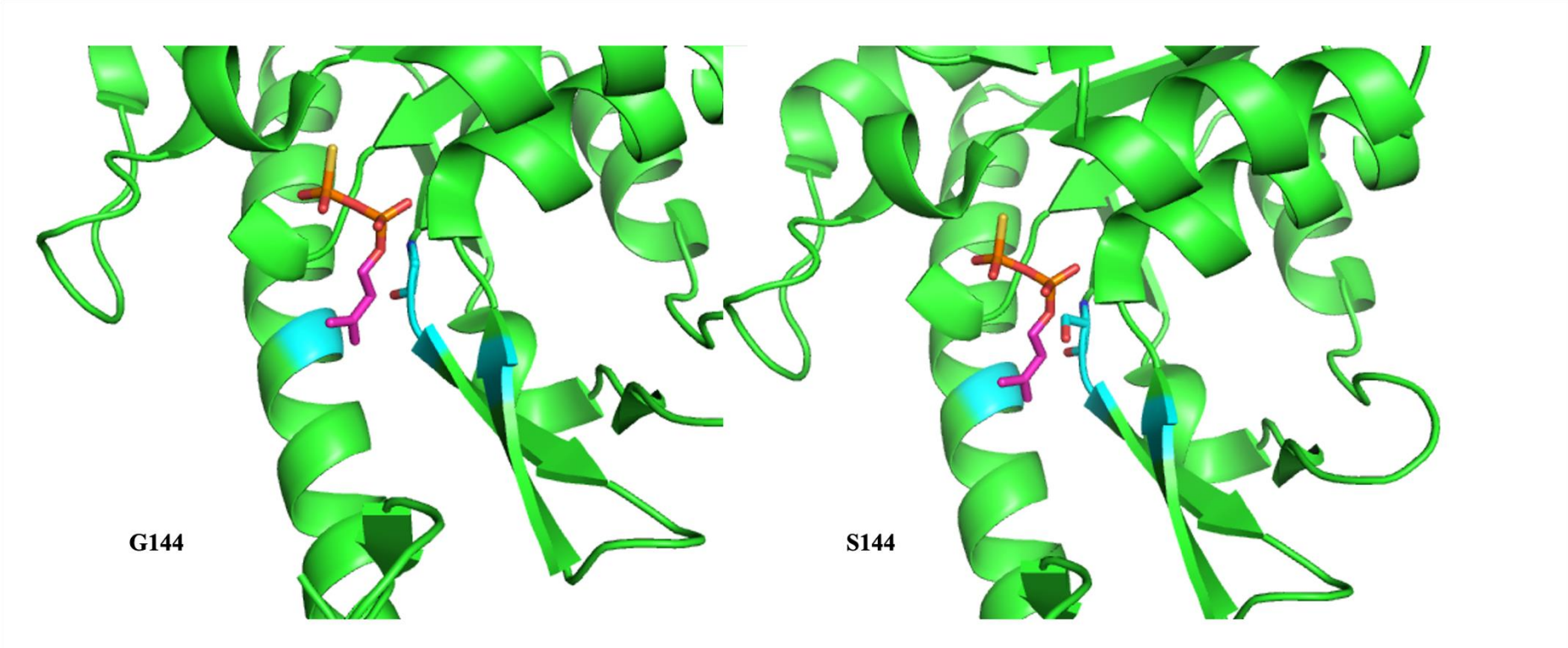


Figure z. Cartoon representation of homology model IPK in green, 3-methylbut-3enyl trihydrogen diphosphate in magenta, and the residue of interest, G144, mutated to serine, in cyan.

D145S

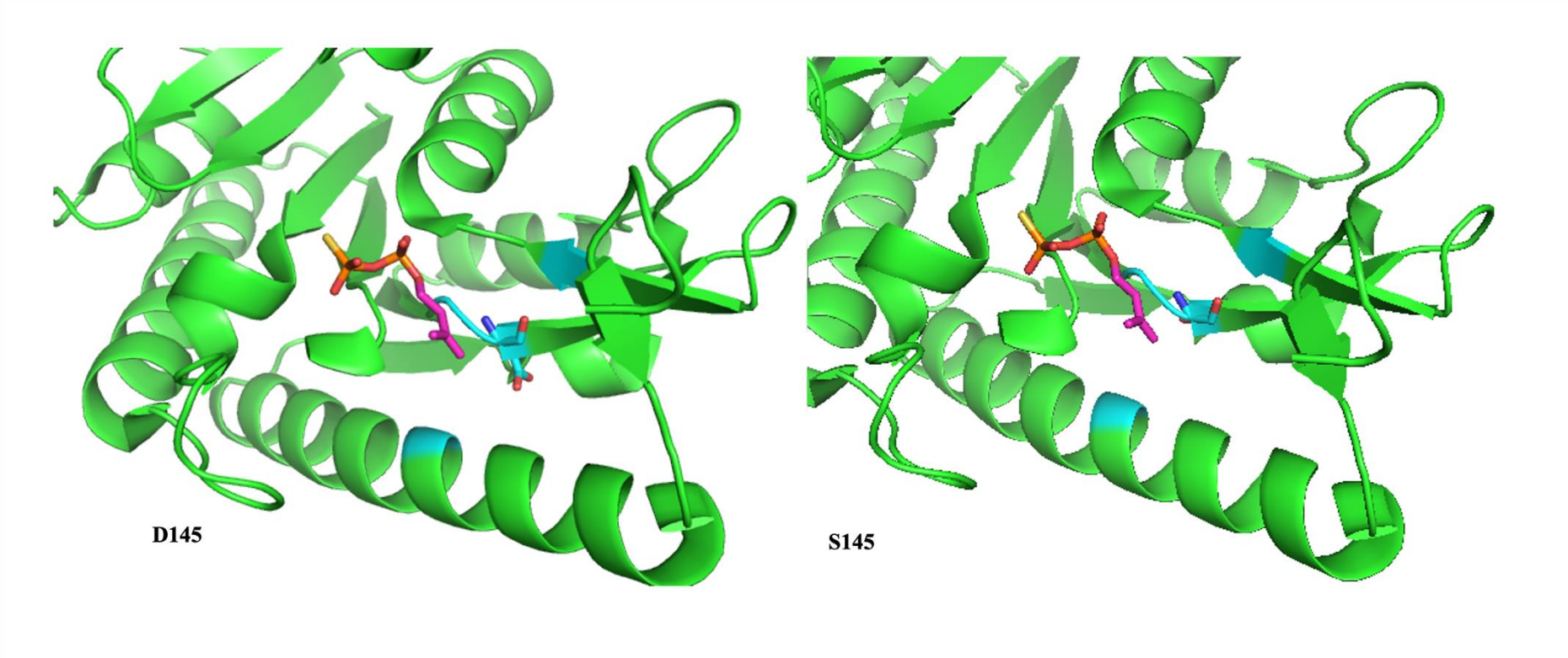


Figure aa. Cartoon representation of homology model IPK in green, 3-methylbut-3enyl trihydrogen diphosphate in magenta, and the residue of interest, D145, mutated to serine, in cyan

I156T

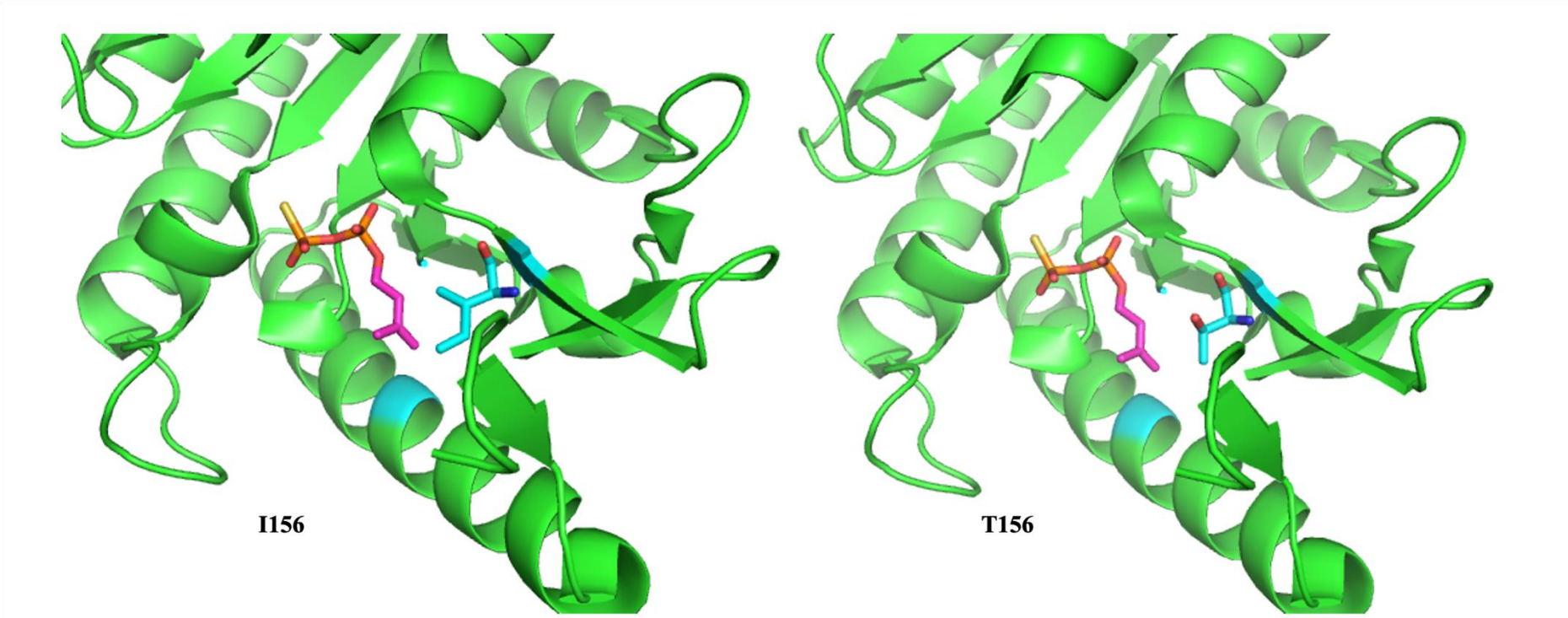


Figure bb. Cartoon representation of homology model IPK in green, 3-methylbut-3enyl trihydrogen diphosphate in magenta, and the residue of interest, I156, mutated to threonine, in cyan.

