

Chemoenzymatic applications and catalytic mechanism of *7-epi*-zingiberene synthase

Jennifer Benton

A thesis submitted to Cardiff University for the
degree of Doctor of Philosophy

January 2022



Abstract

Terpenes are a vast family of natural products, all derived from simple C₅ diphosphate precursors which undergo coupling reactions by prenyltransferases and subsequent cyclisations by terpene synthases. Terpene synthases carry out chemically complex reaction cascades, yielding regio- and stereochemically diverse structures which confer a myriad of physiological functions. Many terpenes have important economic value and applications across the chemical industries as flavours and fragrances, pharmaceuticals and agrochemicals. The abundant volatile C₁₅ sesquiterpene semiochemical of wild tomato, *7-epi*-zingiberene, has been shown to repel whitefly, a significant global crop pest. Its precursor is the unusual (2*Z*,6*Z*)-farnesyl diphosphate ((*Z,Z*)-FDP); the *cis* geometric isomer of the ubiquitous (*E,E*)-FDP. Current syntheses of both compounds are unsuitable for commercial application; hence this work focuses on alternatives based on their biosynthetic enzymes.

A selective three-enzyme chemoenzymatic synthesis of (*Z,Z*)-FDP was established to replace the existing low yielding 11-step chemical synthesis using C₅ alcohol precursors, two kinases and zFPS, the *cis*-prenyltransferase which naturally biosynthesises (*Z,Z*)-FDP. Yields of 92% were achieved, with over 400 mg of pure compound produced per litre of kinase culture. Following expression optimisation, this pathway was combined with *7-epi*-zingiberene synthase (EZS) to selectively make pure sesquiterpene which was isolated and characterised by NMR spectroscopy. This pathway was later adapted for rapid *in vitro* generation of a variety of unnatural diphosphate analogues by expanding the library of alcohol precursors available to the kinases. These substrates were screened with EZS and other semiochemical-producing synthases to produce novel analogues with potential insect bioactivities. Behavioural bioassays furthermore established for the first time that *7-epi*-zingiberene is a repellent of different aphid species, also economically significant crop pests.

EZS structure-function was also investigated. The catalytic mechanism was shown to take place via a [1,3]-hydride shift from the C1 using GC-MS analysis of EZS incubations with deuterated substrate isotopologues, synthesised via a chemoenzymatic approach. Site-directed mutagenesis was used to identify key regions of the synthase responsible for maintaining product fidelity and hydroxylation, with F598 initially identified as an important residue. A G-helix double mutant showed a switch in substrate selectivity, preferentially converting (*E,E*)-FDP into different sesquiterpenes. In general, this work can be used to advance biocatalytic applications of terpene synthases for production of *7-epi*-zingiberene and other high value terpenes.

Acknowledgements

I'd firstly like to thank my supervisor, Prof. Rudolf Allemann, for providing the opportunity to undertake a PhD within his research group and access the research facilities of Cardiff's School of Chemistry. I'd also like to thank the BBSRC's SWBio Doctoral Training Partnership for providing the funding for this work and the team for facilitating a great training and development programme. Special thanks go to Prof. John Pickett for his help and encouragement throughout and sharing his enthusiasm for chemical biology.

I've been fortunate to work with brilliant postdocs: Alan, I appreciate the support you've shown us from the very start, with thanks also extended to Anto, Adura and Zulfa. Particular thanks go to Luke for the opportunity to collaborate and publish, Prabhakar for sharing his extensive terpene synthase expertise, and both for the insightful scientific discussions, project support and thesis proof-reading.

Thank you to the past and present members of the research group who endeavoured to make our lab an enjoyable place to work. To the single cell - Raquel, Florence, Gwawr, Alice - you've been invaluable friends, always around for great chats and importantly keeping everything in perspective. A PhD highlight is our amazing trip to Madrid! Further thanks to Chris for kindly sharing his expertise and precious compounds to continue working with EZS, but also to Eddie, Martin, Tak and Gareth for all the support along the way including laughs at lunch, excellent bake-off contributions and great nights out.

I'm so grateful to have undergone the PhD journey with Gwawr: I will always remember your vivacity and dedication to helping just about anyone, thank you for everything and never change! A huge thanks also to my PhD partner, Tom: I couldn't have wished for a better friend and housemate over the years. Proud of us and excited for what's next!

Finally, I would like to thank my wonderful parents, friends and family for their tireless encouragement of me, and for taking the time to understand what I was doing - this is all for you. All my love especially to Chris, for putting up with the chaos, always building me up and making me feel it's all worthwhile!

Table of Contents

Abstract	i
Acknowledgements	iii
Table of Contents	v
List of Abbreviations	xi
1 Introduction	1
1.1 Terpene natural products and precursor biosynthesis	2
1.1.1 Mevalonate and non-mevalonate biosynthetic pathways	3
1.1.2 Prenyltransferases	6
1.2 Terpene synthases	9
1.2.1 Cyclisation chemistry	9
1.2.2 Structure and function of terpene synthases	11
1.3 Techniques for investigating the mechanisms of terpene synthases	13
1.3.1 Crystallography	14
1.3.2 Mutagenesis	15
1.3.3 Fluorinated substrate analogues	16
1.3.4 Aza substrate analogues	17
1.3.5 Isotopologues	18
1.4 Production of terpenes	19
1.4.1 Semisynthetic and in vivo production of terpenoids	19
1.4.2 Unnatural terpenes from substrate analogues	21
1.5 Chemical ecology applications of terpene semiochemicals	24
1.5.1 Food security	24
1.5.2 Semiochemicals	24
1.5.3 Whitefly pests and wild tomato	25
1.5.4 Atypical cisoid terpene biosynthesis in tomatoes	27
1.6 Project aims	30
1.6.1 Context	30
1.6.2 Aims	32

2	Chemoenzymatic Synthesis of (Z,Z)-FDP and 7-Epi-zingiberene	34
2.1	Introduction	35
2.1.1	Aims of this work	37
2.2	Chemoenzymatic (Z,Z)-FDP synthesis via synthetic IDP and DMADP	38
2.3	Chemoenzymatic synthesis of DMADP and IDP	39
2.3.1	Expression and purification of IPK and THIM	40
2.3.2	Enzymatic diphosphorylation of prenil and isoprenol	41
2.4	Chemoenzymatic (Z,Z)-FDP synthesis via enzyme prepared IDP and DMADP	44
2.4.1	Expression and purification of zFPS	44
2.4.2	Enzymatic production of (Z,Z)-FDP and characterisation	45
2.5	EZS expression optimisation	47
2.5.1	Improving EZS expression using fusion proteins	47
2.5.2	Expression of EZS using C41(DE3) cells	50
2.5.3	EZS activity testing and compound characterisation	53
2.6	Chemoenzymatic 7-epi-zingiberene synthesis	54
2.6.1	Chemoenzymatic production of 7-epi-zingiberene and characterisation	55
2.7	Summary	57
3	Investigation of the Hydride Shifts of the Catalytic Mechanism of EZS	60
3.1	Introduction	61
3.1.1	Aims of this work	61
3.1.2	Proposed reaction mechanism of EZS	61
3.1.3	Understanding the ring closure step using fluorinated substrate analogues	62
3.1.4	Understanding the hydride shifts using deuterated substrate analogues	64
3.2	Preparation of [² H ₁ -6]-(Z,Z)-FDP and incubation with EZS	64
3.2.1	[² H ₁ -6]-(Z,Z)-FDP as a mechanistic probe	64
3.2.2	Synthesis of [² H ₁ -2]-neryl diphosphate and IDP	66
3.2.3	Incubation of [² H ₁ -6]-NDP with IDP, zFPS and EZS	68
3.3	Preparation of [² H ₂ -5]-(Z,Z)-FDP and incubation with EZS	72
3.3.1	[² H ₂ -5]-(Z,Z)-FDP as a mechanistic probe	72
3.3.2	Synthesis of [² H ₂ -1]-neryl diphosphate	73
3.3.3	Incubation of [² H ₂ -1]-NDP with IDP, zFPS and EZS	75

3.4	Preparation of [² H ₂ -1]-(Z,Z)-FDP	79
3.5	Summary	80
4	Structural Determinants of EZS Substrate Selectivity and Product Specificity	82
4.1	Introduction	83
4.1.1	Aims	83
4.2	Selection of residues	84
4.2.1	G-helix residues	86
4.2.2	Other residues of interest	87
4.3	Cloning and expression of EZS mutants	88
4.4	Product specificity	89
4.5	Substrate specificity	92
4.6	Kinetic characterisation	97
4.7	Summary	99
4.8	Supplementary	100
5	Chemoenzymatic Generation of Novel Semiochemical Analogues	107
5.1	Introduction	108
5.1.1	Aims	108
5.2	Alcohol analogue phosphorylation	109
5.2.1	Small scale analogue phosphorylation	109
5.3	FDP analogues	112
5.3.1	Small-scale screening with prenyltransferases	112
5.3.2	(Z,Z)-FDP analogues	112
5.3.3	(E,E)-FDP analogues	115
5.3.4	Chemistries of the <i>cis</i> - and <i>trans</i> -prenyltransferases	118
5.4	Sesquiterpene analogues	120
5.4.1	7- <i>Epi</i> -zingiberene analogues	120
5.4.2	(<i>E</i>)-β-farnesene analogues	122
5.4.3	Germacrene D analogues	123
5.5	Scaled chemoenzymatic synthesis of 7- <i>epi</i> -zingiberene analogues	127
5.5.1	Scaled reaction for analogue production	127
5.5.2	GC-FID yields	128

5.5.3	Chemical synthesis of 3-methyl-penten-1-ol	128
5.6	Summary	130
5.7	Supplementary	131
6	Ecological Applications of Sesquiterpenes for Aphid Pest Management	137
6.1	Introduction	138
6.1.1	Aims	139
6.2	Characterisation and behavioural testing of (<i>E</i>)- β -farnesene	140
6.2.1	Analytical characterisation	140
6.2.2	Behavioural bioassays	140
6.3	Behavioural testing of 7- <i>epi</i> -zingiberene and curcumenes	143
6.4	Release rate and stability testing of germacrene D	145
6.4.1	Release rate	145
6.4.2	Stability	148
6.5	Summary	153
6.6	Supplementary	154
7	Conclusions and Future Work	156
7.1	Chemoenzymatic synthesis of (<i>Z,Z</i>)-FDP and 7- <i>epi</i> -zingiberene	157
7.2	Chemoenzymatic generation of novel semiochemical analogues	158
7.3	Investigation of the hydride shifts of the catalytic mechanism of EZS	159
7.4	Structural determinants of EZS product fidelity and substrate selectivity	160
7.5	Ecological applications of sesquiterpenes for aphid pest management	161
	Materials and Methods	163
8.1	Biological methods	164
8.1.1	Materials	164
8.1.2	Strains and plasmids	164
8.1.3	Growth media, buffers and antibiotic stock solutions	166
8.1.4	<i>E. coli</i> competent cell preparation and transformation	167
8.1.5	Plasmid preparation	168
8.1.6	PCR	169
8.1.7	Golden Gate cloning	170
8.1.8	Site-directed mutagenesis	171

8.1.9	Agarose gel electrophoresis	173
8.1.10	Protein expression and purification	173
8.1.11	SDS-PAGE	177
8.1.12	Western blotting	178
8.1.13	Protein concentration	179
8.1.14	FPLC	180
8.1.15	Steady-state kinetics	181
8.2	Biosynthetic compound preparation	184
8.2.1	Small-scale alcohol diphosphorylation	184
8.2.2	Preparation of diphosphate stocks	184
8.2.3	Screening of diphosphate analogues with zFPS	185
8.2.4	Screening of diphosphate analogues with EZS	185
8.2.5	(<i>Z,Z</i>)-FDP synthesis and purification	185
8.2.6	<i>7-Epi</i> -zingiberene synthesis and isolation	186
8.2.7	GC-MS	186
8.2.8	GC-FID	186
8.2.9	NMR	187
8.3	Synthetic chemistry methods	187
8.3.1	General experimental	187
8.3.2	IDP	188
8.3.3	DMADP	188
8.3.4	NDP	189
8.3.5	[² H ₂ -1]-NDP	189
8.3.6	[² H ₁ -2]-NDP	193
8.3.7	3-Methyl-penten-1-ol	195
8.4	Chemical ecology methods	196
8.4.1	Materials	196
8.4.2	Behavioural bioassays	197
8.4.3	Germacrene D release rate and stability	198
8.4.4	Analytical techniques	199
	References	201

List of Abbreviations

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

A_{280} - absorbance at 280 nm

ADP - adenosine diphosphate

ADS - amorpho-4,11-diene synthase

ANOVA - analysis of variance

AP - alkaline phosphatase

AT-AS - *Aspergillus terreus* aristolochene synthase

ATP - adenosine triphosphate

B. tabaci - *Bemisia tabaci*

β -CD - β -cyclodextrin

CV - column volume

DDXXD - terpene synthase metal-binding motif

DHAAI - dihydroartemisinic aldehyde

DHFR - dihydrofolate reductase

DIBAL-H - diisobutylaluminium hydride

DMAP - dimethylallyl monophosphate

DMADP - dimethylallyl diphosphate

DMSO - dimethylsulfoxide

DNA - deoxyribonucleic acid

DTT - dithiothreitol

EBF - (*E*)- β -farnesene

EBFS - (*E*)- β -farnesene synthase

E. coli - *Escherichia coli*

EDTA - ethylenediaminetetraacetic acid

(*E,E*)-FDP - (*2E,6E*)-farnesyl diphosphate synthase

EI - electron impact

EPICS - *epi*-cubenol synthase

EZS - *7-epi*-zingiberene synthase

FPLC - fast protein liquid chromatography

FPPS - farnesyl diphosphate synthase

GB1 - guanine binding protein

GC-FID - gas chromatography with flame ionisation detection

GC-MS - gas chromatography coupled mass spectrometry

GDP - geranyl diphosphate

GDS - (S)-germacrene D synthase

GGDP - (2E,6E,10E)-geranylgeranyl diphosphate

GST - glutathione-S-transferase

HEPES - 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid

h - hour/s

His - polyhistidine

HPLC- high performance liquid chromatography

HWE - Horner-Wadsworth-Emmons

IDI - isopentenyl diphosphate isomerase

IDP - isopentenyl diphosphate

IP - isopentenyl monophosphate

IPK - isopentenyl phosphate kinase

IPTG - Isopropyl β -d-1-thiogalactopyranoside

k_{cat} - turnover number

KI - Kovats indices

KLD - kinases, ligase, Dpnl

K_M - Michaelis constant

LB - lysogeny broth

LSD - least significant difference

MBP - maltose binding protein

MEP - methylerythritol phosphate pathway

MEV - mevalonate pathway

min - minute

mRNA - messenger ribonucleic acid

M. persicae - *Myzus persicae*

n-BuLi - *n*-butyllithium

NDP - neryl diphosphate

Ni-NTA - nickel coupled nitrilotriacetic acid agarose

NIST - National Institute of Standards and Technology

NMR - nuclear magnetic resonance

NNDP - nerylneryl diphosphate

NSE - terpene synthase metal-binding motif

OD₆₀₀ - optical density at 600 nm

PCR - polymerase chain reaction

PEP - phosphoenolpyruvate

pI - isoelectric point

PK - pyruvate kinase

PMSF - phenylmethylsulfonyl fluoride

ppm - parts per million

PR-AS - *Penicillium roqueforti* aristolochene synthase

PTAD - 4-phenyl-1,2,4-triazoline-3,5-dione

PVC - polyvinyl chloride

Q-seph - quaternary ammonium functionalised sepharose

S. avenae - *Sitobion avenae*

SBS - santalene and bergamotene synthase

S. cerevisiae - *Saccharomyces cerevisiae*

SDM - site-directed mutagenesis

SDS-PAGE - sodium dodecyl sulfate
polyacrylamide gel electrophoresis

SE - standard error

S. lycopersicum - *Solanum lycopersicum*

S. habrochaites - *Solanum habrochaites*

SUMO - small ubiquitin-like modifier

TAE - TRIS-acetate EDTA

TB - terrific broth

TBST - TRIS-buffered saline with Tween 20

TCEP - tris(2-carboxyethyl)phosphine

TEAP - triethylammonium phosphate

TEV - tobacco etch virus

THIM - hydroxyethylthiazole kinase

TLC - thin layer chromatography

TPS - terpene synthase (clade)

TRIS - tris(hydroxymethyl)aminomethane

TRX - thioredoxin

UV - ultraviolet

V_{\max} - maximum (enzymatic) reaction rate

zFPS - (2Z,6Z)-farnesyl diphosphate
synthase

(Z,Z)-FDP - (2Z,6Z)-farnesyl diphosphate

12-OH-FDP - 12-hydroxy-(2E,6E)-farnesyl
diphosphate

Introduction

1.1 Terpene natural products and precursor biosynthesis

Terpenes, also referred to as isoprenoids or terpenoids (modified terpenes), are the largest class of natural products with over 80,000 unique compounds.¹ Their simple five carbon precursors can be linked together and cyclised in a myriad of ways, leading to incredibly stere- and regiochemically diverse structures which serve a multitude of physiological functions across all kingdoms of life.² In primary metabolism, terpene derived compounds are essential in the mediation of cellular growth and development.³ Lipids such as cholesterol are involved in cell membrane biosynthesis and are precursors to steroids used in hormonal signalling,⁴ whilst carotenoids such as lycopene absorb UV light for photosynthesis and prevent photodamage in plants (Figure 1.1).⁵ As secondary metabolites, terpenoids are significant mediators of plant, insect and microorganism chemical ecology.^{6,7} Approximately half the known terpenes originate in plants with their abundance attributed to evolutionary adaptation to their immobility within specific ecological niches, particularly for defence against competitors.⁸

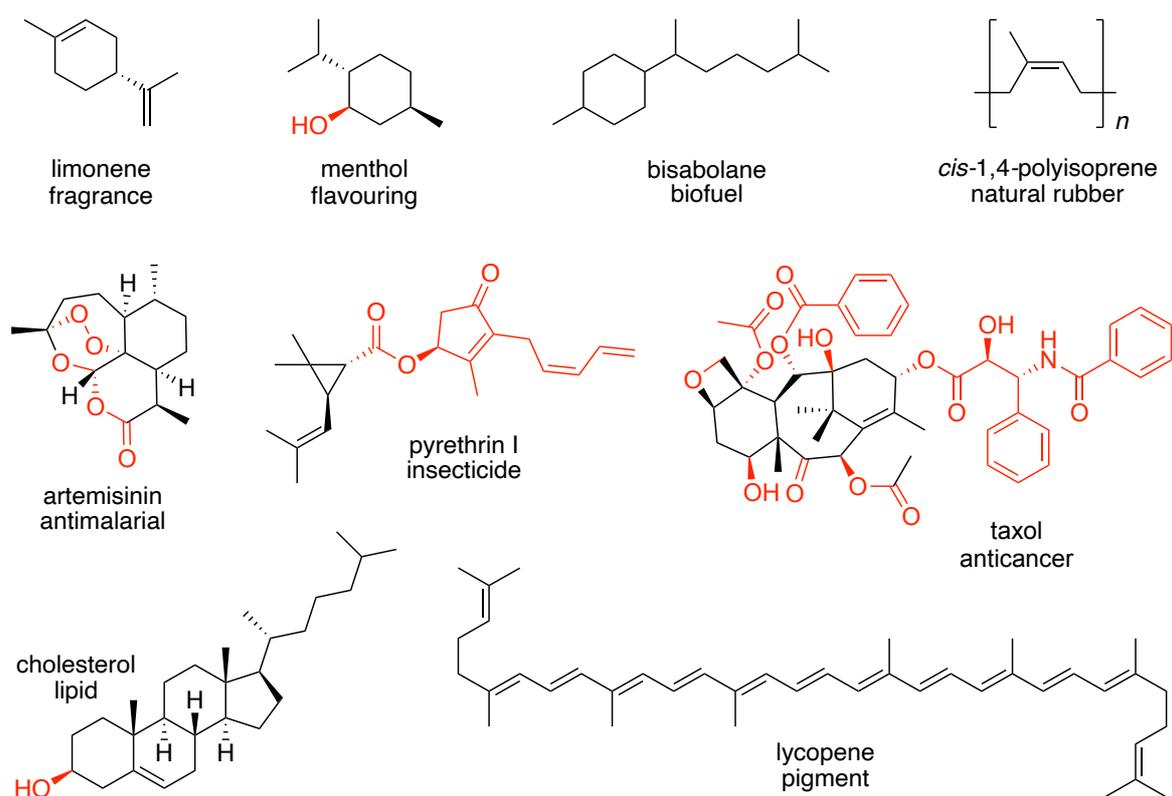


Figure 1.1: Structures of notable terpene natural products. Terpene skeletons derived from fundamental C₅ precursors given in black with red showing later functionalisation.

The useful bioactivities and properties of terpenoids, conferred by their complex molecular scaffolds, have been utilised across the chemical industries. Predominant constituents of plant essential oils, they are widely used as fragrances and flavourings: menthol and limonene for example are produced on a kilotonne scale annually.⁹ Terpene hydrocarbons such as bisabolane are being developed as renewable biofuels with energy densities and other properties (fluidity, hygroscopy) comparable to diesel and petrol.¹⁰ The valuable biopolymer, natural rubber, is biosynthesised from simple terpene monomers, assembled into *cis*-1,4-polyisoprenes of vast chain lengths of C_{100,000} or more.¹¹ Terpenoid fine chemicals of significant economic importance furthermore include the pharmaceuticals artemisinin and taxol, both derived from semisynthetic precursors,^{12,13} as well as the pyrethrin class of insecticides.¹⁴

1.1.1 Mevalonate and non-mevalonate biosynthetic pathways

Despite their incredible structural and functional diversity, all terpenes are fundamentally comprised of multiple five-carbon isoprene (C₅H₈) units (Figure 1.2) arranged in a regular or irregular fashion according to the isoprene rule.¹⁵ The biological precursors to these isoprene units are isopentenyl diphosphate (IDP) and dimethylallyl diphosphate (DMADP) which originate from two distinct biochemical pathways: the mevalonate (MEV) pathway, dependent on acetyl-CoA and the methylerythritol phosphate (MEP) or non-mevalonate pathway, dependent on glyceraldehyde-3-phosphate. In general, the MEP is present in prokaryotes whilst the MEV pathway is present in eukaryotes, although it has been identified in archaea and some bacteria.¹⁶ Plants utilise both: the MEV pathway operates in the cytosol and provides the precursors to sesqui- (C₁₅) and triterpenes (C₃₀), whilst the MEP is localised to plastids and provides the precursors to mono- (C₁₀), di- (C₂₀) and tetraterpenes (C₄₀).¹⁷

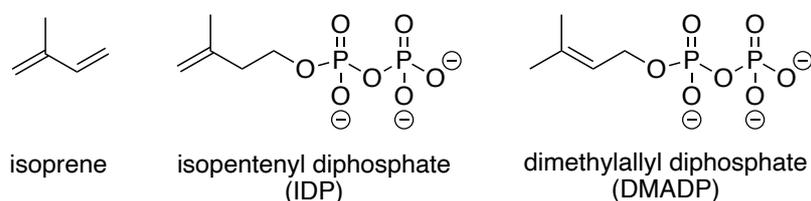


Figure 1.2: Isoprene and the fundamental terpene biosynthetic precursors, DMADP and IDP.

The MEV pathway (Figure 1.3) begins with the Claisen condensation of two acetyl-CoA molecules to form acetylacetyl-CoA. HMG-CoA synthase catalyses an aldol reaction between acetylacetyl-CoA and a further molecule of acetyl-CoA to produce 3-hydroxy-3-methylglutaryl-CoA (HMG-CoA). HMG-CoA reductase (HMGR) utilises two NADPH in the two-step reductive deacylation of HMG-CoA to produce the namesake mevalonate intermediate.^{18,19} HMGR is the target of statins; an important class of drugs involved in lowering cholesterol associated with cardiovascular disease.²⁰ For eukaryotes, mevalonate is diphosphorylated to mevalonate-5-diphosphate by the sequential adenosine triphosphate (ATP)-dependent reactions of mevalonate kinase (MK) and phosphomevalonate kinase (PMK). A third equivalent of ATP is used to initiate the following decarboxylation and dehydration by mevalonate-5-diphosphate decarboxylase (MDD), resulting in the generation of IDP. Finally, DMADP is produced from the reversible tautomerisation of the IDP *proR* C2 proton by isopentenyl diphosphate isomerase (IDI).^{18,19}

Two variations to this pathway have been identified in some archaea and bacteria. Following the Archaea I branch of the pathway, mevalonate-5-phosphate is instead converted to isopentenyl phosphate by mevalonate decarboxylase, which is phosphorylated to become IDP directly via the ATP-dependent isopentenyl kinase (IPK).²¹ For the alternative route known as the Archaea II pathway, mevalonate is diphosphorylated to become mevalonate-3,5-bisphosphate following sequential ATP-dependent conversions by mevalonate-3-kinase and mevalonate-3-phosphate-5-kinase. This substrate undergoes a decarboxylation (of which the related enzyme is yet to be characterised) to yield isopentenyl phosphate, finally joining the Archaea I pathway to produce IDP via IPK.²⁰

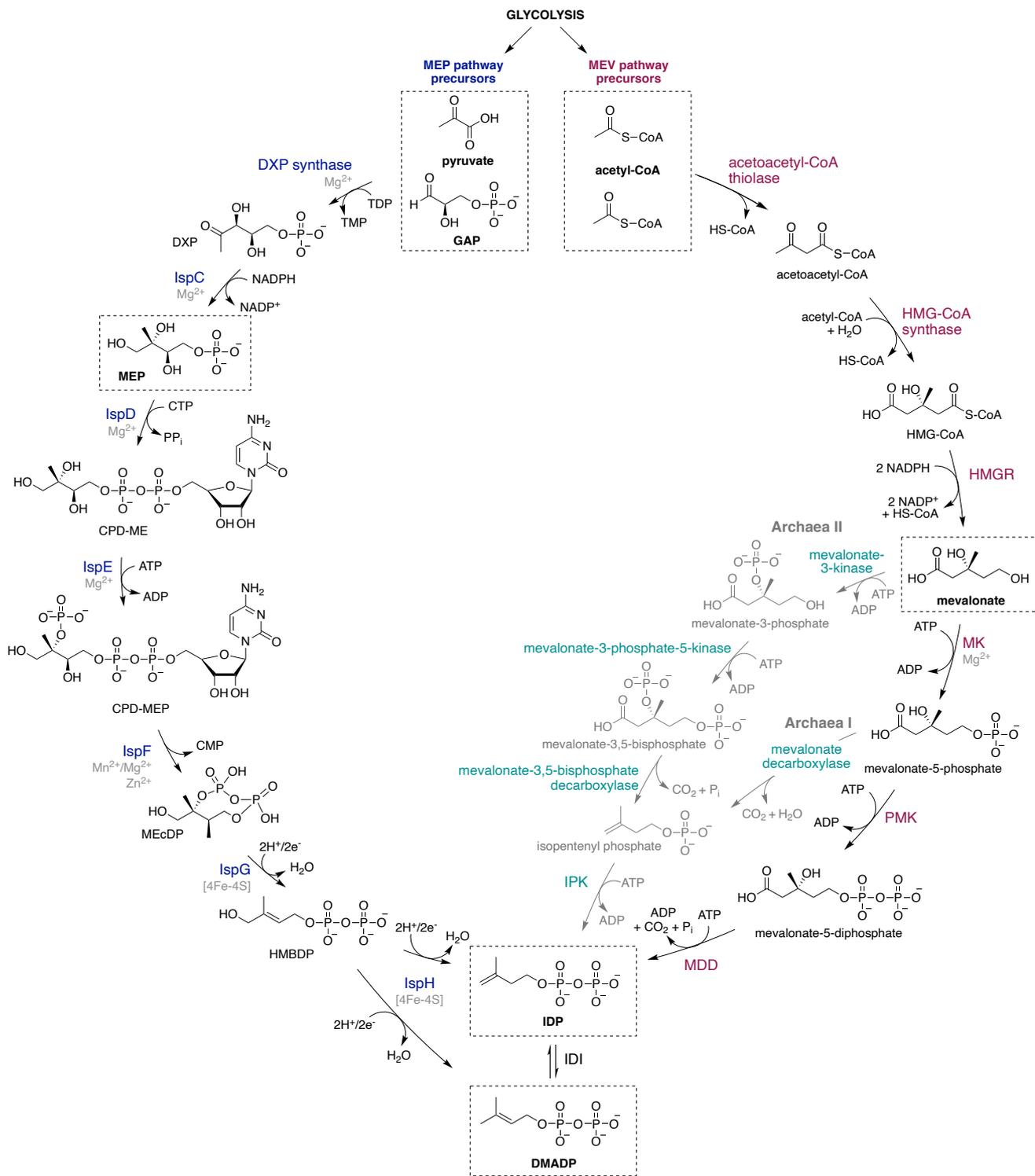


Figure 1.3: A summary of the MEV and MEP pathways leading to the C₅ terpenoid precursors IDP and DMADP. With key substrates highlighted in boxes, the MEV pathway is shown on the right and the MEP is on the left. The alternative Archea I and II pathways are shown from respective branch points of the MEV in grey.

First recognised in the 1950s, the MEV pathway was considered the universal isoprenoid pathway for all domains of life until the discovery of the entirely separate MEP pathway (Figure 1.3) 40 years later.^{22,23} The bacterial pathway alternatively begins with the condensation of pyruvic acid and glyceraldehyde-3-phosphate (GAP) to yield 1-deoxy-D-xylulose-5-phosphate (DXP), catalysed by DXP synthase using the cofactor thiamine diphosphate (TDP). This pathway is often referred to as the DXP pathway because of this intermediate, although DXP is utilised in other biosynthetic pathways so is not considered the first committed precursor to IDP and DMADP of this pathway.¹⁸ DXP reductoisomerase (IspC) then catalyses the isomerisation of DXP and subsequent NADPH-dependent reduction to the key intermediate 2-C-methyl-D-erythritol-4-phosphate (MEP). The next step involves the transfer of cytidine monophosphate (CMP) from the cofactor cytidine 5'-triphosphate (CTP) by MEP cytidyltransferase (IspD), producing 4-(cytidine 5'-diphospho)-2-C-methyl-D-erythritol (CDP-ME). CDP-ME is subsequently phosphorylated by CDP-ME kinase (IspE) at the 2-hydroxy position to give 2-phospho-4-(cytidine-5'-diphospho)-2-C-methyl-D-erythritol (CDP-MEP). In the next reaction, the CMP moiety is eliminated forming an unusual cyclic diphosphate, 2-C-methyl-D-erythritol-2,4-cyclodiphosphate (MEcDP), catalysed by MEcDP synthase (IspF). Two iron-sulfur cluster proteins are responsible for the final two steps of the pathway which both involve a two-electron reduction and elimination of water, although the precise mechanisms are unknown. MEcDP first undergoes a ring-opening to (*E*)-4-hydroxy-3-methylbut-2-enyl diphosphate (HMBDP) by HMBDP synthase (IspG). Lastly, HMBDP is converted to either IDP or DMADP by HMBDP reductase (IspH), with one molecule of DMADP produced for every five to six molecules of IDP.^{24,25}

1.1.2 Prenyltransferases

The relative proportions of DMADP and IDP are maintained in the cell by IDIs in an approximate ratio of 3:7 respectively,²⁶ providing DMADP, the primary terpene building block, with an excess of IDP for successive addition reactions. These lead to the generation of linear polyprenyl diphosphates of a variety of chain lengths which typically span from C₁₅ to C₁₂₀.²⁷ These reactions are catalysed by isoprenyl diphosphate synthases, commonly known as prenyltransferases, by an ionisation–condensation–elimination mechanism (Figure 1.4).²⁸ Magnesium (or manganese) ions bind the diphosphate moiety and enable its ionisation to form an allylic carbocation, which undergoes a nucleophilic attack from double-bond electrons of the incoming substrate. The tertiary carbocation formed is quenched by the removal of a C2 proton to give a double bond in either *trans* (*E*) or *cis* (*Z*) geometry.²⁹ Hence, prenyltransferases are either referred to as *cis*- or *trans*-prenyltransferases, but they are also classified as short, medium or long chain prenyltransferases depending on the length of the

product. The prochirality of the abstracted proton was thought to determine the resulting geometry but this has since been found not to be the case, as for both short chain *cis*- and *trans*-prenyltransferases, the *proR* proton is eliminated to form the double bond.²⁹

Coupling reactions take place most commonly in the 1'-4 regular 'head-to-tail' fashion, but 'head-to-head' and other types of irregular coupling patterns can occur.³⁰ Short chain *trans*-prenyltransferases catalyse the head-to-tail addition of IDP to DMADP to form geranyl diphosphate (GDP, C₁₀) and subsequent elongation reactions to produce (2*E*,6*E*)-farnesyl diphosphate (FDP, C₁₅), (2*E*,6*E*,10*E*)-geranylgeranyl diphosphate (GGDP, C₂₀) and (2*E*,6*E*,10*E*,14*E*)-geranylfarnesyl diphosphate (GFDP, C₂₅).³¹ These compounds are the substrates of terpene synthases (also known as cyclases) which generate the myriad of terpene natural products with the prefixes of mono-, sesqui-, di- and sester- for the respective 10, 15, 20 and 25 carbon structures.³² Recently however (discussed in Section 1.5.4), an unusual family of short chain *cis*-prenyltransferases have been identified which catalyse the formation of neryl diphosphate (NDP, C₁₀),³³ (2*Z*,6*Z*)-farnesyl diphosphate ((*Z,Z*)-FDP, C₁₅)³⁴ nerylneryl diphosphate (NNDP, C₂₀),³⁵ Figure 1.4).

In addition to head-to-tail couplings, other condensation reactions between isoprenyl diphosphates commonly occur. The important metabolites squalene (C₃₀, triterpene precursor to sterols) and phytoene (C₄₀, tetraterpene precursor to carotenoids) are biosynthesised by medium chain *trans*-prenyltransferases via head-to-head 1'-1 condensation reactions between two equivalents of FDP and GGDP respectively.³⁶ Condensation reactions yielding isoprenyl diphosphates with mixtures of double bond geometries are often seen in plants and bacteria, such as C₅₅ undecaprenyl diphosphate, which consists of a *Z*-prenyl chain attached to an (2*E*,6*E*)-farnesyl chain.³⁷ These larger compounds form essential constituents of cell membranes.³⁸ Long chain coupling reactions are typically carried out by *cis*-prenyltransferases resulting in high molecular weight polyprenyl diphosphates numbering many thousands of carbons; as many as C_{>100,000} is seen in natural rubber.³⁹

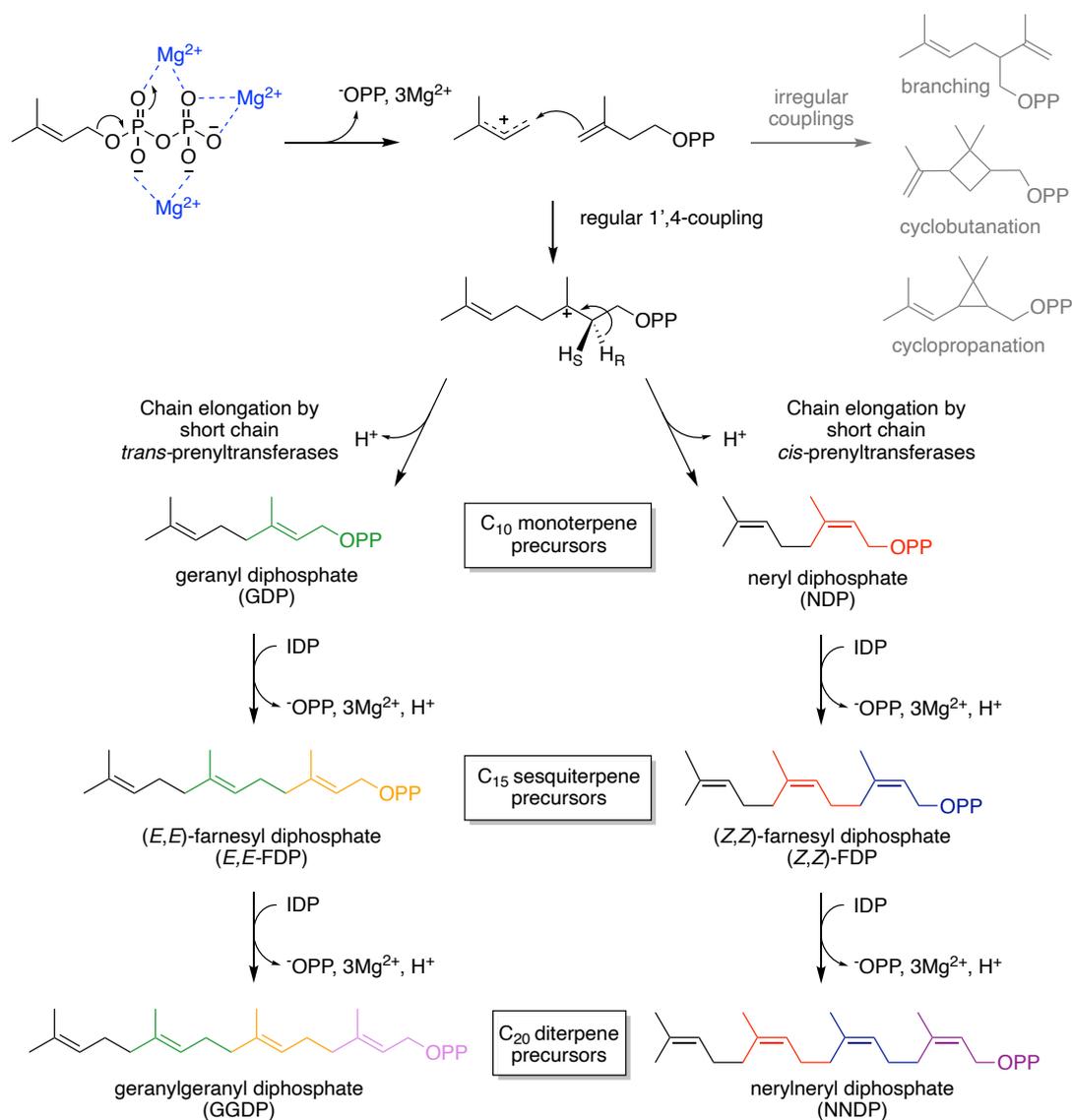


Figure 1.4: Head-to-tail coupling reactions carried out by short chain *cis*- and *trans*-prenyltransferases between DMADP and successive IDP molecules to generate linear polyprenyl diphosphates. The coordination of the diphosphate moiety with three Mg^{2+} ions is representative of the *trans*-prenyltransferase binding mode.

Cis- and *trans*-prenyltransferases belong to two evolutionary and structurally distinct families. The *trans*-prenyltransferases share an α -helical bundle structure with two phosphate binding domains comprised of conserved DDXD motifs, also found in the related class I terpene synthases (Figure 1.5).¹ Contrastingly, *cis*-prenyltransferases have two α -helices and four β -sheets forming the active site with a RX_5RX_nE motif to bind the diphosphate: the arginines bind directly whilst the glutamic acid binds via a Mg^{2+} bridge (like *trans*-prenyltransferases).³¹ Bulky amino acids positioned in the bottom of the hydrophobic active sites of both types of prenyltransferase, upstream from the binding motifs, are important in determining the carbon

chain length of the product.^{37,40} Mutation of the tyrosine (Y81) five positions away from the DDXXD motif of *Bacillus stearothermophilus* (*E,E*)-FDP synthase (FPPS) for example, leads to the formation of C₁₀, C₂₀, C₂₅ and C₃₀ isoprenoid diphosphates; an analogous observation is seen with residues F112 and F113, four and five amino acids away from the metal-binding motif, of avian FPPS.²⁷

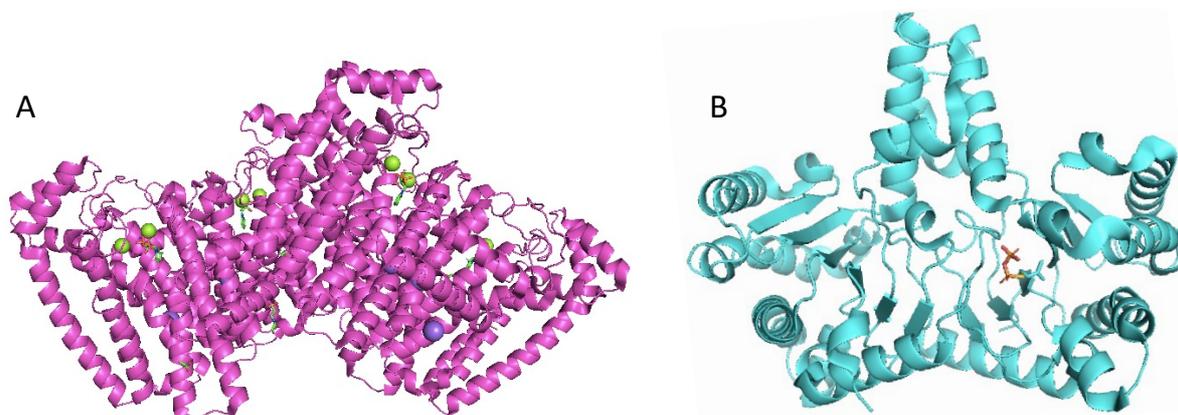


Figure 1.5: Homology models of (A) dimerised *trans*-prenyltransferase, (*E,E*)-FDP synthase (FPPS, PDB: 4K10), comprising of multiple α -helices arranged around a central chamber where the active site is located.⁴¹ (B) Dimerised *cis*-prenyltransferase, (*Z,Z*)-FDP synthase (zFPS, PDB: 5HXQ), comprising of α -helices and β -sheets in a Rossmann fold-like assembly.⁴²

1.2 Terpene synthases

1.2.1 Cyclisation chemistry

The catalytic formation of terpenes from their linear achiral precursors by terpene synthases are widely regarded as some of the most interesting biochemical conversions in nature. Owing to inherent conformational flexibility and reactivity, a huge diversity in structure can be achieved from a single precursor via different cyclisation modes, the number of which increases with each C₅ unit. Monoterpene synthases catalyse the cyclisation of C₁₀ GDP (or NDP) to produce monoterpenes (Figure 1.6). This takes place via an ionisation-recombination step that sees the isomerisation of GDP into a chiral linalyl diphosphate intermediate, enabling the subsequent 1,6-cyclisation to give either the *R* or *S* terpinyl cation intermediate, from which all regular monoterpenes originate.

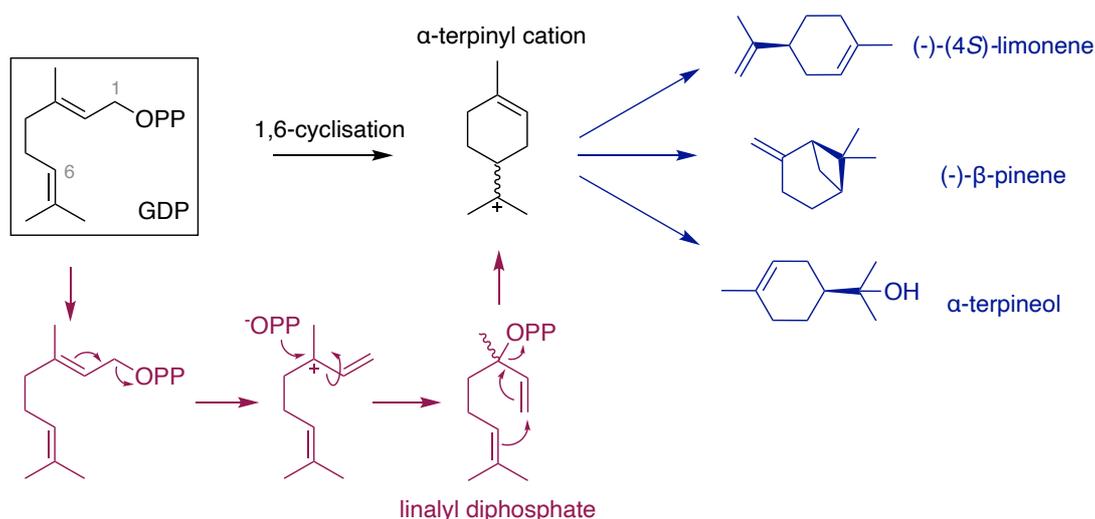


Figure 1.6: C₁₀ GDP undergoes a single 1,6-cyclisation to the terpinyl cation to yield monoterpenes via an ionisation-recombination mechanism with the diphosphate moiety.

For sesquiterpene synthases however, an extra five carbon unit in the precursor FDP facilitates six different cyclisation modes (Figure 1.7). Two occur directly as 1,10- and 1,11-cyclisations, whilst similarly to monoterpenes, the other four occur via an ionisation-recombination sequence, producing a chiral nerolidyl diphosphate intermediate which is oriented to undergo 1,6-, 1,7-, 1,10- and 1,11-cyclisations. The possible cyclisation modes increases again with the additional C₅ subunit of C₂₀ GGDP: diterpenes can be derived from 1,6-, 1,7-, 1,10-, 1,11-, 1,14-, 1,15- and 3,8- cyclisations for class I synthases, whilst for class II synthases, rings are formed simultaneously via 10,15- and 6,11-cyclisations.¹ These cyclised cationic intermediates can then undergo complex reaction cascades where typically over half the substrate carbon atoms undergo changes to bonding, hybridisation and stereochemistry, leading to a diverse myriad of products with multiple rings and stereocentres.³²

These varied hydrocarbon skeletons are often functionalised further (subsequently known as terpenoids) by enzymes in the production of downstream metabolites, greatly expanding the range of structures and functions of these natural products. These decorating reactions are predominantly carried out by cytochrome P450s, producing varieties of oxidised products with epoxide and alcohol functionality.^{43,44} As many terpenes have useful bioactivities and applications, understanding how they are made so efficiently by their respective enzymes is of great interest for chemical and biocatalytic industries.

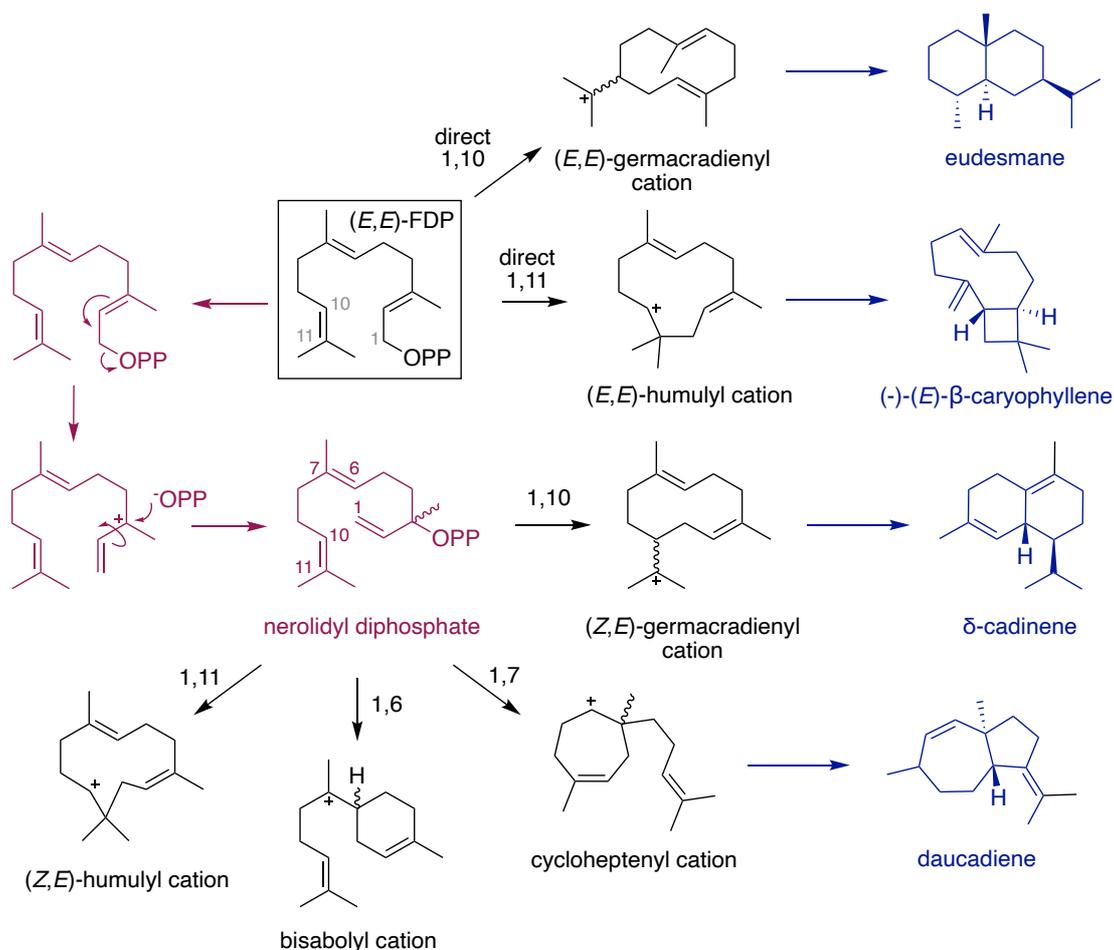


Figure 1.7: C₁₅ FDP has six cyclisation modes for sesquiterpene formation: two occur directly whilst four arise following precursor isomerisation-recombination.

1.2.2 Structure and function of terpene synthases

The relationship between the structure of terpene synthases and their function was first revealed in 1997 with the publication of the crystal structures of pentalenene,⁴⁵ 5-*epi*-aristolochene⁴⁶ and squalene synthases.⁴⁷ They were shown to comprise of evolutionarily distinct architectural domains made of multiple α -helices, known as α , β and γ domains. All terpene synthases contain one or more of these domains and are categorised as either class I or II depending on their mechanism of initial carbocation formation. Class I active sites are located in the α -domain and use three metal cations and three basic amino acids to ionise the diphosphate group of the substrate, hence they are characterised by the presence of two conserved metal-binding motifs: DDXXD of helix D binds two Mg²⁺ cations and (N,D)DXX(S,T)XXXE of helix H binds the third (metal-binding ligands underlined).¹ Alternatively, class II active sites are found at the interface of β and γ domains where a

conserved DXDD metal-binding domain facilitates protonation of an isoprenoid π -bond (or corresponding epoxide⁴⁸) using an aspartate as a general acid.⁴⁹ Bifunctional terpene synthases, which possess both class I and II functionality, are also known.^{50,51}

Class I synthases are anticipated to have evolved from *trans*-prenyltransferases and share the metal-binding α -fold first observed in avian FPPS, known as the terpene fold.⁴¹ Bacterial and fungal class I cyclases comprise solely of an α -domain whilst those of plants are known to be multi-domain (either $\alpha\beta$ or $\alpha\beta\gamma$).⁵⁰ This is exemplified in the first reported structures (Figure 1.8): pentalenene synthase from *Streptomyces exfoliatus* consists of an α domain only;⁵⁰ 5-*epi*-aristolochene synthase from *Nicotiana tabacum* has an $\alpha\beta$ structure, and whilst only the α -domain is catalytically active, the non-functional domain still participates in catalysis by capping the active site when the substrate is bound.⁴⁶ Whereas class I cyclases like these are typically found to cyclise the smaller C_{10} and C_{15} linear substrates, class II cyclases use the C_{20} and C_{30} linear substrates in the majority of cases, although exceptions for both classes are known.⁵² Taxadiene synthase is a diterpene (C_{20}) synthase whose structure comprises of $\alpha\beta\gamma$ domain architecture and the class II $\beta\gamma$ domain is catalytically active but the α is not.^{50,53} Finally, the structures of the class II triterpene (C_{30}) synthases squalene-hopene cyclase⁴⁷ and oxidosqualene synthase⁴⁸ are comprised only of $\beta\gamma$ domains.

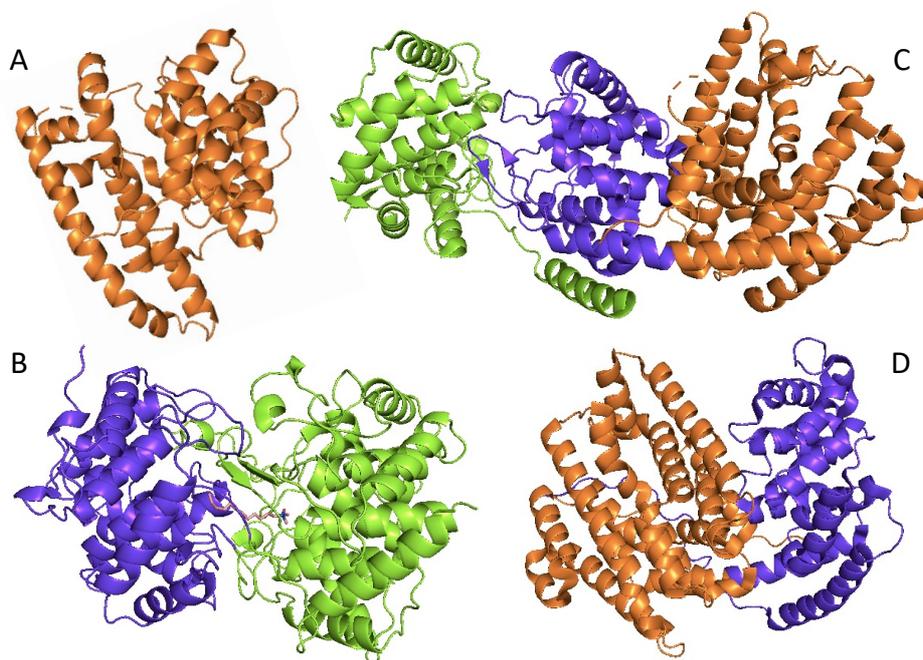


Figure 1.8: The diverse range of structural assemblies of terpene synthases exemplified by homology models of (A) pentalenene synthase (PDB: 1HM4), (B) 5-*epi*-aristolochene synthase (PDB: 5IK6), (C) taxadiene synthase (PDB: 3P5P) and (D) squalene-hopene synthase (PDB: 1SQC). α -Domain architectures coloured in orange, β -domains in purple and γ -domains in green.

Although class I and II synthases differ in respect to metal-binding and mechanism of carbocation formation, they adopt common strategies for controlling substrate cyclisation. The shape of the active site and positioning of the metal-binding domains critically provide a template in which the substrate is orientated and folded into the precise conformation for the ensuing reaction cascade. Many synthases navigate the similar reactive pathways and reactive carbocation intermediates with high fidelity, catalysing formation of a single product. Some however are multi-product, such as γ -humulene synthase of grand fir (*Abies grandis*), the well-known example which produces 52 hydrocarbon products from FDP. Reactive carbocation intermediates are protected from premature quenching by the hydrophobic pocket and stabilised through various side chains interactions such as cation- π or quadrupole interactions with aromatic amino-acid (tryptophan, tyrosine, phenylalanine): at least two or three are present in the active sites of known cyclases.⁵⁴ Reaction cascades typically involve ring closures through attack by double bonds, hydride or alkyl shifts and carbocation rearrangements,⁵⁵ often facilitated by acid-base catalysis with participation from active site water, amino acid side chain or ionised diphosphate.⁵⁶ The reaction cascade is usually terminated via proton elimination, either spontaneously or by a general base, or nucleophilic incorporation of a solvent water molecule. If the latter, this process is tightly controlled to avoid premature quenching of carbocation intermediates.⁵⁷ A variety of biological and chemical methods have been used to study such aspects of the terpene synthase structure-function relationship and their remarkable cyclisation mechanisms.

1.3 Techniques for investigating the mechanisms of terpene synthases

The bond forming and breaking steps of in the conversion of substrate to product can be analysed via substrate analogues which perturb the electronics of the system and mimic hypothesised intermediates, or isotopologues with key atoms radiolabelled which can be traced in the final compound. These can be used alongside site-directed mutagenesis (SDM) of critical amino acids to probe their effect on catalytic efficiency and product outcome, building a detailed picture of how the complex chemical transformations are controlled by the enzyme.

Many interesting mechanistic strategies and the experimental tools employed to study them have been exemplified through detailed study of *Penicillium roqueforti* and *Aspergillus terreus* (+)-aristolochene synthases (PR-AS and AT-AS respectively) by the groups of Allemann, Cane and Christianson. Aristolochene is a sesquiterpene precursor to a large family of fungal metabolites which include PR-toxin, gigantone and bipolaroxin.⁵⁸ The proposed mechanism of formation of aristolochene follows a C1-C10 cyclisation to generate a germacrene A intermediate, the double bonds of which undergo an intramolecular ring closure to generate

the bicyclic structure of the eudesmane cation (Figure 1.9). Further stereospecific hydride and methyl transfers yields the final product which contains three stereocentres and two double bonds.

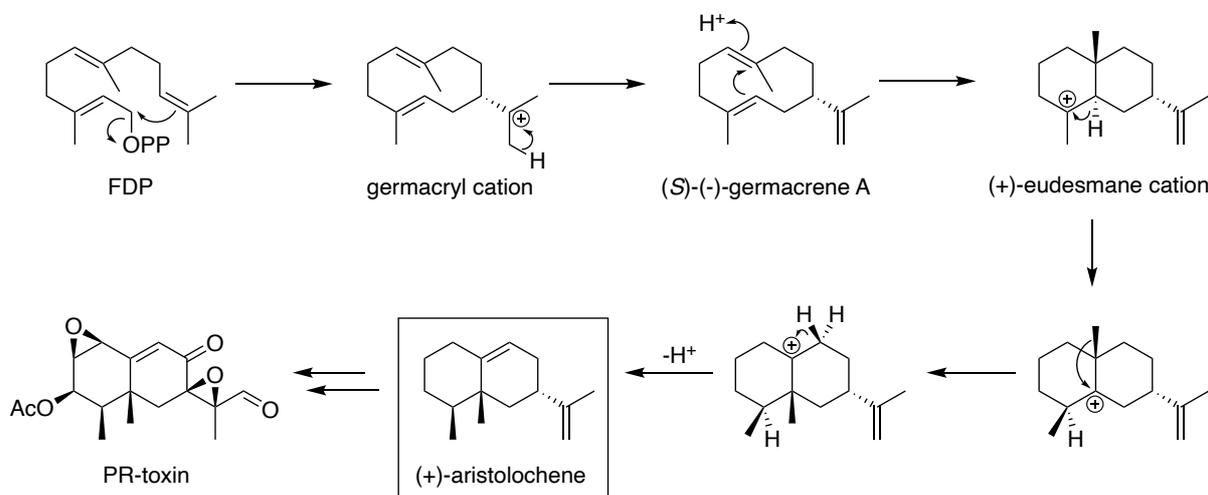


Figure 1.9: Mechanism of (+)-aristolochene formation by aristolochene synthase.

1.3.1 Crystallography

Crystal structures of enzymes complexed with their natural substrates or carefully designed analogues are instrumental for revealing which interactions between an enzyme and its substrate are required for arranging and stabilising its reactive conformation.^{59–61} The determination of crystal structures of both PR-AS⁶² and AT-AS⁶³ provided an opportunity to compare the structure of two separately evolved enzymes which produce the same compound. Although sharing only 61% sequence identity, their α -helical fold active sites are complementary to the shape of aristolochene and to one another, although markedly dissimilar to that of 5-*epi*-aristolochene.⁴⁶ The pocket of their α -helical terpene fold active sites are hydrophobic with a high proportion of aromatic residues, whilst their upper areas are more hydrophilic and contain polar and charged residues.^{62,63}

Crystallographic analysis of AT-AS structures complexed with FDP and fluoro-analogues via crystal soaking provided a series of snapshots which showed that substrate binding occurs with the tetrameric enzyme first in an open conformation, followed by coordination of two Mg^{2+} ions which triggers closure of the active site of one subunit. The FDP then aligns to its reactive conformation upon association of the third Mg^{2+} ion and the diphosphate group is ionised, initiating the reaction sequence.⁵⁸ Optimal overlap between the reacting π bond and C-O bond

orbitals of the diphosphate ester in the C1-C10 concerted ring closure step was shown to be facilitated by key hydrophobic tyrosine and phenylalanine residues.⁵⁸ The subtle conformational rearrangements required to initiate the cyclisation reveal the individual roles played by each metal cation for catalysis, and the stepwise process leading to breaking of the phosphodiester bond.

1.3.2 Mutagenesis

Rational mutation of specific amino acids in the sequence of a protein is a technique widely used to reveal the steric, electronic and functional contributions of individual side chains to catalysis.^{64–66} The PR-AS amino acid residues responsible for binding the metal cofactors and triggering the substrate ionisation were identified by various mutants in the metal-binding motifs, D₁₁₅DVLE and N₂₄₄DIYSYDKE. Some were completely inactive (D115N, N244L), many gave decreased catalytic efficiency (S248A, E252D), whilst some led to increasing amounts of the germacrene A, reaching 100% in the E252Q mutant, providing insight into its intermediary role.⁶⁷

A number of mutagenesis studies were carried out to define to what degree the aromatic side chains of active site W, Y and F provide stabilisation of germacryl and eudesmane cationic intermediates.⁶⁸ PR-AS Y92 was postulated to aid FDP folding and assist the 1,10-cyclisation through steric control as linear sesquiterpene products (primarily farnesenes) were observed when the size of the side chain was reduced by mutation to V (40%), C (70%) and A (80%).⁶⁹ These studies culminated in the generation of a series of W334 PR-AS mutants with non-canonical *para*-substituted phenylalanines with increasing electron withdrawing properties (naphthyl, *p*-(Cl)F, *p*-(CF₃)F, *p*-(NO₂)F), fine-tuned to offer different degrees of cation- π electronic stabilisation. These produced correspondingly increasing amounts of germacrene A when formation of the energetically more demanding eudesmane cation (Figure 1.9) intermediate was less favourable.⁷⁰

1.3.3 Fluorinated substrate analogues

Fluorinated substrate analogues are useful in mechanistic studies for influencing double bond and cation energies. They mimic the size and binding interactions of equivalent hydrogen substituents but significantly perturb the local electronic properties due to increased electronegativity. Cations at an α -carbon will be stabilised by π -donation from the lone pair of the fluorine, whilst cations at β - γ -positions will be destabilised by an inductive effect (Figure 1.10). In some cases, fluoro-analogues may prove more useful than mutagenesis which can alter the template of the active site and provide alternative pathways for product formation.^{71,72}

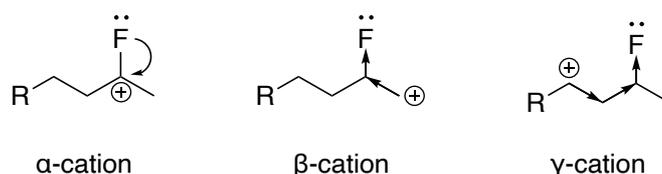
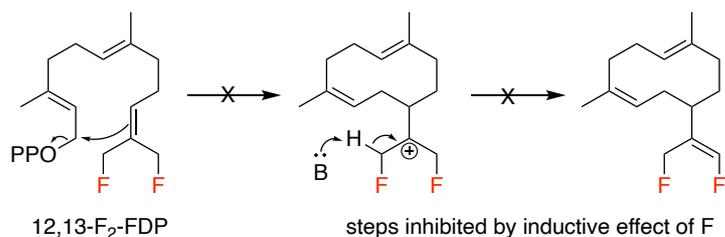


Figure 1.10: Stabilising electron donating and destabilising electron withdrawing effects of α - β - and γ - fluoro-substituted cations.

Fluoro analogues have been used to investigate the mechanism of aristolochene formation. The initial C1-C10 ring closure and generation of the diphosphate ion to give the germacryl cation was revealed to be concerted rather than stepwise by incubation of PR-AS with 12,13-F₂-FDP (Figure 1.11). This substrate acted as a potent inhibitor to AS, forming no enzymatic products upon incubation due to destabilised formation of the germacryl cation intermediate via the inductive effect of fluorine. Had the mechanism been stepwise, following loss of diphosphate, the intermediary carbocation could have been quenched leading to fluorinated farnesene products.⁷³ Another fluoro analogue, 2-F-FDP, was used to provide evidence that the next step in the mechanism goes via a germacrene A intermediate on route to the eudesmane cation. Incubation with AS gave 2-fluorogermacrene A as the sole product as perturbation of the double bond with the fluorine substituent prevented its subsequent intramolecular cyclisation to form the eudesmane cation.⁷¹

A Evidence for a concerted C1-C10 ring closure:



B Evidence for a germacrene A intermediate:

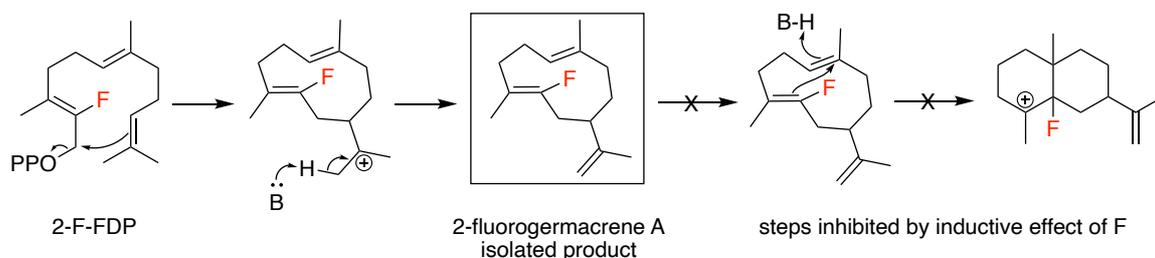


Figure 1.11: (A) Inhibition of aristolochene synthase by 12,13-F₂-FDP gave evidence for a concerted C1-C10 ring closure. (B) Isolation of 2-fluorogermacrene A via 2-F-FDP supports formation of the eudesmane cation intermediate.

1.3.4 Aza substrate analogues

The intrinsic reactivity and transient nature of carbocation intermediates can make it difficult to analyse their role in terpene synthase mechanisms as they cannot be isolated and characterised. Nitrogen containing substrate analogues can be quaternary functionalised to corresponding sp³ ammonium or sp² iminium cations which imitate carbocation intermediates electrostatically and stereochemically.⁷⁴ Due to the aza-stabilisation, cations which mimic genuine mechanistic intermediates will act as inhibitors by binding tightly to the enzyme without being turned over; a useful property for crystallisation studies.⁷⁵ This effect was exploited with the incubation of PR-AS with several eudesmane aza analogues (Figure 1.12) which inhibited different mechanistic steps. Inorganic diphosphate-dependent inhibition for the iminium analogues suggested the diphosphate was playing a stabilising (and potential general-acid) role in those interactions, however diphosphate-independent inhibition was seen for the ammonium analogue implying it binds the active site the way the genuine eudesmane cation intermediate would.^{76,77}

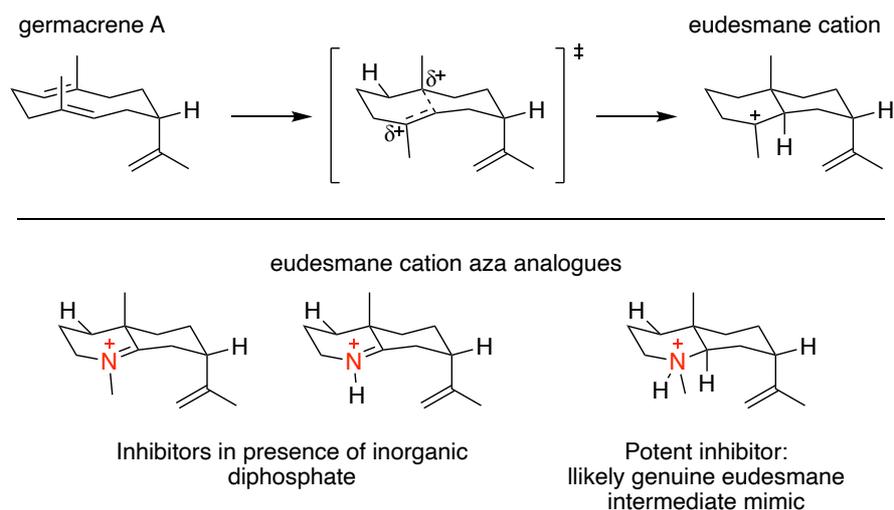


Figure 1.12: Aza analogues of FDP designed to mimic eudesmane cation intermediates in the aristolochene synthase mechanism.

1.3.5 Isotopologues

Another powerful technique for tracing mechanistic steps of terpene synthases involves isotope labelling of key substrate atoms with isotopes (typically ^{13}C and ^2H), as the relative change in atomic environment over the course of a reaction can be readily analysed using standard analytical techniques. Radiolabelled substrates (typically ^3H) furthermore enable measurement of steady-state kinetics with high sensitivities appropriate to the K_M concentration range.⁷⁸ FDP syntheses have been developed for incorporating isotopes throughout the substrate,⁷⁹ which have subsequently uncovered stereo- and regiochemical mechanistic details for production of many sesquiterpenes, including (+)- δ -cadinene,⁸⁰ amorpha-4,11-diene⁸¹ and (+)-germacrene D.⁸²

FDP isotopologues labelled selectively with deuterium were used in early work with AT-AS to study individual stereospecific hydride shifts. Incubation of both [$^2\text{H}_3$ -12]-FDP and [$^2\text{H}_3$ -13]-FDP with the synthase and subsequent ^2H NMR spectroscopic analysis of the product, with comparison to unlabelled aristolochene, revealed that the *cis* methyl group is deprotonated in the reaction step converting the germacryl cation into the proposed germacrene A intermediate (Figure 1.9). The corresponding environments were shifted downfield for alkene ^2H -12 ($\delta = 4.71$ ppm) whilst remaining upfield for unperturbed methyl ^2H -13 ($\delta = 1.69$ ppm).⁸³ Further ^2H NMR spectroscopic analysis of incubations arising from (*R*)- and (*S*)-[$^2\text{H}_1$ -1]-FDP proved that the loss of diphosphate occurs with an inversion of stereochemical configuration. Aristolochene from (*R*)-[$^2\text{H}_1$ -1]-FDP gave a single deuterium

peak at $\delta = 1.76$ ppm relating to the equatorial H-6 (H-6_{re}) whilst the peak at $\delta = 1.17$ ppm derived from (S)-[²H₁-1]-FDP was found at the axial H-6 (H-6_{si}) position (Figure 1.9).⁸⁴

The extensively studied aristolochene synthases exemplifies the complex combinatorial chemistry that is mediated by these and many other related enzymes. The tools outlined - crystallography, substrate analogues, SDM, as well as computational modelling^{54,85} - are used together to provide evidence for precisely how the linear substrates are converted into stereochemically and structurally complex natural products. Understanding terpene synthase structure-function relationships may allow *de novo* enzyme design and advance biocatalysis to produce economically important compounds inspired by nature.

1.4 Production of terpenes

The wide range of complex hydrocarbon scaffolds of terpene natural products which confer a range of properties and bioactivities has meant terpenoids have found uses spanning the entire chemical industries. A significant challenge with using terpene-based compounds is obtaining a sufficient supply. Often chemical syntheses are unfeasible, so compounds are derived from the natural (plant) source which are typically very low yielding and can be unreliable, leading to market instability.^{86,87} Research has sought to overcome these limitations by adopting semisynthetic approaches using increasingly complex terpene precursors which already contain most of the stereochemistry and some functionalisation, installed efficiently and selectively via enzymatic routes. These precursors are then derivatised to the more complex final structures by chemical synthesis.⁸⁸

1.4.1 Semisynthetic and *in vivo* production of terpenoids

In the field of biocatalytic terpenoid production, most approaches have focused on *in vivo* systems. Biotechnological advances in DNA sequencing and synthesis, and engineering tools, have enabled the placement of heterologous genes or whole metabolic pathways into genetically tractable organisms for chemical production. Microbial fermentation processes typically require ambient conditions and inexpensive starting materials, such as simple sugars, are converted into products that are currently derived from non-renewable or limited natural resources.⁸⁹ Preferred microbial hosts include *Escherichia coli* and *Saccharomyces cerevisiae* which are well understood genomically and amenable to engineering due to decades of fundamental research. Metabolically engineered strains may possess modified forms of the MEV and/or non-mevalonate MEP pathways to increase the availability of DMADP and IDP for downstream compound production.⁸⁹⁻⁹¹

The most significant example of microbial engineering is the semisynthetic production of the important anti-malarial sesquiterpenoid drug artemisinin. This research effort highlights the potential of metabolic engineering but also the limitations. Malaria is a devastating disease, suffered by approximately 228 million in 2019, with 67% of the estimated 405,000 deaths being children under the age of five.⁹² Artemisinin (in combination therapies) is the World Health Organisation's (WHO) recommended first-line malaria treatment and is sourced from sweet wormwood (*Artemisia annua*), the supply of which can be unstable. Due to the scale of the disease and it predominantly affecting low-income countries, sourcing artemisinin and derivatives affordably and reliably is the subject of considerable research and funding.¹³

One route uses biologically derived artemisinic acid as a precursor to semisynthetic artemisinin. Artemisinic acid contains four of the seven stereogenic centres and the same basic carbon scaffold as artemisinin but is more accessible. Following success using *E. coli*,⁸⁹ the mevalonate pathway of *S. cerevisiae* was engineered to accumulate FDP which was converted *in vivo* to precursor amorpha-4,11-diene (amorphadiene) by amorphadiene synthase (ADS), achieving titres of 40 g/L⁹³ (Figure 1.13). ADS controls the formation of two 6-membered rings, four stereocentres and two double bonds from FDP in a single step with precise regio- and stereochemical control. With further engineering of cytochrome P450s and an alcohol dehydrogenase from *A. annua*, it was possible to produce the downstream oxidised artemisinic acid in titres of 25 g/L.⁹¹ Following extraction from yeast, this was converted to artemisinin in four chemical steps achieving 55% overall yield. Hailed as a triumph for synthetic biology, this process was later industrialised by Sanofi and used to produce 35 tonnes of semisynthetic artemisinin in 2013.⁹⁴

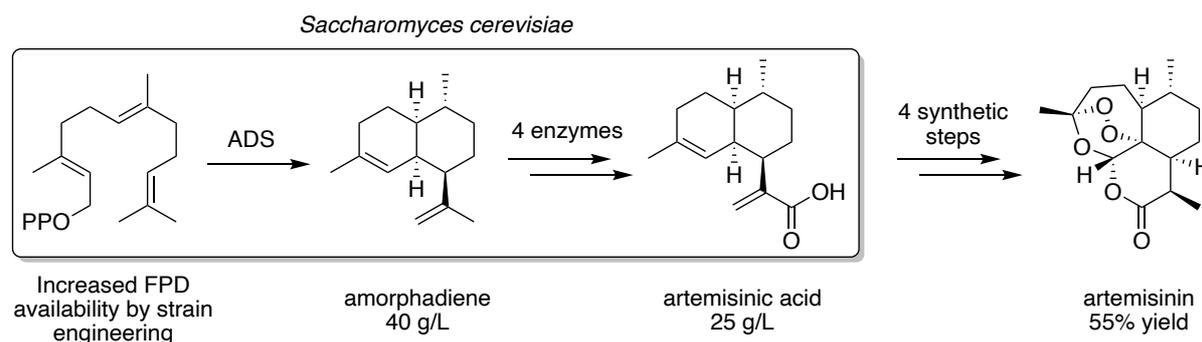


Figure 1.13: Semisynthetic approach to producing artemisinin using metabolically engineered yeast.^{91,93}

Despite the scientific breakthroughs achieved in development of this process, production costs failed to be competitive within the particularly complex market, leading to its eventual abandonment.⁹⁵ With the advent of this process, there was great optimism about the potential of *in vivo* terpene production which has fuelled a large body of research, however, whole-cell systems for fine chemicals are limited beyond the scientific literature.⁸⁶ Compound overproduction competes with an organism's central carbon metabolism in highly regulated systems comprising of many enzymes and cofactors. Microbial engineering is therefore extremely complex and significant time and resource is required to reach production levels that can rival existing production methods. The semisynthetic artemisinin process was supported by \$64 million investment from the Bill & Melinda Gates Foundation and took around 10 years to develop.⁹⁵ *In vivo* production is therefore only possible where value justifies the cost of developing a genetically engineered organism.⁹⁶

1.4.2 Unnatural terpenes from substrate analogues

Alternative semisynthetic routes to terpene scaffolds have been explored which take advantage of inherent terpene synthase plasticity. Although typically terpene synthases catalyse the formation of one major product from a single preferred substrate, some terpene synthases are promiscuous in respect to both substrate acceptance and product profile. For example, β -himachalene synthase accepts GDP, FDP and GGPD⁹⁷ whilst the low-fidelity *epi*-isozizaene synthase produces 79% *epi*-isozizaene as well as various other sesquiterpenes.⁹⁸ Plasticity enables acceptance and conversion of non-natural substrates into complex, enantiopure analogues that would be inaccessible through synthetic chemistry, with potential economic value across industries that already utilise terpene natural products.

This approach was employed by Demiray *et al.* of the Allemann group in the development of an alternative chemoenzymatic synthesis of dihydroartemisinic aldehyde (DHAAI), a key upstream intermediate of artemisinin⁹⁹ (Figure 1.14). Commercially available farnesyl chloride was converted in two steps to an unnatural analogue of FDP - 12-hydroxy-FDP (12-OH-FDP) - which in turn was converted directly into DHAAI by ADS. Artemisinin can then be synthesised from DHAAI in a further four synthetic steps. This short and efficient chemoenzymatic route avoids the redox chemistry required from the natural ADS product, amorphaadiene, by instead increasing the oxidation state of the linear, achiral precursor.¹⁰⁰

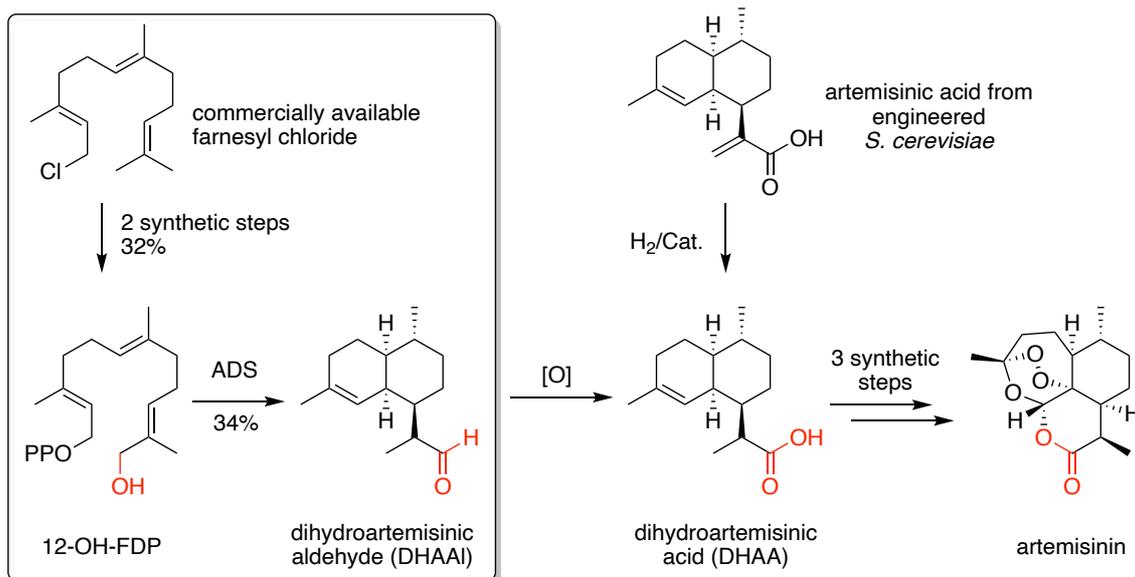


Figure 1.14: A short, efficient chemoenzymatic synthesis of dihydroartemisinic aldehyde via direct conversion of 12-OH-FDP by ADS.⁹⁹

Unnatural substrates are increasingly being used in this way to deliberately explore the chemical space available via terpene synthases. Analogues originally used to examine catalytic mechanisms or act as inhibitors for crystallography were observed to yield terpenoids with unnatural functionality.^{71,101} This has led more recently to the deliberate probing of terpene synthases with heteroatom modified substrates which have created new cyclic ethers¹⁰² and macro- and tricyclic terpenes with novel olfactory properties.¹⁰³

This approach was developed into a methodology for the discovery of novel agrochemicals. As with medicines, the introduction of new agrochemicals has been declining, driving the rise in molecular target-based screening approaches for novel compounds.¹⁰⁴ Bioactivity may arise through the recognition of a compound by an insects' target receptor: for example, the detection of the sesquiterpene (*S*)-germacrene D by arthropods such as aphids (a significant crop pest) causes a repellent response.^{105,106} A hypothesis was developed whereby the active site of the terpene synthase which biosynthesises a semiochemical may act as a template for the chemical space of the target receptor, therefore, unnatural substrates that are accepted by the synthase may be active to the receptor. Building on previous work,¹⁰⁷ this hypothesis was tested by incubating (*S*)-germacrene D synthase (GDS or Y406F mutant) with a series of FDP analogues of varying functionality. The resulting enzymatic products were isolated, characterised and tested for physiological and behavioural activity in aphids. Of the nine substrate analogues, five were converted into derivatives of germacrene D, four of which gave neuronal and behavioural responses (some are shown in Figure 1.15). Significantly,

14,15-dimethylgermacrene D was shown to be a strong aphid attractant, whilst 14-methylgermacrene D retained strong repellent activity.¹⁰⁸ This discovery has prompted development of germacrene D and its dimethylated analogue into a push-pull crop protection system, whereby aphids can be repelled from crops and attracted elsewhere (such as a trap); a highly successful strategy in other applications.^{109,110}

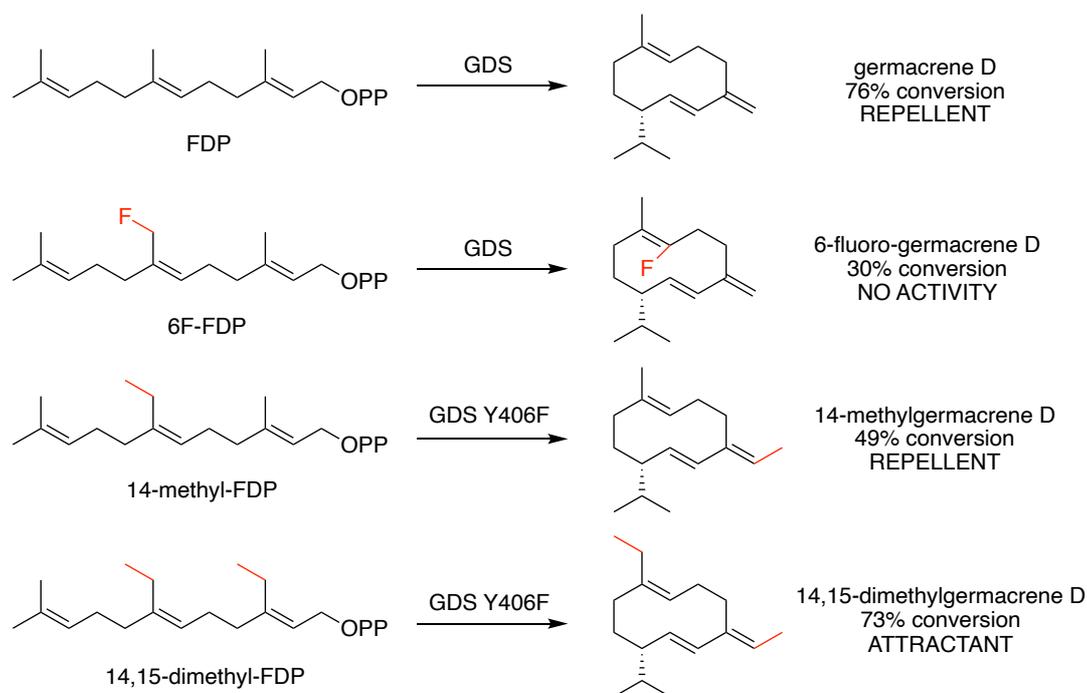


Figure 1.15: Example analogues of FDP which were turned over by GDS or mutant GDS Y406F to give analogues of (S)-germacrene D, some of which were able to influence the physiological and behavioural responses of aphids.

These examples reveal the potential of exploiting the plasticity of terpene synthases with unnatural substrates for the synthesis of valuable chiral intermediates for further synthetic derivatisation and the discovery of novel bioactive compounds. Such an approach may have impact in drug and agrochemical development, particularly if combined with mutagenesis and protein engineering. So far however, these studies have relied on chemical synthesis to create substrate analogues, which can still prove to be lengthy and inefficient and may not be accepted by the enzyme. Access to sufficient amounts of isoprenoid diphosphate precursors, as with terpenoids themselves, is considered a major challenge for this type of research.¹¹¹

1.5 Chemical ecology applications of terpene semiochemicals

The focus of this literature review herein considers the development of terpenes with applications in agricultural pest management.

1.5.1 Food security

Food security is an important global issue: by 2050, the population is expected to reach 9.8 billion¹¹² and studies estimate that current levels of food production will need to increase between 60 - 110% (based on weight¹¹³ and calories¹¹⁴ respectively) to meet the growing demand. As well as food shortages, other significant problems that threaten agriculture is management of pest organisms (including pathogens), to which its estimated that one sixth of agricultural output is lost annually.¹¹⁵ Whilst synthetic insecticides have been extremely effective in dealing with pest insects, their widespread use and indiscriminate toxicity have put selective pressures on pest populations over time, leading to resistance. With parallels to antibiotic resistance, this an increasingly serious issue and relying on the advent of new modes of action is highly problematic.¹¹⁶ The consequences of these problems are particularly severe in developing countries where much of the population are supported by small-holding farms and cereal crops, and crop protection methods and advanced technologies are inaccessible.¹¹⁰ To meet future food demands, whilst mitigating the environmental impact of its production, is therefore an important global challenge. Reducing food waste, improving diets, attaining high yields of existing crops and efficient pest management practices have been identified as critical strategies for a sustainable intensification approach to agriculture.¹¹⁷

1.5.2 Semiochemicals

There has been significant interest in using the natural chemical communication systems of plants and insects to manipulate pest behaviour and reduce our reliance on insecticides.¹¹⁸ Insect and plant ecology is mediated by transient chemical signals (semiochemicals) which elicit behavioural responses related to a variety of functions such as sourcing food, attracting mates, and mediating development. Pheromones are semiochemicals used specifically between organisms of the same species.¹¹⁹ Many of these compounds are terpene secondary metabolites: typically low molecular weight, lipophilic and volatile with high vapour pressures.¹²⁰ Other classifications include phenylpropanoids and derivatives of fatty and amino acids.¹²¹ Relevant to each ecological niche, plants will synthesise their unique subset of these compounds (known as specialised metabolites) which are important for insect ecology, particularly for host selection.⁸ As a result, insects have developed highly specialised

olfactory receptor neurones (ORNs) able to spatially and temporally resolve very low concentrations of a chemical signal (even discerning stereoisomers) from a large amount of background noise. Typically, volatiles of the host plant are detected as compound blends in specific ratios, with an individual ORN typically responding to only one of several compounds.^{122,123}

Due to this specificity, individual pest species can be targeted using low doses of semiochemicals, which are typically non-toxic and therefore considered more sustainable. This however means they are best suited to integration with existing technologies as typically are less efficacious on their own.¹²⁴ applications include population monitoring, mating disruption, natural enemy recruitment and push-pull systems based on companion cropping.^{125,126} Despite their interesting potential, inherent limitations have prevented widespread adoption of semiochemical-based crop protection technologies. Due to their volatility, terpenes can be difficult to handle, whilst they characteristically contain labile double bonds and have conformationally flexible structures which are unstable to air and light (oxidative and thermal degradation).¹²⁷ Therefore the way they are formulated and contained can critically affect their performance in the field.¹⁰⁵

1.5.3 Whitefly pests and wild tomato

Whitefly such as *Bemisa tabaci* (tobacco whitefly, Figure 1.16) and *Trialeurodes vaporariorum* (greenhouse whitefly) are common but particularly virulent agricultural pests, making the top ten of global pest insects.¹²⁸ They are highly polyphagous with a wide variety of host plant species (at least 800) which include many important crops grown in warm and temperate climates and greenhouses, such as cotton, tobacco and pepper. Their direct feeding extracts plant nutrients which inhibits growth, whilst their excreted waste causes moulds to develop adding to the expense of post-harvest processing. Most significantly however, *B. tabaci* vector over 100 plant viruses which can cause total failure of susceptible crops.¹²⁹ Population management has typically relied on broad-spectrum synthetic pesticides to which resistance has rapidly evolved in at least 56 different pesticides.^{128,130} In the United States alone, the pest has reportedly caused economic damage amounting to \$1 billion over the last 30 years.¹³¹

Tomato (*Solanum lycopersicum*) is a highly valuable global crop (second only to potato) which is a favoured host of whitefly and other pests such as spider mites.¹³² Wild tomato species are generally known to have better resistance to pests than domesticated strains, linked to greater genetic diversity and a larger production of defensive compounds such as alkaloids, phenols and terpenes. It is presumed that selective breeding for fruit quality and yield over generations has led to a reduction in the production of such compounds in the commercial varieties.¹³³

Solanum habrochaites is a wild species native to western South America that produce green fruit, which along with the stem and leaves, are covered in small hair-like structures known as trichomes (Figure 1.16).¹³⁴ Trichomes are specialised glands, some of which are known to biosynthesise and excrete an abundance of defensive phytochemicals.¹³⁵

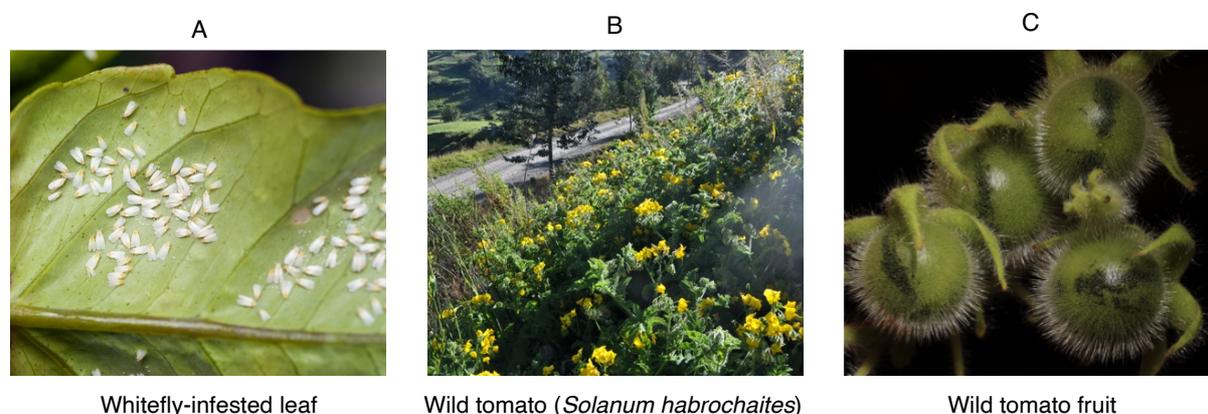


Figure 1.16: (A) *Bemisia tabaci*, or tobacco whitefly occupying the underside of a leaf. (B) Wild Peruvian population of *Solanum habrochaites* with yellow flowers. (C) Green fruit of *S. habrochaites* noticeably covered in hair-like trichomes: specialised tissues associated with the production of defensive compounds.

S. habrochaites has a volatile metabolite profile which is highly abundant in the sesquiterpene 7-*epi*-zingiberene (Figure 1.17). In work by Bleeker *et al.*, 7-*epi*-zingiberene and its oxidation product, (*R*)-curcumene, were shown to significantly repel *B. tabaci* in behavioural bioassays, although other isomers (zingiberene) did not.¹³⁶ In later work, cross breeding *S. habrochaites* with a commercial variety created transgenic strains able to produce high levels of 7-*epi*-zingiberene associated with a toxic environment that led to lower egg oviposition and increased mortality of whitefly and spider mites (*Tetranychus urticae* and *Tetranychus ervasi*). 7-*Epi*-zingiberene also similarly affected the behaviour of insects from other orders, the tobacco hornworm larvae (*Manduca sexta*) and the Colorado potato beetle (*Leptinotarsa decemlineata*).¹³⁷ These findings support the development of 7-*epi*-zingiberene and (*R*)-curcumene as semiochemicals for defence against economically important pests, to reduce reliance on insecticides.

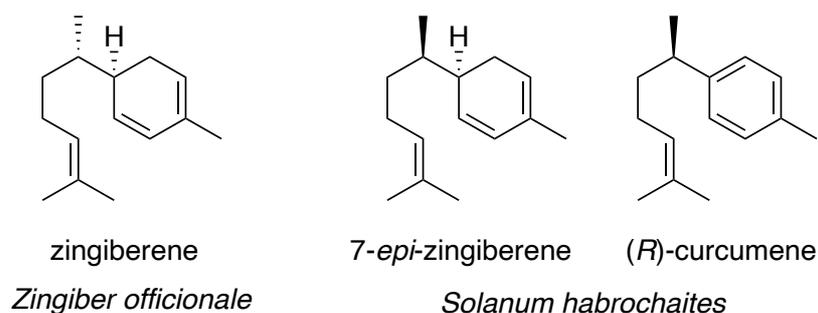


Figure 1.17: Structures of the sesquiterpene metabolite of wild tomato, 7-*epi*-zingiberene, and its oxidised derivative, (*R*)-curcumene. Zingiberene, the diastereomer from ginger (*Zingiber officinale*), is shown for comparison.

1.5.4 Atypical *cisoid* terpene biosynthesis in tomatoes

Bleeker's work included identification of the sesquiterpene synthase that biosynthesises 7-*epi*-zingiberene, 7-*epi*-zingiberene synthase (EZS). Atypically for sesquiterpene synthases, EZS was shown to be of plastidial rather than cytosolic origin: besides the N-terminal chloroplast locating peptide sequence, evidence for this was provided by culturing the plants in the presence of fosmidomycin, an antibiotic which blocks the plastidial MEP biosynthetic pathway. A drop in the production of 7-*epi*-zingiberene (and plastid derived monoterpenes) resulted, indicating that precursors IDP and DMADP were of MEP rather than MEV origin, as is the case for the majority of known sesquiterpenes.¹³⁷ The first plastid localised sesquiterpene synthase was identified as santalene and bergamotene synthase (SBS), also from *S. habrochaites*, and strikingly it shares 91% sequence identity to EZS, although a distinctly different product profile. SBS produces a mixture of five different multicyclic compounds (Figure 1.18),³⁴ whilst EZS produces only 7-*epi*-zingiberene. Santalene and bergamotene are the hydrocarbon precursors of α -santalenoic and α - and β -bergamotenoic acids, previously shown to have strong insecticidal properties.^{138,139} EZS and SBS are also highly unusual because their substrate is exclusively (*Z,Z*)-FDP (Figure 1.4), as opposed to (*E,E*)-FDP from which the majority of known sesquiterpenes are naturally derived.⁵² (*Z,Z*)-FDP is biosynthesised from DMADP and IDP by a short chain *cis*-prenyltransferase known as Z-farnesyl diphosphate synthase (zFPS, Figure 1.5), discovered alongside SBS.³⁴ EZS and SBS have three structural domains ($\alpha\beta\gamma$) which is atypical for sesquiterpene synthases; although the class II active site is non-functional, this indicates diterpene synthase evolutionary origins. It therefore belongs to the TPS-e/f subfamily classification (one of seven clades) which includes diphosphate synthases involved in plant hormone biosynthesis.¹⁴⁰

The full functional characterisation of the terpene synthases of *S. lycopersicum* was recently published, representing the first total analysis of terpene synthases for any plant species.¹⁴⁴ This analysis revealed two more sesquiterpene synthases that exclusively used (Z,Z)-FDP, TPS10 and TPS36. Following characterisation by GC-MS, TPS10 is believed to produce α -bisabolol and TPS36 makes *cis*-muurola-3,5-diene (Figure 1.20). In contrast to *S. habrochaites*, where EZS and SBS are of the TPS-e/f clade and localised to the plastid, these synthases are classified in the TPS-a clade, more commonly associated with sesquiterpene synthases, with TPS36 localised to the mitochondria and TPS10 in the cytosol. Other TPS-a synthases were shown to have activity with (Z,Z)-FDP, producing mixtures of sesquiterpenes, some of which were identified as bisabolene derivatives.¹⁴⁴ It is unclear whether this would be likely to occur *in vivo* however, or whether activity instead arises because of conformational similarity of the substrate to the *cisoid* nerolidyl diphosphate intermediate, through which the reaction mechanisms of many sesquiterpene synthases takes place (Section 1.2.2): from nerolidyl diphosphate, a 1,6-cyclisation leads to the bisabolyl cation.

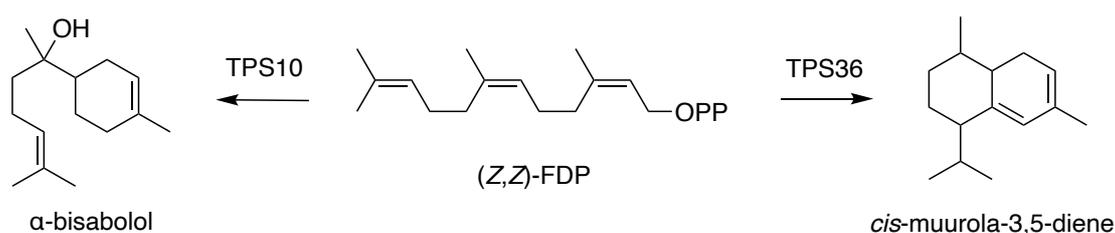


Figure 1.20: Newly characterised sesquiterpenes of *S. lycopersicum* which arise exclusively from (Z,Z)-FDP. Absolute stereochemistry was not defined in the literature.¹⁴⁴

Completing the unique family, a short chain *cis*-prenyltransferase producing C₂₀ NNDP has also been identified in *S. lycopersicum*, known as CPT2.¹⁴⁵ A diterpene terpene synthase (TPS21) and cytochrome P450-oxidoreductase (CYP71BN1) were also found closely associated with the CPT2 gene. TPS21 was shown to catalyse the cyclisation of NNDP into a structure with a tricyclene core which was structurally characterised and named lycosantalene.³⁵ Demonstrated by heterologous expression in *E. coli*, lycosantalene is further oxidised by CYP71BN1 to lycosantalanol through an epoxide intermediate (Figure 1.21). This compound has yet to be identified or isolated from tomato tissues so is suspected to be an intermediate in the biosynthesis of an unknown metabolite. The equivalent genes in wild tomato relatives include deletions which have likely rendered them inactive, a process often

seen in the evolution of specialised metabolite biosynthesis as an organism adapts to environmental stress.¹⁴⁶

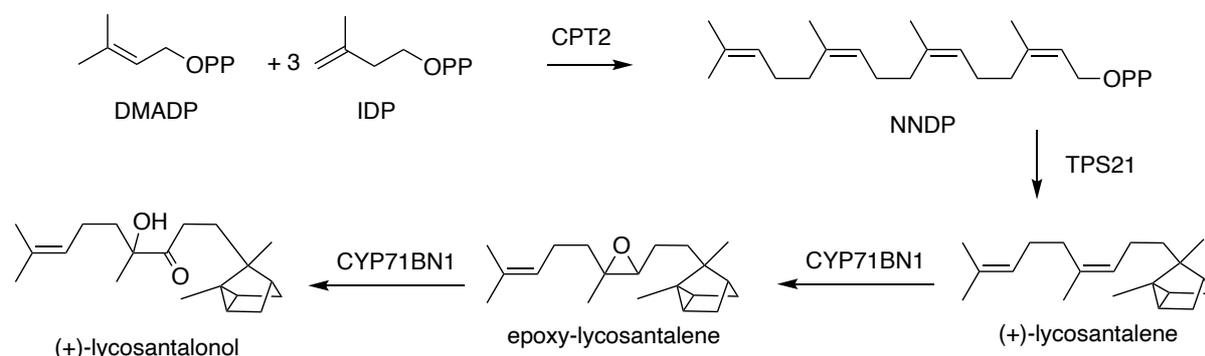


Figure 1.21: *S. lycopersicum* biosynthesis of (+)-lycosantalol, the only known diterpenoid derived from all *cis* NNDP, as opposed to the all *trans* GGDP. Absolute stereochemistry was not defined in the literature.¹⁴⁷

The economic importance of tomato and the occurrence of wild relatives with the ability to synthesise diverse glandular trichome-derived metabolites has generated much interest in utilising its germplasm to provide crops with resistance to insects and plant pathogens.¹⁴¹ Over 100 entire tomato genomic sequences are now known; a key resource for studying these useful phenotypic adaptations which has provided the opportunity to study in detail the occurrence and evolutionary origins of this unusual terpene biosynthesis.

1.6 Project aims

1.6.1 Context

There is desire to develop 7-*epi*-zingiberene and (*R*)-curcumene as natural pest repellents for sustainable control of agricultural pests such as whitefly and spider mites. Terpenoids are typically obtained by extraction from the natural source, but due to low yields, this would be prohibitively expensive on a large-scale. Additionally, it can be difficult to produce sufficiently pure material for semiochemical applications.⁸⁷ Therefore, a more efficient and cost-effective production route is required.

Within the scientific literature, the only devised synthetic route to 7-*epi*-zingiberene is reported by Nicolaou *et al.* as part of the total synthesis of biyouyanagin, an anti-viral natural product. (*R*)-Citronellal and methyl vinyl ketone were reacted in the presence of an asymmetric

organocatalyst to install the correct stereochemistry in an enone precursor, followed by a trifylation and Grignard reaction to obtain the final compound with a 61% overall yield¹⁴⁷ (Figure 1.22). Whilst relatively straightforward and stereospecific, the initial asymmetric reaction is controlled with 5 mol% of an expensive organocatalyst (£285 for 500 mg, *SigmaAldrich*). This route may therefore be unsuitable for large scale manufacturing without process development. There are also several publications of the asymmetric synthesis of (*R*)-curcumene which possesses the more common tertiary benzylic chiral centre, however these similarly adopt organocatalysts¹⁴⁸ or transition metal catalysts¹⁴⁹ for stereo control that may be incompatible with large-scale production.

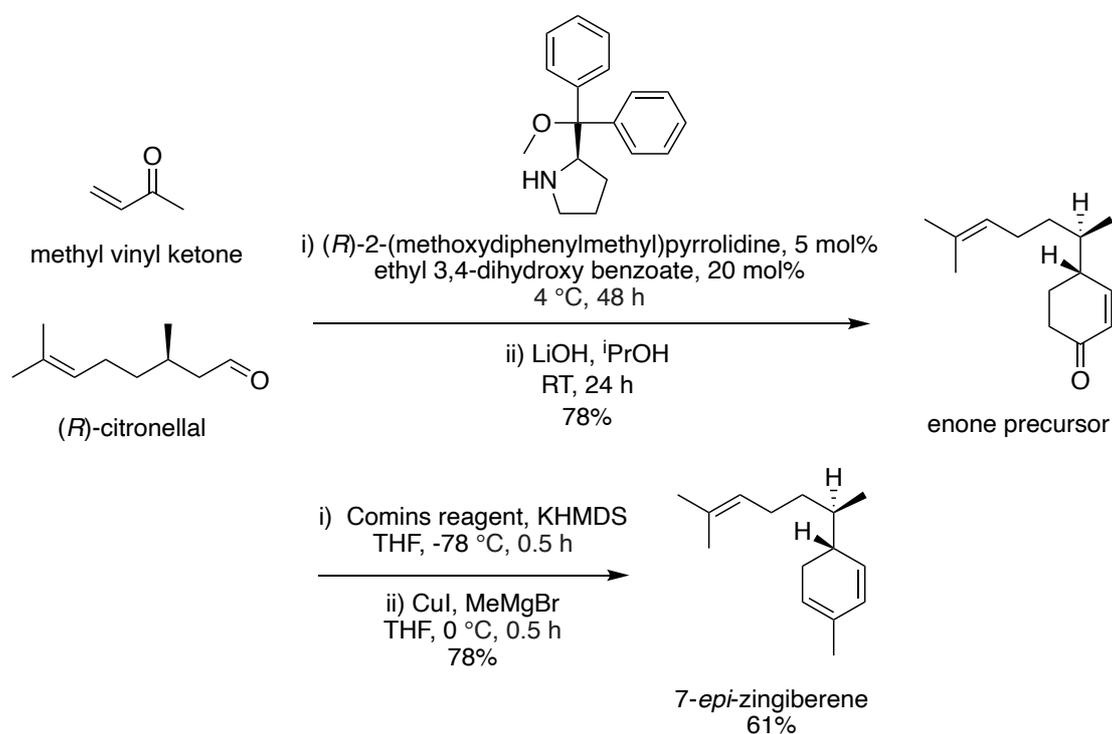


Figure 1.22: Synthetic route to 7-*epi*-zingiberene from chiral precursor, (*R*)-citronellal, using the asymmetric organocatalyst, (*R*)-2-(methoxydiphenylmethyl)pyrrolidine.

1.6.2 Aims

This project has been divided into various parts but maintains an overall focus of developing 7-*epi*-zingiberene into a sustainable agrochemical. The initial aim was to develop an alternative *in vitro* synthesis of 7-*epi*-zingiberene through enzymatic conversion of (Z,Z)-FDP by EZS to give pure product of the desired stereochemistry in a single step. Although Jones *et al.* of the Allemann group had previously developed a synthesis of (Z,Z)-FDP (Section 2.1), it is lengthy and low-yielding. A key objective therefore involved developing a chemoenzymatic route for (Z,Z)-FDP by using zFPS and the simple precursors IDP and DMADP. The heterologous expression of EZS was also optimised to enable its application in biocatalytic compound production.

Another aim of this project was to further understand EZS catalysis by investigating aspects of its structure-function relationship. The catalytic mechanism with respect to hydride shifts was examined by using deuterated isotopologues of (Z,Z)-FDP, also synthesised via a chemoenzymatic approach. Mutagenesis was used to identify which amino acid residues may be important for facilitating the catalysis and maintaining the high-fidelity outcome. Part of this work additionally aimed to investigate how and why EZS is selective for its unusual *cisoid* substrate over the ubiquitous precursor (E,E)-FDP.

The active site of EZS, like other terpene synthases, behaves like a stereochemical template for making its repellent semiochemical product. It was therefore hypothesised that analogues of 7-*epi*-zingiberene produced via EZS may also prove useful agrochemicals. A new chemoenzymatic method was developed with the aim of generating novel semiochemical analogues which could be used for insect bioactivity testing. The final part of this project investigated chemical ecology aspects of sesquiterpenes for their future application in the management of pest aphids.

Chemoenzymatic Synthesis of (Z,Z)-FDP and 7-*Epi*-zingiberene

2.1 Introduction

There is significant interest in using 7-*epi*-zingiberene as a natural pest repellent to aid the control of problematic agricultural pests such as whitefly and spider mites.¹³⁶ To support further research and commercialisation, a cost-effective and reliable way of producing the compound to high levels of purity is required. EZS is the sesquiterpene synthase from wild tomato that produces 7-*epi*-zingiberene with high fidelity from the *cisoid* precursor (Z,Z)-FDP.¹³⁷ Enzymatic production of pure 7-*epi*-zingiberene by EZS offers a highly selective and potentially efficient synthetic route, an equivalent to which would be excessively difficult to achieve using typical organic chemistry methods which may rely on expensive asymmetric catalysts.¹⁴⁷

To exploit the activity of EZS for the biosynthesis of 7-*epi*-zingiberene (Z,Z)-FDP is required in high yields. Unfortunately, (Z,Z)-FDP is not commercially available unlike the universal sesquiterpene precursor and geometric isomer (E,E)-FDP which is but at significant cost. Dr. Chris Jones of the Allemann group developed a synthetic route to (Z,Z)-FDP via the alcohol precursor (2Z,6Z)-farnesol ((Z,Z)-farnesol) based on previous literature (Figure 2.1).^{150–152} It begins with nerol (C₁₀ alcohol with required Z-configured double bond) which is converted in three steps to nerol acetone. Nerol acetone is then used in a Horner-Wadsworth-Emmons (HWE) chain extension reaction requiring a Still-Gennari modified HWE reagent to improve the formation of the Z double bond relative to the more favourable E isomer. This reaction gives ethyl farnesoate in a 1:1 E,Z:Z,Z mixture, compared to 3:1 ratio when using an unmodified HWE reaction. The modified HWE reagent itself is made via a highly air sensitive separate three-step synthesis starting from triethyl phosphonoacetate. The farnesoate esters are then reduced to alcohols using diisobutylaluminium hydride (DIBAL) and the isomers separated to give the desired (Z,Z)-farnesol. This is chemically diphosphorylated (in three further three steps) to give the diphosphate as a white powder. The maximum yield over the 11 steps achieved was 9.5%, but was typically much lower.¹⁵³

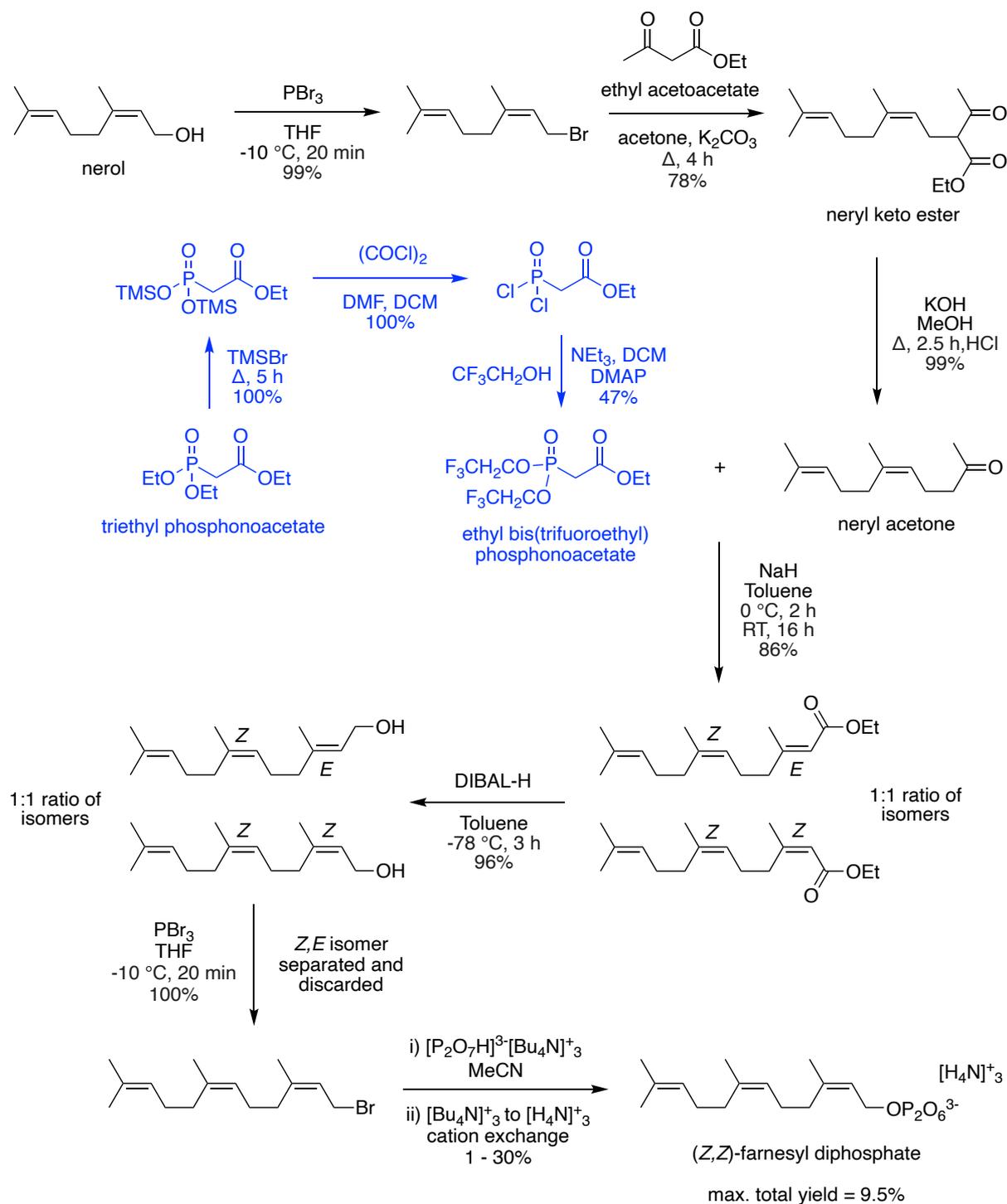


Figure 2.1: The 11 step synthesis of (Z,Z)-FDP starting from commercially available nerol with Z geometry. Formation of the desired Z isomers is disfavoured and the final chemical phosphorylation steps are often problematic resulting in low overall yields.¹⁵³

Despite optimisation, the synthesis is labour intensive and low yielding, particularly due to the phosphorylation reaction. This problem is ubiquitous in isoprenoid chemistry (including for the considerably simpler (*E,E*)-FDP synthesis), as despite there being several established procedures for alcohol diphosphorylation,^{154–156} they are known to suffer from low selectivities, poor yields and difficulties in upscaling.¹⁵⁷ Additionally, in the case of (*Z,Z*)-farnesol synthesis, the desired product is thermodynamically and/or sterically disfavoured in the double bond generation reactions compared to its geometric isomer. This route is also further complicated by modifications when producing analogues of research interest (e.g., isotopologues). As a result, this chemical synthesis is inefficient for larger scale synthesis of (*Z,Z*)-FDP or 7-*epi*-zingiberene.

2.1.1 Aims of this work

This work aimed to develop an alternative, completely selective, and scalable route to (*Z,Z*)-FDP (and consequently 7-*epi*-zingiberene) by exploiting the *cis*-prenyltransferase known as zFPS which naturally catalyses its formation from the universal C₅ terpenoid precursors, DMADP and isopentenyl diphosphate IDP.^{34,42} Both IDP and DMADP can be synthesised from their simple, readily available C₅ alcohol precursors. Initially a chemical diphosphorylation was used to generate these substrates for zFPS catalysed reactions, but a new enzymatic synthesis developed in the Allemann group was later adopted which uses two kinases, hydroxyethylthiazole kinase (THIM) and isopentenyl phosphate kinase (IPK) (Section 2.3). Expression of EZS has historically been very poor therefore this work also aimed to improve its expression to practically useful levels and enable its application in biocatalytic compound production. Exploiting the selectivities of both terpenoid biosynthetic enzymes was envisaged to efficiently lead to enantiomerically pure compound under ambient reaction conditions, saving time and resource compared to the existing organic syntheses of these difficult to make compounds (Figure 2.2).

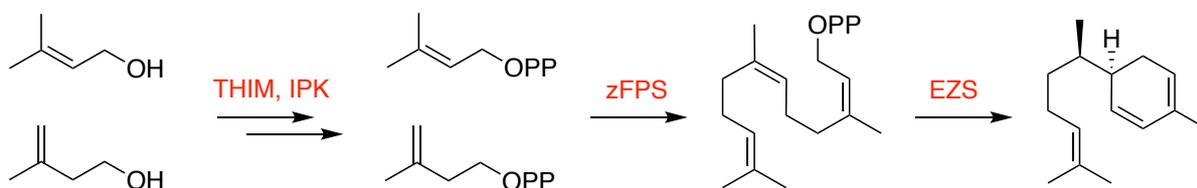


Figure 2.2: Schematic showing an *in vitro* synthesis of (*Z,Z*)-FDP and 7-*epi*-zingiberene from simple C₅ alcohol precursors this work aims to develop.

2.2 Chemoenzymatic (Z,Z)-FDP synthesis via synthetic IDP and DMADP

Development of an *in vitro* route to (Z,Z)-FDP first requires a source of its diphosphate precursors, IDP and DMADP. As mentioned in Section 2.1, there are several established methods in the literature for diphosphorylating alcohols for terpene applications. The prevailing method of Davisson *et al.*,¹⁵⁵ established in 1986, works by introducing the diphosphate moiety to a halogenated form of the allylic precursor via a displacement reaction involving tris(tetrabutylammonium) hydrogendiphosphate ((NBu₄)₃P₂O₇H). This is prepared separately by ion exchange of commercially available sodium dihydrogen diphosphate. After the displacement reaction, the isoprenoid diphosphate undergoes another ion exchange procedure to gain an ammonium counter ion, followed by lyophilisation to yield the product as a white solid. This method is technically challenging as it is sensitive to air and moisture and favours longer alkyl chain starting materials. An alternative method developed by Keller¹⁵⁸ uses triethylammonium phosphate (TEAP) to directly react with the alcohol group, generating a mixture of mono-, di- and triphosphorylated organic and inorganic species from which the diphosphate was isolated. Both methods were tried but the latter found to be more successful, with gram quantities of IDP and DMADP produced (Section 8.3.2). These substrates were then used for *in vitro* condensation reactions in an attempt to produce (Z,Z)-FDP using zFPS prepared according to Section 2.4.1.

A range of different reaction conditions in 100 mM tris(hydroxymethyl)aminomethane (TRIS) pH 8.0, 5 mM MgCl₂ buffer were screened: 1 - 20 mM IDP and DMADP and 0.1 - 2.5 μM zFPS with total volumes between 25 - 100 mL. However, despite many attempts, overall yields were poor at typically 2 - 5%. One reason for this could lie with the very small amounts of mono- and triphosphorylated species (identifiable via ³¹P NMR spectroscopy) which contaminate the diphosphate at levels around 1%. The contaminants are likely to include isopentenyl monophosphate (IP) and dimethylallyl monophosphate (DMAP) which are known to play a role in the regulation of the *trans*-prenyltransferase, FPPS, which makes (E,E)-FDP. These substrates were shown to competitively inhibit FPPS with an IC₅₀ of around 100 μM.¹⁵⁹ Assuming that zFPS would be similarly regulated as its role is comparable to FPPS, if the TEAP prepared diphosphate was contaminated by as little as 0.5% monophosphate, at millimolar reaction concentrations of DMADP and IDP, their respective monophosphate concentrations could exceed the IC₅₀. The presence of such species in the zFPS reaction may therefore have an inhibitory effect on catalysis leading to low yields. Due to the amphiphilic nature of isoprenoid phosphates, methods to dissolve and purify them to separate the mono-, di- and triphosphorylated species are very limited, therefore it may be prohibitively difficult to obtain sufficiently pure diphosphate for prenyltransferase catalysed reactions from this synthetic route.

2.3 Chemoenzymatic DMADP and IDP synthesis

Alongside this project, a new and highly efficient *in vitro* synthesis of DMADP and IDP was under development by Dr. Luke Johnson of the Alleman group with the aim of providing a more cost effective route to (*E,E*)-FDP. Decoupled from central carbon metabolism and either biosynthetic pathway, it takes advantage of two promiscuous kinases to produce DMADP and IDP directly from their cheap and readily available equivalent alcohols, prenil and isoprenol. THIM from *E. coli* has the natural substrate 4-methyl-5-(2-hydroxyethyl)thiazole but was previously reported to phosphorylate prenil,¹⁶⁰ whilst the mevalonate pathway of the archaeon *Methanocaldococcus jannaschii* contains IPK which naturally converts isopentenyl IP to IDP.²¹ When tested with isoprenol and prenil, THIM was found to catalyse their conversion to monophosphates, in turn becoming substrates for IPK which carried out a second phosphorylation to produce the diphosphates IDP and DMADP.¹⁶¹ ATP is the only cofactor required for these two enzymatic reactions and can be used at catalytic quantities when accompanied by an ATP recycling system. In comparison, the MEP and MEV pathways require four cofactor equivalents over seven and six steps respectively in the conversion of their metabolic precursors to IDP and DMADP (Figure 2.3).

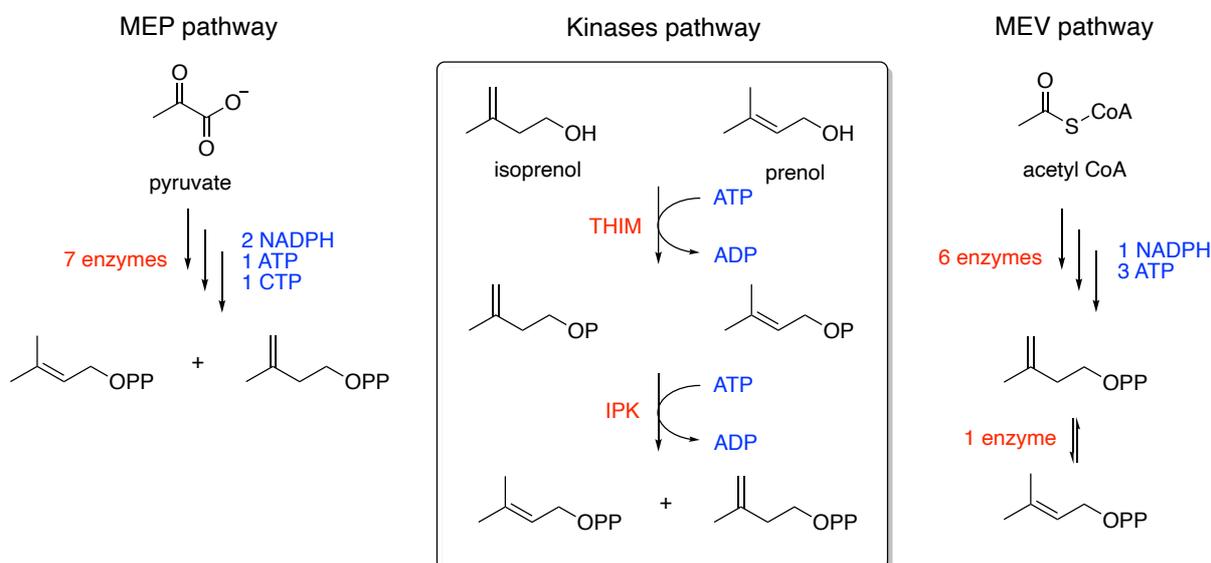


Figure 2.3: The substrates fundamental to all terpene compounds originate from one of two biosynthetic pathways: the MEP and the MEV. A new *in vitro* kinases pathway to the substrates requiring ATP as the only cofactor has recently been devised in the Allemann group using two kinases (THIM and IPK) and readily available alcohol precursors.¹⁶¹

It was therefore envisaged that this *in vitro* pathway could be developed for (Z,Z)-FDP and 7-*epi*-zingiberene production directly from the cheap precursors prenil and isoprenol by including zFPS and EZS: these results were contributed to the publication in *Angewandte Chemie*.¹⁶¹

2.3.1 Expression and purification of IPK and THIM

Kinases THIM and IPK were available in the plasmid library of the group, having been cloned into pET28 expression vectors by Dr. Luke Johnson. Both were overexpressed in *E. coli* BL21-AI cells under tight control of the arabinose inducible promoter (Section 8.1.10). This promoter prevents 'leaky' background expression and allows high levels of heterologous protein to be accumulated.¹⁶² After 18 hours expression at 18 °C, the cells were harvested by centrifugation, lysed in buffer and the soluble fraction purified. As each protein was cloned to contain a N-terminal polyhistidine (His-) tag, Ni²⁺ affinity chromatography was used for purification which was sufficient for obtaining purities of at least 95%, as judged by sodium dodecylsulfate polyacrylamide gel electrophoresis (SDS-PAGE) analysis (Figure 2.4). IPK was found to precipitate readily if cooled so the purification was carried out at room temperature. To remove imidazole, fractions containing the desired protein were desalted into the buffer used for the later assays (100 mM TRIS pH 8.0) using the general method (Section 8.1.14), before concentration using spin ultrafiltration. Final protein concentrations were estimated using absorbance at 280 nm (A_{280}) with the extinction coefficients of 24135 and 30370 M⁻¹ cm⁻¹ for THIM and IPK respectively, estimated using the Benchling webtool under reducing conditions.¹⁶³ Both THIM and IPK were stored at -80 °C until further use with glycerol omitted from the buffer as it was shown to be a substrate for phosphorylation.

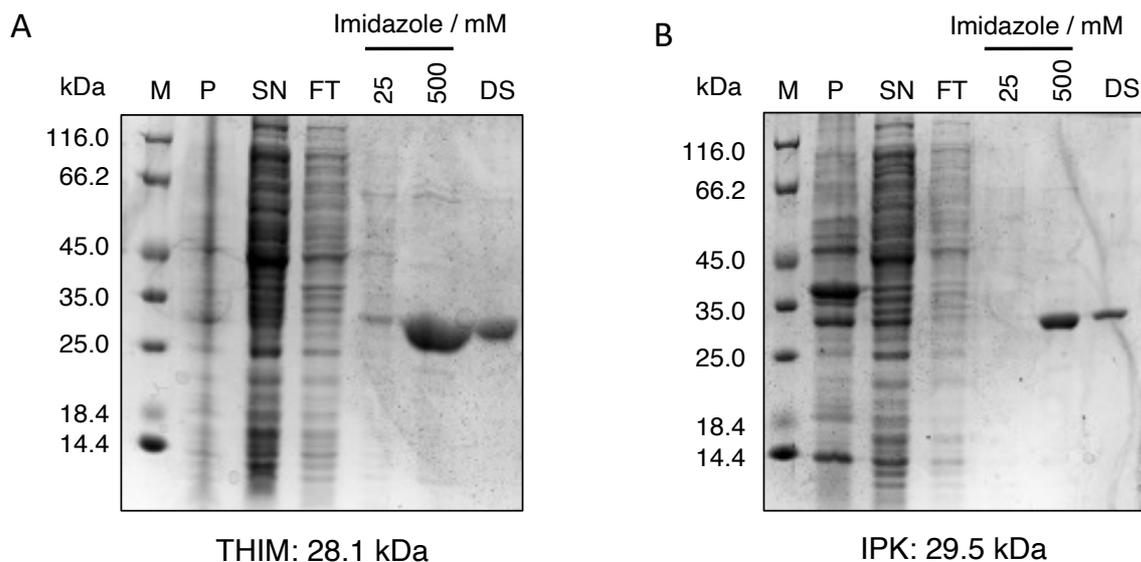


Figure 2.4: SDS polyacrylamide gels showing the purification of THIM (A) and IPK (B) via Ni^{2+} affinity chromatography using an imidazole gradient. Sample abbreviations: M = molecular weight marker, P = pellet, SN = supernatant, FT = flow through and DS = desalt.

2.3.2 Enzymatic diphosphorylation of prenil and isoprenol

Once purified, THIM and IPK were used to diphosphorylate prenil and isoprenol to DMADP and IDP respectively (Figure 2.5). As described in Section 2.1.1, the first phosphorylation is undertaken by THIM followed by the second phosphorylation by IPK. Diphosphorylation reactions were performed by immobilising THIM and IPK together on Ni-NTA resin, enabling their reuse after reaction completion. To achieve this, THIM (10 mg, 65 μM final) and IPK (300 μg , 5 μM final) were incubated with nickel coupled nitrilotriacetic acid agarose resin (Ni-NTA, 1 mL) in a small drip column for at least 30 min before removal of excess buffer. Once both enzymes were immobilised the diphosphorylation reactions were performed using 2 mg/mL substrate (23.6 μL prenil, 23.5 μL isoprenol) in 10 mL reaction buffer (100 mM TRIS pH 8.0, 5 mM ATP, 200 mM KCl, 5 mM MgCl_2). Separate resins were used for prenil and isoprenol due to the difference in substrate preference and catalytic rate of THIM. An *in situ* ATP regeneration system was employed to reduce the ATP concentration required and limit unproductive hydrolysis by the kinases. For this the reaction additionally contained phosphoenolpyruvate (PEP, 75 mM) which provides the source of phosphate, and pyruvate kinase (PK, 10 units/mL) to catalyse the regeneration of ATP from adenosine diphosphate

(ADP, Figure 2.5). This enabled catalytic quantities of ATP to be used for the reactions which is prohibitively expensive to use in equimolar amounts.

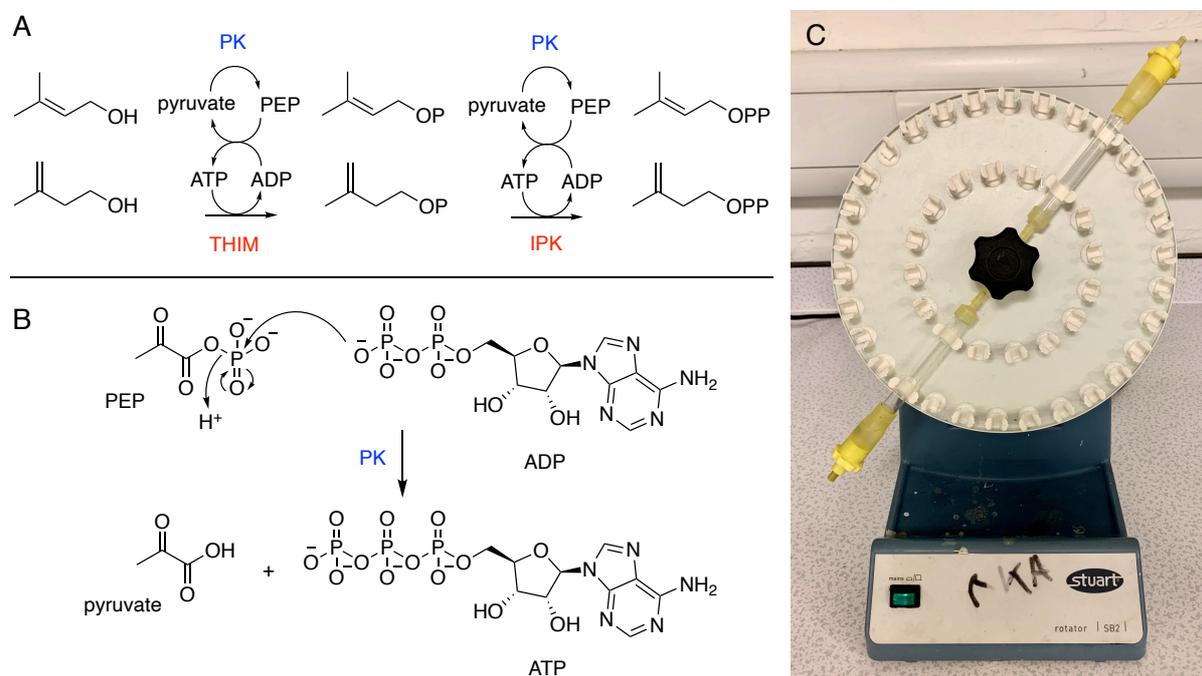


Figure 2.5: (A) The reaction scheme for the sequential phosphorylations of isoprenol and prenol to IDP and DMADP by the ATP-dependent THIM and IPK. (B) The mechanism of ATP recycling via PK and PEP. (C) Reactions set up using drip columns incubated with spinning on a disk rotator.

The reactions were incubated on a spinning disk rotator to gently mix at room temperature and were monitored using ^1H and ^{31}P (^1H decoupled, $\{^1\text{H}\}$) NMR spectroscopy (0.5 mL reaction mix combined with 50 μL D_2O to make the sample). Using the ^{31}P NMR spectra (Figure 2.6), monophosphate formation was observed by the appearance of a single singlet peak at 3.6 ppm. Formation of the diphosphate was indicated by the presence of two doublet peaks around -6.0 and -10.0 ppm, with splitting ($^2J_{\text{PP}} = 21$ Hz) caused by the coupling of the two ^{31}P nuclei to one another. Other signals that were observed include a large singlet around -0.6 ppm from PEP, a small singlet at 2.5 ppm from inorganic phosphate and a set of doublet peaks at -10.7 to -5.6 ppm from ATP (verified by measuring spectra of control samples).

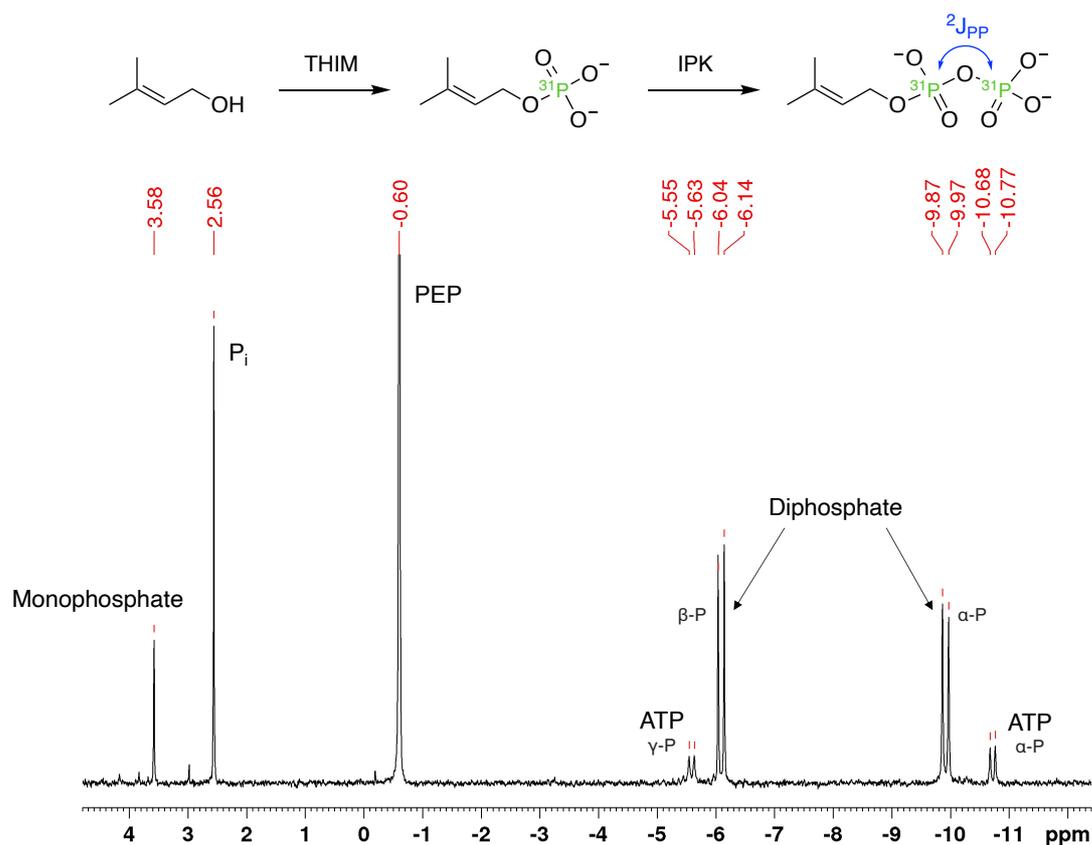


Figure 2.6: Typical ^{31}P - $\{^1\text{H}\}$ NMR spectrum (162 MHz, D_2O) showing the course of the diphosphorylation of prenol by kinases THIM and IPK. A peak relating to dimethylallyl monophosphate (DMAP) produced by THIM can be seen at 3.58 ppm. Two doublet peaks relating to DMADP produced by IPK are seen around -6.1 and -9.9 ppm. Other ^{31}P environments present include phosphate (P_i), PEP and ATP.

The diphosphorylation reactions typically took 18 hours to complete; indicated by the absence of peaks relating to the monophosphorylated product and the presence only of diphosphate peaks. The reactions could also be followed by ^1H NMR spectroscopy with solvent (H_2O) suppression by comparing the multiplicity of the CH_2OH doublet peak splitting to a $\text{CH}_2\text{OP}_2\text{O}_6^{2-}$ triplet accompanied by a downfield shift. Having completed the reaction, the solution containing the diphosphate (at a final concentration of 23.2 mM) was eluted from the resin and stored until required at -20°C . As the kinases remained bound to the resin, the reaction mixes containing the alcohol substrate and ATP regeneration components were regenerated and the diphosphorylation reaction repeated until the kinases lost activity (up to 5 days, followed by NMR spectroscopy). This maximised the amount of diphosphate produced from a single protein purification (particularly of lower activity THIM) and enabled stocks to be accumulated efficiently.

2.4 Chemoenzymatic (Z,Z)-FDP via enzyme-prepared IDP and DMADP

With an efficient biosynthesis of DMADP and IDP established it was next explored whether this could be combined with zFPS to selectively produce (Z,Z)-FDP, with comparison to the previous attempts using chemically prepared substrates.

2.4.1 Expression and purification of zFPS

The zFPS gene was received as a gift from the group of Prof. Andrew H. J. Wang (Academia Sinica, Taiwan) cloned into a pET32-Xa/LIC expression vector which includes an N-terminal thioredoxin tag (additional 11.8 kDa) with a Factor Xa cleavage site.⁴² The full-length enzyme (29.6 kDa) contains a chloroplast targeting sequence predicted to be 45 amino acids in length which was truncated to produce the clone for the expression plasmid, herein named pET32-zFPS. As described in Section 8.1.10, pET32-zFPS was transformed into competent BL21(DE3) pLysS cells and expressed under the control of the T7 promoter then purified by Ni²⁺ affinity chromatography. Fractions containing protein deemed at least 95% pure by SDS-PAGE (Figure 2.7) were desalted into the buffer used for the later assays (100 mM TRIS pH 8.0) with additional 10% glycerol. The protein was concentrated via spin ultrafiltration and the final concentration measured by Bradford assay¹⁶⁴ before storage at -80 °C until required.

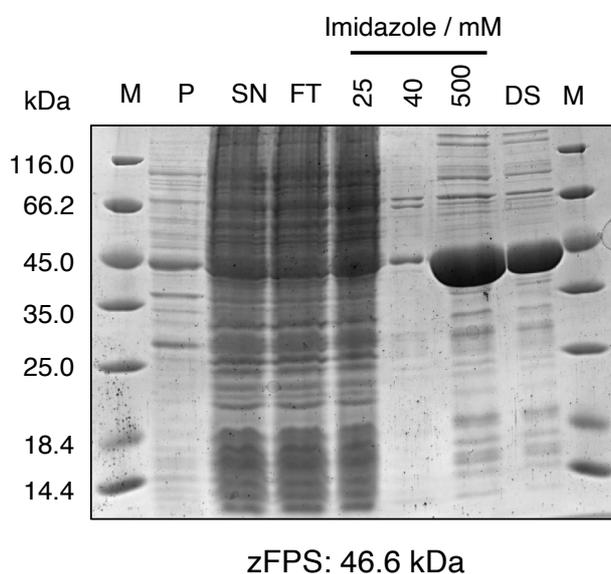


Figure 2.7: SDS-PAGE gel showing the purification of zFPS via Ni²⁺ affinity chromatography using an imidazole gradient. Sample abbreviations: M = molecular weight marker, P = pellet, SN = supernatant, FT = flow through and DS = desalt.

2.4.2 Enzymatic production of (Z,Z)-FDP and characterisation

Purified zFPS was then used to catalyse the condensation reaction between DMADP and IDP to make (Z,Z)-FDP. The reaction was performed in 25 mL total volume using two equivalents of IDP (5 mL stock, 4.6 mM) for each equivalent of DMADP (2.5 mL stock, 2.3 mM) with zFPS (1.25 μ M) in 100 mM TRIS pH 8.0, 5 mM DTT, 5 mM MgCl₂ buffer. The reaction was incubated with shaking (175 rpm) at 37 °C for 18 hours (Figure 2.8).

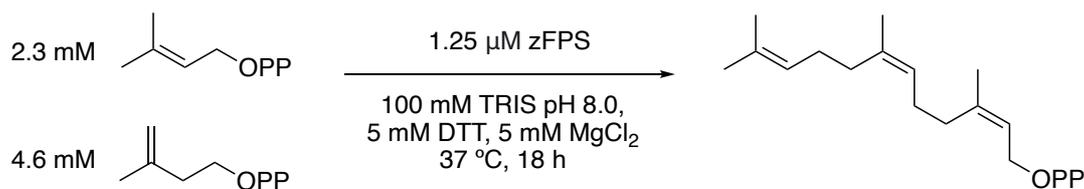


Figure 2.8: Conditions used for the enzymatic preparation of (Z,Z)-FDP from IDP and DMADP produced using the *in vitro* kinases method.

The presence of magnesium in the buffer not only facilitated zFPS catalysis but caused the FDP to precipitate from the reaction mixture at concentrations exceeding approximately 2 mM, enabling it to be easily isolated by centrifugation. Following several washes using deionised H₂O (10 mL) and repeated centrifugation (4000 rpm, 20 min), the precipitate was resolubilised in a solution of sodium diphosphate (250 mM, ~30 mL). Here the excess inorganic phosphate coordinates to the Mg²⁺ allowing the FDP to dissolve. The solution was loaded onto a C₁₈ column and purified using reverse phase flash chromatography with a water:acetonitrile gradient (eluting around 50% acetonitrile). Fractions corresponding to the product were lyophilised to give pure (Z,Z)-FDP (24 mg, 92% yield). A ¹H NMR spectrum is given in Figure 2.9 and full characterisation given in Section 8.2.5 which is consistent with the data recorded from (Z,Z)-FDP produced synthetically.¹⁵³

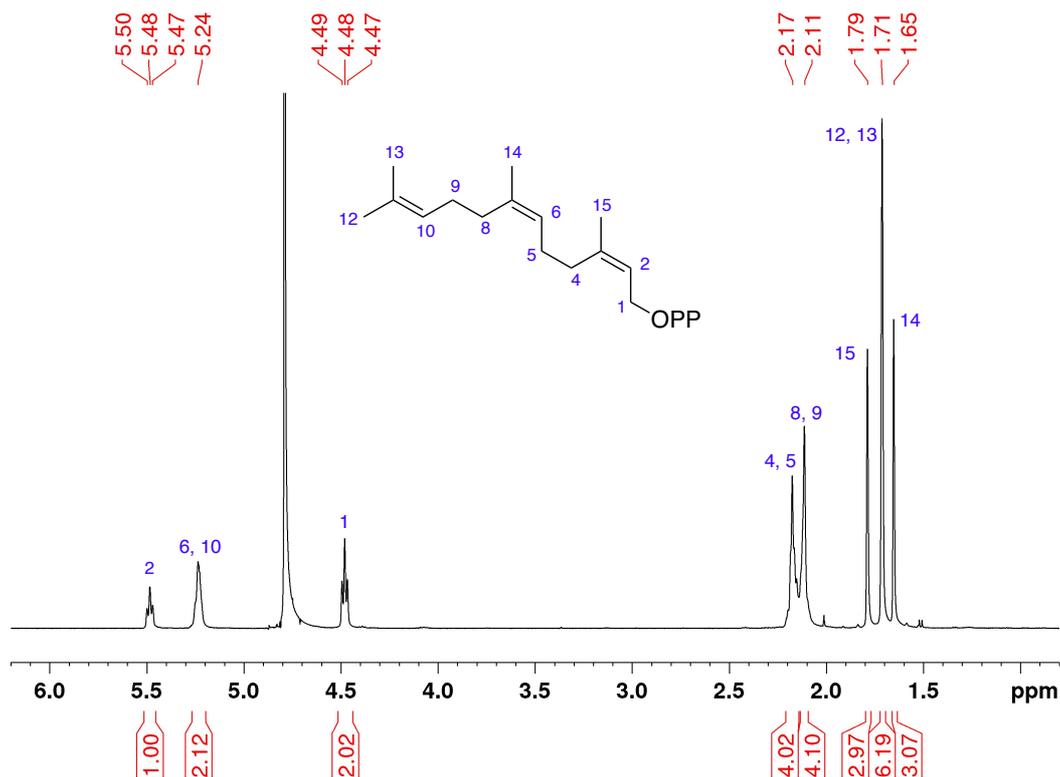


Figure 2.9: Assigned ¹H NMR (500 MHz, D₂O) spectrum of (Z,Z)-FDP produced *in vitro* using its biosynthetic enzyme zFPS and C₅ substrates IDP and DMADP produced enzymatically from isoprenol and prenilol respectively.

This represents the first synthetic route for (Z,Z)-FDP to achieve near quantitative yields and has enabled full structural characterisation of the direct enzymatic product. In the previous literature where zFPS was first characterised, alkaline phosphatase (AP) was used to dephosphorylate the product of zFPS which was confirmed to be (Z,Z)-farnesol following NMR spectroscopic analysis.³⁴ As described in the introduction to this chapter, the previously most optimal route to (Z,Z)-FDP was an 11-step organic synthetic procedure where formation of the desired Z isomer is disfavoured, and typically the majority of the product would be lost at the final phosphorylation step. Using the kinases and prenyltransferase in this way enabled high yields to be achieved (92%) and is completely selective for the desired Z isomer. This new method is also practically simpler and shorter, even when accounting for protein expression and purification, therefore may be more suitable for upscaled compound production.

2.5 EZS expression optimisation

Having established an efficient method for producing (Z,Z)-FDP, the method was further investigated in order to establish whether enantiomerically pure 7-*epi*-zingiberene could be produced by inclusion of EZS. EZS is a large (88 kDa), multi-domain enzyme which has typically suffered from very low expression levels and has therefore been challenging to produce and purify. Previous work in the group had established an expression protocol for EZS. A codon optimised gene for EZS was initially bought from Addgene and cloned into a pET21a vector with T7 promoter, ampicillin resistance gene and a C-terminal His-tag with tobacco etch virus (TEV) endoprotease cleavage site. The effects of different growth temperatures, concentration of IPTG for induction and cell lines were evaluated with expression achieved using C41(DE3) pLysS cells and low temperature overnight expression, however overall yields remained very poor.¹⁵³ Although *E. coli* are highly effective at protein production and have many advantages over other hosts, they can often suffer with low and/or insoluble expression of the target, particularly for eukaryotic proteins.¹⁶⁵ Low levels of protein can arise from a number of different factors: if the problem lies with expression itself, it may be associated with translation initiation, transcriptional regulation, mRNA stability or translation efficiency. The problem may also result from lack of solubility, where poor folding kinetics can lead to the aggregation of partly or mis-folded proteins into inclusion bodies which are subsequently the target of intracellular proteases.¹⁶⁶ Different troubleshooting approaches typically include changing host organisms, cell lines, and cultivation conditions or co-expressing with chaperones and fusion tags. In an attempt to improve expression of EZS in this work, the effect of different fusion protein tags were investigated.

2.5.1 Improving EZS expression using fusion proteins

Fusion tags are stable peptide sequences which are genetically linked to target proteins to enhance yields, folding, solubility and purification. They are used widely in the literature and can prove remarkably successful, although the mechanisms by which they work are not always clear. Proteins which make good fusion partners are often highly conserved, benefiting from strong translational initiation signals, and fold readily thereby acting as nucleation points for subsequent folding of the target, in a similar way to chaperone proteins. Some fusion proteins have been shown to prevent degradation by translocating the target protein to different parts of the cell, away from higher concentrations of proteases found in the cytosol.^{166,167}

A range of fusion proteins differing in origin, function and size were cloned into the pET21-EZS vector at the N-terminus of the gene with inclusion of a TEV protease site before the EZS

peptide sequence site using the Golden Gate assembly technique (Section 8.1.7). The fusion proteins chosen were sourced from the Alleman group plasmid library and included glutathione-S-transferase (GST),¹⁶⁸ small ubiquitin-like modifier (SUMO)¹⁶⁷, thioredoxin (TRX),¹⁶⁸ and dihydrofolate reductase (DHFR) - a small, well-expressing protein used in the group although not known in the literature for this purpose. Maltose binding protein (MBP)¹⁶⁹ and guanidine binding protein (GB1)¹⁷⁰ were also available having previously been cloned into the pET21-EZS vector by Dr. Robert Mart. To enable more efficient cloning, the constructs were designed so each fusion peptide was flanked either side by short linker sequences (five amino acids: SGILV), which remained the same between constructs (Figure 2.10). This meant amplifying a new backbone each time could be avoided as amplifying longer sequences can sometimes prove problematic.

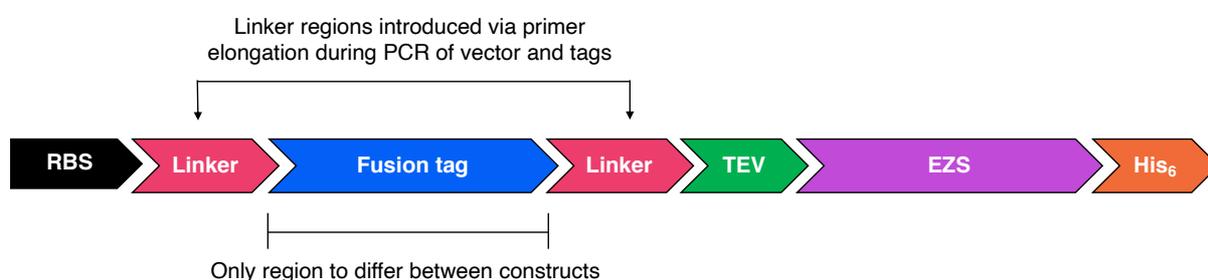


Figure 2.10: Overview of design of EZS plasmids to contain different solubility tags (DHFR, GST, SUMO and TRX) which were assembled using a common vector backbone fragment containing the EZS gene. Regions shown include ribosome binding site (RBS), TEV protease recognition sequence and C-terminal His-tag.

The vector backbone containing the EZS gene was amplified using polymerase chain reaction (PCR) with primers to ensure either end of its linear DNA sequence ended with the linker region, as were the solubility tag fragments which were amplified from existing plasmids of the group library (Section 8.1.2). Products from the PCR reactions were analysed by agarose gel electrophoresis and those corresponding to the desired molecular weight fragments were isolated and purified from the gel. The vector fragment containing EZS was then used for each Golden Gate assembly reaction with the addition of the PCR product relating to each fusion protein. Competent cells were transformed with the assembly mixtures and grown on agar in the presence of ampicillin. From a starter culture of subsequent single colonies, the DNA was extracted and sequenced to confirm the presence and correct insertion of each protein tag. Once successful cloning of the EZS fusions were confirmed, screening of expression conditions was carried out.