7.3 Chapter 4

7.3.1 12-OH farnesol

7.3.1.1 Hydrogen NMR of 12OH-farnesol

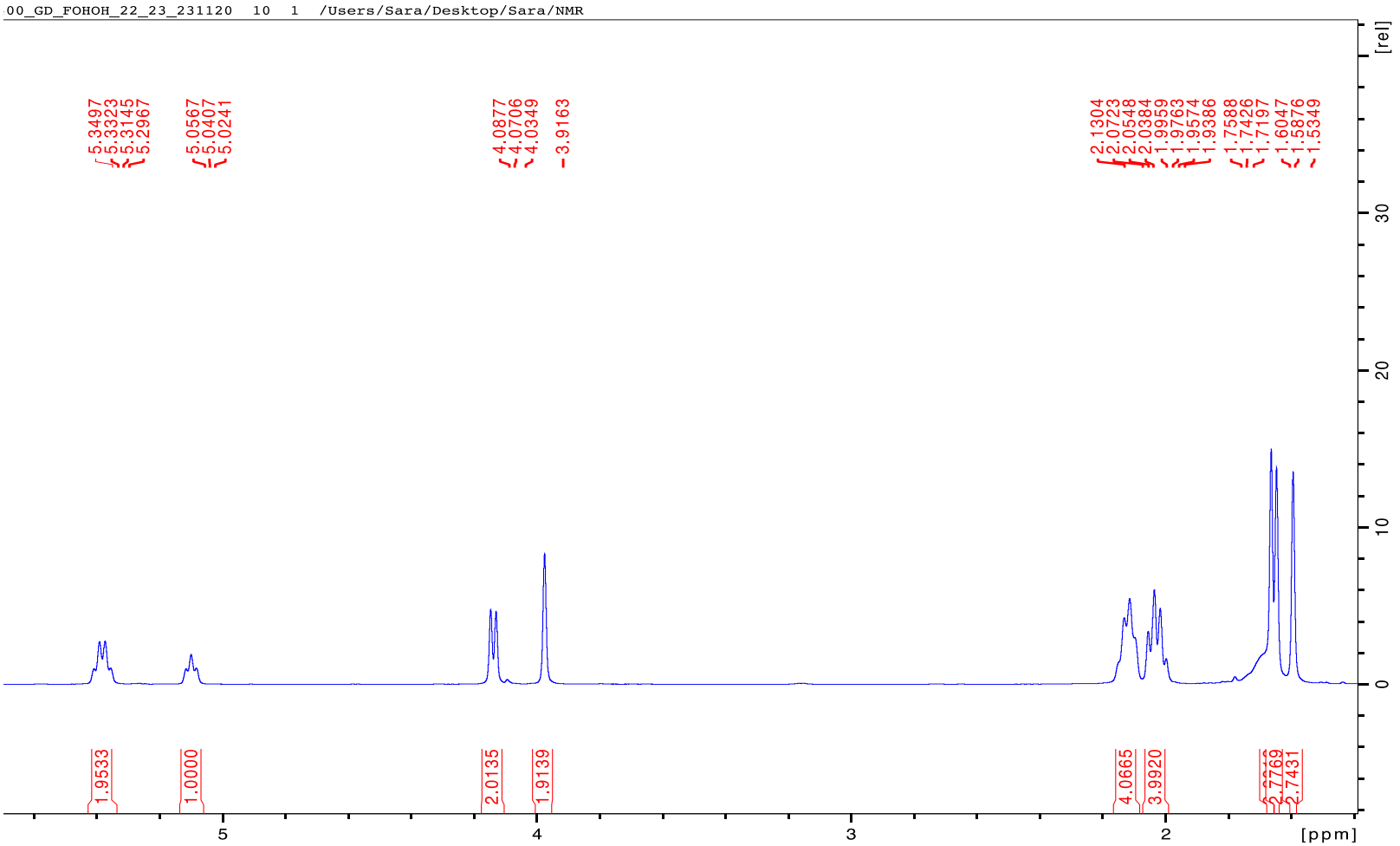


Figure cc. ^{1}H NMR spectrum (400 MHz, $CDCl_3$, 298K) of 12-hydroxy farnesol

7.3.1.2 Carbon NMR of 12OH-farnesol

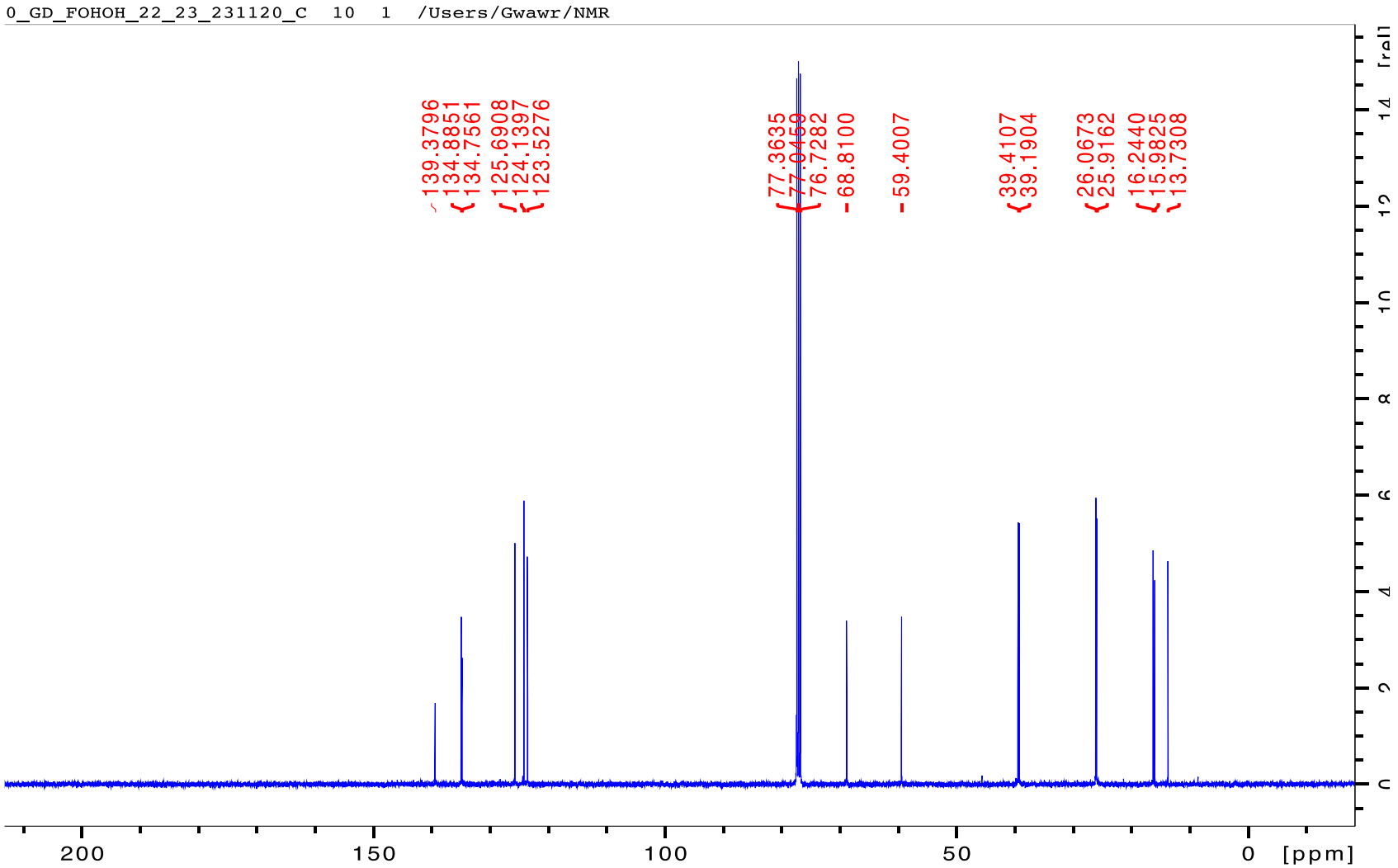


Figure dd. ^{13}C NMR spectrum (100 MHz, $CDCl_3$, 298K) of 12-hydroxy farnesol

7.3.1.3 Mass spectrum of 12OH-farnesol

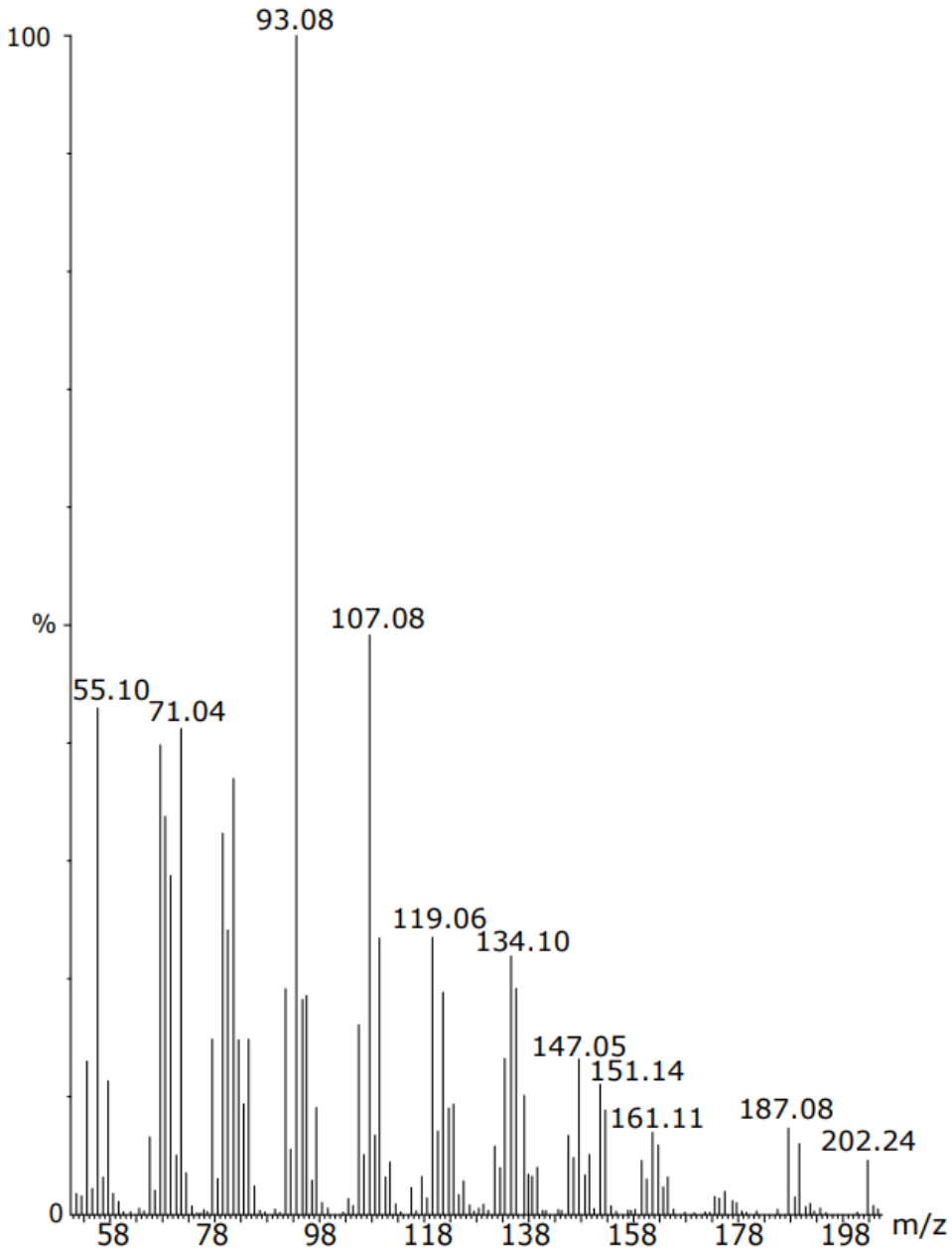


Figure ee. Mass spectrum of peak at 17.06 min of total ion chromatogram of the synthesised compound 12-hydroxy farnesol.