The EZS fusions were expressed on a small scale in the previously most optimal C41(DE3) pLysS cell line using a range of conditions and soluble and insoluble protein content was analysed by SDS-PAGE and western blot. Western blot is a highly sensitive technique which uses antibodies to bind specifically to the target protein: the antibodies can then be detected by techniques such as chemiluminescence enabling very small (up to femtomolar) quantities of target protein to be indirectly detected. Expression conditions varied include post-induction growth temperatures of 16, 20, 25 and 37 °C, and media compositions with different ratios of minerals, nutrients and buffering agents based on lysogeny broth (LB) and terrific broth (TB) medias, as these had not yet been screened. Samples of each expression were taken before and after induction; the post-induction samples lysed by repeated freeze-thawing and the soluble and insoluble protein content separated by centrifugation. Samples of each fraction were analysed by SDS-PAGE and later western blot using an antibody probe which specifically binds to His-tagged proteins. In most expression samples however, across the range of fusion tags and different growing conditions used, His-tagged protein relating to EZS was not detected. Only very small amounts of protein could be identified (Figure 2.11) with the expression samples from the pET21-EZS (no fusion tag) and GB1-EZS vectors, but mostly from their insoluble fractions. The analysis had therefore revealed that the fusion proteins had not improved EZS expression under the conditions tried.

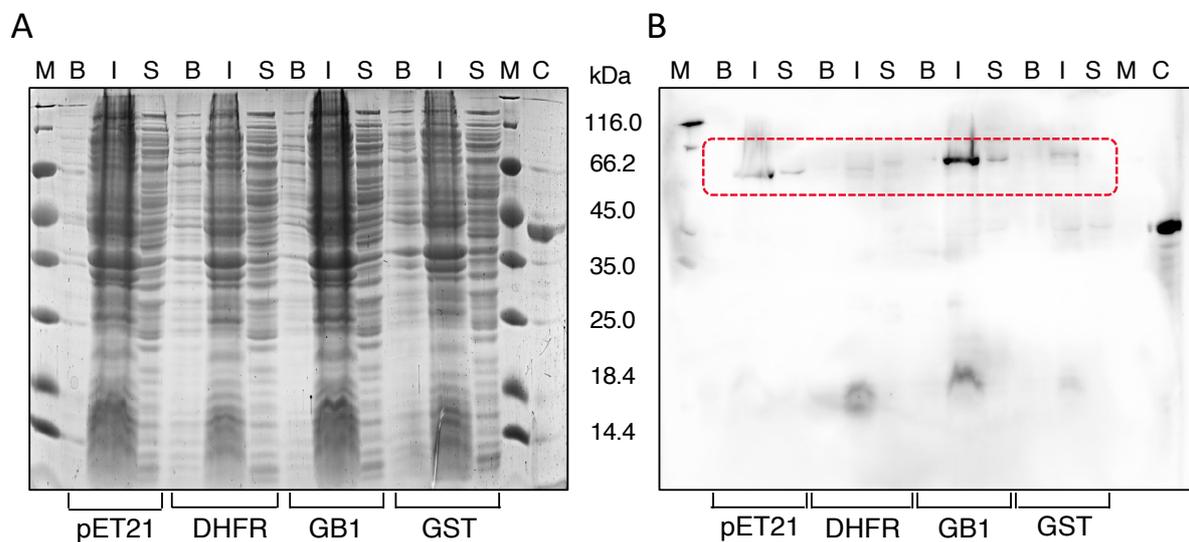


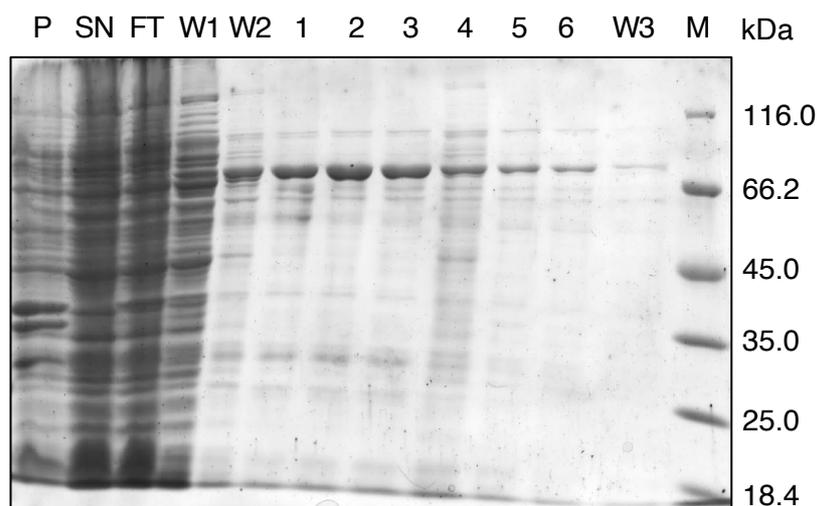
Figure 2.11: Example analysis of a selection of EZS fusions expressed in C41(DE3) pLysS cells with post-induction growth at 16 °C overnight. (A) SDS-PAGE gel analysing total protein content of expression samples. (B) Western blot of gel run in duplicate to (A) using an antibody His-probe. Bands relating to the expected molecular weight of fusion EZS constructs are highlighted in the red box showing small amounts of mostly insoluble protein. Sample abbreviations: M = molecular weight marker, B = before

induction, I = insoluble fraction, S = soluble fraction and C = control of known His-tagged protein.

2.5.2 Expression of EZS using C41(DE3) cells

To further troubleshoot the lack of overexpression, the choice of host strain was reinvestigated. In the original literature EZS was expressed using the C41(DE3) cell line: a strain derived from BL21(DE3) but with genetic mutations that confer tolerance to the expression of toxic proteins. Specifically, the strain has a mutation which reduces the T7 polymerase activity and thereby cell death associated with overexpression of recombinant proteins. The C41(DE3) pLysS strain additionally contains a chloramphenicol resistant plasmid encoding a T7 lysozyme which further inhibits T7 polymerase activity and prevents background expression.¹⁷¹ It was hypothesised that following induction using the pLysS strain, the sudden 'switch on' of T7 polymerase expression, and therefore target protein, was potentially problematic for the cells, and trying the non-pLysS strain may be beneficial. The C41(DE3) cell line was therefore sought and received as a gift from Dr. Petra Bleeker (University of Amsterdam, Netherlands). As the presence of fusion tags were shown to have no apparent effect on expression levels, only the pET21-EZS vector was taken forward.

Competent C41(DE3) cells were prepared and transformed with pET21-EZS. A starter culture was made from a single colony and used to inoculate LB media (500 mL). The cells were cultured at 37 °C until OD₆₀₀ of 0.7 at which point they were subjected to a 30 min cold shock at 4 °C. This was to ensure the cells were sufficiently cooled before induction and to encourage the basal expression of folding chaperones at low temperature expression.¹⁶⁶ Isopropyl β-d-1-thiogalactopyranoside (IPTG, 0.1 mM) was then added to induce the cells and the culture incubated for a further 18 hours at 18 °C. The cells were then harvested by centrifugation, lysed in buffer and the soluble protein content analysed using Ni²⁺ affinity chromatography with an imidazole gradient. In an initial purification attempt, protein bound to the resin was washed with buffer containing imidazole at 10 and 25 mM (10 mL each) before applying buffer containing 250 mM imidazole to elute his-tagged bound protein. Aliquots of 1 mL were collected to ascertain how much buffer was necessary to fully elute the protein and tested for the presence of protein using A₂₈₀ absorbance (via Nanodrop), before applying a final wash of buffer containing 500 mM imidazole (10 mL). Samples from the purification fractions were then analysed by SDS-PAGE (Figure 2.12).



EZS: 88.5 kDa, found approx. 66.2 kDa

Figure 2.12: SDS-PAGE gel showing the first purification of EZS C41(DE3) expressed via Ni^{2+} affinity chromatography using an imidazole gradient. Sample abbreviations: P = pellet, SN = supernatant, FT = flow through, W1 = 10 mM imidazole wash, W2 = 25 mM imidazole wash, 1 - 6 = individual 1 mL fractions collected over 250 mM imidazole wash, W3 = 500 mM imidazole wash and M = molecular weight marker. In later purifications washes 1 and 2 were lowered to 0 mM and 10 mM imidazole respectively.

SDS-PAGE analysis revealed that using the C41(DE3) cell line under these conditions overexpressed EZS which can be identified from the band at approximately 66.2 kDa in the gel (Figure 2.12), negating further western blot analysis. This is consistent with previous work where, despite its predicted molecular mass of 88.5 kDa, EZS has been shown to run at a lower molecular weight.¹⁵³ This observation may be explained considering how SDS-PAGE works: post heat denaturation, negatively charged SDS molecules coat proteins in proportion to their size therefore larger proteins can be separated from smaller by the relative difference in net negative charge. However, proteins with a larger hydrophobic content may become coated by more SDS, resulting in a net negative charge greater than its relative size and therefore migrate further through the gel matrix. This initial purification further revealed apparent loss of EZS in the wash fractions which could result from weak interactions between the His-tag of EZS and the Ni^{2+} resin. The purification protocol was therefore adapted to include initial washes with 0 mM and 10 mM imidazole only.

To estimate the improved protein yield more accurately and further purify the fractions for later enzymatic activity assays fractions containing EZS were purified using anion exchange chromatography with a quaternary ammonium functionalised sepharose (Q-seph) column. The isoelectric point (pI) of a protein is the pH at which the net charge is zero, and for EZS it

is estimated to be 6.02 (calculated by the Benchling webtool).¹⁶³ Fractions from the affinity purification were first desalted using the standard method (Section 8.1.14) into 100 mM TRIS pH 8.0 buffer to remove imidazole and salt and provide EZS with a net negative charge, allowing it to bind strongly to the Q-seph column. EZS was loaded onto the column and a gradient of increasing NaCl concentration applied. The purification was monitored by UV absorbance at 280 nm and fractions relating to peaks in the absorbance were analysed by SDS-PAGE (Figure 2.13). EZS was found to elute at approximately 115 - 185 mM NaCl.

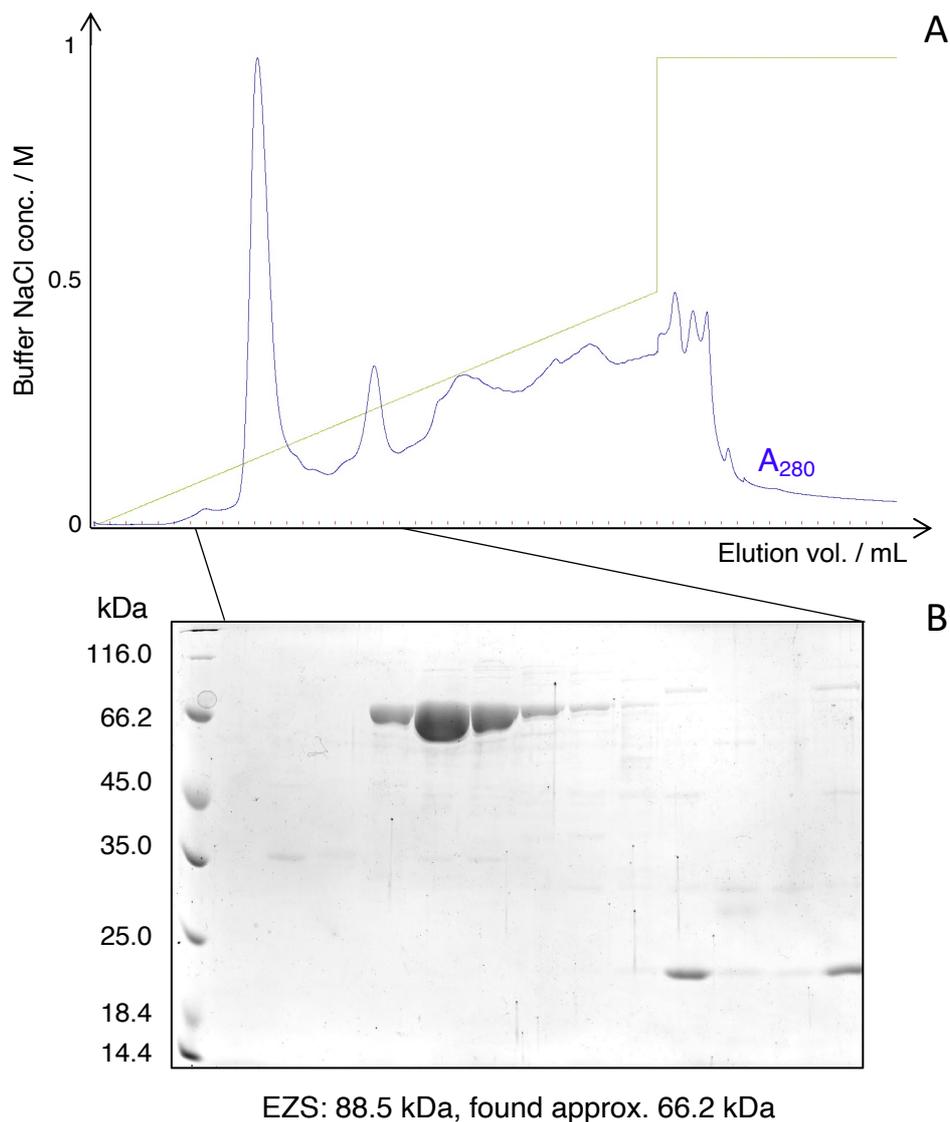


Figure 2.13: (A) Anion exchange chromatogram of EZS purification followed by A_{280} absorbance (blue line) showing the gradient of increasing NaCl concentration (green line). (B) SDS-PAGE gel showing samples taken from fractions collected over the first two peaks in the chromatogram as indicated by connecting lines to (A). Bands of pure protein seen at approximately 66.2 kDa relate to EZS.

Fractions containing pure EZS were then combined and concentrated by spin filtration. The final concentration was measured by Bradford assay¹⁶⁴ from which a yield of 11 mg/L was determined: a considerable improvement over any yields previously achieved in the group. In later expressions, TB media was used instead of LB media and the cells cultured to an OD₆₀₀ of 1.2 – 1.4 before undergoing the 30 min cold shock and induction. Even higher yields of around 15 mg/L were achieved, likely owing to the greater cell density cultivated by the pH buffered TB, and therefore a higher density of protein producing cells per mL of culture. These conditions were used for all further EZS expressions.

2.5.3 EZS activity testing and compound characterisation

EZS activity was tested by taking an aliquot of pure protein and incubating it with a small amount of (Z,Z)-FDP. The product was extracted with pentane (~1 mL) and analysed by gas chromatography coupled mass spectrometry (GC-MS) using electron impact (EI⁺) ionisation. The resulting total ion chromatogram showed a single product peak eluting at 12.46 min (Figure 2.15). The mass spectrum of this compound gave the molecular ion peak at $m/z = 204.4$ (17%, [C₁₅H₂₄]⁺), the base fragment at $m/z = 119.3$ (100%, [C₉H₁₁]⁺) and second most abundant ion at $m/z = 93.2$ (88%, [C₇H₉]⁺). An explanation of how these ions may form can be found in Section 3.2.3. The ion at $m/z = 91.2$ (39%, [C₇H₇]⁺) relates to the aromatically stabilised tropylium ion which forms via a ring expansion mechanism from its benzyl precursor.¹⁷² Structures of these ions and others which could relate to the m/z values found experimentally are given in Figure 2.14. This data is consistent with the literature for 7-*epi*-zingiberene.¹⁷³

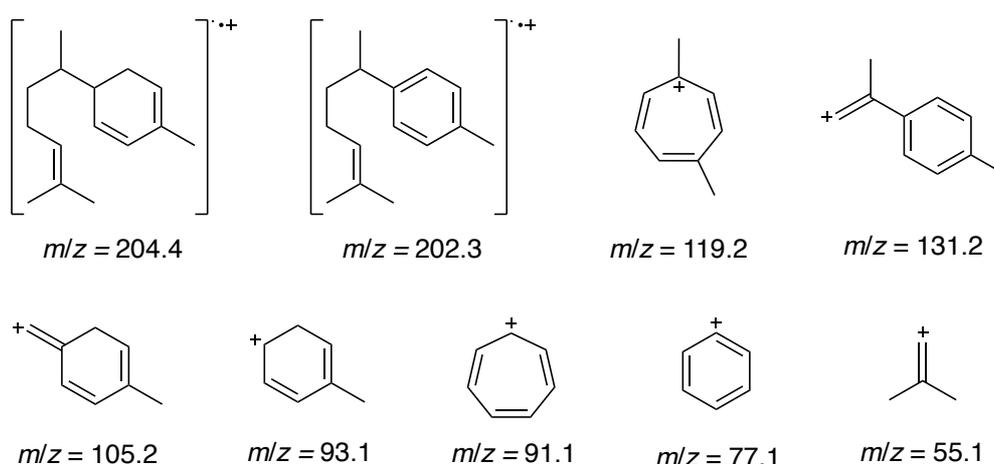


Figure 2.14: Structures of expected and potential ions generated from electron impact ionisation of 7-*epi*-zingiberene.

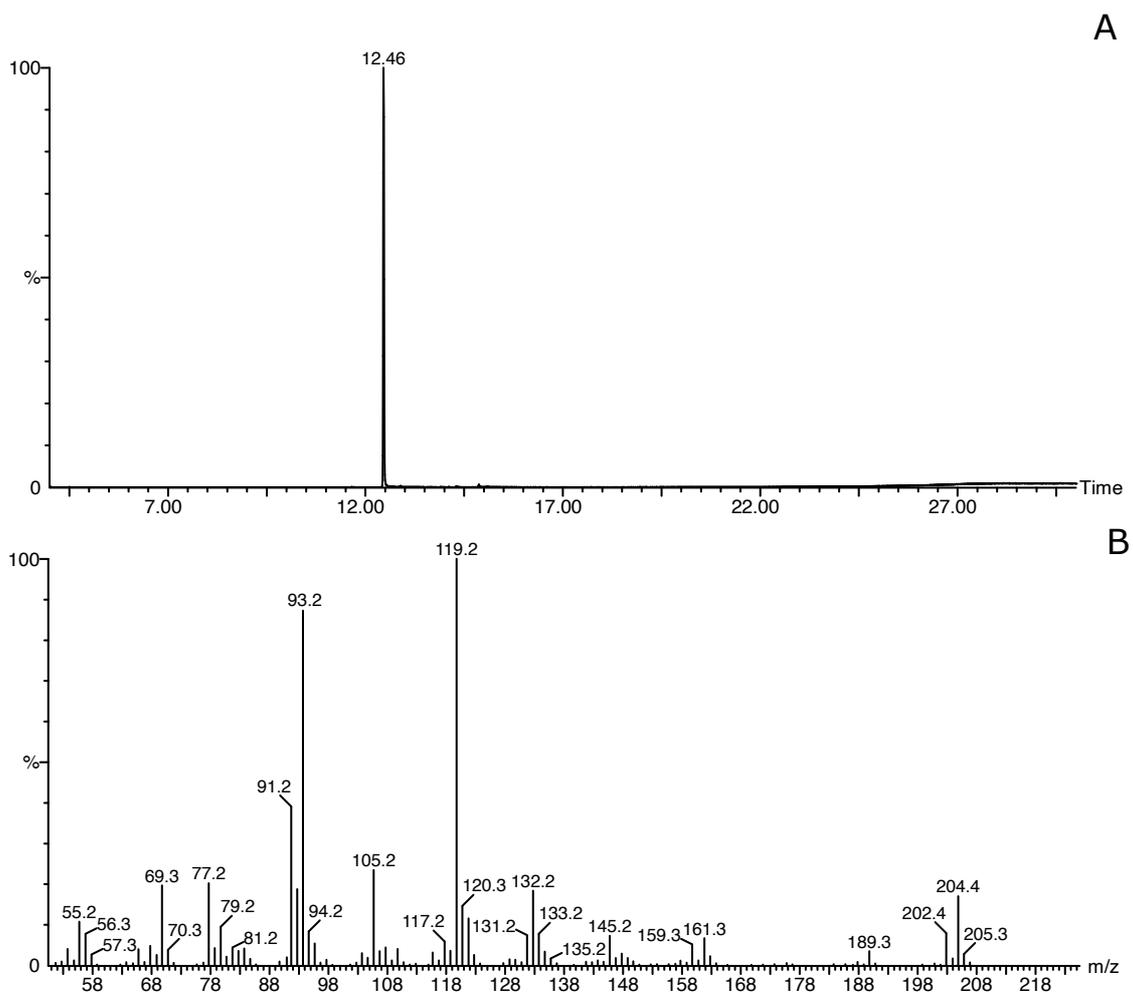


Figure 2.15: (A) Total ion chromatogram of the pentane extractable products resulting from the incubation of (Z,Z)-FDP with EZS. (B) Mass spectrum of the compound eluting at 12.46 min in (A).

2.6 Chemoenzymatic 7-*epi*-zingiberene synthesis

With a reliable way of expressing and purifying adequate quantities of active EZS, and knowing its substrate could readily be produced *in vitro*, a preliminary scaled assay was attempted with the aim of producing enough 7-*epi*-zingiberene to allow NMR characterisation of the direct enzymatic product for the first time and explore how the new chemoenzymatic system may be expanded for sesquiterpene production.

The reaction was prepared in a similar way to the production of (Z,Z)-FDP but additionally contained β -cyclodextrin (β -CD). Work within the Allemann group alongside this project by Florence Huynh found that cyclodextrins (particularly β -CD) aided FDP solubilisation and prevented precipitation. Controls performed without enzyme showed that the cyclodextrin did not contribute to substrate turnover.^{157,161} Historical approaches towards sesquiterpene

isolation by the group typically involved large assay volumes to keep FDP concentration below 2 mM (the approximate threshold at which it precipitates in the presence of the Mg^{2+} cofactor), therefore they were often practically challenging to produce in sufficient quantities in batch. Cyclodextrin structures possess sub nanometre hydrophobic cavities with hydrophilic surfaces which are proposed to interact favourably with the amphiphilic FDP and keep it in solution.^{157,161} For this upscaled assay, β -CD was used meaning that the substrates could be incubated at higher concentrations, enabling smaller assay volumes to be used. Furthermore, for particularly volatile sesquiterpenes, a pentane overlay would be used to trap the hydrocarbon products of enzymatic assays, however this was anticipated to have detrimental effect on terpene synthase activity and was therefore omitted.

2.6.1 Chemoenzymatic production of 7-*epi*-zingiberene and characterisation

The previously prepared phosphorylated substrates stocks of IDP (10 mL) and DMADP (5 mL) were incubated with zFPS (1.25 μ M), EZS (6 μ M) and β -CD (250 μ M) in 30 mL total volume at room temperature for 36 hours with gentle stirring (Figure 2.16). EZS was purified from 2 L of culture to obtain enough for the incubation.

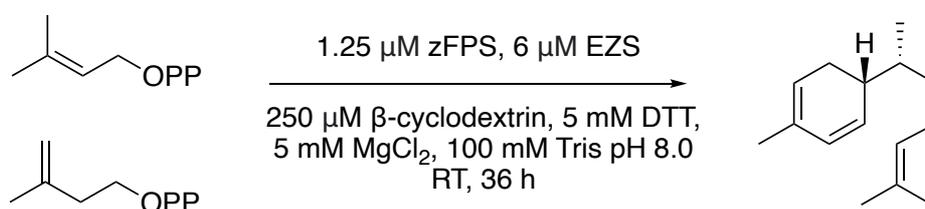


Figure 2.16: Conditions used for an upscaled reaction using IDP and DMADP (prepared via the kinases) with zFPS and EZS to produce sufficient enzymatic product for NMR spectroscopic characterisation.

After 36 hours, the reaction mix was extracted exhaustively using pentane (5 mL portions). The product was then isolated by carefully removing the pentane under reduced pressure and washed with $CDCl_3$ (3 x 0.5 mL) to afford a colourless oil (approximately 1 mg, 12% isolated yield) which was sufficient for further analysis by 1H and ^{13}C NMR spectroscopy. The resulting fully assigned 1H NMR spectrum is shown in Figure 2.17.

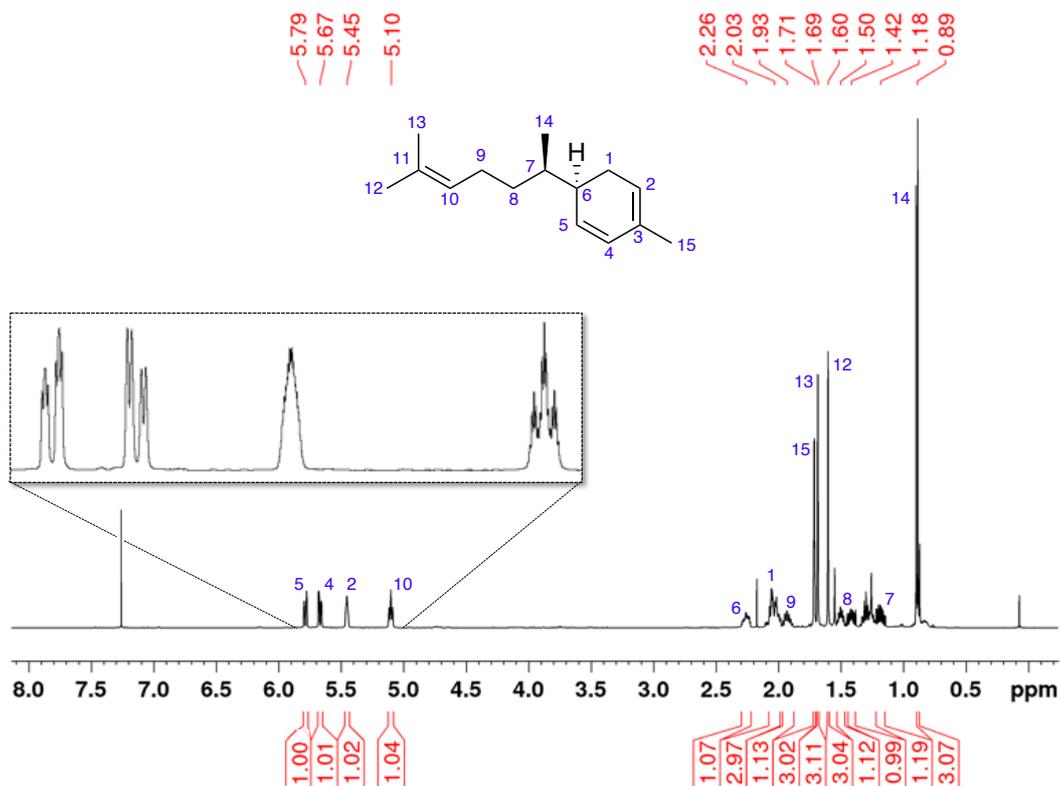


Figure 2.17: Assigned ^1H NMR (500 MHz, CDCl_3) spectrum of *7-epi*-zingiberene produced directly from precursors IDP and DMADP using enzymes zFPS and EZS.

The spectrum shows four peaks between 5.0 and 6.0 ppm which are consistent with the four vinylic protons of the molecule. Protons of the three allylic methyl groups can be found between 1.5 and 2.0 ppm whilst the C7 methyl group doublet is found upfield at 0.89 ppm. Indicative of the stereochemistry of *7-epi*-zingiberene, the allylic ring proton of C6 relates to the peak at 2.26 ppm which is downfield from the C7 proton at 1.18 ppm. The diastereotopic protons of the CH_2 group adjacent to the sp^3 C7 can be seen at 1.42 and 1.50 ppm. Completing the spectrum, the allylic CH_2 group protons at positions 1 and 9 are positioned around 2.0 ppm. This data is in accordance with the literature where the absolute stereochemistry of *7-epi*-zingiberene has previously been determined with comparison to an authentic standard of zingiberene. Building on previous work by Breeden and Coates (who first isolated and characterised *7-epi*-zingiberene from tomato extracts),¹⁷³ Bleeker *et. al* derivatised the sesquiterpene directly with 4-phenyl-1,2,4-triazoline-3,5-dione (PTAD) and obtained an X-ray crystal structure of the subsequent Diels-Alder adduct from which the two chiral centres could be assigned.¹³⁶ The direct enzymatic product of EZS isolated from *in vitro* characterisation

was compared via chiral GC to the product of zingiberene synthase in later work by Bleeker *et. al* to further unambiguously confirm the absolute stereochemistry.¹³⁷

The characterisation data produced in this work shows that pure 7-*epi*-zingiberene was successfully produced via this short, easily assembled, 4-enzyme *in vitro* synthesis, where it is not necessary to isolate the intermediate (Z,Z)-FDP. In this instance, the isolated yield was fairly low so further work could be undertaken to improve this. Comparisons with and without pentane overlays would establish whether it does prevent loss of compound to volatilisation compared to potential enzyme deactivation. Additionally, at this scale, the small quantities of volatile compounds being produced may easily be lost during the removal of pentane and any handling, causing the yield to suffer compared to if a larger assay was trialled. Although β -CD is very effective at preventing precipitation, it may be slowing the reaction rate through its interactions with the FDP, therefore longer reaction times may help. Furthermore, the concentration of EZS could be lowered to enable more protein to be used for substrate conversion per batch of purification, and therefore further increase the assay size. Overall, this initial successful outcome represents a good starting point for continued investigations into chemoenzymatic terpenoid production at a greater scale.

2.7 Summary

This chapter describes the development of a new chemoenzymatic synthesis of (Z,Z)-FDP, the atypical isoprenoid precursor to 7-*epi*-zingiberene, a sesquiterpene of agrochemical importance. It builds upon the efficient chemoenzymatic synthesis of precursors IDP and DMADP recently developed in the Allemann group which makes use of two kinases and an ATP recycling system. By incorporating the *cis*-prenyltransferase zFPS in a single further reaction, (Z,Z)-FDP can be readily made with isolated yields up to 92%. The previously most optimal synthesis of (Z,Z)-FDP was a low yielding, practically challenging 11-step reaction, so this new 3-enzyme route represents a drastically improved synthesis in terms of selectivity, atom economy and labour required. Initial attempts at the zFPS catalysed condensation reaction were not successful when using IDP and DMADP prepared using the synthetic TEAP method (Keller¹⁵⁸), highlighting the challenges faced in producing isoprenoid diphosphates in sufficiently pure quantities and further demonstrating the effectiveness of the new *in vitro* method for producing the universal precursors IDP and DMADP.

As (Z,Z)-FDP could be produced reliably, attention turned to how 7-*epi*-zingiberene could be made with incorporation of EZS into the chemoenzymatic route. EZS expression was previously very poor therefore attempts were made to improve it initially via the use of fusion proteins. A selection of highly soluble proteins were fused to EZS via Golden Gate cloning and

expressed under different conditions, however Western blot analysis revealed that expression levels remained very low. Further troubleshooting involving a change to the C41(DE3) cell line showed a remarkable overexpression of the target to now useful levels. It was then possible to combine EZS with zFPS and the kinase prepared IDP and DMADP to directly produce enough pure *7-epi*-zingiberene for NMR characterisation. The overall route is practically straightforward and has excellent potential for further development to improve yields and generally for scaled production of other high value terpenes.

**Investigation of the Hydride
Shifts of the Catalytic Mechanism
of EZS**

3.1 Introduction

EZS is a plant sesquiterpene synthase from wild tomato which converts the atypical substrate (Z,Z)-FDP into 7-*epi*-zingiberene, a useful repellent semiochemical.^{34,136,137} EZS carefully controls the quenching of multiple reactive carbocationic intermediates through hydride shifts and proton loss with remarkable fidelity to achieve a single product outcome. Due to the transient nature of such intermediates, which are typically unobservable spectroscopically, isotopic labelling experiments have classically been used to study terpene synthase reaction pathways and gain a detailed understanding of their catalysis.⁷⁸

3.1.1 Aims of this work

This chapter describes an investigation of the hydride shifts of the EZS catalytic mechanism using deuterium containing isotopologues of (Z,Z)-FDP, synthesised using a chemoenzymatic approach. As discussed in earlier chapters, total synthetic routes to (Z,Z)-FDP are difficult, time consuming and low yielding and are additionally complicated by the introduction of isotopic atoms.¹⁵³ Steps to incorporate isotopes may need to take place relatively early in the synthesis which is not cost effective, therefore shortening the synthesis of the required deuterium isotopologues is desirable. This worked aimed to develop a simple chemoenzymatic route to the required substrates by exploiting the *cis*-prenyltransferase (zFPS) which naturally synthesises (Z,Z)-FDP from the native isoprenoid precursors IDP and DMADP. C₁₀ NDP is the natural intermediate substrate resulting from the condensation of DMADP and one IDP, with a further addition reaction taking place between NDP and a second IDP to yield the C₁₅ (Z,Z)-FDP.³⁴ C₁₀ substrates labelled with deuterium and IDP were chemically synthesised and combined with zFPS to generate the C₁₅ labelled substrates *in situ* for subsequent catalytic conversion by EZS. Isotopically labelled reaction products were analysed by GC-MS to provide evidence for which hydrides are involved in the mechanism of 7-*epi*-zingiberene formation.

3.1.2 Proposed reaction mechanism of EZS

As with all sesquiterpene synthases, the mechanism by which EZS converts (Z,Z)-FDP into product proceeds via an initial ionisation step where the diphosphate group of the linear isoprenoid substrate is lost through either a stepwise or concerted process and a carbocation is formed at the C1 position (Figure 3.1). The cation is then subject to attack from the middle double bond of the substrate in a 1,6-cyclisation reaction, resulting in the bisabolyll cation intermediate. The reaction could plausibly proceed via a single [1,3]-hydride shift from either

C1 or C5, or via two sequential [1,2]-hydride shifts involving firstly a hydride shift from C6 then a second hydride shift from C1 or C5. To form 7-*epi*-zingiberene, the final carbocation created by the hydride shift is quenched via the loss of a proton to form the second double bond on the ring. Mechanistic investigations have been undertaken to establish whether the ring closure step is stepwise or concerted in nature, and specifically which hydrides move around the ring to quench the carbocation intermediates.

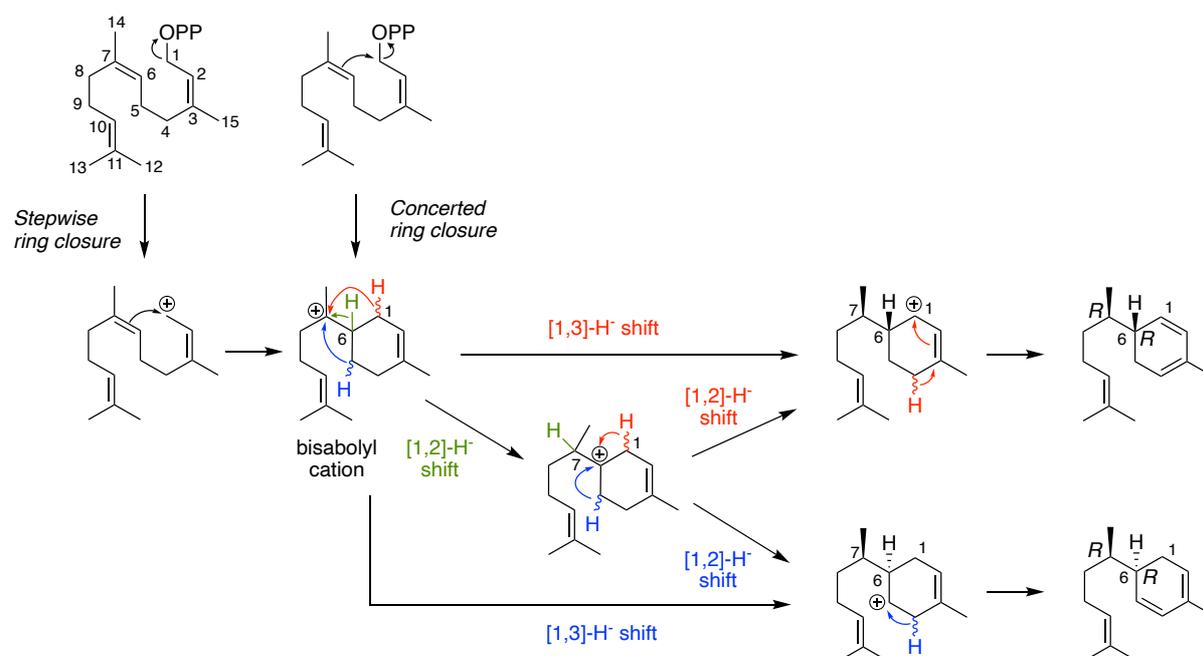


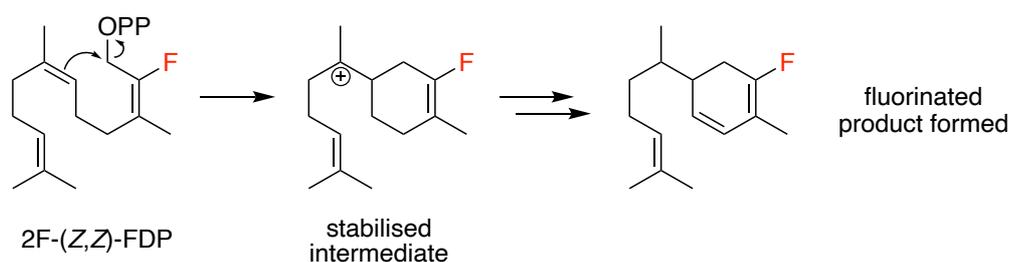
Figure 3.1: The potential reaction pathways catalysed by EZS with substrate (Z,Z)-FDP.

3.1.3 Understanding the ring closure step using fluorinated substrate analogues

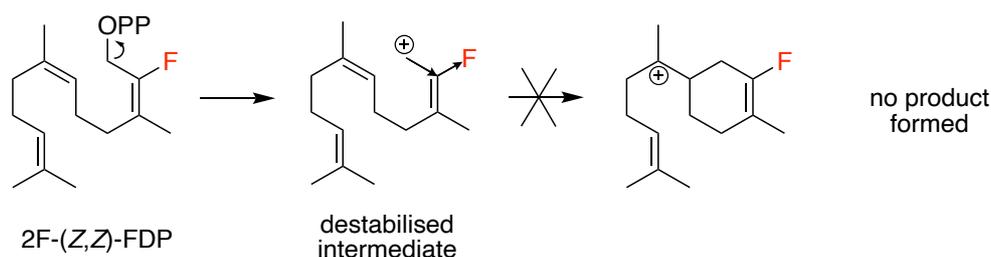
Previous work in the Allemann group on the catalytic mechanism of EZS by Dr. Chris Jones examined the initial reaction steps. Whether the loss of diphosphate and 1,6-ring closure happens in a stepwise or concerted manner was studied using the (2*E*,6*Z*)-2-fluorofarnesyl diphosphate (2F-(Z,Z)-FDP) substrate analogue (Figure 3.2).¹⁵³ It was hypothesised that if the reaction occurs in a stepwise manner, the farnesyl cation intermediate, with localised positive charge at the C1 position, would be destabilised by the presence of an electron withdrawing fluorine at the alpha or C2 position (see Section 1.3.3). However, if the reaction mechanism is concerted, the cyclisation may be expected to occur avoiding the destabilised intermediate,

resulting in product formation with fluorine substitution. No products were detected from the incubation of EZS with 2F-(Z,Z)-FDP which may indicate a stepwise loss of diphosphate and ring closure, as the fluorofarnesyl cation formation is inhibited, preventing the subsequent cyclisation. This substrate was characterised as a competitive inhibitor through kinetic analyses: with a similar K_M value to the native substrate, the fluorine was shown to minimally affect binding, therefore the increased energy of the fluorofarnesyl cation results in the inhibition of EZS.

A Evidence for concerted ring closure:



B Evidence for stepwise ring closure:



C Evidence for bisaboly cation intermediate:

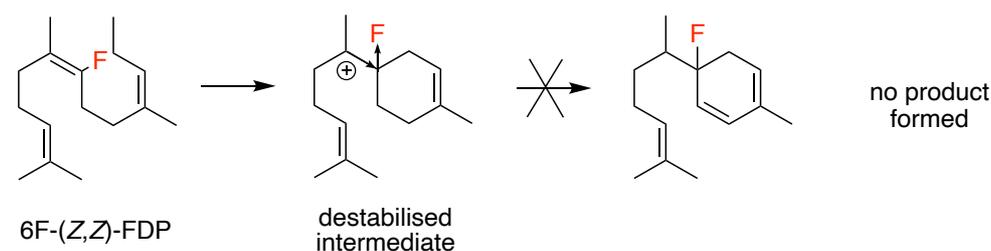


Figure 3.2: The 2F-(Z,Z)-FDP substrate was used to ascertain whether the initial ionisation and ring closure step of the mechanism of EZS is (A) concerted or (B) stepwise by disfavouring the formation of the proposed farnesyl cation intermediate. (C) The 6F-(Z,Z)-FDP substrate was similarly used to indicate whether the bisaboly cation intermediate is formed.

After the diphosphate ionisation and ring closure step, the bisabolyli cation is proposedly generated where the positive centre is located at the tertiary C7. To provide evidence for this intermediate, (2Z,6E)-6-fluorofarnesyl diphosphate (6F-(Z,Z)-FDP) was used (Figure 3.2), as in a similar fashion, the cation generated would be destabilised by the alpha substituted fluorine of the substrate analogue. Upon incubation of the substrate with EZS, no enzymatic products were detected. This implies that bisabolyli cation formation has been prevented and the mechanism is likely to go through this intermediate. Kinetic studies showed the substrate interacts with the enzyme; analogously identified as a competitive inhibitor.

3.1.4 Understanding the hydride shifts using deuterated substrate analogues

After the 1,6-cyclisation and the generation of the bisabolyli cation (Figure 3.1) there are several plausible pathways to the final product when considering possible hydride shifts from positions 6, 5 and 1 of the carbon ring. Generally, [1,2]- and [1,3]-hydride shifts are considered theoretically feasible (with energy barriers around 10 kcal/mol or less), whereas longer range shifts are not.⁵⁴ To examine these possible reaction pathways, (Z,Z)-FDPs labelled with deuterium at C5 and C6 were targeted in this work with the aim of analysing their resulting enzymatic products using GC-MS. Deuterium containing ions have an increased *m/z* value compared to ions of the unlabelled product, therefore their location can be traced to different parts of the molecule by comparing which ion fragments have different *m/z* values compared to the native product.

3.2 Preparation of [²H₁-6]-(Z,Z)-FDP and incubation with EZS

3.2.1 [²H₁-6]-(Z,Z)-FDP as a mechanistic probe

To ascertain whether the hydrogen at the C6 position of the substrate is involved in the mechanism, (Z,Z)-FDP labelled with deuterium at C6 position ([²H₁-6]-(Z,Z)-FDP) was desired. Considering the possible mechanistic routes proposed previously, the outcome of the incubation of EZS with [²H₁-6]-(Z,Z)-FDP could give two chemically different products (Figure 3.3) where the deuterium either stays at the C6 (as in route A or route C) or moves to the C7 (as in route B and route D).

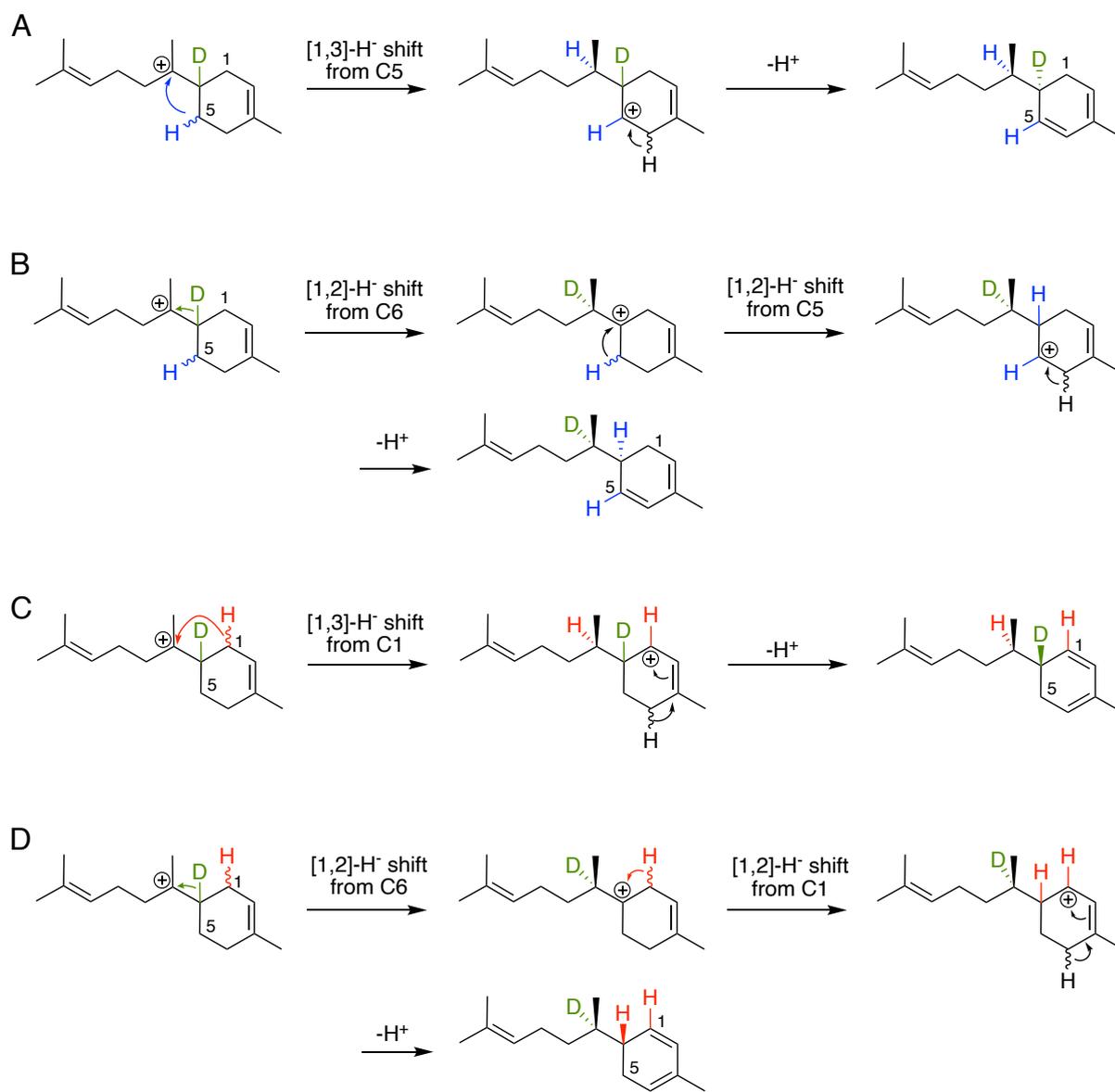


Figure 3.3: Possible reaction pathways of [2H₁-6]-(Z,Z)-FDP from the bisaboly carbocation via: (A) [1,3]-hydride shift from C5, (B) [1,2]-hydride shift from C6 followed by [1,2]-hydride shift from C5, (C) [1,3]-hydride shift from C1 and (D) [1,2]-hydride shift from C6 followed by [1,2]-hydride shift from C1.

3.2.2 Synthesis of [²H₁-2]-neryl diphosphate and IDP

A chemoenzymatic route to [²H₁-6]-(Z,Z)-FDP was devised in an attempt to shorten and improve upon total synthesis. This new route involves introducing the deuterium at the C2 of C₁₀ NDP and carrying out the subsequent chain extension enzymatically with zFPS and IDP to give deuteration at the C6 of FDP. The [²H₁-2]-NDP was in turn synthesised from [²H₁-2]-nerol; accessible via a short synthesis adapted from the Anastasia route to unlabelled (Z,Z)-farnesol which begins from neryl acetone (Figure 3.4).¹⁵³

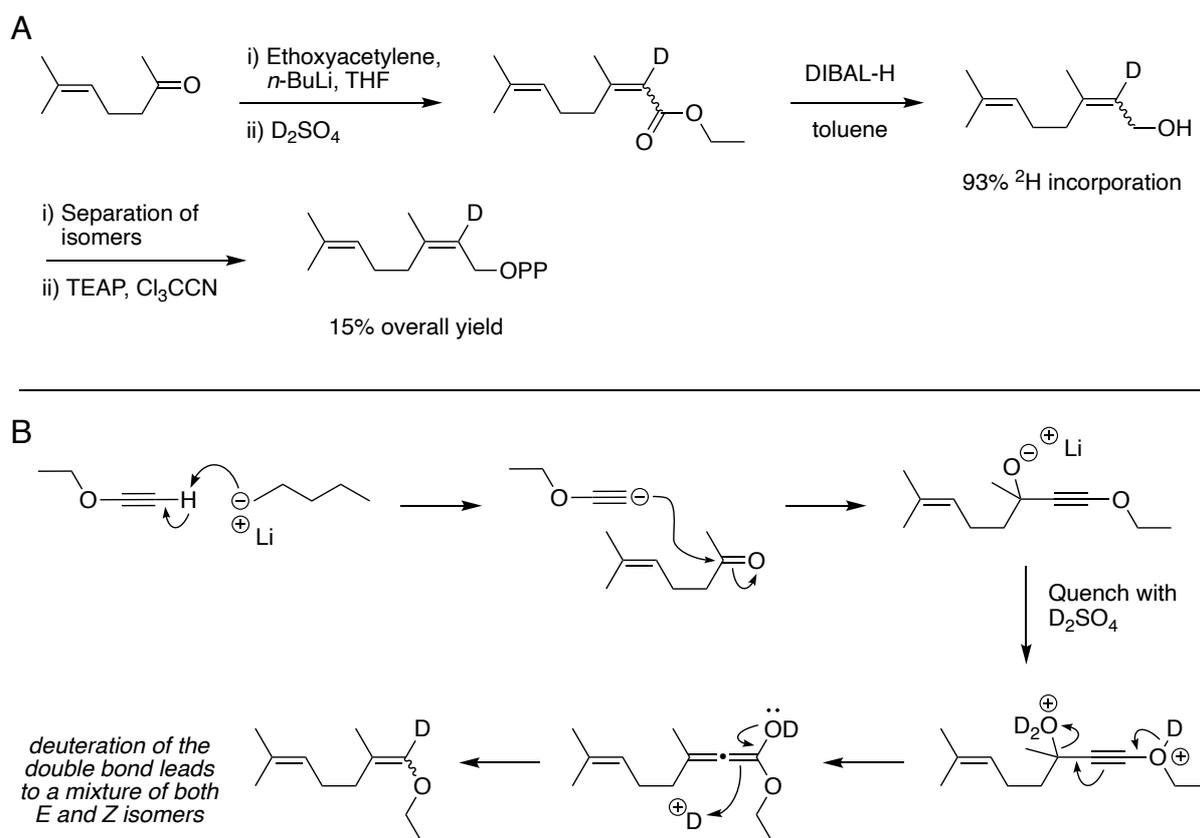


Figure 3.4: (A) Short synthetic route to incorporate deuterium at the C2 position of the C₁₀ substrate [²H₁-2]-NDP. (B) Proposed mechanism for the first reaction where deuterium is incorporated via a quenching step with D₂SO₄ leading to a mixture of isomers.

Ethoxyacetylene was first deprotonated with *n*-BuLi before 6-methylhept-5-en-2-one was added to extend the carbon chain. The deuterium was incorporated during the subsequent quenching step, where D₂SO₄ was used instead of H₂SO₄ to deuterate the intermediate allene and yield a mixture of *E* and *Z* deuterium substituted allylic esters. The mixture was reduced to their respective alcohols from which nerol was purified from the *E* isomer (geraniol) by silica chromatography using an isocratic elution of 10% ethyl acetate in hexane. An overall yield of 40% was achieved for the desired *Z* alcohol (Section 8.3.6). The deuterium incorporation was successfully verified by ²H NMR, but also by comparison of the corresponding proton environments in the ¹H NMR spectra of the labelled and unlabelled nerol (Figure 3.5). The percentage incorporation was established by calculating an integral for the region at 5.44 ppm and found to be 93% ²H.

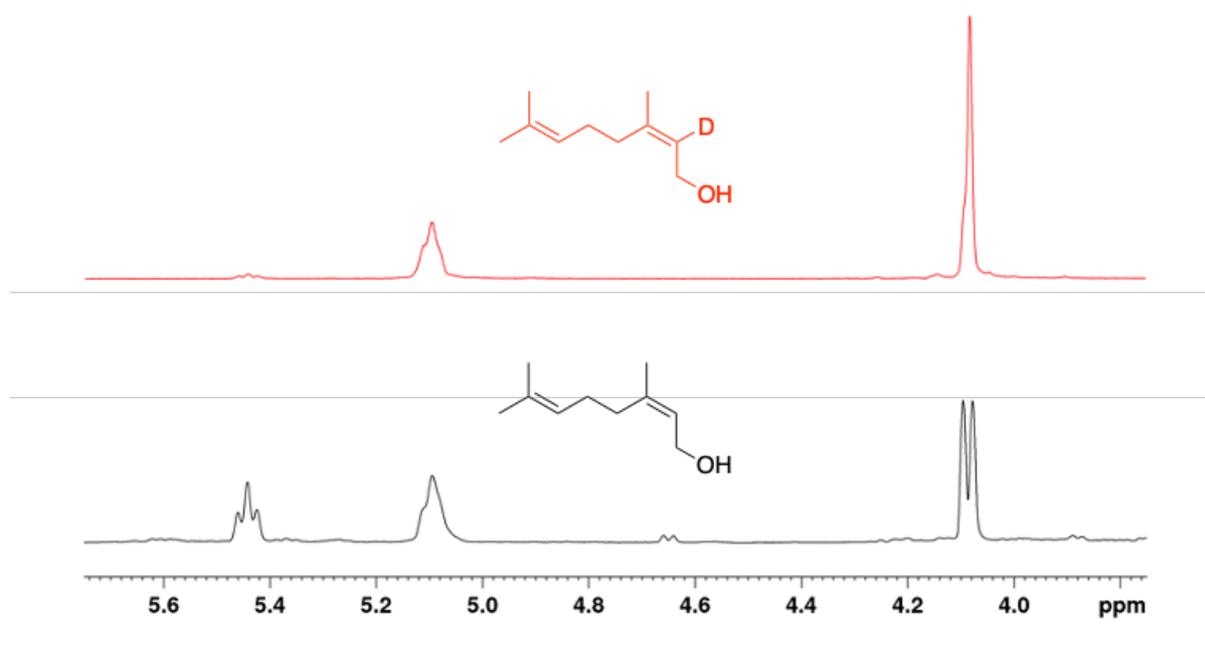


Figure 3.5: Direct comparisons of the ¹H spectra of nerol (black) and [²H₁₋₂]-nerol (red) between 5.75 - 3.75 ppm. The absence of the peak at 5.44 ppm in the red spectrum relates to the now deuterated alkene environment (93% ²H) at C2.

After separation from its isomer, [²H₁₋₂]-nerol was reacted with TEAP (Section 2.2) to diphosphorylate the alcohol group and make it a substrate for zFPS. Solutions of phosphoric acid and triethylamine in acetonitrile were mixed to generate TEAP, which then reacts with the alcohol dissolved in trichloroacetonitrile, generating a mixture of mono-, di- and triphosphates. The mixture of phosphates was separated by silica chromatography using an eluent of 6:2.5:0.5 isopropanol:concentrated ammonium hydroxide:water (v/v/v). Fractions relating to the diphosphate were identified by TLC using the same eluent and combined. Excess

isopropanol and ammonia were removed under reduced pressure and the remaining aqueous solution thoroughly freeze dried to yield the ammonium salt of the diphosphate as a white powder (15% overall yield). Subsequent ^1H , ^{13}C and ^{31}P NMR and HRMS analysis results were consistent with the expected product which was deemed sufficiently pure for use in later enzymatic reactions (Section 8.3.6).

The TEAP reaction was also carried out with unlabelled nerol to produce NDP for control reactions and unlabelled isoprenol to make IDP for use in all prenyltransferase catalysed addition reactions (Section 8.3.2 and 8.3.4).

3.2.3 Incubation of [$^2\text{H}_1$ -6]-NDP with IDP, zFPS and EZS

With [$^2\text{H}_1$ -2]-NDP and IDP prepared, a small-scale reaction was assembled (Figure 3.6) by mixing both compounds with zFPS and EZS in buffer (100 mM TRIS pH 8.0, 5 mM MgCl_2 , 5 mM DTT). The enzymes were produced and stored as previously described enabling use when required (Section 2.4.1 and 2.5.2). Control reactions were carried out in parallel using the unlabelled NDP (Figure 3.7) with zFPS and EZS (and without) to verify products from the labelled incubation result from enzymatic substrate conversion. The reactions were left overnight with shaking (300 rpm) at room temperature, extracted using pentane and analysed by GC-MS (Figure 3.8).

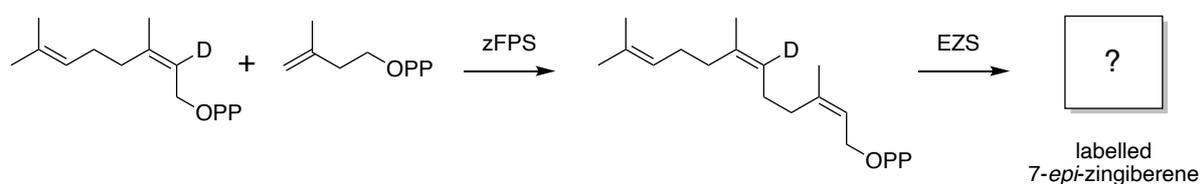


Figure 3.6: The chemoenzymatic preparation of [$^2\text{H}_1$ -6]-(Z,Z)-FDP using the enzyme zFPS and [$^2\text{H}_1$ -2]-NDP with IDP. EZS was used to turn over the *in situ* generated labelled substrate and the pentane-extractable product analysed by GC-MS.

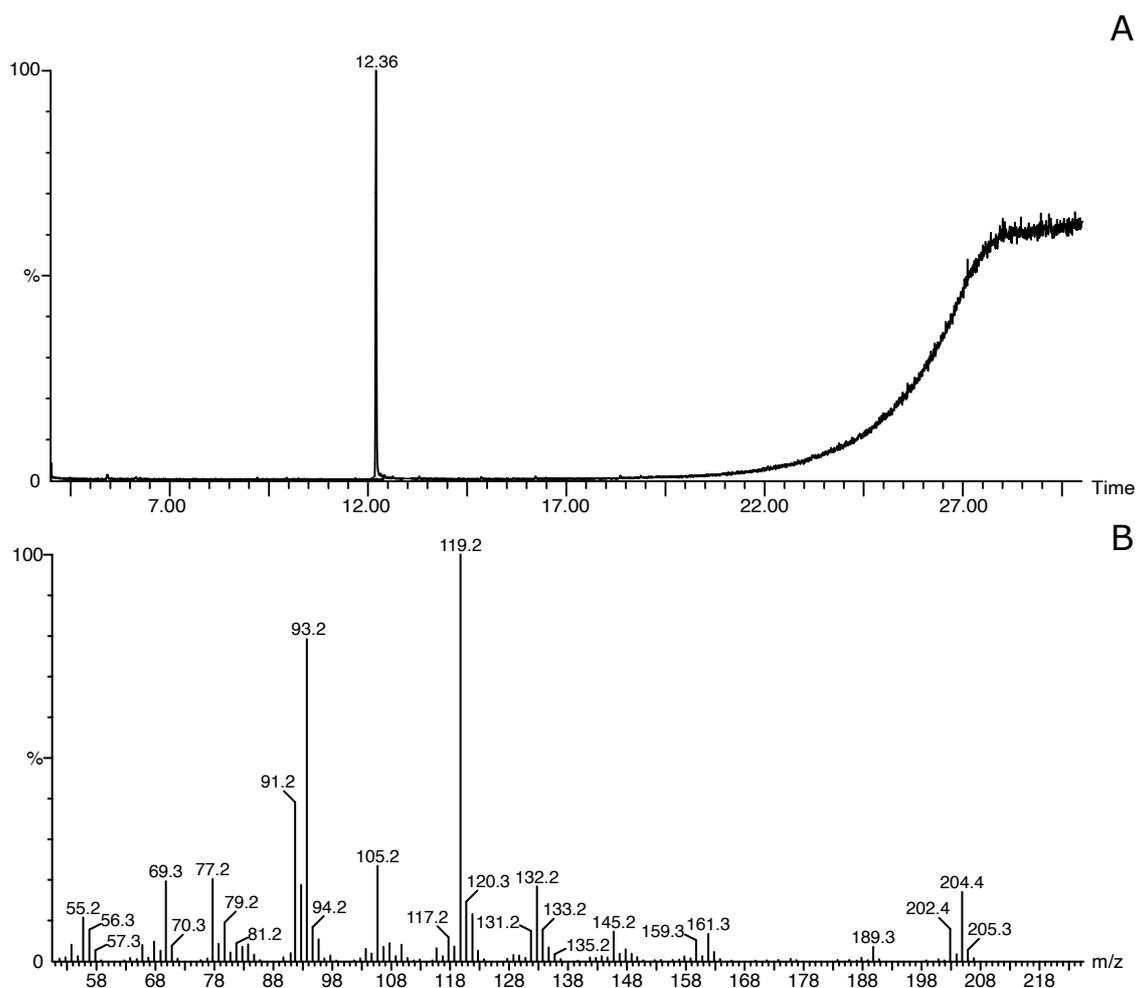


Figure 3.7: (A) Total ion chromatogram of the pentane extractable products resulting from the incubation of the control reaction between NDP with IDP, zFPS and EZS. (B) Mass spectrum of the compound eluting at 12.36 min in (A).

Compared to the mass spectrum of product from the native substrate incubation (Figure 3.7), the product of $[^2\text{H}_1\text{-6}]$ -(Z,Z)-FDP (Figure 3.8) has the parent ion with $m/z = 205.4$ instead of 204.4, showing the incorporation of the single deuterium into the sesquiterpene skeleton giving the overall +1 mass number. The two most abundant ions of the fragmentation pattern of unlabelled 7-*epi*-zingiberene ($m/z = 119.3$ and 93.2) relate to ions which contain atoms of both the C6 and C7 positions which are implicated in the mechanism (see Figures below). Therefore any change to m/z values at these positions can be analysed to reveal whether the deuterium in the substrate $[^2\text{H}_1\text{-6}]$ -(Z,Z)-FDP has stayed in the C6 position or shifted to the C7 during the course of the mechanism.

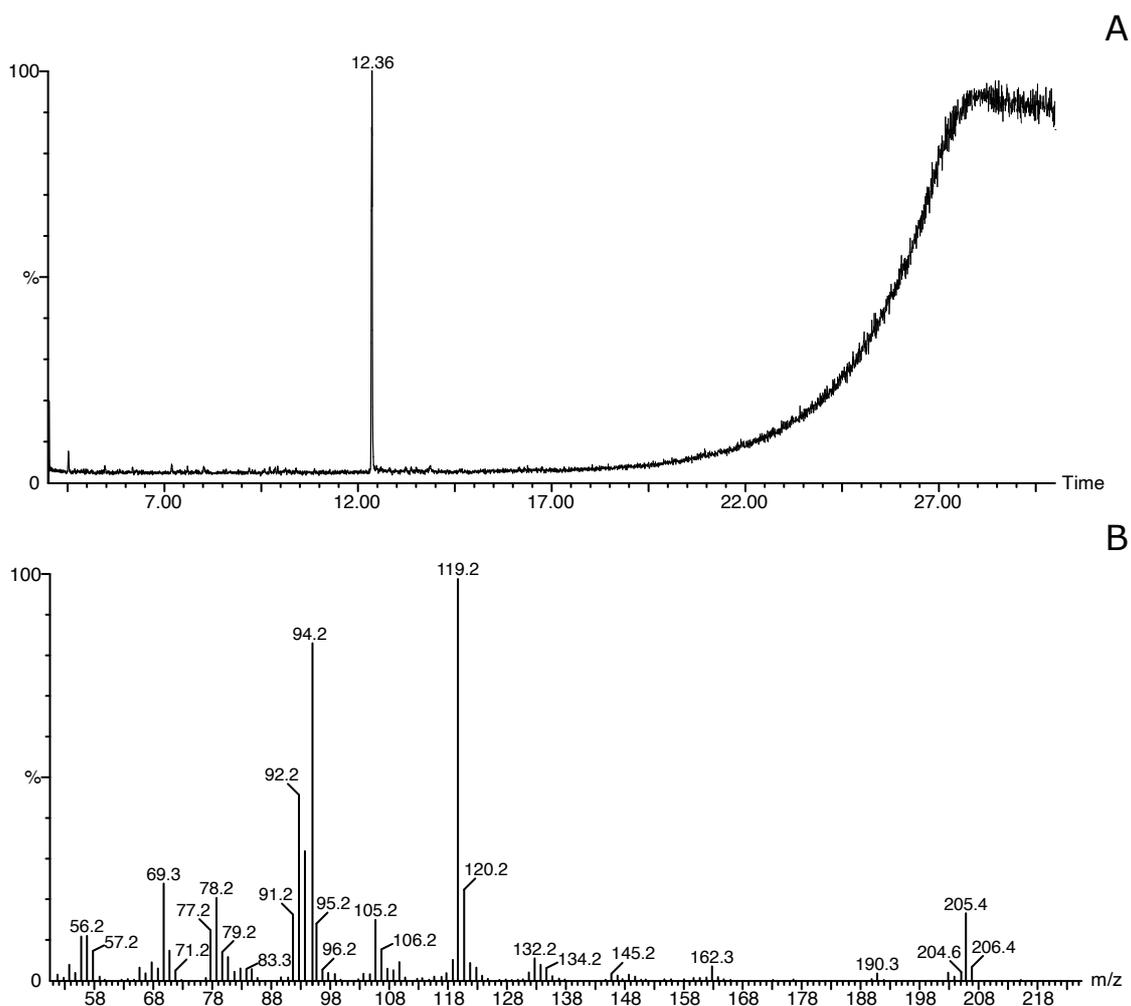


Figure 3.8: (A) Total ion chromatogram of the pentane extractable products resulting from the incubation of [$^2\text{H}_1$ -2]-NDP with IDP, zFPS and EZS. B) Mass spectrum of the compound eluting at 12.36 min in (A).

Firstly, the ion with $m/z = 93.2$ of the mass spectrum of unlabelled 7-*epi*-zingiberene relates to the fragment generated when the bond is broken between C6 and C7 (Figure 3.9). Therefore, if the deuterium has stayed at the C6 position the resulting ion has $m/z = 94$, and if it has moved to C7 the ion will have $m/z = 93$. The fragmentation pattern of the [$^2\text{H}_1$ -6]-(Z,Z)-FDP generated product in Figure 3.8 clearly shows an ion relating to this fragment with a $m/z = 94.2$, as opposed to 93.2 of the native substrate. This is the first indication that the mechanism does not proceed via either of the [1,2]-hydride shift routes.

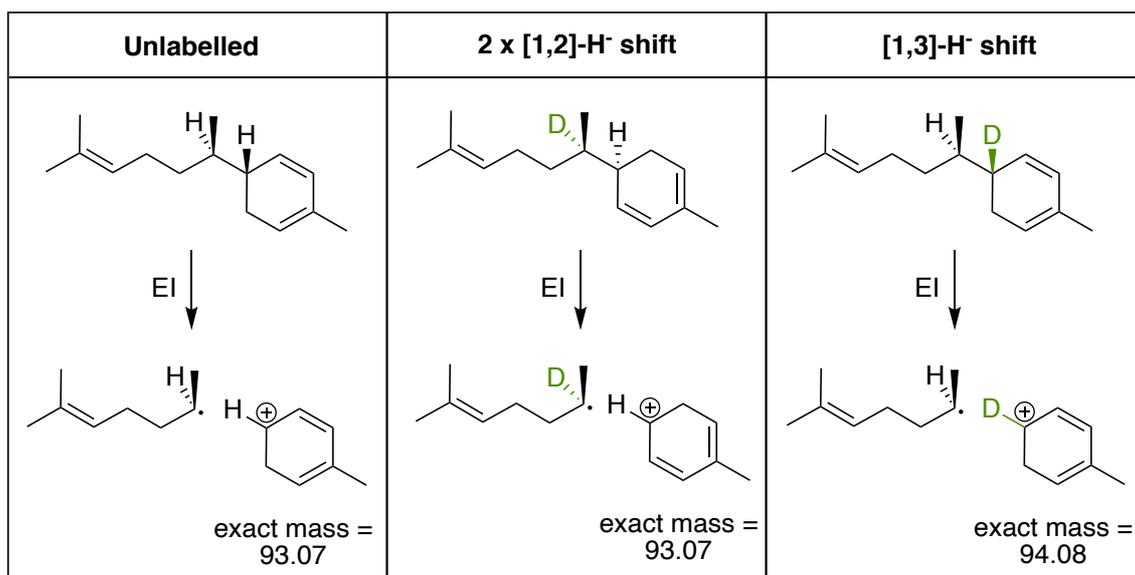


Figure 3.9: Comparison of ions generated by mass spectrometric analysis of 7-*epi*-zingiberene with bond breaking between C6 and C7, generated from both unlabelled substrate and [²H₁₋₆]-(*Z,Z*)-FDP and considering whether sequential [1,2]- or [1,3]-hydride shift pathways were taken.

The most abundant peak in the mass spectrum of unlabelled 7-*epi*-zingiberene is at $m/z = 119.3$ which relates to the ion generated from the bond breaking between C7 and C8 (Figure 3.10). The resulting radical undergoes a dehydrogenation to enable the ring to aromatised and therefore have enhanced stability; formally represented as the tropylium ion and commonly identified in mass spectrometry.¹⁷⁴ In this case, the tropylium ion is methyl substituted, hence leading to a final $m/z = 119.3$. The m/z of this ion is also affected by deuterium substitution at either the C6 or C7. If the deuterium moves via a [1,2]-hydride shift and ends up on C7, the deuterium will be retained during aromatisation and the resulting ion will increase to $m/z = 120$. If the deuterium does not move and stays at C6, when the ion aromatises, the deuterium will be lost generating a final ion with $m/z = 119$. The fragmentation pattern of the [²H₁₋₆]-(*Z,Z*)-FDP generated product in Figure 3.8 shows this ion with a $m/z = 119.2$: consistent with what is expected from the [1,3]-hydride shift routes. This result, taken with the $m/z = 94.2$ of the other fragment described, provides evidence that the hydride at the C6 position remains on the ring and is not involved in quenching the bisabolyl cation at C7. This indicates the pathway likely goes through either of the [1,3]-hydride shift routes from the C1 or the C5 position, rather than sequential [1,2]-hydride shifts described in Figure 3.1.

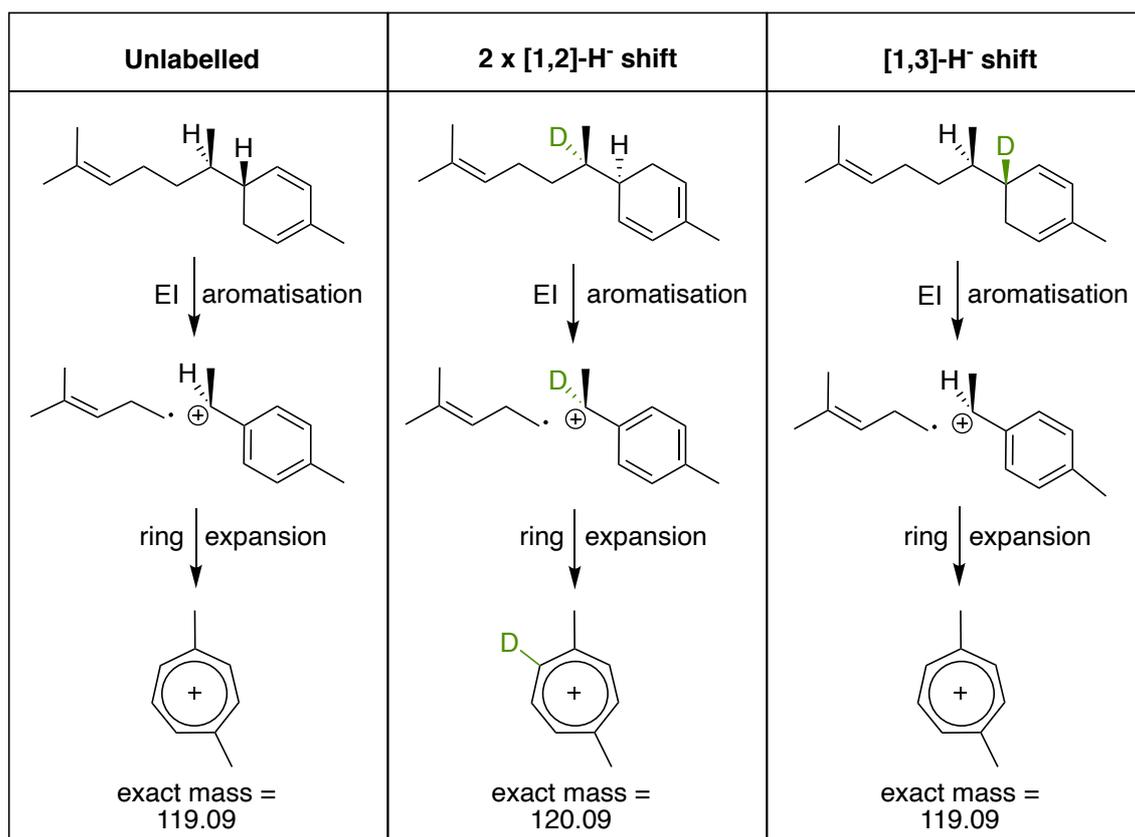


Figure 3.10: Comparison of ions generated by mass spectrometric analysis of 7-*epi*-zingiberene with bond breaking between C7 and C8, generated from both unlabelled substrate and [²H₁₋₆]-(*Z,Z*)-FDP and considering whether sequential [1,2]- or [1,3]-hydride shift pathways were taken.

3.3 Preparation of [²H₂₋₅]-(*Z,Z*)-FDP and incubation with EZS

3.3.1 [²H₂₋₅]-(*Z,Z*)-FDP as a mechanistic probe

With evidence indicating the EZS is more likely to utilise a [1,3]-hydride shift for product formation, the next labelled substrate targeted was [²H₂₋₅]-(*Z,Z*)-FDP, as this substrate could be made readily using [²H₂₋₁]-NDP and IDP in a zFPS catalysed reaction. Using this substrate and examining the [1,3]-hydride shifts possible via C1 or C5, two chemically different product outcomes are plausible (Figure 3.11), which could be distinguished by comparison of the two abundant ions in the resulting mass fragmentation patterns. If the reaction proceeds via the C5 hydride shift, the product will have one deuterium at C7 and one at C5 (route A), or if instead the shifted hydride originates from C1, both deuteriums will remain on C5 (route B).

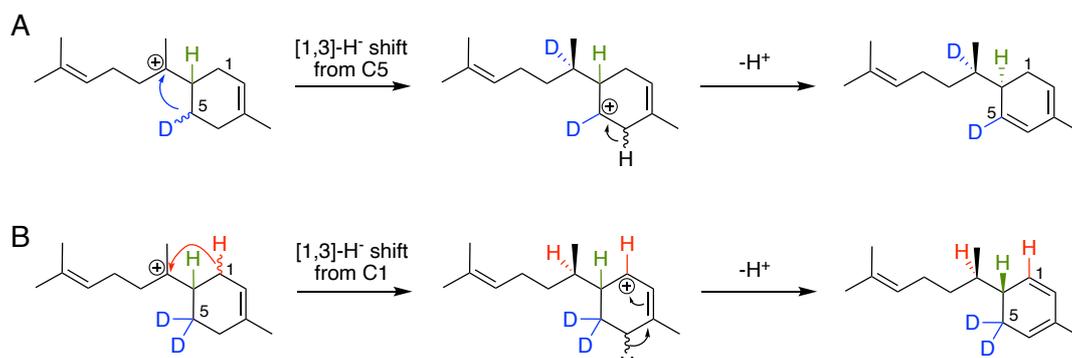


Figure 3.11: Possible reaction pathways of $[^2\text{H}_2\text{-5}]$ -(Z,Z)-FDP from the bisaboly carbocation intermediate given the potential [1,3]-hydride shifts from (A) C5 and B) C1 that lead to the formation of 7-*epi*-zingiberene.

3.3.2 Synthesis of $[^2\text{H}_2\text{-1}]$ -neryl diphosphate

$[^2\text{H}_2\text{-1}]$ -NDP was made from $[^2\text{H}_2\text{-1}]$ -nerol which was in turn made via a simple oxidation and reduction procedure using a deuterated reducing agent, followed by the TEAP diphosphorylation method (Figure 3.12). The devised route involved the oxidation of nerol to its respective aldehyde using MnO_2 followed by reduction with NaBD_4 to introduce the deuteriums at the C1 position.

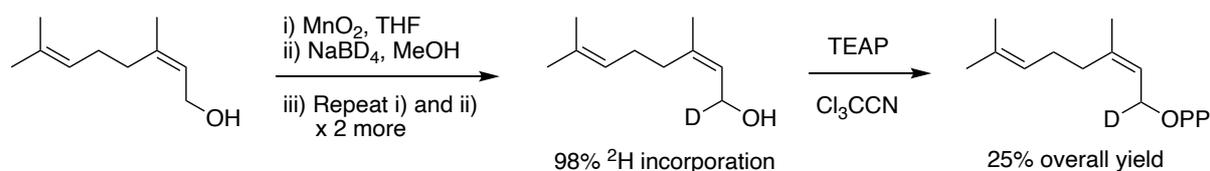


Figure 3.12: Short synthetic route to incorporate deuteriums into the C1 position of the C_{10} substrate $[^2\text{H}_2\text{-1}]$ -NDP.

The oxidation-reduction process was repeated three times to gain a high percentage incorporation of deuterium, as after the first reduction, the incorporation was 50%, going up to around 80% with the second and lastly 98% with the third. This is because when the compound is first oxidised it still retains one hydrogen, and when reduced for the first time with only deuterides provided, a 1:1 ratio of D:H is obtained at the C1. With the next oxidation from this 50% labelled alcohol, oxidation with loss of hydrogen is more favourable than loss of deuterium due to a kinetic isotope effect with the increased mass of deuterium. Therefore, when the aldehyde is reduced with the deuterated reagent, a further increase is seen in the percentage incorporation. At time of synthesis, NaBD₄ was readily available compared to other isotopically labelled reducing agents (such as DIBAL-D), so although these steps had to be repeated to gain the required percentage incorporation, the route was practically simple and high yielding. As before, the percentage of deuterium incorporated was verified by ²H NMR (Section 8.3.5) and by comparison of the corresponding proton environments in the ¹H NMR spectrum of unlabelled nerol (Figure 3.13). In the unlabelled nerol spectrum (black curve) the C1 hydrogens are observed as a doublet at 4.07 ppm, which is essentially absent from the spectrum of [²H₂-1]-nerol (red curve) as the percentage incorporation was found to be 98% ²H.

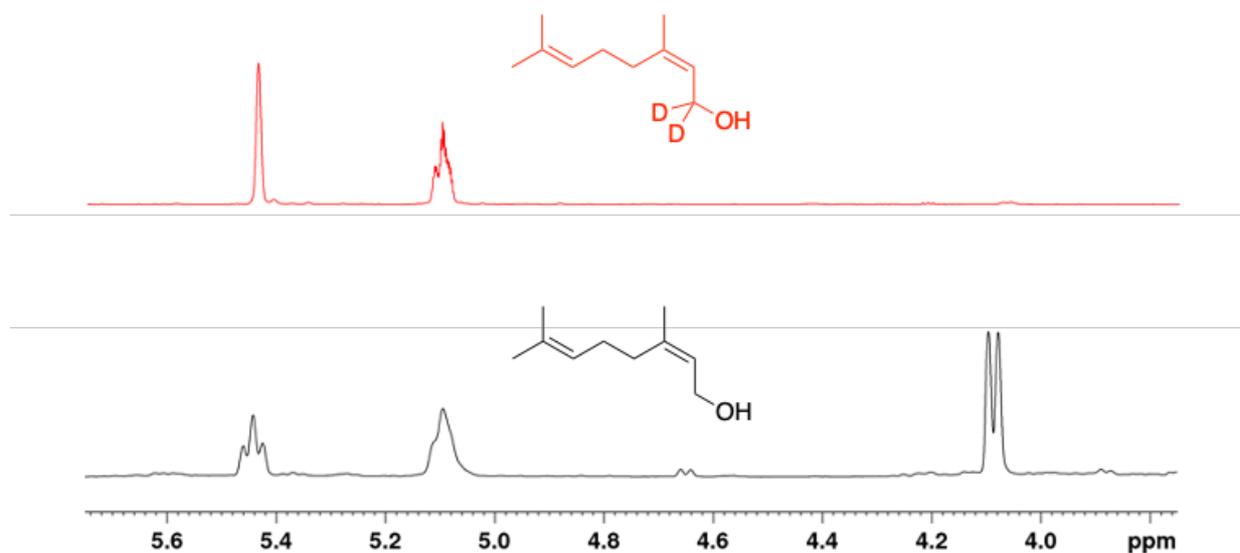


Figure 3.13: Direct comparisons of the ¹H spectra of nerol (black) and [²H₂-1]-nerol (red) between 5.75 - 3.75 ppm. The absence of the peak at 4.07 ppm in the red spectrum relates to the now deuterated environment of the oxidised carbon (98% ²H).

Following synthesis of the labelled alcohol with sufficiently high deuterium incorporation, the TEAP procedure was carried out to generate the ammonium salt of the diphosphate [$^2\text{H}_2$ -1]-NDP as previously described. Excess isopropanol and ammonia were removed under reduced pressure and the remaining aqueous solution thoroughly freeze dried to yield the ammonium salt of the diphosphate as a white powder (15% yield, Section 8.3.5). Results of subsequent analytical characterisation were consistent with the expected product which was deemed sufficiently pure for use in later enzymatic reactions.

3.3.3 Incubation of [$^2\text{H}_2$ -1]-NDP with IDP, zFPS and EZS

With [$^2\text{H}_2$ -1]-NDP produced and IDP available, a small-scale reaction was performed including zFPS and EZS in buffer (100 mM TRIS pH 8.0, 5 mM MgCl_2 , 5 mM DTT), which was left overnight with shaking (300 rpm) at room temperature (Figure 3.14). As before, control reactions were performed simultaneously. The reaction products were extracted into pentane and analysed by GC-MS giving the results shown in Figure 3.15.

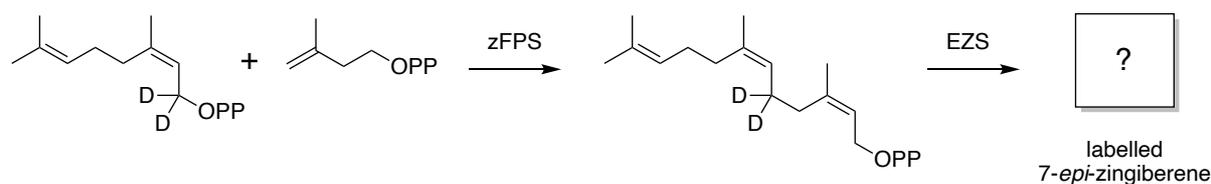


Figure 3.14: The chemoenzymatic preparation of [$^2\text{H}_2$ -5]-(Z,Z)-FDP using the enzyme zFPS and [$^2\text{H}_2$ -1]-NDP with IDP. EZS was used to turn over the *in situ* generated labelled substrate and the pentane extractable product analysed by GC-MS.

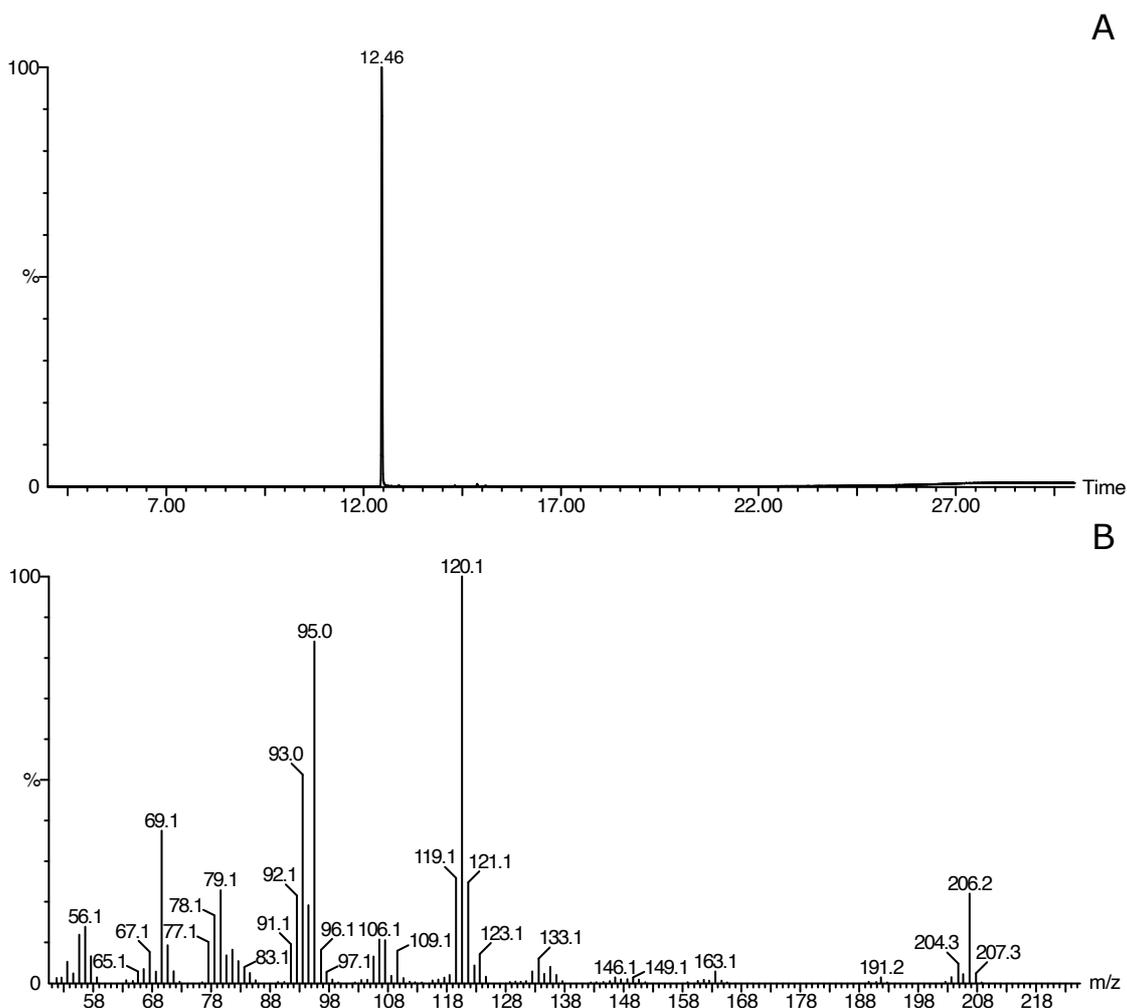


Figure 3.15: (A) Total ion chromatogram of the pentane extractable products resulting from the incubation of [$^2\text{H}_2$ -1]-NDP with IDP, zFPS and EZS. (B) Mass spectrum of the compound eluting at 12.46 min in A.

For this incubation, the labelled product had a retention time of 12.46 min which was consistent with the product of the control reaction performed at the time and identified as 7-*epi*-zingiberene by comparison of the mass spectrum (which is identical to that of Figure 3.7). Small differences to retention times of compounds can often be observed over time as GC-MS equipment undergoes maintenance, hence the difference compared to the retention time of the previously performed incubation. The product of [$^2\text{H}_2$ -5]-(Z,Z)-FDP (Figure 3.15) has the mass fragment ion of $m/z = 206.2$, reflecting the incorporation of two deuteriums into the structure giving a +2 mass number for the parent ion. The ion fragment resulting from bond breaking between C6 and C7 of the labelled product is expected to have an $m/z = 94.08$ if one C5 deuteride shifts, or $m/z = 95.08$ if both remain in position (Figure 3.16). As can be seen

from the mass spectrum of the labelled product (Figure 3.15), the m/z value of this indicative ion compared is now 95.0, consistent with what would be expected if the C5 deuterides were not shifted during the mechanism.

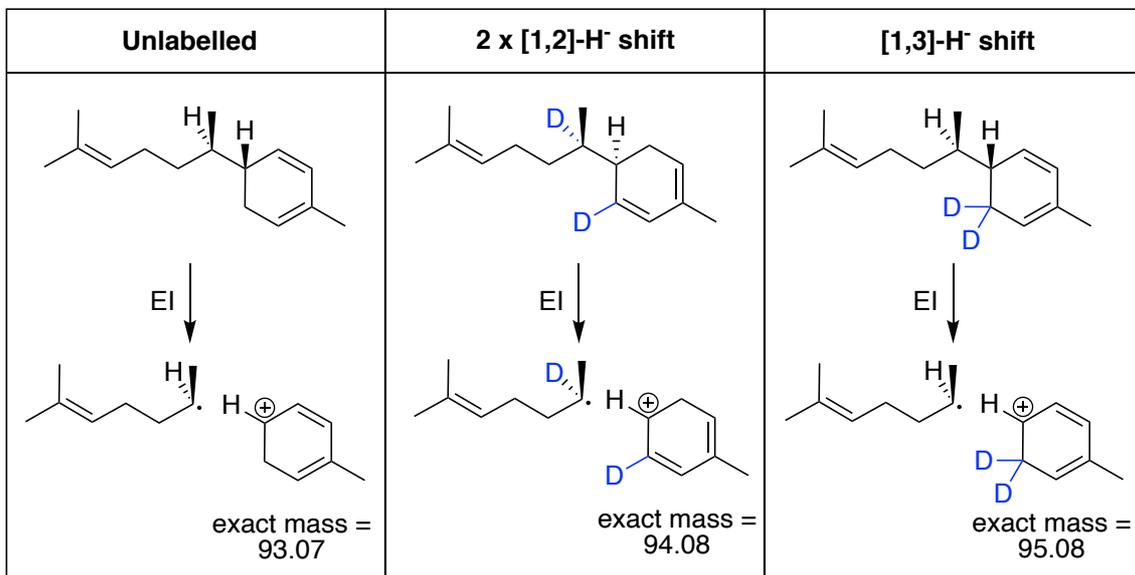


Figure 3.16: Comparison of ions generated by mass spectrometric analysis of 7-*epi*-zingiberene with bond breaking between C6 and C7, generated from both unlabelled substrate and [²H₂-5]-(*Z,Z*)-FDP and considering whether the [1,3]-hydride shift takes place from C1 or C5.

This result is further corroborated by the change in m/z of the other abundant ion. With this substrate, if the bond breaking occurs between C6 and C7, the ion generated would be expected to have $m/z = 121.1$ if the C5 deuterides shift, and $m/z = 120.09$ if they do not (Figure 3.17). The result generated from the labelled product gives a $m/z = 120.1$ which is again consistent with the C5 deuterides not being involved in the mechanism.

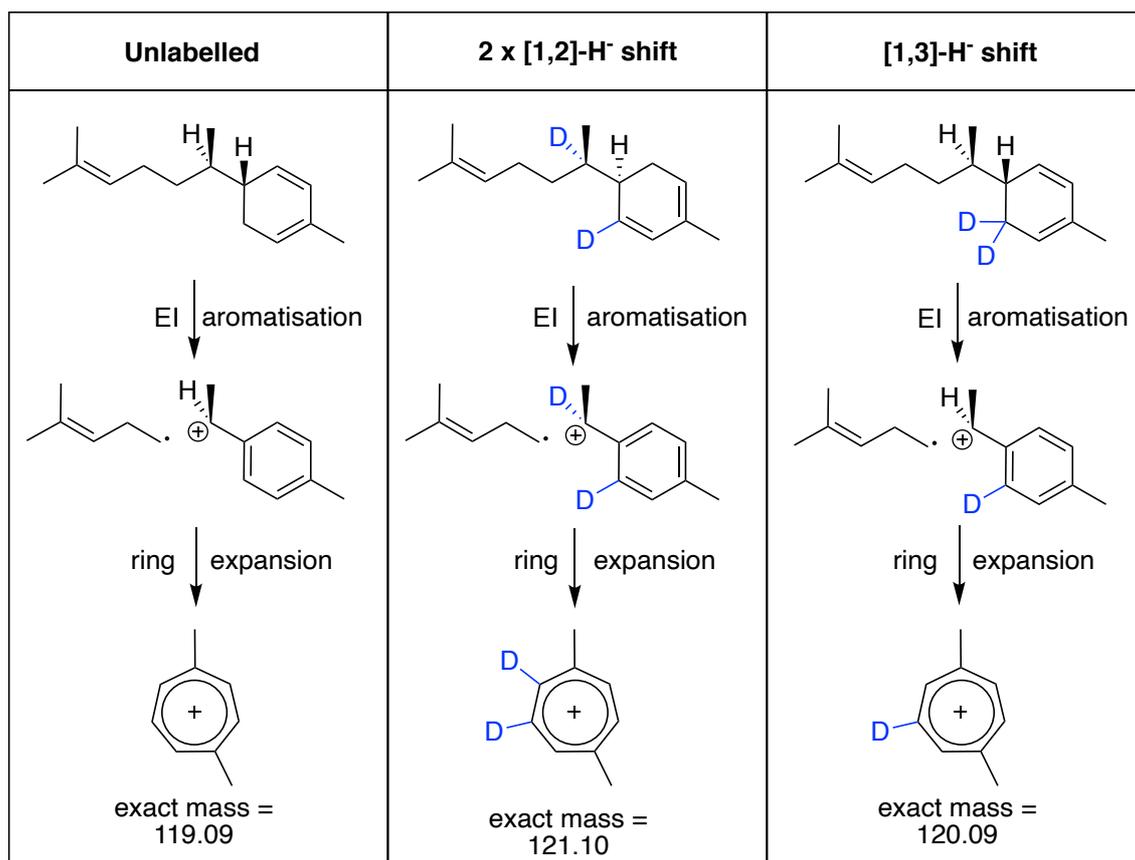


Figure 3.17: Comparison of ions generated by mass spectrometric analysis of 7-*epi*-zingiberene with bond breaking between C7 and C8, generated from both unlabelled substrate and [²H₂-5]-(Z,Z)-FDP and considering whether the [1,3]-hydride shift takes place from C1 or C5.

The results generated from both the [²H₂-5]-(Z,Z)-FDP and [²H₁-6]-(Z,Z)-FDP incubations indicate that the equivalent hydrides at these positions are not likely to be involved in the reaction mechanism to produce 7-*epi*-zingiberene. The catalytic mechanism could therefore be expected to go via the [1,3]-hydride shift from the C1. This outcome may be further rationalised considering the secondary carbocation generated at C1 would be stabilised via delocalisation across the allylic double bond¹⁴ (C2-C3), which isn't possible at the C5 position.

3.4 Preparation of [²H₂-1]-(Z,Z)-FDP

With results indicating that the hydride shifts may take place from C1, (Z,Z)-FDP labelled selectively with deuterium at both the *proR* and *proS* C1 positions ((*R*)-[²H₁-1]-(Z,Z)-FDP and (*S*)-[²H₁-1]-(Z,Z)-FDP) were sought to unequivocally show which hydride shifts (Figure 3.18). A chemoenzymatic approach based on using C1 labelled IDP with the addition of unlabelled NDP in a zFPS catalysed condensation reaction was pursued, deriving the C1 labelled IDPs from their equivalent alcohol precursors.

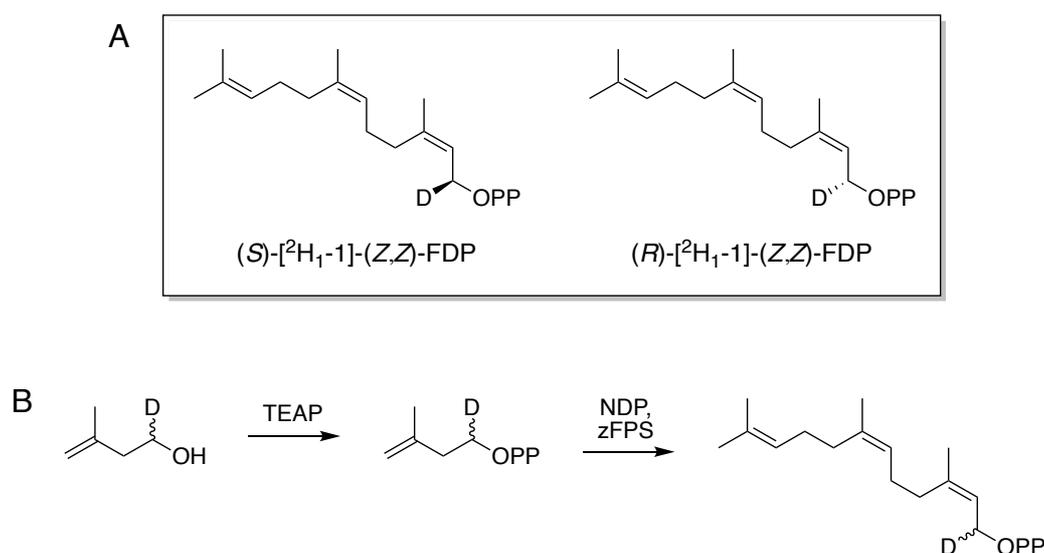


Figure 3.18: (A) Key deuterium labelled substrates that may conclusively establish which hydride is shifted in the mechanism of EZS. (B) The chemoenzymatic route to the C₁₅ substrates via deuterium labelled IDP under investigation. Gaining access to the labelled low molecular weight alcohol precursors may however be prohibitively challenging.

Several synthetic routes to deuterium labelled isoprenol were explored although isoprenol and other similar low molecular weight compounds have proven exceptionally challenging to work with. Initially, a simple oxidation-reduction approach was attempted (comparable to the C1 labelling of nerol discussed in Figure 3.12), where isoprenol was oxidised to isoprenal and subsequently reduced. Different oxidising agents, including MnO₂, Dess-Martin periodinane (DMP) and pyridinium chlorochromate (PCC), were trialled with no success. Crude isoprenal was only detected when 2-iodoxybenzoic acid (IBX) was used, but it was found to be extremely unstable and not amenable to handling or purification using standard techniques. The literature around this type of chemistry is very limited: only one example was found at the time

with routes to the desired (*R*)-[²H₁₋₁]-IDP and (*S*)-[²H₁₋₁]-IDP,¹⁵ and a single further example with a synthesis for [²H₂₋₁]-IDP.¹⁷⁶ The former route involves LiAlD₄ as the deuterating reagent to reduce an ester precursor but was unavailable to purchase whilst this project was taking place and as a result this chemoenzymatic approach was not pursued. Considering these challenges, a total synthesis approach may be the most feasible for obtaining the C1 labelled (*Z,Z*)-FDPs and is suggested for future work.

3.5 Summary

The reaction mechanism of the unusual sesquiterpene synthase, EZS, that selectively converts (*Z,Z*)-FDP into 7-*epi*-zingiberene, has been investigated with a focus on understanding the nature of various hydride shifts as they quench carbocationic intermediates. New chemoenzymatic routes have successfully been developed to produce the deuterium labelled substrates, [²H₂₋₅]-(*Z,Z*)-FDP and [²H₁₋₆]-(*Z,Z*)-FDP. This involved chemically synthesising labelled C₁₀ precursors using short, relatively straightforward, and adequately yielding syntheses. The final asymmetric chain extension step was carried out entirely selectively for the difficult to produce *Z* isomer by using the prenyltransferase zFPS, generating the C₁₅ labelled substrates for EZS. Despite a majority of approaches using established procedures for total synthesis of full-length labelled substrates,⁵⁴ there are several examples where prenyltransferases have similarly been used to make labelled (*E,E*)-FDPs for mechanistic studies from shorter chain length isotopologues.^{79,177–179} This chemoenzymatic methodology has similarly provided fast access to these key substrates, revealing which should be targeted for future work to complete our understanding of the catalytic mechanism of EZS.

Analysis of the mass spectra of the labelled products revealed that the deuterides of both C6 and C5 of the substrate do not move to C7 of the product, as the indicative ion fragments have *m/z* values consistent with no deuterium incorporation. MS is a highly sensitive technique and has been sufficient to analyse 7-*epi*-zingiberene isotopologues in this way given its fragmentation chemistry. The evidence suggests that the mechanism takes place via a [1,3]-hydride shift from C1: both (*R*)-[²H₁₋₁]-(*Z,Z*)-FDP and (*S*)-[²H₁₋₁]-(*Z,Z*)-FDP could be used to unequivocally prove this hypothesis and the stereochemical requisites of this step. Lastly, to complete the picture, (*R*)-[²H₁₋₄]-(*Z,Z*)-FDP and (*S*)-[²H₁₋₄]-(*Z,Z*)-FDP would show specifically which proton is lost in the final mechanistic step. Although chemoenzymatic synthetic routes are suitable for some substrates, challenges faced in dealing with low molecular weight precursors mean total synthetic approaches may be more suitable for producing these final substrates.

**Structural Determinants of EZS
Substrate Selectivity and Product
Specificity**

4.1 Introduction

Terpene synthases behave as stereochemical templates, folding their flexible substrates into specific reactive conformations that precede the multistep reaction cascades leading to product formation.¹ Their active sites are characteristically hydrophobic with conserved metal-binding residues that enable substrate ionisation and typically contain aromatic residues which stabilise carbocation intermediates and facilitate acid/base catalysis.⁵⁵ Over 300 sesquiterpene skeletons can be accessed from the single precursor farnesyl diphosphate (FDP), yet sesquiterpene synthases can navigate the energetically similar reaction pathways and common intermediates with extremely high fidelity.⁷² Whilst isotopic labelling studies are used to investigate the bond breaking and making steps in the conversion of a substrate to product,⁷⁸ rationally altering the primary sequence can reveal which amino acids directly or indirectly facilitate these reaction steps. Mutagenesis has been used to demonstrate key plasticity residues,^{57,180} change substrate specificities¹⁰⁸ and even improve catalytic efficiencies.⁶⁴

4.1.1 Aims

This chapter describes the application of the site-directed mutagenesis technique to EZS with the aim of elucidating which amino acids facilitate formation of 7-*epi*-zingiberene as the sole enzymatic product from (Z,Z)-FDP. Residues belonging to the conserved α -domain metal-binding motifs (DDXXE and NSE) have been identified and shown to be critical for catalysis;¹⁵³ however little else is known about how the structure of EZS influences the other mechanistic steps (ring closure, hydride shifts, deprotonation) whilst quenching of multiple carbocation intermediates is avoided. Interestingly, EZS shares 91% sequence identity to the closely related SBS, which is conversely a multi-product enzyme, despite only a small number of residues differing between the two active sites.³⁴ Hypothesised to be at least partly responsible for the loss of product specificity with SBS, these residues were taken into consideration when using homology modelling to select a variety of amino acids to mutate. Residues with side chains directed into the EZS active site which may interact with the substrate were targeted in this work, with subtle changes made to alter the pocket shape and volume.

This work also aimed to investigate the structural determinants of the specificity of EZS for (Z,Z)-FDP over the universally adopted isomer, (E,E)-FDP: exactly how or why *cis*-prenyl diphosphate utilising terpene synthases exclusively evolved in tomato has remained unclear. However, a recent study¹⁸¹ looking at monoterpene synthases with C₁₀ *trans*-substrate, GDP, selectivity reported the identification of a key residue which behaved like an 'isomeric switch':

mutation of it improved the ability of the synthases to accept the non-canonical C₁₀ *cis*-substrate neryl diphosphate (NDP), over geranyl diphosphate (GDP). Kinetic characterisation data implied the shift in substrate specificity was caused by alterations to the active site geometry, rather than disrupting a direct catalytic role. To investigate whether such a residue may analogously be responsible for EZS selectivity, both all *trans* and *cis* C₁₅ substrates were incubated with the mutants generated in this study, leading to subsequent analysis of how substrate specificity may arise.

4.2 Selection of residues

As no crystal structure of EZS is available, a homology model based on the crystal structure of α -bisabolene synthase was generated by Dr. Prabhakar Srivastava using SWISS-MODEL¹⁸² and visualised using PyMOL molecular graphics software¹⁸³ (Figure 4.1). Like EZS, α -bisabolene synthase from *Abies grandis* is an unusual three-domain plant sesquiterpene synthase that has a functional class I (α -domain) active site, but a non-functional class II active site (interface of β - and γ -domains) associated with diterpene catalysis.¹⁸⁴

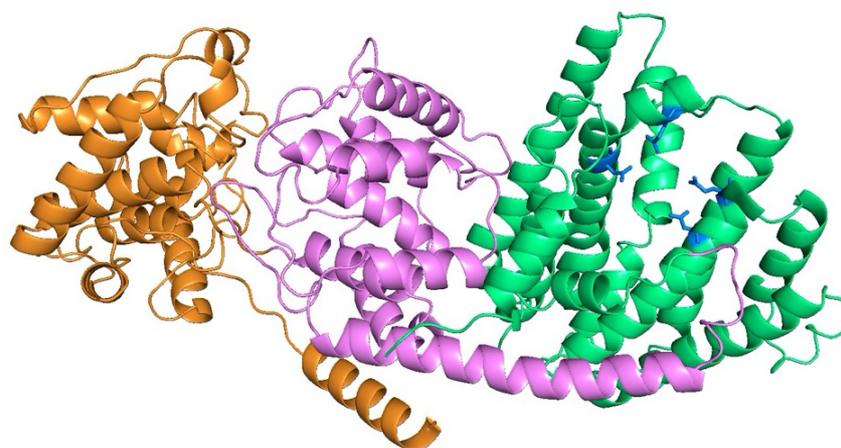


Figure 4.1: Cartoon homology model of EZS based on the crystal structure of α -bisabolene synthase. The catalytic α -domain is shown in green with the side chains of the metal-binding residues of the DDHFE and NSE motifs in the entrance to the active site pocket shown in blue. The β -domain is shown in pink and γ -domain in orange.

Using the PyMOL model and annotations of the α -bisabolene crystal structure, Figure 4.2 was created with comparison to the sequence of SBS to identify precisely which amino acids create the active site contour and other features. The amino acids mutated in this work are shown in red lettering, and where side chains point into the active site, the position is highlighted in grey.

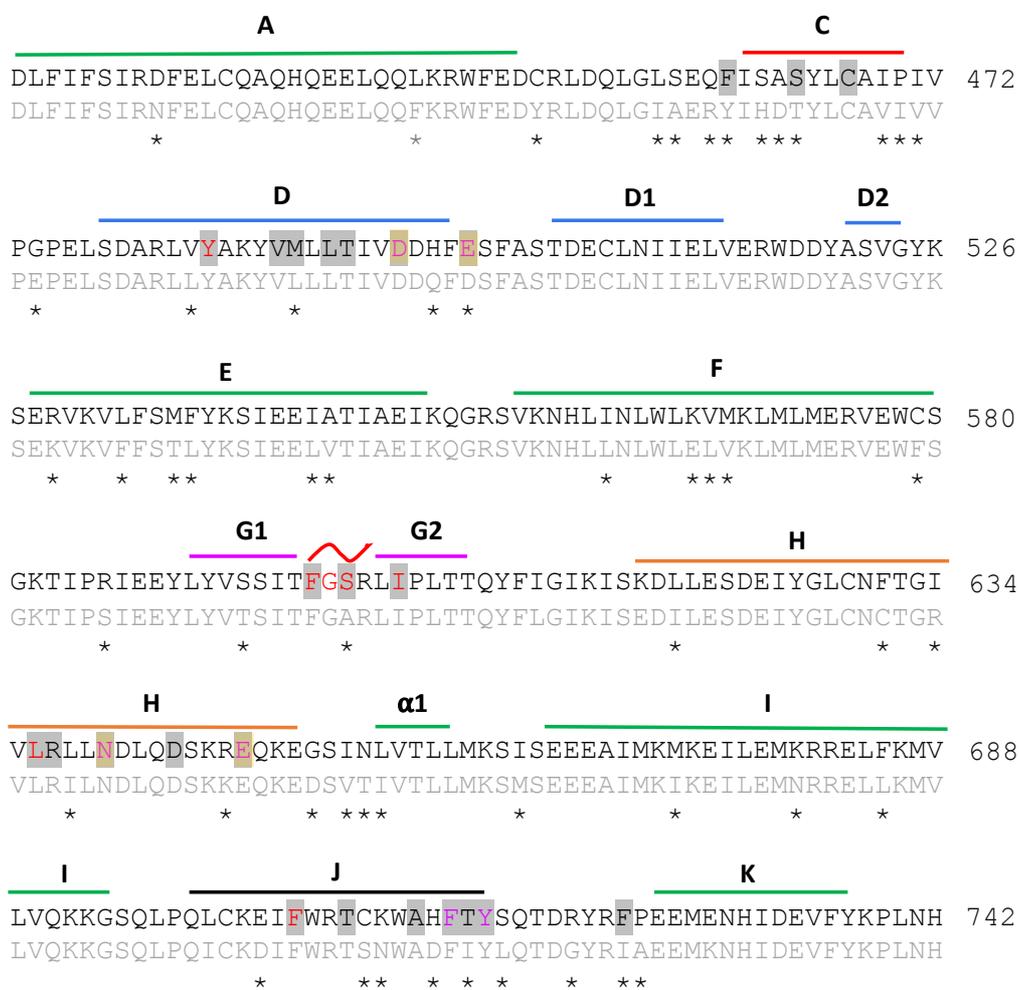


Figure 4.2: Alignment of the structural elements of the catalytically active α -domain of EZS (black lettering) and SBS (grey lettering). The individual α -helices are annotated with the coloured bars and the G-helix break with the red curved line. Amino acids which contour the active site are highlighted in grey and those additionally involved in Mg^{2+} binding highlighted in yellow. Where the sequences differ in amino acid identity the position is highlighted with an asterisk. Amino acids mutated in this work are indicated by red lettering and pink lettering for previous analysis.¹⁵³

4.2.1 G-helix residues

The G-helix is part of the central reaction chamber, highly conserved across many sesquiterpene synthases and shown to change between open and closed conformations upon substrate binding.¹⁸⁵ The helix has two parts (G1/2) connected by a kink region of four amino acids which gives a short break in the α -helical structure and is attributed to enabling helix movement and substrate positioning. This region has been shown in other synthases to be important for catalysis as substitutions here have caused changes to product profile distributions and catalytic efficiencies.^{65,85,186} In EZS, amino acids F598, G599, S600 and R601 form the kink region of helix G and from the model it can be seen, with the exception of R601, that their side chains point into the active site, along with I603 (Figure 4.3). F598, S600 and I603 were initially substituted with A to reduce the size of their side chains and increase flexibility of the G-helix kink. G599C and S600A were selected as the equivalent position in this conserved region of Z-selective enzymes is either G or C, or S or A (Table 4.1).

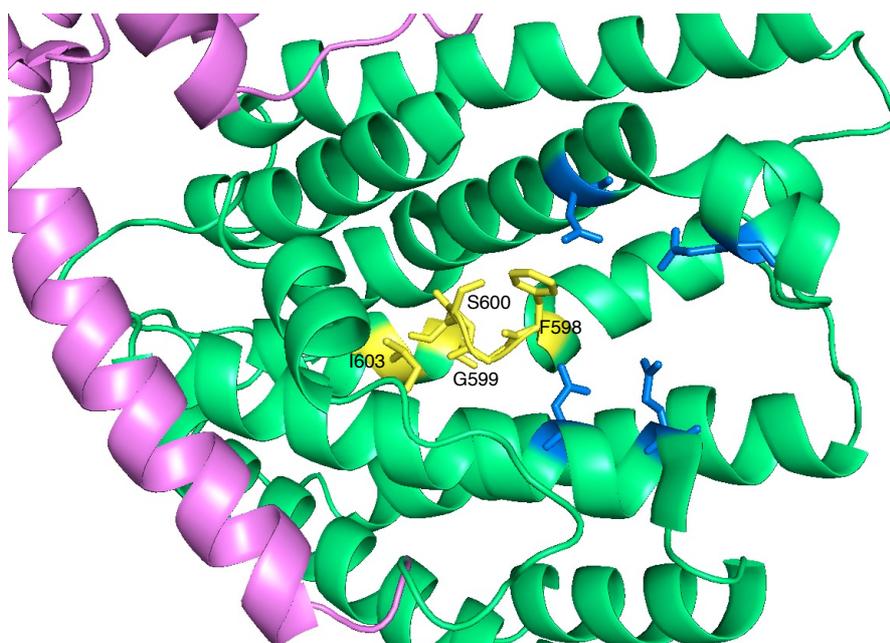


Figure 4.3: Close-up of EZS active site where the G-helix kink residues targeted for mutation in this study have been highlighted in yellow, seen with side chains pointing into the active site. Metal-binding residues are shown in blue.

4.2.2 Other residues of interest

The role of aromatic active site residues is well documented in terpene synthase literature.¹ Near the top of the active site pocket of EZS are residues F714 and Y716. These were both mutated to A in previous work and, whilst the former resulted in product, the latter showed loss of activity. Product formation was maintained after subsequent mutation of Y716 to F, indicating that the size and aromaticity of the side chain is important for catalysis, whereas that of F714 plays a lesser role. To continue investigating, other amino acids contouring the active site were targeted (Figure 4.4). As well as aromatic G-helix F598, these included aromatic Y484 of the D-helix and F705 of the J-helix, which were both substituted for W to retain the side chain hydrophobicity and aromaticity but alter the steric bulk. H-helix L636 was also targeted and substituted with A as when the equivalent mutation was made in SBS, a change to the product profile was observed with an increase in α -santalene formation (Prof. A. Tissier, personal communication).

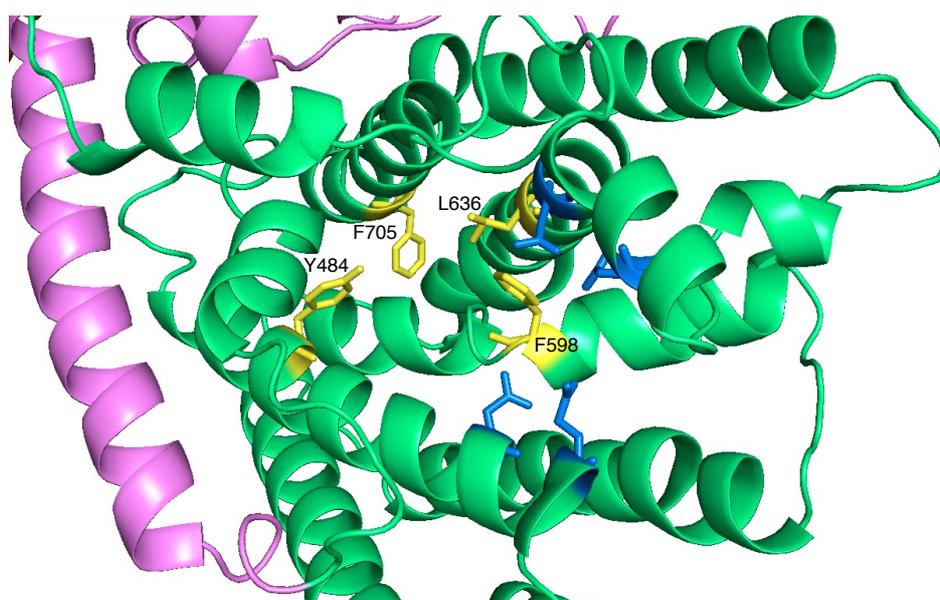


Figure 4.4: Close up of EZS active site where residues targeted for mutation in this study with side chains that point into the active site have been highlighted in yellow. Metal-binding residues are shown in blue.

4.3 Cloning and expression of EZS mutants

A combination of strategies were used to generate point mutations in the EZS gene. The SDM approach uses primers with overlapping parts to generate PCR products which contain the mutation.¹⁸⁷ The PCR products are directly transformed into competent cells for DNA amplification and extraction; however due to the PCR product being linear and unligated, the process can result in low transformation efficiencies. Additionally, requiring relatively long primers, sometimes their properties can be suboptimal, i.e., have secondary structures, making the PCR step challenging. For these reasons, another approach called KLD (kinases, ligase, DpnI) cloning was also used which involves small primers and a short ligation step where the linear, blunt-ended PCR product is phosphorylated, ligated, and the template DNA is digested, in a one-pot 15 min incubation before transformation. Using shorter primers in this approach can however lead to non-specific binding with the template, sometimes problematic as the gene encoding EZS is large. The method chosen was therefore a judgment of which would give primer pairs with the most optimal properties for that region of DNA.

The primers used to generate site-specific mutants of EZS and the cloning technique used can be found in Section 8.1.6 along with the experimental details of the PCR reactions and both assembly techniques. Where necessary, PCR conditions were optimised by changing the annealing temperature and/or template and primer concentrations. PCR products were either assembled and then transformed (KLD) or transformed directly (SDM). Individual colonies were selected for an overnight culture from which the DNA was extracted and sequenced. Once the codon change in the wild-type sequence had been confirmed, the plasmid was transformed into competent C41(DE3) cells and the protein expressed and purified as per the established protocol for EZS (Section 8.1.10).

To characterise each mutant, an assay was carried out using both (*Z,Z*)- and (*E,E*)-FDPs to assess any change to the product profile and establish whether substrate specificity had been affected by the mutations. Substrate (100 μ M) and purified enzyme (approximately 200 ng) were incubated in reaction buffer (0.5 mL total volume; 100 mM TRIS pH 8.0, 5 mM MgCl₂) overnight at room temperature. The organic reaction products were then extracted with pentane and analysed by GC-MS.

4.4 Product specificity

GC-MS analysis of the (Z,Z)-FDP incubations with mutants Y484W, S600A, I603A, L636A and F705W showed no discernible changes to the product profiles in each case, as strong major peaks for 7-*epi*-zingiberene eluting at 12.46 min in the chromatograms were detected, with associated mass spectra identical to the control (Supplementary Figure 4.9). However, the outcomes for both F598A and G599C proved to have diverged from the high-fidelity wild type as multiple enzymatic products were detected from their incubations with (Z,Z)-FDP (Figure 4.6). Homology modelling shows F598 is situated in the middle of the active site with its aromatic side chain pointing directly into the pocket, positioned precisely where the substrate may be expected to fold to enable the 1,6-cyclisation (Figure 4.5).

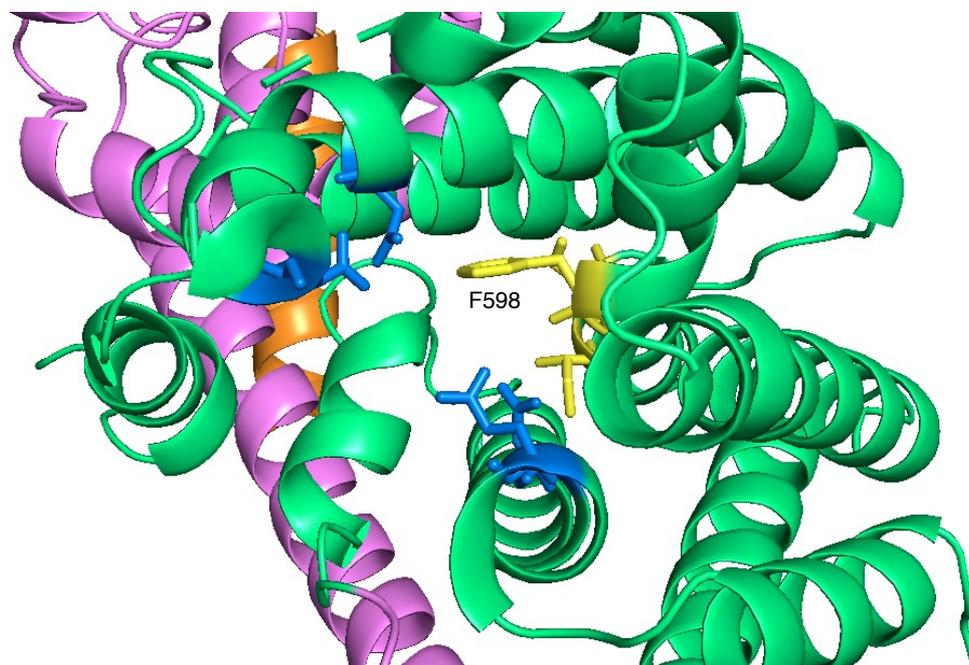


Figure 4.5: Cross section of the catalytic α -domain of EZS where the aromatic side chain of F598 (yellow) can be seen pointing directly into the middle of the active site pocket. Metal-binding residues in blue.

With this result in mind, amino acids with bigger and smaller side chains were chosen to replace F598 to gauge whether sterics, electronics or both may be responsible for maintenance of product specificity. Mutants F598L and F598W were successfully created as well as the double mutant F598I-I603A. The GC-MS product profiles of the enzymatic products of each of these mutants compared to the wild type are also given in Figure 4.6.

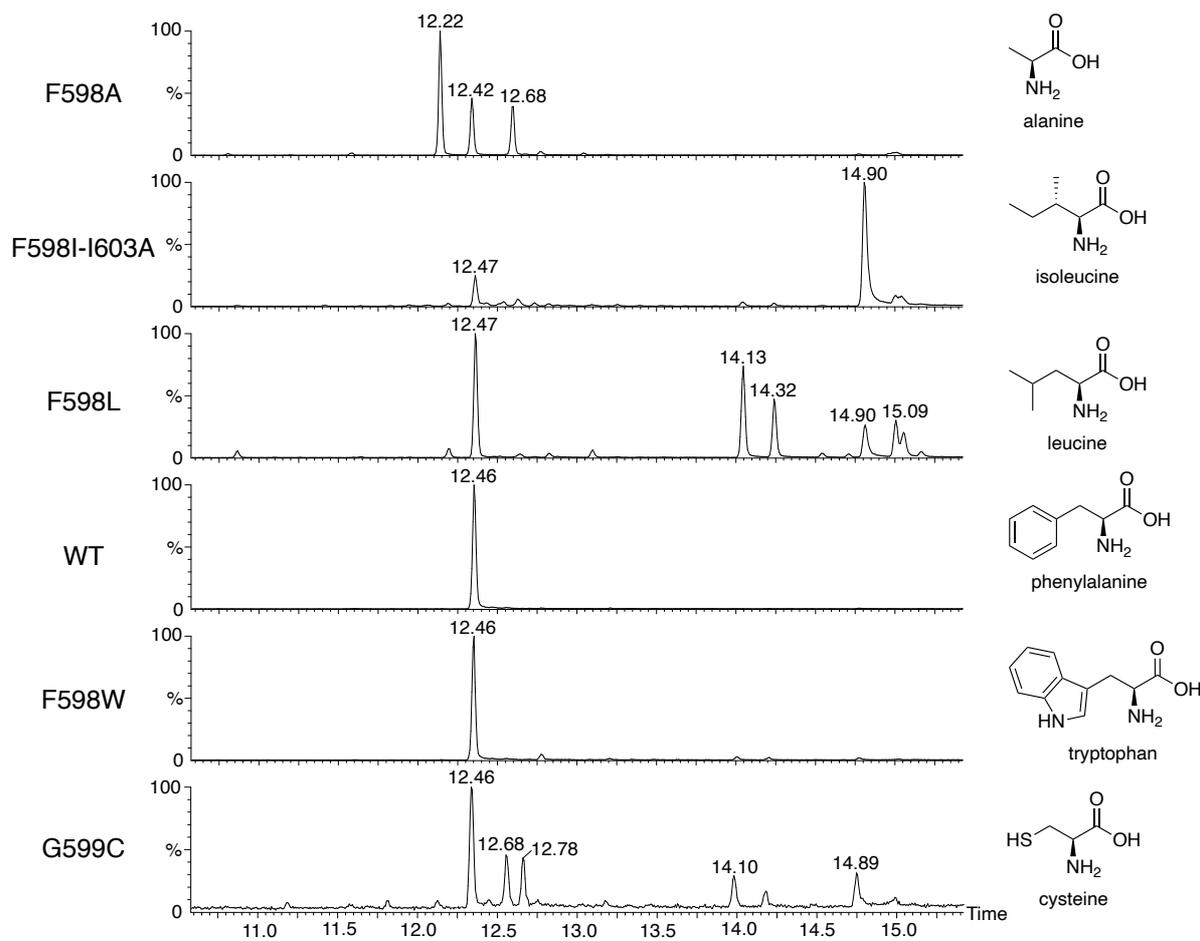


Figure 4.6: Aligned total ion chromatograms of pentane extractable products of the incubations of F598 and G599 mutants with wild type (WT) EZS arising from (Z,Z)-FDP.

For F598A, several products can be seen with retention times around 12 min which differ from 7-*epi*-zingiberene at 12.46 - 12.47 min. The compound eluting at 12.22 min was tentatively identified as β -curcumene as it has good agreement with the NIST mass spectral database and is structurally very close to 7-*epi*-zingiberene, differing only with the position of one of the ring double bonds. The identity of the other compounds is not clear but the ability to form 7-*epi*-zingiberene appears to have been lost with this relatively big change to the side chain.

The major compound identified from the F598I-I603A mutant incubation was (Z,Z)-farnesol at 14.90 min (Figure 4.6), confirmed by comparison to an authentic standard. Replacement of F598 and I603 with smaller residues could result in an increase in the active site volume and a reduction in hydrophobic interaction required for initial cyclisation, thus resulting in the formation of farnesol by premature water capture and quenching the ionised farnesyl cation. Although I603A mutant alone did not change the product profile (Supplementary Figure 4.9),

the double mutation will likely have compounded the effect. Unpublished work from the group of Prof. Alain Tissier further supports these results with the observation of a significant product profile change following mutation of the equivalent F residue in SBS to I, finding the major product was bisabolene. Cloning for the single F598I mutant was not achieved within the time constraints of this project but assay results of this mutant would be necessary to clarify this point.

The F598 change to L has also had a significant impact on the product profile as, in addition to the major peak of 7-*epi*-zingiberene at 12.47 min, multiple significant products with retention times consistent with hydroxylated products (greater than around 14 min) can be detected (Figure 4.6). In the mass spectra of the peaks at 14.13 and 14.32 min, mass ions at $m/z = 222.3$ were detected; indicative of cyclic sesquiterpene alcohols by molecular weight (Supplementary Figure 4.10). Typically these ions are difficult to detect as they rapidly fragment to leave the mass ion equivalent to $[M - H_2O]^+$, although their presence may suggest these unknown compounds are cyclic secondary or tertiary alcohols¹⁷⁴ such as bisabolol (of which there are a variety of isomers). Additional analysis by TLC verified that the sample contained several relatively polar products (Supplementary Figure 4.11). The profile of neighbouring mutant G599C reflects a similar change with the appearance of relatively small proportions of alcohols at the same retention times. Additional cyclic sesquiterpenes are observed eluting after 12.46 min; the compound at 12.68 min was tentatively identified as β -bisabolene. Apart from this, 7-*epi*-zingiberene and (*Z,Z*)-farnesol, it was not possible to further identify the new compounds from retention time and National Institute of Standards and Technology (NIST) mass spectral database comparisons alone, therefore further structural characterisation would be needed to understand the full product profile.

These results show the G-helix region is implicated in preventing the ingress of water into the active site for EZS catalysis. This has been observed in work with other synthases: mutagenesis of the G1/2 helix break in δ -cadinene synthase gave germacradien-4-ol as the major product (93% conversion), with the new function attributed to exposing the active site to a potential water channel.^{186,188} This switch has also been reported in reverse with the mutation of germacradien-11-ol synthase to produce a majority of a non-hydroxylated product, isolepidozene, via targeting the G-helix. Molecular dynamic simulations revealed that water could flow through the active site for the wild type enzyme.⁸⁵

The variety of substitutions at the F598 position have identified its role in mediating active site plasticity and controlling product diversity. When the active site volume is increased, and potentially the freedom of movement of the G-helix kink, multiple products are seen as various reaction pathways of similar energies can now be accessed. For wild type EZS, it may be inferred that F provides additional stabilisation to carbocationic intermediates through the

charge-quadrupole interaction provided by the aromatic ring, as whilst the F598W mutation also gave essentially the same product outcome (Figure 4.6), mutants without aromatic side chains show the loss of specificity.

4.5 Substrate selectivity

EZS is one of a very limited number of enzymes, unique to tomato plants, that accept short chain *cisoid* isoprenoid precursors (C₁₀ - C₂₀). To investigate the structural determinants for isomeric substrate selectivity, all EZS mutants generated in this study were incubated with (*E,E*)-FDP.

For the most part, GC-MS analysis of the enzymatic assays showed no organic products. Small relative amounts of linear farnesol and farnesene products were observed from wild type EZS and mutants F598W and F598A (Supplementary Figure 4.12), in greater abundance than control assays using basally expressed cell line proteins. As mentioned previously, conversion of FDP to farnesol can be observed when the enzyme binds FDP but only cleaves the diphosphate group, and the farnesyl cation is quenched by water. When low levels are detected, however, these compounds should not necessarily be attributed to significant enzymatic activity as the substrates can also hydrolyse to small extents when hydrophobic enzymes are present. Although in this work there are several compounds present in the wild type chromatogram, (*E,E*)-FDP was previously reported to be an inhibitor of EZS with no incubation products detected.^{137,153}

Interestingly, there were several enzymatically generated compounds from (*E,E*)-FDP with the double mutant, F598I-I603A (Figure 4.7) in addition to linear (*E,E*)-farnesol (15.54 min). The chromatogram has been aligned with control incubations to justify compound identification: comparison of the retention time and mass spectra (Supplementary Figure 4.13) to the products of amorphadiene synthase (ADS) and *epi*-cubenol synthase (EPICS) indicate that the compound at 12.25 min is amorphadiene, 13.66 min is germacradien-4-ol and 14.40 min is *epi*-cubenol (although absolute stereochemistry is not determined).

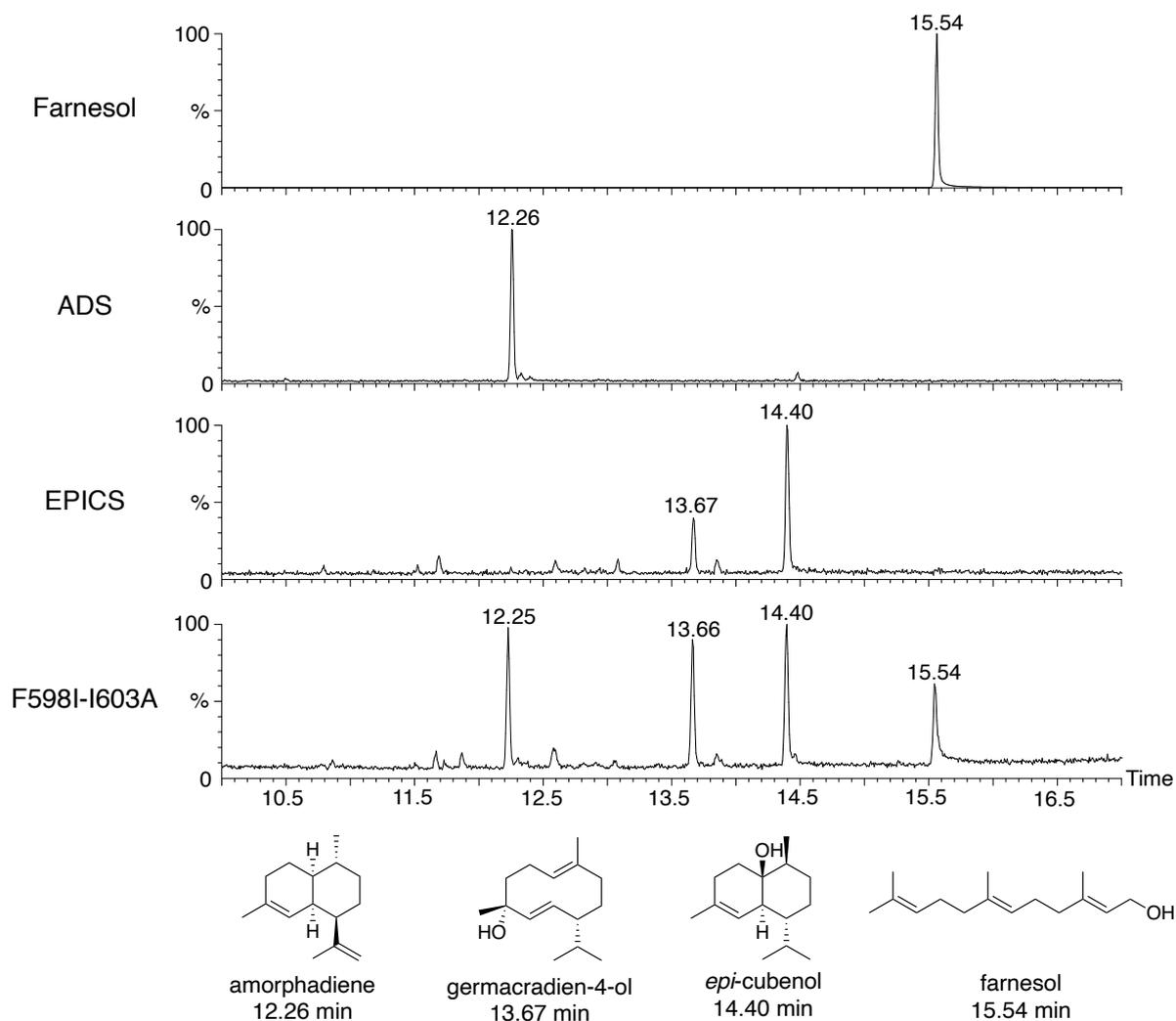


Figure 4.7: Aligned total ion chromatograms of the pentane extractable products from incubation of (*E,E*)-FDP with double EZS mutant F598I-I603A and controls for farnesol at 15.54 min, amorphadiene at 12.26 min (generated from ADS), germacradien-4-ol at 13.67 min and *epi*-cubenol at 14.40 min (generated from EPICS).

The enzymatic reaction mechanisms of both EPICS¹⁸⁹ and ADS⁸¹ with (*E,E*)-FDP proceed after an initial ionisation-isomerisation-recombination step generating nerolidyl diphosphate, the common *cisoid* isomer of (*E,E*)-FDP and sesquiterpene synthase reaction intermediate (Figure 4.8). Similar to the formation of 7-*epi*-zingiberene, amorphadiene is accessed via a 1,6-cyclisation of the substrate to the bisabolylyl cation which then undergoes hydride rearrangements, alternatively followed by a second ring closure to yield the bicyclic final structure. The other two products instead form via a 1,10-cyclisation from nerolidyl diphosphate to yield the germacradienyl cation, which can be quenched before or after various hydride rearrangements to yield germacradien-4-ol and *epi*-cubenol respectively. Chemistry similar to these canonical terpene synthase mechanisms must therefore be facilitated by the

EZS mutant which potentially maintains enough active site character to fold (*E,E*)-FDP in a comparable *cisoid* conformation that is sufficiently stable for product formation, either directly or following isomerisation to a nerolidyl diphosphate-like isomer.

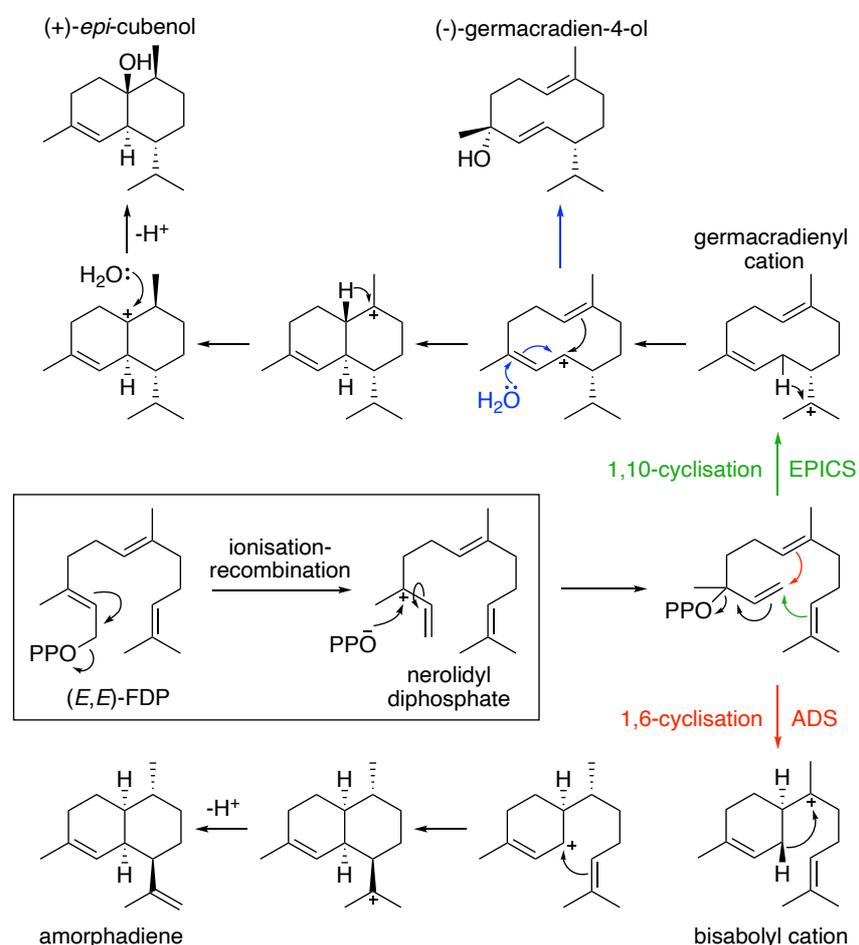


Figure 4.8: Catalytic reaction mechanisms for the formation of *epi*-cubanol and germacradien-4-ol by EPICS¹⁸⁹ and amorphadiene by ADS.⁸¹ Both reactions initially proceed from (*E,E*)-FDP via an ionisation-recombination step that forms the *cisoid* intermediate nerolidyl diphosphate.

Considering that with (*Z,Z*)-FDP the major product is (*Z,Z*)-farnesol, these results could imply that (*E,E*)-FDP is now a better substrate for the double mutant, representing a switch in substrate specificity. As previously mentioned, a comparable observation was made in a 2019 study by Ignea *et al.*¹⁸¹ which set out to engineer monoterpene synthases to accept *cisoid* C₁₀ NDP instead of GDP for improving *in vivo* monoterpene production titres. GDP accepting monoterpene synthases in the study already showed a natural affinity for NDP, largely generating the same product profile when tested, but to improve catalytic efficiencies for the non-canonical substrate, mutants of the model *Salvia fruticosa* 1,8-cineole synthase (*SfCinS1*)

were generated and screened. Residue F547 of SfCinS1 in particular, and corresponding residue of other synthases, was discovered to behave as an isomeric switch: variants of this position showed increased preference for NDP and improved efficiencies with the orthogonal pathway. Although F571 is not located in the equivalent G-helix kink of monoterpene synthases (break between helices α 18 and α 19) but in a different loop,⁶⁵ modelling showed it was positioned in the active site adjacent to where the substrate undergoes *trans* to *cis* inversion during the reaction mechanism (analogous to nerolidyl diphosphate formation, Figure 4.8). The change in apparent K_M values showed improved efficiencies and selectivities with NDP leading to the conclusion that the residue exerts stereochemical control through its size and orientation, as alterations to residue shape disfavoured binding of GDP.¹⁸¹

This finding may not translate as straightforwardly to sesquiterpene synthases, however. When EZS and SBS were first discovered in tomato, the occurrence of short chain *cisoid* substrates was not known elsewhere, but their wider utilisation has since been realised with the full characterisation of the terpene synthases in tomato (*Solanum lycopersicum*) in a recent publication.¹⁴⁴ Through comprehensive testing with C₁₀, C₁₅ and C₂₀ substrates of all *cis* and all *trans* geometries, Zhou *et al.* showed a series of canonical sesquiterpene synthases in the TPS-a gene subfamily had some affinity for (Z,Z)-FDP and produced mixtures of sesquiterpenes, some of which were identified as bisabolene derivatives.¹⁴⁴ Their detailed analysis also revealed two previously uncharacterised sesquiterpene synthases of this clade that are completely selective for (Z,Z)-FDP: cytosol localised TPS10 and mitochondria localised TPS36.¹⁴⁴ What is most interesting is that the Z-selectivity of these TPS-a clade synthases is likely to have evolved entirely separately from the previously known Z-selective TPS-e/f synthases (Table 4.1), which include plastid localised EZS and SBS. These two clades are only distantly related with the TPS-e/f clade diverging early on from diterpene genes involved in primary metabolism whilst the TPS-a evolved separately for secondary metabolism.¹⁴⁰ A comparison of the amino acid conservation across the key G-helix region of Z-selective enzymes is given in Table 4.1, created using the Clustal Omega online webtool (with full sequence alignment given in Supplementary Figure 4.14).

Table 4.1: Comparison of the G-helix region primary sequence of short chain *cisoid* substrate specific enzymes with shading highlighting conserved residues. The residues changed in the (E,E)-FDP accepting double EZS mutant are highlighted with red lettering. NDP = C₁₀ and NNDP = C₂₀ all *cis* isoprenoid diphosphate substrates.

Species	TPS clade	Common name	Preferred substrate	Major product/s	G1	Kink	G2
<i>S. habrochaites</i>	e/f	LMS	NPP	limonene	I E E Y L Y V T S V T	F G S K	L L P L
<i>S. habrochaites</i>	e/f	PIS	NPP	α- & β-pinene	I E E Y L Y V T S I T	F C A K	L I P L
<i>S. lycopersicum</i>	e/f	PHS1	NPP	β-phellandrene	I E E Y L Y V T S I T	F C A K	L I P L
<i>S. pennellii</i>	e/f	TPS20	NPP	α-phellandrene	I E E Y L C V T S I T	F G S R	L L L L
<i>S. lycopersicum</i>	e/f	TPS19	NPP	β-myrcene & β-omicene	I E E Y L Y V T S I T	F C A K	L I P L
<i>S. habrochaites</i>	e/f	EZS	(Z,Z)-FPP	7- <i>epi</i> -zingiberene	I E E Y L Y V S S I T	F G S R	L I P L
<i>S. habrochaites</i>	e/f	SBS	(Z,Z)-FPP	santalene & bergamotene	I E E Y L Y V T S I T	F G A R	L I P L
<i>S. lycopersicum</i>	e/f	TPS21	NNPP	lycosantalene	I E E Y L Y V T S I T	F G S R	L I P L
<i>S. lycopersicum</i>	a	TPS10	(Z,Z)-FPP	α-bisabolol	Y E E H M E I S L V T	A G Y M	M - - G
<i>S. lycopersicum</i>	a	TPS36	(Z,Z)-FPP	cis-muurola-3,5-diene	N C D Y I K N A I V S	T T F M	A - - L

The closely related TPS-e/f enzymes share high sequence identity which is mostly conserved in the G-helix kink region. TPS10 and TPS36 however share little sequence identity in this region to one another, or to the TPS-e/f synthases, including the positions equivalent to F598 and I603. This shows that isomeric selectivity is unlikely to arise from a specific sequence or combination of amino acids which bind *cis* substrates more effectively than *trans*, as seen for selected monoterpene synthases: the overall tertiary active site conformation could instead be responsible, with substrate interactions contributed to from a variety of residues on a synthase and product specific basis. The cyclisation chemistry (Section 1.2.1) possible from the C₁₅ FDP cation is exceedingly more complex than the single 1,6-cyclisation pathway available from the equivalent C₁₀ cations of monoterpene biosynthesis, therefore the active site architecture required to facilitate sesquiterpene formation is expectedly more nuanced.

4.6 Kinetic characterisation

To further understand the role of the F598 and G599 residues for catalysis, kinetic measurements were made for wild type EZS and the mutants at these positions. A radiolabelled substrate assay was used where substrate over a range of concentrations is incubated with a known concentration of enzyme for a defined time. This technique is preferred for terpene synthases as it is highly sensitive, able to detect low levels of radiolabelled substrate in the K_M concentration range for which GC-MS can be insensitive. The radioactivity of the organic reaction products is then measured using scintillation counting and the data are fitted against the Michaelis-Menten equation to determine V_{max} , k_{cat} and K_M .

The radiolabelled substrate used in this work was tritium labelled (Z,Z)-FDP ($[1-^3H_1]$ -(Z,Z)-FDP), obtained from within the research group having previously been synthesised by Dr. Chris Jones, who also optimised the reaction conditions used for the steady state enzymatic assays.¹⁵³ A detailed description of the experimental method, the data processing and calculations used to determine k_{cat} and K_M values can be found in Section 8.1.15. The data is summarised in Table 4.2 with the catalytic efficiencies derived from k_{cat}/K_M . Graphs showing the substrate concentration versus reaction rate for each mutant characterised are given in Supplementary Figure 4.15.

Table 4.2: Experimentally determined kinetic parameters for EZS mutants with radiolabelled (Z,Z)-FDP.

Mutant	k_{cat} / s^{-1}	$K_M / \mu M$	Catalytic efficiency / $\mu M^{-1} s^{-1}$
WT	0.022 ± 0.002	4.99 ± 1.2	4.4×10^{-3}
F598L	$8.77 \times 10^{-4} \pm 4.4 \times 10^{-5}$	1.58 ± 0.4	6.0×10^{-4}
F598W	$0.0079 \pm 9.5 \times 10^{-4}$	14.20 ± 3.7	5.6×10^{-4}
F598A	$0.012 \pm 5.1 \times 10^{-4}$	6.14 ± 0.7	2.0×10^{-3}
F598I, I603A	$0.004 \pm 6.6 \times 10^{-4}$	24.03 ± 6.8	1.7×10^{-4}
G599C	$0.019 s^{-1} \pm 0.001$	11.06 ± 1.8	1.7×10^{-3}

Steady state kinetic parameters were first determined for wild type EZS. The k_{cat} and K_{M} were found to be $0.022 \pm 0.002 \text{ s}^{-1}$ and $4.99 \pm 1.2 \text{ }\mu\text{M}$ respectively, in good agreement with previous work¹⁵³ which recorded a k_{cat} of $0.047 \pm 0.001 \text{ s}^{-1}$ and a K_{M} of $6.23 \pm 0.6 \text{ }\mu\text{M}$. The catalytic efficiencies are also comparable: 4.4×10^{-3} in this work and $7.6 \times 10^{-3} \text{ }\mu\text{M s}^{-1}$ previously, in line with the catalytic efficiencies of other terpene synthases such as Gd11oIS ($1.0 \times 10^{-3} \text{ }\mu\text{M s}^{-1}$)⁸⁵ and GDS ($2.6 \times 10^{-3} \text{ }\mu\text{M s}^{-1}$).¹⁰⁸

Following characterisation of wild type EZS, the other mutants were examined. The k_{cat} values were found to be lower than wild type, most noticeably for F598L with a value of $8.77 \times 10^{-4} \pm 0.001 \text{ s}^{-1}$. This mutant also had a lower K_{M} than the wild type whereas decreased binding affinities were observed for the other mutants, noticeably for F598W at $14.20 \pm 3.7 \text{ }\mu\text{M}$. In terms of catalytic efficiencies, these results are within the same order of magnitude as the wild type, except for F598I-I603A which shows the most significant change. The double mutant has the highest K_{M} value, found to be $24.03 \pm 6.8 \text{ }\mu\text{M}$, and a low k_{cat} at $0.004 \pm 6.6 \times 10^{-4} \text{ s}^{-1}$, representing an approximately 26 times less efficient enzyme. However, as this has been shown to produce mostly (Z,Z)-farnesol, true enzymatic substrate conversion could be orders of magnitude lower again.

Overall, the catalytic performance of the mutants has been shown to be moderately compromised relative to the wild type (except the double mutant). The amino acid substitutions all retain the non-polar aliphatic character of the original residues reflecting the relatively subtle overall changes to catalytic efficiency, despite the significant changes to the product profiles.

4.7 Summary

Homology modelling was used to assist structure-based mutagenesis of active site amino acids of EZS hypothesised to be involved in the catalytic reaction mechanism. Remarkably, half of the mutations made had preserved product specificity with no discernible product profile change, despite the complex reaction mechanism involving at least three carbocations and deprotonation. Key plasticity residue F598 was however identified in the G-helix kink: substitution with other hydrophobic residues was shown to significantly increase product diversity through remoulding of the active site template. These changes also modified the ingress of water to the active site as sesquiterpene alcohols were detected, further highlighting the fine-tuning of this region in protecting reactive carbocation intermediates during catalysis.

G-helix kink residues also proved to be (at least partly) responsible for the unusual substrate selectivity of EZS. The double mutant, F598I-I603A, turned over (*E,E*)-FDP into several non- and hydroxylated sesquiterpenes, whilst it had only poor activity for (*Z,Z*)-FDP and produced mainly the linear farnesol. Further kinetics studies with the double mutant and radiolabelled (*E,E*)-FDP would help clarify the degree to which these residues behave as isomeric switches. To continue investigating the structure-function relationship for isomeric selectivity, other geometric isomers of FDP ((*Z,E*)- and (*E,Z*)-FDPs) could be incubated with mutants of EZS targeting this region, and the structures of evolutionary divergent TPS-a synthases could be examined more broadly. Pursuit of rationally designed mutants around the G-helix region for (*E,E*)-FDP accepting enzymes may similarly lead to reverse engineered enzymes able to accept (*Z,Z*)-FDP. This has enormous potential for future development of orthogonal metabolic pathways for *in vivo* sesquiterpenoid production based on (*Z,Z*)-FDP.