7.3.2 Bromofarnesol

7.3.2.1 Hydrogen NMR of bromo farnesol

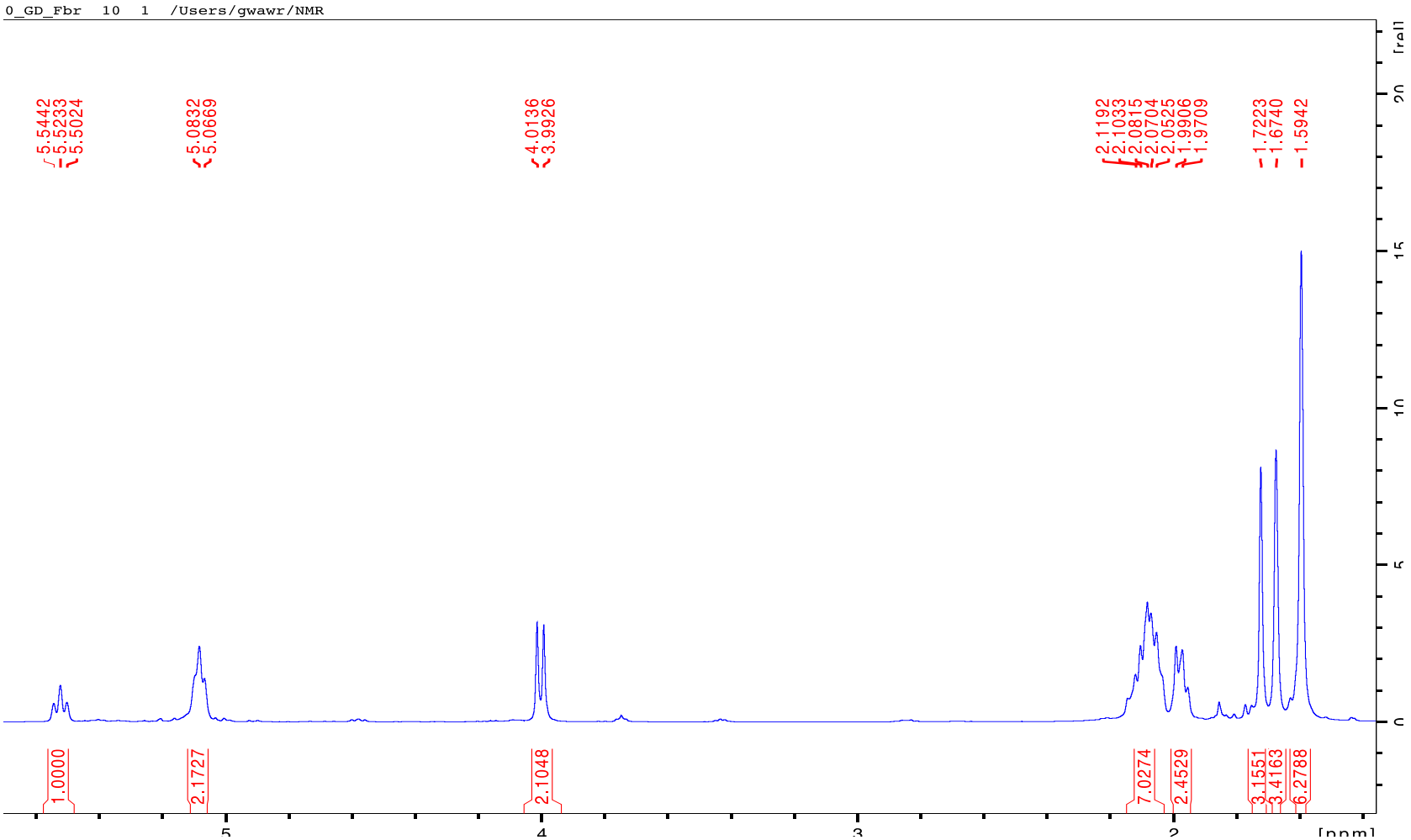


Figure ff. ^{1}H NMR spectrum (400 MHz, $CDCl_3$, 298K) of bromofarnesol

7.3.2.2 Carbon NMR of bromo farnesol

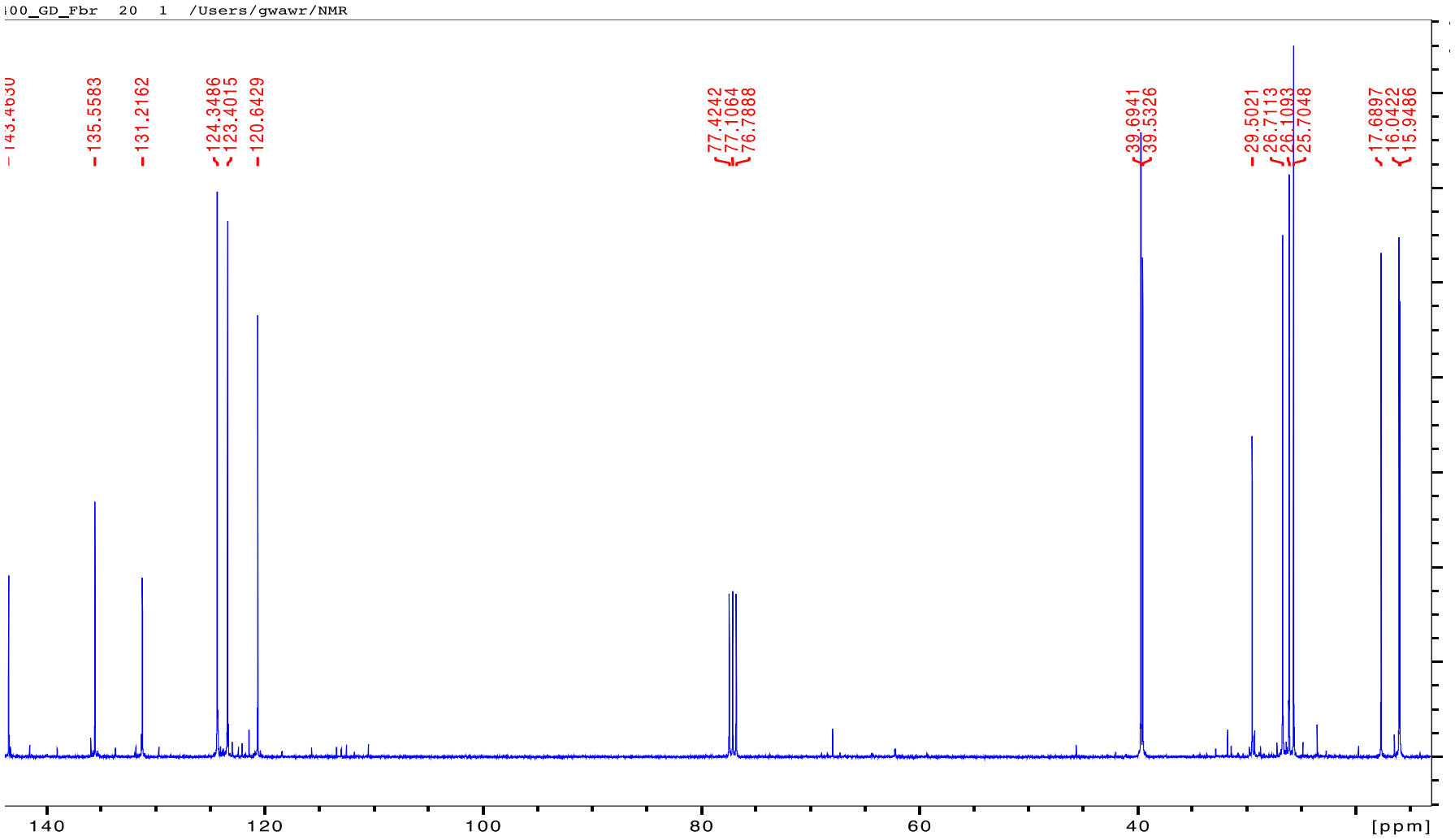


Figure gg. ^{13}C NMR spectrum (100 MHz, CDCl3, 298K) of bromofarnesol

7.3.2.3 Mass spec of bromo farnesol

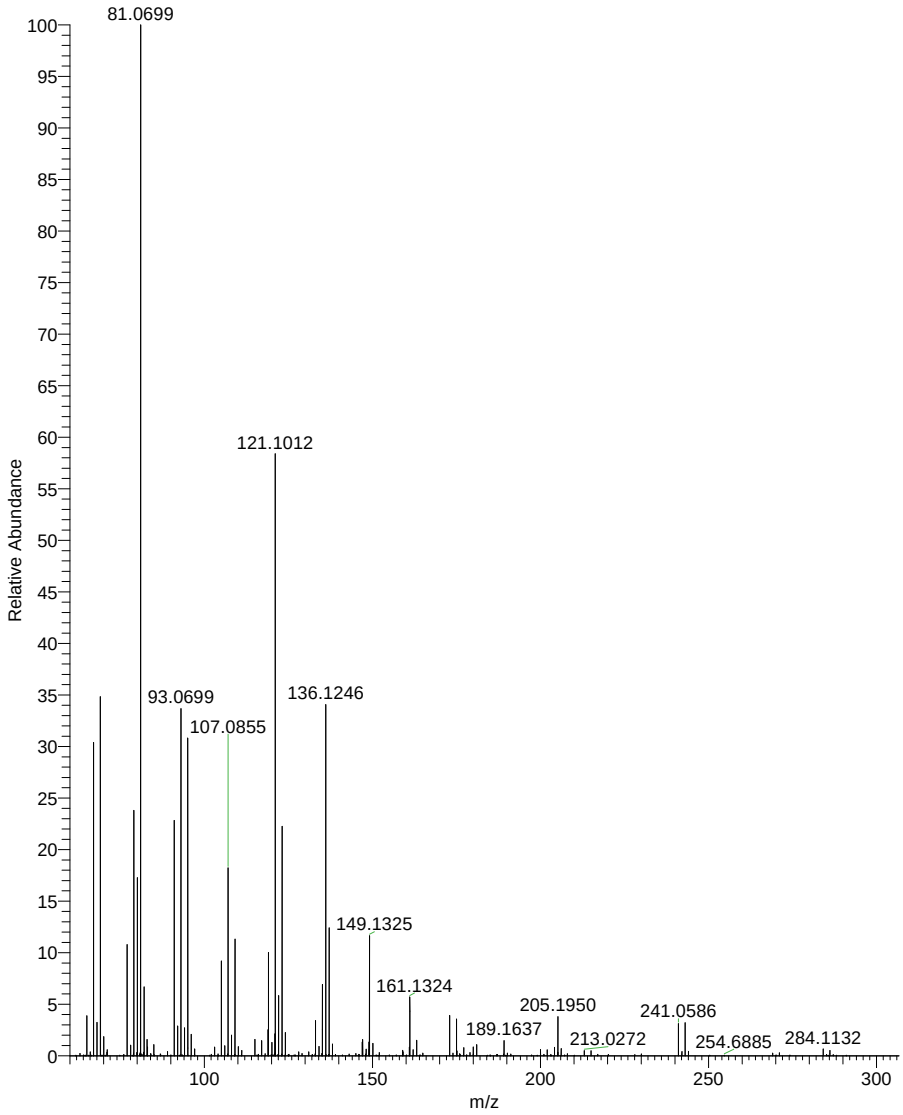


Figure hh. High resolution mass spectrum of peak at 0.69 min of total ion chromatogram of the synthesised compound bromofarnesol