4.8 Supplementary

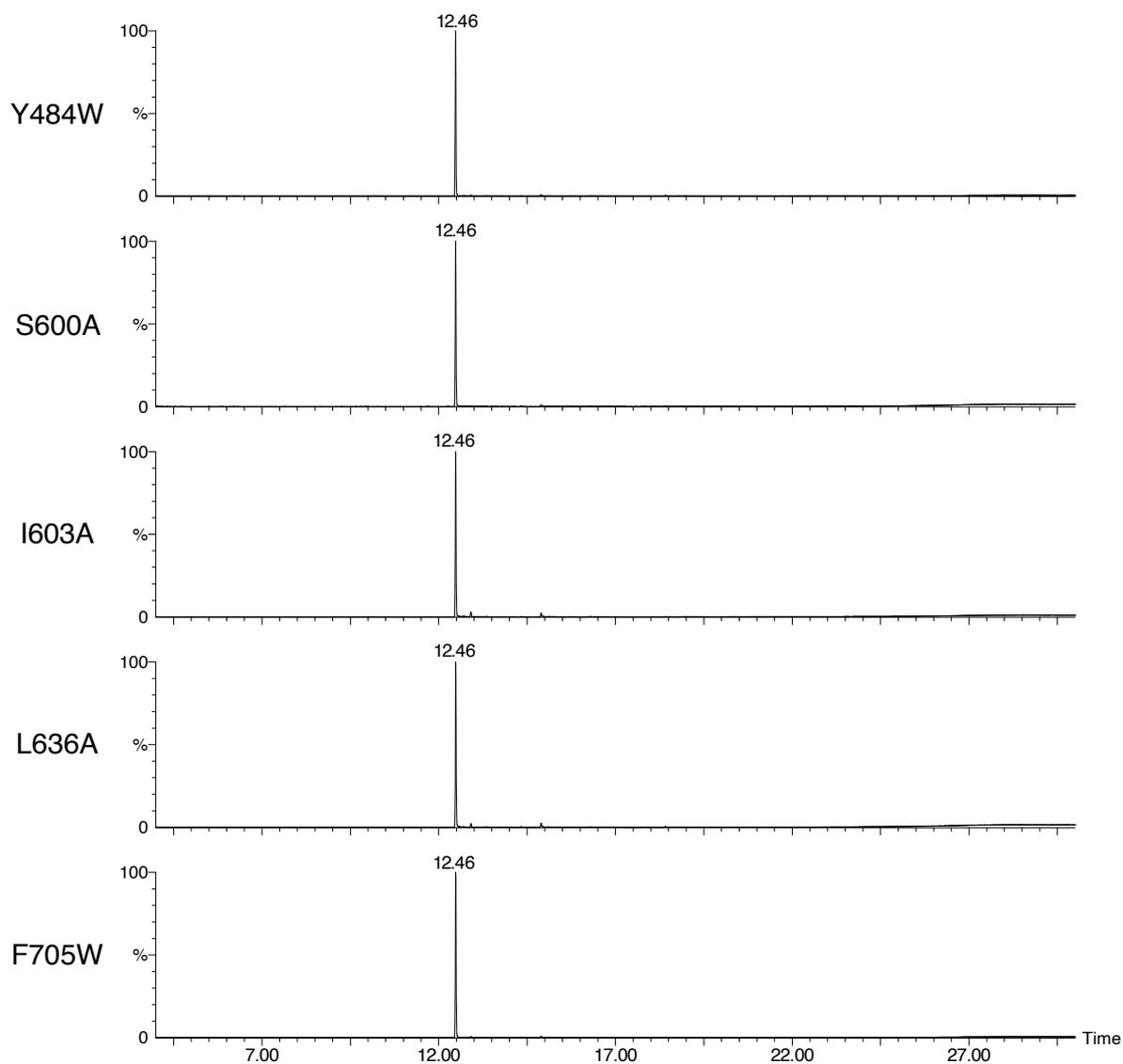


Figure 4.9: Total ion chromatograms of the pentane extractable products from incubations of EZS mutants with (Z,Z)-FDP. The compound eluting at 12.46 min is confirmed to be 7-*epi*-zingiberene by comparison to the product of the wild type incubation.

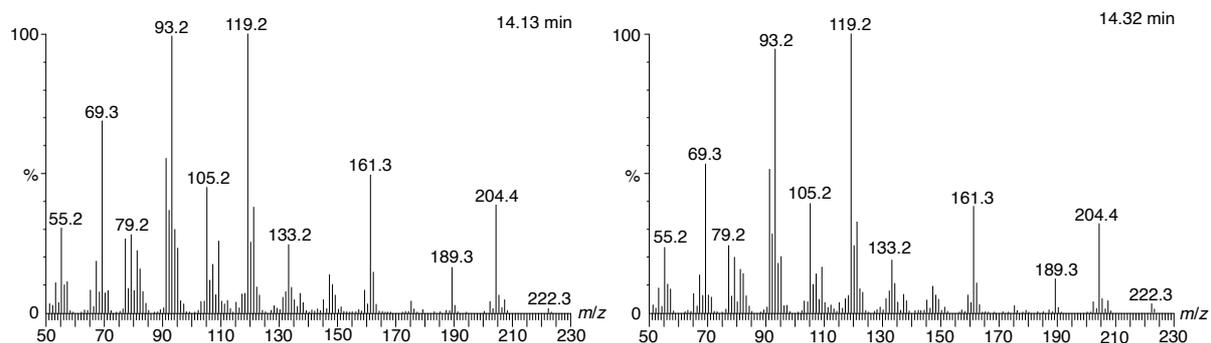


Figure 4.10: Mass spectra of the compounds eluting at 14.13 and 14.32 min in the organic assay products of mutant F598L. Mass ion peaks at $m/z = 222.3$ are consistent with hydroxylated sesquiterpenes.



1 2 3

1. Product of F598L
2. Nerolidol
3. Bisabolene

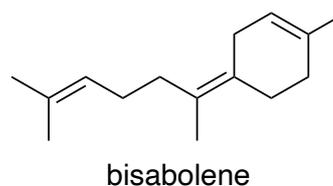
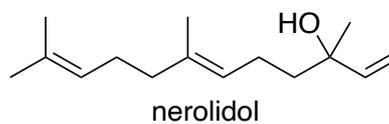


Figure 4.11: Photo of a TLC plate eluted in 10% ethyl acetate/hexane showing a comparison of products from the F598L incubation with (Z,Z)-FDP with nerolidol and bisabolene as representative non- and hydroxylated sesquiterpenes. F598L products anticipated to be hydroxylated can be seen as several spots in the lower left-hand side of the plate.

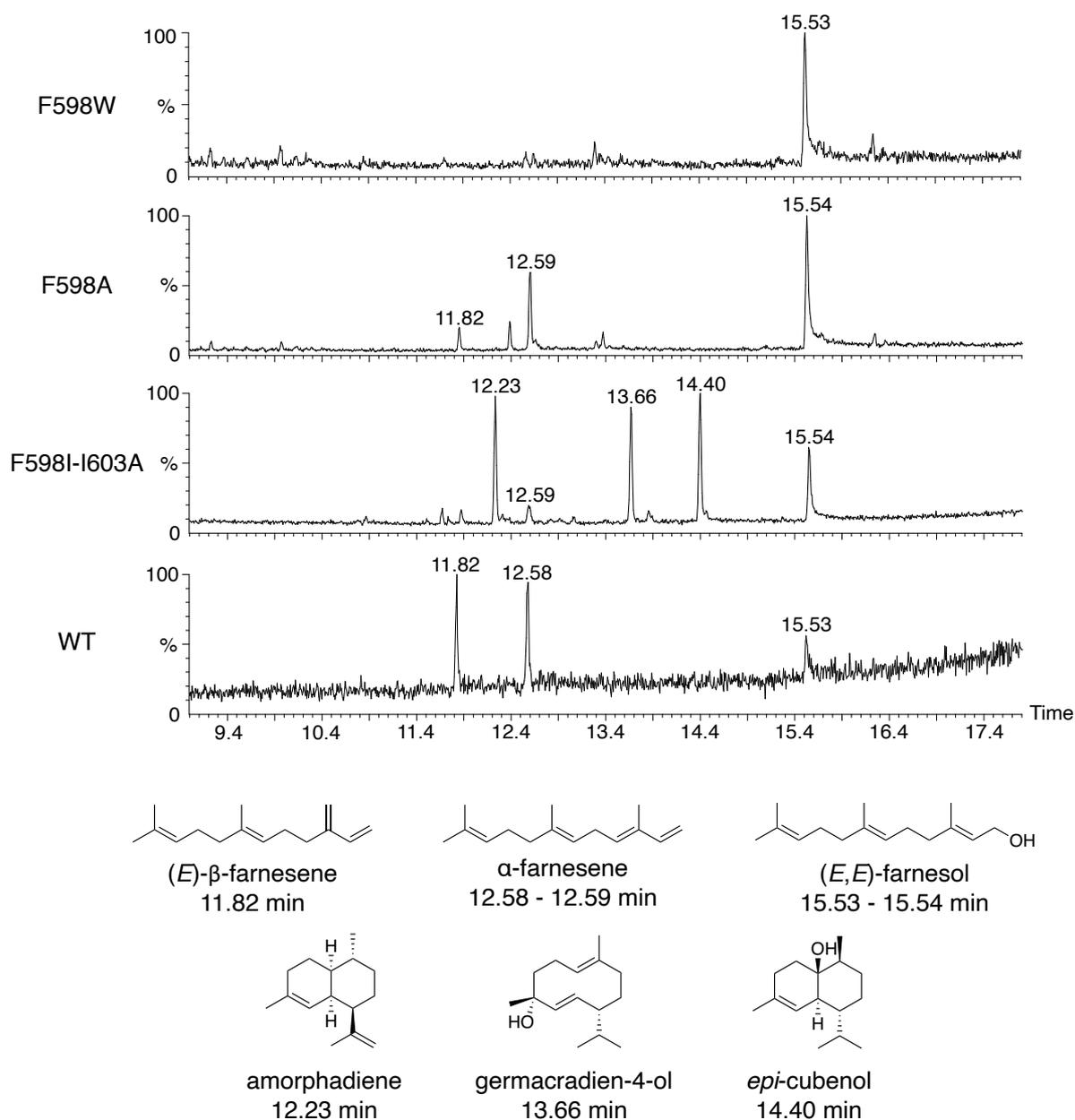


Figure 4.12: Total ion chromatograms of the pentane extractable products from incubations of EZS mutants with (*E,E*)-FDP where compounds were detected. The compound eluting at 15.53 - 15.54 min was identified as (*E,E*)-farnesol by comparison to a synthetic standard whilst compounds eluting at 14.40 min 13.66 min 12.23 and 11.82 min were identified as *epi*-cubenol, germacradien-4-ol, amorphadiene and (*E*)-β-farnesene, respectively, by comparison to enzymatically generated standards (although absolute stereochemistry was not defined). The compound eluting at 12.58 - 12.59 min was tentatively identified as α-farnesene by comparison to the NIST mass spectral database.

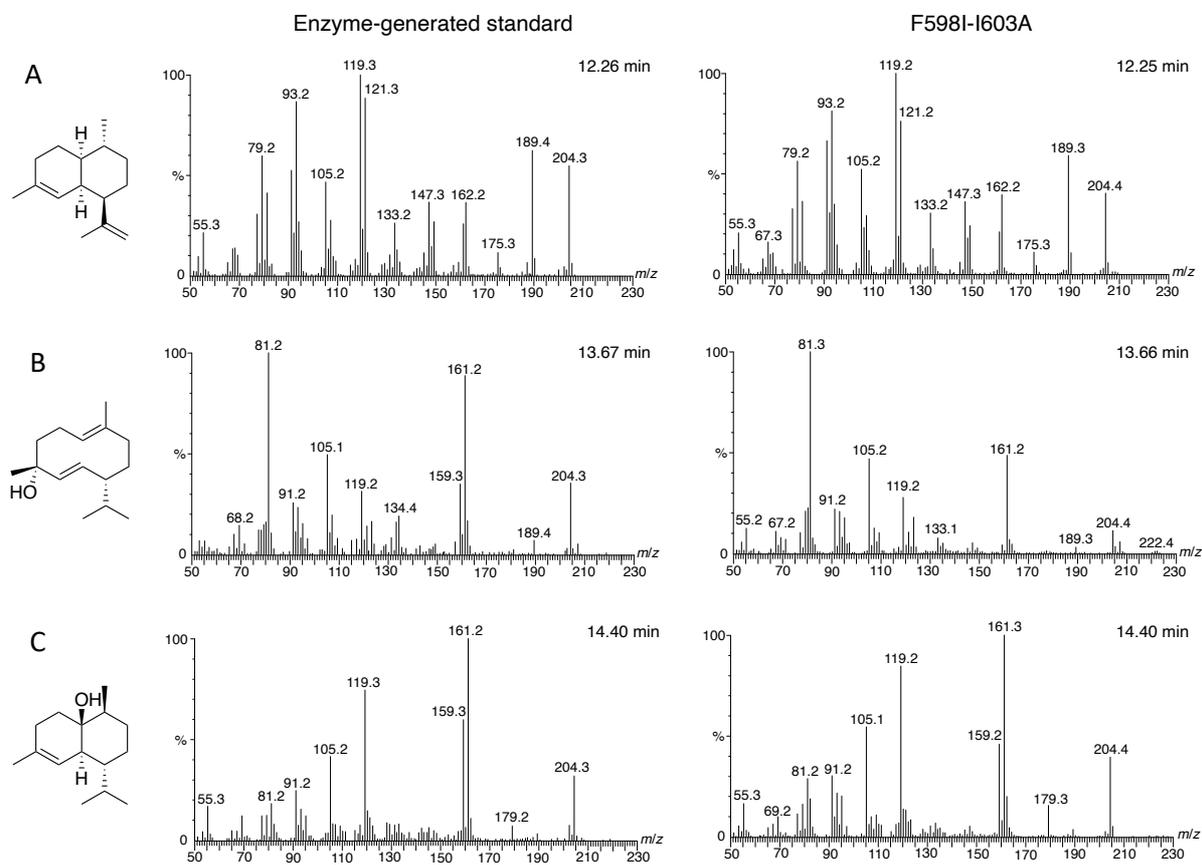


Figure 4.13: Comparison of the mass spectra of reaction products of EZS double mutant F598I-I603A with (*E,E*)-FDP (right-hand side) with confirmed mass spectra of (A) amorphadiene, (B) germacradien-4-ol and (C) *epi*-cubenol (left-hand side).

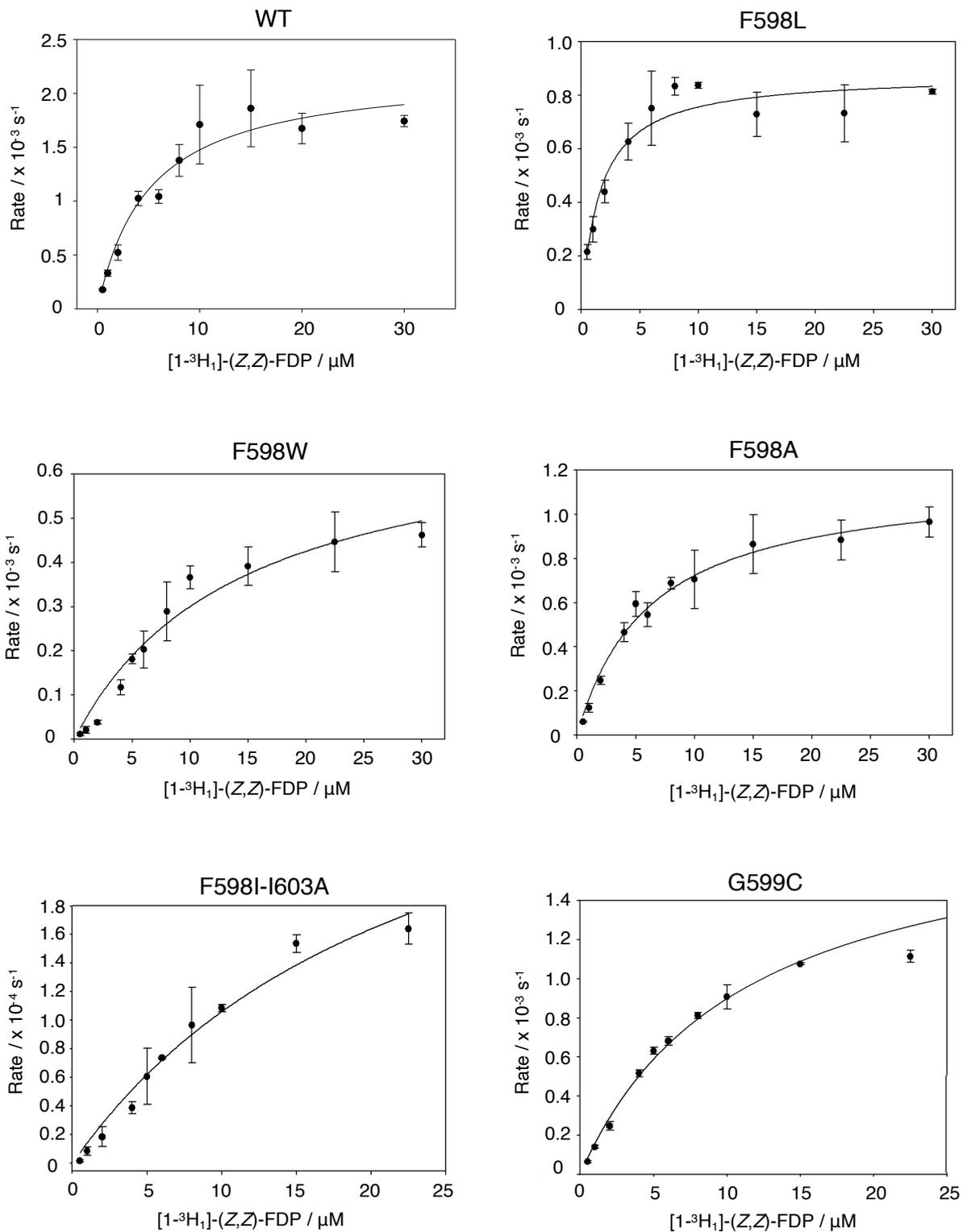


Figure 4.15: Graphs showing the relationship between substrate concentration and reaction rate for wild type EZS and a selection of mutants. Data was fitted against the Michaelis-Menten function using SigmaPlot statistical software (*Systat Software Inc.*). Error bars are expressed as standard error of the mean.

Chemoenzymatic Generation of Novel Semiochemical Analogues

5.1 Introduction

Terpene synthases are fascinating biocatalysts with the ability to generate complex molecular scaffolds which are often inaccessible through synthetic chemical routes. The potential to access new areas of chemical space via the inherent promiscuity and plasticity of these enzymes is becoming increasingly well recognised. Novel terpenoid analogues with useful properties and bioactivities are being reported with increasing frequency, generated from unnatural substrates, engineered enzymes, or both.^{99,103,108}

The research area is currently evolving around two approaches. Primarily, these discoveries have relied on the total synthesis of rationally designed substrate analogues which are then fed to the enzyme. This practice developed as studies using FDP analogues as mechanistic probes led to the discovery of sesquiterpenes with novel functionalities.^{71,111,191} This approach does however have drawbacks: the syntheses of substrate analogues can be excessively challenging and typically only small amounts for research are ever produced, whilst it is also difficult to predict whether the unnatural substrate will be accepted by the enzyme.

The second approach toward expanding the terpenome is based on combinatorial enzymatic approaches, whereby terpenoid biosynthesis genes from different organisms can be assembled in new combinations to generate novel compounds.^{192,193} Some of these methods involve *in vivo* engineering of non-canonical terpene biosynthetic pathways, such as those involving S-adenosyl methionine-dependent methyltransferases^{194,195} and the lepidopteran homo-mevalonate pathway.¹⁹⁶ Whilst large areas of the terpenome can potentially be accessed rapidly, pathways using enzymes from different organisms can be exceedingly difficult to engineer to produce in adequate quantities to enable isolation and characterisation of the novel compounds.

5.1.1 Aims

The work described in this chapter aimed at the development of a new method for the *in vitro* generation of unnatural terpene analogues. It aimed to circumvent *in vivo* pathway engineering and the need for lengthy syntheses of FDP analogues by rapidly generating them using prenyltransferases and analogues of IDP and DMADP (Figure 5.1). This method expands upon the new kinases diphosphorylation method¹⁶¹ for generating the universal C₅ terpene precursors using their respective alcohols, isoprenol and prenil, as described in Section 2.3. The kinases have been shown to be promiscuous which has presented an interesting opportunity to readily produce a variety of diphosphates *in vitro* which can be used as substrates in screening reactions with prenyltransferases and downstream terpene synthases for terpenoid discovery. Epizingiberene synthase (Ezs), (*S*)-germacrene D synthase (GDS)

and (*E*)- β -farnesene synthase (EBFS) were used to explore this system as their sesquiterpene products have potential application as sustainable agrochemicals, functioning as highly specific repellent semiochemicals for a variety of pests.^{105,136,197} Previous testing of the hypothesis that if an unnatural substrate is accepted by the synthase, the product is likely to share sufficient chemical space as the natural ligand to be active in the insect receptor, led to the discovery of germacrene D analogues with remarkable bioactivities.¹⁰⁸ These semiochemical-producing enzymes were therefore used to explore this system with the aim of discovering analogues of their useful products that could be assessed via insect bioactivity testing.

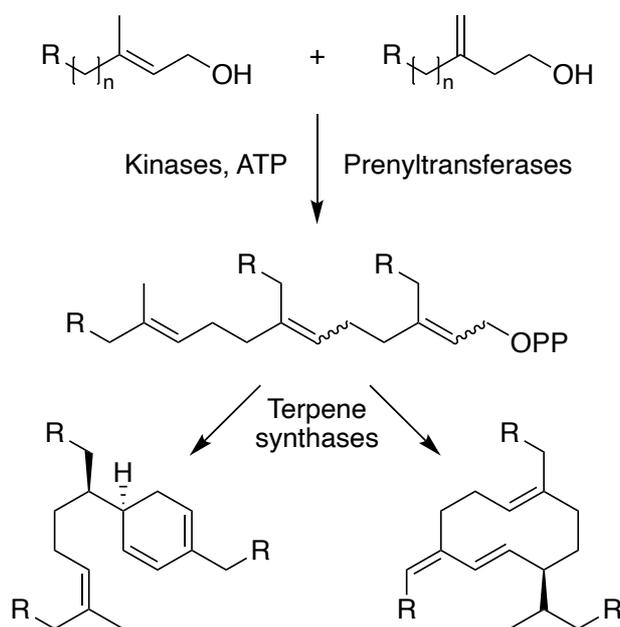


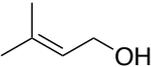
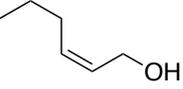
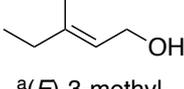
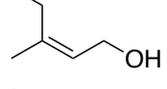
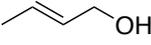
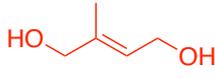
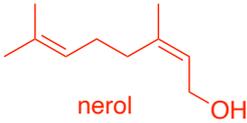
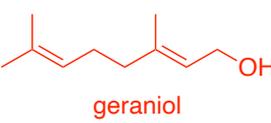
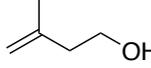
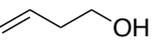
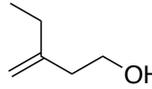
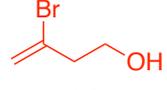
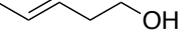
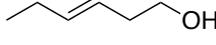
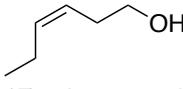
Figure 5.1: Scheme illustrating new chemoenzymatic method explored to discover analogues of useful natural products.

5.2 Alcohol analogue phosphorylation

5.2.1 Small scale analogue phosphorylation

The first objective in developing a system for *in vitro* analogue synthesis was to create a library of IDP and DMADP analogues and to understand the substrate scopes of THIM and IPK. Multiple analogues of prenil and isoprenol with different chain lengths and functionalities were chosen with double bonds positioned either at C2 or C3 of the alcohol to emulate their

respective electronic configurations (Figure 5.2). As well as double bond isomers, these included alcohols with hydroxy and halogen functionalisation which could provide potential opportunities for downstream derivatisation if incorporated into terpene scaffolds. Those annotated with an asterisk were kindly donated from within the research group and the rest available through commercial sources. IPK and THIM were expressed and purified as previously described (Section 8.1.10). Small-scale assays (1 mL total volume) were assembled for each alcohol (2 mg/mL) with THIM (65 μ M) and IPK (5 μ M) in reaction buffer (100 mM TRIS pH 8.0, 5 mM ATP, 200 mM KCl, 5 mM MgCl₂) which included an ATP regeneration system (75 mM PEP and 10 units/mL PK). Reactions were incubated at room temperature with shaking (400 rpm) for 24 hours and then analysed using ³¹P-{¹H} NMR spectroscopy.

	Analogues		
 prenol	 (Z)-2-hexen-1-ol	 ^a (E)-3-methyl-2-penten-1-ol	 ^a (Z)-3-methyl-2-penten-1-ol
	 (E)-2-buten-1-ol	 ^b 4-hydroxyprenol	
	 nerol		 geraniol
 isoprenol	 3-buten-1-ol	 ^c 3-ethyl-3-buten-1-ol	 ^b 3-bromo-3-buten-1-ol
	 3-penten-1-ol	 (E)-3-hexen-1-ol	 (Z)-3-hexen-1-ol

^a Synthesised by Ashley Ramdass

^b Synthesised by Gwawr Davies

^c Synthesised by Dr. Alice Dunbabin

Figure 5.2: Analogues of prenol and isoprenol screened as substrates for diphosphorylation by THIM and IPK. Alcohols which were not successfully diphosphorylated are highlighted in red. (Z)-3-Methyl-2-penten-1-ol was proven to be a substrate as discussed in Section 5.5.3.

All but four of the 13 alcohols (highlighted in Figure 5.2) showed some degree of diphosphorylation after 24 hours, indicated by peaks relating to two doublet peaks around -6.0 and -10.0 ppm in the NMR spectra (Supplementary Section 5.7). From comparison of the abundance of monophosphate species relative to diphosphate species, particularly good substrates include the homo analogues of isoprenol and prenol (3-ethyl-3-buten-1-ol and (*E*)-3-methyl-2-penten-1-ol). Diphosphate peaks were not observed for nerol, geraniol, 4-hydroxyprenol and 3-bromo-3-buten-1-ol. These results were corroborated in the published work alongside this project where further kinetic analyses of the first phosphorylation reaction by THIM found that 4-hydroxyprenol was a substrate albeit a poor one, as the phosphorylation rate was closer to the rate of ATP hydrolysis. A maximum substrate chain length of seven carbons was established.¹⁶¹

These findings may be attributed to the relatively open, solvent exposed sites of the two enzymes (Figure 5.3) which may facilitate predominantly weak and non-specific interactions between a wide range of substrates, represented by K_M values in the millimolar range for THIM which were similar across the substrates.¹⁶¹ Of the alcohols which were not successfully diphosphorylated, longer chain C_{10} geraniol and nerol and bulky hydrophobic 3-bromo-3-buten-1-ol and 4-hydroxyprenol are likely too large for the active site.

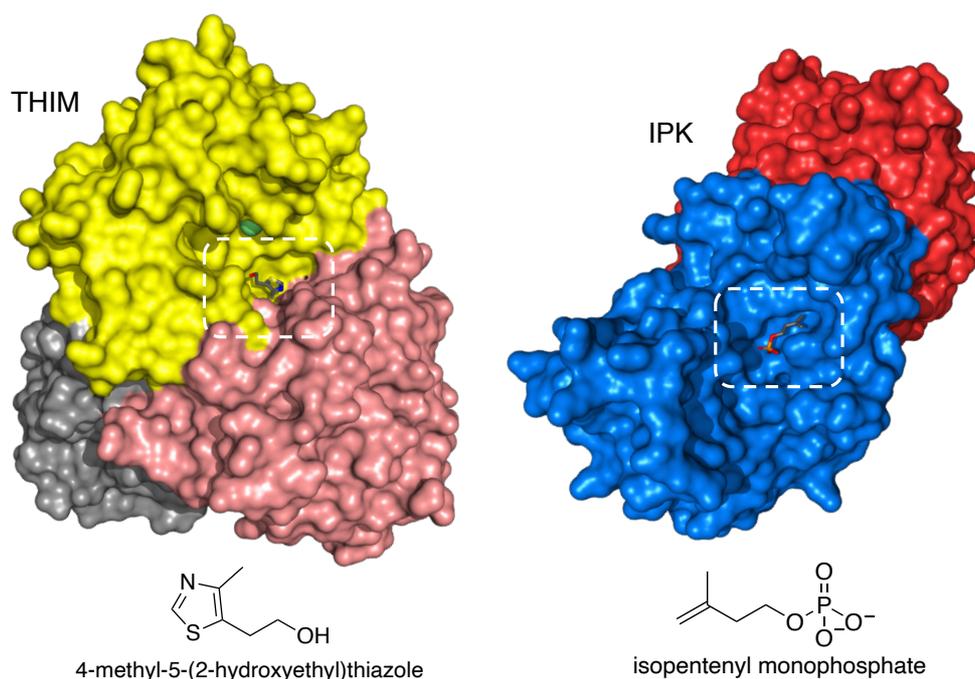


Figure 5.3: Graphic representations of the structures of THIM and IPK showing the small but open, solvent exposed active sites (white dashed boxes) and the structures of their natural substrates.

5.3 FDP analogues

Having efficiently prepared a variety of DMADP and IDP analogues, screening with prenyltransferases zFPS and FPPS could then be carried out to attempt *in vitro* synthesis of new (*Z,Z*)-FDP and (*E,E*)-FDP analogues.

5.3.1 Small-scale screening with prenyltransferases

In native FDP biosynthesis, DMADP is used as the starter for the successive chain extensions with IDP (Section 1.1.2). Two IDP equivalents are required for every equivalent of DMADP to generate FDP. For this reason, the analogues were initially screened using a 1:2 ratio of DMADP with IDP analogue, or 1:2 ratio of DMADP analogue with IDP. Screening was also carried out with NDP and GDP as reaction starters in a 1:1 ratio with IDP analogues. Both NDP and GDP are natural intermediates in the biosynthesis of the respective FDPs, made *in situ* from the first condensation reaction between DMADP and IDP. FDP synthases, zFPS and FPPS, were expressed and purified as previously described (Section 8.1.10). Assays were assembled in 1 mL aliquots containing 50 μ L DMADP or analogue stock solution and 100 μ L IDP or analogue stock solution, 2 μ M prenyltransferase, 2 mM β -CD and 5 mM $MgCl_2$ in 100 mM TRIS pH 8.0. The reactions were incubated at room temperature with shaking (400 rpm) for 16 hours. The following day, alkaline phosphatase was added to a final concentration of 10 mg/mL and the reactions incubated for 2 hours at 37 °C. Organic reaction products relating to the dephosphorylated FDP analogues were extracted from the assays with pentane (1 mL) and analysed by GC-MS.

5.3.2 (*Z,Z*)-FDP analogues

GC-MS analysis of the *cis*-prenyltransferase (zFPS) screening assays provided evidence of the *in vitro* generation of novel (*Z,Z*)-FDP related compounds from the DMADP analogues (*E*)-3-methyl-2-penten-ODP (homo-DMADP) and (*E*)-2-buten-ODP (nor-DMADP). It is first relevant to discuss the mass fragmentation pattern of (*Z,Z*)-farnesol (Figure 5.4) produced from control incubations with the native substrates. The gas chromatogram gives a single compound at 14.88 min and the corresponding mass spectrum fragmentation pattern shows a molecular ion peak at $m/z = 204.3$. This relates to the compound produced following the elimination of water (or the hydroxy group and a proton): this effect is a common artefact of primary alcohol ionisation and fragmentation. The ion abundance “picket fence” peak series pattern with peaks of decreasing abundance separated by 12 - 14 amu is typical of a linear branched hydrocarbon,¹⁷⁴ seen distinctly in the mass spectrum of (*E,E*)-farnesol (Figure 5.7)

but differing where the base peak is at $m/z = 93.2$ compared to $m/z = 69.1$ of (*E,E*)-farnesol. These results have been verified by comparison to a synthetically produced standard of (*Z,Z*)-farnesol.

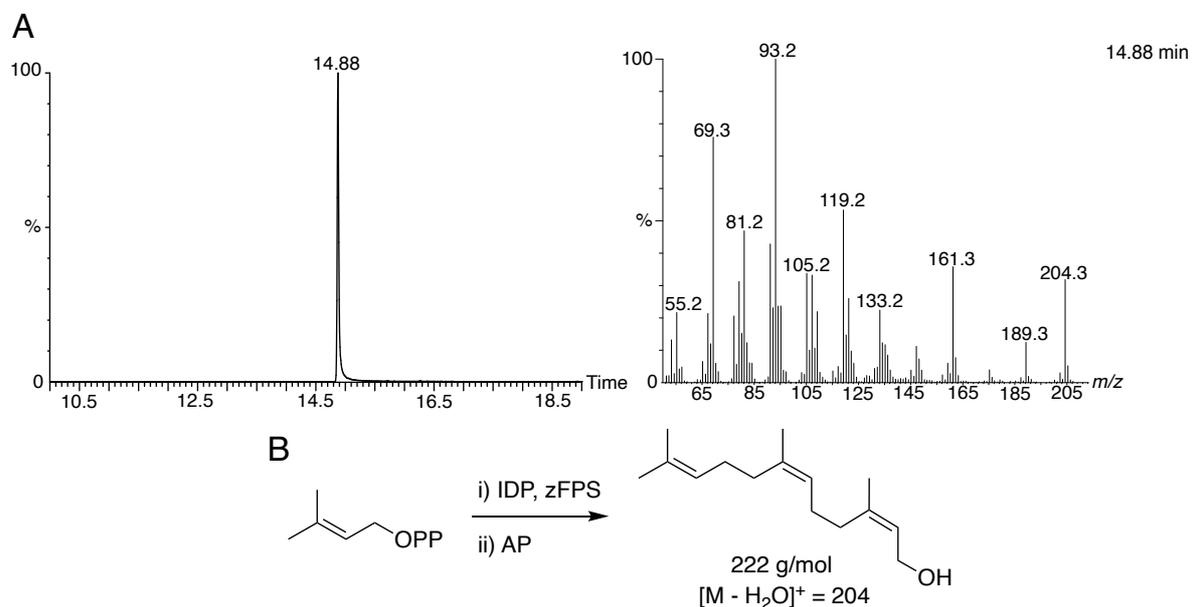


Figure 5.4: (A) Total ion chromatogram of products resulting from the control incubation of DMADP with IDP, zFPS and AP with accompanying mass spectrum of the compound eluting at 14.88 min. (C) Scheme showing reaction product ($m/z = 222$) with mass of the ion of the product of water elimination ($m/z = 204$).

13-Nor analogue

The GC-MS results arising from the nor-DMADP incubation are shown in Figure 5.5. A single peak can be found at 11.11 min which, when considering that the gas-chromatographic separation is boiling point dependent, is consistent with the lower molecular weight C_{14} farnesol analogue (compared to the 14.88 min for (*Z,Z*)-farnesol). The mass spectrum pattern is relatable to that of (*Z,Z*)-farnesol in Figure 5.4: both share base peak at $m/z = 93.2$ and abundant ions at $m/z = 119.2$, although the molecular ion peak is found at $m/z = 190.3$ which would correspond to the expected mass of the nor-analogue with loss of water.

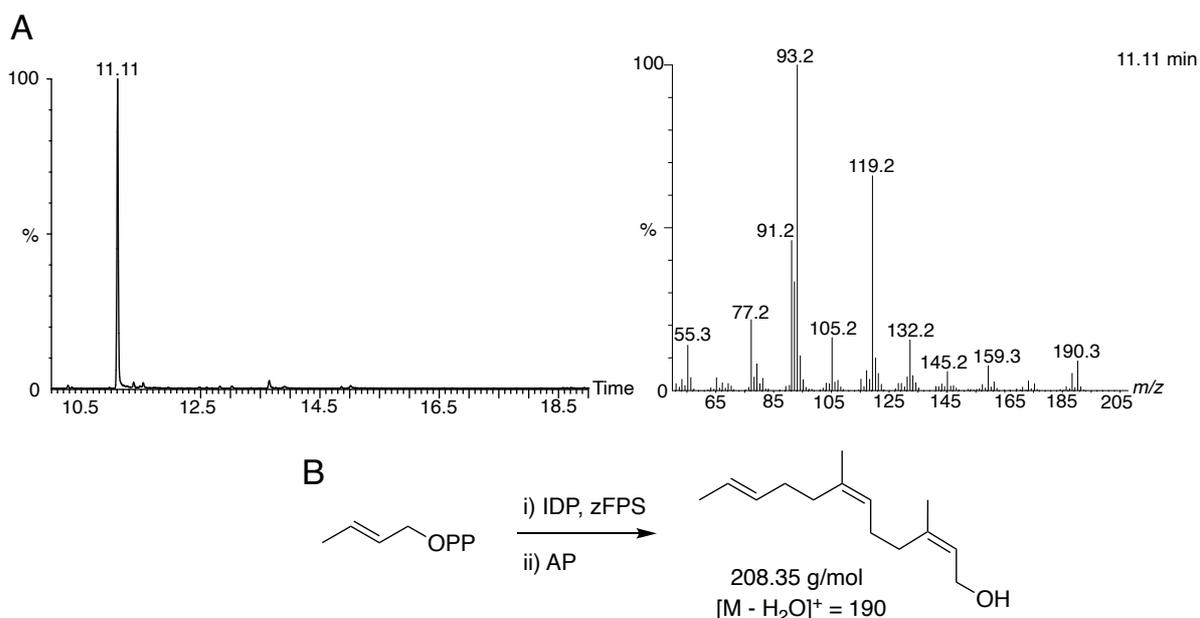


Figure 5.5: (A) Total ion chromatogram of products resulting from the incubation of nor-DMADP with IDP, zFPS and AP with accompanying spectrum of the compound eluting at 11.11 min. (B) Scheme showing reaction and expected product ($m/z = 208$) with mass of the ion of the product of water elimination ($m/z = 190$).

12-Methyl analogue

Figure 5.6 concerns the results generated from incubation of homo-DMADP with IDP and zFPS. The gas chromatogram clearly shows a major compound eluting at 13.88 min and two minor compounds with retention times around 15.46 min. These compounds are clearly quite distinct yet they have very similar mass fragmentation patterns which differ to the ion abundance pattern of (*Z,Z*)-farnesol and the previous analogue. Although the molecular ion peak at $m/z = 218.4$ is indicative of the parent farnesol compound with the extra carbon, its identity is ambiguous and more relatable to the pattern of 7-*epi*-zingiberene and other cyclic terpenes for example. The compound at 15.46 min could relate to the suggested farnesol analogue of higher molecular weight as it is greater than the 14.88 min retention time of (*Z,Z*)-farnesol. The compound at 13.88 min could arise from further chemistry taking place in solution: for example, when dephosphorylated by alkaline phosphatase, NDP is readily converted to terpineol, as detected by GC-MS analysis; similarly, the solvolysis of GDP into cyclic or linear alcohols has been observed in the presence of amorphadiene and *epi*-aristolochene synthase.⁸¹

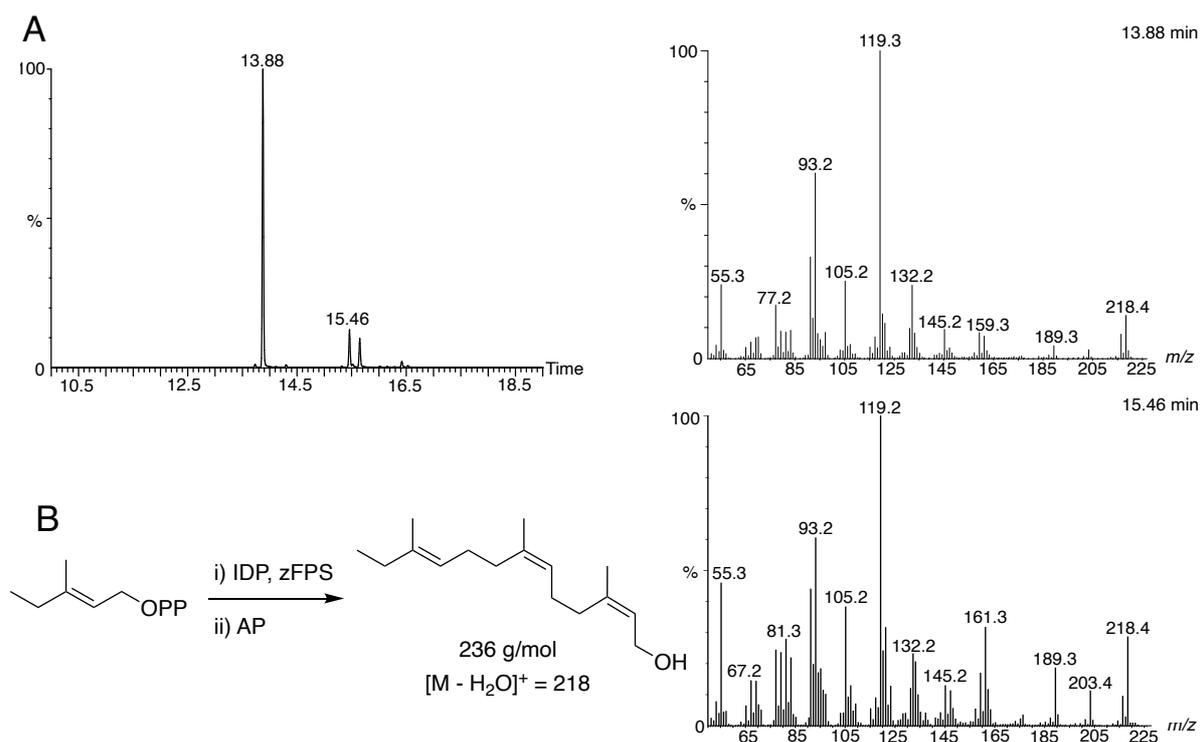


Figure 5.6: (A) Total ion chromatogram of products resulting from the incubation of homo-DMADP with IDP, zFPS and AP with accompanying mass spectra of the compounds eluting at 13.88 min and 15.46 min. (B) Scheme showing reaction and expected product ($m/z = 236$) with mass of the ion of the product of water elimination ($m/z = 218$).

A further analogue was detected when screening with EZS as described below (Section 5.4.1), which arose from the (*Z*)-2-hexen-ODP substrate, although the equivalent farnesol product was not detectable. This is likely due to it being produced in very small amounts that are beyond detection from the assay size used, whilst addition of EZS favours the reaction equilibrium for product formation enabling GC-MS detection of the 7-*epi*-zingiberene analogue.

5.3.3 (*E,E*)-FDP analogues

A variety of (*E,E*)-FDP analogues produced by *trans*-prenyltransferases have been reported in the recent publication alongside this work which include 12-methyl-FDP, 15-methyl-FDP and 14,15-dimethyl-FDP.¹⁶¹ GC-MS results of this combinatorial screening has detected additional unnatural analogues for the first time from 3-buten-ODP (nor-IDP), homo-DMADP and 3-ethyl-3-buten-ODP (homo-IDP). As before, the GC-MS results from the control incubation to generate farnesol is given to enable comparison (Figure 5.7). The mass

spectrum shape is again typical for a linear terpenoid and the molecular ion peak reads $m/z = 204.2$ with the anticipated loss of water from the parent ion.

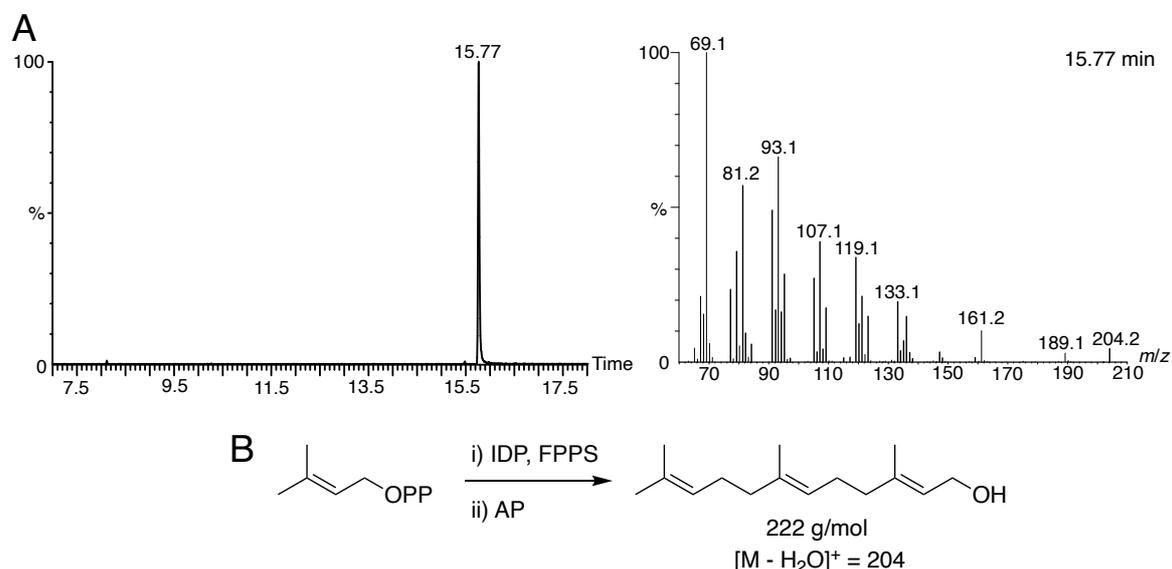


Figure 5.7: (A) Total ion chromatogram of resulting from the control incubation of DMADP with IDP, FPPS and AP with accompanying mass spectrum of the compound eluting at 15.77 min. (B) Scheme showing reaction product ($m/z = 222$) with mass of the ion of the product of water elimination ($m/z = 204$).

15-Nor analogue

Several distinct compounds can be seen in the total ion chromatogram of products from the incubation of GDP and nor-IDP (Figure 5.8). The major product at 14.27 min appears to relate to a 14-carbon farnesol analogue as the fragmentation pattern correlates well to native farnesol, whilst the $m/z = 190.3$ mass peak would again indicate loss of water from the primary alcohol generated in the assay. Also, the retention time is reduced from C_{15} farnesol as would be expected. The compound at 8.49 min was confirmed as C_{10} geraniol by comparison to an authentic standard and would have arisen from excess GDP. The lack of methyl group in the IDP analogue could reduce the stability of intermediate cations during the enzymatic condensation reaction from tertiary to secondary stabilised, and hence overall would be less favourable leading to unreacted GDP.

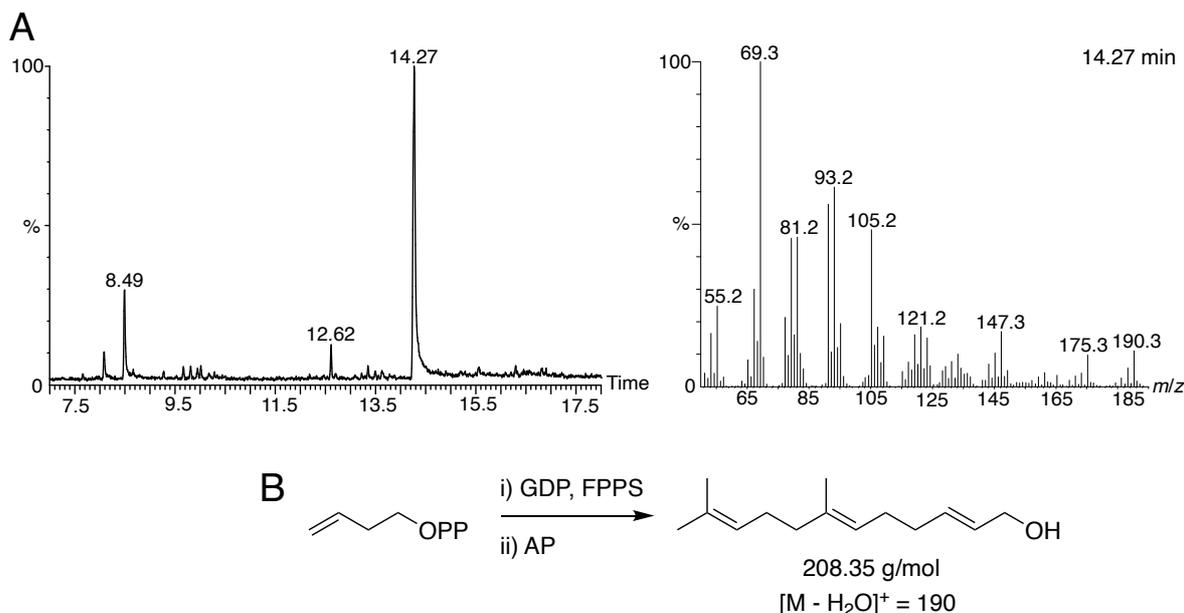


Figure 5.8: (A) Total ion chromatogram of products resulting from the incubation of nor-IDP with GDP, FPPS and AP with accompanying mass spectrum of the compound eluting at 14.27 min. (B) Scheme showing reaction and expected product ($m/z = 208$) with mass of the ion of the product of water elimination ($m/z = 190$).

12,14,15-Trimethyl analogue

Two compounds of interest were also detected by GC-MS from the incubation of homo-DMADP with homo-IDP (Figure 5.9). The peak at 18.58 min has a mass ion at $m/z = 246.3$ which would account for the three extra methyl groups of the trisubstituted farnesol cation (B) following H_2O elimination. The fragmentation pattern relates well to the control although the base fragment is now $m/z = 55.3$ instead of $m/z = 69.1$, and there is a greater abundance of the fragment at $m/z = 83.2$. This is effect is seen again with methyl substitution and further explanation given in Section 5.4.2. The mass spectrum of the compound eluting at 11.26 min has a mass ion at $m/z = 164.3$ which could relate to the product of one addition of homo-IDP to homo-DMADP, resulting in a C_{12} analogue of GDP (C). The fragmentation pattern is less distinctly related to farnesol or geraniol but contains the base ion at $m/z = 55.3$ of the $C_4H_7^+$ species and the $m/z = 135.3$ peak which likely relates to a $C_{10}H_{15}^+$ ion. The conditions used for the assay gave a higher ratio of the one-addition product, but this has potential to be optimised for favour the two-addition product if desired by increasing the ratio of homo-IDP in the reaction,¹⁶¹ or using an additional sesquiterpene synthase to favour the equilibrium.

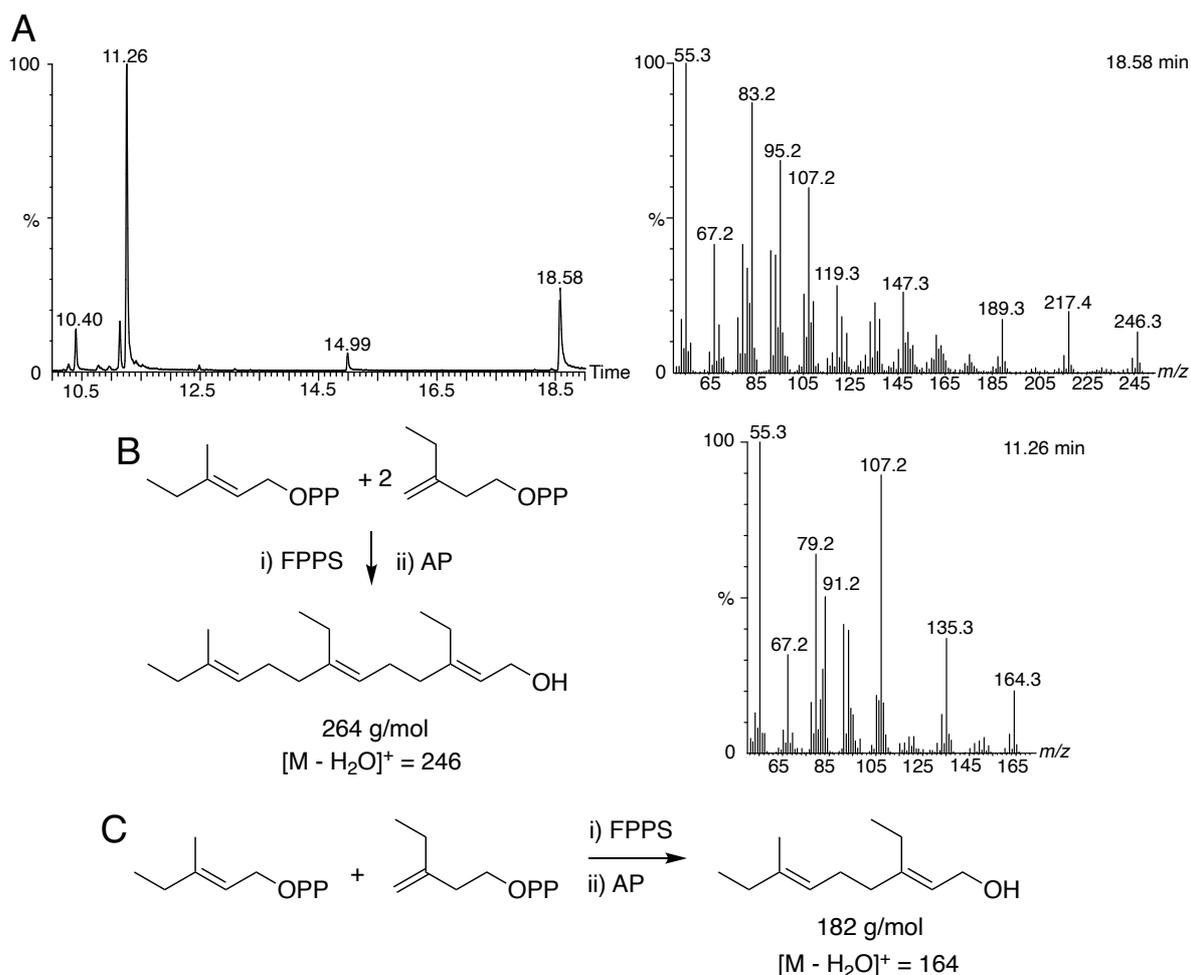


Figure 5.9: (A) Total ion chromatogram of products resulting from the incubation of homo-DMADP with homo-IDP, FPPS and AP with accompanying mass spectra of the compound eluting at 11.26 and 18.58 min. (B) Scheme showing reaction and expected product at 18.58 min with mass of the ion of the product of water elimination ($m/z = 246$). (C) Scheme showing reaction and expected product at 11.26 min with mass of the ion of the product of water elimination ($m/z = 164$).

5.3.4 Chemistries of the *cis*- and *trans*-prenyltransferases

Despite carrying out a similar role *in vivo*, the two prenyltransferases, zFPS and FPPS, have contrasting substrate specificities and chemistries, exemplified by the summary of their substrate scopes in Figure 5.10. Both have moderate specificities for the range of DMADP analogues presented, successfully initiating the reaction from several different and bulkier substrates, although perhaps this is unsurprising given that their active sites have space to facilitate the addition of C₅ IDP to C₁₀ intermediates during the second chain extension reaction. FPPS has however proven more promiscuous with the ability to accept IDP

analogues where zFPS has not, even accepting three C₆ subunits to provide the 12,14,15-trimethyl-FDP analogue. The crystal structure of zFPS reveals secondary IDP binding sites as well as the substrate binding site,⁴² whilst kinetic studies have shown that its catalysis is tightly regulated by IDP first as a substrate and positive effector, then by DMADP.³⁴ The kinetics are modelled by Hill rather than Michaelis-Menten, as is the case for FPPS,^{198,199} where IDP has a Hill number of 4 for zFPS indicating the binding of multiple IDP molecules is required for catalysis. This complex IDP-dependent kinetics may mean zFPS is less active, if at all, with IDP analogues. Additionally, zFPS did not successfully generate compounds from NDP and IDP analogue reactions, whereas FPPS generated compounds from GDP and IDP analogues.

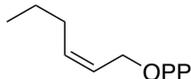
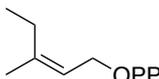
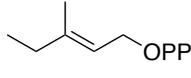
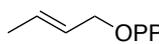
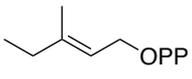
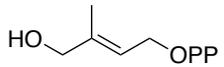
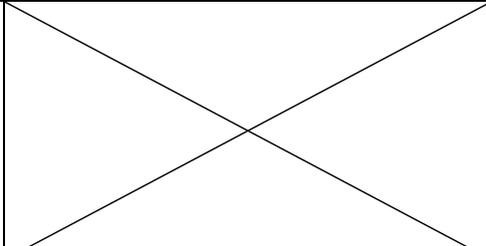
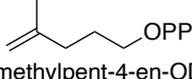
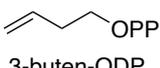
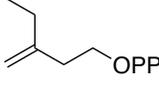
	zFPS substrates	FPPS substrates
DMADP analogues	 (Z)-2-hexen-ODP  (Z)-3-methyl-2-penten-ODP  (E)-3-methyl-2-penten-ODP  3-buten-ODP	 (E)-3-methyl-2-penten-ODP  4-hydroxy-DMADP
IDP analogues		 4-methylpent-4-en-ODP  3-buten-ODP  3-ethyl-3-buten-ODP

Figure 5.10: The substrate scopes of the prenyltransferase enzymes explored within this system. 4-Methylpent-4-en-ODP and 4-hydroxy-DMADP were shown to be FPPS substrates in previous work¹⁶¹ and (Z)-3-methyl-2-penten-ODP and (Z)-2-hexen-ODP were shown to be zFPS substrates later in this work.

5.4 Sesquiterpene analogues

Confirming that both prenyltransferases showed a degree of promiscuity and were amenable to accepting substrate analogues, sesquiterpene synthases EZS, GDS and EBFS were included in the next step for novel compound discovery. Expression plasmids containing the genes for EBFS and GDS were sourced from the Allemann Group library and both enzymes subsequently expressed and purified in addition to EZS (Section 8.1.10). Reactions were assembled using all diphosphate analogues as per Section 5.3.1 but with the addition of 20 μM of the relevant sesquiterpene synthase (zFPS with EZS; FPPS with GDS and EBFS). The assays were incubated as before but instead of the addition of alkaline phosphatase, they were directly extracted with pentane (1 mL) and the organic products analysed by GC-MS.

5.4.1 7-Epi-zingiberene analogues

Three novel analogues of 7-*epi*-zingiberene could be identified from GC-MS analysis of the EZS screening reactions (Figure 5.11). In addition to the two DMADP analogues discussed above which are substrates for zFPS, a third analogue was detected from (*Z*)-2-hexen-ODP incubation with EZS. In comparison to 7-*epi*-zingiberene with a retention time of 12.46 min, as would be expected, products of the higher molecular weight diphosphate analogues have longer retention times (13.82 and 13.78 min for B and C respectively) and shorter retention times for the lower molecular weight compound (10.96 min for D). The mass fragment ion of each product at these retention times additionally corresponds to the molecular weight of the expected natural product. The homo-DMADP product has a very similar fragmentation pattern to 7-*epi*-zingiberene (A) with the same base and abundant ions, whilst the mass ion at $m/z = 218.3$ reflects the additional methyl group in the tail of the molecule, so it could reasonably be assumed the structure of the product is as pictured. The two analogues arising from the (*Z*)-2-hexen-ODP (B) and nor-DMADP (D) have high abundance of the 7-*epi*-zingiberene indicative ions at $m/z = 119$ and 93 (discussed in Section 2.5.3) but in slightly different ratios, implying the ion stability or ease of formation might be altered by the new functionality. Despite this, from the overall pattern, it is still likely that the product for each is also as given. NMR spectroscopic characterisation could be used to conclusively determine the structures but would require more compound. Reactions carried out at a larger scale to produce sufficient quantities of the analogues for attempting this analysis and providing samples for bioactivity testing is described in Section 5.5.

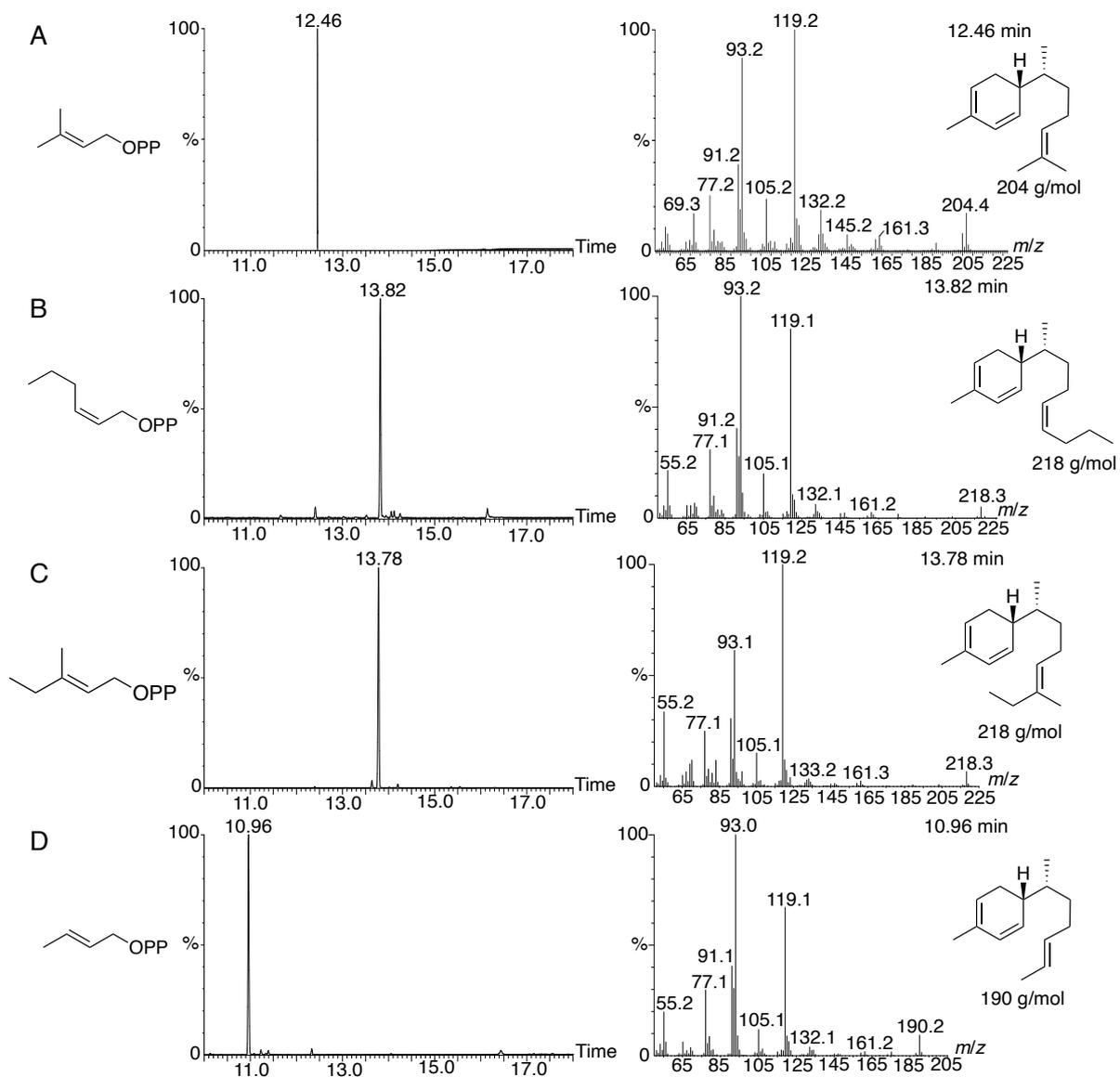


Figure 5.11: Total ion chromatograms and related mass spectra for products resulting from incubations of (A) DMADP, (B) *cis*-hexen-2-ODP, (C) homo-DMADP and (D) nor-DMADP with IDP, zFPS and EZS. Expected products and their molecular weights are given on the right of each mass spectrum for comparison to 7-*epi*-zingiberene in (A).

These results show that EZS, like other terpene synthases, has a degree of plasticity and potential to expand its substrate scope. In each case, the modified position of the DMADP derived substrate analogue is found in the tail of the molecule and not involved in the 1,6-cyclisation mechanism. Therefore, perhaps unsurprisingly, it appears that EZS can tolerate changes in that part of the molecule. As an extra analogue was detected compared to the zFPS screening, it may imply that zFPS has the more limited scope of the two enzymes.

5.4.2 (*E*)- β -Farnesene analogues

One analogue of (*E*)- β -farnesene was identified by GC-MS analysis of the EBFS screening reactions (Figure 5.12), arising from the homo-DMADP analogue giving 12-methyl-FDP. The control incubation with native substrates (A) was authenticated by comparison to a genuine standard. Again, the increased retention time (13.31 min) of the analogue product can be rationalised by the higher molecular weight precursors relative to the control (11.89 min).

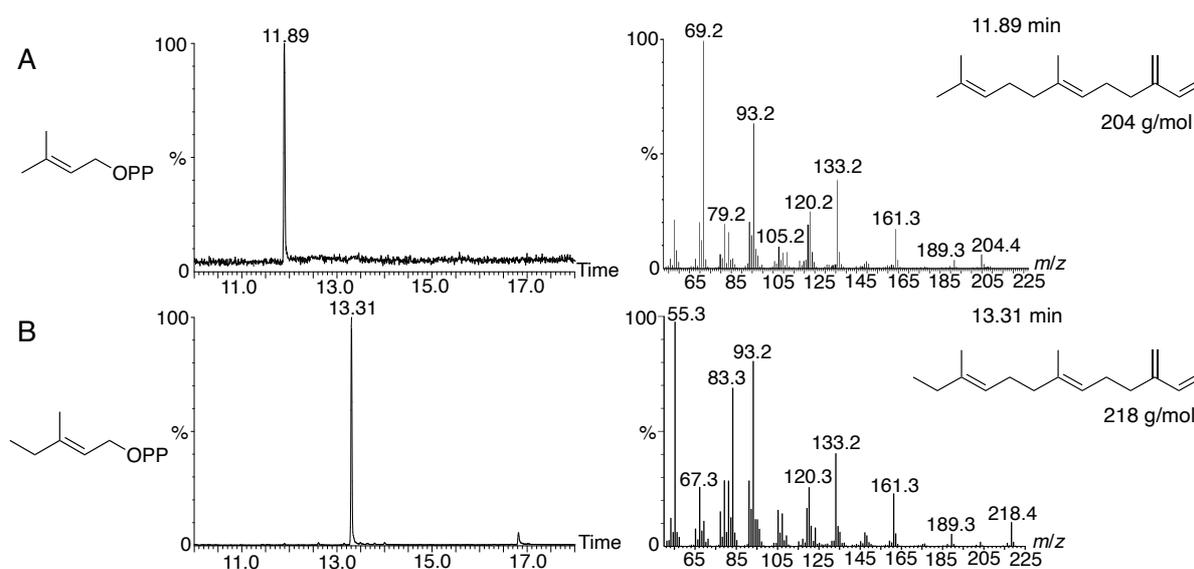


Figure 5.12: Total ion chromatograms and related mass spectra for products resulting from incubations of (A) DMADP and (B) homo-DMADP with IDP, FFPS and EBFS. Expected product and molecular weight of (B) given on the right of the mass spectrum compared to known (*E*)- β -farnesene in (A).

The mass fragment at $m/z = 218.4$ in spectrum B is in accordance with the extra methyl group whilst the rest of the fragmentation pattern is very similar to (*E*)- β -farnesene, differing subtly with the greater abundance of resonance stabilised $m/z = 55.3$ and 83.3 ions compared to the base fragment previously found at $m/z = 69.2$ (Figure 5.13). The $m/z = 83.3$ ion could arise from the cleavage of the tail of the molecule containing the extra methyl group ($C_6H_{11}^+$) which is fragmented further to the $C_4H_7^+$ species giving rise to the new base fragment at $m/z = 55.3$. Further analysis by NMR spectroscopy would conclusively confirm this compound to be 12-methyl-(*E*)- β -farnesene which has not been reported elsewhere in the literature.

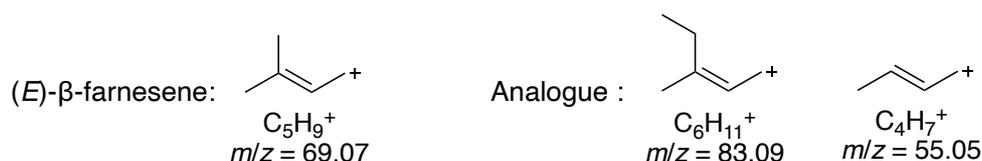


Figure 5.13: Proposed structures of abundant ions in the mass spectra of (*E*)- β -farnesene and the 13-methyl analogue.

As with EZS above, this part of the substrate is not involved in the reaction mechanism of EBFS and the relatively small functionality change is not problematic for substrate turnover as the product was detected at a strong intensity. Compared to the other sesquiterpene synthases discussed, EBFS shows the least promiscuity by only accepting one of the available FDP analogues.

5.4.3 Germacrene D analogues

The recently developed kinases pathway showed it was possible to efficiently make 12-methyl-FDP, 15-methyl-FDP and 14,15-dimethyl-FDP using FPPS,¹⁶¹ all substrates for GDS where the product has been demonstrated to have useful bioactivity toward aphids.¹⁰⁸ The GC-MS results of this screening system with GDS has however revealed a further germacrene D analogue arising from 15-nor-FDP. Control assays show that germacrene D has a retention time of 12.28 min (Figure 5.14) and a distinct fragmentation pattern with base fragment at $m/z = 161.3$ relating to the 10-membered ring fragment which has lost the isopropyl group.

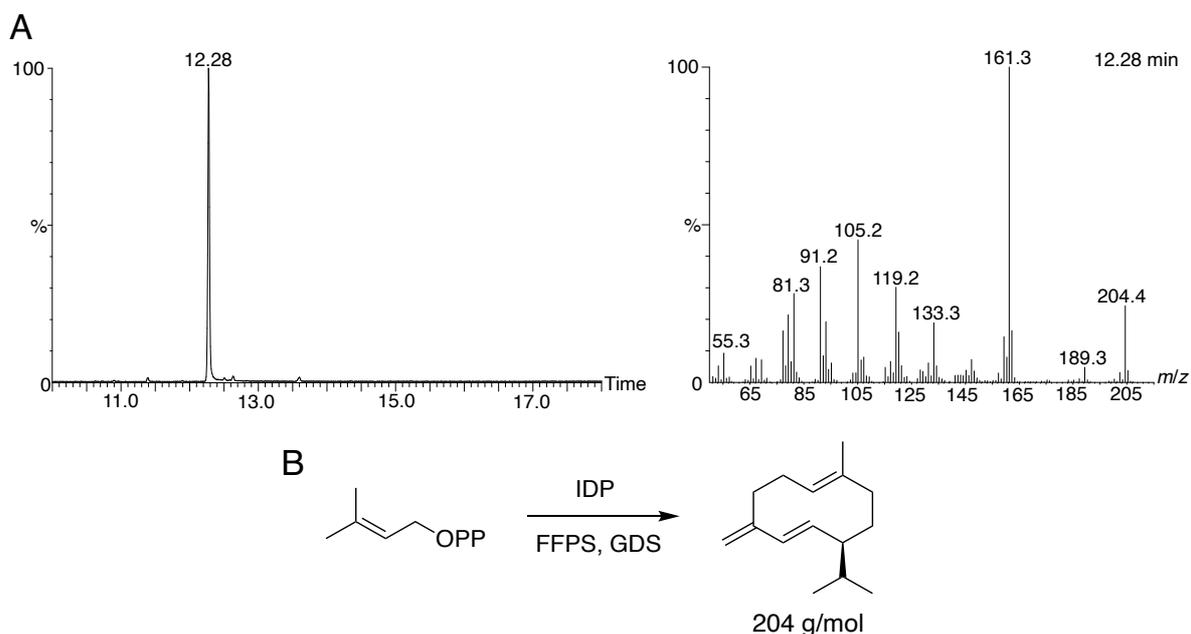


Figure 5.14: (A) Total ion chromatogram of products resulting from the control incubation of DMADP with IDP, FFPS and GDS with mass spectrum of germacrene D eluting at 12.28 min. (B) Scheme showing reaction and product structure (204 g/mol).

Several products were detected from GC-MS analysis of the incubation of GDP with the nor-IDP analogue and GDS (Figure 5.15). The relative intensities of the peaks are fairly low compared to other analogues, perhaps relating to the carbocation stability and favourability of formation of the FDP precursor described above (Section 5.3.3). GC-MS analysis shows that the compound eluting at 11.85 min has the signature germacrene D fragmentation pattern but the mass fragment at $m/z = 190.3$ and base fragment at $m/z = 147.3$ are now consistent with the parent structure lacking a methyl group. However, this has implications for the reaction mechanism as the C15 methyl group which is now absent from the substrate would usually become the exocyclic double bond of germacrene D following proton abstraction (Figure 5.16). This substrate therefore forces the enzyme to take an alternative path.