7.3.3 Farnesyldiphosphate (FDP)

7.3.3.1 Hydrogen NMR of FDP

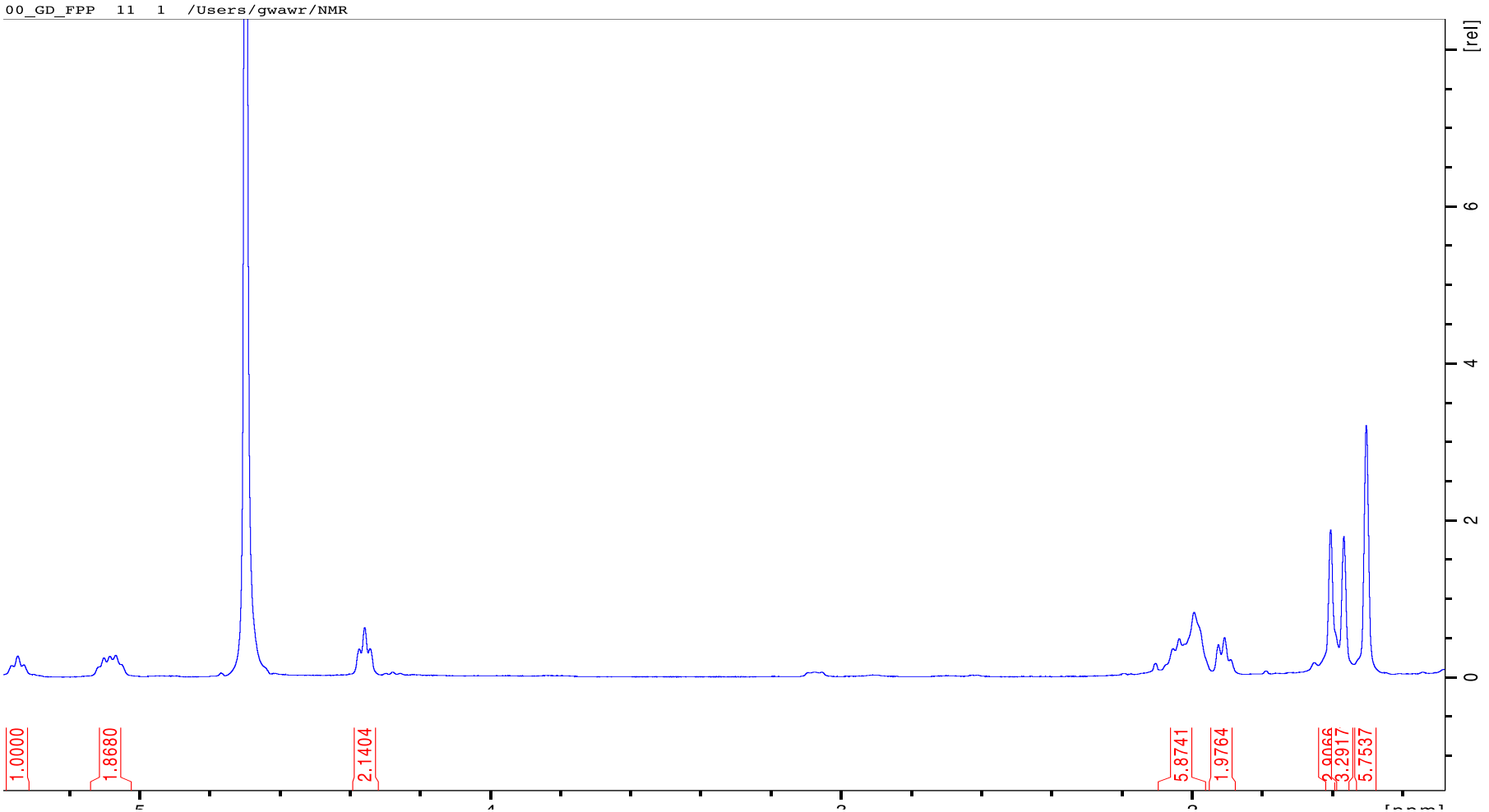


Figure ii. ^{1}H NMR spectrum (400 MHz, CDCl3, 298K) of farnesyldiphosphate

7.3.3.2 Carbon NMR of FDP

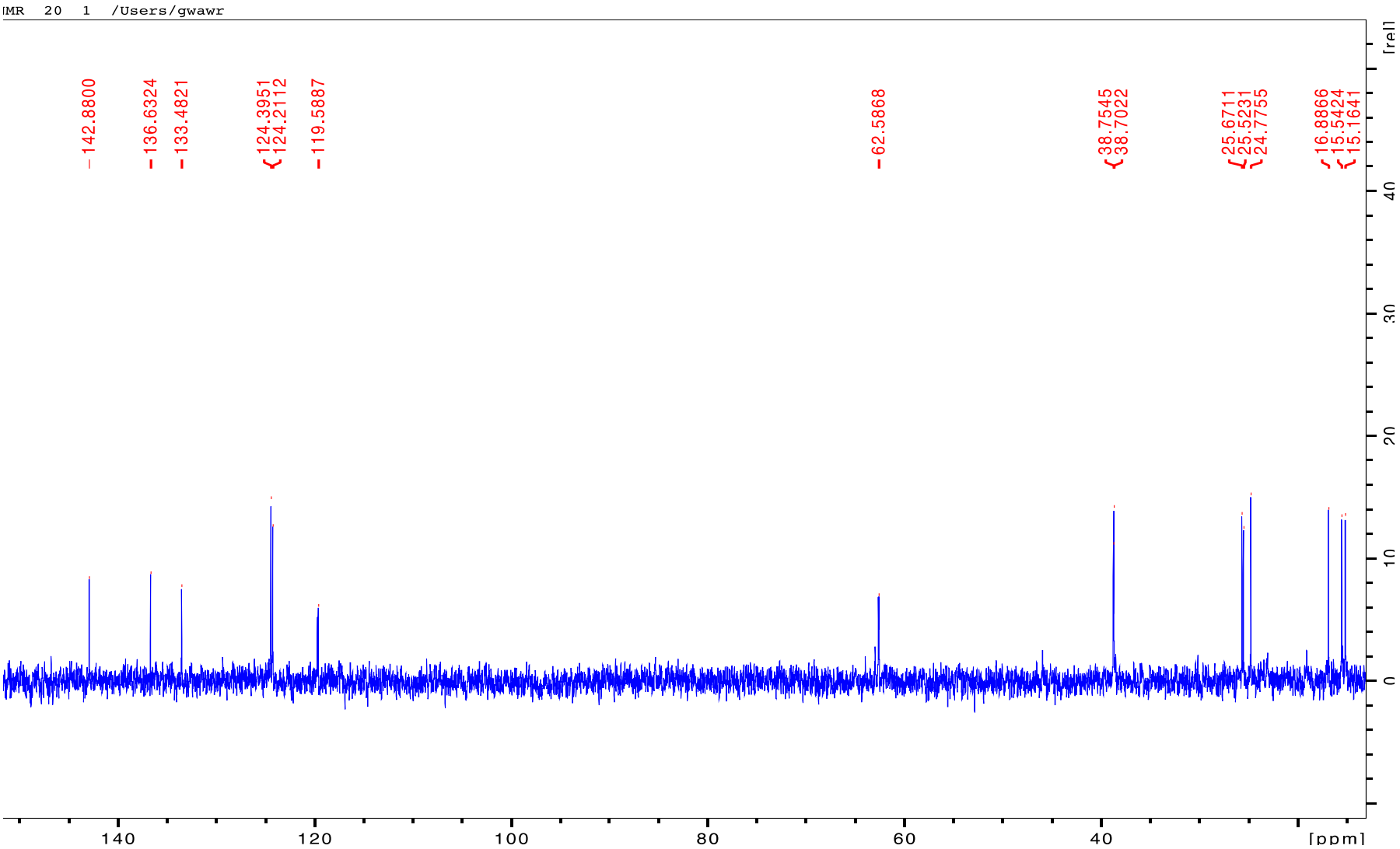


Figure jj. ^{13}C NMR spectrum (100 MHz, CDCl3, 298K) of farnesyldiphosphate

7.3.3.3 Phosphorus NMR of FDP

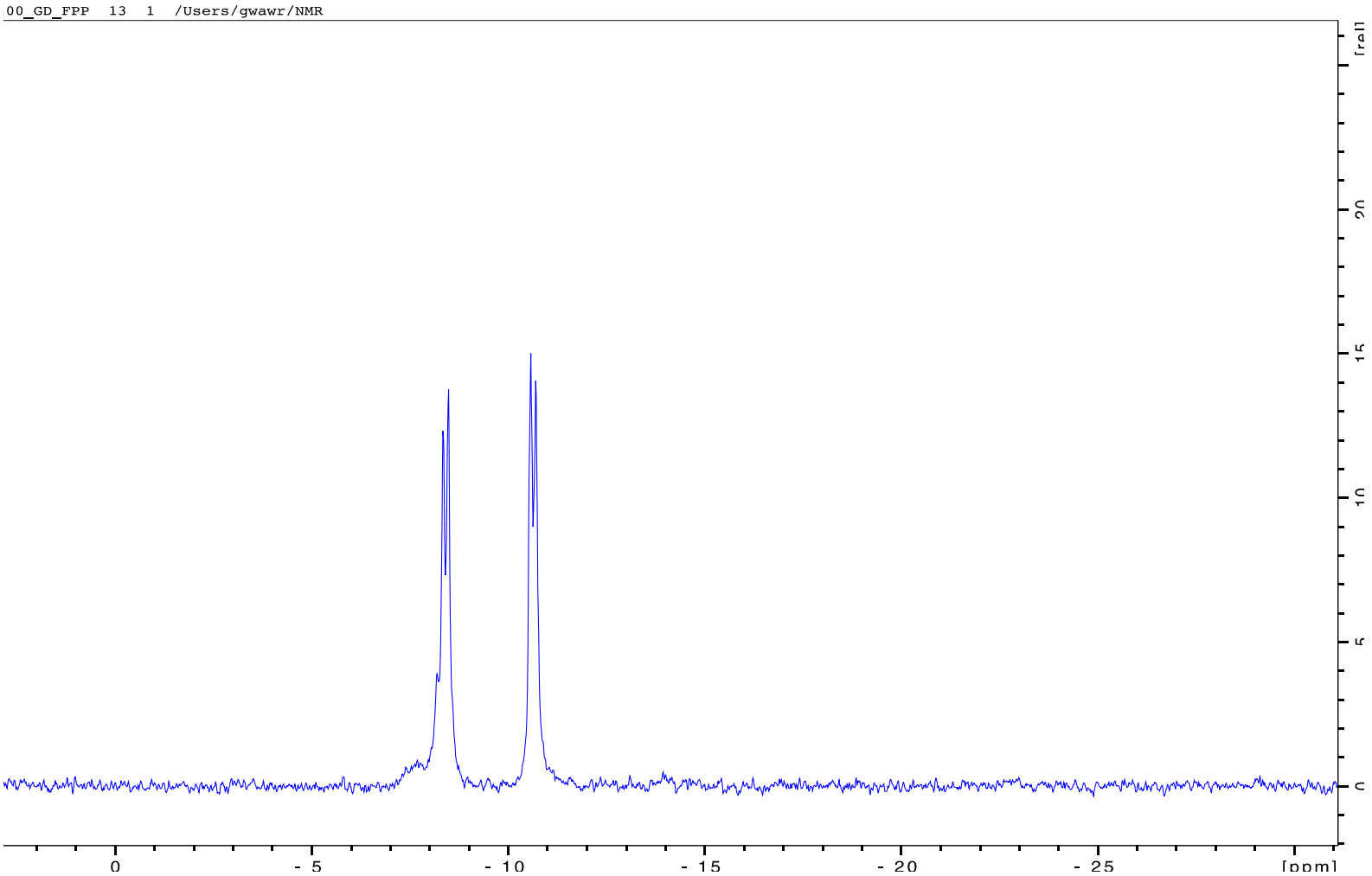


Figure kk. ^{31}P NMR spectrum (243 MHz, D_2O, 298K) of farnesyldiphosphate

7.3.3.4 Mass spectrum of FDP

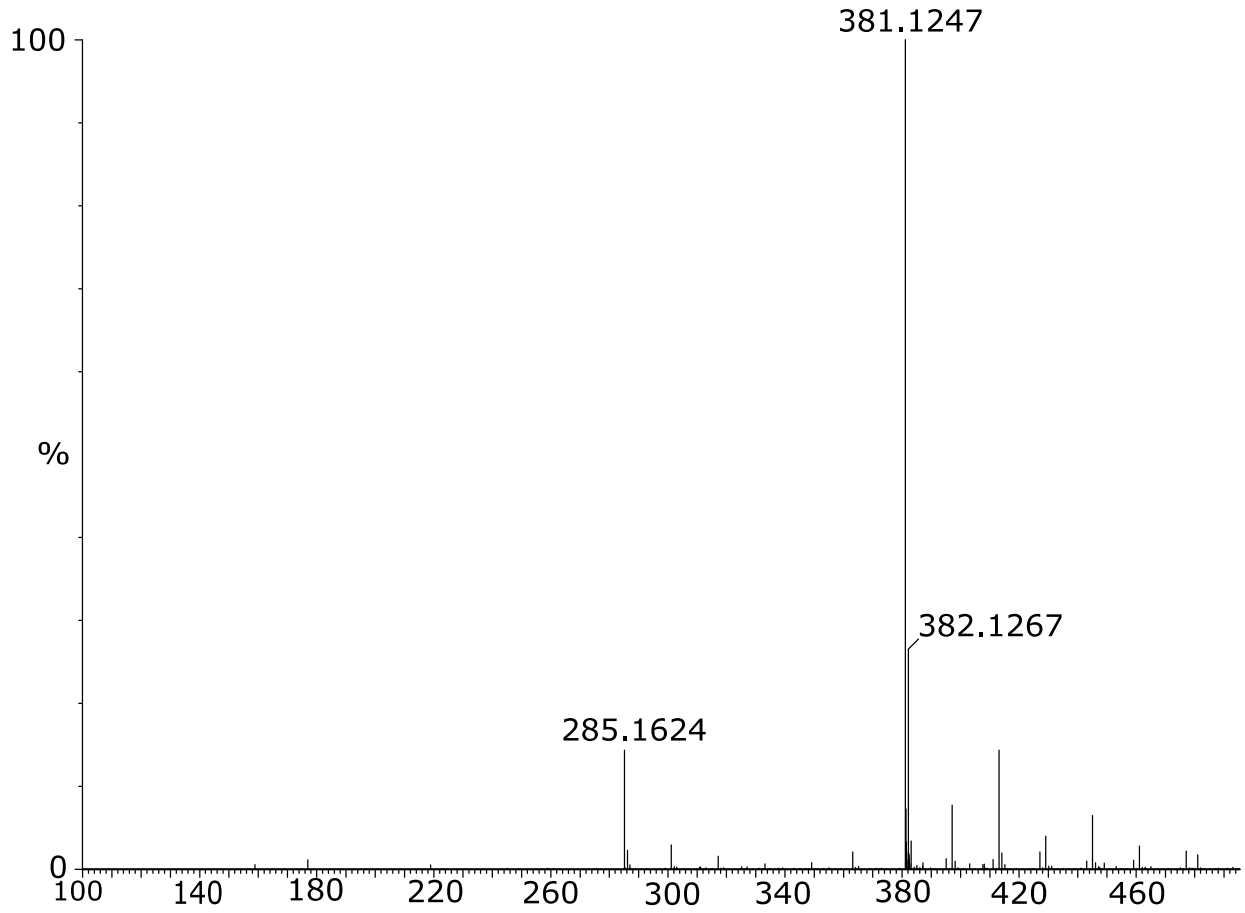


Figure ll. High resolution mass spectrum of peak at 0.79 min of total ion chromatogram of the synthesised compound farnesyldiphosphate

7.3.4 Protein purification

7.3.4.1 SDS-PAGE gel of CYP124A1 wt with no expression

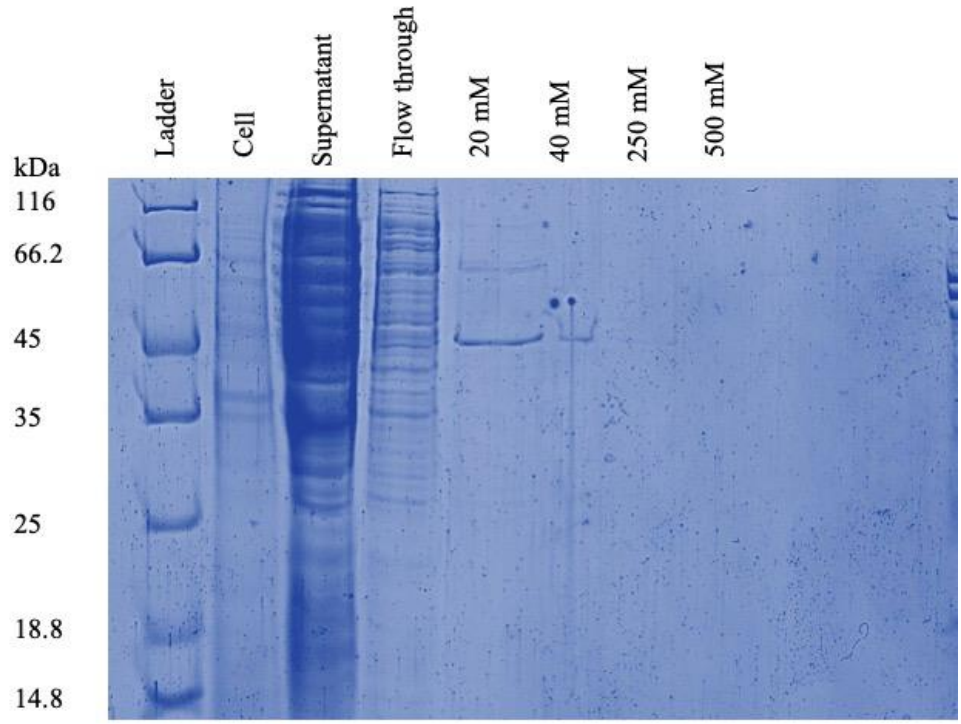


Figure mm. Example SDS-PAGE gel of nickel column purification of an attempted expression of CYP124A1 showing no protein was obtained

7.3.4.2 SDS-PAGE gel of CYP124A1 wt

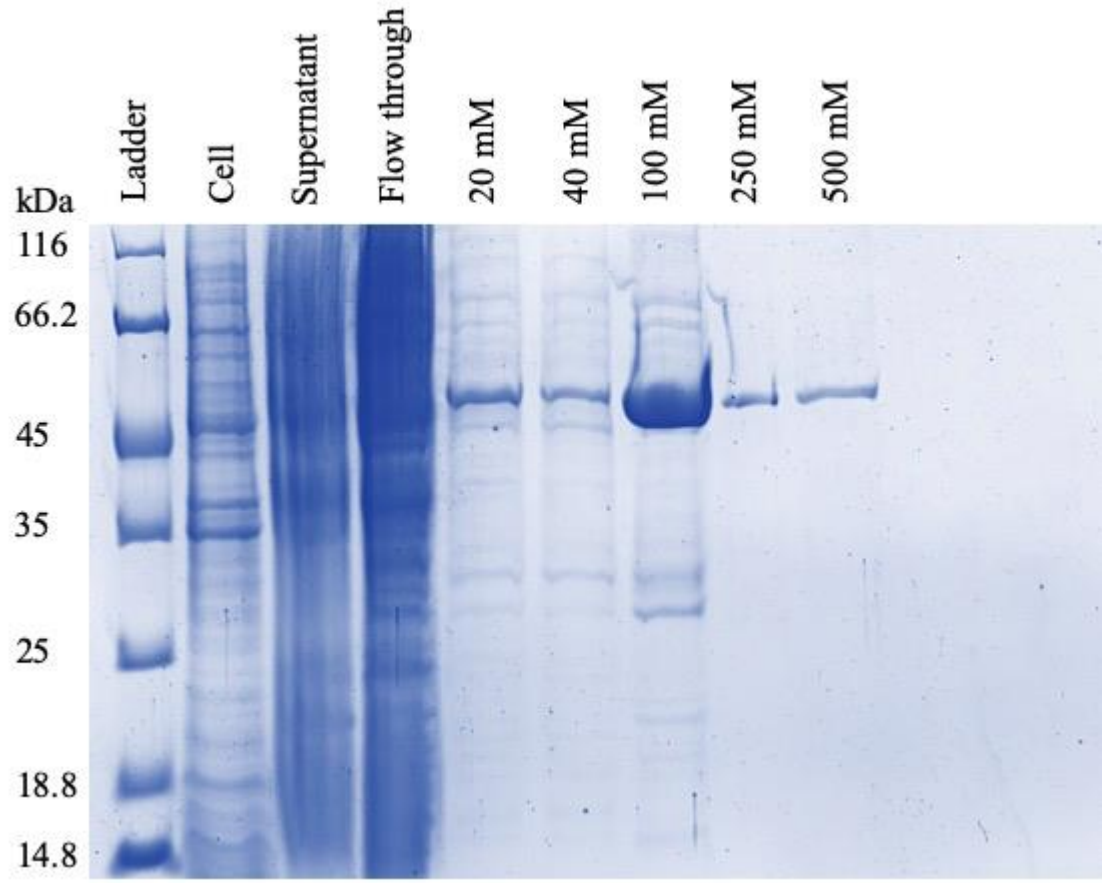


Figure nn. Example SDS-PAGE gel of nickel column purification of expression of CYP124A1 in media containing 1% glucose

7.3.4.3 SDS-PAGE gel of UPK

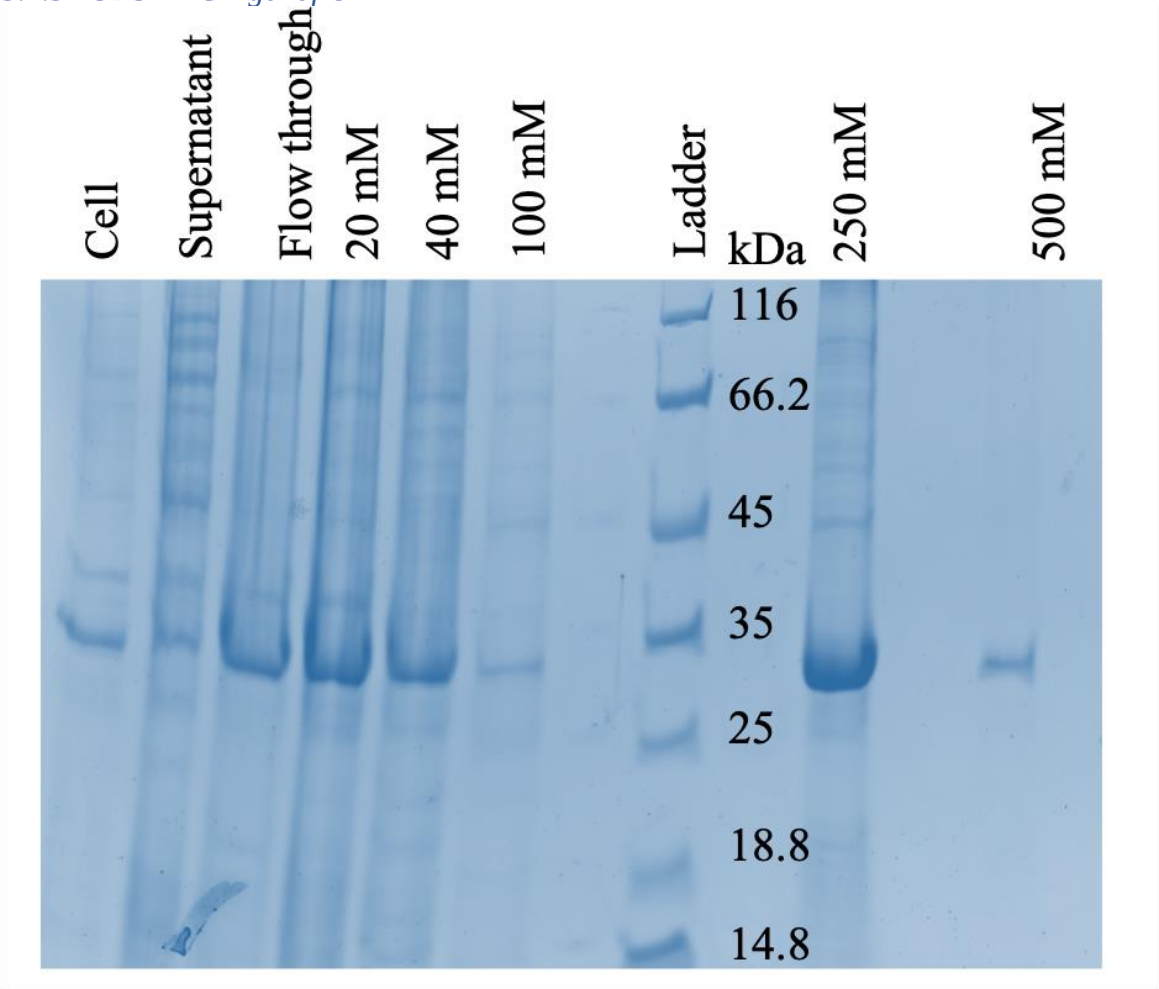


Figure oo. Examples SDS-PAGE gel of nickel column purification of expression of UPK

7.3.5 Images of mutations of CYP124A1

T101S

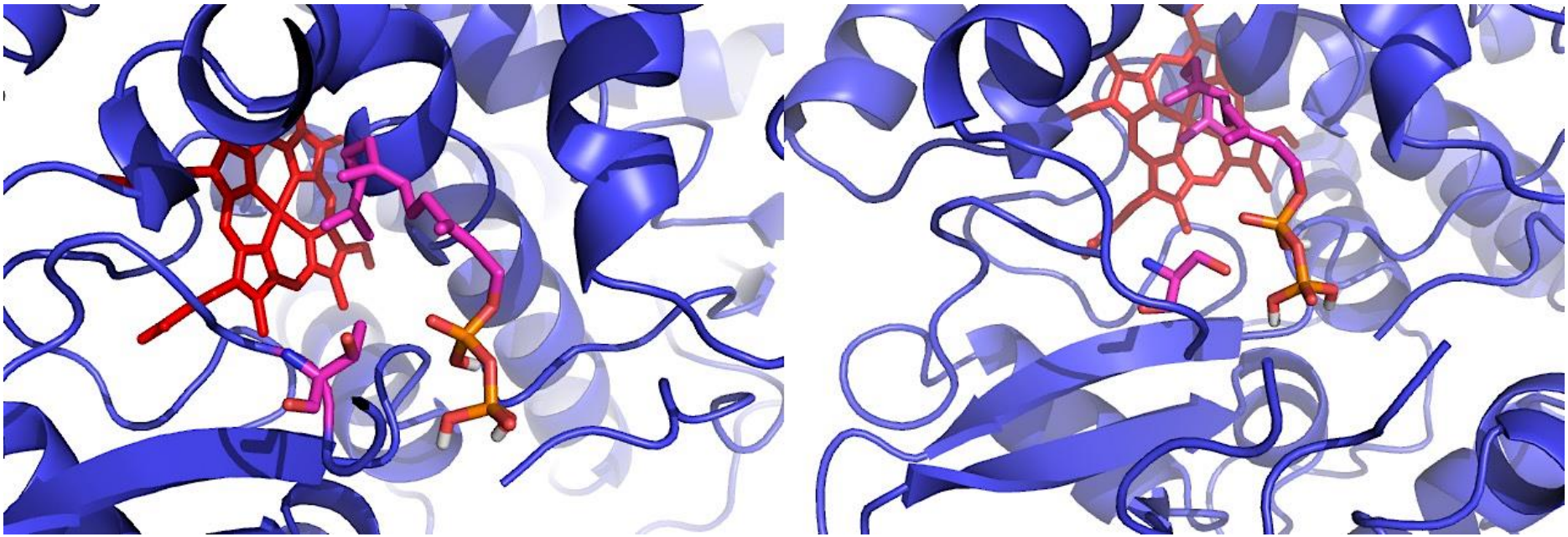


Figure pp. Cartoon representations of a crystal structure of CYP124A1 in blue, with the heme cofactor in red, and a molecule of FDP in magenta modelled into the active site. On the right the threonine residue 101 is in magenta. On the left this residue has been mutated to serine, also in magenta, showing increased space in the active site while maintaining the hydroxy group for hydrogen bonding.