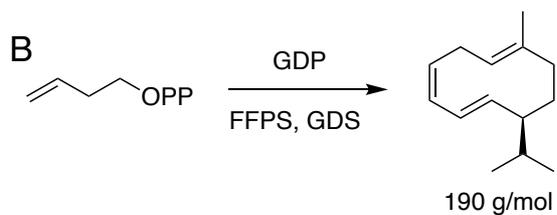
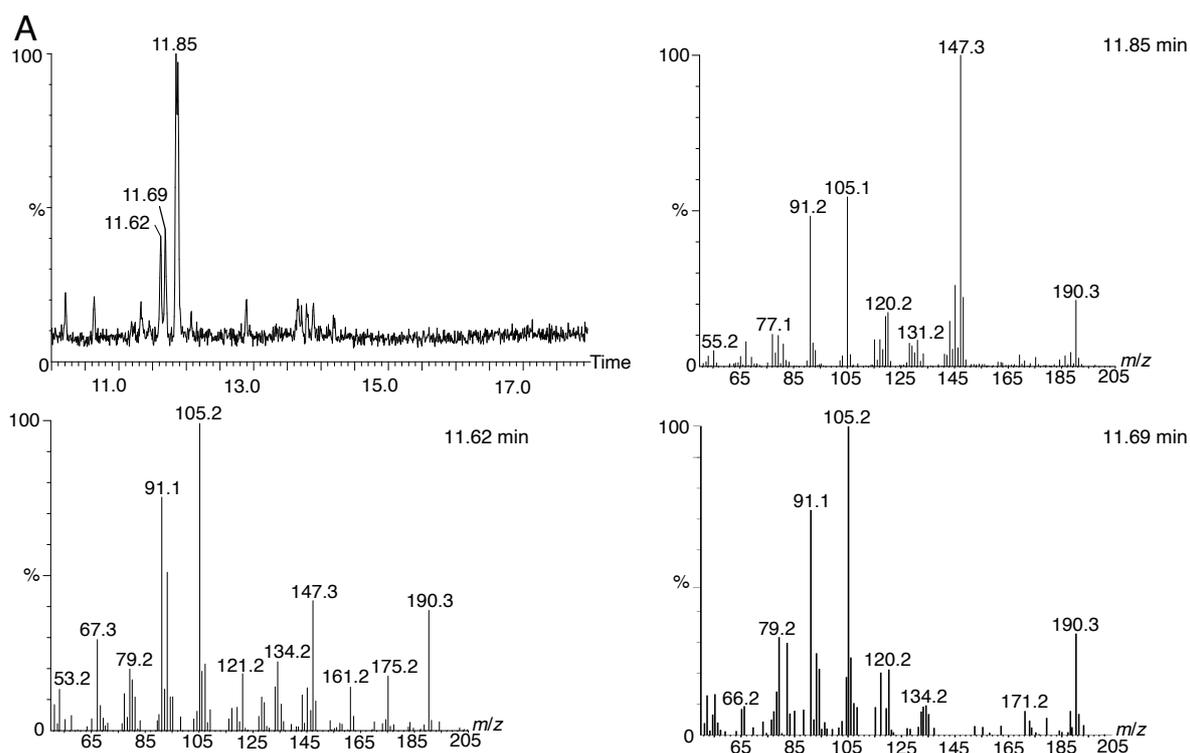


Figure 5.15: (A) Total ion chromatogram of products resulting from the incubation of nor-IDP with GDP, FFPS and GDS accompanied by the mass spectra of the compounds at 11.85, 11.69 and 11.62 min. (B) Scheme showing reaction and expected structure of the compound eluting at 11.85 min (190 g/mol).

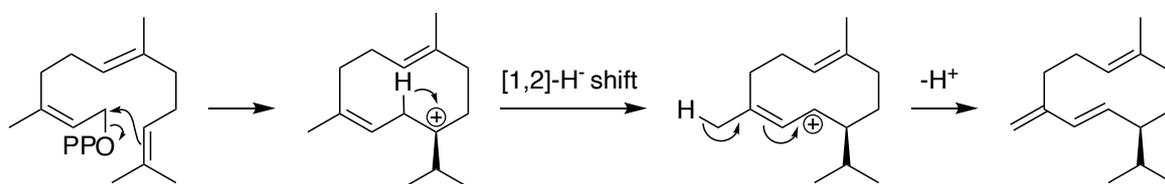


Figure 5.16: Reaction mechanism for the conversion of native substrate FDP into germacrene D by GDS, where the final step involves proton loss from C15.

There are multiple positions from which a proton could be lost from the nor-intermediate equivalent of the second germacradienyl cation (Figure 5.16), several of which are proposed in Figure 5.17. These routes each lead to a different ring double bond configuration in the final product and an argument for which is most likely to form could be made by comparison to the native mechanism. Routes 1 and 3 would lead to resonance stabilised double bonds in the final product but in route 1, the loss of proton leads to conjugated π electron density in the region of the molecule that most represents the equivalent interaction between the native substrate and GDS. In route 1, the isopropyl group remains attached to a sp^3 hybridised carbon (C10), whereas routes 2 and 3 each lead to a sp^2 hybridised C10 which may disrupt the formation of the $m/z = 147.3$ ion. Hence the peak at 11.85 min may be tentatively assigned to the product of route 1.

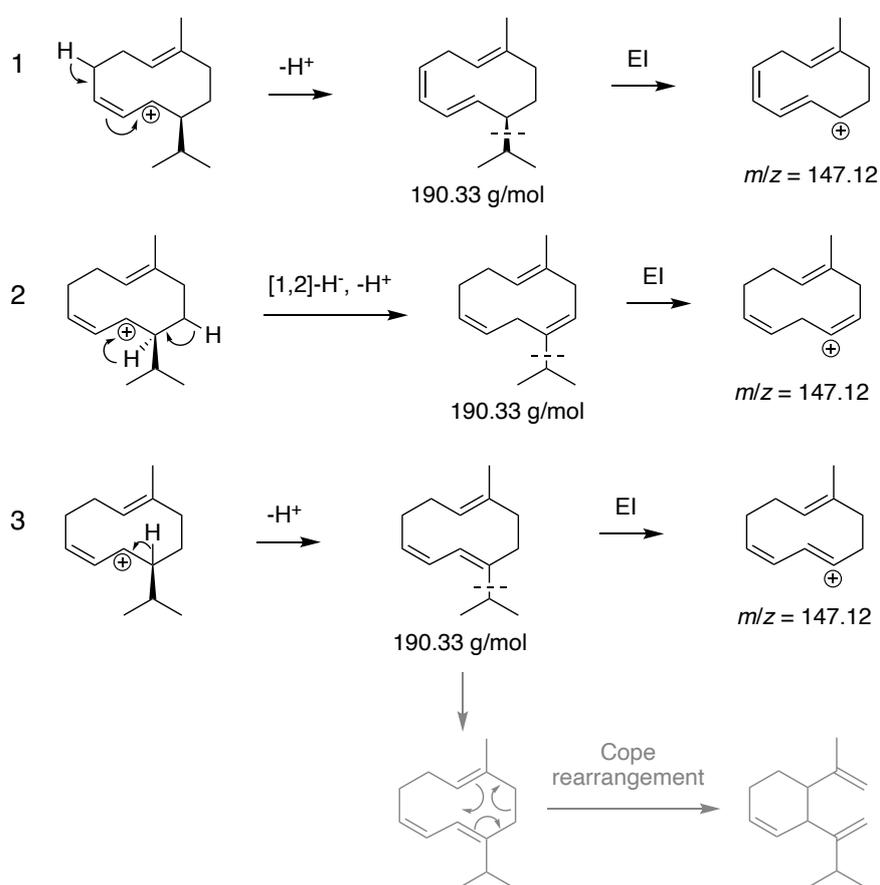


Figure 5.17: Three plausible product forming final steps in the reaction mechanism of GDS with 15-nor-FDP considering proton loss from different ring positions. EI = electron impact ionisation and the dashed line represents fragmentation of the isopropyl group.

Germacrene are frequently found to rearrange via thermally, photochemical or acid-catalysed reactions due to the conformationally flexible structure and labile double bonds.¹²⁷ This type of chemistry, briefly discussed further in Section 6.4.2, may be responsible for the unknown products, a myriad of which are readily accessible via the of the germacrene skeleton. Route 3 for example leads to a product with a configuration prone to undergo a Cope rearrangement (as indicated) which would give an elemene product with a shorter retention time than the cyclododecadiene parent, so may represent a similar rearrangement pathway for the compound at 11.62 min. Lastly, route 2 could be more plausible for the compound at 11.69 min due to the greater abundance of tropylium-like $m/z = 91.2$ and 105.1 ions arising from the triple unsaturation of the ring whilst disrupting $m/z = 147.3$ ion formation. Use of authentic standards of common rearrangement products such as elemenes and cadinenes, and the normalisation of retention times using Kovats indices (KI), would potentially assist further structural identification, however these are much less predictable when arising from these non-canonical substrates.

5.5 Scaled chemoenzymatic synthesis of 7-*epi*-zingiberene analogues

As the GC-MS results show, each sesquiterpene synthase can readily generate unnatural products from substrate analogues. Greater quantities than that produced in the small-scale screening assays was however desired to enable full NMR spectroscopic characterisation and assessment of their biological activity using electroantennography. Due to their availability, alcohol precursors, (*Z*)-2-hexen-1-ol and (*E*)-2-buten-1-ol were taken forward with the kinases, zFPS and EZS to attempt a larger scale synthesis. The third alcohol converted into a 7-*epi*-zingiberene analogue ((*E*)-3-methyl-penten-1-ol) required further synthesis before diphosphate production could be upscaled (see Section 5.5.3).

5.5.1 Scaled reaction for analogue production

The desired analogue diphosphate stocks were produced using fresh preparations of THIM and IPK immobilised on the Ni²⁺ resin (as described in Section 8.2.2) to enable enzyme reuse and stocks to be accumulated (40 mL of 2.3 mM diphosphate stock over 4 days). The reactions were monitored using ³¹P-¹H NMR spectroscopy to ensure full conversion of alcohol to diphosphate before the reactions were regenerated until loss of enzyme activity. Once sufficient stocks had been accumulated, a larger-scale synthesis of each analogue was attempted using zFPS and EZS. Reactions were assembled as previously carried out for 7-*epi*-zingiberene production from the native substrates IDP and DMADP (Section 2.6.1). The reactions were incubated for 48 hours at room temperature with shaking (250 rpm) and

monitored using GC-MS. Pentane portions (5 mL) were then used to extract the reactions exhaustively. The fractions were combined and the concentration of each analogue measured by GC-FID (GC-flame ionisation detection) using a calibration curve established from known concentrations of α -humulene, a C₁₅ sesquiterpene standard. The analogues in question have 14 and 16 carbons but a representative standard for each was not available so the yields discussed herein are approximate.

5.5.2 GC-FID yields

The average peak area of each analogue was calculated from three repeat injections and compared to the calibration curve (Figure 5.28). The C₁₄ nor-analogue preparation was found to contain approximately 0.533 mg (2.8 μ mol) of compound which represents an overall yield of 2.4%. This was lower for the C₁₆ homo-analogue as approximately 0.113 mg (516 nmol) had been synthesised, giving a yield <0.5%. From the same scale reaction with the native substrates IDP and DMADP with zFPS and EZS it was possible to isolate pure 7-*epi*-zingiberene and conduct NMR spectroscopic analysis. However, with yields and quantities of analogue orders of magnitude below the native incubation, compound isolation and further characterisation would be unachievable without repeating multiple large-scale incubations requiring considerable quantities of pure EZS.

Reactions were also attempted to try and isolate the intermediate (Z,Z)-FDP analogues for NMR spectroscopic analysis using the conditions developed in Section 2.4.2 which are sufficient for producing the (E,E)-FDP analogues,¹⁶¹ but this was also unsuccessful. The low yields in both cases indicate that the diphosphate analogues are not favourable substrates for zFPS in particular. As previously discussed (Section 5.3.4), zFPS catalysis is complex and markedly different from FPPS catalysis, ultimately proving less amenable for compound production. For the larger molecular weight analogues, extra functionality could cause repulsive interactions with the hydrophobic active site pocket of the synthase: lower binding affinities would result in lower catalytic activities of the enzyme and a reduced rate of product formation.

5.5.3 Chemical synthesis of 3-methyl-penten-1-ol

Alongside the attempts to upscale analogue production, a synthesis of the third prenil analogue successfully turned over to a 7-*epi*-zingiberene analogue ((E)-3-methyl-penten-1-ol) was carried out. It was made initially via a synthesis devised by Dr. Robert Mart (Figure 5.18) but more was required for an upscaled synthesis of both downstream 7-*epi*-zingiberene and (E)- β -farnesene analogues. The synthesis involves an unselective HWE reaction which

produces both geometric isomers, of which the *Z* had further potential to be used as an analogue precursor in this system. A master's student in the Allemann group, Ashley Ramdass, kindly attempted the synthetic work described herein to produce both alcohols.

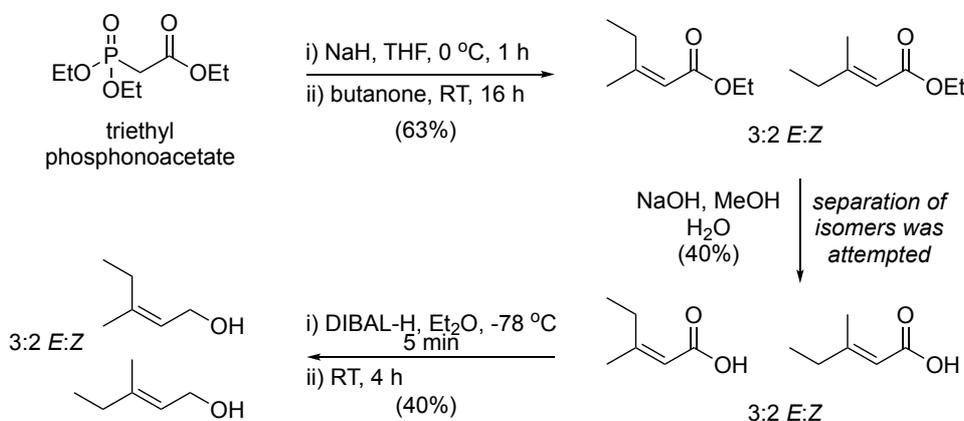


Figure 5.18: Synthetic route to produce both geometric isomers of 3-methyl-penten-1-ol.

The first step in the synthesis involved double bond formation using a HWE reaction between triethyl phosphonoacetate and 2-butanone. Although a moderate yield was achieved, a 3:1 ratio of ester isomers could be observed in the resulting ¹H NMR spectrum (Section 8.3.7). A hydrolysis reaction was then performed with separation of the resulting acid isomers attempted using silica chromatography. Despite various troubleshooting attempts such as extending the column, lowering the concentration gradient of the solvents and reducing the fraction collection size, it was not possible to separate the *Z* isomer from the mixture. The acids were then reduced using DIBAL-H to produce a mixture of both (*Z*)- and (*E*)-3-methyl-penten-1-ol. A small-scale synthesis of 7-*epi*-zingiberene analogues was attempted from the alcohol mixture with the kinases, zFPS and EZS to test the *Z* isomer as a substrate. Two peaks representative of the respective analogues were present following GC-MS analysis of the final organic assay products (Supplementary Figure 5.29) indicating that the *Z* prenol analogue was in fact a substrate for THIM and subsequently IPK, zFPS and EZS. Using a more optimal synthesis to produce both alcohols, for example as described in the recent publication by Hou *et al.*,¹⁹³ could be worth pursuing for chemoenzymatic synthesis of 12- and 13-methyl-FDPs and related analogues.

5.6 Summary

The newly developed *in vitro* kinases diphosphorylation method has been used to generate a library of analogues of the universal terpene precursors DMADP and IDP, subsequent screening of which with prenyltransferases and terpene synthases has successfully generated novel analogues of (*Z,Z*)- and (*E,E*)-FDPs and important semiochemicals 7-*epi*-zingiberene, (*E*)- β -farnesene and germacrene D. With a focus on 7-*epi*-zingiberene in this thesis, attempts were made to upscale the chemoenzymatic analogue syntheses and, whilst small amounts sufficient for further bioactivity testing with insects were produced, the reactions were low yielding, preventing compound isolation and characterisation. If insect testing revealed novel bioactivities, the syntheses could be optimised which may involve mutation of enzymes in the pathway, i.e., zFPS, to increase the favourability of unnatural substrates.

The *in vitro* system is however suitable for rapidly producing and screening isoprenoid diphosphates that would otherwise only be accessible through lengthy and difficult syntheses, with no guarantee that the substrates will be turned over by terpene synthases. Where the intermediate diphosphates can be readily made, as shown by the isolation of 12-methyl-, 15-methyl- and 14,15-dimethyl-(*E,E*)-FDPs, the chemoenzymatic assembly represents a highly efficient alternative to the previously reported chemical syntheses.¹⁶¹ The system has great potential for further development by expanding the library of alcohol precursors and utilising any terpene synthase of interest, including for monoterpene and diterpene production through the inclusion of the relevant prenyltransferases. Natural P450-oxidised derivatives of 7-*epi*-zingiberene (Figure 5.19) have recently been reported which confer considerable toxicity to whiteflies, spider mites and other pest organisms,²⁰⁰ proving that the vast chemical space available from terpene precursors has great potential to provide compounds suitable for pest control.

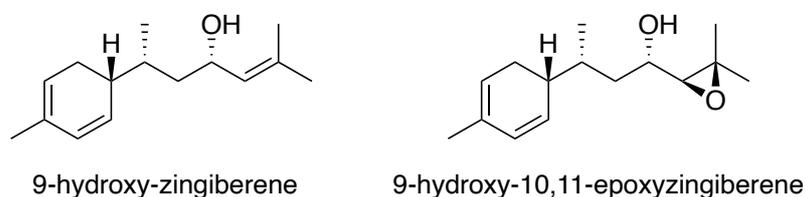


Figure 5.19: Recently discovered natural analogues of 7-*epi*-zingiberene occurring in the glandular trichomes of *Solanum habrochaites* produced by successive cytochrome P450 oxidations of the parent structure.²⁰⁰

5.7 Supplementary

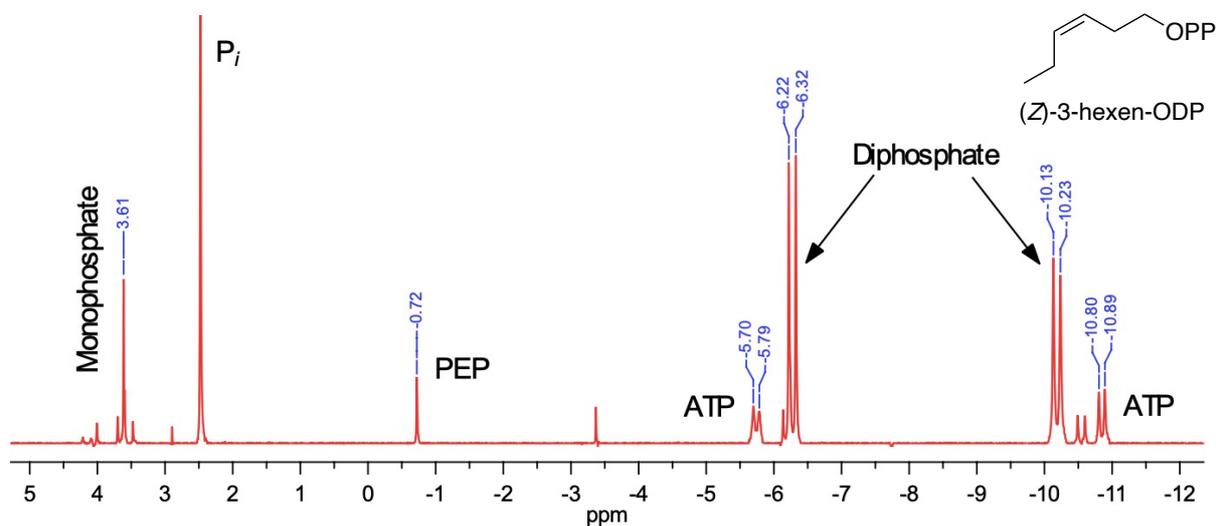


Figure 5.20: $^{31}\text{P}\{-^1\text{H}\}$ NMR spectrum of the chemoenzymatic phosphorylation of (Z)-3-hexen-1-ol.

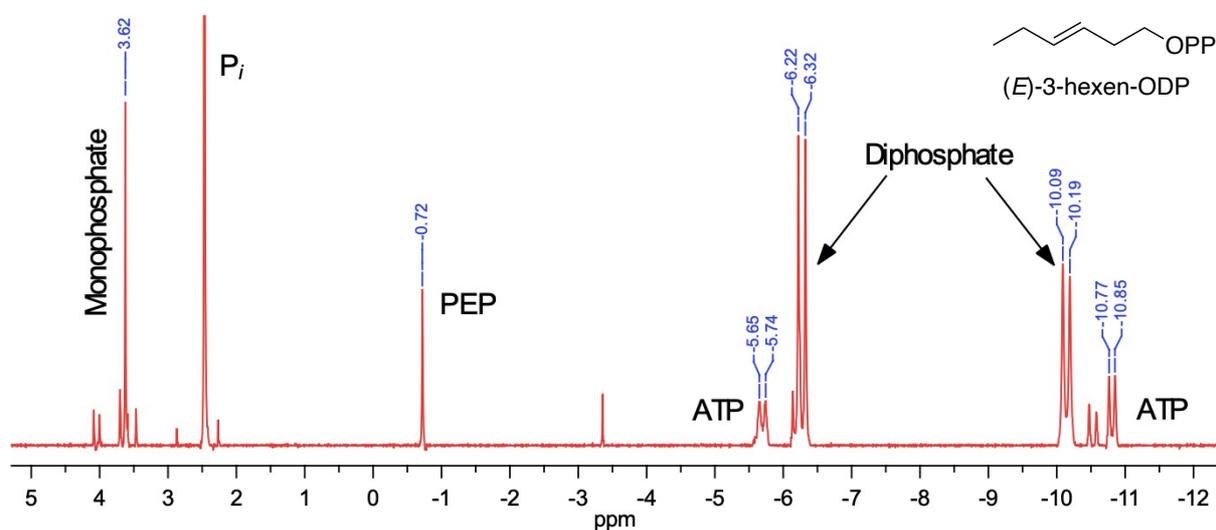


Figure 5.21: $^{31}\text{P}\{-^1\text{H}\}$ NMR spectrum of the chemoenzymatic phosphorylation of (E)-3-hexen-1-ol.

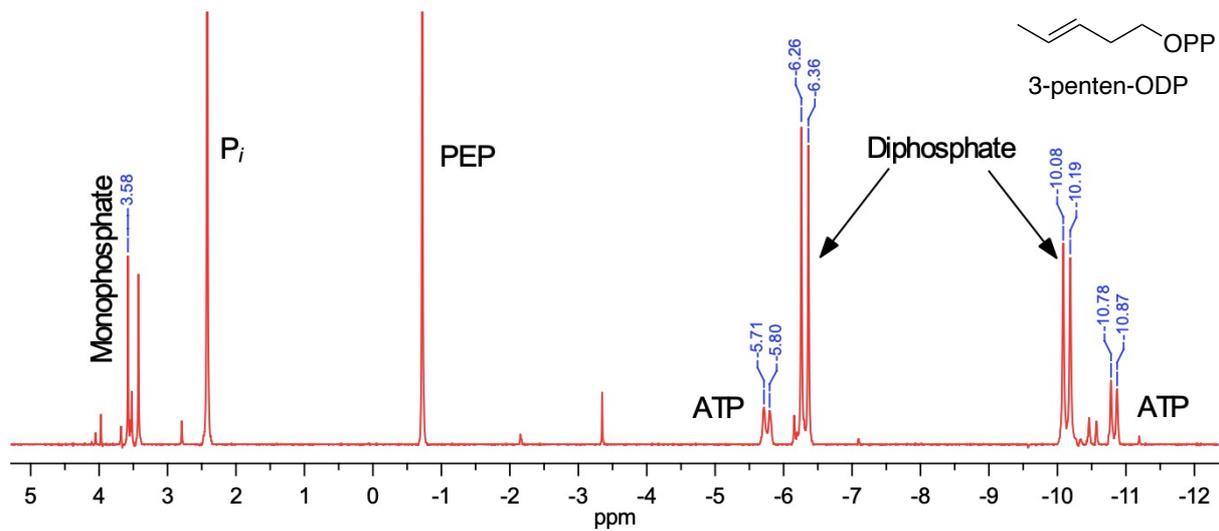


Figure 5.22: $^{31}\text{P}\{-^1\text{H}\}$ NMR spectrum of the chemoenzymatic phosphorylation of 3-penten-1-ol.

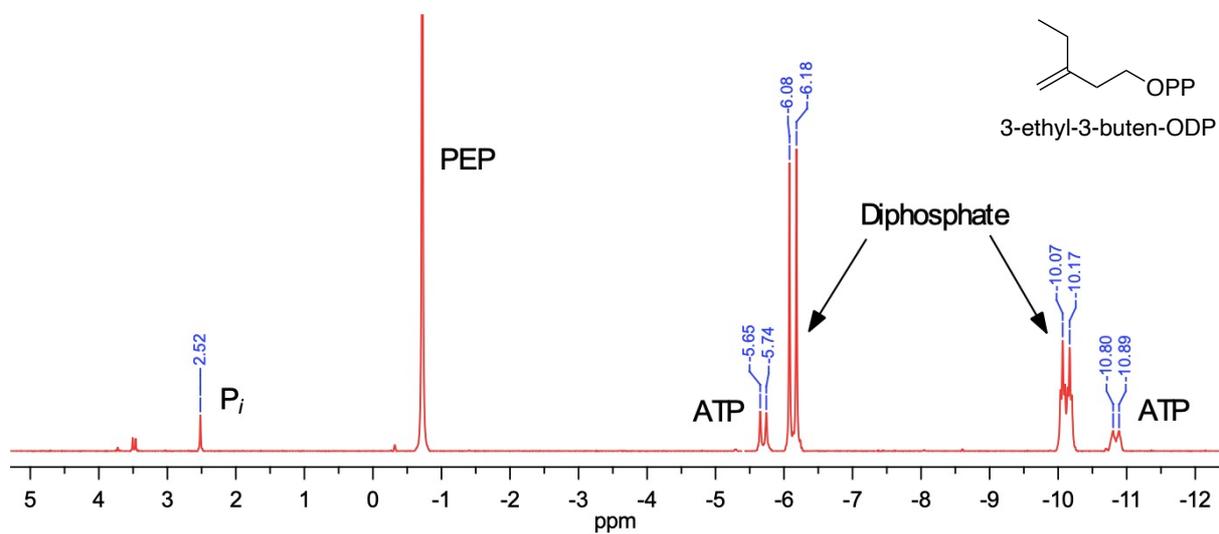


Figure 5.23: $^{31}\text{P}\{-^1\text{H}\}$ NMR spectrum of the chemoenzymatic phosphorylation of 3-ethyl-3-buten-1-ol.

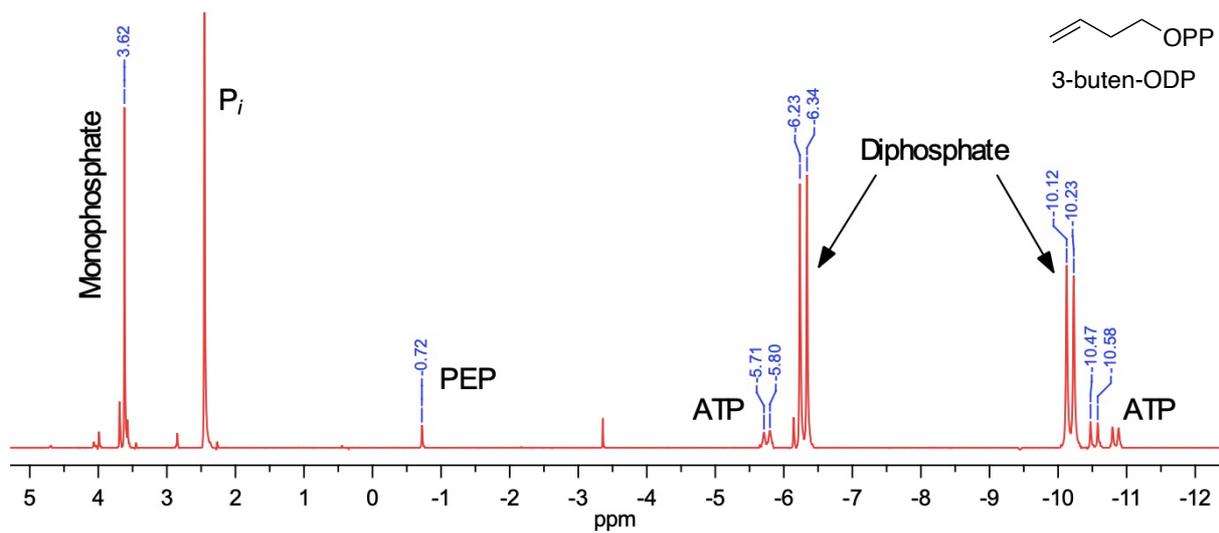


Figure 5.24: $^{31}\text{P}\{-^1\text{H}\}$ NMR spectrum of the chemoenzymatic phosphorylation of 3-buten-1-ol.

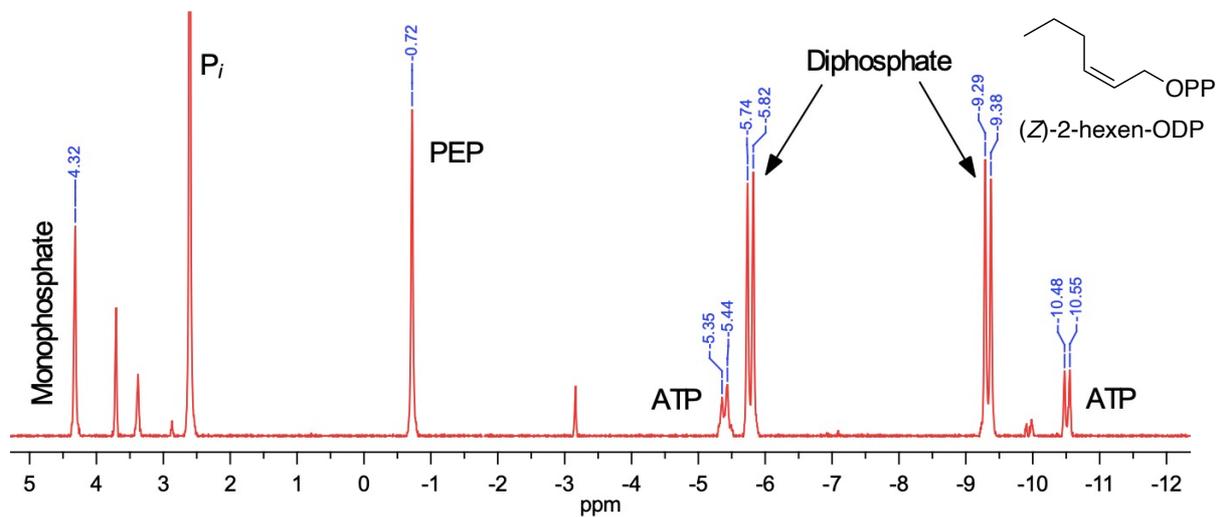


Figure 5.25: $^{31}\text{P}\{-^1\text{H}\}$ NMR spectrum of the chemoenzymatic phosphorylation of (Z)-2-hexen-1-ol.

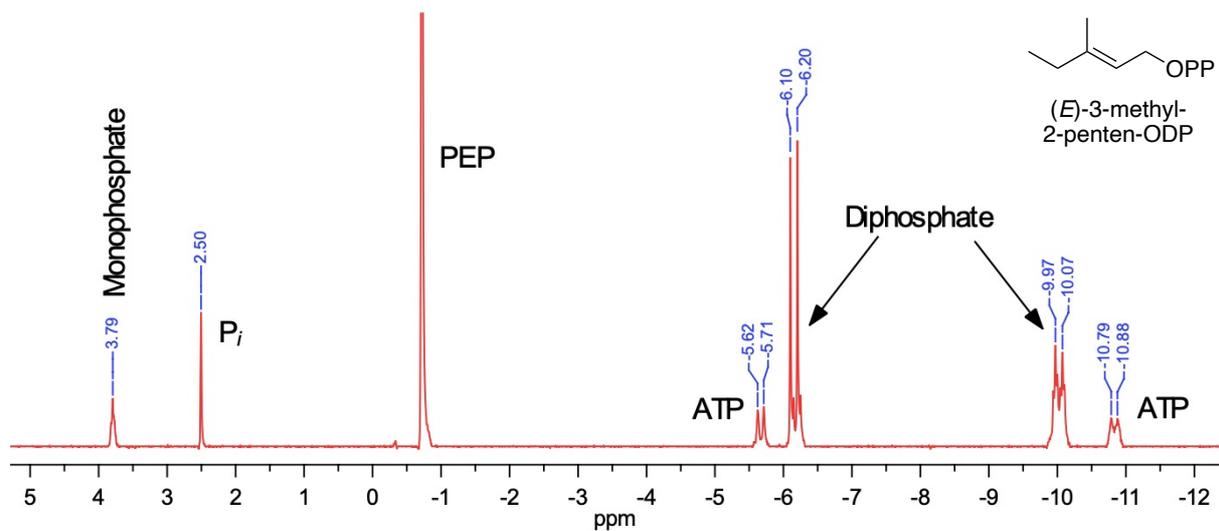


Figure 5.26: $^{31}\text{P}\{-^1\text{H}\}$ NMR spectrum of the chemoenzymatic phosphorylation of (E)-3-methyl-2-penten-1-ol.

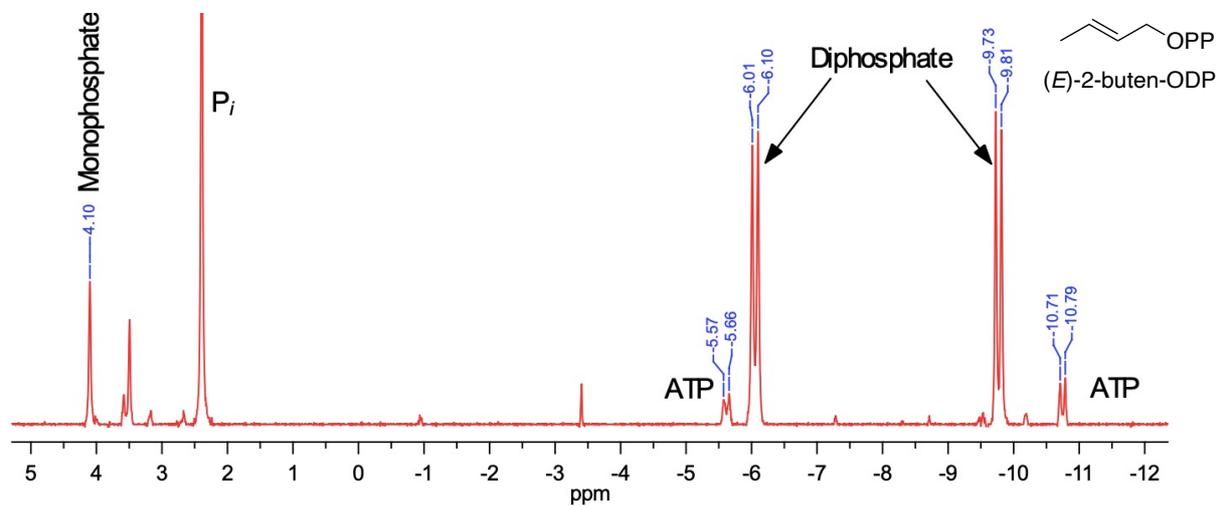


Figure 5.27: $^{31}\text{P}\{-^1\text{H}\}$ NMR spectrum of the chemoenzymatic phosphorylation of (E)-2-buten-1-ol.

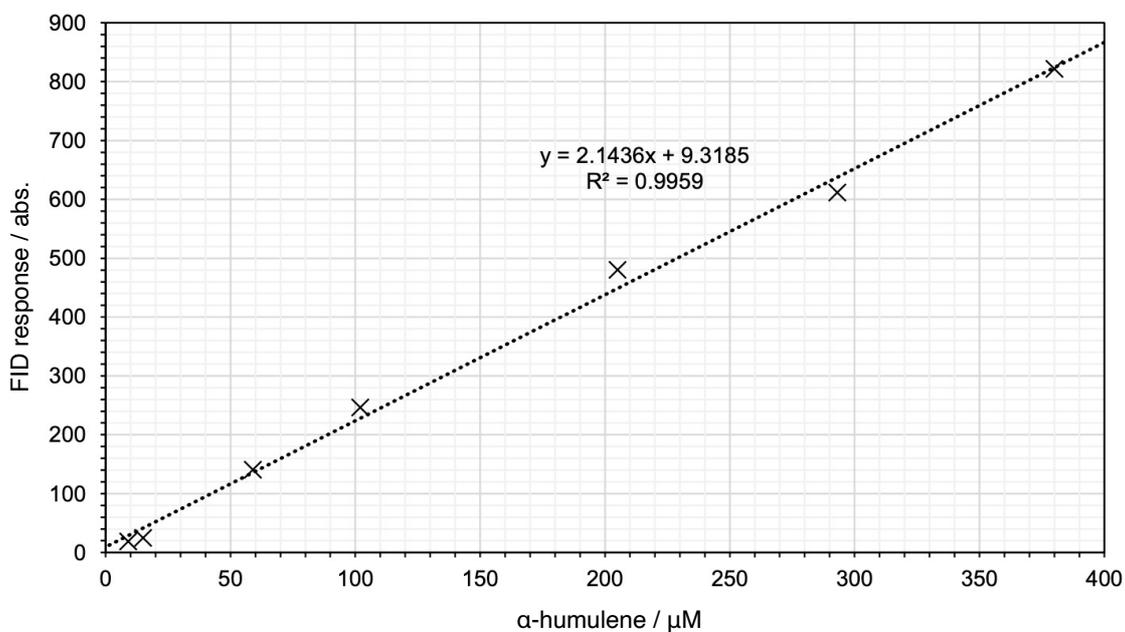


Figure 5.28: GC-FID calibration curve generated from varying known concentrations of a standard of α -humulene.

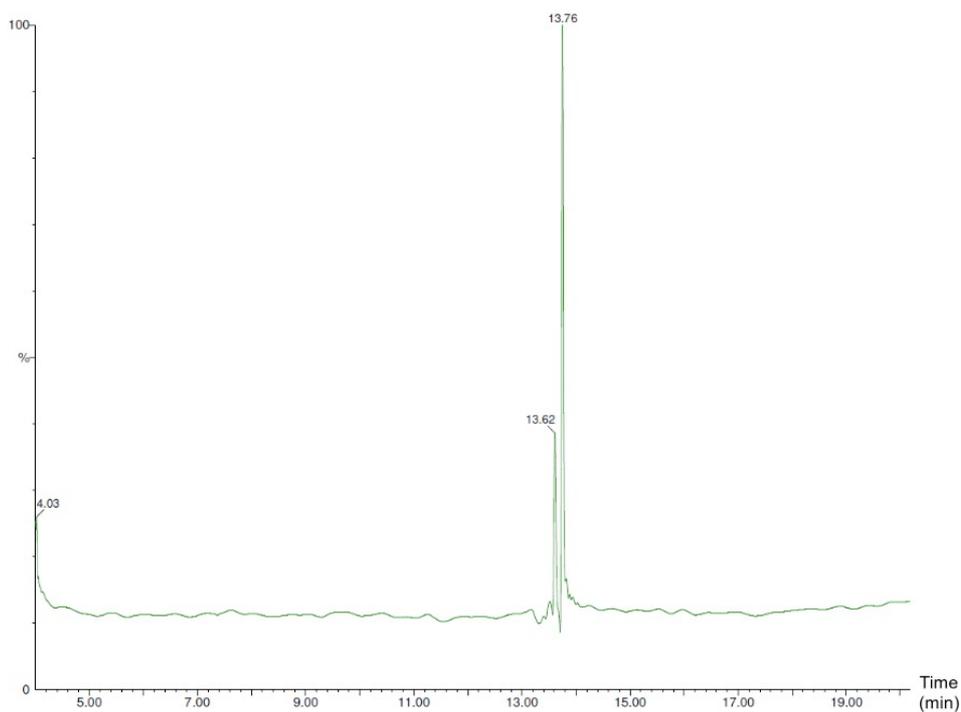


Figure 5.29: Total ion chromatogram produced by Mchem student, Ashley Ramdass, of the pentane extractable products resulting from the incubation of an *E/Z* mixture of 3-methyl-penten-1-ol with THIM, IPK, zFPS and EZS. The two compounds are anticipated to be 12-methyl (13.62 min) and 13-methyl (13.76 min) analogues of 7-*epi*-zingiberene.

Ecological Applications of Sesquiterpenes for Aphid Pest Management

6.1 Introduction

Insect ecology is partly mediated by chemical signals known as semiochemicals, more specifically called pheromones when communicated between members of the same species or allelochemicals when directed to other species.¹¹⁹ Semiochemicals can elicit a wide variety of behavioural responses which serve many functions such as attracting mates, locating food and habitats, warning of predators and mediating changes in development.¹²³ There has been significant interest in using plant and insect semiochemicals for protecting crops from insect pests. Although synthetic insecticides have proven very efficient, negative effects on human health and the environment as well as the increasing problem of insecticide resistance from their continued and widespread use has fuelled the need to develop more sustainable control methods. Naturally occurring semiochemicals are typically non-toxic at the levels deployed but can be used for luring pests into traps, mating disruption and population monitoring for more efficient and sustainable use of insecticides.¹²⁶

Aphids are a large group of highly abundant insects and are important global agricultural pests.²⁰¹ They are particularly pervasive and difficult to control due to their ability to rapidly reproduce both sexually and asexually, the complexity of their life-cycles, their seasonal movement between host plants, as well as causing crop damage from direct feeding and vectoring many plant viruses.²⁰² Many semiochemicals which mediate aphid ecology are members of the terpene family,²⁰³ such as the sesquiterpene (*E*)- β -farnesene, which is the primary or only component of the alarm pheromone for many aphid species.²⁰⁴ It is secreted by aphids under attack from predators, causing those nearby to disperse or detach from their leaves and secrete more in response. Suggested uses for (*E*)- β -farnesene crop protection include increasing aphid mobilisation for better uptake of insecticides, allowing more sustainable and targeted use and to increase susceptibility to attack by natural predators and parasitoids such as ladybirds and parasitic wasps, which show increased foraging in response to (*E*)- β -farnesene,²⁰⁵²⁰⁶ as well as fungal pathogens.²⁰⁷

Although naturally occurring terpenes such as (*E*)- β -farnesene have strong potential for aphid management, they can be highly volatile and unstable to air and light; inherent problems which have so far hampered their development as tools for larger-scale crop protection. Therefore, the way they are formulated and contained can be very important for controlling release and maintaining stability in the field.^{105, 209} Devices such as permeable plastic sachets, vials and rubber septa can be used to release compounds in a predictable way, where the rate can be moderated by the surface area and thickness of the material. Semiochemicals can be formulated with anti-oxidants and UV-stabilisers to improve stability,²⁰⁸ as oil blends and natural plant extracts, however the presence of even trace amounts of other compounds can have significant effects on the elicited response.¹²² Production costs of pure compounds in

large quantities are often prohibitively high as they are typically extracted in mixtures from low yielding natural plant sources and are challenging to produce synthetically due to their complex stereochemical nature.⁸⁷

Germacrene D is another plant sesquiterpene with known repellence toward aphids.¹⁰⁵ Recent work testing the hypothesis that novel substrates accepted by its synthase, GDS, may cover enough of the chemical space of the natural ligand to elicit a behavioural response very interestingly showed that the analogue 14,15-dimethylgermacrene D is a strong aphid attractant.¹⁰⁸ This discovery has prompted the desire to develop the two chemicals into a push-pull crop protection system, whereby aphids can be repelled from crops and simultaneously attracted elsewhere; a technique proven highly successful in other applications.¹¹⁰

The sesquiterpene *7-epi*-zingiberene, the primary focus of this thesis, and its natural derivative by oxidation (*R*)-curcumene, were first isolated and identified from leaf extracts of the wild tomato *Solanum habrochaites* and were shown to repel tobacco whitefly, *Bemisia tabaci*, and other virulent pests.^{136,137} Alongside aphids, whitefly are significant and economically important insects making the top ten of global crop pests,¹²⁸ and are rapidly evolving resistance to many insecticides.¹³⁰ Unusually, *7-epi*-zingiberene is derived from the precursor (*Z,Z*)-FDP as opposed to its all *trans* isomer from which the majority of known sesquiterpenes originate.³⁴ Study of its interesting biosynthesis and behavioural effects may lead to novel natural products with potential for sustainable crop protection against pests, in a similar way to germacrene D.

6.1.1 Aims

This chapter details an investigation into different aspects of the aforementioned repellent sesquiterpenes for their use in mediating pest insect behaviour. Using aphid behavioural bioassays, (*E*)- β -farnesene produced by an industrial manufacturing process (Bedoukian EBF) was assessed for its suitability as a commercial product by comparison to an in-house supply (RRes EBF) of known efficacy. The behavioural responses of aphids to the established whitefly repellent semiochemicals, *7-epi*-zingiberene and curcumenes, were also assessed for the first-time. In addition, the release rate and stability of germacrene D formulated in clear and black polyvinyl chloride (PVC) sachets was investigated to characterise its performance when exposed to light and air in a glasshouse environment. This information will be used to inform future development work of a push-pull crop protection system based on using repellent germacrene D and attractant 14,15-dimethylgermacrene D.

6.2 Characterisation and behavioural testing of (*E*)- β -farnesene

6.2.1 Analytical characterisation

Analysis by NMR spectroscopy and GC showed that the two preparations of (*E*)- β -farnesene (RRes and Bedoukian EBF) were nearly identical, including the trace impurities which are expected to be farnesene isomers from tentative GC-MS identification. This suggests that the synthetic routes or at least the purification methods employed in producing the compound are very similar. The identity and abundance of trace compounds in semiochemical cues can have a significant impact on the behaviour of the organism, even if the major constituent is chemically identical, and organisms are known to respond differently to samples produced directly by the plant or from synthetic sources.¹²² It is therefore important to produce exceptionally pure semiochemicals or to understand fully the effect of different components of a mixture. Confirmed to be of satisfactory purity, behavioural testing was subsequently carried out using the industrially sourced (*E*)- β -farnesene.

6.2.2 Behavioural bioassays

Four-arm olfactometry is one of two behavioural bioassays used to analyse and compare the effects of the RRes and Bedoukian EBFs. This technique was first described by Petterson and is used routinely to elucidate the role of olfactory signals independently from other cues such as visual cues which an organism may use in mediating ecology.²⁰⁹ The four-arm olfactometer (Figure 6.1) uses an airstream to draw air through four inlets to a sealed chamber. The inlets contain samples of the semiochemical, solvent control or air alone. An aphid is then placed inside the chamber where it can move freely and the time it chooses to spend (from a total of 4 min) in each arm of the olfactometer is recorded and statistically analysed. The bioassays were carried out with two representative pest aphid species, grain aphid, *Sitobion avenae*, and peach-potato aphid, *Myzus persicae*.

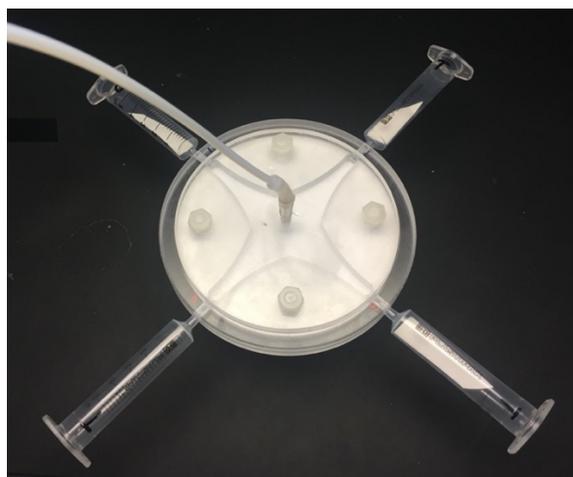


Figure 6.1: A four-arm olfactometer connected to a vacuum pump via the top inlet with four inlets around its circumference where air is drawn across filter paper which are either dosed with compound or solvent as controls.

A more detailed description of the experimental method and data analysis is given in Section 8.4.2. The results summarised in Table 6.1 show that both the RRes and Bedoukian EBF produced a significant repellent effect toward both aphid species. For both compounds, *M. persicae* was repelled more strongly than *S. avenae* with *P*-values of <0.001 and 0.002 compared to 0.006 and 0.030 for RRes and Bedoukian EBF respectively. The repellent effect of each chemical was similar within the species but the RRes EBF had the strongest effect in both cases.

Table 6.1: Response of *M. persicae* and *S. avenae* and to EBF in the four-arm olfactometer with the mean time spent in the control and treated arms compared using a paired *t*-test.

Aphid species	EBF treatment	Conc. / $\mu\text{g}/\mu\text{L}$	Mean time spent in treated arm / min	Mean time spent in control arm / min	<i>P</i> -value*	Reps.
<i>Sitobion avenae</i>	RRes	0.8	2.49	4.51	0.006	9
	Bedoukian	0.8	2.53	4.48	0.030	10
<i>Myzus persicae</i>	RRes	0.8	1.06	4.99	<0.001	10
	Bedoukian	0.8	1.40	4.88	0.002	10

* Paired *t*-test comparing treated and control means. Significance: *P* < 0.05.

The second behavioural bioassay used to assess the efficacy of the different (*E*)- β -farnesene batches was the aphid alarm pheromone bioassay (Figure 6.2). This assay was carried out with alate (wingless) *M. persicae* rather than the winged form used for the four-arm olfactometry, reared on the underside of a single leaf of individual Chinese cabbage plants (Section 8.4.2). During the assay, the colony is exposed to (*E*)- β -farnesene and the number of aphids which move after 1 and 15 min are counted and statistically analysed to determine their overall response.

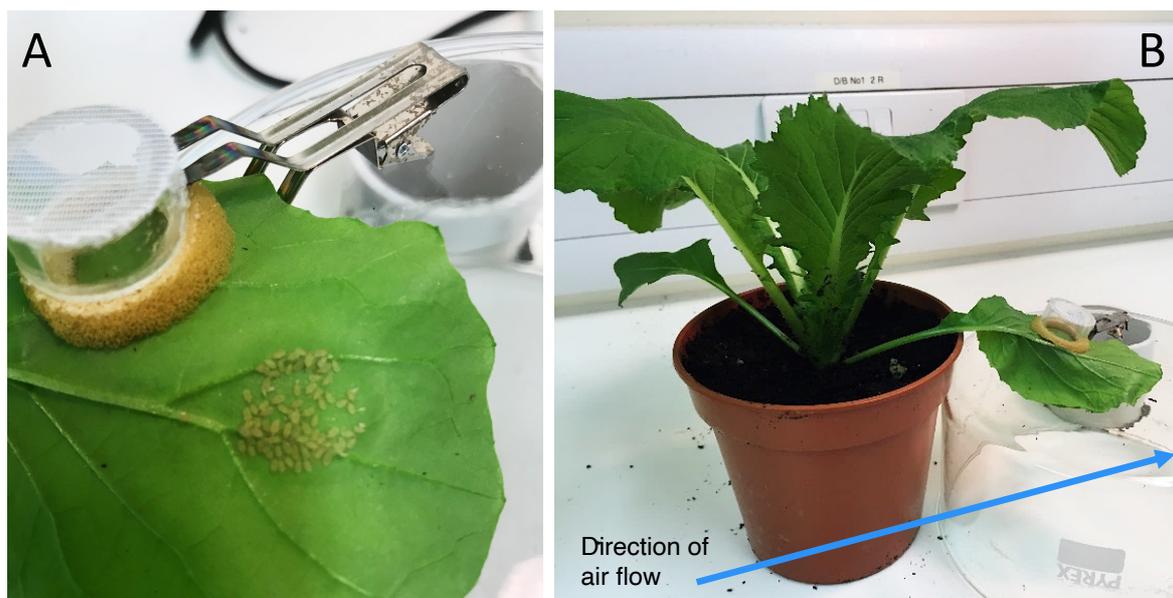


Figure 6.2: Experimental set up for the aphid alarm pheromone bioassay. (A) A colony of apterous *Myzus persicae* reared using a clip cage on the underside of a Chinese cabbage plant leaf. (B) Compound was applied upwind from the colony in the direction of the air flow to ensure exposure to the alarm pheromone.

As shown below in Table 6.2, both the RRes and Bedoukian EBF significantly produced the alarm response behaviour in colonies of *M. persicae* when compared to the treatment of just hexane which was used as a control, with no significant difference between the two according to the LSD test. The response to both EBFs was significant even after only one minute had elapsed after application, whereby most aphids had moved away.

Table 6.2: Summary of results for aphid alarm pheromone bioassay for *M. persicae* with treatments at a concentration of 1 µg/µL.

Treatment	Mean % aphids responding after 1 min (±SE)	Mean % aphids responding after 15 min (±SE)	Replications
Hexane	3.6 (± 2.0) a	7.9 (± 3.1) a	12
RRes EBF	88.9 (± 5.1) b	96.6 (± 1.8) b	11
Bedoukian EBF	88.2 (± 5.7) b	91.8 (± 5.9) b	11

Values in the same column with the same letter do not differ significantly ($P < 0.05$, LSD test). SE = standard error.

These findings confirm the expected significant repellent response of (*E*)-β-farnesene from both sources compared to controls. The results of the two behavioural assays show that the RRes (*E*)-β-farnesene performed marginally better than the Bedoukian EBF, although structural characterisation by NMR spectroscopy and GC-MS analysis revealed no discernible difference between either sample and confirmed both were of very high purity. Additionally, the repellent effect was reproducible in both aphid species and at different stages of *M. persicae* life cycle and development. This suggests that the manufacturing process and subsequent storage is suitable for maintaining its efficacy and support the development of the Bedoukian EBF into a commercial product.

6.3 Behavioural testing of 7-*epi*-zingiberene and (*RS*)-curcumene

In the literature, 7-*epi*-zingiberene and its natural oxidant derivative (*R*)-curcumene have been shown to repel whitefly (*Bemisia tabaci*) and deter spider mites (*Tetranychus urticae*),^{136,137} however behavioural testing with other pest species like aphids has not yet been reported. Using four-arm olfactometry, the behavioural response of *M. persicae* and *S. avenae* toward both semiochemicals were tested (Section 8.4.2). Samples available at the time were 7-*epi*-zingiberene isolated from tomato extracts, and a racemic mixture of curcumene isomers ((*RS*)-curcumene) produced synthetically.

As shown in Table 6.3, *S. avenae* were significantly repelled by both (*RS*)-curcumene and 7-*epi*-zingiberene at treatment levels of 1.0 µg/µL, but there is no clear repellent effect of 7-*epi*-zingiberene when the concentration was lowered to 0.1 µg/µL. With a *P*-value of 0.001, (*RS*)-curcumene is the most repellent for *S. avenae*. Conversely for *M. persicae*, with six

replications, (*RS*)-curcumene did not have a significant repellent effect, although the mean time spent in the treated arm was nearly 2 min lower than the control arm reflecting the variability within the replications. With a *P*-value of <0.001, *M. persicae* gave the most significant repellent response when tested with 7-*epi*-zingiberene.

Table 6.3: Response of *S. avenae* and *M. persicae* to known whitefly semiochemicals in the four-arm olfactometer with the mean time spent in the control and treated arms compared using a paired *t*-test.

Aphid species	Treatment	Conc. / $\mu\text{g}/\mu\text{L}$	Mean time spent in treated arm / min	Mean time spent in control arm / min	<i>P</i> -value*	Reps.
<i>Sitobion avenae</i>	(<i>RS</i>)-curcumene	1.0	1.95	4.68	0.001	10
	7- <i>epi</i> -zingiberene	1.0	2.89	4.65	0.025	14
	7- <i>epi</i> -zingiberene	0.1	3.58	4.12	0.489	12
<i>Myzus persicae</i>	(<i>RS</i>)-curcumene	1.0	2.63	4.44	0.151	6
	7- <i>epi</i> -zingiberene	1.0	1.38	4.87	<0.001	10

* Paired *t*-test comparing treated and control means. Significance: *P* < 0.05.

These results have interestingly shown for the first time that 7-*epi*-zingiberene and curcumene act as repellent signals towards aphids. Aphids are closely related to whitefly therefore may have comparable neurological pathways to detect and respond behaviourally to the chemicals. This revelation provides even more basis for development of 7-*epi*-zingiberene as a commercial crop repellent, and further investigation of its response with aphids which may include electroantennography measurements. Whether the (*S*)-curcumene component of the racemic mixture is also repellent would require examination, as despite both enantiomers being electrophysiologically active in whitefly, only (*R*)-curcumene was found to induce a repellent response.¹³⁶

6.4 Release rate and stability testing of germacrene D

The way a semiochemical is formulated and the device used to release it can have a significant effect on its lifetime and stability in the field. The repellent semiochemical germacrene D and its attractive 14,15-dimethylgermacrene D analogue are under development as a push-pull crop protection system, therefore it is necessary to understand such properties so they can be used effectively in the field following the transition from the lab. The release rate and stability of germacrene D in two types of permeable PVC - one clear and one black - were assessed under different environmental conditions using air entrainment experiments and analytical characterisation. The black PVC was hypothesised to potentially improve the stability of germacrene D by preventing absorption of UV radiation which could lead to degradation.

6.4.1 Release rate

Sachets of both clear and black PVC containing a small (1 cm²) piece of thin sponge dosed with germacrene D (1 µg) were contained in sealed glass vessels (in triplicate) and placed in a greenhouse (Figure 6.3). Volatiles from the samples were collected by air entrainment at weekly time points for a total of 3 weeks using a pump to draw clean, filtered air through an adsorbent material (Porapak Q) from which the compounds could be eluted and analysed by GC (Section 8.4.3). Two separate experiments were carried out approximately 6 weeks apart to assess the PVC release rates under different environmental conditions leading to two sets of results for both the clear and black PVC. Peaks relating to germacrene D were identified in the chromatograms by comparison to a standard and the sample concentration calculated from the peak area using a calibration curve prepared from known concentrations of germacrene D (Supplementary Figure 6.11). These results are summarised in as the total germacrene D accumulated and represented as graphs in Figure 6.4 and Figure 6.5.

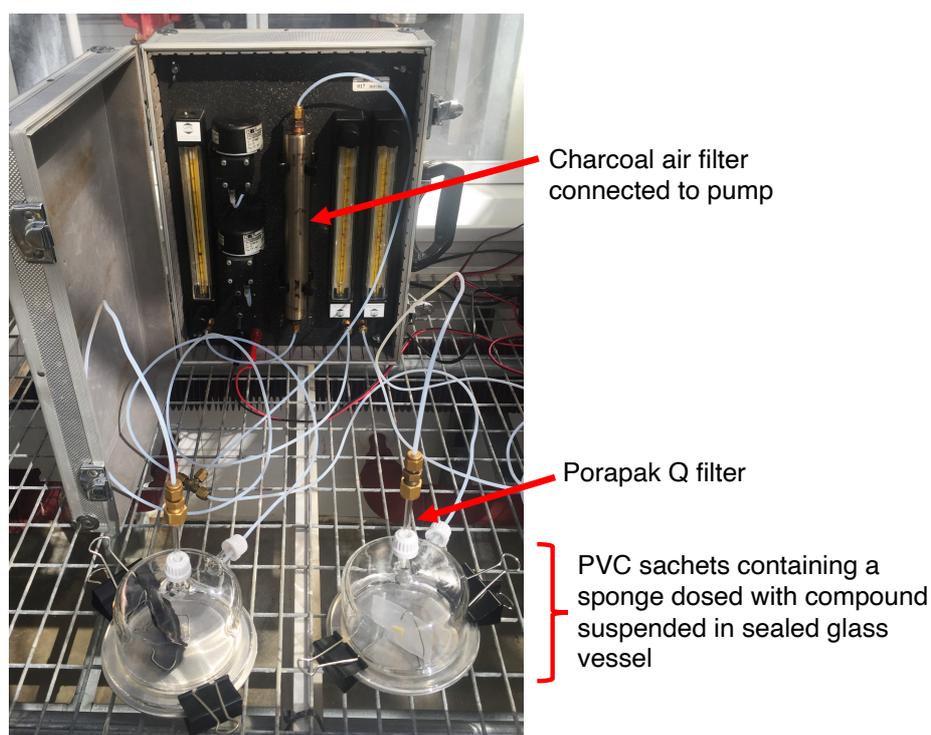


Figure 6.3: Experimental set-up for air entrainment of samples containing (S)-germacrene D enclosed within permeable PVC sachets for studying the rate of release of the compound from the material.

Experiment 1 was carried out in May and over the first 7 days the average release rate for the clear PVC was 0.016 mg/day and for the black PVC it was 0.029 mg/day (Figure 6.4). The clear PVC rate includes a particularly low, potentially anomalous value for one of the replicates where only 5.7 μg of germacrene D was collected by entrainment over the 3 weeks when considering the sponge was initially dosed with around 900 μg (1 μL). For experiment 2, carried out from mid-June, the average release rates over the first 7 days were much lower at 0.014 mg/day for the clear PVC and 0.0019 mg/day for the black PVC (Figure 6.5).

The materials performed similarly within each experiment on average, although considerable sample variation was seen within the replicates resulting in large error margins. The PVC sachets for each experiment were prepared immediately before each experiment using the same method, but the substantial differences in the quantity of germacrene D entrained at each time point imply this type of release device may lack reproducibility. For both experiments, after 7 days, the majority of germacrene D had volatilised therefore, from these results, a 7-day period could be recommended for the application of this dose of compound from this type of PVC under similar conditions. However, considering more air flow and potentially increased volatilisation in the field, this may need to be shortened. More replicates over the time of year relevant to aphid control and a larger loading of germacrene D may help to improve reliability of the results.

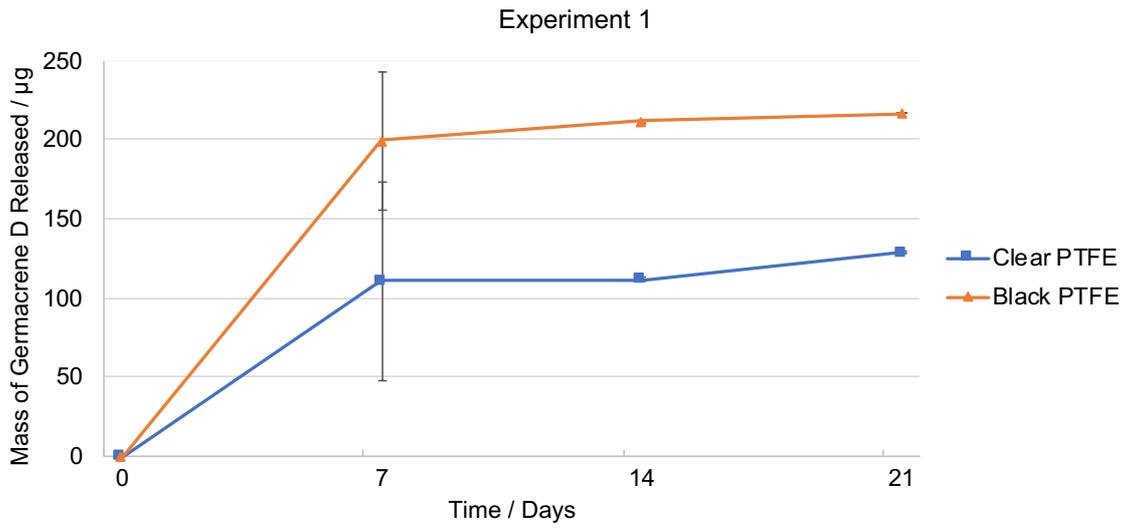


Figure 6.4: Experiment 1 release rate of germacrene D from clear and black PVC (carried out in May).

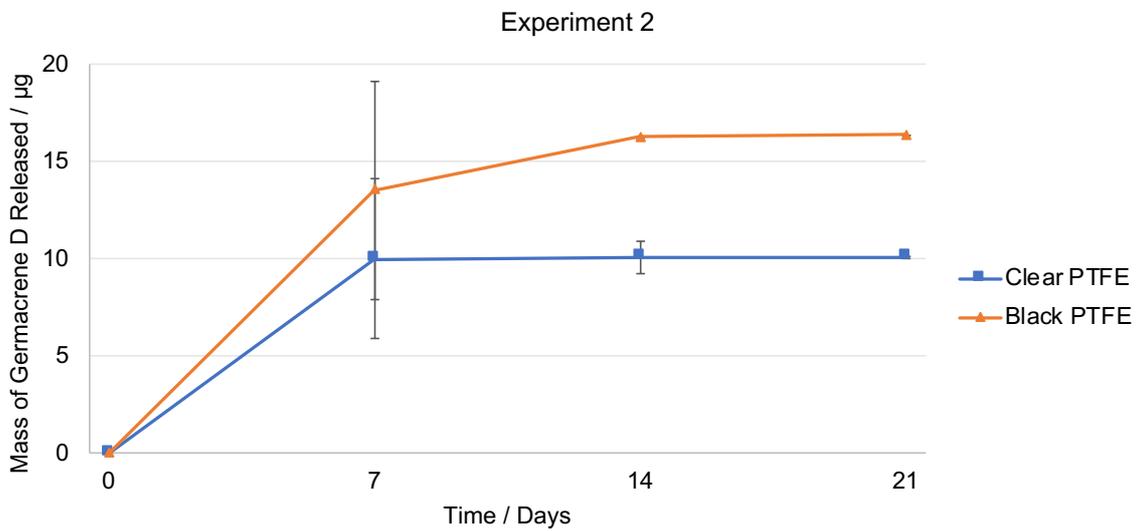


Figure 6.5: Experiment 2 release rate of germacrene D from clear and black PVC (carried out in June).

6.4.2 Stability

The stability of germacrene D when contained in the permeable PVC release device was assessed following exposure to environmental conditions in several ways. Firstly, the GC samples analysed from the air entrainment studies for release rate were also examined regarding their stability profiles, with the relative change in product abundance looked at over weeks 1 - 3. To aid comparison, entrainments of an empty vessel and entrainment of both clear and black sachets which did not contain germacrene D were carried out as controls. The sponges contained within the PVC sachets that were subject to air entrainment (Figure 6.3) were also extracted and analysed by NMR spectroscopy after the second experiment.

Secondly, a separate experiment was set-up alongside the entrainment vessels in the greenhouse whereby PVC sachets containing germacrene D were left exposed to air suspended on wire frames (Figure 6.4). Three sponges from each of the clear or black PVC sachets were extracted at weekly time points to coincide with the entrainment time points (over 3 weeks). These were then analysed by GC to determine whether compound remained on the sponges and was not lost to evaporation.



Figure 6.6: Experimental set-up for stability study where black and clear PVC sachets containing yellow sponges dosed with germacrene D are suspended from wire frames and exposed to air and light in a greenhouse environment.

Entrainment samples

Gas chromatograms of the control entrainment samples where no germacrene D was present revealed the presence of many small peaks relating to compounds which can be directly

attributed to the PVC material itself (Figure 6.7). The clear and black materials also have slightly different profiles. Learning of these results, additional controls were carried out to compare heat sealed and unsealed PVC, but these showed no discernible difference. The presence of these volatiles could be related to the manufacturing and treatment processes used in producing the PVC.

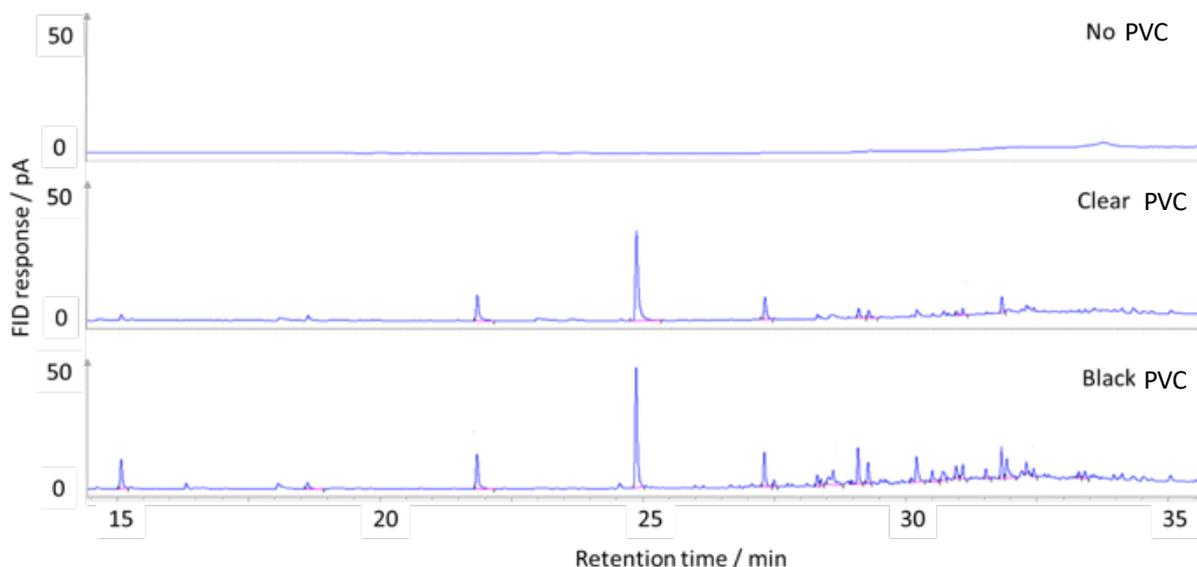


Figure 6.7: Gas chromatograms of controls where no PVC, clear and black PVC were entrained without the sponge dosed with germacrene D. The presence of these compounds are therefore unrelated to germacrene D degradation.

Other various small peaks not seen in the PVC controls are present in all entrainment samples, particularly after the germacrene D peak (29.0 min) with slight variations between the clear and black PVC. These peaks are seen by the first week, generally appearing constant in size however any trends are unquantifiable due to differences in relative sample concentration between replicates. Peak identification by GC-MS analysis was largely inconclusive but fragmentation patterns consistent with alkanes were seen with suggestions of oxidised products whilst KI comparisons were consistent with hydrocarbons.

Example gas chromatograms for volatile samples (Figure 6.8) between release rate experiments 1 and 2 show a clear difference in the relative amounts of germacrene D and any resulting degradants that were recovered; lower by around 85% for experiment 2. They appear similar over the time points within each experiment, with no clear difference in performance between the PVC materials. Caryophyllene (indicated) was previously known to be a small contaminant present in the source of germacrene D used. β -Bourbonene is a known

photochemical product of germacrene D210 and a possible identity of the peak appearing at 27.1 min (KI value of 1391: reference database KI value of 1395). No other compounds were identified using KI comparison.

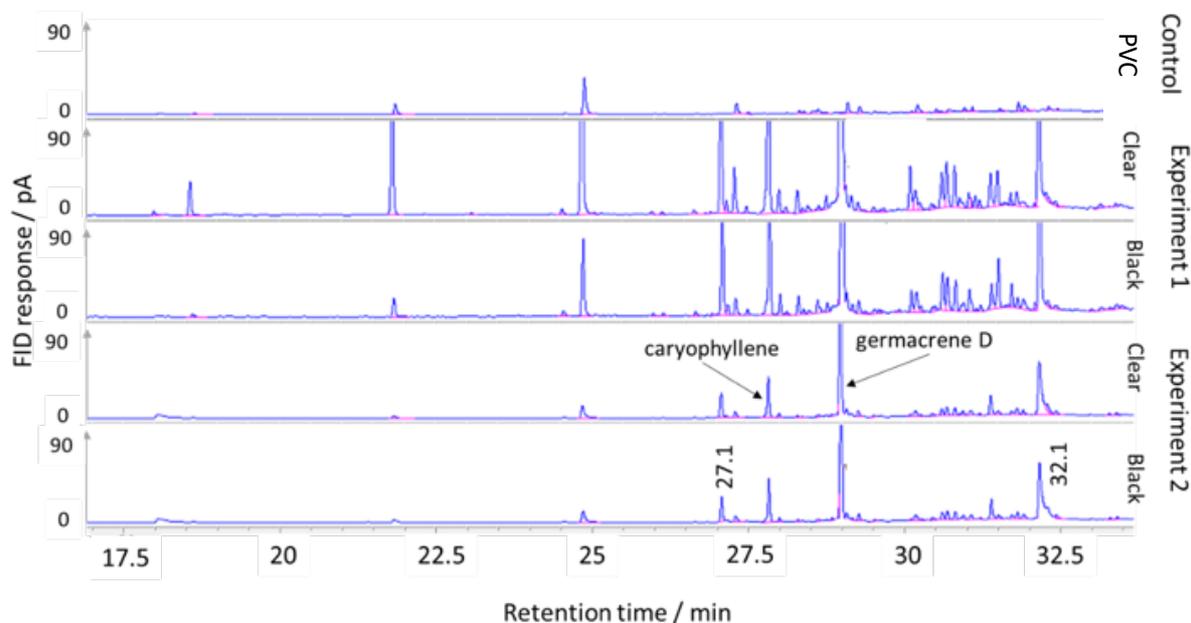


Figure 6.8: Comparison of gas chromatograms of volatile samples after air entrainment over 1-week across all conditions. Potential degradants appear similarly in both clear and black PVC samples.

Atmosphere exposed samples

Gas chromatograms of extracts from the sponges of the un-entrained sachets (Figure 6.4) appear quite different between samples of experiment 1 and experiment 2 (Figure 6.9). As with the volatile samples, they appear similar over the time points within each experiment, with no clear difference in performance between the PVC materials. Notable peaks for experiment 1 include 11.5, 11.6, 12.2, 13.5, 30.1, 30.9 and 31.2 min and for experiment 2 they are 28.5, 30.0, 30.1, 30.6, 31.9, 33.4 and 35.0 min, with the 33.4 min peak of greatest peak area. Pentadecane (indicated) was initially included as an internal standard for experiment 1 but omitted from experiment 2 to allow NMR characterisation of the sponge extracts to conclude the experiment, however compound identification was hindered by the mixture of compounds present.

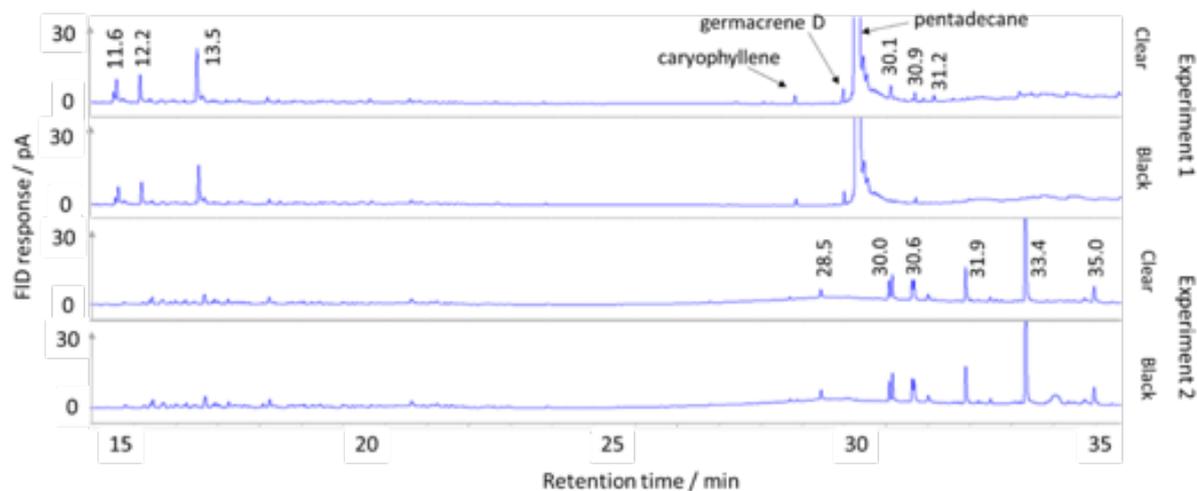


Figure 6.9: Comparison of gas chromatograms of sponge extract samples of both clear and black PVC following 3 weeks exposure to the greenhouse environment showing the contrast in degradants between experiments 1 and 2. No germacrene D (29.0 min) remains after 3 weeks in experiment 2. Pentadecane was initially used as an internal standard in experiment 1.