N103S

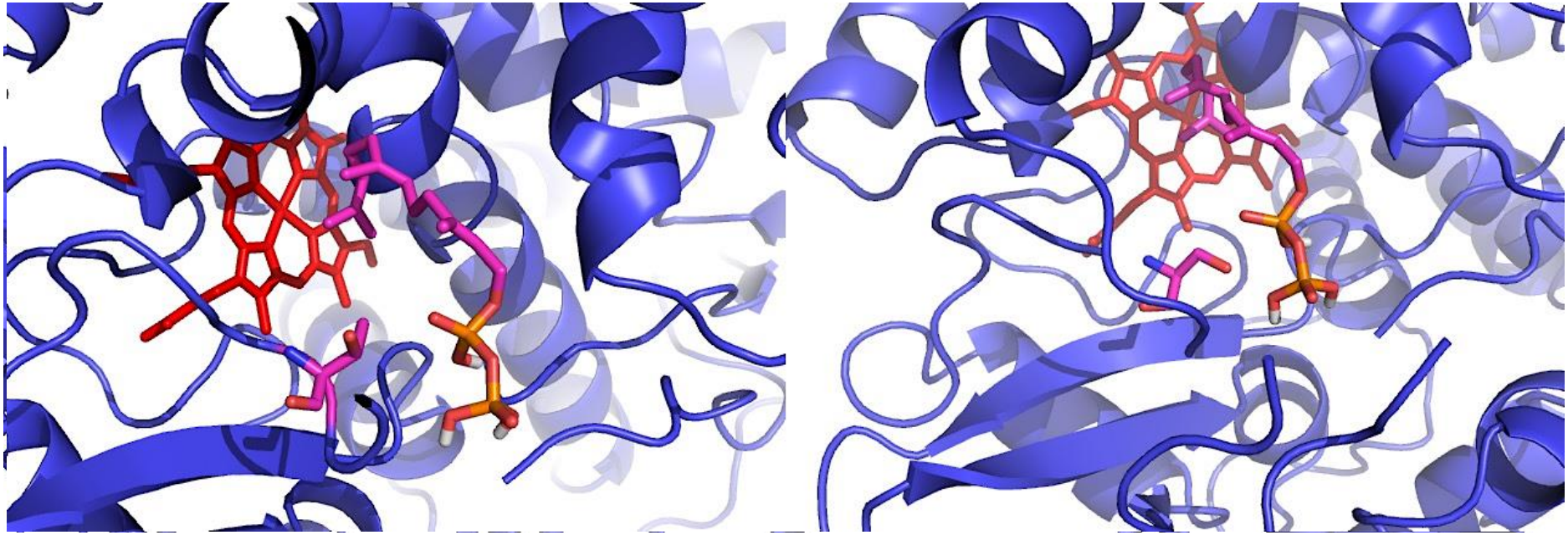


Figure qq. Cartoon representations of a crystal structure of CYP124A1 in blue, with the heme cofactor in red, and a molecule of FDP in magenta modelled into the active site. On the right the asparagine residue 103 is in magenta. On the left this residue has been mutated to serine, also in magenta, showing increased space in the active site.

D104S

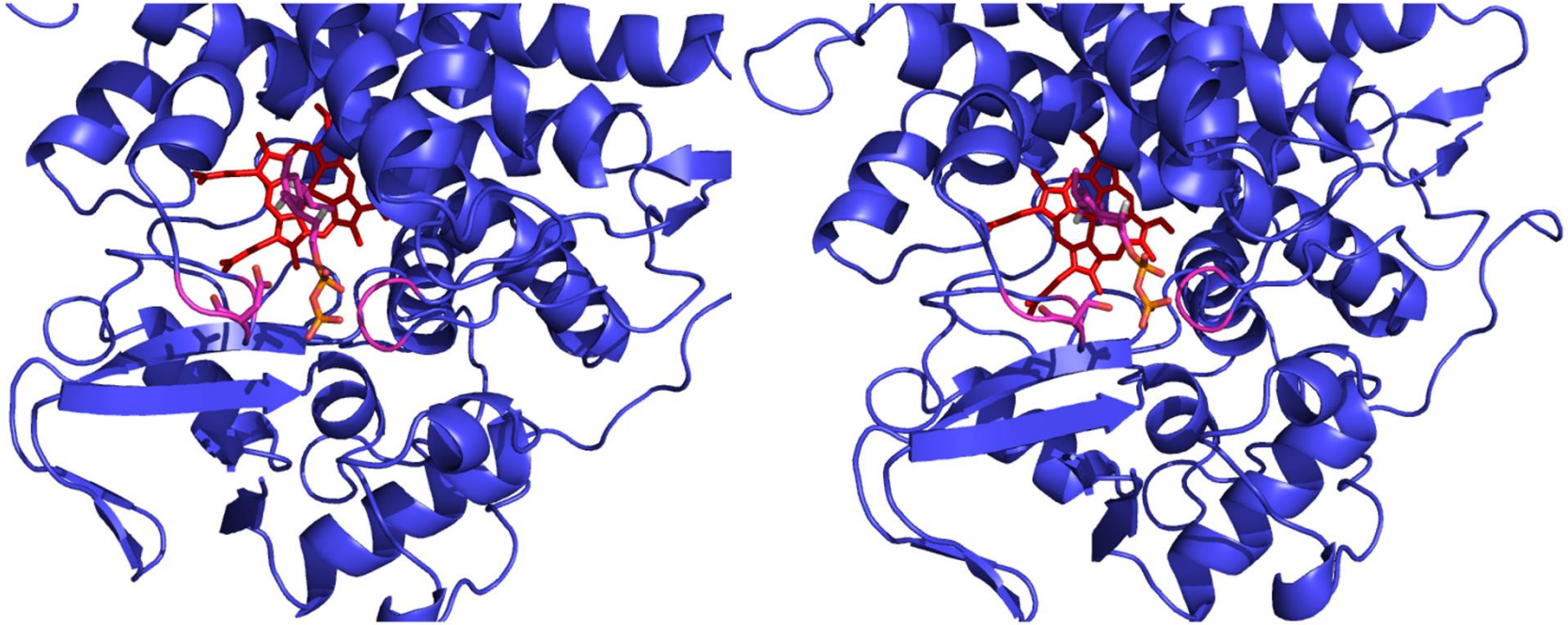


Figure rr.Cartoon representations of a crystal structure of CYP124A1 in blue, with the heme cofactor in red, and a molecule of FDP in magenta modelled into the active site. On the right the aspartate residue 104 is in magenta. On the left this residue has been mutated to serine, also in magenta, showing increased space in the active site.

F422R

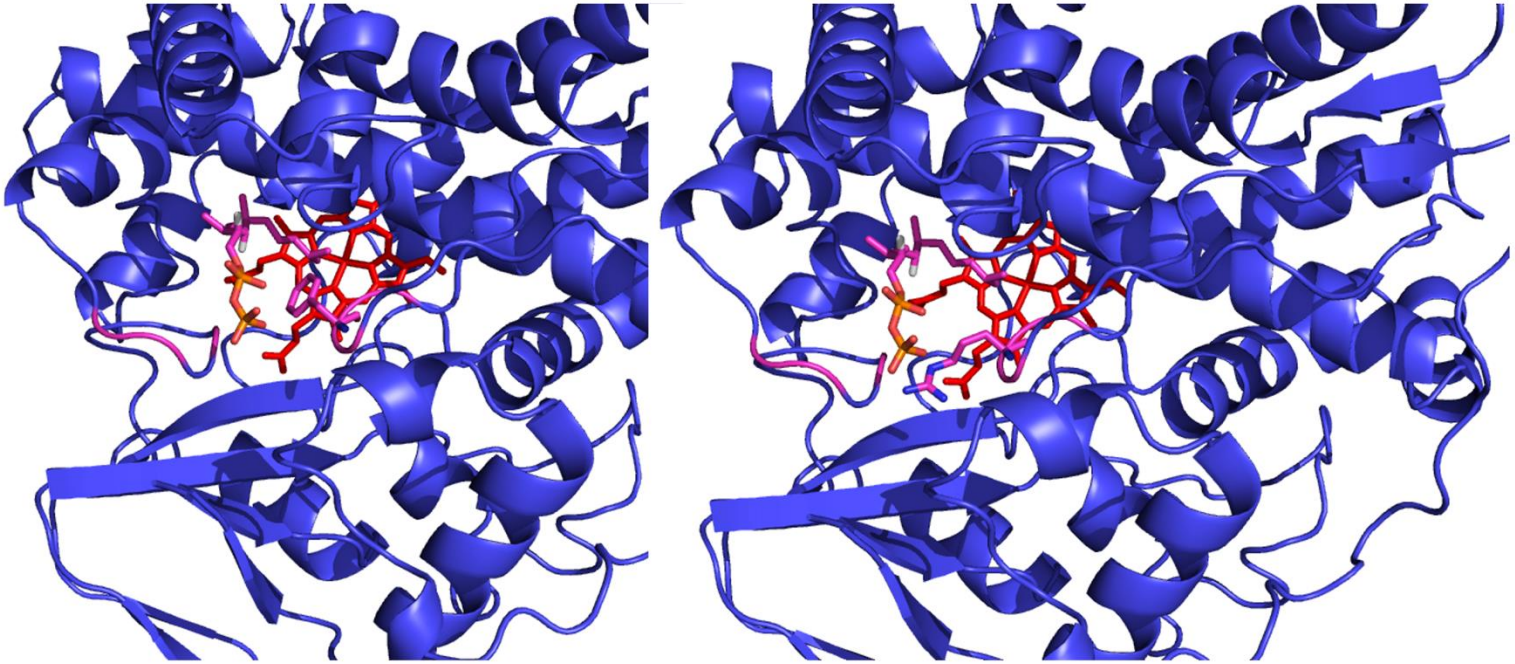


Figure ss. Cartoon representations of a crystal structure of CYP124A1 in blue, with the heme cofactor in red, and a molecule of FDP in magenta modelled into the active site. On the right the phenylalanine residue 422 is in magenta. On the left this residue has been mutated to arginine, also in magenta, showing increased space in the active site.

Q421L

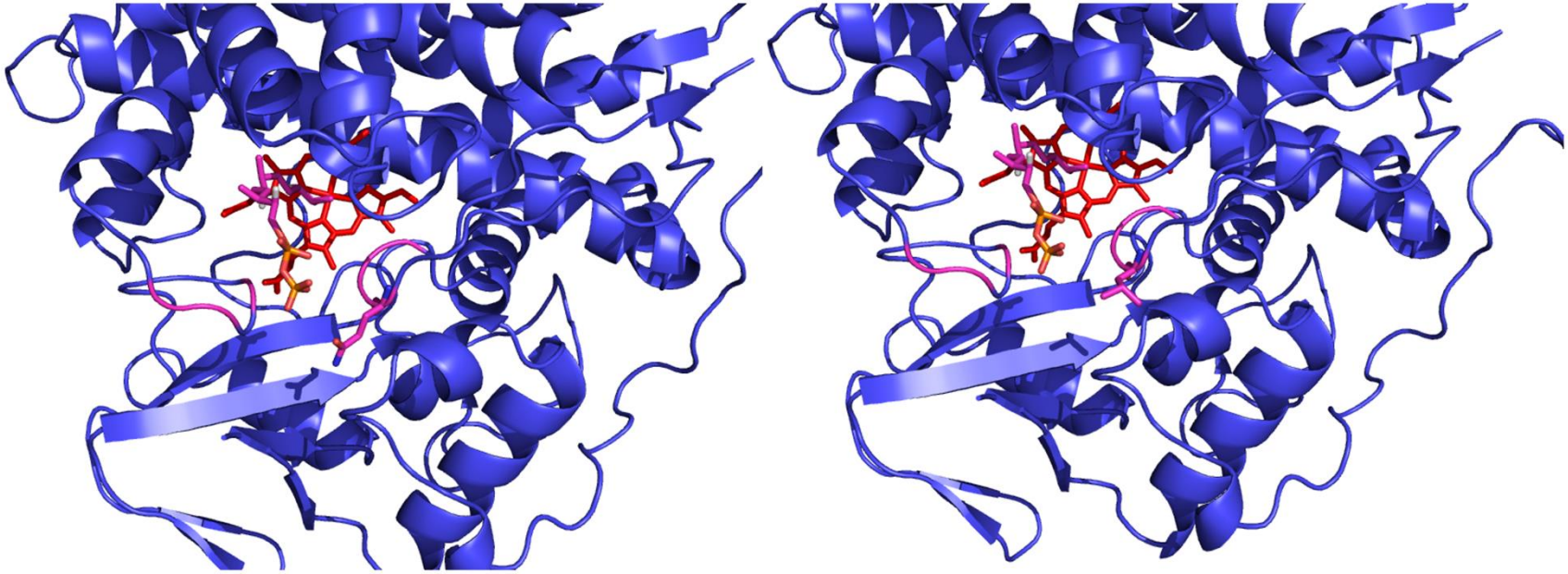


Figure tt. Cartoon representations of a crystal structure of CYP124A1 in blue, with the heme cofactor in red, and a molecule of FDP in magenta modelled into the active site. On the right the glutamine residue 421 is in magenta. On the left this residue has been mutated to leucine, also in magenta, showing increased space in the active site.