Both experiments showed that the materials perform very similarly, and the black PVC does not appear to prevent degradation more than the clear. Environmental conditions play a much more significant role for stability with differences in the PVC material negligible in comparison. The difference in release rate and stability profiles between experiments 1 and 2 largely vary with the time of year, particularly with atmospheric temperature differences: using data provided from the greenhouse monitoring, during experiment 1 the highest average day time temperature was 26.8 °C, peaking at 35.5 °C, and for experiment 2 the highest average day time temperature was 29.5 °C, peaking at 35.5 °C. These analyses highlight the difficulty in achieving reproducible results outside the lab, as well as for controlling effects in the field with many more variables to consider. In future experiments of this type, the Porapak filters could be covered to prevent any *in situ* degradation when the compounds have permeated the PVC and are adhered to the filter. Additionally, a separate investigation may be necessary to find a way of cleaning or preheating the PVC to avoid the release of the volatile compounds from the material during further experiments.

Germacrene D, with its cyclodecadiene ring structure and endocyclic double bonds, is relatively flexible which enables conformational changes, ring closures and rearrangements through acid catalysed, photochemical and thermal reactions.¹²⁷ Although most degradants were unidentified apart from the tentative identification of β -bourbonene, thermal and irradiation rearrangement products have been reported and reviewed in the literature.^{127,211,212}

β -Bourbonene, accessible through a [2+2] photochemical cycloaddition initiated by absorption of UV light, has been observed as a major product, with lower amounts of β -copaene, β -ylangene and isogermacrene D also detected, whilst increased reaction time led to the formation of secondary products. Thermally promoted reactions however lead to the formation of a different range of products. Under higher temperatures, germacrene D has been shown to rearrange into β -copaene, β -ylangene, ϵ -muurolene and ϵ -amorphene as major products.^{127,211,212} A summary of these compounds is shown in Figure 6.10.

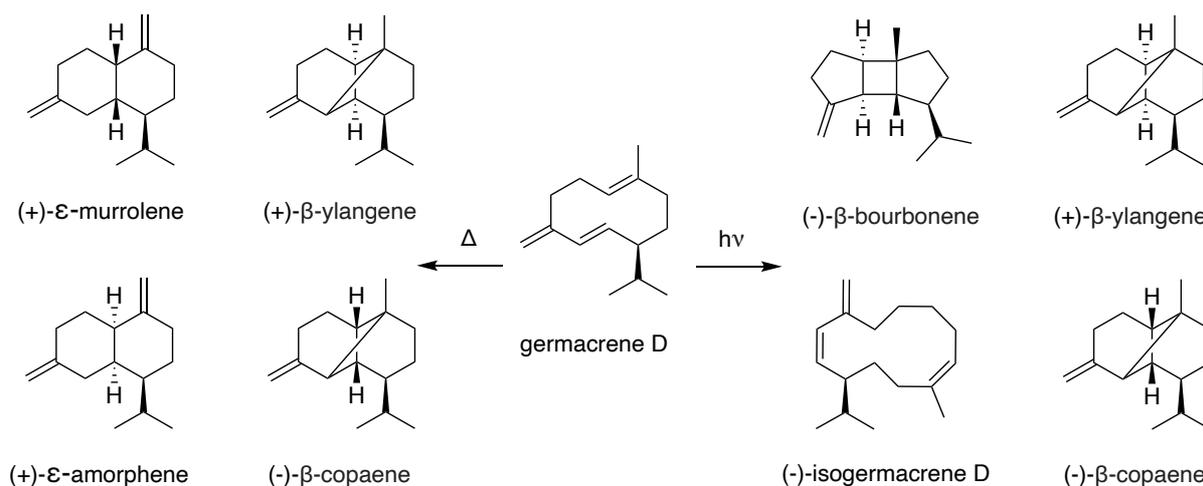


Figure 6.10: Structures of known thermal and irradiation products of germacrene D.

Although not experimentally comparable, in this study germacrene D was exposed to warm and light conditions over several weeks, so occurrence of rearrangement and secondary products is to be expected. Any hydrocarbons present have the potential further to react with atmospheric gases (e.g., NO_x , O_3 , SO_2) and undergo radical rearrangements, expanding the range of possible degradants even more widely. Furthermore, the presence of the caryophyllene impurity which exhibits similar chemistry to germacrene D would likely also lead to many degradants, further convoluting these results.

Although many of the peaks are of a comparable size to that of germacrene D, further investigation would be required to determine if the components were biologically active. Olfactometry could be used to assess potential behavioural effects using the samples collected, including the PVC controls and degraded germacrene D. However, in the field a diverse array of plant volatiles will naturally be present, therefore whether the degradants detected in this study would affect repellent behaviour may be inconsequential.

6.5 Summary

Insects such as aphids are significant agricultural pests, responsible for direct damage and virus transmission to crops resulting in large economic losses globally. Naturally occurring semiochemicals which mediate insect behaviour have potential for improving the sustainability of traditional insecticides which are associated increased pest resistance and possibly environmental damage.

This work investigated different properties of several known repellent semiochemicals using behavioural bioassays with representative pest aphid species and further analytical techniques. Industrially manufactured aphid alarm pheromone (*E*)- β -farnesene was characterised and compared to an in-house supply using behavioural bioassays with *S. avenae* and *M. persicae*. Both olfactometry and the alarm pheromone bioassays showed the external supply was effective at producing the anticipated repellent response for both pest species as well as at different stages of the life cycle of *M. persicae*. As there was no discernible difference between the in-house (*E*)- β -farnesene and the industrial supply, this would suggest the manufacturing process and subsequent storage has no effect on activity of the repellent and would therefore support its commercial development for agricultural scale aphid control.

Whitefly repellents 7-*epi*-zingiberene and curcumene were also assessed for aphid behavioural effects using four-arm olfactometry. Interestingly, 7-*epi*-zingiberene and the (*RS*)-curcumene mixture were shown to be repellent to both aphid species (although the latter not significantly for *M. persicae*). These findings provide a basis for further investigation of these chemicals, at the exploratory stage rather than the commercial stage as with (*E*)- β -farnesene, which may include electroantennography measurements. These finding demonstrate the potential of 7-*epi*-zingiberene beyond the use of sustainable whitefly control, for which there is currently great interest.

In these preliminary studies of germacrene D, clear and black PVC was investigated to determine release rates and potential degradation products when exposed to light and air. Germacrene D was shown to degrade into many minor components but this, as well as the release rates, depended considerably on the environmental conditions. Additionally, the PVC material used for the study was found to be responsible for a range of volatile compounds (likely derived from its manufacturing process) so may require treatment before use. The variability in these results highlight the difficulties involved in translating semiochemical technologies to the field.

6.6 Supplementary

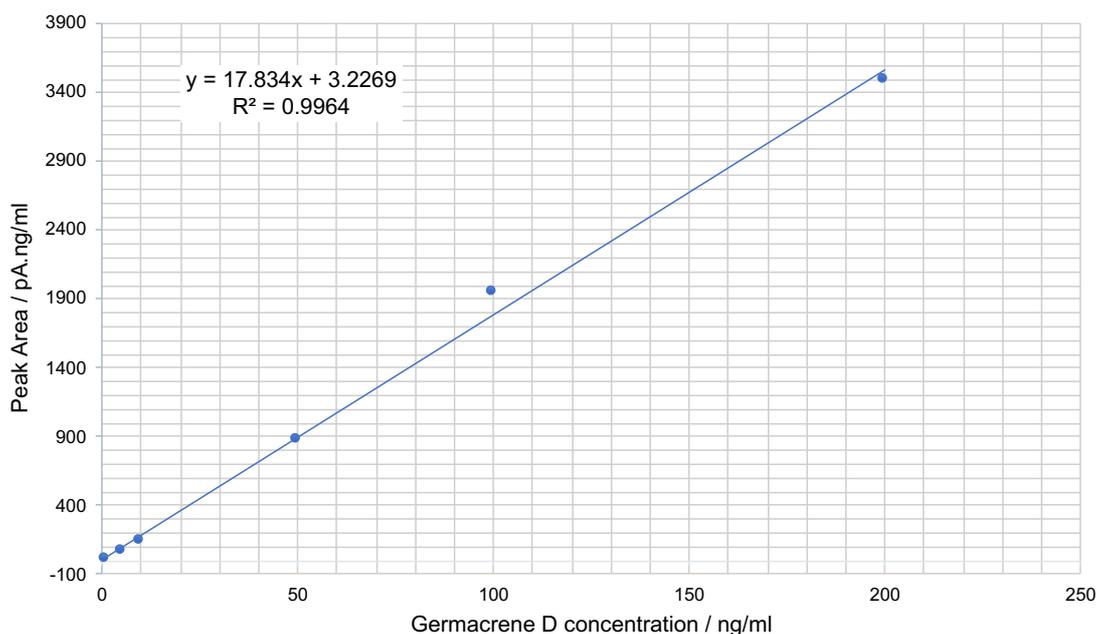


Figure 6.11: GC calibration curve generated from known concentrations of germacrene D used to calculate the amount of germacrene D in entrainment and stability study samples.

Table 6.4: Total germacrene D collected by air entrainment of the clear and black PVC sachets recorded over two experiments. Results expressed accumulatively with averages and standard error calculated from the three repeats taken of each material at each time point.

Exp.	PVC	Total Released / μg			Average Released / μg ($\pm\text{SE}$)		
		7 days	14 days	21 days	7 days	14 days	21 days
1	Clear	147.9	149.1	149.1	110.8 (± 44)	111.7 (± 0.3)	128.8 (± 0.0)
		5.5	5.7	5.7			
		179.0	180.3	180.3			
	Black	351.9	366.9	372.2	199.8 (± 63)	211.8 (± 1.2)	216.4 (± 0.4)
		146.7	156.7	161.6			
		100.7	111.6	115.4			
2	Clear	2.1	0.1	0.0	10.0 (± 5.6)	10.1 (± 0.0)	10.1 (± 0.0)
		23.8	0.0	0.0			
		4.2	0.1	0.0			
	Black	11.7	4.6	0.1	13.5 (± 4.1)	16.3 (± 0.8)	16.4 (± 0.0)
		22.9	2.7	0.0			
		5.9	0.9	0.1			

Conclusions and Future Work

The work described in this thesis focuses on exploration of biosynthetic methods for producing terpenes from isoprenoid diphosphate substrates, with particular focus on the useful repellent semiochemical, 7-*epi*-zingiberene, made by the sesquiterpene synthase, EZS. Efficient chemoenzymatic syntheses of both natural and unnatural substrates were developed to replace difficult and low-yielding existing routes. These were used in combination with EZS and other terpene synthases to produce 7-*epi*-zingiberene and novel semiochemical analogues with applications in crop protection. The bioactivities and physical properties of sesquiterpene semiochemicals were assessed using chemical ecology techniques which established that 7-*epi*-zingiberene produces a repellent response in pest aphid species, in addition to whitefly. Aspects of the structure-function relationship of EZS were also investigated, including determination of the mechanistic hydride shift using deuterium labelling, as well as mutagenesis to investigate structural determinants of product fidelity and substrate selectivity. These findings can be used to support the development of chemoenzymatic approaches to useful and high-value terpenes.

7.1 Chemoenzymatic synthesis of (Z,Z)-FDP and 7-*epi*-zingiberene

Chemical diphosphorylation methods of producing isoprenoid diphosphates from their respective alcohols are widely regarded as difficult and inefficient, generally hampering utilisation of chemoenzymatic methods for terpene production. Synthesis of the atypical (Z,Z)-FDP isomer, and 7-*epi*-zingiberene precursor, is additionally complicated by requiring formation of the less favourable Z double bonds in chain extension reactions to generate the C₁₅ substrate over 11 synthetic steps. An *in vitro* quantitative synthesis of fundamental C₅ terpene precursors, IDP and DMADP, was recently developed in the Allemann group which overcomes the difficult diphosphorylation by using two promiscuous kinases. This highly efficient enzymatic pathway was combined with the (Z,Z)-FDP producing *cis*-prenyltransferase, zFPS, to make the substrate with complete selectivity in yields of 92%, representing the first feasible route to its synthesis. The kinases were reused until loss of activity was observed to increase the amount of compound generated from a single batch of protein, enabling an abundance of stocks to be easily accumulated. Another key advantage of this new route is its modular nature: interchanging alcohol precursors and prenyltransferases enables generation of diphosphates of different chain lengths and functionalities for synthesis of a range of useful natural and unnatural terpenes.

This chemoenzymatic route was further adapted for 7-*epi*-zingiberene production. It was first necessary to optimise EZS expression which has historically been very poor. Initially, a fusion tag approach was tried whereby EZS was N-terminally cloned with various well-expressing

and highly soluble proteins. Western blot analysis of expression samples from a variety of conditions showed that EZS expression levels remained very low, so this approach was not pursued. *E. coli* cell lines were then investigated and a switch from the C41(DE3) pLysS variety to C41(DE3) lead to remarkable overexpression of EZS to practicably useful levels. With a new expression protocol established, EZS was produced at scale and combined with zFPS and kinase produced IDP and DMADP to make 7-*epi*-zingiberene, which was subsequently isolated and structurally characterised.

This initial attempt at chemoenzymatic sesquiterpene synthesis directly from the C₅ diphosphate precursors proved to be considerably more efficient than traditional methods, enabling analytical quantities to be readily produced. It could however be further adapted for larger scale sesquiterpene production in a variety of ways. Of the two kinases, THIM is considerably slower so one approach could involve protein engineering for improved efficiency or sourcing a different enzyme to shorten the time taken for completion of the diphosphorylation. Furthermore, both kinases, the relevant prenyltransferase and a terpene synthase could be combined in a one-pot reaction to produce a sesquiterpene of interest directly from simple C₅ alcohol precursors. A major objective involving biocatalytic application of terpene synthases in this way would involve finding ways to reuse them. Different hydrocarbon extraction systems or enzyme immobilisation methods could be explored and combined with this system, which could prove to be transformative in the field of commercial terpene production.

7.2 Chemoenzymatic generation of novel semiochemical analogues

Unnatural terpenes can be accessed by feeding substrate analogues with modified functionalities to promiscuous synthases. Although this approach of exploring otherwise inaccessible areas of chemical space has yielded novel compounds with useful and commercially relevant properties, such as aphid attractant 14,15-dimethylgermacrene, synthesis of the substrate analogues can be lengthy and cumbersome. This work sought to adapt the new enzymatic diphosphorylation pathway by producing analogues of IDP and DMADP which could rapidly be screened with semiochemical-producing terpene synthases for the discovery of novel agrochemicals *in vitro*. The promiscuity of the two kinases was exploited, and a library of small diphosphate precursors of different chain lengths and functionalities was generated. These substrates were systematically screened with *cis*- and *trans*-prenyltransferases to generate novel FDP analogues, which were subsequently incubated with EZS and (*E*)- β -farnesene and germacrene D synthases. It was hypothesised that unnatural substrates that were successfully turned over to novel product would share

enough similarity to the parent semiochemical to elicit bioactivity in the recipient organism. The modular screening approach successfully generated several novel semiochemicals, identified by GC-MS, which could be used directly for insect testing in future work. Attempts were made to scale the chemoenzymatic syntheses of the novel 7-*epi*-zingberene analogues, however the unnatural substrates were not efficiently turned over. This new method is therefore recommended for an initial screening to identify enzymatic substrate preferences and generate novel compounds which can then undergo further investigation if useful bioactivities are shown. It has provided fast access to a variety of unnatural and high value substrate analogues simultaneously and avoided otherwise lengthy total syntheses, where typically it is not known whether a diphosphate may even be accepted by the intended synthase. This methodology represents a starting platform which could be expanded firstly by increasing the library of simple diphosphate precursors, and secondly by substituting the terpene synthase for others of commercial interest.

7.3 Investigation of the hydride shifts of the catalytic mechanism of EZS

The catalytic mechanism of EZS was previously shown to proceed via the stepwise ionisation of the diphosphate group, followed by cyclisation to the bisaboyl carbocation intermediate. To determine from precisely which carbon the subsequent hydride shifts take place, (*Z,Z*)-FDP was labelled with deuterium at C5 and C6. Due to the challenging syntheses of labelled C₁₅ substrates, a chemoenzymatic approach was adopted utilising zFPS. C1 and C2 deuterium labelled C₁₀ substrates were produced chemically and combined with IDP and zFPS so the chain extension reaction could be carried out enzymatically. The full-length labelled substrates generated *in situ* were subsequently turned over by EZS to produce labelled compound. GC-MS analysis was sufficient to establish the precise location of the deuteriums in diagnostic fragments of the parent ion, leading to the conclusion that both the C5 and C6 deuteriums do not shift during the catalytic mechanism but remain in their original positions. This therefore implies that the catalytic mechanism goes via a 1,3-hydride shift from the C1.

A chemoenzymatic approach to this work avoided capricious, low yielding chemistry and enabled much quicker access to the required substrates. To complete the mechanism, *R* and *S* deuterium labelled (*Z,Z*)-FDP isotopologues could be used to reveal exactly which prochiral hydride of C1 shifts, and similarly, which proton of the C4 is lost in the final mechanistic step. Whilst chemoenzymatic routes to these substrates were looked at, isotopically labelling low molecular weight compounds (e.g., C₄ and C₅) were prohibitively challenging. This type of chemistry could therefore be expanded upon, or total synthetic routes to the final desired substrates could be pursued.

7.4 Structural determinants of EZS substrate selectivity and product specificity

EZS is a member of a small and unique family of terpene synthases which have evolved exclusively in tomato to produce terpenes derived from all *cisoid* precursors, arising atypically from a plastidial biosynthetic pathway directly associated with formation of defensive phytochemicals. EZS is also an example of a high-fidelity synthase, able to navigate its complex reaction mechanism with reactive carbocation intermediates to produce only 7-*epi*-zingiberene from the unusual (Z,Z)-FDP. Point mutations of amino acids in the G-helix of EZS proved this area is critically important in controlling product fidelity and substrate selectivity. Amino acids with side chains pointing into the active site were targeted, hypothesised to play a role in creating the active site contours which are important for folding the substrate into its reactive conformation. GC-MS analysis of incubations arising from (Z,Z)-FDP showed that perturbing the bulk of position F598 in particular lead to the formation of multiple different sesquiterpenes. Sesquiterpene alcohols were also detected, implying this region also functions to protect intermediates from bulk water. Kinetic analyses supported the initial hypothesis by showing that relative changes to K_m and k_{cat} values were less significant than changes to the product profile, therefore the residues at points of mutation provide a passive rather than direct catalytic role.

Mutants generated in the study were also incubated with (E,E)-FDP to understand whether any of changes to the active site shape would impact the substrate selectivity. GC-MS analysis revealed that the double mutant F598I-I603A was able to accept (E,E)-FDP and produce a mixture of amorphadiene, germacradien-4-ol, *epi*-cubenol and (E,E)-farnesol, whilst the product of (Z,Z)-FDP was mostly early exit product, (Z,Z)-farnesol. This finding implies that these residue changes have acted like an isomeric switch and has significant implications: if (E,E)-FDP accepting synthases could be engineered to preferentially accept (Z,Z)-FDP through identification of key amino acid substitutions, orthogonal pathways could be developed for *in vivo* terpene production that decouple sesquiterpene overproduction from native metabolism, which requires (E,E)-FDP, and lead to enhanced yields. To pursue this research, further mutations could initially be made to EZS and (Z,E)- and (E,Z)-FDP isomers used as substrates to understand in more detail how this switch physically occurs, before application to other synthases of commercial interest.

7.5 Ecological applications of sesquiterpenes for aphid pest management

This project also involved assessment of different chemical ecological properties of several sesquiterpene semiochemicals for application in aphid pest management. Four-arm olfactometry bioassays were carried out and 7-*epi*-zingiberene was shown to significantly repel two different pest aphid species for the first time. This highlights the real potential of 7-*epi*-zingiberene as an agrochemical and provides further basis to support its development. The aphid alarm pheromone, (*E*)- β -farnesene, also has prospective application as an aphid repellent. An industrial synthetic source was compared to a natural source to assess the suitability of the former, as even trace differences in formulation can negatively impact performance. Analytical characterisation showed both supplies were essentially identical, and two types of aphid behavioural bioassay gave comparable repellent results, implying that the industrial synthesis is suitable for manufacture of commercial (*E*)- β -farnesene. Lastly, the stability and release rate of another aphid repellent, germacrene D, was assessed using permeable PVC sachets as slow-release formulation devices to inform future use in the field. The performances of black and clear plastic were compared under greenhouse conditions with the black hypothesised to absorb UV radiation and potentially prevent excessive degradation. GC analyses revealed no distinct difference between the two materials, with recommended release rates of 1 week proposed at the doses tested. An abundance of volatiles associated with the plastic itself were detected, meaning it may require pre-treatment before use. Environmental conditions were however significant in affecting compound stability as in warm conditions, germacrene D was shown to break down into many hydrocarbon products. Whether these would be problematic for eliciting a repellent response could be investigated using behavioural testing with the samples collected. These findings highlight the difficulties in developing sesquiterpenes for field applications and the necessity of having reliable release devices. The ability to produce terpene semiochemicals by exploiting biosynthetic routes, as focused upon in this thesis, will only serve to enable more research of this kind and to innovate solutions to these problems and enable future adoption as sustainable agrochemicals.

Materials and Methods

8.1 Biological methods

8.1.1 Materials

All chemicals and general consumables were purchased from *Sigma-Aldrich*, *Apollo Scientific UK*, *Melford Laboratories Ltd* or *Thermo Fisher Scientific Ltd* unless otherwise stated and were used without further purification. Oligonucleotide primers for site-directed mutagenesis were purchased from *Sigma-Aldrich*. All DNA constructs were confirmed by DNA sequence analysis using Eurofins Genomics service (*Eurofins Scientific*). Restriction enzymes were purchased from *New England Biolabs UK Ltd (NEB)*.

8.1.2 Strains and plasmids

Cell lines

A variety of cell lines were used throughout the project according to Table 8.1. The strains were sourced from the Allemann Group library except for the C41(DE3) cell line which was a gift from Dr. Petra Bleeker.¹³⁷

Table 8.1: Cell lines and their specific use during the project.

Cell line	Use
DH5 α	DNA amplification and cloning
XL1-Blue	DNA amplification and cloning
BL21(DE3) pLysS	Expression of zFPS
BL21(AI)	Expression of IPK, THIM, FPPS
C41(DE3) pLysS	Expression of pET21a-EZS historically in the group Screening of clones for EZS expression optimisation
C41(DE3)	Optimised expression of pET21a-EZS
BL21-CodonPlus(DE3)-RP	Expression of EBFS

E. coli expression plasmids

A summary of expression plasmids used throughout this work are given in Table 8.2 with relevant details if known. All plasmids and genes for EZS-fusions were available from the Allemann Group library except for pET32-zFPS,⁴² which was a gift from Prof. Andrew H. J. Wang (Academia Sinica, Taiwan).

Table 8.2: Details of plasmids used in this work.

Plasmid name	Organism of Gene Source	Vector	Protein molecular weight / kDa	Comments	Antibiotic resistance
pET28-THIM	<i>Escherichia coli</i>	pET28a	27.4	N-terminal His-tag with thrombin cleavage site	Kan
pET28-IPK	<i>Methanocaldococcus jannashii</i>	pET28a	32.6	N-terminal His-tag with thrombin cleavage site	Kan
pET21-EZS	<i>Solanum habrochaites</i>	pET21a(+)	88.5	Codon optimised sequence with C-terminal His ₆ tag 45 N-terminal amino acid truncation (chloroplast targeting sequence)	Amp
pET32-zFPS	<i>Solanum habrochaites</i>	pET32-Xa/LIC	46.6	Codon optimised 45 N-terminal amino acid truncation Vector includes N-terminal TRX- and His-tag and Factor Xa cleavage site	Amp
DHFR-EZS	<i>Escherichia coli</i>	pET21a(+)	106.5	EZS with N-terminal dihydrofolate reductase ²¹¹ tag	Amp
GB1-EZS	<i>Streptococcus sp</i>	pET21a(+)	94.5	EZS with N-terminal guanidine binding protein ¹⁷⁰ tag	Amp
GST-EZS	<i>Schistosoma japonicum</i>	pET21a(+)	114.0	EZS with N-terminal glutathione-S-transferase ¹⁶⁸ tag	Amp
MBP-EZS	<i>Escherichia coli</i>	pET21a(+)	131.0	EZS with N-terminal maltose binding protein ⁶ tag	Amp
TRX-EZS	<i>Escherichia coli</i>	pET21a(+)	100.3	EZS with N-terminal thioredoxin ¹⁶⁸ tag	Amp
SUMO-EZS	<i>Saccharomyces cerevisiae</i>	pET21a(+)	99.8	EZS with N-terminal small ubiquitin-like modifier ¹⁶⁷ tag	Amp
pET28-FPPS	<i>Geobacillus stearothermophilus</i>	pET28a	32.3	Thermostable protein variant ²¹²	Kan
pET21-GDS	<i>Solidago canadensis</i>	pET21d	65.0	C-terminal His-tag	Amp
pET32-EBFS	<i>Mentha x piperita</i>	pET32b(+)	63.8	No His-tag and insoluble expression requiring basic extraction ²¹³	Amp

Antibiotics: amp = ampicillin, kan = kanamycin.

8.1.3 Growth media, buffers and antibiotic stock solutions

Luria Bertani (LB) media

Sodium chloride (10 g/L), yeast extract (5 g/L) and tryptone (10 g/L) were dissolved in deionised water and the solution sterilised in an autoclave. The required antibiotics were added before use under sterile conditions.

LB agar

Sodium chloride (10 g/L), yeast extract (5 g/L), tryptone (10 g/L) and agar (20 g/L) were dissolved in deionised water and the solution sterilised in an autoclave. Under sterile conditions, whilst the mixture was still liquid (but not hot, ~40 °C), the appropriate antibiotics were added and the agar set in petri dishes.

Phosphate buffer pH 7.5 (1 M)

Monobasic potassium phosphate (0.17 M, 23.14 g/L) and dibasic potassium phosphate (0.72 M, 125.4 g/L) were dissolved in deionised water (1 L). The pH of the resulting 1 M phosphate buffer solution was ensured to be pH 7.5 before sterilisation in an autoclave.

Terrific broth (TB) media

Tryptone (12 g/L), yeast extract (24 g/L) and glycerol (4 mL/L) were dissolved in deionised water (450 mL) and sterilised in an autoclave. Before use, phosphate buffer (50 mL) was added, alongside the required antibiotics under sterile conditions.

Antibiotics

Stock solutions of ampicillin and kanamycin were prepared at a concentration of 50 mM and used at a 50 µM concentration for bacterial growth. Stocks were made by dissolving the respective compound in deionised water and filter sterilised. Under sterile conditions the stocks were dispensed into 1 mL aliquots and stored at -20 °C.

Medias for screening EZS-fusion clones

Medias of varying ratios of nutrients and buffers used in the optimisation of EZS expression were produced according to the following list. The required antibiotics were added before use under sterile conditions.

Metals mix: a 1000 x solution (100 mL) was prepared by dissolving FeCl₃·6H₂O (0.1 M) in 0.1 M HCl (50 mL). A separate solution of MnCl₂ (0.1 M), ZnSO₄ (0.1 M), CoCl₂ (0.02 M) and NiCl₂ (0.02 M) was prepared in deionised water (50 mL) and filter sterilised.

The two solutions were then combined to give the metals mix solution which was stored at room temperature.

LBE (enhanced LB): yeast extract (5 g/L), tryptone (10 g/L), glucose (5 g/L) and NaCl (5 g/L) were added together and made to volume (allowing for later additions) with deionised water then sterilised by autoclave. Phosphate buffer (100 mM), MgSO₄ (2 mM) and metals mix (1 mL/L) were then added under sterile conditions.

LBE-LP (LBE low phosphate): made according to LBE but with different concentrations of glucose (1 g/L) and phosphate buffer (25 mM).

Autoinduction: yeast extract (5 g/L), tryptone (10 g/L), glycerol (5 g/L), glucose (0.5 g/L), lactose (2 g/L), Na₂SO₄ (0.7 g/L) and NH₄Cl (2.5 g/L) were added together and made to volume (allowing for later additions) with deionised water then sterilised by autoclave. Phosphate buffer (50 mM), MgSO₄ (2 mM) and metals mix (1 mL/L) were then added under sterile conditions.

TBE (enhanced TB): yeast extract (24 g/L), tryptone (12 g/L) and glycerol (4 mL/L) were added together and made to volume (allowing for later additions) with deionised water then sterilised by autoclave. Phosphate buffer (100 mM), MgSO₄ (2 mM) and metals mix (1 mL/L) were then added under sterile conditions.

2 x TYE: made according to LBE but with different concentrations of yeast extract (10 g/L) and tryptone (16 g/L).

8.1.4 *E. coli* competent cell preparation and transformation

Competent cells for general purposes

From stocks, the desired cell line was cultured overnight on LB agar (with antibiotic if appropriate) at 37 °C. A single colony was selected for an overnight starter culture of which 1 mL was used to inoculate 100 mL LB media. The cells were incubated at 37 °C with shaking until optical cell density (OD₆₀₀) had reached precisely 0.4 then incubated on ice for 30 min. After centrifugation (4 °C, 4 krpm, 10 min) cells were re-suspended with sterile ice cold 100 mM CaCl₂ (20 mL, buffer I) and incubated for 20 min. Following centrifugation as before, the cells were resuspended in sterile ice cold 100 mM CaCl₂, 15% v/v glycerol (5 mL, buffer II) and incubated for a final 30 min. The competent cells were flash frozen in 50 µL aliquots in sterile microcentrifuge tubes and stored at -80 °C. With each preparation, cells were plated on agar containing commonly used antibiotics to screen for resistance.

Supercompetent cells for cloning use

Supercompetent cells were prepared as described for competent cells above but using rubidium chloride buffers I and II (prepared as detailed below) in place of buffers I and II respectively.

Rubidium chloride buffer I: potassium acetate (30 mM), rubidium chloride (100 mM) and calcium chloride (10 mM) were dissolved in deionised water, glycerol (15% v/v) was then added and the pH of the solution adjusted to 5.8 before addition of manganese chloride (50 mM). The solution was made to volume using deionised water, filter sterilised and stored at 4 °C

Rubidium chloride buffer II: 3-(N-morpholino)propanesulfonic acid (MOPS, 10 mM), calcium chloride (75 mM) and rubidium chloride (10 mM) were dissolved in deionised water, glycerol (15% v/v) was added and the pH adjusted to 6.5. The solution was made to volume using deionised water, filter sterilised and stored at 4 °C

***E. coli* transformation**

An aliquot of stored competent cells was defrosted on ice for 5 min before adding the required plasmid DNA solution (1 µL). Following incubation on ice for 30 min, the DNA-cell solution was heat shocked in a water bath for 45 s at 37 °C then returned to ice for 5 min. LB media (200 µL) was added to the cells and the tube placed in an incubator with shaking at 37 °C for 1 h. The transformed cells were plated on LB agar containing the appropriate antibiotic and incubated at 37 °C overnight.

8.1.5 Plasmid preparation

Competent cells used for DNA amplification were transformed with the desired plasmid and grown on agar plates with the appropriate antibiotic. From a single colony, an overnight starter culture (4 mL) was prepared. DNA was extracted from the cell pellet using a ZymoPURE Plasmid Miniprep Kit (*Cambridge Bioscience*) following the associated protocol. The concentration of the plasmid solution was determined using a Nanodrop1000 UV Visible spectrophotometer (Nanodrop, *ThermoFisher Scientific*) by measuring 1 µL of sample against a calibrated sample of background solution. Plasmid size was verified using agarose gel electrophoresis and digested by specific restriction enzymes as required.

8.1.6 Polymerase chain reaction (PCR)

Reactions were performed in PCR tubes using the PrimeSTAR® HS DNA Polymerase premix (PrimeSTAR, *TakaraBio*) containing reaction buffer, polymerase and deoxyriboenucleotide triphosphates (dNTPs), and assembled according to Table 8.3. DMSO was added if the primers were anticipated to form undesirable secondary structures. The reactions were incubated in a thermocycler (Biometra Thermocycler T-Gradient Thermoblock, *Biometra*) programmed to run according to

Table 8.4. The annealing temperature typically used was 55 °C but this was changed (50 - 70 °C) if optimisation was required. Other variables adjusted for further PCR optimisation included primer concentration and amount of template DNA (1 - 50 ng).

Table 8.3: Components of PCR reaction mixtures.

Component	Volume or final concentration
PrimeStar	12.5 µL
Primers	0.1 - 0.5 µM
Template DNA	0.5 - 50 ng
DMSO	0 - 5% v/v
Deionised water	Up to 25 µL total volume

Table 8.4: Standard PCR protocol.

Step	Temperature / °C	Time	Number of Cycles
Initial denaturation	95	5 min	1
Denaturation	95	1 min	30
Annealing	55	30 s	
Extension	72	1 min / kb target	
Final extension	72	2 min / kb target	1

8.1.7 Golden Gate cloning

Amplification of vector and inserts

EZS fusion proteins were created as part of work to improve EZS expression using the Golden Gate cloning technique.²¹⁴ The backbone vector for the constructs and the genes for the inserts were amplified using PCR from existing plasmids detailed in Table 8.2. The primers used for the amplification were designed using Benchling software¹⁶³ according to rules relevant to Golden Gate requirements and are listed in Table 8.5. Primers were analysed to check melting temperature (T_m), overall length, mispriming and primer secondary structures. PCR products were analysed by agarose gel electrophoresis and bands relating to the expected DNA fragments were extracted from the gel and their concentrations measured by Nanodrop. The DNA fragments were diluted to final concentrations of 75 ng/ μ L for the assembly reactions.

Table 8.5: Golden Gate primers used in the construction of EZS fusion proteins

Primer	5' to 3' Sequence
EZS FWD	tccaacgggtctcctctgttgaaaacctgtattttcagtcc
EZS REV	tccaacgggtctccgaataaccagagctgctgcccattggtatc
SUMO FWD	tccaacgggtctcgattctgttatgtcggactcagaagtc
SUMO REV	tccaacgggtctcgaagaataaccagaaccaccaatctgttctctg
TRX FWD	gatcgagggtctcgattctgttatgagcgataaaattttcacctgactg
TRX REV	gatcgagggtctcgaagaataaccagaggccaggtagcgtcgaggaac
EcDHFR FWD	gatcaggggtctcgattctgttatgatcagctctgattgcggc
EcDHFR REV	gatcaggggtctcgaagaataaccagaccgctccagaatctcaaag
GST FWD	accgtaggtctcgattctgttatgtcccctatactagggtattg
GST REV	accgtaggtctcgaagaataaccagattttggaggatgggtcgcc

Golden Gate Assembly

Golden Gate assembly reactions were prepared in PCR tubes on ice using the purified DNA and Golden Gate buffer and enzyme mix (*NEB*) according to Table 8.6. Reactions were incubated in a thermocycler for 1 h at 37 °C after which 5 μ L was used to transform supercompetent cells.

Table 8.6: Components of Golden Gate assembly reactions

Component	Volume / μL
Vector (75 ng/ μL)	1
Insert (75 ng/ μL)	1
Golden Gate buffer (10 x)	2
Golden Gate enzyme mix	1
Deionised water	15

8.1.8 Site-directed mutagenesis (SDM)

Amplification of vector

Two cloning techniques were used to generate single point mutations in the protein sequence of EZS for structure-function investigations: a one-step protocol based on overlapping complimentary ends and a kinase-ligase-DpnI (KLD) protocol²¹⁵ based on creating blunt end fragments. Primers were designed using Benchling software¹⁶³ according to rules relevant to each approach and were analysed to check melting temperature (T_m), overall length, mispriming and primer secondary structures. The primer pairs were compared and those with the more optimal properties taken forward. Those used successfully are listed in Table 8.7.

Table 8.7: Primers used for EZS SDM cloning. The codon at the position of the mutation is capitalised.

Primer	5' to 3' Sequence	Cloning Technique
Y484W FWD	ctggtgTGGgcaagtacgttatgctgctgacat	One-step
Y484W REV	ttcgcCCAcaccaggcgtgctgctcagttc	
F598A FWD	attaccGCGggttcccgtctgattccactgactactcag	One-step
F598A REV	ggaaccCGCggaatggaagaaacatacaggtattcttcgatacg	
F598L FWD	CGCcacgataccagtggaag	KLD
F598L REV	gggaaccTAAggaatggaagaaac	
F598I, I603A FWD	cattaccATCggttcccgtctgGCGccactgac	One-step
F598I, I603A REV	gaaccGATggaatggaagaaacatacaggtattcttcg	
F598W FWD	gttcccgtctgattccactgac	KLD

F598W REV	cCCA _g gtaatggaagaaacatacagg	
G599C FWD	TGTtcccgtctgattccactg	KLD
G599C REV	gaaggtaatggaagaaacatacagg	
S600A FWD	GCCcgtctgattccactgactactc	KLD
S600A REV	accgaaggtaatggaagaaacatac	
I603A FWD	gtctgGCCcactgactactcagtacttcattggtatc	One-step
I603A REV	cagtggCGC _c agacgggaaccgaaggtaatgg	
L636A FWD	cgtctgctgaacgacctg	KLD
L636A REV	CGC _c acgataccagtgaag	
F705W FWD	gaaatcTGGtggcgcacctgtaagtgggctcacttc	One-step
F705W REV	cgcagctgtgcaagaaatcTGGtggcg	

SDM method

The pET21-EZS vector was amplified by PCR using primer pairs suitable for the one-step approach. Following the PCR, DpnI (1 μ L) and FastDigest 10 x buffer (2.5 μ L) were added to the reaction mix and incubated for at least 1 h at 37 °C to digest template DNA. An aliquot of reaction mix (10 μ L) was then directly transformed into supercompetent cells.

KLD method

The pET21-EZS vector was amplified by PCR using primer pairs suitable for the KLD approach. The KLD reaction was then assembled in a separate PCR tube using the direct PCR product and KLD Enzyme Mix (*NEB*) reagents according to Table 8.8. The reactions were incubated for at least 5 min at 25 °C in a thermocycler and then transformed (10 μ L) into supercompetent cells.

Table 8.8: Components of KLD assembly reactions

Component	Volume / μ L
PCR product	1
KLD Reaction Buffer (2 x)	5
KLD Enzyme Mix (10 x)	1
Deionised water	3

8.1.9 Agarose gel electrophoresis

Buffers and gel preparation

TRIS-acetate EDTA buffer (TAE, 50 x) was made by dissolving TRIS (242 g), glacial acetic acid (57.1 mL) and 0.5 M EDTA pH 8.0 solution (100 mL) in deionised water (600 mL) then diluted to volume (1 L). The solution was diluted to 1 x for further use.

Agarose gels were prepared by microwaving agarose (0.5 g) with TAE buffer (1 x, 50 mL) and mixed until dissolved. The solution was set in the Sub Cell GT Minicast (*Bio-Rad*) with the addition of SYBR Safe loading dye (5 μ L).

Analysis

Samples containing DNA to be analysed were mixed with FastDigest Green buffer (10 x) before loading onto the gel. A sample of DNA Marker (High Range) 1 kb Ladder (*Melford*, 5 μ L) was loaded alongside to provide a molecular weight reference. TAE buffer was then applied over the gel to fill the tank (*Bio-Rad*), which was operated at 100 V for 60 min. Gels were visualised using ChemiDoc Gel Imaging System (*Bio-Rad*).

Gel extraction

Bands relating to DNA of interest were cut from the gel with a scalpel and transferred into a pre-weighed microcentrifuge tube. The extraction was then performed using a QIAquick Gel Extraction Kit (*Qiagen*) according to the associated instructions. Concentrations of purified DNA were determined by absorbance at 260 nm using the Nanodrop.

8.1.10 Protein expression and purification

Screening EZS clones for expression optimisation

From overnight starter cultures, 20 mL of each media (see section 0 for different medias used) was inoculated with each EZS-fusion clone and cultured at 37 °C until an OD₆₀₀ of 0.8 was reached. At this point, IPTG (0.1 mM) was added to induce expression. These conditions represent the recommended starting conditions in the C41 cell line manufacturers manual. Each culture was then split into 4 x 5 mL identical portions for growth at four different temperatures - 16, 20, 25 and 37 °C - with shaking at 225 rpm. Samples (1 mL) from each temperature were taken before induction and at 2 and 4 h and after incubation overnight following induction. To analyse the soluble and insoluble protein content of each sample, a freeze-thaw approach was taken in account for the small volume. This involved resuspending the cell pellet of each sample in buffer containing lysozyme (100 mM TRIS pH 8.0, 1 mg/mL

lysozyme, 1 mL), followed by incubation at 37 °C for 1 h. The samples were frozen on dry ice (5 min) and then thawed (25 min, 37 °C) for a total of five times, then clarified by centrifugation to obtain a soluble fraction and insoluble pellet. Samples were analysed by SDS-PAGE and western blot using a His-probe to detect the presence of His-tagged EZS-fusion protein.

Expression and purification of THIM, IPK and FPPS

Single colonies of BL21-AI cells transformed with the respective expression plasmid were used to inoculate 10 mL of LB media containing kanamycin (50 µM) and were incubated overnight at 37 °C with shaking (200 rpm). The following day, 5 mL of overnight culture was added to 500 mL of LB media containing kanamycin (50 µM) for each expression. The cells were grown to an OD₆₀₀ between 1.0 - 1.2 at 37 °C with shaking (200 rpm) then induced with L-arabinose (4 g/L) and grown for a further 4 h. The cells were then harvested by centrifugation (6 krpm, 10 min) and the pellets stored in extraction buffer (5 mL, 50 mM TRIS pH 8.0, 300 mM NaCl, 0.5 mM tris(2-carboxyethyl)phosphine (TCEP)) at -20 °C. For purification, the required pellets were thawed and lysed by sonication with the addition of phenylmethylsulfonyl fluoride (PMSF, 5 mg) and lysozyme (5 mg). The cell lysate was clarified by centrifugation (18 krpm, 40 min) and the supernatant incubated with a pre-equilibrated Ni²⁺-NTA resin in a free-flow drip column. The flow through and each fraction was collected as the column was washed with extraction buffer containing 10 mM and 40 mM imidazole (25 mL each). The bound protein was eluted in extraction buffer containing 500 mM imidazole (5 mL, passed twice through the column). The fraction containing protein was exchanged into assay buffer (100 mM TRIS pH 8.0) using a de-salting column (HiPrep 26/10, *GE Healthcare*) and the concentration measured by A₂₈₀ absorbance with extinction coefficients of 24135, 30370 and 23505 M⁻¹ cm⁻¹ for THIM, IPK and FPPS respectively, calculated using the Benchling web tool¹⁶³ under reducing conditions. Samples were taken at each stage of the purification and analysed by SDS-PAGE. If necessary, the proteins were concentrated by centrifugation (4 krpm) using a 15 kDa cut-off spin column. THIM and IPK were stored at -80 °C without glycerol and FPPS with 10% v/v glycerol.

Expression and purification of zFPS

A single colony of BL21-(DE3) cells transformed with pET32-zFPS was used to inoculate 10 mL of TB media containing ampicillin (50 µM) and incubated overnight at 37 °C with shaking (200 rpm). The following day, 5 mL of overnight culture was added to 500 mL of TB media containing ampicillin (50 µM) for each expression. The cells were grown with shaking (200 rpm) to an OD₆₀₀ of 1.2 - 1.4 then induced with IPTG (0.5 mM) and grown overnight at 16 °C. The cells were harvested by centrifugation (6 krpm, 10 min) and the pellets stored in

extraction buffer (5 mL, 25 mM TRIS pH 7.8, 500 mM NaCl, 10 mM imidazole) at -20 °C. For purification, the required pellets were thawed and lysed by sonication with the addition of PMSF (5 mg) and lysozyme (5 mg). The cell lysate was clarified by centrifugation (18 krpm, 40 min) and the supernatant incubated with a pre-equilibrated Ni²⁺-NTA resin in a free-flow drip column. The flow through and each fraction was collected as the column was washed with extraction buffer containing 25 and 40 mM imidazole (25 mL each). The bound protein was eluted in extraction buffer containing 500 mM imidazole (5 mL, passed twice through the column). The fraction containing protein was exchanged into assay buffer (100 mM TRIS pH 8.0, 10% v/v glycerol) using a de-salting column (HiPrep 26/10, *GE Healthcare*) and the concentration measured by Bradford assay.¹⁶⁴ Samples were taken at each stage of the purification and analysed by SDS-PAGE. If necessary, the protein was concentrated by centrifugation (4 krpm) using a 15 kDa cut-off spin column. The protein was stored at -80 °C.

Optimised expression and purification of EZS

A single colony of C41(DE3) cells transformed with pET21-EZS was used to inoculate 10 mL of TB media containing ampicillin (50 µM) and incubated overnight at 37 °C with shaking (200 rpm). The following day, 5 mL of overnight culture was added to 500 mL of TB media containing ampicillin (50 µM) for each expression. The cells were grown with shaking (200 rpm) to an OD₆₀₀ between 1.2 - 1.4 then submitted to a cold shock (30 min at 4 °C). During this time, the incubator was cooled to 18 °C. The cells were then induced with IPTG (0.1 mM) and grown overnight at 18 °C. The cells were harvested by centrifugation (6 krpm, 10 min) and if stored, the pellet resuspended in extraction buffer (5 mL, 100 mM TRIS pH 8.0, 300 mM NaCl, 0.5 mM TCEP) and kept at -20 °C. For purification, the required pellets were resuspended in extraction buffer and lysed by sonication with the addition of PMSF (5 mg) and lysozyme (5 mg). The cell lysate was clarified by centrifugation (18 krpm, 40 min) and the supernatant incubated with a pre-equilibrated Ni²⁺-NTA resin in a free-flow drip column. The flow through and each fraction was collected as the column was washed with extraction buffer containing 0 and 10 mM imidazole (25 mL each). The bound protein was eluted in extraction buffer containing 500 mM imidazole (5 mL, passed twice through the column). The fraction containing protein was exchanged into assay buffer (100 mM TRIS pH 8.0, 10% v/v glycerol) using a de-salting column (HiPrep 26/10, *GE Healthcare*). The protein was concentrated by spin ultrafiltration (4 krpm) using a 30 kDa membrane and the concentration measured by Bradford assay.¹⁶⁴ Samples were taken at each stage of the purification and analysed by SDS-PAGE. The protein was stored at -80 °C.

Expression and purification of GDS

A single colony of C41(DE3) cells transformed with pET21-GDS was used to inoculate 10 mL of TB media containing ampicillin (50 μ M) and incubated overnight at 37 °C with shaking (200 rpm). The following day, 5 mL of overnight culture was added to 500 mL of TB media containing ampicillin (50 μ M) for each expression. The cells were grown with shaking (200 rpm) to an OD₆₀₀ between 1.2 - 1.4 then submitted to a cold shock (30 min at 4 °C). During this time, the incubator was cooled to 18 °C. The cells were then induced with IPTG (0.1 mM) and grown overnight at 18 °C. The cells were harvested by centrifugation (6 krpm, 10 min) and if stored, the pellet resuspended in extraction buffer (5 mL, 50 mM TRIS pH 8.0, 50 mM NaCl, 5 mM β ME) and kept at -20 °C. For purification, the required pellets were resuspended in extraction buffer and lysed by sonication with the addition of PMSF (5 mg) and lysozyme (5 mg). The cell lysate was clarified by centrifugation (18 krpm, 40 min) and the supernatant incubated with a pre-equilibrated Ni²⁺-NTA resin in a free-flow drip column. The flow through and each fraction was collected as the column was washed with extraction buffer containing 0, 25 and 40 mM imidazole (25 mL each). The bound protein was eluted in extraction buffer containing 500 mM imidazole (5 mL, passed twice through the column). The fraction containing protein was exchanged into de-salt buffer (10 mM TRIS pH 7.5, 5 mM β ME, 10% v/v glycerol) using a de-salting column (HiPrep 26/10, *GE Healthcare*). The protein was concentrated by spin ultrafiltration (4 krpm) using a 30 kDa membrane and the concentration measured by Bradford assay.¹⁶⁴ Samples were taken at each stage of the purification and analysed by SDS-PAGE. The protein was stored at -80 °C.

Expression and purification of EBFS

A single colony of BL21-CodonPlus(DE3)-RP cells transformed with pET32-EBFS was used to inoculate 10 mL of LB media containing ampicillin (50 μ M) and incubated overnight at 37 °C with shaking (200 rpm). The following day, 5 mL of overnight culture was added to 500 mL of LB media containing ampicillin (50 μ M) for each expression. The cells were grown with shaking (200 rpm) to an OD₆₀₀ between 0.6 - 0.8 then induced with IPTG (0.1 mM) and grown overnight at 18 °C. The cells were harvested by centrifugation (6 krpm, 10 min) and if stored, the pellet resuspended in extraction buffer (5 mL, 50 mM TRIS pH 8.0, 5 mM β ME) and kept at -20 °C. For purification by basic extraction, the required pellets were resuspended in extraction buffer and lysed by sonication with the addition of PMSF (5 mg) and lysozyme (5 mg). The cell lysate was clarified by centrifugation (18 krpm, 40 min) and the insoluble pellet resuspended in extraction buffer (75 - 100 mL) and stirred slowly on ice until homogeneous (approx. 30 min). With stirring on ice throughout, the solution was titrated carefully to pH 12.0 using NaOH (5 M), incubated for 30 min, then carefully titrated back to pH 8.0 using HCl (1 M) and incubated for

a further 30 min. The solution was clarified by centrifugation (18 krpm, 40 min) and samples taken at each stage of the purification were analysed by SDS-PAGE. Following basic extraction and analysis, the soluble fraction containing EBFS was judged to be of sufficient purity to not require further purification, however if necessary, anion exchange or size-exclusion chromatography would have been used as the protein doesn't contain a His-tag. The protein was concentrated by spin ultrafiltration (4 krpm) using a 30 kDa membrane and the concentration measured by Bradford assay¹⁶⁴ before storage at -80 °C with the addition of glycerol (10% v/v).

8.1.11 SDS-PAGE

Protein samples were analysed by SDS-PAGE using the MiniPROTEAN Tetra Cell system (*Bio-Rad*). Gels were prepared, cast and samples analysed as described below.

Gel preparation

Gels were prepared by first mixing the preprepared components for the resolving gel together in the order listed in Table 8.9 and allowed to set in the cast and covered with isopropanol (0.5 mL). Once set, the isopropanol was removed and the stacking gel components mixed and cast on top of the resolving gel with the inclusion of a comb to create wells to load samples.

Table 8.9: Reagents used in the preparation of gels for SDS-PAGE analysis. Quantities given are for two gels.

Component	Resolving Gel (12%)	Stacking Gel (5%)
Deionised water	3.4 mL	2.85 mL
Resolving buffer (1.5 M TRIS pH 8.0)	2.5 mL	-
Stacking buffer (0.5 M TRIS pH 6.8)	-	1.25 mL
Acrylamide/bis-acrylamide 30% (8% w/v)	4 mL	0.85 mL
10% SDS (w/v)	100 µL	50 µL
10% APS ^a (w/v)	100 µL	50 µL
TEMED ^b	10 µL	5 µL

^aAPS = ammonium persulfate

^bTEMED = *N,N,N',N'*-tetramethylethylenediamine

Buffers and stain

Sample loading buffer (2 x, 10 mL) was prepared by mixing deionised water (4 mL), stacking buffer (1.25 mL), glycerol (1 mL), 10% SDS (0.5 mL), β ME (0.5 mL) and bromophenol blue (0.2 mL, 0.6% w/v solution). The solution was stored in the dark at 4 °C.

Running buffer (10 x, 1 L) was prepared by adding TRIS-base (30.3 g, 0.25 M), glycine (150.14 g, 2 M) and SDS (10 g, 10% w/v) together and slowly stirring until dissolved in deionised water, then making to the final volume. Running buffer was diluted to 1 x before further use.

Gel stain was prepared by dissolving Coomassie Brilliant Blue G-250 (20 mg) in ethanol (2 mL). HCl (2 mL, 11.2 M) was added and the solution made to 200 mL using deionised water and stored in the dark at room temperature.

Analysis

Protein samples (10 μ L) were mixed with sample loading buffer (10 μ L) and incubated at 90 °C for 5 min. After gel polymerisation, the gel was assembled in the tank, the comb removed and the tank filled with running buffer (1 x). Samples were loaded along with 5 μ L of Pierce™ Unstained Protein MW Marker molecular weight marker (14.4 - 116.0 kDa, *Thermo Scientific*). The tank was operated at 200 V for 50 min. Once run, the gels were removed, stained with gel stain and heated for 1 min using a microwave. The stain was incubated for at least 10 min (preferably overnight) before being destained by heating with water and washing repeatedly until protein bands could be seen. Gels were visualised using ChemiDoc Gel Imaging System (*Bio-Rad*).

8.1.12 Western blotting

Protein samples were analysed by western blot when the protein of interest could not be identified by SDS-PAGE analysis. The analysis was carried out using the Trans-Blot® Turbo Blotting System (*Bio-Rad*).

Reagents

Transfer buffer (1 x, 200 mL) was prepared fresh using Trans-Blot Transfer Buffer (4 x) by a 1:4 dilution in deionised water.

TRIS-buffered saline with Tween 20 (TBST, 1 L) buffer was made by dissolving TRIS (3.0 g, 25 mM) with NaCl (8.8 g, 150 mM) in deionised water (800 mL) and adjusting to pH 7.4 using hydrochloric acid (1 M). Tween 20 (0.1% v/v) was added and the solution made to volume with deionised water and stored at 4 °C.

Blocking buffer was prepared by dissolving bovine serum albumin (BSA, 1.5 g) in cold TBST (50 mL) immediately prior to use.

Analysis for His-tagged proteins

Protein samples for analysis by western blot were first resolved by SDS-PAGE with two identical gels run in duplicate. Positive control samples (a protein confirmed to contain a His-tag) and ColorPlus™ Prestained Protein Marker, Broad Range (7 - 175 kDa, *NEB*) were included in addition to aid analysis. Following the manufacturer's protocol, one of the two identical gels was transferred onto a Mini polyvinylidene difluoride (PVDF) membrane using the preprogrammed 'Mixed MW' transfer setting (1.3 A, 25 V, 7 min). The remaining un-transferred gel was visualised to verify success of the transfer. The membrane containing transferred protein was incubated in blocking buffer (20 mL) with gentle agitation at room temperature for 1 hour. The membrane was then washed in TBST buffer (3 x 15 mL, each with 5 min shaking) then incubated overnight with HisProbe™-HRP Conjugate (horseradish peroxidase, 0.8 mg/mL, *Thermo Scientific*) according to the manufacturer's instructions. The next morning, the membrane was washed with TBST buffer (3 x 15 mL, each with 5 min shaking) and the HRP-conjugated His-tagged proteins detected via enhanced chemiluminescence (ECL) using the Clarity Western ECL Substrate (*Bio-Rad*) according to the associated protocol. The membranes were visualised using automatic exposure settings of the ChemiDoc Gel Imaging System (*Bio-Rad*).

8.1.13 Protein concentration

Bradford assay

A modified form of the Bradford assay was used to determine protein concentration in solutions containing glycerol and/or UV active buffering components.¹⁶⁴

Bradford reagent was prepared by dissolving Coomassie Brilliant Blue G-250 (20 mg) in ethanol (2 mL). Phosphoric acid (20 mL, 80% wt.%) was then added and the solution made to 200 mL using deionised water and stored in the dark at 4 °C. The reagent was filtered before use using a syringe filter.

To generate a calibration curve, a stock solution of BSA (1 mg/mL) was used to prepare in triplicate a range of solutions varying between 2 - 100 µg/mL in a final volume of 200 µL in deionised water. Bradford reagent (800 µL) was added to each solution and its absorbance measured between 400 and 600 nm using a microplate reader spectrophotometer (FLUOstar Omega, *BMG Labtech*). The ratio of absorbances at 590 and 450 nm (590/450) were calculated, averaged between concentrations and values used to plot a standard curve.

The concentration of the protein of interest was measured by preparing samples diluted in deionised water (up to 200 μL) in triplicate. Bradford reagent (800 μL) was added and the colour of the resulting solution judged to ensure it was within the range of the samples of the concentration curve prepared previously. The absorbance was measured and normalised as described above and the protein concentration determined by comparison of the ratio of the unknown sample to the calibration curve, accounting for the volume used in the initial sample preparation.

UV-VIS

Absorbance at 280 nm (A_{280}) was used to determine the concentration of protein in solutions which did not contain glycerol and/or UV active buffering components.

A background measurement using the protein sample buffer was used to normalise the absorbance to zero before measuring the A_{280} of samples containing protein. Samples were diluted as appropriate to ensure absorbance values came within the correct sensitivity range (0 - 1 Au) and the absorbance measured using a spectrophotometer (*Shimadzu*). Protein concentrations were determined using Equation 8.1 where [E] is the enzyme concentration (M), A_{280} is the recorded absorbance value and ϵ is the extinction coefficient of 24,135, 30,370 and 23,505 $\text{M}^{-1} \text{cm}^{-1}$ for THIM, IPK and FPPS respectively (estimated using the Benchling web tool¹⁶³ under reducing conditions).

Equation 8.1

$$[\text{E}] = \frac{A_{280} \times \text{dilution factor}}{\epsilon}$$

8.1.14 FPLC

Fast protein liquid chromatography (FPLC) was used in the preparation (desalt) and purification (anion exchange) of proteins using ÄKTA™ (*GE Healthcare*) equipment. All buffers used for FPLC were filtered and degassed before use and the columns washed and stored in 20% ethanol (v/v) solution. UV absorbance at 280 nm was monitored throughout and peaks relating to protein were analysed by SDS-PAGE.

Desalt

Sample buffer exchange was carried out using a HiPrep™ 26/10 Desalting column (*Cytivia*) at a flow rate of 4 mL/min. The column was equilibrated with 100 mL of exchange/ low salt

buffer, the sample loaded onto the column and the flow through relating to the protein of interest manually collected over a further 100 mL elution of buffer.

Anion exchange

Anion exchange chromatography was performed using a Resource™ Q column (*Cytivia*) at a flow rate of 2 mL/min, first equilibrated in low salt buffer (100 mM TRIS pH 8.0, 2 column volumes (CV)). The sample was loaded and a washing step (1 CV) carried out with low salt buffer. A gradient of 0 - 50% high salt buffer (100 mM TRIS pH 8.0, 1 M NaCl) was then applied over 60 min followed by a final wash (1 CV) with high salt buffer. All purification fractions were collected from sample loading.

8.1.15 Steady-state kinetics

Steady state kinetic experiments were performed to characterise EZS and its mutants. Substrate radiolabelled with tritium ((2Z,6Z)-[1-³H]-farnesyl diphosphate, [1-³H]-(Z,Z)-FDP) was previously synthesised by Dr. Chris Jones and kindly made available for this work and conditions used in the subsequent reactions were based on previous work and methods.¹⁵³ The same pipettes were used between experiments to ensure reproducibility as the technique is extremely sensitive.

Stock solutions and buffers

From a master stock of [³H]-(Z,Z)-FDP, a series of 500 μM aliquots (250 μL) were prepared in deionised water and stored at -80 °C. One aliquot was used per experiment to ensure continuity between experiments and to avoid repeated freeze-thawing of the substrate.

The enzyme for the assay was purified by anion exchange to ensure accurate determination of concentration by Bradford assay.¹⁶⁴ The enzyme was then diluted to 500 nM for further use.

Kinetics buffer (15 mM MgCl₂, 25 mM 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES) pH 8.0) and concentrated kinetics buffer (30 mM MgCl₂, 50 mM HEPES pH 8.0) were prepared freshly between experiments from the same stock solutions of 0.5 M MgCl₂ and 1 M HEPES pH 8.0 (stored at 4 °C).

Experimental set-up

An aliquot of 500 μM [³H]-(Z,Z)-FDP was defrosted on ice and used to prepare stocks (1 mL each) of 50 μM and 75 μM substrate in deionised water. Each substrate concentration to be measured in the experiment was assembled according to Table 8.10 in triplicate in microcentrifuge tubes cooled on ice. Substrate background samples were also prepared for

each concentration and were assembled identically but with an extra 15 μL of kinetics buffer to account for the volume of enzyme that would not be added. Throughout the series of kinetics experiments care was taken to ensure experiments were assembled in a consistent order.

Table 8.10: Final substrate concentrations used in the steady state kinetics assays and for each experiment and their components.

[³ H]-(Z,Z)-FDP final conc. / μM	[³ H]-(Z,Z)-FDP stock vol. / μL		Vol. of kinetics buffer / μL	Vol. of concentrated kinetics buffer / μL
	50 μM	75 μM		
0.5	2.5	-	107.5	125
1	5	-	105	
2	10	-	100	
4	20	-	90	
5	25	-	85	
6	30	-	80	
8	40	-	70	
10	50	-	60	
15	-	50	60	
22.5	-	75	35	
30	-	100	10	

Experimental method

Whilst on ice, enzyme (15 μL , 500 nM, giving 30 nM final concentration) was added to each assay (except the blank) followed by hexane (0.9 mL). All assays were then transferred directly to a preheated ThermoMixer Comfort (*Eppendorf*) and incubated at 30 °C for 10 min with shaking (400 rpm). The reactions were immediately placed on ice and EDTA (200 μL , 0.1 M) added followed by vortex mixing (3 s) to stop any further reaction. The hexane overlay was carefully removed to ensure the exclusion of any of the aqueous portion and passed through individual silica columns (~500 mg silica in glass Pasteur pipette plugged with cotton wool) into vials containing the scintillation cocktail (15 mL, OPTI-FLUOR® O, *PerkinElmer*). Extraction of enzymatic product from the assays was repeated with two further portions of hexane followed by vortex mixing (3 s). Each silica column was washed with an additional portion of ethyl acetate in hexane (1 mL, 10% v/v). Following the extractions, each scintillation cocktail mixture was vortexed for a further 5 s to ensure homogeneity of the sample. The

radioactive mixture was analysed using a TRI-CARB 2900TR Liquid Scintillation Analyzer (*PerkinElmer*) in ^3H mode for 4 min per sample.

Data analysis to determine kinetic parameters

The substrate background samples for each concentration were subtracted from the experimental data expressed as counts per minute (cpm) for each corresponding concentration. The corrected cpm data was converted into rate or reaction velocity (V , $\mu\text{M}/\text{s}$) using Equation 8.2 where t (s) relates to the incubation time of enzyme and substrate (in this case 10 min; 600 s) and A (cpm/ μM) is the activity of the radioactive substrate stock determined from measuring the counts of a known 1 μM sample.

Equation 8.2

$$V = \frac{\text{cpm}}{t A}$$

The rate data was used to calculate k_{cat} and K_M by fitting the data to the Michaelis-Menten equation (Equation 8.3) using SigmaPlot (*Systat Software Inc.*). The turnover frequency of the enzyme, k_{cat} (s^{-1}), was derived from the software calculated V_{max} value using Equation 8.4 where V_{max} is the maximum velocity of the reaction ($\mu\text{M}/\text{s}$) and $[\text{E}]$ is enzyme concentration (μM , in this case 0.03 μM).

Equation 8.3

$$V = \frac{V_{\text{max}}[\text{S}]}{K_M + [\text{S}]}$$

Equation 8.4

$$k_{\text{cat}} = \frac{V_{\text{max}}}{[\text{E}]}$$

The error in this work is expressed as standard error of the mean (σ_M), defined by Equation 8.5 where n is the number of samples and σ is the standard deviation, defined by Equation 8.6. The σ_M was calculated from V values of Equation 8.2 and plotted as error bars on the Michaelis-Menten graphs.

Equation 8.5

$$\sigma_M = \frac{\sigma}{\sqrt{n}}$$

Equation 8.6

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

8.2 Biosynthetic compound preparation

8.2.1 Small-scale alcohol diphosphorylation

Each assay was performed in a separate Eppendorf for each alcohol (1 mL total reaction volume) containing buffer (100 mM TRIS pH 8.0, 0.5 mM DTT), IPK (2.1 μ M), THIM (63 μ M), PEP (80 mM) and the required alcohol (2 mg/mL). Isoprenol and prenol are the precursors to IDP and DMADP respectively but analogues of these alcohols were also used as substrates for diphosphorylation. The reactions were incubated at room temperature with shaking for 24 h. To check the reaction progress, a sample (500 μ L) was analysed by NMR with the addition of D₂O (50 μ L). The diphosphorylated substrate solutions were used for further assays without purification.

8.2.2 Preparation of diphosphate stocks

For immobilisation, purified His₆-tagged THIM and IPK were incubated on Ni²⁺ resin (10 mg and 0.3 mg respectively per 1 mL resin) in buffer (10 mL, 100 mM TRIS pH 8.0) with spinning on a disk rotator for 30 min. A separate resin was used for each alcohol to be phosphorylated. After removing the excess buffer, the phosphorylation reactions were performed on the resin with immobilised enzyme in buffer (100 mM TRIS pH 8.0, 10 mL total volume) containing the alcohol (20 mg, 1.72 mmol), KCl (0.2 M), MgCl₂ (5 mM), ATP (5 mM), PEP (80 mM) and PK (10 units/mL). After incubation for 24 hours at room temperature, the reaction solution was removed from the column and a sample of the reaction solution (500 μ L) was analysed by NMR with the addition of D₂O (50 μ L). The immobilised kinases were reused by repeating the above with fresh reagents until loss of activity (determined by NMR spectroscopy). The

resulting diphosphate stock solutions (10 mL) were stored at -20 °C and clarified by centrifugation once defrosted before use.

8.2.3 Screening of diphosphate analogues with zFPS

Each assay was performed in a separate glass vial containing buffer (100 mM TRIS pH 8.0), MgCl₂ (5 mM), β-CD (250 μM), diphosphorylated analogue stock solutions prepared as described in Section 8.2.1, in the ratio of 2:1 IDP or its analogue to DMADP or its analogue, and zFPS (17 μM). Controls with native substrates (IDP and DMADP) were also included. After incubation at room temperature overnight, the reaction mixtures were incubated with alkaline phosphatase (1 mg/mL) for 2 h at 37 °C before extraction in pentane and analysis by GC-MS.

8.2.4 Screening of diphosphate analogues with EZS

Each assay was prepared as described above but with the addition of EZS (12 μM). After incubation at room temperature overnight, the reaction mixtures were extracted in pentane (approximately 1 mL) and analysed by GC-MS.

8.2.5 (Z,Z)-FDP synthesis and purification

The reaction was assembled in buffer (25 mL total reaction volume, 100 mM TRIS pH 8.0), DTT (5 mM), IDP stock solution (5 mL), DMADP stock solution (2.5 mL), zFPS (1.25 μM) and MgCl₂ (10 mM). The reaction was incubated at 37°C overnight with mixing over the course of which the FDP was observed to precipitate. The reaction mixture was pelleted by centrifugation and washed with deionised water (10 mL) five times. The FDP was solubilised in a solution of sodium diphosphate (0.25 M) and purified by reverse phase flash chromatography (gradient of 2 – 50% H₂O in acetonitrile over 5 CV). Fractions corresponding to the product were combined and freeze dried to give (Z,Z)-FDP as a white powder (24 mg, 96 % isolated yield). ¹H NMR (500 MHz, D₂O) δ = 5.47 (t, *J* = 6.9 Hz, 1 H, C=CH), 5.20 (m, 2 H, 2 x C=CH), 4.44 (t, *J* = 9.8 Hz, 2 H, CH₂O), 2.14 (m, 4 H, 2(CH₂)), 2.08 (m, 4 H, 2 x CH₂), 1.75 (s, 3 H, CH₃), 1.68 (s, 6 H, 2 x CH₃), 1.62 (s, 3 H, CH₃) ppm; ¹³C NMR (101 MHz, D₂O) δ = 142.3 (CH₃CCHCH₂O), 137.1 (CH₃CCHCH₂), 133.6 (CH₃CCHCH₂), 124.9 (CCHCH₂O), 121.1 (CCHCH₂), 120.1 (CCHCH₂), 62.1 (CH₂O), 31.4 (CH₂), 31.1 (CH₂), 25.8 (CH₂), 24.8 (CH₂), 22.6 (CH₃), 22.4 (CH₃), 16.9 (CH₃) ppm; ³¹P NMR (243 MHz, D₂O) δ = -6.1 (d, *J* = 22.1 Hz, CH₂OPO₃PO₃), -10.2 (dt, *J* = 22.1, 6.4 Hz, CH₂OPO₃PO₃) ppm; HRMS (ESI⁻) [M-H]⁻

calculated for $C_{15}H_{27}P_2O_7$: 381.1232, found = 381.1219. The characterisation was in accordance with previous work.¹⁵³

8.2.6 7-Epi-zingiberene synthesis and isolation

The reaction was performed in a round bottomed flask containing buffer (100 mM TRIS pH 8.0, 25 mL total reaction volume), IDP stock solution (10 mL), DMADP stock solution (5 mL) purified zFPS (1.8 μ M), $MgCl_2$ (5 mM), purified EZS (5.75 μ M) and β -CD (250 μ M). After incubation at room temperature for 40 h, the reaction mixture was extracted exhaustively in pentane (3 x 5 mL). Samples of each extraction were analysed by GC-MS to ensure full recovery of the natural product. The solvent was removed slowly by reduced pressure to obtain 7-*epi*-zingiberene as a colourless oil (approx. 1 mg, 12% isolated yield). **¹H NMR** (500 MHz, $CDCl_3$) δ = 5.79 (dt, J = 9.8, 2.0 Hz, 1 H, C=CH), 5.67 (dd, J = 9.8, 3.1 Hz, 1 H, C=CH), 5.45 (m, 1 H, C=CH), 5.10 (m, 1 H, C=CH), 2.26 (m, 1 H, CH), 2.06 (m, 3 H, CH_2 + CHH), 1.93 (m, 1 H, CHH), 1.71 (q, J = 1.8 Hz, 3 H, CH_3), 1.68 (d, J = 1.1 Hz, 3 H, CH_3), 1.60 (s, 3 H, CH_3), 1.50 (m, 1 H, CHH), 1.42 (m, 1 H, CHH), 1.18 (m, 1 H, CH_3CH), 0.89 (m, 3 H, CH_3) ppm; **¹³C NMR** (101 MHz, $CDCl_3$) δ = 131.4, 131.3, 129.9, 128.4, 125.0, 120.7, 38.5, 36.2, 34.3, 26.5, 26.1, 25.9, 21.2, 17.8, 16.9 ppm; ***m/z*** (EI⁺): 204.4 (M⁺, 17%), 189.3 (3%), 161.3 (4%), 145.2 (5%), 132.2 (19%), 119.2 (100%), 105.2 (24%), 93.2 (88%), 91.2 (40%), 79.2 (10%), 77.2 (20%), 69.3 (19%), 55.2 (11%). The characterisation was in accordance with the literature.^{147,173}

8.2.7 GC-MS

Gas chromatography coupled mass spectrometry (GC-MS) was performed on a Clarus 680 fitted with an Elite-1 column (30 m \times 0.25 mm internal diameter) and Clarus SQ 8 C mass spectrometer (*Perkin Elmer*). The mass spectrometer was set to detect in the range of 50 - 800 *m/z* in EI⁺ mode. The general method injected 1 or 5 μ L of sample with an injection port temperature of 100 °C and a split ratio of 19:1, an initial temperature of 80 °C for 2 min and a ramp of 8 °C/min to a final temperature of 280 °C which was held for 3 min.

8.2.8 GC-FID

Gas chromatography with flame ionisation detector (GC-FID) was performed on an 7890A GC system (*Agilent*) fitted with a Rt-bDEXsm column (30 m \times 0.32 mm internal diameter, *Restek*). The general method injected 5 μ L of sample with an injection port temperature of 200 °C and

a split ratio of 19:1, an initial temperature of 80 °C for 1 min and a ramp of 8 °C/min to a final temperature of 200 °C which was held for 2 min.

8.2.9 NMR

¹H, ¹³C and ³¹P NMR spectra were recorded using either a Bruker DPX-600MHz, Bruker Avance 500 NMR or Bruker Avance 400 spectrometer (all *Bruker*) at 25 °C unless otherwise stated and are reported as chemical shifts in parts per million downfield from tetramethyl silane (TMS). Deuterated chloroform (CDCl₃) and deuterium oxide (D₂O) were used as sample solvents. Coupling constants (*J*) were calculated to the nearest 0.5 Hz and peak multiplicities were assigned with the following notation: s = singlet, d = doublet, t = triplet, q = quartet and m = multiplet.

8.3 Synthetic chemistry methods

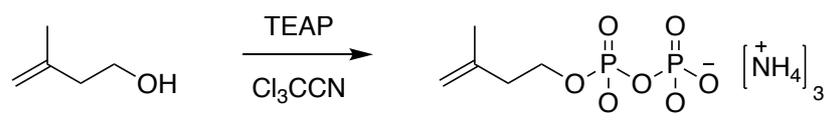
8.3.1 General experimental

All chemicals and general consumables were purchased from *Sigma-Aldrich*, *Alfa-Aesar* or *Fisher Scientific* unless otherwise stated and were used without further purification. Anhydrous tetrahydrofuran (THF) was collected from a SPS800 (*MBraun*) solvent purification system. Room temperature refers to 20 - 25 °C and temperatures of 0 and -10 °C obtained using ice/water and ice/acetone baths respectively.

Thin layer chromatography (TLC) was performed on silica coated aluminium plates (TLC Silica gel 60 F254, *Merck*) and visualised using potassium permanganate stain, made by dissolving KMnO₄ (1.5 g) of, K₂CO₃ (10 g), and NaOH (1.25 mL, 10% w/v) in deionised water (200 mL). Manual column chromatography was carried out using silica gel 60 (230-400 mesh, *Merck*) whilst automated (flash) chromatography and reverse phase purifications were carried out using a Biotage® Isolera Four with SNAP Ultra cartridges (*Biotage*). Lyophilisation was achieved using a Christ Alpha 2-4 LD Plus Freeze Dryer fitted with a RZ6 vacuum pump (*SciQuip*).

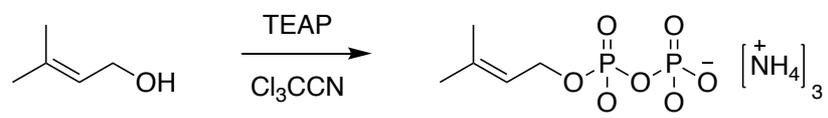
Compounds were characterised by NMR and GC-MS as described in Section 8.2. High- and low-resolution electrospray mass spectrometry (HRMS and LRMS) of small molecules was carried out on GCT Premier time of flight and LCT time of flight mass spectrometers (*Waters*).