I423T

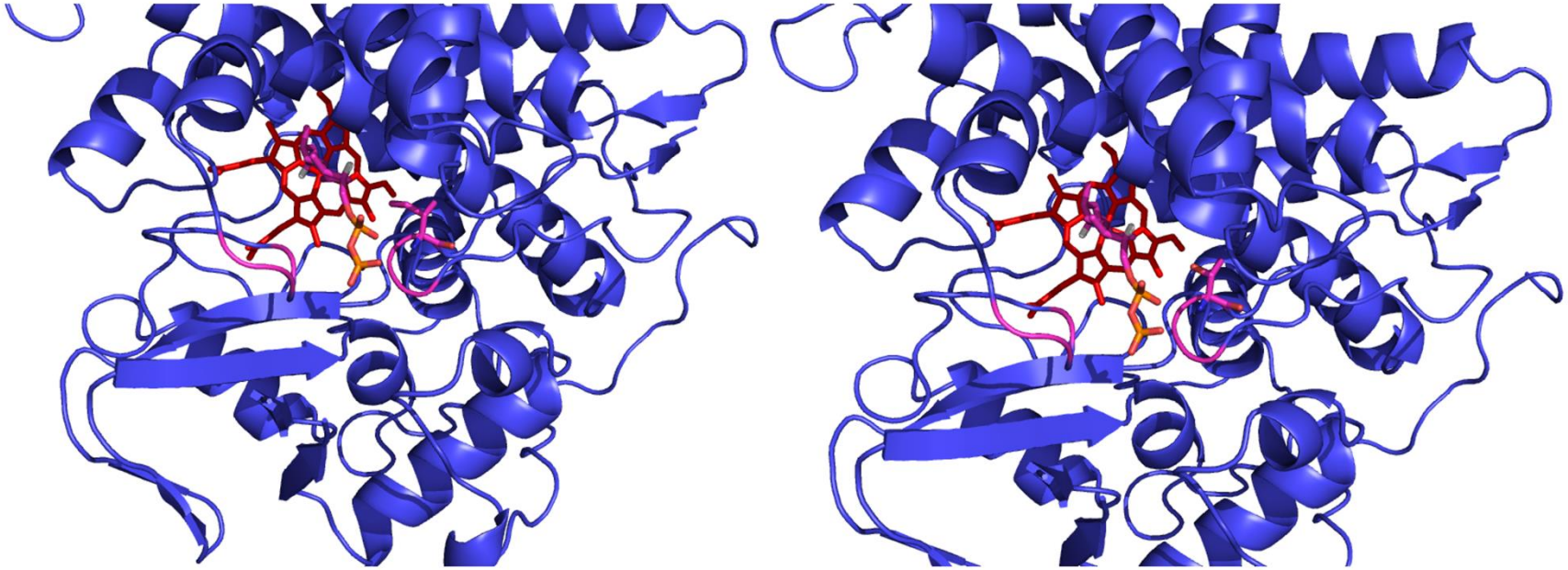


Figure uu.Cartoon representations of a crystal structure of CYP124A1 in blue, with the heme cofactor in red, and a molecule of FDP in magenta modelled into the active site. On the right the isoleucine residue 423 is in magenta. On the left this residue has been mutated to threonine, also in magenta, showing increased space in the active sit

7.3.6 Phosphorylation assays

7.3.6.1 Phosphorus NMR of monophosphorylation of 12-OH Farnesol with UPK

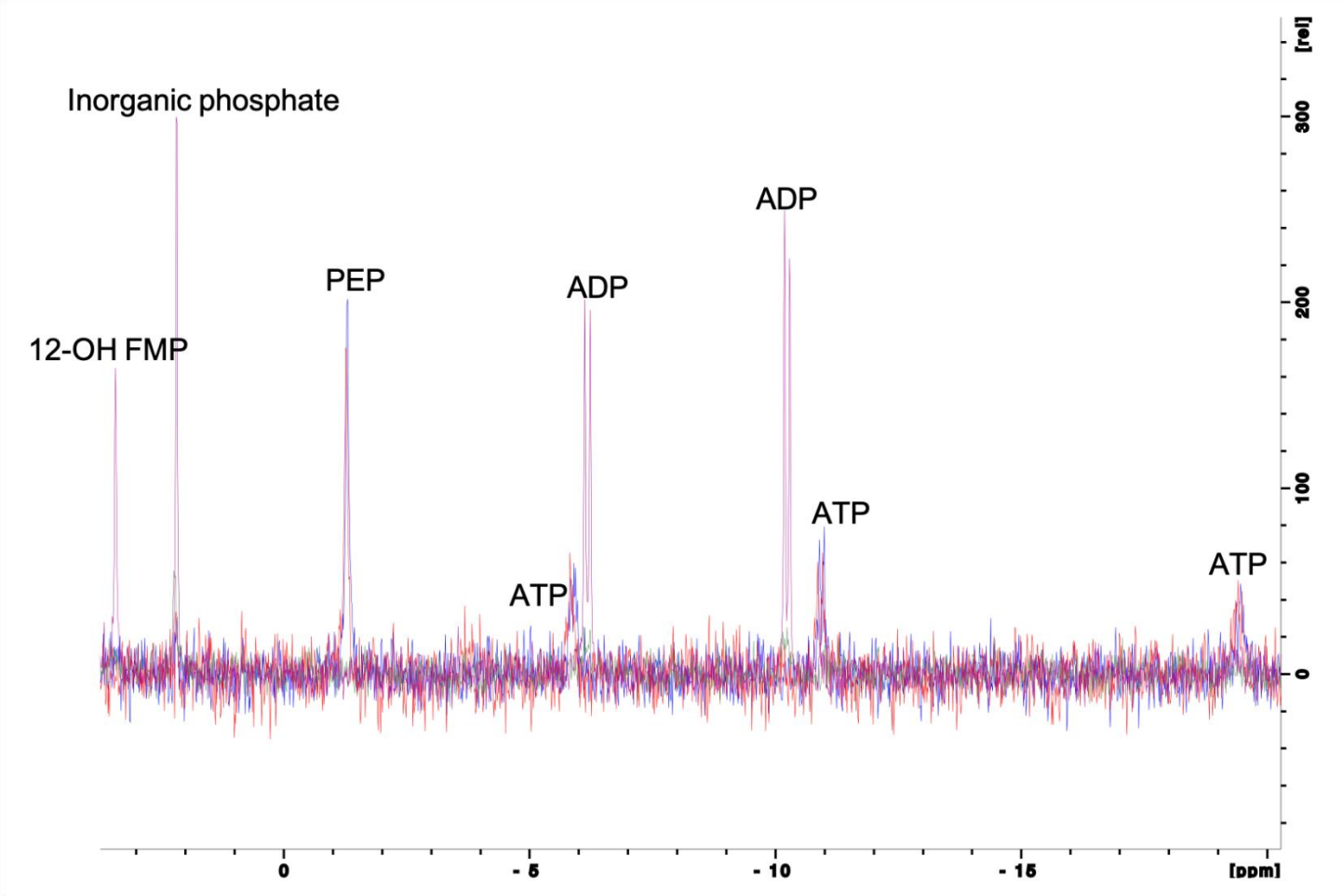
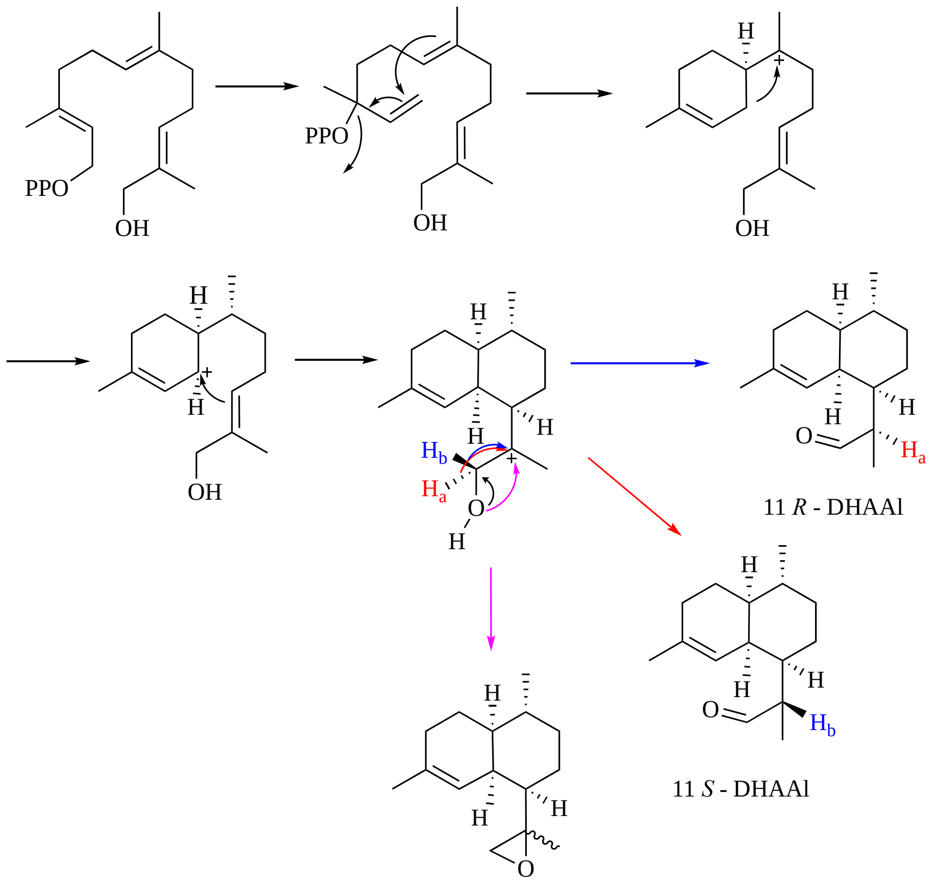


Figure vv. ^{31}P NMR spectrum (243 MHz, D_2O, 298K) of UPK with: blue –no enzyme or substrate, red – substrate and no enzyme, green – enzyme and no substrate, purple - enzyme and substrate

7.3.6.2 Mechanism of conversion of 12-OH FDP to DHAAl



artemisinic-11 *S*, 12- epoxide

Figure ww. Proposed mechanism of formation and products of incubation of 12-OH FDP with ADS (adapted from Dr. F. Huynh, PhD thesis, Cardiff University, 2020).

7.3.6.3 TIC from GC-MS analysis of IPK41-ADS coupled assay

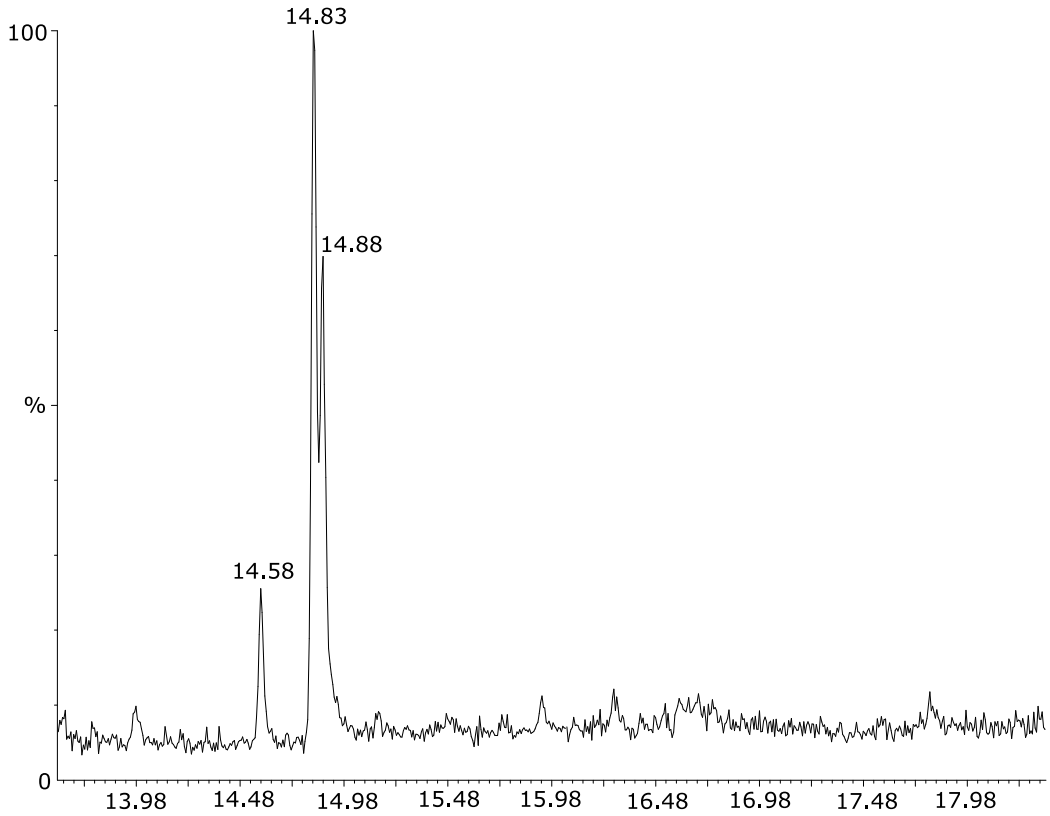


Figure xx.TIC from GC-MS analysis of IPK41-ADS coupled assay treated with alkaline phosphate. IPK41 is in a 1:16 ratio with ADS

7.3.6.4 TIC from GC-MS analysis of IPK41-ADS coupled assay

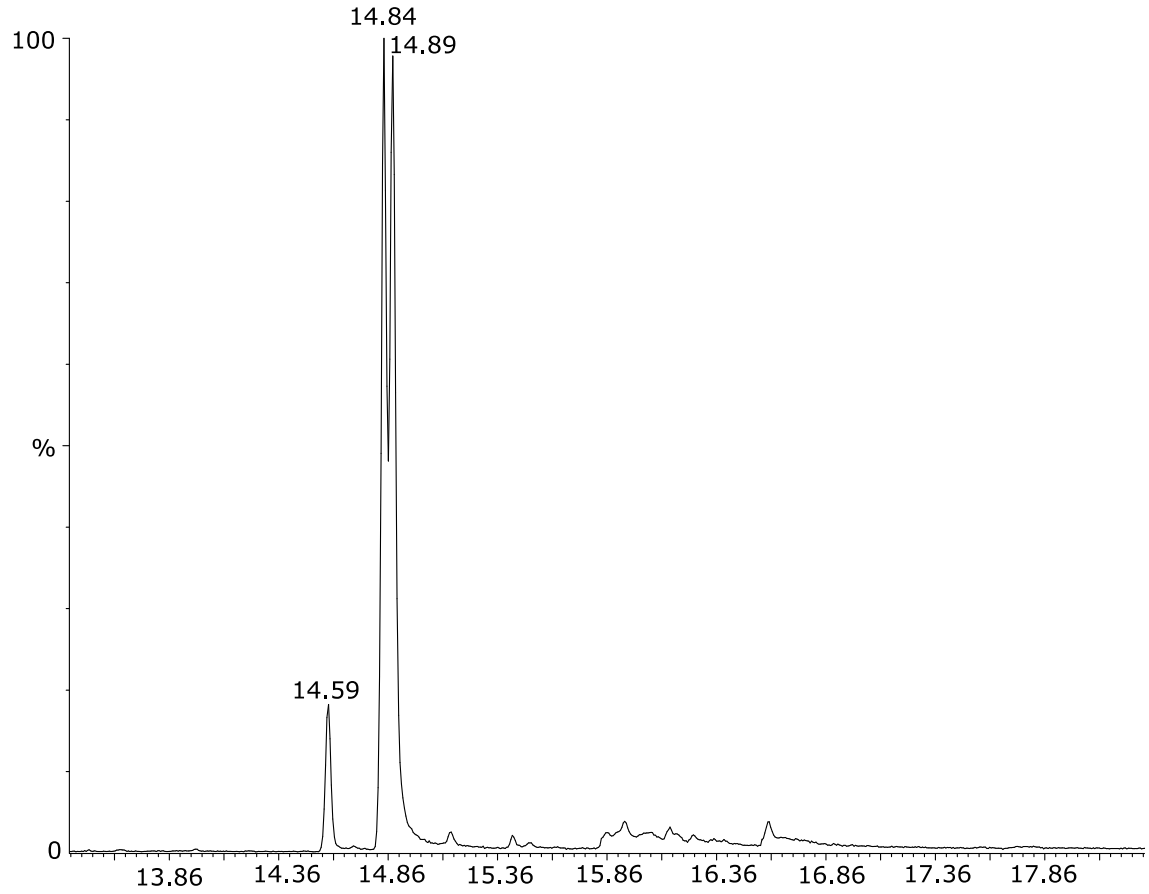


Figure yy. TIC from GC-MS analysis of IPK41-ADS coupled assay treated with alkaline phosphate. IPK41 is in a 12:1 ratio with ADS

7.3.6.5 TIC from GC-MS analysis of IPK41-ADS coupled assay in a ratio of 1:20

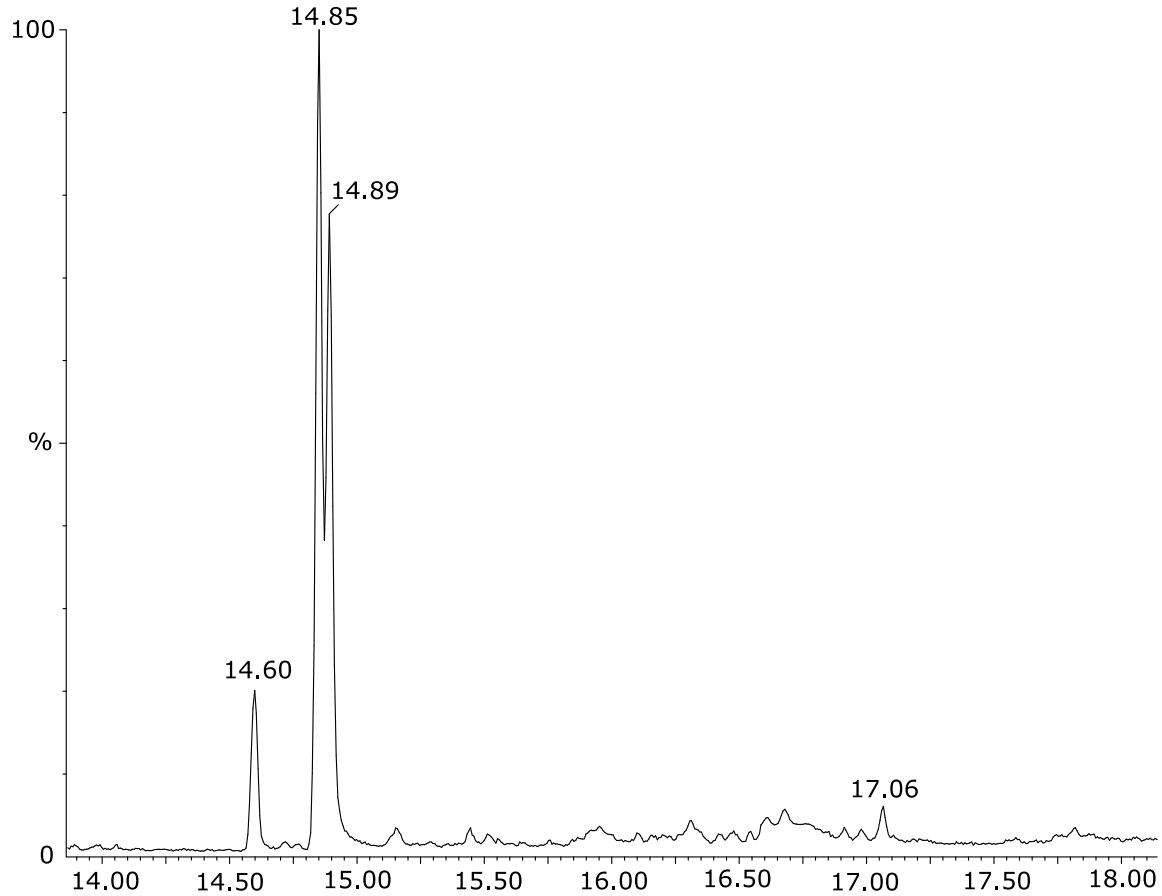


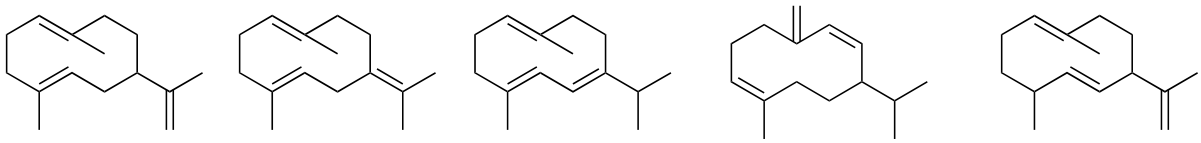
Figure zz. TIC from GC-MS analysis of IPK41-ADS coupled assay treated with alkaline phosphate. IPK41 is in a 1:20 ratio with ADS run in the presence of ammonia, treated with alkaline phosphate

8 Appendix II: Investigation of Germacrene B synthase

8.1 Introduction

Germacrene B synthase is a sesquiterpene synthase found in a variety of plants, such as *Curcuma zedoaria*,[1] *Lycopersicon hirsutum LA1777*,[2] and *Cistus criticus*.[3] The germacrene B synthase used in this work comes from *Curcuma zedoaria*, and is 64.1 kDa.[1] Germacrene B has been found to have a distinctive geranium-like odour that contributes to the scent of lime peel.[4] Many studies on essential oil extractions containing germacrene B have been conducted, and have shown a variety of activities. One such example is the essential oil of *Actinodaphne cupularis* (Hemsl.) Gamble containing 15.2% germacrene B which showed antioxidant behaviours.[5] Germacrene B is one of the most abundant components of the essential oil of leaves from *Carapa guianensis*. This oil was found to have activity against both *S. aureus* and *E. faecalis*.[6] Other activities have be attributed to essential oils containing germacrene B such as the oil extracted from *Ajuga pseudovia*. This oil was found to have antioxidant activity, to be antibacterial against a variety of strains including *S. aureus*, to have antifungal activity against many species including *C. albicans* and exhibit dose dependent ACE (angiotensin I-converting enzyme) inhibition.[7]

Germacrene B is one isomer of a family of sesquiterpenes, the germacrenes which contain germacrene A – E. Germacrene A – E are volatile sesquiterpenes widely found in nature. These compounds share a common cyclodecadiene structure and are part of a wider family of germacrenes.[8]



Germacrene A Germacrene B Germacrene C Germacrene D Germacrene E

Figure ciii Structures of Germacrenes A-E [8]

While this work shows germacrene B is a compound worth investigating, none look at germacrene B as an isolated compound. This may be as germacrene B synthase has yet, to our knowledge, been isolated as a protein and assayed with FDP to obtain a pure product profile. Germacrene B has been expressed in non-native organisms such as *E. coli*, and the cell extract has been assayed.[1–3] This work hoped to purify germacrene B synthase and use it to obtain pure

germacrene B for experimental work, but unfortunately germacrene B synthase proved very difficult to express.

8.2 Results

The gene sequence for germacrene B synthase (GBS) was obtained from Thermofischer with appropriate overhangs to allow cloning into a pET28 plasmid containing a C terminal His tag and preinserted BsaI sites *via* golden gate cloning, see 6.1.16 of materials and methods for further information. Once GBS was cloned, test expression were attempted to try and obtain soluble protein. As this had never been done various cell lines, induction times, and induction temperatures were tested.

8.2.1 Initial screening of induction lengths with BL21 DE3

Initially GBS was transformed into BL21 DE3 cells and it was found that GBS was in the cell pellet when expressed overnight at 20 ° C.

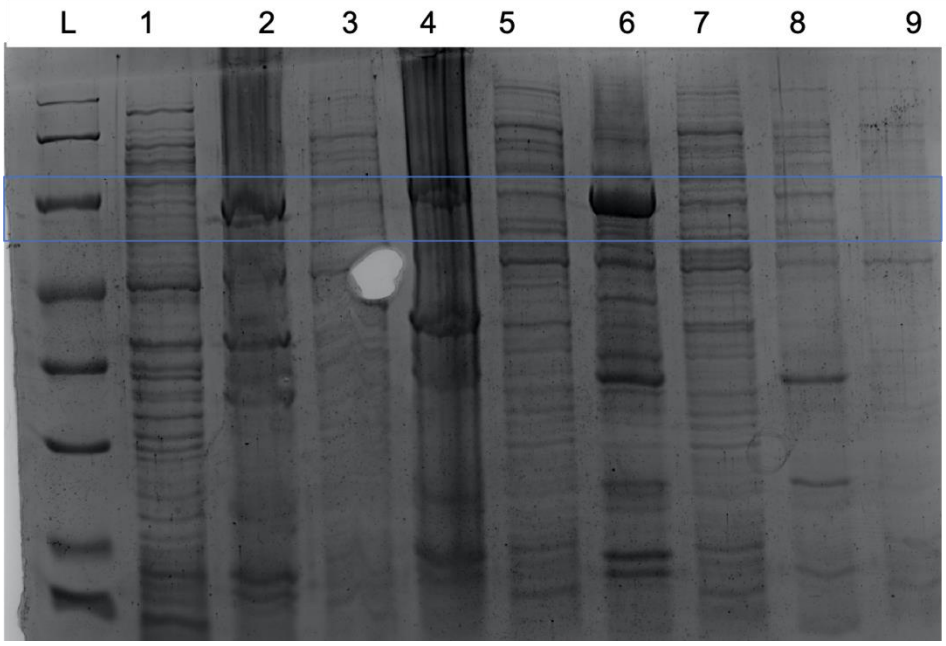


Figure civ Protein gel of GBS 100 mL test expressions. L=ladder, Lane: 1: before induction; 2 - 7 37 ℃ growth OD 0.8, 2: 3h induction 20 ℃ cell pellet, 3: 3h induction 20 ℃ supernatant, 4: 4h induction 20 ℃ cell pellet, 5: 4h induction 20 ℃ supernatant, 6: O/N induction 20 ℃ cell pellet, 7: O/N induction 20 ℃ supernatant; 8 - 9 37 ℃ growth OD 0.2, 20 ℃ growth OD 0.8, 8: 4h induction 20 ℃ cell pellet, 9: 4h induction 20 ℃ supernatant.

As it seemed growth at 37 ℃ until an OD of 0.8, and then overnight induction at 20 ℃ gave overexpression in the cell pellet, these conditions were scaled up to 500 mL. The cells were harvested by centrifugation at (5000 rpm, 20 minutes, 4 °C) and basic extraction attempted on

the cell pellet, as described in materials and methods. Unfortunately, it was seen that GBS stayed in the cell pellet once basic extraction was attempted.

8.2.2 Screening of alternative cell lines

Therefore, other cell lines were screened for overexpression of GBS. The cells lines BL21-AI, C41 (DE3) and C41 (DE3) pLysS were all tested with various growth conditions. 100 mL test expressions were cultured at 37 ℃, one until OD 0.2 which was then cooled to 20 ℃ and then induced at OD 0.8 and left to express for four hours. The other was grown until OD 0.8 before being induced and left to express for four hours.

GBS was transformed into BL21 AI, C41 DE3 and C41 DE3 pLysS cells and were expressed for 6 hours or overnight at 20 ℃. No clear bands were seen for any of the growth conditions, so all were assayed.

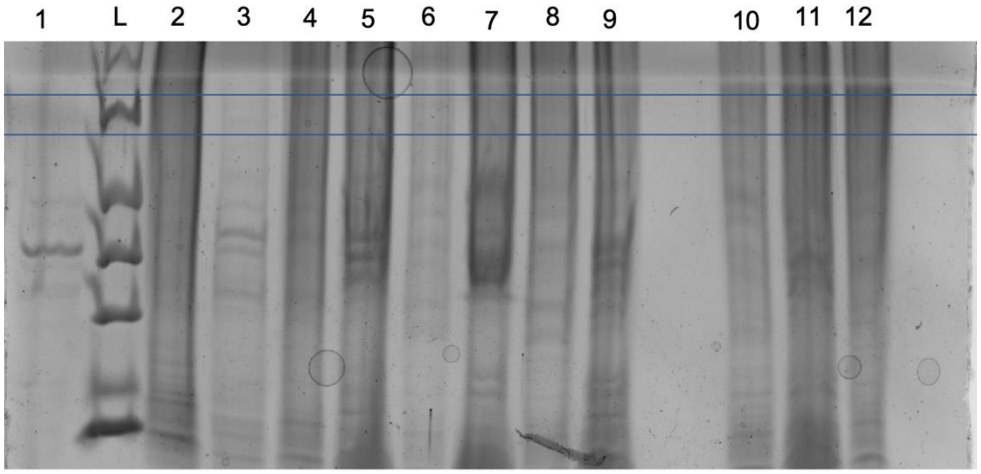


Figure cv Protein gel of GBS 100 mL test expressions induced for 6h. L=ladder, Lane: 1-2: C41 DE3 37 ℃ growth to 0.2 OD, 20 ℃ growth to OD 0.8., 1: cell pellet, 2: supernatant. Lane 3-4 C41 DE3 37 ℃ growth to OD 0.8., 3: cell pellet, 4: supernatant. Lane: 5-6: C41 DE3 pLysS 37 ℃ growth to 0.2 OD, 20 ℃ growth to OD 0.8., 5: cell pellet, 6: supernatant. Lane 7-8: C41 DE3 pLysS 37 ℃ growth to OD 0.8., 7: cell pellet, 8: supernatant. Lane: 9-10: Bl21-AI 37 ℃ growth to 0.2 OD, 20 ℃ growth to OD 0.8., 9: cell pellet, 10: supernatant. Lane 11-12 BL21-AI 37 ℃ growth to OD 0.8., 11: cell pellet, 12: supernatant.