8.3.2 IDP



Triethylammonium phosphate (TEAP) was prepared first by mixing triethylamine (45 mL) with acetonitrile (45 mL) and separately, phosphoric acid (80% wt. %, 15 mL) with acetonitrile (55 mL). The two solutions were slowly added together with stirring to give the TEAP solution. Isoprenol (1.76 mL, 17.9 mmol, 1 eq.) was dissolved in trichloroacetonitrile (45.4 mL, 0.45 mol, 26 eq.) and an equivalent volume of TEAP (47.16 mL) was added slowly and stirred for 5 min. Two further portions of TEAP (of the same volume) were added 5 min apart (15 min total reaction time). The reaction mixture was loaded directly onto silica pre-equilibrated with eluent (6:2.5:0.5 isopropanol:water:concentrated ammonium hydroxide, v/v/v). Care was taken to ensure the mixture was fully absorbed by the silica before adding eluent to prevent excessive contamination from TEAP generated side products. Fractions containing the diphosphate as judged by TLC (using the same eluent) were combined and excess isopropanol removed under reduced pressure. The remaining aqueous solution was lyophilised to give the ammonium salt of the title compound as a white powder (1.55 g, 30%). **$^1\text{H NMR}$** (500 MHz, D_2O): δ = 4.75 - 4.73 (1H, m, C=CH), 4.68 (1H, s, C=CH), 3.95 (1H, q, J = 6.6 Hz, CH_2O), 3.82 (1H, q, J = 6.6 Hz, $\text{CH}_2\text{CH}_2\text{O}$), 2.30 - 2.23 (2H, m, CH_2), 1.67 (3H, s, CH_3) ppm. **$^{13}\text{C NMR}$** (101 MHz, D_2O): δ = 144.4 ($\text{H}_3\text{C}=\text{C}(\text{CH}_2)_2$), 122.2 (C=CH₂), 64.9 (d, J = 6 Hz, CH_2O), 38.5 (d, J = 8 Hz, $\text{CH}_2\text{CH}_2\text{O}$), 22.3 (CH_3) ppm. **$^{31}\text{P NMR}$** (162 MHz, D_2O): δ = -8.1 (d, J = 21 Hz, $\text{CH}_2\text{OPO}_3\text{PO}_3$), -10.5 (d, J = 21 Hz, $\text{CH}_2\text{OPO}_3\text{PO}_3$) ppm. **HRMS** (ESI⁺): 244.9980 calculated for $\text{C}_5\text{H}_{11}\text{O}_7\text{P}_2$, found 244.9985. The characterisation was in accordance with the literature.¹⁷⁸

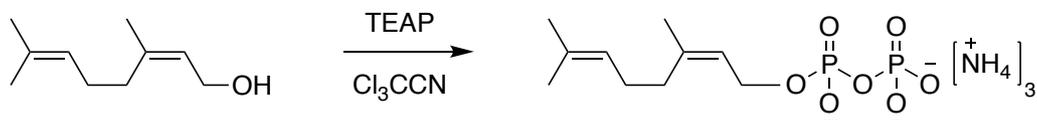
8.3.3 DMADP



DMADP was prepared using the TEAP method described in Section 8.3.1, using prenol (1.0 mL, 9.85 mmol, 1 eq.) in place of isoprenol. The volume of TEAP prepared was scaled as appropriate for the reaction. Fractions containing the diphosphate as judged by TLC were combined and excess isopropanol removed under reduced pressure. The remaining aqueous solution was lyophilised to give the ammonium salt of DMADP as a white powder (0.94 g, 32%). **$^1\text{H NMR}$** (400 MHz, D_2O): δ = 5.36 (1H, t, J = 6.8 Hz, C=CH), 4.37 (2H, t, J = 6.8 Hz, CH_2), 1.68 (3H, s, CH_3), 1.64 (3H, s, CH_3). **$^{31}\text{P NMR}$** (162 MHz, D_2O): δ = -8.3 (d, J = 21.3 Hz,

$\text{CH}_2\text{OPO}_3\text{PO}_3$), -10.5 (d, $J = 21.2$ Hz, $\text{CH}_2\text{OPO}_3\text{PO}_3$) ppm. **HRMS** (ESI⁻): 244.9980 calculated for $\text{C}_5\text{H}_{11}\text{O}_7\text{P}_2$, found 244.9986. The characterisation was in accordance with the literature.¹⁷⁸

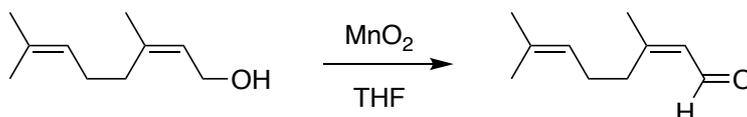
8.3.4 NDP



NDP was prepared using the TEAP method described in Section 8.3.1, using nerol (0.114 mL, 0.65 mmol, 1 eq.) in place of isoprenol. The volume of TEAP prepared was scaled as appropriate for the reaction. Fractions containing the diphosphate as judged by TLC were combined and excess isopropanol removed under reduced pressure. The remaining aqueous solution was lyophilised to give the ammonium salt of NDP as a white powder (68.6 mg, 29 %). **¹H NMR** (400 MHz, D_2O): $\delta = 5.45$ (1H, t, $J = 7.2$ Hz, $\text{C}=\text{CHCH}_2\text{O}$), 5.19 (1H, t, $J = 6.8$ Hz, $(\text{CH}_3)\text{C}=\text{CH}$), 4.42 - 4.49 (2H, m, CCHCH_2O), 2.18 - 2.11 (4H, m, 2 x CH_2), 1.75 (3H, s, CH_3), 1.67 (3H, s, CH_3), 1.61 (3H, s, CH_3). **¹³C NMR** (100 MHz, D_2O): $\delta = 142.8$ ($\text{CH}_3\text{CH}_2\text{C}=\text{C}$), 133.9 ($(\text{CH}_3)_2\text{C}=\text{C}$), 123.9 ($\text{C}=\text{CHCH}_2$), 120.7 ($\text{C}=\text{CHCH}_2$), 62.4 (CH_2O), 31.2 (CH_2), 25.9 (CH_2), 24.8 (CH_3), 22.5 (CH_3), 16.9 (CH_3) ppm. **³¹P NMR** (162 MHz, D_2O): $\delta = -8.3$ (d, $J = 23.0$ Hz, $\text{CH}_2\text{OPO}_3\text{PO}_3$), -10.6 (d, $J = 21.7$ Hz, $\text{CH}_2\text{OPO}_3\text{PO}_3$) ppm. **HRMS** (ESI⁻): 313.0606 calculated for $\text{C}_{10}\text{H}_{19}\text{O}_7\text{P}_2$, found 313.0613. The characterisation was in accordance with previous work.¹⁷³

8.3.5 [²H₂-1]-NDP

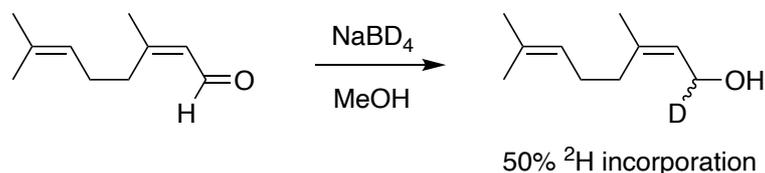
Neral



Neral (1.25 g, 8.1 mmol, 1 eq.) was added to a stirred suspension of activated MnO_2 (7.04 g, 81 mmol, 10 eq.) in THF (50 mL). The reaction was stirred at room temperature until completion (approximately 24 h), as judged by TLC. The mixture was filtered over celite and purified by flash chromatography (10% ethyl acetate in hexane) to give a colourless oil (1.21 g, 97%) which was used immediately. **¹H NMR** (500 MHz, CDCl_3): $\delta = 9.90$ (1H, d, $J = 8.3$ Hz, $\text{HC}=\text{O}$), 5.88 (1H, d, $J = 8.1$ Hz, $(\text{CH}_3)\text{C}=\text{CH}$), 5.10 (1H, t, $J = 7.1$ Hz, $(\text{CH}_3)\text{C}=\text{CH}$), 2.58 (2H, t, $J = 7.5$ Hz, CH_2), 2.23 (2H, q, $J = 7.5$ Hz, CH_2), 1.98 (3H, s, CH_3), 1.68 (3H, s, CH_3), 1.59 (3H, s, CH_3) ppm. **¹³C NMR** (101 MHz, CDCl_3): $\delta = 191.0$ ($\text{HC}=\text{O}$), 164.0 ($\text{HC}=\text{OC}=\text{CH}$), 133.9

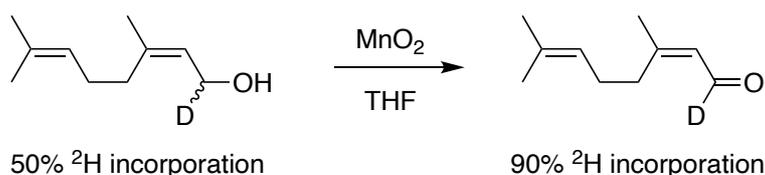
(C=C), 128.8 (C=C), 122.4 (C=C), 32.7 (CH₂), 27.2 (CH₂), 25.8 (CH₃), 25.2 (CH₃), 17.9 (CH₃) ppm. The characterisation was in accordance with previous work.¹⁵³

[²H₁-1]-Nerol (50% ²H incorporation)



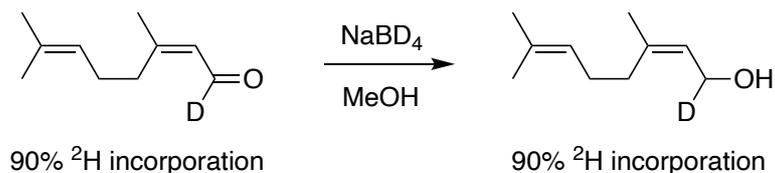
Under inert conditions, sodium borodeuteride (NaBD₄, 1.01 g, 24.3 mmol, 3 eq.) was added to a stirred solution of nerol (1.21 g, 8.1 mmol, 1 eq.) in methanol (60 mL). The reaction was stirred at room temperature until completion (approximately 30 min), as judged by TLC. The mixture was quenched with HCl (1 M, 15 mL) and extracted with dichloromethane (DCM, 3 x 15 mL). Combined organic extracts were washed with water (3 x 15 mL) and brine (2 x 15 mL), dried over anhydrous sodium sulfate and filtered under gravity. The solvent was removed under reduced pressure and the compound purified by flash chromatography (10% ethyl acetate in hexane) to give a colourless oil (1.1 g, 87%, 50% ²H incorporation). **¹H NMR** (600 MHz, CDCl₃): δ = 5.45 (1H, d, *J* = 7.3 Hz, C=CH), 5.10 (1H, m, CDHC=CH), 4.07 (1H, d, *J* = 7.1 Hz, HDCOH), 2.11 - 2.07 (4H, m, 2 x CH₂), 1.75 (3H, s, CH₃), 1.69 (3H, s, CH₃), 1.60 (3H, s, CH₃) ppm. **¹³C NMR** (150 MHz, CDCl₃): δ = 140.2 (CH₃CH₂C=C), 132.6 ((CH₃)₂C=C), 124.5 (C=CHCH₂), 123.9 (C=CHCH₂), 58.9 - 58.6 (m, CHDO), 32.1 (CH₂), 26.63 (CH₂), 25.8 (CH₃), 23.6 (CH₃), 17.8 (CH₃) ppm. **²H NMR** (92 MHz, CDCl₃): δ = 4.10 ppm.

[²H₁-1]-Neral (90% ²H incorporation)



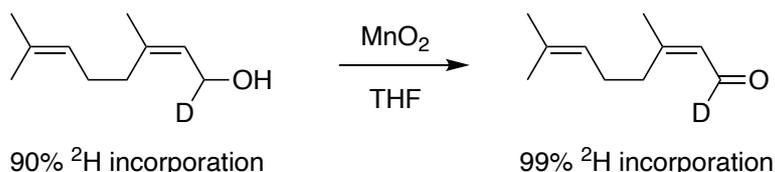
[²H₁-1]-Neral was prepared from [²H₁-1]-nerol (50% ²H incorporation, 1.1 g, 7.1 mmol, 1 eq.) as per the procedure for nerol *vide supra*. The reaction mixture was purified to give [1-²H₁]-neral (1.04 g, 95%, 90% ²H incorporation) which was used immediately. **¹H NMR** (500 MHz, CDCl₃): δ = 9.89 (1H, d, *J* = 8.3 Hz, HC=O), 5.87 (1H, s, (CH₃)C=CH), 5.11 - 5.07 (1H, m, (CH₃)C=CH), 2.58 (2H, t, *J* = 8.2 Hz, CH₂), 2.23 (2H, dd, *J* = 7.5 Hz, CH₂), 1.98 (3H, s, CH₃), 1.68 (3H, s, CH₃), 1.59 (3H, s, CH₃) ppm.

[²H₂-1]-Nerol (90% ²H incorporation)



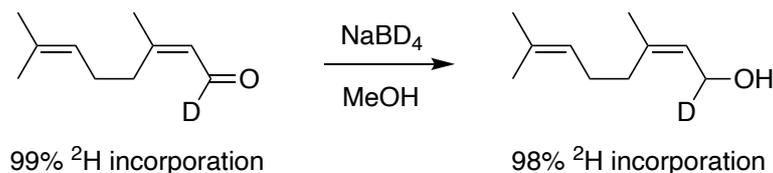
[²H₂-1]-Nerol was prepared from [²H₁-1]-neral (90% ²H incorporation, 1.04 g, 6.7 mmol, 1 eq.) as per the procedure for [²H₁-1]-nerol *vide supra*. The reaction mixture was purified to give a colourless oil (775 mg, 74%, 90% ²H incorporation). **¹H NMR** (500 MHz, CDCl₃): δ = 5.44 (1H, s, C=CH), 5.11 - 5.09 (1H, m, C=CH), 4.07 (1H, d, *J* = 7.1 Hz, HD²COH), 2.12 - 2.07 (4H, m, 2 x CH₂), 1.75 (3H, s, CH₃), 1.69 (3H, s, CH₃), 1.60 (3H, s, CH₃) ppm. **¹³C NMR** (101 MHz, CDCl₃): δ = 140.3 (CH₃CH₂C=C), 132.6 ((CH₃)₂C=C), 124.4 (C=CHCH₂), 124.0 (C=CHCH₂), 53.6 (CHDO), 32.1 (CH₂), 26.7 (CH₂), 25.8 (CH₃), 23.6 (CH₃), 17.8 (CH₃) ppm.

[²H₁-1]-Neral (99% ²H incorporation)



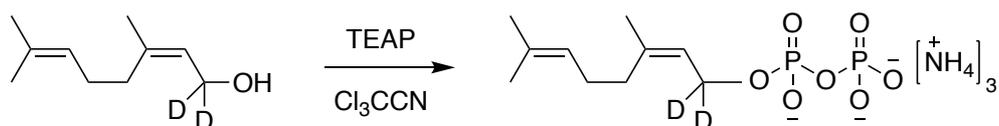
[²H₁-1]-Neral was prepared from [²H₂-1]-nerol (90% ²H incorporation, 500 mg, 3.2 mmol, 1 eq.) as per the procedure for neral *vide supra*. The reaction mixture was purified to give [¹⁻²H₁]-neral (444 mg, 90%, 99% ²H incorporation). **¹H NMR** (500 MHz, CDCl₃): δ = 9.90 (1H, d, *J* = 8.5 Hz, HC=O), 5.87 (1H, s, C=CH), 5.12 - 5.08 (1H, m, C=CH), 2.58 (2H, t, *J* = 8.0 Hz, CH₂), 2.24 (2H, m, CH₂), 1.98 (3H, s, CH₃), 1.68 (3H, s, CH₃), 1.59 (3H, s, CH₃) ppm. **¹³C NMR** (100 MHz, CDCl₃): δ = 190.7 (CO), 164.0 (COC=CH), 133.9 (C=C), 128.8 (C=C), 122.4 (C=C), 32.8 (CH₂), 27.2 (CH₂), 25.8 (CH₃), 25.2 (CH₃), 17.9 (CH₃) ppm.

[²H₂-1]-Nerol (98% ²H incorporation)



[²H₂-1]-Nerol was prepared from [²H₁-1]-neral (99% ²H incorporation, 444 mg, 2.9 mmol, 1 eq.) as per the procedure for [²H₁-1]-nerol *vide supra*. The reaction mixture was purified to give a colourless oil (385 mg, 85%, 98% ²H incorporation). **¹H NMR** (600 MHz, CDCl₃): δ = 5.44 (1H, s, C=CH), 5.12 - 5.09 (1H, m, CD₂C=CH), 2.13 - 2.06 (4H, m, 2 x CH₂), 1.75 (3H, d, *J* = 2 Hz, CH₃), 1.69 (3H, s, CH₃), 1.61 (3H, s, CH₃) ppm. **¹³C NMR** (150 MHz, CDCl₃): δ = 140.2 (CH₃CH₂C=C), 132.6 ((CH₃)₂C=C), 124.4 (C=CHCH₂), 123.9 (C=CHCH₂), 58.9 - 58.1 (m, CD₂O), 32.1 (CH₂), 26.6 (CH₂), 25.8 (CH₃), 23.6 (CH₃), 17.8 (CH₃) ppm. **²H NMR** (92 MHz, CDCl₃): δ = 4.07 ppm.

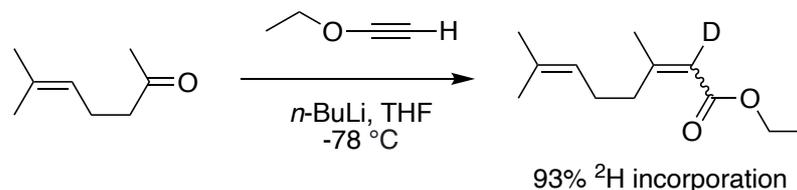
[²H₂-1]-NDP



[²H₂-1]-Neryl diphosphate ([²H₂-1]-NDP) was prepared using the TEAP method described in Section 8.3.1, using [²H₂-1]-nerol (98% ²H incorporation, 100 mg, 0.64 mmol, 1 eq.) in place of isoprenol. The volume of TEAP prepared was scaled as appropriate for the reaction. Fractions containing the diphosphate as judged by TLC were combined and excess isopropanol removed under reduced pressure. The remaining aqueous solution was lyophilised to give the ammonium salt of [²H₂-1]-NDP as a white powder (68.6 mg, 29 %, 98% ²H incorporation). **¹H NMR** (400 MHz, D₂O): δ = 5.46 (1H, s, C=CHCD₂O), 5.21 (1H, m, (CH₃)C=CH), 2.20 - 2.13 (4H, m, 2 x CH₂), 1.78 (3H, s, CH₃), 1.70 (3H, s, CH₃), 1.63 (3H, s, CH₃). **¹³C NMR** (151 MHz, D₂O): δ = 143.1 (CH₃(CH₂)C=C), 134.0 ((CH₃)₂C=C), 123.9 (C=CHCH₂), 120.2 (C=CHCH₂), 62.3 (CD₂O), 31.2 (CH₂), 25.9 (CH₂), 24.8 (CH₃), 22.6 (CH₃), 16.9 (CH₃) ppm. **²H NMR** (92 MHz, CDCl₃): δ = 4.35 ppm. **³¹P NMR** (202 MHz, D₂O): δ = -9.6 (d, *J* = 22.2 Hz, CD₂OPO₃PO₃), -10.6 (d, *J* = 20.2 Hz, CD₂OPO₃PO₃) ppm.

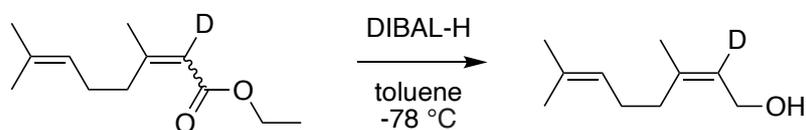
8.3.6 [²H₁-2]-NDP

Ethyl 2-deutero-3,7-dimethylocta-2,6-dienoate



Under inert atmosphere conditions, ethoxyacetylene (2.1 mL, 8.5 mmol, 2.5 eq.) dissolved in anhydrous THF (30 mL) was cooled to -78 °C and *n*-butyllithium (3.4 mL, 8.5 mmol, 2.5 eq.) was added dropwise. The mixture was stirred for 1 h after which 6-methylhept-5-en-2-one (0.5 mL, 3.4 mmol, 1 eq.) was added and the reaction allowed to room temperature with stirring for 16 h. The mixture was then cooled on ice and quenched with D₂SO₄ (1M in D₂O, 10 mL). Reaction products were extracted with diethyl ether (3 x 15 mL), washed with D₂O (3 x 15 mL) and brine (3 x 15 mL), dried over anhydrous Mg₂SO₄, filtered under gravity and concentrated under reduced pressure. The crude oil was purified by flash chromatography on silica gel (10 % ethyl acetate in hexane) to give the title compound (a mixture of isomers) as a clear oil (657 mg, 92 %, 93% ²H incorporation). ¹H NMR (600 MHz, CDCl₃): δ = 5.15 (1H, t, *J* = 7.1 Hz, C=CH), 5.01 (1H, s, C=CH), 4.14 (4H, quintet, *J* = 6.9 Hz, COOCH₂), 2.63 (2H, t, *J* = 8.1 Hz), 2.15 (8H, m), 1.88 (3H, s, CH₃), 1.68 (6H, s, 2 x CH₃), 1.61 (3H, s, CH₃), 1.60 (3H, s, CH₃), 1.27 (6H, m, 2 x CH₂) ppm. ¹³C NMR (150 MHz, CDCl₃): δ = 167.0, 166.5, 160.1, 159.8, 132.6, 132.3, 123.9, 123.2, 59.6, 59.5, 41.0, 33.6, 26.9, 26.2, 25.8, 25.4, 18.9, 17.8, 17.7, 14.5 ppm. ²H NMR (92 MHz, CDCl₃): δ = 5.67 ppm. *m/z* (EI⁺): 197.3 ([M]⁺, 4%), 182.2 (2%), 152.3 (20%), 129.2 (30%), 124.3 (30%), 114.2 (8%), 108.2 (10%), 101.2 (47%), 83.2 (18%), 69.3 (100%), 53.3 (5%).

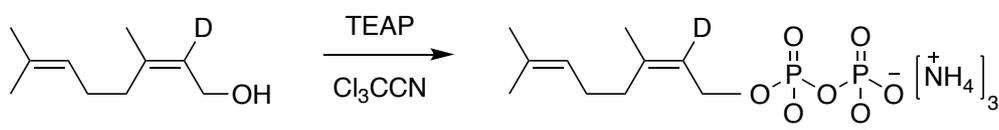
[2-²H₁]-Nerol



Under inert atmosphere conditions, ethyl 2-deutero-3,7-dimethylocta-2,6-dienoate (93% ²H incorporation, 620 mg, 3.14 mmol, 1 eq.) was dissolved in anhydrous toluene (30 mL) and cooled to -78 °C. DIBAL-H (15.7 mL, 15.7 mmol, 5 eq.) was added dropwise with stirring until consumption of the starting material was observed (approx. 2 h) by TLC. The mixture was warmed on ice and quenched with saturated potassium sodium tartrate solution (30 mL). The biphasic mixture was allowed to separate by stirring overnight before the reaction products were extracted with DCM (3 x 20 mL), washed with brine (3 x 15 mL), dried over anhydrous

Mg₂SO₄, filtered under gravity and concentrated under reduced pressure. The crude oil was purified by flash chromatography on silica gel (5% ethyl acetate in hexane), achieving separation of the mixture of alcohol isomers, to give the title compound as a clear oil (210 mg, 43 %, 93% ²H incorporation). **¹H NMR** (400 MHz, CDCl₃): δ = 5.10 (1H, s, C=CH), 4.08 (2H, s, CH₂OH), 2.09 (4H, m, 2 x CH₂), 1.75 (3H, s, CH₃), 1.69 (3H, s, CH₃), 1.60 (3H, s, CH₃) ppm. **¹³C NMR** (151 MHz, CDCl₃): δ = 140.1 (CH₃CH₂C=C), 132.6 ((CH₃)₂C=C), 124.4 (C=CHCH₂), 124.0 (C=CHCH₂), 59.1 (CH₂O), 32.1 (CH₂), 26.7 (CH₂), 25.8 (CH₃), 23.5 (CH₃), 17.8 (CH₃) ppm. **²H NMR** (61 MHz, CDCl₃): δ = 5.50 ppm. **m/z** (EI⁺): 155.3 (M⁺, 4%), 137.2 (15%), 122.2 (19%), 94.1 (65%), 69.2 (100%).

[2-²H₁]-NDP



[2-²H₁]-Neryl diphosphate ([2-²H₁]-NDP) was prepared using the TEAP method described in Section 8.3.1 using [2-²H₁]-nerol (93% ²H incorporation, 100 mg, 0.64 mmol, 1 eq.) in place of isoprenol. The volume of TEAP prepared was scaled as appropriate for the reaction. Fractions containing the diphosphate as judged by TLC were combined and excess isopropanol removed under reduced pressure. The remaining aqueous solution was lyophilised to give the ammonium salt of [1-²H₂]-NDP as a white powder (68.6 mg, 29 %, 93% ²H incorporation). **¹H NMR** (500 MHz, D₂O): δ = 5.21 (1H, m, C=CH), 4.48 - 4.38 (2H, m, H₂CO), 2.18 - 2.14 (4H, m, 2 x CH₂), 1.78 (3H, s, CH₃), 1.70 (3H, s, CH₃), 1.64 (3H, s, CH₃) ppm. **¹³C NMR** (126 MHz, D₂O): δ = 142.9 (CH₃CH₂C=C), 134.0 ((CH₃)₂C=C), 123.9 (C=CHCH₂), 120.9 (C=CHCH₂), 62.7 (CH₂O), 31.2 (CH₂), 25.9 (CH₂), 24.8 (CH₃), 22.5 (CH₃), 17.0 (CH₃) ppm. **³¹P NMR** (202 MHz, D₂O): δ = -10.51 (d, *J* = 20.2 Hz, CH₂OPO₃PO₃), -10.80 (d, *J* = 20.2 Hz, CH₂OPO₃PO₃) ppm. **HRMS** (ESI⁻): 314.0669 calculated for C₁₀¹H₁₈²H₁O₇P₂, found 314.0675.

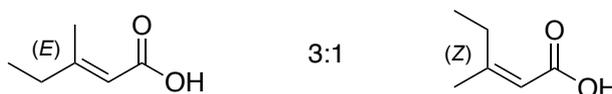
8.3.7 3-Methyl-penten-1-ol

Ethyl 3-methyl-2-pentenoate



Sodium hydride (60% dispersion in mineral oil, 4.2 g, 53.2 mmol, 1.1 eq.) and dry tetrahydrofuran (160 mL) were added to an argon-flushed, flame-dried round bottom flask. Triethyl phosphonoacetate (23 g, 105 mmol, 1.05 eq.) was then added dropwise and the solution was stirred under argon at 0°C for 1 h. To the resulting yellow solution 2-butanone (7 mL, 100 mmol, 1.0 eq.) was added dropwise and the reaction stirred at room temperature for 16 h. The reaction mixture was quenched by addition of water (~60 mL) and stirred for 4 h. The organic phase was separated and the aqueous phase extracted with ethyl acetate (2 x 100 mL). The organic layers were combined, dried with magnesium sulfate, filtered and the solvent removed under reduced pressure to yield a crude 3:1 *E:Z* mixture of ethyl 3-methyl-2-pentenoate as a yellow oil (9 g, 63 mmol, 63%). ¹H NMR (500 MHz, CDCl₃): δ = 5.63 (1H, m, CH), 2.61 (2H, q, *J* = 7.6 Hz, CH₂), 2.26 – 2.09 (5H, m, CH₃ + CH₂), 1.86 (3H, s, CH₃), 1.31 – 1.18 (5H, m, CH₃ + CH₂), 1.05 (3H, t, *J* = 7.4, CH₃) ppm. The characterisation was in accordance with the literature.¹⁶¹

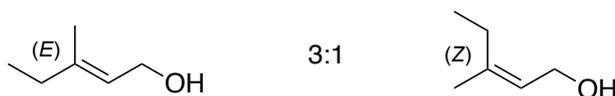
3-Methyl-2-pentanoic acid



Methanol (3.5 mL), 1M NaOH (73 mL, 73 mmol, 1.15 eq.) and the crude ethyl 3-methyl-2-pentenoate (9 g, 63 mmol, eq.) were added to an argon-flushed, flame-dried round bottom flask and stirred at 50°C for 24 h. The reaction mixture was then cooled on ice and diethyl ether (~50 mL) added with stirring, and the subsequent organic layer discarded. The aqueous layer was acidified to pH 1 by dropwise addition of concentrated hydrochloric acid and then extracted with diethyl ether (4 x 40 mL). The organic fractions were combined, dried with magnesium sulfate, filtered and the solvent removed under reduced pressure to give a crude pale-yellow oil. In an attempt to separate the *E* and *Z* isomers, different silica chromatography conditions were tried (including 0-20% EtOAc/cyclohexane, 0-10% EtOAc/cyclohexane, and 1% Methanol/DCM) however no separation was achieved. The product mixture was collected and the solvent removed under reduced pressure to leave a 3:1 *E:Z* mixture of 3-methyl-2-pentenoic acid as a clear oil (2.9 g, 25.2 mmol, 40%). ¹H NMR (500 MHz, CDCl₃): δ = 5.69 (1H, m, CH), 5.66 (1H, m, CH), 2.64 (2H, q, *J* = 7.5 Hz, CH₂), 2.24 – 2.16 (5H, m, CH₃ + CH₂),

1.92 (3H, s, CH_3), 1.08 (3H, t, $J = 7.5$, CH_3) ppm. ^{13}C NMR (125 MHz, $CDCl_3$): $\delta = 172.6$, 171.9, 165.5, 164.9, 115.2, 114.2, 34.2, 26.9, 25.1, 19.2, 12.7, 12.0 ppm. m/z (EI^+): 105.04 (100%), 114.08 (32%), 115.08 (53%). The characterisation was in accordance with the literature.¹⁶¹

3-Methyl-2-penten-1-ol



The *E:Z* mixture of 3-methyl-2-pentenoic acid (700 mg, 6 mmol, 1 eq.) and dry diethyl ether (35 mL) were added to an argon-flushed, flame-dried round bottom flask. The reaction was cooled to $-78^\circ C$ and 1M DIBAL in cyclohexane (8 mL, 7.2 mmol, 1.2 eq.) was added dropwise. The reaction mixture was stirred at room temperature for 4 h and quenched with saturated potassium sodium tartrate solution (~50 mL). The biphasic mixture was allowed to separate by stirring overnight before the reaction products were extracted with diethyl ether (2 x 25 mL). The organic fractions were combined, dried with magnesium sulfate, filtered and the solvent removed under reduced pressure to give the crude product as a pale-yellow oil. The oil was purified using silica chromatography; the relevant fractions were collected, combined and the solvent removed under reduced pressure to yield a 3:1 *E:Z* mixture of 3-methyl-2-penten-1-ol as a volatile, pale-yellow oil (242 mg, 40%). 1H NMR (500 MHz, $CDCl_3$): δ 5.54 – 5.24 (1H, m, CH), 4.13 (2H, d, $J = 7.0$ Hz, CH_2CO), 2.12 – 2.05 (2H, q, $J = 7.5$ Hz, CH_3CH_2), 2.05 – 1.96 (2H, q, $J = 7.5$ Hz, CH_2), 1.73 (3H, s, CH_3), 1.67 (3H, s, CH_3), 1.33 (1H, s, CH), 1.09 – 0.90 (3H, t, $J = 7.5$ Hz, CH_3). ^{13}C NMR (101 MHz, $CDCl_3$): $\delta = 142.3$, 141.8, 123.4, 122.1, 59.6, 59.1, 32.4, 25.1, 23.1, 16.3, 13.3, 12.5. m/z (EI^+): 101.06 (100%), 101.10 (35%). The characterisation was in accordance with the literature.¹⁶¹

8.4 Chemical ecology methods

The chemical ecology work of this project was carried out at Rothamsted Research, Harpenden, from April 2018 – July 2018.

8.4.1 Materials

The aphid cultures used in the laboratory experiments were started from single apterous virginopara collected in the field and were maintained under controlled conditions (16:8 h

light:dark, 20 (\pm 2) $^{\circ}$ C). *Sitobion avenae* were reared on wheat (*Triticum aestivum*) var. Tybalt and *Myzus persicae* on Chinese cabbage (*Brassica pekinensis*) var. Wong Bok.

Laboratory solvents were purchased from *Fisher Scientific* and were of analytical grade and were used distilled as noted. The 'RRes' (*E*)- β -farnesene (EBF, 97.4%) was synthesised at Rothamsted Research by Dr. David Withall and the 'Bedoukian' EBF (97.7%) was supplied from an external grant stakeholder. (*S*)-Germacrene D (germacrene D, 93.3%) was previously purified from Germacrene D Natural (40%, R.C. TREATT & LtD) by silica column chromatography and high-performance liquid chromatography (HPLC) and a small impurity of caryophyllene remained. (*RS*)-Curcumene (90%) was donated from the group of Prof. Allemann, Cardiff University, synthesised according to the method of Du *et al.*²¹⁶ 7-*Epi*-zingiberene (98.7%) was extracted from leaves of *Solanum habrochaites* accession PI12782 and purified by silica column chromatography and HPLC.

8.4.2 Behavioural bioassays

Four-arm olfactometer bioassay

The behavioural responses of individual alate aphids (either *S. avenae* or *M. persicae*) to different compounds were observed using a Perspex four-arm olfactometer (12 cm diameter x 2 cm depth) lit from above by diffuse, uniform lighting and maintained at 23 $^{\circ}$ C, placed within a black box to avoid background distractions. The bottom of the apparatus was lined with filter paper (*Whatman No 1*) and the parts of the olfactometer were fitted tightly together with screws. Disposable 5 mL syringes (*Plastipak*) were inserted into the apparatus at the inlets as the four arms. The test solutions (10 μ L) were applied to small strips of equally sized filter paper and placed in the syringes. Three arms were treated with distilled hexane (10 μ L), whilst one arm contained filter paper treated with the compound of interest in hexane (0.1 - 1 μ g/ μ L). The solvent was allowed to evaporate for 30 s. Each aphid was introduced into the apparatus through the hole in the top and air was then drawn through each arm by attaching a vacuum pump to the hole at 250 mL/min (62.5 ml/min/arm). The time spent in each arm and number of entries into each arm was recorded using *OLFA* software over 16 min. The olfactometer was rotated 90 $^{\circ}$ every 2 min to avoid directional bias. Each assay comprised of at least six replicates and the mean time spent in treated and untreated arms were analysed by a paired t-test (*Genstat, 19th ed.*).

Aphid alarm response bioassay

Individual Chinese cabbage plants were grown to the three- to four-leaf stage and clip cages containing five to seven apterous *Myzus persicae* aphids were secured onto the underside of

the flattest leaves (1 clip cage per plant). The plants were kept under controlled conditions (16:8 h light:dark, 20 (\pm 2) °C) and the adults were removed after 3 days whilst the nymphs were kept in the clip cage for a further 4 days. The number of aphids were counted and the leaf harbouring the aphid colony (20 - 55 individuals) was then placed in a constant flow of air provided by an extraction fan at 22 °C. Hexane (1 μ L) or EBF (RRes or Bedoukian, 1 μ L) at a concentration of 0.8 μ g/ μ l in hexane was applied to the leaf 1 cm from the edge of the colony on the upwind side. The percentage of aphids that moved after 1 and 15 min were counted and the mean percentage response to each treatment was calculated. After conversion into proportion data, the logit function was performed to normalise distributions and the transformed data were then analysed by ANOVA and the LSD test to compare treatments (*Genstat, 19th ed.*).

8.4.3 Germacrene D release rate and stability

Sachet preparation

Clear and black PVC tubing (1000-gauge permeability, *External supplier*) were cut to 4 cm² sized pieces and one end sealed using a heat sealer. Neat germacrene D (1 μ L) was applied to 1 cm² pieces of thin yellow sponge (*Sainsbury's*) then placed inside the sachet, immediately sealed and suspended on small wire frames. Gloves were worn throughout and equipment cleaned with ethanol before use to avoid contamination.

Germacrene D release rate

Clear and black PVC sachets (3 replications respectively) each containing a sponge dosed with germacrene D suspended on wire were placed in individual glass vessels and sealed (Figure 6.3). Before use, all glassware and fittings were cleaned and heated overnight in an oven maintained above 150 °C and the charcoal air filters for were heated for 2 hours in an oven maintained above 150 °C under a constant flow of N₂. Volatiles were collected by air entrainment (500 mL/min air flow in, 350 mL/min air flow out per glass vessel) on Porapak Q 50-80 (50 mg, *Waters*) in glass tubes (Porapak filters) inserted into the collection ports above the vessels over a period of 7 days. Samples were eluted from the Porapak filters with freshly redistilled diethyl ether (750 μ L). This was repeated for a total of 3 weeks from the same sachet, exchanging only the Porapak filter at each time point. Following volatile collection any remaining compounds were extracted from the sponges in redistilled hexane with pentadecane (0.2 μ g/mL, *Koch-Light Laboratories*) used as an internal standard (5 mL) for a period of 24 h, then concentrated to 1 mL under a stream of N₂. Pentadecane was not used in the second experiment as NMR spectroscopic analysis was performed on sponge samples.

Controls were prepared by entrainment of an empty vessel and both heat-sealed and unsealed black and clear sachets that did not contain the sponge dosed with compound.

Germacrene D stability

One stability experiment was carried out alongside the first of the release rate experiments described above. Clear and black PVC sachets (3 replications for each time-point), each containing a sponge dosed with germacrene D suspended on wire, were placed in the same glasshouse compartment as the air entrained sachets (Figure 6.6). After a period of 7 days, any remaining compounds were extracted from the sponges in redistilled hexane (5 mL) containing pentadecane (0.2 µg/µL) as an internal standard for 24 h with the solvent then concentrated to 1 mL under a stream of N₂. For the initial time point, the sponges were dosed with germacrene D and then immediately extracted. As a control, sponges without germacrene D were also extracted.

8.4.4 Analytical techniques

Gas chromatography

Samples were analysed by gas chromatography (GC) using an *Agilent 6890* GC equipped with a non-polar HP1 silica capillary column (50 m length x 0.32 mm inner diameter, 0.52 µm film thickness), cool-on-column injector (inlet at 30 °C) and flame ionization detector (FID). The GC oven temperature was maintained at 30 °C for 0.1 min, increased by 5 °C per min to 150 °C, held for 0.1 min, increased by 10 °C per min to 230 °C and held for 25 min.

GC-MS

GC-MS was performed on a *Hewlett Packard 5972* MSD mass spectrometer attached to a *Hewlett Packard 5890* GC fitted with a non-polar HP1 silica capillary column (50 m length x 0.32 mm inner diameter, 0.52 µm film thickness) and cool-on-column injector. The GC oven temperature was maintained at 40 °C for 1 min, increased by 5 °C per min to 250 °C min and held for 17.2 min. Tentative compound identifications were obtained using the NIST mass spectral database or by spectra interpretation.

Calibration curve

To calculate the mass of germacrene D in samples a calibration curve was established using one replicate per concentration and the resulting line equation ($y = 17.834x + 3.2269$) used where x is equivalent to mass and y is peak area.

NMR spectroscopy

NMR spectra were recorded using a *Bruker* spectrometer with a 5 mm BBO BB-1H probe, set at 500 MHz for ^1H spectra and 125 MHz for ^{13}C spectra. Deuterated chloroform (CDCl_3 , *Sigma Aldrich*) was stored over 4 Å molecular sieves and used as both sample solvent and internal standard. Additional 2D-NMR spectroscopy experiments were performed when necessary to aid in assignments.

References

- 1 D. W. Christianson, *Chem. Rev.*, 2017, **117**, 11570–11648.
- 2 D. J. McGarvey and R. Croteau, *Plant Cell*, 1995, **7**, 1015–1026.
- 3 S. G. Hillier and R. Lathe, *Eur. J. Endocrinol.*, 2019, **242**, 9–22.
- 4 W. D. Nes, *Chem. Rev.*, 2011, **111**, 6423–6451.
- 5 C. I. Cazzonelli, *Func. Plant Biol.*, 2011, **38**, 833.
- 6 E. Pichersky and J. Gershenzon, *Curr. Opin. Plant Biol.*, 2002, **5**, 237–243.
- 7 M. B. Quin, C. M. Flynn and C. Schmidt-Dannert, *Nat. Prod. Rep.*, 2014, **31**, 1449–1473.
- 8 E. Pichersky and R. A. Raguso, *New Phytol.*, 2018, **220**, 692–702.
- 9 W. Dhifi, S. Bellili, S. Jazi, N. Bahloul and W. Mnif, *Medicines*, 2016, **3**, 25.
- 10 S. C. Phulara, P. Chaturvedi and P. Gupta, *Appl. Environ. Microbiol.*, 2016, **82**, 5730–5740.
- 11 S. Cherian, S. B. Ryu and K. Cornish, *Plant Biotechnol. J.*, 2019, **17**, 2041–2061.
- 12 R. Croteau, R. E. B. Ketchum, R. M. Long, R. Kaspera and M. R. Wildung, *Phytochem. Rev.*, 2006, **5**, 75–97.
- 13 C. J. Paddon and J. D. Keasling, *Nat. Rev. Microbiol.*, 2014, **12**, 355–367.
- 14 D. M. Soderlund, *Arch. Toxicol.*, 2012, **86**, 165–181.
- 15 L. Ruzicka, *Experientia*, 1953, **9**, 357–367.
- 16 A. Frank and M. Groll, *Chem. Rev.*, 2017, **117**, 5675–5703.
- 17 E. Vranová, D. Coman and W. Gruissem, *Mol. Plant*, 2012, **5**, 318–333.
- 18 P. M. Dewick, *Nat. Prod. Rep.*, 2002, **19**, 181–222.
- 19 H. M. Mizioroko, *Arch. Biochem. Biophys.*, 2011, **505**, 131–143.
- 20 J. M. Vinokur, T. P. Korman, Z. Cao and J. U. Bowie, *Biochemistry*, 2014, **53**, 4161–4168.
- 21 N. Dellas, S. T. Thomas, G. Manning and J. P. Noel, *eLife*, 2013, **2**, 1–18.

- 22 M. Rohmer, M. Knani, P. Simonin, B. Sutter and H. Sahn, *Biochem. J.*, 1993, **295**, 517–524.
- 23 D. Arigoni, S. Sagner, C. Latzel, W. Eisenreich, A. Bacher and M. H. Zenk, *Biochemistry*, 1997, **94**, 10600–10605.
- 24 M. Rohmer, *Nat. Prod. Rep.*, 1999, **16**, 565–574.
- 25 W. Eisenreich, A. Bacher, D. Arigoni and F. Rohdich, *Cell. Mol. Life Sci.*, 2004, **61**, 1401–1426.
- 26 A. Hemmerlin, J. L. Harwood and T. J. Bach, *Prog. Lipid Res.*, 2012, **51**, 95–148.
- 27 K. Ogura and T. Koyama, *Chem. Rev.*, 1998, **98**, 1263–1276.
- 28 C. D. Poulter and H. C. Rilling, *Acc. Chem. Res.*, 1978, **11**, 307–313.
- 29 Y. Kharel, S. Takahashi, S. Yamashita and T. Koyama, *FEBS J.*, 2006, **273**, 647–657.
- 30 H. V. Thulasiram, H. K. Erickson and C. D. Poulter, *Science*, 2007, **316**, 73–76.
- 31 P. H. Liang, T. P. Ko and A. H. J. Wang, *Eur. J. Biochem.*, 2002, 269, 3339–3354.
- 32 D. W. Christianson, *Chem. Rev.*, 2017, **117**, 11570–11648.
- 33 A. L. Schillmiller, I. Schauvinhold, M. Larson, R. Xu, A. L. Charbonneau, A. Schmidt, C. Wilkerson, R. L. Last and E. Pichersky, *Proc. Natl. Acad. Sci. U. S. A.*, 2009, **106**, 10865–10870.
- 34 C. Sallaud, D. Rontein, S. Onillon, F. Jabes, P. Duffe, C. Giacalone, S. Thoraval, C. Escoffier, G. Herbette, N. Leonhardt, M. Causse and A. Tissier, *Plant Cell*, 2009, **21**, 301–317.
- 35 J. Zi, Y. Matsuba, Y. J. Hong, A. J. Jackson, D. J. Tantillo, E. Pichersky and R. J. Peters, *J. Am. Chem. Soc.*, 2014, **136**, 16951–16953.
- 36 C. D. Poulter, *Acc. Chem. Rev.*, 1990, **23**, 70–77.
- 37 Y. Kharel and T. Koyama, *Nat. Prod. Rep.*, 2003, **20**, 111–118.
- 38 G. Ourisson and Y. Nakatani, *Chemistry and Biology*, 1994, **1**, 11–23.
- 39 Z. Liu, J. Zhou, R. Wu and J. Xu, *J. Chem. Theory Comp.*, 2014, **10**, 5057–5067.
- 40 J. H. Kang, E. Gonzales-Vigil, Y. Matsuba, E. Pichersky and C. S. Barry, *Plant Physiol.*, 2014, **164**, 80–91.
- 41 L. C. Tarshis, J. C. Sacchettini, M. Yan and C. D. Poulter, *Biochemistry*, 1994, **33**, 10871–10877.

- 42 Y. T. Chan, T. P. Ko, S. H. Yao, Y. W. Chen, C. C. Lee and A. H. J. Wang, *ACS Omega*, 2017, **2**, 930–936.
- 43 C. Weitzel and H. T. Simonsen, *Phytochem. Rev.*, 2015, **14**, 7–24.
- 44 U. Bathe and A. Tissier, *Phytochemistry*, 2019, **161**, 149–162.
- 45 C. A. Lesburg, G. Zhai, D. E. Cane and D. W. Christianson, *Science*, 1997, **277**, 1820–1824.
- 46 C. M. Starks, K. Back, J. Chappell and J. P. Noel, *Science*, 1997, **277**, 1815–1820.
- 47 K. U. Wendt, K. Poralla and G. E. Schulz, *Science*, 1997, **277**, 1811–1815.
- 48 R. Thoma, T. Schulz-Gasen, B. D’Arcy, J. Benz, J. Aebi, H. Dehmlow, M. Hennig, M. Stihle and A. Ruf, *Nature*, 2004, **432**, 118–122.
- 49 T. A. Pemberton, M. Chen, G. G. Harris, W. K. W. Chou, L. Duan, M. Köksal, A. S. Genshaft, D. E. Cane and D. W. Christianson, *Biochemistry*, 2017, **56**, 2010–2023.
- 50 M. Köksal, H. Hu, R. M. Coates, R. J. Peters and D. W. Christianson, *Nat. Chem. Biol.*, 2011, **7**, 431–433.
- 51 K. Zhou, Y. Gao, J. A. Hoy, F. M. Mann, R. B. Honzatko and R. J. Peters, *J. Biol. Chem.*, 2012, **287**, 6840–6850.
- 52 D. W. Christianson, *Chem. Rev.*, 2006, **106**, 3412–3442.
- 53 M. Köksal, Y. Jin, R. M. Coates, R. Croteau and D. W. Christianson, *Nature*, 2011, **469**, 116–122.
- 54 D. J. Tantillo, *Nat. Prod. Rep.*, 2011, **28**, 1035–1053.
- 55 D. J. Miller and R. K. Allemann, *Nat. Prod. Rep.*, 2012, **29**, 60–71.
- 56 T. A. Pemberton and D. W. Christianson, *J. Antibiot.*, 2016, **69**, 486–493.
- 57 M. Chen, W. K. W. Chou, N. Al-Lami, J. A. Faraldos, R. K. Allemann, D. E. Cane and D. W. Christianson, *Biochemistry*, 2016, **55**, 45.
- 58 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.
- 59 D. C. Hyatt, B. Youn, Y. Zhao, B. Santhamma, R. M. Coates, R. B. Croteau and C. Kang, *Proc. Natl. Acad. Sci. U. S. A.*, 2007, **104**, 5360–5365.
- 60 P. N. Blank, G. H. Barrow, W. K. W. Chou, L. Duan, D. E. Cane and D. W. Christianson, *Biochemistry*, 2017, **56**, 5798–5811.

- 61 D. J. Grundy, M. Chen, V. González, S. Leoni, D. J. Miller, D. W. Christianson and R. K. Allemann, *Biochemistry*, 2016, **55**, 2112–2121.
- 62 J. M. Caruthers, I. Kang, M. J. Rynkiewicz, D. E. Cane and D. W. Christianson, *J. Biol. Chem.*, 2000, **275**, 25533–25539.
- 63 E. Y. Shishova, L. di Costanzo, D. E. Cane and D. W. Christianson, *Biochemistry*, 2007, **46**, 1941–1951.
- 64 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.
- 65 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.
- 66 P. Baer, P. Rabe, K. Fischer, C. A. Citron, T. A. Klapschinski, M. Groll and J. S. Dickschat, *Angew. Chem., Int. Ed. Engl.*, 2014, **53**, 7652–7656.
- 67 B. Felicetti and D. E. Cane, *J. Am. Chem. Soc.*, 2004, **126**, 7212–7221.
- 68 S. Forcat and R. K. Allemann, *Org. Biomol. Chem.*, 2006, **4**, 2563.
- 69 A. Deligeorgopoulou and R. K. Allemann, *Biochemistry*, 2003, **42**, 7741–7747.
- 70 J. A. Faraldos, A. K. Antonczak, V. González, R. Fullerton, E. M. Tippmann and R. K. Allemann, *J. Am. Chem. Soc.*, 2011, **133**, 13906–13909.
- 71 D. J. Miller, F. Yu and R. K. Allemann, *ChemBioChem*, 2007, **8**, 1819–1825.
- 72 R. K. Allemann, *Pure Appl. Chem.*, 2008, **80**, 1791–1798.
- 73 F. Yu, D. J. Miller and R. K. Allemann, *Chem. Commun.*, 2007, **4**, 2563.
- 74 D. J. Miller and R. K. Allemann, *Nat. Prod. Rep.*, 2012, **29**, 60–71.
- 75 M. Chen, N. Al-Lami, M. Janvier, E. L. D'Antonio, J. A. Faraldos, D. E. Cane, R. K. Allemann and D. W. Christianson, *Biochemistry*, 2013, **52**, 5441–5453.
- 76 J. A. Faraldos, B. Kariuki and R. K. Allemann, *J. Org. Chem.*, 2010, **75**, 1119–1125.
- 77 J. A. Faraldos and R. K. Allemann, *Org. Lett.*, 2011, **13**, 1202–1205.
- 78 J. S. Dickschat, *Eur. J. Org. Chem.*, 2017, **2017**, 4872–4882.
- 79 C. A. Citron, P. Rabe, L. Barra, C. Nakano, T. Hoshino and J. S. Dickschat, *Eur. J. Org. Chem.*, 2014, **2014**, 7684–7691.
- 80 J. A. Faraldos, D. J. Miller, V. González, Z. Yoosuf-Aly, O. Cascón, A. Li and R. K. Allemann, *J. Am. Chem. Soc.*, 2012, **134**, 5900–5908.

- 81 S. Picaud, P. Mercke, X. He, O. Sterner, M. Brodelius, D. E. Cane and P. E. Brodelius, *Arch. Biochem. Biophys.*, 2006, **448**, 150–155.
- 82 C. O. Schmidt, H. J. Bouwmeester, S. Franke and W. A. König, *Chirality*, 1999, **11**, 353–362.
- 83 D. E. Cane, P. C. Prabhakaran, J. S. Oliver and D. B. McIlwaine, *J. Am. Chem. Soc.*, 1990, **112**, 3209–3210.
- 84 D. E. Cane, P. C. Prabhakaran, E. J. Salaski, P. H. M. Harrison, H. Noguchi and B. J. Rawlings, *J. Am. Chem. Soc.*, 1989, **111**, 8914–8916.
- 85 P. L. Srivastava, A. M. Escorcía, F. Huynh, D. J. Miller, R. K. Allemann and M. W. van der Kamp, *ACS Catal.*, 2021, **11**, 1033–1041.
- 86 M. D. Leavell, D. J. McPhee and C. J. Paddon, *Curr. Opin. Biotechnol.*, 2016, **37**, 114–119.
- 87 M. A. Birkett and J. A. Pickett, *Phytochemistry*, 2003, **62**, 651–656.
- 88 D. J. Jansen and R. A. Shenvi, *Future Med. Chem.*, 2014, **6**, 1127–1148.
- 89 V. J. J. Martin, D. J. Pitera, S. T. Withers, J. D. Newman and J. D. Keasling, *Nature Biotech.*, 2003, **21**, 796–802.
- 90 P. K. Ajikumar, W. H. Xiao, K. E. J. Tyo, Y. Wang, F. Simeon, E. Leonard, O. Mucha, T. H. Phon, B. Pfeifer and G. Stephanopoulos, *Science*, 2010, **330**, 70–74.
- 91 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. Lievens, P. S. Covello, J. D. Keasling, K. K. Reiling, N. S. Renninger and J. D. Newman, *Nature*, 2013, **496**, 528–532.
- 92 World Malaria Report 2019, <https://www.who.int/malaria/publications/world-malaria-report-2019/World-Malaria-Report-2019-briefing-kit-eng.pdf>, (accessed April 2020).
- 93 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. U. S. A.*, 2012, **109**, 111–118.

- 94 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.
- 95 M. Peplow, *Nature*, 2016, **530**, 389–390.
- 96 J. D. Keasling, *Science*, 2010, **330**, 1355–1358.
- 97 J. Rinkel and J. S. Dickschat, *Beilstein J. Org. Chem.*, 2019, **15**, 1008–1019.
- 98 X. Lin and D. E. Cane, *J. Am. Chem. Soc.*, 2009, **131**, 6332–6333.
- 99 M. Demiray, X. Tang, T. Wirth, J. A. Faraldos and R. K. Allemann, *Angew. Chem., Int. Ed. Engl.*, 2017, **56**, 4347–4350.
- 100 X. Tang, M. Demiray, T. Wirth and R. K. Allemann, *Bioorg. Med. Chem.*, 2018, **26**, 1314–1319.
- 101 K. A. Rising, C. M. Crenshaw, H. J. Koo, T. Subramanian, K. A. H. Chehade, C. Starks, K. D. Allen, D. A. Andres, H. P. Spielmann, J. P. Noel and J. Chappell, *ACS Chem. Biol.*, 2015, **10**, 1729–1736.
- 102 F. Huynh, D. J. Grundy, R. L. Jenkins, D. J. Miller and R. K. Allemann, *ChemBioChem*, 2018, **19**, 1834–1838.
- 103 C. Oberhauser, V. Harms, K. Seidel, B. Schröder, K. Ekramzadeh, S. Beutel, S. Winkler, L. Lauterbach, J. S. Dickschat and A. Kirschning, *Angew. Chem., Int. Ed. Engl.*, 2018, **57**, 11802–11806.
- 104 C. Lamberth, S. Jeanmart, T. Luksch and A. Plant, *Science*, 2013, **341**, 742–746.
- 105 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 Manage. Sci.*, 2005, **61**, 1115–1121.
- 106 M. A. Birkett, S. Al Abassi, T. Kröber, K. Chamberlain, A. M. Hooper, P. M. Guerin, J. Pettersson, J. A. Pickett, R. Slade and L. J. Wadhams, *Phytochemistry*, 2008, **69**, 1710–1715.
- 107 O. Cascón, S. Touchet, D. J. Miller, V. Gonzalez, J. A. Faraldos and R. K. Allemann, *Chem. Commun.*, 2012, **48**, 9702.
- 108 S. Touchet, K. Chamberlain, C. M. Woodcock, D. J. Miller, M. A. Birkett, J. A. Pickett and R. K. Allemann, *Chem. Commun.*, 2015, **7550**, 7550–7553.
- 109 S. M. Cook, Z. R. Khan and J. A. Pickett, *Annu. Rev. Entomol.*, 2007, **52**, 375–400.
- 110 Z. Khan, C. A. O. Midega, A. Hooper and J. Pickett, *J. Chem. Ecol.*, 2016, **42**, 689–697.
- 111 V. Harms, A. Kirschning and J. S. Dickschat, *Nat. Prod. Rep.*, 2020, **37**, 1080–1097.

- 112 World Population Prospects: The 2017 Revision,
<https://www.un.org/development/desa/publications/world-population-prospects-the-2017-revision.html>, (accessed April 2020).
- 113 World Agriculture Towards 2030/2050: The 2012 Revision,
<https://www.fao.org/3/ap106e/ap106e.pdf> (accessed May 2020).
- 114 D. Tilman, C. Balzer, J. Hill and B. L. Befort, *Proc. Natl. Acad. Sci. U. S. A.*, 2011, **108**, 20260–20264.
- 115 D. P. Bebber, T. Holmes, D. Smith and S. J. Gurr, *New Phytol.*, 2014, **202**, 901–910.
- 116 F. Gould, Z. S. Brown and J. Kuzma, *Science*, 2018, **360**, 728–732.
- 117 B. Bajželj, K. S. Richards, J. M. Allwood, P. Smith, J. S. Dennis, E. Curmi and C. A. Gilligan, *Nat. Clim. Change*, 2014, **4**, 924–929.
- 118 L. E. Smart, G. I. Aradottir and T. J. A. Bruce, in *Integrated Pest Management*, Elsevier, 2014, pp. 93–109.
- 119 D. A. Nordlund and W. J. Lewis, *J. Chem. Ecol.*, 1976, **2**, 211–220.
- 120 J. A. Pickett, R. K. Allemann and M. A. Birkett, *Nat. Prod. Rep.*, 2013, **30**, 1277.
- 121 N. Dudareva, F. Negre, D. A. Nagegowda and I. Orlova, *Crit. Rev. Plant Sci.*, 2006, **25**, 417–440.
- 122 T. J. A. Bruce and J. A. Pickett, *Phytochemistry*, 2011, **72**, 1605–1611.
- 123 C. E. Reisenman, H. Lei and P. G. Guerenstein, *F. Phys.*, 2016, **7**, 1–21.
- 124 J. A. Pickett, L. J. Wadhams and C. M. Woodcock, *Agric., Ecosyst. Environ.*, 1997, **64**, 149–156.
- 125 N. Bakthavatsalam, *Ecofriendly Pest Management for Food Security*, Elsevier Inc., 2016, pp. 563–611.
- 126 P. Witzgall, P. Kirsch and A. Cork, *J. Chem. Ecol.*, 2010, **36**, 80–100.
- 127 A. M. Adio, *Tetrahedron*, 2009, **65**, 1533–1552.
- 128 State of the World's Plants 2017,
https://stateoftheworldsplants.org/2017/report/SOTWP_2017.pdf, (accessed April 2020).
- 129 D. R. Jones, *Eur. J. Plant Pathol.*, 2003, **109**, 195–219.
- 130 S. Schlaeger, J. A. Pickett and M. A. Birkett, *Pest Manage. Sci.*, 2018, **74**, 2405–2411.

- 131 A. M. Dickey, L. S. Osborne, R. G. Shatters, P. M. Hall and C. L. Mckenzie, *J. Econ. Entomol.*, 2013, **106**, 1355–1364.
- 132 Tomato | Land & Water | Food and Agriculture Organization of the United Nations | Land & Water | Food and Agriculture Organization of the United Nations, <http://www.fao.org/land-water/databases-and-software/crop-information/tomato/en/>, (accessed April 2020).
- 133 M. J. Rodríguez-López, E. Garzo, J. P. Bonani, A. Fereres, R. Fernández-Muñoz and E. Moriones, *Phytopathology*, 2011, **101**, 1191–1201.
- 134 Indiana State University, Tomatoes are Wild, https://moylelab.sitehost.iu.edu/tomato_species.html, (accessed May 2020).
- 135 R. W. J. Kortbeek, J. Xu, A. Ramirez, E. Spyropoulou, P. Diergaarde, I. Otten-Bruggeman, M. de Both, R. Nagel, A. Schmidt, R. C. Schuurink and P. M. Bleeker, *Methods Enzymol.*, 2016, **576**, 305–331.
- 136 P. M. Bleeker, P. J. Diergaarde, K. Ament, S. Schütz, B. Johne, J. Dijkink, H. Hiemstra, R. de Gelder, M. T. J. de Both, M. W. Sabelis, M. A. Haring and R. C. Schuurink, *Phytochemistry*, 2011, **72**, 68–73.
- 137 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. U. S. A.*, 2012, **109**, 20124–20129.
- 138 R. M. Coates, J. F. Denissen, J. A. Juvik and B. A. Babka, *J. Org. Chem.*, 1988, **53**, 2186–2192.
- 139 K. Besser, A. Harper, N. Welsby, I. Schauvinhold, S. Slocombe, Y. Li, R. A. Dixon and P. Broun, *Plant Physiol.*, 2009, **149**, 499–514.
- 140 F. Chen, D. Tholl, J. Bohlmann and E. Pichersky, *The Plant Journal*, 2011, **66**, 212–229.
- 141 E. Gonzales-Vigil, D. E. Hufnagel, J. Kim, R. L. Last and C. S. Barry, *The Plant Journal*, 2012, **71**, 921–935.
- 142 M. Gutensohn, T. T. H. Nguyen, R. D. McMahon, I. Kaplan, E. Pichersky and N. Dudareva, *Metab. Eng.*, 2014, **24**, 107–116.
- 143 V. Falara, T. A. Akhtar, T. T. H. Nguyen, E. A. Spyropoulou, P. M. Bleeker, I. Schauvinhold, Y. Matsuba, M. E. Bonini, A. L. Schillmiller, R. L. Last, R. C. Schuurink and E. Pichersky, *Plant Physiol.*, 2011, **157**, 770–789.
- 144 F. Zhou and E. Pichersky, *New Phytol.*, 2020, **226**, 1341–1360.

- 145 T. A. Akhtar, Y. Matsuba, I. Schauvinhold, G. Yu, H. A. Lees, S. E. Klein and E. Pichersky, *The Plant Journal*, 2013, **73**, 640–652.
- 146 Y. Matsuba, J. Zi, A. D. Jones, R. J. Peters and E. Pichersky, *PLoS ONE*, 2015, **10**, 1–16.
- 147 K. C. Nicolaou, S. Sanchini, T. R. Wu and D. Sarlah, *Chem. - Eur. J.*, 2010, **16**, 7678–7682.
- 148 K. Spielmann, R. M. de Figueiredo and J. M. Campagne, *J. Org. Chem.*, 2017, **82**, 4737–4743.
- 149 L. Wu, J. C. Zhong, S. K. Liu, F. P. Liu, Z. D. Gao, M. Wang and Q. H. Bian, *Tetrahedron Asymmetry*, 2016, **27**, 78–83.
- 150 T. Kato, M. Suzuki, T. Kobayashi and B. P. Moore, *J. Org. Chem.*, 1980, **45**, 1126–1130.
- 151 J. S. Yu, T. S. Kleckley and D. F. Wiemer, *Org. Lett.*, 2005, **7**, 4803–4806.
- 152 F. Messik and M. Oberthür, *Synthesis*, 2013, **45**, 167–170.
- 153 C. Jones, PhD thesis, Cardiff University, 2018.
- 154 F. Cramer, W. Rittersdorf and W. Böhm, *Justus Liebigs Ann. Chem.*, 1962, **654**, 180–188.
- 155 V. J. Davisson, A. B. Woodside, T. R. Neal, K. E. Stremmer, M. Muehlbacher and C. D. Poulter, *J. Org. Chem.*, 1986, **51**, 4768–4779.
- 156 *DE Pat.*, DE102004033306 A1, 2004.
- 157 F. Huynh, PhD thesis, Cardiff University, 2020.
- 158 R. K. Keller and R. Thompson, *J. Chromatogr. A*, 1993, **645**, 161–167.
- 159 L. K. Henry, M. Gutensohn, S. T. Thomas, J. P. Noel and N. Dudareva, *Proc. Natl. Acad. Sci. U. S. A.*, 2015, **112**, 10050–10055.
- 160 P. H. Wang, A. N. Khusnutdinova, F. Luo, J. Xiao, K. Nemr, R. Flick, G. Brown, R. Mahadevan, E. A. Edwards and A. F. Yakunin, *Cell Chem. Biol.*, 2018, **25**, 560–570.
- 161 L. A. Johnson, A. Dunbabin, J. C. R. Benton, R. J. Mart and R. K. Allemann, *Angew. Chem., Int. Ed. Engl.*, 2020, **59**, 8486–8490.
- 162 BL21-AI™ Competent Cells,
https://tools.thermofisher.com/content/sfs/brochures/712_021252_BL21AI_bro.pdf,
ed. Invitrogen, (accessed June 2020)

- 163 Benchling, <https://www.benchling.com>, (accessed October 2017).
- 164 O. Ernst and T. Zor, *J. Visualized Exp.*, 2010, 1–6.
- 165 A. L. Demain and P. Vaishnav, *Biotechnol. Adv.*, 2009, **27**, 297–306.
- 166 S. Costa, A. Almeida, A. Castro and L. Domingues, *F. Microbiol.*, 2014, **5**, 1–20.
- 167 T. R. Butt, S. C. Edavettal, J. P. Hall and M. R. Mattern, *Protein Expression Purif.*, 2005, **43**, 1–9.
- 168 C. L. Young, Z. T. Britton and A. S. Robinson, *Biotechnol. J.*, 2012, **7**, 620–634.
- 169 R. B. Kapust and D. S. Waugh, *Protein Science*, 1999, **8**, 1668–1674.
- 170 W. J. Bao, Y. G. Gao, Y. G. Chang, T. Y. Zhang, X. J. Lin, X. Z. Yan and H. Y. Hu, *Protein Expression Purif.*, 2006, **47**, 599–606.
- 171 C41/C43 - OverExpress® Chemically Competent Cells, <https://www.lucigen.com/docs/manuals/MA031-OverExpress-Chemicallycompetent-Cells.pdf>, (accessed October 2018).
- 172 F. Meyer and A. G. Harrison, *J. Am. Chem. Soc.*, 1964, **86**, 4757–4761.
- 173 D. C. Breeden and R. M. Coates, *Tetrahedron*, 1994, **50**, 11123–11132.
- 174 F. W. McLafferty and F. Turecek, *Interpretation of Mass Spectra*, University Science Books, Mill Valley, California, 3rd ed., 1982.
- 175 D. J. Tantillo, *Chem. Soc. Rev.*, 2010, **39**, 2847–2854.
- 176 A. A. Scholte and J. C. Vederas, *Org. Biomol. Chem.*, 2006, **4**, 730–742.
- 177 D. E. Cane, J. S. Oliver, P. M. Harrison, C. Abell, B. R. Hubbard, C. T. Kane and R. Lattman, *J. Am. Chem. Soc.*, 1990, **112**, 4513–4524.
- 178 S. S. Shinde, A. Minami, Z. Chen, T. Tokiwano, T. Toyomasu, N. Kato, T. Sassa and H. Oikawa, *J. Antibiot.*, 2017, **70**, 632–638.
- 179 M. S. Said, G. R. Navale, J. M. Gajbhiye and S. S. Shinde, *RSC Adv.*, 2019, **9**, 28258–28261.
- 180 T. A. Ronnebaum, S. M. Gardner and D. W. Christianson, *Biochemistry*, 2020, **59**, 4744–4754.
- 181 C. Ignea, M. H. Raadam, M. S. Motawia, A. M. Makris, C. E. Vickers and S. C. Kampranis, *Nat. Comm.*, 2019, **10**, 1–15.

- 182 A. Waterhouse, M. Bertoni, S. Bienert, G. Studer, G. Tauriello, R. Gumienny, F. T. Heer, T. A. P. de Beer, C. Rempfer, L. Bordoli, R. Lepore and T. Schwede, *Nuc. Acids Res.*, 2018, **46**, W296–W303.
- 183 Schrödinger LLC, *The PyMOL Molecular Graphics System*, Schrödinger, LLC, 2021.
- 184 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.
- 185 P. Baer, P. Rabe, K. Fischer, C. A. Citron, T. A. Klapschinski, M. Groll and J. S. Dickschat, *Angew. Chem., Int. Ed. Engl.*, 2014, **53**, 7652–6.
- 186 Y. Yoshikuni, V. J. J. Martin, T. E. Ferrin and J. D. Keasling, *Chem. Biol.*, 2006, **13**, 91–98.
- 187 H. Liu and J. H. Naismith, *BMC Biotechnol.*, 2008, **8**, 91.
- 188 M. Loizzi, V. González, D. J. Miller and R. K. Allemann, *ChemBioChem*, 2018, **19**, 100–105.
- 189 D. E. Cane and M. Tandon, *Tet. Lett.*, 1994, **35**, 5355–5358.
- 190 F. Madeira, Y. M. Park, J. Lee, N. Buso, T. Gur, N. Madhusoodanan, P. Basutkar, A. R. N. Tivey, S. C. Potter, R. D. Finn and R. Lopez, *Nuc. Acids Res.*, 2019, **47**, 636–641.
- 191 D. E. Cane and Y. S. Tsantrizos, *J. Am. Chem. Soc.*, 1996, **118**, 10037–10040.
- 192 J. Andersen-Ranberg, K. T. Kongstad, M. T. Nielsen, N. B. Jensen, I. Pateraki, S. S. Bach, B. Hamberger, P. Zerbe, D. Staerk, J. Bohlmann, B. L. Møller and B. Hamberger, *Angew. Chem., Int. Ed. Engl.*, 2016, **55**, 2142–2146.
- 193 A. Hou, L. Lauterbach and J. S. Dickschat, *Chem. - Eur. J.*, 2020, **26**, 2178–2182.
- 194 C. Ignea, M. Pontini, M. S. Motawia, M. E. Maffei, A. M. Makris and S. C. Kampranis, *Nat. Chem. Biol.*, 2018, **14**, 1090–1098.
- 195 L. Drummond, M. J. Kschowak, J. Breitenbach, H. Wolff, Y. M. Shi, J. Schrader, H. B. Bode, G. Sandmann and M. Buchhaupt, *ACS Synth. Biol.*, 2019, **8**, 1303–1313.
- 196 C. B. Eiben, T. de Rond, C. Bloszies, J. Gin, J. Chiniquy, E. E. K. Baidoo, C. J. Petzold, N. J. Hillson, O. Fiehn and J. D. Keasling, *ACS Synth. Biol.*, 2019, **8**, 2238–2247.
- 197 M. H. Beale, M. A. Birkett, T. J. A. Bruce, K. Chamberlain, L. M. Field, A. K. Huttly, J. L. Martin, R. Parker, A. L. Phillips, J. A. Pickett, I. M. Prosser, P. R. Shewry, L. E. Smart, L. J. Wadhams, C. M. Woodcock and Y. Zhang, *Proc. Natl. Acad. Sci. U. S. A.*, 2006, **103**, 10509–10513.

- 198 S. M. S. Fernandez, B. A. Kellogg and C. D. Poulter, *Biochemistry*, 2000, **39**, 15316–15321.
- 199 T. Koyama, S. Obata, M. Osabe, A. Takeshita, K. Yokoyama, M. Uchida, T. Nishino and K. Ogura, *J. Biochem.*, 1993, **113**, 355–363.
- 200 V. I. Babushok, P. J. Linstrom and I. G. Zenkevich, *J. Phys. Chem. Ref. Data*, 2011, **40**, 1–47.
- 201 S. Zabel, W. Brandt, A. Porzel, B. Athmer, S. Bennewitz, P. Schäfer, R. Kortbeek, P. Bleeker and A. Tissier, *The Plant Journal*, 2021, **105**, 1309–1325.
- 202 R. L. Blackman and V. F. Eastop, *Ahipds on the World's Crops*, Wiley-Interscience, Avon, 1st ed., 1985.
- 203 H. van Emden and R. Harrington, *Aphids as Crop Pests*, CABI, Wallingford, 2007.
- 204 S. Vandermoten, M. C. Mescher, F. Francis, E. Haubruge and F. J. Verheggen, *Insect Biochem. Mol. Biol.*, 2012, **42**, 155–163.
- 205 W. S. Bowers, L. R. Nault, R. E. Webb and S. R. Dutky, *Science*, 1972, **177**, 1121–1122.
- 206 S. Al Abassi, M. A. Birkett, J. Pettersson, J. A. Pickett, L. J. Wadhams and C. M. Woodcock, *J. Chem. Ecol.*, 2000, **26**, 1765–1771.
- 207 M. H. Beale, M. A. Birkett, T. J. A. Bruce, K. Chamberlain, L. M. Field, A. K. Huttly, J. L. Martin, R. Parker, A. L. Phillips, J. A. Pickett, I. M. Prosser, P. R. Shewry, L. E. Smart, L. J. Wadhams, C. M. Woodcock and Y. Zhang, *Proc. Natl. Acad. Sci. U. S. A.*, 2006, **103**, 10509–10513.
- 208 D. C. Griffiths and J. A. Pickett, *Entomol. Exp. Appl.*, 1980, **27**, 199–201.
- 209 A. Cork, *A Pheromone Manual*, Natural Resources Institute, Chatham Maritime, 2004.
- 210 M. A. Birkett, C. A. M. Campbell, K. Chamberlain, E. Guerrieri, A. J. Hick, J. L. Martin, M. Matthes, J. A. Napier, J. Pettersson, J. A. Pickett, G. M. Poppy, E. M. Pow, B. J. Pye, L. E. Smart, G. H. Wadhams, L. J. Wadhams and C. M. Woodcock, *Proc. Natl. Acad. Sci. U. S. A.*, 2000, **97**, 9329–9334.
- 211 K. Yoshihara, Y. Ohta, T. Sakai and Y. Hirose, *Tet. Lett.*, 1969, **10**, 2263–2264.
- 212 N. Bülow and W. A. König, *Phytochemistry*, 2000, **55**, 141–168.
- 213 A. F. Scott, L. Y. P. Luk, I. Tuñón, V. Moliner and R. K. Allemann, *ChemBioChem*, 2019, **20**, 2807–2812.
- 214 Y. W. Zhang, Y. Kharel and T. Koyama, *J. Mol. Catal. B: Enzym.*, 2000, **10**, 623–630.

- 215 A. Li, PhD thesis, Cardiff University, 2011.
- 216 C. Engler, R. Kandzia and S. Marillonnet, *PLoS ONE*, 2008, **3**, 3647–3653.
- 217 New England Biolabs, KLD Enzyme Mix,
<https://international.neb.com/products/m0554-kld-enzyme-mix#>, (accessed September 2021).
- 218 Z. T. Du, S. Zheng, G. Chen and D. Lv, *Molecules*, 2011, **16**, 8053–8061.