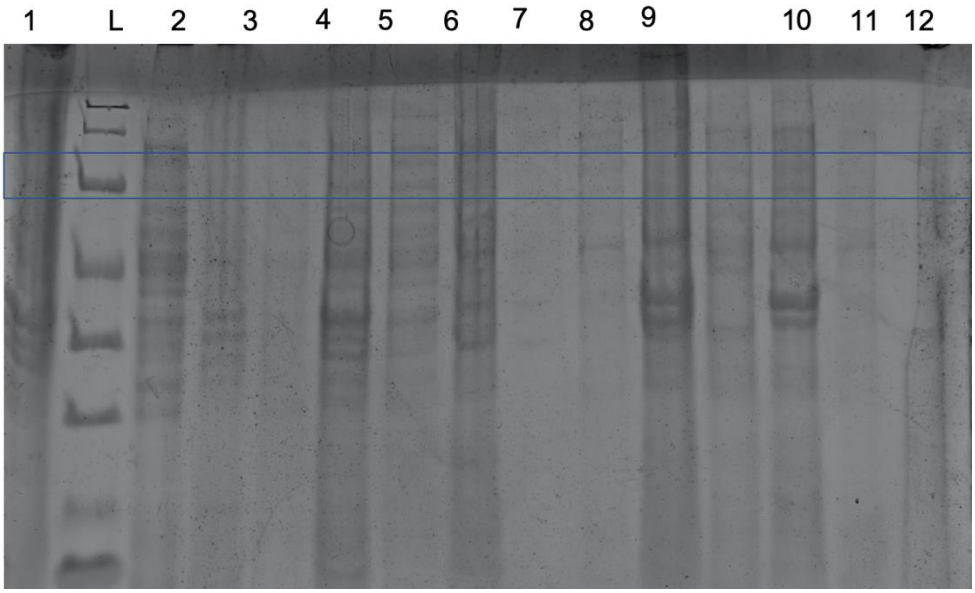


Figure cvi Protein gel of GBS 100 mL test expressions induced overnight. L=ladder, Lane: 1-2: C41 DE3 pLysS 37 ℃ growth to 0.2 OD, 20 ℃ growth to OD 0.8., 1: cell pellet, 2: supernatant. Lane 3-4 C41 DE3 pLysS 37 ℃ growth to OD 0.8., 3: cell pellet, 4: supernatant. Lane: 5-6: C41 DE3 37 ℃ growth to 0.2 OD, 20 ℃ growth to OD 0.8., 5: cell pellet, 6: supernatant. Lane 7-8: C41 DE3 37 ℃ growth to OD 0.8., 7: cell pellet, 8: supernatant. Lane: 9-10: Bl21-AI 37 ℃ growth to 0.2 OD, 20 ℃ growth to OD 0.8., 9: cell pellet, 10: supernatant. Lane 11-12 BL21-AI 37 ℃ growth to OD 0.8., 11: cell pellet, 12: supernatant.

8.2.3 Supernatant activity assays

The supernatant and the cell pellet extract were assayed with FDP (0.4 mM) and buffer (20 mM HEPES, 1 mM DTT and 10 mM $MgCl_2$ pH7.5) with a pentane overlay.

In the assays containing the supernatant from both C41 DE3 cell growths and the AI cells grown until OD 0.8 at 37 ℃ with a 6 hour induction a peak corresponding to a germacrene B authentic standard. A similar peak was seen from the assay containing the supernatant from C41 DE3 cells grown till OD 0.2 at 37 ℃ and until OD 0.8 at 20 ℃ then induced overnight. This showed there was GBS expression, but the level was too low to be seen as an overexpression band on a protein gel.

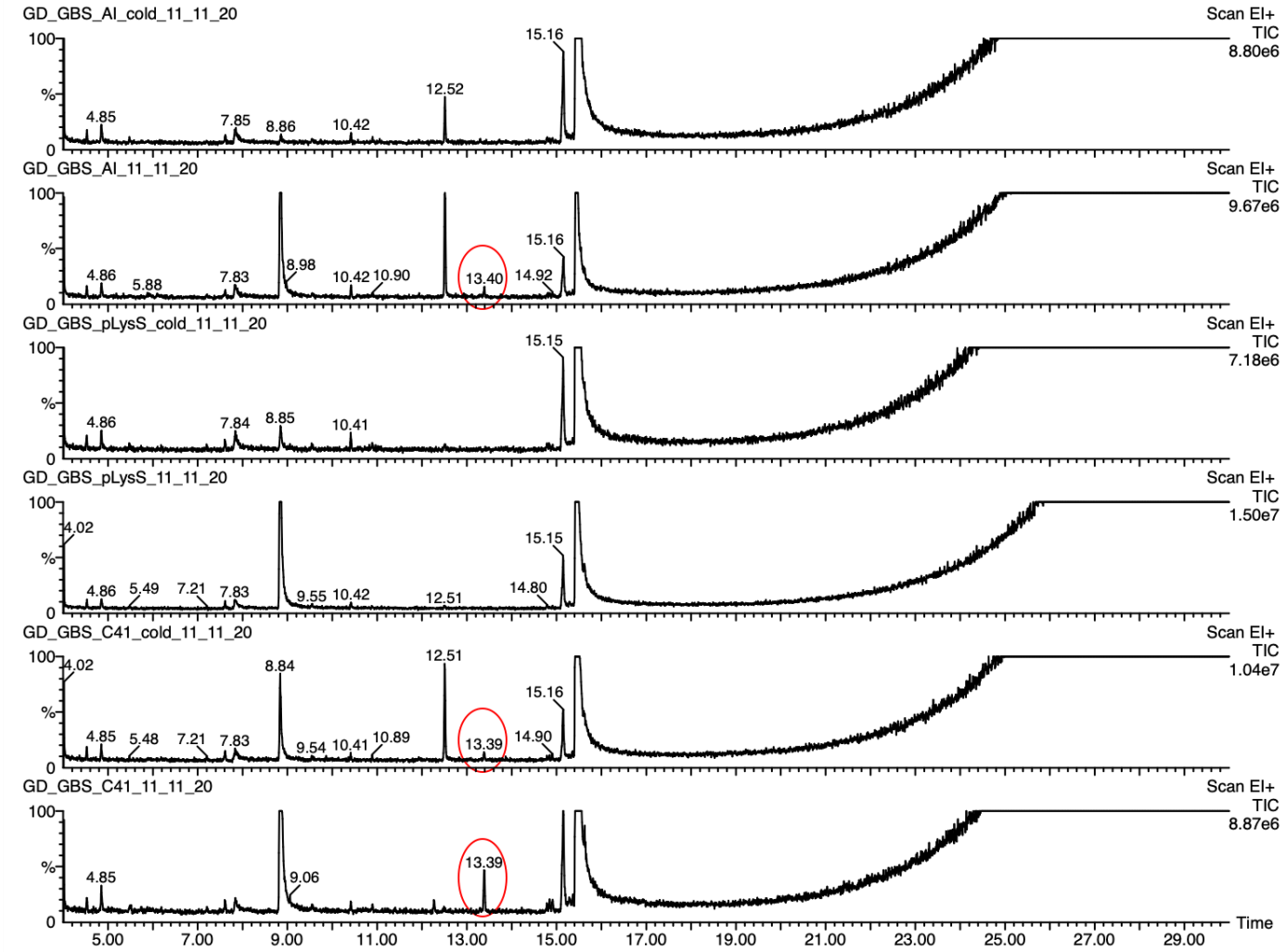


Figure cvii Total ion current chromatogram from GC-MS analysis of pentane extractions of an assay with the supernatant obtained from sonicating various cultures of cells transformed with GBS and induced for 6 hours, FDP (0.4 mM), in buffer (see materials and methods 6.1.26.1 for further details), with the product germacrene B eluting at 13.39-13.40.

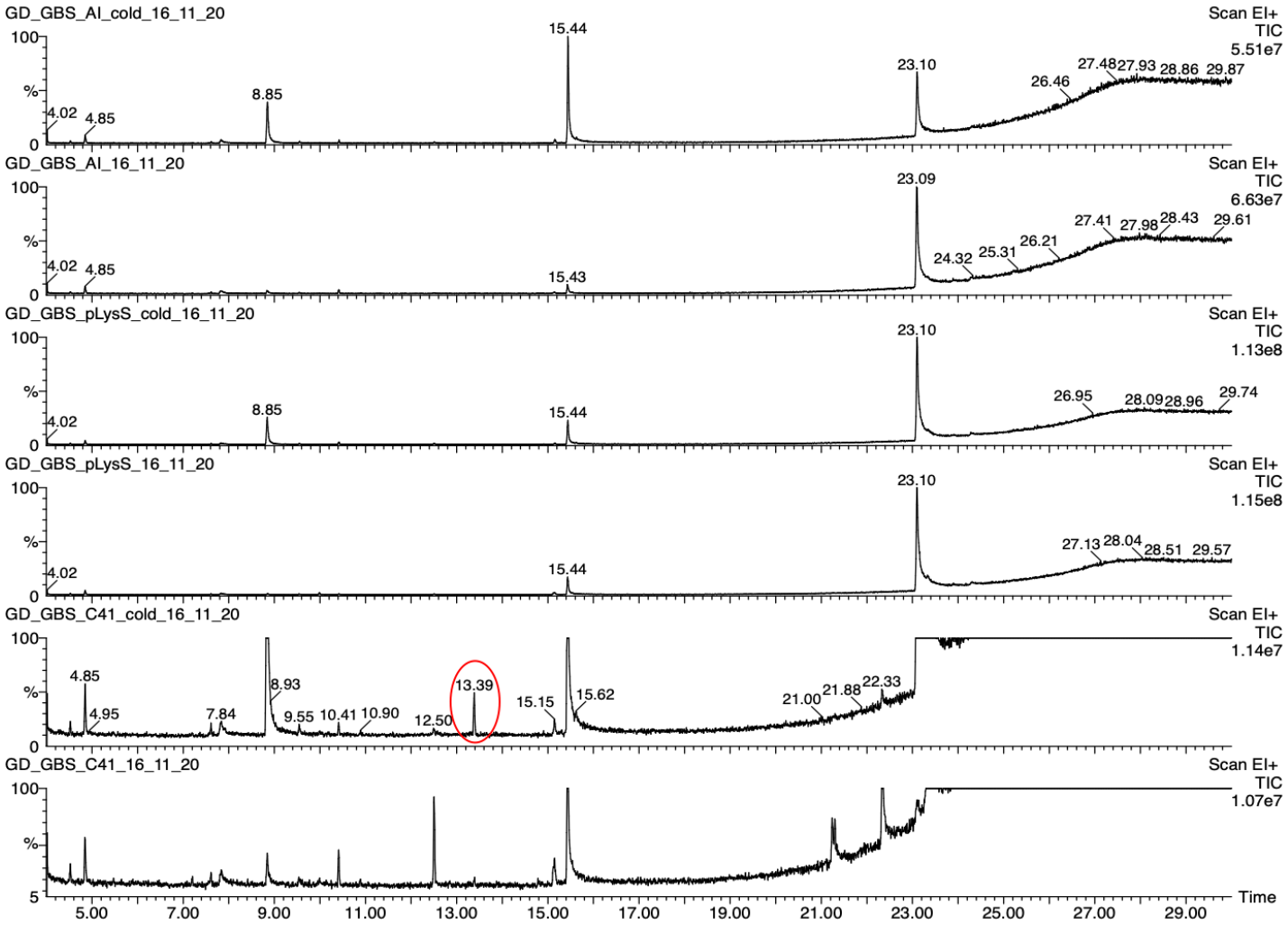


Figure cviii Total ion current chromatogram from GC-MS analysis of pentane extractions of an assay with the supernatant obtained from sonicating various cultures of cells transformed with GBS and induced overnight, FDP (0.4 mM), in buffer (see materials and methods 6.1.26.1 for further details), with the product germacrene B eluting at 13.39-13.40

8.2.4 pET-32 Xa LIC GBS

As various cell lines, induction times and temperatures had not succeeded in obtaining soluble active GBS proteins, the gene was transferred to another plasmid. The GBS gene was transferred to a pET-32 Xa LIC plasmid, which contains a thioredoxin tag for high level expression. This was done *via* golden gate cloning. Once the plasmid was confirmed by sequencing it was transformed into BL21 DE3 and grown at 37 °C until an OD of 0.8, and then the induction conditions described above were used, 8.2.1. Unfortunately, no expression was seen with these system, not even insoluble protein as with the other plasmid. Therefore, the screening system with various cell lines, as described above 8.2.2, was tried. No expression was seen with any of the cell lines tested. The supernatant and the cell pellet extract were assayed with FDP (0.4 mM) and buffer (20 mM HEPES, 1 mM DTT and 10 mM $MgCl_2$ pH7.5) with a pentane overlay. No activity was seen in any of the assays completed.

8.3 Future Work

Germacrene B is clearly a valuable compound and has many reported properties, therefore attempting to obtain it in high yields with pure germacrene B synthase would be very useful. In this work germacrene B synthase proved very difficult to express in a soluble active form in *E. coli*. In the future various solubility tags could be tried with this enzyme *e.g.*, GB1, SUMO; as well as screening other cell lines not tested in this work. In monoterpenes it has been shown that removing particular regions at the N terminus produced more soluble protein.[9] This could be an option to investigate for germacrene B synthase.

1 J. I. Hattan, K. Shindo, T. Sasaki and N. Misawa, *Journal of Oleo Science*, 2018, **67**, 1235–1246.
2 R. S. van der Hoeven, A. J. Monforte, D. Breeden, S. D. Tanksley and J. C. Steffens, *The Plant Cell*, 2000, **12**, 2283–2294.
3 V. Falara, V. Fotopoulos, T. Margaritis, T. Anastasaki, I. Pateraki, A. M. Bosabalidis, D. Kafetzopoulos, C. Demetzos, E. Pichersky and A. K. Kanellis, *Plant Molecular Biology*, 2008, **68**, 633–651.
4 B. C. Clark, T. S. Chamblee and G. A. Iacobucci, *J. Agric. Food Chem*, 1987, **35**, 514–518.
5 B. Yu, D. Zhang, X. W. Yan, J. W. Wang, L. Yao, L. H. Tan, S. P. Zhao, N. Li and W. G. Cao, *Chemistry and Biodiversity*, 2016, **13**, 1573–1583.
6 G. Meccia, P. Quintero, L. B. Rojas, A. Usubillaga, J. Velasco, T. Diaz, C. Diaz, J. Velásquez and M. Toro, *Natural Product Communications*, 2013, **8**, 1641–1642.
7 M. Ben Mansour, R. Balti, L. Rabaoui, A. Bougatef and M. Guerfel, *Process Biochemistry*, 2013, **48**, 723–729.
8 A. M. Adio, *Tetrahedron*, 2009, 65, 1533–1552.
9 M. Back, K. Chappell and J. Noel, *Science (1979)*, 1997, **277**, 1815–1820.